Universidade Federal da Paraíba
Centro de Tecnologia
Programa de Pós-Graduação em Engenharia Mecânica
Doutorado

APLICAÇÃO DE ALGORITMOS DE DEEP LEARNING COMO MODELOS SUBSTITUTOS DE SIMULADORES DE RESERVATÓRIOS DE HIDROCARBONETOS

por

Rafael Marrocos Magalhães

Tese apresentada à Universidade Federal da Paraíba
para obtenção do Grau de Doutorado.

João Pessoa – Paraíba **Março / 2021**

RAFAEL MARROCOS MAGALHÃES

APLICAÇÃO DE ALGORITMOS DE DEEP LEARNING COMO MODELOS SUBSTITUTOS DE SIMULADORES DE RESERVATÓRIOS DE HIDROCARBONETOS

Tese apresentada ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal da Paraíba, em cumprimento às exigências para cumprimento às exigências para obtenção do Grau de Doutorado.

Orientador: Prof. Dr. Moisés Dantas dos Santos

João Pessoa – Paraíba **Março / 2021**

Catalogação na publicação
Seção de Catalogação e Classificação

M157a Magalhães, Rafael Marrocos.
Aplicação de algoritmos de deep learning como modelos substitutos de simuladores de reservatórios de hidrocarbonetos / Rafael Marrocos Magalhães. - João Pessoa, 2021.
112 f. : il.

Orientação: Moises Dantas dos Santos.
Tese (Doutorado) - UFPB/CT.

1. Deep Learning. 2. Proxy Models. 3. Surrogate Models. 4. Simulação de Reservatórios. 5. Redes Neurais Artificiais. I. Santos, Moises Dantas dos. II. Título.

UFPB/BC CDU 004.032:532

Elaborado por Gracilene Barbosa Figueiredo - CRB-15/794

APLICAÇÃO DE ALGORITMOS DE DEEP LEARNING COMO MODELOS SUBSTITUTOS DE SIMULADORES DE RESERVATÓRIOS DE HIDROCARBONETOS

por

Rafael Marrocos Magalhães

Tese apresentada e aprovada em 31 de março de 2021

Período letivo 2020.1

Prof. Dr. Moisés Dantas dos Santos

Presidente da Banca e Orientador

Prof. Dr. Abel Cavalcante Lima Filho

Examinador Interno e Coordenador do Programa

Prof. Dr. Gustavo Charles Peixoto de Oliveira

Examinador Interno

Prof. Dr. Leonardo Vidal Batista

Examinador Externo

Prof. Dr. Igor Fernandes Gomes

Examinador Externo

João Pessoa – Paraíba **Março / 2021**

APLICAÇÃO DE ALGORITMOS DE DEEP LEARNING COMO MODELOS SUBSTITUTOS DE SIMULADORES DE RESERVATÓRIOS DE HIDROCARBONETOS

RESUMO

Um grande desafio da engenharia de reservatórios de óleo e gás e na gestão de produção dos mesmos está relacionado à capacidade de realizar processos de otimização devido ao alto custo computacional exigido pelas simulações numéricas. Apesar de já existirem soluções científicas apresentadas, ainda não foi avaliado o uso de Deep Learning no escopo de predição em baixa granularidade, a nível de célula por exemplo. Neste trabalho são apresentados o procedimento de análise e seleção de descritores, bem como a definição e treinamento de modelos inteligentes com o uso de técnicas de Deep Learning (DNN e CNN), Planejamento Experimental, e toda a avaliação por métricas estatísticas e visuais como solução para criação de modelos substitutos (proxy models) para os simuladores. São apresentados experimentos e resultados para quatro diferentes cenários típicos da indústria, incluindo poços produtores e injetores, e também com diferentes momentos do processo produtivo (poços verdes, em desenvolvimento e maduros). Os resultados indicam taxas de precisão sempre maiores que 80% e chegando em vários casos a 99,9% de acurácia.

PALAVRAS-CHAVES: Deep Learning, Proxy Models, Surrogate Models, Simulação de Reservatórios, Redes Neurais Artificiais

APPLICATION OF DEEP LEARNING TECHNIQUES AS PROXY MODELS ON OIL AND GAS RESERVOIR ENGINEERING

ABSTRACT

A significant challenge on Oil and Gas Industry, especially in reservoir engineering and its management context, is related to the capacity of run optimization strategies. It occurs because the current computational time costs are prohibitive and demand too many resources, even for a medium-size simulation. Perhaps the currently scientific solutions, none of them used Deep Learning techniques in low-level granularity for pre-dictions, especially the grid-cell level size approach. This thesis proposes, analyzes, and states a feature selection, a model design, and a training strategy with the application of Deep Learning techniques (DNN and CNN), the Design of Experiment, and all statistical evaluation-based metrics and its graphic tools. This defined process intends to work as a solution to create a proxy model for reservoir numerical software simulation. Four different classical industrial scenarios, including production and injection wells, are conducted in a diversified temporal sampling to generate proxy models. The achieved results are promising, with accuracy always bigger than 80%, and at specific scenario conditions, it reaches even 99,9% as evaluation criteria.

PALAVRAS-CHAVES: Deep Learning, Proxy Models, Surrogate Models, Reservoir Simulator, Artificial Neural Networks

SUMÁRIO

LISTA SIGLAS

Apesar de todo o empenho e esforço aplicado pela comunidade científica nacional em atribuir ou padronizar traduções adequadas para nossa língua materna, verificamos que ainda não há consenso em traduções de grande parte da tecnologia apresentada em algumas seções deste trabalho. Isto ocorre em especial na temática das Redes Neurais Artificiais mais recentes. Por tanto, com o intuito de ficar mais coerente e facilmente compreensível por parte do leitor, optamos por manter algumas dessas nomenclaturas em Inglês, apresentando sempre que possível a existência de seus equivalentes em Português.

AG – Algoritmo Genético / (GA – Genetic Algorithms)

AI – Artificial Intelligence

CNN – Convolutional Nerual Network

DL – Deep Learning

DLNN – Deep Learning Neural Netwok

E&P – Exploração e Produção

ERT – Estimativa de Recuperação Total / (EUR – Estimated Ultimate Recovery)

LF – Lógica Fuzzy / (FL – Fuzzy Logic)

LSTM – Long Short-Term Memory

P&G – Petróleo e Gás Natural / (O&G – Oil ang Gas Industry)

PM – Proxy Model

RBF – Radial Basis Function / (FBR – Função de Base Radial)

ReLU – Rectified Linear Unit

RL – Reinforcement Learning / (AR – Aprendizagem por Reforço)

RNN – Recurrent Neural Network

RNA – Redes Neurais Artificiais / (ANN – Artificial Neural Network)

SCNN – Sparsely Connected Neural Network

SVM – Support Vector Machine / (MVS – Máquinas de Vetor de Suporte)

SVR – Support Vetor Regressor / (RVS – Regressores de Vetor de Suporte)

TDNN – Time-delayed Neural Network

LISTA FIGURAS

LISTA TABELAS

1 INTRODUÇÃO

A indústria do Petróleo e Gás Natural (P&G) passa por uma transformação significativa relacionada à digitalização de toda sua base de equipamentos e processos, em especial nas etapas de Exploração e Produção (E&P), referenciada eventualmente de Campos Digitais e Indústria 4.0 (CARVAJAL; MAUCEC; CULLICK, 2018a; SAPUTELLI, 2016). Segundo Bravo (BRAVO *et al.*, 2014), para se manter competitiva nas próximas décadas, a indústria do P&G deve adotar o mais rapidamente possível as novas gerações de transformação digital, tecnologias e processos que incluem especialmente a habilidade de captar e lidar com análises avançadas de tendências e modelos de correlação para rapidamente e eficientemente descobrir conhecimento "oculto" em todos os conjuntos de dados, dos pequenos aos grandes e complexos, desde históricos de perfuração aos de produção.

Para efeitos de percepção financeira e prática, em termos globais, se a indústria conseguisse otimizar, através da aplicação destes métodos inteligentes, o desempenho do bombeio dos poços em apenas 1%, isto acarretaria em um aumento de produção de quase meio milhão de barris por dia, consequentemente gerando um faturamento adicional de U$ 19 bilhões de dólares por ano (CARVAJAL; MAUCEC; CULLICK, 2018a). Os mesmos autores destacam ainda que a transformação sustentável da Exploração e Produção no P&G exige uma escalável habilidade analítica e cognitiva, como técnicas de mineração de dados massivamente distribuídas, algoritmos de Aprendizagem de Máquina (AM) e Aprendizagem Estatística, com quase nenhuma ou qualquer supervisão humana.

Alenezi e Mohaghegh (ALENEZI; MOHAGHEGH, 2016a) capitulam que uma das mais importantes ferramentas da atividade de gerenciamento e desenvolvimento de reservatórios de óleo e gás são os simuladores de produção baseados em modelos geofísicos. É uma ferramenta necessária para a engenharia de planejamento de estratégias. Segundo os mesmos autores:

> *"o objetivo principal do simulador de reservatórios é o de predizer o desempenho futuro de um reservatório e auxiliar a encontrar caminhos e meios*

> *de otimizar a recuperação de hidrocarbonetos em diferentes condições operacionais*" (ALENEZI; MOHAGHEGH, 2016a).

Simulações mais robustas exigem modelos de reservatórios mais precisos e fidedignos que envolvem uma abrangente descrição de suas propriedades físicas constitucionais e da dinâmica de seus fluidos. Ferramentas computacionais extremamente robustas implementam sistemas de soluções numéricas para os modelos matemático-físicos que regem os comportamentos desses reservatórios (ALENEZI; MOHAGHEGH, 2017; POULADI *et al.*, 2017; ZUBAREV, 2010). Devido a essa grande complexidade de determinados reservatórios, não é incomum a demanda de semanas de processamento de cluster computacionais para uma única realização de simulação computacional.

Estes softwares compõem as principais ferramentas para operações de predição de produção, avaliação de técnicas de recuperação secundárias e terciárias, planejamento de perfuração, *history matching*, planejamento operacional de campo, planejamento de desenvolvimento e análise de incertezas.

Entretanto, para executar processos de otimização como parte de um plano de gestão, é comum a necessidade de centenas ou milhares de execuções do simulador com diferentes parâmetros para um único reservatório, demandando dias ou mesmo semanas para avaliar diferentes cenários. Tornando, por consequência, os simuladores numéricos inviáveis para determinadas aplicações de otimização e planejamento ou mesmo de gestão de reservatórios mais complexos ou refinados (GOLZARI; HAGHIGHAT SEFAT; JAMSHIDI, 2015a; KRASNOV; GLAVNOV; SITNIKOV, 2018b; NAVRÁTIL *et al.*, 2019).

Para contornar estas limitações, adotou-se como prática na engenharia de reservatórios a substituição dos simuladores por proxy models. Proxy models são construções algorítmicas baseadas em formulações matemáticas, estatísticas ou de inteligência artificial que, devidamente parametrizadas, conseguem substituir em escopo reduzido e com suficiente acurácia determinadas funções normalmente computadas por simuladores de reservatório numéricos. Um sistema tipo proxy model não requer grandes quantidades de recursos computacionais para ser executado, trabalhando por tanto próximo do modo de tempo-real (KRASNOV; GLAVNOV; SITNIKOV, 2018a). Por

consequência, gera resultados imediatos para serem utilizados, por exemplo, em otimizadores, planejadores, análises, *history matching*, dentre outras aplicações.

1.1 ESTADO DA ARTE

Reporta-se em sequência, duas perspectivas do desenvolvimento atual desses modelos aproximativos. A primeira mais geral, em que são sintetizadas informações relativas à nomenclatura, categorias, tipos de dados e escopo de aplicação. E posteriormente uma descrição mais individualizada das contribuições mais relevantes para o desenvolvimento deste trabalho de pesquisa.

1.1.1 SÍNTESE DO SEGMENTO DE ESTUDO

Uma nomenclatura diversa é utilizada para designar este tipo de aplicação ou metodologia. Apesar de tentativas prévias de estabelecer uma nomenclatura padrão (MOHAGHEGH, Shahab D. *et al.*, 2012a), não há unanimidade presente na literatura atual. Nas aplicações de reservatórios de óleo e gás observa-se o uso mais frequente do termo proxy model (JABER; AL-JAWAD; ALHURAISHAWY, 2019) e surrogate reservoir model (AMINI, S. *et al.*, 2014b), nesta ordem. Outras designações existentes na literatura incluem: response surface model, smart-proxy, data-driven model, machine learning model, artificial intelligence model, grid-based proxy, surrogate model, neural network proxy, dynamic surrogate model, cognitive data driven proxy, pattern-recognition based model, adaptative surrogate model, meta model, reduced order model, emulator, end-to-end model.

Um grande esforço científico e técnico tem sido empreendido para desenvolver proxy models para aplicações na indústria do petróleo e gás natural, dentre estes esforços três principais categorias se destacam (AMINI, Shohreh; MOHAGHEGH, 2019; BHOSEKAR; IERAPETRITOU, 2018; NAVRÁTIL *et al.*, 2019; ZUBAREV, 2010). A primeira é denominada Reduced Order Model (ROM) e se desenvolvem alterando a complexidade dos modelos matemáticos e/ou físicos associados aos fenômenos em questão, reduzindo-os através da simplificação geralmente algébrica, de parâmetros específicos ou de precisão, dentro de um escopo pré-definido. O segundo e amplamente utilizado

são os modelos estatísticos nomeados de Response Surface Models (RSM), que substituem os cálculos associados aos processos e focam na replicação das respostas das funções objetivos relacionadas aos processos em estudo, considerando um conjunto reduzido de parâmetros de entradas. A terceira categoria, na qual se encaixa este trabalho, é a denominada Data Driven Models (DDM), fortemente inspirado no paradigma de uso intensivo de dados e na aplicação de algoritmos de Machine Learning (ML) e de Inteligência Artificial (IA), tem como objetivo reproduzir respostas de sistemas complexos considerando suas mais variadas configurações e desdobramentos (BALAJI *et al.*, 2018; CARVAJAL; MAUCEC; CULLICK, 2018b; HEY, TONY; TANSLEY, STEWART; TOLLE, 2009; JORDAN; MITCHELL, 2015).

Dentre os modelos baseados em dados e inteligência artificial, Mohaghegh (MOHAGHEGH, Shahab D. *et al.*, 2012a) considera classificá-los de diferentes formas. Segundo a granularidade e escopo das informações e funções objetivo de saída (grid cell based model, well-based model, full coupled model), basedo no tipo de campo de petróleo em que se desenvolve a solução (green fields, brown fields) baseado na funcionalidade, ou seja, no objetivo definido da aplicação (history matching, predictive problems, development strategies) e, por fim, de acordo com a origem e forma dos dados utilizados para definir o modelo (dados provenientes de simuladores, dados provenientes diretamente de medidas de campo).

Geralmente os modelos baseados em dados são também denominados modelos caixas-pretas (black-box) (BHOSEKAR; IERAPETRITOU, 2018; BRAVO *et al.*, 2014; GORISSEN *et al.*, 2010) em razão da dificuldade em determinar com exatidão os resultados produzidos. Porém, um subconjunto de trabalhos baseados em dados começa a se distinguir dessa denominação, especialmente os modelos desenvolvidos em dados mais refinados, de células de grids, com aplicações relacionadas a predição de estados dinâmicos de reservatórios. É neste tipo de abordagem que se enquadra o presente estudo, tomando como referência e pares os artigos de (ALENEZI; MOHAGHEGH, 2016b; AMINI, S. *et al.*, 2014b; AMINI, Shohreh; MOHAGHEGH, 2019; HAGHSHENAS *et al.*, 2020; MOHAGHEGH, Shahab D. *et al.*, 2012a; SUDAKOV *et al.*, 2019).

Ainda sobre este aspecto, Amini (AMINI, S. *et al.*, 2014a) enfatiza que o desenvolvimento de modelos que seguem essa vertente não ignora os processos físicos associados. Ao contrário, intenta-se exatamente modelar o impacto do processo físico

associado aos fluxos de fluidos em meios porosos em reservatórios simulados, provendo para isso todos os dados correspondentes a este sistema físico. Esta abordagem permite replicar resultados de modelos simulados numericamente não apenas na região de poços ou do reservatório como um todo, podendo predizer valores de saturações e pressões a nível de grid a cada passo de tempo (MOHAGHEGH, Shahab D. *et al.*, 2012b).

1.1.2 CONTRIBUIÇÕES RELEVANTES

Pouladi (POULADI *et al.*, 2017), utiliza um técnica com informações mais físicas do modelo de reservatório, como o efeito de fluxo e de varrido, para criar uma versão de PM baseado em *Fast Marching Method* (FMM) e o aplica com sucesso para determinação de informações de Valor Presente Líquido e com isso aplica em problemas de otimização de *well placement*.

Amirian e Chen (AMIRIAN; JOHN CHEN, 2017) relatam que apesar de serem recentes o uso de criação de modelos baseado em dados utilizando ANN, eles foram empregados com bastante sucesso na predição de recuperação em processos com injeção de água em reservatórios. O modelo foi desenvolvido com dados de análises do reservatório (*core data*) em especial os que descreviam a heterogeneidade do mesmo e concluem que têm grande potencial de serem utilizado em rotinas de tomada de decisão.

Em 2012, Shirangi (SHIRANGI, 2012) construiu um *proxy model* (PM) rápido através da combinação de ANN e *Suport Vector Regressor* (SVR) para permitir a aplicação de otimizadores de produção em um problema robusto, usando também algoritmos não supervisionados (*k-means*) para selecionar um conjunto ótimo de realizações de modelos de reservatórios representativos do ambiente real.

Toal (TOAL, 2016), a exemplo, avalia criteriosamente a contribuição de diferentes arquiteturas computacionais com relação às suas interferências em algoritmos de *surrogate models* que usam *Kriging*.

Golzari (GOLZARI; HAGHIGHAT SEFAT; JAMSHIDI, 2015b), apresenta um muito bem detalhado trabalho de construção de um PM a partir de um modelo adaptativo de extração de dados de treinamento de uma rede neural (ANN). Aplica o estudo para o

reservatório de *benchmark* PUNQ-S3, avalia sua precisão e indica caminhos de otimização do modelo. Wang et al (WANG *et al.*, 2014) abordam de maneira semelhante a obtenção de dados de modo adaptativo e eficiente para estabelecimento do *proxy model*.

Nwachukwu et al (NWACHUKWU *et al.*, 2018), desenvolveram um PM consideravelmente robusto e se destacou dos demais por utilizar um medida relacionada ao comportamento físico de conectividade entre os poços analisados do reservatório. Foram utilizadas as informações padrões de permeabilidade e porosidade, com o uso de uma métrica própria e como algoritmos inteligentes o *XGBoost*.

Em Babaei e Pan (BABAEI; PAN, 2016), são descritos os passos de desenvolvimento de um modelo combinado (ensemble) de algoritmos inteligentes ou modelos estatísticos clássicos como *Kriging*, RBF, regressão multivariada adaptativa, dentre outros para resolver um problema de maximização de Valor Presente Líquido (VPL / NVP) em campos modelados em 2D e 3D.

Em trabalhos que visam aperfeiçoar o desempenho e a qualidade dos resultados de PMs destacam-se os de Zhang et al (ZHANG, J.; CHOWDHURY; MESSAC, 2012), onde se desenvolve um a metodologia de PM híbridos que combinam adaptativamente as características favoráveis de diferentes *proxy models*, incluindo RBFs e *Krigings*. Algo parecido é desenvolvido em (ZHANG, Y. *et al.*, 2016), porém como agregador dos diferentes modelos são utilizado algoritmos não supervisionados. Um *review* mais recente incluindo um detalhamento e avaliação de aplicações de PMs também pode ser encontrada em Bhosekar e Ierapetritou (BHOSEKAR; IERAPETRITOU, 2018).

Algumas ferramentas e *toolbox* foram elaboradas especificamente para trabalhar com desenvolvimento e avaliação de alguns *proxies models* (PMs), duas delas são de código fonte abertos e trabalham com modelos de PMs agregados (ensembles) como os dos trabalhos anteriormente citados. A primeira é a de Müller e Piché (MÜLLER; PICHÉ, 2011) que utilizam uma mistura de PMs e a segunda de Viana (VIANA; HAFTKA; WATSON, 2013) que utilizam uma medida de erro médio quadrático e da raiz das somas dos resultados preditos em uma formulação própria para seleção dos resultados do PMs gerados.

Um dos mais ativos pesquisadores da área, Shahab Mohaghegh, tem se destacado em propor e aplicar *proxy models* na indústria do P&G, fazendo uso especialmen-

te de Redes Neurais Artificiais e outros algoritmos inteligentes desde 2005 (MOHAGHEGH, Shahab D., 2005). Dentre suas contribuições recentes se destacam:

- Um estudo que usa ANN para desenvolver o que ele chama de *Grid-Based Surrogate Reservoir Model*, que funciona como uma réplica de um modelo de simulação de reservatório complexo, treinado calibrado e validado para reproduzir de modo acurado resultado em nível de *grid* (células do modelo em simulador) (AMINI, S. *et al.*, 2014b).
- A partir da técnica desenvolvida, ele aplica um modelo mais refinado em um reservatório gigante e maduro, tendo resultados igualmente significativos fornecendo informações sobre a produção total, individual e acumulada e o corte de água. Torna mais explícito a forma como combina informações das células entre si (MOHAGHEGH, S D *et al.*, 2015)
- Em 2016, já denominado de *smart-proxy model*, a proposta usa informações espaço temporais em conjunto com informações de células da vizinhança e localização dos poços para predizer as funções objetivos desejadas, tudo com base em ANN (ALENEZI; MOHAGHEGH, 2016a).
- Em trabalho mais recente, utilizando desta vez dados mais crus como *well logs*, histórico de produção, dados de completação e de fratura hidráulica, um modelo de reservatório de xisto é desenvolvido com todas as técnicas anteriores de ANN adicionadas a outras de *data-mining* e *machine learning* (ESMAILI; MOHAGHEGH, 2016)

1.2 MOTIVAÇÃO

Uma lacuna observada na literatura anteriormente relatada, diz respeito à ausência do uso de algoritmos do tipo *Deep Learning* (DL) na construção de PMs, em especialmente em granularidades finas (escopo de células). As arquiteturas de Redes Neurais Artificiais (RNA/ANN) conhecidas como *Deep Learning* (DL) que designam, em verdade, uma grande infinidade de diferentes arquiteturas com variados fins e objetivos (NISBET; MINER; YALE, 2018), não são citadas em nenhum artigo recente sobre o

tema de modelos alternativos de reservatórios ou o fazem sem na verdade o implementarem adequadamente no nível de modelagem celular. Sendo algoritmos muito recentes, estando na frente de pesquisa em *machine learning* (CARVAJAL; MAUCEC; CULLICK, 2018a), as DL permitiram a construção de modelos computacionais compostos por múltiplas camadas de processamento a aprenderem representações de dados com múltiplos níveis de abstração (LECUN; BENGIO; HINTON, 2015). É comum na literatura a definição de redes contendo de 5 a 20 camadas ocultas podendo ter, com bastante frequência, redes com centenas de camadas ocultas.

Considerando a assertiva de Goodfellow et al (GOODFELLOW; BENGIO; COURVILLE, 2016), estes métodos incrementaram dramaticamente o estado da arte da resolução de problemas complexos em domínios científicos variados como reconhecimento de voz, reconhecimento de objetos, aplicações farmacêuticas em desenvolvimento de remédios, pesquisa genômica, dentre tantas outras. Yann LeCun (LECUN; BENGIO; HINTON, 2015) e Nisbet (NISBET; MINER; YALE, 2018) acreditam que as *Deep Learnings* (DLs) terão um sucesso ainda maior num futuro próximo porque elas requerem muito menos esforços braçais da engenharia, de forma a facilitar e tirar vantagem dos recentes progressos da capacidade computacional disponível e do crescente e acessível volume de dados de domínios específicos.

Esse espaço em aberto estimulou a realização desta pesquisa, pois, uma investigação do ponto de vista científico é capaz de abrir caminho para todo um novo campo de investigação ainda em franca expansão dentro da própria indústria do P&G.

Corroborando incisivamente essa motivação, pontua-se a seguir, excertos argumentativos que favorecem essa perspectiva:

- Carvajal et al (CARVAJAL; MAUCEC; CULLICK, 2018a) declaram que o "foco deve ser intensificado em aplicações de otimização da produção, isto porque elas se correlacionam muito firmemente com a gestão de reservatórios baseada em dados e por existirem áreas significativas de aplicações nos novos sistemas de campos digitais".
- Balaji (BALAJI *et al.*, 2018) afirma que "métodos baseados em dados e inteligência computacional estão cada vez mais sendo utilizados como

substitutos ou complementares aos modelos baseados em física como os simuladores numéricos."

- Amirian e Chen (AMIRIAN; JOHN CHEN, 2017), também reconhecem o potencial das práticas de modelagem de *proxies* através de dados e computação inteligente, destacam que elas "incorporam técnicas de mineração de dados e de aprendizagem de máquina (*machine learning*) como fascinantes substitutos de modelos explícitos (físicos), especialmente em predições de sistemas extremamente não lineares".
- Nwachukwu (NWACHUKWU *et al.*, 2018) acentua o uso crescente do desenvolvimento de PMs não baseados em modelos físicos sendo construídos com os recentes avanços da inteligência artificial e da aprendizagem estatística, diz ainda que "ferramentas puramente baseada em dados são utilizadas para encontrar padrões complexos entre parâmetros de controle e respostas de reservatórios".
- Por fim, trabalhos recorrentes expõem resultados bem sucedidos mais recentemente fazendo o uso de Redes Neurais Artificiais (RNA/ANN), apresentando resultados superiores a maioria dos outros algoritmos inteligentes (ALENEZI; MOHAGHEGH, 2016a; AMIRIAN; JOHN CHEN, 2017; BALAJI *et al.*, 2018; MOHAGHEGH, S D *et al.*, 2015).

1.3 OBJETIVOS

Esta tese tem como meta a apresentação e desenvolvimento de experimentos que consigam demonstrar a viabilidade e uma avaliação de resultados quantitativos da aplicação de algoritmos e arquitetura do tipo DL no desenvolvimento de modelos substitutos, também denominados de *proxy models* ou *surrogate models*, para os modelos de reservatórios de hidrocarbonetos realizados em simuladores numéricos computacionais convencionais.

A partir do exposto no estado da arte e alavancado pela motivação apresentada, define-se como hipótese avaliada ***a viabilidade do uso de arquiteturas do tipo Deep Learning no projeto, treinamento e aplicação de modelos inteligentes para cria-***

ção de modelos substitutos (proxy models) de simuladores numéricos de reservatórios de óleo e gás, tendo como critério avaliativo principal a sua capacidade de responder em semelhança a senários de produção adequadamente estabelecidos.

Para atingir tal finalidade é estabelecido o seguinte objetivo principal:

- Realizar um conjunto de experimentos de uso de arquiteturas do tipo *Deep Learning*, baseados em dados de *benchmarks* de reservatórios abertos e acessíveis, devidamente documentado e passível de replicação, que demonstram sua viabilidade como modelos substitutos de reservatórios simulados de modo convencional em cenários pré-definidos e limitado por condições de operações igualmente estabelecidas.

Para atingir tal objetivo estabelece-se como metas intermediárias e subsequentes os seguintes objetivos específicos:

1. Investigação a cerca das práticas correntes de desenvolvimento de *proxy models (surrogate models)*, em especial, no domínio de estudos da Indústria do Petróleo e Gás Natural (P&G);
2. Definição e execução de experimentos que demonstrem a construção de PMs com uso de DL a partir de *benchmarks* de reservatórios já aceitos e documentados na literatura científica pertinente; e
3. Avaliação quantitativa e análise crítica dos resultados obtidos, comparando-os sempre que cabível com resultados equivalentes provenientes de simuladores de uso comercial e de qualidade reconhecida.

1.4 METODOLOGIA GERAL DO DESENVOLVIMENTO DO TRABALHO DE TESE

O desenvolvimento desta pesquisa faz uso da metodologia clássica da pesquisa científica, contendo as etapas de busca e fundamentação teórica através da pesquisa bibliográfica, a recuperação de artefatos (experimentos, argumentos, resultados, gráficos) que sustentem o estabelecimento de hipóteses a serem avaliadas por experimentos e a análise de resultados.

Especialmente sobre as etapas de pesquisa e fundamentação teórica foram utilizadas três técnicas de pesquisa. O mapeamento sistemático estabelecido através de um protocolo, a busca por referências pela técnica recursiva ascendente e descendente *(snowball)* e por fim, a busca exploratória livre.

Vale explicitar que os experimentos foram realizados a partir de dados provenientes de simuladores e computação numérica, validados, com uso de técnicas de planejamento experimental *(Design of Experiment)* e métricas pré-estabelecidas para mensuração quantitativa e estatística dos resultados apresentadas no capítulo Materiais e Métodos.

Este documento de tese se desenvolve em mais quatro capítulos. O segundo capítulo trata sobre os aspectos fundamentais teóricos, seguido pelo terceiro sobre materiais e métodos essenciais à execução dos experimentos. O quarto capítulo descreve o conjunto de resultados experimentais alcançados afim de atingir os objetivos propostos e avaliar as hipóteses formuladas e conclui com o quinto e último capítulo com as considerações finais e proposição de encaminhamentos.

2 FUNDAMENTAÇÃO TEÓRICA

A área de Inteligência Artificial (IA) bem como a da Engenharia de Petróleo e Gás (P&G), campos de estudo em que se insere este trabalho, são dois universos de grande magnitude teórica e aplicadas. De naturezas multidisciplinares, ambas se estendem e se conectam com as mais variadas áreas do conhecimento.

Na seção 2.1 deste capítulo apresentam-se formulações importantes para compreensão de como se relacionam os principais dados referentes à simulação de reservatórios de P&G a partir de softwares numéricos, e o porquê de seu uso em trabalhos desta natureza.

A seção seguinte, 2.2, trás uma descrição sobre os Modelos Substitutos (*proxy models / surrogate models*), especialmente sobre como eles são desenvolvidos, avaliados e aplicados. Expõe-se o necessário para compreensão e crítica dos experimentos conduzidos nos capítulos posteriores.

Na seção 2.3 apresenta-se com o detalhamento necessário, mas sem a intenção de esgotar o escopo, as informações mínimas para compreensão de duas sub-áreas da IA conhecidas como Sistemas Inteligentes e Aprendizagem de Máquina (*Machine Learning*). Isto porque os algoritmos mais utilizados nos processos práticos de otimização, análise de dados e auxílio a tomada de decisão na Indústria do P&G fazem interseção nestes dois campos de saberes. Segue-se com a exposição dos algoritmos inteligentes clássicos empregados na área de P&G e de um conjunto de aplicações que usualmente se solucionam com cada tipo. Fechando a seção, descreve-se as arquiteturas de Deep Learning (DL), que são contemporâneas, mais robustas e que, apesar de novas, estão em forte evolução científica e tecnológica.

O capítulo é encerrado com a seção 2.4 sintetizando o conteúdo das referências mais estreitamente ligadas à esta Tese. São relacionados, dentre outras, o foco principal, as técnicas de aprendizagem e amostragem, métricas e dataset utilizado em cada um dos estudos precedentes, seguidos de uma breve avaliação quantitativa sobre o que se destaca em comum.

2.1 SIMULAÇÃO DE RESERVATÓRIOS DE PETRÓLEO E GÁS E FORMULAÇÕES

No estudo, prática e gerenciamento de reservatórios, é comum a aplicação de modelos equivalentes para compreensão dos fenômenos físicos pertinentes a estas estruturas. Estes modelos equivalentes, denominados de simuladores, costumam ser modelados especialmente através de formulações físico/matemáticas e implementados numérica e computacionalmente. Como citado por Alenezi (ALENEZI; MOHAGHEGH, 2016a) no capítulo inicial deste texto, "o objetivo principal do simulador de reservatórios é o de predizer o desempenho futuro de um reservatório e auxiliar a encontrar caminhos e meios de otimizar a recuperação de hidrocarbonetos em diferentes condições operacionais". Os simuladores numéricos de reservatórios, que segundo Rosa (ROSA; CARVALHO; XAVIER, 2013) também são denominados de *simuladores numéricos de fluxo*, são definidos complementarmente como "métodos empregados na engenharia de petróleo para se estimar características e prever o comportamento de um reservatório de petróleo". Ainda segundo o mesmo autor, estes simuladores podem ser classificados em função do seu tratamento matemático (modelos volumétricos, composicionais ou térmicos), pelas dimensões (uni, bi ou tridimensionais) e por fim quantidade de fases materiais consideradas (mono, bi ou trifásicos).

Destes simuladores, obtém-se dados referentes às propriedades estáticas e dinâmicas, referentes a sua caracterização implícita e a sua dinâmica comportamental ao longo de diferentes instantes de tempo.

Destacam-se duas formulações neste trabalho para melhor compreensão das relações físico/matemáticas e dos correspondentes dados utilizados para criação dos modelos proxies. Primeiro, a Equação 1 apresentada em (NAVRÁTIL *et al.*, 2019), transcrita a seguir:

$$\frac{M_{c,f}(x_t) - M_{c,f}(x_{t-1})}{\Delta t} = F_{c,f}(x_t) + Q_{c,f}(x_t, u_t) \tag{1}$$

Onde:

c - Cada célula presente no grid

f - Cada fluido existente no reservatório (óleo, gás, água)

x_t	- Propriedades e saturações relativas (porosidade, permeabilidade, saturações)
$M(x_t)$	- Massa do fluido na célula
$F(x_t)$	- Fluxo de massa de/para as células vizinhas
u_t	- Controles de poços (aberto, fechado, pressão, etc.)
$Q(x_t, u_t)$	- Fluxo de massa nos poços

Desta formulação importa-se destacar a necessidade de se considerar as variações em cada instante de tempo não apenas em cada célula em detalhe, mas também, em qual direção e quantidade considerando as células vizinhas. Além disso, informações características intrínsecas como as petrofísicas e de saturação, e as associadas aos efeitos e consequências em cada poço presente no reservatório (produção ou injeção).

A segunda formulação importante é incorporada a partir do trabalho de Mohaghegh (MOHAGHEGH, Shahab Dean, 2011), e transcrita a seguir:

$$Q = f(x_1, x_2, \dots, x_n \,\&\, y_1, y_2, \dots, y_n \,\&\, w_1, w_2, \dots, w_n) \tag{2}$$

Onde Q, igualmente representando o volume de massa produzido (ou injetado) pelo reservatório, é interpretado como uma relação funcional entre os caracterizadores petrofísicos e dos fluidos do reservatório ($x_1, x_2, \dots, x_n$, como porosidade, permeabilidade, saturações, etc.), suas restrições operacionais ($y_1, y_2, \dots, y_n$, abertura, fechamento, pressão, etc.) e também outras características operacionais ou representativas do reservatório (técnica de completação, configuração dos poços, etc.). Desta interpretação, destaca-se a possibilidade de priorizar o foco no volume do fluido de interesse em função de todos os outros parâmetros, e não apenas a variação do fluxo. Possibilitando uma interpretação a partir dos valores absolutos ou de suas variações.

Assim, dado um fluido de interesse, importa-se para o desenvolvimento da solução a predição de seu volume ou de sua variação em diferentes momentos de tempo, considerando para tanto, informações constitutivas, dinâmicas e operacionais do reservatório.

2.2 PROXIES COMO SUBSTITUTO DE SIMULADORES

Devido a grande complexidade de um reservatório, algumas vezes demanda-se esforço computacional superlativo para desenvolver e executar uma instância de um modelo numérico (computacional) do reservatório correspondente. A exemplo prático, Golzari et al (GOLZARI; HAGHIGHAT SEFAT; JAMSHIDI, 2015b), relatam que uma única execução de um modelo de reservatório em um simulador, desenvolvido a partir de milhares e às vezes milhões de células pode levar centenas de horas.

Predições precisas de comportamentos de reservatórios em respostas às mudanças em parâmetros de controles são atividades rotineiramente efetuadas em gerenciamento de reservatórios em processos de otimização. Logo, para modelos de reservatórios muito grandes e refinados, em efeitos práticos, fica proibitiva computacionalmente a implementação de determinados algoritmos de otimização.

Uma alternativa para superar tais dificuldades é utilizando Proxy Models (PMs). Proxy Models (ou surrogate models) são alternativas computacionalmente mais baratas em comparação aos simuladores computacionais numéricos especialmente nas atividades de history matching, otimização da produção e predição.

Um PM pode ser definido (ALENEZI; MOHAGHEGH, 2017; ZUBAREV, 2010) como um modelo alternativo matemático, estatístico ou baseado em dados que replica a saída de um modelo de simulador para determinados parâmetro de entrada, ou seja, ele replica uma determinada função, dentre inúmeras, das quais um simulador consegue executar. Os resultados provenientes de PMs não são a mimetização perfeita e fidedigna aos resultados de simuladores, elas incluem um determinado grau de erro aceitável dentro de um limite determinado (AMINI, S. *et al.*, 2014b).

Segundo (MOHAGHEGH, S D *et al.*, 2015), engenheiros de reservatórios requerem ferramentas que: 1) disponibilizem uma busca rápida e precisa de uma grande variedade de operações e opções e, 2) sejam capazes de quantificar as incertezas associadas com as decisões gerenciais. Para realizar esta importante tarefa com as tecnologias tradicionais, uma coisa deve ser sacrificada, precisão ou velocidade. Neste sentido, reduzir o tempo computacional para alguns segundos, tem permitido aos PMs se apresentarem competitivos e atrativos para engenheiros de reservatórios.

Quanto à denominação, há uma variedade enorme de nomes associados ao mesmo procedimento de desenvolvimento de modelos substitutos, sendo os termos *proxy models* e *surrogate models* os mais frequentes (BHOSEKAR; IERAPETRITOU, 2018; ZUBAREV, 2010). Na Tabela 1 são apresentados os nomes encontrados no escopo de pesquisa deste trabalho, categorizados em função de sua similaridade com os dois principais e outros tipos.

Tabela 1. Denominações de proxy models encontradas na literatura.

Nome Raiz	Proxy Model	Surrogate Model	Outros
Variações	Proxy-model; proxy-modelling; proxies; smart-proxy	Surrogate-model; surrogate models; surrogate modeling; surrogate reservoir modeling; surrogate reservoir model (SRM)	Reservoir simulation ; reduced order model (ROM); approximation model; response surface model; metamodel; metamodeling; meta model; emulator; kriging model; estimator model; regression model

2.2.1 CRIAÇÃO DE MODELOS ALTERNATIVOS (PROXY MODELS)

Existem várias abordagens para o desenvolvimento de um *proxy model*, com excelentes reviews e resumos sobre eles na literatura (BHOSEKAR; IERAPETRITOU, 2018; FORRESTER; KEANE, 2009; ZHANG, Y. *et al.*, 2016; ZUBAREV, 2010). Uma categorização razoavelmente consistente sobre os modelos de PMs os dividem em três tipos (GORISSEN *et al.*, 2010; WANG *et al.*, 2014):

- **Modelos Matemáticos ou Estatísticos** – são os mais comuns e mais amplamente estudados e utilizados, incluindo os modelos desenvolvidos por regressões lineares, regressão polinomial, métodos de *kriging* e regressão multivariada adaptativa por *splines*. Estes métodos geralmente desenvolvem uma superfície de resposta para uma determinada função desejada de saída fornecida pelo simulador do reservatório.

- **Modelos físico-matemáticos reduzidos** – são menos frequentes devido às dificuldades operacionais em realizá-los. Eles tentam reduzir a precisão dos modelos implementados pelos simuladores através de aproximações dos modelos físico-matemáticos que governam o comportamento dos fluídos dentro do reservatório. Geralmente exigem acesso ao código dos simuladores o que não é trivial.
- **Modelos baseados em dados** – considerados os mais recentes, eles fazem uso de algoritmos inteligentes, utilizam dados de simulações dos reservatórios e nem sempre geram uma superfície simples de resposta de variáveis dos simuladores, podendo inclusive serem desenvolvidos sem informações de simuladores (exigindo tão somente as informações físicas dos processos reais, menos comum). Dentre os principais algoritmos utilizados destacam-se: regression tree, random forest, support vector machines, radial basis functions, artificial neural networks e mais recentemente xgboost.

Há inúmeras variações para cada tipo de algoritmo empregado independente do tipo de modelo (BABAEI; PAN, 2016). Mais recentemente houve uma tentativa de segmentar um grupo de algoritmos que usam as informações do simulador relacionadas aos descritores e parâmetros dos modelos (informações das células) do que apenas do comportamento funcional geral do mesmo. Trabalhos como os de Mohagheg (ALENEZI; MOHAGHEGH, 2017; AMINI, S. *et al.*, 2014b; ESMAILI; MOHAGHEGH, 2016) e o de (AMIRIAN *et al.*, 2018; AMIRIAN; JOHN CHEN, 2017) destacam inclusive a tentativa de emplacar diferentes nomenclaturas para estes modelos como “smart-proxy” e “cognitive data-driven proxy”.

Para realização de construção de um *proxy model*, este trabalho optou pela bem descrita instrução estabelecida em Zubarev (ZUBAREV, 2010), por acreditar que ela engloba os principais trabalhos na área de P&G e permite facilidade de adaptação porventura necessária.

O fluxo de trabalho apresentado na Figura 1 sumariza as principais etapas necessárias para construção e uso de um PM e são descritas a seguir:

- **Definição das Variáveis de Entrada** – A seleção das variáveis de entrada a serem colhidas a partir do modelo real ou do simulador tem natureza fortemente correlacionada com o problema a ser tratado e com o conhecimento teórico e prático do engenheiro. Existe um balanço (*trade-off*) entre a quantidade desejada de parâmetros de incerteza do campo e a quantidade viável para uma solução rápida e simples do modelo. Uma estratégia interessante, no caso de haver a possibilidade, é a de incluir tantas quantas forem possíveis o número de variáveis e ir descartando parâmetros menos importantes na etapa de análise de sensibilidade.
- **Análise de Sensibilidade** – A análise de sensibilidade é a quantificação do impacto que cada parâmetro (variável) de entrada causa na resposta do modelo. Isto é feito através da avaliação do quanto a resposta muda a partir da associação de uma ou mais variáveis em diferentes simulações. Independente do desenvolvimento de PMs, esta etapa é extremamente importante pra a solução de problemas a partir de um simulador de reservatórios. A partir da análise é possível descartar as variáveis que afetam menos o modelo e, com isso, reduzir o esforço computacional total. A análise por variância e análise estatística são as ferramentas mais comuns. É possível utilizar ferramentas mais robustas como Análise de componentes principais (PCA) dentre outras.
- **Amostragem do BD de Entrada** – Considerada a etapa mais importante do processo, a amostragem dos dados de entrada para criação do bando de dados (*dataset*) a ser utilizado pelo PM pode ser realizada de diferentes formas e é alvo frequente de otimização à priori, posteriori ou durante a criação do modelo. Dentre alguns métodos destaca-se o clássico da área de Planejamento Experimental (Design of Experiment – DoE), o Projeto por Hipercubo Latino (Laint Hypercube Design). Outras abordagens incluem: amostragem adaptativa, amostragem geométrica, validação cruzada (*cross-validation*), *jacknifing*, dentre outras. Quanto maior for a variabilidade do espaço amostral, maior será o número de amostras para ser utilizado na adequação do modelo.
- **Estimação do Proxy-Model** – Nesta etapa, dependendo do modelo de PM escolhido, o algoritmo pertinente de ajuste de modelo deve ser empregado utilizando-se para tanto o banco de dados amostrado (*dataset*) na etapa anterior. É

nesta etapa que o modelo escolhido (*proxy model*) deve ser ajustado para mimetizar tanto quanto possível o comportamento apresentado pelo simulador. Por exemplo, se utilizado um ANN, este é o momento de treinamento e ajuste dos pesos em função do *dataset* determinado.

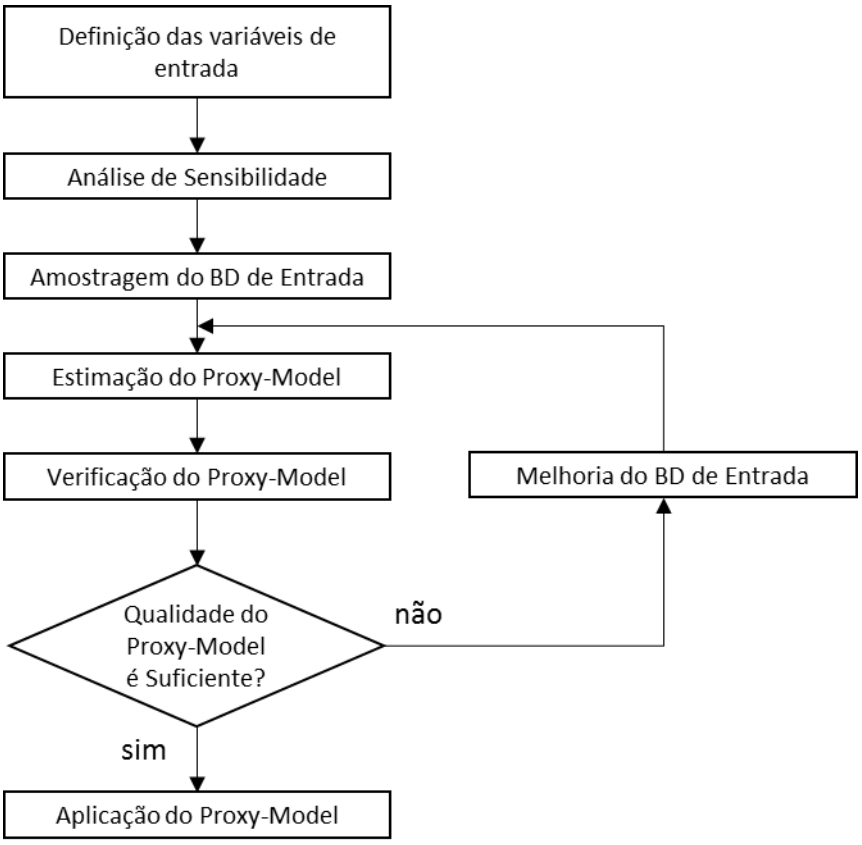


Figura 1. Fluxo de etapas para criação de um Proxy-Model (adaptado de ZUBAREV, 2010).

- **Verificação do Proxy-Model** – A verificação do *proxy model* trata do processo de aferir a acurácia de predição do modelo. Geralmente compara-se os resultados obtidos pelo PM com os resultados obtidos pelo simulador computacional do reservatório, para um conjunto de dados (*dataset*) não utilizados durante o ajuste do PM (etapa anterior), de modo a garantir independência e maior confiabilidade. O nome desse *dataset* independente é *dataset* de teste. Este pode ser construído com alguma técnica de planejamento experimental específica para garantir uma combinação especial de parâmetro de entradas e robustez, ou como uma amostra aleatória do *dataset* original.

- **Melhoria do BD de Entrada** – Na hipótese de o PM não ter atingido um desempenho satisfatório, é possível tomar algumas ações acerca do problema. Dentre as possíveis cita-se a avaliação da complexidade do modelo (se a ANN tem neurônios suficientes, se a RBF tem funções ou parâmetros suficientes, etc.) ou da qualidade e abrangência de representatividade do *dataset* (conjunto de dados do treinamento). No caso da segunda hipótese existem várias formas de incrementar o conjunto de dados, um exemplo seria a aferição da resposta do *proxy model* comparada a do simulador e, para as piores respostas, adicionar os pontos com baixo desempenho ao conjunto de treinamento (GORISSEN *et al.*, 2010; ZUBAREV, 2010).

2.2.2 AVALIAÇÃO DE DESEMPENHO DE PROXY MODELS

A avaliação de desempenho de um PM pode levar em conta questões sobre suas respostas a determinadas funções objetivos e também às medidas de qualidade de cada tipo de algoritmo empregado em sua construção. No âmbito deste trabalho algumas medidas estabelecidas em Bhosekar (BHOSEKAR; IERAPETRITOU, 2018) são empregadas conforme apresentadas na Tabela 2:

Tabela 2. Métricas de avaliação comuns em proxy models (Adaptado de BHOSEKAR, 2018).

Métrica de avaliação	Fórmula
Explained variance score	$1 - \frac{Var\{y - y\}}{Var\{y\}}$
Mean absolute error	$\frac{1}{n}\sum_{i=0}^{n-1} \lvert y_i - \hat{y} \rvert$
Mean squared error	$\frac{1}{n}\sum_{i=0}^{n-1} (y_i - \hat{y})^2$
Median absolute error	$median(\lvert y_1 - \widehat{y_1} \rvert, \dots, \lvert y_n - \widehat{y_n} \rvert)$

R^2 score	$1 - \frac{\sum_{i=0}^{n-1}(y_i - \widehat{y_i})^2}{\sum_{i=0}^{n-1}(y_i - \bar{y})^2}$
Relative average absolute error	$\frac{\sum_{i=0}^{n}\lvert y_i - \widehat{y_i}\rvert}{n \times STD}$
Relative maximum absolute error	$\frac{\max(\lvert y_1 - \widehat{y_1}\rvert, \ldots, \lvert y_n - \widehat{y_n}\rvert)}{STD}$
Onde:	y – saída real $\hat{y}$ – saída do Proxy model (valor predito) n – número de amostras $\bar{y}$ – valor predito médio STD – Desvio Padrão

Além destas métricas, alguns estudos apontam outras mais significativas de acordo com o problema ou domínio a ser avaliado, dentre eles destacam-se o artigo de Fawcett (FAWCETT, 2005) e o relatório técnico de Powers (POWERS, 2007).

2.3 SISTEMAS INTELIGENTES E APRENDIZAGEM DE MÁQUINA (MACHINE LEARNING)

O conjunto de metodologias, algoritmos, técnicas, arquiteturas e estrutura de dados que se aglomeram sob ou se associam ao domínio de estudos da Inteligência Artificial (IA / AI), têm sido ao longo dos anos agrupados em sub-denominações geralmente relacionadas à áreas de aplicação, domínios do saber ou de suas formas de construção (CRANGANU; BREABAN; LUCHIAN, 2015; EMERICK *et al.*, 2009; JORDAN; MITCHELL, 2015; SCHMIDHUBER, 2015; SUTTON; BARTO, 2018).

É comum observar os algoritmos de Lógica Fuzzy (LF / FL) e Redes Neurais Artificiais (RNA / ANN) serem utilizados na área de teoria do controle, os Algoritmos Genéticos e Colônia de Formigas (CF / ACO) na área de problemas de otimização, e novamente as RNA e Máquinas de Vetor de Suporte (MVS / SVM) em problemas de Aprendizagem de Máquina (*Machine Learning* – ML), e toda uma variação de métodos em diferentes agrupamentos.

Alguns autores identificam que o termo "*Intelligent Systems*" (Sistemas Inteligentes) (CARVAJAL; MAUCEC; CULLICK, 2018b; CRANGANU; BREABAN; LUCHIAN, 2015; EMERICK *et al.*, 2009; MOHAGHEGH, S D; KHAZAENI, 2011) foi amplamente utilizado para denominar uma grande gama de algoritmos computacionais que têm por natureza, inspiração, construção ou aplicação algo que os relacione ao comportamento inteligente (humano ou animal) e que sirva para resolver problemas mais complexos e difíceis de serem programados com instruções clássicas, ou seja, problemas menos maquinários ou repetitivos.

Mais recentemente o termo Machine Learning (ML) recebeu uma adesão maior, influenciada por aplicações mais relacionadas a análise e interpretação de dados provenientes da área de ciência dos dados (*data-science*) (JORDAN; MITCHELL, 2015). Neste trabalho optou-se por designar o conjunto desses algoritmos sob o termo guarda-chuva Machine Learning (ML), mesmo entendendo que em contextos distintos eles possam ser referenciados de modos igualmente distintos.

Dois dos autores mais reconhecidos na literatura científica e acadêmica da área de ML, Jordan e Mitchell (JORDAN; MITCHELL, 2015), dissertam que o *Machine Learning*:

> "aborda a questão sobre como construir uma computação que se incremente automaticamente através da experiência. Sendo na atualidade um dos campos técnicos que mais crescem, concentra-se na intersecção da ciência da computação e estatística, e no centro da inteligência artificial e da ciência de dados. Os progressos mais recentes em ML têm sido conduzidos especialmente pelo desenvolvimento de novos algoritmos e teorias de aprendizagem e também pela explosão na disponibilidade de dados online, acessíveis e pela computação de baixo custo. A adoção de métodos de ML baseado em dados pode ser encontrada através da ciência, tecnologia e comércio alavancando mais a decisão baseada em evidências em áreas como engenharia, saúde, manufatura, educação, finanças, políticas e marketing."

Segundo Carvajal (CARVAJAL; MAUCEC; CULLICK, 2018a), "conceitualmente, algoritmos de ML podem ser vistos como uma busca em um grande domínio de programas/soluções candidatos, guiados por uma experiência de aprendizagem, a encontrar

o programa/solução que otimiza uma determinada métrica de desempenho." Ao que Mitchell (JORDAN; MITCHELL, 2015) complementa ao falar que um problema de aprendizagem "pode ser definido como o problema de aprimorar uma determinada métrica de desempenho quando executando uma determinada tarefa, através de uma determinada experiência de aprendizagem".

Uma categorização em três classes distintas de algoritmos geralmente é empregada pela maioria dos autores da área (ANIFOWOSE, Fatai Adesina, 2011; EMERICK *et al.*, 2009; GOODFELLOW; BENGIO; COURVILLE, 2016; JORDAN; MITCHELL, 2015; SCHMIDHUBER, 2015; SHIRANGI, 2012) para separar os diferentes tipos de algoritmos. São elas as de: 1) Algoritmos supervisionados, 2) Algoritmos não supervisionados e, 3) Aprendizagem por Reforço (*Reinforcement Learning* – RL).

Os algoritmos supervisionados são aplicados geralmente em problemas de classificação ou regressão de funções complexas e para que sejam criados modelos adequados é necessário a existência de padrões de entrada e de saída devidamente rotulados para realização do ajuste do modelo (BRAVO *et al.*, 2014).

Já os algoritmos não supervisionados são utilizados mais frequentemente em problemas de agregação de dados (clusterização), redução de dimensionalidade, seleção automatizada, dentre outros. Nestes problemas não há pares de dados correspondentes a entrada e saída do sistema, mas sim definições de quantidades de classes ou padrões distintos para que os algoritmos encontrem similaridades e padrões implícitos de categorização dos dados.

A última classe de algoritmos são os de Aprendizagem por Reforço (RL), que apesar de dependerem de pares de dados de entrada e saída, estas informações geralmente são obtidas a partir da simulação do próprio ambiente a ser modelado. Podem ser definidos a partir de uma metáfora de um agente que atua sobre um ambiente a partir de uma conjunto de regras definidos (política) a fim de obter um retorno de maior valor do ambiente no longo prazo (SUTTON; BARTO, 2018).

2.3.1 TIPOS DE ALGORITMOS E SUAS APLICAÇÕES NA INDÚSTRIA DO P&G

A Tabela 3 sintetiza os principais algoritmos em função de sua categorização, tipo de aplicação, a técnica, o algoritmo, e exemplo de aplicação na indústria P&G.

Uma descrição mais detalhada de cada algoritmo dentre os principais aplicados na indústria do P&G pode ser encontrada em qualquer das seguintes referências (POPA; CASSIDY, 2012)(CARVAJAL; MAUCEC; CULLICK, 2018a)(JORDAN; MITCHELL, 2015)(HOLDAWAY, 2014)(ANIFOWOSE, Fatai A., 2013)(BALAJI *et al.*, 2018)(CRANGANU; BREABAN; LUCHIAN, 2015).

Popa (POPA; CASSIDY, 2012), afirma que as técnicas de Inteligência Artificial têm sido aplicadas com sucesso na área de P&G desde o início da década de 90, inicialmente resolvendo tarefas simples e mais recentemente evoluindo para sistemas híbridos tratando problemas complexos de otimização e de modelagem. Ainda segundo o autor, que trabalha na empresa Chevron, a IA tornou-se parte integral dos negócios há pelo menos 10 anos, com aplicações que variam desde a caracterização de reservatórios, otimização da produção e PMs usados como simulação de reservatórios.

Das principais vantagens e valores gerados pela aplicação dessas tecnologias, identificadas na literatura, estão: o aumento da produção, valor presente líquido (VPL / NPV), e redução de carga horário de homem/hora.

Tabela 3.Tipos de algoritmos por categoria, aplicação, técnica.

Categoria (Paradigma)	**Tipo de Aplicação**	**Técnica**	**Algoritmo**	**Exemplo em P&G**
Aprendizagem Supervisionada	Regressão	Regressão Linear (RL)	Método dos mínimos quadrados; Mínimos quadrados com média móvel	Estimação e predição de dados gerais em praticamente todos os processos do UPSTREAM; em perfuração e completação (logs)
		Regressão Linear Penalizada	Ridge RL; Redes Elásticas	
		Regressão não-linear	MARS; Support Vector Machine (SVM); K-NN; Redes Neurais Aritificiais (ANN)	
		Árvores de Decisão para regressão	CART; árvores de decisão condicionais; Bagging CART; Random Forest (RF); Gradient Boosted Machine (GBM)	
	Classificação	Classificação Linear	Regressão logística; Análise discriminante	Categorização de dados; detecção de falhas; caracterização de reservatórios;
		Classificação não-linear	Análise discriminante, regularizada, quadrática e flexível; SVM; K-NN; Naive Bayes	
		Classificação não-linear com árvores de deci-	CART; Bootstrapped aggregation (Bagging)	

	são	CART; Random Forest (RF); GBM	
Aprendizagem Não Supervisionada	Clustering (agrupamento)	K-means; Hierarchical clustering	Classificação de tipos rochosos; detecção de padrões; tratamentos sísmicos
	Redução de dimensionalidade	Análise de componente principal (PCA); Factor analysis; Escalonamento multidimensional (MDS);	
Aprendizagem por Reforço		Processos de Decisão Markoviano (MDP); Diferenças Temporais; Q-Learning	Otimização de processos; realização de alternativas

Segundo um relatório apresentado por Bravo et al (BRAVO *et al.*, 2014), os termos mais comuns associados à área de Inteligência Artificial que são lembrados ou reconhecidos pelos profissionais da área de P&G são, nesta ordem: Data mining, Neural networks, Workflow automation, Fuzzy logic, Expert systems, Automatic process control, Genetic algorithms, Rule-based on reasoning, Proxy models, Virtual models, Machine learning e, Intelligent agents.

Aparentemente, a ordem de ocorrência e identificação com cada uma destas tecnologias parece estar relacionada com o tempo de aplicação das mesmas na indústria, ou seja, sua maturidade e o aparecimento apenas mais recentemente de outras. Nota-se isso especialmente quando os termos Neural networks, Fuzzy logic e Genetic algorithms, tecnologias amadurecidas na indústria sobressaem-se na memória em relação a Proxy models e Machine learning, sendo estes termos mais recentes.

2.3.2 REDES NEURAIS ARTIFICIAIS (RNA / ANN)

Entendendo que o foco desse trabalho trata da criação de *proxy models* baseados em arquiteturas de Redes Neurais, ainda que utilizando as redes do tipo DL, convém descrever e detalhar melhor as arquiteturas do tipo ANN como subsídio para melhor compreensão do leitor não praticante da técnica.

As ANN são uma das mais populares e utilizadas técnicas de Machine Learning. Segundo Mohaghegh (MOHAGHEGH, Shahab D., 2005), provavelmente uma das pri-

meiras a serem utilizadas na indústria do P&G. As primeiras aplicações foram relacionadas à predição da taxa de penetração em processos de perfuração, diagnóstico de perfuração, de brocas, produção inicial de poços, predição de permeabilidade e/ou porosidade de reservatórios e reconhecimento de cartas de bombeio mecânico.

Simon Haykin, um dos iniciais expoentes da área (HAYKIN, 1999), comenta sobre a inspiração biológica dos algoritmos e que as mesmas provêm uma poderosa ferramenta de interpolação não linear e multidimensional. Através de um processo de treinamento, elas são capazes de capturar potenciais relações não lineares existentes entre uma entrada e saída de sistemas. Para um treinamento bem-sucedido, as redes neurais devem ser expostas a um conjunto de dados representativo e de tamanho suficiente de forma a incorporar conhecimento suficiente e preciso e então ser capaz de detectar situações ainda não previstas. Um ponto frequentemente argumentado é que, sendo um algoritmo do tipo caixa-preta, ela é capaz de encontrar relações não triviais ou de difícil comprovação por métodos matemáticos tradicionais.

Para Popa e Cassidy (POPA; CASSIDY, 2012), o que torna as ANN tão atrativas na indústria do P&G são especialmente sua velocidade, estando aptas a dar respostas praticamente de modo instantâneo, e sua capacidade de lidar com dados incertos e imprecisos, como em problemas de propriedades de reservatórios. Além disso, os autores ainda ressaltam o fato de elas poderem ser treinadas com dados do dia-a-dia do campo e utilizadas como PMs para realizar predições ou classificações de problemas em que geralmente as equações clássicas não funcionariam ou não teriam tempo hábil para tomada de decisão.

A Figura 2 apresenta um neurônio artificial, que neste caso simula em metáfora o neurônio real, mas que funciona na prática como um agregador ponderado que transfere seu resultado para uma função geralmente não linear.

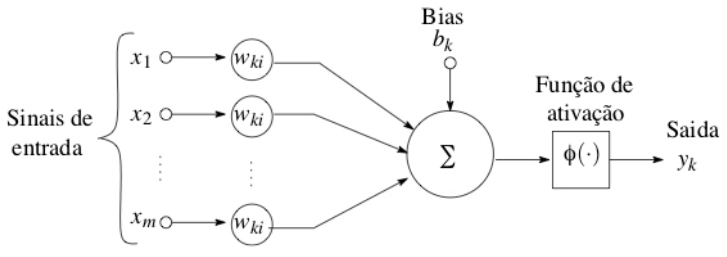


Figura 2. Modelo de neurônio não-linear.

As Redes Neurais Artificiais mais comuns são elaboradas a partir da junção de vários neurônios artificiais em sequências de camadas. Redes do tipo Perceptron de Múltiplas Camadas (PMC) também conhecidas como MLP (do inglês Multi-layer Perceptrons) são um dos tipos mais difundidos de arquitetura de Redes Neurais Artificiais. A arquitetura de uma rede PMC pode ser observada a partir da Figura 3.

Nesse tipo de rede, os dados fluem da camada mais à esquerda, denominada camada de entrada, atravessam cada uma das camadas seguintes, conhecidas como camadas ocultas, e seu fluxo termina na última camada à direita, denominada camada de saída. Em cada passagem por entre as camadas, a entrada é ponderada por um peso sináptico e acumulada juntamente ao bias formando o campo local induzido que é então utilizado pela função de ativação em cada neurônio das várias camadas da rede.

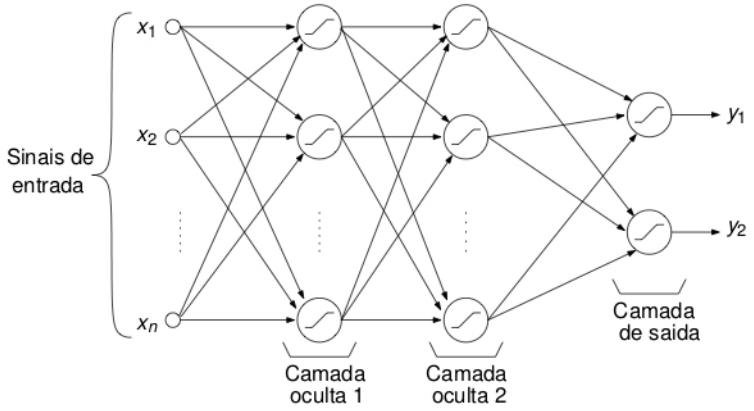


Figura 3. Arquitetura do tipo MLP (Perceptron de múltiplas camadas).

Para que uma rede desse tipo possa ser útil, assim como as demais, ela deve prever um algoritmo de treinamento eficiente. Um algoritmo comumente utilizado e já bem definido na literatura é o algoritmo de retropropagação do erro (*backpropagation*). O algoritmo da retropropagação consiste em ciclos de treinamentos compostos por: apresentações aleatórias de vetores de entradas na RNA, obtenção da diferença entre a resposta fornecida pela rede e a resposta desejada, utilização dessa diferença como o argumento de uma alteração realizada sob o conjunto dos pesos sinápticos da rede, sendo esta alteração baseada no gradiente descendente. O Apêndice [A. Algoritmo da Retropropagação (Backpropagation)] deste trabalho apresenta este algoritmo

e uma leitura mais aprofundada de seu comportamento pode ser obtida em (HAYKIN, 1999; MAGALHÃES, 2007).

2.3.3 DEEP LEARNING (MODELOS CONTEMPORÂNEOS)

A partir de 2012, após um período de aproximadamente uma década sem marcantes inovações, a pesquisa em redes neurais artificiais teve um novo e significativo progresso com a aplicação de redes complexas em problemas de classificação em imagens e também em problemas de maior dimensão em outras áreas como robótica, carros autônomos, medicina, etc. Isto tudo foi possível graças às então denominadas *deep learning*s. As arquiteturas de Redes Neurais Artificiais (RNA / ANN) conhecidas como *deep learning* (DL) designam, em verdade, uma grande infinidade de diferentes arquiteturas com variados fins e objetivos (NISBET; MINER; YALE, 2018).

Sendo algoritmos muito recentes, estando no fronte de pesquisa em machine learning (CARVAJAL; MAUCEC; CULLICK, 2018a), as DL permitiram a construção de modelos computacionais compostos por múltiplas camadas de processamento a aprenderem representações de dados com múltiplos níveis de abstração (LECUN; BENGIO; HINTON, 2015). É comum na literatura a definição de redes contendo de 5 a 20 camadas ocultas podendo ter, com bastante frequência, redes com centenas de camadas ocultas.

Segundo Goodfellow et al (GOODFELLOW; BENGIO; COURVILLE, 2016), estes métodos incrementaram dramaticamente o estado da arte da resolução de problemas complexos em domínios científicos variados como reconhecimento de voz, reconhecimento de objetos, aplicações farmacêuticas em desenvolvimento de remédios, pesquisa genômica, dentre tantas outras. Yann LeCun (LECUN; BENGIO; HINTON, 2015) e Nisbet (NISBET; MINER; YALE, 2018) acreditam que as *deep learning*s terão um sucesso ainda maior num futuro próximo porque elas requerem muito menos esforços braçais da engenharia, de forma a facilitar e tirar vantagem dos recentes progressos da capacidade computacional disponível e do crescente e acessível volume de dados de domínios específicos.

2.3.3.1 Arquiteturas de Redes Deep Learning

Dentre as principais e atuais arquitetura de DL empregadas na resolução de problemas da complexidade dimensional de PMs, citam-se: 1) Deep Neural Networks (DNN) ou ainda Deep Artificial Neural network (menos comum), 2) a Convolutional Neural Network (CNN) e, 3) a Deep Recurrent Neural Networks (DRNN).

O texto das subseções a seguir descreve resumidamente cada uma destas arquiteturas de redes.

2.3.3.2 Deep Neural Network

As Deep Neural Networks (DNN) (ou Redes Neurais Profundas, nomenclatura menos frequente) são estruturas semelhantes às redes convencionais (MLP) porém constituídas de muitas camadas ocultas. A Figura 4 apresenta o diagrama de construção de uma DNN.

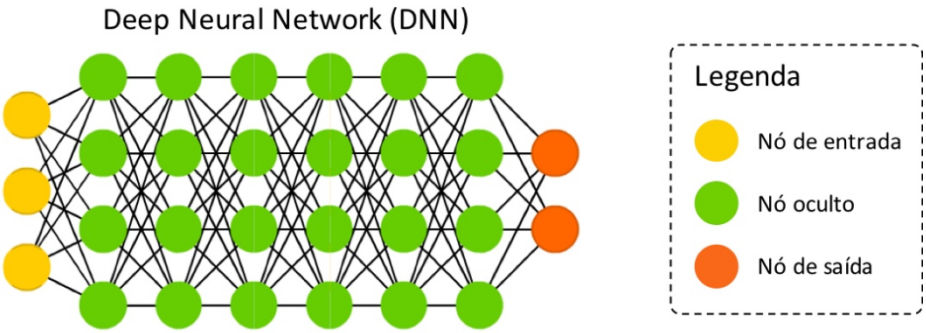


Figura 4. Diagrama de arquitetura de Deep Neural Network (DNN).

O mesmo algoritmo de retropropagação (backpropagation) é utilizado para realizar o treinamento (ajuste dos parâmetros livres) da rede. Uma DNN é aplicada geralmente em problemas de classificação e regressão de maior complexidade, em domínios extremamente não lineares ou para grande quantidade de classes distintas.

A principal dificuldade em utilizar estas estruturas está em compatibilizar sua arquitetura com a dimensão dos dados e a quantidade de parâmetros livres, de tal forma a não dificultar seu treinamento. Algumas técnicas de regularização são aplica-

das com frequência para obter melhores resultados. A referência clássica do Haykin é um bom recurso de compreensão destas arquiteturas (HAYKIN, 1999).

2.3.3.3 Convolutional Neural Network

Responsáveis pela nova revolução na área de pesquisa das ANNs as redes convolucionais foram inicialmente desenvolvidas pelo grupo de pesquisa de LeCun (LECUN; BENGIO; HINTON, 2015) no final da década de 90 para realizar o reconhecimento de dígitos em cheques.

A Figura 5 apresenta um diagrama genérico de CNN. As principais aplicações das redes do tipo CNN estão relacionadas à visão computacional e processamento de imagens. Na área de P&G têm obtido ótimo desempenho em processamento de imagens sísmicas (WALDELAND *et al.*, 2018).

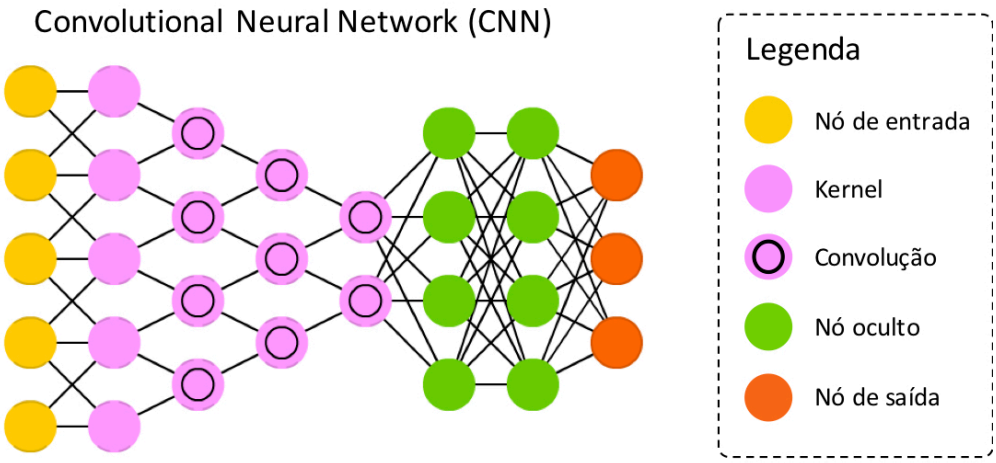


Figura 5. Diagrama de Arquitetura Convolutional Neural Network (CNN).

A ideia das CNNs é a de receber os dado de entrada (geralmente imagens) em forma de matriz ou tensores (matrizes multidimensionais) e então utilizar filtros convolucionais alternados com redutores/aproximadors para gerar um conjunto de caracterizadores mais precisos do dados de entrada. Com estes novos caracterizadores utiliza-se uma ANN convencional para realizar a classificação.

Combinações de duas CNN e outras estruturas são utilizadas para geração de aplicações de codificadores e decodificadores. A Figura 6 apresenta uma versão da arquitetura de uma CNN com imagens de exemplo para melhor compreensão da mesma.

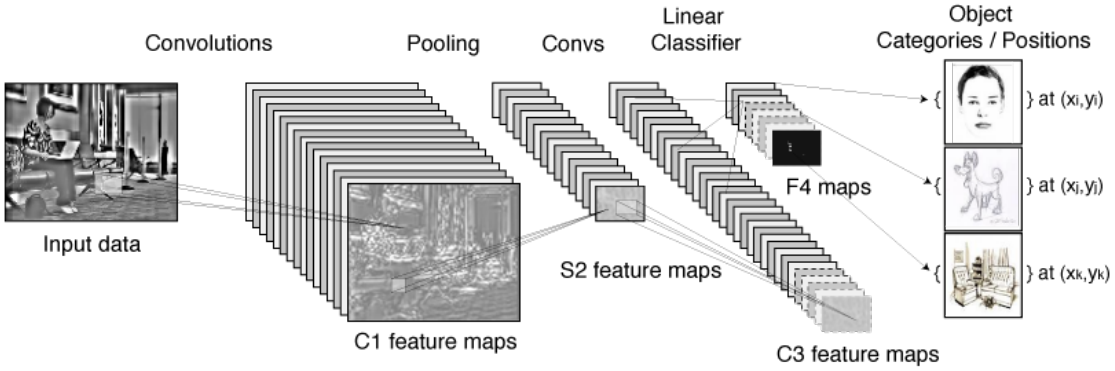


Figura 6. Exemplo de CNN a partir de imagens. Adaptado de (GOODFELLOW, 2106).

As CNNs são costumeiramente aplicadas em sinais bidimensionais mas podem ser aplicados em dados N-dimensionais, como em imagens médicas 3D, vídeos 3D+1(tempo), etc. Referências importantes sobre CNN incluem (GOODFELLOW; BENGIO; COURVILLE, 2016; HUANG; DONG; CLEE, 2017; NISBET; MINER; YALE, 2018; SCHMIDHUBER, 2015).

2.3.3.4 Deep Recurrent Neural Network

A última arquitetura a ser citada é a das redes neurais recorrentes. A Figura 7 demonstra um diagrama reduzido de arquiteturas deste tipo. O diferencial em relação a uma DNN ou mesmo uma MLP simples é que os neurônios das camadas internas (ocultas) podem receber um link que realimentação que representa exatamente a passagem do tempo.

As redes recorrentes são ótimas para tratar problemas de predições de séries temporais, de curto ou longo prazo (com adição de memória, LSTM). São um conjunto de algoritmos muito úteis para processamento de dados sequenciais como som, séries, linguagem, etc.

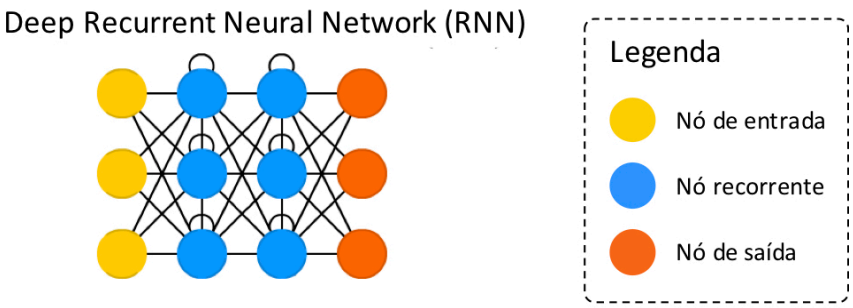


Figura 7. Diagrama de Deep Recurrent Neural Network (DRNN).

Ainda é pouco aplicada no contexto de P&G, aparecendo com maior frequência em predição de produção, previsão de ativos, etc. Em outros domínios de aplicação as redes têm tido excepcional desempenho em agentes autônomos (como em jogos)(MNIH *et al.*, 2013).

Como ela é capaz de incorporar ou aprender a relação entre dados subsequentes, apresenta-se como um ótimo modelo para lidar com problemas dinâmicos, conseguindo responder adequadamente a entradas de dados que dependem do contexto e de seu fluxo antecessor.

Exemplos de aplicações que poderiam se beneficiar destas redes incluem predições e respostas de controle em diferentes perspectivas temporais como os apresentados em Foss et al (FOSS; KNUDSEN; GRIMSTAD, 2018). Para um aprofundamento do tema sugere-se a leitura de (GOODFELLOW; BENGIO; COURVILLE, 2016; MNIH *et al.*, 2013; SERCU *et al.*, 2016).

2.4 SÍNTESE TÉCNICA DAS APLICAÇÕES DE PROXY MODEL EM RESERVATÓRIOS

Um dos resultados da investigação bibliográfica é a consolidação das técnicas, problemas, soluções, escopos e métodos encontrados no corpus em questão. Vinte um (21) trabalhos, se destacaram por se relacionarem mais proximamente com o problema aqui tratado. Destes, três (3) são revisões e dezoito (18) trabalhos primários.

A Tabela 4 a seguir, sintetiza algumas das informações relevantes a cerca de cada um dos trabalhos primários. Seis (6), dos dezoito trabalhos primários, têm aproxi-

mação mais estreita com a proposição desta tese. Destes seis, um terço é publicação em revista científica e os demais são comunicações em conferências. A maior parte dos trabalhos tenta realizar aproximações de respostas a uma granularidade maior das respostas do sistema como as produções individuais dos poços ou as taxas de produção, ou ainda de todo o reservatório. Apesar de alguns utilizarem informações de nível em célula de grid, poucos tentam verdadeiramente predizer a informação futura das propriedades dinâmicas destas.

Com relação ao foco, a maioria tenta predizer saturação de óleo, ocorrendo em menor frequência a produção de água e os valores de pressão. Alguns ainda focam na aplicação de History Matching. Sobre as técnicas, é notável a escolha em maioria absoluta pelo uso de ANN. Com ocorrência pontual de RBFs, Decision Trees, e métodos próprios. Três ocorrências declaram uso de Deep Learning, porém nenhuma delas efetivamente utilizam Percetrons de Múltipas Camadas maiores do que três camadas (incluindo entrada e saída), restando apenas uma aplicação verdadeira de RNN no sentido real de Deep Learning através da técnica de LSTM. Sobre os processos de amostragem, prevalecem as randômicas e a utilização de LHS, ocorrendo em menor frequência a amostragem adaptativa, segmentada, por frequência de erro, pelo método jack-knife e uma com processo próprio e relativo ao objetivo de trabalho. Um trabalho inclui a etapa de otimização e faz uso de Algoritmos Genéticos.

Tabela 4. Sumário das informações mais relevantes dos principais trabalhos relacionados.

Título	T*	Escopo	Foco	Técnica	Métricas	Software	Escopo Dados
A data-driven smart proxy model for a comprehensive reservoir simulation	C	Cell	Simulate reservoir dynamic properties	ANN, Random Sampling	R2, Mapa de propriedades e de erros	-	1,5 milhões de pontos de treinamento, 62 descritores
A Machine Learning Approach to Enhanced Oil Recovery Prediction	C	Field	Predição de EOR	Decision Tree	R2, Curvas de Produção	Eclipse 300, Python	324 simulações, 6 descritores
Accelerating Physics-Based Simulations Using End-to-End Neural Network Proxies: An Application in Oil Reservoir Modeling	J	Field	Resposta a controle e posição de poços	DL/LSTM	Erro relativo (L2)	Eclipse, Tensoflow, OPM	SPE, +200 Mil simulações, 20 descritores
Application of machine learning and artificial intelligence in proxy modeling for fluid flow in porous media	J	Cell	Predição de pressão, saturação e fração CO2	ANN, High Variance Sampling	R2, Mapa de Propriedades e histograma de erros	CMG, GEM	O2CRC Otway Project Pilot Site
Artificial Neural Network Surrogate Modeling of Oil Reservoir: A Case Study	C	Cell	Taxa de produção e produção acumulada para óleo e água	ANN	Histrograma de erros mais mapa de propriedades	-	Campo sintético
Cognitive Data-Driven Proxy Modeling for Performance Forecasting of Waterflooding	J	Well	Predizer produção de óleo cinjeção de água	ANN	Plot de R2 (crossplot) e Treinamento	CMG	-
Converting detail reservoir simulation models into effective reservoir management tools using SRMs; Case study - three green fields in Saudi Arabia	J	Cell/Well	Apresenta o SRM como ferramenta	ANN	Curvas de Produção	Powers (NOC)	Três Campos não nomeados Arábia Saudita
Coupling numerical simulation and machine learning to model shale gas production at different time resolutions	J	Cell/Well	History Matching	ANN	R2, gráfico de produção	Eclipse/Petrel, proprietário ANN	Marcellus shale gas
Developing a Smart Proxy for the SACROC Water-Flooding Numerical Reservoir Simula-	C	Cell	Predizer propriedades de célula em injeção de	ANN, random sampling	Histograma de pressão, mapa de valores so e pressão	CMG	SACROC

tion Model			água		e mapa % erro		
Developing grid-based smart proxy model to evaluate various water flooding injection scenarios	J	Cell	Predição de saturação de óleo e produção com injeção de água	ANN, LHS sampling + análise descritores.	Curva predição para intervalo curto de tempo [80-92] e [92-98] extrapolação	Python	Campo sintético 2D
Development of an adaptive surrogate model for production optimization	C	Well	Predição de NPV	ANN, GA, adaptative sampling, jackknife	Erro relativo, variância e desvio padrão	-	PUNQ3, 3000 simulações
Dynamic Surrogate Reservoir Model with well constraints	C	Well	Predição de produção	ANN, RBF, PCA	R2	BOAST	SPE antigo (1981) 2D
Grid-Based Surrogate Reservoir Modeling (SRM) for Fast Track Analysis of Numerical Reservoir Simulation Models at the Gridblock	C	Cell	Concetual para criação de proxies	-	Mapa de propriedades Pressão e saturação de água	CMG/GEM, ECLIPSE, IDEA	Mattoon Field, Otway CO2, Unamed Giant Oilfield in Middle East
Machine learning for proxy modeling of dynamic reservoir systems: Deep neural network DNN and recurrent neural network RNN applications	C	Field	History Matching	DNN (ANN), RNN	R2, RMSE, CC, MAE	-	SPE Brugge
Pattern recognition and data-driven analytics for fast and accurate replication of complex numerical reservoir models at the grid block	C	Cell	Apenas replicar predições de propriedades	ANN, LHS Sampling, Adaptative Sampling	Erro relativo % e, histograma do erro percentual	CMG, IDEA	Otway Basin CO2
Reservoir simulation and modeling based on artificial intelligence and data mining (AI&DM)	J	Cell/Well	Apenas conceitual	ANN	-	Eclipse, CMG, POWERS	
SimProxy Decision Support System: A Neural Network Proxy Applied to Reservoir and Surface Integrated Optimization	J	Well/Field	Predição de Produção, óleo, gás e água, Otimização poço	DNN (ANN) e RNN, GA, LHS, PCA, LSE, Random	SMAPE, erro das curvas de produção	CMG IMEX, Matlab, C#	104 mil simulações
Time-dependent Neural Network based proxy modeling of SAGD process	C	Field	Predizer taxa E produção acumulado de óleo	ANN (RBF), algoritmo próprio, LHS sampling	Erro relativo % e histograma do erro relativo %	CMG/STARS	Sintético 2D

T* tipo de publicação, J = Journal, C = Conferência.

Sobre as métricas e técnicas de verificação da qualidade dos resultados, a principal é o cálculo de R2 e plotagens de comparação diretas de valores preditos e valores objetivos. A apresentação das curvas de produção para óleo e os mapas de valores por camada e célula. Grande parte apresenta um gráfico de erros absolutos ou erros percentuais e eventualmente um histograma destes valores. Métricas como SMAPE, MAE, RMSE pontualmente.

Sobre os softwares de simulação utilizados para obtenção dos dados e valores objetivos destacam-se: CMG (IMEX/GEM/STARS), Ecplise, Powers (NOC), OPM, IDEA e BOATS. Dentre as linguagens listam-se: Python, Matlab e C#.

Por fim, sobre os datasets, a maioria utiliza modelos numéricos sintéticos artificiais ou baseados em campos reais. Pouco mais da metade com modelos tridimensionais. E aproximadamente metade faz uso de dezenas de milhares de simulações para gerar um único modelo substituto.

Este capítulo apresentou os principais fundamentos teóricos que subsidiam uma melhor compreensão acerca da proposta de tese e experimentos constantes neste documento. Dentre os quais foram tratados: formulações relativas a simulação em reservatório de P&G; os modelos substitutos de simuladores denominados *proxy models*, seus tipos, modelos mais comuns, modo de desenvolvimento e métricas de avaliação; as técnicas de aprendizagem de máquina, seus principais algoritmos e aplicações na indústria do P&G, as mais recentes arquitetura de redes neurais artificiais denominadas *deep learning* (DL) além de algumas de suas principais arquiteturas e, por fim; uma síntese dos trabalhos semelhantes com informações relevantes sobre cada um.

No capítulo seguinte serão tratados os matéria e métodos aplicados na realização dos experimentos, seguido pelos resultados e análise dos mesmos.

3 MATERIAIS E MÉTODOS

Como elementos essenciais para planejamento dos experimentos e consecutiva execução é imperativo o estabelecimento dos materiais necessários associados (modelos, hardware, software) e métodos importantes para condução e avaliação dos mesmos.

Este capítulo se divide em quatro seções. A primeira trata do modelo de reservatório utilizado como campo de estudo dos experimentos. Este campo já foi previamente utilizado em outros estudos do laboratório, sendo bem documentados na literatura e aceitos como *benchmarks* adequados.

A segunda seção trata do fluxo de trabalho e passos a serem empregados na criação e desenvolvimento dos PMs e como eles serão avaliados. A terceira seção trata dos recursos computacionais de hardware e software necessários para concretização dos modelos. A quarta e última explica pormenorizadamente os conjuntos de experimentos realizados, descrevendo motivação, objetivos, materiais, escopo, etc.

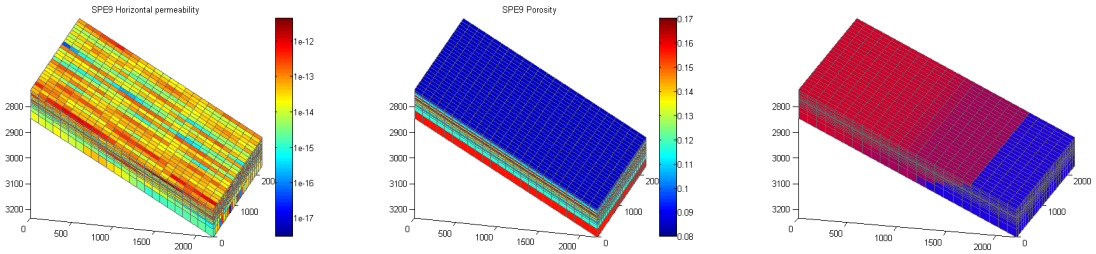
3.1 ESCOPO DE DADOS E MODELOS

Nesta seção apresenta-se o *benchmark* de reservatório de P&G reconhecido e utilizado na literatura científica pertinente e também a estratégia de amostragem de dados a partir de cenários de acordo com os típicos usos na engenharia de reservatórios.

3.1.1 CAMPO SINTÉTICO SPE 9

O Modelo sintético SPE 9 foi criado para desafiar os desenvolvimentos de análises em modelos de software para reservatórios de óleo e gás, como um benchmark aberto. Ele faz parte do SPE – *Comparative Solution Project*, que é uma série de projetos de solução comparativa, organizados pela SPE. A finalidade dos projetos é de proporcionar conjuntos de dados de referência, que possam ser utilizados para comparar o desempenho de diferentes algoritmos ou simuladores.

Em linhas gerais, o SPE 9 é um modelo em grid cartesiano com 24 linhas, 25 colunas e 15 camadas (24 x 25 x 15) totalizando 9000 células ativas. O modelo foi planejado para funcionar em produção por um total de 900 dias, com 25 poços produtores e 1 poço injetor. As figuras (a), (b) e (c) em Figura 8 e a seguir, foram extraídas de [https://www.sintef.no/projectweb/mrst/modules/ad-core/spe9/#7] e apresentam algumas das propriedades petrofísicas ou composicionais do mesmo.



(a) Permeabilidade Hor. (b) Porosidade (c) Sat. Iniciais para óleo e água

Figura 8. Algumas propriedades do modelo SPE 9.

A Figura 9 destaca em especial onde se posicionam os poços de produção e de injeção originalmente, e também destaca que á uma inclinação (dip) em relação ao eixo x de aproximadamente 15 graus.

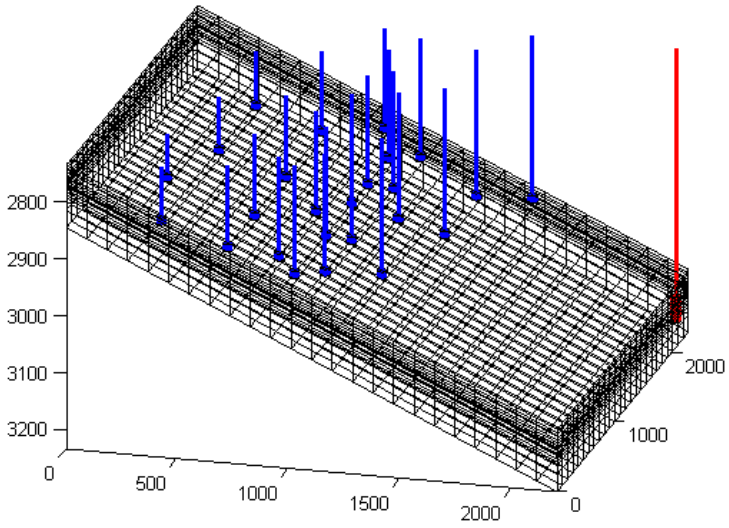


Figura 9. Posicionamento original dos poços produtivos e de injeção, SPE 9.

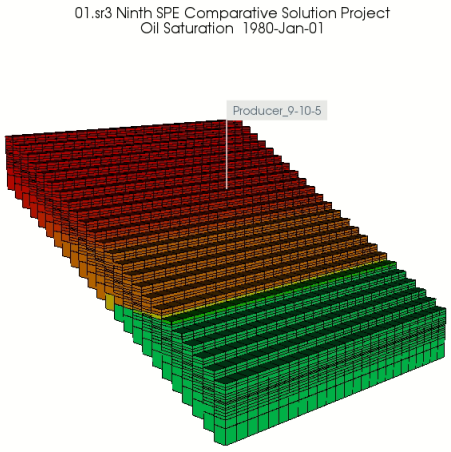
3.1.2 DEFINIÇÃO DE CENÁRIOS

Apesar de o modelo SPE 9 já apresentar configurações iniciais, para este estudo foram preparados quatro (4) diferentes cenários. Em cada cenário um conjunto de realizações de produção (a partir de agora denominadas apenas realizações) foram executadas. Seguem descrições:

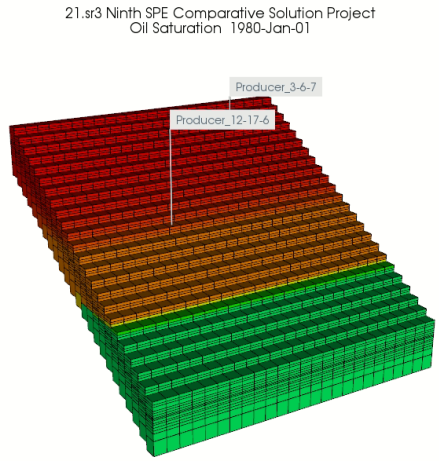
- **Cenário 1: Um Poço Produtor** – neste cenário apenas um poço produtor é introduzido no grid representativo do reservatório. Este primeiro conjunto tem o objetivo de realizar de maneira simples e com uma única direção de fluxo de fluidos. Sendo o conjunto mais simples. A colocação do poço tem as restrições de ser realizada entre as células 1 e 14 do eixo X, 1 e 26 do eixo Y e 2 e 15 do eixo Z. Esta escolha se deve em função da prática lógica de produzir relativamente longe da zona de água.
- **Cenário 2: Dois Poços Produtores** – em sequência ao cenário anterior, este objetiva aumentar a complexidade do sistema ao criar duas direções distintas para o fluxo de fluidos. A localização dos poços segue a mesma restrição com relação às posições i, j, k (eixos X, Y e Z respectivamente), porém acrescida de uma restrição adicional com relação à quantidade mínima de 4 células de distância entre cada uma das coordenadas de cada poço.
- **Cenário 3: Quatro poços Produtores** – para checar o impacto da complexidade ao escalonar a quantidade de poços produtores este cenário foi estabelecido. Além disso com maior quantidade de saída de fluxo a partir do reservatório, torna-se possível melhor perceber a queda de pressão do mesmo e perceber uma maior variação de saturação de todos os fluidos. As restrições são semelhantes ao cenário um, adidas da restrição entre as posições i, j e k de cada um dos poços, neste caso a distância mínima é de duas células para o eixo X (coordenada i) e de três células para o eixo Y (coordenada j).
- **Cenário 4: Um Produtor e Um Injetor** – diferente dos cenários anteriores, neste o objetivo são dois, primeiro de checar o efeito de injeção de

fluxo de fluido, no caso água, afetando a direção de fluxo da produção, e em segundo o efeito da estabilidade da pressão média do sistema, pois com a injeção de líquido o sistema tende a estabilizar melhor.

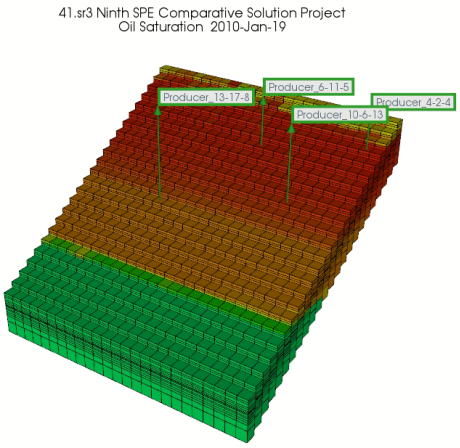
A Figura 10 apresenta exemplos de posicionamentos de poços (produtores e injetor) em cada um dos cenários. Algumas perspectivas foram ajustadas em ângulo para facilitar a identificação das posições dos poços. Em sequência elas representam um exemplo de realização do cenário 1 (a) com poço na posição i, j, k em [9, 10, 5]; cenário 2 (b) com dois poços sendo o primeiro em



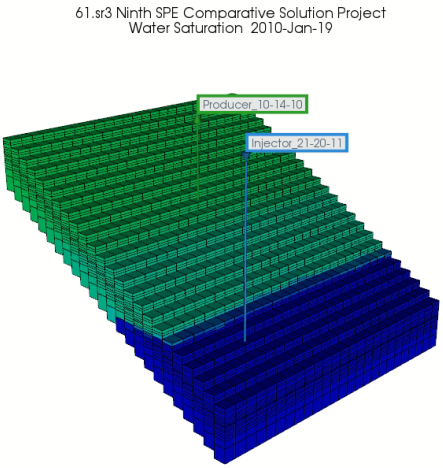
(a) Cenário 1, P1 [9, 10, 15]



(b) Cenário 2, P1 [3, 6, 7], P2 [12, 17, 6]



(c) Cenário 3, P1 [13, 17, 8], P2 [6, 11, 5], P3 [10, 6, 13], P4 [4, 2, 4]



(d) Cenário 4, P1 [10, 14, 10], I1 [21, 20, 11]

Figura 10. Exemplo de realizações para diferentes cenários.

A obtenção de dados para utilização na criação dos experimentos foi realizada a partir do software CMG/IMEX, com auxílio das ferramentas CMG/Report e CMG/Graphics Report. Foi necessário a elaboração de um software para gerir o conjunto de experimentos, a extração e transformação de todos os dados brutos em dados de propriedades estáticas e dinâmicas para cada um dos cenários e das diferentes realizações. Um total de mais de 800GB de armazenamento foram necessários apenas para os dados de simulações.

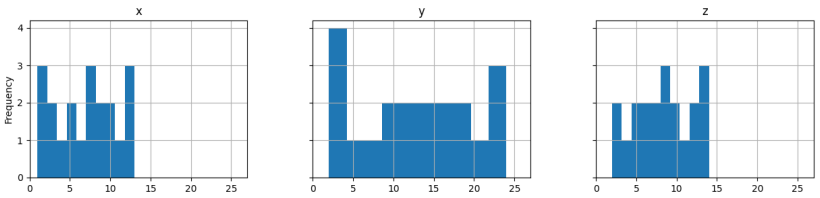
A escolha das posições dos poços em cada cenário foi realizada empregando-se o método de amostragem LHS, pois este tende a garantir uma distribuição uniforme para cada variável livre. Para cada cenário foram realizados vinte (20) realizações, totalizando 80 realizações de produção.

A título ilustrativo a Figura 11 apresenta em (a) os valores para as coordenadas i, j, e k em correspondência aos eixos X, Y e Z do grid, para cada um dos 10 primeiros poços do cenário 1 de simulações e em (b) o histograma de valores ocorridos para cada uma das coordenadas dos 20 poços totais. O Apêndice B. Coordenadas e Histogramas para Poços dos Cenários Experimentais) apresenta todas as coordenadas amostradas com seus respectivos histogramas para cada um dos quatro cenários.

	x	y	z
0	9	10	5
1	9	23	12
2	1	16	2
3	2	12	3
4	2	4	5

	x	y	z
5	5	6	14
6	4	23	9
7	10	17	10
8	11	19	10
9	3	9	14

(a) Coord. cenário 1



(b) Histograma de valores de coordenadas para poços do cenário 1.

Figura 11. Dez primeiras coordenadas de poços cenário 1 e histograma.

3.1.3 AMOSTRAGEM DE DADOS

Para cada uma das realizações são necessários a definição de intervalos de amostragem e quantidade total de dados por intervalo. Neste escopo, foram definidos cinco (5) espaços temporais de trabalho, ou seja, cinco intervalos de tempo e taxa de

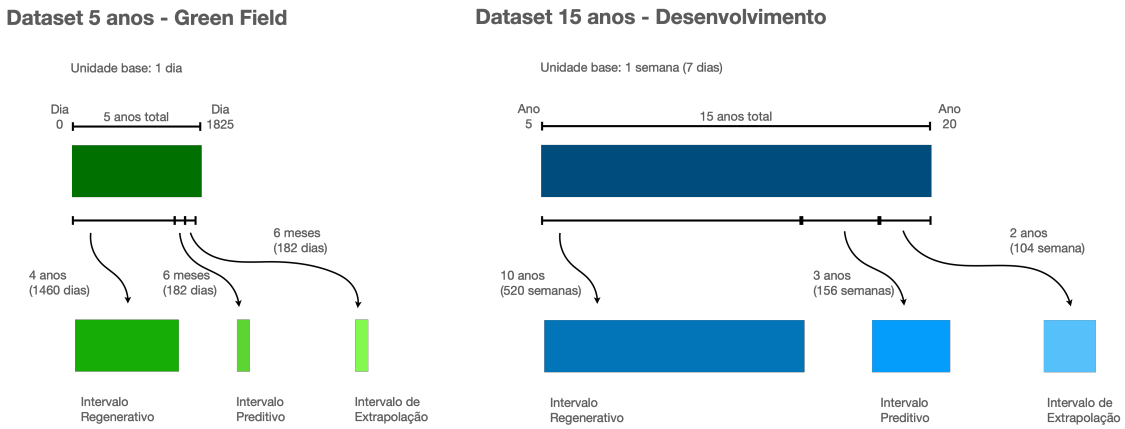
amostragem que se relacionam com os diferentes momentos de desenvolvimento de um reservatório típico.

O intervalo de tempo total ao qual cada simulação (realização) foi executada tem início no dia 0 e finda no dia 10950 (aproximadamente 30 anos). Optou-se por padronizar as contagens em dias para evitar problemas de meses com quantidades de dias distintos e anos bissextos. Desta forma, apenas no escopo de dados do experimento, os intervalos semana tem 7 dias, mês tem 28 dias e ano 364 dias, todos múltiplos de 1 dia e de 7 dias para facilitar efeitos comparativos.

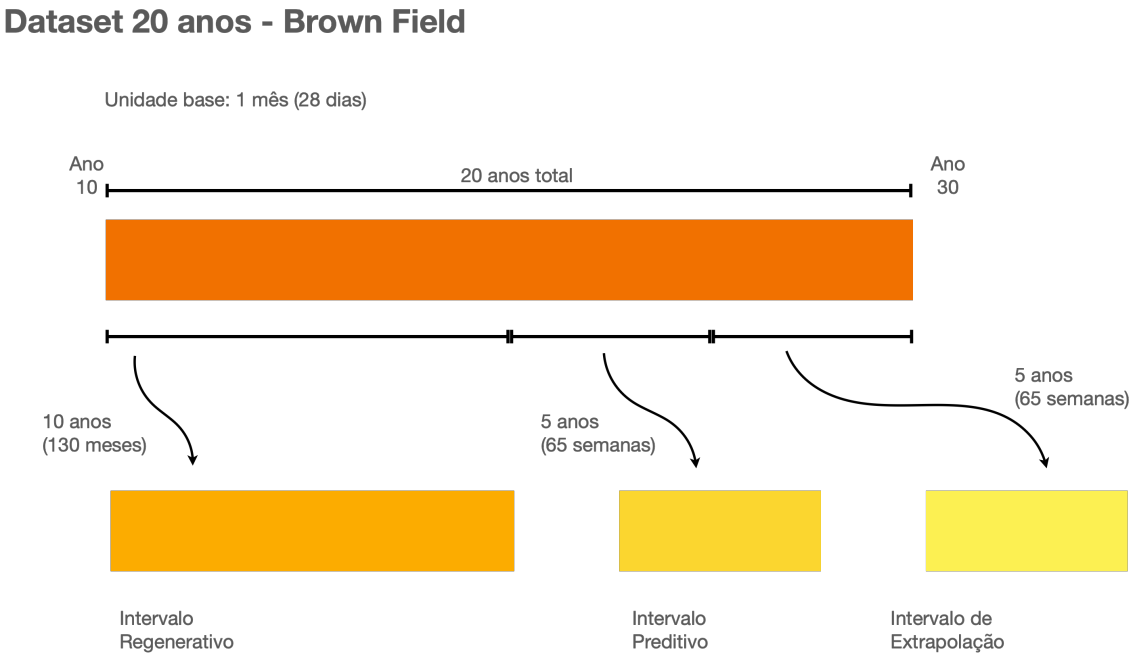
O intervalo total de 30 anos foi dividido em cinco (5) subintervalos, definidos a seguir:

- **Campos Verde** – Intervalo de dados entre o dia 0 e o dia 1825, totalizando 5 anos. Sendo que do dia 0 ao dia 1460 correspondem ao intervalo regenerativo (utilizado para aprender), do dia 1461 ao dia 1642 o intervalo preditivo, e do dia 1643 ao dia 1825 o intervalo extrapolação. A taxa de amostras é diária, totalizando 1825 pontos.
- **Campo em Desenvolvimento** – Intervalo total do dia 1825 ao dia 7300, totalizando 15 anos, do ano 5 ao ano 20. Intervalo regenerativo entre o ano 5 a o ano 15 (520 semanas), preditivo do ano 16 ao 18 inclusive (3 anos, 156 semanas), intervalo extrapolação anos 19 e 20 (104 semanas). A taxa de amostra semanal (7 dias), totalizando 780 pontos.
- **Campo Maduro** – Intervalo total do dia 3650 ao dia 10950, totalizando 20 anos, do ano 10 ao 30. Intervalo regenerativo entre o ano 10 e o 20 (130 meses), preditivo do ano 20 ao 25 (65 meses) e extrapolação do ano 25 ao 30 (65 meses). A taxa de amostra é mensal (28 dias), totalizando 260 pontos.
- **Campo em Desenvolvimento Completo** – Igual ao Campo em Desenvolvimento, porém inicial a partir do dia 0, totalizando 20 anos com taxa amostral semana (7 dias).
- **Campo Maduro Completo** – Igual ao Campo Maduro, porém iniciando a partir do dia 0, totalizando 30 anos com taxa amostral mensal (28 dias).

A Figura 12 ilustra essa segmentação dos dados em cada conjunto de intervalos.



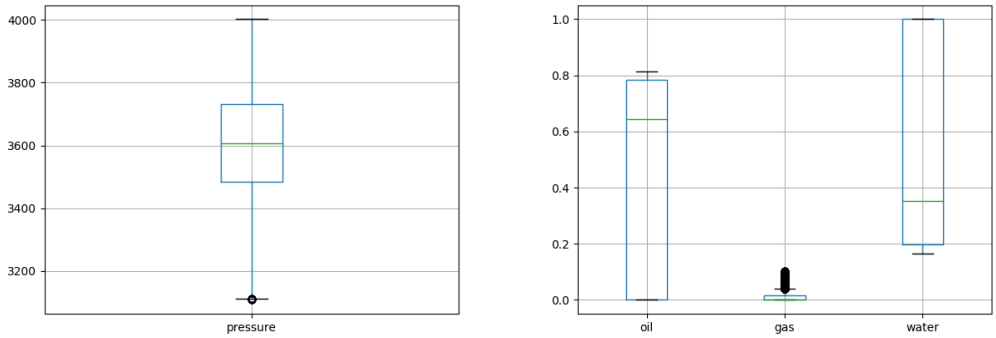
(a) Campo Verde, 5 anos diário (b) Campo em Desenvolvimento, 15 anos semanal



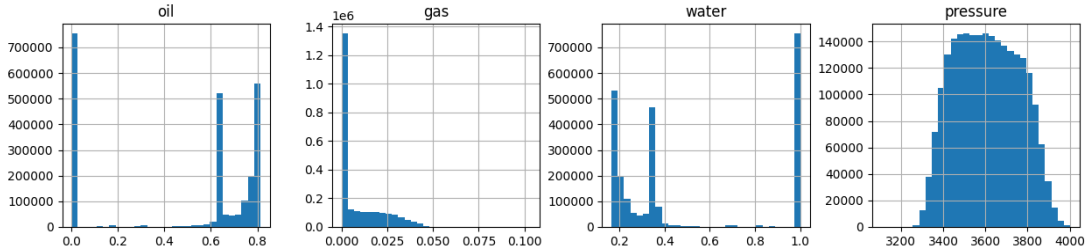
(c) Campo Maduro, 20 anos mensal

Figura 12. Diagramas dos intervalos de segmentação dos dados.

Já na Figura 13 são apresentados (a) os gráficos do tipo boxplot para os valores das propriedades de pressão e saturações de óleo, água e gás, do cenário 1 segmentado pelo subconjunto de amostras do tipo Campo Maduro, e (b) os histrogramas correspondentes. A partir destes gráficos é possível captar informações como intervalo de valores de cada propriedade, se há descontinuidade de valores, ocorrência de pontos fora da distribuição (outliers) e especialmente a compreensão da variação de tais valores ao longo do tempo, quando se compara diferentes segmentos de um mesmo cenário/realização.



(a) Gráfico tipo boxplot para valores de pressão e saturações, cenário 1, segmento campo maduro



(b) histograma para valores de pressão e saturações, cenário 1, segmento campo maduro

Figura 13. Boxplot e Histograma para os valores das propriedades de pressão e saturações do cenário 1 segmentado pelo conjunto amostral de segmento de campo maduro.

3.2 METODOLOGIA A SER EMPREGADA PARA CRIAÇÃO DOS PROXIES

Para o desenvolvimento dos *proxy models* será utilizado o fluxo definido em (ZUBAREV, 2010) adaptado conforme a Figura 14 a seguir. Definido em cinco (5) etapas, o processo se desenvolve em cada uma delas como segue:

1. **Definição das variáveis de entrada –** em cada conjunto de treinamentos o simulador será utilizado com o *benchmark* apropriado (SPE-9) para extrair todo o conjunto de informações da função objetivo desejada correspondente (ex.: produção total acumulada, HBP, NPV, PVT, saturação, etc..) bem como características físicas inerentes a cada célula (saturação, porosidade, permeabilidade) e eventuais informações relacionadas a poços por ventura presentes (distância).

2. **Amostragem do Dataset –** Dado o conjunto total de dados disponibilizado pelos simuladores, um conjunto reduzido denominado de Bando de Dados de Entrada, é desenvolvido utilizando alguma técnica de planejamento experimental (*Design of Experiment*) como sugerido em (ALENEZI; MOHAGHEGH, 2017; GOLZARI; HAGHIGHAT SEFAT; JAMSHIDI, 2015b; MOHAGHEGH, S D *et al.*, 2015) ou utilizando alguma abordagem adaptativa conforme sugerido por (EASON; CREMASCHI, 2014; ZHANG, K. *et al.*, 2017).
3. **Estimação do Proxy Model –** a contribuição principal do trabalho e originalidade encontra-se nesta etapa. Cada conjunto de experimentos utilizará um modelo de arquitetura de DL para validar a viabilidade e qualidade de estimadores robustos. Conforme explanado no capítulo posterior, serão três modelos distintos: Deep Neural Network (DNN), Convolutional Neural Networks (CNN), e Deep Recurrent Neural Network (DRN).
4. **Verificação do Proxy Model –** Das métricas constantes da seção 2.2.2, e considerando as referências (FAWCETT, 2005; GOLZARI; HAGHIGHAT SEFAT; JAMSHIDI, 2015b; ZUBAREV, 2010) a avalição do desempenho utilizará o erro médio quadrático conforme equação: $\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y})^2$.
5. **Melhoria do BD ou do Modelo –** Por fim, em cada conjunto de treinamentos serão avaliados diferentes conjuntos de *datasets* (Bando de Dados) a serem utilizados através do método da validação cruzada (*k-fold*) e diferentes arquiteturas dos modelos propostos também serão avaliadas e definidas a partir de um planejamento experimental k-fatorial.

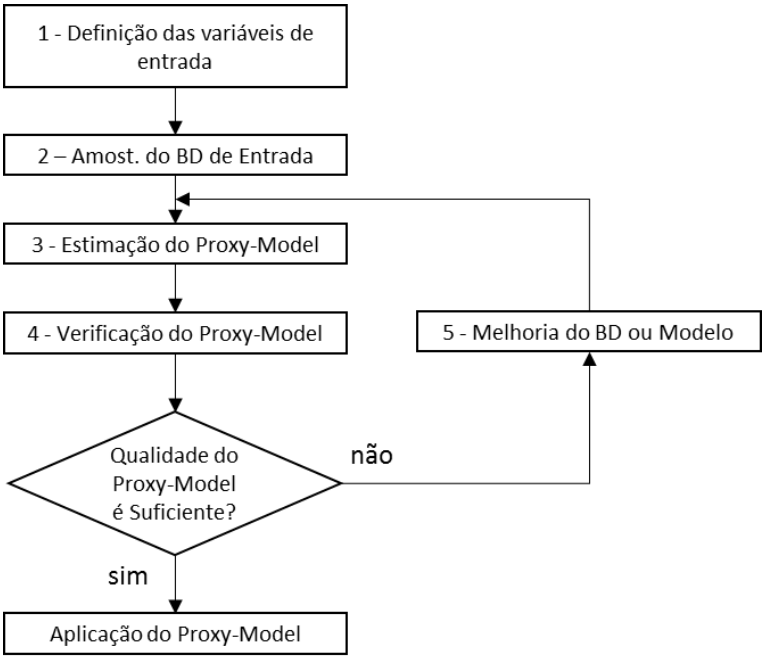


Figura 14. Diagrama do fluxo de etapas na elaboração dos proxy models.

Emprega-se a etapa de análise de sensibilidade tendo em visto que a hipótese contempla a compreensão de robustez dos modelos de estimação empregados. Havendo necessidade de redução do conjunto de dados será empregado, neste caso, algoritmos de análise de componentes principais (PCA – *Principal Component Analysis*).

3.3 FERRAMENTAS COMPUTACIONAIS (HARDWARE E SOFTWARE)

O conjunto de ferramentas computacionais a serem utilizados pelo trabalho compõem-se conforme a Tabela 5 de Hardwares e a Tabela 6 de Softwares.

Tabela 5. Tabela de recursos de Hardware utilizados nos experimentos.

Nome	Descrição	Aplicação / Uso
Workstation de alto desempenho (GPUs) (já adquirido, entrega em janeiro 2019)	192GB RAM DDR4, 4xNVIDIA 1080TI (11GB DDR5), 8-Core i7, 3TB HD	Execução dos simuladores e criação e execução dos *proxy models*
Cluster de alto desempenho (CPUs) (disponível em parceria com UFPB/DCX)	Cluster Xeon 3.2 GHz, 16 núcleos 32 thread, 32GB Ram DDR3	Criação e execução dos *proxy models*

Desktops (CPU+GPU) (já adquiridos, entrega janeiro 2019)	6 X NVIDIA 1060TI 6GB DDR5, 16GB RAM DDR4, i78core	Análise de resultados e execução dos *proxy models*

Tabela 6. Tabela de recursos de Software utilizados nos experimentos.

Nome	Descrição	Aplicação / Uso
Suíte CMG 2015, 2018 e 2020 (já adquirido e disponível)	Suíte simulador de reservatórios completo, multifásico com módulos de modelagem e simulação de fluxos de fluídos pesados e gases.	Geração de dados de entrada (dataset), validação dos modelos, visualização de dados.
Toolbox MRST (opensource disponível)	Toolbox do Matlab especializada em simulação de reservatórios, compatível com modelos comerciais.	Obtenção automatizada de características de células dos reservatórios e visualização de dados.
Matlab 2018 (em aquisição)	Software de computação numérica.	Executar a Toolbox MRST e visualização de dados.
Sci-kit learn (opensource disponível)	Biblioteca de computação de Aprendizagem de Máquina (Machine Learning) em Python	Criação de modelos e fluxos de treinamento junto ao Keras e TensorFlow
Keras (opensource disponível)	Biblioteca de modelos de *deep learning* (alta abstração)	Intermediação para criação de modelos através do TensorFlow
TensorFlow (opensource disponível)	Biblioteca de modelos de DL (baixa abstração)	Modelos de redes neurais DL

Demais bibliotecas e software utilizados são registrados em relatórios técnicos correspondentes e eventualmente relatados em apêndices.

3.4 DEFINIÇÃO DOS EXPERIMENTOS E APLICAÇÕES

Um conjunto de quatro grupos de experimentos e duas aplicações foram planejadas para compor o escopo de análises e resultados. Eles são descritos nas subseções a seguir.

3.4.1 EXPERIMENTO BASELINE

- **Descrição:** Realizar treinamentos de aprendizagem e modelagem de proxies para o cenário 1, no segmento de dados referentes ao campo maduro completo. Para este caso serão utilizados algoritmos clássicos de Machine Learning, incluindo regressão linear e árvore de decisões,

bem como perceptron de múltiplas camadas (simples) para criar uma base de modelos comparativos.

- **Motivação:** Apesar de utilizar um dataset aberto, e este servir como benchmark, não há disponível nenhum conjunto de scripts, dados ou métricas por autores anteriores com o uso do SPE9, inviabilizando a comparação direta. Esta foi a mesma estratégia utilizada por outros autores para avaliar resultados obtidos em relação a alguma métrica razoável.
- **Objetivo:** Criar uma métrica de base para entender se os modelos posteriores são mais robustos e eficazes ou não.
- **Técnicas:** Algoritmos clássicos de Machine Learning, amostragem aleatória, uso de biblioteca em python para agilizar o treinamento
- **Escopo de dados:** Cenário 1, um poço produtor, no segmento de dados campo maduro completo, com amostras a cada 28 dias entre o ano 0 e o ano 25. Tratamento de dados mínimo com escalonamento tipo standard scale.
- **Expectativa de resultado:** Elaborar uma matriz com diferentes algoritmos e métricas de qualidade e erro para cada algoritmo avaliado.

3.4.2 EXPERIMENTO ANÁLISE, SELEÇÃO E CRIAÇÃO DE DESCRITORES

- **Descrição:** Realizar uma análise exploratória dos conjuntos de dados de treinamento, aplicando diferentes técnicas estatísticas e gráficas, de tal maneira a estabelecer critérios e experimentos que possam incrementar a qualidade do treinamento, reduzir o tempo exigido para criação do modelo proxy e também reduzir o consumo de memória ram, tanto no treinamento como na execução.
- **Motivação:** Uma das ausências observadas na literatura foi exatamente um tratamento mais adequado na etapa de pré-processamento e seleção de descritores, abrindo espaço para essa análise no escopo deste estudo.

- **Objetivo:** Avaliar se existe algum processo de pré-processamento, seleção ou criação de descritores no escopo de dados de entrada que melhores o desempenho geral do modelo treinado.
- **Técnicas:** Perceptron de múltiplas camadas (por ter obtido os melhores resultados no experimento anterior e por ser o mais utilizado pelos outros autores), EDA, Gráficos como a correlação cruzada, mapa de calor, análise de componentes principais, escolha sequência de descritores, criação de descritores.
- **Escopo de dados:** Cenário 1, segmento desenvolvimento completo (amostragem semanal – 7 dias).
- **Expectativa de resultado:** A observação de simplificação ou não de descritores, entendimento melhor do conjunto de entrada, e aprofundamento da compreensão dos efeitos físicos do dataset em relação aos efeitos de desenvolvimento do reservatório. Esta é uma contribuição incremental ao corpo de conhecimento da área de estudo da tese.

3.4.3 EXPERIMENTOS DNN

- **Descrição:** Realização modelagem, treinamento e avaliação de modelos Deep Learning, com redes DNN, considerando 3 a 10 camadas ocultas, nos diferentes cenários e segmentos.
- **Motivação:** Inspirando nos experimentos e modelos criados na literatura correlata, este conjunto de experimentos intenta dar um passo além do que já foi realizado, avaliando efetivamente arquiteturas mais complexas de redes neurais.
- **Objetivo:** Checar a capacidade e desempenho de redes do tipo DNN em aprender o escopo de dados em diferentes cenários e segmentos de dados.
- **Técnicas:** DNN, LHS, Random search, Grid search, Keras, Auto-Keras, Plottings. Treinamento considerando regeneração de dados, predição incremental, predição complete, extrapolação.

- **Escopo de dados:** Todos os quatro cenários (1, 2 e 4 produtores, 1 produtor e 1 injetor), todos os segmentos de intervalos de dados (Campos, Verdes, Desenvolvimento, Maduro e Completos). Diferentes amostragens incluindo diária, semanal (7 dias) e mensal (28 dias). Transformar os dados de entrada segundo a descrição N6 apresentada abaixo.
- **Expectativa de resultado:** Métricas de desempenho para modelos DNN, comparação com modelos clássicos (baseline). Gráficos de mapa de propriedades, erros, histograma de erros, qq-plot. Relatar uma contribuição inovadora ao corpo de conhecimento da área de estudo da tese.

Para alimentar os dados de entrada do dataset dos experimentos DNN, foi estabelecido uma estrutura vetorial unidimensional composta de cinquenta e quatro descritores mais quatro descritores por poço presente na realização. Dentre os descritores têm-se relativos a dados estáticos fornecidos (permeabilidades e porosidades), dinâmicos fornecidos (pressão e saturações), calculados em relação à posição da célula dentro do reservatório (distâncias para as fronteiras do reservatório, profundidade da célula em relação ao solo, altura da célula e porosidade líquida), por fim calculados em relação à posição dos poços e célula em questão (distâncias lineares nas direções i, j e k e a distância euclidiana).

Além disso, para cada célula, também foram adicionados ao vetor de entrada, informações relativas às células vizinhas (vizinhança 6) das propriedades estáticas e dinâmicas fornecidas, a saber: permeabilidades i, j, k, porosidade, pressão, saturações de óleo, gás e água. Dentro do grid de células que representam o reservatório, a Figura 15 ilustra cada uma das seis (6) células vizinhas em (a), as vizinhas horizontais (b) e as vizinhas verticais (c).

Como dados de saída foram avaliadas duas possibilidades: predizer o valor futuro das quatro propriedades dinâmicas (pressão e saturações de óleo, gás e água) e; predizer a variação (Δ) da mudança dos valores dessas mesmas propriedades dinâmicas.

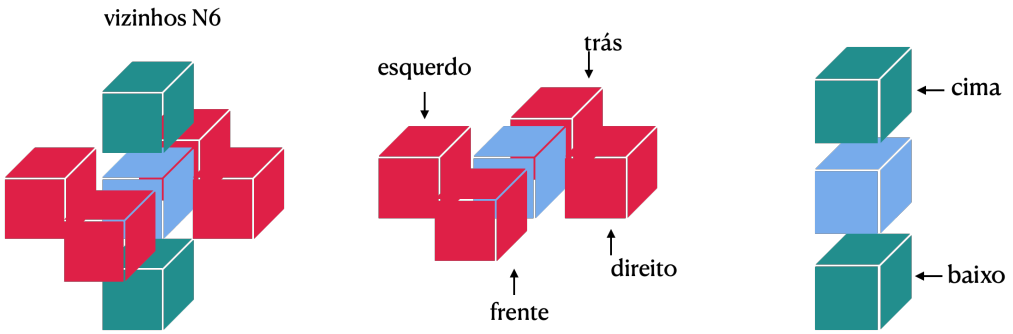


Figura 15. Ilustração de vizinhança 6 para célula em grid cartesiano.

3.4.4 EXPERIMENTOS CNN

- **Descrição:** Realização modelagem, treinamento e avaliação de modelos Deep Learning, com redes CNN, considerando 1 a 6 camadas convolucionais + camada completamente conectada, no cenário 1 e segmento desenvolvimento completo.
- **Motivação:** Inspirando nos experimentos e modelos criados na literatura correlata, este conjunto de experimentos intenta dar um passo além do que já foi realizado, avaliando efetivamente arquiteturas mais complexas de redes neurais.
- **Objetivo:** Checar a capacidade e desempenho de redes do tipo CNN em aprender o escopo de dados em diferentes cenários e segmentos de dados.
- **Técnicas:** CNN, LHS, Random search, Grid search, Keras, Auto-Keras, Plottings. Treinamento considerando regeneração de dados, predição incremental, predição complete, extrapolação.
- **Escopo de dados:** Cenário 1 (1 produtor), no segmento de intervalos de dados (Desenvolvimento). Amostragens mensal (28 dias). Transformar os dados de entrada segundo a descrição C27 apresentada abaixo.
- **Expectativa de resultado:** Métricas de desempenho para modelos DNN, comparação com modelos clássicos (baseline). Gráficos de mapa de

propriedades, erros, histograma de erros, qq-plot. Relatar uma contribuição inovadora ao corpo de conhecimento da área de estudo da tese.

Assim como para a estrutura em DNN, para as estruturas CNN foi elaborada uma estrutura de dados matricial quadri-dimensional para armazenar os dados de entrada da rede. A estrutura escolhida foi uma vizinhança C27, isto é, as células vizinhas ao redor da célula destacada em todas as direções como a vizinhança 6, porém incluindo todas as diagonais, central, inferior e superior, totalizando desta forma 27 células (incluindo a de interesse) para cada descritor de interesse. A Figura 16 ilustra diferentes vistas dessa estrutura matricial.

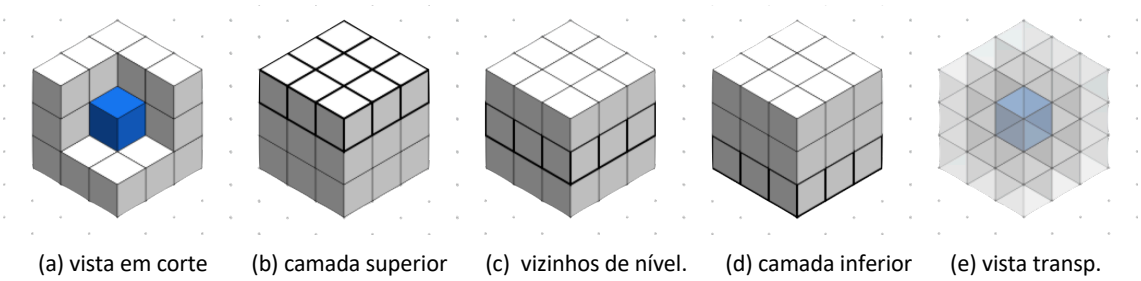


Figura 16. Vista em cortes e de camadas para estrutura C27 para modelos CNN

Para este conjunto de aplicações as variáveis descritivas do modelo, vinte (20) no total, foram agrupadas de acordo com sua origem ou relacionamento, conforme Tabela 7 e estes agrupamentos serviram de critério para avaliação de três proposições de redes CNNs diferentes. Para cada descritor em cada instante de tempo, para cada célula, um cubo de dados (27 células) é instanciado. Ou seja, para uma predição são necessários vinte (20) cubos de dados organizados de uma dentre três proposições descritas em: Figura 17, Figura 18 e Figura 19.

Tabela 7. Descritores por categoria e origem para arquitetura CNN.

Descritor	**Categoria**	**Origem**	**Qt Grupo**
Permeabilidade horiz. (i,j), Permeabilidade k, Porosidade, Profundida-	Física/Petrofísica	Estática fornecida	6

de, Altura da célula e, Volume poroso líquido			
Saturação óleo, gás, água e Pressão	Dinâmicas	Dinâmica fornecida	4
Distância para parede à Esquerda, Direita, Limitação Superior, Inferior, Parede à Frente e ao Fundo	Restrições Reservatório	Estática calculada	6
Distância para o Poço pelo eixo X, eixo Y, eixo Z, e Distância Euclidiana	Fluxo Reservatório	Estática calculada	4

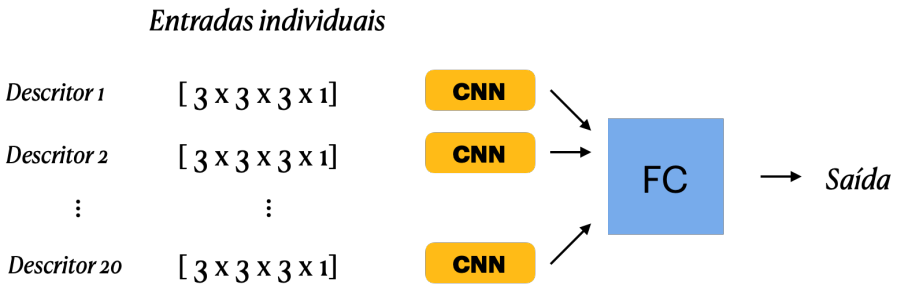


Figura 17. Proposta CNN 1, um conjunto de convoluções para cada cubo descritor.

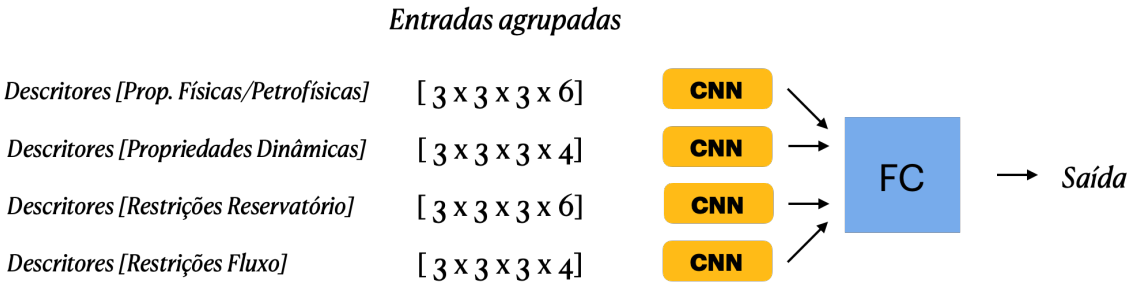


Figura 18. Proposta CNN 2, um conjunto de blocos CNN para cada agrupamento de descritores.

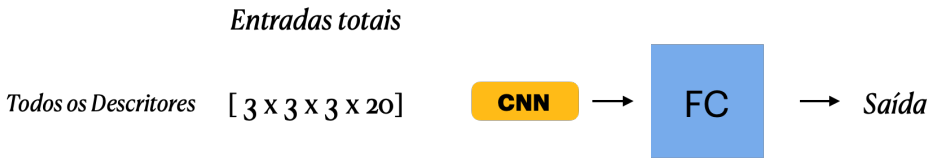


Figura 19. Proposta CNN 3, um único bloco de CNNs para todo o conjunto de descritores.

Ainda para auxiliar no processo de compreensão dos dados, a Figura 20ilustra em formato de blocos 3D como os segmentos cúbicos serão utilizados para treinar uma rede CNN 3D. Uma detalhe importante sobre as CNN, segundo (LECUN; BENGIO; HINTON, 2015), é que elas têm a capacidade de priorizar alterações dos dados de entrada mais importantes do que outros, ou seja, enfatizar o que afeta mais predominantemente a saída objetivo do que variações espúrias ou não correlacionadas diretamente.

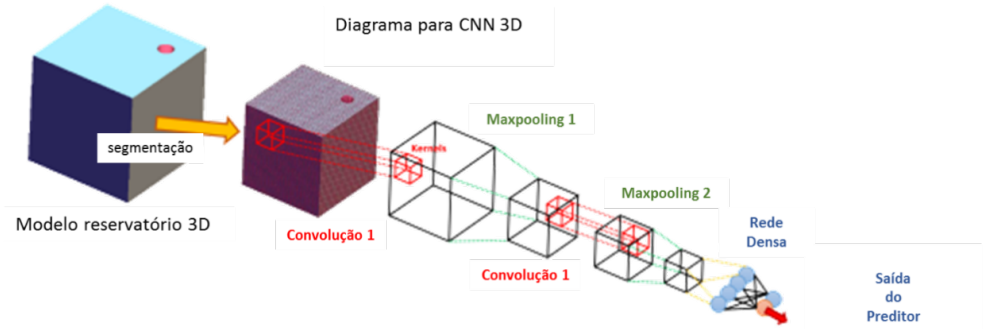


Figura 20. Diagrama ilustrativo de como usar blocos 3D para projeto de CNN 3D do grupo experimental CNN.

3.4.5 APLICAÇÃO PRÁTICA – PREDIÇÃO DE CURVAS DE PRODUÇÃO

- **Descrição:** Utilizando um modelo previamente treinado com bom índice de desempenho, realizar o cálculo e a demonstração gráfica da totalização para as curvas de produção de óleo, gás, água e variação de pressão em nível de reservatório.
- **Motivação:** A criação de um proxy model de um reservatório tem o intuito principal de checar qual o estado futuro de suas propriedades dinâmicas, mas é a partir do cálculo destes valores de variação ao longo tempo que o modelo disponibilizará melhor retorno ao gestor,
- **Objetivo:** Calcular a curva de produção em nível de reservatório a partir das variações de saturações em cada célula, em cada intervalo de tempo.

- **Técnicas:** Uso do modelo, gráficos, funções de totalização e comparação com dados do simulador.
- **Escopo de dados:** Cenário, segmento e taxa de amostragem à escolha baseado no modelo treinado.
- **Expectativa de resultado:** Relatar uma aplicação prática exposta como resultado dos modelos desenvolvidos e rapidamente portável para uso em ambientes de produção, engenheiros e gestores.

Este capítulo apresentou os materiais e métodos associados e necessários à concepção dos conjuntos de experimentos a serem realizados para comprovação das hipóteses. No capítulo seguinte serão apresentadas as propostas de conjuntos de experimentos, seguido pelas considerações finais.

4 RESULTADOS E DISCUSSÃO

Neste capítulo são expostos os experimentos associados ao desenvolvimento do trabalho de Tese. Um primeiro experimento exposto na seção 4.1 trata-se do passo inicial para criação de uma base de resultados para efeito de comparação. A seção 4.2 detalha o experimento de compreensão e análise de descritores. A seção 4.3 e 4.4 na criação de modelos utilizando DNN e CNN respectivamente, e por fim a seção 4.5 demonstra a aplicação dos modelos em efeitos práticos como a predição de produção e a comparação com diferentes posições.

4.1 EXPERIMENTO BASELINE

O primeiro experimento trata-se da criação de um conjunto de resultados para efeito comparativo. Foram criados e treinados diferentes modelos de regressão clássicos da biblioteca scikit-learn. O critério de escolha foi a capacidade de trabalhar com múltiplos descritores (features). Foi utilizado a biblioteca auto-scikitlearn para executar o treinamento e avaliação. O cenário de treinamento foi para o primeiro (1 produtor) no segmento desenvolvimento completo e para realizar a predição completa.

Tabela 8. Resultados de baseline para algoritmos clássicos.

Nome Modelo	**R2 Score Treino**	**R2 Score Test**
Lasso	0.75	0.52
ElasticNet	0.72	0.5
SGDRegressor	0.78	0.63
MLPRegressor	0.83	**0.75**
SVR	0.86	0.72
KernelRidge	**0.87**	0.73

Especificamente para o modelo MLPRegressor um critério de utilizar no máximo duas camadas ocultas foi estabelecido para que o efeito de aprofundamento das camadas internas pudesse ser comparável ao que existe na literatura.

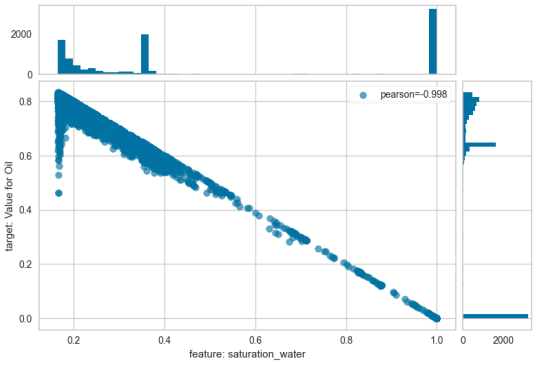
Observa-se uma limitação clara na capacidade desses modelos para conseguir predizer valores mais precisos. Apesar dos primeiros três modelos (Lasso, ElasticNet e SGDRegressor) tentarem reduzir efeito de overfitting, isso aparentemente induz numa redução da capacidade de generalização. Efeito contrário acontece nos três últimos modelos (MLPRegressor, SVR e KernelRidge) quem que um resultado melhor acontece para o conjunto de treinamento e uma redução maior acontece nos dados de teste, indicando overfitting.

4.2 EXPERIMENTO ANÁLISE, SELEÇÃO E CRIAÇÃO DE DESCRITORES

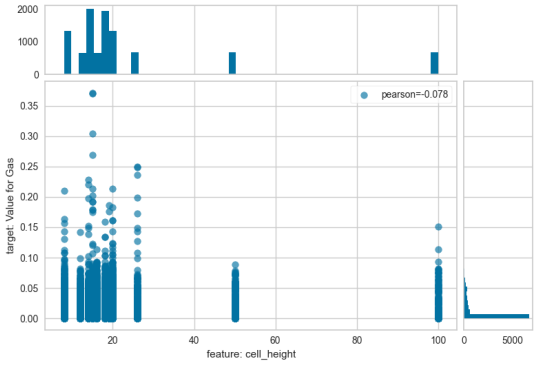
Para cumprimento deste experimento seis (6) técnicas foram empregadas, nomeadamente: plot conjunto (jointplot), plot de relação descritor saída, redução de dimensionalidade manifold, gráficos radviz, ranqueamento bidimensional e, visualização de componentes PCA. Em sua maioria, apenas algumas em casos específicos são relatadas a seguir devido a grande quantidade de resultados obtidos, outros podem ser acessados nos Apêndices.

4.2.1 PLOT CONJUNTO (JOINTPLOT) PARA DESCRITORES E SAÍDA DESEJADA

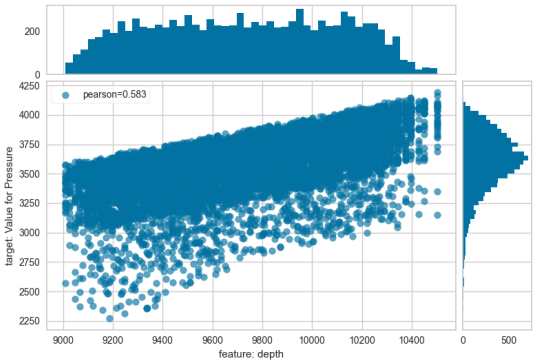
Após a obtenção dos descritores estatísticos do dataset (média, variância, etc.), um dos primeiros passos típicos de uma análise exploratória é a visualização gráfica das relações cruzadas entre os descritores e as saídas objetivos. A Figura 21 a seguir apresenta apenas algumas dessas Relações visuais (dado a enorme quantidade) e o Apêndice E. Plot Conjunto (JointPlot) para Descritores e Saídas Desejada, apresenta apenas para a saída Óleo as relações para vinte e dois (22) descritores.



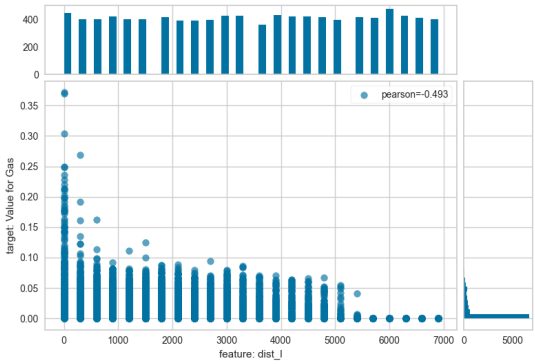
(a) Relação Óleo e Saturação Água



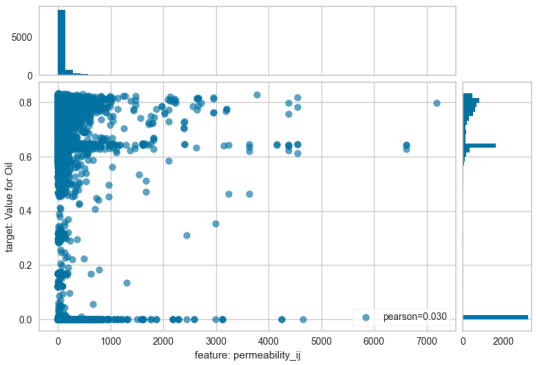
(b) Relação ótleo e Altura da Célula



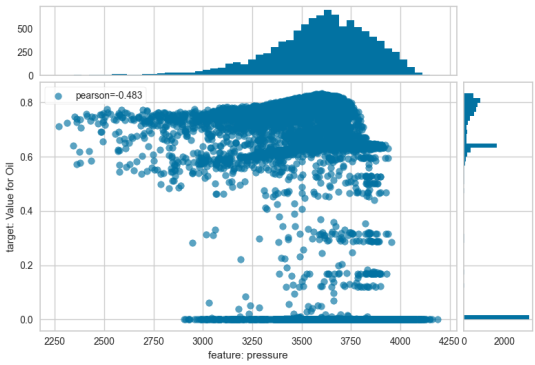
(c) Relação Pressão e Profundidade



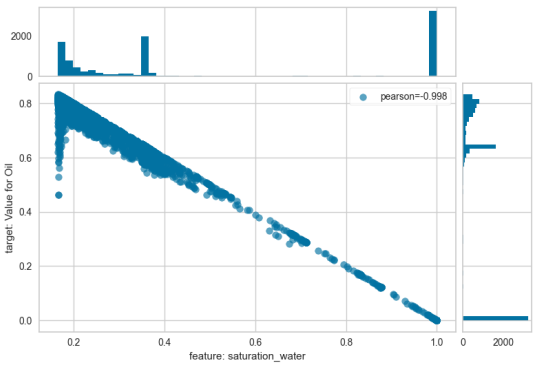
(d) Relação Gás e Distância à Esquerda



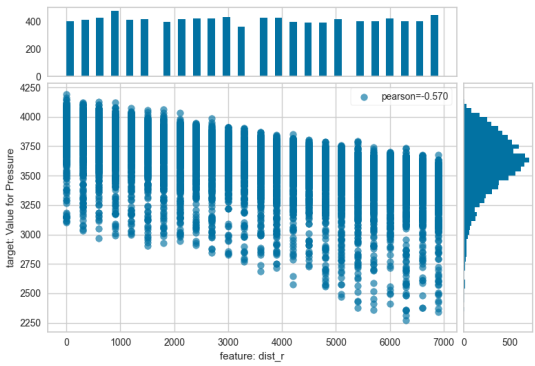
(e) Relação Óleo e Permeabilidade Horizontal



(f) Relação Óleo e Pressão



(g) Relação Óleo e Saturação de Água

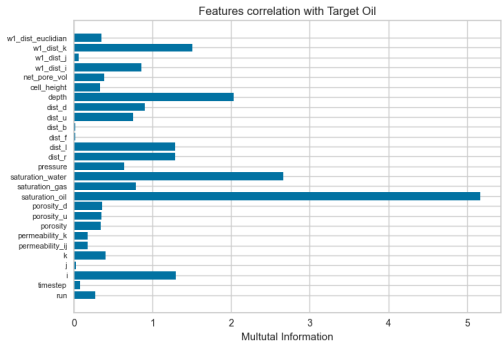


(h) Relação Pressão e Distância à Direita

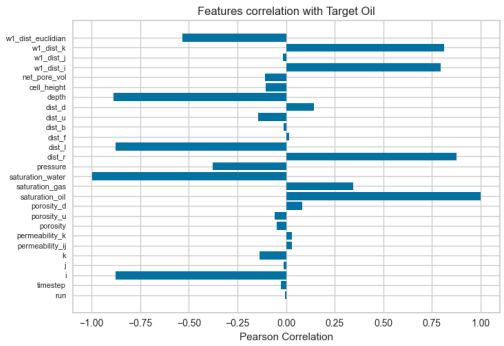
Figura 21. Gráficos de Relação entre Descritores e Saídas selecionadas.

4.2.2 VISUALIZAÇÃO DOS DESCRITORES EM CORRELAÇÃO ÀS SAÍDAS DESEJADAS

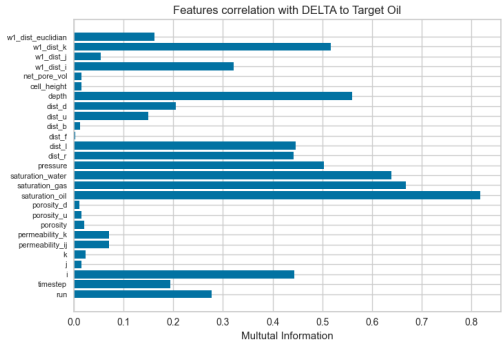
Os seguintes gráficos explicitam a dependência existente entre variáveis dependentes e independentes (descritores e saídas). São utilizados para seleção de descritores e utilizam a correlação de Pearson e Informação Mútua. Foram realizados para cada uma das quatro variáveis de saída tanto para o valor objetivo quanto para o valor de diferença (incremento). A Figura 22 apresenta apenas para saturação de óleo as visualizações para valor de saída, valor de incremento, para correlação de Pearson e informação mútua. Para os demais valores de saída consultar o Apêndice D. Correlação entre Descritores (Pearson e Informação Mútua)



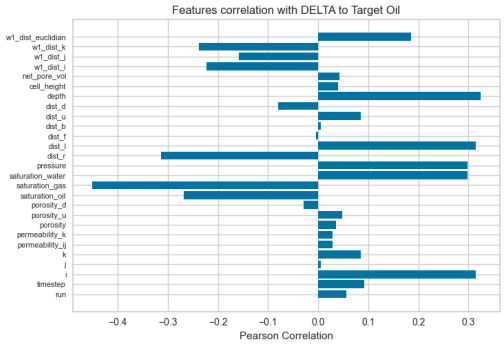
(a) Informação Mútua saturação óleo



(b) Correlação de Pearson saturação óleo



(c) Informação Mútua Delta saturação óleo



(d) Correlação de Pearson Delta saturação óleo

Figura 22. Visualização de correlação Pearson e Informação Mútua.

4.2.3 VISUALIZAÇÃO DE DADOS MULTIDIMENSIONAIS (MANIFOLD)

A técnica de Manifold, permite observar dados multidimensionais em duas dimensões, captando especialmente estruturas não-lineares. As projeções criadas por esta técnica permitem conduzir a uma análise de separabilidade dos dados. A técnica utiliza de algoritmos intermediários como métrica de separação, destes, três retornaram resultado interessantes: Isomap, Multi-dimensional Scaling, t-SNE. As Figura 23, Figura 24 e Figura 25 correspondem respectivamente a visualizações aplicando Isomap, mds e t-SNE. Em todas é possível observar agrupamentos pontuais espacialmente e em escala de valores, sendo especialmente a primeira e a última com maior quantidade de subgrupos.

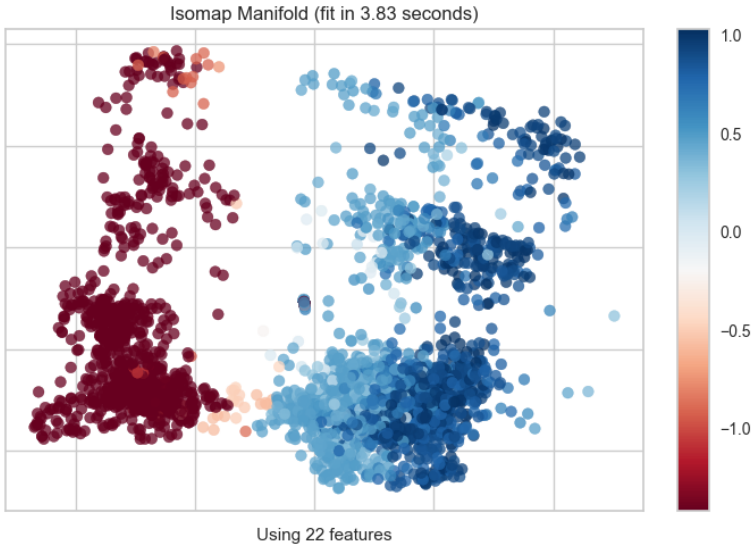


Figura 23. Visualização multidimensional Manifold com aplicação de Isomap.

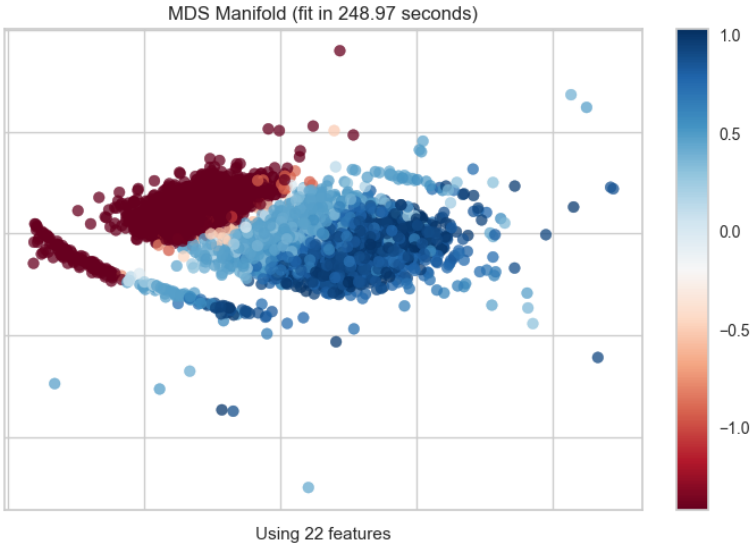


Figura 24. Visualização multidimensional Manifold com aplicação de MDS em 22 descritores.

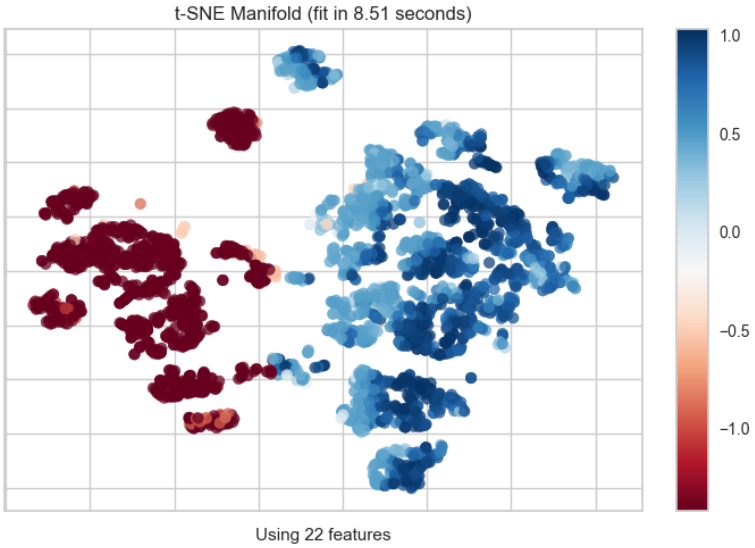
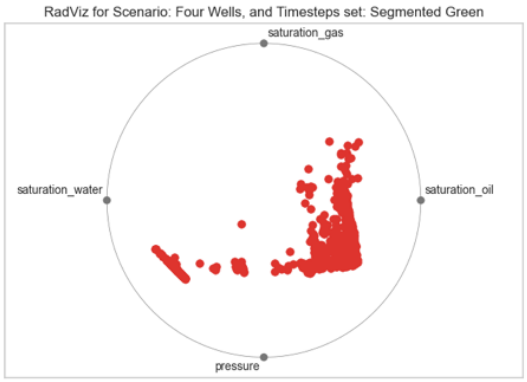


Figura 25. Visualização multidimensional Manifold com aplicação de t-SNE em 22 descritores.

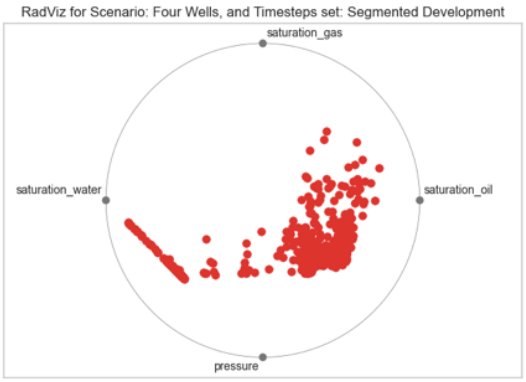
4.2.4 VISUALIZADOR RADVIZ

Esta técnica permite entender se existe uma ou mais tendência de aproximação entre descritores e função objetivo a partir de seus valores. Isto permite checar prioridades ou impactos maiores entre descritores. A Figura 26 apresenta a visualização da

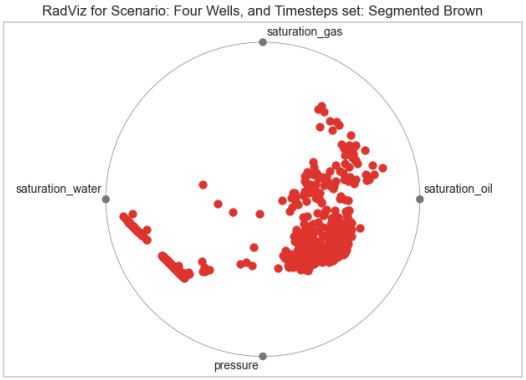
relação dos quatro principais descritores (saturação de óleo, gás e água, pressão) em relação à saída desejada de saturação de óleo para o caso do cenário 1 (um poço produtor).



(a) Radiz Óleo: Cenário 3 Segmento Verde



(b) Radiz Óleo: Cenário 3 Segmento Desenvolvimento



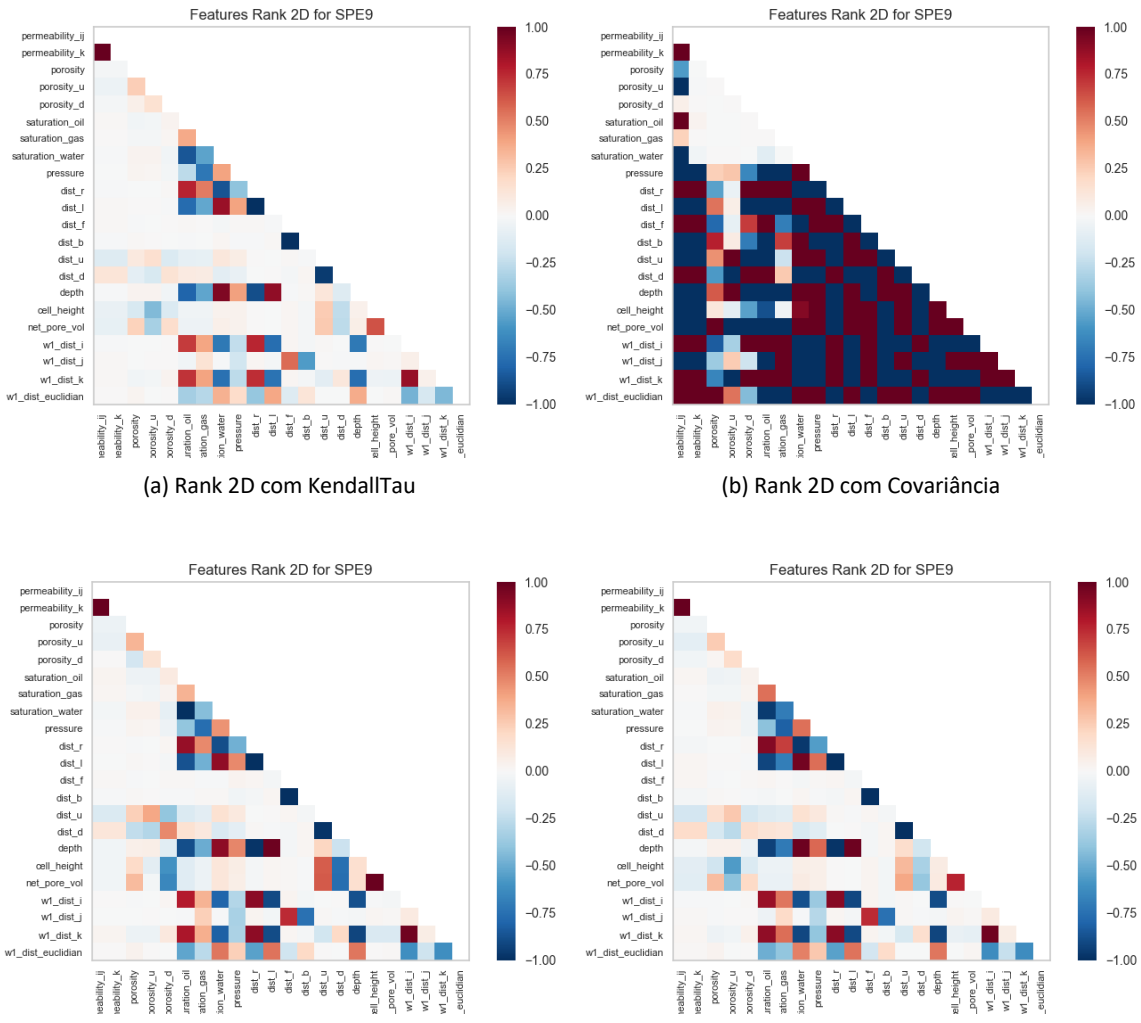
(c) Radiz Óleo: Cenário 3 Segmento Maduro

Figura 26. Radviz para Saturação de Óleo em diferentes Segmentos de Dados para o Cenário 3.

Há uma tendência de aproximação maior entre os valores de saturação de óleo e pressão nas amostras iniciais da produção do reservatório, e uma expansão em direção à saturação de gás e ao centro conforme as amostram alcançam espaço de tempo maior, não chegando a alterar completamente a tendência de aglomeração entre pressão e saturação de óleo. Deixa claro também uma expansão nos valores predominantemente saturados em água.

4.2.5 RANQUE DE DESCRITORES (RANK FEATURES)

O ranqueamento de descritores em função de uma função objetivo de saída, permite perceber correlações par a par (no caso do Rank 2D) entre si e a saída. É interessante para perceber correlações forte na contribuição positiva ou negativa da saída desejada. A **FIGURA** apresenta quatro opções de algoritmos (Pearson, Covariância, Kendall Tau e Spearman) e como pares de descritores se relacionam em função da saturação de Óleo.



(a) Rank 2D com KendallTau

(b) Rank 2D com Covariância

(c) Rank 2D com Pearson

(b) Rank 2D com Spearman

Figura 27. Rank 2D para Saturação de Óleo e 22 descritores.

4.2.6 PROJEÇÃO DE COMPONENTES PRINCIPAIS (2 E 3 COMPONENTES)

A projeção PCA tem como objetivo utilizar o algoritmo PCA para decomposição da alta dimensionalidade de descritores (no caso 22) em duas ou três componentes

(2D e 3D). Ela informa ainda quais componentes têm maior contribuição positiva ou negativa. A Figura 28 apresenta a projeção 3D com três principais componentes e a projeção dos descritores em relação às contribuições. É possível perceber agrupamento por cor e região.

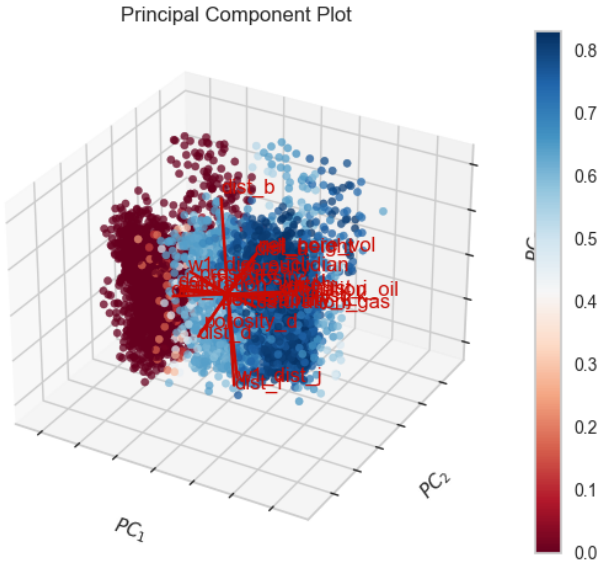


Figura 28. Projeção PCA em 3D com projeção de componentes.

Já Figura 29 é uma projeção 2D, com maior facilidade de detectar as principais contribuições (descritores) positivamente ou negativamente para cada componente. Estas informações foram utilizadas para avaliar experimentos de treinamento com redução de descritores.

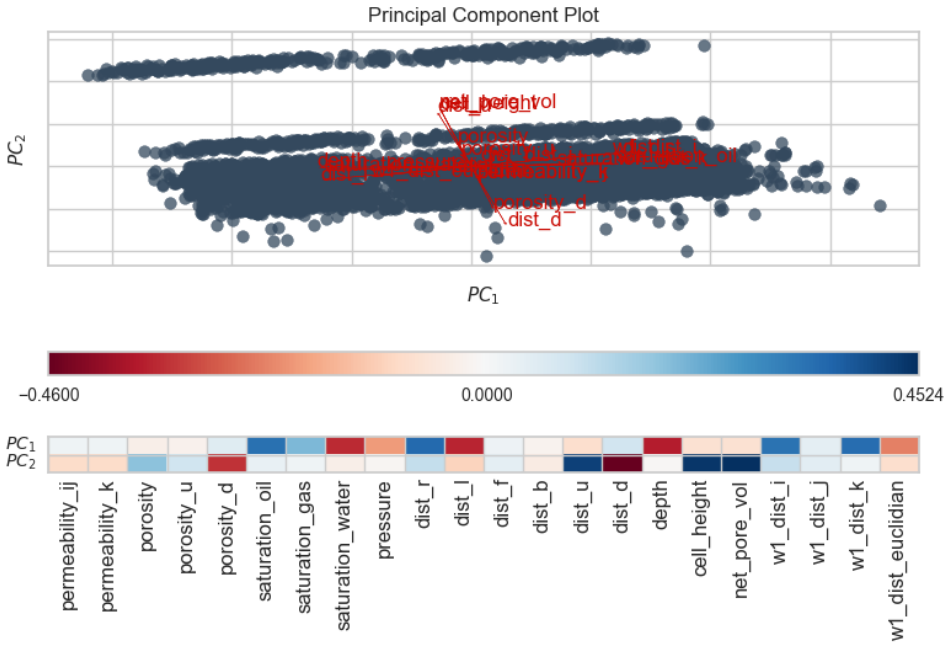


Figura 29. Projeção PCA 2D com projeção dos descritores e legenda de contribuições.

Algo muito relevante foi a compreensão geral sobre as relações entre variáveis e a percepção direta das correlações físicas. Um ponto a destacar é a importância do desacoplamento de variáveis não físicas, das relacionadas a aspectos físicos, em especial as variáveis que indicam índices (i, j, k, timestep), das demais.

A expectativa inicial era de que, utilizando variáveis selecionadas a partir do exposto, fosse possível reduzir o escopo de dados e agilizar o treinamento de modelos. Entretanto, após realizar experimentos de treinamento simples com ANN, não foi possível comprovar através da análise de sensibilidade que: 1) houve aumento de qualidade no resultado final para nenhum das saídas (pressão ou saturações) e, 2) houve redução do erro no processo de treinamento. Restando como vantagem apenas a redução do uso de processamento e de memória, porém isso foi atingido a custo de redução de qualidade dos resultados, inviabilizando, portanto, sua aplicação.

4.3 EXPERIMENTOS DNN

Esta série de experimentos trata do treinamento e avaliação de modelos do tipo Deep Learning DNN. As DNNs se distinguem no contexto contemporâneo das clássicas ANNs por ser constituída por três ou mais camadas ocultas e uso avançado de algoritmos de otimização e funções de ativação lineares ou semi-lineares (relu) ao invés das clássicas (sigmoide, tangente hiperbólica) por serem mais custosas computacionalmente.

Inicialmente foram realizados uma bateria de experimentos utilizando planejamento experimental (DoE) junto ao GridSearch e também RandomSearch para ajuste das arquiteturas e dos hyper-parâmetros de treinamento das Redes. Entretanto, após a avaliação de alguns estudos (BERGSTRA; BENGIO, 2012; JIN; SONG; HU, 2018; LIASHCHYNSKYI; LIASHCHYNSKYI, 2019) a estratégia de treinamento foi alterada para focar na modelagem do problema e entender a etapa de treinamento, ajuste e configuração como processo mecânico inerente ao operador final do modelo. Foram realizadas tentativas de uso com otimizadores (Tspot, nevergrad) porém optou-se pelo uso do Auto-Keras como definidor automatizado dos parâmetros. Para isso foi definido limites para os parâmetros de definição da arquitetura e também dos hyper-parâmetros.

A Tabela 9 apresenta o resultado de um retreinamento realizado com as melhores arquitetura encontradas pelo AutoKeras. Para cada cenário e cada segmento de dados o AutoKeras foi executado com os parâmetros estabelecidos e a partir da arquitetura fornecida foram realizados 10 treinamentos. Nesta tabela constam o valor médio de R2 encontrado, o desvio padrão do mesmo bem como os menores e maiores valores obtidos.

Tabela 9. Resultados para re-treino das melhores redes encontradas no experimento.

Cenário	Segmento Dados	Média R2	Desvio Padrão	(Min, Máx)
1 (um produtor)	Verde	0.95	0.018	
	Desenvolvimento	**0.982**	0.013	(0.9603, **0.999**)
	Maduro	0.975	0.135	(0.952, 0.998)
2 (dois produtores)	Verde	0.916	0.033	(0.866, 0.984)
	Desenvolvimento	0.951	0.012	(0.934, 0.978)
	Maduro	0.96	0.008	(0.945, 0.973)

3 (quatro produtores)	Verde	0.867	0.035	(0.813, 0.908)
	Desenvolvimento	0.900	0.014	(0.858, 0.950)
	Maduro	0.883	0.024	(0.870, 0.905)
4 (um produtor e um injetor)	Verde	0.94	0.031	(0.898, 0.986)
	Desenvolvimento	0.965	0.014	(0.945, 0.985)
	Maduro	0.970	0.010	(0.949, 0.985)

A seguir uma sequência de imagens apresentam gráficos correspondentes ao processo de treinamento e avaliação para um resultado em particular de treinamento para cenário 1 e segmento desenvolvimento. A arquitetura desta rede em especial é de [Entrada, Dense 256, Dense 128, Dense 4], relu como funções intermediárias, linear na saída e otimizador RMSprop(0.001). Trainando todas as saídas ao mesmo tempo.

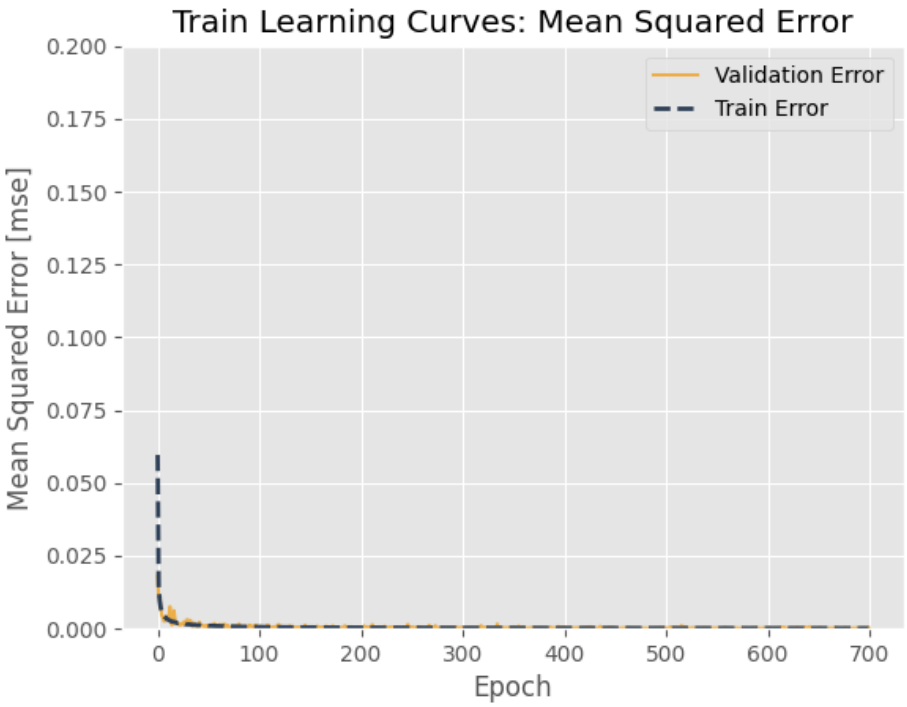


Figura 30. Curva de treinamento e validação para MSE.

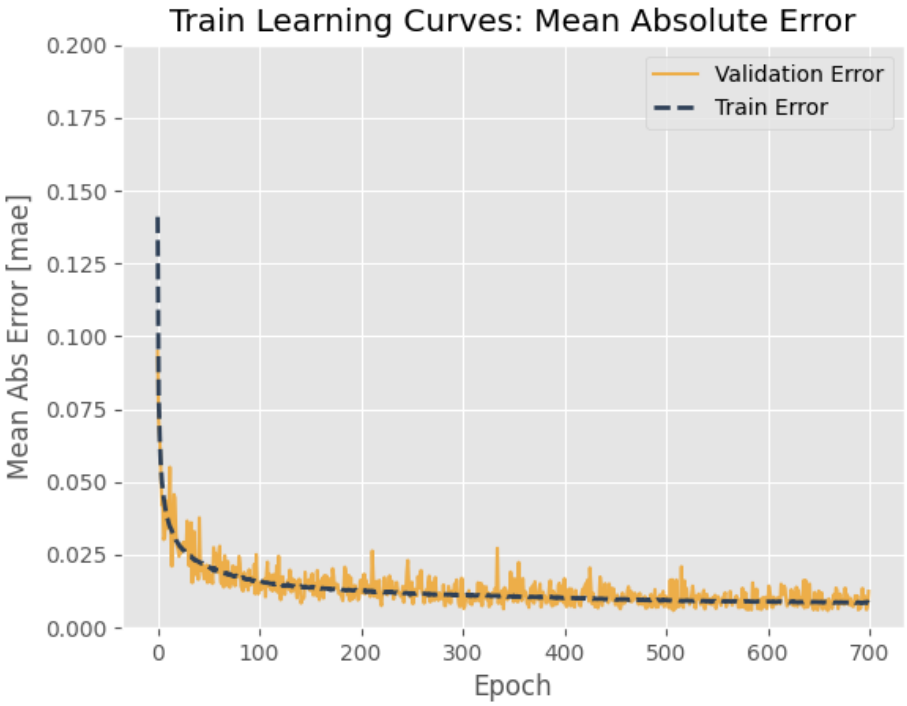


Figura 31. Curva de trainamento e validação para MAE.

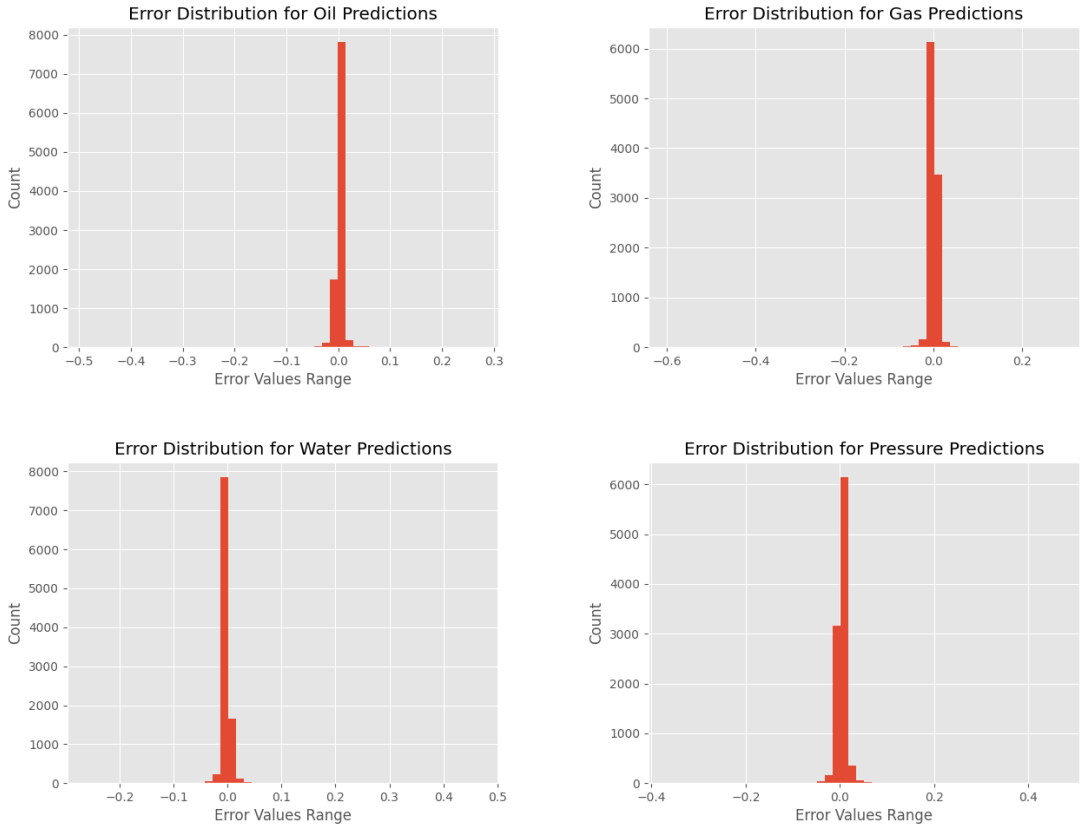


Figura 32. Histograma da distribuição dos valores de erro para cada umas das saídas (Saturações Óleo, Gás, Água e Pressão).

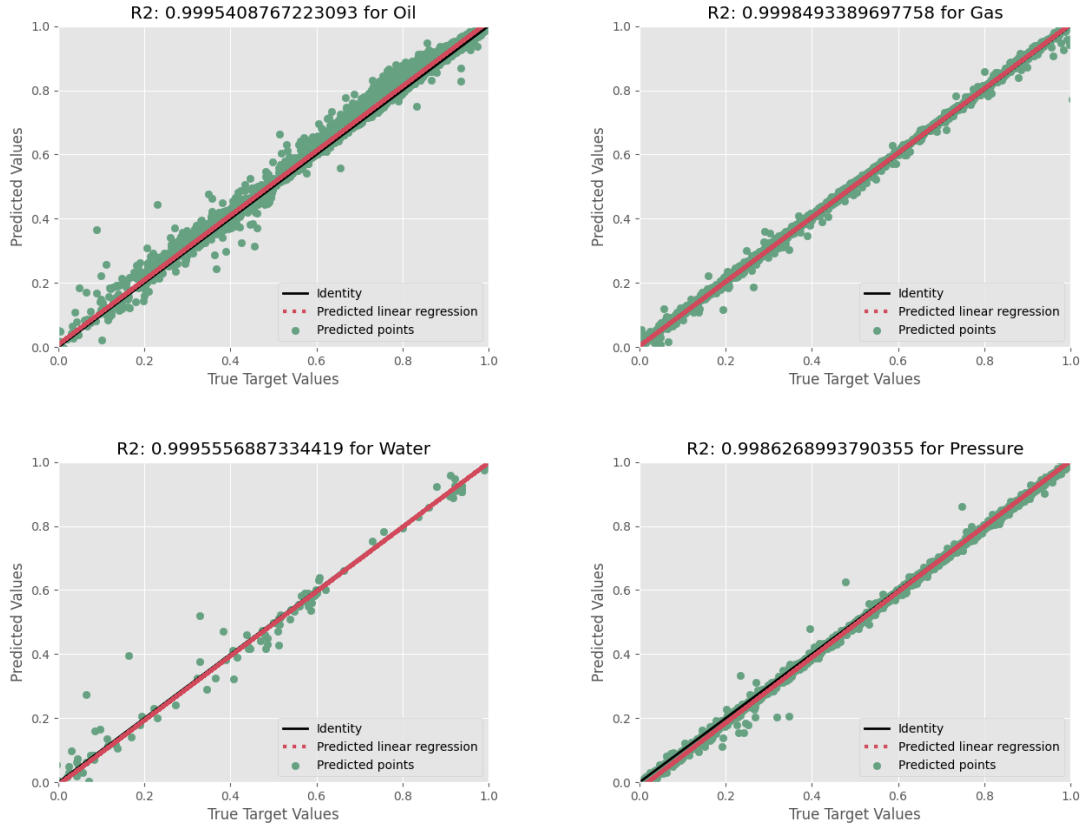
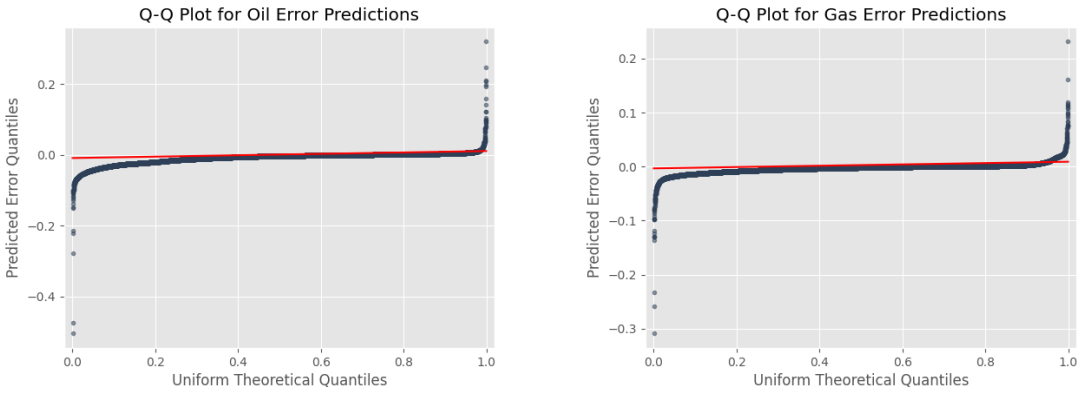


Figura 33. Plots R2 valores reais e preditos para cada umas das saídas (Saturações Óleo, Gás, Água e Pressão), 10 mil pontos.



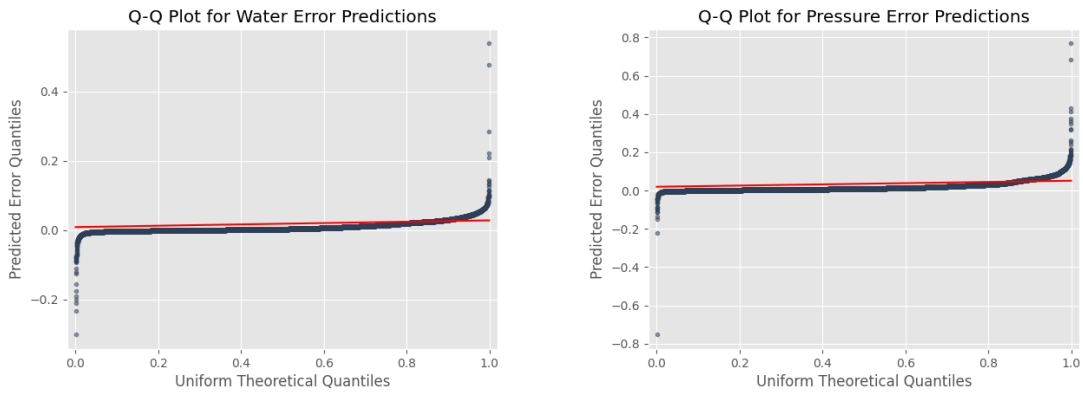


Figura 34. Gráfico Q-Q para comparação dos erros de predição com a distribuição uniforme.

4.4 EXPERIMENTOS CNN

O segundo conjunto de experimentos consistiu da construção de Redes CNN com blocos convolucionais 3D para corresponder à estrutura de input planejada no capítulo anterior correspondente. Foi utilizado a biblioteca TensorFlow com a API Keras e desenvolvido experimentos e ajustes manualmente, sem o uso de automatizadores como a anterior DNN. O que se apresenta na tabela a seguir são resultados de experimentos pontuais com as maiores métricas de R2 encontradas.

Tabela 10 - Arquiteturas e resultados para CNN.

Proposta	**R2 médio**	**R2 individual**	**arquitetura**
P1 Descritores em Cubos individuais [3x3x3]	0.971	Óleo: 0.997 Gás: 0.896 Água: 0.998 Press: 0.994	20x{ Conv2D(32)[2x2], Conv2D(64)[2x2]} + Dense(128) + Dense(64) + + Dense(4)
P2 Descritores agrupados por categoria [3x3x3x4, 3x3x3x6]	0.676	Óleo: 0.867 Gás: 0.844 Água: 0.077 Press: 0.918	4x{Conv3D(32)[2x2x2], Conv3D(64)[2x2x2]} + Dense(128) + Dense(64) + Dense(4)
P3 Descritores agrupados totais [3x3x3x20]	0.683	Óleo: 0.872 Gás: 0.839 Água: 0.095 Press: 0.924	Conv3D(32)[2x2x2] + Conv3D(64)[2x2x2] + Dense(128) + Dense(64) + Dense(4)

Todos os experimentos foram realizados com o cenário 1 e segmento desenvolvimento.

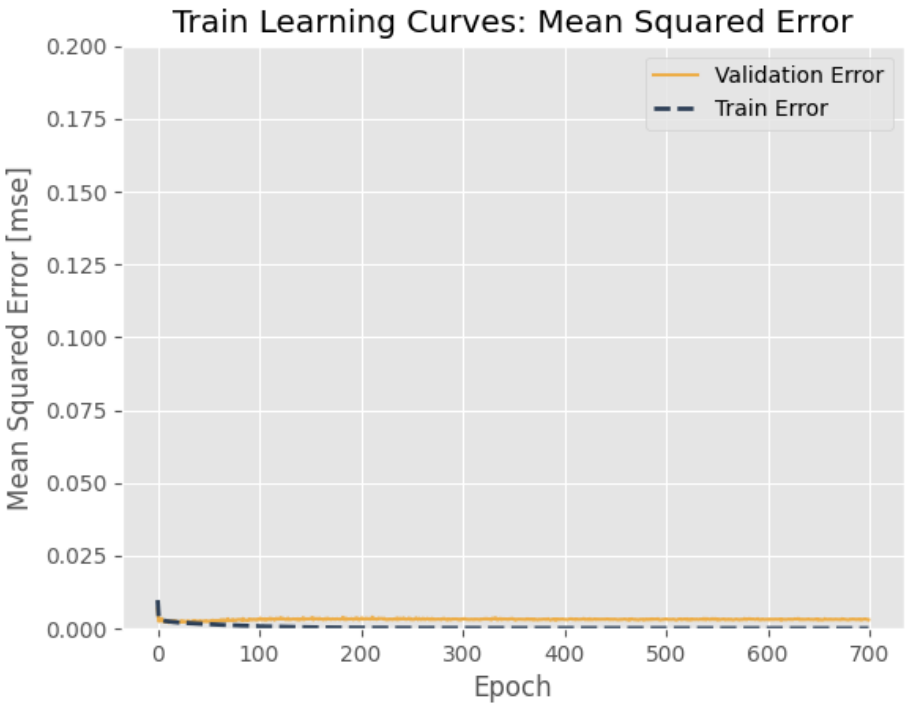


Figura 35. Curva de treinamento e validação para MSE.

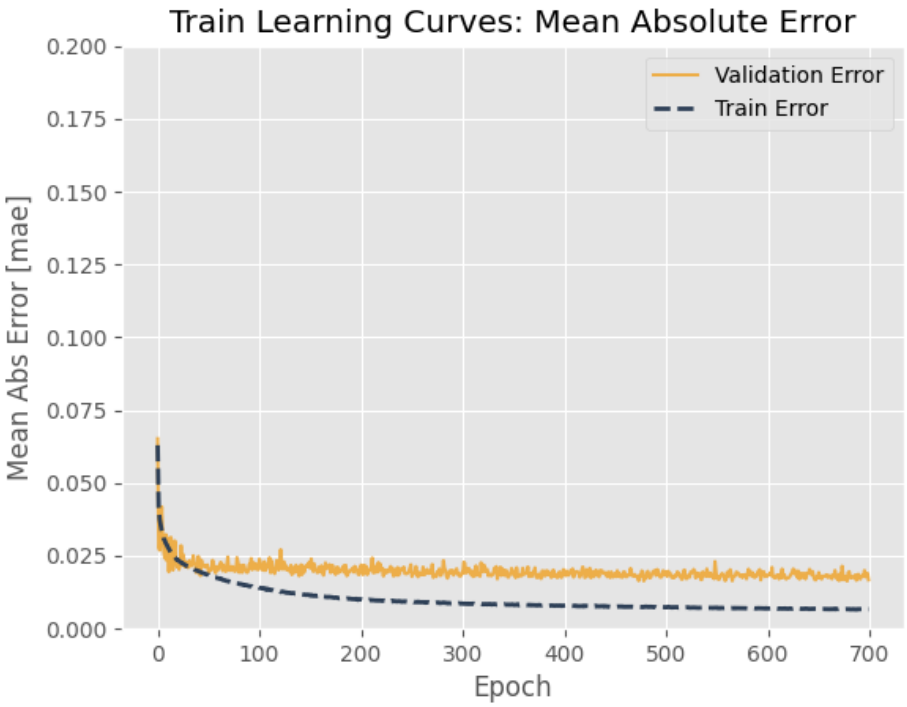


Figura 36. Curva de treinamento e validação para MAE.

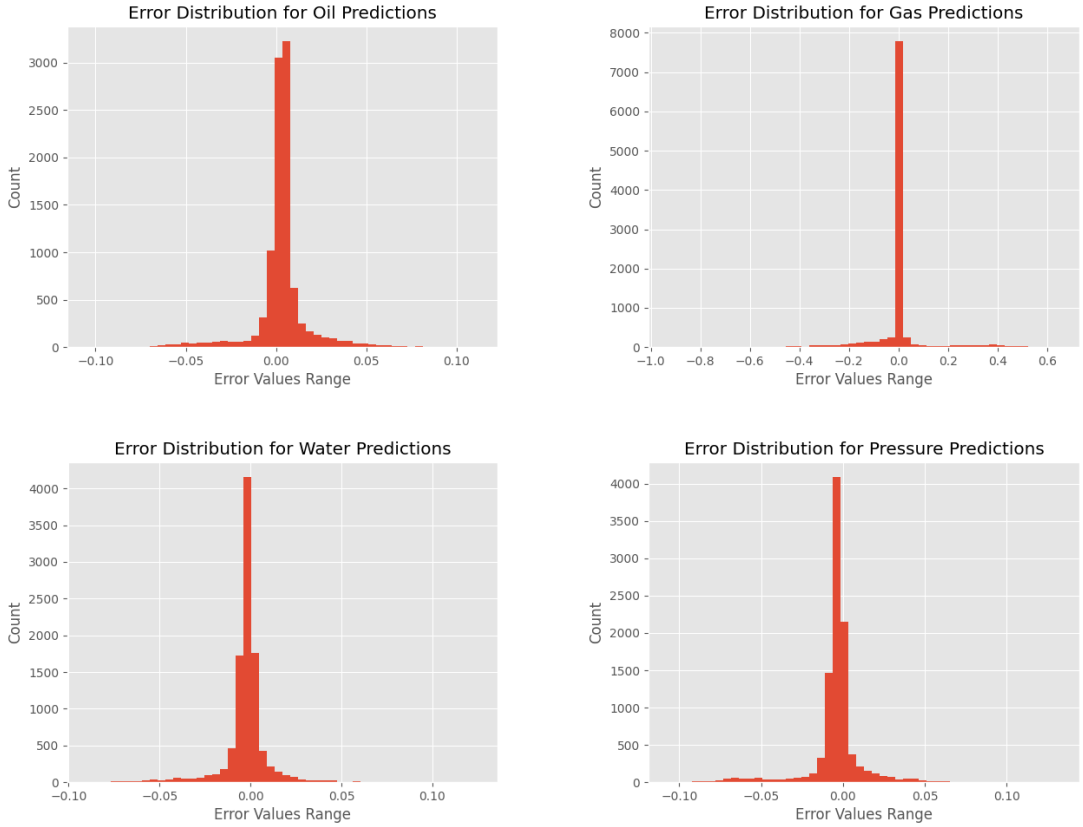
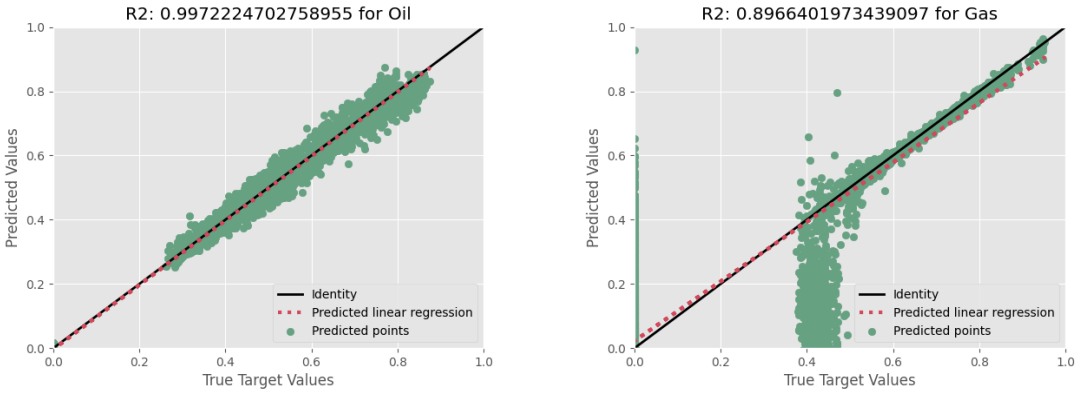


Figura 37. Histograma da distribuição dos valores de erro para cada umas das saídas (Saturações Óleo, Gás, Água e Pressão, 10 mil pontos).



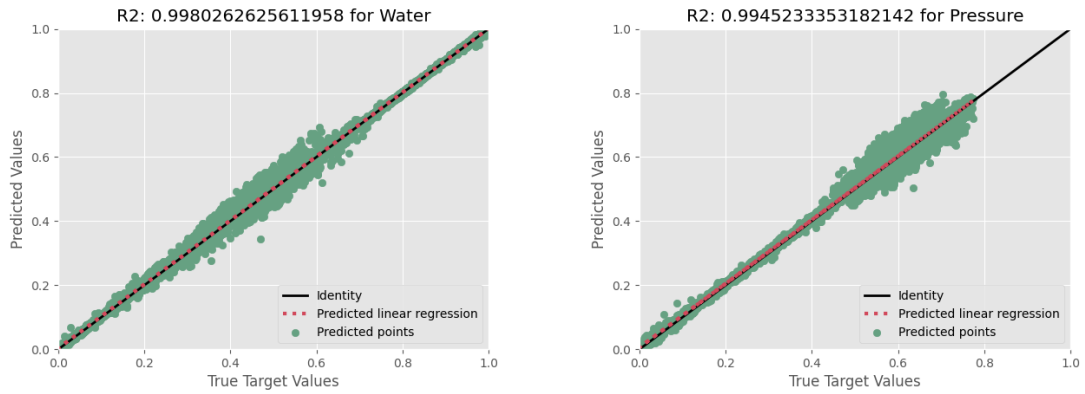


Figura 38. Plots R2 valores reais e preditos para cada umas das saídas (Saturações Óleo, Gás, Água e Pressão), 10 mil pontos.

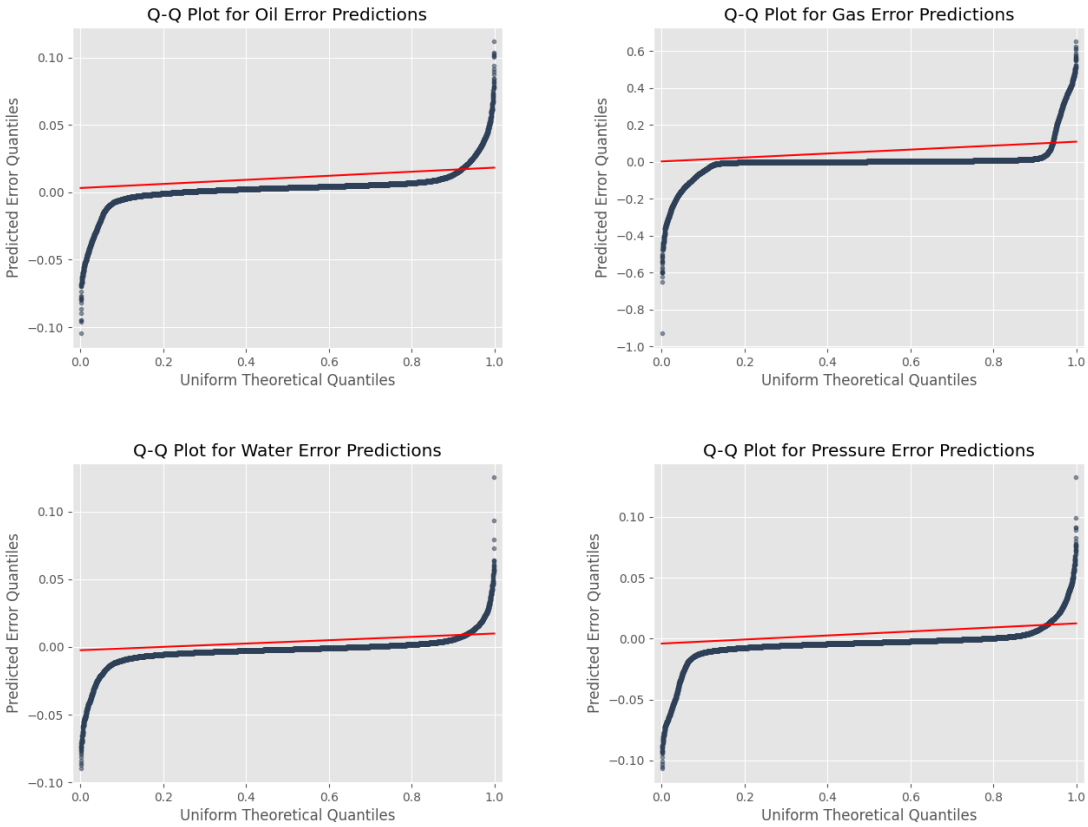


Figura 39. Gráfico Q-Q para comparação dos erros de predição com a distribuição uniforme.

4.5 APLICAÇÃO PRÁTICA – PREDIÇÃO DE CURVAS DE PRODUÇÃO

Esta aplicação apresenta a predição de um passo a frente para cada uma das variáveis de saída com o melhor modelo obtido para DNN. São apresentados as curvas de erro para as predições do cenário 1 e segmento desenvolvimento, assim como os

mapas de predição para determinado passo de tempo e camada (slice i,j) do reservatório. Em seguida a curva de produção para as variáveis de saída.

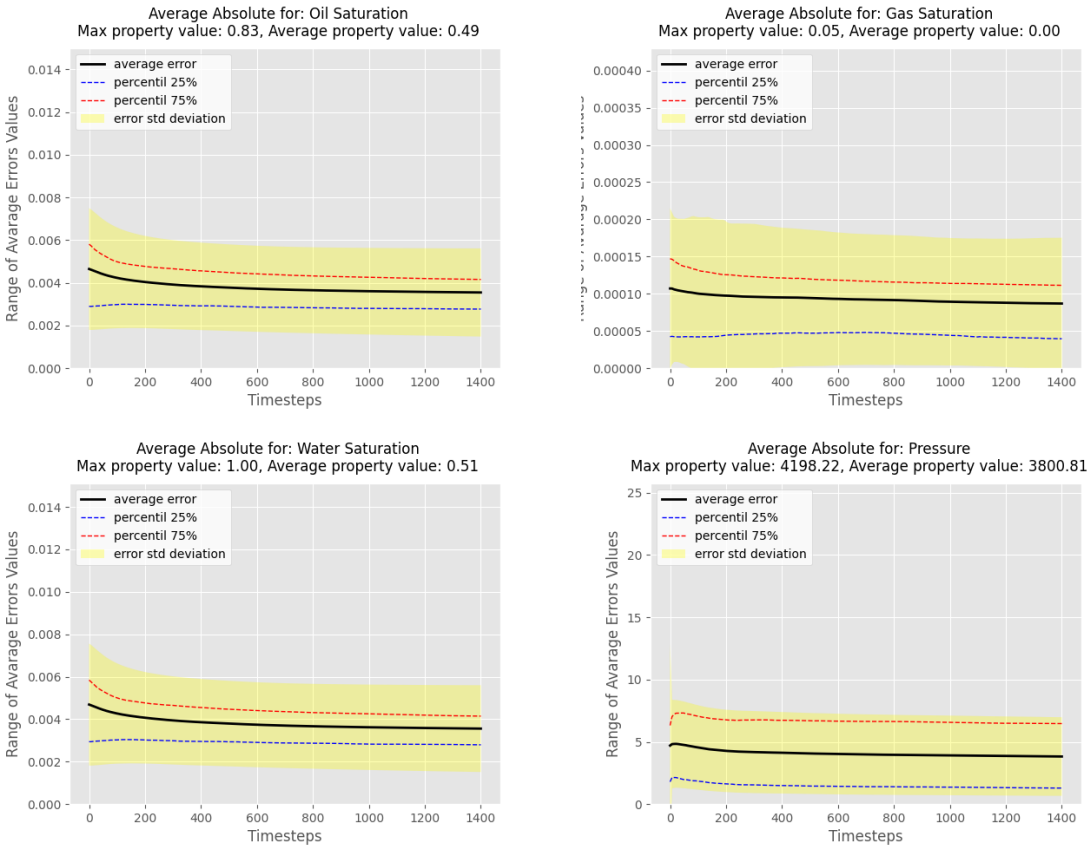


Figura 40. Curvas de Erro Médio Absoluto para predição.

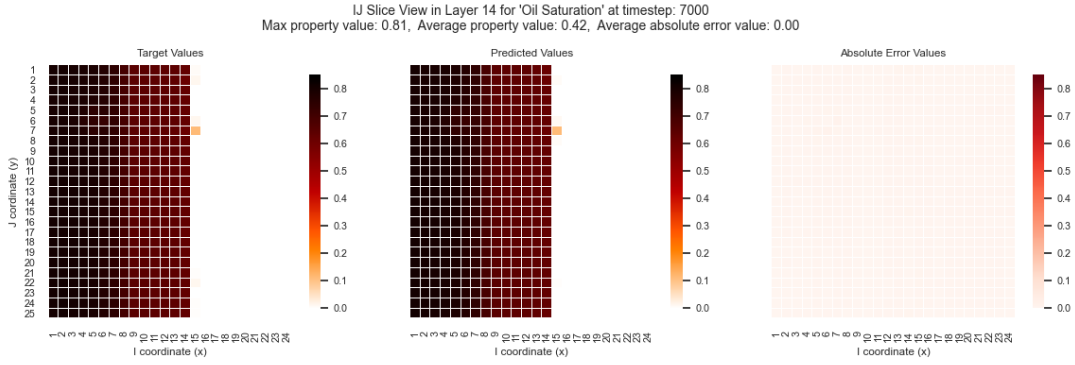


Figura 41. Mapa de valores para saturação de óleo com predição segmento desenvolvimento camada 14.

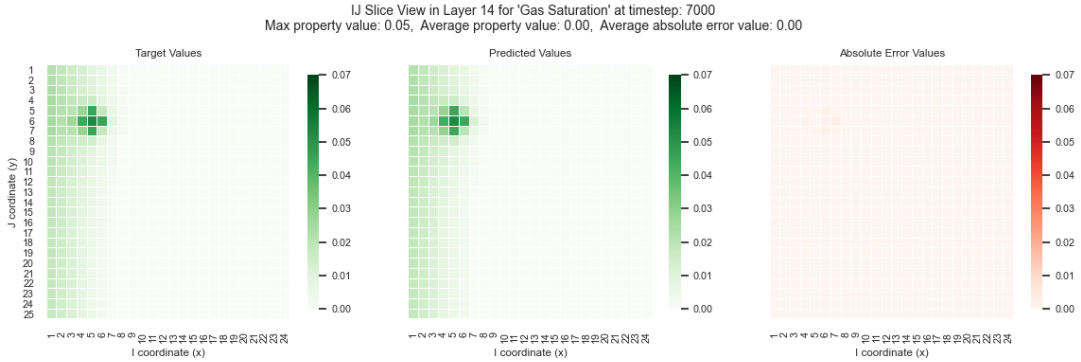


Figura 42. Mapa de valores para saturação de gás com predição segmento desenvolvimento camada 14.

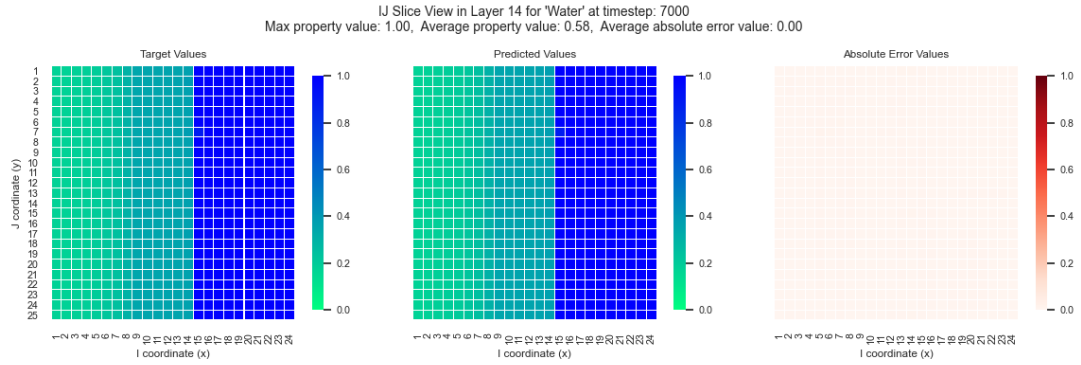


Figura 43. Mapa de valores para saturação de água com predição segmento desenvolvimento camada 14.

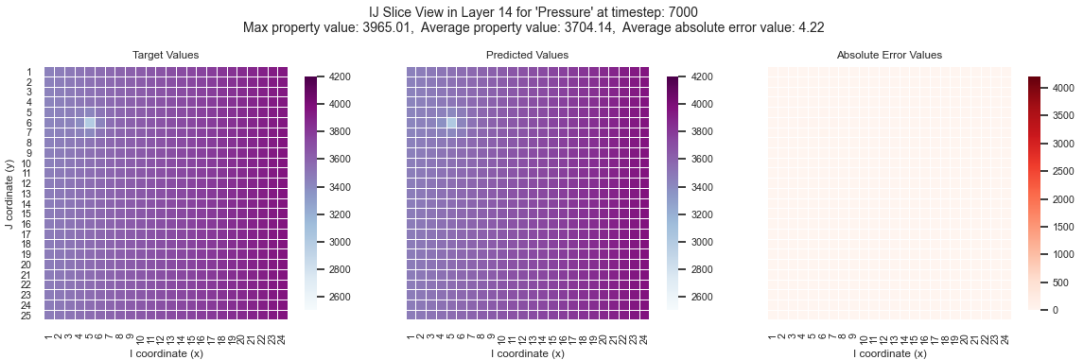


Figura 44. Mapa de valores para saturação de pressão com predição segmento desenvolvimento camada 14.

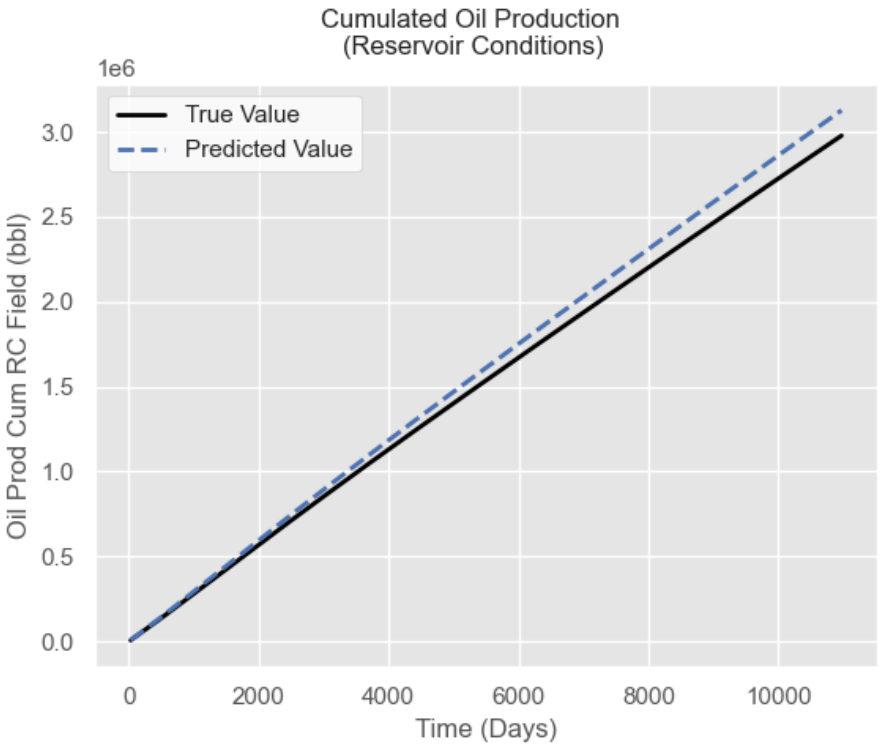


Figura 45. Predição de produção acumulado em barris Erro máximo 5%.

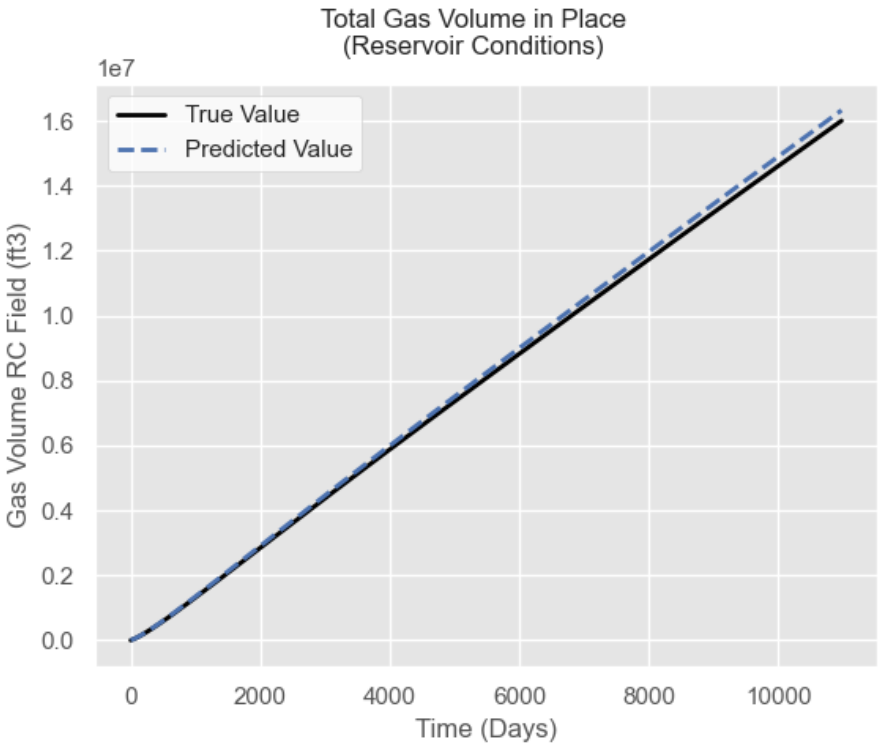


Figura 46. Predição de volume total de gás in locu ft3, erro máximo 4%.

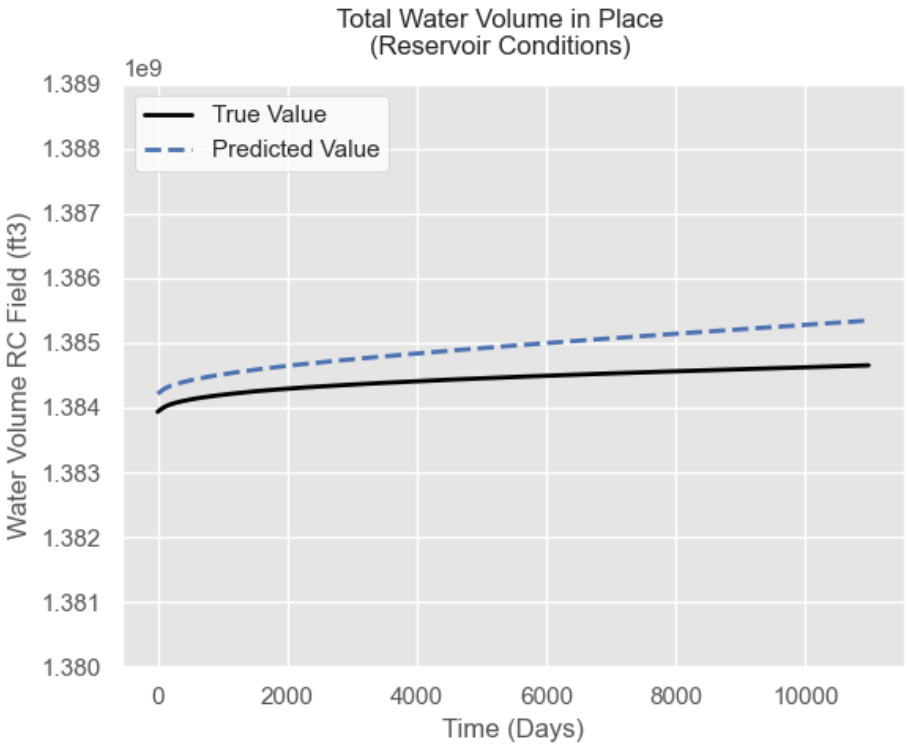


Figura 47. Predição de volume total de água in locu ft3.

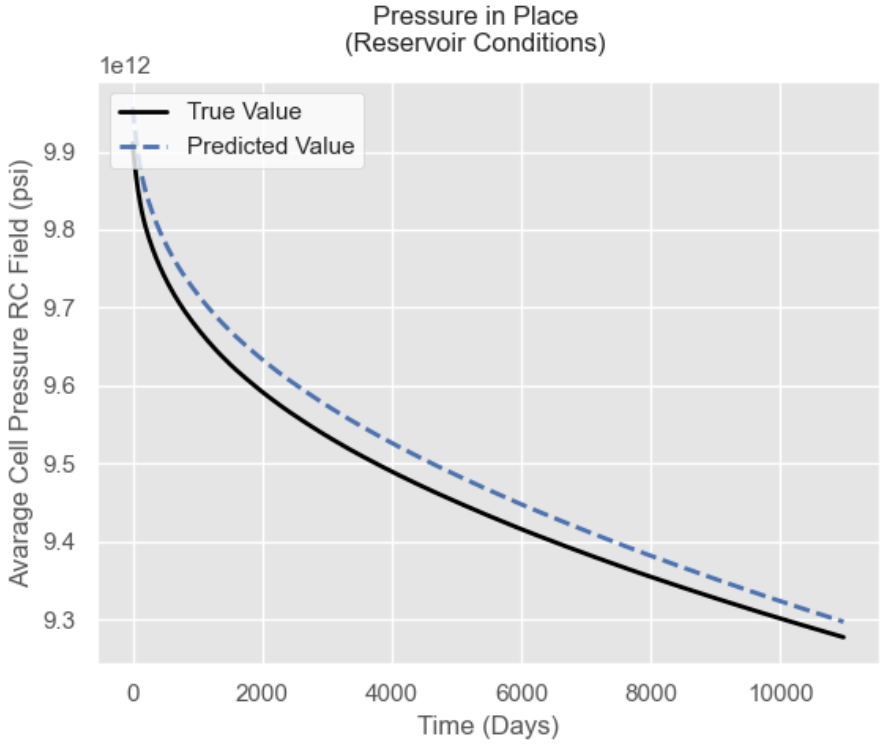


Figura 48. Predição de pressão média por célula, psi.

Este capítulo apresentou os experimentos realizados e descrição dos mesmos a partir de gráficos e tabelas. O capítulo seguinte finaliza este texto com as considerações finais e sugestões de continuidade.

5 CONSIDERAÇÕES FINAIS

Um novo procedimento para construção de proxy models foi apresentado. Este novo conjunto de etapas é capaz de entregar ao engenheiro uma ferramenta robusta e acurada, capaz de substituir com precisão mensurada um simulador numérico de reservatório para contextos específicos e bem condicionados. Em determinados cenários típicos da indústria, o modelo é capaz de prever estados com poucos passos à frente com precisão maior que 99,9% de similaridade com a solução numérica comercial.

A proposição apresentada aprofunda o avanço de modelos substitutos para além da perspectiva de modelos tipo 'caixa-preta' e também avança além dos modelos de superfície de resposta. O uso prioritário de descritores acoplados mais estreitamente com as propriedades físico-espaciais, a exemplo das distâncias para as margens ao invés dos índices de posição (i, j, k) contribui para um melhor resultado e para uma capacidade descritiva melhor.

Das contribuições deste trabalho no corpo de estudo específico destacam-se três principais: incremento no pré-processamento, avaliação de diferentes modelos, uso de Deep Learning. Este último, especialmente através das DNNs, expõe melhores resultados se comparados com o baseline.

A partir do pré-processamento e das estruturas de entrada foi possível realizar a construção e treinamento dos modelos utilizando muito menos dados do que os apresentados na literatura correlata. Sendo comum encontrar treinamento necessitando centenas ou milhares e simulações para construção de um único modelo, a proposta aqui apresentada não requer nem ao menos uma dezena para apresentar resultados potenciais.

Uma extensa revisão bibliográfica e uma síntese dos trabalhos correlacionados foram apresentados no primeiro e segundo capítulo. Este material, pode ser estendido e aprofundado para construção de um review independente e também para disponibilidade de um banco de dados abertos, incentivando pesquisadores na área.

Os experimentos compartilhados foram executados e documentados, tanto quando possível, para que sejam facilmente replicados pelo grupo de pesquisa do qual o autor faz parte e se disponibiliza igualmente para o leitor interessado.

Como em qualquer pesquisa científica, critérios de inclusão e exclusão de escopo, restrições de tempo e recursos além de capacidade técnica são fatores naturais limitadores de qualquer estudo e projeto. Conscientes disso, o autor identifica e aponta potenciais pontos de continuidade deste trabalho:

- Novos métodos para criação de descritores e outros métodos de pré-processamento para escolha e combinação de descritores podem ser empregados para aumentar o escopo de base comparativa e também de avaliação de custo efetivo de processamento e memória em aplicações diretas na indústria ou incorporação em software comercial existente.
- Sobre possibilidades de novos modelos observa-se espaço para uso de técnicas de AutoML na criação e treinamento de novas arquitetura CNN, o uso de métodos híbridos agrupando dados escalares com DNN e propriedades físicas com CNN, e ainda o uso de ensembles.
- A cerca do escopo de dados de treinamento e validação, uma investigação sobre a vizinhança no método DNN, para maior que 6, e um cubo de dados maior que C27 para o caso da CNN, podem trazer significativos incrementos de precisão, especialmente com predição de muitos passos à frente.
- A aplicação do método em novos reservatórios e com resoluções diferentes pode ser preponderante para avaliar a robustez e adaptabilidade do processo aqui definido.

REFERÊNCIAS BIBLIOGRÁFICAS

ALENEZI, F.; MOHAGHEGH, S. A data-driven smart proxy model for a comprehensive reservoir simulation. **2016 4th Saudi International Conference on Information Technology (Big Data Analysis), KACSTIT 2016**, [*s. l.*], 2016a. Available at: https://doi.org/10.1109/KACSTIT.2016.7756063

ALENEZI, F.; MOHAGHEGH, S. A data-driven smart proxy model for a comprehensive reservoir simulation. *In*: , 2016b. **2016 4th Saudi International Conference on Information Technology (Big Data Analysis), KACSTIT 2016**. [*S. l.: s. n.*], 2016. Available at: https://doi.org/10.1109/KACSTIT.2016.7756063

ALENEZI, F.; MOHAGHEGH, S. Developing a Smart Proxy for the SACROC Water-Flooding Numerical Reservoir Simulation Model. *In*: , 2017. **SPE Western Regional Meeting**. [*S. l.*]: Society of Petroleum Engineers, 2017. Available at: https://doi.org/10.2118/185691-MS

AMINI, S. *et al.* Pattern recognition and data-driven analytics for fast and accurate replication of complex numerical reservoir models at the grid block level. **Society of Petroleum Engineers - SPE Intelligent Energy International 2014**, [*s. l.*], p. 781–787, 2014a. Available at: https://doi.org/10.2118/167897-ms

AMINI, S. *et al.* Pattern Recognition and Data-Driven Analytics for Fast and Accurate Replication of Complex Numerical Reservoir Models at the Grid Block Level. *In*: , 2014b. **SPE Intelligent Energy Conference & Exhibition**. [*S. l.*]: Society of Petroleum Engineers, 2014. p. 7. Available at: https://doi.org/10.2118/167897-MS

AMINI, Shohreh; MOHAGHEGH, S. Application of machine learning and artificial intelligence in proxy modeling for fluid flow in porous media. **Fluids**, [*s. l.*], v. 4, n. 3, p. 1–17, 2019. Available at: https://doi.org/10.3390/fluids4030126

AMIRIAN, E. *et al.* Artificial Neural Network Modeling and Forecasting of Oil Reservoir Performance. *In*: APPLICATIONS OFDATA MANAGEMENT AND ANALYSIS. [*S. l.*]: Springer, 2018. p. 43–67. Available at: https://doi.org/10.1007/978-3-319-95810-1_5

AMIRIAN, E.; JOHN CHEN, Z. Cognitive Data-Driven Proxy Modeling for Performance Forecasting of Waterflooding Process. **Global Journal of Technology and Optimization**, [*s. l.*], v. 08, n. 01, p. 1–8, 2017. Available at: https://doi.org/10.4172/2229-8711.1000207

ANIFOWOSE, Fatai A. Ensemble Machine Learning: The Latest Development in Computational Intelligence for Petroleum Reservoir Characterization. *In*: , 2013, Al-Khobar, Saudi Arabia. **SPE Saudi Arabia Section Technical Symposium and Exhibition**. Al-Khobar, Saudi Arabia: Society of Petroleum Engineers, 2013. p. 10. Available at: https://doi.org/10.2118/168111-MS

ANIFOWOSE, Fatai Adesina. Artificial Intelligence Application in Reservoir Characterization and Modeling: Whitening the Black Box. **SPE Saudi Arabia section Young Professionals Technical Symposium**, [*s. l.*], 2011. Available at: https://doi.org/10.2118/155413-ms

BABAEI, M.; PAN, I. Performance comparison of several response surface surrogate models and ensemble methods for water injection optimization under uncertainty. **Computers and Geosciences**, [*s. l.*], v. 91, p. 19–32, 2016. Available at: https://doi.org/10.1016/j.cageo.2016.02.022

BALAJI, K. *et al.* Status of Data-Driven Methods and their Applications in Oil and Gas Industry. *In*: , 2018. **Spe**. [*S. l.*]: Society of Petroleum Engineers, 2018. p. 18. Available at: https://doi.org/10.2118/190812-MS

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, [*s. l.*], 2012.

BHOSEKAR, A.; IERAPETRITOU, M. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. **Computers and Chemical Engineering**, [*s. l.*], v. 108, p. 250–267, 2018. Available at: https://doi.org/10.1016/j.compchemeng.2017.09.017

BRAVO, C. E. *et al.* State of the Art of Artificial Intelligence and Predictive Analytics in the E&P Industry: A Technology Survey. **SPE Journal**, [*s. l.*], v. 19, n. 04, p. 547–563, 2014. Available at: https://doi.org/10.2118/150314-PA

CARVAJAL, G.; MAUCEC, M.; CULLICK, S. Components of Artificial Intelligence and Data Analytics. *In*: INTELLIGENT DIGITAL OIL AND GAS FIELDS. [*S. l.: s. n.*], 2018a. p. 101–148. Available at: https://doi.org/10.1016/B978-0-12-804642-5.00004-9

CARVAJAL, G.; MAUCEC, M.; CULLICK, S. Introduction to Digital Oil and Gas Field Systems. *In*: INTELLIGENT DIGITAL OIL AND GAS FIELDS. [*S. l.*]: Elsevier, 2018b. p. 1–41. Available at: https://doi.org/10.1016/B978-0-12-804642-5.00001-3

CRANGANU, C.; BREABAN, M. E.; LUCHIAN, H. **Artificial intelligent approaches in petroleum geosciences**. [*S. l.: s. n.*], 2015. Available at: https://doi.org/10.1007/978-3-319-16531-8

EASON, J.; CREMASCHI, S. Adaptive sequential sampling for surrogate model generation with artificial neural networks. **Computers and Chemical Engineering**, [*s. l.*], v. 68, p. 220–232, 2014. Available at: https://doi.org/10.1016/j.compchemeng.2014.05.021

EMERICK, A. A. *et al.* **Intelligent Systems in Oil Field Development under Uncertainty**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. ISSN 1860-949X.(Studies in Computational Intelligence).v. 183 Available at: https://doi.org/10.1007/978-3-540-93000-6

ESMAILI, S.; MOHAGHEGH, S. D. Full field reservoir modeling of shale assets using advanced data-driven analytics. **Geoscience Frontiers**, [*s. l.*], v. 7, n. 1, p. 11–20, 2016. Available at: https://doi.org/10.1016/j.gsf.2014.12.006

FAWCETT, T. An introduction to ROC analysis Tom. **Irbm**, [*s. l.*], v. 35, n. 6, p. 299–309, 2005. Available at: https://doi.org/10.1016/j.patrec.2005.10.010

FORRESTER, A. I. J.; KEANE, A. J. Recent advances in surrogate-based optimization. **Progress in Aerospace Sciences**, [*s. l.*], v. 45, n. 1–3, p. 50–79, 2009. Available at: https://doi.org/10.1016/j.paerosci.2008.11.001

FOSS, B.; KNUDSEN, B. R.; GRIMSTAD, B. Petroleum production optimization – A static or dynamic problem? **Computers and Chemical Engineering**, [*s. l.*], v. 114, p. 245–253, 2018. Available at: https://doi.org/10.1016/j.compchemeng.2017.10.009

GOLZARI, A.; HAGHIGHAT SEFAT, M.; JAMSHIDI, S. Development of an adaptive surrogate model for production optimization. **Journal of Petroleum Science and Engineering**, [*s. l.*], 2015a. Available at: https://doi.org/10.1016/j.petrol.2015.07.012

GOLZARI, A.; HAGHIGHAT SEFAT, M.; JAMSHIDI, S. Development of an adaptive surrogate model for production optimization. **Journal of Petroleum Science and Engineering**, [*s. l.*], v. 133, p. 677–688, 2015b. Available at: https://doi.org/10.1016/j.petrol.2015.07.012

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. 1. ed. [*S. l.*]: MIT Press, 2016. v. 1*E-book*.

GORISSEN, D. *et al.* A surrogate modeling and adaptive sampling toolbox for computer based design. **The Journal of Machine Learning Research**, [*s. l.*], v. 11, p. 2051–2055, 2010. Available at: https://doi.org/10.1007/978-1-60761-925-3_30

HAGHSHENAS, Y. *et al.* Developing grid-based smart proxy model to evaluate various water flooding injection scenarios. **Petroleum Science and Technology**, [*s. l.*], v. 38, n. 17, p. 870–881, 2020. Available at: https://doi.org/10.1080/10916466.2020.1796703

HAYKIN, S. **Neural networks - a comprehensive foundation**. [*S. l.: s. n.*], 1999. ISSN 02698889. Available at: https://doi.org/10.1017/S0269888998214044

HEY, TONY; TANSLEY, STEWART; TOLLE, K. **The Fourth Paradigm: Data-Intensive Scientific Discovery**. [*S. l.: s. n.*], 2009. *E-book*.

HOLDAWAY, K. **Harness Oil and Gas big data woth analytics**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2014. v. 136 Available at: https://doi.org/10.1002/9781118910948.fmatter

HUANG, L.; DONG, X.; CLEE, T. E. A scalable deep learning platform for identifying geologic features from seismic attributes. **The Leading Edge**, [*s. l.*], v. 36, n. 3, p. 249–256, 2017. Available at: https://doi.org/10.1190/tle36030249.1

JABER, A. K.; AL-JAWAD, S. N.; ALHURAISHAWY, A. K. A review of proxy modeling applications in numerical reservoir simulation. **Arabian Journal of**

Geosciences, [*s. l.*], v. 12, n. 22, 2019. Available at: https://doi.org/10.1007/s12517-019-4891-1

JIN, H.; SONG, Q.; HU, X. Auto-Keras: **Efficient Neural Architecture Search with Network Morphism**. [*s. l.*], 2018. Available at: http://arxiv.org/abs/1806.10282

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**, [*s. l.*], v. 349, n. 6245, p. 255–260, 2015. Available at: https://doi.org/10.1126/science.aaa8415

KRASNOV, F.; GLAVNOV, N.; SITNIKOV, A. A machine learning approach to enhanced oil recovery prediction. *In*: , 2018a. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [*S. l.: s. n.*], 2018. Available at: https://doi.org/10.1007/978-3-319-73013-4_15

KRASNOV, F.; GLAVNOV, N.; SITNIKOV, A. A Machine Learning Approach to Enhanced Oil Recovery Prediction. *In*: VAN DER AALST, W. M. P. *et al.* (org.). Cham: Springer International Publishing, 2018b. (Lecture Notes in Computer Science).v. 10716, p. 164–171. Available at: https://doi.org/10.1007/978-3-319-73013-4_15

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, [*s. l.*], v. 521, n. 7553, p. 436–444, 2015. Available at: https://doi.org/10.1038/nature14539

LIASHCHYNSKYI, P.; LIASHCHYNSKYI, P. **Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS**. [*S. l.: s. n.*], 2019.

MAGALHÃES, R. M. **Uma Investigação sobre a Utilização de Processamento Paralelo no Projeto de Máquinas de Comitê**. 98 f. 2007. - UFRN, [*s. l.*], 2007.

MNIH, V. *et al.* Playing Atari with Deep Reinforcement Learning. [*s. l.*], p. 1–9, 2013. Available at: https://doi.org/10.1038/nature14236

MOHAGHEGH, S D *et al.* Smart Proxy: An Innovative Reservoir Management Tool; Case Study of a Giant Mature Oilfield in the UAE. *In*: , 2015. **Abu Dhabi International Petroleum Exhibition and Conference**. [*S. l.*]: Society of Petroleum Engineers, 2015. Available at: https://doi.org/10.2118/177829-MS

MOHAGHEGH, S D; KHAZAENI, Y. Application of Artificial Intelligence in the upstream oil and gas industry. **Artificial Intelligence: Approaches, Tools, and Applications**, [*s. l.*], v. 18, n. 3, p. 1–38, 2011. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85016734636&partnerID=40&md5=30767107083ca8cd6cb08971b54c3413

MOHAGHEGH, Shahab D. *et al.* Grid-Based Surrogate Reservoir Modeling (SRM) for fast track analysis of numerical reservoir simulation models at the grid block level. *In*: , 2012a. **Society of Petroleum Engineers Western Regional Meeting 2012**. [*S. l.: s. n.*], 2012. Available at: https://doi.org/10.2118/153844-ms

MOHAGHEGH, Shahab D. *et al.* Grid-Based Surrogate Reservoir Modeling (SRM) for Fast Track Analysis of Numerical Reservoir Simulation Models at the Gridblock Level. **SPE Western Regional Meeting**, [*s. l.*], p. 1–13, 2012b. Available at: https://doi.org/10.2118/153844-MS

MOHAGHEGH, Shahab D. Recent Developments in Application of Artificial Intelligence in Petroleum Engineering. **Journal of Petroleum Technology**, [*s. l.*], v. 57, n. 04, p. 86–91, 2005. Available at: https://doi.org/10.2118/89033-JPT

MOHAGHEGH, Shahab Dean. Reservoir simulation and modeling based on artificial intelligence and data mining (AI&DM). **Journal of Natural Gas Science and Engineering**, [*s. l.*], v. 3, n. 6, p. 697–705, 2011. Available at: https://doi.org/10.1016/j.jngse.2011.08.003

MÜLLER, J.; PICHÉ, R. Mixture surrogate models based on Dempster-Shafer theory for global optimization problems. **Journal of Global Optimization**, [*s. l.*], v. 51, n. 1, p. 79–104, 2011. Available at: https://doi.org/10.1007/s10898-010-9620-y

NAVRÁTIL, J. *et al.* Accelerating Physics-Based Simulations Using End-to-End Neural Network Proxies: An Application in Oil Reservoir Modeling. **Frontiers in Big Data**, [*s. l.*], v. 2, n. September, p. 1–13, 2019. Available at: https://doi.org/10.3389/fdata.2019.00033

NISBET, R.; MINER, G.; YALE, K. Chapter 19 - Deep learning. *In*: HANDBOOK OF STATISTICAL ANALYSIS AND DATA MINING APPLICATIONS. [*S. l.*]: Elsevier, 2018. p.

741–751. Available at: https://doi.org/https://doi.org/10.1016/B978-0-12-416632-5.00019-0

NWACHUKWU, A. *et al.* Fast evaluation of well placements in heterogeneous reservoir models using machine learning. **Journal of Petroleum Science and Engineering**, [*s. l.*], v. 163, n. January, p. 463–475, 2018. Available at: https://doi.org/10.1016/j.petrol.2018.01.019

POPA, A. S.; CASSIDY, S. D. Artificial Intelligence for Heavy Oil Assets: The Evolution of Solutions and Organization Capability. *In*: , 2012. **SPE Annual Technical Conference and Exhibition**. [*S. l.*]: Society of Petroleum Engineers, 2012. p. 1–11. Available at: https://doi.org/10.2118/159504-MS

POULADI, B. *et al.* A robust proxy for production well placement optimization problems. **Fuel**, [*s. l.*], v. 206, p. 467–481, 2017. Available at: https://doi.org/10.1016/j.fuel.2017.06.030

POWERS, D. M. W. **Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation**. Adelaide, Australia: [*s. n.*], 2007. Available at: http://hdl.handle.net/2328/27165.

ROSA, A. J.; CARVALHO, R. de S.; XAVIER, J. A. D. **Engenharia de Reservatórios de Petróleo**. Rio de Janeiro: [*s. n.*], 2013.

SAPUTELLI, L. Technology Focus: Petroleum Data Analytics. **Journal of Petroleum Technology**, [*s. l.*], v. 68, n. 10, p. 66–66, 2016. Available at: https://doi.org/10.2118/1016-0066-JPT

SCHMIDHUBER, J. Deep Learning in neural networks: An overview. **Neural Networks**, [*s. l.*], v. 61, p. 85–117, 2015. Available at: https://doi.org/10.1016/j.neunet.2014.09.003

SERCU, T. *et al.* Very deep multilingual convolutional neural networks for LVCSR. *In*: , 2016. **2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [*S. l.*]: IEEE, 2016. p. 4955–4959. Available at: https://doi.org/10.1109/ICASSP.2016.7472620

SHIRANGI, M. G. Applying Machine Learning Algorithms to Oil Reservoir Production Optimization. **Stanford University**, [*s. l.*], p. 1–5, 2012.

SUDAKOV, O. *et al.* Artificial Neural Network Surrogate Modeling of Oil Reservoir: A Case Study. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [*s. l.*], v. 11555 LNCS, p. 232–241, 2019. Available at: https://doi.org/10.1007/978-3-030-22808-8_24

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning, Second Edition**. 2. ed. [*S. l.: s. n.*], 2018. Available at: https://doi.org/10.1126/science.aaa8415

TOAL, D. J. J. A study into the potential of GPUs for the efficient construction and evaluation of Kriging models. **Engineering with Computers**, [*s. l.*], v. 32, n. 3, p. 377–404, 2016. Available at: https://doi.org/10.1007/s00366-015-0421-2

VIANA, F. A. C.; HAFTKA, R. T.; WATSON, L. T. Efficient global optimization algorithm assisted by multiple surrogate techniques. **Journal of Global Optimization**, [*s. l.*], v. 56, n. 2, p. 669–689, 2013. Available at: https://doi.org/10.1007/s10898-012-9892-5

WALDELAND, A. U. *et al.* Convolutional neural networks for automated seismic interpretation. **The Leading Edge**, [*s. l.*], v. 37, n. 7, p. 529–537, 2018. Available at: https://doi.org/10.1190/tle37070529.1

WANG, C. *et al.* An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. **Environmental Modelling and Software**, [*s. l.*], v. 60, p. 167–179, 2014. Available at: https://doi.org/10.1016/j.envsoft.2014.05.026

ZHANG, J.; CHOWDHURY, S.; MESSAC, A. An adaptive hybrid surrogate model. **Structural and Multidisciplinary Optimization**, [*s. l.*], v. 46, n. 2, p. 223–238, 2012. Available at: https://doi.org/10.1007/s00158-012-0764-x

ZHANG, K. *et al.* A new method for the construction and optimization of quadrangular adaptive well pattern. **Computational Geosciences**, [*s. l.*], v. 21, n. 3, p. 499–518, 2017. Available at: https://doi.org/10.1007/s10596-017-9626-3. Acesso em:

17 set. 2018.

ZHANG, Y. *et al.* Ensemble of surrogates based on error classification by unsupervised learning. **2016 IEEE Congress on Evolutionary Computation, CEC 2016**, [*s. l.*], p. 4344–4349, 2016. Available at: https://doi.org/10.1109/CEC.2016.7744342

ZUBAREV, D. Pros and Cons of Applying a Proxy Model as a Substitute for Full Reservoir Simulations. **Journal of Petroleum Technology**, New Orleans, Louisiana, v. 62, n. 07, p. 23, 2010. Available at: https://doi.org/10.2118/0710-0041-JPT

APÊNDICES

A. ALGORITMO DA RETROPROPAGAÇÃO (BACKPROPAGATION)

Algoritmo de treinamento para ANN de múltiplas camadas conforme relatado pelo autor em (MAGALHÃES, 2007). A rede Perceptron de Múltiplas Camadas (PMC / MLP) genérica considerada aqui é composta de L camadas, sendo uma camada de entrada e uma camada de saída. Cada camada possui $q^{(l)}$ neurônios (indexados pela variável $i, i = 1, \ldots, q^{(l)}$). As funções de ativação de cada neurônio podem ser de qualquer tipo.

Algoritmo de Treinamento para redes ANN tipo MLP

$$v_i^{(l)}(n) = \sum_{j=0}^{m} w_{ij}^{(l)}(n) y_j^{(l-1)}(n)$$

$$y_i^{(l)}(n) = \phi_i^{(l)}\left(v_i^{(l)}(n)\right)$$

$$e_i(n) = d_i(n) - y_i^{(L)}(n)$$

gradiente na camada de saída

$$\delta_i^{(L)}(n) = e_i^{(L)}(n)\phi_i'^{(L)}\left(v_i^{(L)}(n)\right)$$

gradiente nas camadas ocultas

$$\delta_i^{(i)}(n) = \phi_i'^{(l)}\left(v_i^{(l)}(n)\right) \sum_{k=1}^{q^{(l+1)}} \delta_k^{(l+1)}(n) w_{ik}^{(l+1)}(n)$$

$$\boldsymbol{w}_{ij}^{(l)}(n+1) = w_{ij}^{(l)}(n) + \eta\delta_i^{(l)}(n) y_j^{(l-1)}(n)$$

$v_i^{(l)}(n)$ – potencial de ativação do i-ésimo neurônio da l-ésima camada, quando da aplicação do n-ésimo exemplo de treinamento;

$w_{ij}^{(l)}(n)$ – peso sináptico associado à j-ésima entrada do i-ésimo neurônio da l-ésima camada, quando da aplicação do n-ésimo exemplo de treinamento;

$y_j^{(l-1)}(n)$– saída do j-ésimo neurônio da (l-1)-ésima camada, quando da aplicação do n-ésimo exemplo de treinamento;

$e_i^n(n)$ – erro entre a saída do i-ésimo neurônio da camada de saída e a saída desejada da rede;

$d_i(n)$ – saída desejada do i-ésimo neurônio da rede, associada ao n-ésimo exemplo de treinamento;

$\phi_i^{(l)}$ – função de ativação do i-ésimo neurônio da l-ésima camada;

$\phi_i'^{(l)}$– derivada da função de ativação do i-ésimo neurônio da l-ésima camada;

$\delta_i^{(i)}(n)$– gradiente do erro associado ao i-ésimo neurônio da l-ésima camada, quando da aplicação do n-ésimo exemplo de treinamento.

B. COORDENADAS E HISTOGRAMAS PARA POÇOS DOS CENÁRIOS EXPERIMENTAIS

Conforme descrito no texto, as imagens seguintes apresentam os valores correspondentes às coordenadas i, j, k para cada um dos diferentes cenários (1 a 4). Os valores eventualmente são designados de x, y e z em correspondência aos eixos X, Y e Z no grid de células do reservatório.

Na Figura 49 constam os valores de coordenadas respectivamente para Cenário 1 a 4 obtidas a partir da técnica de Amostragem por Hipercubos Latinos (LHS). Figura 50 a Figura 53, apresentam os histogramas para cada um dos vinte (20) realizações de cada cenário.

	x	y	z
0	9	10	5
1	9	23	12
2	1	16	2
3	2	12	3
4	2	4	5
5	5	6	14
6	4	23	9
7	10	17	10
8	11	19	10
9	3	9	14
10	7	3	7
11	8	18	6
12	13	13	7
13	12	3	6
14	7	12	12
15	6	7	11
16	12	24	8
17	3	15	13
18	5	2	9
19	10	21	4

(a) 1 Prod.

	xa	ya	za	xb	yb	zb
0	3	6	7	12	17	6
3	2	4	6	7	9	5
4	12	1	2	4	12	9
15	11	24	9	1	7	8
17	11	9	3	3	4	7
20	4	3	4	12	13	9
22	2	19	14	12	25	11
26	12	1	8	1	16	6
27	7	9	4	12	20	3
29	11	21	2	4	10	3
35	10	7	11	4	17	12
37	1	8	12	13	2	3
38	7	20	8	13	12	4
39	13	16	3	2	6	3
40	10	1	9	1	8	13
41	9	6	6	4	21	14
42	3	4	11	8	25	10
51	3	21	11	11	13	5
52	8	16	14	1	10	13
53	3	13	3	12	8	5

(b) 2 Prod. [a, b]

	xa	ya	za	xb	yb	zb	xc	yc	zc	xd	yd	zd
10	10	6	13	6	11	5	13	17	8	4	2	4
37	13	2	12	6	24	8	11	5	6	1	15	14
70	11	15	14	4	10	3	13	5	5	1	21	4
93	4	10	10	9	4	11	1	21	4	7	14	7
100	3	22	2	9	16	4	7	10	10	11	4	13
113	12	5	9	2	2	6	9	19	9	4	10	3
121	1	9	2	7	18	5	13	12	6	9	22	11
181	7	19	8	10	7	14	1	22	8	12	15	8
182	4	7	14	2	23	6	11	13	13	7	3	5
213	3	9	6	1	15	3	7	20	6	13	2	8
217	5	13	13	7	10	8	10	5	6	2	18	14
246	4	18	5	13	14	2	8	22	5	10	8	10
293	8	17	13	12	13	8	2	23	3	6	1	7
337	12	23	8	2	20	3	5	8	7	8	17	11
382	1	17	12	13	2	6	7	11	13	5	25	2
400	2	22	10	4	1	7	12	19	14	8	14	4
410	7	10	8	1	24	14	12	1	7	10	18	14
418	9	12	10	13	16	10	5	20	9	3	3	11
426	5	8	3	2	18	14	13	15	8	10	25	3
437	11	18	6	3	5	13	5	25	3	9	12	8

(c) 4 Produtores [a, b, c, d]

	xp	yp	zp	xi	yi	zi
0	10	14	10	21	20	11
1	1	9	7	22	24	8
2	6	12	8	20	18	10
3	11	17	13	22	16	3
4	4	3	8	19	11	12
5	8	9	2	19	9	9
6	5	18	13	21	10	2
7	9	8	4	20	8	4
8	7	25	2	23	5	5
9	2	7	5	24	25	9
10	12	11	6	22	1	13
11	3	20	6	22	22	3
12	7	23	11	24	19	6
13	13	2	9	24	14	14
14	11	16	9	21	15	11
15	4	5	14	21	6	7
16	13	21	13	23	3	7
17	10	14	3	19	2	9
18	1	22	12	20	12	14
19	6	1	4	24	21	2

(d) Prod. [p] Injetor [i]

Figura 49. Coordenadas dos cenários planejados.

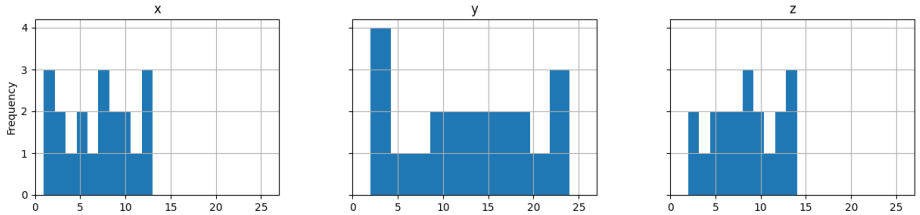


Figura 50. Histograma de valores de coordenadas para cenário 1, 1 produtor.

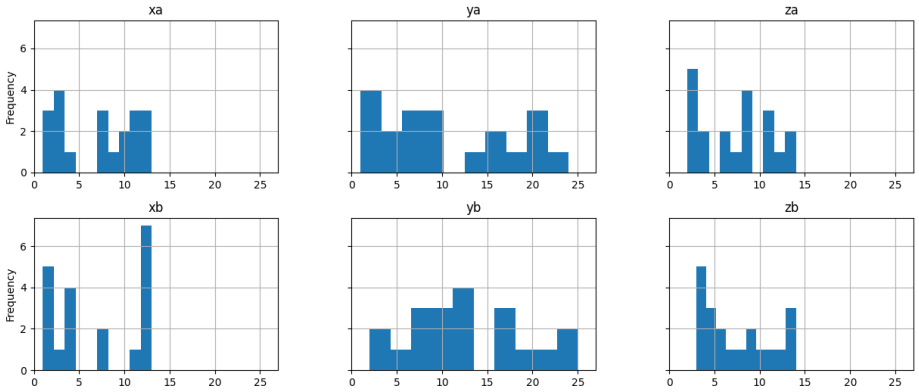


Figura 51. Histograma de valores de coordenadas para cenário 2, 2 produtores.

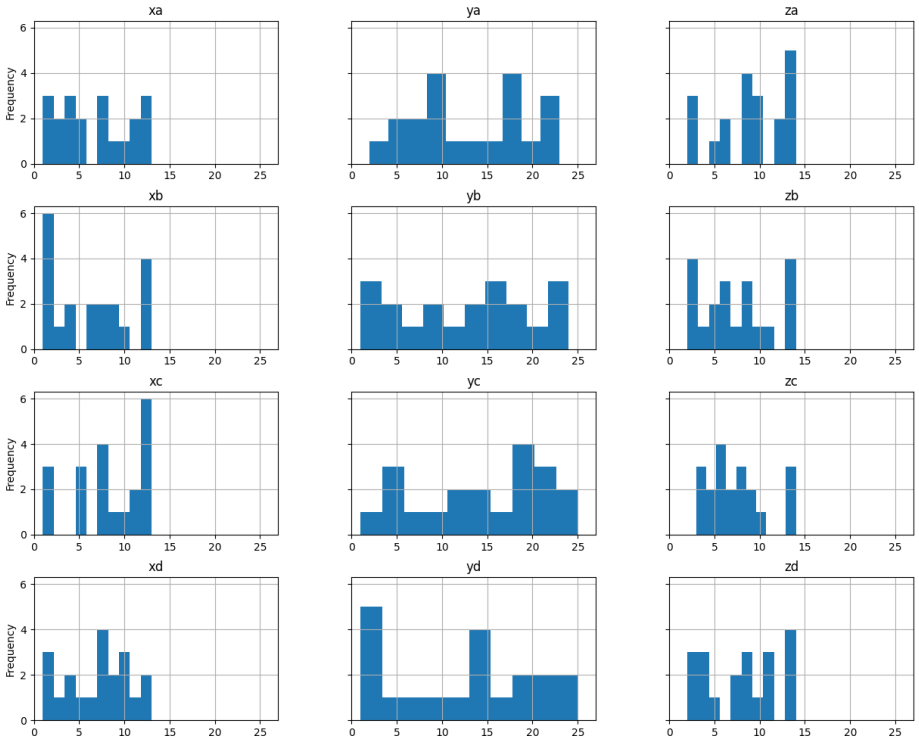


Figura 52. Histograma de valores de coordenadas para cenário 3, 4 produtores.

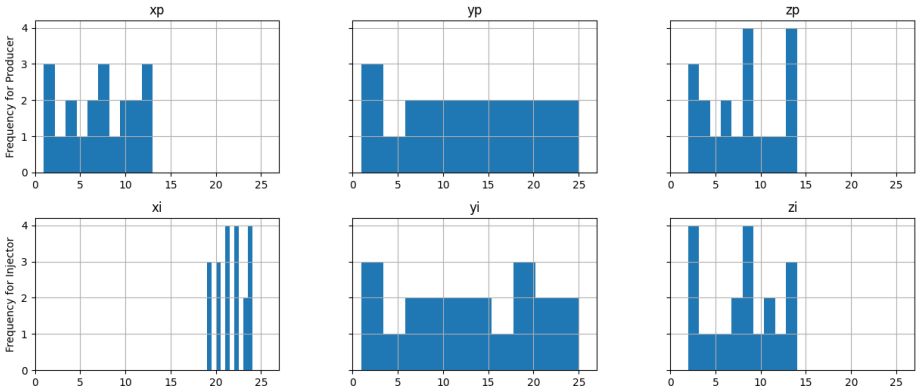
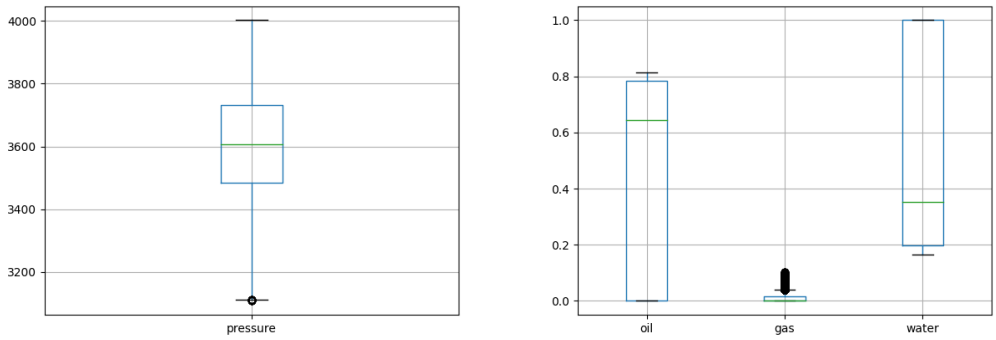


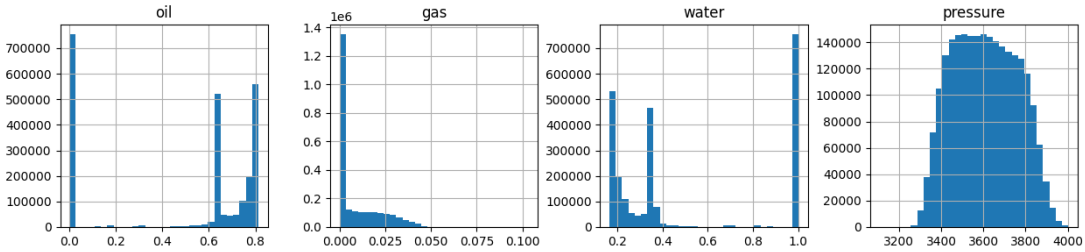
Figura 53. Histograma de valores de coordenadas para cenário 4, 1 produtor e 1 injetor.

C. BOXPLOT E HISTOGRAMA PARA CENÁRIOS SEGMENTADOS

Apresenta-se a seguir uma sequência contendo: gráfico tipo boxplot para valores de pressão, saturação de óleo, gás e água; histograma dos valores em todo o espaço amostral para as propriedades de pressão e saturações. Por limitação de espaço, são apresentados apenas um caso para cada cenário dentre os quatro, alternando os espaços amostrais indistintamente para cada um deles.

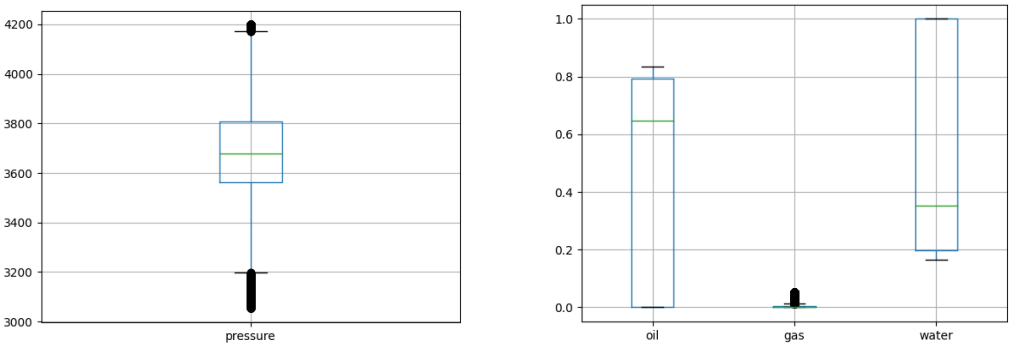


(a) Gráfico tipo boxplot para valores de pressão e saturações, cenário 1, segmento campo maduro

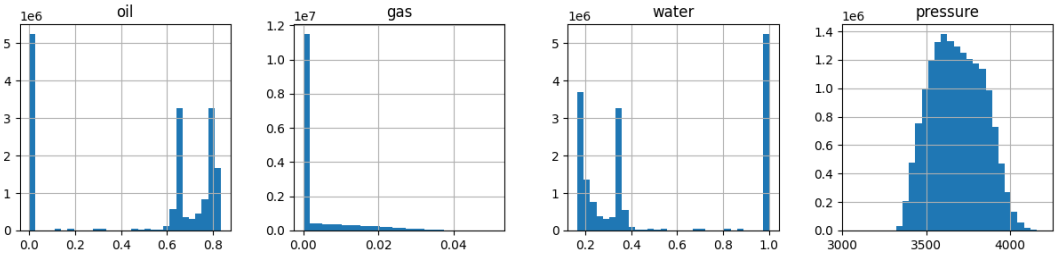


(b) histograma para valores de pressão e saturações, cenário 1, segmento campo maduro

Figura 54. Boxplot e histograma para cenário 1 segmento campo maduro.

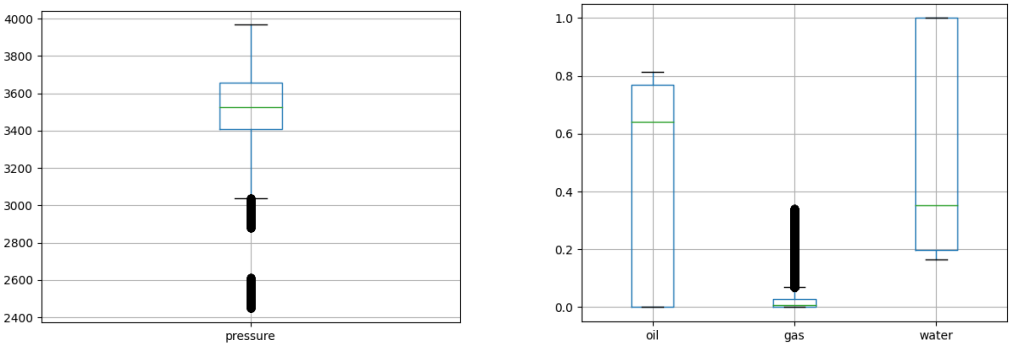


(a) Gráfico tipo boxplot para valores de pressão e saturações, cenário 2, segmento campo verde

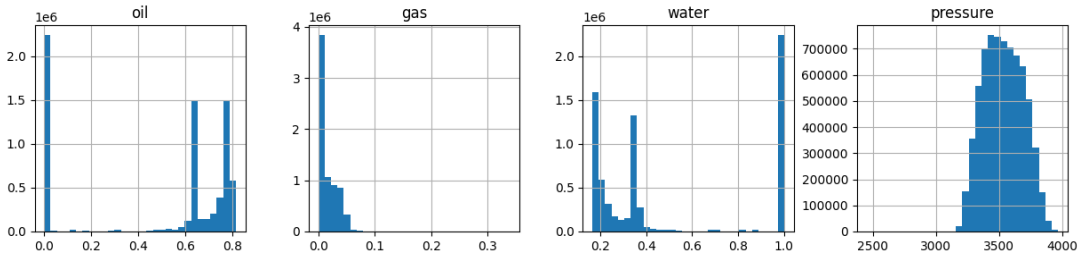


(b) histograma para valores de pressão e saturações, cenário 2, segmento campo verde

Figura 55. Boxplot e histograma para cenário 2 segmento campo verde.

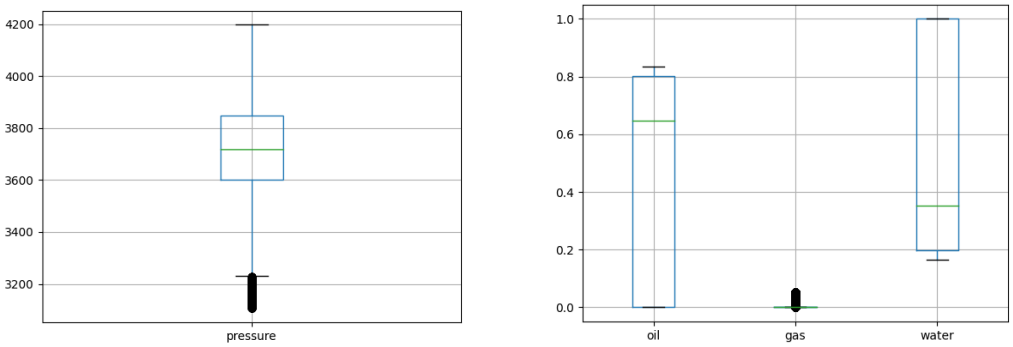


(a) Gráfico tipo boxplot para valores de pressão e saturações, cenário 3, segmento campo desenvolvimento

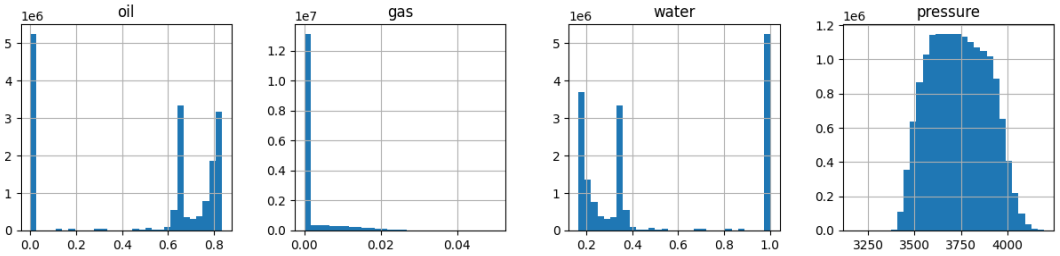


(b) histograma para valores de pressão e saturações, cenário 3, segmento campo desenvolvimento

Figura 56. Boxplot e histograma para cenário 3 segmento campo desenvolvimento.



(a) Gráfico tipo boxplot para valores de pressão e saturações, cenário 4, segmento campo verde



(b) histograma para valores de pressão e saturações, cenário 4, segmento campo verde

Figura 57. Boxplot e histograma para cenário 4 segmento campo verde.

D. CORRELAÇÃO ENTRE DESCRITORES (PEARSON E INFORMAÇÃO MÚTUA)

Os seguintes gráficos explicitam a dependência existente entre variáveis dependentes e independentes (descritores e saídas). São utilizados para seleção de descritores e utilizam a correlação de Pearson e Informação Mútua.

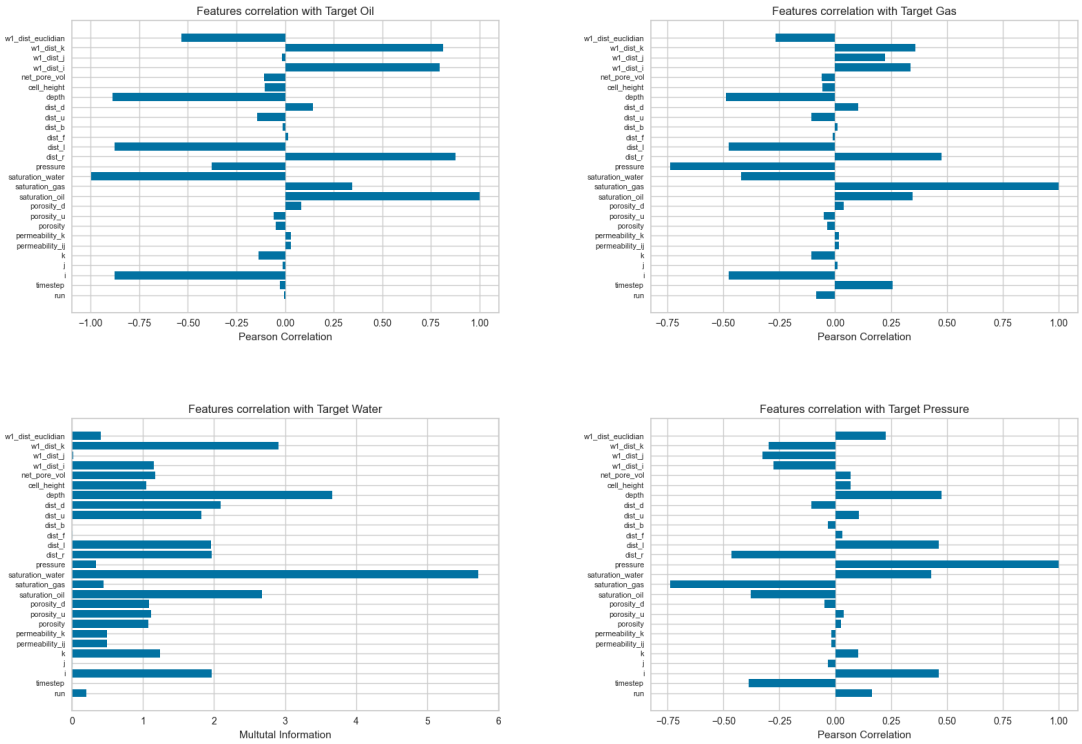
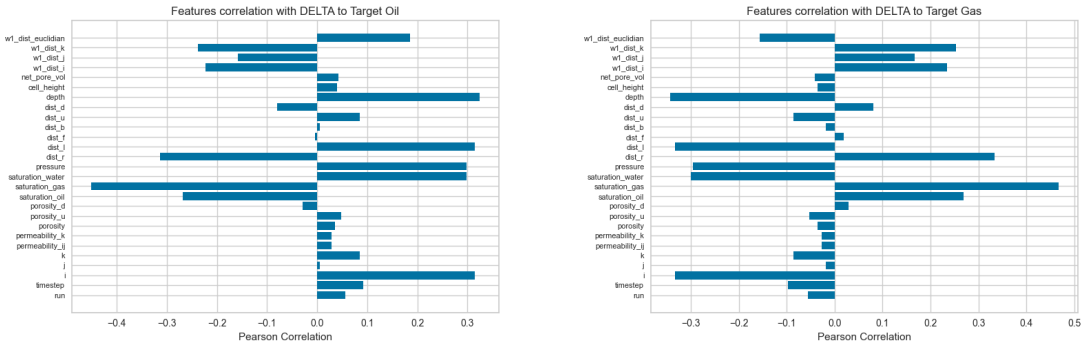


Figura 58. Correlação de Pearson com valor de saída desejado.



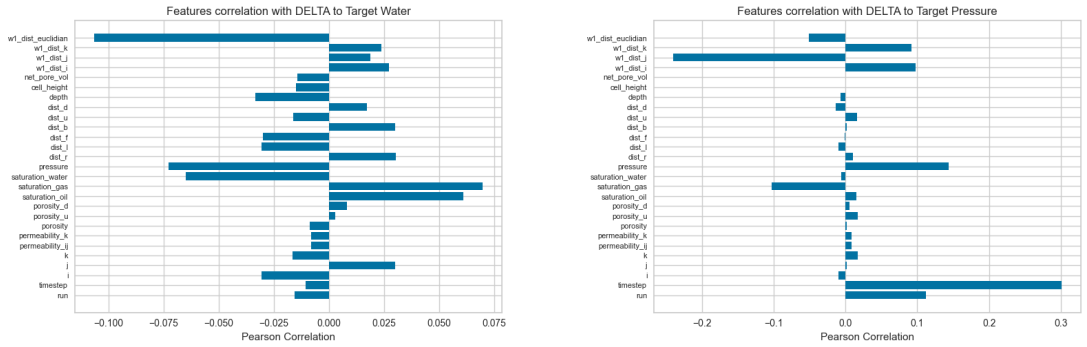


Figura 59. Correlação de Pearson com valor de saída de incremento (Delta).

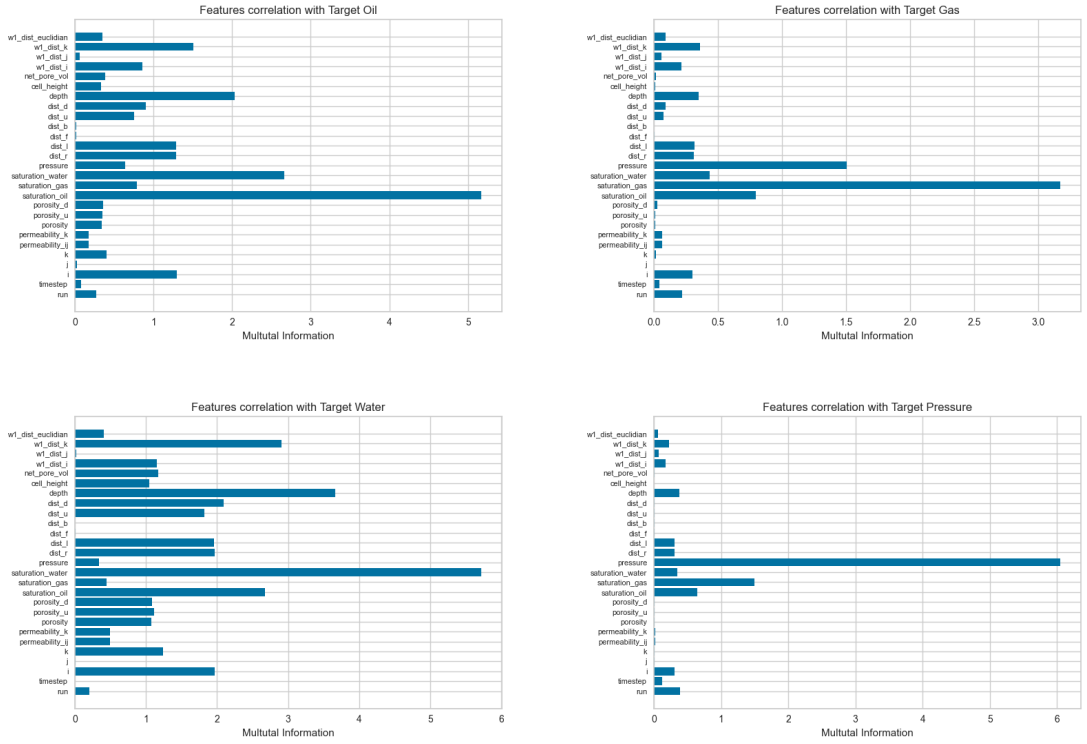
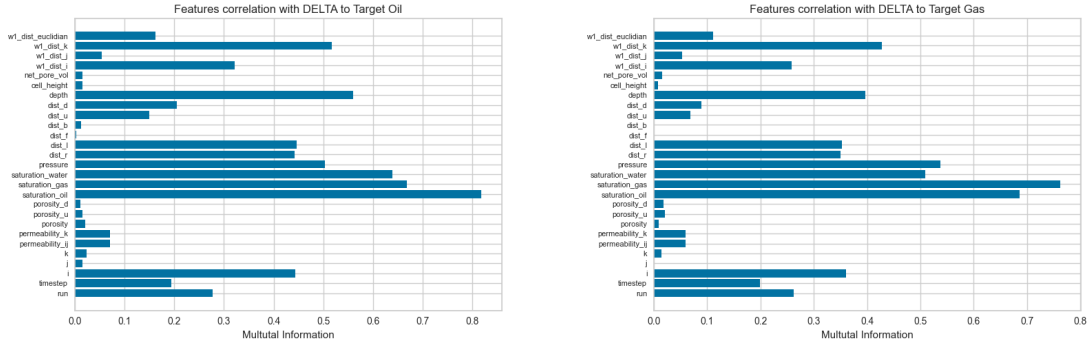


Figura 60. Informação Mútua com valor de saída desejado.



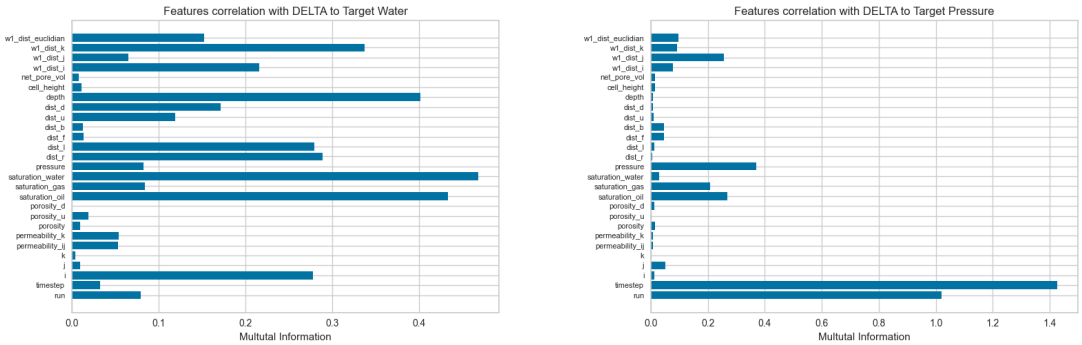
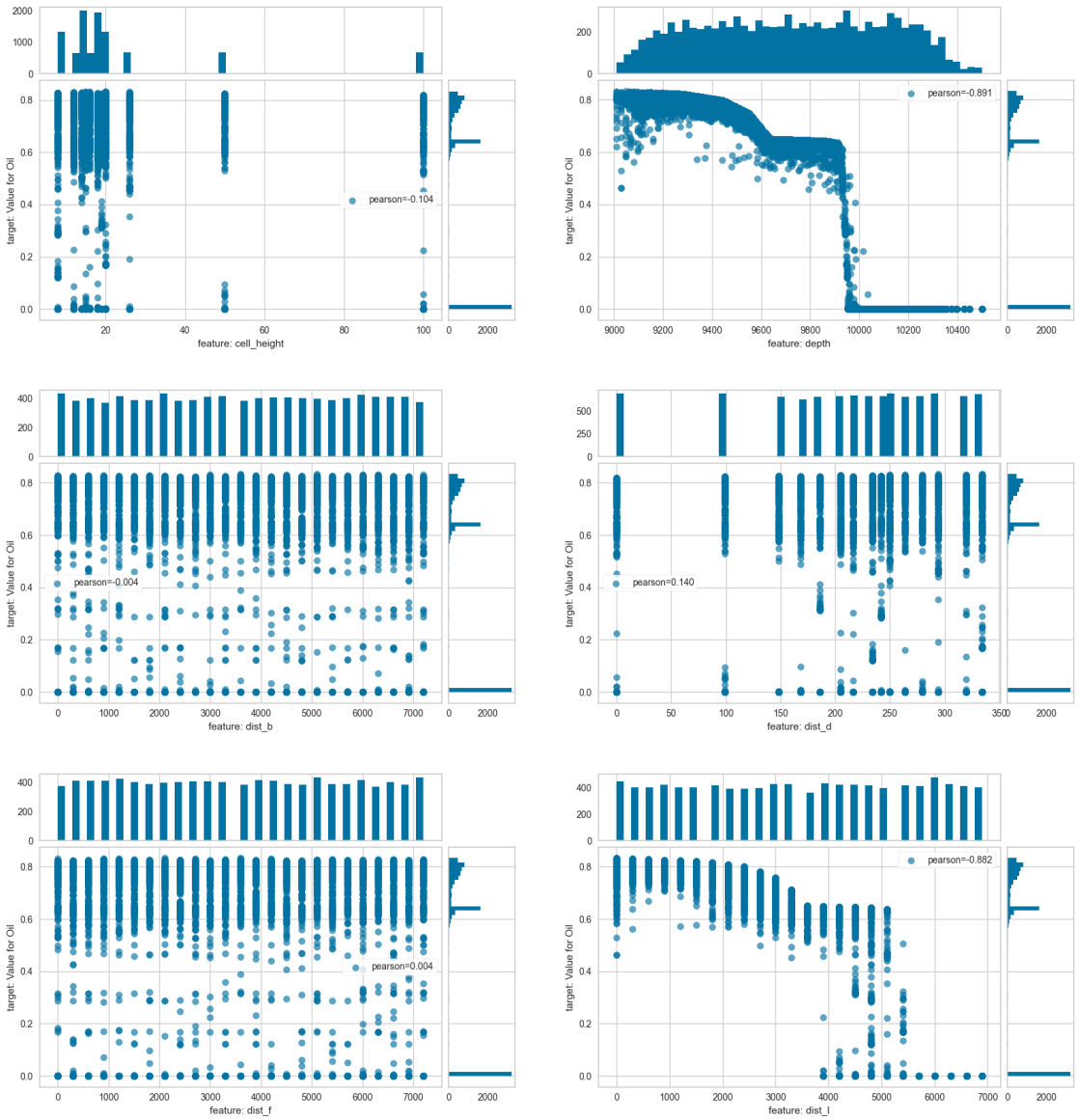
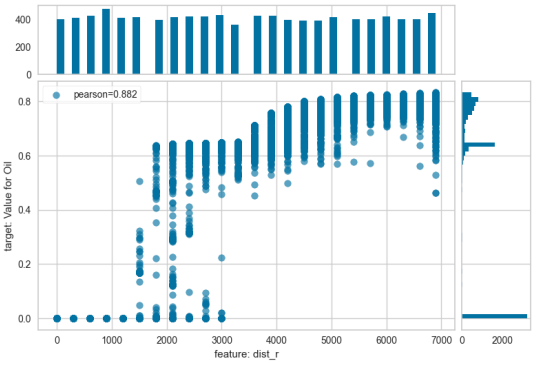


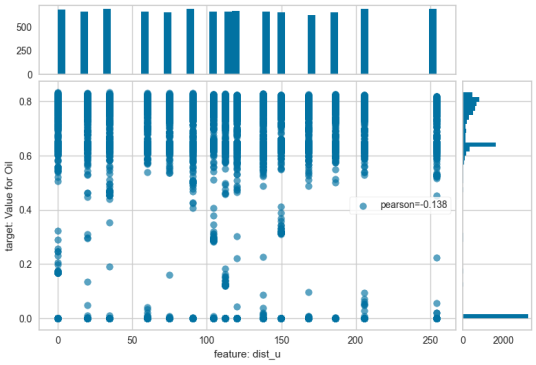
Figura 61. Informação Mútua com valor de incremento de saída (Delta) saída desejado.

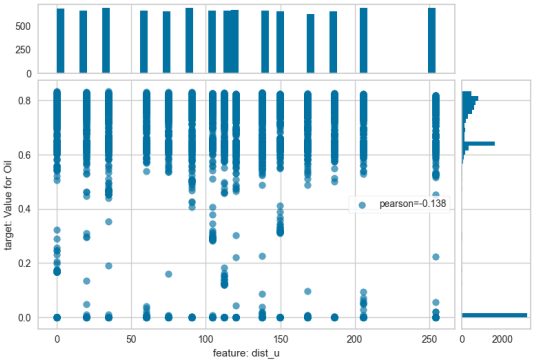
E. PLOT CONJUNTO (JOINTPLOT) PARA DESCRITORES E SAÍDAS DESEJADA

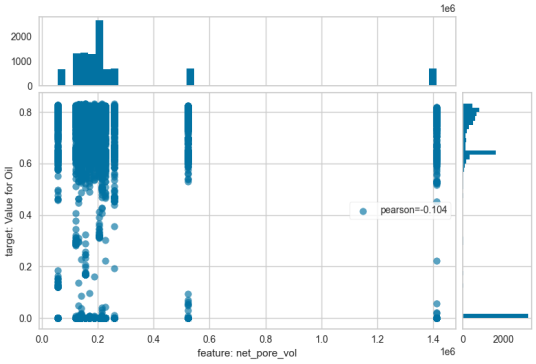
A seguir apresentam-se as correlações visuais para a saída Óleo as relações para vinte e dois (22) descritores.

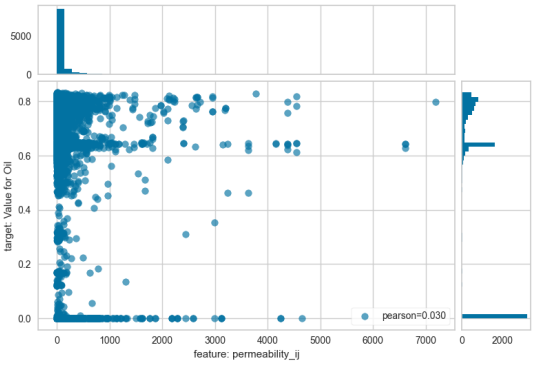


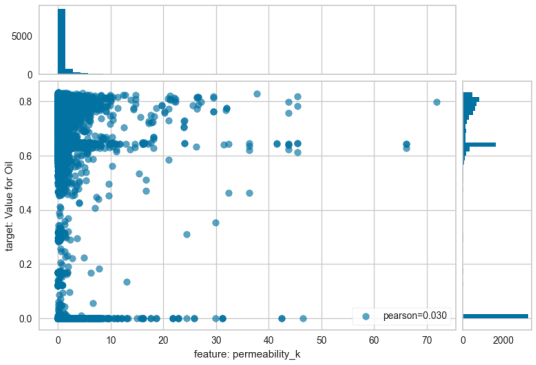

pearson=0.882
target: Value for Oil
feature: dist_r

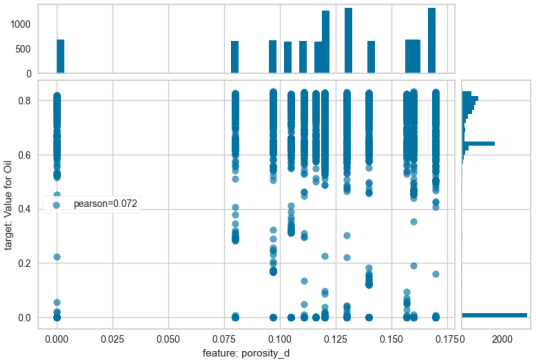

pearson=-0.138
target: Value for Oil
feature: dist_u

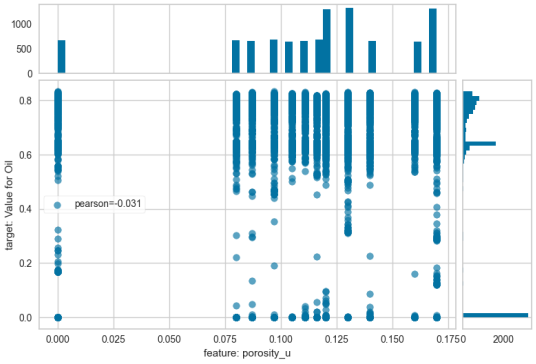

pearson=-0.138
target: Value for Oil
feature: dist_u

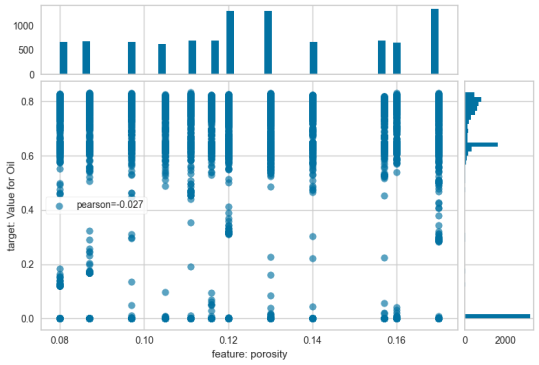

1e6
pearson=-0.104
target: Value for Oil
feature: net_pore_vol
1e6

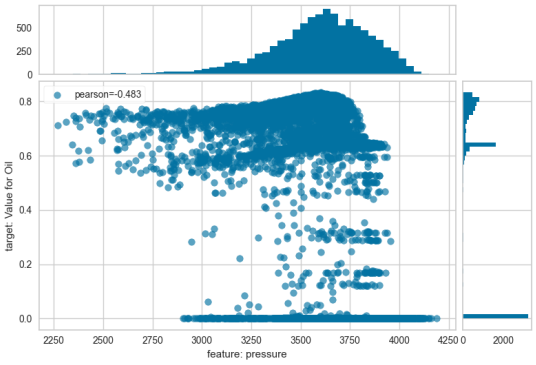

pearson=0.030
target: Value for Oil
feature: permeability_ij

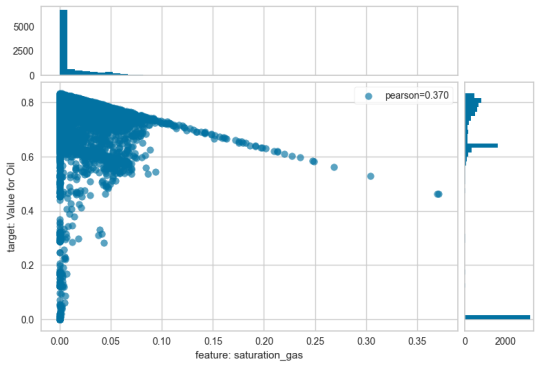

pearson=0.030
target: Value for Oil
feature: permeability_k

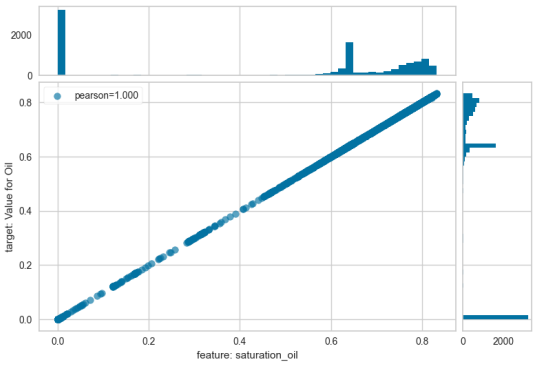

pearson=0.072
target: Value for Oil
feature: porosity_d

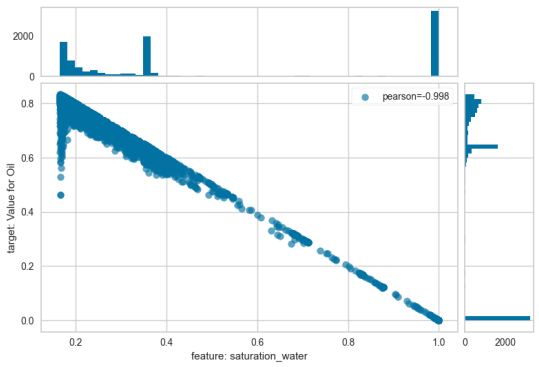

pearson=-0.031
target: Value for Oil
feature: porosity_u

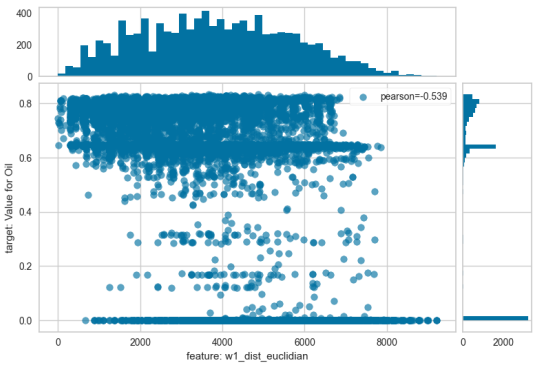

pearson=-0.027
target: Value for Oil
feature: porosity

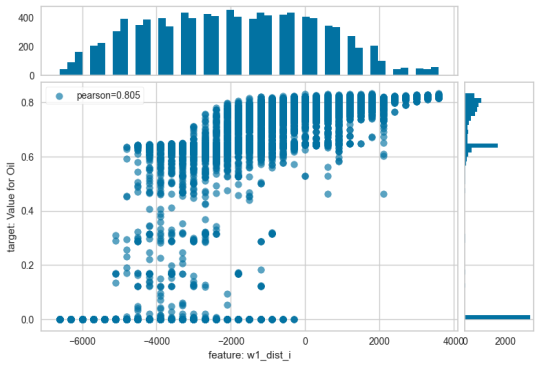

pearson=-0.483
target: Value for Oil
feature: pressure

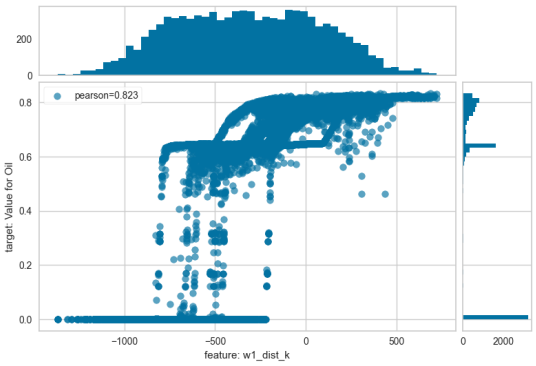

pearson=0.370
target: Value for Oil
feature: saturation_gas


pearson=1.000
target: Value for Oil
feature: saturation_oil


pearson=-0.998
target: Value for Oil
feature: saturation_water


pearson=-0.539
target: Value for Oil
feature: w1_dist_euclidian


pearson=0.805
target: Value for Oil
feature: w1_dist_i


pearson=0.823
target: Value for Oil
feature: w1_dist_k