



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM EM CIÊNCIAS  
DA TERRA E DO AMBIENTE

Doutorado em Modelagem em Ciências da Terra e do Ambiente

**Tese de doutorado**

$\rho_{DCCA}$  e  $DMC_x^2$  : Implementação, Otimização e  
Aplicações

Apresentada por: Fernando Ferraz Ribeiro  
Orientador: Prof. Dr. Gilney Figueira Zebende  
Co-orientador: Prof. Dr. Juan Alberto Leyva Cruz

Maio de 2025

Fernando Ferraz Ribeiro

# $\rho_{DCCA}$ e $DMC_x^2$ : Implementação, Otimização e Aplicações

Tese de doutorado apresentada ao Programa de Pós-graduação em Modelagem em Ciências da Terra e do Ambiente, Curso de Doutorado em Modelagem em Ciências da Terra e do Ambiente da UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA, como requisito parcial para a obtenção do título de **Doutor em Ciências Ambientais**.

Área de conhecimento: Estudos Ambientais e Geotecnologias

Orientador: Prof. Dr. Gilney Figueira Zebende

*UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA*

Co-orientador: Prof. Dr. Juan Alberto Leyva Cruz

*UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA*

Feira de Santana, BA  
UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA  
2025

## **Ficha Catalográfica – Biblioteca Central Julieta Carteado**

Ribeiro, Fernando Ferraz  
R369p      pDCCA e DMC<sup>2</sup><sub>χ</sub>; implementação, otimização e aplicações./  
Fernando Ferraz Ribeiro, 2025.  
106f.: il.

Orientador: Gilney Figueira Zebende  
Coorientador: Juan Alberto Leyva Cruz

Dissertação (mestrado) – Universidade Estadual de Feira de Santana. Programa de Pós-Graduação em Modelagem em Ciências da Terra e do Ambiente, 2025.

1.Séries temporais. 2.DMC<sup>2</sup><sub>χ</sub>. 3.pDCCA. 4.Ciência de dados.  
5.Algorítimos. 6.Otimização. I.Zebende, Gilney Figueira, orient.  
II.Cruz, Juan Alberto Leyva, coorient. IV.Universidade Estadual de Feira de Santana. V.Título.

CDU : 519:004.43

Maria de Fátima de Jesus Moreira – Bibliotecária – CRB5/1120

---

## **Nota sobre o estilo do PPGM-UEFS**

---

Esta tese de doutorado foi elaborada considerando as normas de estilo propostas e aprovadas pelo colegiado do Programa de Pós-graduação em Modelagem em Ciências da Terra e do Ambiente e estão disponíveis no formato eletrônico (<http://ppgm.uefs.br/banco-de-dissertacoes>) ou no formato impresso para consulta.

# UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Programa de Pós-graduação em Modelagem em Ciências da Terra e do Ambiente

Doutorado em Ciências da Terra e do Ambiente

A Banca Examinadora, constituída pelos professores listados abaixo, leu e recomenda a aprovação da Tese de Doutorado, intitulada “ $\rho_{DCCA}$  e  $DMC_x^2$ : Implementação, Otimização e Aplicações”, apresentada no dia 23 de maio de 2025, como requisito parcial para a obtenção do título de **Doutor em Modelagem em Ciências Ambientais**.

Orientador:

---

Prof. Dr. Gilney Figueira Zebende

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Co-Orientador:

---

Prof. Dr. Juan Alberto Leyva Cruz

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Membro externo da Banca:

---

Prof. Dr. Thadeu Josino Pereira Penna

UNIVERSIDADE FEDERAL FLUMINENSE

Membro externo da Banca:

---

Prof. Dr. Jés de Jesus Fiais Cerqueira

UNIVERSIDADE FEDERAL DA BAHIA

Membro externo da Banca:

---

Prof. Dr. Alessandro Guerra Cerqueira

UNIVERSIDADE FEDERAL DA BAHIA

Membro externo da Banca:

---

Prof. Dr. Alex Ferreira dos Santos

UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA

Dedico este trabalho a:

Minha mãe, **Aidil Nunes Ferraz Ribeiro**,  
pelo amor e apoio sempre;  
A meu pai [in memoriam], **Ubirajara Gomes Ribeiro**,  
pelo amor e por ter me mostrado as belezas da ciência;  
A meu irmão, **Rogério Ferraz Ribeiro**,  
pelo amor e por todos os momentos que dividimos;  
A minha esposa, **Carolina de Britto Fernandes**,  
pelo amor e cumplicidade na vida;  
A meu filho, **Gustavo Fernandes Ribeiro** e  
meu sobrinho, **Gabriel Carvalho Ferraz**,  
por serem os caminhos pelos quais este amor continua.

---

## Agradecimentos

---

Ao Prof. Dr. Gilney Figueira Zebende , pela oportunidade de colaborar com uma pesquisa tão interessante. Durante todo o processo minha admiração por você e pelo seu trabalho só cresceu. Ao Coorientador Prof. Dr. Juan Alberto Leyva Cruz pelas orientações e por ter sempre acreditado no nosso trabalho. A todos os membros da banca, pelos comentários, apontamentos, elogios e críticas. A todos os colegas de turma, pelos momentos compartilhados, ainda que circunstâncias do nosso tempo tenha pesado no que diz respeito ao contado presencial.

Agradeço a minha mãe, Aidil Nunes Ferraz Ribeiro, por todo apoio que recebi e ainda recebo na vida; Para minha esposa Carolina de Britto Fernandes pelo esforço para que eu conseguisse tempo para me dedicar à pesquisa; a minha madrinha, Prof. Dra. Sandra Maria Ferraz Mello, pelos conselhos acadêmicos e pelos meus momentos de desabafo; a meu irmão e amigo Rogério Ferraz Ribeiro e sua família, Helena e Gabriel, pelo apoio e torcida; meus sogros, Alfredo e Cláudia Fernandes, pelo incentivo e a grande ajuda com Gustavo que tornaram essa caminhada mais fácil. A meu filho Gustavo, por todas as nossas brincadeiras.

Muitos colegas de trabalho, ensino e pesquisa acompanharam de perto essa caminhada, no começo, meio e fim: Basílio Fernandez, Bruno Leão, Aline de Carvalho, Carlos Alberto Andrade Bomfim e Maurício Felzemburgh também estavam desenvolvendo seus respectivos doutorados neste momento, agradeço pela troca de impressões, torcida e incentivo. Outros dividiram comigo lembranças e reflexões sobre esse momento: Lídia Quièto Viana, Natalie Johanna Groetelaars, Ariane Prevedello Rubin, Luara Batalha, Larissa Scarano, Marcos Rodrigues, Everaldo Guedes e Paolo Bruni, dentre muitos outros que me ouviram e ajudaram em momentos importantes deste processo. Agradeço também aos colegas pesquisadores do LCAD e do Grupo de estudos de sistemas complexos, em particular Florêncio Mendes Oliveira Filho e Andréa de Almeida Brito que acompanharam a maioria das reuniões de orientação, sempre colaborando e apoiando.

Aos amigos de todas as horas: Alessandro Soares, Alexis Guedes, Maurício Cerqueira, Márcio Targa, Juliano Pontes, Fábio Koleski, dentre muitos outros, pelos desabafos que ouviram e pelas boas perspectivas que apontaram.

Por fim agradeço a UEFS e ao PPGM, seus professores e servidores, pelo conhecimento e crescimento, profissional e pessoal, que aqui ganhei.

Feira de Santana, BA, Brasil  
23 de Maio de 2025

Fernando Ferraz Ribeiro

---

## Resumo

---

Esta Tese apresenta uma investigação sobre análise de séries temporais através das funções  $DFA$ ,  $DCCA$  e, principalmente nos coeficientes  $\rho_{DCCA}$  e  $DMC_x^2$ . Apresenta uma pesquisa bibliográfica focada em métodos baseados no  $DFA$  que tratam de multi-correlação e as aplicações destas funções e coeficientes nas Ciências Ambientais. Apresentamos também artigos, produzidos no processo desta tese, tratando da aplicação dos coeficientes  $\rho_{DCCA}$  e  $DMC_x^2$ .

A implementação de uma ferramenta computacional (a biblioteca *Python Zebende*) para a manipulação das séries temporais e cálculo das funções e coeficientes é apresentada, assim como o algoritmo *Detrended Saved*, uma estratégia inovadora no cálculo do  $DCCAd$ e forma geral, mas com vantagens no desempenho quando utilizado para cálculos de muitas séries entre si (como no caso da montagem da matriz do  $\rho_{DCCA}$  para calcular o  $DMC_x^2$ , quando um grande número de séries temporais é utilizado). A capacidade da biblioteca *Zebende* potencializar a utilização destes coeficientes também é apresentada.

**Palavras Chaves:** Séries Temporais,  $DMC_x^2$ ,  $\rho_{DCCA}$ , Ciência de Dados, Algorítimos, otimização

---

## Abstract

---

This thesis presents an investigation into the analysis of time series through the functions  $DFA$ ,  $DCCA$  and, mainly, the coefficients  $\rho_{DCCA}$  and  $DMC_x^2$ . It presents a bibliographic research focused on methods based on  $DFA$  that treat multi-correlation and the applications of these functions and coefficients in Environmental Sciences. We also present articles produced in the process of this thesis, dealing with the application of the coefficients  $\rho_{DCCA}$  and  $DMC_x^2$ .

The implementation of a computational tool (the *Python Zebende* library) for manipulating time series and calculating the functions and coefficients is presented, as well as the *Detrended Saved* algorithm, an innovative strategy in the general calculation of  $DCCA$ , but with advantages in performance when used for calculations of many series among themselves (such as in the case of matrix construction of  $\rho_{DCCA}$  to calculate  $DMC_x^2$ , when a large number of time series is used). The capacity of the *Zebende* library to enhance the use of these coefficients is also presented.

**Keywords:** Time Series,  $DMC_x^2$ ,  $\rho_{DCCA}$ , Data Science, Algorithms, Optimization

---

# Sumário

---

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Definição do problema . . . . .	2
1.2	Objetivos . . . . .	3
1.3	Importância da Pesquisa . . . . .	4
1.4	Viabilidade e Limitações . . . . .	4
1.5	Questões e Hipóteses . . . . .	5
1.6	Metodologia . . . . .	6
1.7	Organização da Tese . . . . .	7
<b>2</b>	<b>Fundamentação Teórica e Metodologia</b>	<b>8</b>
2.1	Definições preliminares . . . . .	8
2.2	Métodos de Análise de Séries Temporais: <i>DFA</i> , <i>DCCA</i> , $\rho_{DCCA}$ e $DMC_x^2$ . . . . .	10
2.3	Fractais e os métodos baseados no <i>DFA</i> . . . . .	14
2.4	Variantes de correlação cruzada baseados no <i>DCCA</i> . . . . .	16
2.5	Aplicações das funções e coeficientes . . . . .	17
<b>3</b>	<b>Produção Científica</b>	<b>19</b>
3.1	Artigo 01: DCCA multi cross-correlation analysis applied on EEG signals to study motor activity (Real/Imaginary) . . . . .	19
3.2	Artigo 02: A Python/Zig optimized and customizable implementation for $\rho_{DCCA}$ and $DMC_x^2$ coefficients . . . . .	43
3.3	Maximização do coeficiente $DMC_x^2$ utilizando matriz $\rho_{DCCA}$ e <i>DPDCCA</i> . . . . .	67
3.3.1	Metodologia . . . . .	69
3.3.2	Resultados . . . . .	71
3.3.3	Conclusão . . . . .	73
<b>4</b>	<b>Conclusão</b>	<b>74</b>
4.1	Validação das hipóteses . . . . .	74
4.2	Trabalhos futuros . . . . .	75
<b>A</b>	<b>Anexo A</b>	<b>76</b>
<b>B</b>	<b>Anexo B: Publicações técnicas</b>	<b>86</b>
	<b>Referências</b>	<b>88</b>

---

## Lista de Tabelas

---

3.1	Maximização do $DMC_x^2$ . $n = 4$ , $count = 8$ , referência= 0.6726 . . . . .	71
3.2	Maximização do $DMC_x^2$ . $n = 69$ , $count = 8$ , referência= 0.9643 . . . . .	71

---

## Lista de Figuras

---

2.1	Análise <i>DFA</i> de sequências de DNA e sequências de controle, resultados apresentando correlação de longo alcance (+ e o) e não apresentando correlação. (x e □) Fonte: (PENG et al., 1994a)	15
3.1	Desvio padrão dos valores do $DMC_x^2$ para cada uma das tarefas. Fonte: Elaborada pelos autores.	20
3.2	Médias dos valores do $DMC_x^2$ para cada uma das tarefas. Fonte: Elaborada pelos autores	20
3.3	Mediana dos valores do $DMC_x^2$ para cada uma das tarefas. Fonte: Elaborada pelos autores	21
3.4	$DMC_x^2$ de todo os canais do experimento[1:63] para cada canal como variável dependente. Fonte: Elaborada pelos autores	67
3.5	Matriz de correlação do $\rho_{DCCA}$	68
3.6	$\rho_{DCCA} n = 4$	68
3.7	$\rho_{DCCA} n = 69$	68
3.8	Matriz de correlação do $\rho_{DCCA}$ . Fonte: Elaborada pelos autores	68
3.9	$DPDCCA n = 4$	69
3.10	$DPDCCA n = 69$ . Fonte: Elaborada pelos autores	69
3.11	Matriz de correlação do $DPDCCA$ . Fonte: Elaborada pelos autores	69
3.12	Channel: T8 $n = 4$	72
3.13	Channel: Tf8 $n = 69$	72
3.14	Gráfico cumulativo do $DMC_x^2$ - Channel: T8 $n = 69$ . Fonte: Elaborada pelos autores.	72
3.15	Channel: Af3 $n = 4$	72
3.16	Channel: Af3 $n = 69$	72
3.17	Gráfico cumulativo do $DMC_x^2$ - Channel: Af3 $n = 69$ . Fonte: Elaborada pelos autores.	72

---

## Lista de Algoritmos

---

1	<i>DFA</i>	10
2	<i>DCCA</i>	12
3	Detrended Saved	44
4	$\Sigma DPDCCA$	70

---

## Lista de Siglas

---

UEFS .....	Universidade Estadual de Feira de Santana
PPGM .....	Programa de Pós-Graduação em Modelagem em Ciências da Terra e do Ambiente
CAPES .....	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
<i>DCCA</i> .....	<i>Detrended Cross-Correlation Analysis</i>
<i>DFA</i> .....	<i>Detrended Fluctuation Analysis</i>
<i>DMC<sub>x</sub><sup>2</sup></i> .....	<i>Detrended Multiple Cross-Correlation Coefficient</i>
$\rho_{DCCA}$ .....	<i>Detrended Cross-Correlation Coefficient</i>
<i>DPDCCA</i> ..	<i>Detrended partial Cross-Correlation Analysis</i>
EEG .....	Eletroencefalograma

## Introdução

“Ordinary life is pretty complex stuff.”

(Harvey Pekar)

Os sistemas complexos compreendem um campo interdisciplinar da ciência, cujo escopo de problemas que trata não possui uma definição exata. 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.

Na área dos estudos ambientais, os sistemas complexos aparecem em problemas de diversas naturezas: sistemas de transportes, redes de energia e comunicação, organizações sociais e econômicas, densidade e ocupação humana do espaço, dentre outras. Os estudos do clima e de variáveis meteorológicas ocupam um espaço de particular relevância na intercessão entre os estudos ambientais e os sistemas complexos. 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 apontando para uma correlação entre as emissões de dióxido de carbono e as mudanças climáticas.

Muitos fenômenos complexos são investigados pela análise de grandes conjuntos de dados. É notável a velocidade e quantidade de dados que são gerados, obtidos 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 (IA) e aprendizado de máquina (AM), dentre outros ([EMC EDUCATION SERVICE, 2015](#), p. 12-13).

Uma das formas em que os dados costumam ser organizados são as denominadas 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](#)).

## 1.1 Definição do problema

As séries temporais registram aspectos fundamentais de uma gama de problemas, de diversas áreas que, devidamente abordados, podem clarificar características e comportamentos dos fenômenos estudados.

Uma das ferramentas utilizadas para analisar correlações entre séries temporais é o coeficiente  $\rho_{DCCA}$ , que permite quantificar a relação entre duas séries, mesmo quando estas apresentam tendências não estacionárias. O  $\rho_{DCCA}$  é especialmente útil em contextos onde os métodos tradicionais, como o coeficiente de Pearson, não são adequados devido à presença de tendências ou ruídos de longa duração nas séries analisadas.

O coeficiente  $DMC_x^2$  é uma ferramenta promissora para a análise de correlações em séries temporais múltiplas, permitindo a identificação de padrões complexos que não são capturados por métodos tradicionais. O cálculo deste coeficiente envolve o cálculo do  $\rho_{DCCA}$  de todas as múltiplas séries em combinação duas à duas.

Os cálculos do  $\rho_{DCCA}$  e por extensão, do  $DMC_x^2$  passam pelos cálculos do  $DCCA$  e  $DFA$  em determinadas etapas, que carregam boa parte do custo computacional do processo todo.

As publicações que fazem estudos com o  $\rho_{DCCA}$ , tem ao seu dispor algumas opções de ferramentas computacionais com esses cálculos, que basicamente pertencem a dois grupos: os implementados em linguagens interpretadas, como Python e R, que apresentam tempos de computação maiores, por outro lado, maior facilidade de manipular e visualizar as entradas e resultados, os implementados linguagens de mais baixo nível, com maior velocidade, mas permitido menos flexibilidade no pre-processamento e na análise dos resultados, sejam esses gráficos ou numéricos. Esses últimos geralmente são desenvolvidos internamente nos grupos de pesquisa e pouco compartilhados entre os pares.

No caso do  $DMC_x^2$  não existem pacotes disponíveis. Este algoritmo, pela necessidade de calcular a combinação dois a dois de todas as séries envolvidas, trás uma nova complexidade para o cálculo e maior custo computacional. As características do cálculo deste coeficiente podem ser estudadas e benefícios podem ser extraídos dos conhecimentos ad-

quiridos no processo.

Além disso, não está disponível uma implementação geral do  $DMC_x^2$  para qualquer número de séries temporais. Os artigos publicados até então utilizam 3 séries apenas. Fazendo com que os pesquisadores tenham que implementar o código para cada número de séries temporais. Isso limita e traz lentidão ao processo.

A integração com eco-sistemas de cálculo estatístico como os do Python e do R é fundamental para o trabalho da ciência de dados de hoje, assim como o desempenho, principalmente para lidar com grandes conjuntos de dados.

Em resumo, por um lado temos coeficientes reconhecidos pela comunidade científica, aplicados em problemas diversos, mas sem ferramentas que consigam ao mesmo tempo entregar um desempenho que viabilize trabalhar com múltiplas séries de longa duração e uma integração com os principais eco-sistemas de ciência de dados.

Essa pesquisa procura entregar uma solução eficiente para o cálculo do  $DFA$ ,  $DCCA$ ,  $\rho_{DCCA}$  e  $DMC_x^2$ , totalmente integrada com as ferramentas do Python de manipulação de dados, matemática, estatística, inteligência artificial e visualização de dados, resolvendo esse diálogo de forma eficiente. Procurando melhorar a velocidade de processamento, principalmente para o processamento de múltiplas séries temporais; oferecendo uma implementação que possa ser aplicada para qualquer número de séries temporais.

## 1.2 Objetivos

Neste trabalho, tem-se por objetivo investigar a aplicabilidade do  $\rho_{DCCA}$  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.

Como objetivos gerais foram elencados:

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.

### 1.3 Importância da Pesquisa

A pesquisa que propõe melhorias no algoritmo e a implementação de uma ferramenta para o cálculo do  $DCCA$ ,  $\rho_{DCCA}$  e  $DMC_x^2$  é de grande relevância, pois aborda um problema computacional desafiador e ainda pouco explorado. A eficiência no cálculo destes métodos é essencial para viabilizar sua aplicação em conjuntos de dados de grande escala, que são cada vez mais comuns em diversas áreas do conhecimento. Além disso, a disponibilização de uma ferramenta computacional robusta e acessível pode democratizar o uso do método, permitindo que pesquisadores de diferentes áreas possam utilizá-lo sem a necessidade de desenvolver suas próprias implementações.

A proposta de melhorias no algoritmo também tem implicações diretas na precisão e na escalabilidade das análises realizadas com o  $DMC_x^2$ . Métodos computacionalmente custosos podem limitar o tamanho dos conjuntos de dados analisados ou mesmo inviabilizar estudos em contextos onde o tempo de processamento é crítico. Ao otimizar o cálculo do  $DMC_x^2$ , a pesquisa contribui para superar essas limitações, abrindo caminho para novas aplicações e para a exploração de padrões complexos em séries temporais múltiplas.

No contexto das ciências ambientais, as implicações são particularmente significativas. A análise de séries temporais de variáveis climáticas e meteorológicas é fundamental para entender fenômenos como mudanças climáticas, padrões de precipitação e distribuição de energia solar. As redes elétrica, de água e de dados são fontes de informações únicas sobre a dinâmica de uma cidade. A mobilidade de pessoas e mercadorias são problemas desafiadores para gestores, planejadores e empresas privadas. Séries temporais também são exploradas no contexto de dados sísmicos, tanto para antevisão e planejamento diante de ua catástrofe natural, quanto para o mapeamento do sub-solo e identificação de recursos geológicos. Uma ferramenta eficiente para o cálculo do  $DMC_x^2$  pode auxiliar na identificação de correlações entre variáveis ambientais ou com forte implicação no espaço e nas dinâmicas locais, contribuindo para o desenvolvimento de modelos analíticos mais precisos e para a formulação de políticas públicas baseadas em dados confiáveis.

### 1.4 Viabilidade e Limitações

A implementação e melhoria dos algoritmos propostos apresentam alta viabilidade, uma vez que não demandam recursos computacionais excepcionais. A maior parte das operações envolvidas, como manipulação de matrizes e cálculos estatísticos, pode ser realizada em computadores domésticos de configurações medianas, resultando em um baixo investimento financeiro.

A viabilidade como análise de hipóteses existe. Certamente limitações como o peso computacional são mais difíceis de antever, mas devem ser levadas em conta desde o princípio.

A obtenção de conjuntos de dados reais e confiáveis é um desafio significativo que pode limitar o progresso de pesquisas como esta. A disponibilidade de dados de qualidade frequentemente depende de fatores externos, como acesso a fontes restritas, custos associados à aquisição ou limitações técnicas na coleta e armazenamento.

De resto, vale lembrar que a honestidade do trabalho científico pode levar a resultados que validam, total ou parcialmente, ou invalidam as conjecturas iniciais. Apenas com a atenta avaliação dos experimentos realizados pode-se entender os ganhos obtidos no percurso. A análise dos dados meteorológicos será aplicada para descrever esse percurso, independente do status da validação dos resultados.

## 1.5 Questões e Hipóteses

Este projeto foi baseado em algumas 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  $\rho_{DCCA}$  para múltiplas séries temporais.
3. O  $\rho_{DCCA}$ , 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](#)).

Partindo destas premissas, procuramos responder perguntas basilares:

1. Poderia-se criar uma ferramenta computacional eficaz para o cálculo dos coeficientes?
2. Poderia-se encontrar uma alternativa mais eficaz o cálculo dos coeficientes, principalmente o recente  $DMC_x^2$ ?
3. A implementação e melhorias, se bem sucedidas, poderiam ampliar o número de possíveis investigações em que os coeficientes podem ser aplicados?

Para orientar o trabalho, as seguintes hipóteses foram formuladas:

1. É possível otimizar os cálculos do  $\rho_{DCCA}$  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.

## 1.6 Metodologia

A metodologia utilizada neste projeto insere-se dentro da abordagem metodológica experimental, estruturada em três etapas principais.

A primeira etapa consiste em uma revisão bibliográfica abrangente, com o objetivo de compreender os fundamentos teóricos e metodológicos relacionados ao *DFA*, *DCCA*,  $\rho_{DCCA}$  e  $DMC_x^2$ . Nesta fase, busca-se consolidar o entendimento sobre os algoritmos e metodologias vigentes, bem como identificar lacunas e oportunidades de melhoria nos métodos existentes.

Na segunda etapa, realiza-se uma avaliação detalhada das possibilidades e limitações das ferramentas computacionais disponíveis para o cálculo dos coeficientes. Essa avaliação se dará através de uma pesquisa que aplica os métodos, utilizando as ferramentas de cálculo existentes. Este procedimento deve embasar o pesquisador para analisar, criticar e avaliar as possibilidades e limitações das ferramentas, das pesquisas e das questões que podem ser formuladas e respondidas através do ferramental vigente.

Com base nessa análise, é proposta a implementação de uma nova ferramenta computacional, com foco na otimização do algoritmo, na praticidade do uso e na eficiência do cálculo. Essa etapa envolve o desenvolvimento, teste e validação da ferramenta, garantindo que ela seja robusta e capaz de lidar com conjuntos de dados de diferentes tamanhos e complexidades.

Por fim, na terceira etapa, a ferramenta desenvolvida é aplicada a um conjunto de dados complexo, com o objetivo de testar os avanços proporcionados pela implementação. Essa aplicação prática permite avaliar a eficácia das melhorias realizadas, bem como explorar novas possibilidades de análise e interpretação de padrões em séries temporais múltiplas.

Essa abordagem metodológica experimental permite não apenas validar as hipóteses formuladas, mas também contribuir para o avanço do estado da arte na análise de correlações em séries temporais, com implicações práticas em diversas áreas do conhecimento.

## 1.7 Organização da Tese

Esta Tese está dividida em capítulos, sendo que um deles versa sobre as pesquisas e artigos produzidos durante o doutorado. Após uma breve introdução do objeto de estudo, realizada no presente capítulo, segue:

Uma revisão da literatura, no Capítulo 2, abrangendo nos métodos de correlação entre séries temporais basados no *DFA*, com um foco específico no estudo de métodos que propõe alguma forma de correlação entre as séries.

O Capítulo 3 apresenta a produção científica em fora de artigos. Uma sequência de pesquisas que compõem o cerne da Tese. São trabalhos independentes porém interconectados que são apresentados em seguir:

- A Sessão 3.1 apresenta um artigo publicado, apresentando uma aplicação do  $DMC_x^2$  na análise de múltiplas séries temporais. Uma contribuição para a análise de séries de ondas cerebrais, captadas por eletroencefalograma (EEG) também é apresentada.
- A Sessão 3.2 apresenta um artigo ainda não publicado que trata da implementação de uma ferramenta computacional que possibilita velocidade e *usabilidade* no cálculo dos algorítimos do *DFA*, *DCCA* e  $\rho_{DCCA}$ . Além da implementação da biblioteca, o artigo apresenta como contribuição, uma mudança na sequência de passos extremamente vantajosa para o cálculo do *DCCA* e  $\rho_{DCCA}$ , principalmente para cálculo com muitas séries paralelas, como no caso da atribuição de valores na matriz do  $\rho_{DCCA}$  para cálculo do  $DMC_x^2$ .
- A Sessão 3.3 apresenta um experimento utilizando a ferramenta computacional desenvolvida. Testa critérios de semelhança entre séries temporais e aponta a possibilidade de se usar esses critérios na seleção de atributos de algoritmos de aprendizado de máquina.

O Capítulo 4 por fim, apresenta a conclusão da Tese.

O Anexo A apresenta um artigo de co-autoria do discente desta Tese. Esse artigo é anterior ao artigo apresentado no Capítulo 3, Sessão 3.1. Apresenta análise de dados oriundos da mesma base utilizada no Capítulo 3, Sessão 3.1 utilizando o coeficiente  $\rho_{DCCA}$  na análise dos dados.

O Anexo B apresenta registros de *sotwares* e de documentação técnica oriundos do desenvolvimento desta Tese.

---

## Fundamentação Teórica e Metodologia

---

“Time is nature’s way of keeping everything from happening at once.”

(John Archibald Wheeler)

O presente capítulo apresenta uma revisão bibliográfica dos métodos utilizados, iniciando com a definições preliminares, seguindo pela explicação passo a passo do método para calcular o *DFA* e o *DCCA*, sem os quais o  $\rho_{DCCA}$  e o  $DMC_x^2$  não podem ser entendidos, seguindo pela explanação do cálculo do  $\rho_{DCCA}$  e do  $DMC_x^2$ , como se dá a leitura dos resultados e como são aplicados em análises.

Em seguida apresenta-se as relações dos métodos com a geometria fractal e multifractal, conceitos importantes nas análises das séries temporais propostas pelos métodos derivados do *DFA*. Apresentamos também metodologia e algorítimo para da determinação da multifractalidade de uma série temporal, segundo o *DFA*.

Na Sessão 2.4, apresenta-se uma seleção de outros algoritmos que, além do  $DMC_x^2$ , trabalham com correlações múltiplas baseadas no *DFA*. Algoritmos selecionados são apresentados.

Na Sessão 2.5, apresenta-se aplicações do que está família de metodologias já foi aplicada, ressaltando a gama de problemas e os mais relacionados com as ciências ambientais.

### 2.1 Definições preliminares

O conceito de correlação pode ser definido como uma medida de associação entre duas ou mais conjuntos de variáveis. Como exemplo podemos citar a correlação de Pearson, uma medida de correlação linear entre dois conjuntos de dados. Para entender o cálculo do coeficiente de Pearson é preciso entender importantes medidas de dispersão: o desvio padrão e a variância.

A média é uma importante medida de posição para os valores de uma determinada variável. Por medida de posição pode-se entender que a média, soma de todos os valores medidos para uma variável, marca um único valor para toda a série. Embora uma

medida importante, a média não apresenta como esses valores se distribuem ao longo das retas. As medidas que procuram capturar a oscilação dos valores são chamadas de medidas de dispersão. As mais importantes destas medidas são a variância ( $\sigma^2$ ) e o desvio padrão ( $\sigma$ ), calculadas de acordo com as Equações 2.1 e 2.2 para uma determinada população.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (2.1)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2.2)$$

As duas medidas partem do mesmo princípio: um somatório da diferença entre os valores e a média (desvio) dividido pelo número de observações. Como a somatória destes valores, considerando os positivos, (à direita do valor médio) e negativos (à esquerda da média), seria igual à zero, no cálculo da variância e do desvio padrão, esses valores são elevados ao quadrado. No caso da variância média do somatório dos desvios é calculada e no caso do desvio padrão temos a raiz quadrada da variância.

$$\text{cov}(X, Y) = \frac{1}{N} \sum_{i=1}^N (X - \mu_X)(Y - \mu_Y) \quad (2.3)$$

É possível estabelecer relações entre dois conjuntos de observações a partir da variância e do desvio padrão, mas a covariância consegue atribuir um valor para essa comparação. A covariância (Eq. 2.3) mede o grau de dependência numérica entre duas variáveis. Quando duas variáveis são independentes entre si o valor da covariância é zero.

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (2.4)$$

A correlação de Pearson é um coeficiente de correlação linear, que normaliza os valores da covariância entre  $-1$  (ante-correlação) e  $1$  (correlação). Seu cálculo é a razão da covariância entre duas variáveis e o produto dos desvios padrões das mesmas (Eq: 2.4).

## 2.2 Métodos de Análise de Séries Temporais: *DFA*, *DCCA*, $\rho_{DCCA}$ e $DMC_x^2$

Assim como o coeficiente de Pearson é baseado no desvio padrão e na covariância, o coeficiente  $\rho_{DCCA}$  (ZEBENDE, 2011) foi formulado tendo como bases o *Detrended Fluctuation Analysis* (*DFA*) (PENG et al., 1994b) e o *Detrended Cross-Correlation Analysis* (*DCCA*) (PODOBNIK; STANLEY, 2008). O *DFA* é um método de análise de uma série temporal que fornece um parâmetro de auto-afinidade. O termo *Detrended* refere-se a eliminação de uma tendência. O processo é executado em 6 passos, como mostrado no Algoritmo 1.<sup>1</sup>

---

### Algoritmo 1 *DFA*

---

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  $F_{DFA} \times n$
- 

O *DFA* caso tenha uma relação de potência (calda longa) entre  $fdfa$  e  $n$ , então haverá uma lei de potência do tipo  $fdfa(n) \propto n^\alpha$  (ZEBENDE; SILVA; FILHO, 2013). Diferentes valores de  $\alpha$  podem ser interpretados como:

- $\alpha < 1/2$ : Anti-correlação.
- $\alpha \simeq 1/2$ : Sem correlação (white noise).
- $\alpha > 1/2$ : Correlação.

---

<sup>1</sup>Aqui considera-se o cálculo com sobreposição de caixas.

- $\alpha \simeq 1$ : Ruido rosa.
- $\alpha > 1$ : Série não estacionária.
- $\alpha \simeq 3/2$ : Ruído browniano.

Para o ajuste do polinômio, definido como a tendência, costuma utilizar um polinômio de primeira ordem, mas existem estudos que comparam o efeito que a ordem do polinômio exerce sobre a análise e a capacidade do polinômio de maior grau capturar a tendência da série (HU et al., 2001).

O passo 4 do Algoritmo do *DFA* (Alg. 1) apresenta semelhanças com o cálculo da desvio padrão (Eq: eq:std). Enquanto o desvio padrão soma os quadrados dos valores subtraídos da média, nesta etapa do algoritmo do *DFA* calcula-se o somatório da diferença entre os valores da série integrada com o valor do polinômio de tendência. Enquanto a medida de dispersão aponta o quanto os valores desviam do valo médio, o passo 4 do *DFA* analisa quanto os valores flutuam em relação à tendência. Essa é a ideia por trás do termo função de flutuação. Como a tendência é calculada para um conjunto de  $n + 1$  pontos, o passo cinco aplica a raiz quadrada da média dos valores do passo 4 obtidos para cada cada cálculo de uma nova tendência em relação aos pontos utilizados para o ajuste da tendência.

O *DCCA* amplia o *DFA* para estabelecer a correlação entre duas séries temporais (PO-DOBNIK; STANLEY, 2008). O valor deste coeficiente tende a ser a média dos valores do *DFA* das duas séries e segue os 6 passos descritos no Algoritmo 2.

O *DCCA* atesta a existência de uma correlação entre duas séries mas não quantifica o nível desta correlação. Assim como a etapa 4 do Algoritmo do *DFA* apresenta uma relação com o desvio padrão, o *DCCA* aproxima-se da covariância.

O *Detrended cross-correlation coefficient* ou  $\rho_{DCCA}$  (equação 2.5) é um coeficiente que, variando entre  $-1$  e  $1$ , aponta ausência de correlação cruzada para valores próximos de zero, sendo maior a correlação quanto mais o valor se aproximar de  $1$  e maior a anti-correlação quanto mais o valor se aproximar de  $-1$  (ZEBENDE, 2011).

$$\rho_{DCCA}(n) = \frac{F_{DCCA(x^{j1}, x^{j2})}^2(n)}{F_{DFA(x^{j1})}(n) \times F_{DFA(x^{j2})}(n)} \quad (2.5)$$

Considerando a proximidade do *DFA* com o desvio padrão e do *DCCA* com a covariância (na etapa número 4), p  $\rho_{DCCA}$  apresenta grande semelhança com o coeficiente de Pearson, substituindo a covariância pelo *DCCA* e os desvios padrões pelos respectivos valores do *DFA*.

---

**Algoritmo 2**  $DCCA$ 

---

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:

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

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

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

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  $F_{DCCA} \times n$
-

O método foi estatisticamente validado (PODOBNIK et al., 2011), testado (VASSOLER; ZEBENDE, 2012; GUEDES et al., 2017; FERREIRA et al., 2018), e critérios para avaliação de relevância estatísticas do resultados foram desenvolvidos (GUEDES et al., 2018a; GUEDES et al., 2018b).

O  $\rho_{DCCA}$  foi estendido para calcular a correlação cruzada de múltiplas séries temporais. Denominado *Detrended Multiple Cross-Correlation Coefficient* ( $DMC_x^2$ ), representa a generalização do  $\rho_{DCCA}$  para múltiplas variáveis (ZEBENDE; SILVA, 2018). Também foi implementado com abordagem de janelas móveis (GUEDES; da Silva Filho; ZEBENDE, 2021) e foi desenvolvido um teste estatístico para o coeficiente múltiplo (da Silva Filho et al., 2021)

A generalização utiliza a equação 2.6. Sendo  $y$  uma série temporal definida como variável dependente,  $x^j$  um conjunto de séries temporais tomadas como variáveis independentes, com  $j$  variando de 1 a  $m$ . Sendo  $\rho_{y,x^j}(n)$  o vetor coluna contendo dos valores dos coeficientes  $\rho_{DCCA}$  entre a série  $y$  e cada uma das séries  $x^j$  (Eq: 2.7).

$$DMC_x^2 \equiv \rho_{Y,X^j}(n)^T \times \rho^{-1}(n) \times \rho_{Y,X^j}(n) \quad (2.6)$$

$$\rho_{Y,X^j}(n)^T = [\rho_{Y,X^1}(n), \rho_{Y,X^2}(n), \dots, \rho_{Y,X^m}(n)] \quad (2.7)$$

A matriz  $\rho^{-1}(n)$ , inversa da matriz apresentada na equação 2.8. A matriz é montada através dos valores dos coeficientes  $\rho_{DCCA}$  para cada par de variáveis dependentes em  $x_w$ . Pela propriedade comutativa presente na equação do cálculo do  $\rho_{DCCA}$ , pode-se afirmar que a matriz é sempre simétrica. O valor do  $\rho_{DCCA}$  de uma série com ela mesma é sempre o de correlação total (1), portanto a diagonal principal apresenta o valor 1 em todas as posições.

$$\rho(n) = \begin{pmatrix} 1 & \rho_{X^1,X^2}(n) & \rho_{X^1,X^3}(n) & \dots & \rho_{X^1,X^m}(n) \\ \rho_{X^2,X^1}(n) & 1 & \rho_{X^2,X^3}(n) & \dots & \rho_{X_2,X^m}(n) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{X^m,X^1}(n) & \rho_{X^m,X^2}(n) & \rho_{X^m,X^3}(n) & \dots & 1 \end{pmatrix} \quad (2.8)$$

## 2.3 *Fractais e os métodos baseados no DFA*

Existe uma relação entre os métodos baseados no *DFA* e a geometria fractal, termo cunhado pelo matemático Benoit Mandelbrot ([MANDELBROT, 1977](#)). A palavra fractal procede do latim *fractus* cujo significado é ”quebrado” ou ”fragmentado”.

A pesquisa que originou a geometria fractal baseou-se em pesquisas anteriores do matemático Gaston Julia, e aplicou computação gráfica para estudar o comportamento de métodos iterativos na resolução de certas equações complexas. O conjunto de images geradas pelo experimento apresentou características e propriedades geométricas particulares, dentre elas a persistência do valor obtido pelo cálculo da dimensão de Hausdorff (que, a partir dos trabalhos de Mandelbrot, passa a ser conhecida também como dimensão fractal), ainda que calculado utilizando diferentes escalas de discretização do espaço de resultados ([MANDELBROT, 1982](#)).

A técnica de contagem de caixas é uma ferramenta utilizada para estimar a dimensão fractal de uma forma. Técnica foi introduzida por Mandelbrot e tem sido amplamente utilizada em estudos de fractais.

A ideia básica da contagem de caixas é dividir o conjunto em caixas de tamanho fixo e contar o número de caixas que o conjunto ocupa. Em seguida, o tamanho da caixa é reduzido e o processo é repetido. A relação entre o número de caixas e o tamanho da caixa pode ser usada para estimar a dimensão fractal do conjunto.

A fórmula para a dimensão fractal utilizando a técnica de contagem de caixas é dada por:

$$D = \lim_{\epsilon \rightarrow 0} \frac{\log(N(\epsilon))}{\log(1/\epsilon)} \quad (2.9)$$

Sendo  $N(\epsilon)$  é o número de caixas de tamanho  $\epsilon$  que preenchem o conjunto.

A ideia de se dividir uma figura 2D em caixas apresenta semelhança com a divisão de uma série temporal (1D) por segmentos de reta de um polinômio de primeiro grau. Mais do que uma semelhança aleatória, em escala log-log, o *DFA*, quando desenha um segmento de reta ligando uma sequência de escalas temporais, representa uma característica de auto-similaridade da série, como mostrado na Figura 2.1 ([PENG et al., 1994a](#)).

Uma generalização multifractal do *DFA* foi proposta por [Kantelhardt et al. \(2002\)](#). Apresentado na Equação 2.10, onde o valor de  $q$  representa a ordem da função *DFA* calculada. Quando  $q = 2$  a equação assume o valor do *DFA*. No caso de  $q = 0$  a média exponencial

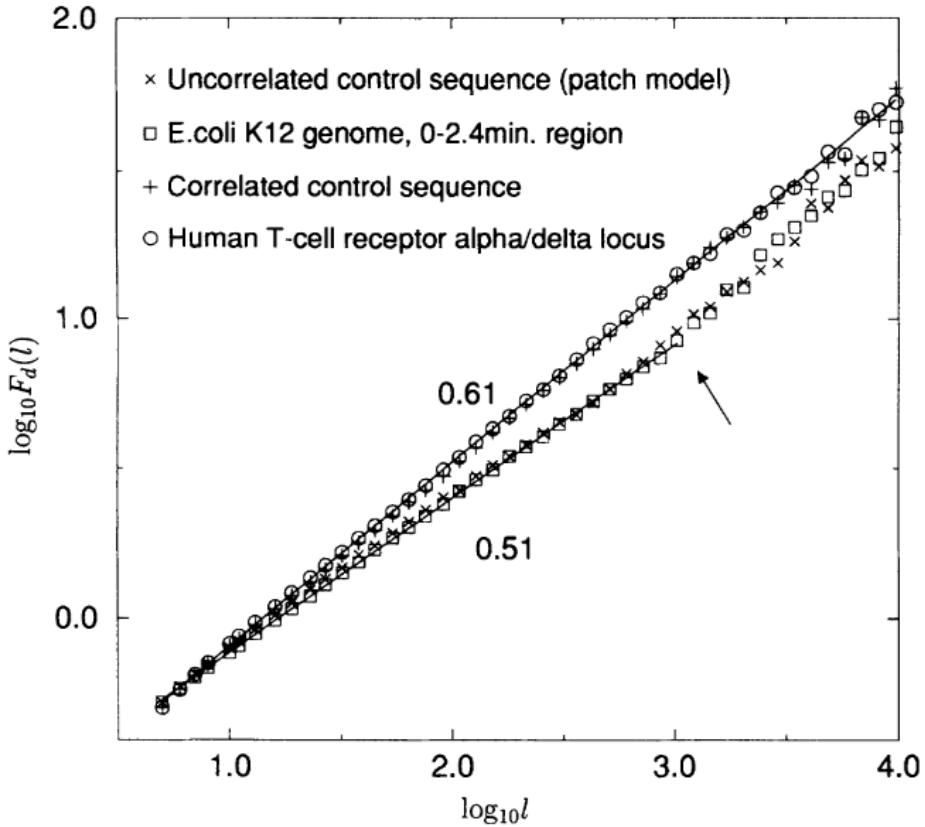


Figura 2.1: Análise *DFA* de sequências de DNA e sequências de controle, resultados apresentando correlação de longo alcance (+ e o) e não apresentando correlação. (x e □)

Fonte: ([PENG et al., 1994a](#))

é utilizada para a medição. Esses resultados também foram expandidos, baseando-se no *DCCA* para a multi-correlação multifractal entre séries temporais ([ZHOU, 2008](#)).

$$q \neq 0,$$

$$F_q(s) \equiv \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} [F^2(v, s)]^{q/2} \right\}^{1/q}, \quad (2.10)$$

$$q = 0,$$

$$F_0(s) \equiv \exp \left\{ \frac{1}{4N_s} \sum_{v=1}^{2N_s} \ln [F^2(v, s)] \right\} \sim s^{h(0)}$$

## 2.4 Variantes de correlação cruzada baseados no *DCCA*

Duas variantes do *DCCA* que lidam com correlação cruzada já foram apresentados nesta pesquisa: O  $\rho_{DCCA}$  e o  $DMC_x^2$ . Além desses métodos, este trabalho destaca mais alguns trabalhos: *Multivariate Detrended Fluctuation Analysis (MVDFA)* ([XIONG; SHANG, 2017](#)), *Generic Multichannel DFA(GMDFA)* ([NAVEED; REHMAN, 2025](#)) e o *Detrended Partial Cross-correlation Coefficient (DPCCA)* ([YUAN et al., 2015](#)).

Diferente do  $DMC_x^2$ , que estabelece um valor de função, onde a correlação é medida entre um conjunto de variáveis independentes e uma variável dependente, o *MVDFA* calcula um valor único para qualquer combinação de séries temporais, representando uma generalização do *DFA* para múltiplas variáveis. Neste método, divide-se a série temporal em um conjunto de caixas não sobreposto, eventualmente perdendo alguns valores no final da série, mas repetindo o processo do fim para o começo da reta para compensar. Após a interpolação dos polinômios, a função local, em cada caixa e dada pela Equação 2.11, one a norma euclidiana (Equação 2.12) é utilizada para o cálculo.

$$F^2(v, s) = \frac{1}{s} \sum_{t=1}^s \left\| (Y_{l_v+t,1}, Y_{l_v+t,2}, \dots, Y_{l_v+t,p}) - (\tilde{Y}_{\cdot,1}, \tilde{Y}_{\cdot,2}, \dots, \tilde{Y}_{\cdot,p}) \right\|^2 \quad (2.11)$$

$$\|X - Y\| = \|(x_1, x_2, \dots, x_n) - (y_1, y_2, \dots, y_n)\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.12)$$

Terminando o cálculo de forma semelhante ao apresentado no algoritmo do *DFA*, como mostrado na Equação 2.13. O *MGDF*A substitui a norma euclideana pela norma de Mahalanobis. Estes métodos não possuem grandes relações com o *DCCA* e, por extensão, com o  $\rho_{DCCA}$  e  $DMC_x^2$ .

$$F_{MVDFA}(s) = \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} [F^2(v, s)] \right\}^{1/n} \quad (2.13)$$

Já o *DPDCCA* aplica a mesma matriz utilizada no cálculo do  $DMC_x^2$ , apresentado na Equação 2.6, cuja forma antes da inversão é apresentada na Equação 2.8. A matriz é denominada de  $C(n)$  (Equação 2.14).

$$C(n) = \boldsymbol{\rho}^{-1}(n) = \begin{pmatrix} C_{1,1}(n) & C_{1,2}(n) & \dots & C_{1,m}(n) \\ C_{2,1}(n) & C_{2,2}(n) & \dots & C_{2,m}(n) \\ \vdots & \vdots & & \vdots \\ C_{m,1}(n) & C_{m,2}(n) & \dots & C_{m,m}(n) \end{pmatrix} \quad (2.14)$$

O cálculo final do *DPDCCA* é realizado pela Equação 2.15, a partir dos valores obtidos pela inversão da matriz do  $\boldsymbol{\rho}_{DCCA}$ .

$$\rho_{DPCCA}(j_1, j_2; n) = \frac{-C_{j_1, j_2}(n)}{\sqrt{C_{j_1, j_1}(n) \cdot C_{j_2, j_2}(n)}} \quad (2.15)$$

O resultado deve ser analisado como a correlação entre as duas variáveis, desconsiderando a atuação das outras variáveis cujos valores de  $\boldsymbol{\rho}_{DCCA}$  foram utilizados na montagem da matriz na Equação 2.8.

## 2.5 Aplicações das funções e coeficientes

Muitas aplicações foram testadas para essas funções e coeficientes. Muitas aplicações na área de medicina e saúde (CHEN et al., 2018; FILHO et al., 2023a; GE; LIN, 2023; GHOSH et al., 2018). Questões de economia e mercado financeiro na China (LI; TIAN, 2024), América Latina (Contreras-Reyes; Jeldes-Delgado; CARRASCO, 2024), Estados Unidos e China (CHEN et al., 2024) e no mercado de energia (ASLAM et al., 2024), detre outros, provando que os estudos de correlação entre mercados costuma tratar de aspectos globais e locais, apresentando forte relação com as ciências ambientais.

Muitos estudos também utilizaram esses métodos em outros aspectos das ciências ambientais: estudos sobre dados sísmicos (SHADKHOO; JAFARI, 2009) e geológicos (BIANCHI; LONGO; PLASTINO, 2018; QIN et al., 2024); transportes (SHI et al., 2018; ZEBENDE; da Silva; FILHO, 2011; ZEBENDE; FILHO, 2009); estudos técnicos que podem trazer melhorias para questões de infra-estrutura (SANTOS et al., 2018); pesquisas sobre violência urbana (FILHO; da Silva; ZEBENDE, 2014); e muitos estudos sobre o clima.

Dentre os estudos sobre o clima, destacamos: correlações entre temperatura (YU; LI; ZHOU, 2019); chuvas (ADARSH et al., 2020), (LI et al., 2024); entre temperatura e umidade (ZEBENDE et al., 2018); entre dados meteorológicos, e poluentes (SHI, 2014), (HE, 2017), (ZHANG; NI; NI, 2015), (SHEN; LI; SI, 2015); Chuva e doenças tropicais (OLIVEIRA et al., 2023); aquecimento global (CHATTERJEE; GHOSH, 2021), (YUAN;

FU, 2014); meteorologia e previsões (AFUECHETA; OMAR, 2021; SHIN et al., 2019); temperatura do solo e composição (KAR; CHATTERJEE; GHOSH, 2019).

Com base nesta pequena amostra dos inúmeros trabalhos que abordam esse conjunto de funções e coeficientes, é possível afirmar que, do ponto de vista das ciências ambientais, o estudos destes algorítimos é uma significativa contribuição.

## Produção Científica

---

Neste Capítulo apresentamos produções científicas em formato de artigos.

Na Sessão 3.1, um artigo publicado que apresenta uma aplicação do  $DMC_x^2$  na análise de dados registrados por um Eletroencefalograma no decorrer de um experimento.

Na Sessão 3.2 temos o artigo, em fase de revisão pelos autores, que deve ser submetido em seguida. O artigo trata da implementação de uma ferramenta computacional para facilitar o uso das funções e coeficientes abordados por pesquisadores.

A Sessão 3.3 apresenta a ferramenta computacional e explora suas potencialidades.

Ressaltamo que o artigo apresentado no Anexo A, foi o primeiro que participei como coautor neste grupo de pesquisa. Ele embasa os trabalhos do artigo apresentado nas Sessões 3.1 e 3.3, e faz parte da trajetória de investigação sobre os algoritmos que levaram ao artigo apresentado na Sessão 3.2.

### **3.1 Artigo 01: DCCA multi cross-correlation analysis applied on EEG signals to study motor activity (Real/Imaginary)**

“That brain of mine is something more than merely mortal, as time will show.”

(Ada Lovelace)

O primeiro artigo apresentado, *DCCA multi cross-correlation analysis applied on EEG signals to study motor activity (Real/Imaginary)* (RIBEIRO et al., 2025), narra uma pesquisa utilizando o  $DMC_x^2$  na busca de padrões cerebrais, através do estudo das séries temporais oriundas de um experimento (de movimentos reais e imaginários dos membros do corpo) monitorado pelas gravações das ondas do cérebro por um aparelho de eletroencefalograma. O artigo trabalha com o  $\rho_{DCCA}$  com a mesma fonte de dados que o artigo intitulado *Statistical study of the EEG in motor tasks (real and imaginary)* (FILHO et al., 2023b), utilizando o  $\rho_{DCCA}$  apresentado no Anexo A, e representa um avanço técnico em relação ao anterior.

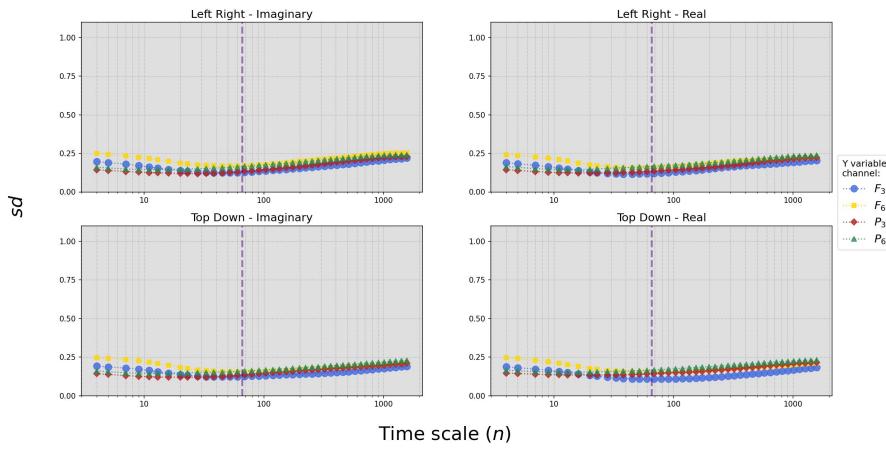


Figura 3.1: Desvio padrão dos valores do  $DMC_x^2$  para cada uma das tarefas.  
 Fonte: Elaborada pelos autores.

No Artigo de 2023, apenas um subconjunto de 11 indivíduos de universo de 109. Já no artigo de 2025, 108 dos 109 indivíduos foram analisados. Encarando o dilema de, trabalhar com um ambiente de mais alto nível, capaz de manipular o conjunto de dados e visualizar os resultados de forma mais prática e automatizada, com sacrifício do desempenho, ou trabalhar em baixo nível, como maior velocidade de cálculos e menor praticidade para trabalhar, uma solução intermediária foi adotada: os dados foram baixados, carregados, editados e preparados utilizando um ambiente Python, os cálculos foram feitos através de um programa escrito em C e a visualização dos resultados novamente em Python. As chamadas do programa em C foram feitas dentro do código Python, utilizando a biblioteca *subprocess*.

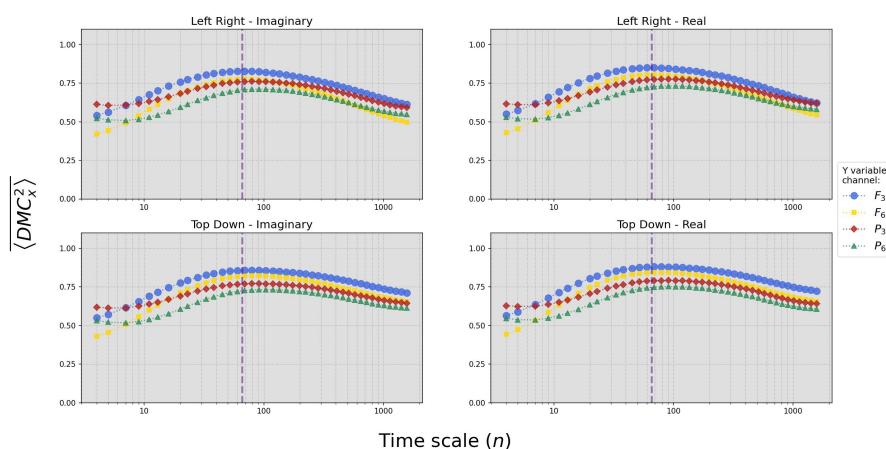


Figura 3.2: Médias dos valores do  $DMC_x^2$  para cada uma das tarefas.  
 Fonte: Elaborada pelos autores

A implementação em C executa a inversão da matriz  $\rho^{-1}(n)$ , apresentada na Equação 2.8 em codificação direta, como mostrado na Equação 3.1. Significando que, para cada quantidade de variáveis independentes que se queira estudar, um novo programa deve ser

escrito.

$$\begin{aligned}
 DMC_x^2 = & \left( \rho_{X_2, X_3}^2 \times \rho_{Y, X_1}^2 - \rho_{Y, X_1}^2 + \rho_{X_1, X_3}^2 \times \rho_{Y, X_2}^2 - \rho_{Y, X_2}^2 \right. \\
 & + 2 \times \rho_{X_1, X_2} \times \rho_{Y, X_1} \times \rho_{Y, X_2} - 2 \times \rho_{X_1, X_3} \times \rho_{X_2, X_3} \times \rho_{Y, X_1} \\
 & + \rho_{X_1, X_2}^2 \times \rho_{Y, X_3}^2 - \rho_{Y, X_3}^2 + 2 \times \rho_{X_1, X_3} \times \rho_{Y, X_1} \times \rho_{Y, X_3} \\
 & - 2 \times \rho_{X_1, X_2} \times \rho_{X_2, X_3} \times \rho_{Y, X_1} \times \rho_{Y, X_3} \\
 & - 2 \times \rho_{X_1, X_2} \times \rho_{X_1, X_3} \times \rho_{Y, X_2} \times \rho_{Y, X_3} \\
 & \left. + 2 \times \rho_{X_2, X_3} \times \rho_{Y, X_2} \times \rho_{Y, X_3} \right) / \\
 & \left( \rho_{X_1, X_2}^2 + \rho_{X_1, X_3}^2 + \rho_{X_2, X_3}^2 - 2 \times \rho_{X_1, X_2} \times \rho_{X_1, X_3} \times \rho_{X_2, X_3}^{-1} \right)
 \end{aligned} \quad (3.1)$$

A pesquisa utilizou-se exaustivamente da análise de gráficos para chegar nas conclusões revisadas pelos pares. A quantidade de figuras excederia o tamanho do artigo, forçando a criação de um repositório *online* no endereço abaixo:

[\(https://255ribeiro.github.io/Multi\\_Cross-correlation\\_EEG/\)](https://255ribeiro.github.io/Multi_Cross-correlation_EEG/)

Os códigos utilizados para o *download* dos dados, tratamento, cálculos e geração das figuras estão disponíveis em [https://github.com/255ribeiro/Series\\_EGG\\_3\\_14](https://github.com/255ribeiro/Series_EGG_3_14).

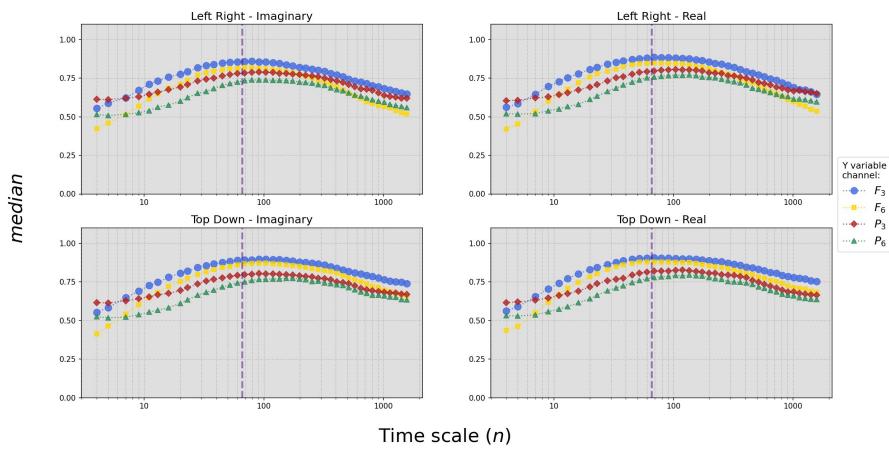


Figura 3.3: Mediana dos valores do  $DMC_x^2$  para cada uma das tarefas.  
 Fonte: Elaborada pelos autores

As Figuras 3.1, 3.2 e 3.3 mostram grande consistência entre a população, no entanto é possível observar características nos gráficos de cada indivíduo que apontam para uma assinatura individual, como uma impressão digital, como explicado no artigo. Para facilitar

o entendimento das análises, alguns vídeos que apresentam sequencialmente os resultados das tarefas de cada sujeito foram criados e colocados no repositório *online*.

Nos termos da quantidade de dados analisados, a implementação carece de melhorias e otimizações. Passar os dados para C através da biblioteca *subprocess* é um processo trabalhoso, exigindo a gravação dos arquivos tratados em arquivos formatados conforme as regras estabelecidas no código em C. Generalizações do algorítimo para um número qualquer de séries temporais e do cálculo da inversão da matriz também devem ser implementados para expandir as possibilidades em trabalhos futuros.

# DCCA multi cross-correlation analysis applied on EEG signals to study motor activity (Real/Imaginary)

Fernando Ferraz Ribeiro<sup>a,b,c</sup>, Andréa de Almeida Brito<sup>d</sup>, Florêncio Mendes Oliveira Filho<sup>a,c</sup>, Juan Alberto Leyva Cruz<sup>e</sup>, Gilney Figueira Zebende<sup>e</sup>

<sup>a</sup>*Earth Sciences and Environment Modeling Program, State University of Feira de Santana, Bahia, Brazil*

<sup>b</sup>*Federal University of Bahia, Salvador, Brazil*

<sup>c</sup>*SENAI CIMATEC University Center, Salvador, Bahia, Brazil*

<sup>d</sup>*Federal Institute of Education, Science and Technology, Salvador, Bahia, Brazil*

<sup>e</sup>*State University of Feira de Santana, Bahia, Brazil*

---

## Abstract

We applied the DCCA multiple cross-correlation coefficient to analyze time series of EEG experiment, where 109 subjects performed four tasks involving real and imaginary motor activities. In this case, four specific channels were selected on the scalp: two on the frontal and two on parietal region. As a result, the DCCA multiple cross-correlation coefficient identified that there is a single signature for each subject. Globally, there is no significant difference between the real and the imaginary task. The frontal channels had greater multiple cross-correlation values than the parietal ones, especially for time scales around 0.42s, and with smaller standard deviations. According to our results, the proposal to study multiple time series at the same time by using the DCCA multiple cross-correlation coefficient is feasible and robust for EEG analysis. Finally, the coefficient was applied in a substantial number of subjects, tasks, and experiments producing high-quality results (figures, movies, and tables), and this study will probably kick off of a new approach to analyzing multiple cross-correlations in EEG signals.

*Keywords:* DFA, DCCA, DCCA Multiple Cross-Correlation Coefficient, EEG signals, Brain

---

## 1. Introduction

An electroencephalogram, or EEG, is a test that evaluates the electrical activity of the brain. It is normally a non-invasive method, where tiny electrodes are placed on the scalp to track brain wave patterns and send them to a computer. The EEG equipment measures the electric potential difference (usually in  $\mu\text{V}$ ) between each of the 64 electrodes and a reference electrode, usually placed in the ear lobe. These impulses are amplified and recorded over time, generating time series for each sensor (with high temporal resolution and low spatial resolution). The international 10 – 10 system was used to map the positions, where 64 electrodes were fixed in the experiments, that collected the data used in this analysis. The nomenclature of the channels in this system is given according to the region in which they are located, namely  $F$  (frontal),  $T$  (temporal),  $C$  (central),  $P$  (parietal), and  $O$  (occipital). Channels located on the midline are indexed by the letter  $z$  (zero), channels located on the left side are represented by odd indexes, and those on the right side are represented by even indexes (see Fig. 1). EEG

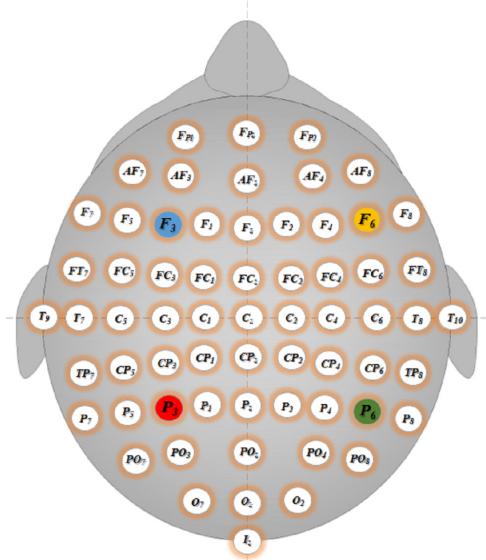


Figure 1: Electrode positions based on international 10 – 10 system for 64 channels. The circles  $F_3$  (blue),  $F_6$  (yellow),  $P_3$  (red), and  $P_6$  (green) identify the channels used for multiple cross-correlation analysis.

analysis is useful for diagnosing problems such as epilepsy, dementia, sleep

disorders, and other health problems. The diagnosis normally focuses on the spectral content of the EEG and on the type of neural oscillations (or brain waves) that can be observed in these signals. Most observed signals are between 1 and 20 Hz. Although EEG is almost a centenary technique, in recent decades, it has addressed new problems, such as brain-triggered neuro-rehabilitation treatments, experimental psychology, or even computational neuroscience, due to its versatility and accessibility, alongside the advances in signal processing [1, 2] that enhance the analysis possibilities in the field. Some new techniques for analyzing EEG signals that we wish to highlight are the root mean square fluctuation function,  $F_{DFA}$ , studying brain activity in the reading task [3, 4], the quantification of long-range correlation of EEG signals [5], and recently in the statistical study of the EEG cross-correlation of two signals using the detrended cross-correlation [6].

In this paper, we will analyze multiple time series of EEG signals produced in a motor activity (Real/Imaginary) by 109 subjects via the Detrended Multiple Cross-Correlation Coefficient [7]. Based on symmetry criteria, we chose the channels  $F_3$ ,  $F_6$ ,  $P_3$ , and  $P_6$  to apply this new statistical tool. In the sections below we present the Data and Method with dataset, methodology, and calculations used to analyze the data, including pre-processing strategies. Afterward, we present the Results and Discussion with its statistics and a discussion and, finally we make a Conclusion.

## 2. Data and Method

### *Data*

The EEG files were downloaded from the *Physionet* open-access website, available at this link:

<https://physionet.org/content/eegmmidb/1.0.0/>

*Physionet* operates under the Open Data Commons Attribution License (ODC-By) v1.0. This website presents a large quantity/quality of EEG experiments, in EDF (European Data Format) format, computed by a Brain-Computer Interface Technology, BCI-2000 [8], on the premise of the international 10 – 10 system. The *Physionet* website was accessed on January 11, 2022 for this study, taking into account that we did not have any information that could identify individual participants. These EEG signals were

performed in 14 experiments on a population of 109 subjects with the objective of record brain signals during various motor stimuli, or tasks, briefly described below.

The first two activities were baseline references: in the first, the subjects were resting with eyes opened, and in the second, with eyes closed (one minute for each). The other four activities (task) were a combination of two categories with two possible options each. In general, the experiments consisted of making the subjects react to a visual stimuli, in other words, a target that appears on a screen. One category is about the target position, i.e., one option was a target appearing on the **Left/Right** of the screen, and the other was a target appearing on the **Top/Down**. The second category determined if the subject actually executed a body movement related to the target position (**Real**) or if the corresponding action was just imagined (**Imaginary**). All experiments for these 109 subjects were downloaded, but we have decided to delete the EEG relating to S106, because the experiment 5 (Top/Down Real - task 3) for this subject have only  $N = 5808$  values.

These tasks, with approximately two minutes of duration, are better summarized in the Table 1, were each task was performed three times (experiments). After introducing our database, we will describe the methodology

Table 1: The activity executed and the experiment number for two one-minute baseline runs (one with eyes open, one with eyes closed) and three two-minute runs of a particular task.

Activity\Experiment	1	2	<b>3</b>	4	5	6	7	8	9	10	11	12	13	14
Baseline 1 (eyes open)	0													
Baseline 2 (eyes closed)	0													
task 1: Real (L/R)			X				X				X			
task 2: Imag (L/R)				X				X				X		
task 3: Real (T/D)					X				X				X	
task 4: Imag (T/D)						X				X				X

applied for multiple cross-correlation analysis.

#### *DCCA multiple cross-correlation methodology*

The coefficient  $DMC_x^2$  is a new statistical tool to analyze non-stationary time series in multiple applications [7]. It starts with the DFA method [9], which was proposed to identify self-affinity in a single time series, and its generalization, the DCCA method [10], for study cross-correlation in two time

series. Some applications in EEG signals, of DFA and DCCA method, can be seen to analyses brain disorder, as in different physiological and pathological states of epilepsy EEG signals [11], Alzheimer disease patients [12], among other [5, 4, 6]. Specifically,  $DMC_x^2$  is a generalization of the detrended cross-correlation coefficient,  $\rho_{X_a, X_b}$  [13] (widely known as a robust statistical tool [14]).  $DMC_x^2$  calculates the multiple cross-correlation of one time series  $\{Y\}$  (dependent variable) in relation to a number  $k$  of others time series  $\{X_1\}, \{X_2\}, \{X_3\}, \dots, \{X_k\}$  (independent variable). DCCA multiple cross-correlation is defined as:

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

This coefficient,  $0 \leq DMC_x^2(n) \leq 1$ , where 0 (1), the weaker (stronger) is the relationship between the dependent and independent variables. The term  $\rho^{-1}(n)$ , in the Eq. 1, represent the inverse matrix of all possible combinations of  $\rho_{X_a, X_b}$  (between the independent variables); in other words:

$$\rho^{-1}(n) = \begin{pmatrix} 1 & \rho_{X_1, X_2}(n) & \rho_{X_1, X_3}(n) & \dots & \rho_{X_1, X_k}(n) \\ \rho_{X_2, X_1}(n) & 1 & \rho_{X_2, X_3}(n) & \dots & \rho_{X_2, X_k}(n) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{X_k, X_1}(n) & \rho_{X_k, X_2}(n) & \rho_{X_k, X_3}(n) & \dots & 1 \end{pmatrix}^{-1} \quad (2)$$

In this equation,  $\rho_{Y, X_i}(n)$  represent the vector of the detrended cross-correlation between the predictor variables (independent) and the target variable (dependent), and  $\rho_{Y, X_i}(n)^T$  it is the transposed.

$DMC_x^2$  has been applied in many problems, specifically for three time series analysis, i.e., one  $\{Y\}$  (dependent) and two  $\{X_1\}, \{X_2\}$  (independent). Examples of these applications can be found applied to study meteorological variables [15], in statistical test [16], with the implementation of sliding windows [17], to measure the contagion effect on stock market indexes [18], to take the statistical analysis between stock market indexes [19], among other applications.

However, for the particular case of four time series, i.e., one  $\{Y\}$  (dependent variable) and three  $\{X_1\}, \{X_2\}, \{X_3\}$  (independent variables), there is still no cases. In this sense, from Eq. 1, the calculus of the  $DMC_x^2$  with three

independent variables can be represented by:

$$\begin{aligned}
DMC_x^2 = & \left( \rho_{X_2, X_3}^2 \times \rho_{Y, X_1}^2 - \rho_{Y, X_1}^2 + \rho_{X_1, X_3}^2 \times \rho_{Y, X_2}^2 - \rho_{Y, X_2}^2 \right. \\
& + 2 \times \rho_{X_1, X_2} \times \rho_{Y, X_1} \times \rho_{Y, X_2} - 2 \times \rho_{X_1, X_3} \times \rho_{X_2, X_3} \times \rho_{Y, X_1} \\
& + \rho_{X_1, X_2}^2 \times \rho_{Y, X_3}^2 - \rho_{Y, X_3}^2 + 2 \times \rho_{X_1, X_3} \times \rho_{Y, X_1} \times \rho_{Y, X_3} \\
& - 2 \times \rho_{X_1, X_2} \times \rho_{X_2, X_3} \times \rho_{Y, X_1} \times \rho_{Y, X_3} \\
& - 2 \times \rho_{X_1, X_2} \times \rho_{X_1, X_3} \times \rho_{Y, X_2} \times \rho_{Y, X_3} \\
& \left. + 2 \times \rho_{X_2, X_3} \times \rho_{Y, X_2} \times \rho_{Y, X_3} \right) / \\
& \left( \rho_{X_1, X_2}^2 + \rho_{X_1, X_3}^2 + \rho_{X_2, X_3}^2 - 2 \times \rho_{X_1, X_2} \times \rho_{X_1, X_3} \times \rho_{X_2, X_3}^{-1} \right)
\end{aligned} \tag{3}$$

The following section presents the data mining and its calculations for the aforementioned case.

#### *Data Mining and Calculations*

The data mining followed the steps presented below. In EEG experiments, usually, the end of each recording is filled with a sequence of zeros, corresponding to the time gap between the EEG machine and the recording system shutting down. In the pre-processing stage, these sequences of zeros were cut. To properly apply the  $DMC_x^2$  and make valid comparisons between subjects and experiments, the time series must have approximately the same length  $N$ . This data mining results in 108 subjects, each performing four tasks, in three consecutive experiments (see Table 1). After, the  $DMC_x^2$  calculation was performed, with all possible combinations between the four time series (our case of study). Due to the number of subjects, tasks, and experiments, we have a great number of results, which can be seen initially by accessing our public repository, through the link:

[https://255ribeiro.github.io/Multi\\_Cross-correlation\\_EEG/](https://255ribeiro.github.io/Multi_Cross-correlation_EEG/)

Basically, we observe the value of  $DMC_x^2$  as a function of  $n$  for each subject, in a given task, by experiments.

For a better understanding of these results, we define a color code to denote the dependent variable used in the applied channels:  $F_3$  (blue),  $F_6$  (yellow),  $P_3$  (red), and  $P_6$  (green) (see Fig. 1). Table 2 summarize this color code, relating to the channel as the dependent variable ( $Y$ ) and the ones

Table 2: The  $Y[\text{Channel}]$  (dependent variable) is represented by a specific color that will be implemented in all figures with the DCCA multiple cross-correlation coefficient.

Color	$Y[\text{Channel}]$	$X[\text{Channel1}, \text{Channel2}, \text{Channel3}]$
Blue	$\mathbf{Y}[F_3]$	$X[F_6, P_3, P_6]$
Yellow	$\mathbf{Y}[F_6]$	$X[F_3, P_3, P_6]$
Red	$\mathbf{Y}[P_3]$	$X[F_3, F_6, P_6]$
Green	$\mathbf{Y}[P_6]$	$X[F_3, F_6, P_3]$

used as the independent variables ( $X$ ). As an example of this application, we randomly selected the S014 subject, to present the first results. Fig. 2 shown the value of  $DMC_x^2 \times n$  for S014 subject carrying out the task 2, that is, Imaginary (Left/Righ) motor activity with: 04, 08, 12 experiments, and the mean value. In this case, the  $F_3$  channel as a  $Y$  variable, had the highest DCCA multiple cross-correlation value, corroborating with [3] for auto-correlation analysis. The complete analysis of the results and their

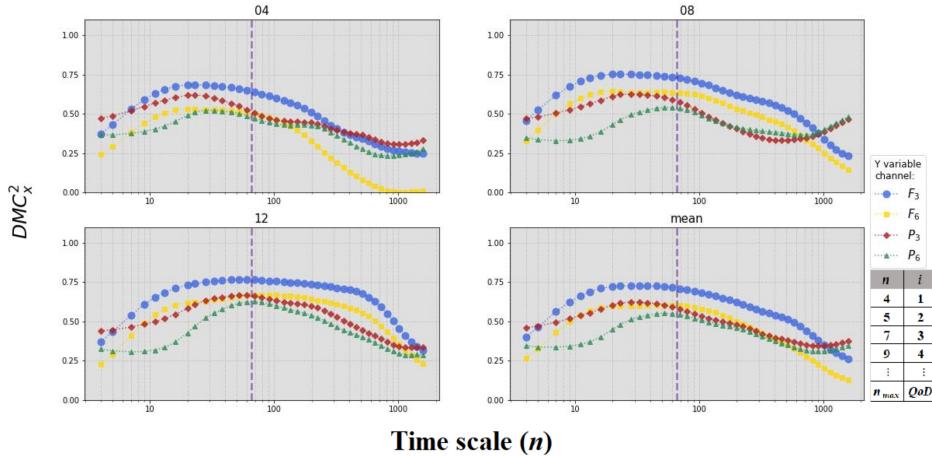


Figure 2:  $DMC_x^2$  as a function of time scale  $n$ . These are the results for S014 subject recordings for task 2, presenting experiments 04, 08, and 12 (Table 1) and the mean values for these experiments. The vertical line represents  $n = 67$ , and  $QoD$  is the total amount of time scales involved in  $DMC_x^2$  calculations.

discussion will be presented below, in the Section 3.

### 3. Results and Discussion

As a better way to compare the stimuli in response to the motor activities, we choose a sample of five randomly subjects: S014, S036, S039, S078, and S099. Therefore, with the mean value of  $DMC_x^2$  as a function of  $n$ , the initial results are presented as listed below:

- S014 in Fig. 3;
- S036 in Fig. 4;
- S039 in Fig. 5;
- S078 in Fig. 6;
- S099 in Fig. 7.

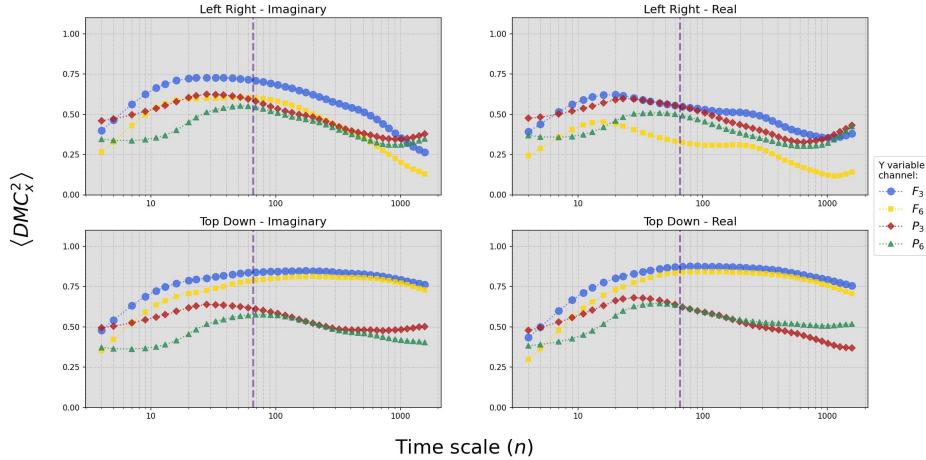


Figure 3: Mean values of  $DMC_x^2 \times n$  for all tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for S014 subject.

The  $F_3$  channel as a  $Y$  variable, has in general the greatest values for  $\langle DMC_x^2 \rangle$ , and the task Top/Down Real and Imaginary are very similar. Also, the Fig. 4 shows that the channel  $F_3$  with the greatest DCCA multiple cross-correlation values, if compared with the channels  $F_6$ ,  $P_3$ , and  $P_6$ . In this case, S036 subject present a distinct behavior in relation to the S014 subject.

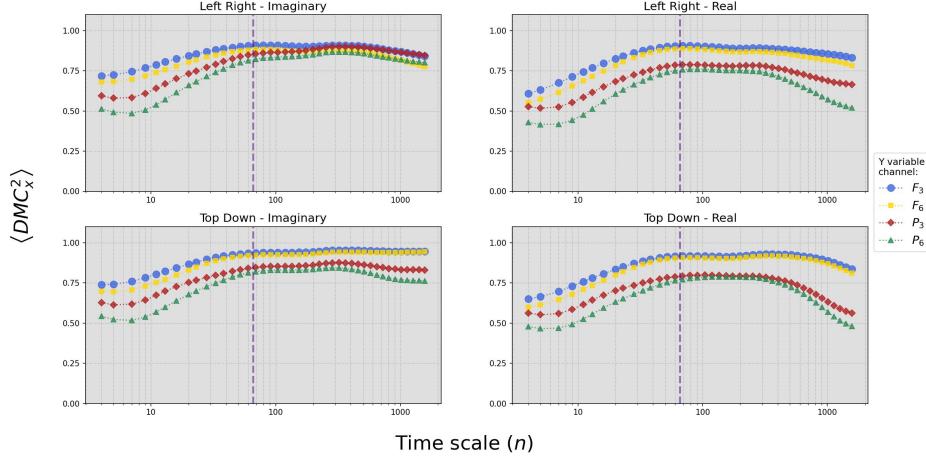


Figure 4: Mean values of  $DMC_x^2 \times n$  for all tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for S036 subject.

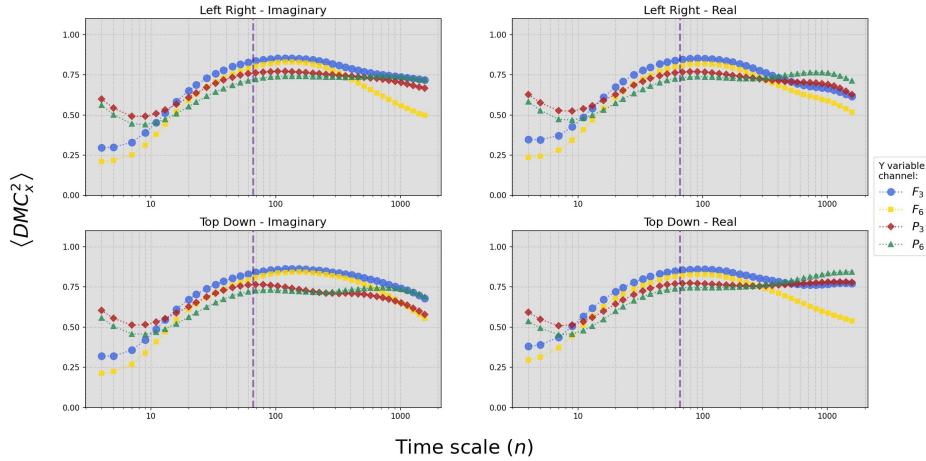


Figure 5: Mean values of  $DMC_x^2 \times n$  for all tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for S039 subject.

The Fig. 5, for the S039 subject, show that for small time scale,  $n \leq 10$ , the parietal channels,  $P_3$  and  $P_6$ , as a  $Y$  variable have values for  $\langle DMC_x^2 \rangle$  higher if compared to the front channels,  $F_3$  and  $F_6$ . The results presented in the

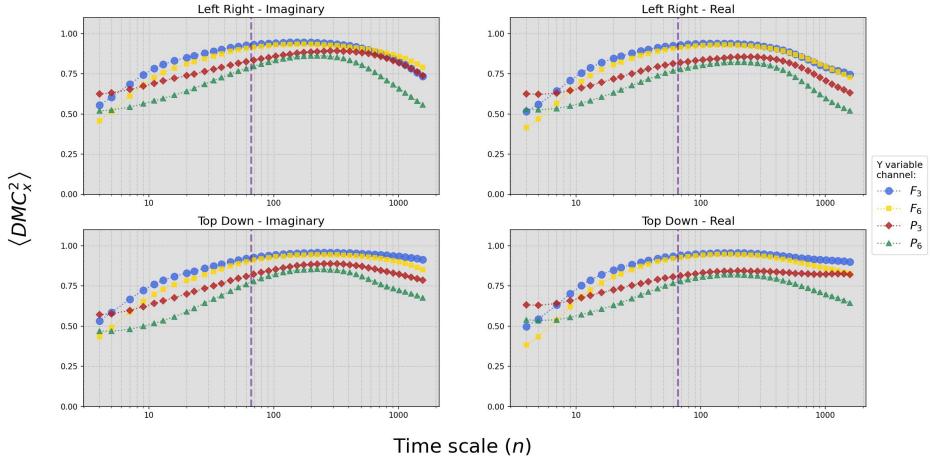


Figure 6: Mean values of  $\langle DMC_x^2 \rangle$  for all tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for S078 subject.

Fig. 6, the values of  $\langle DMC_x^2 \rangle$  are very similar for all tasks, with its values increasing until  $n = 67$ . Finishing this small sample, in the Fig. 7 we present

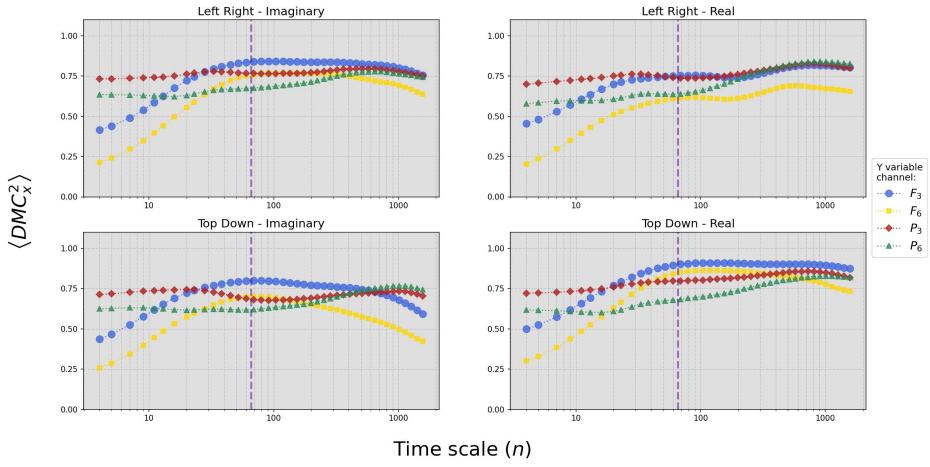


Figure 7: Mean values of  $\langle DMC_x^2 \rangle$  for all tasks: Left/Right (Imaginary), Left/Right (Real), Top/Down (Imaginary), and Top/Down (Real) for S099 subject.

the results for S099 subject. In this figure it is clear that this subject has a

global behavior, in terms of  $\langle DMC_x^2 \rangle$ , different from the previous subjects, thus identifying a possible signature in the EEG stimulus/response measurement, when performing these four tasks (Real/Imaginary or Top/Down).

Next, for a complete visualization of this motor activity (Real/Imaginary) which confirms our assumptions, we created one on-line public repository with  $\langle DMC_x^2 \rangle$  for all subjects (with figures and videos), available at this link:

[https://255ribeiro.github.io/Multi\\_Cross-correlation\\_EEG/web\\_report/web\\_site/subject\\_mean.html](https://255ribeiro.github.io/Multi_Cross-correlation_EEG/web_report/web_site/subject_mean.html)

Our public repository is very complete, for example, we can also see the difference between the Real/Imaginary task in terms of the DCCA multiple cross-correlation coefficient, for each subject  $j$ , that is:

$$diff(i - r)_j(n) \equiv \langle DMC_{x \text{ } imaginary}^2 \rangle_j(n) - \langle DMC_{x \text{ } real}^2 \rangle_j(n) \quad (4)$$

To visualize the difference between real actions and imagining the same action, a set of graphics was developed where the imaginary task of a top-down and left-right  $\langle DMC_{x \text{ } imaginary}^2 \rangle_j$  action are subtracted from the real value of the same action for every subject  $j$  (Eq. 4). See this description at the link:

[https://255ribeiro.github.io/Multi\\_Cross-correlation\\_EEG/web\\_report/web\\_site/diff\\_ir.html](https://255ribeiro.github.io/Multi_Cross-correlation_EEG/web_report/web_site/diff_ir.html)

The values of  $diff(i - r)_j$  are small, denoting that, the application of the  $DMC_x^2$  is equivalent for imaginary and real case, mainly for  $n \leq 67$ . Following our analyses, we also calculated the global mean for DCCA multiple cross-correlation, with 108 subjects:

$$\overline{\langle DMC_x^2 \rangle}(n) = \frac{1}{108} \sum_{i=1}^{108} \langle DMC_x^2 \rangle_i(n) \quad (5)$$

See Fig. 8, were  $\overline{\langle DMC_x^2 \rangle}$  is plotting as a function of  $n$ . In this figure we can see no appreciable differences between these four tasks (Left/Right and Real/Imaginary). The channel  $F_3$  ( $Y$ ) (blue circles), presents the greatest value for DCCA multiple cross-correlation coefficient, mainly around  $n = 67$  or  $0.42s$  ( $\sim 2.38\text{Hz}$ ). It is known that, the most characteristic signals observed on EEG signal are between 1 and 20 Hz.

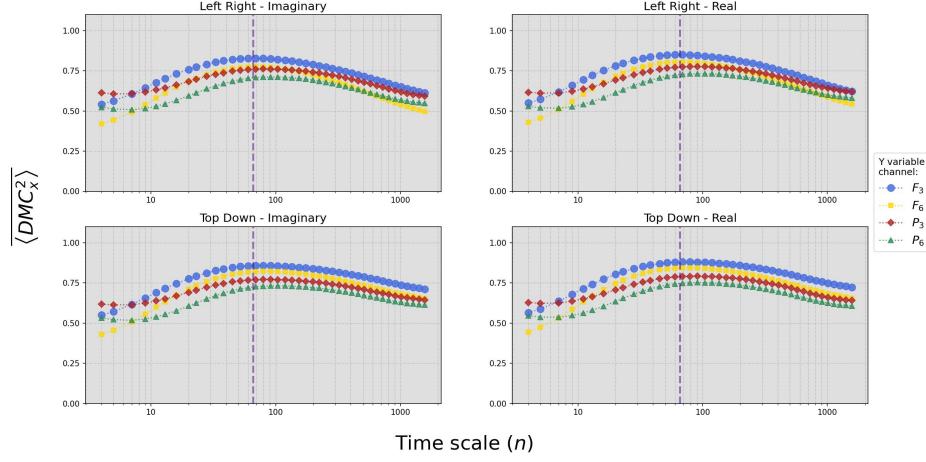


Figure 8:  $\langle DMC_x^2 \rangle \times n$  global mean for all Subjects and task (Left/Right and Real/Imaginary).

To investigate the dispersion in global mean, the standard deviation  $sd$  of  $\langle DMC_x^2 \rangle$  for all 108 subject was calculated. The results are shown in the Fig. 9. It can be seen that the  $sd$  is maximum for the channel  $F_6$  (yellow

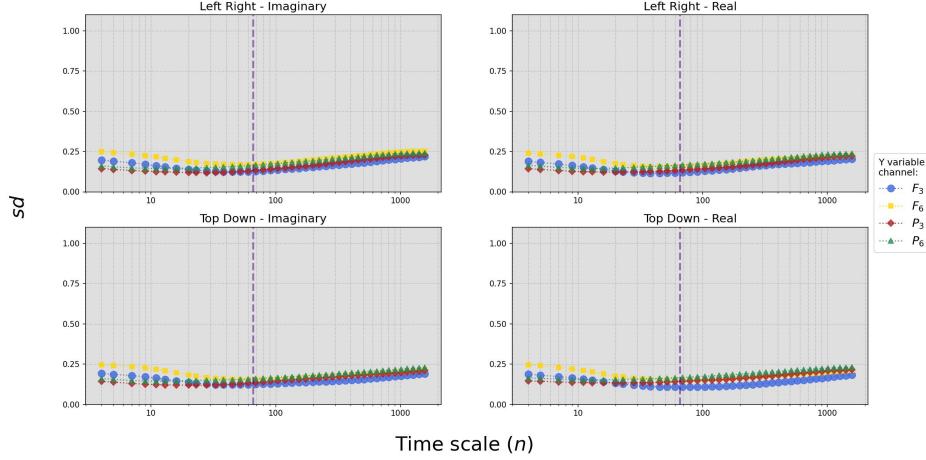


Figure 9:  $sd$  (standard deviation) of  $\langle DMC_x^2 \rangle$  for all Subjects and task.

circles), more noticeable on time scales smaller than  $n = 67$  (minimum value).

The value of  $\overline{\langle DMC_x^2 \rangle}$  is of great importance, because we have it calculated for a great set of subjects. Thus, we can assess how much each subject ( $j$ ) deviates in relation to this global value, highlighting your individual pattern in relation to the aforementioned value, that is, through the equation

$$diff_j(n) \equiv \langle DMC_x^2 \rangle_j(n) - \overline{\langle DMC_x^2 \rangle}(n) \quad (6)$$

To corroborate with this calculation, a complete collection of figures containing the differences between each subject ( $j$ ) and its global mean are available at the link:

[https://255ribeiro.github.io/Multi\\_Cross-correlation\\_EEG/web\\_report/web\\_site/diff\\_mean.html](https://255ribeiro.github.io/Multi_Cross-correlation_EEG/web_report/web_site/diff_mean.html)

Looking at the  $diff_j(n)$ , we clearly see individual patterns, that can be compared with the global mean pattern (now our reference), in the sense for example, to identify in this subject neurological abnormalities in this stimulus/response motor activity. Another way to identify individual patterns based on an average value, can be through by the mean square error function,  $MSE$ , therefore, for each subject  $j$  this function can be written by:

$$MSE(j) = \frac{1}{QoD} \sum_{i=1}^{QoD} \left( \langle DMC_x^2 \rangle[i] - \overline{\langle DMC_x^2 \rangle}[i] \right)^2 \quad (7)$$

Where  $QoD$  is the total amount of time scales used ( $QoD = 42$  in this paper, see insert in the Fig. 2 as an example). The results for  $MSE(j)$  per channel ( $F_3$ ,  $F_6$ ,  $P_3$ , and  $P_6$ ) and task can be seeing in the Figs. 10, 11, 12, and 13.

These figures show how much each subject deviate about the global mean, by channel and task. In this sense, those subjects with the highest  $MSE$  values are those who deserve a more detailed study.

#### 4. Conclusion

This paper proposes one statistical tool to study EEG signals in a motor activity, with the application of the DCCA multiple cross-correlation coefficient taking into account a complete set of 108 subjects in four tasks (Real/Imaginary). In this case, the international 10 – 10 system was used and four specific channels were selected ( $F_3$ ,  $F_6$ ,  $P_3$ , and  $P_6$ ). As a result, the DCCA multiple cross-correlation coefficient identified that, there is a single

signature for each subject and this visual stimulus/response presents a particular behavior as a fingerprint. Globally for these 108 subject, there is no significant difference between the imaginary and the real task. The frontal channels ( $F_3$  and  $F_6$ ) commonly had greater DCCA multiple cross-correlation values than the parietal ones ( $P_3$  and  $P_6$ ), especially for time scales around  $n = 66$  (0.42s or 2.38 Hz) where there is a smaller standard deviations. Finally, according with our results, the proposal of study multiple time series at the same time using the DCCA detrended multiple cross-correlation coefficient is feasible and robust for EEG analysis. The advantage of this method in relation to the others, in frequency domain, is that  $DMC_x^2$  works directly in time-scale. This new coefficient is robust to analyze non-stationary time-series in multiple EEG channels. We have as a novelty the quantification of the multiple cross-correlations of each subject for a large range of time-scales, in a significant sample. As a final result, we obtained a global mean and the individual differences, as well the mean squared errors for each subject (listed in this paper as figures for each channel). This study probably will be the kick-off of a new approach to analyzing multiple cross-correlations in EEG signals.

## 5. Acknowledgments

G. F. Zebende and F. M. Oliveira Filho thanks the Brazilian agency CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) (Grants 310136/2020-2 and 150655/2022-3, respectively). We would also like to thank Paulo Murilo Castro de Oliveira, great statistical physicist (*in memoriam*).

## 6. References

### References

- [1] M. Tudor, L. Tudor, K. I. Tudor, Hans Berger (1873-1941) - the history of electroencephalography, *Acta Medica Croatica* 59 (4) (2005) 307–313.
- [2] A. Biasiucci, B. Franceschiello, M. M. Murray, Electroencephalography, *Curr Biol.* 29 (30) (2019) R80–R85.
- [3] G. F. Zebende, F. M. Oliveira-Filho, J. A. Leyva-Cruz, Auto-correlation in the motor/imaginary human EEG signals: A vision about the FDFA fluctuations, *PLoS One* 12 (9) (2017).

- [4] F. M. Oliveira Filho, J. A. Leyva Cruz, G. F. Zebende, Analysis of the EEG bio-signals during the reading task by DFA method, *Physica A: Statistical Mechanics and its Applications* 525 (2019) 664–671.
- [5] V. B. Mesquita, F. M. Oliveira-Filho, P. C. Rodrigues, Detection of crossover points in detrended fluctuation analysis: An application to EEG signals of patients with epilepsy, *Bioinformatics* (11 2020). doi:10.1093/bioinformatics/btaa955.
- [6] F. Oliveira Filho, F. Ribeiro, J. L. Cruz, A. N. de Castro, G. Zebende, Statistical study of the eeg in motor tasks (real and imaginary), *Physica A: Statistical Mechanics and its Applications* (2023) 128802.
- [7] G. Zebende, A. da Silva Filho, Detrended multiple cross-correlation coefficient, *Physica A: Statistical Mechanics and its Applications* 510 (2018) 91–97.
- [8] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, J. E. Wolpaw, BCI2000: a general-purpose brain-computer interface (BCI) system, *IEEE Trans Biomed Eng.* 51 (6) (2004) 1034–1043.
- [9] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, A. L. Goldberger, Mosaic organization of DNA nucleotides, *Phys. Rev. E* 49 (1994) 1685–1689.
- [10] B. Podobnik, Z.-Q. Jiang, W.-X. Zhou, H. E. Stanley, Statistical tests for power-law cross-correlated processes, *Physical Review E* 84 (2011) 1–8.
- [11] J. C. Zhao, W. H. Dou, H. D. Ji, J. Wang, Detrended Cross-Correlation Analysis of Epilepsy Electroencephalogram Signals, *Advanced Materials Research* 765–767 (2013) 2664–2667.
- [12] Y. Chen, L. Cai, R. Wang, Z. Song, B. Deng, J. Wang, H. Yu, DCCA cross-correlation coefficients reveals the change of both synchronization and oscillation in EEG of Alzheimer disease patients, *Physica A: Statistical Mechanics and its Applications* 490 (2018) 171–184.
- [13] G. F. Zebende, DCCA cross-correlation coefficient: Quantifying level of cross-correlation, *Physica A* 390 (4) (2011) 614–618.

- [14] L. Kristoufek, Measuring cross-correlation between non-stationary series with DCCA coefficient, *Physica A: Statistical Mechanics and its Applications* 402 (2014) 291–298.
- [15] A. A. Brito, H. A. Araujo, G. F. Zebende, Detrended multiple cross-correlation coefficient applied to solar radiation, air temperature and relative humidity, *Sci. Rep.* 9 (2019) 19764. doi:10.1038/s41598-019-56114-6.
- [16] A. M. da Silva Filho, G. F. Zebende, A. P. N. de Castro, E. F. Guedes, Statistical test for multiple detrended cross-correlation coefficient, *Physica A: Statistical Mechanics and its Applications* 562 (2021) 125285.
- [17] E. F. Guedes, A. M. da Silva Filho, G. F. Zebende, Detrended multiple cross-correlation coefficient with sliding windows approach, *Physica A: Statistical Mechanics and its Applications* 574 (2021) 125990.
- [18] E. F. Guedes, A. P. N. de Castro, A. M. da Silva Filho, G. F. Zebende,  $\delta$  dmc x 2: A new approach to measure contagion effect on financial crisis, *Fluctuation and Noise Letters* 21 (04) (2022) 2250026.
- [19] G. F. Zebende, L. C. Aguiar, P. Ferreira, E. F. Guedes, Statistical approach to study the relationship between stock market indexes by multiple dcca cross-correlation coefficient, *Fluctuation and Noise Letters* 21 (5) (2022) 2250045–1314.

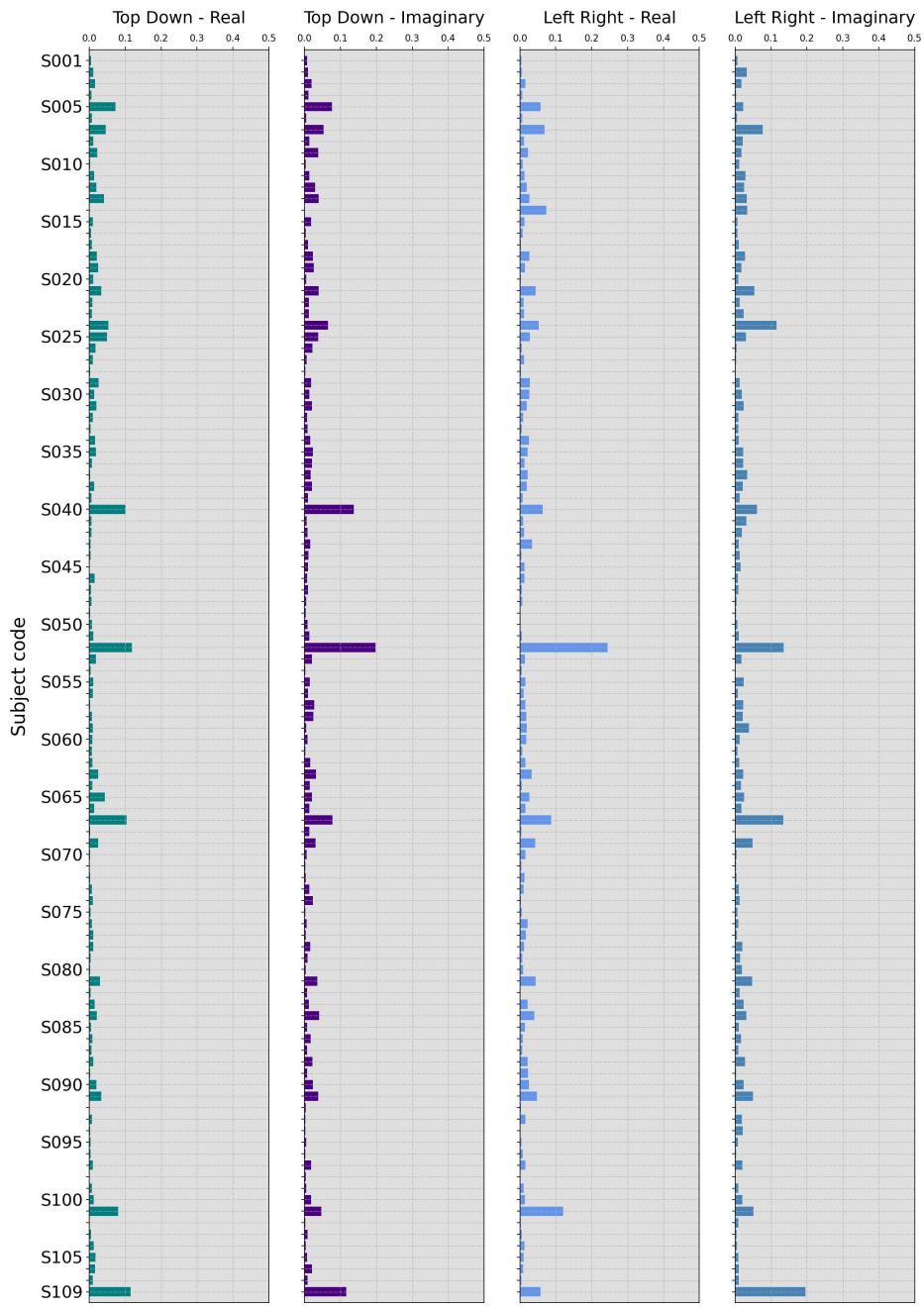


Figure 10:  $MSE$  for the channel  $F_3$ .

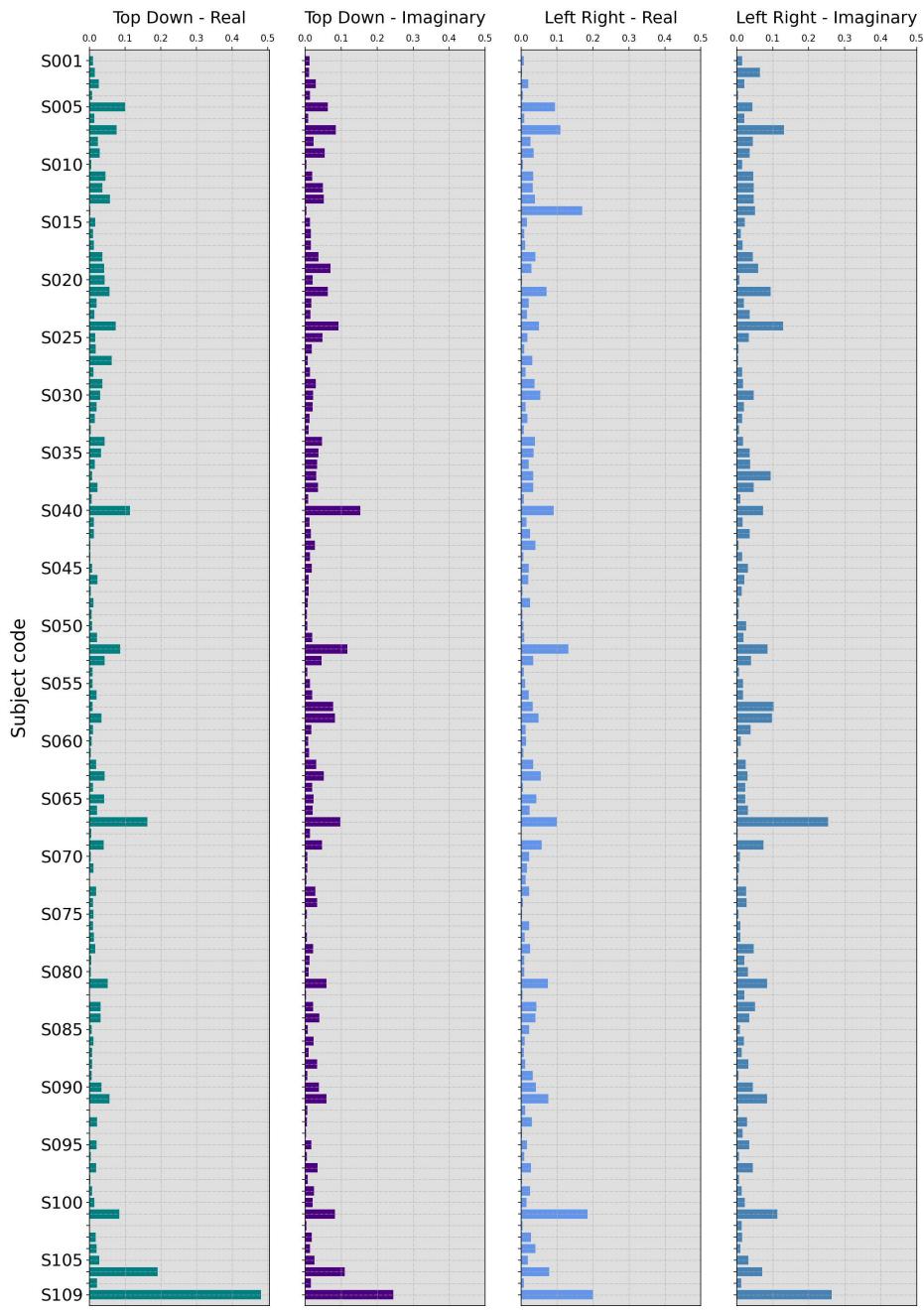


Figure 11:  $MSE$  for the Channel  $F_6$ .

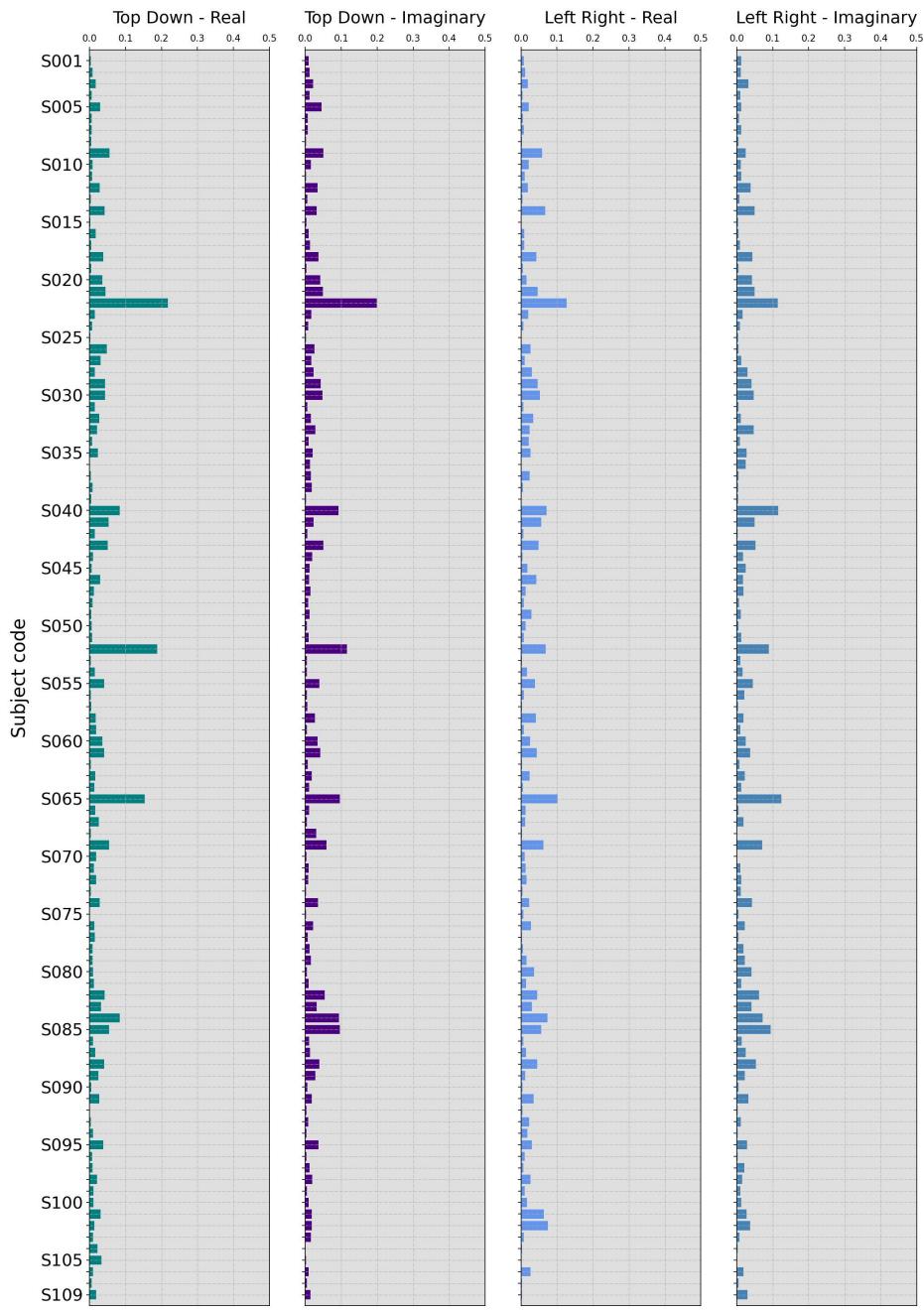


Figure 12:  $MSE$  for the Channel  $P_3$ .

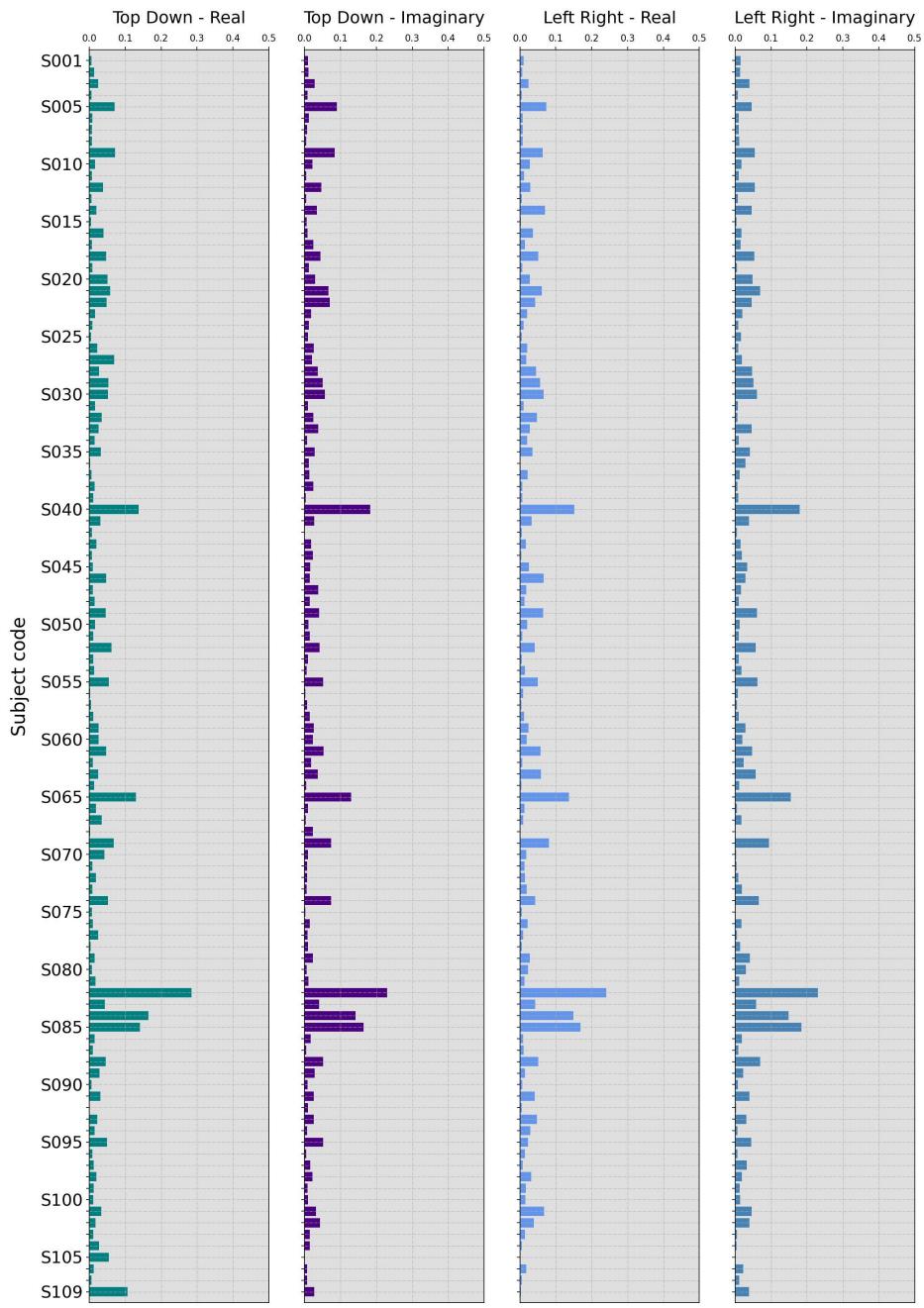


Figure 13:  $MSE$  for the Channel  $P_6$ .

### 3.2 Artigo 02: A Python/Zig optimized and customizable implementation for $\rho_{DCCA}$ and $DMC_x^2$ coefficients

“Just remember, once you’re over the hill  
you begin to pick up speed.”

(Charles M. Schulz)

O artigo *A Python/Zig optimized and customizable implementation for  $\rho_{DCCA}$  and  $DMC_x^2$  coefficients* está sendo revisado com o objetivo de ser publicado na revista *Journal of Statistical Software*. Apresenta um pacote *Python* com métodos de cálculo das funções *DFA* e *DCCA*e dos coeficientes  $\rho_{DCCA}$  e  $DMC_x^2$ . Os cálculos mais custosos foram implementados em programação de nível mais baixo, utilizando uma linguagem de sistemas, nova e promissora, chamada *Zig* (<https://ziglang.org/>). A escolha da linguagem *Zig* se deu por vários motivos:

- Por ser uma linguagem de baixo nível, obtendo velocidades de processamento compatíveis com *C* e *Fortran* (KACS et al., 2024; KACS; BROWN; LEE, 2024).
- Possui compatibilidade com *C* e *C++*, podendo incorporar bibliotecas.
- Possui uma sintaxe moderna e elegante para tratar problemas de baixo nível.
- Tem um eficiente sistema de *cross-compilation*.

O aspecto das *cross-compilation* foi um dos mais considerados na escolha da linguagem. A manutenção de uma biblioteca por um pequeno grupo de pessoas possui diversos desafios. A capacidade de fornecer programas compilados para diversos sistemas operacionais e arquiteturas de processadores é uma etapa crucial para a popularização de um pacote. A possibilidade de se gerar os executáveis desta biblioteca em um único computador pessoal pareceu uma opção vantajosa.

O *software* de cálculo de dinâmica de fluidos computacional *AeroSim* (<https://aerosim.io/>) apresenta desempenho e precisão nos seus cálculos (ROMANUS et al., 2023; LUGARINI; ROMANUS; JUNIOR, 2024) é parcialmente implementado em *Zig* e serviu de incentivo à esta adoção.

Parte da otimização segue a implementação de Hartmann et al. (2013) para o *DFA*, transposta para o *DCCA* e  $\rho_{DCCA}$  por Kaposzta et al. (2022). Onde, durante o cálculo da interpolação da reta de tendência, na primeira caixa de cada séries, pelo método dos mínimos quadrados, valores são gravados em variáveis que podem tornar o *loop* de

### Algoritmo 3 Detrended Saved

---

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  $DV_{k,i}^j = (X_{k,i}^j - \bar{X}_{k,i}^j)$  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  $DCCA$ .

---

somatários dos valores utilizados no cálculo dos coeficientes da equação da reta são armazenados em variáveis. Para a caixa seguinte, o *loop* é substituído por operações de subtração e adição.

A Equação 3.2, mostra que o valor da soma das coordenadas no eixo das abscissas de uma caixa pode ser substituído pelo somatório total na caixa anterior menos o primeiro valor de abscissa da caixa anterior, somado ao último valor de  $x$  para a caixa atual. Raciocínio análogo pode ser aplicado às Equações 3.3, 3.4 e 3.5.

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} T_k = \left( \sum_{j=i-1}^{(i+n)-1} T_j \right) - T_{i-1} + T_{i+n} \quad (3.2)$$

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} T_k^2 = \left( \sum_{j=i-1}^{(i+n)-1} T_j^2 \right) - T_{i-1}^2 + T_{i+n}^2 \quad (3.3)$$

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} S_k = \left( \sum_{j=i-1}^{(i+n)-1} S_j \right) - S_{i-1} + S_{i+n} \quad (3.4)$$

$$\forall 1 < i \leq (N-n), \sum_{k=i}^{i+n} (S_k \times T_k) = \left( \sum_{j=i-1}^{(i+n)-1} (S_j \times T_j) \right) - (S_{i-1} \times T_{i-1}) + (S_{i+n} \times T_{i+n}) \quad (3.5)$$

Quando se pensa no cálculo do  $DMC_x^2$ , e em como otimiza-lo, deve-se pensar em múltiplos cálculos de  $\rho_{DCCA}$ , para a criação da matriz apresentada na Equação 2.8. A ideia é substituir o passo 4 do Algoritmo 2 pelos passos descritos no Algoritmo 3.

Essa pequena mudança, embora possa até aumentar o tempo de processamento para um pequeno número de séries temporais, demonstra-se muito vantajoso quanto maior for o número de séries cujo coeficiente  $\rho_{DCCA}$  precisa ser calculado.

$$\frac{j!}{2 \times (j-2)!} \quad (3.6)$$

Na biblioteca *Zebende* é possível implementar outras versões do código para o cálculo do  $\rho_{DCCA}$  com poucas séries temporais. Mas é preciso entender em que ponto o Algoritmo 3 passa a ser vantajoso. Caso o *DV* não seja armazenado, o número de vezes que ele tem que ser calculado é de duas vezes o número de combinações duas a duas para  $j$  séries (Equação 3.6) e o tempo gasto para calcular o *DV* depende do tamanho das séries.

Vale também ressaltar que as otimizações baseadas no salvamento dos parâmetros do método dos mínimos quadrados pode ser adaptado para diversos graus do polinômio da tendência, mas não pode se adaptar à caixas não sobrepostas.

Como apresentado na Sessão 2.3 do Capítulo 2, alguns autores ainda advogam pela não sobreposição das caixas (ZHOU, 2008), nestes casos, o armazenamento dos parâmetros para o cálculo dos mínimos quadrados não funcionaria. Por outro lado, para uma grande quantidade de séries temporais o Algorítimo 3, *Detrended Saved*, funcionaria também no cenário de caixas não sobrepostas.

Além da otimização, o artigo apresenta o modo de uso da biblioteca e algumas das funções auxiliares já implementadas. As equações destacadas nas Sessões 2.3 e 2.4 do Capítulo 2 estão entre as próximas adições ao pacote.

O código fonte da biblioteca *Zebende* está disponível em <https://github.com/255ribeiro/zebende>.

# A Python/Zig optimized and customizable implementation for $\rho_{DCCA}$ and $DMC_x^2$ coefficients

Fernando Ferraz Ribeiro    
Federal University of Bahia

Gilney Figueira Zebende    
State University  
of Feira de Santana

---

## Abstract

This paper presents **Zebende**, a Python package written in Python and Zig, that calculates the *DFA*, *DCCA* functions, and  $\rho_{DCCA}$  and  $DMC_x^2$  coefficients, along with auxiliary functions to help the calculations process. The package presents an optimized algorithm that significantly improves the calculations speed in large datasets. **Zebende** is the first package that implements the  $DMC_x^2$  and can be applied to any number of time series. The optimized version of the algorithm is compared to the standard implementation, presenting valuable performance increase. A comparison with other Python and R packages that calculates the  $\rho_{DCCA}$  is also presented.

*Keywords:*  $\rho_{DCCA}$ ,  $DMC_x^2$ , optimization, Python, Zig.

---

## 1. introduction

The Detrended Cross-Correlation Coefficient ( $\rho_{DCCA}$ ) (Zebende 2011) is a widely used coefficient that measures the cross-correlation between two non-stationary time series.  $\rho_{DCCA}$  is an extension of the Detrended Fluctuation Analysis (*DFA*) (Peng, Buldyrev, Havlin, Simons, Stanley, and Goldberger 1994) and the Detrended Cross-correlation Analysis (*DCCA*) (Podobnik and Stanley 2008): while the *DFA* calculates the self-affinity and long-memory properties of a time series data, and the *DCCA* analyses power-law cross correlations between two different non-stationary time series, the  $\rho_{DCCA}$  coefficient quantifies this cross-correlation in simple values ranging from  $-1$  to  $1$ , where  $-1$  indicates a perfect anti-correlation between the series,  $1$  a perfect correlation and zero ( $0$ ) no correlation at all.

The method was statistically validated (Podobnik, Jiang, Zhou, and Stanley 2011), applied to climate data Vassoler and Zebende (2012), financial data(Wang, Xie, Chen, Yang, and Yang 2013; Guedes, Dionísio, Ferreira, and Zebende 2017; Ferreira, Dionísio, Guedes, and Zebende 2018), electroencephalogram recorded data (Filho, Ribeiro, Cruz, De Castro, and Zebende 2023), among others. Statistical criteria for evaluating the relevance of the results were also developed Guedes, Brito, Oliveira Filho, Fernandez, de Castro, da Silva Filho, and Zebende (2018a,b).

The Detrended Multiple Cross-Correlation Coefficient (Zebende and Silva 2018) ( $DMC_x^2$ ) is a generalization of the  $\rho_{DCCA}$  coefficient that correlates one time series (dependent variable) a number of time series (independent variables). The  $DMC_x^2$  values ranges from zero ( $0$ ), indicating no multiple cross-correlation to  $1$ , meaning perfect multiple cross-correlation.

The coefficient has been applied to study meteorological data (Brito, Araujo, and Zebende 2019), have been tested statistically (da Silva Filho, Zebende, de Castro, and Guedes 2021), used with the implementation of sliding windows (Guedes, da Silva Filho, and Zebende 2021), to measure the contagion effect on stock market indexes (Guedes, de Castro, da Silva Filho, and Zebende 2022), to produce statistical analysis between stock market behavior (Zebende, Aguiar, Ferreira, and Guedes 2022) or analyze electroencephalogram (EEG) signals (Ribeiro, de Almeida Brito, Filho, Cruz, and Zebende 2025).

This paper presents the **Zebende** Python package, an implementation of the *DFA*, *DCCA*,  $\rho_{DCCA}$ ,  $DMC_x^2$  and utility functions related to the methods. In section 2 the steps for calculating the  $\rho_{DCCA}$  and  $DMC_x^2$  are presented and discussed. Section 3 shows how this library was implemented, the optimization technics and the recommended steps to use the library. In Section 4 the **Zebende** package is compared with other packages for Python and R that calculates the  $\rho_{DCCA}$  in terms of performance and usability, leading to the conclusions in Section 5.

## 2. Algorithms of the coefficients

The algorithms that calculates the  $\rho_{DCCA}$  uses the *DFA* and the *DCCA* steps. The  $DMC_x^2$  coefficient uses the  $\rho_{DCCA}$  coefficient and, consequently, also embraces the *DFA* and the *DCCA*. The *DFA* method is described in six steps:

1. Taking a time series  $\{x_i\}$  with  $i$  ranging from 1 to  $N$ , the integrated series  $X_k$  is calculated by  $X_k = \sum_{i=1}^k [x_i - \langle x \rangle]$  with  $k$  also ranging from 1 to  $N$  (time series length);
2. the  $X_k$  series is divided in  $N - n$  boxes of size  $n$  (time scale), each box containing  $n + 1$  values, starting in  $i$  up to  $i + n$ ;
3. for each box, a polynomial (usually of degree 1) best fit is calculated, getting  $\tilde{X}_{k,i}$  with  $i \leq k \leq (i + n)$ ;
4. in each box is calculated:  $f_{DFA}^2(n, i) = \frac{1}{1+n} \sum_{k=i}^{i+n} (X_{k,i} - \tilde{X}_{k,i})^2$
5. for all the boxes of a time scale, the *DFA* is calculated as:

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

6. for a number of different timescales ( $n$ ), with possible values  $4 \leq n \leq \frac{N}{4}$  the  $F_{DFA}$  function is calculated to find a relation among  $F_{DFA} \times n$

The *DCCA* method is very similar to the *DFA* calculations, with the difference of analyzing two series while the *DFA* evaluate properties of a single time series. It's also a six steps process:

1. Taking two time series of the same length  $\{x_i^{j1}\}$  and  $\{x_i^{j2}\}$  with  $i$  ranging from 1 to  $N$ , the integrated series  $X\alpha_k$  and  $X\beta_k$  are calculated by  $X_k = \sum_{i=1}^k [x_i^{ji} - \langle x \rangle]$  for each series, with  $k$  also ranging from  $i$  to  $N$ ;
2. The series  $X_k^{j1}$  and  $X_k^{j2}$  are divided into  $N - n$  boxes of size  $n$  (time scale), each box containing  $n + 1$  values, starting from  $i$  to  $i + n$ ;

3. For each box, a polynomial (usually of degree 1) is fitted, obtaining  $\widetilde{X^{j1}}_{k,i}$  and  $\widetilde{X^{j2}}_{k,i}$ , for series  $\{x_i^{j1}\}$  and  $\{x_i^{j2}\}$  respectively, with  $i \leq k \leq (i+n)$ ;
4. For each box, one of the  $N - n$  boxes of the same temporal scale, the function is calculated by:

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

5. For all boxes of the same temporal scale, the DCCA is calculated as:

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

6. For a number of temporal scales ( $n$ ), with possible values  $4 \leq n \leq \frac{N}{4}$ , the DCCA is calculated to find a relationship between  $F_{DCCA} \times n$

Comparing the algorithms, the first three are basically identical, the only difference is that the *DCCA* method apply those steps to two series. The step four of the *DFA* can be considered an analogous of the variance, replacing the average subtraction (in the variance) for the values obtained by the polynomial fit (estimated series); and the equivalent step of the *DCCA* is, in the same terms, compared to the covariance between the two series. The technique of fitting a curve, interpreted as a trend inside each box, and subtracting the value estimated by the trend from the actual value in the integrated series ( $X_k - \tilde{X}_k$ ) from now on will be called **detrended values**(*DV*). In the *DFA* algorithm, the  $f_{DFA}^2(n, i)$  function is the mean of the square of the *DV*, in *DCCA* calculations, the  $f_{DCCA}^2(n, i)$  function evaluates the mean of the product of the *DV* of the two series in each box.

Step five of the *DFA* calculates the square root of the mean of the values calculates in the previous step for each box, in the *DCCA*, the mean of the values evaluated for each box is calculated in stead. The last step, in both cases, is more a reminder to repeat the respective previous operations for a number of difference time scales ( $n$ ).

The  $\rho_{DCCA}$  is measured by Eq. 1. Considering the relation between *DFA* and variance and *DCCA* and covariance, the  $\rho_{DCCA}$  resembles Pearson correlation for a time scale  $n$ .

$$\rho_{DCCA}(n) = \frac{F_{DCCA}^2(x_\alpha, x_\beta)(n)}{F_{DFA}(x_\alpha)(n) \times F_{DFA}(x_\beta)(n)} \quad (1)$$

The  $DMC_x^2$  is a generalization of the  $\rho_{DCCA}$  that calculates the correlation between one time-series  $\{Y\}$ , as the dependent variable, and a number  $m$  of time-series  $\{X^1\}$ ,  $\{X^2\}$ ,  $\{X^3\}$ , ...,  $\{X^m\}$  defined as independent variables and represented as  $\{X^j\}$ , with  $j$  ranging from 1 to  $m$ . The coefficient is expressed mathematically as:

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

In Eq. 2,  $\rho^{-1}(n)$  represent the inverse of a matrix populated by all possible combinations of  $\rho_{DCCA}$  between independent variables. In Eq. 3, value  $\rho_{X^1, X^2}(n)$ , for instance, is the  $\rho_{DCCA}$  for independent variables  $X^1$  and  $X^2$  calculated with time scale  $n$ , occupying position  $\rho_{12}$  of the matrix. Two fundamental characteristics: the first is that the main diagonal values are all ones, since it's position in the matrix denotes the calculation of a cross-correlation between

a series and itself. Second, the matrix is symmetric in relation to the main diagonal, as the  $\rho_{DCCA}$  is a commutative operation.

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

At last Eq. 4 represent the transposed vector of the  $\rho_{Y,X_i}(n)$  between the depended variable  $\{Y\}$  and each  $\{X_i\}$  independent variable for a given time scale  $n$ .

$$\rho_{Y,X^j}(n)^T = [\rho_{Y,X^1}(n), \rho_{Y,X^2}(n), \dots, \rho_{Y,X^m}(n)] \quad (4)$$

As the *DFA* and the *DCCA*,  $\rho_{DCCA}$  and  $DMC_x^2$  should be evaluated in a number of time scales ( $n$ ) to analyze the characteristics of each coefficient.

### 3. Zebende package: implementation and optimization

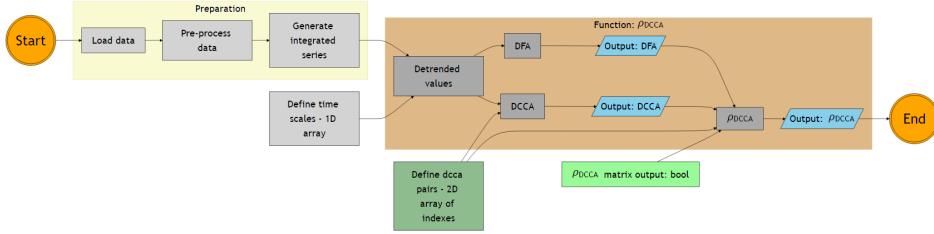
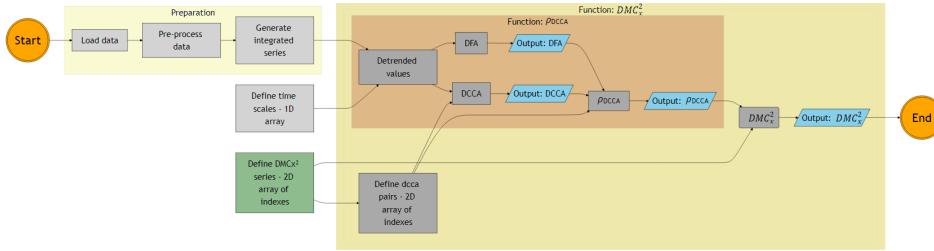
The implementation of the **Zebende** package follows some well defined goals:

1. Enhance performance;
2. avoid redundant calculations;
3. make the outputs compatibles with other data analyses tools (including data manipulation, machine learn and statistical packages);
4. manage multiple time series inputs;
5. operate the  $DMC_x^2$  for any number of series;
6. create a customizable and modular set of tools;
7. facilitate package evolution and maintenance;
8. deliver an easy to use package.

The **Python** language was chosen because it is one of the most used languages in the data analyses field and have a great support for statistical tools and machine learning algorithms. There are a plethora of tools to load and manipulate data (**Pandas**, **Polar**, **PySpark** ...), execute statistical analyzes (**Numpy**, **SciPy**, **StatsModels** ...), machine learning (**Pytorch**, **TensorFlow**, **Scikit Learn** ...) and data visualization (**Matplotlib**, **Seaborn**, **Plotly** ...) among other data related applications.

The first draft of the code was written in pure **Python**, to rapidly prototype the way users will interact with the package. Figures 1 and 2 presents simplified flowcharts illustrating how to use the package and how the main functions ( $\rho_{DCCA}$  and  $DMC_x^2$ ) works.

The preparation steps are the same in both functions. First the data is loaded, and should be analyzed by the researchers. Based on the data characteristics, the set should be treated

Figure 1: Calculating  $\rho_{DCCA}$  with **Zebende** package - Simplified flowchartFigure 2: Calculating  $DMC_x^2$  with **Zebende** package - Simplified flowchart

to ensure the methods requirements in the "Pre-processing" stage. The package functions expects data as a matrix with the columns as the series and the lines as time steps. Columns unwanted in the indented research should also be dropped for better performance of the algorithms in this step. The more common way to do that is to use a data manipulation package. To proceed to the next step, the data table should be in the form of a **Numpy** 2D array. Any data manipulation **Python** package can export a table as a **Numpy** array. The next step is to calculate the integrated series. The package provides a function, named **integrated\_series()**, to calculate that. The code example below show how to load the libraries (using **Pandas** as the data manipulation packages and loading a **.csv** file as a generic example), convert to **Numpy** array and generate the integrated series.

```

# importing packages
import numpy as np
import pandas as pd
import zebende as zb

data = pd.read_csv('path_to_the_file.csv') # loading data
# Pre-processing data
# ...
data = data.to_numpy() #converting data to Numpy array
int_data = zb.integrated_series(data) # calculating the integrated series

```

The option of taking out the integrated series generation from the main methods (**p\_dcca()** and **dmcx2()**) to an independent one was inspired by Peng *et al.* (1994) work, where the way of calculating integrated series was different from the one that is widely used in more recent years. The integration of the series is essentially a pre-processing step, and this approach makes easy to explore alternative ways to integrate the series or compare Peng *et al.* (1994)

approach to the current most used one in different scenarios, or even embrace new proposals for the series integrating step.

The input and output structures of each function are displayed below:

```
def p_dcca(
    input_data: NDArray[float64],
    tws: NDArray[int64] | NDArray[float64],
    DCCA_of: ndarray | Literal['all'] = "all",
    P_DCCA_output_matrix: bool = False
) -> tuple[NDArray[float64], # DFA
           NDArray[float64], # DCCA
           NDArray[float64]    # P_DCCA
          ]

def dmcx2(
    input_data: NDArray[float64],
    tws: NDArray[int64] | NDArray[float64],
    dmcx2_of: ENUM_DMCx2_of | NDArray[float64] | list = 'all-full'
) -> tuple[ NDArray[float64], # DFA
            NDArray[float64], # DCCA
            NDArray[float64], # P_DCCA
            NDArray[float64], # DMC
           ]
```

The first two inputs are the same for functions `p_dcca()` and `dmcx2()`: `input_data` receive the integrated series and the `tws` receives an 1D array representing the time scales (box size) described in the algorithms on Section 2. The `input_data` is a 2D array of 64 bits floating point data. The `tws` accepts integers and, for convenience reasons, also floating points. Since the size of the boxes needs to be integers, in case of floating points, the values will be converted to integers by ignoring the decimal values (truncating). This two inputs are colored in light gray in Figures 1 and 2, indicating mandatory inputs.

With the mandatory steps explained, some very important optional inputs should be addressed. Starting with the  $\rho_{DCCA}$  function (dark green node in Figure 2) represents the input `DCCA_of` of the function. This input requires a 2D array of integers, each row is a pair of index, related to the `data_input` matrix. For example: if the `input_data` receives a four columns matrix, with index ranging from 0 to 3, that is intended to calculate de *DFA* for all the series but the  $\rho_{DCCA}$  between series of index 0 and 1 and also for series 2 and 3, the `DCCA_of` input should receive the array `[[0,1], [2,3]]`. If no value (or the string '`'all'`) is given, the function will calculate all possible combinations of *DCCA* calculations between all the series respecting the index values order, as below:

```
[[0,1], [0,2], [0,3], [1,2], [1,3], [2,3]]
```

### 3.1. Implementation of the Detrended Cross-correlation Coefficient function

Is very important to understand the calculation steps, the role of the `DCCA_of` array and how

it fits in the goals of the package implementation. The code below is part of the pure Python implementation of the  $\rho_{DCCA}$  function,

```

for n_index, n in enumerate(tws): # for each time scale
    # temporary allocation arrays
    f2dfa_n = np.full(shape=(data.shape[0] - n, data.shape[1]),
                      fill_value=np.nan, dtype=data.dtype)
    dcca_n = np.full(shape=(data.shape[0] - n, DCCA_of.shape[0]),
                      fill_value=np.nan, dtype=data.dtype)
    detrended_mat = np.full(shape=(n + 1, data.shape[1]),
                           fill_value=np.nan, dtype=data.dtype)
    for i in range(data.shape[0] - n): # for each box
        detrended_series( # inputs
            time_steps[i : i + (n + 1)], # arr_x
            data[i : i + (n + 1), :], # mat_y
            detrended_mat, # output
        )
        f2dfa_n[i] = np.power(detrended_mat, 2).mean(axis=0)
    for j, pair in enumerate(DCCA_of): # for each DCCA pair
        dcca_n[i, j] = (detrended_mat[:, pair[0]] * detrended_mat[:, pair[1]])
                           .mean(axis=0)
    F_DFA_arr[n_index, :] = np.sqrt(f2dfa_n.mean(axis=0))
    DCCA_arr[n_index, :] = dcca_n.mean(axis=0)
    # calculation of P_DCCA
    P_DCCA_output_function(n_index, DCCA_of, F_DFA_arr, DCCA_arr, # Inputs
                           P_DCCA_arr) # Output

```

The first **for** loop in the code operates over the values of the `tws` input, ensuring that step 6 of the *DFA* and *DCCA* methods, presented in Section 2, is being carried out. In other words, the calculations will be applied, sequentially, to every single value in the `tws` array. Three temporary arrays are allocated and resized for each time scale. The first, `f2dfa_n`, is used to store the calculations of the step 4 of the *DFA* ( $f_{DFA}^2$ ). The number of lines of this array correspond to the number of boxes in the current time scale ( $N - n$ , resized for each time scale  $n$ ) and the columns is the number of series in the analysis (same size for every value of  $n$ ). Second, the array `dcca_n` holds the values calculated in the step 4 of the *DCCA* (to calculate  $f_{DCCA}^2(n, i)$ ). The number of lines also correspond to the number of boxes(resized for each  $n$ ) but the number of columns equals the number of pairs (rows) in the `DCCA_of` input (same size in each  $n$ ). The `detrended_mat` array has a number of lines equal to the number of points in a time scale box ( $n + 1$ , resized for each  $n$ ) and column count also equal to the number of time series (same size in each  $n$ ). The first two temporary arrays will store data for all the boxes in the time scale, the last one will be used in each box and will have the values replaced in the next one, until the last box of the time scale. Than all the arrays will be cleared and recreated with shapes calculated with the new value of  $n$ .

After the allocation of the temporary arrays, the second **for** loop operates in every box for a certain time scale  $n$ . In each box, the `detrended_series()` function execute a one degree polynomial fit, subtract the values of the series with the value given by the interpolated curve(*DV*) and stores this values in the `detrended_mat`. After that, the **mean of the**

**square of each DV in the box** ( $f_{DFA}^2(n, i)$ ) is calculated in the current box for every time series and stored in the `f2dfa_n` array in the line associated with the current box, and the column related to each time series.

The next nested `for` loop operates over the `DCCA_of` array. For each line (pair) in the `DCCA_of` array, the  $f_{DCCA}^2(n, i)$  is evaluated by multiplying the corresponding *DV* from the two series boxes in the current pair, getting the mean, and stores it in the `dcca_n` temporary array.

After all the pairs are calculated, the algorithm goes back to the each box `for` loop and head to the next box for the current time scale. When all the all the boxes are calculated, the final results for the *DFA*, *DCCA* and  $\rho_{DCCA}$  are evaluated and saved in the output vectors `F_DFA_arr`, `DCCA_arr` and `P_DCCA_arr` respectively.

The function named `P_DCCA_output_function()` in the code above, is a pointer to other functions. One that outputs the  $\rho_{DCCA}$  results in the form of a table (rows for each time scale and columns for each `DCCA_of` pair) and the other outputs it the form of a 3D matrix, where each level is the matrix in Equation 3 for one of the time scales. This behavior is driven by the `P_DCCA_output_matrix` input (represented as the light green node in Figure 1), where `False` means table output and `True` matrix output. This is very convenient for calculating the  $DMC_x^2$ . There are two utility functions to transform a table output in a matrix one (`p_dcca_simple_to_matrix()`) and also the other way around (`p_dcca_matrix_to_simple()`).

### 3.2. Implementation of the Detrended Multiple Cross-correlation Coefficient function

The `dmcx2()` function runs the `p_dcca()` with `P_DCCA_output_matrix` set to `True` in the background. There is no `DCCA_of` input for the  $DMC_x^2$  function, instead there is a `dmcx2_of` parameter. This input receives a 2D matrix where each line represents the indexes of the series to be used in Equation 2. For each row, the first elements is the index of the series used as the dependent variable, the others, the index of the independent ones. There are two literal strings that can be used, for convenience as inputs for this parameter: '`all-full`', that generates a 2D array with every series as the dependent variable against all the others; and '`first-full`', with only one row, having the index zero series as the dependent variable in relation with all the others. The '`all-full`' option is conducted by calling `dmc_of_all_as_y()`, also available as a utility function.

With a given `dmcx2_of` an array with all the necessary pairs for the *DCCA* and  $\rho_{DCCA}$  is automatically generated and used for the background `p_dcca()` function to calculate da matrix as in Eq. 3. The `dcca_of_from_dmcx2_of()`, also can be used as an utility function, receives the `dmcx2_of` as input and returns the `dcca_of` array. With the matrix  $\rho_{DCCA}$  assembled, two internal functions, that can also be used as utility functions, calculates the  $DMC_x^2$  for all the lines in the `dmcx2_of` matrix, for all the time scales. The first function, `dmcx2_from_p_dcca_matrix()`, that receives the  $\rho_{DCCA}$  and the `dmcx2_of` array, is presented in the code below.

```
def dmcx2_from_p_dcca_matrix(P_DCCA_arr: NDArray[np.float64] ,
    dmcx2_of: NDArray[np.float64]) -> NDArray[np.float64]:
    # DMCx2 output matrix
    DMCx2_arr = np.full(shape=(P_DCCA_arr.shape[2], dmcx2_of.shape[0]),
```

```

fill_value=np.nan, dtype=P_DCCA_arr.dtype)

for n_index in range(P_DCCA_arr.shape[2]):
    P_DCCA_arr_2D = P_DCCA_arr[:, :, n_index]
    for j, dmcx2_of_1D in enumerate(dmcx2_of):
        DMCx2_arr[n_index, j] = dmcx2_from_p_dcca_matrix_2d(P_DCCA_arr_2D,
                                                               dmcx2_of_1D)

return DMCx2_arr

```

The `dmcx2_from_p_dcca_matrix()` function executes a for loop over the  $\rho_{DCCA}$  that separates the this 3D matrix in to the 2D matrices for each time scale. Nested in this loop, another `for` extracts each line of the `dmcx2_of` and passes, the extracted matrix and the line vector to the `dmcx2_from_p_dcca_matrix_2d()`, displayed here.

```

def dmcx2_from_p_dcca_matrix_2d(P_DCCA_arr_2D: NDArray[np.float64],
                                 dmcx2_of_1D: NDArray[np.float64]) -> NDArray[np.float64]:
    y_index = dmcx2_of_1D[0:1]
    x_indexes = dmcx2_of_1D[1:]

    mat_x = P_DCCA_arr_2D[np.ix_(x_indexes, x_indexes)]
    vec_y = P_DCCA_arr_2D[np.ix_(x_indexes, y_index)]

    return vec_y.T @ np.linalg.inv(mat_x) @ vec_y

```

The `dmcx2_from_p_dcca_matrix_2d()` function uses **Numpy** methods to prepare the data to apply Eq. 2. The list of indexes is divided in `y_index`, holding an one item array whit the index of the dependent variable time series, and `x_indexes`, containing the indexes of the independent ones. The **Numpy** `np.ix_()`, although it's not a very known function of the library constructs index arrays that will use the cross product from a series of 1D arrays as inputs. It's a very convenient way to extract a sub matrix and a vector from the  $\rho_{DCCA}$  matrix. The matrix, as assembled by the `p_dcca()` function, will always need to have extractions of a 2D matrix and a vector, as we can see in Eq. 2. The code below is an example of the extraction process.

```

import numpy as np
arr = np.array([[1,2,3,4],
                [2,1,5,6],
                [3,5,1,7],
                [4,6,7,1]])
print(arr)

[[1 2 3 4]
 [2 1 5 6]
 [3 5 1 7]
 [4 6 7 1]]

```

An illustrative `arr` matrix is defined as  $4 \times 4$  with all ones in the main diagonal and symmetric integer values in the other cells. Since the  $\rho_{DCCA}$  ranges from  $-1$  to  $1$ , those values should be interpreted as place holders.

```
sub_mat_index = np.ix_([1,2,3], [1,2,3])
print("index combination:\n", sub_mat_index)
print("extracted matrix:\n", arr[sub_mat_index])

index combination:
(array([[1,
       [2,
       [3]]), array([[1, 2, 3]]))
extracted matrix:
[[1 5 6]
 [5 1 7]
 [6 7 1]]
```

Above, a sub matrix, holding the positions 1 to 3 is extracted using `np.ix_()` function, and below, the extraction of the vector, first as a line and then as a column vector.

```
sub_mat_vec = np.ix_([0], [1,2,3])
print("line vector:\n", arr[sub_mat_vec])
sub_mat_vec = np.ix_([1,2,3], [0])
print("column vector:\n", arr[sub_mat_vec])

line vector:
[[2 3 4]]
column vector:
[[2]
 [3]
 [4]]
```

In the **Zebende** package, the vector is extracted as a column for better coherence with the  $DMC_x^2$  theory. The method also works for dependent variable different of index 0. The resulting matrix will preserve the diagonal as 1, the symmetry regarding the main diagonal and the order of elements in the column vector respected. the code below extract the index 1 series as the dependent and the others as independent.

```
sub_mat_index = np.ix_([0,2,3], [0,2,3])
print("index combination:\n", sub_mat_index)
print("extracted matrix:\n", arr[sub_mat_index])
sub_mat_vec = np.ix_([0,2,3], [1])
print("column vector:\n", arr[sub_mat_vec])

index combination:
(array([[0],
       [2],
```

```
[3]], array([[0, 2, 3]))
extracted matrix:
[[1 3 4]
 [3 1 7]
 [4 7 1]]
column vector:
[[2]
 [5]
 [6]]
```

The idea of separating the calculations of the  $\rho DCCA$  calculations in three distinct functions aims to different workflows. The `dmcx2()` function calculates and outputs all the prerequisites, as the  $DFA$ ,  $DCCA$ ,  $\rho_{DCCA}$  along with the  $DMC_x^2$ . This is the most practical way of getting all this calculations conducted. But the task can also be divided in two: first use the `p_dcca()` function to generate  $DFA$ ,  $DCCA$ ,  $\rho_{DCCA}$  outputs, analyze the outputs and then get the  $DMC_x^2$  using `dmcx2_from_p_dcca_matrix()`.

Function `dmcx2_from_p_dcca_matrix_2d()` can be used to more customizable applications. Imagine a use case where only the  $\rho_{DCCA}$  anti-correlation, inside a certain range, in relation to the dependent variable, with the `dmcx2_of` set to `all-full`. From previous  $\rho_{DCCA}$  studies is known that this coefficient can vary from positive to negative and vice versa in different time scales. This implies that the  $\rho_{DCCA}$  matrix should be analyzed in every time scale from every line of the `dmcx2_of`, also implies that the `dmcx2_of` may not be a matrix in the since that the rows may have different. The **Numpy** ND Array could not hold that. Many different workarounds could be proposed for that situation. In the implementation of this package, function `dmcx2_from_p_dcca_matrix_2d()` allow the user to make a custom code that extracts the  $\rho_{DCCA}$  matrix for the current time scale and extracts the elements that fit the rules from each `dmcx2_of` row.

### 3.3. Zig implementation

The Python implementation successfully reflects the implementation goals presented in Sec. 3 but the performance could benefit with the integration of a low-level language. Zig was chosen to enhance the algorithms performance. It is a low-level language that gain popularity in recent years and provides performance similar to C and Fortran in some scenarios (Kacs, Lee, Zarins, and Brown 2024).

The interest in using Zig language for this project also relies on the cross-compiling capabilities of the Zig compiler. For a small research group maintaining a package could be challenging and the ability to compile all releases for all platforms in a single machine is a great advantage. The Zig compiler can generate binaries for Windows, Linux and MacOS from a single machine. The Zig compiler is also very fast, and the language is very easy to learn, with a syntax that is very similar to C.

The implementation focus on writing technics  $\rho_{DCCA}$  function in Zig exposed as a C Application Binary Interface (ABI), called by Python using the `ctypes` package. The  $\rho_{DCCA}$  function is the most computational expensive part of the package, and the performance gain was expected to be significant.

The output arrays are allocated in the Python side and passed as pointers to the Zig function

together with the series matrix, `tws` and `DCCA_of` arrays. The Zig function receives the pointers and the size of the arrays as inputs, and the results are stored in the same memory space. Before passing to Zig must be assured that the arrays are contiguous, using the `np.ascontiguousarray()` function. The boolean parameter `P_dcca_matrix_output` is also passed to the Zig function, to determine if the output of the  $\rho_{DCCA}$  should be a table or a matrix.

The Zig implementation follows the same steps as the Python one, but respecting languages differences. In Python, using **Numpy**, all the time series calculations occur in the same line of code, using the **Numpy** broadcasting capabilities. In Zig, the calculations are made in a nested `for` loop.

### 3.4. Algorithm optimization

Although the Zig implementation presents a significant performance gain, the algorithm still can be optimized. The strategy is to focus on avoiding repeated calculations in the process. The code below shows the calculations of the polynomial fit before the optimization.

```
pub fn lin_ls_fit(win: []f64, time: []f64) [2]f64 {
    var x_sum: f64 = 0;
    var y_sum: f64 = 0;
    var xy_sum: f64 = 0;
    var x2_sum: f64 = 0;
    for (win, time) |w, t| {
        x_sum += t;
        y_sum += w;
        xy_sum += t * w;
        x2_sum += pow(f64, t, 2);
    }
    const n: f64 = @as(f64, @floatFromInt(time.len));
    //slope
    const slope: f64 = (((n * xy_sum) - (x_sum * y_sum)) /
        ((n * x2_sum) - (pow(f64, x_sum, 2))));
    //inter
    const inter: f64 = ((y_sum - (slope * x_sum)) /
        (n));
    //result
    return [_]f64{ slope, inter };
}
```

The code above calculates the slope and the intercept of a linear least squares fit. The function runs on every box for every time scale. In the optimized version, for consecutive boxes, the value of the sums from the previous box is used to calculate the next one without a loop for every item in the box. This technic proved to be very efficient in increasing performance. the expressions below show the effectiveness of the optimized algorithm.

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} T_k = \left( \sum_{j=i-1}^{(i+n)-1} T_j \right) - T_{i-1} + T_{i+n} \quad (5)$$

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} T_k^2 = \left( \sum_{j=i-1}^{(i+n)-1} T_j^2 \right) - T_{i-1}^2 + T_{i+n}^2 \quad (6)$$

$$\forall 1 < i \leq (N - n), \sum_{k=i}^{i+n} S_k = \left( \sum_{j=i-1}^{(i+n)-1} S_j \right) - S_{i-1} + S_{i+n} \quad (7)$$

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

The expressions (Eq. 5, Eq. 6, Eq. 7 and Eq. 8) showcases how the optimizations can be implemented in the code. Each expression display that the sum of the values in a box can be calculated using the sum of the previous box and the values of the first and last elements of the current box. Eq.5 presents the sum of the time stamps values and Eq.6 the sum of the squares of the time stamps values. This calculations occur once for every box in every time scale and is stored in variables. Eq.7 and Eq.8 presents the same calculations for the series values. Storing values for each time series.

Also, to optimize the calculations from one time scale to the next, the first sums for the first box are saved in a temporary variable. Considering the current time scale as `tws_prev` and the consecutive as `tws_current`. To calculate the current value, the algorithm takes the sums from the `tws_prev` and add the values with indexes that exceed the `tws_prev` size. The code was also rewritten with structs, to hold the temporary values and the functions for better readability. The code below shows the optimized version of polynomial fitting calculations.

```
fn shiftWindow( self: *MainOperator,
    n: usize, win_start: usize,
    F_DFA_ptr: *allowzero [*c]f64) void {
    self.time_window = self.time[win_start ..][ .. (n + 1)];
    // print("win_start {}\\n", .{win_start});
    if (win_start != 0) {
        // updating sum_x
        self.current.sum_x = self.current.sum_x - self.left_x + self.time_window[n];
        // updating sum x^2
        self.current.sum_x2 = self.current.sum_x2 -
            pow(f64, self.left_x, 2) + pow(f64, self.time_window[n], 2);
        // updating y and y*x for every serie
        for (self.series, 0..) |serie, sr_index| {
            serie.current.sum_y = serie.current.sum_y -
                serie.left_y + serie.serie[win_start + n];

            serie.current.sum_xy = serie.current.sum_xy -
                (self.left_x * serie.left_y) +
                (self.time_window[n] * serie.serie[win_start + n]);
        }
    }
}
```

```

        serie.left_y = serie.serie[win_start];
        self.detrended(serie, win_start, &F_DFA_ptr.*[sr_index]);
    }
} else { // win_start == 0
    self.current.window_len = n + 1;

    for (self.previous.window_len..self.current.window_len) |i| {
        self.previous.sum_x += self.time_window[i];
        self.previous.sum_x2 += pow(f64, self.time_window[i], 2);

        for (self.series) |serie| {
            serie.previous.sum_y += serie.serie[i];
            serie.previous.sum_xy += self.time_window[i] * serie.serie[i];
        }
    }

    // updating current sum values
    self.current.sum_x = self.previous.sum_x;
    self.current.sum_x2 = self.previous.sum_x2;

    for (self.series, 0..) |serie, sr_index| {
        serie.current.sum_xy = serie.previous.sum_xy;
        serie.current.sum_y = serie.previous.sum_y;

        serie.left_y = serie.serie[win_start];
        self.detrended(serie, win_start, &F_DFA_ptr.*[sr_index]);
    }
}
self.left_x = self.time_window[0];
}

```

In the code above, the **MainOperator** is a structure that holds the time series( $Y_n$ ), the time stamps(x), the sums for the time window and for each series and the temporary variables that holds the first values of the previous box. The **shiftWindow()** function is a method of the **MainOperator** structure. The function receives the size of the box, the index of the first element of the box, and a pointer to the  $F_{DFA}^2$  array. The function is called for every box in every time scale. The function is responsible for updating the sums of the time series, the sums of the  $DV$  and the  $DV$  matrix.

The if statement in the code above is responsible for updating the sums for the current box. The first **if** block is executed for every box except the first one in each time scale, where the **else** block is executed instead.

The **if** block first update the values for the sum of the time stamps (**sum\_x**) and the sum of the squares of the time stamps (**sum\_x2**). Then, for every series, the sum of the series values (**sum\_y**) and the sum of the product of the series and the time stamps are updated (**sum\_xy**). Than the values are passed to the **detrended** function that calculates The  $DV$  and the  $DFA$  for the current box.

The `else` block is executed for the first box in each time scale. For consecutive time scales, the values of the sums are updated using the values of the previous box, looping and adding only the values in the current box that don't overlap with the previous one. The `detrended` function is also called for every series in the time scale.

The three implementations, pure Python, Zlg and optimized Zig, were tested and the results are presented in the next section.

## 4. Results

The algorithm was tested with data and results from the [Ribeiro \*et al.\* \(2025\)](#) paper. Details regarding data preparations can be found in the aforementioned paper. A comprehensive dataset of EEG recording, with 108 subjects, executing 12 experiments each, with the  $\rho_{DCCA}$  and  $DMC_x^2$  calculated for all experiments, using signals from four electrodes with 15742 points. The results were compared with the original ones, and the algorithms were validated. The original data came from the [Schalk, McFarland, Hinterberger, Birbaumer, and Wolpaw \(2004\)](#) paper, available on the *Physionet* open-access website, that can be accessed in this link:

<https://physionet.org/content/eegmmidb/1.0.0/>

Regarding performance, the pure Python implementation, the Zig implementation and the optimized Zig were compared with 3 other packages that calculate the  $\rho_{DCCA}$ . Researching the R CRAN repository, the **DFA** ([Mesquita, Filho, and Danilevicz. 2024](#)), a pure R implementation, and the **DCCA** ([Prass and Pumi 2020](#)), a R package implemented in Java where found. From the Python pypi repository, a package also named **DCCA** ([Alchieri, Abdalazim, Alecci, Gashi, Gjoreski, and Santini 2024](#)), a Python implementation with massive use of **Numpy** broadcasting capabilities and memory access with strides to avoid loops, was also used. All those packages have  $\rho_{DCCA}$  functions. Hereafter the **DCCA** R-java package will be referred as **DCCA-R-Java** and the **DCCA** Python package as **DCCA-Python** for better distinction.

The  $dmc_x^2$  is a more recent coefficient, and no other package was found that calculates it besides the **Zebende** package. That alone is a huge contribution of this package. The  $\rho_{DCCA}$  is a fundamental step to calculate the  $DMC_x^2$  and contains the most computational expensive parts. In the  $DMC_x^2$  calculation, The  $\rho_{DCCA}$  is calculated for, at least, a combination of 3 series two by two. With the idea to broadly compare the performance of the **Zebende** package with the others, aiming an overall order of magnitude of the packages compared speed, using the same computer and the same time window scales, the performance comparison was made using data from one of the experiments described in the [Schalk \*et al.\* \(2004\)](#) paper, with 19920 lines and 64 channels. For convenience the data was posted in the link below:

<https://tinyurl.com/zbttestdata01>

None of the comparison packages can calculate multiple series comparison in the same input, so the  $\rho_{DCCA}$  methods were wrapped around a loop that alternate the series used by the function, using a four two by two combination of the data indexes. The **DCCA-Python**

package also didn't accept multiple inputs for the time scale, so another loop was added in that test.

First all the packages were used to calculate the  $\rho_{DCCA}$  of combinations of four time series. The results are presented in the Table 1.

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%

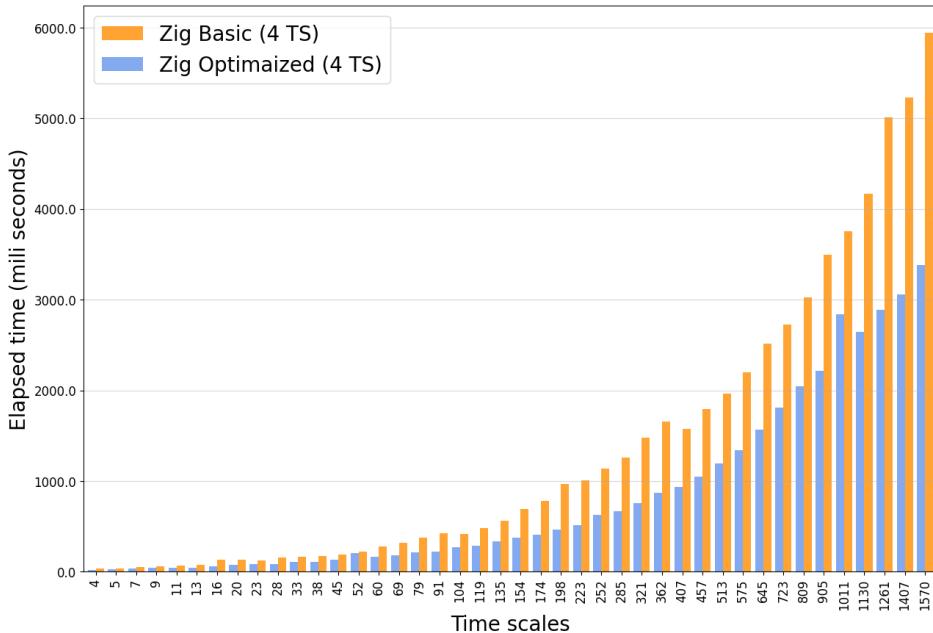


Figure 3: Elapsed Time Comparison for Different Implementations (four time series)

The **Zebende** package, using the optimized Zig implementation, was the fastest, with a 27.52 seconds elapsed time. The basic Zig implementation was the second fastest, with 42.03 seconds. The **DCCA-Python** package was the third fastest, with 370.54 seconds. The **DCCA-R-Java** package was the fourth fastest, with 1987.67 seconds. The **Zebende** package, using the Python implementation, was the fifth fastest, with 2415.94 seconds. The **DFA-R** package was the slowest, with 20173.13 seconds.

The Zig basic and optimized versions occupying the first two positions in the ranking is not unexpected, hence the Zig language is more performative than the other ones used in the comparison. The **DCCA-Python** package in the third position is a surprise, showing that a well implemented Python package can overperform a Java implementation.

The 66.50% increase in performance in the basic compared to the optimized Zig implemen-

tation is a good result, showcasing the effectiveness of the optimizations. Another test was conducted, with the same data and settings as the previous one, but getting the elapsed time inside every time scale. These results are presented in Fig. 3.

Figure 3 shows the elapsed computation time for the Basic and the optimized version of the algorithm. The differences between the two implementations increases as the time scale increases. If the test was conducted with a bigger time series, and with larger time scales the difference would be even more significant.

The **Zebende** package implementations were designed to increase performance in scenarios with large number of points in the series, but also with a large amount of time series used in the  $DMC_x^2$  calculations. To prove that, the same tests were also made using the same data, with the combinations of twelve series, 2 by 2. The results of that second test are presented in Table 2.

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%

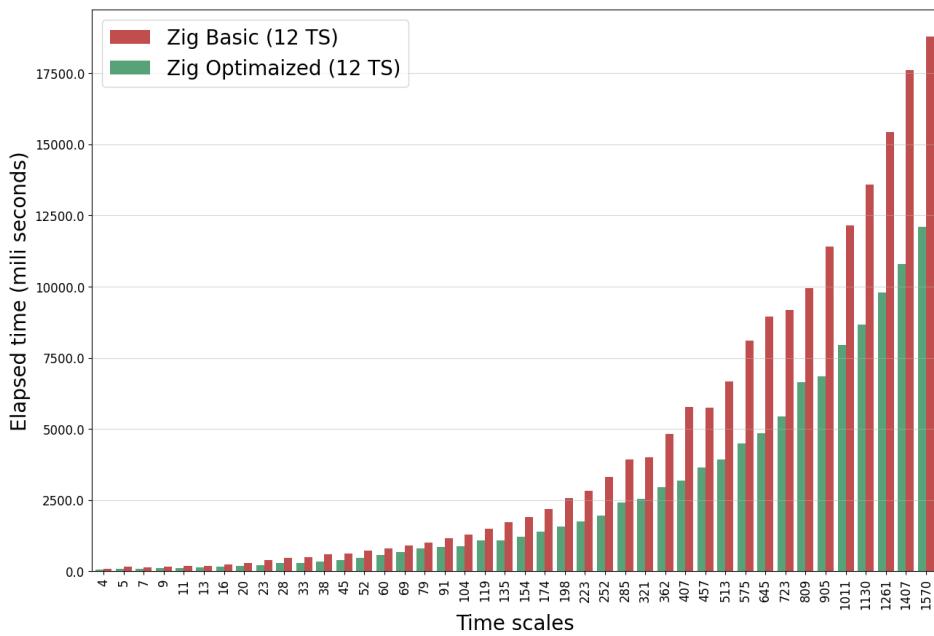


Figure 4: Elapsed Time Comparison for Different Implementations (twelve time series)

With twelve time series, the optimized Zig version still the fastest, followed by the basic Zig implementation. The **Zebende** pure Python implementation ended in the third position, performing better than the **DCCA-Python** (4th) and the **DCCA-R-Java** (5th) packages. The

lists ended with the **DFA** in the 6th position.

Comparing the basic with the optimized Zig algorithms using twelve time series the performance increase is 76.64%. Figure 4 shows the comparison between the basic and optimized Zig implementations with twelve time series. The results are equivalent to the ones obtained with four time series, reinforcing the capabilities of the optimization techniques.

The basic version is available in **Zebende** version 0.1.91 and the optimized implementation in version 0.2.0 of the package. The codes used for testing are available, as jupyter notebooks, in the link below, where they can be tested in Google colab environment:

[https://255ribeiro.github.io/paper\\_zebendelib/notebooks/test\\_nb.html](https://255ribeiro.github.io/paper_zebendelib/notebooks/test_nb.html)

## 5. Summary and discussion

The **Zebende** package is the first tool that calculates the  $DMC_x^2$  coefficient for any number of time series. The package also calculates the  $DFA$ ,  $DCCA$  and the  $\rho_{DCCA}$ , a fundamental step for the  $DMC_x^2$  calculation. The package was tested with data from the Ribeiro *et al.* (2025) paper, and the results were validated. The package was also compared with other packages that calculate the  $\rho_{DCCA}$ , and the results showed that the **Zebende** package is the fastest, making it suitable for large datasets.

The tests with twelve time scales also shows that the **Zebene** implementation, designed to obtain better performance working with greater number of time series, deliver great results even with the pure Python non-optimized version. This package enable the exploration of the  $DMC_x^2$  coefficient in scenarios where it was never tested before. The package is available at the **pypi**.

## References

- Alchieri L, Abdalazim N, Alecci L, Gashi S, Gjoreski M, Santini S (2024). “Lateralization Effects in Electrodermal Activity Data Collected Using Wearable Devices.” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **8**(1), 1–30.
- Brito AA, Araujo HA, Zebende GF (2019). “Detrended Multiple Cross-Correlation Coefficient applied to solar radiation, air temperature and relative humidity.” *Sci. Rep.*, **9**, 19764. doi: [10.1038/s41598-019-56114-6](https://doi.org/10.1038/s41598-019-56114-6).
- da Silva Filho AM, Zebende GF, de Castro APN, Guedes EF (2021). “Statistical test for multiple detrended cross-correlation coefficient.” *Physica A: Statistical Mechanics and its Applications*, **562**, 125285.
- Ferreira P, Dionísio A, Guedes EF, Zebende GF (2018). “A sliding windows approach to analyse the evolution of bank shares in the European Union.” *Physica A: Statistical Mechanics and its Applications*, **490**, 1355–1367. ISSN 03784371. doi:[10.1016/j.physa.2017.08.095](https://doi.org/10.1016/j.physa.2017.08.095).

- Filho FO, Ribeiro F, Cruz JL, De Castro AN, Zebende G (2023). “Statistical Study of the EEG in Motor Tasks (Real and Imaginary).” **622**, 128802. ISSN 03784371. doi: [10.1016/j.physa.2023.128802](https://doi.org/10.1016/j.physa.2023.128802). URL <https://linkinghub.elsevier.com/retrieve/pii/S0378437123003576>.
- Guedes E, Dionísio A, Ferreira PJ, Zebende GF (2017). “DCCA Cross-Correlation in Blue-Chips Companies: A View of the 2008 Financial Crisis in the Eurozone.” **479**, 38–47. ISSN 03784371. doi: [10.1016/j.physa.2017.02.065](https://doi.org/10.1016/j.physa.2017.02.065). URL <http://dx.doi.org/10.1016/j.physa.2017.02.065>.
- Guedes EF, Brito AA, Oliveira Filho FM, Fernandez BF, de Castro AP, da Silva Filho AM, Zebende GF (2018a). “Statistical test for  $\Delta\rho$ DCCA cross-correlation coefficient.” *Physica A: Statistical Mechanics and its Applications*, **501**, 134–140. ISSN 03784371. doi: [10.1016/j.physa.2018.02.148](https://doi.org/10.1016/j.physa.2018.02.148). URL <https://doi.org/10.1016/j.physa.2018.02.148>.
- Guedes EF, Brito AA, Oliveira Filho FM, Fernandez BF, de Castro AP, da Silva Filho AM, Zebende GF (2018b). “Statistical test for  $\Delta\rho$ DCCA: Methods and data.” *Data in Brief*, **18**, 795–798. ISSN 23523409. doi: [10.1016/j.dib.2018.03.080](https://doi.org/10.1016/j.dib.2018.03.080).
- Guedes EF, da Silva Filho AM, Zebende GF (2021). “Detrended multiple cross-correlation coefficient with sliding windows approach.” *Physica A: Statistical Mechanics and its Applications*, **574**, 125990.
- Guedes EF, de Castro APN, da Silva Filho AM, Zebende GF (2022). “ $\Delta$  DMC x 2: A New Approach to Measure Contagion Effect on Financial Crisis.” *Fluctuation and Noise Letters*, **21**(04), 2250026.
- Kacs D, Lee J, Zarins J, Brown N (2024). “Pragma driven shared memory parallelism in Zig by supporting OpenMP loop directives.” In *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 930–938. doi: [10.1109/SCW63240.2024.00132](https://doi.org/10.1109/SCW63240.2024.00132).
- Mesquita VB, Filho FMO, Danilevitz IM (2024). *DFA: Detrended Fluctuation Analysis*. R package version 1.0.0, URL <https://CRAN.R-project.org/package=DFA>.
- Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994). “Mosaic Organization of DNA Nucleotides.” **49**(2), 1685–1689.
- Podobnik B, Jiang ZQ, Zhou WX, Stanley HE (2011). “Statistical tests for power-law cross-correlated processes.” *Phys. Rev. E*, **84**(6), 66118. doi: [10.1103/PhysRevE.84.066118](https://doi.org/10.1103/PhysRevE.84.066118). URL <https://link.aps.org/doi/10.1103/PhysRevE.84.066118>.
- Podobnik B, Stanley HE (2008). “Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series.” *Physical Review Letters*, **100**(8). ISSN 00319007. doi: [10.1103/PhysRevLett.100.084102](https://doi.org/10.1103/PhysRevLett.100.084102). 0709.0281.
- Prass TS, Pumi G (2020). *DCCA: Detrended Fluctuation and Detrended Cross-Correlation Analysis*. R package version 0.1.1, URL <https://CRAN.R-project.org/package=DCCA>.
- Ribeiro FF, de Almeida Brito A, Filho FMO, Cruz JAL, Zebende GF (2025). “DCCA multi cross-correlation analysis applied on EEG signals to study motor activity

- (Real/Imaginary).” *Biomedical Signal Processing and Control*, **103**, 107419. ISSN 1746-8094. [doi:  
https://doi.org/10.1016/j.bspc.2024.107419](https://doi.org/10.1016/j.bspc.2024.107419). URL [https://www.  
sciencedirect.com/science/article/pii/S1746809424014770](https://www.sciencedirect.com/science/article/pii/S1746809424014770).
- Schalk G, McFarland D, Hinterberger T, Birbaumer N, Wolpaw J (2004). “BCI2000: A General-Purpose Brain-Computer Interface (BCI) System.” *IEEE Transactions on Biomedical Engineering*, **51**(6), 1034–1043. [doi:  
10.1109/TBME.2004.827072](https://doi.org/10.1109/TBME.2004.827072).
- Vassoler RT, Zebende GF (2012). “DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity.” *Physica A*, **391**(7), 2438–2443. ISSN 0378-4371. [doi:  
10.1016/j.physa.2011.12.015](https://doi.org/10.1016/j.physa.2011.12.015). URL [http://dx.doi.org/10.1016/j.  
physa.2011.12.015](http://dx.doi.org/10.1016/j.physa.2011.12.015).
- Wang GJ, Xie C, Chen S, Yang JJ, Yang MY (2013). “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*, **392**(17), 3715–3730. ISSN 03784371. [doi:  
10.1016/j.physa.2013.04.027](https://doi.org/10.1016/j.physa.2013.04.027).
- Zebende GF (2011). “DCCA cross-correlation coefficient: Quantifying level of cross-correlation.” *Physica A: Statistical Mechanics and its Applications*, **390**(4), 614–618. ISSN 03784371. [doi:  
10.1016/j.physa.2010.10.022](https://doi.org/10.1016/j.physa.2010.10.022). URL [http://dx.doi.org/10.1016/j.  
physa.2010.10.022](http://dx.doi.org/10.1016/j.physa.2010.10.022).
- Zebende GF, Aguiar LC, Ferreira P, Guedes EF (2022). “Statistical Approach to Study the Relationship Between Stock Market Indexes by Multiple DCCA Cross-Correlation Coefficient.” *Fluctuation and Noise Letters*, **21**(5), 2250045–1314.
- Zebende GF, Silva AM (2018). “Detrended Multiple Cross-Correlation Coefficient.” *Physica A*, **510**, 91–97. ISSN 0378-4371. [doi:  
10.1016/j.physa.2018.06.119](https://doi.org/10.1016/j.physa.2018.06.119).

### Affiliation:

Fernando Ferraz Ribeiro

Faculty of Architecture

Federal University of Bahia

Caetano Moura street, 121,

40210-905, Salvador-BA, Brazil

E-mail: [fernando.ribeiro@ufba.br](mailto:fernando.ribeiro@ufba.br)

URL: <https://arquitetura.ufba.br/>

and

Graduate Program in Modeling in Earth and Environmental Sciences

State University of Feira de Santana

Av. Transnordestina

44036-900, Feira de Santana-BA, Brazil

URL: <https://ppgm.uefs.br/>

and

Senai-Cimatec University

Av. Orlando Gomes, 1845,

41650-010, Salvador-BA, Brazil  
URL: <https://www.senaicimatec.com.br/>

Gilney Figueira Zebende  
Graduate Program in Modeling in Earth and Environmental Sciences  
State University of Feira de Santana  
Av. Transnordestina  
44036-900, Feira de Santana-BA, Brazil  
E-mail: [gfzebende@uefs.br](mailto:gfzebende@uefs.br)  
URL: <https://ppgm.uefs.br/>

### 3.3 Maximização do coeficiente $DMC_x^2$ utilizando matriz $\rho_{DCCA}$ e $DPDCCA$

“Noel Meyerhof consulted the list he had prepared and chose which item was to be first. As usual, he relied mainly on intuition.”

(Isaac Asimov)

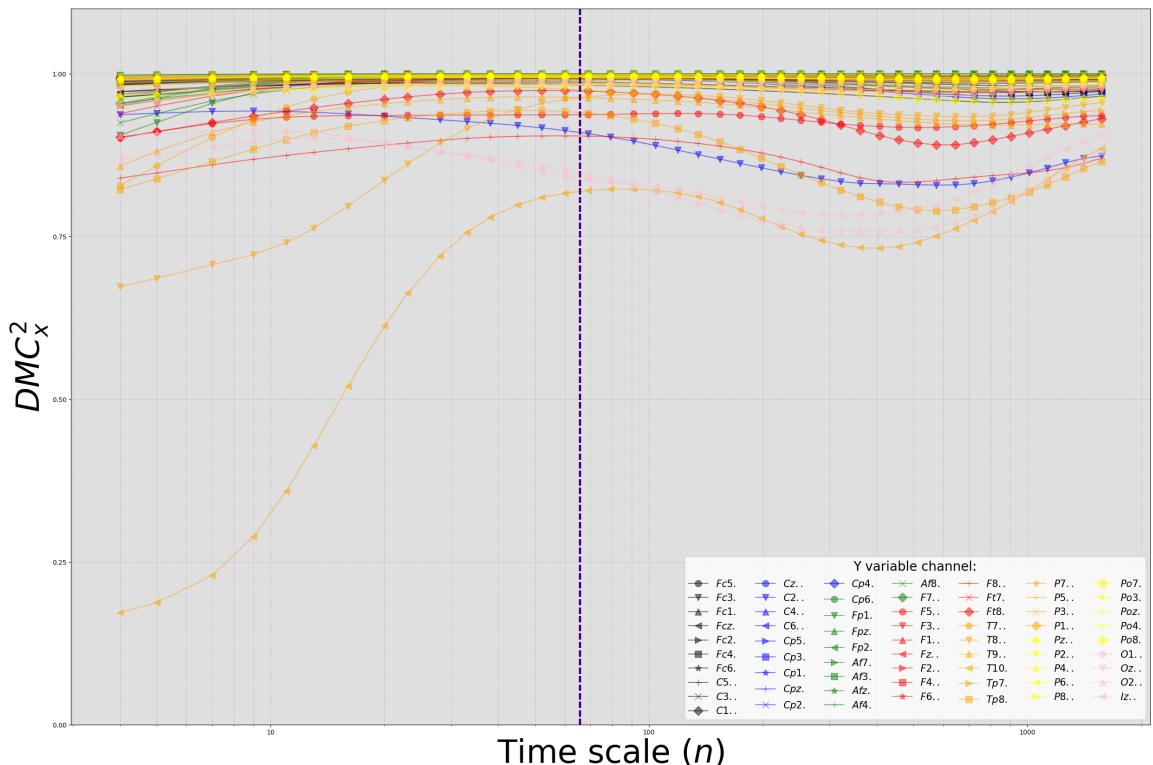


Figura 3.4:  $DMC_x^2$  de todo os canais do experimento[1:63] para cada canal como variável dependente.

Fonte: Elaborada pelos autores

Com a ferramenta desenvolvida e testada, volta-se para os dados utilizados nos artigos do Capítulo 3 Sessão 3.1, e no Anexo A. A matriz do  $\rho_{DCCA}$  para todos os 64 canais é calculada em menos de 17min. Isso se deve ao Algoritmo *Detrended Saved*. Sem a aplicação deste, calculando o número de combinações necessárias para a montagem da matriz (Equação 3.6), levando em conta que, para cada  $DCCA$  aplica-se duas vezes o cálculo do  $DV$ , chega-se ao valor de  $2016 \times 2 = 4032$ . Com o uso do algoritmo, apenas 64 valores são calculados.

Com a matriz montada, o cálculo do  $DMC_x^2$ , envolvendo todas as 64 séries, com cada uma delas como variável dependente, para os 42 valores de  $n$ , utilizados nos referidos artigos, é realizado em 0.7s.

A Figura 3.4 apresenta esses resultados. Apontando para um sistema altamente correlacionado. Também encontramos um pico de correlação entre as escalas  $n = 60$  e  $n = 69$ , confirmando o encontrado nos artigos.

A inversa de cada uma das 42 matrizes do  $\rho_{DCCA}$  também foi calculada em menos de 1s. O  $DPDCCA$  foi implementado e calculado para todas as combinações de canais, para cada escala temporal, levando também menos de 1s na execução.

É evidente o aumento na quantidade de informações que se pode obter e tratar com as ferramentas computacionais desenvolvidas neste trabalho. É necessário desenvolver instrumentos para acessar essas informações.

A Figura 3.8 apresenta as matrizes de correlação do  $\rho_{DCCA}$  para todos os canais, nas escalas temporais  $n = 4$  e  $n = 69$ . A aplicação do coeficiente descreve um sistema altamente correlacionado positivamente, mesmo na escala  $n = 4$ , mas sendo ainda maior na escala  $n = 69$ .

Figura 3.5: Matriz de correlação do  $\rho_{DCCA}$

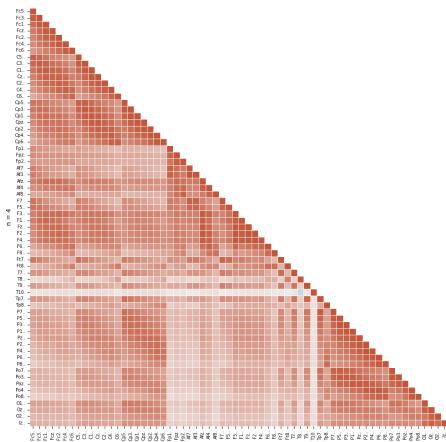


Figura 3.6:  $\rho_{DCCA} n = 4$

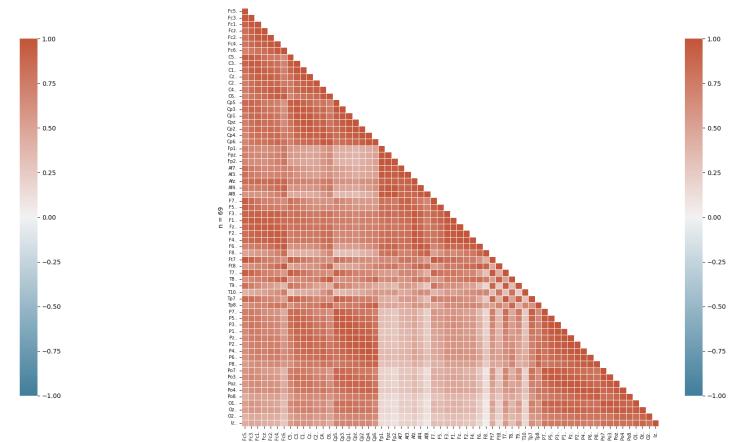
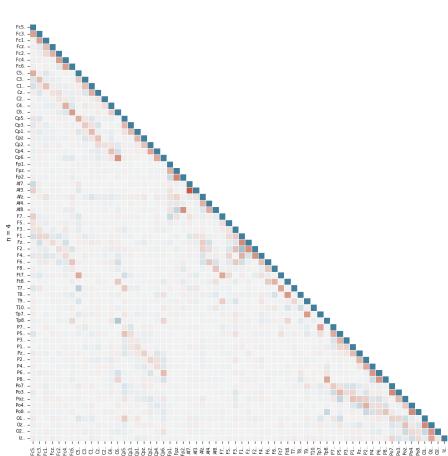
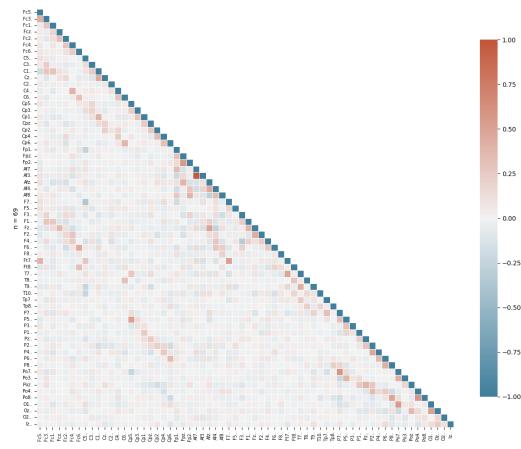


Figura 3.7:  $\rho_{DCCA} n = 69$

Figura 3.8: Matriz de correlação do  $\rho_{DCCA}$ .

Fonte: Elaborada pelos autores

A Figura 3.11, retrata as matrizes de correlação do  $DPDCCA$  para as mesmas escalas temporais. A proposta do  $DPDCCA$  é mostrar o quando duas variáveis se relacionam, levando em conta as outras variáveis do sistema, acaba mostrando uma matriz com a maioria dos valores próximos a zero, com algumas combinações de variáveis se destacando em relação ao sistema. Esses resultados devem ser melhor estudados em trabalhos futuros.

Figura 3.9: *DPDCCA n = 4*

- Aleatório: como um artifício de controle.
- $\rho_{DCCA}$ : Os oito maiores valores do  $\rho_{DCCA}$  de  $T8$  em relação à cada uma das outras variáveis.
- $|\rho_{DCCA}|$ : Os oito maiores valores absolutos do  $\rho_{DCCA}$  de  $T8$  em relação à cada uma das outras variáveis.
- $DPDCCA$ : Os oito maiores valores do  $DPDCCA$  de  $T8$  em relação à cada uma das outras variáveis.
- $\Sigma DPDCCA$ : Um algoritmo baseado nas características do  $DPDCCA$ .

---

**Algoritmo 4**  $\Sigma DPDCCA$ 


---

```

def maximize_dmc_dp(n, index, count, c_mat):
    dmc_of = [index]
    for i in range(count):
        comp_array = np.full(64, fill_value = np.nan, dtype = float)
        for j in range(64):
            if j in dmc_of:
                pass
            else:
                for k in dmc_of:
                    temp = pddcca(j, k, n, c_mat)
                    if np.isnan(comp_array[j]):
                        comp_array[j] = temp
                    else:
                        comp_array[j] += temp
        dmc_of.append(int(np.nanargmax(comp_array)))
    return np.array(dmc_of)

```

---

O critério de seleção  $\Sigma DPDCCA$  aparece está apresentado, em código, no Algoritmo 4. O  $DPDCCA$  tem características particulares. Analisando a Equação 2.15 ve-se que, entre duas séries idênticas, obtém-se o valor mais baixo possível  $-1$ , indicando que a informação daquela série não agrega valor ao todo.

Pelo critério do  $\Sigma DPDCCA$ , a escolha dos 8 canais acontece de forma sucessiva. O primeiro canal é escolhido pelo maior valor do  $DPDCCA$  em relação ao canal  $T8$ . Para os próximos 7 canais, calcula-se o maior valor do  $DPDCCA$  entre o canal candidato e o canal  $T8$  somado com os valores de  $DPDCCA$  entre o canal candidato e os canais escolhidos nas etapas anteriores.

### 3.3.2 Resultados

A Tabela 3.1 apresenta os resultados dos cinco métodos para  $n = 4$ . O valor de referência é o valor do  $DMC_x^2$  de  $T8$  com todos os outros canais. Os métodos aparecem na primeira coluna, Os canais selecionados na segunda, o valor do  $DMC_x^2$  de  $T8$  em relação aos canais selecionados na terceira e o percentual do valor obtido em relação ao valor de referência na última coluna.

Tabela 3.1: Maximização do  $DMC_x^2$ .  $n = 4$ ,  $count = 8$ , referência= 0.6726

Critério	canais selecionados	valor	percentual
$\rho_{DCCA}$	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%

Os valores do  $\rho_{DCCA}$ , do  $|\rho_{DCCA}|$  e do  $\Sigma DPDCCA$  aproximam em torno de 94.1%, com o  $\rho_{DCCA}$ , e o  $|\rho_{DCCA}|$  escolhendo os mesmos canais e performando levemente acima do  $\Sigma DPDCCA$ , que escolheu os canais  $F2$  e  $F8$  em vez dos canais  $P4$  e  $P6$ . O desempenho do critério  $DPDCCA$  fica abaixo dos já citados.

Tabela 3.2: Maximização do  $DMC_x^2$ .  $n = 69$ ,  $count = 8$ , referência= 0.9643

Critério	canais selecionados	valor	percentual
$\rho_{DCCA}$	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%

Na tabela 3.2 ve-se resultados semelhantes. Nesta escala temporal com maior multi-correlação, aparece o  $\Sigma DPDCCA$  um pouco acima dos outros e com o critério  $DPDCCA$  mais próximos dos demais. Diferenças entre os canais escolhidos em uma escala temporal para a outra também foram registrados.

A semelhança entre os critérios  $\rho_{DCCA}$  e  $|\rho_{DCCA}|$  se deve à característica do sistema de além de ser fortemente multi-correlacionado, também é positivamente multi-correlacionado. Desconsiderando o  $\rho_{DCCA}$  de um canal com ele mesmo, os valores de  $\rho_{DCCA}$  variam de 0.9995 até o mínimo de -0.2281. Não existem anti-correlações significativas em nenhuma das escalas temporais calculadas.

A Figura 3.14 apresenta o gráfico do valor do  $DMC_x^2$ , tendo o canal  $T8$  como variável dependente e utilizando o critério  $\rho_{DCCA}$  para ordenar os resultados. O eixo  $x$  apresenta os valores do  $DMC_x^2$  e o eixo  $y$ , a quantidade de canais utilizados para calcular o  $DMC_x^2$  em cada ponto dividido pelo total de canais. Quando o eixo  $y$  tem valor igual a 1 significa

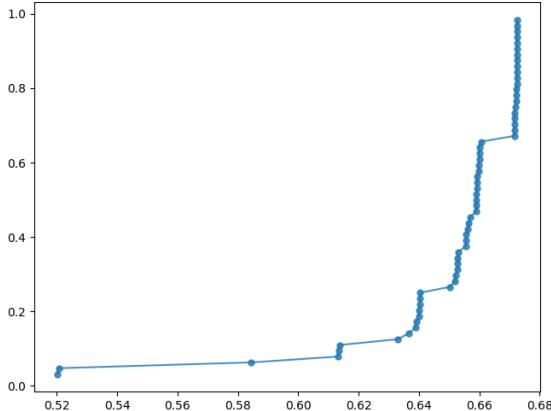
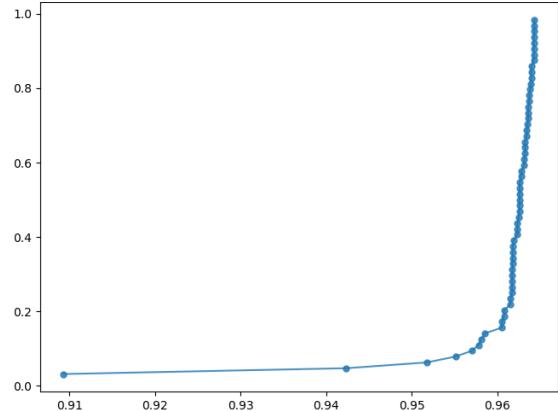
Figura 3.12: Channel: T8  $n = 4$ Figura 3.13: Channel: Tf8  $n = 69$ 

Figura 3.14: Gráfico cumulativo do  $DMC_x^2$  - Channel: T8  $n = 69$ .

Fonte: Elaborada pelos autores.

que todos os 63 canais disponíveis foram utilizados como variáveis independentes. A Figura 3.17 aplica o mesmo procedimento utilizado na Figura 3.14, agora tendo o canal  $Af3$  como variável dependente. É importante notar que, para melhor visualização, os eixos  $x$  e  $y$  estão em escalas diferentes.

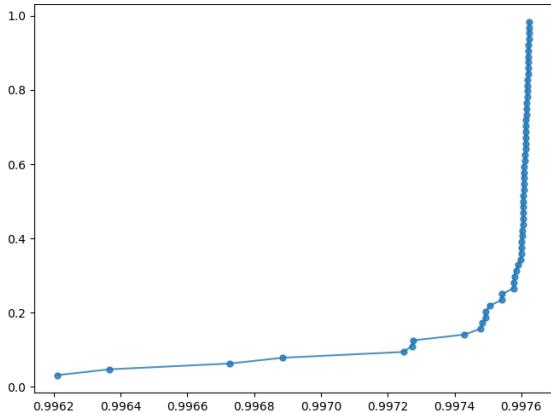
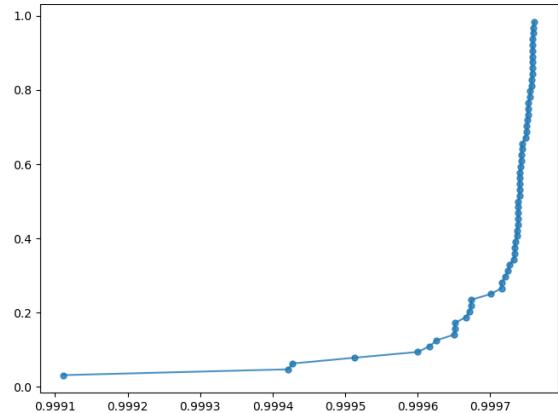
Figura 3.15: Channel: Af3  $n = 4$ Figura 3.16: Channel: Af3  $n = 69$ 

Figura 3.17: Gráfico cumulativo do  $DMC_x^2$  - Channel: Af3  $n = 69$ .

Fonte: Elaborada pelos autores.

Na Figura 3.14, para  $n = 4$ , o valor do  $DMC_x^2$  varia entre 0.52 e 0.68 aproximadamente, para a escala temporal  $n = 69$ , os valores variam apenas na segunda casa decimal. Para o canal  $Af3$ , escolhido por ser o canal de maior  $DMC_x^2$  em relação aos outros 63 canais, apresenta variação na terceira casa decimal para a escala  $n = 4$  e na quarta casa para a escala  $n = 69$ . Para os dados analisados, quanto maior a correlação do canal com o sistema, menor a variação cumulativa do coeficiente  $DMC_x^2$ . Este resultado precisa ser analisado com mais dados e outros conjuntos de variáveis, incluindo conjuntos de séries com características de multi-correlação diferentes.

### 3.3.3 Conclusão

Os critérios  $|\rho_{DCCA}|$  e  $\Sigma DPDCCA$  apresentaram resultados interessantes na maximização do coeficiente  $DMC_x^2$  em relação ao total. É recomendado a repetição do experimento em mais escalas temporais e com outros canais como variável dependente, comparando com o valor máximo do  $DMC_x^2$  calculado para cada combinação de resultado possível.

Os critérios de aproximação também podem ser entendidos como critérios de semelhança entre séries. Se bem entendidos e validados por extensões desta pesquisa.

A ideia de maximizar o coeficiente apresenta similaridades com o processo de seleção de atributos em um algoritmo de aprendizado de máquina. Os resultados preliminares podem ser entendidos como um indício de que o desempenho destes critérios para a seleção de atributos devem ser verificados.

---

## Conclusão

---

“I live on Earth at present, and I don’t know what I am.  
I know that I am not a category. I am not a thing – a noun.  
I seem to be a verb, an evolutionary process (...)"

(R. Buckminster Fuller)

A pesquisa apresenta resultados sólidos e formulações de novas perguntas. Valores de funções e coeficientes foram calculados e testados. Resultados de uma pesquisa, realizada com dados de EEG foram validados pelos pares em publicação.

Uma ferramenta computacional para proporcionar facilidade e velocidade nos cálculos do  $DMC_x^2$  foi implementada e trouxe consigo um novo algoritmo, que apresenta vantagens de desempenho em certos casos de uso.

A mudança na quantidade de informação que pode ser tratada, utilizando a ferramenta, foi ilustrada e critérios para definição de semelhança entre séries temporais foram estabelecidos e avaliados.

Entende-se que, considerando o exposto na Sessão 2.5 do Capítulo 2, pela grande aplicabilidade do método em estudos da área de Ciências Ambientais, os avanços aqui apresentados contribuem com a área.

### 4.1 Validação das hipóteses

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

A hipótese 1 da Tese (É possível otimizar os cálculos do  $\rho_{DCCA}$  e do  $DMC_x^2$ .) foi validada pela implementação do Algoritmo 3.

A hipótese 2 (Com maior poder de cálculo o a aplicação dos coeficientes seria potencializada.) foi validada com o trabalho apresentado na Sessão 3.3 do Capítulo 3.

A hipótese 3 (Uma ferramenta adequada para o cálculo dos coeficientes seria um impulso na utilização e divulgação destes.) só será validada ou descartada com o tempo.

Com duas hipóteses validadas e uma em suspenso, considera-se a validação parcial das intenções da Tese.

## 4.2 Trabalhos futuros

O pacote *Zebende* continua sua evolução, mudanças na usabilidade de certas funções serão implementadas, assim como novos algoritmos devem fazer parte da biblioteca. Tendo os algoritmos destacados nas Seções 2.3 e 2.4 como os primeiros a serem incorporados à ferramenta.

Também é necessário investigar a possibilidade de implementar paralelismo no cálculo das funções e coeficientes, tanto em *CPU* quanto em *GPU*.

A possibilidade de, através da ferramenta computacional criada, explorar critérios de proximidade entre séries temporais e a sua aplicabilidade como seleção de atributos para algoritmos de aprendizado de máquina deve ser melhor explorada.

---

## Anexo A

---

### **Artigo em coautoria**

Este artigo foi produzido, com a coautoria do discente desta tese, durante o período do doutorado. Este artigo embasou a pesquisa apresentada no Capítulo 3, Sessão 3.1.



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

F.M. Oliveira Filho <sup>a,b</sup>, F.F. Ribeiro <sup>a,e</sup>, J.A. Leyva Cruz <sup>c</sup>, A.P. Nunes de Castro <sup>d</sup>,  
G.F. Zebende <sup>c,\*</sup>



<sup>a</sup> Earth Sciences and Environment Modeling Program, State University of Feira de Santana, Bahia, Brazil

<sup>b</sup> SENAI CIMATEC University Center, Salvador, Bahia, Brazil

<sup>c</sup> State University of Feira de Santana, Bahia, Brazil

<sup>d</sup> Jorge Amado University Center, Salvador, Bahia, Brazil

<sup>e</sup> Federal University of Bahia, Salvador, Bahia, Brazil

### ARTICLE INFO

#### Article history:

Received 12 February 2023

Received in revised form 11 April 2023

Available online 5 May 2023

#### Keywords:

DFA

DCCA cross-correlation coefficient

EEG

Motor/imaginary human tasks

### ABSTRACT

EEG is one of the techniques more used to assess the extent of damage from these deficiencies and even to find solutions such as rehabilitation or limb replacement using a bionic prosthesis, through brain-computer interface. That is why it is of vital importance that we understand the functioning of the primary motor cortex of the brain in the control of real/imaginary tasks. From *Physionet* database, with two-minute EEG recordings in three different experiments (real/imaginary), we applied DFA and DCCA methods to find auto-correlations and cross-correlations. DFA method was capable of quantitatively describing similarities when the brain performs the same motor task, and show there are three time-scales. After, in order to compare the fluctuation amplitude of an EEG signal in relation to the other channels and measure these cross-correlations, we applied  $\Delta \log F_{DFA}$  function and  $\rho_{DCCA}$  coefficient. Thus, choosing the  $F_3$  channel (front) as the reference, we identified generally that:  $\Delta \log F_{DFA}[F_3 : xx] \geq 0$  and  $\rho_{DCCA} > 0$ . The channels:  $C_2$ ,  $F_6$ ,  $T_9$ , and  $T_{10}$ , are those that have a higher level of DCCA cross-correlation, if compared to the channel  $F_3$ . The time scale **II**, with  $16 < n \leq 723$ , is the one with  $\rho_{DCCA}$  maximum. Finally, the statistic applied in this paper, based on the DFA-method, proved to be an excellent candidate for studies of motor functions in the brain computer interface area.

© 2023 Elsevier B.V. All rights reserved.

## 1. Introduction

In medicine one of the most challenging problems nowadays is to decipher how the brain functioning, especially in clinically states, as a physical and/or motor impairment. It could allow us to distinguish between realized and imagined activities in the cerebrum. The human brain spontaneously generates neural bio-electric oscillations [1] with a spread range of variability in frequency, amplitude, the scale of duration, and coherence to permit connectivity of different areas involved mainly, in the higher cortical functions (HCF). The HCF is referring to all cognitive mental activity, for example, thinking, remembering, reasoning, and also complex functions, such as speaking and carrying out a premeditated movement.

Two standard techniques that are essential to study these functions, mainly due to their high temporal resolutions (in order of milliseconds), are the Electroencephalography, that use the electroencephalograms (EEG) [2] signals and the

\* Corresponding author.

E-mail addresses: [florenciofh@yahoo.com.br](mailto:florenciofh@yahoo.com.br) (F.M.O. Filho), [gfzebende@uefs.br](mailto:gfzebende@uefs.br) (G.F. Zebende).

Magnetoencephalography that makes use of the so-called magnetoencephalograms (MEG) [3]. The EEG are the temporal series of the bio-electrical voltage signals, measurement on the scalp by electrodes [4]. While the MEGs, are temporal series of the bio-magnetic fields signals, measurement on the scalp using ultra-sensitive magnetometers [5]. Both are signals produced when brain cells or neurons are synchronously activated by external or internal stimuli.

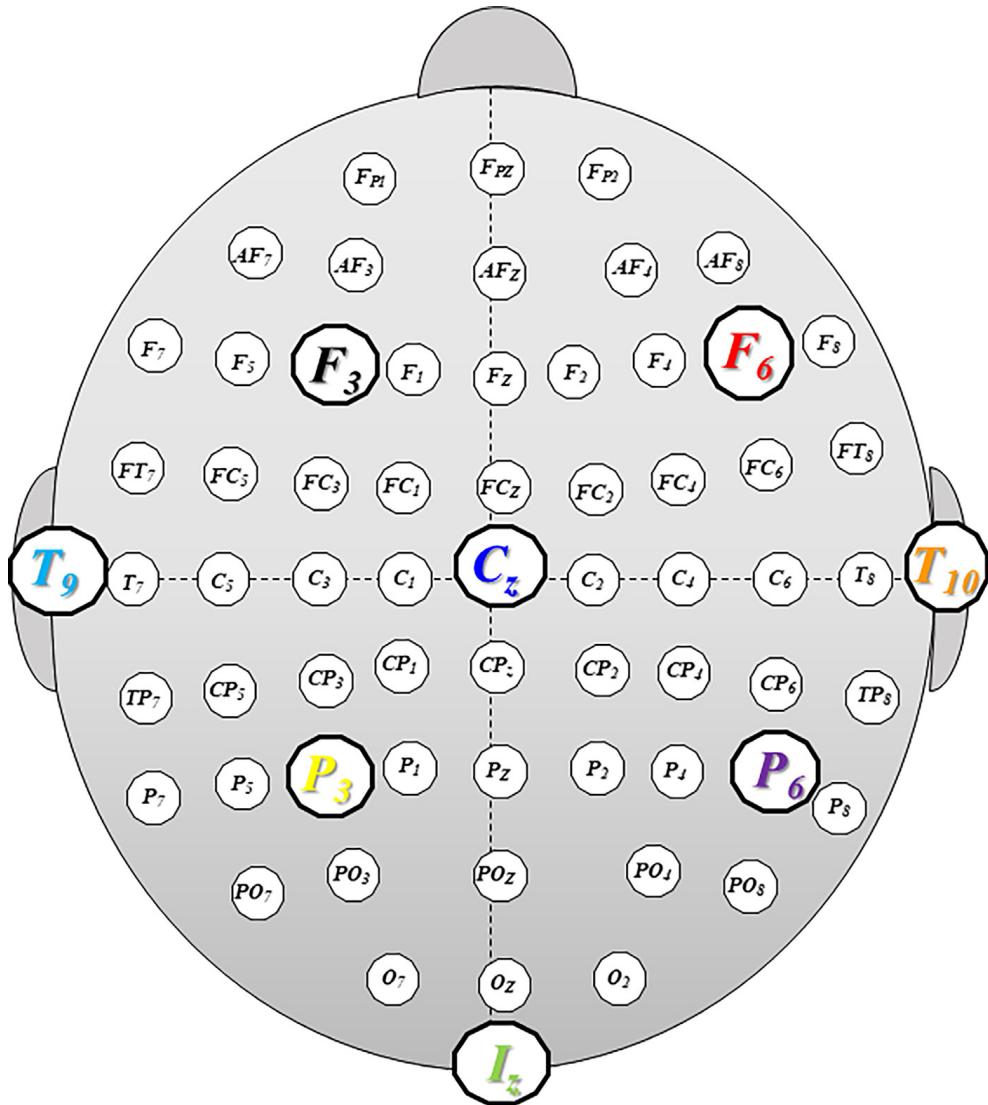
The EEG and the MEG in addition to being considered electrophysiological techniques, it is also seen as neuroimaging techniques, i.e., the reconstruction of electric and magnetic sources inner the cerebral cortex, from the measured on scalp the bio-electromagnetic fields [3,6]. The information that digital EEGs processing reveals in the spectral domain has been used for decades to diagnose several neurological pathologies, such as, in the study of the pain process in humans using evoked potentials [7]; and the diagnosis of epilepsy and autism disorders [8]. The use of a multi-modal approach using the single-photon emission computed tomography (SPECT) and quantitative EEG, have useful to diagnostics of dementia disease [9], and for the location of the epileptogenic zone in pharmacoresistant non-lesional Epilepsy [10]. In other study an integrating approach using EEG, transcranial magnetic stimulation (TMS), and magnetic resonance image (MRI) shown that is a possible way to investigate effective brain connectivity [11]. Also, EEG has been extensively used in the investigations of brain dynamics in pathological states and during cognitive or behavior tasks. Lots of measures have been used to analysis the features of muti-channel EEG signals, such as synchronization, coherence, and functional networks, see [12–15]. Normally, the EEG digital processing, for medical diagnosis purpose using spectral methods, such as Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT), among other [16–18].

The brain is considered a complex and non-linear dynamic system, and the EEG signals are classified as non-stationary with several fractal characteristics. Therefore, the statistical algorithms applied to stationary signals are not appropriated to the processing of these bio-electric signals. Actually, the bio-signal processing state of art have shown an increase of Detrended Fluctuation Analysis (DFA) based methods to study different human physiological process. For example, [1] reported presence of long-range correlations and power-law scaling behavior within the 10 to 20 Hz band (alpha and beta bands) from MEG/EEG data recorded from 10 subjects with eyes open and closed conditions, for a duration of 20 minutes. The power-law behavior was evident in the time-range of 5–300 seconds and the DFA scaling exponent in both the eyes close and open conditions were 0.68 and 0.70, significantly higher than that reference data, which was 0.5. Applying the DFA method on EEG, was demonstrate long-range temporal correlations with an attenuation of the value of  $\alpha_{DFA}$  exponent or lower similarity, in the case of Alzheimer's disease (AD) [19] and schizophrenia [20] diseases. In the case of AD, it was shown that the  $\alpha_{DFA}$  exponent was lower in AD patients compared to controls in the alpha band. While in the case of schizophrenia, the  $\alpha_{DFA}$  exponent was lower in patients in both alpha and beta bands compared to the healthy controls. The reduction of the  $\alpha_{DFA}$  exponent value can be interpreted as a difficulty for patients to remember the past. The value of  $\alpha_{DFA}$  could also potentially used as a bio-marker for neurological disorders like AD and schizophrenia, in future clinical studies.

A study demonstrated that the generalized Hurst exponent estimates differentiate EEG signals of healthy and epileptic patients [21]. In another paper, [22] was applied the DFA method to analyze serial changes in heart rhythm complexity from the acute to chronic phase on patients who suffered an acute myocardial infarction (MI), using electrocardiograms (EEGs). The EEGs were measurement on 27 patients with MI with elevation of ST-segment and 42 control subjects. The DFA  $\alpha_{DFA}$  exponent had a significantly lower value in the acute stage (within 72 hours) and also, at 3 months and 12 months after MI, in comparison with control. In a newly published paper [23], was demonstrated that the DFA method can be used as a data processing algorithm in brain-computer interfaces, to detected real and imaginary arm movements, in most EEG- channels.

Our group reported in [24,25] a new perspective for application of DFA method, with the root mean square of the fluctuation difference function,  $\Delta \log F_{DFA}$ , to study the motor real and imaginary human activity using 64-channel EEGs, and also, to the analysis of the EEG during the reading task [26]. In these papers were verified that the amplitude of the  $F_{DFA}$  rms function is greater for the frontal channels than for the parietal, and the auto-correlation exponent,  $\alpha_{DFA}$ , revealed self-affinity at a specific time scale. The detrended cross-correlation coefficient method,  $\rho_{DCCA}$  [27], was applied to analyze EEG signals from AD patients [28]. The  $\rho_{DCCA}$  presented to be lower in AD patients when compared with the control group, indicating an alteration of both synchronization and oscillation in EEGs of these patients. An attenuation of synchronization in the whole cerebrum and a bigger scale corresponding to maximum correlation was showed in AD patients.

In this paper, we analyze the brain activity in 11 subjects, in three different experiments, using DFA method,  $\Delta \log F_{DFA}$  function, and the cross-correlation coefficient  $\rho_{DCCA}$ . The main idea was to analyze the previous methods, at eight EEG channels, between different areas of the brain, described below: Frontal  $F_3$ ,  $F_6$ ; Temporal  $C_z$ ,  $T_9$ ,  $T_{10}$ ; Parietal  $P_3$ ,  $P_6$  and Occipital  $I_z$  (see Fig. 1 for more details). These electrodes were chosen in this configuration because this symmetry covers the entire scalp, see [24,26]. The Fig. 1 shows the schematic diagram of the arrangement of the 64 electrode locations on the scalp, where the EEGs were measured, in accordance with the International 10–10 system. In this way we can investigate EEGs from different channels and quantify how two areas are associated to the same scale, for a response to a motor stimuli action.



**Fig. 1.** International 10–10 system of 64 EEG electrodes. The highlighted broken lines circles identify the locations of the eight EEGs-channels used in this paper, i.e.: frontal  $F_3$  and  $F_6$ ; temporal  $T_9$  and  $T_{10}$ ; parietal  $P_3$  and  $P_6$ ; occipital  $I_z$ ; and central  $C_z$ .

## 2. Experimental protocol and DFA statistical method

### 2.1. Experimental protocol

The experimental data used in this work were extracted from the well-known *Physionet* database, available at the address [29]:

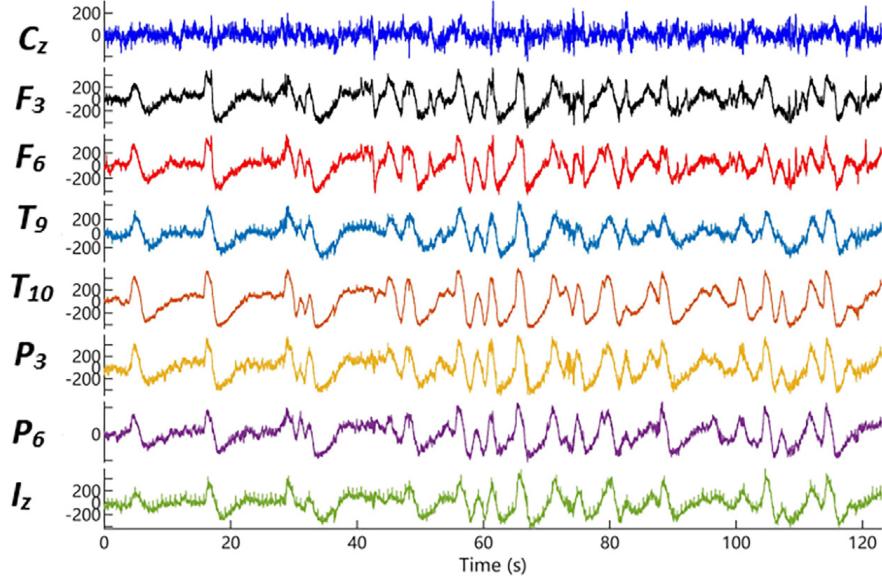
<https://archive.physionet.org/pn4/eegmmidb/>

This dataset consists of over 1500 one and two-minute EEG recordings, obtained from 109 volunteers. From this set to analyze their EEG, we chose 11 subjects randomly, represented here by:

$S_{010}, S_{020}, S_{029}, S_{043}, S_{046}, S_{050}, S_{051}, S_{060}, S_{071}, S_{086}, S_{099}$

Each subject executed different real/imagery motor tasks. These tasks were initialized in answer to different events when the subject observed some targets. In real motor experiments, the target appears on either the left or the right side of the screen. The subject opens and closes the corresponding fist until the target disappears. Then the subject relaxes. At later time, the imagery motor tasks were executed, when the target appears on the same sides on the screen equal to the real motor tasks, the subject imagines opening and closing the corresponding fist until the target disappears. Then the subject relaxes. Within the experimental protocol, the chosen EEGs were measured when the volunteers performed the following real/imagery motor tasks:

**Task 1** open and close left or right fists;



**Fig. 2.** EEG time series from subject  $S_{010}$  performing the **Task 1**, with the following channels:  $C_z$ ,  $F_3$ ,  $F_6$ ,  $T_9$ ,  $T_{10}$ ,  $P_3$ ,  $P_6$  and  $I_z$ .

**Task 2** imagine opening and closing left or right fists;

**Task 3** open and close both fists or both feet;

**Task 4** imagine opening and closing both fists or both feet.

The EEGs were measured using a 64-channel BCI2000 system [30]:

<http://www.bci2000.org>

with a sampled frequency of 160 samples per second and saved in EDF+ format. In Fig. 2, as an example for the EEG signals. We can observe the time series about eight channels from the subject  $S_{010}$  performing the **Task 1**.

## 2.2. DFA statistical method

Proposed by Peng [31], the DFA-based method proved to be a tool capable of providing a priori parameter to detect the auto-correlation of non-stationary temporal signals. The main advantage of the DFA method, in comparison with other fractal analysis methods, is to avoid spurious detection of unreal correlations, which are non-stationarity artifacts [32,33]. The steps to be follow for apply the DFA-based method, considering a signal  $u(i)$  (such as an EEG bio-signal), where  $i = 1, \dots, N_{max}$ , and  $N_{max}$  is the length of the EEG time series, were:

- (1) Obtain the integrated time series by,  $y(k) = \sum_{i=1}^k [u(i) - \langle u \rangle]$ , here  $\langle u \rangle$  represent the average value of  $u(i)$ ;
- (2) The integrated bio-signal  $y(k)$  is decomposed into boxes of equal size  $n$  (time scale);
- (3) For each  $n$  size box, we fit  $y(k)$ , using a polynomial function of order  $\geq 1$ , called by  $y_n(k)$ ;
- (4) The integrated signal  $y(k)$  is detrended by subtracting the local trend  $y_n(k)$  in each box of length  $n$ ;
- (5) The  $F_{DFA}(n)$  root mean square fluctuation (rms) function for this integrated and detrended signal is compute by,

$$F_{DFA}(n) = \sqrt{\frac{1}{N_{max}} \sum_{k=1}^{N_{max}} [y(k) - y_n(k)]^2} \quad (1)$$

These steps are repeated for all time scales between  $4 \leq n \leq N_{max}/4$ .

(6) Finally, we need to investigate if the relationship between  $F_{DFA}(n)$  and the box size ( $n$ ), behaves a power-law, according to Eq. (2) below,

$$F_{DFA}(n) \propto n^{\alpha_{DFA}} \quad (2)$$

In the affirmative case, then  $\alpha_{DFA}$  exponent can be interpreted as a long-range correlation exponent, namely an auto-affinity parameter. This relation is normally illustrated in a  $\log F(n) \times \log(n)$  graph, being represented by a straight line with slope equal to  $\alpha_{DFA}$ . To know more about this coefficient, see the Table 1.

Following the methodology, the approach using the function  $\Delta \log F_{DFA}$ , have the proposal of measure the difference in the fluctuation amplitude between two EEGs channels [24,26]. This tool is an enhancement given to the DFA method [31], and has been shown to be useful when applied to the analysis of electrophysiological signals. Using this function, we can

**Table 1**  
Statistical interpretation of the exponent  $\alpha_{DFA}$ .

$\alpha_{DFA}$ exponent	Signal type
$\alpha_{DFA} < 0.5$	Anti-persistent
$\alpha_{DFA} \simeq 0.5$	Uncorrelated, white noise
$\alpha_{DFA} > 0.5$	Long-range correlated persistent
$\alpha_{DFA} \simeq 1$	1/f noise
$\alpha_{DFA} > 1$	Non-stationary
$\alpha_{DFA} \simeq 3/2$	Brownian noise

study how much two brain regions are auto-correlated on the same scale (coherence temporal). In practice, we applied the DFA-based method on two temporal series and their logarithms individually are calculated by Eq. (3), taking as an example the  $F_3$  and any other  $xx$  channel:

$$\Delta \log F_{DFA}[F_3 : xx] = \log F_{DFA}[F_3] - \log F_{DFA}[xx] \quad (3)$$

In this case we can have three possible conditions, these are:

- (a)  $\Delta \log F_{DFA}[F_3 : xx] > 0$ : the amplitude of  $F_{DFA}[F_3]$  function on channel  $F_3$  is **larger** if compared with  $F_{DFA}_{xx}$  from  $xx$  channel;
- (b)  $\Delta \log F_{DFA}[F_3 : xx] = 0$ : the amplitude of  $F_{DFA}[F_3]$  function on channel  $F_3$  is **similar** with  $F_{DFA}_{xx}$  from  $xx$  channel;
- (c)  $\Delta \log F_{DFA}[F_3 : xx] < 0$ : the amplitude of  $F_{DFA}[F_3]$  function on channel  $F_3$  is **smaller** if compared with  $F_{DFA}_{xx}$  from  $xx$  channel.

It is important to note that the DFA method itself performs only self-affinity analysis for only a single time series, as well as the  $\Delta \log F_{DFA}$  function. With the intention of measuring cross-correlations between two time series, Podobnik [34] created the Detrended Cross-Correlation Analysis (DCCA) method, generalization of the DFA, to estimate the cross-correlation between two non-stationary time series by means of the detrended covariance function,  $F_{DCCA}$ . Therefore, repeating the similar procedure described in DFA method for two time series, it is possible to verify the existence or not of cross-correlation power-law, described by

$$F_{DCCA}^2 \propto n^{2\lambda} \quad (4)$$

Here,  $\lambda$  describes the long-range cross-correlation exponent. The exponent  $\lambda$  identifies and measures long-range cross-correlations between two signals in different time scales  $n$ , but does not quantify the level of cross-correlation. To quantify this level of cross-correlation, Zebende [27] proposed the cross-correlation coefficient,  $\rho_{DCCA}$ . The DCCA cross-correlation coefficient is obtained as the ratio between the DCCA covariance function  $F_{DCCA[X_i, Y_i]}^2(n)$ , and each DFA auto-correlation function by:

$$\rho_{DCCA}(n) = \frac{F_{DCCA[X_i, Y_i]}^2(n)}{F_{DFA[X_i]}(n)F_{DFA[Y_i]}(n)} \quad (5)$$

This DCCA cross-correlation coefficient is dimensionless and range between:

$$-1 \leq \rho_{DCCA} \leq 1.$$

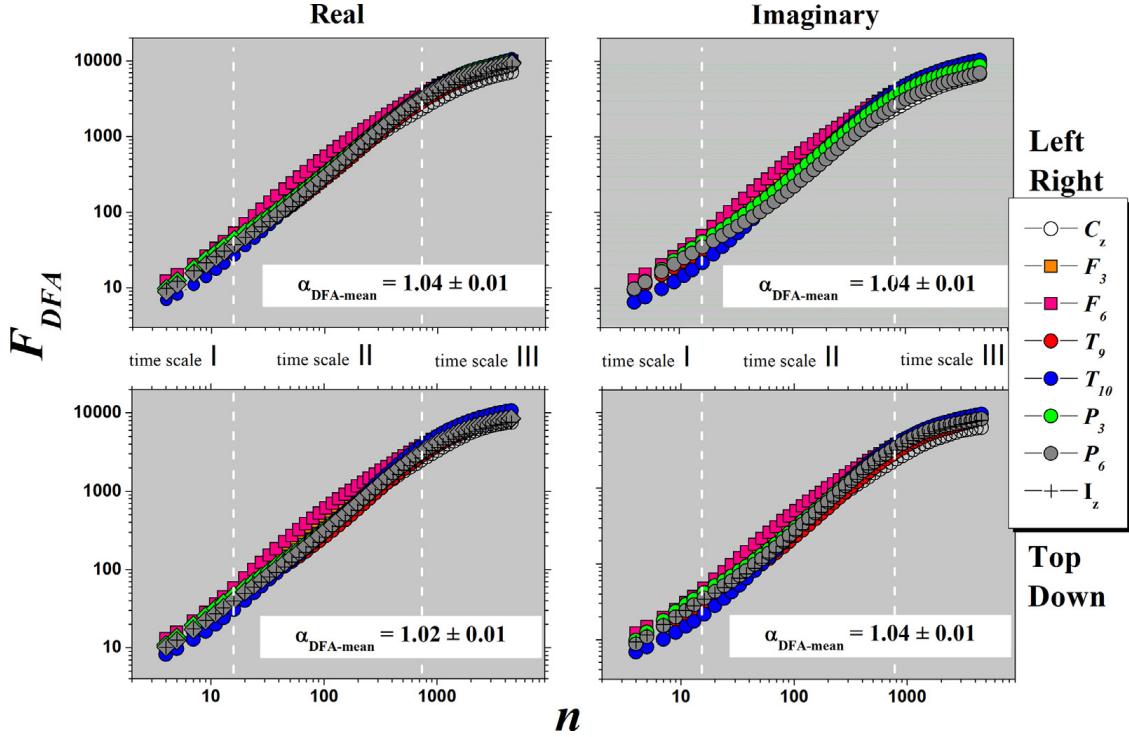
In this case, if  $\rho_{DCCA} = -1$ , we have perfect anti cross-correlation. If  $\rho_{DCCA} = 0$ , then we have no cross-correlation. If,  $\rho_{DCCA} = 1$ , we will get a perfect cross-correlation between the time series  $\{X_i\}$  and  $\{Y_i\}$ . Also, in [35,36] was reported that the DCCA cross-correlation coefficient dominates the Pearson coefficient for non-stationary series analysis.

### 3. Results and discussions

Initially with the time series of the EEG from 11 subjects, we calculated the  $F_{DFA}$  function, for each set of motor tasks (**Real/Imaginary**) in response to the **Right/Left** and **Top/Down** visual stimulus. To detail the influence of the  $F_{DFA}$  method, we calculated the traditional coefficient  $\alpha_{DFA}$  for all possible time scale and for three different time scales (choice made visually and also with the best adjustments):

- all time scale:  $4 \leq n \leq 4564$ ;
- time scale I:  $n \leq 16$ ;
- time scale II:  $16 < n \leq 723$ ;
- time scale III:  $n > 723$ .

Following Fig. 1, we used only 8 channels distributed over the brain: **frontal**  $F_3$  and  $F_6$ ; **temporal**  $C_z$ ,  $T_9$  and  $T_{10}$ ; **parietal**  $P_3$  and  $P_6$ ; **occipital**  $I_z$ . Each time series containing 19.680 points with approximately 2 minutes, and a sampling period of the  $\Delta T = 6.25$  ms. As an example, in Fig. 3 is shown the behavior of  $F_{DFA}$  as a function of  $n$  for all time scale and for the  $S_{010}$ ,



**Fig. 3.**  $F_{DFA}$  as a function of  $n$  for subject  $S_{010}$  performing all 4 Tasks. The continuous yellow line defines the linear fit of the experimental data for all time scale. The different auto-correlation exponents and the associated errors are also presented. Vertical dashed lines represent the time scale (I, II, III).

in this case performing all four tasks: Open/Close (Real/Imaginary) the Left/Right fists, and Open/Close (Real/Imaginary) both fists (Top) or both feet (Down), depending the visual stimulus on screen. The auto-correlation exponent  $\alpha_{DFA}$  values for all tasks are greater than 1. These reveal that the EEGs are non-stationary series with long-range persistence or self affinity, in according with previous reports [35]. Interpreting Fig. 3 more deeply, we can state that the EEG signals in **Task 1** and **Task 2** are very similar, because:

$$\alpha_{DFA}(\text{Real}) = \alpha_{DFA}(\text{Imaginary})$$

(Real) = (Imaginary) which means that **Task 1** and **Task 2** represent the same metal activity. The same conclusion can be made about the **Task 3** and **Task 4**, with  $\alpha_{DFA}(\text{Real}) - \alpha_{DFA}(\text{Imaginary})$  of approximately 0.07. These preliminary results, with  $S_{010}$ , assertively shows that it is indifferent to carry out Real or Imaginary experiments to study brain motor activity.

For better accuracy, we now did it with a sample of 11 subjects and the eight channels. The results for  $\alpha_{DFA}$  (in time scale I, II and III) can be seen in Table 2 . In this Table we present the mean value of  $\alpha_{DFA}$  for the 11 subjects. The results for  $\langle \alpha_{DFA} \rangle$  mean are presented in the Table 2 , indicating a scale-dependent pattern.

Here we can see that in the overall average of the 11 subjects and all channels (last line in the Table 2 ), there are no major changes if the experiment was Real or Imaginary, because the auto-correlation exponent are very similar. Therefore, if we look at each channel separately, we can see that each one has its characteristic value of  $\langle \alpha_{DFA} \rangle$  (higher values for the front channels). Also, we can observe for small ( $n \leq 16$ ) and medium ( $16 < n \leq 723$ ) time scale,  $0.99 < \langle \alpha_{DFA} \rangle < 1.22$ . For large time scales ( $n > 723$ ),  $\langle \alpha_{DFA} \rangle$  presents a value lower than 0.5, except for the channels  $C_z$ ,  $F_3$ ,  $F_6$ , and  $T_9$ , when executed the Real tasks, and the channel  $C_z$ , for the Imaginary task. Based on the Table 1 , in general between the time scale I and time scale II we can observe the transition between non-stationary and noise signals, for Real/Imaginary experiments. Lastly, in the time scale III,  $\langle \alpha_{DFA} \rangle < 0.5$ , showing an anti-persistent behavior on this scale.

Now, knowing that the front channel  $F_3$  has the greatest amplitude of fluctuation, revealed by [24], here was calculated the function  $\Delta \log F_{DFA}[F_3 : xx]$  described in Eq. (3), in order to measure on average where there is greater relative brain activity and in which time scale. This result is presented in Fig. 4. This figure shows the relationship between two points of the brain. Generally, it is possible to observe a positive difference in the fluctuation amplitude for all time scales and channels, in other words,

$$\Delta \log F_{DFA}[F_3 : xx] \geq 0.$$

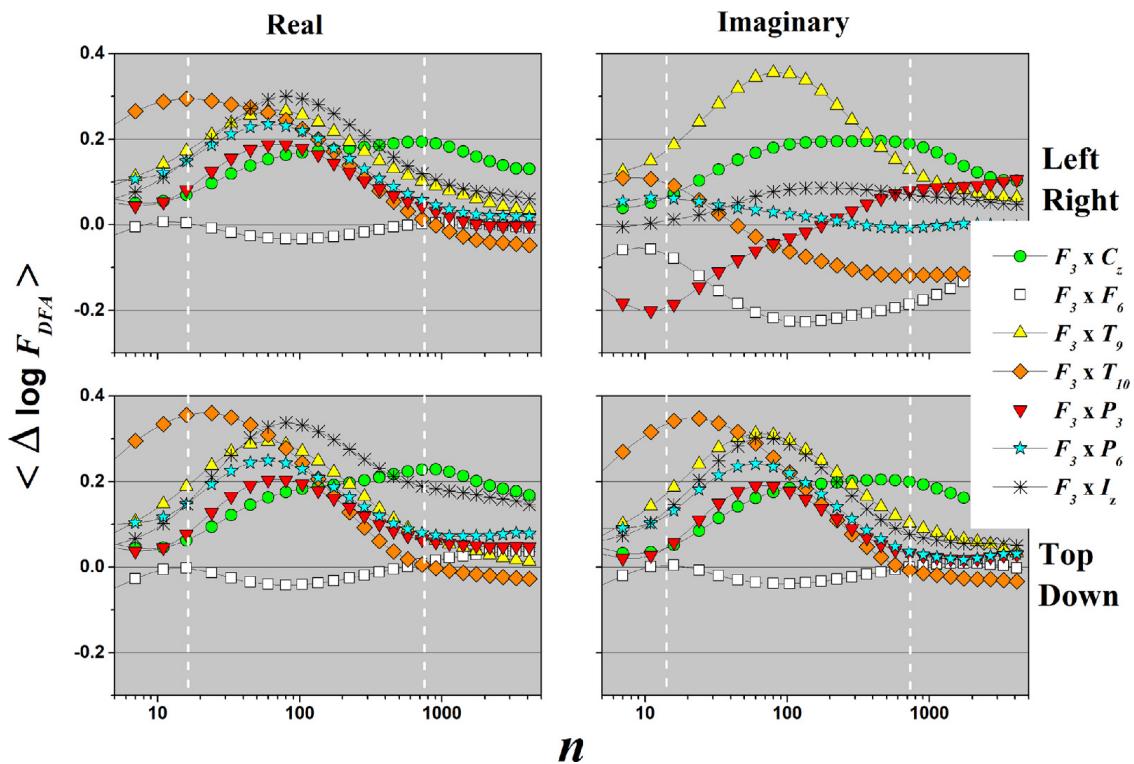
However, if we compare the channel  $F_3$  with the channel  $F_6$  (also on the front of the scalp), we can see that

$$\Delta \log F_{DFA}[F_3 : F_6] \simeq 0.$$

**Table 2**

Mean values of  $\alpha_{DFA}$  for all experiments. The first column contains the stimuli (Left/Right or Top/Down), second column the channel, third (Real case), and fourth (Imaginary case) represent the  $\alpha_{DFA}$  at time scale: **I** ( $n \leq 16$ ), **II** ( $16 < n \leq 723$ ) and **III** ( $n > 723$ ).

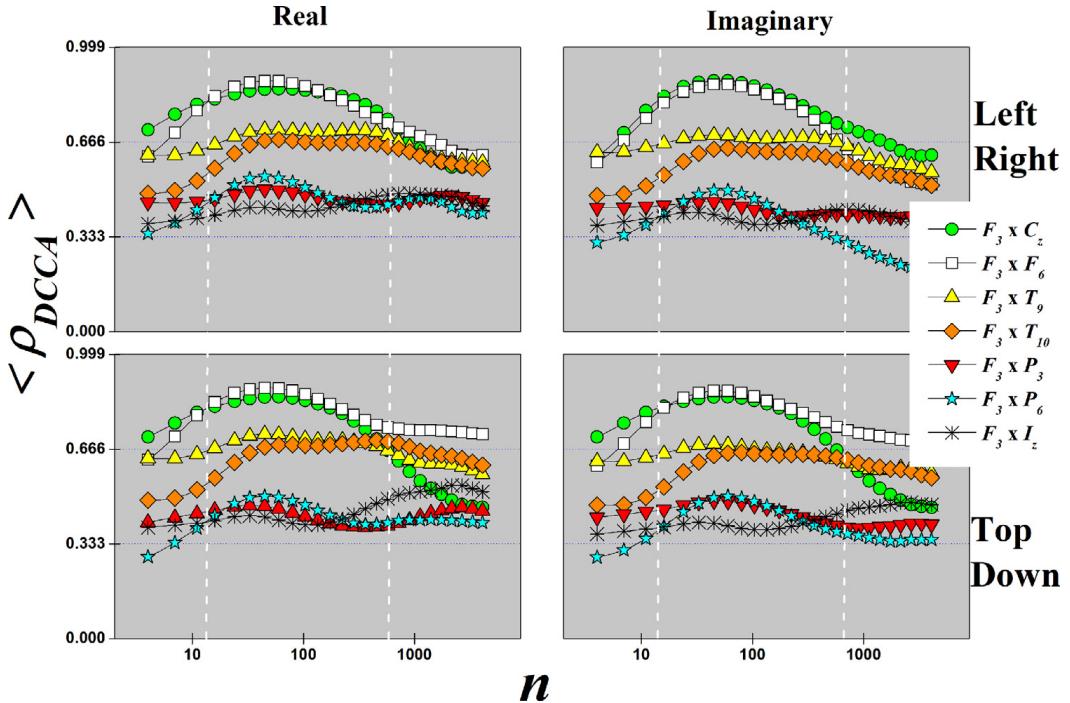
Task	Channel	Real			Imaginary		
		I	II	III	I	II	III
Left/Right	$C_z$	1.10	0.94	0.57	1.07	0.93	0.50
	$F_3$	1.19	1.00	0.51	1.01	0.97	0.42
	$F_6$	1.21	0.94	0.51	1.19	0.89	0.38
	$T_9$	1.04	1.16	0.51	1.08	1.21	0.48
	$T_{10}$	1.14	1.14	0.44	1.18	1.21	0.39
	$P_3$	1.04	1.04	0.40	1.02	1.1	0.40
	$P_6$	1.08	1.11	0.41	0.93	1.23	0.48
	$I_z$	0.99	1.16	0.47	1.05	1.08	0.37
<b>Mean</b>		1.10	1.06	0.48	1.07	1.08	0.43
Top/Down	$C_z$	1.09	0.95	0.39	1.05	0.96	0.39
	$F_3$	1.21	0.90	0.32	1.17	0.91	0.33
	$F_6$	1.22	0.88	0.39	1.18	0.90	0.36
	$T_9$	1.01	1.14	0.42	0.99	1.12	0.41
	$T_{10}$	1.15	1.18	0.38	1.20	1.02	0.35
	$P_3$	1.03	1.06	0.33	1.03	1.02	0.29
	$P_6$	1.06	1.10	0.35	1.05	1.09	0.30
	$I_z$	0.98	1.06	0.37	0.96	1.12	0.34
<b>Mean</b>		1.09	1.03	0.37	1.08	1.02	0.35



**Fig. 4.** Mean value of  $\Delta \log F_{DFA}[F_3 : xx]$  as a function of  $n$ , considering the difference between the channel  $F_3$  and the channels:  $C_z$ ,  $F_6$ ,  $T_9$ ,  $T_{10}$ ,  $P_3$ ,  $P_6$  and  $I_z$ . Vertical dashed lines represent the time scale (I, II, III).

This result evidences the greater activity in the frontal part of the brain. It is verified that for other channels, excepting  $F_6$ , that  $\Delta \log F_{DFA}[F_3 : xx]$  has its maximum value. For large time scale,  $n > 1000$ ,  $\Delta \log F_{DFA}[F_3 : xx]$  tends to a constant value.

As the last case of study in relation to the tasks performed by the 11 subjects, and considering the level of adherence between the EEG channels, we will apply from this point the DCCA cross-correlation coefficient,  $\rho_{DCCA}$ , between the signals (7 channels) having as support the channel  $F_3$ . The results for this application are shown in Fig. 5, with  $\langle \rho_{DCCA} \rangle$  mean. In Fig. 5, there is positive cross-correlation for all channels if compared to the channel  $F_3$ . The channels  $C_z$ ,  $F_6$ ,  $T_9$ , and  $T_{10}$ , are those that have a higher level of DCCA cross-correlation if compared to the channel  $F_3$  (in this order), and the



**Fig. 5.** Mean values for  $\rho_{DCCA}$  as a function of  $n$ . Each curve represent the cross correlation between the channel  $F_3$  with the other channels, represented here by:  $C_z$ ,  $F_6$ ,  $T_9$ ,  $T_{10}$ ,  $P_3$ ,  $P_6$  and  $I_z$ . Vertical dashed lines represent the specific time scale (I, II, III).

time scale II, with  $16 < n \leq 723$  is the time scale with  $\rho_{DCCA}$  maximum. It is possible to notice that the value of  $\rho_{DCCA}$  its similar for all talks, Real/Imaginary, Left/Right or Top/Down. This behavior presented by  $\rho_{DCCA}$  is probably due to the stimulus presentation time and consequently the brain response, when the subjects are performing the tasks

#### 4. Conclusions

Through the statistical analysis of time series, it is proved that the proposed DFA based methods are capable of quantitatively describing similarities when the brain performs the same motor task in a real or imaginary way. Consequently, considering a referenced EEG database, we can assertively conclude that: The statistical results first showed us that there are three important time scales, which are, I ( $n \leq 16$ ), II ( $16 < n \leq 723$ ), and III ( $n > 723$ ), as well as, from the point of view of DFA auto-correlation,  $\alpha_{DFA}(\text{Real}) = \alpha_{DFA}(\text{Imaginary})$ , for all time scales. Also, in order to compare the fluctuation amplitude of an EEG signal in relation to the other channels, we applied the  $\Delta \log F_{DFA}$ . Thus, choosing the  $F_3$  channel (front) as the reference, we identified generally that,  $\Delta \log F_{DFA}[F_3 : xx] \geq 0$  at its average value, except for the channel  $F_6$  (also on the front of the scalp) that has a value for  $\Delta \log F_{DFA}[F_3 : F_6] \simeq 0$ . With DCCA cross-correlation coefficient, we can see that  $\rho_{DCCA} > 0$ , showing that the same brain signal/response was measured for all channels. The channels:  $C_z$ ,  $F_6$ ,  $T_9$ , and  $T_{10}$ , are those that have a higher level of DCCA cross-correlation, if compared to the channel  $F_3$ . The time scale II, with  $16 < n \leq 723$ , is the one with  $\rho_{DCCA}$  maximum. The  $\rho_{DCCA}$  statistic, quantify high temporal coherence similitude of the EEGs performed by the brain, between the Real and Imaginary motor human tasks. For this reason, is plausible choose the  $\rho_{DCCA}$  to quantitative study, this Real/Imaginary brain motor function, or possibly other types of the EEGs. Finally, the statistic applied in this paper, based on the DFA-method, proved to be an excellent candidate for studies of motor functions in the brain computer interface area.

#### CRediT authorship contribution statement

**F.M. Oliveira Filho:** Conceptualization, Methodology, Software, Writing – original draft, Visualization, Validation, Writing – review & editing. **F.F. Ribeiro:** Conceptualization, Visualization. **J.A. Leyva Cruz:** Conceptualization, Methodology, Software, Writing – original draft, Visualization, Validation, Writing – review & editing. **A.P. Nunes de Castro:** Software. **G.F. Zebende:** Conceptualization, Methodology, Software, Writing – original draft, Visualization, Validation, Writing – review & editing.

#### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gilney Figueira Zebende and Florêncio Mendes de Oliveira Filho reports financial support provided by National Council for Scientific and Technological Development.

## Data availability

Data will be made available on request.

## Acknowledgments

G. F. Zebende thanks the Brazilian agency CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil (Grant 310136/2020-2). F. M. Oliveira Filho thanks the Brazilian agency CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil (Grant 150655/2022-3). We would also like to thank Chico Mendes, rubber tapper, political activist and Brazilian environmentalist (*in memoriam*).

## References

- [1] K. Linkenkaer-Hansen, V.V. Nikouline, J.M. Palva, R.J. Ilmoniemi, Long-range temporal correlations and scaling behavior in human brain oscillations, *J. Neurosci.* 21 (4) (2001) 1370–1377.
- [2] A. Biasiucci, B. Franceschiello, M.M. Murray, Electroencephalography, *Curr. Biol.* 29 (3) (2019) R80–R85.
- [3] F.L. da Silva, EEG and MEG: relevance to neuroscience, *Neuron* 80 (5) (2013) 1112–1128.
- [4] H. Berger, Über das elektroenzephalogramm des menschen, *Archiv Für Psychiatrie Und Nervenkrankheiten* 87 (1) (1929) 527–570.
- [5] D. Cohen, Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer, *Science* 175 (4022) (1972) 664–666.
- [6] M.C. Piastra, A. Nüßing, J. Vorwerk, M. Clerc, C. Engwer, C.H. Wolters, A comprehensive study on electroencephalography and magnetoencephalography sensitivity to cortical and subcortical sources, *Human Brain Mapping* 42 (4) (2021) 978–992.
- [7] M. Ploner, E.S. May, Electroencephalography and magnetoencephalography in pain research—current state and future perspectives, *Pain* 159 (2) (2018) 206–211.
- [8] S. Ibrahim, R. Djemal, A. Alsuwailem, Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis, *Biocybern. Biomed. Eng.* 38 (1) (2018) 16–26.
- [9] Y. Höller, A.C. Bathke, A. Uhl, N. Strobl, A. Lang, J. Bergmann, R. Nardone, F. Rossini, H. Zauner, M. Kirschner, et al., Combining SPECT and quantitative EEG analysis for the automated differential diagnosis of disorders with amnestic symptoms, *Front. Aging Neurosci.* 9 (2017) 290.
- [10] K. Batista García-Ramó, C.A. Sanchez Catasus, L. Morales Chacón, A. Aguila Ruiz, A. Sánchez Corneaux, P. Rojas López, J. Bosh Bayard, A novel noninvasive approach based on SPECT and EEG for the location of the epileptogenic zone in pharmacoresistant non-lesional epilepsy, *Medicina* 55 (8) (2019) 478.
- [11] R. Esposito, M. Bortoletto, C. Miniussi, Integrating TMS, EEG, and MRI as an approach for studying brain connectivity, *The Neuroscientist* 26 (5–6) (2020) 471–486.
- [12] H. Yu, J. Liu, L. Cai, J. Wang, Y. Cao, C. Hao, Functional brain networks in healthy subjects under acupuncture stimulation: An EEG study based on nonlinear synchronization likelihood analysis, *Phys. A* 468 (2017) 566–577.
- [13] H. Yu, X. Wu, L. Cai, B. Deng, J. Wang, Modulation of spectral power and functional connectivity in human brain by acupuncture stimulation, *IEEE Trans. Neural Syst. Rehabil. Eng.* 26 (5) (2018) 977–986.
- [14] H. Yu, X. Li, X. Lei, J. Wang, Modulation effect of acupuncture on functional brain networks and classification of its manipulation with EEG signals, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (10) (2019) 1973–1984.
- [15] H. Yu, X. Lei, Z. Song, C. Liu, J. Wang, Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification, *IEEE Trans. Fuzzy Syst.* 28 (1) (2020) 60–71.
- [16] D.G. de Figueiredo, Análise de Fourier E Equações Diferenciais Parciais, IMPA, Projeto Euclides, 1977.
- [17] E. Kanasewich, The fast Fourier transform: Time sequence analysis in geophysics, third ed., The University of Alberta Press, 1974.
- [18] R. Sedgewick, Algorithms in C, Princeton Universit, 1995, Cap.41 - The Fast Fourier Transform p.583.
- [19] T. Montez, S.-S. Poil, B.F. Jones, I. Manshanden, J.P. Verbunt, B.W. van Dijk, A.B. Brussaard, A. van Ooyen, C.J. Stam, P. Scheltens, et al., Altered temporal correlations in parietal alpha and prefrontal theta oscillations in early-stage Alzheimer disease, *Proc. Natl. Acad. Sci.* 106 (5) (2009) 1614–1619.
- [20] V.V. Nikulin, E.G. Jönsson, T. Brismar, Attenuation of long-range temporal correlations in the amplitude dynamics of alpha and beta neuronal oscillations in patients with schizophrenia, *Neuroimage* 61 (1) (2012) 162–169.
- [21] S. Lahmiri, Generalized hurst exponent estimates differentiate EEG signals of healthy and epileptic patients, *Phys. A* 490 (2018) 378–385.
- [22] H.-C. Chiu, H.-P. Ma, C. Lin, M.-T. Lo, L.-Y. Lin, C.-K. Wu, J.-Y. Chiang, J.-K. Lee, C.-S. Hung, T.-D.W. abd Li-Yu Daisy Liu, Y.-L. Ho, Y.-H. Lin, C.-K. Peng, Serial heart rhythm complexity changes in patients with anterior wall ST segment elevation myocardial infarction, *Sci. Rep.* 7 (2017) 43507.
- [23] A. Pavlov, A. Runnova, V. Maksimenko, O. Pavlova, D. Grishina, A. Hramov, Detrended fluctuation analysis of EEG patterns associated with real and imaginary arm movements, *Phys. A* 509 (2018) 777–782.
- [24] G.F. Zebende, F.M. Oliveira Filho, J.A. Leyva Cruz, Auto-correlation in the motor/imaginary human EEG signals: A vision about the FDFA fluctuations, *PLoS ONE* 12 (9) (2017) e0183121.
- [25] V. Mesquita, F. Oliveira Filho, P.C. Rodrigues, Detection of crossover points in detrended fluctuation analysis: an application to EEG signals of patients with epilepsy, *Bioinformatics* 37 (9) (2021) 1278–1284.
- [26] F. Oliveira Filho, J.L. Cruz, G. Zebende, Analysis of the EEG bio-signals during the reading task by DFA method, *Phys. A* 525 (2019) 664–671.
- [27] G.F. Zebende, DCCA cross-correlation coefficient: Quantifying level of cross-correlation, *Phys. A* 390 (4) (2011) 614–618.
- [28] Y. Chen, L. Cai, R. Wang, Z. Song, B. Deng, J. Wang, H. Yu, DCCA cross-correlation coefficients reveals the change of both synchronization and oscillation in EEG of Alzheimer disease patients, *Phys. A* 490 (2018) 171–184.
- [29] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, *Circulation* 101 (23) (2000) e215–e220.
- [30] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, BCI2000: a general-purpose brain-computer interface (BCI) system, *IEEE Trans. Biomed. Eng.* 51 (6) (2004) 1034–1043.
- [31] C.-K. Peng, S.V. Buldyrev, S. Havlin, F. Sciortino, H.E. Stanley, Long-range correlations in nucleotide sequences, *Nature* 356 (1992) 168–170.
- [32] D. Mirzayev, Y. Ashkenazy, Preservation of long range temporal correlations under extreme random dilution, *Phys. A* 389 (24) (2010) 5573–5580.
- [33] C. Heneghan, G. McDarby, Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes, *Phys. Rev. E* 62 (5) (2000) 6103.
- [34] B. Podobnik, H.E. Stanley, Detrended Cross-Correlation Analysis: A new method for analyzing two nonstationary time series, *Phys. Rev. Lett.* 100 (8) (2008) 084102.
- [35] D.A. Blythe, V.V. Nikulin, K.-R. Müller, Robust statistical detection of power-law cross-correlation, *Sci. Rep.* 6 (1) (2016) 1–10.
- [36] L. Kristoufek, Measuring correlations between non-stationary series with DCCA coefficient, *Phys. A* 402 (2014) 291–298.

## Anexo B: Publicações técnicas

---

Este anexo apresenta registros de *softwares* desenvolvidos durante o curso desta pesquisa.



REPÚBLICA FEDERATIVA DO BRASIL  
MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA, COMÉRCIO E SERVIÇOS  
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL  
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS

## Certificado de Registro de Programa de Computador

Processo Nº: **BR512023001376-5**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 15/05/2023, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

**Título:** Multiple EEG cross-correlation

**Data de publicação:** 15/05/2023

**Data de criação:** 01/02/2022

**Titular(es):** GILNEY FIGUEIRA ZEBENDE; ANDREA DE ALMEIDA BRITO; FLORÊNCIO MENDES OLIVEIRA FILHO; FERNANDO FERRAZ RIBEIRO; JUAN ALBERTO LEYVA CRUZ

**Autor(es):** GILNEY FIGUEIRA ZEBENDE; ANDREA DE ALMEIDA BRITO; FLORÊNCIO MENDES OLIVEIRA FILHO; JUAN ALBERTO LEYVA CRUZ; FERNANDO FERRAZ RIBEIRO

**Linguagem:** PYTHON

**Campo de aplicação:** MT-06

**Tipo de programa:** SM-01

**Algoritmo hash:** SHA-256

**Resumo digital hash:** 28a265f9b7e4b4ff4e40d57f8fb671cb2dd941d7c6579e607b4392d362b64cbd

**Expedido em:** 23/05/2023

**Aprovado por:**

Carlos Alexandre Fernandes Silva  
Chefe da DIPTO

---

## Referências Bibliográficas

---

- ADARSH, S. et al. Multifractal description of daily rainfall fields over India. *Journal of Hydrology*, v. 586, p. 124913, 2020. ISSN 0022-1694. [2.5](#)
- AFUECHETA, E.; OMAR, M. H. Characterization of variability and trends in daily precipitation and temperature extremes in the Horn of Africa. *Climate Risk Management*, v. 32, p. 100295, 2021. ISSN 2212-0963. [2.5](#)
- ASLAM, F. et al. Interplay of multifractal dynamics between shadow policy rates and energy markets. *The North American Journal of Economics and Finance*, v. 71, p. 102085, 2024. ISSN 1062-9408. [2.5](#)
- 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. [1](#)
- BIANCHI, S.; LONGO, A.; PLASTINO, W. A new methodological approach for worldwide beryllium-7 time series analysis. *Physica A: Statistical Mechanics and its Applications*, v. 501, p. 377–387, 2018. ISSN 0378-4371. [2.5](#)
- CHATTERJEE, S.; GHOSH, D. Impact of Global Warming on SENSEX fluctuations — A study based on Multifractal detrended cross correlation analysis between the temperature anomalies and the SENSEX fluctuations. *Physica A: Statistical Mechanics and its Applications*, v. 571, p. 125815, 2021. ISSN 0378-4371. [2.5](#)
- CHEN, Y. et al. DCCA cross-correlation coefficients reveals the change of both synchronization and oscillation in EEG of Alzheimer disease patients. *Physica A: Statistical Mechanics and its Applications*, v. 490, p. 171–184, 2018. ISSN 0378-4371. [2.5](#)
- CHEN, Y. et al. Cross-correlation and multifractality analysis of the Chinese and American stock markets based on the MF-DCCA model. *Heliyon*, v. 10, n. 17, p. e36537, 2024. ISSN 2405-8440. [2.5](#)
- Contreras-Reyes, J. E.; Jeldes-Delgado, F.; CARRASCO, R. Jensen-Detrended Cross-Correlation function for non-stationary time series with application to Latin American stock markets. *Physica A: Statistical Mechanics and its Applications*, v. 654, p. 130115, 2024. ISSN 0378-4371. [2.5](#)
- da Silva Filho, A. M. et al. Statistical test for Multiple Detrended Cross-Correlation Coefficient. *Physica A: Statistical Mechanics and its Applications*, Elsevier B.V., v. 562, p. 125285, 2021. ISSN 03784371. [2.2](#)
- 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. [1](#)
- FERREIRA, P. et al. A sliding windows approach to analyse the evolution of bank shares in the European Union. *Physica A: Statistical Mechanics and its Applications*, v. 490, p. 1355–1367, 2018. ISSN 03784371. [2.2](#)

- FILHO, A. M.; da Silva, M.; ZEBENDE, G. Autocorrelation and cross-correlation in time series of homicide and attempted homicide. *Physica A: Statistical Mechanics and its Applications*, v. 400, p. 12–19, 2014. ISSN 0378-4371. [2.5](#)
- 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, 2023. ISSN 0378-4371. [2.5](#)
- 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. [3.1](#)
- GE, X.; LIN, A. Quantifying the direct and indirect interactions for EEG signals by using detrended permutation mutual information. *Chaos, Solitons & Fractals*, v. 176, p. 114155, 2023. ISSN 0960-0779. [2.5](#)
- GHOSH, D. et al. Chaos based nonlinear analysis to study cardiovascular responses to changes in posture. *Physica A: Statistical Mechanics and its Applications*, v. 512, p. 392–403, 2018. ISSN 0378-4371. [2.5](#)
- GUEDES, E. et al. DCCA cross-correlation in blue-chips companies: A view of the 2008 financial crisis in the Eurozone. *Physica A: Statistical Mechanics and its Applications*, Elsevier B.V., v. 479, p. 38–47, 2017. ISSN 03784371. [2.2](#)
- GUEDES, E. F. et al. Statistical test for  $\Delta\rho$ DCCA cross-correlation coefficient. *Physica A: Statistical Mechanics and its Applications*, Elsevier B.V., v. 501, p. 134–140, 2018. ISSN 03784371. [2.2](#)
- GUEDES, E. F. et al. Statistical test for  $\Delta\rho$ DCCA: Methods and data. *Data in Brief*, v. 18, p. 795–798, 2018. ISSN 23523409. [2.2](#)
- GUEDES, E. F.; da Silva Filho, A. M.; ZEBENDE, G. F. Detrended multiple cross-correlation coefficient with sliding windows approach. *Physica A: Statistical Mechanics and its Applications*, Elsevier B.V., v. 574, p. 125990, 2021. ISSN 03784371. [2.2](#)
- 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. [3.2](#)
- HE, H.-d. Multifractal analysis of interactive patterns between meteorological factors and pollutants in urban and rural areas. *Atmospheric Environment*, v. 149, p. 47–54, 2017. ISSN 1352-2310. [2.5](#)
- HU, K. et al. Effect of trends on detrended fluctuation analysis. *Physical Review E*, v. 64, n. 1, p. 011114, jun. 2001. ISSN 1063-651X, 1095-3787. [2.2](#)
- KACS, D.; BROWN, N.; LEE, J. Implementing OpenMP for zig to enable its use in HPC context. In: *The 53rd International Conference on Parallel Processing Workshops*. [S.l.]: ACM, 2024. (ICPP Workshops '24), p. 116–117. [3.2](#)
- KACS, D. et al. Pragma driven shared memory parallelism in Zig by supporting OpenMP loop directives. In: *SC24-w: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*. [S.l.: s.n.], 2024. p. 930–938. [3.2](#)

- KANTELHARDT, J. W. et al. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 2002. [2.3](#)
- 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. [3.2](#)
- KAR, A.; CHATTERJEE, S.; GHOSH, D. Multifractal detrended cross correlation analysis of Land-surface temperature anomalies and Soil radon concentration. *Physica A: Statistical Mechanics and its Applications*, v. 521, p. 236–247, 2019. ISSN 0378-4371. [2.5](#)
- LI, T. et al. Identifying the possible driving mechanisms in Precipitation-Runoff relationships with nonstationary and nonlinear theory approaches. *Journal of Hydrology*, v. 639, p. 131535, 2024. ISSN 0022-1694. [2.5](#)
- LI, Z.; TIAN, Y. Skewed multifractal cross-correlation between price and volume during the COVID-19 pandemic: Evidence from China and European carbon markets. *Applied Energy*, v. 371, p. 123716, 2024. ISSN 0306-2619. [2.5](#)
- LUGARINI, A.; ROMANUS, R.; JUNIOR, W. O. Large Eddy Simulations of Complex Shaped High-Rise Buildings. *SSRN Electronic Journal*, 2024. ISSN 1556-5068. [3.2](#)
- MANDELBROT, B. *Fractals: Form, Chance, and Dimension*. [S.l.]: W. H. Freeman, 1977. (Mathematics Series). ISBN 978-0-7167-0473-7. [2.3](#)
- MANDELBROT, B. B. *The Fractal Geometry of Nature*. 1982. [S.l.]: Henry Holt and Company, 1982. ISBN 978-0-7167-1186-5. [2.3](#)
- NAVEED, K.; REHMAN, N. U. Fully multivariate detrended fluctuation analysis using Mahalanobis norm with application to multivariate signal denoising. *Measurement*, v. 242, p. 116142, jan. 2025. ISSN 02632241. [2.4](#)
- OLIVEIRA, J. B. et al. Paradox between adequate sanitation and rainfall in dengue fever cases. *Science of The Total Environment*, v. 860, p. 160491, 2023. ISSN 0048-9697. [2.5](#)
- PENG, C.-K. et al. Mosaic organization of DNA nucleotides. *Physical Review E*, v. 49, n. 2, p. 1685–1689, fev. 1994. ISSN 1063-651X, 1095-3787. ([document](#)), [2.3](#), [2.1](#)
- PENG, C.-K. et al. Mosaic Organization of DNA Nucleotides. v. 49, n. 2, p. 1685–1689, 1994. [2.2](#)
- PODOBNIK, B. et al. Statistical tests for power-law cross-correlated processes. *Phys. Rev. E*, American Physical Society, v. 84, n. 6, p. 66118, 2011. [2.2](#)
- 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. [2.2](#), [2.2](#)
- QIN, L. et al. A multi-index fusion prediction method for dynamic seismic damage of an arch dam. *Soil Dynamics and Earthquake Engineering*, v. 179, p. 108559, 2024. ISSN 0267-7261. [2.5](#)
- RIBEIRO, F. F. et al. DCCA multi cross-correlation analysis applied on EEG signals to study motor activity (Real/Imaginary). *Biomedical Signal Processing and Control*, v. 103, p. 107419, 2025. ISSN 1746-8094. [3.1](#)

- ROMANUS, R. S. et al. A viable framework for wind load assessments with Large Eddy Simulations. 2023. [3.2](#)
- SANTOS, F. R. et al. Detection of the persistency of the blockages symmetry influence on the multi-scale cross-correlations of the velocity fields in internal turbulent flows in pipelines. *Physica A: Statistical Mechanics and its Applications*, v. 509, p. 294–301, 2018. ISSN 0378-4371. [2.5](#)
- SHADKHOO, S.; JAFARI, G. R. Multifractal detrended cross-correlation analysis of temporal and spatial seismic data. *The European Physical Journal B*, v. 72, n. 4, p. 679–683, dez. 2009. ISSN 1434-6028, 1434-6036. [2.5](#)
- SHEN, C.-h.; LI, C.-l.; SI, Y.-l. A detrended cross-correlation analysis of meteorological and API data in Nanjing, China. *Physica A: Statistical Mechanics and its Applications*, v. 419, p. 417–428, 2015. ISSN 0378-4371. [2.5](#)
- SHI, K. Detrended cross-correlation analysis of temperature, rainfall, PM10 and ambient dioxins in Hong Kong. *Atmospheric Environment*, v. 97, p. 130–135, 2014. ISSN 1352-2310. [2.5](#)
- SHI, K. et al. Detrended cross-correlation analysis of urban traffic congestion and NO<sub>2</sub> concentrations in Chengdu. *Transportation Research Part D: Transport and Environment*, v. 61, p. 165–173, 2018. ISSN 1361-9209. [2.5](#)
- SHIN, K.-H. et al. Neural network and regression methods for optimizations between two meteorological factors. *Physica A: Statistical Mechanics and its Applications*, v. 523, p. 778–796, 2019. ISSN 0378-4371. [2.5](#)
- VASSOLER, R. T.; ZEBENDE, G. F. DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity. *Physica A*, Elsevier B.V., v. 391, n. 7, p. 2438–2443, 2012. ISSN 0378-4371. [2.2](#)
- 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. [3](#)
- XIONG, H.; SHANG, P. Detrended fluctuation analysis of multivariate time series. *Communications in Nonlinear Science and Numerical Simulation*, v. 42, p. 12–21, jan. 2017. ISSN 10075704. [2.4](#)
- YU, D.-L.; LI, W.-J.; ZHOU, Y. Analysis of trends in air temperature at Chinese stations considering the long-range correlation effect. *Physica A: Statistical Mechanics and its Applications*, v. 533, p. 122034, 2019. ISSN 0378-4371. [2.5](#)
- YUAN, N.; FU, Z. Different spatial cross-correlation patterns of temperature records over China: A DCCA study on different time scales. *Physica A: Statistical Mechanics and its Applications*, v. 400, p. 71–79, 2014. ISSN 0378-4371. [2.5](#)
- YUAN, N. et al. Detrended Partial-Cross-Correlation Analysis: A New Method for Analyzing Correlations in Complex System. *Scientific Reports*, v. 5, n. 1, p. 8143, jan. 2015. ISSN 2045-2322. [2.4](#)
- ZEBENDE, G. et al.  $\rho$ DCCA applied between air temperature and relative humidity: An hour/hour view. *Physica A: Statistical Mechanics and its Applications*, v. 494, p. 17–26, 2018. ISSN 0378-4371. [2.5](#)

- ZEBENDE, G.; da Silva, P.; FILHO, A. M. Study of cross-correlation in a self-affine time series of taxi accidents. *Physica A: Statistical Mechanics and its Applications*, v. 390, n. 9, p. 1677–1683, 2011. ISSN 0378-4371. [2.5](#)
- ZEBENDE, G.; FILHO, A. M. Cross-correlation between time series of vehicles and passengers. *Physica A: Statistical Mechanics and its Applications*, v. 388, n. 23, p. 4863–4866, 2009. ISSN 0378-4371. [2.5](#)
- 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. [2.2](#), [2.2](#)
- 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. [2.2](#)
- ZEBENDE, G. F.; SILVA, M. F.; FILHO, A. M. DCCA cross-correlation coefficient differentiation : Theoretical and practical approaches. *Physica A*, v. 392, p. 1756–1761, 2013. [2.2](#)
- ZHANG, C.; NI, Z.; NI, L. Multifractal detrended cross-correlation analysis between PM2.5 and meteorological factors. *Physica A: Statistical Mechanics and its Applications*, v. 438, p. 114–123, 2015. ISSN 0378-4371. [2.5](#)
- ZHOU, W.-X. Multifractal detrended cross-correlation analysis for two nonstationary signals. *Physical Review E*, v. 77, n. 6, p. 066211, jun. 2008. ISSN 1539-3755, 1550-2376. [2.3](#), [3.2](#)

**$\rho_{DCCA}$  e  $DMC_x^2$  : Implementação, Otimização e Aplicações**

Fernando Ferraz Ribeiro

Feira de Santana, BA, Maio de 2025.