

Comparing ARIMA and LSTM models to predict time series in the oil industry

Jaqueline B. Correia¹, Marcos Pivetta¹, Givanildo Santana do Nascimento^{1,2}, and Karin Becker¹

¹Universidade Federal do Rio Grande do Sul, Brazil

{jbcorreia, mvlpivetta, gsnascimento, karin.becker}@inf.ufrgs.br}

²Petrobras, Brazil

Abstract. Monitoring and forecasting oil and gas (O&G) production is essential to extend the life of a well and increase reservoirs' productivity. Popular models for O&G time series are ARIMA and LSTM recurrent networks, and typically several lags are forecasted at once. LSTM models can deploy the recursive prediction strategy, which uses one prediction to make the next, or the multiple outputs (MO) strategy, which predicts a sequence of values in a single shot. This work assesses ARIMA and LSTM models for the forecasting of petroleum production time series. We use time series of pressure and gas/oil flow from actual wells with distinct properties, for which we developed predictive models considering different time horizons. For the LSTM models, we deploy both the recursive and MO strategies. Our comparison revealed the superiority of LSTM models in general, and MO-based models for longer time intervals.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms**.

Keywords: Time series forecast, Oil industry, LSTM, ARIMA

1. INTRODUÇÃO

A indústria de exploração e produção de petróleo *upstream* tem feito grandes investimentos em novas soluções tecnológicas para enfrentar os desafios de produtividade e concorrência no setor. Em particular, tem-se notado 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 também possibilitam o aumento da sua produtividade.

Desenvolver modelos preditivos que resultem em previsões de séries temporais confiáveis é um dos grandes desafios atuais. A complexidade de prever séries temporais está ligada à dificuldade de representá-las, pois características como alta dimensionalidade, dependência entre as observações, e comportamentos (e.g. linearidade) devem ser consideradas. Modelos populares para esse fim, especificamente na indústria de petróleo, são o *Autoregressive Integrated Moving Average* (ARIMA) [Box 1970; Fan et al. 2021; Al-Fattah et al. 2005] e redes recorrentes profundas, em particular o LSTM (*Long-Short Term Memory*) [Fan et al. 2021; Liu et al. 2020; Sagheer and Kotb 2019; Hochreiter and Schmidhuber 1997]. Esses modelos preveem várias etapas dentro de um horizonte de tempo. Dentre as estratégias de previsão de várias etapas no modelo LSTM, a mais utilizada é a recursiva, que consiste no uso da previsão atual como entrada para a próxima previsão. Outra opção é a estratégia de múltiplas saídas (MS), que prevê várias etapas de uma única vez. Na literatura existem trabalhos que obtiveram resultados interessantes tanto com a estratégia recursiva [Sagheer and Kotb 2019; Liu et al. 2020; Fan et al. 2021] quanto com a de MS [Guariso et al. 2020; Garcia et al. 2018].

Pesquisa parcialmente apoiada por CAPES (Código de Financiamento 001) e Projeto PETWIN (financiamento FINEP e Consórcio de LIBRA).

Copyright©2021. Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

No presente trabalho temos por objetivo principal comparar o desempenho de modelos preditivos de múltiplas etapas ARIMA e LSTM, utilizando séries temporais reais de vazão de gás, óleo e pressão de poços com diferentes propriedades. Nos modelos LSTM são utilizadas as estratégias recursiva e de MS. Por meio desse trabalho buscamos responder às seguintes questões de pesquisa: (1) “Os modelos LSTM geram melhores resultados que os modelos estatísticos ARIMA?” (2) “Como é o desempenho dos dois tipos de modelo considerando diferentes horizontes de tempo para previsão?”.

Com base em medidas baseadas em diferenças de previsão, nossos resultados mostram que os modelos LSTM são superiores aos modelos ARIMA. Na média geral dos modelos LSTM com a estratégia de MS, os erros RMSE e MAE são, respectivamente, 35% e 24% menores em relação à média dos erros nos modelos ARIMA. Para os modelos LSTM recursivos, esta diferença é de 21% e 10% (RMSE e MAE, respectivamente). Para os modelos LSTM, observamos a superioridade da estratégia de MS quando comparada aos modelos LSTM recursivos, já que o seu erro médio RMSE é 11% menor, e erro médio MAE 12% menor. Embora a comparação de modelos ARIMA e LSTM tenha sido abordada em trabalhos relacionados [Fan et al. 2021; Siami-Namini et al. 2018], nosso trabalho contribui ao explorar dados de produção de petróleo não limitados à vazão; considerar horizontes de tempo curto, médio e longo; e por explorarmos nos modelos LSTM também a estratégia de MS.

Este trabalho está organizado como segue. Na Seção 2 apresentamos os trabalhos relacionados. Na Seção 3 são discutidos os dados, e o método utilizado para gerar os modelos e compará-los. A Seção 4 apresenta e discute os resultados. Conclusões e direcionamentos futuros são resumidos na Seção 5.

2. TRABALHOS RELACIONADOS

A previsão de produção é fundamental no gerenciamento de um campo comercial de petróleo, a fim de manter rentabilidade, maximizar a drenagem dos hidrocarbonetos do reservatório, cumprir compromissos regulatórios, e operar com eficiência, segurança e responsabilidade socioambiental. As propostas na literatura são centradas principalmente no modelo estatístico ARIMA e em redes neurais recorrentes LSTM.

Identificar o comportamento de uma série temporal permite descobrir se ela é estacionária ou não, se sua tendência é linear ou não, ou se existe autocorrelação entre as observações. O modelo ARIMA lida bem com dados que possuem linearidade positiva ou negativa, e foi utilizado para previsões na área de petróleo [Fan et al. 2021], hidrologia [Wang and Lou 2019] e financeira [Chen et al. 2018].

Na área de produção de petróleo, ainda que as séries tenham comportamento linear, é comum que os poços sejam fechados temporariamente por questões operacionais [Fan et al. 2021]. Nesse caso, há na série alguns intervalos com vazão igual a zero, que caracterizam eventos com comportamento não linear constante nesses intervalos. Trabalhos como [Garcia et al. 2018] costumam retirar os intervalos de fechamento para não afetar os modelos, tratando-os como ruído. Redes neurais recorrentes como o LSTM têm a capacidade de lidar com comportamentos diversos.

Modelos LSTM têm sido explorados para previsão de séries temporais de diversas áreas, como indústria de petróleo [Sagheer and Kotb 2019], energia elétrica [Masum et al. 2018] e saúde [Chimmula and Zhang 2020]. A estratégia recursiva é a mais utilizada para previsão de várias etapas, apresentando resultados promissores na indústria de petróleo, mas por usar cada previsão como entrada para próxima etapa, pode resultar no acúmulo de erros ao longo das iterações. A saída da previsão de várias etapas na estratégia de MS consiste em um vetor de valores futuros [Bontempi et al. 2012]. Além de não gerar acúmulo de erros de previsão, a estratégia de MS possui um tempo de processamento menor, já que prevê vários valores de uma só vez.

Encontramos poucos trabalhos baseados na estratégia de MS na área de petróleo. Em [Garcia et al. 2018], os autores utilizaram-na sobre dados de vazão de óleo, mas além de removerem as observações com valor zero, exploraram apenas horizontes curtos de previsão (12 etapas). Ignorar as observações de valor zero não reproduz o comportamento real de um poço de petróleo e leva à perda de informação, pois quando um poço é reaberto, os valores de pressão, temperatura e vazão se reestabelecem gradativamente. A estratégia de MS também foi aplicada na previsão do preço do

petróleo [Xiong et al. 2013], na qual foram utilizados diferentes tamanhos de horizontes de tempo (de 4 a 24 etapas).

[Fan et al. 2021] e [Siame-Namini et al. 2018] estabelecem comparações do modelo ARIMA e LSTM. Nosso trabalho se diferencia desses por envolver o contexto de produção de petróleo, considerar diferentes horizontes de tempo, e utilizar a estratégia de MS para a previsão no modelo LSTM.

3. MATERIAIS E MÉTODOS

O objetivo neste trabalho é comparar o desempenho preditivo do ARIMA com o LSTM nas estratégias de MS e recursiva, em séries temporais univariadas representando dados de produção de petróleo. O desempenho foi avaliado considerando diferentes condições: (a) horizontes de curto, médio e longo prazo (7, 125 e 764 etapas, respectivamente); (b) diferentes volumes de dados, e (c) comportamentos operacionais comuns da indústria de O&G, como o fechamento temporário de poços.

Os dados utilizados são provenientes de três poços de produção no mar (*offshore*), denominados P1, P2 e P3, os quais contêm séries temporais de vazão de gás, vazão de óleo e de pressão. Os dados do poço P1 referem-se a um poço produtor da região do Polo Pré-Sal da Bacia de Santos (PPBS), contendo 500 registros diários (01/01/2018 e 20/05/2019). Os dados dos poços P2 e P3 são provenientes do *dataset* público Volve¹ do Mar do Norte, contendo 3.056 registros diários (12/02/2008 e 17/09/2016).

Primeiramente analisamos os dados brutos das séries, buscando períodos de fechamento de poços, identificado em todas as séries, já que esses intervalos caracterizam eventos temporários com comportamento constante não linear. No caso específico da série de pressão do poço P2, o intervalo final da série estava totalmente zerado, e concluímos que se tratavam de erros de leitura do sensor, já que nas séries correspondentes de vazão as leituras estavam normais. Portanto, removemos esses dados zerados, reduzindo a série de pressão do poço P2 para um total de 961 registros.

Em seguida exploramos os dados através do gráfico de autocorrelação parcial para identificar a existência de autocorrelação na extensão de *lags*/defasagens. Com isso, conseguimos identificar que todas as séries do P3 e as séries de pressão do poço P1 e P2 possuem autocorrelação significativa em uma extensão maior de *lags* em relação às outras séries. Para mensurar estatisticamente o coeficiente de autocorrelação das séries no primeiro *lag*, utilizamos a função de autocorrelação de Pearson, que retorna valores entre -1 (negativa) e 1 (positiva). Comparativamente, constatamos que apesar das séries temporais de pressão dos poços P1 e P2 terem autocorrelação significativa em uma extensão maior de *lags*, no primeiro *lag* elas têm coeficiente de autocorrelação menor em relação às demais séries. Para identificar o comportamento dos dados, utilizamos o gráfico de dispersão baseado na autocorrelação de Pearson (Figura 1) que detectou que apesar dos pequenos eventos de fechamento de poços, todas as séries temporais possuem um comportamento linear positivo. Finalmente, identificamos que apenas as séries do poço P1 e a série de pressão do poço P2 são estacionárias.

Para a construção dos modelos, definimos para cada série os parâmetros/hiperparâmetros de forma experimental, visando minimizar o valor das métricas escolhidas. Para os modelos ARIMA utilizamos na definição dos parâmetros, a métrica *Akaike Information Criterion* (AIC), que é o padrão para medir quão bem o modelo se ajusta aos dados. Usamos uma função de pesquisa de grade (`auto_arima`), que após testar vários conjuntos de parâmetros indica o melhor conjunto de parâmetros minimizando o AIC. Para a definição dos hiperparâmetros dos modelos LSTM, foram realizados testes exploratórios variando esses hiperparâmetros visando minimizar o *Root Mean Squared Error* (RMSE). Essa métrica calcula a raiz do erro (diferença entre valor real e previsto) quadrático médio.

As divisões dos conjuntos de dados para treinamento e teste foram realizadas considerando os diferentes tamanhos de horizontes de tempo. Devido ao volume de dados, os horizontes de tempo curto e médio foram considerados para todas as séries, enquanto que o longo foi adotado somente para as séries do poço P3, e séries de vazão (óleo e gás) do poço P2. Utilizamos a divisão de 75% para

¹<https://www.equinor.com/en/what-we-do/digitalisation-in-our-dna/volve-field-data-village-download.html>

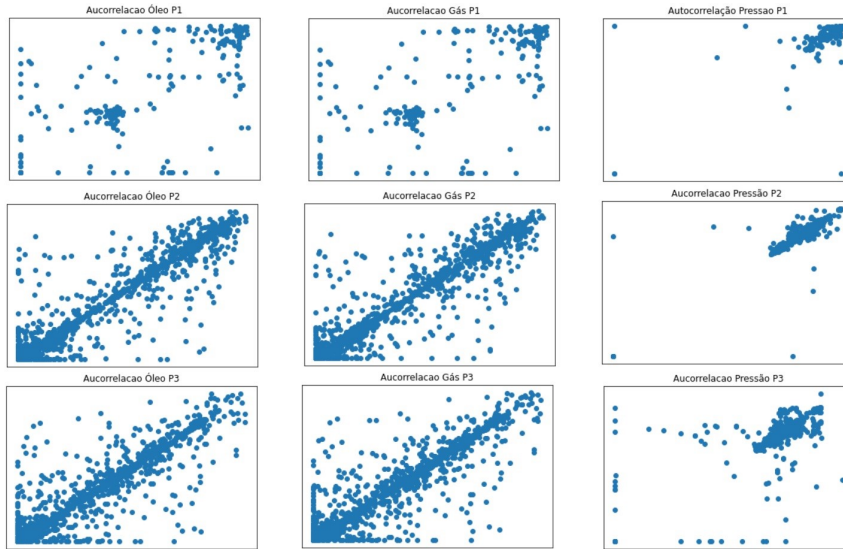


Fig. 1. Gráficos de autocorrelação de todas as séries temporais

treino e 25% para teste dos modelos, sendo que os modelos de curto prazo tiveram uma proporção menor de dados de teste para ser compatível com o horizonte de 7 dias escolhido.

Para comparar os modelos construídos com cada técnica utilizamos as métricas de avaliação de desempenho RMSE e *Mean Absolute Error* (MAE), onde os erros menores indicam melhor desempenho preditivo. Ambas métricas apresentam valor mínimo zero e valor máximo infinito. A MAE calcula a média dos erros absolutos (diferença entre real e previsto), ou seja, os erros contribuem linearmente, fazendo com que os ruídos afetem menos o resultado final. Já a RMSE eleva os erros ao quadrado, o que penaliza as previsões mais distantes do real. Portanto, é mais sensível aos ruídos nos dados.

Todos os modelos foram desenvolvidos utilizando a linguagem Python e as bibliotecas Pandas, Keras, Sklearn e TensorFlow. O código dos experimentos está disponível em um repositório público².

3.1 Modelo ARIMA

O ARIMA depende dos parâmetros (p,d,q) . O parâmetro p representa a autorregressão AR, que utiliza a autocorrelação entre as observações e o número de períodos ou (*defasagens*) associadas a uma observação precedente. O parâmetro d representa o integrado I, que utiliza o grau de diferenciação das observações para tornar uma série temporal estacionária. Já o parâmetro q representa a média móvel MA, que utiliza a dependência entre a observação e o erro residual a partir de um modelo de média móvel aplicado ao número de defasagens. Na Tabela I são apresentados os melhores valores encontrados pela pesquisa de grade para cada série, válidos para todos os horizontes.

Table I. Hiperparâmetros ARIMA

Conjunto de dados	P1			P2			P3		
	p	d	q	p	d	q	p	d	q
Vazão Óleo	1	1	2	1	1	3	5	1	2
Vazão gás	1	1	2	1	1	1	5	1	3
Pressão	6	0	3	5	1	3	4	1	1

Observando o parâmetro p , identificamos que a extensão da autocorrelação tem um *lag* maior no

²<https://github.com/JaquelineBitencourt/Comparing-ARIMA-and-LSTM-models-to-predict-time-series-in-the-oil-industry>

Table II. Hiperparâmetros LSTM

Conjunto de dados	Horizonte curto			Horizonte médio			Horizonte longo			
	Nº epc	Nº und	TL	Nº epc	Nº und	TL	Nº epc	Nº und	TL	
P1	Vazão Óleo	140	1/(5)	15/(1)	50	5	1			
	Vazão Gás	1000/(500)	1/(5)	15/(1)	2500	5	1			
	Pressão	1500	1	1	5000	1	1			
P2	Vazão Óleo	1	1	12	140/(15)	5	12/(1)	350	5	12
	Vazão Gás	1	1	12	1500	5	1	500	100	1
	Pressão	150	5	1	150	5	1			
P3	Vazão Óleo	1	1	12	50	5	12	200	5	12
	Vazão Gás	1	1	12	300	5	1	10000	5	12
	Pressão	1	1	12	400	1	12	100	5	1

poço P3 para todas as séries, e nas séries de pressão para os poços P2 e P3, confirmando o que havia sido identificado através da visualização do gráfico de autocorrelação parcial e validando que a pesquisa de grade realmente escolheu os melhores parâmetros.

3.2 Modelo LSTM

A arquitetura LSTM [Hochreiter and Schmidhuber 1997] é organizada em camadas de unidades LSTM. Cada unidade é uma célula composta de uma porta de entrada, uma porta de saída e uma porta de esquecimento. A célula lembra valores em intervalos de tempo arbitrários e as três portas regulam o fluxo de informações para dentro e para fora da célula. Entre as diversas variações dessa arquitetura, escolhemos a arquitetura empilhada, que possui uma ou mais camadas ocultas seguidas de uma única camada densa de saída para as previsões. O uso dessa arquitetura pode melhorar a habilidade do modelo em alguns problemas e domínios específicos [Graves et al. 2013]. Além do número de camadas e células, há muitos hiperparâmetros que influenciam o desempenho do modelo, bem como o número de épocas, tamanho do lote e o padrão dos dados submetidos para treinamento.

A topologia adotada para as previsões com o LSTM foi configurada com duas camadas ocultas seguidas de uma camada de saída densa, a função de otimização ADAM para minimização da RMSE, a função de ativação tanh, e taxa de abandono de 0.2. Para a implementação da estratégia recursiva a camada densa gera uma única saída, e uma função recursiva faz a sequência de previsões de acordo com o tamanho do intervalo temporal. Já no caso da estratégia de MS, a camada densa gera diretamente um vetor de valores correspondente ao tamanho do intervalo de previsão.

Testamos diferentes números de células dentro das camadas, assim como parâmetros de treinamento (épocas e tamanhos de lote). Na Tabela II são apresentados os melhores hiper-parâmetros para cada série temporal. Esses parâmetros foram utilizados em ambas as estratégias, sendo que, os valores entre parênteses são os parâmetros que foram diferentes para a estratégia recursiva.

Mediante a Tabela II constatamos que para as séries de vazão de gás e pressão do P1, vazão de gás do P2 no horizonte médio e vazão de gás do P3 no horizonte longo, o LSTM obteve melhores resultados com maior número de épocas. Isso sugere que essas séries possuem características intrínsecas que exigem do LSTM mais iterações no conjunto de treinamento para atingir um melhor desempenho. Além disso, observamos que o LSTM obteve melhores resultados de previsão com poucas unidades para cada camada oculta, com exceção da série de vazão de gás do P2 no horizonte longo. Outros testes não são reportados aqui, pois não influenciaram os resultados, tais como maior número de camadas e de células, taxas de abandono diferentes, etc.

É comum para a previsão de séries temporais no LSTM a aplicação de uma normalização na etapa de pré-processamento dos dados [Yunpeng et al. 2017], que consiste em mapear os dados para um valor máximo e mínimo para melhorar o desempenho. Entretanto, não foi identificada melhoria significativa, e por isso, foram submetidos os dados brutos para o treinamento.

4. RESULTADOS E DISCUSSÃO

Nas Tabelas III e IV são apresentados o desempenho dos modelos de acordo com as métricas de avaliação RMSE e MAE, que acumulam de distintas formas as diferenças entre valor real e previsto.

Table III. Resultados RMSE

Conjunto de dados	ARIMA			LSTM MS			LSTM RECURSIVO		
	H. curto	H. médio	H. longo	H. curto	H. médio	H. longo	H. curto	H. médio	H. longo
P1 Vazão óleo	23,77	16,674		22,932	15,271		13,144	16,517	
Vazão Gás	1.503,844	1.075,185		1.372,041	996,744		1.138,673	1.050,536	
Pressão	28,158	20,905		41,537	22,642		41,126	18,676	
P2 Vazão óleo	16,439	111,033	294,191	0,397	106,574	256,572	0,548	100,819	245,171
Vazão Gás	2.092,584	17.625,175	43.450,184	0,461	12.530,760	34.795,417	0,515	12.619,500	43.742,99
Pressão	12,737	17,148		19,316	35,874		18,089	31,582	
P3 Vazão óleo	4,77	74,626	221,773	0,430	55,126	106,36	0,784	56,381	105,235
Vazão Gás	706,641	11.953,199	32.754,973	0,492	9.144,336	23.720,604	0,337	9.183,528	23.837,47
Pressão	9,577	166,704	66,174	0,345	135,11	68,121	0,486	135,774	65,576

Table IV. Resultados MAE

Conjunto de dados	ARIMA			LSTM MS			LSTM RECURSIVO		
	H. curto	H. médio	H. longo	H. curto	H. médio	H. longo	H. curto	H. médio	H. longo
P1 Vazão óleo	16,839	15,224		17,946	13,787		11,586	10,699	
Vazão Gás	1.087,511	981,937		1.133,12	915,989		1.132,104	867,946	
Pressão	27,262	18,577		41,537	15,184		40,999	18,673	
P2 Vazão óleo	15,759	78,374	215,182	0,372	72,493	199,791	0,548	69,517	209,204
Vazão Gás	2.011,66	12.308,427	32.449,08	0,450	11.465,429	28.942,853	0,417	11.330,12	37.829,00
Pressão	10,674	13,083		17,665	33,855		16,672	29,31	
P3 Vazão óleo	4,54	55,996	198,443	0,412	52,182	86,514	0,781	52,717	89,042
Vazão Gás	672,076	8.967,002	29.101,114	0,476	7.947,395	20.146,819	0,264	7.948,748	20.270,45
Pressão	8,987	107,703	23,239	0,345	127,302	42,384	0,486	124,742	36,736

Table V. Diferenças de desempenho em percentuais

Modelos comparados	RMSE			MAE		
	H. curto	H. médio	H. longo	H. curto	H. médio	H. longo
LSTM MS X ARIMA	-202%	-35%	-30%	-218%	-9%	-25%
LSTM RECURSIVO X ARIMA	-262%	-34%	-13%	-220%	-10%	-6%
LSTM MS X LSTM RECURSIVO	17%	-1%	-15%	1%	1%	-18%

Interpreta-se sempre o primeiro tipo de modelo em relação ao segundo. Na primeira linha, por exemplo, observa-se um RMSE 202% menor considerando a média dos resultados do modelos LSTM de MS e ARIMA no horizonte de tempo curto.

Os melhores valores estão destacados em negrito, de acordo com cada série e horizonte de tempo.

A nossa primeira pergunta de pesquisa visa descobrir se os modelos LSTM obtêm melhores resultados preditivos que modelos ARIMA. Segundo a métrica RMSE (Tabela III) os modelos LSTM apresentam melhor desempenho para as séries de vazão (gás e óleo) dos três poços, considerando todos os horizontes de tempo. Para as séries de pressão, os modelos ARIMA apresentaram melhores resultados para o poço P1 no horizonte curto, e P2 nos horizontes curto e médio. Segundo a métrica MAE (Tabela IV) estes resultados se mantêm com pequenas diferenças, pois os modelos ARIMA também apresentam melhor desempenho para séries de pressão do poço P3 em horizontes médio e longo, bem como para a série de vazão de gás do poço P1 no horizonte curto. Os modelos LSTM são melhores nos demais casos. Essas diferenças de desempenho de um mesmo modelo ocorrem devido à forma de cálculo de cada métrica, que refletem a sensibilidade a ruídos. Enquanto na MAE os erros contribuem linearmente, a RMSE é mais afetada pelos ruídos, pois eleva os erros ao quadrado, penalizando as previsões mais distantes do real.

Sob a ótica de uma mesma métrica, o desempenho distinto de um mesmo tipo de modelo pode ser explicado pelas características de cada série. As séries de vazão de gás do P1 no horizonte curto e pressão do P3 no horizonte médio e longo são bem ruidosas. Como a métrica MAE atenua os pesos dos ruídos no resultado final e os modelos ARIMA apresentaram dificuldade em lidar com esses dados ruidosos, esses modelos obtiveram melhor desempenho nesses casos específicos apenas nessa métrica. Com base nessas mesmas constatações, ressaltamos que na série de pressão do P2 e na série de pressão do P1 no horizonte curto o ARIMA obteve melhores resultados em ambas as métricas, o que demonstra que essas séries possuem características estatísticas intrínsecas mais adequadas para o ARIMA do que para os modelos LSTM.

O desempenho distinto de cada tipo de modelo em função da série ainda são influenciados pelas autocorrelações e pelo volume dos dados. Dados temporais que possuem forte autocorrelação são mais fáceis de serem modelados e previstos, pois a autocorrelação representa o padrão dos dados, viabilizando que modelos preditivos extrapolem os dados de maneira mais confiável. A análise exploratória realizada nos dados e descrita na Seção 3 apontou que os resultados da autocorrelação de Pearson mostram que as séries temporais de pressão dos poços P1 e P2 possuem o menor coeficiente

de autocorrelação dentre as demais séries, obtendo valor de 0,65 e 0,85 respectivamente, considerada autocorrelação moderada, que somada a quantidade menor de dados apontam serem os motivos do desempenho preditivo inferior dos modelos LSTM nesses casos em específico.

Para consolidar os resultados, calculamos a média geral de erros de previsão de todos os modelos de cada estratégia preditiva. Constatamos que os modelos LSTM de MS obtiveram a menor média de erros, apresentando portanto melhor desempenho preditivo geral. Comparando ao ARIMA, o LSTM de MS apresenta um erro médio 35% menor na métrica RMSE, e 24% menor na MAE, respectivamente. O LSTM recursivo também foi superior ao ARIMA (RMSE médio 21% menor, e MAE médio 12% menor). Para os modelos LSTM, a estratégia de MS foi superior quando comparada à recursiva, já que o erro médio RMSE é 11% menor, e erro médio MAE 12% menor.

Considerando as 23 situações testadas (combinação de série temporal e horizonte) para cada tipo de modelo, identificamos que, considerando a métrica RMSE, os modelos LSTM de MS, LSTM recursivo e ARIMA obtiveram melhor desempenho em 11, 9 e 3 dos 23 casos, respectivamente. Fazendo a mesma análise para MAE, identificamos novamente a superioridade dos modelos LSTM de MS, que apresentaram melhor resultado em 11 dos 23 casos, seguido pelo LSTM recursivo (7) e ARIMA (6). Portanto, LSTM de MS obteve o melhor desempenho dentre as estratégias de predição avaliadas.

A segunda pergunta de pesquisa busca descobrir se o desempenho dos modelos é afetado quando utilizado tamanhos de horizontes diferentes, e novamente o LSTM demonstra melhor desempenho preditivo. A partir dos resultados nas Tabelas III e IV, é possível identificar que aumentando o tamanho do horizonte de tempo, o erro de previsão de todos os modelos também vai aumentando. Considerando a média dos resultados para cada horizonte de tempo, observamos que o desempenho médio dos modelos LSTM de MS degradaram do horizonte curto para o médio em 94% em ambas métricas, e do horizonte médio para o longo em 78% na RMSE e 77% na MAE. Considerando a estratégia LSTM recursiva, o percentual de degradação de desempenho do horizonte curto para o médio foi em média de 94% na RMSE e 95% na MAE, e do horizonte médio para o longo o aumento médio do erro foi 81% em ambas as métricas. A degradação do horizonte curto para o médio nos modelos ARIMA foi em média de 86% na RMSE e 83% na MAE, e do horizonte médio para o longo média de 78% na RMSE e 80% na MAE. Assim sendo, concluímos que a degradação do desempenho dos modelos quando comparados os horizontes médio e longo é semelhante para as três estratégias de predição, com percentuais muito semelhantes. Já a degradação dos horizontes curto e médio é menor nos modelos ARIMA, comparada à situação dos modelos LSTM.

Finalmente comparamos os três tipos de modelos considerando cada um dos horizontes de tempo. A Tabela V mostra o percentual de redução do erro médio de todos os modelos por horizonte de tempo para cada tipo de modelo. Para ambas as métricas, identificamos que os modelos LSTM de MS e recursivo obtiveram resultados melhores em relação ao ARIMA. Além disso, ao comparar os modelos LSTM com o ARIMA no horizonte curto fica evidente a significativa diferença em virtude do excelente desempenho dos mesmos no horizonte curto dos poços P2 e P3. Com base nessa mesma tabela, também constatamos que as diferenças da comparação dos dois modelos LSTM com o modelo ARIMA diminuem drasticamente nos horizontes médio e longo, o que evidencia que tanto a estratégia recursiva quando a de MS sofrem degradação do desempenho preditivo em horizontes com maior número de etapas no tempo.

Comparando agora os dois tipos de modelo LSTM, observamos que no geral os modelos baseados na estratégia de MS obtiveram melhor desempenho, com algumas exceções, apontadas nas Tabelas III e IV. Podemos observar uma propensão dos modelos LSTM de MS em lidar melhor com os horizontes médio e longo. A maior diferença é observada nos modelos de horizonte longo, onde constatamos que o LSTM de MS produz erros médios 15% e 18% menores que os modelos recursivos para as métricas RMSE e MAE, respectivamente. É possível que o acúmulo de erros da estratégia recursiva seja responsável por esse resultado.

Portanto, com base em todas as análises realizadas, concluímos que os modelos LSTM de MS produziram resultados superiores tanto em comparação com o ARIMA, quanto com o LSTM recursivo.

5. CONCLUSÃO

Este trabalho comparou modelos preditivos ARIMA e LSTM com séries temporais univariadas de produção de petróleo. Nos modelos LSTM, aplicamos tanto a estratégia recursiva, mais explorada na indústria de petróleo, quanto a de MS, menos explorada. As questões de pesquisa direcionaram os experimentos para descobrir se o modelo LSTM gera melhores resultados que o modelo ARIMA, e se o desempenho de ambos os modelos é afetado quando utilizados diferentes horizontes de tempo.

De acordo com as métricas RMSE e MAE, os resultados mostraram que ambos modelos LSTM obtiveram melhor desempenho geral de previsão quando comparados com modelo ARIMA. Os dois modelos LSTM obtiveram melhor desempenho em todos os horizontes em relação ao ARIMA, sendo o desempenho do horizonte curto o mais evidente nos poços P2 e P3. Além disso, constatamos que quando utilizado horizontes com mais etapas de tempo, o desempenho de todos os modelos degrada. Comparamos os dois modelos LSTM para identificar qual deles obteve o melhor desempenho. Baseado na média de erros, o LSTM de MS obteve RMSE 11% e MAE 12% menor que o LSTM recursivo, resultados que o elegem o melhor entre todos os modelos avaliados neste trabalho.

Como trabalhos futuros, pretendemos entre outros, expandir para a previsão de séries temporais multivariadas, bem como modelar a correlação da pressão na vazão de óleo e gás. Pretendemos também, modelar a correlação entre os poços que produzem óleo e gás (produtores) e os poços que são utilizados para estimular a produção (injetores). Modelar esse relevante cenário, pode permitir prever o impacto dos poços injetores na produção e melhorar o planejamento da produção.

REFERENCES

- AL-FATTAH, S. M. ET AL. Time series modeling for us natural gas forecasting. In *International Petroleum Technology Conference*. International Petroleum Technology Conference, 2005.
- BONTEMPI, G., TAIEB, S. B., AND LE BORGNE, Y.-A. Machine learning strategies for time series forecasting. In *European business intelligence summer school*. Springer, pp. 62–77, 2012.
- BOX, G. *Time series analysis; forecasting and control*. Holden-Day, San Francisco, 1970.
- CHEN, G.-Y., GAN, M., AND CHEN, G.-L. Generalized exponential autoregressive models for nonlinear time series: stationarity, estimation and applications. *Information Sciences* vol. 438, pp. 46–57, 2018.
- CHIMMULA, V. K. R. AND ZHANG, L. Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals* vol. 135, pp. 109864, 2020.
- FAN, D., SUN, H., YAO, J., ZHANG, K., YAN, X., AND SUN, Z. Well production forecasting based on arima-lstm model considering manual operations. *Energy* vol. 220, pp. 119708, 2021.
- GARCIA, J., LEVY, A., TUNG, A., YANG, R. M., AND KAYNIG, F. Applying deep learning to petroleum well data. *Harvard University, Harvard John A Paulson School of Engineering and Applied Sciences*, 2018.
- GRAVES, A., MOHAMED, A.-R., AND HINTON, G. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee, pp. 6645–6649, 2013.
- GUARISO, G., NUNNARI, G., AND SANGIORGIO, M. Multi-step solar irradiance forecasting and domain adaptation of deep neural networks. *Energies* 13 (15): 3987, 2020.
- HOCHREITER, S. AND SCHMIDHUBER, J. Long short-term memory. *Neural computation* 9 (8): 1735–1780, 1997.
- LIU, W., LIU, W. D., AND GU, J. Forecasting oil production using ensemble empirical model decomposition based long short-term memory neural network. *Journal of Petroleum Science and Engineering* vol. 189, pp. 107013, 2020.
- MASUM, S., LIU, Y., AND CHIVERTON, J. Multi-step time series forecasting of electric load using machine learning models. In *International conference on artificial intelligence and soft computing*. Springer, pp. 148–159, 2018.
- SAGHEER, A. AND KOTB, M. Time series forecasting of petroleum production using deep lstm recurrent networks. *Neurocomputing* vol. 323, pp. 203–213, 2019.
- SIAMI-NAMINI, S., TAVAKOLI, N., AND NAMIN, A. S. A comparison of arima and lstm in forecasting time series. In *2018 17th IEEE Int. Conference on Machine Learning and Applications (ICMLA)*. IEEE, pp. 1394–1401, 2018.
- WANG, Z. AND LOU, Y. Hydrological time series forecast model based on wavelet de-noising and arima-lstm. In *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, pp. 1697–1701, 2019.
- XIONG, T., BAO, Y., AND HU, Z. Beyond one-step-ahead forecasting: evaluation of alternative multi-step-ahead forecasting models for crude oil prices. *Energy Economics* vol. 40, pp. 405–415, 2013.
- YUNPENG, L., DI, H., JUNPENG, B., AND YONG, Q. Multi-step ahead time series forecasting for different data patterns based on lstm recurrent neural network. In *2017 14th web information systems and applications conference (WISA)*. IEEE, pp. 305–310, 2017.