UNIVERSIDADE FEDERAL FLUMINENSE

INSTITUTO DE GEOCIÊNCIAS

DEPARTAMENTO DE GEOLOGIA E GEOFÍSICA

PROGRAMA DE PÓS-GRADUAÇÃO EM DINÂMICAS DOS OCEANOS E DA TERRA

EBERTON RODRIGUES DE OLIVEIRA NETO

SEISMIC ATTRIBUTE ANALYSIS AND MACHINE LEARNING FOR FRACTURE PREDICTION OF A CARBONATE RESERVOIR FROM ALAGOAS STAGE, SANTOS BASIN PRESALT

Niterói, RJ

2023

EBERTON RODRIGUES DE OLIVEIRA NETO

SEISMIC ATTRIBUTE ANALYSIS AND MACHINE LEARNING FOR FRACTURE PREDICTION OF A CARBONATE RESERVOIR FROM ALAGOAS STAGE, SANTOS BASIN PRESALT

Thesis submitted to the Programa de Pós-Graduação em Dinâmica dos Oceanos e da Terra of Universidade Federal Fluminense in partial fulfillment of the requirements for the degree of Doctor in Science - Geology and Geophysics.

Advisor: Ph.D. Wagner Moreira Lupinacci

Niterói, RJ

2023

SEISMIC ATTRIBUTE ANALYSIS AND MACHINE LEARNING FOR FRACTURE PREDICTION OF A CARBONATE RESERVOIR FROM ALAGOAS STAGE, SANTOS BASIN PRESALT

EBERTON RODRIGUES DE OLIVEIRA NETO

Thesis submitted to the Programa de Pós-Graduação em Dinâmica dos Oceanos e da Terra of Universidade Federal Fluminense in partial fulfillment of the requirements for the degree of Doctor in Science - Geology and Geophysics.

Approved by the Committee on XXX

Committee:

_____

Ph.D. Antonio Fernando Menezes Freire (GIECAR/GGO/UFF and INCT-GP)

_____

Ph.D. Rodrigo Bijani Santos (GIECAR/GGO/UFF)

_____

Ph.D. Leonardo Guimaraes Miquelutti (GIECAR/GGO/UFF)

_____

Ph.D. Leonardo Costa de Oliveira (Petrobras)

_____

Ph.D. Elita Selmara de Abreu (Petrobras)

_____

Ph.D. Wagner Moreira Lupinacci – Advisor (GIECAR/GGO/UFF and INCT-GP)

Niterói, RJ
2023

FICHA CATALOGRÁFICA

Ficha catalográfica automática - SDC/BIG
Gerada com informações fornecidas pelo autor

O48s Oliveira Neto, Eberton Rodrigues de
SEISMIC ATTRIBUTE ANALYSIS AND MACHINE LEARNING FOR FRACTURE PREDICTION OF A CARBONATE RESERVOIR FROM ALAGOAS STAGE, SANTOS BASIN PRESALT / Eberton Rodrigues de Oliveira Neto. - 2023.
201 f.: il.

Orientador: Wagner Moreira Lupinacci.
Tese (doutorado)-Universidade Federal Fluminense, Instituto de Geociências, Niterói, 2023.

1. Aprendizado de Máquina. 2. Atributos Sísmicos. 3. Pré-sal. 4. Reservatórios Fraturados. 5. Produção intelectual. I. Lupinacci, Wagner Moreira, orientador. II. Universidade Federal Fluminense. Instituto de Geociências. III. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

To my mom – Para minha mãe

The dedication and acknowledgments will be in Portuguese. Not only because it is my mother tongue, but because some feelings can only be properly expressed through it.

For example, there is a word in Portuguese that is difficult to translate directly into English and other languages, it is called "saudade". For those who are reading this thesis and do not know Portuguese or the word "saudade", it means a feeling that you have when you miss someone or something. Therefore…

Mãe, esta tese aqui eu dedico totalmente a você. Não que outras pessoas não tenham contribuído para a sua conclusão, é só que ninguém merece isso mais do que a senhora. Por ocasiões da vida, você não poderá estar aqui fisicamente para poder acompanhar o fechamento desta etapa que você sonhou junto comigo, que me encorajou a encará-la, e sempre me suportou para que as tribulações da vida não passassem de pequenas preocupações. Eu não posso ser grato o suficiente para agradecer tudo que você sempre fez por mim; precisaria escrever infinitas teses, com infinitas dedicatórias, e ainda assim não seria suficiente. Talvez a melhor forma de lhe agradecer seja levando a vida da forma que a senhora sempre quis e me ensinou, com alegria e amor. Por isso, nesse momento, não vou lembrar da senhora de forma triste, e sim com felicidade, e muita saudade. Dizem por aí que mães são anjos da guarda que Deus envia para cuidar de nós. Imagino então que, ao alçar o plano espiritual, o seu cuidado, amor e carinho sejam potencializados. Também por esse motivo, não vou lamentar a sua ausência, mas sim me regozijar com a fé de que de algum lugar desse universo você esteja recebendo o meu amor e gratidão. Mas pode ficar feliz, você criou um filho doutor em geofísica, o seu maior orgulho, como você mesma dizia. Espero um dia poder seguir seus passos, e me tornar professor. Só não na mesma área. Herdei parte da sua inteligência, humor e simpatia, porém não com o mesmo traquejo e didática ao ponto de ensinar para adolescentes. Esse é um talento só seu e um desafio do qual não estou a par. Acho que você vai entender. Fica aqui, então, minha dedicatória e agradecimento à pessoa mais especial que eu tive o prazer de conhecer e conviver. Eternamente grato, eternamente seu Dengo, eternamente seu filhote. Com amor, Eberton.

Acknowledgements (Agradecimentos)

Agradeço a Deus por me conceder a oportunidade de terminar esta tese, por ter-me guardado durante esses quatro anos, especialmente nos dois anos de pandemia que o mundo viveu.

À minha família, representada aqui por Lais, Eduardo, Gabriela, Tia Marcinha e Eliane, por sempre torcerem por mim.

Agradeço ao meu orientador e amigo, Professor Wagner Lupinacci. Como eu já lhe disse uma vez, orientar não é só ajudar na realização do trabalho prático e na elaboração da parte teórica; vai muito além disso. E se não fosse por você entender isso, eu jamais teria terminado esta tese. Obrigado por todo apoio, paciência e amizade, e claro, por ter algum dia acreditado que aquele rapaz todo enrolado da graduação iria conseguir bancar um doutorado de cara.

Aos membros da banca de avaliação da minha defesa de doutorado. Por sorte a minha, antes mesmo da defesa, pude contar com importantes contribuições para a construção desta tese, como as colaborações do Dr. Professor Fernando Freire e os doutores Bruno Honório e Leonardo Texeira.

À Universidade Federal Fluminense, ou melhor, UFF, o lugar que aprendi a chamar de casa ao longo de onze anos e meio, primeiro no bacharelado e logo depois no doutorado. Este lugar especial de Niterói que vai ficar para sempre no meu coração. Espero que o destino me conceda a chance de um dia retornar como professor para devolver tudo que esta bela universidade me deu, contribuindo para a formação de novos geofísicos.

Agradeço também aos demais pesquisadores do grupo GIECAR, em especial colaboradores do projeto de pesquisa do qual fiz parte. Esta tese também é resultado da cooperação desse incrível time de geocientistas. Obrigado Maíra, Igor, Thuany, Mariana, Rafaela, Anderson, e Maria Luiza.

Não posso encerrar sem agradecer àqueles que mais torcem por mim. Aos que me conhecem de forma mais íntima e que me suportaram nestes quatro anos conturbados de doutorado, pandemia e outras tempestades. Amigos, ou como costumo dizer, família. Obrigado por tudo, vocês tornaram esta tese possível; sem vocês, não teria chegado até aqui.

Obrigado Lídia, Raquel, Thais e Julia, alguns dos meus presentes da UFF que tenho certeza, levarei para vida. Em especial ao primeiro ano de doutorado, em meio à pandemia. Vocês tornaram um dos anos mais caóticos e trágicos da história da sociedade moderna algo mais leve e suportável.

Obrigado Julia, Amanda e Lorena, por serem algumas das minhas maiores fãs. Minha torcida organizada particular, que torcem, acompanham e suportam, mesmo à distância. E claro, brigam e cobram quando necessário, como toda boa torcida organizada.

Obrigado Paola, Mariana, Victor, Dani e Caio, por serem uma verdadeira família para mim, sempre presentes nos momentos mais cruciais da minha vivência, bons ou ruins, não importa, vocês estiveram e estarão lá. Aqueles que ajudam a trazer sentido para essa experiência louca chamada vida.

Thadeu, Henrique, Elias, Clara e Paula, cada encontro nosso, sempre cômico, leve e familiar, foi crucial para que eu pudesse descansar, me revigorar e encarar a maratona do doutorado, obrigado.

Nilo, Fábio e Gabriela, os mais novos presentes da UFF. Sem vocês, eu não sei se teria concluído esta parte final da maratona. Obrigado.

À Taísa e Alexandre, por sempre me acolherem e por me darem um segundo lar. Um lugar de confraternização e paz, onde parte desta tese também foi escrita.

À Juliana, Lorena, Ricardo, e Sheila, obrigado por me trazerem para dentro da sua família, que também se tornou a minha.

Ana Luiza, que presente a sua amizade. A pessoa que sempre tem um sorriso, e um abraço para o seu amigo. Obrigado por todas as conversas, o sermões, os ouvidos, a torcida, e, óbvio, os barzinhos juntos.

Garreto, meu irmão de espírito, você sempre acreditou em mim e me apoiou. Sempre teve um ombro amigo quando precisei. Você foi uma das pessoas que mais acompanhou essa corrida. Muito obrigado.

À Jeniffer, pela parceria de vida durante este doutorado, pelo carinho, atenção, e por ouvir minhas ideias malucas e acreditar nelas. Você, mais do que ninguém, acompanhou de perto esta jornada e esse trabalho, sempre me apoiando e incentivando. Esta tese não seria a mesma sem você. Obrigado.

“Passion is what gets you through the hardest times that might otherwise make strong men weak, or make you give up.”

– *Neil deGrasse Tyson*

Resumo

Esta tese aborda a estimativa da intensidade de fraturas no pré-sal, uma tarefa crucial para a compreensão, desenvolvimento e segurança da perfuração do reservatório. Dada a dificuldade de confiar apenas em dados de poços para identificar fraturas, o estudo investiga a viabilidade de usar atributos sísmicos. O estudo é dividido em duas partes. A primeira avalia a correlação entre atributos geométricos (curvatura e coerência) e a intensidade da fratura, medida em perfis de imagens de poços. A curvatura k1 mostra uma forte correlação, mas não fornece uma visão completa do fraturamento de rochas, pois reflete apenas os níveis de paleo-tensão e deformação. Outros fatores, como a estratigrafia mecânica e as fases diagenéticas (por exemplo, precipitação de sílica) têm uma influênciam no desenvolvimento de fraturas em carbonatos. A segunda parte do estudo utiliza uma abordagem baseada em dados para prever a intensidade da fratura longe dos poços, incorporando estratigrafia mecânica, análise de curvatura e zonas de danos por falha. Em um estudo preliminar de viabilidade, dois modelos de aprendizado de máquina 1D, baseados em regressores de conjunto de árvores (Random Forest e XGBoost), são treinados. A avaliação de dois cenários possíveis para os parâmetros de entrada dos modelos mostra que um modelo com 3 parâmetros (curvatura k1, distância à falha e impedância P) tem desempenho comparável a um modelo de cinco parâmetros (curvatura k1, distância à falha, módulo de Young, conteúdo de sílica e Razão de Poisson). Em testes estatísticos de importância dos parâmetros, a curvatura k1 apresenta a maior importância. Além disso, o XGBoost supera a Random Forest em termos de generalização. Usando curvatura k1, distância à falha e impedância P invertida, um modelo XGBoost é treinado para prever a intensidade da fratura no volume sísmico. Na análise de teste cego, a intensidade da fratura estimada atinge uma correlação de Pearson de 0,77. A estimativa 3D revela vários graus de fraturamento na Formação Barra Velha, com facies de *build-up* silicificados correlacionados aos níveis mais altos de fraturamento. Este estudo representa uma exploração quantitativa pioneira de fraturas e atributos sísmicos no pré-sal. O estudo destaca a necessidade de integrar diversas fontes de informação nos fluxos de trabalho de caracterização de fraturas e ressalta o potencial do aprendizado de máquina para lidar com as relações complexas entre as variáveis utilizadas.

Keywords: Intensidade de Fratura, Pré-sal, Aprendizado de Máquina, Análise de Curvatura, Atributos Sísmicos

Abstract

This thesis delves into the crucial task of estimating fracture intensity in the presalt, vital for reservoir understanding, development, and drilling safety. While challenging to rely solely on well data for fracture identification, this study explores the viability of using multi-seismic attributes for this purpose. The study is set in two parts, the first evaluates the correlation between standard geometric attributes (curvature and coherence) and fracture intensity in borehole image logs. While k1 curvature shows a strong correlation, it alone doesn't provide a complete picture of rock fracturing, as it only reflects paleo stress and strain levels. Other factors, like mechanical stratigraphy and diagenetic phases (e.g., silica precipitation), significantly influence fracture development in carbonates. The second part of the study employs a data-driven approach to predict fracture intensity away from wells, incorporating mechanical stratigraphy, curvature analysis, and fault damage zones. Two 1D machine learning models, based on tree-ensemble regressors of Random Forest and XGBoost, are trained for feasibility reasons. Evaluation of two scenarios demonstrates that a three-feature model (k1 curvature, distance to fault, and upscaled P-Impedance) performs comparably to a five-feature model (k1 curvature, distance to fault, and upscaled logs of Young's modulus, silica content, and Poisson's Ratio). Notably, k1 curvature emerges as the most important feature, and XGBoost outperforms Random Forest in generalization. Utilizing k1 curvature, distance to fault, and inverted P-Impedance, an XGBoost model is trained to predict fracture intensity in the seismic volume. In blind-test analysis, the estimated fracture intensity achieves a Pearson's correlation of 0.77. The 3D estimation reveals varying degrees of fracturing in the Barra Velha Formation, with intense silicified build-ups correlating to the highest levels of fracturing. This study represents a pioneering quantitative exploration of fractures and seismic attributes in the presalt. It underscores the necessity of integrating diverse information sources in fracture characterization workflows and emphasizes the potential of machine learning in handling the complex relationships among involved features.

Keywords: Fracture intensity, Presalt, Machine learning, Curvature analysis, Seismic attributes

List of Figures

List of Tables

Summary

1. Introduction

My research aims to provide valuable insights into fracture characterization using seismic attributes in Presalt reservoirs. I aim to establish practical guidelines to enhance workflow efficiency, emphasizing the importance of precise attribute parameterization, highlighting the value of curvature compared to coherence in certain scenarios. I also stress that achieving accurate predictions for fractured zones requires considering mechanical stratigraphy and proximity to faults, extending beyond the scope of curvature alone. Ultimately, I aim to assess the potential of data-driven approaches and machine learning in generating consistent, high-resolution results for fracture intensity, surpassing traditional human-driven and model-based methods.

Structural seismic interpretation is an essential task for proper reservoir characterization. Reservoirs that face a geological scenario of intense tectonic activity, either syn- or post-depositional, can have a high intensity of faults and fractured zones.

Brazilian Presalt carbonates are considered one of those cases. These lacustrine carbonates are related to the Gondwana breakup process and were developed both in a syn- and post-rift context. Besides, evidence of several episodes of fault reactivation is present in both formations, as documented by Alves et al. (2017), Zhao et al. (2019), and Mendes et al. (2022).

The faults and fractures can significantly determine poro-perm conditions in the Brazilian Presalt (De Luca et al., 2017; Herlinger et al., 2017). For example, epithermal activity channelized by the faults could have induced the creation of some high-porous deposits, such as hydrothermal build-ups (Souza et al., 2018; Pietzsch et al., 2018). Fault slip can create elevated areas prone to sub-aerial exposure and susceptible to meteoric karsting (Hughes and Darby, 2021). The presence of fracture corridors and fracture networks can enhance the porosity system when opened (Wennberg et al., 2019; De Ros, 2021) and are often associated with excess permeability-related features (Fernandez-Ibánes et al., 2022a). On the other hand, if such discontinuities are associated with hydrothermal fluids, such as silica and calcite precipitation, they can negatively impact the porosity system (Lima and De Ros, 2019; Fatah et al., 2021a; De Ros, 2021; Jesus et al., 2021). Moreover, some works pointed out the fact that proper reservoir modelling in the presalt should indeed account for the porosity and permeability of the fault and

fracture network (Correa et al., 2021; Tanaka et al., 2022; Fernandez-Ibanez et al., 2022a; Fernandez-Ibanez et al., 2022b), especially in the carbonate mounds of the Barra Velha Formation. It is not rare to find excess permeability features in the Brazilian Presalt, with several wells producing above 40,000 boe daily (ANP, 2023), with permeabilities surpassing the thousands or even dozens of thousands of mD (Fernandez-Ibanez, 2022a; Fernandez-Ibanez et al., 2022b) (Figure 1).

The fault and fractured zones are also related to hazard zones for drilling operations. Fernandez-Ibanez et al. (2022a), analysed several drilling reports from the Santos and Campos Presalt, and observed that half (33/65) of the wells reported losses.

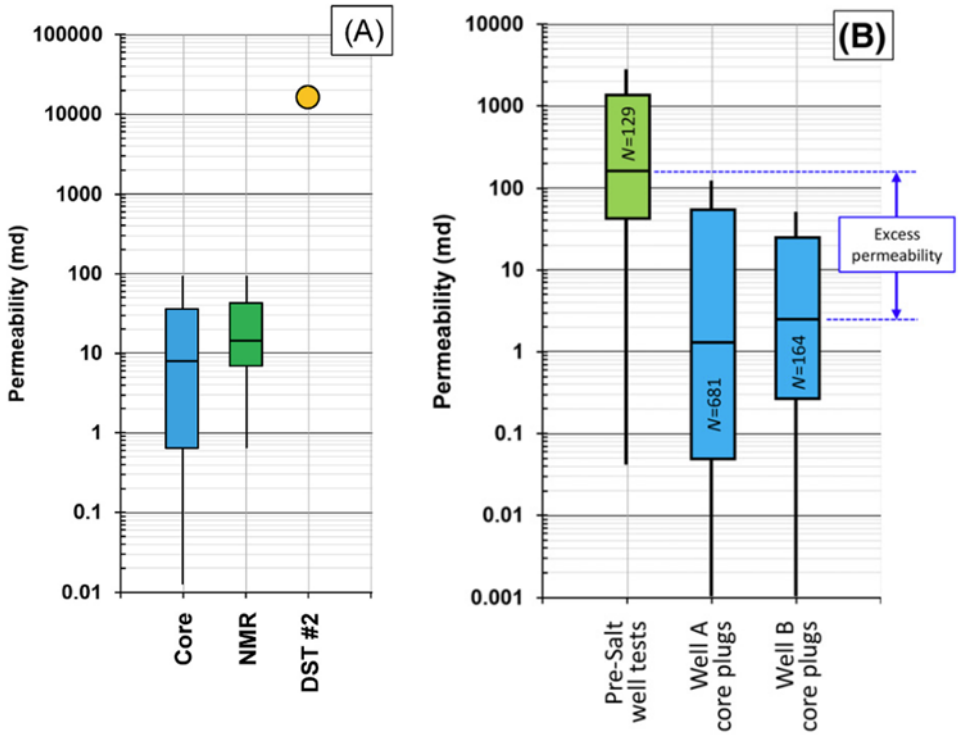

Figure 1: Box plot showing different permeability measurements acquired by Fernandez-Ibanez et al. (2022a; 2022b). The left panel (A) shows the measurements from a single well from Santos Basin. The right panel (B) shows the measurements from several Santos and Campos Presalt wells. Both cases highlighted the excess permeability, the anomalous permeabilities measured from well tests compared to routine core analysis.

To comprehensively characterize presalt fracture systems, relying solely on well logs and rock data proves inadequate due to their limited study scale dictated by the borehole diameter, typically around 12.25 inches (~31.12 cm) in the presalt interval. These data sources lack the spatial scale to fully grasp the complexity of the fracture network, as demonstrated by Fernandez-Ibanez et al. (2022a) (Figure 2).

Core measurements of porosity and permeability offer insights into local fluid transmission and storage capacity but fail to capture the broader spatial variability of surrounding geology. While well logs extend information vertically, they overlook lateral variations in rock properties. To address these limitations, integrating rock, well log, and seismic data is crucial for enhancing reservoir model reliability.

Seismic data plays a pivotal role by providing insights into both vertical and lateral heterogeneity and anisotropy. This comprehensive approach ensures a more accurate and robust reservoir model for presalt formations, incorporating high-resolution information from well logs and rock data alongside the broader perspectives offered by seismic data.

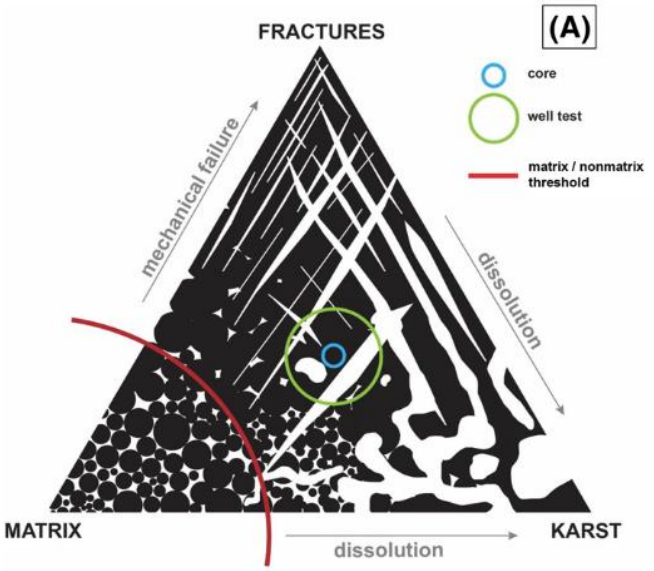

Figure 2: Ternary conceptual diagram illustrating the host rock (depicted in black) and porosity (in white) components of a dual-porosity carbonate pore system. The concentric circles (not to scale) symbolically depict the pore system permeability assessed through a routine core plug analysis (shown in blue) compared to a well test (depicted in green). From Fernandez-Ibanez et al. (2022a).

In light of these critical considerations, it becomes imperative to pinpoint faults and fractured areas for a thorough characterization and an optimal development of reservoirs. An established approach for fracture characterization typically, in the industry, involves the utilization of geostatistical methods, such as Sequential Gaussian Simulation and Discrete Fracture Networks, to either implicitly or explicitly simulate fractures (e.g., Correa et al. 2019; Tanaka et al. 2022). In both methodologies, seismic-scale resolvable faults serve as essential constraints for populating these fractured zones.

Nonetheless, the seismic data acquired from the Presalt region (as highlighted by Zambrini et al. in 2020) confronts significant imaging problems related to illumination issues; leading to inadequate imaging of faults in the data. Consequently, relying solely on faults identified by a seismic interpretation can introduce substantial bias into the results, potentially resulting in underestimating the fracture network's extent.

Furthermore, when determining the appropriate distance from a fault for populating fractures, a pivotal question arises: What constitutes proximity or distance? For instance, Fernandez-Ibanez et al. (2022b) conducted a comprehensive study of fault properties in the Presalt region and their impact on well production performance. They presented an example of a well that, despite not intersecting any seismic-scale fault, was close to a substantial fault zone (almost 1km). This proximity significantly influenced the flow behaviour during the drill-stem test, as illustrated in Figure 3, indicating the presence of non-matrix features related to excess permeability. Subsequently, through numerical modelling, they quantified the width of the damage zone within the fault system necessary to produce the observed pressure transient analysis curve, as depicted in Figure 4.

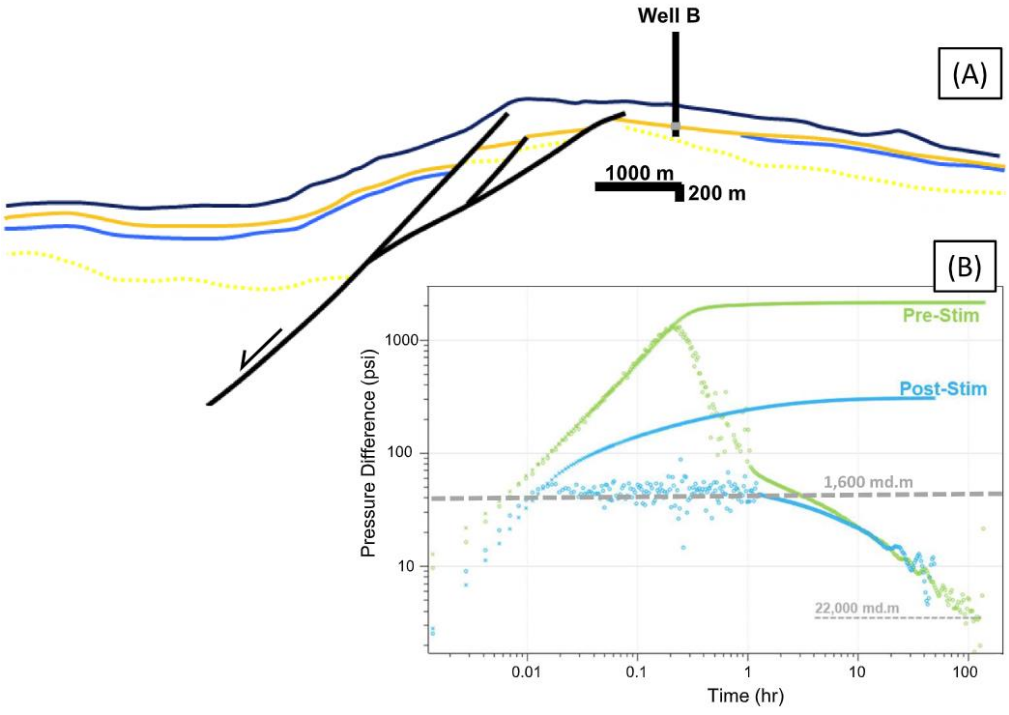

Figure 3: Interpreted cross-section of a faulted carbonate platform in the presalt, with well B located within 1000m of the main fault (A). One fault system controls the field's evolution, active during deposition and reactivated during burial. The well test interval is marked in grey. (B) The log-log plot shows pre- and post-acid stimulation pressure analysis. Skin reduces, and near-wellbore permeability improves post-stimulation.

Both derivatives maintain a consistent trend past 3 hours, indicating increasing transmissivity away from the wellbore, emphasizing higher permeability non-matrix features. From Fernandez-Ibanez et al. (2022b).

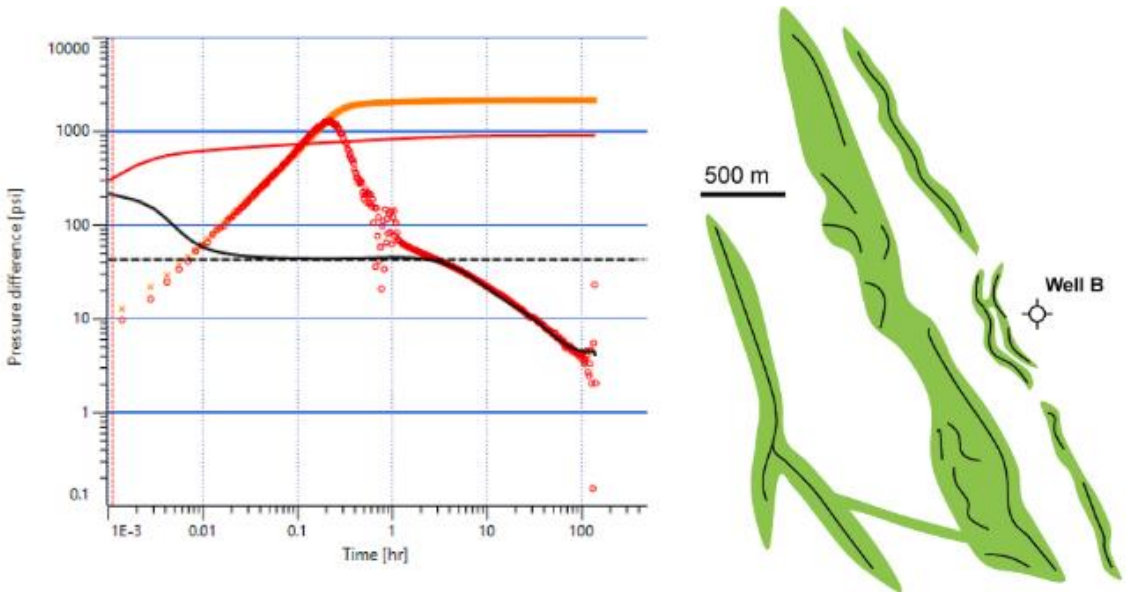

Figure 4: (A) Log-log plots compare the model and measured pressure transient analysis. The model aims to replicate the sustained trend beyond 3 hours. (B) The geologic concept in the model: black lines indicate interpreted fault traces from seismic data, and green regions highlight the necessary fault zone extent for matching the derivative trend. From Fernandez-Ibanez et al. (2022b).

Although identifying the fault location is important, this knowledge alone is insufficient. One should consider other factors, such as the amount of strain imposed by the deformation process, faulting, and folding. One way to overcome both the amount of strain and fault location problems is by using seismic attributes, especially the family of geometric seismic attributes. Seismic interpreters use seismic attributes routinely in interpretation workflows. Coherence, for example, is a traditional geometric attribute that highlights discontinuities in the seismic data associated with faults and other features.

One of the primary geometric seismic attributes used to characterize fractures in carbonates and other stiff rocks is dip-derived attributes (e.g., Astratti et al., 2015); among those stand out the curvature-based attributes. They do so for two main reasons: First, they are orientation-independent measurements (Roberts, 2001; Marfurt, 2018), remaining consistent regardless of surface rotation or tilting. This characteristic sets them apart from dip, edge, and azimuth attributes, making them versatile and applicable in various geological settings. Secondly, curvature-based attributes work with second-order derivatives, unlike other dip-derived attributes that utilize first-order derivatives. This characteristic enhances their ability to emphasize high-wavenumber features, making

them more effective in highlighting sub-seismic strain features. As a result, curvature attributes provide valuable insights into fine-scale structural complexities.

Particularly in the presalt case, where some geological intervals exhibit high-dip events, the "dip saturation effect" (Roberts, 2001) obscures the identification of sub-seismic faults and fractured zones in the first-order derivative dip-based attributes. Despite the advantages of curvature compared to other attributes, the lower frequency content of narrow-band Presalt seismic datasets is a concern for the curvature, but, as for any other geometric attribute. Seismic resolution is linked to seismic bandwidth. Specifically for small-scale features such as sub-seismic faults, fractured zones, and narrow fault, the stronger attenuation of the seismic energy for the high-frequency content hinders the identification of these features. Therefore, preconditioning of seismic data should be performed to enhance resolution and recover energy in the high-frequency portion.

Furthermore, geometric attributes are very noise-sensitive. For instance, noisy regions on the seismic data can be highlighted by the coherence attribute. Also, very low-amplitude seismic facies and noisy events pose challenges in dip estimation, resulting in unstable estimates of dip in those areas. In addition, curvature attributes, as second-derivative-based attributes, can raise noisy content present in the high-end portion of the wavenumber spectrum. Therefore, a less noise-sensitive technique is required for dip estimation since curvature attributes are derived from the dip field.

Recent works in the Presalt applied curvature attributes to infer fractured zones indirectly (Jesus et al., 2019; Penna et al., 2019) or as input to multi-attribute geostatistical and machine learning approaches to estimating fracture density/intensity (Correa et al., 2019; Babasafari et al., 2022). In none of these studies, a comprehensive qualitative and quantitative evaluation of those attributes is presented, and the correlation between the fractured zones and geometric attributes is simply presumed. Also, only Babasafari et al. (2022) address a preconditioning step in the workflow. They used a combination of median and diffusion filters, which attenuate noise while sharpening seismic discontinuities but do not improve vertical and lateral resolution. Furthermore, they used a seismic dataset in the time domain that can have severe problems regarding false structures induced by velocity pull-up and push-down.

The present work is divided into two parts: The first part aims to evaluate the reasoning behind using geometric seismic attributes, such as coherence and curvature, in the Presalt area of the Santos Basin to predict faults and fractured zones identified at the well scale. The second part evaluates a multi-scale (1D and 3D) data-driven workflow that combines seismic attributes and well logs (1D case) to predict fracture intensity in the wells and away from them.

In both parts I use inverse-Q filtering to improve the resolution of the seismic data and, consequently, the geometric attributes. In the first part of the curvature attribute evaluation, I assess three approaches for dip estimation and two post-filtering techniques. Then, through correlation with the upscaled fracture intensity logs, I analyse which parameterization results in the best attribute for fracture characterization. After, two regression models using the curvature attribute to estimate the fracture intensity property are proposed. Finally, I perform a comprehensive study integrating core, well-logs, and seismic attributes to address the causality of fractures at the Barra Velha Formation.

In the second phase of my study, I build upon the insights gained in the initial part to select relevant features for training a machine learning model to estimate fracture intensity within the seismic volume. This part of the work involves conducting a feasibility study where I combine well logs and seismic attributes to predict fracture intensity in the wells. After understanding the key features influencing the predictions by analysing these results, I proceed to train the final model for predicting fracture intensity across the entire seismic volume.

In this thesis, before presenting the data and methods, results, discussions, and conclusions, I will introduce four foundational chapters essential to the development of this work. The first chapter (Chapter 2) explores Santos' Presalt, followed by Chapter 3, focusing on Fractured Reservoirs and Carbonate Brittleness; Chapter 4 delves into Geometric Seismic Attributes, and Chapter 5 addresses Machine Learning. Within each of these chapters, I will comprehensively review key aspects of their respective subjects, aiming to provide the reader with a deeper understanding of the multidisciplinary nature of this study.

2. Santos' Presalt

This chapter review some of the geological aspects of the Presalt from the Santos Basin, highlighting the topics that are inherent to this study, such as tectonic history, fault style, facies association, and diagenesis.

2.1. Santos Basin

Santos Basin, the largest offshore basin in Brazil, stretches from the southern coast of the State of Rio de Janeiro to the northern reaches of the State of Santa Catarina. Situated between the parallels 230 and 280 of south latitude, it lies in the southeastern expanse of the Brazilian continental margin. Bounded to the north by the Cabo Frio high (which demarcates the limit with Campos Basin) and to the south by the Florianopolis high (separating it from the Pelotas Basin), the Santos Basin encompasses an area of approximately 350,000 square kilometres, with bathymetric depths reaching up to 3,000 meters. This expanse is nearly three times the size of the Campos Basin. Refer to Figure 5 for the precise location of the Santos Basin.

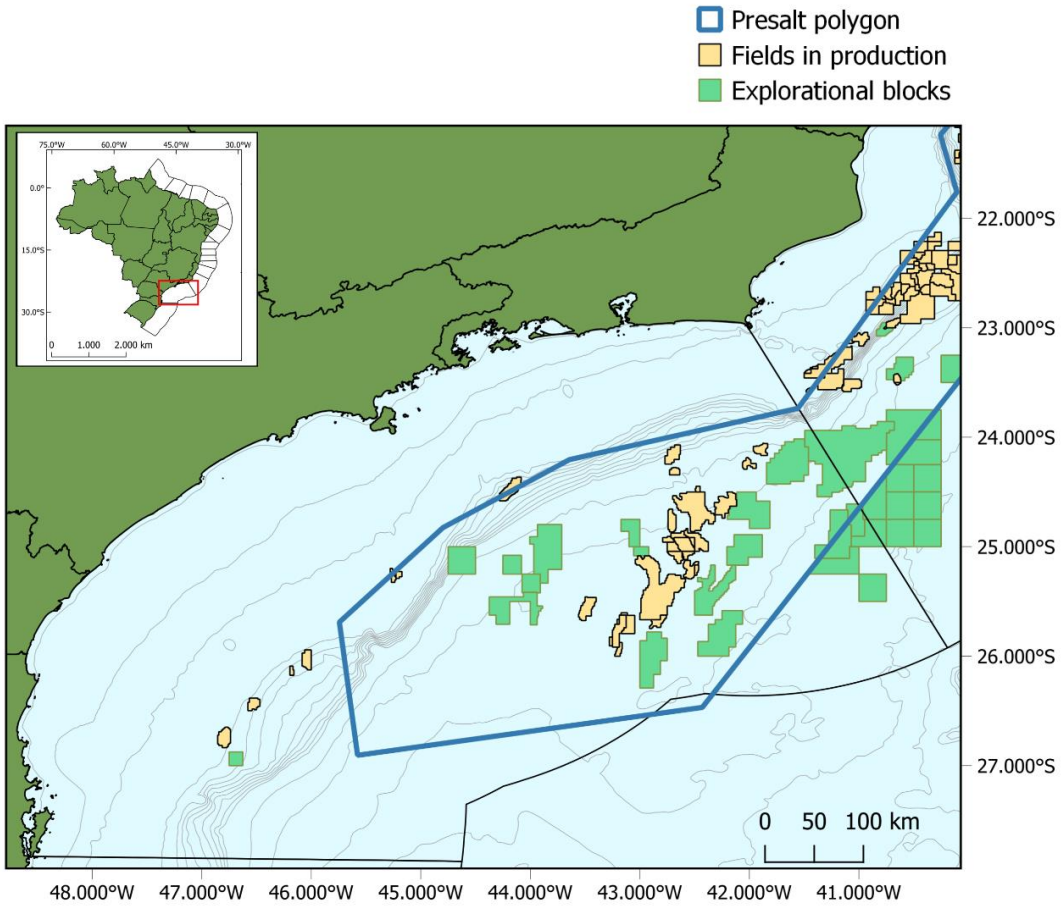

Figure 5: Cartographic map of Santos Basin with the limits of the Basin, the bathymetry, and the location of the presalt polygon, fields in production and explorational blocks.

2.2. Presalt Economic Importance

In 2006, the first exploratory well was drilled in the presalt section of the Santos Basin, targeting the evaluation of the Parati Prospect. While this well yielded only gas and condensate that was commercially unattractive, it substantiated the presence of an active oil system in that sector of the basin. This discovery hinted at a new exploratory opportunity in the Santos Basin: the presalt play (Petersohn, 2013). Simultaneously, in 2006, the Tupi field was unveiled in the deep-water section. This prospect revealed an oil accumulation within Neo-Aptian carbonate reservoirs, ultimately marking the inaugural commercial breakthrough in the presalt deposits of the Santos Basin. In 2010, the commercial viability of the Tupi Field was confirmed. In October 2013, the first bid for the sharing production regime of the presalt occurred, and the Libra Block was bidden. Four years later, the winning consortium declared the commerciality of the NW portion of the Block, knowing as Mero Field.

By July of 2023, presalt reservoirs has generated 75% of Brazil's total oil and gas. The Santos Basin corresponded to 75% of this production. In fact, in the same period, well 3-BRSA-1322-RJS, in the Mero Field, was the most productive in the country with a production of 69,591 boe/d. Mero, Búzios, and Tupi fields, the top 3 most productive fields in presalt, represent 67% of the whole presalt production (Figure 6).

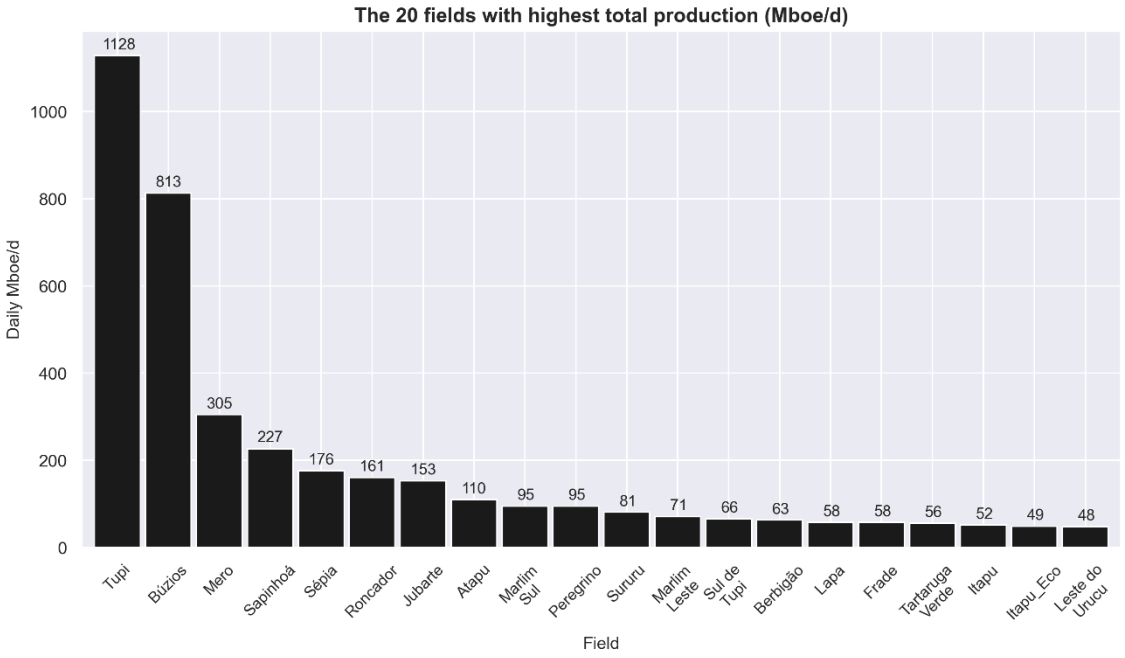

Figure 6: Barplot from Brazil's top twenty most productive hydrocarbon fields for July 2023. Source: ANP, Oil and Gas Monthly Bulletin of Oil and Gas Production (2023).

Carbonates are intricate rocks associated with reservoirs that pose difficulties in characterization, production, and management. Their complexity arises from a combination of depositional and diagenetic processes they undergo (Dunham, 1962). The diverse nature of carbonate facies, with their intricate textures and depositional attributes, leads to considerably heterogeneous reservoir properties, both vertically and laterally (Mohriak et al., 2015). Their substantial challenges in exploration counterbalance the economic significance of presalt carbonate reservoirs.

Within the Santos Basin, the presalt section comprises coquinas from the Itapema Formation and calcareous deposits from the Barra Velha Formation, often called "microbial." Initially, coquinas were viewed as secondary carbonate reservoirs (Formigli, 2008). However, the substantial accumulations discovered in the Mero and Buzios fields elevated the importance of these rocks in the presalt region of the Santos Basin (Kattah, 2015).

Presalt rocks in the Santos Basin are still inadequately understood regarding reservoir characterization. Consequently, more studies are required to better understand the specificities inherent in these deposits, which can exhibit significant variation from one field to another.

Key tools for improving the knowledge about these reservoirs encompass seismic quantitative interpretation and formation evaluation utilizing well logs and rock data. Integrating these tools with diverse data gathered from core samples, drill cuttings, sidewall samples, and laboratory measurements contributes substantially to understanding and characterizing these reservoirs.

2.3. Geological Context

In this topic, I go deeper into the geological characterization of Santos' Presalt. I introduce the lithostratigraphic units, discuss some of the tectonic development of the area, and the genesis of the main geological formations that encompass the reservoirs. I also discuss the complex diagenetic events that these formations went through and how they affect their reservoir properties.

2.3.1. Lithostratigraphy

The lithostratigraphy of the Santos Basin was initially established during the 1970s (Moreira et al., 2007). Given the extensive well and rock data available, Moreira et al. (2007) refined the chronolitostratigraphic framework model that Pereira and Feijó (1994) originally proposed. Their focus was on delineating depositional sequences, and they divided the sedimentary record of the basin into three supersequences: rift, post-rift, and drift. Within this framework, they elevated the Guaratiba Formation (previously designated as a formation by Pereira and Feijó, 1994) to a group category encompassing five formations. Three of these formations belonged to the rift phase (Camboriú, Piçarras, and Itapema formations). The other two were associated with the post-rift phase (Barra Velha and Ariri formations). Consequently, the stratigraphic chart presented by Moreira et al. (2007) serves as the primary reference for describing the stratigraphic composition of the Santos Basin.

Regarding the rift mechanisms, Chang et al. (2008) proposed a subdivision into three rift phases (sin-rift I, sin-rift II, and sin-rift III or Sag), akin to the analysis offered by Karner and Driscoll (1999) for the marginal basins of West Africa. These phases can be correlated with the proposal made by Moreira et al. (2007): sin-rift I and sin-rift II phases align with the rift phase, while sin-rift III or Sag corresponds to the post-rift phase.

In contrast to the positions taken by Moreira et al. (2007) and Chang et al. (2008), Wright and Barnett (2015) alleged that the lower portion of the Barra Velha Formation

should be categorized as part of the rift phase. Accordingly, Wright and Barnett (2015) introduced their own division based on the intra-Alagoas unconformity, distinguishing "microbial" calcareous rocks of the rift phase (situated between the pre-Alagoas and intra-Alagoas unconformities) and "microbial" calcareous rocks of the Sag phase (located above the intra-Alagoas unconformity and extending to the salt base unconformity boundary).

The authors argued that it is possible to discern the transition from the rift phase to a more geologically quiescent phase in seismic cross-sections, accentuating the presence of a Sag phase. They additionally noted that most extensional faults terminated at the base of the Sag phase, which could be correlated with the intra-Alagoas unconformity. Furthermore, they identified two distinct phases of rifting separated by the pre-Alagoas unconformity: a lower phase affecting the Piçarras and Itapema formations and an upper phase affecting the carbonates of the Barra Velha Formation. Consequently, the upper rift section was linked to the carbonate deposition of the Barra Velha Formation, extending up to the intra-Alagoas unconformity, beyond which the sag phase would commence.

The proposed division of the Barra Velha Formation into upper rift and Sag phases, a classification also endorsed by other authors like Pietzsch et al. (2018) and Neves et al. (2019), currently appears to provide a more accurate reflection of the tectonic processes impacting the presalt formations within the Santos Basin. Consequently, this study will adopt the following distribution of tectonic evolution phases for the Santos Basin:

- Lower Rift: Comprising the Camboriú, Piçarras, and Itapema formations.
- Upper Rift: Encompassing the lower portion of Barra Velha Formation up to the intra-Alagoas unconformity.
- Sag: Including the upper portion of the Barra Velha Formation, from the intra-Alagoas unconformity to the salt base, extending to the upper boundary of the Ariri Formation.
- Drift/Passive Margin: Spanning from the top of the Ariri Formation to the present day.

Figure 7 visually depicts the various proposals for tectonic phases and their correspondences with the presalt formations in the Santos Basin.

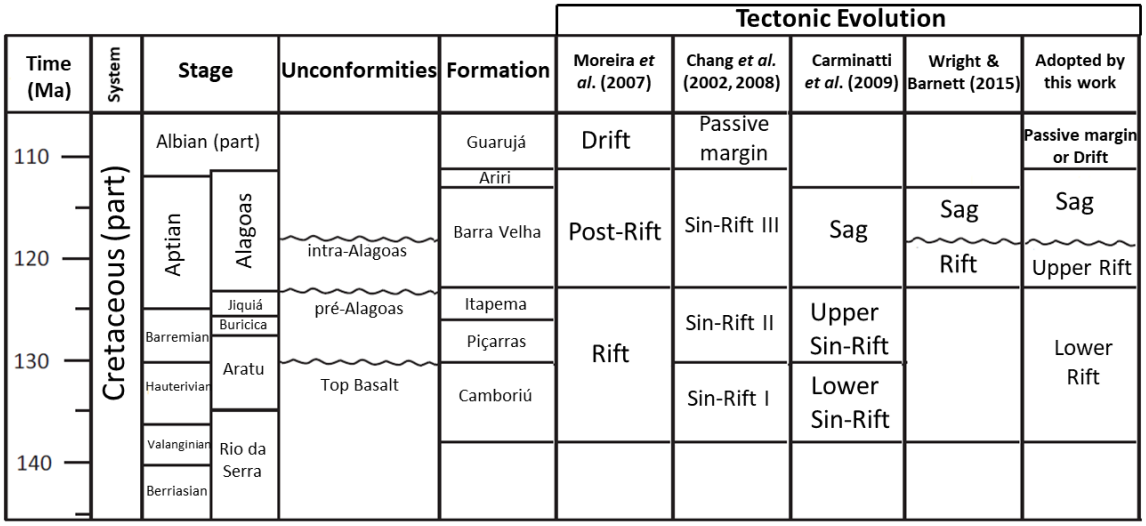

Figure 7: The chronostratigraphic chart correlating the different propositions of tectonic evolution for the Santos Basin (From Castro 2019, adapted from Wright and Barnet, 2015).

2.3.2. Tectonostratigraphic evolution

The Santos Basin's origin is closely linked to the breakup of the Gondwana supercontinent, leading to the formation of the South Atlantic Ocean. The basin's crystalline basement, exposed in São Paulo, comprises pre-Cambrian granites, gneisses, and metasediments (Moreira et al., 2007). The primary weakness zone for the eventual Atlantic Ocean development followed the Ribeira belt's trajectory (Almeida et al., 2013). In the southern region of Western Gondwana, significant basaltic magma outpouring preceded the rifting process. This gave rise to the South Atlantic's marginal basins and the Large Igneous Province (Rigoti, 2015). The initial Gondwana ruptures occurred over a triple junction associated with the Tristan da Cunha hotspot (Standlee et al., 1992).

The sequence of tectonic events leading to Gondwana's fragmentation began with the stretching and thinning of the continental crust, initiating a rifting process. This was influenced by lithospheric stretching (McKenzie, 1978) and thermal anomalies related to the Tristan da Cunha plume (Royden and Keen, 1980). The São Paulo Ridge, primarily composed of stretched continental crust, is a distinctive feature in the Santos Basin (Figure 8). It was formed during the rift process and has a complex configuration in deeper waters (Rigoti, 2015).

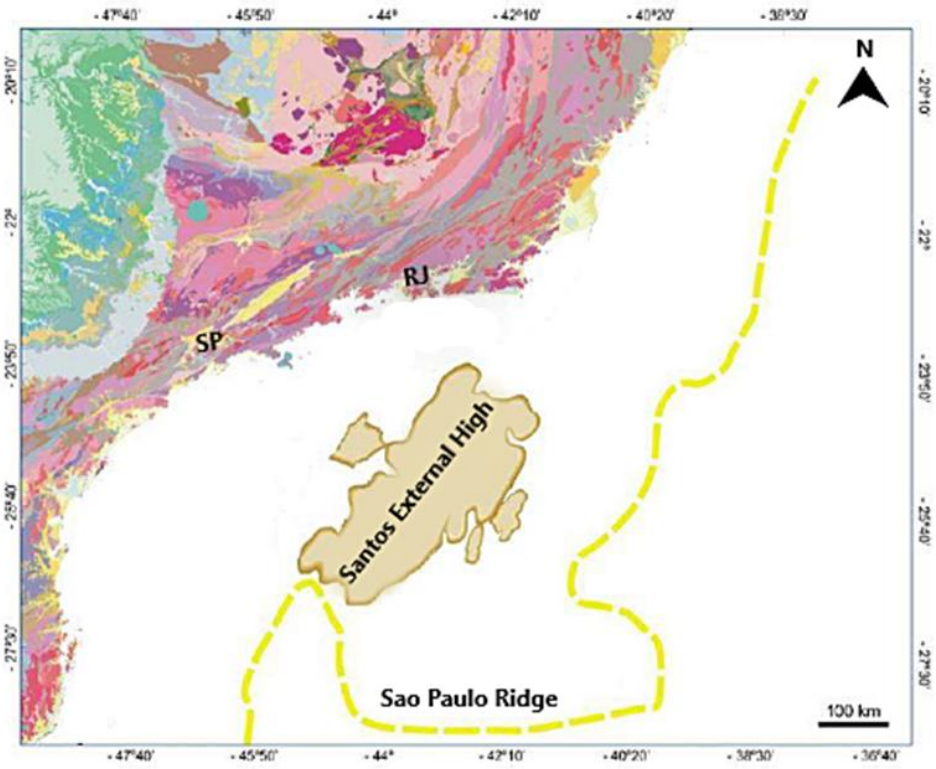

Figure 8: Location map of São Paulo Ridge and Santos External High. Source: adapted from Carminatti et al. (2008).

Around 145 million years ago, the South Atlantic rift began opening along its entire length, with the rate decreasing towards the rotation pole (Szatmari and Milani, 2016). Between 145 and 113 million years ago, the rift along the Equatorial Margin remained closed but expanded southward along the eastern margin of South America. This rotation might have been driven by mantle flow beneath the rift, including processes like mantle plumes. The South Atlantic Magmatic Province (SAMP) formed during this period, differentiating it from other Large Igneous Provinces (Szatmari and Milani, 2016).

The basaltic volcanism of SAMP began as the rift advanced, allowing the deepest faults to reach the hot lithospheric mantle, causing widespread melting. The Camboriú Formation, consisting of Lower Cretaceous basaltic flows, serves as the economic basement of the Santos Basin (Moreira et al., 2007). The Santos External High represents an uplifted structure associated with transitioning from volcanoclastic to carbonate formations (Buckley et al., 2015). The terrigenous sediments of the Piçarras Formation were deposited over the basaltic rocks of the Camboriú Formation. The Itapema Formation comprises bivalve grainstones, bioclastic wackestones and packstones, carbonate mudstones, and organic-rich dark shales (Moreira et al., 2007). The Barremian

sequence, including the Piçarras and Itapema formations, was deposited during the primary deformation period of the rift phase (ANP/LEBAC, 2003). The Proto-Walvis Ridge acted as a barrier between the rift and the opening of the ocean to the south, leading to the formation of an endoreic rift basin (Szatmari and Milani, 2016). The Rio Grande Rise - Walvis Ridge (Figure 9) impeded the free movement of marine waters, which, combined with high evaporation rates due to the dry and hot climate during this period, resulted in the deposition of a thick evaporitic sequence over the Barra Velha Formation (Karner and Gamboa, 2007).

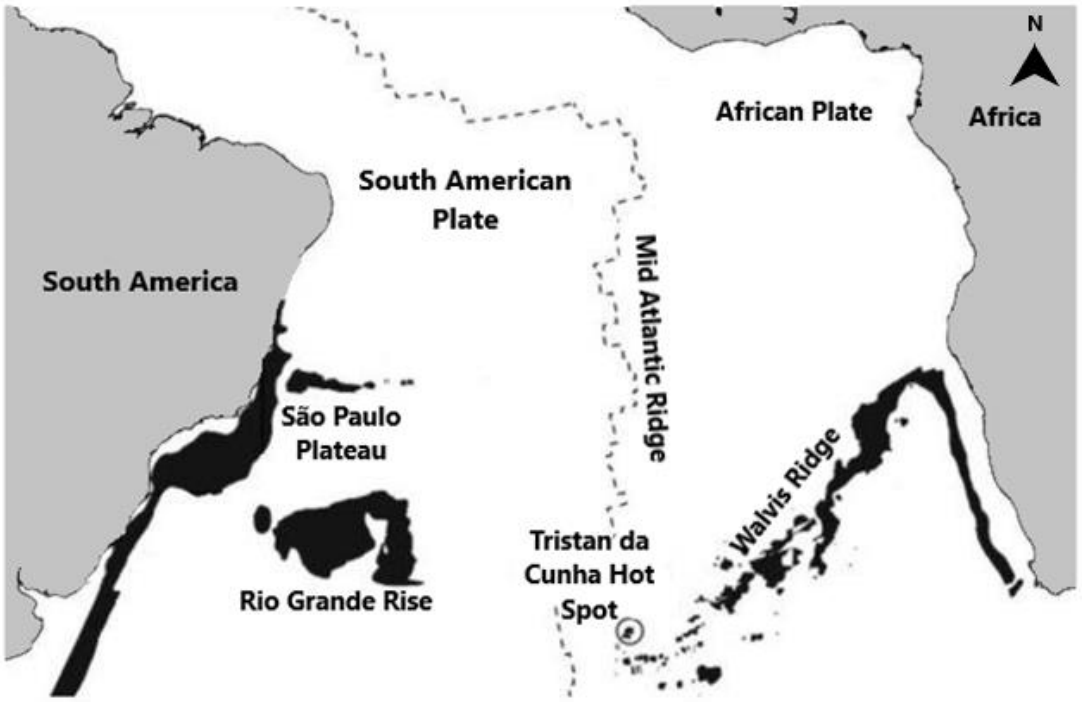

Figure 9: Schematic map showing the current configuration of the main structural elements related to the South Atlantic's opening: Rio-Grande Rise, Walvis Ridge, and Tristan da Cunha Plume (adapted from Schimmel et al., 2003).

The evaporitic sequence corresponds to the Neoaptian (Upper Alagoas stage) lithostratigraphic unit known as the Ariri Formation. This formation primarily consists of thick halite and anhydrite layers, with secondary deposits of more soluble salts like tachyhydrite, carnallite, and sylvite (Moreira et al., 2007). Dias (1998) estimates that these evaporite depositions occurred for approximately 0.7 to 1 million years.

In the Upper Aptian period, the South American continent's clockwise rotation led to a westward movement along the Romanche, Chain, and Saint Paul break zones (Szatmari and Milani, 2016). This shift led to the invasion of the endoreic basin by the

sea, establishing a marine environment. Consequently, the passive margin or drift phase of the Santos Basin, which overlays the evaporitic section, comprises marine sediments deposited from the Albian to the present day (Petersohn, 2013).

2.3.3. Barra Velha Formation

The Barra Velha Formation, a significant presalt reservoir in the Santos Basin, varies in thickness from over 500 meters to less than 55 meters at the top of faulted blocks and may be locally absent (Wright & Barnett, 2015). Initially, there were debates about whether its depositional environment was marine or lacustrine. While early assumptions leaned towards a marine origin, current evidence, such as the absence of marine fossils and the presence of non-marine indicators, supports a lacustrine setting (Wright & Barnett, 2015). Recent geochemical studies also favour a lacustrine interpretation (Tedeschi, 2017; Pietzsch et al., 2018) (Figure 10).

The nature of the carbonates within the Barra Velha Formation has been a subject of discussion. Some researchers initially identified them as microbial in origin (Moreira et al., 2007; Rezende and Pope, 2015; Muniz and Bosence, 2015; Ceraldi and Green, 2017), but further studies suggest that microbially-produced carbonates are actually rare, comprising less than 1% of the sedimentary deposits (Wright & Barnett, 2015). Instead, chemical precipitation appears to be more dominant in carbonate formation (Herlinger et al., 2017).

The carbonates of the Barra Velha Formation are fairly consistent in composition, consisting of three main components, along with detrital materials derived from them: shrubs, fibrous fronds at a millimetre scale, and spherulites. They are organized cyclically, each cycle ranging from 0.75 to 5 meters thick. These carbonates can include remains of fish and non-marine ostracods, and silicification is a common feature. They can also be reworked into various textures, ranging from muddy facies to well-selected grainstones (Wright and Barnett, 2015). Additionally, evidence of travertines and possible hydrothermal activity in the Campos Basin suggests a similar origin for the carbonates in the Barra Velha Formation (Arienti et al., 2018; Schiffer, 2022; Pietzsch et al., 2018).

Magnesian clays, specifically stevensite, are another characteristic feature of the Barra Velha Formation. Stevensite indicates a saline, hyper-alkaline, and evaporitic

depositional environment with magnesium-rich waters. Its presence is crucial for understanding the type of depositional environment and assessing permeability and porosity in reservoirs (Buchheim and Awramik, 2014; Castro and Lupinacci, 2019).

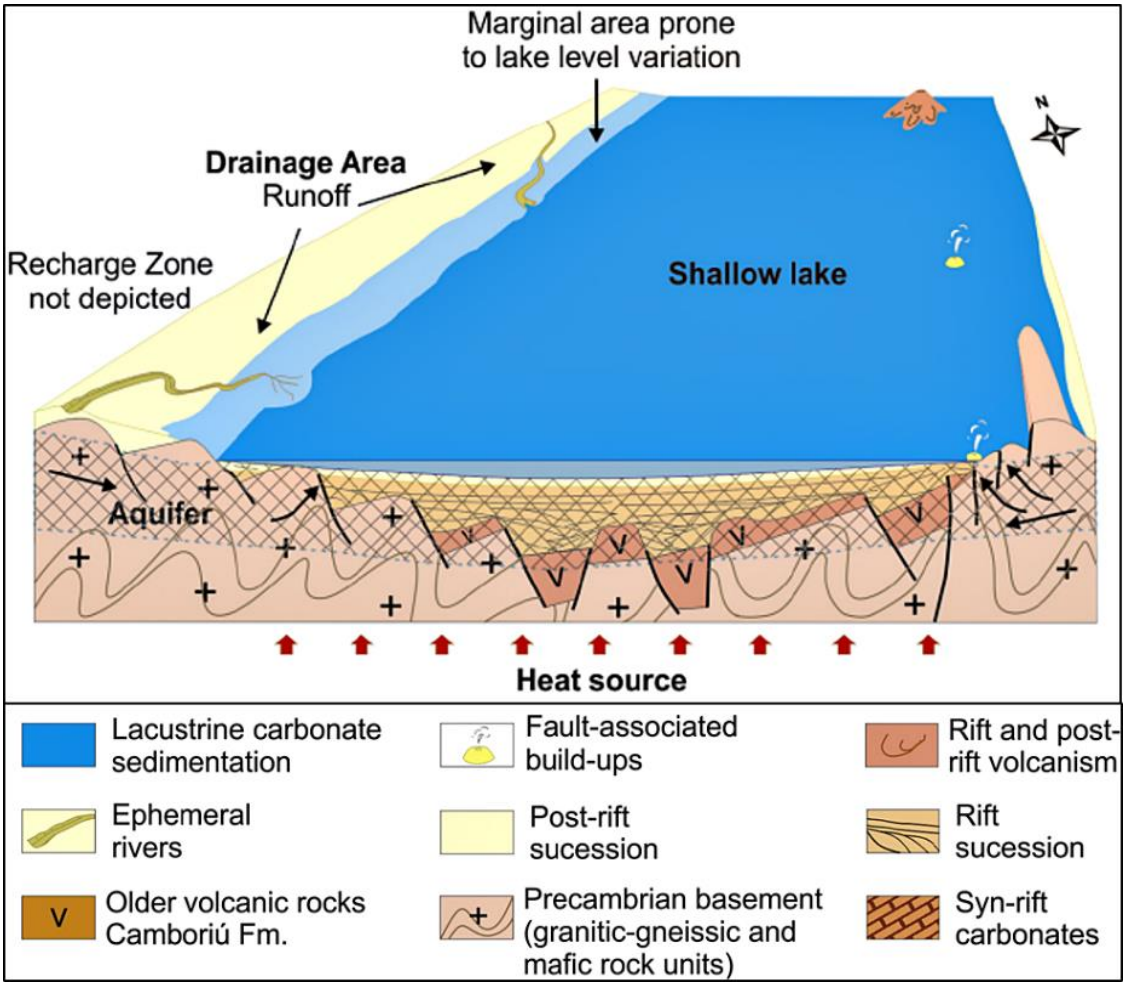

Figure 10: Simplified conceptual model illustrating the Pietzsch's model. There is an endoreic lake. Black arrows indicate the groundwater's infiltration in the lacustrine basin, and red arrows represent geothermal heat. Shaded areas propose the presence of aquifers flowing mainly through the fractured crystalline basement (Adapted from Pietzsch et al., 2018).

In the Sag phase, magnesian clay deposits are frequently replaced by minerals like calcite spherulites, dolomite, chalcedony, pyrite, silica, and svanbergite (Herlinger et al., 2017; Lime and De Ros, 2019). The dissolution of stevensite has a significant impact, potentially giving rise to much of the observed secondary porosity in the samples (Tosca and Wright, 2015). Within the Barra Velha Formation, Mg-clay minerals like stenvensite and kerolite are often layered alongside precipitated calcite spherulites. Additionally, it is common to find fascicular calcite crusts (akin to shrubs) forming at the sediment/water

interface just above the layer of Mg-clay with calcite spherulites (Lima and De Ros, 2019).

Gomes et al. (2020) conducted a data analysis from wells along the horst structure of Santos Outer High. Their findings indicate that the Barra Velha Formation is primarily characterized by three components that were formed in situ: fine-grained mud consisting of microcrystalline calcite and clay minerals, calcite spherulites, and fascicular-optic calcite shrubs. The study introduced a ternary diagram (Figure 11) for classifying these facies, applicable specifically in situations without substantial evidence of significant reworking.

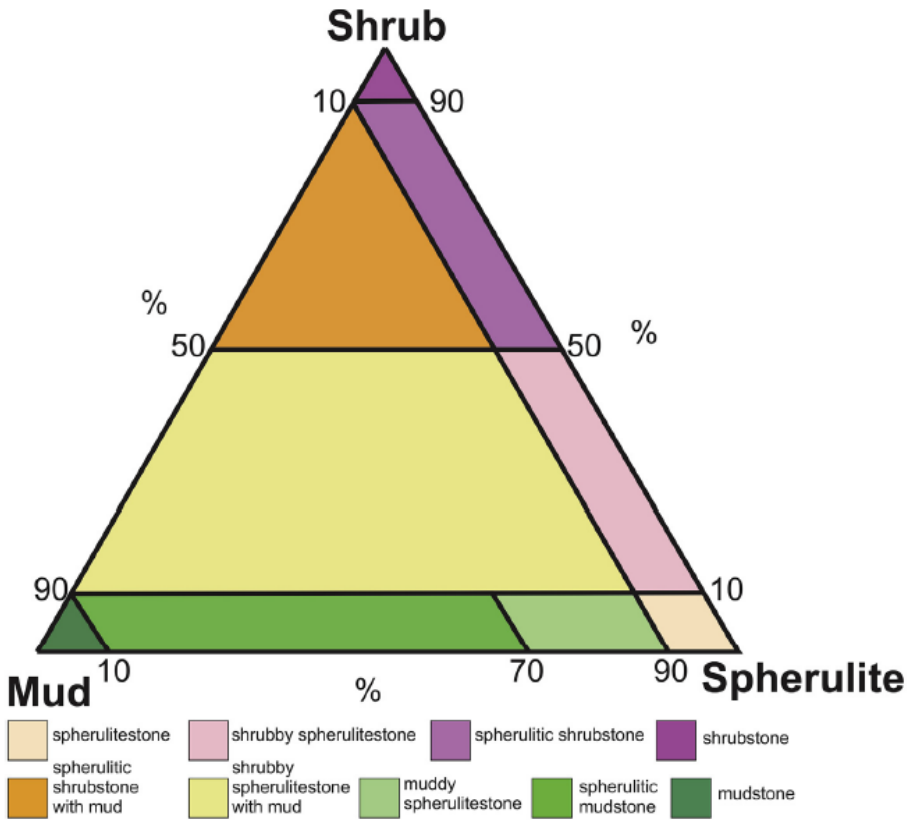

Figure 11: Ternary diagram showing the facies classification for sediment generated in situ for the Barra Velha Formation. Source: Gomes et al. (2020).

These in situ facies developed in lacustrine carbonate aggradational–progradational platforms and build-ups (Ferreira et al., 2019; Wright, 2020; Ferreira et al., 2021) located on structural highs. The reworked facies consist primarily of grainstones and rudstones deposited along shorelines and near main faults such as sandbars, fan-deltas, and debris areas (Ferreira et al., 2019; Wright, 2020; Ferreira et al., 2021).

Gomes et al. (2020) proposed three environmental models for the deposition of the facies succession in the Barra Velha Formation based on traditional stratigraphic principles related to changes in water depth: Humid to Arid Climate Fluctuating Lake-Level Model, Semi-Arid to Arid Climate Shallow-Lake Model, Constant Lake Level Controlled by Spill Point Model. The transition between the climate scenario of each model controls the precipitation rate of Mg-clay and the development of shrubs and spherulites (Figure 12).

Despite recent attention in the literature, many aspects of the Barra Velha Formation's origin and deposits remain uncertain. However, it is widely agreed upon that similar carbonate deposits are rare in geological records. Therefore, further analysis of various aspects of the formation's composition is still necessary due to its complexity.

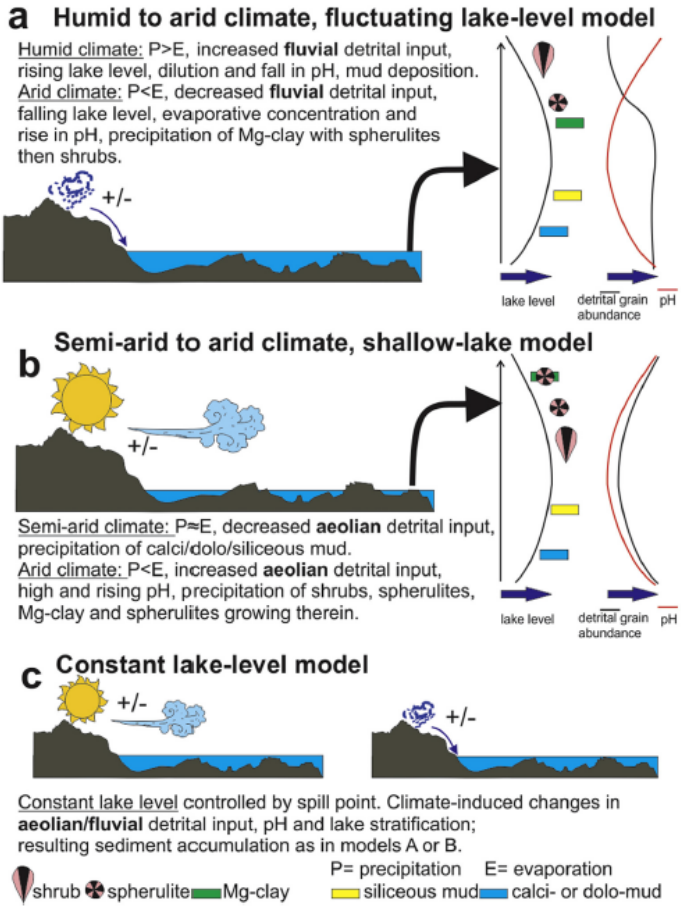

Figure 12: Alternative conceptual models to explain temporal variations of detrital material within the Barra Velha Formation: a) Humid to arid climate, fluctuating lake-level model. b) Semi-arid to arid climate, shallow lake model. c) Constant lake level controlled by spill point. Black arrows point at the stage of the model for the lake x-section. Source: Gomes et al. (2020).

2.3.4. Presalt Diagenesis

Understanding diagenesis in presalt carbonates is crucial for fracture modelling and, consequently, effective production strategies. It involves a complex interplay of physical, chemical, and biological processes during deposition, subsequently intensified by diagenetic processes due to carbonate's chemical reactivity (Herlinger et al., 2017).

Porosity and permeability in these carbonates often exhibit a heterogeneous framework due to intense diagenetic modification, leading to poor correlation between the two parameters. Additionally, fault- and fracture-related alterations contribute to the diagenetic system.

Hydrothermal alteration of carbonate sequences is a critical aspect involving intricate physicochemical processes. It results in the precipitation of "unusual" types of cement, such as saddle dolomite, fluorite, barite, anhydrite, sphalerite, and pyrite (Lima and De Ros, 2019; Carvalho et al., 2023). This alteration is often associated with fluid interaction through faults, fracture systems, and igneous intrusions, promoting processes like dolomitization, silicification, and dissolution to varying degrees and intensities (Lima and De Ros, 2019; Ren et al., 2019). For example, exceptional reservoir performances are found in highly porous dolostones, where intercrystalline porosity is enlarged by dolomite dissolution and porosity enhancement through open fractures (Herlinger et al., 2017).

Silicification is another prominent phenomenon in presalt carbonates, and it is linked to synsedimentary, eodiagenetic, and/or burial diagenetic processes. Various siliceous phases, including chert, chalcedony, moganite, tridymite, cristobalite, and mega-quartz, have been identified in several studies. However, according to Caravalho et al. (2023), crypto-to-microcrystalline silica is the most frequent phase. Fernandez-Ibanes et al. (2022a) defined three main types of silicas in the Barra Velha Formation. They associated these types of silica with nonmatrix excess permeability features (Figure 13). Silica I is the earliest occurrence of silica, and it is strata-bounded. Generally, it occurs near the base of the cycles of depositions defined by Wright and Barnett (2015). Silica II is mainly chalcedony that postdates silica I, and it crosscuts stratigraphy as pore-lining cement. Silica III is the mega-quartz. It is the later stage of silicification, and it occurs mainly in the form of druses along fractures and vug walls. For Fernandez-Ibanes et al. (2022), the predictability of silica is crucial to modelling nonmatrix porosity features, such as fractures and others (Figure 14). Silica I is intimately correlated with opening-

mode fractures, centimetre-scale vugs, and caves. Silica II and III, otherwise, can control the amount of open pore space and overall network connectivity.

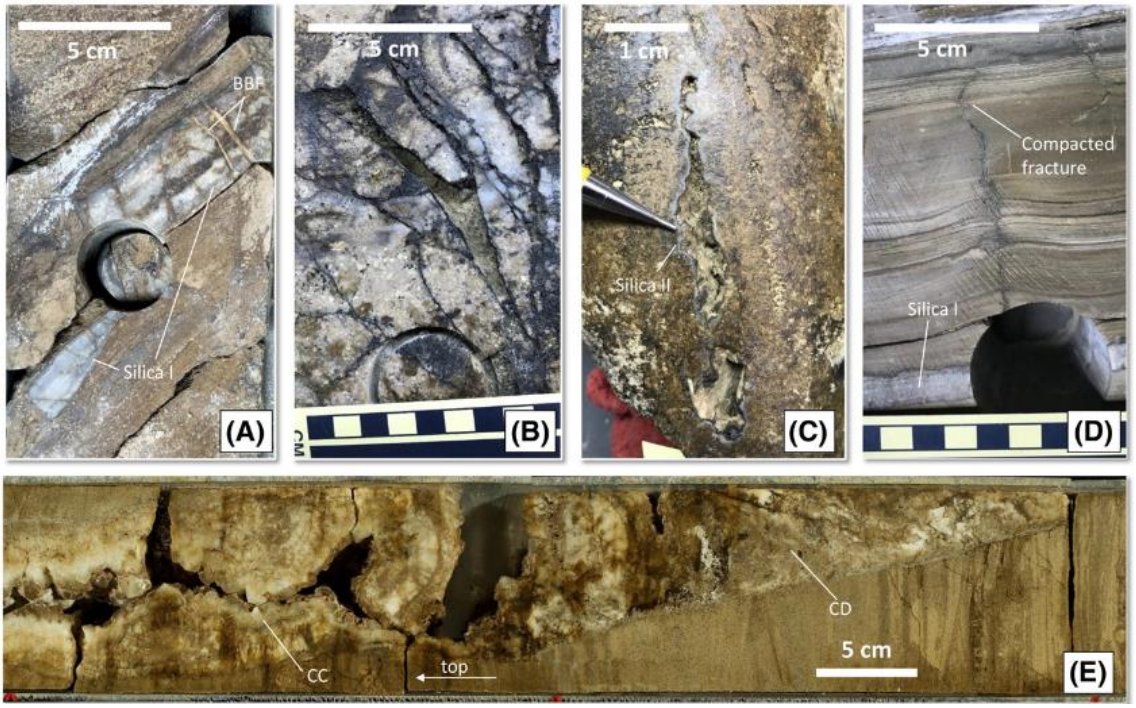

Figure 13: Examples of opening mode fractures in Barra Velha Formation core samples: A) Silica-filled bed-bounded fractures (BBFs) in a 5-cm-thick silicified (silica I) dipping layer. B) Wide-open joints in a silicified layer with bitumen coating. C) Centimeter-wide open fracture with non-matching walls due to dissolution enhancement. Note light blue coating (silica II) and beige-coloured infill of drilling mud. D) Sinuous and filled fractures are interpreted as an early compacted feature. Silica I layer observed at the bottom. E) Meter-scale partly open fracture filled with carbonate debris (CD) and host rock fragments, along with later large white calcite crystals (CC).

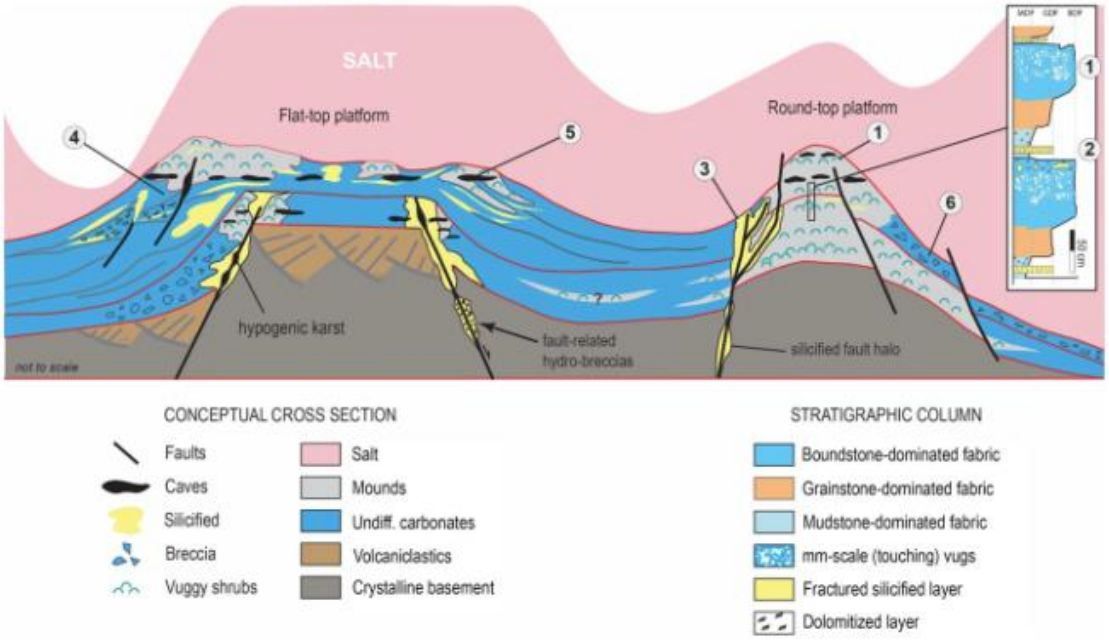

Figure 14: The diagram provides a simplified cross-sectional view (not drawn to scale) of presalt reservoirs, outlining key ideas concerning non-matrix and excess permeability within current reservoir conditions. The inset in the diagram depicts a conceptual representation of the stratigraphic layers at a core scale. Numeric labels on the diagram indicate specific areas associated with various concepts of excess permeability: (1) millimetre-scale touching vugs; (2) bed-bounded fractures; (3) fault zones; (4) fractures and faults because of margin instability;(5) freshwater lens and meteoric exposure-related caves; (6) slope breccias. BDF = boundstone-dominated fabric; GDF = grainstone-dominated fabric; MDF = mudstone-dominated fabric. From Fernández-Ibáñez et al. (2022).

Carvalho et al. (2023) provide further insights into diagenetic alterations, noting the importance of the structural setting (Figure 15). They observe more intense alterations in higher structural areas, likely due to the presence of faults and the occurrence of minerals like saddle dolomite, barite, celestine, and fluorite, which are typical products of hydrothermal alteration. Zhao et al. (2019) observed that igneous intrusions in the Libra block occur mainly associated with deeper faults. This reinforces the link between dissolution processes, hydrothermal influence, and fault presence.

For Carvalho et al. (2023), fractures are mainly found in intraclastic grainstones and fascicular calcite crusts, often filled with quartz and calcite cement. Calcite cementation is typically found in the highest structural parts. While petrographic analysis may indicate low dissolution porosity, core descriptions reveal intense and pervasive dissolution, particularly in structural highs where it is associated with fracturing.

Fernandez-Ibanes et al. (2022b) reinforce the role of early syn-depositional structures, like faults, in enhancing fracture porosity through early karst processes, resulting in cavernous porosity. They note that syndepositional faults repeatedly exposed due to lake level fluctuations are expected to have greater porosity and likely permeability due to developing eogenetic karst features (Figure 16).

In summary, comprehending the multifaceted diagenetic processes in Presalt carbonates is pivotal for optimizing exploration and production in these reserves. The intricate interplay of depositional processes, diagenetic alterations, and the influence of structural position, faults, and fractures shape porosity and permeability, directly impacting production strategies.

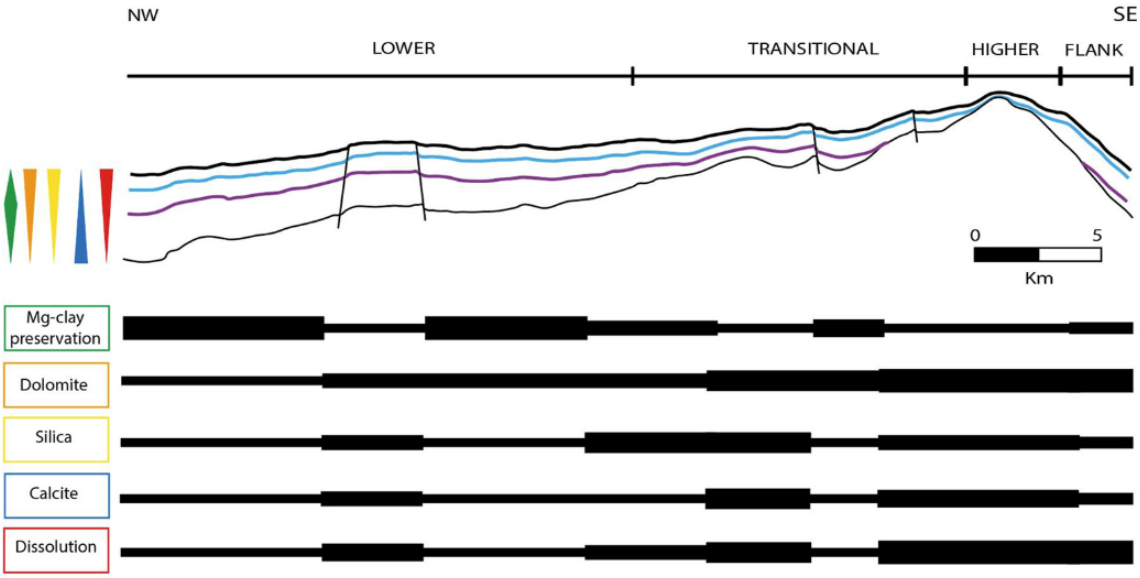

Figure 15: Scheme providing an overview of the lateral and vertical distribution of the principal diagenetic phases within the study area. Notably, the purple line denotes the Intra-Alagoas unconformity, which demarcates the division between Unit 3 and Unit 2. Likewise, the blue line is the boundary between Units 2 and 1. Units defined by Carvalho et al. (2022), following Faria et al. (2017), Artagão (2018), and Gomes et al. (2020). From Carvalho et al. (2022).

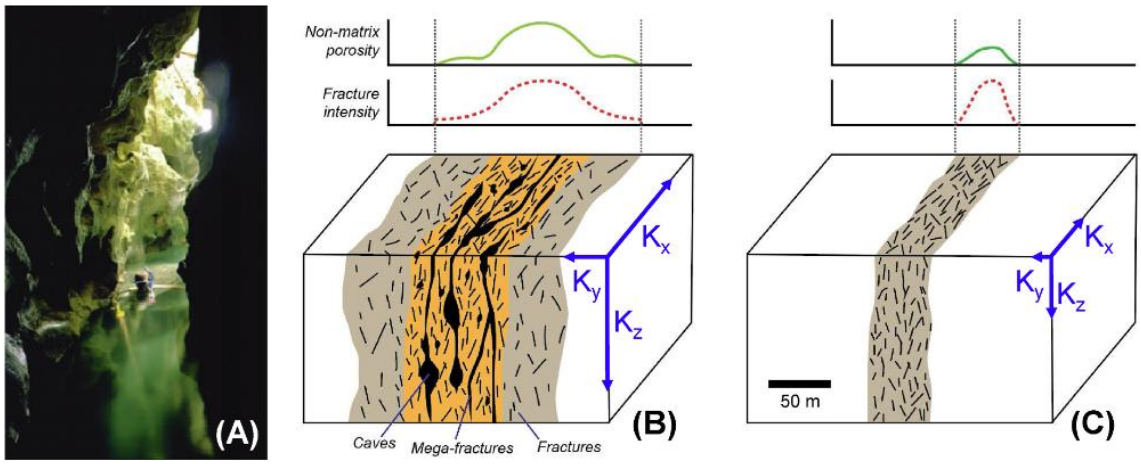

Figure 16: A) Recent dissolution-enhanced fracture shown with scale reference (Photo courtesy of J. Mylroie). B) Conceptual model of a large offset syn-depositional fault zone affected by meteoric dissolution due to lake level changes. Exposure leads to enhanced fracture connectivity, resulting in mega-fracture development and persistent large cavities, as depicted in (A). C) Conceptual model of a small offset burial fault zone with minimal dissolution. Fractures are the primary factor influencing porosity and permeability. (B) and (C) represent extreme fault zone dimensions and properties, impacting reservoir performance significantly. Blue arrows in (B) and (C) indicate relative permeability. From Fernandez-Ibanez et al. (2022b).

3. Fractured Reservoirs and Carbonate Brittleness

This chapter approaches the concept of faults, fractures, and naturally fractured reservoirs. It also reviews a process linked to the heterogeneity of different rock properties and how it affects the formation of fractures in the rock. It also briefly describes how we usually quantify the rock's potential to develop fractures and their actual fracture intensity.

3.1. Faults and Fractures

In this work, it is crucial to understand the concepts of faults and fractures, as it is common to find a misuse of both terms. Both are often used interchangeably in navy form. Although sometimes related, I will review the two terms and how they are formed. At last, I hope the reader understands why they are distinctive features despite being linked.

3.1.1. Fault and Fracture Definitions

A geological fault is a primary geological feature. Before introducing the concept of faults and how they develop, we must define the concept of fractures. Ultimately, the faulting and fracturing processes are intimately linked.

A fracture is any local detachment plane that ‘separates’ a rock, by the process of brittle failure, into two sections (Moore and Wade, 2013). Three types of fractures are extension fractures (mode I), shear fractures (mode II and mode III), and contractional fractures (mode IV) (Figure 17, Figure 18, and Figure 19).

In extension fracture, the fracture plane forms perpendicular to the minimum compressive (or maximum tensile) stress and parallel to the maximum compressive stress; extension fractures exhibit virtually no lateral movement and are commonly termed joints, or fissures, for cases with significant fracture aperture. Fracture aperture (or width) is the degree of openness between the blocks on either side of the fracture; it may be partially or wholly occluded by mineralization (if the latter, it is termed a vein. Shear fractures exhibit significant relative motion between the two rock sections (**Figure 17** and Figure 18). Where shear (Mode II or III) fractures combine with tension (Mode I) fractures are termed mixed fractures or just fractures. Additionally, the term Mode IV (closing mode) is occasionally employed to describe contractional features like stylolites.

Multiple fractures that share similar orientations and occur in proximity are termed a fracture set. Fracture spacing, within a given fracture set, is the average distance between parallel fractures. Its inverse, fracture frequency (or intensity), is the number of fractures per unit length along a profile or within a unit volume.

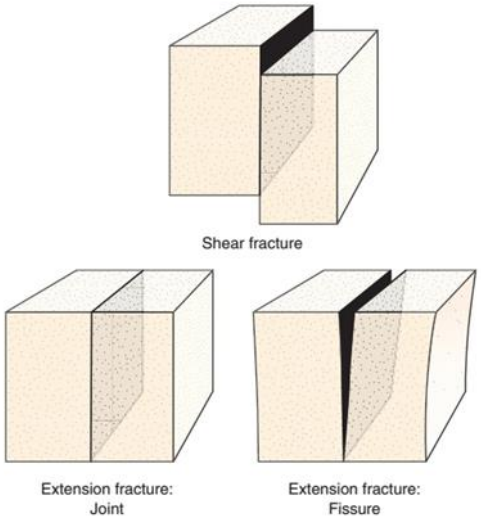

Figure 17: Examples of extension fractures (join and fissure) and a shear fracture. From Fossen (2010)

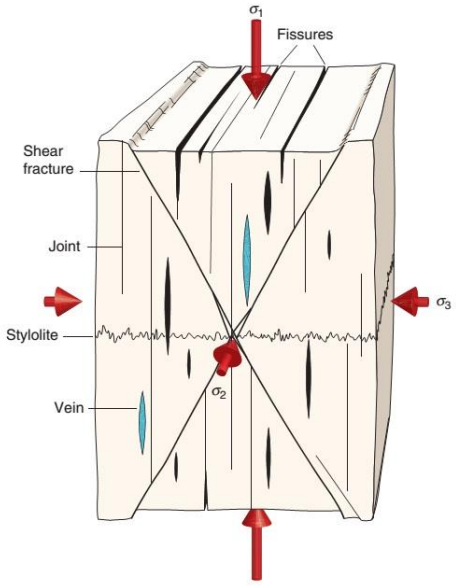

Figure 18: Illustration of the three types of fractures, shear fracture, extension fractures (joint, fissures, and veins), and contractional fractures (stylolite), and how they are related to the stress orientation (sigma 1, sigma 2, sigma 3). From Fossen (2010).

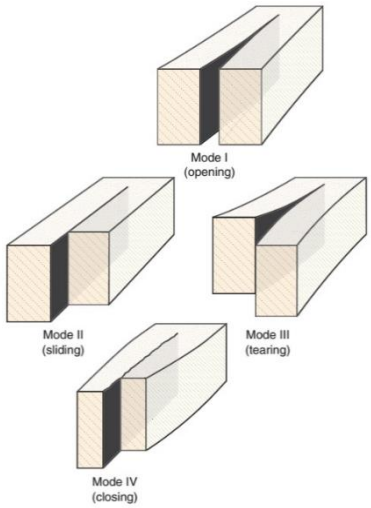

Figure 19: Illustration of the four fracture modes. Mode I, opening; Mode II, sliding; Mode III, tearing; Mode IV, closing. From Fossen (2010).

With the fracture concept introduced, we can define a geological fault primarily as a shear fracture (modelled as a mode II or a mode III crack). However, some (particularly normal) faults contain an opening (mode I) component and are thus mixed mode. However, in the field, a fault rarely is a perfect shearing plane. That is why we use the term fault zone, especially for seismic scale faults. A fault zone is a tabular rock body with some thickness composed of two main hydromechanical units: a core and a damage zone (Figure 22). The core is mainly composed of breccia, gouge, and cataclasite; the damage zone primarily contains fractures of various types and trends.

There are several terms defined to characterize a fault and its geometry. The main ones are fault strike, dip, length, throw, width, and slip direction. Fault strike is the direction defined by the line where the fault plane intersects the horizontal plane, and it is measured in angles with respect to the north. The dip is the orientation of the line on the fault plane that is perpendicular to the strike. It makes the steepest angle concerning the horizontal. The fault length is defined by the dimension of the fault in the strike direction, while the fault width is the length in the dip direction, also called the slip direction. There are several terms, sometimes ambiguous, that define the offset of the fault. Generally, we use the term slip for the offset created by the most recent slip event. The horizontal offset is named heave, while the vertical is named throw. We use the term displacement (or net slip) to refer to the total offset. The term “total” here indicates the vector norm defined by horizontal and vertical components of a single slip event and the accumulated displacement of a series of slip events.

The sense of the slip also determines the style of the fault. Dip-slip faults can be either normal or reverse. If the hanging wall moves upwards relative to the footwall, the fault is called a reverse fault. If it is moving downwards, the fault is called a normal fault. It is called a strike-slip fault if it moves laterally in the strike direction. Oblique-slip faults have a component of dip-slip faults and strike-slip faults. Generally, all faults have components of both dip-slip and strike-slip faults. It is caused by a combination of shearing and tension of compressional forces. These and other types of structures generated by faulting and fracturing, as defined in Peacock et al. (2016), are illustrated in Figure 23.

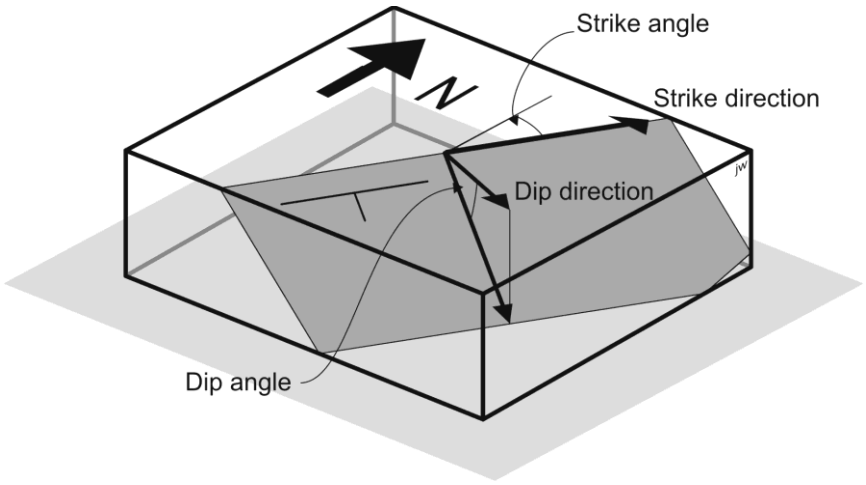

Figure 20: Illustration of the definition of strike, dip, and dip-direction of a plane. Available at https://openeducationalberta.ca/introductorystructuralgeology/chapter/b-orientation-of-structures/.

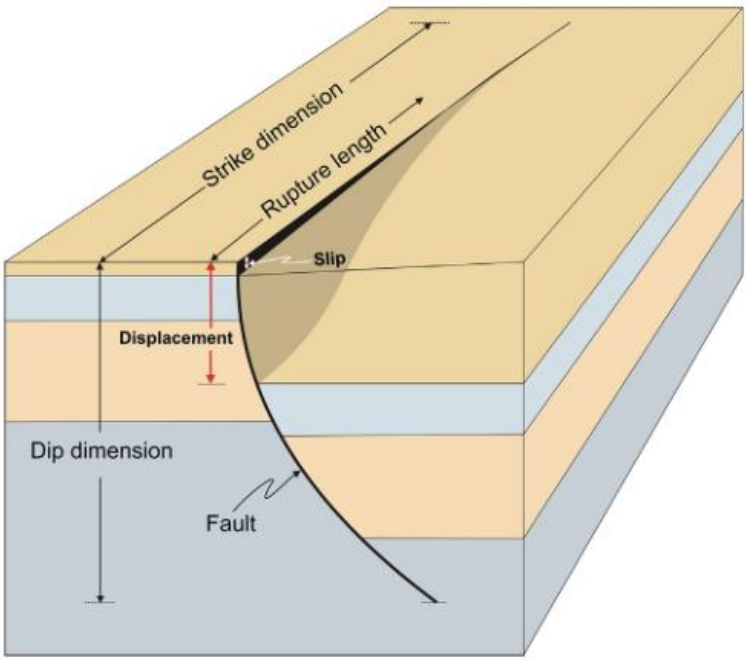

Figure 21: Schematic illustration for the definition of the terms strike dimension (fault-zone length), dip dimension (fault-zone width), displacement (total cumulative fault displacement), rupture length (co-seismic rupture length), and slip (co-seismic slip). See the text for the definitions. From Gudmundsson et al. (2013)

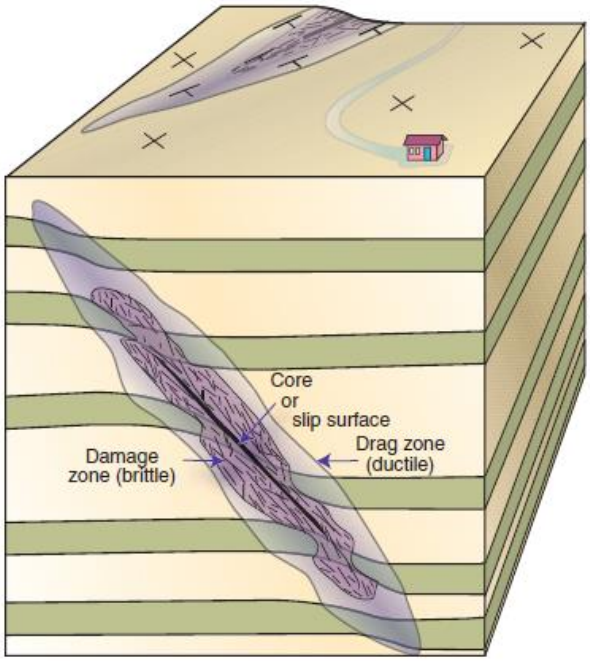

Figure 22: Simplified illustration of the anatomy of a fault. From Fossen (2010).

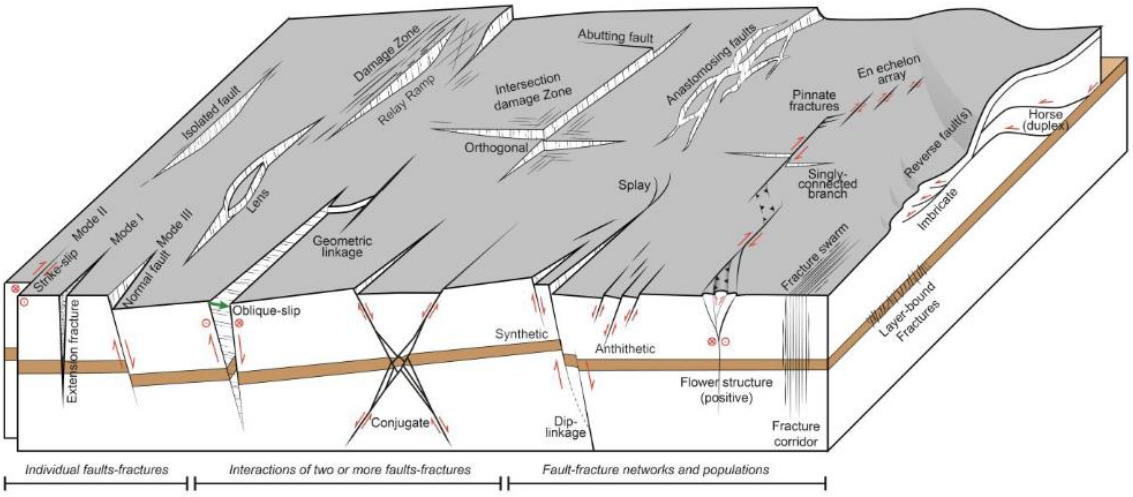

Figure 23: Schematic illustration of the range of fault and other fracture networks defined by Peacock et al. (2016). From Peacock et al. (2016).

3.1.2. Who came first, the faults or the fractures?

A common question when characterizing faults and fractures is regarding each feature produced the other, or in order words, “Who came first?” That is not an easy question with a straightforward answer.

The first difficulty is regarding the concept that faults by themselves can be seen as large shear fractures. However, one main difference is that shear fractures cannot grow in their own plane into a larger structure, while faults can (Fossen, 2016). The second problem is intrinsically related to the birth process of both features.

As Gudmundsson et al. (2013) pointed out, faults normally result from the rocks' flaws or weaknesses. Such flaws include fossils, pores, and principally from all types of fractures. In active areas, joints can commonly be seen to link up into faults, both in lateral and vertical sections, during tectonic events. Seismic slip events are the main contributor to fault growth. Although, some part of the fault growth is due to the accumulation of aseismic slip events. This predominant contribution of seismic slip events is because the main function of faults is to accumulate displacement or strain in order to release stress. That is, to allow the brittle or quasi-brittle crust to deform. Also, most active fault zones are mechanically weak, so they tolerate less shear stress before slipping. That is why most earthquakes and slips occur on existing faults. Consequently, the formation of new faults is less frequent than the slip of old faults.

Concerning the relationship between faults and fractures, fault-related fracturing is a feedback system. As the rock is under some stress, the first deformation features are generally elastic, such as small flexures. In the second stage, fractures such as joints, microfractures, and others start to develop. If the stress is high enough, a fracture swarm is formed. The fracture swarm predates the through-going fault and acts as a process zone that conditions the rock mass for the eventual offset.

In some cases, a large-scale slip does not occur, leaving only the precursive fracture swarm. (Nelson, 2001). This pre-fault fracture swarm within the damage zone is part of the “process zone”, a region where the rock is “processed” before faulting (Figure 24).

The fracture swarm can be predominantly made of shear fractures, but sometimes extension fractures, e.g., joints. Faults formed by faulting of joints inherit some of the features of the original joints. If a fault forms by frictional sliding on a single, extensive joint, the initial fault tends to be a sharp slip surface with almost no fault core and with (almost) no damage zone. If slip accumulates, however, the fault outgrows the joint and links with other joints near its tip zone. The damage zone then thickens, and the fault core may grow. New fractures form in the walls next to the fault core during fault growth. Hence, most faults have a well-defined core of intense cataclastic deformation and a surrounding damage zone of less intense fracturing that diminishes away from the fault (Figure 25). As a final remark, the structures in the damage zone form both before, during, and after the local formation of the slip surface (fault).

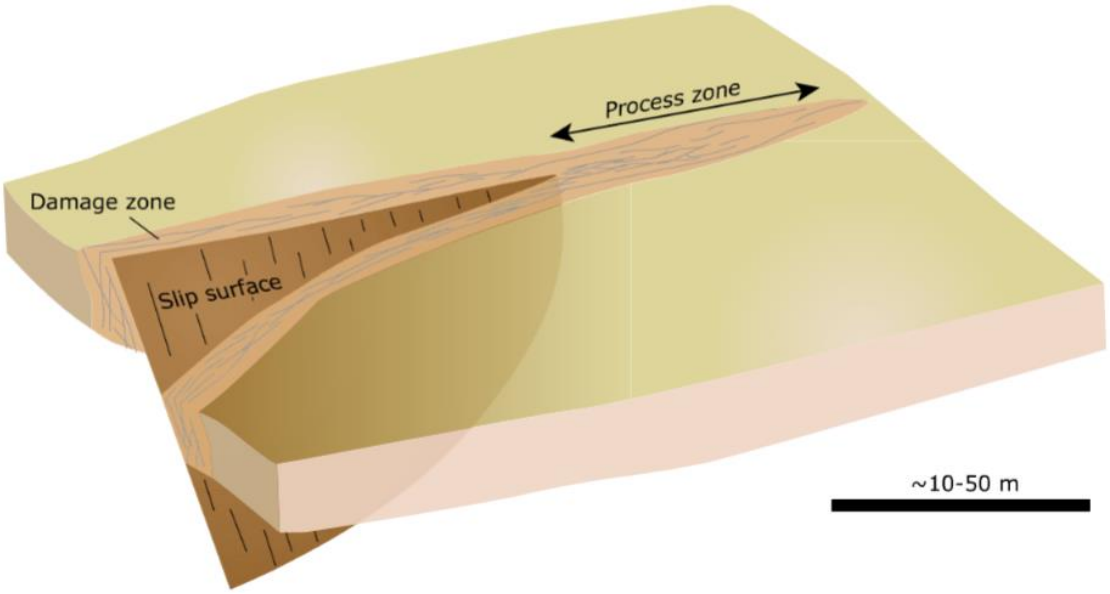

Figure 24: A fault is contained within a damage zone, meaning a (process) zone is ahead of the tip where the rock is "processed" before fault propagation. The process zone may potentially contribute to the compartmentalization of petroleum reservoirs. From Fossen (2010).

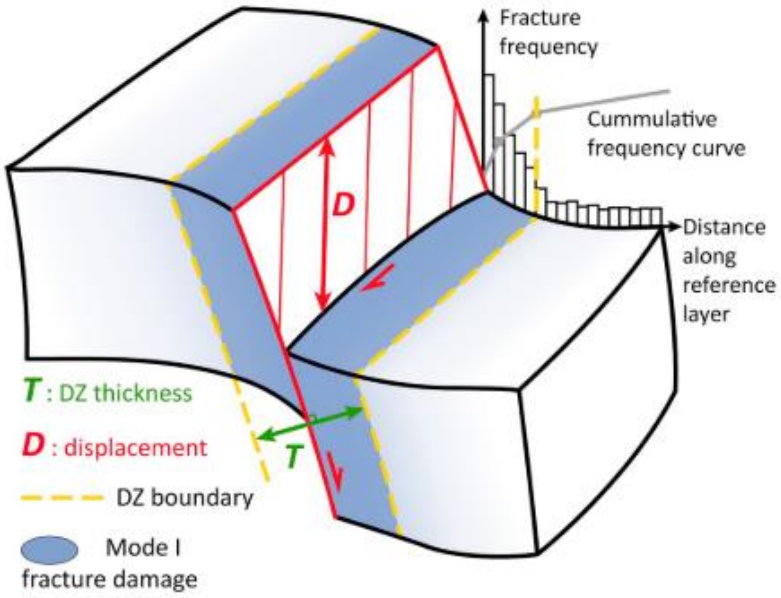

Figure 25: Scheme illustrating the thickness of a fault's damage zone and how the fault's distance is related to the fracture frequency. From Mayolle et al. (2019).

3.1.3. Fold-related fractures

Besides fault-related fractures, there are also fold-related fractures. Actually, as reminded by Moore and Wade (2013), faults and folds often exist in combination with many folds forming in response to movement along faults and many faults rupturing only after intense folding of the strata has taken place. However, as also pointed out by the

authors, fracture patterns associated exclusively with folding are hard to predict. That is because while fault-related fractures formed under the same stress that caused the fault, the same is not true for fold-related ones. During the evolution of the fold, the stress states can change significantly (Stearn and Friendman, 1972). Figure 26 exemplifies two different fold-related fracture patterns defined by (Stearn and Friendman, 1972). While the fracture pattern is totally in the two folds due to different stress orientation conditions, the structure geometry can be the same. Actually, the two patterns can co-exist in the same fold. While reviewing the fold-related fractures, Moore and Wade (2013) used the scheme from Stearn and Friendman (1972) to reinforce that such a difference in the fracture pattern is not simply geological. For example, the drilling strategy for both fracture patterns would be very distinct from each other.

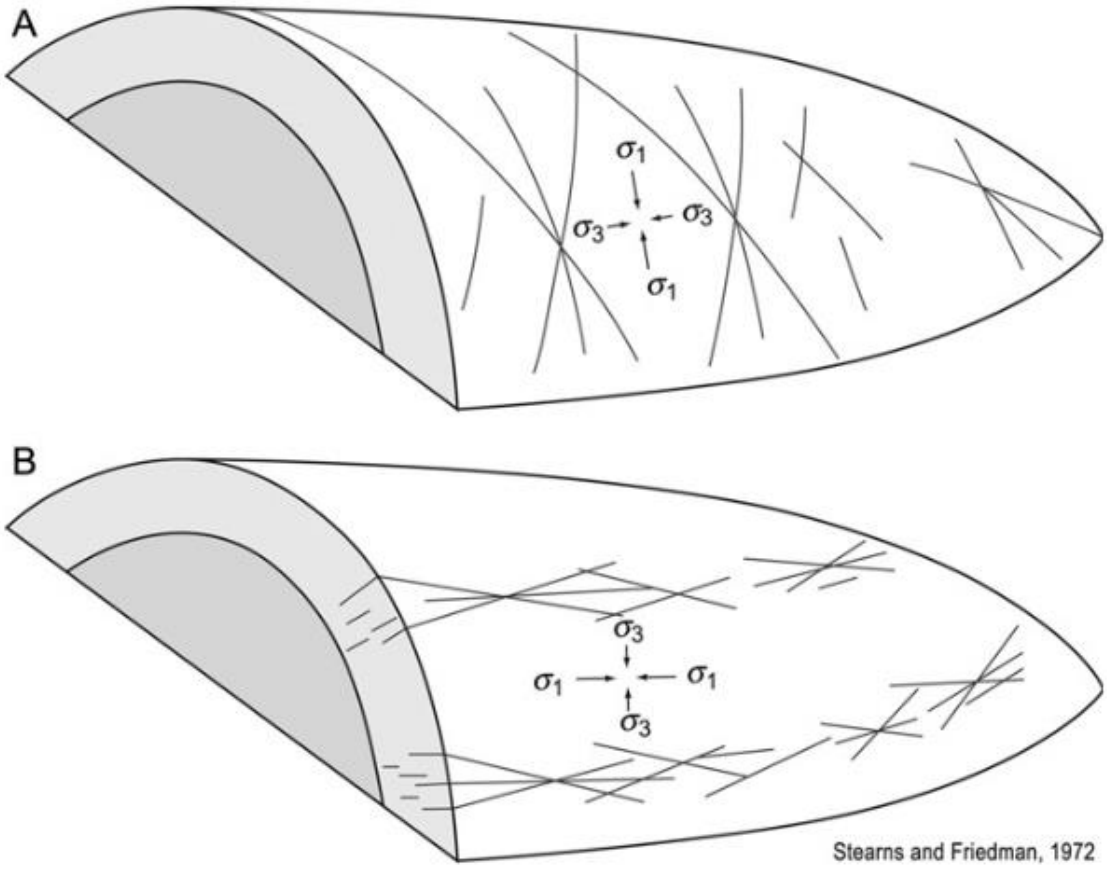

Figure 26: Common primary fracture sets orientations associated with anticlinal folds. From Moore and Wade (2013) after Stearns and Friedman (1972).

3.2. Naturally Fractured Reservoirs

The definition of a naturally fractured reservoir is not entirely clear and depends on the author. Some definitions proposed in the literature are presented by Van Golf-Racht (1996) and Nelson (2001).

For example, Van Golf-Racht (1996) defines a carbonate reservoir as a naturally fractured reservoir "only if a continuous network of various degrees of fracturing is distributed throughout the reservoir." Still, the author proposes some forms of identifying these continuous fracture networks in carbonate reservoirs, and they are: (i) significant mud losses during drilling operations; (ii) special behaviour of transient pressure analysis (e.g., double-slope curves); (iii) cores examination, etc.

Nelson (2001), on the other hand, used a more specific definition, which considers not only the recurrence of the fractures but also its impact on the reservoir properties such as permeability, porosity, and anisotropy. The author defines a naturally fractured reservoir is

> (…) a reservoir in which naturally occurring fractures either have or are predicted to have a significant effect on reservoir fluid flow either in the form of increased reservoir permeability and/or reserves or increased permeability anisotropy.

According to Nelson (2001), fractured reservoirs are difficult to properly evaluate due to (i) A general lack of in-depth quantitative approaches to the description and characterization of highly anisotropic reservoirs, (ii) Failure of geologists and engineers to recognize fractures and/or the regularity of their distribution; (iii) Over-simplistic approaches in the description of fracture distributions and morphologies; (iv) The need for a deterministic solution to modelling fluid flow in fractured porous media, while understanding that our data limitations force us toward stochastic solutions, at best.

In his work, the author reinforces that just "finding fractures is not enough." This means that to evaluate a fractured reservoir's potential successfully, we should be able to identify where the fractures are and other aspects, such as intensity, attitude, width/aperture, length, fracture porosity, fracture permeability, and others.

3.3. Rock Brittleness

In rock mechanics, one of the main objects of study is understanding how a given rock produces brittle deformation after applying some stress and under some circumstances. This ability of the rock to produce brittle deformation is called brittleness. Notice that this is not a formal definition. To Meng et al. (2021), brittleness is an underdefined term in geoscience that is misused with several and sometimes ambiguous definitions.

One way to understand the mechanical behaviour of the rock is through analysis of pre-failure strain (also referred to as ductility, the counterpart of brittleness). Laboratory testing of various common sedimentary rock types demonstrates substantial changes in ductility due to confining pressure and rock type (e.g., Handin and Hager, 1957; Donath, 1970).

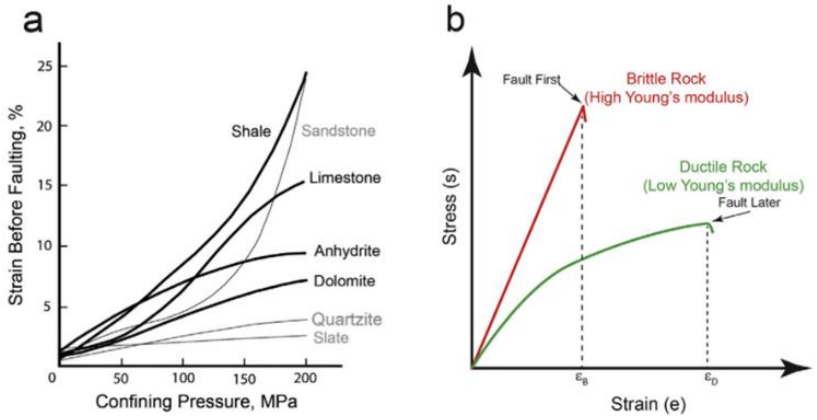

Figure 27: Graph of per cent strain before faulting (ductility) versus confining pressure for a range of common rock types, including lithologies common in carbonate rock sequences. Stress versus strain curves for brittle and ductile rocks illustrate the lower strain at faulting in brittle rocks (ε_B) versus that for faulting in more ductile rocks (ε_D).

These experiments can be analysed under the light of mechanical models. Hooke's Law is a fundamental model in the field of elasticity. It establishes a linear relationship between stress (σ) and strain (ε) within the elastic limit of a material:

$$\sigma = E\,\varepsilon \qquad \text{Eq. (1)}$$

Where E represents Young's Modulus, a measure of a material's stiffness and its ability to resist deformation under an applied load. Young's Modulus is defined as the ratio of stress to strain in the elastic regime:

$$E = \frac{\Delta\sigma}{\Delta\varepsilon} \qquad \text{Eq. (2)}$$

Other elastic parameters are the Shear Modulus, Poissons' Ratio, and Bulk Modulus.

Shear Modulus (G) is of significant importance. It characterizes a material's resistance to shearing deformation when subjected to perpendicular forces. It is defined as:

$$G = \frac{\tau}{\gamma} \qquad \text{Eq. (3)}$$

Where τ is the shear stress and γ is the shear strain. For uniaxial strain, G can be related to E by:

$$E = 2G \qquad \text{Eq. (4)}$$

Furthermore, Poisson's Ratio (ν) is a measure of a material's lateral contraction when subjected to an axial load. It is defined as the negative ratio of transverse strain (ε_t) to axial strain (ε_a):

$$\nu = -\frac{\varepsilon_t}{\varepsilon_a} \qquad \text{Eq. (5)}$$

Lastly, the Bulk Modulus (K) characterizes a material's resistance to uniform compression. It is defined as the ratio of change in pressure to the fractional volume change:

$$K = -V\frac{dP}{dV} \qquad \text{Eq. (6)}$$

Where V is the volume of the material, and P is the applied pressure.

These elastic constants play a crucial role in assessing the mechanical behaviour of rocks. Young's Modulus indicates a material's stiffness, the Shear Modulus suggests resistance to shearing forces, Poisson's Ratio describes lateral contraction, and the Bulk Modulus relates to resistance to uniform compression. A rock with a low E-value is mechanically weak, as its resistance to deformation is small. Since strain is

dimensionless, Young's modulus has the same dimension as stress and is typically given in GPa (10^9 Pa). Table 1 gives some examples of experimentally determined strengths and their characteristic values of E and ν.

Table 1: Representative values of Young's modulus (E) and Poisson's ratio (n) for some rocks and minerals. From Fossen (2010)[1] and Simm and Bacon (2014)[2], after Simmons and Wang (1971).

Medium	$\boldsymbol{E(GPa)}$	$\boldsymbol{\nu}$ $\boldsymbol{(Poissons's\ Ratio)}$
Quartz	95.8	0.064[2]
Salt	40	~0.38[1]
Halite	38.3	~0.25[2]
Anhydrite	87.0	~0.28[2]
Gypsum	77.1	~0.28[2]
Limestone	80	0.15-0.3[1]
Calcite	84.3	0.317[2]
Dolomite	116.6	~0.30[2]
Pyrite	338.7	0.137
Sandstone	10-20	0.21-0.38[1]
Shale	5-70	0.03-0.4[1]
Clays	41.3	0.324[2]
Gabbro	50-100	0.2-0.4[1]
Granite	~50	0.1-0.25[1]
Amphibolite	50-110	0.1-0.33[1]
Marble	50-70	0.06-0.25[1]

In practice, we use some brittleness index to determine the brittleness of the stratigraphic intervals in an oilfield. There are several postulations of the brittleness index. In geophysical studies, we generally use brittleness index based on mineral proportion (Jarvie et al., 2007; Wang and Gales, 2009), elastic parameters ratios (Grieser and Bray, 2007; Rickman et al., 2008; Guo et al. 2012a; Chen et al., 2014; Liu and Sun, 2015; Huang et al., 2015), or by rock physics model (Qian et al., 2020). Although, Meng et al. (2021) pointed out the existence of at least 11 distinct groups of brittleness index. For carbonates, it is common to work on the first two types of brittleness index, mineralogy-based and or

elastic parameters-based. For example, a common process in carbonates is mineral substitution, replacing calcite with quartz or dolomite, which increases the rock's brittleness. As in Table 1, quartz and dolomite have a higher Young's Modulus than calcite. Therefore, substituting calcite for these minerals tends to increase the stiffness of the rock. They also have a lower Poisson's Ratio than the calcite, especially the quartz. More brittle rocks tend to have lower Poisson's Ratio, while more ductile rocks have higher values (Grieser and Bay, 2007; Kibikas et al., 2021).

In the present study, despite not using any brittleness index directly, I took their premises as the basis for selecting some features that can be potential predictors of fracture intensity in the presalt, indicating which stratigraphic intervals are more or less prone to the development of fractures. They are, Silica content, Young's Modulus, and Poisson's Ratio.

3.4. Mechanical Stratigraphy

The degree of fracturing linked with faulting seems to be influenced by factors such as lithology, proximity to the fault plane, the extent of displacement along the fault, overall strain within the rock mass, burial depth, and conceivably the specific type of fault (e.g., thrust, growth, normal, and others). These several factors change across different rock unities and facies. This lateral and vertical variation of fracture according to rock properties is called mechanical stratigraphy.

One of the first studies that started to understand this effect was done by McQuillan (1973). The author observed that the fracture intensity in the Asmari limestone from Iran was more influenced by bed thickness than any other factor. In his case, the fracture intensity had an inverse logarithmic relationship with bed thickness. Besides, this relationship is valid across the whole structure despite the location. However, even in this case, they observed that while tectonics did not create the fractures, they could enhance their size and openness.

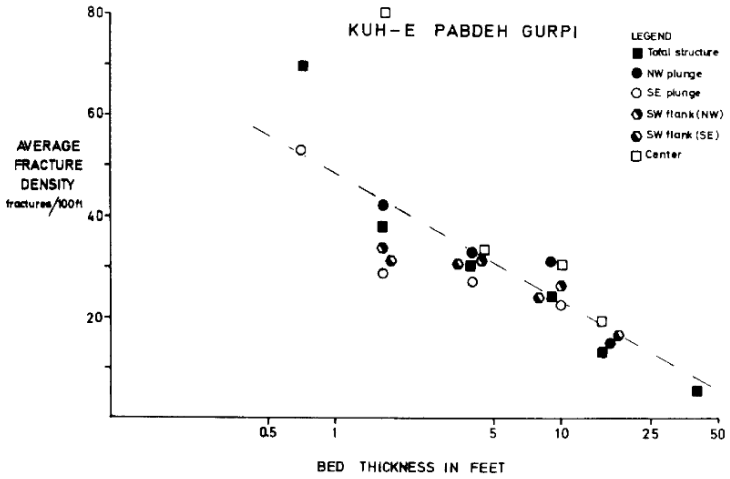

Figure 28: Graph demonstrating the relation between bed thickness and fracture density for various structural realms and the total structure of Kuh-e Pahn anticline. From McQuillan (1973).

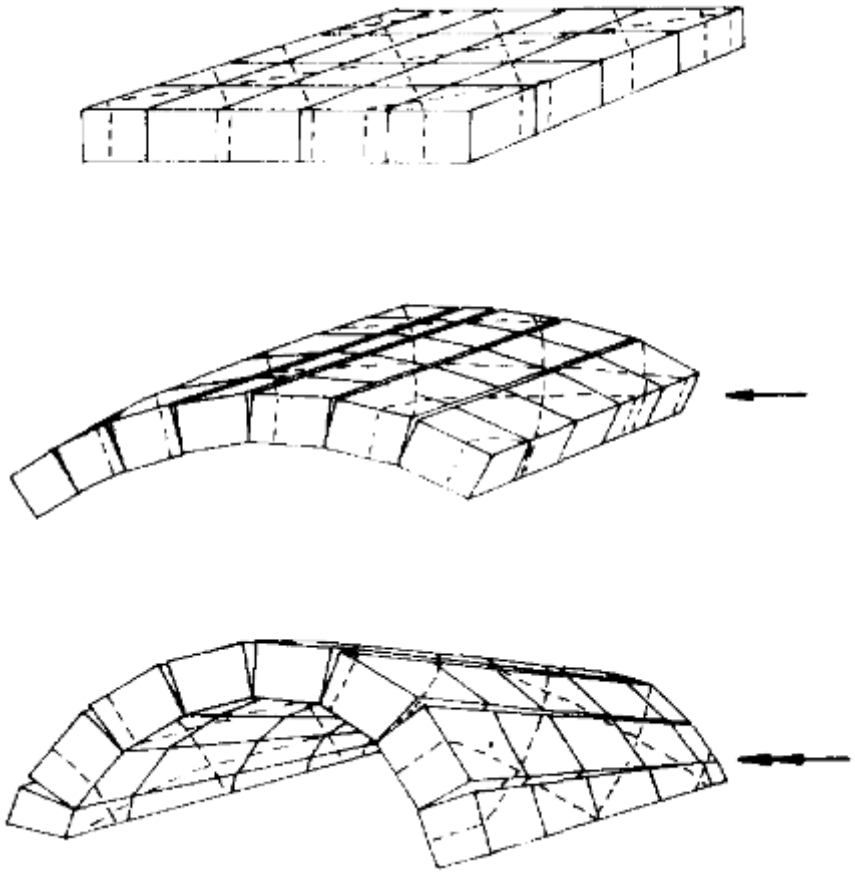

Figure 29: Illustration showing how folding occurs after preexisting fracture sets enhance fracture aperture. This process increases strike-parallel porosity and permeability in limestone formations. From McQuillan (1973).

There are also other features related to mechanical stratigraphy. Wennberg et al. (2006), studying the Asmari Formation at the Khaviz Anticline, show that, in this

location, the principal rock parameter that controls fracture intensity was not bed thickness. Instead, rock texture seems to be leading to the development of fractures. In their case, mud-supported carbonates have a consistent higher degree of fractures than grain-supported carbonates. The authors assigned such behaviour to the fact that grain size is one of the main rock properties that control rock strength. As Hughman and Friedman (1979) reported, the higher the grain size, the lower the rock strength. Wennberg et al. (2006) also concluded that mechanical thickness only exerts a greater control in fracture development for geological scenarios of strong contrast in mechanical properties (shales/marls interbedded with sandstones/limestones) and single-phase tectonic development. In other cases, bed thickness loses importance as a parameter influencing fracture intensity. Another work that reinforces the role of rock texture in mechanical stratigraphy is Askaripour et al. (2022).

Ferril and Morris (2008) and Ferril et al. (2017) demonstrated that mechanical stratigraphy controls not only fractures but also faults. In their case, faults in clay-poor massive limestone tend to be steeper than faults in clay-rich limestones. Also, the rate of incompetent and competent strata controls the displacement configuration and width of the fault zones (Figure 30). In summary, mechanical stratigraphy influences nucleation, failure mode, fault geometry, displacement gradient, displacement distribution, fault core and damage zone characteristics, and fault zone deformation processes.

Watkins et al. (2015), studying the relationship between curvature and fracture intensity, observed that it varies according to structural positioning and mean grain size. The structural position will be the most important parameter controlling fracture intensity, mainly for high-strain regions, while the lithology is the most influential for low-strain regions.

The mineral composition is another important aspect of mechanical stratigraphy, as Du et al. (2022) and Dinç Gogus et al. (2023) reported. Rock beds dominated by brittle materials such as dolomite and quartz tend to develop more fractures than other rocks with different mineral compositions.

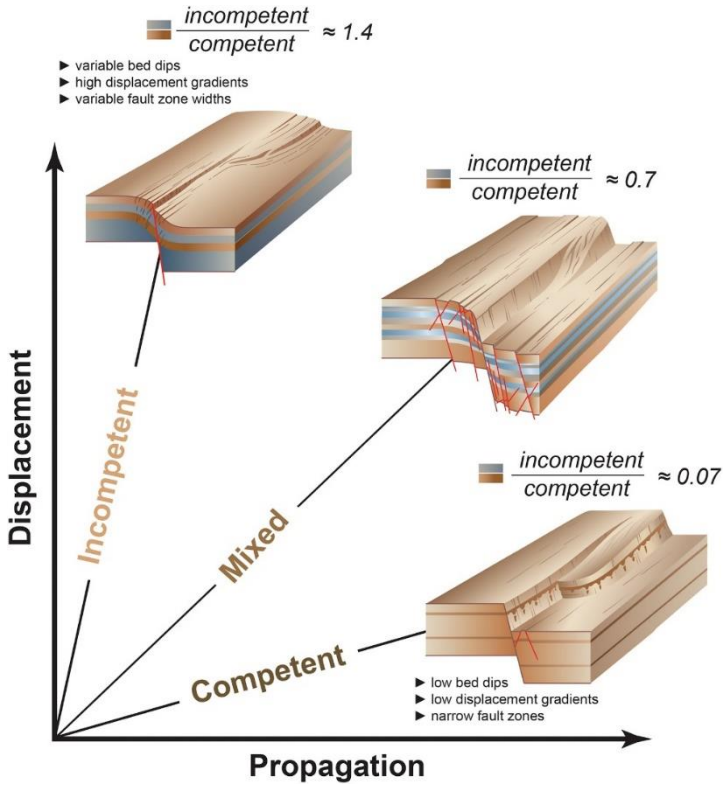

Figure 30: Illustration highlighting how mechanical stratigraphy affects the balance between fault propagation and displacement in units with different thickness ratios of competent to incompetent material. Faults in mechanically competent layers tend to propagate rapidly compared to accumulating displacement over time. In cases with low incompetent to competent thickness ratios (e.g., 0.07), displacement ratios and gradients will be low. Conversely, higher ratios (e.g., 1.4) result in elevated displacement ratios and steeper gradients. Stratigraphies with intermediate ratios exhibit varied behaviour based on local conditions. From Ferril et al. (2017).

3.5. Fracture Measurements

To determine how much a rock mass is fractured, we have to use some measurement system to keep consistency. In that sense, Dershowitz and Herda (1992) developed the Pij system (Figure 31). This measurement system describes the intensity level according to two dimensions. The first is the dimension of the sample, "i". It describes how you are measuring the fractures. The second one describes the dimensions of the measurement, that is, what we are measuring.

For example, suppose we are measuring the fractures in a well. What we generally do is count the fractures in depth interval that has a given length. Then, if we set this length to the unity scale, we will measure the P10 from the Pij system. The "i" equals one

represents the unity length of the well, and the "j" equals 0 describes that we count the number of fractures, which is a dimensionless measurement.

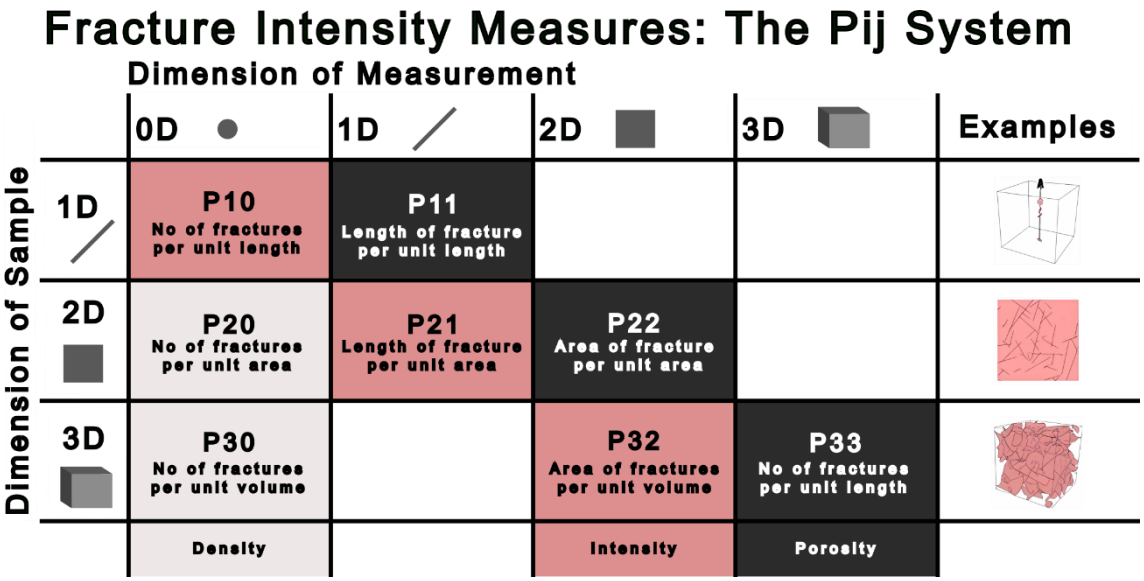

Figure 31: Fracture Intensity Measures using the Pij system (Dershowitz and Herda, 1992). P32 is a valuable measure of fracture intensity with applications in geoscience and engineering studies. An increase in P32 corresponds to a greater fracture area per unit volume, providing important insights.

The most used measurements are from the diagonals of the 3x4 matrix defined by all measurements from the Pij system. Therefore, they are the P10, P21, and P32, or P11, P22 and P33 (Figure 31). In geoscience studies for hydrocarbon exploration, generally, we first estimate the P10 in the wells, convert it to P32 (area of fractures per unit volume, or volumetric fracture intensity) for seismic studies, and finally, using fracture aperture and width, we estimate the P33, which ultimately gives the fracture porosity. We can also estimate the fracture permeability from other information, such as fluid viscosity, fracture orientation, and length. Fracture porosity and fracture permeability are two essential properties for naturally fractured reservoir modelling, for example, to run flow simulation considering that fractures can contribute to the porosity and/or permeability.

As stated above, if we wish to have a volumetric estimation of the fracture in the whole reservoir, we must calculate the P32. Still, the standard measurement in the wells gives us the P10, so we have to convert from one measurement to another. There are several ways to do so. One straightforward way is in the case that we have outcrops from the reservoir. In that case, we can calculate both the P10 and P32 in the field, and then, using regression, we can find a scale factor for the conversion. Although, for cases in which we do not have an outcrop, we have to count on empirical rules, such as the ones

determined by Terzaghi (1965), Zhang and Einstein (2000), Chilès et al. (2008), and Ojeda et al. (2023). Despite some differences in the elaboration made by different constraints, all those methods consider that to convert the P10 to P32, we have to consider the inclination of the wells concerning the fractures' dip or the main dip of the fracture set. That is because, suppose that we have mostly semi-vertical to vertical fractures, a horizontal well would give the optimum sampling orientation. A vertical well would not catch the actual fracture intensity level and underdetermine the number of fractures in the rock mass volume. The Terzaghi correction approximates the true fracture spacing D by the inverse of the sine of the angle α, which is the angle the sampling survey makes with the fractures (Figure 32). So, the corrected distance d is:

$$d = \frac{N}{L\ sin\alpha}$$

Where N is the number of fractures intersected by the survey, and L is the length of the survey.

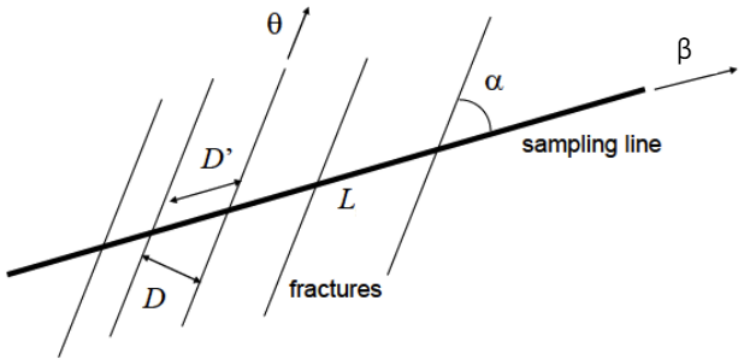

Figure 32: Apparent spacing D' associated with true spacing D when the fractures (direction θ) form an angle α with the scanline (direction β). From Ojeda et al. (2023). Adapted from Chilès et al. (2008).

However, Terzaghi correction assumes a zero-thickness sampling survey with constant orientation, while a borehole has some thickness and varying orientation. Then Chilès et al. (2008) generalized Terzaghi weighting, rewrote it for varying orientation and put it in terms of predicted volumetric fracture intensity, $\hat{P}_{32}$:

$$\hat{P}_{32} = \frac{1}{L}\sum_{1}^{n}\frac{1}{|cos(\omega_i - \beta_i)|}$$

Where L is the length of the scanline, β_i is the local orientation of the scanline at the location of fracture i, N is the Number of fractures, and ω_i the pole of the fracture i.

The same equation can be simplified using the acute angle between the scanline and the fracture i (α_i) as:

$$\hat{P}_{32} = \frac{1}{L}\sum_{1}^{n}\frac{1}{sin\alpha_i}$$

It should be noted that the weight factor tends to infinity for α near zero. Therefore, one should apply a maximum limit to α (minimum bias angle). According to Ojeda et al. (2023), in practice, a limit of 15° is often used to improve the estimator's robustness.

In Figure 33, we illustrate some simulated P10 and P32 measurements to help the writer understand how the fracture intensity given by P32 is related to P10 and is translated into the degree of fracturing of a rock mass.

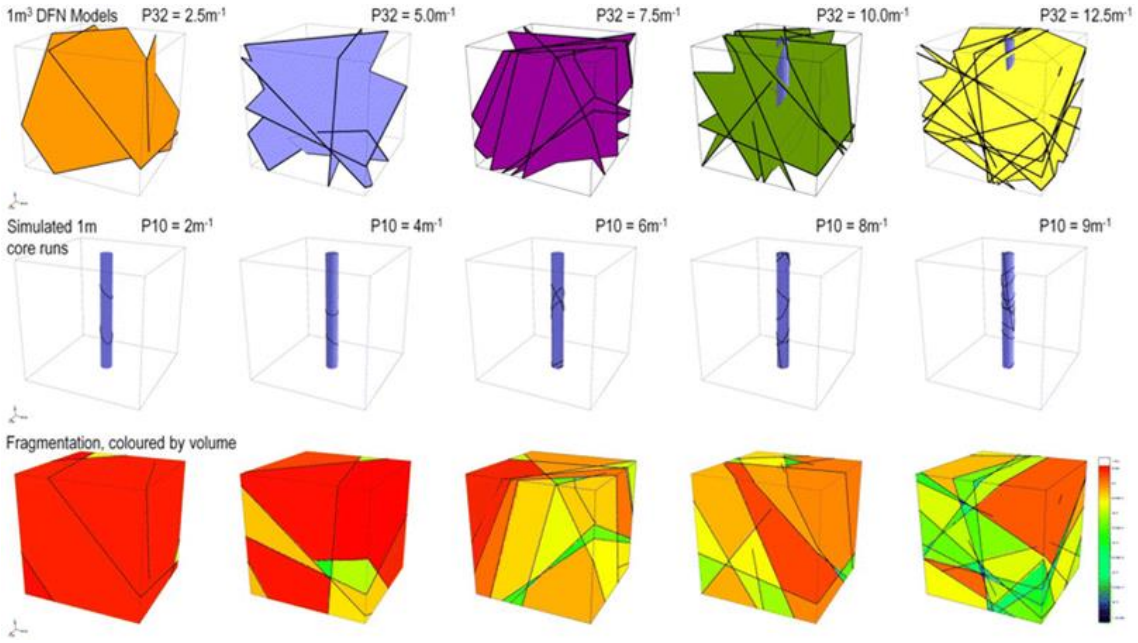

Figure 33: Schematic illustration showing distinct levels of simulated P10 and P32. From Rogers (2023).

4. Geometric Seismic Attributes

In this chapter, we review the basis of seismic attribute. Afterwards, we focus on the geometric seismic attributes used in this study: dip, curvature, coherence, and fault enhancement attributes.

A seismic attribute is any measurement or information obtained directly or indirectly from seismic data (Taner, 2001). The list of seismic attributes can be vast, according Taner (2001) the most traditional attributes are the seismic amplitude, the elevation time (or depth) of a seismic horizon, and the temporal (or depth) thickness between two reflections. These were probably the first seismic attributes used for hydrocarbon exploration. In the late 70s, after Taner and Sheriff (1977) and Taner et al. (1979) worked on complex seismic attributes, when the industry started to understand the power of filters to enhance geological information hidden in the standard seismic amplitude. Since then, countless number of seismic attributes have been created by several authors for different purposes. However, the hype on the subject heavily influenced the outstanding expansion of seismic attributes.

Consequently, many attributes created in that period and available in commercial software are redundant, useless, or have limited applicability. Until today, new ways of using seismic attributes to obtain geological information or better ways of calculating standard attributes to improve its responses are still a topic of research.

Chopra and Marfurt (2007) stated that a seismic attribute can be considered good if it is either directly responsive to the desired geologic feature or reservoir property of interest or if it enables the interpreter to indirectly define the structural or depositional environment in a way that allows for the inference of certain features or properties of interest. However, in the face of multiple attribute options, it may be hard for the interpreter to select a proper list with attributes that would benefit his understanding of the subsurface.

The abovementioned problem leads to a common question when working with seismic attributes: "Which attribute should I use?" The answer to this question is not straightforward and heavily depends on the interpreter's task and knowledge, the characteristics of the seismic data, and the imaged subsurface geology. To help address this question, some authors have proposed taxonomy classifications for seismic attributes.

Those would help interpreters focus on selecting and understanding the family of attributes more suitable for the task rather than solely focusing on a single attribute or a random selection. One of those taxonomy proposals is made by Taner (2001). In his work, he defined two main classes of attributes: pre-stack attributes and post-stack attributes. He further classified the post-stack attributes into other groups, such as instantaneous, wavelet, physical, and geometric attributes. In this work, I am going to focus on geometric attributes.

As Taner (2001) defined, geometrical attributes describe the spatial and temporal relationship of a set of seismic traces. The relationship can be based on any aspect of seismic trace, amplitude, phase, frequency, and others. By doing so, one can measure changes in reflector shape or morphology. Therefore, ultimately, it is a group of attributes that aims to identify and highlight edges, folds, bowls, discontinuities, and reflection patterns.

In the context of this study, I want to evaluate the use of seismic attributes that can help to identify and characterize faults and fracture zones. Faults and fractures can be seen as discontinuities in the subsurface geology. Therefore, it makes sense to select attributes in the group of geometric attributes. They are useful to highlight any seismic discontinuity that is hopefully related to the presence of fault and fracture zones. I will use curvature, coherence, and structurally oriented filters. Although, all three attributes need information from the dip of seismic events. Curvature is an attribute related to the second derivative of the seismic data and, therefore, the first derivative of the dip field. The coherence and structurally oriented filters need the dip information to guide the window-analysis of their algorithms. I will explain more about this in the following topics.

4.1. Dip

The dip attribute is one of the most important attributes because it is the basis for many others. Also, the dip is vital geological information for structural and stratigraphic interpretation studies. Although, it is important to remember that a seismic reflection is simply an image, an imperfect representation of an actual geological reflector. Therefore, any property estimated from this image will represent the actual subsurface geology property as good as the image does. In the case of dip-estimation, it suffers mainly from

poor signal-to-noise ratio and limited seismic resolution (vertical and lateral). For time-domain post-stack migrated seismic data, the dip estimation can also suffer from seismic artefacts induced by velocity distortions.

Barnes (2016) presents a proper definition of the dip attribute. He defines the dip in terms of the slope and azimuth. Let us first define the slope in 2D seismic data. In depth-domain seismic data, the slope is the dimensionless ratio of the change in depth of a reflection over a certain horizontal distance. Therefore, the slope p in the *x*-direction is defined as:

$$p_x = \partial z/\partial x \qquad \text{Eq. (1)}$$

The dip is the angle in degrees, γ, that a seismic reflection makes with the horizontal, which is the arctangent of the slope. In the case of a 3D seismic volume, both the slope and the dip are positive magnitudes, corresponding to the maximum slope and dip that are given according to the azimuth direction. For depth-domain seismic data, the γ angle is defined by:

$$\gamma = \tan^{-1}(p), \qquad \text{Eq. (2)}$$

Where $p = |\partial z/\partial r|$ and ∂r is defined as $1/\partial r^2 = 1/(\partial x^2 + \partial y^2)$.

For time data, the slope becomes inverse apparent velocity with units of milliseconds per meter, and the dip must be estimated using a conversion velocity. The slope is a geophysical property, and the dip is a geological property. Slope in 2D and slope components in 3D are signed according to the convention shown in Figure 34. In 3D, slope and dip are unsigned magnitudes corresponding to the maximum slope or dip in the direction the azimuth gives (Figure 35). Azimuth is the angle measured clockwise from the geographic north of the downdip direction of maximum slope or dip (Figure 36) and is given by

$$\phi = arctan\left(\frac{p_x}{p_y}\right) + \phi_o, \qquad Eq. (3)$$

Where ϕ_o is the angle between the survey y-axis e and true north.

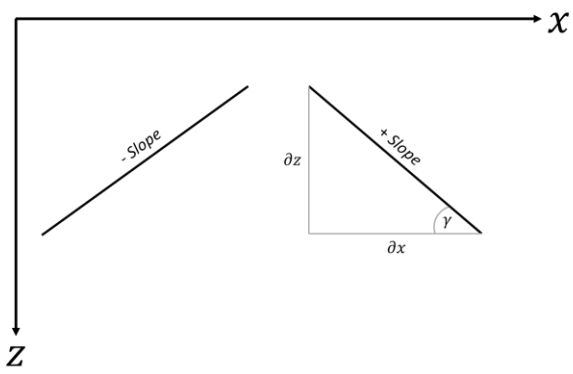

Figure 34: Illustration of the definition of the seismic attributes slope p, and dip γ and their sign conventions in a 2D case. The 3D slope components p_x and p_y follow the same convention. Modified from Barnes (2016).

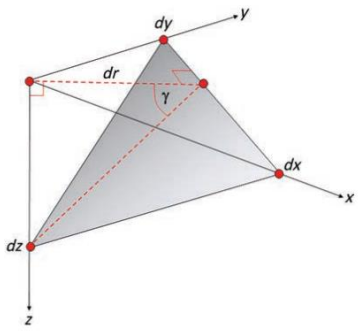

Figure 35: Illustration of the slope p, and dip γ and their sign conventions in a 3D case. From Barnes (2016).

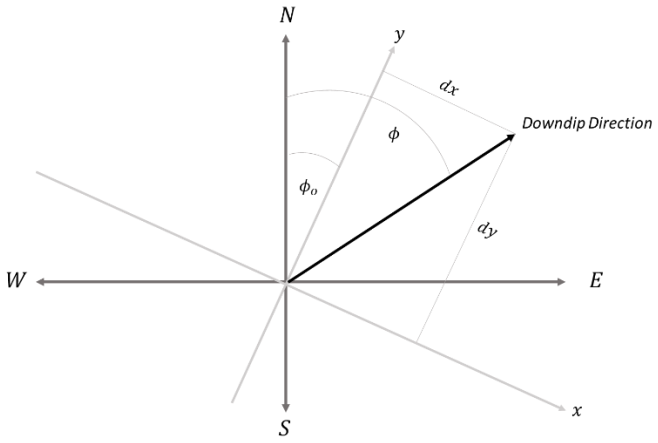

Figure 36: Schematic representation of the azimuth related to the survey's x and y coordinates and true north. Modified from Barnes (2016).

According to Gunther and Marfurt (2016) and Marfurt (2018), there were at least seven different usual approaches to estimating the dip-azimuth and dip magnitude in a seismic volume in the last decade. There were based on (i) derivatives of the phase trough

complex seismic trace analysis using Hilbert Transform, (ii) correlating events with of high energy and constant phase in the frequency domain using Fast-Fourier Transform (FFT), (iii) Principal Component Analysis (PCA) of the amplitude-field, (iv) dip-scanning using semblance method, (v) derivatives of the amplitude-field, (vi) iterative reconstruction of the dip-field using global optimization, (vii) or performing a least-square fitting of the wave equation using a differential operator (Plane-Wave Destructor Filter). Each of the methods mentioned above has its strengths and weaknesses. I explain more about the three methods that were evaluated in this study, which were the (i), (ii), and (iii) methods.

All the methods of dip-estimation are based on the fact that, locally, seismic reflectors tend to fit a plane wave with constant slopes (Figure 37). Therefore, for a 3D wavefield, u, with a constant slope, the wavefield can be approximated by:

$$u(x,\ y,\ z) \approx u\left(z - p_x x - p_y y\right) \qquad Eq.(4)$$

Therefore, what is generally done is running some slope-scanning in a small analysis window.

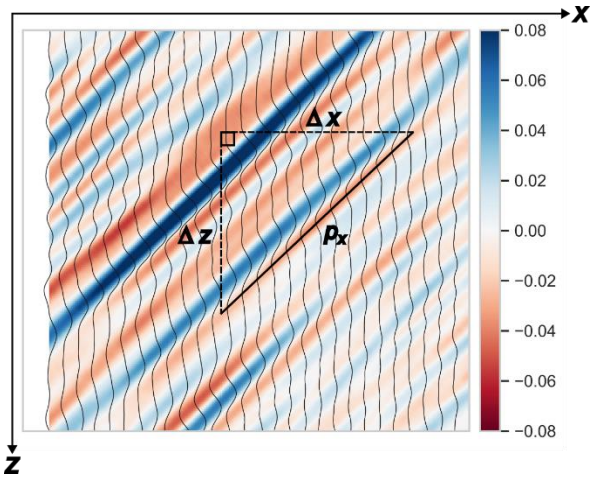

Figure 37**:** Within a limited analysis window, seismic reflection data closely resemble a plane wave characterized by a constant slope. In this two-dimensional scenario, the slope is represented as $px \sim \Delta t\ /\ \Delta x$.

4.1.1. Instantaneous Phase Gradient Dip Estimator

To introduce the dip estimator method based on the gradient of the instantaneous phase, we must define the instantaneous phase vector, k_i. It is the 3D counterpart of the

instantaneous frequency. It is defined as the gradient of the instantaneous phase divided by 2π:

$$k_i = \frac{1}{2\pi}\nabla\theta_i \qquad Eq.(5)$$

Where ∇ is the 3D gradient operator:

$$\nabla = \frac{\partial}{\partial x}\hat{x} + \frac{\partial}{\partial y}\hat{y} + \frac{\partial}{\partial z}\hat{z} \qquad Eq.(6)$$

In a seismic volume, at any given point, the direction of the greatest increase in phase is represented by the vector k_i. When dealing with time data, k_i combines horizontal wavenumbers and vertical frequency, resulting in horizontal units of inverse meters or feet and vertical units of hertz. On the other hand, when working with depth data, the phase vector becomes a pure wavenumber vector, measured in inverse meters or feet. Due to the orthogonal coordinate system of depth data, the phase vector is perpendicular to the local surface with a constant phase (Figure 38). This orthogonality provides mathematical and conceptual advantages, leading many discussions on dip and azimuth attributes to adopt orthogonal coordinate systems.

However, it is important to note that an orthogonal coordinate system is not mandatory for calculating slope and azimuth. In these cases, time data is sufficient. The instantaneous phase vector is expressed in terms of wavenumber components k_{xi} and k_{yi} in the x and y directions, respectively, and instantaneous frequency f_i in the z direction (Figure 39).

$$k_i = k_{xi}\hat{x} + k_{yi}\hat{y} + f_i\hat{z} \qquad Eq.(7)$$

Instantaneous wavenumber k_{xi} is defined as

$$k_{xi} = \frac{1}{2\pi}\frac{\partial}{\partial x}\theta_i \qquad Eq.(8)$$

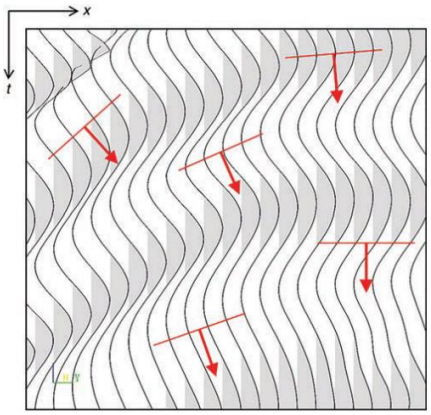

Figure 38: Phase Vector Orientation Principle. In seismic depth data or undistorted displays of seismic data in time, the 2D phase vector (illustrated by red arrows) maintains perpendicularity to the nearby path of constant phase and points downwards, indicating the steepest phase increase. In a three-dimensional context, the phase vector remains orthogonal to the nearby surface of constant phase. While this orthogonality concept supports intuitive understanding and mathematical formulations, it generally does not apply to seismic data in time. However, its application is unnecessary for deriving reflection slope and azimuth. From Barnes (2016).

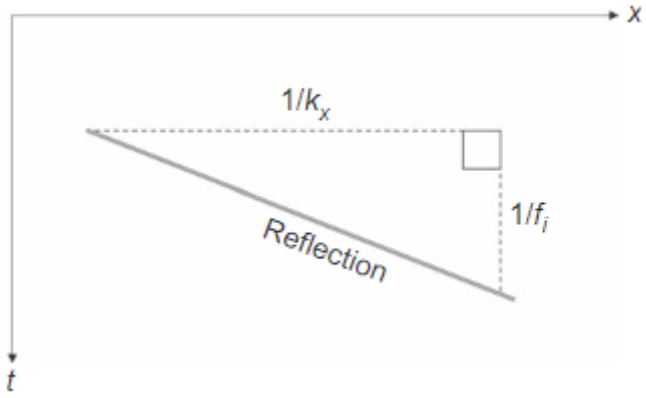

Figure 39: Two-Dimensional Reflection Slope. The reflection slope in a two-dimensional context is determined by the wavenumber component k_x and the instantaneous frequency f_i. In this scenario, the reflection slope exhibits positivity while the wavenumber k_x remains negative. In the three-dimensional realm, the slope incorporates the k_x component as well. From Barnes (2016)

Similar to instantaneous frequency, the instantaneous wavenumber cannot be directly computed from the instantaneous phase due to its lack of continuous differentiability. Instead, wavenumber k_{xi} is calculated using a different approach, defined by:

$$k_{xi} = \frac{1}{2\pi} \frac{u\frac{\partial v}{\partial x} - v\frac{\partial u}{\partial x}}{u^2 + v^2} \theta_i \qquad Eq.(9)$$

Where u represents the real 3D seismic data, while v represents the corresponding quadrature 3D data obtained by applying the Hilbert transform to the real seismic data.

The concepts of Hilbert transform, quadrature traces, and instantaneous attributes (phase, frequency) are defined in complex seismic trace analysis. For reasons of brevity, I will not define those, but more on 1D, 2D, and 3D complex seismic trace analysis can be found in Taner et al. (1979), Barnes (1996), and Barnes (2016).

The wavenumber k_{yi} is computed similarly to k_{xi}, as well as the vertical wavenumber k_{zi} for depth data. Both k_{xi} and k_{yi} wavenumbers are considered negative when measured in a downdip direction and positive when measured in an updip direction. Following the dip sign convention, the sign of wavenumber k_{xi} is opposite to the sign of the slope along the $x - axis$, denoted as p_x. Similarly, the sign of wavenumber k_{yi} is opposite to the sign of the slope along the $y - axis$, represented as p_y.

The instantaneous phase slope of a reflection in the $x - direction$, p_x, is given by:

$$p_x = -\frac{k_{xi}}{f_i} \qquad Eq.(9)$$

Which for time domain seismic data is measured in units of slowness. The negative sign is necessary to adhere to the dip sign convention (refer to Figure 40). The corresponding slope in the $y -$ direction, p_y, is obtained similarly.

Then, the instantaneous slope magnitude p is defined as:

$$p = \frac{\sqrt{k_{xi}^2 + k_{yi}^2}}{|f_i|} \qquad Eq.(10)$$

Instantaneous azimuth, denoted as ϕ, quantifies the dip direction. It is the angle in degrees from geographic north to the downdip direction of the reflection. It can be calculated using the following equation:

$$\phi = arctan\left[\frac{k_{yi}}{k_{xi}}\right] + 180° + \phi_0, \qquad Eq.(11)$$

Instantaneous dip γ is the angle in degrees between the reflection and the horizontal. In 3D, the instantaneous dip is unsigned and given by:

$$\gamma = arctan\left[\frac{\sqrt{k_{xi}^2 + k_{yi}^2}}{|k_{zi}|}\right] \qquad Eq.(12)$$

In 2D, the dip is signed according to the dip sign convention introduced before and becomes:

$$\gamma = -arctan\left(\frac{k_{xi}}{k_{zi}}\right) \qquad Eq.(13)$$

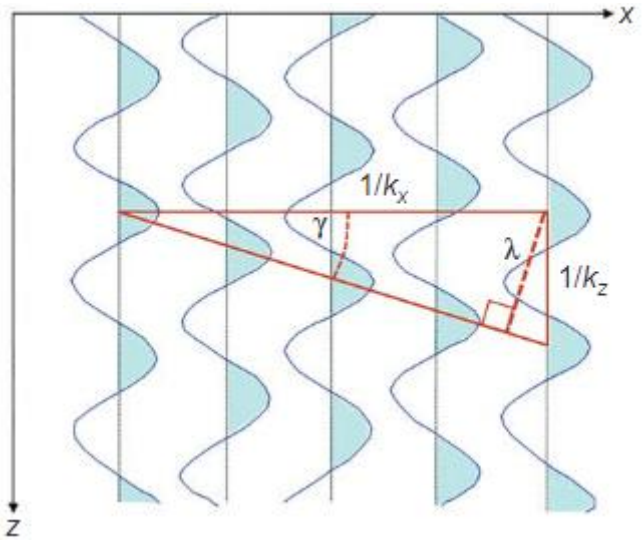

Figure 40: Dip concept in the context of depth data: The 2D instantaneous wavelength (λ) and dip (γ) are functions of the wavenumbers in the x and z directions, represented as k_x and k_z, respectively. λ represents the perpendicular distance between lines of constant phase. Similarly, in a three-dimensional context, the wavelength corresponds to the distance between planes of constant phase. From Barnes (2016).

4.1.2. Fast-Fourier Transform Dip Estimator

FFT steering was created by Tingdahl (1999) and became famous after Tingdahl and De Groot, 2003. It is by far the most complex algorithm defined here in this study. This is because the algorithm evolves three different main concepts in its steps. They are the Radon Transform, the Fast-Fourier Transform, and polynomial interpolation. Explaining in simple terms, the algorithm fits a third-order three-dimensional polynomial to a sub-cube around the sample of the highest energy in the Fourier domain. That

polynomial is then searched for its local maxima, and the dip and azimuth corresponding to that local maxima are set as output.

The 3D Radon Transform of a signal $u(x, y, z)$ can be expressed as:

$$R(u) = \breve{u}(p_x, p_y, \tau) = \int\int (x, y, \tau + p_x \cdot x + p_y \cdot y) dy dx \qquad Eq. (14)$$

Where p_x, and p_y are the slowness in the $x-$ and $y-$ direction, respectively (for time domain seismic data), and τ is the intercept time. The equation above can be expressed as a sum of the signal on lines:

$$\begin{cases} z = \tau + p_x x \\ z = \tau + p_y y \end{cases} \qquad Eq. (15)$$

The lines that intercept the z-axis at the intercept time τ with the angle α_x,

$$\begin{cases} p_x = \dfrac{dt}{dx} = tan(\alpha_x) \\ p_y = \dfrac{dt}{dy} = tan(\alpha_y) \end{cases} \qquad Eq. (16)$$

Since

$$\begin{cases} k_x = p_x f \\ k_y = p_y f \end{cases}, \qquad Eq. (17)$$

Moreover, the Fourier Transform of the same signal u is defined by:

$$F(u) = \hat{u}(k_x, k_y, f) = \int\int\int u(x, y, z) \cdot e^{-j(k_x x + k_y y + f z)} \, dx dy dz \qquad Eq. (18)$$

It can be shown that:

$$\hat{u}(k_x, k_y, f) = \hat{u}(p_x f, p_y f, f) = \hat{\tilde{u}}(p_x, p_y, \tau) = \int e^{-if\tau} \, d\tau \qquad Eq. (19)$$

This implies that the energy associated with various dips, denoted as, $\hat{\tilde{u}}(p_x, p_y, \tau)$, can be readily computed from the time-frequency representation of the signal using equation 19. Additionally, by utilizing the Radon Transform (Eq. 14), it becomes feasible to calculate the dip from the k-f spectrum, thereby providing a dip-interpretation to the k-

f spectrum. Applying the Radon Transform followed by the Fourier Transform in the seismic data has two main advantages for the dip estimation. The first is event enhancement: The Radon transform enhances coherent events, especially in the presence of noise and overlapping reflections, making identifying dips directly from the Radon-transformed data easier. The second is simplicity: Dip estimation becomes more straightforward as the peak frequencies directly correspond to the dip angles in the Radon-transformed data.

An illustration of the Radon Trasnform is depicted in Figure 41.

Using this idea as a basis, Tinghdal (1999) worked on a windowed algorithm of 7x7x7 samples to estimate the dip. Although, he acknowledged that such an approach has some weaknesses. For example, in the case of multiple dips in the window analysis, "which dip should be selected?" A simple average is not reasonable. Picking the maximum value seems to be the right answer. However, by simply picking the highest energy event from the k-f spectrum, the dip estimation would suffer from sampling limitation. That is, the maximum will be restricted to the sampled position of the discretized k-f domain. So, to run any estimation, he should interpolate the k-f domain. He tested three approaches in his work: third-order Lagrange interpolator, local-weighted mean, and normal distribution fitting. In his evaluation with synthetic data, the fitting of a normal distribution curve outperformed the other two. However, in his posterior works, he referenced using the Lagrange interpolator. The entire workflow is summarized in Figure 42.

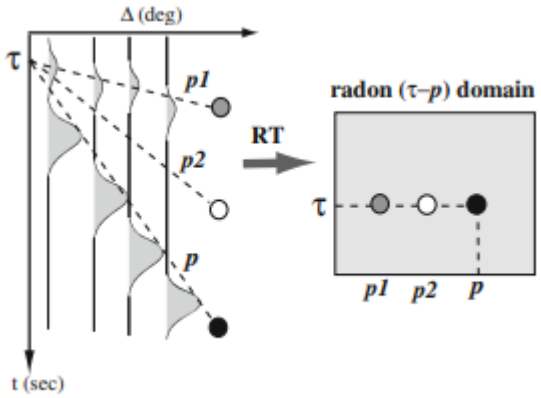

Figure 41: Illustration of the Forward Radon Process. The procedure involves stacking time-domain peaks along the ray parameter p, resulting in a concentrated energy focus in the Radon domain (depicted by the dark solid circle). Conversely, when stacking is done along another ray parameter, p2, Radon energy is

minimal due to significant mismatches with the travel-time slope of the primary arrivals. From Jeffrey Gu and Sacchi (2009).

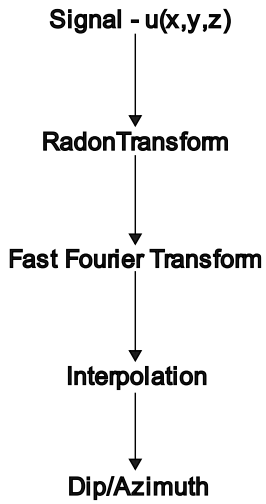

Figure 42: Illustration of the Workflow of the FFT-based dip estimation algorithm.

4.1.3. Gradient Structure-Tensor Dip Estimator

The third method is called the Gradient Structure-Tensor Method, and it was first proposed by Bakker et al. (1999) and Randen et al. (2000). The idea of this method is to determine the principal directions of amplitude variation, where one direction is the direction of maximum variation, normal to the reflection plane, while the other two have minimum variation and follow the reflection. In their work, they used PCA to build a 3x3 tensor from pre-computed derivatives of the amplitude. They estimate the derivatives convolving the seismic data with the derivative of a gaussian filter. In practical terms, to find, for example, the gradient in the x-direction, they convolve a Gaussian gradient operator, g_x, with the seismic data vector, u, obtaining the gradient vector G_x:

$$G_x = g_x * u(x, y, z) \qquad Eq.(20)$$

The same is done for the gradient vectors G_y and G_z. The full gradient tensor, G, is defined by the three gradient vectors of each dimension:

$$G = \begin{pmatrix} G_x \\ G_y \\ G_z \end{pmatrix} \qquad Eq.(21)$$

As pointed out by Randen et al. (2000), the estimation of the three-dimensional gradient vector serves as a representation of the local dip and azimuth of the data.

Nonetheless, it is susceptible to contamination from noise and other artefacts, thereby warranting caution in its interpretation. For instance, in the case of a horizontally oriented reflector exhibiting minor variations in amplitude, the vertically pointing gradient vector fails to portray the reflector's characteristics accurately. Instead, multiple gradient vectors emerge, pointing in diverse directions, as visually demonstrated in Figure 43. This and other cases will suffer from the so-called wrap-around effect, when and angle changes slightly, but its representation changes abruptly. For example, when we change from -180 to +180 degrees. One solution to deal with the noise and artefacts is to smooth the dip estimation. However, as much of the noise is linked to the wrap-around effect, the smoothing is not trivial. Randen et al. (2000) proposed to estimate the principal direction of the gradient vector instead of using a simple average filter.

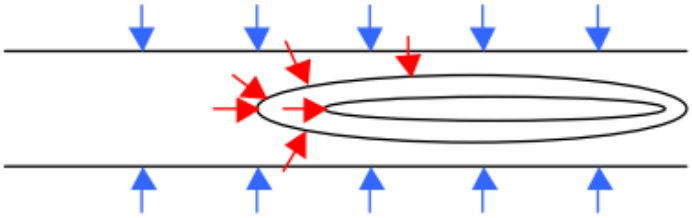

Figure 43: Gradient Estimations for a Horizontal Reflector with Local Amplitude Variation. The depicted black lines indicate contours of reflection amplitude, while the arrows represent gradient vectors. From Randen (2000).

To estimate these directions, they first cross-correlate the gradients, which means calculating the outer product of the full gradient with itself and then producing the gradient-structure tensor:

$$G \cdot G^t = \begin{vmatrix} G_x^2 & G_x G_y & G_x G_z \\ G_y G_x & G_y^2 & G_y G_x \\ G_z G_x & G_y G_y & G_z^2 \end{vmatrix} \qquad Eq.(22)$$

The tensor elements can be locally averaged to further suppress noise and produce more reliable estimations of the gradients than any average filter. However, such an operation implies a loss of resolution. They proposed using windowed Gaussian low-pass filters to better control this trade-off between noise suppression and loss of resolution. Finally, the last step is to use PCA to compute the eigenvectors and eigenvalues of the tensor. The first eigenvector e_1, associated with the largest eigenvalue, gives the direction of the highest gradient, which, as stated earlier, is the direction normal to the reflection

plane. By imposing the constraint that the eigenvector e_1 points in a downward direction and expressing its components in the x, y, and z axes as e_x, e_y and e_z respectively, the slopes p_x and p_y can be determined as follows:

$$p_x = \frac{e_x}{e_z}, \qquad Eq.(23)$$

And

$$p_y = \frac{e_y}{e_z} \qquad Eq.(24)$$

Figure 44 illustrates the definition following the constrains above.

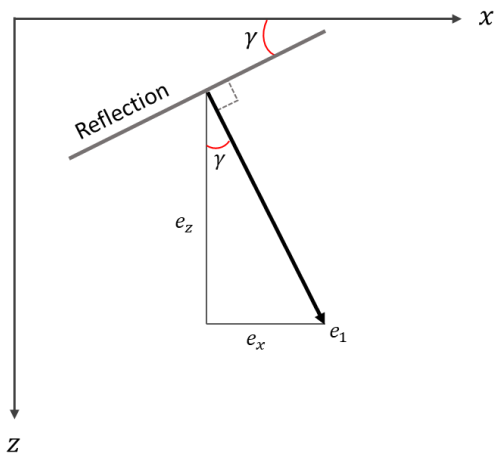

Figure 44: Reflection slope p_x and dip γ are found from the components e_x and e_y of eigenvector e_1, which is normal to the reflection. If the magnitude of the eigenvector e_1 is 1, then $dip = arccos\ e_z$. Modified from Barnes (2016).

Besides the robust estimation of the seismic gradient, or slope, another advantage of this approach is the possibility of combining the information of the eigenvectors to produce other seismic attributes linked to the reliability of the dip estimation and the "texture" of the seismic amplitudes. However, such attributes are out of the scope of this study.

4.2. Curvature

Curvature is a family of geometric attributes widely used for structural evaluation (Chopra, 2009; Jesus et al., 2019; Correa et al., 2021). This attribute is defined as the deviation of a surface, in a particular point, from a straight line. Mathematically, it is

related to the second derivative of a curve (Barnes, 2016). In two dimensions (2D), curvature, k, is the inverse of the circle's radius that tangentially fits a surface (Figure 45). The derivative $d\omega/ds$ represents the 2D curvature, obtained by converting the arc length derivative (d/ds) to the x-axis derivative (d/dx) using the chain rule. The angle ω is the arctangent of the 2D dip $\partial y/\partial x$, and its derivative is computed with the chain rule. Notably, curvature, except for the flat dip case ($\partial y/\partial x = 0$), is not simply the second derivative d^2y/dx^2; it is adjusted for a local dip, ensuring dip invariance.

$$K = \frac{1}{R} = \frac{d\omega}{ds} = \frac{\frac{d^2y}{dx^2}}{\left[1 + \left(\frac{dy}{dx}\right)^2\right]^{3/2}} \qquad Eq.(25)$$

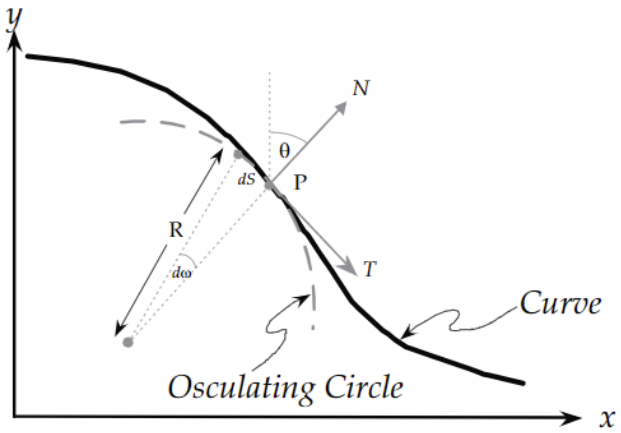

Figure 45: Illustration of the 2D definition of curvature. A point P on a curve is considered. Curvature is explained using the radius of the osculating circle (R), which touches the curve and shares tangent T with it at P. Vector N is perpendicular to the curve, defining local dip angle θ. The curvature at P is the reciprocal of R. From Roberts (2001).

In 3D, as stated by Marfurt (2018), curvature involves calculating second derivatives of a real or implied surface. For horizon-based curvature (e.g., Roberts, 2001), it is performed a local fitting of a second-degree polynomial given by:

$$z(x,y) = a\frac{\partial^2 z}{\partial x^2} + b\frac{\partial^2 z}{\partial x \partial y} + c\frac{\partial^2 z}{\partial y^2} + d\frac{\partial z}{\partial x} + e\frac{\partial z}{\partial y} + f \qquad Eq.(26)$$

Generally, nine or more grid points are used to compute the curvature at the picked horizon (Figure 46). Using more points leads to a smoother curvature estimate, as Rich

(2008) demonstrated. The curvature calculation involves the computation of eigenvalues (k1 and k2) and eigenvectors from a Hessian matrix given by:

$$k_{1,2} = \frac{cde - a(1+e^2) - b(1+d^2) \pm (\alpha - \beta)^{1/2}}{(1+d^2+e^2)^{3/2}}, \qquad Eq.(27)$$

Where:

$$\alpha = [a(1+e^2) - b(1+d^2)]^2, \qquad Eq.(28)$$

And:

$$\beta = [2bde - c(1+e^2)][2ade - c(1+d^2)] \qquad Eq.(29)$$

The corresponding eigenvectors (defining the azimuth perpendicular to a curvature lineament) are:

$$tan\theta_{1,2} = \frac{k_{1,2} - cde + 2b(1+d^2)}{2ade - c(1+d^2)} \qquad Eq.(30)$$

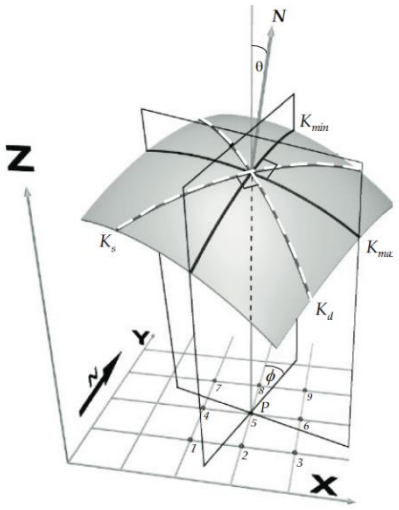

Figure 46: Tridimensional Curvature Analysis. The axes denoted as X and Y correspond to the spatial dimensions, while Z represents the temporal or depth dimension. The intersection of two perpendicular planes with the surface is highlighted, delineating the maximum curvature, Kmax, and the minimum curvature, Kmin. Additionally, the surface showcases two other orthogonal normal curvatures: the dip curvature, Kd, and the strike curvature, Ks. The vector N signifies the surface is normal at point P, forming an angle θ with the vertical axis, referred to as the dip angle. The orientation of any normal curvature can also be extracted, with the angle ϕ being the orientation of minimum curvature. From Roberts (2001).

In which direction we measure the surface deformation defines the curvature measurement. The principal curvature directions are the k1 and k2 curvatures, most-positive and most-negative, respectively.

We can obtain other curvature attributes from this pair of curvatures, such as maximum and minimum curvature, which are the maximum and minimum values in modulus between k1 and k2:

$$k_{max} = \begin{cases} k_1 \; when \; |k_1| \geq |k_2| \\ k_2 \; when \; |k_1| < |k_2| \end{cases}$$
$$k_{min} = \begin{cases} k_1 \; when \; |k_1| < |k_2| \\ k_2 \; when \; |k_1| \geq |k_2| \end{cases} \qquad Eq.(31)$$

There are also the mean curvature, gaussian curvature, and curvedness. The definition of mean curvature is straightforward; the Gaussian curvature, k_{gauss}, is the product between k1 and k2:

$$k_{gauss} = k_1 \cdot k_2 = k_{max} \cdot k_{min} \qquad Eq.(32)$$

The curvedness, C, is the norm of the tensor defined by the k1 and k2 components (Roberts, 2001; Chopra, 2018):

$$C = (k_1^2 + k_2^2)^{1/2} \qquad Eq.(33)$$

Other curvature measurements defined by directions other than k1 and k2 are dip curvature, strike curvature, and Euler (apparent) curvature. Despite this vast number of curvature attributes, several authors highlighted k1 and k2 as the most useful, especially for fault and fracture inference. (e.g., Chopra and Marfurt, 2007b; Bravo and Aldana, 2006; Barnes, 2016; Marfurt, 2018).

In practice, curvature attributes analyze how much a surface is deformed at a given point. They are generally applied to locate dips and structural highs or lows. For example, in 2D, the curvature is zero if the seismic event is flat, positive for an antiform structure, and negative for a synform (Roberts, 2001) (Figure 47).

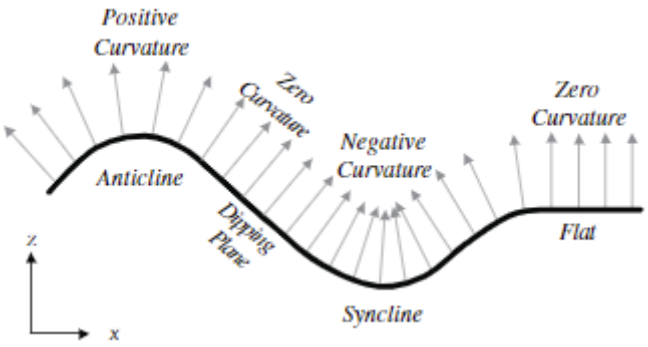

Figure 47: Curvature Attribute Sign Convention. The grey arrows depicted illustrate vectors that are perpendicular to surfaces. The curvature is set to zero on flat or planar-dipping surfaces where these vectors run parallel. Positive curvature is assigned over anticlines, where the vectors diverge, while negative curvature is assigned over synclines, where they converge. From Roberts (2001).

In 3D, to define a surface, we need to have the information of two curvature directions, generally k1 and k2. Figure 48 illustrates the main shapes defined by the values of k1 and k2.

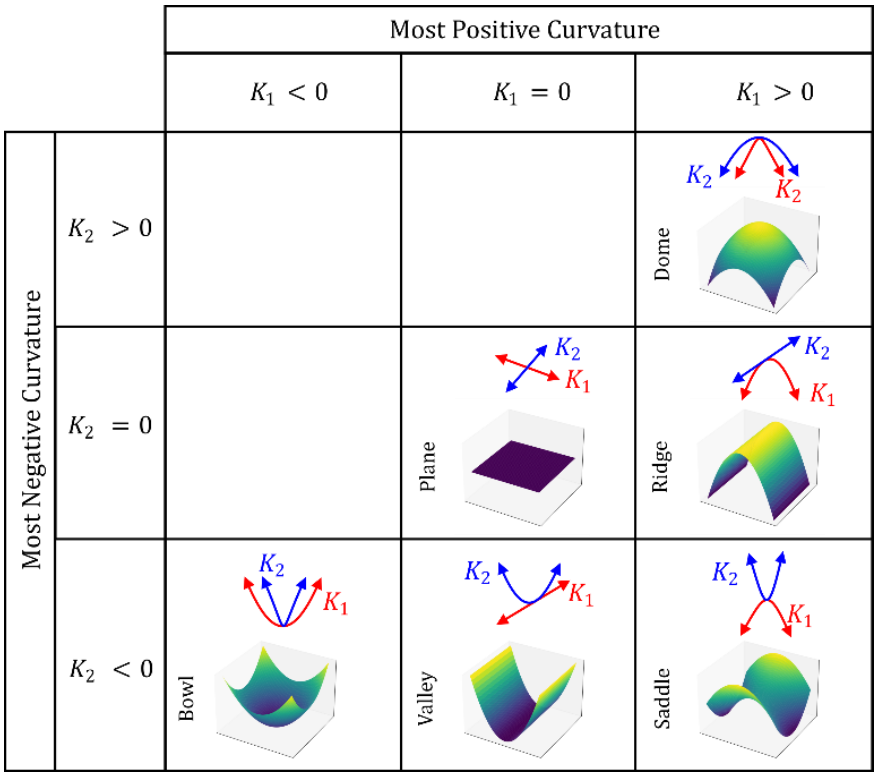

Figure 48: Surface shapes defined by the most-positive (k1) and most-negative (k2) curvature for each case of combined features. When both k1 and k2 are positive, the surface is dome-shaped; a positive positive k1 and a k2 equal to zero define a ridge; a positive k1 and a negative k2 define a saddle. In cases where k1 is zero, the surface is a plane when k2 is also zero, and a valley when k2 is negative. When both k1 and k2 are negative, the surface shape is a bowl. After Roberts (2001).

All the definitions of curvature presented above are related to picked horizons or gridded surfaces. Usually, when working with 3D seismic data, we prefer to compute curvature directly from the seismic volume. Compared to the volumetric-based, horizon-based curvature lacks resolution and is prone to artefacts introduced by erroneous interpretation and gridding interpolation (Roberts, 2001; Klein et al., 2008). Volumetric-based curvature computation is implemented in several ways (Al-Dossary and Marfurt, 2006; Klein et al., 2008; Machado et al., 2013; Di and Gao, 2014; Di and Gao, 2015). Almost all of them are based on a locally implicit reconstruction of a quadratic surface using the information of estimated dip volumes. They use p_x and p_y to approximate the derivatives in equation 26, changing the equation for:

$$z(x,y) = a\frac{\partial p_x}{\partial x} + \frac{b}{2}\left(\frac{\partial p_x}{\partial y} + \frac{\partial p_y}{\partial x}\right) + c\frac{\partial p_y}{\partial y} + dp_x + ep_y + f \qquad Eq.(34)$$

The main difference between these implementations is how to calculate the first and second derivatives. Another difference is that some of those works addressed the problem that the general implementation of curvature is well-behaved for horizontal or semi-horizontal structures. The standard approach will start to fail for complex structures with steeply dipping events, affecting the curvature values. Therefore, some works, such as Machado (2013) and Di and Gao (2015), rotate the fitted surfaces horizontally before computing the curvature. However, this introduces minor changes compared to the first aspect, estimating the dip and its derivatives. Therefore, one should look carefully at the dip estimation workflow when evaluating a curvature attribute. This would make the most impact on the final curvature attribute.

Another concerning point is the set of parameters that controls the investigation scale. Curvature attributes can be used to identify micro- to macro-scale structural features (Bergbauer et al., 2003; Al-Dossary and Marfurt, 2006; Chopra and Marfurt, 2013). Therefore, as in Hunt et al. (2018), one should correctly consider the target's depth and scale to parametrize the attribute window properly. Another alternative to isolating features of different scales is through prior spectral filtering, either of the amplitude or dip volumes (Bergbauer et al., 2003; Al-Dossary and Marfurt, 2006).

4.2.1. Structural and Amplitude Curvatures

The seismic signal can be decomposed in terms of amplitude and phase components. As a derivative-based attribute of the seismic trace, the curvature can use either of the components to calculate the derivatives.

Structural curvature is the classic curvature attribute. It is based on horizontal derivatives of the dip field, commonly obtained by the phase gradient information. In conclusion, it highlights the phase changes and not the amplitude.

Amplitude curvature is based on horizontal second derivatives of the amplitude field. It can be obtained by different means, such as Laplacian operators, principal component analysis (PCA), Fourier transforms, and others (Chopra and Marfurt, 2013). Also, it can be calculated using different amplitude-related data as input. For example, Guo et al. (2010) computed the amplitude curvature on an impedance acoustic volume obtained by model-based inversion. The authors attested that the amplitude curvature on acoustic impedance has greater resolution and sharper responses when compared to seismic amplitude. Since amplitude curvature is sensitive to changes in amplitude, it can be used to identify diagenetic features that cause perturbations in amplitude, such as karsts or cementation.

4.2.2. Curvature for Fracture Characterization

Since the last century, curvature has been linked to increased fracture intensity (Harris et al., 1960; Naar, 1991; Lisle, 1994). Over the past two decades, curvature attributes have primarily been employed for qualitative structural characterization and seismic geomorphology studies and have found significant utility in various geological contexts (Roberts, 2001; Mandujano et al., 2005; Al-Dossary and Marfurt, 2007; Hart and Sagan, 2007; Mai et al., 2009; Guo et al., 2010; Sun et al., 2012; Gao, 2013; Kosari et al., 2015). Moreover, research endeavours have also focused on quantitative analyses using seismic data, digitized outcrops, and "sand-box models" to assess fracture direction and intensity through curvature attributes (Hennings et al., 2000; Allwardt et al., 2007; Keating and Fischer, 2008; Pearce et al., 2011; Hunt et al., 2010; Hunt et al., 2011; Staples and Marfurt, 2013; Chopra et al., 2013; Di and Gao, 2017; Hunt et al., 2018; Watkins et al., 2018).

Hennings et al. (2000) performed curvature and fracture analysis in a reconstructed 3D model of the Oil Mountain Anticline, Wyoming. In their study, they found that gaussian curvature and dip's change rate have a strong correlation with the fracture intensity, even stronger than kinematics-restored attributes such as inclined shear (Buddin et al., 1997) and flexural-flattening strain (Williams et al., 1997).

Allwardt et al. (2007) used GPS data to estimate curvature in Sheep Mountain Anticline, Wyoming, correlating it with fracture analysis results. Their results indicated that higher curvature correlates with greater spherical variance, a fracture analysis metric related to the dispersion of the fracture strikes from the main trend. While they did not find a consistent curvature-fracture intensity correlation, they observed that higher curvature areas have higher fracture intensity, mainly in the fold's hinge line. However, limb areas, which have low curvature, eventually also show significant fracture intensity. Although, they observed that the fractures in the limb were most likely generated from mechanisms other than faulting and folding.

Pearce et al. (2011) used LiDAR data to analyze fractures and curvature in four small UK anticlines. They found no consistent correlation between curvature and fracture intensity. Two folds showed dual fracture populations: one linked to fault and folding, eventually correlating with curvature, and another tied to processes like tectonic loading and unloading, not curvature related. Three folds exhibited a threshold curvature for high fracture intensity; beyond this value, rocks consistently had high fractures. Similarly, as in Allwardt et al.'s (2007) work, the issue lies in low-curvature zones with high fracture intensity.

The research conducted by Hunt and their team in 2010, 2011, and 2018 focused on the Western Canadian Sedimentary Basin. They discovered a significant connection between the volumetric seismic attribute k1 curvature, fracture intensity, and microseismic events. These studies also highlighted several crucial insights. For instance, they demonstrated that k1 curvature surpasses coherence in predicting fractures. Additionally, the studies emphasized the value of integrating curvature with other seismic attributes, such as anisotropic and elastic ones, to represent the fracture system comprehensively.

Staples and Marfurt (2013) evaluated the hypothesis of curvature-fracture intensity correlation using a small-scale clay model. They conclude that k1 curvature

correlates with fracture intensity like in a 3-phase elastic deformation. Moreover, this pattern occurs both in compressional and extensional settings. Later, Chopra et al. (2013) tested these findings in seismic data from the Oklahoma field, Anadarko Basin. Among seven wells evaluated in their study, k1 curvature presented a high correlation (>0.5) with fracture intensity in six of them. Also, they could define critical curvature magnitudes where the correlation is stronger. Below these critical values, the fracture intensity would be too low to be correlated with the curvature.

Finally, a notable observation was made in a study by Watkins et al. (2018) regarding the correlation between curvature and fracture intensity in four anticlines within the Achnashellach Culmination located in NW Scotland. They attributed this observation to the area's extensive history of deformation. It was found that the hinge zones, characterized by high curvature values, exhibited higher fracture intensity. However, it was also noted that elevated values were present in both the backlimb and forelimb. This suggests that in instances where low curvature zones are present, fractures may form along paleo-hinge lines due to the rolling-hinge process associated with thrust belts or other processes unrelated to folding.

In summary, all these studies help to improve our understanding of how the fault-and-fold-related strain process translated into curvature and how this is linked to fracture development. However, many of these studies also remind us that not all faults and folds are equal and that not every fracture is related to this process. Also, scale is crucial. Therefore, when performing curvature analysis on seismic data, one should always try to understand the geological context of the area, the scale of the problem, and, when possible, aggregate other attributes.

4.3. Coherence

Coherence is a commonly used attribute in the workflows for seismic facies identification. This attribute was developed in the early 1990s to quantify the similarity between neighbouring features (Bahorich and Farmer, 1995). It compares the current trace with neighbourhood traces, analyzing whether there is any coherence between the amplitude values.

Since the objective of the coherence attribute is to evaluate the similarity between seismic traces, values from 0 to 1 are used to determine similarity. A discontinuous event

has a coherence value of 0, and continuous events have a value of 1. Thus, a coherence map is used to map geological features such as channels, faults, fractures, and stratigraphic contacts. Two standard approaches to calculate it are by the cross-correlation between two seismic traces or via a semblance panel.

Marfurt et al. (1998) developed a semblance-based based on the coherence to identify structural and stratigraphic discontinuities such as faults and build-ups. Semblance discontinuity measures the variance of a set of traces around the average trace, normalized by the average energy (Barnes, 2016). It is a coherence attribute sensitive to lateral changes in amplitude. Mathematically, the semblance-based coherence is the energy ratio of the average trace to the total energy in a set of M traces:

$$C_{semb} = \frac{ME_a}{E}, \qquad Eq.(35)$$

Where E_a is the energy of the average trace, and E is the total energy.

The similarity attribute (Tingdahl et al., 2001) measures the Euclidian distance between two vectors trace in the hyperspace. It is defined as one minus the Euclidian distance. Therefore, a high similarity value means that both vectors are 'near' each other in amplitude and shape, thus similar. Low values imply the opposite case. In low-frequency bandwidth and low signal-noise ratio, amplitude sensitiveness can be a problem.

One of the disadvantages of any raw implementation of coherence attribute is the lack of accuracy in dipping reflectors. That is because the algorithm does not respect the local dip. So, without implementing dip-steering techniques (Tingdahl and De Groot, 2003), the attribute will identify dipping reflectors as discontinuous zones.

4.4. Fault Enhancement

Many techniques are often used to improve the resolution of the fault planes in the seismic. Usually, these techniques are based on filters that act as fault enhancement attributes. Some examples and definitions are shown in Francelino and Antunes (2013), Odoh et al. (2014), Jibrin and Yelwa (2015), Machado et al. (2015, 2016), Dewett and Henza (2016) and Yu (2016).

In this study, I used a preconditioning filter for fault enhancement. The filter is called an Edge-Preserving Smoothing Filter (EPS). It is a type of Recursive Gaussian Filter (e.g., Deriche, 1993; Hale, 2002) that considers the information of the Fault-Likelihood volume ($1 - coherence^8$) (Hale, 2013). The output resembles results from the literature that used LoG (Laplacian of Gaussian) operators or Kuwahara filters to enhance faults in seismic images (Chopra and Marfurt, 2017). The basic concepts of edge-preserving smoothing filters and their possible applications are well addressed in previous works such as Luo et al. (2002) and AlBinHassan et al. (2006).

5. Machine Learning

This chapter aims to provide an overview of the machine learning techniques used in the study and explain the rationale behind their selection. It begins by introducing the basic concepts of machine learning and convolutional neural networks and reviewing previous works that used them for automatic fault identification. Next, it explores regression methods, which were employed to understand the relationship between fracture intensity and geometric seismic attributes using analytical models, as well as to estimate fracture intensity away from the wells. Finally, it discusses hyperparameter tuning and feature importance, two critical aspects of machine learning that enhance the performance and interpretability of the models. Each topic presents the advantages and disadvantages of the techniques used, providing a comprehensive understanding of the methodology applied in this study.

5.1. Concepts on Machine Learning

Machine Learning is a prominent subfield of artificial intelligence concerned with developing and applying algorithms and models capable of autonomously acquiring knowledge and making predictions or decisions without explicit programming. It leverages statistical techniques and data to facilitate the training of computational systems, enabling them to execute specific tasks or achieve accurate predictions. A notable subset of machine learning is deep learning, which draws inspiration from the structure and functionality of the human brain (LeCun et al., 2015) (Figure 49).

Both machine learning and deep learning domains have a serious of algorithms designed to solve a wide variety of problems. Each one of these algorithms learns to solve a task in a unique way. There are some attempts at categorising these algorithms concerning their type of learning. A common categorization is dividing the machine learning algorithms into supervised, unsupervised, and reinforcement learning. There are other categorizations that introduce more types of learning algorithms, such as self-supervised, semi-supervised, and others. Although, in geoscience, most of the algorithms employed are in the realm of supervised and unsupervised learning. Therefore, I will define only these two types of learning algorithms (Figure 50).

Supervised learning constitutes a machine learning paradigm wherein algorithms are trained using labelled data. Labelled data refers to input instances with their correct

output or target value. The learning process entails the algorithm assimilating knowledge from these exemplars and constructing a predictive model capable of accurately forecasting outcomes or classifying novel, unseen data. This objective is accomplished by minimizing the disparity between predicted and actual targets through an optimization procedure, generally backpropagation with gradient descent (I will define some of those terms later).

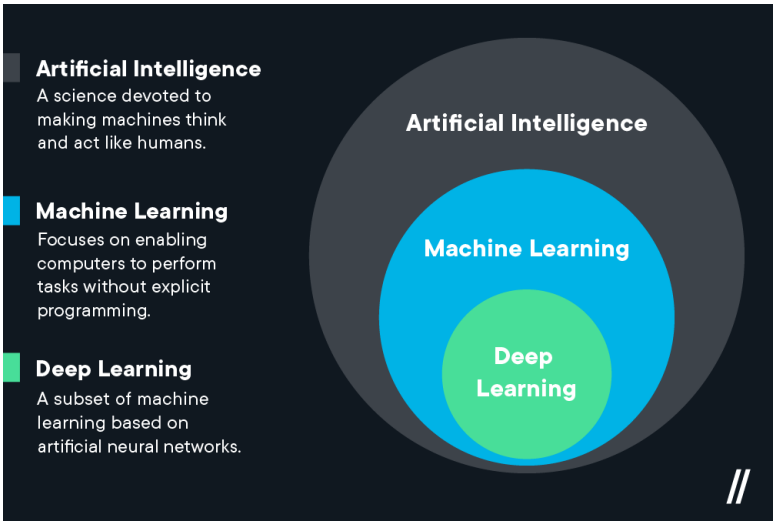

Figure 49: Illustration of the relationship between the fields of Artificial Intelligence, Machine Learning, and Deep Learning. Available at https://flatironschool.com/blog/deep-learning-vs-machine-learning/.

x

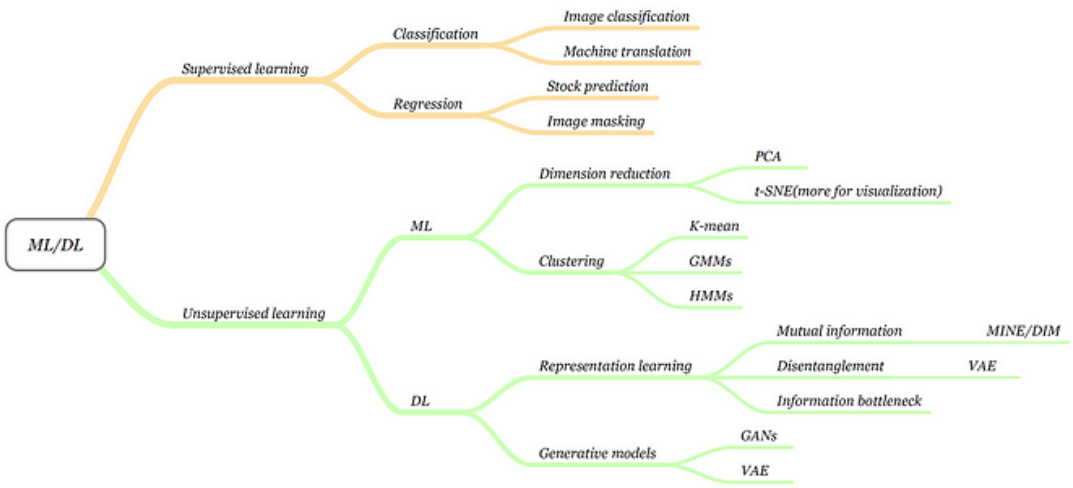

Figure 50: Machine Learning Algorithms Tree. https://pub.towardsai.net/machine-learning-algorithms-for-beginners-with-python-code-examples-ml-19c6afd60daa.

Both machine learning and deep learning domains have a serious of algorithms designed to solve a wide variety of problems. Each one of these algorithms learns to solve a task in a unique way. There are some attempts at categorising these algorithms concerning their type of learning. A common categorization is dividing the machine learning algorithms into supervised, unsupervised, and reinforcement learning. There are other categorizations that introduce more types of learning algorithms, such as self-supervised, semi-supervised, and others. Although, in geoscience, most of the algorithms employed are in the realm of supervised and unsupervised learning. Therefore, I will define only these two types of learning algorithms (Figure 50).

Supervised learning constitutes a machine learning paradigm wherein algorithms are trained using labelled data. Labelled data refers to input instances with their correct output or target value. The learning process entails the algorithm assimilating knowledge from these exemplars and constructing a predictive model capable of accurately forecasting outcomes or classifying novel, unseen data. This objective is accomplished by minimizing the disparity between predicted and actual targets through an optimization procedure, generally backpropagation with gradient descent (I will define some of those terms later).

Regarding the difference between deep learning and machine learning, one can say that deep learning is a specialized domain that concentrates on employing larger models to solve intricate problems. Deep learning is generally related to big data problems and high-performance computing (HPC), where using more complex architectures demands huge computational power to deal with the vast amount of data and complex relationships between the variables. One example of those architectures is deep neural networks. Neural networks consist of interconnected layers of artificial neurons that emulate the cognitive processes observed in the human brain. The perceptron is at the core of neural networks, a basic computational unit that processes inputs and produces an output based on weighted connections and an activation function (Rosenblatt, 1958) (Figure 51). Shallow neural networks, also known as single-layer perceptrons, contain only one layer of perceptrons, limiting their ability to capture complex relationships in data.

As stated above, deep learning is known for using larger architectures, for example, with deeper neural networks, such as multi-layer neural networks or Multi-

Layer Perceptron (MLP). These networks contain multiple hidden layers between the input and output layers (Figure 52). Through the iterative adjustment of parameters, such as weights and biases, and non-linear activation functions, deep neural networks learn to capture complex relationships in the data, enabling them to make accurate predictions or decisions. The advancement of deep neural networks has revolutionized fields such as computer vision, natural language processing, and speech recognition, driving breakthroughs in artificial intelligence research (Bengio et al., 2013; Goodfellow et al., 2016). Neural Networks, particularly deep neural networks, serve as fundamental components within various machine learning and deep learning algorithms.

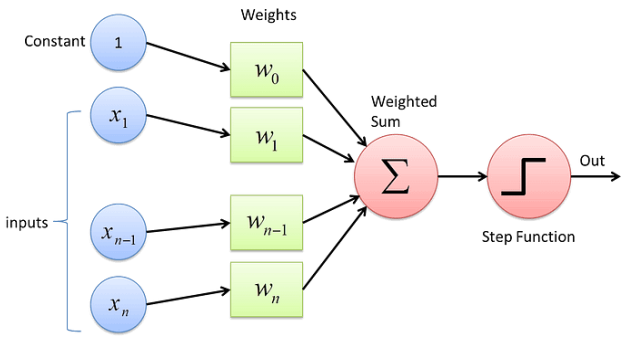

Figure 51: Illustration of a single perceptron's scheme. It used weights and bias to adjust to learn from the inputs and produce a desirable outcome. Available at https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53

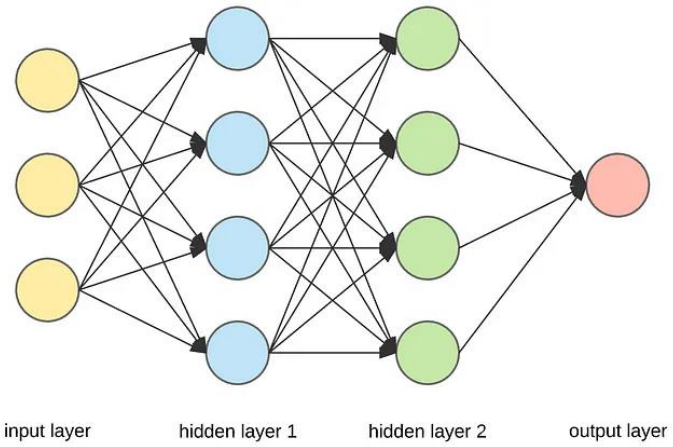

Figure 52: Illustration of a Multi-Layer Perceptron (MLP) network architecture. It has more than one hidden layer. Generally, it is fully connected, meaning it connects all the neurons from one layer to another. Available athttps://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6

5.2. Convolutional Neural Networks

Convolutional neural networks (CNNs) are deep learning algorithms commonly used for image recognition, classification, and segmentation (LeCun et al., 2015). CNNs consist of multiple convolutional and pooling layers, followed by one or more fully connected layers that make predictions based on the extracted features.

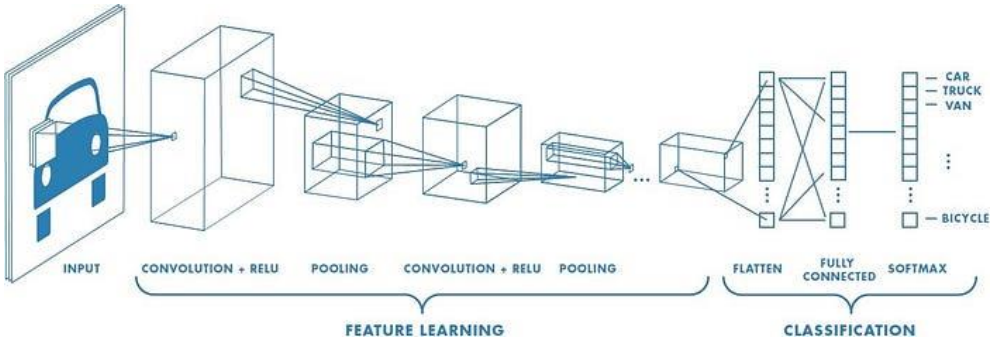

Figure 53: Abstract scheme of the architecture of a fully connected convolutional neural network designed to classify several types of vehicles. In these types of applications, these networks have a series of convolutional layers combined with ReLU activation, followed by a pooling layer to downsample the input image. In the end, they generally have a fully connected network followed by a softmax function to deliver a probability of the predicted class. Available at https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

CNNs were originally proposed in the 80s by works such as Fukushima (1980), Homma et al. (1987), Waibel et al. (1989), and LeCun et al. (1989). However, they received different names at that time, such as neocognitron, shift-invariant neural networks, time-delay neural networks, and other variants. Only in the 90s, after the LeNet-5 model created by LeCun et al. (1998), the term “convolutional neural networks” became publicly adopted. Nevertheless, it took two more decades for CNNs to achieve state-of-the-art performance in image classification and segmentation. The AlexNet model designed by Krizhevsky et al. (2012) outperformed previous standard image recognition methods in the ImageNet competition. For example, compared with the previous winner's solution in the competition, AlexNet represents a 40% improvement.

One of the secrets of this improvement relies on the convolutional filters (or kernel filters). Through them, a CNN learns to extract image features. Those features can be low-level features such as contours, edges, angles, and colours or high-level features such as items, scenarios, interactions, and general context (Figure 54). A CNN identifies both levels of features and learns how to combine low-level features into high-level ones to

reconstruct the image (Simonyan and Zisserman, 2014). Ultimately, this hierarchical learning capacity of a CNN allows it to capture complex patterns and relationships in the image data (Zeiler and Fergus, 2014), such as objects and scenes (Girshick, 2015; Long et al., 2015). This is one of the major strengths of CNNs.

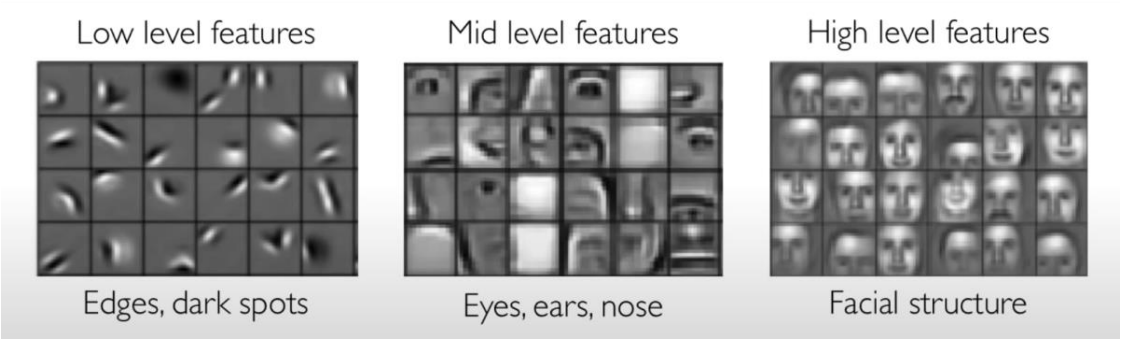

Figure 54: https://www.knime.com/blog/convolutional-neural-networks-computer-vision

If we compare the CNNs to multi-layer perceptrons, the former go-to deep learning algorithm in the recent-past, MLPs struggle to handle spatial information. Also, MLPs cannot learn hierarchical representations of data (Goodfellow et al., 2016). Compared to MLPs, which typically have only a few layers and are limited in their ability to learn complex data representations, CNNs can stack multiple convolutional layers to learn increasingly complex and abstract features. Additionally, CNNs can reduce the number of parameters required to train a model by using pooling layers to downsample feature maps, which reduces the computational requirements without sacrificing accuracy (Szegedy et al., 2015). However, CNNs still being a computationally expensive approach that requires powerful GPUs (or CPUs) and may be susceptible to overfitting if the number of layers is too high, where the model becomes too complex and performs poorly on unseen data (Goodfellow et al., 2016). Despite these limitations, CNNs remain an active area of research in deep learning.

The mathematics behind the convolutional neural networks can be resumed in the convolutional process, ReLu Activation, the pooling operator, a final activation function, and a loss function optimized through backpropagation. For image segmentation problems, generally, the activation function and the loss-function are the softmax and cross-entropy (or binary cross-entropy for binary classification), respectively.

The convolution operation is used to apply a filter to an input image. The filter is a small matrix of weights that slides over the image. At each position, the dot product of

the filter and the image patch under the filter is computed. The result is a feature map that highlights patterns in the input image that the filter is looking for: The analytical convolution operator between two functions f and g is defined as $h(t)$:

$$h(t) = (f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \qquad Eq.(36)$$

Where $*$ denotes the convolutional operator, t and τ are he independent variables of the functions f and g, respectively. The integral is taken over the entire real line, and the convolution result is a new function of the independent variable t.

However, in computational tasks such as machine learning, the convolutional operator acts in the discrete domain. Therefore, the numerical convolution of a 2D filter with a 2D image is defined as:

$$h_{i,j} = \sum_{m=0}^{M-1}\sum_{k=0}^{N-1} w_{m,n} \cdot x_{i+m,j+n} \qquad Eq.(37)$$

where $h_{i,j}$ is the output value at the position (i,j) in the feature map, $w_{m,n}$ is the weight at position m,n in the filter, $x_{i+m,j+n}$ is the input ("pixel") value at the position $(i+m, j+n)$, and M and N are the dimensions of the filter.

The Rectified Linear Unit (ReLU) activation function introduces non-linearity into the network. It applies an element-wise activation to the output of the convolutional layer. It simply returns the input if it is positive and zero otherwise:

$$f(x) = max(0,x) \qquad Eq.(38)$$

Where x is the input of the activation function.

The max pooling operation is used to downsampling the feature maps by keeping only the maximum value within a small rectangular neighbourhood. This helps to reduce the number of parameters in the network and makes it more robust to small translations in the input. The equation for the max pooling operation is:

$$y_{i,j} = max_{m=0}^{l-1} max_{n=0}^{l-1} x_{i\times s+m, j\times s+n} \qquad Eq.(39)$$

Where $y_{i,j}$ is the output value at position i,j in the pooled feature map, $x_{i\times s+m,j\times s+m}$ is the input value at the position $(i \times s + m, j \times s + n)$, l is the length of the pooling window, and s is the stride.

The softmax function converts the network output into a probability distribution over the possible classes. It maps the network output to a set of probabilities that sum to one. The equation for the softmax function is:

$$Softmax(x_i) = p_i = \frac{e^{z_i}}{\sum_j^K e^{z_j}} \qquad Eq.(40)$$

Where p_i is the probability of the $i-th$ class, z_i is the network output for the $i-th$ class, and K is the total number of classes.

The cross-entropy loss measures the difference between the predicted and true probability distributions over the classes. The equation for the cross-entropy loss function is:

$$\mathcal{L} = -\sum_{i=1}^{K} y_i log(p_i) \qquad Eq.(41)$$

Where $\mathcal{L}$ is the cross-entropy loss, K is the total number of classes, y_i is the true probability of the $i-th$ (i.e., it is equal to 1 if the sample belongs to the $i-th$ class and 0 otherwise), and p_i is the predicted probability of the $i-th$ class.

In other words, the cross-entropy loss penalizes the model when it assigns low probability to the true class and high probability to the other classes. It is a widely used loss function in deep learning because it is easy to optimize and has proven to be effective in many classification tasks.

These are some of the key general equations and mathematical foundations behind a CNN. Many other equations are involved in the training and optimization of the network. Still, these are the main ones used in the forward pass of the network.

Finally, in the backward phase, the parameters of the network are optimized trough backpropagation by adjusting the weights and biases of the network based on the error of the output compared to the expected output. The algorithm computes the gradient of the loss function for the weights and biases of each layer using the chain rule of

calculus. This gradient is then used to update the weights and biases of each layer using an optimization algorithm such as gradient descent or Adam optimizer. For the reason of brevity, I will not define the derivatives of the layers. The gradient descent algorithm is defined later in another topic.

5.2.1. U-Net Architecture for Image Segmentation

Image segmentation divides an image into multiple segments or regions, each corresponding to a specific object or part of the image. One type of image segmentation is semantic segmentation, which assigns a class label to each pixel in the image (pixel-wise classification). Several CNN architectures are properly designed to perform pixel-wise classification. Among those, one that stands-out is U-Net. It is a type of CNN architecture introduced by Ronneberger et al. (2015). It has become a popular tool for semantic segmentation tasks. The U-Net was originally designed for biomedical image segmentation, where datasets typically have few training images. However, using image augmentation techniques, U-Net demonstrated an 11% improvement in accuracy over the second-best approach.

U-Net's architecture consists of three main components: the encoder, decoder, and skip connections (Figure 55). The encoder is responsible for extracting and localizing features in an image, which are used to predict the presence of objects in the image. It creates a lower-dimensional representation of the input image, extracting features using convolutional and pooling layers. This way, the encoder gradually reduces the spatial resolution of the input image while increasing the number of channels, allowing it to capture high-level features. However, the encoder output lacks location information, which is essential for segmentation. In contrast, the decoder reconstructs an image from the compact representation using a deconvolution layer while reducing the number of channels. The skip connections pass information from earlier convolutional layers to the deconvolution layers, allowing the network to retain information about the location of features in the image that would otherwise be lost during the downsampling process. U-Net's key to success lies in this part. The skip connections allow for accurate segmentations with fewer data, making it particularly useful for small datasets, such as medical image segmentation datasets, unlike autoencoder architectures, which require an immense amount of data to reconstruct images from compressed representations accurately.

Recent studies have also demonstrated the effectiveness of U-Net in geophysics-related tasks, including seismic interpretation and rock property estimation. For example, Yang et al. (2019) used U-Net to speed-up the process of seismic velocity-model building. Peter et al. designed a U-Net for automatically picking seismic horizons, a labour-intense activity. Another common application of U-Net and other CNN architectures is to automate salt-body detection (e.g., Sen et al., 2020), a critical step in seismic interpretation, especially as an input for seismic imaging. Ao et al. (2021) proposed an innovative workflow using a variant of a U-Net to estimate seismic curvature attributes. In another work aiming to automate seismic interpretation, Vizeu et al. (2021) designed a U-Net to identify intrusive igneous seismic facies in Santos' Presalt. They used sparse-training to allow the network to be trained with just a few examples of human-labelled seismic sections. Those are just some examples that help illustrate the power and flexibility of CNNs and U-Net architecture and how they can be used to achieve state-of-the-art performance in various tasks.

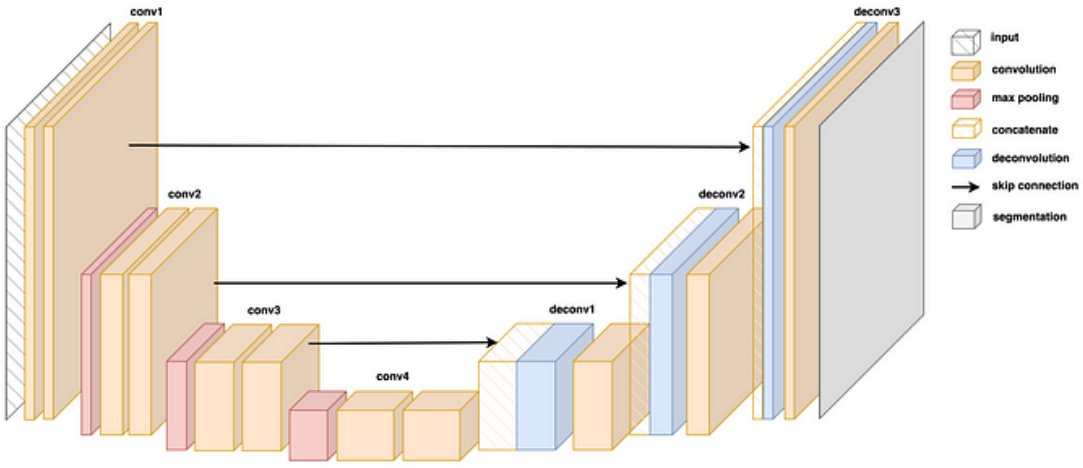

Figure 55: Illustration of the U-Net architecture and its main components, such as the encoder and decoder paths and the skip connections. Available at https://towardsdatascience.com/u-net-explained-understanding-its-image-segmentation-architecture-56e4842e313a.

5.2.2. CNN for Fault Identification

Automatic fault identification is an all-time concerning point that innovative geoscience studies have always sought to resolve. Some reasons for this are mainly: (i) the subjectivity of the traditional manual method; (ii) the amount of time spent on such tasks; (iii) the failure of standard analytical methods (seismic attributes) to isolate fault features from stratigraphic ones. Furthermore, all those problems increase significantly

in the face of noise seismic data with poorly imaged faults and/or with complex fault networks.

Countless works try to automate the fault identification task using several different approaches. Pedersen et al. (2002) used artificial ants to track discontinuities in seismic data, which would later be used as input to extract small fault patches, which in turn were merged to create continuous fault planes. However, this algorithm still needs user input, such as optimum seismic discontinuity attributes and search parameters. Tingdahl and Rooij (2005), inspired by the work of Meldahl et al. (1999), proposed using multiple seismic attributes that highlight discontinuities either in the amplitude, phase, or spectral components of the seismic trace to classify faults. They combined all those attributes through a Multi-Layer Perceptron network to return a binary cube of 0 (non-faults) and 1 (faults). The major pitfall of such an approach is that it still needs expert annotation for the training phase, leading to the subjectivity problem. Besides, it does not resolve the time-consuming point once MLPs are generally slow to train and predict and need a huge amount of labelled data. Some works even combined MLP and ant-tracking approaches to improve the aspect of the final result (Basir et al., 2013; Zheng et al., 2014). Alternatively, some works promote the use of image-processing techniques. For example, Hale (2013) uses fault strikes and dips estimated from the fault likelihood attribute to construct fault surfaces that coincide with ridges of the fault likelihood image. Despite the innovative aspect of all these works, none of those approaches solve entirely the above-mentioned problems of fault identification in seismic data.

In recent years, a major effort has been devoted to developing optimum deep-learning solutions for identifying faults in seismic data. Wu et al. (2019a) made the main reference to this area. Although other works have been previously published (e.g., Huang et al., 2017; Zhao and Mukhopadhyay et al., 2018; Guo et al., 2018; Di et al., 2018; Guitton, 2018) or simultaneously (e.g., Pochet et al., 2019; Zhang et al., 2019; Montaron et al., 2019; Cunha et al., 2019), the open-source nature of the whole study from Wu et al. (2019a) marks their work as the go-to study for such an approach. After that, many studies followed it, and some tried approaching this problematic task with different perspectives. An et al. (2021), Dou et al. (2021a), and Dou et al. (2021b) propose different CNNs with different architectures and loss functions that are more suitable to perform edge-detection tasks and attenuating noise in the final output rather than “simply” classify

pixels in an image. Guo et al. (2020) otherwise think that a CNN model designed to identify both horizons and faults would lead to a more geologically reasonable estimation of the faults.

At last, other works focused on the data rather than on the network's deep learning algorithm and architecture. Otchere et al. (2022) used a pre-trained CNN to predict fault location in preconditioned seismic data. In a way, it implies that they try to fit the aspect of the application data to the training dataset, which generally cannot fully represent the actual aspect of real seismic data. Jing et al. (2022) otherwise try to improve the realism of the training dataset. They use the exact U-Net architecture from Wu et al. (2019a) (Figure 56). However, they made a slight change in the seismic forward modelling process to create the synthetic seismic datasets. Instead of performing a 1D convolution of a given wavelet with a reflectivity model, they used a fast ray-based approach based on points-spread functions (Lecomte, 2004). Such an approach is known for creating synthetic seismic similar to real PSDM seismic data, such as RTM images (Jensen et al., 2021). Both data-focusing approaches mentioned above presented significant improvement compared with their basis case. Focus on data seems to be a fast and efficient way to help CNNs achieve better results regarding fault classification.

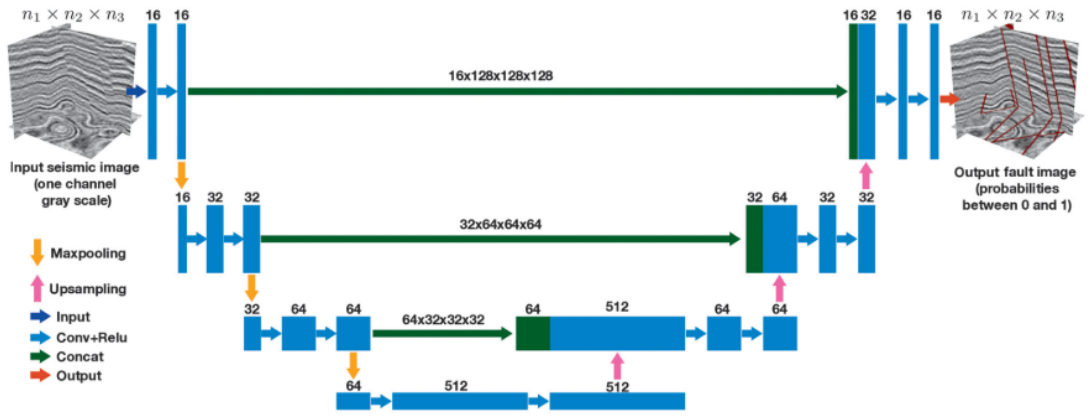

Figure 56: An example of a CNN of an encoder-decoder type with U-Net architecture designed for 3D fault recognition. From Jing et al. (2022), after Wu et al. (2019a).

5.3. Regression Methods

In general, regression analysis is a statistical method used to examine the relationship between a dependent variable (also called the response variable, output feature, target) and one or more independent variables (also called predictor variable, input feature). Regression models are commonly used in various fields, including economics, social sciences, engineering, and geoscience. Regression models can be used to make predictions, estimate the effects of different variables on the response, identify outliers and influential observations, and test hypotheses about the relationship between the variables.

There are several types of regression models, including linear regression, multiple-linear regression, polynomial regression, non-linear regression, tree-based regression, and others. Each of these types of regression methods has its strengths, weaknesses, and applicability. Like any statistical method, regression models also have assumptions and limitations that must be carefully evaluated and tested. Some common issues in regression analysis include overfitting, multicollinearity, heteroscedasticity, and model selection.

This study uses regression methods to model the relationship between fracture intensity and geometric seismic attributes through analytical models. I also used regression methods to estimate the fracture intensity away from the wells. This can be challenging due to the limited well coverage and spatial heterogeneity of fractures in subsurface formations. I carefully evaluated the assumptions and limitations of the models to ensure the validity and reliability of the results.

5.3.1. Non-Linear Regression

Non-linear regression is a statistical technique used to explore and model complex relationships between variables. Unlike linear regression, where a linear equation describes the relationship between the variables, non-linear regression allows the use of non-linear functions. In many cases, linear regression models are appropriate, but sometimes, the data suggest a more complex relationship. In such cases, non-linear regression techniques are more appropriate.

In the context of curve fitting, non-linear regression is a form of curve fitting that involves estimating the parameters of a function that best fits the observed data. In that

case, the fitted function should be non-linear regarding the parameters. Most rock properties and processes studied in geoscience have non-linear patterns and relationships. This, in turn, makes non-linear regression so valuable for this field.

Two common methods for estimating the parameters of non-linear regression models are the Gauss-Newton method and the gradient descent algorithm.

The Gauss-Newton method is an iterative approach that minimizes the sum of squared residuals between the observed data and the predicted values from the model. It starts with an initial estimate of the model parameters. It iteratively updates these estimates until the sum of squared residuals is minimized. It does so by assuming that the least squares function is locally quadratic and then finding the minimum of this quadratic function. This method is commonly used for non-linear least squares problems.

The gradient descent algorithm is another iterative approach that minimizes a cost function by updating the model parameters in the direction of the steepest descent. The cost function is typically defined as the sum of squared residuals between the observed data and the predicted values from the model. In the gradient descent method, the sum of the squared errors is reduced by updating the parameters in the steepest-descent direction. This method is commonly used in machine learning and deep learning applications.

Both methods have their advantages and disadvantages. The Gauss-Newton method uses curvature information (i.e., the second derivative) to take a more direct route. Therefore, it can converge quickly and efficiently for well-behaved problems but can be sensitive to the initial parameter estimates and may not converge for poorly behaved problems. The gradient descent algorithm, which uses first derivatives, is more robust to poorly behaved problems but can be computationally expensive and may require careful selection of learning rates and other hyperparameters.

5.3.1.1. Levenberg-Marquardt Algorithm

One common approach to non-linear regression is the Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963), an optimization technique used to find the parameters of the non-linear function that best fits the data. The method combines the Gauss-Newton and gradient descent methods, and it works by finding an iterative solution to the normal equations system that describes the parameter fitting of the non-linear function. The Levenberg-Marquardt method acts more like a gradient-descent method

when the parameters are far from their optimal value. It also acts more like the Gauss-Newton method when the parameters are close to their optimal value (Figure 57). The system is modified using a regularization parameter. This parameter controls the step size during each iteration, and it is adjusted based on the curvature of the cost function to balance between the efficient steps of the Gauss-Newton method and the conservative steps of the gradient descent algorithm. This approach can help the algorithm converge more quickly and robustly than the Gauss-Newton method alone.

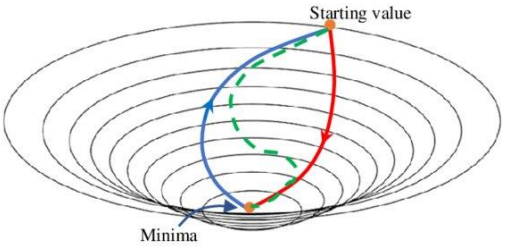

Figure 57: Levenberg-Marquardt method (dashed line) starts like gradient descent (blue) and switches to Gauss-Newton (red) when it becomes in the neighbourhood of the minima. From Mustafa (2020).

The Levenberg-Marquardt method is particularly useful for fitting non-linear models with multiple local minima or noisy data. It is widely used in various fields, such as engineering, physics, biology, and economics.

The method can be described mathematically as follows:

Given a set of data (x_i, y_i), the non-linear function $f(x;\theta)$ with parameters, θ is defined as:

$$f(x;\theta) + \varepsilon \qquad Eq.(42)$$

Where ε is the error term. The objective is to minimize the sum of squared errors between the observed values and the predicted values:

$$S = \sum \left(y_i - f(x_i;\theta)\right)^2 \qquad Eq.(43)$$

The Levenberg-Marquardt algorithm modifies the normal equations system as follows:

$$(J^T J + \lambda I)\Delta\theta = J^T e \qquad Eq.(44)$$

Where J is the Jacobian matrix, λ is the regularization parameter, I is the identity matrix, $\Delta\theta$ is the parameter update vector, and e is the residual vector.

The parameter update is then given by:

$$\theta(k+1) = \theta(k) + \Delta\theta \qquad Eq.(45)$$

The Levenberg-Marquardt algorithm has been extensively used in many fields, and it is efficient and reliable for parameter estimation of non-linear models.

5.3.2. Tree-based Ensemble Regression

Before introducing the basis of tree-ensemble methods, we must explain the bias-variance tradeoff concept. In machine learning, careful model selection is important for attaining optimum outcomes in classification and regression problems. This decision depends on numerous factors, such as the data volume, the feature space's dimensionality, and the underlying distribution assumptions. To train a robust model, we should carefully select the input features, a proper initial model or machine learning algorithm, a suitable way to optimize the parameter of such mode better, and good dataset training that would help our model to generalize. A model will only be considered robust enough if it can perform accurate predictions outside the training dataset. In other words, to ensure the robustness of a model, we have to evaluate the global error (generalization error) in a test dataset.

Imagine that we have a set of given trained models, such as multiple attempts with different parametrizations or different training datasets. The global error of this set of models measured in an independent test dataset can be decomposed into the error given by bias, the variance, and an irreducible error. Let us assume a dependent variable Y defined by an independent variable X and an error e, with e being normally distributed with a mean equal to zero, then Y is defined as follows:

$$Y = f(X) + e \qquad Eq.(46)$$

If we then create a model that approximates the true function $f(X)$ and called it $\hat{f}(X)$, so the expected error at an unseen (test) point x using a squared-error loss is defined as:

$$Err(x) = E\left[\left(Y - \hat{f}(x)\right)^2\right] \qquad Eq.(47)$$

That can be further decomposed as:

$$Err(x) = \left(E\left[\hat{f}(x)\right] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E\left[\hat{f}(x)\right]\right)^2\right] + \sigma_e^2 \qquad Eq.(48)$$

Which can be translated into:

$$Err(x) = Bias^2 + Variance + Irreducible\ error \qquad Eq.(49)$$

Bias is the difference between a model's average prediction and the true value it aims to predict. High bias can cause a model to oversimplify training data and lead to high error rates on both training and test data (Figure 58). Variance refers to the variability in a model's predictions, with high variance causing a model to overfit the training data and perform poorly on new, unseen data (Figure 58). Irreducible error is the remaining error after creating accurate models and reflects the level of noise present in the data.

An optimal model should strike an optimum balance between low bias and low variance. It should possess sufficient degrees of freedom to capture the intricate complexities of the data while avoiding excessive flexibility that can lead to high variance and reduced robustness (Figure 58 and Figure 59). The bias-variance tradeoff, a well-established concept, embodies this delicate equilibrium. This concept is part of the core of tree-based ensemble techniques.

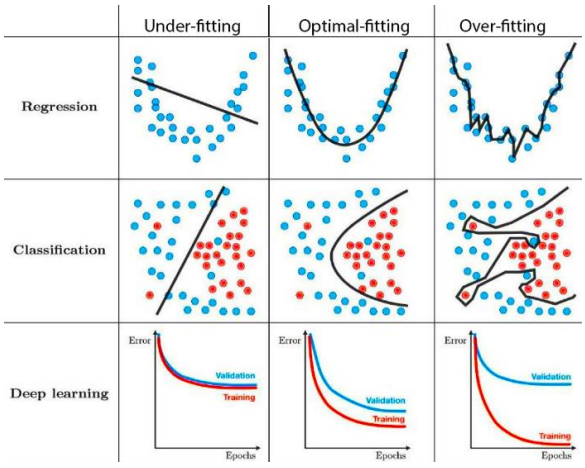

Figure 58: Cartoon illustrating different situations for regressor and classifier machine learning models. The first one is underfitting, meaning it has a high bias. Therefore, it fails to optimize the training and validation data predictions. The second one achieves optimal training and performs considerably well in

training and validation datasets. The third one is overfitting; therefore, it performs extremely well in training data but fails to generalize outside of it. Available at https://towardsdatascience.com/techniques-for-handling-underfitting-and-overfitting-in-machine-learning-348daa2380b9

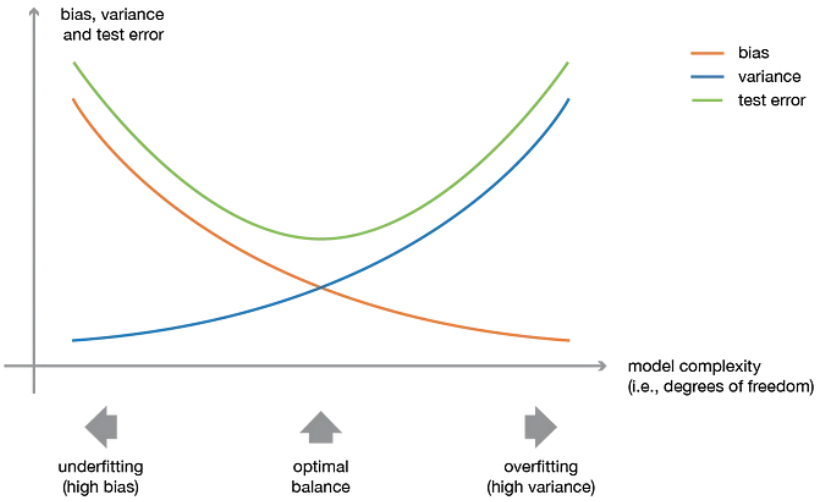

Figure 59: Illustration of the bias-variance tradeoff. The global error of a given set of trained models can be decomposed into the error given by bias, the variance, and an irreducible error. As the model complexity increases, so does the variance. Otherwise, if the model is too simple, the variance is low, but the bias is high. There is an optimum point that balances variance and bias, therefore minimizing the global error. Available at https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205

Tree-based ensemble methods are a category of supervised machine learning methods that aggregate in some manner the multiple results of different decision trees to perform a task, classification, or regression. When dealing with tabular datasets, the tree-based ensemble methods such as Random Forest and eXtreme Gradient Boosting (XGBoost) usually outperform other machine learning techniques, for example, neural networks (Grinsztajn et al., 2022). Furthermore, they are generally faster, more scalable, require less training data, and work better in the face of missing data. Not by coincidence, they are generally the first choice of winning teams of Kaggle competitions and other machine learning conquests (Chen and Guestrin, 2016).

Tree-based ensemble methods outperform single neural nets for a reason. The logic is that even a group comprising the average of many mediocre models can still beat a single good model. Also, tree-based ensembles are less prone to overfitting for this reason. For instance, a Random Forest can reduce the high variance of a flexible model like a decision tree by combining many trees into one ensemble model. Ensembles are

more effective than neural networks, which require much more data than the average person can access. Additionally, neural networks can sacrifice the interpretability of features for the sake of performance, rendering the features meaningless (Montantes et al., 2022). Empirical evidence suggests that ensembles tend to perform better when the models are diverse (e.g., Sollich et al., 1996; Kuncheva et al., 2001).

Some recent works used tree-based ensemble estimators in geoscience. However, the majority of them use it for lithology classification. For example, Kuhn et al. (2018) used non-seismic geophysical data aside from remote-sensing data as input into a random forest classifier to create a lithological map that would help with mining activity. Bhattacharya and Mishra (2018), using well logs as input, evaluated the use of both Bayesian Networks (a type of graph network) and Random Forests to classify facies and fractured/non-fractured zones in three different formations from the Appalachian Basin. In their work, the Random Forest achieved higher accuracy in the classification. Babasafari et al. (2022b) evaluated using several machine learning methods to classify facies in a dataset from Santos' presalt. Again, the tree-based ensemble classifiers (Random Forest and XGBoost in that cases) outperformed the other methods such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Gaussian Naïve Bayes, and Decision Tree. Ngoc Tan et al. (2022) concluded with a similar methodology but in a different area.

Although rarer, some works use tree-ensembled predictors for other tasks, such as regression and forecasting. Bhattacharya et al. (2019) try to forecast the gas production performance in the Marcellus Shale from the Appalachian Basin. They used geophysical well logs added to drilling and production data to do so. They evaluated the use of Random Forest and compared it to other techniques such as Neural Net and Support Vector Machine. The Random Forest achieved the best results regarding accuracy and computational cost. In another work, Man et al. (2021) evaluate a series of machine learning methods to predict flow zone indicator (FZI) in Nam Con Son Basin, Vietnam. Again, the Random Forest outperformed other methods, reaching an R^2 of 0.79 in the test set. Mandal et al. (2021) also evaluated several machine learning regressors to predict missing logs, such as neutron porosity (NPHI) and bulk density (RHOB). Besides, they used the same algorithms to predict the total organic matter (TOC) in an unconventional reservoir. They compared them with traditional empirical methods such as the Passey and

Schmoker methods. Random Forest was the best estimator for both goals, followed closely by a Gradient Boosting Regressor.

Before going deeper into the algorithm explanation of Random Forest and XGBoost, we have to define the method that is the basis from both of those, Decision Tree.

5.3.2.1. Decision Trees

Decision Trees are non-parametric models that partition the predictor space into regions and assign a label to each region. The regions are defined by a set of splitting rules based on the predictors' values. The label of each region is typically the average or median of the target variable values of the observations that belong to that region. A decision tree can be represented as a hierarchical structure, where each node represents a splitting rule, and each leaf represents a region with a label (Figure 60). The splitting rules are chosen to maximize the target variable's impurity or variance reduction. The impurity measure depends on the task. For instance, it can be entropy, Gini index, or classification error for classification, and mean squared error or mean absolute error for regression. Two main hyper-parameters in a decision tree are the maximum depth and minimum samples per leaf. The maximum depth controls how far a tree can grow before making a prediction. Deeper trees have higher accuracy, although higher depths also increase the variance of the model, ultimately leading to overfitting (Figure 61). Minimum samples in a leaf control the number of samples needed in a leaf to validate a prediction. Increasing the number of samples makes the final predictor more rigorous and reduces the variance, but on the other end, it can produce underfitted models with high bias (Figure 62).

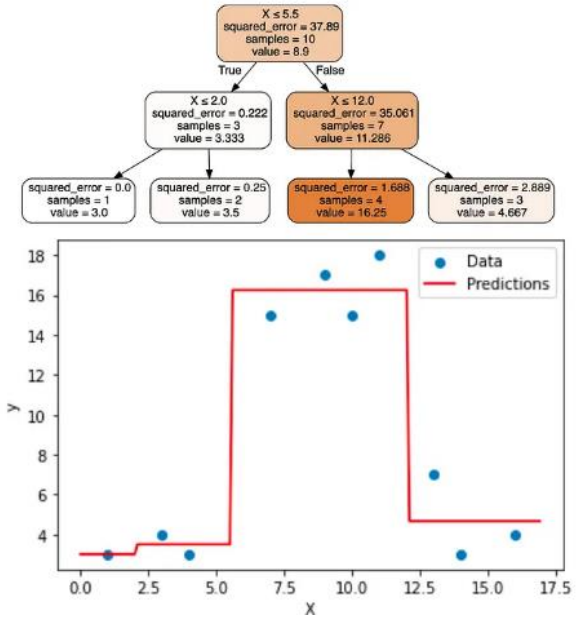

Figure 60: Example of a decision tree with a maximum depth equal to two that achieves its optimum prediction with two nodes and four leaves. Available at https://towardsdatascience.com/decision-tree-regressor-a-visual-guide-with-scikit-learn-2aa9e01f5d7f.

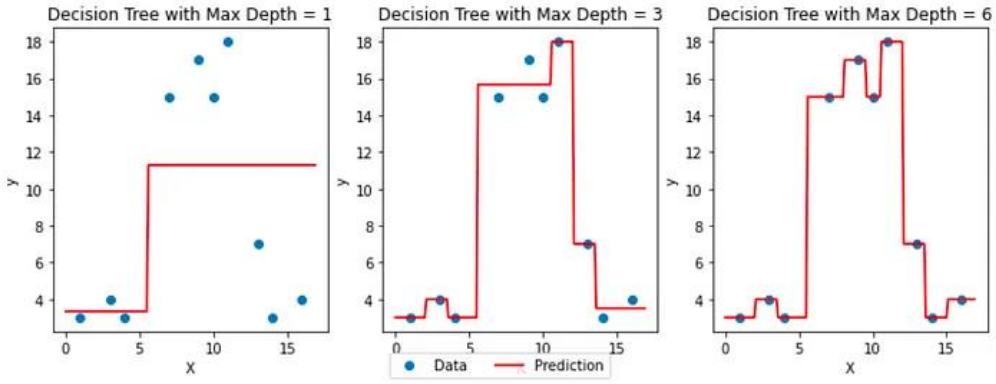

Figure 61: Examples of how increasing the maximum depth of a decision tree affects the final result. With a maximum depth of one, the tree cannot fit the samples (high bias leading to underfitting). Increasing the maximum depth to three increases the variance and improves the model prediction. Ultimately, with a maximum depth of six, the model simply interpolates the points, which means it is overfitting. Available at https://towardsdatascience.com/decision-tree-regressor-a-visual-guide-with-scikit-learn-2aa9e01f5d7f.

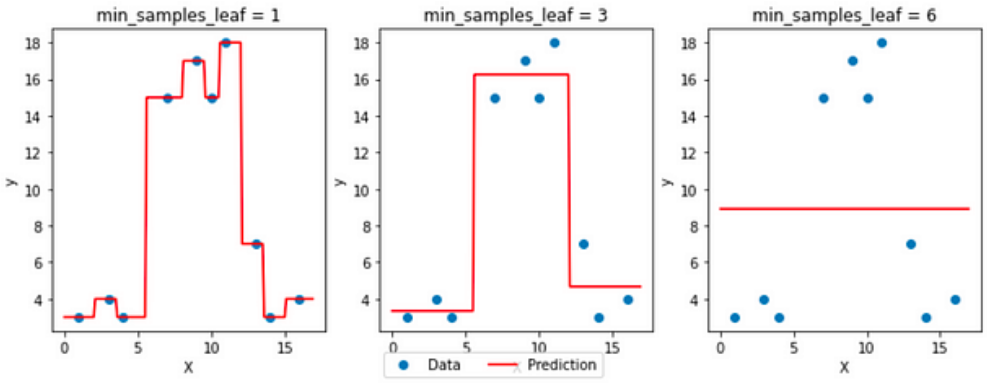

Figure 62: Example of how increasing the minimum samples per leaf can affect the final result. Minimum samples per leaf equal to one make the model flexible, which can lead to overfitting. Increasing its value to three can fine-tune the model, but ultimately, a relatively high value, for example, six, makes the model too rigorous, which leads to underfitting. Available at https://towardsdatascience.com/decision-tree-regressor-a-visual-guide-with-scikit-learn-2aa9e01f5d7f.

5.3.2.2. Random Forest

Random Forest is one type of tree-based ensemble method. It aggregates the results of several decision trees, which are randomly created, then the name Random Forest. The Random Forest bases rely on bootstrapping aggregating, also known as *bagging*, which Breiman first used in decision trees (1996). In his work, Breiman defended the use of bagging to reduce the variance and uncertainty of unstable predictors, such as neural nets and decision trees. An illustration of the bagging procedure is presented in Figure 63 and Figure 64.

Figure 63: Example illustrating the bootstrap processing. Available at https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205.

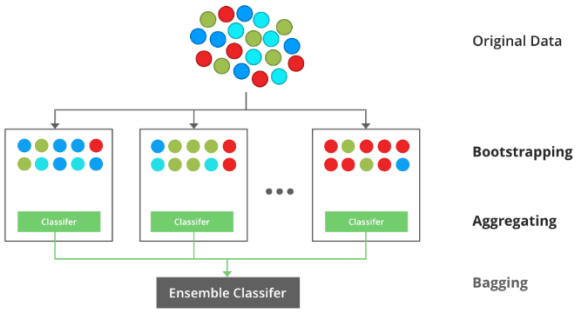

Figure 64: Example illustrating a general bagging classifier https://www.geeksforgeeks.org/bagging-vs-boosting-in-machine-learning/.

The bagging technique used by Breiman in the Random Forest method (Breiman, 2001) creates random sub-samples with replacement of the dataset, both in feature and target variables in every splitting node. Therefore, although similar to bagging, the Random Forest can be seen as an extension of bagging. While Bagging constructs several independent decision models, each trained with a random subset of the training data in tree-root Figure 65. Random Forest adds an extra layer of randomness due to the subsampling of features at each decision node during tree building. This helps to increase the diversity among trees and reduce correlation among them, leading to a more robust and accurate model.

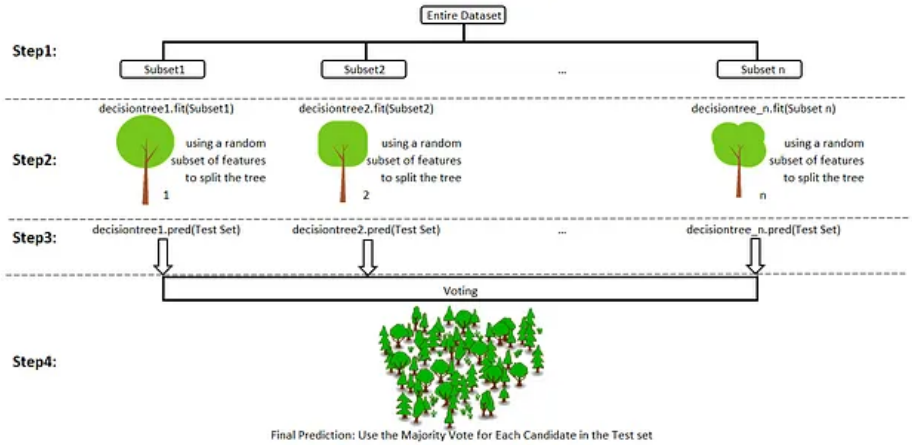

Figure 65: Random Forest scheme. The algorithm uses the random sampling of the features both in the root level of the trees and in the splitting nodes. Available at https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725

Mathematically, the predicted value of an individual regression tree can be represented as:

$$y = f(x) + \varepsilon \qquad Eq.(50)$$

Where y is the predicted value, x is the input variable, $f(x)$ is the unknown function that maps the input variable to the predicted value, and ε is the error term.

The goal is to estimate $f(x)$ as accurately as possible. To do this, we use the bagging method to create multiple regression models on different subsets of the training data. The predicted value for each regression tree can be represented as:

$$y_i = f_i(x) + \varepsilon_i \qquad Eq.(51)$$

For regression problems, each decision tree in the Random Forest predicts a continuous output value. The final prediction of the Random Forest is then the average of the predictions of all the trees. Mathematically, if we have a set of decision trees $T_1, T_2, \dots, T_n$ and a new input sample x, the prediction of the Random Forest is given by:

$$RF(x) = \frac{1}{n}\sum_{i=1}^{n} T_i(x) \qquad Eq.(52)$$

The pseudo-algorithm of Random Forest can be summarized as follows:

i. For each $Tree_i$, take a bootstrap sample D_i of the training data.

ii. Select a random subset of features, $feat_i$, of size max_features.

iii. Grow a decision tree T_i on D_i using only the features in Fi.

iv. Repeat steps 1-3 for n_estimators' trees.

v. For a new input x, predict the output by averaging the predictions of all trees.

5.3.2.3. eXtreme Gradient Boosting (XGBoost)

Boosting (Friedman, 2001) is another type of ensemble method that works by training weak models, or models with only slightly better performance than random guessing, and then combining their predictions to make a final prediction. The weak models are trained sequentially, with each model trying to correct the mistakes made by the previous one (Figure 66). This process is repeated until the desired level of minimum error prediction is achieved. As the ensemble incorporates a new tree, its model

complexity rises while the overall bias diminishes, regardless of the simplicity of the new tree. Compared to bagging, boosting methods aim to reduce the bias of the final prediction by gradually increasing the variance of a series of weak predictors that initially have high bias and using a weighted average of the estimated trees to make the final prediction. On the other hand, bagging methods reduce the variance of the final prediction by averaging weak predictors with high variance (Figure 67). Boosting algorithms are known for their ability to handle complex datasets and high prediction accuracy, but they can also be more computationally expensive and susceptible to overfitting.

Among the several boosting algorithms, one that stands out is XGBoost (Chen & Guestrin, 2016). It is a specific type of boosting algorithm that was developed for efficient and accurate prediction. It stands for "Extreme Gradient Boosting" and is based on the gradient boosting algorithm. It combines boosting with gradient descent to optimize the loss function. XGBoost is widely used in data science competitions and is known for its fast training speed and high prediction accuracy (Jiang et al., 2021; Grinsztajn et al., 2022). It also includes several additional features, such as regularization and parallel processing, which make it more flexible and easier to use than other boosting algorithms. The regularization terms and other features implemented in the algorithm help smooth the final learned weights to avoid overfitting.

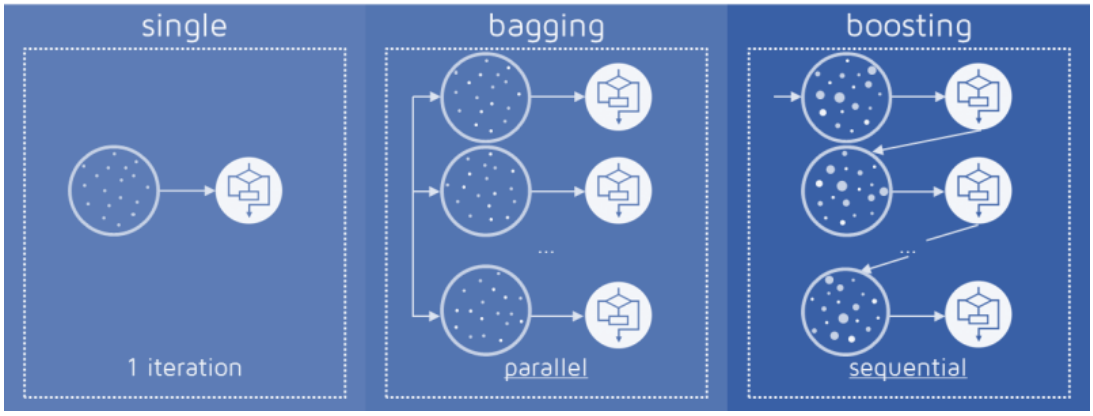

Figure 66: Illustration of the difference between the learning process of Single Decision Trees, Bagging methods (Random Forests), and Boosting. Source: Bagging vs Boosting (Prashant, 2020), available at: https://www.kaggle.com/code/prashant111/bagging-vs-boosting.

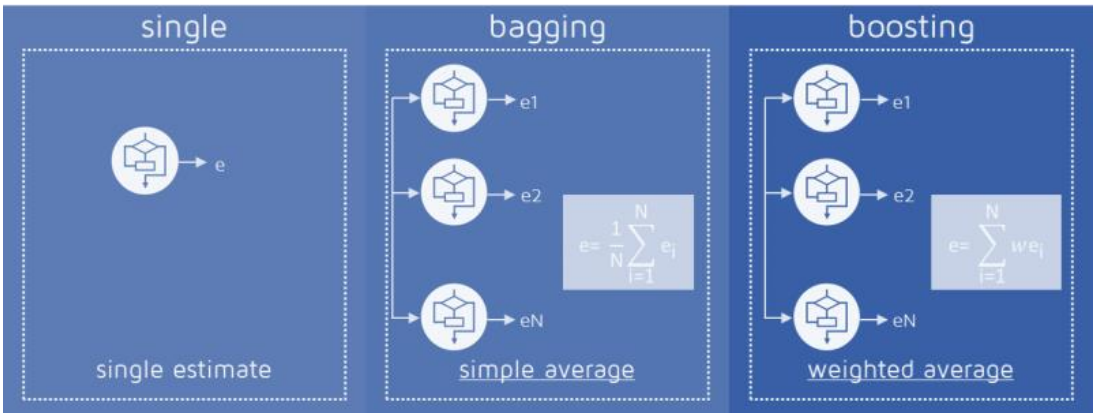

Figure 67: Illustration of the difference between the final estimation of Single Decision Trees, Bagging methods (Random Forests), and Boosting. Source: Bagging vs Boosting (Prashant, 2020), available at: https://www.kaggle.com/code/prashant111/bagging-vs-boosting

The boosting algorithm for regression can be summarized as follows:

i. Initialize the model with a constant value, such as the mean of the target variable.

ii. Train a weak model on the training data.

iii. Add the predictions of the weak model to the current model.

iv. Repeat steps ii-iii for a specified number of iterations.

The weak model can be any regression model, such as a decision tree or linear regression. The goal is to create a model that can accurately predict the target variable. The key to the success of boosting is the ability to learn from the errors made by the previous models. Correcting the errors can make the model more accurate with each iteration.

XGBoost has several hyperparameters that can be tuned to optimize the model's performance. Some of the most important hyperparameters are:

- Learning rate: determines the step size at each iteration. A low learning rate can result in slower convergence and reduce each iteration's impact on the prediction. A high learning rate can result in overshooting the optimal solution.
- Max depth: determines the maximum depth of each tree. A deeper tree can capture more complex patterns in the data but can also lead to overfitting.

- Subsample: determines the fraction of the training data to use for each iteration. A smaller subsample can reduce overfitting but also result in slower convergence.
- colsample by[]: a collection of parameters of subsampling columns. It determines the fraction of features to use. The subsampling can be done for each tree (colsample_bytree), for each depth level reached in a tree (colsample_bylevel), or for every time a new split is evaluated (colsample_bynode). Note that these parameters can work simultaneously: if every parameter has a 0.5 value and the dataset has 32 columns, then each split would use four columns (32/ 2^3). A smaller fraction can reduce overfitting and result in losing essential information.
- Lambda: determines the L2 regularization penalty on the weights. A larger penalty can reduce overfitting but also result in underfitting.
- Alpha: determines the L1 regularization penalty on the weights. A larger penalty can reduce overfitting but also result in underfitting.
- Gamma: minimum loss reduction required to partition a tree leaf node further. The larger the gamma is, the more conservative the algorithm will be.

In practical terms, it is important to tune these hyperparameters to optimize the performance of the model. We can find the optimal values for each hyperparameter using cross-validation and grid search. By understanding the mathematical foundation of XGBoost, we can better use this algorithm to solve real-world regression problems.

Regarding mathematical formulas, XGBoost uses a gradient-boosting framework to minimize a loss function. Let us consider the case of squared error loss, which is commonly used for regression problems. We can write the objective function for XGBoost as:

$$Obj(\Theta) = \sum_{i=1}^{n} [y_i - \hat{y}_i(\Theta)]^2 + \sum_{k=1}^{K} \Omega(f_k), \qquad Eq.(53)$$

Where Θ is the set of parameters to be optimized, y_i is the true value of the target variable for the $i-th$ example on the current set of parameters, K is the number of trees

in the model, f_k is the $k-th$ tree. $\Omega(f_k)$ is a regularization term that penalizes complex trees and is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2 \qquad Eq.(54)$$

Where T is the number of leaves in a Tree, and w is the leaf weights.

The goal of XGBoost is to find the set of parameters Θ that minimize the objective function. This is done through a gradient descent algorithm that updates the parameters at each iteration. Specifically, XGBoost uses a second-order approximation of the loss function to optimize the parameters. This is done by computing the first and second derivatives of the loss function with respect to the predicted values of the target variable and then using these derivatives to update the parameters.

The final prediction of the XGBoost model is obtained by summing the predictions of all the trees in the model, multiplied by a shrinkage factor (i.e., the learning rate). This is given by:

$$(\Theta) = \sum_{k=1}^{K} \eta f_k(x) \qquad Eq.(55)$$

Where η is the learning rate, and $f_k(x)$ is the predicted value of the $k-th$ tree for the input example.

In conclusion, XGBoost is a powerful machine learning algorithm that uses boosting to combine several weak models into a strong model for regression and classification problems.

5.4. Hyperparameter Tuning

Hyperparameter tuning is a critical aspect of building machine learning models, as it involves selecting the optimal values for the hyperparameters that define the model's architecture and behaviour. These hyperparameters, which cannot be learned from the data and must be set before training, significantly impact the model's performance and generalization ability. Researchers have developed various techniques to aid in hyperparameter tuning, including K-Fold Cross-Validation, Random Search, and the Search Grid. These techniques play a pivotal role in enhancing the robustness and

effectiveness of model selection, ultimately leading to improved performance and predictive accuracy.

5.4.1. K-Fold Cross Validation

K-Fold Cross-Validation is a machine learning technique to evaluate a model's performance. It involves dividing the dataset into K subsets, or "folds", where K is usually between 5 and 10. In general, we first split the whole dataset into training and tests. The test set is separate to evaluate the model's generalisation performance further. Then, we use K-Fold Cross Validation on the training set (Figure 68). The algorithm for K-Fold Cross Validation is as follows:

Algorithm 1: K-Fold Cross Validation

Input: Dataset D, number of folds K

Output: Mean accuracy of the model

1 Shuffle dataset D randomly.

2 Split the dataset into K-equal subsets.

3 **For** each fold k in K:

4 | Train the model on all subsets except the k-th subset.

5 | Evaluate the model on the k-th subset and record the accuracy.

6 **End**

7 Calculate the mean accuracy of the model over all folds.

The advantage of using K-Fold Cross-Validation is that it provides a more accurate estimate of the model's performance than simply using a single train-test split. Using multiple splits and averaging the results, K-Fold Cross-Validation reduces the impact of any particular split on the evaluation.

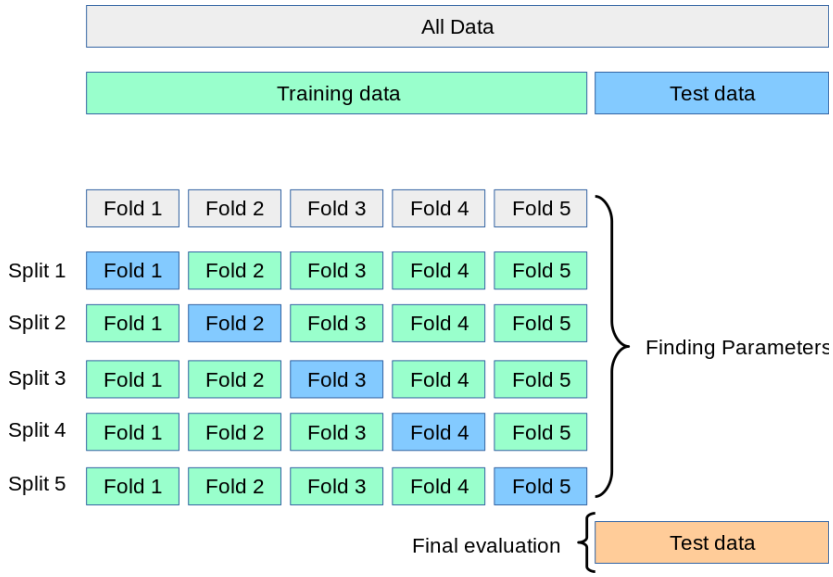

Figure 68: Example of a K-Fold Cross Validation of a dataset, splitting it into training, validation, and test data. Available at https://scikit-learn.org/stable/modules/cross_validation.html.

5.4.2. Search Grid

Search Grid is a machine learning technique to tune a model's hyperparameters. Hyperparameters are values that are set before training the model and cannot be learned from the data. The process of tuning hyperparameters involves trying different combinations of hyperparameter values to find the best set of values that results in the highest performance of the model. The algorithm for Search Grid is as follows:

Algorithm 2: Search Grid

Input: Dataset D, hyperparameter grid G

Output: Best hyperparameter values

1 Define a grid G of hyperparameter values to search over.

2 **For** each set of hyperparameter values g in G:

3 Divide dataset D into training and validation sets.

4 Train the model on the training set using the hyperparameter values g

5 Evaluate the model on the validation set and record the performance.

6 End

7 Select the set of hyperparameter values that achieved the highest score on val. set.

The advantage of using Search Grid is that it helps to identify the best set of hyperparameter values that results in the highest performance of the model. By trying different combinations of hyperparameter values, Search Grid helps avoid overfitting and improves the model's generalisation to new data.

5.4.3. Random Search

Random Search is an alternative technique to Search Grid that involves randomly selecting hyperparameter values to try. The algorithm for Random Search is as follows:

Algorithm 3: Random Search

Input: Dataset D, hyperparameter space H, number of iterations N

Output: Best hyperparameter values

1 Define a grid G of hyperparameter values to search over.

2 **For** each iteration n in N:

3 Randomly sample a set of hyperparameter values h from H.

4 Divide dataset D into training and validation sets.

5 Train the model on the training set using the hyperparameter values h.

Evaluate the model on the validation set and record the performance.

6 End

7 Select the set of hyperparameter values that achieved the highest score on val. set

---

The advantage of using Random Search is that it can be more efficient than Search Grid when the hyperparameter space is large. Instead of exhaustively searching the entire hyperparameter space, Random Search randomly samples hyperparameter values, which

can help to identify good hyperparameters faster. Random Search can also help identify unexpected combinations of hyperparameters that a grid search might overlook.

Overall, K-Fold Cross Validation, Search Grid, and Random Search are powerful techniques that can help improve machine learning models' performance. By using K-Fold Cross Validation to evaluate the model and Search Grid or Random Search to tune the hyperparameters, researchers can develop more robust models, indirectly infer the occurrence of overfitting or underfitting, and gain some insight into how the model performs on unseen data.

5.5. Feature Importance

To finish this machine learning chapter, we introduce the concept of feature importance. There are several types of feature importance measurements. Most tree-based algorithms have a built-in implementation of feature importance based on the mean decrease in impurity (MDI), also called gini importance (Scornet, 2023). It calculates each feature's importance as the sum of the number of splits (across all trees) that include the feature proportionally to the number of samples it splits. This method measures how effective a given feature is at reducing uncertainty, in the case of classifier, or the variance of a prediction, for regressors. Features at the top of the three will consequently contribute to a substantial portion of the input samples. Then, the expected amount of the samples they contribute can thus be used to estimate the relative importance of the features.

The impurity-based feature importance computed on tree-based models suffers from two flaws that can lead to misleading conclusions. First, they are computed on statistics derived from the training dataset and, therefore, do not necessarily inform us of which features are most important to make good predictions on a held-out dataset. Secondly, they favour high cardinality features, which are features with many unique values. Permutation feature importance, also known as mean decrease in accuracy (MDA), is an alternative to impurity-based feature importance that does not suffer from these flaws.

Another alternative approach to feature importance is the Shapley Additive Explanations (SHAP) method. Derived from cooperative game theory, SHAP provides a unified framework for quantifying the contribution of each feature in a prediction by considering all possible combinations of features. SHAP values assign each feature an

importance score based on its impact on the difference between the expected and actual predictions. The SHAP method has gained significant attention and has been successfully applied across various domains, providing interpretable and reliable insights into the importance of features in machine learning models.

5.5.1. Permutation Importance

Permutation Importance is not a novel method that scientists figured out recently. It has been widely used long, even before Random Forest existed. The idea is simple: we train the model on the train set and get the model score on the test set. This score will be our baseline. We will shuffle one feature at a time on the test set and then feed the data to the model to get a new score. If the feature we just shuffled is important, the model should suffer a lot, and the score should drop drastically. The permutation importance strategy does not require retraining the model after permuting each column; we just must re-run the perturbed test samples through the already-trained model. The process is repeated a number of times (defined by the user) to ensure the statistical significance of the results. In the end, the user gets a distribution of importance measured in each iteration (Figure 69).

Breiman (2001) introduced both permutation importance and mean decrease impurity in his 2001 paper "Random Forests". While the two methods are different, they are complementary and can be used together to provide a more comprehensive understanding of variable importance in random forests.

In the paper, Breiman used mean decrease impurity to measure variable importance within the random forest algorithm, and he used permutation importance to assess the external validity of the random forest model by evaluating the importance of each predictor variable concerning the model's predictive accuracy. Breiman argued that permutation importance is a more reliable measure of variable importance than mean decrease impurity, especially when dealing with noisy data or correlated predictor variables. However, mean decrease impurity can still be useful for understanding how the random forest algorithm makes splits and constructing decision trees.

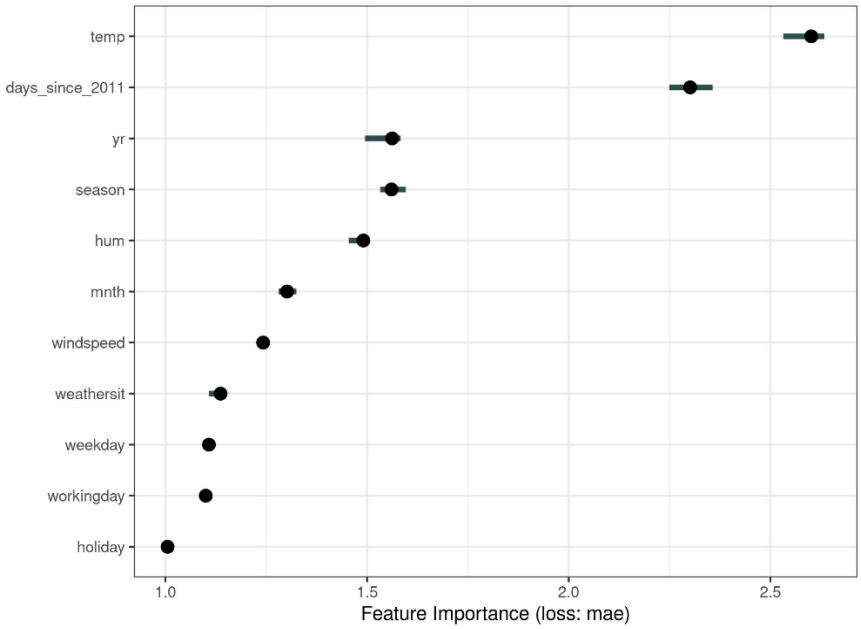

Figure 69: Example of feature importance measured by permutation importance method. The mean absolute error has been used as the evaluation metric in that case. The model was designed to predict the number of rented bikes given weather conditions and calendar information. According to the permutation importance method, the most important feature is temperature. From Molnar (2023). Available at https://christophm.github.io/interpretable-ml-book/index.html.

5.5.2. SHapley Additive exPlanations

The SHapley Additive exPlanations (SHAP) method (Lundberg and Lee, 2017) is a tool for interpreting the predictions of complex machine learning models. The method provides a way to understand how individual features contribute to a model's output and is based on the concept of Shapley values from cooperative game theory.

In a cooperative game, the Shapley value of a player is the average contribution that the player makes to all possible coalitions that could be formed. In the context of feature importance, we can think of the players as the features in our model and the coalitions as subsets of those features. The Shapley value of a feature is then the average contribution it makes to the model's output across all possible subsets of features.

To calculate the Shapley value for a particular feature, we need to consider every possible combination of features and their corresponding predictions. For each combination, we determine the contribution of the feature of interest by comparing the prediction with and without that feature included. The Shapley value is then the average of these contributions across all possible combinations. The SHAP method extends this

idea to a machine learning context using a set of "reference" predictions as a baseline. The method computes the Shapley values for each feature with respect to this baseline and uses them to generate a visual representation of feature importance. This representation, known as a SHAP plot, shows how each feature contributes to the model's output.

Figure 70 shows an example of a SHAP plot used to understand a machine learning model for house-price prediction. The plot is called the waterfall plot. It shows how each feature pushes the model output from the base value (the average prediction over the training dataset) to the model output in a particular instance. Features pushing the prediction higher are shown in red; those pushing the prediction lower are in blue. The bar length gives the SHAP value for each feature in this observation. The sum of all SHAP values will be equal to $E[f(x)] - f(x)$, where $E[f(x)]$ is the expected value of the model predictions, $f(x)$ is the predicted value of the model for the given instance. Figure 71 shows another example of a SHAP plot used in the same model. This plot is called the beeswarm plot and is used to understand the general contribution of each feature and how it changes with respect to the feature value. Therefore, this plot shows that the feature MedInc, which is the median income of the houses in the area, has the highest contribution to the model. It also shows that the higher values of medium income tend to increase the prediction (high positive SHAP values). In contrast, the low median income values tend to lower the final prediction (high negative SHAP values). Those two plots are just a few of a vast variety of plots that can be built with the SHAP method, which helps increase the interpretability of machine learning models.

This different possibility of visualization is one of the main strengths of the SHAP method compared to other feature importance methods. However, the SHAP method is quite different from other feature importance methods in other aspects also. Most traditional feature importance methods do not account for interactions between features and can be sensitive to correlated or redundant features. The SHAP method addresses these issues by providing a more nuanced and model-specific view of feature importance that considers the interactions between features (Lundberg and Lee, 2017). Another advantage of the SHAP method is that it can generate local as well as global explanations of a model's output. In contrast, permutation importance and Gini importance provide only a global view of feature importance and may not reveal the importance of features

for specific cases. The SHAP method can also handle more complex models, such as neural networks and deep learning models (Lundberg et al., 2020).

Despite these advantages, the SHAP method can be computationally expensive, especially for large datasets or complex models. Sometimes, simpler methods like permutation importance may be more practical, especially for exploratory analysis or feature selection. However, Lundberg et al. (2020) have proposed a variant of the SHAP method designed for tree-based ensemble models that are less time-consuming than the original implementation of the SHAP method (Lundberg and Lee, 2017).

Overall, the choice of feature importance method depends on the specific problem and the goals of the analysis. For many applications, the SHAP method offers a powerful and flexible tool for understanding complex machine learning models and building trust in their predictions.

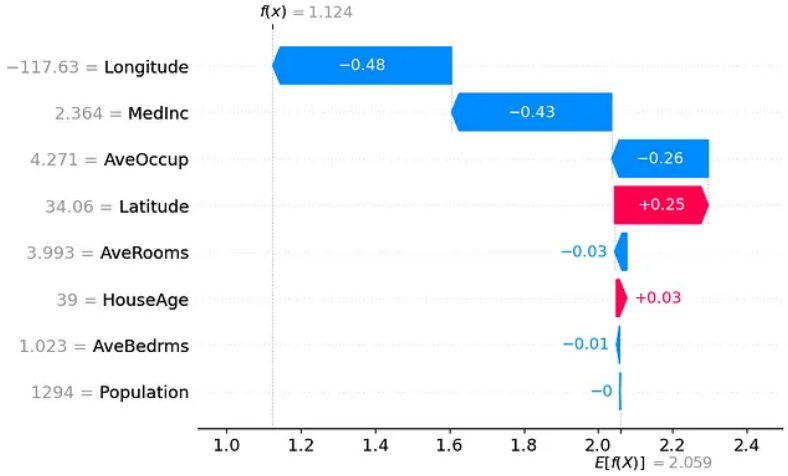

Figure 70: Example of SHAP waterfall plot of a tree-based ensemble machine learning model for house-price prediction. Available at https://towardsdatascience.com/using-shap-values-to-explain-how-your-machine-learning-model-works-732b3f40e137

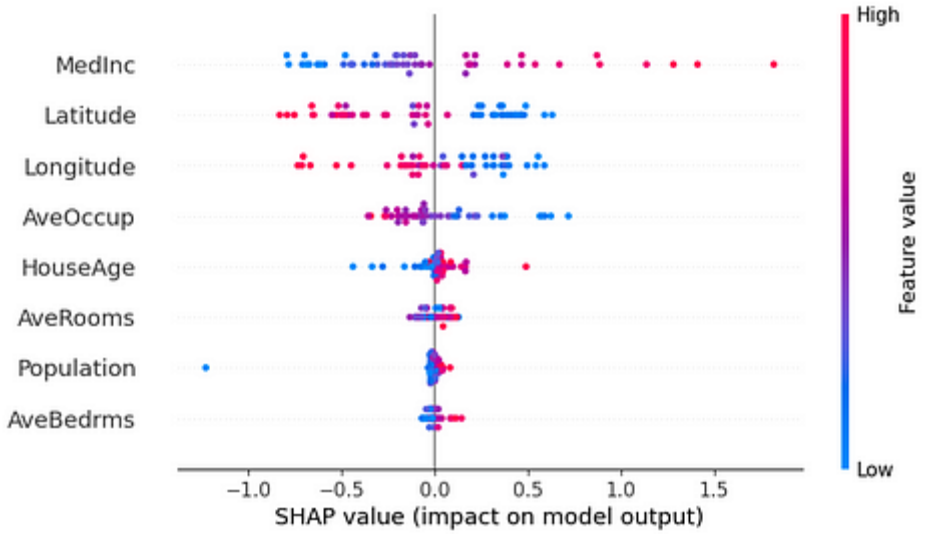

Figure 71: Example of SHAP beeswarm plot of a tree-based ensemble machine learning model for house-price prediction. Each dot corresponds to an instance value of the model. The dot's position on the x-axis shows the impact that feature has on the model's prediction for that person. When multiple dots land at the same x position, they pile up to show density. Available at https://towardsdatascience.com/using-shap-values-to-explain-how-your-machine-learning-model-works-732b3f40e137.

6. Material and Methods

This chapter presents the material used in the study, such as the seismic dataset, the location of the wells, an inventory with additional information. The chapter also presents the applied methodology, which is divide into two parts: 1st part, evaluation of the use of curvature as a proxy for fracture intensity. I work to find an optimum parametrization of curvature that would better fit the upscaled fracture intensity. 2nd part, 1D and 3D fracture intensity prediction using machine learning regression. I aggregate the conclusions of the first part with other concepts, and then I train a machine learning model to predict fracture intensity, both in the wells, and in the seismic volume (away from the wells).

6.1. Dataset

The dataset available for work comprises 9 wells and a 3D Pre-Stack Depth Migrated (PSDM) seismic volume with a bin size of 25 m x 25 m, a sample rate of 5m, zero-phase, and positive SEG Polarity. The seismic processing used Reverse Time Migration and Tilted Transverse Isotropy (TTI) velocity model for the migration. A base map for spatial contextualization is presented in Figure 72. Additionally, seven wells (A - G) with Borehole Image Logs (BHI) were available. I use five of those to build an analytical model correlating the curvature with the fracture intensity. The same wells are used to train a machine learning model of fracture intensity prediction in the second part of the study. However, in this part, I use Well G as a blind test. Additionally, a tenth well, located outside the study area, in a nearby field, was used as an additional blind test to understand if the trained model can be directly applied in other contexts. Three cross-sections intersecting two of the five wells (used in the first part of the study of curvature analysis) are also highlighted with black rectangles in the image (Figure 72).

It is worth noting that we excluded the Well F from the statistical analysis and analytical modelling (1st part of the study), as also from the exploratory data analysis and training of machine learning regressor (2nd part of the study). The Well F is the smallest in number of samples, and almost half of those samples are related to bad recording of the BHI logs due to breakout zones. The few samples that remain with good recording would disturb the statistical analysis and training of the machine learning model.

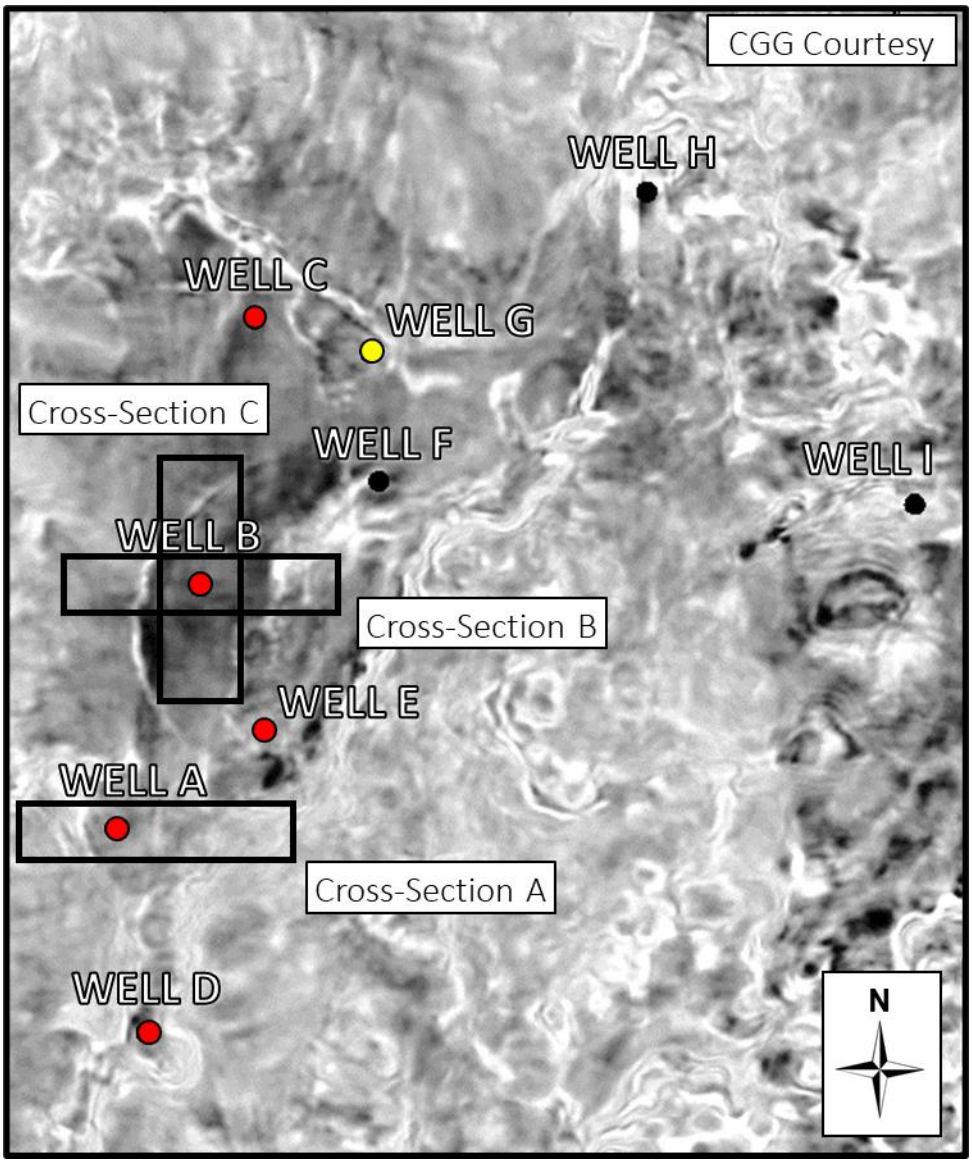

Figure 72: Base map of the study area with seismic amplitude attribute at the top of the reservoir showing the location of the available wells and the seismic sections that are presented. The wells that have Borehole Image Logs (BHI) available and were used to estimate the fracture intensity log are highlighted in red. The blind well used for validation is highlighted in yellow. The seismic sections that I investigated later are highlighted in black rectangles.

Table 2: Well inventory with the information of each wells have BHI logs available and the description of which wells are used in which part of the study. First part: Curvature analysis; Second part: Fracture intensity prediction. For the fracture intensity prediction, one well was used as a blind test. *A second blind test in a well outside the study area from a nearby field.

Well Inventory			
Well	**Logs**		
	BHI Available	**Curvature Analysis**	**Fracture Intensity Prediction**
A	Yes	Yes	Yes
B	Yes	Yes	Yes
C	Yes	Yes	Yes
D	Yes	Yes	Yes
E	Yes	Yes	Yes
F	Yes	No	No
G	Yes	No	Blind test
H	No	No	No
I	No	No	No
X	Yes	No	Blind test*

6.2. Curvature Analysis

Curvature attributes, as identified by Chopra and Marfurt (2007a; 2013), reveal both large- and small-scale structural features. Staples et al. (2013) and Chopra and Marfurt (2015) established correlations between the most-positive (k1) and most-negative curvatures (k2), fractured zones, and fracture density/intensity. In carbonates, Bravo and Aldana (2010) showcased the potential of curvature attributes for identifying sub-seismic faults and fractured zones at the well scale. Hunt et al. (2018) explored various window sizes and post-filtering techniques, emphasizing the need to align parameters with the targets' depth, scale, and geometry.

Therefore, following Hunt et al. (2018), I must consider the ultimate goal of the present study: to evaluate an optimum curvature that best correlates with the upscaled fracture intensity measured in the wells. The wells are located on a 4-way closure structural high with border faults at east and west. Some wells, such as Well B, are in the middle of the structure, in a flatter and relatively undeformed region. Other wells, such as Wells A, E, and D, penetrate build-up carbonate facies located near the border of the

structure and with evidence of intense fault activity. Away from the main structure, there are also smaller and isolated build-ups. Generally, the well paths avoid crossing the main seismic-scale faults for drilling security reasons. However, many of the wells pass through the damage zone of those faults or even smaller sub-seismic faults. Therefore, when parametrizing the curvature, I aim to highlight subtle small-scale features such as the sub-seismic faults and fractured zones. Ultimately, I want to do that both qualitatively and quantitatively. Figure 73 helps to illustrate the geological scenario of this study and what scale of curvature should correlate with the fracture intensity.

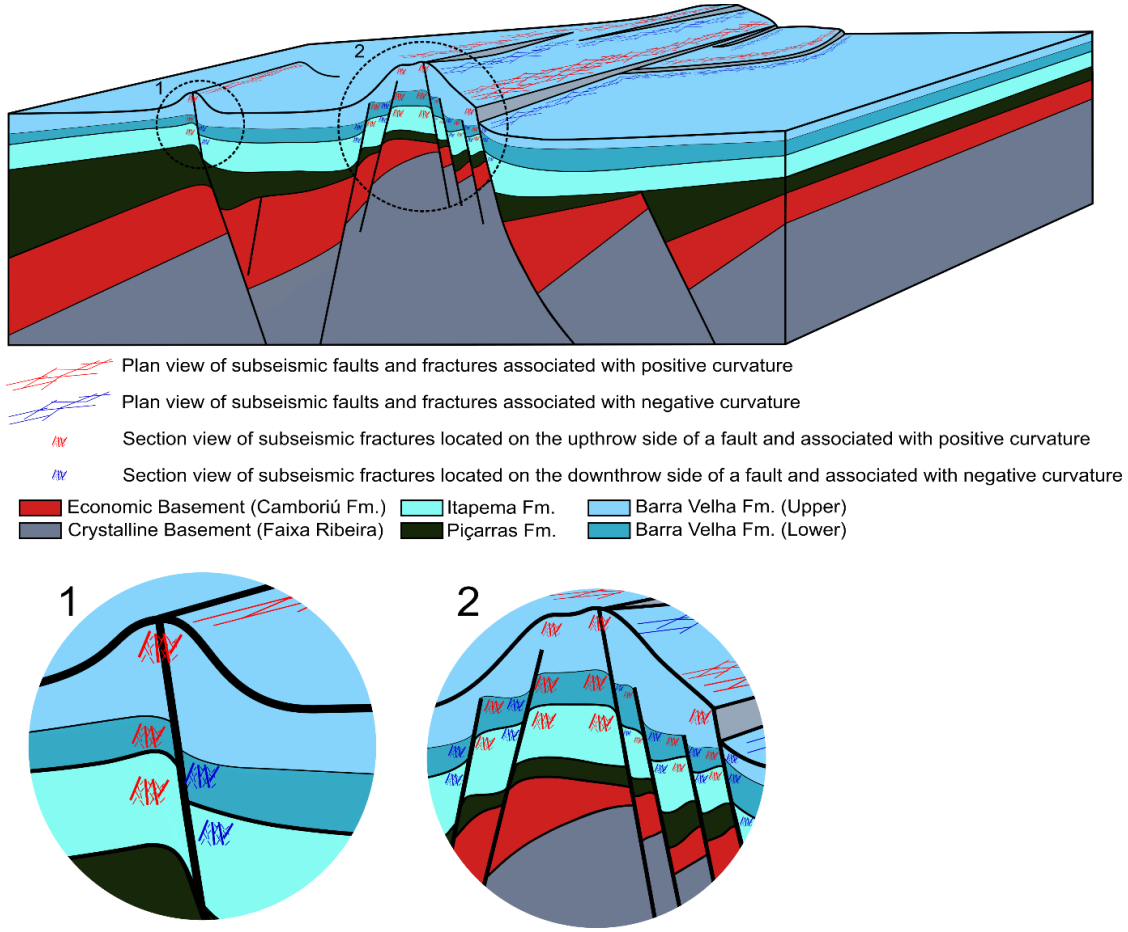

Figure 73: Box model of the geological scenario in the study area. The model helps to illustrate the scale of curvature that the present study aims to highlight. The study area is characterized by a regional 4-way closure structural high with border faults to the east and west. In the northern and southern parts of the high, there are heavily faulted carbonate build-ups. At the basement level, the faults are low-angle faults with a hundred-metre throw. Upwards, the faults are steeper, and the throw decreases significantly to the subseismic scale, making it harder for coherence and other discontinuity attributes to highlight them. The K1 and K2 curvatures will highlight the subtle flexures on the upthrow and downthrow sides of the faults,

respectively, which, in turn, are associated with subseismic fractured zones. A similar pattern occurs for isolated build-ups outside of the main high.

To achieve higher-resolution curvature, various methods are employed, including: (i) utilizing the preconditioned seismic amplitude volume (Chopra et al., 2011), (ii) estimating dip-fields precisely (Marfurt, 2006), or (iii) relying on amplitude curvature rather than structural curvature (Chopra and Marfurt, 2013). In this study, a combination of (i) and (ii) is applied, involving inverse Q filter seismic data preconditioning (Braga and Moraes, 2013) and three dip estimator algorithms using small windows. The dip estimators—Principal Component Analysis (PCA) (Randen et al., 2000), Fast Fourier Transform (FFT) (Tingdahl and De Groot, 2003), and Phase Gradient (PG) (Barnes, 1996)—generate dip volumes in both inline and crossline directions, that can be used to measure the dip in any azimuth. These dip volumes (Dip_PCA, Dip_FFT, Dip_PG) serve as inputs for curvature attribute estimation, as discussed in Chapter 4.

I perform two evaluations to choose the best output. The first is qualitative, based on visually inspecting the dip and curvature responses. For the curvature attribute, I compare the use of small and large window sizes to obtain two different scales of the most-positive curvature. These curvatures are called the short-wavelength curvature and long-wavelength curvature, respectively. These curvature attributes with different scales are referred to in this study, for example, as "Short_k1_FFT" and "Long_k1_FFT" for the most-positive curvatures provided by the dip from FFT (

Table **3**). I evaluate the use of post-filtering techniques based on median filters (MFs) and recursive Gaussian filters (RGFs) to attenuate noise in the dip estimation. To indicate these filtered versions of curvature, I add a lowercase suffix, i.e., mf or rgf, e.g., "Short_k1_FFTmf".

Afterwards, I constrain the analysis to the well information. This means that I extract the trace of the attributes under evaluation at the well location and compare it with the upscaled P32SUM well-log, which is the volumetric estimation of the fracture intensity based on BHI logs (Dershowitz and Einstein, 1988; Niven and Deutsch, 2010). Here, the suffix "SUM" indicates that this fracture intensity is the sum of the fracture densities of all the types of fractures picked in the interpretation of the BHI logs (open, semi-open, closed, partial, and unidentified).

I then measure Spearman's rank correlation coefficient (ρ) between the curvature attributes and the upscaled P32SUM log. These study methods are similar to the target-oriented approach used by Hunt et al. (2018). In addition, these results are compared with a dip-steered coherence attribute, which is one of the most commonly used attributes for seismic fault and fracture studies (e.g., in Liao et al., 2019; Tanaka et al., 2022; Babasafari et al., 2022).

Then, I plot the curvature against the upscaled fracture intensity, select the best final optimum most-positive curvature based on all previous results, and study two possible regression models. The first uses an exponential model, and the second uses a sigmoid function. The workflow is summarized in Figure 74. Finally, with the validation of the curvature as a proper attribute to identify the zones prone to fracturing and careful analysis of some well information, I perform a detailed study on the correlation and causality of the curvature and fracture intensity.

Finally, it is important to mention that while in Figure 73, both k1 and k2 curvature are exemplified, I evaluate only k1 in this study once preliminary analyses indicate a stronger correlation between the fracture intensity and k1. Although I acknowledge that this worse correlation of k2 can potentially be a simple statistical artefact due to biased sampling, I perform the study only with k1 for brevity and conciseness. I understand that such a fact does not affect the results of this study.

Table 3: Nomenclature pattern for each parametrization of the curvature attributes evaluated in the study. I add a lowercase suffix for the filtered curvatures, mf or rgf, e.g., "Short_K1_FFT_mf".

Most-Positive Curvature Attribute Nomenclature		
Dip	**Lateral Window Size (Inline, Crossline)**	
	(5, 5) = 125m x 125m	**(15, 15) = 375m x 375m**
Phase Gradient (PG)	Short Wavelength Most-Positive Curvature from PG Dip (Short_k1_PG)	Long Wavelength Most-Positive Curvature from PG Dip (Long_k1_PG)
Fast Fourier Transform (FFT)	Short Wavelength Most-Positive Curvature from FFT Dip (Short_k1_FFT)	Long Wavelength Most-Positive Curvature from FFT Dip (Long_k1_FFT)
Principal Component Analysis (PCA)	Short Wavelength Most-Positive Curvature from PCA Dip (Short_k1_PCA)	Long Wavelength Most-Positive Curvature from PCA Dip (Long_k1_PCA)

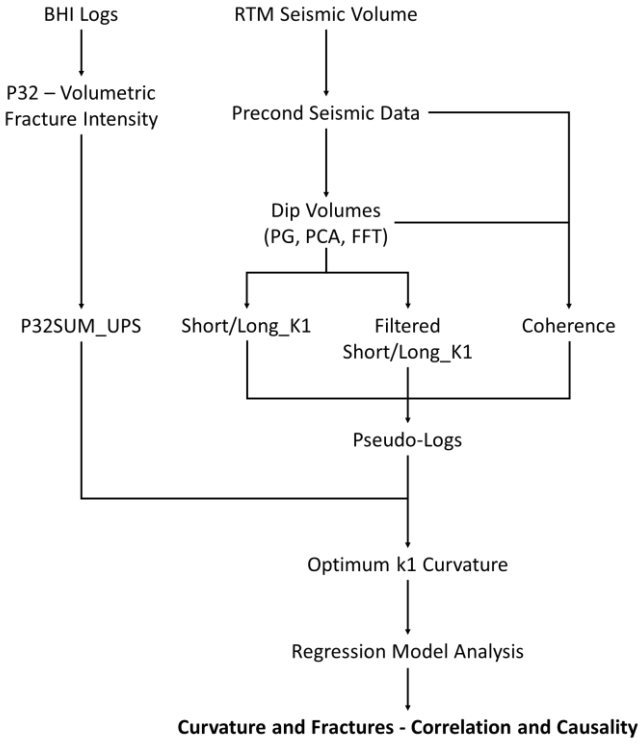

Figure 74: Summary of the workflow used in this study to evaluate the best geometric attributes that fit the upscaled fracture intensity log. By aggregating these attributes with rock and well data, I perform a correlation and causality study of the curvature and fractures.

6.3. Fracture Intensity Prediction

The workflow's second part aims to train a machine learning model that can predict the fracture intensity in the seismic volume. However, due to the time cost of such a task, conducting a 1D feasibility study prior is recommended.

I combined concepts from brittleness studies with curvature analysis and fault-damage zones to select the input features for the 1D fracture intensity prediction model. Silicified zones have a greater fracture intensity due to their greater brittleness and are commonly associated with highly fractured intervals in the presalt (Correa et al., 2019; Fernandez-Ibanez et al., 2022a; Fernandez-Ibanez et al., 2022b; Lupinacci et al., 2023, Wennberg et al., 2023), while muddy facies are less likely to develop fractures (Correa et al., 2019). Highly deformed regions tend to have a higher curvature. Areas close to seismic scale faults will have a higher fracture intensity towards the faults, which will generally decay exponentially away from them (Savage and Brodsky, 2011). Based on those principles, for this study, I chose as input features Young's Modulus (E), Poisson's Ratio (v), Silica Content (Si), Most-Positive ($k1$) Curvature, and a Distance to Fault attribute for predicting fracture intensity. The first three inputs are from the upscaled well logs, while the last two I extracted from seismic data at the well position. I compute the fault distance attribute from the fault probability volume estimated by the UNet model. The probability value 0.7 was the threshold for calculating the distance to a fault.

The faults were identified from a pre-trained 3D UNet fault predictor. The pre-trained network did not demand any human/expert labelling, preventing the process from human bias. However, a preliminary result on this step was considered unsatisfactory, probably due to the quite different aspect of the present seismic dataset from the training data set of the UNet model (Wu et al., 2019a). Thus, I applied the edge-preserving smoothing filter (EPS) (Luo et al., 2002) to the seismic data preconditioned with an inverse Q-filter to further enhance the sharpness of the faults.

I evaluated two tree-based ensemble predictors to predict fracture intensity: XGBoost (Chen and Guestrin, 2016) and Random Forest (Breiman, 2001). Tree-based ensembles are famous for overperforming neural networks and other supervised methods in tabular datasets (Grinsztajn et al., 2022). Previous geoscience studies also came to the

same conclusion (Bhattacharya et al., 2019; Man et al., 2021; Mandal et al., 2022; Nguyen et al., 2022).

I used five wells for this study and excluded intervals with igneous rocks. The number of samples used was 224. I randomly separated the dataset to evaluate the models using the 80/20 rule: 80% for training and 20% for testing.

I evaluate the performance of the models with regression analysis. To understand the importance of each input feature, I tested some feature importance methods. Among several methods of feature importance evaluated in the study, I selected the permutation importance (Altmann et al. 2010) and the SHAP (Lundberg and Lee, 2017) since this method better deals with collinearity between the features, as in the case of k1 curvature and distance to fault. Figure 75 illustrates a resume of the workflow followed in this work. All the workflow of the feasibility study was performed using Python, mainly the Scikit-learn, Pandas, and Numpy libraries.

One can realize that the silica content cannot be directly obtained in the seismic volume. The same can be said for Poisson's Ratio and Young's Modulus once I work with post-stack seismic data. Although, P-Impedance holds a high correlation with Young's Modulus. Therefore, I repeat the feasibility study, this time with k1 curvature, distance to fault, and P-Impedance as input features. Then, I compare the performance of this model to the one with five features using the regressions metrics and residual analysis. The 1D prediction of the feasibility study for all the wells is available in Appendix 11.3.

After validating the workflow, I train and apply a 3D machine learning model on OpendTect using the same three input features to predict fracture intensity in the seismic volume. In this part, I only used XGboost, as the previous stage of the feasibility study indicated it as the best regressor compared to Random Forest. In the next chapter, I evaluate this final result on a map and in some seismic cross-sections of some characteristic wells. The cross-sections of all the wells with the seismic attributes used in the 3D prediction, together with the prediction itself, are exposed in Appendix 11.4

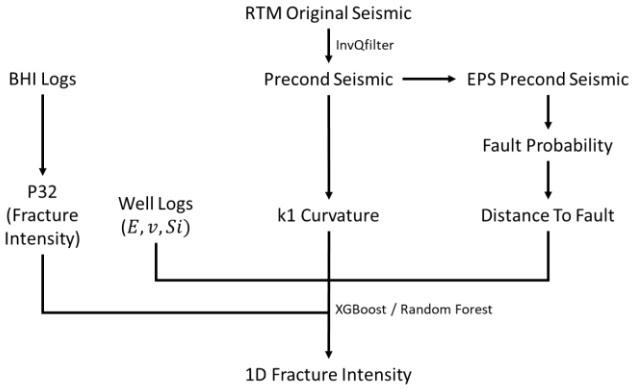

Figure 75: Illustration of the workflow to perform a feasibility study for fracture intensity modelling using machine learning.

7. Results and Discussions

7.1. First Part – Curvature and its Correlation with Faults and Fractures

7.1.1. Results

The seismic preconditioning by inverse Q-filtering enhances the lateral resolution, thereby obtaining a better definition of the seismic discontinuities that can be associated with faults. Before preconditioning, faults were barely inferred by some minor perturbance in the amplitude field and abrupt changes in the dip reflector (Figure 76). Regarding the different approaches for dip estimation and their curvature attributes, FFT (Figure 77-a, -d) presents better resolution and less noisy features than the PG (Figure 77-b, -e). The PCA (Figure 77-c, -f) is the most consistent, with less noise influence. However, it lacks the definition of small features compared to FFT.

Regarding lateral window size, there are two optimum curvatures with half-windows of (2, 2) and (6, 6) in the inline and crossline directions, i.e., (50 m, 50 m) and (150 m, 150 m), respectively. These curvatures are the previously defined short-wavelength curvature and long-wavelength curvature. I also compare the results of the curvature attributes with the coherence attribute (Figure 77-g). The latter highlights only the most prominent seismic discontinuities associated with neatly visible offsets of seismic events. The coherence attribute misses other discontinuity features, such as faults with throws near the limit of the seismic resolution, amplitude dimming, and bending of seismic events.

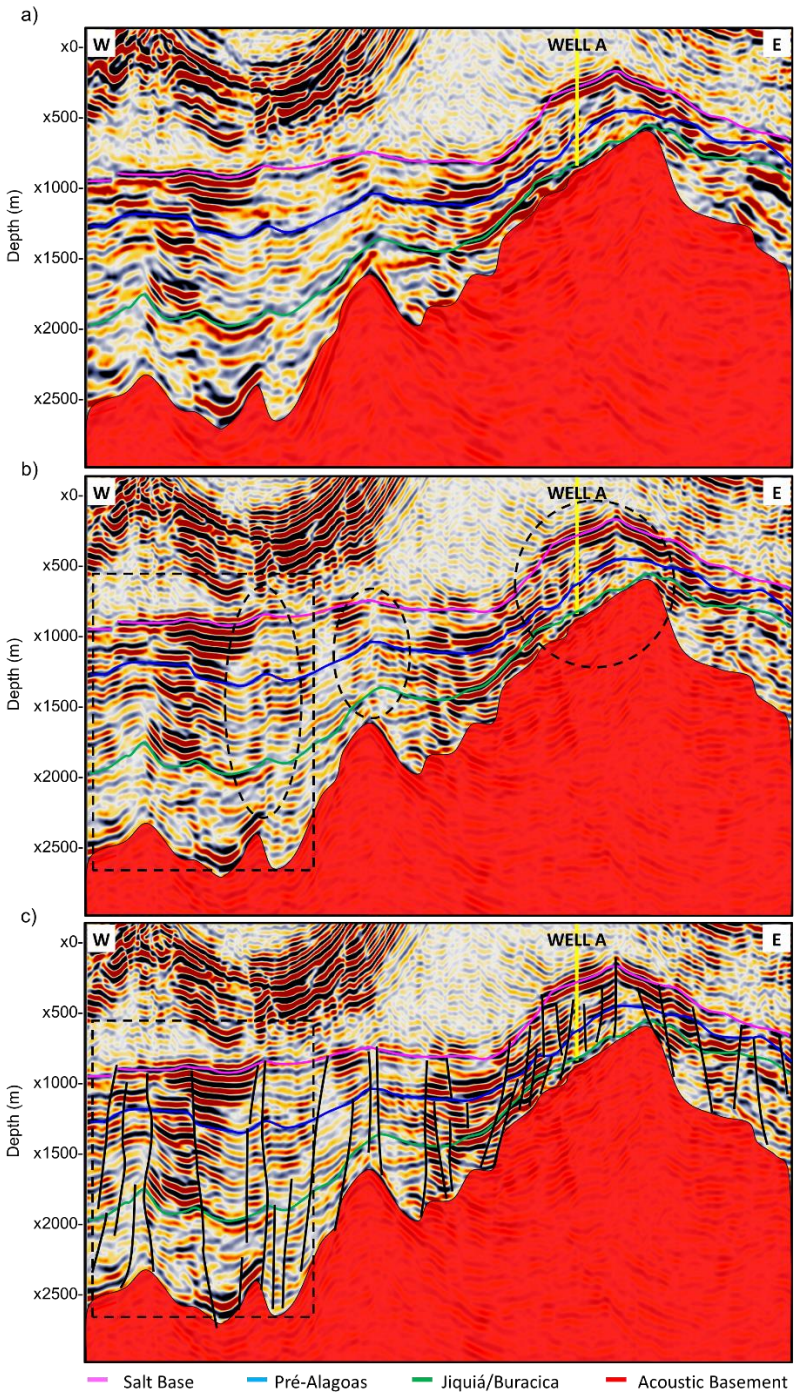

Figure 76: Cross-section A from Figure 72 with the seismic amplitude attribute before (a) and after (b) preconditioning with inverse Q-filtering. The dashed ellipses indicate zones where the seismic discontinuities associated with faults are better defined in the preconditioned seismic volume. The dashed rectangle highlights an area where the different evaluated curvatures are further presented. c) Preconditioned seismic section with interpreted faults for further reference

The filtered curvature attributes with different parameterizations are presented in a seismic section in Figure 78. The RGF (Figure 78-b, -d, -f, -h, -j, -l) produces less noisy

results, highlighting the more prominent trends (longer wavelengths) of the curvature, while the median filters better preserve the small-scale detail (Figure 78-a, -c, -e, -g, -i, -k). This result is expected due to the nature of Gaussian and median filters. Generally, in applications such as the one used in this study, Gaussian filters resemble a moving average filter with a large window size (Hale, 2002). As such, the choice of one over another should be driven by the feature of interest. In the present case, the main goal is to find the most-positive curvature that best fits the fracture intensity property. Therefore, I must move my analysis to the well domain.

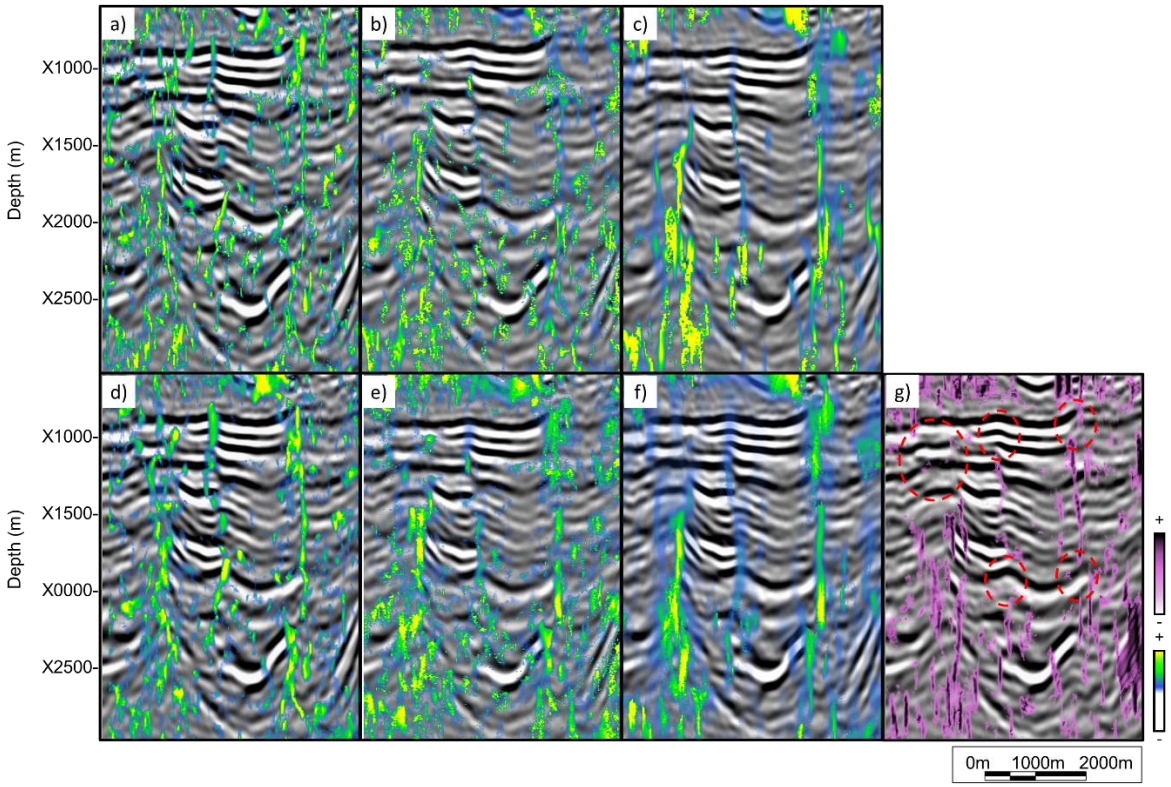

Figure 77: Seismic panels of the area highlighted in Figure 76 with the evaluated geometric attributes a) short_k1_FFT; b) short_k1_PG; c) short_k1_PCA FFT; d) long_k1_FFT; e) long_k1_PG; f) long_k1_PCA; g) coherence. Red dashed circles indicate zones where the coherence has failed to define the seismic discontinuities interpreted as faults or deformed zones prone to the presence of fractures. FFT and PCA dips produced the best curvature attributes.

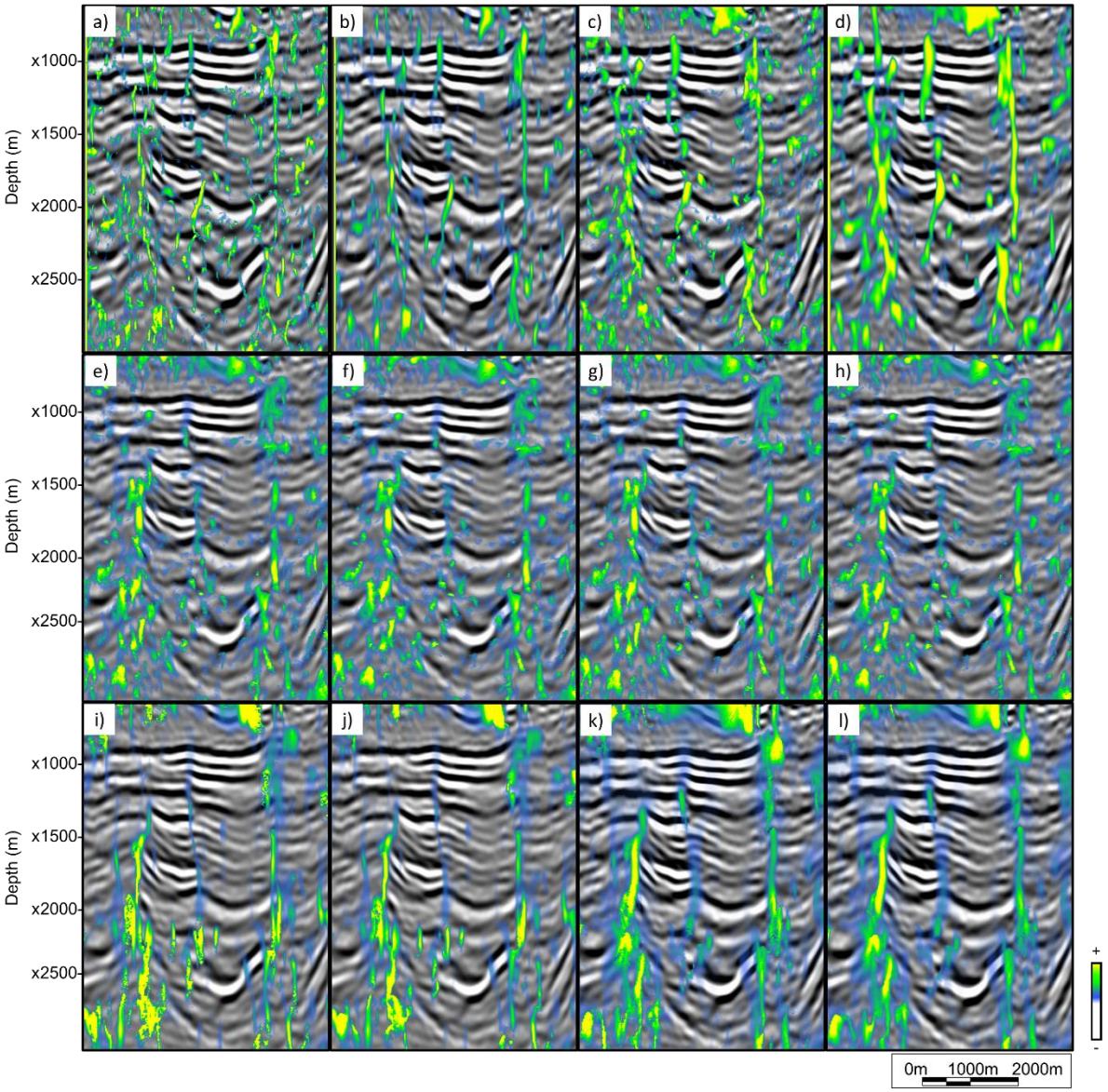

Figure 78: Seismic panels of the area highlighted in Figure 76 with the evaluated filtered curvatures. a) Short_k1_FFTmf, b) Short_k1_FFTrgf, c) Long_k1_FFTmf, d) Long_k1_FFT_rgf, e) Short_k1_PGmf, f) Short_k1_PGmf, g) Long_k1_PGmf, h) Long_k1_PGrgf, i) Short_k1_PCAmf, j) Short_k1_PCArgf, k), Long_k1_PCAmf, l) Long_k1_PCArgf. RGF strongly smooths the curvature response, thus attenuating noise but losing some of the small-scale definition. MF better preserves this level of detail.

This is the most fractured well among the wells that have BHI logs. The coherence attribute highly correlates with the fracture intensity log in the Itapema Formation interval. Previous studies indicated that faults and fractures play a meaningful role in the porosity–permeability system of the Barra-Velha Formation (Correa et al., 2019; De Ros, 2021), while in the Itapema Formation, the permeability and porosity are much more

linked to the matrix and dissolution features, due to the original texture and diagenetic history of those rocks (Chinelatto et al., 2020).

When the analysis is restricted to the Barra-Velha Formation, the curvature attributes correlate more to the fracture intensity log than the coherence. The curvature with the better ratio between the presence of noise and small-detail definition is the long wavelength from FFT (track 5), where we can adequately correlate the picks of the volumetric fracture intensity with the curvature attribute. Nevertheless, the PCA curvatures present satisfactory results to highlight the larger trends. For the filtered curvatures (Figure 80), the better result is achieved with the long wavelength FFT curvature, both the median and recursive Gaussian filter (tracks 7 and 8) and by the short wavelength PCA curvature, also with both filters (tracks 5 and 6). The median filter produces almost the same result for the PCA curvature. This phenomenon can be related to the predominance of the low-frequency content in the PCA curvature (Figure 79 – tracks 4 and 7). Therefore, the attribute has no high-frequency event that the median filter would better preserve, such as with the FFT curvature.

In this particular well (Well A), the filtered PG curvature correlates reasonably with the fracture intensity, especially the long wavelength with a recursive Gaussian filter (Figure 80 – track 10). However, as previously mentioned, this curvature is the noisier curvature in seismic sections. In addition, this good correlation is not constant for all the wells. The statistical analysis later shows that PG curvature has the lowest overall correlation with fracture intensity among the evaluated curvatures.

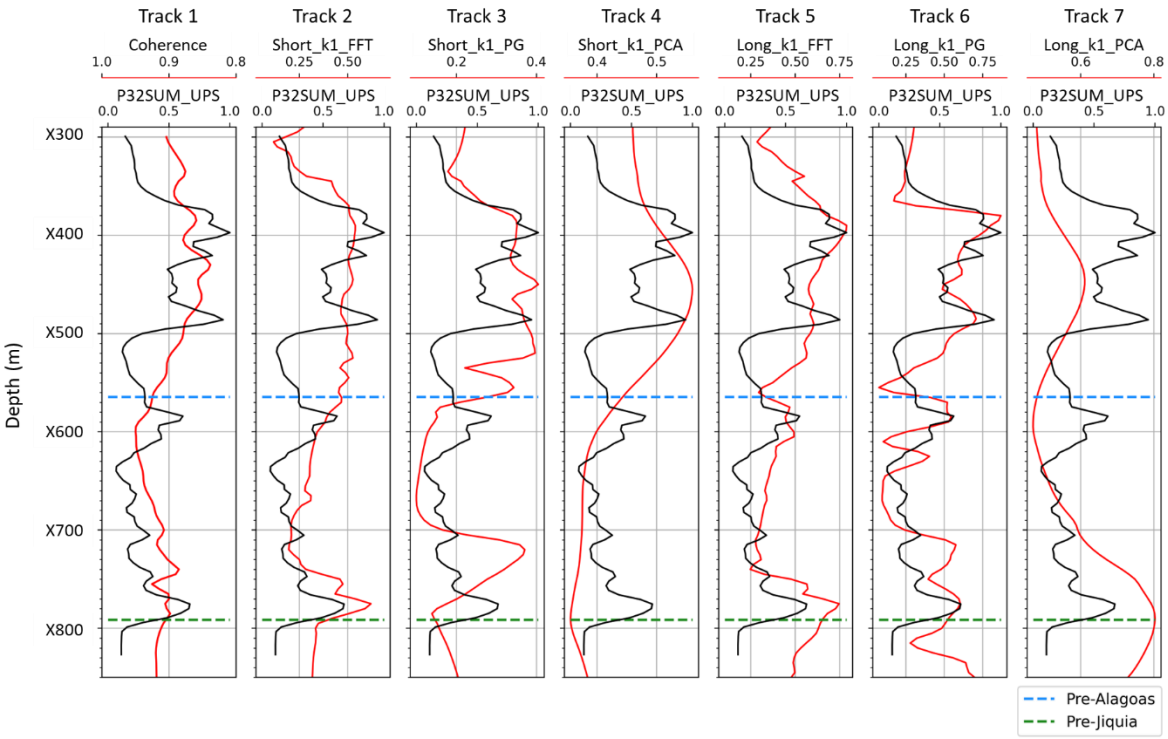

Figure 79: Well-track for the Barra Velha and Itapema Formations in Well A with the pseudologs of the coherence and the unfiltered curvature evaluated (red curves) together with the upscaled volumetric fracture intensity log (P32SUM_UPS) (black curves). From left to right (track 1 to 7): Coherence, Short_k1_FFT, Short_K1_PG, Short_K1_PCA, Long_k1_FFT, Long_k1_PG, Long_k1_PCA.

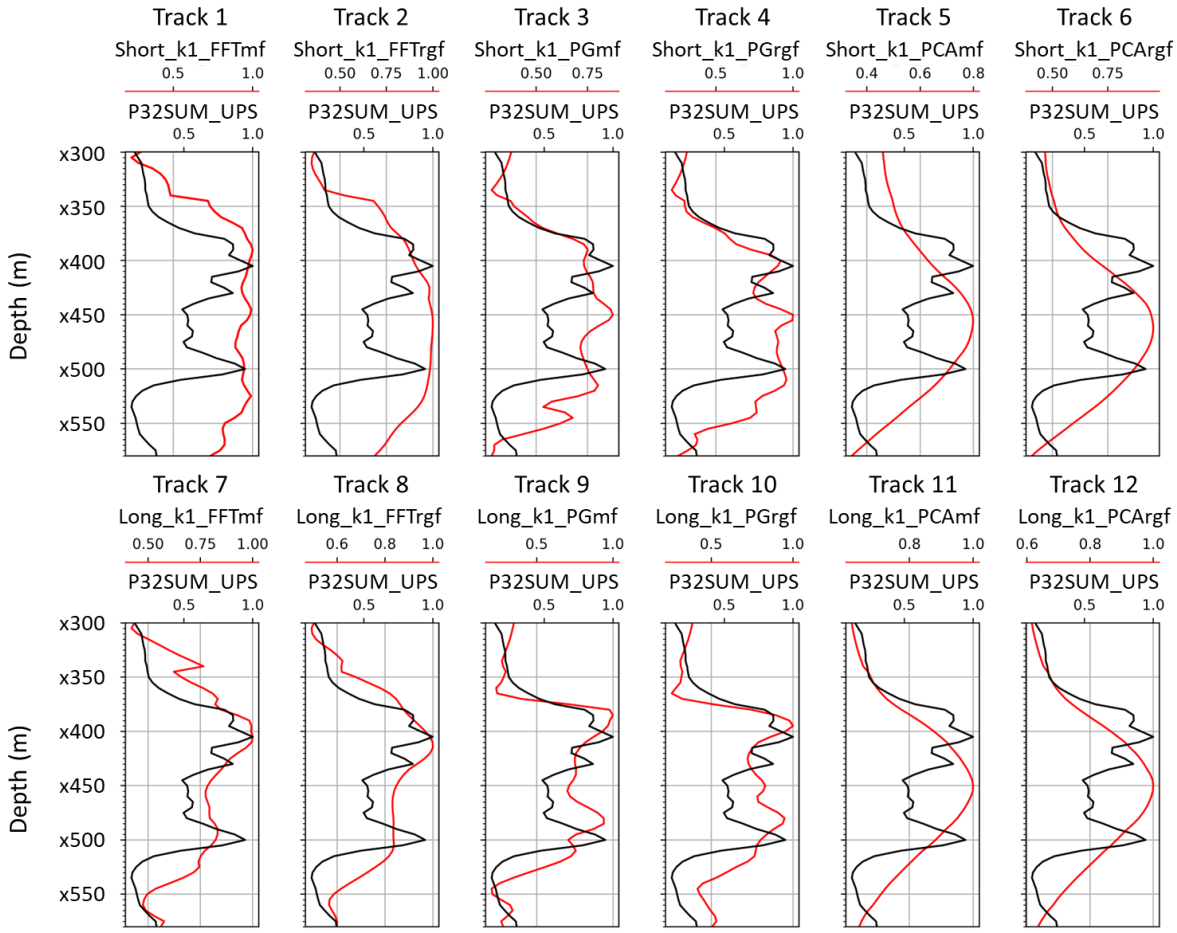

Figure 80: Well-track for the Barra Velha Fm. of the Well A with the pseudologs of the filtered curvature evaluated (red curves) together with the upscaled volumetric fracture intensity log (P32SUM_UPS). From left to right and top to bottom (tracks 1 to 12): Short_K1_FFTmf, Short_k1_FFTrgf, Short_k1_PGmf, Short_k1_PGrgf, Short_k1_PCAmf, Short_k1_PCArgf, Long_k1_FFTmf, Long_k1_FFTrgf, Long_k1_PGmf, Long_k1_PGrgf, Long_k1_PCAmf, Long_k1_PCA_rgf.

I also evaluate the curvature in seismic sections passing through Well B, which is located in an area less affected by fractures, as indicated by the interpretation of BHI logs and the rock samples. Figure 81 shows that the well reaches a carbonate platform with little response from the most positive curvatures, suggesting the absence of fractures. Near the well, there are deformations associated with structures with a most-negative curvature shape. In general, during this study, I observed that the most-positive curvature presents a better correlation with fractured zones. These previous analyses agree with the BHI log information (Figure 82). The most-positive curvature stays at a low level, together with the volumetric fracture intensity. Unlike the Well A case, in Well B, the curvature attribute could not define the picks of fracture intensity.

Nevertheless, the attribute is shown as an envelope surface, indicating a trend to the zones that are more and less prone to develop fractures. Two exceptions are the top and bottom intervals of the formation. These local mismatches between the curvature and the fracture intensity are discussed further in the study.

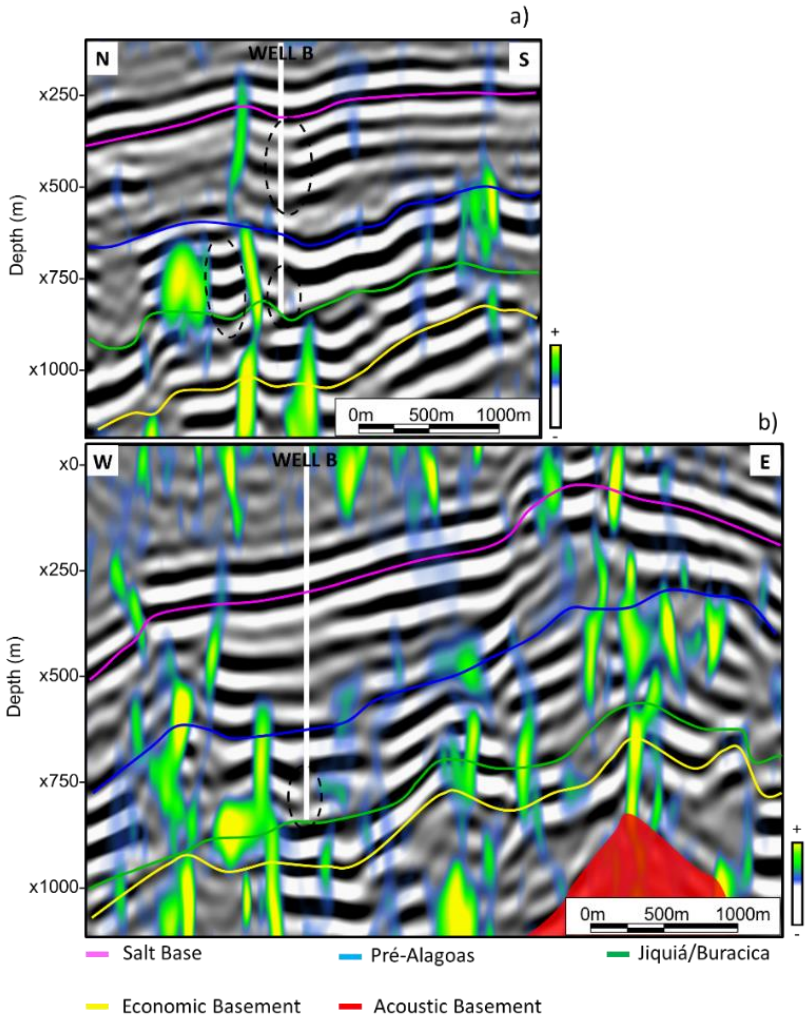

Figure 81: Inline (a) and crossline (b) with the short-wavelength most-positive curvature with a recursive Gaussian filter from FFT in green levels. Dashed circles indicate zones of negative curvature deformation.

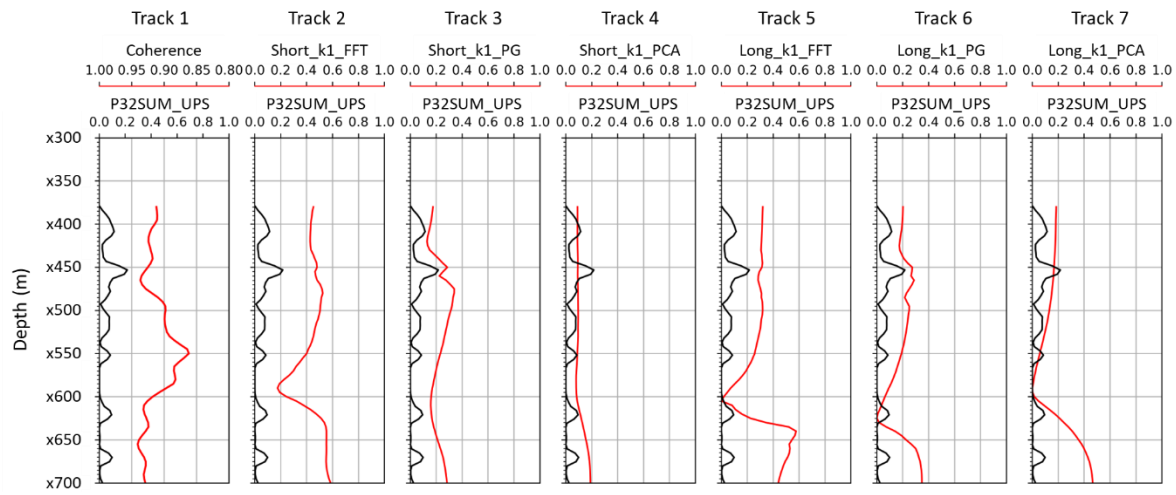

Figure 82: Well track for the Barra Velha Formation of Well B with the pseudologs of the coherence and curvature attributes evaluated with the upscaled volumetric fracture intensity log (P32SUM_UPS). From left to right (track 1 to 7), Coherence, Short_k1_FFT, Short_k1_PG, Short_k1_PCA, Long_k1_FFT, Long_k1_PG, Long_k1_PCA.

Before starting the second part of the curvature attributes and P32SUM evaluation, I removed the samples related to igneous rocks from the data frame. This step had to be performed, as igneous rocks have different rheologies compared to carbonates. The igneous-related samples generally have a high fracture intensity at low curvature attribute values. Additionally, there was no apparent correlation between these two properties for those samples, i.e., the fracture intensity increases with no change in the curvature. Consequently, the correlation coefficients worsen and negatively impact the model fit.

I use Spearman's rank correlation (ρ) instead of the standard Pearson's correlation (r) since there is a nonlinear relationship between the evaluated attributes and the fracture intensity log (Figure 83). The rank correlation confirms that the long wavelength FFT is the best attribute that correlates with P32SUM. Additionally, I demonstrate that the curvature attribute correlates better with fracture intensity than the coherence attribute. The poor correlation of the coherence attribute is basically induced by low to medium levels of fracture intensity in less-fractured wells, such as in Well B, Well D, Well E, and Well F. In all those cases, the correlation of the attribute with the fracture intensity behaved in an opposite sense from the expected. In those wells, the coherence has a weak direct correlation with the fracture intensity. In general, I expect that coherence would be inversely related to the fracture intensity even though part of this effect is not accounted

for in the ρ measurement presented in Figure 83 since I removed the Well F from the analysis.

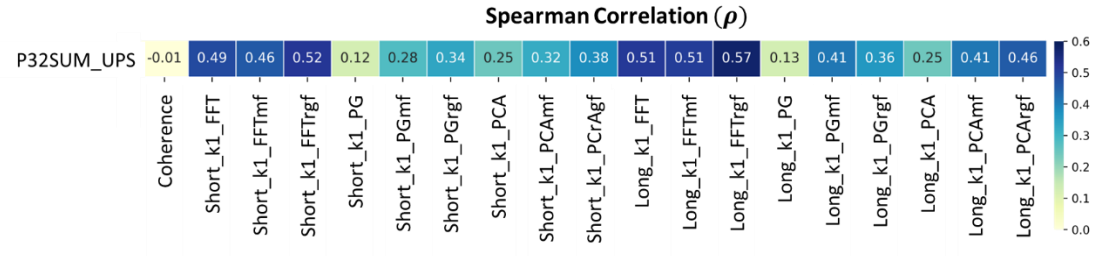

Figure 83: *Spearman's rank correlation (ρ) of the curvature, coherence and the upscaled P32SUM well-log.*

Spearman's correlation analysis and the graph in Figure 84 show that the curvature attribute has a positive monotonic relationship with the fracture intensity log. Two equations were fitted: the first with an exponential function and the second with a sigmoid function. By analysing the Spearman correlation, the well tracks, and the scatter plots, I selected the long wavelength curvature from FFT as an input feature to the models. In both equations, I chose only two parameters, a and b. The P32SUM log and curvature attribute were normalized in this step using min–max scaling to the same scale (0 to 1). I then fit the models using a non-linear least squares problem solver based on the Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) available in the SciPy library of Python.

The optimum parameters for the exponential model are a = 0.019 ± 0.003, b = 4.073 ± 0.186, and for the sigmoid model, they are a = 9.789 ± 0.581, b = 0.731 ± 0.006. The exponential model comprises a large part of the data with the uncertainty band. In contrast, for the sigmoid model, the uncertainty analysis produces a narrow region comprising fewer data. This behaviour is due to the nature of the exponential model, in which the uncertainty band is wider at the right end of the scale, at higher curvature and fracture intensity values. As four wells with distinct characteristics were used, I evaluated the fitting of the models using a standard data science approach, which consists of randomly dividing the dataset into 80% for training and 20% for testing.

The two models present a similar behaviour, with a strong bias in the residuals close to the origin, shifting to a more random distribution for higher values. Nevertheless, there are some differences. For example, the exponential model presents a worse fit

(higher residual values) for the intermediate values of the normalized curvature attribute. These higher residuals are primarily due to a cluster of points that are visible in the scatter plot of the curvature with the P32SUM logs in the region near 0.8–1.0 (y-axis) and 0.6–0.8 (x-axis). These points are detached from the main trend in both models and are not even comprised by the wider uncertainty band of the exponential model. These samples are related to Well A, where in Figure 79, I observe that while the fracture intensity reaches a peak (near X500 m), the curvature is at medium-to-high values. The reasons for those local mismatches between the main trend of fracturing related to curvature are discussed further in the article. I can confirm the previous analysis by adding regression metrics (Table 4), with the sigmoid model performing better. It has better metrics, a better fit, and a better distribution of the residuals. The single region where the exponential model performs slightly better is at low levels of fracture intensity. In contrast, the exponential model accounts for a small quantity of the early increase in the fracture intensity with respect to curvature.

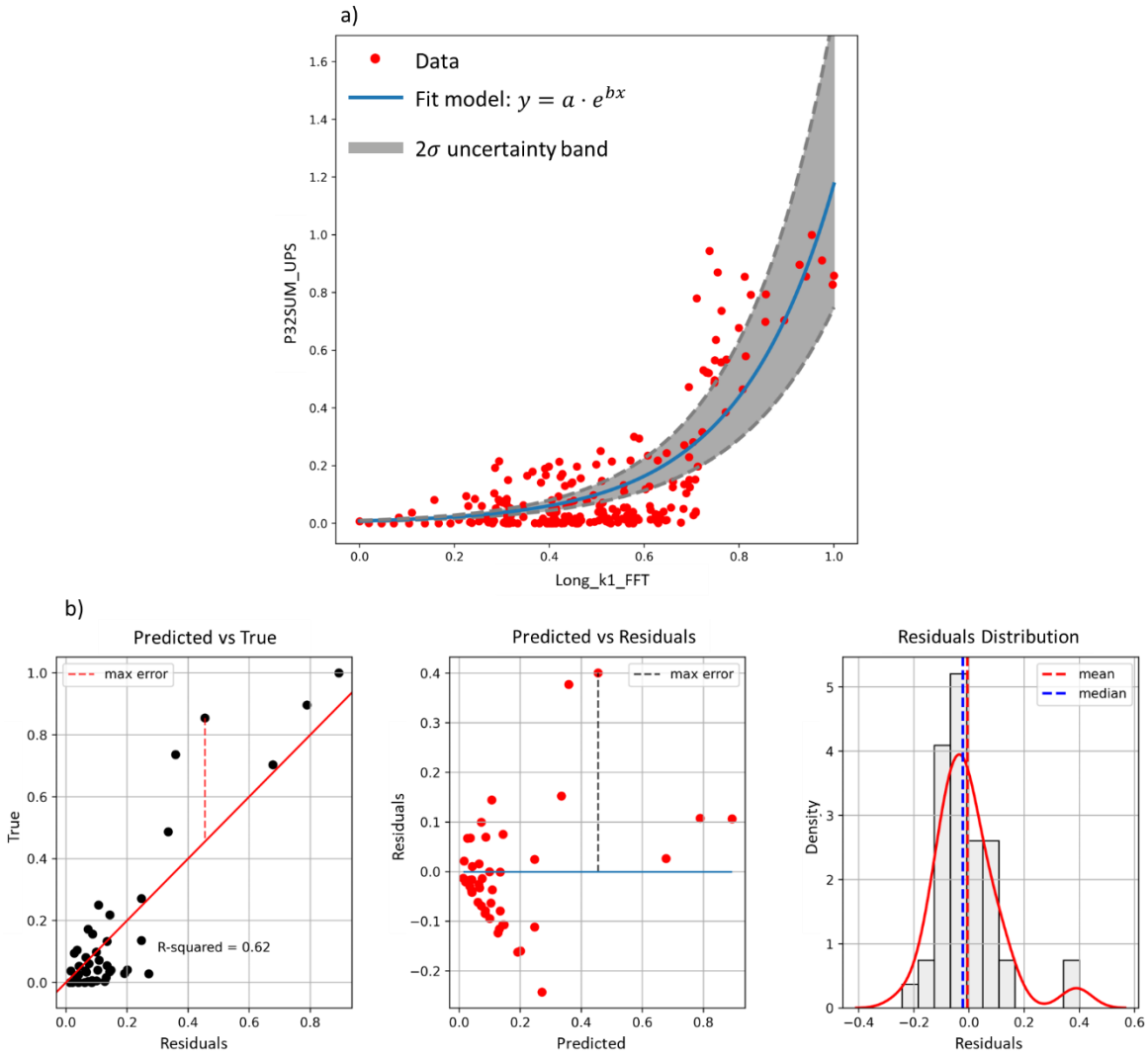

Figure 84: *a) Scatter plot (for the whole dataset) of the long wavelength most positive curvature from FFT and the upscaled P32SUM log with the fitted exponential model and the upper and lower limits of the model produced by the uncertainty of the parameters (plus and minus one standard deviation). b) Residual plots of the fitted exponential model for the test dataset only.*

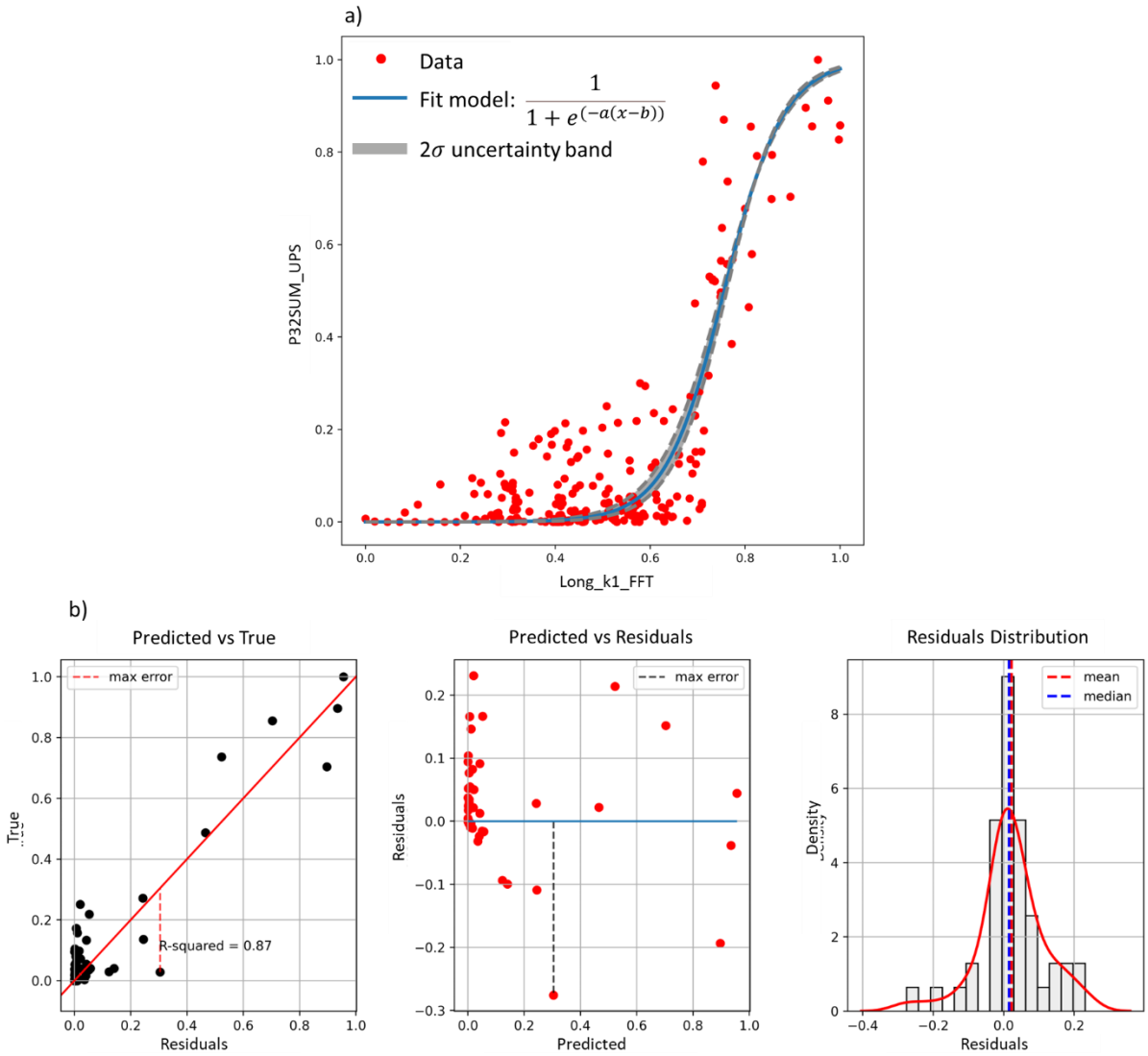

Figure 85: *a) Scatter plot (for the whole dataset) of the long wavelength most positive curvature from FFT and the upscaled P32SUM log with the fitted sigmoid model and the upper and lower limits of the model produced by the uncertainty of the parameters (plus and minus one standard deviation). b) Residual plots of the fitted sigmoid model for the test dataset only.*

Table 4

Regression metrics for the testing set from both models, exponential and sigmoid. MAE: Mean Absolute Error; RMSE: Root Mean Squared Error; R2: R2 score.

Metrics	**Model**	
	Exponential	**Sigmoid**
MAE	0.08	0.06
RMSE	0.12	0.09
R2	0.62	0.87

7.1.2. Discussions

7.1.2.1. Coherence and Curvature Attributes

In seismic attribute workflows for fault and fractured zones recognition, the two main drivers of the quality of the final product are the noise level and frequency content. In general, the signal/noise ratio of the seismic data used in the study is adequate. Therefore, I had to consider only proper seismic preconditioning to enhance the high-frequency content through inverse Q-filtering. The coherence attribute seems to be more influenced by the lower frequency band found in the presalt carbonates than the curvature attributes. In both seismic datasets, the original and preconditioned coherence attribute purely detects faults with a throw close to or above the seismic resolution.

In contrast, the curvature attribute detects small-scale changes in the dip and amplitude fields. Inverse Q-filtering plays an essential part in the presented workflow, improving the definition of the present targets. This better accuracy of faulted and fractured structures by higher frequencies is well demonstrated by previous works such as Botter et al. (2014) and Chen et al. (2020).

7.1.2.2. The role of geological curvature in fracturing

Previous works have studied the role of folding, faulting, and fracturing in the relationship between curvature and fracture density/intensity (Chopra et al., 2013; Keating and Fischer, 2008). Chopra et al. (2013) demonstrated that fracture intensity as a function of k1 curvature behaves as a 3-phase curve of elastic deformation. The first stage is uncorrelated (curvature increases in the absence of fractures), the second stage has a linear correlation, and the third stage is called fracture saturation, where fractures no longer accommodate the strain. The deformation starts to move to the fault domain.

Taking that situation as an analogue, the sigmoid model is more appropriate in the present approach since, in the exponential model, the fracture intensity increases infinitely with the increasing curvature attribute. Such a phenomenon is not geologically possible. These particular situations of the poor correlation of curvature and fractures or curvature as a weak proxy for strain are also well discussed by Keating and Fischer (2008). The authors concluded, similarly to Chopra et al. (2013), that at the initial stages, the maximum curvature does not match perfectly with the presence of fractures. However, they demonstrated that this mismatch is dip- and throw-dependent. Generally, the

correlation is better for vertical to semivertical faults than for low-angle faults and for large-throw than small-throw faults. In the present study area, most of the fractures are vertical to semi-vertical fractures. Based on image logs, only a tiny portion of the fracture set has dips lower than 60° (Fatah et al., 2021b).

7.1.2.3. The poor correlation and non-correlation at all between some fractures and curvature

Some works have pointed out that curvature, in some cases, is not an optimum proxy for fracture intensity (Allwardat et al., 2007; Pearce et al., 2011; Shaban et al., 2011). These previous works argue that curvature is a measurement of the present-day structural form of a surface. Therefore, the curvature would ignore the past location of a maximum strain on a given surface, which, for example, does not necessarily match the actual hinge line of an antiform or a synform. However, this does not seem to be the case for the scenario and scale the present study investigates. As the model in Figure 73 and the results in Figure 77, Figure 78, and Figure 81, the scale of deformation that the k1 curvature is highlighting is more linked to small flexures in the upthrow of a fault (or downthrow of the k2 curvature) than associated with a broad hinge zone of a regional antiform. Additionally, as mentioned previously, Keating and Fischer (2008) addressed that the fault dip is one of the main factors that diminishes the correlation of curvature with the strain process. In detail, this means that curvature correlates poorly with the maximum strain for low-angle faults, such as the basement-involved faults of the lower rift. For steeply dipping faults, such as the vertical to semivertical faults of the Alagoas stage, the location of maximum strain does not change substantially over time. This phenomenon also helps to accumulate the strain in an individual location, which can produce a higher fracture intensity due to the overprinting of the deformation process.

Regarding the deformation process, I already discussed that the curvature and fracture intensity have a poor correlation at the initial stages of deformation. As the deformation increases, a certain level of stress–strain is reached, the curvature follows, and fracturing starts. However, I had to rule out a well (Well F) because the fractures increase without curvature variation. There are several possible explanations for this. The first is the resolution of the seismic attribute, which, despite the preconditioning and the upscaling of the fracture intensity log, still does not match the resolution of the upscaled well information. Most of the fractures in this well are partial fractures (Figure 86); this

could indicate that the well is at the tip of a fractured zone, with smaller faults and fractures that are difficult to associate with dense and well-developed fractured zones that can be recognized at the seismic scale. However, smaller faults could also imply thinner beds. Another plausible reason is that not all fractures are related to faulting and folding. For example, some deformation bands, such as pure-compaction bands, are more of a function of the overburden than tectonics. There are also fractures induced by hydrothermal activity and fluid precipitation, which likely occur more predominantly near the faults, but not always (Ren et al., 2019). All these types of fractures do not have any relationship with the type of stress–strain process that produces the faults and flexures that the curvature attribute highlights. Therefore, I acknowledge that some low-curvature areas can also have moderate to high fracture intensities.

The only rock data I have for Well F are the well cutting samples. Hence, it is difficult to study the nature of the structures in detail. However, I can make some observations by interpreting the well logs. Figure 86 shows that the well has a higher fracture intensity at the base of the formation. Additionally, the fracture intensity peaks are associated with silica content peaks. Moving upwards, the fracture level and the silica content decrease while the gamma ray increases. This suggests higher contents of fine sediments and organic matter, which have a more ductile rheology, thus hindering the development of fractures. The lowest level of fracture intensity at the igneous interval is an artefact after the high presence of breakouts prohibits us from picking the natural fractures with confidence (the same occurs for other wells such as Well B). The fracture intensity increases again near the top of the whole interval as the gamma ray level decreases. This control of fractures due to the change in the facies association is known as mechanical stratigraphy and is discussed in detail in the following section. At this point, what I can conclude from this well is that when the suitable facies condition was present, fractures were present in this interval. This helps explain the medium level of curvature and medium to low fracture intensity and the reason that fractures start to develop without any increase in the curvature. The scale and facies dependence for the correlation between curvature and fractures is well addressed by previous works, such as Stephenson et al. (2007).

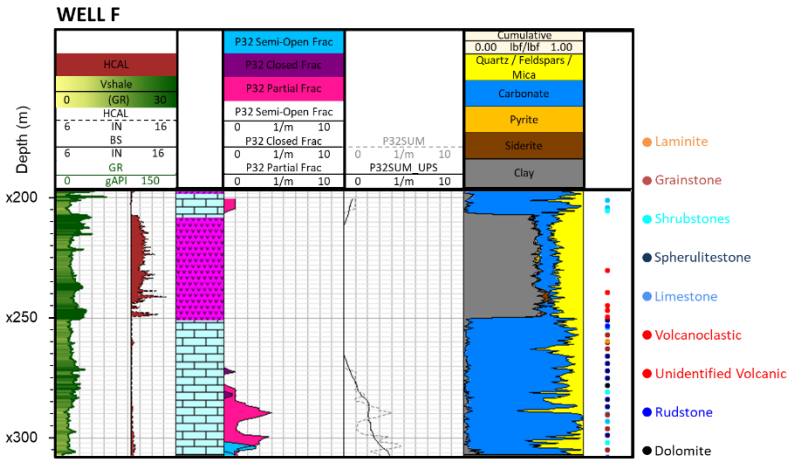

Figure 86: Well log track for Well F showing the calliper (Cal), bit-size (BS), gamma-ray (GR), lithological profile, volumetric fracture intensity (P32) from the classified fractures (semiopen, closed and partial), total volumetric fracture intensity (P32SUM), upscaled total volumetric fracture intensity total porosity (P32SUM_UPS), elemental capture spectroscopy logs (ECS), and the well cutting sample descriptions.

7.1.2.4. Mechanical stratigraphy and the differences between Itapema and Barra Velha Fm.

I do not evaluate the workflow in the Itapema Formation. for two reasons: the first reason, which has already been mentioned, is that faults and fractures play a minor role in its porosity–permeability system, but also because I observed at least three different fracture trends related to the curvature in the Itapema Formation. Therefore, given the curvature, I could not fit only one function that would model the fracture intensity. These trends may indicate a feature associated with the age of the formation and its collection of facies. Being an older formation than the Barra Velha Formation, the Itapema Formation experienced a more significant period of deformation by tectonics and a longer history of diagenesis. These two variables together resulted in a complex arrangement of the fracture network.

Regarding the facies assemblage, at the flanks of the structural high, the formation is predominantly composed of reworked bivalves with medium to low competence for fracturing due to their grainy texture. However, throughout the formation, there are also organic-rich shale rocks, clayey carbonates, and tight carbonates with a cement rich in quartz, calcite, or dolomite (Moreira et al., 2007; Chinelatto et al., 2020; Barnett et al., 2020; Oliveira et al., 2021). This considerable heterogeneity of facies produces a heterogeneity of mechanical properties, which is translated into the mechanical

stratigraphy of the interval, a common phenomenon in carbonates (Wennberg et al., 2006; Ferrill and Morris, 2008; Ferril et al., 2017).

In the Barra Velha Formation, the silica content seems to be one of the main mechanisms controlling fracture development. I already demonstrated that at Well F, the higher levels of fracture intensity are mainly associated with the picks of the silica content. The same pattern is shown in the Well E cores (Figure 87), with fractures and faults associated with silicified intervals. In addition, the fractures and faults are younger than the silicification. This result indicates that the silicification increased the brittleness of those intervals and helped to promote faulting and fracturing. Additionally, Well A has the highest silica content (Figure 88) and, not by coincidence, is the most fractured well. Former works such as Wennberg et al. (2021), Fernandez-Ibánes et al. (2022a), and Lupinacci et al. (2023) also observed the impact of the silica content in the control of fracture development. Therefore, in future works, I must attempt to identify the silicified intervals to predict these anomalous zones of fracture intensity. Mello and Lupinacci (2022) demonstrated that quartz/silica-dominated facies are distinguishable through seismic elastic properties. Therefore, as in Hunt et al. (2011), future works should aim to integrate inverted elastic moduli with curvatures to define the fractured zones properly. If such a workflow succeeds, one can adequately estimate the fractured zones by considering the effect of mechanical stratigraphy. Such a product would be a great advantage in fracture modelling routines to characterize excess-permeability zones and prospect naturally fractured reservoirs.

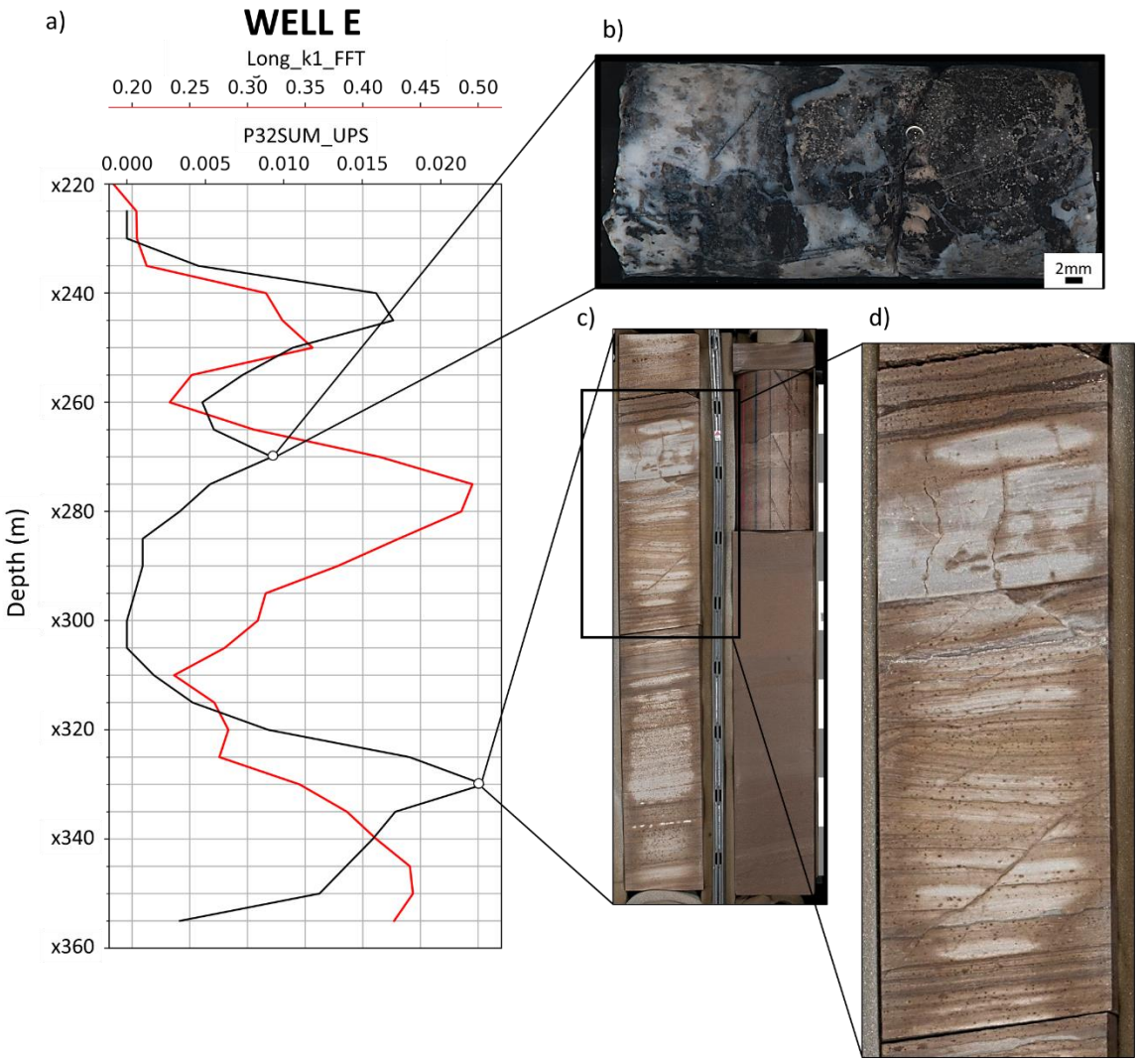

Figure 87: a) Well E, with the P32SUM log and pseudo log plots of the long wavelength most positive curvature from FTT dip. b) Picture of a sidewall core from the upper portion of the interval that is highly silicified with the presence of fractures. c) Picture of the well core showing the effect of mechanical stratigraphy induced by the silica content in the lower portion of the interval. d) Detail of d), highlighting the increased fracturing in the silicified zones.

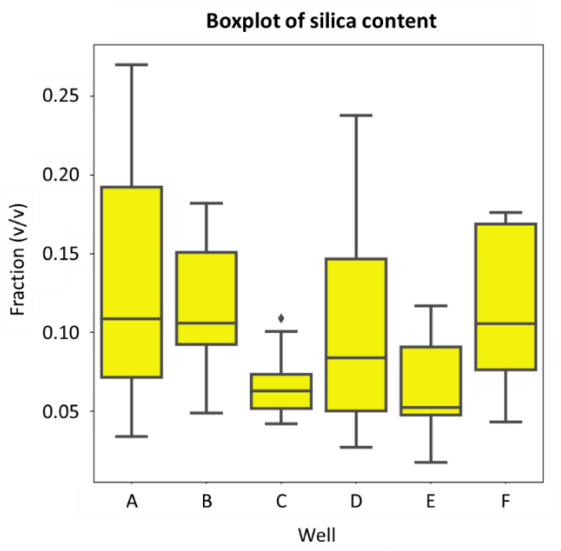

Figure 88*: Boxplot of the silica content grouped by the wells with BHI logs (Well A from Well E). Well A is the richest well in silica content, and also the most fractured well.*

7.1.2.5. The impact of the igneous rocks in the development of fractures, and its association with silicified intervals

The study area exhibits numerous igneous bodies across the entire presalt region. Within the Itapema Formation, igneous extrusions associated with paleo craters and fissure eruptions are prevalent, particularly in close proximity to significant faults (Zhao et al., 2024). In the Barra Velha Formation, intrusive bodies primarily constitute the igneous facies. These intrusive events can induce a high degree of fracturing, especially in the vicinity of the intrusion sites, leading to vertical fractures.

Furthermore, hydrothermal activity linked to the igneous events contributes to an influx of diverse minerals into the lake. This process can promotes the precipitation of silica and other minerals, resulting in an increased stiffness and brittleness of the rock. In the event of subsequent deformation caused by fault reactivation, the rock becomes more conducive to the development of fractures.

The feedback mechanism is suggested in Figure 87, wherein well E is situated near extensive faulted zones and igneous intrusions. As forementioned, this well displays a recurrent presence of fractures that cross-cut silicified intervals. This feedback loop suggests that substantial faults facilitate the intrusion of igneous rocks, which, in turn, promotes silica precipitation, ultimately intensifying the degree of fracture development.

7.1.2.6. How can the models be improved?

There is an issue in the workflow that is intrinsic to the context of the study area. presalt fields are enormous areas of hydrocarbon accumulations that are relatively underdeveloped. Compared to more mature plays, the amount of well data is restricted. For the seismic data, despite recent advances in seismic imaging and velocity modelling workflows, there are still some pitfalls concerning the resolution, seismic acquisition, and illumination in the presalt interval. These questions put some limits on seismic studies that try to explain, on such a scale, features observed in the well. In addition to these more technical problems, the geologic context is overly complex, with high spatial-temporal heterogeneity of facies, diachronic tectonic evolution, and large gaps in the geologic record due to recurrent fault activity or base-level variation with sudden episodes of subaerial exposure and later erosion. Despite all these challenges, I found two possible functions that would explain the correlation between the curvature and fracture intensity of the Barra Velha Formation in the area. The sigmoid model is the optimum model. Nevertheless, an obvious pitfall in this model is the aforementioned need for more data.

The sigmoid model produces an asymptote at y = 1, related to Well A's highest fracture intensity value. However, there are likely regions with higher levels of fracture intensity that the model does not consider. By adding more data, I could better estimate the actual limit of the fracture intensity in the interval of interest. In addition, I could use more complex functions with more samples in the data frame. That way, outliers would interfere less with the model's fit, which is always a concern in seismic attribute analysis. I could also use multivariate regression techniques or machine learning by adding more data (e.g., seismic multiattribute). Chopra and Marfurt (2019; 2020) introduced the concept of multispectral, multiazimuth, and multioffset coherence and curvature. These curvature volumes can capture features of different scales and structures, making it worthwhile to integrate them in the prediction of the fault and fractured zones. Other works indicate the feasibility of using amplitude variance to characterize fault damage zones and subseismic deformation features (e.g., Chen et al., 2020; Liao et al., 2020).

Another area for improvement in the proposed fracture modelling approach is the noise level of the curvature attribute, especially in the curvature from FFT, which, in my analysis, was the best for estimating the fracture. I attempted to solve that point using filtering techniques for the dip volumes. Despite yielding a convenient answer for

qualitative interpretation, when filtered curvatures and the fracture intensity were plotted, I observed that these filters ruined the analytical relationship between these two properties. Thus, I do not suggest filtering techniques in quantitative interpretation workflows of fracture modelling.

Because I am interested in the curvature and fracture response related to fault zones and would like to exclude or at least attenuate the curvature responses related to noise or stratigraphic features, one way to improve the reliability of the curvature should be aggregating any type of fault probability attribute. A suggestion is to use fault probability volumes estimated by Convolutional Neural Networks (CNN) (e.g., Wu et al., 2019a; Dou et al., 2021; An et al., 2021). The use of such information should ponder the curvature response, attenuating nongeological or nonfault-related features. Additionally, using transfer learning, one can fine-tune the output of the CNN, thereby improving the accuracy of this approach (Cunha et al., 2019; Yan et al., 2021).

Finally, the use of pre-stack seismic data, to obtain inverted elastic seismic volumes could help gain some insight regarding the mechanical stratigraphy, once that elastic constants can be used to partially represent the rock brittleness, which ultimately pounder the rock potential to develop fractures.

7.1.2.7. Future works

The results of this work can be used to guide seismic interpreters in how to properly parametrize the curvature attribute to use it as a fracture predictor. Therefore, one example of an application is to use this optimum curvature as a secondary variable in a geostatistics approach, such as a sequential Gaussian simulation of the fracture intensity (e.g., Correa et al., 2019). Alternatively, instead of directly using this optimum curvature, we can use a transformed version of it, applying the fitted equation to the attribute. Although none of these approaches solve the problem that curvature alone cannot always predict the fracture intensity, we have to account for the effect of mechanical stratigraphy. Thus, to solve this issue, one can aggregate other attributes such as elastic moduli, distance to the fault, mechanical facies, and fault throw measurements as input features into a machine learning regressor to estimate a fracture intensity seismic attribute.

7.2. Second Part – Validation and Creation of a Multi-attribute Machine learning Model to Predict Fracture Intensity

7.2.1. Results

The Pearson and Spearman correlation analysis (Figure 89 and Figure 90) revealed important study aspects. The k1 curvature showed a higher correlation with fracture intensity, followed by silica content and distance to fault. Young's Modulus and Poisson's Ratio showed little to no correlation. However, this does not mean that these features cannot be useful for the model or that their importance follows the same order. There are more robust and specific methods for evaluating feature importance, which will be presented later.

In addition to the correlation matrix evaluation, I observed a considerable correlation (>0.5) between k1 curvature and distance to fault, although not high enough to classify them as redundant information. This is confirmed by the cross-plot analysis (Figure 91) between k1 curvature and distance to fault. The two variables have a reasonable correlation with little scattering at the upper-left corner of the plot, which is related to samples of low curvature and higher distances to fault. Towards lower distances to fault, the dispersion increases, with those levels being associated with both low and high values of k1 curvature. This is explained by the fact that distance to fault only measures the proximity of a sample to the nearest fault and does not indicate any information about the amount of deformation associated with this particular fault. On the other hand, the curvature may serve as a proxy for the amount of strain. Therefore, despite being somewhat correlated, distance to fault and k1 curvature are not redundant features and can be used together in the model.

Additionally, the cross-plot analysis confirmed that the highest levels of fracture intensity were found in samples with higher k1 curvature, closer to faults, and with higher silica content.

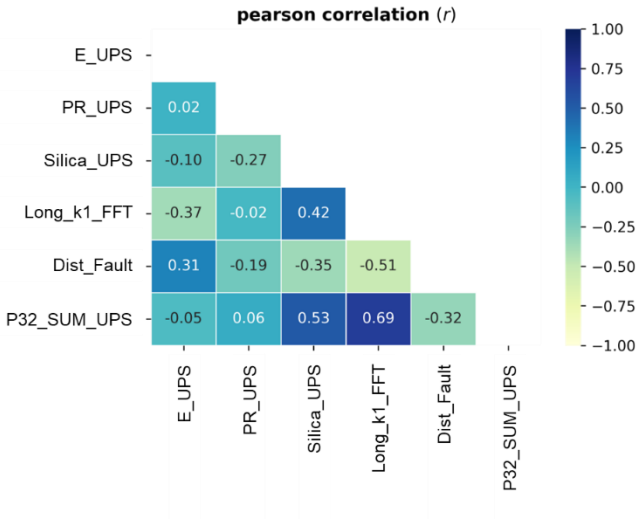

Figure 89: Pearson's correlation matrix showing the nonlinear correlation between the features under study.

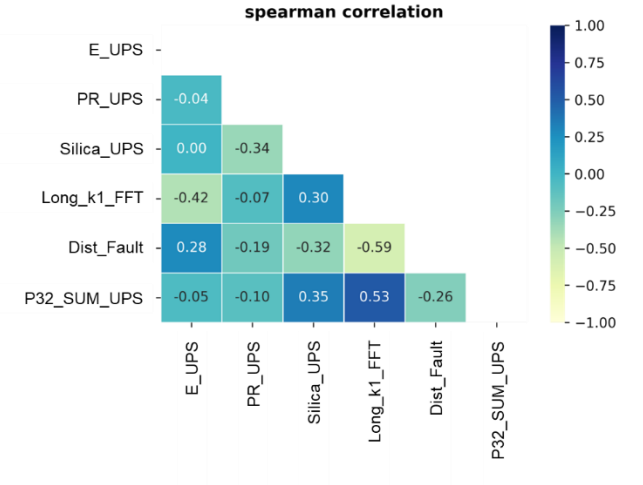

Figure 90: Spearman's correlation matrix showing the non-linear correlation between the features under study.

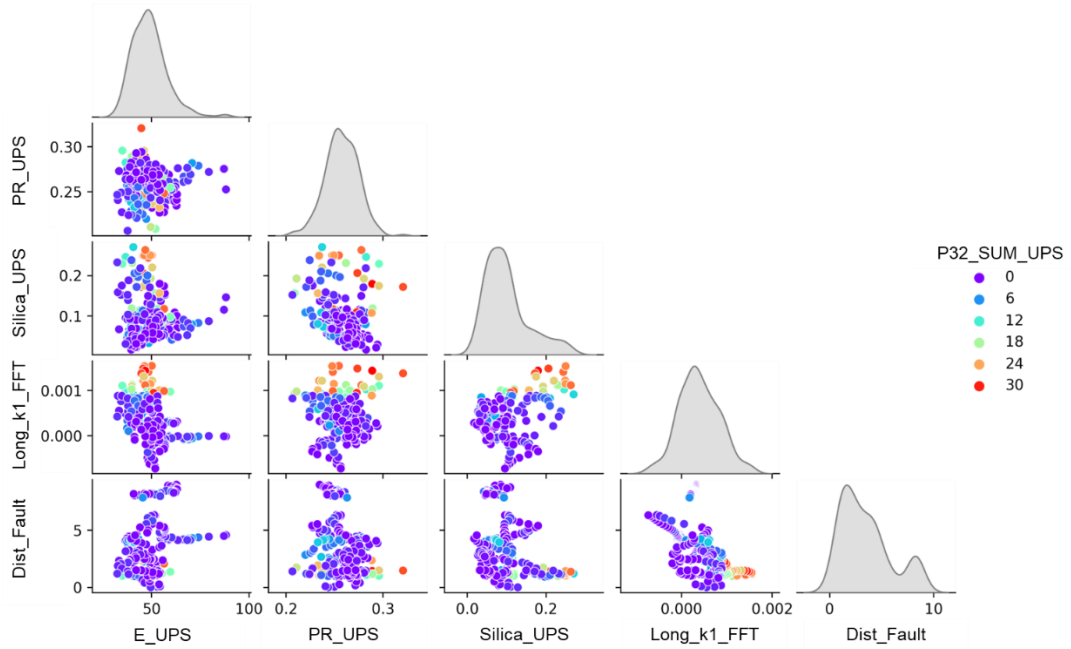

Figure 91: Joint cross-plot analysis between the input features coloured by the levels of fracture intensity. E_UPS: Upscaled Young's Modulus (GPa); PR_UPS: Upscaled Poisson's Ratio (dimensionless), Silica_UPS: Upscaled Silica (v/v); Long_k1_FFT: Long wavelength k1 Curvature (m^{-1}); Dist_Fault: Distance to fault, annotated in sample dimension (each sample represent 25 m).

I evaluated the performance of each model using residual plots (Figure 92 and Figure 93) and regression metrics (Table 5) for fitting with the testing set. The residual plots indicate that XGBoost had slightly better performance than Random Forest, particularly in lower levels of fracture intensity. This is represented by a median and mode closer to zero when compared to Random Forest. Regression metrics also support this analysis, with XGBoost outperforming Random Forest in all three evaluated metrics.

Because the permutation importance method is stochastic, one way to analyze its results is by using boxplots (Figure 94). I ran the permutation importance on the test dataset to evaluate the importance of each feature for the model and its real predictive power. The results showed that k1 curvature has the greatest impact on the R^2 score, followed by distance to fault, and Young's Modulus in third place. Silica content ranked fourth, while Poisson's ratio was the least important feature, with the potential for a null-to-negative impact on the model. This result does not fully agree with the previous results from the correlation analyses (Figure 89 and Figure 90), where silica content presented the third highest correlation with fracture intensity, and Young's Modulus showed the smallest correlation.

These findings support the idea that one can use seismic data to predict fracture intensity away from wells. Among the three most important variables, Young's Modulus is the only feature I cannot use in the present study, but P-Impedance can be used as an approximation. Therefore, I repeated the process to validate this workflow using only the three most important features: k1 curvature, distance to fault, and upscaled P-Impedance log.

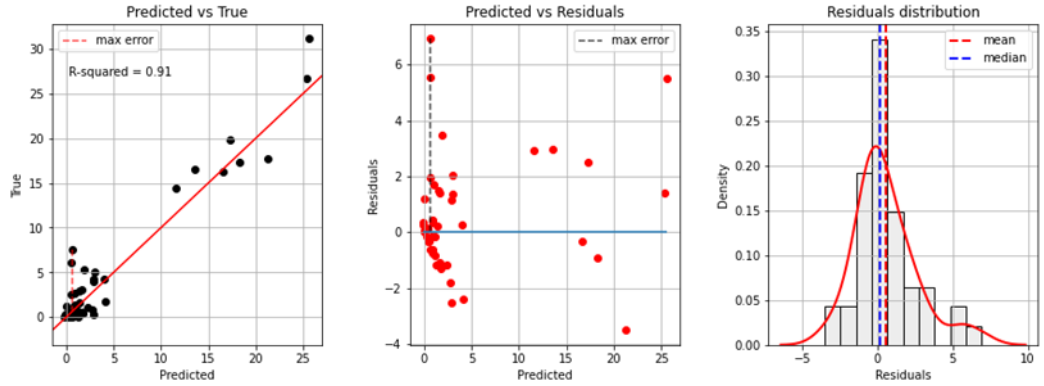

Figure 92: Residual plots of the predicted fracture intensity from the XGBoost model for the test dataset.

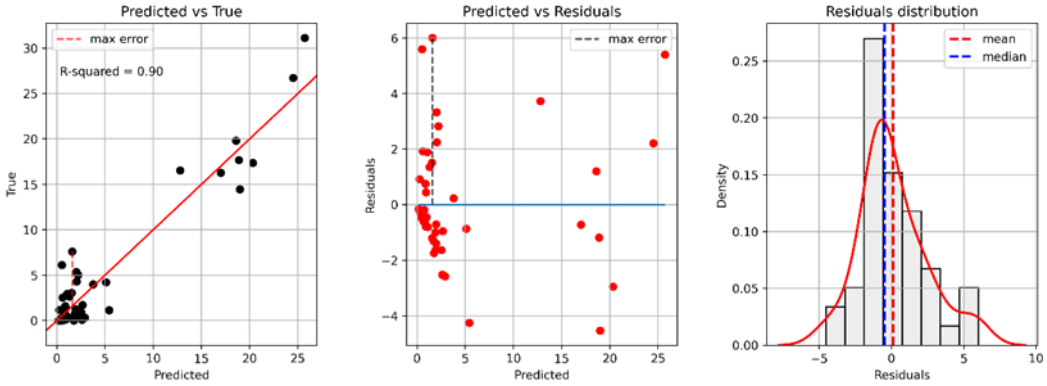

Figure 93: Residual plots of the predicted fracture intensity from the Random Forest model for the test dataset.

Table 5: Regression metrics for the testing set from both models, XGBoost and Random Forest. MAE: Mean Absolute Error; RMSE: Root Mean Squared Error; R2: R2 score.

Metrics	**Model**	
	XGBoost	**Random Forest**
MAE	1.47	1.73
RMSE	2.13	2.29
R2	0.91	0.90

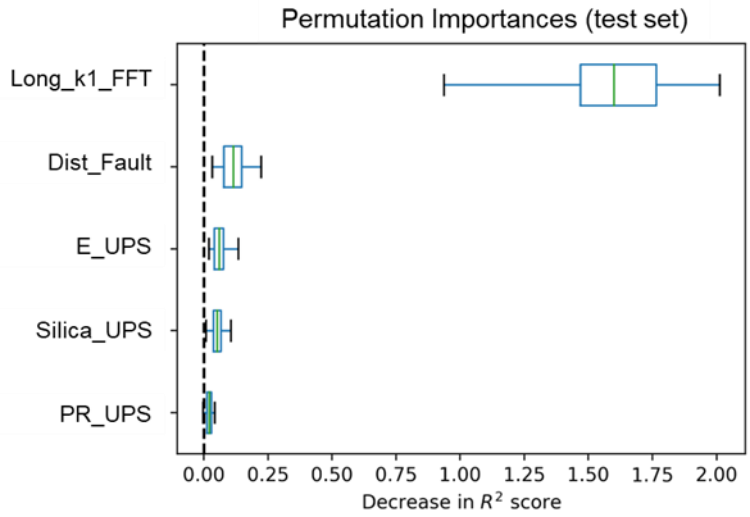

Figure 94: Boxplot analysis from permutation feature importance evaluated in the test set from the XGBoost fitted model. Curvature, Distance to Fault, and Young's Modulus (E) have the most impact on the R^2 score between the predictions and the actual value of the target.

Repeating the exploratory data analysis, I confirmed that, such as Young's Modulus, the P-Impedance has no direct correlation with the fracture intensity (Figure 95). However, it has an important contribution to the model once the samples with the highest fracture intensity are in a limited impedance range (Figure 96). Unsurprisingly, it reasonably impacts the permutation importance evaluation, being near the order of importance as the distance to fault attribute (Figure 97).

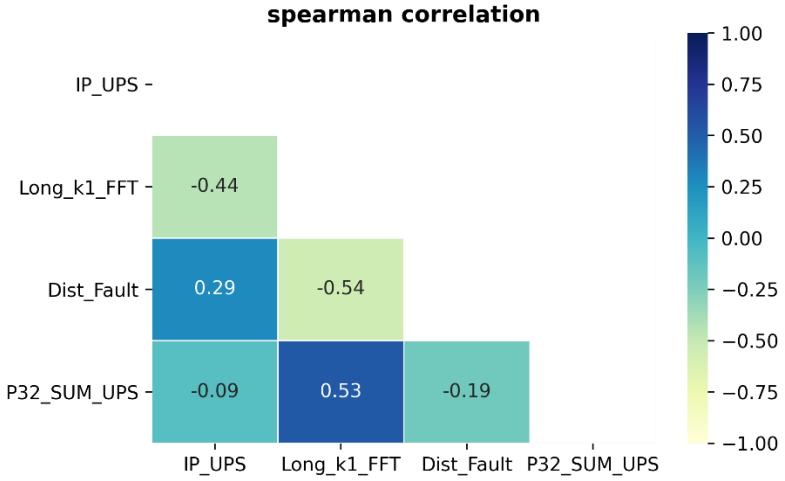

Figure 95: Spearman correlation matrix showing the measured non-linear correlation between the input features under study.

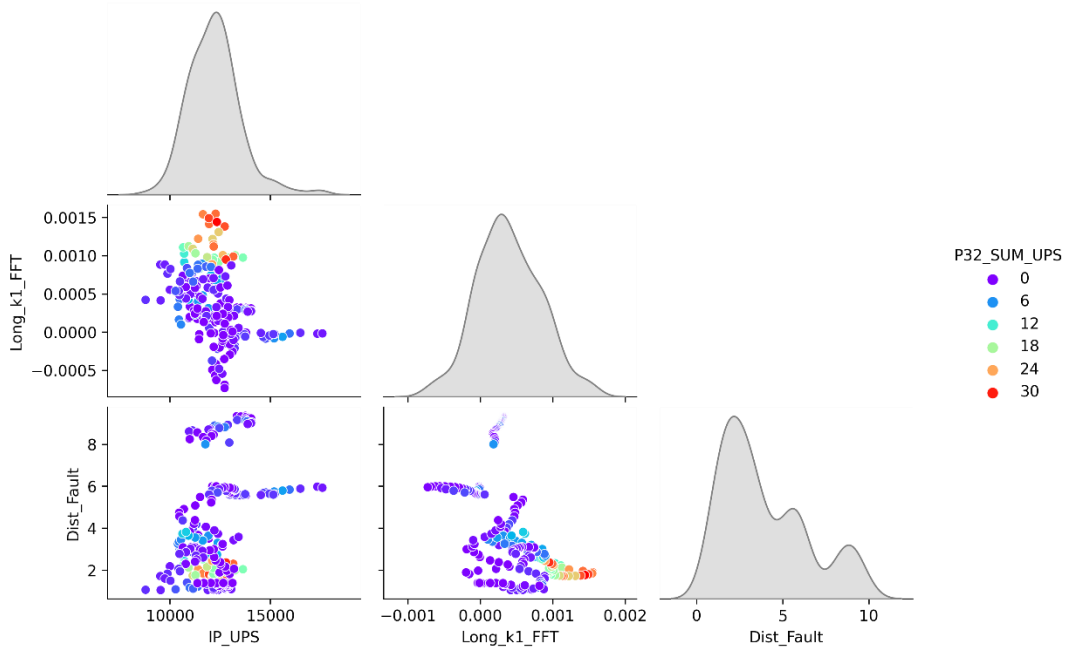

Figure 96: Joint cross-plot analysis between the input features coloured by the levels of fracture intensity. IP_UPS: Upscaled P-Impedance (g/cm³ · m/s); Long_k1_FFT: Long wavelength k1 Curvature (m^{-1}); Dist_Fault: Distance to fault, annotated in sample dimension (each sample represent 25 m).

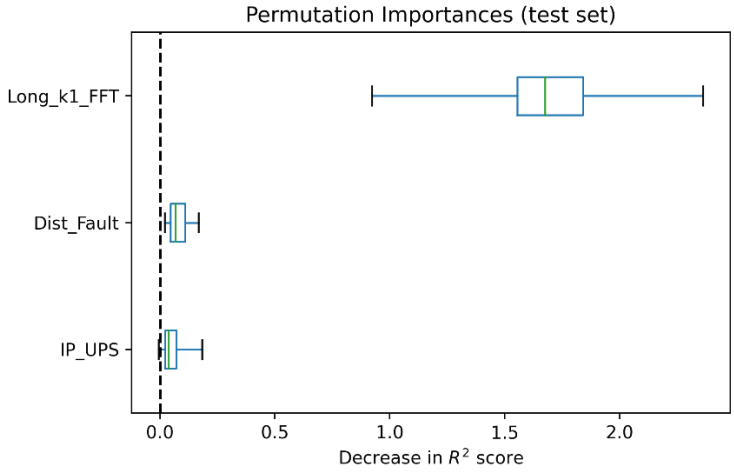

Figure 97: Boxplot analysis from permutation feature importance evaluated in the test set from the XGBoost fitted model using three features. The rank of feature importance is k1 Curvature, Distance to Fault, and P-Impedance, respectively.

The residual plots (Figure 98 and Figure 99) and regression metrics (Table 6) confirm the robustness of the models. Even excluding the silica content and switching Young's Modulus for the P-Impedance, the changes in model performance are minimal, with reasonable residual behaviour and similar metrics. The robustness of the models is also confirmed by the Pearson correlation coefficient measured individually in each well

(Figure 100). The lower Pearson's coefficient is given by the Well E for the XGBoost model, with a correlation of 0.52 between the prediction and the actual values of fracture intensity. In this well, the Random Forest achieved a correlation of 0.85. In the other wells, the correlations given by both models are similar and always above 0.7. Except for Well D, where the XGBoost outperformed the Random Forest by a difference of 0.1. (0.74 to 0.84). In the overall samples, the XGBoost has a greater correlation, 0.99, compared to the 0.98 correlation of Random Forest.

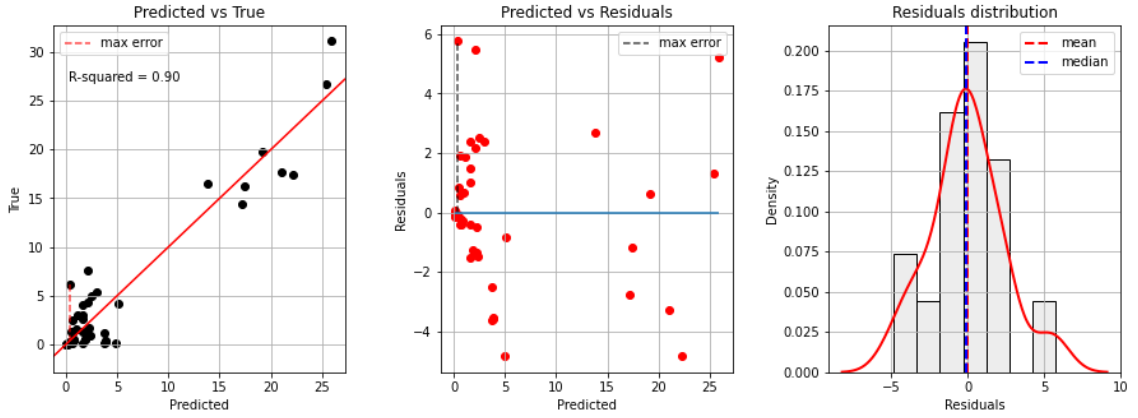

Figure 98: Residual plots of the predicted fracture intensity from the Random Forest model for the test dataset.

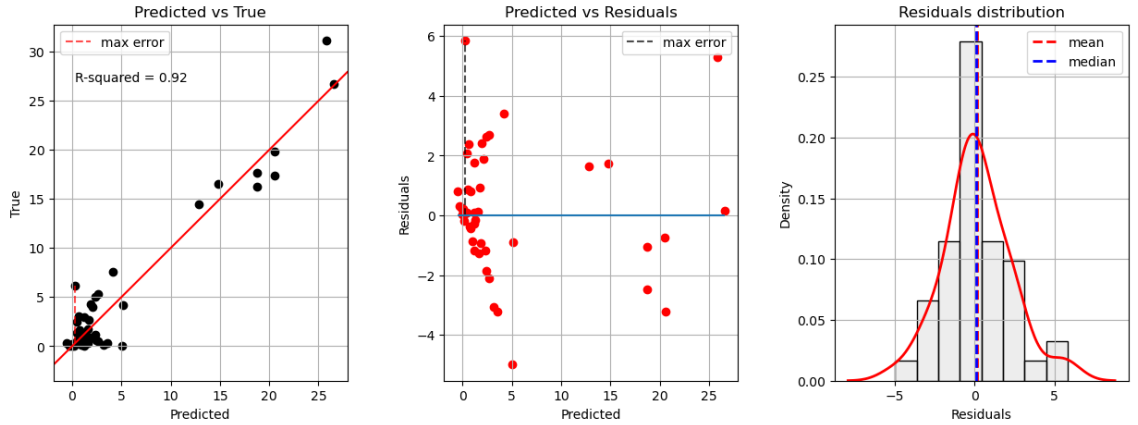

Figure 99: Residual plots of the predicted fracture intensity from the XGBoost model for the test dataset.

Table 6: Regression metrics for the testing set from both models, XGBoost and Random Forest. MAE: Mean Absolute Error; RMSE: Root Mean Squared Error; R2: R2 score.

Metrics	**Model**	
	XGBoost	**Random Forest**
MAE	1.53	1.75
RMSE	2.10	2.37

R2	0.92	0.90

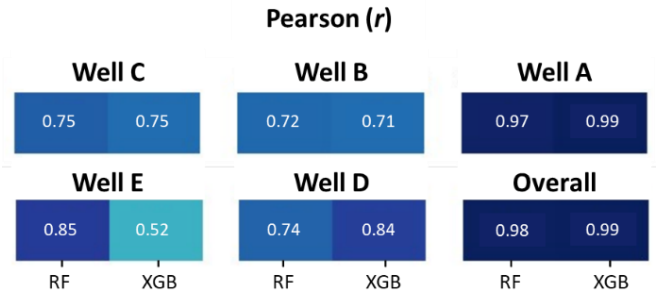

Figure 100: Pearson correlation (*r*) for both Random Forest (RF) and XGBoost (XGB) for each well.

In this section, I applied SHAP plots to conduct an in-depth analysis of the model's behaviour and feature interactions in the XGBoost model. As observed from the violin plot (Figure 101), curvature is the predominant factor influencing the model output, validating previous results obtained through permutation importance analyses. The violin plot offers additional layers of insight. It becomes evident that curvature values do not exert uniform influence. Higher curvature values yield significantly positive SHAP scores, whereas low curvature values lead to a pronounced reduction in prediction (indicating low SHAP scores). Subsequently, Distance to Fault and P-Impedance follow a sequential hierarchy of contribution (based on mean SHAP values). The interpretation of Distance to Fault may present nuances, as diminutive values of this feature are associated with a reduction in output (yielding low SHAP scores), which may seem counterintuitive. Conversely, moderate values exhibit a neutral to positive influence. Regarding P-Impedance, the violin plot underscores that mean values exert a marginal effect on the model's prediction, aligning with theoretical expectations. Interestingly, the extremal lower bounds of P-Impedance manifest an impact associated with low SHAP scores (resulting in a reduction in output). In contrast, the comparatively elevated values of P-Impedance hold notable weight, yielding high SHAP scores (leading to an increase in prediction).

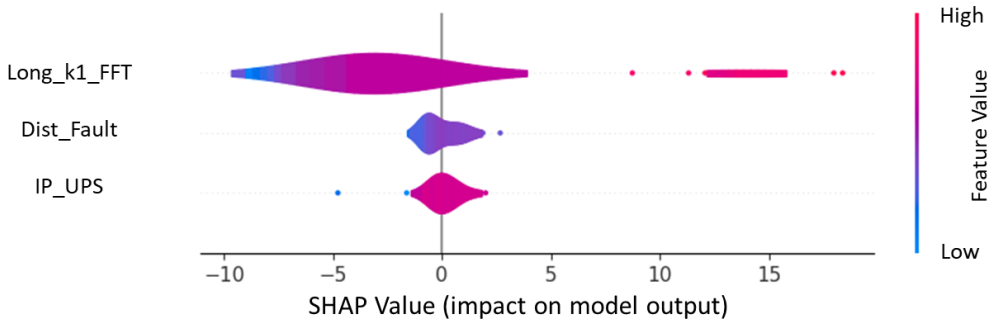

Figure 101: Violin plot representing the distribution of SHAP values across each input feature, demonstrating how these values vary in relation to the feature's values in the XGBoost model. Notably, the k1 curvature exhibits the highest SHAP values on average and at their maximum perspective. The distance to fault ranks second due to its mean SHAP value. While P-impedance exhibits higher absolute SHAP values compared to the distance to fault, its mean SHAP value is comparatively lower, leading to its third-place ranking.

Further into the SHAP plots' analysis, I employed two waterfall plots to illustrate two instances of the model's predictions. Figure 102 illustrates a case with a high predicted outcome. The prediction initiates from the average prediction of 4.70 m^{-1} for P32. Subsequently, it increases by 1.26 m^{-1} due to a value of 47.7 m for distance to fault, followed by an additional 1.42 m^{-1} attributed to the value of 12,226.85 g/cm³ · m/s, and ultimately increases by 18.37 m^{-1} associated with the value of 0.002 m^{-1} for k1 curvature.

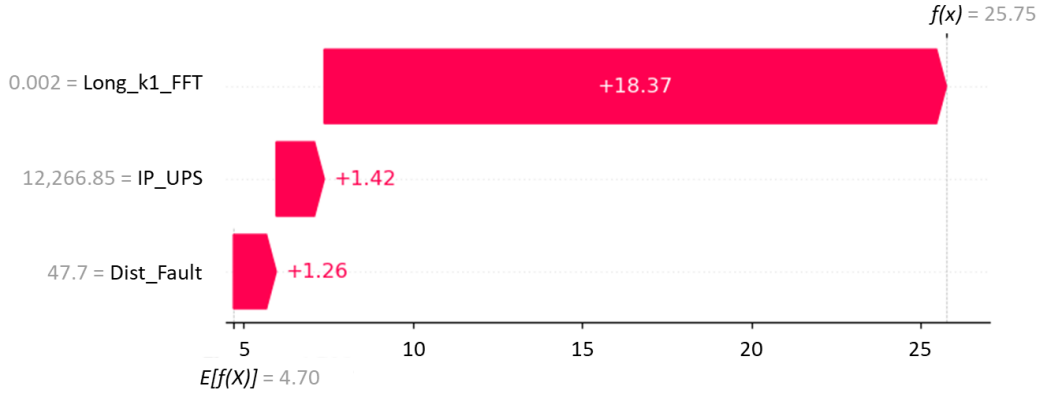

Figure 102: Waterfall plot for a high-value prediction case of fracture intensity. The sample is relatively close to a fault, has a P-impedance slightly higher than the average, and has a very high curvature value. This combination leads to an estimation of 25.75 m^{-1} of fracture intensity. The average prediction is 4.70 m^{-1} for the XGBoost model.

Figure 103 presents a contrasting case to the previous one. The model predicts a lower value for P32, specifically 0.48 m^{-1}. This is primarily due to a lower value of P-Impedance, measuring 12,087.06 g/cm³ · m/s, which reduces the prediction by 0.13 m^{-1}

from the average prediction. Additionally, a higher value for distance to fault, at 144.85 m, further diminishes the prediction by 0.17 m^{-1}. Finally, a lower value of curvature, nearly zero, contributes to a reduction of 3.92 m^{-1} from the prediction.

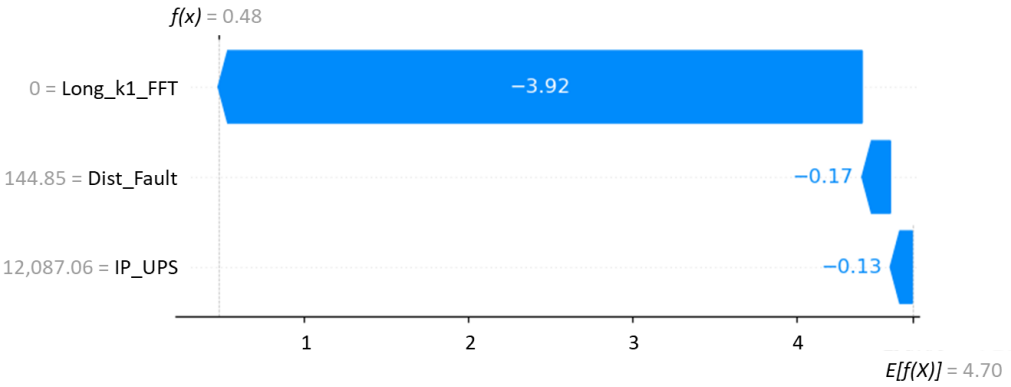

Figure 103: Waterfall plot for a lower value prediction case of fracture intensity. The sample is relatively distant to a fault, has a P-impedance slightly lower than the average, and has a very low curvature value. This combination leads to an estimation of 0.48 m^{-1} of fracture intensity. The average prediction is 4.70 m^{-1} for the XGBoost model.

These two examples help clarify how the model uses the input features to make a prediction. In both cases, curvature has the greatest impact on the outcome. In the higher value case, the sample has a high curvature value, is considerably near a fault, and has a P-Impedance slightly higher than the average (12,164.83 $g/cm^3 \cdot m/s$). In the lower value case, the sample has a lower P-Impedance, near the average, is farther from a fault, and has an extremely low value of k1 curvature.

The predictions of both XGBoost and Random Forest for each well are presented in Appendix A. I plot the prediction for the whole well and highlight the test samples in the well and its forecasting. I also show how the features correlate visually with the target and the forecasting.

The next step of the analysis is to evaluate the performance of the models in blind tests. I used the Well G as a blind well. As illustrated in Figure 104, both models cannot accurately predict the actual P32. Concerning the input features, all the features have a reasonable correlation with the fracture intensity, the target variable. In fact, the models' prediction performed reasonably well in qualitative aspects, which means forecasting the global trend of the fracture intensity along the well.

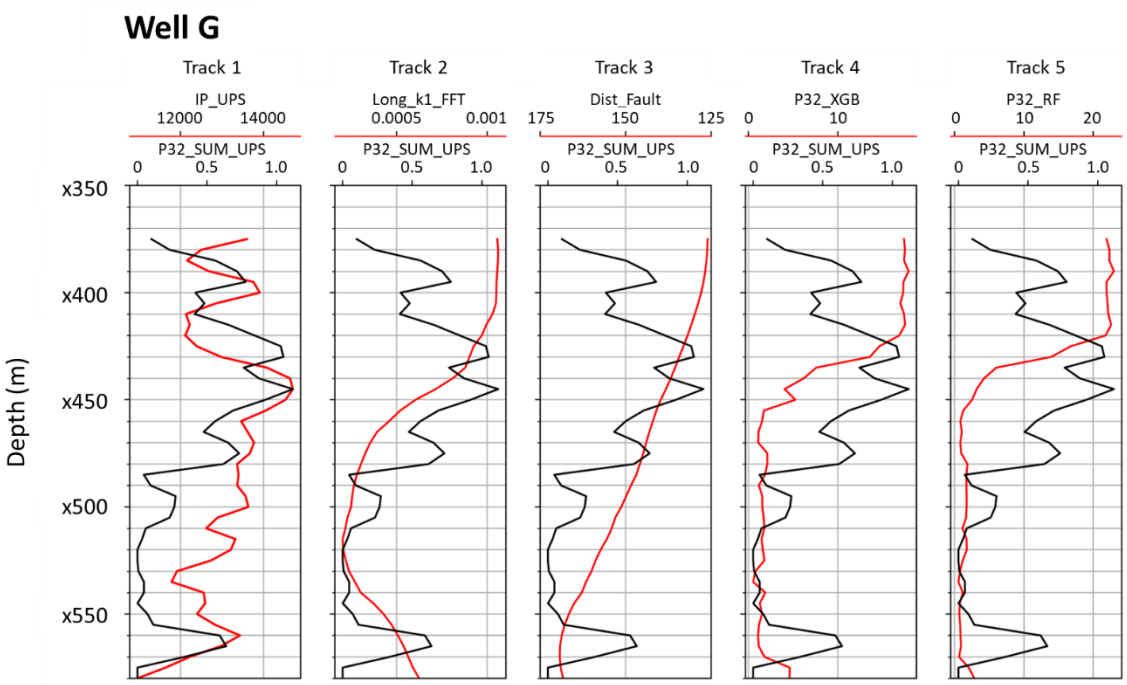

Figure 104: Well track plots showing the visual correlation of the upscaled fracture intensity log (P32_SUM_UPS) with the input features and the predicted fracture intensities in the 1D feasibility study for the Well G (blind test). IP_UPS: Upscaled P-Impedance (g/cm³ · m/s); Long_k1_FFT: Long wavelength k1 from FFT dip (m^{-1}); Dist_Fault: Distance to fault (meters); P32_XGB: Predicted P32 from XGBoost (m^{-1}); P32_RF: Predicted P32 from Random Forest (m^{-1}); P32_SUM_UPS: Upscaled P32 measured in BHI logs. (m^{-1}).

A second blind test was conducted at Well X (Figure 105), located outside the designated study area, within a neighbouring hydrocarbon exploration block that shares the same seismic data. Despite its proximity to the oilfield under study, the geological context at Well X is notably distinct. This area exhibits an exceptional recurrence of intrusion and extrusion igneous events, as documented in three exploratory wells, with Well X being one of them. The analysis from this second blind test yielded results similar to the previous one. Both predictions, generated by the XGBoost and Random Forest models, underestimated the actual scale of fracture intensity by an order of ten. However, it is worth to note that the models performed relatively well in this instance, given that a significant portion of the well comprised igneous rocks. It is pertinent to recall that igneous samples were previously excluded from the study. Therefore, considering that the models' outputs once again capture the overall trend of fracture intensity, even in this challenging scenario involving 'new' lithologies unseen in the training phase, I can regard this outcome as indicative of the robustness of the model. Another point of remark is that,

again, both models produced comparable results. However, the XGBoost prediction appears to have a slightly higher variance than the Random Forest, which has a smoother result.

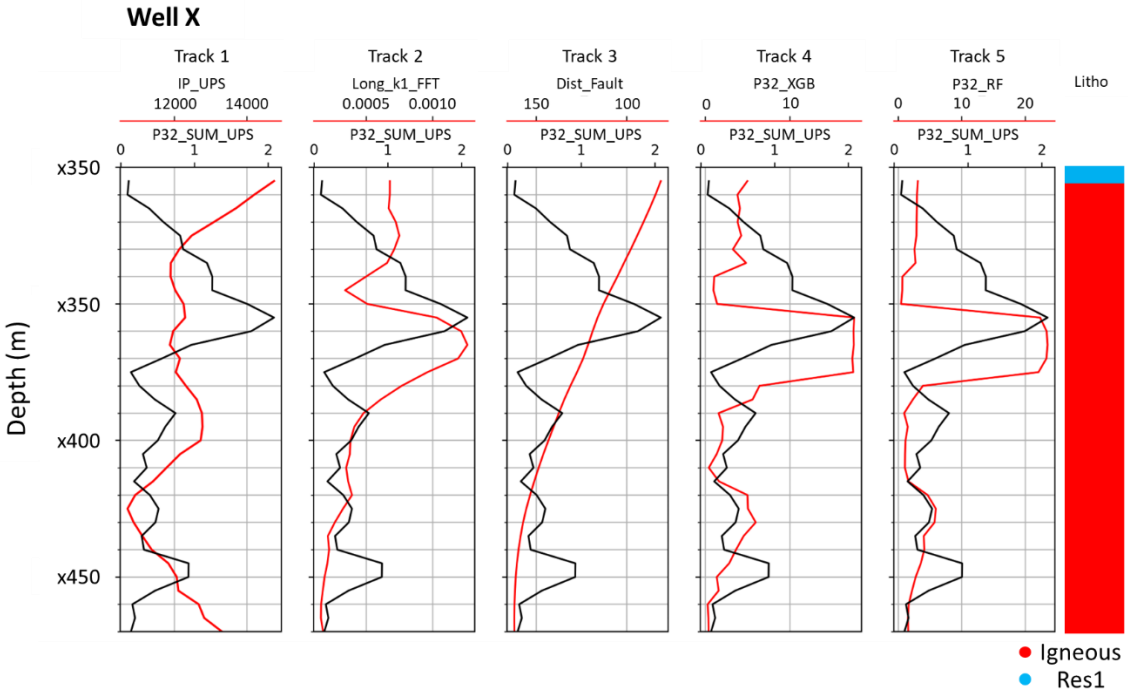

Figure 105: Well track plots showing the visual correlation of the upscaled fracture intensity log (P32_SUM_UPS) with the input features and the predicted fracture intensities in the 1D feasibility study for the Well X (blind test outside the study area). P_UPS: Upscaled P-Impedance (g/cm³ · m/s); Long_k1_FFT: Long wavelength k1 from FFT dip (m^{-1}); Dist_Fault: Distance to fault (meters); P32_XGB: Predicted P32 from XGBoost (m^{-1}); P32_RF: Predicted P32 from Random Forest (m^{-1}); P32_SUM_UPS: Upscaled P32 measured in BHI logs. (m^{-1}). Litho: Eletrofacies determined by the well logs and drill cutting samples. Res 1: Reservoir of carbonate facies with high porosity and low clay volume.

A third viable alternative to evaluate the models' performance and, simultaneously, the applicability of the present workflow was using it to predict the fracture intensity in Well H. This well does not have BHI logs; therefore, one cannot measure its fracture intensity. However, it does have the other necessary information for the prediction (P-Impedance, k1 curvature, and distance to fault). It also has a continuous sampling of rock data. By comparing the model's predictions with the rock sample, I gained insight into the overall performance of the model (Figure 106). The results showed that the predicted fracture intensity aligned with the rock samples, with higher values of fracture intensity corresponding to higher, more fractured samples and the lowest levels

of fracture intensity, indicating fine-grained carbonate rocks without any evidence of fractures.

If the model prediction is right, at least in identifying trends, as indicated by the two previous blind test analyses, there is a thick continuous interval of a fractured reservoir with more than 100m. In that case, this interval can potentially have an anomalous production behaviour associated with the so-called excess permeability zones, as defined by Fernandes-Ibanez et al. (2022a). In fact, this well is one of the most productive wells in the whole Presalt. It is worth noting that I excluded igneous samples from the training phase, and therefore, I cannot evaluate the model's performance in those intervals.

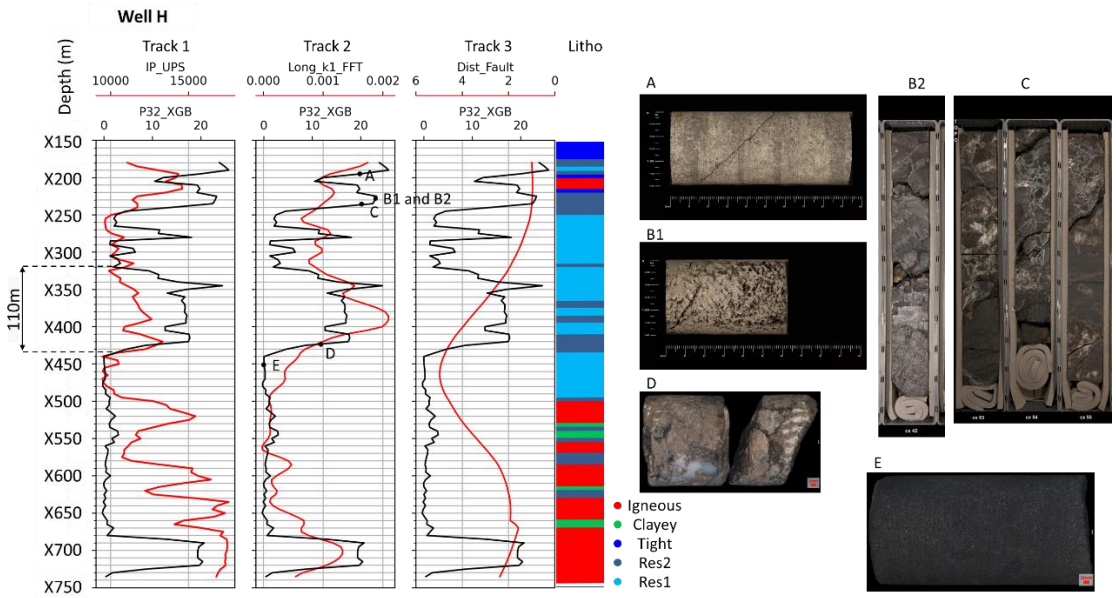

Figure 106: Well log plot from Well H with the 1D fracture intensity predicted by the XGBoost model and the input features, P-Impedance, k1 curvature, and distance to fault (measured in the number of samples). Sidetrack with the upscaled eletrofacies. Points A to E are examples of rock samples (sidewall cores, and well cores) that corroborate the response of the predicted fracture intensity. Samples A to D have a higher degree of fracture intensity with evidence of karstification (B1), silica precipitation (C and D), and dissolution-enhanced fractures (C). Sample E is a fine-grained carbonate with the absence of fractures.

The above analysis shows a potential applicability of the present study. When a reasonable number of wells with BHI logs is available to train a machine learning model, one can use it to run predictions of fracture intensity in other wells without BHI logs, at least to be aware of zones more or less fractured, and to have a quantitative measurement of this potential.

In the concluding phase of this study on 3D fracture intensity prediction, I opted for the XGBoost algorithm to forecast fracture intensity within the seismic volume. The findings of the 1D feasibility study guided this choice. Notably, the XGBoost algorithm demonstrated superior performance over the Random Forest algorithm.

The analytical focus shifted towards visualizing these results on a map. At the salt base level, I extracted the three seismic attributes that served as input features for the XGBoost model. These attributes were complemented by the model's predictions of fracture intensity (refer to Figure 107).

Examining the background trend of predicted fracture intensity (Figure 107-d), it became evident that it primarily correlates with variations in P-Impedance Figure 107-a). Conversely, abrupt shifts in fracture intensity predominantly arise from fluctuations in k1 curvature values (Figure 107-c) and proximity to faults (Figure 107-b), especially noticeable in extremely high k1 curvature and low distance to fault values.

The regions exhibiting heightened fracture intensity are primarily concentrated in the eastern and northern sectors of the map. To delve deeper into these predictions, I detailly analysed specific scenarios across four seismic sections, each intersecting a distinct well.

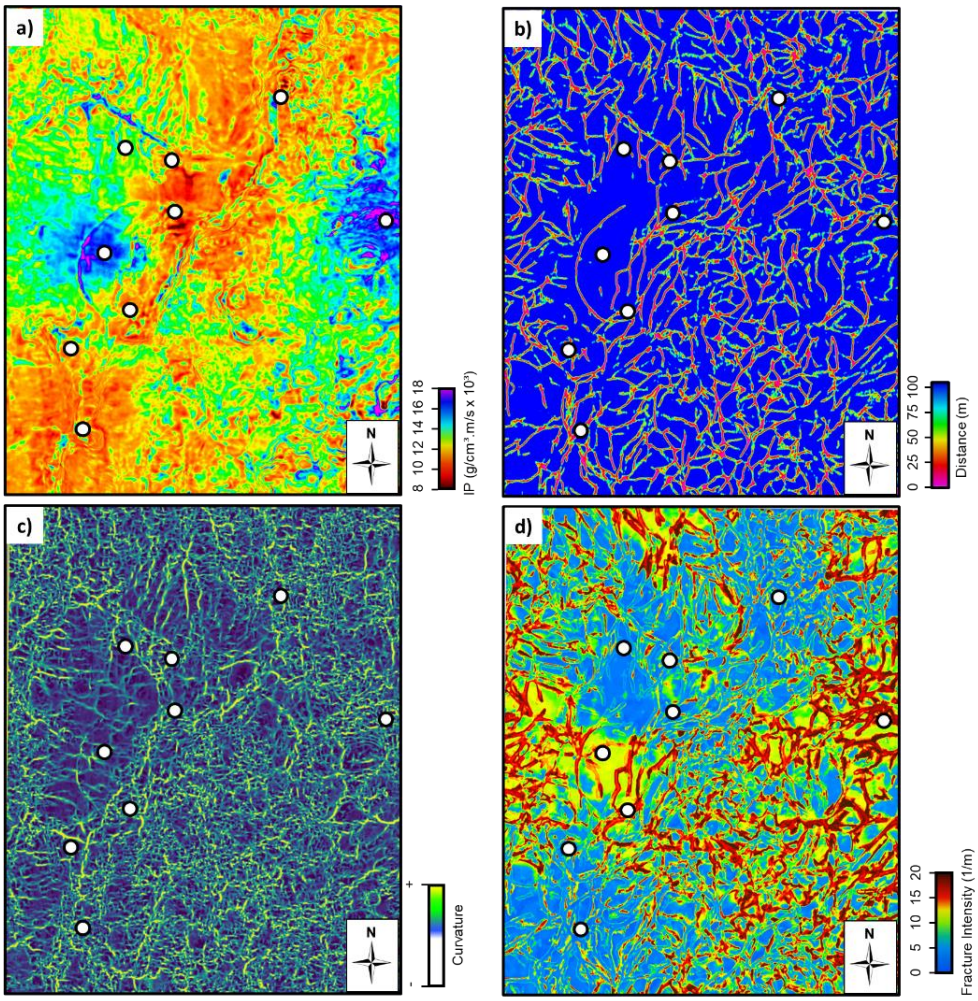

Figure 107: Salt base maps from the three input features and the 3D prediction of fracture intensity. a) Inverted P-Impedance; b) Distance to fault attribute calculated from the CNN fault probability; c) Long wavelength k1 from FFT dip; d) Fracture intensity estimated by the XGBoost in the seismic volume. White dots indicate the location of the wells.

The first seismic section crosses the Well A (Figure 108). As stated before, this is the most fractured well measured by the BHI logs. As shown in Figure 108-b, it has a background of high values of P-Impedance, averaging 14,000 g/cm³ · m/s. The well is also intersected by an expressive fault zone that appears both in the distance to fault (Figure 108-c) and k1 curvature (Figure 108-d) attributes. If the pattern of the well can be extrapolated away from them, this higher value of P-Impedance can be associated with a silicification process that makes the rock stiffer. When associated with the faulting process, this feature helps explain the anomalously high values of fracture intensity encountered in this well.

The case for Well B (Figure 109) is quite conspicuous. This well drilled a local relative structural low associated with a restricted environment of inner platform. Therefore, especially in the base of the well, some argillaceous facies are represented by low values of P-Impedance (Figure 109-b). There is no expressive fault zone near the well, as indicated by the distance to fault attribute (Figure 109-c). Although just aside from the base of the well, some strain is captured by the k1 curvature attribute (Figure 109-d). At the top, the fracture intensity increases associated with increasing P-Impedance values. This is due to the igneous intrusion at the top, which produces some metamorphism and hydrothermal halo.

The next example is from Well D (Figure 110). This well drilled a huge build-up. The background trend of P-Impedance for this particular build-up is of low values of P-Impedance (< 12,000 $g/cm^3 \cdot m/s$) (Figure 110-b). However, this build-up is heavily faulted, as indicated by the distance to fault (Figure 110-c) and k1 curvature (Figure 110-d) attributes. This combination of factors produces a singular pattern regarding the fracture intensity. Contradicting the fracture intensity information given by the BHI logs, this build-up is heavily fractured. Although the fractures, as indicated by the prediction, are restricted to the fault zones. The mechanical characteristics of the facies constituting this build-up rendered it less susceptible to fracture development. Consequently, when drilling a well between two fault zones, it did not intersect the actual range of fracture intensity.

The final example is from Well G, the blind test (Figure 111). Well G is a clear example of underestimating fracture intensity by well data. Aside from the well (<500m), there is a huge fault zone with extremely high fracture intensity values. The 3D prediction produced a better result than the 1D prediction. It correlates very well with the actual fracture intensity measured in the BHI log (Figure 112 and Figure 113).

In Appendix B, I show the same results of the seismic section for the remaining wells.

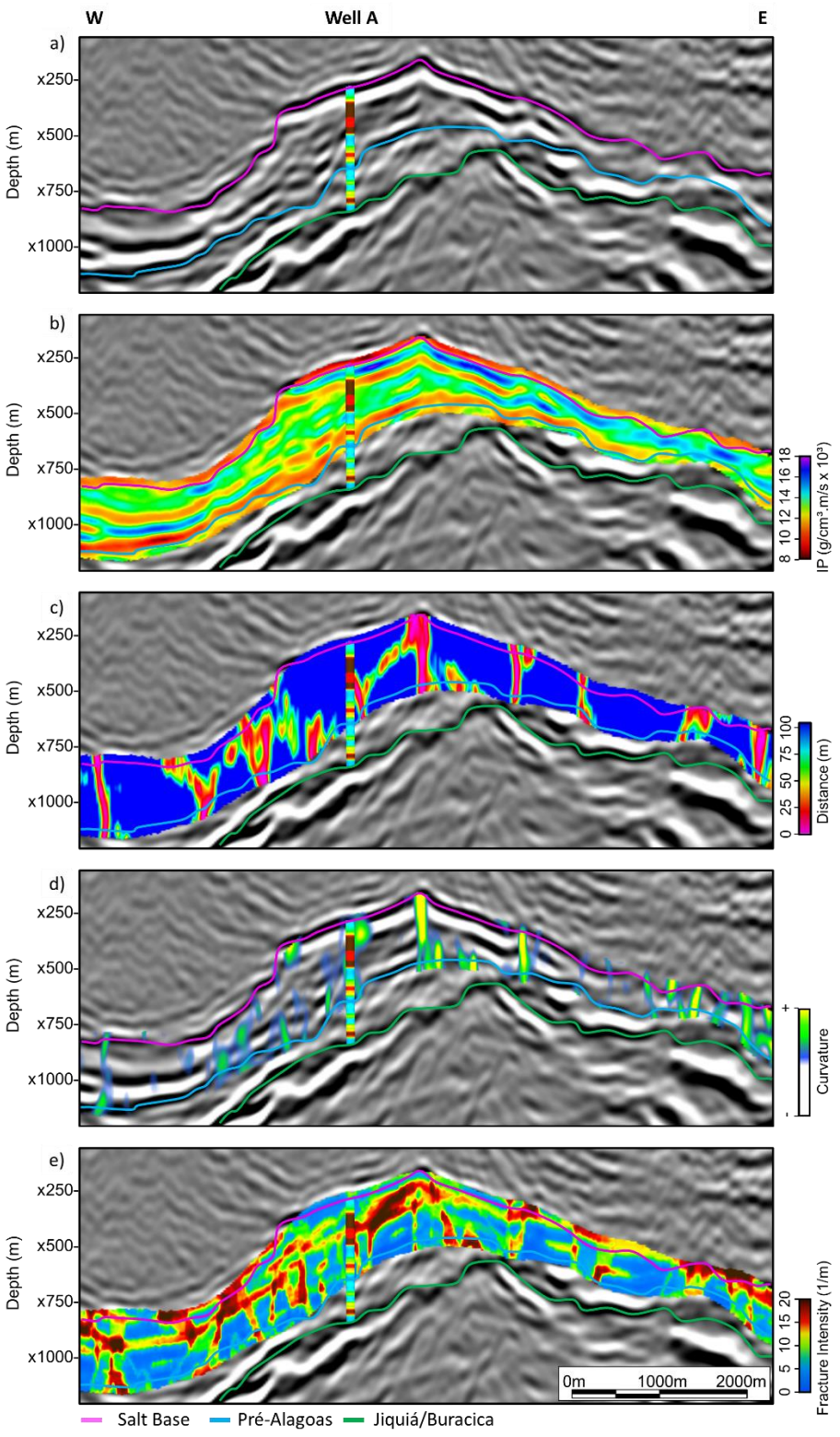

Figure 108: Crossline intersecting the Well A. a) Seismic amplitude; b) Inverted P-Impedance; c) Distance to fault; d) Long wavelength k1 from FFT dip; e) Fracture Intensity predicted by the XGBoost model. The Well A is plotted in each figure for qualitative correlation. The prediction matches the fracture intensity in the well logs. The higher values of fracture intensity extrapolate away from the wells laterally following the seismic beds of higher P-Impedance, close to faults and high strain measured by the k1 curvature.

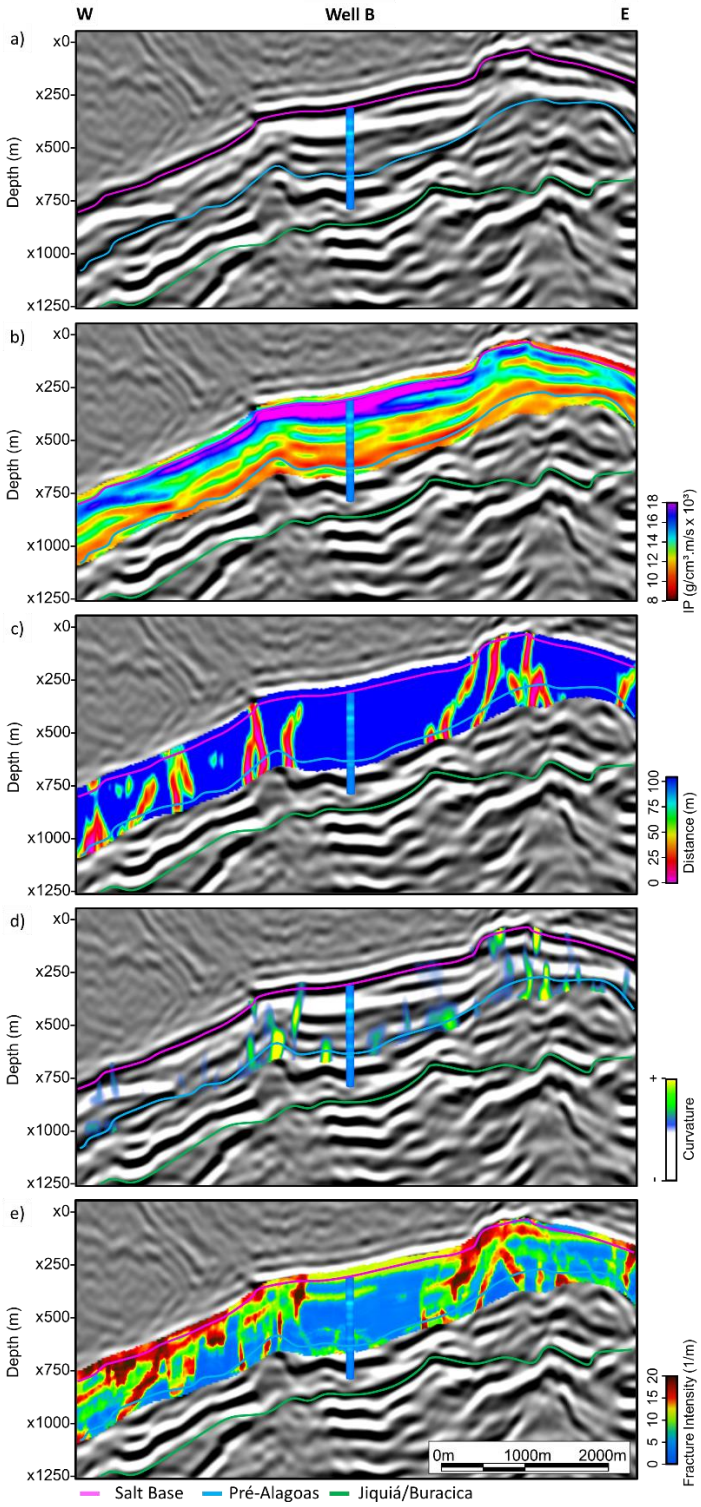

Figure 109: Crossline intersecting the Well B. a) Seismic amplitude; b) Inverted P-Impedance; c) Distance to fault; d) Long wavelength k1 from FFT dip; e) Fracture Intensity predicted by the XGBoost model. The Well B is plotted in each figure for qualitative correlation. The prediction matches the fracture intensity in the well logs. The machine learning model estimates low values of fracture intensity near the well. The higher values of fracture intensity in the well are associated with the presence of a thick igneous intrusion and its hydrothermal and metamorphic halo.

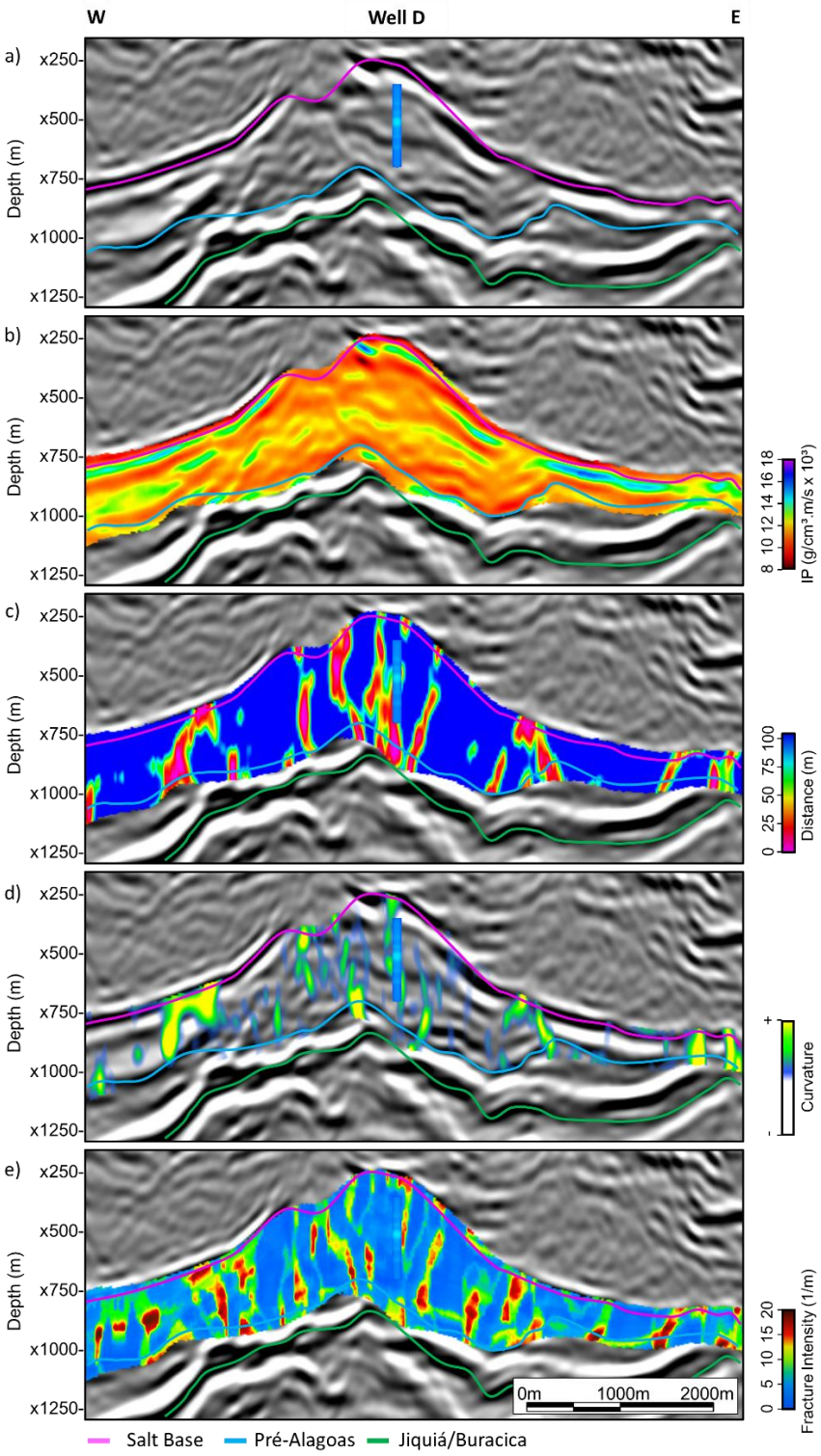

Figure 110: Crossline intersecting the Well E. a) Seismic amplitude; b) Inverted P-Impedance; c) Distance to fault; d) Long wavelength k1 from FFT dip; e) Fracture Intensity predicted by the XGBoost model. The Well E is plotted in each figure for qualitative correlation. The prediction matches the fracture intensity in

the well logs. The higher values of fracture intensity are restricted to faulted zones. The carbonate build-up has a predominance of low P-Impedance values, indicating a less stiff rock not prone to the development of fractures. The highest value of fracture intensity in the well is just aside from a huge fault zone.

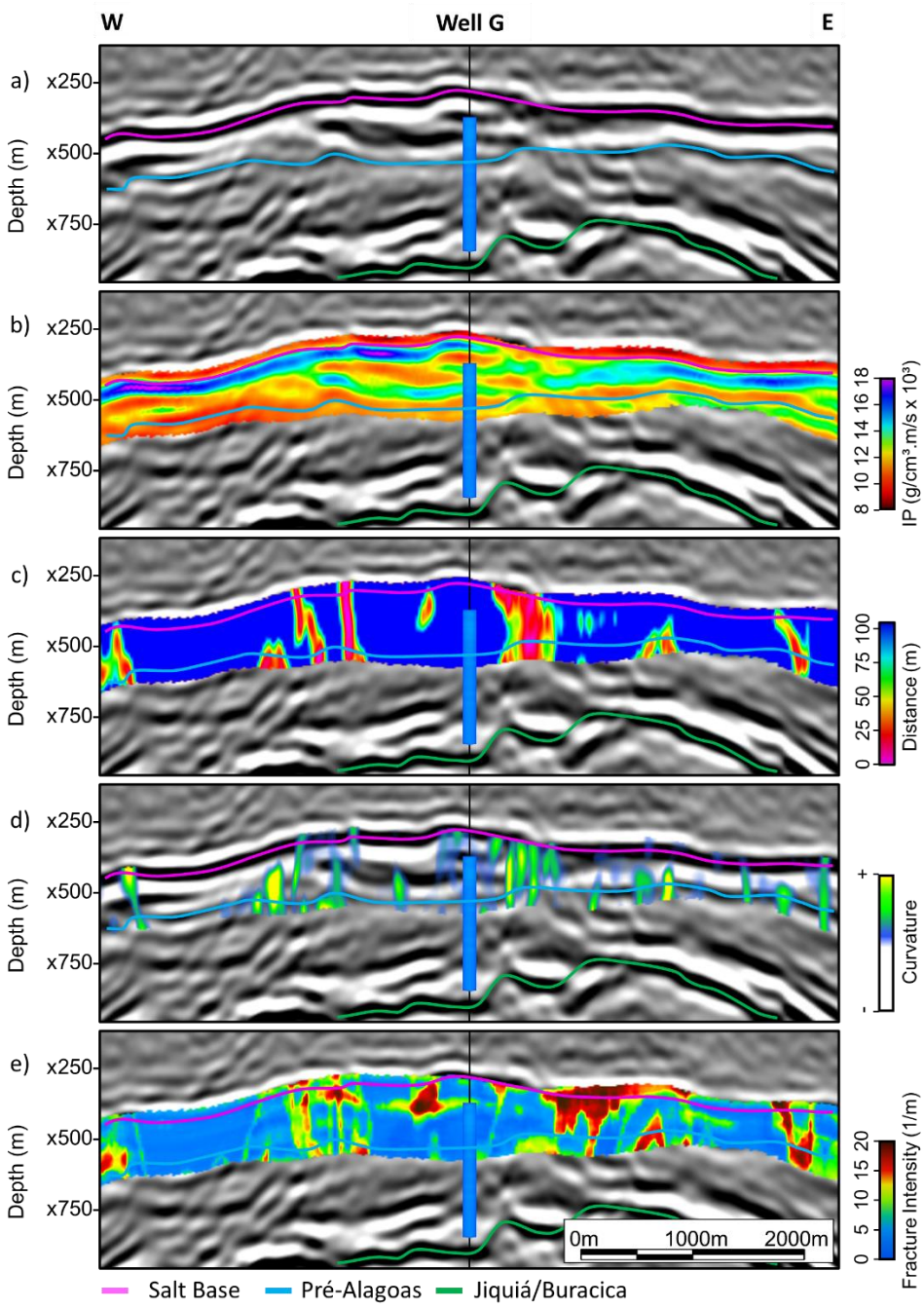

Figure 111: Crossline intersecting the Well G (blind test). a) Seismic amplitude; b) Inverted P-Impedance; c) Distance to fault; d) Long wavelength k1 from FFT dip; e) Fracture Intensity predicted by the XGBoost model. The Well E is plotted in each figure for qualitative correlation. The prediction matches the fracture intensity in the well logs. The region is indeed less fractured than its neighbourhood, where high values of P-Impedance, k1 curvature and the presence of faults are associated with high values of the estimated fracture intensity.

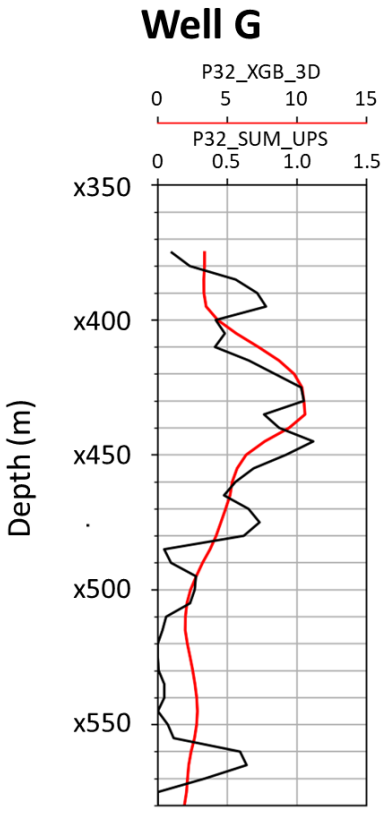

Figure 112: Well track plot showing the visual correlation of the upscaled fracture intensity log (P32_SUM_UPS) with the predicted fracture intensities in the 3D study for the Well G (blind test). P32_XGB_3D: Predicted P32 from 3D XGBoost (m^{-1}); P32_SUM_UPS: Upscaled P32 measured in BHI logs. (m^{-1}).

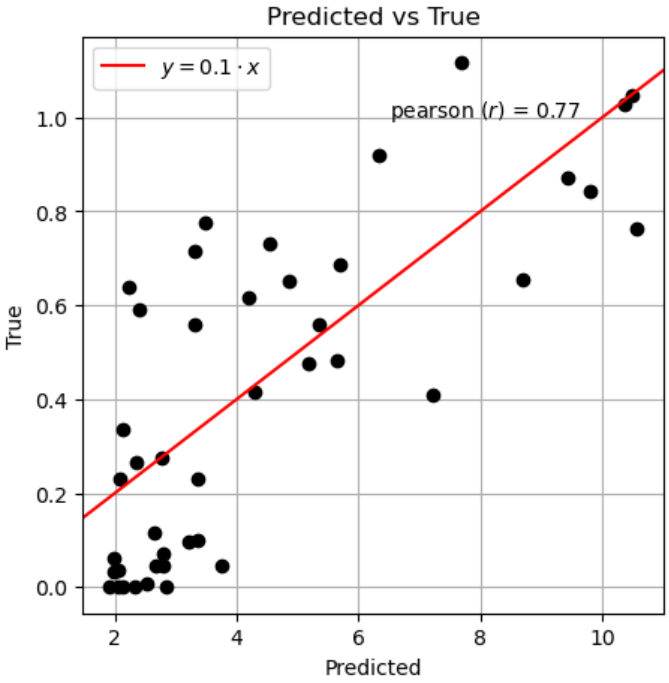

Figure 113: Scatterplot of the fracture intensity predicted by the 3D XGboost model in the seismic volume and the actual fracture intensity measured in the Well G (blind well). The scatterplot helps clarify that the model failed to predict the actual values. However, a reasonable linear correlation exists between the estimation and the actual values as measured by Pearson's coefficient (r) of 0.77.

7.2.2. Discussions

7.2.2.1. 'XGBoost versus Random Forest

In the 1D feasibility study, XGBoost demonstrated superior performance compared to Random Forest across various global metrics, including error levels, the determination coefficient (R^2) on the test set, and the overall correlation in all the wells. As previously mentioned, XGBoost's predictions exhibit a higher degree of variability than Random Forest, resulting in enhanced 'resolution' in some cases. The preliminary tests on 3D predictions reveal a similar result, with Random Forest providing a smoother estimation compared to XGBoost. This observed pattern can be attributed to the inherent algorithms of each model, as discussed in Chapter 5 of the machine learning analysis.

In this context, bagging methods, such as Random Forest, strive to reduce prediction variance, whereas boosting methods, like XGBoost, intrinsically increase variance after each iteration during tree construction. Since fracture intensity is a property characterized by significant discontinuities, a higher variance estimator seems to be better suited for capturing the nuances of this target property. Fortunately, this increased

variance does not appear to lead to overfitting, as evidenced by the metrics on the test set and blind wells, indicating satisfactory generalization capability.

Despite some limitations, the algorithm successfully predicts the general pattern of fracture intensity in previously unseen samples. Looking ahead, I can explore the adoption of other tree-based ensemble regressors. Notably, Light GBM and CatBoost are two examples of boosting algorithms that have been reported to offer improved accuracy and faster performance than XGBoost. They are emerging as promising alternatives, potentially superseding XGBoost as the preferred framework for achieving superior results in managing tabular datasets within machine learning contests (Carlens, 2023).

7.2.2.2. Sampling Bias Effect and Its Impact on the Models

A notable factor prevalent in this study, with the potential to impact the performance of the models, is the issue of sampling bias. This effect is particularly relevant in relation to four key aspects of the present work: the positioning of the well, the upscaling of the fracture intensity, the presence of fault zones, and the k1 curvature.

In the presalt region, wells tend to be strategically positioned away from fault zones for safety during drilling operations. Consequently, the samples often underrepresented areas with higher fault and fracture occurrence. As a result, just one well in the present study is intersected by a seismic-scale fault, with the other wells being at least 75 meters from the closest fault.

Another instance of sampling bias arises when considering the correlation between k1 curvature and distance from a fault. The regions with the highest k1 curvature values tend to be located on the upthrow side of the fault. This leads to a lateral misalignment between the peak k1 curvature value and the precise position of the fault. In this context, the 3D prediction method mitigates this effect by employing a 3D window analysis. This is exemplified in the case of Well G, where the 3D prediction yields a more accurate estimation of fracture intensity compared to the 1D prediction, as it accounts for lateral information.

One potential solution to address this mismatch between curvature and distance from the fault, is to introduce another attribute, flexure (or aberrancy) (e.g., Gao, 2013; Di and Gao, 2015; Di and Gao, 2017; Qi and Marfurt, 2017), which is associated with the gradient of curvature. In the presence of a fault zone, flexure reaches its highest value at

the central position of the fault, aligning with the distance-to-fault information. It is worth noting that while some may question whether incorporating the curvature gradient might introduce redundant information, it is important to highlight that tree-based ensembles, unlike other machine learning methods such as CNNs, do not estimate derivatives of input features internally.

Finally, determining the distance from the fault is achieved by establishing a threshold for the fault probability volume, then by interpolating the calculated distances for each sample and smoothing the result. Each of these steps, from setting the threshold to interpolation and smoothing, can significantly influence the outcomes of this attribute.

7.2.2.3. Poisson's Ratio and other Elastic Moduli

One remarkable result of this study was the unexpected finding that Poisson's ratio had a null to negative impact on the model's predictions. In fact, the model trained without Poisson's ratio exhibited superior metrics compared to the one trained with it. This is a surprising outcome, considering that Poisson's ratio is commonly used to indicate brittleness along with other elastic moduli.

One possible explanation for this phenomenon is the inadequate recording of well logs. This holds especially true for wells in the presalt region, particularly those containing silicified zones and occurrences of vugs (Oliveira, 2023, personal communication). In these intervals, the recording of shear-sonic logs becomes compromised, leading to inaccurate estimations of elastic moduli. Although I did not explicitly assess this issue in my dataset, given the prevalence of silicification and karstification observed in the well cores, it is most likely that the present dataset is affected by the same problem. Consequently, this significantly diminishes the reliability of the information provided by Poisson's Ratio.

7.2.2.4. Mechanical Stratigraphy and Difference Between Well A and Well E

One clear example of mechanical stratigraphy is the difference in the fracture intensity levels of Well A and Well E. Both penetrate carbonate build-ups, which tend to be indicated as heavily fractured facies. However, as I observed both in the BHI logs and in the fracture intensity estimation in the seismic volume. The build-up from Well A is widely more fractured than Well E. This difference is due to the silicification process in

the Well A region. This study does not evaluate the mechanisms of this huge silicification process comprehensively. However, it is important to mention that huge igneous intrusion extended from Well B to the neighbourhood of Well A. Igneous intrusions are often associated with the input of hydrothermal fluids and mineral precipitation in the presalt. Therefore, it is not unlikely that the igneous intrusion at least enhanced the silicification in this region, which can explain the anomalous silica content values registered in Well A.

7.2.2.5. Mechanical Stratigraphy and Timing of Events

One of the focal points of this study lies in addressing the mechanical stratigraphy in fracture modelling. In the latter part of the research, I employed P-Impedance to aggregate information about variations in brittleness within the stratigraphy. This choice significantly influenced the models, as indicated by the permutation importance, SHAP analysis, and my interpretation of the estimated fracture intensity attribute concerning the input seismic attributes. It is worth noting, however, that P-Impedance provides a simplified approximation of rock stiffness.

An additional enhancement could involve incorporating a discrete feature indicating the mechanical facies. To achieve this, for 1D predictions, I could estimate the mechanical facies using data from well logs, rock samples, and other well-related information. For 3D predictions, estimating the cube of mechanical facies would be essential. If these mechanical facies correlate with the elastic moduli, I could employ pre-stack seismic inversion and Bayesian classification to estimate these facies within the seismic volume. As demonstrated by this work, accounting for the influence of mechanical stratigraphy is crucial in performing a robust estimation of fracture intensity in the Santos presalt.

There are several factors controlling mechanical stratigraphy in the presalt. Lithological and depositional aspects such as grain size and facies association are two examples of these factors. However, diagenesis exerts a huge influence also. Diagenesis is a predominant factor in presalt that should always be considered. Among the diagenetic phases occurring in presalt, silicification is often associated with increasing fractures due to higher stiffness. Dolomitization and karstification are other diagenetic phases that can influence the fracturing potential of the rock. For a more robust work concerning

mechanical stratigraphy, it should be important to identify these relevant diagenetic phases throughout the study area and determine their timing relative to the fracturing process. That is, one should map these different diagenetic features and understand if these features occur before or after the fractures. Then, after this work, one should be able to define some diagenetic zones and perform the fracture modelling constrained to these zones. This would also help to determine potential areas of open fractures. For example, as concluded by Fernandez-Ibanez et al. (2022b), fractures that occur relatively early in the diagenetic history of the rock are more likely to be open than fractures that occur lately, when the rock is already buried.

In summary, despite this work's attempt to deal with mechanical stratigraphy for fracture modelling in presalt, further improvements are encouraged regarding this theme in the present workflow.

7.2.2.6. Future works and applicability

The proposed machine-learning-based workflow for fracture modelling relies heavily on data. However, it is important to note that the choice of machine learning framework can also significantly influence the final outcome. For instance, tree-based algorithms are limited because they cannot extrapolate beyond the values encountered during the training phase. To illustrate, if the maximum value of fracture intensity in the dataset is 30 m-1, the model will cap predictions at this value, regardless of the input features provided. To enable extrapolation, alternative methods like neural networks must be considered. Nevertheless, given the commendable performance demonstrated by tree-based ensemble methods in this specific task, I advocate for an ensemble approach, as exemplified by Mandal et al. (2021), rather than favouring one method over the other.

In terms of enhancing results through improved data, I recommend enhancements across all three domains of input features: curvature, fault zones, and rock stiffness. Regarding the curvature, although I exclusively evaluated three methods for dip estimation in the context of k1 curvature, it is worth noting that, as discussed in Chapter 4, there exist at least seven different approaches presently. More recently, Ao et al. (2021) introduced Convolutional Neural Networks (CNNs) for estimating k1 and k2 curvatures, demonstrating their superiority over traditional analytical methods. Additionally, Ao et al. (2022) proposed an even more effective approach involving training a CNN to estimate

dip, which is then used to calculate curvature. Therefore, I propose further evaluating the other curvature (or dip) estimation methods to obtain a better curvature attribute with higher resolution and lower noise content.

In the pursuit of improved data quality, I posit that enhancing the distance-to-fault attribute is pivotal. In the fault probability estimation, I employed a pre-trained CNN by Wu et al. (2019a), which was exclusively trained on synthetic seismic data. I acknowledge that this may not yield realistic seismic examples of the presalt formation. Wu et al.'s (2019a) approach utilises 1D convolution and a simple reflectivity model to simulate seismic data. This results in a relatively monotonous amplitude seismic dataset with well-defined faults. In that sense, I needed to apply two steps of seismic preconditioning in an attempt to approximate the characteristics of actual seismic data to those of the synthetic seismic data used in training. Incorporating more realistic synthetic data in CNN training could potentially obviate the need for preconditioning. As a straightforward improvement, I propose the adoption of 2D (or 3D) point-spread functions over 1D convolution to simulate synthetic seismic data (e.g., Lecomte, 2005). Jing et al. (2022) have demonstrated that CNNs trained on this type of data exhibit superior fault delineation on real seismic datasets.

As previously mentioned, another potential improvement involves including a discrete feature indicating mechanical facies. This discrete feature can significantly enhance the accuracy of the models.

Furthermore, applying the fracture intensity attribute as input into geostatistical methods like Sequential Gaussian Simulation (SGS) or Discrete Fracture Network (DFN) holds promise. These approaches leverage well information as hard data, utilizing the fracture intensity attribute estimated by the machine learning method as a constraint (secondary variable) to extrapolate fracture intensity values away from the well. This is, in fact, a noteworthy aspect of the present results. Geostatistical methods typically require a secondary variable that exhibits linear correlation with the target variable, often not true for most seismic attributes and fracture intensity. As I demonstrated, the estimated fracture intensity attribute holds a strong linear correlation with the actual fracture intensity measured in the wells.

Indeed, this combination of tree-based ensemble regressors and geostatistical methods for estimating rock properties forms the basis of the EMBER method developed

by Daly (2020) and subsequently applied by Ferreira et al. (2021) to model porosity and permeability in Santos' presalt.

Adhering to this proposed workflow of integrating tree-based ensemble regressors with geostatistical methods allows for accurate modelling of porosity and fracture permeability away from wells, while also accounting for multiple aspects of the rock, such as level of strain (e.g., curvature, distance to fault, and fault throw) and rock brittleness (mechanical facies, brittleness index).

8. Conclusions

In the first part, I evaluate the reasonableness of using geometric attributes, especially the k1 curvature, to estimate fracture intensity in the presalt. I showed that k1 curvature is a better proxy for fracture than coherence attributes. However, care must be taken with the parametrization of such attributes when performing quantitative studies of fracture intensity. Depending on the algorithm of dip estimation used prior to the computation of k1 curvature or the parametrization of post-filtering techniques, the correlation of k1 curvature and fracture intensity can be ruined.

I also demonstrate that while having a strong correlation with fracture intensity, curvature alone does not tell the whole story of fracture development. Different rocks, with different stiffness, under the same stress, should deform differently or, in this context, develop fractures with a distinct intensity.

In the second part of the study, I showed that P-Impedance is a reasonable proxy for the mechanical stratigraphy effect once stiffer rocks tend to have higher P-Impedance. Together with distance to fault and k1 curvature, the three features could reasonably predict fracture intensity away from the wells.

Some improvements in the workflow are recommended, especially in the part of mechanical stratigraphy, where the addition of a fourth input feature indicating the mechanical facies has the potential to enhance the final result.

Although not by much, XGBoost outperformed the Random Forest in the present analysis. The evaluation and aggregation of other machine learning regressors is suggested, especially regressors that can extrapolate.

Finally, despite the advances and insights that this workflow brings to the fracture modelling workflow, the outcome of this present research is an intermedium result that does not produce the final output for a fracture modelling workflow. I understand that this result should be used as a secondary variable in a geostatistical approach for fracture simulation to use the well information as hard data.

I expect that the present study helps to highlight the importance of aggregating multiple pieces of information from different natures to properly estimate fracture

intensity in Santos' presalt. In that sense, despite its limitations, data-driven machine learning-based approaches seem to be a good option.

9. References

AlBinHassan, N.M., Luo, Y., Al-Faraj, M.N., 2006. 3D edge-preserving smoothing and applications. Geophysics 71, 5–11. https://doi.org/10.1190/1.2213050

Al-Dossary, S., Marfurt, K.J., 2006. 3D volumetric multispectral estimates of reflector curvature and rotation. Geophysics 71. https://doi.org/10.1190/1.2242449

Allwardt, P.F., Bellahsen, N., Pollard, D.D., 2007. Curvature and fracturing based on global positioning system data collected at Sheep Mountain anticline, Wyoming. Geosphere 3, 408–421. https://doi.org/10.1130/GES00088.1

Alves, T.M., Fetter, M., Lima, C., Cartwright, J.A., Cosgrove, J., Gangá, A., Queiroz, C.L., Strugale, M., 2017. An incomplete correlation between pre-salt topography, top reservoir erosion, and salt deformation in deep-water Santos Basin (SE Brazil). Marine and Petroleum Geology 79, 300–320. https://doi.org/10.1016/j.marpetgeo.2016.10.015

An, Y., Guo, J., Ye, Q., Childs, C., Walsh, J., Dong, R., 2021. Deep convolutional neural network for automatic fault recognition from 3D seismic datasets. Computers and Geosciences 153. https://doi.org/10.1016/j.cageo.2021.104776

Ao, Y., Lu, W., Jiang, B., Monkam, P., 2021. Seismic Structural Curvature Volume Extraction With Convolutional Neural Networks. IEEE Transactions on Geoscience and Remote Sensing 59. https://doi.org/10.1109/TGRS.2020.3042098

Ao, Y., Lu, W., Xu, P., Jiang, B., 2022. Seismic Dip Estimation with a Domain Knowledge Constrained Transfer Learning Approach. IEEE Transactions on Geoscience and Remote Sensing 60. https://doi.org/10.1109/TGRS.2021.3061438

Askaripour, M., Saeidi, A., Mercier-Langevin, P., Rouleau, A., 2022. A Review of Relationship between Texture Characteristic and Mechanical Properties of Rock. Geotechnics 2, 262–296. https://doi.org/10.3390/geotechnics2010012

Astratti, D., Aarre, D., Vejbæk, O. V., White, G., 2015. Mapping and time-lapse analysis of South Arne Chalk fault network using new developments in seismic dip computation. Geological Society Special Publication 406, 331–358. https://doi.org/10.1144/SP406.10

Babasafari, A.A., Campane Vidal, A., Furlan Chinelatto, G., Rangel, J., Basso, M., 2022. Ensemble-based machine learning application for lithofacies classification in a pre-salt carbonate reservoir, Santos Basin, Brazil. Petroleum Science and Technology 0, 1–17. https://doi.org/10.1080/10916466.2022.2143813

Babasafari, A.A., Chinelatto, G.F., Vidal, A.C., 2022. Fault and fracture study by incorporating borehole image logs and supervised neural network applied to the 3D seismic attributes: a case study of pre-salt carbonate reservoir, Santos Basin, Brazil. Petroleum Science and Technology 0, 1–20. https://doi.org/10.1080/10916466.2021.2025072

Bahorich, M., Farmer, S., 1995. 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube. The Leading Edge 14, 1053–1058. https://doi.org/10.1190/1.1437077

Bakker, P., van Vliet, L.J., Verbeek, P.W., 1999. Edge preserving orientation adaptive filtering. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1, 535–540. https://doi.org/10.1109/cvpr.1999.786989

Barnes, A., 2016. Handbook of Poststack Seismic Attributes. https://doi.org/10.1190/1.9781560803324

Barnes, A.E., 1996. Theory of two-dimensional complex seismic trace analysis. Geophysics 61, 264–272. https://doi.org/10.1190/1.1822844

Barnett, A.J., Fu, L., Rapasi, T., Scotellaro, C., Guha, J., Cabolova, A., Domingues, A.L., 2020. Seismic characterization and origin of clinoforms in lacustrine depositional environments: A case study from the cretaceous of the south atlantic. Geological Society Special Publication 509, 127–145. https://doi.org/10.1144/SP509-2019-148

Basir, H.M., Javaherian, A., Yaraki, M.T., 2013. Multi-attribute ant-tracking and neural network for fault detection: A case study of an Iranian oilfield. Journal of Geophysics and Engineering 10. https://doi.org/10.1088/1742-2132/10/1/015009

Bengio, Y., Courville, A., Vincent, P., 2013. Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Bergbauer, S., Pollard, D.D., 2003. How to calculate normal curvatures of sampled geological surfaces. Journal of Structural Geology 25, 277–289. https://doi.org/10.1016/S0191-8141(02)00019-6

Bhattacharya, S., Ghahfarokhi, P.K., Carr, T.R., Pantaleone, S., 2019. Application of predictive data analytics to model daily hydrocarbon production using petrophysical, geomechanical, fiber-optic, completions, and surface data: A case study from the Marcellus Shale, North America. Journal of Petroleum Science and Engineering 176, 702–715. https://doi.org/10.1016/j.petrol.2019.01.013

Bhattacharya, S., Mishra, S., 2018. Applications of machine learning for facies and fracture prediction using Bayesian Network Theory and Random Forest: Case studies from the Appalachian basin, USA. Journal of Petroleum Science and Engineering 170, 1005–1017. https://doi.org/10.1016/j.petrol.2018.06.075

Botter, C., Cardozo, N., Hardy, S., Lecomte, I., Escalona, A., 2014. From mechanical modeling to seismic imaging of faults: A synthetic workflow to study the impact of faults on seismic. Marine and Petroleum Geology 57, 187–207. https://doi.org/10.1016/j.marpetgeo.2014.05.013

Braga, I.L.S., Moraes, F.S., 2013. High-resolution gathers by inverse Q filtering in the wavelet domain. Geophysics 78, V53–V61. https://doi.org/https://doi.org/10.1190/geo2011-0508.1

Bravo, L., Aldana, M., 2010. Volume curvature attributes to identify subtle faults and fractures in carbonate reservoirs: Cimarrona formation, middle Magdalena valley basin, Colombia. SEG Technical Program Expanded Abstracts 2010 2231–2235. https://doi.org/10.1190/1.3513293

Breiman, L., 1996. Bagging predictors. Machine Learning 24, 123–140. https://doi.org/https://doi.org/10.1007/bf00058655

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/https://doi.org/10.1007/9781441993267_5

Buddin, T.S., Kane, S.J., Williams, G.D., Egan, S.S., 1997. A sensitivity analysis of 3-dimensional restoration techniques using vertical and inclined shear constructions. Tectonophysics 269, 33–50. https://doi.org/10.1016/S0040-1951(96)00099-6

Carlens, H., 2023. State of Competitive Machine Learning in 2022. ML Contests Research.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. https://doi.org/10.1145/2939672.2939785

Chen, Y., Liao, Z., Fu, L.Y., Zhou, G., Xu, L., Marfurt, K.J., Mu, X., Zou, H., 2020. Effect of main frequencies on characterizing fault damage-zones using forward modeling and attribute of variance. Interpretation 8. https://doi.org/10.1190/int-2020-0017.1

Chilès, J.-P., Wackernagel, H., Beucher, H., Lantuéjoul, C., Élion, P., 2008. ESTIMATING FRACTURE DENSITY FROM A LINEAR OR AREAL SURVEY. VIII International Geostatistics Congress. Santiago, Chile.

Chinelatto, G.F., Belila, A.M.P., Basso, M., Souza, J.P.P., Vidal, A.C., 2020. A taphofacies interpretation of shell concentrations and their relationship with petrophysics: A case study of Barremian-Aptian coquinas in the Itapema Formation, Santos Basin-Brazil. Marine and Petroleum Geology 116, 104317. https://doi.org/10.1016/j.marpetgeo.2020.104317

Chopra, S., 2009. Interpreting fractures through 3D seismic discontinuity attributes and their visualization., CSEG Recorder.

Chopra, S., Marfurt, K.J., 2015. Is curvature overrated ? No , it depends on the geology. First Break 33, 29–40. https://doi.org/10.3997/1365-2397.2014021

Chopra, S., Marfurt, K.J., 2017. Volumetric fault image enhancement — Some applications. Interpretation 5, 151–161. https://doi.org/10.1190/INT-2016-0129.1

Chopra, S., Marfurt, K.J., 2019. Multispectral, multiazimuth, and multioffset coherence attribute applications. Interpretation 7, SC21–SC32. https://doi.org/10.1190/int-2018-0090.1

Chopra, S., Marfurt, K.J., 2007. Seismic Attributes for Prospect Identification and Reservoir Characterization, Seismic Attributes for Prospect Identification and Reservoir Characterization. https://doi.org/10.1190/1.9781560801900

Chopra, S., Marfurt, K.J., 2007. Volumetric curvature attributes adding value to 3D seismic data interpretation. Society of Exploration Geophysicists - 77th SEG International Exposition and Annual Meeting, SEG 2007 851–855.

Chopra, S., Marfurt, K.J., 2013. Structural curvature versus amplitude curvature. The Leading Edge 37, 178–184. https://doi.org/https://doi.org/10.1190/1.3628237

Chopra, S., Marfurt, K.J., 2020. Adopting multispectral dip components for coherence and curvature attribute computations. The Leading Edge 39, 593–596. https://doi.org/10.1190/tle39080593.1

Chopra, S., Marfurt, K., Staples, E., 2013. Testing Curvature's Impact on Fractures - A primary factor in controlled lab work., AAPG Explorer 42–43.

Chopra, S., Misra, S., Marfurt, K.J., 2011. Coherence and curvature attributes on preconditioned seismic data., The Leading Edge 386–393. https://doi.org/https://doi.org/10.1190/1.3575281

Chopra, S., Staples, E., Marfurt, K., Reches, Z., 2013. Quantitative Curvature Analysis: A Case Study., AAPG Explorer 48–49.

Correa, R.S.M., Pereira, C.E.L., Cruz, F.A.S., Lisboa, S.N.D., Junior, M.P.A., Carvalho, B.R.B.M., Souza, V.H.P., Rocha, C.H.A., Araujo, F.G., 2019. Integrated Seismic-Log-Core-Test Fracture Characterization, Barra Velha Formation, Pre-salt of Santos Basin. Search and Discovery. https://doi.org/10.1306/42425Correa2019

Costa de Oliveira, L., Rancan, C.C., Sartorato, A.C.L., Farias, F.A., Pereira, E., 2021. Drowning unconformities on presalt carbonate platforms – Examples from the Itapema Formation (Lower Cretaceous), Santos Basin, offshore Brazil. Palaeogeography, Palaeoclimatology, Palaeoecology 577, 110570. https://doi.org/10.1016/j.palaeo.2021.110570

Cunha, A., Pochet, A., Lopes, H., Gattass, M., 2019. Seismic fault detection in real data using Transfer Learning from a Convolutional Neural Network pre- trained with synthetic seismic data. Computers and Geosciences 104344. https://doi.org/10.1016/j.cageo.2019.104344

Daly, C., 2021. An Application of an Embedded Model Estimator to a Synthetic Nonstationary Reservoir Model With Multiple Secondary Variables. Frontiers in Artificial Intelligence 4, 1–12. https://doi.org/10.3389/frai.2021.624697

De Luca, P.H.V., Matias, H., Carballo, J., Sineva, D., Pimentel, G.A., Tritlla, J., Esteban, M., Loma, R., Alonso, J.L.A., Jiménez, R.P., Pontet, M., Martinez, P.B., Vega, V., 2017. Breaking barriers and paradigms in presalt exploration: The pão de açúcar discovery (Offshore Brazil). AAPG Memoir. American Association of Petroleum Geologists, 177–193. https://doi.org/10.1306/13572007M1133686

De Ros, L.F., 2021. Syngenetic, diagenetic and hydrothermal processes in the pre-salt sag section of Santos and Campos Basins. 2nd EAGE Conference on Pre-Salt Reservoir. 8–11. https://doi.org/10.3997/2214-4609.202183007

Deriche, R., 1993. Recursively implementating the Gaussian and its derivatives. Rapports de Recherche INRIA 24.

Dershowitz, W.S., Herda, H.H., 1992. Interpretation of fracture spacing and intensity. 33rd US Symposium on Rock Mechanics, Santa Fe, 3–5 June 1992. Publ Rotterdam: A A Balkema, 1992, International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts, P757–766.

Dewett, D.T., Henza, A.A., 2016. Spectral similarity fault enhancement. Interpretation 4, SB149–SB159. https://doi.org/10.1190/INT-2015-0114.1

Di, H., Gao, D., 2014. A new algorithm for evaluating 3D curvature and curvature gradient for improved fracture detection. Computers and Geosciences 70, 15–25. https://doi.org/10.1016/j.cageo.2014.05.003

Di, H., Gao, D., 2015. Efficient volumetric extraction of most positive/negative curvature and flexure for fracture characterization from 3D seismic data. Geophysical Prospecting 64, 1454–1468. https://doi.org/10.1111/1365-2478.12350

Di, H., Gao, D., 2017. 3D Seismic Flexure Analysis for Subsurface Fault Detection and Fracture Characterization. Pure and Applied Geophysics 174, 747–761. https://doi.org/10.1007/s00024-016-1406-9

Di, H., Shafiq, M., AlRegib, G., 2018. Patch-level MLP classification for improved fault detection. SEG Technical Program Expanded Abstracts 2211–2215. https://doi.org/10.1190/segam2018-2996921.1

Dinç Gogus, Ö., Avşar, E., Develi, K., Çalık, A., 2023. Quantifying the Rock Damage Intensity Controlled by Mineral Compositions: Insights from Fractal Analyses. Fractal and Fractional 7. https://doi.org/10.3390/fractalfract7050383

Donath, A., F., 1970. Some Information Squeezed Out of a Rock. AmSci 58, 54–72.

Dou, Y., Li, K., Zhu, J., Li, X., Xi, Y., 2021. Attention-Based 3-D Seismic Fault Segmentation Training by a Few 2-D Slice Labels. IEEE Transactions on Geoscience and Remote Sensing 60, 1–15. https://doi.org/10.1109/TGRS.2021.3113676

Dou, Y.T. of 3D S.I.F.S.N. under S.L. by W.A.A., Li, K., Zhu, J., Li, T., Tan, S., Huang, Z., 2021. Efficient Training of 3D Seismic Image Fault Segmentation Network under Sparse Labels by Weakening Anomaly Annotation. 1–12.

Du, K., Sun, Y., Zhou, J., Khandelwal, M., Gong, F., 2022. Mineral Composition and Grain Size Effects on the Fracture and Acoustic Emission (AE) Characteristics of Rocks Under Compressive and Tensile Stress. Rock Mechanics and Rock Engineering 55, 6445–6474. https://doi.org/10.1007/s00603-022-02980-y

Fatah, T.Y.A., Freire, A.F.M., Gamboa, L.A.P., Lupinacci, W.M., 2021. Estimate Borehole Fracture Density Based on Image Logs : A Case Study on Presalt Carbonate Reservoirs of the Santos Basin Estimate Borehole Fracture Density Based on Image Logs : A Case Study on Presalt Carbonate Reservoirs of the Santos Basin. 17th International Congress of the Brazilian Geophysical Society & Expogef. Universidade Federal Fluminense, Niterói, Rio de Janeiro - Brazil.

Fatah, T.Y.A., Freire, A.F.M., Gamboa, L.A.P., Lupinacci, W.M., 2021. Fracture characterization using borehole images in a presalt field, Santos Basin. 50o Congresso Brasileiro de Geologia.

Fernández-Ibáñez, F., Bowen, M., Jones, G., Mimoun, J., 2022. Excess permeability in the Brazil pre-salt: Non-matrix types and diagnostic indicators. AAPG Bulletin 1–38. https://doi.org/10.3997/2214-4609.201982006

Fernandez-Ibanez, F., Nolting, A., Breithaupt, C.I., Darby, B., Mimoun, J., Henares, S., 2022. The properties of faults in the Brazil pre-salt: A reservoir characterization perspective. Marine and Petroleum Geology 146, 105955. https://doi.org/10.1016/J.MARPETGEO.2022.105955

Ferreira, D.J.A., de Oliveira, G.M.B., Castro, T.M., Dias, R.M., Lupinacci, W.M., 2021. Geostatistics assisted by machine learning for reservoir property modeling: A case study in presalt carbonates of Buzios Field, Brazil. The Leading Edge 40, 876–885. https://doi.org/10.1190/tle40120876.1

Ferrill, D.A., Morris, A.P., McGinnis, R.N., Smart, K.J., Wigginton, S.S., Hill, N.J., 2017. Mechanical stratigraphy and normal faulting., Journal of Structural Geology. https://doi.org/10.1016/j.jsg.2016.11.010

Ferrill, D.A., Morris, A.P., 2008. Fault zone deformation controlled by carbonate mechanical stratigraphy , Balcones fault system , Texas. AAPG Bulletin 92, 359–380. https://doi.org/10.1306/10290707066

Fossen, H., Rotevatn, A., 2016. Fault linkage and relay structures in extensional settings-A review. Earth-Science Reviews 154, 14–28. https://doi.org/10.1016/j.earscirev.2015.11.014

Francelino, A.V.M., Antunes, A.F., 2013. Applying filters and seismic attributes for enhancing faults in the 3D seismic survey of Alto de Siririzinho (Sergipe-Alagoas Basin, northeast Brazil). Revista Brasileira de Geofisica 31, 109–123. https://doi.org/10.22564/rbgf.v31i1.250

Fred P. Wang, J.F.W.G., 2009. Screening Criteria for Shale-Gas Systems.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics 29, 1189–1232. https://doi.org/10.1214/aos/1013203451

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36, 193–202. https://doi.org/10.1007/BF00344251

Gao, D., 2013. Integrating 3D seismic curvature and curvature gradient attributes for fracture characterization: Methodologies and interpretational implications. Geophysics 78. https://doi.org/10.1190/GEO2012-0190.1

Girshick, R., 2015. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision 1440–1448. https://doi.org/https://doi.org/10.48550/arXiv.1504.08083

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT Press. https://doi.org/10.1007/978-981-13-9113-2_16

Grieser, B., Bray, J., 2007. Identification of Production Potential in Unconventional Reservoirs. https://doi.org/10.2118/106623-MS

Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on tabular data? 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Gu, Y.J., Sacchi, M., 2009. Radon transform methods and their applications in mapping mantle reflectivity structure. Surveys in Geophysics 30, 327–354. https://doi.org/10.1007/s10712-009-9076-0

Gudmundsson, A., De Guidi, G., Scudero, S., 2013. Length-displacement scaling and fault growth. Tectonophysics 608, 1298–1309. https://doi.org/10.1016/j.tecto.2013.06.012

Guitton, A., 2018. 3D Convolutional Neural Networks for Fault Interpretation. 80th Conference + Exhibition. Copenhagen, Denmark.

Gunther, M.J., Marfurt, K.J., 2016. A Comparison of Alternative Volumetric Dip Computations. SEG International Exposition and 86th Annual Meeting. 2124–2127.

Guo, B., Liu, L., Luo, Y., 2018. A new method for automatic seismic fault detection using convolutional neural network. SEG Technical Program Expanded Abstracts 1951–1955. https://doi.org/10.1190/segam2018-2995894.1

Guo, Y., Zhang, K., Marfurt, K.J., 2010. Seismic attribute illumination of Woodford Shale faults and fractures, Arkoma Basin, OK. Society of Exploration Geophysicists International Exposition and 80th Annual Meeting 2010. Denver, USA, 1372–1376. https://doi.org/10.1190/1.3513097

Guo, Y., Peng, S., Du, W., Li, D., 2020. Fault and horizon automatic interpretation by CNN: A case study of coalfield. Journal of Geophysics and Engineering 17, 1016–1025. https://doi.org/10.1093/jge/gxaa060

Guo, Z., Chapman, M., Li, X., 2012. A shale rock physics model and its application in the prediction of brittleness index, mineralogy, and porosity of the Barnett Shale. SEG Technical Program Expanded Abstracts 2406–2410. https://doi.org/10.1190/SEGAM2012-0777.1

H. McQuillan, 1973. Small-Scale Fracture Density in Asmari Formation of Southwest Iran and its Relation to Bed Thickness and Structural Setting. AAPG Bulletin 57, 2367–2385. https://doi.org/10.1306/83d9131c-16c7-11d7-8645000102c1865d

Hale, D., 2002. Recursive Gaussian Filters. Golden, CO. https://doi.org/10.1109/icpr.1998.711192

Hale, D., 2013. Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images. Geophysics 78. https://doi.org/10.1190/GEO2012-0331.1

Handin, J., Hager, R. V., 1957. Experimental Deformation of Sedimentary Rocks Under Confining Pressure: Tests at Room Temperature on Dry Samples. AAPG Bulletin 41. https://doi.org/10.1306/5CEAE5FB-16BB-11D7-8645000102C1865D

Hart, B.S., Sagan, J.A., 2007. Curvature for visualization of seismic geomorphology. Geological Society Special Publication 277, 139–149. https://doi.org/10.1144/GSL.SP.2007.277.01.08

Hennings, P.H., Olson, J.E., Thompson, L.B., 2000. Combining Outcrop Data and Three-Dimensional Structural Models to Characterize Fractured Reservoirs: An Example from Wyoming 1, AAPG Bulletin.

Herlinger, R., Zambonato, E.E., De Ros, L.F., 2017. Influence of Diagenesis On the Quality of Lower Cretaceous Pre-salt Lacustrine Carbonate Reservoirs from Northern Campos Basin, Offshore Brazil. Journal of Sedimentary Research 87, 1285–1313. https://doi.org/10.2110/jsr.2017.70

Homma, T., Atlas, L.E., Marks, R.J., 1987. Artificial Neural Network for Spatio-Temporal Binary Patterns: Application To Phoneme Classification. Advances in Neural Information Processing Systems 1, 31–40.

Huang, L., Dong, X., Clee, T.E., 2017. A scalable deep learning platform for identifying geologic features from seismic attributes. Leading Edge 36, 249–256. https://doi.org/10.1190/tle36030249.1

Huang, X.R., Huang, J.P., Li, Z.C., Yang, Q.Y., Sun, Q.X., Cui, W., 2015. Brittleness index and seismic rock physics model for anisotropic tight-oil sandstone reservoirs. Applied Geophysics 12, 11–22. https://doi.org/10.1007/S11770-014-0478-0/METRICS

Hughes, C., Darby, B., 2021. Impact of fault displacement on syn-extensional carbonate deposition and excess permeability. 2nd EAGE Conference on Pre-Salt Reservoir. https://doi.org/10.3997/2214-4609.202183039

Hunt, L., Beshry, B., Chopra, S., Webster, C., 2018. Determination of target-oriented parameters for computation of curvature attributes. Interpretation 6, 485–498. https://doi.org/10.1190/INT-2017-0202.1

Hunt, L., Reynolds, S., Brown, T., Hadley, S., Downton, J., Chopra, S., 2010. Quantitative estimate of fracture density variations in the Nordegg with azimuthal AVO and curvature: A case study. The Leading Edge 29, 1122–1137. https://doi.org/10.1190/1.3485773

Hunt, L., Reynolds, S., Hadley, S., Downton, J., Chopra, S., 2011. Causal fracture prediction: Curvature, stress, and geomechanics. The Leading Edge 30, 1274–1286. https://doi.org/https://doi.org/10.1190/1.3663400

Jensen, K., 2021. Seismic modelling and deconvolution of prestack depth migrated images through ray-based Point-Spread Functions. University of Bergen.

Jesus, C., Olho Azul, M., Moreira Lupinacci, W., Machado, L., 2019. Multiattribute framework analysis for the identification of carbonate mounds in the Brazilian presalt zone. Interpretation 7, T467–T476. https://doi.org/10.1190/int-2018-0004.1

Jesus, I.L. De, Lebre, M.B.S., Carmo, M.C., Fatah, T.Y.A., Freire, A.F.M., Lupinacci, W.M., 2021. Impact of the Silica Content on the Porosity and Permeability in carbonates of the Barra Velha Formation , Santos Basin , and the response in the Acoustic Impedance Impact of the Silica Content on the Porosity and Permeability in carbonates

of the Barra V. 17th International Congress of the Brazilian Geophysical Society & Expogef. Universidade Federal Fluminense, Niterói, Rio de Janeiro - Brazil, 0–4.

Jiang, J., Pan, H., Li, M., Qian, B., Lin, X., Fan, S., 2021. Predictive model for the 5-year survival status of osteosarcoma patients based on the SEER database and XGBoost algorithm. Scientific Reports 11, 1–9. https://doi.org/10.1038/s41598-021-85223-4

Jibrin, B., Yelwa, N.A., 2015. Delineating faults using multi-trace seismic attributes : Example from offshore Niger Delta. IOSR Journal of Applied Geology and Geophysics 3, 40–47. https://doi.org/10.9790/0990-03114047

Jing, J., Yan, Z., Zhang, Z., Gu, H., Han, B., 2022. Fault detection using a CNN trained with PSF-Convolution-based samples. Geophysics 88, 1–64. https://doi.org/10.1190/geo2021-0824.1

John F. Harris, Taylor, G.L., Walper, J.L., 1960. Relation of Deformational Fractures in Sedimentary Rocks to Regional and Local Structure. AAPG Bulletin 44, 1853–1873. https://doi.org/10.1306/0bda6257-16bd-11d7-8645000102c1865d

Keating, D.P., Fischer, M.P., 2008. An experimental evaluation of the curvature-strain relation in fault-related folds. American Association of Petroleum Geologists Bulletin 92, 869–884. https://doi.org/10.1306/03060807111

Kibikas, W.M., Ghassemi, A., Carpenter, B.M., 2021. Evaluating the mechanical properties of carbonate and evaporite caprocks in the Sichuan Basin. Journal of Asian Earth Sciences: X 6, 100063. https://doi.org/10.1016/j.jaesx.2021.100063

Klein, P., Richard, L., James, H., 2008. 3D curvature attributes: a new approach for seismic interpretation., First Break 26, 105–112.

Kosari, E., Ghareh-Cheloo, S., Kadkhodaie-Ilkhchi, A., Bahroudi, A., 2015. Fracture characterization by fusion of geophysical and geomechanical data: A case study from the Asmari reservoir, the Central Zagros fold-thrust belt. Journal of Geophysics and Engineering 12, 130–143. https://doi.org/10.1088/1742-2132/12/1/130

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25, 1097–1105. https://doi.org/10.1201/9781420010749

Kuhn, S., Cracknell, M.J., Reading, A.M., 2018. Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of Australia. Geophysics 83, B183–B193. https://doi.org/10.1190/geo2017-0590.1

Kuncheva, L.I., Whitaker, C.J., 2001. Ten measures of diversity in classifier ensembles: Limits for two classifiers. IEE Colloquium (Digest) 73–82. https://doi.org/10.1049/ic:20010105

Lecomte, I., 2004. Simulating Prestack Depth Migrated Sections. EAGE 66th Conference & Exhibition - Paris, France, 7-10 June, 2004. https://doi.org/10.3997/2214-4609-pdb.3.p071

Lecomte, I., Hamran, S.E., Gelius, L.J., 2005. Improving Kirchhoff migration with repeated local plane-wave imaging? A SAR-inspired signal-processing approach in prestack depth imaging. Geophysical Prospecting 53, 767–785. https://doi.org/10.1111/j.1365-2478.2005.00501.x

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation applied to digit recognition., Neural computation.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE 86, 2278–2324. https://doi.org/10.1109/5.726791

Levenberg, K., 1944. A Method for the Solution of Certain Non-Linear Problems in Least Squares. Quarterly of Applied Mathematics 2, 164–168.

Liao, Z., Li, W., Zou, H., Hao, F., Marfurt, K.J., Reches, Z., 2020. Composite damage zones in the subsurface. Geophysical Journal International 222, 225–230. https://doi.org/10.1093/gji/ggaa158

Lima, B.E.M., De Ros, L.F., 2019. Deposition, diagenetic and hydrothermal processes in the Aptian Pre-Salt lacustrine carbonate reservoirs of the northern Campos Basin, offshore Brazil., Sedimentary Geology. https://doi.org/10.1016/j.sedgeo.2019.01.006

Lisle, R.J., 1994. Detection of zones of abnormal strains in structures using Gaussian curvature analysis. American Association of Petroleum Geologists Bulletin 78, 1811–1819. https://doi.org/10.1306/a25ff305-171b-11d7-8645000102c1865d

Liu, Z., Sun, Z., 2015. New brittleness indexes and their application in shale/clay gas reservoir prediction. Petroleum Exploration and Development 42, 129–137. https://doi.org/10.1016/S1876-3804(15)60016-7

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. IEEE Conference on Computer Vision and Pattern Recognition. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions Scott. Advances in Neural Information Processing Systems 30 (NIPS 2017). 1208–1217. https://doi.org/https://doi.org/10.48550/arXiv.1705.07874

Luo, Y., Marhoon, M., Dossary, S. Al, Alfaraj, M., 2002. Edge-preserving smoothing and applications. The Leading Edge 21, 136–158. https://doi.org/https://doi.org/10.1190/1.1452603

Lupinacci, W.M., Fatah, T.Y.A., Carmo, M.C. do, Freire, A.F.M., Gamboa, L.A.P., 2023. Controls of fracturing on porosity in pre-salt carbonate reservoirs. Energy Geoscience 4, 100146. https://doi.org/10.1016/j.engeos.2022.100146

Machado, G., Alali, A., Hutchinson, B., Olorunsola, O., Marfurt, K.J., 2016. Display and enhancement of volumetric fault images. Interpretation 4, SB51–SB61. https://doi.org/10.1190/INT-2015-0104.1

Machado, G., Alali, A., Hutchinson, B., Oluwatobi, O., Marfurt, K.J., 2015. Improving fault images using a directional laplacian of a gaussian operator. SEG Technical Program Expanded Abstracts. 1851–1855. https://doi.org/10.1190/segam2015-5920781.1

Machado, M., Moliterno, A., de Oliveira, A., Gattass, M., 2013. Computing Curvature by Quadratic Approximation for Steep Dips. Thirteenth International Congress of The Brazilian Geophysical Society. 1607–1611. https://doi.org/10.1190/sbgf2013-329

Mai, H.T., Marfurt, K.J., Chávez-pérez, S., Mexicano, 2009. Coherence and volumetric curvatures and their spatial relationship to faults and folds, an example from Chicontepec basin, Mexico. Mexico SEG Houston 2009 International Exposition and Annual Meeting. 1063–1067.

Man, H.Q., Hien, D.H., Thong, K.D., Dung, B.V., Hoa, N.M., Hoa, T.K., Kieu, N. Van, Ngoc, P.Q., 2021. Hydraulic flow unit classification and prediction using machine learning techniques: A case study from the nam con son basin, offshore vietnam. Energies 14. https://doi.org/10.3390/en14227714

Mandal, P.P., Rezaee, R., Emelyanova, I., 2021. Ensemble learning for predicting TOC from well-logs of the unconventional goldwyer shale. Energies 15. https://doi.org/10.3390/en15010216

Mandujano, J.J., Khachaturov, R. V., Tolson, G., Duncan Keppie, J., 2005. Curvature analysis applied to the Cantarell structure, southern Gulf of Mexico: Implications for hydrocarbon exploration. Computers and Geosciences 31, 641–647. https://doi.org/10.1016/j.cageo.2004.11.018

Marfurt, K.J., Kirlin, R.L., Farmer, S.L., Bahorich, M.S., 2012. 3-D seismic attributes using a semblance-based coherency algorithm. Https://Doi.Org/10.1190/1.1444415 63, 1150–1165. https://doi.org/10.1190/1.1444415

Marfurt, K.J., 2006. Robust estimates of 3D reflector dip and azimuth. Geophysics 71. https://doi.org/10.1190/1.2213049

Marfurt, K.J., 2018. Chapter 2. Seismic Attributes and What They Measure. Seismic Attributes as the Framework for Data Integration Throughout the Oilfield Life Cycle. Society of Exploration Geophysicists, 25–149. https://doi.org/10.1190/9781560803522.ch2

Marquardt, D.W., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. Journal of the Society for Industrial and Applied Mathematics 11, 431–441. https://doi.org/https://doi.org/10.1137/0111030

Mayolle, S., Soliva, R., Caniven, Y., Wibberley, C., Ballas, G., Dominguez, S., Milési, G., 2019. Scaling of fault damage zones and implications for naturally fractured reservoirs. 5th International Conference on Fault and Top Seals 2019. https://doi.org/10.3997/2214-4609.201902342

Mendes, L.D.C., Correia, U.M.C., Rubio, O., Oliveira, F.M., Campane, A., 2022. Topological analysis of fault network in naturally fractured reservoirs : A case study from the pre-salt section of the Santos Basin , Brazil. Journal of Structural Geology 159, 104597. https://doi.org/10.1016/j.jsg.2022.104597

Meng, F., Wong, L.N.Y., Zhou, H., 2021. Rock brittleness indices and their applications to different fields of rock engineering: A review., Journal of Rock Mechanics and Geotechnical Engineering. https://doi.org/10.1016/j.jrmge.2020.06.008

Montaron, B., Montaron, F., Montaron, R., 2019. A DNN trained to pick faults in seismic data: Implementation and results. International Petroleum Technology Conference 2019, IPTC 2019 26–28. https://doi.org/10.2523/iptc-19116-ms

Moore, C.H., Wade, W.J., 2013. Chapter 11 - Natural Fracturing in Carbonate Reservoirs. In: Moore, C.H., Wade, W.J. (Eds.), Developments in Sedimentology. Elsevier, 285–300. https://doi.org/10.1016/B978-0-444-53831-4.00011-2

Moreira, J.L.P., Madeira, C.V., Gil, J.A., Machado, M.A.P., 2007. Bacia de Santos. Boletim de Geociências Petrobras 15, 531–549.

Naar, W., 1991. Fracture Density in the Deep Subsurface: Techniques with Application to Point Arguello Oil Field. AAPG Bulletin 75, 1300–1323.

Nelson, R.A., 2001. 3 - Detecting and Predicting Fracture Occurrence and Intensity. Geologic Analysis of Naturally Fractured Reservoirs. 125–162.

Nelson, R.A., 2001. 1 - Evaluating Fractured Reservoirs: Introduction. Geologic Analysis of Naturally Fractured Reservoirs. 1–100.

Nguyen, N.T., Tran, N.T.H., Hoang, K.S., Tran, V.T., 2022. Supervised machine learning application of lithofacies classification for a hydrodynamically complex gas - condensate reservoir in Nam Con Son basin. Petrovietnam Journal 6, 27–35. https://doi.org/10.47800/pvj.2022.06-03

Odoh, B.I., Ilechukwu, J.N., Okoli, N.I., 2014. The Use of Seismic Attributes to Enhance Fault Interpretation of O T Field , Niger Delta. International Journal of Geosciences 2014, 826–834.

Ojeda, P., Elmo, D., Rogers, S., Brzovic, A., 2023. Discrete Fracture Network (DFN) Analysis to Quantify the Reliability of Borehole-Derived Volumetric Fracture Intensity. Geosciences 13, 187. https://doi.org/10.3390/geosciences13060187

Otchere, D.A., Tackie-Otoo, B.N., Mohammad, M.A.A., Arbi Ganat, T.O., Kuvakin, N., Miftakhov, R., Efremov, I., Bazanov, A., 2022. Improving seismic fault mapping through data conditioning using a pre-trained deep convolutional neural network: A case study on Groningen field. Journal of Petroleum Science and Engineering 213, 110411. https://doi.org/10.1016/j.petrol.2022.110411

Pan, X.P., Zhang, G.Z., Chen, J.J., 2020. The construction of shale rock physics model and brittleness prediction for high-porosity shale gas-bearing reservoir. Petroleum Science 17, 658–670. https://doi.org/10.1007/S12182-020-00432-2/FIGURES/8

Pan, X.P., Zhang, G.Z., Chen, J.J., 2020. The construction of shale rock physics model and brittleness prediction for high-porosity shale gas-bearing reservoir. Petroleum Science 17, 658–670. https://doi.org/10.1007/s12182-020-00432-2

Peacock, D.C.P., Nixon, C.W., Rotevatn, A., Sanderson, D.J., Zuluaga, L.F., 2016. Glossary of fault and other fracture networks. Journal of Structural Geology 92, 12–29. https://doi.org/10.1016/j.jsg.2016.09.008

Pearce, M.A., Jones, R.R., Smith, S.A.F., McCaffrey, K.J.W., 2011. Quantification of fold curvature and fracturing using terrestrial laser scanning. AAPG Bulletin 95, 771–794. https://doi.org/10.1306/11051010026

Pedersen, S.I., Randen, T., Sønneland, L., Steen, Ø., 2002. Automatic fault extraction using artificial ants. SEG Technical Program Expanded Abstracts 21, 512–515. https://doi.org/10.1190/1.1817297

Penna, R., Araújo, S., Geisslinger, A., Sansonowski, R., Oliveira, L., Rosseto, J., Matos, M., 2019. Carbonate and igneous rock characterization through reprocessing, FWI imaging, and elastic inversion of a legacy seismic data set in Brazilian presalt province. The Leading Edge 38, 11–19. https://doi.org/https://doi.org/10.1190/tle38010011.1. Abstract

Pietzsch, R., Oliveira, D.M., Tedeschi, L.R., Queiroz Neto, J. V., Figueiredo, M.F., Vazquez, J.C., de Souza, R.S., 2018. Palaeohydrology of the Lower Cretaceous pre-salt lacustrine system, from rift to post-rift phase, Santos Basin, Brazil. Palaeogeography, Palaeoclimatology, Palaeoecology 507, 60–80. https://doi.org/10.1016/j.palaeo.2018.06.043

Pochet, A., Diniz, P.H.B., Lopes, H., Gattass, M., 2019. Seismic Fault Detection Using Convolutional Neural Networks Trained on Synthetic Poststacked Amplitude Maps. IEEE Geoscience and Remote Sensing Letters 16, 352–356. https://doi.org/10.1109/LGRS.2018.2875836

Qi, X., Marfurt, K.J., 2017. Volumetric aberrancy to map subtle faults and flexures. SEG International Exposition and 87th Annual Meeting. 3443–3447. https://doi.org/https://doi.org/10.1190/segam2017-17632807.1

Qian, K.R., Liu, T., Liu, J.Z., Liu, X.W., He, Z.L., Jiang, D.J., 2020. Construction of a novel brittleness index equation and analysis of anisotropic brittleness characteristics for unconventional shale formations. Petroleum Science 17, 70–85. https://doi.org/10.1007/s12182-019-00372-6

Randen, T., Monsen, E., Signer, C., Abrahamsen, A., Hansen, J.O., Sæter, T., Schlaf, J., Sønneland, L., 2000. Three-dimensional texture attributes for seismic data analysis. 2000 SEG Annual Meeting 2–5. https://doi.org/https://doi.org/10.1190/1.1816155

Ren, K., Oliveira, M.J.R., Zhao, Junfeng, Zhao, Jian, Oliveira, L., Rincan, C., Carmo, I., Deng, Q., 2019. Using Wireline Logging and Thin Sections to Identify Igneous Contact Metamorphism and Hydrothermal Influence on Presalt Limestone Reservoirs in Libra Block, Santos Basin. Offshore Technology Conference Brasil Please. Rio de Janeiro, Brazil. https://doi.org/https://doi.org/10.4043/29818-MS

Rickman, R., Mullen, M., Petre, E., Grieser, B., Kundert, D., 2008. A Practical Use of Shale Petrophysics for Stimulation Design Optimization: All Shale Plays Are Not Clones of the Barnett Shale. Proceedings - SPE Annual Technical Conference and Exhibition 2, 840–850. https://doi.org/10.2118/115258-MS

Roberts, A., 2001. Curvature attributes and their application to 3D interpreted horizons. First Break. https://doi.org/10.1046/j.0263-5046.2001.00142.x

Ronneberger, O., Fischer, P., Brox, T., 2015. Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing Switerzland, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Savage, H.M., Brodsky, E.E., 2011. Collateral damage: Evolution with displacement of fracture distribution and secondary fault strands in fault damage zones. Journal of Geophysical Research: Solid Earth 116. https://doi.org/10.1029/2010JB007665

Scornet, E., 2023. Trees, forests, and impurity-based variable importance in regression. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 59, 1–40. https://doi.org/10.1214/21-aihp1240

Sen, S., Kainkaryam, S., Ong, C., Sharma, A., 2020. SaltNet: A production-scale deep learning pipeline for automated salt model building. Leading Edge 39, 195–203. https://doi.org/10.1190/tle39030195.1

Shaban, A., Sherkati, S., Miri, S.A., 2011. Comparison between curvature and 3D strain analysis methods for fracture predicting in the Gachsaran oil field (Iran). Geological Magazine 148, 868–878. https://doi.org/10.1017/S0016756811000367

Simm, R., Bacon, M., 2014. Seismic Amplitude - An Interpreter's Handbook. Seismic Amplitude - An Interpreter's Handbook. Cambridge University Press, New York. https://doi.org/10.1017/cbo9780511984501

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings 1–14.

Sollich, P., Krogh, A., 1996. Learning with ensembles: How over-fitting can be useful. Proceedings of the 1995 Conference 4–10.

Souza, R.S., Arienti, L.M., Viana, S.M., Falcão, L.C., Cuglieri, M.A., Filho, R.P.S., Leite, C.O., Oliveira, V.C., M., D.O., Anjos, C., Amora, R., Carmo, I.D., Coelho, C.E., 2018. Petrology of the Hydrothermal and Evaporitic Continental Cretaceous (Aptian) Pre-Salt Carbonates and Associated Rocks, South Atlantic Santos Basin, Offshore Brazil. AAPG ACE 2018. Salt Lake City, Utah.

Staples, E., Marfurt, K., 2013. Testing Curvature's Impact on Fractures. AAPG Explorer 41147, 42–43.

Stearns, D.W., Friedman, M., 1972. Reservoirs in Fractured Rock., Stratigraphic Oil and Gas Fields—Classification, Exploration Methods, and Case Histories. https://doi.org/10.1306/M16371C8

Stephenson, B.J., Koopman, A., Hillgartner, H., McQuillan, H., Bourne, S., Noad, J.J., Rawnsley, K., 2007. Structural and stratigraphic controls on fold-related fracturing in the Zagros Mountain, Iran: Implications for reservoir development. Geological Society Special Publication 270, 1–21. https://doi.org/10.1144/GSL.SP.2007.270.01.01

Sun, D.S., Ling, Y., Gao, J., Li, X., Zhang, X., Bai, Y., 2012. Fractured reservoir detection based on fullazimuth seismic data - A case study in south-west China. 74th European Association of Geoscientists and Engineers Conference and Exhibition 2012 Incorporating SPE EUROPEC 2012: Responsibly Securing Natural Resources.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Tanaka, A.P.B., Gomes Borges, J.P., Correa de Matos, G., Relvas Campos, M.T., Cunha, B.M., Brandi de Souza, R., Natan de Moraes Caldeira, J., Soares de Oliveira, T.A., Marçon, D.R., Martins Lima, A.P., 2022. Fault-related fracture modeling in a pre-salt lacustrine carbonate reservoir from Santos Basin, offshore Brazil: Predicting preferential fluid flow paths using 3D geological and flow simulation models. Marine and Petroleum Geology 135, 105392. https://doi.org/10.1016/j.marpetgeo.2021.105392

Taner, M.T., Koehler, F., Sheriff, R.E., 1979. Complex seismic trace analysis. Geophysics 44, 1041–1063. https://doi.org/10.1190/1.1440994

Taner, M.T., Sheriff, R.E., 1977. Application of Amplitude, Frequency, and Other Attributes to Stratigraphic and Hydrocarbon Determination. AAPG Memoir.

Taner, M.T., 2001. SEISMIC ATTRIBUTES, CSEG Recorder.

Terzaghi, R.D., 1965. Sources of error in joint surveys. Geotechnique 15, 287–304. https://doi.org/10.1680/geot.1965.15.3.287

Tingdahl, K.M., 1999. Improving seismic detectability using intrinsic directionality. Goteborge University.

Tingdahl, K.M., Bril, A.H., De Groot, P.F., 2001. Improving seismic chimney detection using directional attributes. Journal of Petroleum Science and Engineering 29, 205–211. https://doi.org/10.1016/S0920-4105(01)00090-0

Tingdahl, K.M., De Groot, P.F.M., 2003. Post-stack-dip- and azimuth processing. Journal of Seismic Exploration 12, 113–126.

Tingdahl, K.M., Rooij, M. de, 2005. Semi-automatic detection of faults in 3-D seismic signals. Geophysical Prospecting 53, 533–542. https://doi.org/10.1190/1.1816520

Van Golf-Racht, T.D., 1996. Chapter 7 Naturally-fractured carbonate reservoirs. In: G.V. Chilingarian, Mazzullo, S.J., Rieke, H.H. (Eds.), Developments in Petroleum Science. Elsevier, 683–771. https://doi.org/10.1016/S0376-7361(96)80029-X

Varughese, G., Kumar, S., 2014. Measurement of Elastic Parameters of Lithium Hydroxylammonium Sulphate Single Crystal, by Ultrasonic Pulse Echo Overlap Technique. Open Journal of Acoustics 04, 138–144. https://doi.org/10.4236/OJA.2014.43014

Vizeu, F., Neto, E.R.D.O., Freire, A.F.M., Lupinacci, W.M., 2021. Convolutional neural network for prediction of igneous seismic facies in the Santos Basin Pre-Salt. 2nd EAGE Conference on Pre-Salt Reservoir 2019–2022. https://doi.org/10.3997/2214-4609.202183008

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., Lang, K.J., 1989. Phoneme Recognition Using Time-Delay Neural Networks. IEEE Transactions on Acoustic, Speech and Signal Processing.

Watkins, H., Butler, R.W.H., Bond, C.E., Healy, D., 2015. Influence of structural position on fracture networks in the Torridon Group, Achnashellach fold and thrust belt, NW Scotland. Journal of Structural Geology 74, 64–80. https://doi.org/10.1016/j.jsg.2015.03.001

Watkins, H., Healy, D., Bond, C.E., Butler, R.W.H., 2018. Implications of heterogeneous fracture distribution on reservoir quality; an analogue from the Torridon

Group sandstone, Moine Thrust Belt, NW Scotlanditle. Journal of Structural Geology 108, 180–197. https://doi.org/https://doi.org/10.1016/j.jsg.2017.06.002

Wennberg, O.P., De Oliveira Ramalho, F., Virgolino Mafia, M., Lapponi, F., Chandler, A.S., Gomis Cartesio, L.E., Hunt, D.W., 2023. The characteristics of natural open fractures in acoustic borehole image logs from the pre-salt Barra Velha formation, Santos Basin, Brazil. Journal of Structural Geology 167. https://doi.org/10.1016/j.jsg.2023.104794

Wennberg, O.P., McQueen, G., de Luca, P.H.V., Hunt, D., Chandler, A.S., Waldum, A., Lapponi, F., 2019. Open fractures in pre-salt reservoirs in the Campos Basin: examples from silicified carbonates in BM-C-33. 1st EAGE Workshop on Pre-Salt Reservoir: From Exploration to Production. https://doi.org/10.3997/2214-4609.201982010

Wennberg, O.P., Svånå, T., Azizzadeh, M., Aqrawi, A.M.M., Brockbank, P., Lyslo, K.B., Ogilvie, S., 2006. Fracture intensity vs. mechanical stratigraphy in platform top carbonates: The Aquitanian of the Asmari Formation, Khaviz Anticline, Zagros, SW Iran. Petroleum Geoscience 12, 235–245. https://doi.org/10.1144/1354-079305-675

Williams, G.D., Kane, S.J., Buddin, T.S., Richards, A.J., 1997. Restoration and balance of complex folded and faulted rock volumes:flexural flattening, jigsaw fitting and decompaction in three dimensions. Tectonophysics 273, 203–218. https://doi.org/10.1016/S0040-1951(96)00282-X

Wu, X., Liang, L., Shi, Y., Fomel, S., 2019. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics 84, IM35–IM45. https://doi.org/10.1190/geo2018-0646.1

Yang, F., Ma, J., 2019. Deep-learning inversion: A next-generation seismic velocity model building method. Geophysics 84, R583–R599. https://doi.org/10.1190/geo2018-0249.1

Yu, Y., 2016. Fault enhancement and visualization with 3D log-gabor filter array. SEG Technical Program Expanded Abstracts 35, 1960–1965. https://doi.org/10.1190/segam2016-11557865.1

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in

Artificial Intelligence and Lecture Notes in Bioinformatics) 8689 LNCS, 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

Zhang, L., Einstein, H.H., 2000. Estimating the intensity of rock discontinuities. International Journal of Rock Mechanics and Mining Sciences 37, 819–37.

Zhang, Z., Lei, Y., Mao, X., Li, P., 2019. CNN-FL: An Effective Approach for Localizing Faults using Convolutional Neural Networks. SANER 2019 - Proceedings of the 2019 IEEE 26th International Conference on Software Analysis, Evolution, and Reengineering 445–455. https://doi.org/10.1109/SANER.2019.8668002

Zhao, Jian, Oliveira, M.J.R., Zhao, Junfeng, Ren, K., Oliveira, L.C., Carmo, I.O., Rancan, C.C., Deng, Q., 2019. Fault activity and its influences on distribution of igneous rocks in Libra block, Santos Basin: Semi-quantitative to quantitative assessment of fault activity based on high-resolution 3D seismic data. Offshore Technology Conference. Rio de Janeiro, Brazil. https://doi.org/10.4043/29691-ms

Zhao, T., Mukhopadhyay, P., 2018. A fault detection workflow using deep learning and image processing. SEG Technical Program Expanded Abstracts 1966–1970. https://doi.org/10.1190/segam2018-2997005.1

Zheng, Z.H., Kavousi, P., Di, H. Bin, 2014. Multi-attributes and neural network-based fault detection in 3d seismic interpretation. Advanced Materials Research 838–841, 1497–1502. https://doi.org/10.4028/www.scientific.net/AMR.838-841.1497

10. Appendices

10.1. A - Well-log Tracks

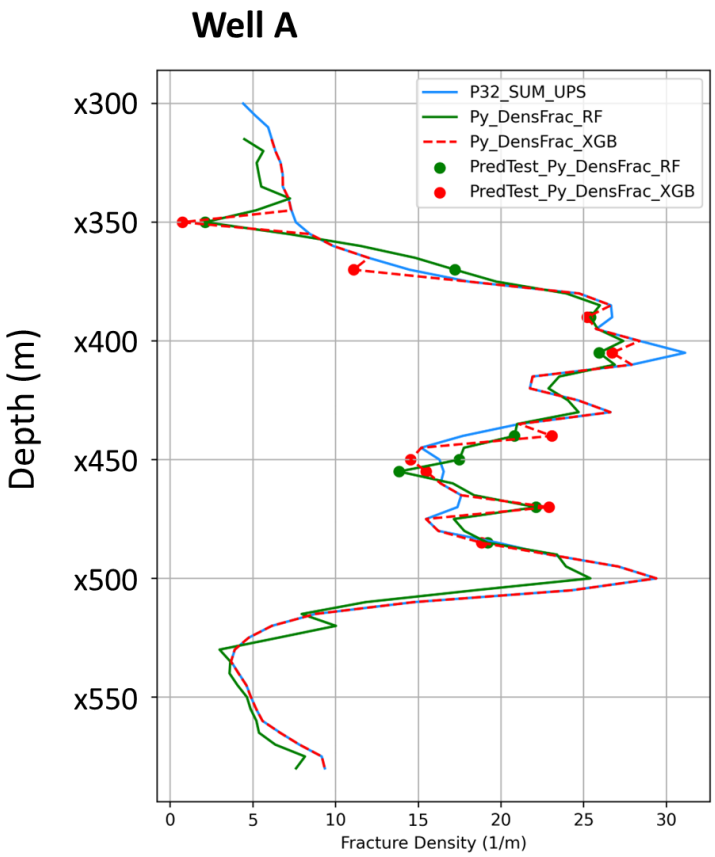

Figure 114: Well log track of Well A showing the actual fracture intensity (P32_SUM_UPS) and the predictions from XGBoost (XGB) and Random Forest (RF), both in the entire well and only for the samples of the test set.

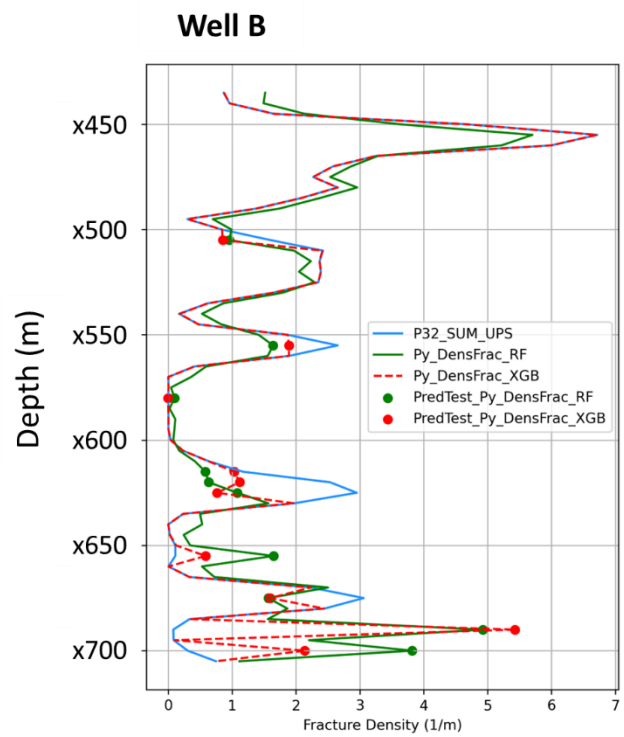

Figure 115: Well log track of Well B showing the actual fracture intensity (P32_SUM_UPS) and the predictions from XGBoost (XGB) and Random Forest (RF), both in the entire well and only for the samples of the test set.

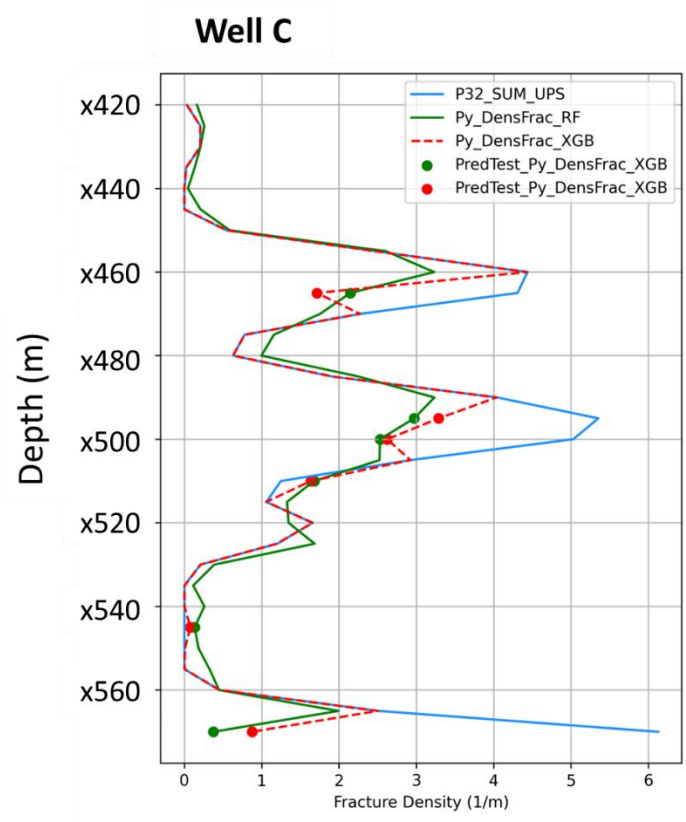

Figure 116: Well log track of Well C showing the actual fracture intensity (P32_SUM_UPS) and the predictions from XGBoost (XGB) and Random Forest (RF), both in the entire well and only for the samples of the test set.

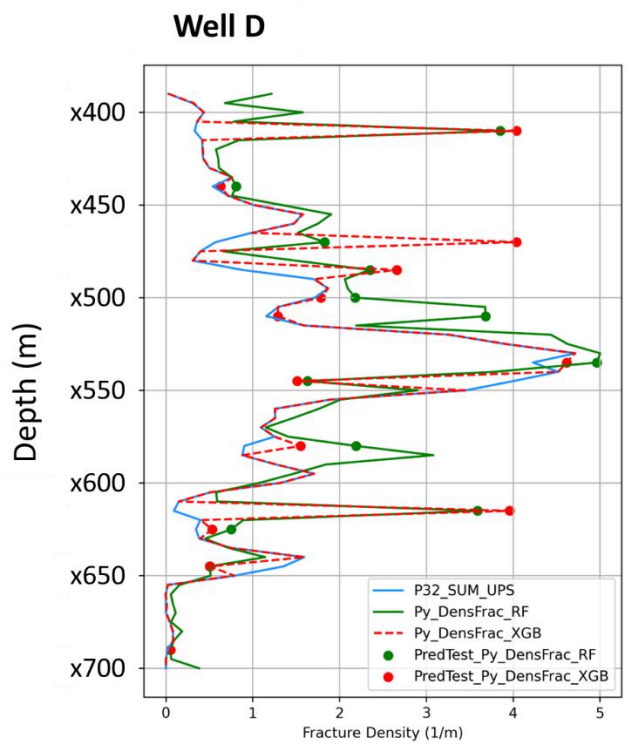

Figure 117: Well log track of Well D showing the actual fracture intensity (P32_SUM_UPS) and the predictions from XGBoost (XGB) and Random Forest (RF), both in the entire well and only for the samples of the test set.

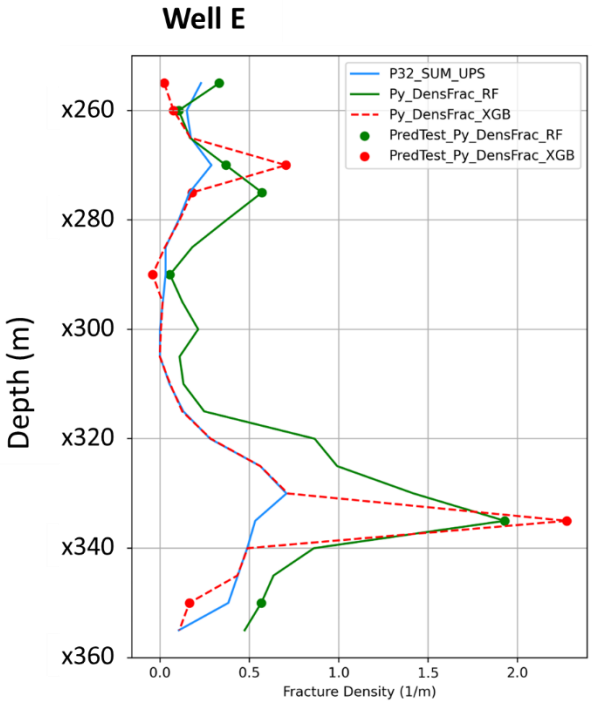

Figure 118: Well log track of Well E showing the actual fracture intensity (P32_SUM_UPS) and the predictions from XGBoost (XGB) and Random Forest (RF), both in the entire well and only for the samples of the test set.

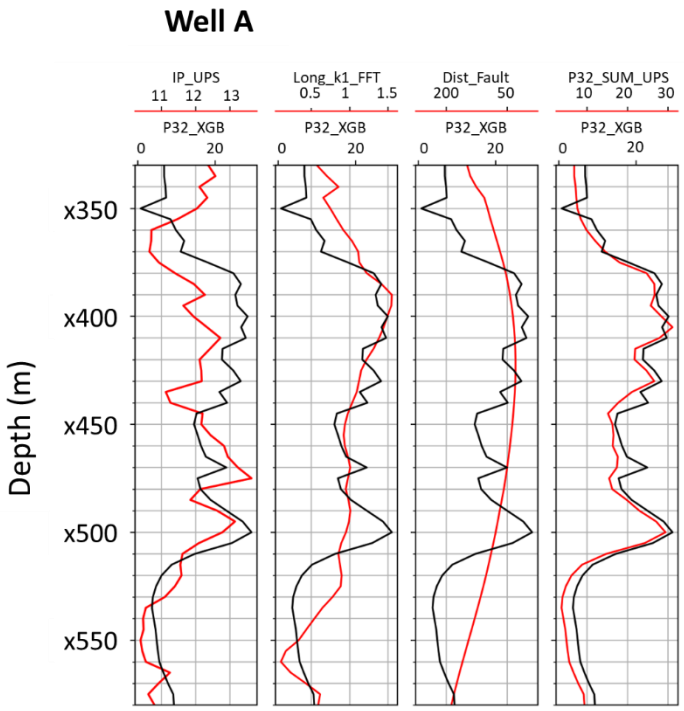

Figure 119: Well log track of Well A showing the correlation visual correlation input (P-Impedance, long wavelength k1 from FFT dip, and distance to fault) and target features (P32_SUM_UPS) with the estimated fracture intensity from XGBoost (P32_XGB).

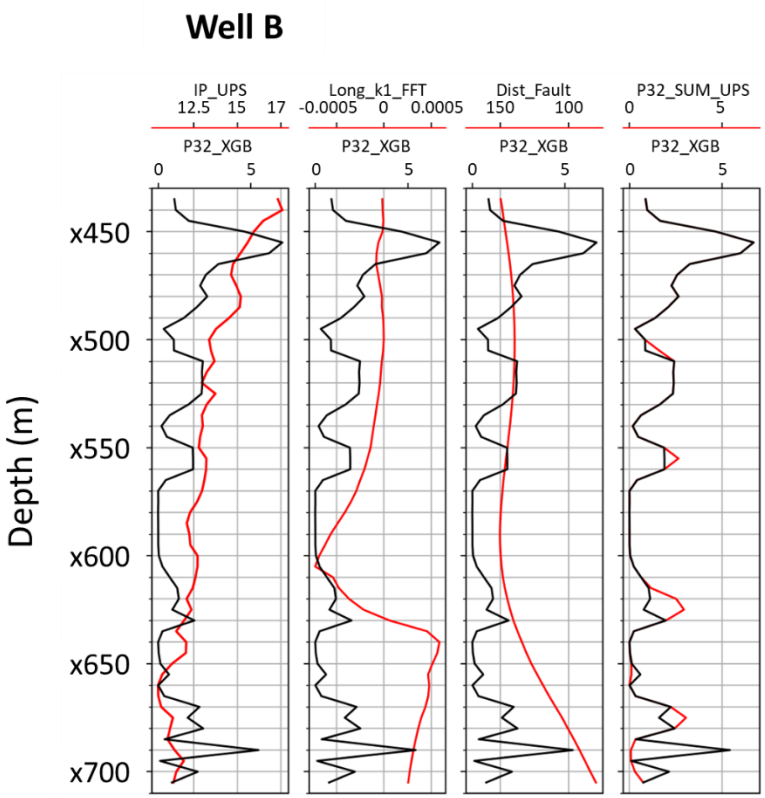

Figure 120: Well log track of Well B showing the correlation visual correlation input (P-Impedance, long wavelength k1 from FFT dip, and distance to fault) and target features (P32_SUM_UPS) with the estimated fracture intensity from XGBoost (P32_XGB).

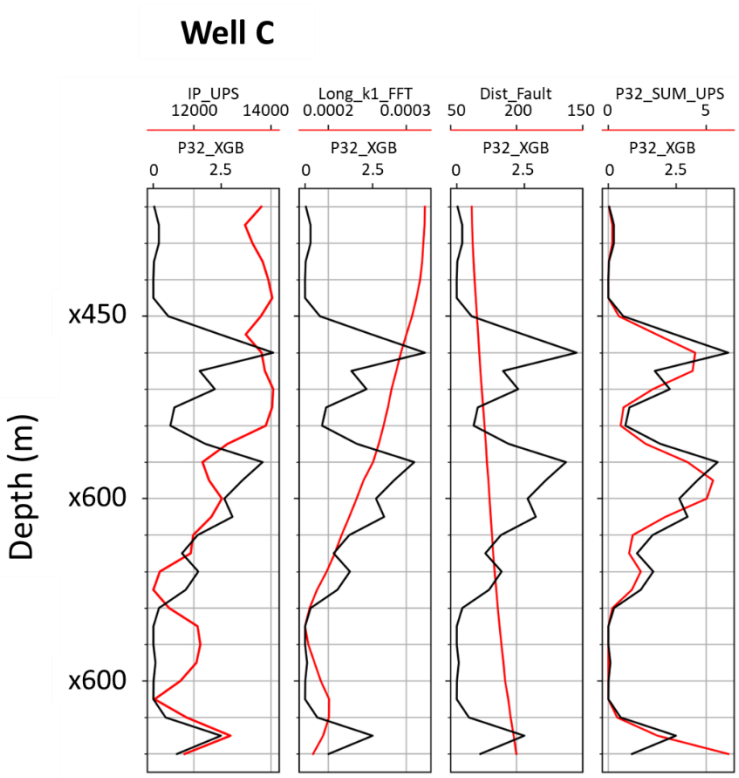

Figure 121: Well log track of Well C showing the correlation visual correlation input (P-Impedance, long wavelength k1 from FFT dip, and distance to fault) and target features (P32_SUM_UPS) with the estimated fracture intensity from XGBoost (P32_XGB).

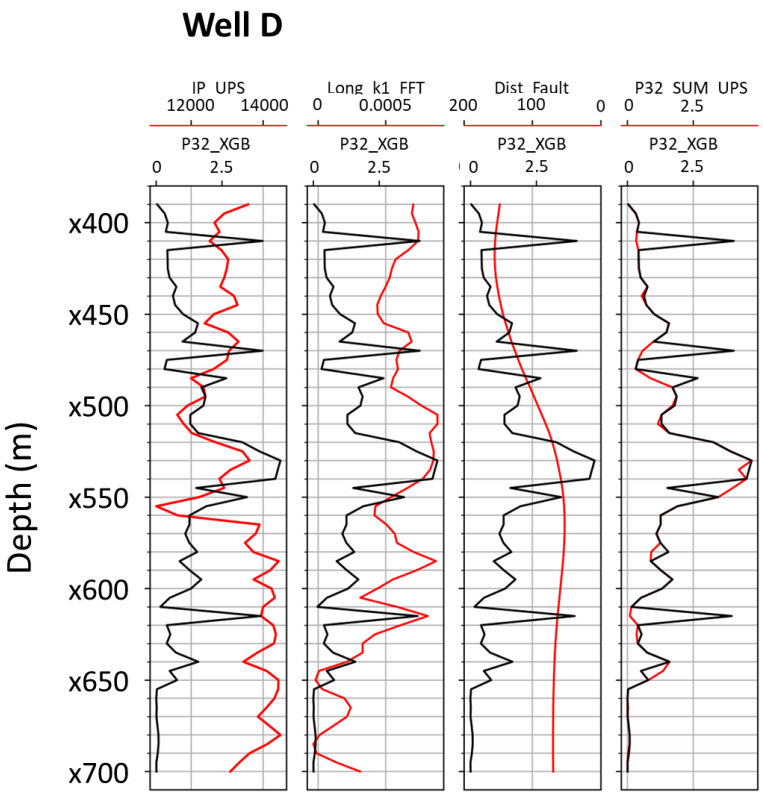

Figure 122: Well log track of Well D showing the correlation visual correlation input (P-Impedance, long wavelength k1 from FFT dip, and distance to fault) and target features (P32_SUM_UPS) with the estimated fracture intensity from XGBoost (P32_XGB).

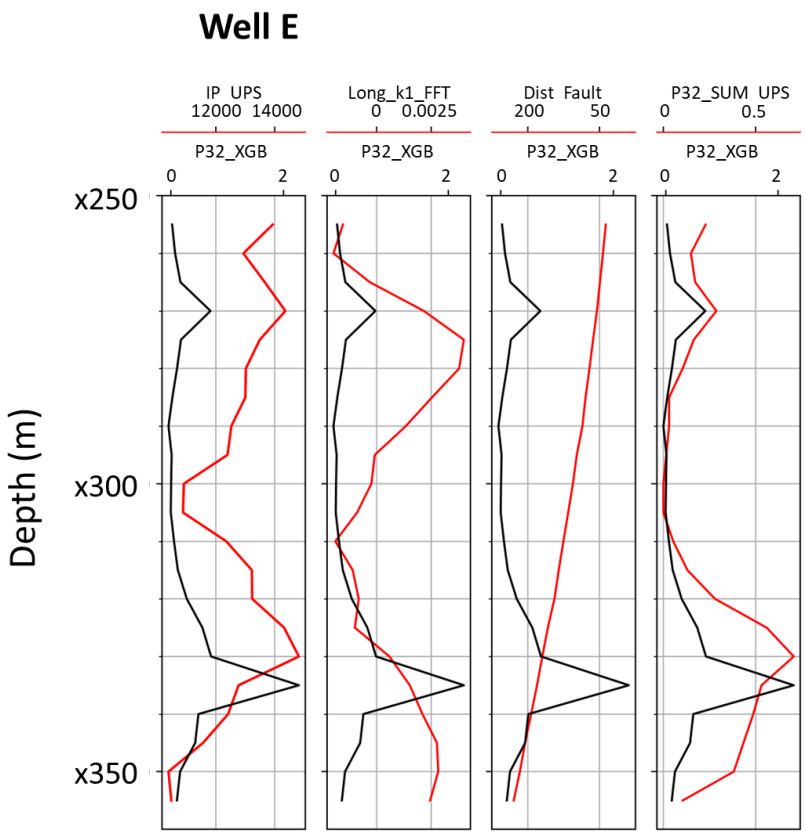

Figure 123: Well log track of Well E showing the correlation visual correlation input (P-Impedance, long wavelength k1 from FFT dip, and distance to fault) and target features (P32_SUM_UPS) with the estimated fracture intensity from XGBoost (P32_XGB).

10.2. B - Seismic Attribute Sections Intersecting the Wells

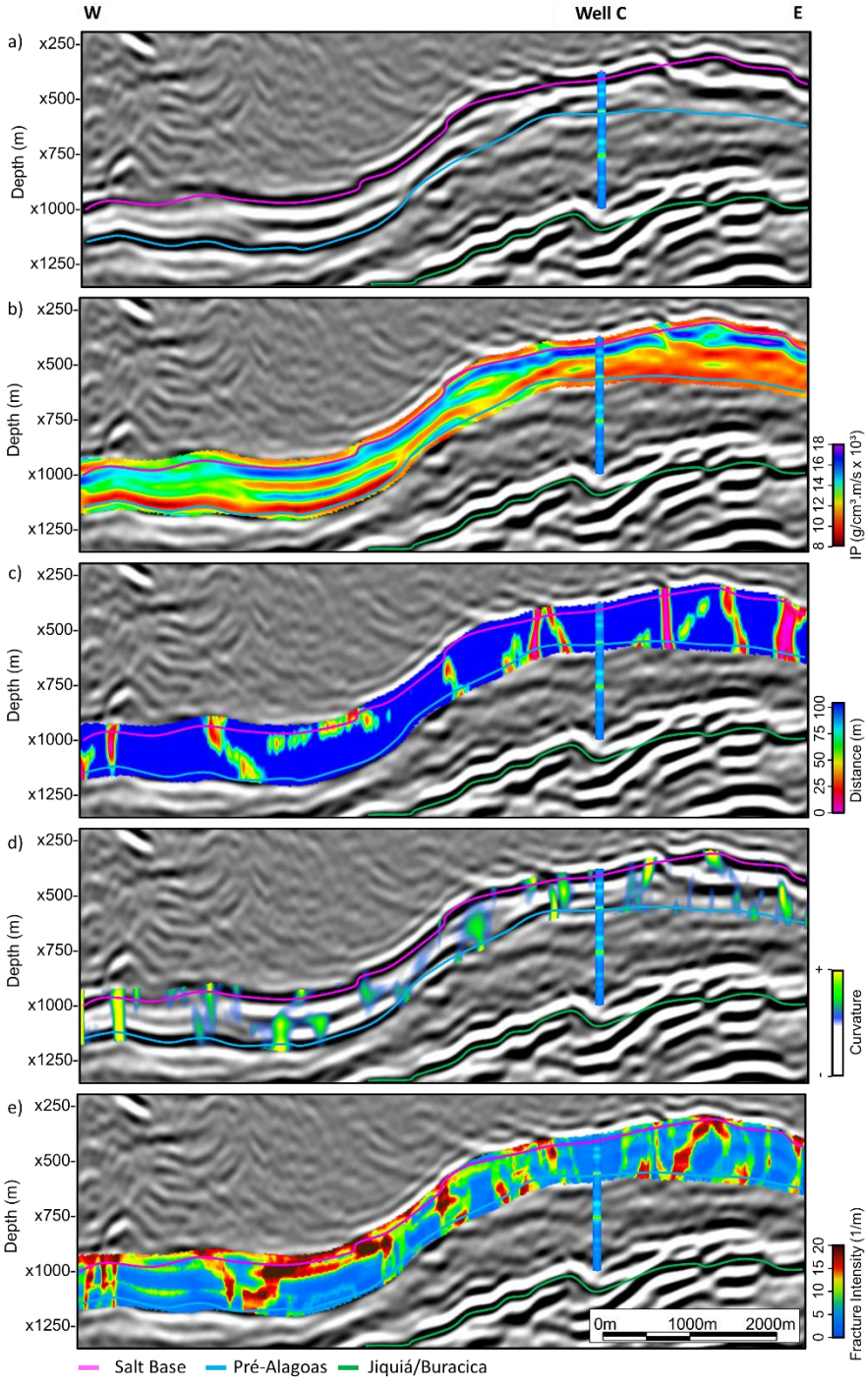

Figure 124: Crossline intersecting the Well C. a) Seismic amplitude; b) Inverted P-Impedance; c) Distance to fault; d) Long wavelength k1 from FFT dip; e) Fracture Intensity predicted by the XGBoost model. The Well C is plotted in each figure for qualitative correlation. The prediction matches the fracture intensity in the well logs. The well drills a carbonate platform facies in the Barra Velha Formation, with faulting occurring at the margins of this platform.

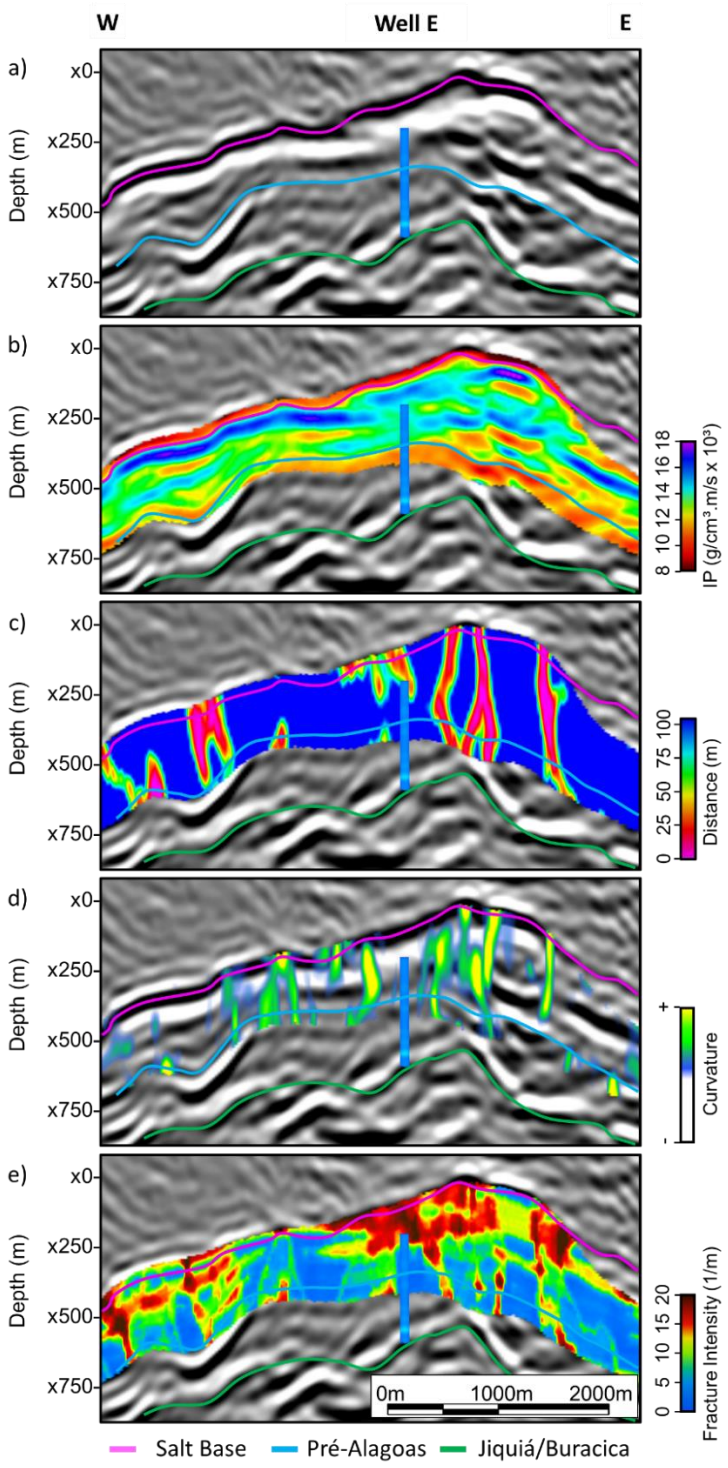

Figure 125: Crossline intersecting the Well E. a) Seismic amplitude; b) Inverted P-Impedance; c) Distance to fault; d) Long wavelength k1 from FFT dip; e) Fracture Intensity predicted by the XGBoost model. The Well E is plotted in each figure for qualitative correlation. The well drills the "protected" flank of a carbonate build-up facies in the Barra Velha formation. Despite the low fracture intensity measured in the wells, the region around it is very fractured, as the estimated fracture intensity indicates. The background of P-Impedance is extremely high, and the k1 curvature and distance to fault attribute indicate an area heavily faulted and deformed.