



TESTE ESTATÍSTICO DO DCCA CORRELOGRAMO

Basílio Fernandez Fernandez¹ - basilio.fernandez@hotmail.com

Gilney Figueira Zebende² - gfzebende@uefs.br

¹Programa de Pós Grad. em Model. em Ciências da Terra e do Ambiente-Feira de Santana, BA, Brazil

²Universidade Estadual de Feira de Santana, BA, Brazil

Resumo. Neste artigo, através de simulação computacional, testamos estatisticamente o coeficiente de autocorrelação proposto por Zebende e Guedes (2022). Para tanto, utilizamos um pseudo gerador de números aleatórios (RAN4) no intuito de concebermos amostras de séries temporais aleatórias, com o propósito verificarmos a hipótese nula do coeficiente de autocorrelação, ou seja, coeficiente de autocorrelação igual a zero. Aqui, cada amostra terá um tamanho de 10.000 séries temporais para dois tamanhos distintos, $N = 250$ e $N = 500$. Assim, aqui mostraremos os resultados do correlograma para esta simulação em termos de suas médias e desvios padrão, bem como de um mapa gradiente de cores.

Palavras Chave: Séries Temporais, Simulação, Correlograma

1. INTRODUÇÃO

Aqui sabemos que a autocorrelação é a correlação cruzada de uma série temporal com uma cópia defasada dela própria em função desta defasagem. Também, a análise de autocorrelação surge como uma ferramenta matemática e estatística na intenção de se descobrir padrões, como a presença de um sinal periódico obscurecido por um ruído.

Diferentemente da análise de Fourier, este tipo de análise é usada no processamento de sinais no domínio do tempo. Diferentes campos de estudo definem a autocorrelação de maneira diferente e nem todas essas definições são equivalentes. Mas, a mais famosa é aquela que utiliza o coeficiente de Karl Pearson Pearson (1920); Rodgers and Nicewander (1988). Processos de raiz unitária, estacionários de tendência, autorregressivos e de média móvel são formas específicas de processos com autocorrelação Box et al. (2015).

Porém, tais análises clássicas são bem aplicadas em séries temporais estacionária, e na sua grande maioria tais séries são não estacionárias. Para estes casos, de comportamento não estacionário, outras alternativas (métodos) de coeficientes deverão ser aplicadas, para a obtenção de resultados mais realistas. Nesta direção, aqui neste artigo propomos realizar uma análise estatística (simulação) em torno do coeficiente de autocorrelação, ρ_{DCCA}^a , proposto por Zebende and Guedes (2022), verificando assim sua robustez para análise de autocorrelação, resultando ao seu final em um mapa de auto-correlação.

Neste estudo, dividimos este artigo em seções, ou seja, na seção 2. apresentamos a metodologia e os dados simulados para o estudo, na seção 3. temos os resultados da simulação e finalmente na seção 4. concluímos o artigo.

2. METODOLOGIA E DADOS

2.1 Auto-correlação clássica de Pearson

A análise de séries temporais tem sido uma importante metodologia para a ciência investigar o comportamento ao longo do tempo de um objeto de estudo Box et al. (2015); Chatfield and Xing (2019). Logo, a partir de uma dada variável $\{u_t\}$ (temperatura, pluviosidade, radiação, mercado de capitais, etc.), sucessivos valores podem ser medidos ao longo do tempo, formando uma série temporal, que muitas das vezes pode estar correlacionada entre si. Tal propriedade ou característica, pode ser analisada no domínio do tempo ou da frequência, por modelos e métodos estatísticos. O coeficiente proposto por Pearson (1895, 1896) base-a se neste tipo de análise. Assim, a autocorrelação clássica, introduzida pelo coeficiente de Pearson, pode ser representado por $\rho(\tau)$, definido como:

$$\rho(\tau) \equiv \frac{Cov(u_t, u_{t+\tau})}{\sqrt{Var(u_t)Var(u_{t+\tau})}} \quad (1)$$

Aqui, τ representa a defasagem temporal, $Cov(u_t, u_{t+\tau})$ a covariância de u_t com $u_{t+\tau}$ e $Var(u_t)$ a sua variância. Como principal característica, tem-se que $-1 \leq \rho(\tau) \leq 1$, identificativo assim o sinal da autocorrelação em negativo/positivo (antipersistência/persistência).

Porém, para séries temporais não estacionárias (caso usualmente mais encontrado no dia a dia da pesquisa), essa análise proposta por Pearson não é robusta. Para analisar estas séries não estacionárias, métodos mais robustos deveriam ser empregados, tais como exemplo: *Detrended Fluctuation Analysis* - DFA Peng et al. (1994, 1995), *Detrended Cross-Correlation analysis Analysis* - DCCA Podobnik and Stanley (2008) métodos, e o *Detrended Cross-Correlation coefficient*, ρ_{DCCA} Zebende (2011); Kristoufek (2014); Zebende and da Silva-Filho (2018) coeficientes. Logo a seguir, apresentamos o novo coeficiente de autocorrelação baseado em robustos métodos estatísticos, que será aqui testado por uma completa simulação computacional.

2.2 Coeficiente de Autocorrelação sem Tendência - $\rho_{DCCA}^a(\tau, n)$

Considere uma dada série temporal não estacionária, $\{u_t\}$, com $t = 1, 2, 3, \dots, N$ (tamanho total da série). Uma nova série temporal é criada (clonada) com os mesmos elementos da série original, porém com seus valores defasados, por um tempo τ . Logo, o coeficiente de autocorrelação sem tendência é definido segundo Zebende and Guedes (2022) por:

$$\rho_{DCCA}^a(\tau, n) \equiv \frac{F_{\{u_t : u_{t+\tau}\}}^2(n)}{F_{\{u_t\}}(n) F_{\{u_{t+\tau}\}}(n)} \quad (2)$$

Abaixo, definimos as funções de flutuação $F_{\{u_t : u_{t+\tau}\}}^2$, $F_{\{u_t\}}$ e $F_{\{u_{t+\tau}\}}$, no algoritmo de cálculo para o coeficiente ρ_{DCCA}^a :

1. Em primeiro lugar clonamos a série original (que será defasada) e escolhemos uma defasagem temporal τ , com $0 \leq \tau \leq N/2$ (ver Fig.
2. A série temporal original e a série temporal defasada são parcialmente integradas:

$$u_k = \sum_{t=1}^k (u_t - \langle u \rangle) \quad (3)$$

e

$$u_{\tau k} = \sum_{t=1}^k (u_{t+\tau} - \langle u_{\tau} \rangle) \quad (4)$$

com $k = 1, 2, 3, \dots, (N - \tau)$. Aqui $\langle u \rangle$ e $\langle u_{\tau} \rangle$ representam os valores médios de $\{u_t\}$ e $\{u_{t+\tau}\}$, respectivamente;

3. Com os valores de u_k e $u_{\tau k}$ formamos duas novas séries temporárias, que serão divididas em caixas sobrepostas de mesmo tamanho n entre 4 e $(N - n - \tau)$. Cada caixa com $(n + 1)$ valores. Para ambas séries temporais, cada caixa inicia em i e finaliza em $(i + n)$. Uma tendência local (na caixa) é obtida pelo método dos mínimos quadrados (aqui com um ajuste linear) por, $u_i(k)$ e $u_{\tau i}(k)$;
4. A linha de tendência de cada caixa é obtida a partir da diferença entre o valor original e o da série defasada. Desse modo é calculado a covariância dos resíduos pela Eq.5,

$$f_{u,\tau}^2(n, i) = \frac{1}{n + 1} \sum_{k=i}^{i+n} [u_k - u_i(k)] [u_{\tau k} - u_{\tau i}(k)] \quad (5)$$

com $4 \leq n \leq \frac{N-\tau}{4}$;

5. A média dos valores para todas as caixas sobrepostas é calculado a fim de se obter um novo valor de função de covariância (Eq.6),

$$F_{\{u_t : u_{t+\tau}\}}^2(n) = \frac{1}{N - n - \tau} \sum_{i=1}^{N-n-\tau} f_{u,\tau}^2(n, i); \quad (6)$$

6. Por fim, o coeficiente de autocorrelação, com o tempo de defasagem τ e a escala temporal n , é definido por:

$$\rho_{DCCA}^a(\tau, n) \equiv \frac{F_{\{u_t : u_{t+\tau}\}}^2(n)}{F_{\{u_t\}}(n) F_{\{u_{t+\tau}\}}(n)} \quad (7)$$

Um exemplo para o cálculo do coeficiente de correlação cruzada, ρ_{DCCA} , proposto por Zebende (2011) foi reproduzido em um vídeo no YouTube (veja *link* abaixo):

<https://www.youtube.com/watch?v=RQL7Db74yG0&t=216s>

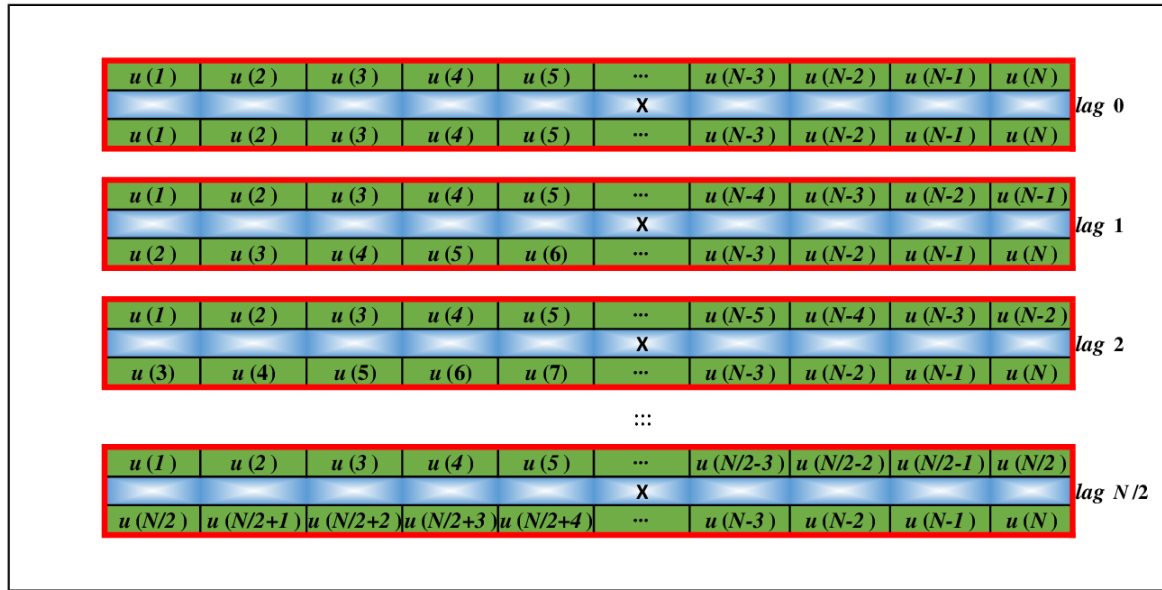


Figura 1: Diagrama para as defasagens temporais, onde lag representa a defasagem temporal da série original em relação a série clonada.

Aqui, neste breve vídeo, alguns passos do algoritmo apresentado logo acima deste texto, pode ser claramente exemplificado.

O coeficiente, $\rho_{DCCA}^a(\tau, n)$ depende da defasagem temporal τ e da escala temporal n (tamanho da caixa), e seus valores variam entre $-1 \leq \rho_{DCCA}^a \leq 1$. O valor de $\rho_{DCCA}^a = 0$ significa que não existe autocorrelação na série, caso contrário ρ_{DCCA}^a se divide em valores positivos e negativos. Assim, o valor de ρ_{DCCA}^a pode ser classificado tendo em consideração o seu nível de intensidade, como por exemplo de acordo com a Tabela 1, veja Zebende et al. (2019); Podobnik et al. (2011). Fechando assim a parte teórica, logo abaixo destacamos a

Tabela 1: Níveis de Correlação Cruzada.

Fraca	ρ_{DCCA}^a Média	Forte
$\pm 0.000 \mapsto \pm 0.283$	$\pm 0.403 \mapsto \pm 0.666$	$\pm 0.666 \mapsto \pm 1.000$

obtenção de nossos dados a partir de simulações computacional.

2.3 Dados Simulados

Afim de testarmos a hipótese nula para o coeficiente, de semelhante modo utilizado em Guedes et al. (2018); Podobnik et al. (2011), ou seja, $\rho_{DCCA}^a = 0$, dado na Eq.7, simulamos duas amostras com 10.000 séries temporais aleatórias cada, cujos valores variam entre 0 e 1 (inclusive). Estas duas amostras foram feitas para que nós testássemos o efeito de tamanho da séries temporais, ou seja, uma amostra tem tamanho $N = 250$ e a outra tem tamanho $N = 500$. Aqui utilizamos a rotina **RAN4**, que gera excelentes números aleatórios a partir de uma dada semente aleatória. Ver tal implementação no capítulo 7, listado em Vetterling et al. (1999). Portanto, para cada série temporal aleatória (nas suas duas amostras), calculamos os valores de ρ_{DCCA}^a como funções de τ e n , obtendo assim um conjunto de resultados que foram modelados

posteriormente, isto no sentido de obtermos seus valores médios

$$\langle \rho_{DCCA}^a(n) \rangle \quad (8)$$

e os respectivos desvios padrões

$$\langle sd(\rho_{DCCA}^a(n)) \rangle \quad (9)$$

como veremos logo próxima seção com os resultados.

3. RESULTADOS

Portanto, para cada um dos 10.000 arquivos, de ρ_{DCCA}^a como função da defasagem temporal τ (lag) e da escala temporal n , calculamos o valor médio e seu respectivo desvio padrão, equações 8 e 9. Logo na Fig.2 apresentamos $\langle sd(\rho_{DCCA}^a(n)) \rangle$ para a amostra de tamanho $N = 250$.

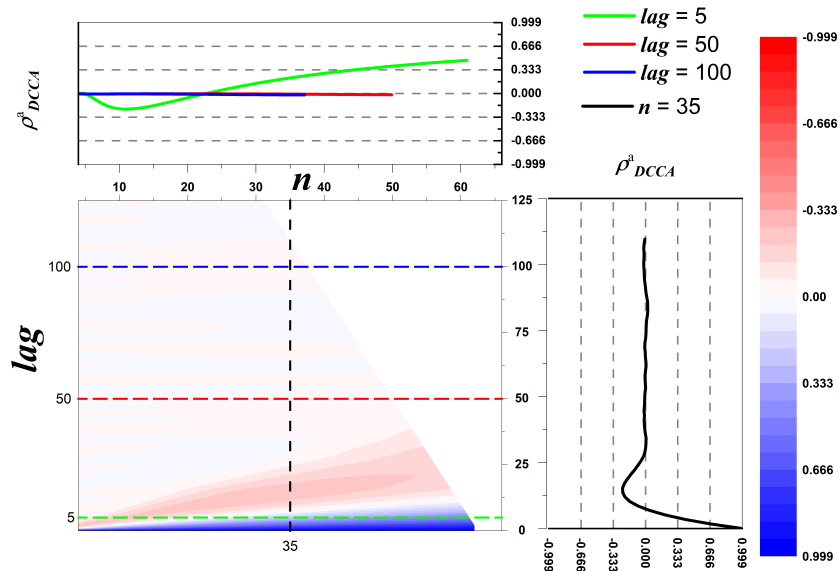


Figura 2: Valores médios de ρ_{DCCA}^a calculados para uma amostra de 10.000 séries aleatórias com $N = 250$ em função da defasagem temporal, lag (eixo vertical), e da escala temporal, n (eixo horizontal). Os painéis, acima e do lado direito do Mapa, são os valores para ρ_{DCCA}^a fixos em $n = 35$ (linha pontilhada preta), ou na defasagem temporal $lag = 5 : 50 : 100$ (linhas pontilhadas verde, vermelha e azul).

Colocamos os resultados do coeficiente de autocorrelação em um mapa de cores, onde: o valor azul significa perfeita autocorrelação, pois $\rho_{DCCA}^a = 1$, o valor vermelho significa a anti-perfeita autocorrelação, pois $\rho_{DCCA}^a = -1$ e o branco a ausência de autocorrelação, pois $\rho_{DCCA}^a = 0$. Como estamos testando a hipótese nula, deveríamos ter na teoria $\rho_{DCCA}^a = 0$ para todas as defasagens e escalas temporais, porém na prática isto não ocorre. Veja que para pequenas defasagens e grandes escalas temporais $\rho_{DCCA}^a \rightarrow 1$, mas no geral $\rho_{DCCA}^a \simeq 0$. Também, além do valor médio, calculamos o seu desvio padrão, para verificarmos a dispersão dos valores de ρ_{DCCA}^a com relação à sua média. Os resultados para o desvio padrão se encontram na Fig. 3.

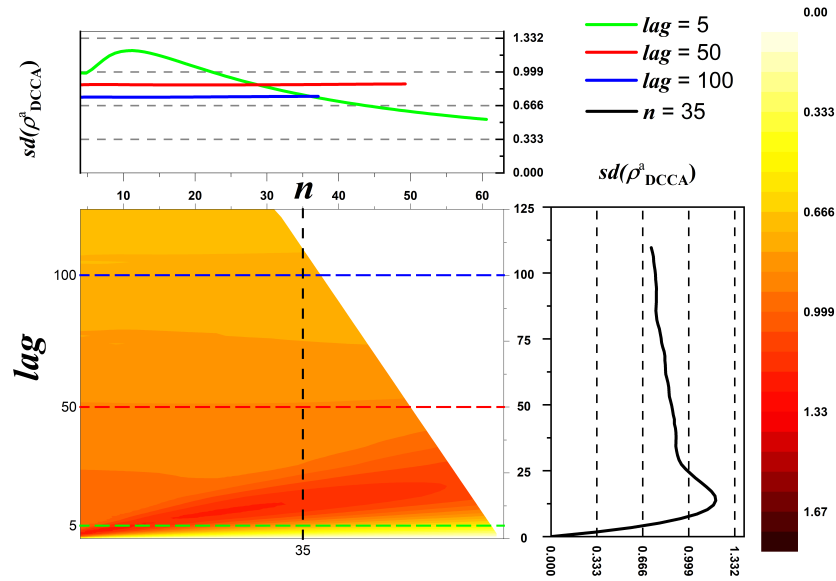


Figura 3: Desvio padrão de ρ_{DCCA}^a para uma amostra de 10.000 séries aleatórias com $N = 250$ em função da defasagem temporal, lag (eixo vertical), e da escala temporal, n (eixo horizontal). Os painéis, acima e do lado direito do Mapa, são os valores para ρ_{DCCA}^a fixos em $n = 35$ (linha pontilhada preta), ou na defasagem temporal $lag = 5 : 50 : 100$ (linhas pontilhadas verde, vermelha e azul).

Neste figura é possível identificar os locais onde há maior dispersão em relação aos valores médios (parte mais vermelha). Agora, para verificar o efeito de tamanho na estatística, calculamos os valores de ρ_{DCCA}^a para a amostra de tamanho $N = 500$. Os resultados para o valor médio podem ser vistos na Fig.4.

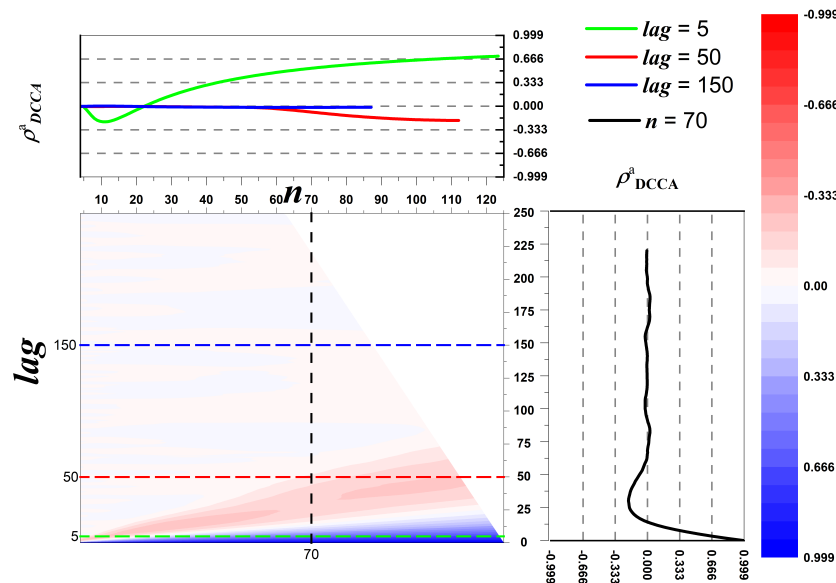


Figura 4: Valores médios de ρ_{DCCA}^a calculados para uma amostra de 10.000 séries aleatórias com $N = 500$ em função da defasagem temporal, lag (eixo vertical), e da escala temporal, n (eixo horizontal). Os painéis, acima e do lado direito do Mapa, são os valores para ρ_{DCCA}^a fixos em $n = 35$ (linha pontilhada preta), ou na defasagem temporal $lag = 5 : 50 : 100$ (linhas pontilhadas verde, vermelha e azul).

Nesta figura vemos que há uma maior parte branca, evidenciando que na sua maior parte o valor de $\rho_{DCCA}^a \simeq 0$, confirmando a hipótese nula do coeficiente para esta simulação. Na Fig.5 temos os resultados para o desvio padrão de ρ_{DCCA}^a .

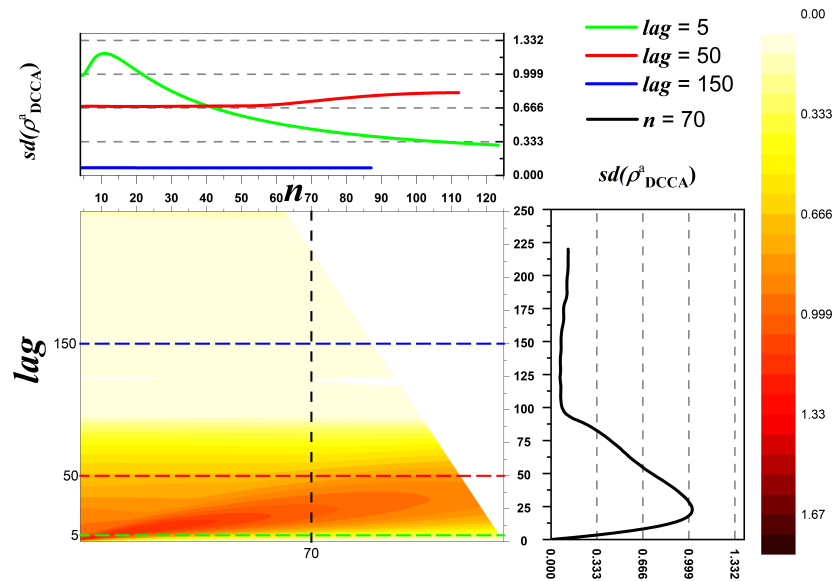


Figura 5: Desvio padrão de ρ_{DCCA}^a calculados para uma amostra de 10.000 séries aleatórias com $N = 500$ em função da defasagem temporal, lag (eixo vertical), e da escala temporal, n (eixo horizontal). Os painéis, acima e do lado direito do Mapa, são os valores para ρ_{DCCA}^a fixos em $n = 70$ (linha pontilhada preta), ou na defasagem temporal $lag = 5 : 50 : 150$ (linhas pontilhadas verde, vermelha e azul).

Que também mostra valores mais perto do branco, ou seja, menos desvio do seu valor médio.

4. CONCLUSÃO

Neste artigo propomos um teste estatístico no coeficiente de autocorrelação, ρ_{DCCA}^a , isto através da hipótese nula do mesmo. Para darmos seguimento a esta proposta, uma simulação computacional foi desenvolvida para se montar duas amostras de tamanhos diferentes (com $N = 250$ e $N = 500$) de séries temporais aleatórias, cada amostra contendo 10.000 séries. Como resultado, calculamos os valores médios e o desvio padrão do ρ_{DCCA}^a calculado para as amostras. Tais resultados foram apresentados em mapas de cores. Identificamos que a hipótese nula foi bem apresentada e robusta nos seus resultados para a amostra de tamanho $N = 500$, ao passo que a amostra de tamanho $N = 250$ ficou deficiente para pequenas defasagens e grandes escalas temporais.

Agradecimentos

B. F. Fernandez agradece à FUNTITEC (Fundação de Tecnologia da Informação Telec. Egberto Costa) e G. F. Zebende agradece ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) (Grant 310136/2020-2). Também agradecemos ao Dr. Paulo Murilo Castro de Oliveira, grande Físico Estatístico (*in memoriam*).

Referências

- George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- Chris Chatfield and Haipeng Xing. *The analysis of time series: an introduction with R*. CRC press, 2019.
- EF Guedes, AA Brito, FM Oliveira Filho, BF Fernandez, APN de Castro, AM da Silva Filho, and GF Zebende. Statistical test for $\Delta\rho_{DCCA}$ cross-correlation coefficient. *Physica A: Statistical Mechanics and its Applications*, 501:134–140, 2018.
- Ladislav Kristoufek. Measuring correlations between non-stationary series with dcca coefficient. *Physica A: Statistical Mechanics and its Applications*, 402:291–298, 2014.
- Karl Pearson. Vii. note on regression and inheritance in the case of two parents. *proceedings of the royal society of London*, 58(347-352):240–242, 1895.
- Karl Pearson. Vii. mathematical contributions to the theory of evolution. iii. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, (187):253–318, 1896.
- Karl Pearson. Notes on the history of correlation. *Biometrika*, 13(1):25–45, 1920.
- C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of DNA nucleotides. *Phys. Rev. E*, 49:1685–1689, 1994.
- C-K Peng, SV Buldyrev, AL Goldberger, S Havlin, RN Mantegna, M Simons, and HE Stanley. Statistical properties of dna sequences. *Physica A: Statistical Mechanics and its Applications*, 221(1-3):180–192, 1995.
- B. Podobnik and H. E. Stanley. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Phys. Rev. Lett*, 100:1–4, 2008.
- Boris Podobnik, Zhi-Qiang Jiang, Wei-Xing Zhou, and H Eugene Stanley. Statistical tests for power-law cross-correlated processes. *Physical Review E*, 84(6):066118, 2011.

- Joseph Lee Rodgers and W. Alan Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988. doi: 10.1080/00031305.1988.10475524.
- William T Vetterling, Saul A Teukolsky, William H Press, and Brian P Flannery. *Numerical recipes in C: the art of scientific computing*, volume 1. Cambridge university press, 2 edition, 1999. ISBN 0-521-43108-5.
- G. F. Zebende. DCCA cross-correlation coefficient: Quantifying level of cross-correlation. *Physica A: Statistical Mechanics and its Applications*, 390(4):614–618, 2011.
- G.F. Zebende and A.M. da Silva-Filho. Detrended multiple cross-correlation coefficient. *Physica A*, 510:91 – 97, 2018. ISSN 0378-4371.
- GF Zebende and EF Guedes. Detrended correlogram method for non-stationary time-series analysis. *Fluctuation and Noise Letters*, 21(02):2250012, 2022.
- GF Zebende, AA Brito, and AP Castro. DCCA cross-correlation analysis in time-series with removed parts. *Physica A: Statistical Mechanics and its Applications*, page 123472, 2019.

STATISTICAL TEST FOR DCCA CORRELOGRAM

Abstract. *In this paper, through computer simulation, we statistically test the auto-correlation coefficient proposed by Zebende and Guedes (2022). For that, we used a pseudo-random number generator (RAN4) to design samples of random time-series, to verify the null hypothesis of the auto-correlation coefficient, that is, the auto-correlation coefficient equal to zero. Here, each sample will have a size of 10,000 time-series for two different sizes, $N = 250$ and $N = 500$. Thus, here we will show the correlogram results for this simulation in terms of their means and standard deviations, as well as a color gradient map.*

Keywords: *Time Series, Simulation, Corrgram*