UNIVERSITY OF MISKOLC

Faculty of Earth and Environmental Sciences and Engineering
Institute of Exploration Geosciences

WELL LOGGING METHOD DEVELOPMENT IN UNCONVENTIONAL HYDROCARBON RESERVOIRS

PhD THESIS

by

Rafael Valadez Vergara

Scientific supervisor:

Prof. Dr. Norbert Péter Szabó

MIKOVINY SÁMUEL DOCTORAL SCHOOL OF EARTH SCIENCES
Head of the Doctoral School: Prof. Dr. Péter Szűcs

Miskolc, 2025

Hungary

SUPERVISOR'S FOREWORD

for the PhD thesis

"WELL LOGGING METHOD DEVELOPMENT IN UNCONVENTIONAL HYDROCARBON RESERVOIRS"

by Rafael Valadez Vergara

The topic of the Candidate's thesis – evaluation of unconventional hydrocarbon formations – is in the focus of international research. The method developments introduced by the Candidate in the thesis belong to the range of modern data processing tools applied to in situ geophysical data. The novel exploratory statistical and inversion techniques allow the determination of such parameters that could not be extracted directly from well logs. Unconventional hydrocarbon reservoirs such as organic-rich shale and tight gas formations are multi-mineral rocks, the matrix volumes, porosity, saturation, and kerogen content of which are delicate parameters to be estimated as accurate as possible for reserve estimations. In the PhD thesis, a new approach is suggested to estimate the total organic carbon content using factor analysis of oilfield wireline logs. In addition to this independent well-log-analysis method, the old ones are also reformed by using the Simulated Annealing method that finds the global optimum of the objective function by minimizing the misfit between the measurements and predictions. One of the most valuable parts of the PhD thesis is the estimation of the so-called level of maturity, which is a zone parameter normally defined arbitrarily based on geological considerations. This is the first report on the automated and objective estimation of this key parameter using geophysical well logs. The study shows how this quantity can be optimally determined to reliably determine the sweet spot intervals and the total organic matter content there.

The interval inversion method was originally developed at the Department of Geophysics of the University of Miskolc. The Candidate followed this research idea and further developed it to evaluate multimineral hydrocarbon formations. In the framework of his inversion methodology, one can estimate the matrix volumes of several mineral types making significant improvement in reservoir modeling and the accuracy of inversion estimates. Such many matrix types and volumetric parameters (e.g., 14 unknowns) cannot be treated in a local inversion process traditionally used in the oil industry. This local inverse problem is underdetermined even by using all types of the observed well logs as input of

inversion. By the interval inversion of wireline logs, it is proven that all the unknowns can be estimated simultaneously in a highly accurate and reliable way. The automated estimation of these matrix volumes makes a significant improvement in reservoir modeling and the accuracy of inversion estimations. This inversion strategy can be used as a powerful tool in petroleum exploration and several other fields of geoscience, too. The extraction of these parameters is not only a scientific innovation, but also it can reduce the financial cost of core sampling operations and lab measurements in practice.

The Candidate's continuous efforts towards scientific research, his creativity, and the results presented in this thesis prove the scientific knowledge and the suitability of the Candidate for independent research. In my opinion, the Candidate's results, especially those related to the estimation of total organic content and reservoir mineralogy based on geophysical inversion tools, are worth publishing in further ranked international journals. The feasibility of the developed inversion method has been tested in various hydrocarbon fields around the world and verified also by core lab measurements. Herewith, I certify that the Candidate's PhD dissertation contains only valid data, and the presented results represent the Candidate's own work. In my opinion, the PhD thesis is fully adequate in scope and quality required by the Mikoviny Sámuel Doctoral School of Earth Sciences. Based on all the above, I recommend the public defense to be carried out to complete a successful process of acquiring a PhD title.

19 November 2024, Miskolc

Prof. Dr. Norbert Péter Szabó

supervisor

TABLE OF CONTENTS

INTRODUCTION

Unconventional reservoirs have appeared as a significant aspect of the global energy landscape, presenting extensive opportunities for resource extraction and reshaping energy markets. Notably abundant in regions such as the United States of America and Canada, these reservoirs have produced deep economic and environmental changes. With daily production of natural gas and oil reaching unprecedented levels, they have facilitated a dramatic departure from coal, leading to reduced CO_2 emissions and enhanced air quality (Kohli and Zoback, 2019a; Rezaee, 2022).

The unconventional oil and gas sector constitutes a fundamental element within the contemporary energy scene, marked by operational complexities and technological complexities. Recent debate in this area has put in the picture various aspects of the industry, reaching from empirical data and academic analyses to reveal key insights. While North America has historically been at the vanguard of unconventional development, recent years have witnessed a growing interest in regions like China and Argentina (Kohli and Zoback, 2019a). The global potential for unconventional reservoirs extends far beyond current production hubs, with numerous countries poised to explore and exploit these resources to satisfy their energy needs.

The exploration and extraction of oil and gas resources demand a refinement comprehension of the geological complexities inherent within formations and basins. A fundamental aspect of this understanding lies in recognizing the layering within formations. Unlike traditional reservoirs, unconventional reservoirs possess distinct characteristics such as high organic content, low permeability, and intricate layering, which present both challenges and opportunities for extraction (Peters et al., 2016; Zee Ma, 2016; Kohli and Zoback, 2019a).

Organic-rich source rocks take a key function in the generation and accumulation of hydrocarbons within sedimentary basins. These source rocks are distributed all over the world (Dong et al., 2016), historically serving as migration pathways for hydrocarbons into conventional reservoirs. However, recent advancements in exploration and extraction techniques have shown the potential for directly targeting these source rocks for unconventional hydrocarbon development.

One prominent example highlighted is the Barnett Formation, located in the Fort Worth Basin (USA), which serves as a paradigmatic case of an organic-rich source rock. Within this formation, the organic material has undergone maturation, predominantly yielding

gas. Despite its traditionally observed role as a source rock rather than a reservoir rock, improvements in extraction techniques, particularly hydraulic fracturing, have enabled economic production from such formations. Conversely, the Bakken Formation exemplifies a different aspect of unconventional resources, representing low-permeability reservoirs where extraction primarily relies on stimulating production through hydraulic fracturing. Additionally, there are layered plays characterized by fracking across various strata, encompassing both organic-rich and conventional reservoirs with low permeability (Peters et al., 2016; Kohli and Zoback, 2019a).

Furthermore, the compositional variability of rocks within formations adds another level of complexity to reservoir characterization. From carbonate-rich to clay-rich compositions, the diverse nature of rock compositions within a single well requires adaptable extraction strategies adapted to the specific geological context (Zee Ma, 2016). For instance, variations in mineralogical composition and maturity across different zones within a formation requires precise well placement and targeting. Thus, identifying zones with optimal mineralogical characteristics and maturity levels is crucial for maximizing resource recovery while minimizing operational costs and environmental impact.

Correlating maturation indicators with hydrocarbon production provides valuable insights for exploration and development activities. In regions such as the Fort Worth Basin, the alignment between indicators like vitrinite reflectance and actual gas production validates the reliability of these metrics in predicting hydrocarbon potential (Kohli and Zoback, 2019b). For this, various indicators, such as vitrinite reflectance, serve as proxy for assessing the maturity level of source rocks, where higher vitrinite reflectance values correspond to increased maturity, with specific thresholds indicating transitions in hydrocarbon composition—from oil to gas, and ultimately to inert carbon. Additionally, the type of gas produced, whether thermogenic or biogenic, is influenced by depth and temperature regime.

Geological layering holds particular significance in petroleum engineering, especially concerning hydraulic fracturing operations (Kohli and Zoback, 2019b). Geological formations like the Eagle Ford and Barnett exhibit distinct layering patterns that profoundly influence rock composition and properties. Variations in Total Organic Carbon (TOC) content within these layers significantly impact the efficacy of hydraulic fracturing activities. Zones characterized by elevated TOC levels introduce complexities in terms of rock porosity, permeability, and mechanical properties (elastic moduli), posing

both challenges and opportunities for hydraulic fracturing operations (Zee Ma, 2016; Kohli and Zoback, 2019b).

Hence, in drilling operations, understanding the mineralogical content and maturity of formations is indispensable for optimizing extraction strategies. Also, a comprehensive modeling approach is imperative to reveal stress variations, TOC fluctuations, and their impact on fracture behavior. Integrating geological, geomechanical, and engineering knowledge is essential for developing fracturing strategies designed to the specific characteristics of layered formations (Kohli and Zoback, 2019b).

Therefore, in my doctoral thesis, the primary objectives revolved around enhancing our understanding of reservoir properties and applying advanced petrophysical techniques to address challenges in petroleum geology, with a specific focus on TOC estimation, factor analysis of wireline logs, interval inversion methodology, and thermal maturity assessment in unconventional hydrocarbon reservoirs.

Moreover, this research delved into the complexities of deep and multi-mineral formations, where the number of applied logging tools is often limited, and more unknowns are explored. These intricate geological settings introduce additional noise and uncertainties into the inversion procedures, necessitating robust and adaptive formation evaluation techniques. Hence, this study addressed these challenges, enabling more reliable and precise characterization of reservoir properties, even in the face of limited data and increased geological complexities.

1. STATE OF THE ART OF DEVELOPED METHODS

1.1. Traditional methodologies for estimation of Total Organic Carbon (TOC) Content

*1.1.1. The Passey method (**ΔlogR**) and level of organic metamorphism (**LOM**)*

A critical indicator of hydrocarbon resources to be studied is TOC, which is very important when evaluating potential organic shales-gas reservoir (Zee Ma, 2016). TOC expresses the amount of organic carbon present in the formation, and it has been shown the direct relationship with porosity and gas saturation (Passey et al., 2010). Diverse well-log interpretation methods have been developed to assess an accurate TOC estimation (Bibor

and Szabó, 2016; Passey et al., 1990; Schmoker, 1979; Zhao et al., 2016; Szabó and Dobróka, 2020).

One widely used method was proposed by Passey et al. (1990). This method is used in shale formation evaluation, since it combines the porosity indicator logs (i.e. sonic, density and/or neutron logs) with resistivity log data and takes maturation into consideration (Jiang, et al., 2018).

Passey et al. (1990) observed that a linear relationship exists between the distance separation of the sonic travel time (Δt) and formation (true) resistivity (R) log curves when both are overlaid on the same track and total organic carbon content, making evident in organic-rich intervals. The distance separation of both curves is also known as delta log R distance ($\Delta\log R$). Passey suggests a mathematical expression Eq. (1.1) to estimate TOC as function of $\Delta\log R$ distance, which is also a function of level of organic metamorphism or maturity (LOM)

$$\mathrm{TOC} = (\Delta\log R) \times 10^{(2.297 - 0.1688 \cdot LOM)} \tag{1.1}$$

The $\Delta \log R$ is calculated by the following equation

$$\Delta\log R = \log_{10}(R/R_{baseline}) + 0.02 \cdot (\Delta t - \Delta t_{baseline}) \tag{1.2}$$

where $R_{baseline}$ is the resistivity corresponding to the $\Delta t_{baseline}$ value when the curves are running at the baseline in fine grained non-source, clay-rich rocks.

Several popular porosity curves (such as bulk density and neutron porosity) have demonstrated excellent results for TOC estimation in addition to the sonic curve. Particularly useful for confirming the $\Delta\log R$ separation found with the sonic/resistivity pair. Nonetheless, sonic/resistivity has produced more accurate findings than density/resistivity or neutron/resistivity, which might be ascribed to the negative influence of hole conditions on density and neutron logging (Passey et al., 1990).

After the accumulation and preservation of organic-rich sediments stage, for hydrocarbon generation; diagenesis, catagenesis and metagenesis processes should take place through millions of years by which the organic matter will go under thermal degradation during burial that increase temperature, leading to the formation of petroleum molecules. The "Level of

Organic Metamorphism" (LOM) describes how far the thermal metamorphism of sedimented organic matter has progressed during subsurface burial (Hood et al., 1975).

There are numerous reported scales that show organic metamorphism level (Figure 1.1). Vitrinite reflectance (*Ro*) is probably a commonly used thermal maturity indicator, thus it is directly applicable to the study of temperature histories of petroleum source rocks, and therefore, there is a relationship between LOM and *Ro*, since vitrinite reflectance is easily obtainable at laboratory studies, it is considered to be the key factor for estimating the level of thermal maturity of organic matter (Khan and Bibi, 2016).

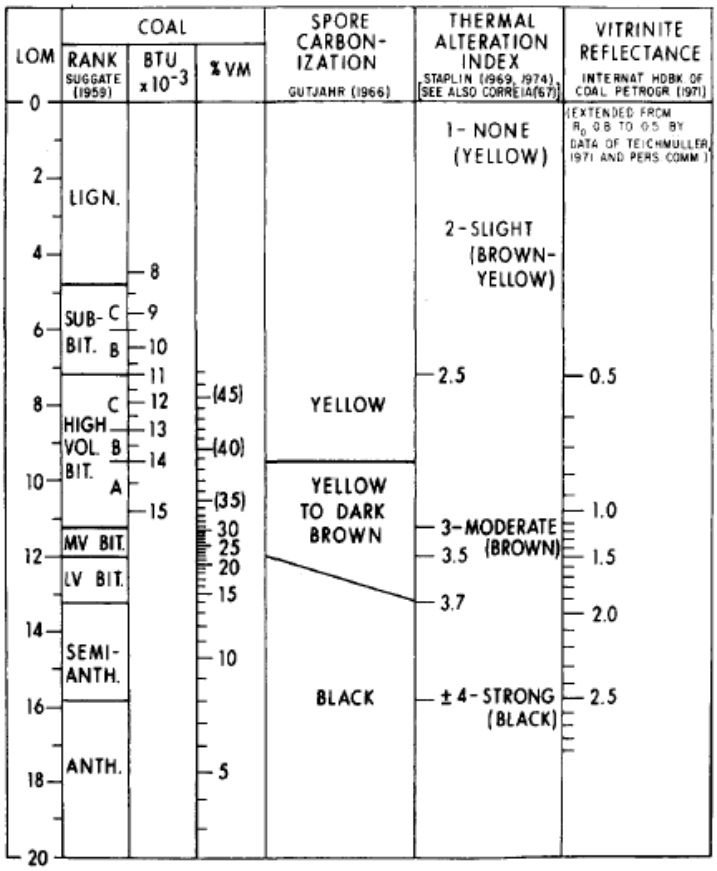

Figure 1.1 Some scales of organic metamorphism that can be converted to LOM values (Hood et al., 1975)

Correct estimation of LOM is important not only because it indicates the current and past maturity level of the formation, but also it is a key tool for estimating the depths (i.e. generation windows) at which oil and gas are produced from the kerogens of probable petroleum source rocks (Kentucky Geological Survey, 2023). Thermal maturity is an

important factor in determining potentially promising shale oil and gas accumulations during the early stages of exploration.

Also, as it can be noted from Eq. (1.1), for the estimation of TOC using $\Delta \log R$ method, a correct estimation of LOM is needed, since if the maturity is incorrect, the absolute TOC values will be mistakenly estimate, however, the vertical variability in TOC will be correctly represented.

1.1.2. The Clay Indicator method (I_{cl})

Zhao et al. (2016) introduced a novel approach to estimate Total Organic Carbon (TOC) using the natural gamma ray log and porosity logs, specifically the neutron porosity and bulk density logs. They named this method the clay indicator method, which utilizes the physical relationship and association between the apparent neutron porosity, ϕ_{Na}, and density porosity, ϕ_{Da}. This correlation was previously identified by Mao (2001) and is referred to by Zhao et al. (2016). The clay indicator (I_{cl}) is computed based on the following physical relationship

$$I_{cl} = \phi_{Na} - \phi_{Da}. \tag{1.3}$$

The clay indicator index is derived from the disparity between the apparent neutron porosity and density porosity. It has been observed that the clay indicator can serve as a means to estimate the clay content in source rocks. This is particularly useful in cases where gas presence is minimal, as the neutron log may be influenced by the excavation effect.

Furthermore, it has been noted that the clay indicator exhibits comparable behavior to natural gamma ray log in non-source rocks. However, the disparity between the gamma ray log and the clay indicator index is more prominent in source rock intervals compared to non-source intervals. As a result, when displayed together and appropriately scaled, the gamma ray log and the clay indicator can effectively differentiate intervals of source rock from those with lower organic content.

The separation between the two curves is expressed as follows

$$\Delta d = GR' - I'_{cl} \tag{1.4}$$

where

$$GR' = \frac{GR - GR_{left}}{GR_{right} - GR_{left}}, \tag{1.5}$$

and

$$I'_{cl} = \frac{I_{cl} - I_{cl_{left}}}{I_{cl_{right}} - I_{cl_{left}}}. \tag{1.6}$$

where GR is the observed log value in API units, GR_{left} and GR_{right} are the left and right scale limits of the GR curve in API, $I_{cl_{left}}$ and $I_{cl_{right}}$ are the left and right scale limits of the clay indicator curve.

According to Zhao et al. (2016) observations there is strong correlation between Δd and the kerogen content – the separation increases as the kerogen increases. Therefore, a linear relationship has been stablished to estimate TOC

$$\text{TOC} = \alpha \Delta d + \beta, \tag{1.7}$$

where α and β are the slope and intercept of the fitted straight line. This method has been proposed to be used where the Delta Log R method (Passey et al., 1990) has presented failure, because of abnormally deep resistivity values has been read. Also, it has the advantage that the natural gamma ray and clay indicator curves overlie in non-source reservoirs with oil or water. Moreover, this method is applicable to formation that present excess radioactivity source rocks with little or no potassium-feldspar.

*1.1.3. The Bibor-Szabó method (**I_{cl}-BS**)*

Bibor and Szabó (2016) aimed to enhance the regression relation between the total organic content and the parameter Δd, which is estimated using Zhao's method. One crucial parameter in this context is the level of maturity, which is assumed to be constant throughout the process. To address this, they proposed estimating the optimal value of the maturity constant by minimizing the discrepancy between the observed well logs and the theoretical logs calculated using the Passey's model. This estimation was performed through an

automated optimization procedure. Subsequently, they compared the outcomes of the improved well-log-analysis methods with the measurements obtained from core samples.

The presented study focuses on enhancing the Δd method and its application to the same well-logging dataset initially utilized by Zhao et al. (2016). Their research aimed to investigate the relationship between Total Organic Carbon (TOC) measurements obtained from core samples and the parameter Δd, estimated through Eq. (1.7). By plotting the associated values of TOC and Δd, it becomes evident that the relationship is characterized by a nonlinear nature, contrary to the initial suggestion by Zhao et al. (2016). To achieve a more precise interpretation, particularly when dealing with significant Δd values, they proposed the adoption of a regression model incorporating site-specific regression coefficients denoted as $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$, and $\hat{\eta}$

$$\mathrm{TOC} = \hat{\alpha}\left(1 - \hat{\beta}e^{-\hat{\gamma}\Delta d\,\hat{\eta}}\right). \tag{1.8}$$

The regression coefficients are determined at the respective well site. The optimal values of the regression coefficients were estimated through the minimization of the discrepancy between the observed TOC values and those calculated using Eq. (1.8) by employing the Marquardt (1959) algorithm for solving a least-squares inverse problem.

1.2. Optimization and statistical data analysis

1.2.1. Linear regression

Linear regression is a widely used statistical method in scientific research to model the relationship between a dependent variable and one or more independent variables. The basic assumption of linear regression is that the relationship between the dependent variable y and the independent variable(s) x can be described by the linear function $y = \beta_0 + \beta_1 x + e$, where β_0 represents the intercept, β_1 the slope of the line, and e is a residual term, which captures the variability in y that cannot be explained by the linear relationship (Montgomery et al., 2012).

The primary objective of linear regression is to estimate the parameters β_0 and, β_1 in a way that minimizes the sum of the squared differences between the observed and predicted

values of the dependent variable. This method, known as Ordinary Least Squares (OLS), ensures that the estimated regression line best fits the data (Freedman, 2009).

For linear regression to produce reliable results, several key assumptions must be met:

1) Linearity: The relationship between independent and dependent variable is linear.
2) Independence of errors: Observations are independent of each other.
3) Homoscedasticity: The variance of the error terms is constant across all levels of the independent variable.
4) Normality of errors: The error terms are normally distributed.

Violations of these assumptions can lead to biased or insufficient estimates, undermining the validity of the regression model's inferences. Researchers can use various diagnostic tools, such as residual plots and Q-Q plots, to check these assumptions (Kutner et al., 2004).

Interpretation of linear regression results typically involves examining the coefficient of determination (R^2), which indicates the proportion of variance in the dependent variable explained by the independent variable(s), as well as the statistical significance of the coefficients, often assessed through p-values and confidence intervals. (R represents the Pearson's correlation coefficient.)

It is important to note that while linear regression is a powerful tool, it may not be appropriate for all types of data or research questions. Non-linear relationships, categorical outcomes, or time series data may require alternative methods such as polynomial regression, logistic regression, or time series analysis, respectively.

1.2.2. Simulated Annealing method

Simulated Annealing (SA) method is probabilistic global optimization method proposed by Metropolis et al. (1953) to model the thermal equilibrium state of solids. This method works based on the idea behind real physical annealing of metals, in which a crystalline solid is heated and then allowed to cool very slowly until it achieves its minimum lattice energy state, and therefore, its most regular crystal lattice configuration. The ultimate combination produces a solid with such improved structural integrity if the cooling schedule is slow enough. The relationship between this kind of thermodynamic behavior and finding the global minimum for a discrete optimization problem is established by SA method (Henderson et al., 2006).

In general, the algorithm of the SA method generates two solutions and evaluates them by means of an objective function (energy function) - the initial solution and a freshly chosen

solution—are compared at each iteration. Improving solutions are always accepted, whereas a subset of inferior solutions is allowed with the goal of escaping local optima and reaching global optima. Temperature as control parameter, which is normally decreasing with each algorithmic iteration, determines the likelihood of adopting a non-improving solution (Henderson et al., 2006).

In this study, a Metropolis scheme of SA was used (Metropolis et al., 1953). At each iteration, the Metropolis SA algorithm adjusts the components of the relevant model parameter vector (**m**), as follows

$$m_j^{(new)} = m_j^{(old)} + b \tag{1.9}$$

where b specifies an actual perturbation term, which is usually a really small value between $[b, b_{max}]$ where b_{max} is often reduced by

$$b_{max}^{(new)} = b_{max}^{(old)} \cdot \epsilon \tag{1.10}$$

and ϵ is known as the decrement factor $(0 \leq \epsilon \leq 1)$.

At each iteration step of the random walk along the parameter space, the energy function or cost function (E) of the relevant model is calculated and compared to the preceding one

$$\Delta E = E(\mathbf{m}^{\text{new}}) - E(\mathbf{m}^{\text{old}}). \tag{1.11}$$

The new model's acceptance probability (P) is determined by the Metropolis criteria

$$P(\Delta E, T) = \begin{cases} 1 & , \text{if } \Delta E \leq 0 \\ \exp\left(-\dfrac{\Delta E}{T}\right), & \text{if } \Delta E > 0 \end{cases} \tag{1.12}$$

where T represents the temperature as important control parameter, and if its value in the current step is lower than it was in the previous one, then the model is approved. If the energy of the new model has been increased, the chance of acceptance would also rely on the amount of energy required to escape from the local minimum. The new model is approved if the following criterion $P(\Delta E) > \alpha_{SA}$, is met; if not, it is rejected (α_{SA} is a random number selected with uniform probability from [0, 1]).

The cooling schedule, which should not be either too quick or too slow, greatly affects the pace of convergence. While slow cooling causes the process to slow down and may not produce a better outcome, fast cooling may lead the process to become trapped in a local minimum.

1.2.3. The Factor Analysis method

Factor analysis is a widely used approach to reduce the dimensionality of multivariate datasets and analyze the interrelationships among a set of measured variables. It encompasses various techniques (Jöreskog et al., 2016) each with distinct objectives and underlying mathematical models. However, they all share a common feature, which is the creation of a small number of hypothetical variables, referred to as factors. These factors are constructed in a manner that captures the essential information contained within a larger set of observed variables or objects, thereby reducing the overall complexity of the data by exploiting inherent interdependencies.

The purpose of factor analysis is to condense a large number of variables into a more manageable set of components. By extracting the largest shared variance from all the variables, factor analysis consolidates this information into a more compacted score, known as factor scores. This score can then be utilized as an index for further investigation of the variables. Factor analysis is an integral part of the General Linear Model (GLM) and operates under several assumptions, including the presence of a linear relationship, the absence of multicollinearity, the inclusion of relevant variables in the analysis, and the existence of genuine correlations between variables and factors (Jöreskog et al., 2016).

Overall, factor analysis serves as a multivariate technique for data reduction, as a small number of factors can effectively convey the same amount of information as a much larger set of original observations. The development of the factor model in relation to the original set of observations, contain on the data matrix is done by assuming the following model

$$\mathbf{X}_{(N\times P)} = \mathbf{F}_{(N\times k_f)}\mathbf{A}^{\mathrm{T}}_{(k_f\times P)} + \mathbf{E}_{(N\times P)}. \tag{1.13}$$

where, **X** represents the observed data matrix, **F** is the matrix of factor scores, **A** is the matrix of factor loadings, and **E** is the matrix of residuals. The subscripts in Eq. (1.13) refer to the dimensions of the matrices. Specifically, k_f is a scalar that indicates the number of factors

to be extracted, which is typically less than P, the number of measured variables or well-logging tools being applied. Constant N represents the number of measurement data for each well-logging tool, or the number of data depth points (i=1,2,…,N).

Factor analysis aims to identify a smaller number of latent factors underlying a given collection of manifest response variables. Each observed variable, $\mathbf{x}_p$, can be expressed as a linear combination of these factors, plus an error term or residual. The product of $\mathbf{F}_{(N\times k_f)}\mathbf{a}^{(p)}_{(k_f\times 1)}$ produces an estimate for $\mathbf{x}_p$, while the term $\mathbf{e}_p$ represents the difference between these estimates and the observed values of $\mathbf{x}_p$. This leads to the notion that each factor is a linear combination of the observed variables. For the general case of p variables and k_f factors, this relationship can be generalized to the following form

$$\mathbf{x}_p = a_{p1}\mathbf{f}_1 + a_{p2}\mathbf{f}_2 + \cdots + a_{pk_f}\mathbf{f}_{k_f} + \mathbf{e}_p \quad (p = 1,2,\ldots,P). \tag{1.14}$$

The factor model can be derived from either the data matrix, or by analyzing the variances and covariances of the variables (Jöreskog et al., 2016). It is more convenient to approach the topic by analyzing the variances and covariances. The assumptions of the factor model include linear independence of factors, uncorrelated terms, and unity variances of the factors in standardized form. By examining the covariance matrix, we can establish the relationship between the data matrix and the population

$$\mathbf{\Sigma} = \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{A}\mathbf{A}^{\mathrm{T}} + \mathbf{\Psi}. \tag{1.15}$$

The data matrix, represented as $\mathbf{X}$, can be treated as a sample matrix from a population with p variables. To obtain more accurate estimates of variances and covariances, the sample size can be increased by adding more row vectors or objects. As the sample size N approaches infinity, it converges towards the population values. Therefore, we have the population covariance matrix $\mathbf{\Sigma}$ (p-by-p) for the observed variables, the factor loadings matrix $\mathbf{A}$ (p-by-k_f), and the residual covariance matrix $\mathbf{\Psi}$ (p-by-p).

Jöreskog (1963) proposed a non-iterative approach for estimating the factor loadings $\mathbf{A}$ from the sample covariance matrix $\mathbf{S}$. Since $\mathbf{S}$ is a consistent estimate of $\mathbf{\Sigma}$. This results in a simple solution for estimating $\mathbf{A}$ as

$$\mathbf{A} = [diag(\boldsymbol{S}^{-1})]^{-\frac{1}{2}}\mathbf{\Omega}(\mathbf{\Gamma} - \theta\mathbf{I})^{\frac{1}{2}}\mathbf{U} \tag{1.16}$$

where $\mathbf{\Gamma}$ is the diagonal matrix of the first k_f number of sorted eigenvalues $\left(\lambda_{k_f}\right)$ of the sample covariance matrix $\mathbf{S}$, $\mathbf{\Omega}$ is the matrix of the first k_f number of eigenvectors and $\mathbf{U}$ is an arbitrary chosen k_f-by-k_f orthogonal matrix. And θ is calculated as follows

$$\theta = \frac{1}{p - k_f}\left(\lambda_{k_f+1} + \lambda_{k_f+2} + \cdots + \lambda_{k_f}\right). \tag{1.17}$$

Once the factor loadings $\mathbf{A}$ have been calculated by Eq. (1.16), the matrix of factor scores $\mathbf{F}$ can be computed. Factor scores $\mathbf{F}$, which describe the measure of an individual datapoint's score on a factor, can be estimated using several methods.

Factor scores, which describe the measure of an individual datapoint's score on a factor, can be estimated using several methods. One common and easily understood method is Bartlett's formula (1937), which produces unbiased scores that are correlated only with their own factor (Eq. 2.2)

$$\mathbf{F}^{\mathrm{T}} = (\mathbf{A}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{X}^{\mathrm{T}}. \tag{1.18}$$

Other methods include the maximum likelihood method, where factor scores and loadings are simultaneously estimated to maximize the likelihood objective function in an iterative procedure.

To enhance the interpretability of factors in factor analysis, an orthogonal transformation of factor loadings is often applied, as recommended by Lawley and Maxwell (1962). This transformation helps simplify the structure of the factors, making it easier to understand their relationships with observed variables while maintaining their statistical properties. Orthogonal rotation techniques (e.g. varimax) are commonly employed to achieve this simplification, aiding researchers in gaining meaningful insights from their factor analysis results.

Reservoir characterization forms the foundation for effective reservoir management, enabling the evaluation of reservoir properties and the optimization of hydrocarbon production. Accurate estimation of parameters such as shale volume, TOC, and permeability is essential for understanding reservoir quality and fluid flow characteristics. Factor analysis,

a powerful statistical technique, offers a comprehensive approach to unraveling reservoir complexities and estimating some of these critical parameters.

Factor analysis is a powerful statistical tool that has been applied to the evaluation of petrophysical parameters in unconventional reservoirs (Szabó et al., 2021). Due to the variations in unconventional reservoir types, no single method can be satisfactorily applied for the estimation of important parameters, such as Total Organic Carbon (TOC) content, which aid in the proper appraisal of resources. Factor analysis offers a multivariate statistical method of estimation by exploring the interdependencies between well logs and model parameters.

Various approaches have been developed to solve the factor analysis problem and enhance the accuracy of reservoir characterization (Abordán and Szabó, 2018; Szabó and Dobróka, 2013; Szabó and Dobróka, 2018; Abordán and Szabó, 2020; Szabó et al., 2022a). Inversion techniques, such as maximum likelihood estimation and Bayesian inversion, aim to refine factor scores by minimizing the misfit between observed well logs and calculated data derived from the factor model. Artificial Neural Networks offer a powerful tool for capturing complex relationships between well logs and reservoir parameters, enhancing the accuracy of factor analysis. Simulated Annealing, inspired by the process of annealing in metallurgy, optimizes factor analysis by searching for the global minimum between observed and calculated data. Other optimization techniques, such as particle swarm optimization and genetic algorithms, can refine factor scores and improve the accuracy of the factor model.

In this study, case studies from different geological basins serve as compelling examples of the practical implementation of factor analysis in reservoir characterization. These studies encompass a range of hydrocarbon formations, including tight sand and shale gas reservoirs. The results obtained through factor analysis are validated against independent estimates derived from core data, deterministic modeling, and other indirect estimation techniques. The consistency and accuracy of the estimated parameters, including shale volume, and permeability, demonstrate the effectiveness of factor analysis in reservoir characterization.

Therefore, factor analysis is presented as a powerful tool in reservoir characterization, enabling accurate estimation of crucial petrophysical parameters, offering an alternative to other methods for evaluating unconventional resources. The various approaches employed to solve the factor analysis problem including inversions, neural networks, simulated annealing, and other optimization techniques, contribute to refining factor scores and improving the accuracy of the factor model. Further research can be conducted to analyze

effects of TOC, level of maturity, kerogen type, and depositional environments on extracted organic factor.

1.3. Well logging inversion

A crucial point in geophysics is to make inferences about the physical parameters of geological structures from the observed data. This can be done based on the laws of physics, which provide the means for computing the data values given a mode. This is called the forward problem. Inverse theory requires a knowledge of the forward model capable of predicting data if the model parameters were, in fact, already known. A good choice of the forward model is needed because inverse theory is inherently mathematical and as such does have its limitations, it is best suited to estimating the numerical values of model parameters for some known or assumed mathematical model.

Well logging data is used to determine essential petrophysical parameters of geological formations that underlie the calculation of hydrocarbon reserves, such as effective porosity, water saturation, shale content, and fractional volume of rock matrices, and they can be determined with their estimation errors from the joint inversion of suitable well logs. Said the above, any type of inversion method must start with the definition of the forward problem. In future research, a modern method called interval inversion will be further investigated (Dobróka and Szabó, 2001, Dobróka et al., 2016).

Dobróka and Szabó (2001) introduced the interval inversion method that uses depth-dependent response functions (instead of local ones) in solving the forward problem to improve inversion accuracy. They suggested increasing the overdetermination (data-to-unknowns) ratio by inverting data from a longer processing interval in a joint inversion procedure. This is achieved through a series expansion-based model discretization technique, in which at least an order of magnitude more data is inverted than unknowns, compared to local inversion. This approach significantly improves the accuracy and reliability of model parameters extracted by inversion (Dobróka et al., 2016). The interval inversion method allows treating increasing number of unknowns without the significant decrease of overdetermination, making possible to estimate additional unknowns of interest, for example the simultaneous determination volumetric properties and textural parameters (e.g. cementation exponent, saturation exponent, and tortuosity factor) (Dobróka and Szabó, 2011).

The interval inversion approach combines all data from an arbitrary logging interval inverting it all at once in the same inversion procedure, instead of estimating the petrophysical parameters in a depth-by-depth inversion by the used of all available data in just one measure point, making the inverse problem highly overdetermined and the interpretation process less sensitive to data disturbances by carefully picking the number of unknowns in the same interval. The great boost in estimation accuracy provided by interval inversion has previously been demonstrated for conventional hydrocarbon deposits (Dobróka et al., 2016). Another, important future of interval inversion is that since there are more data than unknowns in the processed interval, it is possible to automatically calculate not only a restricted number of volumetric quantities, but also layer-thicknesses and other well log constants as key calibration parameters.

Different studies (Szabó et al., 2022b; Abordán and Szabó, 2020; Szabo and Dobróka, 2020; Dobróka and Szabó, 2012; Dobróka et al., 2012) have proven that multi-mineral structures and unconventional formations may be reliably assessed in the context of interval inversion, which has previously been used to analyze shaly sand sequences, carbonate and metamorphic reservoirs, and shale gas formations.

1.3.1. The forward problem of well logging

As a first step in well-logging inversion, it is necessary to set up the forward problem based on the tool response functions, presented in Eqs. (1.19.) to (1.27), to produce wireline logging data from the model parameters in shale gas formations

$$\mathrm{GR} = V_{cl}GR_{cl} + V_kGR_k + \sum_{i=1}^{p} V_{ma,i}GR_{ma,i} \tag{1.19}$$

$$\rho_\mathrm{b} = \phi[S_w\rho_w + S_h\rho_h] + V_{cl}\rho_{cl} + V_k\rho_k + \sum_{i=1}^{p} V_{ma,i}\rho_{ma,i} \tag{1.20}$$

$$\mathrm{U} = V_{cl}U_{cl} + V_kU_k + \sum_{i=1}^{p} V_{ma,i}U_{ma,i} \tag{1.21}$$

$$\mathrm{K} = V_{cl}K_{cl} + V_kK_k + \sum_{i=1}^{p} V_{ma,i}K_{ma,i} \tag{1.22}$$

$$\mathrm{Th} = V_{cl}Th_{cl} + V_kTh_k + \sum_{i=1}^{p} V_{ma,i}Th_{ma,i} \tag{1.23}$$

$$\mathrm{P_e} = \phi\left[S_w P_{e,w} + S_h P_{e,h}\right] + V_{cl} P_{e,cl} + V_k P_{e,k} + \sum_{i=1}^{p} V_{ma,i} P_{e,ma,i} \tag{1.24}$$

$$\phi_\mathrm{N} = \phi\left[S_w \phi_{N,w} + S_h \phi_{N,h}\right] + V_{cl} \phi_{N,cl} + V_k \phi_{N,k} + \sum_{i=1}^{p} V_{ma,i} \phi_{N,ma,i} \tag{1.25}$$

$$\Delta\mathrm{t} = \phi[S_w \Delta t_w + S_h \Delta t_h] + V_{cl} \Delta t_{cl} + V_k \Delta t_k + \sum_{i=1}^{p} V_{ma,i} \Delta t_{ma,i} \tag{1.26}$$

$$R_d = \frac{a R_w}{\phi^m S_w^{n^*}} - R_{cl}(V_{cl} - V_k)^2 + V_k^2 K_{RF} \tag{1.27}$$

In Eqs. (1.19)-(1-27) one can find two types of parameters, the petrophysical parameters, such as porosity (ϕ), water saturation (S_w), hydrocarbon saturation (S_h), clay content (V_{cl}), kerogen volume (V_k) , and matrix volumes ($V_{ma,i}$) (p is the number of matrix components), and the second type of parameters are the so-called zone parameters that do not have a significant change with depth interval, for instance, textural parameters (i.e. saturation exponent (n), cementation exponent (m), tortuosity factor (a)). The observed quantities are natural gamma ray intensity (GR_i), potassium (K_i), uranium (U_i) and thorium (Th_i) spectral intensity, bulk density (ρ_i).

And other parameters for each component (i) of the modelled formation (i.e., water (w), hydrocarbon (h), pyrite (py), etc.) noted as a subscript. Usually, these zone parameters are considered constants in the processed interval (Table 1.1) and are obtained from special and conventional core analysis tests, drilling information, theoretical graphics.

The petrophysical parameters in Eqs. (1.19.)-(1.27) are governed by the material balance relation,

$$\phi + V_{cl} + V_k + \sum_{i=1}^{p} V_{ma,i} = 1, \tag{1.28}$$

which ensures that the total volume of all components within a formation equals to 1. This relation represents a complete volumetric balance of the rock's petrophysical properties and is a fundamental physical requirement that must be strictly satisfied.

Total organic carbon (TOC) estimation, crucial for understanding shale properties, can be achieved by applying the formula of Tissot and Welte (1978),

$$\mathrm{TOC} = \frac{\rho_k}{K_\alpha} V_k \rho_b^{-1}, \tag{1.29}$$

where K_α is known as kerogen conversion factor that refers to the quantity and quality of kerogen, and it is dependent on the type and concentration of kerogen present in the rock, as well as its thermal maturity.

Table 1.1 Zone parameters used for calculating synthetic well logging data

Wireline log	Unit	Pore-water (*w*)	Hydrocarbon (*h*)	Clay (*cl*)	Kerogen (*k*)	Pyrite (*py*)	Quartz (*qz*)	Carbonate (*c*)
Natural gamma ray (GR_i)	API	0	0	200	500	0	2	8
Potassium (K)	%	0	0	4.4	5.9	0	0.5	0.3
Uranium (U)	ppm	0	0	1.5	150	0	0	0
Thorium (Th)	ppm	0	0	8	12	0	0	0
Photoelectric absorption index (P_e)	barn/e	0.81	0.09	3.5	0.24	16.97	1.81	4.11
Bulk density (ρ_b)	g/cm^3	1.09	0.016	2.58	1.45	5.01	2.65	2.79
Sonic transit-time (Δt)	µs/ft	200	305	108	135	36.2	56	46
Neutron-porosity (ϕ_N)	v/v	0.92	0.26	0.3	0.67	0	-0.02	0.01
Deep resistivity (R_d)	Ωm	0.015	0	1	3300	0.27	0	0
Textural factor (*m, n, a*)	-	2	2	1	---	---	---	---

1.3.2 Point-by-point inversion

In geophysical studies, inverse problems are often formulated as linear systems involving a sensitivity matrix. This linear operator captures the relationship between the unknown model parameters to estimate, and the observed data measurements. A common application is well log data inversion, where the goal is to determine petrophysical properties like porosity, saturation, mineral fractions etc. at different depths from logs measuring various physical quantities.

The conventional approach is to perform a point-by-point joint inversion, where at each depth level, all the log measurements from the different tools are combined to estimate the localized petrophysical parameters. Mathematically, this leads to an overdetermined system since there are slightly more data points than unknowns. However, the inversion must account for noise corrupting the data.

While this seems reasonable, a key limitation arises from the inherently underdetermined nature of the inverse problem. There are typically fewer log measurements than the total

number of petrophysical parameters requiring estimation. This low data-to-unknown ratio makes the parameter estimation less accurate and reliable. However, a new method has been proposed (Dobróka and Szabó, 2001) at the Department of Geophysics of the University of Miskolc to enhance the data-to-unknown ratio for well logging inverse problems.

1.3.3. Interval inversion of wireline logging data

Petrophysical parameters are assumed to be functions of depth overcoming the small overdetermination of the local inverse procedures that leads to a relatively noise sensitive inversion results (Dobróka and Szabó, 2001). The modified response function is fixed so that it is depth-dependent

$$d_s^{(c)} = g_s\big(m_1(z), \dots, m_i(z), \dots, m_M(z)\big), \tag{1.30}$$

where g_s denotes the response function of the *s*-th logging tool (*s*=1, 2, ..., *ST*, where *ST* is the number of logging instruments) and m_i is the *i*-th petrophysical property (there is *M* number of petrophysical parameters).

In Eq. (1.30), model parameters are represented by continuous functions that must be discretized for numerical calculations (Dobróka et al., 2016). The interval inversion approach is based on model parameter discretization by series of expansion as follows

$$m_i(z) = \sum_{q=1}^{Q_i} B_q^{(i)} \Psi_q(z) \tag{1.31}$$

where m_i represents the *i*-th unknown model parameter, $B_q^{(i)}$ represents the *q*-th series of expansion coefficient, and Ψ_q represents the *q*-th basis function (Q_i is the sufficient number of expansion coefficients characterizing the *i*-th model parameter) (Dobróka et al., 2012).

The basis functions are considered to be known quantities, the choice of which is not tightly limited. For example to describe a layer wise homogenous model, it can be using a combination of unit step functions (Heaviside function) (Dobróka et al., 2012), while if the petrophysical parameter shows vertical variation within a layer, the use of power or other relevant basis function types as orthogonal set of polynomials can be selected as basis

function for inhomogeneous formations models (Szabó and Dobróka, 2020; Dobróka et al., 2016).

However, it has recently been shown that a large number of unknown expansion coefficients reduces the stability of the inversion operation in some instances, and this phenomenon is strongly reliant on the choice of the initial model. Although regularization can numerically stabilize the inversion technique, it may nevertheless trap in a local minimum or become divergent and remain sensitive to the starting model (Szabó and Dobróka, 2020).

1.3.4. Layer-wise homogeneous model in interval inversion

As mentioned before, for a layer wise homogeneous model, we may characterize it using a combination of unit step functions (Dobróka and Szabó, 2001), also known as Heaviside function, with the fewest unknown parameters

$$\Psi_q(z) = u\left(z - z_{q-1}\right) - u\left(z - z_q\right) \tag{1.32}$$

$$\Psi_q(z) = \begin{cases} 0, \text{if} & z < z_{q^-} \\ 1, \text{if} & z_{q-1} \le z \le z_q \\ 0, \text{if} & z > z_q \end{cases} \tag{1.33}$$

where z_{q-1} and z_q are, respectively, the upper and lower depth coordinates of the q-th rock unit in meters, respectively. Coefficient $B_q^{(i)}$ in Eq. (1.31) corresponds to the i-th petrophysical parameter of the q-th layer since the q-th basis function is always zero except in the q-th layer

$$m_i\left(z_{q-1} \le z \le z_q\right) = \sum_{q=1}^{Q_i} B_q^{(i)}. \tag{1.34}$$

Therefore, each petrophysical parameter is specified by one series expansion coefficient using the above orthogonal-function expansion. The inversion technique is used to determine the series expansion coefficients, which now constitute the model vector of the interval inversion problem. The interval-wise homogeneous model has the benefit of requiring the

inversion technique to identify a substantially smaller number of unknowns than data. The substantially overdetermined inverse problem yields a precise and trustworthy estimate over the whole recorded interval rather than individual measuring points (Abórdan and Szabó, 2020; Dobróka et al., 2012).

Substituting Eq. (1.34) into Eq. (1.30), which represents the connection between the data and the model parameters changes with this series expansion, it is now a response function evaluated in a depth interval. The derived synthetic data from the y'-th log is

$$d_s^{(c)}(z) = g_s[z, B_1^{(1)}, \dots, B_{Q1}^{(1)}, \dots \dots, B_1^{(M)}, \dots, B_{QM}^{(M)}]^{\mathrm{T}}. \tag{1.35}$$

To estimate the $B_{QM}^{(M)}$ expansion coefficient, the following objective function was optimized for fitting the theoretical data to observed data by minimizing the relative data distance between them. Equation (1.35) is used to compute the calculated data with that the objective function is

$$\mathrm{E} = \sum_{p=1}^{PM} \sum_{j=1}^{ST} \left(\frac{d_{pj}^{(m)} - d_{pj}^{(c)}}{d_{pj}^{(m)}} \right)^2 = \min \tag{1.36}$$

where $d_{pj}^{(m)}$ and $d_{pj}^{(c)}$ indicate the j-th measured and computed data in the p-th depth (PM is the total number of measurement points in the processed depth interval). Basically, two types of optimization approaches are used to solve the inverse problem: linear and global optimum seeking methods.

1.3.5. Linear optimum seeking method for inversion

Linear optimization algorithms are the most commonly used inversion problem solvers in practice because they are fast and effective if an initial model is set sufficiently close to the optimum. The regularized variation of the Gaussian Least Squares approach, as proposed by Marquardt (1959), is widely utilized. However, linear optimization methods are not absolute minimum seeking methods since they tend to assign the solution to the objective function's local optimum.

The solution of the nonlinear inverse problem is reduced to a succession of linear problems using ordinary search methods. The problem should be linearized before utilizing these approaches. Based on Marquardt algorithm, the vector of model correction for the series expansion coefficients is computed by

$$\mathbf{B} = (\mathbf{G}^{\mathrm{T}}\mathbf{G} + \varepsilon^2\mathbf{I})^{-1}\mathbf{G}^{\mathrm{T}}\,\mathbf{d}^{(\mathrm{m})} \tag{1.37}$$

where $\mathbf{G}$ is a linear operator, called Jacobian or sensitivity matrix, $\mathbf{I}$ is the identity matrix, and ϵ is a properly chosen small positive constant to numerically stabilize the inversion procedure. When there is no substantial change in the fit between the measured and computed data, as it is calculated in Eq. (1.38) based on Eq. (1.36), it can be considered that the already estimated expansion coefficients as solution and may readily deduce the depth variation of petrophysical characteristics by inserting them into Eq. (1.31)

$$D = \sqrt{\frac{1}{PM \cdot ST}\sum_{p=1}^{PM}\sum_{j=1}^{ST}\left(\frac{d_{pj}^{(m)} - d_{pj}^{(c)}}{d_{pj}^{(m)}}\right)^2} \cdot 100\ \%. \tag{1.38}$$

The inversion kernel matrix or Jacobian of the problem (Eq. 1.30) describes the relationship between the unknow model parameters $\mathbf{m}$ and the observed data $\mathbf{d}$. In vector calculus, the Jacobian matrix of a vector function of several variables is the matrix whose elements are the first-order partial derivatives of said function, in the geophysical inversion problem the Jacobian matrix is form the partial derivatives of the data, with respect to the parameters.

To obtain the partial derivatives in the present work, the partial derivates to form the Jacobian matrix (Eq. 1.39) are approximated by finite differences

$$\mathbf{G} = \begin{bmatrix} \left.\frac{\partial f_1}{\partial m_1}\right|_{\mathbf{m}_0} & \left.\frac{\partial f_1}{\partial m_2}\right|_{\mathbf{m}_0} & \left.\frac{\partial f_1}{\partial m_M}\right|_{\mathbf{m}_0} \\ \left.\frac{\partial f_2}{\partial m_1}\right|_{\mathbf{m}_0} & \ddots & \vdots \\ \left.\frac{\partial f_N}{\partial m_1}\right|_{\mathbf{m}_0} & \cdots & \left.\frac{\partial f_N}{\partial m_M}\right|_{\mathbf{m}_0} \end{bmatrix} \tag{1.39}$$

where the f_N function is any of the Eqs. (1.19)-(1.27).

Centered finite difference is often used as an approximation of the derive typically in numerical differentiation using the Eq. (1.40), in which the computation of f_N is straight forward

$$\left.\frac{\partial f_N}{\partial m_M}\right|_{\mathbf{m}_0} \approx \frac{f_N(\mathbf{m}_0 + \Delta\mathbf{m}_0) - f_N(\mathbf{m}_0 - \mathbf{\Delta m}_0)}{2\Delta\mathbf{m}_0} \tag{1.40}$$

Menke's (1984) theory states that the covariance matrix of petrophysical characteristics may be determined from the observed data at near-zero regularization damping factor. The connection between the data and model can be expressed using their covariance matrices. Menke (1984) proposed the following matrix equation

$$\mathbf{COV}(\mathbf{m}) = \mathbf{G}^{-g}\mathbf{COV}\left(\mathbf{d}^{(\mathrm{m})}\right)[\mathbf{G}^{-g}]^{\mathrm{T}} \tag{1.41}$$

where $\mathbf{d}^{(m)}$ is the vector of measured data and $\mathbf{G}^{-g}$ is the general inverse matrix of the linearized problem. The estimation error of the i-th model parameter is derived from the i-th element for the main diagonal of the model covariance matrix

$$\sigma_i^{(m)} = \sqrt{cov_{ii}(\mathbf{m})} \qquad (i = 1,2,\dots,M). \tag{1.42}$$

The correlation matrix characterizes the reliability of the inversion estimate to the effect that strongly correlated model parameters normally die not yield stable and acceptable accurate parameter estimations. Therefore, to quantify the strength of correlation between model parameters the correlation matrix is used

$$corr_{ii}(\mathbf{m}) = \frac{cov_{ij}(\mathbf{m})}{\sqrt{cov_{ii}(\mathbf{m})cov_{jj}(\mathbf{m})}}. \tag{1.43}$$

The correlation matrix can be characterized by one scalar, which is called the mean spread

$$S_C = \sqrt{\frac{1}{M(M-1)}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(corr_{ij}(\mathbf{m}) - \delta_{ij}\right)^2}. \quad (1.44)$$

1.4. Geological information of study areas

1.4.1. Norway: North Sea Central Graben Area

The analyzed well logging dataset is collected in the Norwegian Continental Shelf at the North Sea, which it is part of the most productive hydrocarbon provinces in the world (Brennand et al., 1998). Geologically, the North Sea is an intracratonic basin, in other words a basin which lies on continental crust. This basin has been subject to periods of stretching, thinning and subsidence during the late Carboniferous, Permian-Early Triassic, and Late Jurassic times. It has been seen that the most important source rocks in the North Sea province are marine shales from the Upper Jurassic. These rocks correspond to the Triassic to Jurassic transition, which is characterized by a change from continental to shallow marine depositional environment due to a transgression in the Early Jurassic leading to the accumulation of black shales over large parts of Northwest Europe with favorable organic richness, organic matter quality and maturation characteristics, being good source rocks for oil and gas in southern parts of the North Sea.

The dataset corresponds to a well drill on the Central Graben area, crossing three distinctive lithological units. The lithological description is based on information reported by the Norwegian Petroleum Directorate (2023). The topmost unit has been described as a marlstone with some content of glauconite and pyrite, local occurrences of sandstones and siltstones has been reported. Also, it has been suggested that this unit were deposited in a well-oxygenated shallow-marine environment with limited supply of clastic and which have a rich faunal assemblage. The following geological unit consists of a slightly to non-calcareous, carbonaceous claystone, rich in organic carbon content. It has been described as contain thin stringers of carbonate rocks and, in some areas, sandstone. It is suggested that the formation was deposited in a restricted, low-energy, oxygen-deficient (i.e. anoxic) marine environment distal from any significant terrestrial influence. For the last layer, it is a generally massive, fine to medium grained, sandstone. Although a thin siltstone layer is

present in the formation. The sandstones are arkosic to subarkosic, glauconitic and micaceous. Thin, nodular calcite-cemented bands are common.

1.4.2. Alaska: North Slope Region

The Kingak Formation, situated in Alaska's North Slope region, is one of three oil and gas source rock systems in the area. This formation is part of the Jurassic to Lower Cretaceous Kingak sequence, it is buried at a depth of over 2700 m and has a basal sequence that is between 300-380 m thick, also known as the K1 sequence. The basal sequence is made up of both marine and terrigenous materials, with organic matter from both sources being deposited in a marine siliciclastic environment during the opening of the Canadian basin (Houseknecht and Bird, 2004; Rouse and Houseknecht, 2016) Wire-line logs through K1 show an off scale gamma-ray response (hot shale) in a thin interval of uniform silty mudstone (North Inigok and Inigok well) (Houseknecht and Bird, 2004).

The Kingak Formation is composed of dark gray to dark-olive-gray shale, siltstone, claystone, and clay ironstone (Detterman et al., 1975). The upper portion of the formation consists of clay shale, silty shale, and siltstone with red ironstone beds that weather to a rusty color. The lower section, K1, is made up of dark gray to black fissile paper shale, dark-gray clay shale, minor claystone, and beds and nodules of red-weathering ironstone (Reiser et al., 1980). The most distal well penetrations of sequence set K1 are located near or beyond the toes of clinoforms in eastern NPRA (e.g., Inigok and North Inigok wells) that exhibit a strong gamma-ray response, indicating a hot shale, in a thin interval of silty mudstone. This is interpreted as a basinal condensed section and an important petroleum source rock (Houseknecht and Bird, 2004). In most cases, the base of sequence set K1 marks the boundary between the Kingak Shale and the underlying Triassic strata of the Sag River Sandstone or Shublik Formation.

A feasibility study was carried out by Detterman et al. (1975) using well logs from the North Inigok 1 well, classifying it as a potential source rock, according to Passey's method. The results of the study were confirmed using core data from the section. The Kingak Shale was also found to be thermally spent (postmature) in the North Inigok-1 well, which is consistent with the reduction of the HIo in the basal organic-rich shale to measured values in the range of about 20 –30 mg HC/g TOC and also at the North Inigok-1 well, TOC values were found to be about 4 wt% in the basinal organic-rich shale (Peters et al., 2006).

1.4.3. Africa: Lower Congo Basin

The Lower Congo Basin (LCB) stands out as a prolific hydrocarbon-bearing region along the west coast of Africa, positioned in the South Atlantic Ocean as the northernmost offshore extension of the Angolan margin. Spanning an expansive area exceeding 115,000 km² with an average water depth of approximately 3000 m, the LCB is a Cretaceous passive continental margin basin shaped by the Gondwana breakup during the Early Cretaceous (Jatiault et al., 2019). This geological setting has led to the formation of the second-largest oil and gas field in Africa (Yang et al., 2023).

A focal point in the LCB is the Upper Cretaceous–Paleogene Madingo Formation, a sedimentary unit extending over a 52-Myr period. Within this formation, the fine-grained rocks present significant targets for analysis in marginal marine environments and post-salt hydrocarbon exploration. The Madingo Formation witnessed two distinct marine environments during its deposition: the Early and Middle Madingo section, a semi-restricted sea bay environment from the Turonian to lower Campanian, correlated with the Coniacian–Santonian oceanic anoxic event, and the Late Madingo section, an open marine environment from the upper Campanian to Eocene, correlated with the Paleocene–Eocene thermal maximum (PETM) (Zeng et al., 2024).

The Early and Middle Madingo Period experienced abundant terrigenous nutrient substances and flourishing radiolarians, deposited in a partially enclosed sea bay. This era corresponds to a significant geological event known as the Coniacian–Santonian Oceanic Anoxic Event 3 (OAE-3). Organic content during this time reached up to 4 wt% total organic carbon (TOC), primarily composed of types II_2 and II_1 organic matter (Yang et al., 2023; Zeng et al., 2024).

Transitioning to the Late Madingo period, a shift occurred with an abundance of foraminifera and phosphatic mudstone in a more open marine setting, featuring lithological combinations of argillaceous mudstone and silty mudstone. This phase correlates with the Paleocene-Eocene Thermal Maximum (PETM). The organic matter composition during this later period was predominantly types I and II_1 (Zeng et al., 2024).

The tectonic evolution and sedimentary infill of the LCB can be divided into four main stages: pre-rift, syn-rift, transitional, and drift stages. Oil and gas generation within the basin primarily occurred during the syn-rift stage from pre-salt Lower Cretaceous lacustrine source rocks. The post-salt Upper Cretaceous–Paleogene Madingo Formation, characterized by mud-rich turbidite deposition, is a crucial target for marine source rock analysis, displaying

high total organic carbon (TOC) content (>2 wt%) and hydrocarbon generation potential (S_1 + S_2) (>10 mg/g) (Yang et al., 2023; Zeng et al., 2024).

Drilling and well logging results indicate that, in general, the Madingo Formation consists of thick mudstone containing foraminifera and radiolarians, alongside minor amounts of limestone and siltstone. The formation exhibits a relatively low-medium maturity or maturity stage (0.30% < vitrinite reflectance (*Ro*) < 1.0%). However, limited drilled wells and sparse geochemical data in the post-salt section constrain a comprehensive understanding of the Madingo Formation's potential (Zeng et al., 2024).

2. TRADITIONAL LINEAR ESTIMATION TECHNIQUE FOR TOTAL ORGANIC CARBON CONTENT PREDICTION

2.1. Proposed modification for improving TOC estimation precision (I_{cl}-BS-SA)

In obtaining the optimal values for the regression coefficients (Eq. 1.8), two approaches were compared: (1) the minimization of the discrepancy between observed Total Organic Carbon (TOC) values and calculated values using the Marquardt algorithm for solving the least squares inverse problem as proposed by Bibor and Szabó (2016), and (2) the utilization of Simulated Annealing.

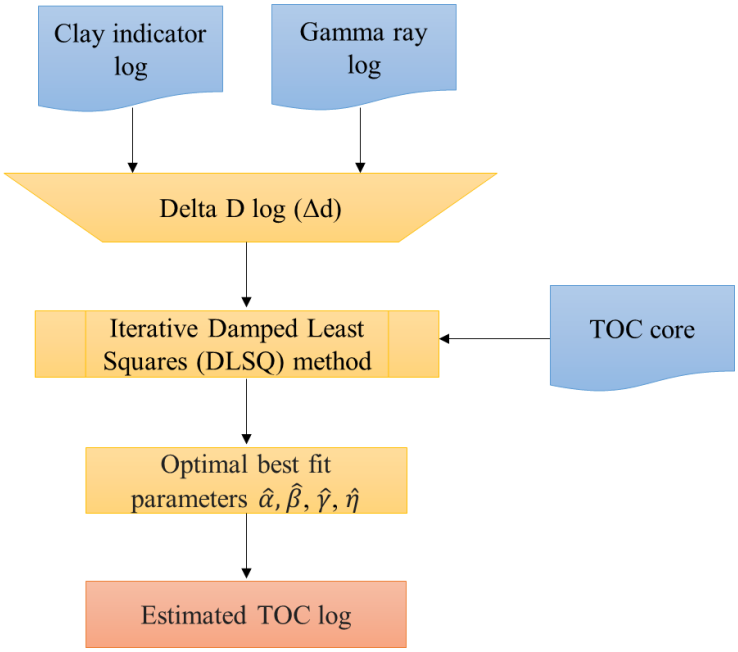

Figure 2.1 The flowchart of the Bibor and Szabó (2016) modified traditional linear estimation technique for TOC prediction (Icl-BS)

The first approach involved employing the Marquardt algorithm, which is a widely used optimization technique for solving nonlinear least squares problems. The algorithm iteratively adjusts the regression coefficients to minimize the difference between the observed TOC values and the values predicted by the regression model. This method aims to find the best-fitting coefficients that yield the lowest overall discrepancy.

On the other hand, the second approach utilized Simulated Annealing, an optimization algorithm inspired by the physical annealing process. Simulated Annealing explores the parameter space by gradually accepting new candidate solutions, even if they are initially worse, to avoid being trapped in local optima. By treating the regression coefficients as parameters to be optimized, Simulated Annealing searches for the set of coefficients that minimizes the discrepancy between observed TOC values and model predictions.

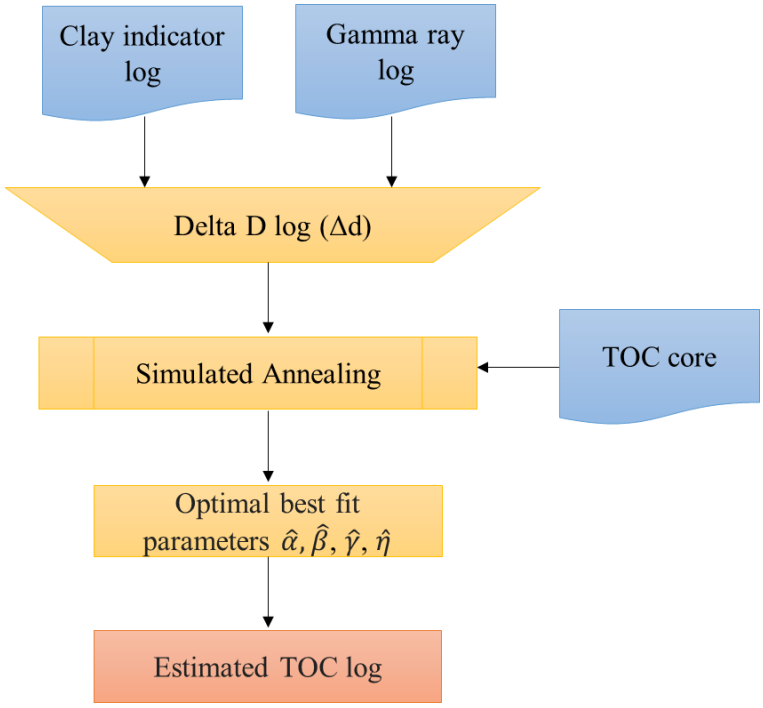

Figure 2.2 The flowchart of the proposed modification of Icl-BS using Simulated Annealing as optimization method (Icl-BS-SA)

By comparing the results obtained from both approaches, we assessed their effectiveness in finding the optimal regression coefficients. Evaluating factors such as convergence speed, accuracy, and robustness, we aimed to determine which method provided a better fit to the observed TOC values and enhanced the overall performance of the regression model. This comparison allowed us to draw insights into the suitability

and advantages of each optimization technique in improving the accuracy and reliability of TOC estimation.

2.2. Comparative analysis TOC estimation using I_{cl}-BS-SA and I_{cl}-BS methods

2.2.1. Field test: Norway

I tested the proposed **I_{cl}-BS-SA** method and compared against **I_{cl}** and **I_{cl}-BS** ones, using one data corresponding to well-2 from the wells of the North Sea Central Graben area. As input for the three procedures, I used the natural gamma-ray intensity (GR), and clay indicator (I_{cl}) logs, which was derived from the neutron-porosity (NEU) and density derived porosity logs using Eq. (1.3), covering a depth interval of 100 m with a sampling space of 0.153 m. Also, TOC values from 22 core samples were used as calibration parameters.

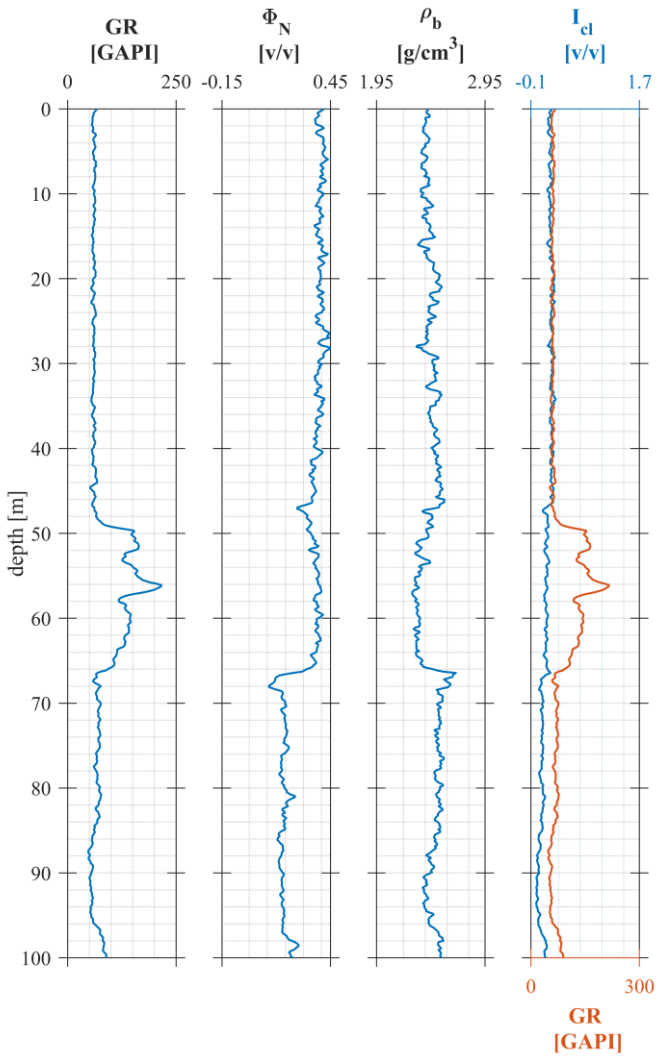

Figure 2.3 Observed well logs in Well-2 in the North Sea Central Graben. GR = natural gamma-ray intensity log; Φ_N = neutron-porosity log; ρ_b = bulk density log; I_{cl} = clay indicator curve

As for the traditional methodology (**I_{cl}**), a linear model was adjusted as Eq. (1.4) suggested using the standard methodology of the Ordinary Least Squares (OLS) method. The regression parameters of the OLS estimates are equal to 11.6465 and -0.0159, for coefficients α and β, respectively. From the coefficient of determination, R^2, it is suggested that approximately 77.28 % of the variability in the dependent variable can be explained by the independent variable in the regression model, indicating a moderately string relationship between the variables of the model. The root-mean-square deviation (RMSE) of the estimates is equal to 1.3177. As for the relative distance error of this model, it is equal to 177 %.

The **I_{cl}-BS** methodology gave the following estimate nonlinear regression parameters, according to the Eq. (1.7), that is for $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\eta}$, which values are 7.3670, 0.9621, 349.9163, 5.10364, respectively. For solving the nonlinear regression, an inversion procedure was performed as shown in the Fig. 2.1. The initial regression parameters were set to 10, 1, 100 and 10 values, giving an initial relative data distance of 95.40 %, with 40 iterations with a damping factor of 0.5, and a decrement of 0.8 of it per iteration, a final relative data distance value of 51.74 % was achieved. The RMSE of the estimates was equal to 0.4438.

The proposed **I_{cl}-BS-SA** method was performed considering 30 runs, with 100,000 iterations each run. An initial temperature (T_0) of 3.3905 and changing the temperature in a multiplicative monotonic cooling scheme, $T_0 * k$, with k equals to 0.9. The initial regression parameters for the **I_{cl}-BS-SA** scheme were set as for the **I_{cl}-BS**. The final average relative data distance of the procedure achieved after the 30 runs was 15.09 %, while the RMSE of the estimates was 0.1749: considering 7.0824. 0.9445, $1.3500 \cdot 10^6$, 11.7005, for α, β, γ, and η, correspondingly.

Table 2.1. Results of the different methods for TOC estimation in Well-2

Method	Estimated regression coefficients				Relative Data Distance (%)	RMSE
	$\alpha \mid \hat{\alpha}$	$\beta \mid \hat{\beta}$	$\gamma \mid \hat{\gamma}$	$\eta \mid \hat{\eta}$		
I_{cl}	11.6465	-0.0159			177	1.3177
I_{cl}-BS	7.3670	0.9621	349.9163	5.10364	51.74	0.4438
I_{cl}-BS-SA	7.0824	0.9445	$1.3500 \cdot 10^6$	11.7005	15.09	0.1749

The results of the comparison between the **I_{cl}**, **I_{cl}-BS**, and **I_{cl}-BS-SA** methods provide valuable insights into their performance and effectiveness in modeling the relationship between the natural gamma-ray intensity (GR) and clay indicator (I_{cl}) logs, expressed through the Δd distance as variables for TOC estimation.

Starting with the traditional **I_{cl}** method, which employed linear regression solved by OLS. Although R^2 indicated that the variability in the dependent variable could be explained by the independent variable, suggesting a moderate relationship, the relatively high RMSE and a substantial relative distance error indicate limitations in accurately predicting the dependent variable based on the independent variable. To address the limitations of the linear approach, the **I_{cl}-BS** method employed a nonlinear regression model, employing an inversion procedure based on the Levenberg–Marquardt algorithm

or DLSQ method. The resulting relative data distance indicated an improvement compared to the linear model, suggesting a better fit to the data. Additionally, the reduced RMSE indicated a more accurate estimation of the dependent variable. However, the **I_{cl}-BS** method still exhibited a relatively high relative data distance and lacked the robustness to achieve optimal parameter estimation. In light of these limitations, the proposed **I_{cl}-BS-SA** method introduced a novel approach by incorporating simulated annealing into the nonlinear regression framework. The resulting relative data distance indicated a substantial improvement compared to both the **I_{cl}** and **I_{cl}-BS** methods. The remarkably low RMSE further demonstrated the enhanced accuracy achieved by the **I_{cl}-BS-SA** method.

The performance of the **I_{cl}-BS-SA** method can be attributed to the introduction of Simulated Annealing, which effectively optimized the regression parameters $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\eta}$ to minimize the discrepancy between predicted and observed values. The ability of simulated annealing to escape local optima and explore the parameter space likely played a crucial role in achieving this improved performance. The lower relative data distance and RMSE values attained by the **I_{cl}-BS-SA** method suggest a stronger and more precise relationship between the Δd distance, highlighting its superior capability in capturing the underlying geological characteristics.

In conclusion, the comparative analysis demonstrates that the proposed **I_{cl}-BS-SA** method surpasses the traditional linear regression approach (**I_{cl}**) and the nonlinear regression method with inversion procedure (**I_{cl}-BS**) in terms of accuracy and parameter estimation. The incorporation of Simulated Annealing enables the **I_{cl}-BS-SA** method to achieve significantly improved results, with a lower relative data distance and RMSE.

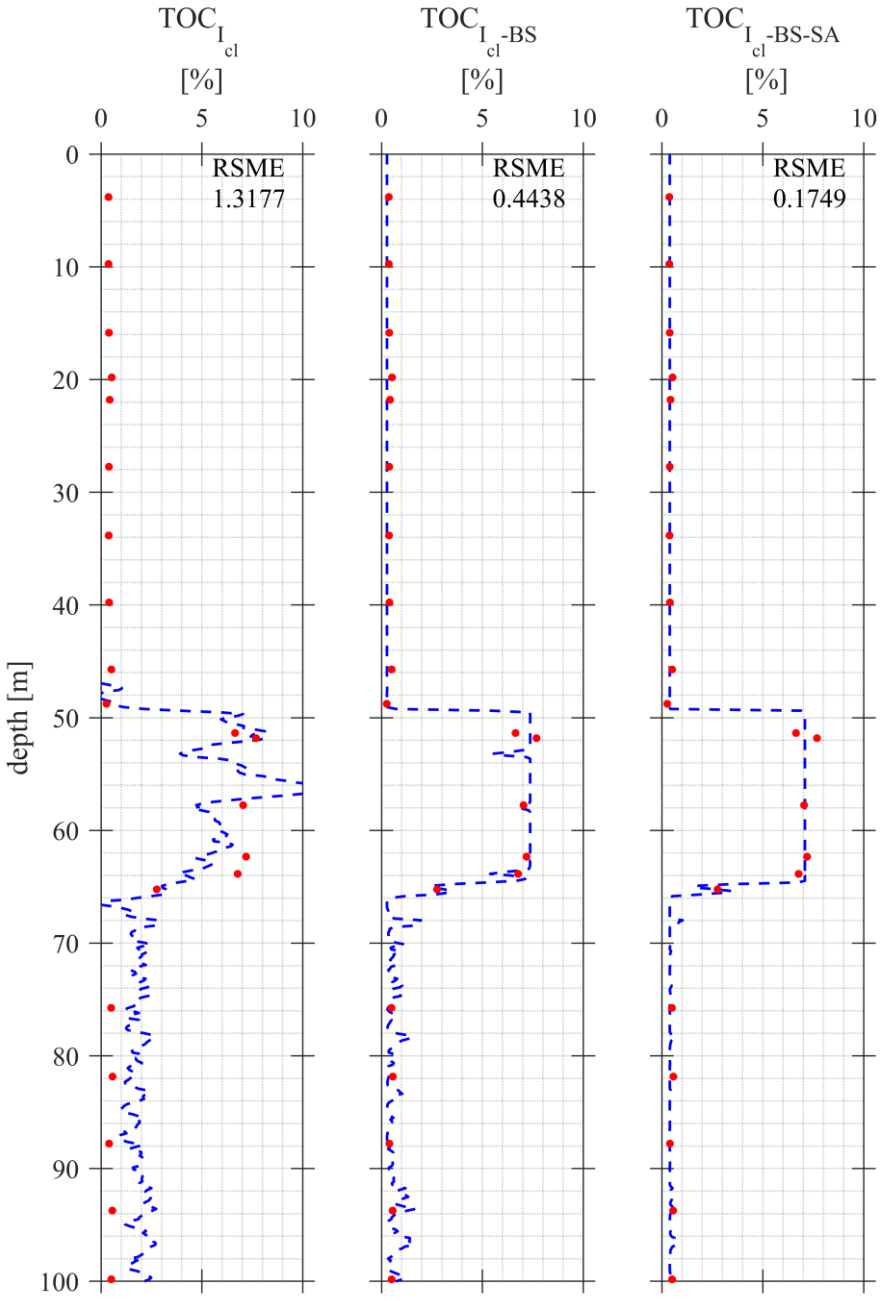

Figure 2.4 The resultant TOC logs and their RSME in Well-2 in the North Sea Central Graben, Norway. From top to bottom the TOC calculated using **I_{cl}-BS-SA**, **I_{cl}-BS** and **I_{cl}** methods, respectively

2.3. Summary

The comparison between three methods for predicting Total Organic Carbon (TOC) content, denoted as the **I_{cl}**, **I_{cl}-BS** and **I_{cl}-BS-SA** methods, highlights significant advancements in predictive accuracy and parameter estimation.

The original linear estimation technique, **I_{cl}**, utilizes Ordinary Least Squares (OLS) to establish a linear relationship between the natural gamma-ray intensity (GR) and clay indicator (I_{cl}) logs for TOC prediction. However, this method exhibits limitations due to a relatively high Root Mean Square Error (RMSE) and significant relative distance error. To address these drawbacks, the **I_{cl}-BS** method introduces a nonlinear regression model employing the Levenberg–Marquardt algorithm. Despite improvements in relative data distance and reduced RMSE, it still lacks robustness in achieving optimal parameter estimation. As an improvement, the **I_{cl}-BS-SA** method, integrates Simulated Annealing into the nonlinear regression framework, significantly improving parameter estimation and showcasing a remarkable decrease in relative data distance and RMSE compared to both the **I_{cl}** and **I_{cl}-BS** methods. Simulated Annealing's efficacy in exploring parameter space effectively and escaping local optima plays a pivotal role in achieving better performance. Overall, the **I_{cl}-BS-SA** method exhibits higher accuracy and parameter estimation compared to traditional linear and previous nonlinear regression techniques.

The comparison emphasizes the limitations of the traditional linear I_{cl} method in accurately predicting TOC values, leading to the introduction of the **I_{cl}-BS-SA** modification. This enhancement integrates Simulated Annealing into the nonlinear regression framework to optimize regression coefficients and improve predictive accuracy. Testing against traditional and modified methods demonstrates that **I_{cl}-BS-SA** outperforms both, displaying substantially lower relative data distance and RMSE. This superiority is attributed to Simulated Annealing, which effectively optimizes regression parameters and captures geological characteristics more precisely. Consequently, the **I_{cl}-BS-SA** method represents a significant advancement over traditional linear and modified nonlinear approaches, offering improved accuracy and parameter estimation for TOC content prediction.

In review, the comparative analysis strongly supports the superiority of the **I_{cl}-BS-SA** method over traditional linear and previous nonlinear techniques in predicting TOC content. The incorporation of simulated annealing stands out as a key factor enhancing accuracy and parameter estimation, suggesting its potential for reliable TOC estimations.

2.4. Thesis I

I compared two classical well logging methods, the clay indictor (**I_{cl}**) and Bibor-Szabó (**I_{cl}-BS**) methods and further developed the latter to improve the estimation of Total Organic Carbon (TOC) content in hydrocarbon formations.

The proposed **I_{cl}-BS-SA** method integrates Simulated Annealing into the nonlinear regression framework, a global optimum search method, to avoid the limitations of local optimization methods. To achieve optimal parameter estimation, the energy function was introduced utilizing normalized root mean squared error minimization.

This key addition led to significant advancements in parameter estimation, showcasing significant decreases in both relative data distance (91.5% improvement) and RMSE compared to its predecessors - the **I_{cl}** (86.6% improvement) and **I_{cl}-BS** methods (60.6% improvement) - when tested in real well-log data collected from North Sea Central Graben, Norway.

3. EVALUATION OF UNCONVENTIONAL HYDROCARBON RESERVOIRS APPLYING FACTOR ANALYSIS

3.1. Total Organic Carbon content evaluation through factor analysis

Factor analysis has been successfully used in extracting new variables or factor logs which correlate highly to organic content (kerogen) of the studied formations. Of the extracted two factors, the first factor (F_1) defines the shale volume (Szabó, 2011; Szabó and Dobróka, 2013) while the second factor (F_2) correlates highly to the organic content of the formation, termed the "organic factor" for the estimation of TOC. An empirical Eq. (3.1) was suggested (Szabó et al., 2021) to estimate the TOC continent directly from the statistical factor (F_2), where α_F and β_F are site-specific regression coefficients obtained from the calibration procedure to minimize the distance between the observed (e.g. core data) and theoretical TOC (e.g. linear regression based)

$$\mathrm{TOC}(F_2) = \alpha_F F_2 + \beta_F. \tag{3.1}$$

In my investigations, the second factor (F_2) extracted from well logging data often exhibiting a strong correlation with TOC values. Regression analysis and cross-plots further refine the relationship between the second factor and TOC, allowing for accurate

estimation of organic richness within unconventional formations. This proposed approach (Fig. 3.1) permitted the estimation of both shale volume and organic content in unconventional reservoirs and highlighted the importance and robustness of the second factor for the estimation of TOC, and the significance of using F_2 conjunctly with F_1 to highlight sweet spots, as F_1 acts as a lithological indicator.

It is important to mention that the factor used to reveal the TOC petrophysical information from the well-logging datasets needs to be scaled into an interval [0 1] using the Eq. (3.2) suggested by Szabó (2011), so that, interdependencies between the extracted factor (F_2) and the estimated petrophysical parameters can be confirmed by the newly scaled factor (F_2')

$$F_2' = \frac{F_2 - F_{2,\min}}{F_{2,\max} - F_{2,\min}}. \tag{3.2}$$

This method, named the organic factor method, improves our knowledge on application of statistical methods in well logging analysis, enhances the vertical resolution of in situ TOC estimation necessary for reservoir characterization, offers a non-destructive method for TOC estimation, also making possible reducing the number of petrophysical unknowns encountered in inversion procedures and thereby, increasing the overdetermination ratio in well log inversion.

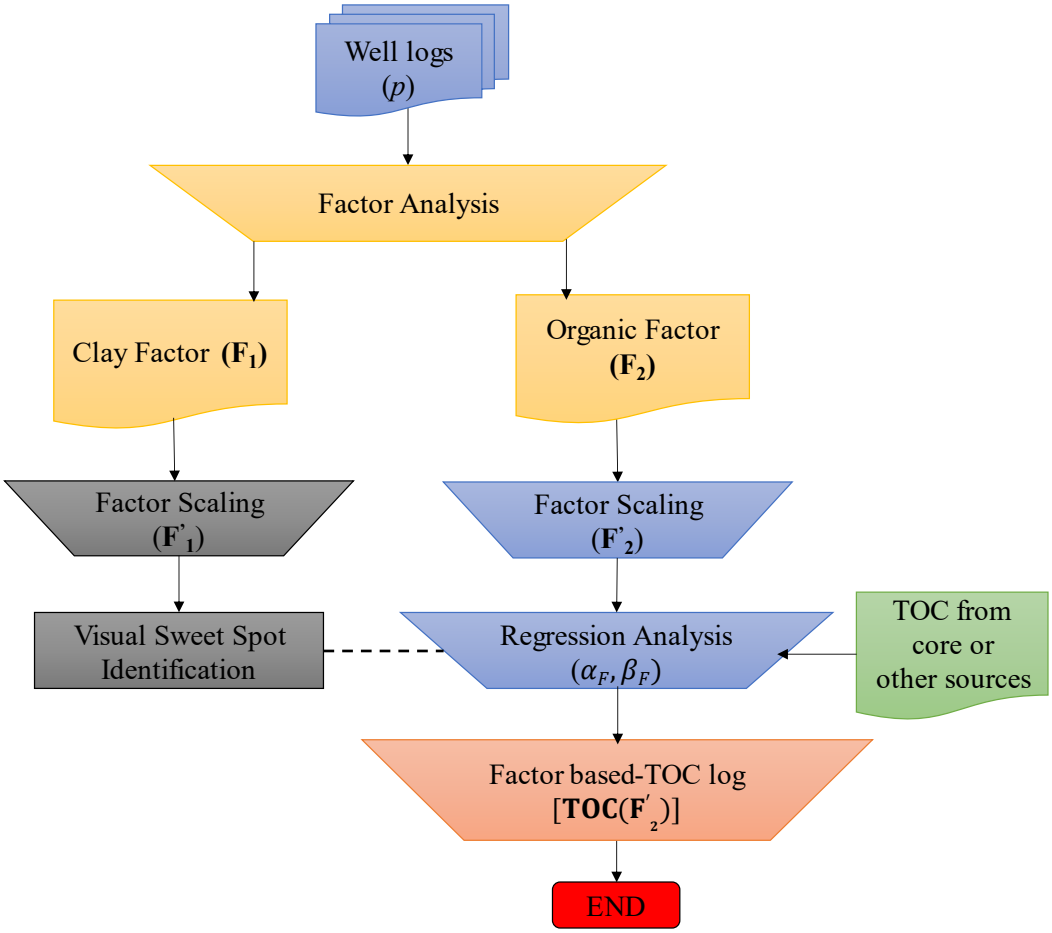

Figure 3.1 The flowchart of the proposed technique for factor based-TOC prediction (**FA-TOC**)

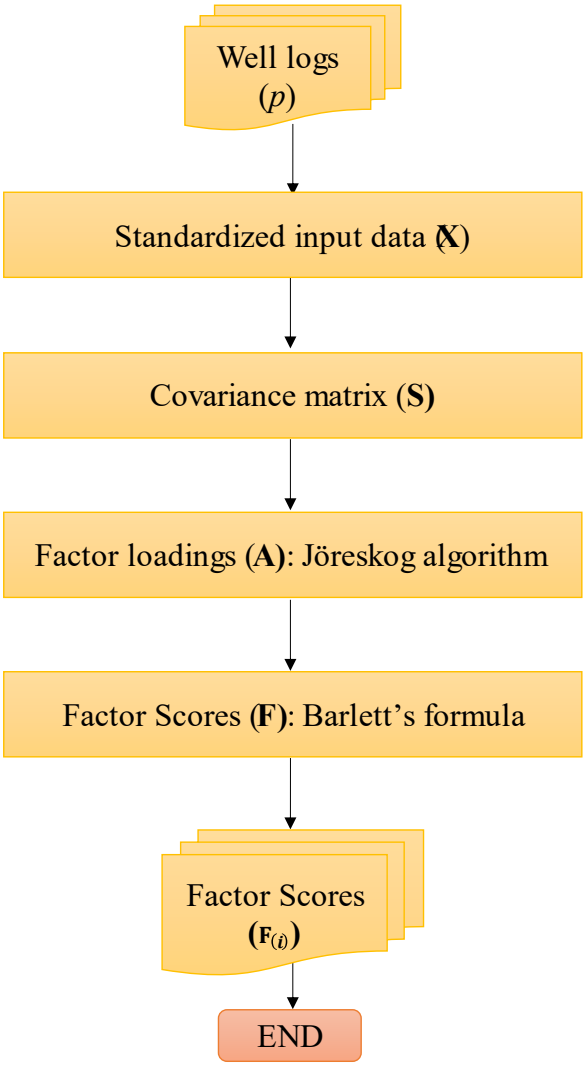

Figure 3.2 The flowchart of the factor analysis phase of the FA-TOC procedure

3.2 Case studies

3.2.1 Synthetic modeling test

I tested the **FA-TOC** method first on a synthetic well logging dataset (Fig. 3.3), generated in a well named Well-X, and the inferences drawn from the results obtained were used later to analyze other wells from real boreholes drilled in oil fields in different basins. The geological setting assumed for the syntenic dataset tried to emulate a shale interval of 40 m in depth, modeled after an unconventional reservoir based on distinctive petrophysical characteristics of Barnett Shale reservoir from the USA (Lee and Kim, 2019). The theoretical probe response functions used are given in Eq. (1.19) to (1.27). The results from the modeling are displayed in Figure 3.3, also a 4 % Gaussian distributed

noise ($\sigma_{d,k}$) was added to the synthetic data (d_k^{obs}) using the formula $d_k^{noisy} = d_k^{obs}\left(1 + \mathrm{N}_D(0, \sigma(0, \sigma_{d,k})\right)$.

The synthetic wireline logs build considered a sampling rate of data is 0.1 m, and it includes the natural gamma-ray intensity; spectral gamma ray intensity logs such as uranium, thorium, and potassium concentrations; photoelectric absorption index; bulk density; P wave sonic interval time; neutron porosity; and deep (laterolog) resistivity logs, given a matrix of 401-by-9 noisy synthetic data.

After standardizing the input data from Well-X, and obtaining its covariance matrix, the factor loadings and the residual covariance matrix were calculated by Jöreskog's non-iterative approach using Eq. (1.16). For this, it was necessary to choose the number of factors to be extracted. Based on Jöreskog's work, Eq. (1.17) can be used to select the number of factors, that is, the value of factors (k_f) should be such that the value of parameter θ in Eq. (1.17) is less than 1. Therefore, the number of factors selected was set to 3 (θ_3= 0.9450, θ_2= 1.2594). Table 3.2 contains the resultant factor loadings, which indicate how strongly each variable (i.e. well logs) is associated with the factors. Factor loadings close to 1 or -1 indicate a strong correlation relation, while loadings close to 0 suggest a weak relationship. Based on the obtained loadings, we can make some initial interpretation.

Table 3.2. Factor loadings rotated using the varimax algorithm for the synthetic dataset in Well-X.

Logs / ***Loadings***	GR	ρ_b	U	Th	K	PEF	ϕ_N	Δt	R_d
Loadings (F_1)	0.94	-0.08	-0.06	0.98	0.99	0.07	0.57	0.53	0.45
Loadings (F_2)	0.31	-0.94	0.98	0.13	0.02	-0.25	0.79	0.83	-0.06
Loadings (F_3)	0.13	-0.22	0.16	0.09	0.10	-0.94	-0.18	-0.02	0.86

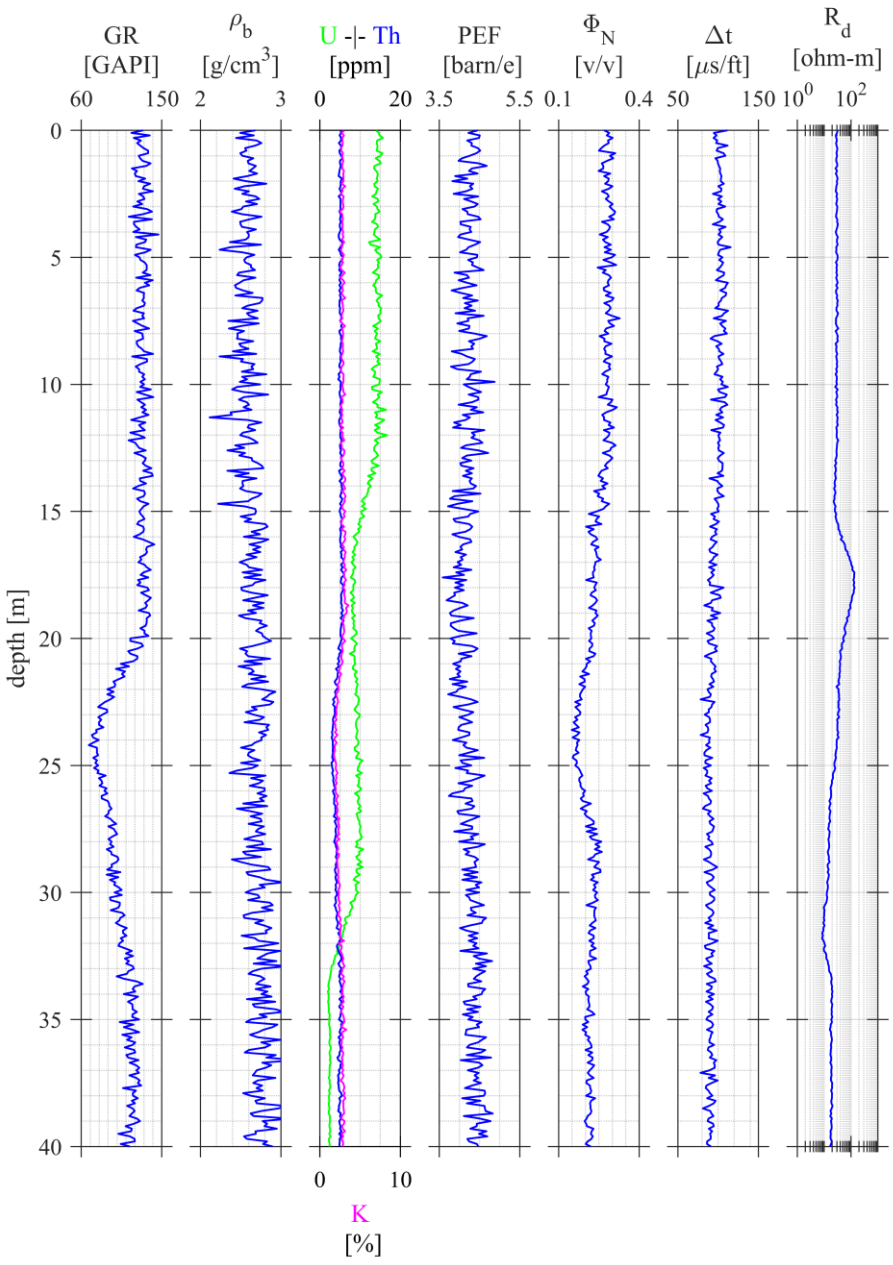

Figure 3.3 Synthetic dataset in well X. Denotations are as follows: natural gamma-ray intensity (GR), spectral gamma-ray intensity (K, U, TH), photoelectric absorption index (PEF), bulk density (ρ_b), neutron-porosity (Φ_N), acoustic travel-time (Δt), and deep resistivity (R_d)

As it is observed in Table 3.2, the natural gamma ray, and potassium - thorium spectral gamma ray logs have the highest loadings on the first factor, while there is a visible relationship between the uranium spectral gamma ray, neutron porosity and sonic logs. Meanwhile, the third factor does not show a clear combination of other logs, rather than the resistivity log. Thus, the presence of uranium (U) in the second factor with a high loading indicates a potential link to organic matter.

Previous work on the field of factor analysis applied to the study of conventional reservoirs has proven that the first factor strongly correlates with the shale volume of the investigated formations, and an empirical relationships was established for the estimation of shale volume (Szabó, 2011). Therefore, the second factor's loadings given in this unconventional formation called the attention to my work, since it is notable that there is a relationship between the organic content and the second factor (F_2).

Furthermore, after obtaining the factor loadings and the residual covariance matrix, it was possible to used then as inputs to estimate the factor scores of the dataset and analyzing the correlation between the second factor and the volumetric parameters, providing insights into the relationships between the extracted factors and the petrophysical parameters. There is a highly positive relation between F_2 and volume of kerogen (R= 0.98), making feasible to be established an empirical relationship to the TOC content, obtained by different means (i.e. cores samples, inversion, Passey's method, etc.). Also, it should be mentioned that with organic-rich related mineralogy components, such as, pyrite (R=0.16) and carbonates (R=0.73), there is a moderate to weak positive correlation, while in the case of a quartz lithology exhibits a negative association (R=-0.77).

These remarks suggest that Factor 2 exhibits a strong correlation with the volume of kerogen. This indicates that high Factor 2 values are associated with significantly higher kerogen content. Factor 2 might represent geological conditions or lithologies that are more favorable to the accumulation of kerogen while, there is a negative relationship between Factor 1 and the volume of kerogen (R=-0.09), implying that high Factor 1 values are associated with lower kerogen content, further confirmed by the low negative correlation between F_2 and volume of clay (R=-0.36).

Likewise, the correlation of the second factor between the borehole logs readings was analyzed, obtaining the following remarks. F_2 has a strong positive correlation with spectral gamma ray for uranium (R=0.98), sonic travel-time (R=0.83), and neutron-porosity (R=0.78), and a strong negative correlation is reflected by the density log (R=-0.95).

Based on response of well-logs to organic matter, and the tendency of correlation (signal), it is clear that F_2 reflects the nature of the organic matter in the formation. For instance, the enrichment of organic matter in a formation interval can lead to the increased of sonic interval times values, because it often has lower acoustic velocity than minerals, and neutron porosity readings, due to the lower neutron absorption of organic matter. The

positive correlation between F_2 and sonic and neutron-porosity logs supports this idea; while the bulk density log response may show lower values in the presence of organic-rich material due to its low density, approximated to 1 (g/cm^3), related also by the negative correlation between F_2 and bulk density logs.

Following the correlation analysis it was clear that there is a strong positive correlation (R=0.98) with the measured TOC log (Fig. 3.4), obtained from the volume of kerogen of the known synthetic model. Then, the extracted Factor 2 was then scaled into a maximum and minimum interval (Eq. 3.2) to obtain a model that utilized the exiting interdependencies between the extracted factor and the organic matter in the formation, to approximate the values of organic matter, as levels of TOC content. Thus, a linear regression model for the F_2 vs. TOC relation was established (Eq. 3.1), giving the following regression coefficients, $\alpha_F = 4.1092$ and $\beta_F = 0.4816$.

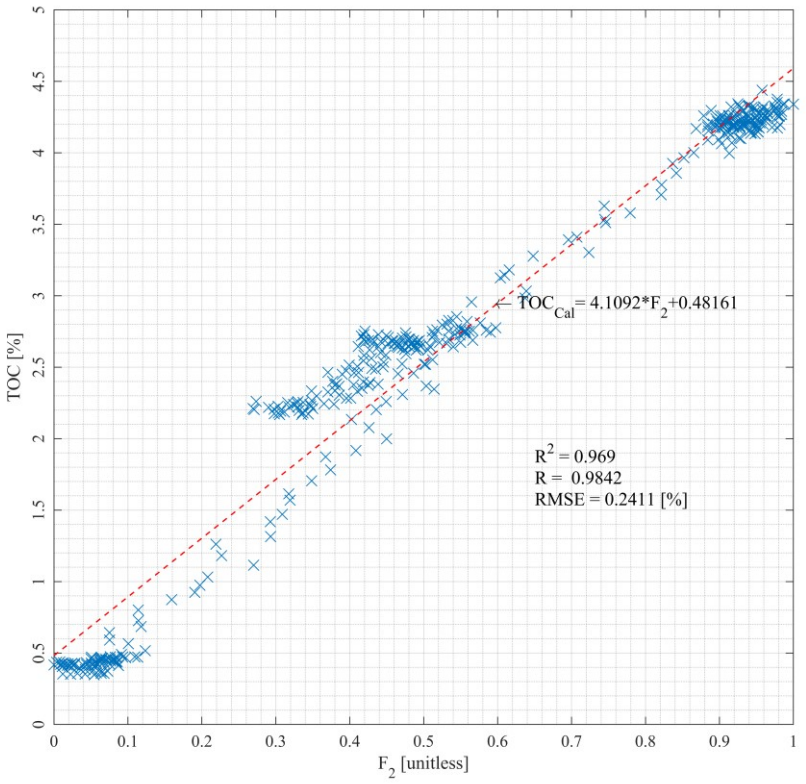

Figure 3.4 F_2 versus TOC plot showing a strong linear relationship between the second extracted factor and the total organic carbon content

The proposed model explains a significant portion of the variance in TOC based on F_2, as indicated by the R^2 value of 0.97, with a strong positive Pearson's correlation coefficient (R) equals to 0.98, and a Root Mean Squared Error (RMSE) of 0.2411 (%). The high positive Spearman's correlation coefficient of 0.97 – which indicates the strength of correlation in case of nonlinear relation – further supports the relationship between F_2 and TOC. Both the linear regression and Spearman's correlation results suggest a substantial association between F_2 and TOC in the dataset. However, the main difference is in the assumption about linearity. Pearson's R assumes a linear relationship, whereas Spearman's correlation coefficient does not make any assumptions about linearity.

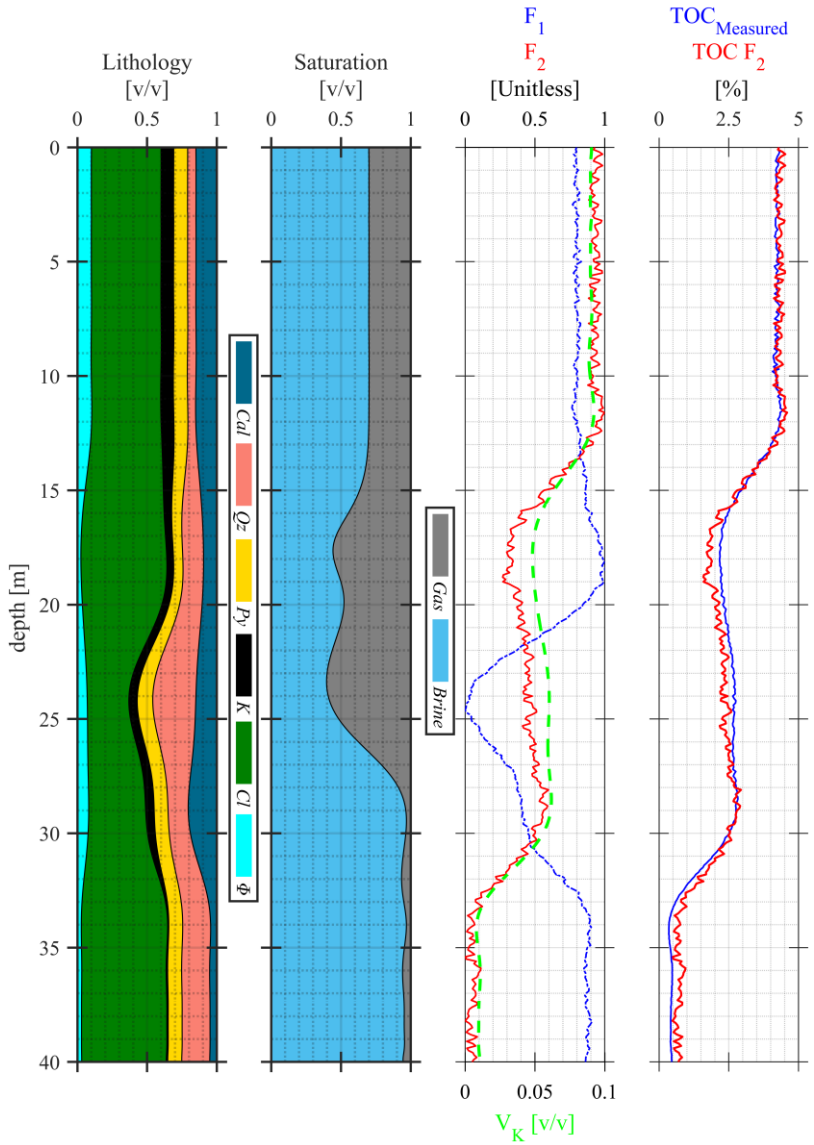

Figure 3.5 Results from the factor analysis of the synthetic logs in well X. On the first and second track the proportions of lithology and fluids in the formation are shown. On the third

track, extracted factors are shown along with the volume of kerogen. On the last track, the TOC measured and calculated using the second factor are displayed. Denotations are as follows: calcite (*Cal*), quartz (*Qz*), pyrite (*Py*), kerogen (*K*), clay (*Cl*) and porosity (*Φ)*

3.2.2 Field test: Norway

I tested the proposed **FA-TOC** method, using the data corresponding to Well-2 from the wells of the North Sea Central Graben area. (Figure 3.6), such dataset was also introduced in chapter 2. Therefore, the geological description of the study area can be found in subsection 1.4.1. As input for the factor analysis procedure, the natural gamma-ray intensity (GR), bulk density (ρ_b), neutron-porosity (Φ_N), sonic travel-time (Δt), and deep resistivity (R_d) logs were used, covering a depth interval of 100 m with a sampling space of 0.153 m.

The resulting factor loadings from the factor analysis procedure are shown in Table 3.3. The dataset analyzed does not contain spectral gamma ray logs for uranium, thorium, or potassium concentrations (Figure 3.6). However, the results are consistent with the ones observed for the synthetic dataset. As the factor loadings show, the first factor is highly influenced by the natural gamma ray log, strongly related to the volume of shale as proofed before, while the volume of shale has an important impact on the resistivity log as an increase in the volume of shale is strongly associated with higher resistivity values. As for the second factor, it seems to be more related to those logs that are associated with the volume of kerogen, particularly due to its strong positive loadings for Φ_N and Δt, followed by a high negative loading for the ρ_b. As it was previously discussed, the geological implication of such tendency is related to the organic matter effect on the well log's readings, where high positive loadings for Φ_N and Δt logs indicate that as the volume of kerogen increases within a subsurface formation, the neutron-porosity readings increase, and the transit time decreases, implying that formations with higher kerogen content tend to have higher porosities and lower transit times.

Table 3.3 Factor loadings rotated using the varimax algorithm in Well-2, Norway

Logs / ***Loadings***	**GR**	$\boldsymbol{\rho_b}$	$\boldsymbol{\Phi_N}$	**Δt**	$\mathbf{R_d}$
Loadings (F_1)	0.88	-0.32	-0.18	0.35	0.85
Loadings (F_2)	0.37	-0.68	0.92	0.93	-0.03

The correlation between the observed well logs and the two factors is also congruent with the expected impact of organic rich intervals on borehole readings. Factor 1 has a strong and positive correlation to natural gamma ray (R=0.90) and deep resistivity (R=0.88) logs, and Factor 2 is highly correlated to the neutron-porosity (R=0.93) and acoustic travel-time (R=0.93) logs. These relationships can be indicative of hydrocarbon-rich source rocks, as kerogen is a precursor to hydrocarbons.

After analyzing the factor scores, TOC values from 22 core samples were used as calibration parameters to obtain an estimated TOC log (Figure 3.6). The resulting initial best fit model according to the proposed Eq. (3.1) has α_F and β_F values of 10.998, and -0.7237, respectively. However, such model, although it is the one that best minimizes the misfit between the calculated and measured TOC data, would be meaningfully wrong due to the fact that for a portion of the range of function $\mathrm{TOC}(F_2)$, $\{F_2 \in \mathbb{R} \mid 0 \leq F_2 \leq 0.066\}$, the estimated organic content would be negative; which is physically impossible. Therefore, a constrained to the model was proposed such that β_F should be greater or equal to zero. Then, the final best fit model has parametric constant of $\beta_F = 0$, and a $\alpha_F = 9.3597$, with 0.8954 correlation coefficient and R^2 equals to 0.80, and a RMSE of 1.322 (%).

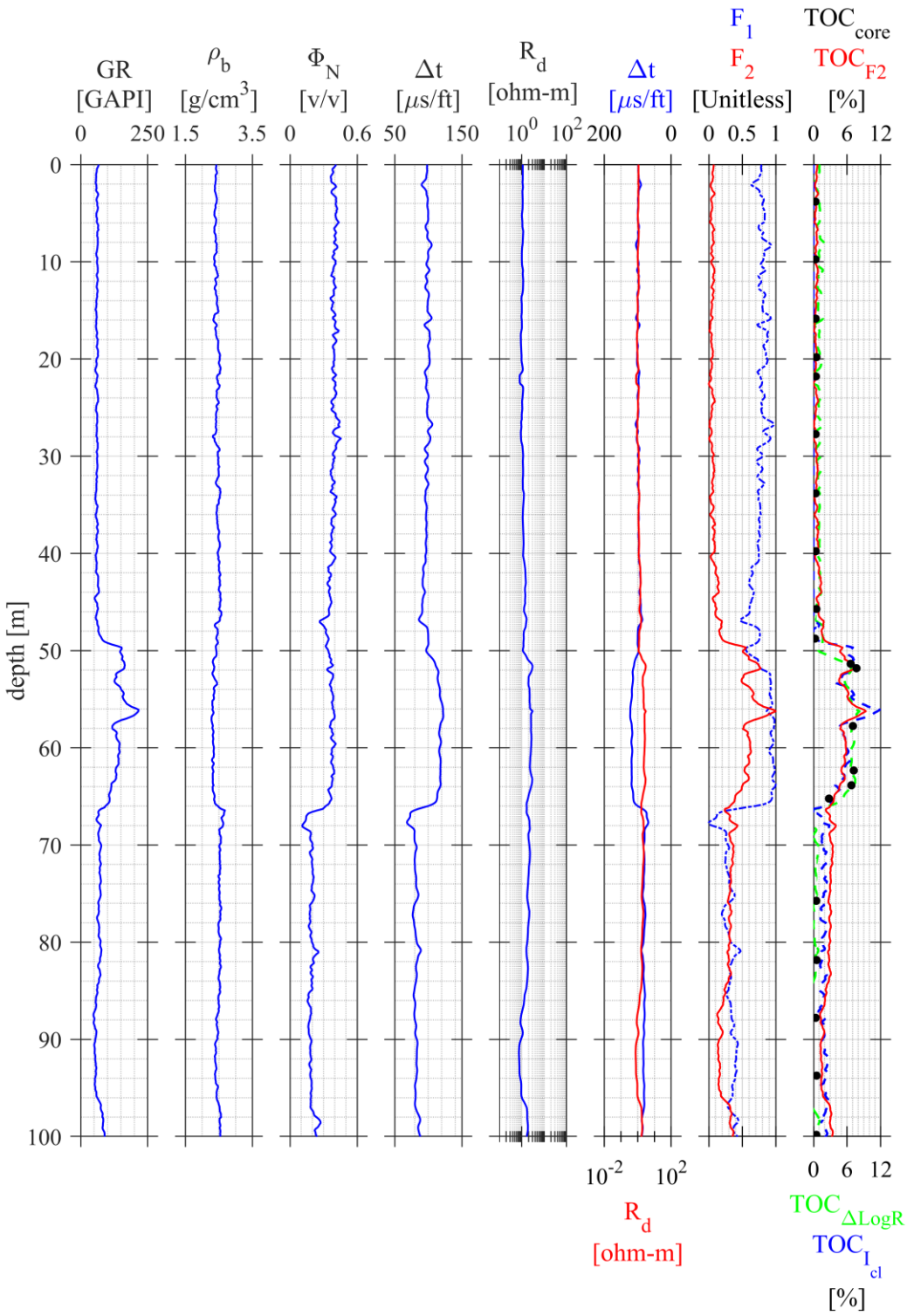

Figure 3.6 Well logging dataset in Well-2 in the North Sea, Norway. Denotations are as follows: natural gamma-ray (GR), bulk density (ρ_b), neutron-porosity (Φ_N), acoustic travel-time (Δt), and deep resistivity (R_d), Factor 1 (F_1) and Factor 2 (F_2). Total organic carbon logs from core analysis (black dots), F_2 (red line), Passey's method (green dashed line) and clay index method (blue dashed line)

Comparing these results with the ones obtained using other classical estimation methods, for example, the clay indicator method, introduced in chapter 2, we can see that the estimated model output does not vary much respect to the RMSE, 0.0044 percentage points; comparing it to the model obtained previous by Valadez Vergara (2020) using the Passey's method, the RMSE was reported as 1.36 (%), having an absolute difference of 0.043 points. Therefore, it can be stated that the results obtained from factor analysis are not only consistent but also demonstrate a high degree of accuracy and reliability when compared to other established estimation methods (Figure 3.6).

3.2.3 Field test: Alaska

I also tested the proposed **FA-TOC** method, using data corresponding to a 90 m long interval from the well North Inigok 1, situated in Alaska's North Slope region (USA), more specifically at the Kingak geological formation (See subsection 1.4.2). As input for the factor analysis procedure, the natural gamma-ray intensity (GR), bulk density (ρ_b), neutron-porosity (Φ_N), sonic travel-time (Δt), and deep resistivity (R_d) logs were used, covering a depth interval of 90 m with a sampling distance of 0.153 m.

The resulting factor loadings are shown in Table 3.4. The factor analysis result showed that, unlike in the previous two cases, where the highest loadings on the second factor were directly related to the neutron-porosity and sonic logs, while the natural gamma ray log readings were associated with the first factor (volume of shale), here we can see that first factor had its highest correlation with the neutron-porosity log, while that with the natural gamma-ray is low.

Table 3.4 Factor loadings rotated using the varimax algorithm for North Inigok 1, Alaska, USA

Logs / ***Loadings***	**GR**	**ρ_b**	**ϕ_N**	**Δt**	**RD**
Loadings (F_1)	0.12	-0.73	0.91	0.84	-0.90
Loadings (F_2)	0.96	-0.49	0.37	0.39	0.05

Although neutron-porosity, sonic and bulk density logs are both highly sensitive to kerogen content, which is evident from the previous results of factor analysis, it is important to consider the geological evolution of the formation. In this case, North Inigok 1 well is interpreted as a basinal condensed section (Houseknecht and Bird, 2004) that

was a significant contributor of oil to the Prudhoe Bay field based on oil–source rock correlation (Peters et al., 2006) and which current TOC values are about 4 wt% in the basinal organic-rich shale. However, the Kingak Shale is thermally spent in the North Inigok-1 well (Peters et al., 2006), which means that the generation and expulsion of petroleum has already occurred, after a long thermal maturation period. Therefore, the well logging tool is slightly affected by the little remaining organic matter in the formation.

To further confirm the relationship between the second factor and the organic matter content in the Alaska formation (Figure 3.7), I calculated the correlation between the available 13 core samples, and also, as an additional evaluation tool, I used the Zhao's method to calculate a TOC log, which model parameters (Eq. 1.7) are 12.606 and 1.4736, respectably, having a RMSE with respect to the core samples of 0.813 (%).

The correlation coefficient between F_2 and the calculated TOC log is 0.98, which can be said that there is a high correlation between the studied variables. For Zhao's method, I needed to obtain the clay indicator index (Eq. 1.3), which has a strong correlation with the first factor (R=0.82) as it was expected.

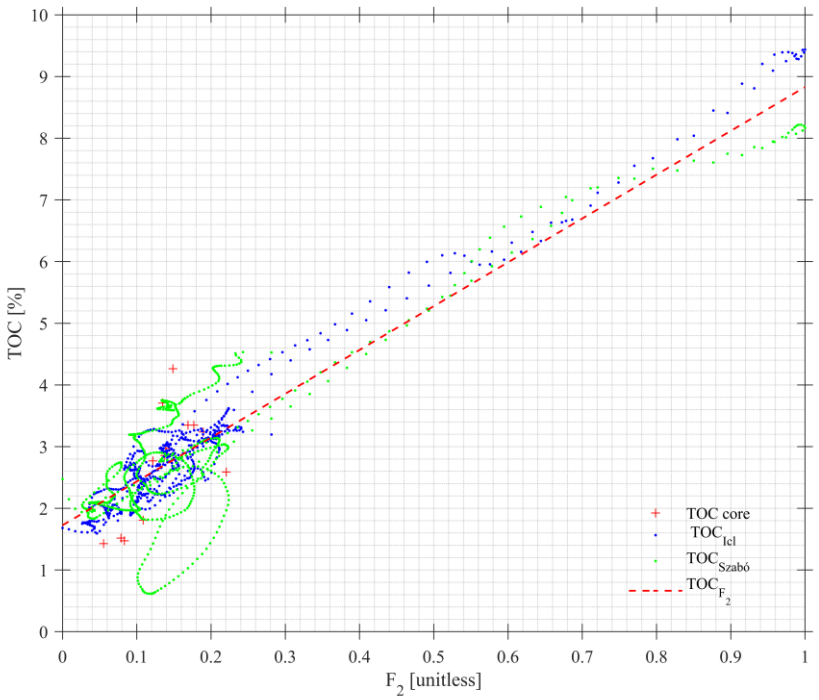

Figure 3.7 Factor 2 vs. TOC cross-plot. Different TOC measurements are displayed. TOC from core measurement is shown by a red cross, while measurements from Zhao's method is displayed by green dots, and TOC estimated by Szabó and Dobróka (2021) using inversion

techniques are shown in red dots. The red dashed line shows the model obtained from F_2 and Zhao's TOC method

Then, I proceeded to calculate the FA vs. TOC model parameters using as calibration measurements the available core samples; however, the results were poor due to the low number of samples. Thus, I used as input the obtained TOC log. The final model parameters then are $\alpha_F = 7.1035$ and $\beta_F = 1.7256$, with a RMSE of 0.7655 (%) and R equal to 0.6709, with respect to the measured (core) TOC values (Figure 3.7). As a comparative independent calculation, an estimated TOC log from Szabó and Dobróka (2020) was used, which compared to FA-TOC shows a high correlation between the variables (R=0.883), and the difference between the predicted values and the actual value is just better for 0.0076 units difference respect to the FA-TOC (Figure 3.8).

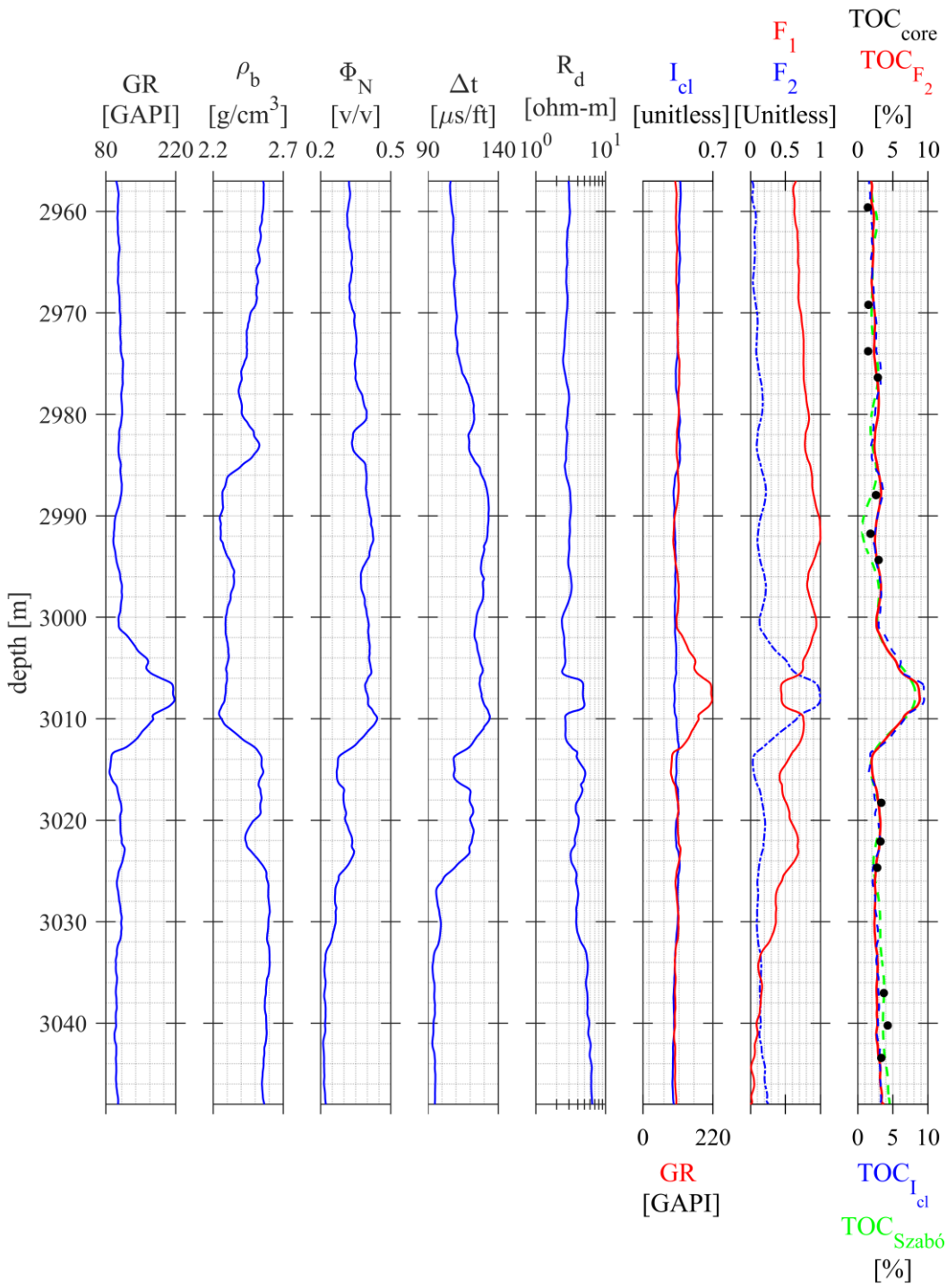

Figure 3.8 Well logging dataset of North Inigok-1, Alaska. Denotations are as follows: natural gamma-ray (GR), bulk density (ρ_b), neutron porosity (Φ_N), acoustic travel-time (Δt), and deep resistivity (R_d) log. The clay index (I_{cl}) is displayed in the sixth track along with GR log. Factor 1 (F_1) and Factor 2 (F_2) are shown at the 7th track. Total organic carbon logs from core (black dots), F_2 (red line), clay index method (blue dashed line), Szabó and Dobróka (2021) method (green dashed line) are displayed in the last track

These analyses confirmed the reliability of factor analysis in providing consistent results when compared to the existing linear methods and the core samples of the well-logs. Furthermore, some other wells were analyzed in Szabó et al. (2021), showing various cases across different geological formations (i.e. shale gas formations, tight gas sands sourced, a conventional reservoir) and well log suites.

3.3 Summary

This chapter discusses the application of Factor Analysis (FA) in estimating Total Organic Carbon (TOC) content in unconventional hydrocarbon reservoirs. It involves the extraction of factors from well logging data to define relationships between factors and organic content, enhancing the understanding of reservoir characteristics and allowing non-destructive estimation of TOC.

The statistical method involves defining two factors: Factor 1 (F_1) representing shale volume, while Factor 2 (F_2) representing organic content. F_2, particularly, exhibits a strong correlation with TOC, enabling the estimation of organic richness within unconventional formations. This approach uses statistical factors to estimate TOC and improves vertical resolution in characterizing reservoirs while reducing unknowns encountered in inversion procedures.

Case studies were conducted to validate this **FA-TOC** methodology. Synthetic modeling tests mimicked shale intervals based on Barnett Shale characteristics, demonstrating the relationship between factor loadings and petrophysical parameters like natural gamma-ray intensity, neutron-porosity, and resistivity logs. This correlation indicates F_2's association with kerogen content, crucial for estimating TOC.

Field tests made both in Norway and Alaska corroborated these findings, showing consistent relationships between factor loadings and logs associated with kerogen volume. The results demonstrated strong correlations between F_2 and TOC, with the FA-TOC method providing accurate estimations compared to other established techniques. Even in a thermally spent formation like Alaska's Kingak Shale, **FA-TOC** showed reliable estimations despite the limited number of core samples.

The **FA-TOC** method not only enhances understanding of reservoir properties - but also offers a robust and accurate means of estimating TOC in unconventional reservoirs across various geological formations. Furthermore, the interpretation of both factors

simultaneously aids quantitative analysis of well logs, enabling differentiation between sweet spots and organic matter-free shale intervals.

3.4. Thesis II

I proposed a new alternative methodology to the estimation of total organic content in unconventional hydrocarbon reservoirs based on the factor analysis. By extracting the statistical factors from well logging data, I corroborated that Factor 1 (F_1) correlates significantly to shale volume, and I identified Factor 2 log (F_2) as the "*organic factor*" linked to TOC.

Through synthetic modeling experiments and field tests, I demonstrated the feasibility of the proposed *FA-TOC* method. F_2 log, particularly, showed strong correlation with TOC, enabling reliable estimations of organic richness within different unconventional formations.

Furthermore, by parallelly interpreting F_1 and F_2, I successfully identified sweet spots and refined quantitative log interpretations, setting a more accurate and effective approach to assess organic content in unconventional reservoirs, providing a reliable means of *in situ* estimating TOC in unconventional reservoirs across diverse geological formations.

4. WELL LOGGING INVERSION IN UNCONVENTIONAL HYDROCARBON RESERVOIRS

4.1 Interval inversion method

The primary goal is to evaluate the volume of multiple matrix components, including the kerogen volume, which plays a significant role in unconventional hydrocarbon reservoirs. The inversion procedure is driven by a synthetic geological model divided into multiple layers, where each layer corresponds to a distinct geological interval, and the model parameters are assumed to remain constant within these intervals.

To execute the inversion, I proposed to utilize the Heaviside basis function for interval inversion (Eq. 1.34), which helps to increase the overdetermination ratio (ODR)—the ratio of observed data to unknown parameters, compared to traditional point-by-point inversion. This ratio is a crucial factor, as a higher ODR allows for better-constrained model parameters and reduces the fitting errors in the inversion process.

I created a synthetic petrophysical model considering the fractional volumes of 12 mineral types forming the rock matrix (i.e., kerogen (V_k), pyrite (V_{py}), illite (V_{ill}), chlorite (V_{clo}), kaolinite (V_{kao}), quartz (V_{qz}) , potassium feldspar (V_{kfelds}), plagioclase (V_{plg}), dolomite (V_{dol}), calcite (V_{cal}), biotite (V_{bio}), and apatite (V_{ap})), and two types of fluids (i.e., water (S_w) and hydrocarbon (S_h) saturations) filling the pore space of the rock (Φ), distributed in 6 proto layers. The petrophysical model is based on binomial distribution of mineral species, trying to replicate the inherent distribution of minerals within a formation (Ma et al., 2016). This choice is particularly important due to the frequent heterogeneity observed in shales, potentially resulting in diverse multimodal distributions. This heterogeneity manifests notably in the histograms of geological logs, displaying either a skewed long-tail or a multimodal pattern. The presence of multiple lithofacies within a mixture distinctly contributes to this non-normality, introducing elements of both multimodality and skewness in wireline log histograms. For instance, tight sandstone formations often exhibit bimodal histograms, and it is essential to note that this bimodality might obscure the presence of more than two distinct mixtures of lithofacies within the formation (Ma et al., 2016).

Each layer consists of 1000 depth points that were later assigned to a depth resolution of 0.01 m; the model parameters was randomly generated from binomial distribution considering 10 trials and a probability of success for each petrophysical parameter as shown in Table 4.1. Also, every output was normalized to ensure the fulfillment of the material balance equation (Eq. 1.27). Additionally, a continuous function of the volumetric parameter was generated by fitting them to a Legendre polynomial of degree 15. Then, from the 6000 samples, an equal distributed sampling was performed to reduce the data set to a resolution of 0.1 m between samples. Finally, the sample was constrained to 501 data points by cutting the top and bottom values of the formation, to avoid the effects of the fitting at the end of the dataset (Fig. 4.1).

Table 4.1. Volumes for each mineral class in the synthetic case.

Layer	Φ	S_w	V_k	V_{py}	V_{qz}	V_{cal}	V_{dol}	V_{plg}	V_{kfelds}	V_{ill}	V_{kao}	V_{clo}	V_{bio}	V_{ap}
1	0.15	0.70	0.02	0.01	0.09	0.08	0.04	0.02	0.04	0.09	0.11	0.1	0.1	0.15
2	0.112	0.80	0.03	0.01	0.03	0.07	0.05	0.05	0.02	0.135	0.16	0.153	0.07	0.11
3	0.08	0.90	0.09	0.02	0.05	0.05	0.04	0.01	0.03	0.16	0.17	0.18	0.05	0.07
4	0.08	0.95	0.15	0.02	0.03	0.04	0.04	0.02	0.03	0.16	0.17	0.17	0.04	0.05
5	0.08	0.90	0.09	0.02	0.03	0.085	0.065	0.03	0.02	0.14	0.165	0.155	0.05	0.07
6	0.100	0.89	0.02	0.02	0.02	0.12	0.09	0.03	0.03	0.1504	0.16	0.1345	0.055	0.07

The volumetric parameter generated (Figure 4.1) was used as input to generate the response function of 9 borehole logging tools, according to Eqs. (1.19) -(1.27). The

selected values of the of the zone parameters in tool response function are given in Table 1.1. and Table 4.2. To simulate the circumstances of in situ observations, the generated well log suite was contaminated with Gaussian distributed noise. The k-th noisy data is computed as, $d_k^{(noisy)} = d_k^{(obs)}\left(1 + \mathrm{N}(0, \sigma(0, \sigma_{d,k})\right)$, where σ_{dk} is the standard deviation proportional to the desired noise level (i.e., 2%, 4% ,6%, 8% ,10%, 13% and 15%).

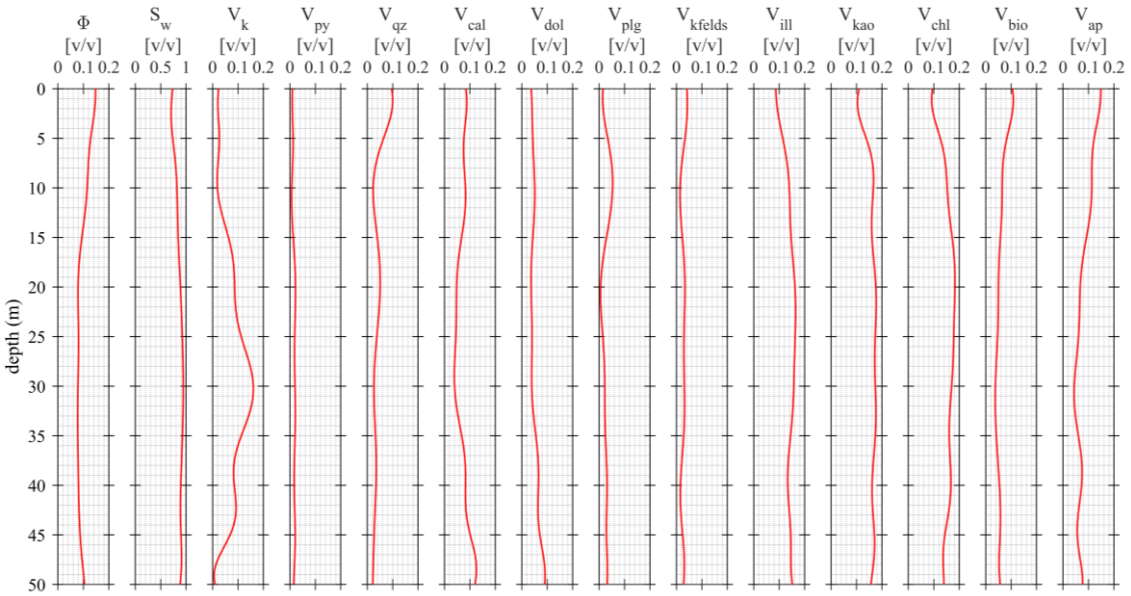

Figure 4.1 Petrophysical parameters of the synthetic model, integrating 12 matrix types, porosity, and water saturation, across 6 proto layers, derived from a fitted Legendre polynomial of degree 15. The dataset resolution is 0.1 m, set to 501 points

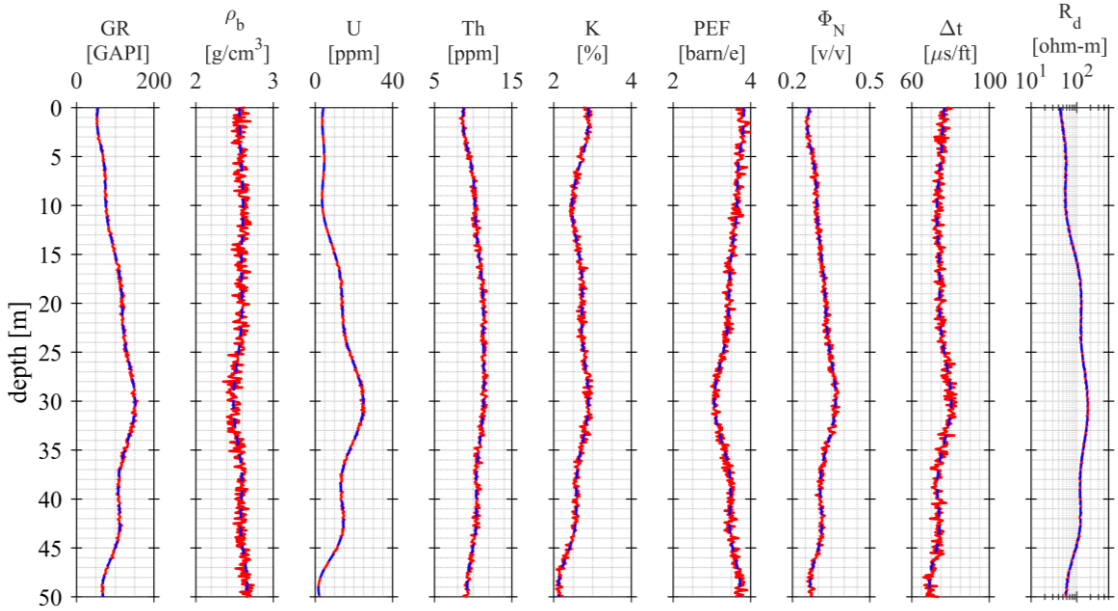

Figure 4.2 Wireline logging dataset modelled from the response function for 9 borehole logging tools. The generated well log suite underwent contamination with varying levels of Gaussian distributed noise. Here 2 % Gaussian noise was added to the synthetic logs

Table 4.2 Zone parameters used for calculating synthetic well logging data											
Wireline log	**Unit**	**Illite**	**Chlorite**	**Kaolinite**	**k-feldspar**	**Plagioclase**	**Dolomite**	**Calcite**	**Biotite**	**Apatite**	**Quartz**
Natural gamma ray (GR)	API	140	130	150	2.5	1.5	0	0	7.5	12.5	2
Potassium (K)	%	5.9	0.15	0.3	16	3	0	0	5	4	0.5
Uranium (U)	ppm	1	0	4	0.2	0	0	0	0	0	0
Thorium (Th)	ppm	17.5	4.5	26.5	10	2	0	0	25	0	0
Photoelectric absorption index (P_e)	barn/e	3.03	4.77	1.47	2.87	3.14	3.13	5.09	8.59	5.87	1.81
Bulk density (ρ_b)	g/cm^3	2.77	2.85	2.64	2.54	2.74	2.87	2.71	3.2	3.17	2.65
Sonic transit-time (Δt)	µs/ft	64.6	55.5	64.3	68.9	45.1	43.9	47.2	55.3	42.9	56
Neutron-porosity (ϕ_N)	v/v	0.158	0.428	0.491	-0.011	-0.018	0.005	0.005	0.225	0.08	-0.02
Deep resistivity (R_d)	Ωm	1	2	1	-	-	-	-	-	-	-

4.2. Synthetic case: Multimineral evaluation

After the quantitative analysis a final solution was given to one of the proposed syntenic logs sets, which well logs are shown in Figure 4.2. Each probe tool (i.e. 9 of them) has a total number of 501 measured points, spaced into 0.1 m, giving a total of 4509 inverted points. The petrophysical variables to be inverted were eight, that is, porosity, water saturation, volumes of kerogen, pyrite, shale, sand, carbonates, and other lithology.

The formation was divided into 20 different layers, giving a total number of 160 unknowns, and a quite high overdetermination ratio 28.18; for which an initial value (Eq. 1.37) of ε was set at 69.07, and a decrement of 0.55 per every iteration (i.e. 25) was selected, reaching a final value of $4.0 \cdot 10^{-5}$ (Figure 4.3).

The initial model parameters for the interval were set as Φ=0.1, S_w=0.8, V_k=0.05, V_{py}=0.02, V_{sh}=0.4, V_{sd}=0.15, V_{carb}=0.20 and V_{other}=0.25; given an initial modeling error of 94 % and reaching a final modeling error of around 13 %, after the 25 iterations; while the initial data error was 371.58 % and reached a final value of 4.64 % (Figures 4.4 and 4.5). The average standard deviation for the estimated petrophysical parameters is as follows, σ_{Φ} =±0.08, σ_{Sw} =±0.01, σ_{Vk} =±0.004, σ_{Vpy} =±0.02, σ_{Vsh} =±0.007, σ_{Vsd} =±0.05, σ_{Vcarb}=±0.07, σ_{Vother} =±0.05; assuming the standard deviations of data proportional to the

noise level (σ_d=0.02). The spread value of the inversion is 0.1035, indicating a low correlation between the coefficients representing the model parameters.

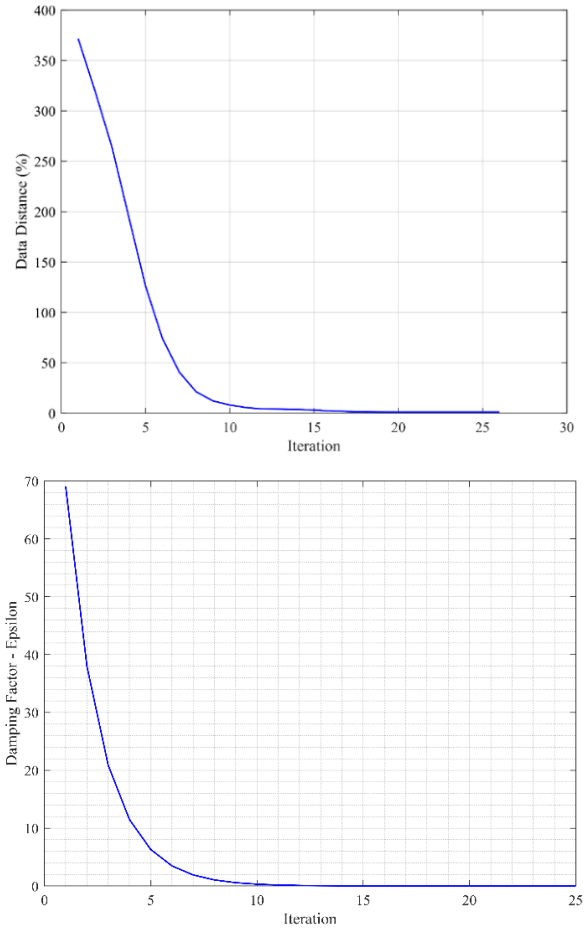

Figure 4.3 Left panel: data distance behavior over the 25 iterations. Right panel: Damping factor (ε) value along the inversion procedure. A 69.07 value was chosen as initial estimation, and a decrement of 0.55 per every iteration was selected, reaching a final value of $4.0 \cdot 10^{-5}$

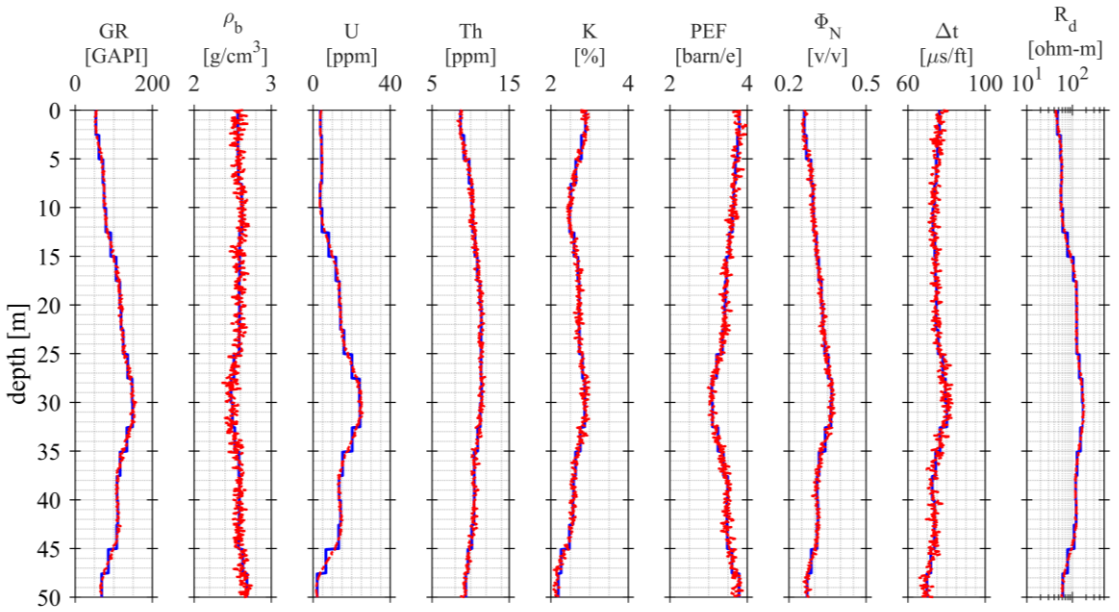

Figure 4.4 Fitted well logs to noisy synthetic data. The formation was divided into 20 different layers, giving a total number of 160 unknowns, and an overdetermination ratio of 28.1

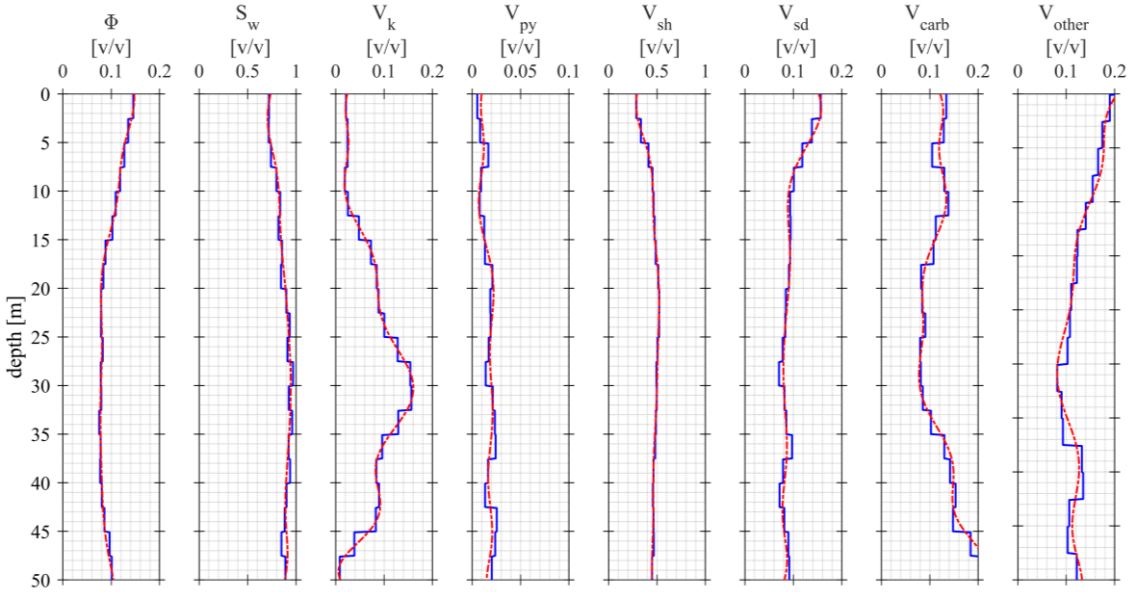

Figure 4.5 Petrophysical parameters estimated by inversion. The initial model parameters for the interval were set as Φ=0.1, S_w=0.8, V_k=0.05, V_{py}=0.02, V_{sh}=0.4, V_{sd}=0.15, V_{carb}=0.20 and V_{other}=0.25; given an initial modeling error of 94 % and reaching a final modeling error of approx. 13 %, after the 25-iteration step

The synthetic case of multimineral evaluation presented here provides an analysis of a set of synthetic logs. The power of the methodology employed in this study is strongly demonstrated by the adjustment of the estimated parameters over the course

of the iterations, resulting in a significant reduction in both modeling and data error rates. These reductions in error rates are indicative of the successful application of the inversion process in refining the model parameters. Also, the average standard deviation for the estimated petrophysical parameters provides a further insight into the accuracy of the model. The mean spread value of the inversion result was found to be relatively low, indicating a poor correlation between the coefficients representing the model parameters, suggesting that the model parameters are independent of each other, which is a desirable outcome in this context.

4.3. Synthetic case: Augmented multimineral evaluation

The original well logging dataset responded not only to a few broad classes of lithologies but mainly on different types of minerals, as in nature. Therefore, I tried to evaluate the 14 components using the synthetic dataset (Figure 4.1). However, it should be mentioned that in real cases, it would be challenging to assume all zone parameters to evaluate in an augmented way the mineral components of the lithology, since the data available is limited to set all the assumptions for the forward modeling.

To maintain a proper overdetermination ratio the formation number of layer was decreased, only 11 different layers were considering this time, since the number of unknows have being increased for each layer to 14, giving a total number of 501 unknows and a ratio of underdetermination of 29.27, slightly higher than the last case (Figure 4.6). The initial value of ε was set to 57.39, and a decrement of 0.55 per every iteration (i.e. 20) was selected, reaching a final value of $6.6 \cdot 10^{-4}$.

The 14 petrophysical variables to be inverted were porosity, water saturation, volumes of kerogen, pyrite, illite, chlorite, kaolinite, quartz, potassium-feldspar, plagioclase, dolomite, calcite biotite, and apatite (Fig. 4.7). The initial model parameters for the interval were set as Φ=0.1, S_w=0.8, V_k=0.05, V_{py}=0.02, V_{ill}=0.13, V_{clo}=0.15, V_{kao}=0.16, V_{qz}=0.04, V_{kfelds}=0.02, V_{plg}=0.03, V_{cal}=0.07, V_{dol}=0.04, V_{bio}=0.06 and V_{ap}=0.09 (v/v). The initial data distance was 218.31 % and reached a final value of 3.11 %, and a final modelling error of around 18 %. Assuming that the standard deviation of data is proportional to level of data noise (σ_d=0.02) the average standard deviation for the estimated petrophysical parameters were obtained as follows: σ_Φ=±0.08, σ_{Sw}=±0.02 σ_{Vk}=±0.004, σ_{Vpy}=±0.02, σ_{Vill}=±0.02, σ_{Vclo}=±0.02, σ_{Vkao}=±0.02, σ_{Vqz}=±0.02, $\sigma_{Vkfelds}$=±0.04, σ_{Vplg}=±0.01, σ_{Vcal}=±0.03, σ_{Vdol}=±0.02, σ_{Vbio}=±0.03 and σ_{Vap}=±0.04 (v/v). The mean spread

value of the inversion result is 0.13, indicating an increase respect the previous case, but still showing low correlation between the coefficients representing the model parameters.

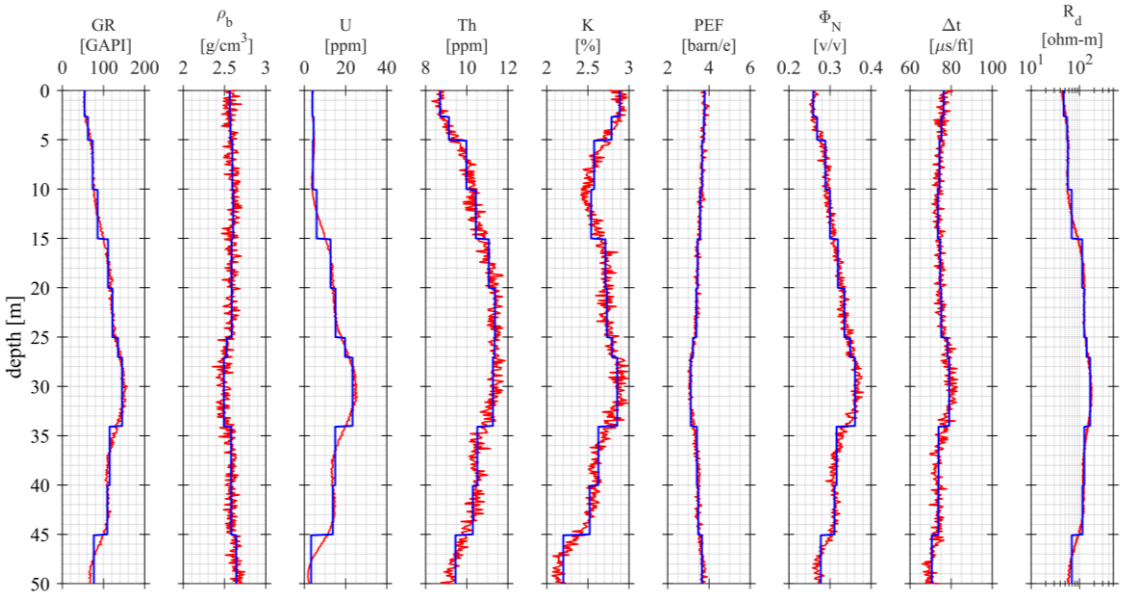

Figure 4.6 Calculated data fitted to noisy synthetic well logs. 11 different layers were considered since the number of unknowns have being increased for each layer to 14, giving a total number of 501 unknows and an overdetermination ratio of 29.27.

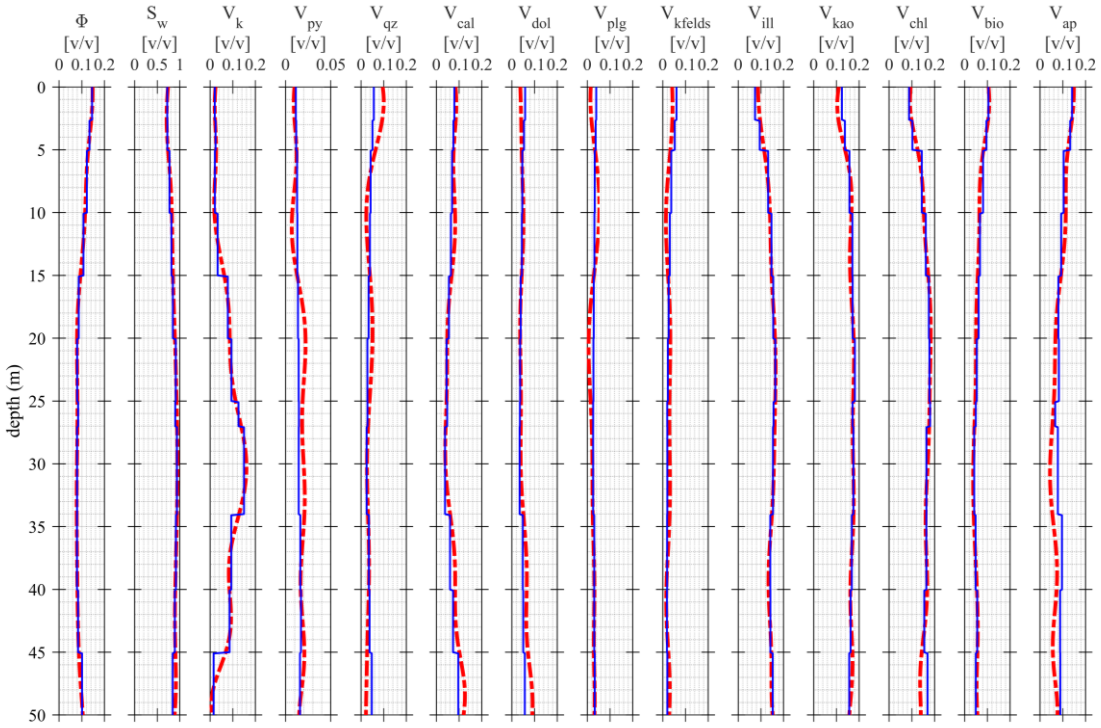

Figure 4.7 The 14 petrophysical variables to be inverted were porosity, water saturation, volumes of kerogen, pyrite, illite, chlorite, kaolinite, quartz, potassium-feldspar, plagioclase, dolomite, calcite biotite, and apatite.

The analysis of this dataset presents unique challenges due to the complexity of the geological formation and the richness of the mineral data. Despite these challenges, I was able to maintain a good overdetermination ratio and reduce data errors significantly. The robustness of inverse modeling and the low correlation between the coefficients suggest that the model is reliable and can be used as a valuable tool in future studies, always that important a priori information is available, if not, only a few parameters should be considered for inversion, the most important for unconventional hydrocarbon systems. This experience highlights the importance of adaptability and precision in geological data analysis.

4.4. Field Test: Lower Congo Basin

An inversion test was performed using data corresponding to a 120 m long interval from Well-1, situated in Lower Congo Basin in Africa, corresponding to the Early-Middle period of the Madingo formation, considered as a tight reservoir (Yang and Yang, 2020). As input for the inversion, the natural gamma-ray intensity (GR), bulk density (ρ_b), neutron-porosity (Φ_N), sonic travel-time (Δt), and deep resistivity (R_d) logs were used, with a sampling space of 0.15 m.

The geological description of the study area can be found in subsection 1.4.3. Furthermore, TOC values from 7 core samples were available, as well as the volume of clay, sandstone and carbonate as petrophysical components of the formation obtained by Yang et al. (2023) using laboratory-supervised analysis with the Strata and Emerge modules of Hampson-Russell software (Figura 4.8).

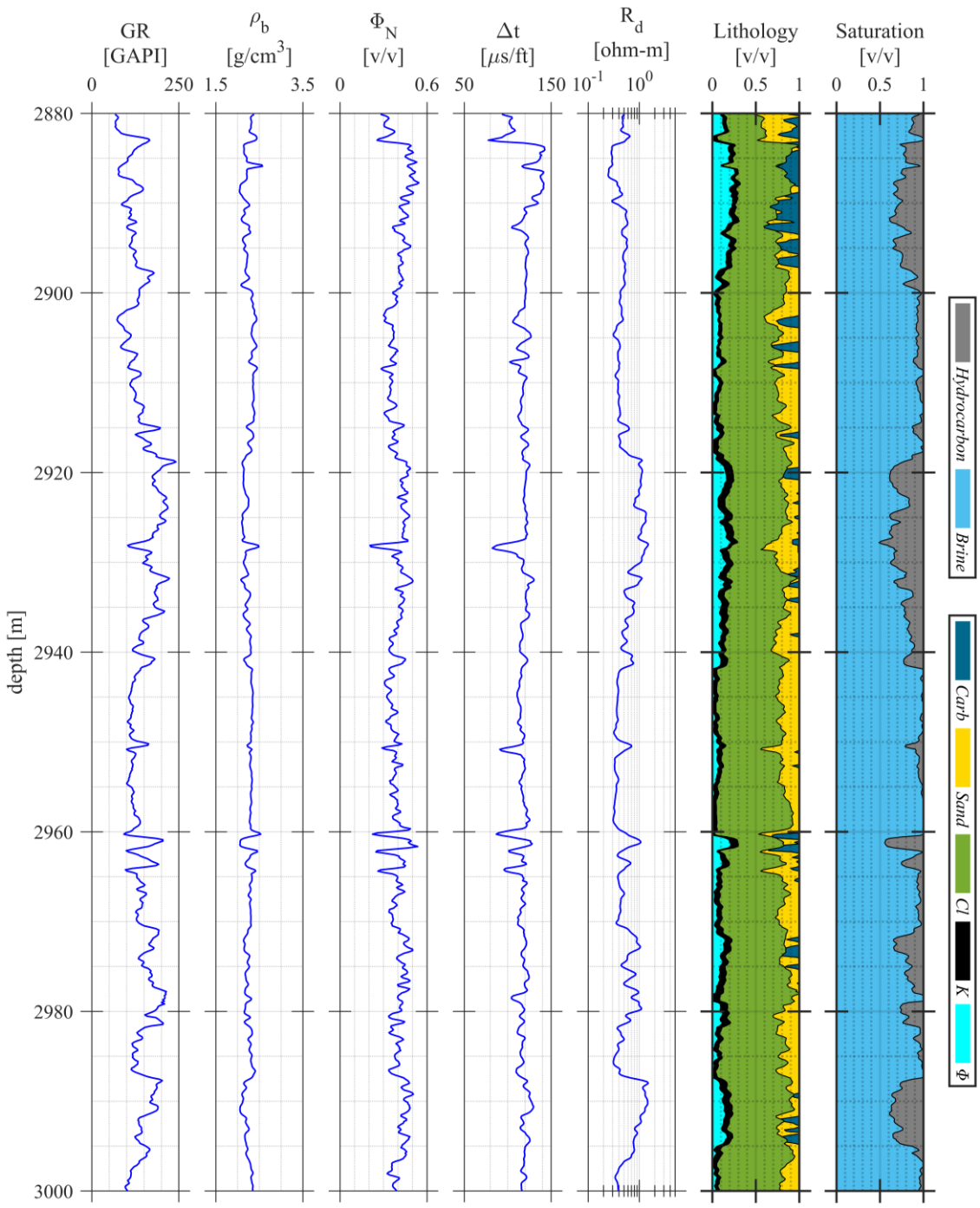

Figure 4.8 Input dataset in Well-1, Madingo formation at Lower Congo Basin, Africa. Denotations are as follows: natural gamma-ray (GR), bulk density (ρ_b), neutron porosity (Φ_N), acoustic travel-time (Δt), and deep resistivity (R_d). On the sixth and seventh track the proportions of lithology and fluids obtained by Yang et al. (2023) are shown. Abbreviations are as follows: carbonate (*Carb*), sandstone (*Sand*), kerogen (*K*), clays (*Cl*) and porosity (*Φ)*

Following the interval inversion procedure the examination of the 5 distinct types of measured well logs against 6 petrophysical parameters were performed. In a classical

inversion procedure, the ODR would be low, leading to an underdetermined inverse problem, providing high uncertainties and inaccuracies solutions. Therefore, the use of interval inversion provides a solution to overcome this problem.

The total number of data points in the analyzed interval was 788, leading to the possibility to increase the ODR up to 10000 % compared to the traditional method, if 5 unknowns are considered to be inverted per layer, that is, porosity, volume of water saturation, kerogen, clays, and carbonates (the sand volume was derived deterministically from the material balance equation).

The analyzed interval was meticulously divided into 18 layers based, mainly, on the natural gamma ray and resistivity logs, and on few geological reports (Yang et al., 2023; Zeng et al., 2024) Consequently, a total of 90 unknown expansion coefficients were applied. The observed dataset and log probes summed up to 3,940 data points, resulting in an overdetermination ratio stands impressively at 43.77, indicating a robust foundation for a stable inversion procedure.

For resolving the forward problem, tool response functions Eqs. (1.19), (1.20), (1.25)-(1.27) were applied. Also, zone parameters were sourced from established literature (Table 4.3). The initial value of epsilon was chosen as 51.93, and a decrement of 0.55 per every iteration was selected, reaching a final value of $6.05 \cdot 10^{-4}$. The initial model parameters for the interval were set as Φ=0.1, S_w=0.9, V_k=0.01, V_{sh}=0.5, V_{carb}=0.3; given initial data error of 40.5 % and reached a final value of 10 % after 20 iterations.

The average standard deviation for the estimated petrophysical parameters is as follows, σ_{Φ} =±0.06, σ_{Sw} =±0.1, σ_{Vk} =±0.005, σ_{Vsh} =±0.02 and σ_{Vcarb} =±0.17; assuming the standard deviations of data a, σ_{GR}=0.08, $\sigma_{\rho b}$=0.05, $\sigma_{\Phi N}$ =0.09, $\sigma_{\Delta t}$=0.06, σ_{RD}=0.06. The mean spread value of the inversion estimates is 0.14, indicating a low correlation between the expansion coefficients representing the model parameters.

The resultant petrophysical parameters of the Early-Middle period of the Madingo formation are graphically presented in Fig. 4.9. A close alignment between observed wireline logs and synthetic logs calculated from the petrophysical model is evident across Tracks 1–5. Furthermore, Tracks 6-11 showcased porosity, fractional volumes of water, and rock constituents, respectively. TOC estimation, crucial for understanding shale properties, was achieved by applying the formula Eq. (1.29) of Tissot and Welte (1978), with a typical kerogen conversion factor (K_{α}) set at 1.2. Notably, the inversion-derived

TOC values closely resembled those obtained through core laboratory measurements, as depicted in Track 12.

Table 4.3. Zone parameters for the Madingo formation, Africa.							
Wireline log	**Unit**	**Brine**	**Hydro-carbon**	**Clays**	**Kerogen**	**Sand**	**Carbonate**
Natural Gamma Ray	API	-	-	64	1592	5	0
Bulk density	g/cm^3	1.09	0.95	2.32	1.19	2.74	2.97
Sonic transit-time	μs/ft	230	149	127	82	53	57
Neutron-porosity	v/v	0.94	0.63	0.43	0.72	-0.03	0.06
Deep resistivity	Ωm	0.022	-	0.189	278	-	-
Textural factor (m, n, a)	-	2	2	1	-	-	-

While the results of the interval inversion are considered satisfactory, a more refined understanding of petrophysical properties, particularly with higher vertical resolution, could be attained by augmenting the number of expansion coefficients and refining zone parameters through additional laboratory measurements. Such enhancements may reduce data distance, though at the expense of increased CPU time, necessitating a delicate balance between precision and computational efficiency.

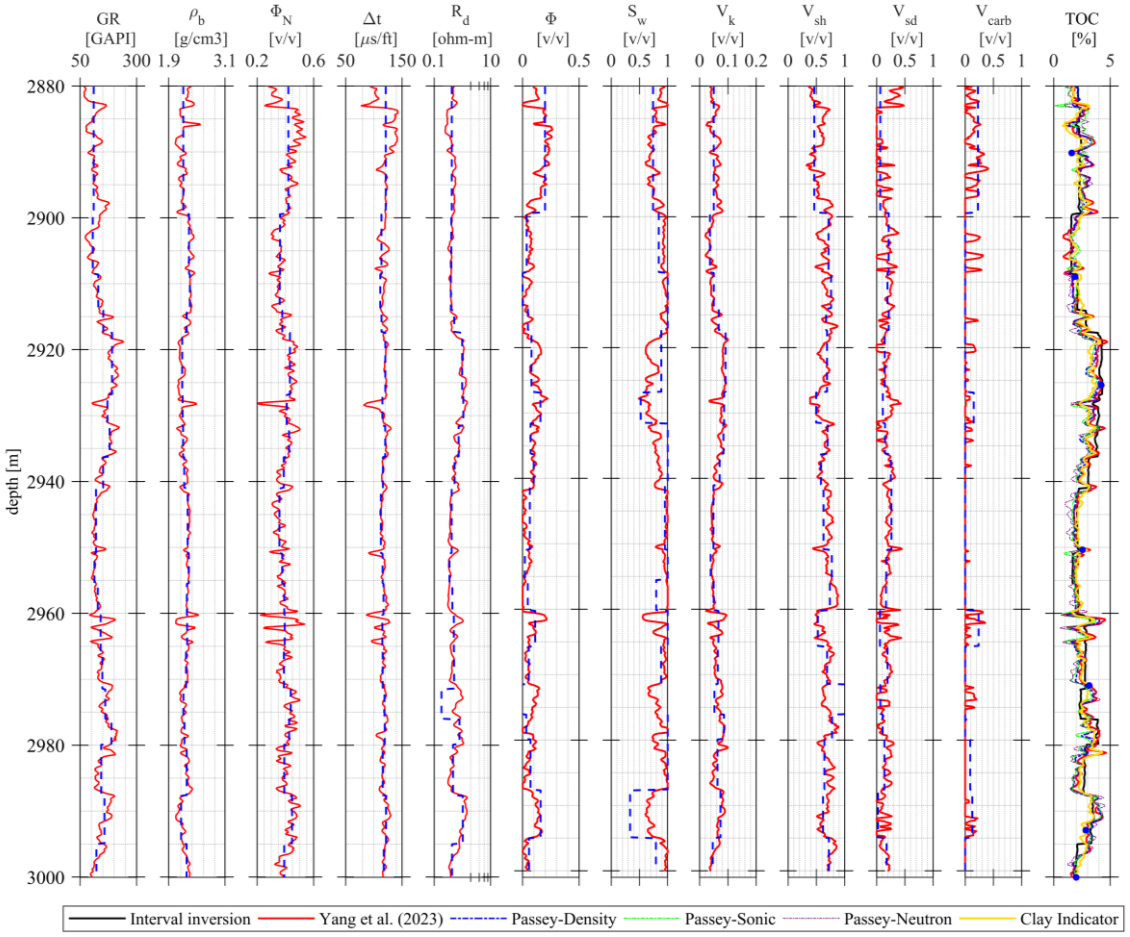

Figure 4.9 Inversion result in Well-1, Madingo formation at Lower Congo Basin, Africa. In tracks 1-11, results of layer-interval inversion are shown in blue, and in read the curves from the dataset. In tract 12, different TOC measurements are shown.

4.5. Summary

This chapter studies well-logging data inversion in unconventional hydrocarbon reservoirs. It introduces the appropriate version of the interval inversion methodology and continues by creating a geological model representing 12 mineral species and two fluids in six proto-layers, essential for heterogeneous shales. The data is then pre-processed, normalized, and reduced to create input for nine borehole log tools, considering noise levels in synthetic data sets.

The synthetic case studies evaluate multimineral scenarios, demonstrating the methodology's success in refining model parameters, reducing errors, and illustrating low correlation between model coefficients. Augmented multimineral evaluation faces

challenges due to geological complexity, yet maintains a good overdetermination ratio, reducing data errors significantly.

Also, the theory was proved by conducting a comprehensive field test in the Lower Congo Basin in Africa, focusing on a 120-meter interval from the Early-Middle period of the Madingo formation. Utilizing natural gamma-ray intensity (GR), bulk density (ρ_b), neutron-porosity (Φ_N), sonic travel-time (Δt), and deep resistivity (R_d) logs and core samples and volumetric parameters of clays, sandstone, and carbonates obtained by Yang et al. (2023) using Hampson-Russell software modules to validate the results obtained from the inversion procedure.

While the results of interval inversion were adequately assessed, I recognized the potential for further refinement, particularly in achieving higher vertical resolution. Augmenting expansion coefficients and refining zone parameters through additional laboratory measurements could enhance the understanding of petrophysical properties, though at the expense of increased computational resources. Striking a delicate balance between precision and computational efficiency remains essential in advancing well-logging inversion in unconventional hydrocarbon systems.

4.6 Thesis III

I implemented a series expansion-based inversion method, traditionally used in conventional reservoirs, for evaluating complex multimineral scenarios in unconventional reservoirs. I estimated the volume of several matrix components and the kerogen volume, that is an important parameter, simultaneously in a joint inversion procedure.

The study assumed a multimineral formation, encompassing 11 mineral species alongside organic matter (kerogen) and two fluid types. Employing an interval inversion approach, I conducted calculations for several numbers of layers —up to 20—involving a substantial petrophysical parameters, reaching up to 11. This process involved accounting for different noise levels in log readings. Utilizing a layer-wise homogeneous model, the methodology enabled accurate estimation of volumetric parameters, including Total Organic Carbon (TOC), ensuring stability in the process and accuracy in the inversion results.

A field case study was conducted to give a better understanding of the inversion methodology, showing an improvement (around 20 %) in TOC estimation based on volume of kerogen inversion compared to classical linear inversion.

5. ESTIMATION OF LEVEL OF MATURITY (LOM) VIA SIMULATED ANNEALING AND TOC CONTENT FROM INTERVAL INVERSION

5.1. Interval Inversion and Simulated Annealing LOM estimation method

Level of Maturity (LOM) is traditionally approximated using the TOC-Δlog$R$ diagram (Figure 5.1) proposed by Passey et al. (1990). The best fitting straight line to the data is given by $\mathrm{TOC} = \alpha_P * (\Delta\mathrm{logR}) + \beta_P$, where ΔlogR is calculated as mentioned in section 1.1.1. in Eq. (1.2).

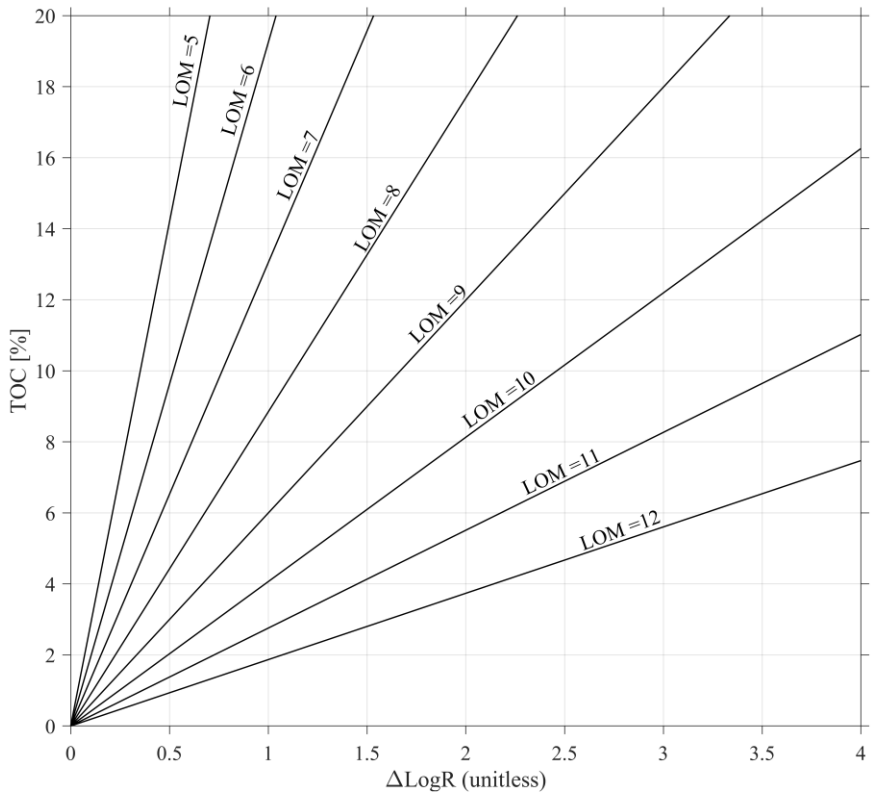

Figure 5.1 Δlog$R$ diagram relating the Δlog$R$ to TOC at different maturity levels (LOM)

The empirical equation for calculating TOC from Δlog$R$ suggested by Passey et al. (1990) Eq. (1.1) does not consider the intercept term of the fitting model equation, represented as β_P. Then, β_P is assumed as zero, which implies a perfect baselining

between the sonic transit-time and resistivity curves in non-source intervals. Then, when equaling the fitting model equation and Eq. (1.1), and solving for LOM, it is obtained, $\mathrm{LOM} = (2.297 - \log_{10} \alpha_P)/0.1688$.

However, the above diagram technique is only possible to be performed if there is a fair amount of TOC values from geochemistry data to calibrate the linear. Therefore, since geochemistry data is scarce and expensive to acquire in sufficient amount, we propose combining interval inversion technique and the SA method to solve the problem (Valadez Vergara and Szabó, 2024). Our inversion workflow is presented in Figure 5.2.

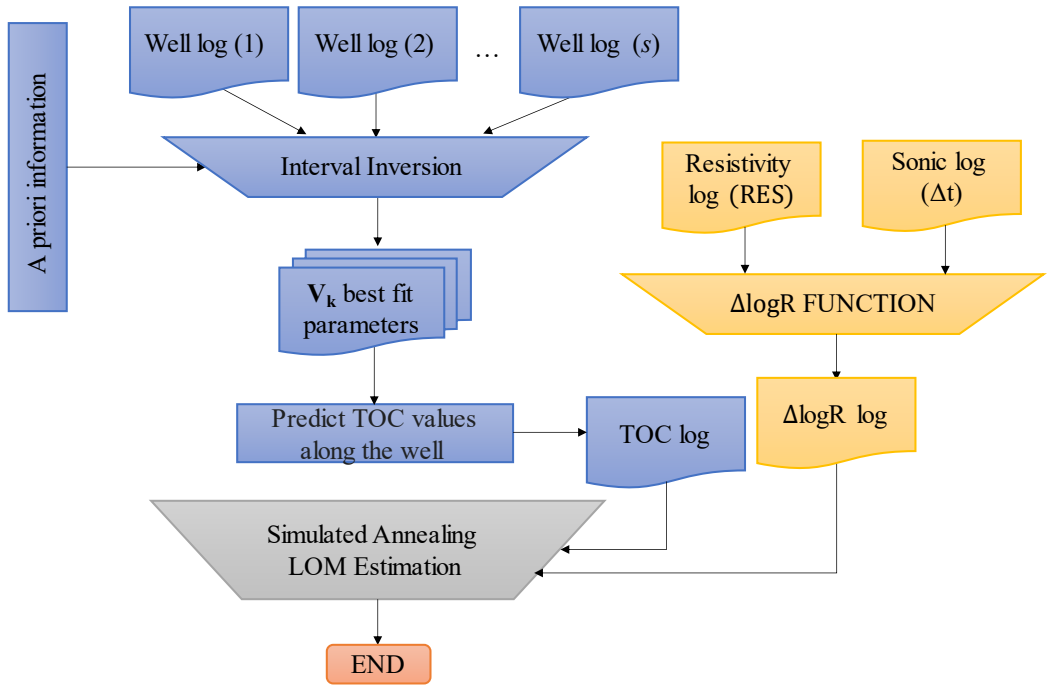

Figure 5.2 Flowchart of the inversion-based Level of Maturity estimation method (LOM-SA)

Through interval inversion, first we estimate the petrophysical parameters of the studied formations. The TOC is derived from the kerogen volume (V_k) using Eq. (1.29), where ρ_k is the kerogen density ($\frac{g}{cm^3}$), K_α is a constant, usually varying from 1.18 to 1.48, and ρ_b is the bulk density measurable using the bulk density (gamma-gamma) log. Then, after obtaining a predicted TOC log curve, it is sought to minimize the objective or energy function that is defined as

$$E = \left(\frac{1}{NS} \sum_{i=1}^{NS} \left(TOC_i^{(m)} - TOC_i^{(c)} \right)^2 \right)^{\frac{1}{2}}, \tag{5.1}$$

where TOC measured ($TOC_i^{(m)}$) is taken from Eq. (1.29) and TOC calculated ($TOC_i^{(c)}$) is evaluated according to Eq. (1.1).

5.2. Case Studies

5.2.1. Field test: Norway

For the suggested inversion procedure, five wireline logs were inverted such as Natural Gamma Ray (GR), Bulk Density (ρ_b), Neutron-porosity (Φ_N), Sonic transit-time (Δt) and Deep resistivity (R_d) logs, also the interval was divided into four layers, marked by the natural gamma-ray intensity changes along the borehole. Boundaries of the 4-layers were chosen at the depth of 328, 439, and 564 m. For the interval inversion procedure, 20 iterations were performed over the 4-layers' intervals with a damping factor (ϵ) of 1000 with a decrement of 0.30 in each iteration. The initial model ($\mathbf{m}_o$) was set at 50% water saturation, 10% porosity, 10 % clay, 10% quartz and 10% carbonate content, and kerogen and pyrite content 1% each. Therefore, seven model parameters were estimated jointly in the inversion procedure. The data distance values calculated using Eq. (1.38) at the last iteration for the 4 layers are given respectively 0.44, 1.62, 0.87, and 1.6. The data calculated by the resultant model is approximated properly as can be seen in Figure 5.3.

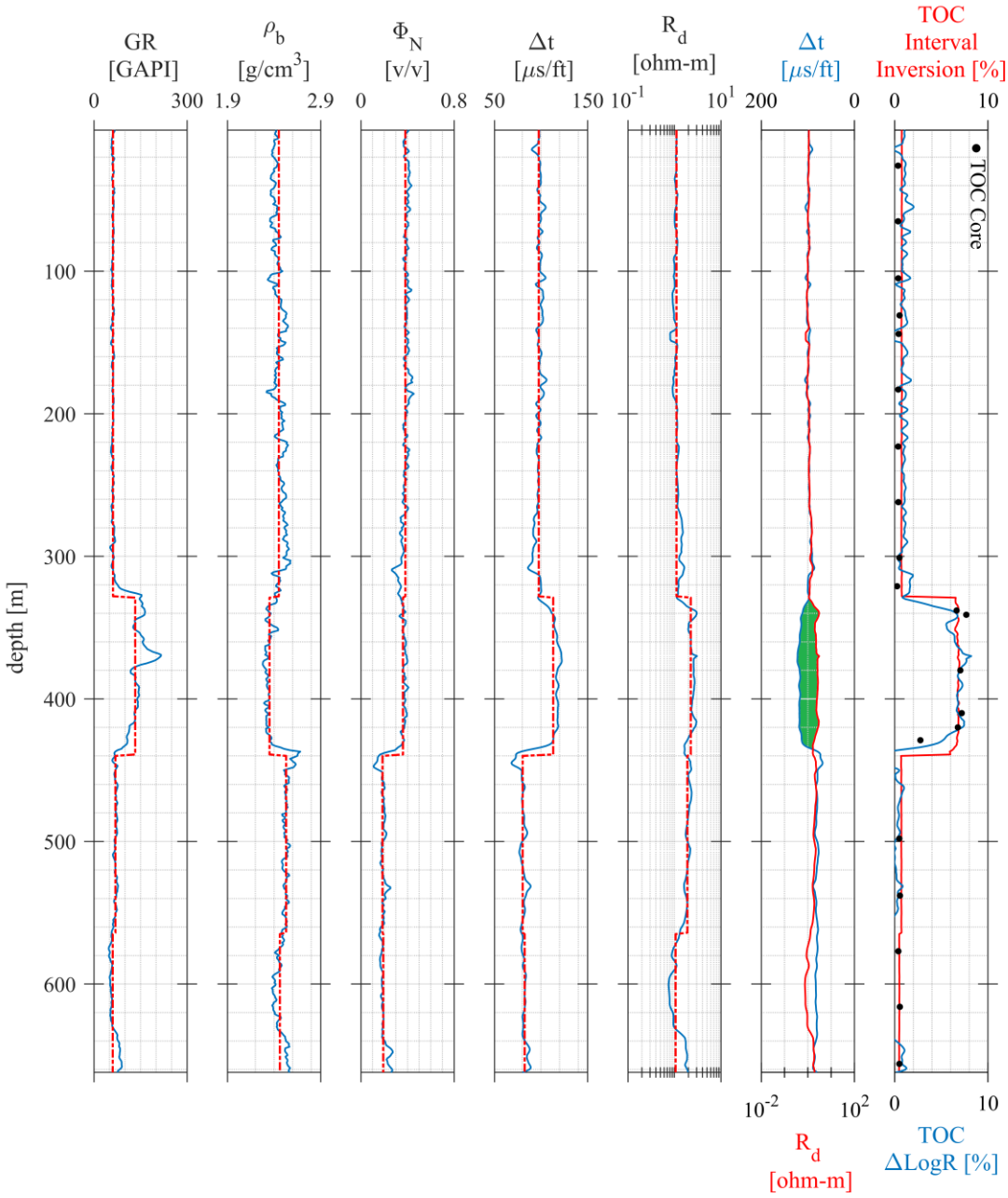

Figure 5.3 Borehole logs originating from the North Sea, Norway. Measured (blue line) and interval inversion estimated (red dashed line) logs are displayed from track 1 to 5. The overlaid Acoustic (Δt) and Resistivity (R_d) logs are displayed on track 6. The separation of the curves (green area) indicates the source rock zone. On track 7, Total organic carbon content logs are presented. In red, the inverted TOC log. In blue, the TOC was obtained by Passey's method. And the black dots indicate the core TOC values

The ΔlogR distance was previously calculated by Valadez Vergara (2020) establishing the baseline interval around 480 m depth (Figure 5.3), and considering the overlap between resistivity and sonic log, with baseline values of 1.99 ohm-m and 81.1 μs/ft,

respectively. TOC content was determined using Eq. (1.29), from the V_k obtained by the interval inversion procedure considering $\rho_k = 1.0$ $(\frac{g}{cm^3})$, $K_\alpha = 1.2$, and ρ_b from the bulk density log (Figure 5.3). In Figure 5.3, the TOC log from the inversion procedure follows a good estimation from other estimations obtained before by Valadez Vergara (2020). The root mean square deviation from the TOC inversion is 0.91, while from the Passey's method or ΔlogR distance method was reported as 1.36 (Valadez Vergara, 2020).

After obtaining the input logs (Figure 5.2) for the Simulating Annealing process, thirty runs were performed with 100,000 iterations in each run (Figure 5.4). The temperature was changed in the q-th iteration as $T^{(new)} = T_0/\lg(q)$, where the initial temperature T_0 is experimentally set to $1.5 \cdot 10^{-1}$. The result from minimizing the energy function established in Eq. (5.1) gave an average of 8.2334 for the value of LOM (~0.6 *Ro*).

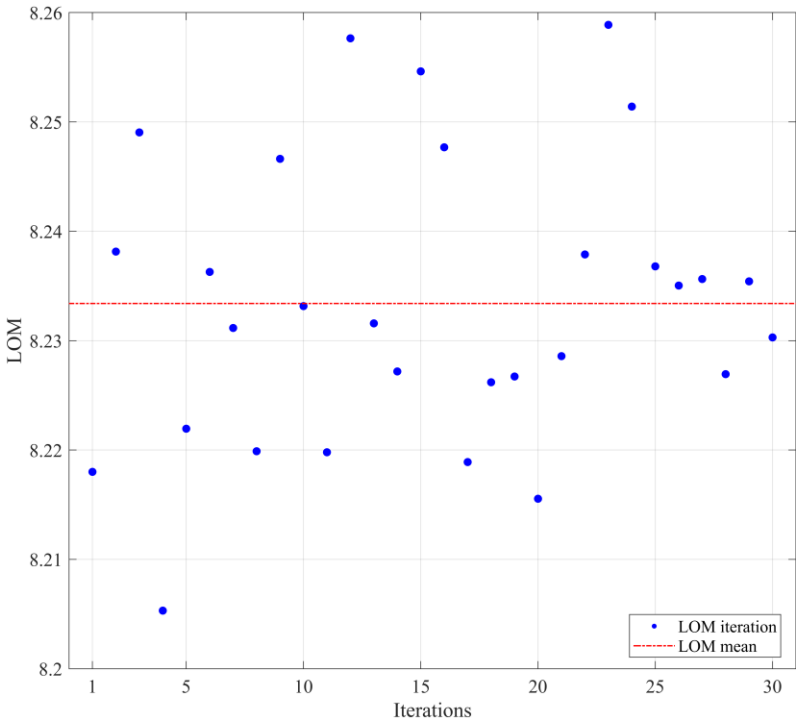

Figure 5.4 Average value of LOM resulting from the 30 LOM-SA runs (blue dots), the red line marks the average value of the LOM estimation results

The measure of central tendency of the estimated LOM are mean equal to 8.2334, the most frequent value (Steiner, 1991) is 8.2053, and the median 8.2324, which indicates a positively skewed distribution. The skewness value is 0.2024, which indicates closely Gaussian distributed data; also, its positive value confirmed that the data is skewed right,

which means that the right tail is long relative to the left tail, while the kurtosis value is 2.5697, indicating more peaked distribution compared to the normal distribution (Figure 5.5).

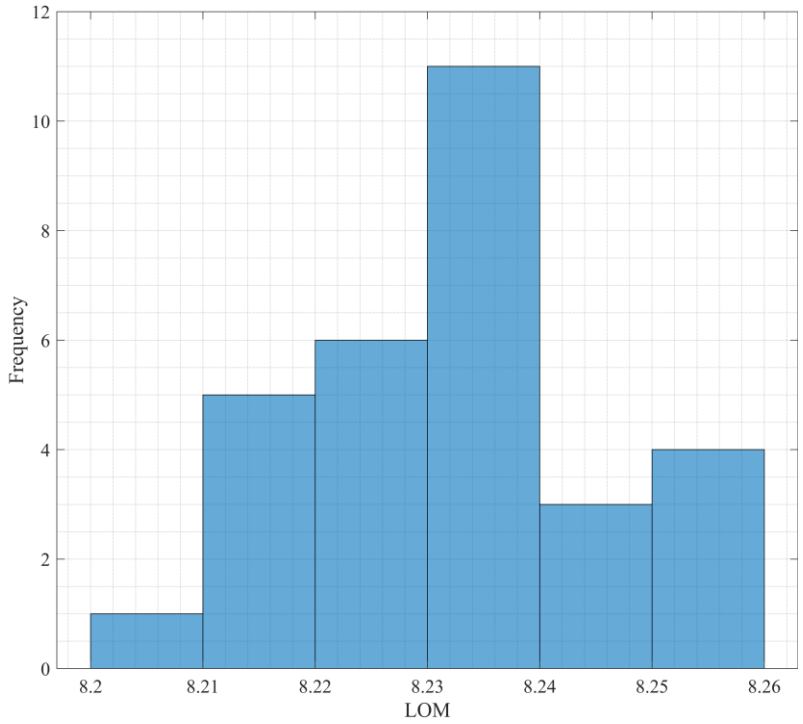

Figure 5.5 Histogram of the LOM sample values after 30 iteration runs of LOM-SA method.

The 95% confidence interval of the calculated LOM values around the mean is LOM=8.2334±0.0047. The standard deviation calculated from the thirty runs is 0.0131 and it has a variance of $1.718 \cdot 10^{-4}$. The minimum LOM value is 8.2053 and the maximum 8.2589 giving a range of 0.0536. The 25^{th} and 75^{th} percentile values are 8.2262 and 8.2381, respectively, and the second (50) percentile is 8.2324. Outliers can be considered as elements more than 1.5 interquartile ranges above the upper quartile (75 %) or below the lower quartile (25 %), given a total of three outliers to the trend (Figure 5.6). However, if we consider Grubbs' method (Grubbs, 1969) to identifies an outlier, we have zero, since it calculates the difference between the value and the mean, and then dividing that difference by the standard deviation and compared each sample so when that ratio is too large, the value is defined to be an outlier. This method is more appropriate since it is applicable to data that is normally distributed, as it is the case.

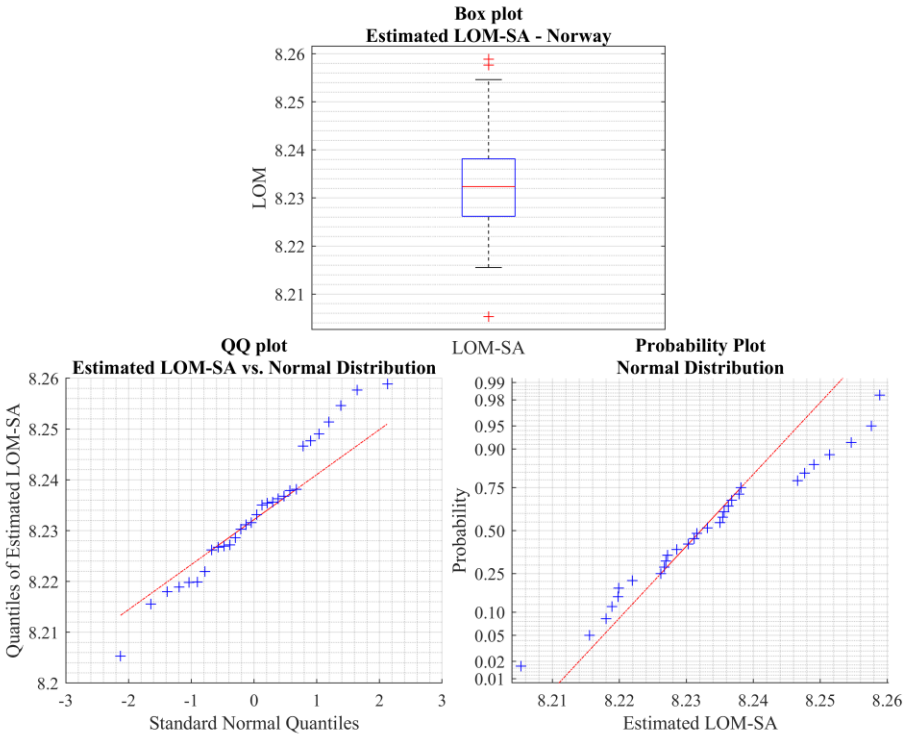

Figure 5.6 Top panel: box plot showing the first (8.2262) and third quartile (8.2381). The median is 8.2324. Bottom panels: QQ and probability plot showing normal distribution of LOM-SA

Quantile-quantile plot (QQ-plot) was used to evaluate how well the distribution of the dataset matches a standard normal distribution, and it can be confirmed that the data tends to the normal distribution although it is skewed and with a heavy tails' behavior (Figure 5.6). The normal distribution parameters were estimated (Figure 5.6), that is, the mean (μ) and standard deviation (σ), and the 95% confidence intervals for the parameters, given the following values respectably, 8.2334, 0.0131, and [8.22851, 8.2383], [0.0104386, 0.0176201].

The average data distance after the thirty runs with 100,000 iterations is 30.68 % (Figure 5.7). The root-mean-square deviation of the differences between sample values predicted by the model and the values observed is 0.7916. The residuals have a mean of 0.3621, a standard deviation of 0.8069, and a variance of 0.6511, a mode of 0.4862, and a median of 0.2699.

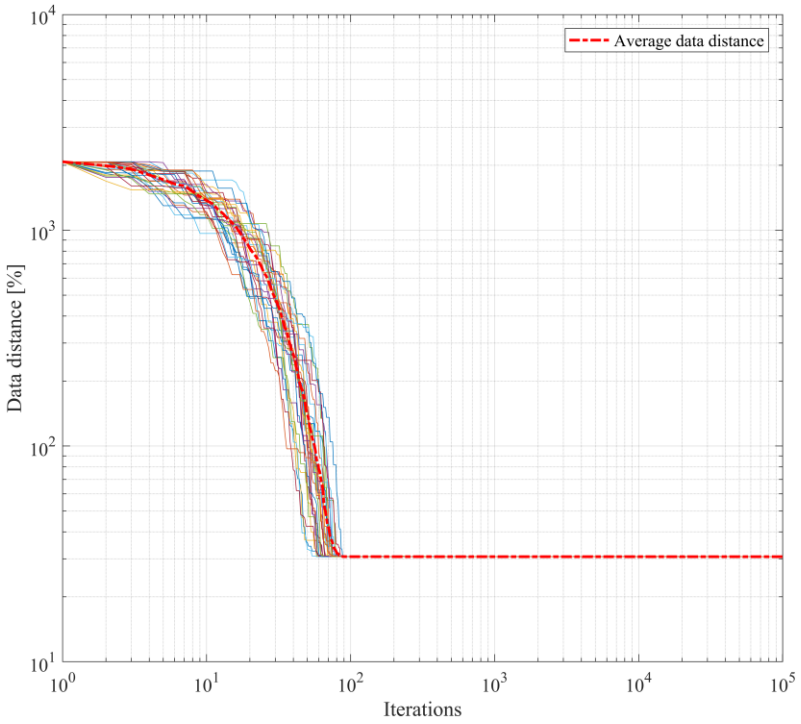

Figure 5.7 Data distance curves calculated during the LOM-SA procedures are obtained over 100,000 iterations per procedure. The strong red dashed line is the average data distance from the thirty runs

The organic richness or TOC content, and the current and past maturity level of the formation are two crucial factors that determine whether a particular rock will be a good source rock, and they are needed to assess the capacity of a source rock to produce hydrocarbon (Passey et al., 2010). However, borehole geophysical instruments have not been used to evaluate these petrophysical variables directly, including some other volumetric metrics (e.g., porosity, shale volume, water saturation, etc.).

There have been many attempts to overcome the problem of lack of direct measurements of petrophysical parameters and other relevant ones, as, textural properties and zone parameters (e.g., LOM), and innovative techniques have been developed (Szabó and Dobróka, 2020; Szabó et al., 2021; Szabó et al., 2022b) showing to be useful tools for detecting and characterizing unconventional reservoirs. The preceding results here demonstrate that the interval inversion technique method enables an accurate estimation of TOC, but also fractional volumes of pore spaces, water, hydrocarbon, and mineral composition of unconventional formations. Although, there is a lack of core data on the fractional volumes of the mineral composition, porosity, and water saturation; the results obtained correlated quite well with the ones that are found in the literature of the area.

For instance, Storebø (2021) reported that water saturations at the North Sea Lower clay rich carbonate reservoirs are ranging from 60 to 80 percent, and porosities around 25 percent, while Worden H., et al., (2020) reported porosities values around 20 percent, which could reach up to 30 percent in some areas. The mineral composition of the formations in that area is mainly dominated by calcium minerals and clays (Worden et al., 2020), with a significant amount of pyrite (~8%) reaching up to 17 percent in some samples (Storebø, 2021).

In practice, LOM is calculated from core samples on lab (vitrinite reflectance) or using the well-known relationship of Passey's method by establishing an empirical relationship that relays TOC from core data. Our proposed method for predicting LOM is based on the estimation of TOC through interval inversion and approximating the global optimum of the energy function proposed in Eq. (5.1). Using the suggested methodology, it is provided a straightforward estimation of LOM that can be independent of the TOC from cores, since it is proven that TOC from the inversion procedure is a reliable tool for its estimation and the characterizing of shale gas formations. The result from the SA procedure on the Norway region is confirmed by previous studies in the area, which shows a precise correlation with the Vitrinite reflectance gradients for the Norwegian sectors of the Central Graben in the North Sea (Petersen et al., 2013)

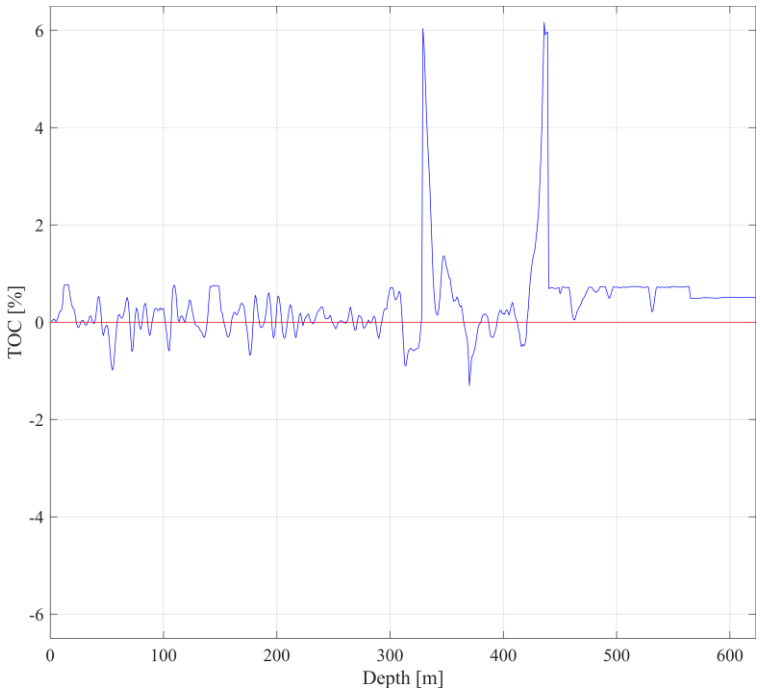

Figure 5.8 8. The blue curve shows the calculated residuals of TOC values for each data point of Norway well. The residuals are calculated from the difference between TOC estimated from LOM-SA and Interval Inversion. The red line marked the deviation zero or perfect estimation

5.2.2 Field test: Alaska, USA

Real data measured at North Inigok 1 well in the set K1 organic-rich shale reservoir in the Kingak Formation was inverted using the interval inversion procedure (Figure 5.9). The TOC results showed a good agreement between the estimated and measured data, indicating that this method can be a valuable tool for evaluating organic-rich shale formations (Figure 5.9). Therefore, we used the estimated TOC log as input for the SA procedure. To estimate the $\Delta \log R$ distance (Figure 5.9), the values for RT and sonic travel-time baseline were chosen as 2.46 (ohm-m) and 112.4 (µs/ft), respectively; from the average measurements of the interval located between 2960 and 2970 m (Figure 5.9).

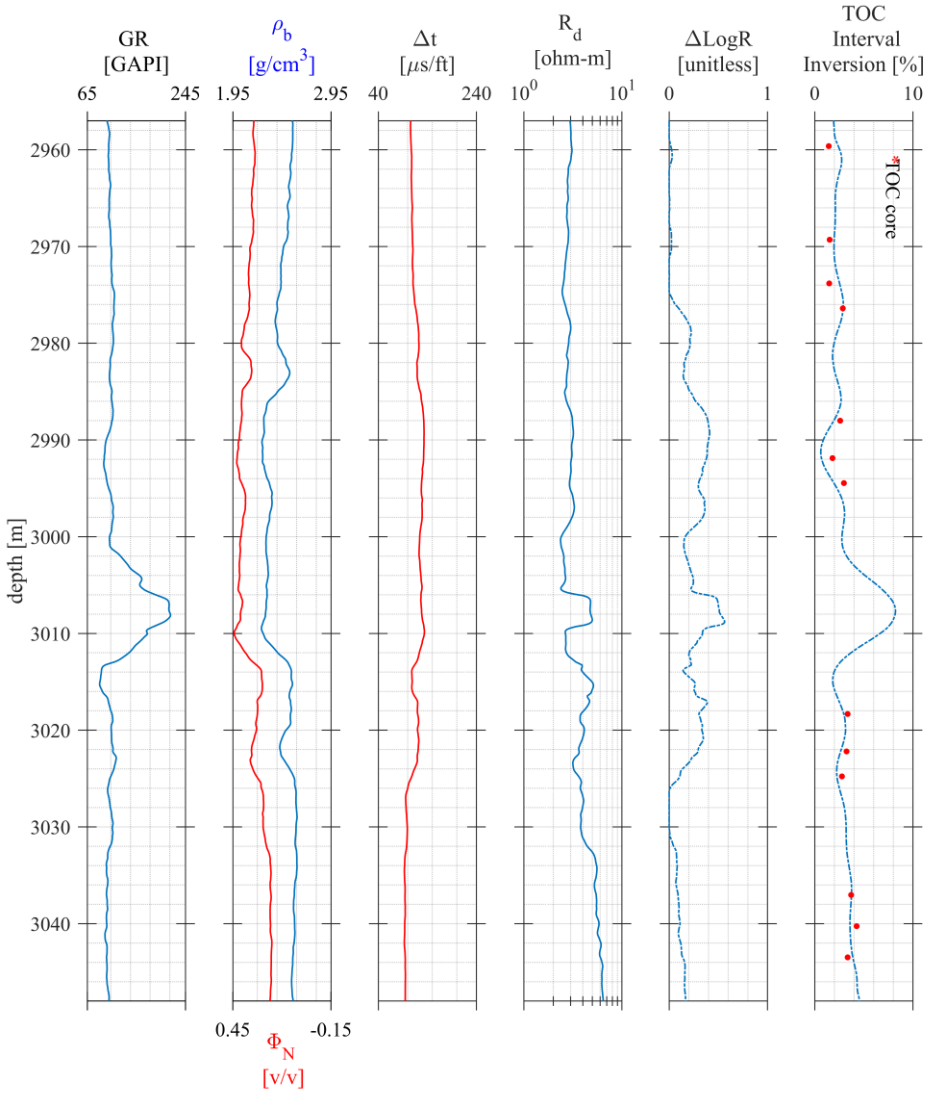

Figure 5.9 12. Borehole geophysical logs from the North Inigok 1 well, Alaska. At the first track Natural Gamma Ray (GR) is displayed. Bulk Density (ρ_b) and Neutron-Porosity (Φ_N)

logs are presented at the second track. Sonic or Acoustic (Δt) log is displayed in the third track. And Deep Resistivity (R_d) is shown on the fourth track. The $\Delta \log R$ log is displayed on the fifth track. Finally, on track sixth, TOC content log (blue dashed line) from interval inversion is presented. In red dots the core TOC values are shown

As in the cases analyzed before, the SA was run multiple times to offer and statistical analysis. However, in this case the number of runs were doubled (60 runs) to increase the significance of the results. Then, after running the Simulating Annealing algorithm the result from minimizing the energy function gave an average LOM value of 36.8026±2.1655 (~3.58 ± 0.07 Ro). The range of the results goes from 21.6183 to 55.7545 LOM, with a median of 37.4593 LOM and a mode of 21.6183.

The normality tests distribution of the results was analyzed using the Anderson-Darling and the Jarque-Bera Test, at a 5% significance level, given both as a result that the null hypothesis is not rejected. That is, for the Jarque-Bera test that the data does not significantly deviate from a normal distribution in terms of these statistical measures (i.e., skewness and kurtosis); as for the Anderson-Darling test, it suggests that the data closely follows the expected distribution (i.e., the normal distribution).

The 25th and 75th percentile have values of -0.0057 and 0.5787, respectively, and the second (50) percentile is 0.2699. They have a kurtosis of 27.8602 and a skewness of 4.2256, and the Quantile-quantile plot (QQ-plot) and the normal distribution plot ensures the normality of the residuals. The residual sum of squares is 517.1630, and the residual standard error is equal to 0.8845, the total sum of squares is $3.1725 \cdot 10^3$, giving coefficient of determination of 0.8370, and a Person's correlation coefficient of 0.9149.

The Alaska dataset results indicate an agreement with the data reporter from the lower section of the Kingak Formation, K1 (Houseknecht and Bird, 2004; Peters et al., 2006; Bird et al. 1999), such section is an over mature formation that has been indicated as the source of the heavy crude oil in Alpine field (Hannon et al., 2000). Also, Peters et al. (2006) mentioned that on the well North Inigok-1 at depth of the section K1, the reduction of the HIo in the basal organic-rich shale to measured values is consistent with a postmature or thermally spent unit. Furthermore, from the USGS report on Vitrinite Reflectance Data, the Kingak Shale shows a diverse distribution of vitrina reflectance values, which goes from 2.01 (14.21 LOM) to 4.26 (63.49 LOM).

It is worthy to mention that Passey's method (1990) was calibrated and tested on worldwide type of lithologies, however, it has been suggested that this technique should

be primary applicable to formations that spawn between the onset of oil generation and over maturity, that is, LOM is between 7 and 12. Even though it can be applied for lower limits it should be cautious on the results, since this means that it might not be as accurate or applicable to source rocks with significantly lower or higher maturity levels, mostly when it comes to TOC estimation, since it under-predicts TOC values in over mature shale gas systems (Sondergeld et al., 2010; Passey et al., 2010).

5.3. Summary

This chapter presents a comprehensive approach for estimating the Level of Maturity (LOM) of geological formations utilizing Simulated Annealing (SA) in conjunction with the Interval Inversion technique to assess Total Organic Carbon (TOC) content. Traditionally, LOM determination relies on the TOC-Δlog$R$ diagram, a method proposed by Passey et al. (1990). However, this approach necessitates sufficient geochemistry data for calibration, which can be costly and challenging to obtain. To avoid this limitation, the proposed methodology integrates interval inversion, which estimates petrophysical parameters including TOC from borehole logs, and SA, which minimizes an energy function to optimize LOM estimation.

Different borehole logs were meticulously analyzed and processed. In the Norway field test, the analysis yielded an average LOM value of 8.2334, which closely aligns with previous studies done in the area. The successful validation of the proposed method in the Norway test highlights its effectiveness in accurately estimating LOM, even in regions with limited geochemical data availability.

Similarly, in the Alaska field test, the proposed approach demonstrated promising results. By utilizing borehole logs and applying interval inversion followed by SA, the estimated LOM for the geological formations in question averaged at 36.8026. This result is consistent with known geological characteristics of the region, validating the reliability of the methodology.

This study features the critical importance of accurately estimating TOC and LOM in evaluating the source rock potential of geological formations. While traditional methods such as the Passey's method have inherent limitations, the proposed technique offers a robust alternative, particularly suited for assessing unconventional reservoirs. By exploring advanced computational algorithms like SA alongside interval inversion, this methodology provides a sophisticated yet accessible means of estimating LOM, paving

the way for more precise characterization of hydrocarbon reservoirs. Furthermore, the findings highlight the potential of this approach to significantly enhance the exploration and development of oil and gas resources, particularly in regions where conventional methods may fall short.

5.4. Thesis IV

I proposed an innovative methodology that combines Interval Inversion and Simulated Annealing techniques to estimate the Level of Maturity (LOM) of geological formations by assessing Total Organic Carbon (TOC) content. By reducing reliance on expensive geochemical data and improving accessibility, my methodology offers a robust and reliable means of LOM estimation, particularly for unconventional reservoirs.

The proposed method centers on increasing the available TOC data by deriving it from kerogen volume (V_k) via Interval Inversion. This expanded data then serves as the foundation for minimizing the RMSE based objective function for LOM values from the Passey's equation, employing randomness within the search process, specifically Simulated Annealing. The feasibility of this method constitutes its primary strengths, allowing not only accurate approximations of LOM values, typically showing a deviation of less than ±1 LOM unit. Furthermore, it offers an automated solution for this zone parameter in the ΔlogR equation compared to arbitrarily selected LOM values.

For validating the well log analysis results, field tests were conducted in both Norway and Alaska, where I have demonstrated the effectiveness and applicability of this methodology across diverse geological contexts.

6. SUMMARY

My doctoral thesis focused on the quantitative analysis of reservoir properties and the application of newly developed petrophysical techniques to address challenges in petroleum geology, particularly emphasizing Total Organic Carbon (TOC) estimation, factor analysis interpretation of wireline logs, interval inversion methodologies, and the assessment of thermal maturity in unconventional hydrocarbon reservoirs. By meticulously addressing the ideas outlined in each thesis, this research has not only

achieved significant goals but has also paved the way for future advancements in hydrocarbon exploration and reservoir characterization.

The theory presented in Thesis I was diligently analyzed, leading to the development of a comprehensive understanding of simulated annealing as a global optimization method for enhancing TOC estimation accuracy. Through comparative analyses with traditional linear (LSQ based) estimation techniques, the efficacy of simulated annealing in improving precision and reducing computational requirements was demonstrated. Practical recommendations for researchers and industry professionals were provided, guiding the selection and application of linear estimation methods for TOC prediction in geological studies. The integration of simulated annealing represents a significant step forward in TOC estimation accuracy and reliability, promising enhanced decision-making capabilities in hydrocarbon exploration efforts.

Thesis II addressed the complex task of interpreting wireline logs and estimating organic matter content in unconventional reservoirs through factor analysis. A novel statistical method and algorithm were developed, demonstrating its effectiveness across diverse geological contexts. The feasibility and reliability of this method were highlighted through comprehensive analyses and validations, laying a solid foundation for its practical application in reservoir characterization problems. The insights gained from Thesis II not only enhance our understanding of unconventional reservoirs but also offer practical solutions for industry professionals seeking to optimize reservoir evaluation processes.

My research results summarized in Thesis III have significantly advanced the field of well logging data inversion for unconventional hydrocarbon reservoirs by exploring into the practical aspect of interval inversion using layer-wise modeling techniques with a primary focus on quantitative analysis. This contributes to enhancing the accuracy and reliability of reservoir characterization, supporting informed decision-making in hydrocarbon exploration and production activities. Through both synthetic modeling experiments and field tests, we have validated the effectiveness and adaptability of this inversion approach across diverse geological settings, analyzing different factors affecting the development and refinement of this method (i.e., overdetermination ratio, number of layers, regularization factor, noise levels). Moving forward, my work sets the stage for further research in refining computational algorithms and extending the applicability of interval inversion to more complex geological scenarios.

The exploration of thermal maturity estimation presented in Thesis IV offers valuable insights into the assessment of hydrocarbon reservoirs. By introducing the innovative approach of interval inversion, linked with simulated annealing-based LOM estimation, this research addresses critical challenges in reservoir evaluation. The methodology's capability to derive in situ estimates of LOM without the need for costly laboratory measurements represents a significant advancement in reservoir characterization. Moreover, the study's future directions offer promising opportunities for further research, including the extension of interval inversion techniques to the evaluation of other reservoir properties related to thermal maturity.

The culmination of my doctoral work captures the interdisciplinary nature of reservoir evaluation and development. By exposing the complexities of reservoir quality assessment, completion quality, and thermal maturity estimation, this research contributes to a holistic understanding of unconventional hydrocarbon reservoirs. The proposed methodologies not only streamline the prospecting phase but also offer practical implications for operational efficiency and cost reduction. Moreover, the study's future directions highlight the ongoing commitment to advancing reservoir characterization methodologies and fostering a more sustainable approach to hydrocarbon exploration.

In the pursuit of a deeper understanding of petroleum geology, this doctoral thesis has explored various methods and techniques to estimate Total Organic Carbon (TOC) content and assess the Level of Maturity (LOM) in unconventional hydrocarbon reservoirs. Through thorough research, careful experimentation, and collaboration, significant progress has been made in improving our ability to predict and understand these vital aspects of oil and gas exploration.

Finally, there exist abundant opportunities for enhancing and implementing these methodologies further. It is imperative that academia, industry, and government stakeholders collaborate closely to propel these advancements and guarantee the sustainable exploration and utilization of energy resources. By improving our ability to estimate TOC content and assess reservoir maturity, we are better equipped to make informed decisions in oil and gas exploration. As we persist in refining and deploying these techniques, we edge closer to realizing a future characterized by sustainable and efficient energy practices.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude and appreciation to all those who have supported and guided me throughout my PhD journey.

I am deeply grateful to my supervisor, Dr. Norbert Péter Szabó, for his unwavering support, insightful guidance, and constant encouragement throughout this research project. His expertise and constructive feedback have been instrumental in shaping this work.

I would like to express my sincere gratitude to István Szabó, my industry supervisor from MOL Plc., for his guidance and support throughout this research project.

I would like to extend my sincere thanks to the MOL Group and the National Research, Development, and Innovation Office (NKFIH) of Hungary for providing financial support through the MOL Grant and Project No. K-135323, which made this research possible.

The research was partly funded by the Sustainable Development and Technologies National Programme of the Hungarian Academy of Sciences (FFT NP FTA).

I am grateful to my colleagues at the Faculty of Earth and Environmental Sciences and Engineering, especially those from the Department of Geophysics, for their insightful contributions, stimulating discussions, and valuable feedback that enriched this research.

Finally, I would like to express my deepest gratitude to my family, especially my parents María de los Angeles and Rafael, whose deep unconditional love and prayers have always been with me; my brother Diego and sister Alejandra for always supporting me and cheering me up; and my friends for their encouragement throughout this challenging yet rewarding journey.

With gratitude and blessings to all.

REFERENCES

Abordán, A., Szabó, N.P., 2018. Metropolis algorithm driven factor analysis for lithological characterization of shallow marine sediments. Acta Geod Geophys 53, 189–199. https://doi.org/10.1007/s40328-017-0210-z

Abordán, A., Szabó, N.P., 2020. Uncertainty reduction of interval inversion estimation results using a factor analysis approach. Int J Geomath 11, 11. https://doi.org/10.1007/s13137-020-0144-4

Bartlett, M.S., 1937. The Statistical Conception of Mental Factors. British Journal of Psychology. General Section 28, 97–104. https://doi.org/10.1111/j.2044-8295.1937.tb00863.x

Bibor, I., and Szabó, N. P., 2016. Unconventional shale characterization using improved well logging methods. Geosciences and Engineering. 5. 32-50.

Bird K., et al. 1999. Chapter VR Thermal Maturity. In The Oil and Gas Resource Potential of the 1002 Area, Arctic National Wildlife Refuge, Alaska. US Geological Survey Open File Report. 98-34

Brennand, T.P., Hoorn, B. van, James, K.H., Glennie, K.W., 1998. Historical Review of North Sea Exploration, in: Petroleum Geology of the North Sea. John Wiley & Sons, Ltd, pp. 1–41. https://doi.org/10.1002/9781444313413.ch1

Deo, A.S., Walker, I.D., 1995. Overview of damped least-squares methods for inverse kinematics of robot manipulators. J Intell Robot Syst 14, 43–68. https://doi.org/10.1007/BF01254007

Detterman, R.L.; Reiser, H.N.; Brosge, W.P.; Dutro, J.T., Jr. Post-Carboniferous Stratigraphy, Northeastern Alaska; US Geological Survey Professional Paper 886; USGS Publications Warehouse: Reston, VA, USA, 1975; 46p.

Dobróka, M., Szabó, N., Turai, E., 2012. Interval inversion of borehole data for petrophysical characterization of complex reservoirs. Acta Geodaetica et Geophysica Hungarica 47, 172–184.

Dobróka, M., Szabó, N.P., 2001. The inversion of well log data using Simulated Annealing method. Publs. Univ. Miskolc Geosci. A Min 59, 115–137.

Dobróka, M., Szabó, N.P., 2012. Interval inversion of well-logging data for automatic determination of formation boundaries by using a float-encoded genetic algorithm. Journal of Petroleum Science and Engineering 86–87, 144–152. https://doi.org/10.1016/j.petrol.2012.03.028

Dobróka, M., Szabó, N.P., Tóth, J., Vass, P., 2016. Interval inversion approach for an improved interpretation of well logs. Geophysics 81, D155–D167.

Dong, Z., Holditch, S.A., John Lee, W., 2016. Chapter 2 - World Recoverable Unconventional Gas Resources Assessment, in: Ma, Y.Z., Holditch, S.A. (Eds.), Unconventional Oil and Gas Resources Handbook. Gulf Professional Publishing, Boston, pp. 53–70. https://doi.org/10.1016/B978-0-12-802238-2.00002-X

Freedman DA. Statistical Models: Theory and Practice. 2nd ed. Cambridge University Press; 2009.

Grubbs FE (1969) Procedures for Detecting Outlying Observations in Samples. Technometrics, 11(1):1–21. https://doi.org/10.2307/1266761

Hannon, R. C., D. A. Gingrich, S. R. Redman, K. P. Helmold, W. J. Campaign, and C. B. Dotson, 2000, The discovery and delineation of a new Alaskan reservoir: The Alpine field, North Slope, Alaska— U.S.A. (abs.): AAPG Annual Convention Program, v. 9, p. A63.

Henderson, D., Jacobson, S., Johnson, A., 2006. The Theory and Practice of Simulated Annealing, in: Handbook of Metaheuristics. pp. 287–319. https://doi.org/10.1007/0-306-48056-5_10

Hood, A., Gutjahr, C.C.M., Heacock, R.L., 1975. Organic Metamorphism and the Generation of Petroleum1. AAPG Bulletin 59, 986–996. https://doi.org/10.1306/83D91F06-16C7-11D7-8645000102C1865D

Houseknecht, D.W., Bird, K.J., 2004. Sequence stratigraphy of the Kingak Shale (Jurassic-Lower Cretaceous), National Petroleum Reserve in Alaska. American Association of Petroleum Geologists Bulletin 88, 279–302.

Jatiault, R., Loncke, L., Dhont, D., Imbert, P., Dubucq, D., 2019. Geophysical characterisation of active thermogenic oil seeps in the salt province of the lower Congo basin part I: Detailed study of one oil-seeping site. Marine and Petroleum Geology 103, 753–772. https://doi.org/10.1016/j.marpetgeo.2018.11.026

Jiang, S., Mokhtari, M., Borrok, D., and Lee, J.,2018. Improving the Total Organic Carbon Estimation of the Eagle Ford Shale with Density Logs by Considering the Effect of Pyrite. Minerals, 8(4), 154. https://doi.org/10.3390/min8040154

Jöreskog, K.G., 1963. Statistical Estimation in Factor Analysis: A New Technique and Its Foundation. Stockholm.

Jöreskog, K.G., Olsson, U.H., Wallentin, F.Y., 2016. Exploratory Factor Analysis (EFA), in: Jöreskog, K.G., Olsson, U.H., Y. Wallentin, F. (Eds.), Multivariate Analysis with

LISREL, Springer Series in Statistics. Springer International Publishing, Cham, pp. 257–282. https://doi.org/10.1007/978-3-319-33153-9_6

Kentucky Geological Survey, University of Kentucky (2023, March 3). Vitrinite reflectance. https://www.uky.edu/KGS/coal/coal-analyses-vitrinite.php

Khan, S., Bibi, Z., 2016. Shale Gas Characterization of Sembar Formation, Khipro Area, Pakistan. International Journal of Geosciences 07, 1009–1019. https://doi.org/10.4236/ijg.2016.78076

Kiss, K., Magyar, I., Nusszer, A., Szuromin´e Korecz, A., Vincze, M., Papp, K., 2012. New Research Results on the Miocene Formations in Derecske Depression (In Hungarian). Nosztalgeo Konferencia Szeged (oral presentation)

Kohli, A.H., Zoback, M.D. (Eds.), 2019a. Composition, Fabric, Elastic Properties and Anisotropy, in: Unconventional Reservoir Geomechanics: Shale Gas, Tight Oil, and Induced Seismicity. Cambridge University Press, Cambridge, pp. 31–64. https://doi.org/10.1017/9781316091869.003

Kohli, A.H., Zoback, M.D. (Eds.), 2019b. Introduction, in: Unconventional Reservoir Geomechanics: Shale Gas, Tight Oil, and Induced Seismicity. Cambridge University Press, Cambridge, pp. 3–30. https://doi.org/10.1017/9781316091869.002

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. (2004) Applied Linear Statistical Models. 5th Edition, McGraw-Hill, Irwin, New York.

Lawley, D.N., Maxwell, A.E., 1962. Factor Analysis as a Statistical Method. Journal of the Royal Statistical Society. Series D (The Statistician) 12, 209–229. https://doi.org/10.2307/2986915

Lee, K.S., Kim, T.H., 2019. Chapter 2 - Petrophysical Characteristics of Shale Reservoirs, in: Lee, K.S., Kim, T.H. (Eds.), Transport in Shale Reservoirs. Gulf Professional Publishing, pp. 7–34. https://doi.org/10.1016/B978-0-12-817860-7.00002-4

Ma, Y.Z., Moore, W.R., Gomez, E., Clark, W.J., Zhang, Y., 2016. Chapter 14 - Tight Gas Sandstone Reservoirs, Part 1: Overview and Lithofacies, in: Ma, Y.Z., Holditch, S.A. (Eds.), Unconventional Oil and Gas Resources Handbook. Gulf Professional Publishing, Boston, pp. 405–427. https://doi.org/10.1016/B978-0-12-802238-2.00014-6

Mao, Z.-Q., 2001. The Physical Dependence and The Correlation Characteristics Of Density And Neutron Logs. Petrophysics 42

Marquardt, D. W., 1959. Solution of non-linear chemical engineering models: Chemical Engineering Progress, 55, 65–70.

Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press Inc.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. 21, 1087–1092. https://doi.org/10.1063/1.1699114

Moler, C.B., 2010. Numerical Computing with MATLAB: Revised Reprint. SIAM.

Montgomery, D.C., Peck, E.A. and Vining, G.G. (2012) Introduction to Linear Regression Analysis (5th ed.), Wiley.

Norwegian Petroleum Directorate (2023, March 15). FactPages: Wellbore.

Passey, Q. R., Bohacs, K., Esch, W. L., Klimentidis, R., and Sinha, S., 2010. From Oil-Prone Source Rock to Gas-Producing Shale Reservoir—Geologic and Petrophysical Characterization of Unconventional Shale Gas Reservoirs. SPE-131350-MS, 29. https://doi.org/10.2118/131350-MS

Passey, Q.R., Creaney, S., Kulla, J.B., Moretti, F., and Stroud, J.D., 1990. A Practical Model for Organic Richness from Porosity and Resistivity Logs.

Peters, K.E., Magoon, L.B., Bird, K.J., Valin, Z.C., Keller, M.A., 2006. North Slope, Alaska: Source rock distribution, richness, thermal maturity, and petroleum charge. AAPG Bulletin 90, 261–292. https://doi.org/10.1306/09210505095

Peters, K.E., Xia, X., Pomerantz, A.E., Mullins, O.C., 2016. Chapter 3 - Geochemistry Applied to Evaluation of Unconventional Resources, in: Ma, Y.Z., Holditch, S.A. (Eds.), Unconventional Oil and Gas Resources Handbook. Gulf Professional Publishing, Boston, pp. 71–126. https://doi.org/10.1016/B978-0-12-802238-2.00003-1

Petersen, H.I., Holme, A.C., Andersen, C., Whitaker, M.F., Nytoft, H.P., Thomsen, E., 2013. The source rock potential of the Upper Jurassic– lowermost Cretaceous in the Danish and southern Norwegian sectors of the Central Graben, North Sea. First Break 31. https://doi.org/10.3997/1365-2397.2013011

Reiser, H.N., Brosge, W.P., Jr, J.T.D., Detterman, R.L., 1980. Geologic map of the Demarcation Point Quadrangle, Alaska. IMAP. https://doi.org/10.3133/i1133

Rezaee, R., 2022. Editorial on Special Issues of Development of Unconventional Reservoirs. Energies 15, 2617. https://doi.org/10.3390/en15072617

Rouse, W.A., Houseknecht, D.W., 2016. Modified method for estimating petroleum source-rock potential using wireline logs, with application to the Kingak Shale, Alaska North Slope (No. 2016–5001), Scientific Investigations Report. U.S. Geological Survey. https://doi.org/10.3133/sir20165001

Schmoker, J. W., 1979. Determination of Organic Content of Appalachian Devonian Shales from Formation-Density Logs: GEOLOGIC NOTES. AAPG Bulletin, 63(9), 1504–1509. https://doi.org/10.1306/2F9185D1-16CE-11D7-8645000102C1865D

Sondergeld, C.H., Newsham, K.E., Comisky, J.T., Rice, M.C., Rai, C.S., 2010. Petrophysical Considerations in Evaluating and Producing Shale Gas Resources. Presented at the SPE Unconventional Gas Conference, OnePetro. https://doi.org/10.2118/131768-MS

Steiner F. 1991. The Most Frequent Value: Introduction to a Modern Conception of Statistics. Akadémiai Kiadó.

Storebø, E. M., 2021. Effect of pyrite on fluid distribution and permeability modelling of North Sea Lower Cretaceous clay rich carbonate reservoirs. Technical University of Denmark, Department of Civil Engineering.

Szabó, N.P., 2011. Shale volume estimation based on the factor analysis of well-logging data. Acta Geophys. 59, 935–953. https://doi.org/10.2478/s11600-011-0034-0

Szabó, N.P., Abordán, A., Dobróka, M., 2022a. Permeability extraction from multiple well logs using particle swarm optimization based factor analysis. Int J Geomath 13, 10. https://doi.org/10.1007/s13137-022-00200-x

Szabó, N.P., Dobróka, M., 2013. Extending the Application of a Shale Volume Estimation Formula Derived from Factor Analysis of Wireline Logging Data. Math Geosci 45, 837–850. https://doi.org/10.1007/s11004-013-9449-2

Szabó, N.P., Dobróka, M., 2018. Exploratory Factor Analysis of Wireline Logs Using a Float-Encoded Genetic Algorithm. Math Geosci 50, 317–335. https://doi.org/10.1007/s11004-017-9714-x

Szabó, N.P., Dobróka, M., 2020. Interval inversion as innovative well log interpretation tool for evaluating organic-rich shale formations. Journal of Petroleum Science and Engineering 186, 106696. https://doi.org/10.1016/j.petrol.2019.106696

Szabó, N.P., Dobróka, M., Drahos, D., 2012. Factor analysis of engineering geophysical sounding data for water-saturation estimation in shallow formations. GEOPHYSICS 77, WA35–WA44. https://doi.org/10.1190/geo2011-0265.1

Szabó, N.P., Remeczki, F., Jobbik, A., Kiss, K., Dobróka, M., 2022b. Interval inversion based well log analysis assisted by petrophysical laboratory measurements for evaluating tight gas formations in Derecske through, Pannonian basin, east Hungary. Journal of Petroleum Science and Engineering 208, 109607. https://doi.org/10.1016/j.petrol.2021.109607

Szabó, N.P., Valadez-Vergara, R., Tapdigli, S., Ugochukwu, A., Szabó, I., Dobróka, M., 2021. Factor Analysis of Well Logs for Total Organic Carbon Estimation in Unconventional Reservoirs. Energies 14, 5978. https://doi.org/10.3390/en14185978

Tissot, B.P., Welte, D.H., 1978. Petroleum Formation and Occurrence. Springer.

Ugochukwu, A.A. (2021) Application of Factor Analysis for Estimation of Shale Volume and Total Organic Carbon in Unconventional Hydrocarbon Reservoirs. Master Thesis. University of Miskolc. (Accessed: 2nd February, 2023)"

Valadez Vergara, Rafael., 2020. Quantification of Source Rock Potential Using Wireline Logs [Master dissertation, University of Miskolc]. MIDRA a Miskolci Egyetem Digitális Raktára és Adattára

Valadez Vergara, R., Szabó, N.P., 2024. Level of thermal maturity estimation in unconventional reservoirs using interval inversion and simulating annealing method. Acta Geophys. https://doi.org/10.1007/s11600-024-01413-4

Worden, R.H., Allen, M.J., Faulkner, D.R., Utley, J.E.P., Bond, C.E., Alcalde, J., Heinemann, N., Haszeldine, R.S., Mackay, E., Ghanbari, S., 2020. Lower Cretaceous Rodby and Palaeocene Lista Shales: Characterisation and Comparison of Top-Seal Mudstones at Two Planned CCS Sites, Offshore UK. Minerals 10, 691. https://doi.org/10.3390/min10080691

Yang, Y., Yang, X., 2020. Data from W-2 and W-4 well in the Lower Congo Basin 1. https://doi.org/10.17632/dvzbhsn7z5.1

Yang, Y., Yang, X., Shi, W., Zhu, H., Wang, W., Kang, H., Pang, L., 2023. Recognition and Prediction of Source Rocks of the Madingo Formation in the Lower Congo Basin. J. Earth Sci. 34, 232–241. https://doi.org/10.1007/s12583-020-1354-2

Zee Ma, Y. (2016). Chapter 1—Unconventional Resources from Exploration to Production. In Y. Z. Ma and S. A. Holditch (Eds.), Unconventional Oil and Gas Resources Handbook (pp. 3–52). Gulf Professional Publishing. https://doi.org/10.1016/B978-0-12-802238-2.00001-8

Zeng, Z., Zhu, H., Yang, X., Ji, S., Zhang, Z., Huang, X., 2024. Depositional characteristics of fine-grained sedimentary rocks and the links to OAE-3 and PETM

of the Upper Cretaceous-Paleogene Madingo Formation, lower Congo Basin, West Africa. Marine Geology 467, 107206. https://doi.org/10.1016/j.margeo.2023.107206

Zhao, P., Mao, Z., Huang, Z., Zhang, C., 2016. A new method for estimating total organic carbon content from well logs. AAPG Bulletin 100, 1311–1327. https://doi.org/10.1306/02221615104

NOMENCLATURE

Constants, scalars and functions		
Symbol	**Description**	**Unit**
a	Tortuosity coefficient	-
B	Series of expansion coefficient	-
b	Perturbation term	Variable dependent
D	Relative data distance	%
E	Energy or cost function	-
F_1	First factor or shale related factor	-
F_1'	Scaled first factor	-
F_2	Second factor or organic factor	-
F_2'	Scaled second factor	-
GR	Natural gamma ray intensity	GAPI
GR'	Scaled natural gamma ray log	-
GR_{left}	Left scale limit of the GR curve	GAPI
GR_{right}	Right scale limit of the GR curve	GAPI
$\mathbf{g}$	Logging tool response functions	Variable dependent
I_{cl}	Clay indicator index	v/v
I_{cl}'	Scaled clay indicator index	v/v
$I_{cl_{left}}$	Left scale limit of the clay indicator curve	v/v
$I_{cl_{right}}$	Right scale limit of the clay indicator curve	v/v
K	Potassium concentration	%
K	Constant used for parameter ε	-
K_{RF}	Kerogen resistivity factor	Ohm·m
K_α	Kerogen conversion factor	-
k	Changing rate of the temperature parameter	
k_f	Number of factors to be extracted	-
K_s	condition number of the product of the Jacobian matrix (**G**) and its transpose ($\mathbf{G}^T$).	-
LOM	Level of organic metamorphism or maturity	-

M	Total number of petrophysical parameters in the inversion procedure	-
m	Cementation exponent	-
MKs	mean of the condition number of the $\mathbf{G^TG}$ matrix	-
$m(z)$	Depth dependent petrophysical parameter	Variable dependent
N	Number of data measurements for variable being analyzed in factor analysis	-
n	Saturation exponent	-
P	Acceptance probability function	-
p	Number of variables to being analyzed in factor analysis	-
P_e	Photoelectric absorption index	barn/electron (b/e)
PM	Total number of measurement points in the processed interval	-
Q	Number of expansion coefficients characterizing the model parameter	-
R	Resistivity	Ohm·m
$R_{baseline}$	Resistivity value corresponding to deep resistivity log curve when baselined in fine grained non-source, clay-rich rocks.	Ohm·m
R_d	Deep resistivity	Ohm·m
S	Fluid saturation	% or fraction (v/v)
S_c	Mean spread of the correlation matrix	-
ST	Number of logging instruments	-
T	Temperature	-
T_0	Initial temperature in SA algorithm	-
Th	Thorium concentration	ppm (parts per million)
TOC	Total organic carbon content	%
U	Uranium concentration	ppm (parts per million)
V	Lithological or mineral volume	v/v

z_q	Lower depth coordinates of the q-th rock unit	m (meters)
z_{q-1}	Upper depth coordinates of the (q-1)-th rock unit	m (meters)
α	Slope value in $\mathbf{I_{cl}}$ method	-
$\hat{\alpha}$	Proportionality constant in $\mathbf{I_{cl}}$**-BS** and $\mathbf{I_{cl}}$**-BS-SA** methods	-
α_F	Site-specific slope coefficient for based factor analysis TOC	-
α_P	Slope value in Passey's method	-
α_{SA}	Random number selected with uniform probability distribution from [0, 1]	-
β	Intercept value in $\mathbf{I_{cl}}$ method	-
$\hat{\beta}$	Exponential scaling factor in $\mathbf{I_{cl}}$**-BS** and $\mathbf{I_{cl}}$**-BS-SA** methods	-
β_F	Site-specific intercept coefficient for derive based factor analysis TOC	-
β_P	Intercept value in Passey's method	-
$\hat{\gamma}$	Decay rate constant in $\mathbf{I_{cl}}$**-BS** and $\mathbf{I_{cl}}$**-BS-SA** methods	-
Δd	Delta d distance, separation between the scaled gamma ray and clay indicator logs	-
ΔE	Change in energy or cost function	-
Δt	Slowness (P wave acoustic interval time)	µs/ft
Δt	Sonic interval transit time	µs/ft
$\Delta t_{baseline}$	Slowness value corresponding to acoustic log curve when baselined in fine grained non-source, clay-rich rocks	µs/ft
ΔlogR	Delta log R distance	-
δ_{ij}	Kronecker delta symbol	-
ϵ	Decrement factor	-
ε	Regularization (damping) factor	-
$\hat{\eta}$	Exponent factor in $\mathbf{I_{cl}}$**-BS** and $\mathbf{I_{cl}}$**-BS-SA** methods	-
θ	Constant smaller than 1	-

λ_{k_f}	Sorted eigenvalues of the sample covariance matrix	-
ρ_b	Bulk density	g/cm³
σ	Standard deviation	Variable dependent
$\sigma_{max}(\mathbf{G}^T\mathbf{G})$	largest singular value of the **G'G** matrix	-
$\sigma_{min}(\mathbf{G}^T\mathbf{G})$	smallest singular value of the **G'G** matrix	-
ϕ	Total porosity	% or fraction (v/v)
ϕ_{Da}	Apparent density porosity	% or fraction (v/v)
ϕ_N	Neutron porosity	% or fraction (v/v)
ϕ_{Na}	Apparent neutron porosity	% or fraction (v/v)
Ψ	Basis function (Heaviside step function)	-

Vectors and matrices	
$\mathbf{A}$	Matrix of factor loadings
$\mathbf{B}$	Vector of expansion coefficients
$\mathbf{b}$	Perturbation vector
$\mathbf{CORR}$	Pearson's correlation matrix
$\mathbf{COV}$	Covariance matrix
$\mathbf{d}$	Data vector
$\mathbf{E}$	Matrix of residuals
$\mathbf{F}$	Matrix of factor scores
$\mathbf{G}$	Jacobian or sensitivity matrix
$\mathbf{I}$	Identity matrix
$\mathbf{m}$	Model parameters' vector, observed in factor analysis
$\mathbf{m}_0$	Initial model parameter vector
$\mathbf{\Delta m}_0$	Small perturbation term of the model parameter vector
$\mathbf{U}$	arbitrary orthogonal matrix
$\mathbf{X}$	Data matrix
$\mathbf{S}$	Sample covariance matrix
$\mathbf{\Gamma}$	Diagonal matrix of the first k_f number of sorted eigenvalues
$\mathbf{\Sigma}$	Population covariance matrix for the observed variables
$\mathbf{\Psi}$	Residual covariance matrix
$\mathbf{\Omega}$	The matrix of the first k_f number of eigenvectors

Lower and upper scripts	
cl	Clay
h	Hydrocarbon
i	The *i*-th petrophysical property
k	Kerogen

ma	matrix
pt	Total number of petrophysical parameters
s	The s-th logging tool
T	Transposed
w	Water
c	Calculated value
m	Measured or observed value
pj	j-th data point in the p-th depth