



Monique Gomes de Araujo

**Previsão da curva de produção para projeto
exploratório utilizando redes neurais artificiais**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre pelo Programa de Pós-graduação
em Engenharia Mecânica do Departamento de Engenharia
Mecânica do Centro Técnico Científico da PUC-Rio.

Orientador: Prof. Ivan Fabio Mota de Menezes

Rio de Janeiro
Setembro de 2020



Monique Gomes de Araujo

**Previsão da curva de produção para projeto
exploratório utilizando redes neurais artificiais**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Mecânica do Departamento de Engenharia Mecânica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo:

Prof. Ivan Fabio Mota de Menezes

Orientador

Departamento de Engenharia Mecânica – PUC-RIO

Prof. Marcos Vitor Barbosa Machado

Departamento de Engenharia Mecânica – PUC-RIO

Dr. Rafael de Oliveira Valle dos Santos

Petróleo Brasileiro S.A. - PETROBRAS

Rio de Janeiro, 14 de Setembro de 2020

Todos os direitos reservados. A reprodução, total ou parcial, do trabalho é proibida sem autorização do autor, do orientador e da universidade.

Monique Gomes de Araujo

Graduou-se em Engenharia de Produção pela Universidade Federal do Rio Grande do Norte (Natal/RN) em 2010. Durante o período da faculdade, estudou por um ano Engenharia Industrial na *École de Mines* de Nancy (França) através de um programa da CAPES/BRAFITEC. Em 2010, ingressou na PETROBRAS onde trabalha até momento exercendo suas atividades na área de Avaliação Econômica na Exploração.

Ficha Catalográfica

Araujo, Monique Gomes de

Previsão da curva de produção para projeto exploratório utilizando redes neurais artificiais / Monique Gomes de Araujo ; orientador: Ivan Fabio Mota de Menezes. – 2020.

88 f. : il. color. ; 30 cm

Dissertação (mestrado)—Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Mecânica, 2020.

Inclui bibliografia

1. Engenharia Mecânica – Teses. 2. Reservatórios de petróleo. 3. Exploração. 4. Curva de produção. 5. Redes neurais artificiais. 6. Simulador numérico. I. Menezes, Ivan Fabio Mota de. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Mecânica. III. Título.

CDD: 621

Agradecimentos

Em primeiro lugar a Deus por me dar forças e sabedoria para enfrentar os desafios e alcançar os meus objetivos.

À minha família que sempre me apoiou e me fez acreditar que eu seria capaz. Em especial ao meu noivo Fabrício, que sempre esteve ao meu lado, até mesmo nos finais de semana e nas madrugadas de estudo. A sua parceria me fez ganhar forças para conseguir realizar esse sonho.

Ao professor Ivan pelas orientações e por confiar no meu trabalho.

Ao doutorando Renan Finotti pelos ensinamentos em redes neurais que foram imprescindíveis para a construção de todo o conhecimento adquirido durante o trabalho. Sua ajuda, dedicação e disponibilidade foram primordiais para a condução e conclusão deste estudo.

Aos colegas da Petrobras, em especial, pelos ensinamentos em engenharia de reservatório e na utilização do simulador numérico.

Aos amigos, pelo carinho e momentos de descontração durante todo o período do mestrado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Araújo, Monique Gomes de; Menezes, Ivan Fabio Mota de. **Previsão da curva de produção para projeto exploratório utilizando redes neurais artificiais**. Rio de Janeiro, 2020. 88p. Dissertação de Mestrado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

A estimativa de produção de petróleo é um dos parâmetros essenciais para mensurar a economicidade de um campo e, para tanto, existem várias técnicas convencionais na área da engenharia de petróleo para prever esse cálculo. Essas técnicas abrangem desde modelos analíticos simplificados até simulações numéricas mais complexas. Este trabalho propõe o uso de Redes Neurais Artificiais (RNA) para prever uma curva de produção de óleo que mais se aproxime da obtida por um simulador numérico. A metodologia consiste na utilização da rede neural do tipo *feedforward* para a previsão da vazão inicial e da curva de produção ao longo de dez anos para um poço produtor de óleo. Essa metodologia tem aplicação prática na área da exploração, visto que, nessa fase, ainda há muita incerteza sobre a acumulação de petróleo e, portanto, os modelos de reservatório tendem a não ser complexos. Os resultados foram obtidos a partir do treinamento de RNAs com dados coletados do simulador numérico IMEX, cujas saídas foram posteriormente comparadas com os dados originais da simulação numérica. Foi possível obter uma precisão de 97% na estimativa da vazão inicial do poço produtor de óleo. A previsão da curva de produção apresentou um erro percentual médio absoluto inferior a 10% nos dois primeiros anos. Apesar dos valores de erro terem crescido ao longo dos últimos anos, eles são menores quando comparados com a metodologia de declínio exponencial e com a regressão linear múltipla.

Palavras-chave

Reservatórios de Petróleo. Exploração. Curva de Produção. Redes Neurais Artificiais. Simulador Numérico.

Abstract

Araujo, Monique Gomes de; Menezes, Ivan Fabio Mota de (Advisor). **Production Forecast for Exploratory Project Using Artificial Neural Networks**. Rio de Janeiro, 2020. 88p. Dissertação de Mestrado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

Production forecasting is one of the essential parameters to measure the economics of an oil field. There are several conventional techniques in petroleum engineering to estimate the production curve. They range from simplified analytical models to complex numerical simulations. This study proposes the use of Artificial Neural Networks (ANN) to predict an oil production curve that approximates to a numerical simulator curve. The methodology consists of using a feedforward neural network to predict the initial flow and the production forecast over ten years of an oil well. This methodology has practical application in the exploration area, since, at this stage, there is still much uncertainty about the oil accumulation, so the reservoir models tend not to be complex. The results were obtained from the ANN training with data collected from the numerical simulator IMEX, whose outputs were later compared with the original data of the numerical simulation. It was possible to get an estimate for the oil initial flow forecast with an accuracy of 97%. The production forecast had a mean absolute percentage error of less than 10% in the first two years. Despite the increasing error values over the years, they are smaller when compared to those obtained from the exponential decline and multiple linear regression.

Keywords

Oil Reservoirs. Exploration. Production Forecast. Artificial Neural Networks. Numerical Simulator.

Sumário

1. Introdução.....	14
2. Conceitos Teóricos Fundamentais	18
2.1. Engenharia de Reservatório	18
2.2. Rede Neural Artificial.....	23
2.2.1. Rede Neural <i>Feedforward</i>	26
2.2.2. Algoritmo de treino das RNA's.....	29
2.3. Divisão do Banco de Dados.....	30
2.4. Normalização do Banco de Dados.....	33
2.5. Métricas Estatísticas	34
3. Metodologia	37
3.1. Método Proposto.....	37
3.2. Estrutura do Estudo.....	40
4. Resultados Numéricos Experimentais.....	42
4.1. Definição dos Parâmetros de Entrada.....	42
4.2. Banco de Dados	45
4.3. Treinamento	49
4.4. Definição dos Parâmetros da RNA.....	53
4.4.1. Caso UR.....	54
4.4.2. Caso MR.....	59
4.5. Resultados Finais	71
5. Conclusão.....	80
Referências Bibliográficas.....	83

Lista de Figuras

Figura 2.1. Fluxo radial em um reservatório.	20
Figura 2.2. Representação de uma acumulação em modelo tridimensional no simulador numérico. Fonte: ERTEKIN et al. (2001).	22
Figura 2.3. Célula neural com a seqüência de propagação do sinal.	24
Figura 2.4. Analogia da estrutura de uma rede neural com a célula neural.....	24
Figura 2.5. Estrutura de uma rede neural <i>perceptron</i>	26
Figura 2.6. Esquemático de um caso de parada no treinamento devido ao crescimento do erro de validação.....	31
Figura 2.7. Ilustração de casos com <i>underfitting</i> e <i>overfitting</i>	31
Figura 2.8. Esquemático da metodologia <i>5-fold</i>	32
Figura 3.1. Sugestão de utilização da metodologia proposta neste trabalho.	38
Figura 3.2. Etapas para a condução do trabalho.	40
Figura 3.3. Foram coletados dados dos elementos por onde o poço foi perfurado e dos elementos adjacentes a ele.	41
Figura 4.1. Definição de elementos do poço e adjacentes.....	43
Figura 4.2. Perfil médio da curva de produção de óleo dos poços utilizados nessa análise	46
Figura 4.3. Ilustração de seis modelos de reservatório hipotéticos com os respectivos poços produtores (em vermelho) alocados.	46
Figura 4.4. Ilustração de seis modelos de reservatório hipotéticos com os respectivos poços produtores (em vermelho) e sua região adjacente (em verde).	47
Figura 4.5 Distribuição das vazões iniciais que compõem o banco de dados.	48
Figura 4.6 Distribuição dos fluidos que compõem o banco de dados.	48
Figura 4.7 Distribuição da permeabilidade média na região do poço produtor utilizado no banco de dados.	49
Figura 4.8 Representação esquemática da construção da rede para a previsão da vazão inicial e curva de produção no caso UR.	50

Figura 4.9 Representação esquemática da construção da rede para a previsão da vazão inicial no caso MR.....	50
Figura 4.10 Representação esquemática da construção das redes para a previsão da curva de produção no caso MR.	51
Figura 4.11 Representação da utilização de cada grupo na definição da RNA.....	52
Figura 4.12 Configuração da rede utilizada para avaliar qual a melhor RNA para a curva de produção do caso UR.	55
Figura 4.13. Configuração da rede utilizada para avaliar qual a melhor normalização e o melhor algoritmo de treinamento para a RNA da vazão inicial.	60
Figura 4.14 Configuração da rede utilizada para avaliar qual a melhor normalização e o melhor algoritmo de treinamento para as RNAs de geração da curva de produção no caso MR.....	65
Figura 4.15 Comparativo do MAPE entre o método do declínio exponencial e o método de RNA para MR.....	73
Figura 4.16 Comparação da produção acumulada [MMbbl] obtida de uma curva média dos poços usados para treinamento da RNA, com a produção acumulada estimada pela RNA em modo <i>closedloop</i> para esse grupo de treino.	77
Figura 4.17 Comparação da produção acumulada [MMbbl] obtida de uma curva média dos poços usados para testar a RNA, com a produção acumulada estimada pela RNA em modo <i>closedloop</i> para esse grupo de teste.....	78
Figura 4.18 Comparação entre o MAD [m ³ /d] e seu DP na visão <i>closedloop</i> da RNA da curva de produção.....	79

Lista de Tabelas

Tabela 4.1 Cenários simulados para definir a melhor topologia da RNA.....	54
Tabela 4.2 Comparação da <i>performance</i> da RNA entre dois tipos de normalização para o caso UR.	55
Tabela 4.3 Comparação do MAPE ano a ano entre os tipos de normalização para o grupo de treino e teste no caso UR.....	56
Tabela 4.4 Comparação do MAD [m^3/d] ano a ano entre os tipos de normalização para o grupo de treino e teste no caso UR.....	56
Tabela 4.5 Comparação do MAPEam para dois tipos de algoritmo de aprendizado para o caso UR.	56
Tabela 4.6 Comparação do MAPE ano a ano entre os tipos de algoritmo de aprendizado para o grupo de treino e teste no caso UR.	57
Tabela 4.7 Comparação do MAD [m^3/d] ano a ano entre os tipos de algoritmo de aprendizado para o grupo de treino e teste no caso UR.....	57
Tabela 4.8 Resultado do MAPEam para os cenários de variação no número de neurônios no caso UR.....	58
Tabela 4.9 MAPE ano a ano no melhor cenário simulado para o caso UR.	59
Tabela 4.10 DP do MAPE ano a ano no melhor cenário simulado para o caso UR. ..	59
Tabela 4.11 MAD [m^3/d] ano a ano no melhor cenário simulado para o caso UR.	59
Tabela 4.12 Comparação do MAPE da RNA da vazão inicial entre dois tipos de normalização.....	61
Tabela 4.13 Comparação do MAD [m^3/d] da RNA da vazão inicial entre dois tipos de normalização.....	61
Tabela 4.14 Comparação do MAPE da RNA da vazão inicial entre dois tipos de algoritmo de treino.	62
Tabela 4.15 Comparação do MAD [m^3/d] da RNA da vazão inicial entre dois tipos de algoritmo de treino.	62
Tabela 4.16 MAPE dos cenários simulados para definição da melhor topologia para a RNA da vazão inicial.....	63
Tabela 4.17 Comparação do desempenho da RNA entre dois tipos de normalização no caso MR para geração da curva de produção.	65

Tabela 4.18 Comparação do MAPE ano a ano entre dois tipos de normalização para o grupo de treino e teste no caso MR.	66
Tabela 4.19 Comparação do MAD [m ³ /d] ano a ano entre os tipos de normalização para o grupo de treino e teste no caso MR.	66
Tabela 4.20 Comparação do MAPEam para dois tipos de algoritmo de aprendizado no caso MR para geração da curva de produção.	67
Tabela 4.21 Comparação do MAPE ano a ano para dois tipos de algoritmo de aprendizado no caso MR.	67
Tabela 4.22 Comparação do MAD [m ³ /d] ano a ano para dois tipos de algoritmo de aprendizado no caso MR.	67
Tabela 4.23 Resultado do MAPEam para os cenários de variação no número de neurônios do caso MR para geração da curva de produção.	68
Tabela 4.24. Análise do MAPE ano a ano do cenário com 4 neurônios em cada camada oculta no caso MR.	69
Tabela 4.25 Comparação do MAD [m ³ /d] ano a ano do cenário com 4 neurônios em cada camada oculta no caso MR.	69
Tabela 4.26 MAPE ano a ano obtido na visão <i>closedloop</i> do caso MR para a curva de produção.	70
Tabela 4.27 DP do MAPE ano a ano obtido na visão <i>closedloop</i> do caso MR para a curva de produção.	70
Tabela 4.28 MAD ano a ano obtido na visão <i>closedloop</i> do caso MR para a curva de produção.	70
Tabela 4.29. MAPEam e DPam do caso MR (incluindo vazão inicial e curva de produção).	70
Tabela 4.30 . Comparativo do MAPEam e DPam entre os casos UR e MR.	71
Tabela 4.31 MAPE ao longo dos anos quando utilizando o método do declínio exponencial.	72
Tabela 4.32 DP do MAPE ao longo dos anos quando utilizando o método do declínio exponencial.	72
Tabela 4.33. Comparativo do MAPEam e DPam entre o método do declínio exponencial e o método de RNA para o caso MR.	74

Tabela 4.34. Comparação entre os resultados da RNA e regressão linear múltipla para a obtenção da vazão inicial.....	74
Tabela 4.35 MAPE ao longo dos anos quando utilizada a regressão linear múltipla..	75
Tabela 4.36 DP do MAPE ao longo dos anos quando utilizada a regressão linear múltipla.....	75
Tabela 4.37. Comparativo do MAPEam e DPam entre o método da regressão linear múltipla e o método de RNA para o caso MR.....	75

Lista de Abreviaturas

ANP - Agência Nacional do Petróleo

BFGS - Broyden-Fletcher-Goldfarb-Shanno

CMG - *Computer Modelling Group*

CAPEX - *Capital Expenditure*

DP - Desvio Padrão

DPam - Desvio Padrão Atualizado Médio

IA - Inteligência Artificial

LM - Levenberg-Marquardt

MAD - *Mean Absolute Deviation*

MAPE - *Mean Absolute Percentage Error*

MAPEam - MAPE Atualizado Médio

MLP - *Multilayer Perceptron*

MR - Múltiplas Redes

MSE - *Mean Square Error*

OPEX - *Operational Expenditure*

RNA - Neurais Artificiais

RNR - Rede Neural Recorrente

UR - Única Rede

VPL - Valor Presente Líquido

1. Introdução

A indústria de petróleo e gás natural desempenha um papel importante para a economia mundial e tem sido essencial para o desenvolvimento socioeconômico do Brasil. Com a descoberta do pré-sal, essa relevância no cenário nacional se tornou mais evidente.

Segundo Petrobras (2015), existem várias etapas exploratórias anteriores a efetiva produção do campo:

- 1) Aquisição de blocos exploratórios;
- 2) Estudos da bacia;
- 3) Perfuração;
- 4) Plano de avaliação;
- 5) Avaliação da descoberta.

Após a aquisição do bloco exploratório em leilões promovidos pela Agência Nacional do Petróleo (ANP), as empresas utilizam estudos, dados geológicos e geofísicos para determinar qual a região com maior probabilidade de ter petróleo. Em seguida, é definido a posição do primeiro poço (pioneiro). Após a sua perfuração, pode-se confirmar ou não a presença de petróleo ou gás natural naquela região. Em geral, após essa perfuração ainda será necessário perfurar mais poços para melhor delimitar a acumulação de petróleo ou gás natural. Dessa forma, o consórcio pode solicitar um tempo adicional à ANP para melhor avaliar a nova acumulação. Essa solicitação deve ser concretizada em um documento chamado plano de avaliação, que consiste em um programa de trabalho com investimentos necessários à avaliação de uma descoberta. Após a aquisição dessas novas informações técnicas, a empresa deverá ter condições de considerar a acumulação como economicamente viável e, em seguida, efetuar a declaração de comercialidade. Posteriormente, o campo passa para a fase de desenvolvimento e novos investimentos serão necessários para a etapa da produção.

Segundo Souza & Silva (2016), a indústria do petróleo é caracterizada por ser tipicamente de capital intensivo, com altos riscos nas atividades de exploração, altos investimentos em desenvolvimento da produção, e longos prazos de maturação do investimento. Portanto, cada decisão, desde a aquisição de um bloco exploratório até o início da produção do campo, deve ser muito bem embasada em avaliações econômicas robustas. Obidike et al. (2014), afirma que a previsão da produção de um campo é um dos dados essenciais para determinar a economicidade do projeto e subsequentemente para tomadas de decisão na empresa. Segundo Oyekunle (2011), além da curva de produção, existem outros parâmetros importantes para a definição da economicidade de um campo de petróleo como a previsão de despesas (CAPEX e OPEX), modelo fiscal e previsão de preços de venda dos produtos (óleo e gás). Esse trabalho se concentrará na estimativa da curva de produção.

Existem soluções analíticas que podem ser usadas para estimar a curva de produção. Uma das equações de maior utilização em meios porosos e que embasa os métodos de previsão de curva de produção é a lei de Darcy, que estabelece que a vazão de um fluido é proporcional à área de fluxo, ao diferencial de pressão, e inversamente proporcional ao comprimento (Rosa et al., 2006). Nos dias de hoje, com o avanço computacional, é possível ter modelos mais complexos de geração de curva de produção, buscando sempre o máximo grau de semelhança com a realidade (Thomas, 2001). Esse avanço tem impulsionado pesquisas de desenvolvimento de simuladores numéricos de reservatório. Contudo, esses simuladores requerem um grande poder de processamento para obter resultados confiáveis em um tempo aceitável (Kohler, 2013).

Em paralelo, há o crescimento em pesquisas na área da Inteligência Artificial (IA), que é um conceito que abrange diversas tecnologias entre elas o *machine learning*. Shaeffer et al. (2020) explica que esses algoritmos possuem a habilidade de compreender e aplicar o comportamento de um banco de dados com rapidez e eficiência. Um dos tipos de *machine learning* é a Rede Neural Artificial (RNA) cuja característica mais marcante é a semelhança da sua estruturação com o cérebro humano. A RNA é bastante flexível aos diversos tipos de problemas e dados, tendo um grande potencial de resolver desafios e problemas complexos. Sua aplicação não é recente, e Ali (1994) já

comentava sobre o poder das RNAs na indústria de petróleo e citava aplicações como reconhecimento de padrões sísmicos, previsões de permeabilidade, identificação de litofácies de arenito, diagnóstico de brocas, análise e melhoria da produção de poços de gás, previsão e otimização de desempenho de poço, caracterização integrada de reservatório e gerenciamento de portfólio. Modelos de IA não dependem muito de significado físico, possuindo a capacidade de lidar com problemas não lineares tais como de previsão, que são influenciados por vários fatores complexos. Por esse motivo, os modelos de IA têm sido aplicados com sucesso na área de reservatórios (Zang, 2018).

O objetivo deste trabalho é propor o uso de RNAs para a previsão da vazão inicial de produção de óleo por poço produtor e da sua respectiva curva de produção de óleo por um horizonte de 10 anos. Utilizou-se a biblioteca *Neural Network Toolbox* do MATLAB para o treinamento. E os dados foram coletados de simulações de reservatório geradas para projetos exploratórios através do simulador IMEX (CMG, 2015), desenvolvido pela empresa CMG (*Computer Modelling Group*). O IMEX é um simulador black-oil de três fases que leva em consideração a gravidade e efeitos capilares. O IMEX suporta configurações de *grid* cartesiana, cilíndrica e com profundidade e espessura variáveis. Desta forma é possível mapear as configurações de campos em 2D e 3D (Almeida, 2007).

A metodologia do trabalho seguiu algumas etapas que consistiram desde a definição dos parâmetros que seriam utilizados como dados de entrada para o treinamento da RNA, passando pela coleta de dados no IMEX, treinamento da RNA no MATLAB e terminando com a estimativa da curva de produção do poço produtor de óleo. Os parâmetros de entrada foram: permeabilidade, volume de óleo, propriedades do fluido e pressão. Essas informações foram coletadas na região onde os poços produtores de óleo foram localizados e se referem às condições iniciais da simulação. As duas saídas alvos (*targets*) da RNA são: vazão inicial de óleo por poço produtor e a curva de produção de óleo por um horizonte de 10 anos. A partir desses dados de entrada, foram montados diversos cenários de RNA, alterando os critérios de normalização, o algoritmo de aprendizado, número de neurônios nas camadas ocultas e até mesmo a

forma como representar a rede. Ao final, buscou-se obter a rede que apresentasse o melhor desempenho na previsão da vazão inicial e da curva de produção.

A pretensão do trabalho não é substituir os diversos modelos convencionais existentes. A proposta é apresentar uma nova metodologia de previsão de curva de produção de óleo que apresente vantagens em relação às atuais. O ganho dessa técnica estaria em ter a uma resposta extremamente próxima a de um simulador numérico e levar o mesmo tempo e esforço computacional de um modelo analítico.

A estrutura do trabalho está organizada em 5 capítulos:

- O primeiro é a introdução;
- No segundo capítulo será dado um embasamento teórico sobre engenharia de reservatório e redes neurais artificiais. Serão descritos alguns parâmetros de reservatórios utilizados como base para o treinamento da rede neural. Em seguida, será apresentado o conteúdo teórico de um modelo analítico e um numérico de geração de curva de produção. Por fim, será detalhada a caracterização de uma rede neural artificial e alguns métodos estatísticos para avaliar o desempenho da rede;
- No capítulo 3 será descrita a metodologia do trabalho, começando com o detalhamento do método proposto e como o estudo foi estruturado;
- O quarto capítulo (“Resultados Numéricos Experimentais”) discutirá os principais resultados obtidos que servirão como base na escolha da rede com melhor desempenho para o problema estudado neste trabalho;
- Finalmente, o quinto capítulo (“Conclusões”) apresentará as conclusões do trabalho e apontará sugestões de desenvolvimento em trabalhos futuros.

2. Conceitos Teóricos Fundamentais

2.1. Engenharia de Reservatório

Conforme descrito por Silva (2006), antes da perfuração do primeiro poço (pioneiro), as informações disponíveis a respeito do reservatório são originárias de métodos indiretos. Desta maneira, ainda não há comprovação da existência de uma acumulação de petróleo, havendo apenas um conjunto de evidências na região pesquisada, comparando-se seus dados sísmicos e resultados de reservatórios localizados nas proximidades, aos quais se acredita que possuam características similares às do reservatório que está sendo estudado.

À medida que vai se adquirindo novas informações, elas vão sendo incorporadas aos estudos sobre aquela acumulação. E em etapas futuras, já é possível ter um modelo de reservatório mais complexo. A complexidade desses modelos está diretamente associada à quantidade de detalhes que se deseja considerar e com a quantidade de informações a serem incorporadas.

Dentre as diversas informações a serem obtidas a respeito de uma acumulação de petróleo após sua descoberta, a existência e a quantidade de hidrocarbonetos que se pode retirar dessa jazida e o tempo em que essa produção e vazão se efetuarão são, sem qualquer dúvida, dos mais importantes. O conhecimento das leis e equações que regem o movimento dos fluidos nos meios porosos é fundamental para a obtenção dessas informações (Rosa et al., 2006).

Em uma abordagem volumétrica e simplificada (quando não há dados dinâmicos para confirmar) são necessárias as seguintes informações para o cálculo do volume (Silva, 2006):

- Volume total de rocha (Vr): obtida por meio de sísmica de reflexão;
- Porosidade (Φ) – obtida por interpretação de perfis ou em ensaios de laboratório;
- Saturação de água (S_w) – obtida por interpretação de perfis ou em ensaios de laboratório;
- Fator volume de formação do fluido (B_o) - obtido por meio de análise feita em laboratório. Serve para a conversão do volume de óleo no reservatório para as condições de superfície.

Assim, o volume do fluido (N) é calculado por

$$N = \frac{Vr \Phi (1 - S_w)}{B_o}. \quad (2.1)$$

Segundo Thomas (2001), mesmo que uma rocha contenha uma quantidade apreciável de poros e dentro desses poros existam hidrocarbonetos em quantidade razoável, não há a garantia de que eles possam ser extraídos. Para que isso ocorra, é necessário que a rocha permita o fluxo de fluidos através dela. E essa medida é chamada de permeabilidade (K). Um outro fator que influencia no escoamento é a viscosidade do fluido (μ) que pode ser definida como a resistência de um fluido ao escoamento para uma dada temperatura.

A equação de deslocamento de fluidos em meios porosos foi formulada por Henry D'Arcy (1803-1858). Ele afirmou que a vazão Q de um fluido em um meio poroso é proporcional à permeabilidade K , à diferença de pressão ($P_A - P_B$), à área transversal A e inversamente proporcional à viscosidade do fluido μ e a distância L por onde ele é percorrido. Portanto, em um meio poroso horizontal e plano, a vazão Q é dada por

$$Q = \frac{KA}{\mu L} (P_A - P_B). \quad (2.2)$$

O movimento dos fluidos do reservatório para o poço é melhor caracterizado pelo fluxo radial (Rosa et al., 2006). Considerando um sistema homogêneo e isotrópico, as linhas de fluxo convergem para o eixo central do poço, conforme ilustrado na Figura 2.1.

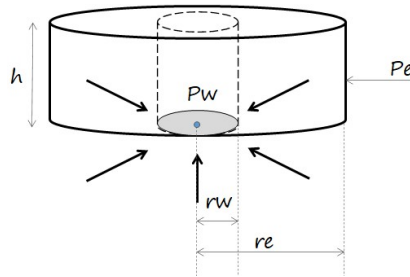


Figura 2.1. Fluxo radial em um reservatório.

Na Figura 2.1, P_e representa a pressão externa, P_w a pressão interna, R_e o raio externo, R_w o raio interno e h a espessura do reservatório.

Considerando um fluxo isotérmico, radial, horizontal, fluido incompressível e regime permanente, a expressão da vazão em superfície se reduz a

$$Q = \frac{2\pi kh}{B_o \mu \ln \left(\frac{r_e}{r_w} \right)} (P_e - P_w). \quad (2.3)$$

Conforme descrito por Thomas (2001), o fluxo radial apresentado na Equação (2.3) é o que melhor caracteriza o movimento dos fluidos do reservatório para o poço na maioria das situações. Contudo, possui algumas premissas básicas:

- O reservatório tem espessura constante;
- É considerado homogêneo em toda a sua extensão, com relação a todas as propriedades da rocha;
- É isotrópico em relação à permeabilidade;
- Está saturado com um único fluido;
- Poço é completado em todo o intervalo produtor para assegurar o fluxo radial em toda a espessura do reservatório.

A previsão do comportamento do reservatório ao longo dos anos também é um grande desafio. E para isso, há vários métodos de previsão. Abaixo serão descritos dois: análise de declínio e a simulação numérica de reservatório.

Análise de Declínio

A análise de curvas de declínio, conforme descrito por Rosa et al. (2006), é um método bastante simplificado para ajuste de histórico, previsão do comportamento de poços isolados, de reservatórios ou mesmo de campos de óleo. Esse método é útil nas seguintes situações:

- Realização de estudos preliminares ou de rotina, pois sua aplicação demanda menos tempo que outros métodos;
- Realização de ajuste histórico quando não há dados suficientes para utilização de outros métodos analíticos;
- Estimativa de comportamento (produção e recuperação) quando há pouco ou nenhum histórico de produção, embora muitas vezes seja empregado também quando está disponível um histórico razoável do comportamento do poço, reservatório ou campo.

Segundo Arps (1945), a equação que define o comportamento da curva de produção é definida por

$$q_t = q_i(1 + bD_it)^{-\frac{1}{b}}, \quad (2.4)$$

onde q_t é a vazão de produção no tempo t , q_i é a vazão inicial da produção, b é o expoente de declínio e D_i é a taxa de declínio.

No método de declínio não são utilizadas informações sobre as propriedades da rocha reservatório, sobre o comportamento dos fluidos ou sobre as relações rocha-fluido. Também, não são levadas em consideração as leis de fluxo e o mecanismo responsável

pela produção do reservatório (Marsili, 2008). Os tipos de declínio podem ser classificados em: exponencial, harmônico e hiperbólico. No caso específico do declínio exponencial, um dos mais usados na indústria de petróleo, a taxa de declínio é considerada constante e pode ser expresso por

$$q_t = q_i e^{-D_i t} . \quad (2.5)$$

Simulação de Reservatório

Na engenharia de petróleo, uma das tarefas mais importantes é a previsão da produção de hidrocarbonetos de reservatórios. Essa tarefa permite quantificar reservas, avaliar e priorizar projetos de exploração e dimensionar sistemas de produção para otimizar a produção de um reservatório. Nesse sentido, a simulação computacional de reservatórios de petróleo vem se tornando uma importante ferramenta para a indústria de exploração de petróleo e gás natural (Kohler, 2013). Na simulação numérica, o reservatório é subdividido em células (ver Figura 2.2) com propriedades diferentes e envolve a solução simultânea de um grande número de equações que representam o fluxo de fluidos no meio poroso para cada uma destas células. Um simulador de reservatórios leva em consideração também a localização dos poços de produção e injeção e as suas condições operacionais (Marsili, 2008).

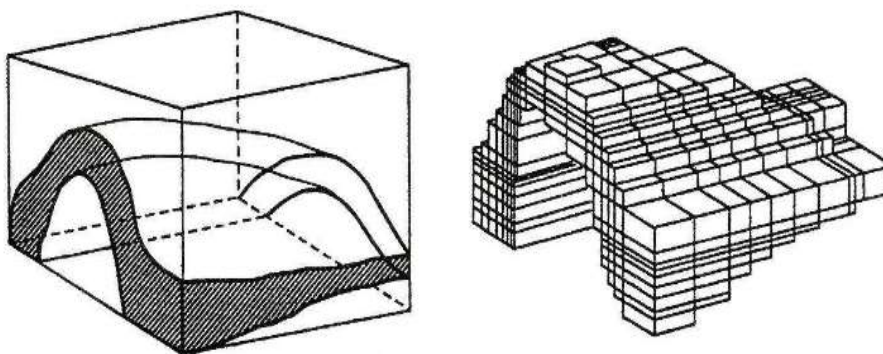


Figura 2.2. Representação de uma acumulação em modelo tridimensional no simulador numérico. Fonte: ERTEKIN et al. (2001).

Este processo geralmente exige um alto custo computacional, uma vez que os métodos numéricos utilizados (elementos finitos, volumes finitos ou diferenças finitas) conduzem a sistemas algébricos, não lineares, de grande porte para resolver um conjunto de equações diferenciais que modelam as transferências de fluidos dentro do reservatório simulado (Tupác, 2005).

Os simuladores numéricos permitem bons resultados nos estudos dos reservatórios e estes vão ficando cada vez mais precisos a cada refinamento do modelo, onde novos e melhores dados da rocha, dos fluidos, da geologia, do histórico de produção vão tornando o modelo cada vez mais próximo das características do reservatório (Araujo Junior, 2010).

2.2. Rede Neural Artificial

Redes Neurais Artificiais (RNA) são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neuronal de organismos inteligentes e que adquirem conhecimento através da experiência. O cérebro é formado por uma rede com uma grande quantidade de processadores simples, chamados de neurônios (ver Figura 2.3), interconectados por estruturas ramificadas, no caso os axônios e os dendritos. As sinapses fazem a ligação dos axônios e dendritos entre os neurônios (Silva, 2006).

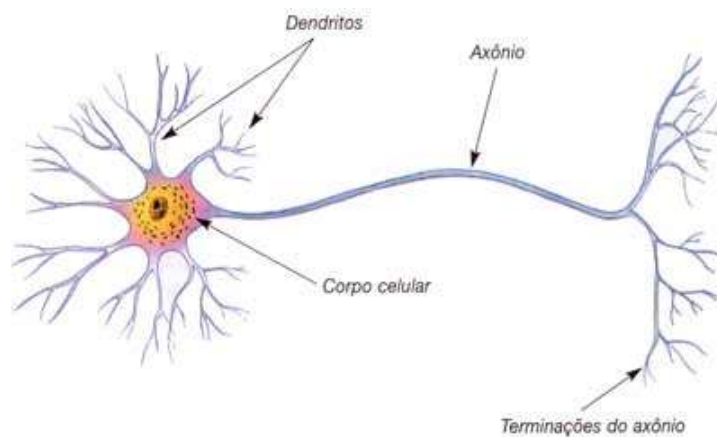


Figura 2.3. Célula neural com a sequência de propagação do sinal.

Os elementos de processamento em uma RNA são os neurônios, que recebem os sinais dos dados de entrada, processam e geram um sinal de saída. Por meio das interconexões, esta saída pode ser um sinal de entrada para outros neurônios.

Uma analogia entre a estrutura da RNA e uma célula neural, é mostrada na Figura 2.4. Os dados de entrada (x_1, x_2, x_3) são ponderados por pesos (w_1, w_2, w_3), somados a um *bias* (b) e ligados a uma função de ativação. Em seguida, essa resposta é conectada ao neurônio posterior onde esse pode ser o produto final ou a entrada para outro neurônio.

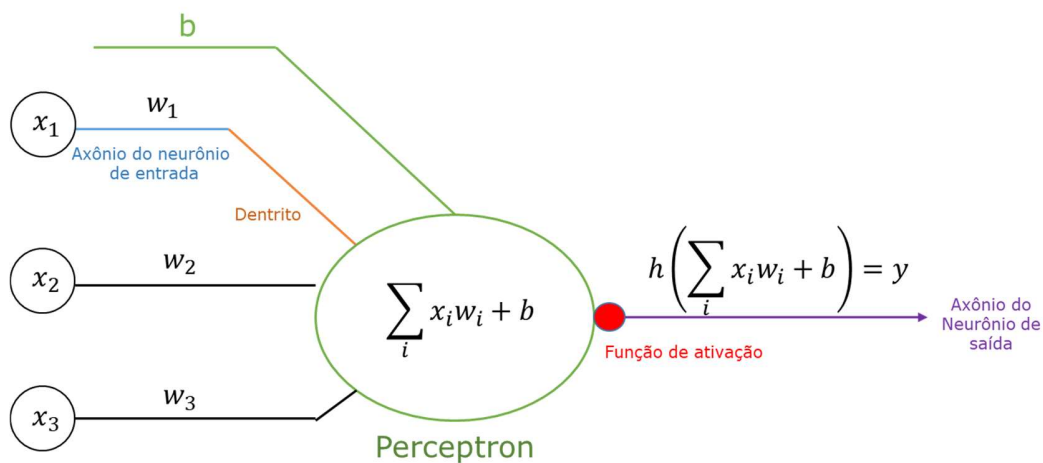


Figura 2.4. Analogia da estrutura de uma rede neural com a célula neural.

O *bias* (*b*) tem a função de aumentar ou diminuir a influência do valor da entrada para a ativação do neurônio. Já a função de ativação, determina a amplitude em que um sinal deve progredir na rede para se chegar no dado de saída.

Os neurônios em uma rede estão organizados em níveis que se interligam formando uma espécie de grafo. A rede é composta por:

- Camada de entrada: onde são informados os dados de entrada;
- Camadas ocultas: onde as conexões são ponderadas e a maior parte do processamento é realizado. Essa camada tem por finalidade extrair as informações associadas ao sistema;
- Camada de saída: onde é apresentado o resultado final.

O número de camadas ocultas e a quantidade de neurônios por camada caracterizam a topologia da rede neural. A escolha da topologia para cada problema não é uma tarefa simples pois não existem regras para essa definição. Muitas vezes ela é baseada em tentativa e erro, ou com base em trabalhos anteriores.

Segundo Rebeschini et. al. (2013), RNA tem uma propensão natural para armazenar conhecimento experimental e disponibilizá-lo para uso. A base de dados usada para o treinamento da rede é uma das etapas mais importantes para que ela tenha o desempenho esperado. Por isso, é fundamental analisar quais serão os dados de entrada e dados de saída esperados, e garantir que todos eles estejam bem estruturados.

A principal tarefa das RNAs é definir pesos e *biases* a partir dos dados de entrada e reproduzir um dado de saída igual ou simular ao que foi usado para treinamento. Dessa forma, quando uma amostra futura for apresentada à RNA treinada, ela reproduza uma saída próxima ao desejado. Para que esse objetivo seja atingido, existem vários tipos de redes, entre elas estão as redes neurais *feedforward*. De maneira geral, essas redes não têm conexões de retorno, ou seja, a saída dos neurônios de uma camada serve de entrada para as demais camadas, seguindo uma sequência até a saída. Essa rede é capaz de modelar problemas em que os dados de entrada não têm relação temporal com os dados de saída.

2.2.1. Rede Neural *Feedforward*

Dentre as redes do tipo *feedforward* as mais comuns são:

- Única camada (*perceptron*): Conforme ilustrado na Figura 2.5, consiste nos dados de entrada e uma camada de saída. Em geral, os neurônios de entrada passam por uma combinação linear usando pesos.

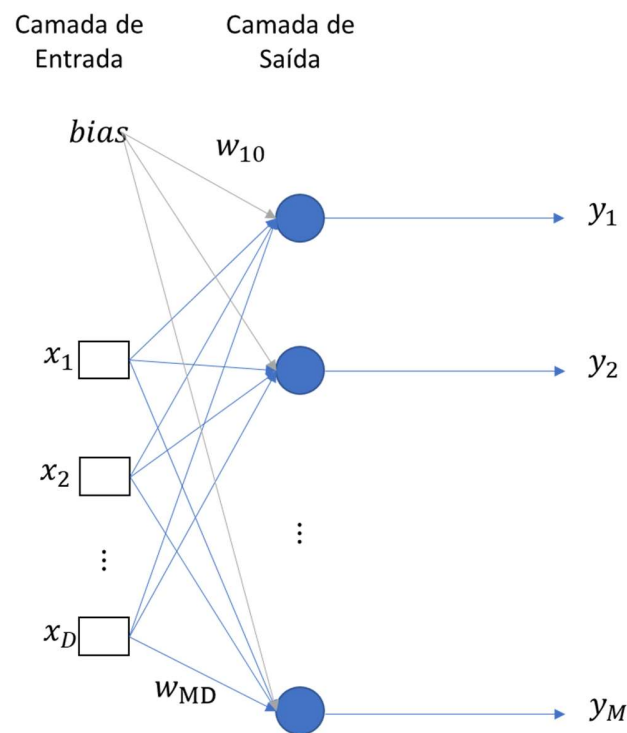


Figura 2.5. Estrutura de uma rede neural *perceptron*.

Na Figura 2.5 acima, temos uma rede formada pelos dados de entrada x_1, \dots, x_D e, a partir deles, são feitas combinações lineares conforme mostrada na Equação (2.6) (Bishop, 2006)

$$a_j = \sum_{i=1}^D w_{ji} x_i + w_{j0} , \quad (2.6)$$

onde D é o número total de dados de entrada, $j = 1, \dots, M$, com M sendo o número total de saídas, w_{ji} representa o peso que liga o dado de entrada x_i ao neurônio posterior a_j e w_{j0} representa o peso que liga o *bias* ao neurônio a_j . Cada a_j é transformado por meio de uma função de ativação $h(\cdot)$ dada por

$$y_j = h(a_j). \quad (2.7)$$

A seguir serão mostradas as principais funções de ativação usadas em aplicações de aprendizado de máquina. Uma delas é a função linear, geralmente aplicada na camada de saída, em que a saída do neurônio simplesmente repete o sinal que entra

$$h(a) = a. \quad (2.8)$$

Um outro tipo é a função de ativação logística, a qual possui uma faixa de valores do sinal de saída variando de 0 e 1, e escrita como

$$h(a) = \frac{1}{1 + e^{-a}}. \quad (2.9)$$

Por fim, uma outra função de ativação bastante utilizada é a tangente hiperbólica. Ela preserva a forma sigmoideal da função logística, mas assume valores que podem variar entre -1 a 1

$$h(a) = \tanh(a), \quad (2.10)$$

onde

$$\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}. \quad (2.11)$$

- **Múltiplas Camadas (*Multilayer Perceptron*-MLP):** esse tipo de rede possui uma ou mais camadas ocultas. A saída de uma camada oculta é utilizada na próxima camada seguindo assim até a camada de saída. Em geral, o treinamento

mais comum dessa rede utiliza um algoritmo chamado retro-propagação do erro (*backpropagation*).

Basicamente, esse processo de retro-propagação do erro é constituído de duas fases (Silva, 1998):

- 1) Etapa de propagação do sinal (*feedforward*): o efeito dos dados de entrada segue até a camada de saída, definindo um peso fixo para as conexões;
- 2) Etapa de retro-propagação do erro (*backpropagation*): Nessa etapa é calculado um erro com base na resposta da rede com a saída desejada. O sinal do erro é propagado da saída para a entrada e os pesos das conexões são ajustados com o objetivo de reduzir o erro.

Fazendo analogia com a Equação (2.6), pode-se assumir que o *bias* será absorvido pela matriz de pesos \mathbf{W} por meio da adição de uma variável de entrada x_0 cujo valor será fixado em $x_0 = 1$. Considerando uma MLP com L camadas, onde $M^{(l)}$ representa o número de neurônios da camada l , então o neurônio de saída da camada l é dado por

$$a_j^{(l+1)} = \sum_{i=0}^{M^{(l)}} w_{ji}^{(l)} x_i^{(l)}, \quad (2.12)$$

onde $x_i^{(l)}$, para $i = 0, 1, \dots, M^{(l)}$, são os neurônios de entrada da l -ésima camada e $w_{ji}^{(l)}$, para $j = 1, 2, \dots, M^{(l+1)}$, são os pesos associados que conectam o neurônio $x_i^{(l)}$ com $a_j^{(l+1)}$. É importante mencionar que $w_{j0}^{(l)}$ associa o *bias* ($x_0^{(l)} = 1$) da camada l ao neurônio $a_j^{(l+1)}$. Portanto, o neurônio de entrada para a próxima camada será expresso por

$$x_j^{(l+1)} = h(a_j^{(l+1)}), \quad (2.13)$$

sendo o $h(\cdot)$ a função de ativação. Onde a saída da RNA é calculada usando os neurônios da última camada

$$y_j = h(a_j^{(L)}). \quad (2.14)$$

2.2.2. Algoritmo de treino das RNA's

O processo de aprendizado de uma rede neural utiliza algoritmos de otimização, para ajustar os parâmetros de pesos e *biases* da rede visando a minimização de uma dada função objetivo. Seja o vetor \mathbf{x} representando as variáveis de entrada, o vetor \mathbf{w} composto pelos pesos da RNA, o vetor \mathbf{y} cujos elementos y_j , para $j = 1, \dots, M^{(L)}$, representam as variáveis de saída da RNA e \mathbf{z} o vetor cujos elementos z_j são as respostas desejadas ou saídas-alvo, uma das formas de escrever a função erro é (Bishop, 2006)

$$E(\mathbf{w}) = \frac{1}{2} \sum_{j=1}^{M^{(L)}} |y_j(\mathbf{x}, \mathbf{w}) - z_j|^2. \quad (2.15)$$

Essa função erro, através de um método de otimização, ajusta os pesos e *biases* dos neurônios da rede com o objetivo de aproximar cada vez mais a saída com a resposta desejada. Dessa forma, o algoritmo de otimização da rede busca minimizar a função objetivo descrita abaixo:

$$\min E(\mathbf{w}) = \frac{1}{2} \sum_{j=1}^{M^{(L)}} |y_j(\mathbf{x}, \mathbf{w}) - z_j|^2. \quad (2.16)$$

Há vários tipos de algoritmos de otimização e cada um tem características e desempenhos diferentes. Alguns dos algoritmos mais utilizados para RNA's são:

- BackPropagation;
- Levenberg-Marquardt (LM);
- Broyden-Fletcher-Goldfarb-Shanno (BFGS).

Em geral, esses algoritmos precisam de um ponto inicial e estabelecem um critério para escolha de uma direção de busca, um método de busca de mínimo nesta direção e um critério de parada. A principal diferença entre esses métodos (e de tantos outros baseados em derivadas) é a forma como essas direções são determinadas (Xavier, 2009).

2.3. Divisão do Banco de Dados

O objetivo do treinamento da RNA é obter um ótimo mapeamento de entrada-saída. Uma rede bem treinada significa que aprendeu o modelo suficientemente bem para ter boas estimativas no futuro (Túpac, 2005). Para que isso ocorra, o conjunto de dados é geralmente dividido em três grupos: treino, validação e teste.

O grupo de treino deve abranger todos os dados pertencentes ao domínio do problema e é usado na fase de treinamento para atualizar os pesos da rede. O grupo de validação é usado como possível critério de parada, examinando ou confirmando a precisão da melhor rede selecionada antes de implementá-la no sistema neural e/ou ser entregue ao usuário final. O grupo de teste é usado para verificar a resposta da rede com relação a dados não treinados. (Basheer, 2000).

Para cada ciclo de treinamento, denominado época de treino, os erros do grupo de treino e validação são monitorados. Conforme ilustrado na Figura 2.6, tais erros tendem a diminuir com o processo de treinamento. Contudo, se em algum momento o erro do grupo de validação começar a crescer de forma consistente, o treinamento da rede é interrompido (Kohler, 2013). Isso significa que a rede estava perdendo a sua habilidade de generalização e estava se ajustando de forma excessiva ao grupo de treino.

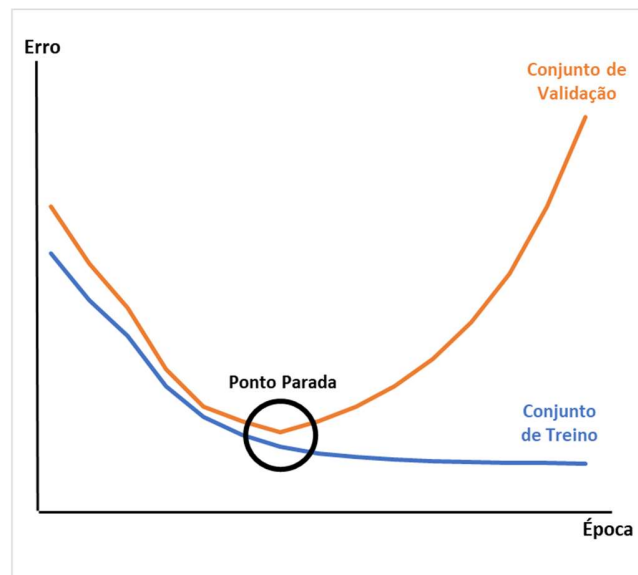


Figura 2.6. Esquemático de um caso de parada no treinamento devido ao crescimento do erro de validação.

Um bom modelo não pode sofrer de *underfitting* e nem de *overfitting*. No primeiro caso, a rede não consegue encontrar relação entre as variáveis do modelo e não obtém um bom desempenho na etapa de treinamento. Já no segundo caso, a rede consegue uma relação ótima apenas para os dados de treino, com isso, ela não conseguirá fazer uma boa estimativa quando um novo dado for acrescentado. A Figura 2.7 demonstra esses dois casos e compara com um exemplo de *fitting* apropriado para o modelo.

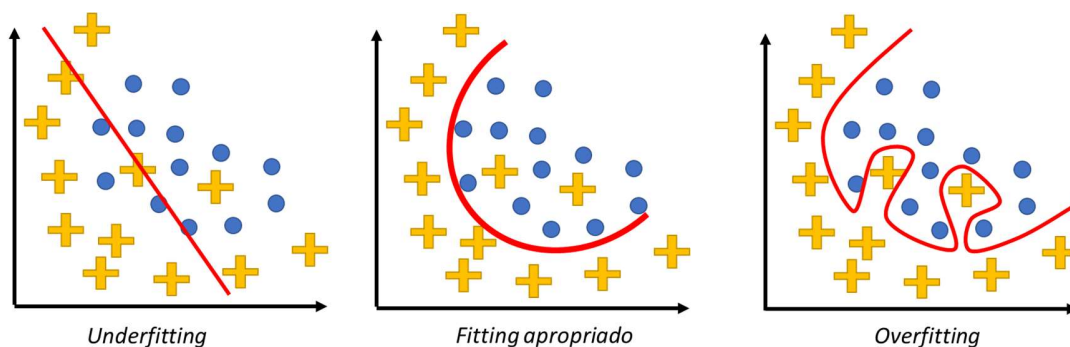


Figura 2.7. Ilustração de casos com *underfitting* e *overfitting*.

Uma maneira de avaliar a generalização de uma RNA, especialmente quando o banco de dados é pequeno, é utilizando o método de validação cruzada, denominado *K-fold*.

Ela consiste em dividir aleatoriamente o banco de dados em k grupos mutualmente excludentes de tamanhos aproximadamente iguais. Um desses grupos é utilizado para teste e o $k-1$ grupos restantes utilizados para treinar e validar a rede. Esse processo de treino e teste é repetido k vezes. A acurácia do modelo é medida a partir da média dos erros de todas as k rodadas (Kohavi, 1995). Na Figura 2.8, há um exemplo do 5-fold, em que os retângulos em azul claro representam o grupo que irá testar a RNA resultante em cada rodada.

TODAS OS DADOS					
Treino					Teste
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
Fold 1	Treino	Treino	Treino	Treino	Teste
Fold 2	Treino	Treino	Treino	Teste	Teste
Fold 3	Treino	Treino	Teste	Treino	Teste
Fold 4	Treino	Teste	Treino	Treino	Teste
Fold 5	Teste	Treino	Treino	Treino	Teste

Figura 2.8. Esquemático da metodologia 5-fold.

O conjunto de treino é dividido adicionalmente em dois subconjuntos disjuntos: subconjunto de treinamento e subconjunto de validação (Haykin, 2001). Com isto, é possível verificar se o modelo, com um conjunto de dados diferente daquele usado para estimar os parâmetros da rede, responde corretamente aos padrões apresentados (Kohler, 2013). Para cada *fold* (conforme descrito na Figura 2.8), são calculados os erros da saída e avaliada a robustez da rede treinada. Essa metodologia garante que

todo o banco de dados seja, em alguma rodada, considerado um dado desconhecido à rede e utilizado para testá-la. Dessa forma, podemos avaliar a generalização da rede e evitar o *overfitting*. O erro final após a etapa de treinamento é obtido pela média do erro de todos os *folds*.

2.4. Normalização do Banco de Dados

Segundo Basheer (2000), a normalização do banco de dados é um processo essencial para que não haja números com ordens de grandeza muito distintas. Dessa forma, evita-se que grandes números se sobreponham aos pequenos e o algoritmo fique prejudicado. Não há um procedimento único para normalização dos dados de entrada e saída. Abaixo serão mostrados dois tipos de normalização.

- Normalização por desvio-padrão,

$$n_s = \frac{x_s - \mu}{\sigma}, \quad (2.17)$$

onde \mathbf{X} é o vetor cujos elementos x_s , para $s = 1, \dots, S$, representam o conjunto das amostras de uma determinada variável de entrada, μ representa a média do vetor \mathbf{X} e σ o desvio padrão de \mathbf{X} , n_s os seus respectivos valores normalizados.

- Normalização Min-Max,

$$n_s = \frac{x_s - X_{min}}{X_{max} - X_{min}}, \quad (2.18)$$

onde \mathbf{X} é o vetor cujos elementos x_s , para $s = 1, \dots, S$, representam o conjunto das amostras de uma determinada variável de entrada, X_{min} é o menor valor de \mathbf{X} , X_{max} o maior valor de \mathbf{X} , n_s os seus respectivos valores normalizados.

2.5. Métricas Estatísticas

As medições dos erros estatísticos são um dos pontos mais importantes para monitorar a precisão do modelo que está sendo proposto. Algumas métricas de erro são:

- MSE

Normalmente esse é o indicador de erro usado durante o treinamento da RNA. O Erro Quadrático Médio (*Mean Square Error* - MSE) começa alto no início e tende a reduzir ao longo do processo de treinamento. Essa é uma medida da média do quadrado do desvio entre as saídas \mathbf{y} da RNA e os valores \mathbf{z} fornecidos pelo simulador, ou seja

$$MSE = \frac{1}{M^{(L)}} \left(\frac{1}{2} \sum_{j=1}^{M^{(L)}} (y_j - z_j)^2 \right), \quad (2.19)$$

onde $M^{(L)}$ é o total de saídas da rede. A Equação (2.19) é aplicada a cada uma das amostras dos *folds*, e, portanto, o MSE final será um valor médio entre eles.

- MAPE

Uma outra forma de avaliar os resultados da rede é por meio do erro percentual absoluto médio (*Mean Absolute Percentage Error* - MAPE), o qual fornece o erro médio em termos percentuais, independentemente de ser positivo ou negativo. O fato de ser expresso em percentual, torna fácil a sua interpretação. Esse indicador é escrito como

$$MAPE = \frac{1}{M^{(L)}} \sum_{j=1}^{M^{(L)}} \left| \frac{(y_j - z_j)}{z_j} \right|, \quad (2.20)$$

onde $M^{(L)}$ é o total de saídas da rede. O MAPE final é calculado como sendo a média do MAPE de todas as amostras dos *folds*.

- MAD

O Desvio Absoluto Médio (*Mean Absolute Deviation* - MAD) é utilizado como forma de medir o erro em unidades, da seguinte forma

$$MAD = \frac{1}{M^{(L)}} \sum_{j=1}^{M^{(L)}} |y_j - z_j|, \quad (2.21)$$

onde $M^{(L)}$ é o total de saídas da rede. O MAD final é calculado como sendo a média do MAD de todas as amostras dos *folds*

- DESVIO PADRÃO (DP)

O desvio padrão é uma medida do grau de dispersão de um conjunto de dados, ou seja

$$DP = \sqrt{\frac{\sum_{j=1}^{M^{(L)}} (y_j - \bar{y})^2}{M^L}}, \quad (2.22)$$

onde $M^{(L)}$ é o total de saídas da rede e \bar{y} é o valor médio de \mathbf{y} .

- MAPE ATUALIZADO MÉDIO (MAPEam)

Para a monetização da curva de produção, os valores iniciais têm um impacto bem maior do que os valores finais. Comercialmente, as avaliações econômicas dos projetos são referenciadas em valores presentes, que consiste em analisar um valor considerando o tempo. Portanto, este trabalho desenvolveu a métrica estatística MAPE atualizado médio (MAPEam), com o objetivo de facilitar a comparação entre as diversas abordagens apresentadas, permitindo analisar projetos sob a mesma perspectiva mesmo que estejam em períodos diferentes. Esse conceito é bastante utilizado em finanças no cálculo do valor presente líquido (VPL). O VPL é uma técnica de análise de fluxos de caixa que consiste em calcular o valor presente de uma série de

pagamentos (ou recebimentos) iguais ou diferentes a uma taxa conhecida, e deduzir deste o valor do fluxo inicial (valor do empréstimo, do financiamento ou do investimento)" (Vieira Sobrinho, 1997).

A atualização dos números em uma série temporal é importante quando os valores do início da série são mais representativos do que os últimos. Para tal, o processo de atualização consiste em aplicar uma taxa de desconto sobre as parcelas da série e trazê-las a uma visão de data zero, ou seja

$$MAPE_{am} = \frac{1}{T} \sum_{t=1}^T \frac{MAPE_t}{(1 + taxa)^t}, \quad (2.23)$$

onde $MAPE_{am}$ é o valor médio do MAPE atualizado, $MAPE_t$ é o MAPE no período t , T representa o número de períodos da série temporal e, por fim, $taxa$ é a taxa de desconto. Da forma que esse indicador foi idealizado, só é possível utilizá-lo para comparar duas redes que possuam o mesmo número de período.

- DESVIO PADRÃO ATUALIZADO MÉDIO (DPam)

Seguindo a mesma linha do $MAPE_{am}$, foi criado o desvio padrão atualizado médio. Sendo DP_t é o desvio padrão no período t , T o número de períodos da série temporal e, por fim, $taxa$ é a taxa de desconto, então o DPam pode ser definido como

$$DP_{am} = \frac{1}{T} \sum_{t=1}^T \frac{DP_t}{(1 + taxa)^t}. \quad (2.24)$$

3. Metodologia

3.1. Método Proposto

Como a produção de petróleo, principalmente *off-shore*, é bastante dispendiosa, faz-se necessário um cuidadoso estudo de sua viabilidade econômica. Para isso, torna-se necessário a obtenção de uma previsão confiável de produção esperada para o campo, o que possibilitará uma acurada avaliação do retorno dos investimentos (Magalhaes, 2007). A escolha entre utilizar um modelo complexo de simulação de reservatório ou um modelo simplificado analítico está diretamente ligado à disponibilidade de recursos de informática, ao tempo disponível e a quantidade de informações que se tem daquela acumulação.

Durante a fase exploratória do projeto, há muitas incertezas associadas aos parâmetros de entrada do modelo de simulação devido às heterogeneidades ainda não descobertas do reservatório. Dessa forma, modelos simplificados, porém robustos, de geração de curva podem agregar bastante valor à avaliação econômica do projeto. Até que uma acumulação de petróleo seja conhecida em detalhes, é necessário tempo e aquisição de informação. Antes da perfuração do poço pioneiro na fase exploratória, o conhecimento da acumulação fica restrito a estudos regionais e sísmicos e analogia com reservatórios já descobertos.

Nos modelos analíticos, o reservatório é tratado como se fosse um único bloco com propriedades médias e uniformes. Já nos modelos numéricos, existe a possibilidade de se subdividir o reservatório em células com propriedades diferentes, envolvendo assim, uma solução simultânea de um grande número de equações que representam o fluxo em meio poroso (Silva, 2006). Esse refinamento acontece à medida em que vão sendo obtidas mais informações sobre a acumulação.

A proposta desde trabalho é criar uma metodologia que, a partir das informações de rocha e fluido desse modelo discretizado, seja possível prever uma curva de produção para um poço produtor de óleo que seja próxima à curva que seria obtida caso esse mesmo modelo fosse levado para o simulador numérico IMEX (CMG, 2015). A metodologia aqui proposta pode ser bastante útil em estágios intermediários de conhecimento de um reservatório conforme é ilustrado na Figura 3.1. Nessa fase, já é possível construir um modelo discretizado do reservatório, mesmo que simplificado. A geração da curva de produção será feita a partir de RNAs treinadas. Dessa forma, a curva será estimada em menor tempo e com capacidade computacional inferior quando comparados aos necessários para a execução de um simulador numérico.

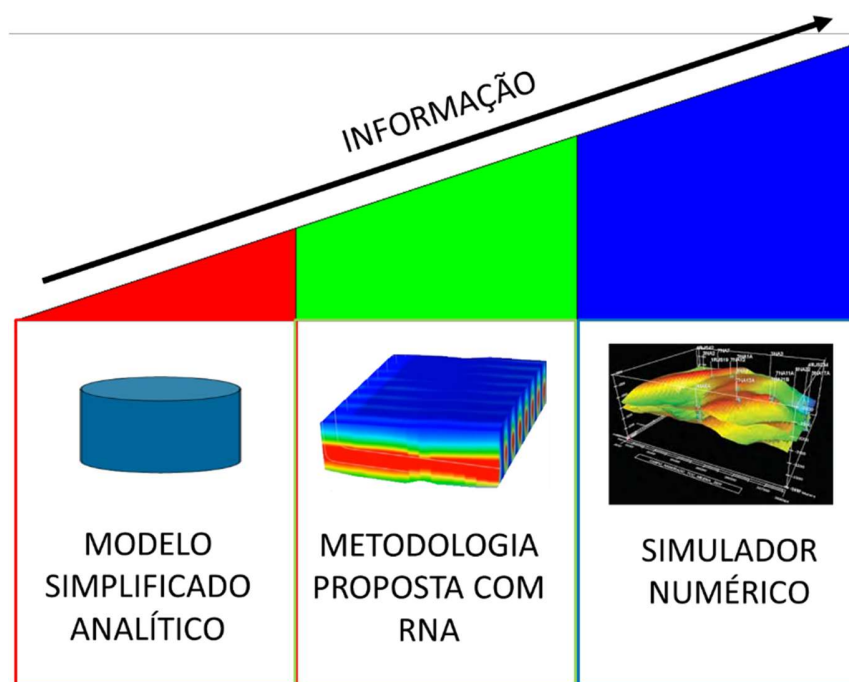


Figura 3.1. Sugestão de utilização da metodologia proposta neste trabalho.

A escolha de uma RNA se dá devido à sua natureza, que lhe permite ser aplicada em estudos multidisciplinares. Sua aplicação não é recente, e Al-fattah (2001) já verificava o crescente interesse em aplicar RNA nas diversas áreas da ciência, engenharia e finanças. Dentre as várias aplicações na engenharia de reservatório, as redes neurais têm sido usadas para reconhecimento de padrões e na predição do comportamento dos

reservatórios. Neste trabalho, foi utilizado RNA do tipo *Multilayer Perceptron* para a construção de um aproximador de função não linear para a previsão da vazão inicial e da curva de produção de um poço produtor de óleo obtida por um simulador numérico de reservatório.

Alguns trabalhos anteriores seguem a linha de estudo da Jia (2016) e Cao (2016), em que uma RNA é usada para fazer previsão da curva a partir do histórico de produção para um poço ou campo específico. A metodologia aqui proposta se diferencia pelo fato de não haver histórico da produção para projetos na fase exploratória. Outros trabalhos como o de Araujo Junior (2010), verificam o aprendizado de uma rede neural recorrente na previsão da curva de produção ao variar alguns parâmetros de entrada como vazão de injeção e intervalo de completação. Já Kohler (2013), treinou uma rede para prever a curva de produção fornecendo como entrada a coordenada do poço. Nesses casos, o reservatório já era conhecido e a previsão foi em nível do campo. No presente trabalho, buscou-se gerar uma curva de produção por poço sem se fixar a um reservatório específico. Foram utilizados como parâmetros de entrada como pressão, características do fluido, volume de óleo e permeabilidade na proximidade do poço. Cada um dos estudos listados buscava estimar uma curva de produção, porém, em situações diferentes. Cada estudo tem sua relevância e propósito.

Já era esperado que a previsão de uma curva de produção não seria simples, por esse motivo, decidiu-se também tentar estimar a vazão inicial do poço. A partir de uma vazão inicial bem calibrada, seria possível utilizar metodologias de curva de declínio para fazer uma estimativa simplificada.

A proposta de metodologia neste trabalho não inviabiliza as demais metodologias convencionais. Apenas apresenta uma nova técnica capaz de agregar à curva de produção informações da heterogeneidade do reservatório (quando comparado aos modelos analíticos que usam características homogêneas) e que possa apresentar uma resposta rápida e com baixos recursos computacionais (quando comparados aos simuladores numéricos). O foco do trabalho não é especificar e detalhar quando uma metodologia deve ser usada em detrimento da outra. Cada uma tem seu valor e propósito e cabe ao avaliador a tomada de decisão de qual é a mais apropriada.

3.2. Estrutura do Estudo

Neste trabalho, a sigla RNA será utilizada para se referir uma rede neural do tipo *feedforward*. O estudo foi conduzido em etapas conforme descrito na Figura 3.2, que se estende desde a definição dos parâmetros, coleta de dados, passando pelo treinamento da RNA e terminando com a estimativa da curva de produção do poço produtor de óleo.

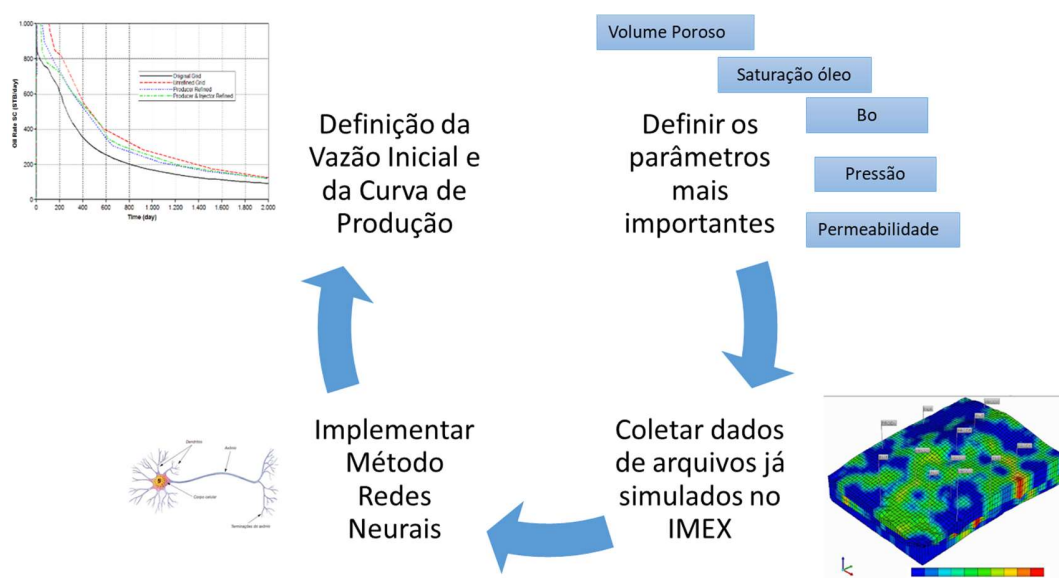


Figura 3.2. Etapas para a condução do trabalho.

A metodologia consistiu em coletar dados de simulações no simulador IMEX (CMG, 2015) realizadas para projetos exploratórios cujo fluido principal tenha sido o óleo. Nos modelos escolhidos, já tinham alguns poços posicionados no reservatório e, os arquivos de saída da simulação continham a curva de produção e a vazão inicial projetada para cada poço.

A simulação numérica consiste em representar o reservatório em pequenos pedaços, conhecidos como elementos, bem definidos e conectados entre si. Cada elemento é composto por diversas variáveis. Para este estudo, coletaram-se as características

predominantes dos elementos por onde o poço produtor foi perfurado e dos elementos adjacentes.

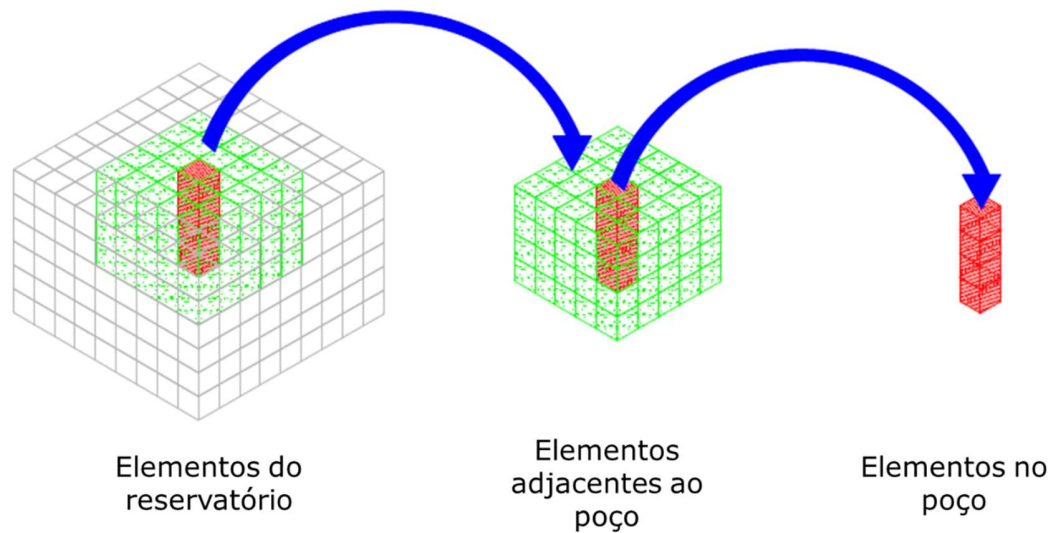


Figura 3.3. Foram coletados dados dos elementos por onde o poço foi perfurado e dos elementos adjacentes a ele.

De todas as variáveis disponíveis em cada elemento, foram selecionadas apenas algumas para treinar a RNA. O objetivo é, a partir dessas variáveis, estimar o comportamento da curva de produção resultante da simulação de fluxo.

Como premissa, assumiu-se como saídas esperadas da RNA:

- Vazão inicial do poço;
- Curva de produção por poço durante os 10 primeiros anos.

4. Resultados Numéricos Experimentais

4.1. Definição dos Parâmetros de Entrada

Buscou-se coletar informações que fossem representativas para a caracterização do reservatório e do fluido esperados para a acumulação e, ao mesmo tempo, que fossem características possíveis de se obter nos estudos preliminares da acumulação. Portanto, foram escolhidos os seguintes parâmetros da rocha e do fluido:

- Volume poroso;
- Saturação do óleo;
- Pressão do reservatório;
- Permeabilidade;
- Fator volume de formação.

Essas informações são referentes às condições iniciais da simulação e foram coletadas dos elementos por onde o poço foi perfurado e dos elementos adjacentes, conforme ilustrado na Figura 4.1.

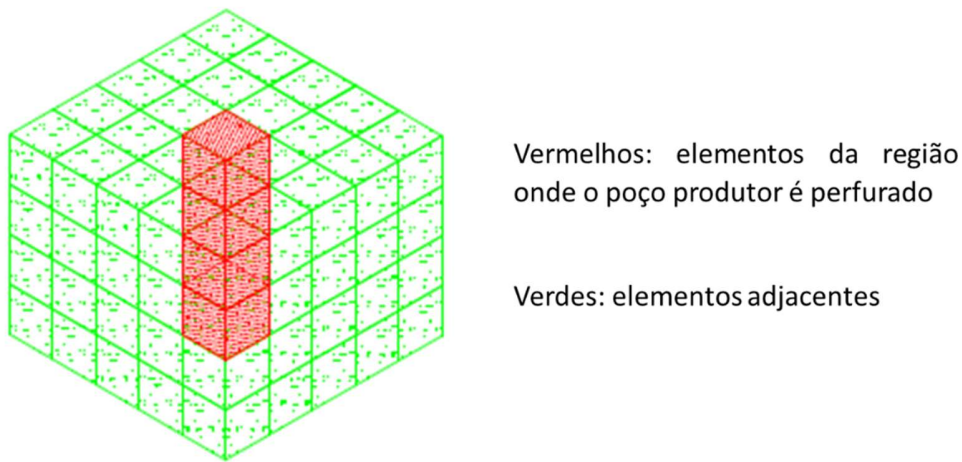


Figura 4.1. Definição de elementos do poço e adjacentes.

Optou-se por desmembrar os elementos que compõem o poço dos elementos que integram a área adjacente por se acreditar que os primeiros irão descrever melhor a vazão inicial de óleo (q_{oi}) do poço e os elementos adjacentes explicarão o perfil do declínio da produção. Os elementos adjacentes ao poço estariam relacionados ao seu raio de drenagem. No caso em estudo, está sendo investigado um raio de aproximadamente 700m ao redor do poço. A partir desses itens, foram derivados alguns parâmetros que pudessem captar a heterogeneidade do reservatório. Importante lembrar que o volume de óleo em cada elemento é calculado conforme descrito na Equação (2.1) e o valor aqui utilizado se refere às condições iniciais. Sendo R o número de elementos na região onde o poço produtor é perfurado e G o número de elementos adjacentes, os parâmetros que servirão como entrada serão:

- Média da permeabilidade na região do poço

$$\text{Entrada}_1 = \frac{\sum_{r=1}^R \text{permeabilidade}_r}{R}; \quad (4.1)$$

- Média da (permeabilidade x volume óleo) na região do poço

$$\text{Entrada}_2 = \frac{\sum_{r=1}^R (\text{permeabilidade}_r \times \text{volume óleo}_r)}{R}; \quad (4.2)$$

- Soma da permeabilidade na região do poço

$$\text{Entrada}_3 = \sum_{r=1}^R \text{permeabilidade}_r; \quad (4.3)$$

- Soma da (permeabilidade x volume óleo) na região do poço

$$\text{Entrada}_4 = \sum_{r=1}^R (\text{permeabilidade}_r \times \text{volume óleo}_r); \quad (4.4)$$

- Média da permeabilidade na região adjacente ao poço

$$\text{Entrada}_5 = \frac{\sum_{g=1}^G \text{permeabilidade}_g}{G}; \quad (4.5)$$

- Média da (permeabilidade x volume óleo) na região adjacente ao poço

$$\text{Entrada}_6 = \frac{\sum_{g=1}^G (\text{permeabilidade}_g \times \text{volume óleo}_g)}{G}; \quad (4.6)$$

- Soma da permeabilidade na região adjacente ao poço

$$\text{Entrada}_7 = \sum_{g=1}^G \text{permeabilidade}_g; \quad (4.7)$$

- Soma da (permeabilidade x volume óleo) na região adjacente ao poço

$$Entrada_g = \sum_{g=1}^G (permeabilidade_g \times volume\ óleo_g); \quad (4.8)$$

- $Entrada_9$ = Pressão média do reservatório na proximidade do poço nas condições iniciais;

- $Entrada_{10}$ = Fator volume de formação do fluido nas condições iniciais. Nos modelos de simulação utilizados nesse estudo, o fluido é uniforme em todo o reservatório nas condições iniciais.

4.2. Banco de Dados

Utilizou-se como base oito modelos de simulação feitos para um projeto exploratório situado no pré-sal da bacia de campos. Por questões de confidencialidade, não serão mostrados em detalhes as características dos reservatórios estudados. Em cada um desses modelos, foram variadas características de rocha e fluido. Foi utilizada a injeção de água como método de recuperação suplementar, logo, em cada modelo havia poços produtores de óleo e os injetores de água posicionados. Em todos os modelos simulados, os reservatórios encontravam-se inicialmente subsaturados, ou seja, a pressão inicial do reservatório encontrava-se acima ou igual à pressão de bolha do fluido (pressão mínima em que o gás dissolvido no óleo começa a sair de solução). Foram simulados dois tipos de fluidos, um com Bo de $1,39\text{ m}^3/\text{m}^3\text{ std}$ (Tipo 1) e o outro com um Bo de $1,59\text{ m}^3/\text{m}^3\text{ std}$ (Tipo 2).

Para essa análise, não foi colocada nenhuma restrição de produção aos poços, como por exemplo, restrição na capacidade de produção das plataformas para onde os poços foram conectados. Dessa forma, a produção dos poços cresce ao longo do ano 1 e declina em seguida, não havendo patamar de produção. A Figura 4.2 mostra um perfil médio da curva de produção de óleo dos poços produtores utilizados como amostra para esse estudo.

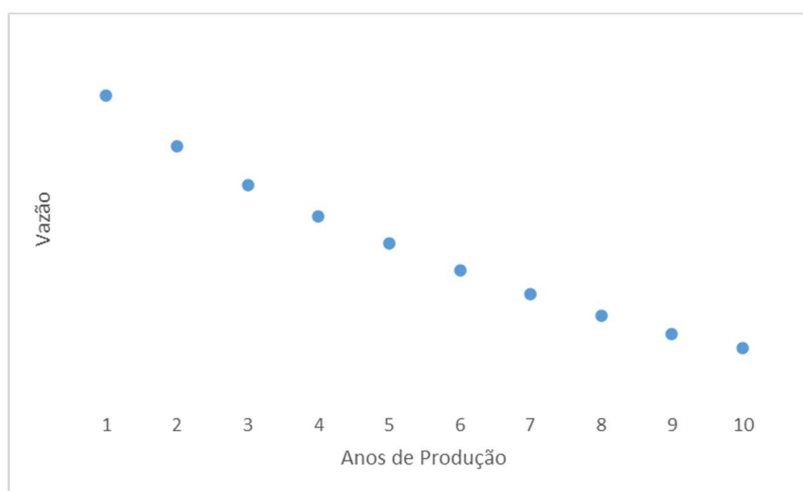


Figura 4.2. Perfil médio da curva de produção de óleo dos poços utilizados nessa análise

Como não será possível mostrar o desenho dos modelos de reservatório utilizados, será feita uma analogia com alguns modelos simplificados e hipotéticos para fins de demonstração da metodologia. A Figura 4.3 é apenas ilustrativa e representa seis modelos de reservatórios hipotéticos com os respectivos poços produtores (em vermelho) alocados.

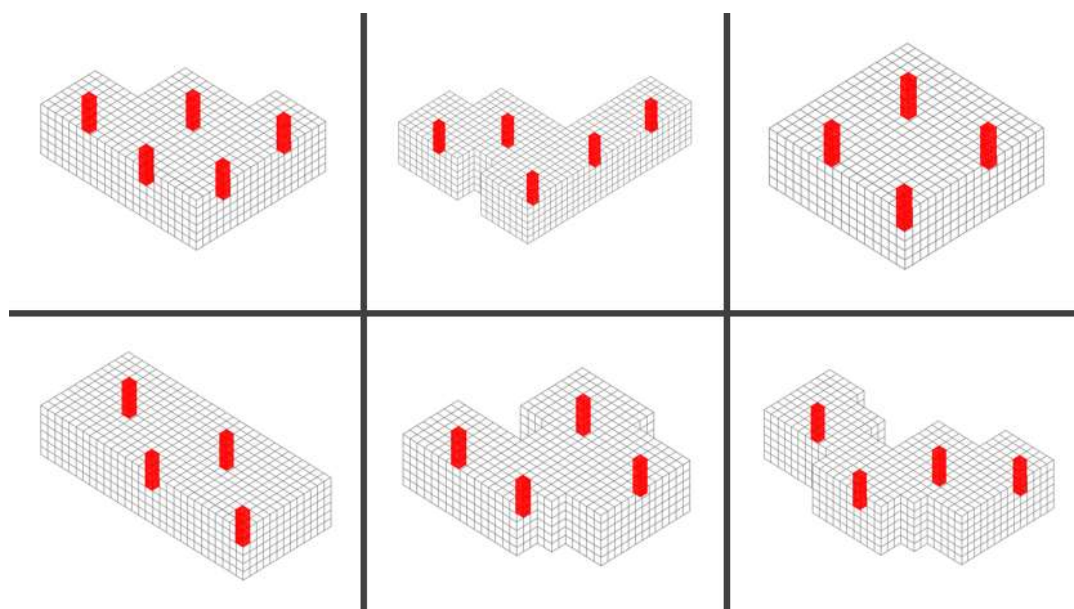


Figura 4.3. Ilustração de seis modelos de reservatório hipotéticos com os respectivos poços produtores (em vermelho) alocados.

Fazendo analogia com a Figura 4.1, foram coletadas as informações dos elementos da região do poço (em vermelho) e da região adjacente ao poço (em verde). A Figura 4.4 ilustra as regiões investigadas de cada um dos modelos de reservatório hipotéticos.

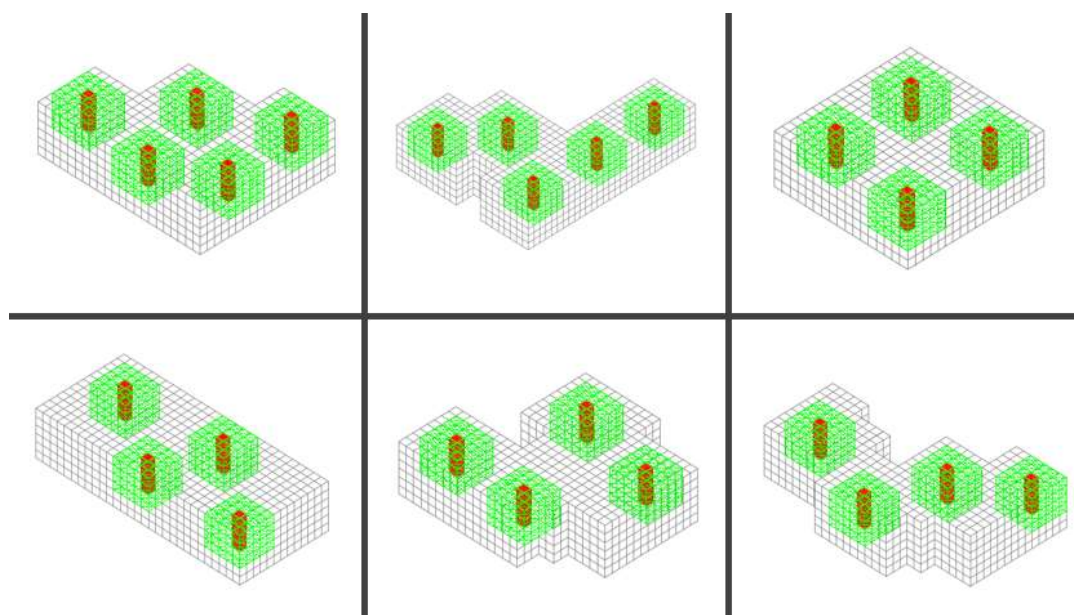


Figura 4.4. Ilustração de seis modelos de reservatório hipotéticos com os respectivos poços produtores (em vermelho) e sua região adjacente (em verde).

Ao total, foram coletados dados de 91 poços com características diversas de forma a gerar uma base heterogênea e ampla para treinar a rede neural. Os gráficos da Figura 4.5, Figura 4.6 e Figura 4.7 ilustram a diversidade dos elementos considerados. Dos poços que foram coletados, as vazões iniciais de óleo variam conforme a Figura 4.5. A Figura 4.6 mostra a distribuição dos tipos de fluidos entre os poços.

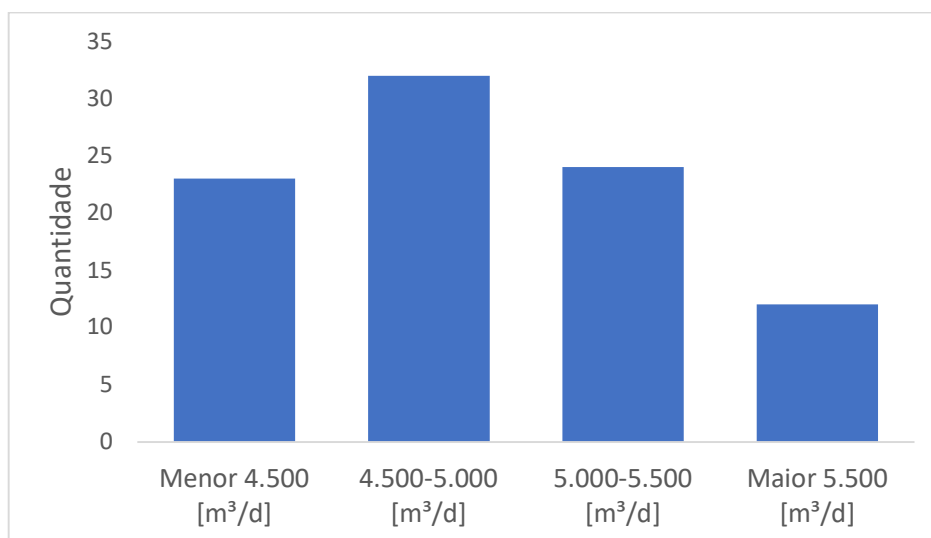


Figura 4.5 Distribuição das vazões iniciais que compõem o banco de dados.

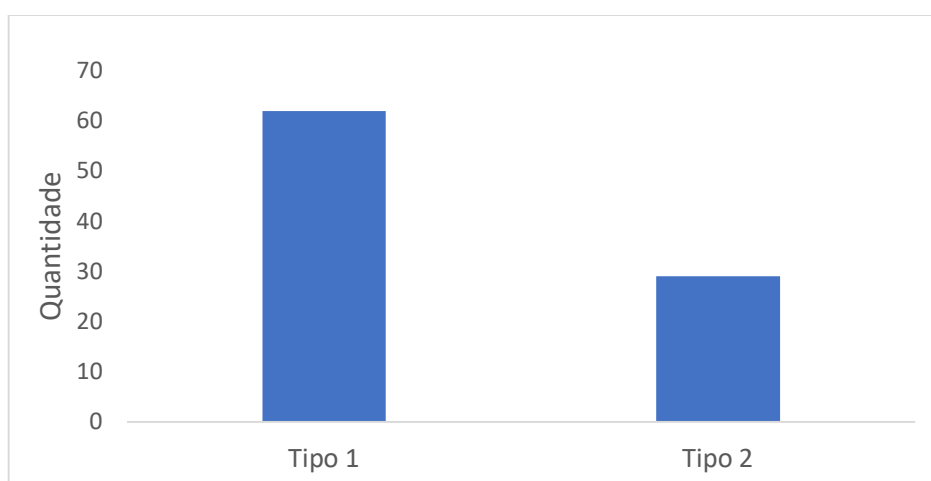


Figura 4.6 Distribuição dos fluidos que compõem o banco de dados.

No gráfico mostrado na Figura 4.7, pode-se observar a distribuição das permeabilidades médias encontradas nas regiões dos poços produtores.

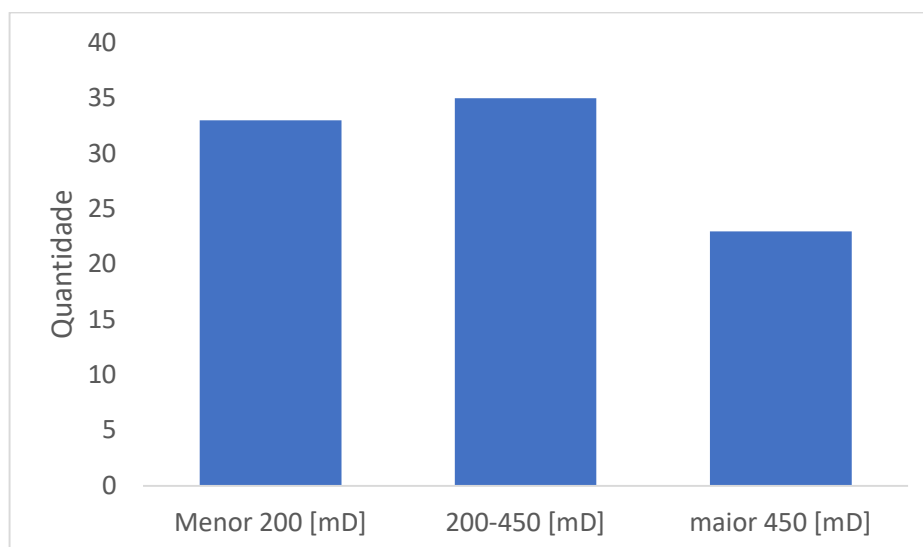


Figura 4.7 Distribuição da permeabilidade média na região do poço produtor utilizado no banco de dados.

Foi fixado um horizonte de previsão da curva de produção de 10 anos, visto que, após esse período, acredita-se que haverá uma grande influência dos poços injetores próximos na curva de produção. Esse estudo está restrito aos poços produtores, ou seja, não estão sendo considerados os poços injetores.

4.3. Treinamento

A etapa de processamento dos dados por meio da rede neural iniciou-se com a inserção no MATLAB dos dados normalizados. Utilizou-se a biblioteca *Neural Network Toolbox* para a construção do modelo de inteligência artificial.

Foram propostas duas abordagens de treinamento:

- 1) Caso única rede (UR): uma única rede MLP é usada para prever a vazão inicial e a curva de produção ao longo dos 10 anos

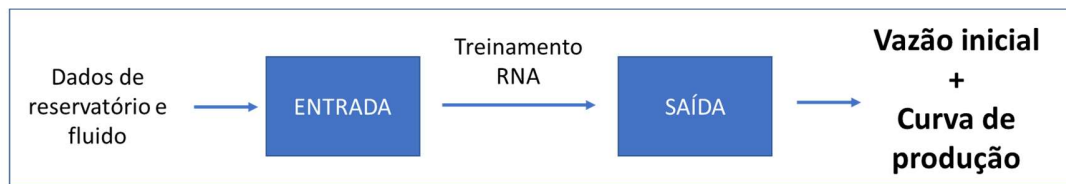


Figura 4.8 Representação esquemática da construção da rede para a previsão da vazão inicial e curva de produção no caso UR.

- 2) Caso múltiplas redes (MR): são treinadas várias redes, sendo uma para vazão inicial e uma para cada ano de produção. Ou seja, nesse caso, serão 11 redes do tipo MLP.

Na RNA criada para a previsão da vazão inicial, os dados de entrada são os parâmetros do reservatório e do fluido nas condições iniciais e a saída esperada é a vazão inicial. Essa vazão inicial será um dado de entrada na RNA de previsão da vazão média do ano 1 (ver Figura 4.10).

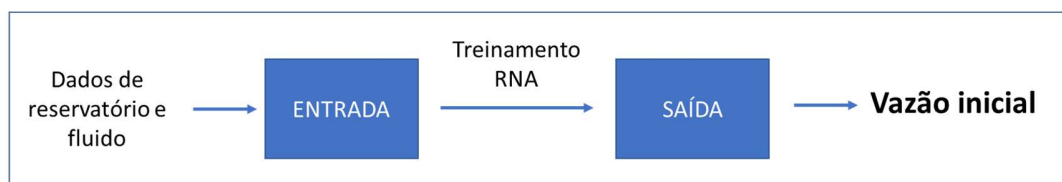


Figura 4.9 Representação esquemática da construção da rede para a previsão da vazão inicial no caso MR.

Para a previsão da vazão média dos anos seguintes, foi criada uma RNA para cada ano, de forma que cada rede usasse como dados de entrada os parâmetros do reservatório e do fluido nas condições iniciais e a previsão de vazão média do ano anterior para prever a vazão média do ano seguinte.



Figura 4.10 Representação esquemática da construção das redes para a previsão da curva de produção no caso MR.

Em ambos os casos (UR e MR), utilizou-se a metodologia do *K-fold* para verificar o nível de generalização da RNA. Neste trabalho, k assumirá o valor de 10, tornando assim, a divisão dos dados por *10-fold*. Para cada uma das k simulações, o grupo destinado ao treino teve uma subdivisão onde uma parte será utilizada para treinamento e outra para validação. Por fim, o grupo de teste será utilizado para realizar o teste final da RNA. Neste trabalho, adotou-se 80% para treinamento e 20% para validação (Figura 4.11)

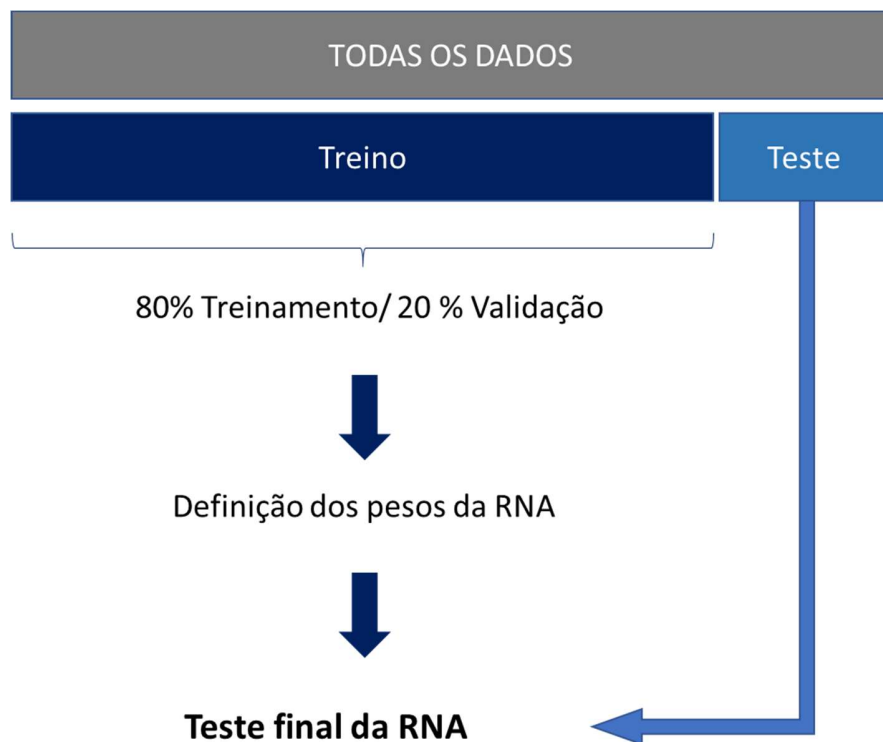


Figura 4.11 Representação da utilização de cada grupo na definição da RNA.

Idealmente, o processo de otimização buscará o valor mínimo global do problema. Contudo, nem sempre é possível e o algoritmo precisa de algum critério de parada caso ele encontre um mínimo local. Neste trabalho, foram empregados os seguintes critérios de parada:

- Número máximo de iterações: 1000;
- Parada quando o erro (MSE) do processo de validação aumenta por 15 interações seguidas durante o treinamento (exemplificado na Figura 2.6).

Em todas as redes, o segundo critério de parada foi o responsável pela interrupção do treinamento. Esse critério evitou que ocorresse um *overfitting* durante o processo de otimização.

4.4. Definição dos Parâmetros da RNA

Conforme descrito anteriormente, foram propostas duas metodologias de treinamento (caso UR e caso MR). Em todas as redes testadas neste trabalho foi utilizada a função de ativação linear na saída da rede e a função tangente hiperbólica nas camadas intermediárias. Com base em estudos preliminares, observou-se que redes com duas camadas ocultas apresentavam melhores resultados, portanto, foi definido como premissa inicial a utilização de RNAs com duas camadas ocultas. Para cada uma das RNAs, foram testados qual a melhor normalização, o melhor algoritmo de treinamento e qual a melhor topologia da rede para o problema estudado.

Os tipos de normalização verificados foram:

- Normalização por desvio padrão (DP);
- Normalização Min-Max.

Para essa análise, foi considerada uma rede com duas camadas ocultas, sendo 10 neurônios na primeira e 5 neurônios na segunda. Para treinamento utilizou-se o Método de Levenberg–Marquardt como algoritmo de aprendizado.

Os tipos de algoritmo de aprendizado verificados foram:

- BFGS;
- Método de Levenberg–Marquardt (LM).

Nessa análise, já foi considerada a normalização que teve o melhor resultado da verificação anterior. Além disso, assumiu-se uma RNA com duas camadas ocultas, sendo 10 neurônios na primeira e 5 neurônios na segunda.

Para definir qual a melhor topologia para este estudo foi considerada uma RNA com a melhor normalização e o melhor algoritmo de aprendizado obtidos nas avaliações anteriores. Como padrão, foram simulados os cenários apresentados na Tabela 4.1 variando o número de neurônios nas duas camadas ocultas.

Tabela 4.1 Cenários simulados para definir a melhor topologia da RNA.

Cenário	Neurônios camada oculta 1	Neurônios camada oculta 2
1	4	2
2	4	4
3	4	6
4	4	8
5	4	10
6	8	2
7	8	4
8	8	6
9	8	8
10	8	10
11	12	2
12	12	4
13	12	6
14	12	8
15	12	10
16	16	2
17	16	4
18	16	6
19	16	8
20	16	10
21	20	2
22	20	4
23	20	6
24	20	8
25	20	10

O indicador MAPEam é importante para essas avaliações, visto que, ele sinaliza qual das redes, em valor presente, possui o menor erro. Dessa forma, os erros dos anos iniciais têm maior peso do que os finais. Da mesma forma que POSTALI (2009), padronizou-se nesse trabalho 10% a.a. como taxa de desconto.

4.4.1. Caso UR

Para as análises da melhor normalização e algoritmo de aprendizado, foi utilizada uma RNA com a configuração ilustrada na Figura 4.12

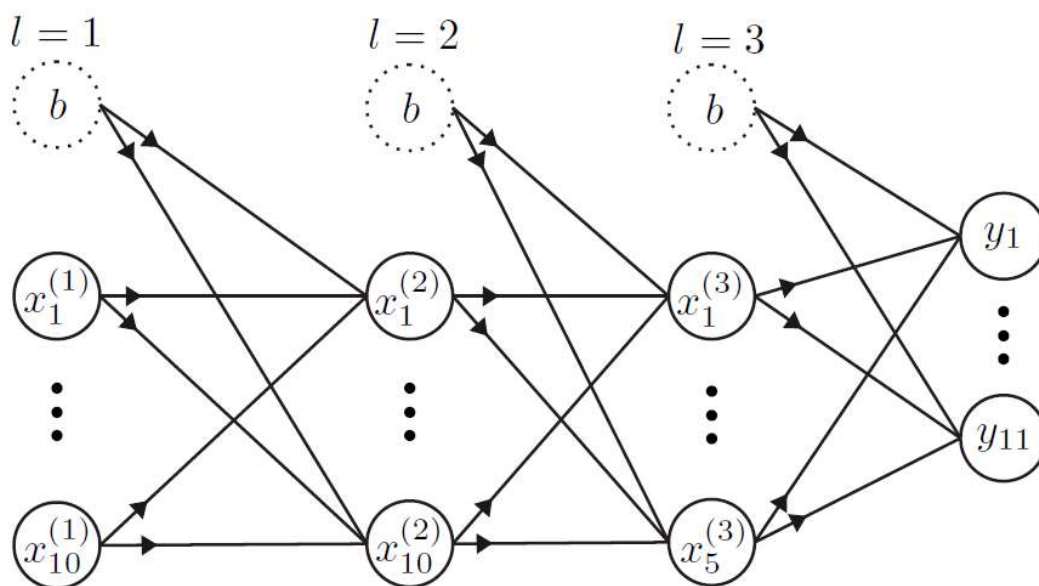


Figura 4.12 Configuração da rede utilizada para avaliar qual a melhor RNA para a curva de produção do caso UR.

- Definição da normalização

Pela Tabela 4.2, a normalização Min-Max obteve um melhor resultado tanto no grupo de treino como no grupo de teste.

Tabela 4.2 Comparação da *performance* da RNA entre dois tipos de normalização para o caso UR.

	Normalização DP	Normalização Min-Max
MAPEam Treino	8%	7%
MAPEam Teste	12%	10%

A Tabela 4.3 e a Tabela 4.4 mostram respectivamente o MAPE e o MAD [m^3/d] ano a ano para ambos os grupos de treino e teste.

Tabela 4.3 Comparação do MAPE ano a ano entre os tipos de normalização para o grupo de treino e teste no caso UR.

		Qoi	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	DP	4.7%	5.0%	7.4%	9.7%	12.7%	13.5%	14.8%	17.9%	19.6%	21.9%	24.7%
	Min-Max	4.5%	4.2%	6.3%	8.7%	10.6%	11.6%	12.7%	14.8%	16.5%	18.7%	20.9%
Teste	DP	5.2%	6.0%	9.2%	15.1%	20.8%	23.6%	27.1%	29.7%	30.5%	31.6%	34.3%
	Min-Max	5.3%	5.6%	9.3%	13.8%	17.0%	19.0%	21.6%	23.7%	25.3%	27.7%	29.9%

Tabela 4.4 Comparação do MAD [m^3/d] ano a ano entre os tipos de normalização para o grupo de treino e teste no caso UR.

		Qoi	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	DP	237	231	282	306	333	320	321	343	344	351	356
	Min-Max	213	181	229	274	287	274	272	292	292	303	309
Teste	DP	257	267	338	458	538	554	564	544	508	477	442
	Min-Max	248	247	333	432	472	465	471	474	454	455	472

A partir desse estudo, optou-se por utilizar a normalização Min-Max como sendo a normalização padrão do caso UR.

- Definição do algoritmo de aprendizado

Para essa análise foi utilizada a mesma topologia feita para a análise anterior, ou seja, duas camadas ocultas: 10 neurônios na primeira e 5 neurônios na segunda, e a normalização tipo Min-Max. A partir da comparação feita na Tabela 4.5, conclui-se que o algoritmo Levenberg–Marquardt se mostrou como o melhor método de aprendizado para esse problema.

Tabela 4.5 Comparação do MAPEam para dois tipos de algoritmo de aprendizado para o caso UR.

	Levenberg– Marquardt	BFGS
MAPEam Treino	7%	9%
MAPEam Teste	10%	11%

A Tabela 4.6 e a Tabela 4.7 mostram respectivamente o MAPE e o MAD [m^3/d] ano a ano para ambos os grupos, treino e teste.

Tabela 4.6 Comparação do MAPE ano a ano entre os tipos de algoritmo de aprendizado para o grupo de treino e teste no caso UR.

		Qoi	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	LM	4.5%	4.2%	6.3%	8.7%	10.6%	11.6%	12.7%	14.8%	16.5%	18.7%	20.9%
	BFGS	6.3%	6.5%	8.6%	11.2%	14.8%	16.8%	18.0%	19.8%	21.6%	23.2%	25.2%
Teste	LM	5.3%	5.6%	9.3%	13.8%	17.0%	19.0%	21.6%	23.7%	25.3%	27.7%	29.9%
	BFGS	6.8%	7.9%	11.1%	14.7%	18.5%	21.3%	22.7%	24.8%	27.0%	27.3%	30.2%

Tabela 4.7 Comparação do MAD [m^3/d] ano a ano entre os tipos de algoritmo de aprendizado para o grupo de treino e teste no caso UR.

		Qoi	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	LM	213	181	229	274	287	274	272	292	292	303	309
	BFGS	298	283	310	334	380	388	376	383	372	366	366
Teste	LM	248	247	333	432	472	465	471	474	454	455	472
	BFGS	315	340	393	435	470	481	468	472	458	423	427

A partir dessa análise, definiu-se utilizar o algoritmo Levenberg–Marquardt como algoritmo de aprendizado da RNA padrão para o caso UR.

- Definição da topologia da RNA

Para essa análise foi utilizada a normalização tipo Min-Max e o método Levenberg–Marquardt. Por meio da Tabela 4.8, percebe-se que não há grandes diferenças entre as topologias simuladas. O melhor cenário foi o que representa duas camadas ocultas, sendo 8 na primeira e 10 na segunda.

Tabela 4.8 Resultado do MAPEam para os cenários de variação no número de neurônios no caso UR.

Cenário	Neurônios camada oculta 1	Neurônios camada oculta 2	MAPEam Treino	MAPEam Teste
1	4	2	7.76%	10.11%
2	4	4	7.27%	10.27%
3	4	6	6.70%	10.12%
4	4	8	7.22%	9.92%
5	4	10	7.65%	10.44%
6	8	2	7.20%	10.02%
7	8	4	7.22%	10.70%
8	8	6	7.42%	9.76%
9	8	8	6.50%	10.10%
10	8	10	6.23%	9.39%
11	12	2	7.60%	9.80%
12	12	4	7.03%	10.51%
13	12	6	8.73%	11.01%
14	12	8	6.68%	10.19%
15	12	10	6.53%	9.68%
16	16	2	7.37%	10.24%
17	16	4	7.80%	10.51%
18	16	6	6.47%	9.87%
19	16	8	6.66%	10.25%
20	16	10	6.20%	10.18%
21	20	2	7.75%	9.86%
22	20	4	7.71%	10.08%
23	20	6	8.77%	10.04%
24	20	8	6.24%	9.65%
25	20	10	7.38%	10.51%

A Tabela 4.9, Tabela 4.10 e a Tabela 4.11, mostram o MAPE, com seu respectivo DP, e MAD [m³/d] do cenário com 8 neurônios na primeira camada oculta e 10 na segunda.

Tabela 4.9 MAPE ano a ano no melhor cenário simulado para o caso UR.

	Qoi	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	3.5%	3.8%	5.9%	8.1%	10.0%	11.1%	13.2%	14.0%	15.5%	17.7%	19.6%
Teste	4.1%	5.0%	9.0%	13.5%	15.3%	16.8%	18.8%	22.2%	26.4%	24.6%	29.7%

Tabela 4.10 DP do MAPE ano a ano no melhor cenário simulado para o caso UR.

	Qoi	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	±1%	±1%	±2%	±3%	±4%	±4%	±6%	±4%	±4%	±4%	±4%
Teste	±2%	±1%	±4%	±8%	±9%	±7%	±7%	±8%	±8%	±5%	±7%

Tabela 4.11 MAD [m³/d] ano a ano no melhor cenário simulado para o caso UR.

	Qoi	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	166	164	211	251	261	257	283	272	272	288	290
Teste	190	214	310	395	386	387	394	424	444	394	428

- Conclusão

Finalmente, a partir desse estudo realizado, pode-se concluir que o melhor resultado para o caso UR foi obtido com a seguinte configuração de RNA:

- Normalização Min-Max
- Algoritmo de treinamento Levenberg–Marquardt
- Duas camadas ocultas sendo, 8 na primeira camada e 10 na segunda camada.

4.4.2. Caso MR

Como descrito acima, nesse caso há uma RNA para previsão da vazão inicial e outras 10 para previsão de cada um dos anos da curva de produção.

4.4.2.1. RNA da vazão inicial

Para os testes da melhor normalização e algoritmo de treinamento, foi considerada a RNA com a seguinte topologia:

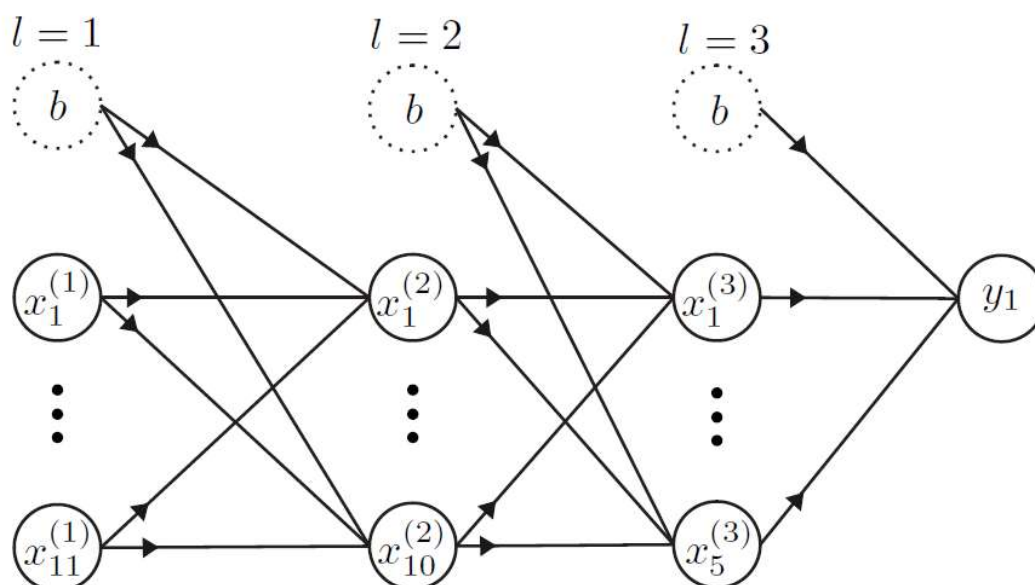


Figura 4.13. Configuração da rede utilizada para avaliar qual a melhor normalização e o melhor algoritmo de treinamento para a RNA da vazão inicial.

- Definição da normalização

Para essa análise, como não se trata de uma série temporal, será utilizado o MAPE para avaliar a melhor alternativa. De acordo com a Tabela 4.12, a normalização Min-Max obteve o melhor resultado tanto no grupo de treino como no grupo de teste. O mesmo resultado é obtido quando se observa o MAD mostrado na Tabela 4.13

Tabela 4.12 Comparação do MAPE da RNA da vazão inicial entre dois tipos de normalização.

	Normalização DP	Normalização Min-Max
MAPE Treino	1,5%	1,5%
MAPE Teste	3,1%	2,6%

Tabela 4.13 Comparação do MAD [m³/d] da RNA da vazão inicial entre dois tipos de normalização.

	Normalização DP	Normalização Min-Max
MAD Treino	72	68
MAD Teste	156	123

Após a comparação, optou-se por adotar a Min-Max como padrão para a RNA da vazão inicial.

- Definição do algoritmo de aprendizado

Para definir o melhor algoritmo de aprendizado, adotou-se como padrão a normalização Min-Max para gerar os resultados. Foram utilizadas duas camadas ocultas com 10 neurônios na primeira e 5 neurônios na segunda. Analisando o MAPE (Tabela 4.14) e o MAD (Tabela 4.15), conclui-se que o algoritmo Levenberg–Marquardt se mostrou o mais adequado para a RNA da vazão inicial.

Tabela 4.14 Comparação do MAPE da RNA da vazão inicial entre dois tipos de algoritmo de treino.

	BFGS	Levenberg–Marquardt
MAPE Treino	2,1%	1,5%
MAPE Teste	2,9%	2,6%

Tabela 4.15 Comparação do MAD [m³/d] da RNA da vazão inicial entre dois tipos de algoritmo de treino.

	BFGS	Levenberg–Marquardt
MAD Treino	100	68
MAD Teste	136	123

- Definição da topologia da RNA

Foram adotados a normalização Min-Max e o método o Levenberg–Marquardt para gerar os resultados dos cenários para a escolha da melhor topologia. Analisando o MAPE (Tabela 4.16) e o MAD (Tabela 4.17), conclui-se que a melhor topologia para a RNA da vazão inicial é composta por 16 neurônios na primeira camada oculta e 6 neurônios na segunda.

Tabela 4.16 MAPE dos cenários simulados para definição da melhor topologia para a RNA da vazão inicial.

Cenário	Neurônios camada oculta 1	Neurônios camada oculta 2	MAPE Treino	MAPE Teste
1	4	2	1.4%	2.4%
2	4	4	1.5%	2.2%
3	4	6	1.4%	2.3%
4	4	8	1.5%	2.6%
5	4	10	1.4%	2.3%
6	8	2	1.3%	3.3%
7	8	4	1.4%	2.2%
8	8	6	1.4%	2.9%
9	8	8	2.0%	3.2%
10	8	10	1.9%	2.6%
11	12	2	1.3%	2.7%
12	12	4	1.8%	3.0%
13	12	6	1.9%	3.0%
14	12	8	2.2%	3.3%
15	12	10	1.6%	2.9%
16	16	2	1.9%	3.1%
17	16	4	1.5%	2.7%
18	16	6	1.3%	2.2%
19	16	8	1.4%	3.2%
20	16	10	1.5%	3.0%
21	20	2	1.7%	2.7%
22	20	4	1.7%	2.8%
23	20	6	1.7%	3.3%
24	20	8	1.8%	3.1%
25	20	10	1.7%	3.0%

- Conclusão

Portanto, os próximos passos do trabalho serão feitos com as seguintes premissas para a RNA da vazão inicial:

- Normalização Min-Max
- Algoritmo de treinamento Levenberg–Marquardt
- Duas camadas ocultas sendo, 16 na primeira camada e 6 na segunda camada.

4.4.2.2. RNAs da geração da curva de produção

Para a análise dos resultados das RNAs responsáveis pela geração da curva de produção, serão aplicados alguns parâmetros empregados em Rede Neural Recorrente (RNR). Esse tipo de rede é caracterizado por ter um laço de realimentação (Haykin, 2001), em que a saída de neurônios de determinada camada é introduzida como entrada de neurônios nas camadas anteriores ou da própria.). A análise de RNRs é feita utilizando o conceito de *openloop* e *closedloop*:

- *Openloop*: é o modo em que a rede é treinada, em que as saídas são calculadas considerando como entradas as vazões reais do simulador no tempo t .
- *Closedloop*: nesse modo é utilizado a rede treinada no *openloop*. A vazão no tempo t é estimada a partir da vazão calculada pela rede no tempo $t - 1$.

A escolha dos melhores cenários aqui avaliados será feita na visão *openloop*. A avaliação será feita em cima da curva de produção total estimada ao longo dos 10 anos. A Figura 4.14 mostra a topologia utilizada para avaliar a melhor normalização e o melhor algoritmo de treinamento.

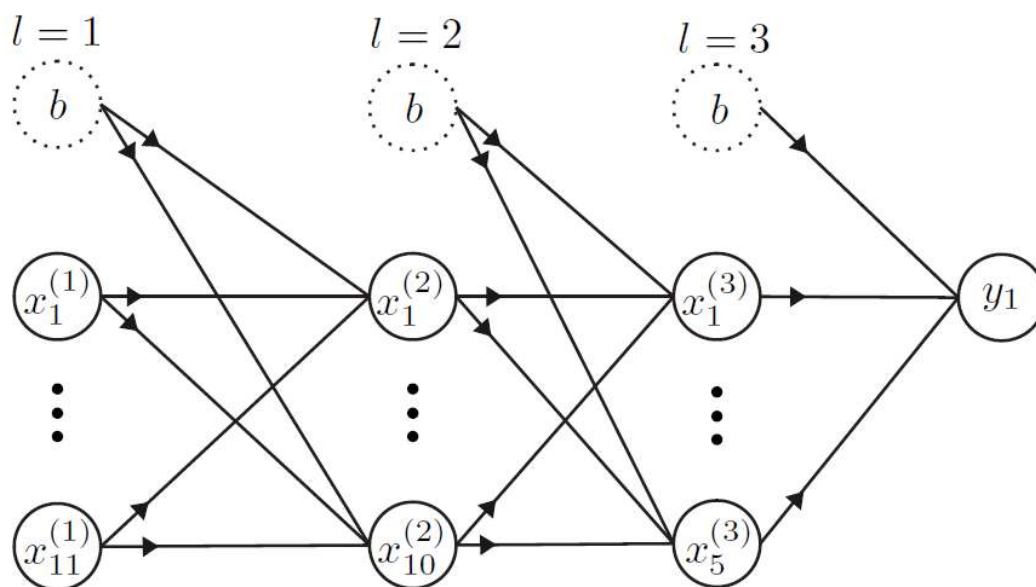


Figura 4.14 Configuração da rede utilizada para avaliar qual a melhor normalização e o melhor algoritmo de treinamento para as RNAs de geração da curva de produção no caso MR.

- Definição da normalização

Por se tratar de uma série temporal, será utilizado o indicador MAPEam, dessa forma, os erros dos anos iniciais terão maior peso do que os finais. De acordo com a Tabela 4.17, Tabela 4.18 e a Tabela 4.19, conclui-se que a normalização Min-Max obteve um melhor resultado tanto no grupo de treino como no grupo de teste.

Tabela 4.17 Comparação do desempenho da RNA entre dois tipos de normalização no caso MR para geração da curva de produção.

	Normalização DP	Normalização Min-Max
MAPEam Treino	2,20%	1,90%
MAPEam Teste	4,32%	3,63%

Tabela 4.18 Comparação do MAPE ano a ano entre dois tipos de normalização para o grupo de treino e teste no caso MR.

		Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	DP	2.1%	2.1%	3.0%	5.2%	3.6%	4.9%	4.4%	5.0%	4.6%	2.7%
	Min-Max	2.0%	2.1%	2.3%	3.2%	4.0%	4.1%	3.6%	4.0%	4.1%	3.4%
Teste	DP	3.5%	6.1%	5.6%	9.9%	7.3%	8.2%	8.2%	8.2%	9.5%	7.2%
	Min-Max	3.5%	4.8%	5.8%	7.1%	5.9%	8.1%	6.6%	6.7%	6.5%	6.4%

Tabela 4.19 Comparação do MAD [m³/d] ano a ano entre os tipos de normalização para o grupo de treino e teste no caso MR.

		Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	DP	98	84	98	143	90	112	92	93	76	44
	Min-Max	84	74	73	87	95	91	76	77	70	52
Teste	DP	159	221	180	249	173	198	174	159	158	115
	Min-Max	149	164	187	183	145	176	140	132	111	108

A partir desse estudo, definiu-se utilizar a normalização Min-Max como sendo a base para esse problema por apresentar a melhor *performance*.

- Definição do algoritmo de aprendizado

Adotou-se a normalização tipo Min-Max como padrão para essa análise. Analisando o MAPEam (Tabela 4.20) conclui-se que o algoritmo Levenberg–Marquardt se mostrou como o melhor método de aprendizado para esse problema.

Tabela 4.20 Comparação do MAPEam para dois tipos de algoritmo de aprendizado no caso MR para geração da curva de produção.

	Levenberg–Marquardt	BFGS
MAPEam Treino	1,90%	3,10%
MAPEam Teste	3,63%	4,47%

A Tabela 4.21 e a Tabela 4.22 mostram respectivamente o MAPE e o MAD [m³/d] ano a ano para ambos os grupos de treino e teste.

Tabela 4.21 Comparação do MAPE ano a ano para dois tipos de algoritmo de aprendizado no caso MR.

		Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	LM	2.0%	2.1%	2.3%	3.2%	4.0%	4.1%	3.6%	4.0%	4.1%	3.4%
	BFGS	3.3%	4.1%	4.7%	5.9%	4.8%	5.6%	6.2%	5.5%	6.4%	6.4%
Teste	LM	3.5%	4.8%	5.8%	7.1%	5.9%	8.1%	6.6%	6.7%	6.5%	6.4%
	BFGS	4.3%	5.9%	6.5%	9.2%	7.4%	8.5%	8.7%	7.1%	10.7%	8.0%

Tabela 4.22 Comparação do MAD [m³/d] ano a ano para dois tipos de algoritmo de aprendizado no caso MR.

		Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	LM	84	74	73	87	95	91	76	77	70	52
	BFGS	142	151	145	162	119	131	129	103	106	98
Teste	LM	149	164	187	183	145	176	140	132	111	108
	BFGS	182	208	193	255	182	191	187	136	182	125

A partir desse estudo, optou-se por utilizar o algoritmo Levenberg–Marquardt como método de aprendizado da RNA construída para este problema.

- Definição da topologia da RNA

Assumiu-se como padrão a normalização Min-Max e o método Levenberg–Marquardt. Ao analisar os dados da Tabela 4.23, percebe-se que não há uma diferença muito grande

entre os cenários simulados. A topologia que apresentou melhor resultado foi a de 4 neurônios em cada uma das duas camadas ocultas.

Tabela 4.23 Resultado do MAPEam para os cenários de variação no número de neurônios do caso MR para geração da curva de produção.

Cenário	Neurônios camada oculta 1	Neurônios camada oculta 2	MAPEam Treino	MAPEam Teste
1	4	2	2.00%	3.10%
2	4	4	1.80%	3.00%
3	4	6	1.90%	3.30%
4	4	8	2.10%	3.50%
5	4	10	1.90%	3.60%
6	8	2	2.00%	3.30%
7	8	4	2.10%	3.70%
8	8	6	2.10%	4.00%
9	8	8	2.00%	3.60%
10	8	10	2.10%	4.20%
11	12	2	1.90%	3.50%
12	12	4	1.80%	3.60%
13	12	6	2.00%	3.90%
14	12	8	2.10%	3.80%
15	12	10	2.10%	4.10%
16	16	2	2.10%	4.00%
17	16	4	2.20%	4.00%
18	16	6	2.00%	4.20%
19	16	8	2.00%	3.90%
20	16	10	2.20%	4.20%
21	20	2	2.20%	4.00%
22	20	4	2.20%	3.90%
23	20	6	2.10%	4.10%
24	20	8	2.10%	4.30%
25	20	10	2.00%	4.40%

A Tabela 4.24 e a Tabela 4.25 apresentam respectivamente o MAPE e o MAD [m³/d] ano a ano do melhor cenário obtido das simulações.

Tabela 4.24. Análise do MAPE ano a ano do cenário com 4 neurônios em cada camada oculta no caso MR.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	2.1%	2.4%	2.6%	2.7%	3.6%	4.3%	3.0%	3.7%	3.7%	3.0%
Teste	3.3%	3.8%	4.9%	4.7%	5.1%	5.8%	5.7%	5.6%	6.9%	5.5%

Tabela 4.25 Comparação do MAD [m^3/d] ano a ano do cenário com 4 neurônios em cada camada oculta no caso MR.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	92	86	81	76	92	99	65	71	65	47
Teste	143	130	138	124	137	143	115	110	112	90

- Conclusão

Finalmente, os próximos passos do trabalho serão feitos com as seguintes premissas para a RNA da curva de produção no caso MR:

- Normalização Min-Max
- Algoritmo de treinamento Levenberg–Marquardt
- Duas camadas ocultas sendo, 4 na primeira camada e 4 na segunda camada.

Para se ter números comparáveis com o caso UR, é necessário analisar o modo *closedloop*, em que as vazões da curva de produção no tempo t são estimadas a partir de previsões feitas pela própria RNA no tempo $t - 1$. Conforme ilustrado na Figura 4.10, será utilizada a saída da RNA da vazão inicial como entrada para estimativa da vazão do ano 1.

Já é esperado que a rede nesse modo tenha um desempenho inferior ao encontrado no modo *openloop*, visto que, a previsão da vazão no tempo t é prejudicada pelos erros das previsões das vazões nos tempos anteriores. O MAPE, com seu respectivo DP, e o MAD [m^3/d] anual no modo *closedloop* são mostrados respectivamente na Tabela 4.26, Tabela 4.27 e a Tabela 4.28.

Tabela 4.26 MAPE ano a ano obtido na visão *closedloop* do caso MR para a curva de produção.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	3%	6%	8%	10%	11%	13%	14%	16%	19%	21%
Teste	4%	9%	12%	14%	16%	18%	20%	22%	25%	27%

Tabela 4.27 DP do MAPE ano a ano obtido na visão *closedloop* do caso MR para a curva de produção.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	$\pm 0,5\%$	$\pm 1\%$	$\pm 1\%$	$\pm 2\%$	$\pm 1\%$	$\pm 2\%$	$\pm 2\%$	$\pm 2\%$	$\pm 2\%$	$\pm 3\%$
Teste	$\pm 2\%$	$\pm 4\%$	$\pm 8\%$	$\pm 9\%$	$\pm 9\%$	$\pm 8\%$	$\pm 7\%$	$\pm 8\%$	$\pm 9\%$	$\pm 9\%$

Tabela 4.28 MAD ano a ano obtido na visão *closedloop* do caso MR para a curva de produção.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	121	197	243	261	265	274	279	287	298	307
Teste	180	306	358	365	377	391	385	390	389	399

Incorporando o MAPE da vazão inicial (Tabela 4.16) e considerando que esse valor estaria no tempo zero, o MAPEam e o DPam da série completa é mostrado na Tabela 4.29.

Tabela 4.29. MAPEam e DPam do caso MR (incluindo vazão inicial e curva de produção).

	MAPEam	DPam
Grupo Treino	6,00%	$\pm 0,9\%$
Grupo Teste	8,53%	$\pm 3,81\%$

4.5. Resultados Finais

Conforme descrito, o trabalho foi dividido em duas metodologias diferentes de treino para a obtenção da previsão da vazão inicial e da curva de produção. No caso UR, é utilizada apenas uma rede para prever todos os dados de saída. No caso MR, foi criada uma rede para cada dado de saída (uma para vazão inicial e uma para cada vazão anual ao longo de 10 anos).

Vale destacar aqui a grande precisão da RNA para estimar a vazão inicial de produção de óleo dos poços. No caso de MR, o erro absoluto médio foi de apenas 1,3% para o grupo de treino e 2,2% para o grupo de teste quando comparado com os resultados obtidos pelo simulador numérico IMEX. Na previsão da curva de produção de óleo, foi observado um aumento do MAPE ao longo dos anos em ambos os casos simulados. Isso pode demonstrar que, para se prever um horizonte maior da curva de produção, seria necessário ter mais dados de entrada.

A partir das análises dos diversos cenários estudados acima, percebe-se que, para o presente problema, o caso de MR apresentou o melhor desempenho quando comparamos o indicador MAPEam. Também foi feita uma análise do desvio padrão do MAPE ano a ano. E para que esse valor seja comparável ao MAPEam, utilizou-se o indicador DPam. A Tabela 4.30 apresenta cada um dos casos.

Tabela 4.30 . Comparativo do MAPEam e DPam entre os casos UR e MR.

	MAPEam treino	DPam treino	MAPEam teste	DPam teste
Caso UR	6,23%	$\pm 1,96\%$	9,39%	$\pm 3,58\%$
Caso MR	6,00%	$\pm 0,90\%$	8,53%	$\pm 3,81\%$

Portanto, todas as análises a seguir, serão feitas com base na RNA obtida por meio do caso MR.

Uma outra forma de avaliar as curvas de produção obtidas pela RNA é compará-las com as curvas que seriam obtidas caso elas fossem estimadas utilizando-se o método simplificado de declínio exponencial. De acordo com um relatório de 2019 do *Geological Survey of Finland*, a taxa média mundial de declínio na produção pós-pico é de 5 a 7%. Para essa análise comparativa, foi utilizado uma taxa de declínio de 6% e ela foi aplicada sobre vazão inicial obtida pelo simulador. A Tabela 4.31 mostra o MAPE calculado quando comparamos as curvas obtidas por meio do declínio exponencial com as curvas do simulador numérico e a Tabela 4.32 mostra o seu respectivo desvio padrão.

Tabela 4.31 MAPE ao longo dos anos quando utilizando o método do declínio exponencial.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	4%	7%	11%	16%	19%	22%	26%	30%	33%	38%
Teste	4%	8%	14%	18%	20%	23%	26%	28%	32%	37%

Tabela 4.32 DP do MAPE ao longo dos anos quando utilizando o método do declínio exponencial.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	±0,2%	±1%	±1%	±1%	±1%	±1%	±1%	±2%	±2%	±2%
Teste	±1%	±4%	±5%	±5%	±3%	±3%	±3%	±5%	±5%	±5%

Abaixo será mostrado o MAPE ano a ano comparando as duas metodologias.

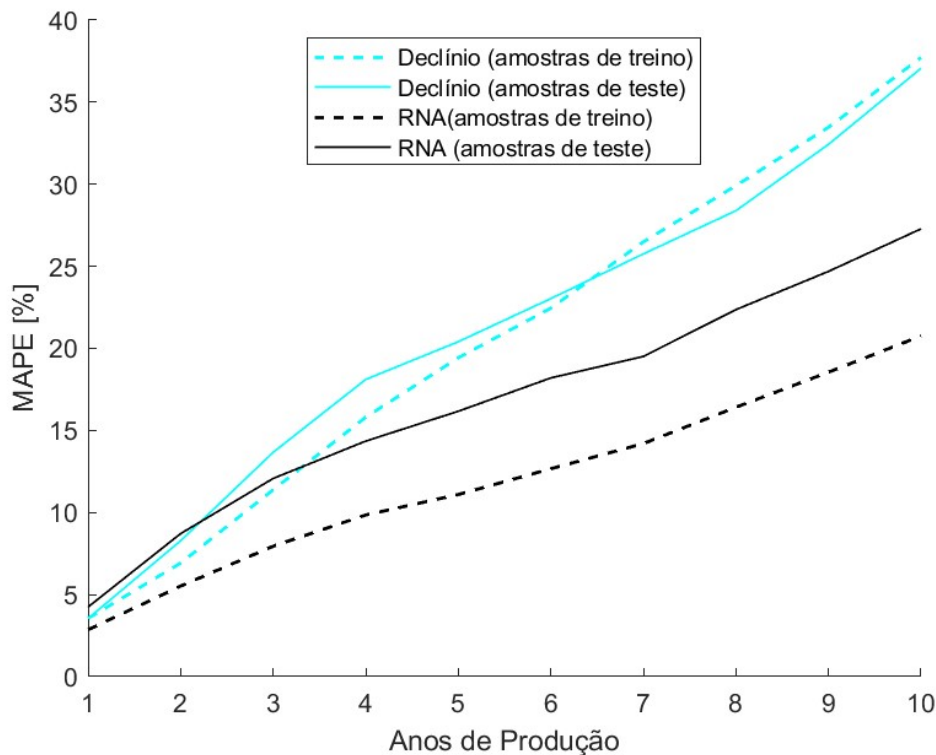


Figura 4.15 Comparativo do MAPE entre o método do declínio exponencial e o método de RNA para MR.

De acordo com a Figura 4.15, percebe-se que nos primeiros anos de produção, ambos os métodos são similares. Contudo, à medida que o horizonte de análise cresce, o MAPE da curva de produção gerada a partir da RNA, é bem inferior àquele calculado para a curva de produção estimada via declínio exponencial.

Em relação ao MAPE_{am}, foi assumida que a vazão inicial no método do declínio exponencial é conhecida e, por isso, foi usada vazão inicial calculada pelo simulador. Dessa forma, o MAPE do Q_{oi} é zero. Conectando essa informação ao MAPE da Tabela 4.31, chega-se ao MAPE_{am} do declínio exponencial conforme mostrado abaixo

Tabela 4.33. Comparativo do MAPEam e DPam entre o método do declínio exponencial e o método de RNA para o caso MR.

	MAPEam treino	DPam treino	MAPEam teste	DPam teste
Declínio Exponencial	9,94%	$\pm 0,55\%$	10,26%	$\pm 2,15\%$
RNA - Caso MR	6,00%	$\pm 0,90\%$	8,53%	$\pm 3,81\%$

A partir desse comparativo, percebe-se que a metodologia proposta de geração de curva de produção por meio de RNA possui um ganho de precisão em relação à metodologia de declínio exponencial.

Com o mesmo intuito de melhorar a avaliação das curvas de produção obtidas pela RNA, também foi implementada a técnica estatística de regressão linear múltipla para obter a vazão inicial e as curvas de produção. A regressão foi estimada por meio do módulo de análise de dados do *software* Excel. Aplicou-se a metodologia *10-folds* e também houve a separação em grupo de teste e treino. A regressão foi gerada utilizando o grupo de treino e, em seguida, aplicada ao grupo de teste. Ao final, foi feita uma média dos resultados de todos os *folds*. Os resultados obtidos para a estimativa da vazão inicial estão descritos na Tabela 4.34:

Tabela 4.34. Comparação entre os resultados da RNA e regressão linear múltipla para a obtenção da vazão inicial.

	RNA	Regressão linear múltipla
MAPE Treino	1,3%	2,8%
MAPE Teste	2,2%	3,3%

Para a previsão da curva de produção por meio da regressão linear múltipla, utilizou-se a metodologia utilizada pela abordagem do caso MR. Os resultados, tanto para o grupo de teste como para o grupo de treino, foram resumidos na Tabela 4.35 e na Tabela 4.36.

Tabela 4.35 MAPE ao longo dos anos quando utilizada a regressão linear múltipla.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	5%	9%	12%	16%	18%	19%	22%	24%	26%	27%
Teste	5%	10%	15%	19%	22%	24%	27%	30%	32%	34%

Tabela 4.36 DP do MAPE ao longo dos anos quando utilizada a regressão linear múltipla.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10
Treino	±0,1%	±1%	±1%	±1%	±2%	±2%	±2%	±1%	±1%	±1%
Teste	±0,8%	±1%	±3%	±6%	±10%	±13%	±12%	±11%	±10%	±9%

Ao comparar esses resultados com os obtidos pela RNA no caso MR, percebe-se por meio da Tabela 4.37 que a metodologia da regressão linear múltipla possui um erro superior ao da RNA. A estimativa da vazão inicial e da curva de produção por meio da RNA não necessita de linearidade entre as variáveis, isso é uma vantagem e que pode justificar os seus resultados terem sido melhores do que os obtidos pela regressão linear múltipla.

Tabela 4.37. Comparativo do MAPEam e DPam entre o método da regressão linear múltipla e o método de RNA para o caso MR.

	MAPEam treino	DPam treino	MAPEam teste	DPam teste
Regressão linear múltipla	9,10%	±0,75%	11,10%	±4,38%
RNA - Caso MR	6,00%	±0,90%	8,53%	±3,81%

Uma outra forma de analisar o desempenho dos resultados obtidos com a RNA é verificando o impacto na produção acumulada de óleo do poço ao final do décimo ano. Apesar de percebermos um aumento do erro percentual médio absoluto ao longo dos anos, os maiores desvios se dão nos últimos anos em que a vazão de óleo já é bem inferior à dos primeiros anos. Para exemplificar esse ponto, calculou-se uma curva média obtida a partir dos dados reais usados no grupo de treino (de todos os *folds*) e uma outra curva média dos dados reais do grupo de teste. A partir dessas curvas, calculou-se a produção acumulada de óleo em milhões de barris [MMbbl] no décimo ano. Para estimar uma curva média gerada pela RNA na visão *closedloop*, foi aplicado o MAPE sobre a curva média dos dados reais. Como o MAPE é um indicador que mede o erro médio percentual em termos absolutos, a curva média gerada pela RNA deverá ser informada como uma dispersão. Seja $q_t^{(real)}$ a vazão média obtida dos dados reais no ano t , logo, a vazão média da RNA $\left(q_t^{(RNA)}\right)$ no ano t poderá variar entre $q_t^{(real)}(1 \pm MAPE_t)$. Ou seja,

$$\text{Limite Superior } q_t^{(RNA)} = q_t^{(real)}(1 + MAPE_t), \quad (4.9)$$

$$\text{Limite Inferior } q_t^{(RNA)} = q_t^{(real)}(1 - MAPE_t). \quad (4.10)$$

Esse cálculo foi utilizado para a o grupo de treino e teste e suas curvas podem ser visualizadas na Figura 4.16 e Figura 4.17, respectivamente.

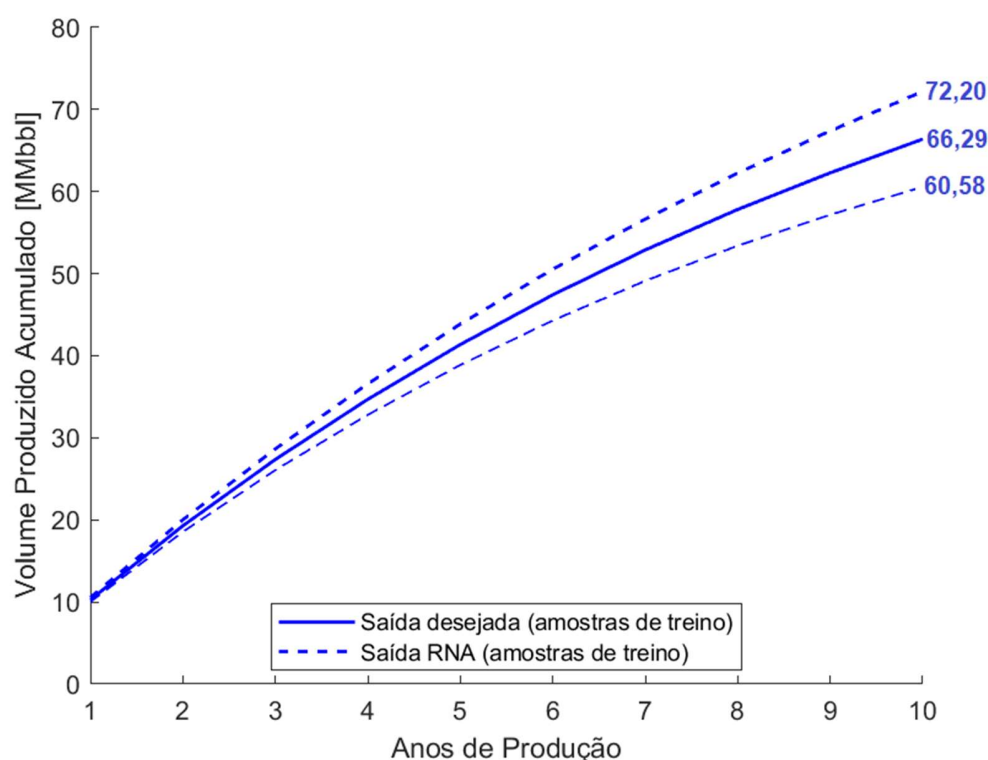


Figura 4.16 Comparação da produção acumulada [MMbbl] obtida de uma curva média dos poços usados para treinamento da RNA, com a produção acumulada estimada pela RNA em modo *closedloop* para esse grupo de treino.

Conforme mostrado no gráfico da Figura 4.16, a produção acumulada média em 10 anos obtida pela RNA para o grupo de treino tem um erro médio percentual de 8,8% quando comparada com a produção média dos dados reais do simulador. A curva de produção média dos poços do grupo de treino obtida do simulador teria uma produção acumulada média em 10 anos de 66,39 MMbbl enquanto a curva de produção média dos poços estimados pela RNA teria uma produção acumulada média variando entre 72,20 - 60,58 MMbbl no ano 10. A seguir, será feita essa mesma análise com o grupo de teste.

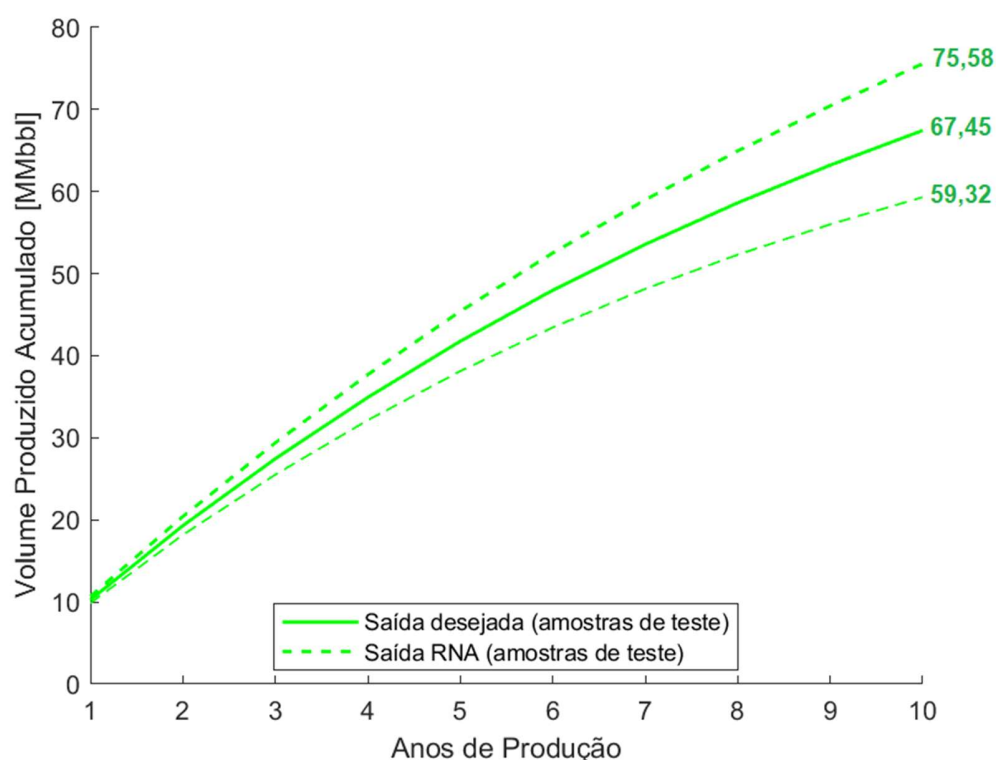


Figura 4.17 Comparação da produção acumulada [MMbbl] obtida de uma curva média dos poços usados para testar a RNA, com a produção acumulada estimada pela RNA em modo *closedloop* para esse grupo de teste.

Conforme mostrado na Figura 4.17, o erro médio absoluto da produção acumulada no último ano é 12% para o grupo de teste. A curva de produção média dos poços do grupo de teste obtida do simulador teria uma produção acumulada em 10 anos de 67,45 MMbbl enquanto a curva de produção média dos poços estimados pela RNA teria uma produção acumulada média de variando entre 75,58 – 59,32 MMbbl no ano 10.

Uma outra questão importante é avaliar o comportamento do MAD [m^3/d] ao longo dos anos. Para complementar a análise, foi calculado o desvio padrão do MAD dentro dos 10 *folds* utilizados para treino e teste de cada ano. Na Figura 4.18, a curva representa o MAD e as barras verticais o seu desvio padrão por ano. O MAD segue a

mesma conclusão do MAPE, em que há uma tendência de aumento do erro ao longo dos anos.

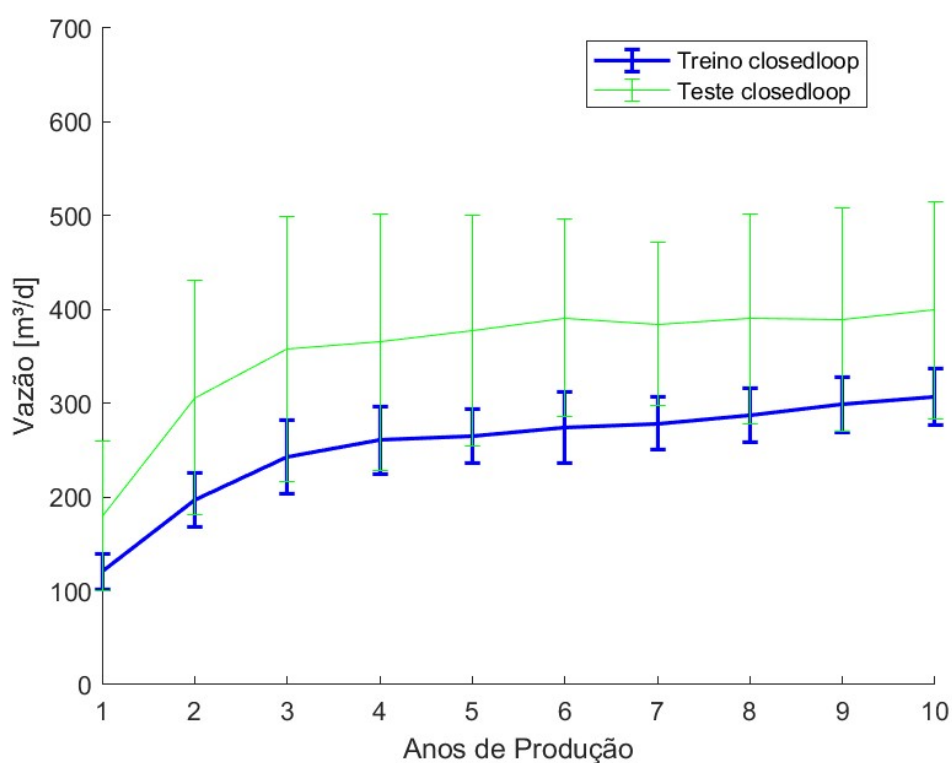


Figura 4.18 Comparação entre o MAD [m³/d] e seu DP na visão *closedloop* da RNA da curva de produção.

Sabendo da complexidade do modelo matemático por trás de um simulador numérico, o trabalho buscou ter resultados bem ambiciosos ao tentar se aproximar da resposta do simulador por meio de 10 parâmetros de entrada. Sabe-se que simuladores são modelos numéricos bastante complexos e, principalmente para esses casos, uma RNA apresentará melhores resultados quanto mais informações forem fornecidas. Portanto, acredita-se que com a inclusão de informações de novos poços no banco de dados será possível melhorar consideravelmente a resposta da RNA.

5. Conclusão

Quando um projeto está na fase exploratória, o conjunto de informações disponíveis sobre o reservatório é restrito de forma que, é necessário assumir diversas simplificações quanto à sua caracterização. A possibilidade de ter uma ferramenta de geração de curva de produção que se aproxime de um simulador numérico e que tenha uma resposta rápida tem muito a agregar.

Esta dissertação apresentou uma nova metodologia de geração de curva de produção de óleo para ativos exploratórios a partir de Redes Neurais Artificiais. Tal método tem grande relevância na geração da curva de produção de um poço produtor de óleo, principalmente, por trazer a robustez de um simulador numérico de reservatório em um tempo consideravelmente menor. Enquanto no simulador numérico as simulações duraram em torno de 1 a 2 horas, com as redes neurais o tempo gasto ficou em torno de 10-15 segundos. Uma outra vantagem na utilização de RNAs na previsão da curva de produção é que esses modelos são capazes de encapsular a complexidade do problema tornando a aplicação bastante ampla. Esse encapsulamento da complexidade torna a utilização da rede treinada acessível a usuários com diversos níveis de conhecimento sobre o tema. Contudo, é sempre bom reforçar a importância de uma análise crítica dos resultados por profissionais qualificados.

É importante reforçar que tal metodologia não substitui os atuais métodos de geração de curva de produção. Cada um tem seu valor e caberá ao avaliador a escolha do método que julgar ser o mais adequado naquele momento.

Para a obtenção dos resultados, foram assumidas algumas simplificações como: os dados de entrada ficaram restritos às condições iniciais da simulação, não foi avaliada a interação entre um poço e qualquer produção/injeção próxima, não foi considerado o efeito do limite do reservatório e os estudos englobam apenas poços produtores de óleo.

Este trabalho utilizou a biblioteca *Neural Network Toolbox* do MATLAB e usou para treinamento das redes informações coletadas do simulador numérico IMEX (CMG, 2015), desenvolvido pela empresa CMG. Conclui-se que, para esse problema, o algoritmo de aprendizado Levenberg–Marquardt e a normalização do tipo Min-Max mostraram-se mais adequadas. Além disso, foram testadas duas abordagens de estruturação da RNA para o treinamento: UR e MR. A que se mostrou mais adequada ao problema estudado foi a MR, em que, para a previsão de vazão do ano t , é utilizada a estimativa da vazão do ano $t - 1$.

Os resultados se mostraram bastante satisfatórios na previsão da vazão inicial de produção de óleo com um erro percentual médio absoluto variando de 1,3% a 2,2%. Para a previsão da curva de produção de óleo, a RNA conseguiu um erro percentual médio absoluto inferior a 10% nos anos iniciais, porém esse erro foi crescente ao longo dos anos. Foram geradas curvas de produção a partir do método declínio exponencial e comparadas com as curvas obtidas pelo simulador numérico IMEX. O erro percentual médio absoluto do declínio exponencial foi superior ao da RNA. Concluiu-se que, para os poços analisados, a metodologia baseada em RNA apresentou um ganho de precisão em relação a esse método analítico. Os resultados também foram comparados com uma metodologia bastante utilizada na literatura que é a regressão linear múltipla. Tanto para a estimativa da vazão inicial como para a previsão da curva de produção, a metodologia da RNA apresentou erros menores do que a regressão. Isso evidencia uma vantagem do modelo da RNA que é a não necessidade de linearidade entre as variáveis.

Devido à complexidade de um simulador numérico de reservatório, há diversas possibilidades de aprimorar os estudos relacionados ao tema. Estabelecer um procedimento iterativo para obtenção de curvas de produção englobando mais dados e mais variáveis de entrada poderá permitir uma maior adequação e generalização do método. Além disso, com o aumento do banco de dados, mais tipos de fluidos serão englobados e, com isso, pode ser necessário incluir mais características do fluido como entrada. Portanto, como sugestão para trabalhos futuros, propõe-se incluir novos parâmetros de entrada, coleta de dados de outros modelos de simulação e estudar formas de estender a curva de produção por um horizonte de longo prazo. Além disso,

pode-se verificar como os resultados se comportariam caso fosse utilizada a abordagem MR com outros períodos de recorrência, como por exemplo, utilizar as estimativas da vazão dos anos $t - 2$ e $t - 1$ para prever a vazão do ano t .

Referências Bibliográficas

ALI, J. K., **Neural Networks: A New Tool for the Petroleum Industry?**. Society of Petroleum Engineers. European Petroleum Computer Conference, Aberdeen, Reino Unido, 15-17 Março de 1994. Disponível em <<https://www.onepetro.org/conference-paper/SPE-27561-MS>>. Acesso em 21 jun. 2020

ALMEIDA, L. F. **Sistema Híbrido de Otimização de Estratégias de Controle de Válvulas de Poços Inteligentes sob Incertezas**. Tese de Doutorado - Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio, Rio de Janeiro, 2007.

AL-FATTAH, S. M., & STARTZMAN, R. A., **Predicting Natural Gas Production Using Artificial Neural Network**. Society of Petroleum Engineers. SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas, 2-3 de abril de 2011. Disponível em < <https://www.onepetro.org/conference-paper/SPE-68593-MS>>. Acesso em 16 jun. 2020.

ARAUJO JUNIOR, A. D., **Predição não-linear de curvas de produção de petróleo via redes neurais recursivas**. Dissertação de Mestrado – Programa de Pós-Graduação em Ciência e Engenharia de Petróleo Universidade Federal do Rio Grande do Norte, Natal, 2010.

ARPS, J. J. **Analysis of Decline Curves**, Houston, EUA, 1945.

BASHEER, I.; HAJMEER, M. **Artificial neural networks: fundamentals, computing, design, and application**. Journal of Microbiological Methods, v. 43, p. 3–31, 2000.

BENINI, F. A. V., **Rede neural recorrente com perturbação simultânea aplicada no problema do caixeiro viajante**. Dissertação (Mestrado) -Programa de Pós-

Graduação em Engenharia. Escola de Engenharia de São Carlos da Universidade de São Paulo. São Carlos, 2008.

BISHOP, C. M., **Pattern Recognition and Machine Learning**. 1 ed. Nova York, Springer, 2006. 738p.

CAO, Q., et al., **Data Driven Production Forecasting Using Machine Learning**. Society of Petroleum Engineers. SPE Argentina Exploration and Production of Unconventional Resources Symposium, 1-3 Junho 2016, Buenos Aires, Argentina. Disponível em: < <https://www.onepetro.org/conference-paper/SPE-180984-MS>>. Acesso em 09 ago. 2020.

CMG. **IMEX Black-Oil Reservoir Simulator Version 2015 User's Guide**. Computer Modelling Group LTD., Calgary, Canada, 1214pp. 2015

ERTEKIN, T., ABOU-KASSEM, J. H., KING, G. R., **Basic Applied Reservoir Simulation**. Richardson, Texas: Society of Petroleum Engineers, Textbook Series, v. 7, 2001.

GOMES, D. T., **Modelos de redes neurais recorrentes para previsão de séries temporais de memórias curta e longa**. Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica. Campinas, 2005.

HAYKIN, S. **Redes Neurais: Princípios e Prática**, 2. ed. Porto Alegre: Bookman, 2001.

JIA, X., ZHANG, F., **Applying Data-Driven Method to Production Decline Analysis and Forecasting**. Society of Petroleum Engineers. SPE Annual Technical Conference and Exhibition, 26-28 Setembro de 2016, Dubai, UAE. Disponível em: < <https://www.onepetro.org/conference-paper/SPE-181616-MS>>. Acesso em 20 jun. 2020.

KOHAVER, R., **A study of cross-validation and bootstrap for accuracy estimation and model selection**. International Joint Conference on Artificial Intelligence (IJCAI), Stanford, California, 1995.

KOHLER, M. R., **Redes Neurais Aplicadas à Construção de Aproximadores para Simulação Integrada entre Reservatório e Sistema de Produção**. Dissertação (Mestrado) –Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica. Rio de Janeiro, 2013.

MAGALHAES, R. J. R. de O., **Avaliação Do Impacto Econômico Do Desenvolvimento Da Produção Offshore**. Dissertação (mestrado) – Programa de Pós-Graduação em Engenharia Mecânica. Pontifícia Universidade Católica do Rio de Janeiro - PUC-RIO. Rio de Janeiro, 2007.

MANTOVANI, W. A., **Utilização de redes neurais recorrentes na caracterização de cargas não lineares em sistemas elétricos**. Dissertação (mestrado) - Universidade Estadual Paulista. Faculdade de Engenharia Elétrica. Ilha Solteira, São Paulo, 2011

MARSILI, M. D., **Simulação numérica da injeção de água como método de recuperação suplementar para um campo com alto grau de exploração**. Monografia (Graduação) - Curso de engenharia do petróleo. Escola politécnica da universidade federal do rio de janeiro. Rio de Janeiro, 2008.

MICHAUX, Simon. **Oil from a Critical Raw Material Perspective**. Espoo: *Geological Survey of Finland*, 2019. 510p. (Report 70/2019).

OBIDIKE, P., EMELLE, C., VLAARDINGERBROEK, R., **Integrated Production Forecasting for Bussiness Planning**. Society of Petroleum Engineers. SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 5-7 de agosto de 2014. Disponível em: <<https://www.onepetro.org/conference-paper/SPE-172479-MS>>. Acesso em 16 jun. 2020.

OYEKUNLE, A. A. **Impact of the Petroleum Industry Bill on Deepwater Economics**. Society of Petroleum Engineers. SPE Nigeria Annual International Conference and Exhibition, Abuja, Nigeria, 30 jul – 3 ago de 2011. Disponível em: <<https://www.onepetro.org/download/conference-paper/SPE-150774-MS>>. Acesso em 10 nov. 2020.

PERALES, T., **Utilização de Redes Neurais Artificiais no Diagnóstico de Cardiopatias**. Dissertação (Mestrado) –Universidade Estadual de Londrina. Departamento de Engenharia Elétrica, Londrina, 2011.

PETROBRAS. Fatos e dados: **A descoberta de um campo de petróleo e gás natural em 5 passos**, 20 de fev de 2015. Disponível em: <<https://petrobras.com.br/fatos-e-dados/a-descoberta-de-um-campo-de-petroleo-e-gas-natural-em-5-passos.htm>>.

Acesso em 10 jun. 2020.

POSTALI, F. A. S., **Regime fiscal, investimento em petróleo e opções reais**. Economia Aplicada, v. 13, n. 2, 2009, pp. 207-230. Disponível em <<https://doi.org/10.1590/S1413-80502009000200002>>. Acesso em 23 ago. 2020.

RAMGULAM, A., ERTEKIN, T., FLEMINGS, P. B., **An Artificial Neural Network Utility for the Optimization of History Matching Process**. Society of Petroleum Engineers. Latin American & Caribbean Petroleum Engineering Conference, 15-18 Abril de 2007, Buenos Aires, Argentina. Disponível em: <<https://www.onepetro.org/conference-paper/SPE-107468-MS>>. Acesso em 13 jun. 2020.

REBESCHINI, J., QUERALES, M., CARVAJAL, G. A., VILLAMIZAR, M., ADNAN, F. M., RODRIGUEZ, J., KNABE, S., RIVAS, F., SAPUTELLI, L., AL-JASMI, A., NASR, H., GOEL, H. K., **Building Neural-Network-Based Models Using Nodal and Time-Series Analysis for Short-Term Production Forecasting**. Society of Petroleum Engineers. SPE Middle East Intelligent Energy Conference and Exhibition, Manama, Bahrain, 28-30 Outubro de 2013. Disponível em: <<https://www.onepetro.org/download/conference-paper/SPE-167393-MS>>. Acesso em 20 jun. 2020.

ROSA, A. J., CARVALHO, R. S., XAVIER, J. A. D., **Engenharia de reservatório de petróleo**. Editora Interciência, 2006. Rio de Janeiro (RJ, Brasil).

SHAEFFER, A. K., WILSON, W., YANG, C. **Application of Machine Learning to Early-Stage Hull Form Design**. The Society of Naval Architects and Marine Engineers. SNAME Maritime Convention, 2 outubro de 2020, virtual. Disponível em:

<<https://www.onepetro.org/conference-paper/SNAME-SMC-2020-098>>. Acesso em 20 out. de 2020

SILVA, L. N. de C., **Industrial Análise e Síntese de Estratégias de Aprendizado para Redes Neurais Artificiais**. Universidade Estadual de Campinas (UNICAMP), Faculdade de Engenharia Elétrica e de Computação. Campinas, Brasil, 1998.

SILVA, L. C. F. da, **Inteligência Computacional para Predição de Produção de Reservatórios de Petróleo**. Tese de Doutorado - Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, 2006.

SOUZA, I. T., SILVA, L. A., **Estudo de viabilidade econômica de projetos para desenvolvimento da produção de campos petrolíferos com alto grau de exploração**. In: XXXVI ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO (ENEGEP), João Pessoa, Brasil, 2016.

TEIXEIRA, F. E. de S., **Implementação de um Modelo em Elementos Finitos para a Avaliação do Comportamento Geomecânico na Estabilidade de Poços de Petróleo**. Dissertação (mestrado) – Programa de Engenharia Civil, Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, 2011.

THOMAS, J. E., **Fundamentos de Engenharia de Petróleo**. 2. ed. Rio de Janeiro: Editora Interciência, 2001.

TÚPAC, Y. J. V., **Sistema Inteligente de Otimização de Alternativas de Desenvolvimento de Campos Petrolíferos**. Tese (Doutorado) –Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, 2005.

VIEIRA SOBRINHO, J. D., **Matemática Financeira**. Editora Atlas, 1997.

XAVIER, C. R., **Comparação de métodos de otimização para o problema de ajuste de histórico em ambientes paralelos**. Dissertação (mestrado) – Universidade Federal de Juiz de Fora, Juiz de Fora, 2009.

ZANG, D., LIN, J., PENG, Q., WANG, D., YANG, T., SOROOSHIAN, S., LIU, X., ZHUANG, J., **Modeling and simulating of reservoir operation using the artificial**

neural network, support vector regression, deep learning algorithm. Journal of Hydrology, v. 565, p. 720-736, 2018. Disponível em:<<https://doi.org/10.1016/j.jhydrol.2018.08.050>>. Acesso em 09 ago. 2020.