Data science applied to oil wells' behavior prediction in the Estructura Cruz de Piedra - Lunlunta oil field, Cuyana Basin, Argentina

Micaela Báez[1], Luis P. Stinco[2], Silvia P. Barredo[1] and Hernán D. Merlino[3]

[1]*Instituto del Gas y del Petróleo, Departamento de Energía, Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires, Argentina*

[2]*Departamento de Ciencias Geológicas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina*

[3]*Laboratorio de Sistemas de Información Avanzados, Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires, Argentina*

Abstract

The Cuyana Basin is one of the six Argentinian productive basins, recording historical productions of more than 210 million cubic meters of oil. 5% of that production comes from the Potrerillos formation, a continental triassic sequence that includes alluvial, fluvial, sub-deltaic and lacustrine facies. Porosity, permeability, saturation, thickness and depth data from this unit in different oil fields located on the north of Mendoza province was analyzed in order to evaluate if similarities between these features could relate to production performances. An exploratory data analysis was carried out among production data from over 130 oil wells, so as to recognize the productive wells behavior and develop models that are able to predict performance patterns according to the oil field, considering their geographic setting and the basin evolution.

Keywords

Geology, Cuyana Basin, Potrerillos Formation, Exploratory Data Analysis

1. Introduction

1.1. Theoretical framework

Hydrocarbons constitute the main energy sources in the Argentinian energy matrix [1]. These resources are found on sedimentary basins, lithospheric depressions generated by different geological mechanisms that enable sediment accumulation during long periods of time [2]. As a result, different strata are formed within the basins, relatively homogeneous rock layers with thicknesses over 1 cm. This homogeneity implies that the sedimentary components of the unit were deposited under constant physical conditions. Each stratum is characterized by a specific lithology and sedimentary structures that allow differentiation from the layers placed above

ICAIW 2021: Workshops at the Fourth International Conference on Applied Informatics 2021, October 28–30, 2021, Buenos Aires, Argentina
mibaez@fi.uba.ar (M. Báez); lstinco@fi.uba.ar (L. P. Stinco); sbarredo@fi.uba.ar (S. P. Barredo); hmerlino@fi.uba.ar (H. D. Merlino)
0000-0002-8775-3499 (M. Báez); 0000-0002-5157-6136 (L. P. Stinco); 0000-0002-8805-7778 (H. D. Merlino)

and below; they are separated by surfaces that represent sedimentation interruptions or erosive events.

The sedimentary filling will suffer modifications over time due to tectonic processes, sea level changes, erosive events or the deposition itself that increases pressure and temperature in the basin as it subsides along sediment deposition. One of the most affected properties due to the basin infilling (and the correspondent depth increase) is porosity. Porosity is a measurement of the empty spaces found surrounding the grain framework within a rock, and so it is closely related to its storage capacity. It is an expression of the percentage of fluids by volume compared to the total rock volume [3]. In an early stage, porosity is associated with the depositional setting. This primary porosity can be altered as a consequence of post-depositional changes linked to depth and overburden increase, resulting in secondary porosity [4]. Such porosity can be higher or lower than primary one, depending on multiple variables. Cement precipitation, particle filtration into pores and compaction result in the diminishing of pore space, whereas processes like dissolution, dolomitization and grain fracture enlarge it. Another interesting property to observe when analyzing a sedimentary basin infilling is permeability, the rock capacity to allow fluids flow through them [5]. This property depends on effective porosity, the interconnection between pores. Therefore, permeability is affected by grain size and shape, sorting, packing, cementing, compaction grade and clay mineralogy, being smectite and illite the ones that mainly modify sandstones permeability and porosity. Fluid circulation within the basin is also worth considering. Water may come from surface (meteoric water) and subsurface, as a sub-product from volatile loss in magmas during solidification, clay dehydration or mineral recrystallization, among other processes. This mineralized fluid alters chemical composition of rocks, due to grain dissolution and cements precipitations; therefore, also affects rocks porosity. Besides, it plays an essential role in hydrocarbon generation, for it accelerates organic matter decomposition.

In this framework, the petroleum system is formed. According to the definition published by Magoon and Dow (1994) [6], a petroleum system includes all essential elements and processes required for the existence of a petroleum and gas accumulation. The word ‘system’ describes the interdependence between the elements and processes involved in this hydrocarbon accumulation. On the other side, the ‘petroleum’ word refers to any of the following substances:

- Biogenic or thermal gas, located on conventional reservoirs as methane hydrates, tight reservoirs, fractured shales and coal.
- Condensate oil.
- Crude oil.
- Natural bitumen on clastic or carbonate reservoirs.

The essential elements of a petroleum system are the source, reservoir, seal and overburden rocks. Processes include trap formation and hydrocarbon generation, migration and accumulation. These elements and processes need to be correctly placed in time and space, so that organic matter contained on a source rock can be turned into an oil/gas accumulation.

The ‘source rock’ term defines a unit with considerable amounts of organic matter within its sediments, so oil and gas can be generated. They are sedimentary rocks, either clastic or carbonates, formed by less than 0.06 mm particles and with good porosity but no permeability, since pore space is reduced and does not allow fluid circulation.

Organic matter preservation for hydrocarbon generation depends on an anaerobic environment where there are no organisms or bacteria capable of destroying it. Besides, the sedimentation rate needs to be fast enough to guarantee its quick cover [7]. Deep lakes and oceans constitute classic environments where these conditions are often fulfilled, so their deposits may end up as excellent source rocks [8]. The organic matter accumulated at the bottom of these large bodies of water usually consists of phytoplankton, zooplankton, spores, pollen, exoskeleton and plant fragments.

Burial leads to physicochemical changes resulting in various products: kerogen, bitumen and hydrocarbons. Once hydrocarbons are generated, they are expelled into a reservoir through migration process. Hydrocarbons take advantage of rock discontinuities, such as fault planes or interstratal planes, to migrate through the basin deposits.

A reservoir rock is a rock of any lithological type that has the capacity to store hydrocarbons (porosity) and allow their production (permeability). In the case of conventional reservoirs, these are clastic or carbonate rocks whose particles have grain sizes greater than 0.06 mm, where the pores are interconnected, which allows the circulation of hydrocarbons and its subsequent extraction by mechanical pathways. Naturally fractured reservoirs are rock bodies of any lithological type where cracks provide the sufficient porosity and permeability to carry a removable accumulation of hydrocarbons [9].

The reservoir rock has the ability to accumulate hydrocarbons as long as it is covered by a seal rock, which prevents the hydrocarbons from continuing the migration toward less pressured areas. For this reason, the seal rock must be a laterally continuous body of impermeable rock, with sufficient capillary pressure to resist the rising pressure of the hydrocarbons, and with a certain degree of ductility (so that it does not fracture), such as evaporites or shales [10].

Both the reservoir rock and the seal must be contained in a trap, that is, a three-dimensional geometric configuration, either structural or stratigraphic, that allows the accumulation of the resource [11]. A structural type trap is formed by tectonic deformation during or after the accumulation of sediments in the basin. Generally, they are associated with folds or faults. In contrast, stratigraphic traps are formed by variations in sedimentation rates, regardless of structural deformation. The processes involved in the formation of this type of traps can be depositional, erosive or diagenetic.

Finally, above all these elements must be the overburden rocks, that is the overlying column of sediments that fills the basin. The function of these rocks is to provide the pressure and temperature necessary for the hydrocarbons to originate as the basin subsides (that is, as the basin deepens).

In unconventional shale reservoirs, on the other hand, the hydrocarbons generated are retained in the source rock, from where they cannot be mobilized due to the low permeability present in these units. Although they cannot be extracted using the classical methods for the conventional reservoirs' operations, technological advances in recent years allowed the implementation of more complex techniques that allowed unconventional reservoirs to be put into production. This is the case of hydraulic fracturing (fracking), the mechanism by which hydraulic fractures are induced in the source rock to create the necessary permeability pathways to extract the gas or oil contained inside. Shale gas/oil are fine-grained sedimentary rocks that have high organic matter content, so they can function as source and reservoir rocks, as well as seals. There, the hydrocarbons generated cannot migrate, but remain stored in the pores, in natural fractures and adsorbed on the organic matter. Any source rock in a conventional system is a potential shale-type reservoir. Another case of unconventional reservoir are Tight reservoirs. These are rocks of any lithology that present less than 0.1 mD of permeability. This feature is given by conditions inherent to the depositional fabric or due to diagenetic processes. There are also other types of unconventional reservoirs, such as coal bed methane, oil shale, heavy oils, tar sands and methane hydrates. Either case, the evolution of the basin predetermines the geographical conditions that must be dealt with in order to exploit the resource. The deposits are not always in easily accessible places, and it is also necessary to adapt to the regulations imposed by the local authorities. In recent years, it is essential as well to consider the environmental factor, taking into account how hydrocarbon production activities influence the environment and the populations surrounding the operation sites.

2. Background

Supervised and unsupervised automatic learning models [12, 13] are increasingly used in the oil and gas industry [14, 15] due to the impact of the contributions data science is making to the sector. Upstream industry has been particularly affected [14]. During the last decade, the Argentinian oil and gas industry increased artificial intelligence application in almost every phase of the hydrocarbon exploitation chain, from exploration to distribution. However, most of these techniques are implemented on the Neuquina Basin, for it bears the Vaca Muerta Formation. Software are being developed to optimize oil production predicting the most efficient water injection method [16] and the best fracture design [17], schedule revision dates before incidents, supervise interferences between wells [18] and diagnose troubles early on classifying the complications that may appear [17, 18]. Besides, real-time data is used to prevent upwellings, obstructions or any kind of issues during drilling activities, pre–detect screen outs during non-conventional reservoirs stimulation [17], and supervise wells operations and functioning [19]. Deep learning and video analytics processes are also useful to estimate petrophysical parameters from well logging data [17], generate recommendations for future wells terminations and production optimization [20, 21].

The combination of supervised and unsupervised algorithms has allowed predictions of, among other things, future well production rates [22]. Nevertheless, challenges arise regarding the necessary amounts of data for training the models and making further predictions, especially since public data is limited. Focusing in this particular issue, a data extrapolation model based on unsupervised learning and expert's judgment [23] is proposed.

3. Automated machine learning application for data analysis

3.1. Data gathering

The Potrerillos Formation is one of the productive units of the Cuyana Basin. It is a purely continental triassic sequence formed by alluvial, fluvial, sub-deltaic and lacustrine sediments

[24]. The best oil reservoirs of the formation are found in the sandstones and conglomerates belonging to low sinuosity fluvial channels, and are exploited in several oil fields in the north of the province of Mendoza, such as Barrancas, Cacheuta, Estructura Cruz de Piedra - Lunlunta, Chañares Herrados, Puesto Pozo Cercado, Piedras Coloradas, La Ventana, Mesa Verde and Lunlunta Carrizal [25].

Oil fields are places where one or more reservoirs are contained within the same underground trap; in the case of the Cuyana Basin, many of the oil fields produce hydrocarbons from more than one formation.

First, as much information as possible was sought about the basin and the formation. This data was extracted from Chapter IV of the Secretaría de Energía de la Nación Argentina [26]. The information found corresponds to the last 15 years of production, and include datasets with the monthly production of the entire basin, the monthly production of each oil field, the monthly production of each productive formation, the production of each formation in each oil field, the average daily production for each month of the entire basin and each oil field, and for some reservoirs we had data on estimated resources and reserves, both proven and proven as possible.

Regarding the Potrerillos Formation, data were found about the depth at which it is located in each oil field, the thickness of the reservoir, and the lithology, porosity, permeability and

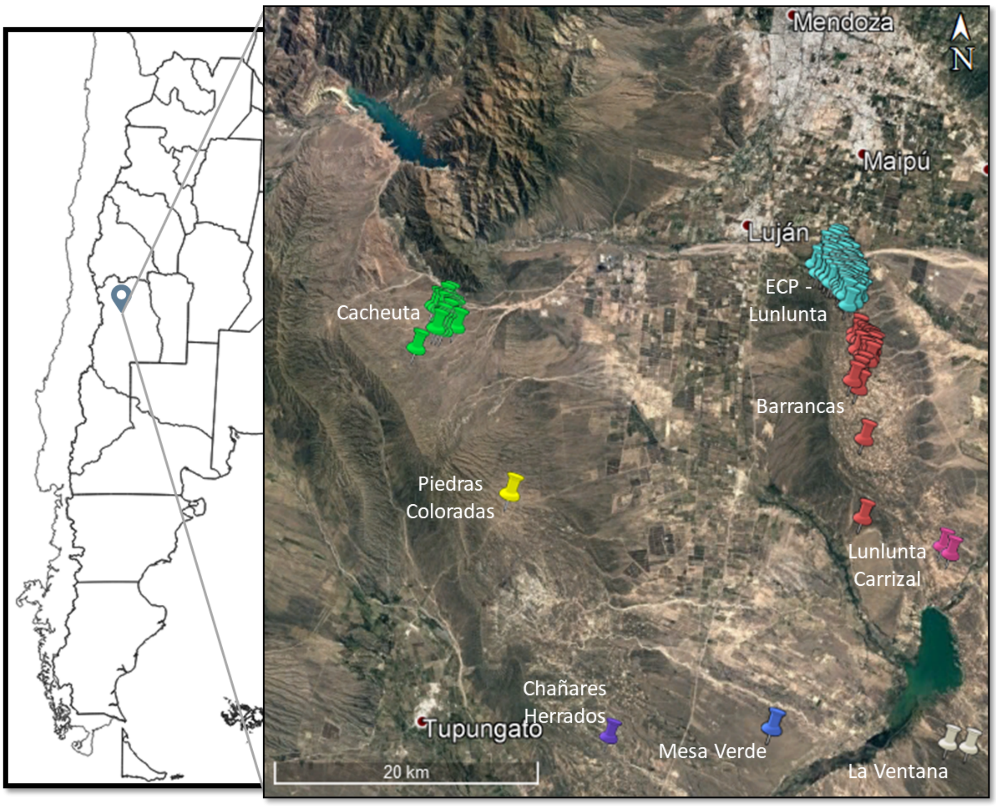

Figure 1: Well location by oil field

saturation. In addition, there are data on viscosity, density, pressure and salinity of the fluids [27]. There is also data from each well in the Cuyana basin that produced oil from this formation in the last 15 years, such as the monthly production of oil, gas and water, if any type of injection was necessary, effective extraction times and the type of extraction, the current state of the wells, the depth and their location [26].

The wells are located on the S-SE of the city of Mendoza (Figure 1). There are 130 wells drilled between 1959 and 2016, operated by different companies and with different extraction mechanisms. Besides, the productive formation is found at very different depths; the shallowest appearance of the formation is 60 m deep, while the deepest appearance is beneath 4500 m.

3.2. Data features

Two main datasets were used for this initial phase of this project. The first one includes geological and petrophysical data from the Potrerillos Formation in every oil field within the Cuyana Basin. These features are depth (m), thickness (m), porosity (%), permeability (mD) and saturation (%) of the reservoir in the following oil fields: Cacheuta, Puesto Pozo Cercado, Barrancas, Estructura Cruz de Piedra – Lunlunta, Chañares Herrados, Piedras Coloradas and La Ventana. Most of the values were obtained from specific academic publications [27, 28], and the missing data was completed using expert's judgement.

The second dataset contains the total oil production from the Potrerillos Formation by month in cubic meters, from January 2006 to April 2021 [26]. The 74 oil wells considered are currently operating in the Estructura Cruz de Piedra – Lunlunta oil field.

3.3. Exploratory data analysis

First, an exploratory data analysis was carried out on the data about the formation in each oil field, in order to find any patterns or relationships between properties using Pandas Python library.

All properties were taken to a logarithmic scale and plotted to recognize the configuration of the curves (Figure 2). It is clear that the depth distribution has no relation with the behavior of the other features, so at least for this instance of the work, it will not be considered as a significant factor. Another interesting thing that can be pointed out is that saturation values present small variations; even though there is a peak that coincides with the rest of the graph, the rest is constant, so for now it will not be considered either.

This primary analysis was useful to define that, in the first place, the relationship between thickness, porosity and permeability was going to be an object of analysis within this study. Plotting these curves separately (Figure 3) shows that thickness maintains a good relationship with permeability (although this can't be observed with the porosity curve), and that the porosity and permeability curves have a similar shape.

Therefore, thickness against porosity and permeability was plotted, and so porosity against permeability. Graphs against thickness did not return a good fitted relation, but the porosity vs permeability plot allows a relationship to be glimpsed (Figure 4). Since the data belongs to fluvial clastic sediment deposits, this relationship observed between porosity and permeability is geologically consistent, independently of the computational analysis made.

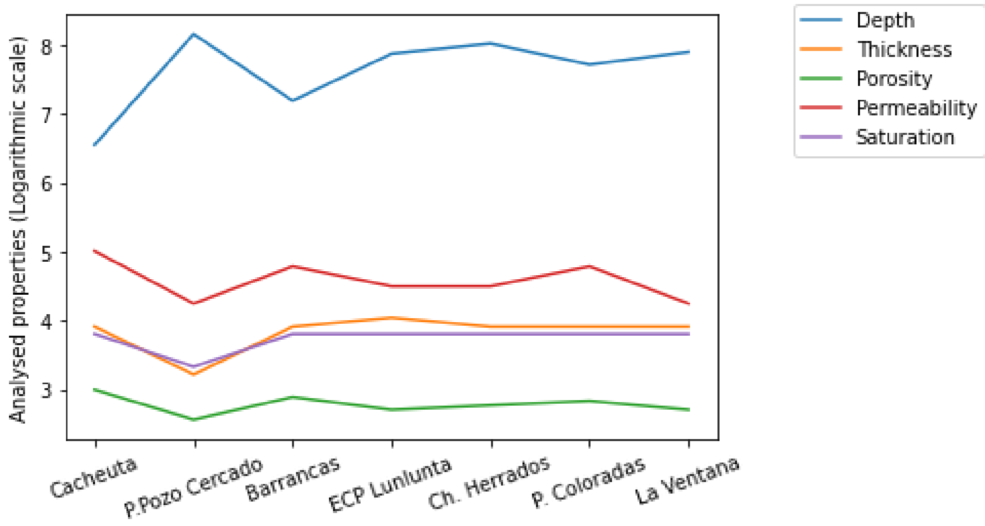

Figure 2: Oil fields features distribution

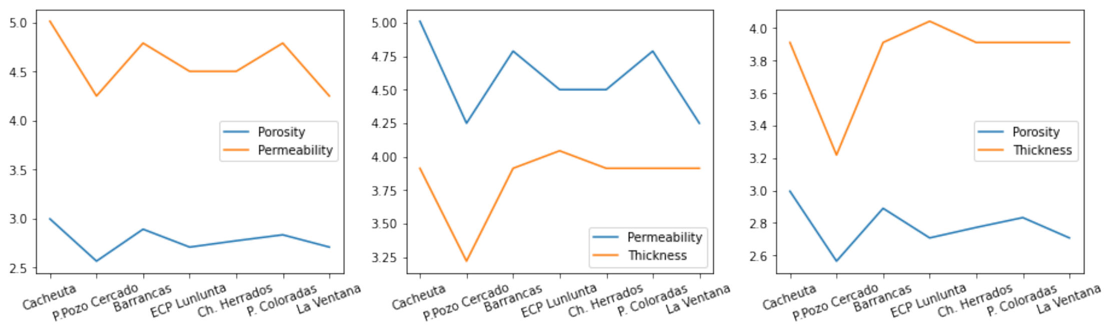

Figure 3: Comparison between selected features

With further studies, this result may make it possible to estimate permeability using porosity data and vice versa. Beyond the prediction, it will also permit filling in missing data based on a regression.

Also, radar graphs were made for each oil field (Figure 5). These radar charts are good visual methods for comparing sets of different features. Each oil field was compared to the Estructura Cruz de Piedra - Lunlunta oil field, the one with the most productive wells and more data available.

It is clear that the behavior of the oil fields, at least regarding these properties, is very similar. This observation raises the question that if the oil fields are similar, they might have similar production yields. Therefore, the analysis of the wells' behavior was conducted to see their performances.

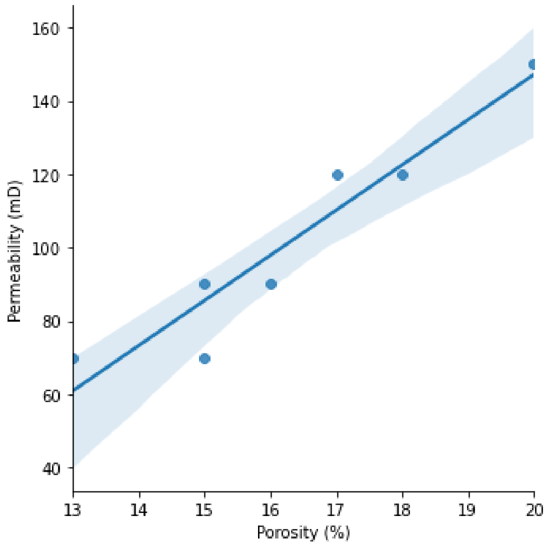

Figure 4: Porosity vs Permeability

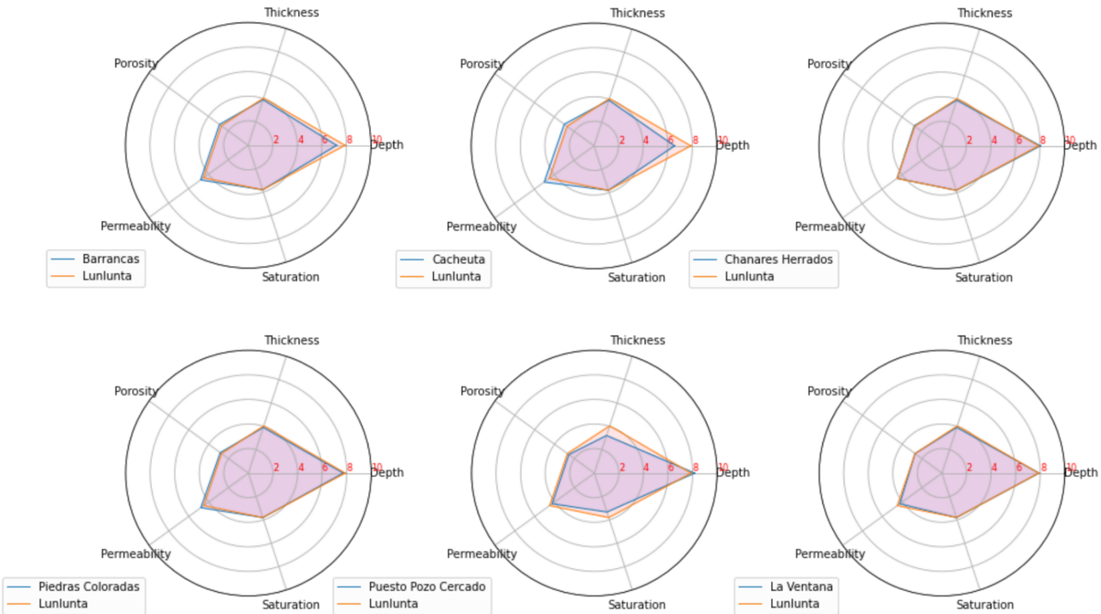

Figure 5: Radar charts comparing oil fields

3.4. Yacimiento Estructura Cruz de Piedra - Lunlunta

For now, until more information is available, this work is limited to the Estructura Cruz de Piedra - Lunlunta oil field because it presents the most complete datasets and also has most of the productive wells. Oil production from all the wells for the first 100 months of exploitation (Figure 6) was plotted in order to take a general look; the decline of the yield curve is evident.

That decline trend is also observed when plotting the average production of all the wells (Figure 7). Besides the general tendency, it is easy to find peaks of yield improvement; questions arise regarding why this variation of the curve is occurring. It could be due to a campaign to start production and completion of wells, drilling of new wells, or some secondary recovery projects. This is something to take into account on further analysis, it would be interesting to see if the other deposits present similar behaviors.

This decreasing trend is also observed in other graphs such as these box plots (Figure 8),

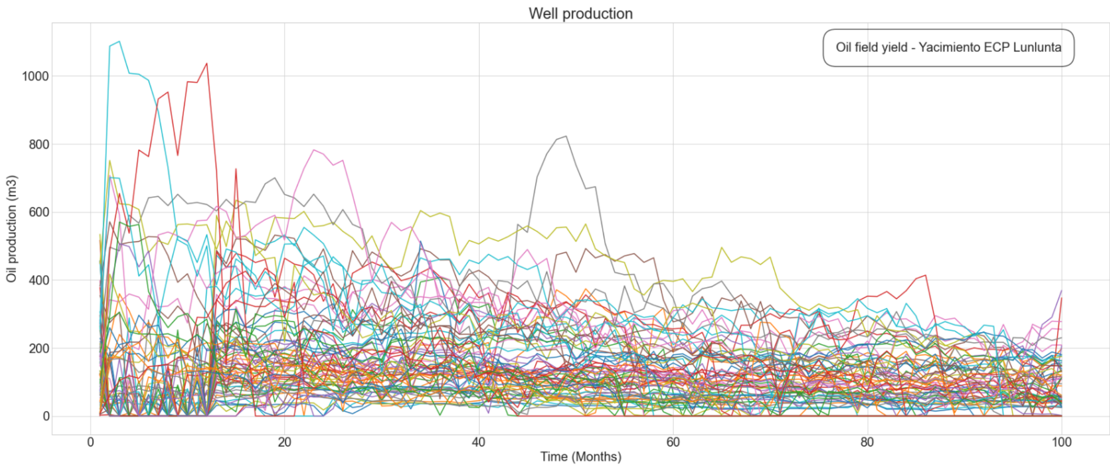

Figure 6: Oil production for the first 100 months

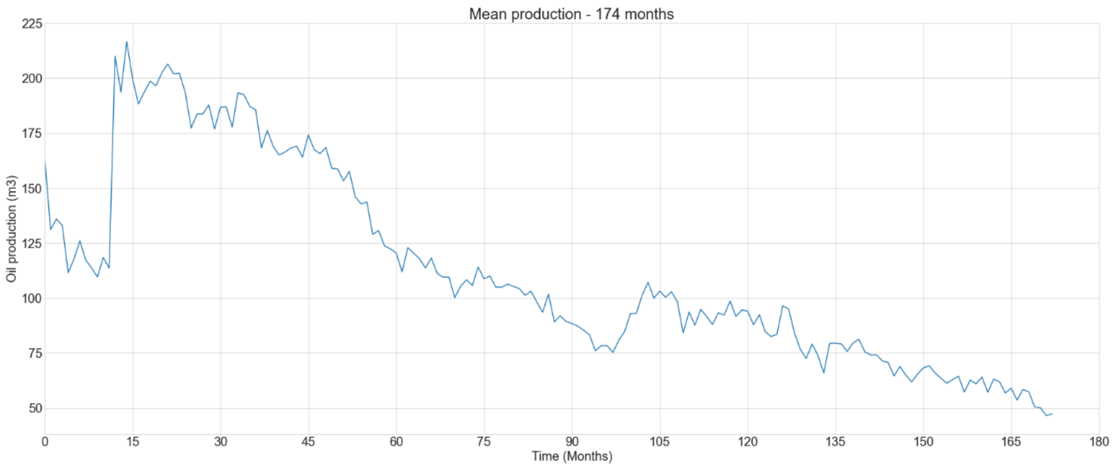

Figure 7: Mean production

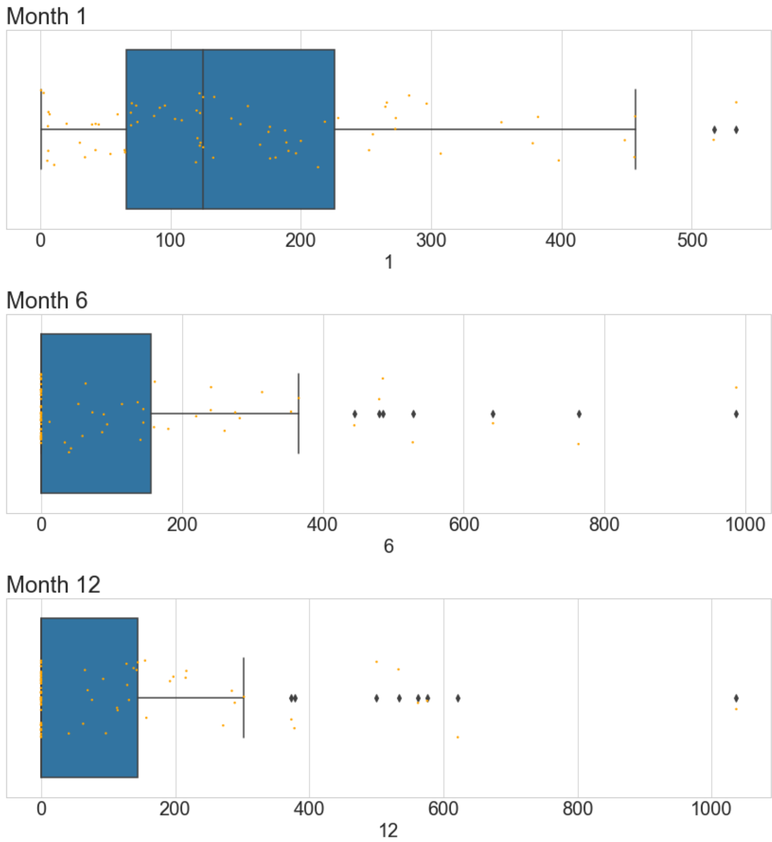

Figure 8: Box plots for production at months 1, 6 and 12

where production from each well for the first month, at six months and after one year was plotted. On the other hand, heat maps (Figure 9) were made, comparing the production of the first month against month number 2, as to see the distribution of the relationship between the first and second month of production. The same plots were made comparing the first month with months 12, 24, 36, 48 and 56. The distribution of the data is similar, the highest density is always in the same quadrant. This could indicate that the wells that had the best initial flow continue with the same disposition and are those that at least until month 56 maintain the best flow. Deeper analysis on the future should corroborate whether this behavior is maintained throughout the period of time analyzed.

3.4.1. Projection

An interesting addition to this part of the study was a projection for the subsequent 12 months of oil production based on the average yield curve over time for all the wells (Figure 10). This prediction was made using Prophet, an open-source library designed for automatic forecasting on Python.

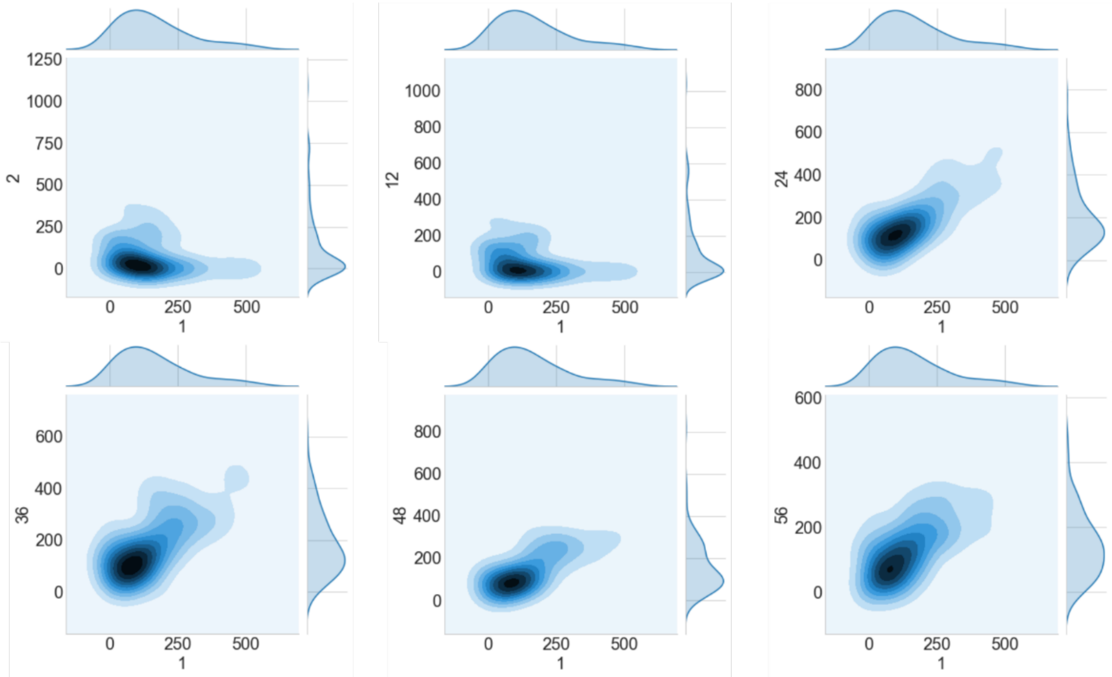

Figure 9: Heat maps comparing production rates

3.4.2. Unsupervised learning methods

The model proposed is based on wells clustering using the unsupervised learning algorithm K-Means from the Pycaret library in Python. Several models were created, but the K-Means algorithm presented the best Silhouette metric. 95% of the wells were used for training and

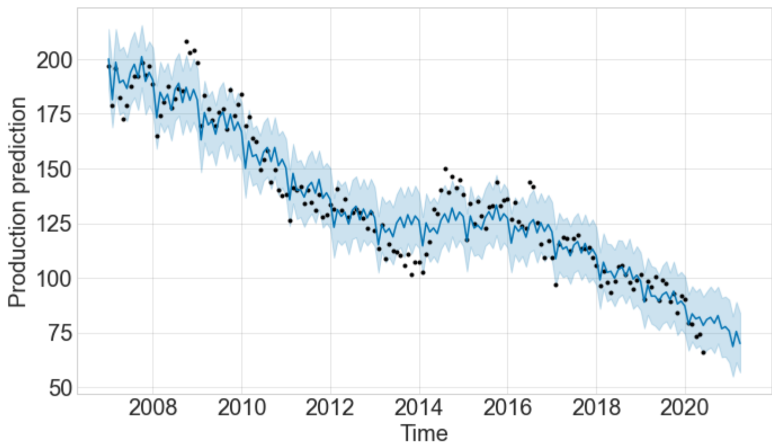

Figure 10: Projection based on mean production

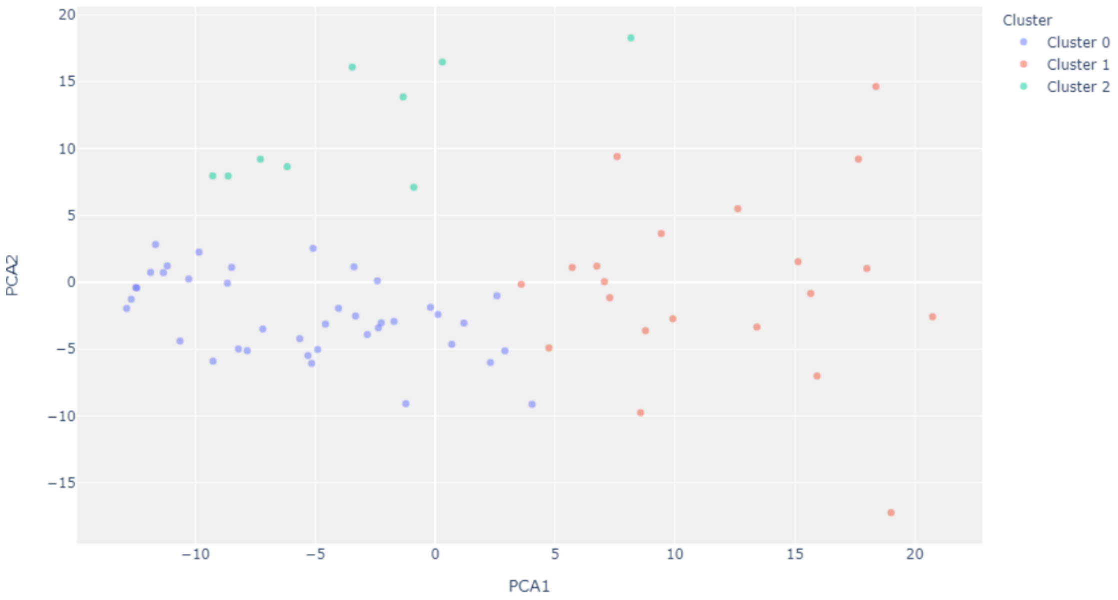

Figure 11: Wells distribution and classification by cluster

the other 5% remained as the testing data set. 3 clusters were found: cluster 0 with 40 wells, cluster 1 with 21 wells and cluster 2 with 9 wells. For a clearer understanding, the model was reduced to two dimensions (Figure 11). Another thing done in order to predict yield curves of oil production wells was an anomaly detection to identify those wells whose behavior does

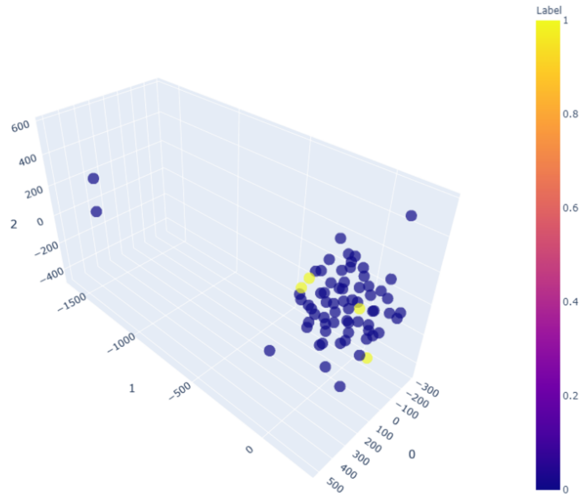

Figure 12: Anomaly detection plot. Notice the 4 wells detected as anomalies in yellow

not match the others (Figure 12). 4 wells were pointed out as anomalies, so further stages of the work will analyze the reasons why they were considered as such. The aim of the proposed process is to deal with oil fields as if they were black boxes only evaluated by their production, given the lack of public data for traditional analysis. This method has been inspired by artificial neuronal networks since they provide results but cannot necessarily explain them, although there is still a debate about this.

4. Preliminary Results

Even though this is a current work in progress, some interesting things could be observed.

- Significant relationships were found between porosity and permeability, which is geologically consistent.
- All the different oil producing fields present very similar features, which raises the question if their production curves are similar as well.
- The performance analysis in the Estructura Cruz de Piedra - Lunlunta oil field showed the typical production rate declination.
- A projection for the wells' performance was made using Prophet library on Python.
- K-Means algorithm from Pycaret library on Python was used to classify wells into 3 clusters according to their production rates.
- Anomaly detection using Pycaret library on Python found 4 wells that did not match the mean behavior.

These achievements demonstrate that the research is progressing with the expected results, and outline new goals for the study.

5. Future Work

In future analysis, the first thing to be done is to check whether the projection made matches the actual production rates for that considered time range.

Then, the oil production curves from this reservoir on the other oil fields should be analyzed and compared in order to corroborate if the fact that they share some specific features means similar production yields. Based on that, models could be generated for each oil field, to finally be able to characterize the petroleum system of the Cuyana Basin. Reinforcement learning models would be added for adjustment, once the models are elaborated.

Finally, those models shall be used to build an expert system that generates models that work as inputs in a neural network, from which possible general models for the basin could be obtained. These models would allow, knowing the location of the best-performing wells, to extrapolate all the considered features in order to propose locations for new wells with a certain degree of certainty.

6. Acknowledgements

Authors would like to thank the Instituto del Gas y del Petróleo from Universidad de Buenos Aires.

References

[1] Secretaría de Energía de la Nación Argentina, Balances energéticos nacionales, https://www.argentina.gob.ar/economia/energia/hidrocarburos/balances-energeticos, 2021.
[2] S. P. Barredo, E. Sharkov, et al., Geodynamic and tectonostratigrafic study of a continental rift: The triassic cuyana basin, argentina, Tectonics—Recent Advances 346 (2012) 99–130.
[3] K. Bjorlykke, Petroleum geoscience: From sedimentary environments to rock physics, Springer Science & Business Media, 2010.
[4] F. Pettijohn, Rocas sedimentarias, 1963.
[5] J. Serra, Introduction to mathematical morphology, Computer vision, graphics, and image processing 35 (1986) 283–305.
[6] L. B. Magoon, W. G. Dow, The petroleum system-from source to trap, AAPG Bulletin (American Association of Petroleum Geologists);(United States) 75 (1991).
[7] D. Welte, P. Tissot, Petroleum formation and occurrence, Springer, 1984.
[8] H. Klemme, G. F. Ulmishek, Effective petroleum source rocks of the world: stratigraphic distribution and controlling depositional factors, AAPG bulletin 75 (1991) 1809–1851.
[9] L. P. Stinco, Introducción al la caracterización de reservorios de hidrocarburos. Empleo de Técnicas de subsuelo en la evaluación de formaciones, Asociación Geológica Argentina, 2001.
[10] R. M. Sneider, Petrophysical properties of seals, 1991.
[11] R. R. Vincelette, E. A. Beaumont, N. H. Foster, Treatise of petroleum geology/handbook of petroleum geology: Exploring for oil and gas traps. chapter 2: Classification of exploration traps, 1999.
[12] P. Bangert, Machine Learning and Data Science in the Oil and Gas Industry: Best Practices, Tools, and Case Studies, Gulf Professional Publishing, 2021.
[13] E. Q. Wu, Z.-R. Tang, R. Hu, M. Zhang, G.-J. Li, L.-M. Zhu, G.-R. Zhou, Flight situation recognition under different weather conditions, IEEE Transactions on Aerospace and Electronic Systems 57 (2021) 1753–1767.
[14] S. Misra, H. Li, J. He, Machine learning for subsurface characterization, Gulf Professional Publishing, 2019.
[15] Y. N. Pandey, A. Rastogi, S. Kainkaryam, S. Bhattacharya, L. Saputelli, Reservoir engineering, in: Machine Learning in the Oil and Gas Industry, Springer, 2020, pp. 195–222.
[16] C. Calad, F. Gutiérrez, P. Pastor, Aplicación de data physics, combinación de física con inteligencia artificial (ia) en la optimización de los campos maduros, 2019.
[17] E. Ferrigno, L. Gonzalez, A. Romero, J. A. Alvarez Claramunt, D. Gallart, Experimentando con inteligencia artificial para optimizar el desarrollo de vaca muerta, 2019.
[18] A. Romero, C. Feldmann, K. Silva, J. Barros, J. Martinez, M. Montero, J. Alvarez, E. Ferrigno,

Modelo prescriptivo para la optimización de pozos no convencionales con sistema plunger lift, 2020.
[19] V. Girardi, B. Buteler, Monitoreo en tiempo real del ciclo de vida de pozos, 2020.
[20] D. Gallart, A. Lopez Gibson, Data analytics en tight gas, 2020.
[21] A. G. Mercader, B. Pogorelc, Video analytics en oil & gas: consideraciones prácticas, 2020.
[22] G. D. Rottoli, Resúmenes técnicos de validación mediante juicios de expertos de procesos de explotación de información espacialmente referenciada, https://www.researchgate.net/publication/346611919_Resumenes_tecnicos_de_validacion_mediante_juicios_de_expertos_de_procesos_de_explotacion_de_informacion_espacialmente_referenciada, 2020.
[23] P. C. Sen, M. Hajra, M. Ghosh, Supervised classification algorithms in machine learning: A survey and review, in: Emerging technology in modelling and graphics, Springer, 2020, pp. 99–111.
[24] D. Kokogian, F. Fernandez Seveso, A. Mosquera, Las secuencias sedimentarias triásicas. 65–78, in: Geologıa y Recursos Naturales de Mendoza. 12 Congreso Geológico Argentino y 2 Congreso de Exploración de Hidrocarburos (Mendoza, 1993), Relatorio, volume 1, 1993, pp. 65–78.
[25] J. H. Moratello, V. A. Ramos, Proceeding of the 12th congreso geológico argentino, 1993.
[26] Secretaría de Energía de la Nación Argentina, Producción de petróleo y gas por pozo (capítulo iv), http://datos.minem.gob.ar/dataset/produccion-de-petroleo-y-gas-por-pozo, 2021.
[27] D. Boggetti, J. Scolari, C. Regazzoni, Cuenca cuyana: Marco geologico y resena historica de la actividad petrolera, in: M. Schiuma, G. Hinterwimmer y G. Vergani (Eds.), Rocas Reservorio de las Cuencas Productivas de la Argentina, 5º Congreso de Exploración y Desarrollo de Hidrocarburos, Instituto Argentino del Petróleo y Gas, 2002, pp. 585–607.
[28] V. A. Ramos(Ed.), Geología y recursos naturales de mendoza, 1993.