

Inferência do teor de óleos e graxas em água produzida via convolutional LSTM network

Karina dos Reis Teixeira^{*,**} José Marques Oliveira Júnior^{*}
Ricardo Emanuel Vaz Vargas^{*} Patrick Marques Ciarelli^{**}
Celso J. Munaro^{**}

^{*} *Petróleo Brasileiro S.A., ES (e-mails: karinareisteixeira@gmail.com; josemarques.oliveira@petrobras.com.br; ricardo.vargas@petrobras.com.br)*

^{**} *Departamento de Engenharia Elétrica, Universidade Federal do Espírito Santo, ES (e-mails: patrick.ciarelli@ufes.br; celso.munaro@ufes.br)*

Abstract: Offshore oil and gas production units need to treat and dispose their produced water respecting the limits and reference measurement methods imposed by current laws. The main index to be controlled of this treatment process is the Total Oil and Grease (TOG) in the water to be discarded, which has its official value available for only about 20 days after the disposal is carried out. This work evaluates two neural network models with Long Short-Term Memory convolutions to estimate the TOG value from process variables, laboratory analysis and other data. Due to the dynamics of the process, data windows of 48 hours are used before the TOG value to be estimated. The results obtained indicate that the proposed models are as good as those presented in the literature, and better than simpler models to estimate the TOG value. These results corroborate the feasibility of using methods based on recurrent neural networks in the industry as a means to implement an *online* sensor capable of estimating the TOG and assisting in the decision-making of a platform operators regarding the continuity of water disposal to the sea.

Resumo: As unidades marítimas de produção de óleo e gás têm a necessidade de tratar e descartar água produzida respeitando os limites e métodos de referência de medição impostos pelas leis vigentes. O principal parâmetro a ser controlado desse processo de tratamento é o Teor de Óleo e Graxas (TOG) na água a ser descartada, o qual tem seu valor oficial disponibilizado por apenas cerca de 20 dias após a realização do descarte. Este trabalho avalia dois modelos de redes neurais com convoluções em módulos Long Short-Term Memory para estimar o valor do TOG a partir de variáveis do processo, análises laboratoriais e outros dados. Devido a dinâmica do processo, são utilizadas janelas de dados de 48 horas antes do valor do TOG a ser estimado. Os resultados obtidos indicam que os modelos propostos são tão bons quanto os apresentados na literatura, e melhores do que modelos mais simples para estimar o valor TOG. Tais resultados corroboram a viabilidade de utilização dos métodos baseados em redes neurais recorrentes na indústria como meio para implementar um sensor *online* capaz de estimar o TOG e auxiliar na tomada de decisão dos operadores de uma plataforma quanto a continuidade do descarte de água para o mar.

Keywords: Total oil and grease; Recurrent neural networks; ConvLSTM; Soft sensor; Fault detection; Process monitoring.

Palavras-chaves: Teor de óleo e graxas; Redes neurais recorrentes; ConvLSTM; Sensor virtual; Detecção de falhas; Monitoramento de processo.

1. INTRODUÇÃO

As unidades marítimas de produção e processamento de petróleo produzem água de seus reservatórios. Denominada como água produzida, esta é trazida à superfície juntamente com o óleo e o gás durante as atividades de produção desses fluidos (Motta et al., 2013). Essa água contém inúmeros contaminantes, tais como o próprio óleo, formações minerais dissolvidas, produtos químicos e gases dissolvidos (Fakhru'l-Razi et al., 2009).

Para tratar esta água, removendo o óleo disperso e dissolvido que está contido nela, tanto para descarte no mar ou reinjeção em poços, existem diversas tecnologias, dentre as quais pode-se citar o uso de ar dissolvido, que facilita a flotação do óleo e conseqüentemente separação gravitacional água/óleo (Bayati et al., 2011).

Para medir a qualidade da água tratada nas unidades *offshore*, um dos parâmetros de mensuração utilizado é o Teor de Óleos e Graxas (TOG). Este, por sua vez, é estimado ou medido por diversos métodos (Cirne et al.,

2016), realizados tanto a bordo quanto em laboratórios em terra.

A medida considerada oficial atualmente é obtida com o método gravimétrico SM 5520-B (também chamado neste trabalho de somente método gravimétrico), realizada apenas em laboratório em terra, uma vez que depende do uso de balança analítica de precisão. Esta, que deve ser controlada dentro dos limites vigentes da lei, se torna disponível para as Plataformas de Petróleo apenas após alguns dias a partir da coleta da amostra.

O documento que define a metodologia para medição do TOG gravimétrico SM 5520-B é a Resolução CONAMA Nº 393, de 8 de agosto de 2007 (Brasil, 2007). Esta resolução estabelece em seu artigo 5º que o descarte de água produzida deve obedecer a concentração média aritmética simples mensal de óleos e graxas (O&G) de até 29 mg/L, com valor máximo diário de 42 mg/L.

Neste trabalho, propõe-se aplicar modelos baseados em dados que possam auxiliar as unidades operacionais a monitorar o TOG gravimétrico dentro dos limites legais. Mais especificamente, foram avaliados dois modelos de redes neurais com camadas ConvLSTM utilizando dados de processo e análises laboratoriais de uma plataforma. As ConvLSTM são uma variante das LSTM *Long Short-Term Memory* com operações internas de convolução (Hassan e Mahmood, 2017). Isto pode permitir que tais camadas possam aprender dependências de longo prazo com resultados melhores do que uma simples LSTM. Os resultados obtidos neste trabalho foram comparados com os resultados obtidos pelos métodos empregados por Júnior et al. (2021), que utilizaram o mesmo conjunto de dados empregado aqui.

O restante deste artigo está organizado da seguinte forma: a Seção 2 discorre sobre a importância do monitoramento do TOG para unidades operacionais; a Seção 3 traz referências de trabalhos relacionados ao tema; a Seção 4 apresenta a metodologia utilizada durante a coleta, o tratamento e a análise dos dados, assim como os modelos propostos; a Seção 5 apresenta os resultados obtidos neste trabalho e algumas discussões; a Seção 6 traz as conclusões e trabalhos futuros sugeridos.

2. IMPORTÂNCIA DO MONITORAMENTO DO TOG PARA UNIDADES OPERACIONAIS

As unidades marítimas de produção e tratamento de petróleo e gás utilizam diversas configurações para tratar as águas produzidas por elas. Um sistema típico de tratamento de água produzida em instalações *offshore* apresenta tanques de separação água/óleo, hidrociclones e flutadores, conforme descrito por Júnior et al. (2021).

O controle de uma planta de tratamento de água é importante para garantir a continuidade da operação e produção de óleo e gás de uma plataforma de petróleo, além de manter a qualidade do produto e limites das variáveis dentro das faixas aceitáveis.

Os processos industriais são equipados com uma grande variedade de sensores que informam de forma *online* as grandezas medidas do processo. Contudo, existem variáveis do processo que devido à falta de sensores ou alto custo de medição em tempo real, são medidas por meio

de coleta manual e análise em laboratório de forma *offline* (Souza et al., 2016).

Para medir o Teor de Óleos e Graxas (TOG) em laboratório a bordo, é feita uma estimativa a cada certa quantidade de horas com o método de medição fotométrico e também pelo método de medição infravermelho. Também são utilizados analisadores em linha que estimam esse parâmetro de forma automatizada e com maior frequência (segundos ou poucos minutos) (Araújo Filho et al., 2020).

Por sua vez, a legislação vigente estabelece, conforme Resolução CONAMA Nº 393, de 8 de agosto de 2007, os seguintes critérios:

Art. 5º. O descarte de água produzida deverá obedecer à concentração média aritmética simples mensal de óleos e graxas de até 29 mg/L, com valor máximo diário de 42 mg/L.

Art. 6º. A concentração de óleos e graxas a que se refere o Art. 5º desta Resolução deverá ser determinada pelo método gravimétrico.

Art. 11. Os métodos de coleta e de análise são os especificados em normas técnicas cientificamente reconhecidas.

Para atender aos critérios descritos acima, conforme já relatado na Seção 1, as medições oficiais do TOG são obtidas em laboratório em terra e disponíveis apenas após alguns dias a partir da coleta da amostra. Desta forma, torna-se extremamente importante inferências confiáveis para viabilizar o controle e monitoramento do TOG durante a operação da plataforma, a partir de variáveis de processo e outras medições de TOG realizadas a bordo da unidade.

Destaca-se que há distinções entre as capacidades de detecção de componentes químicos de alguns métodos (SM5520-B e Fotométrico), conforme Tabela 1.

Tabela 1. Capacidades de detecção dos métodos de medição/estimativa de TOG analisados.

Óleo Total	Classe	SM5520-B	Fotométrico
Dissolvido	Aromáticos	Detecta	Não detecta
	Ácidos Fenóis		
Disperso	Aromáticos	Detecta	Detecta
	Ácidos Alifáticos		

Como pode ser verificado na Tabela 1, diferente do método SM5520-B, a medição de TOG utilizando o método Fotométrico não consegue detectar o óleo dissolvido de nenhuma das classes de compostos orgânicos relatadas (hidrocarbonetos aromáticos, ácidos e fenóis). Desta forma, observa-se que ocorrerão diferenças entre os valores obtidos entre esses métodos.

É importante esclarecer que o método de medição por infravermelho, já utilizado nas plataformas, consegue detectar ambos os tipos de óleos, dissolvido e disperso. No entanto, o problema torna-se ainda mais complexo quando se percebe que os resultados de TOG obtidos em laboratório a bordo não são semelhantes aos obtidos pelo método SM5520-B, uma vez que podem ocorrer oscilações no processo e a coleta das amostras pode não ser representativa, pois não ocorrem no exato momento.

É relevante citar que a influência do óleo dissolvido nos resultados do TOG gravimétrico é particular a cada unidade de produção de óleo e gás, pois é dependente das características do petróleo produzido. De acordo com as características geológicas do local de onde é extraído, o petróleo bruto pode variar quanto à sua composição química e ao seu aspecto (Delgado e Gauto, 2021).

Com a finalidade de analisar a correlação entre diversos métodos de medição de TOG entre diferentes amostras de petróleo, Queiroz et al. (2005) encontraram valores bem distintos entre aqueles avaliados. Isto corrobora a dificuldade de se encontrar modelos para viabilizar a resolução do problema.

Para corroborar ainda mais com a importância deste trabalho, destaca-se que o descarte de água produzida desenquadrada quanto ao teor de óleo e graxas pode introduzir impactos ambientais à fauna e flora. Desta forma, prever a medida de TOG obtida pelo método SM 5520-B antes que ocorra o descarte de água produzida pode evitar penalidades legais, prejuízos financeiros, danos à imagem de uma empresa e ao meio ambiente.

Não existe uma regra definida pelos órgãos reguladores, com especial atenção ao Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) e Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), para o cálculo das penalidades a serem aplicadas nos casos de desenquadramento dos valores medidos de TOG. No entanto, de acordo com o artigo 38 do Decreto nº 4136/2000, as sanções pecuniárias para um caso deste podem variar entre R\$1.000,00 e R\$30.000.000,00 (Brasil, 2002). Vale ressaltar que, em caso de reincidências, o IBAMA prevê a aplicação de fator multiplicador nas penalidades.

Um dos riscos também associados a recorrência dos desenquadramentos das medições de TOG é a interdição das Unidades de Produção de Óleo e Gás. É possível calcular o lucro cessante devido à interdição de uma unidade. De acordo com os dados de Única dos Petroleiros (2022), o custo médio de extração de petróleo e produção de derivados da Petrobras, em 2021, foi de R\$114,89 por barril. Ao passo que sua venda para o mercado interno foi de R\$416,40 por barril. Desta forma, o lucro cessante calculado foi de R\$301,51 por barril, ou seja, se uma plataforma produz 100.000 barris por dia, o lucro cessante dela em 2021 foi de R\$30.151.000,00 por dia em que sua produção estava parada.

3. TRABALHOS RELACIONADOS

Uma maneira de medir de forma *online* variáveis importantes para o processo, como o TOG, mas de difícil aquisição, pode ser por meio de um sensor virtual (*soft sensor*) construído a partir de modelos computacionais. Tal sensor pode inferir o valor de uma variável a partir de outras variáveis de processo que estão disponíveis em tempo real (Souza et al., 2016).

Sensores virtuais são usados em uma grande variedade de tarefas. Hadjiski et al. (2016) utilizaram esse tipo de sensor para fazer controle de queima dentro de uma planta de geração de energia. O mesmo foi desenvolvido a

partir de um modelo matemático analítico, redes neurais e modelos baseados em lógica *fuzzy*. Por sua vez, Yuan et al. (2020) propuseram uma rede LSTM para modelar um sensor virtual que prevê os pontos de ebulição iniciais de nafta pesada e querosene de aviação de um processo de hidrocrackeamento industrial.

Para monitorar a qualidade de produtos, cita-se o trabalho de Pani e Mohanta (2016), os quais desenvolveram um sensor virtual para supervisionar a qualidade do clínquer do cimento usando estatística multivariada e técnica de inferência *fuzzy Takagi-Sugeno*. Pani et al. (2016) também criaram um sensor virtual para monitorar a qualidade do produto de uma coluna debutanizadora utilizando análise de componentes principais e rede neural artificial *feed-forward*.

Oliveira Júnior e Pereira (2020) encontraram uma forma alternativa de estimar o TOG fotométrico por meio do uso de dados do processo e algoritmos de aprendizado de máquinas (*Machine Learning*), ao passo que Araujo Filho et al. (2020) foram capazes de criar modelos a partir de *Machine Learning* baseados em árvores de decisão ou redes neurais para estimar o TOG SM5520-B.

Júnior et al. (2021) também propuseram dois modelos baseados em redes neurais recorrentes para prever o valor do TOG em unidades de produção e tratamento de óleo e gás. O primeiro usou *Long Short-Term Memory* (LSTM) e o segundo adicionou uma camada convolucional (*Convolutional Neural Network*) antes da camada LSTM, sendo chamada de CNN+LSTM.

Hassan e Mahmood (2017) desenvolveram um modelo de rede neural chamado ConvLSTM. A ConvLSTM é uma variante de LSTM que contém uma operação de convolução dentro da célula LSTM. Ambos os modelos são um tipo especial de Rede Neural Recursiva (RNN), capaz de aprender dependências de longo prazo. Na ConvLSTM a multiplicação de matrizes é substituída pela operação de convolução em cada porta na célula LSTM. Ao fazer isso, ela captura recursos espaciais subjacentes por operações de convolução em dados multidimensionais, o que é capaz de conferir um maior poder de extração de características.

Diferente de uma ConvLSTM, em um modelo CNN+LSTM a entrada primeiro passa por camadas de convolução. O resultado desta convolução é um conjunto de características ao longo do tempo e esta é usada como entrada para a camada LSTM (Xavier, 2019). Na ConvLSTM este processo já ocorre internamente na camada LSTM.

O uso do modelo ConvLSTM tem-se mostrado cada vez mais frequente em aplicações industriais. Khan et al. (2021) utilizaram uma arquitetura híbrida baseada em ConvLSTM para previsão do estado de integridade de baterias e do consumo de energia. Por sua vez, Yu et al. (2020) propuseram uma rede neural profunda (DNN) *autoencoder* de memória de longo prazo convolucional (CLSTM-AE) para detecção de falhas de sinais de processo.

4. METODOLOGIA

Para este trabalho foi utilizado um conjunto de dados de processo proveniente de uma plataforma de produção

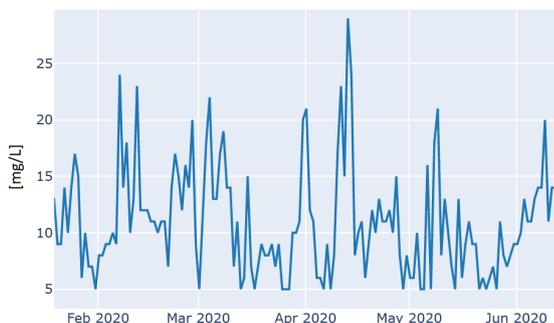


Figura 1. Gráfico com valores de TOG SM5520-B no tempo

offshore com operação entre janeiro de 2018 e fevereiro de 2021. Este conjunto de dados foi o mesmo utilizado por Júnior et al. (2021). A Figura 1 exibe valores de TOG gravimétrico em um certo período de tempo, variável alvo do problema. Observa-se que a tarefa de estimar o correto valor da variável é desafiadora, pois esta possui uma alta variância e uma grande quantidade de variações bruscas ao longo do tempo.

O conjunto de dados usado possui um total de 162 variáveis, incluindo análises de TOG gravimétrico e de TOG fotométrico, realizadas em laboratório em terra e pelo laboratório da plataforma, respectivamente. Também possui valores de TOG obtidos do analisador *online* e variáveis de processo que influenciam na qualidade da água tratada, obtidas de dados do sistema historiador (PIMS) da unidade de produção. Adicionalmente, traz variáveis obtidas por leitura manual de operadores, análises laboratoriais e outras que dependem de ajustes por meio de cálculos e simulações, como: vazão de produção de poços, vazão de produção líquida e bruta, BSW (*Basic Sediments and Water*) e volumes de injeção de produtos químicos.

Houve necessidade de realizar tratamento de dados previamente à aplicação dos modelos propostos. No conjunto de dados utilizado existem valores de TOG gravimétrico descritos no formato “<10mg/L”, por exemplo. Estes estão desta forma pois o TOG gravimétrico possui uma maior incerteza para baixas concentrações encontradas. Desta forma, substituiu-se valores que estão neste formato pelos respectivos valores inteiros, por exemplo: “<7mg/L” foi trocado para 7.

Também foram descartados os dados do conjunto de variáveis dos dias em que não ocorreu descarte de água produzida, ou seja, dos dias em que não houve coleta de amostra de TOG. Adicionalmente, não foram considerados os valores de TOG *online* nos dias em que o analisador entrou em falha.

É importante destacar que a variável inferida no treinamento do modelo proposto é o TOG gravimétrico e as demais variáveis citadas foram utilizadas como entradas. Para estimar a variável alvo são consideradas as variáveis de entrada em um período de 48 horas anterior à estimativa.

Foi utilizado um período de 48 horas anterior à estimativa pois a maioria das variáveis de processo não influenciaram no resultado da qualidade da água produzida após este período de tempo. Ressalta-se que cada variável de

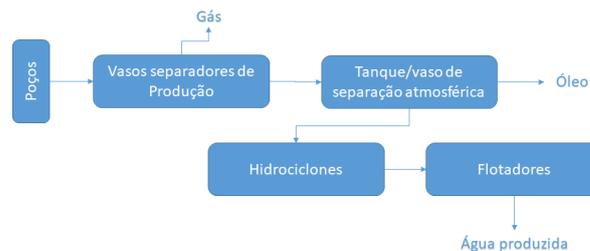


Figura 2. Configuração típica de uma planta de tratamento de água produzida

processo influencia de forma distinta no valor do TOG. Por exemplo, os produtos químicos, quando injetados após o tanque de separação gravitacional, têm uma influência muito rápida no resultado do teor de óleo e graxas da água tratada, podendo atuar em até 30 minutos. Por sua vez, a injeção de um produto antes desse tanque contribui para o resultado apenas após um certo período de tempo, o qual é dependente da vazão de água tratada, ou seja, do tempo de residência da água produzida nestes tanques de separação, que pode até ultrapassar 48 horas.

Essas diferentes contribuições das variáveis de processo são um desafio para os trabalhos nesta área. O uso de métodos mais eficazes, que possam considerar pesos e contribuições em tempos distintos de cada uma delas, torna-se um caminho promissor para obter melhores resultados.

As influências distintas das variáveis de processo no resultado da qualidade da água tratada também é dependente da configuração das plantas de processamento de cada unidade de produção de óleo e gás. Desta forma, o modelo proposto é aplicável a uma planta que contém um tanque de separação anterior à planta de tratamento de água, o qual favorece valores mais altos de tempo de residência e torna o controle do TOG mais complexo. A Figura 2 mostra uma configuração típica e simplificada de uma planta de tratamento de água produzida de uma plataforma, desde o poço até a obtenção do óleo e água produzida.

Para os modelos propostos, foram utilizadas variáveis de processo com amostragem a cada 5 minutos. Variáveis obtidas por boletins diários de produção, como volume injetado de produtos químicos e vazão de produção de óleo e água, tiveram seus valores referentes ao mesmo *data index* redimensionados para a mesma frequência de 5 minutos. Este procedimento também foi aplicado para os valores de TOG fotométrico, que são obtidos por análises laboratoriais a cada 2 horas.

A matriz de entrada dos modelos ConvLSTM é uma combinação de amostras no tempo na forma de um tensor 5D que contém: número de amostras, *frames*, canais, *time steps*, número de variáveis.

Neste trabalho, a matriz utilizada como entrada contém 162 variáveis divididas em 576 *time steps* de 5 minutos cada, totalizando uma janela temporal de 48 horas que se associa a um único valor de TOG gravimétrico (Figura 3). Ao todo, foram utilizadas 960 janelas temporais neste trabalho, associadas a 960 valores de TOG gravimétrico.



Figura 3. Exemplo de matriz de entrada

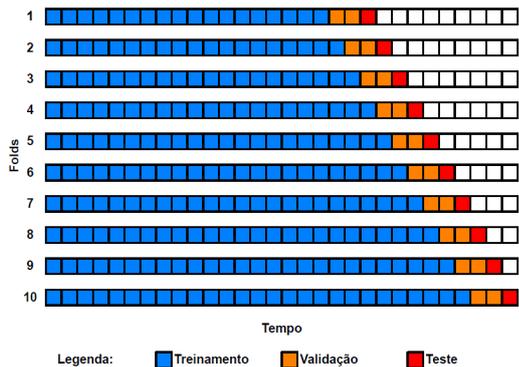


Figura 4. Separação dos dados entre treinamento, validação e teste para os 10 *fold*s. Fonte: Júnior et al. (2021)

A separação dos dados em treinamento, validação e teste foi realizada de acordo com a Figura 4, semelhante ao procedimento adotado em Júnior et al. (2021), para fins de comparação dos resultados obtidos por diferentes modelos. Cada bloco apresentado é equivalente a um conjunto com 32 valores de TOG Gravimétrico e 32 matrizes de entrada, iguais a representada na Figura 3.

O conjunto de validação é subsequente ao de treinamento, e o de teste subsequente ao de validação. Desta forma, o modelo não é treinado com dados futuros e nem testado com dados passados. Tal escolha visou não violar a ordem temporal da série, uma vez que existe dependência temporal da variável alvo com as demais variáveis de processo. Ressalta-se que o tamanho do conjunto de treinamento cresce para cada *fold* executado.

Foram utilizados 2 modelos com arquiteturas baseadas em ConvLSTM (Figura 5), sendo uma com complexidade maior (Arquitetura A) que a outra (Arquitetura B), de forma a verificar qual é a mais adequada para aplicar sobre o conjunto de dados utilizado.

A seguir, descreve-se as camadas utilizadas nas arquiteturas propostas:

- **ConvLSTM**: camada recursiva que contém operação de convolução dentro da célula LSTM, capaz de aprender dependências de longo prazo (Hassan e Mahmood, 2017);
- **Batch Normalization**: torna o treinamento das redes neurais mais estável a partir da normalização dos dados de entrada (Alzubaidi et al., 2021);
- **Max Polling**: são geralmente usadas após camadas convolucionais para simplificar as informações da saída, uma vez que substitui a saída da rede em um

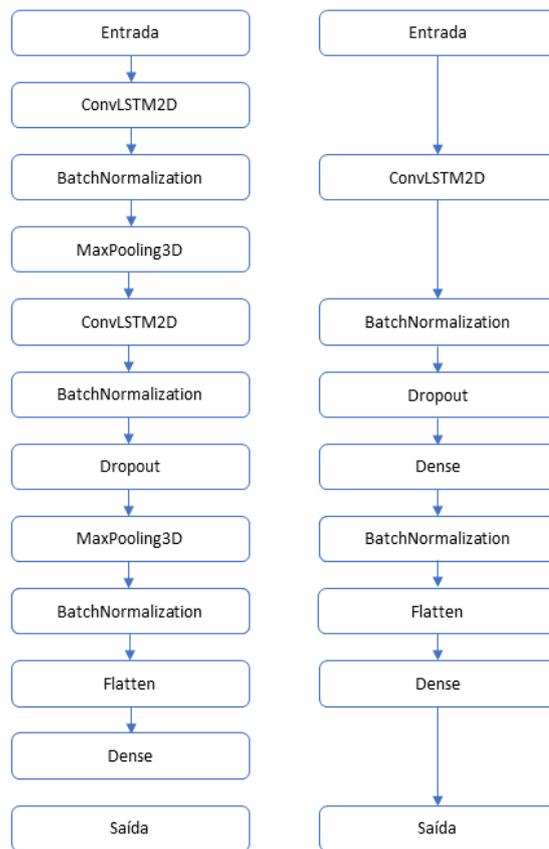


Figura 5. Arquiteturas propostas: Arquitetura A à esquerda e B à direita

determinado local, por uma estatística resumida das saídas próximas (Goodfellow et al., 2016);

- **Dropout**: utiliza um método que desconecta aleatoriamente um percentual de neurônios durante o treinamento, de forma a reduzir ocorrência de *overfitting* (Srivastava et al., 2014);
- **Flatten**: camada que transforma a sua entrada em um vetor, de forma que possa ser usada como entrada de uma camada totalmente conectada (*dense*);
- **Dense**: camada totalmente conectada por neurônios, similar ao de uma rede neural clássica, que consegue se ajustar a mapeamentos não lineares entre entrada e saída.

Para o treinamento, foi usada a função de ativação sigmóide nas camadas de ConvLSTM e a função de perda MSE (*Mean Squared Error*) com um número máximo de épocas de 2000 e *early stopping* para interromper o treinamento com paciência de validação igual a 200 épocas. Foram utilizados valores altos para evitar um término prematuro do treinamento.

Os valores de hiperparâmetros mostrados na Tabela 2 foram escolhidos pelo melhor resultado obtido para ambos os modelos testados, a partir de uma busca aleatória dentro da faixa de busca exposta. Estes valores de hiperparâmetros foram definidos buscando melhorar o resultado sobre o conjunto de validação do primeiro *fold*. Para evitar a definição de novos hiperparâmetros para cada *fold* de teste e, conseqüentemente, reduzir o esforço computacional

dos testes realizados, os melhores hiperparâmetros foram mantidos para os demais *folds*.

Tabela 2. Hiperparâmetros utilizados para criação dos modelos.

Hiperparâmetros	Modelo A	Modelo B	Faixa
<i>Dropout</i>	0,1	0,1	[0, 0,5]
<i>Kernel Size</i>	(5,5)	(3,3)	[(3,3), (5,5)]
Número de filtros	15;10	15	[5, 30]
<i>Batch Size</i>	8	8	[4, 32]

Foi utilizada a métrica RMSE (*Root Mean Square Error*) para avaliar o modelo, conforme Equação (1):

$$RMSE = \sqrt{\frac{\sum_{k=1}^n (x_k - \hat{x}_k)^2}{n}} \quad (1)$$

em que n é o número de matrizes de entrada (semelhante à da Figura 3) que estão sendo avaliadas, x_k é o valor do TOG gravimétrico associado à matriz k e \hat{x}_k é o valor estimado por um dos modelos. Quanto menor o valor dessa métrica, mais adequado o modelo gerado, o qual compara o valor de referência do TOG gravimétrico e o valor gerado pelo algoritmo em questão.

De forma a complementar a análise do trabalho, aplicou-se também o modelo estatístico *Autoregressive Integrated Moving Average* (ARIMA), utilizando o algoritmo *auto* ARIMA. Este possui um procedimento interno para auto ajustar os seus hiperparâmetros. Correia et al. (2021) também fazem uso desse modelo para comparar com arquiteturas que utilizam LSTM para prever séries temporais na indústria de óleo e gás.

5. RESULTADOS E DISCUSSÃO

As arquiteturas propostas A e B foram aplicadas com os hiperparâmetros mostrados na Tabela 2 para cada *fold* exibido na Figura 4 e foi obtida a métrica RMSE (Equação (1)) para cada teste realizado. A mediana dos resultados de RMSE dos dados de teste dos *folds*, obtida para cada modelo, é exposta na Tabela 3.

Tabela 3. Resultados obtidos.

Arquitetura	<i>Kernel size</i>	Filtros	<i>Batch size</i>	RMSE
A	(5,5)	15; 10	8	5,7
B	(3,3)	15	8	5,2

Os *boxplots* dos valores de RMSE calculados para os 10 *folds* de teste para os dois modelos propostos são mostrados na Figura 6. Desta forma, é possível comparar a dispersão dos erros e das suas medianas. Pode ser observado que a Arquitetura B apresentou uma menor mediana e uma menor dispersão que da Arquitetura A. Esses achados corroboram para afirmar que o modelo mais simples se mostrou mais adequado à proposta deste trabalho.

Durante os testes observou-se *overfitting* no modelo mais complexo, provavelmente devido a quantidade limitada de amostras da base de dados, o que pode justificar o fato do modelo mais simples alcançar melhores resultados neste trabalho.

Júnior et al. (2021) aplicaram 2 modelos distintos baseados em CNN+LSTM e LSTM para estimar o valor do TOG,

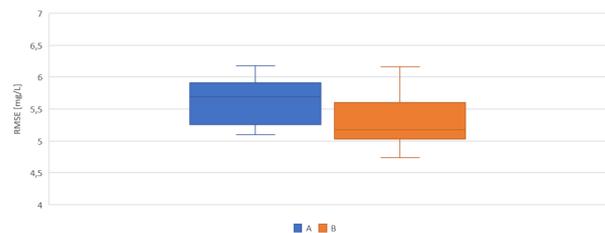


Figura 6. Resultados obtidos para arquiteturas A e B

usando o mesmo conjunto e separação de dados deste trabalho. Os erros RMSE relatados por Júnior et al. (2021) foram comparados com um preditor *baseline* ingênuo, o qual forneceu uma previsão da média dos valores de TOG gravimétrico dos dados de treinamento. Também foram comparados com outros dois preditores que utilizaram o valor médio do TOG fotométrico e do valor médio do TOG infravermelho no dia. A Tabela 4 apresenta os valores de mediana, média e desvio padrão do RMSE obtidos por este trabalho e pelo trabalho de Júnior et al. (2021). Em negrito estão os resultados obtidos neste trabalho.

Tabela 4. Resultados obtidos neste trabalho e por Júnior et al. (2021)

Modelo	Mediana [mg/L]	Média [mg/L]
ConvLSTM B	5,2	5,3 ± 0,44
LSTM	5,3	5,4 ± 1,27
ARIMA	5,4	5,3 ± 1,10
CNN + LSTM	5,5	5,2 ± 1,13
ConvLSTM A	5,7	5,8 ± 0,75
Fotométrico	7,5	7,9 ± 1,86
<i>Baseline</i>	8,3	8,1 ± 0,91
Infravermelho	12,6	12,8 ± 3,85

Observa-se que os valores de média e de mediana do RMSE encontrados para ambas as arquiteturas propostas neste trabalho, em especial a Arquitetura B (ligeiramente inferior), são semelhantes aos encontrados pelos modelos CNN+LSTM, LSTM e ARIMA. Um fato relevante é que ambas as arquiteturas que utilizaram ConvLSTM apresentaram menores variâncias. Assim como também é possível afirmar que ambos os modelos usados neste trabalho apresentaram erros inferiores aos modelos preditor *baseline* ingênuo, TOG fotométrico e TOG infravermelho, o que também ratifica a viabilidade de se utilizar ConvLSTM para o fim proposto.

Embora o modelo ARIMA apresente valores bem semelhantes aos modelos propostos, ele apresenta maior variância e ao longo dos testes pôde ser observado uma piora quando se utilizou quantidades de dados de treino e teste bem maiores que os utilizados neste trabalho, o que não se observou para as arquiteturas que utilizaram ConvLSTM.

É importante citar que este trabalho não fez uso de engenharia de atributos, o que foi feito em Júnior et al. (2021), e mesmo assim alcançou resultados no mesmo patamar. O não uso da engenharia de atributos facilita a replicação do método para outras plataformas, uma vez que traz uma maior independência das especificidades de cada planta, o que é bastante desejável para aplicação em maior escala industrial.



Figura 7. Resultado do modelo

A Figura 7 mostra um gráfico comparando os valores do TOG gravimétrico verdadeiro (TOG medido) e a saída gerada (TOG estimado) pelo modelo ConvLSTM B dos dados de teste do primeiro *fold*.

Para o uso em aplicação como sensor *online* sugere-se considerar a soma do valor do TOG inferido pelo método (\hat{x}_k) ao desvio padrão ($\hat{\sigma}$) estimado entre o valor real e o valor estimado, ou seja, $\hat{x}_k + \hat{\sigma}$. Embora esta abordagem não garanta estimativas mais precisas, ela aumenta as chances de detectar situações em que o TOG seja superior ao definido pela Resolução.

6. CONCLUSÃO

Os modelos criados com base em redes neurais recorrentes utilizando ConvLSTM se mostraram aplicáveis para estimar, a partir de dados de variáveis de processo, o TOG da água tratada e descartada de uma plataforma de produção de óleo e gás.

Os resultados obtidos mostraram que a aplicação do modelo sugerido é promissora para estimar o valor do TOG e assim auxiliar os operadores de uma plataforma de produção no controle da qualidade da água tratada. Isto é muito importante, visto que os valores de referência do TOG só são conhecidos após um longo período de tempo pelo método de referência aceito pelo IBAMA, e que valores acima do previsto em Resolução podem resultar em penalidades legais, danos ao meio ambiente e prejuízos a imagem da empresa. O uso dos modelos que utilizam camadas de ConvLSTM contribuem para corroborar o fato de serem também aplicáveis a indústria de processos e que podem ser ainda mais explorados para seu uso como alternativa de criação de sensores *online*.

É importante a continuidade de pesquisas para o desenvolvimento de modelos com erros médios ainda menores, de forma que seja possível validar a utilização de métodos baseados em redes neurais recorrentes como ferramenta oficial de inferência da qualidade da água tratada das plataformas de produção, tornando o processo de identificação de desvios no descarte ainda mais eficiente. Possíveis trabalhos futuros incluem a utilização combinada de outros modelos, uso de técnicas de aumento de dados para conseguir treinar os modelos de forma mais adequada, análise de diferentes períodos de tempo para estimação do TOG e também utilização de métodos mais eficientes para otimizar hiperparâmetros para cada *fold*.

AGRADECIMENTOS

Agradecimentos à Petróleo Brasileiro S.A. (Petrobras) pelo encorajamento e fornecimento dos dados necessários

para a realização deste trabalho, bem como ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Espírito Santo pelo incentivo e compartilhamento do conhecimento.

REFERÊNCIAS

- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., e Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8(1), 1–74.
- Araújo Filho, C.F., Vargas, R.E.V., Bucker, E.B., e Delaiba, V.H.B. (2020). Monitoramento de teor de óleos e graxas em água descartada no mar usando ciência de dados. *Rio Oil & Gas Conference*.
- Bayati, F., Shayegan, J., e Noorjahan, A. (2011). Treatment of oilfield produced water by dissolved air precipitation/solvent sublation. *Journal of Petroleum Science and Engineering*, 80(1), 26–31.
- Brasil (2002). Decreto nº 4.136, de 20 de fevereiro de 2002. *Diário Oficial [da] República Federativa do Brasil*. URL https://www.planalto.gov.br/ccivil_03/decreto/2002/d4136.htm.
- Brasil (2007). Resolução conama nº 393/2007. *Diário Oficial da União*, (153), 72–73. URL <http://www2.mma.gov.br/port/conama/legiabre.cfm?codlegi=541>.
- Cirne, I., Boaventura, J., Guedes, Y., e Lucas, E. (2016). Methods for determination of oil and grease contents in wastewater from the petroleum industry. *Chemistry and Chemical Technology*, 10(4), 437–444.
- Correia, J.B., Pivetta, M., do Nascimento, G.S., e Becker, K. (2021). Comparing arima and lstm models to predict time series in the oil industry. In *Anais do IX Symposium on Knowledge Discovery, Mining and Learning*, 129–136. SBC.
- Delgado, F. e Gauto, M.A. (2021). Petróleo: qualidade físico-químicas, preços e mercados: o caso das correntes nacionais.
- Fakhru'l-Razi, A., Pendashteh, A., Abdullah, L.C., Biak, D.R.A., Madaeni, S.S., e Abidin, Z.Z. (2009). Review of technologies for oil and gas produced water treatment. *Journal of hazardous materials*, 170(2-3), 530–551.
- Goodfellow, I., Bengio, Y., e Courville, A. (2016). *Deep learning*. MIT press.
- Hadjiski, M., Doukovska, L., Nikolov, M., e Koynov, S. (2016). Intelligent multi-soft sensing for flame position of steam boilers. *Cybernetics and Information Technologies*, 16(1), 158–172.
- Hassan, A. e Mahmood, A. (2017). Deep learning approach for sentiment analysis of short texts. In *2017 3rd international conference on control, automation and robotics (ICCAR)*, 705–710. IEEE.
- Júnior, J.M.O., dos Reis Teixeira, K., Vargas, R.E.V., Ciarelli, P.M., e Munaro, C.J. (2021). Estimação do teor de óleos e graxas em água descartada no mar usando redes neurais recorrentes. In *Simpósio Brasileiro de Automação Inteligente-SBAI*, volume 1.
- Khan, N., Haq, I.U., Ullah, F.U.M., Khan, S.U., e Lee, M.Y. (2021). Cl-net: ConvLstm-based hybrid architecture for batteries' state of health and power consumption forecasting. *Mathematics*, 9(24), 3326.

- Motta, A.R.P.d., Borges, C.P., Kiperstok, A., Esquerre, K.P., Araujo, P.M., e Branco, L.d.P.N. (2013). Tratamento de água produzida de petróleo para remoção de óleo por processos de separação por membranas: revisão. *Engenharia Sanitária e Ambiental*, 18, 15–26.
- Oliveira Júnior, J.M. e Pereira, M.d.A. (2020). Forecasting Total Oil and Grease in produced water using Machine Learning methods in an oil extraction plant. *Marine Systems & Ocean Technology*, 15, 124–134.
- Pani, A.K., Amin, K.G., e Mohanta, H.K. (2016). Soft sensing of product quality in the debutanizer column with principal component analysis and feed-forward artificial neural network. *Alexandria Engineering Journal*, 55(2), 1667–1674.
- Pani, A.K. e Mohanta, H.K. (2016). Online monitoring of cement clinker quality using multivariate statistics and takagi-sugeno fuzzy-inference technique. *Control Engineering Practice*, 57, 1–17.
- Queiroz, Y.G.C., Clarisse, M.D., Oliveira, B.D., Reis, B.D., Travalloni, A.M.L., e Lucas, E.F. (2005). Determinação e correlação do teor total de óleos e graxas por meio de diferentes técnicas espectroscópicas e gravimétrica. In *3º Congresso Brasileiro de P&D em Petróleo e Gás*, 6.
- Souza, F.A., Araújo, R., e Mendes, J. (2016). Review of soft sensor methods for regression applications. *Chemometrics and Intelligent Laboratory Systems*, 152, 69–79.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., e Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929–1958.
- Xavier, A. (2019). An introduction to convlstm. URL <https://medium.com/neuronio/an-introduction-to-convlstm-55c9025563a7>.
- Yu, J., Liu, X., e Ye, L. (2020). Convolutional long short-term memory autoencoder-based feature learning for fault detection in industrial processes. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–15.
- Yuan, X., Li, L., Shardt, Y.A., Wang, Y., e Yang, C. (2020). Deep learning with spatiotemporal attention-based lstm for industrial soft sensor model development. *IEEE Transactions on Industrial Electronics*, 68(5), 4404–4414.
- Única dos Petroleiros, F. (2022). Petrobras vende derivados de petróleo 3 vezes e meia a mais do que custo de produção. URL <https://www.cut.org.br/noticias>.