ALINE: A Computational System Based on Seismic Data and Machine Learning for Gas Reservoir Detection

Luiz Santos[1], Felipe Jordão[1], Marcelo Gattass[1], Roberto Quevedo[1], Maria Julia Lima[1], Diogo Michelon[2], Carlos Siedschlag[2], Roberto Ribeiro[2], Sebastião Pereira[2]

[1]*Tecgraf Institute/PUC-Rio*
Prédio Pe. Laércio Dias de Moura - R. Marquês de São Vicente, 225, 22451-900, Gávea, Rio de Janeiro - RJ
lsantos@tecgraf.puc-rio.br, felipejordao@tecgraf.puc-rio.br, mgattass@tecgraf.puc-rio.br, mjulia@tecgraf.puc-rio.br
[2]*Eneva*
R. Praia de Botafogo, 501, 22250-040, Botafogo, Rio de Janeiro - RJ
diogo.michelon@eneva.com.br, carlos.sied@eneva.com.br, roberto.ribeiro@eneva.com.br, sebastiao.pereira@eneva.com.br

Abstract. Reflection seismic is one of the most used geophysical methods by the O&G industry for subsurface imaging. Through the processing and interpretation of seismic data, geoscientists infer the positioning and geometry of potential hydrocarbon accumulations. However, the reflection seismic method can produce ambiguous data owing to similar signatures in natural bodies with different physical properties. Moreover, onshore seismic data have in general less quality when compared to offshore seismic data, making the interpretation process even more difficult. Artificial intelligence (AI) techniques have been adopted in several applications, particularly for the interpretation of salt bodies and geological faults. However, for identification of hydrocarbon reservoirs, AI techniques are still under development, particularly owing to the great amount of seismic data to be processed. Recently, Eneva and Tecgraf/PUC-Rio developed the computational system ALINE (Automated Learning Intelligence for Exploration) based on Machine Learning techniques and seismic data to generate indicators of potential gas accumulations in on-shore fields. In this study, we focus on the description of ALINE's system, its current capabilities and methods, advantage and limitations, and future developments. The current methodology uses modern neural network architectures through the analysis 1D of seimic traces to identify specific signatures of gas accumulation. Several onshore seismic sections from Parque dos Gaviões at Parnaíba's Basin were provided by Eneva during the first validation tests. The results obtained for that region showed an accuracy of 75 - 80% of the gas class and 90 - 95% of the non-gas class. Although other approaches for similar applications are not available in the literature for comparisons, the global average of success shows that the system has a significant potential for exploratory purposes. Moreover, ALINE's system also can be adopted for predictions considering offshore 3D seismic data, as performed on the Block F3 in the North Sea that enabled better accuracy rates owing the best quality of data. Those results highlight the potential of ALINE as a computational tool for the interpretation of 2D or 3D seismic data, onshore or offshore, boosting the value of seismic data and minimizing uncertainties, representing an effective technological advance in the sector O&G.

Keywords: Machine Learning; Seismic Acquisitions; Gas reservoirs; ALINE.

1 Introduction

The identification of certain anomalies in seismic data can be strong indications of hydrocarbon reservoirs. For their evaluation, amplitude anomaly visualization (Direct Hydrocarbon Indicator) approaches, joint and separate analysis of geophysical good log information, and seismic inversion algorithms are commonly used. However, the use of DHI is inefficient considering seismic data with targets with low signal-to-noise ratios, where often, the application of seismic attributes and inversion processes can reveal inaccurate information. Furthermore, these methodologies are characterized by being time-consuming and human-intensive, taking into account the high range of attributes that can be applied and inspected and the additional data processing required for the application of inversion algorithms.

Thus, several methodologies to semi-automate and assist geoscientists in interpretation processes similar to identifying possible gas signatures have been proposed in the literature, such as geological fault detection and salt dome design. Among them, the one gaining more prominence due to the results presented and the efficiency in handling large volumes of data are deep neural network algorithms (deep-learning). For the most part, such algorithms are applied through image analysis, convolutional networks, which extract features, and then make use of this information to find similarities in new data. Some more recent works already deal with such contrast between deep neural networks in the image domain and the seismic trace domain, presenting consistent results. However, to the best of our knowledge, no system present in the literature addresses the use of artificial intelligence techniques to identify possible gas accumulations in seismic data using seismic trace data.

To improve efficiency in the interpretation of seismic data and reduce exploration costs, Eneva and the Tecgraf Institute PUC-Rio have developed the ALINE system, which embeds a post-stack and seismic imaging methodology. Machine Learning algorithms to support the definition of regions with potential for gas accumulation.

The current version of the ALINE system supports 2D seismic data analysis specific to the collection of potential gas accumulations from Machine Learning or “Machine Learning” methods. The methodology currently implemented within the system considers three types of network architecture which can be used, namely: LSTM [1], [2] and Encoder-Decoder LSTM. From seismic data, the definition of regions of interest through the horizons that define the reservoir layer, the knowledge of regions with gas in the field, each architecture is able to learn to identify certain patterns associated with the regions with and without gas through a training process. From this process, each architecture defines a trained network with a set of weights which are used to predict new gas accumulations through an inference process.The trained networks can then be used to identify gas in new prospects through heat maps that identify regions with “gas” and “non-gas,” helping to define the location of new exploratory wells. The networks were trained based on information from various fields belonging to the “Parque dos Gaviões” in the Parnaíba Basin.

2 System Overview

ALINE (Automated Learning Intelligence for Exploration) system was developed from a partnership between Eneva and the Tecgraf/PUC-Rio Institute to improve the seismic interpretation processes of new exploratory blocks by identifying potential gas accumulations. The system is based on a pioneering methodology that uses seismic sections and Machine Learning techniques to recognize patterns associated with gas and non-gas regions. The architecture that ALINE is based on will be presented below, as well as some features.

2.1 Architecture

The system was created from the Soma platform [3] that uses modern technology based on the concepts of native cloud architecture and web interfaces, providing an environment for sharing data and solution methods with geographically distributed teams. The architecture is based on three main components: a web system, a repository of algorithms and projects, and an execution agent, all interconnected through a set of microservices, as shown in the Figure 1 below.

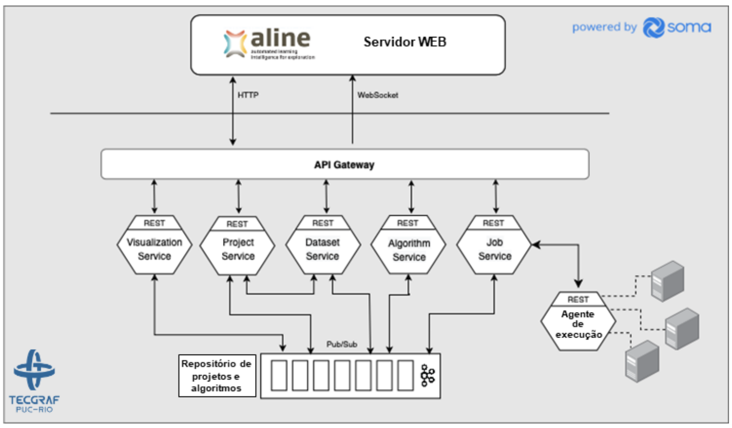

Figure 1. SOMA Architecture.

The system presents a friendly graphical interface that provides the user with visualization tools that allow using the tools and algorithms available in the system. The repository of projects and algorithms is a physical space in which all the algorithms and projects of all ALINE users are stored. In turn, the execution agent collects the algorithms that are part of the methodology and is in charge of executing these in graphics processing units (GPU).

All system functionality need just a browser. In this sense, the ALINE system does not require any installation process and does not depend on the processing capacity of the device in use.

2.2 Dataset Manager

In data science, control and storage are essential for building robust models with high generalizability. To ensure these objectives, a concept was implemented in the Aline system that allows the creation of datasets (data sets) practically and dynamically for training and testing the neural network models present in the system. Furthermore, the idea behind this concept is to guarantee the reproducibility of experiments carried out previously since each dataset is saved for future use.

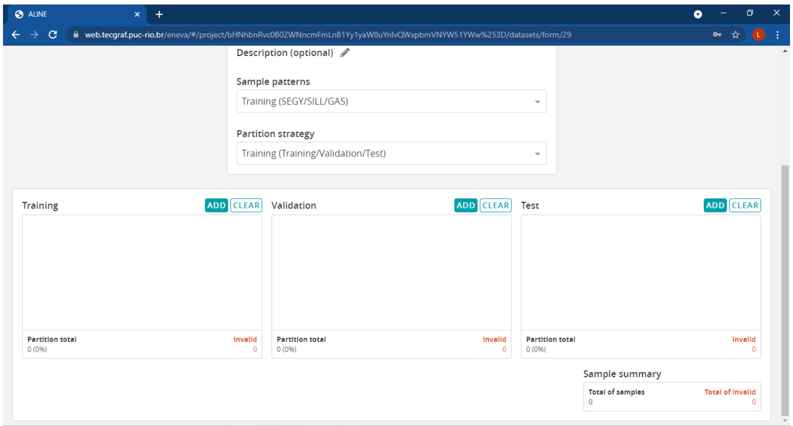

Figure 2. SOMA Dataset Manager.

2.3 Visualization

The seismic viewer is the web application integrated into the ALINE environment, which allows the user to visualize the seismic data, the gas and sill markings, and the inference result. The viewer also allows the user to change the gas and threshold markings. The primary user interaction elements that are part of the seismic viewer are presented below.

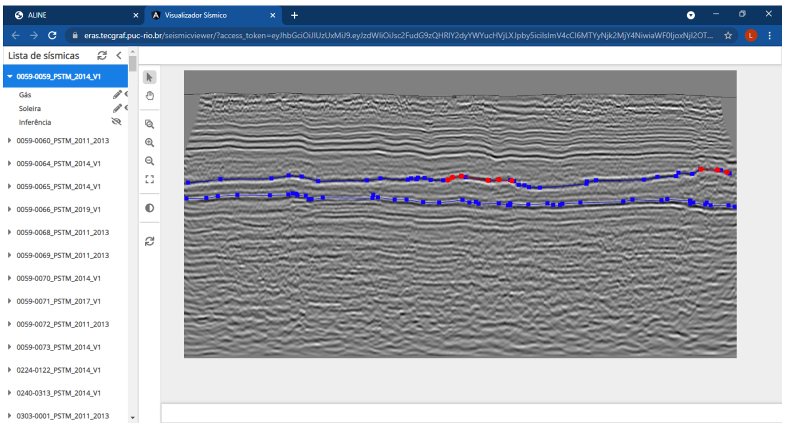

Figure 3. Seismic Viewer.

3 Experiments and Discussion

This section presents a brief description of the pre-processing carried out on the data, the database used, the experiments, and the results obtained.

3.1 Dataset

The data available for this work correspond to onshore 2D seismic lines from the "Parque dos Gaviões" [4], consisting of nine gas fields located in the Parnaíba Basin, a Paleozoic basin in northeastern Brazil. Eneva has been acquiring more seismic data in subsequent years to support its successful Reservoir-to-Wire business model. A significant difference is that the 2D seismic lines were processed from 2011 to 2020, a period in which the seismic acquisition parameters and the technology used in the Parnaíba Basin improved significantly, resulting in differences in the quality of the seismic line. These improvements can impact the results and define the approaches of our proposed method.

The difference in seismic datasets represents a significant challenge to the generalization of Deep Neural Networks. To compose our training and validation datasets, we created two scenarios to test our machine learning algorithm. The ultimate goal is to help the company improve its exploration success.

3.2 Pre-processing

Different onshore seismic surveys acquired over the years can differ from each other in terms of quality due to acquisition parameters and technology limitations. Also, older seismic surveys tend to have more noise, especially in the deepest part of the seismic section. For these reasons, we emphasize the need for the pre-processing steps described below:

- **Definition of region of interest (ROI):** this step focuses on the part of the seismic data that contains relevant information, ignoring parts of the seismic that we do not expect to contribute to our learning and decision process, as shown in Figure 4. It is an interactive step, where the interpreter provides a top polyline and a time interval in milliseconds delimiting the region of interest. The polyline marks the top sample of the ROI and the time interval defines the lowest sample. With two horizontal clips, we end up with a rectangular ROI.
- **Dataset Balance:** In a typical seismic survey, non-gaseous samples vastly outnumber those with gas accumulation. This discrepancy can cause the machine learning model to learn a more biased representation of the predominant class. To avoid this bias, for each processed 2D line, we randomly select the same number of ROI traces with and without the presence of gas.
- **Standardization:** customarily used when analyzing time series. We use it to minimize the effect that outliers have on our data. Here, this step is performed for each ROI trait.
- **Normalization min. max .:** normalize the range of values between -1 and +1. This step is often used in machine learning applications to help decision models better estimate weights since all observed samples are in the same range of values. Here, this step is performed for each ROI extracted.

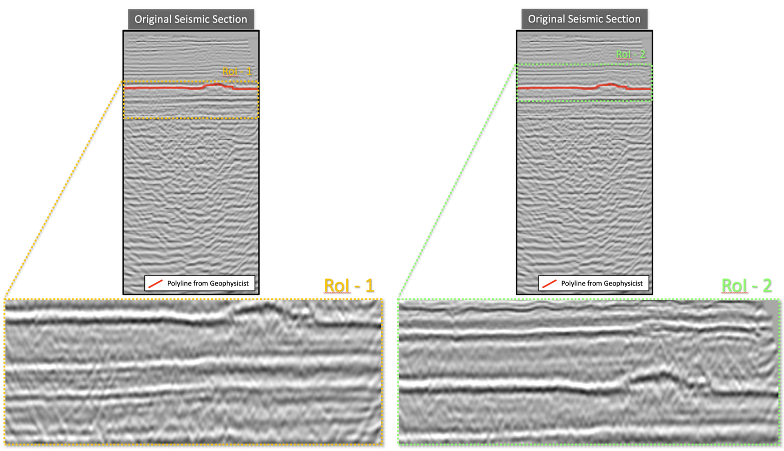

Figure 4. Example of ROI delimitation. The red horizon provided by geophysicist used to define the top of ROI.

3.3 Results

To compare the neural network models, currently present in the system, we used two datasets of interest from fields that have already been discovered containing gas reservoirs were selected in this work. The first dataset consists of a set of 10 2D lines, 8 lines were used for training, and two were reserved for testing.

Figure 5 shows the result of the models presents in the system for the first data set of Gavião Vermelho. We can see that from the first line, 0303-0291, both methods identified the location of gas accumulations observed in the field, even with several missing gas pixels, represented in yellow. However, this fact does not seem to affect the interpreter's perception. As the lines belong to the same field, they could see the accumulations more accurately because they present more similarities with the training data.

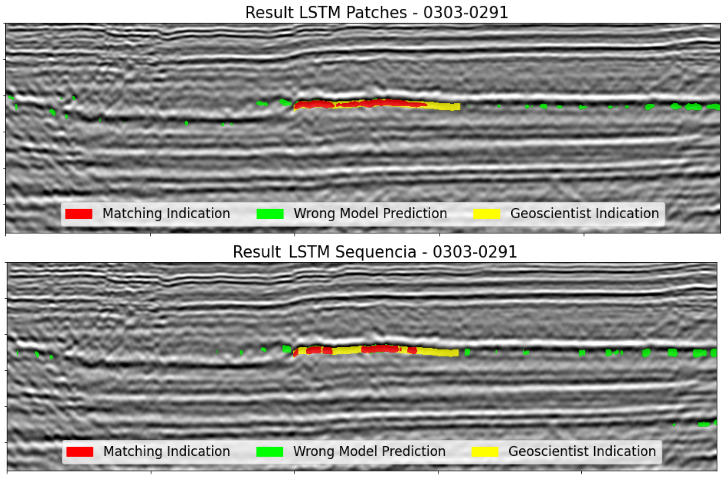

Figure 5. Model prediction result for Line 0303-0291.

The second dataset consists of two fields called Gavião Branco (GVB) and Gavião Preto (GVP). To train and validate the models the GVB field lines were used. In the testing phase, the entire set of GVP lines were used.

Figure 6 shows the results for the second dataset. To compare the methods, line 0303-0355 was selected. Observing Figure 6, it can be seen that both methods can adequately delineate the gas accumulation points. It is also noticed that in some cases, each model has more false positives than others.

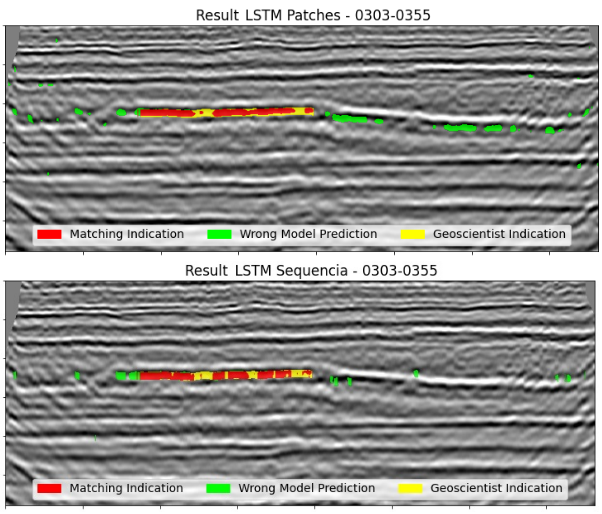

Figure 6. Model prediction result for Line 0303-0355.

4 Conclusions

To identify viable gas accumulation regions for development, several wells are drilled in the field, several without success, as reflection seismic (used in hydrocarbon prospecting) can generate ambiguous data, producing similar signatures in regions with different properties, or still, data with dimly lit/low-resolution targets. Eneva and Instituto Tecgraf PUC-Rio have developed a first version of the ALINE system to support seismic data interpretation and reduce exploration costs, which incorporates a post-stacking seismic data processing methodology and Machine Learning algorithms for identifying regions with potential for gas accumulation. The networks were trained based on information from a set of fields belonging to the “Parque dos Gaviões” in the Parnaíba Basin, providing very promising results in identifying gas in seismic sections not used during the training phase.

The presented methods in Aline proved to be robust, with good qualitative results. Thus, it can be concluded that the proposed method can help assist specialists in the task of detecting possible accumulations of gas in seismic surveys. And it was then optimizing the seismic interpretation procedure, reducing costs and time of a gas exploratory phase. As can be seen, this approach can become a powerful tool when it comes to seismic surveys with a lot of noise, such as onshore seismic, where the more traditional 2D techniques become very difficult or even unfeasible. Another critical point is that the method is based on post-stack surveys, thus dealing with 2D and 3D surveys.

Acknowledgements. The authors acknowledge the support from ANEEL (Agência Nacional de Energia Elétrica) through the project ”Evolução do Sistema computacional ALINE para detecção de acúmulos de gás destinado ao complexo termelétrico do Paranaíba, empregando dados de linhas sísmicas e algoritmos de Machine Learning” at Eneva in collaboration with Tecgraf Institute (PUC-Rio).

Authorship statement. This section is mandatory and should be positioned immediately before the References section. The text should be exactly as follows: The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] L. F. Santos, R. M. G. E. Silva, M. Gattass, and A. C. Silva. Direct hydrocarbon indicators based on long short-term memory neural network. In *SEG Technical Program Expanded Abstracts 2019*, pp. 2373–2377. Society of Exploration Geophysicists, 2019.

[2] L. Fernando Santos, M. Gattass, A. Correa Silva, F. Miranda, C. Siedschlag, and R. Ribeiro. Natural gas detection in onshore data using transfer learning from a lstm pre-trained with offshore data. In *SEG Technical Program Expanded Abstracts 2020*, pp. 1190–1195. Society of Exploration Geophysicists, 2020.

[3] J. M. and T. PUC-Rio. Documentação da plataforma soma. `https://soma-guide.readthedocs.io/en/latest/`, 2019.

[4] de F. S. Miranda, A. L. Vettorazzi, da P. R. Cruz Cunha, F. B. Aragão, D. Michelon, J. L. Caldeira, E. Porsche, C. Martins, R. B. Ribeiro, A. F. Vilela, and others. Atypical igneous-sedimentary petroleum systems of the parnaíba basin, brazil: seismic, well logs and cores. *Geological Society, London, Special Publications*, vol. 472, n. 1, pp. 341–360, 2018.