

Detectando e Interpretando Sinais Periódicos em Séries Temporais Paleobiológicas

Helena Baptista Reis 14577622, Luca Marinho Nasser Valadares Paiva 13691375

Maio de 2025

Resumo:

Este relatório descreve a replicação dos métodos utilizados por [Robert A Rohde et al., 2005a](#) para investigar ciclos periódicos na diversidade de fósseis marinhos ao longo do Fanerozoico. A série temporal de diversidade foi inicialmente normalizada e submetida à remoção de tendência via ajuste polinomial de terceira ordem. Após o cálculo dos resíduos e aplicação de zero-padding, foi realizada a Transformada Rápida de Fourier (FFT) para obter o espectro de potência da série. Dois picos principais foram identificados nos períodos de aproximadamente 62 e 140 milhões de anos. Para avaliar sua significância estatística, foram implementados dois modelos nulos baseados em simulações de Monte Carlo: um embaralhamento de incrementos (Modelo R) e um bootstrap por blocos (Modelo W). O teste de significância indica que o pico de 62 Myr é estatisticamente relevante, enquanto o pico de 140 Myr não difere significativamente das flutuações esperadas sob os modelos nulos. Os resultados reforçam a hipótese de periodicidade de longa duração na biodiversidade, possivelmente associada a fatores externos.

1 Introdução

Este relatório descreve os procedimentos realizados no EP1 da disciplina de Computação IV. O objetivo principal foi replicar os métodos utilizados no artigo [Robert A Rohde et al., 2005a](#), com foco na análise de periodicidades em séries temporais de diversidade fóssil.

2 Pré-processamento dos Dados

Os dados de diversidade de gêneros marinhos foram carregados de um arquivo CSV e normalizados para milhares de gêneros. A série temporal apresentou tendência decrescente de longo prazo, com flutuações superpostas.

Remoção de Tendência

Nesta etapa, o objetivo principal foi isolar as flutuações periódicas da diversidade ao longo do tempo, eliminando variações que se devem a tendências de longo prazo (como o declínio geral da diversidade visto no gráfico original).

Para isso, ajustou-se um polinômio de 3ª ordem aos dados, uma função suave que representa a tendência global da série (Figura 1a). Após o ajuste do polinômio de 3ª ordem, a tendência estimada foi subtraída dos dados originais de diversidade, gerando uma nova série composta apenas pelos resíduos, isto é, as variações em torno da tendência (Figura 1b).

Esses resíduos são mais apropriados para análise espectral, pois removem componentes que poderiam mascarar os padrões cíclicos

que estamos tentando detectar.

Padding

Antes de aplicar a transformada de Fourier, estendemos a série com zeros nas extremidades (zero-padding) [Robert A. Rohde et al., 2005b](#). Esse procedimento suaviza os efeitos de borda na análise espectral e melhora a resolução em frequência do espectro resultante.

3 Análise Espectral

A análise espectral permite decompor a série de diversidade fóssil em componentes de frequência, revelando padrões periódicos que não são evidentes na série temporal bruta.

Transformada de Fourier e Espectro de Potência

Após a remoção da tendência polinomial, aplicamos a Transformada Rápida de Fourier (FFT) aos resíduos da série temporal. A FFT converte a série do domínio do tempo para o domínio da frequência, permitindo quantificar a contribuição de oscilações periódicas de diferentes escalas. O espectro de potência é obtido pelo quadrado do módulo dos coeficientes de Fourier, refletindo a intensidade de cada frequência presente nos dados.

Seleção de Frequências Relevantes

O espectro (Figura 1c) foi analisado apenas para frequências positivas e biologicamente relevantes, restringindo a análise a $0 < f < 0.05$ ciclos/Myr (correspondente a períodos superiores a 20 milhões de anos). Essa faixa cobre os ciclos de interesse paleobiológico e evita artefatos de alta frequência [Robert A. Rohde et al., 2005b](#).

Identificação e Caracterização de Picos

A identificação dos picos espectrais foi feita de forma automática, utilizando um threshold de 25% da potência máxima do espectro. Apenas picos suficientemente separados (distância mínima proporcional ao intervalo de frequências) foram considerados, evitando múltiplos picos próximos [Robert A Rohde et al., 2005a](#). Os principais picos detectados correspondem a períodos de aproximadamente 62 e 140 milhões de anos.

4 Modelos Nulos: Simulações de Monte Carlo

Para evitar a superestimação da significância de picos que podem surgir por acaso, é essencial comparar o espectro real com espectros simulados. Com esse objetivo, implementamos dois modelos nulos baseados em simulações de Monte Carlo, projetados para

preservar diferentes aspectos dos dados originais. Esses modelos permitiram estimar a probabilidade de que picos com intensidade semelhante à observada ocorressem apenas por flutuações aleatórias, fornecendo uma base estatística para avaliação da robustez dos padrões detectados.

Modelo R

Este modelo parte da suposição de que as possíveis oscilações periódicas observadas na série podem ser explicadas apenas pela variabilidade aleatória da magnitude e direção das mudanças sucessivas. Para testá-la, embaralhamos aleatoriamente os incrementos da série temporal original, ou seja, as diferenças entre valores consecutivos de diversidade. Esse embaralhamento destrói completamente a estrutura temporal dos dados, eliminando qualquer dependência sequencial ou padrão recorrente, mas preserva a distribuição estatística dos tamanhos dos saltos (amplitude das variações locais).

A cada simulação, uma nova série foi reconstruída acumulando esses incrementos a partir do valor inicial da série real. Para manter a comparabilidade com a série empírica, a tendência de longo prazo foi novamente removida por ajuste polinomial de 3ª ordem, produzindo uma série de resíduos simulada que pôde ser submetida à análise espectral.

Modelo W

Neste modelo, busca-se preservar parte da estrutura local de dependência temporal, especialmente autocorrelações de curto alcance, que são comuns em séries paleobiológicas [Robert A. Rohde et al., 2005b](#). Para isso, a série de resíduos (após remoção da tendência) foi segmentada em 20 blocos de tamanho aproximadamente igual.

Durante cada simulação, esses blocos foram embaralhados aleatoriamente e recombinaados para formar uma nova série. A ordem interna de cada bloco é mantida nesse procedimento, o que garante a preservação de padrões locais, mas a ordem global da série é destruída, eliminando a possibilidade de ciclos de longa duração se manterem.

Esse método é útil para testar a hipótese de que picos espectrais podem surgir da agregação de padrões de curto prazo, sem a necessidade de estruturas periódicas extensas. Ao aplicar a FFT nas séries simuladas, foi possível comparar a distribuição de potências espectrais obtidas com as observadas nos dados reais, controlando para a autocorrelação local.

4.1 Procedimento das Simulações

Para avaliar a significância estatística dos picos espectrais observados, realizamos 30.000 simulações independentes para cada modelo nulo (R e W). Cada simulação reproduz o mesmo fluxo de processamento aplicado aos dados reais, assegurando comparabilidade direta entre espectros reais e simulados. O procedimento adotado em cada simulação foi o seguinte:

- Uma nova série sintética é gerada de acordo com as regras do modelo nulo (R ou W), embaralhando incrementos ou blocos conforme descrito anteriormente.

- Nas simulações do modelo R, aplicou-se novamente a remoção da tendência polinomial de 3ª ordem, a fim de obter resíduos comparáveis aos da série original [Robert A. Rohde et al., 2005b](#).
- A Transformada Rápida de Fourier (FFT) foi então aplicada à série simulada, com o mesmo fator de zero-padding utilizado na análise do dado real. Isso assegurou que a resolução espectral entre espectros reais e simulados seja idêntica.
- O espectro de potência resultante foi normalizado para que sua média fosse igual à média do espectro real, permitindo comparações diretas entre as alturas dos picos, independentemente da escala absoluta da série simulada [Robert A. Rohde et al., 2005b](#).

A repetição desse processo gera, ao final, um conjunto de 30.000 espectros de potência para cada modelo nulo. A média desses espectros fornece uma estimativa do comportamento espectral esperado caso não exista nenhum padrão periódico real nos dados, ou seja, representa o fundo espectral sob a hipótese nula. Além disso, a variação entre os espectros simulados em cada frequência permite quantificar quão incomum é a altura dos picos observados no espectro real. Essa distribuição serve de base para calcular dois tipos de p-valor: ponto a ponto (avaliando a raridade de um pico em uma frequência específica) e global (avaliando a chance de aparecer um pico tão alto em qualquer parte do espectro analisado).

5 Teste de Significância Estatística dos Picos Espectrais

Após gerar os espectros simulados a partir dos modelos nulos, o passo seguinte foi verificar se os picos observados no espectro real são estatisticamente relevantes ou se poderiam ocorrer apenas por acaso. Para isso, comparamos diretamente a altura dos picos reais com a distribuição das potências espectrais obtidas nas simulações de Monte Carlo.

Metodologia do Teste de Significância

Para avaliar a relevância estatística dos picos identificados no espectro real (62 Myr e 140 Myr), aplicamos o teste pontual (Pointwise), utilizando os espectros simulados de ambos os modelos nulos (R e W). Esse teste avalia, para cada frequência específica de interesse, a proporção de simulações em que a potência espectral é igual ou superior à observada no dado real, medindo o quão incomum é a presença de um pico em uma frequência fixa.

Resultados Quantitativos do Teste

Os resultados do teste de significância para os dois principais picos espectrais estão sumarizados na Tabela 1. Os valores apresentados correspondem aos p-valores obtidos para o teste e modelo nulo.

Tabela 1

P-valores do teste de significância para os picos espectrais principais (62 Myr e 140 Myr) nos modelos R e W.

Período	R (freq)	W (freq)	Significância
62 Myr	0.000033	0.000133	99,99%
140 Myr	0.124600	0.006567	87.54%

6 Discussão

A análise de significância estatística dos dois principais picos espectrais revelou contrastes importantes. O pico em torno de 62 milhões de anos apresentou p-valores de 0.000033 no Modelo R e 0.000133 no Modelo W, ambos bem abaixo do limiar de 0,005 correspondente ao nível de 99,5% de confiança. Isso indica que, sob os dois modelos nulos considerados, há uma probabilidade inferior a 0,5% de que um pico com aquela intensidade ocorra por acaso, resultado que reforça a interpretação de que esse ciclo reflete um padrão recorrente com baixa probabilidade de origem estocástica.

Já o pico em aproximadamente 140 milhões de anos mostrou p-valores de 0.1246 no Modelo R e 0.006567 no Modelo W. No primeiro caso, o valor está claramente acima do limiar de significância, enquanto no segundo aproxima-se dele, mas ainda excede o critério adotado. Isso sugere que a evidência para a existência de um ciclo em 140 Myr é menos consistente e pode ser explicada pelas propriedades aleatórias dos dados, especialmente quando a estrutura de autocorrelação local é considerada, como faz o Modelo W.

A interpretação desses resultados deve levar em conta as limitações dos dados e dos modelos. Os registros de diversidade fóssil estão sujeitos a diversas fontes de incerteza, como lacunas no registro geológico, vieses de amostragem, redefinições taxonômicas e a resolução variável ao longo do tempo. Essas limitações afetam diretamente a precisão da série e, por consequência, a análise espectral. Quanto aos modelos nulos, o Modelo R, ao embaralhar os incrementos, elimina completamente a dependência temporal, enquanto o Modelo W, embora preserve alguma estrutura local, rompe a continuidade global da série. Assim, ambos simplificam a complexidade dos dados originais e podem não representar fielmente certos aspectos estruturais relevantes.

Apesar dessas restrições, a recorrência de um sinal forte em 62 Myr, detectado de forma consistente nos dois modelos, sugere que esse padrão pode ter uma origem sistemática. Diversas hipóteses já foram propostas para explicar essa periodicidade, incluindo variações astronômicas, eventos de impacto recorrentes, ciclos tectônicos e mudanças ambientais globais que afetam diretamente os ecossistemas marinhos. A natureza exata desse mecanismo ainda é objeto de debate, mas a presença de um sinal estatisticamente significativo e persistente reforça a necessidade de investigá-lo com outras abordagens independentes.

7 Conclusão

A replicação da análise espectral proposta por [Robert A Rohde et al., 2005a](#) confirmou a presença de um ciclo estatisticamente significativo com período de aproximadamente 62 milhões de anos na diversidade de fósseis marinhos ao longo do Fanerozoico. Esse sinal se destaca tanto visualmente no espectro quanto pelos testes estatísticos de Monte Carlo aplicados a dois modelos nulos distintos, com p-valores baixos

O segundo pico identificado, em torno de 140 Myr, embora presente no espectro de potência, não demonstrou significância estatística robusta frente aos modelos nulos utilizados. Isso sugere que sua ocorrência pode estar associada ao ruído inerente à estrutura temporal da série.

Análise Espectral Paleobiológica

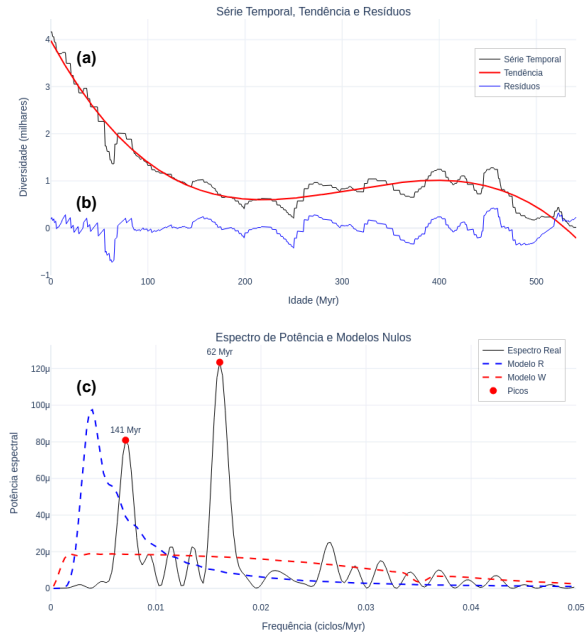


Figura 1. Resumo visual das etapas da análise espectral. (a) Série temporal original de diversidade de gêneros marinhos ao longo do Fanerozoico (linha preta), com a tendência polinomial de 3ª ordem ajustada (linha vermelha). A curva suavizada representa variações de longo prazo na diversidade, como o declínio observado após o Paleozoico. (b) Série de resíduos obtida pela subtração da tendência da série original (linha azul). Esta série contém apenas as flutuações em torno da tendência, que são a base da análise espectral. (c) Espectro de potência dos resíduos (linha preta), mostrando os picos identificados em aproximadamente 62 Myr e 140 Myr (círculos vermelhos). As linhas tracejadas coloridas representam o envelope espectral médio dos modelos nulos: azul para o Modelo R e vermelho para o Modelo W.

Os resultados reforçam a hipótese de que flutuações periódicas de longa escala podem estar embutidas nos padrões macroevolutivos de biodiversidade, possivelmente refletindo influências externas recorrentes, como mudanças tectônicas, climáticas ou astrofísicas. No entanto, a ausência de significância para o ciclo mais longo ressalta a importância de testes estatísticos rigorosos na interpretação de padrões espectrais.

A metodologia implementada mostrou-se eficaz na detecção de periodicidades em séries paleobiológicas e ilustra o papel da estatística computacional na inferência sobre processos evolutivos de longa duração.

Referências

- Rohde, Robert A e Richard A Muller (2005a). "Cycles in fossil diversity". Em: *Nature* 434.7030, pp. 208–210.
- (2005b). *Supplementary Information for "Cycles in Fossil Diversity"*. <https://www.nature.com/articles/nature03339>. *Nature*, Vol. 434, pp. 208–210 (2005), Supplementary Material.