

12º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS



TÍTULO DO TRABALHO:

INFLUÊNCIA DA QUALIDADE DE DADOS DE ENTRADA NO DESEMPENHO DE MODELOS DIGITAIS DE DEGRADAÇÃO DE EQUIPAMENTOS SUBMARINOS NA INDÚSTRIA DE ÓLEO E GÁS

AUTORES:

Luiz Eduardo de Farias, Miguel Burg Demay, Pedro Henrique Zomer Machado, Suélen Seledes Rodrigues

INSTITUIÇÃO:

Centro de Metrologia e Instrumentação, Fundação Certi.

INFLUÊNCIA DA QUALIDADE DE DADOS DE ENTRADA NO DESEMPENHO DE MODELOS DIGITAIS DE DEGRADAÇÃO DE EQUIPAMENTOS SUBMARINOS NA INDÚSTRIA DE ÓLEO E GÁS

Resumo

Empresas do setor de Óleo & Gás tradicionalmente direcionam seu desenvolvimento para a maximização da eficiência econômica da produção, impulsionadas pela busca por maior lucratividade. Reforçando esta tendência, recentemente, a indústria vem se alterando em uma direção pautada pela transformação digital de procedimentos e processos. É fato que o mundo virtual está cada vez mais presente, com simulações e gêmeos digitais baseados em modelos digitais representativos de fenômenos físicos. Como entrada, estes modelos recebem traduções de variáveis como vazão, temperatura e pressão por meio de sensores e sistemas de medição. Entretanto, dados originados de sensores podem possuir falhas, tais como, dados faltantes e outliers, os quais têm o potencial de impactar significativamente os resultados finais dos cálculos do mecanismo de degradação. Se os dados de sensores não se apresentarem confiáveis, os modelos estatísticos, empíricos ou de inteligência artificial desenvolvidos, assim como os resultados de simulações, avaliações de estado, predições e descrições, também não o serão. Este artigo investiga a importância da avaliação da qualidade dos dados para aprimorar a confiabilidade dos resultados de um modelo digital de degradação, destacando os desafios enfrentados na medição e na garantia dessa qualidade. Através do método desenvolvido neste trabalho, demonstramos como a qualidade dos dados impacta diretamente na exatidão das saídas de modelos digitais, com enfoque no modelo de erosão DNV-RP-O501, evidenciando os efeitos de dados de baixa qualidade na confiabilidade do gêmeo digital. Essa análise reforça a importância da avaliação da qualidade dos dados para garantir a confiabilidade dos resultados de degradação do ativo. Por fim, sugerimos que os métodos de avaliação de qualidade deste estudo possam servir como ferramentas de apoio à tomada de decisões em nível descritivo, prescritivo e preditivo.

Palavras-chave: qualidade de dados, modelos digitais, gêmeo digital

Abstract

Oil & Gas companies have traditionally directed their development towards maximizing production efficiency, driven by the pursuit of greater profitability. Reinforcing this trend, the industry has recently been shifting towards the digital transformation of procedures and processes. It is evident that the virtual world is becoming increasingly present, with simulations and digital twins based on digital models representative of physical phenomena. These models receive input from variables such as flow rate, temperature, and pressure through sensors and measurement systems. However, data originating from sensors can have failures, such as missing data and outliers, which have the potential to significantly impact the final results of degradation mechanism calculations. If sensor data is not reliable, the developed statistical, empirical, or artificial intelligence models, as well as the results of simulations, condition assessments, predictions, and descriptions, will also be unreliable. This article investigates the importance of data quality assessment to enhance the reliability of the results of a digital degradation model, highlighting the challenges faced in measuring and ensuring this quality.

Through the method developed in this work, we demonstrate how data quality directly impacts the accuracy of digital model outputs, focusing on the DNV-RP-O501 erosion model, and highlighting the effects of poor-quality data on the reliability of the digital twin. This analysis reinforces the importance of data quality assessment to ensure the reliability of asset degradation results. Finally, we suggest that the quality assessment methods in this study can serve as support tools for decision-making at descriptive, prescriptive, and predictive levels.

Keywords: data quality, digital models, digital twin

Introdução

A indústria de Óleo & Gás busca sempre a máxima eficiência durante o processo de produção. Para isso, a modernização de procedimentos e processos acontece com o objetivo de produzir mais com menor custo. Nesse escopo, a garantia da integridade de ativos se insere no contexto da transformação digital, com investimentos cada vez maiores em Pesquisa e Desenvolvimento (P&D).

Monitorar a integridade de ativos submarinos exige medições locais realizadas por um veículo subaquático operado remotamente (ROV). No entanto, além de ser um processo caro, é muito demorado para planejar e executar. Por isso, modelos digitais representativos dos principais fenômenos de degradação que acometem ativos submarinos podem servir como ferramentas importantes na tomada de decisões (DEMAY, 2023). Entretanto, se os dados fornecidos ao modelo não possuírem boa qualidade, a resposta pode estar comprometida, induzindo o usuário ao erro e, conseqüentemente, à falhas na integridade dos ativos, com conseqüências como derramamentos de óleo e vazamentos de gás, o que resultaria em perdas de segurança e financeiras, danos ambientais e à vidas humanas.

De acordo com a norma DNV-RP-A204 (2021a), para que um gêmeo digital seja confiável e agregue valor ao negócio da empresa, é imprescindível operar com dados válidos e precisos. Em outras palavras, os dados devem ter boa qualidade, a qual pode ser definida como a capacidade dos dados de satisfazer os requisitos técnicos, de sistema e de negócios, ou seja, é uma visão ou uma avaliação da adequação dos dados para servir a seu propósito em um determinado contexto (MAHANTI, 2019)

A norma DNV-RP-0497 (2021b) define a qualidade dos dados em três categorias distintas: sintática, semântica e pragmática.

A qualidade sintática refere-se ao grau de conformidade dos dados com a sua sintaxe especificada; ou seja, o grau de conformidade com que os dados estão expressos de acordo com as expectativas estabelecidas. Este tipo de análise pode ser automatizada, verificando a estrutura dos dados em relação às regras previamente definidas. Um exemplo de dado de baixa qualidade sintática seria quando um valor que deveria ser numérico (por exemplo, um número decimal representado como um *float*) é registrado como uma *string* (texto), causando erros em cálculos ou análises subsequentes.

A qualidade semântica está relacionada ao grau de correspondência entre os dados e o fenômeno que eles representam. Em outras palavras, trata-se da precisão com que os dados refletem a realidade observada. A verificação semântica geralmente é realizada por meio de métodos de amostragem estatística. Exemplo de um dado de baixa qualidade semântica: uma temperatura negativa de óleo produzido na ANM, pois temperaturas negativas não representam a realidade associada a esse dado.

A qualidade pragmática avalia o grau em que os dados são apropriados e úteis para um determinado propósito; ou seja, a adequação dos dados aos requisitos da aplicação em questão. A validação pragmática é, muitas vezes, baseada na percepção dos usuários sobre a utilidade e relevância dos dados, sendo medida por meio de grupos de usuários, questionários e *loops de feedback*. Como exemplo de baixa qualidade pragmática, o registro de pressão que ocorre semanalmente, porém a frequência necessária é de pelo menos um registro diário.

Para mensurar a qualidade dos dados são empregadas métricas comumente agrupadas em dimensões de qualidade de dados. Considerando que os dados medidos correspondem a dados de sensores e podem ser conceituados a partir de um ponto de vista estatístico e temporal, algumas destas dimensões de qualidade se destacam: exatidão, completude, confiabilidade e temporalidade (International Organization for Standardization, 2015).

O presente artigo tem como objetivo avaliar o desempenho de um modelo digital de degradação de ativos submarinos, com foco no modelo de erosão DNV-RP-O501, frente à utilização de dados de alta e baixa qualidade. Para isso, foi desenvolvido um método de avaliação da qualidade dos dados de entrada baseado nas normas e nos métodos de falha dos sensores, com ênfase nas qualidades sintática e semântica. Posteriormente, os parâmetros de avaliação de qualidade foram utilizados como forma de identificar as falhas nos sensores, podendo gerar análises e avisos ao usuário.

Metodologia

Existem normas, como a ISO 8000-8, a DNV-RP-A204 e a DNV-RP-0497, que oferecem recomendações para a avaliação da qualidade de dados em gêmeos digitais, entre as quais se destaca o uso de indicadores de qualidade para os dados. No entanto, nenhuma dessas normas apresenta um método claro para a criação de tais indicadores. Desta forma, esse trabalho propõe desenvolver os parâmetros de avaliação de qualidade com base nas categorias apresentadas na DNV-RP-0497, com enfoque nas categorias sintática e semântica.

Os dados serão aplicados no modelo de erosão descrito na DNV-RP-O501 (2021c). Este modelo utiliza parâmetros de entrada, tais como: propriedades do fluido, taxa de fluxo, geometria do duto e condições ambientais para calcular a energia cinética do fluido e estimar a taxa de erosão em dutos e equipamentos. Baseado em dados empíricos de experimentos de laboratório e medições de campo, o modelo auxilia engenheiros e operadores a avaliar a probabilidade e gravidade da erosão, informando cronogramas de manutenção e prevenção de falhas. Reconhecido e amplamente utilizado

na indústria de óleo e gás, o modelo DNV-RP-O501, apresentado na equação 1, é considerado uma ferramenta confiável para previsão de erosão e é atualizado periodicamente com novos dados e desenvolvimentos (DNV, 2021c).

$$E = K \cdot F(\alpha) \cdot U_p^n \cdot m_p \quad (1)$$

- E : taxa de desgaste erosivo [kg/s];
- K, n : constantes empíricas associadas à relação de erosão entre o material das partículas e o material da superfície erodida, prescritas na DNV-RP-O501 para os valores mais comuns;
- U_p : velocidade média de impacto das partículas [m/s];
- α : ângulo médio de impacto [rad];
- $F(\alpha)$: fator de ductilidade do material da superfície;
- m_p : taxa de produção de areia [kg/s].

Os parâmetros que integram o modelo não correspondem diretamente aos parâmetros de vazão e geometria. Então, para que o modelo seja aplicado, são necessárias etapas de pré-processamento onde são calculados os parâmetros intermediários. Para o desenvolvimento do estudo, foi utilizado o conjunto de dados públicos de Volve, fornecido pela empresa Equinor em 2018. Os parâmetros de entrada analisados, são apresentados na Tabela 1 (EQUINOR, 2018).

Tabela 1 – Parâmetros de produção obtidos do conjunto de dados Volve.

Parâmetro	Unidade
Vazão de óleo	m³/dia
Vazão de água	m³/dia
Vazão de gás	S. m³/dia
Pressão	psi
Temperatura	°C

Além dos dados apresentados na Tabela 1, existem parâmetros cujos valores foram estimados com base nos valores presentes na DNV-RP-O501, exceto a concentração de partículas, que foi estimada dentro de um intervalo de confiança. Não foi possível encontrar informações detalhadas sobre o equipamento, por isso, foi utilizado uma geometria fictícia de um tubo de aço padrão com curvatura, dimensionado por Pauli (2023). Tais valores, apresentados na Tabela 2, são necessários para a realização dos cálculos e não estão explicitamente disponíveis no conjunto de dados Volve.

Tabela 2 – Parâmetros estimados utilizados para cálculos de erosão.

Parâmetro	Valor	Unidade
Diâmetro do duto	162	mm
Raio de curvatura	1,5D	-

Espessura inicial	24	mm
Concentração de partícula	0 - 0,005	%
Densidade da partícula	2650	kg/m ³
Diâmetro da partícula	150	µm
Densidade do material	7800	kg/m ³
Densidade do óleo	880	kg/m ³
Viscosidade do óleo	8×10^{-3}	Pa.s
Densidade da água	1020	kg/m ³
Viscosidade da água	1×10^{-3}	Pa.s
Massa molar do gás	22	kg/kmol
Viscosidade do gás	$3,5 \times 10^{-5}$	Pa.s
Fator de compressibilidade do gás	0,9	-

O modelo “*Black Oil*” é utilizado para estimar a velocidade e as propriedades de mistura do fluido, seguindo a formulação sugerida na DNV-RP-O501 (2021c). Todos os parâmetros necessários para a aplicação do modelo DNV para curvas são derivados dos parâmetros listados acima.

Parâmetros de grande influência na resposta do modelo DNV-RP-O501 são a vazão, a pressão operacional e o teor de areia. Em condições de operação, cada uma dessas variáveis apresenta um número diferente de dados válidos, com períodos de aquisição, comportamentos no tempo em frequência diferentes. Ainda, é comum a existência de *outliers* e dados inválidos (NaN) nos sinais. Tudo isso afeta a qualidade dos dados de entrada, e impacta a saída do modelo digital.

Para cada uma dessas três variáveis, foi determinado um indicador de qualidade, responsável por representar a degradação da qualidade do sinal de acordo com diferentes causas de degradação, como incertezas instrumentais, variações naturais e anômalas, e falhas em sensores.

Nesse processo, cada uma das três variáveis monitoradas inicia com uma qualidade de 100 %. Essa qualidade é ajustada ao longo do tempo, conforme as condições específicas são avaliadas. As avaliações de qualidade ocorrem por meio de um monitoramento contínuo, considerando três condições principais: a presença de valores nulos (NaN), a faixa de valores fisicamente possíveis (FFP) e a mediana do coeficiente de variação dos dados em um determinado período passado (CVP), como dia ou semana. Se qualquer uma dessas condições for violada, a qualidade sofre decrementos temporais pré-definidos.

O objetivo dessas avaliações é estimar a qualidade conforme o método de falha do sinal. Se um sensor falhar, possivelmente a quantidade de NaN se tornará maior. Caso o sinal de saída do sensor congele ou apresente variações anômalas, isso será indicado pelo CVP. Se os dados estiverem fora dos limites fisicamente possíveis, como pressões negativas, a FFP detecta essa inconsistência.

Os dados de cada sensor são avaliados de acordo com esses critérios de qualidade, e comparados com valores limite previamente definidos. Caso um limite seja transposto, aplica-se a taxa de decremento sobre a qualidade avaliada do dado em questão. Os limites utilizados nos parâmetros de avaliação de qualidade foram escolhidos arbitrariamente para cada variável e são apresentados na tabela 3:

Tabela 3 – Parâmetros estimados utilizados para avaliação da qualidade.

Variável	Max. NaN/dia	FFP (min, max)	CVP (max)	Taxa de decremento
Vazão de óleo	50 %	0 , 2000	0,25	0,05
Pressão	50 %	0 , 400	0,082	0,08
Teor de Areia	50 %	0 , 100	0,0033	0,015

O teor de areia é um dado aferido em laboratório e, por isso, não apresenta registro diário. Diferentemente da vazão e da pressão, em que se espera dados diários ou horários. Os dados provenientes de análises realizadas em laboratório têm frequências de aquisição menores quando comparados àqueles não obtidos em análises laboratoriais. Para que um dado laboratorial não fique com a qualidade prejudicada, o decréscimo de qualidade ocorre de acordo com o período de aquisição da variável. Por exemplo, se um dado de teor de areia é registrado a cada quatro meses em média, depois de quatro meses sem dados registrados, a qualidade destes dados começará a decrescer.

Resultados e Discussão

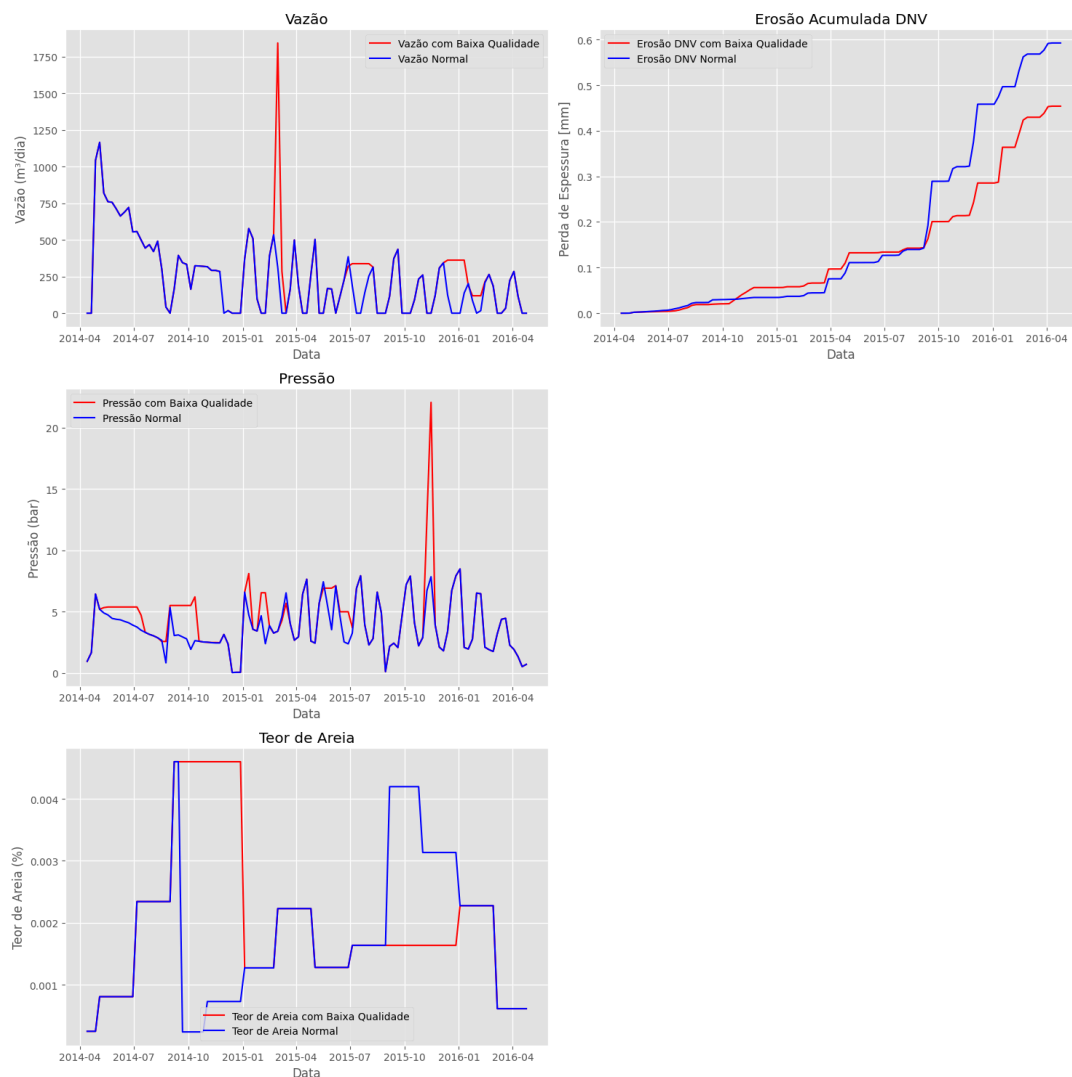
Neste capítulo, apresenta-se os resultados obtidos a partir do estudo de caso, que envolve a realização de cálculos do modelo de erosão DNV-RP-O501 utilizando dados de alta e baixa qualidade. O objetivo é analisar como a qualidade dos dados de entrada influencia a precisão e confiabilidade dos cálculos do modelo.

Primeiramente, os cálculos foram realizados com dados de alta qualidade, considerados confiáveis e livres de anomalias. Em seguida, repetimos os cálculos utilizando dados modificados, de qualidade inferior, para identificar as discrepâncias e avaliar o impacto no desempenho do modelo.

Para definir os dados de baixa qualidade, consideramos premissas que refletem situações comuns de falhas em medições. Esses dados foram caracterizados pela ausência de variação, variação excessiva, presença de valores ausentes (NaN) e valores fora de uma faixa fisicamente possível. A inclusão dessas características permite simular condições reais onde sensores podem falhar, fornecer leituras inconsistentes ou dados fora dos limites esperados.

Os resultados dessa análise estão representados na figura 1. Na primeira coluna, os gráficos exibem a vazão, a pressão e o teor de areia com dados de alta e baixa qualidade. Na segunda coluna, o gráfico apresenta os resultados da erosão do equipamento utilizando dados de alta e baixa qualidade.

Figura 1 – Resultados do modelo de erosão DNV-RP-O501 para Vazão, Pressão e Teor de Areia ao longo do período analisado .



A análise das curvas do gráfico "Erosão Acumulada DNV" revela uma diferença significativa entre os resultados obtidos com dados de alta e baixa qualidade. Utilizando dados de alta qualidade, o desgaste estimado do duto foi de 0,59 mm, enquanto com dados de baixa qualidade, o desgaste estimado foi de 0,45 mm; aproximadamente 23,7 % de diferença, em dois anos. Um erro de 11,8 % ao ano.

A subestimação do desgaste real do duto, resultante do uso de dados de baixa qualidade, pode levar o usuário a inferir incorretamente que o duto sofreu menos desgaste do que realmente ocorreu. Essa situação aumenta significativamente os riscos de vazamentos, pois o equipamento pode sofrer danos muito antes do esperado, podendo resultar em grandes danos ambientais. Além disso, prejudica o desenvolvimento de planos de manutenção e reduz a sua confiabilidade, influenciando assim a assertividade de decisões baseadas em dados, como a de realizar inspeções locais.

Conclusões

Este estudo teve como objetivo avaliar o impacto da qualidade dos dados no desempenho de modelos digitais de degradação de ativos submarinos na indústria de óleo e gás. Para isso, realizamos cálculos do modelo de erosão da DNV-RP-O501 (2021c) com dados de alta e baixa qualidade. A análise dos resultados revelou que os dados de baixa qualidade, caracterizados por ausência de variação, variação excessiva, presença de valores ausentes (NaN) e valores fora de uma faixa fisicamente possível, levaram a uma subestimação significativa do desgaste real do duto, com aproximadamente 23,7 % de diferença.

Esse resultado evidencia a importância de se utilizar dados de alta qualidade e de implementar avaliações da qualidade dos dados para garantir a precisão dos modelos digitais e, consequentemente, a integridade dos ativos submarinos. Dados imprecisos ou de baixa qualidade podem induzir erros graves na interpretação dos resultados, levando a decisões inadequadas que podem comprometer a segurança e a eficácia das operações. Portanto, a aplicação de metodologias de avaliação da qualidade dos dados, como as apresentadas neste estudo, é essencial para mitigar riscos e assegurar a confiabilidade dos gêmeos digitais utilizados na gestão de ativos submarinos.

Agradecimentos

Os autores agradecem à Fundação Certi pela oportunidade e pelo incentivo na elaboração deste estudo.

Referências Bibliográficas

- DEMAY, M. B., et al. **Trustworthiness of subsea assets degradation model estimations and its influence on the reliability of risk-based decisions**. 2023.
- DNV. **Qualification and assurance of digital twins**. DNV-RP-A204 (2021a).
- DNV. **Data quality assessment framework**. DNV-RP-0497 (2021b).
- DNV. **Managing sand production and erosion**. DNV-RP-O501 (2021c).
- EQUINOR. Volve Field Data. 2018. Disponível em: <https://www.equinor.com/energy/volve-data-sharing>. Acesso em: 10/07/2024.
- ISO. **ISO 8000-8: Data quality - Part 8: Information and data quality: Concepts and measuring**. 2015.
- MAHANTI, Rupa. **Data Quality: Dimensions, Measurement, Strategy, Management, and Governance**. 2019.
- PAULI, G.; et al. **On the Impact of Temporal Resolution on Nonlinear Model Accuracy for Predicting Wear Due to Solid Particle Erosion in Digital Twins of Oil and Gas Equipment**. In: OFFSHORE TECHNOLOGY CONFERENCE. 2023.