



CENTRO UNIVERSITÁRIO SENAI CIMATEC

PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM

COMPUTACIONAL E TECNOLOGIA INDUSTRIAL

Doutorado em Modelagem Computacional e Tecnologia Industrial

Tese de Doutorado

Modelagem Computacional de Auto-correlações e
Correlações Cruzadas Aplicada em Variáveis
Meteorológicas e Simuladas

Apresentada por: **Andrea de Almeida Brito**

Orientador: Prof. Dr. Gilney Figueira Zebende

Co-orientador: Prof. Dr. Ivan Costa da Cunha Lima

29 de janeiro de 2021

Andrea de Almeida Brito

**Modelagem Computacional de Auto-correlações e
Correlações Cruzadas Aplicada em Variáveis
Meteorológicas e Simuladas**

Tese de Doutorado apresentada ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial, Curso de Doutorado em Modelagem Computacional e Tecnologia Industrial do Centro Universitário SENAI CIMATEC, como requisito parcial para a obtenção do título de Doutora.

Área de conhecimento: Sistemas Complexos

Orientador: Prof. Dr. Gilney Figueira Zebende

Universidade Estadual de Feira de Santana

Co-Orientador: Prof. Dr. Ivan Costa da Cunha Lima

Centro Universitário SENAI CIMATEC

Salvador, BA
Centro Universitário SENAI CIMATEC
2021

Ficha catalográfica elaborada pela Biblioteca do Centro Universitário SENAI CIMATEC

B862m Brito, Andrea de Almeida

Modelagem Computacional de Auto-correlação e Correlações Cruzadas Aplicada em Variáveis Meteorológicas e Simuladas/ Andrea de Almeida Brito. – Salvador, 2021.

117 f.: il. Color.

Orientador: Prof. Dr. Ivan Costa da Cunha Lima.

Tese (Doutorado em Modelagem Computacional) – Programa de Pós-Graduação, Centro Universitário SENAI CIMATEC, Salvador, 2021.
Inclui referências.

1. Modelagem Computacional. 2. Séries Temporais. 3. Variáveis Meteorológicas. I. Centro Universitário SENAI CIMATEC. II. Lima, Ivan Costa da Cunha. III. Título.

CDD 003.3



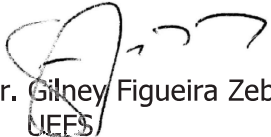
Centro Universitário SENAI CIMATEC


Doutorado em Modelagem Computacional e Tecnologia Industrial

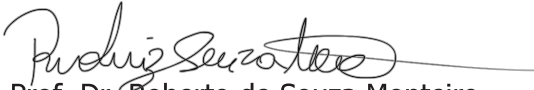
A Banca Examinadora, constituída pelos professores abaixo listados, leu e aprovou a Tese de doutorado, intitulada "**Modelagem Computacional de Autocorrelação e Correlações Cruzadas Aplicada em Variáveis Meteorológicas e Simuladas**", apresentada no dia 29 de janeiro de 2021, como parte dos requisitos necessários para a obtenção do Título de Doutor em Modelagem Computacional e Tecnologia Industrial.

Ivan da Cunha Lima Digitally signed by Ivan da Cunha Lima
Date: 2021.03.15 17:33:44 -03'00'

Orientador: Prof. Dr. Ivan Costa da Cunha Lima
SENAI CIMATEC


Coorientador: Prof. Dr. Gilney Figueira Zebende

UEFS


Membro Interno: Prof. Dr. Davidson Martins Moreira

SENAI CIMATEC

Membro Interno: Prof. Dr. Roberto de Souza Monteiro

SENAI CIMATEC

Membro Externo: Prof.^a Dr.^a Olga de Castro Vilela

UFPE

Membro Externo: Prof. Dr. Alberto Sebastião de Arruda

UFMT

Membro Externo: Prof. Dr. Joseph Nee Anyah Yartey

UFBA

Dedico esta Tese à **Deus** e aos meus pais,
Gizélia Maria de Almeida Brito e **Antonio Brito**.

Ainda que tivesse o dom da profecia,
e conhecesse todos os mistérios e toda a ciência
e ainda que tivesse toda a fé,
de maneira tal que transportasse os montes,
e não tivesse **AMOR**, nada seria.
(1 Coríntios 13:1-2)

Agradecimentos

Agradeço em primeiro lugar à **DEUS** pelo dom da vida e pela sabedoria para chegar até aqui e com a certeza de que seguirei ir mais adiante.

Aos meus pais, irmãos, sobrinhos e cunhados, em especial **Adriana Almeida Brito**, por estar presente em minha vida o tempo todo.

Ao meu professor, Orientador, amigo **Prof. Dr. Gilney Figueira Zebende**, pelo seu carinho, atenção, ensinamentos e cuidados.

Ao **Prof. Dr. Ivan Costa da Cunha Lima**, por acreditar e confiar em mim ao longo desta pesquisa, meu muito obrigado.

Ao **Prof. Dr. Alex Álisson Bandeira Santos**, pelo apoio científico e confiança.

Ao grupo de pesquisa *ORIENTANDOS de ZEBENDE*, em especial aos amigos **Prof. Dr. Arleys Pereira Nunes de Castro**, **Prof. Msc. Basílio Fernandez Fernandez**, **Prof. Dr. Florêncio Mendes Oliveira Filho**, **Prof. Dr. Aloísio Machado da Silva Filho**, **Prof. Dr. Everaldo Freitas Guedes** e ao querido **Prof. Msc. Moíses Domingos Namila da Costa** pela parceria, apoio e incentivo.

Ao meu amigo **Prof. Dr. Urbano Uelligton Secundes**, parceiro e companheiro nas horas difíceis, complexas e complicadas. Amigo de todas as horas.

Ao meu amigo **Prof. Msc. Cesar Andrey Gomes Ferreira**, companheiro, sempre disposto a ouvir meus desabafos nos momentos de fúrias porque doutoranda também surta.

Ao amigo **Prof. Dr. Paulo Henrique Farias Xavier** de longas datas, desde a graduação de Matemática na UNIFACS, passando pela pós-graduação em Educação Matemática na UCSAL, no Mestrado e atualmente Doutorado SENAI CIMATEC.

À querida **Julie Miranda dos Santos** pela presteza e eficiência ao longo desses anos.

Ao querido **Msc. Heráclio Alves de Araújo**, meteorologista do INMET (DF), pelas contribuições a esta pesquisa.

Ao **INMET**, pela disponibilização da base de dados das estações meteorológicas que foram modeladas nesta pesquisa.

À **FAPESB**, pelo suporte científico financeiro.

Ao **IFBA**, pelo apoio na realização deste projeto científico, em especial ao **Prof. Arleno José de Jesus**.

E, por fim, a todos que de algum modo contribuíram para a realização deste trabalho.

Salvador, Brasil
29 de janeiro de 2021

Andrea de Almeida Brito

Resumo

Nas últimas décadas, diversos cientistas, dentre eles, Físicos, Matemáticos e Estatísticos, têm dado atenção especial ao estudo e a análise de sistemas complexos, de comportamento não linear. Geralmente tais sistemas são caracterizados por meio de variáveis que são medidas ou simuladas ao longo tempo, formando desta maneira séries temporais que serão analisadas no intuito de mensurar eventuais propriedades de escalas. Esse grande volume de dados e informações têm sido produzido em diversas áreas do conhecimento, onde a aplicação e o desenvolvimento de métodos estatísticos inovadores são muito bem vindos. Na literatura existem diversos métodos de análise de séries temporais, que objetivam a quantificação das auto-correlações e correlações cruzadas.

Neste sentido, nesta Tese estudamos e aplicamos os métodos **DFA** e **DCCA**, bem como os coeficientes de correlações cruzadas ρ_{DCCA} e DMC_x^2 , para o estudo de séries temporais de variáveis meteorológicas (temperatura do ar, umidade relativa do ar e radiação solar) e de dados simulados produzidos por processos estatísticos bem definidos.

Inicialmente, fizemos uma análise entre duas importantes variáveis meteorológicas (temperatura do ar e a umidade relativa do ar) e com esta análise propusemos um novo mapa de correlações. Em seguida, aplicamos a generalização do coeficiente ρ_{DCCA} para o estudo de múltiplas séries temporais meteorológicas, tendo a radiação solar como base de estudo, e desta forma, propusemos uma nova modelagem para captação de energia solar. Para referendarmos nossa análise, simulamos computacionalmente séries temporais com partes removidas (com memória de longo alcance), no intuito da verificação da robustez do coeficiente ρ_{DCCA} , para cortes e remoções. Desta simulação, observamos que o ρ_{DCCA} mostrou-se eficiente para até o limite de 50% de remoção da série temporal original.

Por fim, destacamos que novos trabalhos já estão finalizados e submetidos para publicação e outros estão em via de finalização, mostrando que tal projeto de Tese terá destacada continuidade.

Palavras chaves: Séries Temporais, Variáveis Meteorológicas, **DFA**, **DCCA**, ρ_{DCCA} , DMC_x^2 .

Abstract

In the last decades, several scientists, such as Physicists, Mathematicians and Statisticians, have given special attention to the study and analysis of complex systems, with non-linear behavior. Generally, such systems are characterized by means of variables that are measured or simulated over time, thus forming time series that will be analyzed in order to measure any scale properties. This large volume of data and information has been produced in several areas of knowledge, where the application and development of innovative statistical methods are very welcome. There are several methods of analyzing time series in the literature, which aim to quantify the auto-correlations and cross-correlations.

In this sense, in this Thesis we study and apply **DFA** and **DCCA** the methods, as well as, the cross-correlation coefficients ρ_{DCCA} and DMC_x^2 , for time series study of meteorological variables (air temperature, relative humidity and solar radiation) and also simulated data, here produced by statistical processes.

Initially, we performed an analysis between two important meteorological variables (air temperature and relative humidity) and with this analysis a new map of cross-correlations was proposed. Then, we applied the ρ_{DCCA} generalization for study multiple meteorological time series, using solar radiation as the base, and in this way, a new model for capturing solar energy was defined. To endorse our analysis, we computationally simulate time series with removed parts (with long-range memory), in order to verify the robustness of the ρ_{DCCA} for cuts and removals parts. From this simulation, we observed that the ρ_{DCCA} proved to be efficient up to the limit of 50% of removal parts of the original time series.

Finally, we highlight that new works are already finalized and submitted for publication and others are in the process, showing that such a Thesis project will have outstanding continuity.

keywords: Time Series, Weather Variables, **DFA**, **DCCA**, ρ_{DCCA} , DMC_x^2 .

Sumário

1	INTRODUÇÃO	1
2	FUNDAMENTAÇÃO TEÓRICA	7
2.1	Variáveis Meteorológicas	7
2.1.1	Dados utilizados e área de estudo	12
2.1.2	Descrição das variáveis meteorológicas	13
2.2	Variáveis Simuladas - Processo ARFIMA	16
2.3	Séries Temporais	18
2.3.1	Objetivo da Análise de Séries Temporais	19
2.3.2	Estacionariedade	21
2.3.3	Processos Estocásticos	22
2.3.3.1	Tendências em Séries Temporais	24
2.3.3.2	Sazonalidade	25
2.4	Coefficiente de Correlação de Pearson	26
2.5	Outros Métodos de Análise de Séries Temporais Estacionárias	28
2.5.1	Função de Auto-correlação - FAC	28
2.5.2	Análise R/S	29
2.6	Métodos Estatísticos para Análise de Séries Temporais não-estacionárias	31
2.6.1	Detrended Fluctuation Analysis - DFA	32
2.6.2	Detrended Cross-Correlation Analysis - DCCA	34
2.6.3	Coefficiente de Correlação Cruzada ρ_{DCCA}	35
2.6.4	$\Delta\rho_{DCCA}$	36
2.6.5	Coefficiente de Correlação Cruzada Múltiplo - DMC_x^2	36
3	PUBLICAÇÕES	38
3.1	ρ_{DCCA} aplicado entre a temperatura do ar e a umidade relativa: uma visão hora/hora	38
3.2	DMC_x^2 aplicado à radiação solar global, temperatura do ar e umidade relativa do ar	58
3.3	Análise de Correlação Cruzada em séries temporais com partes removidas	70
4	CONSIDERAÇÕES FINAIS	85
4.1	Conclusões	85
4.2	Perspectivas Futuras	86
	REFERÊNCIAS BIBLIOGRÁFICAS	87
A	APÊNDICE I	93
A.1	Registro de Software	93
A.2	Capítulos de Livro	98
B	APÊNDICE II	99
B.1	Outras Publicações	99

Lista de Tabelas

2.1	Descrição das estações meteorológicas.	12
2.2	Parâmetros para o Processo Simulado de Série Temporal.	17
2.3	Propriedade do expoente de HURST	31
2.4	Expoente de auto-correlação α_{DFA} e suas característica.	33
2.5	ρ_{DCCA} , coeficiente de correlação cruzada de Zebende, e suas propriedades. . .	36
2.6	Possíveis intervalos para o Coeficiente de Correlação Cruzada ρ_{DCCA}	36
2.7	DMC_x^2 , coeficiente de correlação cruzada múltiplo e possíveis níveis de correlação cruzada.	37

Lista de Figuras

1.1	Dados atmosféricos de CO ₂ e temperatura derivados de Medições do núcleo de gelo antártico	5
2.1	Radiação solar global média anual (MJm ⁻² por dia)	14
2.2	Variação horária da temperatura do ar e umidade relativa do ar, para a estação meteorológica de Salvador na data de 24/06/2019	16
2.3	Algoritmo de corte e remoção de partes das séries temporais originais geradas por um processo ARFIMA com $N = 10.000$ (canto superior): $\{x_i\}$ é o ARFIMA 01 com $\delta_1 = 0,1$ e $\{y_i\}$ é o ARFIMA 02 com $\delta_2 = 0,4$. A força das correlações cruzadas, $W = 0,5$, com $\varepsilon_{x_i} \neq \varepsilon_{y_i}$	17
2.4	Sistema dinâmico	20
2.5	Interpretação de um processo estocástico como uma família de variáveis aleatórias	22
2.6	Um processo estocástico interpretado como uma família de trajetórias	23
2.7	Diagrama de Dispersão referente a estação de Salvador (BA) em 24/06/2019. (a) Umidade Relativa do Ar \times Temperatura do Ar, (b) Temperatura do Ar \times Temperatura do Ar e (c) Umidade Relativa do Ar (aleatório) \times Temperatura do Ar	26
2.8	Modelo do reservatório de Hurst para o cálculo do $R/S(\tau)$	30
2.9	Modelo do Reservatório de Hurst (1951)	31
2.10	Algoritmo para o cálculo da Função de Flutuação DFA	33

Lista de Siglas

AR	<i>Modelo Auto-regressivo</i>
ARFIRMA ..	<i>Modelo Auto-regressivo Fracionário Integrado de Média Móvel</i>
ARIMA	<i>Modelo Auto-regressivo Integrado e de Média Móvel</i>
ARMA	<i>Modelo Auto-regressivo e de Média Móvel</i>
ASAS	<i>Alta Pressão do Atlântico Sul</i>
AST	<i>Análise de Séries Temporais</i>
DCCA	<i>Detrended Cross-Correlation Analysis</i>
DFA	<i>Detrended Fluctuation Analysis</i>
FAC	<i>Função de Auto-correlação</i>
GCM	<i>Modelos de Circulação Geral</i>
INEMA	<i>Instituto do Meio Ambiente e Recursos Hídricos</i>
INMET	<i>Instituto Nacional de Meteorologia</i>
INPE	<i>Instituto Nacional de Pesquisas Espaciais</i>
INPI	<i>Instituto Nacional da Propriedade Industrial</i>
MA	<i>Modelo de Média Móvel</i>
OMM	<i>Organização Meteorológica Mundial</i>
PDF	<i>Função de Distribuição de Probabilidade</i>
PNT	<i>Previsão Numérica do Tempo</i>
PPGM	<i>Programa de Pós-graduação em Modelagem em Ciências da Terra e do Ambiente</i>
PPGMCTI ..	<i>Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial</i>
SBF	<i>Sociedade Brasileira de Física</i>
ST	<i>Série Temporal</i>
v.a.	<i>variável aleatória</i>
VCAN	<i>Vórtices Ciclônicos de Altos Níveis</i>
ZCAS	<i>Zona de Convergência do Atlântico Sul</i>
ZCIT	<i>Zona de Convergência Intertropical</i>

INTRODUÇÃO

Nas últimas décadas, diversos cientistas, dentre eles Físicos, Matemáticos e Estatísticos, têm dado atenção especial ao estudo e análise de sistemas complexos de comportamento não linear. Em geral, esses sistemas são caracterizados por meio de variáveis que são medidas ou simuladas ao longo do tempo, formando desta maneira séries temporais, que serão analisadas no intuito de mensurar eventuais propriedades de escala. Entretanto, segundo Filho (2014), devido ao grande volume de dados e informações produzidos nas diversas áreas científicas, faz-se necessário a modelagem e o desenvolvimento de métodos estatísticos no auxílio da tabulação e estatística dos mesmos.

Na literatura existem vários métodos estatísticos que objetivam quantificar as auto-correlações e correlações cruzadas entre séries temporais, tais como: o expoente α_{DFA} (PENG et al., 1994), o expoente λ_{DCCA} (PODOBNIK; STANLEY, 2008), o coeficiente de Pearson (STIGLER, 1989) e (STANTON, 2001), o coeficiente parcial sem tendência (YUAN et al., 2015), o coeficiente de Spearman (SPERMAN, 1904), o coeficiente ρ_{DCCA} (ZEBENDE, 2011), o coeficiente $DMCA$ (KRISTOUFEK, 2014), o coeficiente DMC_x^2 (ZEBENDE; FILHO, 2018), dentre outros. Nesta Tese de Doutorado, limitaremos nossa pesquisa ao estudo da auto-correlação e da correlação cruzada para a análise de séries temporais pelos métodos DFA, DCCA e pelos coeficientes de correlação cruzada, ρ_{DCCA} e DMC_x^2 .

Assim, para o estudo da auto-correlação de séries temporais, em 1994 foi publicado o *Detrended Fluctuation Analysis - DFA*, proposto por Peng et al. (1994) para a análise das flutuações sem tendências. O método **DFA** é utilizado na investigação e identificação de auto-correlações de longo alcance em série temporais não estacionárias. Este método possui citações em diversas áreas científicas, tais como: no estudo de séries cronológicas de acidentes (ZEBENDE; SILVA; FILHO, 2011) e (ZEBENDE; SILVA; FILHO, 2011), na computação científica (ZEBENDE, 1999) e (ZEBENDE; OLIVIERA; PENNA, 1998), no estudo de indicadores de criminalidade (FILHO, 2009), (ZEBENDE; SILVA; FILHO, 2011) e (ZEBENDE; FILHO; SILVA, 2014), nos sistemas astrofísicos (FERNANDEZ, 2014) e (ZEBENDE; FERNANDEZ; PEREIRA, 2017), na análise de atividade cerebral estimulada pela leitura (cognição), realizada por indivíduos usando um aparelho de EEG (ZEBENDE; FILHO; CRUZ, 2017) e (FILHO; CRUZ; ZEBENDE, 2019), na análise de cortes e remoções de séries temporais (ZEBENDE; BRITO; CASTRO, 2020), dentre outros.

Para o estudo da correlação cruzada entre séries temporais, em 2008 surge então o método

DCCA - *Detrended Cross-Correlation Analysis*, proposto por [Podobnik e Stanley \(2008\)](#) e que trata-se da generalização do método **DFA**. O **DCCA** tem como objetivo de investigar e identificar correlações cruzadas de longo alcance entre duas séries temporais distintas e com o mesmo número de pontos. Diversas áreas são encontradas na literatura utilizando o **DCCA** como por exemplo: na climatologia ([PODOBNIK; STANLEY, 2008](#)), ([VASSOLER; ZEBENDE, 2012](#)), ([VASSOLER, 2012](#)) e ([BRITO; MACHADO; ZEBENDE, 2015](#)), nos movimentos complexos na deglutição de um indivíduo ([CASTRO, 2012](#)), em séries de veículos e passageiros do sistema *Ferry-Boat* da cidade de Salvador/BA ([ZEBENDE; FILHO, 2009](#)), nos indicadores de criminalidade e violência ([JUNIOR et al., 2008.](#)) e ([FILHO, 2009](#)), em série temporal de acidentes de Táxi ([ZEBENDE; SILVA; MACHADO, 2011](#)), no teste de aleatoriedade de séries temporais simuladas ([PODOBNIK et al., 2011](#)), entre outros.

É sabido que, o **DCCA** mensura se há ou não a correlação cruzada de longo alcance entre duas séries temporais distintas através do seu expoente λ_{DCCA} , porém o mesmo não quantifica o nível de correlação entre as referidas séries. Daí surge a pergunta: como quantificar e mensurar o nível de correlação cruzada entre séries temporais distintas?

Partindo deste questionamento, surge então, em 2011 o coeficiente ρ_{DCCA} , também conhecido como **Coeficiente de ZEBENDE**, proposto por [Zebende \(2011\)](#), com o objetivo de quantificar e mensurar o nível de correlação cruzada entre duas séries temporais distintas.

O coeficiente ρ_{DCCA} tem despertado o interesse de muitos pesquisadores, que por sua vez provaram a sua robustez na análise de séries temporais especialmente se comparado com o coeficiente de Pearson, ver ([KRISTOUFEK, 2014](#)). Para comprovar a robustez, eficiência e eficácia do coeficiente ρ_{DCCA} , em 2011 [Podobnik et al. \(2011\)](#), propuseram um teste estatístico com o objetivo de quantificar a significância das correlações cruzadas entre diferentes séries temporais.

Logo, a análise das correlações cruzadas pelo coeficiente ρ_{DCCA} , tornou-se um tópico frequente nos últimos anos, em diversas áreas, tais como: na meteorologia ([VASSOLER, 2012](#)) e ([ZEBENDE et al., 2018](#)), na criminalidade ([OLIVEIRA; GUEDES; ZEBENDE, 2020](#)), na dinâmica de fluídos ([BRITO et al., 2018](#)) e ([SANTOS et al., 2018](#)), (ver Apêndice B). Outros trabalhos poderão ser consultados através do link abaixo:

<https://scholar.google.com/scholar?cites=6351233990136499122&as_sdt=2005&scioldt=0,5&hl=en>

O coeficiente ρ_{DCCA} também tem sido uma ferramenta bastante aplicada na economia e finanças nos últimos anos e segundo [Guedes \(2018\)](#), a propagação das crises financeiras tem sido um tema que tem despertado o interesse de pesquisadores, formuladores de políticas e órgãos reguladores e também de investidores, uma vez que desequilíbrios

iniciados em um país podem ser sentidos em outros países. Dessa forma, a econofísica introduziu os métodos utilizados por físicos estatísticos para o estudo econômico, o que trouxe uma nova visão aos problemas de séries temporais, o que poderá possibilitar a criação de cenários em curvas que também não possuem as características de previsibilidade, ver (PEREIRA; SILVA; PEREIRA, 2017). Dentro deste universo de pesquisa podemos citar algumas publicações: (SILVA et al., 2015), (BASHIR et al., 2016), (SILVA et al., 2016), (GUEDES et al., 2017), (GUEDES; ZEBENDE; LIMA, 2017), (HUSSAIN et al., 2017), (FERREIRA et al., 2018), (PEREIRA et al., 2018), (NASCIMENTO et al., 2018), (SILVA et al., 2019), (GUEDES; ZEBENDE, 2019), (GUEDES et al., 2019), (BASHIR et al., 2019), (ABBAS et al., 2019), (FERREIRA et al., 2019), (FERREIRA; KRISTOUFEK; PEREIRA, 2020), (PEREIRA et al., 2019), (PEREIRA; FERREIRA, 2019), (FERREIRA; PEREIRA; SILVA, 2020).

Porém, para a análise da propagação das crises financeiras, a abordagem linear é a mais utilizada e é fundamentada na comparação dos coeficientes de correlação linear do período que antecede a crise com o período pós crise, ver (LI; ZHU, 2014). Segundo Guedes (2018), os métodos lineares como por exemplo, o coeficiente de Pearson, não é apropriado para séries com características não lineares, como a da volatilidade. Portanto, em 2016, surge um novo método, o $\Delta\rho_{DCCA}$, proposto por Silva et al. (2016), com o objetivo de mensurar os efeitos da interdependência e contágio em mercados financeiros por meio da diferença das correlações entre o pré e o pós-crise, método este derivado do coeficiente ρ_{DCCA} . A significância estatística do método $\Delta\rho_{DCCA}$, foi aferida em 2018, por Guedes et al. (2018a), (ver Apêndice B) e teve como base o teste estatístico para o coeficiente ρ_{DCCA} , proposto por Podobnik et al. (2011).

Para o teste estatístico $\Delta\rho_{DCCA}$, analisamos séries simuladas e reais. Os resultados do teste estatístico mostraram que a significância estatística do $\Delta\rho_{DCCA}$ depende do tamanho N da série temporal e da escala temporal n , definindo um valor crítico para essa dependência em 90%, 95% e 99% do nível de confiança. Dessa forma, o $\Delta\rho_{DCCA}$ introduziu uma nova maneira de analisar os efeitos da interdependência e do contágio na economia, sendo bastante aplicado e referenciado.

Em 2018, apresentamos o algoritmo para o teste estatístico $\Delta\rho_{DCCA}$, ver (GUEDES et al., 2018b) (Apêndice B). O algoritmo consiste na simulação de pares de séries temporais pelo processo ARFIMA. Essas séries temporais possuem tamanhos distintos, sendo $N = 250$, $N = 500$, $N = 1000$ e $N = 2000$ pontos. A função de distribuição de probabilidade (PDF) é disponibilizada para todas as 10.000 amostras geradas, que partem da série temporal original, em material suplementar permitindo que outros pesquisadores estendam suas análises. Como já mencionado acima, o coeficiente ρ_{DCCA} quantifica o nível de correlação cruzada entre duas séries temporais, não sendo possível sua aplicação para mais de duas séries. Assim, com o objetivo de generalizar a análise do coeficiente ρ_{DCCA} para múltiplas

séries, em 2018 [Zebende e Filho \(2018\)](#), propuseram um novo coeficiente, denominado de Coeficiente de Correlação Cruzada Múltipla, denotado por DMC_x^2 , que mensura a correlação cruzada entre múltiplas séries temporais. A grande relevância de se estudar mais de duas séries temporais são as múltiplas relações existente entre as mesmas. Por se tratar de uma metodologia recente, o coeficiente DMC_x^2 é ainda pouco explorado na literatura.

O artigo “Detrended Multiple Cross-Correlation Coefficient applied to solar radiation, air temperature and relative humidity”, publicado na revista **Scientific Reports** relata a aplicação e os resultados da técnica proposta neste trabalho de doutorado, ([BRITO; Araújo; ZEBENDE, 2019](#)).

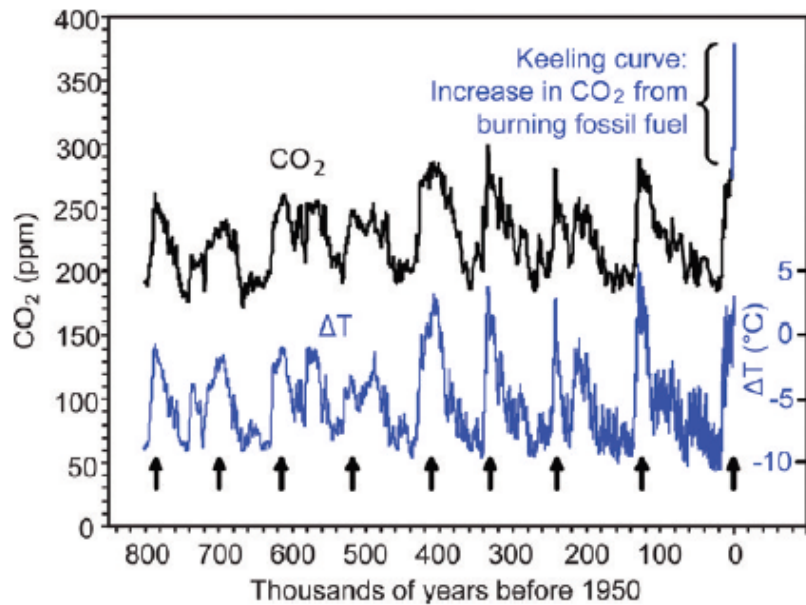
Através destes métodos de análise de séries temporais mencionados acima, a presente pesquisa procura responder a seguinte questão: É possível modelar os dados empíricos e simulados pelos referidos métodos estatísticos de tal maneira a encontrar resultados científicos inéditos? E com o intuito de discorrer sobre esta questão, nesta Tese temos como objetivo principal realizar a modelagem computacional da auto-correlação e correlações cruzadas através dos métodos DFA, DCCA e os coeficientes de correlação cruzada ρ_{DCCA} e DMC_x^2 , para mensurar as correlações cruzadas entre variáveis meteorológicas e dados simulados, entre outros.

Este presente projeto de Doutorado é a continuidade da pesquisa desenvolvida no Mestrado onde foram realizados o estudo da auto-correlação e correlação cruzada em séries temporais de dados meteorológicos para 31 estações meteorológicas localizadas no estado da Bahia (Brasil), ver ([BRITO; MACHADO; ZEBENDE, 2015](#)) e ([BRITO, 2016](#)).

Seguindo esta lógica, faremos a abordagem de algumas variáveis meteorológicas e a importância de sua modelagem estatística e computacional. Como é sabido, há tempos o homem vem registrando dados meteorológicos, como por exemplo, o valor da temperatura ambiente de milhares de anos atrás, ([HARRIS, 2010](#)). Isto se deu por um grande esforço de corajosos cientistas e por modernas técnicas, que foi possível ter um valor estimado da temperatura ambiente e ver que esta está diretamente ligada com o aumento (diminuição) do CO_2 na atmosfera (Figura 1.1).

Segundo [Barry e Chorley \(2013\)](#), o armazenamento dos valores da temperatura ambiente ao longo do tempo começou durante o século *XIX* com estudo da climatologia onde foi possível montar um grande banco de dados climatológicos e usá-los para generalizar regiões. Atualmente, modernas estações climatológicas (convencionais ou automáticas) são construídas com o intuito de deixar como legado os valores das variáveis climatológicas (temperatura do ar, umidade relativa do ar, pressão atmosférica, direção e velocidade do vento, precipitação, radiação solar, entre outras) como função do tempo e da sua locali-

Figura 1.1: Dados atmosféricos de CO₂ e temperatura derivados de Medições do núcleo de gelo antártico



Fonte: (HARRIS, 2010)

zação. Com esta grande “massa de dados” (*Big Data*), técnicas e ferramentas computacionais são de grande importância, permitindo assim uma análise cada vez mais acurada das condições meteorológicas (na escala de dias).

Especificamente, lincaremos esta grande base de dados (simulado e empírico) com os modelos estatísticos e computacionais mencionados acima, com o objetivo de se obter resultados inovadores para que futuramente os mesmos possam, por exemplo, ser empregados nos atuais modelos meteorológicos, contribuindo assim com uma nova visão na área.

Para atender o objetivo geral desta Tese, definimos como objetivos específicos:

- i Analisar as correlações cruzadas entre as variáveis climatológicas (radiação solar, temperatura do ar e umidade relativa do ar), via os coeficientes ρ_{DCCA} e DMC_x^2 ;
- ii Propor uma nova visão para as variáveis climatológicas (temperatura do ar e umidade relativa do ar) através do mapa de cores das correlações cruzadas;
- iii Identificar locais com maior eficiência para captação de energia solar tendo como parâmetro três variáveis climatológicas conjugadas simultaneamente;
- iv Averiguar a robustez do coeficiente de correlação cruzada ρ_{DCCA} no corte e remoção nas séries temporais.

Vale destacar que projetos, dentro do escopo da Tese, estão em fase de finalização e versam sobre a auto-correlação e correlação cruzadas em dados horários da temperatura para algumas estações meteorológicas localizadas no estado da Bahia (Brasil).

Por se tratar de um estudo de natureza interdisciplinar, direcionada para área de sistemas complexos, como aspecto metodológico, esta pesquisa tem caráter quali-quantitativo, explorado por meio de bibliografia e pesquisa em laboratório, com a realização de simulações e aplicação de modelos computacionais em dados meteorológicos e simulados. Esta Tese está organizada em capítulo, da seguinte forma:

- **Capítulo 1 - INTRODUÇÃO** - Apresentamos a definição do problema, o objetivo geral, os objetivos específicos e os aspectos metodológicos.
- **Capítulo 2 - FUNDAMENTAÇÃO TEÓRICA** - Neste capítulo abordaremos a teoria relativa ao estudo de séries temporais e os métodos aplicados na pesquisa, tais como: **DFA**, **DCCA**, ρ_{DCCA} , $\Delta\rho_{DCCA}$ e DMC_x^2 .
- **Capítulo 3 - PUBLICAÇÕES** - Neste capítulo apresentamos os resultados no formato de artigos científicos publicados em periódicos indexados.
- **Capítulo 4 - CONSIDERAÇÕES FINAIS** - Apresentamos as conclusões, contribuições e propostas futuras.

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo faremos um breve resumo sobre variáveis meteorológicas e os métodos estatísticos para análise de séries temporais estacionárias e não estacionárias, dentre eles: o Coeficiente de Correlação de Pearson, σ_{xy} ; a Função de Auto-correlação **FAC**; a Análise *R/S* de Hurst; o método *Detrended Fluctuation Analysis*, **DFA**; o método *Detrended Cross-Correlation Analysis*, **DCCA**; os coeficientes de correlações cruzadas, ρ_{DCCA} e DMC_x^2 e o coeficiente $\Delta\rho_{DCCA}$.

2.1 Variáveis Meteorológicas

Segundo [Pereira, Angelocci e Sentelhas \(2007\)](#), a atmosfera é uma massa de ar em contínuo movimento e isto induz variações nas condições meteorológicas predominantes em uma região. O estado da atmosfera pode ser descrito por variáveis que caracterizam sua condição energética. Para um determinado local, essa descrição pode ser tanto em termos instantâneos, definindo sua condição atual, como termos estatísticos, definindo uma condição média, introduzindo uma escala temporal na descrição das condições atmosféricas.

Dentre as principais variáveis meteorológicas que influenciam o comportamento termodinâmico da atmosfera podemos citar: a temperatura do ar, a umidade relativa do ar, a radiação solar, a pressão atmosférica, o vento (velocidade, direção), a evaporação e a precipitação. O estudo do comportamento dessas variáveis está relacionado a uma série de fatores, também climatológicos, e que serão apresentados nesta seção a fim de justificar o comportamento e as possíveis diferenças das condições climáticas analisadas nas cidades monitoradas nesta pesquisa de Tese.

Partindo de uma escala macro, a fim de chegarmos às variáveis meteorológicas analisadas, definimos a meteorologia como a ciência que estuda a atmosfera terrestre e seus fenômenos, enquanto a climatologia trata do estudo científico do clima, ([VIEIRA, 2009](#)). Na meteorologia o conceito entre tempo e clima são distintos. Denomina-se tempo a descrição instantânea das variáveis meteorológicas, enquanto a descrição média dessas variáveis é denominada clima. Logo, tempo é o estado atual da atmosfera em um local e instante e o clima é a descrição média (ou comportamento médio) das variáveis atmosféricas nesse mesmo local ou região. Quando se trata de uma climatologia (ou normal climatológica), se refere às características da atmosfera inferidas de observações contínuas das variáveis meteorológicas durante um longo período de tempo, como por exemplo 30 anos.

As observações meteorológicas consistem na medição, registro ou determinação das variáveis (elementos) que, em seu conjunto, representam as condições meteorológicas em um dado momento e em determinado local, utilizando instrumento adequado e valendo-se da vista do observador. Estas observações realizadas de forma sistemática, uniforme, ininterrupta e em horas estabelecidas de acordo com a Organização Meteorológica Mundial (OMM), permitem conhecer as características e variações dos elementos atmosféricos, os quais constituem os dados básicos para confecção de mapas, representando as condições do tempo e do clima, bem como, para a investigação de leis gerais que regem os fenômenos meteorológicos.

Também é importante mencionar nessa pesquisa a influência dos fatores climáticos que condicionam os elementos (variáveis), tais como os geográficos: latitude, longitude, altitude, continentalidade/oceanidade, tipo de correntes oceânicas. Por exemplo, é sabido que a temperatura do ar varia com a altitude, sendo ela menor nas suas maiores escalas. A radiação solar recebida em um determinado ponto, depende da latitude, altitude e também da época do ano.

De acordo com [Vieira \(2009\)](#), as observações de superfície são procedimentos sistemáticos e padronizados pela OMM no que diz respeito ao tipo de equipamento, às técnicas de calibração, à aferição, aos ajustes, ao manuseio, a procedimentos observacionais, aos horários de observação, ao tratamento dos dados, às correções, às estimativas, à transmissão e ao uso operacional. Tais medidas visam à obtenção de informações qualitativas e quantitativas referentes às variáveis meteorológicas capazes de serem comparadas e de caracterizarem plenamente o estado instantâneo da atmosfera. Os dados meteorológicos são obtidos mediante leituras ou registros contínuos, observados diretamente dos instrumentos instalados nas estações.

Existem dois tipos de estações meteorológicas de superfície: as convencionais e as automáticas. Para as convencionais, o seu funcionamento exige a presença diária do observador meteorológico para coleta dos dados, enquanto nas automáticas, a coleta de dados é feita de forma automatizada, sem a interferência humana, com menores intervalos de coletas e de acordo com a necessidade do segmento. Aqui nesta Tese, nossa pesquisa foi realizada em estações meteorológicas de superfície automáticas localizadas no Estado da Bahia, Brasil.

O estado da Bahia é uma região de aproximadamente 565km², localizado entre as latitudes de aproximadamente 9°S e 18°S e longitudes de 37°O e 46°O, tem como principal característica climática a alta variabilidade espacial e temporal da precipitação, determinada principalmente por padrões de grande da circulação geral da atmosfera, associada a características locais como topografia e forma geográfica, ver ([PARRY; CARTER; KONIJN, 1988](#)). Ao contrário do norte do Nordeste, trabalhos científicos que tratam da

variabilidade dos parâmetros meteorológicos, como precipitação, radiação solar, umidade do ar e temperaturas no sul da Região ainda são poucos.

Segundo [Parry, Carter e Konijn \(1988\)](#), a variabilidade climática no Nordeste brasileiro é resultado de efeitos combinados da ação de vários sistemas meteorológicos e das variações na intensidade e posição das circulações de Hadley e Walker. A circulação de Walker com ramo ascendente sobre a bacia Amazônica, associada aos elevados índices de temperatura e umidade do ar, causando a convecção intensa sobre esta e, a circulação de Hadley cujo ramo ascendente está associada a Zona de Convergência Intertropical (ZCIT), apresentam um ramo descendente sobre o Nordeste brasileiro a adjacências do Oceano Atlântico subtropical sul.

Fatores locais como características topográficas, alto valores de albedo e forma geográfica, também contribuem com a aridez do Nordeste brasileiro, sobretudo na porção semiárida. Segundo [Filho \(1979\)](#), por meio de modelagem numérica, mostra que a topografia da Região tende a intensificar os movimentos subsidentes sobre o semiárido nordestino, enquanto que o albedo diferencial não interferiria nos resultados. Para [Aragão \(1975\)](#), não existe déficit de umidade sobre a região semiárida, e a ausência de um mecanismo dinâmico que provoque movimentos ascendentes seria a principal razão da existência do semiárido nordestino.

No estado da Bahia, atuam vários sistemas meteorológicos que influenciam as condições do tempo e do clima, contribuindo no aumento da umidade do ar, cobertura de nuvens, gerando precipitação, variabilidade das temperaturas e incidência da radiação solar. Tais sistemas de origem tanto tropicais quanto extratropicais, como a Zona de Convergência Intertropical (ZCIT), Zona de Convergência do Atlântico Sul (ZCAS); de escala sinótica, como os sistemas frontais e os Vórtices Ciclônicos de Altos Níveis (VCAN) e, os sistemas de escala local, como as brisas. A atuação destes sistemas determina três períodos chuvosos e a distribuição espacial da precipitação sobre o sul do Nordeste, que inclui a Bahia.

O principal período chuvoso no sul do Nordeste (que inclui o estado da Bahia) ocorre nos meses de novembro a janeiro, sendo no mês de dezembro quando ocorrer os maiores volumes, ([KOUSKY; CHUG, 1978](#)) e ([RAO; HADA, 1990](#)). A precipitação neste período está associada a passagem das frentes frias, oriundas do sul do País, ou seus vestígios em latitudes mais baixas, ver ([KOUSKY, 1979](#)). Estes sistemas frontais fornecem os mecanismos de indução à convecção tropical e estão, algumas vezes, relacionadas à Zona de Convergência do Atlântico Sul (ZCAS) e aos Vórtices Ciclônicos de Altos Níveis (VCAN).

A Zona de Convergência do Atlântico Sul (ZCAS) é caracterizada por uma persistente banda de nebulosidade orientada no sentido noroeste-sudeste, que se estende desde a Amazônia até o oceano Atlântico sudoeste, ([KODAMA, 1992](#)), ([QUADRO, 1994](#)), ([KOUSKY,](#)

1988) e (CARVALHO; JONES; LIEBMANN, 2004), influenciando as precipitações desde o centro-sul da Região Norte até o norte do Paraná, incluindo a totalidade das Regiões Centro-Oeste e Sudeste e parte do Nordeste do Brasil, principalmente no sul do Maranhão e do Piauí, além de grande parte da Bahia. A ZCAS está associada com uma sequência de dias chuvosos que em muitas ocasiões provocam episódios de chuva extrema e, por consequência, desastres naturais como por exemplo inundações e deslizamentos de terra, (REBOITA et al., 2017) e (SELUCHI; CHOU, 2009).

Já o VCAN se caracteriza por um centro de baixa pressão frio restrito à média e alta troposfera, originando um escoamento descendente de ar seco em direção aos baixos níveis da atmosfera, reduzindo a umidade do ar, a formação de nuvens e, conseqüentemente as chuvas. Kousky e Gan (1981) e Ramirez, Kayano e Ferreira (1999), mostraram que os VCAN são mais comuns durante a estação do verão no Hemisfério Sul, com tempo bom em seu centro (área de subsidência) e convecção com chuvas intensas em sua borda. No Nordeste do Brasil, estes sistemas são mais frequentes no estado da Bahia. O segundo período chuvoso da Bahia se concentra no norte do Estado entre os meses de fevereiro a abril, como consequência do deslocamento latitudinal da Zona de Convergência Intertropical (ZCIT), quando alcança sua posição mais ao sul. A ZCIT é o fenômeno meteorológico mais importante para provocar chuva sobre o norte do Nordeste do Brasil e é caracterizada por várias bandas de nuvens carregadas que se formam a partir do encontro (convergência) dos ventos alísios na faixa equatorial. De acordo com Uvo e Nobre (1989), o principal sistema gerador de precipitação norte do Nordeste está relacionado à proximidade da ZCIT e, tal fato é observado quando se considera que o pico de precipitação sobre a região (nos meses março e abril) ocorre exatamente na época em que a ZCIT atinge suas posições mais ao sul. Por isso, a importância de identificar o quanto tempo este sistema ficará posicionada mais ao sul de sua posição normal, podendo chegar até o norte da Bahia.

Na parte leste da Bahia, o principal período chuvoso acontece nos meses de abril a junho. Este período chuvoso é associado aos sistemas de vento local e aos vestígios de sistemas frontais, (KOUSKY, 1979). Rao, Lima e Franchito (1993) verificaram que as precipitações nesta faixa estão também associadas com os ventos alísios de sudeste, que são modulados pelo sistema de Alta Pressão do Atlântico Sul (ASAS). Quanto a variabilidade diária, nas áreas litorâneas a máxima atividade convectiva (com chuva mais intensa) acontece no período noturno, como consequência de ventos locais que se formam ao longo do litoral, principalmente no período de abril a julho. Nas áreas adjacentes a costa ocorre entre 15 e 21 horas (hora local), devido à penetração de brisa associada ao aquecimento superficial diurno. Na região central da Bahia, devido ao efeito topográfico, a atividade convectiva é máxima entre 09 e 15 horas (hora local), associada aos ventos anabáticos. No Vale do São Francisco a máxima atividade convectiva ocorre entre 21 e 06 horas, associada aos ventos catabáticos, que promovem o levantamento do ar no centro do vale, (KOUSKY, 1980).

A maior parte da Bahia apresenta baixo índice pluviométrico e alta variabilidade espacial e temporal. Na faixa litorânea os totais anuais de precipitação excedem os 1.000 mm, decrescendo, com forte gradiente, em direção ao interior do Estado, onde o total anual apresenta valores entre 400 mm e 800 mm, valores que caracterizam as regiões semiáridas, (KOUSKY; CHUG, 1978). Em áreas isoladas da região conhecida como “polígono das secas” encontram-se valores de precipitação anual acima de 1.000 mm associados com a formação rochosa da Chapada Diamantina. Em direção Oeste do Estado, o volume da precipitação anual volta a aumentar alcançando valores acima de 1.000 mm, onde a convecção em muitos casos está associada à convecção oriunda da Amazônia.

Durante os meses de verão, praticamente todos os sistemas meteorológicos que atuam sobre a América do Sul, têm influência sobre a precipitação no estado da Bahia. Entre estes, encontra-se a Alta da Bolívia, caracterizada por um centro de alta pressão em altos níveis da atmosfera e circulação anticiclônica (contrária a direção dos ponteiros do relógio), situada em média entre 15°S e 65°O, (TANAJURA, 1996). Esta circulação é resultado das altas taxas de calor latente liberadas sobre a Amazônia devido à convecção intensa nesta região, (DIAS; SCHUBERT; DEMARIA, 1983) e (FIGUEROA; SATYAMURTY; DIAS, 1995). O aquecimento atmosférico diurno pela emissão de calor sensível pelo platô da Bolívia também contribui para a formação deste sistema, (SCHWERDTFEGER, 1976).

Sobre o Nordeste brasileiro e áreas oceânica adjacentes, há também ocorrência Vórtices Ciclônicos de Altos Níveis (VCAN), que atuam principalmente nos meses de primavera, verão e outono, apresentando maior frequência no mês de janeiro. Estes penetram no interior da Região, na maioria das vezes próximo a cidade de Salvador, desintensificando-se à medida que se deslocam para oeste, produzindo céu claro na parte central da Bahia e chuvas nos estados do norte do Nordeste, (KOUSKY; GAN, 1981). No inverno os vestígios dos sistemas frontais tendem a provocar chuvas praticamente só nas áreas litorâneas, enquanto que no verão estes têm papel fundamental na formação da convecção tropical no interior da Região, onde está localizada a maior parte do semiárido nordestino. Em um estudo climatológico sobre os sistemas frontais que passaram por Caravelas (sul da Bahia) no período de 1961 a 1970. Kousky (1979) mostra que estas penetram no sul da região durante todo o ano e tem importância fundamental na formação nos sistemas convectivos (chuvas mais intensas) no período de novembro a janeiro, algumas vezes afetando também os estados do norte do Nordeste.

A Zona de Convergência do Atlântico Sul (ZCAS) tem papel importante na precipitação do sul do Nordeste (inserindo todo o estado da Bahia) durante os meses de primavera e verão. De acordo com Nobre (1988), nos anos considerados como chuvosos, a ZCAS situa-se ao norte (atingindo o sul do Piauí e do Maranhão, além de grande parte da Bahia) e nos anos considerados secos, mais ao sul. Esta Zona é caracterizada por uma faixa de nebulosidade convectiva (grande dimensão vertical e associada a chuvas intensas) que se

estende da região Amazônica até o Atlântico subtropical, nos meses de primavera e verão.

2.1.1 Dados utilizados e área de estudo

Segundo Araújo (1996), o estado da Bahia tem como principal característica a alta variabilidade espacial e temporal da precipitação determinada, principalmente, por padrões de grande escala da circulação geral da atmosfera, como também por características locais. Esta variabilidade climática é resultado de efeitos combinados da ação de vários sistemas meteorológicos, bem como, das variações e intensidade de cada um deles. Além disso, fatores locais como características topográficas e forma geográfica, também contribuem para essa variabilidade.

Nesta Tese foram analisados os dados horários de temperatura do ar, umidade relativa do ar e radiação solar global, coletados pelas estações meteorológicas automáticas pertencentes ao Instituto Nacional de Meteorologia (INMET), instalada nas cidades de Salvador (litoral), Paulo Afonso (semiárido), Cruz das Almas (agreste) e Barreiras (cerrado) no estado da Bahia (Brasil). A Tabela 2.1 apresenta o resumo das características físicas das estações utilizadas nessa pesquisa.

Tabela 2.1: Descrição das estações meteorológicas.

Estação	Código	Lat. (S)	Long. (O)	Alt.	Abertura	Situação
Salvador	A401	13°0'21"	38°30'21"	47	05 de 2000	operando
Barreiras	A402	12°7'29"	45°1'37"	474	12 de 2001	operando
Paulo Afonso	A411	9°22'41"	38°13'36"	219	02 de 2003	fechada (2017)
Cruz das Almas	A406	12°40'31"	39°5'23"	255	06 de 2003	operando

Fonte: INMET

A seleção das localidades deu-se pelo fato destas possuírem características climatológicas distintas (conforme Tabela 2.1) entre elas, de acordo com as microrregiões nas quais essas cidades estão localizadas, bem como, por dispor de uma série histórica de dados representativa (período de 10 anos, aproximadamente).

As estações automáticas utilizadas nesta pesquisa operam de forma contínua na coleta dos dados das variáveis meteorológicas, como: temperatura do ar (instantânea, máxima e mínima), umidade do ar (instantânea, máxima e mínima), pressão atmosférica, precipitação pluviométrica, direção e velocidade do vento, radiação solar global, dentre outras. Tais dados são coletados por meio de sensores eletrônicos, armazenados na memória da estação e transmitidos via satélite em intervalos de 01 (uma) hora. Após transmissão, estes dados são disponibilizados no site institucional do INMET, no *link*:

<<https://portal.inmet.gov.br/>>

em dados Meteorológicos.

2.1.2 Descrição das variáveis meteorológicas

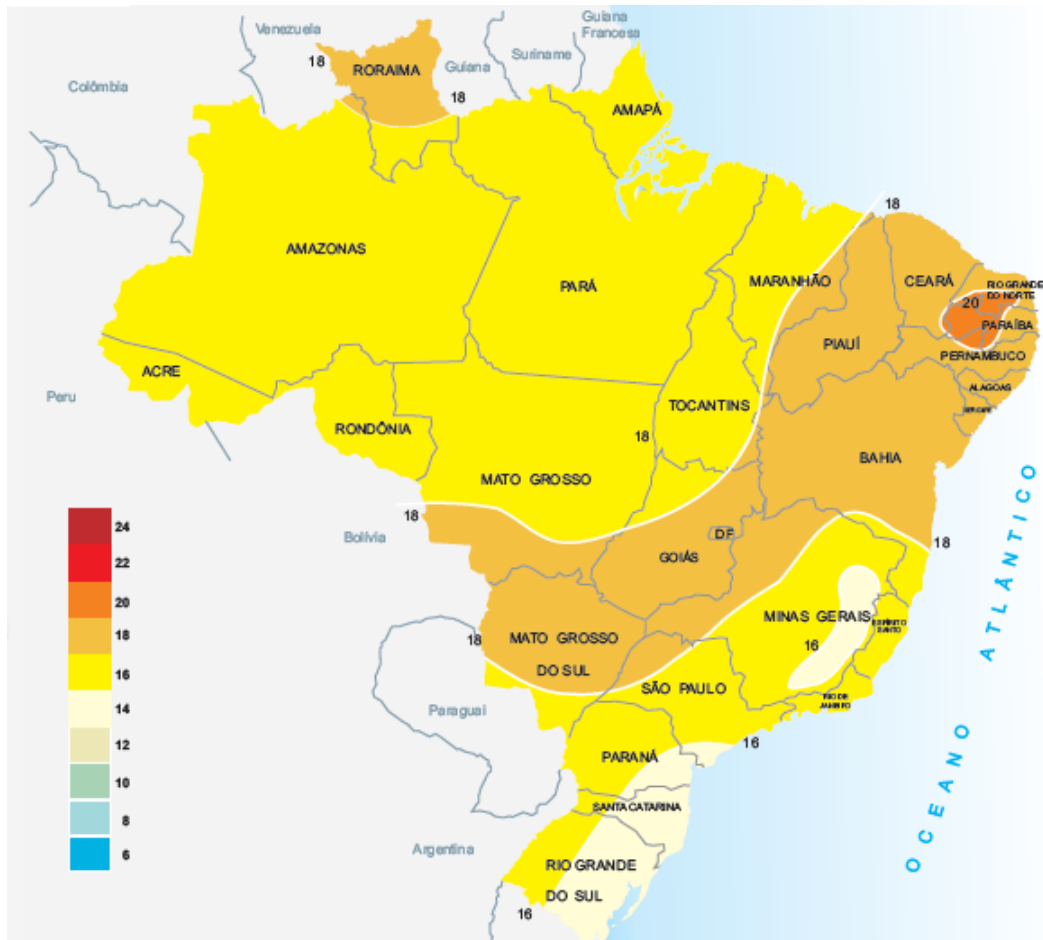
Resultados preliminares, obtidos por Brito (2016), mostraram características estatísticas distintas nas localidades de Salvador, Paulo Afonso, Cruz das Almas e Barreiras, analisando as variáveis temperatura do ar e umidade relativa do ar. Ainda de acordo com os autores, o comportamento da temperatura do ar para uma mesma localidade varia ao longo do dia, bem como ao longo do ano. Tais variações dependem de outras variáveis meteorológicas, a exemplo, da radiação solar, deslocamento das massas de ar, correntes oceânicas, posição geográfica, altitude, umidade do ar, dentre outros.

Análises estatísticas poderão descrever a atuação e influência de fenômenos atmosféricos, caracterizando o clima em função da localização geográfica, estações do ano e as variáveis meteorológicas utilizadas. De acordo com estudos realizados por Brito (2016), o comportamento da temperatura do ar para uma mesma localidade varia ao longo do dia, bem como ao longo do ano. Tais variações dependem de outras variáveis meteorológicas, a exemplo, da radiação solar, deslocamento das massas de ar, correntes oceânicas, posição geográfica, altitude, umidade do ar, dentre outros.

A radiação solar, segundo Angelocci, Pereira e Sentelhas (2002), é a maior fonte de energia a Terra, sendo também o principal elemento meteorológico, pois é ela que desencadeia todo o processo meteorológico afetando todos os outros elementos (temperatura, pressão, vento, chuva, umidade, etc). A energia solar é a fonte primária de energia para todos processos terrestres, desde a fotossíntese, responsável pela produção vegetal e manutenção da vida na presente forma, até o desenvolvimento de furacões, tempestades, enfim, pela circulação geral da atmosfera e oceanos. Além da sua importância em Meteorologia, a energia radiante do Sol é um elemento fundamental em estudos ecológicos e de disponibilidade energética, pois a maior parte da energia disponível na Terra tem origem na radiação solar.

A radiação solar é a principal variável de influência sobre a temperatura ambiente na Terra. A disponibilidade do recurso em um dado local dependerá da posição do sol em relação à superfície terrestre. Localidades com menores latitudes, o seja próximas à linha do Equador, contam com maior incidência da radiação solar, a exemplo da Região Nordeste do Brasil, do que naquelas localidades em altas latitudes, como nos Polos Terrestre. A Figura 2.1 apresenta a espacialização radiação solar anual média no território brasileiro

Figura 2.1: Radiação solar global média anual (MJm^{-2} por dia)



Fonte: Atlas Solarimétrico do Brasil - UFPE (2000)

A temperatura ambiente é influenciada pela movimentação das massas de ar na atmosfera, sendo natural o processo físico em que as correntes de convecção acontecem pelo fato de o vento soprar de uma área mais fria (alta pressão) para uma área mais quente (baixa pressão). A movimentação de correntes oceânicas também influencia a temperatura ambiente, ou seja, as correntes frias migram dos Polos para o equador, enquanto as correntes quentes fazem o caminho inverso. Essa movimentação das correntes oceânicas afeta a velocidade, direção e sentido dos ventos, causando uma interferência direta na temperatura local, (SILVA, 2006).

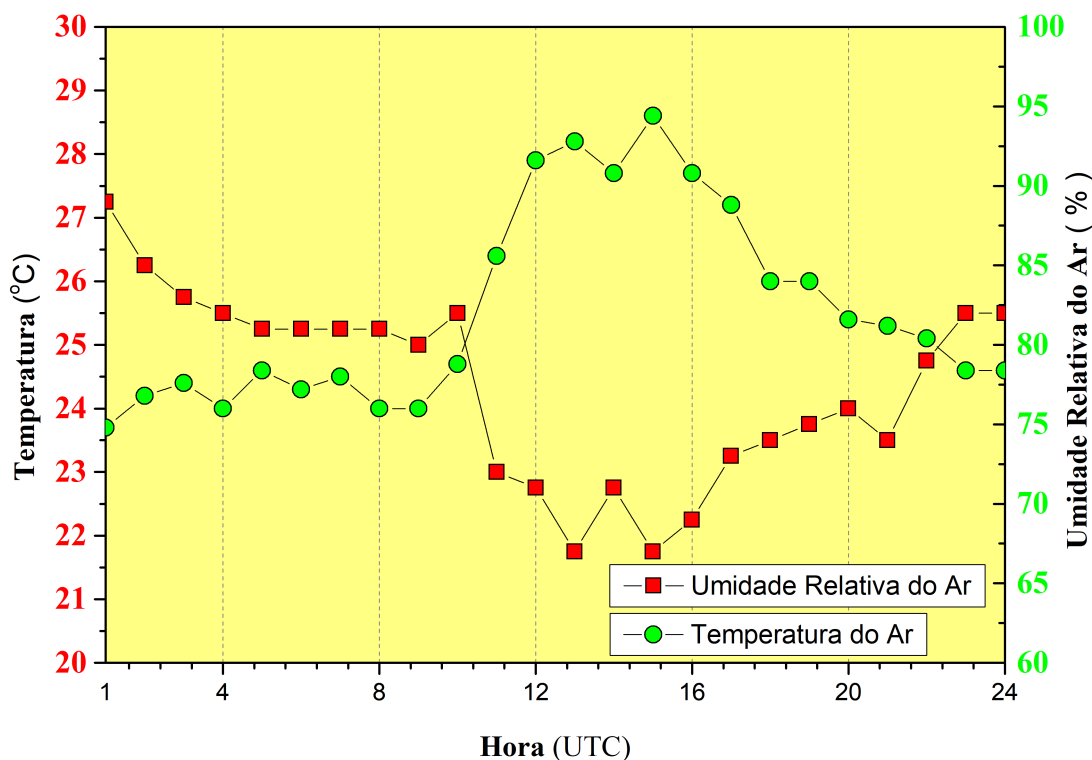
Outro fator também determinante na temperatura ambiente é a posição geográfica da localidade. Percebe-se que em regiões costeiras, onde os ventos são dirigidos do mar para o continente, a temperatura do ar pode ser diferente de uma localidade continental em que o vento não obedeça a essa direção. Nota-se também que as regiões com menor influência da brisa marítima, sofrem uma maior variação de temperatura entre as estações do inverno e do verão. O relevo, por sua vez, influenciam de forma significativa na variável temperatura, a exemplo de cadeias montanhosas que pode dificultar ou, até mesmo, alterar, a circulação do vento, influenciando diretamente na temperatura local.

O comportamento da umidade relativa do ar, outra variável analisada nesta pesquisa, está relacionado à quantidade de vapor de água existente no ar, sobretudo nas camadas mais próximas a superfície. Segundo estudos realizados por Brito (2016), a umidade relativa do ar é uma variável muito utilizada na climatologia, até mais usada que a própria umidade absoluta do ar, pois ela indica a proximidade que o ar se encontra em sua saturação, ou seja, 100%, percentual que ocorrem as precipitações. No entanto, a quantidade de umidade necessária para ocorrer a saturação do ar dependerá também da temperatura em que o ar se encontra. Em síntese, o comportamento da variação da umidade relativa do ar está relacionado a variação da temperatura ambiente ao longo do dia, ao movimento horizontal da massa de ar e movimento vertical do ar pelas correntes de convecção.

Segundo Silva (2006), a temperatura do ar traduz a temperatura reinante em um ponto da atmosfera próximo a superfície da Terra. Para fins de análise do estado físico da atmosfera ou climatológicos, as observações da temperatura do ar à superfície são efetuadas a uma altura de 1,25 a 2,00 m acima do terreno. De acordo com o autor, a mais elevada e a mais baixa temperaturas observadas em um determinado intervalo de tempo, constituindo as temperaturas extremas desse mesmo intervalo, são conhecidas como máxima e mínima no referido intervalo.

Sabemos que há uma relação entre a temperatura e a umidade do ar, mesmo mantendo-se constante a quantidade de vapor de água real. Por exemplo, a Figura 2.2, mostra o comportamento diário da umidade relativa do ar em função da temperatura ao longo de um dia.

Figura 2.2: Variação horária da temperatura do ar e umidade relativa do ar, para a estação meteorológica de Salvador na data de 24/06/2019



2.2 Variáveis Simuladas - Processo ARFIMA

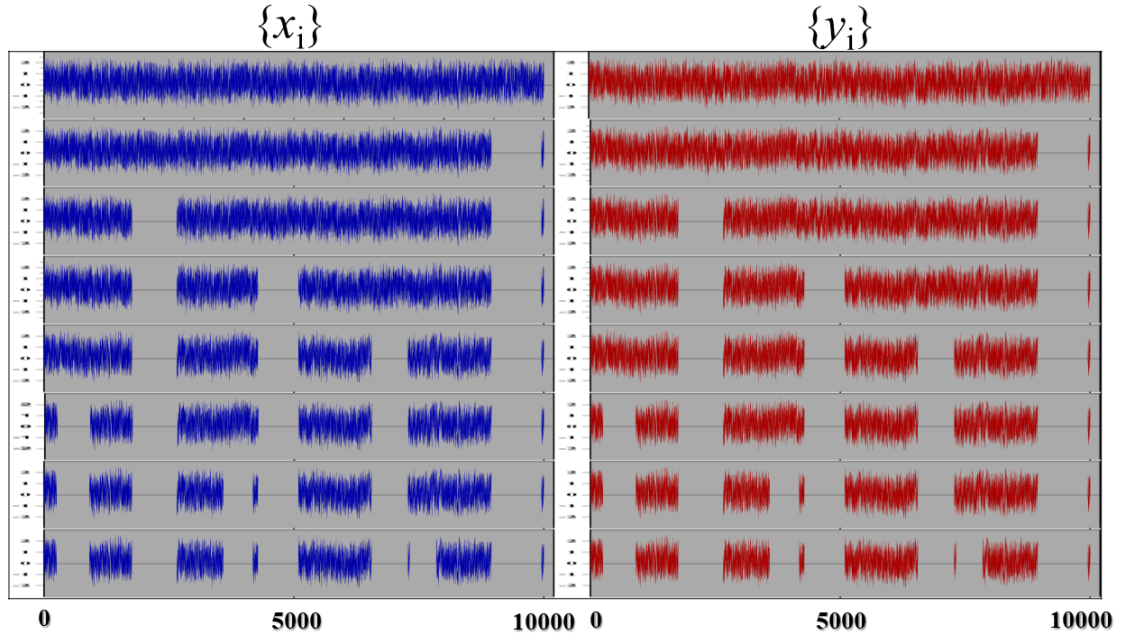
Nesta Tese, para o nosso estudo de caso, simulamos duas séries temporais com processo de memória longa e para isso usamos o processo Auto-regressivo Fracionário de Média Móvel (ARFIMA), onde cada variável depende do próprio passado e dos valores passados da outra variável, ver Equação 2.1.

$$\begin{aligned}
 x_i &= W \sum_{m=1}^{\infty} a_m(\delta_1) x_{i-m} + (1 - W) \sum_{m=1}^{\infty} a_m(\delta_2) y_{i-m} + \varepsilon_{x_i}, \\
 y_i &= (1 - W) \sum_{m=1}^{\infty} a_m(\delta_1) x_{i-m} + W \sum_{m=1}^{\infty} a_m(\delta_2) y_{i-m} + \varepsilon_{y_i}.
 \end{aligned} \tag{2.1}$$

Na Figura 2.3 apresentamos o algoritmo descrito acima com $N = 10.000$ e fizemos cortes até atingir 50 % do valor inicial (neste exemplo com N final igual a 5.000).

Nossa simulação foi realizada da seguinte forma: primeiramente, simulamos duas séries temporais auto-afins, $\{y_{rand1}\}$ and $\{y_{rand2}\}$, com números aleatórios entre 0 e 1 e $N =$

Figura 2.3: Algoritmo de corte e remoção de partes das séries temporais originais geradas por um processo ARFIMA com $N = 10.000$ (canto superior): $\{x_i\}$ é o **ARFIMA 01** com $\delta_1 = 0,1$ e $\{y_i\}$ é o **ARFIMA 02** com $\delta_2 = 0,4$. A força das correlações cruzadas, $W = 0,5$, com $\varepsilon_{x_i} \neq \varepsilon_{y_i}$



1.000.000. Essas séries temporais são independentes e não correlacionadas, com expoentes $\alpha_{DFA\{1\}} = \alpha_{DFA\{2\}} = 0,5$. Em seguida, aplicamos o processo ARFIMA (Equação 2.1) e introduzimos um comprimento de corte $l = 990.000$ em nossas simulações. Com a soma de 1 a l . Realizamos com essas etapas, um par de séries temporais com $N = 10.000$ e memória de longo alcance foram criados, ver Tabela 2.2. Apresentamos essas duas séries temporais na Figura 2.3.

Tabela 2.2: Parâmetros para o Processo Simulado de Série Temporal.

	ARFIMA 01	ARFIMA 02
δ	0,1	0,4
W	0,5	
	$\varepsilon_{x_i} \neq \varepsilon_{y_i}$	

Para nosso estudo de caso, ver **artigo 3.3**, através do processo ARFIMA, geramos duas séries temporais x_i (**ARFIMA 01**) and y_i (**ARFIMA 02**), com valores diferentes para $a_m(\delta_{1,2})$, W e comprimento N seguindo os seguintes passos:

1. A partir das séries temporais originais, calculamos $F_{DFA}(n)$, $F_{DCCA}^2(n)$ e $\rho_{DCCA}(n)$ (ver seção 2.6);
2. Escolhemos uma posição aleatória entre 1 e $N - (10\% \text{ de } N)$ (novo tamanho) na série temporal. Dessa posição, 10% da série é removida (corte i);
3. Os valores de $F_{DFA}(n)$, $F_{DCCA}^2(n)$ e $\rho_{DCCA}(n)$ são calculados para a nova série;

4. Repetimos o procedimento (b) e (c) novamente;
5. O algoritmo é implementado até que o comprimento a última série atinja $N/2$ ou seja, 50% do comprimento da série temporal original.

2.3 Séries Temporais

Segundo [Pereira \(1980\)](#), série temporal é uma coleção de observações feitas sequencialmente no tempo e para [Morettin e Tolo \(2006\)](#), série temporal é qualquer conjunto de observações ordenadas no tempo. São exemplos de séries temporais:

- valores horários das temperaturas do ar na cidade de Salvador-BA;
- casos de Chikungunya por mês no estado da Bahia;
- índices diários da Bolsa de valores de São Paulo.

Série temporal é denotada pelo conjunto dos $Z_{t_1}, Z_{t_2}, \dots, Z_{t_N}$ onde Z_{t_i} indica a variável observada e N o tamanho da série, com $i = 1, 2, 3, \dots, N$.

O estudo com séries temporais é aplicado em várias áreas do conhecimento como: na medicina (níveis de eletrocardiograma), na meteorologia (temperatura diária, umidade relativa do ar, velocidade do vento), epidemiologia (casos semanais de dengue, casos anuais de AIDS), economia (desemprego semanal, preços diários de uma bolsa de valores), mercadologia (vendas diárias, vendas semanais, gastos mensais com propaganda), demografia (população anual, nascimento e mortes semanais). Podemos classificar a série temporal como: contínua, discreta, multivariada ou multidimensional, isto é:

- é contínua quando as observações são realizadas continuamente no tempo, ou seja, T é um intervalo finito, como exemplo temos: medições durante três minutos de um eletrocardiograma;
- é discreta quando as observações são realizadas em tempos específicos, tempo igualmente espaçado, ou seja, T é um conjunto finito de pontos, como exemplo temos: o valor de importações mensais entre 1960 a 1980;
- multivariada, quando a série é discreta ou contínua, ou seja, $Y_1(t), \dots, Y_k(t), t \in T$ ou $Y(t), t \in T$, como exemplo: vendas e gastos com propaganda mensais de um determinado produto;

- multidimensional, quando temos $Z(t), t \in T$ (um vetor), por exemplo temos $Z(t, r, l)$ com $t \in T$ onde Z é o valor da temperatura em um determinado instante de tempo t , a latitude r e a longitude l .

Considera-se aqui dois aspectos para o estudo de séries temporais, a análise e a modelagem. Segundo [Pereira \(1980\)](#), o principal objetivo para se modelar uma série é tornar possível a previsão de valores futuros, procurando obter um modelo matemático que descreva o sistema de maneira simples para encontrar o desejado, enquanto que o objetivo da análise de uma série temporal é sumarizar as propriedades da série e caracterizar seu comportamento identificando ou sugerindo um modelo adequado.

Existem dois enfoques na análise de séries temporais: a análise no domínio do tempo e a análise no domínio das frequências. No primeiro, o interesse está no tamanho (intervalo) de eventos que ocorrem em determinado tempo e na relação entre as observações em diferentes tempos. A função de auto-correlação é o método clássico utilizado, enquanto que no segundo o interesse está na frequência com que certos eventos ocorrem em um período de tempo, neste caso, a transformada de Fourier é o método clássico utilizado.

2.3.1 Objetivo da Análise de Séries Temporais

O objetivo da análise de séries temporais é descrever o comportamento passado da série, com o intuito de compreender o comportamento da série e a conseqüentemente a previsão de movimentos futuros. Para [Pereira \(1980\)](#), o objetivo de analisar uma série temporal é estudar a distribuição do número de eventos em determinado período, assim como a distribuição dos intervalos entre ocorrências de eventos, entretanto, segundo [Morettin e Tolo \(2006\)](#), existem basicamente dois objetivos para análise de séries temporais: entender o mecanismo gerador da série, renunciando o comportamento futuro da série e construir modelos para as séries, com o propósitos determinados.

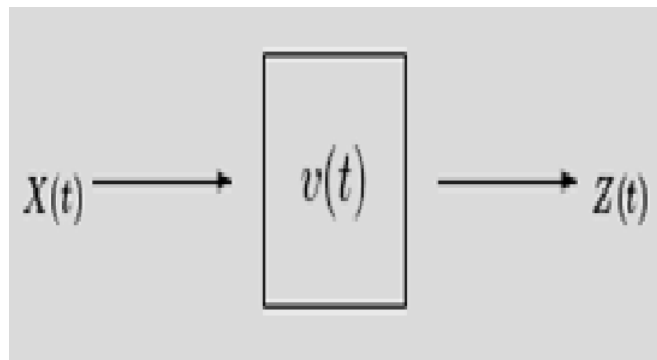
Portanto, dada uma série temporal $Z_{t_1}, Z_{t_2}, Z_{t_3}, \dots, Z_{t_N}$ observada nos instantes t_1, t_2, \dots, t_N , podemos estar interessados em:

- investigar seu mecanismo gerador;
- procurar comportamentos relevantes nos dados;
- fazer previsões a partir de valores passados, que podem ser de curto, médio e longo prazo;
- descrever o seu comportamento através da construção de gráficos, construção de

histogramas e diagramas de dispersão, obtenção da estatística descritiva e verificação de tendências, ciclos e sazonalidades.

Segundo [Pereira \(1980\)](#), as mais diversas situações em ciências físicas, biológicas e sociais, assim como na engenharia, envolvem o conceito de sistema dinâmico, caracterizado por uma série de entrada $X(t)$, uma série de saída $Z(t)$ e uma função de transferência $v(t)$, ver [Figura 2.4](#).

Figura 2.4: Sistema dinâmico



Fonte: [Pereira \(1980\)](#)

Desta forma, uma importante particularidade descrita por [Pereira \(1980\)](#) e por [Morettin e Toloi \(2006\)](#), são os sistemas lineares, onde a saída é relacionada com a entrada através de um funcional linear envolvendo $v(t)$. Como exemplo, temos o chamado modelo da função de transferência proposto por [Box, Jenkins e Reinsel \(2008\)](#), cuja expressão é dada pela [Equação 2.2](#).

$$Z(t) = \sum v(t)X(t - u) \quad (2.2)$$

Na função de transferência, com u sendo a defasagem no tempo, temos como interesses:

- estimar a função de transferência $v(t)$, conhecendo-se as séries de entrada $X(t)$ e saída $Z(t)$;
- fazer previsões da série $Z(t)$, com o conhecimento de observações da série de entrada $X(t)$ e $v(t)$;
- estudar o comportamento do sistema, simulando-se a série de entrada;
- controlar a série de saída $Z(t)$, de modo a trazê-la o mais próximo possível de um valor desejado, ajustando-se convenientemente a série de entrada $X(t)$. Este controle é necessário, devido perturbações que normalmente afetam um sistema dinâmico.

2.3.2 Estacionariedade

Segundo [Pereira \(1980\)](#), a série temporal é dita estacionária se não existe uma mudança sistemática na média (tendência) e na variância e, se as variações cíclicas forem removidas. A maior parte da teoria de séries temporais trata-se de séries estacionárias, por esta razão, a análise é feita transformando séries não estacionárias em estacionárias, para aplicar a teoria.

Para [Morettin e Tolo \(2006\)](#), uma das suposições mas frequentes que se faz a respeito de uma série temporal é a estacionariedade, ou seja, ela se desenvolve aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Geralmente, as séries são apresentadas na forma não estacionária, a exemplo, podemos citar algumas séries não estacionárias que apresentam tendência, como as séries financeiras e econômicas, onde a série flutua ao redor de uma reta, com inclinação positiva ou negativa (tendência linear). Uma outra forma encontrada de série não estacionária é a explosiva, como o crescimento de uma colônia de bactérias. Uma série temporal, poderá mudar de nível e/ou de inclinação durante sua trajetória, podendo ser estacionária por um período longo, mas também pode ser por períodos curtos.

A maioria dos procedimentos da análise estatística de séries temporais supõe que estas sejam estacionárias, sendo necessário transformar os dados originais. Esta transformação consiste em calcular as diferenças sucessivas da série original, até obter uma série estacionária. Assim, a primeira diferença de $Z(t)$ é definida pela Equação 2.3:

$$\Delta Z(t) = Z(t) - Z(t - 1) \quad (2.3)$$

a segunda diferença é expressa pela Equação 2.4:

$$\Delta^2 Z(t) = \Delta[\Delta Z(t)] = \Delta[Z(t) - Z(t - 1)] \quad (2.4)$$

ou seja,

$$\Delta^2 Z(t) = Z(t) - 2Z(t - 1) + Z(t - 2) \quad (2.5)$$

onde, a n -ésima diferença de $Z(t)$ é dada pela Equação 2.6.

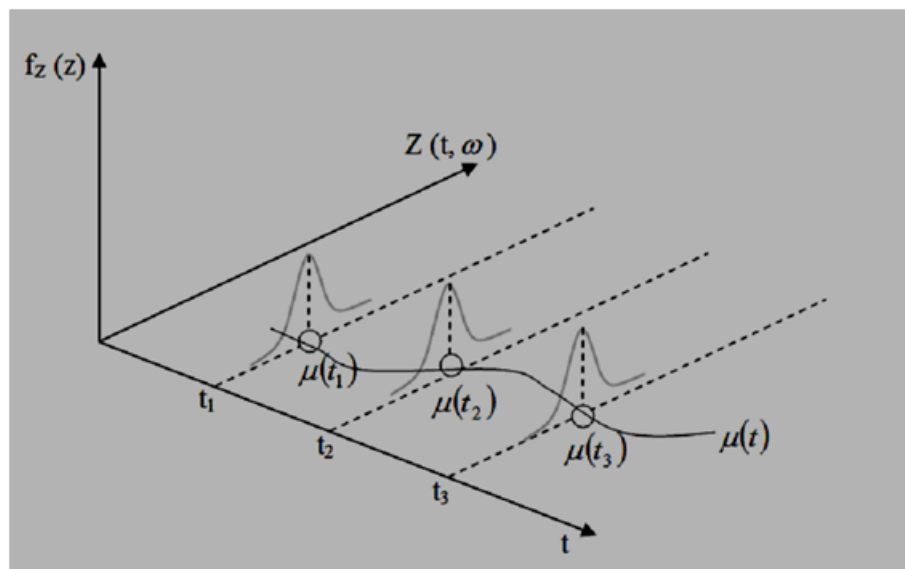
$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)] \quad (2.6)$$

De modo geral, calcula-se uma ou duas diferenças para que a série se torne estacionária.

2.3.3 Processos Estocásticos

Processos estocásticos tem grande relevância para o estudo da análise de série temporais e segundo [Pereira \(1980\)](#), processo estocástico é uma família $Z(t)$ com $t \in T$. Neste caso $Z(t)$ é uma variável aleatória e T um conjunto arbitrário. Assim, podemos entender um processo estocástico como uma família de variáveis aleatórias, definidas num mesmo espaço probabilístico Ω . Desta forma, $Z(t)$ é uma função de dois argumentos $Z(t, \omega), t \in T, \omega \in \Omega$, ver [Morettin e Toloi \(2006\)](#).

Figura 2.5: Interpretação de um processo estocástico como uma família de variáveis aleatórias

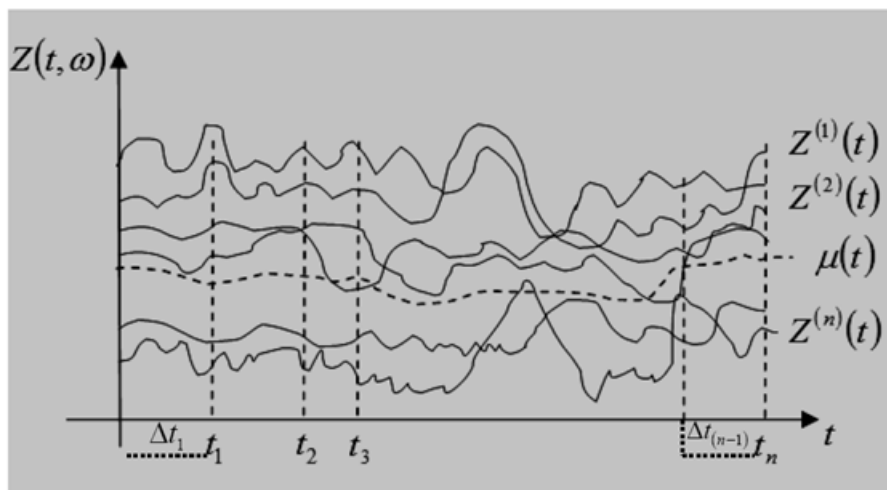


Fonte: MORETTIN (2006)

Na Figura 2.5 podemos perceber que para cada $t \in T$ teremos uma v.a. $Z(t, \omega)$ como uma distribuição de probabilidade, neste sentido, é possível que a função densidade de probabilidade $f_z(Z)$ no instante t_1 seja distinta da $f_z(Z)$ no instante t_2 . Desta forma, para cada $\omega \in \Omega$ obtemos uma função de t , isto é, uma trajetória do processo, ou ainda, uma série temporal. Assim, o conjunto de todas as trajetórias é chamado de *ensemble* e definimos por $Z(t, \omega)$ onde $Z^{(1)}(t), Z^{(2)}(t), Z^{(3)}(t) \dots Z^{(n)}(t)$ são as trajetórias (realizações) do processo. Cada realização $Z^{(j)}(t)$ é uma função do tempo t não aleatória e para cada t fixo, $Z^{(j)}(t)$ é um número real. ([MORETTIN; TOLOI, 2006](#)), propõem que uma maneira de encarar uma distribuição de probabilidade $Z(t, \omega)$ para um t fixo, ou seja, considerar

a proporção de trajetórias que passam por uma “janela” de amplitude Δ . Esta proporção é dado por $f_Z(z)\Delta$, sendo $f_Z(z)$ função de distribuição de probabilidade de $Z(t, \omega)$, ver Figura 2.6.

Figura 2.6: Um processo estocástico interpretado como uma família de trajetórias



Fonte: Morettin (2006).

Logo, o conjunto de valores $Z(t), t \in T$ será chamado de *espaço dos estados*, ϵ , do processo estocástico, e os valores de $Z(t)$ são chamados de *estados*. Caso T seja um conjunto finito ou enumerável, como $T = 1, 2, 3, \dots, N$ ou $T = \mathbb{Z}$ este processo será chamado de parâmetro discreto, se T for um intervalo de \mathbb{R} teremos um processo com parâmetro contínuo. Os espaço dos estados, ϵ 's, poderão ser discretos ou contínuos.

Segundo [Morettin e Tolo \(2006\)](#), um processo $Z(t)$ é estacionário se ele desenvolve no tempo de modo que a escolha de uma origem dos tempos não é importante, ou seja, as características de $Z(t+r)$, para todo r , são as mesmas de $Z(t)$. Existe duas formas de estacionariedade, a fraca e a estrita. Dizemos que um processo estocástico $Z(t), t \in T$ é estritamente estacionário se todas as distribuições finitos-dimensionais permanecem as mesmas sob as translações no tempo para quaisquer t_1, t_2, \dots, t_n, r de T , ou seja, $F(z_1, \dots, z_n; t_1+r, \dots, t_n+r) = F(z_1, \dots, z_n; t_1, \dots, t_n)$ e é fracamente estacionário, se e somente se:

- O valor esperado, $E[Z(t)] = \mu$, constante, para todo $t \in T$;
- A variância, $E[Z^2(t)] < \infty$, para todo $t \in T$;
- A covariância, $\gamma(t_1, t_2) = Cov(Z(t_1), Z(t_2))$ é uma função de $|t_1 - t_2|$.

Os modelos aplicados em séries cronológicas tem como fundamento o conceito de estacionariedade. Assim, os modelos considerados lineares estacionários fracos são: os modelos auto-regressivos (AR), de média móvel (MA) e misto (ARMA). Segundo [Filho \(2014\)](#), existem modelos que se destinam a descrever estatisticamente séries estacionárias e não estacionárias, como exemplo, o modelo de auto-regressivos integrado médias móveis (ARIMA).

Um modelo clássico relativo a séries temporais supõe que a série $Z(t)$ pode ser decomposta em tendência T_t , ciclo C_t , sazonalidade S_t e termo aleatório E_t , ver ([FILHO, 2014](#)).

2.3.3.1 Tendências em Séries Temporais

Dentre os vários métodos para estimar tendência T_t , os mais utilizados destinam-se a:

- ajustar uma função do tempo, com um polinômio, uma função exponencial ou outra função de t ;
- suavizar (ou filtrar) os dados da série ao redor de um ponto, para estimar a tendência naquele ponto;
- suavizar os dados da série através de sucessivos ajustes de retas de mínimos quadrados ponderados.

Uma técnica utilizada para estimação de tendência T_t em uma série temporal é o ajuste de uma curva aos valores observados da série para fazer previsões. A função exponencial, a logarítmica e a polinomial de grau m , são as mais utilizadas.

Neste trabalho, vamos nos limitar a descrever resumidamente o ajuste de um polinômio. Suponhamos que a tendência \hat{T}_t de uma série é representada pelo polinômio dado na Equação 2.7 de grau m , onde o grau m do polinômio é bem menor que o número de observações N . Desta forma, para estimar os parâmetros β_j , o método utilizado é o método dos mínimos quadrados onde é minimizada a Equação 2.8, obtendo-se os valores estimados de mínimos quadrados para $\beta_0, \beta_1, \dots, \beta_m$.

$$\hat{T}_t = \beta_0 + \beta_1 t + \dots + \beta_m t^m \quad (2.7)$$

$$f(\beta_0, \dots, \beta_m) = \Sigma(Z_t - \beta_0 - \beta_1 t - \dots - \beta_m t^m)^2 \quad (2.8)$$

Assim, podemos realizar previsões para valores futuros de uma série por meio de um modelo estimado através do método dos mínimos quadrados para T_t , utilizando \hat{T}_t como estimador linear para série.

Para [Morettin e Tolo \(2006\)](#), outro procedimento utilizado para eliminar a tendência de uma série temporal é o procedimento de tomarmos as diferenças, utilizando-se a Equação 2.9.

$$\Delta Z_t = Z_t - Z_{t-1} \quad (2.9)$$

2.3.3.2 Sazonalidade

Sazonalidade é caracterizada por padrões cíclicos de flutuações que se repetem em intervalos relativamente constantes de tempo, ver [Filho \(2014\)](#). Portanto, podemos denominar sazonalidade como uma série temporal em que todos os eventos ocorrem em intervalos constantes e regulares. Desta forma, consideremos uma série temporal como a adição de três componentes, isto é, $Z_t = T_t + S_t + a_t$ com $t = 1, 2, \dots, N$, T_t e S_t representam a tendência e a sazonalidade, respectivamente e a_t componente aleatória, de média zero e variância constante σ_a^2 . O objetivo aqui é estimar S_t e construir uma série com uma sazonalidade ajustada, para isso, utilizamos \hat{S}_t como uma estimativa de S_t através da Equação 2.10 que é a série sazonalmente ajustada.

$$Z_t^{SA} = Z_t - \hat{S}_t \quad (2.10)$$

Quando estimamos a componente sazonal estamos cometendo erro de ajuste sazonal, dado pela Equação 2.11. Para esse ajuste em uma série temporal, é de suma importância minimizar $E(\delta_i^2)$ que é o erro de ajustamento sazonal.

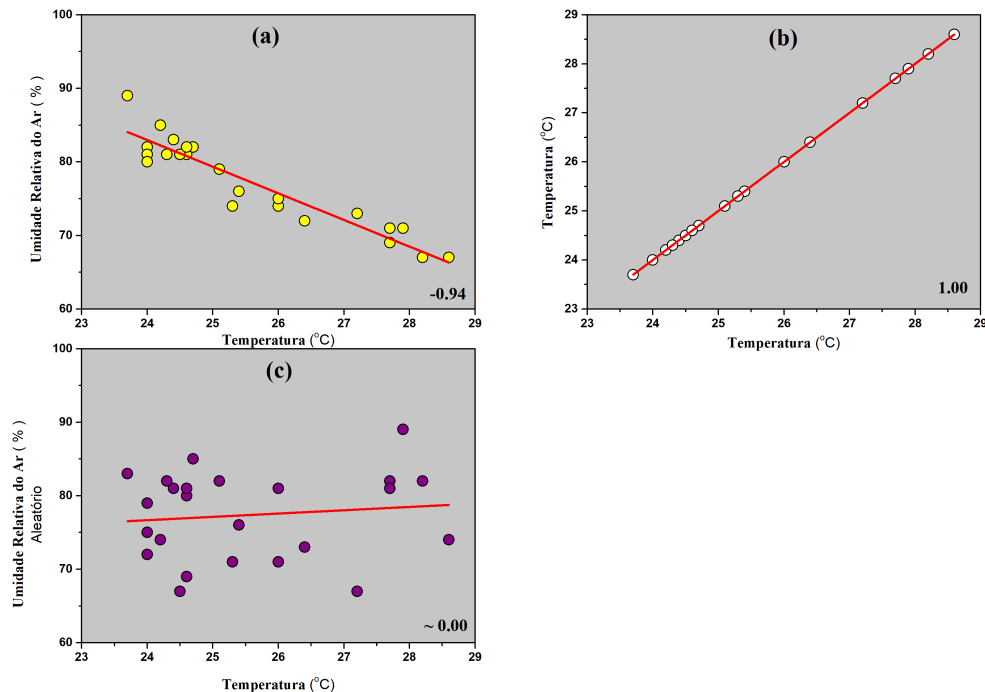
$$\delta_t = S_t - \hat{S}_t \quad (2.11)$$

Quando a série temporal apresenta sazonalidade estocástica, os métodos mais utilizados para a sua análise são, método de regressão e método da média móvel.

2.4 Coeficiente de Correlação de Pearson

Segundo Filho (2014), é possível identificar e mensurar a relação entre duas ou mais variáveis e, em estatística, esse relacionamento é denominado de correlação cruzada. O diagrama de dispersão é uma ferramenta utilizada para análise das correlações cruzadas lineares, é a representação dos pares de valores em um sistema cartesiano, ver exemplo na Figura 2.7. As análises feitas através dos diagramas de dispersão tendem a conclusões subjetivas e devido a este fato métodos mais robustos tornam-se necessários.

Figura 2.7: Diagrama de Dispersão referente a estação de Salvador (BA) em 24/06/2019. (a) Umidade Relativa do Ar \times Temperatura do Ar, (b) Temperatura do Ar \times Temperatura do Ar e (c) Umidade Relativa do Ar (aleatório) \times Temperatura do Ar



O coeficiente de correlação de Pearson, também conhecido como correlação produto-momento, tem com objetivo mensurar o grau e a direção (se positiva ou negativa) da correlação entre duas variáveis Pearson (1895). O coeficiente de Pearson, σ_{xy} , é a razão entre a covariância e a raiz quadrada do produto das variâncias de x e de y definido pela expressão 2.12:

$$\sigma_{xy} = \frac{Cov(x, y)}{\sqrt{Var(x)} \cdot \sqrt{Var(y)}} \quad (2.12)$$

em que a covariância entre x e y é definida pela expressão 2.13:

$$Cov(x, y) = \sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y}) \quad (2.13)$$

e a variância de x e de y são definidas pelas expressões 2.14 e 2.15 respectivamente:

$$Var(x) = \sum_{i=1}^N (x_i - \bar{x})^2 \quad (2.14)$$

$$Var(y) = \sum_{i=1}^N (y_i - \bar{y})^2 \quad (2.15)$$

em que as expressões 2.16 e 2.17 são as médias de x e de y , respectivamente.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.16)$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (2.17)$$

O Coeficiente de Pearson tem como propriedade:

- $-1 \leq \sigma_{xy} \leq 1$ é o campo de variação do coeficiente de Pearson;
- Se permutarmos os valores de x e de y , o valor de σ_{xy} não será alterado.

A interpretação do coeficiente de correlação de Pearson dependerá do valor numérico e do sinal que ele assume. Classificamos a correlação cruzada por Pearson quanto ao valor e ao sinal da seguinte forma, correlação linear:

- **positiva**, quando o valor de σ_{xy} estiver entre 0 e 1;
- **positiva perfeita**, quando o valor de σ_{xy} for igual a 1;
- **negativa**, quando o valor de σ_{xy} estiver entre -1 e 0;

- **negativa perfeita**, quando o valor de σ_{xy} for igual -1;
- **nula**, quando não houver correlação linear entre as variáveis x e y .

2.5 Outros Métodos de Análise de Séries Temporais Estacionárias

2.5.1 Função de Auto-correlação - FAC

A função de auto-correlação, definida pela Equação 2.18, é a covariância amostral da série na defasagem k , dividida pela variância amostral da série, em que N denota o tamanho da série e \bar{X} o valor médio das observações.

$$\sigma_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{X}) \cdot (x_{t+k} - \bar{X})}{\sum_{t=1}^N (x_t - \bar{X})^2} \quad (2.18)$$

A **FAC** tem como objetivo medir a auto-correlação entre as observações, defasadas por períodos de tempo k de uma série temporal com N observações, medindo assim a interdependência existente entre elas. Devido ao fato de que tanto a covariância como a variância são mensuradas nas mesmas unidades de medida, σ_k , é um número adimensional com a seguinte variação, $-1 \leq \sigma_k \leq 1$, em que:

- $0 < \hat{p}_k \leq 1$ a ST possui autocorrelação positiva;
- $-1 \leq \hat{p}_k \leq 0$ a ST possui autocorrelação negativa;
- $\hat{p}_k = 0$ a ST não tem autocorrelação.

Segundo [Box, Jenkins e Reinsel \(2008\)](#), estas correlações devem ser calculadas com variações de $k = 1, 2, \dots, j$, em que j não deve exceder o valor de $\frac{N}{4}$. Desta forma, podemos representar $\sigma_k \times k$ graficamente através de um correlograma amostral, procedimento estatístico padrão muito utilizado para caracterizar uma série temporal. Tendo como método de análise a **FAC** em diferentes defasagens, podemos caracterizar uma série temporal como:

- estacionária, se os valores de σ_k decrescem rapidamente para zero;
- não estacionárias, se os valores de σ_k decrescem lentamente para zero;
- periódica, quando a série apresenta característica sazonal de um dado período mudando de sinal.

Através da análise da função de auto-correlação podemos também identificar a memória da série, ou seja, analisar se a ST possui memória longa ou de curto alcance. Se os valores σ_k decrescem rapidamente para zero, a série é caracterizada como de memória curta e sua dependência temporal permanece presente por um período curto, (FILHO, 2014).

2.5.2 Análise R/S

A análise R/S foi proposta por Harold E. Hurst (1880 – 1978), (HURST, 1951). Hurst foi um hidrólogo que dedicou uma parcela de sua vida no projeto de construção de uma represa para o Rio Nilo, chegando à região por volta de 1907 e permanecendo por lá quase 6 (seis) décadas. Hurst tinha como objetivo otimizar os fluxos de água do rio Nilo de modo que a represa não transbordasse nem ficasse muito vazia.

Para a construção do seu modelo, Hurst assumiu que o fluxo de águas da chuva, seguia um processo estocástico (ver Seção 2.3.3), devido à ecologia da região do Rio Nilo, que tinha um sistema com muitos graus de liberdade. Hurst testou a hipótese de que o regime de chuvas no Rio Nilo seguia um processo aleatório e criou assim uma nova análise de flutuação, em que o interesse era a medida do chamado expoente de **expoente de Hurst**, H , que servia para distinguir uma série aleatória de uma série não aleatória.

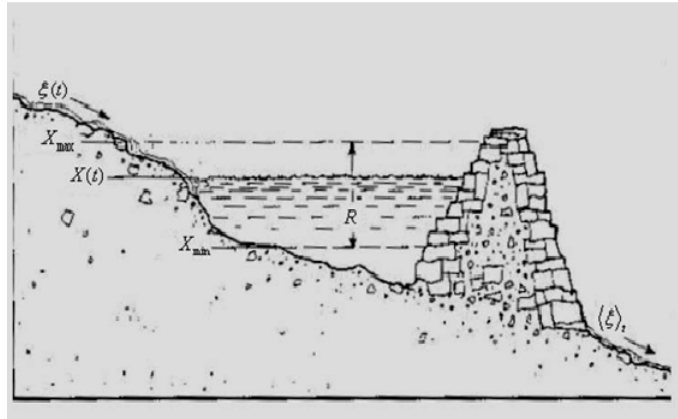
Dessa forma, Hurst descobriu que a maioria dos fenômenos naturais, tais como o fluxo dos rios, a dinâmica das chuvas e da temperatura, seguiam como leis de potência com característica de persistência. A análise da tendência e o nível do ruído poderiam ser mensurados pelo valor do expoente H . Esta análise é encontrada com mais detalhes nos livros *Long-Term Storage: An Experimental Study* de (HURST, 1951) e *Fractals* de (FEDER, 1988).

O método R/S de Hurst consiste em mensurar os volumes máximos e mínimos do reservatório (amplitude) e através do cálculo do desvio-padrão dos influxos de água em um determinado período de tempo τ . Assim, R/S é a razão entre a amplitude R_τ e o desvio padrão S_τ das vazões de água, ver Figura 2.8.

A modelagem R/S consiste nos seguintes passos:

Primeiramente, determinamos a média $\langle \xi \rangle_\tau$ do influxo de água no período τ dada pela equação 2.19.

$$\langle \xi \rangle_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi(t) \quad (2.19)$$

Figura 2.8: Modelo do reservatório de Hurst para o cálculo do $R/S(\tau)$ 

Fonte: Fractals (FEDER, 1988)

aqui ξ apresenta o influxo de água no reservatório no instante de tempo t .

Em seguida, determinamos a variação da média para v , de modo que a soma da variação acumulada da média, para os períodos de 1 à τ seja dada pela Equação 2.20, isto é,

$$X(t, \tau) = \sum_{v=1}^t [\xi(v) - \langle \xi \rangle_{\tau}] \quad (2.20)$$

em que $\langle \xi \rangle_{\tau}$ representa a média de todos os pontos da série no período τ e t representa um tempo entre o primeiro período, ($t = 1$), e o último, ($t = \tau$).

Calculamos a diferença entre o valor máximo e o mínimo de $X(t)$, no período τ (ver Modelo do reservatório de Hurst na Figura 2.9) dada pela Equação 2.21:

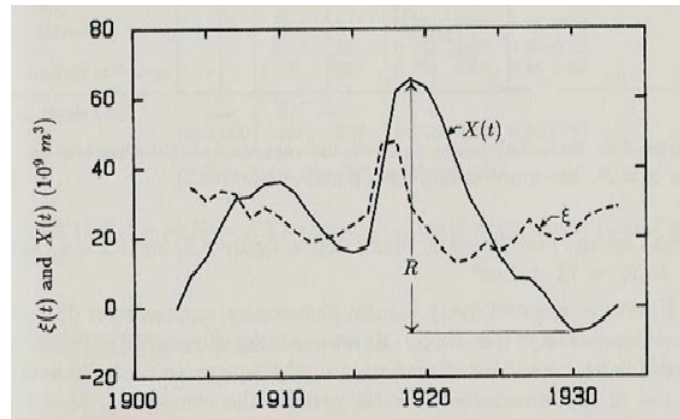
$$R(\tau) = \text{Max}[X(t, \tau)] - \text{Min}[X(t, \tau)] \quad (2.21)$$

para $1 \leq t \leq \tau$.

E, por último, dividimos $R(\tau)$ pelo seu desvio padrão $S(\tau)$, ou seja, $R/S = \frac{R(\tau)}{S(\tau)}$ em que $S(\tau)$ é denotada pela expressão 2.22.

$$S(\tau) = \sqrt{\frac{1}{\tau} \sum_{t=1}^{\tau} [\xi(t) - \langle \xi \rangle_{\tau}]^2} \quad (2.22)$$

Figura 2.9: Modelo do Reservatório de Hurst (1951)



Fonte: (FEDER, 1988)

Assim, para tamanhos de tempo distintos, a relação R/S poderá, ou não, seguir uma lei de potência em τ , isto é, ver Equação 2.23, onde H é o chamado **expoente de Hurst**.

$$R/S \sim \tau^H \quad (2.23)$$

O **Expoente de Hurst** tem uma variação entre 0 e 1, conforme a Tabela 2.3:

Tabela 2.3: Propriedade do expoente de HURST

Valor	Característica da série temporal
$H < 0,5$	comportamento anti-persistente
$H = 0,5$	descorrelacionada ou aleatória
$H > 0,5$	comportamento persistente

2.6 Métodos Estatísticos para Análise de Séries Temporais não-estacionárias

Dentre os diversos modelos para análise de séries não estacionárias, apresentamos aqui nesta seção o método **DFA** - *Detrended Fluctuation Analysis*, o **DCCA** - *Detrended Cross-Correlation Analysis*, os coeficientes correlações cruzadas ρ_{DCCA} , DMC_x^2 e $\Delta\rho_{DCCA}$, métodos que têm como objetivo quantificar e mensurar correlação cruzada em séries temporais não estacionárias em escalas de tempo distintas, que apresentados a seguir.

2.6.1 Detrended Fluctuation Analysis - DFA

O método **DFA** - *Detrended Fluctuation Analysis* (em inglês), análise das flutuações sem tendências, proposto por Peng et al. (1994), tem como objetivo identificar e mensurar a auto-correlação de longo alcance em séries temporais não-estacionárias. O método **DFA** é uma evolução da análise estatística clássica R/S , (HURST, 1951), possuindo mais de 2000 citações em diversas áreas científicas, para análise de auto-correlações de séries temporais.

<https://scholar.google.com.br/scholar?hl=pt-BR&as_sdt=0%2C5&q=detrended+fluctuation+analysis&oq=detrende>

O algoritmo para o cálculo da função de flutuação para o método **DFA** consiste nos seguintes passos, listados logo abaixo:

Passo 1: Considere uma série temporal $\{x_i\}$, com i variando de 1 a N (número total de pontos). Integra-se x_i para obter X_K , equação 2.24:

$$X_K = \sum_{i=1}^K [x_i - \langle x \rangle] \quad (2.24)$$

sendo $\langle x \rangle$ o valor médio de x_i e $K = 1, 2, \dots, N$.

Passo 2: Dividimos a série integrada X_K em $(N - n)$ caixas de igual comprimento n (escala temporal) sobrepostas. Cada caixa contém $n + 1$ valores, que começam em i e terminam em $i + n$.

Passo 3: Para cada caixa, ajustamos em X_K um polinômio de grau ≥ 1 , denotado por $\tilde{X}_{K,i}$, com $i \leq K \leq (i + n)$, ver Figura 2.10.

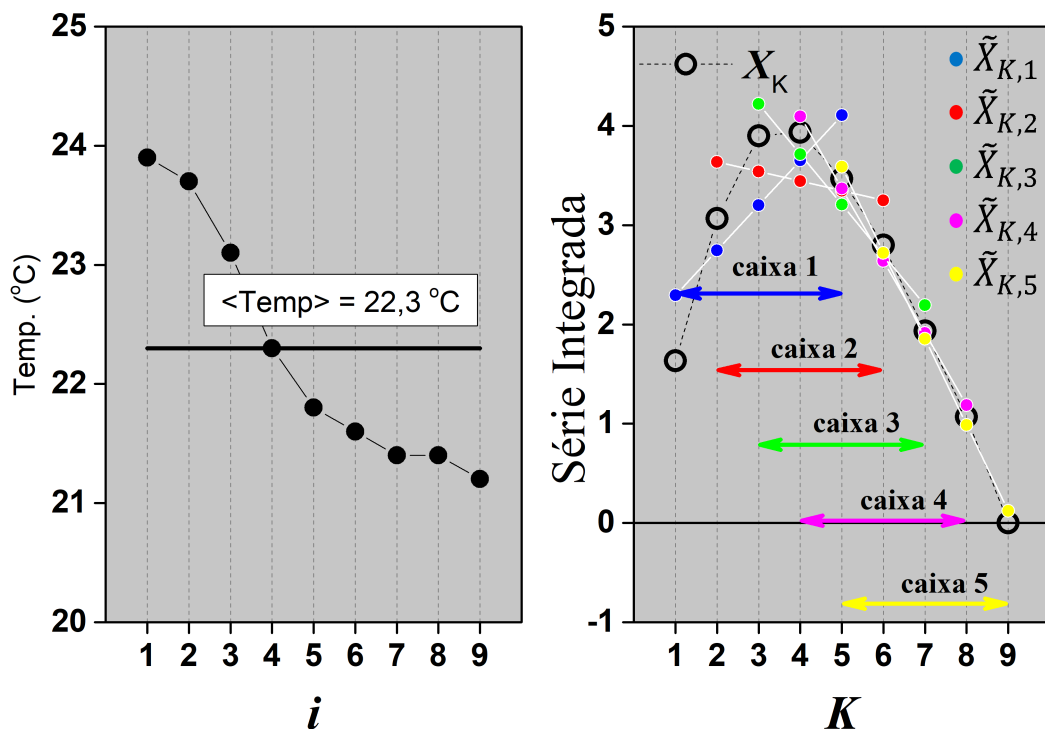
Passo 4: Subtraímos a série integrada X_K da série ajustada $\tilde{X}_{K,i}$, em cada caixa sobreposta de tamanho n , e após a subtração, calculamos a função $f_{DFA}^2(n, i)$ pela Equação 2.25.

$$f_{DFA}^2(n, i) \equiv 1/(n + 1) \sum_{K=i}^{i+n} (X_K - \tilde{X}_{K,i})^2 \quad (2.25)$$

Passo 5: Depois para uma dada escala temporal de comprimento n a função de flutuação, $F_{DFA}(n)$, é calculada por:

$$F_{DFA}(n) = \sqrt{\frac{1}{N - n} \sum_{K=1}^{N-n} f_{DFA}^2(n, i)} \quad (2.26)$$

Figura 2.10: Algoritmo para o cálculo da Função de Flutuação DFA



Passo 6: Os passos acima são repetidos para diferentes caixas de tamanho n , com $4 \leq n \leq \frac{N}{4}$ (valores mínimo e máximo de n). Por fim, verificamos a relação entre n e F_{DFA} .

Se a relação $F_{DFA} \times n$ seguir uma lei de potência, ou seja, $F_{DFA}(n) \sim n^{\alpha_{DFA}}$, então neste caso α_{DFA} será o expoente de auto-correlação de longo alcance. O expoente α_{DFA} quantifica a força empírica do sinal das auto-correlações de longo alcance (lei de potência), um parâmetro de auto-afinidade, ver [Zebende, Fernandez e Pereira \(2017\)](#). Na Tabela 2.4 apresentamos as características para o expoente α_{DFA} , segundo [Kantelhardt \(2008\)](#).

Tabela 2.4: Expoente de auto-correlação α_{DFA} e suas característica.

Expoente	Tipo de sinal
$\alpha_{DFA} < 0,5$	Anti-persistente
$\alpha_{DFA} \simeq 0,5$	Ruído branco não correlacionado
$\alpha_{DFA} > 0,5$	Persistente correlacionado de longo alcance
$\alpha_{DFA} \simeq 1,0$	Ruído 1/f
$\alpha_{DFA} > 1,0$	Não estacionário
$\alpha_{DFA} \simeq 3/2$	Ruído browniano

A vantagem do método **DFA**, em relação a outros métodos, consiste em permitir a detecção de auto-correlações de longo alcance embutidas em séries temporais aparentemente

não estacionárias e também evitar a detecção espúria de auto-correlações aparentes de longo alcance, que são um artefato de estacionariedade.

Segundo Filho (2014), através do expoente α_{DFA} , obtido por meio do método **DFA** é possível avaliar em que medida a tendência observada na série temporal passada implica em manutenção do comportamento no futuro, indicando um efeito de memória de longa duração na série.

2.6.2 Detrended Cross-Correlation Analysis - DCCA

Proposto por Podobnik e Stanley (2008), o método **DCCA** - *Detrended Cross-Correlation Analysis* (em inglês), é uma generalização do **DFA** e tem como objetivo analisar e quantificar as correlações cruzadas entre duas séries temporais distintas não estacionárias com mesmo tamanho N . A finalidade do método **DCCA** é estimar a existência ou não da correlação cruzada de longo alcance entre séries temporais isto por meio da função de covariância sem tendência $F_{DCCA}^2(n)$.

A análise da correlação cruzada de duas séries temporais, $\{x_i\}$ e $\{y_i\}$, pelo método **DCCA** é feita seguindo o algoritmo para o cálculo da função de flutuação para o método **DFA**. O algoritmo consiste nos seguintes passos, listados logo abaixo:

Passo 1: Considere duas séries temporais $\{x_i\}$ e $\{y_i\}$, com i variando de 1 a N (número total de pontos). Integra-se x_i e y_i para se obter X_K e Y_K respectivamente, equação 2.27:

$$\begin{aligned} X_k &= \sum_{i=1}^K [x_i - \langle x \rangle] \\ Y_k &= \sum_{i=1}^K [y_i - \langle y \rangle] \end{aligned} \tag{2.27}$$

sendo $\langle x \rangle$ e $\langle y \rangle$ os valores médios de x_i e y_i , com $K = 1, 2, \dots, N$.

Passo 2: Dividimos as séries integradas X_K e Y_K em $(N-n)$ caixas de igual comprimento n (escala temporal) sobrepostas. Cada caixa contém $n+1$ valores, que começam em i e terminam em $i+n$.

Passo 3: Para cada caixa, ajustamos em X_K e Y_K polinômios de grau ≥ 1 , denotado por $\tilde{X}_{K,i}$ e $\tilde{Y}_{K,i}$, com $i \leq K \leq (i+n)$.

Passo 4: Subtraímos X_K e Y_K das séries ajustadas $\tilde{X}_{K,i}$ e $\tilde{Y}_{K,i}$, em cada caixa sobreposta

de tamanho n , e então calculamos a função $f_{DCCA}^2(n, i)$ pela Equação 2.28.

$$f_{DCCA}^2(n, i) \equiv 1/(n+1) \sum_{K=i}^{i+n} (X_K - \tilde{X}_{K,i}) (Y_K - \tilde{Y}_{K,i}) \quad (2.28)$$

Passo 5: Depois para uma dada escala temporal de comprimento n a função de flutuação, $F_{DCCA}^2(n)$, é calculada por:

$$F_{DCCA}^2(n) = \frac{1}{(N-n)} \sum_{K=1}^{N-n} f_{DCCA}^2(n, i) \quad (2.29)$$

Passo 6: Os passos acima são repetidos para diferentes caixas de tamanho n , com $4 \leq n \leq \frac{N}{4}$ (valores mínimo e máximo de n).

Quando repetido o procedimento anteriormente descrito para tamanhos diferentes, verificar-se a existência ou não da lei de potência, através da expressão 2.30:

$$F_{DCCA}^2(n) \sim n^{2\lambda} \quad (2.30)$$

em que λ descreve o expoente de correlação cruzada de longo alcance.

2.6.3 Coeficiente de Correlação Cruzada ρ_{DCCA}

Desenvolvido por [Zebende \(2011\)](#), o coeficiente de correlação cruzada ρ_{DCCA} , tem como objetivo quantificar o nível de correlação cruzada entre duas séries temporais não estacionárias, definido pela razão entre a função de covariância sem tendência, F_{DCCA}^2 e a função de variância sem tendência, F_{DFA} , dada pela Equação 2.31 logo abaixo:

$$\rho_{DCCA}(n) \equiv \frac{F_{DCCA}^2(n)}{F_{DFA_x}(n) F_{DFA_y}(n)} \quad (2.31)$$

Este coeficiente de correlação cruzada é adimensional e com variação no intervalo $-1 \leq \rho_{DCCA} \leq 1$. Na Tabela 2.5 apresentamos suas características, segundo [Zebende \(2011\)](#).

Podemos também definir o coeficiente de correlação cruzada ρ_{DCCA} em seis níveis de correlação cruzada, sendo eles, três positivos e três negativos, ver Tabela 2.6, segundo

Tabela 2.5: ρ_{DCCA} , coeficiente de correlação cruzada de Zebende, e suas propriedades.

Característica	ρ_{DCCA}
1	correlação cruzada perfeita
0	sem correlação cruzada
-1	anti correlação cruzada perfeita

Silva et al. (2016).

Tabela 2.6: Possíveis intervalos para o Coeficiente de Correlação Cruzada ρ_{DCCA} .

Condição	ρ_{DCCA}
Forte	$\pm 0,666 \mapsto \pm 0,999$
Médio	$\pm 0,333 \mapsto \pm 0,666$
Fraca	$\pm 0,000 \mapsto \pm 0,333$

2.6.4 $\Delta\rho_{DCCA}$

O $\Delta\rho_{DCCA}$, proposto por Silva et al. (2016), tem como objetivo medir os efeitos da interdependência e do contágio entre variáveis financeira e é definido por meio da diferença das correlações entre o pré e o pós- crise, derivado do método ρ_{DCCA} , (ZEBENDE, 2011), cuja expressão é 2.32:

$$\Delta\rho_{DCCA}(n) \equiv \rho_{DCCA}^{after}(n) - \rho_{DCCA}^{before}(n) \quad (2.32)$$

Em que $\rho_{DCCA}^{after}(n)$ é a correlação cruzada após a crise e $\rho_{DCCA}^{before}(n)$ é a correlação cruzada antes da crise.

O $\Delta\rho_{DCCA}$ introduziu uma nova forma de analisar os efeitos da interdependência e do contágio no índice econômico.

2.6.5 Coeficiente de Correlação Cruzada Múltiplo - DMC_x^2

Proposto por Zebende e Filho (2018), o coeficiente de correlação cruzada múltiplo, DMC_x^2 , é a generalização da análise do coeficiente ρ_{DCCA} , tem como objetivo mensurar as correlações cruzadas de duas ou mais séries temporais não estacionárias e é definido pela Equação 2.33.

$$DMC_x^2(n) \equiv \rho_{y,x_i}(n)^T \rho^{-1}(n) \rho_{y,x_i}(n) \quad (2.33)$$

Em que:

$$\rho_{y,x_i}(n)^T = [\rho_{y,x_1}(n), \rho_{y,x_2}(n), \dots, \rho_{y,x_i}(n)] \quad (2.34)$$

é o vetor transposto do vetor coluna das correlações cruzadas entre a variável dependente e as variáveis independentes, e $\rho^{-1}(n)$ é matriz quadrada inversa das correlações cruzadas entre as variáveis independentes, definida como:

$$\rho^{-1}(n) \equiv \begin{pmatrix} 1 & \rho_{x_1,x_2}(n) & \rho_{x_1,x_3}(n) & \cdots & \rho_{x_1,x_i}(n) \\ \rho_{x_1,x_2}(n) & 1 & \rho_{x_2,x_3}(n) & \cdots & \rho_{x_2,x_i}(n) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{x_1,x_i}(n) & \rho_{x_2,x_i}(n) & \rho_{x_3,x_i}(n) & \cdots & 1 \end{pmatrix}^{-1} \quad (2.35)$$

com as seguintes propriedades: $\rho_{x_i,x_j}(n) = \rho_{x_j,x_i}(n)$ e $\rho_{x_i,x_i}(n) = 1$.

Tendo como exemplo duas variáveis independentes, $\{x_1\}$ e $\{x_2\}$, e uma variável dependente, $\{y\}$, o DMC_x^2 assume a seguinte expressão abaixo:

$$DMC_x^2(n) = \frac{\rho_{y,x_1}^2(n) + \rho_{y,x_2}^2(n) - 2\rho_{y,x_1}(n)\rho_{y,x_2}(n)\rho_{x_1,x_2}(n)}{1 - \rho_{x_1,x_2}^2(n)} \quad (2.36)$$

De forma análoga, podemos estender a Equação 2.36 para três ou mais variáveis independentes, em todos os casos $0 \leq DMC_x^2(n) \leq 1$.

O valor do coeficiente de correlação cruzada múltiplo, DMC_x^2 , significa que, quanto mais próximo de 1, maior é a correlação cruzada múltipla entre a variável dependente e as variáveis independentes e quanto mais próximo de 0, menor é a correlação cruzada múltipla entre as mesmas. Zebende e Filho (2018), dividiram o intervalo do coeficiente de correlação cruzada múltiplo, DMC_x^2 e propuseram 05 (cinco) níveis de intervalos com as seguintes características, (ver Tabela 2.7):

Tabela 2.7: DMC_x^2 , coeficiente de correlação cruzada múltiplo e possíveis níveis de correlação cruzada.

Características	DMC_x^2
Muito forte	0,8 \mapsto 1,0
Forte	0,6 \mapsto 0,8
Médio	0,4 \mapsto 0,6
Fraco	0,2 \mapsto 0,4
Muito Fraco	0,0 \mapsto 0,2

PUBLICAÇÕES

Neste capítulo, apresentamos as publicações realizadas neste projeto de Tese, os quais estão relacionados com os objetivos específicos listados no Capítulo 1.

3.1 ρ_{DCCA} aplicado entre a temperatura do ar e a umidade relativa: uma visão hora/hora

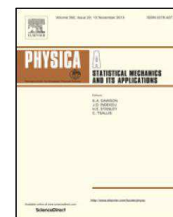
Neste artigo intitulado “ ρ_{DCCA} applied between air temperature and relative humidity: An hour/hour view ”, publicado na revista **Physica A: Statistical Mechanics and its Applications**, v. 494, p.17-26, 2018, analisamos as correlações cruzadas entre as variáveis climatológicas, temperatura do ar e a umidade relativa do ar, utilizando o coeficiente de correlação cruzada ρ_{DCCA} .

O **artigo 3.1** (em formato PDF) é apresentado a seguir.



Contents lists available at ScienceDirect

Physica A

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

ρ_{DCCA} applied between air temperature and relative humidity: An hour/hour view

G.F. Zebende^{a,d,*}, A.A. Brito^{b,c}, A.M. Silva Filho^d, A.P. Castro^e^a Department of Physics, State University of Feira de Santana, Bahia, Brazil^b Federal Institute, Paulo Afonso, Bahia, Brazil^c SENAI CIMATEC, Salvador, Bahia, Brazil^d Earth Sciences and Environment Modeling Program, State University of Feira de Santana, Bahia, Brazil^e Jorge Amado University Center, Salvador, Bahia, Brazil

HIGHLIGHTS

- We quantify cross-correlations between air temperature and relative humidity.
- Our data were available hourly by the Brazilian Institute of Meteorology (INMET).
- We present a new vision about these variables at every scales with a detrended cross-correlation map.

ARTICLE INFO

Article history:

Received 25 August 2017

Received in revised form 13 November 2017

Available online 11 December 2017

Keywords:

Air temperature

Relative humidity

Detrended cross-correlation coefficient

Contour map

ABSTRACT

In this paper we propose to study the cross-correlation between the air temperature and the air relative humidity using the DCCA cross-correlation coefficient ρ_{DCCA} . We choose four cities (meteorological stations) located in the State of Bahia (Brazil), as our case study. Our data were available hourly from the Brazilian Institute of Meteorology. Accordingly, in addition to showing that the variables have a negative cross-correlation, we present a new vision about these variables, producing the detrended cross-correlation contour map. As a result, we can see (or not) islands of cross-correlations, that depend on the local time and also the time scale involved.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Global warming can be caused by human activities (anthropogenic causes), as well as, by natural phenomena. The climate components are never in equilibrium and are constantly varying, like a complex system, see [1–4]. Thereby, any change in the components may result in a considerable variation in the climate [5]. From the perspective of climatology, mathematical models are important, because they can be applied to study the dynamics of the weather, projections of the future climate, and changes in the air temperature, among others. Meteorological data accumulated over many years, such as records of rainfall, temperature, and atmospheric composition can be analyzed by modern statistical techniques and mathematical models [6–10]. Thus, to study these complex systems we must properly define the fundamental variables, such as: air temperature, pressure, wind direction and speed, relative humidity, and many others, see chapter five of [11]. Most of these models use systems of differential equations based on the laws of physics, fluid motion, and chemistry. But, [12] recently proposed studying the cross-correlation between two time series by applying the DCCA method, designed to investigate

* Corresponding author at: Department of Physics, State University of Feira de Santana, Bahia, Brazil.

E-mail address: gfzebende@pq.cnpq.br (G.F. Zebende).

ρ_{DCCA} applied between air temperature and relative humidity: An hour/hour view

G. F. Zebende^{a,d}, A. A. Brito^{b,c}, A. M. Silva Filho^d, A. P. Castro^e

^a*Department of Physics, State University of Feira de Santana, Bahia, Brazil*

^b*Federal Institute, Paulo Afonso, Bahia, Brazil*

^c*SENAI CIMATEC, Salvador, Bahia, Brazil*

^d*Earth Sciences and Environment Modeling Program, State University of Feira de Santana, Bahia, Brazil*

^e*Jorge Amado University Center, Salvador, Bahia, Brazil*

Abstract

In this paper we propose to study the cross-correlation between the air temperature and the air relative humidity using the DCCA cross-correlation coefficient ρ_{DCCA} . We choose four cities (meteorological stations) located in the State of Bahia (Brazil), as our case study. Our data were available hourly from the Brazilian Institute of Meteorology. Accordingly, in addition to showing that the variables have a negative cross-correlation, we present a new vision about these variables, producing the detrended cross-correlation contour map. As a result, we can see (or not) islands of cross-correlations, that depend on the local time and also the time scale involved.

Keywords: Air Temperature, Relative Humidity, Detrended Cross-Correlation Coefficient, Countour Map

1. Introduction

Global warming can be caused by human activities (anthropogenic causes), as well as, by natural phenomena. The climate components are never in equilibrium and are constantly varying, like a complex system, see [1, 2, 3, 4]. Thereby, any change in the components may result in a considerable variation in the climate [5]. From the perspective of climatology, mathematical models are important, because they can be applied to study the dynamics of the weather, projections of the future climate, and changes in the air temperature,

Tabela 1: Geographical information for the stations.

Station	Lat.(S)	Long.(E)	Elev.(m)
Barreiras	12°07'29.0"	45°01'37.4"	474
Cruz das Almas	12°40'31.5"	39°05'22.5"	220
Paulo Afonso	09°22'41.2"	38°13'36.4"	255
Salvador	13°00'19.9"	38°30'20.7"	48

among others. Meteorological data accumulated over many years, such as records of rainfall, temperature, and atmospheric composition can be analyzed by modern statistical techniques and mathematical models [6, 7, 8, 9, 10]. Thus, to study these complex systems we must properly define the fundamental variables, such as: air temperature, pressure, wind direction and speed, relative humidity, and many others, see chapter five of [11]. Most of these models use systems of differential equations based on the laws of physics, fluid motion, and chemistry. But, [12] recently proposed studying the cross-correlation between two time series by applying the DCCA method, designed to investigate power-law cross-correlations between different simultaneously recorded time series in the presence of non-stationarity. However, the DCCA method does not quantify the level of cross-correlation. In order to solve this problem [13] defined the DCCA cross-correlation coefficient, ρ_{DCCA} , which allows the analysis of non-stationary time series, as in climatological variables. To show the efficiency of the DCCA cross-correlation coefficient, we analyze the relationship between the air temperature and the relative humidity. Our database comes from four stations located in Bahia State (BR), with data measured hourly by the Brazilian Institute of Meteorology (INMET).

With an area of 564,733.177 km², Bahia is one of the 27 federated Brazilian States, bordering eight other States. On the east it is bathed by the Atlantic Ocean, having the most extensive coast in Brazil. Below, in Fig 1, we present the location of the meteorological stations, and their description in Table 1.

The importance of these stations (cities) comes from the fact that they are the ones that have the best databases, and because they have the following characteristics (see [14] and Table 2):

- **Barreiras:** This is the most populous city and the most important agricultural center in the western region of Bahia. The city is pierced by the Rio Grande river, the main affluent on the left of the São Fran-

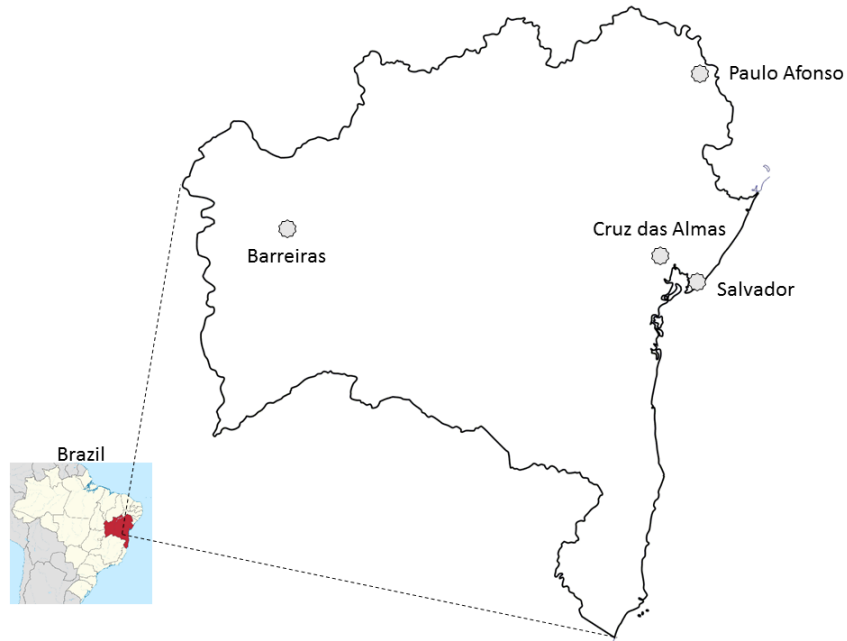


Figura 1: State of Bahia (BR) and four climatological stations located in cities of: Barreiras, Cruz das Almas, Paulo Afonso, and Salvador.

Tabela 2: Data information.

Station	Start	End	points
Barreiras	12/22/2001	09/30/2010	2047
Cruz das Almas	01/26/2003	09/30/2010	1919
Paulo Afonso	06/18/2003	09/30/2010	2065
Salvador	10/07/2000	09/30/2010	2616

cisco river (one of the most important in Brazil and South America). Barreiras stands out in national (international) agribusiness as a great producer of cotton and soybeans.

- **Cruz das Almas:** This is considered an important center in the *Recôncavo*, a geographical region located around the *Todos os Santos* bay, covering not only the coast but also the entire region of the interior surrounding the bay. The city has some agricultural research centers, such as Brazilian Company of Agricultural Research (EMBRAPA), and its industry is known for well made footwear and textiles.
- **Paulo Afonso:** This is the city where the São Francisco Hydroelectric Company (CHESF) plant complex is located. It has four large plants, PA I, PA II, PA III, and PA IV. This makes Paulo Afonso known as the “energy capital”. The city has a planned structure, the center is inside an artificial island that was built with the implantation of the channel of the PA IV plant. The main characteristic of the hydroelectric plant is to have been the first underground plant installed in Brazil, because its turbines are more than 80 meters below the level of the São Francisco river.
- **Salvador:** This is the capital and economic center of Bahia, and is notable for its food, music and architecture, with strong African and Portuguese influences. With more than 2.9 million people, it is the most populous city in the Northeast and the third most populous in Brazil.

Thus, taking into account the number of meteorological variables and the wide possibility of its applications, in this paper we restrict ourselves to studying the cross-correlation between the air temperature and the air relative humidity [15, 16, 17, 18, 19, 20]. We adopt the DCCA cross-correlation coefficient ρ_{DCCA} [13], which is defined in terms of the DFA [21] and the DCCA method [12], briefly described in Section 3.

2. Data and Descriptive Statistics

Our data was made available by INMET [22], and measured hourly over a period that ranges generally from 2001 to 2010, see Table 2. Initially we present the descriptive statistics of the data with: mean, minimum, and

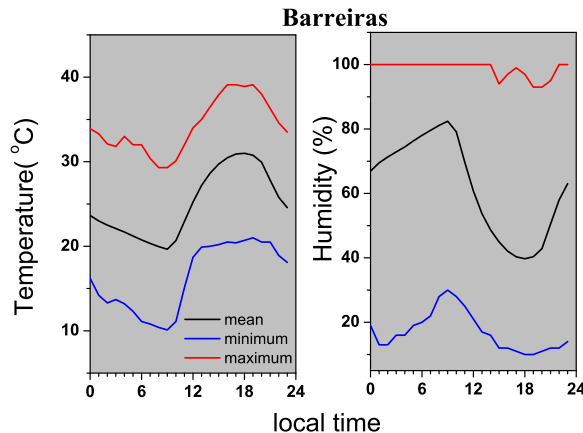


Figura 2: (color on-line) Descriptive statistic with mean, minimum, and maximum values of the air temperature and the relative humidity for Barreiras .

maximum values of the air temperature and relative humidity, presented in Figs. 2, 3, 4, and 5 as a function of the local time. We can see that there is an inverse trend between these variables, an increase (decrease) in the air temperature leads to a decrease (increase) in the relative humidity.

Specifically for Barreiras, Fig. 2, we note a smaller air temperature and a greater relative humidity at 9:00. But, at approximately 17:00, the air temperature is maximal and the relative humidity is minimal. Cruz das Almas and Paulo Afonso, Fig. 3 and Fig. 4, are similar to Barreiras. However, there is also present an inverse form for Salvador, the air temperature and the relative humidity have a lower variability, see Fig. 5. These variables can be seen also a functions of time with a contour plot, see Figs. 6, 7, 8, and 9.

In general, we can see that the air temperature ranges from 15°C to 40°C and the relative humidity from 10% to 100%. From these figures, we can conclude that there is apparently an inverse relation between the air temperature and the relative humidity, and that Salvador presents less variability, if compared to the other three cities. However, if we want to measure the cross-relation between these two variables and verify whether or not there an inverse relation between them, we should apply a statistical test that is well-adjusted for non-stationary time series analysis. Thus, the proposal in this paper is to apply the detrended cross-correlation coefficient, ρ_{DCCA} , which measures the cross-correlation between two non-stationary time series,

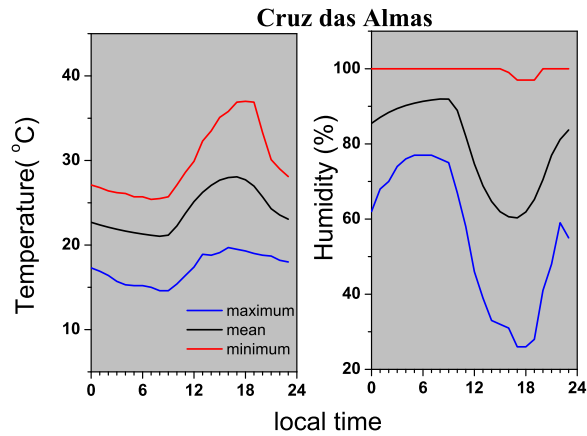


Figura 3: (color on-line) Descriptive statistic with mean, minimum, and maximum values of the air temperature and the relative humidity for Cruz das Almas.

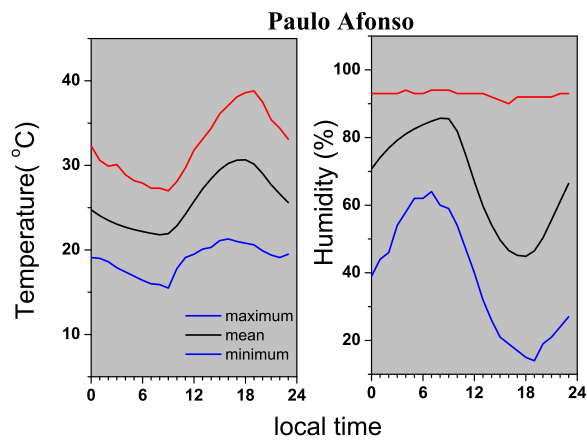


Figura 4: (color on-line) Descriptive statistic with mean, minimum, and maximum values of the air temperature and the relative humidity for Paulo Afonso.

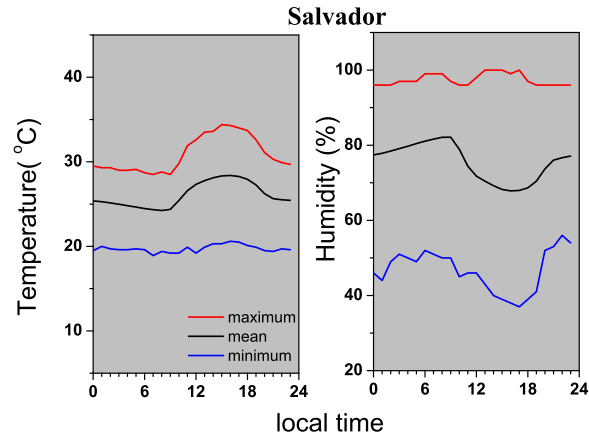


Figura 5: (color on-line) Descriptive statistic with mean, minimum, and maximum values of the air temperature and the relative humidity for Salvador.

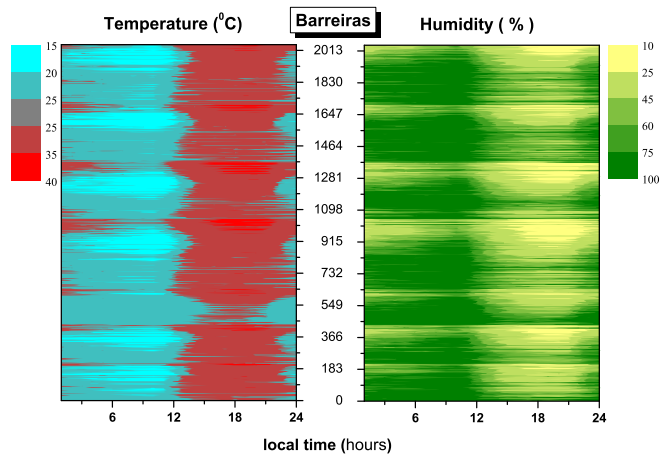


Figura 6: (color on-line) Contour plot of the air temperature and relative humidity for Barreiras. The vertical axis shows the time (in days) and horizontal axis shows the local time.

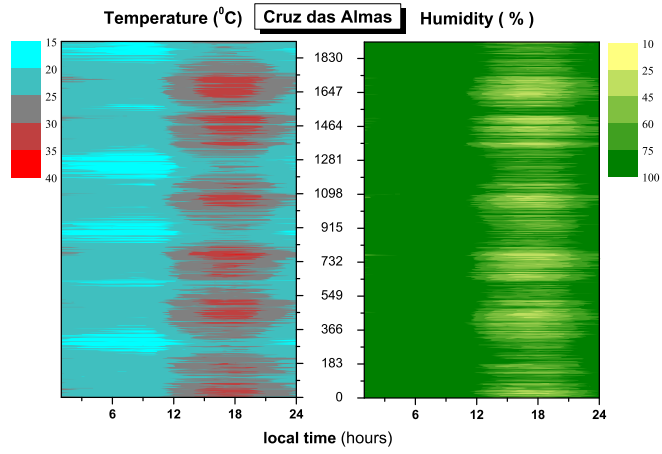


Figura 7: (color on-line) Contour plot of the air temperature and relative humidity for Cruz das Almas. The vertical axis shows the time (in days) and the horizontal axis show the local time.

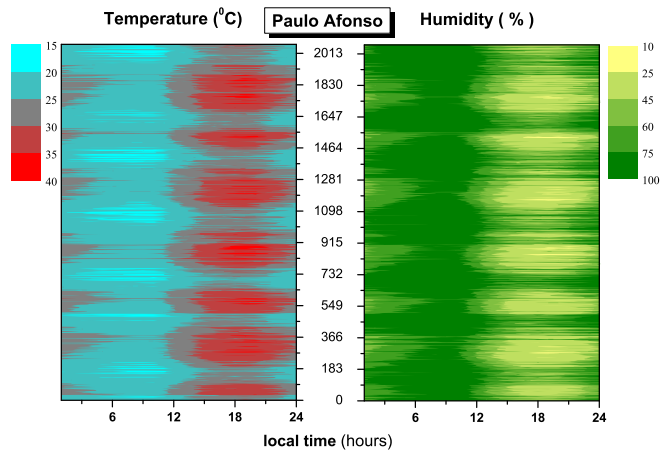


Figura 8: (color on-line) Contour plot of the air temperature and the relative humidity for Paulo Afonso. The vertical axis shows the time (in days) and the horizontal axis show the local time.

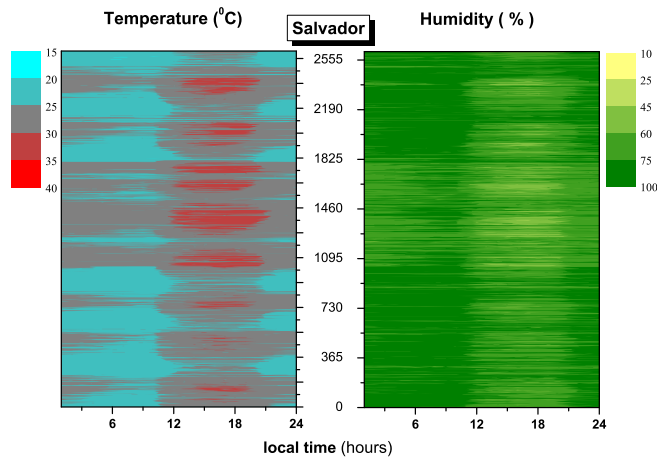


Figura 9: (color on-line) Contour plot of the air temperature and the relative humidity for Salvador. The vertical axis shows the time (in days) and the horizontal axis show the local time.

briefly described below.

3. Discussion and Methods

In time series analysis there are some well-known paths to follow, [23, 24]. Thus, if the time series exhibit complex behavior such as self-affinity, we can apply new strategies for its analysis [25, 26, 27, 28, 29, 30]. From this point of view, one of the most frequently cited methods for analyzing the time series of complex problems is detrended fluctuation analysis (DFA) [21]. This method provides a relation between $F_{DFA}(n)$ (root mean square fluctuation) and the scale n . The DFA method has been very efficient at detecting long-range auto-correlations embedded in a patch landscape and also avoiding the spurious detection of apparent long-range auto-correlations. This can be proved by a great number of applications and citations [31, 32, 33, 34]. However, if we have two time series, $\{y_i\}$ and $\{y'_i\}$, the analysis of cross-correlation can be applied, as in [35, 36, 37]. Here, we will adopt the detrended cross-correlation coefficient of [13], based on the Detrended Cross-Correlation Analysis method (DCCA) [12].

The DCCA method is a generalization of the DFA method and is based on detrended covariance. This method is designed to investigate power-law

cross-correlations between different simultaneously recorded time series in the presence of nonstationarity. Therefore, for two time series of equal length N , we compute two integrated signals $R_k \equiv \sum_{i=1}^k y_i$ and $R'_k \equiv \sum_{i=1}^k y'_i$, where $k = 1, \dots, N$. In the next step we divide the entire time series into $(N - n)$ overlapping boxes, each containing $(n + 1)$ values. For both time series, in each box that starts at i and ends at $(i + n)$, we define the local trend, $\tilde{R}_{k,i}$ and $\tilde{R}'_{k,i}$ ($i \leq k \leq i + n$), to be the ordinate of a linear least-squares fit. We define the detrended walk as the difference between the original walk and the local trend. Next we calculate the covariance of the residuals in each box $f_{DCCA}^2(n, i) \equiv 1/(n + 1) \sum_{k=i}^{i+n} (R_k - \tilde{R}_{k,i})(R'_k - \tilde{R}'_{k,i})$. Finally, we calculate the detrended covariance function by summing over all overlapping $N - n$ boxes of size n :

$$F_{DCCA}^2(n) \equiv (N - n)^{-1} \sum_{i=1}^{N-n} f_{DCCA}^2(n, i). \quad (1)$$

When only one random walk is analyzed, ($R_k = R'_k$), the detrended covariance $F_{DCCA}^2(n)$ reduces to the detrended variance $F_{DFA}^2(n)$, used in the DFA method. If there is self-affinity, then $F_{DCCA}^2(n) \sim n^{2\lambda}$. DCCA has been applied in many situations [38, 39, 40, 41, 42]. The λ exponent quantifies the long-range power-law cross-correlations and also identifies seasonality, [42], but λ does not quantify the level of cross-correlation.

To quantify the level of cross-correlation, we can apply the DCCA cross-correlation coefficient, [13], defined as the ratio between the detrended covariance function F_{DCCA}^2 and the detrended variance function F_{DFA} , i.e.,

$$\rho_{DCCA}(n) \equiv \frac{F_{DCCA}^2(n)}{F_{DFA_{\{y_i\}}}(n) F_{DFA_{\{y'_i\}}}(n)}. \quad (2)$$

Equation (2) leads us to a new scale of cross-correlation in non-stationary time series. The value of ρ_{DCCA} ranges between

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

In this case, $\rho_{DCCA} = 0$ means there is no cross-correlation, and it splits the level of cross-correlation between positive and the negative case (see Table1 in [13]). The detrended cross-correlation exponent, ρ_{DCCA} , has been tested on selected time series, simulated and real cases, and has proved to be quite robust. See the link below for more details:

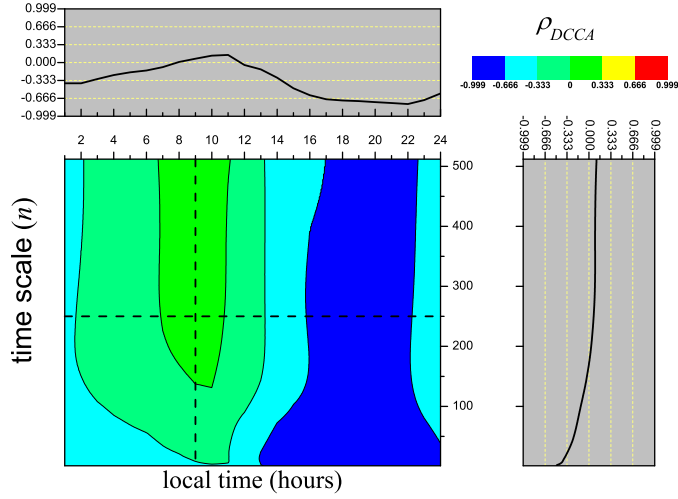


Figura 10: (color on-line) Contour plot of ρ_{DCCA} for Barreiras. Horizontal line represent the local time at 9:00 and vertical line represent $n = 250$ (time scale).

Tabela 3: Detrended cross-correlation coefficient intervals.

condition	ρ_{DCCA}
weak	$\pm 0.000 \mapsto \pm 0.333$
medium	$\pm 0.333 \mapsto \pm 0.666$
strong	$\pm 0.666 \mapsto \pm 0.999$

<http://scholar.google.com.br/scholar?oi=bibs&hl=pt-BR&cites=6351233990136499122>

4. Results

In Section 2, we can see directly that the air temperature and the relative humidity are anti cross-correlated, but how much? The answer to this question will be given by calculating ρ_{DCCA} for our variables. Here we will present the final results in a contour map (a new idea), where the x -axis represents the local time and the y -axis the time in days, see Figs. 10, 11, 12, and 13.

Thus, based on Table 3 we can define, by colors, where there is a weak, medium or strong cross-correlation (negative or positive).

According to the six levels defined (three positive and three negative) in this paper, we can associate one color to a range of values of ρ_{DCCA} . In our

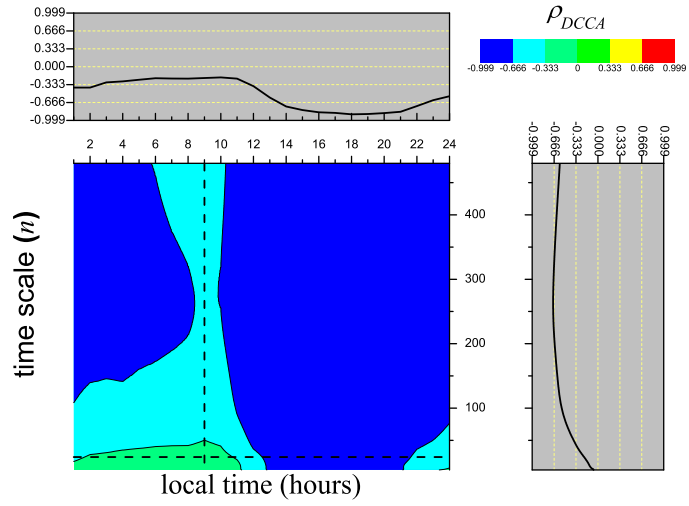


Figura 11: (color on-line) Contour plot of ρ_{DCCA} for Cruz das Almas. Horizontal line represent the local time at 9:00 and vertical line represent $n = 24$ (time scale).

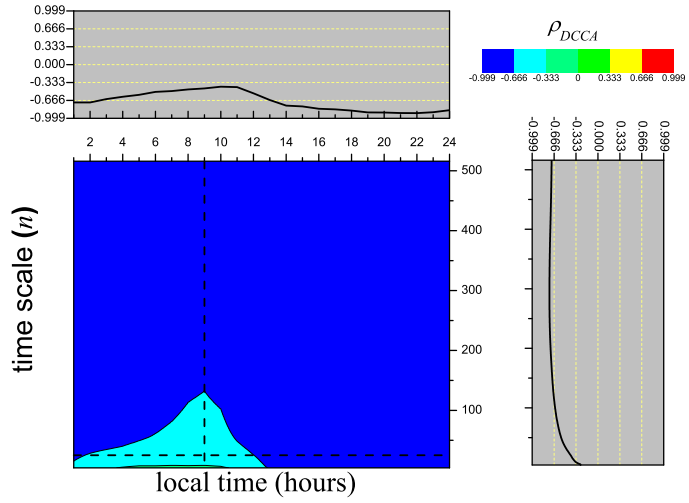


Figura 12: (color on-line) Contour plot of ρ_{DCCA} for Paulo Afonso. Horizontal line represent the local time at 9:00 and vertical line represent $n = 25$ (time scale).

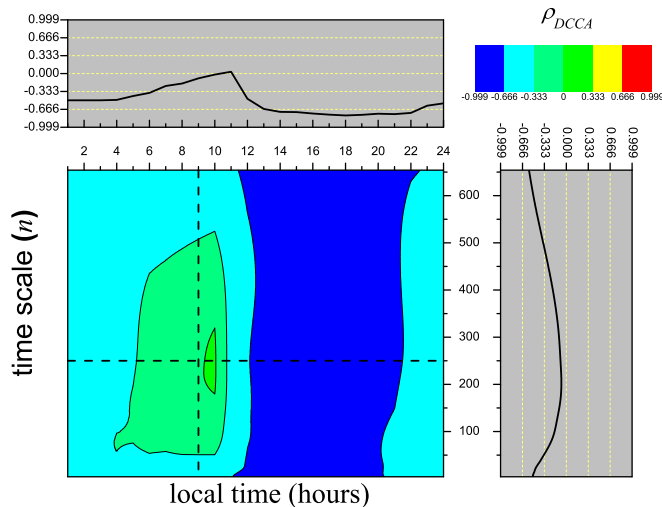


Figura 13: (color on-line) Contour plot of ρ_{DCCA} for Salvador. Horizontal line represent the local time at 9:00 and vertical line represent $n = 250$ (time scale).

case, the value of the detrended cross-correlation coefficient goes from the most negative value (blue) to the most positive (red). In this way, we have a complete climatological map using the level of cross-correlation between these two variables, the air temperature and the relative humidity.

Looking at the climatological map of ρ_{DCCA} and depending on the color position, we can initially clearly see that there are “islands” of cross-correlations, a fact not yet seen in the previous literature. For example, Fig. 10 shows these islands of cross-correlation mainly in the vertical direction (temporal scale). In the horizontal direction (local time), we can see transitions of anti cross-correlation (negative case) from weak (small portion), medium (during the morning) to strong (afternoon and evening). Furthermore, as an example, we present ρ_{DCCA} as a function of a specific scale: horizontal (local time) and vertical (time scale) dashed lines. This highlights the cross-correlation transition between the levels.

For Cruz das Almas, Fig. 11, there is a larger island of strong anti cross-correlation (blue) level, and a small portion of a weak anti cross-correlation level, located at $n \approx 20$ and between 0h and 10h. For Paulo Afonso station, Fig. 12, what is present is almost entirely an anti cross-correlation strong (blue), except for small temporal scales at the beginning of the day where medium anti-cross-correlation show up. Salvador, Fig. 13, has some simila-

rities with Barreiras, regarding vertical islands. The period between 12h and 22h shows a strong anti cross-correlation. For this station, near the coast, a small island appears around 10h and $n \approx 250$, where the value of ρ_{DCCA} is slightly positive and weak.

5. Conclusions

In this paper we can see the temporal evolution of the air temperature and the relative humidity measured hourly for four automatic meteorological stations, located in Bahia State (Brazil). Comparing their mean values with approximately 10 years of measurements, we can see that there is an inverse relation between them. However, to quantify how the variables are anti cross-correlated, we have applied the detrended cross-correlation coefficient, ρ_{DCCA} , designed to measure the cross-correlation between two non-stationary time series. But the interesting and new fact is that we were able to see the value of ρ_{DCCA} in a climatological map. From this map and taking into account some ranges of cross-correlation range, we can see directly in which regions there was strong, medium, and weak cross-correlation (negative and positive).

Specifically, for Barreiras, a strong anti cross-correlation can be seen between 16h and 22h. But, for other local times, this behavior tends to a smaller values, e.g., for mornings. If we compare Cruz das Almas with Barreiras, there is a larger portion of anti cross-correlation levels. Paulo Afonso station was the one that presented the highest anti cross-correlation (strong) levels for all time scales. Salvador, the station with the lowest variability in the air temperature and relative humidity, has a behavior similar to that of Barreiras in this sense of ρ_{DCCA} .

Finally, in this paper we could see the new climatological map based on the detrended cross-correlation coefficient between the air temperature and the relative humidity, two important climatological variables. This map can give us a complete view of these variables at all temporal scales.

6. Acknowledgments

We wish to thank the Brazilian Meteorological Institute, CNPq and FAPESB (Grant BOL 0262/2017).

7. References

- [1] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.* 74 (2002) 47–97.
- [2] J. Ladyman, J. Lambert, K. Wiesner, What is a complex system?, *European Journal for Philosophy of Science* 3 (1) (2013) 33–67.
- [3] H. Stanley, V. Afanasyev, L. Amaral, S. Buldyrev, A. Goldberger, S. Havlin, H. Leschhorn, P. Maass, R. Mantegna, C.-K. Peng, P. Prince, M. Salinger, M. Stanley, G. Viswanathan, Anomalous fluctuations in the dynamics of complex systems: from DNA and physiology to econophysics, *Physica A: Statistical Mechanics and its Applications* 224 (1) (1996) 302–321.
- [4] M. Ruth, *Smart Growth and Climate Change: Regional Development, Infrastructure and Adaptation*, Edward Elgar Publishing Limited, Cheltenham, 2006.
- [5] National geographic 2011.
URL <http://environment.nationalgeographic.com/environment/natural-disasters>
- [6] D. J. Gaffen, R. J. Ross, Climatology and trends of u.s. surface humidity and temperature, *J. Climate* 12 (1999) 811 – 828.
- [7] X. Shi, C. Lu, X. Xu, Variability and trends of high temperature, high humidity, and sultry weather in the warm season in china during the period 1961-2004, *Journal of Applied Meteorology and Climatology* 50 (1) (2011) 127–143.
- [8] L. Benyahya, D. Caissie, N. El-Jabi, M. G. Satish, Comparison of microclimate vs. remote meteorological data and results applied to a water temperature model (miramichi river, canada), *Journal of Hydrology* 380 (3-4) (2010) 247–259.
- [9] B. Podobnik, D. Horvatic, A. M. Petersen, H. E. Stanley, Cross-correlations between volume change and price change, *Proceedings of the National Academy of Sciences* 106 (52) (2009) 22079–22084.
- [10] B. Podobnik, I. Grosse, D. Horvatić, S. Ilic, C. P. Ivanov, E. H. Stanley, Quantifying cross-correlations using local and global detrending approaches, *The European Physical Journal B* 71 (2) (2009) 243–250.

- [11] W. M. O. WMO. (Ed.), Calculation of Monthly and Annual 30-Year Standard Normals, Vol. WCDP-No. 10, WMOTD/No. 341, Geneva, 1989.
- [12] B. Podobnik, H. E. Stanley, Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series, *Phys. Rev. Lett* 100 (2008) 1–4.
- [13] G. F. Zebende, Dcca cross-correlation coefficient: Quantifying level of cross-correlation, *Physica A: Statistical Mechanics and its Applications* 390 (4) (2011) 614–618.
- [14] Ibge 2017.
URL <http://www.ibge.gov.br/home/>
- [15] D. S. Gutzler, Climatic variability of temperature and humidity over the tropical western pacific, *Geophys. Res. Lett.* 19 (1595) (1992) 1598.
- [16] G. Kaser, A review of modern fluctuations of tropical glaciers, *Global and Planetary Change* 22 (1-4) (1999) 93–103.
- [17] H. Elminir, Dependence of urban air pollutants on meteorology, *Science of the Total Environment* 350 (2005) 225–237.
- [18] R. F. M. d. Oliveira, J. L. Donzele, M. L. a. T. d. Abreu, R. A. Ferreira, R. G. M. V. Vaz, P. S. Cella, Efeitos da temperatura e da umidade relativa sobre o desempenho e o rendimento de cortes nobres de frangos de corte de 1 a 49 dias de idade, *Revista Brasileira de Zootecnia* 35 (2006) 797 – 803.
- [19] M. E. Mann, Climate during the past millennium, *Weather* 56 (2001) 91.
- [20] R. Vassoler, G. Zebende, Dcca cross-correlation coefficient apply in time series of air temperature and air relative humidity, *Physica A: Statistical Mechanics and its Applications* 391 (7) (2012) 2438 – 2443.
- [21] 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.

- [22] Inmet 2017.
URL <http://www.inmet.gov.br/portal/>
- [23] P. J. Huber, *Robust Statistics*, Wiley, New York, 1981.
- [24] R. R. Wilcoxon, *Introduction to Robust Estimation and Hypothesis Testing*, 2nd Edition, Academic Press, San Diego, CA, 2005.
- [25] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H. E. Stanley, Long-range correlations in nucleotide sequences, *Nature* 356 (1992) 168 – 170.
- [26] G. F. Zebende, P. M. C. de Oliveira, T. J. P. Penna, Long-range correlations in computer diskettes, *Phys. Rev. E* 57 (3) (1998) 3311–3314.
- [27] Z.-Q. Jiang, W.-X. Zhou, Multifractal detrending moving-average cross-correlation analysis, *Phys. Rev. E* 84 (1) (2011) 016106.
- [28] B. Podobnik, Z.-Q. Jiang, W.-X. Zhou, H. E. Stanley, Statistical tests for power-law cross-correlated processes, *Phys. Rev. E* 84 (2011) 066118.
- [29] L. Kristoufek, Measuring cross-correlation between non-stationary series with dcca coefficient, *Physica A: Statistical Mechanics and its Applications* 402 (2014) 291–298.
- [30] G.-J. Wang, C. Xie, S. Chen, J.-J. Yang, M.-Y. Yang, 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) (2013) 3715–3730.
- [31] A. Király, I. M. Jánosi, Detrended fluctuation analysis of daily temperature records: Geographic dependence over australia, *Meteorology and Atmospheric Physics* 88 (3) (2005) 119–128.
- [32] E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, H.-J. Schellnhuber, Indication of a universal persistence law governing atmospheric variability, *Phys. Rev. Lett.* 81 (3) (1998) 729–732.
- [33] K. Fraedrich, R. Blender, Scaling of atmosphere and ocean temperature correlations in observations and climate models, *Phys. Rev. Lett.* 90 (10) (2003) 108501.

- [34] E. Guo, J. Zhang, H. Si, Z. Dong, T. Cao, W. Lan, Temporal and spatial characteristics of extreme precipitation events in the midwest of jilin province based on multifractal detrended fluctuation analysis method and copula functions, *Theoretical and Applied Climatology* (2016) 1–11.
- [35] Y. Wang, H. Cheng, R. L. Edwards, Y. He, X. Kong, Z. An, J. Wu, M. J. Kelly, C. A. Dykoski, X. Li, The holocene asian monsoon: Links to solar changes and north atlantic climate, *Science* 308 (5723) (2005) 854–857.
- [36] J. D. Pelletier, D. L. Turcotte, Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment, *Journal of Hydrology* 203 (1) (1997) 198 – 208.
- [37] M. P. Baldwin, T. J. Dunkerton, Propagation of the arctic oscillation from the stratosphere to the troposphere, *Journal of Geophysical Research: Atmospheres* 104 (D24) (1999) 30937–30946.
- [38] A. Machado Filho, M. F. da Silva, G. F. Zebende, Autocorrelation and cross-correlation in time series of homicide and attempted homicide, *Physica A: Statistical Mechanics and its Application* 400 (2014) 12–19.
- [39] G.-J. Wang, C. Xie, Cross-correlations between WTI crude oil market and US stock market: A perspective from econophysics, *Acta Physica Polonica-Series B Elementary Particle Physics* 43 (10) (2012) 2021–2036.
- [40] G. F. Zebende, M. F. Silva, A. M. Filho, Dcca cross-correlation coefficient differentiation: Theoretical and practical approaches, *Physica A: Statistical Mechanics and its Applications* 392 (8) (2013) 1756–1761.
- [41] D. Horvatic, H. E. Stanley, B. Podobnik, Detrended cross-correlation analysis for non-stationary time series with periodic trends, *Euro. Phys. Lett.* 94 (1) (2011) 18007.
- [42] G. Zebende, A. M. Filho, Cross-correlation between time series of vehicles and passengers, *Physica A: Statistical Mechanics and its Applications* 388 (23) (2009) 4863 – 4866.

3.2 DMC_x^2 aplicado à radiação solar global, temperatura do ar e umidade relativa do ar

Neste artigo intitulado “Detrended Multiple Cross-Correlation Coefficient applied to solar radiation, air temperature and relative humidity”, publicado na revista **Scientific Reports**, v. 9, p. 1-10, 2019, estudamos e analisamos as correlações cruzadas entre as variáveis meteorológicas, radiação solar global, temperatura do ar e umidade relativa do ar, numa perspectiva global de correlação cruzada através dos coeficientes de correlações ρ_{DCCA} e DMC_x^2 .

O [artigo 3.2](#) (em formato PDF) é apresentado a seguir.

OPEN

Detrended Multiple Cross-Correlation Coefficient applied to solar radiation, air temperature and relative humidity

Andrea de Almeida Brito^{1,2}, Heráclio Alves de Araújo³ & Gilney Figueira Zebende^{4*}

Due to the importance of generating energy sustainably, with the Sun being a large solar power plant for the Earth, we study the cross-correlations between the main meteorological variables (global solar radiation, air temperature, and relative air humidity) from a global cross-correlation perspective to efficiently capture solar energy. This is done initially between pairs of these variables, with the Detrended Cross-Correlation Coefficient, ρ_{DCCA} , and subsequently with the recently developed Multiple Detrended Cross-Correlation Coefficient, DMC_x^2 . We use the hourly data from three meteorological stations of the Brazilian Institute of Meteorology located in the state of Bahia (Brazil). Initially, with the original data, we set up a color map for each variable to show the time dynamics. After, ρ_{DCCA} was calculated, thus obtaining a positive value between the global solar radiation and air temperature, and a negative value between the global solar radiation and air relative humidity, for all time scales. Finally, for the first time, was applied DMC_x^2 to analyze cross-correlations between three meteorological variables at the same time. On taking the global radiation as the dependent variable, and assuming that $DMC_x^2 = 1$ (which varies from 0 to 1) is the ideal value for the capture of solar energy, our analysis finds some patterns (differences) involving these meteorological stations with a high intensity of annual solar radiation.

Nowadays, we know of the requirement to seek sustainable energy sources in order to preserve the environment for future generations on Earth. As is known, the Sun with its radiation can be increasingly used in order to generate energy renewably and sustainably. There are essentially two main ways of harnessing solar energy: direct electricity generation and heating water in a boiler. From the point of view of electric energy, the two systems that are the most widespread are the heliothermic system (the radiation is converted first into thermal energy and then into electric energy) and the photovoltaic system (the radiation is converted directly into electric energy). But, realistically speaking, there are several ways of generating electricity. For example, according to the Brazilian Ministry of Mines and Energy¹, the electrical matrix is made up of parts with the following percentages: Hydroelectric (63.7%), Thermal (27.2%), Wind (8.1%), and Solar (1.0%). We can see that most of the energy generation comes from hydroelectric plants, such as Itaipu (the world's largest hydroelectric power plant). Of the 27% of non-renewable thermal sources, 8.1% come from natural gas, 9.1% biomass, 6.2% oil, 2.3% coal, 1.2% nuclear, and 0.2% from others types¹.

Although hydroelectric plants use a renewable and a low-cost resource, they change the landscape, cause major deforestation, dam can break, as in Brumadinho², and in general they cause damage to the fauna and the flora also, in many cases, families are displaced from their homes³⁻⁶.

Thus, there remains an incentive for research and development in this area for the furtherance of the policy of cleaner electricity generation, in order to replace non-renewable thermal energy by a cleaner renewable source such as solar radiation. Specifically, in Brazil the incentive for solar power energy is justified by the fact that the country has large areas with a high incidence of solar radiation; the semi-arid regions of the Northeast have the highest intensity of solar radiation over the year^{7,8}. Presently, in the case of solar energy powered by photovoltaic cell panels, there is the disadvantage of their high initial cost and their low process efficiency, from 15% to 25%. Another aspect

¹Federal Institute of Education, Science and Technology of Bahia, Paulo Afonso, Bahia, Brazil. ²SENAI CIMATEC, Salvador, Bahia, Brazil. ³National Institute of Meteorology, Brasília, Brazil. ⁴State University of Feira de Santana, Bahia, Brazil. *email: gfzebende@uefs.br

Detrended Multiple Cross-Correlation Coefficient Applied to Solar Radiation, Air Temperature and Relative Air Humidity: A Global DCCA Cross-correlation Vision

Andrea de Almeida Brito^{1,2}, Heráclio Alves de Araújo³, and Gilney Figueira Zebende^{4,*}

¹Federal Institute of Education, Science and Technology of Bahia, Paulo Afonso, Bahia, Brazil

²Senai Cimatec, Salvador, Bahia, Brazil

³National Institute of Meteorology, Brasília, Brazil

⁴State University of Feira de Santana, Bahia, Brazil

*gfzebende@uefs.br

ABSTRACT

Due to the importance of generating energy sustainably, with the Sun being a large solar power plant for the Earth, we study the cross-correlations between the main meteorological variables (global solar radiation, air temperature, and relative air humidity) from a global cross-correlation perspective to efficiently capture solar energy. This is done initially between pairs of these variables, with the Detrended Cross-Correlation Coefficient, ρ_{DCCA} , and subsequently with the recently developed Multiple Detrended Cross-Correlation Coefficient, DMC_x^2 . We use the hourly data from three meteorological stations of the Brazilian Institute of Meteorology located in the state of Bahia (Brazil). Initially, with the original data, we set up a color map for each variable to show the time dynamics. After, ρ_{DCCA} was calculated, thus obtaining a positive value between the global solar radiation and air temperature, and a negative value between the global solar radiation and air relative humidity, for all time scales. Finally, for the first time, was applied DMC_x^2 to analyze cross-correlations between three meteorological variables at the same time. On taking the global radiation as the dependent variable, and assuming that $DMC_x^2 = 1$ (which varies from 0 to 1) is the ideal value for the capture of solar energy, our analysis finds some patterns (differences) involving these meteorological stations with a high intensity of annual solar radiation.

Introduction

Nowadays, we know of the requirement to seek sustainable energy sources in order to preserve the environment for future generations on Earth. As is known, the Sun with its radiation can be increasingly used in order to generate energy renewably and sustainably. There are essentially two main ways of harnessing solar energy: direct electricity generation and heating water in a boiler. From the point of view of electric energy, the two systems that are the most widespread are the heliothermic system (the radiation is converted first into thermal energy and then into electric energy) and the photovoltaic system (the radiation is converted directly into electric energy). But, realistically speaking, there are several ways of generating electricity. For example, according to the Brazilian Ministry of Mines and Energy¹, the electrical matrix is made up of parts with the following percentages: Hydroelectric (63.7%), Thermal (27.2%), Wind (8.1%), and Solar (1.0%). We can see that most of the energy generation comes from hydroelectric plants, such as Itaipu (the world's largest hydroelectric power plant). Of the 27% of non-renewable thermal sources, 8.1% come from natural gas, 9.1% biomass, 6.2% oil, 2.3% coal, 1.2% nuclear, and 0.2% from others types¹.

Although hydroelectric plants use a renewable and a low-cost resource, they change the landscape, cause major deforestation, dam can break, as in Brumadinho², and in general they cause damage to the fauna and the flora also, in many cases, families are displaced from their homes³⁻⁶.

Thus, there remains an incentive for research and development in this area for the furtherance of the policy of cleaner electricity generation, in order to replace non-renewable thermal energy by a cleaner renewable source such as solar radiation. Specifically, in Brazil the incentive for solar power energy is justified by the fact that the country has large areas with a high incidence of solar radiation; the semi-arid regions of the Northeast have the highest intensity of solar radiation over the year^{7,8}. Presently, in the case of solar energy powered by photovoltaic cell panels, there is the disadvantage of their high initial cost and their low process efficiency, from 15% to 25%. Another aspect to be considered is the environmental impact caused by the use of the silicon in the production chain of the photovoltaic cells. However, the current and recurring challenge is to establish

innovative manufacturing processes and use high-performance materials in order to obtain more efficiency in the collecting surface of the photovoltaic cells.

The efficiency of the photovoltaic cells depends not only on internal manufacturing factors, but also on external factors. External factors include shading by trees and/or clouds, rain, dust, solar radiation, air temperature, relative air humidity, and the wind speed and direction, among others. Since the external factors are not controllable, there has been research to understand their effects. Due to its strong influence on the performance of photovoltaic cells, the solar radiation and the air temperature have been the most studied external factors⁹⁻¹². There have also been many studies of the impact of the air temperature on photovoltaic cells, but there are still only a few studies that address the impact of these environmental factors in regions with a tropical climate, such as the Northeast of Brazil or that study how the variables are inter-related and interact¹³. Thus, not only should our attention be focused on energy generation: it is also important to have robust statistical tools in the area of the environmental sciences, with the purpose of analyzing how the external meteorological variables are related, as will be proposed in this paper with the DCCA multiple cross-correlation coefficient¹⁴.

Data set

Taking into consideration that the global solar radiation is measured while we have sunlight on the sensor (pyranometer), hence the data was taken hourly from 10 to 21h UTC (Coordinated Universal Time). Our data were obtained from meteorological stations managed by Brazilian Institute of Meteorology (INMET)¹⁵. Therefore, in order to study its potential for solar power we chose three meteorological stations localized in the state of Bahia (Brazil), see Fig 1. These stations are important because they

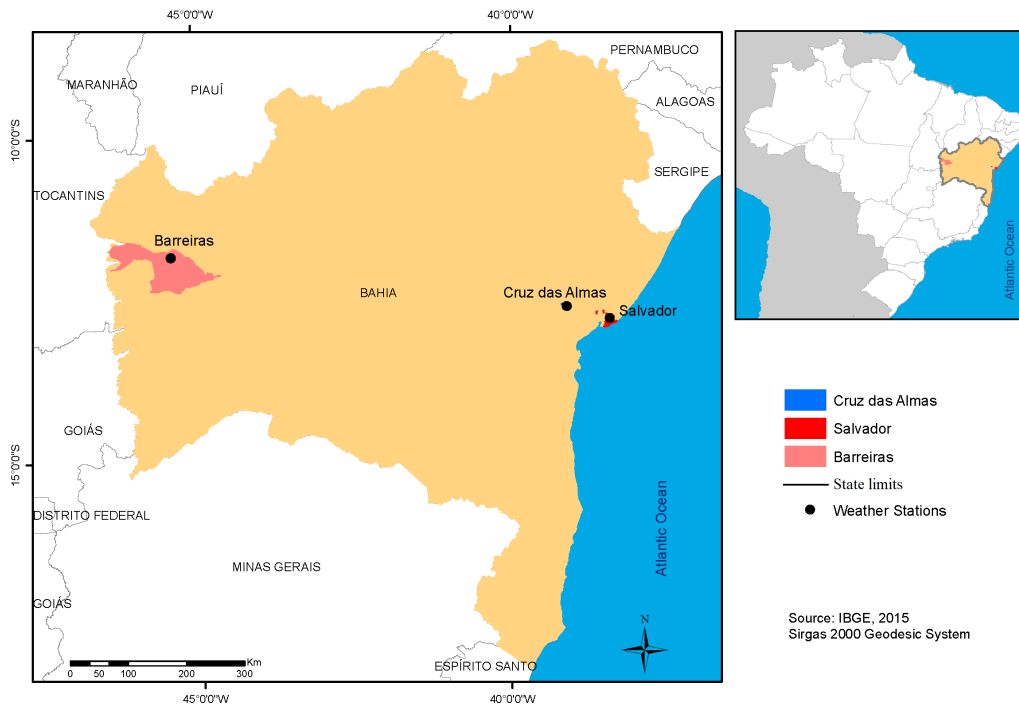


Figure 1. INMET automatic stations located in Bahia State (Brazil) at the cities of: Barreiras (Elev: 474 m), Cruz das Almas (Elev: 220 m), and Salvador (Elev: 48 m).

are the ones that have the best databases (about global solar radiation)¹⁵, and because they have the following characteristics, see Table 1.

Our first choice of analysis was to temporally order the data in a color map, with the three meteorological variables side by side, which are:

- (a) Global Solar Radiation (KJ/m^2);
- (b) Air Temperature ($^{\circ}\text{C}$);
- (c) Relative Air Humidity (%).

Table 1. Station Info: Barreiras the most populous and important agricultural center in the western region, is an important producer of cotton and soybeans. Cruz das Almas important center in the *Recôncavo*, region with some agricultural research centers such as EMBRAPA. Salvador the capital and economic center of Bahia with more than 2.9 million inhabitants.

Station	INMET code ¹⁵	Start	End	<i>N</i> points
Barreiras	A402	12/22/2001	12/11/2018	62328
Cruz das Almas	A406	10/04/2005	12/11/2018	48084
Salvador	A401	10/06/2000	12/11/2018	63264

For each station, these variables can be seen in Fig. 2 (Barreiras), Fig. 3 (Cruz da Almas), and Fig. 4 (Salvador). In these

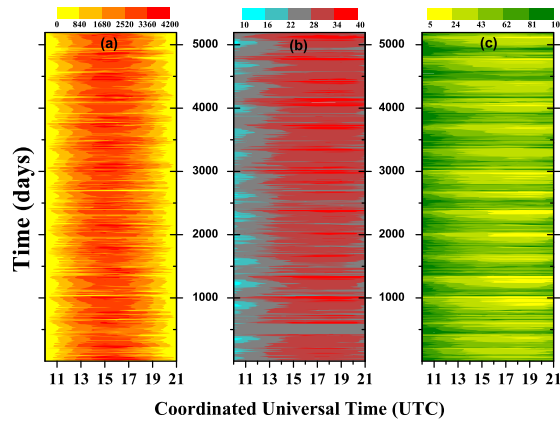


Figure 2. (color on-line) Color map for Barreiras station with: (a) global solar radiation (KJ/m^2), (b) air temperature ($^{\circ}\text{C}$), and (c) relative air humidity (%).

figures, a day (with 12h) start at 10h and ends at 21h (UTC). It can be seen that the maximum global solar radiation is usually concentrated around the peak of sunlight ($\approx 15\text{h}$). Logically, there are fluctuations in this value, depending on the season of the year or the time of day. We can also note in the color map a direct relation between the variables, but the color map is not able to robustly quantify that value; it gives us a beautiful visual display of information as to the dynamic changes in the variables. The next step is to quantitatively describe the features of the data for the whole of the period: for example, if we take the minimum and maximum values one can see:

- **BAR:** $0 \leq \text{Rad.} \leq 4051 \text{ (KJ}/\text{m}^2) \therefore 11 \leq \text{Temp.} \leq 40 \text{ (}^{\circ}\text{C)} \therefore 8 \leq \text{Hum.} \leq 100 \text{ (%)};$
- **CRUZ:** $0 \leq \text{Rad.} \leq 4148 \text{ (KJ}/\text{m}^2) \therefore 14 \leq \text{Temp.} \leq 38 \text{ (}^{\circ}\text{C)} \therefore 15 \leq \text{Hum.} \leq 100 \text{ (%)};$
- **SSA:** $0 \leq \text{Rad.} \leq 4088 \text{ (KJ}/\text{m}^2) \therefore 19 \leq \text{Temp.} \leq 35 \text{ (}^{\circ}\text{C)} \therefore 33 \leq \text{Hum.} \leq 100 \text{ (%)}.$

But, these statistics can be more refined if the seasons are considered. More details about these statistics can be seen in the results section below.

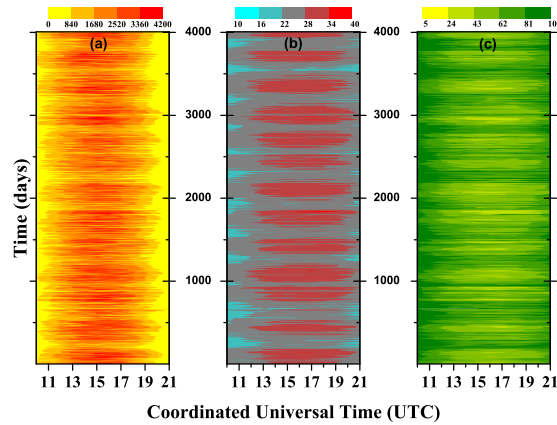


Figure 3. (color on-line) Color map for Cruz das Almas station with: (a) global solar radiation (KJ/m^2), (b) air temperature ($^{\circ}\text{C}$), and (c) relative air humidity (%).

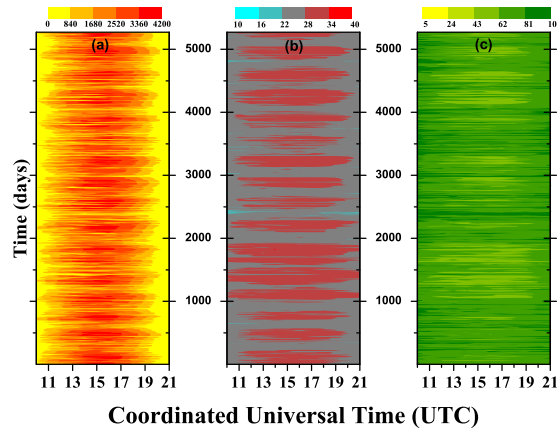


Figure 4. (color on-line) Color map for Salvador station with: (a) global solar radiation (KJ/m^2), (b) air temperature ($^{\circ}\text{C}$), and (c) relative air humidity (%).

Results

Descriptive Statistical

Initially as the results, we compute the mean values in the point of view of the annual seasons in the southern hemisphere. In this sense to simplify the climatological calculations and keep them uniform we choose the meteorological definition for seasons¹⁶ with:

- **Spring** from September/01 to November/30;
- **Summer** from December/01 to February/28;
- **Autumn** from March/01 to May/31;
- **Winter** from June/01 to August/31.

Fig. 5 present these mean values (performed at every time (UTC)) for the global solar radiation, air temperature, and relative air humidity at each season (Spring, Summer, Autumn, Winter). It is possible to observe that the radiation in the sensor has a distribution (apparently normal) characterized by a mode varying with the season. The value for maximum solar global radiation is around 15h (UTC) and this intensity depends on the location (see Table 2).

The maximum global solar radiation is at 15h, except at Salvador in the spring and summer, that is at 16h. The peak of maximum air temperature and minimum relative air humidity has a greater variation depending on the annual season. There is a

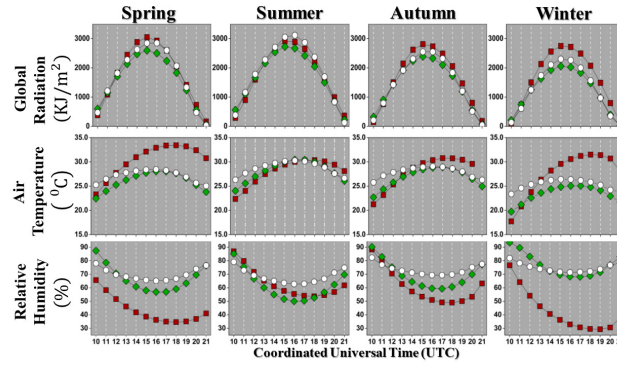


Figure 5. (color on-line) Mean values for global solar radiation, air temperature, and relative air humidity measured at every hour (UTC) for the four seasons (Spring, Summer, Autumn, Winter). The results are displayed for Barreiras (■), Cruz das Almas (◆), and Salvador (○) stations.

Table 2. Mean of global solar radiation with maximum values. First column shows the time and subsequent columns shows air temperature and relative air humidity.

	Time (UTC)	Max. Rad. (KJ/m ²)	Temp. (°C)	Hum. (%)	
BAR	15	3048	32	39	Spring
	15	2914	29	58	Summer
	15	2808	29	53	Autumn
	15	2747	30	36	Winter
CRUZ	15	2595	28	58	Spring
	15	2722	30	52	Summer
	15	2388	28	61	Autumn
	15	2055	25	70	Winter
SSA	16	2849	28	65	Spring
	16	3115	30	63	Summer
	15	2550	29	70	Autumn
	15	2299	26	72	Winter

clear inverse relation between air temperature and relative air humidity, more evidenced for Barreiras station, as shown in¹⁷. But, depending on the location (northern hemisphere) such a relation is not always true¹⁸. In this paper we also measured other moments, that are:

- Standard deviation (%): $sd(\%) = \frac{sd \cdot 100}{\langle x \rangle}$
- Skewness: $sk \equiv \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \langle x \rangle}{sd} \right)^3$
- Kurtosis: $K_e \equiv \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \langle x \rangle)^4}{\left[\frac{1}{n} \sum_{i=1}^n (x_i - \langle x \rangle)^2 \right]^2} - 3$

in this case, $\langle x \rangle$ is the mean and sd is the sample standard deviation, see Table 3 with the results. We can see that the highest relative standard deviation is for the global solar radiation ($\approx 35\%$), except in the Spring and Winter for Barreiras station, where the relative air humidity has the highest. For the skewness, we note in general values different from 0 (but close), indicating that our data-set diverges a little from the mean with positive or negative values. There is an excess Kurtosis, $K_e \neq 0$ for most of the values found in the Table 3, with $K_e < 0$ (platykurtic distribution) and with $K_e > 0$ (leptokurtic distribution). These results indicate that the meteorological time series is non-stationary. But, these descriptive statistics are well known and we want to

Table 3. Descriptive statistics of the variables with: Standard deviation (%), Skewness, and Kurtosis.

		Spring			Summer			Autumn			Winter		
		Rad.	Tem.	Hum.	Rad.	Tem.	Hum.	Rad.	Tem.	Hum.	Rad.	Tem.	Hum.
B	sd(%)	33.1	10.1	44.3	40.3	9.3	23.1	33.3	8.0	21.9	18.3	7.8	21.7
	sk	-1.0	-0.8	0.9	-0.6	-0.4	0.2	-0.6	-0.7	0.4	-1.5	-0.3	0.7
	K_c	1.0	0.7	0.1	-0.4	0.1	0.0	0.7	0.6	0.1	5.4	0.5	1.9
C	sd(%)	37.7	8.6	16.4	32.4	5.8	16.6	42.5	9.3	19.1	43.9	6.5	12.1
	sk	0.3	-0.5	0.5	-0.4	-0.8	0.8	0.0	-0.2	-0.1	1.0	-0.2	0.0
	K_c	8.2	0.0	0.1	0.1	1.5	1.0	-0.3	-0.3	-0.4	18.8	0.2	-0.1
U	sd(%)	35.8	6.7	12.0	33.0	5.6	12.0	46.6	7.0	12.6	39.0	6.0	12.0
	sk	-0.6	-0.5	0.5	-0.8	-0.6	0.6	-0.1	-0.4	0.3	0.0	-0.3	0.0
	K_c	0.2	0.4	0.6	0.5	1.1	0.6	0.3	-0.2	-0.5	0.8	0.1	0.0

propose something innovative in the study of direct or indirect relations between the main meteorological variables. With ρ_{x_i, x_j} and DMC_x^2 , we will succeed.

DCCA Cross-Correlation Coefficient ρ_{x_i, x_j}

Fig. 6 presents the values of ρ_{x_i, x_j} for the cross-correlations between global solar radiation \times air temperature, global solar radiation \times relative air humidity, and air temperature \times relative air humidity (these are particular cases of DMC_x^2 for a pair of time series). We can see clearly that the variables are related for all time scales n and for all meteorological stations, because

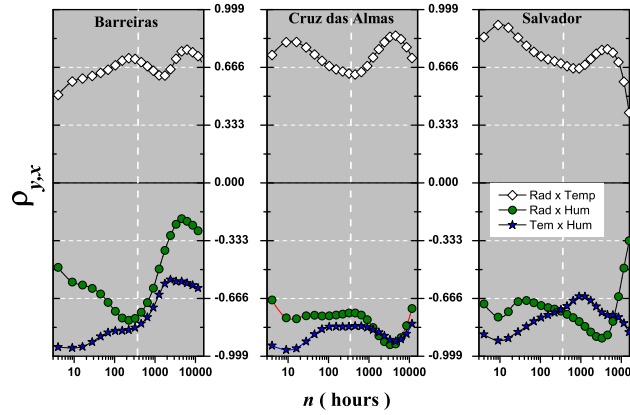


Figure 6. (color on-line) ρ_{x_i, x_j} as a function of n for all stations and combinations: Solar Radiation \times Air Temperature (\diamond), Solar Radiation \times Relative Humidity (\bullet), and Air Temperature \times Relative Humidity (\star). Vertical lines show $n = 360$ (30 days).

$\rho_{x_i, x_j} \neq 0$. Thus, due to the importance of the air temperature as a meteorological variable to solar power energy, Fig. 6 (\diamond) shows ρ_{x_i, x_j} between global solar radiation and air temperature. In this case $\rho_{x_i, x_j} > 0$, with a strong (mostly) DCCA cross-correlation for all time scales n and meteorological stations. Specifically, for $n < 360$ (30 days), Fig. 6 (\diamond) shows that ρ_{x_i, x_j} is lower for Barreiras than for with Cruz das Almas and Salvador at the same time scale. For $n = 360$, ρ_{x_i, x_j} has approximately the same value of 0.666 for these three stations. But, for $n > 360$, there is a maximum value for ρ_{x_i, x_j} around $n \approx 4400$ (≈ 365 days); this probably shows the cyclical annual effect. Analyzing the DCCA cross-correlation between the global solar radiation and the relative air humidity in Fig. 6 (\bullet) it is possible to observe that $\rho_{x_i, x_j} < 0$ for all time scales n and stations. Again, if we take as reference $n = 360$, patterns for each of the stations can be identified. For the DCCA cross-correlation between air temperature and relative air humidity, Fig. 6 (\star), $\rho_{x_i, x_j} < 0$ which agrees with¹⁷.

But, if we want to study a cross-correlation between three (global solar radiation, air temperature and relative air humidity) or more variables, we must apply a method that generalizes ρ_{x_i, x_j} , such as DMC_x^2 . The results of its application will be presented in the following.

DCCA Multiple Cross-Correlation Coefficient DMC_x^2

Fig. 7 presents $DMC_x^2(n)$ for global solar radiation, air temperature and relative air humidity at the same time. From this figure, it can be seen that the variables are related globally, because they have DCCA multiple cross-correlations that range from weak to very strong. In an intuitive way, as introduced here, this result depends on the dependent variable $\{y\}$, the station, and the time scale involved. Fig. 7 show that DMC_x^2 for $\{\text{Air Temperature}; (\text{Global Radiation} \times \text{Relative Humidity})\}$ (\bullet) behave

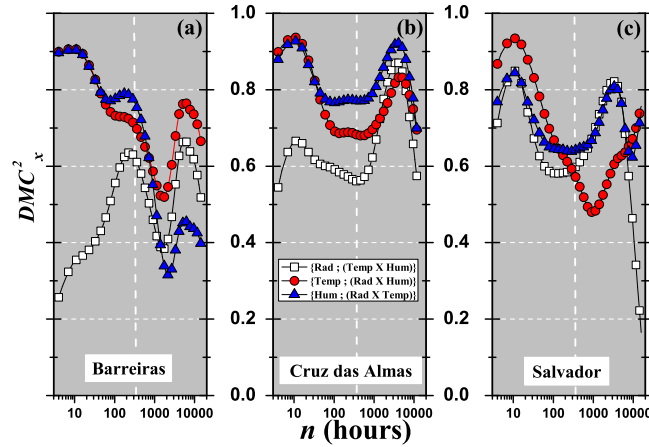


Figure 7. (color on-line) DMC_x^2 as a function of n between {Global Radiation; (Air Temperature \times Relative Humidity)} (\square), {Air Temperature; (Global Radiation \times Relative Humidity)} (\bullet), and {Relative Humidity; (Global Radiation \times Air Temperature)} (\blacktriangle) for all stations: (a) Barreiras, (b) Cruz das Almas, and (c) Salvador. Vertical line show $n = 360$ (30 days).

similarly, for $n < 360$, for all stations, but differ for $n > 360$. Up to the value $n \simeq 70$ {Air Temperature; (Global Radiation \times Relative Humidity)} and {Relative Humidity; (Global Radiation \times Air Temperature)} yield approximately the same value of DMC_x^2 , going from very strong to strong DCCA multiple cross-correlation. For the Salvador station, DMC_x^2 (\square) are closer to (\bullet and \blacktriangle) than they are for the Barreiras station, mainly for small time scales n .

Our goal here is to study how the meteorological variables are related. We believe that this can help promote the efficiency of capturing solar energy in photovoltaic cells located in a certain region with climatic variations. To this end, looking at $DMC_x^2(n)$ and taking as a dependent variable the global solar radiation, we have the following results, Fig. 8. Setting $DMC_x^2 = 1$ for

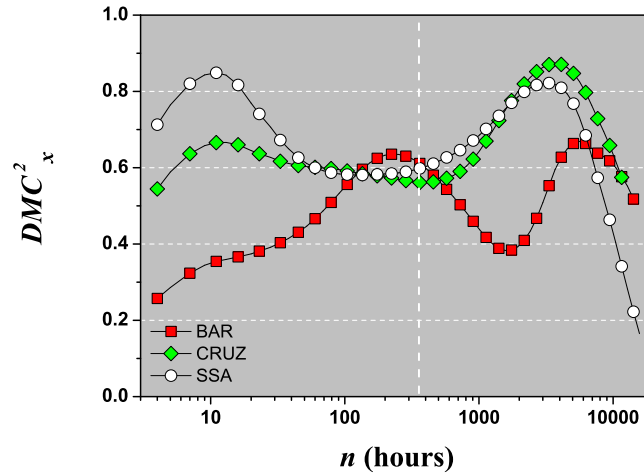


Figure 8. (color on-line) DMC_x^2 as a function of n between {Global Radiation; (Air Temperature \times Relative Humidity)} for the Barreiras (\blacksquare), Cruz das Almas (\blacklozenge), and Salvador (\circ) stations. Vertical line show $n = 360$ (30 days).

maximum efficiency in capturing solar energy for all time scales, both in short and long-term, we can analyze separately each location and observe the dependence of DMC_x^2 on the time scale and location and define in this way the potential efficiency of each location. In our case study, all stations have $DMC_x^2 > 0.2$ with levels of the multiple cross-correlation lying between weak and very strong (see Table 4). For small time scales, $n < 10$, we have, in decreasing order of their multiple cross-correlation, Salvador (\circ), Cruz das Almas (\blacklozenge), and Barreiras (\blacksquare). At $n \simeq 360$ (30 days), all stations have approximately the same value, $DMC_x^2 \simeq 0.6$. For long time scales, there is a peak of DMC_x^2 for Salvador and Barreiras at $n \approx 3652$ (one year), of approximately 0.8, whereas for Barreiras this value is below 0.6.

Discussion

Taking into account the global solar radiation, the air temperature and the relative air humidity, we have studied the cross-correlations from a global perspective, by using the multiple detrended cross-correlation coefficient, DMC_x^2 . Initially we provide, for better data visualization, the descriptive statistics of these time series. Then, with their mean values and with the capture and or generation of solar energy in view, we found the maximum global solar radiation value. It can be seen here that usually this maximum is concentrated at 15h (UTC), but with different air temperature and relative air humidity depending on the season. Logically, with standard deviation, skewness, and kurtosis, we can infer whether or not the probability distribution function approach the normal distribution. In this paper, it can be seen that depending on the season, the distributions deviate from the normal, characterizing these time series as non-stationary. But, this classical statistical analysis only takes into consideration each variable separately. To analyze the relations between them in pairs (or more) we must apply a statistical tool that has this capability.

Developed by Zebende¹⁹, Detrended Cross-correlation Coefficient, ρ_{x_i, x_j} , was constructed in order to analyze the cross-correlations between pairs of non-stationary time series. It is robust if compared, for example, with the Pearson's coefficient²⁰. In this sense, the values of ρ_{x_i, x_j} between the global solar radiation, air temperature, and relative air humidity (in pairs) depend on the geographical location and the time scale, for these three meteorological stations chosen here. There is a positive relation between global solar radiation and air temperature and an inverse relation between global solar radiation and relative air humidity.

As mentioned in the introduction, the efficiency of a photovoltaic cell depends mainly on internal factors, but that the external factors are also important, as well as their inter-relations. Our goal here was to apply the multiple DCCA cross-correlation, DMC_x^2 , to study globally the relation between three (main) variables involved in solar energy. It is noteworthy that such application has not yet been performed and this paper is the pioneer to treat together these three variables at the same time with DMC_x^2 . Thus, assuming $DMC_x^2 = 1$ to be the ideal value for solar energy capture, if we have global solar radiation as the dependent variable, than for the meteorological stations of Barreiras, Cruz das Almas, and Salvador, located in the northeast of Brazil, an area with a high intensity of annual solar radiation, we did our analysis and noticed some patterns (differences). Because DMC_x^2 is a function of the time scale n , we can determine this multiple coefficient for small, medium, and long time scales. For example, for $n = 12$ (small time scale), the Salvador station has the best value for capture solar energy according to DMC_x^2 , with very-strong value. For $n = 360$ (one month) all stations have an intermediate value for DMC_x^2 , that is, between medium and strong. But, for long time scales, Barreiras and Cruz das Almas have the best values for capture solar energy if compared to Salvador.

Finally, it is worth pointing out that in a certain way the stations are close to each other, and that a study for other stations around the planet would be very welcome. But, the purpose of this paper was to apply a new method to the analysis of multiple cross-correlations between meteorological variables in a global (innovative) way. In conclusion, as the expression for multiple correlation is quite general, other variables can be employed, such as atmospheric pressure and wind speed, among others, adding even more information to the calculation of DMC_x^2 .

Methods

For DCCA multiple cross-correlation coefficient¹⁴ presentation, we employ the DCCA cross-correlation coefficient ρ_{x_i, x_j} ¹⁹, which is defined in terms of the $F_{DFA}(n)$ ²¹ and the $F_{DCCA}^2(n)$ fluctuation function²²:

$$\rho_{x_i, x_j}(n) \equiv \frac{F_{DCCA}^2(n)}{F_{DFA\{x_i\}}(n)F_{DFA\{x_j\}}(n)}. \quad (1)$$

The DCCA cross-correlation coefficient in Eq.1 ranges between $-1 \leq \rho_{x_i, x_j} \leq 1$, and has been applied in several papers, such as^{20, 23–27}, among many others²⁸. It is possible to generalize the idea behind ρ_{x_i, x_j} to more than two variables, and such a new multiple coefficient is referred to as the DCCA multiple cross-correlation coefficient, denoted by DMC_x^{214} :

$$DMC_x^2(n) = \Phi_{y,x}(n)^T \rho^{-1}(n) \Phi_{y,x}(n) \quad (2)$$

where,

$$\Phi_{y,x}(n)^T \equiv [\rho_{y,x_1}(n), \rho_{y,x_2}(n), \dots, \rho_{y,x_i}(n)] \quad (3)$$

is the detrended cross-correlations vector between the independent variables (x_i) and the dependent variable (y), and $\rho^{-1}(n)$ is the matrix inverse of the detrended cross-correlation matrix of the independent variables, defined as

$$\rho^{-1}(n) \equiv \begin{pmatrix} 1 & \rho_{x_1,x_2}(n) & \rho_{x_1,x_3}(n) & \cdots & \rho_{x_1,x_i}(n) \\ \rho_{x_1,x_2}(n) & 1 & \rho_{x_2,x_3}(n) & \cdots & \rho_{x_2,x_i}(n) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{x_1,x_i}(n) & \rho_{x_2,x_i}(n) & \rho_{x_3,x_i}(n) & \cdots & 1 \end{pmatrix}^{-1} \quad (4)$$

This matrix is symmetric because $\rho_{x_i,x_j}(n) = \rho_{x_j,x_i}(n)$, $\rho_{x_i,x_i}(n) = 1$, and $0 \leq DMC_x^2(n) \leq 1$, see Table 4 for more details about the levels. For example, if the global solar radiation is $\{y\}$, the air temperature is $\{x_1\}$, and the relative air humidity is $\{x_2\}$,

Table 4. DCCA multiple cross-correlation coefficient levels.

condition	DMC_x^2
very-weak	0.0 \mapsto 0.2
weak	0.2 \mapsto 0.4
medium	0.4 \mapsto 0.6
strong	0.6 \mapsto 0.8
very-strong	0.8 \mapsto 1.0

then

$$DMC_x^2(n) = \frac{\rho_{y,x_1}^2(n) + \rho_{y,x_2}^2(n) - 2\rho_{y,x_1}(n)\rho_{y,x_2}(n)\rho_{x_1,x_2}(n)}{1 - \rho_{x_1,x_2}^2(n)} \quad (5)$$

References

1. Ministério de Minas e Energia. <http://www.mme.gov.br/> (2019).
2. Brumadinho dam collapse (guardian news on youtube). <https://www.youtube.com/watch?v=sKZUZQytads> (2019).
3. Marta-Almeida, M., Mendes, R., Amorim, F. N., Cirano, M. & ao M. Dias, J. Fundão Dam collapse: Oceanic dispersion of River Doce after the greatest Brazilian environmental accident. *Mar. Pollut. Bull.* **112**, 359 – 364 (2016).
4. do Carmo, F. F. *et al.* Fundão tailings dam failures: the environment tragedy of the largest technological disaster of Brazilian mining in global context. *Perspectives Ecol. Conserv.* **15**, 145 – 151 (2017).
5. Guerra, M. B. B. *et al.* Post-catastrophe analysis of the Fundão Tailings Dam Failure in the Doce River System, Southeast Brazil: Potentially Toxic Elements in Affected Soils. *Water Air Soil Pollut* 228 (2017).
6. Omachi, C. Y. *et al.* Atlantic forest loss caused by the worlds largest tailing dam collapse (Fundão Dam, mariana, brazil). *Remote. Sens. Appl. Soc. Environ.* **12**, 30 – 34 (2018).
7. Tiba, C. & Fraidenraich, N. *Atlas Solarimétrico do Brasil: banco de dados solarimétricos* (Recife, UFPE, 2000).
8. for Space Research, I. N. I. *LABREN - Laboratory of Modeling and Studies of Renewable Energy Resources* (INPE, 2019).
9. Rehman, S. & Mohandes, M. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy* **36**, 571 – 576 (2008).
10. Bindi, M. & Miglietta, F. Estimating daily global radiation from air temperature and rainfall measurements. *Clim. Res.* **1**, 117–124 (1991).
11. Almorox, J., Hontoria, C. & Benito, M. Models for obtaining daily global solar radiation with measured air temperature data in madrid (spain). *Appl. Energy* **88**, 1703 – 1709 (2011).
12. Fernandes, D. S., Heinemann, A. B., de Oliveira Amorim, A. & da Paz, R. L. F. Estimativa da radiação solar global com base em observações de temperatura para o estado de goiás. *Revista Brasileira de Meteorol.* **33**, 558 – 566 (2018).
13. Cantor, G. A. R. *Influência dos fatores climáticos no desempenho de módulos fotovoltaicos em regiões de clima tropical*. Master's thesis, UFPB (2017).
14. Zebende, G. & da Silva Filho, A. Detrended multiple cross-correlation coefficient. *Phys. A: Stat. Mech. its Appl.* **510**, 91 – 97 (2018).

15. National Institute of Meteorology (inmet). <http://www.inmet.gov.br/portal/> (2019).
16. Seasons: Meteorological and astronomical (time and date as 1995–2019). <https://www.timeanddate.com/calendar/aboutseasons.html> (2019).
17. Zebende, G., Brito, A., Filho, A. S. & Castro, A. ρ_{DCCA} applied between air temperature and relative humidity: An hour/hour view. *Phys. A: Stat. Mech. its Appl.* **494**, 17 – 26 (2018).
18. Vassoler, R. & Zebende, G. DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity. *Phys. A: Stat. Mech. its Appl.* **391**, 2438 – 2443 (2012).
19. Zebende, G. F. DCCA cross-correlation coefficient: Quantifying level of cross-correlation. *Phys. A: Stat. Mech. its Appl.* **390**, 614–618 (2011).
20. Kristoufek, L. Measuring cross-correlation between non-stationary series with DCCA coefficient. *Phys. A: Stat. Mech. its Appl.* **402**, 291–298 (2014).
21. Peng, C.-K. *et al.* Mosaic organization of DNA nucleotides. *Phys. Rev. E* **49**, 1685–1689 (1994).
22. Podobnik, B. & Stanley, H. E. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Phys. Rev. Lett* **100**, 1–4 (2008).
23. ZHAO, X., SHANG, P. & HUANG, J. Several fundamental properties of DCCA cross-correlation coefficient. *Fractals* **25**, 1750017 (2017).
24. Zebende, G. F., Silva, M. F. & Filho, A. M. DCCA cross-correlation coefficient differentiation: Theoretical and practical approaches. *Phys. A: Stat. Mech. its Appl.* **392**, 1756–1761 (2013).
25. Wang, G.-J., Xie, C., Chen, S. & Han, F. Cross-correlations between energy and emissions markets: New evidence from fractal and multifractal analysis. *Math. Probl. Eng.* **2014**, 1–13 (2014).
26. Marinho, E., Sousa, A. & Andrade, R. Using detrended cross-correlation analysis in geophysical data. *Phys. A: Stat. Mech. its Appl.* **392**, 2195 – 2201 (2013).
27. Kwapien, J., Oświęcimka, P. & Drożdż, S. Detrended fluctuation analysis made flexible to detect range of cross-correlated fluctuations. *Phys. Rev. E* **92**, 052815 (2015).
28. DCCA cross-correlation coefficient: quantifying level of cross-correlation (citations on academic google). <https://scholar.google.com.br/scholar?oi=bibs&hl=pt-BR&cites=6351233990136499122> (2019).).

Acknowledgements

A. A. Brito thanks the FAPESB (Fundação de Amparo à Pesquisa do Estado da Bahia) (Grant BOL 0262/2017) and G. F. Zebende thanks the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) (Grant 304362/2017-4), Brazilian agencies. We all thanks the National Institute of Meteorology for the availability of the data, the UFPE Renewable Energy Center for the rich discussions and the Earth Science and Environment Modeling Program (UEFS) by the scientific support. We also thanks Dra. Jocimara Lobão of PPGM UEFS for help in the Fig. 1.

Author contributions statement

A. A. Brito, H. A. de Araújo and G. F. Zebende did a data mining, analysed the results and reviewed all the manuscript.

Additional information

Competing interests The authors declare no competing interests.

3.3 Análise de Correlação Cruzada em séries temporais com partes removidas

Neste artigo intitulado “DCCA cross-correlation analysis in time-series with removed parts”, publicado na revista **Physica A: Statistical Mechanics and its Applications**, v. 545, p. 123472, 2020, analisamos o efeito de cortes e remoções de partes de uma série temporal, com memória de longo alcance, através dos métodos **DFA**, **DCCA** e do coeficiente de correlação cruzada ρ_{DCCA} com o intuito de mostrar que para até 50 % das partes removidas, comparadas com as séries temporais iniciais, não há alterações nos resultados finais para as correlações cruzadas para os métodos analisados.

O **artigo 3.3** (em formato PDF) é apresentado a seguir.



DCCA cross-correlation analysis in time-series with removed parts

G.F. Zebende ^{a,*}, A.A. Brito ^{b,c}, A.P. Castro ^d

^a State University of Feira de Santana, Bahia, Brazil

^b Federal Institute of Education, Science and Technology, Paulo Afonso, Bahia, Brazil

^c Senai, Salvador, Bahia, Brazil

^d Jorge Amado University Center, Salvador, Bahia, Brazil



ARTICLE INFO

Article history:

Received 22 September 2019

Received in revised form 4 November 2019

Available online 15 November 2019

Keywords:

DFA

DCCA

Detrended cross-correlation coefficient

Long-range memory time-series

ARFIMA process

ABSTRACT

In this paper we analyze the effect of removing pieces in time-series with long range-memory by DFA method and detrended cross-correlation coefficient. To achieve this purpose, initially simulated time-series are produced by ARFIMA process (long time dependence). From these simulated time-series cuts and removals are produced. The results show that for up to 50% of removed parts, compared to the original time-series, there is no change in the final results for detrended auto and cross-correlations. Therefore, with this paper we show that the DFA method and the detrended cross-correlation coefficient are robust for time-series analysis even for time-series with removed parts. This result ensures that these methods can be applied to real time-series, which in many cases lacks measurement for a variety of reasons and causes.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Usually in the study of the real time-series, depending on the time scale involved, data loss proves to be an unfortunate reality for the researcher, for example: an automatic weather station is equipped with sensors to measure some variables as air temperature, wind speed/direction, relative humidity, atmospheric pressure, solar radiation, among others. In this station data is automatically collected at every minute, but sometime for some unexpected reason the operation of data collection can stop, impairing future statistics analysis. Although, even in the absence of some data, time-series analyzes by statistical models are performed [1,2]. Taking this fact into consideration, this paper study the influence of removed parts in time-series with long-range memory by the application of the Detrended Fluctuation Analysis, DFA [3], and the Detrended Cross-Correlation Coefficient, ρ_{DCCA} [4], see also [5,6]. As a case of study, two time-series with long-memory process (autocorrelation functions decay slowly) were simulated. This stochastic process is said Auto Regressive Fractionally Integrated Moving Average, ARFIMA, and each variable depends not only on its own past, but also on the past values of the other variable [7–10], in other words:

$$\begin{aligned} x_i &= W \sum_{m=1}^{\infty} a_m(\delta_1) x_{i-m} + (1-W) \sum_{m=1}^{\infty} a_m(\delta_2) y_{i-m} + \varepsilon_{xi}, \\ y_i &= (1-W) \sum_{m=1}^{\infty} a_m(\delta_1) x_{i-m} + W \sum_{m=1}^{\infty} a_m(\delta_2) y_{i-m} + \varepsilon_{yi}. \end{aligned} \quad (1)$$

* Corresponding author.

E-mail address: gfzebende@uefs.br (G.F. Zebende).

DCCA cross-correlation analysis in time-series with removed parts

G. F. Zebende^a, A. A. Brito^{b,c}, A. P. Castro^d

^a*State University of Feira de Santana, Bahia, Brazil*

^b*Federal Institute of Education, Science and Technology, Paulo Afonso, Bahia, Brazil*

^c*Senai, Salvador, Bahia, Brazil*

^d*Jorge Amado University Center, Salvador, Bahia, Brazil*

Abstract

In this paper we analyze the effect of removing pieces in time-series with long-memory by DFA method and detrended cross-correlation coefficient. To achieve this purpose, initially simulated time-series are produced by ARFIMA process (long time dependence). From these simulated time-series cuts and removals are produced. The results show that for up to 50% removed parts, compared to the original time-series, there is no change in the final results for detrended auto and cross-correlations. Therefore, with this paper we show that the DFA method and the detrended cross-correlation coefficient are robust for time-series analysis even for time-series with removed parts. This result ensures that these methods can be applied to real time-series, which in many cases lacks measurement for a variety of reasons and causes.

Keywords: DFA, DCCA, Detrended cross-correlation coefficient, long-memory time-series, ARFIMA process

1. Introduction

Usually in the study of the real time-series, depending on the time scale involved, data loss proves to be an unfortunate reality for the researcher, for example: an automatic weather station is equipped with sensors to measure some variables as air temperature, wind speed/direction, relative humidity, atmospheric pressure, solar radiation, among others. In this station data is

Email address: gfzebende@uefs.br (G. F. Zebende)

automatically collected at every minute, but sometime for some unexpected reason the operation of data collection can stop, impairing future statistics analysis. Although, even in the absence of some data, time-series analyzes by statistical models are performed [1, 2]. Taking this fact into consideration, this paper study the influence of removed parts in time-series with long-range memory by the application of the Detrended Fluctuation Analysis, DFA [3], and the Detrended Cross-Correlation Coefficient, ρ_{DCCA} [4]. As a case of study, two time-series with long-memory process (autocorrelation functions decay slowly) were simulated. This stochastic process is said Auto Regressive Fractionally Integrated Moving Average, ARFIMA, and each variable depends not only on its own past, but also on the past values of the other variable [5, 6, 7, 8], in other words:

$$\begin{aligned} x_i &= W \sum_{m=1}^{\infty} a_m(\delta_1)x_{i-m} + (1 - W) \sum_{m=1}^{\infty} a_m(\delta_2)y_{i-m} + \varepsilon_{xi}, \\ y_i &= (1 - W) \sum_{m=1}^{\infty} a_m(\delta_1)x_{i-m} + W \sum_{m=1}^{\infty} a_m(\delta_2)y_{i-m} + \varepsilon_{yi}. \end{aligned} \quad (1)$$

ε_{xi} and ε_{yi} denote two independent and identically distributed (i.i.d.) Gaussian variables with zero mean and unit variance, $a_m(\delta)$ are statistical weights defined by $a_m(\delta) \equiv \Gamma(m - \delta)/(\Gamma(-\delta)\Gamma(1 + m))$, where Γ denotes the Gamma function, δ are parameters ranging from -0.5 to 0.5 (related to the DFA exponent (see equation 4), $\alpha_{DFA} = 0.5 + \delta$, and W is a free parameter ranging from 0.5 to 1.0 and controlling the strength of the power-law cross-correlations between x_i and y_i [5, 6]).

By this process, we generate two time-series x_i (ARFIMA 01) and y_i (ARFIMA 02) characterized by different values of $a_m(\delta_{1,2})$, W and length N with the following steps performed:

- (a) From the original time-series, it is calculated $F_{DFA}(n)$, $F_{DCCA}^2(n)$ and $\rho_{DCCA}(n)$ (see section 2);
- (b) A random position is chosen between 1 and $N - (10\% \text{ of } N)$ (new length) in the time-series. From this position, 10% of the time-series is removed (cut i);
- (c) $F_{DFA}(n)$, $F_{DCCA}^2(n)$ and $\rho_{DCCA}(n)$ are calculated for this new time-series;

- (d) Repeat the procedures (b) and (c) again;
- (e) The algorithm is implemented until the length of the last time-series reaches $N/2$ or 50% of the original time-series length.

The algorithm described above is presented in the Fig. 1, with $N = 10000$. Looking downwards, we can see the cuts until they reach 50% of the initial value (in this example, $N' = 5000$).

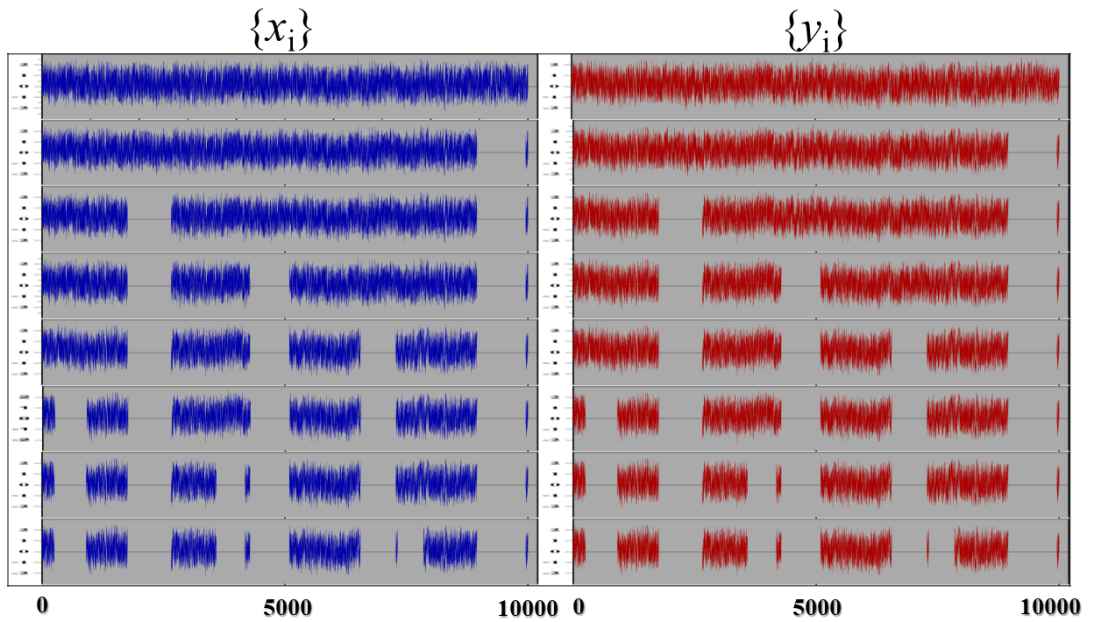


Figure 1: (Color on-line) The algorithm of cutting and removal pieces from the original time-series generated by an ARFIMA process with $N = 10000$ (top corner): $\{x_i\}$ is the [ARFIMA 01](#) with $\delta_1 = 0.1$ and $\{y_i\}$ is the [ARFIMA 02](#) with $\delta_2 = 0.4$. The strength of the power-law cross-correlations W is 0.5 and $\varepsilon_{x_i} \neq \varepsilon_{y_i}$.

and y_i

In the case of cut out time series it is noteworthy that the remaining pieces will be joined in the temporal sequence, in such a way that we can apply the statistical methods, DFA auto-correlation and DCCA cross-correlation, as we will see shortly below.

2. Methodology for Time-series Analysis

2.1. Detrended Fluctuation Analysis, DFA [3]

This method is based in the fluctuation function, F_{DFA} , defined from a time-series $\{x_i\}$ with $i = 1, 2, 3, \dots, N$ (length). Therefore, the integrated time-series is calculated by:

$$X_k = \sum_{i=1}^k (x_i - \langle x \rangle) \quad (2)$$

where, $\langle x \rangle$ represent the mean, $k = 1, 2, 3, \dots, N$. The cumulative sum, X_k , is divided into box windows of equal length n , and the local trend of each box, $X_{k,n}$, is obtained by a least-squares fit of each series. The fluctuation function is calculated by:

$$F_{DFA}(n) = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - X_{i,n})^2} \quad (3)$$

If there is power-law auto-correlations, then

$$F_{DFA}(n) \sim n^{\alpha_{DFA}} \quad (4)$$

Exponent α_{DFA} quantifies the empirical strength of the long-range (power-law) auto-correlations of the signal, a self-affinity parameter [9] (see Table 1 for α_{DFA} properties). The advantage of DFA over other methods lies in

Table 1: α_{DFA} exponent and properties [10].

Exponent	Type of signal
$\alpha_{DFA} < 0.5$	Anti-persistent
$\alpha_{DFA} \simeq 0.5$	Uncorrelated, white noise
$\alpha_{DFA} > 0.5$	Long-range correlated persistent
$\alpha_{DFA} \simeq 1.0$	1/f noise
$\alpha_{DFA} > 1.0$	Non-stationary
$\alpha_{DFA} \simeq 3/2$	Brownian motion

allowing the detection of long-range auto-correlations embedded in seemingly non-stationary time-series, and also avoiding spurious detection of apparent long-range auto-correlations, which are an artifact of non-stationarity. It is possible to generalize the DFA method to measure cross-correlation, and this generalization will be introduced below.

2.2. Detrended Cross-Correlation Analysis, DCCA [7]

Considering two time-series, $\{x_i\}$ and $\{y_i\}$, with $1 = 1, 2, \dots, N$, and the integrated time-series as equation 2, we obtain two new time-series, $\{X_k\}$ and $\{Y_k\}$. Thus, if we have overlapping boxes, these two integrated time-series, $\{X_k\}$ and $\{Y_k\}$, are divided into $(N - n)$ boxes of equal length n , with $4 \leq n \leq \frac{N}{4}$. The local trend of each box, $X_{k,n}$ and $Y_{k,n}$, are obtained by a least-squares fit of each series, and the covariance function of the residuals in each box are calculate the by:

$$f_{DCCA}^2(n, i) = \frac{1}{(n + 1)} \sum_{k=i}^{i+n} (X_k - X_{k,n})(Y_k - Y_{k,n}) \quad (5)$$

Finally, the average over all $(N - n)$ overlapping boxes is calculated to obtain the detrended covariance function:

$$F_{DCCA}^2(n) = \frac{1}{N - n} \sum_{i=1}^{N-n} f_{DCCA}^2(n, i) \quad (6)$$

If power-law cross-correlation is present between these time-series, then:

$$F_{DCCA}^2(n) \sim n^{2\lambda} \quad (7)$$

Exponent λ quantifies long-range (power-law) cross-correlations and also identifies seasonality [11, 12], but λ does not quantify the level of cross-correlations. In this sense, it is possible to quantify the level of cross-correlation with the DFA and the DCCA methods as will see now.

2.3. DCCA cross-correlation coefficient, ρ_{DCCA} [4]

This cross-correlation coefficient is defined as the ratio between the detrended covariance function F_{DCCA}^2 (equation 6) and the detrended variance function F_{DFA} (equation 3):

$$\rho_{DCCA}(n) = \frac{F_{DCCA}^2(n)}{F_{DFA_x}(n) F_{DFA_y}(n)} \quad (8)$$

It is observed that, $-1 \leq \rho_{DCCA} \leq 1$, and:

- i. $\rho_{DCCA} = 1$ means a perfect cross-correlation;
- ii. $\rho_{DCCA} = 0$ there is no cross-correlation;

iii. $\rho_{DCCA} = -1$ means a perfectly anti cross-correlation.

The advantage of this detrended cross-correlation coefficient lies in measuring cross-correlations between two non-stationary time-series at different time scales, see [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23], among others papers [24, 25, 26, 27, 28, 29, 30, 31, 32]. Also, Zebende and Machado proposed the generalization of ρ_{DCCA} , a Detrended Multiple Cross-Correlation coefficient DMC_x^2 , to be applied in order to quantify the relations between non-stationary signals for more than two independent variables [33].

3. Results and Discussions

3.1. Time-Series Simulation

Firstly, it was simulated two self-affine times-series, $\{y_{rand1}\}$ and $\{y_{rand2}\}$, with random numbers between 0 and 1 and $N = 1000000$. These time-series are independent and uncorrelated, with DFA exponents $\alpha_{DFA\{1\}} = \alpha_{DFA\{2\}} = 0.5$. After, we apply the ARFIMA process (Equation 1), and we introduce a cutoff length $l = 990000$ in our simulations, with the sum from 1 to l . Performed these steps, a pair of time-series with $N = 10000$ and long-range memory were created, see Table 2. These two time series are shown in

Table 2: Time series parameters for the ARFIMA process, [5, 6].

	ARFIMA 01	ARFIMA 02
δ	0.1	0.4
W	0.5	
	$\varepsilon_{xi} \neq \varepsilon_{yi}$	

the Fig. 1, top corner.

3.2. DFA results

Fig. 2 show $F_{DFA} \times n$ for all time series, original and with one (\square), two (\diamond), three (\circ), four (\triangleright), five (∇), six (\triangleleft), and seven (\bullet) cuts. By this figure, $F_{DFA}(n) \sim n^{\alpha_{DFA}}$, regardless of the analyzed time-series. For each curve, the values of α_{DFA} exponent (linear adjustment) was calculated, and their values are presented in the Table 3. The values of α_{DFA} exponents for all seven broken time-series has the same type of power-law, if compared to the original time-series. The values for α_{DFA} demonstrate this fact, Table 3. As

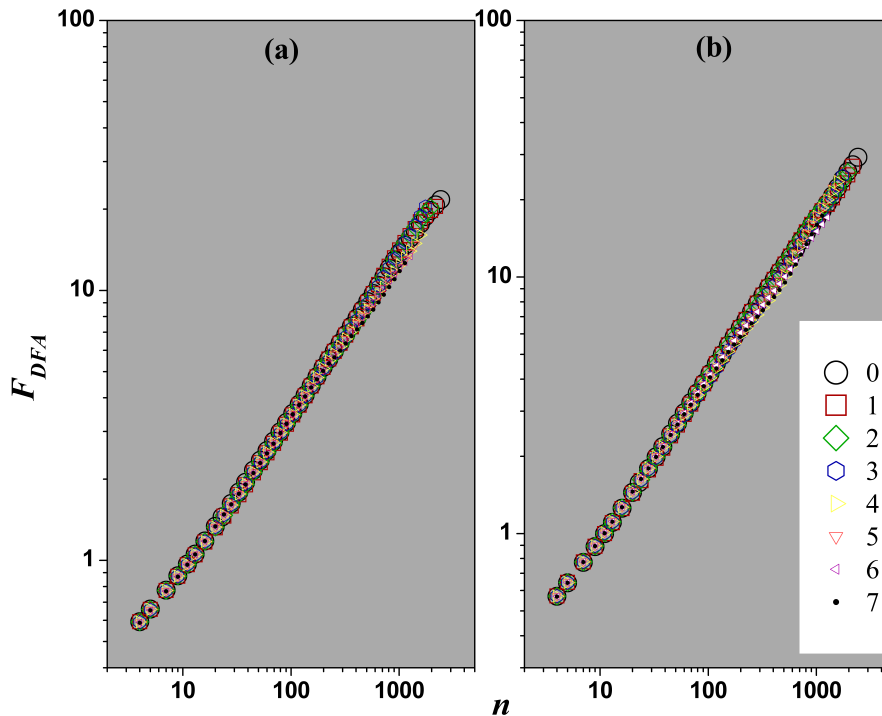


Figure 2: (Color on-line) F_{DFA} as a function of time scale n : (a) ARFIMA 01 and (b) ARFIMA 02. The numbers/symbols represents the DFA analysis for the original (\circ) time-series, with one (\square), two (\diamond), three (\odot), four (\triangleright), five (∇), six (\triangleleft), and seven (\bullet) cuts.

expected, DFA method which allows analysis in non-stationary time-series, showed robustness for statistics analysis with removed parts or missing data. To study the effect of cross-correlation between these time-series, we can apply the generalization of the DFA method, which will be introduced with sliding windows in the section below.

3.3. DCCA results

Fig. 3 presents the results for cross-correlation analysis using the DCCA method. We can observe that F_{DCCA}^2 does not follow a power-law for short time scales n , logically because of the simulated ARFIMA process, but the values for DCCA fluctuation function are equivalent regardless of the original or the cutoff time-series. To finish our goal, if we want to look at the cross-correlation level, we must calculate ρ_{DCCA} , that will be implemented in the next section.

Table 3: DFA exponent for: the original time-series (0), with one, two, three, four, five, six, and seven cuts. Linear adjustments for α_{DFA} with $R^2 > 0.99$.

	ARFIMA 01			ARFIMA 02		
0	α_{DFA}	0.581	± 0.002	α_{DFA}	0.623	± 0.001
1	α_{DFA}	0.586	± 0.003	α_{DFA}	0.622	± 0.002
2	α_{DFA}	0.586	± 0.003	α_{DFA}	0.625	± 0.001
3	α_{DFA}	0.588	± 0.004	α_{DFA}	0.620	± 0.002
4	α_{DFA}	0.563	± 0.002	α_{DFA}	0.615	± 0.004
5	α_{DFA}	0.561	± 0.002	α_{DFA}	0.618	± 0.003
6	α_{DFA}	0.559	± 0.002	α_{DFA}	0.596	± 0.003
7	α_{DFA}	0.551	± 0.002	α_{DFA}	0.595	± 0.003

3.4. ρ_{DCCA} results

Fig. 4 show the results for ρ_{DCCA} between the ARFIMA 01 and ARFIMA 02. We can see that there is cross-correlation between these time-series, but because $W = 0.5$ this cross-correlation is not perfect ($\rho_{DCCA} = 1$) for all time scale n . This cross-correlation increases its value as larger the time-scales are analyzed, due to the long-range memory generation process. Again, it can be seen that regardless of the time-series analyzed, in relation to the original time-series, all time-series had the same type of behavior, especially for time-scales $n < 1000$, that is, $n_{max} = 5000/4 \approx 1250$ points.

4. Conclusions

In this paper we analyze the effect on the final result due to the removal of pieces from the time-series itself in the point of view of auto-correlation and cross-correlation analysis. This objective was realized by the DFA and the DCCA methods. The motivation for this proposal comes from the fact that, for a variety of reasons, in many real cases, time-series may present data loss and that the DFA and DCCA methods have already been shown to be robust for non-stationary time-series analysis, if compared to the others classical methods.

Therefore, we test the DFA, the DCCA and the ρ_{DCCA} coefficient for time-series analysis with removed parts from a simulated original time-series with long-range memory, accomplished by an ARFIMA process. In this case

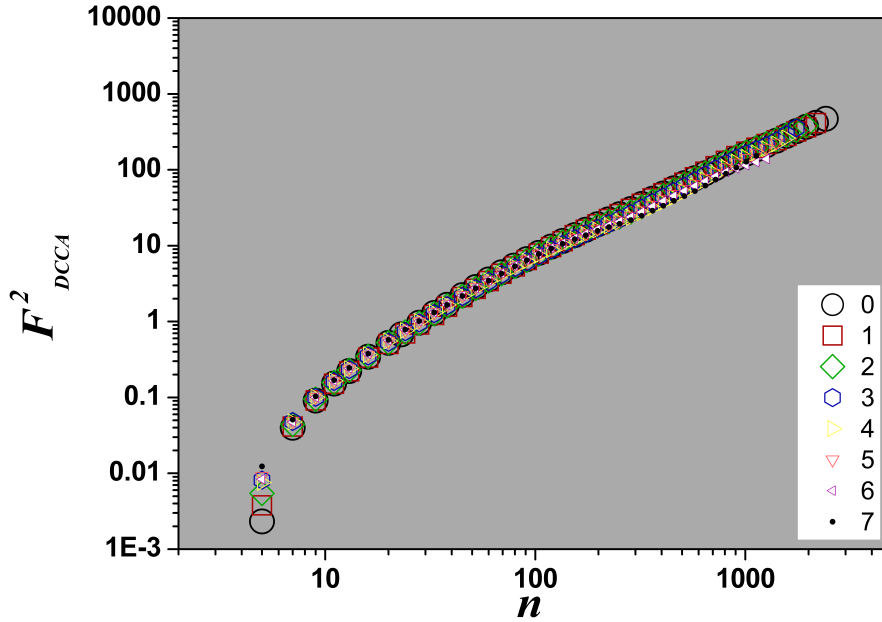


Figure 3: (Color on-line) F_{DCCA}^2 as a function of n for cross-correlation between [ARFIMA 01](#) and [ARFIMA 02](#) process. The numbers/symbols represents the DCCA analysis for the original (\circ) time-series, with one (\square), two (\diamond), three (\triangle), four (\triangleright), five (∇), six (\triangleleft), and seven (\bullet) cuts.

we could prove that the results for DFA, DCCA and ρ_{DCCA} coefficient not changed, if compared with the results for the original time-series. The results for the auto-correlation analysis, by F_{DFA} , and the cross-correlation analysis, by F_{DCCA} , show equivalence between the original time-series and the all time series with removed parts. Also, for ρ_{DCCA} the results are unaffected and remain equivalent regardless of the cutoff time-series, this up to the limit of 50 % of random removed parts. Finally, such methodology proved to be robust for time-series analysis with removed parts and we can say that such time-series analysis will be robust even if applied to the real non-stationary time-series with data loss in its generation.

5. Acknowledgments

Brito thanks the FAPESB (Fundação de Amparo à Pesquisa do Estado da Bahia) (Grant BOL 0262/2017) and Zebende thanks the CNPq (Conselho

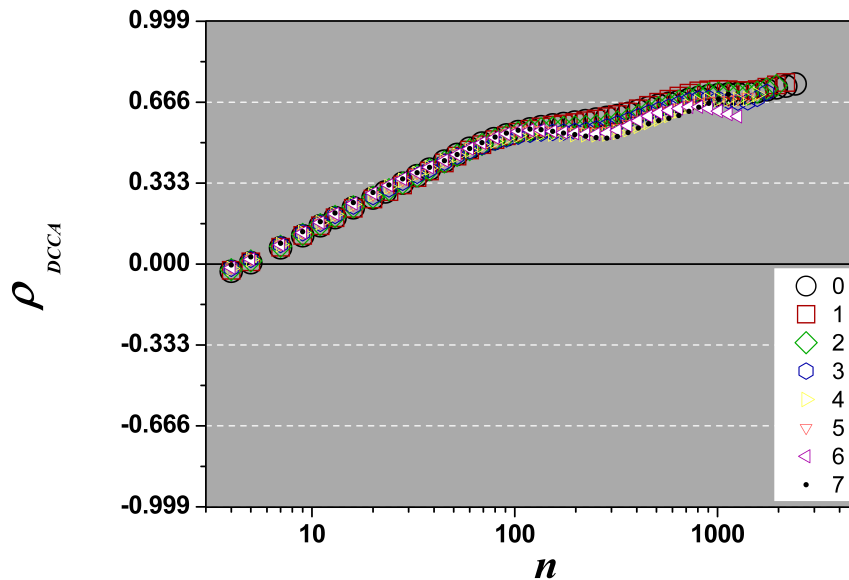


Figure 4: (Color on-line) ρ_{DCCA} as a function of time scale n for DCCA cross-correlation between [ARFIMA 01](#) and [ARFIMA 02](#). The numbers/symbols represents the ρ_{DCCA} analysis for the original (\circ) time-series, with one (\square), two (\diamond), three (\hexagon), four (\triangleright), five (∇), six (\triangleleft), and seven (\bullet) cuts.

Nacional de Desenvolvimento Científico e Tecnológico) (Grant 304362/2017-4).

6. References

- [1] J. D. Hamilton, Time Series Analysis, Vol. 2, Princeton university press Princeton, NJ, 1994.
- [2] W. W. S. Wei, Time Series Analysis. In The Oxford Handbook of Quantitative Methods in Psychology, Vol. 2, Oxford University Press, 2013.
- [3] 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.
- [4] G. F. Zebende, DCCA cross-correlation coefficient: Quantifying level of cross-correlation, Physica A 390 (4) (2011) 614–618.

- [5] J. R. M. Hosking, Fractional differencing, *Biometrika* 68 (1) (1981) 165–176.
- [6] B. Podobnik, I. Grosse, D. Horvatić, S. Ilic, P. C. Ivanov, H. E. Stanley, Quantifying cross-correlations using local and global detrending approaches, *The European Physical Journal B* 71 (2) (2009) 243.
- [7] B. Podobnik, H. E. Stanley, Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series, *Phys. Rev. Lett* 100 (2008) 1–4.
- [8] E. S. Trevisan, R. C. Souza, L. R. Souza, Estimação do parâmetro “d” em modelos arfima, *Pesquisa Operacional* 20 (2000) 73 – 82.
- [9] G. F. Zebende, B. F. Fernandez, M. G. Pereira, Analysis of the variability in the sdb star KIC 10670103: DFA approach, *Monthly Notices of the Royal Astronomical Society* 464 (2017) 2638 – 2642.
- [10] J. Walleczek, *Self-organized biological dynamics and nonlinear control: toward understanding complexity, chaos and emergent function in living systems*, Cambridge University Press, 2000.
- [11] G. Zebende, M. da Silva, A. M. da Silva-Filho, DCCA cross-correlation coefficient differentiation: Theoretical and practical approaches, *Physica A* 392 (2013) 1756–1761.
- [12] G. Zebende, P. A. da Silva, A. M. Filho, Study of cross-correlation in a self-affine time series of taxi accidents, *Physica A* 390 (2011) 1677–1683.
- [13] L. Kristoufek, Measuring cross-correlation between non-stationary series with DCCA coefficient, *Physica A* 402 (2014) 291–298.
- [14] L. Kristoufek, Detrending moving-average cross-correlation coefficient: Measuring cross-correlations between non-stationary series, *Physica A* 406 (2014) 169–175.
- [15] L. Kristoufek, Testing power-law cross-correlations: Rescaled covariance test, *The European Physical Journal B* 86 (2013) 418.
- [16] G.-J. Wang, C. Xie, S. Chen, J.-J. Yang, Ming-YanYang, 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* 392 (17) (2013) 3715–3730.
- [17] L. Piao, Z. Fu, Quantifying distinct associations on different temporal scales: comparison of DCCA and Pearson methods, *Scientific Reports* 6 (36759) (2016) 1 – 11.
- [18] J. C. Reboredo, M. A. Rivera-Castro, G. F. Zebende, Oil and US dollar exchange rate dependence: A detrended cross-correlation approach, *Energy Economics* 42 (c) (2014) 132–139.
- [19] E. Marinho, A. Sousa, R. Andrade, Using detrended cross-correlation analysis in geophysical data, *Physica A* 392 (2013) 2195–2201.
- [20] U. Bashir, Y. Yu, M. Hussain, G. F. Zebende, Do foreign exchange and equity markets co-move in Latin American region? Detrended cross-correlation approach, *Physica A* 462 (2016) 889 – 897.
- [21] M. Hussain, G. F. Zebende, U. Bashir, D. Donghong, Oil price and exchange rate co-movements in Asian countries: Detrended cross-correlation approach, *Physica A* 465 (1) (2017) 338 – 346.
- [22] B. Podobnik, Z.-Q. Jiang, W.-X. Zhou, H. E. Stanley, Statistical tests for power-law cross-correlated processes, *Phys. Rev. E* 84 (2011) 066118.
- [23] X. Jin, Time-varying return-volatility relation in international stock markets, *International Review of Economics & Finance* 51 (2017) 157 – 173.
- [24] E. Guedes, A. Brito, F. Oliveira Filho, B. Fernandez, A. de Castro, A. da Silva Filho, G. Zebende, Statistical test for $\delta\rho_{DCCA}$ cross-correlation coefficient, *Physica A: Statistical Mechanics and its Applications* 501 (2018) 134–140.
- [25] G. F. Zebende, F. M. Oliveira Filho, J. A. L. Cruz, Auto-correlation in the motor/imaginary human eeg signals: A vision about the F DFA fluctuations, *PloS one* 12 (9) (2017) e0183121.
- [26] G. Zebende, A. Brito, A. Silva Filho, A. Castro, ρ_{dcca} applied between air temperature and relative humidity: An hour/hour view, *Physica A: Statistical Mechanics and its Applications* 494 (2018) 17–26.

- [27] U. Bashir, G. F. Zebende, Y. Yu, M. Hussain, A. Ali, G. Abbas, Differential market reactions to pre and post brexit referendum, *Physica A: Statistical Mechanics and its Applications* 515 (2019) 151–158.
- [28] E. Guedes, G. Zebende, Dcca cross-correlation coefficient with sliding windows approach, *Physica A: Statistical Mechanics and its Applications* 527 (2019) 121286.
- [29] L. A. da Silva, E. Guedes, P. Ferreira, A. Dionísio, G. Zebende, $\rho_{x, y}$ between open-close stock markets, *Physica A: Statistical Mechanics and its Applications* 534 (2019) 122152.
- [30] J. C. Reboredo, M. A. Rivera-Castro, G. F. Zebende, Oil and us dollar exchange rate dependence: A detrended cross-correlation approach, *Energy Economics* 42 (2014) 132–139.
- [31] R. T. Vassoler, G. F. Zebende, DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity, *Physica A* 391 (7) (2012) 2438–2443.
- [32] G. Zebende, M. Da Silva, A. Machado Filho, DCCA cross-correlation coefficient differentiation: Theoretical and practical approaches, *Physica A: Statistical Mechanics and its Applications* 392 (8) (2013) 1756–1761.
- [33] G. Zebende, A. da Silva Filho, Detrended multiple cross-correlation coefficient, *Physica A: Statistical Mechanics and its Applications* 510 (2018) 91–97.

CONSIDERAÇÕES FINAIS

4.1 Conclusões

Neste projeto de Tese, realizamos uma pesquisa de natureza interdisciplinar, direcionada para a área de Sistemas Complexos, com o objetivo de fazer uma modelagem computacional de auto-correlação e correlações cruzadas, através dos métodos DFA e DCCA e dos coeficientes de correlações cruzadas ρ_{DCCA} e DMC_x^2 , no intuito de mensurar as correlações cruzadas entre séries temporais de variáveis meteorológicas e dados simulados.

Nossos dados empíricos foram disponibilizados pelo INMET, para quatro estações automáticas localizadas no Estado na Bahia, nas cidades de Barreiras, Cruz das Almas, Paulo Afonso e Salvador e os dados simulados foram obtidos pelo processo ARFIMA, para séries temporais de memória longa. Os principais resultados encontrados são: artigos publicados em revistas científicas qualificadas, sendo 03 (três) deles relacionados ao projeto inicial do Doutorado e registros de programa de computador no INPI, que permitiu a modelagem computacional dos dados.

Especificamente, no **artigo 3.1**, apresentamos uma nova maneira de visualizar as variáveis meteorológicas, temperatura do ar e umidade relativa do ar, produzindo assim um novo mapa de cores das correlações cruzadas. Este mapa nos fornece uma visão completa das correlações cruzadas entre as variáveis para todas as escalas e a hora do dia. Com este mapa é possível identificar padrões para as diferentes estações.

No **artigo 3.2**, identificamos via o coeficiente de múltiplas correlação cruzada, DMC_x^2 , potenciais locais com maior eficiência para captação de energia solar, isto através da análise das variáveis radiação solar, temperatura do ar e umidade relativa do ar. Resumidamente é sabido que, a eficiência de uma célula fotovoltaica depende principalmente de fatores internos, mas fatores externos são muito importantes, bem como suas inter-relações. Considerando a radiação solar global como a variável dependente, e a temperatura do ar e a umidade relativa do ar como variáveis independentes, temos o seguinte: assumindo $DMC_x^2 = 1$ como valor ideal para a máxima captação de energia solar, pudemos testar esta nova hipótese neste artigo.

Finalmente no **artigo 3.3**, analisamos o efeito no resultado final da remoção de partes de séries temporais do ponto de vista da análise de auto-correlação e correlação cruzada utilizando os métodos DFA, DCCA e o coeficiente ρ_{DCCA} , isto para séries temporais

simuladas. Identificamos que os resultados do DFA, do DCCA e do coeficiente ρ_{DCCA} não sofreram alterações (qualitativas e quantitativas) se comparados com os resultados da série temporal original, até o limite de 50% de corte e remoção. Por fim, concluímos que essa metodologia mostra-se robusta para análise de séries temporais com partes removidas, e assim podemos dizer que tais análises de séries temporais são robustas, mesmo se aplicadas a séries temporais com perda de dados em sua geração.

4.2 *Perspectivas Futuras*

Como perspectivas futuras, deixamos como motivação um trabalho submetido para publicação na **Revista Brasileira de Climatologia** e um outro em fase de conclusão da escrita, parte em conjunto com um projeto de Mestrado na área de Ciências Ambientais no PPGM respectivamente descritos logo abaixo:

Uma visão hora a hora da auto-correlação em dados de temperatura e umidade relativa do ar na Bahia

Propomos neste trabalho analisar empiricamente as auto-correlações em dados hora/hora de temperatura e umidade relativa do ar para as quatro estações meteorológicas automáticas localizadas nas cidades de Barreiras, Cruz das Almas, Paulo Afonso e Salvador no Estado da Bahia. Esse estudo foi realizado pelo método DFA e o seu expoente de auto-correlação, α_{DFA} .

Tendências, auto-correlação e correlação cruzada: um estudo sobre a temperatura do ar na BA

Neste projeto proporemos, em parceria com o programa PPGM, analisar o efeito climático antropogênico, a ser implementado ainda este ano.

Referências Bibliográficas

- ABBAS, G. et al. The return and volatility nexus among stock market and macroeconomic fundamentals for China. *Physica A: Statistical Mechanics and its Applications*, v. 526, p. 121025, 2019.
- ANGELOCCI, L. R.; PEREIRA, A. R.; SENTELHAS, P. C. *Agrometeorologia - Fundamentos e Aplicações Práticas*. one. [S.l.: s.n.], 2002.
- Aragão, J. O. R. *Um estudo das estruturas das perturbações sinóticas do Nordeste do Brasil*. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais: INPE, 1975. Disponível em: <<http://mtc-m16b.sid.inpe.br/col/sid.inpe.br/jeferson/2005/05.20.18.35/doc/INPE-789-TPT-017.pdf>>.
- Araújo, H. A. de. *Padrões de Ecos de Radar Meteorológico no Semi-árido Brasileiro e Condições Atmosféricas*. Dissertação (Mestrado) — UFPB, 1996.
- BARRY, R. G.; CHORLEY, R. J. *Atmosfera, Tempo e Clima*. [S.l.]: Porto Alegre - RS, 2013.
- BASHIR, U. et al. Do foreign exchange and equity markets co-move in Latin American region? detrended cross-correlation approach. *Physica A: Statistical Mechanics and its Applications*, v. 462, p. 889–897, 2016.
- BASHIR, U. et al. Differential market reactions to pre and post Brexit referendum. *Physica A: Statistical Mechanics and its Applications*, v. 515, p. 151–158, 2019.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. *Times Series Analysis: Forecasting and Control*. [S.l.]: Englewood Cliffs: Prentice Hall, 2008.
- BRITO, A. A. *Temperatura ambiente e Umidade relativa do ar na Bahia: uma visão hora a hora via Coeficiente de Correlação Cruzada ρ_{DCCA}* . Dissertação (Mestrado) — SENAI CIMATEC, 2016.
- BRITO, A. A.; Araújo, H. A.; ZEBENDE, G. F. Detrended multiple cross-correlation coefficient applied to solar radiation, air temperature and relative humidity. *Scientific Reports*, v. 9, p. 1–10, 2019.
- BRITO, A. A.; MACHADO, A.; ZEBENDE, G. F. Análise temporal de dados climatológicos de Feira de Santana (BA). *Conjuntura & Planejamento*, v. 189, p. 50–59, 2015.
- BRITO, A. A. et al. Cross-correlation in a turbulent flow: Analysis of the velocity field using the DCCA coefficient. *Europhysics Letters*, v. 123, p. 20011, 2018.
- CARVALHO, L. M. V.; JONES, C.; LIEBMANN, B. The south atlantic convergence zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall. *Journal of Climate*, v. 17, p. 88–108, 2004.
- CASTRO, A. P. N. de. *Método de detecção de padrões na dinâmica dos movimentos complexos na deglutição de um indivíduo*. Dissertação (Dissertação de Mestrado) — SENAI CIMATEC, 2012.

- DIAS, P. L. S.; SCHUBERT, W. H.; DEMARIA, M. Large scale response of the tropical atmosphere to transient convection. *Journal of Atmospheric Sciences*, v. 40, p. 2689–2707, 1983.
- FEDER, J. *Fractals*. [S.l.]: Boston, MA, 1988.
- FERNANDEZ, B. F. *Análise da Auto-afinidade do Comportamento de Variabilidade Temporal da Estrela Pulsante KIC10670103*. Dissertação (Dissertação de Mestrado) — SENAI CIMATEC, 2014.
- 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.
- FERREIRA, P.; KRISTOUFEK, L.; PEREIRA, E. J. A. L. DCCA and DMCA correlations of cryptocurrency markets. *Physica A: Statistical Mechanics and its Applications*, v. 545, p. 123803, 2020.
- FERREIRA, P.; PEREIRA, E. J.; SILVA, M. F. The relationship between oil prices and the Brazilian stock market. *Physica A: Statistical Mechanics and its Applications*, v. 545, p. 123745, 2020.
- FERREIRA, P. et al. Detrended correlation coefficients between oil and stock markets the effect of the 2008 crisis. *Physica A: Statistical Mechanics and its Applications*, v. 517, p. 86–96, 2019.
- FIGUEROA, S. N.; SATYAMURTY, P.; DIAS, P. L. S. Simulations of the summer circulation over south american region with na eta coordinate model. *Journal Atmosphere Physics*, v. 52, p. 1573–1584, 1995.
- FILHO, A. M. da S. *Correlação de longo alcance em indicadores de criminalidade de Salvador-BA: DFA e DCCA*. Dissertação (Dissertação de Mestrado) — SENAI-CIMATEC, 2009.
- FILHO, A. M. da S. *Autocorrelação e Correlação Cruzada: Teoria e Aplicações*. Tese (Doutorado) — SENAI CIMATEC, 2014.
- FILHO, F. M. O.; CRUZ, J. A. L.; ZEBENDE, G. F. Analysis of the EEG bio-signals during the reading task by DFA method. *Physica A: Statistical Mechanics and its Applications*, v. 525, p. 664–671, 2019.
- FILHO, M. F. G. *Um estudo sobre a influência do albedo diferencial e da orografia na circulação atmosférica: uma aplicação para o Nordeste brasileiro*. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais: INPE, 1979.
- GUEDES, E. F. *Um Modelo Computacional para o Monitoramento em Tempo Real da Eficiência e da Interdependência entre Mercado Financeiros*. Tese (Doutorado) — SENAI CIMATEC, 2018.
- GUEDES, E. F. et al. Statistical test for $\Delta\rho$ cross-correlation coefficient. *Physica A: Statistical Mechanics and its Applications*, v. 501, p. 134–140, 2018.
- GUEDES, E. F. et al. Statistical test for $\Delta\rho_{DCCA}$: Methods and data. *Data in Brief*, v. 18, p. 795 – 798, 2018. ISSN 2352-3409.

- GUEDES, E. F. 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*, v. 479, p. 38 – 47, 2017.
- GUEDES, E. F. et al. An econophysics approach to study the effect of BREXIT referendum on european union stock markets. *Physica A: Statistical Mechanics and its Applications*, v. 523, p. 1175–1182, 2019.
- GUEDES, E. F.; ZEBENDE, G. F. DCCA cross-correlation coefficient with sliding windows approach. *Physica A: Statistical Mechanics and its Applications*, v. 527, p. 121286, 2019.
- GUEDES, E. F.; ZEBENDE, G. F.; LIMA, I. C. da C. Quantificação dos efeitos do câmbio na produção da indústria de transformação baiana: uma abordagem via coeficiente de correlação cruzada RHO-DCCA. *CONJUNTURA & PLANEJAMENTO*, v. 01, p. 75–89, 2017.
- HARRIS, D. C. Charles David Keeling and the Story of Atmospheric CO₂ Measurements. *Analytical Chemistry*, v. 82, n. 19, p. 7865–7870, 2010. PMID: 20536268.
- HURST, E. Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engenieers.*, v. 116, p. 770–799, 1951.
- HUSSAIN, M. et al. Oil price and exchange rate co-movements in asian countries: Detrended cross-correlation approach. *Physica A: Statistical Mechanics and its Applications*, v. 465, p. 338–346, 2017.
- JUNIOR, J. S. S. et al. Uma análise estatística nos indicadores de criminalidade de salvador. *Conjuntura & Planejamento*, v. 161, p. 40–49, 2008.
- KANTELHARDT, J. W. *Fractal and Multifractal Time Series*. [S.l.]: Germany, 2008.
- KODAMA, Y. M. Large-scale common features of subtropical precipitation zones (the baiu frontal zone, the spcz, and the sacz). part i: Characteristics of subtropical frontal zones. *Meteorological Society Japan*, v. 70, p. 813–835, 1992.
- KOUSKY, V. E. Frontal influences on northeast brazil. *Monthly Weather Review*, v. 107, p. 1140, 1979.
- KOUSKY, V. E. Diurnal rainfall variation in northeast brazil. *Monthly Weather Review*, v. 108, p. 488, 1980.
- KOUSKY, V. E. Pentad outgoing longwave radiation climatology for the south american sector. *Revista Brasileira de Meteorologia*, v. 3, p. 217, 1988.
- KOUSKY, V. E.; CHUG, P. S. Fluctuations in annual rainfall for northeast brazil1. *Journal of the Meteorological Society of Japan. Ser. II*, v. 56, n. 5, p. 457–465, 1978.
- KOUSKY, V. E.; GAN, M. A. Upper tropospheric cyclonic vortices in the tropical south atlantic. *Tellus*, v. 33, n. 6, p. 538–551, 1981.
- KRISTOUFEK, L. Measuring correlated between non-stationary series with DCCA coefficient. *Physica A: Statistical Mechanics and its Applications*, v. 402, p. 291–298, 2014.
- LI, F.; ZHU, H. Testing for financial contagion based on a nonparametric measure of the cross-market correlation. *Review of Financial Economics*, v. 23, p. 141, 2014.

- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. [S.l.]: São Paulo: ABE - Projeto FISHER, 2006.
- NASCIMENTO, A. et al. Cross-correlation analysis on brazilian gasoline retail market. *Physica A: Statistical Mechanics and its Applications*, v. 508, p. 550–557, 2018.
- NOBRE, C. A. Ainda sobre a zona de convergência do atlântico sul: A importância do oceano atlântico. *Climanálise*, v. 3, p. 30, 1988.
- OLIVEIRA, S. G. S.; GUEDES, E. F.; ZEBENDE, G. F. Vulnerabilidade social e criminalidade na região metropolitana de salvador: Uma abordagem com métodos quantitativos. *Conjuntura & Planejamento*, v. 01, p. 79–93, 2020.
- PARRY, M. L.; CARTER, T. R.; KONIJN, N. T. The climatology of droughts and drought prediction. In: _____. *The Impact of Climatic Variations on Agriculture*. Springer, 1988. cap. 2. Disponível em: <https://link.springer.com/chapter/10.1007/978-94-009-2965-4_12>.
- PEARSON, K. Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London Series I*, v. 58, p. 240–242, jan. 1895.
- PENG, C. K. et al. Mosaic organization of dna nucleotides. *Physical Review E*, v. 49, p. 1685–1689, 1994.
- PEREIRA, A. R.; ANGELOCCI, L. R.; SENTELHAS, P. C. *Meteorologia Agrícola 306*. [S.l.], 2007. Disponível em: <http://www.leb.esalq.usp.br/leb/aulas/lce306/MeteorAgricola_Apostila2007.pdf>.
- PEREIRA, B. B. *Tópicos em Séries Temporais: Métodos Automáticos de Previsão*. [S.l.]: Rio de Janeiro, 1980.
- PEREIRA, E. J.; FERREIRA, P. The impact of the Brexit referendum on british and european union bank shares: a cross-correlation analysis with national indices. *Economics Bulletin*, v. 39, p. 335–346, 2019.
- PEREIRA, E. J. et al. Multiscale network for 20 stock markets using DCCA. *Physica A: Statistical Mechanics and its Applications*, v. 529, p. 121542, 2019.
- PEREIRA, E. J. A. L.; SILVA, M. F.; PEREIRA, H. B. B. Econophysics: Past and present. *Physica A: Statistical Mechanics and its Applications*, v. 473, p. 251–261, 2017.
- PEREIRA, E. J. L. et al. Trumps effect on stock markets: A multiscale approach. *Physica A: Statistical Mechanics and its Applications*, v. 512, p. 241–247, 2018.
- PODOBNIK, B. et al. Statistical tests for power-law cross-correlated processes. *Physical Review E*, v. 84, p. 066118, 2011.
- PODOBNIK, B.; STANLEY, H. E. Detrended cross-correlation analysis: A new method for analyzing two nonstationary times series. *PHYSICAL REVIEW LETTERS*, v. 100, p. 084102, 2008.
- QUADRO, M. F. D. L. *Estudo de Episódios de Zona de Convergência do Atlântico Sul (ZCAS) sobre a América do Sul. INPE-6341-TDL/593*. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais: INPE, 1994.

- RAMIREZ, M. C. V.; KAYANO, M. T.; FERREIRA, N. J. Statistical analysis of upper tropospheric vortices in the vicinity of northeast brazil during the period 1980-1989. *Atmosfera*, v. 12, p. 75–88, 1999.
- RAO, V. B.; HADA, K. Characteristics of rainfall over brazil: Annual variations and connections with the southern oscillation. *Theoretical and Applied Climatology*, v. 42, p. 81, 1990.
- RAO, V. B.; LIMA, M. C. de; FRANCHITO, S. H. Seasonal and variations of rainfall over eastern northeast brazil. *Journal of Climatology*, v. 6, p. 1754–1763, 1993.
- REBOITA, M. S. et al. Caracterização atmosférica quando da ocorrência de eventos extremos de chuva na região sul de minas gerais. *Revista Brasileira de Climatologia*, v. 21, 2017.
- 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.
- SCHWERDTFEGGER, W. High thunderstorm frequency over the subtropical andes during the summer, causes and effects. climate of central and south american. *World Survey of Climatology*, v. 12, p. 192–195, 1976.
- SELUCHI, M. E.; CHOU, S. C. Synoptic patterns associated with landslide events in the serra do mar, brazil. *Theoretical and Applied Climatology*, v. 98, p. 67–77, 2009.
- SILVA, L. S. A. Almeida da et al. $\rho(x, y)$ between open-close stock markets. *Physica A: Statistical Mechanics and its Applications*, v. 534, p. 122152, 2019.
- SILVA, M. A. V. *Meteorologia e Climatologia*. [S.l.]: Versão Digital 2, 2006.
- SILVA, M. F. da et al. Quantifying cross-correlation between ibovespa and brazilian blue-chips: The DCCA approach. *Physica A: Statistical Mechanics and its Applications*, v. 424, p. 124–129, 2015.
- SILVA, M. F. da et al. Quantifying the contagion effect of the 2008 financial crisis between the G7 countries (by GDP nominal). *Physica A: Statistical Mechanics and its Applications*, v. 453, p. 1–8, 2016.
- SPERMAN, C. The proof and measurement of association between two things. *The American Journal of Psychology*, v. 15, p. 72–101, 1904.
- STANTON, J. M. Galton, Pearson, and the Peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, v. 9, p. 3, 2001.
- STIGLER, S. M. Francis Galton's account of the invention of correlation. *Statistical Science*, v. 4, p. 73–79, 1989.
- TANAJURA, C. A. S. *Modeling and Analysis os the South American Summer Climate*. Dissertação (Mestrado) — Universidade de Maryland, 1996.
- UVO, C. R. B.; NOBRE, C. A. A zona de convergência intertropical (ZCIT) e a precipitação no norte do nordeste do brasil. parte i: A posição da zcitr no atlântico equatorial. *Climanálise*, v. 3, n. 7, p. 34–40, 1989.
- VASSOLER, R. T. *Estudo de Auto-Afinidade e Correlações Cruzadas em Dados Climatológicos*. Dissertação (Dissertação de Mestrado) — SENAI CIMATEC, 2012.

VASSOLER, R. T.; ZEBENDE, G. F. DCCA cross-correlation coefficient applied in time series of air temperature and air relative humidity. *Physica A: Statistical Mechanics and its Applications*, v. 391, p. 2438–2443, 2012.

VIEIRA, L. *Meteorologia e Climatologia Agrícola - Notas de aula Agrometeorologia*. [S.l.], 2009.

YUAN, N. et al. Detrended partial-cross-correlation analysis: A new method for analyzing correlations in complex system. *Scientific Reports*, v. 05, p. 8143, 2015.

ZEBENDE, G. F. *Correlação de informação em disquetes de computador*. Tese (Tese de Doutorado) — Universidade Federal Fluminense, 1999.

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

ZEBENDE, G. F.; BRITO, A. A.; CASTRO, A. P. N. de. DCCA cross-correlation analysis in time-series with removed parts. *Physica A: Statistical Mechanics and its Applications*, v. 545, p. 123472, 2020.

ZEBENDE, G. F. et al. DCCA cross-correlation coefficient applied air temperature and relative humidity: An hour/hour view. *Physica A: Statistical Mechanics and its Applications*, v. 494, p. 17–26, 2018.

ZEBENDE, G. F.; FERNANDEZ, B. F.; PEREIRA, M. G. Analysis of the variability in the sdB star KIC10670103: DFA approach. *Monthly Notices of the Royal Astronomical Society*, v. 464, p. 2638–2642, 2017.

ZEBENDE, G. F.; FILHO, A. M. D. S. Detrended multiple cross-correlation coefficient. *Physica A: Statistical Mechanics and its Applications*, v. 510, p. 91–97, 2018.

ZEBENDE, G. F.; FILHO, A. M. da S. Cross-correlation between time series of vehicles and passengers. *Physica A: Statistical Mechanics and its Applications*, v. 88, p. 4863–4866, 2009.

ZEBENDE, G. F.; FILHO, A. M. da S.; SILVA, M. F. da. 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.

ZEBENDE, G. F.; FILHO, F. M. de O.; CRUZ, J. A. L. Auto-correlation in the motor/imaginary human EEG signals: A vision about the FDF fluctuation. *PLoS One*, v. 12, p. e0183121, 2017.

ZEBENDE, G. F.; OLIVIERA, P. M. C. de; PENNA, T. J. P. Long-range correlation in computer diskettes. *Physical Review E*, v. 57, p. 3311–3314, 1998.

ZEBENDE, G. F.; SILVA, P. A.; FILHO, A. M. da S. Study of cross-correlation in a self-affine time series of taxi accidents. *Physica A: Statistical Mechanics and its Applications*, v. 390, p. 1677–1683, 2011.

ZEBENDE, G. F.; SILVA, P. A.; MACHADO, A. Study of cross-correlation in a self-affine time series of taxi accidents. *Physica A: Statistical Mechanics and its Applications*, v. 390, p. 1677–1683, 2011.

APÊNDICE I

A.1 Registro de Software

Registramos no INPI - Instituto Nacional de Propriedade Industrial, os seguintes programas de computador:

- **RHO-DCCA-MATRIX**, sob número *BR*51 2016 000375 – 8, que calcula em diferentes escalas de tamanhos os expoentes de autocorrelação, α_{DFA} , o expoente de correlação cruzada, λ_{DCCA} e o coeficiente de correlação cruzada, ρ_{DCCA} . Com os resultados deste software, desenvolvemos a dissertação defendida em 2016, “Temperatura ambiente e Umidade relativa do ar na Bahia: uma visão hora a hora via Coeficiente de Correlação Cruzada ρ_{DCCA} ” ver [Brito \(2016\)](#) e também publicamos o paper intitulado “ ρ_{DCCA} applied between air temperature and relative humidity: An hour/hour view”, publicado na revista *Physica A: Statistical Mechanics and its Applications*, v. 494, p. 17-26, 2018, ver [Zebende e Filho \(2018\)](#).
- **RHO-FLUIDO**, sob número *BR*51 2017 000744 – 6, que calcula em diferentes escalas as correlações cruzadas entre pontos distintos, internos e externos, de dutos. Este software foi desenvolvido para a publicação dos artigos intitulados: “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”, publicado na revista *Physica A: Statistical Mechanics and its Applications*, v. 509, p. 294-301, 2018 e “Cross-correlation in a turbulent flow: Analysis of the velocity field using the ρ_{DCCA} coefficient”, publicado na revista *Europhysics Letters*, v. 123, p. 20011, 2018, ver artigos: [Brito et al. \(2018\)](#) e [Santos et al. \(2018\)](#).
- **RHO-CUT**, sob número *BR*51 2020 001149 – 7, que calcula em diferentes escalas as correlações cruzadas dos cortes e remoções da série temporal de modo aleatório no limite de até 50% da série analisada. Com este software, publicamos o paper intitulado “DCCA cross-correlation analysis in time-series with removed parts” na revista *Physica A: Statistical Mechanics and its Applications*, v. 545, p. 123472, 2020 ver, [Zebende, Brito e Castro \(2020\)](#).
- **RHO BA MAP**, sob número *BR*51 2020 001150 – 0, que calcula as correlações cruzadas da variável meteorológica temperatura do ar entre diversas estações meteorológicas no estado da Bahia (Brasil). Artigo submetido.



INPI INSTITUTO
NACIONAL
DA PROPRIEDADE
INDUSTRIAL

REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA E COMÉRCIO EXTERIOR
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE CONTRATOS, INDICAÇÕES GEOGRÁFICAS E REGISTROS

CERTIFICADO DE REGISTRO DE PROGRAMA DE COMPUTADOR

Processo: BR 51 2016 000375-8

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 criação indicada, em conformidade com o art. 3º da Lei Nº 9.609, de 19 de Fevereiro de 1998, e arts. 1º e 2º do Decreto 2.556 de 20 de Abril de 1998.

Título: **RHO_DCCA_MATRIX**
Criação: 15 de junho de 2015
Titular(es): SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL-SENAI/DR-BAHIA (03.795.071/0001-16)
Autor(es): ANDREA DE ALMEIDA BRITO (876.091.425-49)
ARLEYS PEREIRA NUNES DE CASTRO (595.821.465-91)
GILNEY FIQUEIRA ZEBENDE (022.396.277-51)
Linguagem: C, C++
Aplicação: MT-01, MT-06
Tipo Prog.: AP-01, IT-04, SM-01

DOCUMENTAÇÃO TÉCNICA EM DEPÓSITO SOB SIGILO ATÉ 30/03/2026.

Os Direitos Patrimoniais relativos ao programa de computador objeto do presente registro foram cedidos dos Criadores para o Titular, na data de 29 de março de 2016, conforme documentação

A exclusividade de comercialização deste programa de computador não tem a abrangência relativa à exclusividade de fornecimento estatuída pelo art.25, I, da Lei nº8.666, de 21 de Junho de 1993, para fins de inexigibilidade de licitação para compras pelo poder público.

Expedido em 02 de agosto de 2016

Assinado digitalmente por:

Breno Bello de Almeida Neves

Diretor de Contratos, Indicações Geográficas e Registros



INPI INSTITUTO
NACIONAL
DA PROPRIEDADE
INDUSTRIAL

REPÚBLICA FEDERATIVA DO BRASIL
Ministério Da Indústria, Comércio Exterior e Serviços
Instituto Nacional da Propriedade Industrial

Diretoria de Patentes, Programas de Computador e Topografias de Circuitos Integrados

Certificado de Registro de Programas de Computador

Processo nº: BR 51 2017 000744-6

O Instituto Nacional da Propriedade Industrial expede o presente certificado de Registro de Programas de Computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de Criação: 15 de abril de 2017, em conformidade com o parágrafo 2º, artigo 2º da Lei Nº 9.609, de 19 de Fevereiro de 1998.

Título: **RHO-FLUIDO**

Data de Criação: 15 de abril de 2017

Titular(es): SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL-SENAI/DR-BAHIA

Autor(es): ANDREA DE ALMEIDA BRITO
/ ARLEYS PEREIRA NUNES DE CASTRO
/ GILNEY FIGUEIRA ZEBENDE
/ IVAN COSTA DA CUNHA LIMA

Linguagem: C##

Campo de Aplicação: MT-01, MT-06

Tipo Programa: AP-01, IT-04, SM-01

Expedido em: 11 de julho de 2017

Aprovado por Liane Elizabeth Caldeira Lage

Liane Elizabeth Caldeira Lage

Diretor(a) Substituto(a) de Patentes, Programas de Computador e Topografia de Circuitos Integrados

15 de Novembro
REPÚBLICA FEDERATIVA DO BRASIL
de 1889



INPI INSTITUTO
NACIONAL
DE PROPRIEDADE
INDUSTRIAL
Assinado
Digitalmente

REPÚBLICA FEDERATIVA DO BRASIL

MINISTÉRIO DA ECONOMIA

INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512020001149-7**

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 01/07/2019, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: RHO_CUT

Data de publicação: 01/07/2019

Data de criação: 01/07/2019

Titular(es): ARLEYS PEREIRA NUNES DE CASTRO; GILNEY FIGUEIRA ZEBENDE; ANDREA DE ALMEIDA BRITO; FLORÊNCIO MENDES OLIVEIRA FILHO

Autor(es): ARLEYS PEREIRA NUNES DE CASTRO; GILNEY FIGUEIRA ZEBENDE; ANDREA DE ALMEIDA BRITO; FLORÊNCIO MENDES OLIVEIRA FILHO

Linguagem: C++

Campo de aplicação: MT-06

Tipo de programa: SM-01

Algoritmo hash: SHA-512

Resumo digital hash:

f835830504244c70281cc115b54bd9938a10162f6aabb7229a2712e5f9b47db417c3569b00c11da4b6a6239440cf7ec6a762e12d4fce897fe75546fa4cbe440f

Expedido em: 30/06/2020

Aprovado por:

Helmar Alvares

Chefe da DIPTO - Portaria/INPI/DIRPA Nº 09, de 01 de julho de 2019



INPI INSTITUTO
NACIONAL
DE PROPRIEDADE
INDUSTRIAL
Assinado
Digitalmente

REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DA ECONOMIA
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512020001150-0**

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 01/01/2020, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: RHO BA MAP

Data de publicação: 01/01/2020

Data de criação: 01/01/2020

Titular(es): ARLEYS PEREIRA NUNES DE CASTRO; GILNEY FIGUEIRA ZEBENDE; ANDREA DE ALMEIDA BRITO; FLORÊNCIO MENDES OLIVEIRA FILHO

Autor(es): ARLEYS PEREIRA NUNES DE CASTRO; GILNEY FIGUEIRA ZEBENDE; ANDREA DE ALMEIDA BRITO; FLORÊNCIO MENDES OLIVEIRA FILHO; MOÍSES DOMINGOS NAMILA DA COSTA

Linguagem: C++

Campo de aplicação: MT-06

Tipo de programa: SM-01

Algoritmo hash: SHA-512

Resumo digital hash:

132cc5856368b6a426e2cff4118d656101b1d468194bd291226982fc8af1413a977da579b8ba91f84fd9576d0eee0faba5c0df02f2780171396f069a9d89aaa9

Expedido em: 30/06/2020

Aprovado por:

Helmar Alvares

Chefe da DIPTO - Portaria/INPI/DIRPA Nº 09, de 01 de julho de 2019

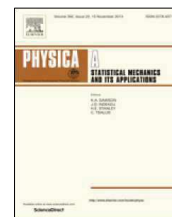
A.2 *Capítulos de Livro*

- Silva Filho, Aloísio Machado da; Zebende, Gilney Figueira; Guedes, Everaldo Freitas; Santos, Elvira Catiana de Oliveira; **Brito, Andréa de Almeida**; Castro, Arleys Pereira Nunes de; Oliveira Filho, Florêncio Mendes . **MODELAGEM DA DIFERENCIAÇÃO DO COEFICIENTE DE CORRELAÇÃO CRUZADA SEM TENDÊNCIA: UMA APLICAÇÃO NA SÉRIE TEMPORAL DE VEÍCULOS FURTADOS E VEÍCULOS ROUBADOS EM SALVADOR-BA**. Ciências Tecnológicas, Exatas e da Terra e seu Alto Grau de Aplicabilidade. 1a. ed.: Atena Editora, 2020, v. 1, p. 24-36.
- **Brito, Andréa de Almeida**; Silva Filho, Aloísio Machado da ; Lima, Ivan Costa da Cunha; Zebende, Gilney Figueira.**DCCA APLICADO ENTRE TEMPERATURA AMBIENTE E UMIDADE RELATIVA DO AR: MÉDIAS DISTINTAS**. As Diversidades de Debates na Pesquisa em Matemática 2. 1^a. ed.: Atena Editora, 2019, v. 1, p. 162-166.
- **Brito, Andréa de Almeida**; Cruz, Dênio Oliveira; Lima, Ivan Costa da Cunha; Zebende, Gilney Figueira. **MÉTODOS ESTATÍSTICOS APLICADOS A TEMPERATURA AMBIENTE E UMIDADE RELATIVA DO AR NA CIDADE DE PORTO SEGURO (BA)**. Ciência, Tecnologia e Inovação: Desafio para um Mundo Global 2. 2^a. ed.: Atena Editora, 2019, v. 2, p. 185-193.

APÊNDICE II

B.1 Outras Publicações

- E. F. Guedes; **A. A Brito**; F, M. Oliveira Filho; B. F. Fernandez; A. P. N. Castro; A. M. da Silva Filho; G. F. Zebende. “Statistical test for $\Delta\rho_{DCCA}$ cross-correlation coefficient ”, publicado na revista **Physica A: Statistical Mechanics and its Applications**, v. 501, p. 134–140, 2018.
- E. F. Guedes; **A. A Brito**; F, M. Oliveira Filho; B. F. Fernandez; A. P. N. Castro; A. M. da Silva Filho; G. F. Zebende. “Statistical test for $\Delta\rho_{DCCA}$: Methods and data ”, publicado na revista **Data in Brief**, v. 18, p. 795–798, 2018.
- F. Rodrigues Santos; **A. A Brito**; A. P. N. Castro; M. P. Almeida; A. T. da Cunha Lima; G. F. Zebende; I. C. da Cunha Lima. “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 ”, publicado na revista **Physica A: Statistical Mechanics and its Applications**, v. 509, p. 294–301, 2018.
- **A. A Brito**; Rodrigues Santos; A. P. N. Castro; A. T. da Cunha Lima; G. F. Zebende; I. C. da Cunha Lima. “Cross-correlation in a turbulent flow: Analysis of the velocity field using the ρ_{DCCA} coefficient ”, publicado na revista **Europhysics Letter**, v. 123, p. 20011, 2018.



Statistical test for $\Delta\rho_{DCCA}$ cross-correlation coefficient

E.F. Guedes^a, **A.A. Brito**^{a,b}, F.M. Oliveira Filho^{a,c}, B.F. Fernandez^d,
A.P.N. de Castro^e, A.M. da Silva Filho^f, G.F. Zebende^{f,g,*}

^a Computational Modeling Program, SENAI CIMATEC, Salvador, Bahia, Brazil

^b Federal Institute of Paulo Afonso, Bahia, Brazil

^c Estácio de Sá University, Gilberto Gil Campus, Salvador, Bahia, Brazil

^d Municipal Foundation Egberto Costa, Feira de Santana, Bahia, Brazil

^e Jorge Amado University Center, Bahia, Brazil

^f Earth Sciences and Environment Modeling Program, State University of Feira de Santana, Bahia, Brazil

^g Department of Physics, State University of Feira de Santana, Bahia, Brazil



H I G H L I G H T S

- We proposed a new statistical test for $\Delta\rho_{DCCA}$.
- We have analyzed simulated and real time series.
- We define a new critical value for this dependency at specific confidence level.

A R T I C L E I N F O

Article history:

Received 2 June 2017

Received in revised form 13 November 2017

Available online 15 March 2018

Keywords:

Time series

DCCA cross-correlation coefficient

Statistical test

A B S T R A C T

In this paper we propose a new statistical test for $\Delta\rho_{DCCA}$, Detrended Cross-Correlation Coefficient Difference, a tool to measure contagion/interdependence effect in time series of size N at different time scale n . For this proposition we analyzed simulated and real time series. The results showed that the statistical significance of $\Delta\rho_{DCCA}$ depends on the size N and the time scale n , and we can define a critical value for this dependency in 90%, 95%, and 99% of confidence level, as will be shown in this paper.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

The Detrended cross-correlation coefficient, ρ_{DCCA} , proposed by Zebende [1] intended to measure the level of cross-correlation between two non-stationary times series of size N . Since its implementation, ρ_{DCCA} has become an important tool for cross-correlation analysis and has been applied in many areas of knowledge, such as: climatology [2–4], crime indicators [5,6], economy [7–20], in geology [21], among others.

Unlike other statistical methods (like Pearson correlation coefficient or [22,23]), ρ_{DCCA} algorithm takes into account the temporal order of data, which does not contradict the basic principle of time series analysis, e.g., the temporal dependence between the variables. In this way, ρ_{DCCA} has some advantages over other cross-correlation coefficients, such as:

- measures cross-correlation between two time series at different time scales n [9];
- is a robust tool for nonlinear time series analyses [6,24].

* Corresponding author.

E-mail address: gfzebende@pq.cnpq.br (G.F. Zebende).

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Data in Brief

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

Data Article

Statistical test for $\Delta\rho_{DCCA}$: Methods and data

E.F. Guedes^a, **A.A. Brito**^{a,b}, F.M. Oliveira Filho^{a,c},
 B.F. Fernandez^d, A.P.N. de Castro^e, A.M. da Silva Filho^f,
 G.F. Zebende^{g,f}

^a Computational Modeling Program, SENAI CIMATEC, Salvador, Bahia, Brazil^b Federal Institute of Paulo Afonso, Bahia, Brazil^c Estácio de Sá University, Gilberto Gil Campus, Salvador, Bahia, Brazil^d Municipal Foundation Egberto Costa, Feira de Santana, Bahia, Brazil^e Jorge Amado University Center, Bahia, Brazil^f Earth Science and Environment Modeling Program, State University of Feira de Santana, Bahia, Brazil^g Department of Physics, State University of Feira de Santana, Bahia, Brazil

ARTICLE INFO

Article history:

Received 24 February 2018

Received in revised form

10 March 2018

Accepted 16 March 2018

Available online 22 March 2018

ABSTRACT

In this paper the algorithm for $\Delta\rho_{DCCA}$ statistical test (Guedes et al., 2018) [1] is presented. Our test begins with the simulation of four time series pairs, by an ARFIMA process. These time series has $N = 250, 500, 1000,$ and 2000 points, see Guedes et al. (2018) [1]. The probability distribution function (PDF) is made available for all 10,000 samples, that start from the original time series, in supplementary material.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Physics and Astronomy
More specific subject area	General physics and methods
Method name	$\Delta\rho_{DCCA}$ Statistical test

DOI of original article: <https://doi.org/10.1016/j.physa.2018.02.148>E-mail address: gfzebende@pq.cnpq.br (G.F. Zebende).<https://doi.org/10.1016/j.dib.2018.03.080>2352–3409/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



Contents lists available at ScienceDirect

Physica A

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

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



F. Rodrigues Santos^{a,b}, **A.A. Brito**^{a,c}, A.P.N. de Castro^d, M.P. Almeida^{e,f},
A.T. da Cunha Lima^{a,f,g}, G.F. Zebende^{h,i}, I.C. da Cunha Lima^{a,f,j,*}

^a Computational Modeling Program, Centro Universitário SENAI CIMATEC, Salvador, Bahia, Brazil

^b Federal Institute of Bahia, Camaçari, Bahia, Brazil

^c Federal Institute of Bahia, Paulo Afonso, Bahia, Brazil

^d Unijorge University Center, Salvador, Bahia, Brazil

^e Federal University of Ceará, Fortaleza, Ceará, Brazil

^f National Institute for Science and Technology–Petroleum Geophysics, INCT-GP, Salvador, Bahia, Brazil

^g Federal University of Bahia, Salvador, Bahia, Brazil

^h Department of Physics, State University of Feira de Santana, Bahia, Brazil

ⁱ Earth Sciences and Environment Modeling Program, State University of Feira de Santana, Bahia, Brazil

^j Pursuelife—Consultancy on Applied Science, Salvador, Bahia, Brazil

H I G H L I G H T S

- The flow near the walls preserve memory of blockage symmetry for longer distances.
- The correlations at different time scales are sensitive to the blockages symmetry.
- The cross-correlations in mid-channel are stronger than close to the pipeline wall.
- Correlations increase at large time scales as the points move away from blockages.

A R T I C L E I N F O

Article history:

Received 10 April 2018

Received in revised form 30 May 2018

Available online 13 June 2018

Keywords:

Turbulence

Navier–Stokes equation

Pipeline

ρ_{DCCA}

A B S T R A C T

In this paper we analyze the influence of obstacles symmetry on the development of the turbulent flow of a fluid through a pipeline. The analysis is based on the numerical solutions of the Navier–Stokes equations for the velocity field. The influence of the obstacle symmetry on the turbulence is detected by changing their shape while keeping the blockage ratio constant and calculating velocity field cross-correlations on the time series resulting from the simulation. The Detrended Cross-Correlation coefficient (ρ_{DCCA}) is applied to obtain two-point correlations located at different regions of the channel: at mid-channel and near the walls (above and below). With this cross-correlation coefficient we quantify how far from the obstructions these coefficients become independent on the obstructions shape, establishing a scale for the obstruction symmetry memory loss.

© 2018 Elsevier B.V. All rights reserved.

* Corresponding author at: Computational Modeling Program, Centro Universitário SENAI CIMATEC, Salvador, Bahia, Brazil.
E-mail address: ivandacunhalima@pq.cnpq.br (I.C. da Cunha Lima).

Cross-correlation in a turbulent flow: Analysis of the velocity field using the ρ_{DCCA} coefficient

A. A. BRITO^{1,5}, F. RODRIGUES SANTOS^{2,5}, A. P. N. DE CASTRO³, A. T. DA CUNHA LIMA^{4,5,6}, G. F. ZEBENDE^{7,8} and I. C. DA CUNHA LIMA^{5,6,9}

¹ Federal Institute of Bahia - Paulo Afonso, Bahia, Brazil

² Federal Institute of Bahia - Camaçari, Bahia, Brazil

³ Unijorge University Center - Salvador, Bahia, Brazil

⁴ Federal University of Bahia - Salvador, Bahia, Brazil

⁵ Computational Modeling Program, SENAI CIMATEC - Salvador, Bahia, Brazil

⁶ National Institute for Science and Technology - Petroleum Geophysics, INCT-GP - Salvador, Bahia, Brazil

⁷ Department of Physics, State University of Feira de Santana - Feira de Santana, Bahia, Brazil

⁸ Earth Sciences and Environment Modeling Program, State University of Feira de Santana
Feira de Santana, Bahia, Brazil

⁹ Pursuelife - Consultancy on Applied Science - Salvador, Bahia, Brazil

received 24 May 2018; accepted in final form 25 July 2018

published online 24 August 2018

PACS 05.20.Jj – Statistical mechanics of classical fluids

PACS 02.60.-x – Numerical approximation and analysis

PACS 05.45.Tp – Time series analysis

Abstract – The stochastic process of a turbulent flow in a pipeline provides a time series of the velocity field at any point of the domain by solving numerically the Navier-Stokes equation. The turbulent flow was produced by obstacles near the inlet, injecting eddies into the current. Moving downstream, these eddies evolve to a fully turbulent flow. Many length and time scales are involved in this process. We explore the cross-correlations of the velocity field time series at different points and also at different time scales using the detrended cross-correlation coefficient, ρ_{DCCA} , designed to analyze the cross-correlations in non-stationary time series. Thus, the results with ρ_{DCCA} allow interpreting how these eddies propagate downstream, and also quantify how adherent the velocity fields are with respect to the pipeline position.

Copyright © EPLA, 2018

Introduction. – It is well known that the great majority of flows in nature are turbulent. A flow in a pipeline past an obstruction has complex structures just downstream from the obstruction, which evolves to a fully turbulent flow [1]. The observation of the flow in the fully turbulent region for a large enough duration determines the velocity field as being a stationary stochastic process. In this turbulent regime, the averages of functions of the velocity fields taken from a given initial time over a large time interval are independent of the initial time [2]. The Probability Density Function (PDF) describing the stochastic velocity field at a given position is time independent, *i.e.*, stationary. Even before reaching the fully turbulent regime, a flow in a pipeline supplied by a steady inflow is necessarily stationary, but the process of obtaining a time-independent PDF takes a very long measurement interval.

Advances in high-performance computers over the last decade have provided the opportunity of calculating the velocity field of turbulent flows in complex domains with great accuracy. The area of Computational Fluid Dynamics (CFD) has benefitted from these advances [3]. Flows of classical fluids where momentum conservation is described by the Navier-Stokes equations [4] can be simulated even at high Reynolds numbers, where the ratio of inertial to viscous forces is very high. The transition to turbulence occurs when the Reynolds number reaches a threshold which depends on the nature of the structure containing the flow. Random perturbations near this threshold lead to the transition to turbulence. A numerical solution of the Navier-Stokes equation, even without including such a random perturbation *ad hoc*, has an intrinsically stochastic nature. In fact, starting from an initial condition known at a mesh of points in the domain

Modelagem Computacional de Auto-correlações e Correlações Cruzadas Aplicada em Variáveis Meteorológicas e Simuladas

Andrea de Almeida Brito

Salvador, 29 de janeiro de 2021.