



ρ_{DCCA} e DMC_x^2 : Implementação, Otimização e Aplicações

Defesa de Tese

Discente: Fernando Ferraz Ribeiro

Orientador: Prof. Dr. Gilney Figueira Zebende

Coorientador: Prof. Dr. Juan Alberto Leyva Cruz

23/05/2024

PPGM UEFS - Feira de Santana, BA

Sumário

1. Introdução
2. Metodologia
3. Fundamentação Teórica
4. Resultados
5. Referências

Introdução

Sistemas Complexos

Este conjunto amplo de fenômenos é comumente identificado e agrupado por algumas de suas características: são formados pela contribuição de um conjunto (geralmente grande) de componentes (muitas vezes simples) que, interagindo, estruturam-se de forma auto-organizada, gerando resultados inesperados, que não podem ser previstos pelos estudos estatísticos e/ou matemáticos tradicionais dos elementos formadores do sistema.

Reconhecimento

Em 2021, a Academia Real das Ciências da Suécia concedeu metade do Prêmio Nobel de Física para Syukuro Manabe e Klaus Hasselmann, cujos estudos apresentam modelos complexos para a análise do clima. Em particular apontam uma correlação entre as emissões de dióxido de carbono e as mudanças climáticas.

Sistemas Complexos e Ciência de Dados

Muitos fenômenos complexos são investigados pela análise de grandes conjuntos de dados. É notável a velocidade e quantidade de dados que são gerados e armazenados pela humanidade atualmente. A aquisição, manipulação, gestão, armazenamento e criação de valor a partir de dados, através de ambientes computacionais, tem-se apresentado como um novo paradigma tecnológico. ...

Um campo do conhecimento que recebeu a denominação de **Ciência de Dados**, conceito que envelopa alguns termos frequentemente associados à inovação científica, técnica e social como **Big Data, mineração de dados, Business Intelligence internet das coisas, inteligência artificial e aprendizado de máquina(AM)**, dentre outros (EMC EDUCATION SERVICE, 2015, p. 12-13).

Séries Temporais

As séries temporais são definidas como um conjunto de observações (numéricas ou categóricas) ordenado no tempo. Embora muitos dos dados que descrevem as dinâmicas espaciais podem ser registrados na forma de séries temporais (abastecimento de água nas tubulações, consumo de energia elétrica nos imóveis, fluxos de pessoas e veículos pela cidade, casos de uma doença por dia, etc.), contudo as técnicas de medição de correlações, bem como a devida exploração destas para inferir novos conhecimentos, permanecem como perguntas abertas em muitas sub-áreas das ciências ambientais(Bermudez-Edo; BARNAGHI; MOESSNER, 2018).

Definição do Problema

O coeficiente DMC_x^2 é uma ferramenta promissora para a análise de correlações em séries temporais múltiplas. No entanto, sua aplicação ainda é limitada devido à falta de implementações computacionais eficientes e estudos que explorem seu potencial em diferentes contextos.

Objetivo Principal

Investigar a aplicabilidade do ρ_{DCCA} e do DMC_x^2 com o objetivo, não só de identificar correlações entre variáveis, mas também de entender o algoritmo, propor e aperfeiçoar ferramentas computacionais para a realização dos cálculos.

Objetivos Gerais

1. Testar os algoritmos e os métodos.
2. Implementar um algoritmo computacional geral para calcular o DMC_x^2 para qualquer número de séries temporais.
3. Testar as ferramentas implementadas em novas análises.

Premissas

1. O DMC_x^2 , pelas características de análise do método, pode ajudar a entender características de séries temporais e aplicado em problemas de diversas áreas.
2. O DMC_x^2 é uma generalização do método ρ_{DCCA} para múltiplas séries temporais.
3. O ρ_{DCCA} , em determinadas condições testadas, apresentou resultados mais interessantes (como melhor descrição dos fenômenos) que os apresentados pelo coeficiente de Pearson quando aplicado à séries temporais não estacionárias (WANG et al., 2013).

Hipóteses

1. É possível otimizar os cálculos do ρ_{DCCA} e do DMC_x^2 .
2. Com maior poder de cálculo a aplicação dos coeficientes seria potencializada.
3. Uma ferramenta adequada para o cálculo dos coeficientes seria um impulso na utilização e divulgação destes.

Metodología

Etapas do processo

Abordagem experimental.

Etapas:

- Revisão bibliográfica.
- Análise das ferramentas computacionais existentes.
- Proposta de uma nova ferramenta computacional.
- Implementação.
- Testes.
- Exploração das novas possibilidades de análise.

Fundamentação Teórica

1. **Cálculo da série integrada:** dada uma série temporal $\{x_i\}$, com i variando de 1 a N , a série integrada X_k é calculada por $X_k = \sum_{i=1}^k [x_i - \langle x \rangle]$, com k também variando de 1 a N ;
2. **Divisão da série em caixas:** a série integrada X_k é dividida em $(N - n)$ caixas de tamanho n (escala temporal), cada caixa contendo $n + 1$ valores, começando em i até $i + n$;
3. **Cálculo do ajuste polinomial:** para cada caixa, é calculado polinômio (geralmente de grau 1) que melhor se ajusta, obtendo $\tilde{X}_{k,i}$ com $i \leq k \leq (i + n)$;

4. **Cálculo da função f_{DFA}^2 para cada caixa:** para cada caixa de uma escala temporal é calculada a função de DFA pela expressão:

$$f_{DFA}^2(n, i) = \frac{1}{1+n} \sum_{k=i}^{i+n} (X_{k,i} - \tilde{X}_{k,i})^2;$$

5. **Cálculo da função de flutuação para uma escala temporal:** para todas as caixas de uma escala de tempo, o DFA é calculado como:

$$F_{DFA}(n) = \sqrt{\frac{1}{N-n} \sum_{i=1}^{N-n} f_{DFA}^2(n, i)};$$

6. **Análise em diferentes escalas temporais:** para diferentes escalas de tempo (n), com valores possíveis $4 \leq n \leq \frac{N}{4}$, é calculada a função F_{DFA} para encontrar uma relação entre $F_{DFA} \times n$

1. **Cálculo das séries integradas:** tomando duas séries temporais com a mesma extensão $\{x_i^{j1}\}$ e $\{x_i^{j2}\}$ com i variando de 1 a N , as séries integradas $X\alpha_k$ e $X\beta_k$ são calculadas por $X_k = \sum_{i=1}^k [x_i - \langle x \rangle]$ para cada série, com k também variando de i a N ;
2. **Divisão das séries em caixas:** as séries X_k^{j1} e X_k^{j2} são divididas em $(N - n)$ caixas de tamanho n (escala de tempo), cada caixa contendo $n + 1$ valores, começando em i até $i + n$;
3. **Cálculo dos ajustes de polinômios:** para cada caixa, um polinômio (geralmente de grau 1) melhor se ajusta, obtendo $\widetilde{X^{j1}}_{k,i}$ e $\widetilde{X^{j2}}_{k,i}$, para séries $\{x_i^{j1}\}$ e $\{x_i^{j2}\}$ respectivamente, com $i \leq k \leq (i + n)$;

4. **Cálculo da função f_{DCCA}^2 em cada caixa:** para cada caixa uma das $N - n$ caixas de uma mesma escala temporal a função é calculada por:

$$f_{DCCA}^2(n, i) = \frac{1}{1+n} \sum_{k=i}^{i+n} (X_{k,i}^{j1} - \widetilde{X_{k,i}^{j1}}) \times (X_{k,i}^{j2} - \widetilde{X_{k,i}^{j2}})$$

5. **Cálculo do DCCA para toda a escala temporal:** para todas as caixas de uma mesma escala temporal, o DCCA é calculado como:

$$F_{DCCA}(n) = \sqrt{\frac{1}{N-n} \sum_{i=1}^{N-n} f_{DCCA}^2(n, i)};$$

6. **Análise em diferentes escalas temporais:** para um número de escalas de tempo (n), com valores possíveis $4 \leq n \leq \frac{N}{4}$, o DCCA é calculado para encontrar uma relação entre $F_{DCCA} \times n$

ρ_{DCCA} - (ZEBENDE, 2011)

$$\rho_{DCCA}(n) = \frac{F_{DCCA \ (x^{j1}, \ x^{j2})}^2(n)}{F_{DFA \ (x^{j1})}(n) \times F_{DFA \ (x^{j2})}(n)} \quad (1)$$

$$DMC_x^2 \equiv \rho_{Y,X^j}(n)^T \times \rho^{-1}(n) \times \rho_{Y,X^j}(n) \quad (2)$$

$$\rho_{Y,X^j}(n)^T = [\rho_{Y,X^1}(n), \rho_{Y,X^2}(n), \dots, \rho_{Y,X^m}(n)] \quad (3)$$

$$\rho(n) = \begin{pmatrix} 1 & \rho_{X^1,X^2}(n) & \rho_{X^1,X^3}(n) & \dots & \rho_{X^1,X^m}(n) \\ \rho_{X^2,X^1}(n) & 1 & \rho_{X^2,X^3}(n) & \dots & \rho_{X_2,X^m}(n) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{X^m,X^1}(n) & \rho_{X^m,X^2}(n) & \rho_{X^m,X^3}(n) & \dots & 1 \end{pmatrix} \quad (4)$$

Resultados

Figure 1: (FILHO et al., 2023)

 ELSEVIER

Contents lists available at [ScienceDirect](#)

Physica A

journal homepage: www.elsevier.com/locate/physa



Statistical study of the EEG in motor tasks (real and imaginary)

F.M. Oliveira Filho ^{a,b}, F.F. Ribeiro ^{a,e}, J.A. Leyva Cruz ^c, A.P. Nunes de Castro ^d,
G.F. Zebende ^{c,*}



^a Earth Sciences and Environment Modeling Program, State University of Feira de Santana, Bahia, Brazil
^b SENAI CIMATEC University Center, Salvador, Bahia, Brazil
^c State University of Feira de Santana, Bahia, Brazil
^d Jorge Amado University Center, Salvador, Bahia, Brazil
^e Federal University of Bahia, Salvador, Bahia, Brazil

Algoritmo registrado

Figure 2: Registro de Software



Algoritmo registrado - 01

$$\begin{aligned} DMC_x^2 &= \left(\rho_{X_2, X_3}^2 \times \rho_{Y, X_1}^2 - \rho_{Y, X_1}^2 + \rho_{X_1, X_3}^2 \times \rho_{Y, X_2}^2 - \rho_{Y, X_2}^2 \right. \\ &\quad + 2 \times \rho_{X_1, X_2} \times \rho_{Y, X_1} \times \rho_{Y, X_2} - 2 \times \rho_{X_1, X_3} \times \rho_{X_2, X_3} \times \rho_{Y, X_1} \\ &\quad + \rho_{X_1, X_2}^2 \times \rho_{Y, X_3}^2 - \rho_{Y, X_3}^2 + 2 \times \rho_{X_1, X_3} \times \rho_{Y, X_1} \times \rho_{Y, X_3} \\ &\quad - 2 \times \rho_{X_1, X_2} \times \rho_{X_2, X_3} \times \rho_{Y, X_1} \times \rho_{Y, X_3} \\ &\quad - 2 \times \rho_{X_1, X_2} \times \rho_{X_1, X_3} \times \rho_{Y, X_2} \times \rho_{Y, X_3} \\ &\quad \left. + 2 \times \rho_{X_2, X_3} \times \rho_{Y, X_2} \times \rho_{Y, X_3} \right) \Bigg/ \\ &\quad \left(\rho_{X_1, X_2}^2 + \rho_{X_1, X_3}^2 + \rho_{X_2, X_3}^2 - 2 \times \rho_{X_1, X_2} \times \rho_{X_1, X_3} \times \rho_{X_2, X_3}^{-1} \right) \end{aligned} \tag{5}$$

Multi Cross-correlation Analysis in a Multi-channel EEG applied in Motor Activity (Real/Imaginary)

[⟨https://255ribeiro.github.io/Multi_Cross-correlation_EEG/⟩](https://255ribeiro.github.io/Multi_Cross-correlation_EEG/)

Artigo 02

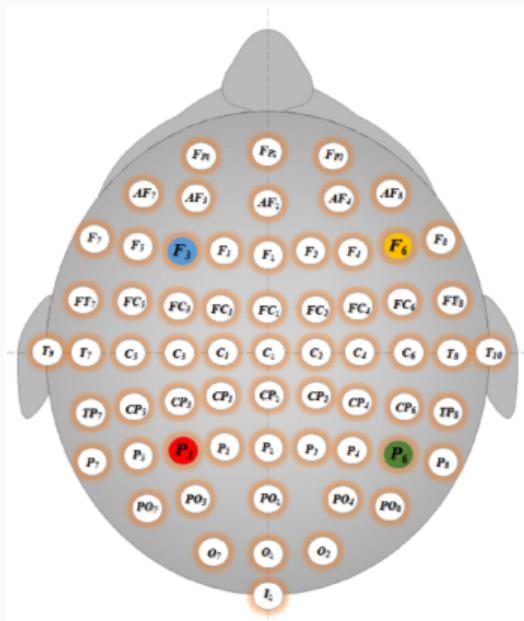


Figure 3: Posição dos canais de EEG e canais utilizados

Artigo 02

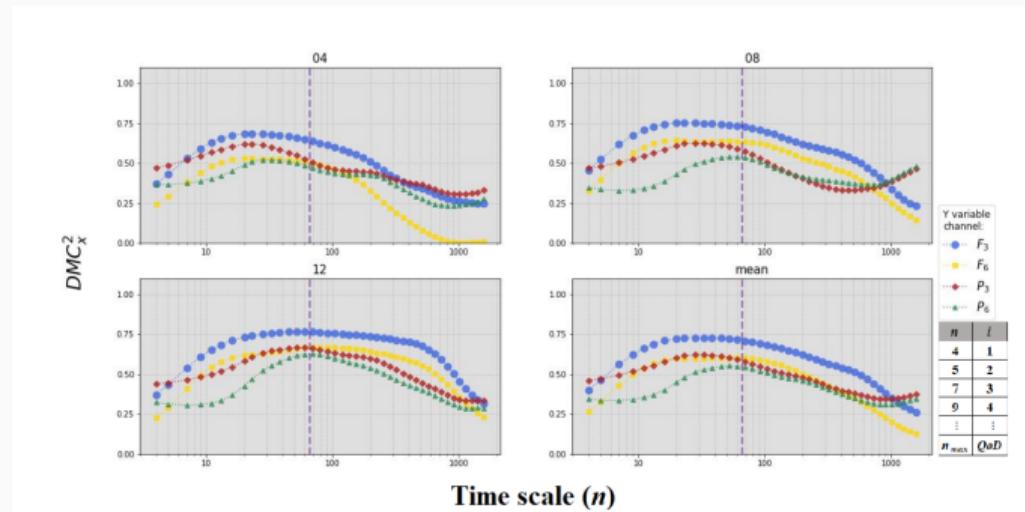


Figure 4: DMC_x^2 as a function of time scale n . Here is showing the results for subject S014 recordings for Task 2, presenting experiments 04, 08, 12 and the mean values for these experiments. The vertical line represents $n = 67$ and QoD is the total amount of time scales involved in DMC_x^2 calculations.

Artigo 02

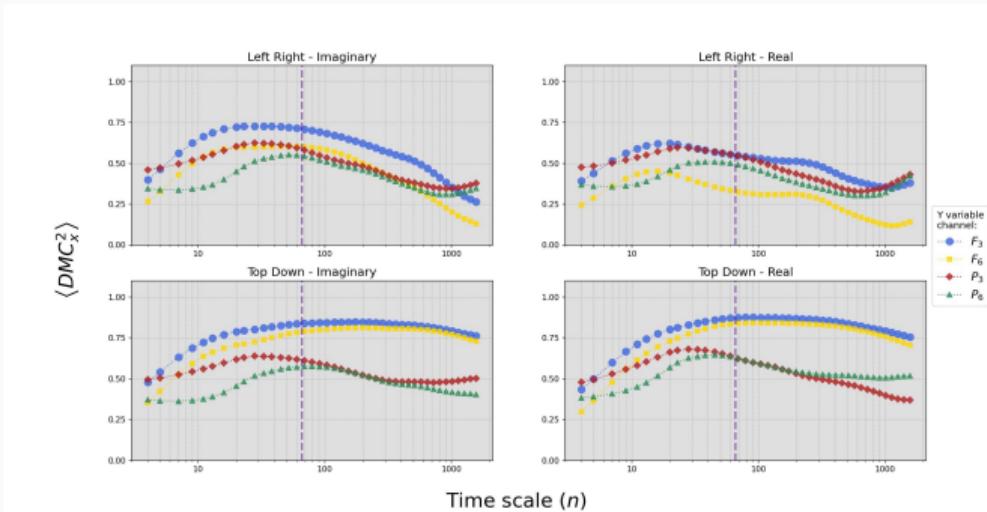


Figure 5: Mean values of $DMC_x^2 \times n$ for all Tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for the S014 subject.

Artigo 02

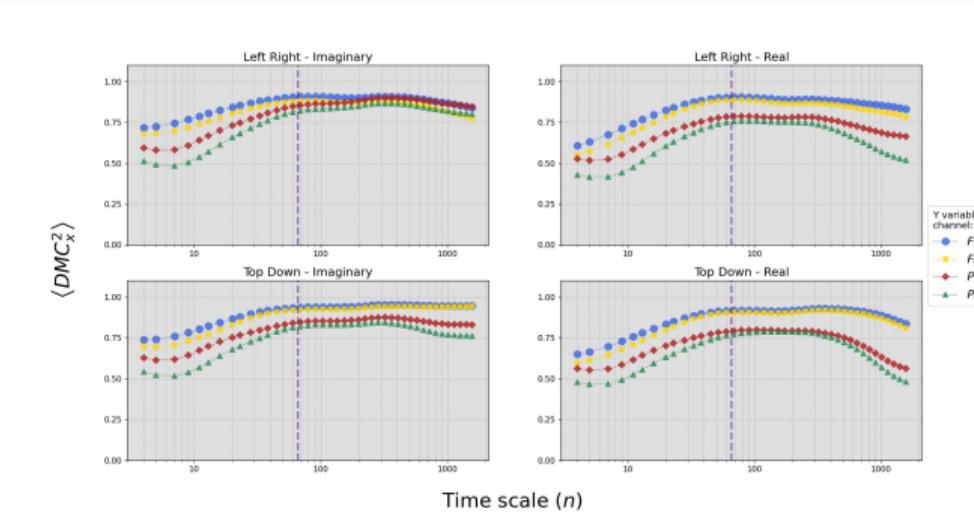


Figure 6: Mean values of $DMC_x^2 \times n$ for all Tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for the S036 subject.

Artigo 02

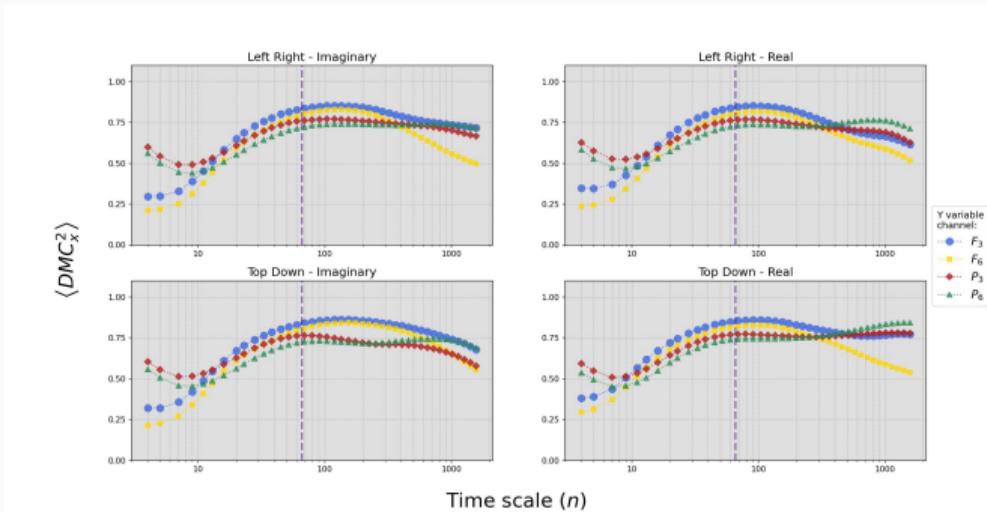


Figure 7: Mean values of $DMC_x^2 \times n$ for all Tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for the S039 subject.

Artigo 02

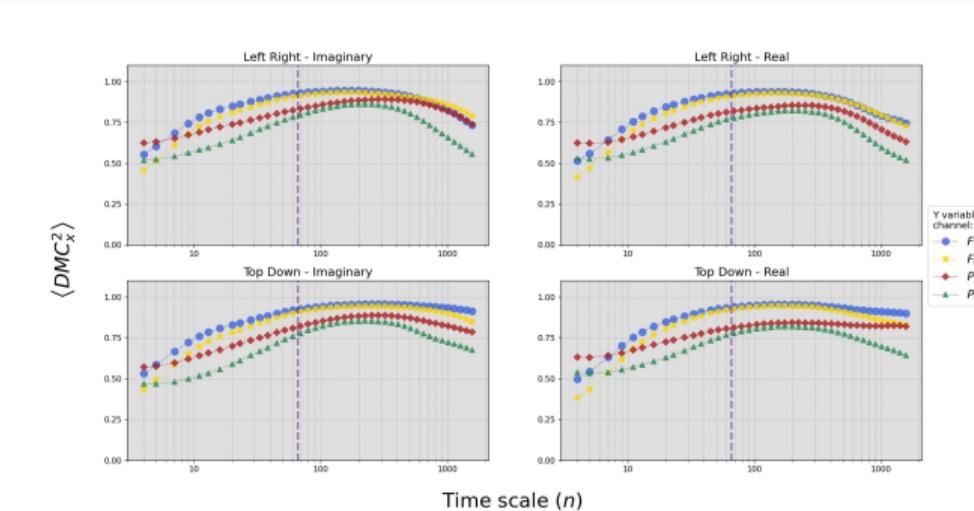


Figure 8: Mean values of $DMC_x^2 \times n$ for all Tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for the S078 subject.

Artigo 02

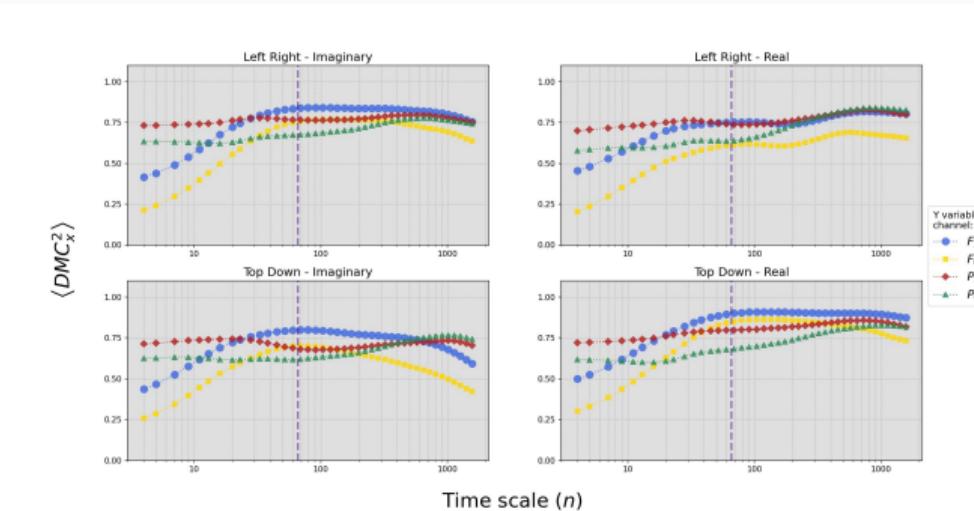


Figure 9: Mean values of $DMC_x^2 \times n$ for all Tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for the S099 subject.

Artigo 02

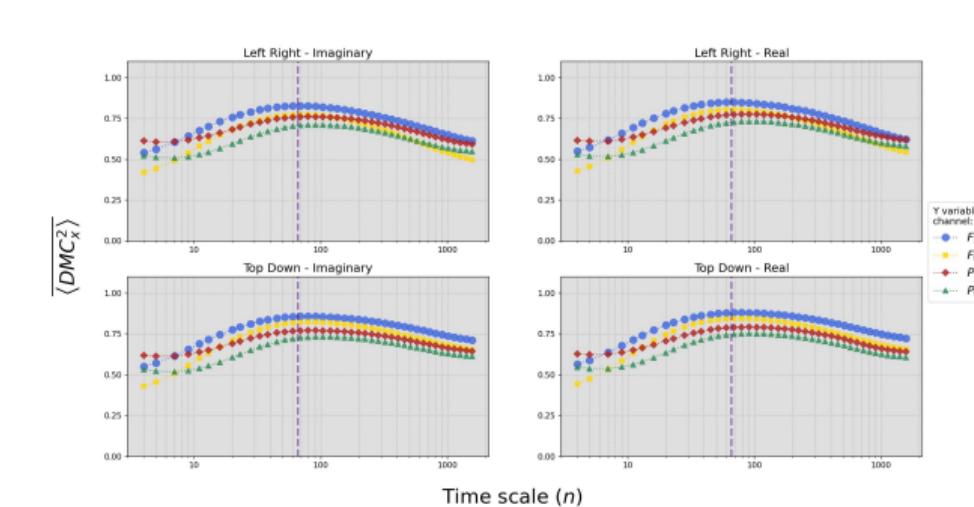


Figure 10: $DMC_x^2 \times n$ mean global for all Subjects and Tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real).

Artigo 02

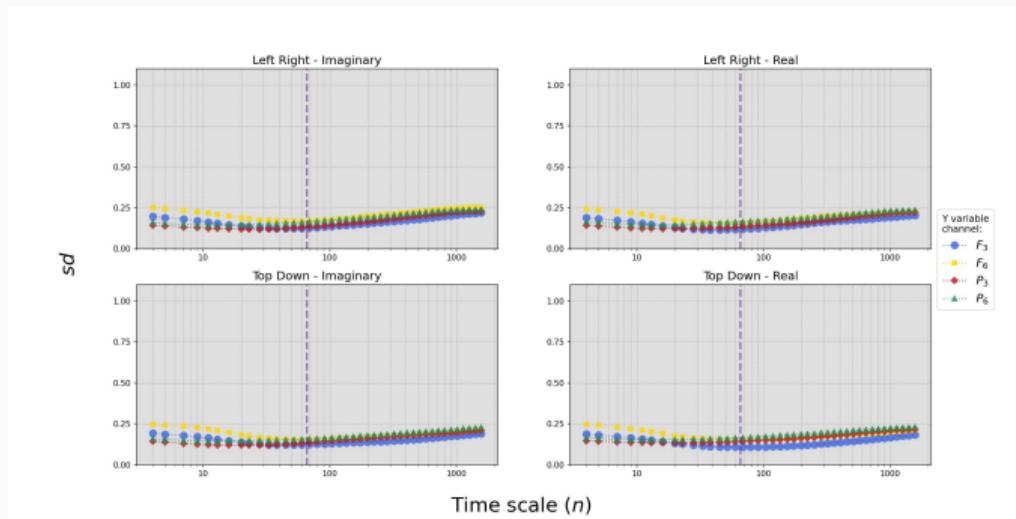


Figure 11: Standard deviation, *sd*, of the global mean for all Subjects and Tasks.

Artigo 02

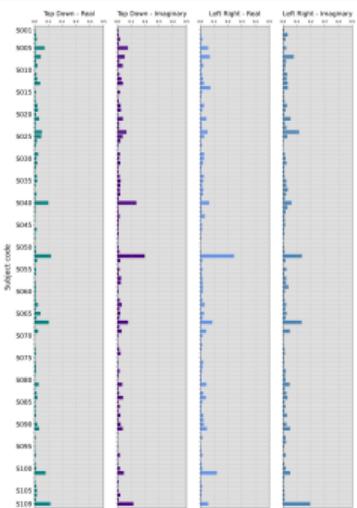


Figure 12: MSE
for the Channel F_3 .

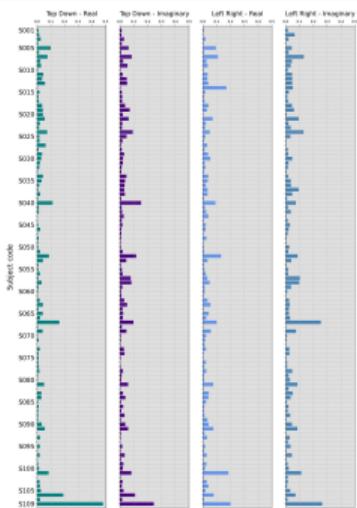


Figure 13: MSE
for the Channel F_6 .

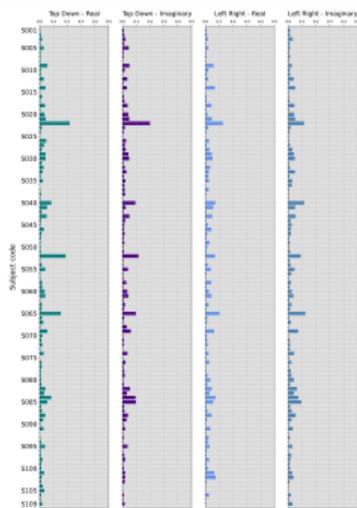


Figure 14: MSE
for the Channel P_3 .

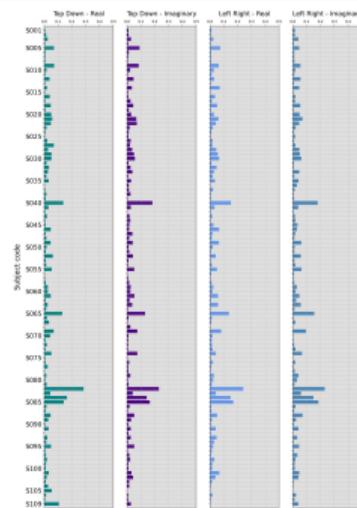


Figure 15: MSE
for the Channel P_6 .

Algoritmo - biblioteca Python/Zig

```
pip install zebende  
(https://pypi.org/project/zebende/)
```

Algoritmo - biblioteca Python/Zig

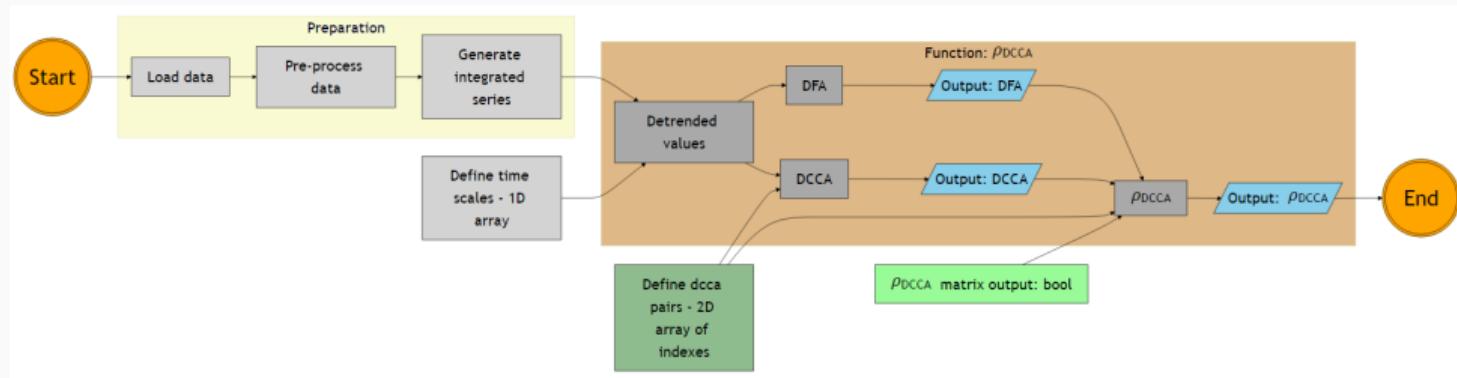


Figure 16: ρ_{DCCA} flowchart.

Algoritmo - biblioteca Python/Zig

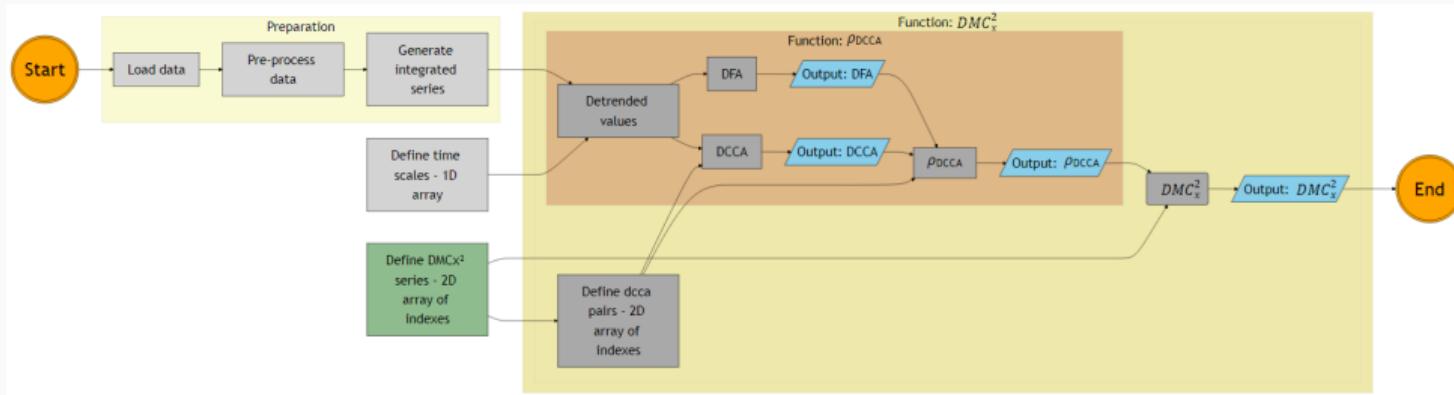


Figure 17: DMC_x^2 flowchart.

Otimização

Proposta por Hartmann et al. (2013) para o *DFA*, transposta para o *DCCA* e ρ_{DCCA} por Kaposzta et al. (2022).

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} T_k = \left(\sum_{j=i-1}^{(i+n)-1} T_j \right) - T_{i-1} + T_{i+n} \quad (6)$$

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} T_k^2 = \left(\sum_{j=i-1}^{(i+n)-1} T_j^2 \right) - T_{i-1}^2 + T_{i+n}^2 \quad (7)$$

Otimização

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} S_k = \left(\sum_{j=i-1}^{(i+n)-1} S_j \right) - S_{i-1} + S_{i+n} \quad (8)$$

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} (S_k \times T_k) = \left(\sum_{j=i-1}^{(i+n)-1} (S_j \times T_j) \right) - (S_{i-1} \times T_{i-1}) + (S_{i+n} \times T_{i+n}) \quad (9)$$

Otimização Detrended Saved i

- 4.a **Calculando o Detrended Value:** Para cada série X^j (incluindo a variável dependente) onde $1 \leq j \leq m$, sendo m o número de séries temporais; para cada escala temporal, em cada caixa i , calcula-se o valor de $DV_{k,i}^j = (X_{k,i}^j - \widetilde{X}_{k,i}^j)$ e armazena-se em uma matriz. A matrix tem por dimensões $m, n + 1$, onde cada linha corresponde a uma série temporal e as colunas correspondem ao número de pontos em cada caixa.
- 4.b **Cálculo da função f_{DFA}^2 para cada caixa:** Calcula-se o valor do DFA :

$$f_{DFA}^2(n, i) = \frac{1}{1+n} \sum_{k=i}^{i+n} (DV_{k,i}^{j1})^2;$$

- 4.c **Cálculo da função f_{DCCA}^2 em cada caixa:** com a matriz DV devidamente preenchida, para uma das $N - n$ caixas de uma mesma escala temporal a função é calculada para todas as combinações de séries temporais duas à duas por:

$$f_{DCCA}^2(n, i) = \frac{1}{1+n} \sum_{k=i}^{i+n} (DV_{k,i}^{j1}) \times (DV_{k,i}^{j2})$$

- 4.d **Retomando o algorítimo padrão:** Após o cálculo de todas as caixas, aplica-se o passo 5 do *DFA* e *DCCA*.

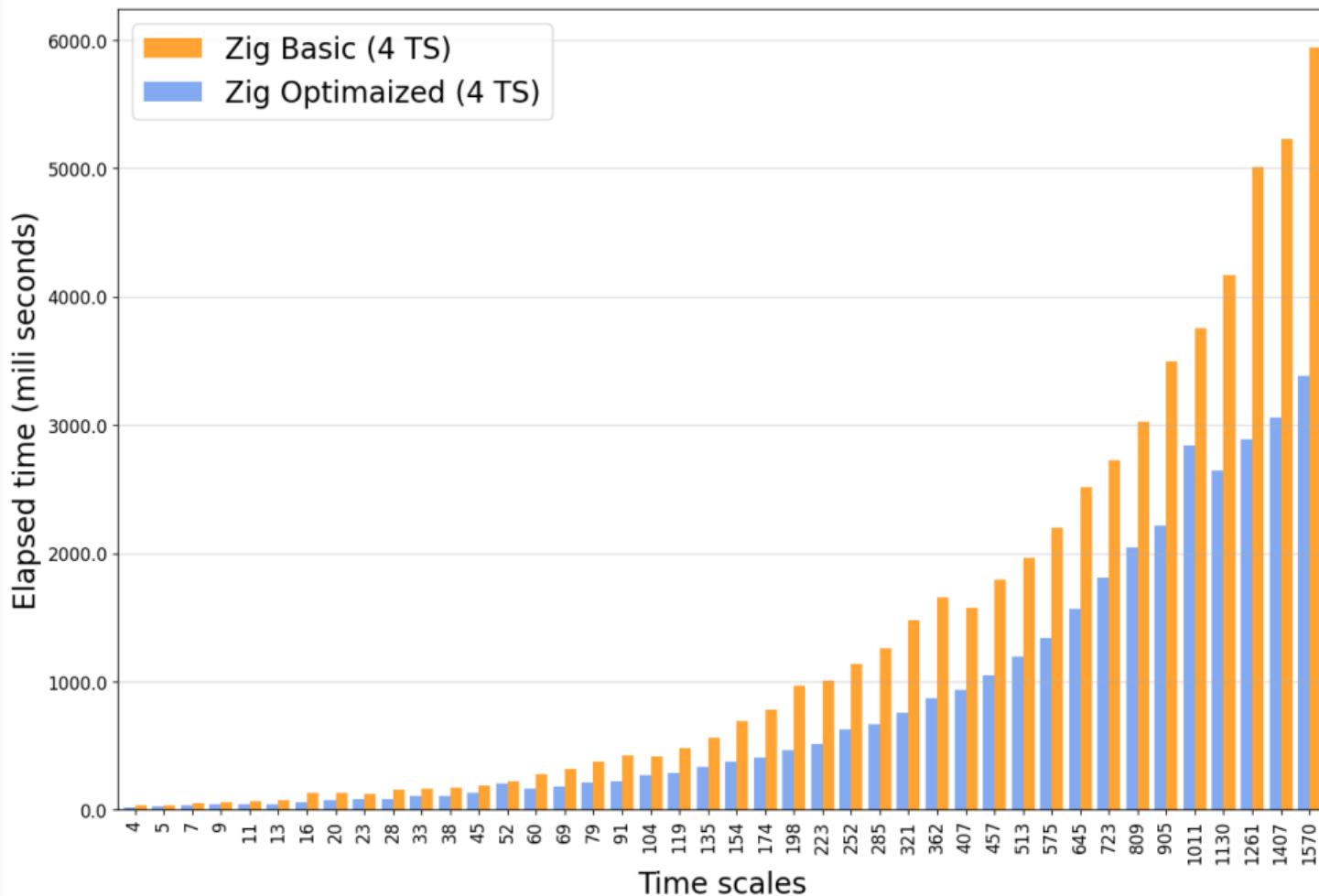


Table 1: Elapsed Time Comparison with four series

| Package | Implementation | Elapsed Time(s) | Performance Increase |
|---------|----------------|-----------------|----------------------|
| Zebende | Zig (opt) | 35.88 | - |
| Zebende | Zig (basic) | 59.74 | 66.50% |
| DCCA | Python | 447.67 | 1,147.69% |
| DCCA | R-Java | 2,206.2 | 6,048.83% |
| Zebende | Python | 3,076.23 | 8,473.66% |
| DFA | R | 28,116.00 | 78,261.20% |

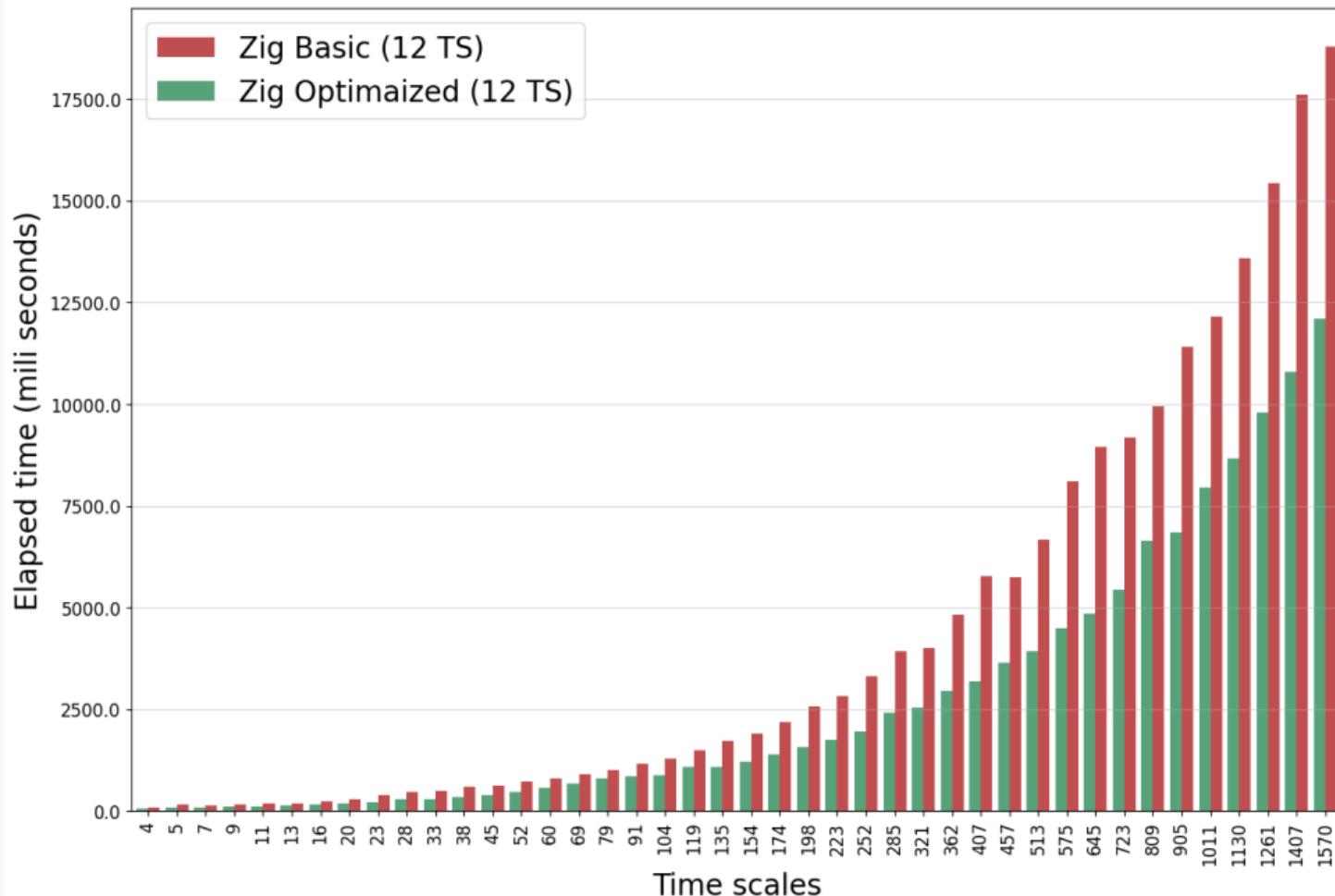


Table 2: Elapsed Time Comparison with twelve series

| Package | Implementation | Elapsed Time(s) | Performance Increase |
|---------|----------------|-----------------|----------------------|
| Zebende | Zig (opt) | 120.55 | - |
| Zebende | Zig (basic) | 213.08 | 76.64% |
| Zebende | Python | 3,761.40 | 3,020.20% |
| DCCA | Python | 4,675.17 | 3,778.20% |
| DCCA | R-Java | 28,980.00 | 23,939.82% |
| DFA | R | 309,276.00 | 256,454.13% |

Maximização do coeficiente DMC_x^2 utilizando matriz ρ_{DCCA} e $DPDCCA$

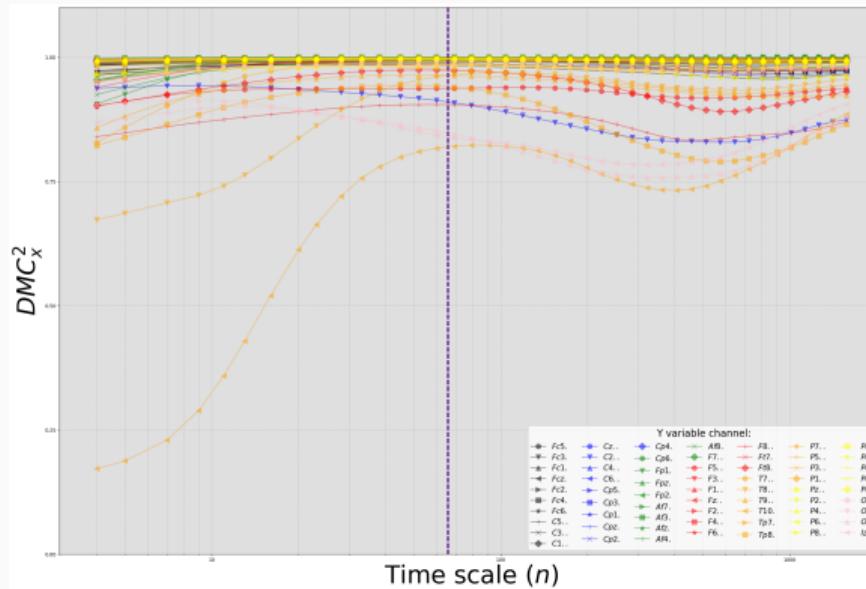


Figure 20: DMC_x^2 de todo os canais do experimento para cada canal como variável dependente.

Maximização do coeficiente DMC_x^2 utilizando matriz ρ_{DCCA} e $DPDCCA$

Table 3: Maximização do DMC_x^2 . $n = 4$, $count = 8$, referência= 0.6726

| Critério | canais selecionados | valor | percentual |
|-----------------|--------------------------------|--------|------------|
| ρ_{DCCA} | T8 Fc6 C6 Cp6 F6 F8 Ft8 P4 P6 | 0.6330 | 94.1108% |
| $ \rho_{DCCA} $ | T8 Fc6 C6 Cp6 F6 F8 Ft8 P4 P6 | 0.6330 | 94.1108% |
| $\Sigma DPDCCA$ | T8 Fc6 C6 Cp6 F2 F4 F6 F8 Ft8 | 0.6329 | 94.0908% |
| $DPDCCA$ | T8 C6 Cp6 Fp2 Afz F3 F4 Ft8 P6 | 0.5686 | 84.5341% |
| Random | T8 P6 P5 C1 F1 F4 Cp6 Fc1 F6 | 0.3138 | 46.6517% |

Maximização do coeficiente DMC_x^2 utilizando matriz ρ_{DCCA} e $DPDCCA$

Table 4: Maximização do DMC_x^2 . $n = 69$, $count = 8$, referência= 0.9643

| Critério | canais selecionados | valor | percentual |
|-----------------|---------------------------------|--------|------------|
| ρ_{DCCA} | T8 Fc4 Fc6 C4 C6 Cp6 F6 Ft8 Tp8 | 0.9581 | 99.3548% |
| $ \rho_{DCCA} $ | T8 Fc4 Fc6 C4 C6 Cp6 F6 Ft8 Tp8 | 0.9581 | 99.3548% |
| $\Sigma DPDCCA$ | T8 Fc6 C6 Cp6 Ft8 T10 Tp8 P6 P8 | 0.9597 | 99.5189% |
| $DPDCCA$ | T8 C6 Cp6 Ft8 T7 T10 Tp7 Tp8 P8 | 0.9572 | 99.2568% |
| Random | T8 Fp1 Po7 F7 P6 Cp3 C3 Po4 P2 | 0.8083 | 83.8165% |

Referências

 Bermudez-Edo, M.; BARNAGHI, P.; MOESSNER, K. Analysing real world data streams with spatio-temporal correlations: Entropy vs. Pearson correlation. *Automation in Construction*, v. 88, n. May 2017, p. 87–100, 2018. ISSN 09265805.

8

 EMC EDUCATION SERVICE. *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Indianapolis, Indiana, USA: JOHN WILEY \& SONS, 2015. ISBN 978-1-118-87605-3 1-118-87605-9 978-1-118-87613-8 1-118-87613-X 978-1-118-87622-0 1-118-87622-9.

7

 FILHO, F. O. et al. Statistical study of the EEG in motor tasks (real and imaginary). *Physica A: Statistical Mechanics and its Applications*, v. 622, p. 128802, jul. 2023. ISSN 03784371.

24

 HARTMANN, A. et al. Real-time fractal signal processing in the time domain. *Physica A: Statistical Mechanics and its Applications*, v. 392, n. 1, p. 89–102, jan. 2013. ISSN 03784371.

41

 KAPOSZTA, Z. et al. Real-Time Algorithm for Detrended Cross-Correlation Analysis of Long-Range Coupled Processes. *Frontiers in Physiology*, v. 13, p. 817268, mar. 2022. ISSN 1664-042X.

41

 PENG, C.-K. et al. Mosaic Organization of DNA Nucleotides. v. 49, n. 2, p. 1685–1689, 1994.

17, 18

 PODOBNIK, B.; STANLEY, H. E. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Physical Review Letters*, v. 100, n. 8, 2008. ISSN 00319007.

19, 20

 WANG, G. J. et al. Random matrix theory analysis of cross-correlations in the US stock market: Evidence from Pearson's correlation coefficient and detrended cross-correlation coefficient. *Physica A: Statistical Mechanics and its Applications*, Elsevier B.V., v. 392, n. 17, p. 3715–3730, 2013. ISSN 03784371.

12

 ZEBENDE, G. F. DCCA cross-correlation coefficient: Quantifying level of cross-correlation. *Physica A: Statistical Mechanics and its Applications*, Elsevier B.V., v. 390, n. 4, p. 614–618, 2011. ISSN 03784371.

21

 ZEBENDE, G. F.; SILVA, A. M. Detrended Multiple Cross-Correlation Coefficient. *Physica A*, Elsevier B.V., v. 510, p. 91–97, 2018. ISSN 0378-4371.

22

Obrigado
