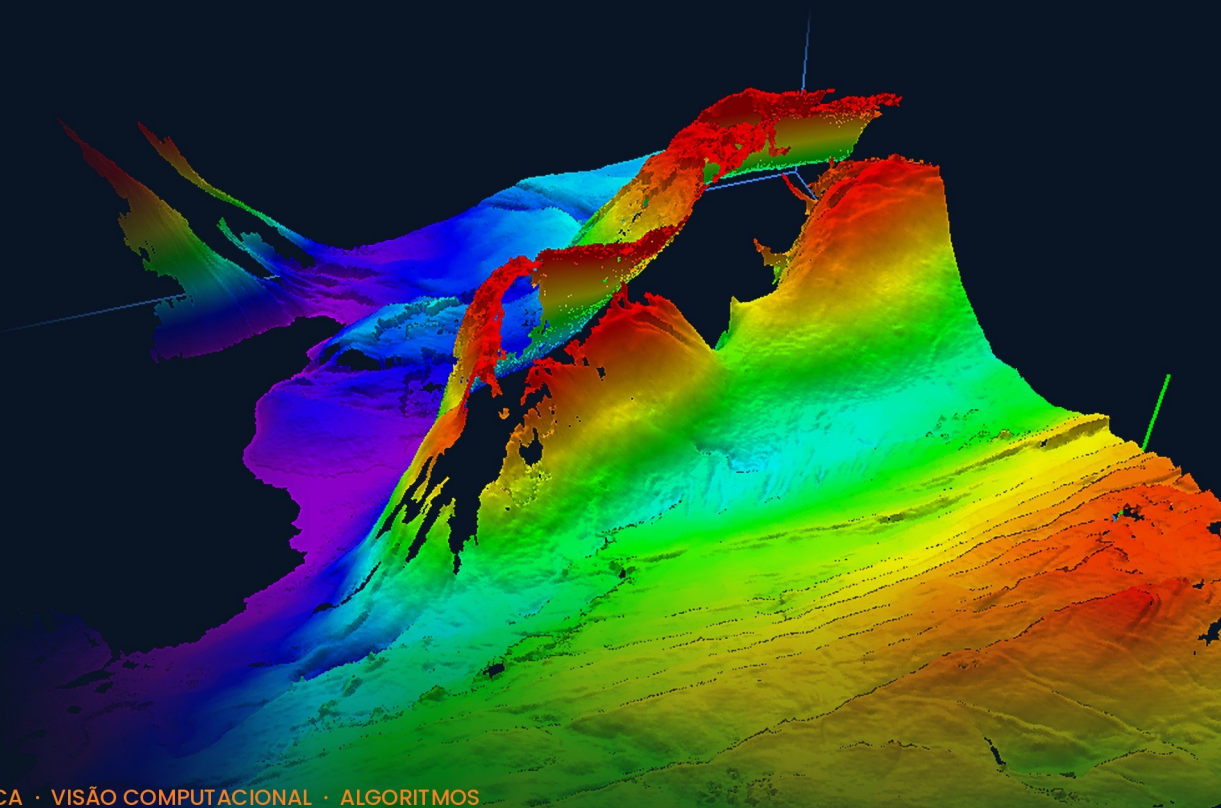


INTERPRETAÇÃO SÍSMICA · VISÃO COMPUTACIONAL · ALGORITMOS

99,9 % localmente certo — e porque *isso não chega*

O salto de ciclo silencioso no rastreio 3-D de horizontes sísmicos: referência de comparação, ordem de crescimento e localidade de acesso

99,9 %
células em extremos locais — numa superfície suspeita

511 / 276
ms: intervalo da superfície contra o das marcações

600 → 3 s
custo de corrigir a ordem, recuperado pela fila de baldes

91,5 → 38 %
cobertura ao comparar com a semente e não com o progenitor

Edmilson Delfim Praia
COPIA Seis

Luanda · Outubro de 2026

Dados: levantamento público F3, Mar do Norte (dGB Earth Sciences)

ARTIGO DE INVESTIGAÇÃO Interpretação sísmica · Visão computacional · Algoritmos

99,9 % localmente certo — e porque isso não chega

O salto de ciclo silencioso no rastreio 3-D de horizontes sísmicos: referência de comparação, ordem de crescimento e localidade de acesso

Edmilson Delfim Praia
COPIA Seis, Luanda, Angola

Resumo — O rastreio automático de horizontes a partir de sementes falha, sobretudo, por salto de ciclo: a superfície resultante é suave e plausível, e cada célula satisfaz os critérios locais de aceitação. Sobre o levantamento público F3 (619 101 traços), uma superfície com **99,9 %** das células em extremos locais da polaridade correcta abrangia **511 ms**, contra **276 ms** nas 43 marcações de origem, todas numa inline. A superfície é suspeita, não refutada — um horizonte que mergulhe fora dessa linha pode abranger legitimamente mais tempo —, mas o ponto central não depende disso: **os testes locais não distinguem as duas hipóteses**. Modelamos a deriva como um termo sistemático proporcional ao comprimento do caminho mais um passeio aleatório na árvore de crescimento, e verificamos quatro consequências no código e nos dados. (i) Comparar cada candidato com o traço da **semente**, e não com o do progenitor, retira o caminho da decisão de aceitação: a cobertura desceu de **91,5 % para 38,0 %** e o intervalo temporal de 511 para 396 ms. (ii) Uma faixa de tolerância em torno da tendência das marcações não recusa saltos de um laço quando a sua meia-largura excede meio período (≈ 16 ms no F3); a ±60 ms só reduziu a cobertura, e está desligada por omissão. (iii) A janela de correlação e a ordem de crescimento controlam os degraus anómalos (> 40 ms): de 15 para 0 ao passar a janela de 4 para 12 amostras, e de 40 para 0 ao trocar a prioridade por amplitude pela prioridade por proximidade. (iv) Corrigir a ordem de crescimento levou o rastreio de um horizonte de 5 s para mais de 600 s; uma fila de baldes alinhada com os blocos do formato de armazenamento trouxe-o para **3,0 s** com a mesma aritmética. Identificamos a experiência que falta — a comparação com horizontes do F3 interpretados manualmente e publicados — e propomos, sem resultados, uma integração com aprendizagem automática.

Palavras-chave: *interpretação sísmica; rastreio de horizontes; salto de ciclo; tensor de estrutura; crescimento de região; localidade de referência; supervisão fraca.*

Abstract — Seeded horizon tracking fails chiefly through cycle skipping, a silent failure: the output surface is smooth and plausible and every cell passes the local acceptance tests. On the public F3 survey, a surface with 99.9 % of its cells at local extrema of the correct polarity spanned 511 ms where the 43 seed picks, all on one inline, spanned 276 ms. The surface is suspect, not refuted — a horizon dipping away from that line may legitimately span more time — but the central point does not depend on it: local tests cannot tell the two cases apart. Modelling drift as a systematic term proportional to path length plus a random walk on the growth tree, we verify four consequences in code and data: (i) comparing candidates with the seed trace rather than the parent removes the path from the acceptance decision, cutting coverage from 91.5 % to 38.0 % and span from 511 to 396 ms; (ii) a trend tolerance band cannot reject one-loop skips when wider than half a period (≈ 16 ms on F3), and at ±60 ms only reduced coverage; (iii) correlation window and growth order control anomalous steps (> 40 ms): 15 → 0 from 4 to 12 samples, 40 → 0 from amplitude-first to proximity-first; (iv) correcting growth order took full-horizon tracking from 5 s to over 600 s, and a brick-aligned bucket queue brought it to 3.0 s with the same arithmetic. The missing experiment — comparison with published, manually interpreted F3 horizons — is identified.

Keywords: *seismic interpretation; horizon tracking; cycle skipping; structure tensor; region growing; locality of reference; weak supervision.*

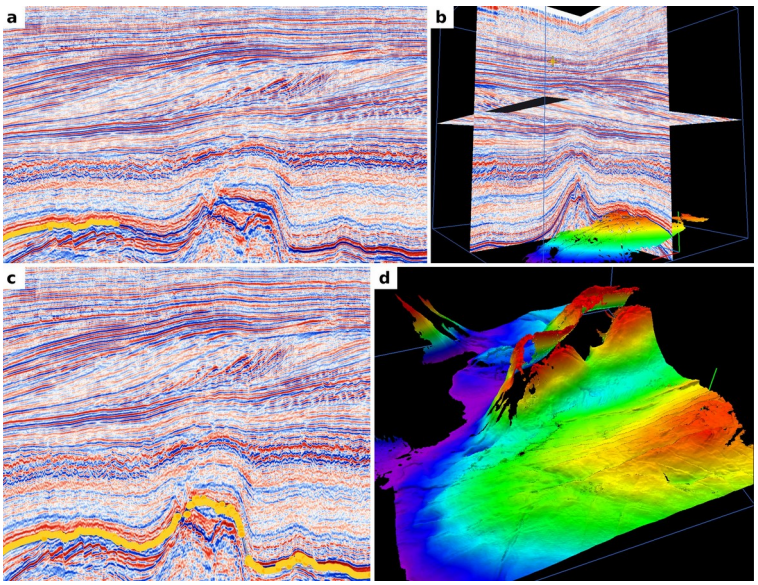


Figura 1. Rastreio semeado em modo volumétrico no F3. (a, b) Três sementes numa única secção: cobertura pequena, confinada ao flanco marcado. (c, d) Vinte e duas sementes distribuídas por várias secções, incluindo a crista e o flanco abrupto da estrutura central: a aplicação reporta **75,9 %** de cobertura. Escala de cores em (b, d): tempo duplo do horizonte. A cobertura é a reportada pela aplicação numa sessão interactiva, não uma experiência controlada. As lacunas em (d) são recusas deliberadas (falhas, baixa semelhança, viagem de volta). As paredes quase verticais junto à estrutura central podem ser rejeito real ou as lâminas da secção 4.6; distingui-las exige horizontes de referência (secção 7.2).

1 INTRODUÇÃO

Um levantamento sísmico tridimensional é uma função amostrada $a(i, c, s)$ de duas coordenadas de mapa — inline i e crossline c — e de uma coordenada vertical s em tempo duplo de percurso. Interpretá-lo começa por identificar **horizontes**: superfícies $z(i, c)$ que representam fronteiras cronoestratigráficas contínuas. No conjunto F3, usado ao longo do trabalho, isso significa 651 × 951 = 619 101 posições de mapa sobre 462 amostras, ou $2{,}86 \times 10^8$ amostras. A marcação manual é inviável; a prática corrente é o **rastreio a partir de sementes**, em que o intérprete marca poucos pontos e um algoritmo propaga a superfície (Hoyes e Cheret, 2011).

A dificuldade não está em encontrar um extremo local. Um traço sísmico é quase periódico, e o laço adjacente ao correcto assemelha-se fortemente a ele. Seguir o laço errado — o **salto de ciclo** — tem três propriedades que organizam este trabalho: é *localmente consistente* (a posição errada é um extremo genuíno, da polaridade correcta e bem correlacionado); é *cumulativo* (um deslizamento a cada duzentos traços é imperceptível em cada passo); e é *observável apenas globalmente*. Daqui decorre uma tese metodológica que atravessa o artigo: **a inspecção visual não é verificação**. Uma superfície errada é tão suave como uma correcta.

As contribuições são as seguintes:

(i) **Quantificação do modo de falha silencioso** e um modelo de deriva — termo sistemático mais passeio aleatório sobre a árvore de crescimento — que explica por que razão nenhum critério local o detecta (secções 3 e 6.1).

(ii) **Uma restrição não local que funciona e outra que não, ambas explicadas pelo modelo**: comparar com o traço da semente retira o caminho da decisão de aceitação; a tolerância à tendência não recusa saltos de um laço acima de meio período. A validação de ida e volta detecta saltos assimétricos (secções 3.2, 4.4 e 4.5).

(iii) **A ordem de crescimento como parâmetro do algoritmo**, com efeito medido sobre a geometria (secções 4.6 e 6.2).

(iv) **O custo de corrigir a ordem, e a sua recuperação**: a ordem correcta multiplicou o tempo por mais de 100; uma fila de baldes alinhada com a organização em blocos trouxe-o abaixo do original (secções 4.8 e 6.4).

(v) Contribuições de engenharia validadas — núcleos de Scharr com escala interna, barreira de falha em vista de mapa, custo convexo e corrigido pelo mergulho na marcação guiada — e uma **proposta**, explicitamente sem resultados, de integração com aprendizagem automática (secção 8).

(vi) **Uma delimitação explícita do que falta demonstrar** e da experiência que o resolve com dados públicos (secção 7.2).

2 TRABALHO RELACIONADO

Rastreio local. O seguimento de eventos por correlação normalizada entre traços adjacentes, organizado como crescimento de região a partir de sementes (Adams e Bischof, 1994), é a base dos rastreadores comerciais. A implementação de referência aberta foi lida como literatura e não reutilizada, por incompatibilidade de licença. A literatura descreve em detalhe os critérios de aceitação, mas raramente trata a *ordem* de visita como variável experimental, que é o objecto da secção 4.6.

Métodos globais. Uma linha alternativa evita a propagação passo a passo: o aplanamento sem marcação resolve um problema de mínimos quadrados sobre o campo de mergulho (Lomask *et al.*, 2006), os volumes de tempo geológico relativo atribuem uma idade a cada amostra (Stark, 2004) e as restrições do intérprete podem ser impostas ao campo global (Wu e Hale, 2015). Estes métodos não acumulam deriva da mesma forma, mas sacrificam a operação interactiva central — «este evento, a partir daqui» — e dependem da qualidade do campo de mergulho.

Orientação. O tensor de estrutura (Bigün e Granlund, 1987; Kass e Witkin, 1987) e os filtros de destruição de onda plana (Fomel, 2002) são as ferramentas correntes para estimar mergulho. A sua utilização como previsor do passo seguinte expõe sensibilidades — enviesamento do operador de derivada e aliasing em mergulhos acentuados — que a secção 4.3 quantifica.

Verificação de consistência. A validação de ida e volta é o análogo sísmico da verificação esquerda–direita em correspondência estereoscópica e do erro *forward–backward* no seguimento de objectos em vídeo (Kalal *et al.*, 2010). Não encontrámos a sua descrição explícita como critério de aceitação no rastreio de horizontes, mas não fizemos uma revisão sistemática dos rastreadores bidireccionais e comerciais, cuja documentação é frequentemente incompleta; a novidade deve ser lida com essa reserva.

Aprendizagem automática. As redes convolucionais concentraram-se na segmentação de falhas treinada em dados sintéticos (Wu *et al.*, 2019) e na regressão do tempo geológico relativo (Geng *et al.*, 2020), e o F3 é um conjunto de referência estabelecido (Alaudah *et al.*, 2019). Estas abordagens substituem o rastreador; a secção 8 propõe, ao contrário, informá-lo componente a componente.

3 FORMULAÇÃO

3.1 Problema

Seja Ω o conjunto das posições de mapa $x = (i, c)$ e $S = \{(x_k, z_k)\}$ o conjunto de sementes. Procura-se $z\colon \Omega' \to \mathbb{R}$, com $\Omega' \subseteq \Omega$ tão grande quanto possível, tal que **todas as posições determinadas pertençam ao mesmo evento geológico**. Sem esta cláusula qualquer extremo local serve; com ela, o problema é de identidade, não de detecção. Os índices verticais são fraccionários e todos os tempos são tempos duplos.

3.2 Deriva como passeio aleatório na árvore de crescimento

Um rastreador semeado constrói uma floresta: cada célula aceite y tem um progenitor x e herda dele a posição prevista. Seja $\pi(x)$ o caminho da semente até x, de comprimento $L(x)$, e $\Delta_x = z(x) - z^*(x)$ o erro em relação ao evento verdadeiro z^*. Decompomos o erro de cada passo em $\varepsilon_e = b + \eta_e$: um termo sistemático b — o enviesamento do previsor de mergulho — e um termo η_e de média nula (quantização residual, ruído). Admitindo b aproximadamente constante ao longo do caminho e η_e com correlação de curto alcance e variância de longo prazo σ^2, tem-se $\Delta_x = b\,L(x) + \Sigma\eta_e$ e, para duas células vizinhas,

$$E\left(\Delta_x - \Delta_y\right)^2 = b^2\left(L_x - L_y\right)^2 + \sigma^2\,\ell\left(x, y\right) \qquad (1)$$

em que $\ell(x, y)$ é o comprimento da diferença simétrica entre $\pi(x)$ e $\pi(y)$. Com correlação de curto alcance a variância continua a crescer linearmente com ℓ, sendo σ^2 a soma das autocovariâncias; só uma componente persistente, como b, muda a escala. A expressão (1) tem três consequências. Primeiro, o erro cresce com o comprimento do caminho — linearmente pelo termo sistemático, com a raiz pelo aleatório — e nenhum teste aplicado a um passo isolado observa essa soma: **a deriva é invisível localmente por construção**. Segundo, a descontinuidade entre vizinhos depende de $L_x - L_y$ e de ℓ, propriedades da *forma da árvore*, isto é, da ordem de visita, e não do critério de aceitação. Terceiro, uma restrição $|z(x) - T(x)| \le \tau$ em torno de uma tendência T com erro $e_T = T -$

z^* garante $|\Delta_x| \le \tau + |e_T(x)|$ e $|\Delta_x - \Delta_y| \le 2\tau + |e_T(x) - e_T(y)|$, seja qual for a ordem. **A restrição limita o erro relativamente à tendência, não o erro geológico**: se a tendência estiver errada, a superfície pode cumpri-la e seguir outro evento.

O salto de ciclo é o caso discreto do mesmo mecanismo: com probabilidade q por passo, η_e toma o valor de um período da ondícula, T_w, e o número esperado de saltos num caminho de comprimento L é qL. Reduzir q (secções 4.1, 4.4 e 4.5) atrasa a falha. A tolerância à tendência só recusa saltos isolados se $\tau < T_w/2$; com τ maior, trava apenas a deriva de vários laços. No F3, $T_w \approx 32$ ms, pelo que o limiar é $\tau \approx 16$ ms.

Previsão e decisão. O caminho entra no rastreador em dois sítios: na *previsão* da posição $\hat{z}$ e na *decisão* de aceitação, quando o candidato é comparado com o traço do progenitor. A segunda entrada é evitável: comparando com o traço da semente, a pontuação deixa de depender do comprimento do caminho, embora a posição dos candidatos continue a depender. É a quarta consequência de (1) — a restrição não local eficaz não é uma faixa em torno da tendência, mas uma referência de comparação fixa. Não elimina o salto de ciclo, porque o laço vizinho também se parece com a semente; impede que a referência derive com ele.

3.3 *Condicionamento da tendência*

A tendência T é interpolada a partir das marcações. Se estas estiverem numa única secção, T é extrapolada perpendicularmente sem informação, e qualquer restrição baseada nela perde significado. O mesmo vale para os diagnósticos: a razão R da secção 6.1 compara a superfície com marcações de uma só linha. A figura 1 mostra que a cobertura depende muito da distribuição das sementes — três numa linha contra vinte e duas em várias secções —, mas é uma sessão interactiva, não uma experiência controlada.

4 MÉTODO

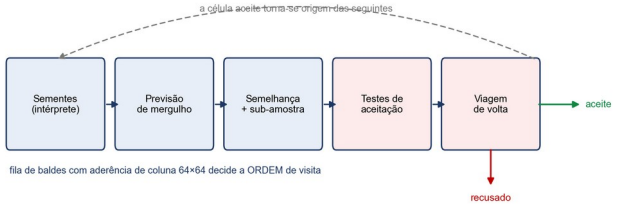


Figura 2. Fluxo de rastreio. A vermelho, os testes de aceitação; a seta tracejada indica que cada célula aceite se torna origem das seguintes. A fila de baldes com aderência de coluna determina a ordem de visita (secções 4.6 e 4.8).

O algoritmo 1 resume o rastreador; as subsecções seguintes justificam cada passo.

Algoritmo 1 Rastreio semeado com fila de baldes

```
entrada: a, S, barreira B, tendência T, τ, W
 1 Q ← fila de baldes; ∀(x,z)∈S: Q.insere(x,0)
 2 enquanto Q ≠ ∅:
 3   x ← Q.extrai()  ▷ banda mínima; mesmo
                     ▷ bloco 64×64 se possível
 4   para y ∈ N4(x), z(y) indefinido:
 5     se B(y)=1: rejeita(falha); continua
 6     ẑ ← z(x) + p(x)            ▷ (5)
 7     d* ← argmax_{|d|≤W} ρ(d)   ▷ (2)
 8     z′ ← ẑ + d* + δ            ▷ (4)
 9     se ¬aceita(ρ,pol,a,|z′−T(y)|≤τ):
10        rejeita(motivo); continua
11     se volta(y→x) ≠ z(x): rejeita(volta)
12     senão z(y) ← z′; Q.insere(y, banda(y))
saída: z, contadores de rejeição por motivo
```

4.1 *Medida de semelhança e o seu limiar*

Dado um traço de referência a centrado em t_0 e um candidato b centrado em t_1, procura-se o deslocamento d, $|d| \le W$, que maximiza

$$\rho(d) = \frac{\sum_{k=-M}^{M} a(t_0+k)\,b(t_1+k+d)}{\sqrt{\sum_k a^2(t_0+k)}\,\sqrt{\sum_k b^2(t_1+k+d)}} \qquad (2)$$

A expressão (2) não subtrai a média: é a similaridade do cosseno, que coincide com o coeficiente de Pearson apenas para dados de média nula. Em dados processados a componente contínua é removida e a distinção é irrelevante; após certos atributos, a componente comum infla o numerador. É uma limitação declarada.

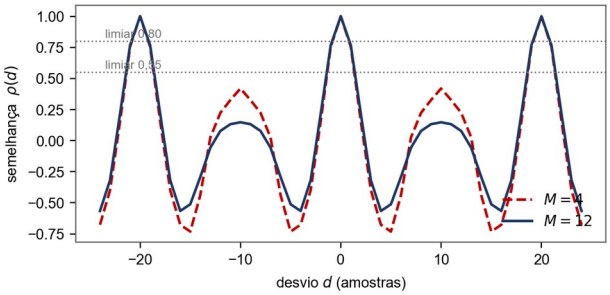


Figura 3. Semelhança em função do deslocamento num traço quase periódico. Com $M = 4$ os laços adjacentes pontuam quase como o correcto; com $M = 12$ o máximo correcto destaca-se. As linhas horizontais marcam os limiares 0,80 e 0,55 (secção 6.1).

Janela. A correlação de um sinal quase periódico sobre pouco mais de um período é ela própria quase periódica (figura 3); a janela tem de abranger vários ciclos para que a assinatura conjunta distinga um laço do vizinho. Adoptamos $M = 12$ (±48 ms, janela de 100 ms) e $W = 5$; a secção 6.2 mede o efeito da janela.

Limiar sob a hipótese nula. Para dois segmentos independentes de ruído gaussiano isotrópico com n amostras efectivas, a estatística $t = \rho\sqrt{(n-2)}/\sqrt{(1-\rho^2)}$ segue uma distribuição t de Student com $n-2$ graus de liberdade (Fisher, 1915). Como o rastreador escolhe o melhor de $2W+1$ deslocamentos, a probabilidade de aceitar ruído é majorada por

$$P(\max \rho > \theta) \le (2W+1)\,P\left(t_{n-2} > \frac{\theta\sqrt{(n-2)}}{\sqrt{(1-\theta^2)}}\right) \qquad (3)$$

Com $n = 25$ e $2W+1 = 11$, o majorante é $2{,}4 \times 10^{-2}$ para $\theta = 0{,}55$ e $8{,}7 \times 10^{-6}$ para $\theta = 0{,}80$. Como o ruído sísmico é de banda limitada, o número efectivo de amostras independentes é inferior a $2M+1$; com $n = 11$ os valores sobem para 0,36 e 0,017. Em ambos os casos $\theta = 0{,}55$ aceita ruído com frequência relevante à escala de 6×10^5 células, enquanto $\theta = 0{,}80$ não. O argumento tem um alcance preciso: trata da aceitação de **ruído**, não do salto de ciclo. Um laço vizinho do mesmo evento correlaciona acima de 0,80 (figura 3), pelo que nenhum limiar de semelhança o recusa; essa é a função da tolerância de amplitude e da ida e volta.

4.2 *Refinamento abaixo da amostra*

A posição inteira carrega a quantização do volume: ±2 ms a 4 ms de amostragem, cerca de ±2,5 m a 2500 m/s. Manifesta-se como degraus e propaga-se aos mapas derivados — em particular às isócoras, onde a diferença de duas superfícies quantizadas produz um padrão em xadrez sem significado

geológico. Ajusta-se uma parábola aos valores y_1, y_2, y_3 nas amostras anterior, central e seguinte ao extremo:

$$\delta = \frac{1}{2}\,\frac{y_1 - y_3}{y_1 - 2\,y_2 + y_3}, \quad |\delta| \le ½ \tag{4}$$

com retorno a zero para denominador nulo ou resultado não finito; um vértice fora de [−½, ½] indica que o extremo verdadeiro está noutra amostra.

4.3 Previsão de orientação

Sem previsão, a pesquisa é centrada na posição anterior e, num reflector inclinado, o evento correcto pode sair da janela enquanto o laço vizinho entra. O mergulho local p (amostras/traço) é estimado pelo tensor de estrutura das derivadas g_t (ao longo da direcção de avanço) e g_z (vertical), suavizado por $\langle\cdot\rangle$:

$$p = -\frac{J_{tz}}{J_{zz}}, \quad J_{uv} = \langle g_u\, g_v \rangle \tag{5}$$

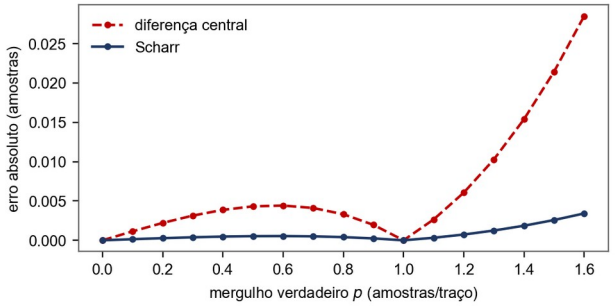


Figura 4. Erro absoluto da orientação estimada num campo sintético de mergulho conhecido. A diferença central enviesa sistematicamente, e o enviesamento cresce com o mergulho — é máximo onde a previsão mais importa.

Núcleos de derivada. Uma diferença central atenua as altas frequências de forma desigual nos dois eixos, porque a ondícula varia depressa ao longo do traço e devagar através dos traços; o quociente (5) sai enviesado sempre no mesmo sentido (figura 4). Os núcleos de Scharr (2000), optimizados para o erro angular, separam derivada e suavização transversal:

$$k_d = (-½, 0, ½), \quad k_s = \frac{1}{16}(3, 10, 3) \tag{6}$$

Escala interna. Um molde de 3 × 3 não resolve mergulhos superiores a cerca de uma amostra por traço: a mergulho 2,5 os gradientes sofrem aliasing e a orientação colapsa para a horizontal, com 1,77 amostras de erro medido no flanco de um domo. Suavizar ao longo do traço antes de derivar, com um núcleo binomial de cinco pontos, reduz o erro para 0,56; como a mergulho 3 o erro permanecia em 1,9, aplicam-se duas passagens. É um compromisso explícito entre alcance angular e resolução vertical.

Confiança. O contraste de valores próprios fornece, sem custo adicional, uma medida de linearidade

$$lin = \frac{\sqrt{(J_{tt} - J_{zz})^2 + 4\,J_{tz}^2}}{J_{tt} + J_{zz}} \tag{7}$$

que vale 1,000 num campo estratificado sintético e 0,191 em ruído gaussiano puro, com o mesmo molde e janela.

4.4 Testes de aceitação

Cada candidato atravessa os testes da tabela 1, com um contador por motivo de rejeição. A **tolerância de amplitude** é o critério que detém o salto de ciclo propriamente dito: o laço acima e o laço abaixo são a mesma ondícula e passam o limiar de semelhança sem dificuldade, porque (2) é invariante à escala; o que os distingue é a amplitude. A **referência de comparação** decide contra que traço se medem semelhança e amplitude — o do progenitor ou o da semente (secção 3.2).

Tabela 1. Testes de aceitação e valores por omissão, por ordem de custo crescente após a barreira de falha.

Teste	Valor por omissão	Detém
Semelhança mínima	ρ ≥ 0,80	Correlação baixa
Polaridade	a da semente	Evento oposto
Amplitude mínima	0,30 × semente	Reflector extinto
Passo máximo	2 am./traço	Mergulho irrealista
Tolerância de amplitude	$\lvert a/a_{\mathrm{ref}} - 1\rvert \le 0{,}60$	Salto de ciclo
Tolerância à tendência	desligada (τ = 0)	Saltos só se τ < 16 ms

Uma versão anterior usava a tendência como *guia*, centrando a pesquisa numa interpolação por inverso da distância, o que prendia o resultado às próprias sementes: a tendência não é evidência. Como *tolerância* é coerente, mas a medição mostra o seu limite: com τ = ±60 ms a cobertura desceu de 20,8 % para 15,1 %, com o mesmo pior degrau e nenhum degrau anómalo em qualquer dos casos. A faixa recusou células sem recusar saltos — o que a secção 3.2 prevê, porque 60 ms está muito acima de $T_w/2 \approx 16$ ms. Por isso está desligada por omissão.

4.5 Validação de ida e volta

Todos os testes anteriores são afirmações *sobre o candidato*, e um rastreador que deslize um laço pode satisfazê-los a todos. A validação de ida e volta olha na direcção oposta: encontrado z' em y a partir de x, procura-se o evento em x a partir de z'; se a resposta não for $z(x)$, o passo é recusado. Como (2) é simétrica — para as mesmas janelas, $\rho_{AB}(d) = \rho_{BA}(-d)$ —, o teste aceita exactamente os pares que são **melhor correspondência mútua** dentro das duas janelas de pesquisa. Detecta, por isso, os saltos assimétricos, como o da figura 5, em que o laço escolhido aponta de volta para outro laço; **não detecta** um salto em que o laço errado seja o melhor par mútuo. É um detector de inconsistência, não uma prova de correcção, e a sua taxa de detecção tem de ser medida contra uma referência (secção 7.2). O custo é uma segunda pesquisa por passo aceite.

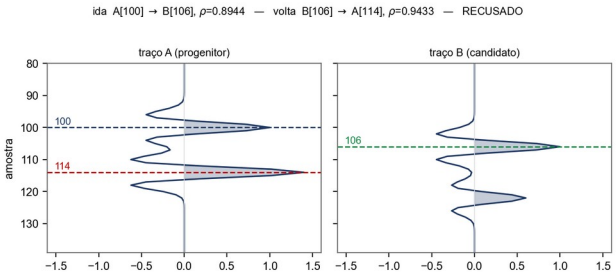


Figura 5. Um passo aceite por todos os testes de ida e recusado pela volta. A ida, de A(100), escolhe B(106) com ρ = 0,8944, acima do limiar; a volta, de B(106), escolhe A(114) com ρ = 0,9433. Os dois traços discordam sobre a identidade do evento.

4.6 Ordem de crescimento

A prioridade de crescimento decide que célula é visitada a seguir e é independente do critério de aceitação. A **prioridade por amplitude** escolhe sempre a melhor correlação e produz dedos longos e estreitos que se afastam do controlo; onde dois dedos se encontram, $\ell(x, y)$ em (1) é da ordem do comprimento dos dedos e a junção é uma lâmina quase vertical. A **prioridade por proximidade** cresce uma frente de onda — tudo a um

passo das marcações, depois tudo a dois — e nenhum ramo ultrapassa a frente, o que mantém ℓ pequeno entre vizinhos. O defeito não está em nenhum passo: cada célula de uma lâmina passa todos os testes. Está na ordem. Em termos de (1), as lâminas resultam sobretudo do termo sistemático: vizinhos alcançados por caminhos de comprimentos muito diferentes acumulam $b(L_x - L_y)$, que cresce linearmente com a diferença.

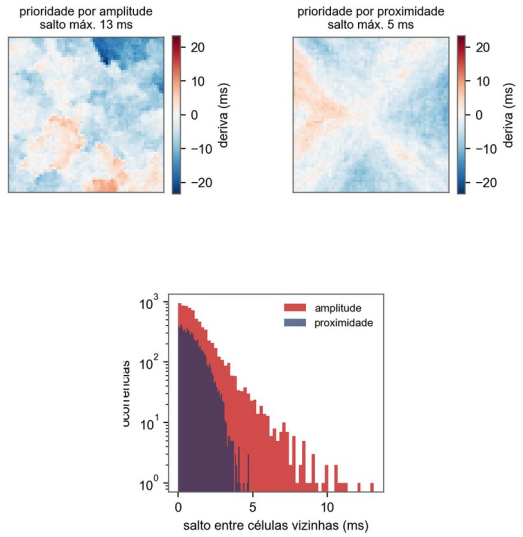


Figura 6. Simulação do mecanismo da equação (1): cada passo aceite acrescenta uma deriva independente. Em cima, deriva final sob as duas prioridades; em baixo, distribuição dos saltos entre células vizinhas (escala logarítmica). Salto máximo 13,1 ms contra 4,8 ms. A simulação demonstra o mecanismo; as magnitudes medidas no F3 estão na secção 6.2.

4.7 Barreira de falha em vista de mapa

Um rastreador que ignore uma falha atravessa-a e produz uma superfície suave precisamente onde a estrutura deveria estar partida. A abordagem convencional consulta um volume de coerência (Bahorich e Farmer, 1995; Marfurt *et al.*, 1998). Para a pergunta que o rastreador faz — *posso passar desta posição de mapa para aquela?* — a informação tridimensional é largamente redundante, porque uma falha de alto ângulo intersecta o mapa segundo uma linha. A barreira é por isso um octeto por posição de mapa, com razão de memória

$$\frac{N_i N_c N_s \times 4}{N_i N_c \times 1} = 4 N_s = 1848 \qquad (8)$$

ou 619 kB contra 1,14 GB no F3. Esta razão é uma identidade de construção, não um resultado; o ganho real é que o teste passa a ser uma leitura num vector contíguo, executada antes de qualquer correlação. A barreira não dispensa a interpretação das falhas — tem de ser derivada de um atributo ou de marcações — e não é equivalente ao volume: numa falha inclinada o traço em mapa varia com a profundidade, pelo que a barreira tem de ser construída à profundidade do horizonte (por exemplo, ao longo da tendência T), ou bloqueia em excesso. As falhas de baixo ângulo são o caso extremo.

4.8 Localidade de acesso

As amostras de um traço são contíguas em memória, com índice linear

$$idx(i, c, s) = i\, N_c N_s + c\, N_s + s \qquad (9)$$

Uma frente de onda pura visita células dispersas por todo o mapa antes de regressar à vizinhança de qualquer uma — o pior padrão possível num volume comprimido em blocos, em que cada acesso pode obrigar a descomprimir um bloco inteiro. A história do código ilustra-o. A primeira versão corria em 5 s, mas por acidente: uma prioridade constante tornava a ordem de visita arbitrária e local em cache — boa localidade, sem garantia de geometria. Corrigir a ordenação para a frente de onda da secção 4.6 levou o mesmo rastreio a mais de 600 s. Substituímos então a fila de prioridade por uma **fila de baldes** (Dial, 1969), com bandas de confiança em vez de ordenação total e **aderência a colunas de 64 × 64**, igual à dimensão dos blocos do formato ZGY, de modo que cada bloco descomprimido seja esgotado antes do seguinte (figura 7). A aritmética é a mesma, mas o resultado não é garantidamente idêntico: reordenar visitas dentro de uma banda pode mudar o progenitor de uma célula e, com ele, a decisão de aceitação.

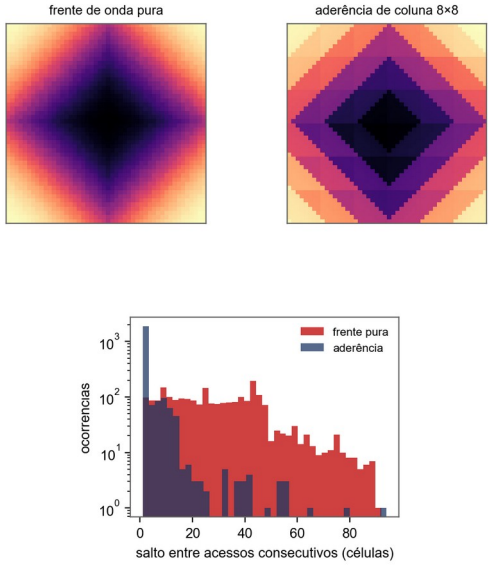


Figura 7. Ordem de visita (em cima; cor = instante de visita) e distribuição do salto entre acessos consecutivos (em baixo, escala logarítmica). A frente pura salta por todo o mapa; a aderência esgota um bloco antes do seguinte. Ilustração numa grelha reduzida com blocos de 8 × 8; em produção, 64 × 64.

4.9 Marcação guiada por programação dinâmica

Entre dois pontos conhecidos na mesma secção, o caminho óptimo obtém-se por programação dinâmica com retrocesso (cf. Hale, 2013), com custo de transição

$$C = \alpha(1-\rho) + \beta \frac{(z - z' - p)^2}{\sigma^2} + \gamma\, |\ln(a/a')| \qquad (10)$$

Duas falhas da formulação inicial, que não encontrámos documentadas, tornavam-na inutilizável. **Termo linear:** com custo proporcional a $|z - z'|$, um salto de doze amostras custa o mesmo que doze passos de uma; o termo quadrático torna-o doze vezes mais caro — a convexidade é o mecanismo. **Referência a mergulho nulo:** sem o termo $-p$, no flanco de um domo com mergulho de 3 amostras/traço seguir o reflector custava $3^2 = 9$ e um caminho horizontal custava 0, pelo que atravessar a estrutura era *mais barato* do que segui-la. Subtraindo a previsão, os custos invertem-se exactamente: 0 e 9.

5 DADOS E PROTOCOLO

Todas as medições usam o levantamento F3, da plataforma continental holandesa (dGB Earth Sciences): 651 inlines × 951 crosslines, 462 amostras a 4 ms, células de 25 m, azimute inline 358,4°. Os tempos são medianas de cinco repetições, em compilação optimizada e cache do sistema quente; medições em compilação de depuração não são reportadas. Os parâmetros — τ, κ, W, M, distribuição das sementes e restrições activas — dependem da experiência e são indicados em cada uma. Plataforma: Intel Core Ultra 9 185H, 31,4 GB de

memória, g++ 13.1.0 (MinGW) com -O3, cache de colunas de 256 MB; o crescimento corre numa única thread.

Dado o modo de falha silencioso, cada componente é verificado contra um caso de **solução analítica conhecida**, não contra a execução do próprio programa: campo sinusoidal de mergulho conhecido (orientação), ondícula avaliada fora da grelha (refinamento), falha sintética em que o rastreio tem de parar (barreira) e ida e volta exacta ao bit (persistência). A suite tem 209 testes — 203 unitários e 6 programas de ensaio, com 1080 asserções — e inclui um conjunto de referência publicado por terceiros, cujos valores esperados não provêm deste código.

Definições operacionais. Uma célula está num *extremo local da polaridade correcta* se a amplitude em z for o extremo da polaridade da semente em relação às amostras adjacentes (±1). É o teste mais fraco possível: mostra apenas que a superfície assenta num pico ou numa cava do traço, e é com esse alcance que os 99,9 % devem ser lidos. Um par de vizinhos tem um *degrau anómalo* se $|z(x) - z(y)| > 40$ ms. Nenhuma destas definições usa uma referência geológica: ambas medem consistência interna.

6 RESULTADOS

Tabela 2. Síntese dos resultados medidos no F3. Cada linha isola um mecanismo; as restantes condições mantêm-se. Nenhum valor é medido contra uma referência geológica independente (secção 7.2).

Mecanismo	Antes	Depois
Cobertura, progenitor → semente[a]	91,5 %	38,0 %
Intervalo temporal, idem[a]	511 ms	396 ms
idem, marcações de origem	276 ms	—
Cobertura, tendência desligada → ±60 ms	20,8 %	15,1 %
Degraus > 40 ms, amplitude → proximidade	40	0
Degraus > 40 ms, $M = 4 \rightarrow 12$	15	0
Pior degrau, $M = 4 \rightarrow 12$	72,7 ms	34,5 ms
Erro do campo de orientação	0,094 am.	0,004 am.
Deriva do rastreador (sintético)	17,4 am.	0,046 am.
Tempo, fila de prioridade → baldes[b]	> 600 s	3,0 s

[a] θ = 0,55, 43 sementes numa inline. [b] a versão original, com ordem degenerada, corria em 5 s. am. = amostras.

6.1 O modo de falha, quantificado

Este é o resultado central. Com θ = 0,55 e comparação com o progenitor, o rastreador preencheu **91,5 %** do levantamento a partir de **43 marcações numa única inline**, recusando apenas 2 863 de 619 101 células, com a tolerância à tendência desligada. 99,9 % das células assentavam num extremo local da polaridade correcta, e a superfície abrangia 511 ms quando as marcações abrangiam 276 ms.

Suspeita, não refutada. A razão 511/276 não demonstra o erro: com sementes numa só linha, um horizonte que mergulhe fora dela pode abranger legitimamente mais tempo. O que está demonstrado é mais limitado e basta para a tese: **os testes locais não distinguem as duas superfícies** — a correcta e uma que tenha mudado de laço passariam ambas com 99,9 %. Afirmar mais seria cometer o erro que o artigo descreve. Mudar apenas a referência de comparação para a semente reduziu a cobertura para 38,0 % e o intervalo para 396 ms. Menos cobertura não prova correcção — 396 ms continua acima das marcações —, mas mostra quanto da superfície original dependia da decisão acumulada ao longo do caminho.

A consequência prática é contra-intuitiva: **um limiar permissivo não produz erros visíveis — produz cobertura**. A desproporção entre marcações e cobertura é, ela própria, o aviso. Propomos por isso a razão de amplitude temporal $R = \Delta z_{\text{superfície}} / \Delta z_{\text{marcações}}$ como sinal de alerta — não como prova; aqui $R = 1{,}85$. O valor θ = 0,80 é adoptado e é consistente com o majorante (3).

6.2 Ordem de crescimento

Sob prioridade por amplitude, 40 pares de células vizinhas apresentavam degraus acima de 40 ms, até 275 ms, numa superfície cujo degrau médio é 1,15 ms; sob prioridade por proximidade, com os mesmos critérios, nenhum. A janela de correlação tem um efeito da mesma ordem: passar de $M = 4$ para $M = 12$ reduziu os degraus acima de 40 ms de 15 para 0 e o pior degrau de 72,7 para 34,5 ms. Os dois factores foram medidos um de cada vez e a sua interacção não foi medida, pelo que não atribuímos a eliminação dos degraus a nenhum deles em exclusivo. O degrau médio é cego a este defeito: é a cauda que o revela.

6.3 Orientação

A substituição da diferença central por núcleos de Scharr reduziu o erro do campo de orientação de 0,094 para 0,004 amostras e a deriva final do rastreador num caso sintético de 17,4 para 0,046 amostras — uma redução de duas a três ordens de grandeza com o mesmo custo computacional.

6.4 Custo computacional

Três tempos contam a história: 5 s na versão original, com ordem degenerada; mais de 600 s depois de corrigir a ordem; 3,0 s com a fila de baldes. O número honesto não é «200 × mais rápido», mas «a ordem correcta custou mais de 100 × e a fila de baldes recuperou esse custo». A aritmética total — 25 amostras, 11 deslocamentos, 619 101 células — é de cerca de 5×10^8 operações, menos de um segundo: **o tempo estava nos acessos**. As taxas de leitura não se comparam directamente com os tempos. A fila de baldes lê cerca de $4{,}1 \times 10^6$ amostras/s; a fila de prioridade começava em $2{,}18 \times 10^5$ (19 × menos) e caía para $8{,}9 \times 10^4$ às 200 000 células (46 × menos). Como a taxa decai à medida que a frente cresce, o tempo total integra a curva inteira e a razão acumulada excede ambas. A diferença célula a célula entre as duas filas não foi medida.

6.5 Cobertura

As coberturas reportadas só se comparam com a sua configuração: 91,5 % e 38,0 % (θ = 0,55, 43 sementes numa inline, progenitor ou semente como referência); 20,8 % e 15,1 % (tendência desligada ou a ±60 ms, restantes parâmetros iguais); 32,3 % com sementes de intérprete e 5,1 % com sementes automáticas; e 75,9 % reportados pela aplicação na sessão da figura 1. A cobertura depende sobretudo da distribuição das sementes (secção 3.3) e **não é uma medida de qualidade**: mais cobertura pode ser pior, e menos cobertura não prova que seja melhor. A colocação automática não substitui o intérprete.

7 DISCUSSÃO

7.1 O que se generaliza

Três resultados excedem este software. **Critérios locais são insuficientes em problemas de propagação com acumulação:** qualquer algoritmo que propague estado passo a passo, aceitando cada passo por critérios locais, obedece a (1) — o seguimento de objectos em vídeo, o desdobramento de fase e o *picking* de primeiras chegadas partilham a estrutura. **A ordem de visita é um parâmetro do algoritmo:** determina a forma da árvore e, portanto, onde as descontinuidades aparecem. **Em volumes grandes, a localidade domina a aritmética:** a ordem que a geometria exige é a pior para a memória, e a mesma aritmética variou entre 3 s e mais de 600 s só pela ordem de acesso.

7.2 Ameaças à validade

Validade interna. As comparações isolam um factor de cada vez, mas nenhum resultado foi ainda medido contra uma referência geológica. Essa referência existe: o repositório do F3 inclui horizontes interpretados, e Silva *et al.* (2019) publicaram nove horizontes reinterpretados manualmente por dois geocientistas, em acesso aberto. Rastrear esses horizontes a partir de sementes retiradas deles mede o erro absoluto em milissegundos, a taxa real de saltos e a fracção detectada pela ida e volta. **É a experiência que falta e a principal lacuna deste trabalho. Validade externa.** Um único levantamento, de geologia relativamente simples e boa qualidade: os mecanismos são gerais, as magnitudes podem não ser transferíveis. **Validade de construto.** O modelo (1) é uma hipótese explicativa, não um modelo estatístico ajustado ao rastreador: a constância de b ao longo do caminho e o curto alcance da correlação de η não foram verificados. **Ambiguidade da figura 1(d).** As paredes quase verticais junto à estrutura central tanto podem ser rejeito real como lâminas de crescimento; só a comparação com a referência as distingue. **Limitações de desenho.** A semelhança sem remoção de média sobrestima com dados com tendência; a barreira bidimensional exige construção à profundidade do horizonte; a tolerância à tendência só é útil abaixo de meio período e com marcações distribuídas; a comparação com a semente pressupõe que a ondícula muda pouco lateralmente.

7.3 Reprodutibilidade

As figuras 3 a 7 são geradas por código a partir das mesmas expressões do texto, não traçadas à mão: uma figura calculada mudaria se o algoritmo mudasse. Durante a preparação deste manuscrito, duas discrepâncias entre o texto e o gerador foram detectadas por este mecanismo e corrigidas.

8 PROPOSTA: INTEGRAÇÃO COM APRENDIZAGEM AUTOMÁTICA

> *Esta secção é uma proposta. Descreve uma arquitectura, um esquema de supervisão e um protocolo de avaliação; não apresenta modelos treinados nem desempenho medido.*

Treinar uma rede para produzir directamente o tempo geológico relativo (Geng *et al.*, 2020) enfrenta duas objecções. Exige volumes integralmente interpretados, escassos e de qualidade desconhecida — pela secção 6.1, uma interpretação publicada pode conter saltos de ciclo que ninguém detectou. E substitui o rastreador em vez de o informar, perdendo a interactividade. Propomos antes **substituir componentes, mantendo interfaces** (tabela 3).

Tabela 3. Pontos de integração propostos e origem dos rótulos.

Ponto	Componente aprendido	Rótulos
P1 Sementes	campo de aptidão (CNN 3-D) ou política	cobertura final
P2 Mergulho	regressão do campo p	sintéticos exactos
P3 Semelhança	rede siamesa (Hadsell *et al.*, 2006)	pares ida e volta
P4 Aceitação	classificador binário	decisões ida e volta
P5 Ordem	política de visita	cobertura e custo

8.1 Supervisão fraca pelo próprio rastreador

A validação de ida e volta é um oráculo que não depende de interpretação humana: correndo o rastreador sobre volumes não interpretados, cada passo tentado gera um exemplo rotulado (y = 1 se ida e volta concordam, 0 caso contrário), à maneira dos esquemas de supervisão fraca programática (Ratner *et al.*, 2017). Um único horizonte do F3 produz centenas de milhares de pares. A objecção central tem de ser enunciada: **um modelo treinado nas decisões do rastreador aprende o rastreador, não a geologia**. Há duas respostas parciais — se a viagem de volta for conservadora, hipótese ainda não medida, um modelo que a reproduza com maior cobertura pode ser melhor; e os rótulos de P2 são sintéticos e independentes do rastreador —, mas nenhuma dispensa a validação contra os horizontes de referência da secção 7.2. Pela secção 4.5, estes rótulos herdam ainda a cegueira da ida e volta aos saltos mutuamente consistentes.

8.2 Riscos específicos

P3 corrige uma limitação real — a invariância à escala de (2) torna-a cega à diferença de amplitude que denuncia o salto de ciclo — mas P4 é *menos auditável* do que cinco limiares nomeados e perde a explicação ao intérprete que os contadores de rejeição fornecem; propomos manter os contadores em paralelo. P5 tem uma restrição dura e não óbvia: qualquer política aprendida tem de incluir o custo de acesso na função objectivo, sob pena de perder o ganho da secção 6.4.

8.3 Protocolo de avaliação

Comparar cobertura é insuficiente: a secção 6.1 mostra que mais cobertura pode ser pior. Propomos quatro métricas por ordem de precedência: (1) a razão de amplitude temporal R, próxima de 1; (2) a cauda da distribuição de saltos entre vizinhos; (3) **validação cruzada por retenção** — sobre um horizonte de referência (no F3, os de Silva *et al.*, 2019), marcam-se N linhas, rastreia-se com $N - k$ e mede-se o erro em milissegundos nas k retidas; e (4) só então a cobertura. A métrica 3 é a única comparação honesta contra uma resposta conhecida e penaliza directamente o salto de ciclo, que erra por cerca de um período da ondícula nas linhas retidas. Qualquer afirmação de superioridade de um componente aprendido deve ser sustentada por ela, com as linhas retidas escolhidas antes do treino.

O risco maior é a **degradação silenciosa**. Toda a arquitectura das secções 4 a 6 existe porque este domínio não dá sinal quando erra; um componente aprendido não altera essa propriedade — amplifica-a, porque é mais difícil de inspeccionar. Qualquer integração tem de trazer consigo o seu próprio mecanismo de detecção de erro silencioso; sem ele, não deve ser integrada.

9 CONCLUSÕES

O principal modo de falha do rastreio semeado de horizontes é silencioso e, pelo modelo (1), indetectável por construção por critérios locais: uma superfície com 99,9 % das células em extremos locais e uma razão de amplitude temporal de 1,85 passa todos os testes locais, esteja certa ou errada. A restrição não local que funcionou foi comparar com o traço da semente e não com o progenitor; a faixa em torno da tendência não recusa

saltos de um laço acima de meio período e ficou desligada. A janela de correlação e a ordem de crescimento controlam os degraus anómalos, e a ordem correcta, cara para a memória, tornou-se viável com uma fila de baldes: de mais de 600 s para 3,0 s. O passo seguinte está definido e é realizável com dados públicos: comparar com os horizontes do F3 interpretados manualmente e mostrar, em secções sísmicas, os casos de salto e de lâmina. Só então a superfície de 511 ms passa de suspeita a refutada — ou a correcta.

DISPONIBILIDADE DE DADOS E AGRADECIMENTOS

O conjunto F3 é disponibilizado publicamente pela dGB Earth Sciences. As medições sobre dados de poço usaram o projecto de formação do Roxar RMS, publicado de forma independente. O trabalho não utiliza nem liga contra código sob licença recíproca; as dependências são permissivas: OpenZGY (Apache-2.0), glm e spdlog (MIT) e Qt (LGPL-3.0, ligação dinâmica). O gerador das figuras acompanha o trabalho. Horizontes de referência: Silva *et al.* (2019), doi:10.5281/zenodo.1422787.

REFERÊNCIAS

Adams, R. e Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), 641–647.

Alaudah, Y., Michałowicz, P., Alfarraj, M. e AlRegib, G. (2019). A machine-learning benchmark for facies classification. *Interpretation*, 7(3), SE175–SE187.

Bahorich, M. e Farmer, S. (1995). 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube. *The Leading Edge*, 14(10), 1053–1058.

Bigün, J. e Granlund, G. H. (1987). Optimal orientation detection of linear symmetry. *Proc. 1st ICCV*, 433–438.

dGB Earth Sciences. *F3 Demo dataset*, offshore Netherlands. Dados públicos.

Dial, R. B. (1969). Algorithm 360: Shortest-path forest with topological ordering. *Communications of the ACM*, 12(11), 632–633.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4), 507–521.

Fomel, S. (2002). Applications of plane-wave destruction filters. *Geophysics*, 67(6), 1946–1960.

Geng, Z., Wu, X., Shi, Y. e Fomel, S. (2020). Deep learning for relative geologic time and seismic horizons. *Geophysics*, 85(4), WA87–WA100.

Hadsell, R., Chopra, S. e LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping. *Proc. IEEE CVPR*, 1735–1742.

Hale, D. (2013). Dynamic warping of seismic images. *Geophysics*, 78(2), S105–S115.

Hoyes, J. e Cheret, T. (2011). A review of "global" interpretation methods for automated 3D horizon picking. *The Leading Edge*, 30(1), 38–47.

Kalal, Z., Mikolajczyk, K. e Matas, J. (2010). Forward–backward error: Automatic detection of tracking failures. *Proc. 20th ICPR*, 2756–2759.

Kass, M. e Witkin, A. (1987). Analyzing oriented patterns. *Computer Vision, Graphics, and Image Processing*, 37(3), 362–385.

Lomask, J., Guitton, A., Fomel, S., Claerbout, J. e Valenciano, A. A. (2006). Flattening without picking. *Geophysics*, 71(4), P13–P20.

Marfurt, K. J., Kirlin, R. L., Farmer, S. L. e Bahorich, M. S. (1998). 3-D seismic attributes using a semblance-based coherency algorithm. *Geophysics*, 63(4), 1150–1165.

Projecto OpenZGY. *Open source ZGY accessor library*. Licença Apache-2.0.

Ratner, A., Bach, S. H., Ehrenberg, H., Fries, J., Wu, S. e Ré, C. (2017). Snorkel: Rapid training data creation with weak supervision. *Proc. VLDB Endowment*, 11(3), 269–282.

Scharr, H. (2000). *Optimale Operatoren in der digitalen Bildverarbeitung*. Tese de doutoramento, Universidade de Heidelberga.

Silva, R. M., Baroni, L., Ferreira, R. S., Civitarese, D., Szwarcman, D. e Vital Brazil, E. (2019). Netherlands dataset: A new public dataset for machine learning in seismic interpretation. *arXiv*:1904.00770.

Stark, T. J. (2004). Relative geologic time (age) volumes — Relating every seismic sample to a geologically reasonable horizon. *The Leading Edge*, 23(9), 928–932.

Wu, X. e Hale, D. (2015). Horizon volumes with interpreted constraints. *Geophysics*, 80(2), IM21–IM33.

Wu, X., Liang, L., Shi, Y. e Fomel, S. (2019). FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics*, 84(3), IM35–IM45.