



**Vitor Hugo Pinheiro Marques**

**Modelos de simulação para análise de incerteza na previsão  
de produção de óleo em plataformas da bacia de Campos**

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia de Produção, do Departamento de Engenharia Industrial da PUC-Rio.

Orientador: Prof. Fernando Luiz Cyrino Oliveira  
Coorientador: Prof. Antônio Orestes de Salvo Castro

Rio de Janeiro  
agosto de 2023



**Vitor Hugo Pinheiro Marques**

**Modelos de simulação para análise de incerteza na previsão  
de produção de óleo em plataformas da bacia de Campos**

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

**Prof. Fernando Luiz Cyrino Oliveira**

Orientador

Departamento de Engenharia Industrial - PUC-Rio

**Prof. Antônio Orestes de Salvo Castro**

Coorientador

UERJ

**Prof. Paula Medina Maçaira Louro**

Departamento de Engenharia Industrial – PUC-Rio

**Prof. José Francisco Moreira Pessanha**

UERJ

Rio de Janeiro, 31 de agosto de 2023.

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

### **Vitor Hugo Pinheiro Marques**

Graduou-se em Engenharia de Produção pela Universidade Federal do Rio Grande do Norte – UFRN em 2012. É engenheiro de produção na Petrobras desde 2012, atuando atualmente na área de exploração e produção, com ênfase em planejamento e controle da produção de E&P.

#### Ficha Catalográfica

Marques, Vitor Hugo Pinheiro

Modelos de simulação para análise de incerteza na previsão de produção de óleo em plataformas da bacia de Campos / Vitor Hugo Pinheiro Marques; orientador: Fernando Luiz Cyrino Oliveira; coorientador: Antônio Orestes de Salvo Castro. – 2023.

68 f.: il. color.; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2023.

Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Produção de petróleo. 3. Análise de incerteza na previsão de produção. 4. Simulação. 5. Séries temporais. 6. Predição com julgamento humano. I. Oliveira, Fernando Luiz Cyrino. II. Castro, Antônio Orestes de Salvo. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. IV. Título.

CDD: 658.5

## **Agradecimentos**

Agradeço aos meus pais que me educaram e ensinaram a sempre buscar maiores conhecimentos.

À minha esposa Beatriz pelo apoio constante, fundamental para essa realização, e à minha filha Melina, benção de Deus que nasceu no período desta pesquisa.

Agradeço à toda minha família, amigos e a todas as pessoas que incentivaram na realização deste trabalho.

Ao meu orientador Fernando Luiz Cyrino, meu agradecimento por seu profissionalismo e contribuições que me ajudaram na construção desse trabalho.

Ao meu coorientador Antônio Orestes de Salvo Castro, agradeço o incentivo e pelas ideias que nortearam o desenvolvimento do trabalho.

À Petrobras, por incentivar a capacitação dos seus funcionários, com destaque aos gestores e funcionários da minha área.

Aos professores e demais funcionários da PUC-Rio, pelos ensinamentos, orientações, organização e convivência ao longo 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

Marques, Vitor Hugo Pinheiro; Oliveira, Fernando Luiz Cyrino (Orientador); Castro, Antônio Orestes de Salvo (Coorientador). **Modelos de simulação para análise de incerteza na previsão de produção de óleo em plataformas da bacia de Campos.** Rio de Janeiro, 2023. 68 p. Dissertação de Mestrado - Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro

A produção de petróleo possui alta relevância em âmbito brasileiro e mundial. Por outro lado, a incerteza do setor presume alta variabilidade nas previsões de produção de óleo, e exerce um impacto significativo nas decisões. O estudo contempla analisar o cenário da bacia geográfica de Campos, em estudo de caso aplicado em empresa nacional de óleo e gás, com objetivo de aprimorar a previsão de produção de óleo. Para isso, são empregados métodos de simulação, clusterização e previsão, sendo integrados com julgamento humano. Busca-se inferir as incertezas inerentes às atividades, analisar os principais riscos envolvidos e subsidiar a definição das metas de produção. Com esse intuito, foi desenvolvida uma modelagem orientada a dados, por meio da criação de um simulador com linguagem de programação em R. Os dados compreendem os anos de 2017 a 2021, e a projeção é realizada para o ano de 2022. O modelo incorpora julgamento humano durante o processo, permitindo que os especialistas realizem modificações no resultado das previsões, agregando sua experiência e informações exclusivas. A análise de série temporal avalia oito métodos de previsão, seu resultado mostra que a entidade do potencial produtivo apresenta menor erro do que na eficiência, e o método TBATs obteve o menor erro na predição. A análise do planejamento das paradas e entrada dos novos poços é realizada por meio de análise gráfica, observando os principais riscos relacionados. Por fim, o simulador apresenta proposta para auxiliar na definição das metas de produção, ele verifica a probabilidade para alcançar a meta com base nos resultados das simulações.

## Palavras-chaves

Produção de Petróleo; Análise de incerteza na previsão de produção; Simulação; Séries Temporais; Predição com julgamento humano.

## Abstract

Marques, Vitor Hugo Pinheiro; Oliveira, Fernando Luiz Cyrino (Orientador); Castro, Antônio Orestes de Salvo (Coorientador). **Simulation models for uncertainty analysis in oil production forecasting on platforms in the Campos basin.** Rio de Janeiro, 2023. 68 p. Dissertação de Mestrado - Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro

Oil production has Brazilian and World importance. However, the randomness of the sector results a high variability in oil production forecasts. This variability has a significant impact on decisions. The study analyzes the challenging scenario at geographic Campos basin, in a case applied in a national energy company. The objective is to improve the risk analysis associated with the achievement of oil production targets. Simulation, clustering, and time series forecasting methods are employed, integrating into human judgment. It tries to infer the uncertainties inherent of the activities to increase the accuracy of oil production forecasts, analyze the main risks involved, and subsidize the definition of production targets. A data-driven model is developed, creating a simulator with R language. The data used the years 2017 to 2021, and the projection is made for the year 2022. Human judgment is incorporated into the model during the process, specifying the input parameters to enable experts to make modifications based on the predictions, adding their unique experience and information. The time series analysis eight prediction methods, the results show that the oil potential presents less error than in the production efficiency, and TBATS was the prediction method that obtained the lowest prediction error. The main risks related to the maintenance planning and the entry of new wells are identified through graphical analysis. Finally, the simulator presents a possible solution to help define production goals, it verifies the probability of reaching the goal based on the simulation results.

## Keywords

Oil Production; Production forecasting uncertainty analysis; Simulation; Time Series; Prediction with human judgments.

# Sumário

1. Introdução	11
1.1 Objetivo Geral	12
1.2 Objetivos Específicos	12
1.3 Delimitação do estudo e justificativa	13
1.4 Estrutura da dissertação	15
2. Revisão bibliográfica	17
2.1 Modelagem da Incerteza e simulação em exploração e produção	17
2.2 Análise de risco da previsão de produção de óleo	18
2.3 Simulação com análise preditiva e julgamento humano	20
3. Metodologia	23
3.1 Metodologia de pesquisa	23
3.2 Metodologia para cálculo da produção probabilística	24
3.3 Modelo atual de análise de risco da previsão da produção	27
3.4 Metodologia de desenvolvimento do aplicativo	29
3.4.1 Objetivo do domínio	29
3.4.2 Pré-processamento	29
3.4.3 Modelagem da Simulação	30
3.4.4 Plano por entidades	33
3.4.5 Desenvolvimento da aplicação	34
3.4.6 Avaliação dos Modelos	35
4. Desenvolvimento da aplicação	36
4.1. Potencial Implantado e Eficiência Operacional	36
4.2. Poços Novos	39
4.3. Paradas programadas para manutenção	41
5. Apresentação do aplicativo	44
5.1 Disponibilizar aplicação	44
5.2 Analisar dados de entrada	46
5.3 Simular Entidades	47
5.4 Simular Produção	51
6. Análise dos resultados	54
6.1 Comparação dos erros das séries temporais e produção	54
6.2 Análise dos principais riscos	55
6.3 Definição da meta anual	58
7. Considerações Finais	61
Referências	63
Apêndice	66

## Lista de Figuras

Figura 1 – Modelo conceitual da metodologia de previsão da produção de petróleo. ....	27
Figura 2 – Estruturação do modelo para desenvolvimento do aplicativo .....	29
Figura 3 – Modelagem de simulação do aplicativo .....	31
Figura 4 – Etapas do plano por entidade para simulação do Modelo .....	33
Figura 5 – Exemplos de cenários nos experimentos: Método MLP (esq.) e RBF (dir.) ..	38
Figura 6 – Número ótimo de clusters: Silhouette (esq.); Gap stat (centro); WSS (dir.)...	41
Figura 7 – Plataformas clusterizadas pelos eventos de paradas programadas .....	42
Figura 8 – Tela inicial da aplicação desenvolvida em <i>Shiny</i> .....	45
Figura 9 – Menu de opções do aplicativo.....	45
Figura 10 – Dados históricos do potencial de produção da bacia geográfica de Campos.	46
Figura 11 – Eficiência Histórica e Poços Novos da bacia geográfica de Campos .....	47
Figura 12 – Total histórico de Perda (colunas) e Duração (linhas) por Plataforma .....	47
Figura 13 – Parâmetros disponíveis para a simulação das entidades.....	49
Figura 14 – Simulação das entidades com análise de séries temporais: eficiência .....	50
Figura 15 – Simulação das entidades com análise de séries temporais: potencial .....	51
Figura 16 – Simulação da Produção: comparar erro pelo número de iterações.....	51
Figura 17 – Simulação da Produção: mesclar métodos de previsão .....	52
Figura 18 – Simulação da Produção: julgamento dos especialistas nos dados de saída ..	53
Figura 19 – Comparativo entre os erros (MAPE) das previsões .....	54
Figura 20 – Principais riscos para as paradas planejadas.....	56
Figura 21 – Principais riscos para os poços novos .....	57
Figura 22 – Parâmetros para os poços novos.....	57
Figura 23 – Probabilidade de superar a meta de 559 Mbpd em relação a simulação .....	59
Figura 24 – Probabilidade de superar a meta de 526 Mbpd em relação a simulação .....	59



## Lista de Tabelas

Tabela 1 - Métodos de integração entre julgamento humano e análise preditiva.....	21
Tabela 2 - Plano de experimentos para identificação dos hiperparâmetros ótimos .....	37
Tabela 3 – Aplicação da abrangência dos experimentos: Exemplo na eficiência .....	38
Tabela 4 – Resultados dos experimentos para selecionar métodos de predição .....	39
Tabela 5 – Resultados da variação entre simulação e realização no ano .....	60

## Lista de Equações

Equação 1 – Cálculo da entidade potencial produtivo implantado.....	25
Equação 2 – Cálculo da entidade eficiência operacional. ....	25
Equação 3 – Cálculo utilizado na simulação da data de entrada do poço novo .....	25
Equação 4 – Cálculo utilizado na simulação da perda do evento de parada. ....	26
Equação 5 – Cálculo da produção a partir das entidades de entrada do modelo.....	26
Equação 6 – Cálculo da data conservadora para a entrada de poços novos .....	58
Equação 7 – Cálculo da data otimista para a entrada de poços novos .....	58

# 1. Introdução

A formação do portfólio de uma empresa de Petróleo envolve complexas decisões devido ao ambiente de incertezas e é de extrema importância na definição do futuro estratégico (ABATH, 2020).

Para uma empresa do setor, explorar óleo, prever a vazão dos reservatórios, implantar e manter os equipamentos associados às plataformas e poços marítimos representam a resolução de problemas complexos em exploração e produção (E&P). Todo esse gerenciamento caracteriza um ambiente de alta incerteza.

A necessidade de quantificar as incertezas em projetos de exploração da produção surge naturalmente a partir do processo de modelagem, pois é impossível representar com precisão todas as propriedades de um sistema dinâmico real (GUIMARÃES, 2016).

O presente estudo avalia as incertezas geológicas do reservatório e operacionais de produção, envolvendo necessidade de conhecimento técnico a fim de resolver problemas complexos em E&P.

A partir do progresso de ferramentas analíticas, essa pesquisa estabelece procedimentos para gerenciar riscos associados à previsão de produção de óleo, com foco em aplicações de simulação, análise de séries temporais, e clusterização.

Segundo Bomfim (2017) a previsão de produção é uma das etapas mais críticas do processo de planejamento de curto prazo das empresas de exploração e produção de petróleo (E&P). O volume de petróleo que será produzido, denominado meta de produção, influencia diretamente todas as ações das empresas e tem um crítico impacto em relação ao mercado.

O alto número de parâmetros a serem considerados na simulação do fluxo de reservatórios para mensurar a vazão de produção, aliada à complexidade substancial dos equipamentos inerentes a um sistema de produção, que engloba a plataforma, seus poços e a infraestrutura para o escoamento do óleo, contribuem para uma

variabilidade significativa nas projeções de produção. Essa variabilidade, combinada ao substancial montante de investimento envolvido, ocasiona repercussões em decisões de natureza tática e estratégica, as quais se encontram intrinsecamente vinculadas às estimativas de produção e às metas estabelecidas.

A referida variabilidade eleva a necessidade de empregar métodos de previsão estocástica, e requer uma análise criteriosa das incertezas que permeiam as variáveis de entrada para diferentes plataformas de produção. É fundamental explorar os dados para verificar as entidades de entrada que possuem maior correlação com as oscilações na produção, e, enfim, disponibilizá-los em um modelo objetivo, a fim de: obter previsões assertivas, avaliando a adequabilidade entre diferentes métodos de previsão; mapear os principais riscos associados, facilitando interpretar os padrões de oportunidades e ameaças relacionados aos parâmetros do simulador; e subsidiar a definição da meta de produção, identificando a probabilidade de ocorrência das metas pretendidas em relação a distribuição da simulação.

### **1.1 Objetivo Geral**

O trabalho objetiva desenvolver modelagem baseada em dados, identificando incertezas, com o intuito de subsidiar a definição de metas e aumentar a acurácia das previsões de produção de óleo em plataformas marítimas da bacia de Campos.

### **1.2 Objetivos Específicos**

O objetivo geral pode ser traduzido para os seguintes objetivos específicos:

- Desenvolver modelagem de simulação baseado em dados históricos de potencial de produção, eficiência operacional, paradas para manutenção, e de projeções em poços novos;
- Inferir principais incertezas e riscos associados aos dados de entrada;
- Avaliar modelos de simulação que utilizam diferentes métodos de previsão, objetivando utilizar modelos com maior acurácia;
- Integrar análise preditiva com julgamento humano, diminuindo subjetividade nos dados de entrada a partir de dados históricos, mas também permitindo agregar conhecimentos dos especialistas durante o processo;
- Subsidiar à definição das metas previstas baseado nos resultados simulados;
- Desenvolver protótipo de aplicação para ser usado na previsão de produção.

### 1.3 Delimitação do estudo e justificativa

O estudo foi desenvolvido em empresa nacional que atua no ramo de energia em diversos setores, estando dentre eles a área objeto deste estudo: a exploração e produção de petróleo.

De acordo com a *International Energy Agency* (2021) o Brasil terminou o ano de 2021 como sétimo maior produtor de petróleo do mundo. No Brasil, os campos marítimos já produzem 97% do petróleo nacional (ANP, 2022).

A alta proporção da produção marítima destaca a relevância nacional e mundial relacionada à produção de óleo. O Brasil, por muito tempo, teve foco na exploração de campos terrestres. Atualmente o foco tem sido nos campos marítimos, em campos do pré-sal e pós-sal.

Este trabalho analisa a bacia de Campos, que é a mais madura entre os campos marítimos brasileiros. Essa bacia iniciou a produção comercial em 1977 e, portanto, possui uma extensa quantidade de dados históricos, podendo servir como benchmark para a previsão dos novos campos que podem enfrentar desafios semelhantes em termos de declínio de potencial e ineficiência. A bacia de Campos ocupa uma área marinha de aproximadamente 100 mil quilômetros quadrados, localizada na região que se estende do centro-norte do Estado do Rio de Janeiro até o sul do Estado do Espírito Santo.

Ferreira (2016) discute a maturidade dos campos de produção maduros, analisando suas projeções de produção e taxas de declínio em diferentes campos concentrados nesta região.

A previsão da produção é definida pela meta de produção probabilística de óleo. Ao utilizar o intervalo de confiança entre a P10 (projeção mais desafiadora, alcançável em 10% dos casos) e a P90 (projeção mais conservadora, alcançável em 90% dos casos), é adotado um intervalo de confiança de 80%.

Diferentemente de outras áreas de exploração e produção, a região geográfica da bacia de Campos apresentou dificuldades nos últimos anos para atingir a meta de produção, em parte explicada por eventos disruptivos, como a COVID-19, mas também em parte devido à metodologia utilizada e aos dados de entrada determinados pelos especialistas da produção.

A empresa enfrenta o desafio de planejar a meta de produção periodicamente. O acompanhamento da produção é atualizado regularmente para subsidiar o planejamento operacional (elaborado semanalmente para um horizonte de 4 meses),

tático (elaborado mensalmente para um horizonte de 15 meses) e estratégico (elaborado anualmente para um horizonte plurianual) da companhia. Este último corresponde ao planejamento plurianual divulgado ao mercado.

Esses planejamentos impactam decisões estratégicas de longo prazo, como manutenção, venda, revitalização de campos de produção ou descomissionamento de plataformas. Além disso, eles também afetam as decisões táticas, como o planejamento de datas de manutenções programadas, que podem exigir a alteração do volume de produção de acordo com movimentos externos, como alta pressão da concorrência, oligopólios, crise política, ou até guerra. Vale ressaltar que a vulnerabilidade do mercado de óleo e gás tem maior representatividade que os demais mercados, dado que o preço do barril de óleo é altamente volátil.

Os planejamentos também guiam as decisões em áreas relacionadas, interferindo no planejamento e controle das refinarias, logística e comercialização. Eles têm impacto nos indicadores de custos, impacto ambiental e no EVA (*economic value added*) da empresa. Além disso, afetam projetos em andamento, como paradas programadas para manutenção, perfuração de novos poços, construção de novas plataformas, desinvestimentos, entre outros.

Dessa forma, este estudo propõe o desenvolvimento de um aplicativo para uso no planejamento operacional, tático e estratégico da empresa, com implementação na região que atualmente indica maior necessidade de controle da produção.

Este trabalho também busca aprimorar o modelo, buscando soluções que aumentem a precisão da meta de produção, por meio da seleção dos melhores métodos de previsão integrados com simulação, análise de séries temporais e julgamento humano. Além disso, ele mapeia o estado atual *as is* do processo de simulação e análise de risco atualmente utilizado na empresa em estudo, e estabelece um procedimento com uma descrição do modelo *to be* proposto, visando proporcionar objetividade na inserção dos atributos de entrada para o modelo, como data de início, duração, entre outros.

Outro aspecto importante é que existem estudos relevantes na área de reservatórios que abordam simulações e comparam previsões para verificar o modelo de séries temporais que melhor prevê o declínio do potencial dos poços ao longo do tempo. No entanto, o grande número de variáveis consideradas na simulação estocástica de fluidos demanda alto esforço computacional e tempo de

execução, o que geralmente leva à não utilização direta das informações do reservatório no modelo corporativo de análise de risco para determinar as metas de produção. Portanto, essa informação costuma não ser considerada nas decisões estratégicas. Este trabalho se diferencia dos principais estudos de séries temporais em óleo e gás, pois integra a análise da previsão do potencial ao modelo corporativo de simulação e análise de risco das metas de produção.

A proposição deste estudo é estruturada em uma metodologia com encadeamento lógico orientado à análise de dados, com pré-tratamento dos dados, processamento através de algoritmos de séries temporais e clusterização, e análise dos resultados com uso do julgamento humano. O processo utiliza os insumos já utilizados pela empresa e reproduz com alta velocidade, adequando às particularidades das atividades, por meio do código desenvolvido na linguagem de programação R.

O modelo desenvolvido sugere contribuições para a incorporação do novo modelo de risco que está em desenvolvimento pela empresa, principalmente no sentido de utilizar abordagem mais objetiva através dos dados históricos, análises estatísticas e *machine learning*. Estuda-se um modelo que seja capaz de simular adequabilidades para diferentes plataformas, e que possa conciliar o trade-off entre traçar metas desafiadoras, porém, ao mesmo tempo, tangíveis.

#### **1.4 Estrutura da dissertação**

Este trabalho é composto por sete capítulos, incluindo a presente introdução. No Capítulo 2, descrevem-se os principais estudos identificados relacionados à modelagem da incerteza e simulação, bem como identificam pesquisas em análise de risco da meta de produção em óleo e gás, e no que tange a integração entre análise preditiva e julgamento humano. No Capítulo 3, é apresentada a metodologia de pesquisa, lógica de cálculo da produção probabilística, características do modelo atual de análise de risco, e a metodologia da aplicação em proposição, descrevendo a estruturação do modelo. No Capítulo 4, são expostas as regras lógicas que foram aplicadas a cada entidade durante a fase de desenvolvimento da aplicação. Detalhes específicos sobre como essas regras influenciam o funcionamento do modelo são fornecidos. No Capítulo 5, é apresentado o simulador desenvolvido, explorando o uso da interface elaborada em software R. No Capítulo 6, são apresentados os resultados alcançados, demonstrando as principais análises para os resultados

processados pelo simulador. Por fim, no Capítulo 7, são consolidadas as observações mais relevantes extraídas ao longo do estudo e são feitas proposições para possíveis trabalhos futuros.



## **2. Revisão bibliográfica**

Esta seção apresenta trabalhos relacionados ao tema, identificados através de pesquisas bibliográficas. São avaliados modelos implantados e tendências para modelagem de incertezas da previsão de produção, bem como métodos e aplicações de análise de risco e predição da produção de óleo. Também são abordados estudos que exploraram a integração da análise preditiva com o julgamento humano de especialistas.

### **2.1 Modelagem da Incerteza e simulação em exploração e produção**

Becerra (2012) analisa como as incertezas geológicas, de reservatórios, econômicas e tecnológicas afetam a tomada de decisões. Porém uma dificuldade é a variabilidade e complexidade dos dados históricos disponíveis para gerenciar a incerteza usando simulação numérica. Seu trabalho integra análise histórica com a análise probabilística, identificando os parâmetros incertos críticos e sua possível variação para estimar a faixa de reserva representativa.

O estudo de Becerra objetiva mostrar como o processo de calibração das incertezas permite reduzi-las dos parâmetros críticos e aumentar a confiança na previsão de produção, enaltecendo a importância de calibração das incertezas. O presente estudo busca explorar os parâmetros de incertezas nesse mesmo sentido, compartilhando do objetivo de aumentar a acurácia ao mensurar e calibrar bem seus parâmetros e as entidades de entrada.

Lechner (2005) comenta que modelos numéricos de simulação são ferramentas preditivas poderosas para responder a muitas questões de engenharia submarina. No entanto, as incertezas são grandes e os resultados da previsão abrangem uma ampla faixa. Para levar em conta toda a gama de resultados possíveis da simulação de reservatórios, rotinas de otimização teriam que ser executadas nos modelos numéricos, talvez milhares de vezes. Construir uma superfície de resposta capaz de prever um amplo número de resultados possíveis de uma simulação numérica por meio do processamento de uma variedade de parâmetros de entrada pode fornecer uma solução para esse problema. Ele propõe modelo analisando os

atributos de entrada, e combinando uso de modelo de predição e simulação Monte Carlo, assim como a presente pesquisa.

Guerillot e Bruvelle (2017) verificam que as decisões em reservatórios de petróleo e gás são frequentemente baseadas na avaliação de incertezas sobre previsões de produção. Ele combina a simulação Monte Carlo com análise de séries temporais, e observa que modelos geoestatísticos requerem milhares de simulações de fluxo para considerar as incertezas geológicas. Para contornar a restrição devido ao alto tempo computacional, apresenta uma abordagem para substituir o simulador por um modelo aproximado, com uso de simulação Monte Carlo, e aplica método de previsão de produção com uso de redes neurais artificiais (ANN).

Seguindo a mesma lógica que Lechner (2005) e Guerillot e Bruvelle (2017), o presente estudo desenvolve modelagem de simulação para análise de incertezas, simplificado em relação ao modelo de simulação de reservatório, e integra análise preditiva com simulação para inferir as incertezas e aumentar a acurácia das previsões.

## **2.2 Análise de risco da previsão de produção de óleo**

Castro *et al.* (2010) propõem modelo para avaliar o risco da previsão de produção considerando quatro entidades de entrada: potencial produtivo implantado, entrada de poços novos, paradas programadas para manutenção e eficiência operacional. O simulador apresenta modelagem que identifica e quantifica as incertezas para suportar o processo de determinação da meta de produção de óleo.

Bomfim (2017) sugeriu melhorias para o mesmo modelo, principalmente para poços novos, avaliando a data prevista de entrada em produção a partir de um sistema lógico de inferência *fuzzy* que interpreta riscos associados a cada poço. O sistema é concebido e integrado ao modelo de simulação. Os resultados mostram maior acurácia das previsões de produção.

O modelo proposto por Castro *et al.* (2010), e desenvolvido por Bomfim (2017), constitui a base do modelo de simulação do presente estudo que tece análise das mesmas quatro entidades de entrada, e pretende-se aprimorar o modelo.

Zhong (2016) estuda o problema de previsão de risco de desenvolvimento de campos petrolíferos e alertas imediatos para manter a produção de petróleo segura

e reduzir os erros de tomada de decisão causados pelo risco, apresentando um método de rede *Bayesiana* aprimorado com rede neural. Também comenta que não existe um método apropriado para medir o risco de desenvolvimento de campos petrolíferos, destacando como principais métodos de previsão de risco: o método de redes neurais, Monte Carlo, *fuzzy* e o método de rede *bayesiana*. A previsão de produção necessita fazer uso de métodos para avaliação do risco e auxiliar na tomada de decisão. A presente pesquisa faz uso dos métodos de Redes Neurais e Monte Carlo que são os mais comuns para reduzir erros de decisão baseados na previsão de produção.

Mamudu (2020) apresenta um modelo de predição de produção de óleo híbrido que combina *multilayer perceptron* (MLP) e *métodos Bayesianos*. Ele busca capturar incertezas, avaliar a dinâmica de riscos do sistema de produção, como perdas de produção. O objeto da presente pesquisa segue no mesmo sentido, explora métodos de *machine learning*, aplicados em análise de séries temporais, para gerenciar riscos na previsão de óleo.

Correia *et al.* (2021) compara a aplicação de séries temporais especificamente em 3 poços marítimos de petróleo, comparando o modelo estatístico ARIMA com redes neurais recorrentes. Ele ressalta que a área de óleo e gás tem apresentado progressiva aderência à ciência de dados para tornar as tomadas de decisões orientadas a dados. O monitoramento e previsão da produção de óleo e gás proporcionam informações úteis para o melhor manejo econômico de um poço ou de um campo de petróleo. Essas ações, na prática, prolongam o ciclo de vida do ativo e possibilita o aumento da sua produtividade. Os resultados mostraram que os modelos de redes neurais obtiveram melhor desempenho de previsão quando comparados com o modelo ARIMA.

Yanrui *et al.* (2022) utiliza previsão de séries temporais, motivado pelas flutuações históricas nas operações de poços de produção e reservatórios, com os dados de produção de petróleo em 65 poços localizados na Bacia Denver-Julesburg (EUA). Ele compara ARIMA, *LSTM Network* e Prophet, verificando que o ARIMA e LSTM têm desempenho melhor que o Prophet sob a suposição que nem todos os dados de produção de petróleo incluem influências sazonais.

Como Correia *et al.* (2021) e Yanrui *et al.* (2022) o estudo avalia análise preditiva, comparando erros de previsão entre os métodos de redes neurais e

estatísticos, avaliando os métodos com o intuito de obter maior acurácia no planejamento de produção em óleo e gás.

### **2.3 Simulação com análise preditiva e julgamento humano**

De acordo com Bomfim (2017) em condições de incerteza, uma alternativa para análise de risco é a utilização da simulação estocástica, que visa construir uma distribuição de probabilidade para a variável que se está analisando, possibilitando avaliar o risco (probabilidade) do valor assumido por esta ser inferior a um determinado limite estabelecido.

Estudos de simulação com séries temporais buscam melhorar conjunto de dados desbalanceados, redimensionando amostras de treinamentos. Métodos de simulação modificam conjuntos de dados desbalanceados, equilibrando distribuições (VANHOEYVELD e MARTENS, 2018).

Neste contexto, o presente estudo também aplica a simulação Monte Carlo com o objetivo de balancear a distribuição dos dados e evitar o uso de valores discrepantes (*outliers*), e concebe um modelo de simulação aproximado que integra de forma rápida todas as variáveis de produção ao modelo de risco. Os dados que fazem uso de análise de séries temporais são projetados e, em seguida, simulados a fim de obter previsões estocásticas que representem a distribuição de probabilidade das previsões, juntamente com os riscos associados.

A simulação é realizada por meio da integração entre o modelo e a opinião dos especialistas. Estudos examinam como as empresas devem integrar o julgamento humano com modelos preditivos para melhorar o desempenho dos modelos. Existem diferentes metodologias para essa integração, e a Tabela 1 destaca quatro dos principais métodos.

Estudos comparativos de métodos quantitativos e julgamento humano mostram que nenhum método é universalmente superior, e o melhor método varia em função de fatores como disponibilidade, qualidade, extensão e formato dos dados, sugerindo que as duas abordagens podem se complementar para produzir modelos mais precisos e resilientes (Zellner *et al.*, 2021).

No caso da análise dos especialistas na saída de um modelo, são considerados tanto a experiência dos especialistas quanto informações confidenciais que

englobam previsões de investimentos ou desinvestimentos, planos de ação para melhorias identificadas e a abrangência de outros casos semelhantes.

Tabela 1 - Métodos de integração entre julgamento humano e análise preditiva

<b>Métodos de integração</b>	<b>Definição</b>
Julgamento na entrada e desenvolvimento do modelo	O especialista é usado para desenvolver um modelo quantitativo, selecionando variáveis, especificando o modelo e estimando parâmetros de entrada.
Julgamento na saída do modelo	Ajustes realizados pelo especialista na saída do modelo.
Correção quantitativa	Sistema automatizado que monitora previsões de julgamento e usa qualquer viés detectado nas previsões anteriores para ajustar a previsão do próximo período.
Combinação da previsão	Previsão combina média ponderada entre especialistas e modelo de análise preditiva.

Fonte: Adaptado de Brau *et al.* (2023).

O presente estudo também avalia a performance de modelos, e ao mesmo tempo avalia disponibilizar opções para que o especialista possa ter liberdade de definir o melhor modelo para seu caso. A disponibilização para o especialista pode superar a seleção automática de variáveis em relação à variabilidade sistemática, como tendência e sazonalidade (Baets,2020; Petropoulos et al., 2018; Petropoulos e Siemsen, 2023).

Petropoulos (2018) sugere que as empresas devem considerar a seleção de previsões por julgamento como uma ferramenta complementar à seleção de modelos estatísticos, e ressalta que os sistemas de suporte a previsão que incorporam interfaces gráficas e recursos de séries temporais são um pré-requisito para a implementação bem-sucedida.

Vale ressaltar que também existem aperfeiçoamentos desses modelos, principalmente contemplando aplicações com maior interação entre modelo e usuário. Experimentos mostram que os especialistas têm a capacidade de usar e aprender com registros visuais de desempenho passado para selecionar e ajustar previsões baseadas em modelos de forma adequada (Baets *et al.*, 2020).

O presente estudo combina simulação com análise preditiva, utilizando a simulação Monte Carlo com o objetivo de obter previsões com menor margem de erro. O modelo emprega a contribuição dos especialistas na definição dos parâmetros de entrada; aplica uma abordagem com correções quantitativas na variável de parada programada; e utiliza a análise dos especialistas para avaliar e poder modificar a previsão de eficiência operacional e do potencial produtivo.

### 3. Metodologia

Esta seção apresenta a metodologia de pesquisa detalhada para o desenvolvimento do modelo de simulação para análise de risco probabilístico, abrangendo:

- a metodologia de pesquisa aplicada,
- a metodologia para o cálculo da produção probabilística,
- a metodologia do modelo atualmente em uso (a ser aperfeiçoado), e
- a metodologia do desenvolvimento da aplicação (nova proposição).

No último tópico, são explicados os detalhes do desenvolvimento em cada etapa do modelo orientado a dados.

#### 3.1 Metodologia de pesquisa

Adota-se a classificação da pesquisa conforme a taxonomia proposta por Vergara (2007), que divide as pesquisas segundo dois critérios básicos: em relação aos fins e aos meios.

Em relação à finalidade, a pesquisa consiste, primeiramente, como uma pesquisa aplicada, pois tem finalidade prática, motivada pela necessidade de se minimizar o problema concreto de previsão da meta de produção enfrentado por uma empresa de petróleo.

Ainda quanto aos fins, pode ser considerada uma pesquisa exploratória, pois aborda temas na literatura para testar um modelo matemático desenvolvido como ferramenta de solução para o problema da empresa. Foram levantados trabalhos relacionados às bases: Scopus, Science Direct, e Google Scholar sobre temas como: *Simulation, Stochastic, Forecasting, Oil & Gas, production planning with uncertainty, prediction with human judgment*, e temas relacionados aos métodos de previsão e simulação.

Em relação aos meios, esta pesquisa pode ser descrita como: Bibliográfica pois teve fundamentação teórica realizada com base em material publicado e

acessível ao público, como: livros, teses, dissertações e artigos; Documental visto que os materiais são de fontes primárias, compreendendo documentos oficiais da companhia; e estudo de caso por se tratar da observação de uma empresa única, apesar desse trabalho poder gerar benefícios para várias empresas com aplicação de análise de risco de atingimento da meta prevista, ou que estão aplicando técnicas de simulação, séries temporais, *machine learning*, e métodos estatísticos, principalmente em aplicações no ramo da produção *offshore*.

### 3.2 Metodologia para cálculo da produção probabilística

As projeções da produção são realizadas com base nos especialistas que buscam entender o histórico para estimar as projeções. Uma das proposições do modelo proposto é aumentar a interação entre os especialistas e o modelo, utilizando análise preditiva com base em dados estatísticos históricos, a fim de reduzir o uso subjetivo de dados puramente provenientes dos especialistas. Que por observação prática apresentam viés otimista o que aumenta a previsão de produção.

Os dados históricos são extraídos através da ferramenta Power BI, que está integrada ao banco de dados oficial da empresa. Isso significa que a produção e o potencial de óleo, tanto dos poços já instalados quanto dos novos poços, são alimentados pelos dados fornecidos pela área de reservatório da empresa. Os dados de perdas, incluindo eventos de paradas para manutenção, são extraídos de relatórios das ferramentas corporativas. É importante destacar que esses relatórios possuem um processo bem definido de preenchimento e controle, e reuniões diárias envolvendo engenheiros de diferentes áreas são realizadas para garantir o correto registro das perdas. Quanto aos dados projetados pelos especialistas, eles são disponibilizados em um arquivo específico no formato de planilha Excel.

A análise das incertezas para cada uma das quatro entidades dos dados de entrada resulta no cálculo da produção resultante, conforme proposto por Castro *et al.* (2010).

As entidades de entrada são detalhadas abaixo para que o modelo seja de fácil interpretação:

**Potencial produtivo implantado:** Essa entidade representa a capacidade de potencial de produção de óleo nos reservatórios dos poços atualmente instalados.



Ela incorpora as incertezas relacionadas aos reservatórios. O potencial já implantado de uma plataforma é representado pela curva de potencial estimada para cada poço “ $p$ ” no momento “ $t$ ”, sendo denotado pela equação 1.

Equação 1 – Cálculo da entidade potencial produtivo implantado.

$$\sum_{t=1}^T \sum_{p=1}^P Potencial_p [t] = \sum_{t=1}^T \sum_{p=1}^P Produção_p [t] + \sum_{t=1}^T \sum_{p=1}^P Perdas_p [t] p[t] .$$

A unidade de medida do potencial é de barris por dia (bpd), e o momento é determinado mensalmente.

**Eficiência Operacional:** entidade abrange eventos que acarretam a não realização da produção (perdas) relacionadas a falhas nos equipamentos, resultantes de eventos inesperados e não previstos antecipadamente. Essas falhas podem levar a interrupções ou redução da produção dos poços marítimos, devido a problemas nos sistemas operacionais que impedem que a produção realizada seja igual ao seu potencial produtivo.

O percentual de eficiência é uma medida de desempenho operacional, sendo o quociente da produção (potencial produtivo subtraindo as perdas de produção) pelo potencial produtivo, para cada plataforma “ $i$ ” no momento “ $t$ ”, como mostra a equação 2.

Equação 2 – Cálculo da entidade eficiência operacional.

$$\sum_{t=1}^T \sum_{i=1}^I Eficiência_i [t] = \sum_{t=1}^T \sum_{i=1}^I \frac{Potencial_i[t] - Perdas_i[t]}{Potencial_i[t]}$$

**Entrada de poços novos:** entidade representa a quantidade de potencial de produção que será adicionada ao potencial produtivo da plataforma existente devido ao início da operação de novos poços produtores de óleo.

O deslocamento da contribuição de potencial de um poço é influenciado pela sua data (Dt.) de entrada em operação, para cada poço novo “ $p$ ” no momento “ $t$ ”. O cálculo do desvio para início de produção é realizado conforme especificado na equação 3.

Equação 3 – Cálculo utilizado na simulação da data de entrada do poço novo

$$\sum_{t=1}^T \sum_{p=1}^P Dt. Calculada_p [t] = \sum_{t=1}^T \sum_{p=1}^P Dt. Inicio_p [t] + \sum_{t=1}^T \sum_{p=1}^P \Delta Dt. Simulada_p [t]$$

Onde  $\Delta Data simulada$  é obtido através da distribuição triangular.

De posse do momento inicial da entrada do poço, é considerado o potencial de produção (em barris por dia):  $Potencial_p[t]$ , onde  $Potencial_p[t]$  representa o potencial do poço “p” no momento “t”.

**Paradas programadas para manutenção:** entidade corresponde a manutenções planejadas que paralisam as plataformas completamente. Essas paradas programadas são a perda mais significativa e com maior previsibilidade de eficiência. A partir do início da parada, a produção é reduzida de acordo com o cálculo especificado na equação 4.

Equação 4 – Cálculo utilizado na simulação da perda do evento de parada.

$$\sum_{t=1}^T \sum_{i=1}^I Parada_i[t] = \sum_{t=1}^T \sum_{i=1}^I Perda\ de\ produção_i[t] * Duração_i[t]$$

Onde a perda de produção é representada em barris por dia, e a duração corresponde ao número de dias no momento considerado (normalmente em mês). Então,  $Parada_i[t]$  representa a perda com parada da plataforma “i” no momento “t”.

**Simulação da produção:** é realizada por meio de um simulador que considera diversos atributos e incertezas associadas a cada plataforma. O processo de simulação envolve a geração de números aleatórios para os atributos de cada entidade de entrada, como datas de entrada de poços e durações de manutenções.

O simulador utiliza um processo iterativo e simulações por amostragem para obter diferentes valores para cada entidade em cada rodada. Em cada iteração do simulador, é estimada uma projeção de produção, levando em conta as características específicas de cada plataforma.

O resultado da simulação é uma distribuição de probabilidade que representa a curva probabilística da produção de óleo. Essa distribuição é construída agregando os resultados de todas as iterações, sendo determinada através da equação 5.

Equação 5 – Cálculo da produção a partir das entidades de entrada do modelo.

$$\sum_{n=1}^N \sum_{t=1}^T \sum_{p=1}^P \left[ \frac{((Potencial_p[t] + Potencial_i[t]) - Parada_i[t]) * Eficiência_i[t]}{T} \right]$$

Dado que:

$Potencial_p[t]$  – potencial de produção do poço novo p no momento t;

$Potencial_i[t]$  – potencial de produção já implantado da plataforma i no momento t;

$Parada_i[t]$  – volume de perdas devido à parada da plataforma  $i$  no momento  $t$ ;

$Eficiência_i[t]$  – eficiência da plataforma  $i$  no momento  $t$ ;

$p$  – indicador de número de poços;

$P$  – Número total de poços;

$n$  – indicador do número de plataformas marítimas de produção;

$N$  – número total de plataformas marítimas de produção;

$t$  – indicador do momento do tempo (Mês);

$T$  – Horizonte de tempo do período de planejamento (Mês).

$i$  – indicador de número de plataformas;

$I$  – Número total de plataformas.

A partir curva probabilística da produção de óleo, é possível obter diversas informações, como a média anual de produção diária de óleo, medida em barris por dia (bpd), e os percentis representativos da distribuição, como o P50 (mais provável) e o intervalo de confiança entre o P10 e o P90.

A simulação permite quantificar as incertezas existentes na produção de óleo. Ela estabelece intervalos de confiança, e auxilia no processo de determinação da meta de produção para um horizonte de doze meses.

### 3.3 Modelo atual de análise de risco da previsão da produção

O modelo atual de análise de risco da projeção de produção utilizado pela empresa em estudo segue o modelo representado pelo esquema da figura 1.

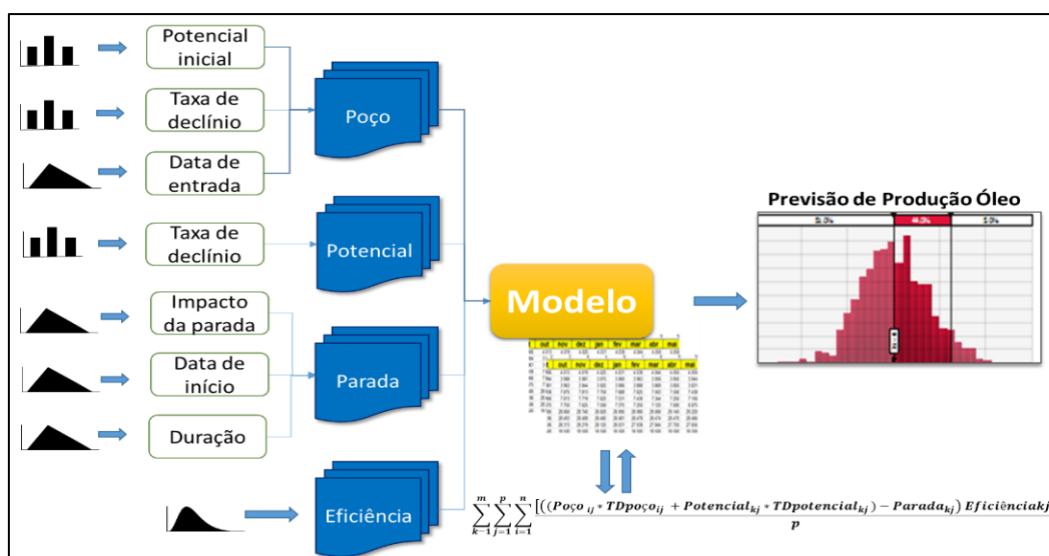


Figura 1 – Modelo conceitual da metodologia de previsão da produção de petróleo.

Fonte: Bomfim (2017)

O simulador é executado usando software específico, o @Risk. O modelo gera as iterações e fornece a curva probabilística da previsão de produção.

No simulador do modelo atualmente em uso, o potencial de produção é parametrizado com base em três cenários que buscam representar as distribuições P10, P50 e P90, sendo determinadas pelas estimativas dos especialistas. Para isso, os padrões de declínio de cada plataforma são identificados por meio de observação gráfica, com três estimativas tentando extrapolar o histórico realizado. A distribuição utiliza os percentuais 30-40-30 conforme proposto por Hurst *et al.* (2000). Ele comenta que essa distribuição estava sendo usada na indústria de exploração de petróleo.

As projeções de eficiência são estimadas pelos especialistas, levando em conta a média e o desvio padrão mensal para cada plataforma, parametrizando uma distribuição normal para a simulação. Essas estimativas são baseadas na observação do histórico e em tendências futuras.

As contribuições dos poços novos são estimadas considerando o potencial inicial pleno, sem usar a projeção mensal do reservatório, sendo aplicadas três taxas de declínio (TD) estimadas para projetar a diminuição do potencial (mais conservadora, mais provável e mais otimista), e são estimadas três datas para a data de início de produção, utilizando uma distribuição triangular.

As paradas programadas para manutenção contemplam três estimativas dos especialistas para determinar os parâmetros de uma distribuição triangular que representa a data de início, impacto e a duração da parada.

O modelo explicado se configura como o modelo base para a atual análise de risco da previsão de produção probabilística corporativa da empresa em estudo.

É importante melhorar o modelo no que tange a diminuir a subjetividade dos dados inseridos pelos especialistas. Para tanto, busca-se uma abordagem com maior uso dos dados históricos para os parâmetros de entrada. Por outro lado, busca-se ter maior integração entre o modelo de simulação com o julgamento dos especialistas durante a modelagem e para a análise dos resultados.

Outros pontos de melhorias a serem aperfeiçoados são: maior flexibilidade na modelagem das entidades de entrada, maior integração com banco de dados corporativo e do reservatório, maior interpretação da análise, redução do tempo de processamento da simulação, e maiores opções para análise dos resultados.

### 3.4 Metodologia de desenvolvimento do aplicativo

A metodologia de desenvolvimento do aplicativo pretende se aproximar das etapas utilizadas na descoberta do conhecimento baseado em banco de dados (KDD), conforme preconizada por Fayyad *et al.* (1996) que o descreve como o processo não trivial de identificar padrões potencialmente úteis e compreensíveis em dados.

A figura 2 representa o framework da metodologia desenvolvida para o modelo proposto do simulador.

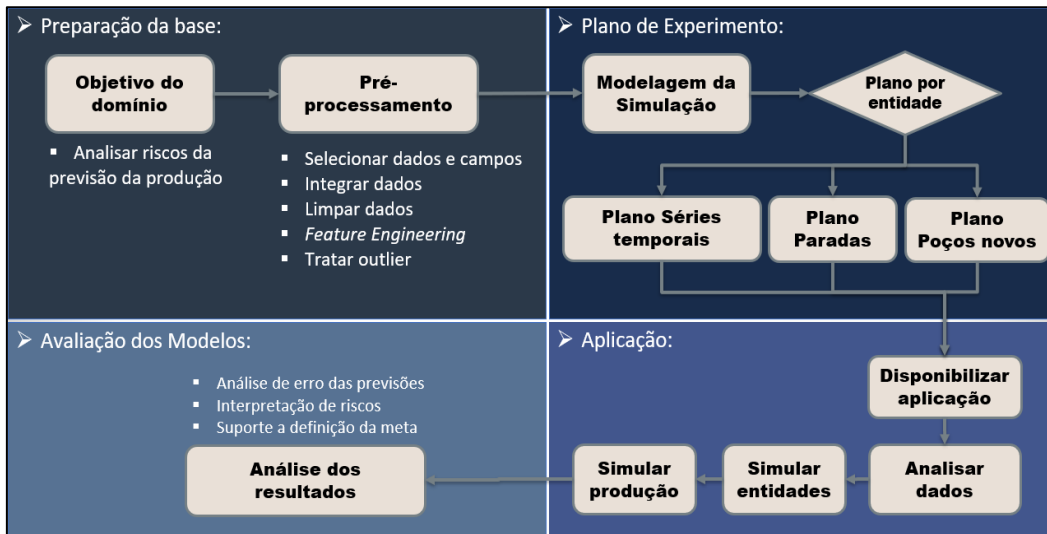


Figura 2 – Estruturação do modelo para desenvolvimento do aplicativo

Fonte: Elaboração própria.

O processo busca descrever todo o ciclo da descoberta de informações. O pré-processamento tem análise desde a seleção dos dados, com tratamento, limpeza, e integração entre os dados para preparar a base. O processamento implementa os algoritmos da aplicação conforme plano definido. O pós-processamento identifica padrões e avalia os modelos, com a análise dos resultados e o julgamento humano dos especialistas realimentando o processo para verificar possíveis adequabilidades dos eventos inseridos e atributos do modelo em uma nova simulação.

#### 3.4.1 Objetivo do domínio

O objetivo do domínio é o ponto de partida para o desenvolvimento da aplicação. Ele está plenamente alinhado aos objetivos transcritos neste trabalho.

#### 3.4.2 Pré-processamento

No contexto do pré-processamento, as atividades realizadas nessa etapa correspondem a: seleção dos dados; integração; limpeza; extração de características/*feature engineering*; e tratamento de outlier.

**Seleção dos dados:** o conjunto de dados compreende o histórico das entidades de entrada. Compreendem os anos de 2017 a 2021 (60 meses), e a projeção é realizada para o ano de 2022 (12 meses).

Os especialistas em planejamento da produção da área de planejamento e controle da produção da empresa em estudo, sinalizam investimentos significativos nos últimos anos a fim de melhorar o desempenho operacional, não sendo indicado trabalhar com períodos anteriores a 2017. Os dados compreendem 22 plataformas marítimas da Bacia de Campos.

**Integração dos dados:** destaca-se a padronização de nomenclatura das plataformas e poços a fim de unir os dados de diferentes fontes.

**Limpar dados:** como pretende-se analisar o histórico das plataformas atualmente em produção, então foram retirados dados de 29 unidades pois foram vendidas ou hibernadas no período.

**Feature engineering:** destaque para o tratamento no potencial de produção com a divisão da variável (*splitting variable*) em potencial Pleno, sem desinvestimento, e potencial implantado (retirando os poços novos no período).

**Tratar Outlier:** o tratamento é feito em abril/20 devido ao fechamento de poços no início da COVID-19, e que tiveram como tratamento imediato não contar no potencial. Como o intuito é avaliar o potencial em mesmas condições que os demais períodos, foi necessário descartar os valores de potencial do período e imputar novos valores, utilizando média móvel exponencialmente ponderada (EWMA), em seu uso os fatores de ponderação diminuem exponencialmente para cada período do tempo. (LIN *et al.*, 2020).

A escolha pela técnica foi por ter ampla aplicação em séries temporais na indústria e nos negócios (HAGHIGHATI e HASSAN, 2018).

Foi definido, junto aos especialistas, usar a média móvel dos últimos quatro meses, pois o declínio do potencial pode ser significativo em períodos maiores.

### 3.4.3 Modelagem da Simulação

As projeções probabilísticas são obtidas para cada entidade de entrada do modelo: potencial implantado, eficiência, poços novos, e paradas. A execução do

algoritmo segue a ordem citada, são realizadas simulações com replicações independentes para cada entidade de entrada, gerando números aleatórios a partir dos parâmetros detalhados no plano de cada entidade.

A entidade de paradas é a única entidade que possui dependência de outra, ela depende do potencial simulado para o período (mês), a fim de realizar o cálculo da perda para o mesmo período.

A figura 3 mostra a modelagem da simulação, com a utilização da distribuição triangular para todas as entidades. os dados simulados, o método utilizado e os resultados disponibilizados pela simulação.

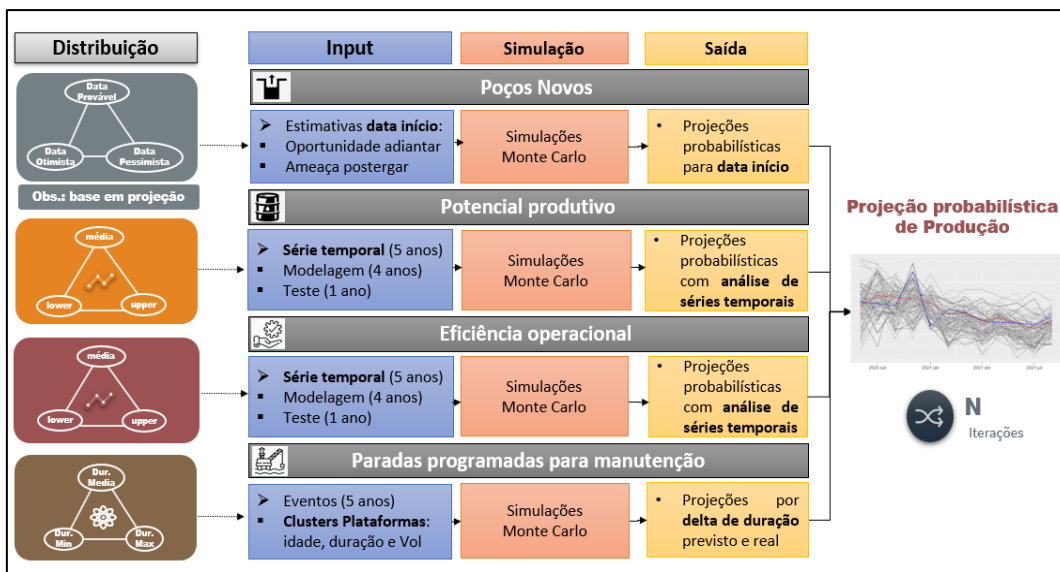


Figura 3 – Modelagem de simulação do aplicativo

Fonte: Elaboração própria.

É parametrizado para que todas as entidades utilizem a mesma distribuição que já é utilizada no modelo atual, com distribuições triangulares, e, ao invés de terem seus parâmetros estimados pelos especialistas, cada qual tem seu processo aperfeiçoado de modo específico. O plano de aplicar métodos diferentes para cada dimensão de entrada é justamente para possibilitar avaliar diferentes ideias de aplicações, e aumentar a possibilidade de diferentes análises por entidade.

O estudo tem grande foco em desenvolver o simulador através de análise de séries temporais. Primeiramente, tem-se a definição dos modelos a serem analisados, através de um plano de experimentos, onde são avaliados diferentes métodos de previsão. A análise de séries temporais pode utilizar diferentes métodos de previsão, correspondendo a diferença entre os modelos. O plano de séries temporais detalha a avaliação entre os métodos de previsão do modelo. Foram

avaliados oito métodos, dos quais quatro foram selecionados para uso constante no simulador.

As melhorias baseadas em séries temporais apenas foram aplicadas nas dimensões de potencial e eficiência. A aplicação dos métodos de previsão indica a média e o intervalo de confiança projetado. Esses três parâmetros são exatamente os insumos utilizados para a distribuição triangular da simulação para o potencial e eficiência, possibilitando visualizar os resultados mais esperados (P50), e obter um intervalo de confiança devido as incertezas (entre a P10 e a P90).

No caso dos poços novos, apenas foram sugeridos aperfeiçoamentos nos atributos e parâmetros de entrada do modelo, alinhando-os com a identificação de ameaças e oportunidades relacionadas à curva projetada, bem como suas probabilidades de ocorrência. Desse modo, ela é a única entidade que permanece fazendo uso da projeção dos especialistas, sua permanência foi adotada para que também seja possível avaliar, em uma entidade, como está a acurácia das estimativas dos especialistas.

Para as paradas programadas para manutenção, foi realizado um estudo do histórico, incluindo a clusterização e análise estatística dos agrupamentos para comparar o desvio entre o realizado e o planejado. Sua implementação busca aproveitar o uso do julgamento humano no que tange a uma correção quantitativa histórica, buscando que os dados históricos ajustem vieses do julgamento humano, através de sistema automatizado.

As simulações Monte Carlo são realizadas de acordo como descrito no plano detalhado para cada entidade, cada qual com seus parâmetros para utilizar na distribuição triangular da simulação. Assim, a partir de três estimativas dos especialistas são realizadas iterações que determinam nova data de entrada para cada poço novo; com base na média e intervalo de confiança das séries temporais obtém-se novo patamar de potencial e eficiência das plataformas; e com base no viés histórico entre o planejamento dos especialistas para cada parada, em relação ao máximo, médio e mínimo da duração realizada para o cluster de plataformas associado, obtém-se a duração das manutenções programadas por plataforma.

O número de iterações é um parâmetro variável no simulador, e o cálculo da produção segue a equação 5 mencionada na metodologia de cálculo da produção probabilística.



### 3.4.4 Plano por entidades

Para as dimensões de entrada ficarem disponíveis para simulação, cada qual tem seu processo de tratamento e parâmetros definidos, sendo importante explicar suas etapas para tornar clara a interpretação do modelo e facilitar a explicação dos resultados.

As atividades adotadas para cada entidade de entrada do modelo são resumidas na Figura 4:

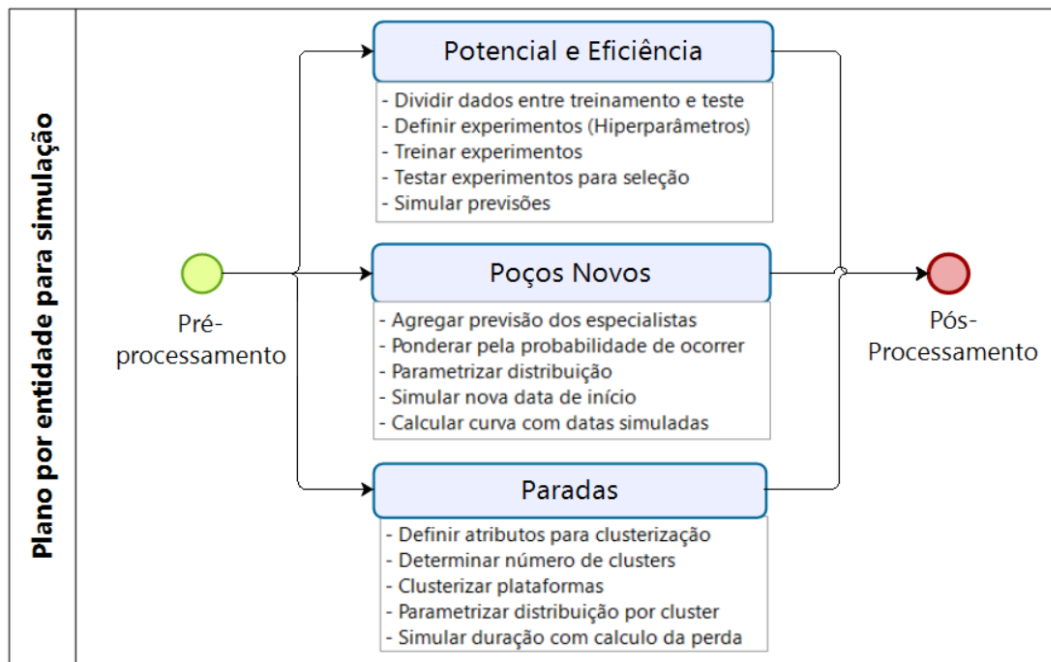


Figura 4 – Etapas do plano por entidade para simulação do Modelo

Fonte: Elaboração própria.

**Potencial Implantado e Eficiência Operacional:** As entidades usam séries temporais para análise. A aplicação de séries temporais foi modelada para comparar oito métodos de previsão, sendo selecionados quatro métodos estatísticos e quatro métodos de aprendizado de máquina para comparação. Os experimentos são realizados para testar a aplicação dos métodos e selecionar para o simulador apenas os quatro com maior acurácia nos experimentos.

Foram utilizados métodos de previsão comumente identificados na literatura. Buscou-se avaliar 4 métodos estatísticos e 4 métodos de *Machine Learning*. Os métodos em análise são detalhados no apêndice.

**Poços Novos:** No contexto dos poços novos, propõe-se utilizar as projeções dos especialistas sem análise da base histórica. Isso ocorre porque cada projeto de

interligação de novo poço já realiza uma análise detalhada do cronograma de entrada. Dessa forma, justifica priorizar a análise histórica para as outras entidades.

O plano de poços novos compreende aos especialistas em projetos da companhia avaliarem a existência de oportunidades ou riscos relacionados a data de entrada de cada poço, avaliando a probabilidade de cada evento, servindo como fator de ajuste para o impacto projetado. O impacto é combinado com as datas esperadas para a data de entrada dos poços, parametrizadas em distribuição triangular, resultando em novas projeções simuladas.

**Paradas programadas para manutenção:** A proposta é analisar o histórico, apurando diferenças entre duração prevista dos eventos em relação ao realizado. Um projeto de parada programada tem alta complexidade, por isso é comum precisar aumentar o escopo, com consequente realização de paradas além do previsto.

O histórico de 5 anos por plataforma é considerado baixo para análises estatísticas, devido à baixa quantidade de paradas programadas anuais. Portanto, propõe-se agrupar os dados por plataformas similares usando clusterização.

A simulação baseada nas diferenças históricas entre planejado e realizado, devem permitir que os especialistas avaliem melhor o próprio planejamento, considerando correções quantitativas do simulador para cada cluster.

### 3.4.5 Desenvolvimento da aplicação

O aplicativo em *Shiny* é desenvolvido a partir da linguagem de programação em R (versão 4.1.2; R Core Team, 2021). Os pacotes utilizados foram: “tidyverse” (Wickham, 2017), lubridate (Grolemund, 2011), “imputeTS” (Moritz, 2017), “forecast” (Hyndman, 2021), “nnfor” (Kourentzes, 2017), “timetk” (Dancho, 2023), “extraDistr” (Wolodzko, 2017), “Shiny” (Chang, 2021), “shinydashboard” (Chang, 2021), “DT” (Xie, 2022), e “Plotly” (Sievert, 2022).

O desenvolvimento tem suas regras lógicas definidas para cada entidade de entrada, conforme atividades do plano por entidade, ele visa traduzir as regras de negócio da empresa, conforme detalhado no capítulo 4. Ao final, a aplicação é desenvolvida para uso, conforme capítulo 5.

### **3.4.6 Avaliação dos Modelos**

A avaliação dos modelos abrange desde a análise dos experimentos realizados, os quais selecionam os principais modelos para posterior uso no simulador, explicado no capítulo 4, e a análise de casos de simulação após a disponibilização da aplicação, conforme capítulos 5 e 6.

## 4. Desenvolvimento da aplicação

### 4.1. Potencial Implantado e Eficiência Operacional

A implementação dos modelos de séries temporais compreende, principalmente, a implementação e análise dos experimentos realizados, os quais selecionam os principais modelos para posterior uso no simulador.

A principal diferença entre a análise de séries temporais realizada durante os experimentos e após implantação no simulador, é que os experimentos usam parâmetros manuais, e o simulador parâmetros automáticos. Como os métodos automáticos possuem diferentes modos de avaliação para definir os parâmetros ótimos, os experimentos usam parâmetros manuais para avaliar diferentes combinações de parâmetros, todos eles sendo mensurados em relação ao erro percentual absoluto médio (MAPE).

As atividades executadas para realizar a análise com série temporal, e disponibilizá-las no simulador, podem ser divididas nos passos a seguir:

**Dividir dados entre treinamento e teste:** A análise de séries temporais envolve as etapas de treinamento e teste.

Para os experimentos, os dados das séries são divididos em treinamento, quatro primeiros anos de 2017 a 2020, e testes com o último ano dos dados históricos, 2021, buscando manter proporção adequada (80% e 20%), e manter a caracterização das perdas de todo um ano, de janeiro a dezembro, para ambos os casos.

Para o simulador, o treinamento contempla todos os 5 anos de histórico, 2017 a 2021, e a projeção é realizada para o ano de 2022.

**Definir experimentos com Hiperparâmetros:** Esta etapa planeja as alterações dos hiperparâmetros para identificar os ótimos. Assim, são identificados os melhores ajustes em cada método analisado. A tabela 2 mostram as alterações realizadas para cada método de série temporal.

Tabela 2 - Plano de experimentos para identificação dos hiperparâmetros ótimos

Método	Modificações nos hiperparâmetros
MLP	Número de camadas, <i>lags</i> , iterações
RBF	Número de Centros e alcance ( <i>gamma</i> )
ELM	Número de camadas, Tipo de regressão para pesos da camada de saída; <i>lags</i> históricos e número de iterações
ARIMA	Autoregressões, diferenciação e média móvel
ETS	Amplitude histórica ( <i>alpha</i> ) e amortecimento da tendência ( <i>beta</i> )
Prophet	Sazonalidade semanal, intervalo de confiança, número de iterações.
TBATs	Nível ( <i>alpha</i> ), Tendência ( <i>beta</i> ), sazonalidade ( <i>gamma</i> ) uso de transformação <i>Box-Cox</i> , erros <i>ARMA</i> .

Fonte: Elaboração própria.

**Treinar experimentos:** os dados de treinamento são baseados no histórico de cada plataforma, abrangendo o período de 2017 a 2020. As séries temporais (potencial, numerador da eficiência e denominador da eficiência) são treinadas para cada método de previsão.

**Testar experimentos para seleção:** Os testes dos experimentos são conduzidos para selecionar os métodos de previsão com melhores performances. As previsões são testadas em relação aos dados reais de 2021, e são comparadas usando o erro percentual absoluto médio (MAPE), devido a facilidade de interpretar porcentagens, sendo mais comum em aplicações corporativas.

As previsões com menores erros permitem selecionar os quatro métodos a serem disponibilizados no simulador.

Plano dos experimentos: o plano é registrado em uma planilha de Excel, contendo informações como o ID do cenário, método de previsão, hiperparâmetros definidos no cenário do experimento, e erro obtido. Foram avaliados 141 cenários no total. Outros testes também foram realizados, porém não foram sinalizados devido à baixa alteração no resultado dos erros.

A figura 5 mostra dois exemplos de modificações dos parâmetros, apresentando seus cenários registrados no plano de experimentos, ambos aplicados na entidade do potencial produtivo: MLP à esquerda, indica ajustes nas camadas escondidas (hd), e RBF à direita, mostrando alterações no número de centros (k) e alcance do  $\gamma$  (y). A cor em verde mostra os parâmetros ótimos encontrados. O caso da cor laranja, também mostrado na figura, apenas é representado pois durante o processo, ao encontrar valores próximos ao valor ótimo, eles também eram destacados para ter maior foco em buscar alterações diante de parâmetros próximos, a fim de verificar se teria diminuição do erro.

Plano de Experimento (Potencial de produção)				Plano de Experimento (Potencial de produção)			
Cenário	Modelos		Métrica de erro	Cenário	Modelos		Métrica de erro
	MLP	Parâmetros	MAPE		RBF	Parâmetros	MAPE
36	x	hd=6,7	4,3%	1	x	k=9	9,64%
37	x	hd=6,8	6,53%	2	x	k=8	9,68%
38	x	hd=6,10	7,03%	3	x	k=7	9,72%
39	x	hd=7,2	9,14%	4	x	k=6	9,73%
40	x	hd=7,5	8,73%	5	x	k=5	9,73%
41	x	hd=7,8	7,26%	6	x	k=4	9,68%
42	x	hd=7,9	5,89%	7	x	k=3	9,55%
43	x	hd=7,10	5,44%	8	x	k=2	9,6%
44	x	hd=7,11	4,93%	9	x	k=11	9,73%
45	x	hd=7,12	6,88%	10	x	k=12	9,77%
46	x	hd=8,2	7,65%	11	x	k=10,y=1	9,59%
47	x	hd=8,5	8,92%	12	x	k=10,y=1,5	9,61%
48	x	hd=8,6	6,83%	13	x	k=10,y=2,0	9,64%

Figura 5 – Exemplos de cenários nos experimentos: Método MLP (esq.) e RBF (dir.)

Fonte: Elaboração própria.

Abrangência dos experimentos: Os testes analisam todos os métodos sendo aplicados para as seis seguintes unidades: total de produção de todas as plataformas da bacia de Campos, e para as cinco plataformas (Plat.) com maior potencial da região, visando ter maior generalização dos testes, e evitando *overfitting*.

A tabela 3 mostra, como exemplo, o resultado para a entidade de eficiência. As médias dos erros são aplicadas, tanto para a eficiência como para o potencial.

Tabela 3 – Aplicação da abrangência dos experimentos: Exemplo na eficiência

Modelo	Plat. 1	Plat. 2	Plat. 3	Plat. 4	Plat. 5	Total Bacia	Média Eficiência
TBATS	7,76%	7,78%	14,97%	7,18%	7,03%	2,68%	<b>7,90%</b>
ETS	7,85%	7,78%	14,96%	7,19%	7,02%	2,67%	<b>7,91%</b>
ARIMA	7,85%	7,92%	12,52%	7,20%	10,38%	2,64%	<b>8,09%</b>
RNAR	11,68%	8,30%	12,92%	7,86%	10,56%	2,13%	<b>8,91%</b>
ELM	13,33%	8,63%	13,22%	7,88%	10,35%	2,82%	<b>9,37%</b>
RBF	9,42%	11,85%	16,44%	11,33%	6,88%	1,89%	<b>9,64%</b>
MLP	10,81%	17,60%	15,49%	8,14%	6,10%	2,69%	<b>10,14%</b>
Prophet	8,93%	14,56%	18,33%	11,30%	8,56%	2,07%	<b>10,63%</b>

Fonte: Elaboração própria.

Resultado dos experimentos: A média dos erros encontrados para as unidades na entidade de eficiência e Potencial, corresponde ao resultado para a definição dos métodos escolhidos. O resultado é detalhado na tabela 4. Os métodos selecionados e disponibilizados para o modelo foram os quatro primeiros da tabela: TBATs, ELM, ARIMA e Suavização exponencial.

Tabela 4 – Resultados dos experimentos para selecionar métodos de predição

MODELO	Média Eficiência	Média Potencial	Média Geral	Classificação
TBATS	7,90%	5,18%	<b>6,54%</b>	<b>1</b>
ELM	9,37%	4,94%	<b>7,16%</b>	<b>2</b>
ARIMA	8,09%	6,49%	<b>7,29%</b>	<b>3</b>
ETS	7,91%	7,19%	<b>7,55%</b>	<b>4</b>
RNAR	8,91%	6,91%	<b>7,91%</b>	<b>5</b>
RBF	9,64%	9,49%	<b>9,56%</b>	<b>6</b>
Prophet	10,63%	8,78%	<b>9,70%</b>	<b>7</b>
MLP	10,14%	15,29%	<b>12,72%</b>	<b>8</b>

Fonte: Elaboração própria.

**Simular previsões:** Os métodos selecionados nos experimentos são disponibilizados para rodar a simulação do modelo. O treinamento é realizado com todos os dados dos cinco anos (2017 a 2021), e as previsões são feitas para o ano de 2022.

Durante a simulação, os dados são primeiramente projetados e, em seguida, são executadas as iterações da simulação. Uma distribuição triangular é utilizada, sendo parametrizada pela média e intervalo de confiança projetado (*lower e upper*).

No simulador, foram utilizadas funções automáticas, como ets, autoarima, tbats da biblioteca *forecast* e elm da biblioteca *nnfor*, visando automatizar a otimização dos parâmetros e garantir maior eficácia no ajuste dos parâmetros. Isso contribui para reduzir o tempo de processamento da simulação e permite uma análise mais ampla dos resultados.

## 4.2. Poços Novos

As atividades implementadas para desenvolver o modelo de poços novos podem ser divididas nos passos a seguir:

**Agregar previsão dos especialistas:** O incremento desse estudo à modelagem de poços novos, é dada pelo mapeamento de oportunidades ou riscos, quantificando em dias e em probabilidade de ocorrência (em escala de 1 a 5), associadas a cada poço pelos especialistas em planejamento e controle da produção.

O mapeamento de oportunidades ou riscos identifica principalmente fatores relacionados a: documentação de projeto, anuência do IBAMA, priorização de perfuração de poços ou barcos de lançamento de linhas de escoamento da produção (PLSVs), e suprimentos.

Assim, o cenário conservador considera a data com maior risco de atraso; data mais provável de início da produção alinhada à data do planejamento determinístico; e otimista considera plano de ação com maior oportunidade de antecipação.

**Ponderar pela probabilidade de ocorrência:** Para cada oportunidade ou ameaça identificada, sugere-se avaliar a probabilidade de ocorrência do evento, classificando-a como baixa (fator multiplicativo de 50%), média (75%) ou alta (100%). Essa probabilidade serve como fator de ajuste para o impacto considerado, ponderando o resultado com a probabilidade de ocorrência.

**Parametrizar a distribuição:** É utilizada uma distribuição triangular (conservador, mais provável e otimista) para parametrizar a incerteza na data de entrada de cada novo poço. Assim, o conservador considera a data impactada pelo maior risco a atrasar, e otimista considera a data impactada pela maior oportunidade a antecipar.

As curvas são diretamente impactadas pelo deslocamento da curva de cada poço, resultado da incerteza na data de entrada de operação. O cálculo do deslocamento para início de produção segue conforme apresentado na equação 3.

**Simular nova data de início:** a simulação é realizada com as iterações da Simulação Monte Carlo em relação à data de entrada, gerando novas datas probabilísticas de entrada dos poços.

**Calcular curva com datas simuladas:** os resultados estimam novas datas probabilísticas de entrada dos poços, causando deslocamento da contribuição de potencial atual. Dessa forma, existe uma simplificação para o deslocamento da curva de potencial do poço. Essa prática de deslocamento não é a mais precisa por



existir fatores de reservatório que interferem no potencial de um poço com o tempo, porém para decisões corporativas entende-se ser possível adotar a simplificação.

### 4.3. Paradas programadas para manutenção

As etapas desenvolvidas para a entidade de paradas podem ser divididas nos passos a seguir:

**Definir atributos para clusterização:** os atributos considerados para a clusterização são indicados pelos especialistas em produção: o volume de perda, a duração média dos eventos e a idade atual da plataforma. A idade é considerada fundamental, pois está correlacionada com as necessidades de manutenção das plataformas.

**Determinar número de clusters:** a definição do número de clusters foi com base nos índices do pacote NbClust. Os índices indicam o número ideal de 1 ou 4 clusters. Como a necessidade é de formar agrupamentos, a possibilidade de 1 agrupamento é descartada, sendo utilizada a indicação da largura média da silhueta (ASW) com 4 centros, conforme figura 6.

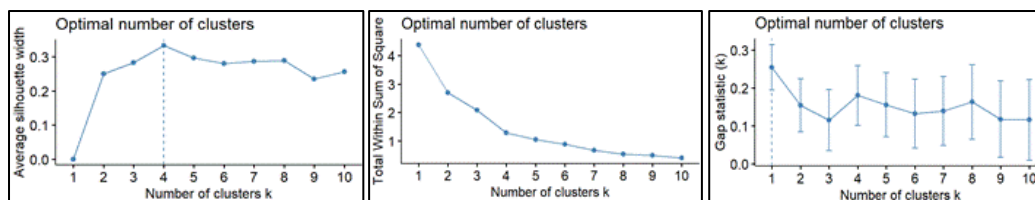


Figura 6 – Número ótimo de clusters: Silhouette (esq.); Gap stat (centro); WSS (dir.)

Fonte: Elaboração própria em software R.

**Clusterizar Plataformas:** a clusterização foi realizada com o algoritmo k-means. Entre as formulações de agrupamento baseadas na minimização da função objetivo, talvez a mais usada e estudada seja o agrupamento k-means. (KANUNGO *et al.*, 2002).

Vysala e Gomes (2020) explica o algoritmo que procede da seguinte forma: 1. Inicia aleatoriamente com um número de “centroids”; 2. Atribui cada ponto ao centróide de cluster mais próximo; 3. Calcula o centróide da nova partição formada tomando a média dos pontos atribuídos a esse cluster; 4. Repete as etapas 2 e 3 até que a convergência seja obtida.

O resultado da clusterização é apresentado na figura 7. Consta a duração normalizada no eixo X, a perda normalizada no eixo Y, e a idade como o tamanho da circunferência.

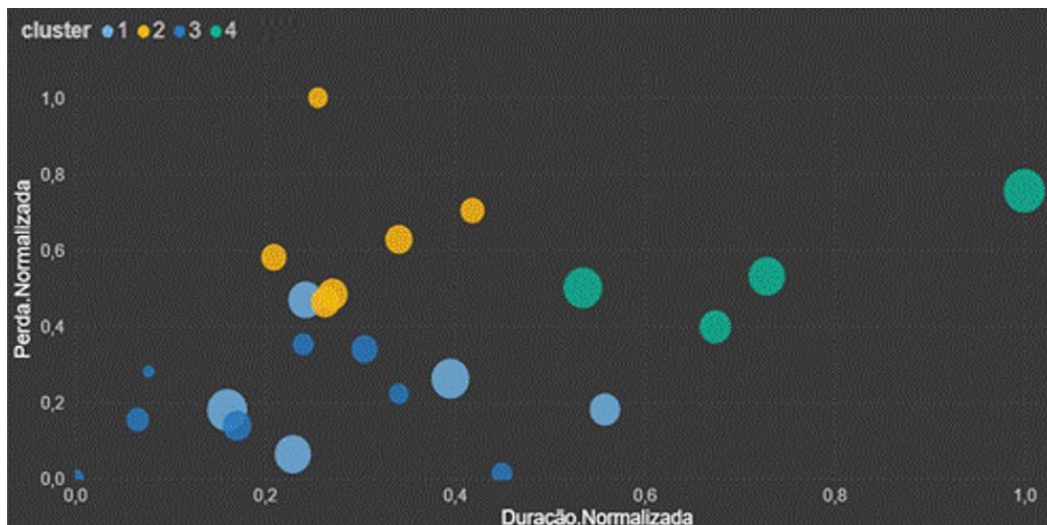


Figura 7 – Plataformas clusterizadas pelos eventos de paradas programadas

Fonte: Elaboração própria em software Power Bi.

Verifica-se a seguinte maior abrangência das plataformas por cluster:

- **Cluster 1:** idade mais avançada, mas não possuem alta duração, e volume de perdas.
- **Cluster 2:** alto potencial de produção, porém com idade e duração não relevantes. Merece atenção devido à possibilidade de alto impacto.
- **Cluster 3:** menores durações médias, perdas, e plataformas mais jovens. Possui menor desvio entre planejado e realizado.
- **Cluster 4:** maior duração de perdas e com idades mais avançadas. Possui maior desvio entre planejado e realizado.

**Parametrizar distribuição por cluster:** Com base no resultado da clusterização, os eventos semelhantes passam a ter histórico suficiente (entre 26 e 40 eventos por cluster), e permitem indicar possíveis riscos de postergações em relação aos eventos para aquele cluster. Dessa forma, é utilizada a diferença entre o realizado e o planejado de cada evento. Para o simulador, os parâmetros estatísticos da triangular fazem uso do mínimo, da média, e máximo por cluster, utilizando o desvio entre o realizado e planejado da duração dos eventos históricos.

**Simular duração com cálculo da perda:** a simulação é feita com as iterações da Simulação Monte Carlo em relação a duração da parada. Com base nos

dias simulados para um evento, multiplica-se pela variável aleatória simulada do potencial da mesma iteração e mês, obtendo a projeção da perda do evento.

## 5. Apresentação do aplicativo

Este capítulo apresenta a lógica e as telas desenvolvidas para utilizar o aplicativo disponibilizado para simular análise de risco da previsão de produção.

Foram utilizados casos de cenários como exemplo, cada qual com seus parâmetros específicos bem definidos, a fim de explicar as aplicações em cada uma das visões desenvolvidas, buscando utilizar diferentes métodos e iterações. Abaixo é explicado os casos que foram utilizados, seguindo o seguinte modelo:

**Caso X:** finalidade utilizada no trabalho (parâmetros utilizados).

- **Caso 1:** Simulação da entidade Eficiência Operacional (10 iterações; Método ELM; e ajustes de modificadores).
- **Caso 2:** Simulação da entidade Potencial Produtivo (10 iterações; Método ETS; e ajustes de modificadores).
- **Caso 3:** Simulação da Produção (10 iterações; Método ARIMA).
- **Caso 4:** Simulação da Produção (1000 iterações; Método ARIMA).
- **Caso 5:** Simulação da Produção (100 iterações, Método TBATs no potencial e ETS na eficiência).
- **Caso 6:** Simulação da Produção (100 iterações, Método com cenário modificado do Caso 1 na eficiência e do Caso 2 no potencial).

Os casos exemplificados comparam a previsão de produção para o total da Bacia de Campos, a fim de avaliar a produção de todas as plataformas da região, e visando simplificar a análise para focar no processo de simulação e na comparação dos resultados, em um mesmo escopo. Ao mesmo tempo, o simulador também se preocupa em disponibilizar a possibilidade de avaliar a previsão de cada plataforma, permitindo comparar a adequabilidade dos modelos, modificadores e parâmetros de entrada para cada uma das plataformas.

### 5.1 Disponibilizar aplicação

O aplicativo é iniciado com a apresentação da inicial “sobre o aplicativo” onde contém uma breve introdução do objetivo da aplicação, e com uma diagramação que explica a modelagem elaborada, ambos alinhados à explanação apresentada neste presente trabalho. É possível observar a tela inicial na figura 8.

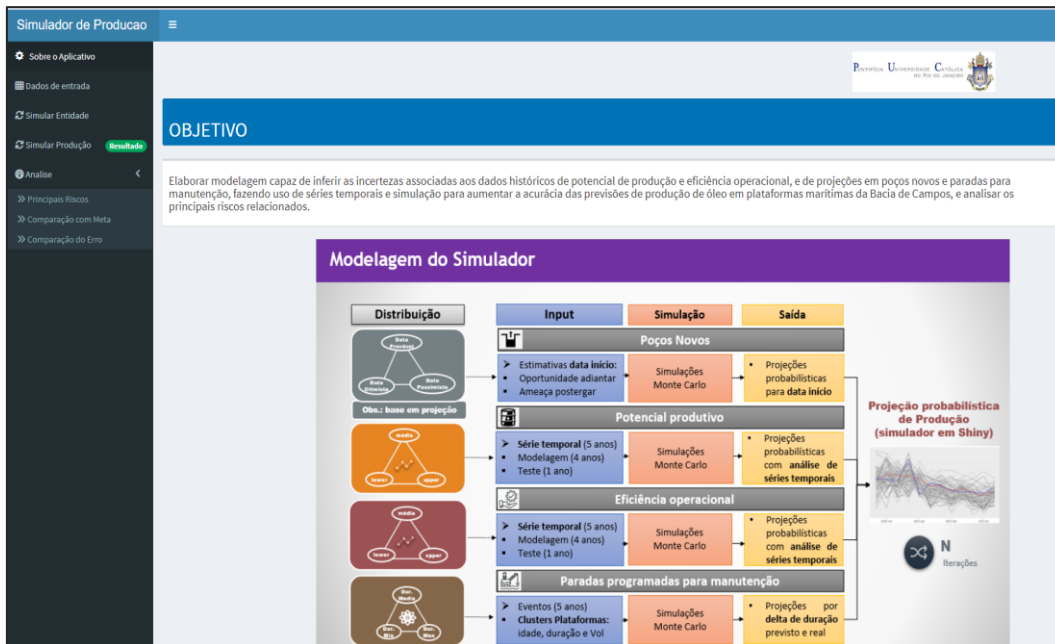
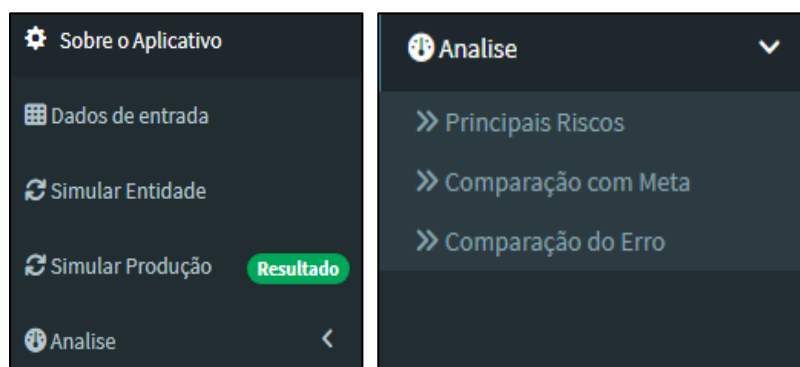


Figura 8 –Tela inicial da aplicação desenvolvida em *Shiny*.

Fonte: Elaboração própria em software R.

Conforme o menu de opções da figura 9, a aplicação possibilita ao usuário:

- **Dados de entrada:** analisar dados históricos das entidades de entrada do modelo;
- **Simular entidade:** testar a predição das entidades que fazem uso da análise de séries temporais (eficiência e potencial);
- **Simular produção:** avaliar resultado da simulação das quatro entidades de entrada, resultando na curva mensal de produção;
- **Realizar análise:** verificar os principais riscos, comparar resultado anual com a meta pretendida, e comparar o erro de predição entre os métodos de séries temporais.



(a) Menu de opções

(b) Submenu da “análise”

Figura 9 – Menu de opções do aplicativo.

Fonte: Elaboração própria em software R.

## 5.2 Analisar dados de entrada

A análise dos dados de entrada (2017 a 2021) é disponibilizada para ser visualizada e analisada para cada entidade. O potencial produtivo da região, correspondente à soma de todas as plataformas, é mostrado na figura 10.

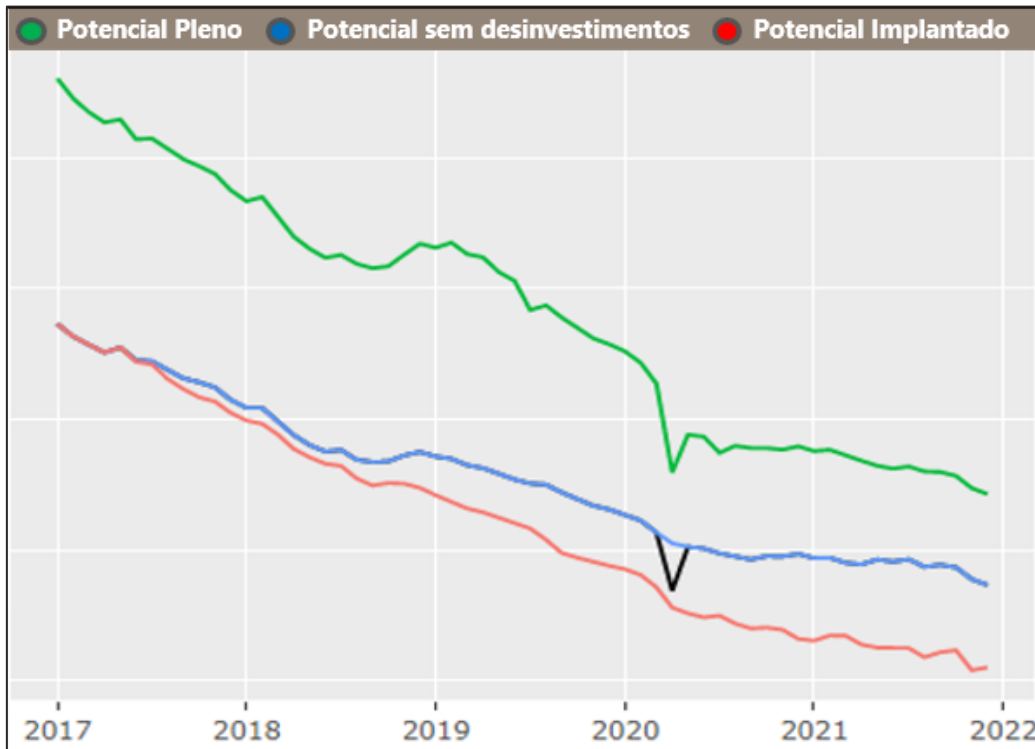


Figura 10 – Dados históricos do potencial de produção da bacia geográfica de Campos.

Fonte: Elaboração própria em software R.

O potencial pode ser dividido em:

- **Potencial Pleno:** potencial total no horizonte da análise.
- **Potencial sem desinvestimentos:** potencial pleno subtraído pelas 29 unidades vendidas ou hibernadas no período. Esse é o potencial utilizado para predição e simulação do cálculo da produção.
- **Potencial Implantado:** potencial sem desinvestimento subtraído pela parcela dos poços novos que iniciaram no horizonte da análise. Foi utilizado apenas para testes preliminares que definiram o uso dos 4 métodos de séries temporais a serem disponibilizados no aplicativo. O uso do potencial implantado permite analisar o efetivo declínio do potencial pois retira melhorias de performance dos novos poços.

O histórico da eficiência e dos poços novos podem ser visualizados na figura 11. É possível verificar a tendência de declínio da eficiência que possuiu maior significância por volta de 2019, mas que os atuais investimentos mostram diminuir o declínio atual.

Em relação aos novos poços, três plataformas tiveram investimentos significativos em 2019, elevando a proporção de novos poços, por outro lado, é normal que esses poços iniciem com potencial maior que o normal, então seus incrementos tiveram relativa diminuição em 2020, mas que também voltaram a ter incremento em seguida, devido à novos investimentos.

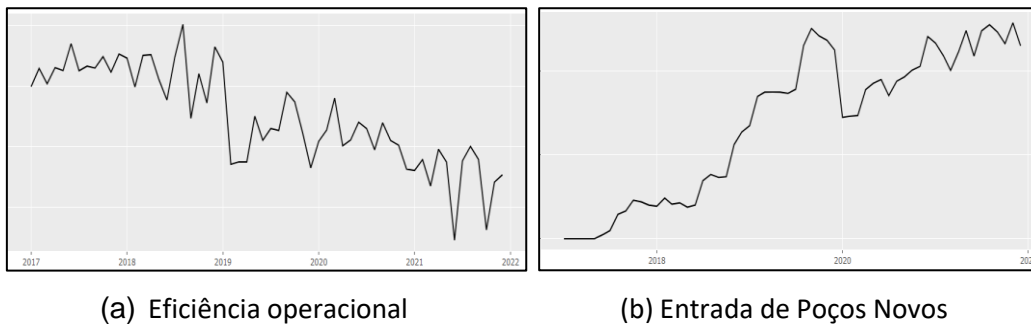


Figura 11 – Eficiência Histórica e Poços Novos da bacia geográfica de Campos

Fonte: Elaboração própria em software R.

O histórico de Perdas com Paradas é mostrado em gráfico específico, na figura 12, ele permite mostrar as diferenças entre os quatros clusters obtidos. O gráfico mostra o total de perdas em média diária (em gráfico de coluna), e de duração histórica em dias (gráfico de linhas). É possível verificar a maior proporção das durações para o cluster 4. Em relação às perdas médias, o cluster 2 tem maiores valores. Ambos estão em consonância à análise dos resultados verificados no momento da clusterização das paradas.

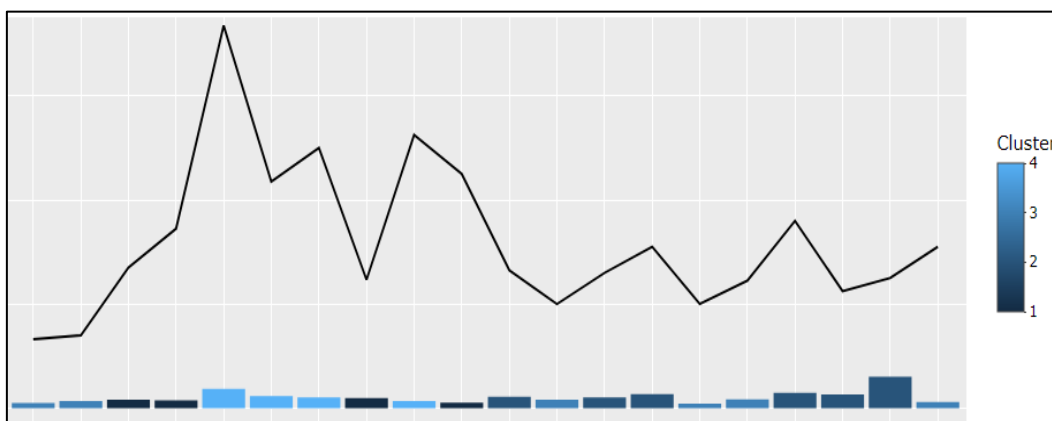


Figura 12 – Total histórico de Perda (colunas) e Duração (linhas) por Plataforma

Fonte: Elaboração própria em software R.

### 5.3 Simular Entidades

O primeiro ambiente de simulação disponível, é exclusivo para a análise das entidades que fazem uso de análise de séries temporais: eficiência ou potencial.

É possível executar análises sobre a previsão por plataforma, ou para o total da região, em cada entidade. Os testes são realizados com apenas 10 iterações para ter resultados preliminares rápidos.

Nesse momento o planejador pode observar a análise da contribuição mensal, com entendimento da previsão da série temporal, e é possível agregar o julgamento humano no *output* da previsão. Dessa forma, o julgamento também é realizado na saída do modelo. Essa possibilidade é bastante relevante ao modelo pois o especialista pode agregar seus conhecimentos na previsão em modo conjunto ao modelo preditivo, principalmente em relação a conhecimentos futuros que são diferentes do desempenho no passado, permitindo calibrar os dados, principalmente em relação a: ajuste de perdas históricas que têm projeção de mitigação; expurgos de perdas disruptivas, sem previsão de recorrências; previsão de abandono de poços implantados ou casos inesperados no potencial (ex.: fechamento por COVID-19).

O simulador permite parametrizar até duas modificações para a projeção. Após calibrar os dados, é possível salvar esse novo cenário para poder utilizar na simulação da produção.

Nessa etapa também é possível comparar particularidades entre os métodos de previsão, principalmente averiguando alguma especificidade por plataforma, permitindo ao especialista de cada plataforma analisar a adequabilidade dos métodos disponíveis no simulador e incluir suas particularidades na análise.

### **Simulação da entidade eficiência operacional**

A simulação realizada para a entidade de eficiência operacional é apresentada na figura 13. O simulador disponibiliza a opção de selecionar a plataforma, o método de série temporal e o número de modificações a serem realizadas no cenário.

Os modificadores podem ser 0, 1 ou 2. Sinalizar 0 significa que não terá nenhuma alteração em relação a previsão. Assim, pode ser determinado até 2 níveis diferentes de impacto na projeção anual, cada qual impactando em diferentes meses.



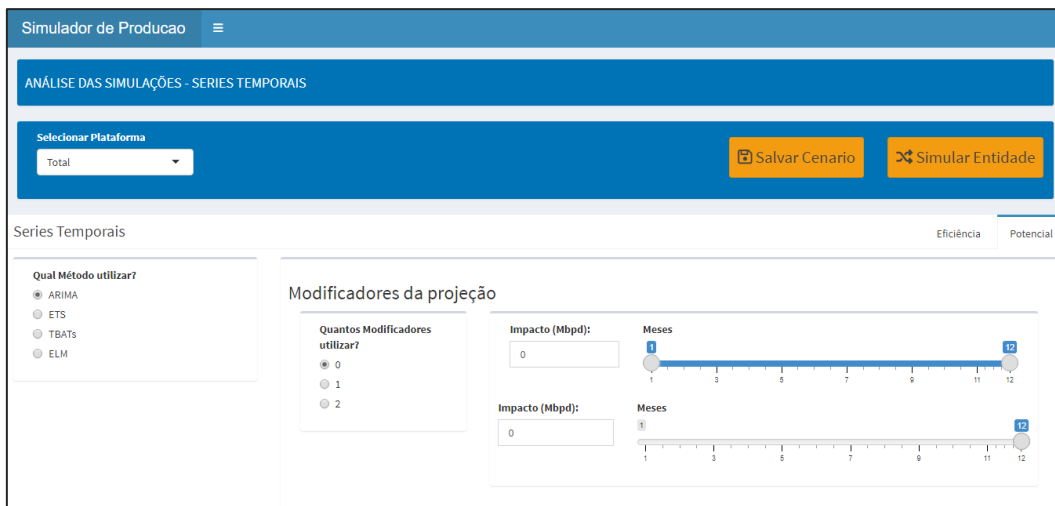


Figura 13 – Parâmetros disponíveis para a simulação das entidades

Fonte: Elaboração própria em software R.

A caracterização do caso utilizado como exemplo corresponde aos parâmetros definidos para o “Caso 1”, com uso do método de predição ELM e ajustes nos modificadores. As alterações caracterizam um novo cenário com base em alterações na projeção influenciada pelos conhecimentos dos especialistas.

No exemplo apresentado na figura 14, o cenário modificado simulado considera uma situação em que, de janeiro a abril, o potencial apresentará uma projeção 4% menor que a projeção mais provável. A partir de maio, espera-se uma projeção 8% acima da projeção mais provável. Esse caso hipotético pode corresponder a uma nova possível perda de produção devido a problemas na capacidade de compressão de gás natural, e que tenha um plano de ação para realizar a manutenção do compressor até maio. Conseqüentemente, a plataforma operará com perdas nos primeiros meses, mas espera-se que sejam mitigadas quando a compressão for restaurada.

O caso rodou as 10 iterações de teste em 86 segundos, e a previsão da eficiência para o novo cenário configurado teve erro de 9,18%. A visualização da simulação permite analisar as contribuições mensais simuladas (em cinza), a P50 com o cenário base simulado (em azul escuro), a P50 com as modificações propostas pelos especialistas (em azul claro) e compará-las com os resultados realizados (em laranja). Essa análise resulta em indicadores que demonstram o erro calculado pelo MAPE (representado por um cartão na cor roxa) e o tempo de processamento (representado por um cartão na cor azul).

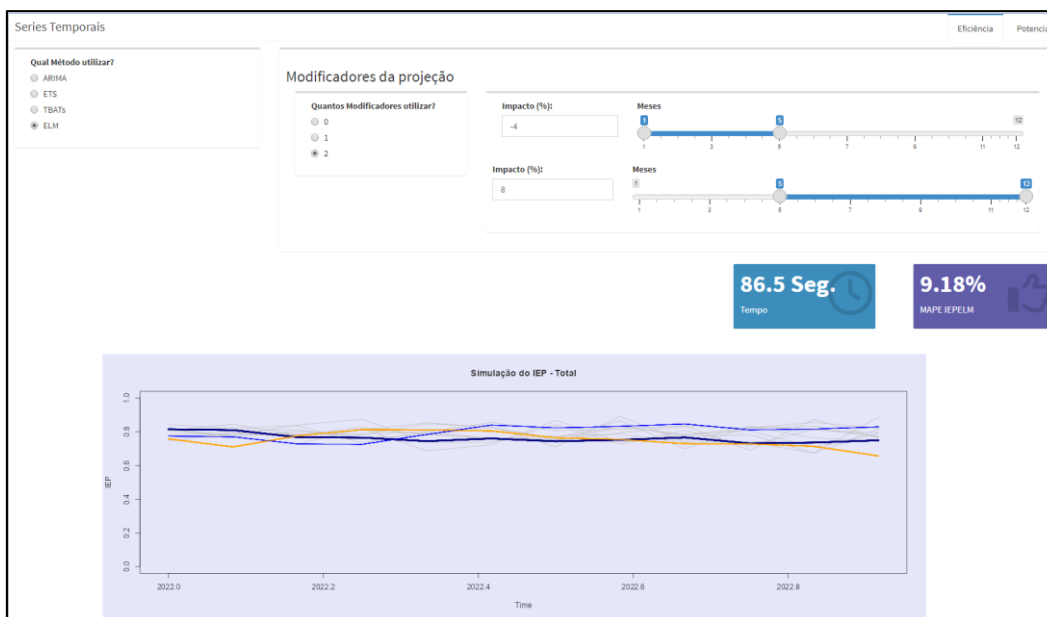


Figura 14 – Simulação das entidades com análise de séries temporais: eficiência

Fonte: Elaboração própria em software R.

### Simulação da entidade potencial produtivo

A diferença entre a visualização da entidade de eficiência e a do potencial existe apenas na unidade de referência. Enquanto a eficiência apresenta mudanças nas projeções em pontos percentuais, o potencial é calculado em Mbpd (Mil barris por dia). As demais lógicas entre as duas visualizações são semelhantes.

A caracterização do caso corresponde aos parâmetros definidos para o “Caso 2”, com uso do método de predição com a suavização exponencial (ETS) e ajustes nos modificadores, conforme apresentado na Figura 15.

O especialista também pode determinar um novo cenário de projeção para o potencial produtivo. No exemplo ilustrado na figura, existe um único modificador no intervalo entre os meses de abril e outubro, em que a projeção do potencial terá um decréscimo de -42 Mbpd. Essa situação pode ser caracterizada pelo conhecimento de poços que estavam em operação, mas que precisarão ser fechados devido à necessidade de intervenção durante esse período. Portanto, a projeção já pode ser ajustada levando em consideração essa modificação. Ademais, a projeção representada rodou as 10 iterações de teste em 6 segundos, e a previsão do potencial teve erro de 3,46%.



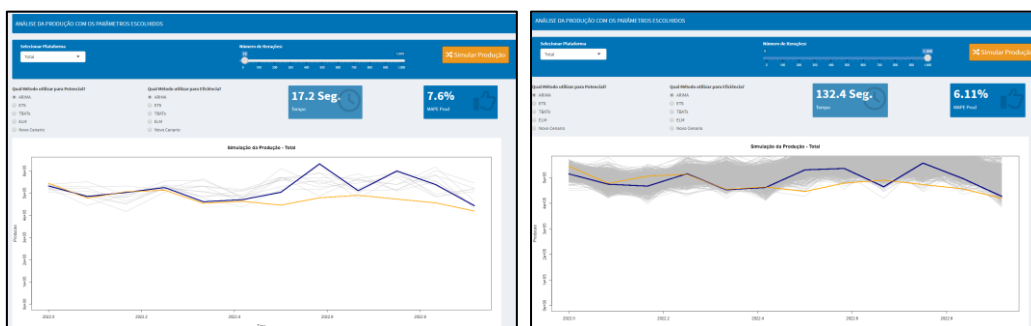
Figura 15 – Simulação das entidades com análise de séries temporais: potencial

Fonte: Elaboração própria em software R.

## 5.4 Simular Produção

A simulação da produção agrega os resultados obtidos nas simulações das entidades de entrada, resultando na curva mensal de produção, em Mil barris por dia (Mbpd).

Conforme ilustrado na Figura 16, é possível selecionar a plataforma, o método de séries temporais ou o cenário modificado pelos especialistas, e o número de iterações, permitindo simular diferentes cenários de produção. A disponibilização dessas opções permite que o especialista tenha a liberdade de definir o melhor método para o seu caso, dando-lhe autonomia e discernimento em relação ao método mais adequado para utilizar.



(a) ARIMA e 10 iterações

(b) ARIMA e 1000 iterações

Figura 16 – Simulação da Produção: comparar erro pelo número de iterações

Fonte: Elaboração própria em Software R.

Recomenda-se, ao final do processo de calibração das curvas, executar a simulação da produção com 1000 iterações para verificar o erro com precisão.

Os exemplos mostram uma comparação dos resultados entre a caracterização do “Caso 3” e “Caso 4”, ambos fazem uso do método ARIMA e possuem variação no número de iterações. Os casos avaliam o valor total das plataformas, sendo a única diferença o número de iterações. Nesse contexto, o intervalo do risco é representado (linhas em cinza), e o cenário mais provável (representado em azul) é comparado com o realizado (representado em laranja).

É observado que com 10 replicações o simulador processa rapidamente em 17 segundos, e com 1000, ainda com performance aceitável, em 4 minutos e 23 segundos. O uso de 1000 iterações calibra bem a projeção mais provável do parâmetro, e representa melhor o túnel de risco da projeção.

O simulador também permite combinar diferentes métodos, conforme representado na figura 17 que utiliza os parâmetros do “Caso 5”. A produção é simulada a partir de combinações, utilizando um método para a eficiência e outro para o potencial. De posse dos testes das simulações das entidades, facilita saber os melhores modelos de simulação para cada plataforma, e utilizar essa combinação no momento de simular a produção.

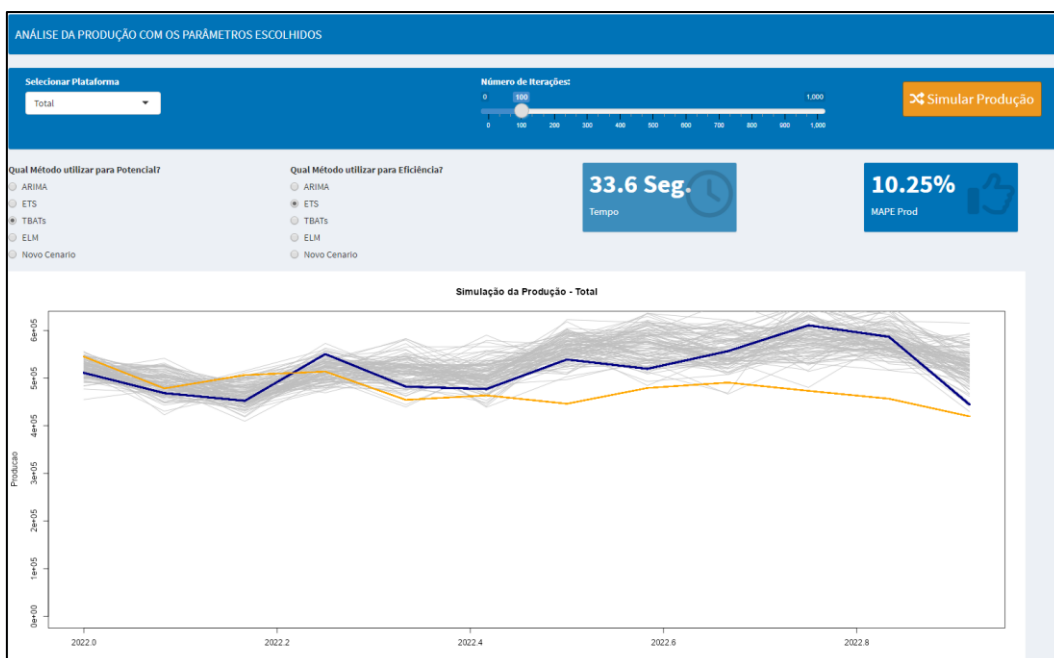


Figura 17 – Simulação da Produção: mesclar métodos de previsão

Fonte: Elaboração própria em Software R.

A combinação do método preditivo com o julgamento humano nos dados de saídas do modelo é disponibilizada através da opção de seleção do método “novo cenário”. Neste sentido, é possível utilizar o último cenário armazenado na simulação de uma entidade, seja com modificações no percentual de eficiência ou utilizando o cenário modificado para o potencial produtivo, a fim de simular a produção, conforme figura 18. Ele roda a simulação com os parâmetros do “Caso 6”, usando o cenário modificado do “Caso 1” na eficiência e “Caso 2” no potencial.

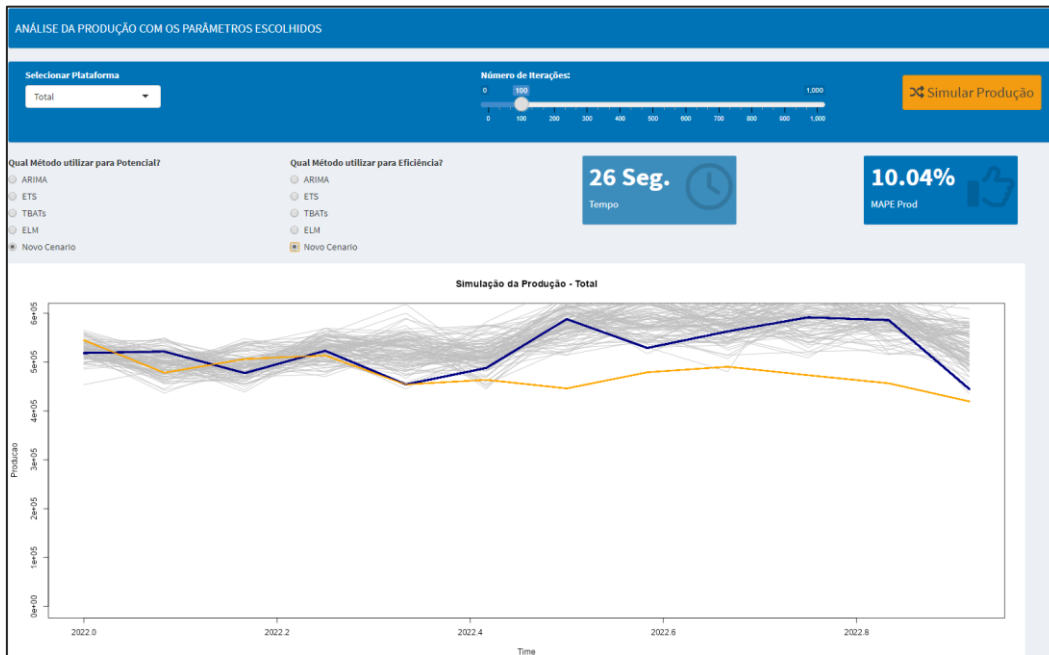


Figura 18 – Simulação da Produção: julgamento dos especialistas nos dados de saída

Fonte: Elaboração própria em Software R.

## 6. Análise dos resultados

Este capítulo tece análise a respeito dos resultados obtidos, tanto em relação à comparação da acurácia entre os métodos quanto à identificação dos principais riscos e à determinação da meta prevista.

Para a análise dos resultados, a comparação dos erros foi realizada simulando todos os quatro possíveis métodos com 1000 iterações a fim de comparar o resultado para todos disponibilizados. Para o resultado dos principais riscos e definição da Meta, foi utilizado o caso 4 como caso base, o qual faz uso do método ARIMA no potencial e na eficiência.

### 6.1 Comparação dos erros das séries temporais e produção

O aplicativo permite comparar o erro entre o previsto e a curva mais provável simulada. O erro é calculado para as variáveis que analisam as séries temporais, como a entidade de eficiência de produção e potencial produtivo, além de verificar o erro para a produção resultante.

Com base no cálculo do erro, é possível inferir qual dos quatro métodos de série temporal apresenta maior acurácia, seja para uma plataforma específica ou para o total da Bacia de Campos. A Figura 19 evidencia os resultados com 1000 iterações para o total da Bacia, sinalizando em destaque a principal informação que é o erro para a previsão de produção resultante.

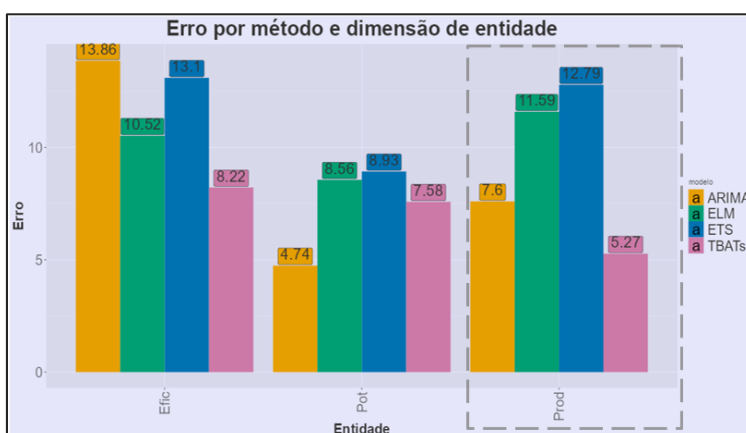


Figura 19 – Comparativo entre os erros (MAPE) das previsões

Fonte: Elaboração própria em Software R.

O resultado mostra que o método TBATs apresentou a maior acurácia no cálculo da produção, (MAPE de 5,3%), enquanto o método ARIMA é o segundo maior (MAPE de 7,6%). Uma segunda análise é que, para todos os métodos utilizados, a variável de potencial apresenta um erro menor do que a eficiência, o que denota maior assertividade do seu processo de previsão.

## **6.2 Análise dos principais riscos**

Em relação às entidades que não fazem uso de séries temporais, é importante verificar o resultado das simulações e se aprofundar em seus detalhes, por meio de gráficos que comparam os planejamentos determinísticos para o ano com os resultados de suas variáveis simuladas.

### **Análise dos riscos em paradas programadas**

A simulação, baseada em dados históricos estatísticos agrupados em clusters, pode indicar durações bem diferentes das planejadas no planejamento determinístico. A ideia é que as simulações subsidiem possíveis adequações nos planejamentos dos especialistas. Essa abordagem de correção quantitativa possibilita que a simulação auxilie no julgamento humano para a definição do planejamento, evidenciando se a parada deveria ter uma duração maior ou menor do que o planejado.

A figura 20 mostra os principais riscos identificados para o planejamento de paradas, após simulação com 1000 iterações. Os pontos em verde representam o planejamento determinístico para o ano, enquanto a densidade em azul representa o que foi simulado.

Como exemplo, a plataforma mais à esquerda da figura 19 indica a necessidade de um planejamento com duração maior, uma vez que as simulações baseadas no histórico mostram densidades em azul com durações maiores do que o planejado em verde. Por outro lado, as simulações para as duas plataformas mais à direita indicam considerar uma duração um pouco menor do que a estimada pelos especialistas.



Figura 20 – Principais riscos para as paradas planejadas

Fonte: Elaboração própria em Software R.

#### Análise dos riscos em poços novos

A análise gráfica também é utilizada na simulação para a outra entidade que não faz uso de séries temporais: poços novos. O gráfico da figura 21 facilita mapear riscos através do detalhamento da simulação para cada poço. São destacadas as datas planejadas pelos especialistas, como a data mais provável (em verde), a data conservadora (em vermelho) e a data otimista (em azul). Além disso, são mostradas as simulações (em cinza).

Para essa análise, observa-se que o número de iterações foi reduzido para 10, visto que sem a iteratividade do gráfico, apenas com a imagem estática, dificultaria a visualização de todos os pontos para um caso de 1000 iterações.

Em relação aos resultados da simulação deste caso, é possível observar, por exemplo, que o poço 2 possui uma maior possibilidade de antecipação, com simulações entre a data provável e a data otimista, evidenciando uma maior chance de ser implementado antes do prazo. Por outro lado, os poços 1, 3, 8 e 11 apresentam maiores possibilidades de serem implementados mais tarde, indicando um maior risco de postergação associado a eles. Portanto, esses poços demandam maior atenção e a elaboração de um plano de ação para mitigar esses riscos. Vale ressaltar que a incerteza da simulação dos poços está de acordo com as principais oportunidades e ameaças identificadas, e a respectiva probabilidade de ocorrência.



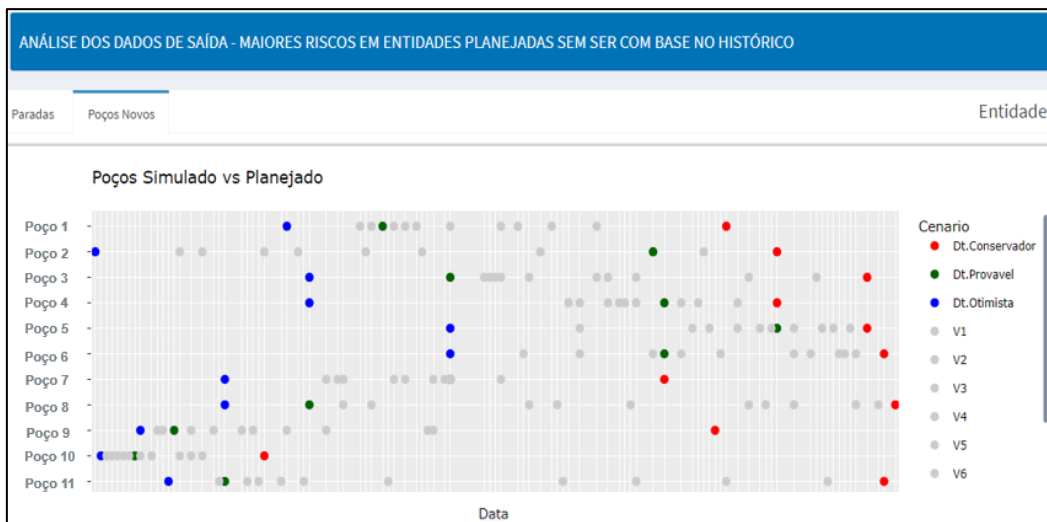


Figura 21 – Principais riscos para os poços novos

Fonte: Elaboração própria em Software R.

Os parâmetros para a simulação também são disponibilizados na parte inferior da tela do simulador, conforme mostrado na Figura 22.

Poço	Cat.Otimista	Cat.Conservador	Otimista	Conservador	Mais.Provavel	Data Entrada (Plan. Det.)	Data Entrada (real)	Delta
Poço 1	Recurso	Recurso	-30	60	0	21-07-2022	30-12-2022	162
Poço 2	Suprimentos	Suprimentos	-211	31	0	31-08-2022	08-05-2023	250
Poço 3	Recurso	Anuência	-31	92	0	01-08-2022	06-08-2022	5
Poço 4	Anuência	Anuência	-62	30	0	01-09-2022	13-03-2022	-172
Poço 5	Recurso	Suprimentos	-61	31	0	01-10-2022	18-10-2022	17
Poço 6	Anuência	Projeto	-31	91	0	01-09-2022	10-12-2022	100
Poço 7	Anuência	Anuência	-61	31	0	01-08-2022	09-03-2022	-145
Poço 8	Suprimentos	Projeto	-30	184	0	01-07-2022	06-01-2023	189
Poço 9	Recurso	Projeto	-30	120	0	16-05-2022	07-08-2022	83
Poço 10	Suprimentos	Suprimentos	-30	60	0	15-04-2022	12-12-2022	241
Poço 11	Recurso	Recurso	-31	183	0	01-06-2022	28-08-2022	88

Figura 22 – Parâmetros para os poços novos

Fonte: Elaboração própria em Software R.

As colunas com as categorias ("Cat.") indicam os parâmetros otimista ou conservador, permitindo interpretar o risco associado a cada poço. Como exemplo, o maior valor presente na coluna "conservador" relacionada ao poço 8, representa a possibilidade de postergar em 184 dias a data de entrada do poço, esse valor é o utilizado no parâmetro mais conservador da distribuição triangular simulada, e a explicação para essa possibilidade de entrar após 184 dias é devido a mudança no escopo do projeto para evitar o fechamento de outro poço associado. Nesse mesmo caso, caso a definição do projeto seja rápida, o poço pode ser interligado até mais cedo, a depender de acessório necessário para interligação que estava sendo

adquirido, correspondendo a outro ponto crítico para a determinação na data de entrada.

Nessa mesma tabela, também foram inseridos os dados efetivamente realizados em 2022, o que permite comparar melhor o resultado do modelo em relação à simulação para os poços novos. As últimas três colunas permitem observar o que estava planejado deterministicamente pelos especialistas para a meta da companhia, alinhado ao parâmetro da data mais provável do modelo, com o que realmente realizou.

O desvio mostra a diferença entre o efetivamente realizado para o planejamento mais provável. Além disso, foi utilizada as equações 6 e 7 para avaliar o intervalo possível considerado no planejamento pelos especialistas:

Equação 6 – Cálculo da data conservadora para a entrada de poços novos

$$\sum_{t=1}^T \sum_{p=1}^P Dt. Conservadora_p [t] = \sum_{t=1}^T \sum_{p=1}^P Dt. Inicio_p [t] + \sum_{t=1}^T \sum_{p=1}^P \Delta Dt. Simulada_p [t]$$

Equação 7 – Cálculo da data otimista para a entrada de poços novos

$$\sum_{t=1}^T \sum_{p=1}^P Dt. Otimista_p [t] = \sum_{t=1}^T \sum_{p=1}^P Dt. Inicio_p [t] + \sum_{t=1}^T \sum_{p=1}^P \Delta Dt. Simulada_p [t]$$

Onde o intervalo entre a data conservadora e a data otimista são avaliadas em relação a data de entrada realizada para cada poço “p”.

Observa-se que 64% das estimativas não tiveram a data efetiva de entrada dos poços dentro do intervalo considerado, mostrando ser importante o aprofundamento nas estimativas dos especialistas. Utilizando o mesmo exemplo do poço 8, a estimativa de entrada do poço era entre 01/07/22 e 01/10/22, e mesmo assim ele só entrou em operação no dia 06/01/23.

### 6.3 Definição da meta anual

Além da curva mensal de produção, é imprescindível analisar a média anual e sua distribuição simulada. Tal análise auxilia na definição da meta embasadas em dados numéricos e de modo objetivo.

A figura 23 auxilia na definição das metas de produção neste sentido. No caso, ao estabelecer o valor de uma meta, o gráfico retorna o grau de probabilidade de superá-la em relação a distribuição simulada para a produção.

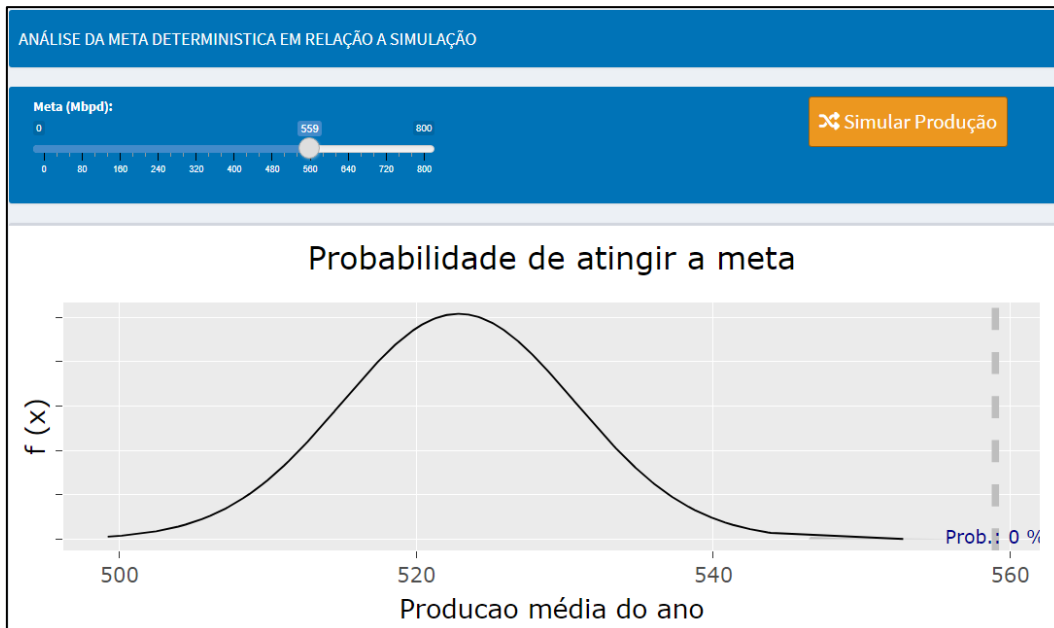


Figura 23 – Probabilidade de superar a meta de 559 Mbpd em relação a simulação

Fonte: Elaboração própria em Software R.

Em uma situação hipotética, ao considerar uma meta de produção anual de 559 Mbpd, de acordo com os riscos identificados no simulador, a probabilidade de alcance seria de 0%.

Em outro caso hipotético, uma meta de 526 Mbpd corresponderia a uma probabilidade de alcançar a meta em 31%, como ilustrado na figura 24.

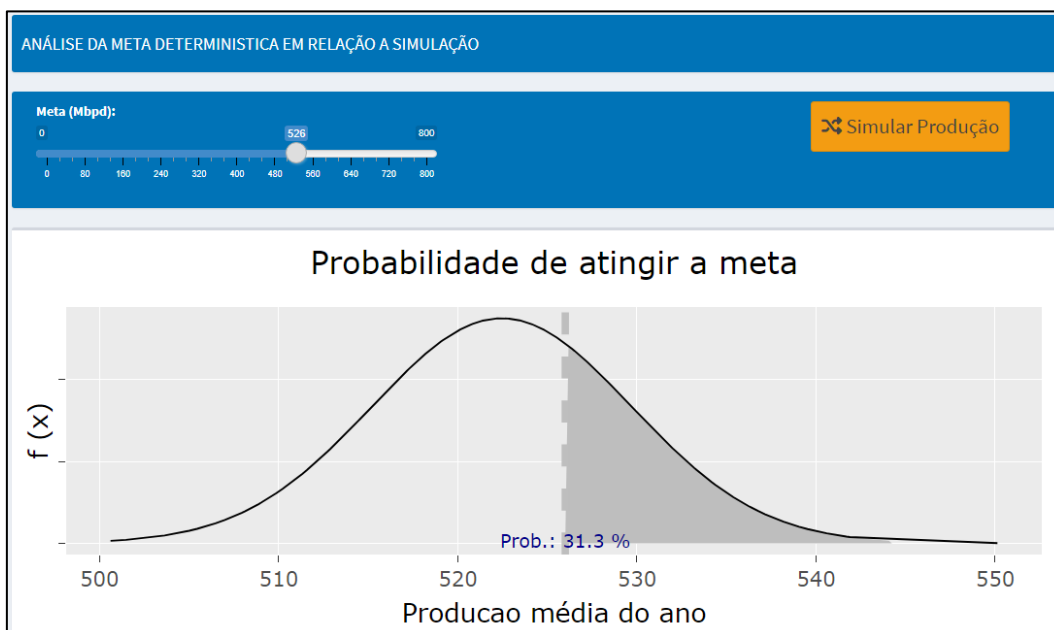


Figura 24 – Probabilidade de superar a meta de 526 Mbpd em relação a simulação

Fonte: Elaboração própria em Software R.

Ao efetuar uma análise do resultado anual para a bacia de Campos em comparação com o valor mais provável de produção resultante do simulador, observa-se resultados próximos ao valor efetivamente realizado em 2022. Utilizando dois casos como exemplo, verifica-se uma mínima variação com valores próximos a 1%, conforme tabela 5.

Tabela 5 – Resultados da variação entre simulação e realização no ano

<b>Caso (exemplo)</b>	<b>Variação: realização e simulação (P50)</b>
Caso 4	0,9%
Caso 5	1,3%

Fonte: Elaboração própria.

O caso 4 que foi utilizado como base para a análise dos resultados, apresentou diferença em 0,9%. Esse resultado evidencia a alta precisão do modelo, a qual pode ser ainda mais aprimorada ao considerar os julgamentos humanos dos especialistas, que ainda não foram considerados nesse caso. Essa possibilidade passa a ser disponibilizada aos especialistas no simulador e pode contribuir para refinar os resultados.

## 7. Considerações Finais

O trabalho alcança seus principais propósitos, uma vez que a nova modelagem proposta torna a aplicação mais flexível ao negócio da empresa, baseando-se em dados objetivos e eliminando subjetividades decorrentes de estimativas pessoais.

As comparações dos resultados demonstram possíveis melhorias em seu principal objetivo, ao permitir avaliar diferentes modelos e indicar qual obtém maior acurácia. É verificado o benefício da aplicação de séries temporais para avaliar a tendência de declínio de potencial e eficiência, permitindo realizar análise crítica das projeções dos especialistas. A disponibilização do simulador permite verificar com frequência, diante de novos dados e particularidades das plataformas.

Ele também facilita a definição objetiva da meta de produção. Ao considerar os principais riscos, simular, e fornecer a probabilidade de retorno de possíveis metas, é possível subsidiar o estabelecimento de diferentes níveis para a meta anual da região, com uma clara probabilidade de alcançá-los, além de facilitar a subdivisão das metas por subconjuntos, baseando-se nos resultados mais esperados da simulação para cada plataforma.

A modelagem também possibilita ultrapassar os resultados analisados. Ao disponibilizar a possibilidade do simulador para avaliar os diferentes métodos, é possível avaliar não somente os dados de toda a Bacia, mas também avaliar especificidades e avaliar diferentes adequabilidades de casos para cada plataforma, com possibilidade de uso de diferentes casos de cenário e parâmetros de entrada, além de poder ter uma maior interação entre o simulador e o especialista durante e ao final do processo.

Outro benefício identificado é analisar os principais riscos, de modo a explorar as variáveis que não fazem uso da análise de séries temporais. Suas simulações permitem interpretar os riscos relevantes, seja por plataforma ou poço, identificando os principais eventos que podem provocar variações nas projeções, permitindo priorizar e traçar planos de ações para mitigá-los.

O protótipo do simulador foi disponibilizado para a empresa e testado pelos especialistas, permitindo o feedback para o aprimoramento do processo de modelagem, conforme disseminado pela metodologia de descoberta do conhecimento baseado em banco de dados. Ele serve de base, como protótipo, para a área de tecnologia da informação da empresa condicionar infraestrutura, e aprimorar modelo mais robusto para uso pelos especialistas em planejamento e controle de produção da Bacia de Campos, e posteriormente com possibilidade de implantação para toda a empresa.

Recomenda-se, para trabalhos futuros, explorar variações nos dados de entrada e análise de sensibilidade para o modelo. Também é sugerido extrapolar a análise com base histórica para a variável de poços novos, uma vez que o resultado do ano analisado extrapolou as estimativas dos especialistas na maior parte dos casos. Para as paradas de manutenção, é interessante realizar a clusterização incluindo outros parâmetros técnicos, como: quantidade de serviços de manutenção (danos no *flare*, trechos retos, furo ou vazamentos em linhas), além de aspectos como condições de mar e climáticas.

Em relação às séries temporais, sugere-se avaliar uso de séries hierárquicas para a previsão do nível das plataforma influenciar diretamente na previsão do total da Bacia, e diante da análise dos resultados, recomenda-se melhorar o processo de elaboração da projeção da entidade com maior erro, a eficiência, fazendo uso de nova possibilidade implementada, na qual os especialistas podem realizar modificações nas projeções, de modo a mapear as mudanças nas projeções futuras com base em detalhamentos tangíveis, como ações e investimentos planejados.

## Referências

Abath, J. R. **Formação de Portfólio sob Incerteza de uma Empresa de Produção e Refino de Petróleo**. Dissertação de Mestrado – PUC-RIO. Rio de Janeiro. 2020.

ANP. **Brasil produz cerca de 3 milhões de bbl/d de petróleo em abril**. *Boletim Mensal de Produção de Petróleo e Gás Natural da Agência Nacional de Petróleo, Agência Nacional do Petróleo, Gás Natural e Biocombustíveis*. 2022. Publicado em 03/06/2022. Disponível em: [https://www.gov.br/anp/pt-br/canais\\_atendimento/imprensa/noticias-comunicados/brasil-produz-cerca-de-3-milhoes-de-bbl-d-de-petroleo-em-abril](https://www.gov.br/anp/pt-br/canais_atendimento/imprensa/noticias-comunicados/brasil-produz-cerca-de-3-milhoes-de-bbl-d-de-petroleo-em-abril).

Baets, Shari De. Harvey, Nigel. **Using judgment to select and adjust forecasts from statistical models**. *European Journal of Operational Research*. Vol. 284. p. 882-895. 2020.

Becerra et al. **Uncertainty History Matching and Forecasting, a Field Case Application**. SPE Latin America and Caribbean Petroleum Engineering Conference. Mexico. 2012.

Bomfim, G. A. **Modelagem da Data de Entrada em Produção de Poços de Petróleo Utilizando Inferência Fuzzy**. PUC-Rio, Rio de Janeiro. 2017.

Box, G. e Jenkins, G. **Time Series Analysis: Forecasting and Control**. Holden-Day, San Francisco. 1970.

Brau, Rebekah., Aloysius, John., Siemsen, Enno. **Demand planning for the digital supply chain: How to integrate human judgment and predictive analytics**. *Journal of Operations Management*. 2023.

Broomhead, D. S. e Lowe, David. **Multivariable Functional Interpolation and Adaptive Networks**. *Complex Systems*, Vol.2, p. 321-355. 1988.

Buhmann, Martin Dietrich. **Radial Basis Functions: Theory and Implementations**. Vol. 12 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press. 2003.

Castro, A. O. de S., Mattos, J. R., Silva, S. B. P. **Job Risk - Análise Probabilística da Previsão de Produção de Óleo**. Seminário de Planejamento, Controle e Avaliação do Desempenho do E&P. Anais...Rio de Janeiro. 2010.

Chang, Winston et al. **Shiny: Web Application Framework for R**. 2021.

Correia, J. B., Pivetta, M., Nascimento, G. S. N., Becker, K. **Comparing ARIMA and LSTM models to predict time series in the oil industry**. p. 129-136. 2021.

Dancho, Matt, Vaughan D. **timetk: A Tool Kit for Working with Time Series**. 2023.

De Livera A. M., Hyndman R. J. & Snyder R. D. **Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing**. *Journal of the American Statistical Association*, 106:496, p. 1513-1527. 2011.

Fayyad, U, Piatetsky-Shapiro, G, Smyth, P and Uthurusamy, R (eds), **Advances in Knowledge Discovery and Data Mining**. AAAI Press. 1996.

Ferreira, V. M. (2016). **Avaliação de métodos de recuperação melhorada de petróleo para campos marítimos no Brasil: o caso da Bacia de Campos**. Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, RJ; Joelian, Albert. "Engine replacement scheduling optimization using Data Mining." *Journal of Physics: Conference Series*. Vol. 1500. No. 1. IOP Publishing. 2020.

Goodfellow, I., Bengio, Y., & Courville, A. **Deep Learning**. MIT Press. 2016.

Grolemund, G., Wickham, H. **Dates and Times Made Easy with lubridate**. *J. Stat. Softw.* Vol. 40, p. 1–25. 2011.

- Guerillot, D. e Bruyelle, J. **Uncertainty Assessment in Production Forecast with an Optimal Artificial Neural Network**. 2017.
- Guimarães, A. Wilmersdorf, R. Araújo, E. **Análise de incertezas em reservatório de petróleo: Uma abordagem estocástica usando métodos não intrusivos**. Dissertação de mestrado. Pós-graduação em Engenharia Civil. UFPE. 2016.
- Haghighati, R. e Hassan, A. **Recognition performance of imputed control chart patterns using exponentially weighted Moving average**. *European J. Industrial Engineering*, n. 5, Vol. 12. 2018
- Hurst, A., Brown, G. C., Swanson, R.I. **Swanson's 30-40-30 rule**. *AAPG Bulletin*, v. 84, no. 12, December 2000. p. 1883–1891. 2000.
- Huang, G., Huang, G. B., Song, S., You, K. **Trends in Extreme Learning Machines: A Review**. *Neural Networks*. 2014.
- Huang, Guang-Bin., Zhu, Qin-Yu., Siew, Chee-Kheong., **Extreme learning machine: Theory and applications**. *Neurocomputing*. Vol. 70, 2006.
- Hyndman, R.J., e Athanasopoulos, G. **Forecasting: principles and practice**. 3rd edition, OTexts: Melbourne, Australia. Disponível em: [OTexts.com/fpp3](https://www.otexts.com/fpp3). 2021. Acesso em: 22/11/2022.
- International Energy Agency. **World Energy Outlook 2021, IEA**, Paris, França. 2021. Disponível em: <https://www.iea.org/reports/world-energy-outlook-2021>. Acesso em: 22/11/2022.
- Kanungo, T., Mount, D.M., Netanyahu N.S., Piatko C.D., Silverman R., Wu A.Y.. et Al. **An efficient k-means clustering algorithm: analysis and implementation**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 24, no. 7, p. 881-892. 2002.
- Kourentzes, Nikolaos. **Nnfor: Time Series Forecasting with Neural Networks**. 2017.
- Lechner, J. P.; e G. Zangl. **Treating Uncertainties in Reservoir Performance Prediction with Neural Networks**. Paper presented at the SPE Europec/EAGE Annual Conference, Madrid, Spain. 2005.
- Lin, Suwen., Wu, Xian., Martinez, Gonzalo., e Chawla, Nitesh V. **Filling Missing Values on Wearable-Sensory Time Series Data**. *Proceedings of the 2020 SIAM International Conference on Data Mining (SDM)*. p. 46-54. 2020.
- Makridakis, S. G., Wheelwright, S. C. and Hyndman, R. J. **Forecasting: Methods and Applications**. 3 ed. John Willey & Sons. 1997.
- Mamudu, A. et al. **A hybrid intelligent model for reservoir production and associated dynamic risks**. *Journal of Natural Gas Science and Engineering*, Vol. 83. 2020.
- Mehdi Khashei, Mehdi Bijari. **An artificial neural network (p,d,q) model for timeseries forecasting**. *Expert Systems with Applications*. Vol. 37, Issue 1, p. 479-489. 2010.
- Meira, R., Gama, J., Mendonca, G. (2021). **Métodos de previsão em séries temporais: uma revisão**. *Revista Brasileira de Estatística*, vol. 76, nº 3, p. 61-78.
- Moritz Steffen, Bartz-Beielstein Thomas. **imputeTS: Time series missing value imputation in R**. Vol. 9, p. 207–218. Rio de Janeiro. 2017.
- Nunes, I.; Silva, H. S. D. **Artificial neural networks: a practical course**. Springer. 2018.
- Petropoulos, F., Kourentzes, N., Nikolopoulos, K., & Siemsen, E. **Judgmental selection of forecasting models**. *Journal of Operations Management*. Vol. 60, p. 34–46. 2018.
- Petropoulos, F., e Siemsen, E. **Forecast selection and representativeness**. *Management Science*. Vol. 69, p. 2672–2690. 2023.
- R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2021. Disponível em: <https://www.R-project.org>.
- Rosenblatt, F. **The perceptron: A probabilistic model for information storage and organization in the brain**. *Psychological Review*, Vol. 65, p. 386–408. 1958.
- Rumelhart, D. E., Hilton, G. E. e Williams, R. J. **Learning representations by backpropagation errors**. *Nature*. Vol. 323, p. 533–536. 1986.



- Suslick, S. B.; Schiozer, D. J. **Risk analysis applied to petroleum exploration and production: An overview.** Journal of Petroleum Science and Engineering, v. 44, n. 1–2, p. 1–9. 2004.
- Sievert, C. R **Package: Plotly.** Version 4.10.0. 2019.
- Taylor, S.J. Letham, B. **Forecasting at scale.** The American Statistician, 72:1, p. 37-45. 2017.
- Tealab, A.; Hefny, H.; Badr, A. **Forecasting of nonlinear time series using ann.** **Future Computing and Informatics Journal.** Elsevier, v. 2, n. 1, p. 39–47. 2017.
- Vanhoeyveld, J.; Martens, D. **Imbalanced classification in sparse and large behaviour datasets.** Data Mining and Knowledge Discovery, v. 32, n. 1, p. 25-82. 2018.
- Vergara, S. C. **Projetos e relatórios de pesquisa em administração.** 9 ed. São Paulo: Atlas, 2007.
- Vogt, M. **Combination of radial basis function neural networks with optimized learning vector quantization.** Proc. IEEE, vol. 83, p. 1841–1846. 1993.
- Vysala, A., Gomes, Dr. **Evaluating and Validating Cluster Results.** 2020.
- Wickham, Hadley. **Tidyverse: Easily Install and Load 'Tidyverse' Packages.** 2017.
- Wolodzko, T and Wolodzko, MT. **Package 'extraDistr'.** 2017.
- Xia, L. W. et al. **Prediction of CO2 Solubility in Ionic Liquids Based on Multi-Model Fusion Method.** Processes Journal. Vol. 7. 2019.
- Xie, Y.; Cheng, J.; Tan, X. **R Package: DT.** Version 0.22. 2022.
- Yan, M. et al. **Optical fiber sensor system for monitoring leakage current of post insulators based on RBF neural network.** ICHVE International Conference on High Voltage Engineering and Application, p. 1-4. 2014.
- Yanrui N., Hossein K., Pejman T. **A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet.** Computers & Geosciences, Vol. 164. 2022.
- Zellner, M., Abbas, A. E., Budescu, D. V., Galstyan, A. **A survey of human judgement and quantitative forecasting methods.** R Soc Open Sci. 2021.
- Zhang, G., Patuwo, B. E., e Hu, M.Y. **Forecasting with artificial neural networks: the state of the art.** international journal of forecasting. Vol. 14, p. 36-62. 1998.
- Zhong, Y., Liu, Y., Lin, X., Luo S., **The Method of Oilfield Development Risk Forecasting and Early Warning Using Revised Bayesian Network.** Mathematical Problems in Engineering. 2016.

## Apêndice

Esta seção apresenta a caracterização e conceitos fundamentais para compreensão sobre os oito métodos de previsão em análise.

**ETS:** Autores como Hyndman e Athanasopoulos (2021) desenvolveram métodos de suavização exponencial. O método de previsão ETS usa trinta modelos de suavização exponencial e os distingue de modo automático, avaliando o erro (E), tendência (T) e sazonalidade (S).

Os métodos de suavização exponencial atribuem pesos decrescentes exponencialmente à medida que a observação envelhece, ou seja, as observações recentes recebem um peso relativamente maior na previsão do que as mais antigas (Meira et al., 2021).

**ARIMA:** O *Autoregressive Integrated Moving Average* (ARIMA), conforme desenvolvido por Box e Jenkins (1970), combina componentes Autorregressivos (AR) entre uma observação e um número de observações defasadas (p), integrado (I) para tornar a série temporal estacionária medindo as diferenças de observações em diferentes tempos (d), e processos de Média Móvel (MA) levando em consideração a dependência entre as observações (q).

De modo geral, ARIMA é um modelo flexível e poderoso que pode ser usado para capturar padrões complexos em séries temporais, tais como tendências, sazonalidade e autocorrelação. (Hyndman e Athanasopoulos, 2021)

**TABTs:** O *Exponential Smoothing Method, Box-Cox Transformation, ARMA errors, Trend and Seasonal components* (TBATs) conforme proposto por De Livera et al. (2011), em que T significa termos trigonométricos para sazonalidade, B para transformações Box-Cox, A para Média Móvel Autoregressiva (ARMA) em relação aos erros de curto prazo, T para tendência e S para períodos sazonais.

O autor explica que as vantagens do TBATS, em relação a outros modelos sazonais, são que ele pode contabilizar vários dados sazonais dentro da mesma série temporal, capturam alta frequência, e é eficiente no manuseio de séries temporais não lineares.

**Prophet:** O método proposto por Taylor e Letham (2017) é apresentado como um modelo que funciona bem com parâmetros padrão, *outliers*, mudanças na sazonalidade e tendências (*changepoints*).

O método permite aos analistas selecionar os componentes que são relevantes para seu problema de previsão e facilita fazer ajustes ao usuário.

**MultiLayer Perceptron (MLP):** conforme ideia do perceptron de Rosenblatt (1958), com incremento da capacidade do algoritmo de retropropagação proposta por Rumelhart et al. (1986), o qual ajusta repetidamente os pesos das conexões na rede de modo a minimizar a diferença entre o vetor de saída real da rede e o vetor de saída desejado.

O MLP consiste em uma camada de entrada, uma camada de saída e uma ou mais camadas escondidas. Os neurônios em um MLP são unidades de processamento que recebem entradas ponderadas e aplicam uma função de ativação não linear para produzir uma saída. Essa não linearidade é essencial para seu aprendizado, e modele relações complexas em dados Goodfellow et al. (2016)

**Radial Basis Function (RBF):** desenvolvido por Broomhead e Lowe (1988), utiliza funções de base radial como função de ativação. As redes de função de base radial têm muitos usos, incluindo aproximação de função, previsão de séries temporais, classificação e controle de sistema devido às suas boas capacidades de aproximação, algoritmos de aprendizado mais rápido e estruturas de rede mais simples (BUHMANN, 2003).

O RBF contém a camada de entrada, saída e função não linear. A primeira camada é a camada de entrada e o número de nós é determinado pelo número de parâmetros de entrada. A segunda camada é a camada escondida e o número de nós é determinado pelo treinamento real. A terceira camada é a camada de saída e o número de nós é determinado de acordo com a saída. A relação entre as duas primeiras camadas é não linear e a relação entre as duas últimas camadas é linear (Yan, 2014).

**Autoregressive neural network (ARNN):** também conhecido como Rede Neural Autoregressiva (RNAR), corresponde a uma combinação entre o modelo de rede neural artificial (ANN) com o modelo autorregressivo (AR), sendo implementado um modelo de regressão não-linear para a série histórica, como explicado na literatura por Zhang et al. (1998), Khashei (2010) e Tealab et al. (2017).

***Extreme Learning Machine (ELM)***: método proposto conforme Huang (2006). Huang et al. (2014) comenta que ELM foi proposta para treinar redes neurais com uma única camada escondida *feedforward*, denominada por *single-hidden layer feedforward neurais networks* (SLFNs). Os nós da camada escondida são iniciados aleatoriamente e, em seguida, fixados. O único parâmetro que precisa ser aprendido são os pesos entre a camada escondida e a camada de saída. O autor cita que ELM têm três vertentes: menos invenção humana, alta precisão de aprendizagem e velocidade de aprendizagem rápida.

De acordo com Xia et al. (2019) comparado com algoritmos tradicionais, o método ELM é fácil de usar e alcança uma solução globalmente ótima com alta velocidade de aprendizado e boa capacidade de generalização. No entanto, parâmetros relevantes são fornecidos aleatoriamente, o que pode invalidar alguns nós da camada oculta e afetar a predição do modelo.