

Detectando e Interpretando Sinais Periódicos em Séries Temporais Paleobiológicas

Computação IV - EP1

Maio de 2025

1 Introdução

A diversidade da vida ao longo do tempo geológico apresenta padrões complexos de variação, incluindo eventos de extinções em massa e recuperações evolutivas. Em 2005, Robert Rohde e Richard Muller [1] publicaram um estudo revelando um ciclo estatisticamente significativo de aproximadamente 62 milhões de anos na diversidade de gêneros marinhos ao longo do fanerozoico, utilizando dados do compêndio de Sepkoski [2]. Este ciclo representa uma excelente ocasião para aplicar técnicas de análise de séries temporais e inferência estatística.

Neste exercício, você deverá reproduzir, interpretar e discutir os resultados centrais do estudo, com foco na identificação de sinais periódicos em dados reais. O projeto propõe a utilização de análise espectral via transformada de Fourier, remoção de tendências por ajustes polinomiais e testes de significância com simulações de Monte Carlo, todos baseados na inferência frequencista.

Além de aprender ferramentas e técnicas para lidar com séries temporais, queremos que você reflita criticamente sobre o que significa detectar um padrão estatístico em um contexto científico: até que ponto podemos confiar nos dados? Como saber se um ciclo aparente é real ou fruto de vieses? Ao final deste trabalho, esperamos que você esteja mais preparado para navegar entre análises quantitativas e interpretações científicas cuidadosas.

2 Etapas do Projeto

I) Exploração Inicial

- Carregue os dados e visualize a série temporal da diversidade.
- Verifique tendências de longo prazo, picos e vales.

II) Remoção da Tendência

- Ajuste um polinômio de 3ª ordem aos dados para modelar a tendência de longo prazo.
- Subtraia essa tendência da série original para obter os resíduos (série sem tendência).

III) Análise espectral:

- Aplique a transformada de Fourier à série sem tendência.
- Plote o espectro de potência e identifique os picos relevantes.

IV) Construção de modelos nulos:

- Implemente simulações de Monte Carlo para gerar séries aleatórias:
Modelo R: embaralhamento dos incrementos (random walk).
Modelo W: embaralhamento de blocos (preserva autocorrelações locais).
- Para cada simulação, aplique a transformada de Fourier e guarde o espectro.

V) Teste de significância:

- Compare o pico de 62 Myr do espectro real com os espectros simulados.
- Estime um valor de p com base na fração de simulações que produziram picos iguais ou maiores.

VI) Discussão e Interpretação

- Os resultados obtidos corroboram a existência de um ciclo real?
- Quais são as limitações dessa análise?
- Você acredita que o sinal é de origem biológica, geofísica ou artefato de amostragem?

3 Instruções Gerais

3.1 Sobre a Entrega

- O trabalho deverá ser desenvolvido individualmente ou em duplas.
- A entrega deverá ser feita no Google Classroom.
- O arquivo entregue deverá ser uma pasta compactada .zip nomeada como: comp4_ep1_NOME (individual) ou comp4_ep1_NOME1_NOME2 (em duplas).

- Esta pasta deverá conter:
 1. `ep1.ipynb`: código comentado com a análise completa;
 2. `relatorio.pdf`: extraído do notebook com resultados e interpretação;
 3. `requirements.txt`: com as bibliotecas utilizadas;
 4. `README.md`: com instruções para reproduzir os resultados, referências e comentários gerais;
 5. `data/`: pasta contendo:
 - (a) `raw/`: arquivos originais utilizados (ex: `genera_data.csv`);
 - (b) `processed/`: arquivos gerados após qualquer tipo de transformação ou limpeza.

3.2 Sobre os Materiais Fornecidos

Os seguintes materiais estão disponíveis no Classroom como uma pasta `comp4_ep1.zip`:

- Artigo original: *Cycles in Fossil Diversity* (Rohde & Muller, Nature, 2005) [1]
- Dentro de `data`, há um arquivo de dados `genera_data.csv` contendo a série temporal de diversidade de gêneros marinhos.
- (Opcional) Dentro de `data`, há um arquivo complementar com diversidade separada por tempo de duração (`short_long_lived.csv`).
- Arquivo `ep1.ipynb` que vcs deverão utilizar para realizar o projeto.
- Um `README.md` com algumas instruções.

3.3 Sobre os Critérios de Avaliação

A avaliação do EP será feita com base nos seguintes critérios:

Bônus poderá ser atribuído para extensões adicionais bem implementadas, como comparação entre diferentes conjuntos de dados (short-lived vs long-lived), gráficos interativos (ex: plotly), repetir a análise com outros modelos de tendência, análise de sensibilidade dos modelos nulos.

Nota sobre a análise de sensibilidade dos modelos nulos:

Analisar a sensibilidade dos modelos nulos significa testar o quanto os resultados obtidos (especialmente os picos espectrais) mudam quando você altera os parâmetros do modelo nulo. Por exemplo, no modelo W, mudar o tamanho dos blocos embaralhados pode alterar o fundo espectral esperado. Se o pico de 62 Myr permanece significativo mesmo após várias configurações de simulação, ele é considerado robusto. Caso contrário, pode ser um artefato de como o modelo nulo foi construído.

Critério	Descrição	Peso
Clareza e organização do código	Código bem estruturado, comentado e reutilizável.	20%
Correção técnica	Aplicação adequada das etapas: ajuste polinomial, detrending, Fourier, simulações e cálculo do valor de p .	25%
Qualidade das visualizações	Gráficos informativos, legíveis e bem apresentados.	15%
Interpretação crítica dos resultados	Discussão sobre a significância estatística, limitações dos dados e possíveis origens do ciclo.	25%
Apresentação e comunicação	Clareza na redação da conclusão e justificativa das escolhas metodológicas.	15%

Tabela 1: Critérios de avaliação do EP.

Referências

- [1] Robert A Rohde and Richard A Muller. Cycles in fossil diversity. *Nature*, 434(7030):208–210, 2005.
- [2] John J Sepkoski Jr. A compendium of fossil marine animal genera. *Bulletins of American paleontology*, 363:1–560, 2002.