

Identificação de Estáticas em Poços de Petróleo utilizando *Motifs*

Daniel Folador Rossi,¹ Karin Satie Komati¹, Mateus Conrad Barcellos da Costa¹

¹ Programa de Pós-Graduação em Computação Aplicada (PPComp)
Instituto Federal do Espírito Santo (IFES), Campus Serra
Avenida dos Sabiás, 330 - Morada de Laranjeiras – 29166-630 – Serra – ES – Brazil

dan.folador@gmail.com, {kkomati,mcosta}@ifes.edu.br)

Abstract. *The identification of build-ups is a crucial step in the analysis of formation damage in oil wells, which can be considered a laborious and error-prone process. This article discusses an approach to this process based on the search for Motifs in time series of pressure data. Three distinct methods were evaluated in the study: a pure Motifs detection algorithm and combinations of the Motifs detection algorithm with the Derivative method and the Convolution Filter method. Tests were performed on ten sets of synthetic data. The method that achieved the best performance in the $F_{0.5}$ metric was the Motifs detection algorithm combined with the derivative method.*

Resumo. *A identificação de estáticas é uma etapa da análise de dano de formação em poços de petróleo considerada trabalhosa e suscetível a erros. Neste artigo é discutida uma abordagem para esta etapa baseada na busca por Motifs em séries temporais de dados de pressão de fundo de poço. No estudo foram avaliados três métodos distintos: um algoritmo de detecção de Motifs puro e combinações do algoritmo de detecção de Motifs com o método da Derivada e com o método do Filtro de Convolução. Os testes foram realizados com dez conjuntos de dados sintéticos. O método que obteve o melhor desempenho na métrica $F_{0.5}$ foi o algoritmo de detecção de Motifs combinado com o método da derivada.*

1. Introdução

Ao longo da sua vida operacional, um poço de petróleo pode sofrer dano de formação, que é qualquer processo físico, químico ou mecânico que cause a redução na produtividade de um reservatório [Bennion et al. 2002]. Neste sentido, a Gestão de Dano de Formação é uma atividade fundamental para garantir o bom desempenho dos poços [Civan 2016]. Uma das formas de se avaliar o dano de formação é por meio da análise de transientes de pressão, e o principal equipamento utilizado para obtenção de dados de pressão é denominado de PDG — *Permanent Downhole Gauge* (medidor de fundo de poço) [Houze et al. 2020].

Na análise de transientes, pode-se analisar tanto os períodos de fluxo quanto os de estática. É preferível a utilização de estáticas por apresentarem menos ruído e menos incerteza na vazão [Houze et al. 2020]. Um período de estática é caracterizado por uma subida abrupta de pressão de fundo. A Figura 1 apresenta um exemplo didático de um gráfico 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 há sete períodos de transientes, sendo dois deles períodos de estáticas, marcados pelo fundo hachurado em amarelo. 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. 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. Tipicamente, não há medidores de vazão disponíveis por poço.

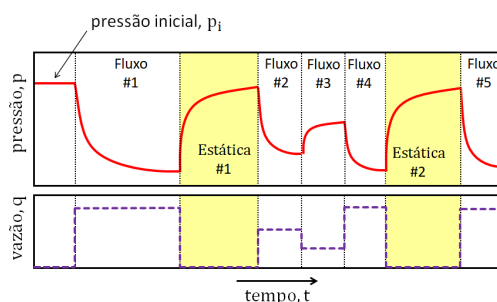


Figura 1. Desenho esquemático do comportamento de pressão de fundo (p) em um poço de petróleo com a variação de vazão (q).

Segundo dados da ANP referente a 2021, o Brasil contava com um total 6.954 poços de petróleo, sendo 6.191 em terra e 763 em mar [ANP 2022]. Atualmente, a metodologia de gerenciamento de dano realizado por unidades de avaliação de reservatórios da Petrobras não utiliza técnicas automatizadas. Os métodos baseiam-se em técnicas típicas de interpretação de transientes e indicadores propostos por alguns autores [Proano et al. 2004, Haddad et al. 2004, Holicek et al. 2008].

De acordo com o departamento de Avaliação de Reservatórios de Águas Profundas da Petrobras, estima-se que um estudo completo de análise de dano para um único poço leva, em média, de dois a três dias, sendo que a detecção de estáticas ocupa uma grande fatia deste tempo. Com um método de detecção automatizado e disponível, pode-se detectar as estáticas necessárias em questão de minutos, permitindo que o especialista direcione seu tempo para tarefas mais nobres, como a interpretação dos dados.

Ao longo dos anos, foram propostas várias metodologias com a finalidade de automatizar total ou parcialmente a interpretação de transientes de pressão [Athichanagorn et al. 1999, Olsen et al. 2005, Fair et al. 2014, Suzuki et al. 2018, Chang et al. 2019]. Entretanto, apesar das contribuições destes trabalhos, as metodologias apresentaram limitações na sua utilização prática. As técnicas de identificação de estáticas necessitam da intervenção do especialista para determinar limiares ou descartar falsos positivos, as técnicas de reconstrução de vazões possuem alto custo computacional, e as técnicas de identificação de modelos necessitam do auxílio do especialista, seja na detecção de assinaturas, seja na estimativa inicial de parâmetros.

O presente trabalho propõe uma solução para automatizar a etapa de identificação de estáticas utilizando *motifs*. *Motifs* são grupos de séries temporais (ou segmentos de uma série maior) que são muito similares entre si [Lin et al. 2002]. Comparamos três métodos, o primeiro método utiliza apenas o algoritmo de detecção de *motifs*, enquanto o

segundo e o terceiro métodos combinam o algoritmo com o método da derivada primeira [Chang et al. 2019] e o método do filtro de convolução [Suzuki et al. 2018], respectivamente. Testamos esses métodos em dez conjuntos de dados sintéticos, usando a métrica $F_{0.5}$ para comparação quantitativa.

O restante do trabalho está organizado da seguinte forma: na Seção 2 são apresentados os trabalhos relacionados com a identificação de estáticas; na Seção 3 são detalhados o conjunto de dados e os métodos usados nos experimentos; a Seção 4 apresenta os resultados os experimentos, e as discussões associadas, e a Seção 5 suas conclusões e trabalhos futuros.

2. Trabalhos Correlatos

Embora a identificação de estáticas por meio da análise de dados de pressão do fundo possa ser considerado um problema de reconhecimento de padrões, a maioria dos métodos propostos para este fim não opera automaticamente, sendo necessário a intervenção do especialista para determinar limiares ou descartar falsos positivos. Nesta seção são discutidos brevemente trabalhos que ilustram este problema, em particular com o uso de técnicas de Derivada e Convolução. Durante a pesquisa bibliográfica realizada não foram encontrados trabalhos que utilizem *motifs* para detecção de estáticas. Sendo assim nesta seção são apresentados trabalhos em que o conceito de Motif é utilizado com objetivos semelhantes no contexto de reconhecimento de padrões.

2.1. Identificação de estáticas usando derivada

[Chang et al. 2019] utilizam diretamente a derivada dos dados de pressão pelo tempo, 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. O método aplica um limiar na derivada para isolar as variações altas positivas. Para agrupar os pontos acima deste limiar, utilizaram o algoritmo de clusterização K-Means [Kodinariya and Makwana 2013], um método de aprendizado de máquina não supervisionado. Cada agrupamento de pontos é considerado como uma estática, e o início desta estática é identificada pelo máximo local do grupo.

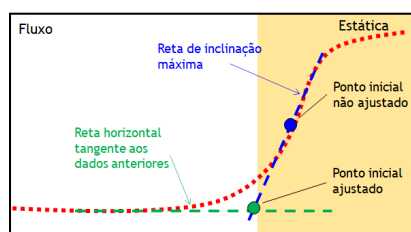


Figura 2. Método de ajuste fino do início do transiente elaborado pelo autor feito de acordo com a proposta de [Chang et al. 2019].

Para o ajuste fino do ponto inicial, os autores 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. A Figura 2 ilustra este ajuste, os dados de pressão são indicados pelos pontos vermelhos e a posição do máximo da derivada (ponto não ajustado) é indicado pelo ponto azul. Traça-se então uma reta de inclinação máxima (linha

tracejada azul) e uma reta horizontal (linha tracejada verde). O ponto de interseção das retas (ponto verde) é o ponto ajustado de início da estática. Os autores testaram a técnica para duas bases de dados reais, ambas com cinco anos de duração, tendo sido identificadas 80 estáticas em cada. Segundo os autores a técnica ainda seleciona alguns falsos positivos, mas que são descartados em seu próximo passo, que utiliza a similaridade de imagens nos gráficos especializados. Os autores concluem que a técnica auxilia na automatização da análise de transientes de pressão.

Uma limitação do método é a definição do valor limiar, que deve ser estimado pelo especialista, e a definição do número k de *clusters* para o algoritmo K-Means. Para contornar essa dificuldade, [Rossi et al. 2021] sugerem a substituição do K-means pela clusterização DBSCAN, que calcula automaticamente o número de *clusters* com base em dois parâmetros: o número mínimo de pontos de cada *cluster*, e a distância ϵ usada para considerar um dado como pertencente a um *cluster* ou como ruído.

2.2. Identificação de estáticas usando convolução

[Suzuki et al. 2009] propuseram o uso de um filtro de convolução para identificar estáticas. A proposta inicia com a definição da função pulso $F(t)$ que indica quando a derivada fica positiva (acima de um certo limiar de derivada, λ_{der}) e o início da estática ocorre quando $F(t)$ muda de zero para um. Em seguida, é aplicada uma função filtro $G(t)$ normalizada à função pulso, sendo w metade do tamanho de $G(t)$.

$$F(t) = \begin{cases} 1, & \text{caso } \frac{dp}{dt} > \lambda_{der} \\ 0, & \text{caso contrário} \end{cases}, \quad G(t) = \begin{cases} \frac{2}{w^2}t + \frac{2}{w}, & -w \geq t \geq 0 \\ \frac{2}{w^2}t - \frac{2}{w}, & 0 \geq t \geq +w \end{cases} \quad (1)$$

$$F * G = \int_{-\infty}^{\infty} F(\tau)G(t - \tau) d\tau \quad (2)$$

Calcular a convolução $F * G$ 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 . Valores positivos de $F * G$ indicam aumento na frequência, e quanto maior o valor absoluto $F * G$, mais abrupta é a mudança na derivada. Ao final, aplica-se um limiar aos dados da convolução, os dados abaixo do limiar são descartados e os dados acima são agrupados. O valor do limiar depende dos dados e deve ser estimado caso a caso pelo especialista. Os autores testaram a técnica para os dados de pressão de três poços, variando de 2.200 a 3.300 horas. Os dados foram tratados para ruído e espúrios pelo aplicativo Saphir, da KAPPA¹, sendo em seguida suavizados pela média móvel antes da aplicação do filtro e do limiar. Segundo os autores, os transientes foram identificados de forma satisfatória.

Posteriormente, [Suzuki et al. 2018] apresentaram um aprimoramento, substituindo a média móvel pelo filtro de mediana de Hampel [Hampel 1974]. O método foi testado para seis casos de dados reais, com duração variando de um ano a 4,5 anos. A técnica conseguiu detectar estáticas representativas de todos os casos, removendo automaticamente aquelas que não seriam adequadas para análise. Algumas estáticas tiveram que ser removidas em uma etapa de pós-tratamento.

¹<https://www.kappaeng.com/software/saphir/overview>

Os autores utilizaram o algoritmo de busca de *motifs* neste trabalho, não na etapa de detecção de estáticas, mas numa etapa posterior da interpretação de transientes, durante o ajuste de modelos de interpretação.

2.3. *Motifs*

Motifs são pares de séries temporais individuais, ou subsequências de uma série temporal mais longa, que são muito semelhantes entre si [Lin et al. 2002]. A abordagem mais simples, na “força-bruta”, compara todas as distâncias de todos os pares de cada subsequência dentro de uma série temporal. Para uma série temporal com comprimento n e um comprimento de comparação de subsequência m , a complexidade computacional do algoritmo de detecção é de $O(n^2m)$.

No artigo de [Mueen et al. 2009], é feita a proposta do algoritmo Mueen-Keogh (MK). A ideia intuitiva apresenta um conjunto de objetos 2D que são arranjados em uma ordenação 1D a partir de um objeto de referência. As distâncias entre pares adjacentes ao longo da projeção linear é um limite inferior para a distância real entre eles. O algoritmo usa essa ordenação linear de dados como heurísticas, a observação é que se dois objetos estão próximos no espaço original, eles também devem estar próximos na ordenação linear. O inverso não é verdadeiro, dois objetos podem estar arbitrariamente próximos na ordenação linear, mas muito distantes no espaço original. Com essa heurística, o algoritmo realiza uma poda e não avalia todas as possibilidades, mas as que se apresentarem promissoras na ordenação linear. O algoritmo se mostra muito sensível com o ponto de referência selecionado, e portanto são usados mais de um ponto de referência. A complexidade final no melhor caso é $O(R)$, onde R é a quantidade de pontos de referência e $O(m^2)$ para o pior caso.

Posteriormente, [Yeh et al. 2016] aprimorou a proposta do algoritmo MK, gerando a proposta STAMP. Uma das mudanças foi que o algoritmo de busca de similaridade usa distância euclidiana normalizada como uma sub-rotina, explorando a sobreposição entre subsequências usando o algoritmo FFT (Fast Fourier Transform). A complexidade final do STAMP é linear e além disso, é paralelizável. [Law 2019] publicou a biblioteca STUMPY² que calcula eficientemente o “matrix profile”, que é um vetor que representa as distâncias entre todas as subsequências dentro de uma série temporal e seus vizinhos mais próximos [Yeh et al. 2016, Zhu et al. 2016]. A biblioteca fornece funções para identificação de *motifs*, busca por um *motif* específico, descoberta de anomalias, segmentação semântica e estimativa de densidade.

3. Materiais e Métodos

Esta seção é dividida em três partes: a descrição dos conjuntos de dados dos experimentos, os métodos usados e quais as métricas de avaliação para comparação dos resultados.

3.1. Base de Dados

Nos experimentos foram usados os mesmos conjuntos de dados do artigo de [Rossi et al. 2021]³. São dez conjuntos de dados de pressão e vazão que foram gerados usando o *software* comercial de interpretação de transientes de pressão: o aplicativo

²<https://github.com/TDAmeritrade/stumpy>

³disponível em: https://github.com/danfolador/SBAI2021_dataset

Saphir, da KAPPA, o mesmo do artigo de [Suzuki et al. 2009]. As propriedades do reservatório e fluido utilizadas na simulação são apresentadas na Tabela 1(a). As vazões foram geradas por uma distribuição normal, definidas pela vazão média e o desvio padrão conforme Tabela 1(b). Para cada base de dados variou-se a permeabilidade da rocha. Os valores de ambas as tabelas foram obtidos do conhecimento do especialista.

Tabela 1. Propriedades de reservatório e fluido utilizados na simulação dos dados de pressão e vazão.

(a)		(b)				
Propriedade do reservatório		Base de dados	Vazão média (m ³ /d)	Desvio padrão da vazão (m ³ /d)	Permeabilidade simulada (mD)	Número de picos de pressão
Límites do reservatório	Infinito	BD-01	500	50	500	1
Pressão inicial	351,535 kgf/cm ²	BD-02	1000	100	500	2
Espessura do reservatório	30,0 m	BD-03	1500	150	500	1
Raio do poço	0,108 m	BD-04	2500	250	500	2
Porosidade da rocha	25%	BD-05	3000	300	500	1
Propriedade do fluido		BD-06	500	50	1000	2
Fluido produzido	Apenas óleo	BD-07	1000	100	1000	0
Coefficiente de estocagem	0,05 (m ³)/(kgf/cm ²)	BD-08	1500	150	1000	0
Densidade do óleo	28°API	BD-09	2500	250	1000	1
Fator volume-formação	1,2755 m ³ /m ³	BD-10	3000	300	1000	5
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 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 [-0,02; 0,02] kgf/cm² (baseado em conhecimento do especialista). Em algumas bases foram incluídos picos de pressão (variações abruptas e curtas), simulando perturbações nas válvulas do poço, conforme última coluna da Tabela 1(b). Assim, a quantidade de transientes totais na base BD-01 é de 66 (65 gerados mais um pico de pressão incluído), na base de BD-02 é de 67 e assim por diante.

3.2. Métodos

Foram desenvolvidos três métodos: (1) o método de *motifs* puro, (2) o método de *motifs* combinado com a derivada [Chang et al. 2019] e (3) o método de *motifs* combinado com a convolução [Suzuki et al. 2018]. Depois de identificadas as estáticas pelos métodos, o ponto inicial é ajustado pela técnica das tangentes (Figura 2 à direita). Todas as bases e *motifs* foram interpolados linearmente para obter espaçamento uniforme de cinco segundos. Como o STUMPY compara o *motif* com as pressões ponto a ponto, a interpolação auxilia a preencher vazios de dados que distorceriam as respostas. As bases foram previamente tratadas para ruído com a média móvel e uma janela de dez pontos, o mesmo sendo aplicado aos *motifs*.

3.2.1. Identificação de *motifs*

Para este método foi usada a biblioteca STUMPY. A biblioteca fornece funções para detectar *motifs* recorrentes e desconhecidos numa série temporal, como também funções para identificar padrões previamente determinados. Como já se sabe qual é a característica de uma estática, já se conhece previamente o padrão do *motif*.

Para isso, foram utilizados dois *motifs*, apresentados na Figura 3. O *motif1* é a estática #1 da base BD-04 e o *motif2* é estática #3 da base BD-06. Ambos os *motifs*

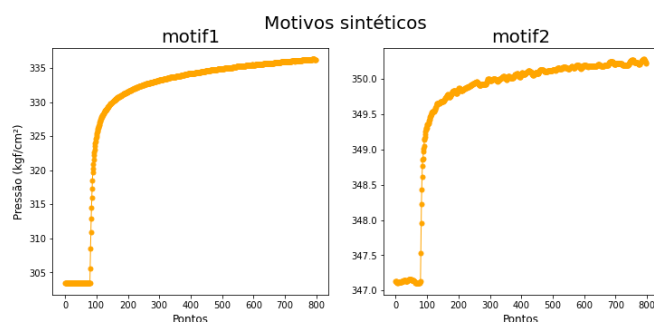


Figura 3. Motifs sintéticos utilizados para a busca de estáticas.

foram interpolados para espaçamento de cinco segundos, possuindo pouco mais que cinco minutos antes do início da estática e um hora depois, totalizando 800 pontos. O algoritmo buscará os três trechos nos dados de pressão que apresentem a maior similaridade com o *motif* usado (o STUMPY utiliza a distância euclidiana como métrica de similaridade).

3.2.2. Combinação de identificação de *motifs* com derivada

Para combinar o método de *motifs* com a técnica da derivada usada por [Chang et al. 2019], utilizou-se a curva de derivada positiva como um “peso” com o qual a métrica de similaridade do STUMPY será ponderada. Para cada *motif* encontrado, divide-se sua similaridade pela maior derivada positiva naquele trecho. Para limitar os valores, a curva de derivada positiva foi normalizada para ficar entre zero e um.

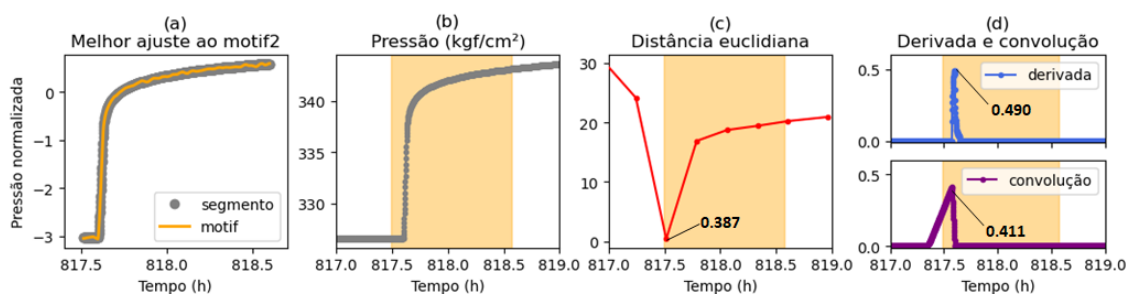


Figura 4. Método dos *motifs* combinado com a derivada e a convolução.

A Figura 4 ilustra este método. Na Figura 4(a), temos um dos segmentos da BD-10 que apresentou a maior similaridade com o *motif2*. O *motif2* está representado na linha laranja sobreposta ao segmento (pontos cinza). Na Figura 4(b) é mostrado o trecho dos dados (pontos cinza) de onde o segmento foi retirado, com o local do ajuste marcado pelo fundo amarelo. Na Figura 4(c) temos as distâncias euclidianas (linha vermelha) para vários ajustes neste trecho de dados. O segmento analisado é o que tem a menor distância (maior similaridade), indicado pelo mínimo em 0,387. Na Figura 4(d) (superior, curva azul) temos as derivadas para o mesmo trecho, com um pico em 0,490, a convolução (inferior, curva roxa), com um pico em 0,411, sendo o local do ajuste marcado pelo fundo amarelo. Para combinar o método do *motif* com a derivada, dividiremos a distância do ajuste (0,387) com a maior derivada na região de ajuste (0,490). Assim, privilegiamos os ajustes com as maiores derivadas.

3.2.3. Combinação de identificação de *motifs* com convolução

De forma análoga à combinação com o método da derivada, para combinar com a identificação de estáticas pelo operador de convolução proposto por [Suzuki et al. 2018], utilizou-se a curva de convolução para ponderar os valores de similaridade do STUMPY. Para cada *motif* encontrado, divide-se sua similaridade pelo maior valor de convolução positiva naquele trecho. A convolução positiva se encontra entre zero e um.

Como exemplo, tomemos a Figura 4(d). (inferior, curva roxa), onde temos as convoluções para o trecho de dados da Figura 4 (já comentado na seção anterior). Para combinar o método do *motif* com a convolução, dividiremos a distância do ajuste (0.387) com a maior convolução na região de ajuste (0.411). Foi utilizado um limiar de derivada normalizada igual a $\lambda_{der} = 0,05$ e uma meia-janela de convolução de $w = 15\text{min}$. Assim, privilegiamos os ajustes com as maiores valores de convolução.

3.3. Métricas de avaliação

A tarefa de identificar inícios de estáticas pode ser tratado como um problema de detecção de pontos de mudança (*change point detection*) em séries temporais, sendo o início da estática o ponto a ser detectado. Segundo [van den Burg and Williams 2020], o número de “pontos de mudança” costumam ser muito menores do que o número de pontos da série temporal, resultando num problema com classes desbalanceadas (a classe positiva é muito menor do que a classe negativa). Para estes casos, pode-se utilizar a métrica F_β . Como neste trabalho deseja-se priorizar a *precisão*, será utilizado $\beta = 0.5$, sendo a $F_{0.5}$ a principal métrica de avaliação. Considera-se que os verdadeiros positivos são as estáticas identificadas corretamente; os falsos positivos são os pontos que não são inícios de estáticas classificados como estáticas; os verdadeiros negativos são os pontos que não são inícios de estáticas classificados corretamente e os falsos negativos são os inícios de estáticas não identificados como tal.

Ao se utilizar métricas de classificação para este tipo de problema, é comum permitir que a resposta varie dentro de uma margem de tolerância [van den Burg and Williams 2020]. Entretanto, deve-se evitar uma múltipla contagem, de modo que se vários pontos forem identificados próximos do ponto verdadeiro, apenas um seja contado. Para este trabalho, foi utilizada uma margem de tolerância de $\delta = 0,5h$ ao redor do ponto verdadeiro na comparação dos métodos. O primeiro ponto predito que esteja situado dentro da margem de tolerância será considerado um verdadeiro positivo e os demais que se situarem dentro da margem serão descartados.

4. Resultados e Discussão

Foram realizados 60 experimentos no total, os três métodos (*motif*, *motif* com derivada e *motif* com convolução) com dois modelos de estáticas (*motif1* e *motif2*), sobre dez bases de dados. Os resultados foram consolidados na Tabela 2, com os métodos em cada linha e as bases em cada coluna. As cores de cada coluna variam entre vermelho (menor valor), amarelo (valor mediano) e verde (melhor valor). A última coluna indica a classificação do método, sendo #1 o melhor método e #6 o pior método.

Conforme tabela, os métodos que obtiveram o melhor resultado na identificação de estáticas foram os métodos de *motif* combinado com a derivada, que ficaram em 1º e

Tabela 2. Resultados dos experimentos para a métrica F0.5.

Métodos	Resultados para F0.5										Méd	Classif.
	Bases											
	BD-01	BD-02	BD-03	BD-04	BD-05	BD-06	BD-07	BD-08	BD-09	BD-10		
Motif1	0,67	1,00	0,67	1,00	0,67	0,67	0,33	0,67	0,67	0,00	0,63	#5
Motif2	0,67	0,33	0,00	0,00	0,00	1,00	0,67	0,33	1,00	0,67	0,47	#6
Motif1 + derivada	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,67	1,00	0,97	#1
Motif2 + derivada	1,00	1,00	1,00	1,00	0,67	1,00	1,00	0,67	0,67	1,00	0,90	#2
Motif1 + convolução	1,00	1,00	0,67	1,00	0,67	1,00	0,33	1,00	1,00	1,00	0,87	#3
Motif2 + convolução	0,67	1,00	0,67	1,00	0,67	1,00	0,67	0,67	1,00	0,67	0,80	#4

2º lugares. Sendo que o *motif1* obteve resultados melhores do que o *motif2*, em qualquer combinação, o que leva a se considerar que a forma do *motif1* modela melhor uma estática. Considera-se também que os métodos são sensíveis ao modelo de *motif* que é usado.

A Figura 5 ilustra a aplicação dos métodos e seus resultados para a base BD-08. Na primeira linha de gráficos temos: (a) os dados de pressão, em cinza; (b) as distâncias euclidianas em vermelho, (c) a derivada normalizada, em azul, e (d) a convolução em roxo. Nos gráficos seguintes temos os dados pressão, em cinza, e as estáticas identificadas pelos métodos, marcadas em laranja, verde e azul. Na segunda linha de gráficos temos: (e) os resultados para o *motif1*; (f) para o *motif1* com a derivada; e (g) para o *motif1* com a convolução. Na terceira linha de gráficos temos: (h) os resultados para o *motif2*; (i) para o *motif2* com a derivada; e (j) para o *motif2* com a convolução. A base BD-08 contém três estáticas, sendo que os resultados (f) e (g) apresentam identificação correta das três. Em (e), a estática marcada em verde está incorreta, embora seja um pico de pressão, não chega a ser uma estática. Os resultados (e) e (f) apresentam o mesmo erro. E (h) identifica incorretamente estáticas em amarelo e em verde.

Os métodos que obtiveram o pior desempenho foram os que utilizaram apenas os *motifs* sem combinação. O bom desempenho dos métodos combinados se explica porque, enquanto a busca de *motifs* se concentra na *forma* a ser buscada, a derivada e convolução indicam *regiões mais prováveis* de se encontrar um estática, aumentando a eficácia da busca. Apesar da convolução apresentar resultados piores que a derivada, é possível que possa ser melhorado ao se ajustar seus parâmetros (o limiar da derivada λ_{der} e a janela de convolução w). A otimização destes parâmetros pode ser buscada em trabalhos futuros. Ao mesmo tempo, o fato do método da derivada não possuir parâmetros de entrada é uma vantagem perante à convolução.

Por se tratarem de bases sintéticas, a assinatura de pressão durante o fluxo apresenta-se mais “comportada” do que um fluxo real, isto é, sem ruído excessivo e sem grandes perturbações de pressão. Isso fez com que muitos períodos de fluxo se parecessem com estáticas, o que impactou na busca do algoritmo. Acredita-se que, para bases reais, onde o comportamento das estáticas dificilmente pode ser confundido com fluxos, o método dos *motifs* ponderado apresente resultados ainda melhores. Vale ressaltar, também, que o número de estáticas buscadas é um parâmetro de entrada do método e, se for buscado um número maior de estáticas do que as existentes na base, teremos mais *falsos positivos*, que devem ser descartados em algum pós-tratamento.

Buscou-se fazer uma comparação dos resultados com outros métodos de identificação de estáticas [Athichanagorn et al. 1999, Viberti et al. 2007, Houze et al. 2011, Suzuki et al. 2018, Chang et al. 2019], entretanto, embora os métodos

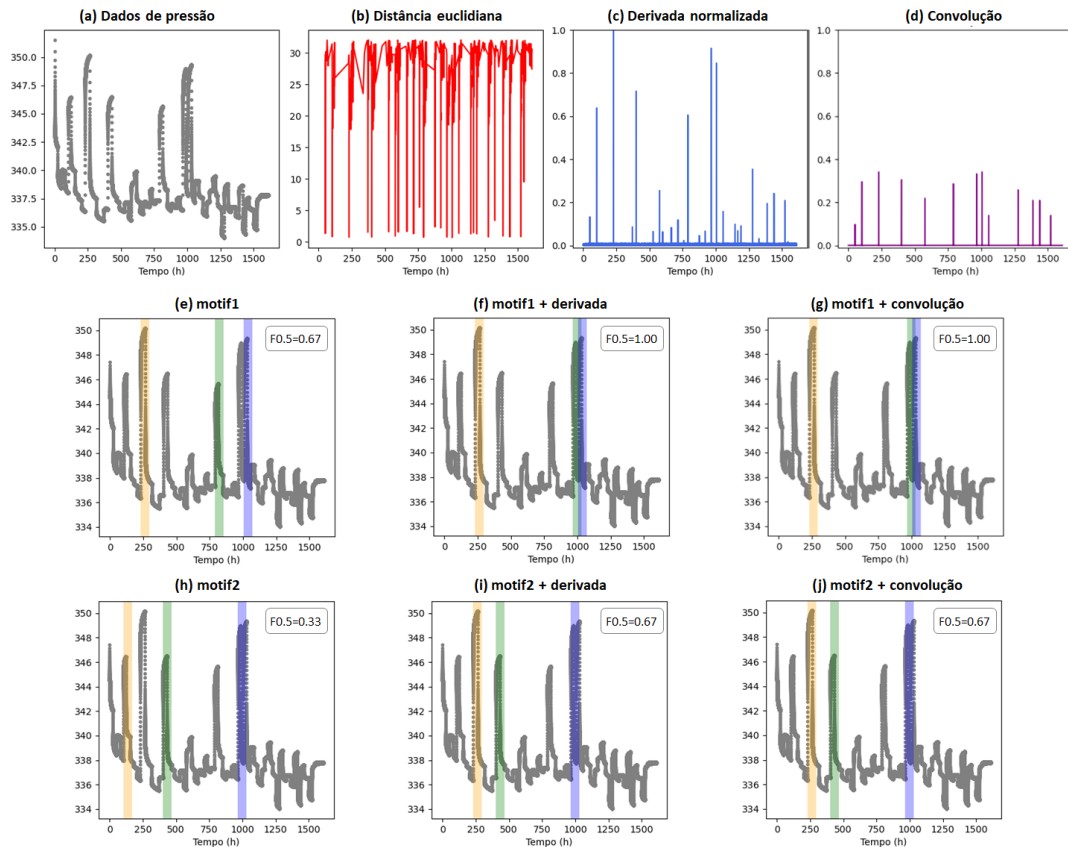


Figura 5. Resultados para a base BD-08.

apresentem resultados satisfatórios segundo seus autores, não foram apresentadas métricas que permitissem realizar a comparação.

No artigo de [Rossi et al. 2021] foram testados os métodos de [Chang et al. 2019] e [Suzuki et al. 2018] para as mesmas dez bases deste trabalho. Os resultados destes métodos, porém, dependem do limiar de derivada utilizado. Com o limiar adequado, todas as estáticas são detectadas, obtendo-se um $F_{0.5} = 1,00$. Entretanto, este limiar “ótimo” deve ser obtido manualmente pelo usuário após várias tentativas, para cada base de dados. Esse ajuste é uma limitação destes métodos, e também impede que seus resultados sejam comparados com os métodos de *motifs*. Os métodos com *motifs*, por sua vez, têm a vantagem de não depender da base de dados.

5. Conclusão

A automatização da análise de dano de formação em poços de petróleo é fundamental para a efetiva gestão de dano e a consequente garantia de maior produtividade desses poços. Neste trabalho foi proposta uma solução para automatizar a etapa de identificação de estáticas na análise de transientes de pressão usando a busca de *motifs*. Esta técnica foi combinada com dois outros métodos, o método da derivada de pressão e o método do filtro de convolução. As três combinações foram testadas para dez bases de dados sintéticos, e o método que obteve o melhor resultado foi o método de *motifs* combinado com a derivada. Conclui-se, então, que o método de *motifs* ponderado se mostrou uma técnica bastante promissora na detecção de estáticas, além de não necessitar da intervenção do

especialista na determinação de limiares, como acontece com os métodos originais.

Para trabalhos futuros, deseja-se testar o desempenho das técnicas de *motifs* para bases reais rotuladas, testar o impacto do *motif* escolhido nos resultados, e testar o impacto dos parâmetros de cada método. Pretende-se aplicar os métodos de outras propostas aos dados deste trabalho para comparação de resultados, tal como o método de similaridade de imagens proposto por [Olsen et al. 2005].

6. Agradecimentos

Os autores agradecem à FAPES e CAPES pelo PDPG (Programa de Desenvolvimento de Pós-Graduação - Parcerias Estratégicas nos Estados, processo 2021-2S6CD, FAPES nº132/2021). A segunda autora agradece ao CNPq pela Bolsa de Produtividade DT-2 (308432/2020-7) e pelo projeto 407742/2022-0, agradece à FAPES pelo Auxílio Taxa de Pesquisa (nº 293/2021) e pelo projeto nº1023/2022 P:2022-8TZV6.

Referências

- ANP (2022). Anuário estatístico brasileiro do petróleo, gás natural e biocombustíveis 2022.
- 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).
- 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. (2016). *Reservoir formation damage: fundamentals, modeling, assessment, and mitigation*. Gulf Professional Publishing, an imprint of Elsevier.
- Fair, C., Flores, R., Hakim, B., Nande, S., et al. (2014). Using the results from automated petroleum engineering calculations to accelerate decision workflows. In *SPE Asia Pacific Oil & Gas Conference and Exhibition*. Society of Petroleum Engineers.
- Haddad, S., Proano, E., Patel, Y., et al. (2004). A method to diagnose depletion, skin, and drive mechanism effects using reservoir monitoring data. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.
- Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393.
- Holicek, R., Adachi, J., Vilorio, L., Mese, A., and Traore, Y. (2008). Surveillance and diagnostics of permanent bottomhole gauge data coupled with geomechanical modeling to identify sources of formation damage. volume All Days of *U.S. Rock Mechanics/Geomechanics Symposium*. ARMA-08-322.
- 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.
- Kodinariya, T. M. and Makwana, P. R. (2013). Review on determining number of cluster in k-means clustering. *International Journal*, 1(6):90–95.
- Law, S. M. (2019). STUMPY: A Powerful and Scalable Python Library for Time Series Data Mining. *The Journal of Open Source Software*, 4(39):1504.
- Lin, J., Keogh, E., Lonardi, S., and Patel, P. (2002). Finding motifs in time series. In *Proceedings of Workshop on Temporal Data Mining, 2002*, pages 53–68.
- Mueen, A., Keogh, E., Zhu, Q., Cash, S., and Westover, B. (2009). Exact discovery of time series motifs. In *Proceedings of the 2009 SIAM international conference on data mining*, pages 473–484. SIAM.
- 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.
- Proano, E., Patel, Y., Thawer, R., Al-Harthy, S., Gadiyar, B., John, C., Morales, H., et al. (2004). Life-of-well approach for optimizing production-gom case histories. In *SPE International Symposium and Exhibition on Formation Damage Control*. Society of Petroleum Engineers.
- Rossi, D. F., Komati, K. S., and da Costa, M. C. B. (2021). Identificação de estáticas em poços de petróleo via sensor de pressão de fundo. In *Simpósio Brasileiro de Automação Inteligente-SBAI*, volume 1.
- 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.
- van den Burg, G. J. and Williams, C. K. (2020). An evaluation of change point detection algorithms. *arXiv preprint arXiv:2003.06222*.
- 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.
- Yeh, C.-C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, A., Silva, D., Mueen, A., and Keogh, E. (2016). Matrix profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. pages 1317–1322.
- Zhu, Y., Zimmerman, Z., Senobari, N. S., Yeh, C.-C. M., Funning, G., Mueen, A., Brisk, P., and Keogh, E. (2016). Matrix profile ii: Exploiting a novel algorithm and gpus to break the one hundred million barrier for time series motifs and joins. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 739–748.