



DESENVOLVIMENTO DE MODELAGEM ORIENTADA A DADOS PARA ANÁLISE DE INCERTEZA E RISCO NA PREVISÃO DE PRODUÇÃO DE OLEO: SIMULAÇÃO DE SÉRIES TEMPORAIS EM PLATAFORMAS MARÍTIMAS

Vitor Hugo Pinheiro Marques (PUC-RIO) marques.hugo@petrobras.com.br
Fernando Luiz Cyrino (PUC-RIO) Cyrino@puc-rio.br
Antônio Orestes de Salvo Castro (UFRJ) Orestes@ime.uerj.br
Gabriel Alcântara Bomfim (Petrobras) Alcantara.ep@gmail.com

Resumo

O trabalho almeja prever a produção de óleo em plataformas marítimas, fazendo uso de técnicas de simulação e análise de séries temporais. O estudo contempla o desafiador cenário de amadurecimento da bacia geográfica de Campos de grande empresa nacional de energia. A elaboração da modelagem orientada a dados é apresentada, observando as transformações realizadas durante o processo. Compara-se a aplicação de métodos estatísticos (ETS e ARIMA) e simulação (Bootstrap e Monte Carlo com decomposição e aplicação no resíduo) com a metodologia atualmente utilizada que não faz uso de simulação do histórico, realizando simulação Monte Carlo nas projeções dos especialistas. É verificado que os modelos identificam as incertezas e a escolha do modelo pode variar com a necessidade de tempo e acurácia. Os resultados obtiveram menores erros (MAPE) do que a metodologia atual, principalmente a simulação *bagging* com previsão ARIMA.

Palavras-Chaves: Simulação. Séries Temporais. Produção de óleo.

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 da empresa. (ABATH, 2020).

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



Segundo Keith (2019) os modelos determinísticos ignoram a incerteza, e as distribuições de probabilidade são usadas para modelar a incerteza de modo estocástico.

Apesar do conceito de incerteza apresentar variações, deve-se avaliar os atributos de entrada de modo a caracterizá-las em um modelo, a fim de gerir o risco. Nesse estudo são avaliadas às incertezas geológicas de reservatório e operacionais de produção.

Suslick e Schiozer (2004) ressaltam que a maioria das organizações petrolíferas decidem usar procedimentos consistentes de análise de risco, buscando ferramentas para modelos de simulação mais complexos, estes irão estimular novos progressos à medida que melhores modelos e métodos tornam as ferramentas analíticas mais flexíveis e precisas.

A partir do progresso de ferramentas analíticas, essa pesquisa estabelece procedimento para gerenciar riscos associados à previsão de produção de óleo.

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.

A arbitrariedade dos reservatórios, aliada a alta complexidade de equipamentos de uma plataforma, provocam alta variabilidade para a previsão de produção, impactando em importantes decisões táticas e estratégicas, necessitando de uma análise de risco assertiva.

1.1. Contextualização

O estudo foi desenvolvido em empresa brasileira 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).

O Brasil atualmente tem maior foco na exploração de petróleo dos campos marítimos, no pré-sal e pós-sal. O trabalho analisa a região geográfica da Bacia de Campos, campo marítimo mais maduro brasileiro, com início da produção comercial em 1977, logo tem grande quantidade de dados históricos e pode servir como benchmarking para previsão das novas plataformas que devem enfrentar os mesmos desafios de declínio de potencial e ineficiência.



A Bacia Sedimentar de Campos ocupa uma área marinha de aproximadamente 100 mil quilômetros quadrados, localizada na região que vai do centro-norte do Estado do Rio de Janeiro até o sul do Estado do Espírito Santo.

2. Objetivo

O trabalho objetiva elaborar modelagem capaz de inferir as incertezas associadas aos dados históricos de potencial de produção e eficiência operacional, e de projeções em poços novos e paradas para manutenção, fazendo uso de séries temporais e simulação para aumentar a acurácia da distribuição de probabilidade que represente o risco em superar ou não as metas previstas relacionadas à produção de óleo em plataformas marítimas da Bacia de Campos.

3. Revisão bibliográfica

3.1. Análise de risco de previsão da meta de produção de óleo

Castro et al. (2010) propôs 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 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.

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. Como o presente estudo, objetiva mostrar como o processo de calibração permite reduzir a incerteza dos parâmetros críticos e aumentar a confiança na previsão de produção.

3.2. Simulação de séries temporais

Estudos de simulação com séries temporais buscam melhorar conjunto de dados desbalanceados, redimensionando amostras de treinamentos. (VANHOEYVELD, 2018).

Guerillot (2017) observa que modelos geoestatísticos requerem milhares de simulações de várias horas, sendo assim, ele apresenta abordagem que consiste em substituir o simulador por um modelo aproximado, calculando a produção de petróleo com uso de Simulação Monte Carlo para aplicar método de previsão de produção, em seu caso com redes neurais.

Outro trabalho que propõe combinação de simulação com séries temporais foi desenvolvido por Dantas & Oliveira (2018). Ele utiliza técnica de *bootstrap aggregating (bagging)* aplicada no resíduo da decomposição STL e transformação Box-Cox, para gerar séries sintética, e aplicar previsão, conforme proposto por Breiman (1996) e Bergmeir et al. (2016).

O presente estudo combina séries temporais com simulação Monte Carlo e Bootstrap a fim de balancear a distribuição dos dados, almejando obter previsões com menores erros.

4. Metodologia

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

Em relação à finalidade, consiste como uma pesquisa aplicada pois tem finalidade prática, motivada pela necessidade de minimizar o problema de previsão da meta de produção enfrentado por uma empresa de petróleo. Ainda quanto aos fins, caracteriza-se como pesquisa exploratória pois aborda temas na literatura para testar um modelo matemático desenvolvido como ferramenta de solução para o problema. Foram levantados trabalhos relacionados nas bases: Scopus, Science Direct, e Google Scholar, sobre os temas *Simulation, Forecasting, Oil & Gas*, e outros ligados às técnicas e métodos aplicados.

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 da meta prevista, ou que estão aplicando técnicas de simulação, séries temporais, principalmente em aplicações no ramo da produção *offshore*.

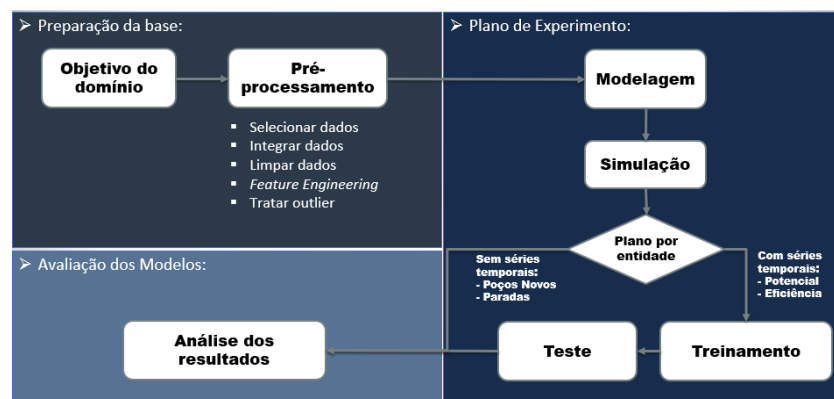
A aplicação é desenvolvida em software R, versão 4.1.2. Os pacotes utilizados foram: dplyr, lubridate, ImputeTS, tidyverse, tsibble, forecast, extraDistr, triangle, ggplot2 e plotly.

Os dados utilizados são extraídos da ferramenta Power Bi que é integrada ao banco de dados corporativo. As entidades de entrada podem ser divididas em quatro:

- Potencial implantado: capacidade de produção de óleo dos poços instalados;
- Eficiência: eventos com perdas relacionados às falhas dos equipamentos que impedem que a sua produção seja igual ao Potencial. O percentual de eficiência é uma medida de desempenho operacional, sendo o quociente da produção pelo potencial;
- Paradas: eventos de manutenção programadas. Corresponde a perda mais significativa e com maior previsibilidade de eficiência, sendo analisado a parte;
- Poços novos: poços em implantação a serem interligados em plataforma já em produção. A contribuição desses poços aumenta o potencial produtivo da plataforma.

A modelagem é estruturada baseada nas etapas da 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. O uso de um processo estruturado é fundamental para garantir o devido tratamento dos dados e extração do conhecimento. A figura 1 representa o framework da metodologia desenvolvida.

Figura 1 –Metodologia proposta para análise de risco da previsão de produção.



Fonte: Adaptado de Dantas (2020).

4.1 Objetivo do domínio

Analisar riscos da previsão da produção, conforme alinhado aos objetivos desse trabalho.

4.2 Pré-processamento

As atividades dessa etapa correspondem a: seleção dos dados; integração; limpeza; *feature engineering*; e tratar outlier.

Selecionar dados: Os conjuntos de dados podem ser divididos em: realizado histórico das 4 entidades de entrada desde 2017 até 2022 (60 meses); e planejamento dos especialistas para 2022 (12 meses) em relação as mesmas variáveis (dados que o modelo atual já utiliza).

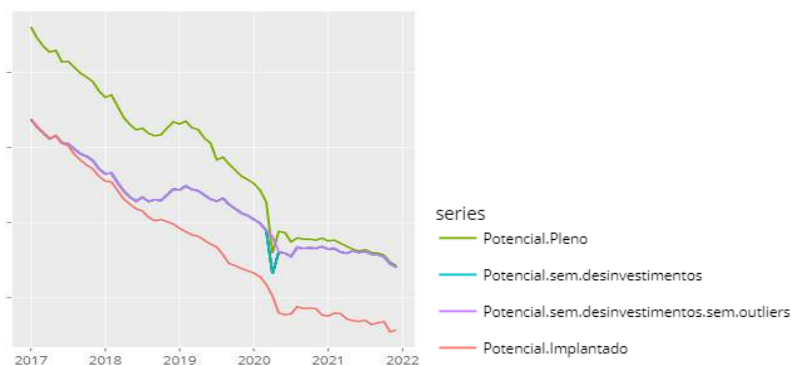
Os especialistas em planejamento da produção da empresa em estudo sinalizam investimentos significativos nos últimos 5 anos a fim de melhorar o desempenho operacional, não sendo indicado trabalhar com períodos anteriores. Os dados são de 24 plataformas, porém ao longo do trabalho foi decidido analisar as 5 de maior potencial produtivo, reduzindo o alto tempo de processamento para rodar as simulações.

Integrar dados: é necessário tratamento para padronizar nomenclatura das plataformas e poços a fim de unir os dados de diferentes fontes.

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

Feature Engineering: Para todas as dimensões ocorrem extrações de novas características. Opta-se em destacar o tratamento no potencial de produção com a divisão da variável (*splitting variable*) para detalhar a sua análise. É possível verificar diferentes níveis de potencial na figura 2. O eixo Y foi retirado para garantir sigilo das informações da empresa.

Figura 2 – Dados histórico do Potencial de produção da Bacia de campos.



Fonte: Elaboração própria em Software R.

O potencial pleno dos poços é mensurado pela área de reservatório, sendo representado pela linha verde. O potencial sem desinvestimentos é representado pela linha azul clara. A linha roxa é obtida após os ajustes do mês com outliers por COVID. A linha vermelha representa o potencial sem a entrada dos poços novos, apenas com a parcela já implantada, esse é o potencial simulado na entidade do potencial. Percebe-se alto desinvestimento até 2020, e maior investimento em poços novos a partir de 2019.

Tratar outlier: Quatro plataformas tiveram poços fechados em abril/20 devido a casos com Covid. O tratamento imediato foi indicar não contar com seus potenciais. 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). Nela, os fatores de ponderação diminuem exponencialmente para cada período do tempo. (LIN, 2020).

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

Foi definido usar média móvel dos últimos 4 meses em conjunto aos especialistas, os quais indicaram ter declínio relevante do potencial em períodos maiores.

4.3 Modelagem

O planejamento dos experimentos começa com a definição dos modelos a serem analisados.

Em relação aos poços novos e paradas programadas, optou-se por manter a simulação Monte Carlo na projeção dos especialistas. Para potencial e eficiência são avaliadas combinações de simulações com métodos de previsão.

Dois métodos de previsão estatísticos (ARIMA e ETS) são utilizados, sendo combinados com as seguintes possibilidades de simulação:

- Monte Carlo: aplicada nos resíduos após de decomposição STL, com uso da distribuição normal.
- *Bagging*: conforme proposta por Bergmeir et al. (2016), com aplicação da transformação Box-Cox e decomposição STL, e o resíduo sendo simulado através de “blocos móveis” (*moving block bootstrap*).
- Sem simular histórico: previsão automática do método de previsão.

As 3 possibilidades de simulação são combinadas com os 2 métodos de previsão, sendo avaliados 6 modelos em comparação ao modelo atual, totalizando 7 modelos em análise.

A metodologia atual realiza simulação Monte Carlo com distribuição 30%/40%/30% nos dados de projeção dos especialistas, sem simulação dos dados históricos.

4.4 Simulação

As projeções probabilísticas são obtidas para cada entidade de entrada do modelo: eficiência, potencial implantado, poços novos, e paradas. O cálculo da produção é aplicado para todas as iterações, e plataformas, seguindo a mesma equação do modelo atual:



Equação 1 – Equação de cálculo para obter produção a partir das 4 entidades de entrada do modelo.

$$\text{Produção} = ((\text{Potencial Poços Novos} + \text{Potencial Implantado}) - \text{Paradas}) * \text{Eficiência}$$

A simulação resulta na média de produção diária calculada.

4.5 Plano por entidade

4.5.1 Potencial Implantado e Eficiência Operacional

As entidades usam séries temporais, com aplicação de 2 métodos de previsão estatísticos bastante identificados na literatura: suavização exponencial e ARIMA.

- Suavização exponencial: Autores como Hyndman e Athanasopoulos (2021) desenvolveram métodos com pesos exponenciais. Observações recentes recebem um peso relativamente maior do que as mais antigas (MEIRA, 2021).

O método automático de previsão ETS usa trinta modelos e os distingue avaliando o erro (E), tendência (T) e sazonalidade (S);

- 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).

Para uma maior eficácia no tempo de ajustes dos parâmetros foi utilizada a função ets e autoarima da biblioteca *forecast* da linguagem R que automatiza a otimização dos parâmetros.

Treinamento: Após simular 100 amostras para os 60 meses do histórico, de cada plataforma, os dados são disponibilizados para cada método de previsão treinar as replicações.

Testes: As previsões são testadas para a soma das 5 plataformas de maior potencial em relação ao realizado de 2022, indicando os melhores modelos. A comparação das previsões é verificada através do erro percentual absoluto médio (MAPE), devido a facilidade de interpretar porcentagens, sendo mais comum em aplicações corporativas.

4.5.2 Poços Novos

Propõe-se usar as projeções dos especialistas, sem análise da base histórica, mantendo a lógica já em uso pela empresa. Como cada projeto de interligação de novo poço já realiza análise criteriosa do cronograma de entrada, justifica priorizar outras análises.

Desse modo, mantém-se o processo que faz uso de Simulação Monte Carlo com distribuição triangular, sendo parametrizada pelas estimativas das datas de entradas para cada poço, no caso: conservador considera a data com maior risco de atraso; data mais provável de início da produção; e otimista considera plano de ação com maior oportunidade de antecipação.

Os resultados estimam novas datas probabilísticas de entrada dos poços, causando deslocamento da contribuição de potencial atual. 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.

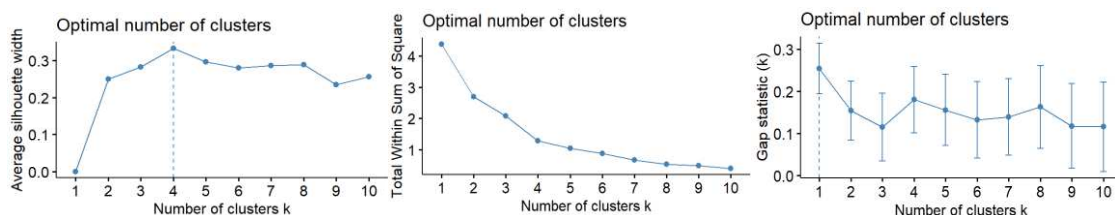
4.5.3 Paradas programadas para manutenção

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 conseqüente realização de paradas além do previsto. O histórico de 5 anos por plataforma é baixo para análises estatísticas, dado que a quantidade de paradas programadas anual é pequena.

Propõe-se agrupar os dados por plataformas similares através de clusterização. Para tanto, os especialistas indicaram analisar os dados de: volume de perda, duração média dos eventos, e idade atual da plataforma (alto correlacionamento com necessidades de manutenção).

A definição do número de clusters foi de 4 clusters, com base nos índices do pacote NbClust. O ótimo foi recomendado pela largura média da silhueta (ASW) e o WSS, conforme figura 3.

Figura 3 – Avaliação do número ótimo de clusters: Silhouette (esq.); Gap stat (centro); WSS (dir.)

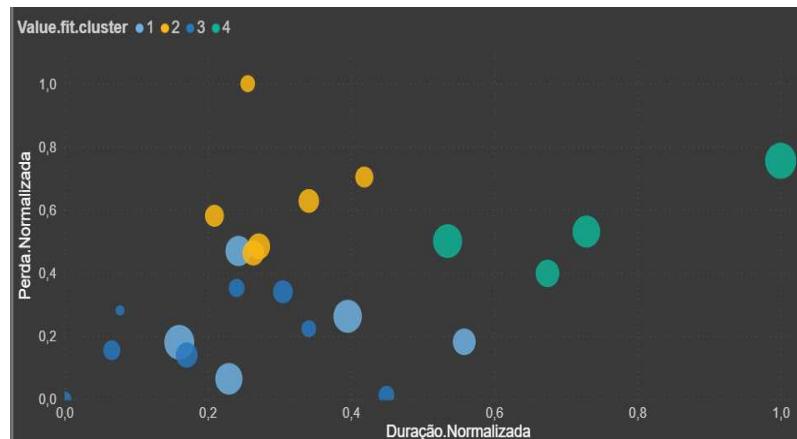


Fonte: Elaboração própria em software R.

A clusterização foi realizada com o algoritmo k-means. Entre as formulações de agrupamento baseadas na minimização de uma função objetivo formal, talvez a mais amplamente usada e estudada seja o agrupamento k-means. (KANUNGO, 2002).

O resultado é apresentado na figura 4, plotando a duração normalizada no eixo X, perda normalizada no eixo Y, e a idade como o tamanho da circunferência:

Figura 4 – Resultado da clusterização dos dados de paradas por plataforma



Fonte: Elaboração própria em software Power Bi.

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

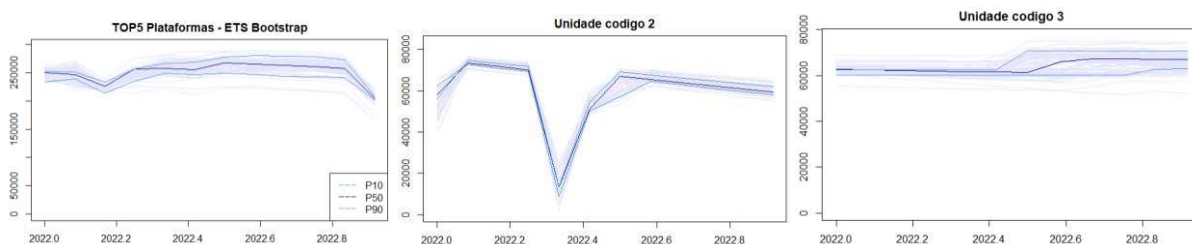
- Cluster 1: idade mais avançada, mas ainda não possuem duração alta de perdas.
- Cluster 2: alto potencial de produção, porém com idade e duração não relevantes, merece atenção devido à possibilidade de alto impacto.
- Cluster 3: menores durações médias, perdas, e plataformas mais jovens. Possui menor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

Os eventos semelhantes passam a ter histórico satisfatório, entre 26 e 40 por cluster. Para o simulador, os parâmetros estatísticos da triangular fazem uso do mínimo, média, e máximo por cluster, utilizando o delta entre o realizado e planejado da duração dos eventos históricos. 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. Análise dos resultados

Os resultados foram analisados para os sete modelos. A figura 5 mostram as 100 iterações projetadas com a contribuição P50 (mais provável), e o intervalo de confiança entre a P10 (otimista com 10% de chance de atingir) e P90 (conservadora com 90% de chance de atingir).

Figura 5 – Resultado da previsão de produção: Top 5 (esq.); Unidade 2 (centro); Unidade 3 (dir.)



Fonte: Elaboração própria em software R.

O gráfico à esquerda mostra, como exemplo, os resultados das simulações com *bagging* ETS para a soma das 5 plataformas. Verifica-se paradas no início e final do ano, e incremento do potencial no meio do ano devido a entrada de poço novo. O detalhamento por plataforma (gráfico central e direita) evidencia diferentes projeções, principalmente para: a parada da unidade 2, com base no histórico de seu cluster; e o início da produção do novo poço da unidade 3, alinhado aos riscos e oportunidades do projeto de interligação.

Os resultados da acurácia são visualizados na figura 6. A projeção atualmente utilizada pela empresa apresenta erro de 17.5%, e as demais simulações com análise do histórico permitiu melhorar o erro das previsões. Destaque para o *bagging* com ARIMA com menor erro.

Figura 6 - Resultados dos Modelos

Modelo	Produção MAPE (%)	Tempo Médio (minutos)	Potencial MAPE (%)	Eficiência MAPE (%)
ARIMA <i>bagging</i>	6.8	17	2.8	6.7
ARIMA (sem simular histórico)	10.2	4	4.3	6.7
ETS <i>bagging</i>	10.9	55	7.6	7.7
ETS (sem simular histórico)	11.4	4	7.6	7.6
ARIMA SMC	16.6	16	12.8	7.1
ETS SMC	17.3	56	10.5	7.4
SMC projeções (método atual)	17.5	6	3.0	18.2

Fonte: Elaboração própria em software Excel.

A análise das projeções evidencia que o potencial obteve menores erros do que a eficiência (prioridade para melhorias), e mostra maior otimismo nas projeções de eficiência do modelo atual em relação aos modelos que projetam com base no histórico.

Os resultados também permitem analisar as principais incertezas. Como exemplos: poço com alto potencial previsto entre abril e setembro, iniciou efetivamente em agosto; outro poço com previsão de postergação em até 6 meses, começou a produzir no quarto mês. Ambos evidenciam que as principais incertezas estão sendo realizadas dentro dos intervalos previstos.



6. Considerações Finais

As comparações dos resultados mostram possibilidades de melhorias da acurácia do modelo. 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.

Outro benefício identificado é simular antes de realizar a previsão, permitindo balancear os dados e melhorar a precisão das projeções. Por outro lado, também deve ser avaliado o aumento no tempo de processamento, variando a recomendação do modelo de acordo com a necessidade de tempo e acurácia.

A análise exploratória das simulações permite interpretar as incertezas relevantes por plataforma, identificando os principais eventos que podem provocar variações nas projeções, permitindo priorizar e traçar planos de ações que os mitiguem.

Recomenda-se para trabalhos futuros explorar a acurácia de diferentes métodos de séries temporais, modelos de simulação com menor tempo de processamento, e simulador para visualizar variações nos dados de entrada e análise de sensibilidade.

REFERÊNCIAS

Abath, Juliana Regueira. **Formação de Portfólio sob Incerteza de uma Empresa de Produção e Refino de Petróleo**. Dissertação (Mestrado). Rio de Janeiro: PUC-RIO, 2020.

ANP. **Brasil produz cerca de 3 milhões de bbl/d de petróleo em abril**. Boletim Mensal de Produção de 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>. Acesso em: 02 jan. 2023.

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

Bergmeir, Christoph.; Hyndman, Rob. J. and Benítez, José. M. **Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation**. *International Journal of Forecasting*, vol.32. P. 303–312. 2016.

Bomfim, Gabriel Alcântara. **Modelagem da Data de Entrada em Produção de Poços de Petróleo Utilizando Inferência Fuzzy**. Dissertação (Mestrado). PUC-Rio, Rio de Janeiro, RJ. 2017.

Box, George. E. P. and Jenkins, Gwilym. M. **Time series analysis: Forecasting and control**. San Francisco: Holden-Day. 1970.

Breiman, Leo. **Bagging predictors**. *Machine Learning* vol.24. P. 123–140. 1996.

Castro, Antônio Orestes de S.; Mattos, J. R. and 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.



Dantas, Thiago Mendes and Oliveira, Fernando Luiz Cyrino. **Improving time series forecasting: An approach combining bootstrap aggregation, clusters and exponential smoothing.** International Journal of Forecasting, vol.34. P. 748–761. 2018.

Dantas, Leila Figueiredo. **Predicting the Acquisition of Resistant Pathogens in ICUs using Machine Learning Techniques.** Tese (doutorado). Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2020.

Fayyad, Usama. et. Al. **Advances in Knowledge Discovery and Data Mining.** AAAI Press. 1996.

Guerillot, Dominique. and Bruyelle, Jeremie. **Uncertainty Assessment in Production Forecast with an Optimal Artificial Neural Network.** 2017.

Guimarães. André Menezes Oliveira. **Análise de incertezas em reservatórios de petróleo: uma abordagem estocástica usando métodos não intrusivos.** Dissertação de mestrado. Pós-graduação em Engenharia Civil. UFPE. 2016.

Haghighati, Raziieh. and Hassan, Adnan. **Recognition performance of imputed control chart patterns using exponentially weighted Moving average.** European J. Industrial Engineering, vol.12, No. 5, 2018.

Hyndman, Rob J., and Athanasopoulos, George. **Forecasting: principles and practice.** 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. 2021. Acessado em: 27/01/2023.

International Energy Agency. **World Energy Outlook 2021, IEA,** Paris, França. 2021. Disponível em: <<https://www.iea.org/reports/world-energy-outlook-2021>>. Acesso em: 20 jan. 2023.

Kanungo, Tapas. 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.

Keith, Andrew J., Ahner, Darryl K. **A survey of decision making and optimization under uncertainty.** Ann Oper Res 300, P. 319–353. 2021.

Lin, Suwen. **Filling Missing Values on Wearable-Sensory Time Series Data.** Proceedings of the 2020 SIAM International Conference on Data Mining (SDM). P. 46-54. 2020.

Meira, Erick; Oliveira, Fernando Luiz Cyrino and Menezes, Lilian M. **Point and interval forecasting of electricity supply via pruned ensembles.** Point and interval forecasting of electricity supply via pruned ensembles, Energy, vol.232. 2021.

Suslick, Saul Barisnik; Schiozer, Denis José. **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.

Vanhoeyveld, Jellis; Martens, David. **Imbalanced classification in sparse and large behaviour datasets.** Data Mining and Knowledge Discovery, v. 32, n. 1, p. 25–82, jan. 2018.

Vergara, Sylvia Constant. **Projetos e relatórios de pesquisa em administração.** 9 ed. São Paulo: Atlas, 2007.