

Identificação de Estáticas em Poços de Petróleo via Sensor de Pressão de Fundo^{*}

Daniel Folador Rossi, Karin Satie Komati,
Mateus Conrad Barcellos da Costa

*Programa de Pós-Graduação em Computação Aplicada (PPComp)
Instituto Federal do Espírito Santo (IFES), Campus Serra
(e-mail: dan.folador@gmail.com, {kkomati,mcosta}@ifes.edu.br)*

Abstract: The management of oil-producing reservoirs is a set of activities that aim to ensure and enhance their productivity. The formation damage management is one of these activities, whose objective is to identify and mitigate the reduction of reservoir permeability next to the well. Formation damage management techniques include analyzing well pressure and oil rate data collected continuously, mostly in periods when the well is closed to production, called build-ups. The identification of build-ups in the pressure data requires efficient and precise methods that reduce analysis efforts. Considering this context, this work discusses the semi-automatic identification of build-ups through two methods: using the first derivative and using a convolution filter, both of them with the support of DBSCAN clustering algorithm. Both methods were tested for synthetic pressure and oil rate data, having the convolution filter method showed better results. The obtained results suggest the method proved to be a reliable tool to aid in the identification of well build-ups.

Resumo: O gerenciamento de reservatórios produtores de petróleo é um conjunto de atividades que visa garantir e melhorar sua produtividade. A gestão de dano de formação é uma destas atividades, cujo objetivo é identificar e mitigar a redução de permeabilidade do reservatório próximo ao poço de petróleo. Técnicas de gestão de dano incluem a análise de dados de pressão e vazão colhidos continuamente do poço, principalmente em períodos em que o poço esteja fechado, denominados estáticas. A identificação destes períodos nos dados de pressão requer métodos rápidos, precisos e que reduzam esforços de análise. Considerando este contexto, este trabalho discute a identificação semi-automática de estáticas por meio de dois métodos: utilizando a derivada primeira e utilizando um filtro de convolução, ambos com o apoio do algoritmo de clusterização DBSCAN. Os dois métodos foram testados para dados sintéticos de pressão e vazão, tendo o método do filtro de convolução apresentado os melhores resultados. Os resultados obtidos sugerem que o método seja uma ferramenta confiável para auxiliar na identificação de estáticas.

Keywords: formation damage; build-up identification; signal processing; DBSCAN; Permanent Downhole Gauge.

Palavras-chaves: dano de formação; identificação de estáticas; processamento de sinais; DBSCAN; medidor de fundo de poço.

1. INTRODUÇÃO

Ao longo da sua vida operacional, um poço de petróleo pode sofrer do chamado dano de formação, caracterizado como qualquer processo mecânico, químico, biológico ou térmico que cause redução de produtividade e injetividade (Bennion et al., 2002). Neste sentido, a Gestão de Dano de Formação é uma atividade fundamental para garantir o bom desempenho dos poços (Civan, 2015).

Uma das formas de se avaliar o dano de formação é por meio da análise de vazão e pressão do poço, realizada durante períodos de fluxo, quando o poço é aberto, ou períodos de estática, quando o poço é fechado. O principal equipamento utilizado nestas análises é um registrador de pressão denominado PDG — *Permanent Downhole Gauge* (em português, medidor de fundo de poço) (Houze et al., 2020), introduzido dentro do poço. Tipicamente, os períodos mais utilizados para a análise são estáticas, por apresentarem maior controle na vazão (vazão nula) e menor ruído nos dados (Bourdet, 2002).

Um período de estática é caracterizado por uma subida abrupta de pressão de fundo após um período de fluxo. A interpretação desses dados requer que sejam identificados tanto a estática, quanto o instante exato em que ela se inicia. Entretanto, a identificação de estáticas nem sempre

^{*} Agradecemos à FAPES (Fundação de Amparo à Pesquisa e Inovação do Espírito Santo) e à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro dado por meio do PDPG (Parcerias Estratégicas nos Estados da CAPES). Também agradecemos ao Propós (Programa Institucional de Apoio à Pós-Graduação Stricto Sensu) do IFES pelo apoio financeiro.

é uma tarefa trivial (Suzuki et al., 2009). Uma dificuldade é que os dados de vazão estão frequentemente defasados dos dados de pressão, fazendo com que o ponto inicial possa não ser de fácil identificação; outra dificuldade é que a assinatura da pressão (forma da variação de subida de pressão) de uma estática é similar à assinatura de um fluxo de baixa vazão.

Neste contexto, este trabalho discute duas abordagens para a identificação de estáticas em poços produtores de petróleo por meio da análise de dados de pressão de fundo, avaliando também a identificação de seu ponto inicial. A primeira delas utiliza a derivada primeira com limiar (*threshold*), baseado no trabalho de Chang et al. (2019), e a segunda abordagem utiliza um filtro de convolução com limiar, baseado no trabalho de Suzuki et al. (2018). Em ambas as abordagens, segue-se um passo de agrupamento, em que foi usado o método DBSCAN (do inglês, *Density-Based Spatial Clustering of Applications with Noise*) (Ester et al., 1996). Os experimentos usam dados simulados que descrevem situações reais encontradas no ambiente de produção. Os dados foram disponibilizados de forma pública para possibilitar a realização de outros trabalhos.¹

O restante do trabalho está organizado da seguinte forma: a Seção 2 contextualiza a identificação de estáticas dentro do processo de interpretação como um todo; na Seção 3 são apresentados os trabalhos relacionados com a identificação de estáticas; na Seção 4 é mostrada a metodologia dos experimentos realizados; a Seção 5 traz as discussões levantadas por este trabalho, e a Seção 6 suas conclusões.

2. CONTEXTO

Poços de petróleo são perfurações na crosta terrestre visando produzir óleo, gás natural ou ambos. Após ser perfurado, são descidos tubos de aço dentro do poço (tubos de revestimento) que serão assentados com cimento em suas paredes, visando garantir a integridade estrutural. Em seguida, é descida a coluna de produção, que irá comunicar a zona produtora, no fundo do poço, e a superfície. A coluna de produção é uma coluna de vários tubos de aço, podendo conter válvulas e outros equipamentos utilizados para o controle do poço e gerenciamento da produção. Entre esses equipamentos, encontra-se o PDG, responsável por registrar dados de pressão de fundo, bastante utilizados para gerenciamento do reservatório.

O processo típico de análise utilizando PDGs pode ser dividido nas etapas abaixo:

1. Extração de dados de pressão e vazão;
2. Pré-processamento de dados (filtragem de dados, remoção de ruído e valores espúrios);
3. Identificação de períodos de estática interpretáveis;
4. Ajuste de pressão inicial e final das estáticas;
5. Sincronização da pressão com a vazão;
6. Gerar gráficos especializados de análise;
7. Identificar assinaturas típicas (estocagem e radial);
8. Interpretar os parâmetros obtidos;
9. Gerar indicadores para monitoramento no tempo.

A utilização de dados PDG, principalmente para a gestão de dano de formação, tem crescido nos últimos anos, mas

¹ https://github.com/danfolador/SBAI2021_dataset

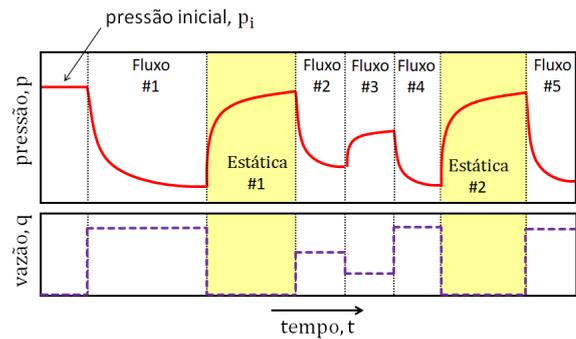


Figura 1. Desenho esquemático do comportamento de pressão de fundo (p) em um poço de petróleo com a variação de vazão (q).

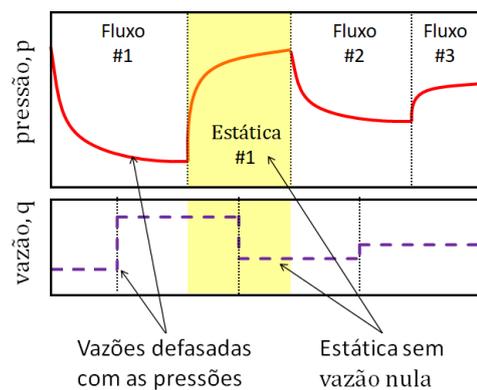


Figura 2. Desenho esquemático mostrando a defasagem entre os registros de vazão e pressão.

por ser uma tarefa trabalhosa e demorada, não é realizada em tempo real, por vezes realizada muito tempo depois da aquisição dos dados.

Ao longo dos anos, foram propostas várias metodologias com a finalidade de automatizar total ou parcialmente a análise e interpretação desses dados (Athichanagorn et al., 1999; Viberti et al., 2007; Rai et al., 2007; Olsen et al., 2005; Suzuki et al., 2009; Houze et al., 2011; Suzuki et al., 2018; Chang et al., 2019). Neste trabalho, busca-se automatizar a Etapa 3 — identificação de estáticas e parte da Etapa 4 com a identificação do ponto de início de cada estática.

Um período de estática é caracterizado por uma subida abrupta de pressão de fundo quando o poço é fechado. As Figuras 1 e 2 apresentam gráficos de pressão de fundo (p) e a vazão (q) variando com o tempo (t). Quando a vazão (curva tracejada em roxo) apresenta uma subida, a pressão de fundo (curva contínua em vermelho) responde com uma queda, e vice-versa. Na Figura 1 há sete períodos de transientes, sendo dois deles períodos de estáticas, marcados pelo fundo hachurado em amarelo. Esse gráfico é uma visão didática do processo, que é bem mais ruidoso com dados reais.

Como a pressão de fundo responde de maneira inversa à vazão, reduções na vazão também apresentam subidas de pressão (em menor escala), como pode ser observado no Fluxo #3 da Figura 1. Deste modo, simples subidas de pressão não indicam necessariamente a ocorrência de

estáticas, exemplificando a dificuldade relatada de que a assinatura da pressão durante uma estática é similar à assinatura de uma restrição de fluxo.

Além disso, tipicamente não há medidores de vazão disponíveis por poço, de modo que as vazões são rateadas para cada poço utilizando uma média, normalmente diária. Os dados de pressão, por outro lado, são individuais por poço e possuem frequência de amostragem na ordem de segundos, o que faz com que as mudanças de vazão não acompanhem de forma sincronizada as respostas de pressão. E se uma estática ocorre entre dois registros de vazão, a vazão nula, que evidenciaria a estática, não será registrada. A Figura 2 exemplifica este problema: as mudanças de vazão, representadas pela curva tracejada roxa, ocorrem em momentos diferentes da mudança de pressão, representada pela curva contínua vermelha, e portanto a Estática #1 (fundo hachurado de amarelo) apresenta vazões não-nulas.

Fechamentos de poços podem ocorrer de forma planejada ou não-planejada. Para identificar estáticas planejadas bastaria um registro de eventos operacionais, mas esses, quando realizados, muitas vezes não estão em forma estruturada, sendo de difícil compilação. Para as estáticas não-planejadas, que ocorrem devido a alarmes de segurança, por exemplo, curtas e antigas, há ainda menos informações disponíveis.

As dificuldades aumentam quando se trata de dados muito antigos e de históricos longos de produção. A quantidade significativa de dados produzidos por PDGs é um desafio para as técnicas tradicionais de interpretação de transientes de pressão, tornando a análise trabalhosa, e impedindo uma resposta imediata quando um problema é detectado. Deste modo, a busca pela automatização se faz necessária.

3. TRABALHOS RELACIONADOS

Nos trabalhos correlatos foram selecionados os artigos em que a identificação de estáticas foi realizada via abordagem de processamento digital de sinais em conjunto com outras técnicas, tais como técnicas de agrupamento não supervisionado, além de calibração de parâmetros via definição de limiares.

Athichanagorn et al. (1999) propuseram a utilização de filtros de *wavelet* para a detecção de transientes de pressão causados por mudanças de vazão. O ponto de início é ajustado para a interseção de duas retas: a tangente aos dados anteriores ao início do transiente, e a tangente aos dados posteriores ao início do transiente.

Viberti et al. (2007) apontaram que a metodologia de Athichanagorn et al. (1999) é sensível à amostragem da vazão, então propuseram combiná-la com valores de limiares. Entretanto, tais valores dependem de cada caso, precisando ser definidos por tentativa e erro.

Suzuki et al. (2009) propuseram o uso de um filtro de convolução para identificar estáticas. O início de uma estática é identificado pelo aumento da frequência de sinais positivos da derivada da pressão, ao invés da mudança do próprio sinal. O método também pode ser combinado com outras técnicas de redução de ruído, como média móvel e o uso de limiares. Posteriormente, Suzuki et al.

(2018) apresentaram um aprimoramento desta metodologia. Nesta nova proposta, utilizam um algoritmo iterativo para ajustar o tempo correto de início da estática, buscando o ponto que resulte na assinatura esperada para o gráfico especializado.

Rai et al. (2007) também concluíram que o método de Athichanagorn et al. (1999) não seria adequado, propondo três algoritmos alternativos: os filtros de Savitzky-Golay (SVG) (Savitzky and Golay, 1964), uma nova abordagem chamada de método da Segmentação e uma variante da Segmentação. O ponto de início da estática é ajustado para valores altos da derivada de SVG utilizando limiares.

Houze et al. (2011) sugeriram as derivadas geradas pelo filtro SVG como alternativa aos filtros de *wavelet*. As derivadas são calculadas e combinadas em uma única curva, e o início das estáticas são definidos por limiares aplicados aos valores mais altos positivos. Para refinar o tempo de início, utilizam um algoritmo heurístico para estáticas com variação abrupta de pressão, e para as de variação suave, utilizando a interseção de duas retas: a tangente aos dados de pressão antes da estática e a reta vertical posicionada na subida de pressão do fechamento.

Chang et al. (2019) propuseram a clusterização K-Means (Kodinariya and Makwana, 2013) para a identificação de estáticas, um método de aprendizado de máquina não supervisionado. Primeiramente, aplica-se a derivada nos dados de pressão, que irão indicar valores altos positivos para início de estáticas, e em seguida, aplica-se um limiar para separar variações menores, provenientes de ruído e fluxo. Por fim, o K-Means é aplicado ao sinal, agrupando os pontos, de modo que cada *cluster* identifica uma estática. Quanto ao ajuste do ponto inicial, utilizam a interseção de duas retas: a reta horizontal tangente aos dados de fluxo antes da estática e a reta de inclinação máxima dos pontos durante a subida de pressão. Os autores comentam que a técnica ainda seleciona alguns falsos positivos, mas que serão descartadas em seu próximo passo, que utiliza a similaridade de imagens nos gráficos especializados.

A Tabela 1 resume as técnicas dos trabalhos descritos quanto à identificação de estáticas. Os primeiros trabalhos usavam filtro *wavelets*, mas que Viberti et al. (2007), Rai et al. (2007) e Houze et al. (2011) verificaram não ser uma técnica viável por apresentarem muitos falsos positivos. As técnicas de tratamento de sinais utilizando limiares apresentaram resultados melhores do que sem o seu uso, embora seja um parâmetro que dependerá de calibração.

Tabela 1. Resumo dos métodos de identificação de estáticas dos trabalhos correlatos.

Autores	Principais métodos
Athichanagorn et al. (1999)	Filtro <i>wavelet</i>
Viberti et al. (2007)	Filtro <i>wavelet</i> com limiar
Suzuki et al. (2009)	Filtro de convolução com limiar
Suzuki et al. (2018)	Filtro de convolução com limiar
Rai et al. (2007)	Filtro SVG ou Segmentação
Houze et al. (2011)	Derivadas SVG
Chang et al. (2019)	Derivada primeira e K-means

Como proposto por Chang et al. (2019), o uso posterior de um algoritmo de clusterização melhora a identificação de estáticas para grandes volumes de dados, pois diminuem os ruídos dos dados de entrada. O uso do K-Means para

tal objetivo é bem difundido para tratar de dados de EEG (Khanna et al., 2014). No entanto, testes preliminares com os dados deste trabalho não apresentaram bons resultados e optou-se pela técnica DBSCAN. Os trabalhos de Dudik et al. (2015) e Lu et al. (2016) indicam que o DBSCAN teve resultados mais corretos que o K-Means na separação dos agrupamentos.

Quanto ao ajuste do início da estática, de forma resumida, os trabalhos listados utilizam duas técnicas de interseção de duas retas: (i) interseção da tangente aos dados antes do início do transiente com a tangente depois do transiente (Athichanagorn et al., 1999) (Houze et al., 2011) (Chang et al., 2019), e (ii) inclinação unitária no gráfico especializado (Suzuki et al., 2018). Foi escolhido o método (i), por ser o mais utilizado dentre os trabalhos.

4. MATERIAIS E MÉTODOS

A metodologia dos experimentos tem os seguintes passos: (i) a filtragem de dados para suavização dos dados de pressão; (ii) a identificação de estáticas, testando os métodos de Derivada Primeira (Método 1) e Convolução (Método 2); (iii) a clusterização, onde os dados são tratados e agrupados por estáticas; e (iv) o ajuste do ponto inicial de cada estática. A base de dados e os passos listados são descritos nas subseções a seguir.

4.1 Base de dados

Foram geradas dez bases de dados sintéticos de pressão e vazão² usando *software* comercial de interpretação de transientes de pressão: a ferramenta Saphir, da KAPPA³, ferramenta disponível no ambiente experimental. As propriedades de reservatório e fluido utilizados na simulação seguem na Tabela 2. As vazões foram geradas por uma distribuição normal, definidas pela vazão média e o desvio padrão conforme Tabela 3. Além disso, para cada base de dados ainda se variou a permeabilidade da rocha. Os valores de ambas as tabelas foram obtidos do conhecimento tácito do especialista.

Tabela 2. Propriedades de reservatório e fluido utilizados na simulação dos dados de pressão.

Propriedade do reservatório	
Limites do reservatório	Infinito
Pressão inicial	351,535 kgf/cm ²
Espessura do reservatório	30,0 m
Raio do poço	0,108 m
Porosidade da rocha	25%
Propriedade do fluido	
Fluido produzido	Apenas óleo
Coefficiente de estocagem	0.05 (m ³)/(kgf/cm ²)
Densidade do óleo	28°API
Fator volume-formação	1,2755 m ³ /m ³
Viscosidade do óleo	1,18896 cP
Compressibilidade total	1,62895E-4 (cm ² /kgf)

Cada base de dados foi gerada inicialmente com 65 transientes de duração média de 24h, totalizando pouco mais de dois meses de produção. Dentre essas, existem três estáticas verdadeiras e três “estáticas falsas” (períodos de

Tabela 3. Parâmetros de cada base de dados.

Dataset	Vazão média (m ³ /d)	Desvio padrão da vazão (m ³ /d)	Permeabilidade simulada (mD)	Número de picos de pressão
dataset_0	500	50	500	1
dataset_1	1000	100	500	2
dataset_2	1500	150	500	1
dataset_3	2500	250	500	2
dataset_4	3000	300	500	1
dataset_5	500	50	1000	2
dataset_6	1000	100	1000	0
dataset_7	1500	150	1000	0
dataset_8	2500	250	1000	1
dataset_9	3000	300	1000	5

fluxo onde a vazão cai para 30% da vazão média). Os dados de pressão foram simulados no próprio Saphir, utilizando como ruído um valor aleatório entre [-0,04; 0,04] kgf/cm². Em seguida, os dados de vazão foram re-amostrados para simular a defasagem com a pressão.

Algumas bases de dados possuem períodos sem dados, a fim de simular problemas no registrador. Em algumas bases foram incluídos picos de pressão (variações abruptas e curtas), simulando perturbações nas válvulas do poço, conforme última coluna da Tabela 3. Assim, a quantidade de transientes totais na base dataset_0 é de 66 (65 gerados mais 1 pico de pressão incluído), na base de dataset_1 é de 67 e assim por diante.

4.2 Filtro Savitzky-Golay (SVG)

Os dados de pressão devem passar por um filtro de suavização, de forma a reduzir ruído, valores espúrios e variações abruptas do fluxo. Este é um passo necessário, sem o qual os valores da derivada primeira e da convolução apresentam oscilações, dificultando a identificação das estáticas.

Para a filtragem, foi utilizado o filtro Savitzky-Golay (SVG) (Savitzky and Golay, 1964) que é um filtro de suavização que consegue manter informações de baixa frequência. O filtro realiza um ajuste polinomial do sinal, podendo ser utilizado tanto para suavização dos dados (Rai et al., 2007), quanto para calcular as derivadas suavizadas deste sinal (Houze et al., 2011). Neste trabalho, o filtro foi utilizado tanto como filtro de dados quanto para gerar a derivada primeira.

Para um *dataset* formado por um conjunto de pontos x_j, y_j , com $j = 1, 2, \dots, n$, sendo y_j o valor medido e x_j uma variável independente, os dados podem ser tratados como um conjunto de m coeficientes de convolução, C_i , de acordo com a Equação 1.

$$Y_j = \sum_{i=\frac{1-m}{2}}^{\frac{m-1}{2}} C_i y_{j+i}, \quad \frac{m-1}{2} \leq j \leq n - \frac{m-1}{2} \quad (1)$$

Para calcular a derivada nos experimentos deste trabalho foi utilizado um polinômio de grau cinco.

² https://github.com/danfollador/SBAI2021_dataset

³ <https://www.kappaeng.com/software/saphir/overview>

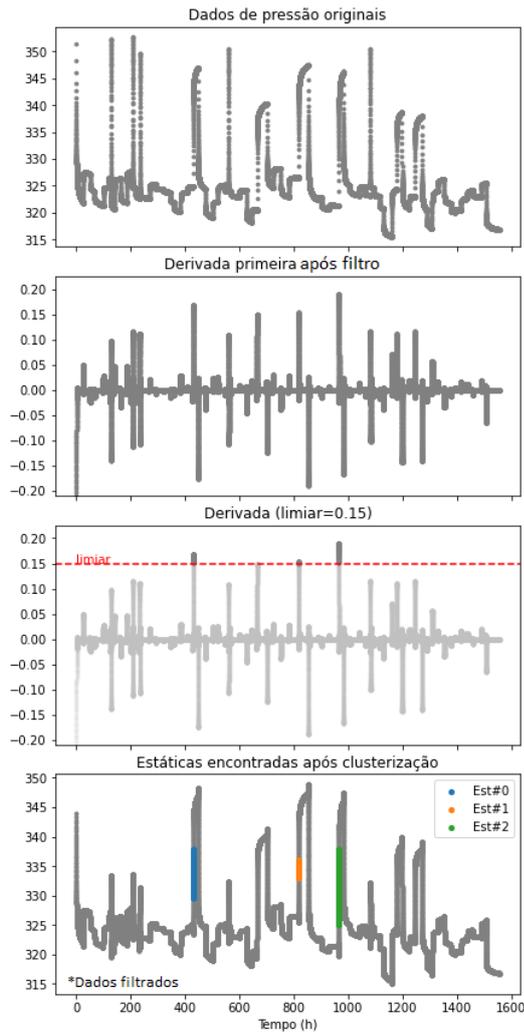


Figura 3. Método 1 — Derivada primeira, aplicado aos dados do dataset_9, usando limiar=0,150 e janela de filtro=351.

4.3 Identificação de estáticas pelo Método 1

O primeiro método, da derivada primeira com limiar, utiliza a derivada dos dados de pressão pelo tempo, conforme proposto por Chang et al. (2019), de modo a evidenciar as mudanças abruptas de assinatura. Tipicamente, inícios de estática apresentam alto valor positivo de derivada e inícios de fluxos apresentam alto valor negativo. Quanto maior o valor absoluto da derivada, mais abrupta é a mudança de pressão. Assim, aplica-se um limiar na derivada para isolar as variações altas positivas.

Conforme ilustrado na Figura 3, primeiro calcula-se a derivada dos dados de pressão usando o filtro SVG (o segundo gráfico da figura). Em seguida, é aplicado um limiar (linha vermelha pontilhada) à derivada (terceiro gráfico da figura). Os dados abaixo do limiar são descartados e os dados acima são projetados no eixo x, ficando separados por valores nulos. Sobre estes, é aplicada a clusterização DBSCAN, de modo que cada *cluster* identifica uma estática, marcada por cores diferentes no último gráfico. O valor do limiar é dependente dos dados e deve ser ajustado por tentativa e erro para gerar o melhor resultado.

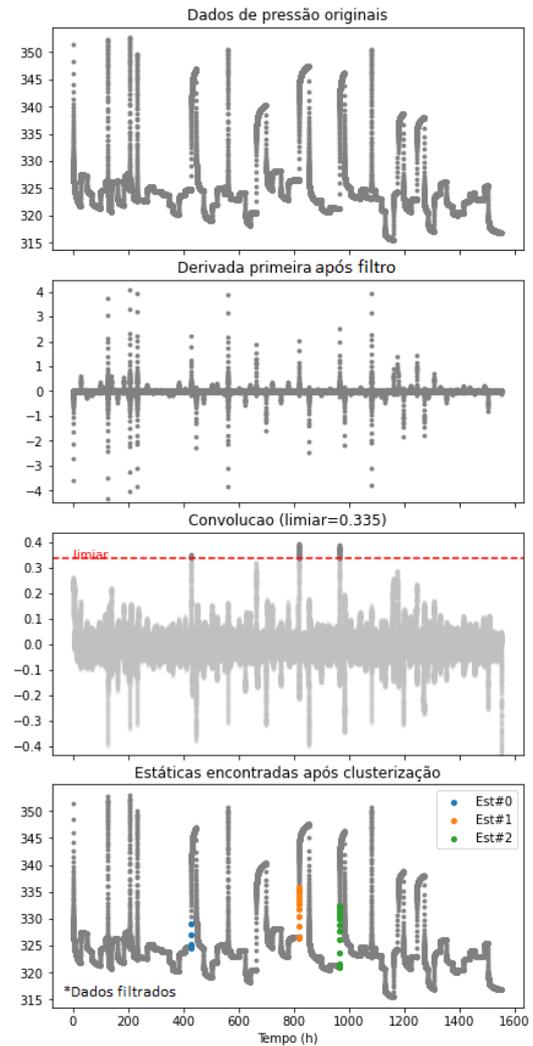


Figura 4. Método 2 — Convolução, aplicado aos dados do dataset_9, usando limiar=0,335, janela de filtro=7 e meia janela de convolução=301.

4.4 Identificação de estáticas pelo Método 2

O método da convolução com limiar utiliza o filtro de convolução proposto por Suzuki et al. (2009) e Suzuki et al. (2018). Como a derivada apresenta valores positivos para o início da estática e negativos para inícios de fluxo, os autores propuseram uma função de pulso, com valor 1 caso a derivada seja positiva e 0 caso contrário, conforme (2).

$$F(t) = \begin{cases} 1, & \text{caso } \frac{dp}{dt} > 0 \\ 0, & \text{caso contrário} \end{cases} \quad (2)$$

Para este trabalho, a derivada dp/dt foi calculada pelo filtro SVG, utilizado um polinômio de grau cinco.

A função pulso $F(t)$ apenas indica quando a derivada fica positiva. O início da estática ocorre quando $F(t)$ muda de 0 para 1, mas a existência de ruído faz com que o sinal da derivada alterne muitas vezes, principalmente durante o fluxo. Para suavizar este comportamento, os dados são previamente filtrados e, em seguida, a função $F(t)$ é convoluída com uma função filtro $G(t)$, conforme

(3), sendo w metade do tamanho da janela da função $G(t)$. A função já se encontra normalizada para que o valor da convolução fique entre -1 e 1.

$$G(t) = \begin{cases} \frac{2}{w^2}t + \frac{2}{w}, & -w \geq t \geq 0 \\ \frac{2}{w^2}t - \frac{2}{w}, & 0 \geq t \geq +w \end{cases} \quad (3)$$

Calcular a convolução de $F * G$, de acordo com (4), equivale a mover uma janela de tamanho $2w$ ao longo de $F(t)$ e, para cada ponto t , somar a área sob $F(t)$ à esquerda de t e subtrair da área à direita de t , ponderando cada área linearmente com a distância de t .

$$F * G = \int_{-\infty}^{\infty} F(\tau)G(t - \tau) d\tau \quad (4)$$

Ou seja, a convolução $F * G$ calcula a frequência de ocorrências de derivadas positivas. Valores positivos de $F * G$ indicam aumento na frequência de derivadas positivas, e quanto maior o valor absoluto $F * G$, mais abrupta é a mudança na derivada.

A Figura 4 ilustra o Método 2. Após o cálculo da derivada e da convolução, aplica-se um limiar (linha vermelha pontilhada) aos dados da convolução (terceiro gráfico da figura). Os dados abaixo do limiar são descartados e os dados acima são projetados no eixo x, ficando separados por valores nulos. Sobre estes, é aplicada a clusterização DBSCAN, de modo que cada *cluster* identifica uma estática, marcada por cores diferentes no último gráfico. Como no Método 1, o valor do limiar é dependente dos dados, e deve ser ajustado por tentativa e erro para gerar o melhor resultado.

4.5 Clusterização DBSCAN

Sigla inglesa para Clusterização Espacial Baseada em Densidade para Aplicações com Ruído, o DBSCAN é um algoritmo de aprendizado não supervisionado utilizado para formar agrupamentos ou *clusters* (Ester et al., 1996). Trata-se de um método baseado em densidade: o algoritmo classifica *clusters* como áreas de alta densidade separadas por áreas de baixa densidade. Por esse motivo, seus *clusters* podem ter qualquer formato, ao contrário de outros métodos.

O algoritmo DBSCAN utiliza dois parâmetros: uma distância, ϵ , e o mínimo de pontos necessários para formar uma região densa, min_pts . Para um ponto não-visitado, o algoritmo busca seus pontos vizinhos a uma distância ϵ . Caso o ponto tenha pelo menos um número de vizinhos igual a min_pts , ele é classificado como um ponto central e parte de um *cluster*. Quando um ponto é classificado como parte de um *cluster*, seus vizinhos em uma distância ϵ também são parte deste *cluster*. Por outro lado, se um ponto não é parte de um *cluster*, é classificado como ruído (podendo ser reclassificado para outro *cluster*). Este processo é repetido até que se encontrem todos os *clusters* do conjunto.

4.6 Ajuste de ponto inicial

Depois de identificada a estática, é preciso ajustar o ponto exato de início da mudança de pressão. Foi escolhido um método presente na literatura (Athichanagorn et al., 1999; Houze et al., 2011; Chang et al., 2019), podendo ser aplicado para os Métodos 1 e 2. O método é o de interseção da tangente aos dados antes do início do transiente com a tangente depois do transiente.

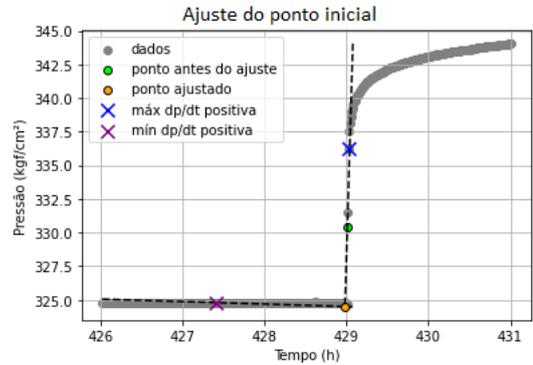


Figura 5. Ajuste do ponto inicial da estática.

A Figura 5 apresenta graficamente como é feito o ajuste. No gráfico de pressão pelo tempo, aplica-se uma janela de tempo ao redor da estática identificada, e são traçadas duas retas (retas tracejadas): a tangente aos pontos de mínima derivada positiva (marcado com um x roxo), e a tangente aos pontos de máxima derivada positiva (marcado com um x azul). O ponto inicial é então ajustado do ponto anterior (ponto verde) para a interseção destas duas retas (ponto laranja).

4.7 Métricas de avaliação

O problema abordado neste trabalho pode ser tratado como uma classificação binária, que avalia quando um transiente é ou não uma estática. Problema similar ao de detecção de *onsets* (início das batidas) em estruturas rítmicas, em que são usadas três métricas usuais de avaliação (Degara et al., 2010): precisão, revocação e acurácia.

As métricas são calculadas conforme as Equações 5 e 6:

$$\text{Precisão} = \frac{VP}{VP + FP}, \text{ Revocação} = \frac{VP}{VP + FN}, \quad (5)$$

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (6)$$

Sendo VP o número de verdadeiros positivos (isto é, estáticas corretamente identificadas); FP o número de falsos positivos (isto é, fluxos e picos identificados como estáticas); VN o número de verdadeiros negativos (isto é, fluxos e picos não identificados como estáticas) e FN o número de falsos negativos (isto é, estáticas que não foram identificadas).

Tomando como exemplo o resultado da Figura 4, teríamos: $VP = 3$ (todas as três estáticas verdadeiras foram identificadas), $FP = 0$ (nenhum fluxo ou pico de pressão foi identificado como estática), $FN = 0$ (nenhuma estática

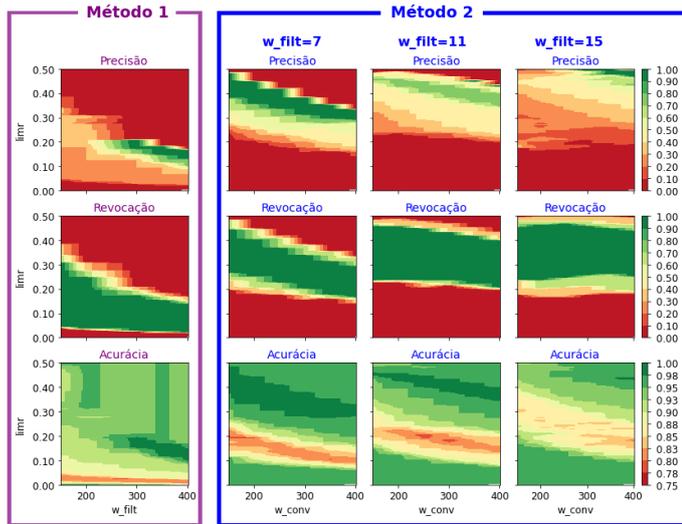


Figura 6. Resultados da avaliação para o dataset_9, com variações de limiar e tamanho de janela do filtro para o Método 1 e com variações de limiar, tamanho da janela do filtro e tamanho da meia janela de convolução para o Método 2.

verdadeira foi descartada) e $VN = 67$ (conforme Tabela 3, são 70 transientes, sendo que 3 deles são estáticas, logo 67 não são estáticas). Os resultados, então, seriam Precisão=1,0; Revocação=1,0 e Acurácia=1,0. Se o limiar tivesse sido um pouco menor, seriam identificadas 4 estáticas, sendo uma incorreta, com $VP = 3$, $FP = 1$, $FN = 0$, $VN = 66$, Precisão=0,75; Revocação=1,0 e Acurácia=0,98.

5. RESULTADOS E DISCUSSÃO

Os dois métodos foram testados para as dez bases de dados. As Figuras 3 e 4 ilustram um dos resultados para a base de dados dataset_9. No primeiro gráfico de ambas as figuras temos os dados originais de pressão contra o tempo; no segundo gráfico temos a derivada da pressão calculada após a filtragem de dados, no terceiro gráfico temos a derivada/convolução com o limiar aplicado (linha vermelha tracejada). No último gráfico, temos todas as três estáticas corretas, identificadas por cores diferentes (azul, laranja e verde).

Para suavizar os picos de pressão no Método 1 (Figura 3), foi necessária uma janela de filtro com valor de $w_filt = 351$. Porém, isso fez com que os dados também fossem excessivamente atenuados, distorcendo os dados originais, como pode ser verificado ao se comparar o primeiro com o último gráfico da figura. Essa atenuação impactou no ajuste do ponto inicial das estáticas, que teve que ser realizado nos dados originais, sem filtro. Na presença de ruído expressivo, será necessária a re-filtragem antes deste ajuste.

O Método 2 (Figura 4) utilizou um valor da janela de filtro igual a $w_filt = 7$ e valores da meia-janela de convolução de $w_conv = 301$ para suavizar os picos. Isto fez com que os dados não fossem excessivamente atenuados, como se pode notar ao se comparar o primeiro com o último gráfico da figura. Também por essa razão, o ajuste do ponto inicial pôde ser realizado diretamente nos dados filtrados.

Nos exemplos das Figuras 3 e 4, o valor do limiar para o Método 1 foi de 0,150 e no Método 2 foi de 0,355. Em ambos os casos, se o valor do limiar diminui, são detectadas mais estáticas incorretas. Para analisar essa variação foram elaborados os gráficos ilustrados na Figura 6. Para o Método 1, variou-se o valor do limiar ($limr$) na faixa de 0,00 à 0,50 e janela de filtro (w_filt) de 150 à 400. Para o Método 2, foram variados os seguintes parâmetros: valor do limiar ($limr$) de 0,00 à 0,50, janela da convolução (w_conv) de 150 à 400 e, janela do filtro (w_filt) para os valores 7, 11 e 15.

Os resultados do Método 1 estão dentro do quadro roxo e do Método 2 no quadro azul. Os resultados de precisão estão nos gráficos da primeira linha, seguido pelos de revocação na segunda linha e de acurácia nos gráficos da última linha. Regiões em verde indicam os melhores resultados, com valores próximos à 1,0 para todas as métricas. Regiões em vermelho indicam resultados ruins, com valores próximos de 0,0 para precisão e revocação e valores próximos à 0,75 para acurácia.

Quanto maior a região verde-escura no gráfico, mais tolerante é o método com relação aos picos de pressão e vazões baixas e menos sensível às variações de limiar. No gráfico de acurácia, este aspecto significa facilidade para identificar apenas as estáticas verdadeiras. É possível notar que as maiores regiões verde-escuras nos gráficos estão no Método 2 (gráficos do quadro azul), para $w_filt = 7$. Para $w_filt = 7$, para quaisquer valores de tamanho da janela da convolução usada, há sempre um valor de limiar em que as métricas são próximas de 1,00. Esta tolerância do Método 2 ocorre devido à convolução, que considera a duração das estáticas, o que pode ser um atributo interessante para descartar estáticas que, mesmo verdadeiras, seriam curtas demais para serem interpretadas. Por outro lado, à medida que o tamanho da janela de filtro aumenta, as regiões verde-escuras diminuem, tanto para a precisão quanto para a acurácia. No Método 1, as regiões verde-escuras são menores, indicando uma menor possibilidade na seleção de parâmetros para se obter os melhores resultados.

O método utilizado para ajuste do ponto inicial funcionou corretamente para as estáticas identificadas. Infelizmente, não foi possível comparar os resultados dos trabalhos correlatos, pois nenhum deles forneceu a base de dados de forma pública.

6. CONCLUSÃO

Este trabalho discutiu a detecção semi-automática de períodos de estática em poços produtores de petróleo, por meio de dois métodos analíticos: o método da derivada primeira da pressão com limiar e o de filtro de convolução com limiar. Como mecanismo de detecção foi adotado a clusterização DBSCAN, a partir da qual os períodos são classificados ou não como estáticas. Os dois métodos foram implementados e avaliados sobre um conjunto de dados sintéticos que reproduziram situações realísticas e onde a identificação da estática apresentava dificuldades. Sendo encontrada a melhor calibração, ambos os métodos foram capazes de ignorar problemas como a ausência de dados em alguns instantes de tempo, picos de pressão e “estáticas falsas” induzidas por períodos de vazão muito baixa.

Foi também realizada uma análise sobre a seleção de parâmetros e o impacto de sua variação nos resultados. O método baseado em filtro de convolução apresentou maior facilidade de ajuste, com melhores resultados para uma maior variedade de parâmetros, conforme pode ser observado pela presença de maiores áreas verde-escuras na Figura 6. Isso se deve ao fato de que a janela de convolução permite que se descartem variações curtas nos dados de pressão. O método da derivada primeira apresentou dificuldades para descartar picos de pressão, exigindo janelas de filtro maiores que o primeiro método para suavizá-los. Esse tamanho de janela pode atenuar excessivamente os dados e impactar no ajuste do ponto inicial da estática. Por este motivo, para este método, o ajuste do ponto inicial teve que ser feito diretamente nos dados não-filtrados.

Os estudos realizados se mostraram satisfatórios e sugerem que os métodos investigados são apropriados para apoiar esta etapa da análise do dano. Porém, ambos os métodos são baseados em limiares, cujos valores dependem dos dados e precisam do ajuste manual de parâmetros. Dessa forma, o processo requer ainda o aperfeiçoamento do método de ajuste para a sua total automatização.

Para trabalhos futuros, pretende-se aplicar os métodos dos trabalhos correlatos aos dados deste trabalho para comparação de resultados. Pretende-se, também, avaliar todos os métodos com um banco de dados reais. Além disso, deseja-se investigar a aplicação de métodos que não usem limiares, tal como o método de similaridade de imagens proposto por Olsen et al. (2005). Outra possibilidade que será investigada com este propósito é a abordagem baseada em modelo neural NARX (*Nonlinear Autoregressive with eXogenous inputs*), que já foi usado em sistema de detecção de distúrbios de qualidade da energia elétrica com resultados promissores (Alcântara et al., 2019).

REFERÊNCIAS

- Alcântara, I., Barbosa, B.H.G., and Ferreira, D. (2019). Detecção de distúrbios em qualidade de energia usando modelos narx neurais. In *Anais eletrônicos do 14^o Simpósio Brasileiro de Automação Inteligente (SBAI 2019)*, 1707–1712. SBA. doi:10.17648/sbai-2019-111384.
- Athichanagorn, S., Horne, R.N., Kikani, J., et al. (1999). Processing and interpretation of long-term data from permanent downhole pressure gauges. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.
- Bennion, D.B. et al. (2002). An overview of formation damage mechanisms causing a reduction in the productivity and injectivity of oil and gas producing formations. *Journal of Canadian Petroleum Technology*, 41(11).
- Bourdet, D. (2002). *Well test analysis: the use of advanced interpretation models*. Elsevier, Amsterdã, Holanda.
- Chang, O., Pan, Y., Dastan, A., Teague, D., Descant, F., et al. (2019). Application of machine learning in transient surveillance in a deep-water oil field. In *SPE Western Regional Meeting*. Society of Petroleum Engineers.
- Civan, F. (2015). *Reservoir formation damage: fundamentals, modeling, assessment, and mitigation*. Gulf Professional Publishing, Elsevier.
- Degara, N., Pena, A., Davies, M.E., and Plumbley, M.D. (2010). Note onset detection using rhythmic structure. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 5526–5529. IEEE.
- Dudik, J.M., Kurosu, A., Coyle, J.L., and Sejdić, E. (2015). A comparative analysis of dbSCAN, k-means, and quadratic variation algorithms for automatic identification of swallows from swallowing accelerometry signals. *Computers in biology and medicine*, 59, 10–18.
- Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, volume 96, 226–231.
- Houze, O., Viturat, D., Fjaere, O.S., et al. (2020). *Dynamic Data Analysis: The Theory and Practice of Pressure Transient, Rate Transient Analysis, Formation Testing, Production Logging and the use of Permanent Downhole*, volume v5.30. Kappa Engineering, Paris.
- Houze, O.P., Allain, O.F., Josso, B., et al. (2011). New methods enhance the processing of permanent-gauge data. In *SPE Middle East Oil and Gas Show and Conference*. Society of Petroleum Engineers.
- Khanna, A., Pascual-Leone, A., and Farzan, F. (2014). Reliability of resting-state microstate features in electroencephalography. *PLoS one*, 9(12), e114163.
- Kodinariya, T.M. and Makwana, P.R. (2013). Review on determining number of cluster in k-means clustering. *International Journal*, 1(6), 90–95.
- Lu, C., Shi, Y., Chen, Y., Bao, S., and Tang, L. (2016). Data mining applied to oil well using k-means and dbSCAN. In *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, 37–40. IEEE.
- Olsen, S., Nordtvedt, J.E., et al. (2005). Automatic filtering and monitoring of real-time reservoir and production data. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.
- Rai, H., Horne, R.N., et al. (2007). Analyzing simultaneous rate and pressure data from permanent downhole gauges. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.
- Savitzky, A. and Golay, M.J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8), 1627–1639.
- Suzuki, S., Chorneyko, D.M., et al. (2009). Automatic detection of pressure-buildup intervals from permanent downhole pressure data using filter convolution. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.
- Suzuki, S. et al. (2018). Using similarity-based pattern detection to automate pressure transient analysis. In *Abu Dhabi International Petroleum Exhibition & Conference*. Society of Petroleum Engineers.
- Viberti, D., Verga, F., Delbosco, P.F., et al. (2007). An improved treatment of long-term pressure data for capturing information. *SPE Reservoir Evaluation & Engineering*, 10(04), 359–366.