

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE GEOCIÊNCIAS

0440500 – Trabalho de Formatura (2023)

Monografia (MG)

Ellen Caroline Ferreira de Carvalho 6846508

São Paulo

2023

IDENTIFICAÇÃO

INSTITUIÇÃO: Instituto de Geociências da Universidade de São Paulo

NÚMERO DE PROJETO: TF-23/15

TÍTULO DO PROJETO: Interpretações litológicas e correlação estratigráfica da Fm. Solimões na Bacia do Acre e Bacia do Solimões com base em perfis de poços (ANP) e '*Machine Learning*'

ALUNO(A): Ellen Caroline Ferreira de Carvalho

ORIENTADOR(A): Renato Paes de Almeida, Departamento de Geologia Sedimentar e Ambiental (GSA)

CO-ORIENTADOR(A): Larissa Natsumi Tamura (Instituto de Geofísica - IAG)

SÃO PAULO

2023.

DEDICATÓRIA

Dedico este trabalho à minha família, meu pai, minha mãe e meus irmãos, e amigos feitos ao longo da jornada, minha família não sanguínea, por todo o apoio dado, principalmente nos momentos mais críticos.

E a todos aqueles que ousam sonhar em mudar o mundo mudando a si mesmos, pois é no desejo dos inconformes que o mundo de fato evolui.

Agradecimentos

Primeiramente ao meu pai, Valdemir Ferreira de Carvalho, e minha mãe, Elisabete Souza de Carvalho, por toda dedicação, amor e sacrifício para que eu e meus irmãos, Karen Ferreira de Carvalho e Gustavo Ferreira de Carvalho, pudéssemos romper as barreiras impostas pela sociedade para nossa origem e termos um futuro diferente. Doeu, demorou, mas valeu a pena!

Às minhas melhores amigas, Larissa Mazaia Graziel e Monica Laraia Leonardo da Silva, que nos momentos de maior escuridão nesses mais de 23 anos e 13 anos de amizade, respectivamente, foram a luz e o suporte que eu precisei. Quando eu perdi a fé em mim mesma e na vida, vocês estavam lá para me guiar e mostrar que viver vale a pena, e sem vocês duas nada disso seria possível.

Aos amigos Renegados feitos nessa última etapa, Nicoli Ariel sant'Ana da silva, Diego Fernandes Custodio, Raphael de Oliveira Vianna, Cauby Barboza Vasconcelos, Lucas Almeida Barbosa. No meio de tanto surto e cansaço, vocês alegraram meus dias, me forneceram alguns dos dias mais felizes deste ano e me deram forças para continuar um dia após o outro e amar o Ligea demais.

À Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), por fornecer todos os dados necessários para a realização deste projeto, bem como todo o apoio material e financeiro e todas as oportunidades de contato com profissionais e conteúdos relacionados à área da indústria de petróleo e reservatórios realizadas durante a execução da proposta de Iniciação Científica correlata por intermédio do PRH-43.1.

Ao Laboratório de Informática Geológica, ao Antônio Tadashi Kikuda e à empresa Geoative e Martin Hester que viabilizaram a licença acadêmica do software IP fundamental para a extração dos dados de poços disponibilizados pela ANP.

Ao professor Renato Paes de Almeida, orientador do meu trabalho de formatura e possivelmente uma das pessoas mais inteligentes que já conheci e dono de um dos humores mais refinados, muito obrigada pela orientação e apoio técnico em todas as etapas do projeto, e por ter sido um dos docentes mais especiais que tive contato. Você foi fundamental para que essa etapa pudesse ser concluída com êxito e despertou em mim o interesse e a curiosidade por esse universo amplo e necessário da Análise de Dados e da Linguagem de Computação.

À Larissa Natsumi Tamura, co-orientadora do projeto, que me acompanha desde o início da Iniciação Científica, e tornou possível a realização não apenas deste projeto e da Iniciação, desde a aquisição dos dados e orientação dos procedimentos até as oportunidades realizadas com o PRH, como também a realização de um sonho,

com dedicação, empatia, carinho, suporte de todas as formas e motivação quando eu pensei em desistir no caminho. O IGc ganhou demais com a sua presença e tenho certeza de que muito mais será feito nos próximos anos.

À Geovana Leite Geraldo, co-orientadora do projeto de iniciação e que junto com a Lari conseguiu os dados para a pesquisa e organizou toda a base do projeto de iniciação. Você pavimentou a via para que essa monografia se concretizasse, obrigada.

A todos os funcionários responsáveis por organizar e manter o instituto funcional, que muitas vezes não são reconhecidos pelo excelente trabalho e por todo suporte fornecido. Em especial ao Ubiratan (setor de transportes), Lívia (secretária da Graduação) e José Carlos (audiovisual), por toda ajuda e todos os momentos especiais compartilhados.

A todos os colegas de curso e de turma que me acompanharam ao longo desses cinco anos. A decisão de vir para a geologia mudou a minha vida para melhor e cada um tem participação importante nesse novo caminho, pude aprender demais com cada um. Um agradecimento em especial à Marggiore, que ainda que nossos caminhos tenham se separado um pouco por causa da grade obrigatória, sempre esteve presente tornando meus dias mais agradáveis.

Aos professores que se dedicaram ao trabalho de ensino e tornaram essa formatura possível, muito obrigada. A educação no país é paradoxalmente o maior bem que produzimos e o maior gargalo ao nosso desenvolvimento desde a formação da nossa sociedade, então poder aprender com os melhores da área é um privilégio. Em especial aos professores Sergio Russo Matioli (IB), Viviana Giampaoli (IME), Fabio Ramos Dias de Andrade, Adriana Alves, Rafael Rodrigues de Assis, Claudio Alejandro Salazar Mora, Brenda Chung da Rocha, Renato de Moraes, e Camila Duelis Viana do IGc, não apenas pelas aulas incríveis como também pelo carinho e suporte.

Aos professores Christine Laure Marie Bourotte, Gaston Eduardo Enrich Rojas, Veridiana Teixeira de Souza Martins com quem pude trabalhar durante a monitoria da disciplina 0440108 - Dinâmica do Sistema Terra II (2023) para o curso de Licenciatura em Geociências e Educação Ambiental (Ligea), que reativaram o amor pela docência em mim e demonstraram aula após aula que a arte de ser professor vai muito além de ser perito em determinada aula, mas comprehende também ter empatia suficiente para entender as realidades díspares que vivem os alunos e com sagacidade saber a linguagem correta para transmitir esses conhecimentos e despertar a curiosidade e o amor pela ciência, que foi tão prejudicada principalmente nos últimos anos em nosso país. Se o país tivesse mais docentes como vocês, com o devido investimento governamental, nossa realidade com certeza seria outra.

Ao professor Paulo César Boggiani e a conselheira Rosana de Cássia de Oliveira Oba responsável pelo AcolheGeo. Mais que docente e funcionária vocês foram meus amigos, meu apoio mental e financeiro, conselheiros e um sopro de alívio quando a situação se tornou mais complicada nesses cinco anos. A realização deste sonho tem participação direta de vocês dois e jamais poderei agradecer propriamente por tudo que fizeram por mim. Jamais esquecerei, muito obrigada.

Ao professor José Renato Baptista de Lima (Poli), por servir como um exemplo negativo, tanto enquanto docente e como ser humano. É impossível apagar a experiência que tive nos longos anos durante a Poli, ou esquecer de toda violência e traumas sofridos, mas até mesmo nas piores situações há algo a se aprender, e, apesar de tudo, pude me recompor e crescer, pessoalmente, profissionalmente e emocionalmente. Contra todas as possibilidades e todos os anseios contra isso, a mulher preta periférica formou sim na USP e provou que não era um gasto de dinheiro público.

E por último, na contramão do agradecimento anterior por justamente ter me salvado dele, à professora Alexandra Vieira Suhogusoff, que além de docente ímpar, foi o início dessa jornada. Quando a Poli já havia consumido toda a minha alegria, toda a minha esperança e desejo de permanecer viva, você me encontrou. E foi a primeira pessoa em anos a me ver como um ser humano e a me convencer a não desistir de mim mesma quando eu estava prestes a fazê-lo. Todas as vezes que me lembro de que mais do que uma docente você foi o ser humano mais necessário que encontrei naquele momento, meus olhos se enchem de água como agora escrevendo este agradecimento, e eu jamais poderei agradecer por ter sido aquela que me deu uma segunda chance de viver e me convidou a vir para o IGc. Essa formatura é por você também e se algum dia eu puder ser um décimo do ser humano que você demonstrou ser, eu já terei cumprido meu papel no mundo. Obrigada, de coração.

“Isto é para os loucos. Os desajustados. Os rebeldes. Os criadores de caso. Os que são peças redondas nos buracos quadrados. Os que veem as coisas de forma diferente. Eles não gostam de regras. E eles não têm nenhum respeito pelo status quo. Você pode citá-los, discordar deles, glorificá-los ou difamá-los. Mas a única coisa que você não pode fazer é ignorá-los. Porque eles mudam as coisas. Eles empurram a raça humana para frente. Enquanto alguns os veem como loucos, nós vemos gênios. Porque as pessoas que são loucas o suficiente para achar que podem mudar o mundo são as que, de fato, mudam.”

Steve Jobs

Resumo

O presente estudo concentra-se na formação Solimões nas bacias do Acre e do Solimões, na região amazônica ocidental. Enfatiza a importância da crosta terrestre e dos sedimentos ao longo do Cenozoico. São resumidas as características geográficas e geológicas da bacia enfatizando sua importância geopolítica devido às suas fronteiras com países vizinhos e riqueza de recursos naturais. As pesquisas históricas e geológicas na região são realizadas desde o século XIX e intensificadas no século XX com a perfuração de poços exploratórios.

A classificação da ANP categoriza ambas as bacias em Nova Fronteira, o que significa que há um interesse em aumentar a exploração de hidrocarbonetos para as duas Bacias nos próximos anos.

A falta de padronização e dados geofísicos limitados nos poços, além dos desafios inerentes às diferenças geológicas nas duas bacias e ambientais da região são identificados como desafios, sugerindo a necessidade de abordagens avançadas. Propõe-se a aplicação de ferramentas de análise de dados, programação em Python e *Machine Learning* para aprimorar a resolução estratigráfica e delimitar utilizando inteligência artificial a base da Formação Solimões. Esta abordagem não apenas poderá melhorar a interpretação dos dados nessas bacias estratégicas, mas também poderá ser replicada como modelo em futuros projetos em outras regiões.

Abstract

The present study focuses on the Solimões Formation in the Acre and Solimões basins in the western Amazon region. It emphasizes the importance of the Earth's crust and sediments throughout the Cenozoic era. The geographical and geological characteristics of the basin are summarized, highlighting its geopolitical significance due to its borders with neighboring countries and rich natural resources. Historical and geological research in the region has been conducted since the 19th century and intensified in the 20th century with the drilling of exploratory wells. The Agência Nacional de Petróleo, Gás e Biocombustíveis (ANP) classification categorizes both basins as New Frontier, indicating an interest in increasing hydrocarbon exploration in the coming years.

Challenges, such as the lack of standardization and limited geophysical data in the wells, along with inherent geological and environmental differences between the two basins, are identified, suggesting the need for advanced approaches. The application of data analysis tools, Python programming, and Machine Learning is proposed to enhance stratigraphic resolution and delineate the base of the Solimões Formation using artificial intelligence. This approach would not only improve data interpretation in these strategic basins but could also serve as a model for future projects in other regions.

1. INTRODUÇÃO

A Formação Solimões, unidade mais superficial presente tanto na Bacia do Acre quanto na Bacia do Solimões, registra a história tectonossedimentar do oeste da Amazônia ao longo do Cenozóico. A unidade é a formação aflorante que recobre quase todo o território da Bacia do Acre e a porção oeste da Bacia do Solimões, atingindo espessuras de até 2.200 m e 1.800 m nessas bacias, respectivamente (Cunha, 2007; Loureiro et al., 2019).

A Bacia do Solimões é uma bacia intracratônica localizada na região norte do Brasil no estado do Amazonas com cerca de 950.000 km² de área total, com produção de hidrocarbonetos (óleo + gás + condensado). É limitada ao norte pelo Escudo das Guianas, ao sul pelo Escudo Brasileiro, a leste pelo Arco de Purus e a oeste pelo Arco de Iquitos (Barata e Caputo, 2007).

A Bacia do Acre está situada no extremo oeste do Brasil, representando uma continuidade das bacias de foreland peruanas, como Marañón, Ucayali e Madre de Dios, abrangendo aproximadamente 97.552 km² de área (Alevato, 2019), cuja deposição ocorreu do Cretáceo até o Plioceno (Cunha, 2007). É limitada no território brasileiro a leste pelo Arco de Iquitos - que separa a Bacia do Acre da Bacia do Solimões (Cunha, 2007).

A região norte do país que envolve a Bacia do Acre e a Bacia do Solimões é uma área de alto interesse geopolítico, não apenas por ser região fronteiriça com países vizinhos na América Latina como Peru e Bolívia, com efetivo do Exército Brasileiro na proteção de fronteiras e combate ao garimpo e comércios ilegais, como também pela abundância de recursos minerais e alta biodiversidade, além de ser área de proteção ambiental e residência de diversas comunidades indígenas (Junior, 2022).

Desde o final do século XIX houve investimentos de órgãos públicos e pesquisas de etnógrafos e naturalistas, como a Comissão Rondon (Sá et al., 2008), especialmente na região do estado do Acre. No início do século XX houve um aumento considerável das pesquisas na região devido à criação do Serviço Geológico e Mineralógico do Brasil (SGB), além da atuação de outros órgãos públicos como o Departamento Nacional de Pesquisas Minerais (DNPM) e a Petrobrás e Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), com destaque para os investimentos na década de 70, conduzidos pela Companhia de Pesquisas de Recursos Minerais (CPRM).

Estes investimentos, iniciados na década de 30, foram responsáveis pela instalação de 11 poços exploratórios na Bacia do Acre e 172 poços exploratórios na Bacia do Solimões, de acordo com o site e o banco de dados fornecido pela ANP.

Além disso, pelo contexto geológico que a Bacia do Acre está inserida, na Bacia Hidrográfica do Amazonas como uma bacia de foreland andina multicíclica localizada entre as bacias subandinas Marañon e Ucayali no Peru e na Bolívia e a Bacia do Solimões intracratônica, possuindo algumas características dos dois sistemas geológicos, as recentes descobertas de grandes reservas de hidrocarbonetos nas áreas vizinhas da Bacia de Marañon e na Bacia de Ucayali, ambas no Peru, e a produção atual de hidrocarbonetos na Bacia do Solimões fornecem expectativas de possíveis reservas (Alevato, 2017) e aumentam a pressão para a realização de estudos mais aprofundados na área.

De acordo com a classificação fornecida pela ANP no estudo de bacias, tanto a Bacia do Acre quanto a Bacia do Solimões são consideradas Bacias de Nova Fronteira (Vasconcellos & Coutinho, 2013), isto é, são bacias sedimentares ainda não exploradas ou com insuficiente delimitação total de seu potencial de exploração e produção (ANP, 2022).

Apesar da ocorrência da Formação Solimões tanto na Bacia do Acre como do Solimões, o contexto tectônico - a Bacia do Acre é uma bacia de *foreland* (Hoorn et al., 2010; Wanderley-Filho et al., 2010) e a Bacia do Solimões encontra-se em contexto intracratônico (Shephard, et al., 2010; Flament et al., 2015) –, diferente registro paleoestratigráfico (Negri et al., 2010; Maia et al., 1977) e diferente contexto de ambiente deposicional (Hoorn et al., 2010) gera características distintas na estratigrafia da formação nas duas bacias.

Além disso, há pouca informação concreta sobre a geologia local. Dentre os dados dos poços atualmente instalados nas duas bacias cedidos pela ANP para este projeto, apenas 09 dos 11 poços apresentam dados de profundidade das formações na Bacia do Acre e 45 dos 171 poços disponíveis na Bacia do Solimões, além de não haver uma padronização dos dados geofísicos coletados entre os diferentes poços.

Portanto, a possibilidade de se obter maior resolução da estratigrafia local através do uso de ferramentas de análise de dados e programação com base na linguagem de alto nível Python para análise estatística e propagação de erros, e principalmente no uso de um ramo da inteligência artificial *Machine Learning*, cada vez mais essencial nos estudos das ciências exatas tornando a interpretação dos dados menos subjetiva, é altamente desejável, não apenas no contexto de trazer novas possibilidades de análise para esse setor estratégico do território brasileiro, como também podendo ser utilizado como modelo para outros projetos na região ou em outras localidades no futuro.

2. METAS E OBJETIVOS

O objetivo do presente trabalho de formatura é de delimitar a Formação Solimões na Bacia do Acre e na Bacia do Solimões, definindo a base da mesma em diferentes poços, a partir da análise de dados por *Machine Learning* (método *Random Forest*) e propagação de erros no Python para cálculo de incertezas, para construir um mapa de isópacas.

As metas estabelecidas para o projeto foram:

- 1) Extração dos dados brutos “.dlis” para “.asc”, através do software Interactive Petrophysics (IP), próprio para análise de perfis de poços e muito utilizado na indústria do petróleo e gás, para coleta dos valores medidos de perfilagem geofísica, e conversão para “.csv” que é o formato mais adequado para manipulação dos dados no aprendizado de máquina;
- 2) Realizar a análise dos dados, desenvolvida no Python com uso de códigos no Google Colab, a fim de entender como se dá a distribuição desses dados nos diferentes poços;
- 3) Criar uma base de dados treino adequada, a fim de averiguar a possibilidade de realizar a correlação padrão e delimitar a base da Formação Solimões em poço(s) teste, através do uso de *Machine Learning*;
- 4) Comparar com a informação representada nos perfis compostos deles através de matriz de confusão para avaliação da qualidade do modelo proposto e sua respectiva interpretação

3. JUSTIFICATIVA

De acordo com a nota técnica Nº 8/2022/SAG/ANP-RJ publicada pela ANP em setembro de 2022, a região do norte do Brasil apresenta elevado interesse em pesquisa, incluindo a Bacia do Acre e a Bacia do Solimões nos 25 setores terrestres considerados como Bacias de Nova Fronteira.

Pela definição fornecida pela nota técnica, Bacias de Novas Fronteiras são bacias ou áreas que apresentam “estágios iniciais de conhecimento, com poucos poços perfurados, ausência de produção ou produção limitada a uma pequena porção da bacia, conhecimento insuficiente dos sistemas petrolíferos, muitas vezes sem

qualquer descoberta ou indícios de petróleo e ausência de infraestrutura ou mesmo apresentando desafios tecnológicos para a indústria. Estas bacias, por outro lado, apesar do elevado risco geológico que representam, oferecem oportunidades únicas para empresas de grande porte ou com domínio de tecnologias de ponta. Para estas empresas, as bacias de Novas Fronteiras (**Figura 01**) constituem a oportunidade de grandes descobertas e de grandes projetos.”

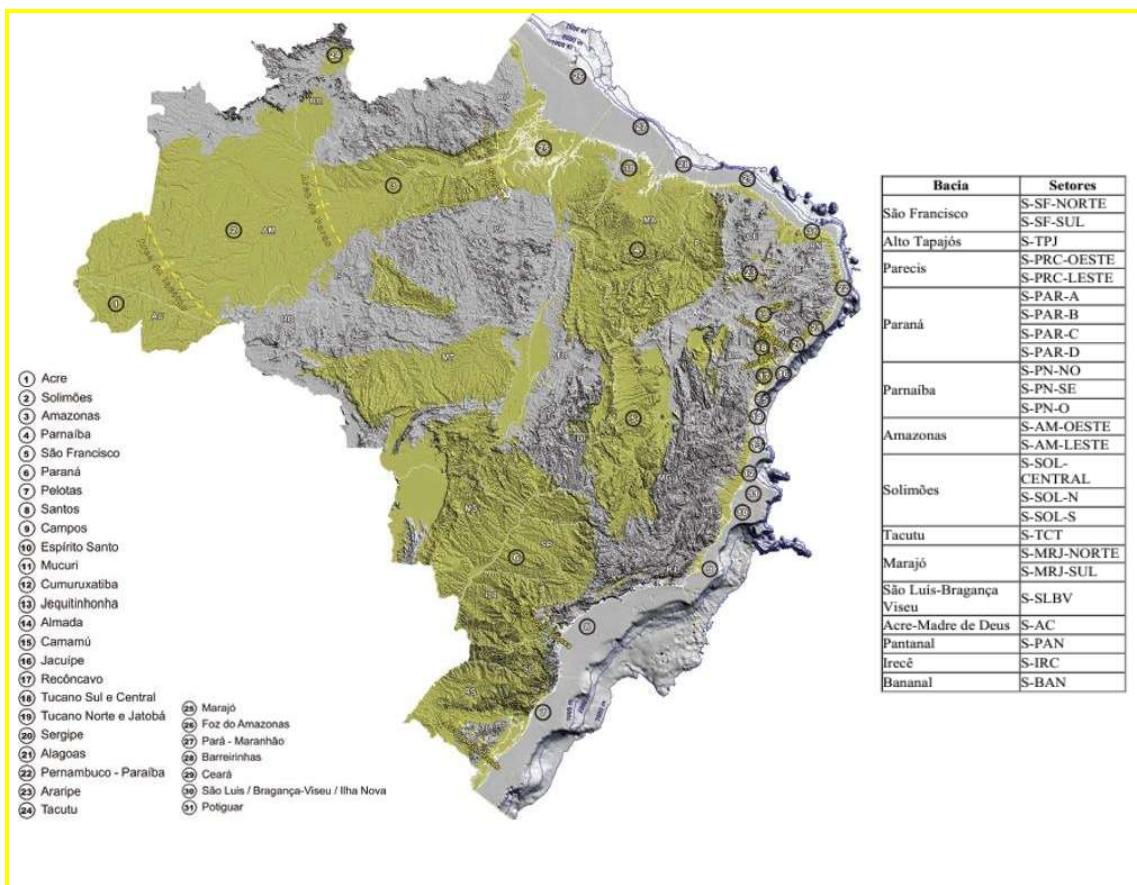


Figura 01. Localização das bacias sedimentares em território brasileiro e os 25 setores denominados Bacias de Nova Fronteira definidos pela nota técnica Nº 8/2022/SAG/ANP-RJ da ANP, que incluem blocos da Bacia do Acre e da Bacia do Solimões. Adaptado de Milani et al. (2007).

Ainda, pelas diretrizes da Resolução CNPE 17/2017, a promoção de exploração em bacias de novas fronteiras visando obter maior conhecimento geológico e a descoberta de novas áreas produtoras é previsto no inciso IV do Art 3º, o que demonstra um elevado interesse da ANP em adquirir novos dados nessas regiões e a possibilidade de novas reservas futuras.

Nesse sentido, o projeto temático Trans-Amazon Drilling Project, coordenado pelo prof. André Sawakuchi vai ao encontro de tais anseios e como parte da equipe de desenvolvimento do projeto, a proposta de se delimitar com maior precisão as

formações que compreendem tais bacias, de maneira mais imparcial com base nos dados adquiridos, além de compreender melhor as diferenças geológicas, de registros paleontológicos e ambientes de deposicionais entre as duas Bacias colabora para o sucesso do projeto.

Diante disso, esse projeto apresenta fundamental importância para se obter uma melhor resolução da Formação Solimões nas duas bacias em sua profundidade, com delimitação de topo e base por inteligência artificial por classificação supervisionada, com a possibilidade inclusive no futuro em próximos projetos de se obter um modelo mais realista para rochas mais profundas análogas que possam apresentar potencial para reserva de hidrocarbonetos.

Dentre as formações disponíveis, a Formação Solimões foi escolhida tendo em vista o fato de ser a sequência mais rasa do Oeste e ser a única comum às duas Bacias (se estendendo até mesmo a partes da Bacia do Amazonas).

A proposta de delimitar a Formação Solimões com base em resultados gerados por *Machine Learning* apresenta alta relevância uma vez que pode ser pioneira para auxiliar em projetos futuros tanto em poços que tal informação não foi disponibilizada - dentre os 182 poços fornecidos pela ANP, excetuando-se poços sem informação de perfilagem geofísica, apenas 52 apresentam detalhamento da profundidade de topo e base da Formação Solimões e de outras formações nos poços, o que corresponde a menos de 30% de todos os poços na área – quanto para a aplicação de uma metodologia robusta em outros poços e bacias, com menor subjetividade na interpretação dos dados litológicos e maior suporte na informação objetiva dos dados geofísicos coletados, utilizando um código aberto e gratuito.

Ressalta-se ainda, a multidisciplinaridade do projeto, utilizando os conhecimentos obtidos durante o curso de Graduação em Geologia de diferentes áreas como Sedimentologia e Estrutural, e os conhecimentos de análises de dados e principalmente *Machine Learning*, que está alinhado com o futuro dos estudos em geologia.

4. FUNDAMENTAÇÃO BIBLIOGRÁFICA

A fundamentação bibliográfica para o presente trabalho se baseou nos estudos do contexto geológico, estratigrafia, geologia estrutural e histórico exploratório para a Bacia do Acre e a Bacia do Solimões, bem como características importantes da Formação Solimões e base teórica de estatística, Python e *Machine Learning*.

4.1 Bacia do Acre: Contexto Geológico

4.1.1. Localização

A Bacia do Acre localiza-se na região norte do Brasil, no extremo oeste brasileiro, entre os paralelos 6°S e 9°S e meridianos 72°30'W e 74°W (Cunha, 2007). É limitada a leste/nordeste pelo Arco de Iquitos (limite com a Bacia do Solimões), a norte pela Bacia de Maraón (Peru), a oeste/sudoeste pela Bacia de Ucayali (Peru) e a norte pela Bacia Madre de Dios (Bolívia e Peru) (Alevato, 2019).

A delimitação em área da Bacia do Acre evoluiu ao longo dos anos: segundo Miura (1972), Campos et al. (1975) e outros autores, a Bacia do Acre corresponderia a toda cobertura sedimentar Cretáceo-Terciária delimitada pela fronteira Brasil-Peru e o Arco de Iquitos atingindo cerca de 300.000 km²; posteriormente, Arana et al. (1990 a, b e c) a definiu como uma área 40.000 km² onde há a preservação dos sedimentos paleozóicos; Cunha (2007) estabeleceu que a Bacia do Acre compreenderia uma área de quase 150.000 km² com 40.000 km² de porções não aflorantes; e o trabalho mais recente, que será utilizado como base deste projeto, foi definido por Alevato (2019) à partir dos estudos de Vasconcellos et al. (2013) e da definição de bacias sedimentares brasileiras de acordo com a ANP, que compreende a Bacia do Acre como a área de 97.552 km² (projeção UTM, Datum SIRGAS 2000), de acordo com a **Figura 02**.

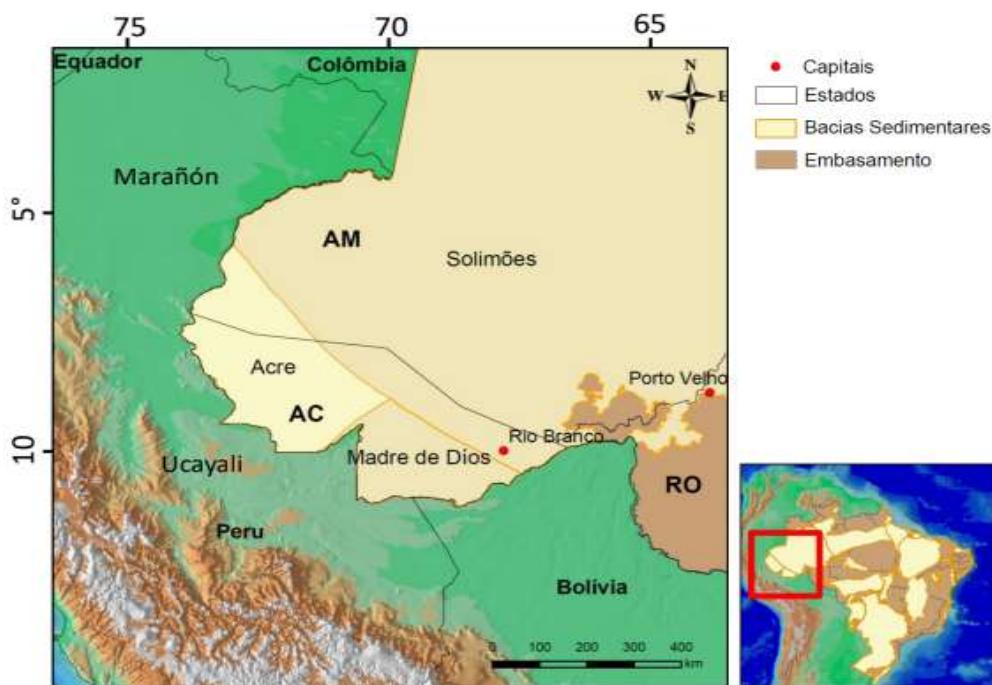


Figura 02. Região definida por Alevato (2019) como a área total da Bacia do Acre realizada em levantamento no QGIS, com destaque para as bacias sedimentares vizinhas. Fonte: Alevato (2019).

4.1.2. Litoestratigrafia e Evolução Geotectônica

A compreensão da evolução tectônica de bacias sedimentares é crucial não apenas para o conhecimento das formações e elementos estruturais que a compõem como também para a modelagem de seu aporte sedimentar e de possíveis sistemas petrolíferos a ela pertinentes, como caracterização de rocha geradora, rocha reservatório, selantes e trapas, e ainda a acumulação e migração primária e secundária de hidrocarbonetos (Alevato, 2019).

Deste modo, conhecer a geologia da Bacia do Acre é elemento fundamental deste projeto (Figura 03). A geologia da Bacia do Acre é complexa, pois envolve uma bacia de foreland multicíclica situada entre as bacias subandinas vizinhas e o cráton amazônico. Segundo a classificação proposta por Bally & Snelson *apud* Raja Gabaglia e Figueiredo (1990), a Bacia do Acre é uma perissutura ou uma bacia foredeep sobre litosfera rígida, diretamente relacionada com a megassutura compressional do cinturão andino (Cunha, 2007), formando o que Klemme (1980) propôs como uma plataforma de exogeossinclíneo, resultado da transformação de bacia marginal aberta de tipo V para tipo andino, e posteriormente passando a bacia do tipo II-B ou Interior Composta Móvel (Cunha, 2007).

Com relação aos elementos estruturais e geotectônicos, a Bacia do Acre, situa-se entre dois trends estruturais, NNW-SSE das bacias mesozóicas e cenozóicas

subbandinas vizinhas, e o trend E-W da sinéclise amazônica paleozóica, que compõe a Bacia do Solimões e a Bacia do Amazonas (Oliveira et al. 1994), apresentando elementos estruturais das duas unidades mencionadas ao longo do tempo geológico (Alevato, 2019).

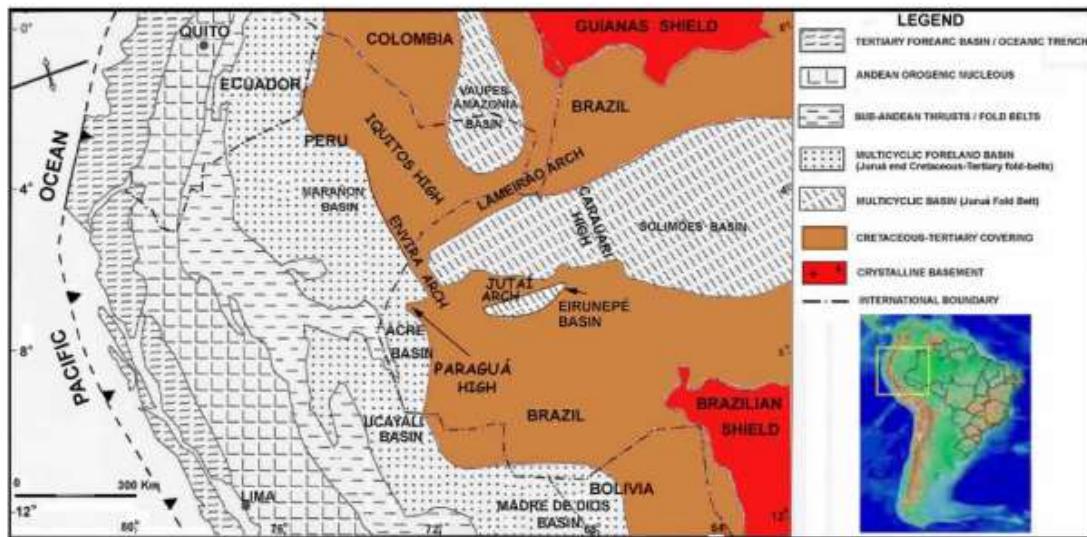


Figura 03. Mapa geológico da região NW da América do Sul. Fonte: Alevato (2019).

A formação da região Amazônica inicia-se ainda no Proterozóico, com a fusão de terrenos cratônicos de granitos e greenstones belts submetidos em seguida a metamorfismo de alto grau (Wanderley-Filho et al., 2010), e após o assentamento da litosfera, inicia-se a deposições sedimentar das bacias paleozóicas da região, que incluem a Bacia do Acre, a Bacia do Solimões e a Bacia do Amazonas (Alevato, 2019).

A evolução tectônica da Bacia do Acre pode ser dividida em 11 sequências distintas de acordo com Cunha (2007):

- I) Sequência Siluriano - Devoniano Inferior (?): representada por folhelhos, siltitos e arenitos de plataforma marinha rasa, cujos sedimentos terrígenos provém, segundo Dalmayrac (1978), do Grupo Contaya da Bacia do Maraño devido a um soerguimento da região central do Peru.
- II) Sequência Devoniano Inferior - Carbonífero Inferior a Médio (?): com base em novos dados sísmicos e inferências a partir de correlações aos arenitos conglomeráticos sobrepostos por uma sequência espessa de folhelhos nas bacias vizinhas (bacias de Ucayali no Peru e Solimões no Brasil), principais potenciais de rochas geradoras nessas bacias, pode

ser constituída por folhelhos marinhos de potencial moderado para hidrocarbonetos.

- III) Sequência Carbonífero Inferior a Médio - Permiano Médio a Superior: período em que se forma a bacia intracratônica/marginal com subsidência flexural para oeste, dividida em dois estágios, o primeiro sob tectônica distensiva com sedimentação continental após a grande regressão arenosa com base arenosa conglomerática (Formação Apuí, Mississipiano terminal), e o segundo com sedimentação carbonática (Formação Cruzeiro do Sul, final do Pensilvaniano) em deposição transgressiva, com fim no Eopermiano com a deposição de rochas calcáreas, concordantes com a deposição da Formação Cruzeiro do Sul acima.

A formação Apuí (Feijó e Souza, 1994), é formada por conglomerados castanhos, de granulometria grossa (seixos e grânulos de quartzo, granito, feldspato) com matriz areno-argilosa, depositada em ambiente de leques aluviais, do Neocarbonífero. Apresenta contato discordante com o embasamento proterozóico - ou coberturas sedimentares inferidas para a sequência do Devoniano Inferior (II), em relação concordante com a Formação Cruzeiro do Sul e máxima superfície de inundação, com presença também de arenitos marinhos e continentais, e folhelhos permianos, maiores potenciais de rochas geradoras de hidrocarbonetos até o momento na Bacia do Acre. Diretamente correlacionada às formações Juruá na Bacia do Solimões e Monte Alegre na Bacia do Amazonas.

A Formação Cruzeiro do Sul é formada por espessas camadas de calcarenitos bioclásticos de conteúdo fossilífero marinho (palinomorfos de ambiente deposicional marinho raso, Feijó e Souza (1994)), com intercalações de anidritas, folhelhos calcíferos e calcilutitos, em relação concordante com a Formação Apuí e com a Formação Rio do Moura acima.

- IV) Sequência Permiano Médio a Superior - Triássico Inferior: composta na base por arenitos marinhos transgressivos com boa qualidade para rochas reservatórios em concordância aos arenitos da sequência anterior (III), seguidos por alternância de siltitos, arenitos, folhelhos

potenciais para rochas geradoras, e carbonatos em continuação à fase transgressiva anterior. No horizonte peruano, caracteriza-se pelos depósitos de *red beds* associados a evaporitos do Grupo Mitu nas bacias de Ucayali e Marañon, com intercalações de riolitos e andesitos em fase de distensão tectônica

Formação Rio do Moura: composta por rochas clástico carbonáticas, com camadas espessas de arenitos intercalados a siltitos e folhelhos na porção basal e carbonatos no topo em ambiente marinho raso, seguidos por arenitos avermelhados e evaporitos em ambiente de maior continentalização no Permo-Carbonífero com participação dos efeitos de soerguimento da Orogenia Herciniana, marcando o fim do Sistema Pré Andino e início do Sistema Andino, relacionado com a subducção da Placa de Nazca e a quebra do Supercontinente Pangea (Barros e Carneiro, 1991). Apresenta discordância com a Formação Juruá Mirim (Eotriássico).

- V) Sequência Triássico Inferior - Jurássico Inferior: representada pela metade inferior da formação Juruá-Mirim, composta por arenitos na base seguidos por siltitos vermelhos com intercalações de halita, anidrita, arenitos betuminosos, carbonatos (calcários e dolomitos escuros), folhelhos, e rochas ígneas ácidas extrusivas similares às rochas intrusivas básicas da Bacia do Solimões durante o magmatismo Penatecaua, como traquitos nefelínicos de idade 194 ± 12 Ma (K-Ar). Posteriormente ocorreu deposição de Evaporitos (Eojurássico). Durante essa sequência ocorre a maior fase de subsidência flexural para oeste da bacia intracratônica seguido por dobras de baixa amplitude e longo período em evento compressivo, gerando halocinese (ascenção de rochas evaporíticas por ação de forças gravitacionais (Castro & Holz, 2005)).
- Formação Juruá Mirim: composta por siltitos avermelhados, com intercalações de anidritas e halitas sobrepostos a arenitos espessos eotriássicos, depósitos em ambiente fluvial, lacustrino, sabkha e marinho restrito, e eólico no topo.
- VI) Sequência Jurássico Inferior - Jurássico Superior: porção superior da Formação Juruá Mirim, composta por arenitos avermelhados/rosáceos em ambiente eólico, com intercalações de folhelhos vermelhos, e

basaltos de idade 177 ± 8 Ma (método K- Ar em amostra de poço). O Diastrofismo Juruá (Campos e Teixeira, 1988) provocou forte compressão da bacia, gerando inversões e falhamentos, halocinese e reativação de falhas antigas. Esse ciclo é encerrado por erosão e peneplanização dos sedimentos superiores.

- VII) Sequência Aptiano - Albiano Superior/Cenomaniano: fase de menor atividade tectônica, com deposição de sedimentos clásticos originados do Escudo Brasileiro e do Escudos das Guianas. Corresponde à porção basal do Grupo Jaquirana, porção inferior da Formação Moa, com arenitos finos a médios flúvio-deltaicos e neríticos depositados em bacia de foreland do cinturão andino.
 - VIII) Sequência Albiano Superior/Cenomaniano - Turoniano Superior/Coniaciano: porção superior da Formação Moa, composta por arenitos finos a médios e folhelhos relacionados a superfície de inundação máxima desde o Triássico, marcando a transição do Cenomaniano/Turoniano.
 - IX) Sequência Turoniano Superior/Coniaciano - Campaniano Inferior: composta por pelitos marinhos (folhelhos cinzentos e castanhos) da Formação Rio Azul do Neosantoniano, com arenitos finos intercalados.
 - X) Sequência Campaniano Inferior - Eoceno Inferior: corresponde aos arenitos fluviais da base da Formação Divisor, em contato abrupto aos folhelhos marinhos e transicionais da Formação Rio Azul e aos folhelhos e carbonatos da Formação Ramon com superfície máxima de inundação do Paleoceno.
- Grupo Jaquirana: de idade Neocretácea dada pelo conteúdo palinomórfico (apesar disso Barros et al. (1974) datam idade terciária à Formação Ramon). Apresenta relação discordante à Formação Solimões.
- XI) Sequência Eoceno Inferior - Plioceno: formado por deposição de folhelhos de ambiente marinho raso e lacustrino, intercalado a níveis

arenosos e red-beds dentro Sistema Pebas assentados em onlap contra a borda leste devido ao tectonismo terciário dos Andes.

A **Figura 04** demonstra as principais unidades aflorantes da Bacia do Acre no perfil A-B traçado:

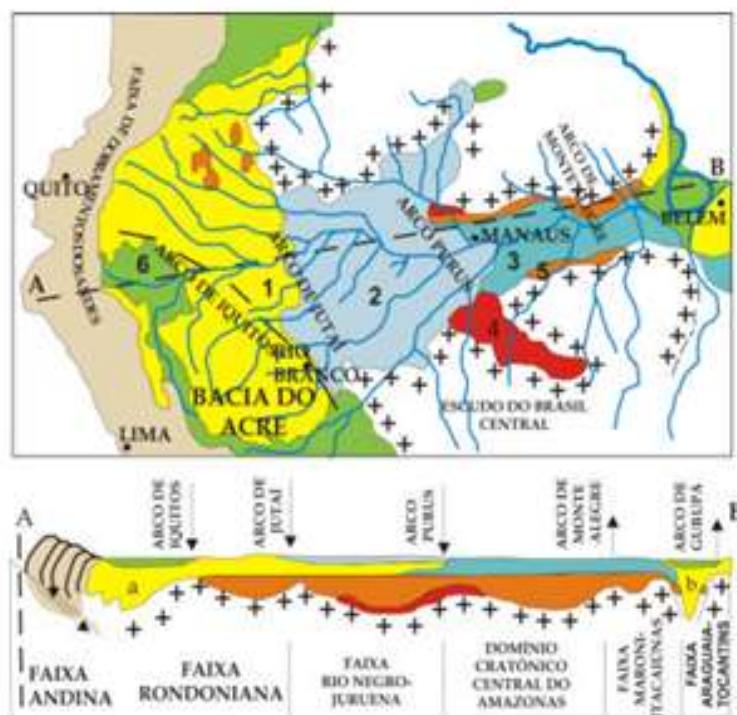


Figura 04. Mapa de localização da Bacia do Acre e principais unidades no perfil A-B, sendo: 1) A – Formação Solimões; B – sedimentos terciários; 2) Formação Içá; 3) Formação Alter do Chão; 4) coberturas proterozoicas; 5) rochas Paleozóicas; 6) coberturas do quaternário. Fonte: Cavalcante (2006).

A Bacia do Acre é única em território brasileiro que é influenciada pelo processo de soerguimento andino (Feijó e Souza, 1994; Cunha, 2007), recebendo um volume expressivo de sedimentos oriundos dos Andes. Desta forma, o conhecimento de seu preenchimento e produção sedimentar - com fluxo de sedimentos da ordem 10^8 a 10^9 ton/ano (Hoorn et al., 2010) - ao longo do tempo geológico significa uma interpretação mais realista do registro sedimentar da Margem Equatorial Brasileira (Hoorn et al., 2017).

A carta cronoestratigráfica proposta pela Petrobrás para a Bacia do Acre pode ser visualizada a seguir na **Figura 05**:

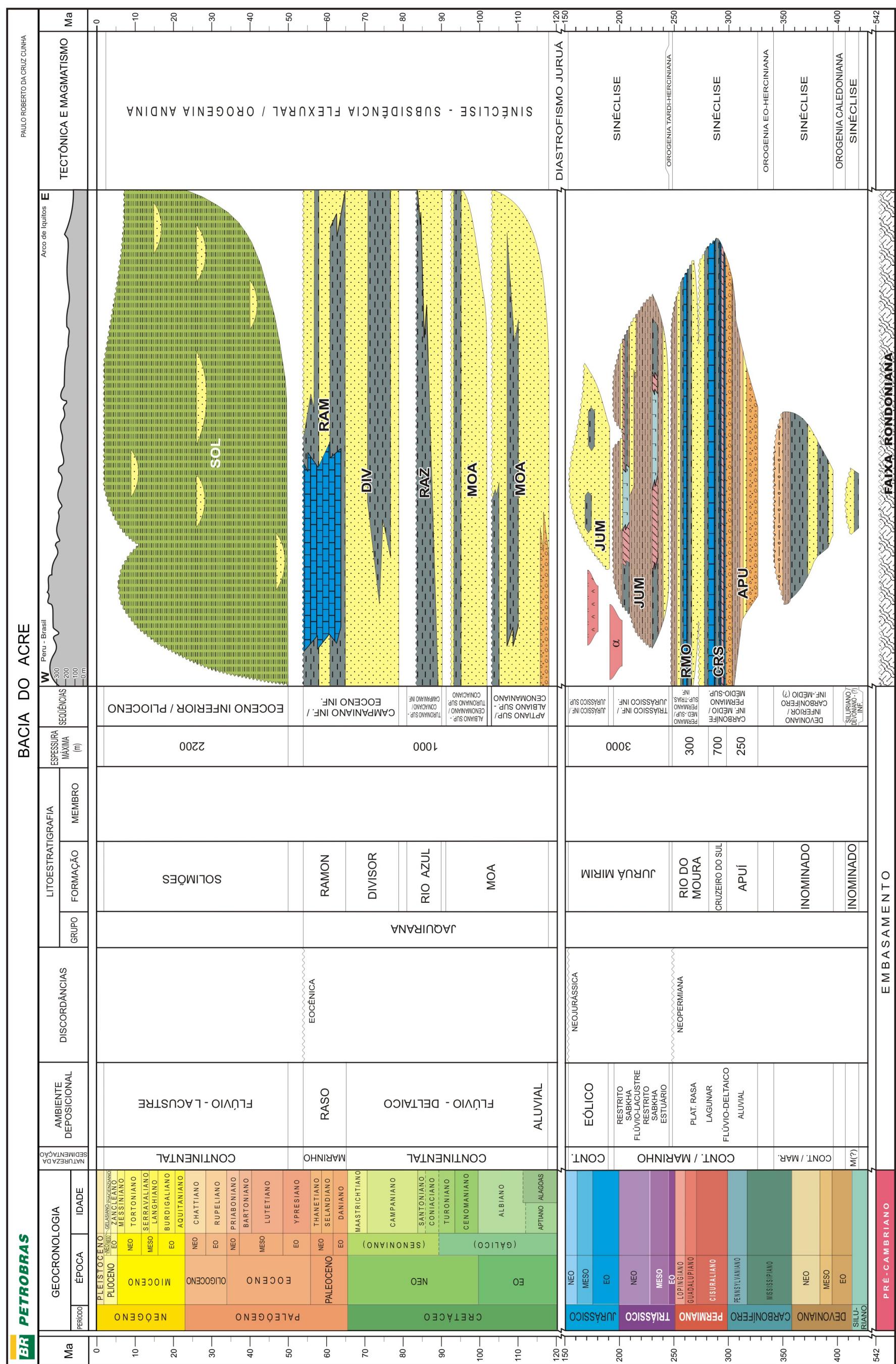


Figura 05. Carta cronoestratigráfica da Bacia do Acre (Milani, 2007).

A evolução geotectônica da Bacia do Acre descrita por Oliveira (1994), Oliveira et al. (1997) e Wanderley-Filho et al. (2010), por meio de análise das seções sísmicas produzidas ao longo dos anos, demonstra algumas características principais estruturais de acordo com a **Figura 06**.



Figura 06. Mapa estrutural para a Bacia do Acre contendo as principais estruturas indicadas na literatura. Fonte: Wanderley-Filho et al., 2010.

Destaca-se a falha reversa de trend norte-sul de Batã, localizada ao leste; o Sistema de Falhas Reversas de Oeste de Batã, associados à Serra do Divisor; o Arco de Paraguá-Jutaí, que separa dois depocentros: a depressão Jaquirana ao norte e a depressão Central; e a plataforma Cruzeiro do Sul a leste (Wanderley-Filho et al., 2010).

Os campos tensionais de alternância entre caráter extensional e caráter compressional, associados a deslocamentos strike-slip, halocinese e intrusões ígneas afetaram a evolução estrutural na Bacia.

Notadamente há dois eventos tectônicos principais (**Figura 07**). Um extensional com falhas normais de alto ângulo que afetam o embasamento, com

formação de grábens e hemi-grábens onde as primeiras unidades sedimentares se depositaram (Oliveira et al., 1997). Na segunda fase extensional há a um conjunto de falhas normais de aspecto variado cortando a seção pré-Cretácica e o embasamento (Oliveira et al., 1997), com magmatismo básico (basalto) e formação de dobras extensionais forçadas.

O segundo evento, compressivo, se relaciona a três fases distintas, sendo a primeira restrita à porção norte da bacia, com a formação de falhas reversas desde o embasamento que mergulham para oeste e afetam apenas as primeiras unidades basais. A segunda fase, de maior distribuição na bacia, afeta toda a seção pré-Cretácica e é limitada no topo pela discordância do Diastrofismo Juruá, compreendendo extensas falhas reversas de alto ângulo que mergulham em direção a Oeste, além de dobras que vão desde o embasamento até a porção pré-Cretácica. A terceira fase engloba toda a seção do Cretáceo e partes do Terciário, responsável pelo soerguimento da Serra do Divisor e pela geometria atual da falha de Batã (Oliveira et al., 1997).

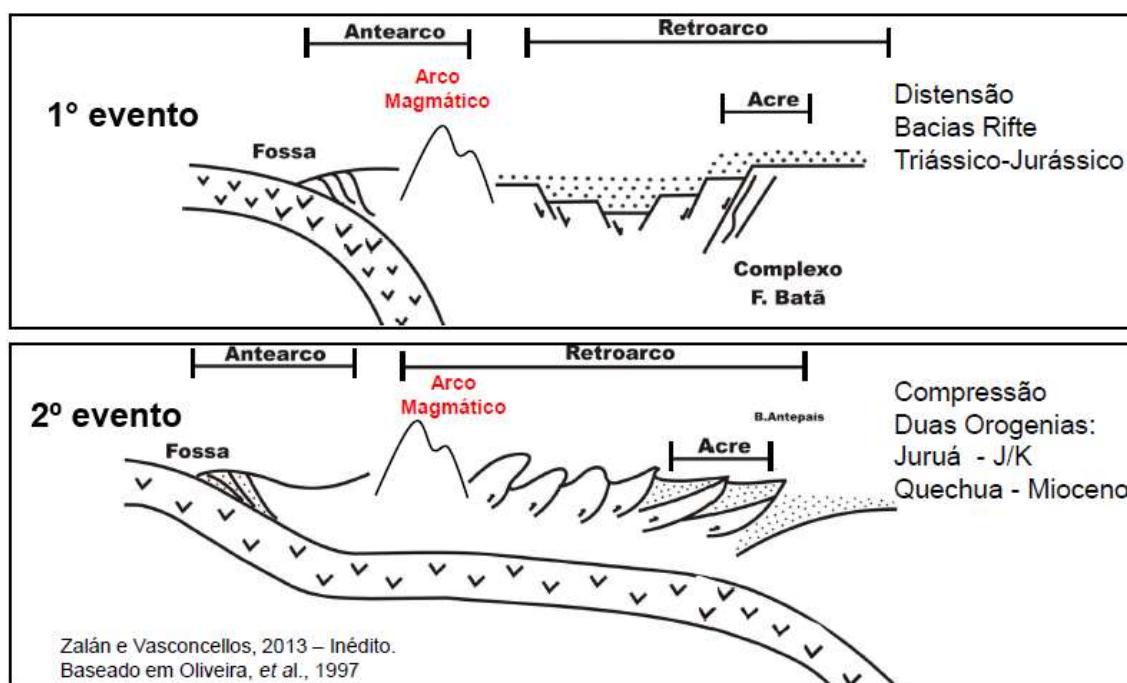


Figura 07. Esquematização dos dois eventos, o primeiro distensivo e o segundo compressional.
Fonte: Oliveira et al., 1997

Outras estruturas também presentes são empurões do tipo back-thrusts, estruturas em triângulo, blind thrusts e pop-ups e sequência imbricada de falhas reversas (Oliveira et al., 1997).

Ainda, observa-se a presença de dois eventos magmáticos discerníveis nas seções sísmicas: um de natureza intrusiva alcalina com área de aproximadamente 100 km², que corta a seção sedimentar pré-Cretácica, e um segundo evento relacionado a lacólitos basálticos.

4.1.3. Histórico Exploratório

O histórico exploratório da Bacia do Acre teve início em 1935, e divide-se em cinco fases de acordo com Vasconcellos & Coutinho (2013) de acordo com a **Figura 08**.

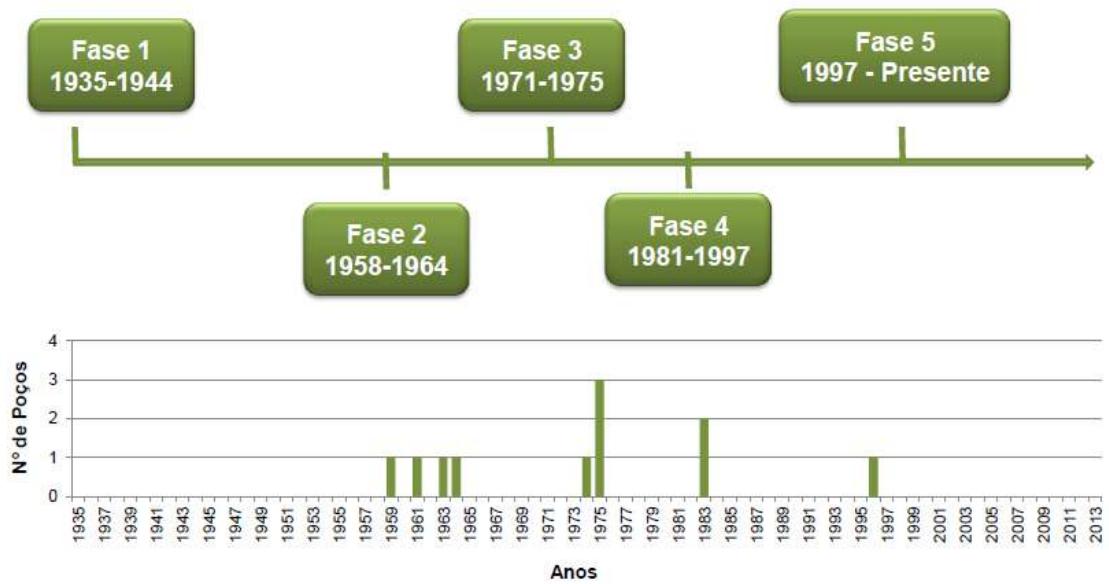


Figura 08. histórico exploratório da Bacia do Acre. Fonte: 12^a Rodada de Licitações (ANP, 2013)

A primeira fase, entre 1935 e 1944, inicia-se com trabalhos realizados pelo Departamento Nacional de Produção Mineral (DNPM) e Conselho Nacional de Petróleo (CNP) para reconhecimento geológico da região, com levantamento estratigráfico-estrutural para a avaliação do potencial petrolífero.

A segunda fase, entre 1958 e 1964, após a criação da Petrobrás em 1953, engloba a descoberta de acumulações no Peru e a perfuração de quatro poços exploratórios, além de levantamento gravimétrico e de fotografias aéreas.

A terceira fase, de 1971 a 1975, compreendeu a perfuração de quatro poços pioneiros pela Petrobras, com objetivo de atingir a seção Cretácica, e a contribuição de vários autores devido a novas descobertas no Peru e na Colômbia para correlação com as bacias Peruanas.

A quarta fase, de 1981 a 1997, focou a investigação da seção pré-Cretácica, com a perfuração de três novos poços e a descoberta do campo gás de Camisea, na

bacia de Ucayali (Peru), com cerca de 426.1 Bm³ de reserva provada segundo o World Energy Council (2014).

A última fase, de 1997 aos dias atuais, após a criação da Agência Nacional de Petróleo (ANP) em 1998, compreende a concessão de 11 blocos exploratórios (**Figura 09**), aquisição de dados sísmicos, gravimétricos e magnetométricos 2D.

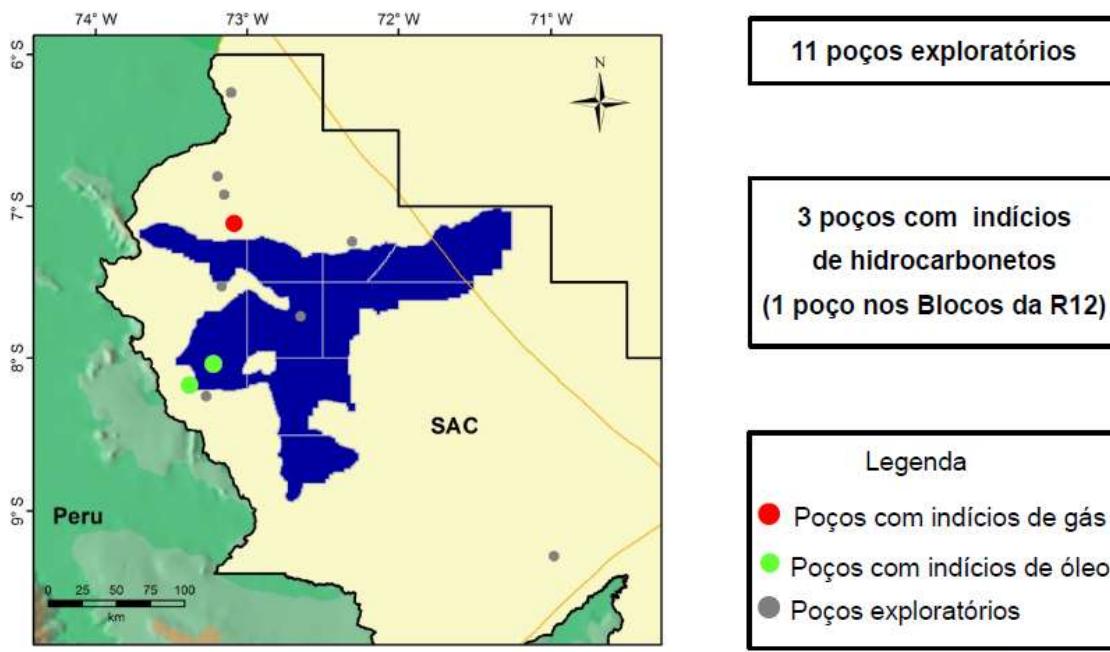


Figura 09. localização dos blocos exploratórios na 12^a Rodada de Licitações de Blocos Exploratórios. Fonte: ANP, 2013.

4.1.4. Geologia do Petróleo

Os possíveis sistemas petrolíferos estimados para a Bacia do Acre foram propostos em alguns trabalhos como Cunha (2007) e Vasconcellos & Coutinho (2013):

- a) Rocha Geradora: rocha depositada em ambiente marinho ou lacustre anóxico onde os hidrocarbonetos são gerados, de granulação fina como folhelhos, calcilutitos e margas. A concentração de matéria orgânica da rocha geradora (Carbono Orgânico Total - COT) depende da produtividade sedimentar, do soterramento e aumento de temperatura, e da maturação do material orgânico até a evolução para hidrocarbonetos por degradação térmica e sua migração para rochas reservatório.

Na Bacia do Acre os autores propõem que os folhelhos e margas da Formação Cruzeiro do Sul podem possuir COT de até 6% e até 225 miligramas de hidrocarboneto por grama de carbono orgânico total (mgHC/gCOT), que indica um potencial de geração para gás e

condensado. Os folhelhos da sequência Devoniano Inferior a Carbonífero Inferior a Médio e os folhelhos da Formação Rio do Moura também são cogitados como possíveis rochas geradoras.

- b) Rocha Reservatório: são rochas com boa qualidade de porosidade e permeabilidade capazes de armazenar hidrocarbonetos e outros fluídos, como arenitos e carbonatos (por porosidade secundária). Para os autores, os arenitos da Formação Juruá-Mirim e os arenitos e carbonatos do Permo-Carbonífero podem apresentar acumulações.
- c) Rocha Selante: são rochas pouco ou nada permeáveis e plásticas (a fim de manter a sua impermeabilidade quando submetida a esforços tectônicos) que impedem a migração dos hidrocarbonetos presentes nas rochas reservatórios para outras localidades. Em geral, evaporitos e folhelhos pouco permeáveis são considerados os principais selos de sistemas petrolíferos. Para os autores, os folhelhos, margas e calcilutitos da Formação Cruzeiro do Sul, as anidritas e halitas da Formação Juruá-Mirim e os folhelhos da Formação Moa comporiam as rochas selantes de possíveis sistemas petrolíferos na Bacia.
- d) Trapas/Armadilhas: estruturas geológicas que tornam possível a acumulação de hidrocarbonetos e impedem a migração deles na rocha reservatório para outras rochas porosas ou para a superfície, como sistemas arranjados de rochas reservatórios e rochas selantes. A mais tradicional observada em sistemas petrolíferos são dobras anticlinais e sistemas de falhas. Segundo os autores, as falhas estruturais a noroeste da falha reversa de Batã podem constituir armadilhas estruturais do sistema.
- e) Plays Exploratórios: os autores especulam para a Bacia do Acre os seguintes sistemas petrolíferos: Inominado (Devoniano) - Cruzeiro do Sul (Permo-Carbonífero) (?); Cruzeiro do Sul (Permo-Carbonífero) - Cruzeiro do Sul (Permo-Carbonífero) (?); Inominado (Devoniano) - Juruá Mirim (Juro-Triássico) (?); Cruzeiro do Sul (Permo-Carbonífero) - Juruá Mirim (Juro Triássico) (?)

Devido ao fato de a Bacia do Acre apresentar sedimentação paleozóica semelhante a das bacias de Marañon e de Ucayali, de acordo com Alves (1989),

existe a possibilidade de haver acumulações na Bacia do Acre. Na Bacia de Ucayali há um imenso campo de gás e condensado denominado Camisea, com volume estimado de 16,6 trilhões de pés cúbicos de gás seco e 970 milhões de barris de condensado, de acordo com dados da Perupetro (2013) e Vasconcellos & Coutinho (2013), e volume comprovado de 426.1 Bm³ de gás em Camisea, na Bacia de Ucayali no Peru (Alevato, 2019). Além disso, descobertas recentes da Petrobrás no bloco SOL-T-171 no Município de Coari no Amazonas descrevem uma nova acumulação de petróleo e gás na Bacia do Solimões por meio do poço 1-BRSA-961-AM (site da ANP), alimentam a pesquisa por hidrocarbonetos na região buscando nos sistemas petrolíferos propostos semelhanças com a Bacia de Solimões vizinha.

4.2. Bacia do Solimões: Contexto Geológico

4.2.1. Localização

A Bacia do Solimões, localizada entre os paralelos de 2° e 8°S e meridianos 62° e 72°, é uma bacia sedimentar paleozóica intracratônica limitada a norte pelo Escudo das Guianas, a Leste pelo Arco de Purus, a oeste pelo Arco de Iquitos, e a sul pelo Escudo Brasileiro, com área aproximada de 950.000 km². Importante região produtora de óleo, condensado e gás do país, estima-se que cerca de 480.000 km² sejam prospectáveis para hidrocarbonetos (Barata, 2007).

4.2.2. Litoestratigrafia e Evolução Geotectônica

Cordani et al. (1984) consideraram os escudos da Guiana e do Brasil divididos em quatro províncias geotectônicas, mostrando continuidade abaixo das bacias do Solimões e do Amazonas. A Bacia do Solimões se sobrepõe às províncias do Rio Negro Juruena e Cinturão Rondoniense, o que lhe confere três feições morfológicas principais que controlaram a sedimentação paleozóica (**Figura 10**):

- a) Sub-bacia do Juruá – sobrepõe-se à província Rio Negro-Juruena e possui eixo sedimentar orientado NE-SW. Alto de Carauari, que a separa da sub-bacia de Jandiatuba;
- b) Sub-bacia de Jandiatuba – sobrepõe-se à zona Rondoniense, segue no sentido leste-oeste e é limitado a oeste pelo Arco de Iquitos;
- c) Alto de Carauari – alto dentro da bacia hidrográfica, voltada NO-SE, com 120 km de largura. Localiza-se na zona de sutura entre as províncias do Rio Negro-Juruena e a zona Rondoniense. Esta estrutura tem forte controle sobre a

sedimentação do Carbonífero Ordoviciano, Devoniano e Permiano. Divide a Bacia do Solimões nas duas sub-bacias mencionadas acima (**Figura 10**). Ao longo da deposição paleozoica da Bacia do Solimões, as intrusões marinhas vieram da bacia ocidental e oriental do Peru, com sobreposição das costas leste, norte e sul.

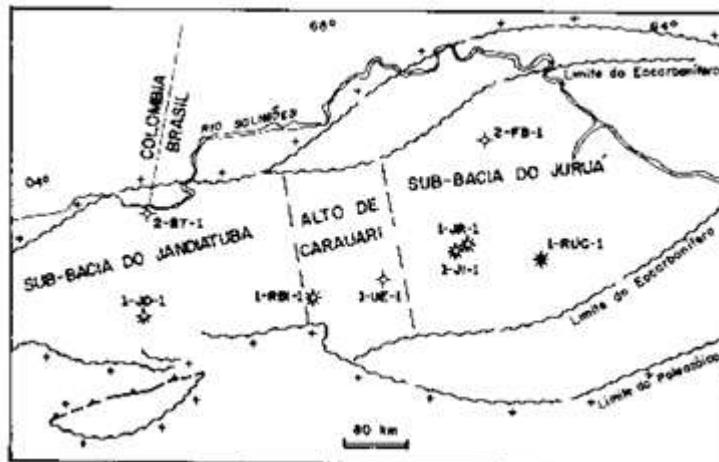


Figura 10. feições morfológicas que controlam a sedimentação na Bacia do Solimões. Extraído de Caputo (1991).

A deposição da Bacia do Solimões pode ser dividida em sete sequências sedimentares, sendo as quatro primeiras restritas ao Paleozoico, onde são observadas condições oceânicas registradas na história geológica da bacia (Silva, 1987). A distribuição espaço-temporal das unidades litoestratigráficas que compõem a sequência sedimentar se dá da seguinte forma:

Sequência 1 - Ordoviciano Inferior (Arreginiano/Ranvirniano) consiste em sedimentos marinhos siliciclosos da Formação Benjamin Malar, confinados à Sub-bacia de Jandiatuba e temporalmente correlacionados com sedimentos da Formação Kontaya da Bacia Ukajali (Quadros, 1988). Os estágios de erosão que ocorreram após a deposição da Sequência 1 abrangem um intervalo de tempo de aproximadamente 100 milhões de anos.

Sequência 2 - O primeiro registro desta sequência do Emsiano, com sedimentos marinhos costeiros da Formação Jandiatuba e praias basais da Formação Biá, que ainda estão restritos à Sub-bacia de Jandiatuba. A elevação relativa do nível do mar ocorreu com pulsos retrógrados-regressivos até o Frasniano, quando se iniciou a deposição nas sub-bacias Alto de Carauari e Jurua dos depósitos siliciclásticos da Formação Uerê. A explosão da fauna única de espongiários na Formação Uere (Silva, 1987) sugere um clima periglacial nesta bacia, o que reforça a paleontologia de alta latitude do supercontinente Gondwana ocidental, como sugerido por Caputo (1984). A próxima fase glacial é representada pelos tilitos da Formação Jaraqui Mesofameniana

(Caputo, 1984), depositados em condições glaciais na parte sul da sub-bacia do Juruá e no flanco ocidental do Alto de Carauari. A regressão no final do ciclo Devoniano foi caracterizada por progradação fluviódéltica com retrabalhamento eólico, sendo que a parte inferior da Formação Jurua cruzou com os folhelhos regressivos da Formação Jandiatuba em direção à Sub-bacia de Jandiatuba (Silva, 1987a). Pode haver inconformidades não descobertas nesta parte que a separam da ordem superior.

Sequência 3 - A Bacia do Solimões é caracterizada pela continentalização parcial em decorrência do processo de desertificação durante a deposição do alto Formação Jurua. Os desertos carboníferos expandiram a sobreposição nas margens, criando inconformidades na bacia, incluindo o Arco Purus. A retomada da sedimentação marinha na bacia ocorreu de forma pulsante, o que foi responsável pela fase espessa do ciclo carbonato-evaporito da Formação Carauari. Os primeiros 200 metros desta unidade litoestratigráfica estão confinados à Bacia do Solimões (Caputo, 2001) e depositados no Namuriano^(?), com sucessivas deposições nas bacias do Solimões e do Amazonas. O caráter destas unidades é geralmente progradacional, com extensas rochas sedimentares depositadas por deposição transgressiva em áreas da bacia até então não alcançadas. Ao final desta sequência, uma fase regressiva é marcada pelo aumento dos clastos pelíticos de coloração vermelha (Silva, 1987a - ciclos V e V1).

Sequência 4 - Deposição dos *red beds* da Formação Fonte Boa marca a completa continentalização da bacia. Esses depósitos foram em grande parte erodidos e permanecem apenas nas partes centrais das bacias do Juruá e Jandiatuba. Esta unidade pode estar associada à Formação Andira da Bacia Amazônica e está datada da idade Permiana^(?).

Sequência 5 - preservada apenas na parte sul da sub-bacia de Jandiatuba em sedimentos mais antigos que o Cretáceo. É constituído por sedimentos que podem estar relacionados à Formação Juruá Mirim e são datados do vulcanismo basal da Bacia do Acre (Neotriássico).

Sequências 6 e 7 - Sedimentos Cretáceo e Terciário da Formação Alter do Chão e Solimões, respectivamente. No Juro-Triássico, a bacia experimentou intenso magmatismo basal com intrusões de soleira no Paleozóico.

O magmatismo da região, com intrusões de diques e soleiras de diabásio, contribuíram para a evolução térmica da matéria orgânica da rocha geradora do sistema petrolífero da região, com destaque para os sistemas Jandiatuba-Juruá(!) e Jandiatuba-Uerê (.)¹, que apresentam pico máximo de carbono orgânico total (COT) de 8,25% (Barata e Caputo, 2007) (**Figura 11**).

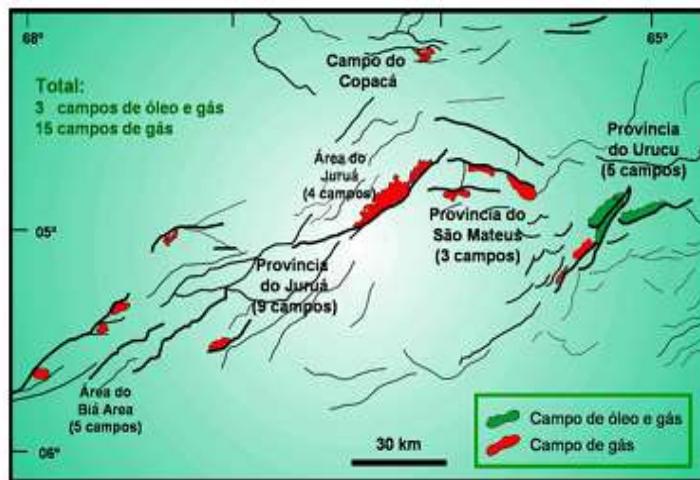


Figura 11. Distribuição dos principais campos de produção de hidrocarbonetos da Bacia do Solimões. Fonte: Eiras (1999).

Segundo Carneiro e Schneider (1976), a estrutura recentemente mapeada da bacia pode ser resultado de esforços compressivos associados à orogenia andina (Cretáceo-Terciário). Szatmari et al. (1975) explicou a origem da estrutura do Juruá como uma orogenia hercínica ativa durante o período Permo-Carbonífero, o que pode ter resultado em forças de compressão na Bacia do Solimões devido à subducção da placa do Pacífico. A acumulação de gás no Juruá pode estar na intersecção de duas tendências. Uma é a direção leste-oeste reativada (direção axial da bacia Amazônica) e a outra é NNE-SSO (provavelmente direção do vale da bacia do Solimões).

Segundo Esteves (1984), durante o período Jurássico a litosfera se desintegraria ao longo da junção do continente sul-americano com a bacia do Pacífico, formando a Placa de Nazca, e teria início o processo de subducção na borda noroeste da América do Sul. Quando esses limites convergentes aparecem, o fluxo tectônico do Juruá é formado por fenômenos compressivos.

Szatmari (1984) reconstruiu e ampliou seu modelo para explicar a presença de extensas falhas desde o Graben de Tacutu até Pizco (Peru), que ocorreram durante o Triássico e o Jurássico, quando o Golfo do México e o oceano se abriram. Durante a abertura do Golfo do México, o bloco noroeste desta falha se moverá para sudoeste. O movimento do bloco noroeste desta possível falha estaria associado a uma ligeira rotação anti-horária do bloco deslocado, causando extensão do Graben de Tacutu e compressão da área do Juruá. A intensidade desta compressão aumentará de nordeste para sudoeste.

Em 1984, Estevez modificou seu modelo e concluiu que as forças compressivas NE-SE atuando durante os períodos Meso-Jurássico-Neo-Jurássico e Eo-Cretáceo determinaram a formação e individualização das placas Caribenha,

Cocos e Nazca. A dobra do Jurua com tendência NE-SW é provavelmente resultado desse esforço de compressão. A extensão das dorsais meso-oceânicas da América Central e do Pacífico Sul terá um papel importante na geração de forças compressivas atuantes na Bacia do Solimões, que serão responsáveis pela formação do trend do Jurua.

Caputo (1984) apresentou evidências que contradiziam parcial ou totalmente as interpretações anteriores e propôs um modelo em que a estrutura da Bacia do Solimões era formada por forças compressivas atuantes no Neo-Jurássico-Eocretáceo. Posteriormente, Caputo (1985) apresentou uma interpretação alternativa envolvendo deformação, formação de zonas de impacto e falhas de soerguimento, com base em sua análise das diversas estruturas da Bacia do Solimões.

A carta cronoestratigráfica da Bacia do Solimões encontra-se a seguir (**Figura 12**).

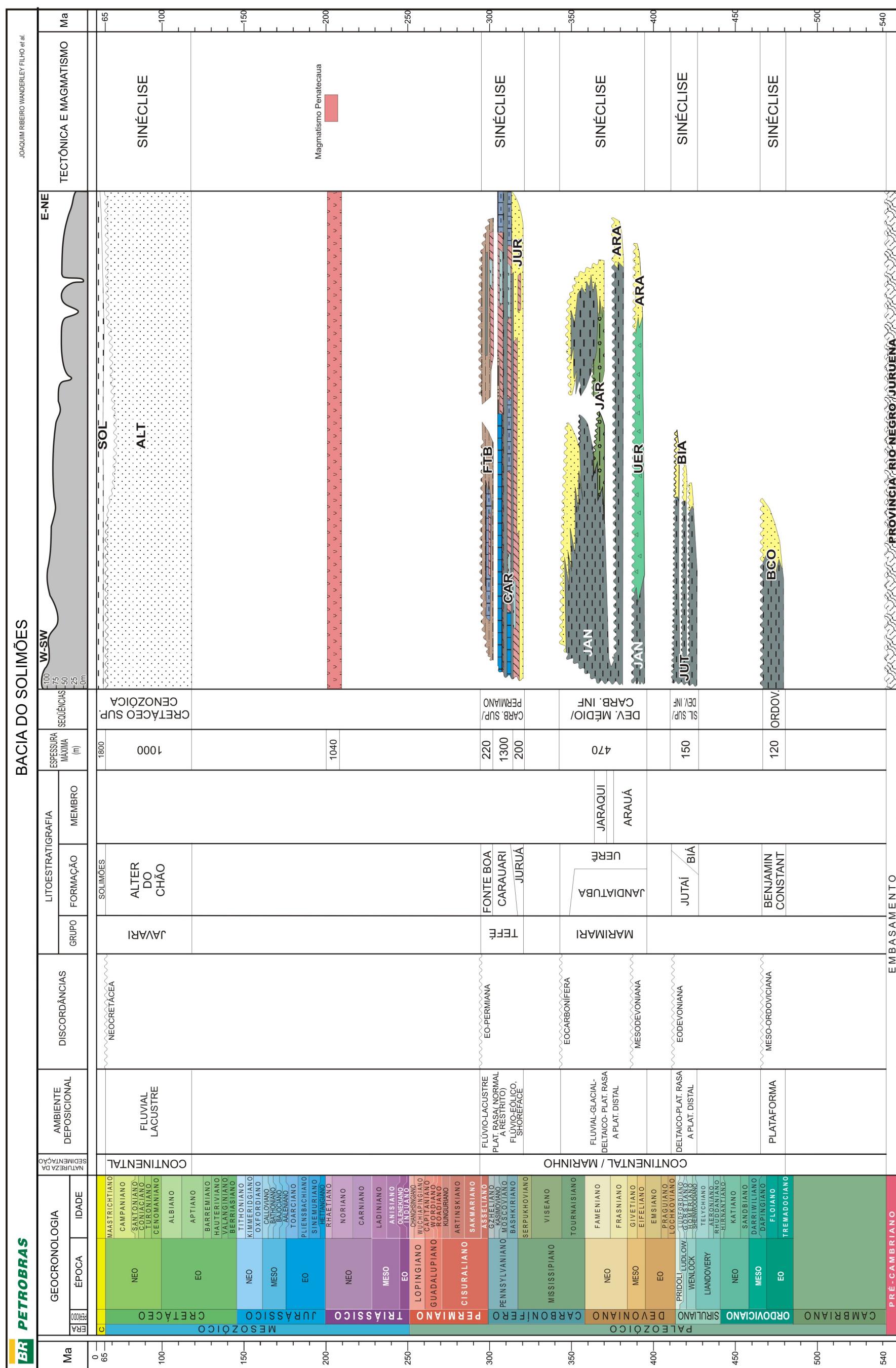


Figura 12. Carta cronoestratigráfica da Bacia do Solimões (Milani, 2007)

4.2.3. Geologia do Petróleo

Quanto à geologia de petróleo na região, a orogenia neojurássica foi responsável por parte do espessamento tectônico das margens continentais, criado por uma série de falhas de empuxo. A sobrecarga de múltiplas fatias de rocha empilhadas na margem continental teria causado forte soerguimento periférico na região do Arco de Iquitos durante o Cretáceo, causando erosão e retirada da cobertura sedimentar da região. Este mecanismo de formação de arco de bacia foi proposto por Walcot (1970).

O arco Assunção, no oeste da bacia do Paraná, pode ter a mesma origem, ou seja, teria resultado de sobrecarga na margem continental. Com a contínua elevação da cordilheira andina durante os períodos Cretáceo e Terciário, a área influenciada pelo grande manto andino se expandiu, causando subsidência parcial do pico original do arco de Iquitos, deslocando o topo da elevação periférica mais para o leste. Na análise macroscópica da Bacia do Solimões, as principais reservas de hidrocarbonetos estão geneticamente ligadas à fase de regressão e continentalização que a bacia experimentou no final da sequência 2 (Formação do Baixo Juruá) e início da sequência 3 (formação do Alto Juruá e base da formação Caruari).

As características de permeabilidade e desenvolvimento diagenético de cada reservatório dependem diretamente do ambiente sedimentar em que se desenvolveram, tais como:

I) Formação Juruá Inferior – arenito deltaico fluvial cimentado por diagênese. O desenvolvimento sintático do quartzo, essencialmente porosidade secundária, é produzido pela dissolução dos grãos de feldspato. Na área do campo do Rio Urucu, este trecho apresenta porosidade não superior a 9 a 11%, com permeabilidade variando de 1 a 350 mD, sem relação entre as duas grandezas.

II) Formação Juruá Superior – ambiente desértico, onde as melhores rochas reservatório estão associadas a dunas, com valores de porosidade de até 22,5% e permeabilidades de 100 a 316 mD (Lanzarini, 1984). A porosidade é de origem secundária, com empacotamento heterogêneo, poros gigantes e alongados, resultado da dissolução do cimento de calcita e anidrita (Lanzarini, 1984).

III) Formação Basal Caruari – os reservatórios mais importantes desta unidade encontram-se na região do Rio Urucu, também depositado em ambiente eólico, porém mais restrito às margens, devido às intrusões marinhas do Carbono tomou conta de grande parte da bacia. Sete reservatórios produtores de petróleo, condensado e gás estão zoneados nos campos petrolíferos do Rio Urucu, na parte basal da Formação Caruari, sendo o maior deles conhecido como R-7 (Cunha et al., 1988). As melhores

condições de permeabilidade estão associadas à fácie duna e banco de areia, embora se verifique uma diminuição dos valores de permeabilidade verticalmente relativamente à horizontal, devido à intensa cimentação de anidrite eodiagénica que esta fácie tem experimentado (Cunha et al., 1988).

O modelo de sistemas petrolíferos da Bacia do Solimões compreende:

- Rochas Geradoras: a) Formação Jandiatuba - geneticamente, o xisto espesso que compõe esta unidade está relacionado a uma fase progradacional em que no topo desta fase foi formado o protólito contendo até 8,25°/70 de matéria orgânica e uma fase regressiva onde o nível médio não atingiu 1.091°; b) Formação Jaraqui – folhelhos offshore em conjunto com Tilita formam esta unidade com teor de matéria orgânica variando de 0,65 a 1,45%. Dados orgânicos mostram que 40 a 70% da matéria orgânica é amorfa e herbácea, adequada para a produção de petróleo (Cunha et al., 1988);; e c) Formação Uerê - O sílex acumulado na parte leste do Alto de Caruari e sub-bacia do Juruá possui teor médio de carbono orgânico de 1,48 a 3,07%, com proporção de carbono orgânico amorfado e/ou de 75 a 95 % de matéria orgânica vegetal.

Embora nenhum campo produtor de hidrocarbonetos tenha sido descoberto até o momento na sub-bacia de Jandiatuba, esta parte da bacia do Solimões contém a rocha geradora mais importante em termos de conteúdo de carbono orgânico (a Formação Jandiatuba). Na sub-bacia do Juruá, os excelentes teores e características da matéria orgânica utilizada na produção de óleo, gás e condensado dos xistos da Formação Uerê permitem que sejam considerados a principal fonte de produção da região.

- Rochas Reservatório: As rochas reservatório são arenitos carbonáceos da Formação Juruá, com até 50 m de espessura, onde ocorrem arenitos eólicos da Formação Juruá. A parte superior apresenta porosidade de até 22,5% e permeabilidade de 100 a 320 mD e arenitos fluviodeltaicos da parte basal, possuindo porosidade de 9 a 11% e permeabilidade de 1 a 350 mD.

- Rochas selantes: As rochas selantes são evaporitos (anidrita e halita) e xistos, também carbonáceos, presentes na própria formação Juruá e na base da formação Caruari.

- Trapas/Armadilhas: As armadilhas são estruturais e compostas por linhas reflexivas formadas internamente de blocos de falhas reversas altas na direção geral NE-SW, desenvolvidos durante o Megacisalhamento do Solimões.

A **Figura 13** demonstra um perfil geológico da Bacia do Solimões com as rochas do sistema petrolífero dos Campos de Juruá e Urucu.

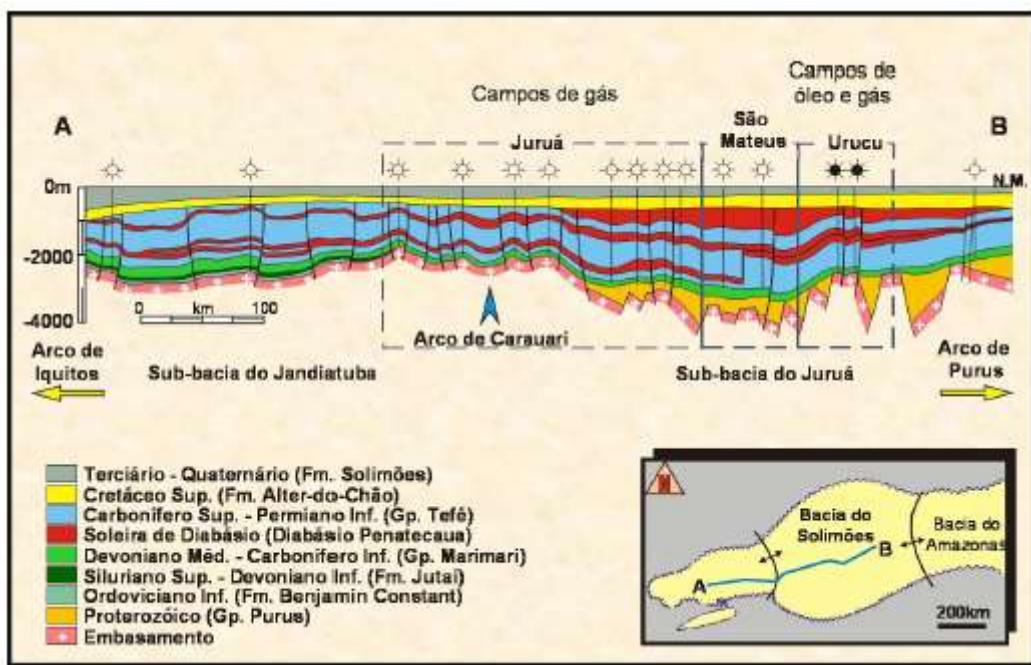


Figura 13. Seção geológica longitudinal da Bacia do Solimões. Extraído de Caputo (2007).

4.3. Formação Solimões

A Formação Solimões registra a história tectonossedimentar do oeste da Amazônia ao longo do Cenozóico. A unidade recobre quase todo o território da Bacia do Acre e a porção oeste da Bacia do Solimões (**Figura 14**), tendo espessuras de até 2.200 m e 1.800 m nessas bacias, respectivamente (Cunha, 2007; Loureiro et al., 2019).

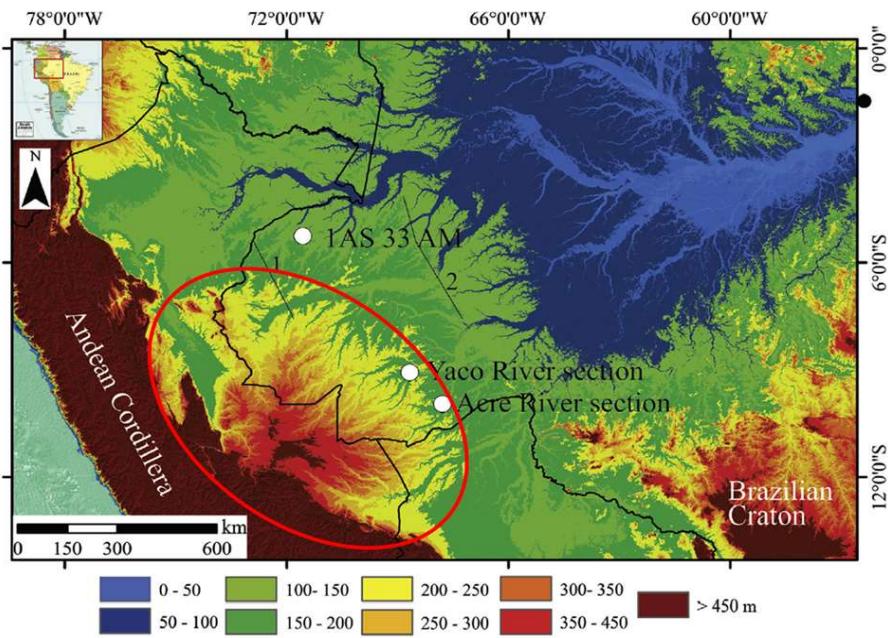


Figura 14. Mapa hipsométrico da região do oeste da Amazônia. Destacado em vermelho encontra-se a região onde ocorre o Arco de Fitzcarrald, notar as maiores elevações que ocorrem devido ao domo. Imagem: Horbe et al., 2019.

Com deposição do Mioceno e Plioceno, a porção basal da Formação Solimões apresenta predominância de folhelhos e lamitos, com o aumento de corpos de areia em direção ao topo. Isso reflete a evolução da paleogeografia ao longo do tempo: do Mioceno Inferior a Médio, a paisagem era composta por grandes áreas inundadas, com sistemas lacustres por vezes com a influência marinha, o que foi denominado como Sistema Pebas (Hoorn et al., 2010; Hoorn et al., 2017). Já do Mioceno Médio a Tardio, ocorreu a progressiva redução da área de terras alagadas, dando lugar à predominância de sistemas fluviais ou fluviais influenciados por maré, fase chamada de Sistema Acre (Hoorn et al., 2010).

Do ponto de vista estratigráfico há uma correlação das litologias típicas para a Formação Solimões nas duas bacias, sendo composta por folhelhos e lamitos, argilitos vermelhos e siltitos, intercalados a arenitos oxidados - depósitos red-beds (Cunha, 2007) - depositados do Mioceno e Plioceno, com aumento de corpos de areia em direção ao topo, caracterizando o paleoambiente que a formou: do Mioceno Inferior a Médio, a paisagem era composta por grandes áreas inundadas, com sistemas lacustres por vezes com a influência marinha, o que foi denominado como Sistema Pebas (Hoorn et al., 2010; Hoorn et al., 2017). Já do Mioceno Médio a Tardio, ocorreu a progressiva redução da área de terras alagadas, dando lugar à predominância de sistemas fluviais ou fluviais influenciados por maré, fase chamada de Sistema Acre (Hoorn et al., 2010).

Apesar da ocorrência da Formação Solimões tanto na Bacia do Acre como do Solimões, tais bacias estiveram sob contextos tectônicos diferentes durante o Cenozóico. A Bacia do Acre é uma bacia de foreland, sendo a única em território brasileiro afetada diretamente pela tectônica andina. Sua subsidência foi resultado pela flexura regional da litosfera devido ao seu espessamento na região dos Andes (Hoorn et al., 2010; Wanderley-Filho et al., 2010).

Por outro lado, a Bacia do Solimões encontra-se em contexto intracratônico. Estudos recentes apontam que a subsidência dessa bacia foi intensificada por processos mantélicos devido a perturbações causadas pela subducção da Placa de Nazca sob a Placa da América do Sul (Shephard, et al., 2010; Flament et al., 2015; Tacio et al., 2000). Os efeitos causados na superfície por esse fenômeno são denominados como topografia dinâmica.

Além disso, parte da Bacia do Acre foi soerguida no Plioceno, originando a estrutura dômica denominada Arco de Fitzcarrald (Espurt et al., 2010). Desse modo, as unidades aflorantes na região seriam mais antigas em relação às seções-tipo da Bacia do Solimões que representam a unidade (**Figura 15**).

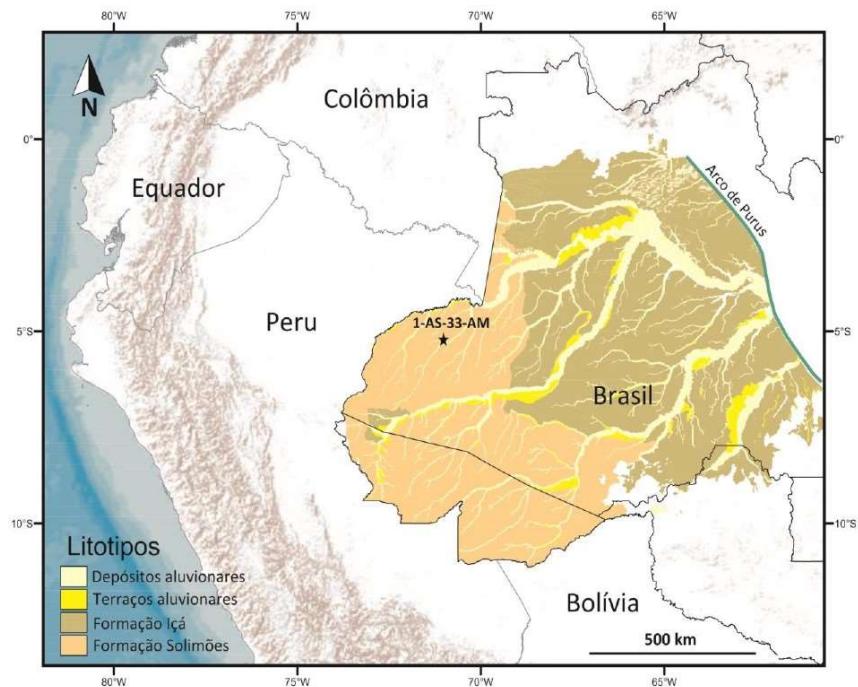


Figura 15. Região na qual aflora a Formação Solimões. Notar que a Fm. Solimões corresponde quase a totalidade da unidade aflorante na região do estado do Acre, assim como grande parte da porção oeste da Bacia do Solimões, localizada no estado do Amazonas. Imagem: Medeiros, 2017.

O registro paleontológico da Formação Solimões também varia. Na Bacia do Acre existem diversas ocorrências de fósseis de vertebrados terrestres e de água doce (Negri et al., 2010), enquanto que no Vale do Javari e em exposições no Alto Solimões

foram encontrados somente invertebrados e matéria orgânica lenhosa (Maia et al., 1977).

Outro elemento importante para entender as diferenças da Formação Solimões nas duas bacias é o Arco Fitzcarrald: com uma área de 400.000 km² e orientação WSW-ESE, a feição geomorfológica está localizada no sudeste do Peru e na porção oeste da planície Amazônica (Espurt et al., 2010), destaca-se na região por ser uma porção sobrelevada, com altitude média de 615 m, enquanto as bacias adjacentes de Ucayali e Madre de Dios apresentam respectivamente altitude média de 120 m e 150 m. Existem muitas discussões para a causa do soerguimento dessa porção do foreland dos Andes e a mais aceita atualmente, com base em evidências geomorfológicas, geofísicas e geodinâmicas é de que a subducção sub-horizontal da Placa de Nazca, iniciada há aproximadamente 4 m.a. seria a responsável para a origem da feição no relevo (Espurt et al., 2007).

O Arco de Fitzcarrald possui influências na configuração atual da drenagem da Bacia Amazônica, criando um padrão de drenagens radial assimétrico, com a maior parte do fluxo para NE, das bacias Ucayali e Madre de Dios em direção à Bacia do Acre (Espurt et al., 2010). Os depósitos que compõem o arco são do Mioceno e do Plioceno. A origem e evolução da feição no Plioceno-Pleistoceno teriam resultado em mudanças no padrão de drenagem, de erosão e de deposição da região, de modo que o Arco de Fitzcarrald deve ser levado em consideração para a reconstituição da história tectono-sedimentar, biológica e hidrológica da região (Espurt et al., 2010).

A partir desse quadro regional de evolução geológica e de reconstituições de ambientes deposicionais no Cenozoico, diversos autores propuseram cenários paleogeográficos distintos para a Bacia do Acre e a Bacia do Solimões (Latrubbles et al., 1997; Hoorn, 2010)..

Devido à grande extensão da Fm. Solimões, que abrange duas bacias com descrições litológicas distintas para cada uma delas, existe uma grande importância em estudar essa Formação para o maior entendimento do contexto fluvial na evolução paleoambiental amazônico. Atualmente, há estudos em andamento do grupo de pesquisa do IGc relacionado ao Trans-Amazon Drilling Project, sob comando do professor André Oliveira Sawakuchi, sobre o contexto fluvial da Formação Solimões.

4.4. Análise de exploratória de dados (Bussab e Moretin, 2017)

As medidas estatísticas desempenham um papel importante na análise e interpretação de dados. É uma ferramenta crucial para compreender corretamente um determinado conjunto de dados. No contexto deste trabalho, merecem destaque

quatro medidas estatísticas fundamentais: a média, o desvio padrão, a curtose e a assimetria (*skewness*) e o coeficiente de variação (CV).

Essas medidas estatísticas são amplamente utilizadas em várias áreas, como ciências sociais, economia, saúde e pesquisa de mercado. Elas fornecem informações valiosas sobre a distribuição dos dados e ajudam a identificar padrões, tendências e anomalias nos conjuntos de dados.

a) Média: a média aritmética é uma métrica central que representa o valor médio de um conjunto de dados. É calculado somando todos os valores e dividindo pelo número de observações. A fórmula é expressa como:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

(equação 01)

onde \bar{x} é a média e x_i são os valores individuais.

A média é uma estimativa sensível à presença de valores extremos, que pode gerar distorções em distribuições que sejam assimétricas.

b) Desvio padrão: o desvio padrão quantifica a dispersão dos dados em torno da média. Quanto menor a variância, menor o desvio padrão e vice-versa. A fórmula do desvio padrão é definida da seguinte forma:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

(equação 02)

onde σ representa o desvio padrão.

c) Assimetria: A assimetria mede a forma e a inclinação das caudas de uma distribuição em relação à distribuição normal. Uma curtose mais alta (positiva) indica caudas mais pesadas, enquanto uma curtose mais baixa (negativa) indica caudas mais leves. A fórmula da curtose é intrincada e contém elementos estatísticos, mas a interpretação baseia-se basicamente na comparação com a distribuição normal.

A curtose fornece informações sobre a concentração ou espalhamento dos valores ao redor da média. A fórmula para cálculo da curtose é definida como:

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2} - 3$$

(equação 03)

onde K é a curtose.

d) Assimetria: A assimetria (skewness) mede a assimetria da distribuição. Uma skewness positiva indica uma cauda longa à direita (valores maiores), enquanto uma skewness negativa sugere uma cauda longa à esquerda (valores menores). A skewness ajuda a identificar se há assimetria nos dados e qual direção ela segue.

A fórmula assimétrica é definida da seguinte forma:

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{3}{2}}}$$

(equação 04)

sendo S a assimetria.

e) Coeficiente de variação: O coeficiente de variação (CV) é uma medida estatística que expressa a variabilidade relativa de um conjunto de dados em relação à sua média. Essa medida é especialmente útil quando se deseja comparar a dispersão de conjuntos de dados que possuem diferentes unidades de medida ou escalas. O coeficiente de variação é calculado pela divisão do desvio padrão pela média e, em seguida, multiplicado por 100 para expressar o resultado em termos percentuais.

A fórmula para o coeficiente de variação (CV) é dada por:

$$CV = \left(\frac{\sigma}{\bar{x}} \right) \times 100$$

(equação 05)

O resultado do coeficiente de variação é expresso como uma porcentagem e fornece uma medida relativa da variabilidade em relação à média. Quanto menor o CV, menor a variabilidade em relação à média, indicando uma maior consistência nos dados. Por outro lado, um CV mais elevado sugere uma maior variabilidade em relação à média.

Por fim, as análises gráficas podem ajudar nas interpretações geológicas de maneira quantitativa feita através de análises estatísticas, auxiliando na compreensão da Fm. Solimões através das tendências e comportamentos gráficos dos dados de perfilagens de poços e a modelagem correta do código pelo método *Random Forest* (*Machine Learning*).

4.5. Machine Learning

O *Random Forest* (“floresta aleatória”) é um algoritmo robusto de *Machine Learning* que opera por árvores de decisão de forma aleatória através de votação preditiva com regras de classificação para o resultado, a partir de uma base de dados amostral treino, utilizando Bootstrap para a construção das árvores de decisão. É uma técnica poderosa de aprendizado de máquina ensemble que combina a força de múltiplas árvores de decisão para melhorar a precisão e a robustez das previsões (Breiman, 2001)

O *Random Forest* apresenta alta capacidade de lidar com variedade de dados e problemas, desde classificação até regressão, capaz de melhorar o desempenho geral., com uma coleção de árvores de decisão independentes com uma amostra aleatória de dados de treinamento combinando as previsões em um resultado final (Frajacomo, 2020) – neste caso, utilizando os parâmetros geofísicos (melhor explicado no item *Machine Learning* na seção “Metodologia”) extraídos dos poços que apresentam as informações de topo e base.

O primeiro passo na criação de uma *Random Forest* envolve a construção de árvores de decisão individuais. Cada árvore é treinada em uma amostra aleatória do conjunto de dados original, utilizando um subconjunto aleatório de características em cada divisão do nó. Este processo introduz diversidade nas árvores, tornando-as menos propensas a overfitting e mais robustas em relação a diferentes padrões nos dados (Frajacomo, 2020).

Uma característica fundamental do *Random Forest* é a amostragem Bootstrap. O Bootstrap é uma ferramenta de reamostragem proposta por Bradley Efron (1979) muito útil para estatística não paramétrica, como por exemplo em casos em que a população em estudo não segue uma distribuição normal, que é um caso frequente na Geologia, como por exemplo na análise de dados azimutais, dados composicionais, dentre outros. O princípio compreende a reamostragem aleatória com reposição de dados (conteúdo de aula de GSA0200).

Em cada iteração, uma amostra é retirada aleatoriamente do conjunto de dados original e substituída, permitindo que alguns pontos de dados apareçam várias vezes na amostra e outros pontos de dados nem apareçam. Este método de amostragem aumenta a variação das árvores e contribui para a diversidade geral (Frajacomo, 2020).

Depois que as árvores são construídas, o *Random Forest* combina suas previsões. Para problemas de classificação é realizada votação majoritária, enquanto para problemas de regressão é calculada a média. Esta agregação de resultados

reduz o impacto de outliers e erros individuais, tornando as florestas aleatórias mais resistentes ao ruído nos dados (Frajacomo, 2020).

Uma árvore de decisão é uma estrutura hierárquica composta por nós (que representam decisões) e arestas (que representam condições que guiam para decisões subsequentes). O algoritmo constrói essas árvores por meio de processos como particionamento recursivo e seleção de atributos (Frajacomo, 2020).

A construção de uma *Random Forest*, portanto, envolve os seguintes passos principais: a) Amostragem Aleatória: Para cada árvore de decisão no conjunto, uma amostra aleatória (com substituição) dos dados de treinamento é selecionada; b) Construção das Árvores: Cada árvore é construída usando a amostra aleatória e um subconjunto aleatório de atributos em cada nó; c) Combinação de Previsões: As previsões de todas as árvores são combinadas para determinar a classe final (no caso de classificação) ou valor (no caso de regressão).

O *Random Forest* apresenta várias vantagens, incluindo alta precisão de previsão - produz modelos mais precisos, especialmente em conjuntos de dados complexos e grandes -, resistência a overfitting (sobreajuste) - ao construir árvores rasas e utilizar amostragem aleatória, a *Random Forest* é menos suscetível a overfitting em comparação com uma única árvore de decisão -, capacidade de lidar com dados faltantes - pode lidar efetivamente com valores ausentes e características categóricas, sem a necessidade de pré-processamento extensivo-, e avaliação da importância dos atributos - fornece uma medida da importância de cada característica no processo de tomada de decisão, auxiliando na interpretabilidade do modelo (adaptado do site da IBM).

No entanto, seu principal desafio reside na interpretabilidade das decisões individuais de cada árvore e na complexidade computacional, uma vez que várias árvores precisam ser construídas e combinadas. Além disso, para conjuntos de dados extremamente grandes, o treinamento de muitas árvores pode exigir tempo significativo e alto espaço de armazenamento para múltiplas árvores, demandando elevado processamento computacional (IBM).

Por este motivo, com sua abordagem *ensemble* inovadora, a *Random Forest* é uma popular escolha em *Machine Learning* devido à sua flexibilidade, precisão e resistência a overfitting. Compreender seus princípios fundamentais e características distintivas é essencial para aplicar eficazmente este método em uma variedade de contextos. Embora apresente algumas limitações, a *Random Forest* permanece uma ferramenta valiosa no arsenal de algoritmos de *Machine Learning*, tendo sido escolhido como base para proposta de delimitação da Formação Solimões.

5. MATERIAIS E MÉTODOS

A metodologia a seguir foi utilizada ao longo do projeto para o cumprimento das metas estabelecidas nos objetivos, e engloba desde a aquisição e tratamento de dados à conclusão e interpretação dos resultados. O fluxograma com as etapas desenvolvidas por ser visto na **Figura 16** e o detalhamento das atividades nos tópicos subsequentes.

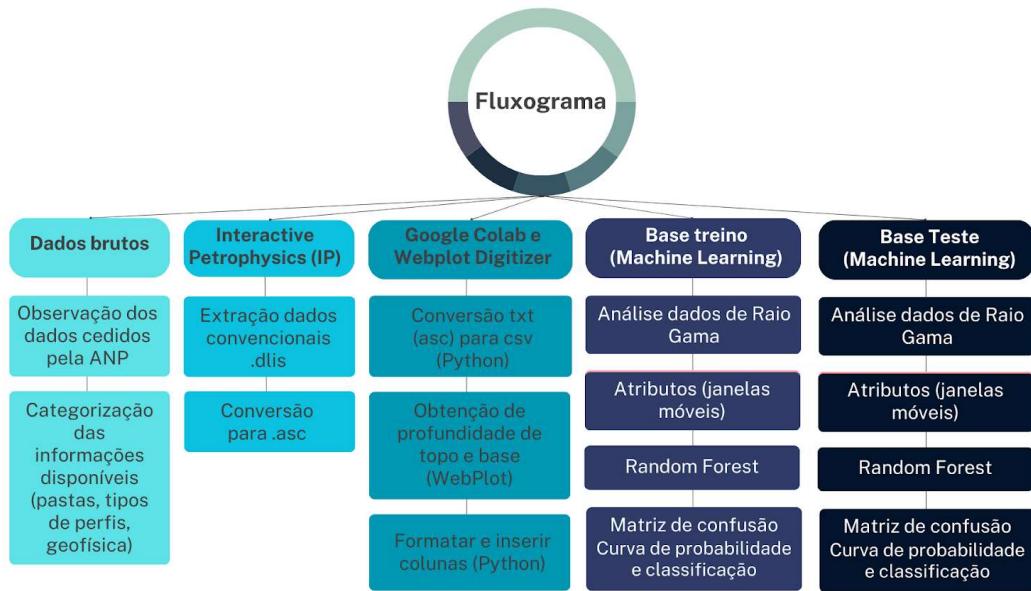


Figura 16. Fluxograma com os procedimentos a serem realizados durante todo o projeto.

5.1. Revisão bibliográfica e organização dos dados

Foi realizada uma revisão bibliográfica dos modelos estratigráficos propostos de evolução tectono-sedimentar para a Bacia do Acre e Bacia do Solimões e das características esperadas para a Formação Solimões com base na literatura disponível, além de organização dos dados disponibilizados pela ANP dos poços alocados na região por meio do programa do pacote Office Excel e escolha dos poços a serem utilizados, tendo em vista a limitação temporal do projeto e ausência de dados nas pastas convencionais dos poços e/ou na Formação Solimões. Os critérios utilizados para seleção dos poços e parâmetros utilizados foram definidos após a extração dos dados convencionais e devida formatação.

A revisão bibliográfica compreendeu também tópicos fundamentais no aprendizado introdutório à Linguagem Python e ao algoritmo Random Forest utilizado no *Machine Learning*, com base no material disponibilizado durante a disciplina de

GSA0200 – Introdução à Análise de Dados em Geologia Sedimentar (2023) e na literatura disponível.

A organização dos poços foi disposta em uma tabela de acordo com a **Figura 17**.

Bacia	Coordenada (UTM)	Coordenadas (geo) - Boca do furo						Prof. (m)	Reclassificação	Arquivos					
		N	E	MC	DATUM	lat	long				AGP	Amostra doras laterais	Arquivo geral de poço	Dados de canhone de direcionais	Dados de Rochas e Fluidos
pocos com pasta extraída															
1-AJ-1-AM	SO	9449310.6	783445.6	69°	SAD-69	4°58'37.60"S	66°26'38.80"W	3340.0	Seco s/ indicio de	66					
1-BRSA-56-AM	SO	9464090.2	180535.1	63°	SAD-69	4°49'33.57"S	65°53'11.80"W	3461.0	Portador Gás	52					
1-BRSA-58-AM	SO	9471641.4	338509.7	63°	SAD-69	4°46'42.78"S	64°27'22.10"W	2000.0	Seco c/ indicio de	63					
1-BRSA-372-AM	SO	9464058.8	283194.2	63°	SAD-69	4°49'51.46"S	64°37'39.55"W	2610.0	Seco c/ indicio de	63					
1-BRSA-714-AM	SO	9444956.3	367746.3	63°	SAD-69	5°01'13.50"S	64°11'34.76"W	2170.0	Seco com indicio	62					
1-BRSA-738-AM	SO	9454819.1	359806.4	63°	SAD-69	4°55'55.89"S	64°15'51.94"W	2513.0	Seco com indicio	62					
1-BRSA-769-AM	SO	9488034.7	231869.7	63°	SAD-69	4°37'39.59"S	65°25'00.22"W	3485.0	Descobridor de c	15					
1-BRSA-835-AM	SO	9517454.1	361822.7	63°	SAD-69	4°21'02.57"S	64°14'01.73"W	2421.0	Seco c/ indicio de	63					
1-BRSA-892-AM	SO	9470970.6	275263.9	63°	SAD-69	4°46'59.40"S	65°01'34.43"W	3240.0							
1-BRSA-961-AM	SO	9485379.9	247164.2	63°	SAD-69	4°39'07.63"S	65°16'44.49"W	3290.0	Descobridor de c	15					
1-BRSA-984-AM	SO	9460551.1	238374.1	63°	SAD-69	4°52'35.27"S	65°21'32.33"W	2855.0							
1-BRSA-985-AM	SO	9429789.7	215601.3	63°	SAD-69	5°09'13.39"S	65°33'54.96"W	2600.0	Seco c/ indicio de	61					
1-BRSA-1059-AM	SO	9497368.2	241032.9	63°	SAD-69	4°32'36.87"S	65°20'02.07"W	3546.0	Portador Oleo. G	55					
1-BRSA-1189-AM	SO	9481033.2	289579.1	63°	SAD-69	4°41'33.20"S	64°53'49.08"W	2966.0	Seco com indicio	62					
1-BRSA-1236-AM	SO	9394430.5	169403.8	63°	SIRGAS2000	5°28'16.25"S	65°58'59.37"W	2064.0	Seco s/ indicio de	66					
1-BRSA-1311-AM	SO	9474982.7	248875.0	63°	SIRGAS2000	4°44'46.22"S	65°15'50.13"W	3042.0	Seco s/ indicio de	66					
1-BT-1-AM	AC	9247600.0	699500.0	75°	SAD-69	6°48'12.35"S	73°11'41.53"W	3749.0	Seco s/ indicio de	66					
1-BT-2-AM	AC	9234314.3	704323.5	75°	SAD-69	6°55'24.17"S	73°09'02.81"W	2790.0	Seco s/ indicio de	66					
1-BV-1-AM	SO	9536201.7	231469.4	63°	SAD-69	4°11'32.14"S	65°25'07.82"W	3410.0	Seco c/ indicio de	63					
1-CAT-1-AM	SO	9494103.0	292947.7	63°	SAD-69	4°34'27.95"S	64°51'58.43"W	2719.0	Seco s/ indicio de	66					
1-Cl-1-AM	SO	9453917.9	744170.5	69°	SAD-69	4°56'12.30"S	66°47'53.56"W	2651.7	Seco com indicio	62					
1-CP-1-AM	SO	9453456.5	2222697.7	63°	SAD-69	4° 56.24.2	65 30 2.5	2946	Seco com indicio de óleo e gás						
1-CR-1-AM	SO	9515129.2	290839.8	63°	SAD-69	4°23'03.34"S	64°53'05.03"W	2721.0	Seco s/ indicio de	66					
1-CRP-1-AM	SO	9440816.6	230165.1	63°	SAD-69	5°03'15.66"S	65°26'00.67"W	2675.0	Descobridor de c	14					
1-CS-3-AM	AM	9580658.8	601048.0	63°	SAD-69	3°47'55.99"S	62°05'24.00"W	1320.5	Seco s/ indicio de	66					
1-CUP-1-AM	SO	9445441.1	236095.2	63°	SAD-69	5°00'45.90"S	65°22'47.71"W	2683.6	Descobridor de c	14					
1-GV-1-AM	SO	9463351.4	721162.5	69°	SAD-69	4°51'07.60"S	67°00'21.02"W	2508.0	Seco com indicio	62					
1-HRT-1-AM	SO	9481536.5	213895.2	63°	SAD-69	4°41'07.32"S	65°34'43.82"W	3462.2	Portador Gás	52					
1-HRT-2-AM	SO	9473199.5	831107.6	69°	SAD-69	4°45'33.03"S	66°00'53.97"W	3470.0							
1-HRT-3-AM	SO	9462091.1	726271.9	69°	SAD-69	4°51'48.26"S	66°37'34.88"W	2680.0							
1-HRT-5-AM	SO	9436244.7	787504.3	69°	SAD-69	5°05'22.64"S	66°24'25.12"W	2983.0							
1-HRT-6-AM	SO	9471923.3	219681.8	63°	SAD-69	4°46'22.41"S	65°31'37.32"W	3384.0							
1-HRT-8-AM	SO	9452092.1	819894.6	69°	SAD-69	4°57'02.42"S	66°07'03.06"W	3345.0							
1-HRT-9-AM	SO	9443371.7	763045.4	69°	SAD-69	5°01'53.48"S	66°37'39.63"W	2960.0							
1-HRT-10-AM	SO	9415339.1	796312.8	69°	SAD-69	5°17'01.25"S	66°19'36.46"W	2385.0							
1-HRT-11-AM	SO	9491158.5	327537.3	63°	SAD-69	4°36'06.61"S	64°33'16.78"W	2441.0							
1-IA-1-AM	SO	9507227.9	339315.7	63°	SAD-69	4°27'24.12"S	64°26'53.41"W	2166.6	Seco s/ indicio de	66					

Figura 17. Cópia de trecho da tabela de controle de poços.

Pela análise das informações disponibilizadas, dentre os 11 poços na Bacia do Acre, 3 apresentam indícios de hidrocarbonetos – um com indício de gás e dois com indícios de óleo -, sem exploração comercial no momento -; e na Bacia do Solimões, dentre os 172 poços instalados há 24 poços descobridores de campo (códigos ANP 12, 14, 44 e 45), 4 poços portadores de hidrocarbonetos (códigos ANP 52, 54 e 55), 31 poços secos com indícios de hidrocarbonetos (códigos ANP 61, 62, 63, 64 e 65) e 23 poços produtores comerciais ou sub-comerciais ou ainda como extensão de poço produtor de pelo menos um dos hidrocarbonetos (óleo, gás e condensados) (códigos ANP 1, 2, 5, 32, 34, 71, 74, 75), de acordo com os dados de poços fornecidos pela ANP para este projeto, totalizando 85 poços de interesse comercial dos 183 poços implantados nas duas Bacias.

De forma resumida, a classificação dos poços quanto à presença de hidrocarbonetos (óleo, gás e condensados), de acordo com a PORTARIA Nº 76, DE 3 DE MAIO DE 2000 da ANP, pode se dividir em 10 tipos diferentes, sendo os tipos citados acima definidos como:

- Poço descobridor de campo: poço cujo resultado foi a descoberta de uma nova área produtora ou potencialmente produtora de petróleo e/ou gás natural;
 - Poço de extensão: poço com petróleo e/ou gás natural para a delimitação ou a ampliação de uma jazida, podendo ou não ser aproveitado economicamente para produção;
 - Poço produtor comercial: poço que possibilite a drenagem econômica de petróleo e/ou gás natural de um reservatório;
 - Poço produtor sub-comercial: poço que apresenta uma quantidade de hidrocarbonetos, porém a produção de petróleo e/ou gás natural é considerada inviável economicamente à época de sua avaliação, podendo se tornar produtor no futuro caso as condições econômicas do mercado e locais se alterem;
 - Poço portador de petróleo e/ou gás natural: é todo poço incapaz de permitir a produção em quantidades comerciais, independentemente das facilidades de produção na área;
- VII - Poço seco: é todo poço onde não se caracterizou a presença de petróleo móvel e/ou gás natural, podendo apresentar indícios ou não de hidrocarbonetos.

A dispersão espacial destes poços pode ser observada na **Figura 18** a seguir:

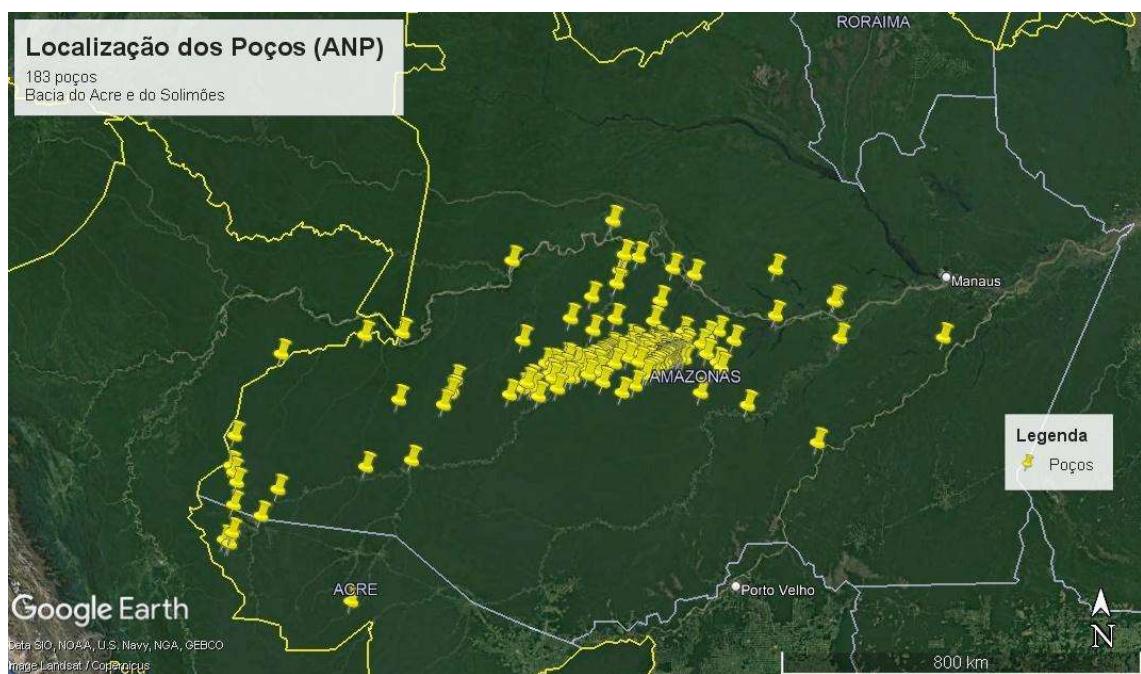


Figura 18. Localização geográfica dos poços atuais na Bacia do Acre e Bacia do Solimões.
Fonte: gerado a partir dos dados de poços e coordenadas fornecidos pela ANP.

5.2. Interactive Petrophysics (IP)

Utilizando do software Interactive Petrophysics (IP) por meio da licença fornecida pelo Laboratório de Informática Geológica (LIG), extraiu-se os dados convencionais na extensão “.dli” disponibilizados pela ANP dos poços locados na Bacia do Acre e Solimões, convertendo-os em “.asc” a ser utilizado como base para análise de dados no Python utilizando Google Colab para conversão das tabelas em “.csv”.

O IP é um software pago muito utilizado por geólogos e engenheiros de reservatório na indústria de petróleo por oferecer funcionalidades de interesse como cálculo de volume de argila, porosidade, saturação de água, elaboração de perfis compostos, dentre outras.

Atualmente o LIG apresenta uma licença disponível para uso de docentes e alunos (**Figura 19**).

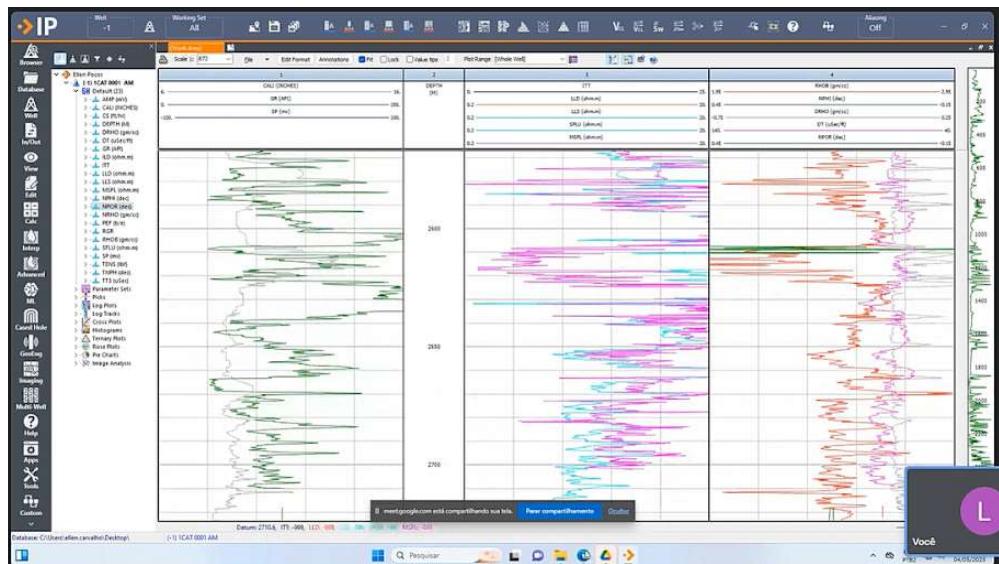


Figura 19. Interface do Interactive Petrophysics (IP).

A perfilagem de poços permite, mesmo na ausência de testemunhos, mensurar propriedades físicas das rochas como: condutividade, porosidade e permeabilidade (medidas indiretas), % de U, Th e K, densidade, velocidade e outras propriedades. Através dessas propriedades é possível interpretar litotipos e padrões estratigráficos, além de permitir correlações entre poços. As perfilagens mais comuns para encontrar as propriedades físicas das rochas citadas acima são: perfilagem de resistividade, perfilagem de raios gama (natural e densidade) e perfil sônico. Normalmente são

requisitados em qualquer tipo de perfilagem de poço, por isso são métodos geofísicos extremamente importantes para exploração de petróleo.

Dentre estes perfis, destacam-se os perfis radiométricos (Keary, Brooks e Hill, 2002). Os perfis radiométricos são utilizados para identificar a litologia, densidade e porosidade ao redor do poço através da emissão natural de elementos instáveis como ^{238}U , ^{232}Th e ^{40}K (perfil raios gama) ou pode ser induzido um bombardeio radiométrico de núcleos estáveis com raio gama ou nêutrons (perfil densidade ou neutrão). A ferramenta utilizada para detectá-los normalmente é um contador de cintilação e a unidade de medida da radiação nos poços é expressa pela unidade API (American Petroleum Institute).

O perfil de raios gama detecta raios gama natural emitidos pelas rochas e identifica camadas com níveis elevados de alguns dos três elementos medidos (^{238}U , ^{232}Th ou ^{40}K). Por exemplo, um folhelho com mica, feldspato alcalino ou argilominerais terá um valor elevado de potássio (^{40}K) e alguns elementos traços de ^{238}U e ^{232}Th . Essas características aumentam substancialmente o valor no perfil, assim como evaporitos (rochas selantes) que contém altas quantidades de potássio, o perfil de raios gama natural pode ajudar a refinar a interpretação dados pelos poços de resistividade.

Já no perfil de densidade de raios gama emite-se uma radiação artificial de ^{60}Co ou ^{137}Cs (fonte) nas rochas, os fótons de raios gamas emitidos colidem sobre os elétrons das rochas e reduzem sua energia (espalhamento de Compton). O número de colisão em um determinado tempo depende da quantidade de elétrons da rocha, que são dependentes da densidade de formação. A relação entre a densidade de formação (ρ_f) e a densidade dos elétrons (ρ_e) é dado por:

$$\rho_f = \rho_e w / (2 \sum N), \quad (\text{equação 06})$$

onde w é o peso molecular dos constituintes da formação e N o número atômico dos elementos presentes. A porosidade pode ser estimada através da densidade de formação (ρ_f), densidade da matriz (ρ_m) e densidade de fluidos da matriz (ρ_w):

$$\Phi = (\rho_m - \rho_f) / (\rho_m - \rho_w). \quad (\text{equação 07})$$

Observando os dados geofísicos disponíveis (**Figura 20**), optou-se por utilizar a Curva de Raio de Gama, por ser o perfil mais presente na profundidade explorada para a Formação Solimões (os outros perfis geofísicos ou eram ausentes em boa

parte dos poços ou foram medidos apenas em profundidades maiores), e por ser diretamente relacionado com a litologia das rochas e em consequência com a formação em análise como demonstrado anteriormente.

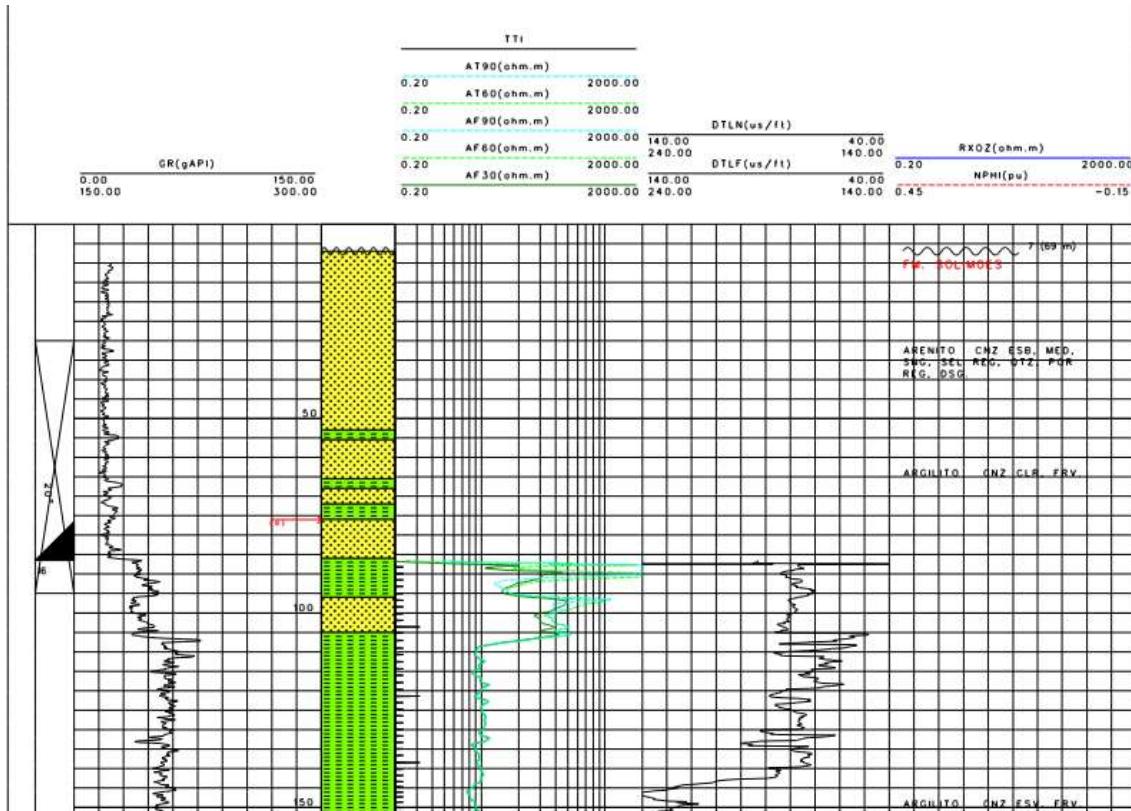


Figura 20. Trecho extraído como exemplo a partir do perfil composto do poço 1-BRSA-55-AM, contendo 3 rampas de perfilagem: a curva de Raio Gama (Gamma Ray - GR) na primeira rampa, perfis de resistividade em escala logarítmica na segunda, perfil sônico (DTL) e microresistividade (RXOZ) na terceira, além da coluna de litologia, legenda e escala dos perfis, marcação do topo da Formação Solimões e características do poço e das rochas presentes. Fonte: dados fornecidos pela ANP.

Dentre os 182 poços fornecidos pela ANP – dos 183 poços existentes, um dos poços não veio -, 9 poços da Bacia do Acre e 45 poços na Bacia do Solimões apresentavam perfis compostos com os dados de profundidade da Formação Solimões. Dos 128 poços restantes, 27 poços da Bacia do Solimões não apresentavam perfil composto, e ainda 2 poços da Bacia do Acre e 99 poços da Bacia do Solimões apresentavam perfis compostos sem a profundidade da Formação Solimões.

Dentre os 54 poços com informações, foram encontrados erros importantes: 3 poços apresentavam apenas dados em ".tiff"; 1 poço foi extraído e gerou um arquivo grande o suficiente para não conseguir ser processado no Colab posteriormente, 12 poços também foram descartados pois apesar de apresentarem curva de gama no

perfil composto, ao abrirmos os dados no IP não havia dado convencional no trecho em que a formação estava presente e 5 poços foram descartados pois no perfil composto apresentavam formações com espessura menor que a resolução do perfil, de modo que ao extraímos as informações de litologia não foi possível classificar qual a litologia associada a essas camadas muito finas, o que poderia alterar os resultados da comparação litologia e valores de GR, e por consequência a análise da base da Formação neste poços. Portanto, para fins de ajustes para o projeto, foram utilizados efetivamente 33 poços para formação da base de treino e teste.

A extração dos dados, portanto, concentrou-se na obtenção dos valores de gama para trechos de profundidades que compreendessem a Formação Solimões nesses 33 poços selecionados, convertendo para “.asc” a fim de realizar o devido tratamento e conversão no Python para continuação das etapas seguintes do projeto.

5.3. Formatação dos dados (Google Colab) e coleta de profundidades (WebPlot Digitizer)

Após a extração dos dados de perfilagem em “.asc” – profundidade e GR – os arquivos foram convertidos manualmente para “.txt”, e em seguida convertidos para “.csv” e analisados estatisticamente no Python utilizando o Google Colab de acordo com o código mostrado na **Figura 21**.

O Google Colaboratory, também denominado como Google Colab, é um serviço de nuvem gratuito da Google para criação de códigos por meio de linguagem python, com uma ampla disponibilidade pacotes e bibliotecas nas mais variadas funcionalidades, desde cálculos numéricos, elaboração de tabelas (dataframes), gráficos, edição de textos e imagens, dentre outras, por meio de notebooks virtuais colaborativos.

```

✓ 1s [1] import pandas as pd
    from pandas import DataFrame
    import numpy as np
    import matplotlib.pyplot as plt
    import matplotlib.mlab as mlab
    import seaborn as sns
    import scipy
    import re
    import csv
    from io import StringIO
    from google.colab import files

✓ 0s [2] from google.colab import drive

✓ 45s ① files.upload()

[ ] nome_arquivo='1-BRSA-55-AM'
arquivo='1-BRSA-55-AM'+'.txt'
csvfile = open(arquivo) #open text file

reader = csv.reader(csvfile)
new_list=[]

for line in reader:
    for i in line:
        new_list.append(re.findall(r'(\d*\.\d+|\d+)', i))

table = pd.DataFrame(new_list)
display(table.iloc[:20,:])
#conferir quantas linhas temos que tirar
tirar_linhas=3
manter_colunas=46
coluna_profundidade_duplicada=1 #0 se não e 1 se sim

table=table.iloc[tirar_linhas:,coluna_profundidade_duplicada:manter_colunas]

#print(table) # output will be pandas DataFrame with values
display(table)

② #Conferir se a lista de colunas é mesmo esta
table.columns=['ABS','AFCD','GR','DEPTH']
table.replace('999.0000',np.NaN,inplace=True)

table.to_csv(nome_arquivo+".csv")

```

Figura 21. código utilizado no Google Colab para conversão dos dados “.txt” para “.csv”.

Alguns poços tiveram problemas de leitura ao ler uma das linhas tabeladas, e por isso foi necessário fazer um tratamento prévio de cada um para exclusão das linhas extras (que continham informações de data ou valores nulos de parágrafo do cabeçalho) conforme mostrado acima.

Na sequência, foram coletadas para cada um dos poços selecionados a profundidade das formações presentes e a litologia definida pela ANP nos poços, com as profundidades de topo e base de cada uma, utilizando o software WebPlot Digitizer, disponível de forma gratuita e online no site oficial. A interface geral do site pode ser observada na **Figura 22**.

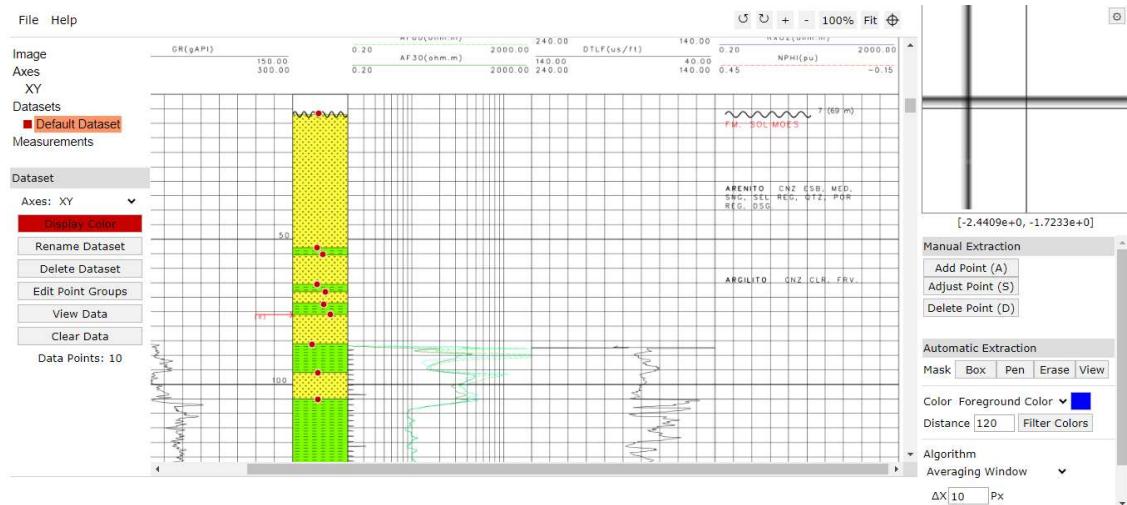


Figura 22. Interface do WebPlot Digitizer utilizando o poço 1-BRSA-55-AM como exemplo.

Os valores da profundidade das litologias foram exportados arquivo “.csv”, e a correspondência entre a profundidade de topo e base e a litologia foi feita manualmente no Excel, criando as colunas de litologia e formação.

A partir do arquivo salvo anteriormente com os dados de profundidade e GR, e do código presente na **Figura 23**, foi realizado um mapeamento dos pares ordenados “(DEPTH;GR)” do arquivo base do poço e atribuídos os valores de ‘LITO’ e ‘FORMATACAO’ obtidos do “.csv” de saída do Webplot.

```

1s  ✓  import pandas as pd
      from pandas import DataFrame
      import numpy as np
      import matplotlib.pyplot as plt
      import matplotlib.mlab as mlab
      import seaborn as sns
      import random
      from scipy import stats
      from cycler import cycler
      from google.colab import drive
      from google.colab import files

[ ] files.upload()

[ ] caminho = '/content/1-HRT-9-AM.csv'
df = pd.read_csv(caminho)

print(df.head())

#excluir primeira coluna que foi gerada como posição apenas em alguns poços
df = df.drop(df.columns[0], axis=1)
print(df.head())

[ ] #criando os limites de litologia
litologias = {
(7.0473644, 35.06493506): 'argilito',
(35.06493506, 72.02062643): 'arenito',
(72.02062643, 105.5385791): 'argilito',
(105.5385791, 107.4293354): 'arenito',
(107.4293354, 127.8838808): 'argilito',
(866.9977082, 891.0618793): 'siltito',
(1692.398778, 1696.867838): 'anidrita',
(1696.867838, 1699.961803): 'folhelho',
(1699.961803, 1702.024446): 'calcarenito',
(1702.024446, 1705.977846): 'folhelho',
(1705.977846, 1707.524828): 'calcilitito',
}

[ ] def mapear_litologias(row):
    for (inicio, fim), litologia in litologias.items():
        if inicio <= row['DEPTH'] <= fim:
            return litologia
    return 'Desconhecida'

df['LITOLOGIA'] = df.apply(mapear_litologias, axis=1)

print(df)

[ ] #criando os limites de formacao
formacoes = {
(8, 399): 'solimoes',
(399, 794): 'alter_do_chao',
(794, 1262): 'soleira',
(1262, 1543): 'carauari',
(1543, 1701): 'soleira',
(1701, 2463): 'carauari',
(2463, 2045): 'soleira',
}

[ ] def mapear_formacoes(row):
    for (inicio, fim), formacao in formacoes.items():
        if inicio <= row['DEPTH'] <= fim:
            return formacao
    return 'Desconhecida'

df['FORMACAO'] = df.apply(mapear_formacoes, axis=1)

print(df)

[ ] nome_arquivo='1-HRT-9-AM'
df.to_csv(nome_arquivo+"_pronto.csv", index=False)

```

Figura 23. Código utilizado para adição das colunas de litologia de formação para os pontos obtidos de profundidade e GR.

Como alguns poços iniciais foram coletados com mais informações que o necessário ou com as colunas em ordens diferentes, foi aplicado um novo código para efeito de padronização dos poços a serem utilizados na análise de dados e na etapa do *Machine Learning* (**Figura 24**).

```
[ ] import pandas as pd
from pandas import DataFrame
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.mlab as mlab
import seaborn as sns
import random
from scipy import stats
from cycler import cycler
from google.colab import drive
from google.colab import files

[ ] uploaded=files.upload()

[ ] #criar novo dataframe só com as colunas de interesse

filename = list(uploaded.keys())[0]

df = pd.read_csv(filename)

df = df[['DEPTH', 'GR', 'LITOLOGIA', 'FORMACAO']]

print(df)

      DEPTH      GR    LITOLOGIA  FORMACAO
0     0.7620  7.2957  Desconhecida  Desconhecida
1     0.9144  7.2957  Desconhecida  Desconhecida
2     1.0668 18.3493  Desconhecida  Desconhecida
3     1.2192  9.7889  Desconhecida  Desconhecida
4     1.3716  9.5819  Desconhecida  Desconhecida
...
19144 2918.3076  0.0000  Desconhecida  Desconhecida
19145 2918.4600  0.0000  Desconhecida  Desconhecida
19146 2918.6124  0.0000  Desconhecida  Desconhecida
19147 2918.7648  0.0000  Desconhecida  Desconhecida
19148 2918.9172  0.0000  Desconhecida  Desconhecida

[ ] #remover linhas que tenham desconhecida simultaneamente para litologia e formação,
#por serem mais rasas que a coluna do perfil,
#ou por serem muito profundas em relação à Fm Solimões

df = df[(df['LITOLOGIA'] != 'Desconhecida') & (df['FORMACAO'] != 'Desconhecida')]

print (df)

      DEPTH      GR    LITOLOGIA  FORMACAO
42    7.1628  3.9413  argilito  solimoes
43    7.3152  3.5258  argilito  solimoes
44    7.4676  2.9026  argilito  solimoes
45    7.6200  1.8694  argilito  solimoes
46    7.7724  2.6962  argilito  solimoes
...
11195 1706.8800  78.9912  calcilutito  carauri
11196 1707.0324  77.8111  calcilutito  carauri
11197 1707.1848  86.7410  calcilutito  carauri
11198 1707.3372 103.9007  calcilutito  carauri
11199 1707.4896 112.3601  calcilutito  carauri

[11158 rows x 4 columns]

[ ] #remover pontos de GR com valores nulos

df = df.dropna(subset=['GR'])
print (df)

      DEPTH      GR    LITOLOGIA  FORMACAO
42    7.1628  3.9413  argilito  solimoes
43    7.3152  3.5258  argilito  solimoes
44    7.4676  2.9026  argilito  solimoes
45    7.6200  1.8694  argilito  solimoes
46    7.7724  2.6962  argilito  solimoes
...
11195 1706.8800  78.9912  calcilutito  carauri
11196 1707.0324  77.8111  calcilutito  carauri
11197 1707.1848  86.7410  calcilutito  carauri
11198 1707.3372 103.9007  calcilutito  carauri
11199 1707.4896 112.3601  calcilutito  carauri

[11158 rows x 4 columns]

[ ] nome_arquivo='6-BRSA-24-AM'
df.to_csv(nome_arquivo+"_formatado.csv", index=False)
```

Figura 24. Código no Google Colab para formatação dos dados finais a serem utilizados na análise de dados e *Machine Learning*.

5.4. Análise de exploratória de dados

Para a correta análise estatística dos dados disponíveis foi observada a distribuição dos dados de GR nos poços. Para isso, foram criados histogramas a fim de reconhecer a distribuição dos valores, como mostrado nas **Figuras 25 a 27** três exemplos desses histogramas para análise.

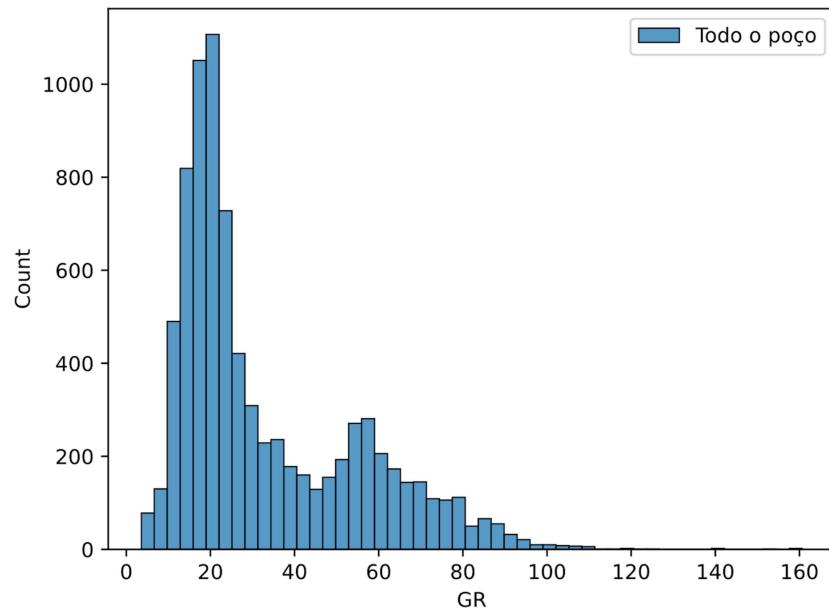


Figura 25. Histograma de contagem de dados GR para o poço 1-BRSA-55-AM.

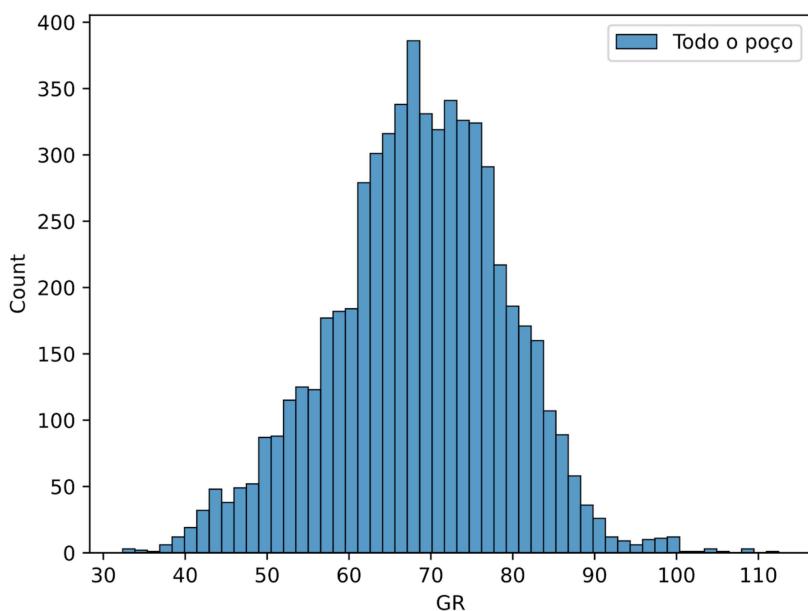


Figura 26. Histograma de contagem de dados GR para o poço 1-BT-1-AM.

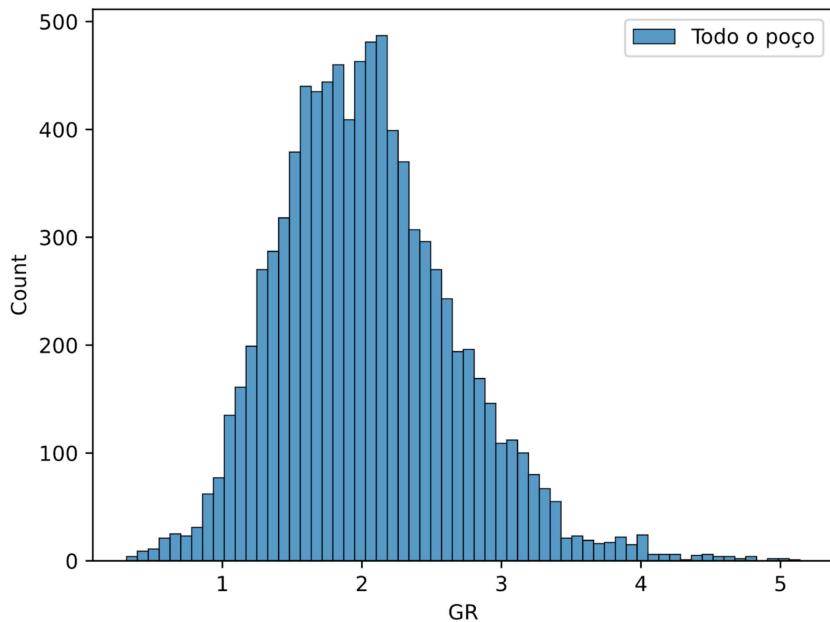


Figura 27. Histograma de contagem de dados GR para o poço 2-CDST-1-AC.

É possível observar que a distribuição dos valores de GR não é uniforme entre os poços. O primeiro ponto é a escala linear, alguns poços possuem escala para GR no perfil composto variando de 0 a 150 unidade API, enquanto outros apresentam escala medida em grau API de 0 a 5, como no caso do poço 2-CDST-1-AC, de modo que analisar a variável GR se torna mais complexo. Uma análise adimensional,

utilizando o coeficiente de variação – dividindo-se a média pelo desvio padrão – torna-se mais interessante a fins de comparação entre poços.

Outra observação interessante é com relação aos valores de gama. O poço 1-BRSA-55-AM apresenta uma distribuição bimodal e assimétrica, com o primeiro pico mais intenso em 20 API e o segundo pico bem menor próximo de 60 API. Já os poços 1-BT-1-AM e 2-CDST-1-AC demonstram uma distribuição normal, com média de 70 API e 2° API, respectivamente.

5.5 Machine Learning

Pelo fato de não haver uma regularidade nos dados coletados durante as sondagens dos diferentes poços fornecidos – alguns poços apresentam uma ampla variedade de dados de perfilagem geofísica, como perfil SP, GR, resistividade, ROHB e perfil sônico abrangendo a Fm. Solimões, enquanto outros apresentam tais curvas apenas em trechos específicos de profundidade para outras formações ou se encontram ausentes nos perfis – optou-se por escolher a curva de Raio Gama (Gamma Ray) para abordagem, tanto pela relação direta com a litologia quanto pela maior disponibilidade de dados nos diferentes perfis compostos.

Após o tratamento dos dados iniciais, possuímos 33 poços com informações sobre a delimitação de topo e base da Formação Solimões de acordo com a interpretação inicial fornecida pela ANP dos perfis compostos, com base nos dados obtidos da perfilagem geofísica e dados de amostras dos poços, com os dados dispostos em um dataframe único composto pelas colunas ‘DEPTH’, ‘GR’, ‘LITOLOGIA’ e ‘FORMACAO’, exportado em “.csv”.

A partir desses dados em “.csv”, para o reconhecimento automático da base da Formação Solimões em diferentes poços das bacias do Solimões e Acre, utilizou-se uma abordagem de *Machine Learnig* tendo como base o algoritmo *Random Forest* em linguagem Python.

Com a obtenção de dados de interesse a partir dos perfis disponíveis dos poços da ANP em estudo, nesta etapa utilizamos um ramo da inteligência artificial para classificação supervisionada de poços sem informações dos limites de topo e base das formações nas duas Bacias – com foco na Fm. Solimões em específico - a partir dos dados analisados comparativos em poços que apresentam tal informação. O método definido para a classificação foi *Random Forest*.

O professor Renato Paes de Almeida, orientador do projeto, foi responsável pela elaboração dos códigos “AM_model.pkl” e “Find_base.py” (**Anexos I e II**, respectivamente), utilizados no *Machine Learning*, uma vez que o conhecimento

necessário para sua execução era complexo e foge do escopo do curso de graduação abordado na disciplina GSA0200 – Introdução à Análise de Dados em Bacias Sedimentares.

Os códigos elaborados têm como base a utilização de algumas bibliotecas comuns no *Machine Learning*, como *Scikit-Learn* – biblioteca gratuita em Python com recursos eficientes para modelagem estatística, análise e mineração de dados além de suporte ao aprendizado supervisionado e não supervisionado com interação com bibliotecas mais populares como SciPy, Matplotlib e Numpy.

O primeiro código, “AM_model.pkl” é o código base para modelagem dos poços, e foi calibrado por tentativa e erro, utilizando alguns poços que pudessem servir como modelo de ajuste para cálculo da base da Fm Solimões em outros poços, comparando com o valor reconhecido no perfil composto pela ANP. Já o segundo código, “Find_base.py” aplica a classificação gerada no poço modelo para os outros poços, de modo a utilizar os critérios estabelecidos pela Random Forest para avaliar a probabilidade da profundidade mais plausível para a Fm Solimões nestes poços, à partir dos parâmetros treinados.

Dentre os poços selecionados para poço modelo, o poço que mais se adequou, por tentativa e erro, e forneceu respostas mais próximas dos valores observados nos perfis compostos dos poços selecionados foi o poço 1-BT-1-AM e por esse motivo o código “AM-model.pkl” utilizou os dados deste poço como base para a Random Forest.

De forma resumida, o código “Find_base.py” carrega o poço modelo 1-BT-1-AM e atribui o modelo para cada poço a fim de calcular a base para estes poços. Depois, à partir da coluna ‘DEPTH’, interpola o valor de GR para profundidades menores que 0,2 m (resolução mínima da perfilagem de GR nos poços na indústria do petróleo), gera o histograma dos valores, as janelas móveis no valor atribuído, a curva de probabilidade e faz a predição para a base da Formação Solimões no poço selecionado.

A maioria dos intérpretes utiliza não apenas os valores em unidades API como também a forma da curva de Gamma Ray (Raio Gama) em segmentos de metros a dezenas de metros de espessura. Com o objetivo de inserir informações compatíveis com esse tipo de interpretação do modelo de Machine Learning, foi adotada uma estratégia de criação de atributos derivados da curva original, considerando a variabilidade espacial em janelas móveis (composta por 5, 21, 41, 61, 81 dados consecutivos, respectivamente, capaz de atribuir uma medida estatística a uma posição dos dados da sequência numérica⁽¹⁾), após todos os valores serem amostrados para o espaçamento mais comum nos poços, de 20 cm.

(¹) De modo simples, pode-se explicar utilizando de forma análoga o conceito de uma Progressão Aritmética: suponha a sequência de números 2, 4, 6, 8, 10, 12. O cálculo da média a uma janela móvel de 4 dados consecutivos, por exemplo, seria atribuído a uma posição inexistente entre o segundo (4) e o terceiro (6) dados, depois entre o terceiro (6) e o quarto (8) e assim em diante. Já no caso de uma janela móvel de 5 dados sequenciais iria atribuir a média a uma posição determinada, primeiro para o terceiro valor, depois para o quarto e assim em diante.

Os atributos escolhidos foram os parâmetros das distribuições obtidas nas janelas móveis, incluindo o coeficiente de variação (média / desvio padrão), assimetria e excesso de curtose. Esses parâmetros adimensionais trazem informações sobre a vizinhança de cada ponto, gerando dados na mesma escala espacial da interpretação geológica tradicional.

As dezesseis colunas de atributos criadas, que incluem o dado original de 'Gamma Ray', cinco janelas de coeficiente de variação, cinco de assimetria e cinco de excesso de curtose, foram então analisadas para a avaliação de sua correlação. A **Figura 26** mostra que a correlação significativa ocorre apenas entre atributos derivados de um mesmo parâmetro estatístico e em janelas móveis de tamanho semelhante, especialmente o coeficiente de variação.

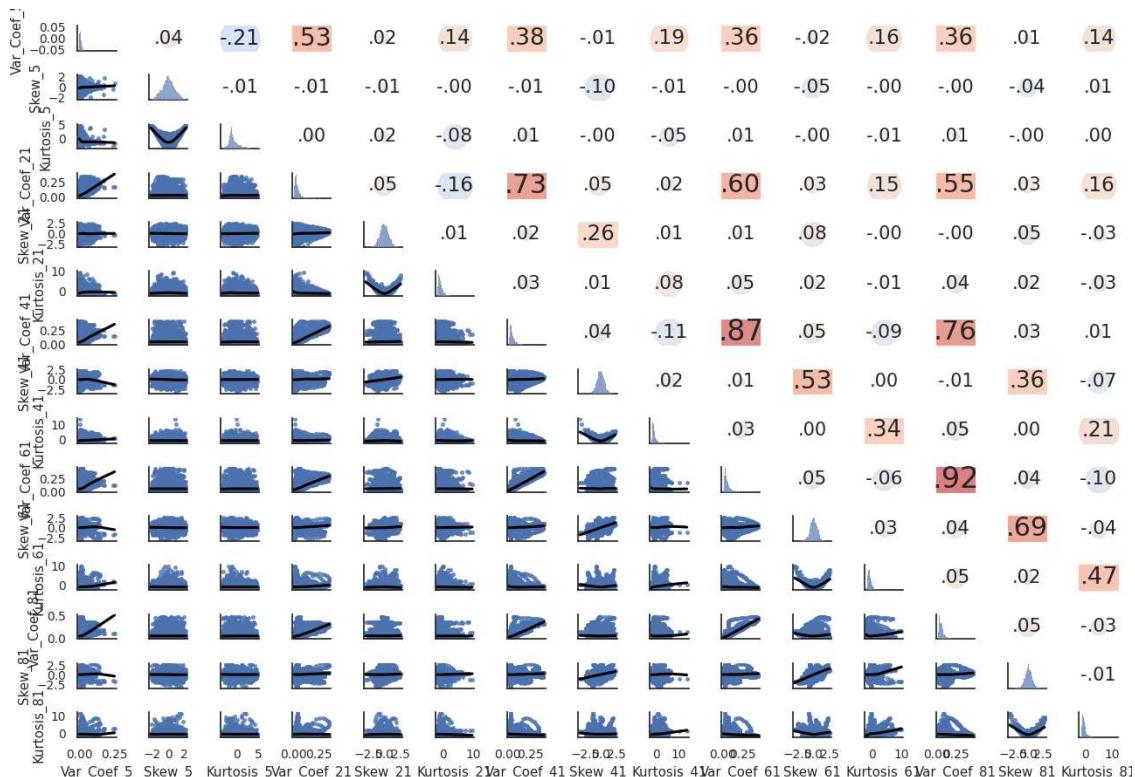


Figura 26 - Diagonal central - histogramas mostrando a distribuição de cada um dos atributos utilizados. Abaixo da diagonal central – gráficos de dispersão de pares de atributos. Acima da diagonal

central - Valores dos coeficientes de correlação de pares de atributos, indicando grande independência entre os atributos utilizados.

Após a confirmação da pequena correlação entre os atributos, treinou-se um modelo “*Random Forest*” no poço 1-BT-1-AM com 30% para treino 70% para teste. A variável alvo foi a unidade litoestratigráfica, cuja interpretação encontra-se no perfil composto, reclassificada como “Formação Solimões” ou “Outra”. O modelo treinado foi avaliado através das métricas de *Machine Learning* e da matriz de confusão, revelando boa previsibilidade (**Figuras 27 e 28**), e exportado em formato “.pkl” com auxílio do módulo “joblib”, para aplicação aos demais poços.

Formacao	outra	solimoes
outra	421	125
solimoes	6	3235

Figura 27. Matriz de confusão para o modelo “*Random Forest*” treinado no poço 1-BT-1-AM.

Formacao	Recall	Precision	F1	Accurac
y				
outra	0.771	0.986	0.865	0.965
solimoes	0.998	0.963	0.98	0.965

Figura 28. Tabela com métricas para o modelo “*Random Forest*” treinado no poço 1-BT-1-AM

Uma avaliação da importância relativa de cada atributo para o modelo revela que os coeficientes de variação das maiores janelas móveis (81 e 61 dados espaçados de 20 cm) têm relevância até maior que o valor absoluto de “Gamma Ray” (**Figura 29**).

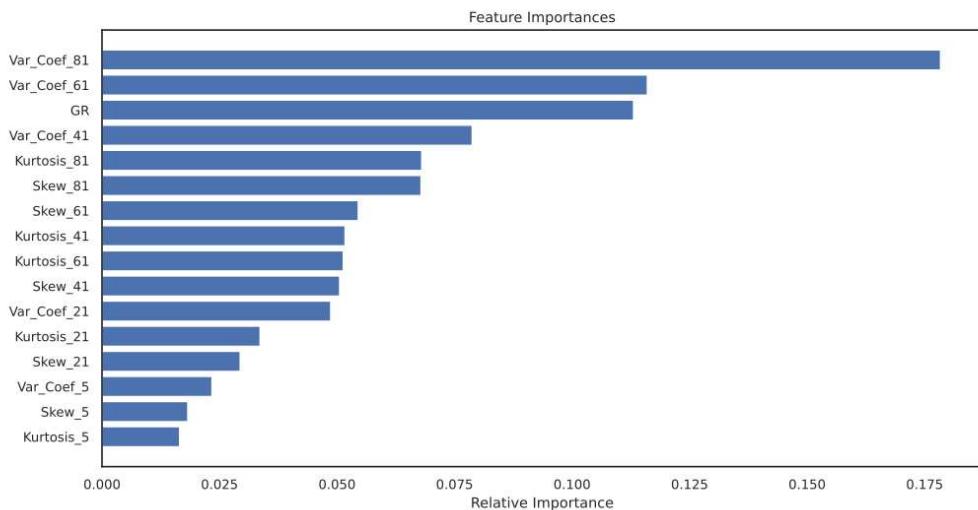


Figura 29. Importância relativa dos atributos do modelo.

6. RESULTADOS E DISCUSSÕES

Como o objetivo deste trabalho é o de encontrar a base da Formação Solimões, a classificação ponto-a-ponto gerada pelo *Random Forest* não é de grande utilidade. Alternativamente, optou-se por gerar uma curva contínua com a probabilidade de classificação (resultante do “*Random Forest*”) como Formação Solimões com a aplicação do modelo treinado (**Figura 30**) utilizando-se o código desenvolvido *Find_base.py*.

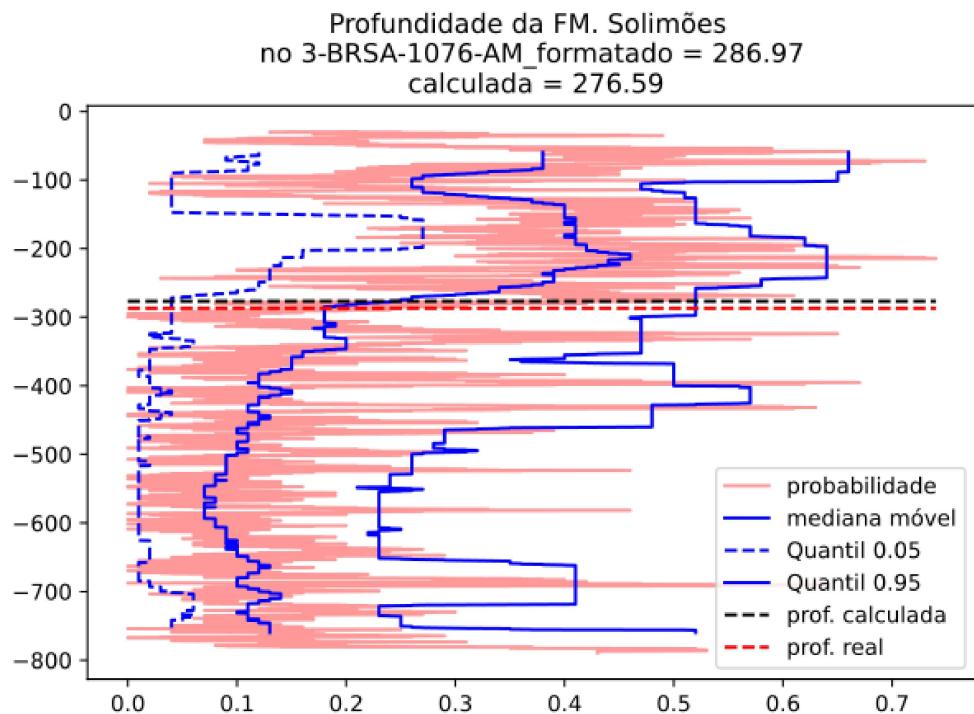


Figura 30 - Curva de probabilidade de classificação como Formação Solimões a diferentes profundidades, obtida pela aplicação do modelo treinado ao poço 3-BRSA-1076-AM.

Com base nessa curva de probabilidade, adotou-se um critério para separar conjuntos, posteriormente validado pela aplicação a diversos poços também previamente classificados em termos de unidades litoestratigráficas. Observou-se que não apenas a mediana como também os percentis 0.95 e 0.05 da curva de probabilidades respondem à classificação da Formação Solimões. Assim, criou-se uma nova curva com a média geométrica entre esses dois percentis (curva “Produto” na **Figura 31**). O critério para definição da base da Fm. Solimões é triplo: ser um trecho de queda da curva “Produto”, separar um conjunto com média geométrica dos percentis acima de 0.32, e ter profundidade maior que 150 m.

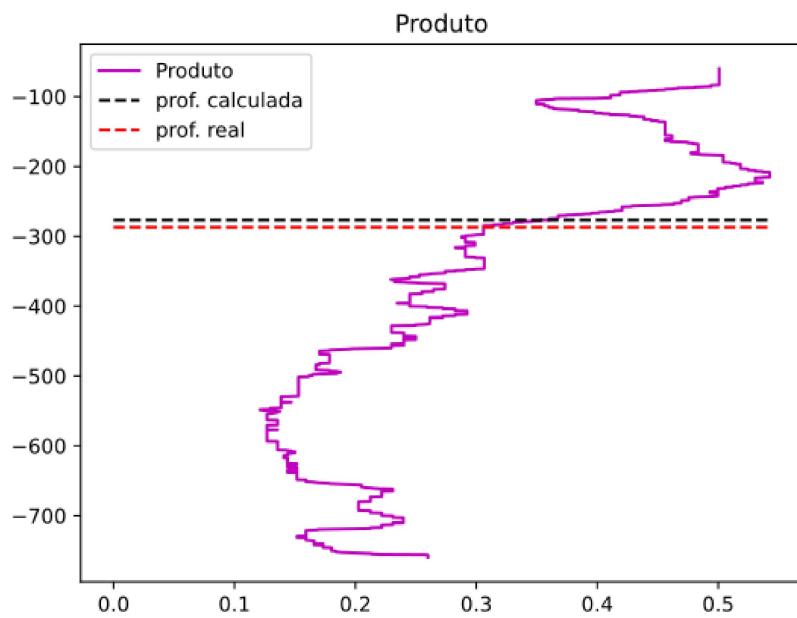


Figura 31 - Curva da média geométrica dos percentis 0.95 e 0.05 da curva de probabilidades.

7. INTERPRETAÇÃO E DISCUSSÃO DOS RESULTADOS

O método adotado permitiu a identificação da posição da base da Fm. Solimões com boa precisão, chegando-se a um erro padrão de 27.5 m em 22 poços com perfil composto interpretado (**Figura 32** e **Figura 33**).

Poço	Profundidade_	Porfundiade_
	_real	modelada
1-HRT-1-AM_formatado	327,00	305,60
1-HRT-3-AM_formatado	346,00	330,00
4-HRT-7D-AM_formatado	365,00	310,50
3-BRSA-1139-AM_formatado	289,97	286,22
4-BRSA-1203-AM_formatado	347,98	323,23
3-BRSA-1113D-AM_formatado	300,99	330,94
3-BRSA-515-AM_formatado	314,93	338,35
3-BRSA-1076-AM_formatado	286,97	276,59
4-BRSA-1252-AM_formatado	338,99	319,89
6-BRSA-24-AM_formatado	324,92	242,36
1-BRSA-55-AM_formatado	412,85	374,21
1-BRSA-714-AM_pronto	233,98	197,80
1-BRSA-738-AM_pronto	270,00	270,75
1-BRSA-835-AM_pronto	254,00	270,75
1-BRSA-984-AM_formatado	319,96	297,84
1-BRSA-985-AM_formatado	351,97	376,57
1-BRSA-1059-AM_formatado	313,94	297,84
1-BRSA-1189-AM_formatado	293,98	297,84
1-BRSA-1311-AM_formatado	309,98	298,62
1-BRSA-1311-AM_pronto	309,98	268,40
1-RA-2-AM_formatado	314,86	289,21
1-TAQ-1-AM_formatado	359,91	321,51

Figura 32 - Tabela de resultados da validação do modelo preditivo sobre a profundidade da base da Fm. Solimões em 22 poços satisfatoriamente modelados dentre os 33 poços selecionados com perfil composto interpretado.

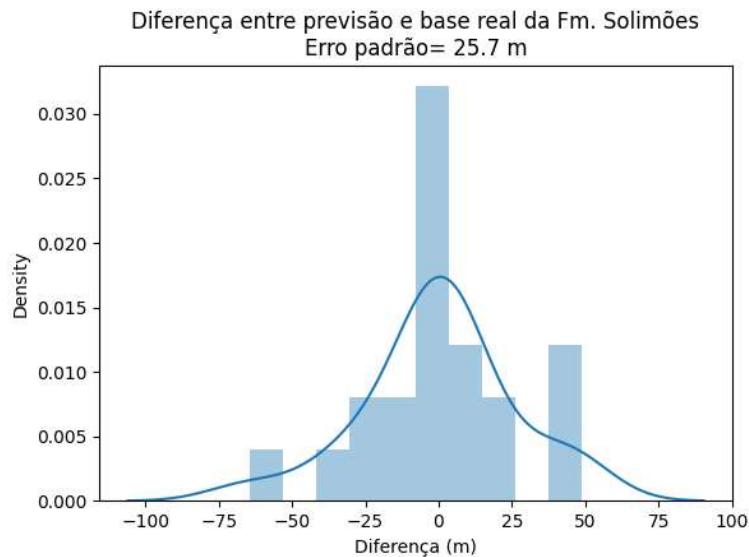


Figura 33 - Erro padrão da previsão do modelo em relação à profundidade interpretada em 22 poços.

Exemplos de resultados das previsões (Figuras 34 a 37):

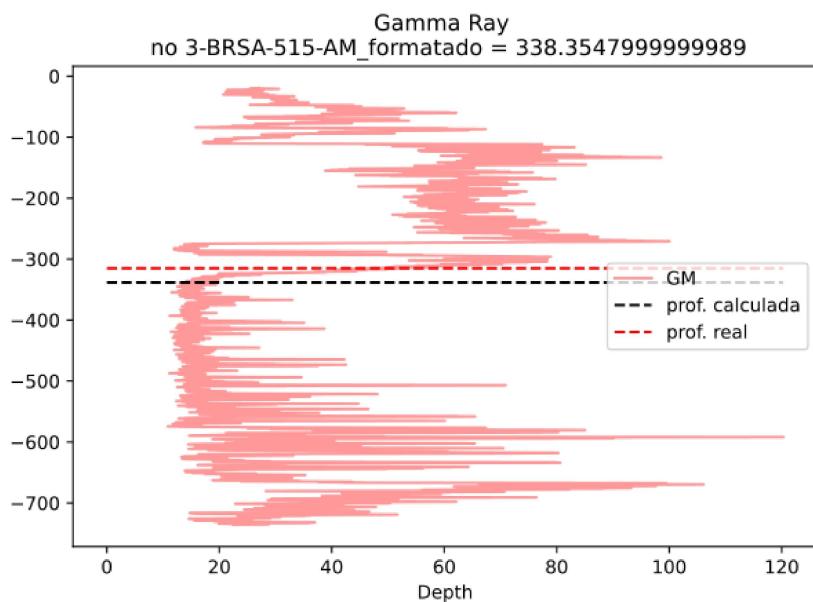


Figura 34 – Curva de probabilidade para profundidade estimada x profundidade real para o poço 3-BRSA-515-AM

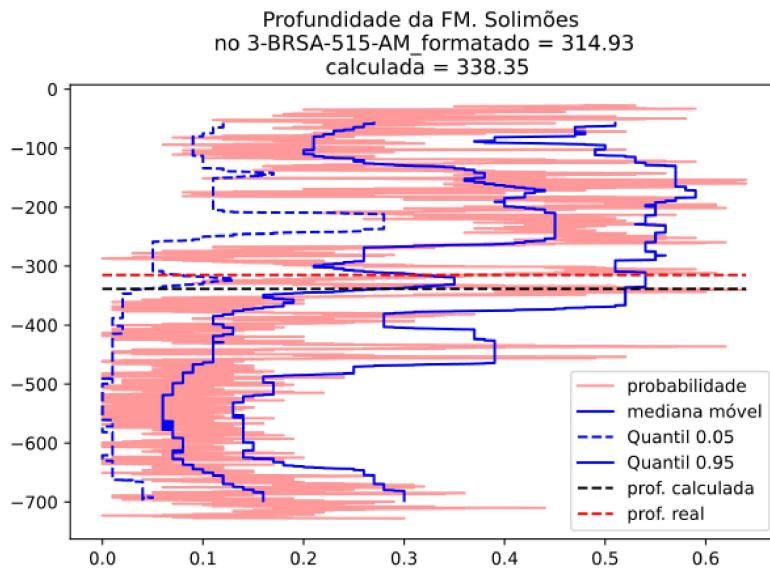


Figura 35 – Curva de probabilidade para profundidade estimada x profundidade real com a mediana móvel e quartis para o poço 3-BRSA-515-AM

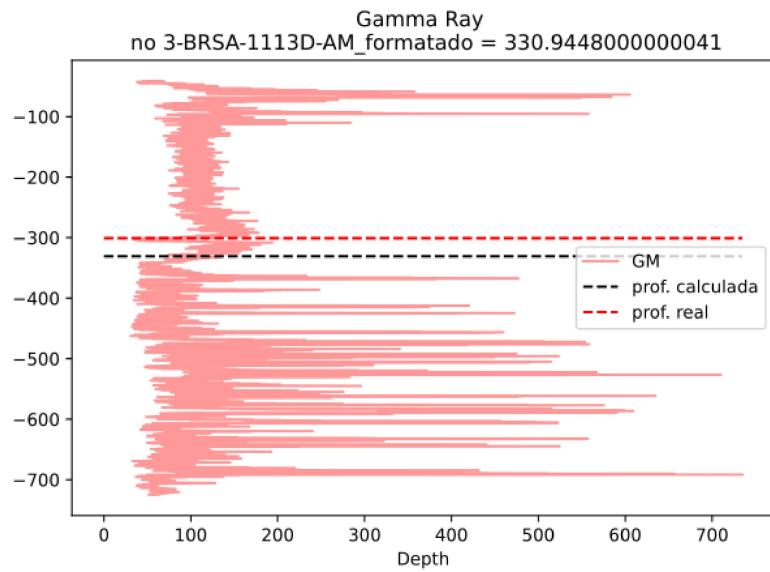


Figura 36 – Curva de probabilidade para profundidade estimada x profundidade real com a mediana móvel e quartis para o poço 3-BRSA-1113D-AM

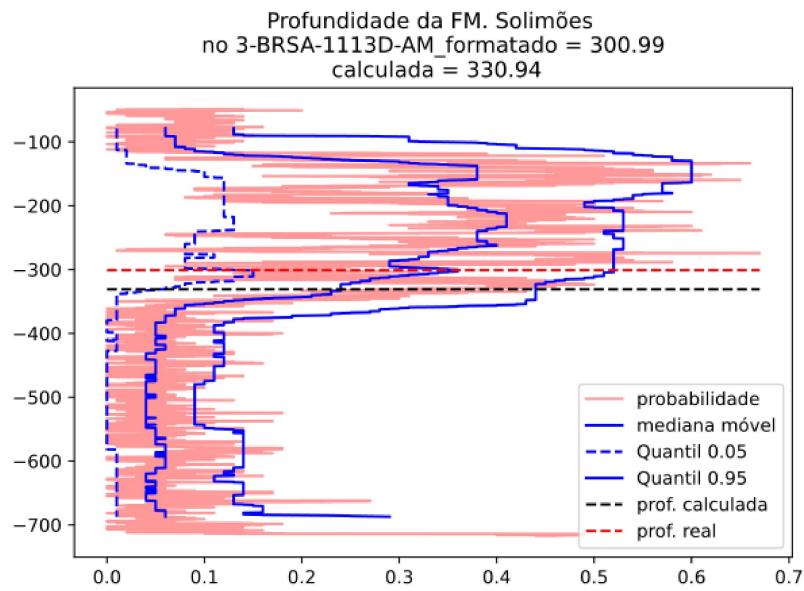


Figura 37. Curva de probabilidade para profundidade estimada x profundidade real com a mediana móvel e quartis para o poço 3-BRSA-1113D-AM.

A seguir apresenta-se os mapas com a localização dos poços utilizados (**Figura 38**) e os mapas classificados de acordo com a qualidade dos resultados, no segundo mapa mostrando os poços adequadamente modelados com valores estimados para a base da Fm Solimões com erro menor que 22 m (**Figura 39**) e no terceiro mapa os poços que não apresentaram resultados satisfatórios (**Figura 40**):

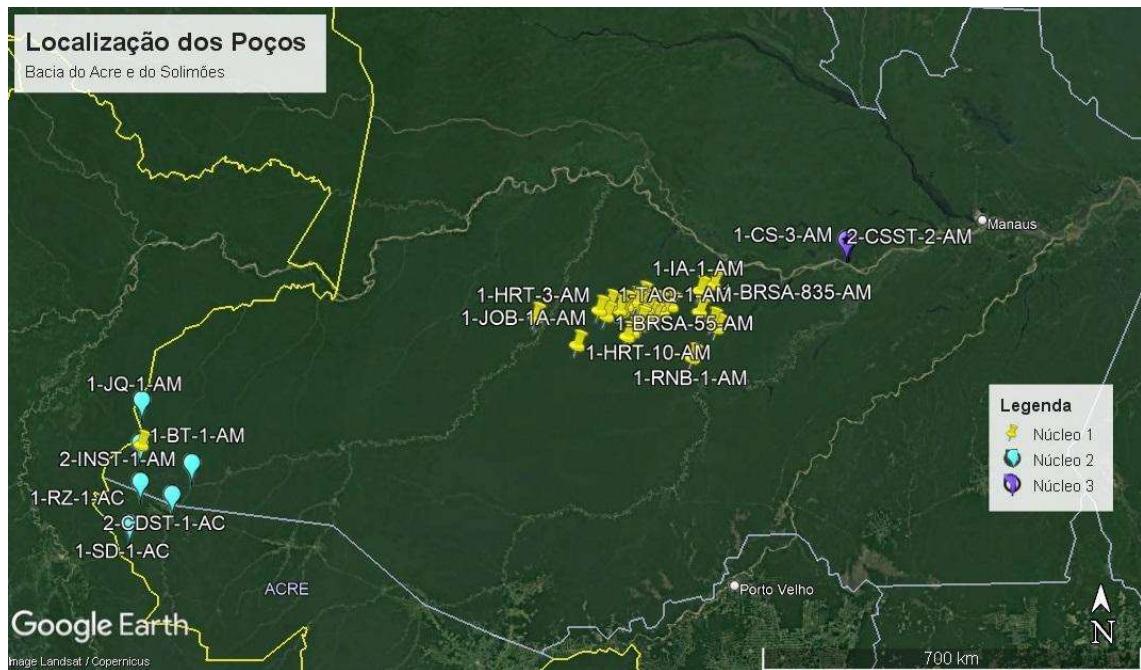


Figura 38. Mapa de localização com os 33 poços selecionados para base treino no *Machine Learning*.

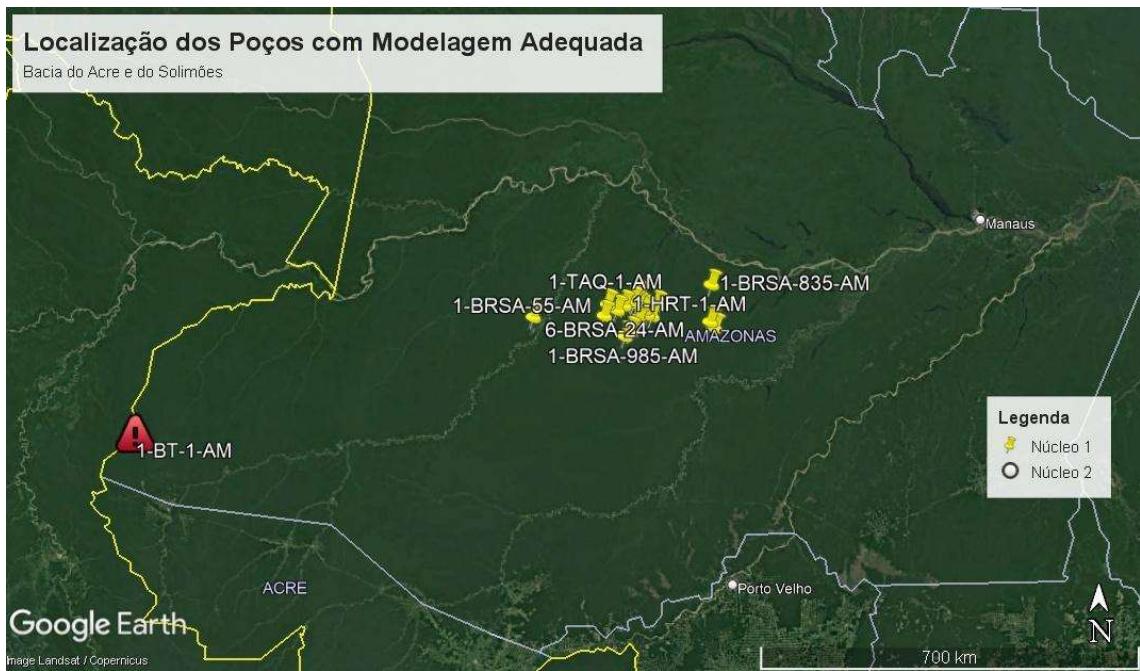


Figura 39: mapa de localização dos 22 poços que apresentaram uma modelagem para a base da Formação Solimões com erro inferior a 22m. Em destaque com símbolo em vermelho no mapa o poço modelo 1-BT-1-AM.



Figura 40: Mapa de localização dos outros poços que não apresentaram resultado satisfatório. Em destaque com símbolo em vermelho no mapa o poço modelo 1-BT-1-AM.

8. CONCLUSÕES

Observando-se os resultados obtidos, nota-se que a proposta foi bem eficiente: dos 34 poços utilizados como base para o projeto para delimitar a base da Formação Solimões, com o topo por padrão variando entre 6 e 8 m na totalidade dos poços, 22 poços obtiveram um resultado com erro inferior a 22 m, como demonstrado anteriormente.

A fonte desta diferença pode ser atribuída tanto pela estrutura das duas bacias, quanto por problemas na aquisição dos dados durante a perfilagem geofísica dos poços.

O poço 1-BT-1-AM mostrou-se eficiente como modelo para a maior parte dos poços, mas não para todos os poços selecionados para estudos, por isso mais testes e refinamentos precisarão ser feitos no futuro, inclusive para compreensão do efeito das distâncias nos resultados, e talvez a escolha de outros poços também seja uma alternativa para melhorar os resultados obtidos para os 11 poços em que o poço modelo não foi adequado e futuramente ser aplicado nos poços que realmente não apresentam informações sobre a base da Fm. Solimões para realização de mapa de isópacas.

Desta forma, a aplicação da metodologia detalhada neste projeto pode colaborar de forma robusta tanto na aquisição de novas informações para outros poços disponibilizados pela ANP em projetos futuros, quanto para propostas de construção de isópacas e correlações estratigráficas, de forma mais imparcial e automatizada para a modelagem de outras bacias sedimentares.

9. REFERÊNCIAS BIBLIOGRÁFICAS

ALEVATO, R.C. Bacia do Acre: Interpretação Sísmica e Integração com Dados de Métodos Potenciais. Trabalho Final de Graduação UERJ, 2019, 121 p.

BARATA, C.F.; CAPUTO, M.V. Geologia do Petróleo da Bacia do Solimões. O “Estado da Arte”. 4o PDPEIRO, Campinas, SP, 2007. p. 21-24.

BREIMAN, L. Random Forests. Machine Learning, v. 45, n. 1, p. 5–32, 2001.

BUSSAB, W.O.; MORETIN, P.A. Estatística Básica. 9^a Edição. Saraiva, 2017, 554 p.

CAVALCANTE, L.M. Aspectos geológicos do estado do Acre e implicações na evolução da paisagem. Embrapa Acre, 2006, 25 p.

CAVALLANTE, L.; SCHNEIDER, R.L. Reavaliação das possibilidades petrolíferas da Bacia do Alto Amazonas. Belém, PETROBRÁS/RENOR, 1976, 25 p.

CUNHA, P.R.C. Bacia do Acre. Boletim de Geociências da Petrobras, v. 15, p. 207-215, 2007.

EFRON, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.*, v. 7, n. 1, p. 1-26, 1979.

ESPURT, N.; BABY, P.; BRUSSET, S.; RODDAZ, M.; HERMOZA, W.; BARBARAND, J. The Nazca Ridge and uplift of the Fitzcarrald Arch: implications for regional geology in northern South America. In: HOORN, C.; WESSELINGH, F.P. (Eds.). *Amazonia: Landscape and Species Evolution: A Look into the Past*. Wiley Online Library, 2010, p. 89-100.

FEIJÓ, F.J.; SOUZA, R.G. Bacia do Acre. Boletim de Geociências da Petrobras, v. 8, n. 1, p. 9-16, 1994.

FLAMENT, N.; GURNIS, M.; MÜLLER, D.; BOWER, D.J.; HUSSON, L. Influence of subduction history on South American topography. *Earth Planet. Sci. Lett.*, v. 430, p. 9–18, 2015.

HOORN, C.; BOGOTÁ-A, G.R.; ROMERO-BAEZ, M.; LAMMERTSMA, E.I.; FLANTUA, S.G.A.; DANTAS, E.L.; DINO, R.; CARMO, D.A.; CHEMALE JR, F. The Amazon at sea: Onset and stages of the Amazon River from a marine Record, with special reference to Neogene plant turnover in the drainage basin. *Global and Planetary Change*.

HOORN, C.; WESSELINGH, P.; TER STEEGE, H.; BERMUDEZ, M.A.; MORA, SEVINK, J.; SANMARTÍN, I.; SANCHEZ-MESEGUER, A.; ANDERSON, C.L.; FIGUEIREDO, J.P.; JARAMILLO, C.; RIFF, D.; NEGRI, F.R.; HOOGHJEMSTRA, LUNDBERG, J.; STADLER, T.; SÄRKINEN, T.; ANTONELLI, A. Amazonia Through Time: Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. *Science*, v. 330, p. 927-931, 2010.

HORBE, A.M.C.; RODDAZ, M.; GOMES, L.B.; CASTRO, R.T.; DANTAS, E.L.; CARMO, D.A. Provenance of the Neogene sediments from the Solimões Formation

(Solimões and Acre Basins, Brazil. *Journal of South American Earth Science*, v. 93, p. 232-241, 2019.

KEAREY, P.; BROOKS, M.; HILL, I. *Geofísica de Exploração*. Oficina de Textos, 2002, p. 399-442.

LATRUBESSE, E.M.; BOCQUENTIN, J.C.; SANTOS, J.C.; RAMONELL, C. Paleoenvironmental Model for the Late Cenozoic of Southwestern Amazonia: Paleontology and Geology. *Acta Amazonica*, v. 27, n. 2, p. 103-118, 1997.

LOUREIRO, E.M.L.; TAVARES, B.L.; JANNUZI, J. Bacia do Solimões: Sumário Geológico e Setores em Oferta. Superintendência de Definição de Blocos. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Rodada Permanente de Licitações, 2019.

MAIA, R.G.N.; GODOY, H.K.; YAMAGUTI, H.S.; MOURA, P.A.; COSTA, F.S.F.; HOLANDA, M.A.; COSTA, J.A. Projeto Carvão no Alto Solimões: Relatório Final. Ministério de Minas e Energia, Departamento Nacional de Produção Mineral, Convênio DNPM-CPRM, 1977.

MEDEIROS, C.G. Ostracodes da Formação Solimões, Brasil: Contribuição à Bioestratigrafia do Néogeno da Amazônia. Dissertação de Mestrado. Instituto de Geociências, Universidade de Brasília: Brasília, 2017.

MIURA, K. Possibilidades petrolíferas da Bacia do Acre. In: SBG, Congresso Brasileiro de Geologia, 26, Anais de Resumo, p. 15-20, 1972.

NEGRI, F.R.; VILLANUEVA, J.; B.; FERIGOLO, J.; ANTOINE, P.O. A review of Tertiary mammal faunas and birds from western Amazonia. In: HOORN, C.; WESSELINGH, F.P. (Eds.). *Amazonia: Landscape and Species Evolution: A Look into the Past*. Wiley Online Library, 2010, p. 245-258.

NASCIMENTO JUNIOR, Edvaldo Nunes. O emprego das Companhias e Pelotões Especiais de Fronteira do Comando de Fronteira Acre na defesa da Amazônia no contexto dos crimes transfronteiriços na atualidade. Trabalho de Conclusão de Curso (Curso de Especialização em Ciências Militares, com ênfase em Defesa) - Escola de

Comando e Estado-Maior do Exército, Escola Marechal Castello Branco, Rio de Janeiro, RJ, 2022.

NEVES, C.A.O.; CAMPOS, J.N.P.; TEIXEIRA, L.B.; RODRIGUES, R.; TRIGUIS, J.A. Integração geológica e exploratória da Bacia do Amazonas. Manaus, PETROBRÁS/DENOC, 1989, 113 p.

OLIVEIRA, C.M.M. Estilos estruturais e evolução tectônica da bacia do Acre. Dissertação (Mestrado em Evolução Crustal e Recursos Naturais) – Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto, 1994.

PERUPETRO. Public maps and drilling information delivered in digital format by Perupetro, 2013.

RAJA GABAGLIA, G.P.; FIGUEIREDO, A.M.F. Evolução dos conceitos acerca das classificações de bacias sedimentares. In: RAJA GABAGLIA, G.P.; MILANI, E.J. (Ed.). Origem e evolução de bacias sedimentares. Rio de Janeiro: Petrobras. SEREC. CEN-SUD, 1991, p. 31-45.

SÁ, D.M.; SÁ, M.R.; LIMA, N.T. Telégrafos e inventário do território no Brasil: as atividades científicas da Comissão Rondon (1907-1915). História, Ciências, Saúde – Manguinhos, v. 15, n. 3, p. 779-810, 2008.

SHEPHARD, G.; MÜLLER, D.; LIU, L.; GURNIS, M. Miocene drainage reversal of the Amazon River driven by plate-mantle interaction. *Nat. Geosci.*, v. 3, p. 870–875, 2010.

SILVA, O.B. Análise da Bacia do Solimões (Revisão estratigráfica, magmatismo e geoquímica). Universidade Federal de Ouro Preto, Ouro Preto, Dissertação de Mestrado, 1987.

SZATMARI, P.; SIMÕES, I.A.; CARVALHO, R.S. Evaporitos da Bacia do Amazonas. Rio de Janeiro, PETROBRÁS, 1975, 128 p.

VASCONCELLOS, R.V.A.; COUTINHO, C.I. Bacia do Acre. Brasil 12^a Rodada Licitacões de Petróleo e Gás, ANP 15 Anos. Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP), 2013.

WALCOTT, R.I. Flexural rigidity, thickness, and viscosity of the lithosphere. *J. Geophys. Res.*, v. 75, p. 3941-3954, 1970.

WANDERLEY-FILHO, J.R.; EIRAS, J.F.; CUNHA, P.R. da C. The Paleozoic Solimoes and Amazonas basins and the Acre foreland basin of Brazil. In: HOORN, C.; WESSELINGH, F.P.; VONHOF, H.B.; KROONENBERG, S.B., HOOGHIEMSTRA, H. (Ed.). *Amazonia, landscape and species evolution: a look into the past*, 2010, p. 29-37.

Sites:

IBM - Random Forest:

IBM. Random Forest.

Disponível em: <https://www.ibm.com/topics/random-forest>. Acesso em: 25/11/2023.

ANP - Agência Nacional do Petróleo:

Agência Nacional do Petróleo. Concessão de Blocos Exploratórios - 4^a Rodada de Licitações de Blocos.

Disponível em:
<https://www.gov.br/anp/pt-br/rodadas-anp/rodadas-concluidas/concessao-de-blocos-exploratorios/4a-rodada-licitacoes-blocos/bacias>. Acesso em: 08/06/2023.

Breiman, L.:

BREIMAN, L. Breiman's Random Forests. Disponível em:

<https://www.stat.berkeley.edu/~breiman/RandomForests/>. Acesso em: 28/11/2023

ANEXOS

ANEXO I – CÓDIGO POÇO MODELO

```

import pandas as pd
from pandas import DataFrame
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.mlab as mlab
import seaborn as sns
import scipy
from scipy import stats
from scipy import stats
import sklearn
from numpy import unique
from numpy import where
from sklearn.datasets import make_classification
from sklearn.mixture import GaussianMixture
# import skbio.stats.composition as comp
import plotly.express as px
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn import svm # SVC para classificação (target discreto) e SVR para
regressão (target contínuo)
from sklearn.ensemble import RandomForestClassifier # só classificação
from sklearn import metrics
import joblib # para exportar o modelo treinado

#import c as gs

#df_original=pd.read_csv('1-RZ-1-AC_pronto.csv')
#df_original=pd.read_csv('1-BRSA-985-AM_formatado.csv')
#df_original=pd.read_csv('4-BRSA-1252-AM_formatado.csv') # Modelo Amazonas
#df_original=pd.read_csv('2-CDST-1-AC_pronto.csv')
#df_original=pd.read_csv('1-JOB-1A-AM_formatado.csv')
df_original=pd.read_csv('1-BT-1-AM_formatado.csv') # modelo AC

#só o gama
df=pd.DataFrame({'GR':df_original.GR, 'Formacao':df_original.Formacao})#
df.Formacao.loc[df.Formacao!='solimoes']='pre_solimoes'
df=df.dropna()
#df.GR=df.GR/df.GR.mean()

plt.figure('Histogramas')
sns.histplot(df.GR.loc[df.Formacao=='solimoes'],label='Solimoes')
sns.histplot(df.GR.loc[df.Formacao=='not_solimoes'],label='Other')
plt.legend()
plt.savefig('Histograma_Modelo_Acre2.svg')

#df['first_derivative']=df.groupby('Formacao')['GR'].diff().fillna(0)
#df['second_derivative']=df.groupby('Formacao')['first_derivative'].diff().fillna(0)
#df['third_derivative']=df.groupby('Formacao')['second_derivative'].diff().fillna(0)
#df.GR=df.GR/df.GR.mean()
#df['mean11']=df.GR.rolling(11).mean()
df['Var_Coef_5']=df.GR.rolling(5,center=True).std()/df.GR.rolling(5,center=True)

```

```

.mean()
df['Skew_5']=df.GR.rolling(5,center=True).skew()
df['Kurtosis_5']=df.GR.rolling(5).kurt()

df['Var_Coef_21']=df.GR.rolling(21,center=True).std()/df.GR.rolling(21,center=True).mean()
df['Skew_21']=df.GR.rolling(21).skew()
df['Kurtosis_21']=df.GR.rolling(21).kurt()

df['Var_Coef_41']=df.GR.rolling(41,center=True).std()/df.GR.rolling(41,center=True).mean()
df['Skew_41']=df.GR.rolling(41,center=True).skew()
df['Kurtosis_41']=df.GR.rolling(41,center=True).kurt()

df['Var_Coef_61']=df.GR.rolling(61,center=True).std()/df.GR.rolling(61,center=True).mean()
df['Skew_61']=df.GR.rolling(61,center=True).skew()
df['Kurtosis_61']=df.GR.rolling(61,center=True).kurt()

df['Var_Coef_81']=df.GR.rolling(81,center=True).std()/df.GR.rolling(81,center=True).mean()
df['Skew_81']=df.GR.rolling(81,center=True).skew()
df['Kurtosis_81']=df.GR.rolling(81,center=True).kurt()

#df=df.drop(columns=['GR'])

df=df.dropna()

def corrdot(*args, **kwargs):
    corr_r = args[0].corr(args[1], 'pearson')
    corr_text = f"{corr_r:2.2f}".replace("0.", ".")
    ax = plt.gca()
    ax.set_axis_off()
    marker_size = abs(corr_r) * 5000
    ax.scatter([.5], [.5], marker_size, [corr_r], alpha=0.6, cmap="coolwarm",
               vmin=-1, vmax=1, transform=ax.transAxes)
    font_size = abs(corr_r) * 10 + 15
    ax.annotate(corr_text, [.5, .5], xycoords="axes fraction",
                ha='center', va='center', fontsize=font_size)

def model_analysis(df,target):
    print("Data Types:")
    print(df.dtypes)
    print("Rows and Columns:")
    print(df.shape)
    print("Column Names:")
    print(df.columns)
    print("Null Values:")
    print(df.apply(lambda x: sum(x.isnull()) / len(df)))

```

```

print(df.describe())

sns.set(style='white', font_scale=0.9)
g = sns.PairGrid(df, aspect=.3, diag_sharey=False)
g.fig.set_size_inches(13,9)
g.map_lower(sns.regplot, lowess=True, ci=False, line_kws={'color':
'black'}, scatter_kws={'s':7, 'alpha':0.7})
g.map_diag(sns.histplot, kde_kws={'color': 'black'})
g.map_upper(corrplot)
plt.tight_layout()
plt.savefig('Corrdot_Acre2.png')

print('AQUI',df[target].dtype)
if df[target].dtype=='object' or df[target].dtype=='category':


    p=sns.pairplot(df,
hue=target,corner=True,kind='reg',plot_kws={'line_kws':{'lw':2}, 'scatter_kws':
{'alpha': 0.5, 's': 7}})
    p.fig.set_size_inches(13,9)
    plt.tight_layout()
    plt.savefig('Scatter_pairs.png')

else:


    correlations = df.corrwith(df[target]).iloc[:-1].to_frame()
    correlations['abs'] = correlations[0].abs()
    sorted_correlations = correlations.sort_values('abs',
ascending=False)[0]
    fig, ax = plt.subplots(figsize=(10,20))
    sns.heatmap(sorted_correlations.to_frame(), cmap='coolwarm',
    annot=True, vmin=-1, vmax=1, ax=ax);
    plt.savefig('Corr_target_Acre2.png')


model_analysis(df,target='Formacao')

#Montando o modelo
#df.target=df.Sinuosity_categories
df.target=df.Formacao
#print(df.target)
print('Target',df.Formacao.unique()) #
#df.data=df.drop(columns=['Sinuosity_categories']) #NChannel_categories
dfData=df.drop(columns=['Formacao']) # tirando a coluna de variáveis categóricas
#print(df.data)
pd.set_option("display.max_rows", None, "display.max_columns", None)
print('ESTE',dfData.columns)

# Normalizar
# para escolher qual método
https://stackoverflow.com/questions/30918781/right-function-for-normalizing-input-data

```

```

t-of-sklearn-svm

#sklearn.preprocessing.scale(X) # standard scaler - assume distribuições normais
#e normaliza com média 0 e std 1

#sklearn.preprocessing.normalize(X, axis=0) # normalizer - mantém os ângulos no
#espaço multidimensional mas não as magnitudes dentro de cada variável

#sklearn.preprocessing.MinMaxScaler().fit_transform(X) - transforma cada
#variável em valores entre 0 e 1, sensível a outliers.

min_max_scaler = sklearn.preprocessing.MinMaxScaler()

def scaleColumns(dfDat, cols_to_scale):

    n_test = dfDat[cols_to_scale]
    # print('Cols',n_test)
    cols_to_norm = cols_to_scale
    x = n_test.values
    min_max_scaler = sklearn.preprocessing.MinMaxScaler()
    x_scaled = min_max_scaler.fit_transform(x)
    n_test = pd.DataFrame(x_scaled, columns=cols_to_norm)
    # l_test = dfDat.drop(cols_to_norm, axis=1)
    # df_out = pd.concat([n_test, l_test], axis=1)
    df_out=n_test
    return(df_out)

dfData=scaleColumns(dfData,dfData.columns)
#print(df.data)

    #criar dataset train e target

X_train, X_test, y_train, y_test = train_test_split(dfData, df.target,
test_size=0.3,random_state=130) # 70% training and 30% test

#print(X_train)
#print(y_train)

#Create a Gaussian Classifier
clf=RandomForestClassifier(n_estimators=100)

#Train the model using the training sets y_pred=clf.predict(X_test)
clf.fit(X_train,y_train)

y_pred=clf.predict(X_test)
#print(y_pred)

print("\nRandom Forest")

```

```

df_confusion = pd.crosstab(y_test, y_pred)
print(df_confusion)
df_confusion.to_csv('confusion_janelas_moveis_AC2.csv')

print('\n')

# Model Accuracy: how often is the classifier correct?
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

df_metrics =
pd.DataFrame({'Formacao':np.unique(y_pred),"Recall":metrics.recall_score(y_test,
y_pred,average=None),"Precision":metrics.precision_score(y_test,
y_pred,average=None),"F1":metrics.f1_score(y_test,
y_pred,average=None),"Accuracy":metrics.accuracy_score(y_test, y_pred)})
print(df_metrics)
df_metrics.to_csv('metrics_janelas_moveis_AC2.csv')

print('\nFeature importance')

feature_imp =
pd.Series(clf.feature_importances_,index=dfData.columns).sort_values(ascending=F
alse)
print('\n',feature_imp)

importances = clf.feature_importances_
indices = np.argsort(importances)
features = X_train.columns
plt.figure('Importances',figsize=(12,6))
plt.title('Feature Importances')
plt.barh(range(len(indices)), importances[indices], color='b', align='center')
plt.yticks(range(len(indices)), [features[i] for i in indices])
plt.xlabel('Relative Importance')
plt.savefig('importances_janelas_moveis_AC2.svg')

#exportando o modelo terminado
joblib.dump(clf, 'AC_model.pkl')

```

ANEXO II – CÓDIGO FIND_BASE

```
import pandas as pd
from pandas import DataFrame
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.mlab as mlab
import seaborn as sns
import scipy
from scipy import stats
from scipy import stats
import sklearn
from numpy import unique
from numpy import where
from sklearn.datasets import make_classification
from sklearn.mixture import GaussianMixture
# import skbio.stats.composition as comp
import plotly.express as px
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn import svm # SVC para classificação (target discreto) e SVR para regressão (target contínuo)
from sklearn.ensemble import RandomForestClassifier # só classificação
from sklearn import metrics
import joblib # para exportar o modelo treinado
import ruptures as rpt
from scipy import interpolate
import csv
```

✓ import skbio.stats.composition as comp

```
import plotly.express as px
from imblearn.over_sampling import SMOTE
from sklearn import svm # SVC para classificação (target discreto) e SVR para regressão (target contínuo)
from sklearn.ensemble import RandomForestClassifier # só classificação
from sklearn import metrics
import joblib # para exportar o modelo treinado
import ruptures as rpt
from scipy import interpolate
import csv
```

```
def AchaBase(mediana,quant95,quant05):
```

```
import pandas as pd
from pandas import DataFrame
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.mlab as mlab
import seaborn as sns
import scipy
from scipy import stats
from scipy import stats
import sklearn
from numpy import unique
from numpy import where
from sklearn.datasets import make_classification
from sklearn.mixture import GaussianMixture
# import skbio.stats.composition as comp
import plotly.express as px
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn import svm # SVC para classificação (target discreto) e SVR para regressão (target contínuo)
from sklearn.ensemble import RandomForestClassifier # só classificação
from sklearn import metrics
import joblib # para exportar o modelo treinado
import ruptures as rpt
from scipy import interpolate
import csv
```

```
#def AchaBase(mediana,quant95,quant05):
```

```
#Aplicando a um poço
```

```
#Bacia do Acre
#nome_poco='1-JQ-1-AM_pronto' # 2.6
#nome_poco='1-RZ-1-AC_formatado' # erro
#nome_poco='1-BT-1-AM_formatado' # erro
#nome_poco='1-BT-2-AM_formatado' # erro
#nome_poco='2-CDST-1-AC_formatado' #870
#nome_poco='2-CSST-2AM_formatado'
#nome_poco='2-INST-1-AM_pronto' #erro
```

```
#Sem localização
#nome_poco='1-RNB-1AM_formatado'
#nome_poco='1-JOB-1A-AM_formatado' #159
nome_poco='1-IA-1-AM_formatado' #293
```

```
#Bacia Solimões
```

```
#nome_poco='1-HRT-1-AM_formatado' # -4
#nome_poco='1-HRT-3-AM_formatado' # 15
#nome_poco='4-HRT-7D-AM_formatado' #247
```

```

#nome_poco='3-BRSA-1139-AM_formatado' #3.8
#nome_poco='4-BRSA-1203-AM_formatado' #-3.8
#nome_poco='3-BRSA-1113D-AM_formatado' #-24
#nome_poco='3-BRSA-515-AM_formatado'# -17
#nome_poco='3-BRSA-1076-AM_formatado'
#nome_poco='4-BRSA-1252-AM_formatado'
#nome_poco='6-BRSA-24-AM_formatado'

#Sem localização bo kmz

#nome_poco='1-BRSA-55-AM_formatado' # 59.8
#nome_poco='1-BRSA-714-AM_pronto' #-251
#nome_poco='1-BRSA-738-AM_pronto' #8.8
#nome_poco='1-BRSA-835-AM_pronto' #-7.1
#nome_poco='1-BRSA-984-AM_formatado' # 38
#nome_poco='1-BRSA-985-AM_formatado' #-121
#nome_poco='1-BRSA-1059-AM_formatado' #32
#nome_poco='1-BRSA-1189-AM_formatado' #12
#nome_poco='1-BRSA-1311-AM_formatado' #28
#nome_poco='1-BRSA-1311-AM_pronto' #58
#nome_poco='1-RA-2-AM_formatado' # # -206
#nome_poco='1-TAQ-1-AM_formatado'

#Modelo

clf = joblib.load('AM_model.pkl') # AM_model.pkl para bacia do Amazonas ou AC_model.pkl para bacia do Acre
#clf = joblib.load('AC_model.pkl') # AM_model.pkl para bacia do Amazonas ou AC_model.pkl para bacia do Acre
#clf = joblib.load('AC2_model.pkl') # AM_model.pkl para bacia do Amazonas ou AC_model.pkl para bacia do Acre
#clf = joblib.load('JOB_model.pkl') # AM_model.pkl para bacia do Amazonas ou AC_model.pkl para bacia do Acre

df_original=pd.read_csv(nome_poco+'.csv')
Profundidade_Solimoes= df_original.DEPTH[df_original.Formacao=='solimoes']
Profundidade_Solimoes=Profundidade_Solimoes.iloc[-1]

#interpolando para 0.2 m

f = interpolate.interp1d(df_original.DEPTH,df_original.GR)
xnew = np.arange(min(df_original.DEPTH), max(df_original.DEPTH), 0.2)
ynew=f(xnew)

df=pd.DataFrame({'DEPTH':xnew,'GR':ynew})#
df=df.dropna()

plt.figure('Histogramas')
sns.histplot(df.GR,label='Todo o poço')
plt.legend()
plt.savefig('histograma_'+nome_poco+'.svg')

df['Var_Coef_5']=df.GR.rolling(5,center=True).std()/df.GR.rolling(5,center=True).mean()
df['Skew_5']=df.GR.rolling(5,center=True).skew()
df['Kurtosis_5']=df.GR.rolling(5,center=True).kurt()

df['Var_Coef_21']=df.GR.rolling(21,center=True).std()/df.GR.rolling(21,center=True).mean()
df['Skew_21']=df.GR.rolling(21,center=True).skew()
df['Kurtosis_21']=df.GR.rolling(21,center=True).kurt()

df['Var_Coef_41']=df.GR.rolling(41,center=True).std()/df.GR.rolling(41,center=True).mean()
df['Skew_41']=df.GR.rolling(41,center=True).skew()
df['Kurtosis_41']=df.GR.rolling(41,center=True).kurt()

df['Var_Coef_61']=df.GR.rolling(61,center=True).std()/df.GR.rolling(61,center=True).mean()
df['Skew_61']=df.GR.rolling(61,center=True).skew()
df['Kurtosis_61']=df.GR.rolling(61,center=True).kurt()

df['Var_Coef_81']=df.GR.rolling(81,center=True).std()/df.GR.rolling(81,center=True).mean()
df['Skew_81']=df.GR.rolling(81,center=True).skew()
df['Kurtosis_81']=df.GR.rolling(81,center=True).kurt()

```

```

df=df.dropna()
profundidades=df.DEPTH
df=df.drop(columns=['DEPTH'])
#df=df.drop(columns=['GR'])

y_pred=clf.predict(df)
pred_probab = clf.predict_proba(df)

df['Previsao']=y_pred
df['Probabilidade']=pred_probab[:,1]
df['DEPTH']=profundidades

df.to_csv(nome_poco+'_previsto.csv')

#achando o ponto que marca a maior diferença entre a probabilidade dos dois conjuntos

...
prob=np.array(df.Probabilidade)
algo = rpt.Pelt(model="l2", min_size=int(len(df.DEPTH)/8)) # esta função identifica intervalos em séries temporais
algo.fit(prob)
result = algo.predict(pen=1)
plt.savefig('Pelt_'+nome_poco+'.svg')

rpt.display(prob, [], result)
...

...
GR=np.array(df_original.GR)
algo = rpt.Pelt(model="l2", min_size=int(len(df_original.DEPTH)/18)) #acha limites possíveis na sequência de probabilidade
algo.fit(GR)
result2 = algo.predict(pen=1)

rpt.display(GR, [], result2)
...
...
cont=0
RASOS=[]
FUNDOS=[]
diferença_médias=[]
conts=[]
while cont<len(result):

    RASO=df.Probabilidade.iloc[:result[cont]].mean()
    FUNDO=df.Probabilidade.iloc[result[cont]:].mean()
    RASOS=np.append(RASOS,RASO)
    FUNDOS=np.append(FUNDOS,FUNDO)
    diferença_médias=RASOS-FUNDOS
    conts=np.append(conts,cont)
    cont+=1
...
...
Cont_limite=conts[(RASOS>0.5)&(FUNDOS<0.5)]
Diffs_limite=diferença_médias[(RASOS>0.5)&(FUNDOS<0.5)]

Limite1=df.DEPTH.iloc[Cont_limite[Diffs_limite==max(Diffs_limite)]]]

...
...
res2=np.array(result)
Segmentos=[]
Grupos=[]
ini=0
cont=0
cont2=0
Segmento_Anterior=0.

while cont2<len(res2):
    Este_Segmento=df.Probabilidade.iloc[ini:res2[cont2]].median()
    print(cont,Este_Segmento)
    ...

```

```

Este Grupo=df.Probabilidade.iloc[:res2[cont2]].median()
print('Este Grupo',Este Grupo)
print('Segmento_Anterior',Segmento_Anterior)
if Este_Segmento-Segmento_Anterior>-0.2 : # o segmento tem que ser ter probabilidade maior que 0.5 e não pode ser muit
    ini=res2[cont]
    cont+=1
Segmento_Anterior=Este_Segmento
cont2+=1

Limite2=df.DEPTH[res2[cont-1]]
...
#cumulativas
window=301
roll_median=df.Probabilidade.rolling(window,center=True).median()
roll_quant05=df.Probabilidade.rolling(window,center=True).quantile(.05)
roll_quant95=df.Probabilidade.rolling(window,center=True).quantile(.95)

cum_median=roll_median.cumsum()
cum_quant05=roll_quant05.cumsum()
cum_quant95=roll_quant95.cumsum()

produto=(roll_median*roll_quant95)**(1/2)

Calculada=0
cont=0
while cont<len(produto-1) and Calculada==0:
    if produto.iloc[cont]>produto.iloc[cont+1] and produto.iloc[cont]<0.32 and df.DEPTH[cont]>150.:
        Calculada=df.DEPTH[cont]
    else:
        cont+=1
    print(cont)

plt.figure('Produto')
plt.title(f'Produto')
plt.plot(produto,-df.DEPTH,'m-',label='Produto')
# plt.plot([produto.quantile(0.54),produto.quantile(0.54)],[-min(df.DEPTH),-max(df.DEPTH)],'k--')
plt.plot([0.,produto.max()],[-Calculada,-Calculada],'k--',label='prof. calculada')
plt.plot([0.,produto.max()],[-Profundidade_Solimoes,-Profundidade_Solimoes],'r--',label='prof. real')
plt.legend()
plt.savefig(nome_poco+ '_produto.svg')

...
plt.figure('Cum_Probabil')
plt.title(f'CUM')
plt.plot(cum_median,-df.DEPTH,'b-',label='mediana móvel')
plt.plot(cum_quant05,-df.DEPTH,'r--',label='Quantil 0.05')
plt.plot(cum_quant95,-df.DEPTH,'r-',label='Quantil 0.95')
...


plt.figure('Probabil')
plt.title(f'Profundidade da FM. Solimões \n no {nome_poco} = {np.round(Profundidade_Solimoes,2)} \n calculada = {np.rou
plt.plot(df.Probabilidade,-df.DEPTH,'r-',alpha=0.4,label='probabilidade')
plt.plot(roll_median,-df.DEPTH,'b-',label='mediana móvel')
plt.plot(roll_quant05,-df.DEPTH,'b--',label='Quantil 0.05')
plt.plot(roll_quant95,-df.DEPTH,'b-',label='Quantil 0.95')
plt.plot([0.,df.Probabilidade.max()],[-Calculada,-Calculada],'k--',label='prof. calculada')
plt.plot([0.,df.Probabilidade.max()],[-Profundidade_Solimoes,-Profundidade_Solimoes],'r--',label='prof. real')
plt.legend()
plt.tight_layout()
plt.savefig(nome_poco+ '_probabil.svg')

...
cont_plot=0
while cont_plot<len(res2):
    Limite=df.DEPTH[res2[cont_plot]]
    print('Limite',Limite)
    plt.plot([0,1],[-Limite,-Limite],'r--',alpha=0.3)
    cont_plot+=1

```

```
#plt.plot([0,1],[-Limite1,-Limite1],'r--')
plt.plot([0,1],[-limite2,-Limite2],'r--')
#plt.plot(df_original.GR,df_original.DEPTH,'g-',label='GR')
plt.legend()
plt.xlabel('Depth')
plt.savefig('Probabilidade_'+nome_poco+'.svg')
'''

plt.figure('Gamma Ray')
plt.title(f'Gamma Ray \n no {nome_poco} = {Calculada}')
plt.plot(df_original.GR,-df_original.DEPTH,'r-',alpha=0.4,label='GM')
#plt.plot(df_original.GR.rolling(201).median(),-df_original.DEPTH,'b-',label='mediana móvel')
#plt.plot(df_original.GR.rolling(201,center=True).quantile(.05),-df_original.DEPTH,'b--',label='Quantil 0.05')
#plt.plot(df_original.GR.rolling(201,center=True).quantile(.95),-df_original.DEPTH,'b-',label='Quantil 0.95')
plt.plot([0.,df_original.GR.max()],[-Calculada,-Calculada],'k--',label='prof. calculada')
plt.plot([0.,df_original.GR.max()],[-Profundidade_Solimoes,-Profundidade_Solimoes],'r--',label='prof. real')

#plt.plot([0.5,0.5],[-min(df.DEPTH),-max(df.DEPTH)],'k--')
#plt.plot([0,1],[-Limite1,-Limite1],'r--')
#plt.plot([0,1],[-Limite2,-Limite2],'r--')
# plt.plot(df_original.GR,df_original.DEPTH,'g-',label='GR')
plt.legend()
plt.xlabel('Depth')
plt.tight_layout()
plt.savefig(nome_poco+'_gamma.svg')

#res2=res2[:-1]
#print(df.DEPTH[res2])
print('Profundidade_Solimoes',Profundidade_Solimoes)
print('Calculada',Calculada)
print('Erro',Profundidade_Solimoes-Calculada)

fields=[nome_poco,Profundidade_Solimoes,Calculada]
with open(r'Resultados_AC_model.csv', 'a') as f:
    writer = csv.writer(f)
    writer.writerow(fields)

plt.show()
```