UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

VINICIUS MEDEIROS GRACIOLLI

Bedding Contact Detection: A moving mean-based approach

Work presented in partial fulfillment
of the requirements for the degree of
Bachelor in Computer Science

Advisor: Prof. Dr. Mara Abel
Coadvisor: Prof. Dr. Luis Fernando De Ros

Porto Alegre
December 2014

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Reitor: Prof. Carlos Alexandre Netto
Vice-Reitor: Prof. Rui Vicente Oppermann
Pró-Reitor de Graduação: Prof. Sérgio Roberto Kieling Franco
Diretor do Instituto de Informática: Prof. Luis da Cunha Lamb
Coordenador do Curso de Ciência de Computação: Prof. Raul Fernando Weber
Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

“Eppur si muove.”

— GALILEO GALILEI

THANKS

Joel Carbonera
Sandro Fiorini
Ana Julia Magnus de Assis
Luan Garcia
Dean Pereira de Melo
Andrius Jonuska
Nenad Jaksic
Petrobrás and ANP for providing testing data

CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

FMS	Formation MicroScanner, a resistivity based image logging tool
UBI	Ultrasonic Borehole Imager, a ultrasound based image logging tool
ANN	Artificial Neural Network
BNN	Bayesian Neural Network
LAS	Log ASCII Standard

LIST OF SYMBOLS

w_l	Size of the long moving mean window.
w_s	Size of the short moving mean window.
aw	Agreement window
at	Agreement threshold
nt	Noise threshold
l	Number of logs
d	Total length of logs in increments
dc	Total length of cropped logs in increments
c	Total number of facies on the core sample
ss	Vector containing a log with a moving mean of size w_s applied
sl	Vector containing a log with a moving mean of size w_l applied
BM	Break matrix of size $l \times dc$

ABSTRACT

This work discusses the importance of core-log integration and how the problem of core sample depth matching is such an integral part of this process. In order to align the depth of cores and logs, it is required to identify some datum, or lithological contacts, that are strongly marked both in cores and and geophysical logs. We revise here the current techniques used to perform depth matching and proposes a new method that can derive information about position of contacts that can be used to perform core depth matching based on wireline log analysis. This method is based on moving means and works by detecting bedding planes across multiple logs and by creating a unified assessment of lithology contacts, which can then be compared to core sample descriptions. While the method shows some promise, the datasets used for testing were were ill-suited for wireline log analysis and testing with data from logs with more well defined and contrasting rocks is required for a more definitive proof of viability of this method as a tool for core-log integration.

Keywords: Wireline log. core sample. depth matching. Strataledge. lithology prediction.

Detecção de planos de acabadamento: Uma abordagem com médias móveis

RESUMO

Este trabalho apresenta a importância da integração entre dados de perfil e de testemunho, e como o ajuste de profundidade do testemunho é uma parte integral deste processo. Para alinhar a profundidade de perfis e testemunhos, é necessário identificar algum datum, ou contato litológico, que é fortemente marcado em ambos os testemunhos e os perfis geofísicos. São discutidas as técnicas atuais utilizadas para ajuste de profundidade de testemunho e proposto um novo método que deriva informação sobre a posição de contatos de perfis geofísicos que pode ser utilizada para realizar o ajuste de profundidade de testemunho. Esse método é baseado em médias móveis, e trabalha detectando limites entre camadas de rochas em múltiplos perfis; então unificando esta informação em uma avaliação total dos limites entre todos os perfis, que pode então ser comparada com as camadas de rocha descritas no testemunho. O método mostra-se promissor, porém, os dados disponíveis para teste são pouco adequados à análise de perfis, portanto, mais testes com dados mais adequados são necessários para se avaliar a viabilidade do método e sua utilidade para a tarefa de ajuste de profundidade.

Palavras-chave: perfil geofísico. testemunho. ajuste de profundidade. Strataledge. predição litológica.

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

In the field of petroleum exploration and development, exploration wells provide important information about the potential production of reservoirs, since they allow direct access to the rock properties. These properties define the quality of a particular reservoir, as well as provide us with geological information regarding the surrounding areas, which can be useful for geological and stratigraphic interpretation of the surrounding region.

As a common process, exploration logs provide several objects of study. The first are the geophysical logs, which are physical measures taken along of the whole deep of the well, such as emission of gamma rays, variation of electrical resistivity of the rocks, acoustic transmission and many others. The second is the drill cuttings that are taken out of the well along with the perforation water. The small cuts of rock provide a general view of the lithology that is being cutting by the perforation. The third object of study are core samples, that due to the price of recovery, are taken only from intervals of interest along the well, usually the reservoir rock and the rock seal that traps the oil in place (HYNE, 2012).

A core sample is defined as a cylindrical section taken out of some material for further study. Coring is a technique applied in many fields, in both naturally occurring and man-made materials. Core samples from a metal alloy or ceramic can allow the properties of the material to be tested, and core samples from bones can be used in medicine in diagnostics. In the context of this paper, core sample will refer to a cylindrical section of rock taken during the drilling of a borehole. Coring a borehole is an expensive and time-consuming process that requires a special drill, so only a few small sections are cored along the whole length of the borehole. Another problem with core samples is that it is not a foolproof procedure; rocks can be brittle, and more often than not, there are fractures in the core sample and loss of material during the recovery process, leading to possible sections of missing material (HYNE, 2012). Furthermore, since the only depth information available about the core is the depth interval that was drilled for it, there is no way to know right away what depth each individual core fragment comes from because of the slack space introduced into the drilled depth by loss of material. This possible depth variance is called the depth uncertainty of a core fragment and is illustrated in figure 1.1 (MALINVERNO, 2008). The data obtained from core samples used in this work are descriptions of the cores made by a geologist using the Strataledge[1] software.

Geophysical data can be presented in a wide range of resolutions, from surface measurements such, as seismic logs, which can give an overall view of the subsurface of an area hundreds of square meters, to analysis of rock cuttings, which can characterize the rock formation down to the grain size. While all these individual measurements can provide a good idea of the subsurface makeup of the subject area, integrating all these measurements can show us a more complete picture of the reservoir (WORTHINGTON et al., 1991). When integrating this data, however, we run into the problem caused by the difference between the measurement scales.

[1]Strataledge is a registered trademark of ENDEEPER.

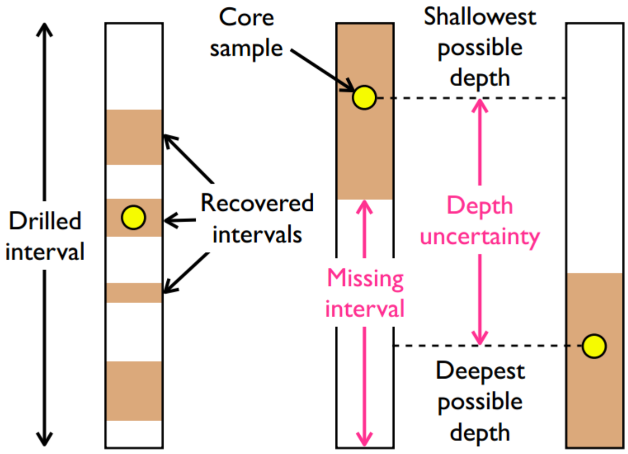

Figure 1.1: Illustration exemplifying the depth uncertainty of core fragments within a drilled interval, although the total length of unrecovered material is known, the size of each individual gap in the core section is not. Most commonly, losses occur on the top and bottom of the drilled interval, although losses between recovered core fragments can also occur. source: (MALINVERNO, 2008)

Figure 1.2: Examples of core samples, source: (DURANTI et al., 2002)

The information that may seem a small offset between two sets of data from the point of view of the larger scale measurement may line up the smaller scale with a reading that makes no sense. These offsets are very common since measurements are taken at different times and using different tools (LOFTS; BRISTOW, 1998).

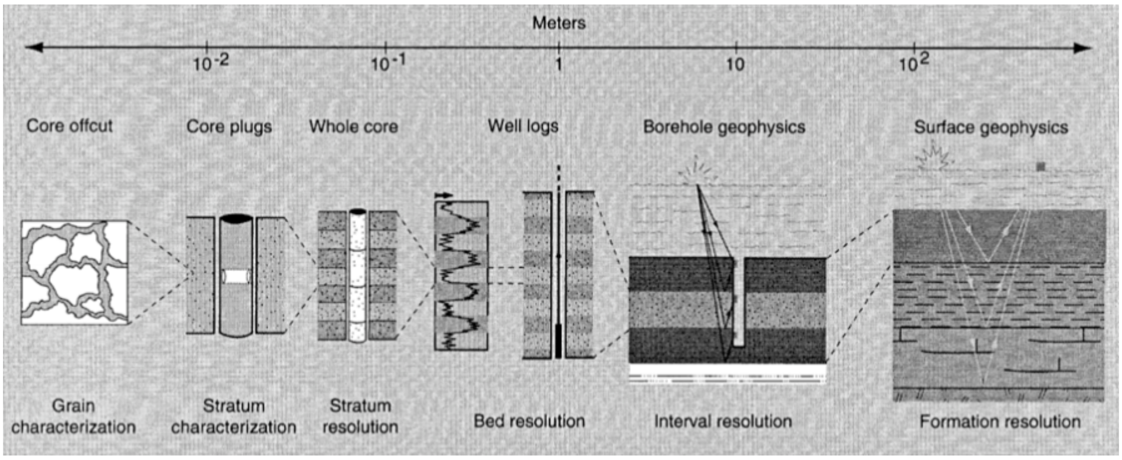

Figure 1.3: Diagram illustrating the range of measurement scales used in geophysics, source: (MALINVERNO, 2008)

Wireline logs are measurements taken along the length of the borehole, either during drilling by tools that are attached to the drilling head (LWD) or after it is drilled by tools lowered into the borehole (Wireline). For simplicity purposes, we will refer to measurements taken with any of these tools as wireline logs. These tools can be classified as passive, where they just take measurements from the borehole such as natural gamma ray emissions, or active, where they excite the rock in some way and then measure a response, like the neutron porosity tool, which emits neutrons and records the returning neutrons or gamma radiation. These log measurements may be further classified in uncased and cased measurements, with the former being taken from an open borehole, and the latter being taken after the borehole has been cased with piping. Wireline logs are presented as a table of depths and the values measured by each of the logging tools used at each depth increment (KRYGOWSKI, 2003).

Depth data from logging runs is still prone to error, the tool may get stuck while it is being pulled up from the well and record data from the same depth multiple times, after getting unstuck it can also come up too fast due to cable tension and miss certain depths. Additional problems arise on underwater wells, where the waves can cause the logging ship to heave and alter the depth of the logging tool. Furthermore, if multiple logging runs are done on the same well, some processing must be done in order to line up the depths recorded by the different runs, usually one of the tools is used across all the runs and used as reference, most frequently, the gamma ray tool. These problems are best and usually handled by the logging company itself during data acquisition, so, the data used in this work will be considered correct (LOFTS; BRISTOW, 1998).

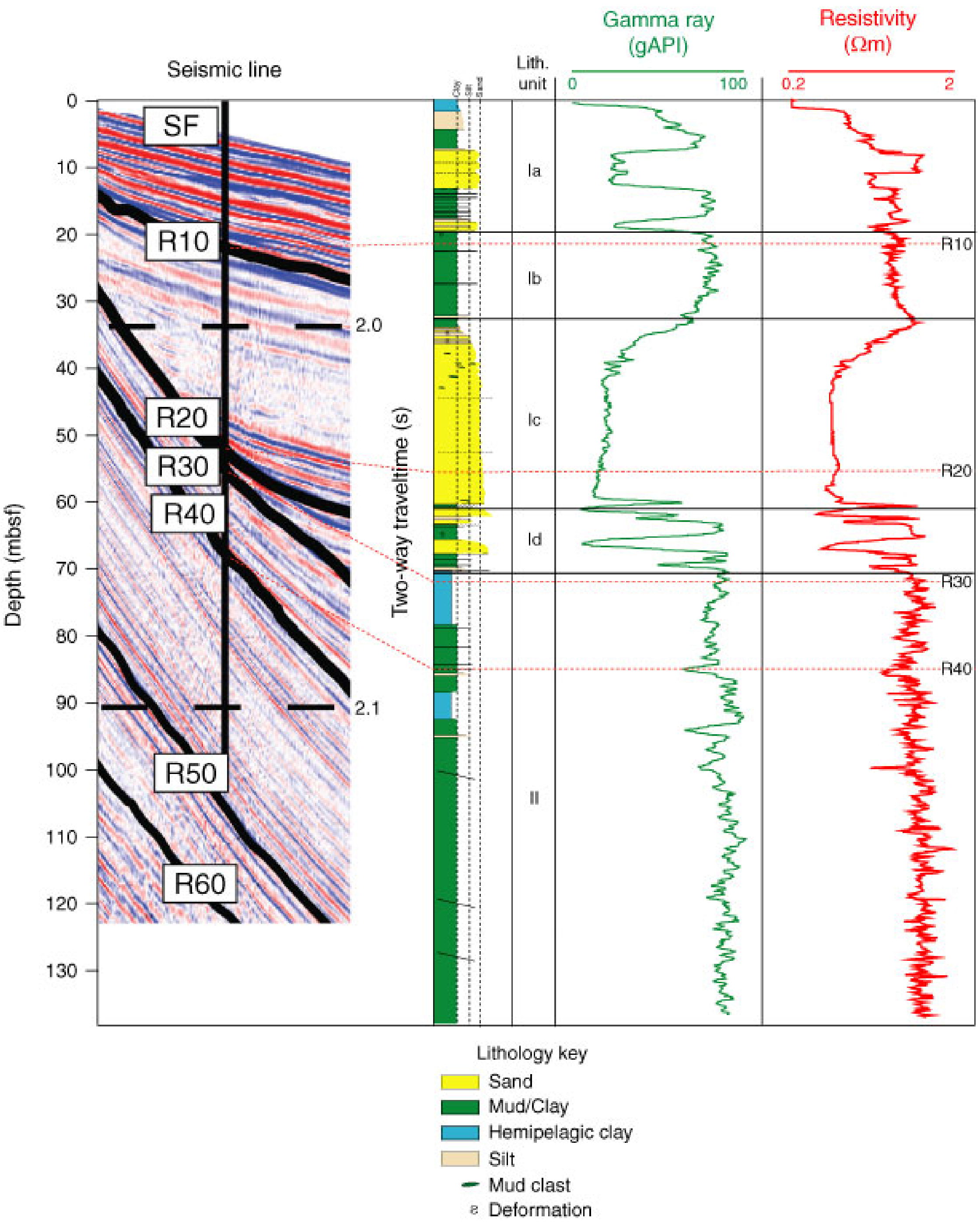

Figure 1.4: Example of gamma ray and resistivity wireline logs plotted next to lithological interpretation of well data, note how the changes in lithology correspond to changes in the log levels. Source: (FLEMINGS; BEHRMANN; JOHN, 2006)

In this work, I will focus on the depth matching step of integration of core sample descriptions and wireline-log data. We can define the problem of core depth-matching as assigning to each core sample fragment it's real depth within the drilled depth, using the wireline logs as reference to make that assessment. Although there are plenty of efforts done by software provider companies, to this day, this is still a task performed manually by a specialist, relying on more information than what is being used in this work, such as geophysical measurements of the core sample, or resistivity images from the borehole wall.

The aim of this work is not to replace the expert guided work, but to offer an alternative method of adjustment to support non-specialized geologists in adjusting the core depth and correlating wells by providing qualified rock data and semi-automatic methods of depth adjustment.

2 ROCK-LOG CORRELATION METHODS

This chapter revises the main methods described in literature for core depth adjustment and rock-log correlation in exploration wells. The description of the approach used in each method is followed by an evaluation of the effectiveness of the proposal.

2.1 Data Formats Used in This Work

This section describes the data formats used in this work to retrieve data regarding well logs and core sample descriptions. Their usage will be described in the next chapter.

2.1.1 The LAS File

The Log ASCII Standard (LAS) file is the current industry standard for storing and transmitting borehole log information. A LAS file usually consists of the following (CRANGLE, 2007):

1. A header detailing the LAS version the file is formatted as.
2. The well information section, dealing in general well information such as geographical coordinates, depth drilled and companies responsible for drilling and logging.
3. The curve information section, which lists which measurements are logged in the file.
4. The ASCII log data section, which lists the values measured at each depth increment for every curve in the log, in the same order as described in the curve information section.

2.1.2 The Strataledge Software

Core sample descriptions are the result of visual inspection by a geologist, who will draw and write down what attributes he sees on the sample. This can create a lot of ambiguity, since different geologists and companies can have their own way of drawing structures and a different lexicon to describe them.

The Strataledge software (PERRIN et al., 2012) provides a standardized ontology based environment where the geologist can describe the core sample using standardized names and representations based on a well-founded ontology. This description can then be exported as an SVG file for human inspection, or as an XML file for processing.

2.2 Core Depth Adjustment

Since core depth matching is a problem under study for decades, many techniques have been proposed over the years to tackle this issue. Most of these methods disregard the core sample descriptions used in this work and instead opt to extract geophysical data from the core samples using objective measurement tools. These kind of approaches give you the advantage of comparing two sets of numerical data, instead of one set of numerical data and a qualitative description that is open to interpretation.

The current standard procedure in the petroleum industry consists in taking the gamma ray measurements from the core sample using a gamma ray logger and comparing it with the gamma ray measurements obtained from the wireline tools. It is up to the geologist to find the depths where the two curves have the best match. (MORTON-THOMPSON; WOODS; GEOLOGISTS, 1993)

Another method, patented by Vinegar et al. (VINEGAR; WELLINGTON, 1985) revolves around the use of a computerized axial tomographic scanner (CAT) to determine the attenuation coefficients at a plurality of cross sections along the core sample. Since the attenuation coefficients are directly proportional to the density values of the core, these values can be interpolated and convolved with the response function of a density logging tool to generate convolved density values that can be compared to bulk density data obtained from wireline tools. Another embodiment of the same method involves convolving effective atomic numbers with the response function of the tool to obtain convolved effective atomic numbers that are cross correlated with the photoelectric log values. In either case, values on both sets of data are compared by a specialist who determines the best depth shift on the core.

A patent made in 2005 by Siddiqui (SIDDIQUI, 2005) builds upon Vinegar's work. It aims to simplify the calculations involved by using standardized CT number data and statistical analysis to calculate the core's bulk density, instead of using the tool response function.

Techniques that do not rely on core geophysical data have also been developed over the years, Fontana et al. (FONTANA; ITURRINO; TARTAROTTI, 2010) describes a way to derive a core image from a mathematical representation of the core sample:

> *We assume that the geological structures described in the structural logs are planar features and the borehole wall is a cylinder. Considering the unrolled surface of the borehole wall, any intersection of a structure and the borehole wall can then be represented by:*
>
> 1. *a sinusoid (structure neither normal nor parallel to the borehole axis),*
> 2. *a horizontal straight line (horizontal structure, normal to the borehole axis),*
> 3. *or a vertical straight line (vertical structure, parallel to the borehole axis)*

The derived core image is then compared to an image created by a logging tool such as FMS or UBI. This method has the advantage of not only providing depth correction, but also core orientation. Once again, the matching is made manually by a specialist. Usage of a possible automated correlation process is discussed briefly in the article, and the authors believe it would be less reliable than manual matching.

While the methods discussed so far provide good results, they all rely in specialized tools or data that are outside of the scope of the resources available for this work. During this research,

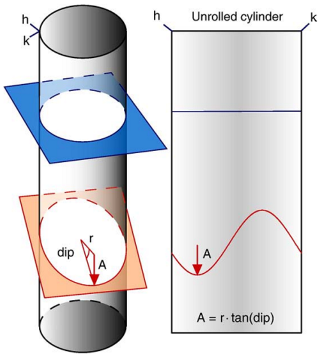

Figure 2.1: Illustration exemplifying the construction of the structural log discussed in (FONTANA; ITURRINO; TARTAROTTI, 2010). The geological features present in the core sample can be represented as planes intersecting the core, when the core wall is unrolled and represented as a plane, the intersections between the core sample and the planes can be seen as sinusoids or straight lines.

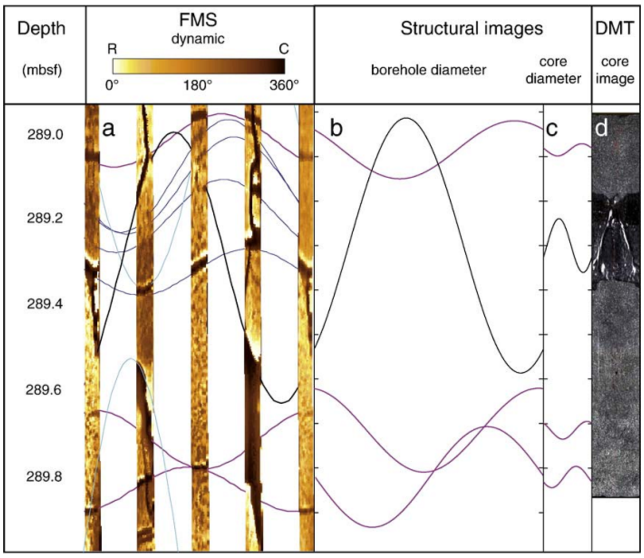

Figure 2.2: Comparison between the structural logs derived in Fontana's article and FMS images, note how the structures observed on the image log can be correlated to the curves in the structural log. Source (FONTANA; ITURRINO; TARTAROTTI, 2010)

we found no literature on solving depth matching using only wireline logs and core descriptions, so a new method has to be proposed.

2.3 Deriving Rock Information From Wireline Logs

The main obstacle to overcome when integrating wireline data and core descriptions rests in bridging the gap between a numeric value and a qualitative description. To do that, we need to process one of the data sets and make it comparable to the other. Taking a qualitative description and simulating a wireline log based on the description is not a reliable process, since not only readings for the same type of rock can vary wildly between different wells, but also some readings may have no direct correlation to the lithology observed (MANN; LEYTHAEUSER; MÜLLER, 1986). Therefore, it is a more interesting approach to process the wireline logs and look for markers that can be correlated to the core description.

Neural networks have been used in this field to predict lithology with varying degrees of success. An artificial neural network (ANN) is a computational model inspired by an animal's nervous system. In its most basic form, an ANN is presented as a system of interconnected nodes (neurons). The connections between neurons have assigned weights that are tuned by a learning algorithm in such a way that the final output of the neural network matches the expected result. In the context of this work, an example would be a neural network whose input neurons receive the measured wireline log values at a certain depth, apply an evaluation function to these values and propagate the results to the adjacent neurons. These neuron, in turn, do the same process until an output neuron is activated, whose final value is the estimated lithology at that depth.

The learning algorithms used in neural networks can be supervised or unsupervised. On supervised learning, the algorithm is fed a set of sample pairs consisting of previously determined inputs and their associated outputs. The aim of the learning algorithm is then to estimate which function that correctly maps the input value to the output value. On unsupervised learning, the algorithm receives input data and a cost function to minimize.

Ojha (OJHA; MAITI, 2013) shows an approach using a Bayesian Neural Network (BNN) that differs from a regular ANN by virtue of the heuristic used to choose the starting values for the network weights. While an ANN picks a single set w_0 of starting weighs randomly, a BNN considers the entire space w of possible weight assignments and performs a weighted average across the whole domain using supervised learning. The data sets used for supervised learning on Ojha's work are derived from clustering and statistical analysis methods applied to wireline log data. The article claims an average accuracy of 67.38% in the presence of 10% red noise. The problem in using this approach as a way of comparing wireline and core data is that the training process must be redone for every individual well. Also, that training requires some prior knowledge about the well, such as how many different lithologies are represented on the logs so that the clustering methods may be applied.

Brereton (BRERETON; GALLOIS; WHITTAKER, 2001) employs a purely clustering based approach to determine lithologies based on wireline log data. This method consists of using the log readings at a given depth as coordinates on a color space, and then classifying the lithology at that depth based on what area of the color the coordinates fall in. One of the biggest problems with this approach is that while it detects the most obvious changes in lithology reasonably well, it also detects numerous changes not described in the lithology data used for validation. The article claims that it can detect subtle variations not readily detected by the specialist, however, as our work is trying to reconcile the specialist derived data and the wireline log readings, we need a method that finds similarities between both instead of differences such as seen on figure 2.3.

Gifford (GIFFORD; AGAH, 2010) implements a multi-agent based method to classify facies from wireline logs. In a multi-agent system, multiple methods try to solve a problem independently, each one of these independent methods is what is known as an agent, their results are then combined into the final output of the multi-agent system. In Gifford's implementation, the agents used are a series of learning algorithms, such as different implementations of neural networks. While this approach manages up to 84.3% accuracy, it has the same pitfalls associated with Ojha's work, namely the need for training with pre-processed training data.

Reid (REID; LINSEY; FROSTICK, 1989) describes what he has called in his work the *Automatic Bedding Discriminator*, a method to detect lithology boundaries based on gamma ray logs. This approach consists of using moving means to detect sudden changes in the log data, which characterize a change in lithology. Bedding planes can be used as strong markers that can be easily distinguished in core samples. Since this method offers relatively good accuracy in determining strong markers using minimal input data, it was chosen as the basis for this work and it will be explained in detail in the next chapter.

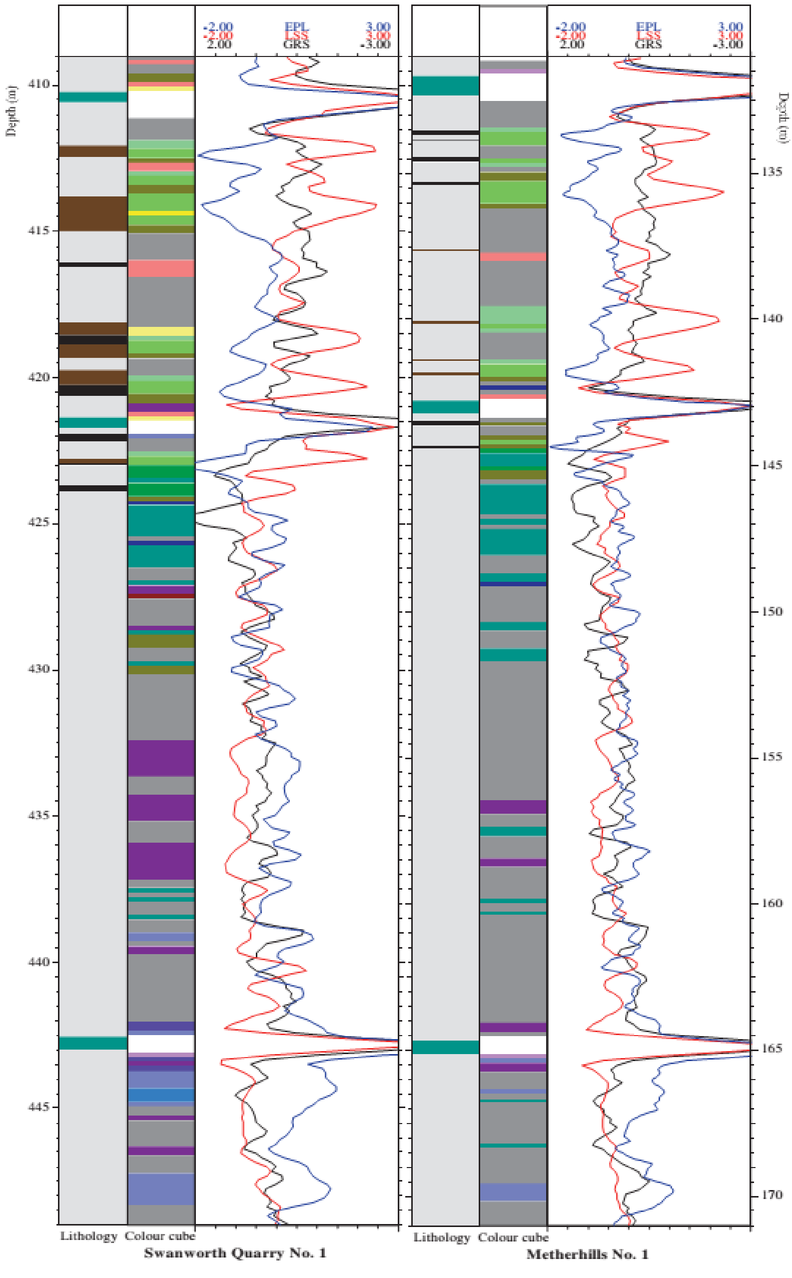

Figure 2.3: Result from Bereton's method, note the high discrepancy between the derived lithology and the actual core description. Source (BRERETON; GALLOIS; WHITTAKER, 2001)

3 AUTOMATIC BEDDING DISCRIMINATOR

In this chapter, we will describe the implementation of the *Automatic Bedding Discriminator* proposed by Reid (REID; LINSEY; FROSTICK, 1989), the improvements upon it that were proposed in our work, and how it can be used as a tool in core sample depth matching.

3.1 Method

In simple terms, a wireline log consists of a series of measurements sampled along the length of the borehole, these measurements can then be displayed as a curve. A change in lithotype can be detected by a sudden increase or decrease in the values measured by the logging tool. Certain types of logs are more indicative of a change in lithology than others. Chief among them is the gamma ray log, which is also the only type of log used in the original iteration of the Automatic Bedding Discriminator implemented by Reid. However, logs that deal with resistivity, density and porosity can also serve as good indications of a change in lithology (OJHA; MAITI, 2013).

The first step necessary in order to detect bedding contacts is dealing with the signal noise, since log measurements can be greatly affected by well conditions; such as variations in borehole width and the type of drilling mud used. In order to reduce the effects of noise on the log values, a simple centered moving mean is applied to each of the logs. Aside from reducing perceived noise, the smoothing resulting from the moving mean also has the benefit of filtering out log variations resulting from small-scale changes in lithology. The size of the moving mean window determines how much log will be smoothed by this process. A small window will combine each of the sampled values with fewer of its neighbors than a larger window, preserving the shape of the curve better, but will not be as effective at filtering out noise. On the other hand, if the size of the window is too large, you can filter out important features from the log as if they were caused by noise. In Reid's research (REID; LINSEY; FROSTICK, 1989) it is demonstrated that information is lost quickly once the smoothing interval is increased beyond 2m, thus, a smoothing interval 1.05m is used, which translates to a moving mean window of 7 samples in a log with 15cm steps between each measurement.

Once we have the smoothed log, the next step is identifying the where the boundaries between facies are represented on the log. Since lithological changes are characterized by a sudden increase or decrease in wireline log readings, the change points lie between local minimum and maximum points. However, we are not dealing with an idealized log. Small fluctuations in the log that happen between readings from the same bed create points of minima and maxima that do not correspond to a change in lithology. Therefore, a methodology has to be used which detect bedding contacts between the most significant peaks and valleys in the log curve while ignoring meaningless signal variation.

The approach used to find these change points consists of applying a second moving mean to the original log with a much greater window than the one used to smooth the log, with the goal of extracting a curve that shows the general trend of the log. We can then determine the locations of the bedding contacts by identifying at which depths this curve intersects the smoothed log. As with the log smoothing operation, the size of the long smoothing window is also an important parameter in determining the sensitivity of the algorithm's detection: windows that are too short will tend to recognize noise as legitimate bedding contacts, and windows that are too long will ignore important signal variations. As extreme examples, we have a long window that is the same length as the short window, both curves will be the same, and thus, intersect at every point. We can also have a long window that covers the entire length of the log, which will create a flat line that will intersect with the smoothed log only when it crosses the average value of the whole log. Reid's research suggests that an interval of 10m be used for the long window. In this work, we will call the size of the short window w_s, and the size of the long window w_l.

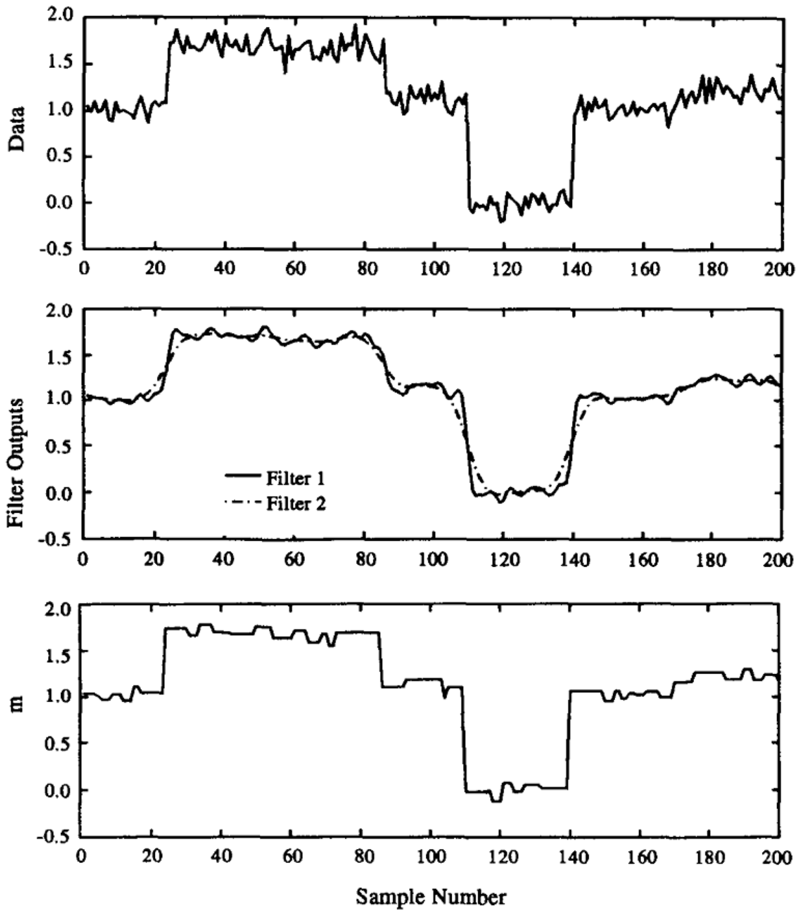

Figure 3.1: Example of moving means being used to detect change points in an arbitrary dataset. Filter 1 is obtained by applying a small moving window to the original data, Filter 2 by applying a larger one. The final result displayed is obtained by averaging all values between each intersection. Source (MACDOUGALL; NANDI, 1997)

Since this method works based on relative differences between log values, special consideration must also be made in cases where the log in question is monotonous, showing little difference between absolute log values along a considerable depth, most likely this means a large depth consisting of the same lithology. In these cases, we can establish a threshold and tell the algorithm to ignore intersections resulting from changes in value that don't exceed this threshold. Reid's original implementation (REID; LINSEY; FROSTICK, 1989) focuses exclusively in the application of this method to gamma ray logs, and uses a threshold of 4 API to demarcate bedding planes. We will call this threshold the noise threshold, or nt.

In this work, we decided to apply Reid's method not only to gamma ray logs, but to create a generic application that can take any number of logs from the same well as input, apply the Automatic Bedding Discriminator described in Reid's paper to each one of them and then corroborate bedding contact evidence across multiple logs for a more accurate assessment.

It is important to keep in mind that wireline logs have different degrees of representativeness when it comes to lithology. Logs measuring gamma ray emissions can tell the amount of organic content in the rock and differentiate between shale and sandstone. Porosity is also an important lithological factor, so logs that are used to infer density, such as sonic and neutron logs are also strongly tied to lithological characteristics. Resistivity logs are also used as an indication of lithology, but often to a lesser degree compared to the logs previously discussed (KRYGOWSKI, 2003).

LAS files also come with logs that have little to no lithological significance, the caliper log for instance, that measures the width of the borehole can provide important information to the engineering crew tasked with developing the well. While it is true that a friable rock crack more easily during drilling, leading to a larger borehole size, the caliper log is still not a reliable lithological marker. We also have logs related to the logging process on a LAS file, such as tool tension and transit time, which are used by the logging company to correct and compensate measurements from other logs.

Due to these issues, it is important to assign weights to the logs being analyzed. For example: a bedding contact reported by a gamma ray log should be taken more seriously than one reported by a resistivity log in a shale and sandstone rich environment. These weights have to reflect the importance of each log when used to perform lithological assessment. After inquiries made with geologists regarding the lithological representativeness of each log, the following general guidelines for assigning log weights were established and validated for the specific environment represented by the data sets available for testing:

- Gamma Ray logs are the most representative, and should have the highest weight.
- Logs dealing in measures of density or porosity should have weights close or equal to the gamma ray logs.
- Logs dealing in measures of resistivity should have weights around half the value used for gamma ray logs.
- Other logs are not representative enough for lithological assessment, and thus, should have weights equal to zero.

Once the Automatic Bedding Discriminator has been applied to each of the logs and the weights have been assigned, we can integrate these separate results into a unified assessment of bedding contacts. This is done by going through each sampled depth and verifying how many logs accuse a break at that depth; if the sum of the value of the weights of those logs is equal or greater than a user-defined threshold, that depth is declared as a bedding contact. This threshold will be referred in this work as the agreement threshold, or at.

3.2 Implementation

In this section, I will detail the algorithm employed to implement the method described in the last section. First, we will take a look at the files and data structures used to store the data being analyzed.

3.2.1 Loading Data Structures

In this work, wireline log data is loaded into the program from a modified LAS file: the first two sections are stripped out and log weights are added to the curve information section, next to the name of each of the corresponding curves. The data for all l logs across d depth increments is loaded and stored in a matrix of $l \times d$ elements. Weights into a vector of size l.

The core data is loaded from XML files generated by Strataledge. Strataledge can represent a wide variety of core sample characteristics, but since the goal is comparing the core to bedding contacts inferred from wireline log data, the only information loaded from the Strataledge files are the depth of the top of the core sample and the length of each facies described by the geologist in order of depth. The result is stored in a vector of size c, where c is the number of facies described in the XML file.

3.2.2 Cropping Logs

Since we are interested in comparing bedding contacts derived from wireline log data to bedding contacts described in the core sample, it makes little sense to apply the algorithm to the entire length of the borehole since wireline log data can span kilometers, where typical core sample intervals are around 20 meters.

Thus, the first step done is to crop the log to a more manageable size, the new starting depth of the logs should be at least $w_l/2$ samples before the start of the core depth, and the new final depth at least $w_l/2$. This ensures the moving means will be operating at their full length during the cored section of the borehole.

Having eliminated superfluous depths, we can now discard logs that are deemed useless to the bedding contact assessment. This is done by declaring certain logs "dead" and registering this in an array of size l containing flags which say if the corresponding log is dead or not. A log is considered dead if it meets any of the following conditions:

1. Its assigned weight is zero.
2. More than 20% of its values are null.

A dead log is ignored by the rest of the algorithm, but can still be displayed in the interface at the discretion of the user. The length of the cropped logs will be referred to as dc, note that the logs declared dead are not discarded, but simply considered inactive, so, the first dimension of the matrix storing the cropped log remains the same, resulting in a matrix of $l \times dc$.

During this operation, we also calculate the noise thresholds of each log by finding the highest and lowest non-null values on the log and multiplying the difference between those measurements by a value between 0 and 1. This value can be changed in the program interface.

Naturally, if the user desires to use the application to detect bedding contacts across the whole length of the borehole, the depth cropping step of this section can be skipped, and the complete log be used in place of its cropped version in the rest of the application. In this case, the rule regarding declaring logs dead if it contains too many null values should be rescinded, since every log will be relevant at some point in the well.

3.2.3 Moving Means

The next step is applying the moving means to the log, a centered moving mean with a window of size w at the point d_n is the average of w closest elements to d_n including d_n and can be calculated as:

$$\frac{d_{(n-\lceil (w-1)/2 \rceil)} + d_{(n-\lceil (w-1)/2 \rceil+1)} + d_{(n-\lceil (w-1)/2 \rceil+2)} + ... + d_{(n+\lfloor w/2 \rfloor)}}{w}$$

For each log, the moving mean is applied with a size of $w_(s)$ and the result is saved in a vector of size dc, then, a second moving mean with a window size of $w_(l)$ is applied to the cropped log and the result is saved in a second vector of size dc. These smoothed vectors, resulting from the application of a short and a long moving window from here on will be referred as ss and sl respectively, and are fed into the procedure described in the next section.

3.2.4 Finding Break Points

Once in possession of both smoothed curves, the next step is comparing them both and finding the intersection points. It is important to notice that while wireline log data is frequently referred as a curve, it is not actually a curve that is mathematically defined as an equation; it is a series of sampled measurements, and thus, not continuous. Therefore, the bedding contacts will not be marked on the exact depth it manifests itself on the log as if it was a continuous curve, but on the closest sampled depth.

This procedure is done first by ascertaining whether the first value of ss is lower than the first value of sl, if it is, a flag is turned on signaling ss is currently "under" sl, else the flag is turned off. Additionally, if the first value of ss is equal to the first value of sl, a bedding contact is marked on the first position.

After this setup procedure, we start traversing the vectors. At each depth increment after the first, the algorithm looks for a break by checking if the "under" flag matches the measurements at the current depth, this is also the stage where the noise tolerance loaded from the log file is checked. When comparing points ss_n and sl_n on a log with a noise tolerance of nt a bedding contact is marked:

1. When $ss_n \geq sl_n$ and $under = true$ and $|ss_n - ss_{n-1}| > nt$
2. When $ss_n \leq sl_n$ and $under = false$ and $|ss_n - ss_{n-1}| > nt$

When one of the above conditions are met, to ensure the bedding contact is marked on sampled depth closest to it, it is marked:

1. On position $n - 1$ when $|ss_{n-1} - sl_{n-1}| < |ss_n - sl_n|$
2. On position n otherwise

The return value of this procedure is a binary vector of size dc with a 0 value on depths with no bedding contacts, and 1 values on depths where there is a bedding plane. This will be referred to as the break vector. Once the break vectors for all non-dead cropped logs are calculated, they are stored in a vector of size l, resulting in a $l \times dc$ sized matrix, referred to as the break matrix.

3.2.5 Comparing Break Points

After the bedding contacts have been detected in each of the non-dead cropped logs, we have to compare these results in order to reach a unified assessment of bedding contact locations. This is done by going through every sampled depth in the break matrix and adding up the values of the weights of every log that accuses a bedding contact at that position. If the sum of these weights meets or exceeds the agreement factor, we consider that position as a legitimate bedding contact.

It is important to notice that there is no guarantee that the same break will manifest itself across multiple logs at the same exact depth. Signal noise can easily cause the computations done up to this point to nudge the demarcation of a bedding contact up or down, so we have to take this in account when comparing the break vectors.

Since the break vectors are iterated through sequentially, this is a matter of including in the bedding contact assessment the next depth increments after finding a break in one of the logs. This is done by establishing a vertical threshold that determines across how many increments a single bedding contact can manifest itself, we will call this threshold the agreement window, or aw. This threshold should be kept fairly small, usually not more than one or two additional depths, since the goal is just to account for noise, and we assume the wireline logs provided to the application are already depth-matched. Application of this technique in a size $l \times dc$ matrix storing breaks detected in l logs across dc depth samples is detailed in algorithm 1.

```
Input: A matrix BM of size l × dc. The position (n, m) contains a 1 if there was a break detected on the
       depth m of log n, and a 0 otherwise.
Input: A vector weight of size l, where the weight of log n is contained in position (n)
Input: A numeric agreement threshold value at
Input: A numeric agreement window value aw
Output: A vector of size dc. The position (n) contains a 1 if there was a break detected at that position and
        a 0 otherwise.
begin
    /* This will be used to tell which logs have breaks inside the
       agreement window                                                      */
    Initialize a vector inWindow of size l with zeroes;
    /* This will store the position in which the last break occured on
       each log                                                              */
    Initialize a vector lastBreak of size l with zeroes;
    /* This will store the unified change point assessment                   */
    Initialize a vector result of size l with zeroes;
    foreach m = 0 to m < dc do
        foreach n = 0 to n < l do
            if BM_(n,m) == 1 then
                inWindow_(n) = aw;
                lastBreak_(n) = m;
        breakPosition = 0;
        totalWeight = 0;
        foreach n = 0 to n < l do
            if inWindow_(n) > 0 then
                totalWeight = totalWeight + weight_(n);
                breakPosition = m * weight_(n);
                inWindow_(n)–;
        if totalWeight ≥ at then
            result_(breakPosition/totalWeight) = 1;
    return result;
```

Algorithm 1: Algorithm for integration of change point detection across multiple logs

An example illustrating the application of algorithm 1 is shown in figure 3.2. The agreement window used in the example is 2, and the agreement threshold is 5. A break is declared once the algorithm reaches depth 2, since at that depth the total weight is greater or equal to the agreement threshold (4 from breaks at that depth, 1 from a break in a previous depth, but still within the agreement window). The weighted average between the position of all 3 breaks involved results in 1.8, so a break is marked at position 2, the closest integer index.

Once this procedure is done, the result is a break vector similar to the ones returned by the procedure described in the last subsection, but derived from a unified assessment of bedding contacts across multiple logs. This is the final result of the method, and what should be compared to the core sample description.

Bedding contacts are usually unambiguous markers in a core sample, and comparing the results from the application of our method to core log descriptions should be a reasonably straightforward approach. The work left to the expert is shifting the core fragments up and down the borehole and finding the best match between the bedding contacts described in the

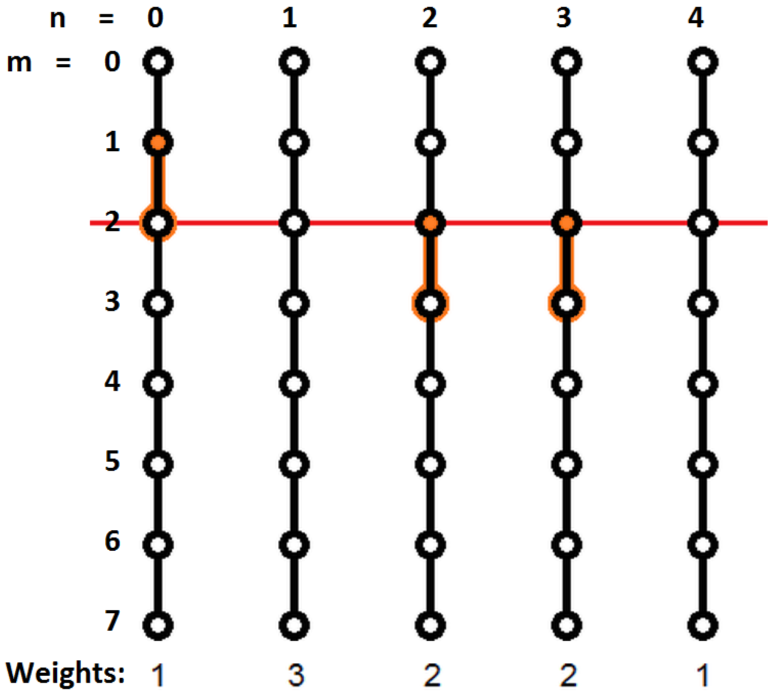

Figure 3.2: Example illustrating the previously defined algorithm for comparing change points. Orange points indicate a break, and an orange outline indicates the agreement window that extends from that break.

core and the bedding contacts located by our application.

4 RESULTS

In this chapter, I will present the results obtained by applying the bedding detection method to two sets of wireline log data procured from two wells in the Espirito Santo basin. The bedding contacts are plotted in yellow, and the core sample beddings are displayed to the right. The cores in these cases have already been depth matched by using core gamma readings as described in 2.2

Unfortunately, the data provided for testing was ill-suited to the task at hand. The carbonate lithology described in the core samples show a high degree of homogeneity and the similar lithotypes translate into monotonous logs, consequently, the noise threshold has to be set to a very low value for any bedding contacts to be detected. A more ideal case would be a core showing contrasting lithotypes with distinctive log readings, such as shale and sandstone.

A breakdown of the results obtained from the data provided for testing can be seen in table 4.3. This table show how many bedding contacts present on the core were detected in the core within one and two increments of accuracy. The number of bedding contacts detected on the logs that were not present on the core sample (false positives), the number of bedding contacts present on the core not detected in the logs (false negatives) and the number of different lithotypes present on the core sample. Bedding planes detected outside of the cored interval are not considered in this calculation. It is also important to know that the intercalation of thin facies on core ESS-0023 T1 is supposed to be a single conglomerate lithotype. It can be seen that when detection occurs, it is usually accurate, but due to the similarity between lithotypes in these wells, there is a large number of false negatives, since these similarities translate to minute changes on the log readings.

4.1 Well ESS-0023

Well ESS-0023 is an offshore well located in the Espirito Santo basin approximately 6 kilometers away from the shore. Wireline log data from this well is taken in steps of 20 centimeters and has been described by an inquired expert as of poor quality. A total of 26 curves are present on the LAS file, however, most of them are null for large lengths of data, indicating they are originated from multiple logging runs done at different depths. Cores T1 and T2 were redescribed in Strataledge based on previously made descriptions done by hand. These core descriptions can be seen in figures 4.1 and 4.4 respectively. Weights used for the logs in well ESS-0023 are seen in table 4.1. These weights were defined by following the guidelines described in section 3.1. The results from the application of the method on cores T1 and T2 are shown in figures 4.2 and 4.5, with parameters used in each case shown in figures 4.3 and 4.6.

These were the best results obtained on those cores after experimentation with different parameters. While a substantial number of bedding contacts described on the log were not detected, most of the bedding contacts that were detected by the algorithm are marked within less than one or two samples to bedding contacts on the core, especially on core T1. It is also of note that when presented with these results, an inquired expert could not detect any bedding contacts from visual inspection of the logs, while the application did detect bedding contacts that are confirmed in the core sample description.

4.2 Well LPN-002

Well LPN-002 is located inland, approximately 5 kilometers from the shore in the Espirito Santo basin. Wireline log data from this well is taken in steps of 15.24 centimeters and have been described by an expert as of higher quality than the data from well ESS-0023. Core samples however, reveal an extremely homogeneous carbonate lithology, which is reflected in the monotonous lines plotted by the wireline log data. Well LPN-002 has a much shallower depth than well ESS-0023, and data from this well is presented in 12 logs, most of which are active for most of the logged depth, indicating they were likely taken in a single logging run. Cores T1 and T3 can be seen in figures 4.7 and 4.10. Weights used for the logs in well LPN-002 are seen in table 4.2. These weights were defined by following the guidelines described in section 3.1. Results from applying parameters shown in figures 4.9 and 4.12 to cores T1 and T3 can be seen in figures 4.8 and 4.11.

Despite the higher quality logging data, results on this well were noticeably worse than on well ESS-0023, mostly due to the fact that the extremely similar lithotypes make it hard to distinguish bedding contacts. Parameters have to be adjusted in a way as to consider slight variations as a possible bedding contact, and that causes noise to register as bedding contacts.

4.3 Effects of Parameters in Contact Detection

As expected, changing the parameters can lead to wildly different results. For the carbonate lithology present in these wells, parameters that control noise filtering such as the noise threshold and the size of the short moving mean window had to be kept low in order to detect a reasonable number of bedding contacts. The noise threshold in particular had to be kept under 0.5% for any detection to be made, this translates to a number much lower than the 4 API noise threshold used in Reid's research (REID; LINSEY; FROSTICK, 1989).

Controlling the size of the long moving mean window also allows increasing or decreasing the detection of bedding contacts. A long window that is close in size to the short window makes it easier for the means to intersect, resulting in more detections.

The agreement factors are especially important in increasing detection accuracy in data sets with more monotonous logs. Decreasing the agreement factor needed to declare a break can help detection considerably by having logs with lower weights declare breaks without needing corroboration from other logs, while increasing it can reduce the number of false positives in sets with many bedding contacts detected.

The agreement window can create a huge number of false positives if used irresponsibly, values greater than two will start showing large sections of the log with multiple consecutive bedding contacts. Increasing it can be useful in cases where that is the behavior that is expected however, such as when the core sample has a series of thin facies.

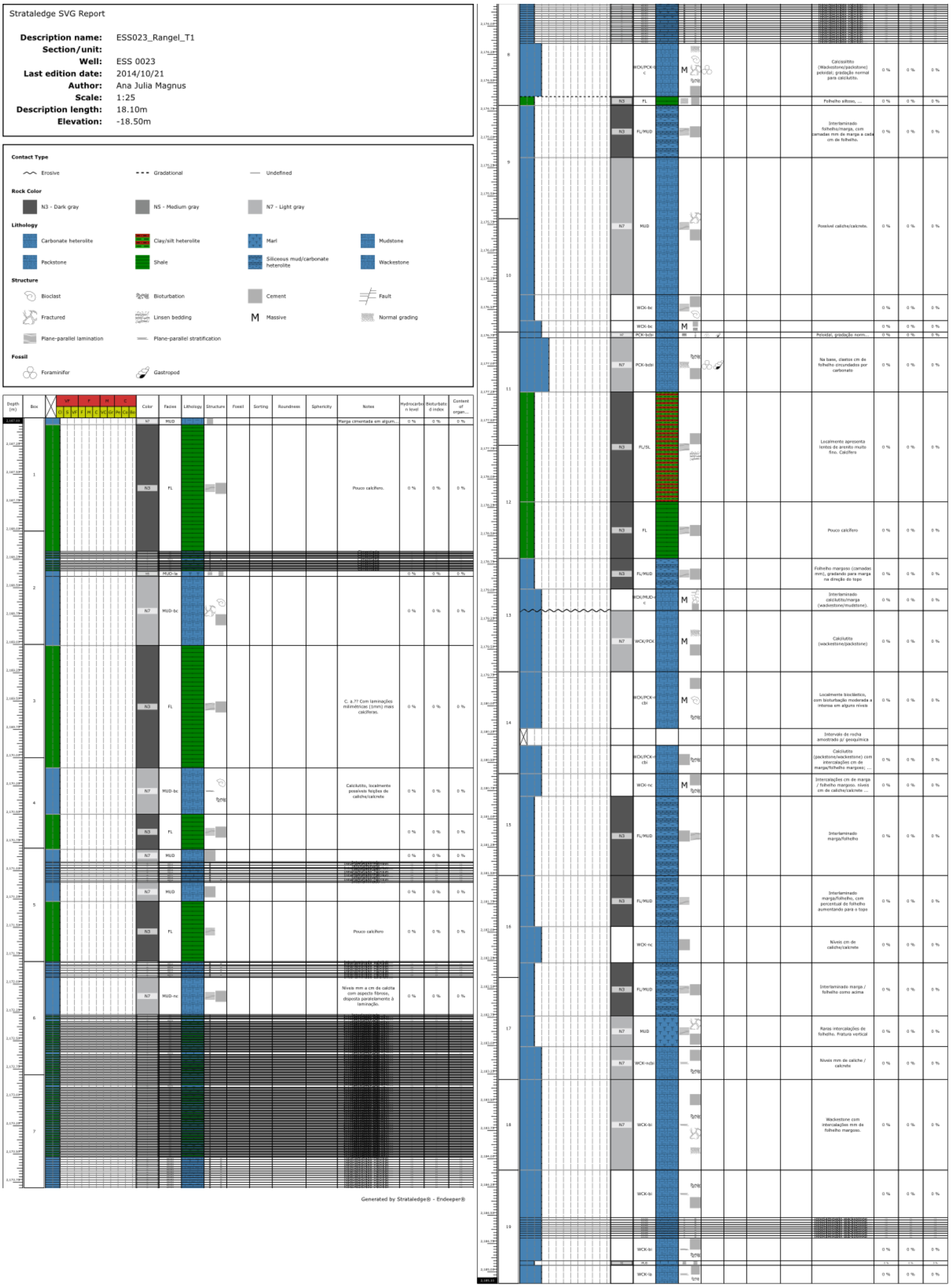

Figure 4.1: Strataledge description of core sample T1 of well ESS-0023

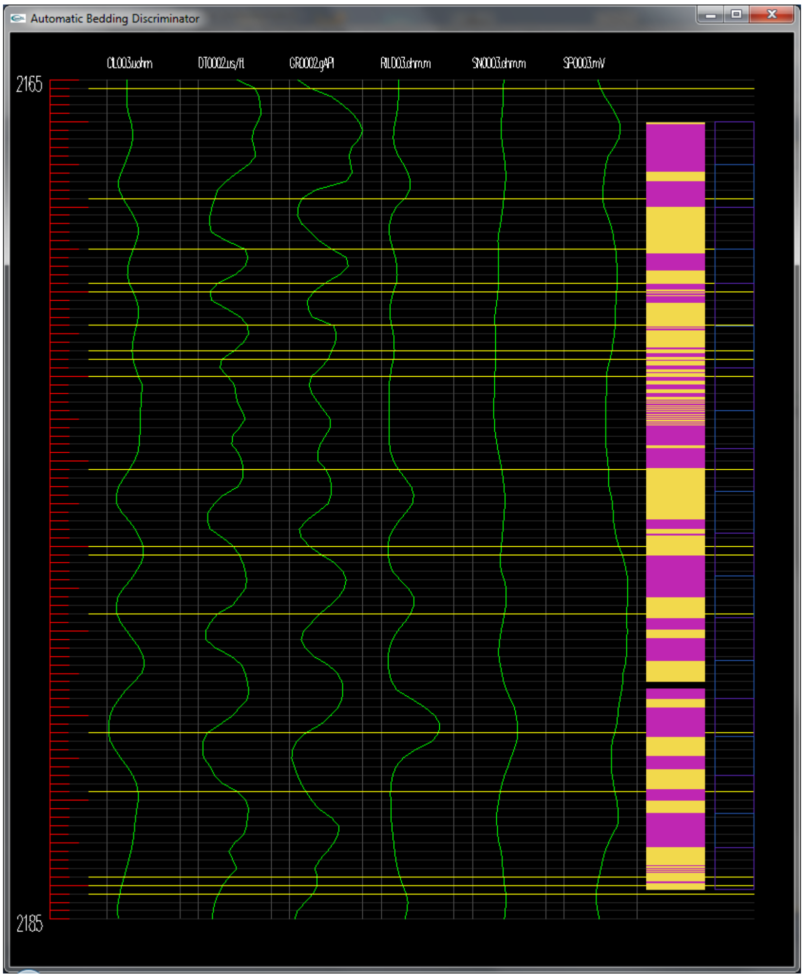

Figure 4.2: Bedding planes inferred by analysis of wireline logs in well ESS-0023, core T1 is displayed on the right. Note that only the bedding planes are shown on the core, the colors on the core are not representative of lithotype, they only show where the lithology changes.

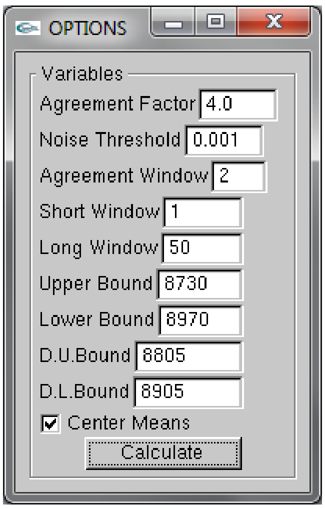

Figure 4.3: Parameter used in core T1, well ESS-0023

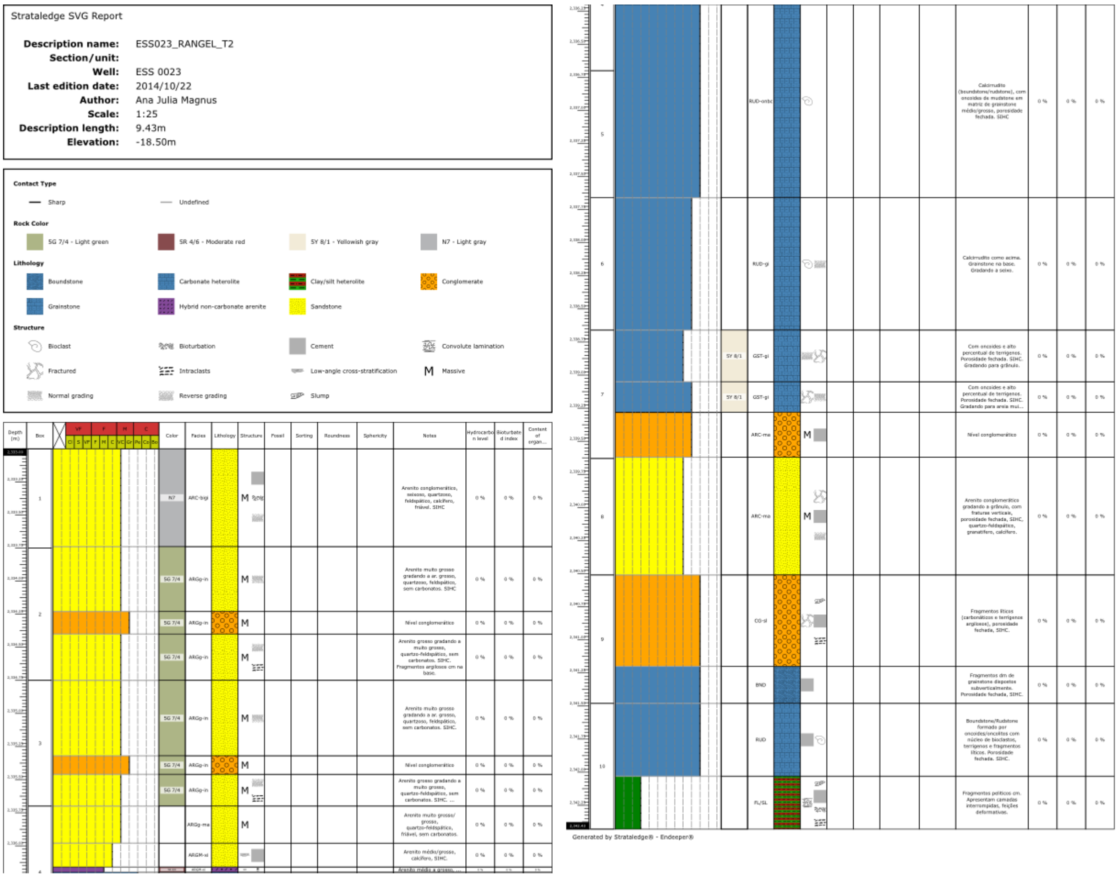

Figure 4.4: Strataledge description of core sample T1 of well ESS-0023

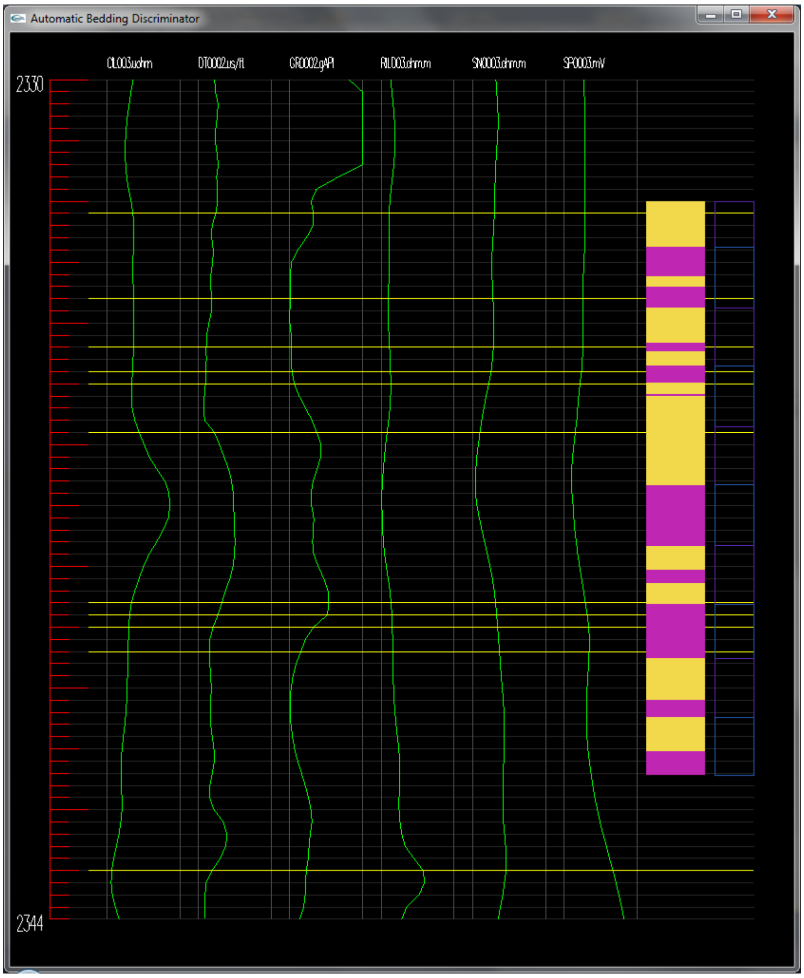

Figure 4.5: Bedding planes inferred by analysis of wireline logs in well ESS-0023, core T2 is displayed on the right. Note that only the bedding planes are shown on the core, the colors on the core are not representative of lithotype, they only show where the lithology changes.

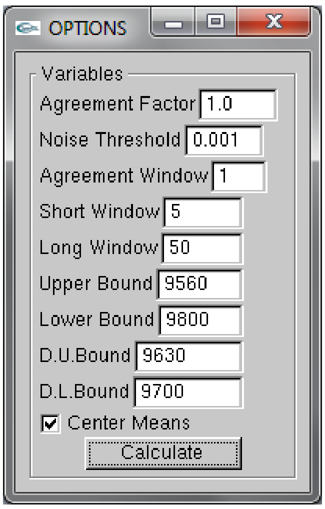

Figure 4.6: Parameter used in core T1, well ESS-0023

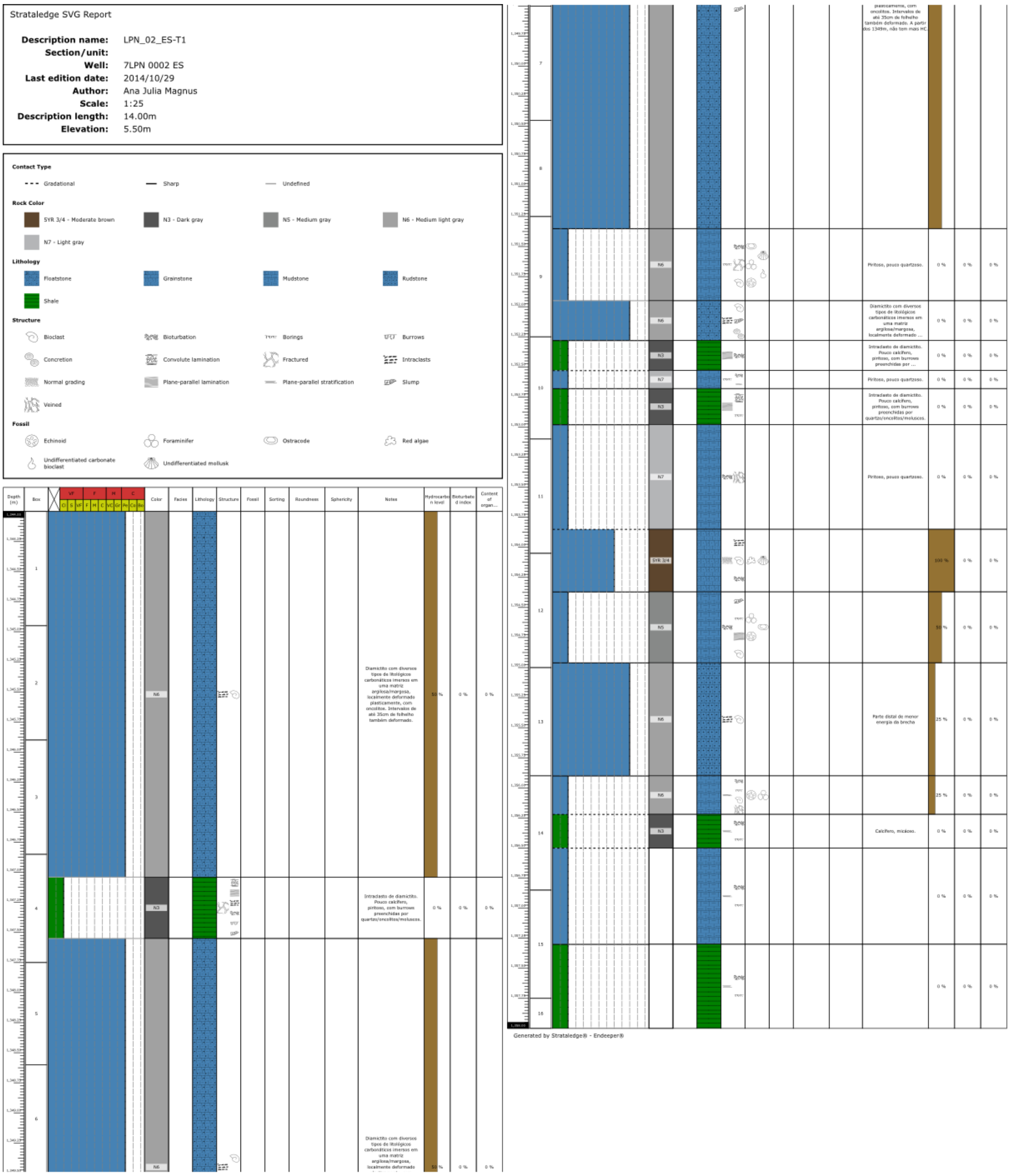

Figure 4.7: Strataledge description of core sample T1 of well LPN-02

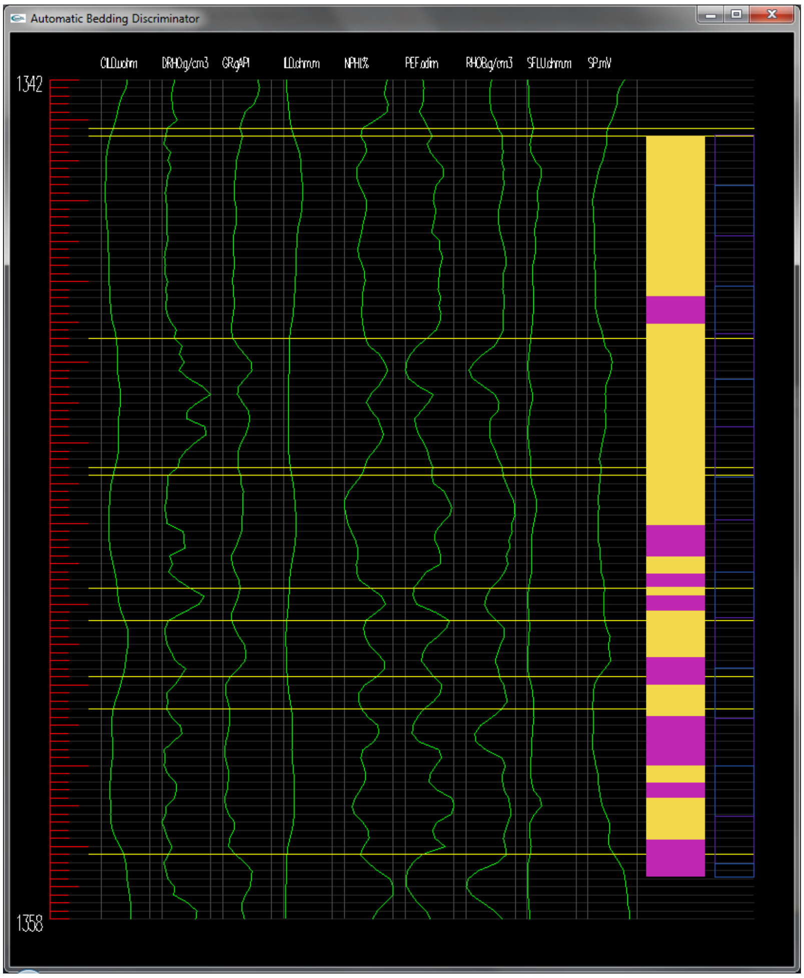

Figure 4.8: Bedding planes inferred by analysis of wireline logs in well LPN-02, core T1 is displayed on the right. Note that only the bedding planes are shown on the core, the colors on the core are not representative of lithotype, they only show where the lithology changes.

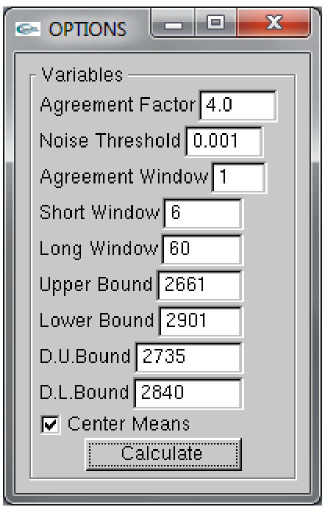

Figure 4.9: Parameter used in core T1, well LPN-02

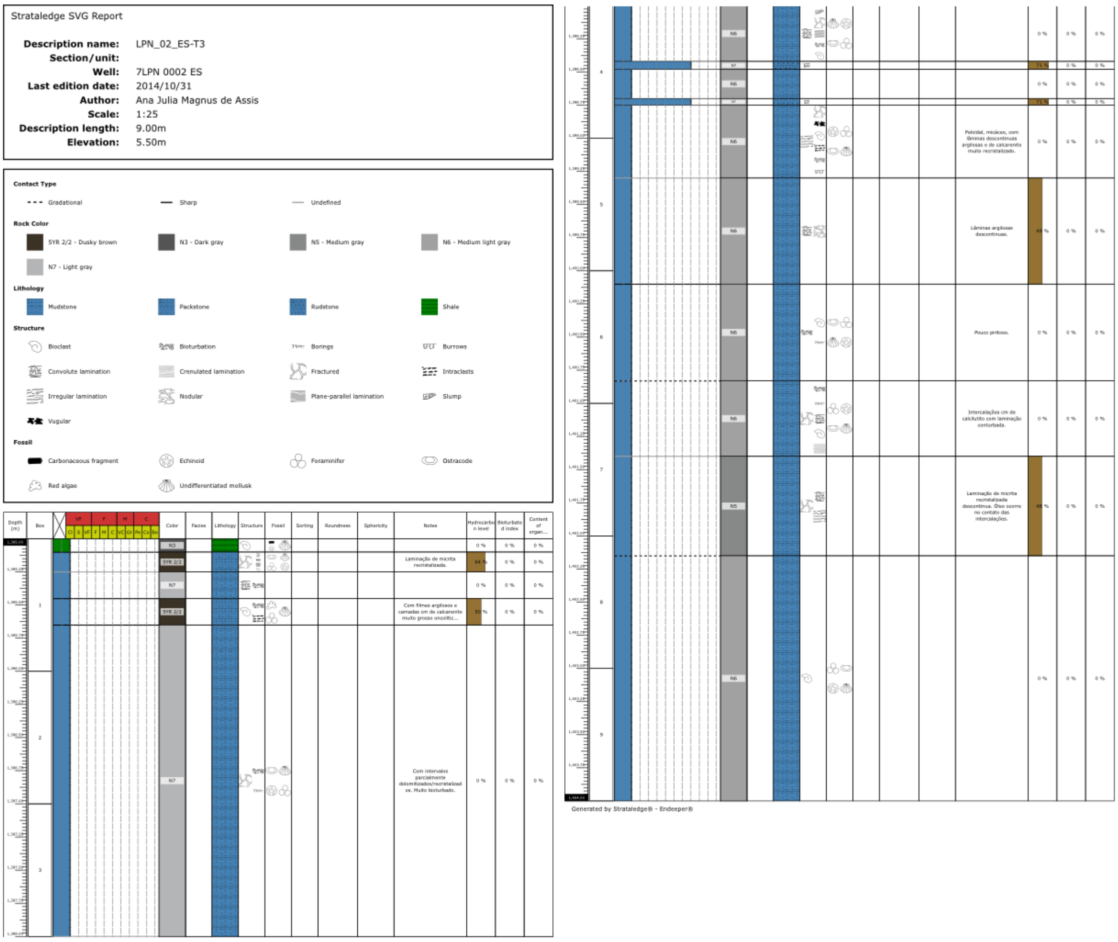

Figure 4.10: Strataledge description of core sample T3 of well LPN-02

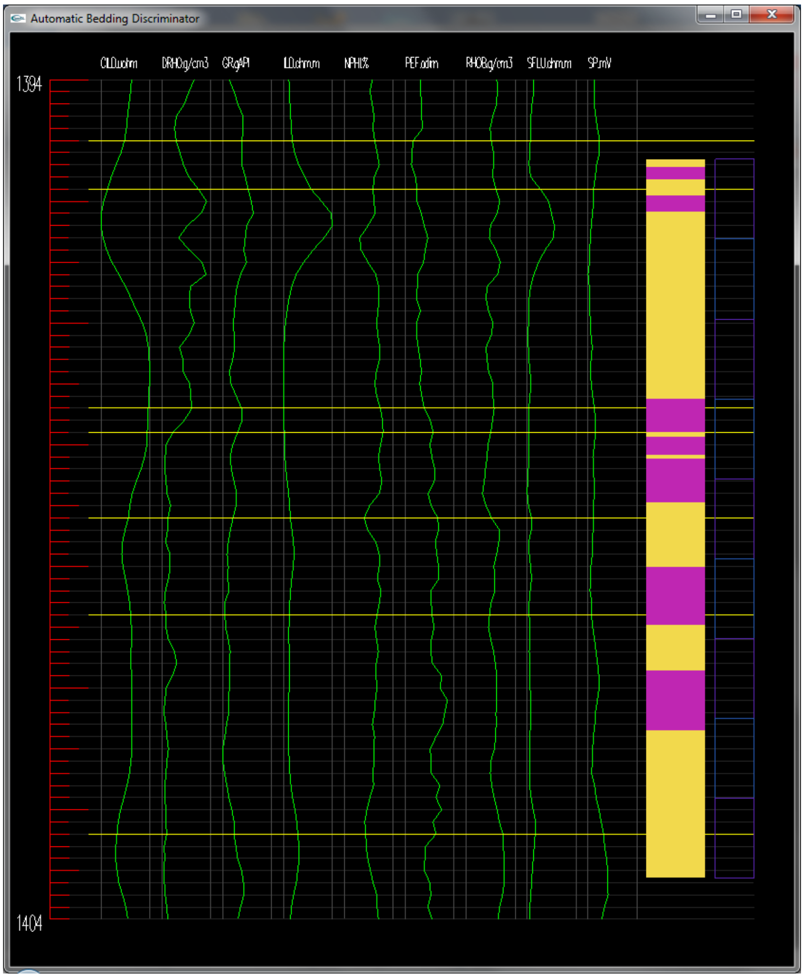

Figure 4.11: Bedding planes inferred by analysis of wireline logs in well LPN-02, core T3 is displayed on the right. Note that only the bedding planes are shown on the core, the colors on the core are not representative of lithotype, they only show where the lithology changes.

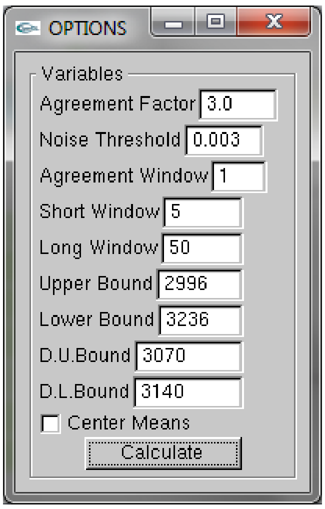

Figure 4.12: Parameter used in core T3, well LPN-02

Curve Mnemonic	**Weight**
DEPT.M	0
CALI.in	0
CALI02.in	0
CIL.uohm	1.5
CIL003.uohm	1.5
DRHO.g/cm3	3
DT.us/ft	3
DT0002.us/ft	3
DT0003.us/ft	3
GR.gAPI	3
GR0002.gAPI	3
GR0003.gAPI	3
RHOB.g/cm3	3
RILD.ohm.m	1.5
RILD03.ohm.m	1.5
N.ohm.m	1.5
SN0003.ohm.m	1.5
SP.mV	1.5
SP0003.mV	1.5
TOT.pol	0
TOT002.pol	0
TOT003.pol	0
TTI.data	0
TTI002.data	0
TTI003.data	0
LITO.	0

Table 4.1: Weights used for logs on well ESS-0023

Curve Mnemonic	**Weight**
DEPT.M	0
CALI.in	0
CILD.uohm	1.5
DRHO.g/cm3	3
GR.gAPI	3
ILD.ohm.m	1.5
NPHI.%	3
PEF.adim	1.5
RHOB.g/cm3	3
SFLU.ohm.m	1.5
SP.mV	1.5
TENS.lbf	0

Table 4.2: Weights used for logs on well LPN-02

Core	Lithotypes	Breaks Detected Within 1 Interval	Breaks Detected Within 2 Intervals	False Positives	False Negatives
ESS-0023 T1	8	15	2	0	21
ESS-0023 T2	7	8	0	4	11
LPN-02 T1	4	4	2	2	9
LPN-02 T3	5	6	1	1	7

Table 4.3: Breakdown of the results shown in this chapter.

5 CONCLUSION

Despite being based on a method that was proven to provide reliable results, and with improvements based upon strong theoretical foundations validated by specialists in the field of geology, the results of my work proved lacking in providing sufficient information to perform core depth-matching on the data sets that were provided for testing. This can be ascribed to the fact that the data sets used were ill-suited to wireline log analysis; the abundant carbonate rocks present on the wells used can not be easily distinguished in wireline log data.

However, even in those cases the algorithm is still capable of detecting the more well-defined bedding planes. This information can still be useful to a specialist to perform core depth matching assuming the specialist can identify which of the bedding planes described on the core sample are the most obvious and thus, more likely to be detected by the application.

It is also worth noting that in order to achieve these results, it was necessary to experiment with the parameters until a better fit was reached, in a real situation, where the core was not adjusted, comparing the application output with the core description would be harder, since there would be no direct correlation between core and log depths.

Limitations arise with this technique's high dependence on quality log data, the size of the step between measurements can have a deep impact on the results. Not only a geological feature thinner than the step size can be skipped over by the measuring tool, accuracy suffers due to the fact that breaks can only be marked on the sampled depths, when most likely, they occur between them. Interpolating points between sampled depths can help with the latter problem, but not with the former.

It is also of note that the method does not look specifically for lithology changes, but for any heterogeneity. While the most common heterogeneities will be changes in lithology, they can also be caused by other features, such as presence of fluid in the rock formation or diagenesis. This is still useful information however, as these features can be identified in the core sample.

Enhancements to the method can be made, such as varying the parameters dynamically based on the nearest data; for example, it can be a good idea to lower the noise threshold in a particular monotonous section of the log or decrease the size of the short moving means window on a section where many thin facies are suspected. It can also be an interesting idea to analyze the core sample data and adjust the parameters in a way that the application finds facies of thickness similar to the ones described on the core.

A procedure to adjust the core depth based on bedding plane assessment automatically can also be developed, but viability and accuracy would be highly dependent on results from the bedding plane assessment.

In conclusion, while the method is sound and shows some promise, more extensive testing with better datasets is required in order to prove its viability. Validation must also be made with specialists in order to verify whether or not the results are useful for the task of core depth-matching, and how accurate is matching done with this data.

REFERENCES

BRERETON, N.; GALLOIS, R.; WHITTAKER, A. Enhanced lithological description of a Jurassic mudrock sequence using geophysical wireline logs. **Petroleum Geoscience**, [S.l.], v.7, n.3, p.315–320, 2001.

CRANGLE, R. D. Log ASCII Standard(LAS) Files for Geophysical Wireline Well Logs and Their Application to Geologic Cross Sections through the Central Appalachian Basin. **Open-file Report. U. S. Geological Survey**, [S.l.], p.14, 2007.

DURANTI, D. et al. Injected and remobilized Eocene sandstones from the Alba Field, UKCS: core and wireline log characteristics. **Petroleum Geoscience**, [S.l.], v.8, n.2, p.99–107, 2002.

FLEMINGS, P.; BEHRMANN, J. H.; JOHN, C. Gulf of Mexico Hydrogeology (Expeditions 308 of the riserless drilling platform from Mobile, Alabama, to Balboa, Panama, Sites U1319-U1324). **Proceedings of the Integrated Ocean Drilling Program: Scientific Results**, [S.l.], v.308, p.1–70, 2006.

FONTANA, E.; ITURRINO, G. J.; TARTAROTTI, P. Depth-shifting and orientation of core data using a core–log integration approach: a case study from odp–iodp hole 1256d. **Tectonophysics**, [S.l.], v.494, n.1, p.85–100, 2010.

GIFFORD, C. M.; AGAH, A. Collaborative multi-agent rock facies classification from wireline well log data. **Engineering Applications of Artificial Intelligence**, [S.l.], v.23, n.7, p.1158–1172, 2010.

HYNE, N. J. **Nontechnical guide to petroleum geology, exploration, drilling, and production**. [S.l.]: PennWell Books, 2012.

KRYGOWSKI, D. A. Guide to petrophysical Interpretation. **Austin Texas USA**, [S.l.], 2003.

LOFTS, J. C.; BRISTOW, J. Aspects of core-log integration: an approach using high resolution images. **Geological Society, London, Special Publications**, [S.l.], v.136, n.1, p.273–283, 1998.

MACDOUGALL, S.; NANDI, A. K. Hybrid Bayesian procedures for automatic detection of change-points. **Journal of the Franklin Institute**, [S.l.], v.334, n.4, p.575–597, 1997.

MALINVERNO, A. **Well Logging Principles and Applications G9947 - Seminar in Marine Geophysics**. 2008.

MANN, U.; LEYTHAEUSER, D.; MÜLLER, P. Relation between source rock properties and wireline log parameters: an example from lower jurassic posidonia shale, nw-germany. **Organic geochemistry**, [S.l.], v.10, n.4, p.1105–1112, 1986.

MORTON-THOMPSON, D.; WOODS, A.; GEOLOGISTS, A. **Development Geology Reference Manual**: aapg methods in exploration series, no. 10. [S.l.]: American Association of Petroleum Geologists, 1993. (Methods in Exploration Series).

OJHA, M.; MAITI, S. Sediment classification using neural networks: an example from the site-u1344a of {IODP} expedition 323 in the bering sea. **Deep Sea Research Part II: Topical Studies in Oceanography**, [S.l.], n.0, p.–, 2013.

PERRIN, M. et al. Ontology-based rock description and interpretation. **Shared Earth Modeling: Knowledge based solutions for building and managing subsurface structural models.**, [S.l.], 2012.

REID, I.; LINSEY, T.; FROSTICK, L. E. Automatic bedding discriminator for use with digital wireline logs. **Marine and petroleum geology**, [S.l.], v.6, n.4, p.364–369, 1989.

SIDDIQUI, S. **Method for depth-matching using computerized tomography**. US Patent 6,876,721.

VINEGAR, H. J.; WELLINGTON, S. L. **Method of correlating a core sample with its original position in a borehole**. US Patent 4,542,648.

WORTHINGTON, P. F. et al. Reservoir characterization at the mesoscopic scale. **Reservoir Characterization II**, [S.l.], p.123–165, 1991.