Semi-automatic well-log correlation using the Dynamic Time Warping Algorithm

Matheus Nilo*[1,2], Alexsandro G. Cerqueira[1,2], Vinicius Carneiro[1], Michael Holz[1], [1]GAIA-UFBA, [2]INCT-GP

This paper was prepared for presentation during the 18th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 October 2023.

Abstract

Well log correlation is an essential step in geophysical interpretation. However, as the number of wells increases, the complexity of the interpretation process and subjectivity also increases. To overcome this drawback, this work proposed a methodology based on the Dynamic Time Warping algorithm to establish a correlation between well logs, without prior interpretation. From there, it was possible to identify and correlate well log patterns that allow, in some cases, to identify a set of corresponding depths for each pair of profiles. Sometimes, the depth that better align the log – here treated as a time series – can be enormous, without any geological meaning. Therefore, to avoid this problem, a restriction limit was established relating the value of the cost function of the curves with the band restriction parameter *Sakoe-Chiba* through graphical analysis. In this way, it is possible to determine the displacement depth between the wells that best influence the correlation using a method based on the derivative of the DTW cost function and optimize the correlation process by establishing the best plateaus. To show that the correlation via the DTW algorithm is an adequate alternative, applications were performed on synthetic datasets that simulate four different patterns associated with distinct depositional environments. Once confirmed that the DTW can be used to perform correlation between synthetic datasets, We employed the methodology in well logs located in the Miranga Low, one of the deepest areas in the Reconcavo basin. This methodology showed promise in correlation workflow due to the new insights that can be brought to geophysicists and geologists.

Introduction

The correlation of geological layers using well logs is a challenging step in the area of petroleum geology since understanding the lithofacies distribution of a reservoir is essential for assessing hydrocarbon production. Geophysical data such as gamma-ray logs, slowness, and others can be used as inputs for the geological correlation of reservoir intervals. However, interpreting these logs requires matching the depth of each layer in different wells. Therefore, the analysis of these correlations can be limited by the uncertainty in the data and by the variability of the position of the layers — since they are done individually, the final result cannot be identical from interpreter to interpreter. Automating this process can substantially reduce the time spent on this task and help geophysicists and geologists eliminate ambiguity in their interpretations.

An alternative approach for correlation between the logs is using the Dynamic Time Warping (DTW) algorithm, a time alignment technique that allows the comparison of time series of different lengths and with different offsets. To use this algorithm to correlate wells, it is necessary to consider that the well logs behave like a regularly sampled time series. This algorithm demonstrates excellent ability in handling logs characterized by relatively simple stratigraphy (Wheeler* & Hale). However, it presents difficulties in more complex situations requiring data processing. One of the ways to use the DTW algorithm to minimize uncertainties in borehole correlation is to use the Sakoe-Chiba parameter, which limits the search window for this alignment to reduce the noise effect and increase the correlation accuracy.

This work proposes using the Dynamic Time Warping algorithm to determine a correlation between a pair of wells from the gamma-ray sonic records. For this, the Sakoe-Chiba parameter was used to avoid unrealistic comparisons between the correlated depths, and an automation of the choice of this parameter was proposed, considering the rate of change of the function at the cost of the DTW algorithm. The experiments were reproduced on three synthetic logs and six wells from the Reconcavo sedimentary basin located in the region of Cantagalo, Rio Fundo, and Polo Miranga 3D seismic surveys. Although the generated results show correlations for all samples, the stratigraphic intervals of the rift section where there is a variation from a deltaic depositional environment to gravitational flows are marked for better visualization.

Well log interpretation

According to Kendall and Pomar (2005), possible patterns of stratigraphic stacking and the interpretations of depositional environments that result from them can be inferred by analyzing the behavior of the gamma-ray logs. The main forms of GR profile (Figure 1) that have frequently been used to interpret the sedimentary setting are:

- **Cylindrical** - Low gamma-ray value, sharp edges, and no internal change: this tendency is predominant in braided river channel sands and aeolian sands.
- **Funnel** - Gradual upward decrease in gamma response. This trend in shallow marine environments reflects a lithology shift from shale to sandstone and an coarsening upward in depositional energy.

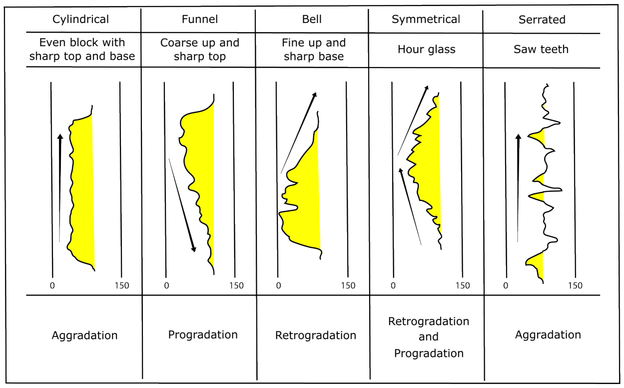

Figure 1 – Definition of stratal stacking patterns using the gamma-ray logs. Modified from Kendall and Pomar (2005).

- **Bell** - Gradual upward increase in gamma response: This trend may reflect a fining upward interval (e.g., a lithological shift from sand to shale) or the gradual thinning of sand layers in a shale-sand intercalated unit.
- **Symmetrical** - Gradual decrease succeeded by a gradual increase in the gamma response: this is usually the result of progradation and retrogradation of clastic sediments or the intercalation of thin layers of turbidites.
- **Serrated** - This trend represents the aggradation of shales or silts and can occur in other settings.

The previously quoted standards were used to simulate synthetic gamma-ray logs to employ the proposed methodology.

Dataset

In order to certify that the proposed methodology is viable, three synthetic gamma-ray logs were generated (Figure 2) to simulate sedimentary environment patterns similar to those shown in Figure 1. Furthermore, in order to simulate a more realistic situation, a normal noise was added to the data to check if the alignment obtained by the DTW algorithm can be capable of handling with the presence of noise in addition to problems related to the lag between signatures of interest, absence of layers and changing the amplitude of the signal.

Log A shows a sequence considered standard. Log B shows the same sequence A but with a depth shift of 500m, simulating a situation where the wells are located in blocks separated by a normal fault or reverse. Finally, log C shows a succession where the funnel, bell, and cylindrical patterns varied the absolute values but followed the same curve trend.

Then, logs from six different boreholes located in the sedimentary Reconcavo basin were used to attest the algorithm's robustness and behavior on real data. The wells show a siliciclastic sedimentary succession that includes the main formations of the basin from the turbidites and debrites of Candeias Formation at the bottom to the fluvial sandstones of the São Sebastião Formation at the top. The correlation from the DTW algorithm can be carried out from any of the measures of petrophysics, e.g., gamma rays, electrical resistivity, slowness, density, neutronic, etc. Nevertheless, this work proposes understanding the automatic correlation from the point of view of standards related to sedimentary environments, and it was decided to use data from gamma rays and slowness.

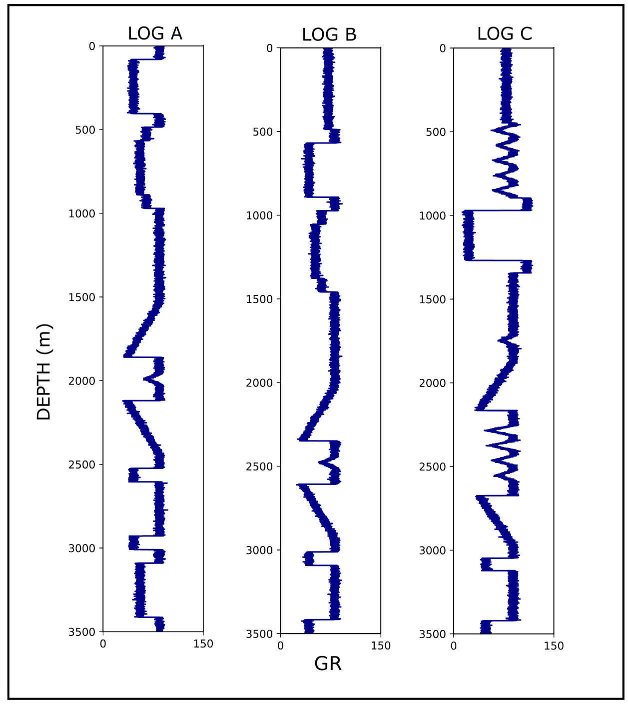

Figure 2 – Synthetic gamma-ray logs simulating the features presented by stratal stacking patterns, similar to those shown in Figure 1.

Methodology

The approach adopted in this work was implemented using the Python programming language, making use of the packages: tslearn (Tavenard et al.), numpy (Harris et al.), scipy (Virtanen et al.), pandas (Wes McKinney) and matplotlib (Hunter). The workflow adopted is shown in Figure 3.

Median filter

Median filtering is performed by moving a window over the points in a sequence and replacing the value in the center of the window with the median of the original values in a given range, which produces an output sequence that is generally smoother than the original data. For example, Figure 4 illustrates the smoothing caused by a median filter with a window of 101 samples in wells 1-BRSA-502-BA (4a) and 4-BRSA-895D-BA (4b). Interestingly, although the number of samples is reduced, the curve maintains the general profile variation behavior.

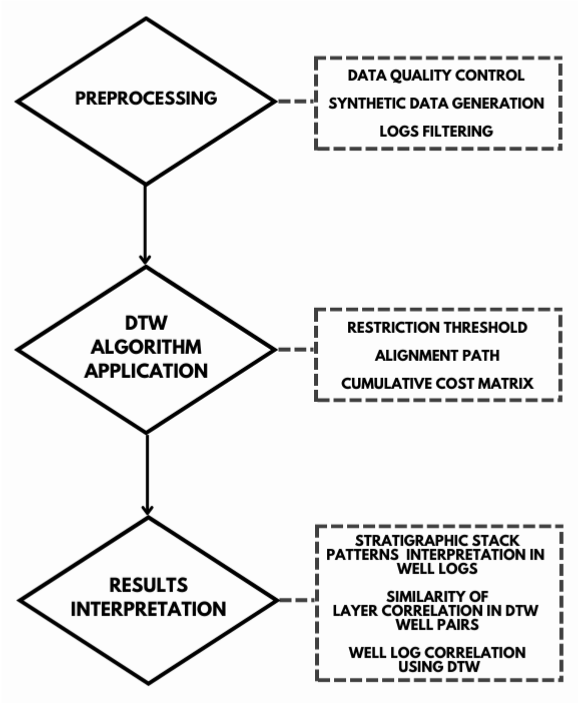

Figure 3 – Workflow employed to perform the well-log correlations using the Dynamic Time Warping Algorithm.

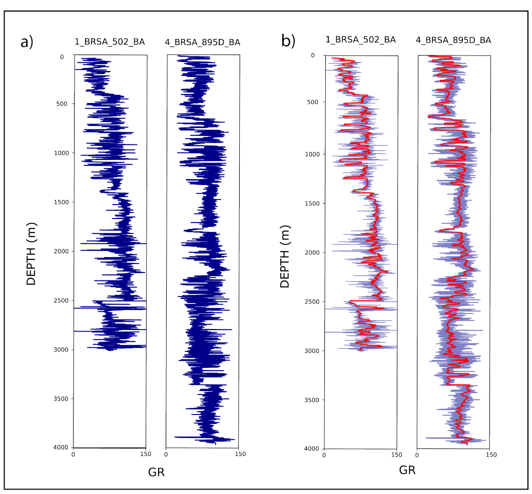

*Figure 4 – Figure **a)** gamma-ray logs in wells 1-BRSA-502-BA and 4-BRSA-895D-BA without smoothing; **b)** shows the original curve (blue) and the smoothed measure (in red) after filtering using the 101-sample window median filter.*

Well log correlation

The computationally most expensive part of the correlation process is the execution of the DTW algorithm. Therefore, correlating dozens of wells simultaneously requires significant time and computational resources, in addition to adaptations in the algorithm. Therefore, this work was restricted to comparing pairs of wells.

Let us consider two-time series, but with different numbers of elements, $\mathbf{x} = x_1, ..., x_N$ e $\mathbf{y} = y_1, ..., y_M$ to demonstrate the operation of the DTW algorithm. The first step is to build the local cost matrix, given by:

$$\mathbf{C} \in R^{NxM}: \quad c(i,j) = ||x(i) - y(j)||, \quad i \in [1:N],\ j \in [1:M]. \quad (1)$$

Once the local cost matrix has been defined, it is necessary to define the accumulated cost matrix, which will be defined with **D** of dimensions $Nx \times N_y$, which is built using the following steps:

1. **First row**: $D(1,j) = \sum_{k=1}^{k} c(1,k)$;
2. **First column**: $D(i,1) = \sum_{k=1}^{i} c(k,1)$
3. **All other elements**: $D(i,j) = c(i,j) + min(D(i-1,j-1), D(i-1,j), D(i,j-1))$

The algorithm finds the best alignment, which is the one that runs through the lowest ”cost” areas within the matrix. For this, some criteria must be followed, such as boundary condition, monotonicity, and step size condition. Senin (2008).

Considering the sequence of points $p = (p_1, p_2, ..., p_K)$ with $p_l = (p_i, p_j)$, where p_i and p_j represent the alignment indices of the matrix, the cost function associated with the alignment path is given by the sum of the value of the cost function accumulated over the alignment, and it can be estimated through Eq. 2.

$$D_{dtw} = \sum_{l=1}^{K} D(p_l) \quad (2)$$

Figures 5 and 6 show the cumulative cost matrix and correlation, respectively, obtained by the Dynamic Time Warping algorithm, applied to synthetic logs A and B displayed in Figure 6. Due to the large number of samples, the well logs were resampled to show how each sample relates to the well to be correlated (on the right) from the lines in black.

Global constraint: Sakoe-Chiba

Although the algorithm returns the lowest value of the cost function and correlation between each sample, it was observed that samples of the Reconcavo basin data were presenting a very ”distorted” result, i.e., there were alignments with a considerable difference in depth. In order to solve this problem, it was necessary to use a local constraint so that the algorithm could execute the best

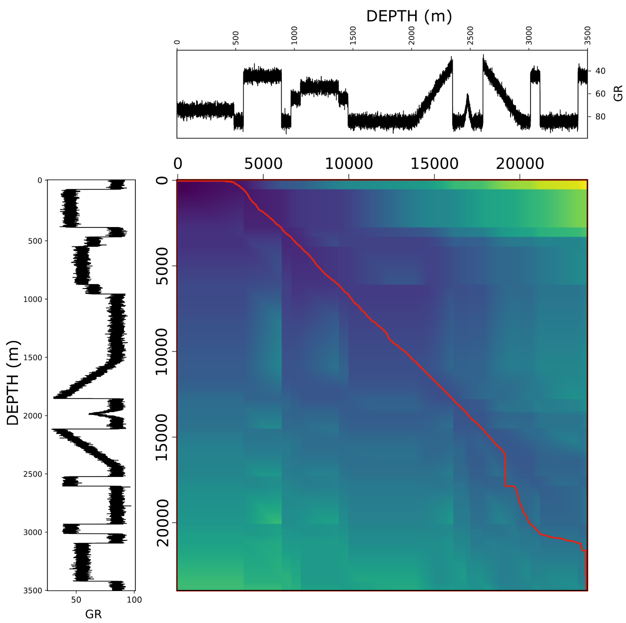

Figure 5 – Cumulative cost matrix and alignment path (in red) of A (left) relative to B (above the cost matrix).

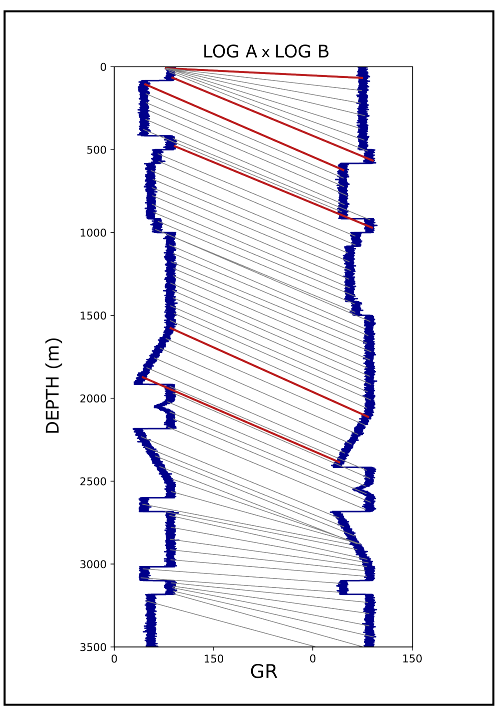

Figure 6 – Correlation lines obtained by the DTW algorithm between synthetic logs A and B, indicating which samples in depth correlate with each other, according to the alignment path obtained. The red lines highlight the interfaces between the patterns considered key in the correlation between the profiles.

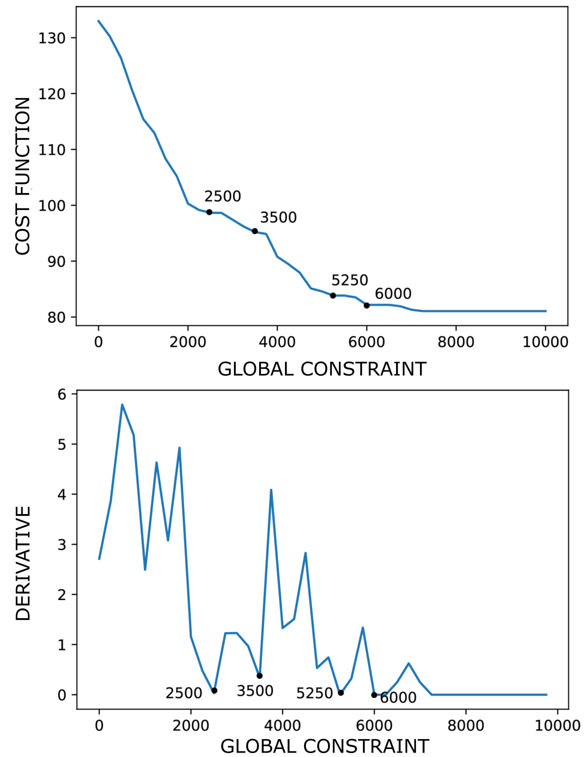

Figure 7 – Figure (a) Graph of the cost function DTW versus the constraint parameter of Sakoe-Chiba (ω), and (b) first derivative of the cost function. The black dots show the plateaus found in the graph, which were the subject of study to define the optimal parameters for performing the correlation between the 1-BRSA-502-BA and 4-BRSA-895D-BA logs.

alignment path within a limited window in the accumulated cost matrix and analyze the behavior of the cost values concerning the constraint parameter from Sakoe-Chiba. A graphical analysis of the value of the cost function of the DTW algorithm in relation to the Sakoe-Chiba band was used to find the value of the parameter less subjectively. The Sakoe-Chiba band restricts the comparison interval between the time series through a positive real number ω, which limits the interval traversed in the accumulated cost matrix, preventing it from deviating too much from the diagonal main.

In order to determine which value of the ω parameter should be adopted, the graph of the rate of change (first derivative) versus the Sakoe-Chiba parameter was scrutinized to establish plateaus in this curve in order to indicate a point of significant change in the function behavior. Figure 7 shows a graph with the value of the cost function concerning the parameter adopted in the Sakoe-Chiba constraint (ω) and the possible "plateaus" that can be the choice of parameter restriction.

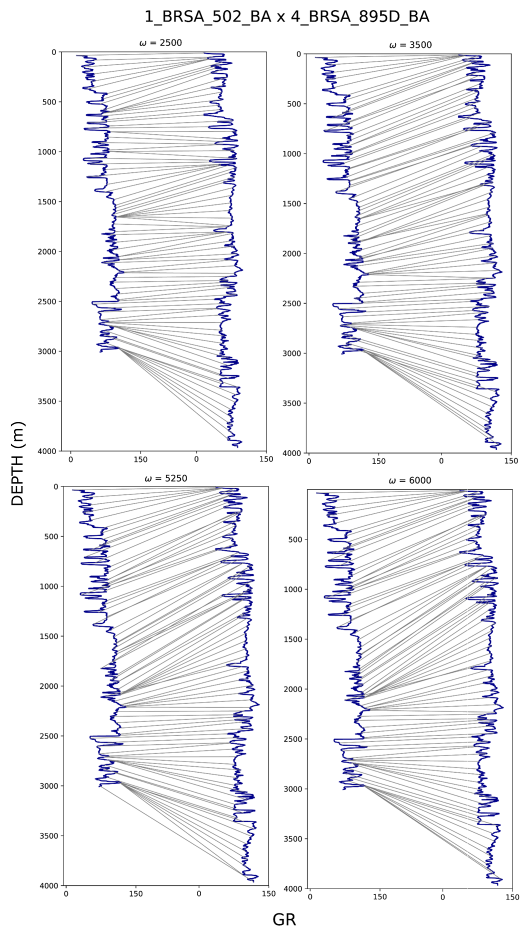

Figure 8 – Correlation lines between the well logs from 1-BRSA-502-BA and 1-BRSA-895D-BA, with different values of the restriction parameter ω.

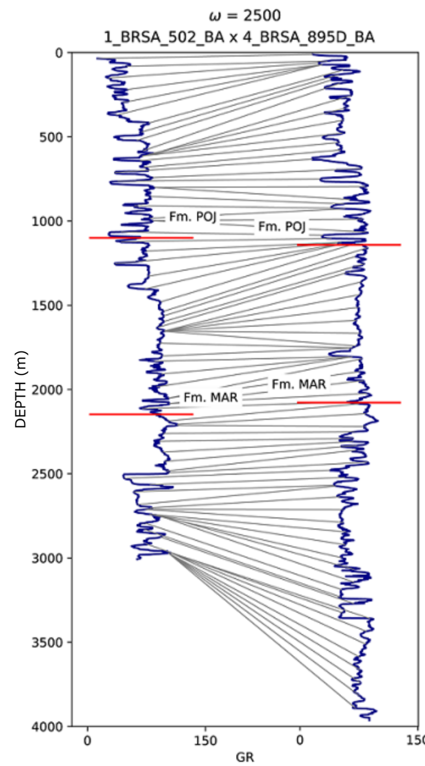

Figure 9 – Correlation lines obtained by the DTW algorithm between the gamma-ray logs of wells 1-BRSA-502-BA and 4-BRSA-895D-BA, indicating which samples in depth correlate with each other, according to the alignment path obtained. The red lines delimit the top of the Pojuca and Maracangalha formations according to the lithostratigraphy. The Sakoe-Chiba parameter adopted was 2500.

Results and Discussions

Once the possible parameters to be chosen were defined, a qualitative analysis was carried out based on the geophysical interpretation of the wells to understand which is the restriction parameter that best correlates the geological information between the logs, as it is possible to see in (Figure 8). It is important to emphasize that the curves vary in the number of samples due to the greater number of measurements in depth and, consequently, the restriction parameter can change for each situation.

The DTW algorithm does not always do the best alignment when no constraint parameters are imposed. However, the algorithm can be seen as an excellent tool that assists the interpreter in extracting insights to initiate correlations between wells. For this reason, to minimize the effects of distortions by samples with significant depth differences, the results obtained from candidates for appropriate Sakoe-Chiba parameters were analyzed. By choosing a parameter considered suitable for correlation by the interpreter, it was possible to observe a significant improvement in the correlation between the wells, as shown in Figure 9, where the adopted analysis parameter was $\omega = 2500$. It means that the temporal sample x_n can only be compared up to the samples y_{n-2500} and y_{n+2500}. To verify the reliability of the result, the tops of the Pojuca and Maracangalha Formations were used, according to the lithostratigraphy of the wells, to confirm if the algorithm could correlate this sedimentary package. It is possible to notice that even with a lateral variation between the logs, the algorithm identifies the events and correlates them clearly, since the trends of the curves are similar and the DTW is not sensitive to the variation of the absolute values.

Another way to validate the algorithm's efficiency is to examine its result for other geophysical logs, such as the sonic log. Once a first correlation has been made using the gamma-ray log, these results can be used to verify if it is possible to use another tool, in the absence of the GR, to perform the correlation through the DTW algorithm.

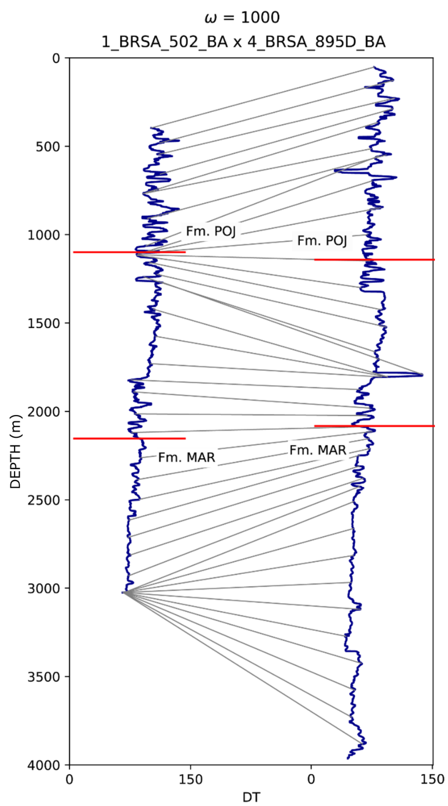

Figure 10 – Correlation lines obtained by the DTW algorithm between the seismic slowness logs of wells 1-BRSA-502-BA and 4-BRSA-895D-BA indicating which samples in depth correlate with each other, according to the alignment path obtained using the sonic log. The red lines highlight the patterns considered key in the correlation between the logs. The Sakoe-Chiba parameter adopted was 1000.

The Gamma Rays logs available are more sensitive to layer variations within an interval; that is, they are more subject to smaller scale variations in the measurements carried out in the well. This is due to the fact that it has a higher vertical resolution allowing to identify thinner layers. However, the sonic does not reflect these changes clearly, since the average of the transit time measurements does not highlight these variations in an interval so much, mainly if the layers are very thin. Figure 10 shows the results obtained for the same pairs of logs used to correlate through the DTW algorithm from GR measurements, now using the slowness log. It is important to note the similar alignment obtained by the DTW when analyzing the gamma-ray and seismic slowness logs.

In situations where the curves were with intervals without sampling, it was noticed that the algorithm had difficulty in correlating the samples since the points in gaps became biased concerning the depth variation. Thus, estimating the best parameter value was essential, as it prevented the algorithm from correlating samples with unreasonable depths. In this way, the large variations of the patterns in the profiles correspond with greater reliability and generate better results in the correlations.

Acknowledgments

We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Instituto Nacional de Ciência e Tecnologia de Geofísica do Petróleo (INCT-GP) for the support provided to the research. We also thank the Group for the Study and Application of Artificial Intelligence in Geophysics at UFBA and the Centro de Pesquisa em Geofísica e Geologia (CPGG-UFBA) for providing the appropriate infrastructure conditions for the development of this work. Co-Author M.H acknowledges CNPq for personal study grant, and co-author V.C. thanks the Fundacao de Amparo a Pesquisa do Estado da Bahia (FAPESB) for doctoral scholarship.

References

Harris, C. R. et al., 2020. Array programming with NumPy, Nature, vol. 585(7825): 357–362, doi:10.1038/s41586-020-2649-2, URL `https://doi.org/10.1038/s41586-020-2649-2`.

Hunter, J. D., 2007. Matplotlib: A 2d graphics environment, Computing in Science & Engineering, vol. 9(3): 90–95, doi: 10.1109/MCSE.2007.55.

Wes McKinney, 2010. Data Structures for Statistical Computing in Python, in: Stéfan van der Walt & Jarrod Millman (Eds.), Proceedings of the 9th Python in Science Conference, 56 – 61, doi:10.25080/Majora-92bf1922-00a.

Senin, P., 2008. Dynamic time warping algorithm review, Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA, vol. 855(1-23): 40.

Tavenard, R. et al., 2020. Tslearn, a machine learning toolkit for time series data, Journal of Machine Learning Research, vol. 21(118): 1–6, URL `http://jmlr.org/papers/v21/20-091.html`.

Virtanen, P. et al., 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, vol. 17: 261–272, doi:10.1038/s41592-019-0686-2.

Wheeler*, L. & Hale, D., 2014. Simultaneous correlation of multiple well logs, 618–622, doi:10.1190/segam2014-0227.1.