

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Análise tempoescala de séries temporais de geofísica espacial com lacunas: estudo de caso

Luciano Aparecido Magrini¹

Programa de Pós Graduação em Computação Aplicada

Instituto Nacional de Pesquisas Espaciais - INPE

Instituto Federal de Educação, Ciência e Tecnologia - IFSP

Margarete Oliveira Domingues²

Laboratório Associado de Computação e Matemática Aplicada - INPE

Odim Mendes Junior³

Divisão de Geofísica Espacial - INPE

Resumo. Séries temporais que não estejam regularmente distribuídas ao longo do tempo não podem ser tratadas pela análise espectral padrão. Nestes casos, é prática comum usar algum método interpolatório para estimar os valores nas lacunas e na sequência fazer a análise tempoescala ou, como alternativa, aplicar uma técnica *wavelet* adaptada e desenvolvida especialmente para estes casos. Neste trabalho, utilizando uma série temporal típica de dados medidos em satélites para aplicações da Geofísica Espacial e a construção de uma série sintética derivada para testes analisam-se as alterações que o uso da interpolação e da técnica *wavelet* adaptada introduzem na energia calculada do sinal e no comportamento do escalograma e do espectro global *wavelet*. Os resultados são bastante esclarecedores dos efeitos das técnicas de tratamento.

Palavras-chave. Séries temporais. *Wavelet* adaptativa. Interpolação. Lacunas em dados.

1 Introdução

A análise de *wavelets* tem sido amplamente empregada por pesquisadores ligados a diversas áreas que possuem interesse em extrair informações relevantes de séries temporais. Neste caso, uma das poucas exigências é de que as séries temporais sejam regularmente distribuídas no tempo, o que principalmente no caso das ciências espaciais pode ser um problema devido a limitações naturalmente impostas, como por exemplo, as condições atmosféricas e de clima espacial, funcionamento dos equipamentos de medição e capacidade de armazenamento e transmissão dos dados. Tais limitações podem gerar séries temporais com dados irregularmente distribuídos no tempo, ou seja, séries que apresentam falhas. As regiões com falha causam dificuldades para a análise tempoescala, porque nessas regiões ocorre a perda da condição de admissibilidade da base *wavelet* escolhida.

¹magrini@ifsp.edu.br

²margarete.domingues@inpe.br

³odim.mendes@inpe.br

Um possível tratamento nesses casos consiste em aplicar algum método interpolatório para estimar numericamente os registros nas regiões que apresentam lacunas, sendo comumente empregados *splines* cúbicos neste processo. Alternativamente, *Frick et al* [1998] propuseram a introdução de uma família de *wavelets* adaptadas, baseadas na *wavelet* de Morlet e caracterizada pela introdução de um fator de correção a ser computado para cada tempo e escala considerados de modo a restaurar a condição de admissibilidade nas regiões de lacuna.

Neste trabalho, apresenta-se um exemplo sintético construído de modo a reproduzir um comportamento bastante observado nas séries de Geofísica Espacial. O objetivo é analisar séries temporais com lacunas para investigar as variações provocadas no escalograma e no espectro global *wavelet* quando se estuda o desempenho, no sentido de reprodução tempoescala, dos dados tratados por três técnicas. A primeira técnica é a do método interpolatório por *splines* cúbicos [1, 6], escolhida por ser de uso habitual. A segunda é a do método interpolatório dos polinômios cúbicos de Hermite [1, 6], que permitirá uma comparação de desempenho com relação aos *splines*. A terceira técnica diz respeito a *wavelets* adaptadas às regiões com falhas [2–4], sendo que aqui a *wavelet* de Morlet é usada como função base para as análises espectrais apresentadas.

2 Metodologia

Fazendo uso de uma série temporal de dados sem falhas derivou-se um exemplo sintético com intervalos de falhas artificialmente introduzidos. O exemplo foi tratado com interpolações via *splines* cúbicos e via polinômios cúbicos de Hermite, aplicando na sequência a transformada *wavelet* contínua. Fez-se também o tratamento desta série com falhas com a transformada *wavelet* contínua adaptativa. Os resultados de análise espectral obtidos foram comparados com a análise espectral da série sem falhas, de forma que fosse possível comparar as técnicas de tratamento empregadas. A seguir estão descritas, resumidamente, as três técnicas.

Polinômios Cúbicos de Hermite A interpolação de Hermite permite aproximar um sinal f por um polinômio H (chamado polinômio de Hermite) que coincide com f nos pontos conhecidos, exigindo além disso que as derivadas também coincidam nestes mesmos pontos. Suponha f conhecido em $n + 1$ pontos t_0, t_1, \dots, t_n com $t_i \neq t_j$ se $i \neq j$. Por construção, para se obter H impõe-se

$$H(t_i) = f(t_i), i = 0, 1, \dots, n$$

$$H'(t_i) = f'(t_i), i = 0, 1, \dots, n$$

que é equivalente a um sistema com $2n + 2$ equações; tal sistema conduz a um polinômio interpolador de grau menor ou igual a $2n + 1$, conforme [1].

Nota-se que o grau de H tende a ser relativamente alto quando n é grande, e isto, pode conduzir a erros causados pelo excessivo número de oscilações deste polinômio; para contornar tal dificuldade, se forem conhecidos os valores de f e f' nos pontos $t_0 \leq t_1 \leq$

... $\leq t_n$ faz-se a construção de um polinômio cúbico de Hermite em cada subintervalo $[t_0, t_1], [t_1, t_2], \dots, [t_{n-1}, t_n]$ de modo a se obter uma função contínua (por partes) no intervalo $[t_0, t_n]$.

A construção do polinômio H pode ser feita por meio da interpolação por diferenças divididas de Newton para o polinômio de Lagrange e a demonstração deste fato pode ser encontrada em [6]. Em termos de diferenças divididas o polinômio cúbico de Hermite no intervalo $[t_k, t_{k+1}]$ para $k = 1, 2, \dots, n - 1$ se escreve

$$H_3(x) = f(t_k) + f[t_k, t_k](t - t_k) + f[t_k, t_k, t_{k+1}](t - t_k)^2 + f[t_k, t_k, t_{k+1}, t_{k+1}](t - t_k)^2(t - t_{k+1}).$$

Splines Cúbicos Menos restritiva que a interpolação de Hermite, a interpolação por *splines* cúbicos não exige que as derivadas da função interpoladora S coincidam com as derivadas do sinal aproximado em ponto algum; esta técnica assegura, contudo, que o sinal S seja continuamente diferenciável e tenha derivada segunda contínua [6].

Considere o sinal f definido no intervalo $[t_0, t_n]$ e conhecido no conjunto de pontos t_0, t_1, \dots, t_n . Define-se um *spline* cúbico S para f como a função que satisfaz as seguintes condições:

1. $S(t)$ tem grau três (denotado $S_j(t)$ no intervalo $[t_j, t_{j+1}]$ para $j = 0, 1, \dots, n - 1$);
2. $S_j(t_j) = f(t_j)$ e $S_j(t_{j+1}) = f(t_{j+1})$ para $j = 0, 1, \dots, n - 1$;
3. $S_{j+1}(t_{j+1}) = S_j(t_{j+1})$ para $j = 0, 1, \dots, n - 2$;
4. $S'_{j+1}(t_{j+1}) = S'_j(t_{j+1})$ para $j = 0, 1, \dots, n - 2$;
5. $S''_{j+1}(t_{j+1}) = S''_j(t_{j+1})$ para $j = 0, 1, \dots, n - 2$;
6. $S'''(t_0) = S'''(t_n) = 0$

CWT Adaptada A transformada *wavelet* contínua (CWT) de um sinal f é definida como

$$\psi(a, b) = C_\psi^{-1/2} a^{-1/2} \int_{-\infty}^{\infty} \overline{\psi\left(\frac{t-b}{a}\right)} f(t) dt,$$

em que C_ψ é uma constante, a é um número real positivo e $\bar{\psi}$ é o conjugado da função ψ que satisfaz as condições de energia unitária e de admissibilidade. A primeira condição garante a existência de um suporte compacto para a função ψ fora do qual acontece uma rápida queda em sua amplitude e a segunda garante que ψ é uma função de média zero com caráter oscilatório e inversível. *Frick et al* [1998] apresentaram uma técnica *wavelet*, conhecida como *wavelet gapped* que adapta a função *wavelet* para o caso em que f apresenta falhas; para tal, é necessário construir um sinal auxiliar $f_{aux} = f(t)G(t)$ definido pelo produto de f por

$$G(t) = \begin{cases} 1, & \text{se } f(t) \text{ é corretamente conhecido} \\ 0, & \text{se } f(t) \text{ é uma falha ou está fora da janela de observação} \end{cases}$$

A CWT deste sinal auxiliar é $\psi(a, b) = C_\psi^{-1/2} a^{-1/2} \int_{-\infty}^{\infty} \overline{\psi_{aux}\left(\frac{t-b}{a}\right)} f_{aux}(t) dt$, em que se define $\psi_{aux}\left(\frac{t-b}{a}\right) = \psi\left(\frac{t-b}{a}\right) G(t)$, exigindo deste modo que a função G seja "absorvida" pela função de base wavelet.

Durante a análise nas regiões com falha emprega-se ψ_{aux} ao invés de ψ ; deste modo, o problema de se analisar o sinal f com falhas equivale ao problema de se encontrar uma *wavelet* adaptada $\tilde{\psi}$ que substitua ψ_{aux} e que garanta a validade da condição de admissibilidade. A construção de $\tilde{\psi}$, baseada na *wavelet* de Morlet, pressupõe ser possível escrever $\psi(t) = h(t)\Phi(t)$, em que h e Φ são responsáveis pela parte oscilatória e pelo envelope de ψ , respectivamente. De acordo com [3], a função $\tilde{\psi}$ é dada explicitamente por

$$\tilde{\psi}(t, b, a) = \left[h\left(\frac{t-b}{a}\right) - \frac{\int_{-\infty}^{\infty} h\left(\frac{t-b}{a}\right) \Phi\left(\frac{t-b}{a}\right) G(t) dt}{\int_{-\infty}^{\infty} \Phi\left(\frac{t-b}{a}\right) G(t) dt} \right] \Phi\left(\frac{t-b}{a}\right) G(t),$$

3 Dados

Para demonstrar a importância prática dos cuidados na utilização de métodos de tratamentos de dados escolheu-se, para a série temporal em análise, os dados do meio interplanetário relacionados a um importante efeito geofísico da interação do vento solar com a atmosfera da Terra: o evento geofísico HILDCAA (*High-intensity, long-duration, continuous auroral electrojet activity*). Foram analisadas 120 séries temporais referentes à componente B_z , de direção norte-sul, do campo magnético interplanetário no período de 1975 a 2012 [5]. Os dados foram obtidos do site OMNIWeb⁴ usando a resolução de um minuto. Devido a propagação do sinal do campo magnético da posição do satélite em que é medido até o seu efeito na interação com a atmosfera magnetizada da Terra, considerou-se um tempo de deslocamento de 30 e 90 minutos com relação aos horários de início e término dos eventos HILDCAAs, respectivamente. Isso definiu o intervalo da série a ser analisada.

Com o objetivo de analisar o desempenho dos *splines* cúbicos, dos polinômios cúbicos de Hermite e das *wavelets* adaptadas no tratamento das lacunas, bem como as alterações que o uso destas técnicas provocam no espectro global, gerou-se um exemplo sintético por meio da introdução artificial de lacunas na série temporal correspondente ao evento HILDCAA ocorrido entre 28/08/1996 e 30/08/1996, por estar livre de falhas. As falhas introduzidas reproduzem um dos padrões observados no conjunto das séries, pois em aproximadamente 53,3% delas encontram-se falhas do mesmo tipo (cujo comprimento é superior a 10% do comprimento do sinal). Foram introduzidas quatro lacunas de comprimentos 14,60%, 12,81%, 12,81% e 11,28% nos intervalos [5, 461], [1000, 1400], [2200, 2600] e [2747, 3099] respectivamente. Na Figura 1, vê-se a componente magnética B_z referente ao evento HILDCAA e a série sintética, com falha, dela derivada.

⁴<http://omniweb.gsfc.nasa.gov/>

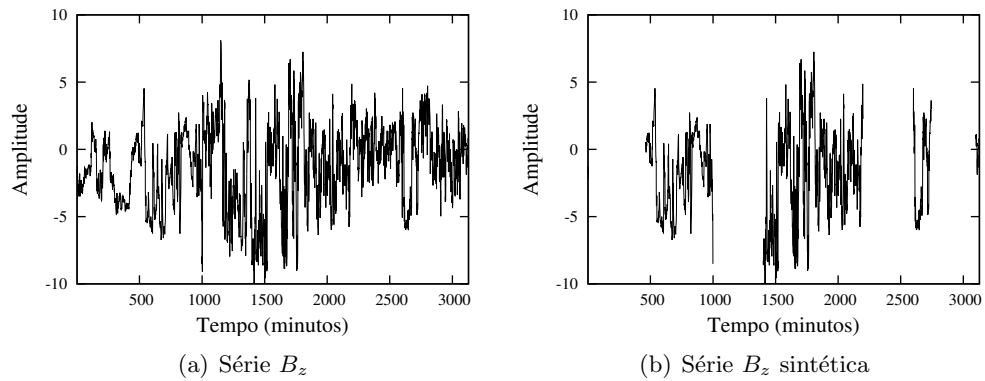


Figura 1: Componente magnética B_z do plasma solar referente a evento HILDCAA e a série derivada sintética com falhas.

4 Resultados e discussões

Na Figura 2 é possível ver os escalogramas do evento original HILDCAA e do caso sintético tratado com as técnicas anteriormente descritas.

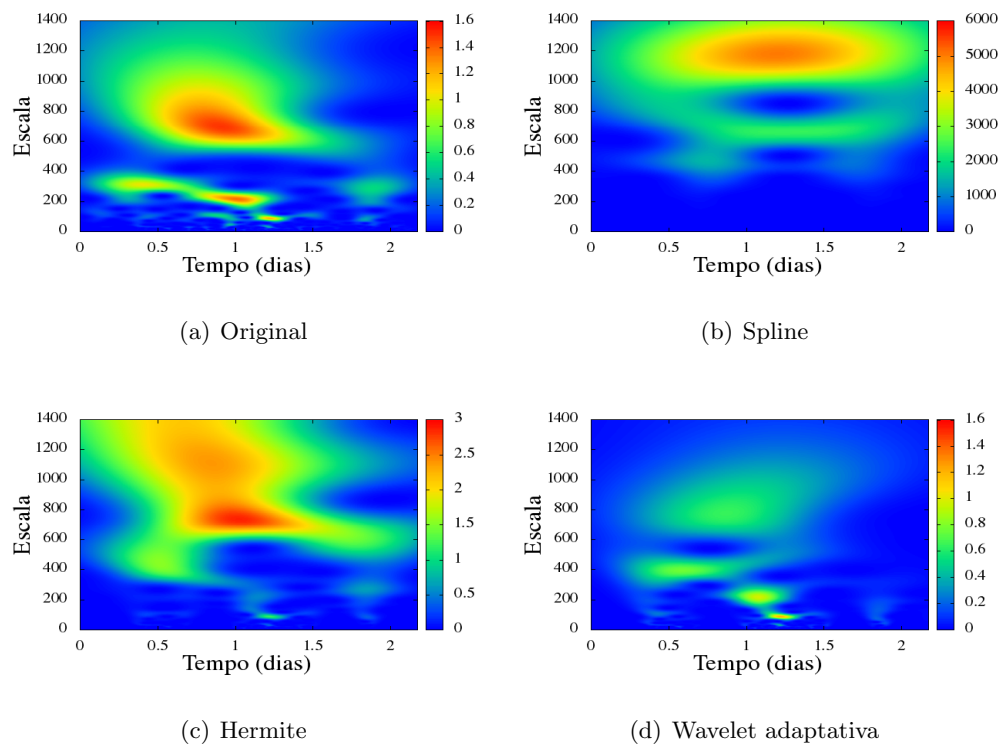


Figura 2: Escalogramas de energia: sinal original (a) e caso sintético: spline (b), Hermite (c) e *wavelet* adaptativa (d).

A interpretação dos escalogramas mostra que dentre os métodos empregados para o tratamento das falhas, somente o baseado na *wavelet* adaptada foi capaz de preservar as estruturas de energia presentes no sinal original, ainda que seja possível notar uma pequena perda da energia calculada no contexto do Teorema de Parseval⁵ Também se nota que o uso de *splines* cúbicos foi responsável por superestimar a energia em 3750 vezes. É possível verificar o mesmo no caso em que o tratamento se deu por polinômios cúbicos de Hermite, ainda que em menor proporção, pois neste caso a energia foi superestimada em 1,875 vezes. De forma importante, observa-se que nenhum dos dois métodos interpolatórios preservou as estruturas de energia do sinal original.

A análise do espectro global (Figura 3) indica que em altas frequências os resultados apresentados pela *wavelet* adaptada e pelo uso dos polinômios de Hermite são bastante próximos dos valores calculados para o sinal original; em baixas frequências, contudo, verifica-se que o erro apresentado pelo uso dos dois tratamentos tende a aumentar. Em consonância com o verificado no escalograma, o espectro *wavelet* obtido via uso de *splines* apresenta-se superestimado com relação ao original, sendo maior o erro em baixas frequências.

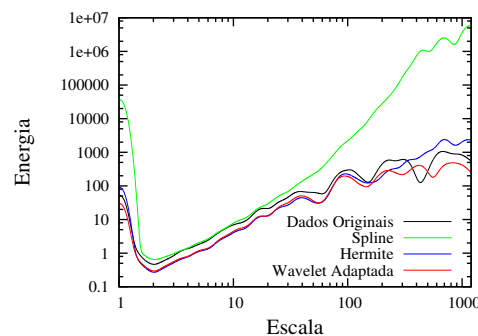


Figura 3: Espectros globais *wavelet*: Sinal original e casos de estudo.

Identificam-se no espectro global do sinal original cinco picos de energia que não foram, na totalidade, corretamente identificados pelas técnicas de tratamento estudadas. A *wavelet* adaptada foi capaz de identificar corretamente três picos de energia, apresentando dois falsos positivo; o tratamento com *splines* não permitiu identificar nenhum dos picos exatos e apresentou quatro falsos positivo; e o tratamento com polinômios cúbicos de Hermite identificou três dos picos e apresentou apenas um falso positivo. Em outras palavras, no caso da identificação dos picos de energia, os resultados apresentados pela *wavelet* adaptada e os resultados apresentados pelo uso da interpolação de Hermite estão bem próximos. Contudo, as estruturas identificadas pela *wavelet* estão melhor caracterizadas que as por Hermite.

⁵O Teorema de Parseval implica na conservação da energia nos espaços físico e de Fourier. Em termos matemáticos, tem-se $\int_{-\infty}^{\infty} |f(t)|^2 dt = \int_{-\infty}^{\infty} |F(\xi)|^2 d\xi$, onde f e F indicam, respectivamente, um sinal qualquer e sua Transformada de Fourier.

5 Conclusão

A qualidade dos sinais disponíveis para análise espectral apresenta grande importância. Quando, por alguma razão, há restrições a essa qualidade, os tratamentos a que são submetidos os dados podem criar artefatos nas análises. A análise local da energia de um sinal que apresente falhas similares às presentes no caso sintético desenvolvido neste trabalho mostrou ser mais adequadamente tratada por meio do método da *wavelet* adaptada. Para a análise global, os resultados indicam que em altas frequências o espectro é aproximado de modo satisfatório pela *wavelet* adaptada e pelo uso dos polinômios de Hermite. A interpolação por *splines* não constitui uma boa abordagem para o tratamento das falhas do tipo apresentado, pois provoca mudanças bastante significativas na energia local e global. Este tipo de estudo revela-se de grande importância para as aplicações espaciais, sobretudo para as análises feitas na Geofísica Espacial.

Agradecimentos

Agradecemos os apoios e auxílios recebidos do CNPq (processos 306038/2015-3 e 312246/2013-7), da CAPES e também à NASA pelos dados disponibilizados via OMNIWeb (<http://omniweb.gsfc.nasa.gov>).

Referências

- [1] R. L. Burden, J. D. Faires. *Análise Numérica*. CENGAGE Learning, São Paulo, 2015.
- [2] J. E. Castilho, M. O. Domingues, O. Mendes, A. Pagamisse. *Introdução ao mundo das wavelets*. Sociedade Brasileira de matemática Aplicada e Computacional, São Carlos, 2012.
- [3] P. Frick, A. Grossmann and P. Tchamitchian. Wavelet analysis of signals with gaps. In *Journal of Mathematical Physics*, 39:4091-4107, 1998. DOI: 10.1063/1.532485.
- [4] P. Frick, S. L. Baliunas, D. Galyagin, D. Sokoloff and W. Soon. Wavelet analysis of stellar chromospheric activity variations. In *The Astrophysical Journal*, 483:426-434, 1997.
- [5] R. Hajra, E. Echer, B. T. Tsurutani, W. D. Gonzalez. Solar cycle dependence of High-Intensity Long-Duration Continuous AE Activity (HILDCAA) events, relativistic electron predictors?. In *Journal of Geophysical Research: Space Physics*, 118:5626-5638, 2013. DOI: 10.1002/jgra.50530.
- [6] M. J. D. Powell. *Approximation theory and methods*. University Press, Cambridge, 1981.