UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

Vagner Vilela de Oliveira

MODELAGEM DA PRESSÃO DE FUNDO DE POÇO EM SISTEMAS DE ESCOAMENTO MULTIFÁSICO: UMA ABORDAGEM UTILIZANDO MODELOS DE REDES NEURAIS POLINOMIAIS

Juiz de Fora
2024

Vagner Vilela de Oliveira

MODELAGEM DA PRESSÃO DE FUNDO DE POÇO EM SISTEMAS DE ESCOAMENTO MULTIFÁSICO: UMA ABORDAGEM UTILIZANDO MODELOS DE REDES NEURAIS POLINOMIAIS

Dissertação apresentada ao Programa de Pós Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Modelagem Computacional.

Orientador: Doutor Leonardo Goliatt da Fonseca

Coorientadora: Doutora Camila Martins Saporetti

Juiz de Fora

2024

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Sobrenome, Nome do autor.
MODELAGEM DA PRESSÃO DE FUNDO DE POÇO EM SISTEMAS DE ESCOAMENTO MULTIFÁSICO: UMA ABORDAGEM UTILIZANDO MODELOS DE REDES NEURAIS POLINOMIAIS / Vagner Vilela de Oliveira. – 2024.
57 f. : il.

Orientador: Leonardo Goliatt da Fonseca
Coorientadora: Camila Martins Saporetti
Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós Graduação em Modelagem Computacional, 2024.

1. Palavra-chave. 2. Palavra-chave. 3. Palavra-chave. I. Sobrenome, Nome do orientador, orient. II. Título.

Vagner Vilela de Oliveira

MODELAGEM DA PRESSÃO DE FUNDO DE POÇO EM SISTEMAS DE ESCOAMENTO MULTIFÁSICO: UMA ABORDAGEM UTILIZANDO MODELOS DE REDES NEURAIS POLINOMIAIS

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Modelagem Computacional. Área de concentração: Modelagem Computacional.

Aprovada em 26 de setembro de 2024.

BANCA EXAMINADORA

Prof. Dr. Leonardo Goliatt da Fonseca - Orientador

Universidade Federal de Juiz de Fora

Prof.ª Dr.ª Camila Martins Saporetti - Coorientadora

Universidade do Estado do Rio de Janeiro

Prof. Dr. Raul Fonseca Neto

Universidade Federal de Juiz de Fora

Prof. Dr. Iago Augusto de Carvalho

Universidade Federal de Alfenas

Juiz de Fora, 18/09/2024.

Documento assinado eletronicamente por **Raul Fonseca Neto**, **Professor(a)**, em 26/09/2024, às 16:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Leonardo Goliatt da Fonseca**, **Professor(a)**, em 26/09/2024, às 17:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Iago Augusto de Carvalho**, **Usuário Externo**, em 26/09/2024, às 17:16, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Camila Martins Saporetti**, **Usuário Externo**, em 26/09/2024, às 17:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **1992103** e o código CRC **31833BA0**.

Dedico este trabalho à Paula e ao Davi, pela força e inspiração. À Dani e à Dê, pela força e alegria. À Tia Auxiliadora e à Tia Altivina, pela energia. Ao meu pai e à minha mãe.

AGRADECIMENTOS

Agradeço à Jotaha na figura de seu Diretor, Sr. Luiz pela liberação para assistir as aulas ainda nas disciplinas isoladas. Bem como a Nova Tendência na figura de meu Coordenador Augusto Beviláqua e a Luiz Severo meu Líder Técnico, pela compreensão na gestão de meus horários.

Ao meu amigo Deivid pelo apoio no tema e pelas cervejadas (que deveriam ser mais frequentes).

À Paula, pelo apoio inconteste e ao Davi pela inspiração de sempre.

Ao *primovski* Adauto Villela pelos *master minds* que sempre me fizeram pensar à frente.

Aos meus orientadores Leonardo Golliat e Camila Saporetti pela confiança.

deus o vento tão sem vento quanto o mundo por trás de uma tela de computador (Jane Miller, "Dias Santos", em Pai Nosso Computador de John Updike)

RESUMO

No complexo processo de extração de petróleo e gás, há muitos indicadores importantes que devem ser acompanhados para manter a eficiência e segurança dos poços, do meio-ambiente e principalmente dos profissionais envolvidos na operação. Um desses indicadores, é a Pressão de Fundo de Poço (FBHP). Esse indicador influencia diretamente o controle de produção e a integridade dos poços, sendo um parâmetro crucial para evitar falhas operacionais e minimizar riscos. Tanto a FBHP baixa (que reduz a produção) quanto alta (que acarreta riscos diversos, entre eles o risco de explosão) devem ser foco constante de atenção. Nesse sentido, modelos computacionais têm sido desenvolvidos para prever a FBHP, fornecendo formas adicionais ou alternativas às medidas tradicionais desse indicador. Essa dissertação explora a técnica do GMDH (*Group Method of Data Handling*) na modelagem preditiva da FBHP. O GMDH se refere a uma família de algoritmos ainda pouco explorada, sobretudo na área de petróleo e gás. Essa pesquisa visa preencher essa lacuna, avaliando a técnica como uma alternativa possível na modelagem de FBHP. É feita uma análise comparativa entre quatro diferentes tipos de algoritmos GMDH, apontando vantagens e limitações em termos de precisão, sendo gerados modelos interpretáveis relativos aos algoritmos em sua melhor configuração de desempenho.

Palavras-chave: pressão de fundo de poço; fbhp; rede neural polinomial; gmdh; modelagem computacional; modelo interpretável; inteligência artificial; aprendizado de máquina.

ABSTRACT

In the complex oil and gas extraction process, there are many important indicators that must be monitored to maintain the efficiency and safety of the wells, the environment and especially the professionals involved in the operation. One of these indicators is Flowing Bottom Hole Pressure (FBHP). This indicator directly influences production control and well integrity, being a crucial parameter to avoid operational failures and minimize risks. Both low FBHP (which reduces production) and high FBHP (which entails various risks, including the risk of explosion) must be a constant focus of attention. In this sense, computational models have been developed to predict FBHP, providing additional or alternative forms to traditional measurements of this indicator. This dissertation explores the GMDH (Group Method of Data Handling) technique in FBHP predictive modeling. GMDH refers to a family of algorithms that is still little explored, especially in the oil and gas area. This research aims to fill this gap, evaluating the technique as a possible alternative in FBHP modeling. A comparative analysis is made between four different types of GMDH algorithms, pointing out advantages and limitations in terms of precision, generating interpretable models relating to the algorithms in their best performance configuration.

Keywords: flowing bottom hole pressure; fbhp; polynomial neural network; gmdh; computational modeling; interpretable model; artificial intelligence; machine learning.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ANN	Rede Neural Artificial (*Artificial Neural Network*)
API	Ídicie de Gravidade Específica do Petróleo (*American Petroleum Institute Gravity*)
BPNN	Rede Neural de Retropropagação (*Backpropagation Neural Network*)
ELM	*Extreme Learning Machine*
FBHP	Pressão de Fundo de Poço (*Flowing Bottom-hole Pressure*)
GFR	Taxa de Fluxo de Gás (*Gas Flow Rate*)
GMDH	Método de Agrupamento de Dados (*Group Method of Data Handling*)
IC	Inteligência Computacional
ID	Diâmetro Interno do Tubo (*Internal Diameter of Pipe*)
KNN	k-Vizinhos mais Próximos (*k-Nearest Neighbors*)
LSTM	Modelo de Memória de Longo Prazo (*Long Short-Term Memory*)
MARS	*Multivariate Adaptive Regression Splines*
MAE	Erro Médio Absoluto (*Mean Absolute Error*)
MAPE	Erro Percentual Médio Absoluto (*Mean Absolute Percentage Error*)
ML	Aprendizado de Máquina (*Machine Learning*)
MSE	Erro Quadrático Médio (*Mean Squared Error*)
OFR	Taxa de Fluxo de Óleo (*Oil Flow Rate*)
PG	Programação Genética
PSO	Otimização por Enxame de Partículas (*Particle Swarm Optimization*)
R	Coeficiente de correlação
R^2	Coeficiente de determinação
RF	Floresta Aleatória (*Random Forest*)
RMSE	Erro Quadrático Médio Residual (*Root Mean Squared Error*)
RBFNN	Rede Neural de Função de Base Radial (*Radial Basis Function Neural Network*)
SVR	*Support Vector Regression*
WBHT	Temperatura na cabeça do poço (*Wellbore Head Temperature*)
WFR	Taxa de Fluxo de Água (*Water Flow Rate*)
WHP	Pressão na Cabeça do Poço (*Wellhead Pressure*)
WPD	Produção Diária de Água (*Water Production Rate*)

SUMÁRIO

1 INTRODUÇÃO

O petróleo é uma das *commodities* mais importantes e valiosas do mundo, desempenhando um papel crucial na economia global. Ele não só é a principal fonte de energia para o transporte, aquecimento e geração de eletricidade, mas também é um componente essencial na fabricação de produtos como plásticos, produtos químicos e farmacêuticos (9). O setor de petróleo e gás contribui significativamente para o Produto Interno Bruto (PIB) de muitos países e é uma fonte vital de receitas fiscais e criação de empregos.

Atualmente, o petróleo representa cerca de 31% do consumo total de energia mundial, fazendo dele a maior fonte de energia, seguida pelo carvão e pelo gás natural (10). Apesar dos esforços globais para transição para energias renováveis, a demanda por petróleo continua alta devido ao crescimento econômico e ao aumento da urbanização, especialmente em países em desenvolvimento. Essa dificuldade de transição para energias renováveis é uma preocupação mundial a ponto de em 2011, a Agência Internacional de Energia (IEA) sugerir que os países deveriam interromper a construção de novos campos de exploração de petróleo e gás para manter as temperaturas globais dentro de limites seguros e alcançar as metas de emissões líquidas zero até 2050 (13). A motivação para essa sugestão foi a necessidade urgente de reduzir as emissões de gases de efeito estufa e combater a crise climática (11) (12).

A Petrobras, empresa estatal brasileira responsável por toda atividade de exploração, avaliação, desenvolvimento e produção de óleo gás no Brasil, prevê em seu Plano Estratégico 2024-2028 a perfuração de 50 novos poços em áreas onde possui direito de exploração (34). Essas atividades relativas à cadeia produtiva de óleo e gás (O&G) são extremamente complexas e um desafio constante para diversas áreas do conhecimento tais como Geologia, Física e Engenharias, onde são extensamente aplicadas ferramentas matemáticas e computacionais (47) (48) (49). O chamado poço de produção ou desenvolvimento, que é o tipo em que se drena o petróleo de um campo, é submetido constantemente a várias medidas de monitoração de sua saúde (1). Algumas dessa métricas são vazão, temperatura, monitoramento de composição de fluidos, monitoramento de intergridade do poço, e a pressão no fundo do poço em escoamento. Essa última é a pressão conhecida como FBHP (*Flowing Bottom Hole Pressure*), que é o objeto de modelagem desse trabalho.

A medição da FBHP é um aspecto crítico na gestão da produção de petróleo e gás. Tradicionalmente, a FBHP é medida usando manômetros de fundo de poço ou calculada através de modelos empíricos e simuladores de reservatório. Vários tipos de bombas de pressão têm sido usadas para medir a pressão no fundo dos poços. Uma delas é um pedaço de tubo de aço com uma válvula de retenção no fundo e uma conexão para um manômetro no topo (36). No entanto, essas técnicas apresentam diversos desafios e limitações. A instalação e manutenção de manômetros de fundo de poço podem ser

perigosas, expondo os trabalhadores a condições adversas e ambientes de alta pressão (50). O custo de instalação e manutenção dos equipamentos de medição é elevado, além dos custos associados à interrupção da produção durante a instalação e manutenção. A precisão das medições pode ser comprometida por diversos fatores, como incrustações nos equipamentos, mudanças nas condições do reservatório e falhas nos sensores (51).

1.1 MOTIVAÇÃO

Falhas na leitura precisa da FBHP podem resultar em desastres significativos, tanto em termos de segurança quanto de impacto ambiental. Alguns exemplos incluem *blowouts*, que é a incapacidade de monitorar e controlar a pressão do poço, que pode levar a erupções descontroladas de petróleo ou gás, causando explosões e incêndios. Falhas na gestão da pressão do poço podem resultar em vazamentos e derramamentos de óleo, causando danos ambientais severos e custos elevados de limpeza e recuperação (11). A Figura 1.1 demonstra a magnitude do dano que o excesso de pressão pode causar em uma coluna de extração. A imagem se refere a um revestimento de 16" colapsado no poço Pompano A-31 no Golfo do México (53).

Figura 1.1 – Revestimento de 16"colapsado

Fonte: Adaptada de (54)

Um exemplo contundente dessa falha na leitura foi o acidente ocorrido em 20 de abril de 2010, em um poço em desenvolvimento em águas profundas no campo de Marlin, no Golfo do México (14). A explosão da plataforma Deepwater Horizon foi um dos piores desastres ambientais da história. Estima-se que aproximadamente 4,9 milhões de barris de petróleo foram derramados no Golfo do México durante os 87 dias em que o poço esteve descontrolado. Milhares de aves, peixes, tartarugas marinhas e mamíferos marinhos morreram devido à exposição ao petróleo. Manguezais, pântanos, e recifes de corais sofreram danos severos, prejudicando os ecossistemas que dependem desses habitats. O petróleo contaminou a cadeia alimentar marinha, afetando tanto a vida marinha quanto

os seres humanos que dependem dos frutos do mar para alimentação. A recuperação total dos ecossistemas afetados pode levar décadas. Estudos indicam que os impactos do derramamento de óleo e dos dispersantes ainda são visíveis muitos anos após o desastre (15), (16), (17).

Essa situação coloca o tema da dissertação em alinhamento com vários dos 17 Objetivos de Desenvolvimento Sustentável (ODS) estabelecidos pela Organização das Nações Unidas (ONU). Dois desses objetivos são bastante aderentes ao trabalho que são o ODS 9 (Indústria, Inovação e Infraestrutura) e ODS 12 (Consumo e Produção Responsáveis) (18). O ODS 9 tendo como meta "Construir infraestruturas resilientes, promover a industrialização inclusiva e sustentável e fomentar a inovação"se alinha ao trabalho na medida em que a utilização de técnicas ML representa alguma inovação no setor de petróleo e gás. Melhorar a previsão de FBHP pode levar a operações mais seguras e eficientes, promovendo uma indústria mais resiliente. Já a meta do ODS 12 ("Assegurar padrões de produção e de consumo sustentáveis") também demonstra um alinhamento com este estudo no momento em que a produção de petróleo e gás através de previsões precisas propicia uma redução da pegada ambiental das operações e garante que os recursos sejam usados de forma mais eficiente e sustentável. Além desses dois ODS, vários outros como ODS 7: Energia Limpa e Acessível, ODS 13: Ação Contra a Mudança Global do Clima, ODS 14: Vida na Água e ODS 15: Vida Terrestre podem também de alguma forma ter algum alinhamento com esta dissertação. Esses alinhamentos demonstram a relevância e a contribuição potencial da pesquisa em um contexto mais amplo de desenvolvimento sustentável.

A crescente demanda por métodos eficazes e precisos na previsão de parâmetros críticos de produção petrolífera, como a FBHP, que influencia diretamente a eficiência e a segurança das operações de extração de petróleo e gás é uma das motivações do trabalho. A previsão precisa da FBHP é crucial para otimizar a produção, reduzir custos operacionais e minimizar riscos ambientais, o que será detalhado na Seção 1.1. Com o desenvolvimento das tecnologias de Inteligência Artificial, mais especificamente nas áreas de Aprendizado de Máquina e Redes Neurais, surgem oportunidades significativas para aplicar técnicas que possam superar as limitações dos métodos tradicionais de modelagem.

Modelos como RNA (Rede Neural Artificial), *Deep Learning*, Lógica Fuzzy, Árvore de Decisão, *Random Forest* ou modelos híbridos tem sido usados para modelar o domínio O&G. A Figura 1.2 mostra um aumento nesta área de pesquisa nos últimos anos (42). Esse estudo coletou artigos indexados no *Web of Science*, *Science Direct*, *Scopus* e IEE, especificamente sobre uso de modelagem preditiva utilizando *Machine Learning* para modelos na área de O&G. Somente um artigo foi encontrado ulilizando a técnica de GMDH. Trata-se de (27) GAO, em um trabalho de predição de pressão em poros. Nota-se que há espaço para mais estudos para avaliar novos modelos e conjuntos de dados com relação à adequação, assertividade e diversidade técnica.

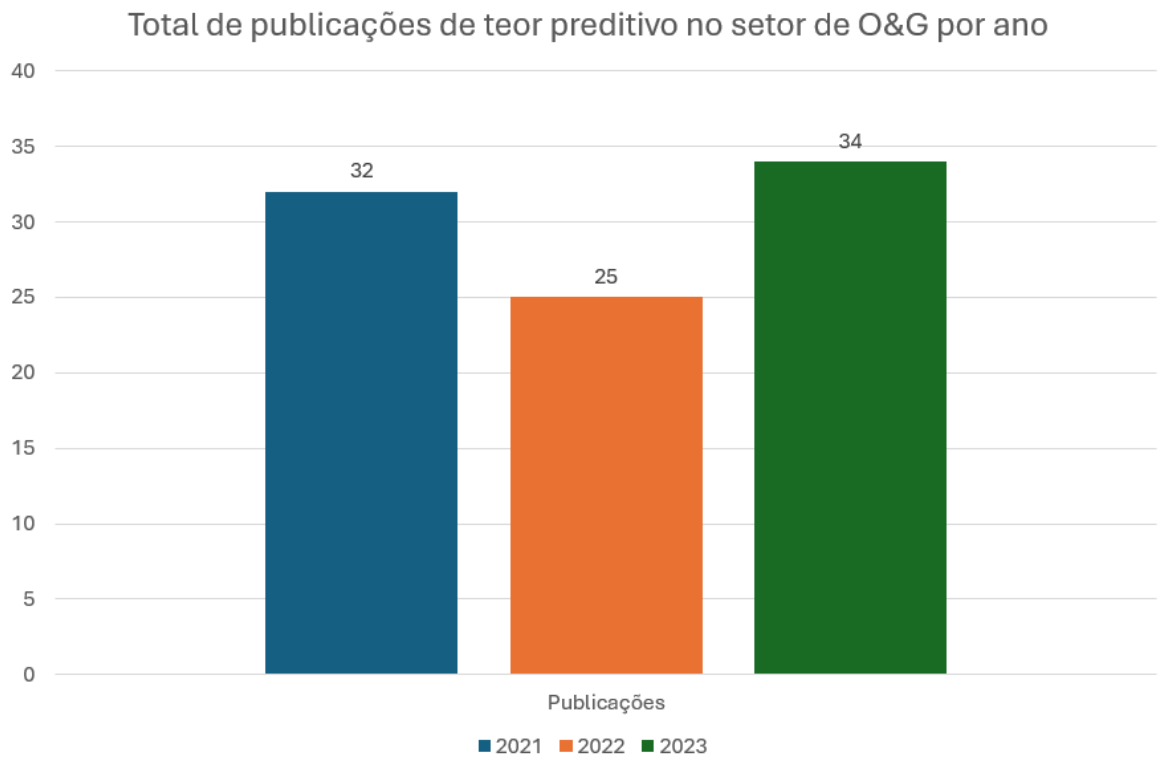

Figura 1.2 – Publicações de teor preditivo no setor de O&G

Fonte: Elaborada pelo autor (2024)

Embora os métodos de Inteligência Artificial ofereçam um grande potencial para a modelagem de sistemas complexos, existem várias limitações a serem consideradas. Primeiramente, a qualidade e a quantidade de dados disponíveis que podem impactar significativamente a precisão dos modelos preditivos. Dados incompletos ou ruidosos podem levar a resultados menos confiáveis. Além disso, a complexidade dos modelos pode aumentar substancialmente com a adição de mais variáveis, exigindo maior poder computacional e tempo de processamento (42).

Comumente, os primeiros métodos empregados no cálculo da FBHP eram baseados em correlações empíricas como em Ros (37), Hagerdon (38) e Orkiszeski (39). Em poços de produção de petróleo ou gás, o fluxo multifásico geralmente consiste em óleo, gás e água. Devido à complexidade do fluxo multifásico, várias abordagens foram adotadas para explicar e avaliar o comportamento do fluxo. O foco se voltou para técnicas de modelagem mecanicista baseadas em princípios hidrodinâmicos (40) (41). Nos últimos dois anos, os modelos de algoritmos de ML (Machine Learning) têm sido amplamente usados em análises preditivas de O&G para abordar as limitações dos modelos matemáticos clássicos. A Figura 1.3 mostra um gráfico representando a distribuição do modelo de análise preditiva (42), ênfase em abordagens de classificação relativas a reservatórios e poços.

Em suma, esta dissertação busca contribuir para a melhoria das técnicas de previsão de FBHP, reconhecendo as limitações inerentes e propondo soluções para superar esses

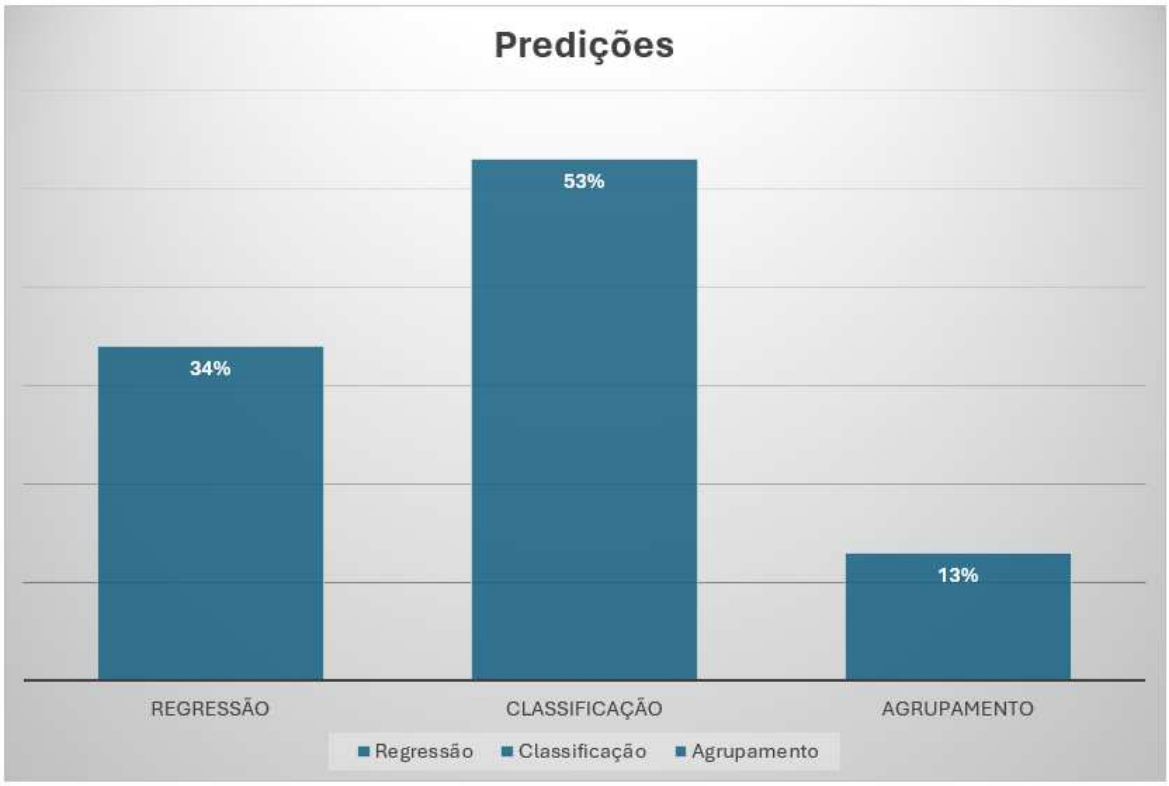

Figura 1.3 – Distribuição das predições na área de O&G
Fonte: Elaborada pelo autor (2024)

desafios. Através da aplicação de métodos indutivos e da análise dos resultados obtidos, espera-se oferecer *insights* valiosos para pesquisadores da indústria de petróleo e gás. Utilizou-se, no trabalho, uma técnica de Redes Neurais Polinomiais (RNP) conhecida com GMDH (*Group Method of Data Handling*). Apesar de já utilizada na área de O&G, como demonstrado na Seção 1.5, a técnica foi pouco utilizada para previsão de FBHP. O trabalho pretende, nesse sentido, colocar o GMDH como mais uma alternativa a ser explorada para a previsão desse indicador.

1.2 A PRESSÃO DE FUNDO DE POÇO (FBHP)

O petróleo, principal alvo da produção, pode existir em diversas viscosidades e composições dependendo das características específicas do reservatório. A água também está presente em muitos poços de petróleo, seja na forma de água de formação (uma fase) ou de água injetada para métodos de recuperação secundária. No contexto dos poços de petróleo, o termo “fases” refere-se aos diferentes estados em que os hidrocarbonetos e outras substâncias podem existir nas condições encontradas no reservatório e no poço. Essas fases podem incluir hidrocarbonetos líquidos (petróleo), hidrocarbonetos gasosos (gás natural) e água. Compreender os tipos de fases presentes em um poço é essencial para os fenômenos físicos inerentes ao sistema como temperatura e pressão (46) (35) (36).

Além destas fases primárias (óleo, gás e água), os poços de petróleo também podem conter outras substâncias, como o sulfureto de hidrogênio (H_2S) e o dióxido de carbono (CO_2), que devem ser cuidadosamente geridas devido ao seu potencial impacto na segurança e na produção (43).

Quando a pressão natural do reservatório não é suficiente para elevar os fluidos até a superfície, é necessário o uso de sistemas de elevação artificial, como bombas submersas elétricas (ESP), bombas de cavidade progressiva (PCP) ou gás-lift. As bombas, ao alterarem o gradiente de pressão ao longo do poço, podem mudar o comportamento do fluxo multifásico e, por consequência, a distribuição das fases ao longo do poço (44).

O fluxo vertical refere-se ao movimento do fluido dentro do poço, que é influenciado por uma variedade de mecanismos e variáveis. Esses mecanismos incluem fluxo natural, elevação artificial já citada e a interação entre o fluido e o reservatório circundante. As variáveis que impactam o fluxo vertical incluem a viscosidade do fluido, os diferenciais de pressão dentro do poço e a presença de quaisquer obstruções ou irregularidades no poço (37).

No caso do fluxo natural, o movimento de fluidos dentro do poço é impulsionado principalmente pelos diferenciais de pressão entre o reservatório e a cabeça do poço. Isto pode ser influenciado por fatores como a permeabilidade da rocha reservatório, a viscosidade do fluido e a geometria geral do poço. Na tubulação, por exemplo, a dinâmica do fluxo é influenciada por fatores como taxa de fluxo, propriedades do fluido e geometria do poço. No revestimento e no anel, a dinâmica do fluxo é afetada por fatores como viscosidade do fluido, diferenciais de pressão e presença de fases gasosas ou líquidas (45).

Um fenômeno crítico é a pressão do ponto de bolha. Refere-se à pressão na qual a primeira bolha de gás é formada no fluido do reservatório à medida que sofre separação de fases. Este fenômeno é de grande importância na produção de petróleo e gás, pois afeta diretamente o comportamento do reservatório e o desempenho do poço. Compreender a pressão do ponto de bolha é crucial para determinar a estratégia de produção ideal, estimar as reservas recuperáveis e garantir a produtividade do poço a longo prazo. A pressão do ponto de bolha de um fluido de reservatório é influenciada por vários fatores, incluindo a composição do fluido, temperatura e condições de pressão. Em termos práticos, a determinação da pressão do ponto de bolha envolve testes e análises laboratoriais, utilizando equipamentos e procedimentos especializados (51).

A interação de diversos parâmetros e variáveis como temperatura, pressão e composição influencia significativamente o comportamento não linear dos fluidos em reservatórios de petróleo. Para compreender o comportamento não linear dos fluidos, é essencial considerar a influência de parâmetros e variáveis no desempenho geral do reservatório. A não linearidade de deve à complexa interação de fatores como comportamento de fases, composição de fluidos e heterogeneidade e permeabilidade do reservatório (52).

A permeabilidade do reservatório, que é a capacidade da rocha de permitir o fluxo de fluidos através de seus poros, é um fator determinante na FBHP. Reservatórios com alta permeabilidade permitem que o fluido flua mais livremente, o que pode resultar em uma menor FBHP, enquanto reservatórios com baixa permeabilidade podem causar um aumento na pressão devido à resistência ao fluxo. A diagênese, que é o processo de alteração das rochas após sua deposição, pode modificar significativamente a permeabilidade do reservatório. Processos diagenéticos, como a compactação, cimentação ou dissolução de minerais, podem reduzir a conectividade dos poros, resultando em uma menor permeabilidade e, consequentemente, em uma maior FBHP (52). À medida que o sedimento é enterrado, a compactação e a cimentação reduzem o espaço dos poros, levando à diminuição da porosidade e ao aumento da pressão de formação. A relação entre diagênese e pressão de formação é crucial para entender o comportamento dos reservatórios de hidrocarbonetos. Na avaliação da pressão de formação, vários dados petrofísicos de registro de poço, como densidade, nêutrons, sônicos, raios gama e registros de resistividade, são usados para avaliar e prever indiretamente a pressão dos poros. Essa abordagem envolve comparar as propriedades da formação subterrânea com as propriedades de pressão normal para identificar formações sobrepressurizadas, que exibem características distintas, como porosidades mais altas, densidades de massa mais baixas e velocidades mais baixas (55).

Além disso, as características diagenéticas e a evolução das bacias sedimentares desempenham um papel vital no controle da qualidade do reservatório. Os processos diagenéticos em bacias sedimentares são influenciados por fatores como transferência de massa e sistemas geoquímicos. Entender o tempo e os processos de diagênese é essencial para avaliar a qualidade do reservatório e prever a porosidade em reservatórios (56). O impacto da diagênese na redução da porosidade ressalta ainda mais a importância de estudar a relação entre diagênese e pressão de formação em reservatórios de hidrocarbonetos.

O comportamento da FBHP em sistemas de escoamento multifásico é inerentemente não linear, devido à essa complexa interação entre as fases de fluido, a permeabilidade do reservatório e as condições operacionais, como a elevação artificial.

1.3 ORIGENS DO GMDH

A teoria do *Group Method of Data Handling* (GMDH), tem como base a teoria neuronal do perceptron e é baseada no princípio de auto-organização. Foi formalizada por Ivakhnenko (4), portanto ainda antes dos trabalho se Geoffrey Hinton sobre MLP (*Multi-Layer Perceptron*) e *back-propagation* (5). Posteriormente, em 1994, em seu livro *Inductive Learning Algorithms for Complex Systems Modeling* (3), Ivaknenko já apresenta algorimos multi-camadas, multi-camadas com aspectos combinatórios e diversos outros algoritmos que viriam a ser a base conceitual das várias implementações da família de

algoritmos baseados em GMDH. Por conta de sua forte natureza algébrica, os algoritmos GMDH são também conhecidos por Redes Neurais Polinomiais. Na Seção 2.1 são expostos detalhes do GMDH.

1.4 MODELOS INTERPRETÁVEIS

Um modelo interpretável é um modelo de ML (*Machine Learning*) que pode ser entendido por humanos em termos de suas previsões e decisões. Isso significa que o modelo permite que os usuários compreendam e sigam o raciocínio por trás de suas previsões. Modelos como regressões lineares e árvores de decisão são considerados interpretáveis porque suas estruturas e processos de tomada de decisão são transparentes e compreensíveis. Em contraste, modelos mais complexos, como redes neurais profundas e o próprio GMDH, são frequentemente vistos como "caixas-pretas"devido à sua complexidade, dificultando a interpretação direta de suas operações internas (19).

Tais modelos, também denominados IA Explicável (*Explainable AI*, ou *XAI*) são em geral produzidos por Programação Simbólica, uma subárea da Computação Simbólica, envolvendo o uso de métodos para encontrar expressões matemáticas que melhor se ajustam a um conjunto de dados. A técnica mais comum é a regressão simbólica, que busca identificar a relação matemática subjacente entre variáveis de entrada e saída. Pode ou não, utilizar algoritmos evolutivos, como a programação genética, para gerar modelos matemáticos interpretáveis. (20).

1.5 REVISÃO BIBLIOGRÁFICA

A utilização da técnica de GMDH tem sido exitosa na predição de vários indicadores na área petrolífera. Algumas dessas aplicações podem ser conferidas na Tabela 1.1.

Dentre as aplicações mais recentes, Mulashani (26), Gao (27) e Guo (28), tratam da permeabilidade, pressão do poro e produção respectivamente. Mulashani (26) aborda o desafio de prever a permeabilidade de reservatórios de petróleo, uma variável crucial para a caracterização de reservatórios e estimativa do fluxo e volume de hidrocarbonetos. A permeabilidade é geralmente avaliada a partir de amostras de rochas de laboratório ou informações de testes de poços, mas também pode ser estimada indiretamente usando dados de registros de poços assim como a FBHP no presente trabalho. No trabalho de Mulashani foi feita uma comparação entre o GMDH-LM e o BPNN (*Backpropagation Neural Network*). O GMDH-LM, uma aplicação de GMDH desenvolvido com base no algoritmo de otimização de Levenberg-Marquardt, utilizado por Mulashani, mostrou melhor desempenho em termos de precisão e eficiência de tempo comparado ao BPNN. Foram obtivos um RMSE de 0,679 e MAE de 0,056. Gao (27) estudou a pressão de poro (PP), em formações subterrâneas, um parâmetro crucial na indústria de petróleo e gás. A PP é fundamental em várias etapas da exploração e produção de petróleo, incluindo perfuração, planejamento de reservatórios,

Tabela 1.1 – Aplicações de GMDH na área de petróleo e gás.

Referência(s)	**Aplicação**	**Notas**
Alvin (26)	Predição da permeabilidade do reservatório	A rede neural GMDH foi usada para prever com sucesso a permeabilidade dos registros do poço
Mesbah (29)	Predição da temperatura de formação de hidrato	O GMDH híbrido foi usado na previsão da temperatura de formação de hidrato de uma ampla gama de misturas de gás natural, incluindo gás doce e ácido.
Guo (28)	Previsão da produção de campos petrolíferos em séries temporais	GMDH modificado usado para prever séries temporais de produção de campos petrolíferos usando parâmetros de reservatório
Mahdavi-Meymand (30)	Previsão do fluxo de ar	Um novo GMDH integrado foi desenvolvido e aplicado para estimar a demanda de ar no vertedouro utilizando parâmetros de vazão.
Amar (31)	Modelagem e previsão da capacidade de adsorção de metano em gás de xisto	O GMDH foi utilizado para fornecer expressões matemáticas explícitas precisas e confiáveis para prever a adsorção de metano.
Mahdaviara (32)	Previsão de permeabilidade de formação	O GMDH foi utilizado para estimar a permeabilidade de formação do reservatório heterogêneo de óleo carbonatado a partir de propriedades petrofísicas.
Youcefi (33)	Previsão de pressão do tubo vertical (SPP)	O GMDH estendido foi usado para prever a pressão do tubo vertical (SPP) em tempo real como uma função do fluxo de lama a partir dos parâmetros de perfuração.

Fonte: Elaborada pelo autor (2024).

e prevenção de acidentes como explosões (*blowouts*). Essa pressão se refere à pressão exercida pelos fluidos dentro dos poros de uma rocha reservatório. A medição direta da PP, realizada com ferramentas como o *Modular Formation Dynamics Tester* (MDT) ou o *Repeat Formation Tester* (RFT), é cara e demorada, limitando a determinação precisa da PP em diferentes profundidades. O estudo utilizou dados petrofísicos de três poços em um campo petrolífero no Oriente Médio. O GMDH foi treinado com dados de dois poços e validado/testado com dados do terceiro poço. O modelo apresentou alta precisão, com um erro médio (RMSE) de 1,83 Psi e um coeficiente de determinação (R^2) de 0,9994. O artigo apresenta um modelo híbrido de previsão de produção de campos de petróleo, combinando o GMDH modificado com BPNN. O objetivo do estudo é melhorar a precisão na previsão da produção de petróleo ao explorar as vantagens de ambos os métodos: o

GMDH modificado para a seleção eficaz de variáveis e a BPNN para o mapeamento não linear. E Guo, (28) apresenta um interessante exemplo de um modelo testado usando dados de séries temporais da produção de um campo de petróleo, e seu desempenho foi comparado com outros modelos, incluindo regressão linear múltipla (MLR), GMDH tradicional, GMDH modificado, BPNN, e uma combinação de GMDH tradicional com BPNN (GMDH-BP). Os resultados indicaram que o modelo híbrido GMDH-BP modificado apresentou maior precisão na previsão da produção de petróleo em comparação com os outros modelos testados, R = 0,9986, RMSE = 13,7979 e MAPE = 2,9889.

Já para a previsão especificamente de FBHP, em 2005, Osman (25) conseguiu com modelos de ANN (*Artificial Neural Networks*) um Coeficiente de Determinação (R^2) de 0,97, um RMSE (Erro Quadrático Médio Relativo) de 2,80, com um Desvio Padrão de 66,24. O modelo desenvolvido usa redes *feedforward* com múltiplas camadas, sendo a camada de entrada composta por nove neurônios que representam variáveis como taxa de produção de óleo, água e gás, diâmetro e comprimento do tubo, pressão na cabeça do poço, gravidade do óleo (API), temperatura de superfície e de fundo de poço. Após otimizações, a arquitetura final do modelo inclui três camadas ocultas e uma camada de saída que prevê a pressão de fundo de poço. Ahmad Al-Shammari (1) (2011), em sua tese utilizou um modelo ANFIS (*Adaptive Neuro Fuzzy Inference System*) atingindo um R^2 de 0,93, MAE (Erro Médio Absoluto) de 4,93%, RMSE de 6,03% e Desvio Padrão dos Erros de 5,87%. foi treinado usando 596 conjuntos de dados de testes de poços também do Oriente Médio e posteriormente testado com 199 conjuntos de dados. Tariq (22) (2020), desenvolveu um modelo baseado em PSO-ANN (*Particle swarm optimization – Artificial Neural Network*) atingindo um R^2 de 0,98. O estudo utiliza dados de produção na superfície para treinar o modelo. Testes de análise de tendência mostram que o modelo captura com precisão a física subjacente ao problema. A metodologia inclui a aquisição e pré-processamento de dados, a aplicação do modelo ANN e a otimização dos parâmetros do modelo com PSO. Em uma publicação no ano de 2023, Goliatt et al. (23) (2023) desenvolveu um modelo híbrido integrando um algoritmo de seleção de característica com experimentos em quatro algoritmos de ML (*Machine Learning*). O melhor resultado, com o algoritmo MARS (*Multivariate Adaptive Regression Splines*) alcançou um R^2 de 0,88, Coeficiente de Correlação (R) = 0,94, RMSE = 97,88, MAE (*Mean Absolute Error*) de 74,69 e Erro Percentual Médio Absoluto (MAPE) = 3,12%.

Notas-se que estudos com boa assertividade de previsão FBHP tem sido conduzidos nos últimos tempos. Tais estudos têm utilizado diversas técnicas de Inteligência Artificial. Mas a utilização da técnica de GMDH, especificamente para previsão de FBHP são raros. Estudo com técnica de GMDH e com a geração de modelos simbólicos é uma inovação do presente trabalho.

Um trabalho inovador na utilização de programação simbólica na previsão de FBHP, foi conduzido por Campos (8) em que foi utilizada a Programação Genética Simbólica

(SGP) para desenvolver modelos preditivos de FBHP. O estudo alcançou um R^2 de 0,756, mostrando que a programação simbólica pode ser uma abordagem eficaz para previsões de FBHP, embora com uma precisão ligeiramente inferior em comparação com as técnicas baseadas em redes neurais, mas com bons índices de RMSE de 0,1963 e MAPE de 6,334.

Em um dos raros trabalhos utilizando GMDH, para previsão de FBHP, Ayoub (2014) (24) atingiu um índice notável de R^2 de 0,92 e um RMSE de 5,86 com um Desvio Padrão de 3,8. Nesse trabalho, Ayoub afirma que a abordagem GMDH foi utilizada para desenvolver, pela primeira vez, um modelo para estimar com sucesso a pressão em poços verticais.

Tabela 1.2 – Resumo do levantamento dos trabalhos relacionados.

Autor	**Ano**	**Técnica Utilizada**	**Resultados**
Osman (25)	2005	ANN	R^2 = 0,97 RMSE 2,80 Desvio Padrão de 66,24
Shammari (1)	2011	ANFIS	R^2 de 0,93 MAE 4,93% RMSE de 6,03% e Desvio Padrão dos Erros de 5,87%
Ayoub (24)	2014	GMDH	R^2 de 0,92 RMSE de 5,86 com um Desvio Padrão de 3,8
Tariq (22)	2020	PSO-ANN	R^2 de 0,98
Goliatt (23)	2023	MARS	R^2 de 0,88 R = 0,94 RMSE = 97,88 MAE de 74,69 MAPE = 3,12%
Campos (8)	2024	SGP	R^2 de 0,756 RMSE de 0,1963 e MAPE de 6,334

1.6 OBJETIVO

Esta dissertação tem como objetivo principal desenvolver e avaliar modelos preditivos para a Pressão de Fundo do Poço (FBHP - *Flowing Bottom Hole Pressure*) utilizando Redes Neurais Polinomiais, mais especificamente a GMDH (*Group Method of Data Handling*). A pesquisa busca contribuir com mais um método preditivo de bom nível de assertividade que gere também uma representação simbólica interpretável do modelo. Essas duas características (boa assertividade e geração de modelo interpretável) pretendem ser o ponto focal do trabalho. Propõe-se também disponibilizar um estudo sintetizado sobre quatro tipos de algoritmos GMDH como não encontrado em nenhuma bibliografia estudada para essa dissertação.

1.6.1 Objetivos específicos

- Desenvolver Modelos Preditivos Baseados em GMDH: utilizar a técnica de GMDH criar modelos preditivos que possam capturar as relações complexas entre as variáveis

envolvidas na determinação da FBHP.

- Implementar e Treinar Redes Neurais Polinomiais: implementar redes neurais polinomiais como parte da técnica GMDH, ajustando e treinando esses modelos com conjuntos de dados reais provenientes de operações de petróleo e gás.
- Validar e Avaliar os Modelos: realizar validações cruzadas rigorosas e testes empíricos para avaliar a precisão, robustez e generalização dos modelos desenvolvidos, comparando-os com modelos preditivos tradicionais.
- Analisar o Impacto de Variáveis Críticas: investigar como diferentes variáveis, como profundidade do poço, densidade do óleo (API), e condições operacionais, influenciam a FBHP e a precisão das previsões.
- Publicar os Resultados: disseminar os achados da pesquisa por meio de publicações em revistas científicas, contribuindo para o avanço do conhecimento na área de modelagem preditiva e inteligência artificial aplicada à indústria de petróleo e gás.

1.7 VISÃO GERAL DA DISSERTAÇÃO

No Capítulo 1, estabelece-se as diretrizes do trabalho bem como expondo para o leitor requisitos e ferramentais básicos para acompanhar o restante do texto. No Capítulo 2, é feita uma dissertação sobre o método GMDH, situando entre as RNP e discorrendo sobre os quatro algoritmos que serão alvo de implementação no capítulo seguinte. A Metodologia do trabalho é apresentada no Capítulo 3, onde se encontra detalhado o experimento objeto da dissertação. O Capítulo 4 apresenta os resultados do trabalho ressaltando os objetivos atingidos com a análise dos erros e medidas de desempenho. O trabalho se encerra no Capítulo 5 com a conclusão e sugestão de trabalhos futuros.

2 REDES NEURAIS POLINOMIAIS

Uma Rede Neural Polinomial (RNP) é um tipo de Rede Neural Artificial (RNA) onde a função de ativação é sempre uma função polinomial em vez das funções de ativação não lineares tradicionais. Ela utiliza polinômios de múltiplas variáveis para modelar as relações entre as entradas e as saídas (3). Seu treinamento é feito através da seleção de termos polinomiais que melhor explicam os dados, o que pode incluir uma busca heurística para determinar a estrutura da rede. Sua estrutura é construída de forma incremental, selecionando os melhores neurônios (as funções polinomiais) com base em critérios de desempenho, como minimização do erro. Pode resultar em modelos mais interpretáveis, que é o caso deste trabalho, pois a saída de cada neurônio é uma expressão polinomial explícita. As RNP possuem aplicações semelhantes às das RNAs convencionais, mas especialmente úteis em problemas onde a modelagem explícita de relações polinomiais é vantajosa.

As principais diferenças entre uma RNP e uma RNA estão nas funções de ativação, no processo de aprendizado, na estrutura, na complexidade e nas aplicações, conforme exposto na Tabela 2.1.

Tabela 2.1 – Comparação Resumida entre RNA e RNP

Característica	**RNA Convencional**	**Rede Neural Polinomial (RNP)**
Funções de Ativação	Não lineares (sigmoide, tanh, ReLu, ELU, softmax)	Polinômios de múltiplas variáveis
Processo de Aprendizado	Algoritmos de otimização (retropropagação)	Seleção heurística de termos polinomiais
Estrutura	Flexível, múltiplas camadas e neurônios	Incremental, baseado em desempenho
Complexidade	Pode ser elevada com muitas camadas	Pode ser mais interpretável
Aplicações	Classificação, regressão, visão, NLP	Modelagem de relações explícitas, previsão

Fonte: Elaborada pelo autor (2024).

2.1 VISÃO GERAL DO MÉTODO

O método GMDH é técnica de aprendizado indutivo e baseado no princípio da auto-organização para a modelagem de sistemas complexos (3). Fundamentado na construção de redes neurais polinomiais, são utilizados algoritmos auto-organizáveis para selecionar, dentre um conjunto de candidatas, as funções polinomiais que melhor representam a relação entre variáveis de entrada e saída. Esse processo é iterativo e orientado por critérios de

desempenho, garantindo que a complexidade do modelo se ajuste de maneira otimizada ao problema específico (7). Matematicamente, o GMDH opera através da construção de polinômios de Volterra e Kolmogorov-Gabor (VKG), que são utilizados como funções base nas camadas da rede neural. Esse polinômio é uma série que expande a saída do sistema em termos das funções de entrada e é descrito por

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots \tag{2.1}$$

Onde:

- y é a variável de saída,
- $X(x_1, x_2, x_3, ..., x_n)$ é um vetor de variáveis de entrada,
- a_0 representa o componente de viés (bias),
- $A(a_i, a_{ij}, a_{ijk}, ...,)$ é o vetor dos coeficientes ou pesos em cada neurônio e
- n é o número de termos de entrada.

Os componentes do vetor de entrada X podem ser variáveis independentes, funções ou termos de diferenças finitas. Outras funções de referência não lineares, como diferença probabilística, harmônica, logística também podem ser usadas. O método permite encontrar simultaneamente a estrutura do modelo e a dependência da saída do sistema modelado com os valores das entradas mais significativas do sistema. Esses polinômios são combinados e testados em cada iteração do processo de modelagem, com a seleção dos melhores candidatos baseada em técnicas de validação cruzada.

A cada etapa, novos polinômios são gerados como combinações não-lineares das variáveis selecionadas anteriormente, permitindo a construção de modelos hierárquicos que capturam de forma precisa a dinâmica do sistema modelado. É um algoritmo que se mostra bastante eficiente e escalável. Ele emprega um procedimento de busca heurística para explorar o espaço de soluções possíveis, e utiliza algoritmos de otimização para ajustar os parâmetros dos polinômios.

A estrutura do modelo é determinada automaticamente, o que reduz a necessidade de intervenção manual e facilita a aplicação do método a diferentes tipos de dados e problemas. Além disso, o GMDH pode ser combinado com outras técnicas de *machine learning*, como algoritmos genéticos e redes neurais profundas, para aprimorar ainda mais sua capacidade de modelagem. Essa flexibilidade e eficiência computacional tornam o GMDH uma ferramenta valiosa para aplicações em áreas como previsão econômica, controle de processos, e análise de séries temporais. A figura 2.1 ilustra de forma macro a dinâmica de uma rede GMDH.

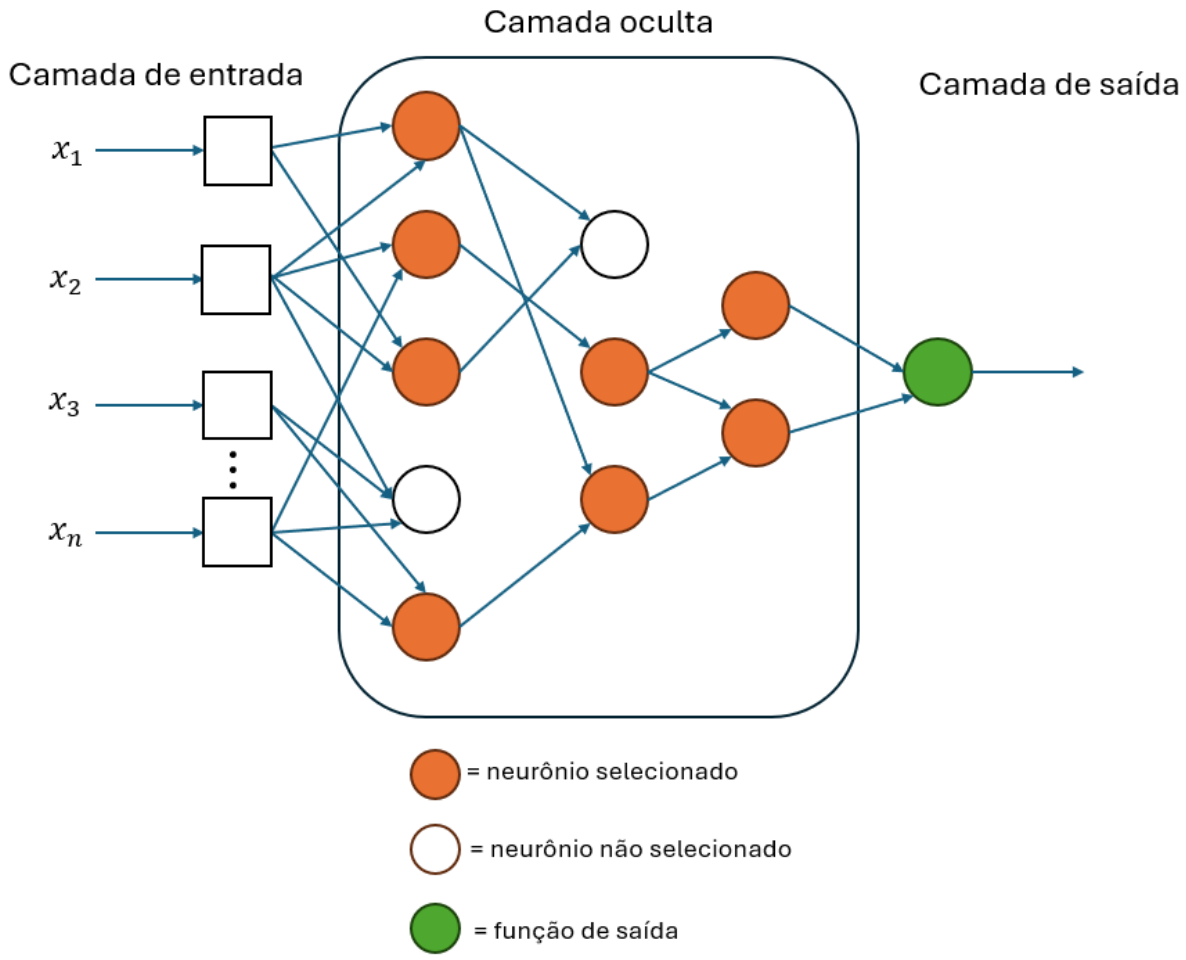

Figura 2.1 – Rede Neural Polinomial GMDH

Fonte: Elaborada pelo autor (2024).

De forma geral, GMDH se refere a uma família de algoritmos que implementam redes neurais polinomiais. Esse trabalho se concentra na exploração de dois dos principais algoritmos dessa família: o modelo COMBI(*Combinatorial Algorithm*) e o modelo MIA (*Multilayer Iterative Algorithm*) e suas principais variações: o RIA(*Recursive Iterative Algorithm*) e o MULTI(*Multilayer Combinatorial Algorithm*) que serão vistos nas seções seguintes.

2.2 O MODELO COMBI

O algoritmo COMBI é o algoritmo GMDH básico e é iniciado com um conjunto de dados X e uma variável de saída y. Os dados são divididos em treinamento e teste. Com esses dados são gerados os primeiros polinômios VKG. Usando a parte de treinamento, os coeficientes (pesos) a_i são estimados por métodos como o dos mínimos quadrados. Cada modelo é avaliado com base em seu desempenho no conjunto de teste usando critérios como regularidade, regularidade simétrica ou estabilidade estruturando os modelos nesse momento. Os melhores modelos são selecionados e usados para formar a próxima camada

de modelos. O processo de geração, avaliação e seleção de modelos continua até que um critério de parada seja atingido (por exemplo, erro mínimo). Este processo iterativo e exaustivo garante que o COMBI encontre a melhor estrutura de modelo polinomial possível para os dados fornecidos (3) e (6). O algoritmo segue os seguintes passos:

1. **Dados de Entrada**: O algoritmo começa com um conjunto de dados $\mathbf{X}$ e uma variável de saída y.

2. **Divisão da Amostra**: Os dados são divididos em duas partes: treinamento $(\mathbf{X}_{train}, y_{train})$ e teste $(\mathbf{X}_{test}, y_{test})$.

3. **Geração de Modelos**:

   - Modelos polinomiais são gerados usando combinações das variáveis de entrada $\mathbf{X}$.
   - Cada modelo polinomial é expresso como polinômios VKG:

$$y = a_0 + \sum_i a_i x_i + \sum_i \sum_j a_{ij} x_i x_j + \cdots$$

4. **Estimativa de Coeficientes**: Usando a parte de treinamento, os coeficientes a_i são estimados por métodos como mínimos quadrados.

$$\text{Erro} = \frac{1}{N} \sum_{i=1}^{N} (y_{test} - \hat{y})^2$$

   Onde $\hat{y}$ é a predição do modelo.

5. **Avaliação de Modelos**: Cada modelo é avaliado com base em seu desempenho no conjunto de teste utilizando vários critérios. Nesse trabalho, não só o algoritmo COMBI, como também os outros algoritmos, utilizaram critérios baseados na Tabela 3.3 de hiperparâmetros.

6. **Seleção de Modelos**: Os melhores modelos são selecionados e usados para formar a próxima camada de modelos.

7. **Iteração**: O processo de geração, avaliação e seleção de modelos continua até que um critério de parada seja atingido (por exemplo, erro mínimo).

A rede COMBI é ilustrada na Figura 2.2.

No combi cada perceptron é uma combinação de todas variáveis com os coeficientes (determinados via LSM). Então cada modelo é uma combinação. Como exemplo, de acordo com Figura 2.2 a combinação da variável x_1 com seu peso w_1, da variável x_2 com seu peso w_2 e assim por diante atá a variável x_n com seu peso w_n. Em seguida teremos as

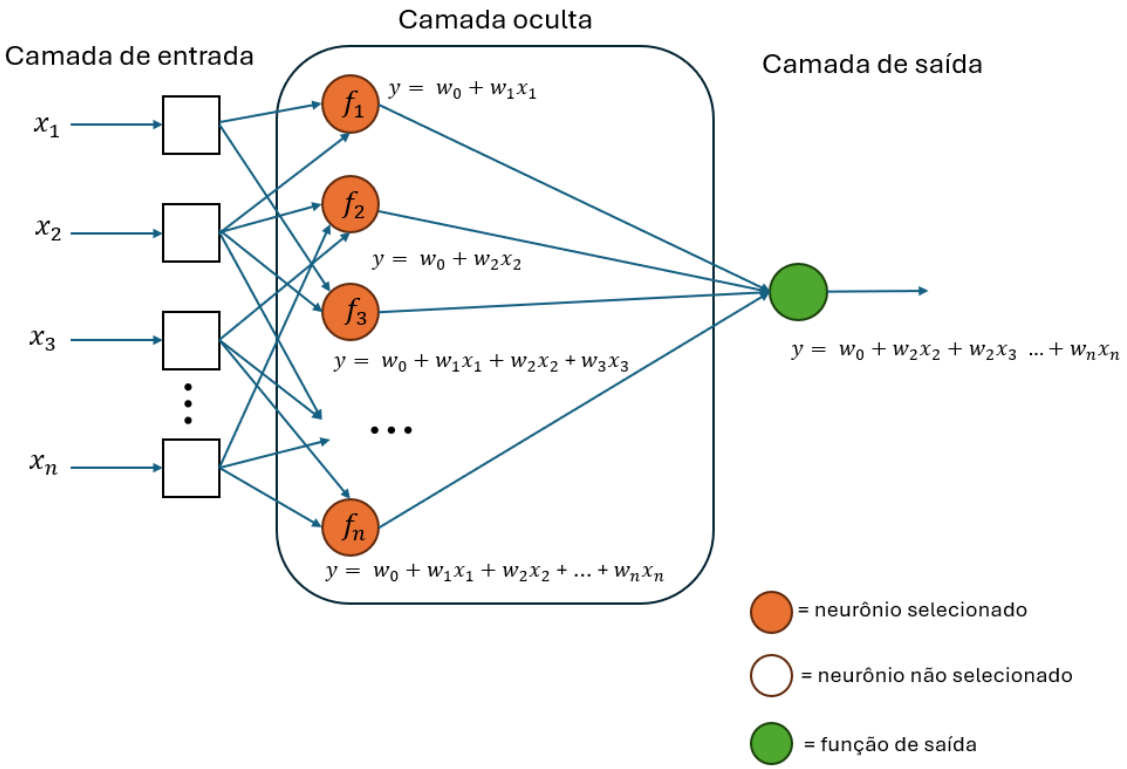

Figura 2.2 – Rede neural COMBI

Fonte: Elaborada pelo autor (2024).

combinações das variáveis x_1 com x_2, x_2 com x_3, até x_2 com x_n. Em seguida x_1 com x_2 e x_3, assim por diante. O algoritmo faz todas as combinações possíveis. Todas as combinações são avaliadas e o algoritmo elege a melhor. Isso nos dá $2^m - 1$ combinações onde m é o número de variáveis. No caso do presente estudo com 8 variáveis teremos 255 combinações.

2.3 O MODELO MULTI

O Multi é uma variação com COMBI. É um combinatório *multilaeyer*. Ele não implementa todas as possíveis combinações. Ele obtem as melhores combinações em uma camada, e combina somente essas melhores com todas as outras da camada anterior. Temos então as funções SELECIONADAS de uma camada combinadas com TODAS as funções da camada anterior. Sempre um par de (*selecionada*, *selecionada ou não*). Por exemplo: supondo um domínio de 4 variáves de x_1 a x_4, e supondo que na primeira camada as melhores funções sejam as funções com x_2 e x_3, ou seja $y = f(x_2)$ e $y = f(x_3)$. Na segunda camada teremos as seguintes combinações: $y = f(x_2, x_1)$, $y = f(x_2, x_3)$, $y = f(x_2, x_4)$ e $y = f(x_3, x_1)$ e $y = f(x_3, x_4)$. Não haverá por exemplo a combinação $f(x_1, x_4)$, pois a função com a variável x_1 não foi selecionada na primeira camada.

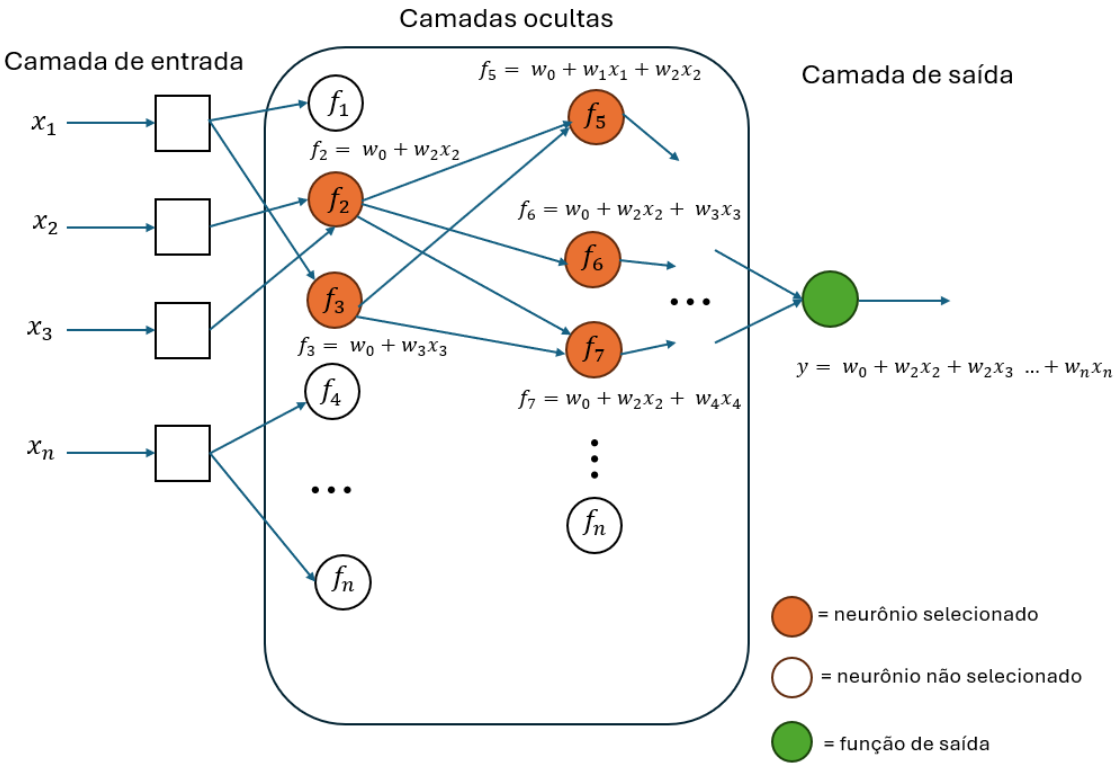

Figura 2.3 – Rede neural MULTI

Fonte: Elaborada pelo autor (2024).

2.4 O MODELO MIA

Assim como o modelo COMBI, o modelo MIA também executa as fases de treinamento, teste, geração dos polinômios VKG, estimativa dos coeficientes a_i e avaliação de seu desempenho. A diferença aqui é que há um processo de iteração que gera novas camadas de modelos usando as melhores combinações das camadas anteriores, continuando até que um critério de parada seja atingido, como o erro mínimo. A cada combinação os elementos combinados são polinômios $f(x_i, x_j)$. Esses polinômios podem ser não lineares. O tipo de polinômio pode ser especificado usando um hiper-parâmetro informando o tipo de polinômio a ser usado. No MIA duas variáveis são passadas para cada unidade, compondo as funções de cada camada. Cada neurônio desse que é ativado para a próxima camada irá ser um parâmetro para as funções da próxima camada. Então a partir daí cada camada é feita por funções que são combinações 2 a 2 de funções ativadas na camada anterior. Supondo que as melhores combinações na primeira camada sejam $y = f_1(x_1, x_2)$, $y = f_2(x_1, x_3)$ e $y = f_3(x_2, x_4)$. A segunda camada irá considerar as possíveis combinações: $y = f_4(f_1, f_2)$, $y = f_5(f_1, f_3)$ e $y = f_6(f_2, f_3)$. Se temos m variáveis de entrada então a primeira camada gerará C_m^2 funções.

A Figura 2.4 ilustra uma rede GMDH MIA.

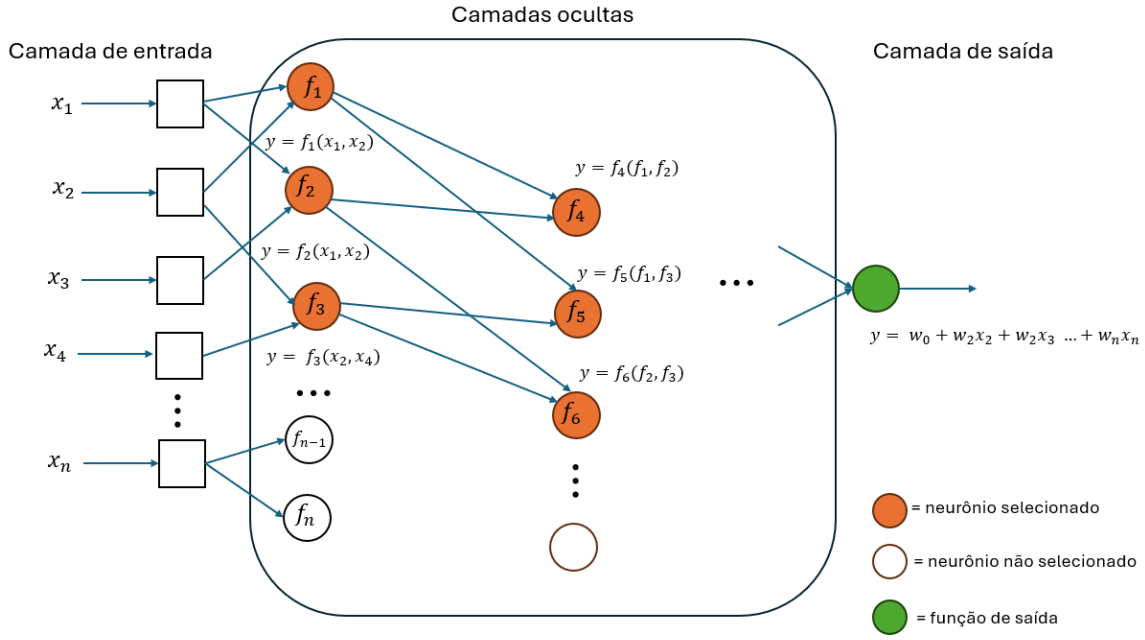

Figura 2.4 – Rede neural MIA

Fonte: Elaborada pelo autor (2024).

2.5 O MODELO RIA

O modelo RIA é uma variação do modelo MIA, implementando uma fase de iteração recursiva. Ele gera novas camadas de modelos utilizando os melhores modelos das camadas anteriores de forma recursiva, continuando até que um critério de parada seja atingido, como o erro mínimo (3) e (6). A rede RIA é ilustrada na Figura 2.5. É uma variação do modelo MIA. Duas variáveis são passadas para cada unidade, compondo as funções da primeira camada. Cada neurônio (função) desse que é ativado irá ser um parâmetro para as funções da próxima camada, juntamente com uma variável. Então a partir daí cada camada é feita por funções que são combinações 2 a 2 de **funções** ativadas na camada anterior e **variáveis** utilizadas na camada anterior. Novamente, supondo que as melhores combinações na primeira camada sejam $y = f_1(x_1, x_2)$ e $y = f2(x_1, x_3)$, na segunda camada teremos como combinações do RIA: $y = f_3(f_1, x_1)$, $y = f_4(f_1, x_2)$, $y = f_5(f_1, x_3)$, $y = f_6(f_2, x_1)$, $y = f_7(f_2, x_2)$, $y = f_8(f_2, x_3)$.

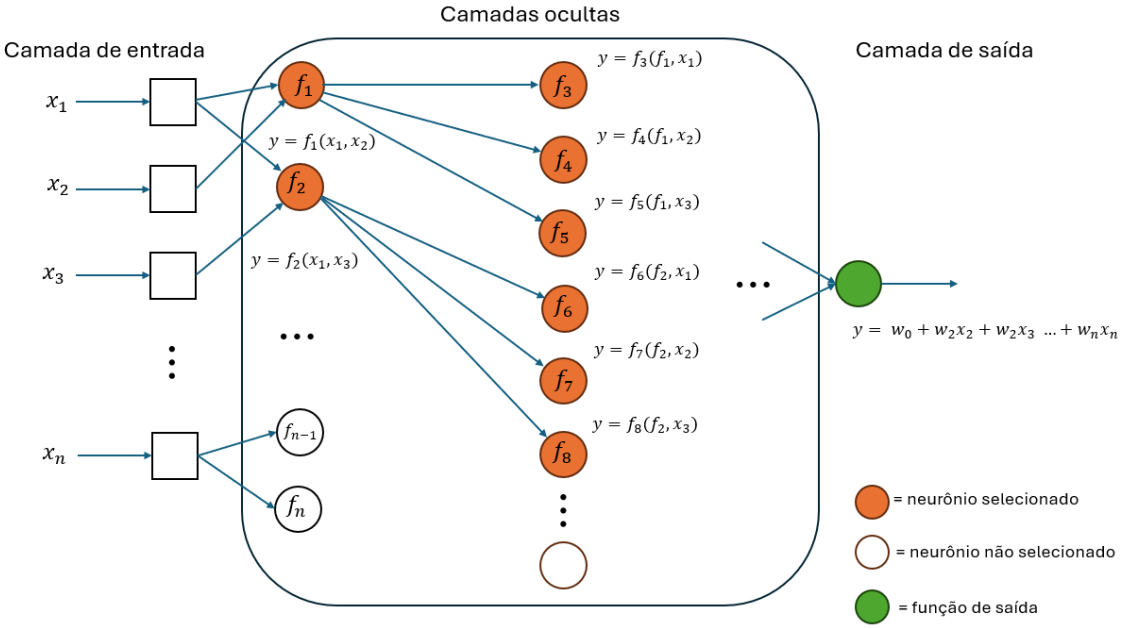

Figura 2.5 – Rede neural RIA

Fonte: Elaborada pelo autor (2024).

3 METODOLOGIA

Neste capítulo apresenta-se a metodologia utilizada na pesquisa, detalhando os aspectos relativos à obtenção, tratamento e processamento dos dados, as métricas de avaliação utilizadas, os recursos computacionais com os quais foram executados os modelos e detalhada também a aplicação de quatro algoritmos da família GMDH.

3.1 BASE DE DADOS

Os dados representam a base dos trabalhos relacionados a Inteligência Artificial, mais especificamente nas áreas de Aprendizado de Máquina e Ciência de Dados. Especificamente neste trabalho a base dados utilizadas já estava razoavelmente limpa e exigiu pouco trabalho de preparação e formatação dos dados para execução dos algoritmos. A base de dados se refere a 795 amostras de medições de pressão de fundo de poço em vários campos do Oriente Médio. Essa base é detalhada por Al-Shammari (1). Além da variável de saída FBHP, Pressão de Fundo de Poço (*Flowing Bottom-hole Pressure*) a base apresenta outras sete variáveis conforme lista abaixo:

- Pressão na cabeça do poço (*Wellhead Pressure* - WHP)
- Taxa de fluxo de água (*Water Flow Rate* - WFR)
- Taxa de fluxo de óleo (*Oil Flow Rate* - OFR)
- Taxa de fluxo de gás (*Gas Flow Rate* - GFR),
- Produção diária de água (*Water Production Rate* - WPD)
- Gravidade API (*American Petroleum Institute Gravity* - API)
- Diâmetro interno do tubo (*Internal Diameter of Pipe* - ID)
- Temperatura na cabeça do poço (*Wellbore Head Temperature* - WBHT)

A Figura 3.1 exibe os coeficientes de correlação entre as variáveis de entrada e a pressão de fundo de poço (FBHP). Os coeficientes variam entre +1 e -1, onde +1 representa uma correlação direta entre as variáveis e -1 uma relação indireta entre as duas variáveis. Três variáveis apresentam correlação importante (acima de 0,4) com a medição de FBHP: a pressão na cabeça do poço (WHP) com 0,53, a produção diária de água (WPD) com 0,47 e a gravidade API (API) com 0,42.

Em relação ao WHP é natural que haja relevância na relação com FBHP, pois é esperado que a perda seja a menor possível entre a pressão no fundo e na cabeça do poço. Ao passo que a variação na pressão no fundo também impacte diretamente na

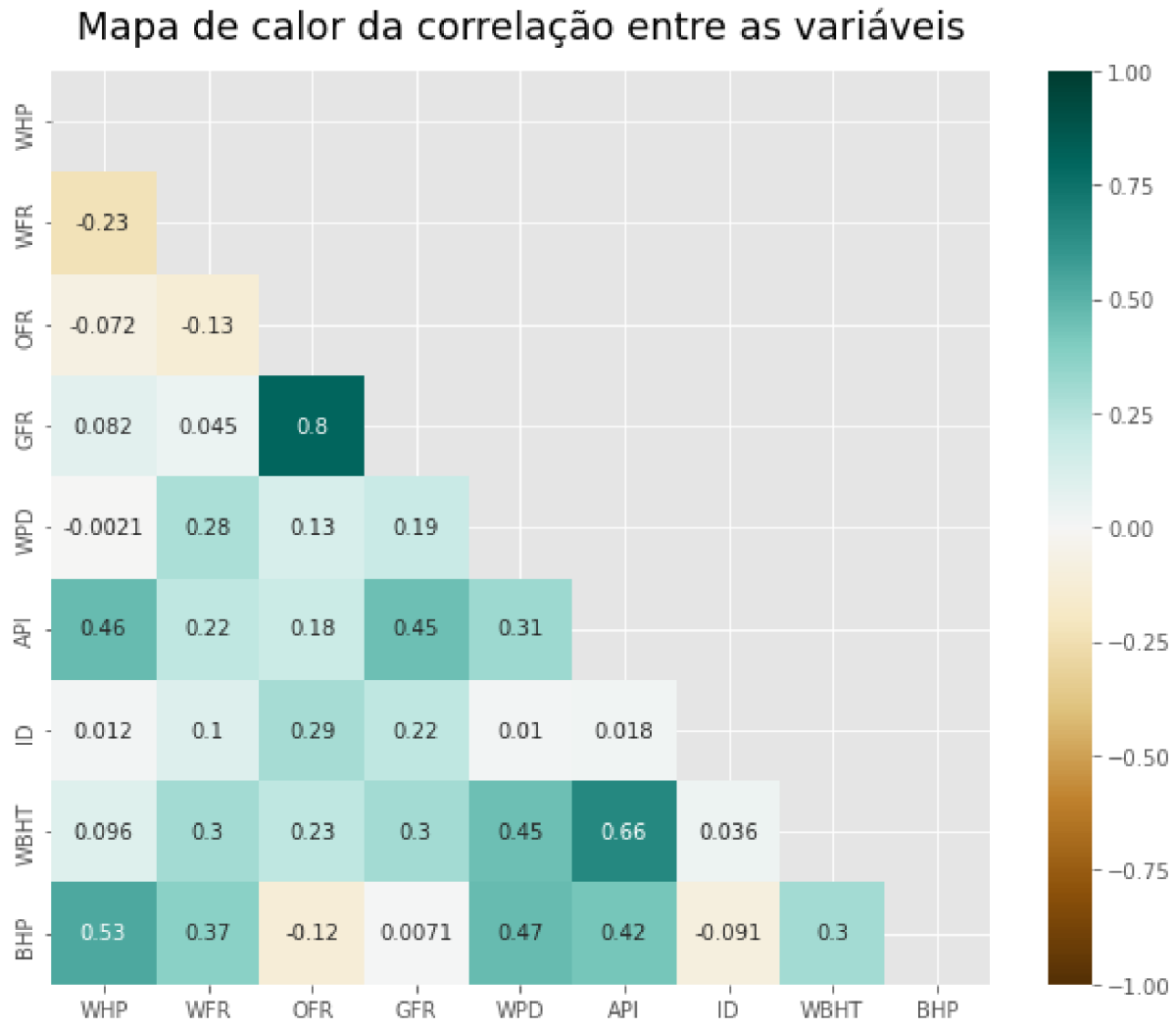

Figura 3.1 – Mapa de calor da correlação entre as variáveis

Fonte: Elaborada pelo autor (2024).

pressão na cabeça do poço, mesmo levando em consideração as perdas com atrito ou temperatura, por exemplo. Da mesma forma, a produção diária de água (WPD) apresenta uma relação positiva com o FBHP, explicada pela necessidade proporcional de água para elevação artificial dos fluidos, levando a uma pressão maior no fundo do poço. A correlação positiva entre FBHP e gravidade API, dada pela Fórmula 3.1 precisa ser melhor explorada, considerando que, em geral, óleos com maior API (menos densos) fluem mais facilmente e exigem menos pressão para serem produzidos. Pode haver outros fatores contextuais ou específicos do campo de petróleo que precisam ser considerados para explicar essa correlação.

$$\text{API} = \frac{141,5}{\gamma} - 131,5 \tag{3.1}$$

onde γ representa a densidade do fluido em relação à água.

Tabela 3.1 – Faixas de dados coletados de parâmetros de entrada e saída

		Mínimo	**Máximo**	**Média**	**Desvio Padrão**
Entrada	WHP (psi)	92	1550	423,82	253,74
	WFR (bpd)	0	11395	2215,13	2294,8
	OFR (bdp)	176	17663	2215,13	3722,13
	GFR (scf/stb)	9	17859	2699,15	2370,41
	WPD (L/dia)	4243	8620	6326,89	511,37
	API	25,4	47,5	33,86	3,11
	ID (in)	1,995	6,276	3,95	0,57
	WBHT (ºF)	160	233	210,25	18,26
Saída	FBHP (psi)	1198	3698	2469,73	387,23

Fonte: Elaborada pelo autor (2024).

As variáveis estão distribuídas nas faixas de valores apresentadas conforme Tabela 3.1.

Em função do alto desvio padrão e diferenças de escala apresentadas pelas variáveis, foi utilizado um escalonamento utilizando a técnica de padronização (*standardization*).

Não havia dados faltantes e a quantidade de outliers não se mostrou relevante para variáveis com alta correlação com a variável alvo, conforme Tabela 3.2 e Figuras 3.2, 3.3 e 3.4.

Tabela 3.2 – Percentual de outliers por variável

Variável	Outliers(%)
WHP	5,03
WFR	3,27
OFR	2,76
GFR	2,89
WPD	6,91
API	2,51
ID	13,20
WBHT	12,95
FBHP	0,25

Fonte: Elaborada pelo autor (2024).

3.2 AVALIAÇÃO DO MODELO

O desempenho do modelo foi avaliado usando as métricas Coeficiente de Determinação (R^2), Erro Quadrático Médio (MSE), Erro Quadrático Médio Relativo (RMSE) e Erro Percentual Médio Absoluto (MAPE). A escolha se deu para garantir uma avaliação abrangente e robusta dos modelos de previsão de FBHP. Cada métrica fornece uma perspectiva diferente sobre o desempenho do modelo, permitindo uma compreensão completa

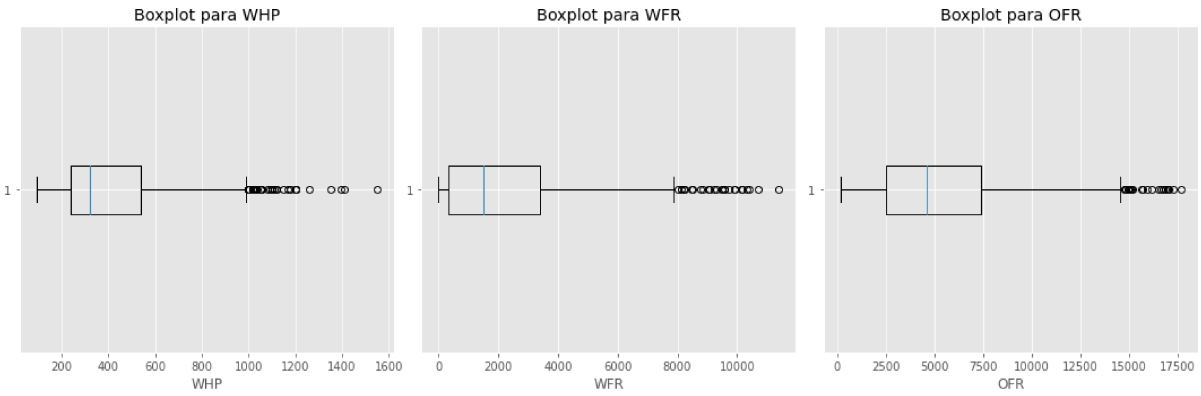

Figura 3.2 – Outliers das variáveis WHP, WFR e OFR

Fonte: Elaborada pelo autor (2024).

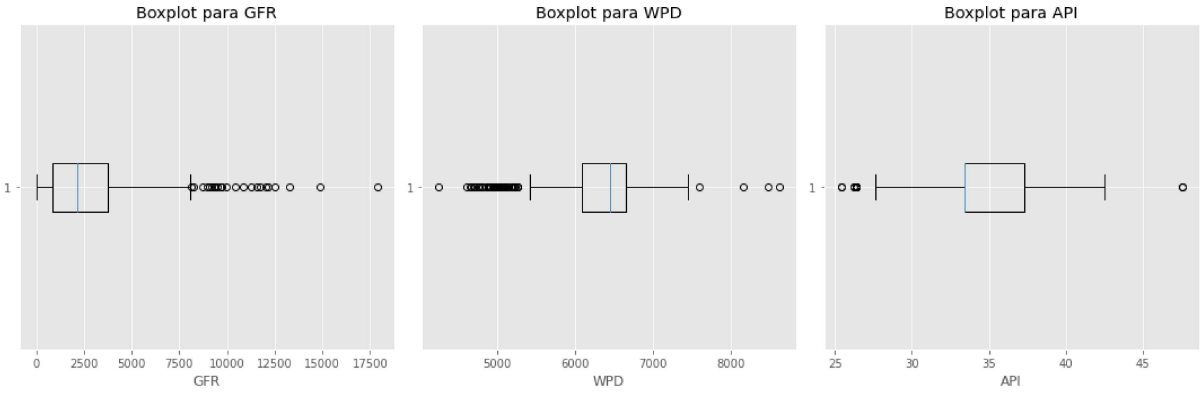

Figura 3.3 – Outliers das variáveis GFR, WPD e API

Fonte: Elaborada pelo autor (2024).

de suas capacidades e limitações. Essas métricas são fundamentais para garantir que os modelos desenvolvidos sejam precisos, confiáveis e aplicáveis em contextos operacionais críticos na indústria de petróleo e gás.

O Coeficiente de Determinação (R^2) indica o quão bem o modelo ajusta os dados (Fórmula 3.2). Um R^2 próximo de 1 indica um bom ajuste, enquanto um R^2 próximo de 0 indica que o modelo não explica bem a variabilidade dos dados. Ele é calculado como a soma dos quadrados das diferenças entre os valores observados e os valores previstos, normalizada pela soma dos quadrados das diferenças entre os valores observados e a média dos valores observados.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{3.2}$$

O Erro Quadrático Médio (MSE) (Fórmula 3.3) é uma métrica que calcula a média dos quadrados das diferenças entre os valores observados e os valores previstos. Mede a

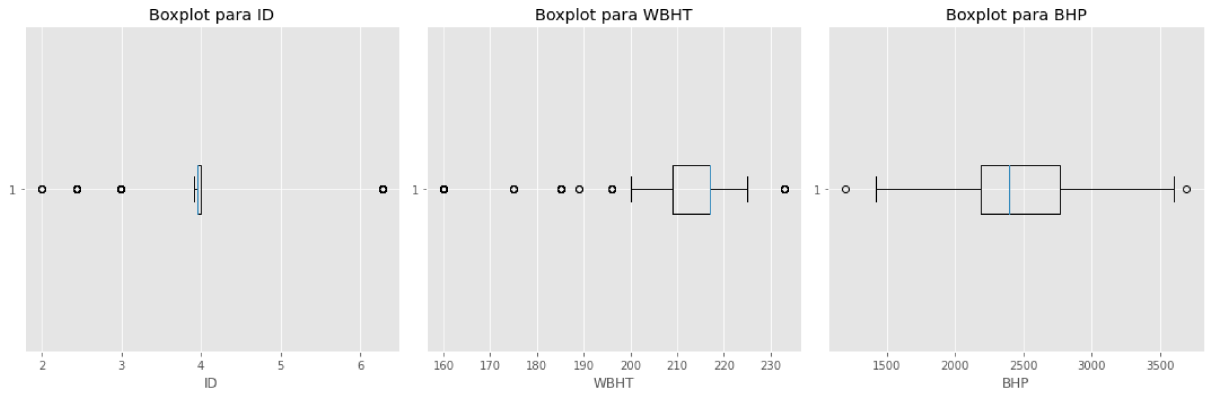

Figura 3.4 – Outliers das variáveis ID, WPHD e FBHP

Fonte: Elaborada pelo autor (2024).

qualidade do ajuste do modelo. Penaliza erros grandes de forma mais severa do que erros pequenos, devido ao termo quadrático.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3.3}$$

O Erro Quadrático Médio Relativo (RMSE) (Fórmula 3.4) fornece uma medida da magnitude média do erro de previsão. O RMSE está na mesma unidade dos valores previstos, tornando a interpretação mais intuitiva. Um RMSE menor indica que as previsões do modelo estão mais próximas dos valores reais.

$$RMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}}{\bar{y}} \tag{3.4}$$

O Erro Percentual Médio Absoluto (MAPE) mede a precisão das previsões em termos percentuais. É fácil de interpretar e útil para comparar o desempenho de previsões em diferentes escalas.

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}}{y_i}\right| \tag{3.5}$$

Onde N é o número total de elementos testados, y são os valores reais e $\hat{y}$ são os valores previstos e $\bar{y}$ é média dos valores observados.

3.3 RECURSOS COMPUTACIONAIS

Os experimentos realizadas nesta pesquisa foram feitos em um *laptop* com um processador Intel Core I7-1355U, 13^a geração, com 1.70 GHz de frequência, memória RAM de 16 GB DDR3, 2400 MHz. Os códigos foram gerados em linguagem de programação Python (Python versão 3.11). A execução dos algoritmos GMDH bem como a geração das

expressões simbólicas representando os polinômios, foram feitas utilizando a biblioteca *gmdh* (versão 1.0.3) Para as métricas de avaliação foi utilizada a biblioteca do módulo *scikit-learn* (versão 1.3).

3.4 EXPERIMENTO COMPUTACIONAL

Para o experimento, inicialmente os dados foram submetidos a uma divisão de treino e teste na razão de 70/30 (treinamento e teste). Com essa divisão, os dados foram submetidos a um algoritmo de *grid search* com todas as combinações possíveis de todos os hiperparâmetros disponíveis para cada modelo (que estão expostos na Tabela 3.3).

O hiperparâmetro *criterion* se refere a uma função utilizada para avaliar a qualidade dos modelos gerados durante o processo de modelagem. Serve como uma base para selecionar os melhores modelos em cada etapa do algoritmo, garantindo que a complexidade do modelo seja controlada enquanto se maximiza a precisão e a capacidade de generalização. São funções matemáticas ou heurísticas que avaliam a qualidade dos modelos gerados em cada iteração do algoritmo, selecionam os melhores polinômios ou redes que serão mantidos e usados para gerar novos modelos em iterações subsequentes. O valor *STABILITY* avalia a estabilidade do modelo em relação a pequenas perturbações nos dados de entrada. Um modelo estável é menos sensível a variações nos dados, o que significa que ele deve apresentar desempenho consistente, mesmo quando submetido a pequenas mudanças nas entradas. *SYM_STABILITY* é um *criterion* de estabilidade simétrica. Ele considera não apenas a estabilidade do modelo, mas também como o modelo se comporta de forma simétrica em relação a pequenas perturbações. Isso significa que ele avalia se o modelo mantém sua precisão de forma consistente quando os dados são perturbados simetricamente. *UNBIASED_OUTPUTS* verifica se as saídas do modelo não apresentam viés. Um modelo com saídas não tendenciosas deve produzir previsões cujas médias estejam alinhadas com a média dos valores reais observados. Em outras palavras, a diferença média entre as previsões e os valores reais deve ser próxima de zero, indicando que o modelo não está consistentemente superestimando ou subestimando os resultados. O valor *SYM_UNBIASED_OUTPUTS* é uma extensão do *criterion UNBIASED_OUTPUTS*, mas com um foco em verificar se as saídas são simetricamente não tendenciosas. Isso significa que o critério não apenas avalia se o modelo é tendencioso em termos gerais, mas também se essa tendência se mantém de forma simétrica em relação a pequenas variações nos dados de entrada. O objetivo é garantir que o modelo não favoreça sistematicamente uma direção (superestimação ou subestimação) quando os dados são ligeiramente alterados. Já o valor *UNBIASED_COEFFS* é um critério que avalia os coeficientes do modelo, verificando se eles são tendenciosos. Um modelo com coeficientes não tendenciosos deve apresentar coeficientes que reflitam corretamente as relações entre as variáveis de entrada e a variável de saída, sem inclinações sistemáticas em nenhuma direção. Este

critério é importante para garantir que o modelo represente fielmente as dependências nos dados. *ABSOLUTE_NOISE_IMMUNITY*: este critério avalia a capacidade do modelo de ser imune ao ruído absoluto nos dados. Um modelo com alta imunidade ao ruído absoluto deve ser capaz de manter sua precisão e estabilidade mesmo quando os dados de entrada contêm variações aleatórias significativas (ruído). Isso é crucial para garantir que o modelo seja robusto e confiável em condições de dados imperfeitos ou ruidosos. Finalmente, o valor *SYM_ABSOLUTE_NOISE_IMMUNITY* é semelhante ao critério *ABSOLUTE_NOISE_IMMUNITY*, este critério considera a imunidade ao ruído, mas com foco em como o modelo responde a ruídos simetricamente distribuídos. Ele verifica se o modelo pode resistir ao ruído de maneira uniforme em todas as direções, o que é importante para garantir que o modelo não se torne instável ou impreciso quando os dados contêm ruído que afeta as entradas de forma simétrica (3), (4), (6).

Esses critérios mencionados são essenciais no processo de seleção e otimização de modelos no algoritmo GMDH. Eles permitem que o algoritmo escolha os modelos mais adequados para prever os dados de maneira precisa, robusta e generalizável. Cada critério tem um foco específico, seja na estabilidade, na ausência de viés, ou na imunidade ao ruído, garantindo que o modelo final seja o mais adequado possível para os dados disponíveis.

Os hiperparâmetros *p_average*, *n_jobs* e *test_size* possuem a mesma faixa de valores para todos os algoritmos. *p_average* especifica o número máximo de combinações para o cálculo do erro médio aceitável em cada nível, foram testadas até 16 combinações. O hiperarâmetro *n_jobs* define o número de processos (*threads*) a serem usados no processamento do algoritmo. No trabalho foi configurado *n_jobs* = -1 indicando o uso máximo possível de processos. Por fim, o parâmetro *test_size* determina a proporção dos dados de entrada a serem incluídos no conjundo de testes, usado pra calcular os valores dos critérios. Foi usado o conjunto de valores $\{0,01; 0,25; 0,5; 0,75; 0,99\}$.

A quantidade das melhores combinações, em cada camada que irão ser combinadas com outras variáveis não utilizadas na combinação anterior é determinada pelo hiperparâmetro *k_best*.

O hiperparametro *polynomial_type* define o tipo de polinômio a ser usado para construir o polinômio final. Os algoritmos COMBI e MIA são lineares por padrão e não possuem esse parâmetros. Os outros modelos utilizam polinômios lineares, quadráticos ou lineares com termo de covariância, conforme o valores sejam *QUADRATIC*, *LINEAR* ou *LINEAR_COV* que são as equações lineares incluem variáveis covariáveis. Uma covariável é uma variável que pode influenciar ou estar associada à variável dependente, e seu efeito é considerado no modelo.

Após a obtenção da melhor combinação de hiperparâmetros para cada algoritmo (conforme métricas colocadas na Seção 3.2), os modelos foram novamente executados em k-folds e kk-folds nos valores de 5-folds, 5x5-fold e 10x5-folds, nas divisões de treino e

teste nas proporções de 60/40, 70/30 e 80/20.

Para cada execução, também foi feita a medição de tempo e extraída a expressão polinomial que representa o modelo de forma simbólica interpretável.

A Figura 3.5 ilustra os passos do experimento computacional.

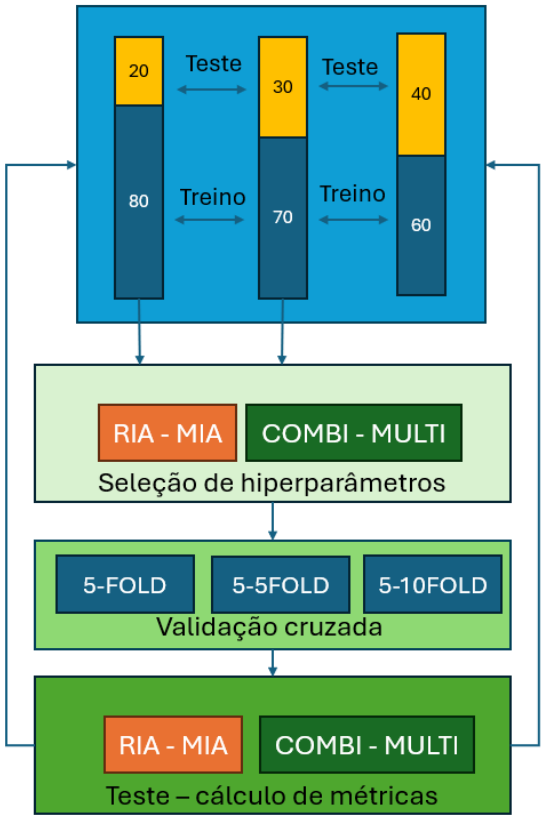

Figura 3.5 – Fluxo macro da execução do experimento

Fonte: Elaborada pelo autor (2024).

Tabela 3.3 – Tabela de Hiperparâmetros dos Algoritmos

	Hiperparâmetros					
Algoritmo	**criterion**	**k_best**	**p_average**	**n_jobs**	**test_size**	**polynomial_type**
RIA	REGULARITY SYM_REGULARITY STABILITY SYM_STABILITY UNBIASED_OUTPUTS SYM_UNBIASED_OUTPUTS UNBIASED_COEFFS ABSOLUTE_NOISE_IMMUNITY SYM_ABSOLUTE_NOISE_IMMUNITY	[1,26]	[1,16]	-1	[0.01, 0.25, 0.5, 0.75, 0.99]	QUADRATIC LINEAR LINEAR_COV
MIA	REGULARITY	[3,6]	[1,16]	-1	[0.01, 0.25, 0.5, 0.75, 0.99]	QUADRATIC LINEAR LINEAR_COV
COMBI	REGULARITY SYM_REGULARITY STABILITY SYM_STABILITY UNBIASED_OUTPUTS SYM_UNBIASED_OUTPUTS UNBIASED_COEFFS ABSOLUTE_NOISE_IMMUNITY SYM_ABSOLUTE_NOISE_IMMUNITY	-	[1,16]	-1	[0.01, 0.25, 0.5, 0.75, 0.99]	-
MULTI	REGULARITY SYM_REGULARITY STABILITY SYM_STABILITY UNBIASED_OUTPUTS SYM_UNBIASED_OUTPUTS UNBIASED_COEFFS ABSOLUTE_NOISE_IMMUNITY SYM_ABSOLUTE_NOISE_IMMUNITY	[1,26]	[1,16]	-1	[0.01, 0.25, 0.5, 0.75, 0.99]	-

Fonte: Elaborada pelo autor (2024).

4 RESULTADOS E DISCUSSÕES

Neste capítulo, serão apresentados os resultados e discussões obtidos por meio da aplicação de modelos GMDH na previsão dos valores de FBHP.

De acordo com o exposto no Capítulo 3, os dados foram submetidos a divisões de treino e teste de nas razões de 60/40, 70/30 e 80/20 (treinamento/teste). E cada uma das divisões executadas 5, 25 e 50 vezes. As Tabelas 4.1, 4.2 e 4.3 apresentam resultados médios obtidos por cada modelos de GMDH. O desempenho dos modelos é avaliado através das métricas apresentadas na Seção 3.2. Os indicadores na divisão 60/40 apresentaram exatamente os mesmos resultados que a divisão 70/30 em todos os *folds* para todos os algoritmos.

Tabela 4.1 – Resultados médios para os modelos de GMDH usados para prever os valores de FBHP no conjunto de teste divido na razão 70/30 com 5 execuções. Valores entre parênteses indicam o desvio padrão. Valores destacados em negrito indicam os melhores valores médios.

Modelo	R^2	MSE	RMSE	MAPE
RIA	**0,8161 (0,0298)**	**27200,1389 (6445,3726)**	**163,7129 (19,9560)**	**5,1379 (0,5564)**
MULTI	0,6696 (0,0384)	48472,9689 (7607,3357)	219,3927 (18,4346)	7,4130 (0,9088)
COMBI	0,6696 (0,0384)	48472,9689 (7607,3357)	219,3927 (18,4346)	7,4130 (0,9088)
MIA	0,5612 (0,0987)	64150,7782 (15766,2555)	251,2992 (31,6149)	8,1766 (0,8921)

Fonte: Elaborada pelo autor (2024).

Tabela 4.2 – Resultados médios para os modelos de GMDH usados para prever os valores de FBHP no conjunto de teste dividido na razão 80/20 com 5 execuções. Valores entre parênteses indicam o desvio padrão. Valores destacados em negrito indicam os melhores valores médios.

Modelo	R^2	MSE	RMSE	MAPE
RIA	**0,8256 (0,0266)**	**25725,5576 (5762,9172)**	**159,3830 (17,9615)**	**5,2100 (0,5549)**
MULTI	0,5438 (0,0901)	66702,3256 (14702,8219)	256,6853 (28,5479)	8,3345 (0,8679)
COMBI	0,5479 (0,0936)	66175,2228 (15357,4388)	255,4794 (30,0917)	8,2738 (0,9711)
MIA	0,6483 (0,0532)	51605,4144 (10331,1230)	226,0056 (22,9537)	7,5228 (0,7182)

Fonte: Elaborada pelo autor (2024).

A Tabela 4.4 registra quais os hiperparâmetros foram determinados no melhor cenário que é o apontado na Tabela 4.3.

A partir de 25 execuções, cada métrica de cada modelo, convergiu para o mesmo valor, dentro de suas partições de 70/30 ou 80/20. Os testes foram realizado com a precisão de 10^{-13}.

Tabela 4.3 – Resultados médios para os modelos de GMDH usados para prever os valores de FBHP no conjunto de teste dividido na 70/30 com 25 execuções. Valores entre parênteses indicam o desvio padrão. Valores destacados em negrito indicam os melhores valores médios.

Modelo	R^2	MSE	RMSE	MAPE
RIA	**0,8256 (0,0266)**	**25725,5576 (5762,9172)**	**159,3830 (17,9615)**	**5,2100 (0,5549)**
MULTI	0,6696 (0,0384)	48472,9689 (7607,3357)	219,3927 (18,4346)	7,4130 (0,9088)
COMBI	0,6696 (0,0384)	48472,9689 (7607,3357)	219,3927 (18,4346)	7,4130 (0,9088)
MIA	0,6483 (0,0532)	51605,4144 (10331,1230)	226,0056 (22,9537)	7,5228 (0,7182)

Fonte: Elaborada pelo autor (2024).

Tabela 4.4 – Tabela de Melhores Hiperparâmetros dos Algoritmos

	Hiperparâmetros					
Algoritmo	**criterion**	**k_best**	**p_average**	**n_jobs**	**test_size**	**polynomial_type**
RIA	SYM_REGULARITY	4	1	-1	0.25	QUADRATIC
MIA	REGULARITY	4	2	-1	0.25	QUADRATIC
COMBI	STABILITY	-	1	-1	0.25	-
MULTI	STABILITY	1	1	-1	0.25	-

Fonte: Elaborada pelo autor (2024).

Cabe observar que na divisão 70/30, com 5 execuções os modelos combinatórios (MULTI e COMBI) já haviam atingido seus melhores resultados. Cenário em que os modelos RIA e MIA apresentaram seus piores desempenhos (RIA ainda superior aos outros). Ainda com 5 execuções, mas com divisão 80/20 o cenário se inverteu com os modelos RIA e MIA já apresentando seus melhores resultados e MULTI e COMBI apresentando uma queda no desempenho conforme Tabelas 4.1 e 4.2

Com 50 execuções seja com divisão 70/30 ou 80/20 o desempenho se mantém inalterado, seguindo os mesmos valores da Tabela 4.3.

Apesar de as Tabelas acima demonstrarem uma sensibilidade dos modelos combinatórios (MULTI e MIA) à proporção de divisão entre treino e teste, seus valores não se alteraram para a divisão de 60/40 nas diversas execuções. Os outros modelos também não se alteraram com essa divisão.

Percebe-se uma sensibilidade dos modelos combinatórios à mudança partição, naturalmente, uma vez que quanto mais combinações puderem ser feitas maior a probabilidade de acerto. A quantidade de combinações está ligada à disponibilidade maior de dados de teste, caso em que 70/30 se mostrou mais favorável que 80/20 para esses algoritmos. Interessante notar que disponibilizando uma partição de 60/40 tais algoritmos não sofreram alterações em seus indicadores.

Já os modelos iterativos (RIA e MIA) se mostraram mais sensíveis ao número de execuções. Ambos atingiram seu melhor desempenho somente em 25 *folds* ao passo que os combinatórios já não mostraram seu melhor desempo com 5 *folds*. A Figura 4.1 ilustra essas situações.

Na Figura 4.1 os círculos vermelhos representam o pior desempenho e os símbolos verdes representam o melhor desempenho.

Particionamento	*Folds*	Modelos				Ref. no texto
		RIA	MULTI	COMBI	MIA	
60/40 – 70/30	5	●	✓	✓	●	Tabela 4.1
	25	✓	✓	✓	✓	Tabela 4.3
	50	✓	✓	✓	✓	-
80/20	5	✓	●	●	✓	Tabela 4.2
	25	✓	●	●	✓	-
	50	✓	●	●	✓	-

Figura 4.1 – Representação esquemática de desempenho x particionamentos x folds

Fonte: Elaborada pelo autor (2024).

Nota-se claramente um performance superior do modelo RIA em relação aos outros modelos, em todas as métricas. Esse desempenho (R^2 mais alto, MSE, RMSE e MAPE mais baixos), indica tanto precisão quanto robustez na predição. MULTI e COMBI têm desempenhos idênticos, mas ambos significativamente inferiores a RIA em todas as métricas mas melhores que o modelo MIA que apresenta o pior desempenho intermediário.

Ainda que o modelo RIA tenha ficado com um R^2 mais alto que os outros modelos, ainda ficou bem abaixo do valor de 0,98 apresentado por Tariq (22) utilizando PSO-ANN, conforme apresentado na Seção 1.5. Ainda na revisão bibliográfica nota-se que os melhores modelos são aqueles em que foram aplicadas técnicas híbridas, como o proprio Tariq (22), Goliatt (23). Por outro lado o presente trabalho apresenta um valor de R^2 acima de Campos(8) com um valor de 0,756. O RMSE de 159,3830 do presente trabalho é um valor alto se comparado com o maior valor RMSE encontrado na bibliografia que é de Golliat (23).

Em relação ao desvio padrão, também o modelo RIA apresenta os menores valores, indicando uma maior consistência e confiabilidade em suas previsões. MIA apresenta alto desvio padrõe em todas as métricas, sugerindo maior variabilidade e menos consistência

em suas previsões. MULTI e COMBI, com desvios padrões intermediários, mostram maior consistência que MIA, mas ainda menos consistentes que RIA.

Efetuando uma análise de erros tendo em vista o erro médio de previsão, a largura da faixa de incerteza e o intervalo de confiança observa-se que o modelo RIA apresenta o menor erro médio de previsão, seguido pelo modelo MIA, enquanto os modelos MULTI e COMBI possuem o maior erro médio da previsão. Esses resultados são exibidos na Tabela 4.5. O modelo RIA também tem a menor largura da faixa de incerteza, o que significa que suas previsões são mais consistentes em comparação aos outros modelos. O modelo MIA tem uma faixa de incerteza menor do que os modelos MULTI e COMBI. O intervalo de confiança do modelo RIA é o mais estreito e mais próximo ao seu erro médio de previsão, seguido pelo modelo MIA. Os modelos MULTI e COMBI têm intervalos de confiança mais amplos, indicando maior variabilidade nas previsões. Em resumo, o modelo RIA se destaca como o melhor devido à sua maior precisão e consistência nas previsões, enquanto os modelos MULTI e COMBI apresentam desempenho inferior em todos os critérios analisados.

Tabela 4.5 – Análise de erros para os modelos RIA, MULTI, COMBI e MIA

Modelo	Erro médio da previsão (psi)	Larg. faixa de incerteza (psi)	Intervalo de confiança (psi)
RIA	65,5328	589,6765	42,1505 a 88,9149
MULTI	101,3528	799,9336	69,6333 a 133,0721
COMBI	101,3528	799,9336	69,6333 a 133,0721
MIA	77,0920	748,0375	47,4303 a 106,7536

Fonte: Elaborada pelo autor (2024).

4.1 ANÁLISE DE INCERTEZA

Uma análise de incerteza para a modelagem do FBHP é exibida na Tabela 4.6. O erro médio absoluto (MAD)

$$\text{MAD} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} |\text{FBHP}_i - mediana(\text{FBHP})| \tag{4.1}$$

é usado para o cálculo da incerteza dado por

$$\text{Incerteza } \% = \frac{\text{MAD}}{mediana(\text{FBHP})} \times 100 \tag{4.2}$$

onde $N_{MC} = 159$ e $FBHP_i$ é a pressão de fundo de poço prevista para a i-ésima amostra.

Tendo o menor RMSE entre todos, o modelo RIA tem previsões mais próximas dos valores observados. No entanto possui maior MAD e maior incerteza. Já os modelos

Tabela 4.6 – Análise de incerteza para os modelos RIA, MULTI, COMBI e MIA (Resultados médios).

Modelo	Mediana (psi)	MAD (psi)	Incerteza (%)	RMSE (psi)
RIA	2442,4490	299,7079	12,2831	159,3830
MULTI	2425,2755	257,1305	10,6046	219,3927
COMBI	2425,2755	257,1305	10,6046	219,3927
MIA	2443,4703	262,6479	10,7417	226,0056

Fonte: Elaborada pelo autor (2024).

MIA e COMBI apresentaram boa precisão com a menor incerteza dentre os modelos e também o menor índice MAD.

O modelo COMBI apresentou os mesmos valores que o modelo MULTI, incerteza baixa (10,6046%), MAD baixo (257,1305), e RMSE igual a 218,3927.

Já o modelo MIA tem uma incerteza relativamente baixa (10,7417%), próxima à dos modelos MULTI e COMBI. Apresenta um MAD mais alto (262,6479) e um RMSE maior que os modelos MULTI e COMBI (226,0056), indicando previsões menos precisas. Apesar do MAD ser menor que o do RIA, ele ainda é maior comparado aos modelos MULTI e COMBI.

Em termos de precisão das previsões (RMSE) o modelo RIA é o melhor, pois tem o menor RMSE (159,3830), apesar de ter maior incerteza e variabilidade dos dados. Em termos de incerteza e a variabilidade dos dados (MAD) os modelos MULTI e COMBI apresentam desempenho idênticos e seriam escolhas melhores devido à sua menor incerteza (10,6046%) e menor MAD (257,1305).

Considerando todas as métricas até o momento, o modelo RIA pode ser considerado o melhor para previsões precisas, enquanto os modelos MULTI e COMBI são mais adequados se a precisão relativa (baixa incerteza) for mais valorizada. O modelo MIA se destaca como tendo a pior performance.

A Figura 4.2 ilustra o gráfico relacionando RMSE e incerteza demonstrando que o modelo RIA possui menor RMSE e maior incerteza, em contraposição aos três outros modelos.

O diagrama de Taylor (Figura 4.3) foi elaborado com os dados de previsão e testes. Esse diagrama corrobora as métricas explicitadas até agora, na medida em que revela o modelo RIA com o maior R^2 e menor RMSE.

A Figura 4.4 exibe a dispersão dos valores preditos do modelo RIA indicando uma coerência com o valor de 0,82 para R^2, sobretudo nos valores próximos à mediana (2401,8030 psi).

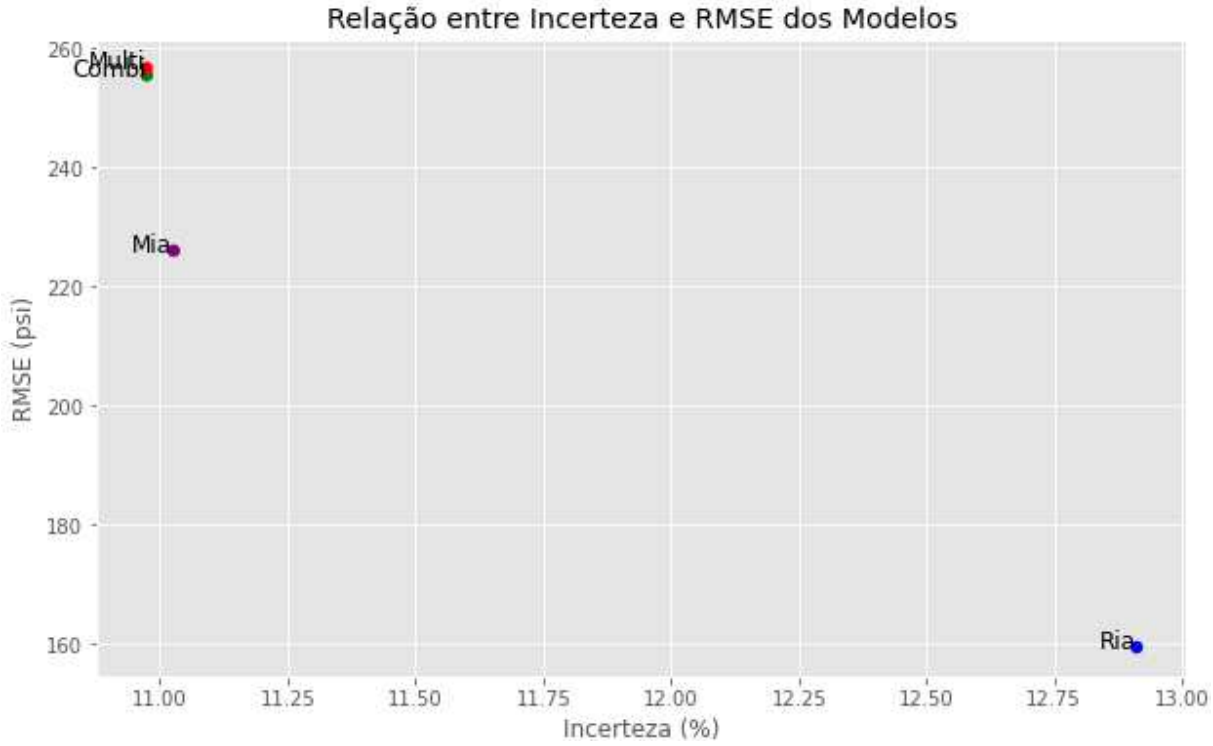

Figura 4.2 – Comparação gráfica da incerteza x RMSE dos modelos avaliados

Fonte: Elaborada pelo autor (2024).

4.2 MODELOS POLINOMIAIS

Para cada modelo em sua melhor configuração foram geradas as seguintos formulações polinomiais, representadas pela equações abaixo:

Equação 4.3, referente ao modelo RIA:

$$y = -1.15661e-13 * x_6 + f_{51} + 4.35659e-17 * x_6 * f_{51} - 8.33846e-15 * x_6^2 \\ +4.702e-2 * f_{51}^2 + 1.28495e-12 \quad (4.3)$$

Onde as funções de f_1 a f_{51} são explicitadas no Apêndice A.

Equação 4.4, referente ao modelo COMBI:

$$y = 274.2747 * x_1 + 186.0511 * x_2 + 73.1367 * x_3 - 93.9845 * x_4 + 152.5278 * x_5 \\ -19.1894 * x_6 - 62.9113 * x_7 - 4.0596 * x_8 + 2489.7719 \quad (4.4)$$

Equação 4.5, referente ao modelo MULTI:

$$y = 274.2747 * x_1 + 186.0511 * x_2 + 73.1367 * x_3 - 93.9845 * x_4 + 152.5278 * x_5 \\ -19.1894 * x_6 - 62.9113 * x_7 - 4.0596 * x_8 + 2489.7719 \quad (4.5)$$

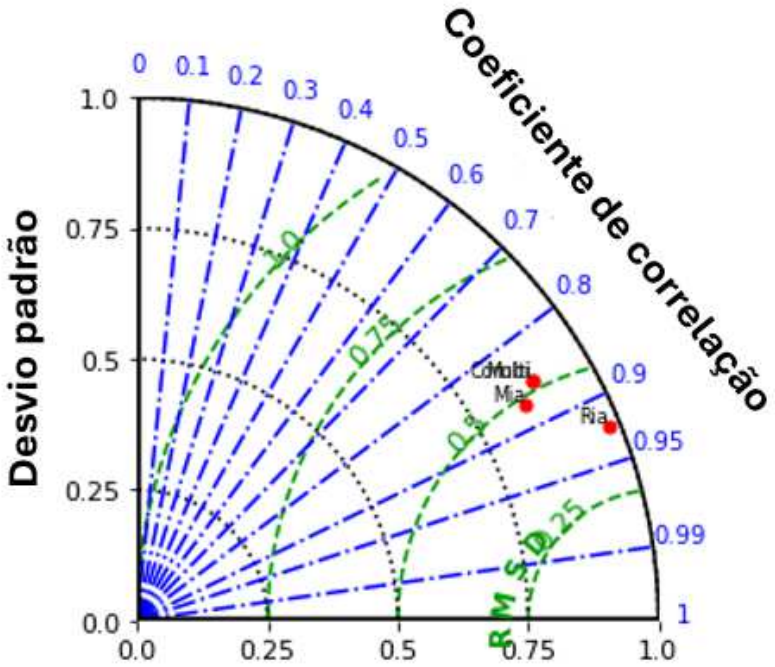

Figura 4.3 – Diagrama de Taylor

Fonte: Elaborada pelo autor (2024).

Equação 4.6, referente ao modelo MIA:

$$y = 3.6568 f_1^3 - 2.6698 f_2^3 - 0.0044 f_1^3 f_2^3 + 0.0016 (f_1^3)^2 + 0.0027 (f_2^3)^2 + 19.1692 \quad (4.6)$$

onde as funções de f_1^n a f_3^n são dadas por:

$$f_1^1 = 240.8531 x_1 + 177.0585 x_5 + 21.1086 x_1 x_5 - 30.85 x_1^2 - 0.8812 x_5^2 + 2532.8779$$

$$f_1^2 = 307.1902 * x_1 + 237.1291 * x_2 + 25.4238 * x_1 * x_2 - 22.1628 * x_1^2 - 23.7596 * x_2^2 + 2558.6253$$

$$f_1^3 = 237.2173 * x_1 + 202.4862 * x_8 - 14.713 * x_1 * x_8 - 24.3136 * x_1^2 + 59.6844 * x_8^2 + 2464.3955$$

$$f_1^4 = 223.3231 * x_5 + 218.2477 * x_8 + 20.0841 * x_5 * x_8 + 22.7496 * x_5^2 + 80.5464 * x_8^2 + 2388.0514$$

$$f_2^1 = -0.9055 * f_1^2 + 2.3997 * f_1^4 + 0.0005 * f_1^2 * f_1^4 + 8.70503e-05 * (f_1^2)^2 - 0.0006 * (f_1^4)^2 - 1215.4082$$

$$f_2^2 = 0.0005 * f1_1 + 0.373 * f_1^2 + 0.0013 * f1_1 * f_1^2 - 0.0005 * (f_1^1)^2 - 0.0006 * (f_1^2)^2 + 585.6934$$

$$f_2^3 = 0.5489 * f_1^1 + 1.7848 * f_1^4 - 0.0013 * f_1^1 * f_1^4 + 0.0007 * (f_1^1)^2 + 0.0004 * (f_1^4)^2 - 1838.7813$$

$$f_2^4 = -2.0676 * f_1^3 + 2.2005 * f_1^4 + 0.0006 * f_1^3 * f_1^4 + 0.0003 * (f_1^3)^2 - 0.0006 * (f_1^4)^2 + 510.8552$$

$$f_3^1 = 0.8802 * f_2^1 - 0.6947 * f_2^3 + 0.0015 * f_2^1 * f_2^3 - 0.0007 * (f_2^1)^2 - 0.0006 * (f_2^3)^2 + 932.3497$$

$$f_3^2 = -1.2724 * f_2^2 + 2.8012 * f_2^4 + 0.0005 * f_2^2 * f_2^4 + 0.0001 * (f_2^2)^2 - 0.0007 * (f_2^4)^2 - 749.259$$

A Equação 4.3, do modelo RIA possui uma riqueza estrutural que reflete na precisão do modelo. A não linearidade associada à utilização de 51 sub-funções consegue refletir a complexidade da previsão do FBHP. A presença somente de exponenciações sem a utilização de funções mais complexas como seno ou cosseno é um grande diferencial ao

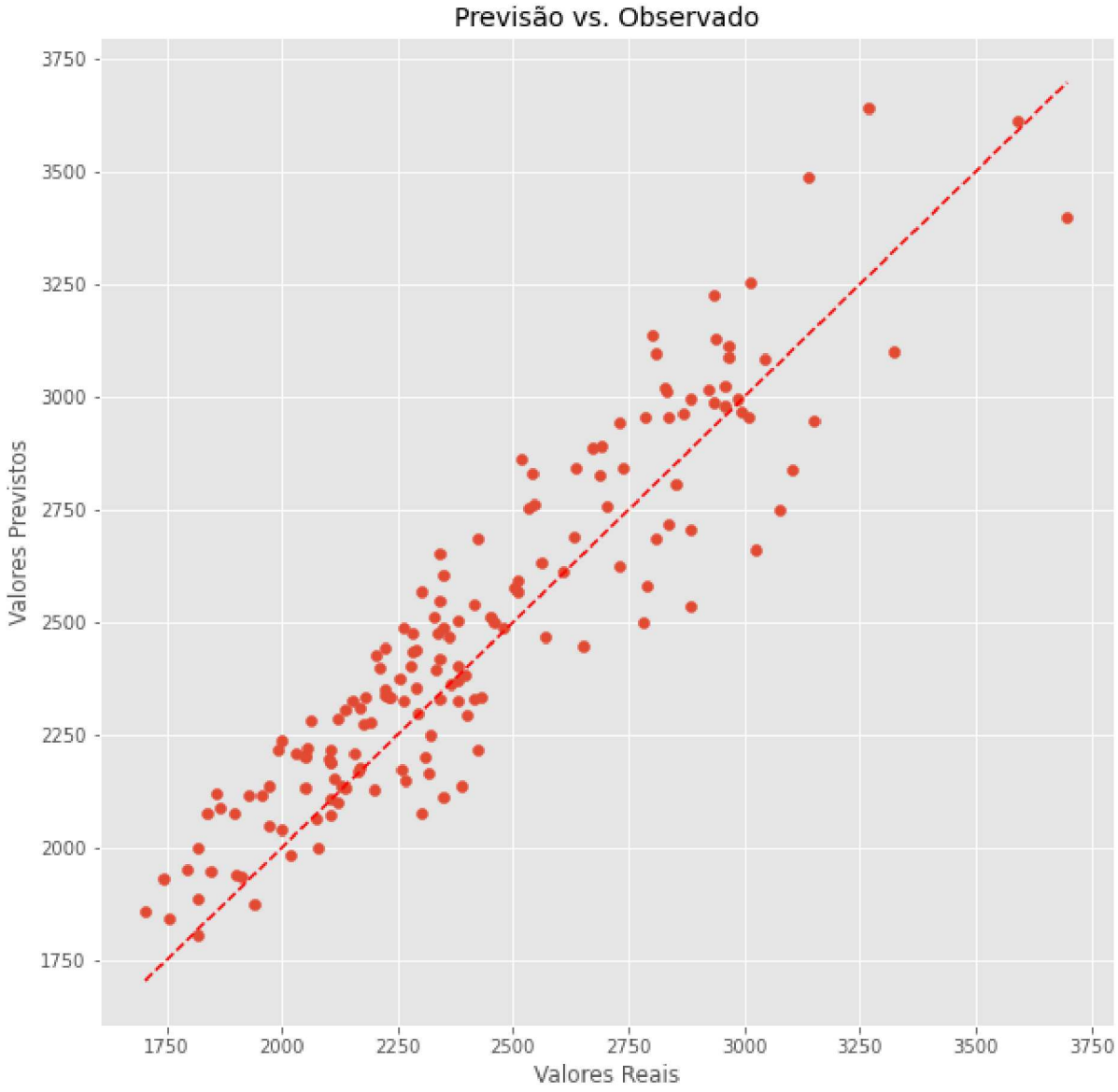

Figura 4.4 – Gráfico de dispersão modelo RIA

Fonte: Elaborada pelo autor (2024).

introduzir de certa forma uma simplificação na formulação do problema, podendo refletir inclusive em ganho de performance computacional em relação a outras formulações.

Os modelos COMBI e MULTI apresentaram exatamente as mesmas equações (equações 4.4 e 4.5). Como foi colocado na seção 2, sendo o MULTI um aperfeiçoamento do COMBI, os recursos adicionais do MULTI não agregraram valor em performance e convergiram ao estágio que o COMBI.

O modelo MIA (Equação 4.6) apesar de também possuir uma natureza indutiva assim como o RIA, comprova que a não linearidade por si só, não é garantia de melhor performance, ainda que utilizando sub-funções como o modelo RIA.

A performance e a modelagem polinomial dos modelo COMBI e MULTI explicitam

que o problema pode ser abordado tanto da forma indutiva quanto combinatorial. A formulação polinomial apresentada nessa seção explicita e complementa bastante as análises efetuadas em torno das métricas de erro. O algoritmo mais promissor (RIA) é representado por uma Equação mais sofisticada ao passo que os algoritmos com menor performance (COMBI e MULTI) possuem formulação linear.

4.3 TEMPOS DE EXECUÇÃO

Para cada modelo em cada execução, foram compuatados seus tempos conforme Tabela abaixo:

Tabela 4.7 – Tempo médio de processamento em milisegundos com desvios padrão (calculado em 50 execuções independentes).

Modelo	Tempos (ms)
RIA	14,4137 ± 8,0810
MULTI	3,3533 ± 6,1692
COMBI	4,9909 ± 7,4695
MIA	5,6419 ± 7,1049

Fonte: Elaborada pelo autor (2024).

Apesar do alto desvio padrão de todos os modelos, o tempo de execução se mostrou como um diferencial deste trabalho com GMDH. RIA com maior tempo de execução e MULTI com menor tempo, como era esperado pela natureza do próprio algoritmo. E MULTI e COMBI em uma posição intermediária de custo computacional. Em comparação com outros estudos onde houve aferição de tempo, como Campos (8) e Mulashani (26) onde os melhores modelos apresentaram 2900 segundos e 1,15 segundos respectivamente, o presente trabalho se destaca com custo computacional na ordem de milissegundos.

5 CONCLUSÃO E TRABALHOS FUTUROS

A abordagem efetuada no trabalho acrescenta mais uma visão a trabalhos já existentes ao redor do método GMDH ao explorar quatro algoritmos diferentes desse método e também propõe formulação em programação simbólica com indicador R^2 relevante.

O estudo demonstra que o método GMDH, mais especificamente utilizando o modelo RIA se mostrou eficiente para modelagem do valor de FBHP. Além disso foi gerado o modelo polinomial relativo não somente ao RIA, mas também a outros três modelos, o que fornece uma opção interpretável do problema.

O trabalho aponta também a sensibilidade do GMDH a variações na relação da divisão da base em treino e teste, apontando uma perfomance melhor com percentuais de 80% para treinamento.

Foi demonstrado também que abordagens tanto polinomiais quanto combinatórias podem ser exploradas e otimizadas no tratamento da previsão de FBHP.

A característica de custo computacional também se mostrou relevante, ao apresentar valores muito baixos de execução, ainda considerando um alto desvio padrão.

Como trabalhos futuros, pode ser feita a exploração de outros algoritmos da família GMDH como *Objective System Analysis* (OSA), *Multiplicative-Additive* (MAA) ou *Objective Computer Clusterization* (OCC). Outros dados também podem ser testados representando campos de petróleo com características diferentes, bem como pode-se acrescentar a técnica de seleção de características nos algoritmos já testados. Outras faixas de particionamento entre treino e teste podem ser exploradas na tentativa de sensibilizar os diversos algoritmos.

REFERÊNCIAS

1 AL-SHAMMARI, Ahmed. **Accurate Prediction of Pressure Drop in Two-Phase Vertical Flow Systems using Artificial Intelligence**. volume All Days of SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, pages SPE–149035–MS, 05 2011. doi: 10.2118/149035-MS. URL https://doi.org/10.2118/149035-MS.

2 TRUJILLO, Leonardo, Stephan M Winkler, Sara Silva, Wolfgang Banzhaf. **Genetic Programming Theory and Practice** XIX. Springer Nature, 2023.

3 MADALA, Hema Rao, Alexey G. Ivakhnenko, **Inductive learning algorithms for complex systems modeling**. Boca Raton: CRC, 1994.

4 IVAKHNENKO A.G. Polynomial Theory of Complex Systems. **IEEE Transactions on Systems, Man and Cybernetics**, Vol.SMC-1, no.4, October 1971

5 RUMELHART, D. E., Hinton, G. E., and Williams, R. J.. Learning representations by back-propagating errors. **Nature**, , 323, 533–536, 1986.

6 Ucrânia. Instituto Nacional para Assuntos Estratégicos. Centro Internacional para Tecnologia da Informação e Sistemas da Academia Nacional de Ciências da Ucrânia. **GMDH**.Disponível em: https://www.gmdh.net/index.html. Acesso em: 8 ago. 2024.

7 FARLOW, S.J., **Self-organizing Methods in Modeling: GMDH Type Algorithms.**. New York: CRC, 1984

8 CAMPOS, Deivid **Modelagem da pressão de fundo de poço em sistemas de escoamento multifásico: uma abordagem utilizando programação genética**. Dissertação (Mestrado em Modelagem Computacional)Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Modelagem Computacional, 2024.

9 Fórum Econômico Mundial. **Why do oil prices matter to the global economy? An expert explains**, Maciej Kolaczkowski e Amy White. Genebra. Disponível em: https://www.weforum.org/agenda/2022/02/why-oil-prices-matter-to-global-economy-expert-explains/. Acesso em: 3 ago. 2024.

10 IEA (International Energy Agency), **Oil 2024**, Paris. Disponível em: https://www.iea.org/reports/oil-2024, Licence: CC BY 4.0. Acesso em 3 ago. 2024.

11 IEA (International Energy Agency), **World Energy Outlook 2011**, Paris. Disponível em https://www.iea.org/reports/world-energy-outlook-2011, Licence: CC BY 4.0. Acesso em 3 ago. 2024.

12 Offshore Engineer Asia, **Australia Snub IEA's Call to Halt new Fossil Fuel Investments**, Sonali Paul. Melbourne. Disponível em https://www.oedigital.com/news/487760-asia-australia-snub-iea-s-call-to-halt-new-fossil-fuel-investments. Acesso em 3 ago. 2024.

13 IEA (International Energy Agency), **Net Zero by 2050 - A Roadmap for the Global Energy Sector**, Disponível em https://www.iea.org/reports/net-zero-by-2050. Acesso em 3 ago. 2024.

14 ALCOFRA, Elisa Lage Modesto **Aumento de pressão de fluido confinado no anular de poços de petróleo**. Dissertação (Mestrado) Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Mecânica, 2014.

15 NOAA (National Oceanic and Atmospheric Administration), **Deepwater Horizon Oil Spill**, Disponível em https://darrp.noaa.gov/oil-spills/deepwater-horizon. Acesso em 3 ago. 2024.

16 EPA (United States Environmental Protection Agency), **Deepwater Horizon - BP Gulf of Mexico Oil Spill**, Disponível em https://www.epa.gov/enforcement/deepwater-horizon-bp-gulf-mexico-oil-spill. Acesso em 3 ago. 2024.

17 United States Department of the Interior, **Restoring the Gulf of Mexico After the Deepwater Horizon Oil Spill**, Disponível em https://www.doi.gov/deepwaterhorizon. Acesso em 3 ago. 2024

18 ONU (Organização das Nações Unidas), **Objetivos de Desenvolvimento Sustentável**, Brasil, Disponível em https://brasil.un.org/pt-br/sdgs. Acesso em 3 ago. 2024.

19 MAKKE, N., Chawla, S. **Interpretable scientific discovery with symbolic regression: a review**. Artificial Intelligence Review 57, 2 (2024). https://doi.org/10.1007/s10462-023-10622-0

20 BUCHBERGER, B. **Automated programming, symbolic computation, machine learning: my personal view**. Annals of Mathematics and Artificial Intelligence, 91, 569–589 (2023). https://doi.org/10.1007/s10472-023-09894-7

21 LA CAVA, W.G., Lee, P.C., Ajmal, I. et al. **A flexible symbolic regression method for constructing interpretable clinical prediction models**. npj Digital Medicine. 6, 107 (2023). https://doi.org/10.1038/s41746-023-00833-8

22 TARIQ, Z., Mahmoud, M. Abdulraheem, A. **Real-time prognosis of flowing bottom-hole pressure in a vertical well for a multiphase flow using computational intelligence techniques**. Journal of Petroleum Exploration and Produtcion Technology 10, 1411–1428 (2020). https://doi.org/10.1007/s13202-019-0728-4

23 GOLLIAT, Leonardo, Reem Sabah Mohammad, Sani I. Abba, Zaher Mundher Yaseen. **Development of hybrid computational data-intelligence model for flowing bottom-hole pressure of oil wells: New strategy for oil reservoir management and monitoring**. Fuel, 350:128623, 2023. ISSN 0016-2361. doi: https://doi.org/10.1016/j.fuel.2023.128623. URL https://www.sciencedirect.com/science/article/pii/S001623612301236X

24 AYOUB, M.A., Negash, B.M., Saaid, I.M. (2015). **Modeling Pressure Drop in Vertical Wells Using Group Method of Data Handling (GMDH) Approach**. In: Awang, M., Negash, B., Md Akhir, N., Lubis, L. (eds) ICIPEG 2014. Springer, Singapore. https://doi.org/10.1007/978-981-287-368-26

25 OSMAN, E. A., Ayoub, M. A., and M. A. Aggour. **Artificial Neural Network Model for Predicting Bottomhole Flowing Pressure in Vertical Multiphase Flow**, Paper presented at the **SPE Middle East Oil and Gas Show and Conference**, Kingdom of Bahrain, March 2005. doi: https://doi.org/10.2118/93632-MS

26 MULASHANI, Alvin K. , Chuanbo Shen, Baraka M. Nkurlu, Christopher N. Mkono, Martin Kawamala, **Enhanced group method of data handling (GMDH) for permeability prediction based on the modified Levenberg Marquardt technique from well log data**, Energy, Volume 239, Part A, 2022, 121915, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2021.121915.

27 GAO Guozhong, Hazbeh Omid, et. al. **Application of GMDH model to predict pore pressure**, Frontiers in Earth Science vol. 10, 2023, doi 10.3389/feart.2022.1043719, ISSN: 2296-6463

28 GUO, J., Wang, H., Guo, F. et al. **The back propagation based on the modified group method of data-handling network for oilfield production forecasting**. J Petrol Explor Prod Technol 9, 1285–1293 (2019). https://doi.org/10.1007/s13202-018-0582-9

29 MESBAH M, Habibnia S, Ahmadi S, Dehaghani AHS, Bayat S. **Developing a robust correlation for prediction of sweet and sour gas hydrate formation temperature**. Petroleum 2020. https://doi.org/10.1016/j.petlm.2020.07.007.

30 MAHDAVI-MEYMAND A, Zounemat-Kermani M. **A new integrated model of the group method of data handling and the firefly algorithm (GMDH-FA): application to aeration modelling on spillways**. Artificial Intelligence Review 2020;53(4): 2549e69.

31 AMAR MN, Larestani A, Lv Q, Zhou T, Hemmati-Sarapardeh A. **Modeling of methane adsorption capacity in shale gas formations using white-box supervised machine learning techniques**. J Petrol Sci Eng 2021:109226.

32 MAHDAVIARA M, Rostami A, Shahbazi K. **State-of-the-art modeling permeability of the heterogeneous carbonate oil reservoirs using robust computational approaches**. Fuel 2020;268:117389.

33 YOUCEFI MR, Hadjadj A, Boukredera FS. **New model for standpipe pressure prediction while drilling using Group Method of Data Handling**. Petroleum 2021. https://doi.org/10.1016/j.petlm.2021.04.003.

34 **Plano Estratégico 2024-2028 Petrobrás**, Disponível em https://petrobras.com.br/quem-somos/estrategia?_gl=1*1xtvrg9*_gcl_au*MTI4NTUwMzg5Ny4xNzIzOTM1MDgz*_ga*MjEzNj . Acesso em 17 ago. 2024

35 AHMED, T. **Reservoir Engineering Handbook**. Gulf Professional Publishing, 2010.

36 MILLIKAN, Charles V.; SIDWELL, V. **Bottom-hole pressures in oil wells**. Transactions of the AIME, v. 92, n. 01, p. 194-205, 1931.

37 ROJ, D.J., **Fluxo vertical de misturas de gases e líquidos em poços**. No 6º Congresso Mundial do Petróleo. 1963

38 HAGERDON, A.R., Brown, KE. **Estudo experimental de gradientes de pressão que ocorrem durante fluxo bifásico contínuo em conduítes verticais de pequeno diâmetro**. Jornal de Tecnologia de Petróleo. 1965. 17(04). 475-484.

39 ORKISZEWSKI, J. **Previsão de quedas de pressão bifásicas em tubos verticais**. Jornal do Petróleo Tecnologia, 1967. 19(06). 829-838.

40 ANSARI, A.M., Sylvester, N.D., Sarica, C., Shoham, O., Brill, JP. **Um modelo mecanístico abrangente para fluxo ascendente de duas fases em poços**. Produção e instalações de SPE. 1994. 9(02). 143-151.

41 GOMEZ, LE, Shoham, O., Schmidt, Z., Chokshi, RN, Northug, T. **Modelo mecanístico unificado para fluxo bifásico em estado estacionário: fluxo ascendente horizontal a vertical**. Revista SPE. 2000. 5(03). 339-350.

42 AZMI R.P.A.; Yusoff, M.; Mohd Sallehud-din, M.T. **A Review of Predictive Analytics Models in the Oil and Gas Industries**. Sensors 2024, 24, 4013. https://doi.org/10.3390/s24124013

43 CHOQUETTE, P. W., Pray, L. C. (1970). **Geologic nomenclature and classification of porosity in sedimentary carbonates**. AAPG Bulletin.

44 SCHRAMM, L. L. **Foams: Fundamentals and Applications in the Petroleum Industry. American Chemical Society**. 1994

45 MUKHERJEE, H., Brill, J. P. **Pressure drop correlations for inclined two-phase flow**. Journal of Energy Resources Technology. 1985

46 **Dicionario do Petróleo e Gás em Língua Portuguesa**. Disponível em https://dicionariodopetroleo.com.br/manometro-fundo/, Acesso em 20 ago. 2024.

47 ABDULRAUF R Adebayo, Abdulazeez Abdulraheem, Sunday O Olatunji. **Artificial intelligence based estimation of water saturation in complex reservoir systems**. Journal of Porous Media, 18(9), 2015.

48 AHMED Buhulaigah, Ali S Al-Mashhad, Sulaiman A Al-Arifi, Mohammed S Al-Kadem, Mohammed S Al-Dabbous. **Multilateral wells evaluation utilizing artificial intelligence**. In SPE Middle East Oil and Gas Show and Conference, page D031S028R005. SPE, 2017.

49 MAHDY, Ali, Wael Zakaria, Ahmed Helmi, Ahmad Sobhy Helaly, Abdullah M.E. Mahmoud. **Machine learning approach for core permeability prediction from well logs in sandstone reservoir, mediterranean sea, egypt**. Journal of Applied Geophysics, 220: 105249, 2024. ISSN 0926-9851. doi: https://doi.org/10.1016/j.jappgeo.2023.105249.

50 AGGREY G. H., D. R. Davies. **Tracking the State and Diagnosing Downhole Permanent Sensors in Intelligent-Well Completions With Artificial Neural Network**. volume All Days of SPE Offshore Europe Conference and Exhibition, pages SPE–107198–MS, 09 2007. doi: 10.2118/107198-MS.

51 ALTON R Hagedorn, Kermit E Brown. **Experimental study of pressure gradients occurring during continuous two-phase flow in small-diameter vertical conduits**. Journal of Petroleum Technology, 17(04):475–484, 1965. doi: https://doi.org/10.2118/940-PA.

52 HAGER, B. H. (1991). **Diagenesis and fluid flow in the subsurface**. Reviews of Geophysics.

53 ALCOFRA, Elisa Lage Modesto. **Aumento de pressão de fluido confinado no anular de poços de petróleo**. Orientador: Ângela Ourivio Nieckele. 2014. Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Mecânica, 2014.

54 PATTILLO P.D., COCALES, B.W, MOREY, S.C.,2004 **Analysis of an Annular Pressure Buildup Failure During Drill Ahead**. SPE Deepwater Drilling and Completions Conference, 20-21 June, Galveston, Texas, USA. Paper no. 90151.

55 ROTIMI, J., O., Adeoye, T. O., Ologe, O., Onuh, C. Y. (2015). **Pore Pressure and Fracture Gradientassessment from Well Log Petrophysical Data of Agbada Formation**, South-East Niger-Delta.

56 LUO, L., Meng, W., Gluyas, J., Tan, X., Gao, X., Feng, M., Kong, X., Shao, H. (2019). **Diagenetic characteristics, evolution, controlling factors of diagenetic system and their impacts on reservoir quality in tight deltaic sandstones : typical example from the Xujiahe Formation in Western Sichuan Foreland Basin**, SW China.

APÊNDICE A – Funções componentes do modelo RIA

$$f_1 = 292.542*x_1+222.854*x_2+36.1403*x_1*x_2-18.4365*x_1^2-10.4838*x_2^2+2523.5853$$

$$f_2 = -493.8028 * x_5 + 1.6068 * f_1 + 0.2521 * x_5 * f_1 - 4.0388 * x_5^2 - 0.0001 * f_1^2 - 654.8963$$

$$f_3 = -129.0458 * x_4 + 2.1651 * f_2 + 0.0134 * x_4 * f_2 + 26.9644 * x_4^2 - 0.0002 * f_2^2 - 1548.6547$$

$$f_4 = 82.0141 * x_3 + 0.7654 * f_3 - 0.0486 * x_3 * f_3 + 60.2186 * x_3^2 + 4.56204e-05 * f_3^2 + 231.1158$$

$$f_5 = -145.9778 * x_7 + 1.6811 * f_4 + 0.0309 * x_7 * f_4 + 5.6539 * x_7^2 - 0.0001 * f_4^2 - 885.6034$$

$$f_6 = 114.5424 * x_1 + 1.0556 * f_5 - 0.0385 * x_1 * f_5 + 20.8081 * x_1^2 - 1.82227e-05 * f_5^2 - 36.6165$$

$$f_7 = -140.1323*x_8+0.642*f_6+0.1044*x_8*f_6+46.2092*x_8^2+6.68772e-05*f_6^2+406.9022$$

$$f_8 = 77.4465 * x_6 + 1.1831 * f_7 - 0.0383 * x_6 * f_7 - 0.429 * x_6^2 - 2.97573e-05 * f_7^2 - 259.9308$$

$$f_9 = 241.6074 * x_3 + 1.2907 * f_8 - 0.0864 * x_3 * f_8 - 3.0571 * x_3^2 - 5.51709e-05 * f_8^2 - 377.3244$$

$$f_{10} = 81.6674 * x_5 + 0.9521 * f_9 - 0.0272 * x_5 * f_9 + 12.2381 * x_5^2 + 8.4064e-06 * f_9^2 + 58.581$$

$$f_{11} = -16.9646*x_1+1.1548*f_{10}+0.0103*x_1*f_{10}+0.0235*x_1^2-3.37845e-05*f_{10}^2-173.4602$$

$$f_{12} = -133.5816*x_8+1.0657*f_{11}+0.0547*x_8*f_{11}-1.1279*x_8^2-1.32416e-05*f_{11}^2-86.227$$

$$f_{13} = 13.0935 * x_6 + 1.018 * f_{12} - 0.0082 * x_6 * f_{12} - 1.463 * x_6^2 - 1.54676e-06 * f_{12}^2 - 31.8951$$

$$f_{14} = -60.832*x_2+1.2141*f_{13}+0.0143*x_2*f_{13}+7.2274*x_2^2-3.83681e-05*f_{13}^2-299.0972$$

$$f_{15} = 13.7345 * x_6 + 1.0021 * f_{14} - 0.0052 * x_6 * f_{14} - 1.8264 * x_6^2 - 5.66161e-07 * f_{14}^2 + 0.9743$$

$$f_{16} = 47.9881 * x_7 + 1.0005 * f_{15} - 0.024 * x_7 * f_{15} + 2.334 * x_7^2 + 4.16269e-07 * f_{15}^2 - 6.6389$$

$$f_{17} = -17.3843*x_3+0.943*f_{16}+0.0098*x_3*f_{16}-3.4784*x_3^2+1.26102e-05*f_{16}^2+66.3378$$

$$f_{18} = -0.4667*x_6+1.0065*f_{17}-0.0007*x_6*f_{17}+0.2848*x_6^2-6.68488e-07*f_{17}^2-11.9264$$

$$f_{19} = -0.7152*x_3+0.9953*f_{18}+0.0007*x_3*f_{18}-0.2402*x_3^2+1.05185e-06*f_{18}^2+5.3652$$

$$f_{20} = 0.8219 * x_6 + 1.0005 * f_{19} - 0.0005 * x_6 * f_{19} + 0.0737 * x_6^2 + 9.07473e-09 * f_{19}^2 - 1.1948$$

$$f_{21} = -0.047*x_3+0.9991*f_{20}+7.93496e-05*x_3*f_{20}-0.0382*x_3^2+1.86803e-07*f_{20}^2+0.9809$$

$$f_{22} = 0.1742*x_6+1.0001*f_{21}-9.27583e-05*x_6*f_{21}+0.0143*x_6^2+4.03497e-09*f_{21}^2-0.1911$$

$$f_{23} = -0.0095 * x_3 + 0.9998 * f_{22} + 1.43404e-05 * x_3 * f_{22} - 0.0067 * x_3^2 + 3.42192e-08 * f_{22}^2 + 0.1813$$

$$f_{24} = 0.0317 * x_6 + f_{23} - 1.67071e-05 * x_6 * f_{23} + 0.0025 * x_6^2 + 6.85432e-1 * f_{23}^2 - 0.0337$$

$$f_{25} = -0.0018 * x_3 + f_{24} + 2.57842e-06 * x_3 * f_{24} - 0.0012 * x_3^2 + 6.06707e-09 * f_{24}^2 + 0.0322$$

$$f_{26} = 0.0056 * x_6 + f_{25} - 2.96995e-06 * x_6 * f_{25} + 0.0004 * x_6^2 + 1.21179e-1 * f_{25}^2 - 0.0059$$

$$f_{27} = -0.0003 * x_3 + f_{26} + 4.56168e-07 * x_3 * f_{26} - 0.0002 * x_3^2 + 1.07028e-09 * f_{26}^2 + 0.0057$$

$$f_{28} = 0.001*x_6+f_{27}-5.25341e-07*x_6*f_{27}+7.71377e-05*x_6^2+2.15463e-11*f_{27}^2-0.001$$

$$f_{29} = -5.83141e-05 * x_3 + f_{28} + 8.04836e-08 * x_3 * f_{28} - 3.6408e-05 * x_3^2 + 1.88754e-1 * f_{28}^2 + 0.001$$

$$f_{30} = 0.0002*x_6+f_{29}-9.27284e-08*x_6*f_{29}+1.35964e-05*x_6^2+3.81092e-12*f_{29}^2-0.0002$$

$$f_{31} = -1.02985e-05 * x_3 + f_{30} + 1.41944e-08 * x_3 * f_{30} - 6.41686e-06 * x_3^2 + 3.32843e-11 * f_{30}^2 + 0.0002$$

$$f_{32} = 3.11637e-05 * x_6 + f_{31} - 1.63555e-08 * x_6 * f_{31} + 2.39699e-06 * x_6^2 + 6.72559e-13 * f_{31}^2 - 3.23702e-05$$

$$f_{33} = -1.81688e-06 * x_3 + f_{32} + 2.50297e-09 * x_3 * f_{32} - 1.13126e-06 * x_3^2 + 5.86882e-$$

$12 * f_{32}^2 + 3.11638e - 05$

$f_{34} = 5.49564e - 06 * x_6 + f_{33} - 2.88411e - 09 * x_6 * f_{33} + 4.22612e - 07 * x_6^2 + 1.18619e - 13 * f_{33}^2 - 5.70704e - 06$

$f_{35} = -3.20413e - 07 * x_3 + f_{34} + 4.41333e - 1 * x_3 * f_{34} - 1.99451e - 07 * x_3^2 + 1.03479e - 12 * f_{34}^2 + 5.49482e - 06$

$f_{36} = 9.69034e - 07 * x_6 + f_{35} - 5.08541e - 1 * x_6 * f_{35} + 7.45128e - 08 * x_6^2 + 2.09175e - 14 * f_{35}^2 - 1.00622e - 06$

$f_{37} = -5.64983e - 08 * x_3 + f_{36} + 7.78155e - 11 * x_3 * f_{36} - 3.51661e - 08 * x_3^2 + 1.8245e - 13 * f_{36}^2 + 9.6883e - 07$

$f_{38} = 1.7086e - 07 * x_6 + f_{37} - 8.96655e - 11 * x_6 * f_{37} + 1.31378e - 08 * x_6^2 + 3.68929e - 15 * f_{37}^2 - 1.77405e - 07$

$f_{39} = -9.96291e - 09 * x_3 + f_{38} + 1.37207e - 11 * x_3 * f_{38} - 6.20033e - 09 * x_3^2 + 3.21705e - 14 * f_{38}^2 + 1.70828e - 07$

$f_{40} = 3.01259e - 08 * x_6 + f_{39} - 1.58097e - 11 * x_6 * f_{39} + 2.31643e - 09 * x_6^2 + 6.49265e - 16 * f_{39}^2 - 3.12864e - 08$

$f_{41} = -1.75697e - 09 * x_3 + f_{40} + 2.41936e - 12 * x_3 * f_{40} - 1.09327e - 09 * x_3^2 + 5.67092e - 15 * f_{40}^2 + 3.01151e - 08$

$f_{42} = 5.31206e - 09 * x_6 + f_{41} - 2.78764e - 12 * x_6 * f_{41} + 4.08407e - 1 * x_6^2 + 1.15471e - 16 * f_{41}^2 - 5.51174e - 09$

$f_{43} = -3.09349e - 1 * x_3 + f_{42} + 4.2637e - 13 * x_3 * f_{42} - 1.92702e - 1 * x_3^2 + 1.00408e - 15 * f_{42}^2 + 5.3315e - 09$

$f_{44} = 9.36379e - 1 * x_6 + f_{43} - 4.91432e - 13 * x_6 * f_{43} + 7.20352e - 11 * x_6^2 + 1.69996e - 17 * f_{43}^2 - 9.90464e - 1$

$f_{45} = -5.37502e - 11 * x_3 + f_{44} + 7.48883e - 14 * x_3 * f_{44} - 3.39894e - 11 * x_3^2 + 1.78346e - 16 * f_{44}^2 + 9.45777e - 1$

$f_{46} = -1.53102e - 11 * x_3 + f_{45} + 6.10948e - 15 * x_3 * f_{45} - 3.01894e - 13 * x_3^2 - 9.69964e - 18 * f_{45}^2 - 5.87017e - 11$

$f_{47} = 1.66283e - 1 * x_6 + f_{46} - 8.71318e - 14 * x_6 * f_{46} + 1.24337e - 11 * x_6^2 + 1.54572e - 17 * f_{46}^2 - 9.70722e - 11$

$f_{48} = 5.1083e - 12 * x_3 + f_{47} + 7.34262e - 15 * x_3 * f_{47} - 5.64015e - 12 * x_3^2 + 2.80858e - 17 * f_{47}^2 + 1.4729e - 1$

$f_{49} = 2.84586e - 11 * x_6 + f_{48} - 1.49344e - 14 * x_6 * f_{48} + 2.45167e - 12 * x_6^2 - 5.22331e - 19 * f_{48}^2 - 3.53946e - 11$

$f_{50} = -1.98461e - 12 * x_3 + f_{49} + 2.49039e - 15 * x_3 * f_{49} - 1.10753e - 12 * x_3^2 + 7.52221e - 18 * f_{49}^2 + 4.00077e - 11$

$f_{51} = 5.21294e - 12 * x_6 + f_{50} - 2.71851e - 15 * x_6 * f_{50} + 3.91859e - 13 * x_6^2 - 5.32306e - 19 * f_{50}^2 - 7.94334e - 12$