



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM EM CIÊNCIAS
DA TERRA E DO AMBIENTE
Mestrado em Modelagem em Ciências da Terra e do Ambiente

Dissertação de Mestrado

**Análise das correlações em séries temporais de
acidentes de trânsito e de outras séries**

Apresentada por: ROBERTA PIMENTA CUNHA DOS SANTOS
Orientador: Prof. Dr. Gilney Figueira Zebende
Co-orientador: Prof. Dr. Edrian Mania

Agosto de 2022

ROBERTA PIMENTA CUNHA DOS SANTOS

Análise das correlações em séries temporais de acidentes de trânsito e de outras séries

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Modelagem em Ciências da Terra e do Ambiente da UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA, como requisito parcial para a obtenção do título de **Mestre em Ciências Ambientais**.

Área de conhecimento: Estudos Ambientais e Geotecnologias

Orientador: Prof. Dr. Gilney Figueira Zebende

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Co-orientador: Prof. Dr. Edrian Mania

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Feira de Santana, BA
UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA
2022

Ficha Catalográfica - Biblioteca Central Julieta Carteadó - UEFS

S238 Santos, Roberta Pimenta

Análise das correlações em séries temporais de acidentes de trânsito e de outras séries / Roberta Pimenta Santos. – 2022.

138 f.: il.

Orientador: Gilney Figueira Zebende.

Coorientador: Edrian Mania.

Dissertação (mestrado) – Universidade Estadual de Feira de Santana, Programa de Pós-graduação em Modelagem em Ciências da Terra e do Ambiente, Feira de Santana, 2022.

1. Acidentes de trânsito. 2. Rodovias federais – Bahia. 3. Séries temporais. I. Título. II. Zebende, Gilney Figueira, orient. III. Mania, Edrian, coorient. IV. Universidade Estadual de Feira de Santana.

CDU 614.86(814.2)

Roberta Pimenta Cunha dos Santos

**“ANÁLISE DAS CORRELAÇÕES EM SÉRIES TEMPORAIS DE
ACIDENTES DE TRÂNSITO E DE OUTRAS SÉRIES”**

Dissertação apresentada ao Programa de
Pós-Graduação em Modelagem em Ciências da
Terra e do Ambiente da Universidade Estadual
de Feira de Santana.

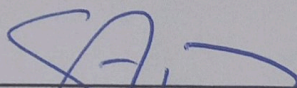
Orientador: Prof. Dr. Gilney Figueira Zebende.

Coorientação: Prof. Dr. Edrian Mania

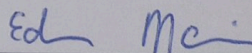
Linha de Pesquisa: Estudos Ambientais e
Geotecnologias.

Data de aprovação: 24 de agosto de 2022.

BANCA EXAMINADORA:



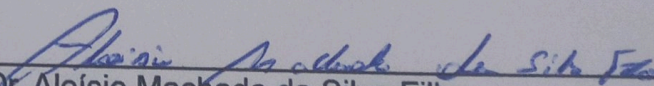
Prof. Dr. Gilney Figueira Zebende
Universidade Estadual de Feira de Santana - UEFS



Prof. Dr. Edrian Mania
Universidade Estadual de Feira de Santana - UEFS



Prof. Dr. Cassiano Augusto Isler
Universidade de São Paulo - USP



Prof. Dr. Aloísio Machado da Silva Filho
Universidade Estadual de Feira de Santana - UEFS

Para Sérgio e Raphael, meus tesouros e a Ricardo,
por ter me convencido de que é possível alcançar
qualquer objetivo quando nos dedicamos de verdade.

Agradecimentos

Agradeço a Deus, meu Consolador, que de maneira extraordinária se revela por meio da ciência, e assim nos encantamos pela vida.

À minha família, que sempre me apoiou em todos os projetos, vocês são a minha base. Em especial, a meu irmão Ricardo, pelo incentivo constante, seu exemplo de esforço e dedicação me encorajam a ir além.

Minha gratidão pelo apoio e dedicação dos meus orientadores, professor Prof. Dr. Gilney Figueira Zebende que acreditou e confiou a mim a ideia original do projeto para que juntos pudéssemos desenvolver este trabalho inovador, e professor Prof. Dr. Edrian Mania que esteve sempre disponível a contribuir e a me desafiar.

Ao professor Dr. Everaldo Guedes, a quem sou muitíssimo grata pelo incentivo e auxílio tão importantes para a realização deste projeto.

Aos colegas e professores do PPGM, a troca de conhecimento foi essencial para eu chegar aqui. Obrigada pelas contribuições para concretizar esta dissertação, compartilhando o que sabem e me impelindo a pensar.

Jamais poderei agradecer o suficiente ao meu esposo Sérgio e ao meu filho Raphael que, amorosamente, suportaram minha ausência e agenda apertada em alguns momentos divertidos que passaríamos juntos. Vocês são incríveis!

*“A Matemática é enfim uma das verdades eternas
e, como tal, produz a elevação do espírito
- a mesma elevação que sentimos ao contemplar
os espetáculos da Natureza, através dos quais
sentimos a presença de Deus, Eterno e Onipotente!
...Conto os versos de um poema, calculo a altura
de uma estrela, avalio o número de franjas, meço a área
de um país, ou a força de uma torrente - aplico, enfim,
fórmulas algébricas e princípios geométricos
- sem me preocupar com os louros que possa tirar
de meus cálculos e estudos! Sem o sonho e a fantasia
a ciência se abastarda. É ciência morta!”*

(Malba Tahan - O Homem que Calculava, cap. XIV)

Resumo

Acidentes de trânsito representam a oitava causa de morte para todas as faixas etárias no mundo. No Brasil, os fluxos de bens e pessoas ocorrem, majoritariamente, por meio do sistema rodoviário. Ao mesmo tempo, o país ocupa o quinto lugar no *ranking* dos países com maiores índices de morte no trânsito, atrás apenas da Índia, China, EUA e Rússia. Nesse sentido, os acidentes de trânsito representam um problema de saúde pública e também, porque não dizer, um problema ambiental. Em decorrência do aumento absoluto de vítimas fatais dos acidentes ao longo dos anos, faz-se necessário o embasamento científico, para direcionar as ações de enfrentamento a essa problemática. Com o intuito de contribuir com a formulação de políticas públicas para a redução de acidentes de trânsito nas rodovias federais, este trabalho tem como objetivo geral analisar as variáveis dos acidentes referentes ao número total de: *peças, feridos, feridos leves, feridos graves, ilesos, ignorados e veículos* envolvidos nos acidentes registrados nas rodovias federais da Bahia. Pretendemos também representar a estrutura, além de caracterizar o comportamento dessas variáveis em função do tempo, por meio da aplicação do Coeficiente de Correlação Cruzada (ρ_{DCCA}). Esse coeficiente se baseia nos métodos da mecânica estatística: o DFA (*Detrended Fluctuation Analysis*), usado para identificar autocorrelação de longo alcance e o DCCA (*Detrended Cross-Correlation Analysis*) que estima o expoente para caracterizar a correlação cruzada entre duas séries temporais não estacionárias de mesmo tamanho N . Diversos campos do conhecimento já foram modelados por meio do ρ_{DCCA} , a exemplo do estudo de acidentes de táxi, fenômenos climáticos, mercado financeiro, indicadores de criminalidade, comportamento da produção industrial, dinâmica temporal dos casos de dengue, dentre outros. Para alcançar esse objetivo, analisamos as séries de dados referentes aos acidentes reportados pela Polícia Rodoviária Federal (PRF) da Bahia, entre 2007 e 2021. De maneira geral, analisaremos a existência de correlação cruzada de longo alcance entre a variável *mortos* com as demais variáveis visto que, reduzir o número de vítimas fatais é um dos principais objetivos da PRF para mitigar a violência no trânsito. Os resultados do ρ_{DCCA} indicaram correlação cruzada em todos os cruzamentos realizados, o que mostra que as oscilações na quantidade de *Mortos* é afetada pelas oscilações nas quantidades das demais variáveis, segundo uma correlação positiva (altas seguidas de altas ou baixas seguidas de baixas). Dentre outras características, os acidentes tiveram como causa principal a postura indevida do condutor, sendo que a maioria deles ocorreu nas três principais rodovias do estado, BR 101, BR 324 e BR 116 e a maior parte das pessoas envolvidas eram do sexo masculino.

Palavras Chave: Séries temporais, DFA, DCCA, ρ_{DCCA} , acidentes de trânsito.

Abstract

Traffic accidents represent the eighth leading cause of death for all age groups in the world. In Brazil, the flows of goods and people occur mainly through the road system. At the same time, the country occupies fifth place in the ranking of countries with the highest rates of road traffic deaths, behind only India, China, USA, and Russia. In this sense, traffic accidents represent a health problem public and also, why not say, an environmental problem. As a result of the increased number of fatal accident victims over the years, it is necessary to scientific basis, to direct the actions to face this problem. To contribute to the formulation of public policies for the reduction of traffic accidents on federal highways, this work has the general objective of analyzing the accident variables refer to the total number of people, injured, injured light, seriously injured, unharmed, ignored and vehicles involved in accidents recorded on the federal highways of Bahia. We also intend to represent the structure, in addition to characterize the behavior of these variables as a function of time, through the application of the cross-correlation coefficient (ρ_{DCCA}). This coefficient is based on the methods of Statistical mechanics: DFA (Detrended Fluctuation Analysis), used to identify long-range autocorrelation and DCCA (Detrended Cross-Correlation Analysis), estimates the exponent that characterizes the cross correlation between two non-stationary time series of the same size N . Several fields of knowledge have already been modeled through of the ρ_{DCCA} , such as the study of taxi accidents, climatic phenomena, financial market, crime indicators, industrial production behavior, dynamics of dengue cases, among others. To achieve this objective, we analyzed the data series referring to accidents reported by the Federal Highway Police (PRF) of Bahia, between 2007 and 2021. In general, we analyzed the existence of long-range cross-correlation between the dead variable with the others variables since reducing the number of fatal victims are one of the main objectives of PRF to mitigate traffic violence. The ρ_{DCCA} results indicated a correlation in all the crossings carried out, which shows that the fluctuations in the amount of Mortos is affected by fluctuations in the quantities of the other variables, according to a positive correlation (highs followed by highs or lows followed by lows). Among other characteristics, accidents were mainly caused by improper posture of the driver, most of which took place on the three main highways in the state, BR 101, BR 324, and BR 116, and most of the people involved were male.

Keywords: Time series, DFA, DCCA, ρ_{DCCA} , accidents traffic.

Sumário

1	Introdução	1
2	Fundamentação Teórica	6
2.1	Acidentes de Trânsito	6
2.2	Iniciativas para promover segurança no tráfego	8
2.3	Análise estatística de acidentes de trânsito	13
2.3.1	Trabalhos correlatos	13
3	Materiais e Método	18
3.1	Base de dados	18
3.2	Área de estudo	20
3.3	Metodologia	21
3.3.1	Análise de séries temporais	21
3.3.2	Contribuições de um SIG para monitoramento e prevenção de acidentes de trânsito	27
3.3.3	Método <i>R/S</i>	28
3.3.4	Método DFA	29
3.3.5	Método DCCA	33
3.3.6	Coefficiente de Correlação Cruzada	35
3.4	Percurso metodológico	36
4	Resultados e Conclusões	38
4.1	Caracterização dos acidentes	38
4.2	Séries temporais dos acidentes	49
4.2.1	Análise exploratória	49
4.2.2	Resultados 2007-2021	57
4.2.2.1	Autocorrelação das séries	57
4.2.2.2	Correlação cruzada entre as séries	62
4.2.3	Resultados ano a ano	65
4.2.3.1	Autocorrelação das séries	65
4.2.3.2	Correlação cruzada entre as séries	67
4.2.4	Variação relativa entre os anos	72
4.2.4.1	Autocorrelação das séries	72
4.2.4.2	Correlação cruzada entre as séries	74
4.3	Conclusões	75
A	Artigos publicados	78
	Referências	116

Lista de Tabelas

3.1	Classificação da faixa dinâmica e valores do expoente de flutuação α_{DFA}	33
3.2	Níveis de correlação cruzada	35
4.1	Análise descritiva das séries temporais diárias das variáveis de acidentes na Bahia, de 1 ^o de janeiro de 2007 a 31 de dezembro de 2021	49
4.2	Resultados do Teste Jarque-Bera, Bahia, 2007-2021	50
4.3	Resultados do ADF, Bahia, 2007-2021	52
4.4	Resultados do Teste Ljung-Box, Bahia, 2007-2021	54
4.5	Expoente de autocorrelação, Bahia, 2007-2021	58
4.6	Mortos: valores médios do coeficiente ρ_{DCCA}	65
4.7	Expoentes α_{DFA} subtraído de 0,5 ($\alpha_{DFA}-0,5$), 2007-2021	66
4.8	Mortos: Valor médio por ano do ρ_{DCCA} , Bahia, 2007-2021	68

Lista de Figuras

2.1	Mapa conceitual das ações de segurança para o trânsito.	12
3.1	Variáveis dos acidentes.	19
3.2	Exemplo de planilha com a base de dados. Note que de 2017 em diante temos a localização georreferenciada das ocorrências.	20
3.3	Localização da área de estudo com aplicação às rodovias federais.	21
3.4	Totais diários de vítimas fatais em acidentes nas rodovias federais da Bahia, 2007-2021. Fonte: PRF	22
3.5	Totais mensais de acidentes nas rodovias ferais da Bahia, 2007-2021. Fonte: PRF	22
3.6	Totais por ano de feridos em acidentes nas rodovias federais da Bahia, 2007-2021. Fonte: PRF	22
3.7	Totais por ano de ilesos em acidentes nas rodovias federais da Bahia, 2007-2021. Fonte: PRF	22
3.8	Bahia: Série temporal das pessoas envolvidas acidentes de trânsito nas rodovias federais, 2007-2021.	30
3.9	Bahia: Série integrada das pessoas envolvidas acidentes de trânsito $\{X_k\}$ nas rodovias federais, 2007-2021. Nota: Cálculo do autor.	31
3.10	Divisão do sinal integrado X_k em $N-n$ boxes sobrepostos de tamanho $n = 4$, contendo $n+1$ valores.	32
3.11	Comportamento do expoente α das pessoas envolvidas em acidentes de trânsito nas rodovias federais da Bahia em 2007-2021.	33
3.12	Esquema do processo metodológico empregado.	37
4.1	Bahia: Percentuais de pessoas envolvidas nos acidentes, segundo a situação da vítima, 2007-2021.	38
4.2	Bahia: Evolução dos acidentes de trânsitos registrados nas rodovias federais do estado da Bahia, 2007-2021. Fonte: PRF. Nota: Dados processados pelo autor.	39
4.3	Bahia: Evolução da quantidade de pessoas envolvidas em acidentes de trânsito segundo a situação de gravidade, 2007-2021.	40
4.4	Bahia: Média diária anual das séries de acidentes, 2007-2021.	41
4.5	Evolução dos acidentes por municípios da Bahia, 2017-2020.	44
4.6	Concentração de mortos nos municípios com maior volume de ocorrências registradas em 2017-2020.	46
4.7	Feira de Santana, Bahia: Concentração de mortos nos municípios com maior volume de ocorrências registradas em 2017-2020.	47
4.8	Vitória da Conquista, Bahia: Concentração de mortos nos municípios com maior volume de ocorrências registradas em 20017-2020.	48
4.9	Histogramas das séries temporais de acidentes, Bahia, 2007-2021.	51
4.10	Correlogramas das séries temporais e histogramas dos resíduos das séries de acidentes, Bahia, 2007-2021.	53
4.11	Decomposição das séries temporais de acidentes	56

4.12	Bahia: Quantidade de pessoas envolvidas em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de pessoas envolvidas em acidentes, 2007-2021 (b)	58
4.13	Bahia: Quantidade de vítimas fatais em acidentes, 2007-2021 (a). F_{DFA} em função de n no total de vítimas fatais em acidentes, 2007-2021 (b)	59
4.14	Bahia: Quantidade de feridos leves em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de feridos leves em acidentes, 2007-2021 (b)	59
4.15	Bahia: Quantidade de feridos graves em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de feridos graves em acidentes, 2007-2021 (b)	60
4.16	Bahia: Quantidade diária de feridos em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de feridos em acidentes, 2007-2021 (b)	60
4.17	Bahia: Quantidade diária de ilesos em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de ilesos em acidentes, 2007-2021 (b)	61
4.18	Bahia: Quantidade diária de ignorados em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de ignorados em acidentes, 2007-2021 (b)	61
4.19	Bahia: Quantidade diária de veículos envolvidos em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de veículos envolvidos em acidentes, 2007-2021 (b)	62
4.20	Correlação entre <i>Mortos</i> e as demais séries de acidentes reportados pela PRF, Bahia, 2007 a 2021.	63
4.21	Correlação cruzada de <i>Mortos</i> com <i>Pessoas</i> , <i>Feridos leves</i> , <i>Feridos graves</i> , <i>Feridos</i> , <i>Ignorados</i> e <i>Veículos</i> , Bahia, 2007-2021.	64
4.22	Variação anual do coeficiente de autocorrelação das séries temporais de acidentes, 2007-2021.	67
4.23	Variação anual do coeficiente de autocorrelação das séries temporais de acidentes, 2007-2021.	69
4.24	Correlação cruzada de <i>Mortos</i> com <i>Pessoas</i> (●), <i>Feridos leves</i> (◀), <i>Feridos graves</i> (◆), <i>Feridos</i> (★), <i>Ilesos</i> (●), <i>Ignorados</i> (■) e <i>Veículos</i> (▼), 2007 a 2021.	72
4.25	Expoente DFA: variação relativa entre os anos, Bahia, 2007-2021	74
4.26	<i>Mortos</i> : correlação cruzada - variação relativa anual do coeficiente de correlação cruzada, Bahia, 2007-2021.	75

Lista de Siglas

ABNT	Associação Brasileira de Normas Técnicas
ABRAMET	Associação Brasileira de Acidentes de Tráfego
ABS	<i>Anti-lock Braking System</i>
ARMA	<i>Autoregressive moving average</i>
ARIMA	<i>Autoregressive integrated moving average</i>
AST	Análise de Séries Temporais
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CONTRAN	Conselho Nacional de Trânsito
CTB	Código de Trânsito Brasileiro
DCCA	<i>Detrended Cross-Correlation Analysis</i>
DFA	<i>Detrended Fluctuation Analysis</i>
FIBA	Federação Internacional de Basquete
Hab.	Habitantes
INMET	Instituto Nacional de Meteorologia
NBR	Norma Brasileira
ODM	Objetivos de Desenvolvimento do Milênio
ODS	Objetivos de Desenvolvimento Sustentável
OMS	Organização Mundial de Saúde
ONU	Organização das Nações Unidas
OPAS	Organização Pan-Americana de Saúde
PIB	Produto Interno Bruto
PPGM	Programa de Pós-Graduação em Modelagem em Ciências da Terra e do Ambiente
PRF	Polícia Rodoviária Federal
SIG	Sistema de Informação Geográfica
ST	Séries Temporais
UEFS	Universidade Estadual de Feira de Santana
UTC	<i>Coordinated Universal Time</i>
ρ_{DCCA}	Coefficiente de Correlação Cruzada

Introdução

As relações humanas têm se tornado cada vez mais complexas, sobretudo com o advento das novas tecnologias. Dentre outros espaços, o trânsito também abriga reflexos dessa complexidade e o comportamento do motorista tem sido considerado uma das principais causas de acidentes nas rodovias. Diversos estudos apontam que os acidentes de trânsito, em muitos casos, são evitáveis e, portanto, estão relacionados ao comportamento do condutor, ver (BOATENG, 2021; FEBRES et al., 2020; PEREYRA, 2013). Assim, a maneira como dirigimos pode estar associada ao nosso funcionamento psicológico, revelando escolhas, hábitos e crenças e até mesmo aspectos de nossa personalidade (ONSV, 2021). Nessa perspectiva, sabe-se que o trânsito possui muitas unidades em interação cujos agentes (motorista + carro) podem ou não ser idênticos. As interações ocorrem entre elas e com o meio externo (acelerar, frear, mudar de faixa) das quais originam fenômenos emergentes (engarrafamentos, acidentes). No entanto, tais fenômenos são auto-organizados podendo ocorrer sem a necessidade de agentes externos, sendo caracterizado portanto, como um sistema complexo (GARCIA, 2017).

Uma pesquisa realizada em 2018 (WHO, 2018) revelou que acidentes de trânsito representam a oitava causa de morte para todas as faixas etárias no mundo e a principal causa de morte de crianças e jovens de 5 a 29 anos. O fardo das lesões e mortes causadas pelo trânsito é desproporcionalmente suportado pelos usuários vulneráveis das estradas e pelos que vivem em países de baixa e média renda, onde o número crescente de mortes é alimentado por transportes que são cada vez mais motorizados. Segundo a Organização Mundial de Saúde (OMS), acidentes de trânsito representaram o sétimo lugar no *ranking* de causa de morte nos países considerados de baixa renda em 2019 e os acidentes e mortes apresentam um padrão em países em desenvolvimento (WAISELFISZ, 2013; BAHADO-RIMONFARED et al., 2013; MARÍN; QUEIROZ, 2000; MARÍN-LEÓN et al., 2012). No grupo dos países de categorias de renda média-baixa e média-alta, lesões na estrada ocupa a décima posição. A alta renda é a única categoria de países por grupo de renda em que esta causa de morte não está entre as dez principais (OPAS, 2019).

Diante desse cenário, as consequências dos acidentes de trânsito refletem nos setores econômico e social (CARMO; NERY; ROCHA, 2019; N.SODERLUND; ZWI, 1995). A perda econômica com o alto custo e as consequências deixadas pelos acidentes, sobretudo no tocante ao valor da vida, refletem a relação complexa do homem com o ambiente, bem como as transformações feitas por ele. Apesar do aprofundamento deste tema mostrar-se presente em diversos trabalhos de relevância social e acadêmica (RIBEIRO et al., 2020; WHO, 2018; PEREYRA, 2013; WAISELFISZ, 2013; CARVALHO, 2005), a segurança no

trânsito ainda denota uma problemática carente de soluções mais eficazes, revelando-se como um aspecto essencial para a continuidade da vida. Nesse contexto, os acidentes de trânsito representam um problema de saúde pública (BAHADORIMONFARED et al., 2013; JORGE; LATORRE, 1994) e porque não dizer, um problema ambiental, visto que fatores humanos e ambientais contribuem para os acidentes de trânsito (BOATENG, 2021). Paralelamente, segurança para estrada e veículos, legislação, atendimento profissional de emergência, além do conhecimento sobre a correlação a longo prazo das variáveis de acidentes são elementos que contribuem para a mitigação da violência no trânsito, já que a análise dos fatores influenciadores de acidentes podem auxiliar no estudo de comportamento dos motoristas (JIN et al., 2021).

A taxa de mortes em relação ao tamanho da população mundial se estabilizou e diminuiu em relação ao número de veículos motorizados nos últimos anos (WHO, 2018). Entretanto, mesmo com as modificações na legislação em diversos países e a implementação de itens de segurança pelos fabricantes de veículos, o volume de vítimas fatais nas estradas continua alto. De acordo com o último relatório sobre segurança no trânsito elaborado pela Organização Pan-Americana da Saúde (OPAS), intitulado *Estado de segurança viária na região das Américas*, o Brasil ocupa a nona posição entre os países com maior número de mortes nas estradas nas Américas. O país fica atrás apenas de Santa Lúcia, República Dominicana, Venezuela, Belize, Guiana, Paraguai, El Salvador e Equador (OPAS, 2019).

Em 2019, o Brasil registrou 16 mortes por 100 mil pessoas causadas por acidentes de trânsito, sendo que, somente entre os homens este indicador atingiu a marca de 26,9 mortes por mil pessoas, representando a 6^a principal causa de morte entre brasileiros do sexo masculino (ORGANIZATION, 2020). Os dados de acidentes disponibilizados no portal virtual da Polícia Rodoviária Federal (PRF) mostram que a Bahia se destaca por possuir o segundo maior volume de vítimas fatais. O estado é seguido apenas por Minas Gerais, que possui a malha rodoviária mais extensa do país (CNT, 2019).

No mesmo ano, 2019, a Bahia foi o terceiro estado com maior número de mortes por acidentes nas estradas, com 507 óbitos, ficando atrás apenas de Minas Gerais (717) e Paraná (526). Como característica viária, a Bahia concentra duas rodovias importantes que compõem um dos principais eixos rodoviários e as maiores em extensão do país: a BR 101 (4,77 mil km) e a BR 116 (4,70 mil km), atravessando 10 e 12 estados, respectivamente. Segundo o anuário estatístico elaborado pela PRF, em 2020 essas rodovias estão entre as dez com maior frequência de mortos do país, sendo que a BR 101/BA ocupa a terceira posição.

Essas duas estradas também registraram o maior volume de vítimas fatais em acidentes no ano de 2020, sendo Feira de Santana o segundo município do estado com maior número de acidentes nesse período (258). No ano anterior, em 2019, o município ocupou a primeira

posição (366) no *ranking* nacional. Importante ressaltar que esses dados se referem aos acidentes ocorridos somente nas rodovias federais, o que representa apenas 5,3% de um total de 1.720.700 km de malha rodoviária no país (CNT, 2019).

Analisando dados mais recentes, em 2021, do total de 5.381 vítimas que perderam suas vidas em acidentes nas rodovias federais brasileiras reportados pela PRF, metade (50,9%) estão concentradas somente em 6 estados do país (Minas Gerais, (12,9%), Paraná (10,6%), Bahia (9,3%), Santa Catarina (6,6%), Rio Grande do Sul (5,9%) e Rio de Janeiro (5,7%). Este *ranking* se repete em 2020, mudando apenas a quinta posição que passou a ser ocupada pelo estado de Pernambuco, praticamente, o mesmo cenário da ano anterior. Em 2019, Pernambuco e Rio de Janeiro assumiram a quinta e sexta posições, respectivamente.

Considerando o desafio e escassez de estudos relacionadas ao estado da Bahia, faz-se necessária a realização de mais pesquisas que abordem essa problemática. Para isso, a análise aprofundada dos dados de acidentes e seu comportamento em função do tempo e longo prazo podem ser uma possível solução para compreensão e, sobretudo, para a mitigação dos acidentes fatais. Diante dessa limitação, a motivação dessa pesquisa surgiu a partir de uma demanda da PRF junto à Universidade Estadual de Feira de Santana (UEFS). Assim, com o intuito de contribuir para soluções desenvolvendo estudo e geração de informações, o presente projeto procura responder a seguinte questão de pesquisa: Existe correlação cruzada de longo alcance de *mortos* com as demais variáveis de acidentes *peçoas, feridos, feridos leves, feridos graves, ilesos, ignorados e veículos* nas rodovias federais da Bahia entre 2007-2021?

A pesquisa é baseada nos bancos de dados fornecidos pela PRF, cujos resultados procuram fornecer subsídios para auxiliar a formulação de políticas públicas. Como consequência, a necessidade de compreender a interação das variáveis presentes nas séries temporais de acidentes trânsito nas rodovias fomentou essa investigação científica.

A maioria dos estudos que envolve a previsão de acidentes de trânsito é baseada em dados históricos relacionados a intervalos temporais regulares. Segundo [Commandeur et al. \(2013\)](#), algumas técnicas estatísticas usadas, a exemplo dos modelos de regressão tradicionais (lineares, lineares generalizados e não-lineares), implicam suposições que muitas vezes são violadas pelas propriedades especiais de dados de séries temporais, ou seja, dependência serial entre os distúrbios associados às observações. Nesse sentido, tais violações podem conduzir a uma sub ou superestimação do erro padrão e, conseqüentemente, podem produzir inferências errôneas.

Em [Commandeur et al. \(2013\)](#), o autores consideram ainda que, para lidar adequadamente com as dependências inerentes aos dados de séries temporais e colocar os resultados da análise em uma base estatística sólida, devem ser usadas técnicas de análise de séries

temporais. Os modelos ARMA (no caso de dados estacionários) e ARIMA (no caso mais geral de dados não estacionários, que é o caso da segurança viária) permitem descrever a dinâmica de processo em uma época e extrapolá-lo para o futuro, sem visitar as variáveis adicionais e com a única suposição de que o processo dinâmico permanecerá inalterado no horizonte previsto.

Existem modelos estatísticos de análise de séries temporais (AST) com a premissa das séries como objeto de estudo serem estacionárias. A característica de estacionariedade implica um tipo de equilíbrio estatístico ou estabilidade dos dados, isto é, uma série temporal é estacionária quando possui média e variância constantes (MONTEGOMERY, 2008). Nesse caso, é necessário transformar os dados originais, caso não se comportem como tal, por meio de procedimentos estatísticos. Entretanto, a maioria das séries cronológicas a serem modeladas apresenta alguma característica de não-estacionariedade (tendência/sazonalidade). Neste trabalho, propomos analisar séries cronológicas de acidentes de trânsito visando compreender, interpretar e prever tendências de segurança no trânsito.

O diferencial deste projeto está no emprego do *Coefficiente de Correlação Cruzada* ρ_{DCCA} proposto por Zebende (2011), para análise de séries temporais não estacionárias. Nessa perspectiva, pretendemos contribuir com o estudo dos acidentes de trânsito, fornecendo subsídios para a formulação de políticas públicas e/ou divulgação científica. Para tanto, tivemos como objetivo geral descrever e analisar o comportamento das variáveis envolvidas nos acidentes de trânsito nas rodovias federais da Bahia, em função do tempo e do espaço, com técnicas estatísticas, retratando o perfil das vítimas e os locais de maior densidade dos acidentes com morte.

Especificamente vamos:

- (i) Estimar o nível de correlação cruzada entre as variáveis de acidentes ao longo do tempo empregando o coeficiente de correlação cruzada (ρ_{DCCA});
- (ii) Identificar os acidentes com morte entre 2017 e 2020, nos locais de maior incidência por meio da visualização de dados espaciais a partir de um Sistema de Informação Geográfica (SIG);
- (iii) Descrever o perfil dos acidentes e o dos vitimados por meio da análise descritiva dos dados.

Com esta proposta, a dissertação foi organizada da seguinte forma, a contar deste capítulo:

✓ **Capítulo 1 - Introdução;**

- ✓ **Capítulo 2 - Fundamentação Teórica:** Aborda o embasamento teórico da pesquisa, destacando as ações voltadas para a mitigação do problema;
- ✓ **Capítulo 3 - Materiais e Método:** Expõe a metodologia empregada, com vistas a alcançar os objetivos propostos;
- ✓ **Capítulo 4 - Resultados e Conclusões:** Apresenta os resultados das análises obtidos e a conclusão.

Fundamentação Teórica

Neste capítulo abordaremos a relação do homem com o ambiente na concepção da psicologia. Em seguida, apresentaremos ações desenvolvidas nos âmbitos nacional e internacional voltadas para a segurança no trânsito, além das contribuições do uso de um SIG para o estudo de acidentes de trânsito.

2.1 *Acidentes de Trânsito*

O ser humano é um ser ambiental e a maneira como ele interage e explora o ambiente determina seu estilo de vida e as consequências de suas intervenções. A dependência do ambiente para a vida humana fez com que o homem buscasse transformá-lo a fim de solucionar seus problemas cotidianos. Nesse sentido, [Rozestraten \(2003\)](#), considera que

“É fato que com o homem o ambiente mudou. Enquanto os outros animais se satisfaziam com seus abrigos em cavernas, cavando suas pequenas tocas, construindo seus ninhos nas árvores, o homem começou a cultivar a terra, domesticar os animais, construir suas habitações. Com o homem, a terra começou a mudar, transformando o ambiente para servir às suas necessidades.”

De maneira sintetizada, o trânsito é um ambiente material modificado pelo homem com sinalizações e regulamentações para enfrentar condições adversas, a fim de garantir a segurança das vias e evitar acidentes. A via, por sua vez, se traduz como sendo um ambiente de trânsito, caracterizado por situações adversas onde facilmente pode ocorrer acidentes.

A partir da reforma sanitária ocorrida em diversos países inclusive no Brasil, os acidentes de trânsito passaram a ser percebidos como um aspecto de saúde pública, bem como os acidentes de trabalho, o uso de drogas, o comportamento sexual e outros aspectos que, direta e indiretamente afetam a saúde da população ([MAURO, 2001](#)).

A definição “acidente de trânsito” adotada pela OMS é aquela que se refere ao contexto,

“Um acidente de trânsito é qualquer acidente de veículo que ocorre na via pública [ou seja, originando, terminando ou envolvendo um veículo parcialmente na rodovia], sendo que via pública ou rua é toda a largura entre as

linhas de propriedade (ou outras linhas de limite) de um terreno aberto ao público por uma questão de direito ou costume para o propósito de mover pessoas ou propriedades de um lugar para outro (OMS, 2016).”

Em 2020, no Brasil, a ABNT publicou a revisão da NBR 10697/2018 que redefiniu os termos técnicos usados na preparação e execução de pesquisas relativas e na elaboração de relatórios estatísticos e operacionais a incidentes de trânsito. A norma corrige a expressão “acidente de trânsito”, substituída por “sinistro de trânsito”, e suprime o entendimento de “sinistro não premeditado” (ABNT, 2020).

Para a Associação Brasileira de Acidentes no Tráfego (ABRAMET), a requalificação dessa norma traduz-se como uma ação importante para o favorecimento da implantação de políticas voltadas à preservação da vida, visto que o termo “acidente” pode denotar algo imprevisível e incontrolável, que na verdade, na sua maioria, não são acidentes do acaso, mas sim provocados.

Nesse contexto, a ABNT defende sinistro de trânsito como sendo:

“Todo evento que resulte em dano ao veículo ou à sua carga e /ou lesões a pessoas e/ou animais, e que possa trazer dano material ou prejuízos ao trânsito, à via ou ao meio ambiente, em que pelo menos uma das partes está em movimento nas vias terrestres ou em áreas abertas ao público”.

O primeiro acidente de trânsito da história do mundo com vítima fatal ocorreu em agosto de 1869, no *Condado de Offaly*, Irlanda. O acidente foi provocado por um parente de *Mary Ward* que construiu um dos primeiros automóveis movidos a vapor do mundo. Certa tarde, para passear de carro o primo de *Ward* fez uma curva e a fez cair do veículo sob o volante. Ela morreu instantaneamente de pescoço quebrado (POST, 2016). No Brasil, o registro do primeiro acidente automobilístico foi datado de 1897, no Rio de Janeiro onde nem todas as ruas eram pavimentadas. O carro a vapor, um *Serpollet*, era guiado por Patrocínio que convidou seu amigo Olavo Bilac para dirigir o veículo mesmo sem nunca ter contato com um automóvel. Patrocínio deu algumas instruções ao amigo, que partiu de Botafogo rumo à Estrada Velha da Tijuca, no Alto da Boa Vista. O carro a vapor teria chegado a 4 km/h antes da primeira curva, onde Olavo Bilac perdeu o controle e bateu em uma árvore. Bilac e Patrocínio não se machucaram, mas o carro teve perda total (RODRIGUEZ, 2020).

Importante ressaltar que desde o surgimento do primeiro veículo motorizado, segurança no trânsito é objeto de estudo para diferentes situações. Nas últimas décadas, apesar dos esforços para a construção de vias seguras e bem sinalizadas, além da inovação de

itens de segurança pelos fabricantes de veículos e da elaboração de uma legislação que visa garantir a integridade da vida humana, cerca de 1,35 milhões de pessoas em todo o mundo perdem a vida a cada ano em decorrência de acidentes de trânsito (OPAS, 2019).

Empregamos nesse estudo a classificação “acidente de trânsito” apenas com o intuito de familiaridade com o referido termo. No entanto, consideramos sua definição com base na atualização da NBR 10697 definida pela ABNT em 2020, a qual a substitui pela terminologia “sinistro de trânsito”.

2.2 Iniciativas para promover segurança no tráfego

A segurança no trânsito é uma preocupação que atinge praticamente todos os países do mundo. Segundo Marín e Queiroz (2000), essa questão só passou a ser examinada no mundo com a preocupação devida a partir da década de 1950. Entretanto, somente a partir de 1974, a Assembleia Mundial da Saúde da ONU adotou a Resolução WHA 27.59, declarando os acidentes rodoviários como um importante problema de saúde pública. Diante dessa problemática global, em setembro de 2000, 191 nações firmaram um compromisso para combater a extrema pobreza e outros males da sociedade. Esse compromisso se concretizou nos 8 Objetivos de Desenvolvimento do Milênio (ODM), que deveriam ser alcançados até 2015.

A partir de março de 2003 a OMS designou o tema “Estradas seguras” como tema do Dia Mundial da Saúde de 2004, além de adotar diversas resoluções em prol das ações: conscientização mundial sobre o problema generalizado de mortes e lesões evitáveis no trânsito inclusive aos jovens nos estabelecimentos de ensino; comprometimento dos governos para o cumprimento das Leis de Trânsito; relatório sobre a crise global de segurança no trânsito expressando opinião de Estados-membros a serem consideradas pela Assembleia Geral em sessão posterior à publicação do documento.

Ainda em novembro de 2003, a OMS adotou a resolução intitulada *Global Road Safety Crisis*, cujo texto reconhecia a importância de fortalecer os esforços de países em desenvolvimento, com recursos financeiros e suporte técnico em favor de capacidades no campo da segurança no trânsito.

No ano seguinte, em abril de 2004, a OMS em parceria com o Banco Mundial lançam o Relatório Mundial sobre Tráfego Rodoviário Prevenção de Lesões, em comemoração ao Dia Mundial da Saúde, com o tema “Segurança viária não é acidente”. O documento contém uma série de recomendações, prestando atenção especial aos cinco fatores de risco identificados, nomeadamente: a não utilização de cintos de segurança e de sistemas de retenção para crianças; álcool; o não uso de capacetes; velocidade inadequada e excessiva

e a falta de infraestrutura.

Em dezembro de 2005, por meio da Resolução 60/5, a ONU designou o terceiro domingo de novembro como o Dia Mundial da Memória das Vítimas do Trânsito, em reconhecimento à perda e sofrimento das vítimas do trânsito e de suas famílias e destaca preocupação com o aumento contínuo, em particular em países em desenvolvimento, de fatalidades e lesões no trânsito em todo o mundo. Dentre outras ações, o documento também incentiva os Estados-membros a se esforçarem para reduzir as lesões causadas pelo trânsito e a mortalidade, a fim alcançar os ODM.

Uma iniciativa importante foi realizada em 2009 quando a Assembleia Geral da ONU convocou o mundo por meio de seus 192 países membros a reduzirem em 50 por cento o número de vítimas fatais, realizando em Moscou a Primeira Conferência Ministerial Global sobre Segurança Viária. Nesse mesmo ano, foram registradas campanhas e eventos em defesa da segurança no trânsito realizados por diversas instituições em vários países da Europa, a exemplo da Federação Internacional de Basquetebol (FIBA) durante o Campeonato EuroBaskete 2009. Ainda em 2009, o Brasil implementou uma alteração do Código de Trânsito Brasileiro (CTB) por meio da lei nº 11.910/2009, determinando a carros novos o uso de equipamento suplementar de retenção: *air bag* frontal para o condutor e para o passageiro do banco dianteiro, o que começou a valer de fato somente a partir do ano de 2014, quando o percentual de carros novos que deveriam ter esses itens aumentou gradativamente até chegar a 100 por cento nesse ano.

Um ano depois, em 2010, a ONU aprovou a Resolução 64/255, instituindo a “Década de Ação pela Segurança no Trânsito” como uma maneira de deter e reduzir globalmente os níveis previstos das mortes por acidentes de trânsito. Em setembro de 2010, o mundo renovou o compromisso dos ODM para acelerar o progresso em direção ao cumprimento desses objetivos (PNUD, 2016).

Durante a 51ª Reunião do Conselho Diretor, em 2011, a Organização Pan-Americana da Saúde (OPAS) adotou o “Plano de Ação para a Segurança no Trânsito”, cujos objetivos foram baseados na situação Região e, aderindo à Década da Ação, estipulou as diretrizes para os Estados Membros à medida que estes se aproximavam da meta de evitar e controlar as mortes no trânsito. Como medida de segurança, em âmbito nacional, o CONTRAN (Conselho Nacional de Trânsito) publicou a Resolução nº 380, de 28/04/2011 estabelecendo como obrigatória a utilização do sistema antitravamento de rodas - ABS (*Anti-lock Braking System*) em veículos nacionais e importados, sob o argumento de que a instalação desse sistema de freios melhora a estabilidade e a dirigibilidade do veículo durante o processo de frenagem permitindo ao condutor manter o controle do veículo principalmente em pista escorregadia com possibilidade de evitar acidentes causados pelo travamento das rodas.

Em maio de 2013, durante a Semana Global de Segurança Viária das Nações Unidas, em Seul, foi adotada uma declaração conjunta sobre a segurança rodoviária na Ásia e no Pacífico, e a assistência técnica fornecida aos países membros para desenvolver e refinar as metas e objetivos nacionais de segurança no trânsito e indicadores em apoio à Década de Ação.

Entendendo a urgência para consolidação das metas estabelecidas anteriormente em um acordo global, em setembro de 2015, representantes dos 193 Estados-membros da ONU adotaram a Agenda 2030 para o Desenvolvimento Sustentável, que inclui os 17 Objetivos do Desenvolvimento Sustentável (ODS), substituindo os ODM estabelecidos em 2000 após a Declaração do Milênio das Nações Unidas. A segurança viária que não estava incluída na agenda dos ODM, agora faz parte da nova Agenda de 2030. A Meta 3.6 do Objetivo 3 (ODS3) visa até 2030, reduzir pela metade as mortes e lesões por acidentes no trânsito. Paralelamente, a Meta 11.2 (ODS11) procura proporcionar o acesso a um sistema sustentável de transporte para todos, expandir o transporte público e melhorar a segurança no trânsito até 2030 (OPAS, 2016).

A elaboração das metas globais dos ODS inclui 70 países. Iniciativas como essa têm contribuído para reduzir o número de acidentes por meio de intervenções nas cidades. As áreas urbanas são regiões onde abrigam mais da metade da poluição mundial e funcionam como vitrine para soluções promovidas pelos governos. A exemplo disso é possível citar: Accra (Gana) está usando dados para melhorar a segurança de pedestres, e *Ho Chi Minh City* (também conhecida como Saigon, Vietnã) tem realizado medidas econômicas para tornar mais de 150 cruzamentos de alto risco mais seguros, além de cidades em países como o Brasil e a Índia que reduziram os acidentes de trânsito por meio de campanhas na mídia e fiscalização mais forte em cidades que implementaram a "Lei Seca", que prevê punição para o condutor que dirigir embriagado (WHO, 2018). Apesar disso, para que ações e medidas de segurança sejam amplamente efetivas minimizando o impacto do número de mortes no trânsito, é essencial que os governos priorizem a segurança nas estradas, além de conhecer a magnitude do problema e dos riscos causados por ele. Em outubro de 2021, a OMS deu início à "Década de Ação pela Segurança no Trânsito 2021-2030", com a meta de prevenir ao menos 50 por cento das mortes e lesões no trânsito até 2030.

No âmbito nacional, diversas ações voltadas para a segurança no trânsito têm contribuído com essas iniciativas. No Brasil essa era foi inaugurada por campanhas em nível federal, estadual e municipal e, sobretudo a partir de 1998 com implantação do Código de Trânsito Brasileiro (CTB). Anos depois alguns projetos foram desenvolvidos com o intuito de mitigar a violência nas estradas.

O projeto *Vida no Trânsito* é a denominação, no Brasil, do *Road Safety in Ten Countries*, voltado à redução das mortes e lesões causadas no trânsito em 10 países, com o financi-

amento da Fundação Bloomberg e a coordenação global da OMS e suas agências regionais. No Brasil, o Projeto é desenvolvido desde 2010 em cinco cidades: Belo Horizonte; Campo Grande; Curitiba; Palmas e Teresina e conta, além do suporte da Organização Pan-americana de Saúde (OPAS/OMS no Brasil), com o aporte técnico e financeiro do Governo Federal ([WHO, 2021](#)).

O *Movimento Maio Amarelo* nasceu com a proposta de chamar a atenção da sociedade para o alto índice de mortes e feridos no trânsito em todo o mundo. O movimento consiste numa campanha de conscientização sobre segurança no trânsito, cuja primeira ação ocorreu em 2014. O objetivo do movimento é uma ação coordenada entre o Poder Público e a sociedade civil. A intenção é colocar em pauta o tema segurança viária e mobilizar toda a sociedade, envolvendo os mais diversos segmentos: órgãos de governos, empresas, entidades de classe, associações, federações e sociedade civil organizada para, fugindo das falácias cotidianas e costumeiras, efetivamente discutir o tema, engajar-se em ações e propagar o conhecimento, abordando toda a amplitude que a questão do trânsito exige, nas mais diferentes esferas ([ONSV, 2021](#)).

A Figura [2.1](#) apresenta uma síntese cronológica das principais ações de promoção de segurança para os trânsito.

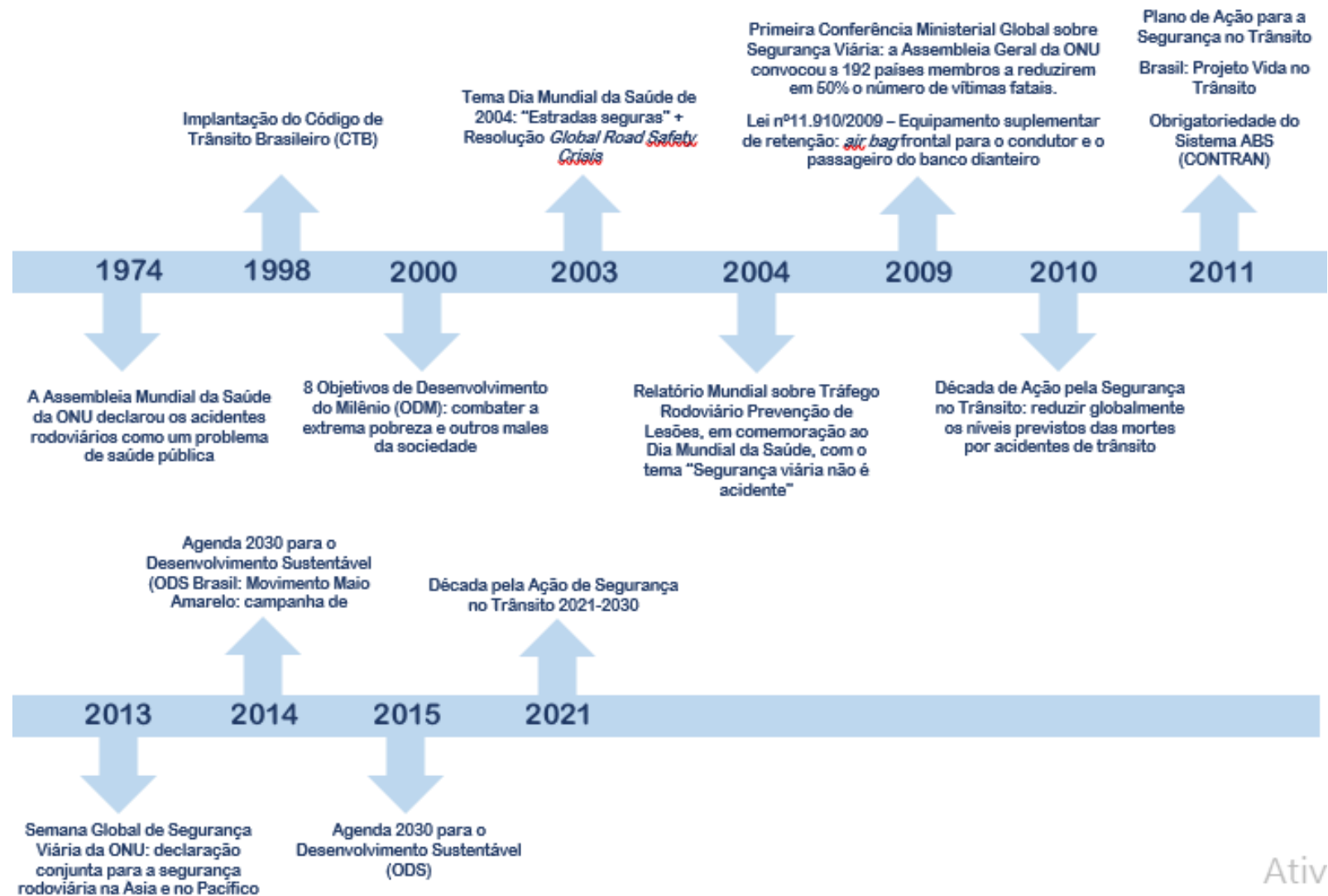


Figura 2.1: Mapa conceitual das ações de segurança para o trânsito.

2.3 Análise estatística de acidentes de trânsito

2.3.1 Trabalhos correlatos

O trânsito é considerado um sistema dinâmico complexo com grande número de agentes interagindo entre eles e com o meio externo (BAR-YAM, 2002). Uma das características importantes na dinâmica do trânsito é a presença de correlação cruzada entre as variáveis envolvidas nos acidentes. Para investigar a ocorrência dos acidentes de trânsito sob diferentes aspectos, quer seja utilizando análise descritiva ou técnicas estatísticas mais elaboradas, a análise estatística está presente na maioria dos estudos sobre o tema. A análise de séries temporais se aplica a diversos campos da ciência, a exemplo de finanças, economia, epidemiologia, demografia, meteorologia, acidentes de trânsito, dentre outros.

Em estudos anteriores, diversas abordagens metodológicas foram propostas para o estudo de acidentes de trânsito. Ivajnsič et al. (2021a) empregaram regressão geograficamente ponderada para a análise de dados históricos de segurança no tráfego e modelos de árvore de regressão ajustados, a fim de identificar um padrão espacial dos pontos críticos de acidentes de trânsito na Eslovênia. Na África do Sul, Sukhai et al. (2011) estudaram modelos de regressão exógena de auto-regressão não linear (NARX) ajustados para explicar as variações nas fatalidades do trânsito e para avaliar o grau em que essas variações entre as nove províncias do país estavam associadas aos fatores de risco. Bahadorimonfared et al. (2013) aplicaram o modelo de regressão de Poisson para estimar as mudanças na frequência de eventos ao longo do tempo para fatores de risco conhecidos e associados, usando do modelo SARIMA para ajustar à série temporal de índice de mortalidade em acidentes de trânsito no Irã. A respeito de estudos relacionados a acidentes registrados no estado da Bahia, podemos destacar o trabalho de Rios et al. (2020), que tiveram como objetivo identificar fatores associados ao envolvimento em acidentes de trânsito (AT) entre condutores de veículos, na cidade de Jequié, Bahia por meio de um modelo conceitual hierárquico e de regressão logística multinível para as estimativas de Razão de Chances (RC).

Em outros trabalhos foram utilizados testes estatísticos. Na Finlândia, Malin, Silla e Mladenovic (2020) examinaram a prevalência de mortes de pedestres e lesões graves no trânsito, por meio da investigação dos principais fatores e as diferenças entre mortes de pedestres e lesões graves onde a outra parte era um veículo motorizado aplicando um teste de independência χ^2 . Feng, Wang e Quddus (2020) conduziram um estudo usando técnicas de séries temporais multivariadas para analisar as relações contemporâneas e interações dinâmicas entre violações policiais e acidentes de trânsito em Xangai, China. O estudo de Barros et al. (2003), mostrou a aplicação do teste qui-quadrado de Pearson para avaliar a associação entre os tipos de acidente e comparar os

riscos de lesão e morte de diferentes tipos de veículos e pedestres no Brasil.

Modelos autorregressivos também foram aplicados em vários estudos. [Li, Wu e Yang \(2018\)](#) estudaram a relação de longo e curto prazo entre acidentes de trânsito em Hong Kong e variáveis socioeconômicas por um modelo autorregressivo de retardo distribuído (ARDL) e um modelo de correção de erro vetorial (VECM). [Ramstedt \(2008\)](#) estimou a associação entre o consumo de álcool *per capita* e acidentes fatais nos Estados Unidos usando a técnica de Box-Jenkins para análise de séries temporais. [Quddus \(2008\)](#) apresentou a classe de modelos autoregressivos de valor inteiro (INAR) para analisar séries temporais de acidentes de trânsito na Grã-Bretanha. [Lobato \(2018\)](#) analisou a evolução dos acidentes nas rodovias federais do Brasil concedidas pela união utilizando o teste de *Wilcoxon* e dos Mínimos Quadrados Ordinários (MQO). Sobre a influência dos efeitos acidentes na vida das pessoas, [Ferreira \(2016\)](#) examinou os efeitos dos acidentes de trânsito sobre a qualidade de vida e a funcionalidade dos indivíduos afetados tendo como base análise de variância (ANOVA), teste de qui-quadrado de Pearson, teste Exato de Fisher, estatística Gamma de Goodman e Kruskal, teste t, modelos de regressão logística e de regressão logística ordinal.

Algumas pesquisas foram conduzidas com base em Cadeias de Markov.

[Malyshkina e Mannering \(2009\)](#) propõem modelos *logit* multinomiais de comutação Markov para modelagem estatística de lesões graves por acidente de trânsito. [Xiong, Tobias e Mannering \(2014\)](#) aplicaram Markov comutando parâmetros aleatórios ordenados *probit* aplicado à análise de dados de gravidade de lesões em acidentes de veículo. [Oliveira e Achcar \(2020\)](#) sugeriram o uso de um modelo estatístico composicional sob uma abordagem Bayesiana usando métodos de simulação de *Markov Chain Monte Carlo* aplicados a vítimas de tráfego rodoviário ocorrendo em âmbito federal estradas do Brasil em um determinado período de tempo.

[Sebego et al. \(2014\)](#) examinaram o impacto da implementação de políticas relacionadas ao álcool e à segurança no trânsito nas taxas de acidentes em Botswana, a partir de uma análise de série temporal interrompida. [Fei et al. \(2020\)](#) realizaram um estudo semelhante, ao avaliar a eficácia da lei de dirigir embriagado na China na redução de acidentes de trânsito, lesões e mortalidade. [Minervi \(2013\)](#) investigou o mecanismo gerador da série de acidentes ocorridos entre janeiro de 2007 e julho de 2011 nas rodovias federais brasileiras para explicar a variabilidade dos dados através do uso de Análise Fatorial, técnica de análise multivariada, e a aplicação dos modelos Autorregressivos, Integrados e de Médias Móveis (ARIMA). [Andrade e Antunes \(2020\)](#) calcularam a variável percentual mensal do números de vítimas e o número de infrações devidos a distração ao volante, com base no método *Prais-Winsten* na análise de tendência, realizando uma correção da correlação residual de primeira ordem. Eles também analisaram a tendência do número de mortos, feridos graves e feridos leves por acidentes de trânsito nas rodovias federais brasileiras,

segundo macrorregião, antes e depois do início da Década de Ação pela Segurança no Trânsito (DAST) por meio da aplicação do método desse mesmo método para o cálculo da variação percentual mensal (VPM) do número de mortos, feridos graves e feridos leves (ANDRADE; ANTUNES, 2019).

Martins (2020) desenvolveu um Sistema de Apoio à Decisão (SAD) para auxiliar os interessados no ordenamento em criticidade de acidentes de trânsito dos trechos de rodovias federais a partir abordagens multicritério e multiperíodo através do método MUPOM (Multi-criteria multi-Period Outranking Method) para considerar diferentes indicadores de segurança desenvolvidos no estudo. Oliveira et al. (2019) avaliaram o impacto que a concessão de rodovias federais trouxe na redução de acidentes de trânsito, feridos leves, feridos graves e mortos nas rodovias federais do Distrito Federal e seu entorno por meio do cálculo das Diferenças em Diferenças, com a utilização de dados dos anos de 2010 a 2016.

Estatísticas descritivas também foram conduzidas em diversos trabalhos. Ferreira (2017) apresenta um estudo sobre o desenvolvimento de uma metodologia de análise e previsão de quantidades e severidade dos acidentes de trânsito para o Brasil utilizando um tratamento matemático dos fatores relacionados, no caso os Volumes Diários Médios Anual (VDMA) por faixa e extensão urbana e/ou rural dos trechos, critérios de parametrização da severidade dos acidentes e coeficiente de correlação de Pearson. Andrade e Mello-Jorge (2016) estimaram os anos potenciais de vida perdidos por acidentes de transporte terrestre e o perfil dessas vítimas após três anos do início da Década de Ação pela Segurança no Trânsito. O estudo mostrou que a Região Centro-Oeste apresentou a taxa mais elevada. A maioria dos óbitos por acidentes de transporte terrestre foi observada no sexo masculino. Mais da metade das pessoas que faleceram em decorrência de acidentes de transporte terrestre eram da raça/cor da pele negra, adultos jovens indivíduos com baixa escolaridade e motociclistas. A taxa de mortalidade, no triênio 2011 a 2013 apresentou redução, mas aumentou entre os motociclistas. Em todo o País, mais de um milhão de anos potenciais de vida foram perdidos em 2013, devido aos acidentes de transporte terrestre, especialmente na faixa etária de 20 a 29 anos.

Rocha e Nassi (2012) aplicaram modelagem estatística e informações georreferenciadas em ambiente de Sistema de Informações Geográficas (SIG) para compreender a distribuição espacial de acidentes, gerar novas variáveis explicativas e testá-las por meio de modelos de regressão multivariada, modelos de regressão *Spatial Autoregressive* (SAR) e modelos lineares generalizados.

A partir de uma revisão de literatura Bacchieri e Barros (2011) descreveram a situação dos acidentes de trânsito no Brasil, desde a implementação do Código de Trânsito Brasileiro de 1998 a 2010. O estudo revelou que os principais problemas identificados foram aumento do

número absoluto de mortos e das taxas de mortalidade, ampliação da frota de motocicletas e o uso de álcool.

Ribeiro et al. (2020) apresentaram uma análise descritiva dos dados referentes às principais causas dos acidentes automobilísticos nas rodovias federais da Bahia, com base no relatório produzido pela Polícia Rodoviária Federal no período de 2014 a 2017. O estudo de Bastos, Oliveira e Barreto (2016) teve como objetivo informar os dados de mortalidade na Bahia, no período de 2008 a 2012, analisando os números de mortes causadas por acidentes de trânsito, segmentados em ano, faixa etária, gênero, escolaridade e macrorregião de saúde. WEST (2016) procurou caracterizar de forma descritiva os acidentes de trânsito ocorridos em dez meses do ano de 2014 na região oeste da Bahia a partir de dados documentais fornecidos pelo Batalhão da Polícia Militar da cidade de Barreiras. Dentre outros resultados, os autores concluíram que existe uma real necessidade dos serviços públicos e demais atores da área desenvolver ações que reduzam os acidentes de trânsito e melhorem a segurança viária. Também com base em uma revisão de literatura, Rios (2010) abordou a mortalidade por acidentes de trânsito a partir do estudo da série temporal de casos compreendidos entre 1996 a 2007, em diferentes regiões do estado da Bahia, mostrou que há a real necessidade dos serviços públicos e demais atores da área desenvolver ações que reduzam os acidentes de trânsito e melhorem a segurança viária na Bahia. Carvalho (2005) utilizou variáveis sociodemográficas para desenvolver um estudo transversal a partir dos registros de acidentes de trânsito envolvendo adolescentes e taxas de incidência e mortalidade em Feira de Santana, Bahia em 2001.

Importante destacar que a análise estatística está presente em quase todos os estudos que tratam de acidentes de trânsito mencionados nesse estudo. Entretanto, poucos trabalhos estudaram a existência de dependência de memória de longo alcance. A memória longa descreve a “persistência” das séries temporais. Assim, se uma série exibe memória longa, a dependência temporal entre as observações distantes é estatisticamente significativa e persistente (VLAHOIANNI; KARLAFTIS, 2011).

Dentre os trabalhos que abordam a dependência de longo alcance em dados de acidentes, podemos citar o trabalho de Jianjun e Huijun (2014). O estudo mostra a aplicação do método de análise de flutuação destendenciada (DFA) para investigar a complexidade da série de dados no tempo para fluxo de tráfego misto com bloqueio induzido por acidente. Em um outro estudo, Remy (2011) monitorou a auto-semelhança estatística de canais fisiológicos, ambientais e veiculares de dados, por meio da aplicação de Análise de Flutuação Destendenciada (DFA).

A análise de séries temporais e a previsão de valores futuros estão entre os problemas mais importantes que os analistas enfrentam em diversas áreas do conhecimento, desde finanças e economia, a gestão das operações de produção, até a análise de políticas públicas, isto é,

sempre que fenômenos podem ser ordenados no tempo ou no espaço ([MONTEGOMERY, 2008](#)), ([MORETIN; TOLOI, 2008](#)) para auxiliar na tomada de decisão.

O conhecimento da correlação a longo prazo entre as variáveis envolvidas é de grande relevância em diversos aspectos. O principal deles tem a ver com a saúde pública, sobretudo com a preservação da vida. Compreender a distribuição espacial de acidentes, conhecer o perfil dos vitimados, gerar novas variáveis explicativas podem auxiliar no comportamento do condutor no sentido de adotar uma conduta preventiva criando uma cultura de segurança nas estradas ([MOLAN; MOLAN, 2021](#)), além de contribuir para a elaboração de políticas públicas e ações de mitigação da violência no tráfego com vítimas fatais.

Materiais e Método

Neste capítulo apresentaremos a área de estudo que abrange a pesquisa, as séries temporais e a metodologia utilizada para alcançar os resultados.

3.1 Base de dados

A análise dos resultados deste estudo se baseia em estatísticas relacionadas aos acidentes de trânsito ocorridos nas rodovias federais da Bahia disponíveis no portal virtual da PRF no endereço:

<https://www.gov.br/prf/pt-br/aceso-a-informacao/dados-abertos/dados-abertos-acidentes>

Os dados disponíveis no referido *link* representam o registro dos acidentes ocorridos nos municípios brasileiros, reportados diariamente pela PRF durante atendimento presencial para acidentes considerados relevantes que são caracterizados por, pelo menos, uma das seguintes situações: vítima (lesionada ou morta); danos a bens públicos não concedidos à iniciativa privada (como veículos oficiais, sinalizações, mobiliário, entre outros); danos ao meio ambiente. Em casos como esses a PRF registra o acidente no Boletim de Acidente de Trânsito (BAT). Para os acidentes que não se enquadram nessas características, o envolvido no evento pode registrar a ocorrência pela internet por meio de preenchimento da Declaração de Acidente de Trânsito (DAT).

O banco de dados foi agrupado a partir de quatro arquivos, sendo dois que foram extraídos do referido portal virtual da PRF, *Agrupados por ocorrência* (para analisar a flutuação das séries de acidentes); *Agrupados por pessoa* (para identificar o perfil das vítimas), e outros dois, *Malha dos municípios* e *Malha das rodovias federais* (ambos para localizar a incidência dos acidentes), extraídos das páginas do IBGE (Instituto Brasileiro de Geografia e Estatística) e do Ministério da Infraestrutura, respectivamente. A análise das séries de acidentes, principal objetivo deste trabalho, foi elaborada a partir das oito variáveis, ver Figura 3.1:



Figura 3.1: Variáveis dos acidentes.

representando as seguintes quantidades:

- **peessoas**: Total de pessoas envolvidas na ocorrência;
- **mortos**: Total de pessoas mortas envolvidas na ocorrência;
- **feridos**: Total de pessoas feridas envolvidas na ocorrência (é a soma dos feridos leves com os graves);
- **feridos leves**: Total de pessoas com ferimentos leves envolvidas na ocorrência;
- **feridos graves**: Total de pessoas com ferimentos graves envolvidas na ocorrência;
- **ilesos**: Total de pessoas ilesas envolvidas na ocorrência;
- **ignorados**: Total de pessoas envolvidas na ocorrência que não se soube o estado físico;
- **veículos**: Total de veículos envolvidos na ocorrência.

Desse arquivo, construímos uma base de dados (Figura 3.1), sendo que as informações georreferenciadas foram registradas somente a partir do ano de 2017, o que nos limitou a estudá-las apenas para o período 2017-2021. Para analisar a correlação entre as séries (correlação cruzada), o dia 29/02 foi suprimido de todos os anos bissextos, a fim de torná-las de mesmo tamanho, requisito necessário para o cálculo do coeficiente de correlação cruzada (ρ_{DCCA}). A análise do arquivo “Agrupados por pessoa” viabilizou caracterizar os acidentes de um ano para o outro, segundo estado físico das vítimas envolvidas, gênero e idade do condutor. A partir das informações georreferenciadas (latitude e longitude)

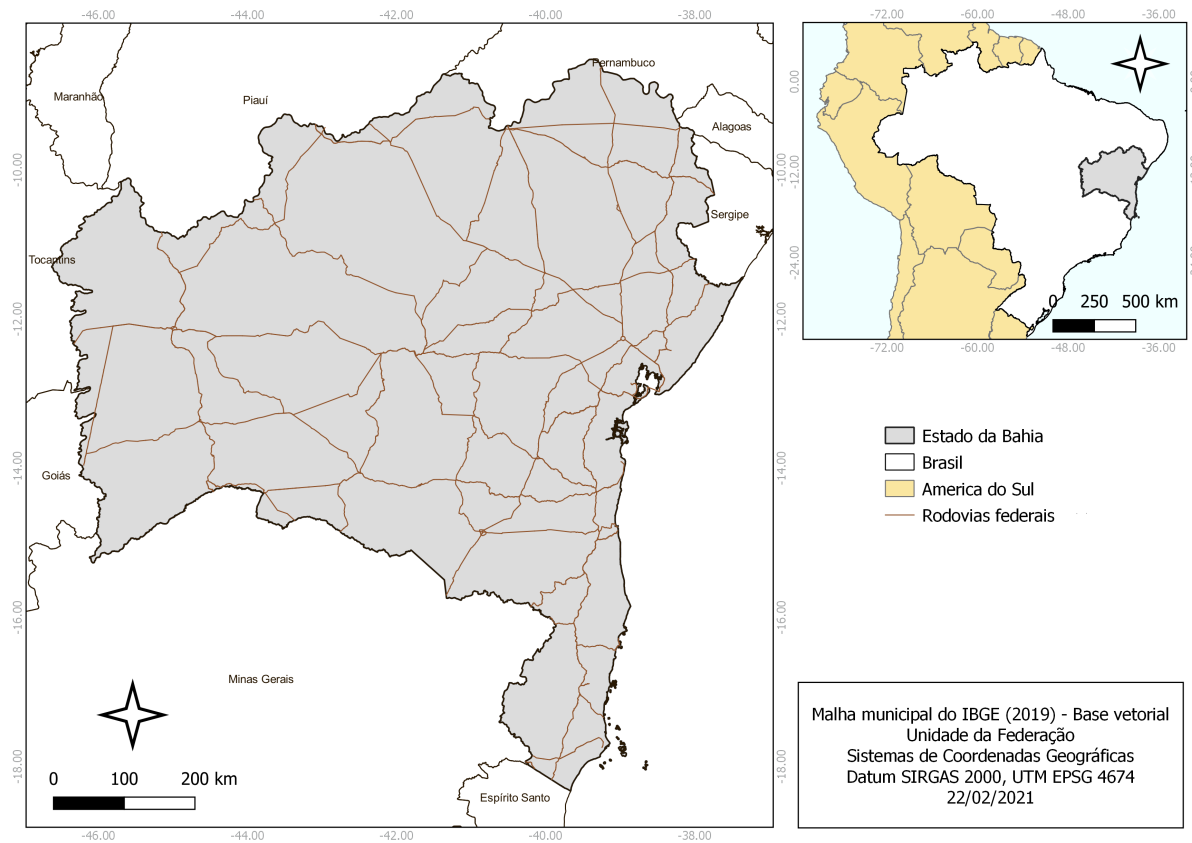
associadas a um SIG, foram elaborados mapas para identificar a incidência local dos acidentes.

	B	D	S	T	U	V	W	X	Y	Z	AA	AB
1	data_inversa	horario	pessoas	mortos	feridos_leves	feridos_graves	ileso	ignorados	feridos	veiculos	latitude	longitude
548	28/02/2020	20:00:00	1	1	0	0	0	0	0	1	-12,23	-43,83
549	28/02/2020	10:15:00	4	0	1	1	2	0	2	3	-12,46	-38,93
550	01/03/2020	14:15:00	2	0	0	0	2	0	0	1	-17,56	-39,73
551	01/03/2020	11:20:00	1	0	0	1	0	0	1	1	-16,01	-39,60
552	01/03/2020	13:30:00	2	0	1	0	1	0	1	2	-12,16	-38,41
553	01/03/2020	15:30:00	2	0	0	1	0	1	1	1	-11,59	-38,99
554	01/03/2020	18:20:00	2	0	0	1	0	1	1	1	-12,25	-38,99
555	01/03/2020	18:10:00	2	0	2	0	0	0	2	1	-13,83	-40,11
556	01/03/2020	18:30:00	4	0	3	0	1	0	3	2	-13,85	-46,18
557	01/03/2020	18:45:00	2	2	0	0	0	0	0	1	-14,96	-39,31
558	01/03/2020	17:30:00	1	0	1	0	0	0	1	1	-14,14	-39,35

Figura 3.2: Exemplo de planilha com a base de dados. Note que de 2017 em diante temos a localização georreferenciada das ocorrências.

3.2 Área de estudo

O Estado da Bahia está localizado na Região Nordeste do Brasil (Figura 3.2). Ele ocupa a quarta posição dentre os estados brasileiros com maior população do país, são aproximadamente 14,9 milhões de habitantes (IBGE, 2020). Segundo os dados mais recentes publicados pela Superintendência de Estudos Econômicos e Sociais da Bahia (SEI), no que tange à posição relativa do produto interno bruto (PIB) entre 2002-2018, a Bahia está em sétimo lugar dentre os estados brasileiros com maior PIB do país, sendo a UF com maior PIB da Região Nordeste.



Fonte: Elaborado pelo autor.

Figura 3.3: Localização da área de estudo com aplicação às rodovias federais.

A malha rodoviária federal do estado é composta por vinte e cinco rodovias federais ou BRs (020, 030, 101, 110, 116, 122, 135, 242, 251, 234, 330, 342, 349, 367, 407, 410, 415, 418, 420, 423, 430, 439, 489 e 498). Como característica viária, a Bahia concentra duas rodovias importantes que compõem um dos principais eixos rodoviários do país, a BR 116 e a BR 101, estradas com o maior volume de vítimas fatais em acidentes registrados no ano de 2021.

3.3 Metodologia

3.3.1 Análise de séries temporais

Uma série temporal (ST) é uma coleção de observações feitas sequencialmente ao longo do tempo, podendo também a variável *tempo* ser substituída por uma outra, como espaço, profundidade, etc. (EHLERS, 2007). Sua representação é dada por $[X(t), t \in T]$, de uma variável X , sendo T o total de valores.

São exemplos de séries temporais:

- (i) Quantidade diária de acidentes com vítimas (mortos ou feridos) registrados nas rodovias federais da Bahia;
- (ii) Taxa anual de mortalidade por acidentes de trânsito no Brasil;
- (iii) Evolução mensal dos acidentes por tipo (com vítimas e sem vítimas);
- (iv) Custo anual com acidentes de trânsito nas rodovias federais;
- (v) Participação anual das rodovias federais nos acidentes com vitimas fatais.

Aqui seguem alguns exemplos gráficos de séries temporais de acidentes, em que o eixos horizontais representam a escala temporal de observação e os eixos verticais seu referido total.

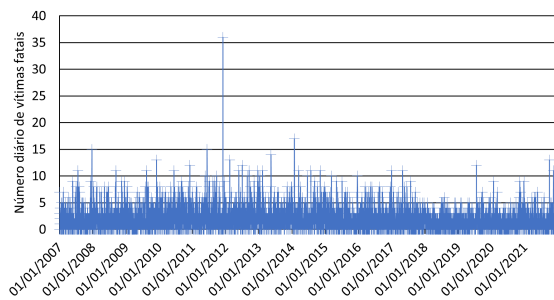


Figura 3.4: Totais diários de vítimas fatais em acidentes nas rodovias federais da Bahia, 2007-2021. Fonte: PRF

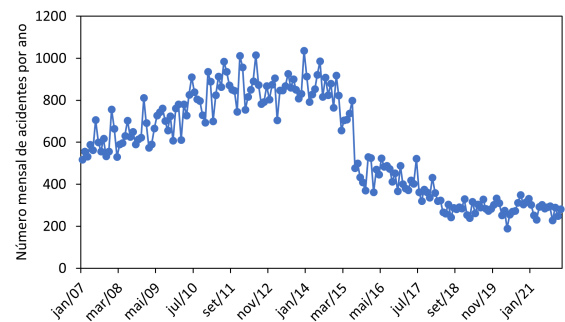


Figura 3.5: Totais mensais de acidentes nas rodovias federais da Bahia, 2007-2021. Fonte: PRF

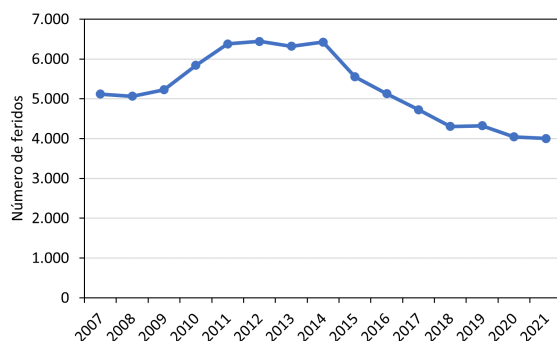


Figura 3.6: Totais por ano de feridos em acidentes nas rodovias federais da Bahia, 2007-2021. Fonte: PRF

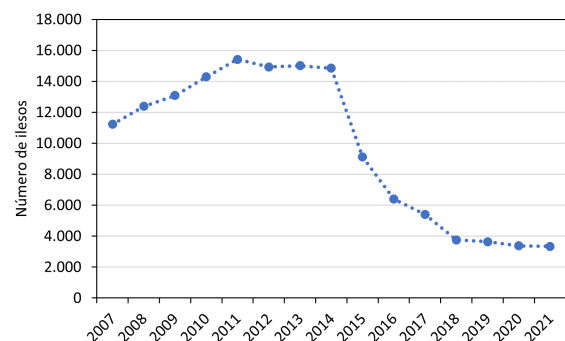


Figura 3.7: Totais por ano de ilesos em acidentes nas rodovias federais da Bahia, 2007-2021. Fonte: PRF

Alguns objetivos da análise de séries temporais podem ser destacados, como por exemplo: investigar o mecanismo gerador da série; fazer previsões de valores futuros; apenas descrever o comportamento da série ou ainda procurar prioridades relevantes nos dados. De modo geral, procura-se construir modelos probabilísticos (quando os documentos são pontuados por sua probabilidade estimada de relevância) ou modelos estocásticos (uma família de variáveis aleatórias representando a evolução de um sistema de valores no tempo) a partir dessas análises.

Os gráficos de uma série temporal, além de caracterizar a distribuição dos dados, podem também revelar padrões como aleatoriedade, tendência, sazonalidade e períodos ou ciclos, observações incomuns ou uma combinação de padrões (MONTGOMERY, 2008). Nesse sentido, para Ehlers (2007), o gráfico temporal deve ser o primeiro passo que antecede qualquer análise. Assim, uma série temporal X_t pode ser representada como uma *soma* ou um *produto* de pelo menos três componentes: Tendência (T_t), Sazonalidade (S_t) e Resíduos (γ_t), e a maneira clássica de decomposição baseia-se em, pelos menos, dois modelos (GUEDES, 2014), o Aditivo e o Multiplicativo, conforme as equações 3.1 e 3.2, respectivamente.

$$X_t = T_t + C_t + \gamma_t \quad (3.1)$$

$$X_t = T_t \cdot C_t \cdot \gamma_t \quad (3.2)$$

em que T_t é a componente de tendência, C_t é uma componente cíclica ou sazonal e γ_t é uma componente aleatória ou ruído (a parte não explicada, que espera-se ser puramente aleatória) (EHLERS, 2007). Segundo Guedes (2014), enquanto o modelo aditivo é usado nas séries de dados cujas taxas de variação são constantes, o multiplicativo é útil naquelas em que as taxas são crescentes ou decrescentes. De acordo com Milone (2006), enquanto o modelo aditivo pressupõe componentes não correlacionadas, o multiplicativo, é usado quando as variáveis são correlacionadas atuando de modo proporcional às respectivas forças.

Ao analisar uma série temporal primeiramente é preciso caracterizar o comportamento gráfico dos dados por meio de técnicas descritivas. O teste de aderência de Jarque-Bera (JB) para verificar normalidade de amostras é um critério objetivo que possibilita analisar se uma série de dados segue ou não uma distribuição normal. Quanto mais próximo for do zero a estatística JB, maior a probabilidade de a distribuição ser normal. Este teste foi proposto por Jarque e Bera (JARQUE; BERA, 1980) e se baseia na diferença entre os coeficientes de assimetria e curtose dos dados x_1, x_2, \dots, x_n e àqueles da distribuição

assumida Normal (LUCAMBIO, 2008). O resultado desse teste pode ser representado pela estatística JB (3.3);

$$JB = n \left(\frac{\alpha_3^2}{6} + \frac{(\alpha_4 - 3)^2}{24} \right) \sim \chi_{(2)}^2, \quad (3.3)$$

em que

$$\alpha_3 = \frac{\sum_{i=1}^n (y_i - \bar{y})^3}{ns^3}, \quad (3.4)$$

$$\alpha_4 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{ns^4}, \quad (3.5)$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}. \quad (3.6)$$

Sendo n o número de observações, \bar{x} a média das observações e s^2 , α_3 e α_4 o segundo, terceiro e quarto momentos centrais, respectivamente. Assim, no teste de normalidade JB , testam-se as hipóteses H_0 : $X(t)$ segue distribuição Normal versus H_1 : $X(t)$ não é uma distribuição Normal. Logo, rejeita-se a hipótese nula sempre que a estatística JB extrapolar valores extremos da distribuição qui-quadrado com 2 graus de liberdade, a um certo nível de significância. Apesar do teste de normalidade Shapiro-Wilk ser um dos testes de aderência mais usados, o RStudio possui a limitação de analisar esse teste para, no máximo, 5.000 observações, tornando um empecilho para empregarmos o referido teste neste trabalho fazendo com que optássemos para a estatística JB .

Um outro aspecto relevante no estudo de ST é identificar se a série é estacionária, isto é, se é caracterizada por uma variável que se comporta de forma aleatória ao longo do tempo em torno de uma média constante. Uma ST é considerada estritamente estacionária se suas propriedades não forem afetadas por uma mudança na origem do tempo (MORETIN; TOLOI, 2008). Isto significa que os dados variam em torno de uma média constante, independente do tempo, sendo que a variância das flutuações permanece essencialmente a mesma. De maneira resumida, o teste Dickey-Fuller (DF) possibilita verificar se o processo é estacionário ou não-estacionário por meio de uma regressão linear ($Y = a + bx + e$), de uma série Y_{t-1} , isto é, $\Delta Y_t = \phi Y_{t-1} + u_t$. Em seguida, analisa-se a significância do coefi-

ciente ϕ , se $\phi = 0$, dizemos que o processo é não estacionário, se não, é estacionário. Para investigar a suposição de estacionariedade nesta dissertação, aplicamos o teste Dickey-Fuller Aumentado (ADF), que tem como base o DF, sendo que introduz um operador de defasagens $\sum_1^n \beta_k \Delta X_{t-k}$ para resolver o problema da autocorrelação no termo do erro u_t tal como: $\Delta X_t = \alpha X_{t-1} + \sum_1^n \beta_k \Delta X_{t-k} + u_t$. O testes DF e ADF podem ser vistos mais amplamente em [Fuller \(2009\)](#).

No entanto, a característica mais importante da uma série temporal é a autocorrelação, que se refere à memória ou correlação interna da série. Ou seja, é quando o efeito de um evento ocorrido em um instante pode ser detectado em muitos *lags* (unidades de tempo) depois ([SOUZA; TABAK; CAJUEIRO, 2006](#)). A função de autocorrelação (FAC) é eficiente para mensurar a interdependência de uma ST. Segundo [Ehlers \(2007\)](#), a autocorrelação quantifica a correlação entre as observações de uma mesma variável em diferentes horizontes de tempo, isto é, correlações entre observações defasadas 1, 2, ... períodos de tempo. Assim, a correlação entre observações x_i defasadas de k períodos de tempo pode ser calculada pela equação 3.7:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}, k = 0, 1, 2, \dots \quad (3.7)$$

A equação 3.7 fornece o coeficiente de autocorrelação adimensional e $-1 \leq r_k \leq 1$. [Guedes \(2014\)](#) simplificou a interpretação de r_k como sendo a covariância da série na defasagem k dividida pela variância da série, de modo que N representa o comprimento da série e \bar{x} é o valor médio das observações, determinadas a partir da variação temporal k . Assim, podemos considerar que se:

- (ii) $-1 \geq r_k > 0$ a série possui autocorrelação negativa;
- (ii) $r_k = 0$ não existe autocorrelação na série;
- (i) $0 < r_k \leq 1$ a série possui autocorrelação positiva.

O correlograma é o gráfico com os r_k primeiros coeficientes de autocorrelação como função de k , também conhecido como função de autocorrelação amostral (FAC). Segundo [Filho \(2009\)](#), a FAC possibilita caracterizar uma série temporal como sendo:

- i) Estacionária, se os valores r_k decrescem rapidamente para zero; não estacionária, se

as estatísticas r_k decrescem lentamente para zero e periódica, se a série apresentar característica de sazonalidade. Divergindo, assim, das séries estacionárias, que apresentam correlações que divergem de zero, para k grande, oscilando indefinidamente;

ii) Ruído branco, diante de uma série de várias aleatórias independentes, identicamente distribuídas, as correlações seriais serão nulas, exceto r_0 que é igual a 1;

iii) De memória curta, se as estatísticas de r_k (FAC) decrescem rapidamente para zero e sua dependência temporal desaparece rapidamente; de memória longa, quando a FAC decresce lentamente para zero e sua dependência temporal permanece presente por um longo período.

De acordo com [Gujarati e Porter \(2011, p.747\)](#), a significância estatística de qualquer r_k pode ser definida pelo seu erro padrão. Desta forma, se uma série apresentar comportamento de ruído branco (puramente aleatório), os coeficientes de autocorrelação amostral serão normalmente distribuídos com média zero e variância $1/n$, sendo n o tamanho da amostra. Nesse sentido, seguindo as propriedades da distribuição normal padrão, o intervalo de confiança de 95% é mostrada como duas linhas horizontais tracejadas.

Dito isto, uma vez calculados os coeficientes de correlação r_k , é importante investigar se todos eles são simultaneamente iguais a zero para verificar se são correlacionados ou não ([GREENE, 2002](#)). Esse cálculo pode ser feito através da estatística LB de [Ljung e Box \(1978\)](#), denotada por (3.8):

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{r_k^2}{n-k} \right) \sim \chi_m^2 \quad (3.8)$$

em que n = tamanho da amostra e m = duração da defasagem (*lag*).

A estatística LB segue a distribuição qui-quadrado com m graus de liberdade em que, se seu valor calculado exceder o valor crítico da tabela $\sim \chi_m^2$ ao nível escolhido de significância, rejeita-se a hipótese nula de que todos os r_k são iguais a zero, concluindo-se, portanto, que pelo menos alguns deles devem ser diferentes de zero. Neste caso, os coeficientes de correlação amostral r_k não são normalmente distribuídos logo, a série é provavelmente não estacionária.

3.3.2 Contribuições de um SIG para monitoramento e prevenção de acidentes de trânsito

A segurança viária tem sido uma grande preocupação para o Brasil e para diversos outros países do mundo. Uma das maiores preocupações da comunidade científica é reduzir o número de acidentes com vítimas fatais no trânsito. Uma maneira de investigar os altos índices de mortos por acidentes de trânsito é o estudo da análise espacial estatística baseada em ferramentas de um SIG (Sistemas de Informação Geográfica). A análise espacial tem se mostrado bastante útil para o estudo de padrões espaciais de distribuição de dados das áreas de Transportes, Recursos Naturais, Energia, Planejamento Urbano, Cartografia, no entanto, nem sempre foi assim. Até a década de 1950 a coleta das informações geográficas de recursos minerais, animais e de vegetação era feita apenas em documentos e mapas em papel. Com o desenvolvimento da tecnologia, o Geoprocessamento¹ ganhou espaço e possibilitou o armazenamento e a representação dessas informações em ambientes computacionais, com amplo potencial, sobretudo, devido ao seu custo relativamente baixo e ao acesso local do conhecimento (CÂMARA; DAVIS; MONTEIRO, 2001).

Segundo Câmara, Davis e Monteiro (2001), nos anos 1950, a Inglaterra e os Estados Unidos inauguraram as primeiras tentativas de automatização parcial de dados espaciais com estudos de botânica (Inglaterra) e volume de tráfego (Estados Unidos). Somente a partir da década de 1960, no Canadá, surgiram os primeiros SIG, com objetivo de criar um inventário de recursos naturais, porém ainda pouco acessíveis dada as dificuldades de uso, necessidade de mão de obra especializada e altíssimo custo. Durante os anos 1970, foram desenvolvidos novos e mais acessíveis *hardwares* com alguns fundamentos matemáticos de geometria computacional voltados para a Cartografia. Mas foi na década de 1980, com o avanço da microinformática e o estabelecimento de centros de estudos em informações geográficas, além do surgimento e evolução dos computadores pessoais e dos sistemas de gerenciadores de bancos de dados relacionais, que o uso de SIG sofreu uma grande difusão com ritmo crescente até os dias atuais e alternativas menos custosas para a construção de bases de dados geográficas.

Desta forma, diversos estudos de âmbito nacional e internacional vêm sendo desenvolvidos com o intuito de compreender o padrão de distribuição espacial dos acidentes, como por exemplo Ivajnsič et al. (2021b), Diesel (2005), Rocha, Fernandes e Lustosa (2011), França (2008), Blazquez e Celis (2013), Betgamaschi (2010), Souza, Cavenaghi e Alves (2016), Costa e Moraes (2009), Dai (2012). Nessa perspectiva, Alkhadour et al. (2021) argumentam que análises desenvolvidas através de um SIG podem ajudar na formulação de políticas voltadas para a segurança no trânsito e encontrar melhores soluções pra fatores que causam esses acidentes.

¹Segundo Câmara (2001), Geoprocessamento denota a disciplina do conhecimento que utiliza técnicas matemáticas e computacionais para o tratamento da informação.

Neste projeto, nos atemos ao estudo dos dados espaciais por meio da análise exploratória sem a intenção de nos aprofundarmos em qualquer de suas técnicas. Segundo [Monteiro et al. \(2004\)](#) essas técnicas são essenciais ao desenvolvimento das etapas da modelagem estatística espacial. Contudo, pretendemos ilustrar os acidentes com morte nos locais de maior incidência segundo o tipo de acidentes, com o intuito de introduzir a ideia de técnicas para modelagem de processos espaciais não estacionários.

3.3.3 Método R/S

A análise de Hurst ou método R/S é um dos mais conhecidos métodos usados para estimar memória de longo alcance ([MANDELBROT, 1997](#)). Hurst passou a vida estudando o rio Nilo e os problemas relacionados ao armazenamento de água. Ele inventou um novo método estatístico - o intervalo reescalado (análise R/S) - que descreveu em detalhes no livro, *Long-Term Storage: An Experimental Study* ([BLACK; HURST; SIMAIKA, 1965](#)). Como uma introdução a este método, [Feder \(1988\)](#) considerou o Lago Albert, um exemplo discutido por Hurst cujo problema era determinar o projeto de um reservatório ideal, isto é, que nunca transborda ou esvazia, baseado em registro de descargas observadas no lago. Segundo [Souza, Tabak e Cajueiro \(2006\)](#), a ideia consiste em determinar quais eram os valores máximo e mínimo no reservatório, de maneira que a diferença entre eles seria o intervalo (*range*) de volumes do reservatório. Assim, a estatística R/S seria a razão entre o *range* e o desvio-padrão das vazões da água, sendo este adimensional. Ao analisar essa estatística para diferentes períodos de tempo, Hurst descobriu que havia uma função que relacionava o valor de R/S ao número de observações envolvidas no cálculo.

O algoritmo para calcular o expoente de Hurst, seguem os passos:

1^o) Calcular a média $\langle \xi \rangle_\tau$ do fluxo de água no período τ determinada por:

$$\langle \xi \rangle_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi(t) \quad (3.9)$$

sendo que $\xi(t)$ denota o influxo da água no reservatório no instante t .

2^o) Determinar a variação acumulada da média v para os períodos de 1 a t como:

$$X(t, \tau) = \sum_{v=1}^t [\xi(v) - \langle \xi \rangle_\tau] \quad (3.10)$$

onde $\langle \xi \rangle_\tau$ representa a média de todos os pontos da série no primeiro período ($t = 1$) ao

último ($t = \tau$)

3^o) Calcular a diferença entre os valores máximo e mínimo de $X(t)$ no período τ

$$R(\tau) = \text{Max}[X(t, \tau)] - \text{Min}[X(t, \tau)], \quad (3.11)$$

para $1 \leq t \leq \tau$

4^o) Dividir $R(t)$ pelo desvio padrão $S(t)$,

$$R/S = \frac{R(t)}{S(\tau)}, \quad (3.12)$$

sendo

$$S(\tau) = \sqrt{\frac{1}{\tau} \left(\sum_{t=1}^{\tau} [\xi(t) - \langle \xi \rangle_{\tau}]^2 \right)}, \quad (3.13)$$

Ao repetir o procedimento sistematicamente para tamanhos distintos de (τ) , se a relação R/S seguir uma lei de potência, então teremos:

$$R/S = (\tau/2)^H \quad (3.14)$$

em que H representa o expoente de Hurst.

O expoente de Hurst possui campo de variação entre 0 e 1, possibilitando informar a persistência ou não do sinal: se $H = 0,5$ indica que a série é antipersistente. Para $0,5 < H < 1$ a série é persistente. Caso $0 \leq H < 0,5$, diz-se que a série é antipersistente.

Além da análise R/S, diversos métodos possibilitam mensurar a existência de autocorrelação em séries temporais (TAQQU; TEVEROVSKY; WILLINGER, 1995). No presente trabalho aplicaremos o método *Detrended Fluctuation Analysis* (DFA) Peng et al. (1994) para estudar a presença (ou não) de autocorrelação das séries de acidentes de trânsito.

3.3.4 Método DFA

O método DFA (*Detrended Fluctuation Analysis*) foi desenvolvido por Peng et al. (1994) ao estudar correlação de leis de potência de longo alcance para sequências de DNA. Trata-se de uma evolução da análise R/S para verificar correlações de longo alcance em séries temporais não estacionárias, mostrando-se mais robusto (HU et al., 2001). É um método de análise de escala utilizado para estimar expoentes que caracterizam correlação de longo

alcance e facilita a acurácia das previsões. O DFA permite também encontrar correlações para escalas de diferentes tamanhos e evita a detecção de correlações aparentes de longo alcance, o que demonstra a robustez do método (FILHO, 2014). Diversos trabalhos nacionais e internacionais que já foram desenvolvidos em diferentes campos de interesse adotaram o método DFA, a exemplo de séries biológicas (PENG et al., 1993), (VISWANATHAN et al., 1997), (LJUBISAVLJEVIC, 1999), (ZEBENDE; FILHO; CRUZ, 2017), (ARSAC; DESCHODT-ARSAC, 2018), epidemiologia (AZEVEDO; SABA; MORET, 2014), mercado financeiro (GUEDES et al., 2022), (DIAS et al., 2021), climatologia (ZEBENDE et al., 2021), (FIGUEIREDO, 2014), (LIMA et al., 2017), séries temporais sonoras (LACZKOWSKI, 2013).

Segundo Peng et al. (1994), o algoritmo para a aplicação do método DFA segue seis passos:

1^o) Considerando uma série temporal $\{x_i\}$ com $i = 1, \dots, N$ número total de pontos da série temporal, como por exemplo a quantidade de pessoas envolvidas em acidentes de trânsito nas rodovias federais da Bahia em 2020, conforme a Figura 3.8,

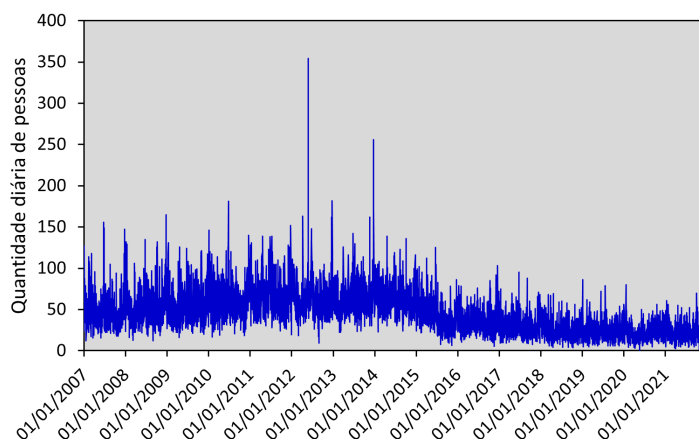


Figura 3.8: Bahia: Série temporal das pessoas envolvidas acidentes de trânsito nas rodovias federais, 2007-2021.

é preciso determinar os desvios de cada observação em relação à incidência média da série completa $\langle x \rangle$, calculada pela equação 3.16. Assim eliminamos o grande número de oscilações, integrando $\{x_i\}$ e obtendo a série integrada denotada por X_k (Figura 3.9), conforme a equação 3.16;

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.15)$$

$$X_k = \sum_{i=1}^k [x_i - \langle x \rangle] \quad (3.16)$$

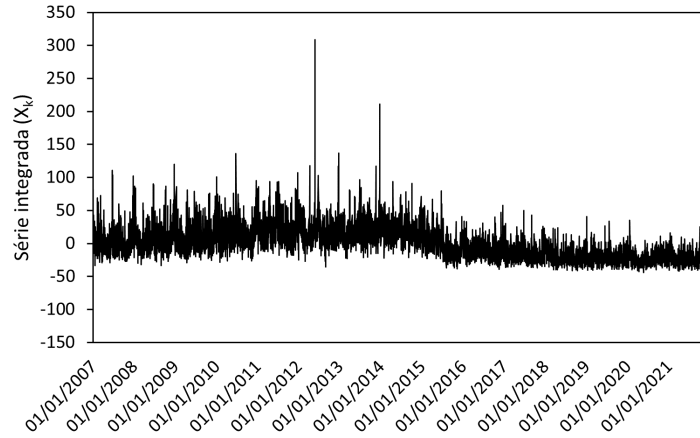


Figura 3.9: Bahia: Série integrada das pessoas envolvidas acidentes de trânsito $\{X_k\}$ nas rodovias federais, 2007-2021. Nota: Cálculo do autor.

^{2º}) O sinal X_k é dividido em $N - n$ intervalos (caixas ou box) sobrepostos (Figura 3.10) de igual comprimento n (escala temporal). Cada caixa contém $(n+1)$ valores, que começam em i e terminam em $(i+n)$;

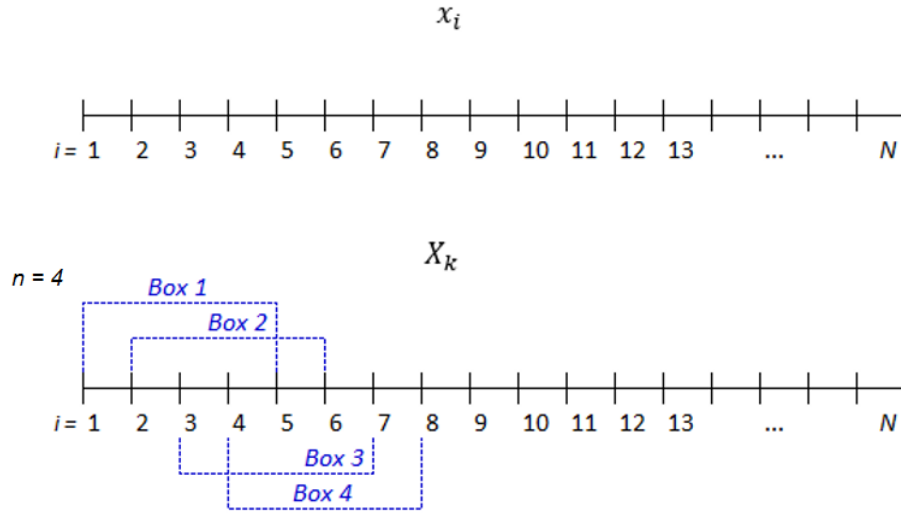


Figura 3.10: Divisão do sinal integrado X_k em $N-n$ boxes sobrepostos de tamanho $n = 4$, contendo $n+1$ valores.

Nota: Elaboração do autor.

3^o) Para cada tamanho n de caixa, é necessário ajustar um polinômio de grau ≥ 1 em X_k . Esse ajuste representa a tendência local $\tilde{X}_{k,i}$ em cada box de tamanho n , com $i \leq k \leq i + n$ (valor ajustado);

4^o) A série integrada X_k é subtraída da série ajustada $\tilde{X}_{k,i}$ em cada caixa de tamanho n , obtendo-se assim:

$$f_{DFA}^2(n, i) = \frac{1}{n+1} \sum_{k=i}^{i+n} [X_k - \tilde{X}_{k,i}]^2 \quad (3.17)$$

5^o) Em seguida, para cada box de tamanho n calcula-se a função de flutuação, $F_{DFA}(n)$ denotada pela equação 3.18:

$$F_{DFA}(n) = \sqrt{\frac{1}{(N-n)} \sum_{i=1}^{N-n} f_{DFA}^2(n, i)} \quad (3.18)$$

Repete-se o cálculo acima para diferentes escalas (caixas de tamanho n), de modo que $4 \leq n \leq \frac{N}{4}$.

6^o) Assim, verifica-se se a função $F(n)$ segue o comportamento de uma lei de potência do tipo $F(n) \sim n^\alpha$, onde α é o expoente de correlação de longo alcance podendo ser obtido

a partir de uma relação linear $\log F(n) \times \log(n)$. O coeficiente angular da reta representa o expoente α (Figura 3.11).

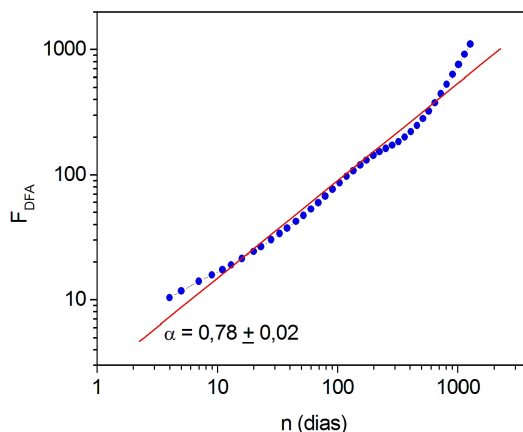


Figura 3.11: Comportamento do expoente α das pessoas envolvidas em acidentes de trânsito nas rodovias federais da Bahia em 2007-2021.

Desta forma, α é o expoente de escala de autocorrelação de longo alcance, de modo que se $\alpha = 0,5$ o comportamento da série é dito aleatório, o sinal não está correlacionado. Se $\alpha < 0,5$, a série apresenta comportamento antipersistente. Por fim, se $\alpha > 0,5$ o sinal é persistente, conforme mostra a Tabela 3.1. De maneira resumida, Filho (2009) afirma que o expoente α obtido pelo método DFA permite avaliar em que medida a tendência observada na série temporal passada implica em manutenção do comportamento futuro, indicando um efeito de memória de longa duração na série.

Tabela 3.1: Classificação da faixa dinâmica e valores do expoente de flutuação α_{DFA}

Expoente	Tipo de sinal
$\alpha_{DFA} < 0,5$	Antipersistente de longo alcance
$\alpha_{DFA} \cong 0,5$	Não correlacionado, ruído branco
$\alpha_{DFA} > 0,5$	Persistente de longo alcance

Fonte: Dias et al. (2021)

3.3.5 Método DCCA

Quando o resultado de determinada série sofre intervenção não somente dela mesma, mas também de outras séries, é importante analisar a existência ou não de correlação cruzada

entre as séries envolvidas no sistema estudado. Nessa perspectiva, no caso de analisarmos duas séries temporais, podemos aplicar uma generalização do método DFA, denominado método DCCA (*Detrended Cross-Correlation Analysis*), ou análise de correlação cruzada retirando tendências. O DCCA foi desenvolvido por [Podobnik e Stanley \(2008\)](#) e sua metodologia consiste em investigar leis de potência de correlação cruzada simultânea, entre duas séries temporais não estacionárias com o mesmo número de observações, caracterizada por um expoente estimado, além de identificar componentes sazonais ou periodicidade ([ZEBENDE; FILHO, 2009](#)).

O DCCA também é calculado a partir de seis passos, como mostra o algoritmo logo abaixo:

1^o) Considerando duas séries temporais de mesmo tamanho $\{x_i\}$ e $\{y_i\}$ com $i = 1, \dots, N$ o comprimento das séries, integram-se x_i e y_i para obter os sinais $X_k = \sum_{n=1}^k [x_i - \langle x \rangle]$ e $Y_k = \sum_{n=1}^k [y_i - \langle y \rangle]$, respectivamente, sendo $\langle x \rangle$ e $\langle y \rangle$ os valores médios das séries $\{x_i\}$ e $\{y_i\}$;

2^o) Os sinais X_k e Y_k são divididos em $N - n$ caixas sobrepostas de mesma amplitude n (escala temporal), sendo que cada caixa contém $(n + 1)$ valores que começam em i e terminam em $i + n$;

3^o) Para cada caixa, é calculada a tendência local por meio, por exemplo, de um ajuste linear em X_k e Y_k , obtendo assim $\tilde{X}_{k,i}$ e $\tilde{Y}_{k,i}$, sendo $i \leq k \leq (i + n)$ (valor ajustado);

4^o) Os sinais X_k e Y_k são subtraídos de $\tilde{X}_{k,i}$ e $\tilde{Y}_{k,i}$, nessa ordem, calculando-se assim a covariância dos resíduos em cada intervalo por meio da equação [3.19](#):

$$f_{DCCA}^2(n, i) = \frac{1}{n + 1} \sum_{k=i}^{i+n} (X_k - \tilde{X}_{k,i})(Y_k - \tilde{Y}_{k,i}) \quad (3.19)$$

5^o) Assim, para uma dada escala temporal de comprimento n determina-se a função de flutuação (*detrended covariance*), expressa pela equação [3.20](#):

$$F_{DCCA}^2(n) \equiv \frac{1}{N - n} \sum_{i=i}^{N-n} f_{DCCA}^2(n, i) \quad (3.20)$$

6^o) Repetindo os passos acima para uma gama de escalas temporais compreendidas por $4 \leq n \leq N/4$, é possível verificar a existência ou não de uma lei de potência, representada pela equação [3.21](#):

$$F_{DCCA} \approx n^{2\lambda} \quad (3.21)$$

O expoente de escala λ representa a correlação cruzada de longo alcance entre duas séries temporais $\{x_i\}$ e $\{y_i\}$. Além de estimar o expoente λ que caracteriza a correlação de longo alcance entre duas séries, é possível identificar através do DCCA sazonalidade das séries (ZEBENDE; FILHO, 2009). O DCCA tem sido aplicado em diferentes áreas da ciência, dentre os quais sistemas financeiros (PODOBNIK; STANLEY, 2008), (LIN; SHANG; ZHAO, 2012), fenômenos físicos (ZEBENDE; SILVA; FILHO, 2013), climatologia (BRITO et al.,), fluxo de tráfego (ZHAO, 2012).

3.3.6 Coeficiente de Correlação Cruzada

O método DCCA pode ser aplicado em diversos campos do conhecimento. Apesar dessa aplicação demonstrar ampla utilidade em diferentes campos de estudo, o DCCA não quantifica o nível de correlação cruzada entre as séries (VASSOLER; ZEBENDE, 2012). Nesse contexto, Zebende (2011) propôs o Coeficiente de Correlação Cruzada ou DCCA Cross-Correlation Coefficient (ρ_{DCCA}), um novo procedimento estatístico que permite mensurar o nível de correlação entre duas ST não-estacionárias de mesmo tamanho n .

O ρ_{DCCA} se baseia nos métodos DFA e DCCA, é definido pela razão entre a função de covariância sem tendência F_{DCCA}^2 e as funções de variância sem tendência F_{DFA} das duas séries analisadas $\{x_i\}$ e $\{y_i\}$, sendo formalmente expresso pela equação 3.22:

$$\rho_{DCCA}(n) \equiv \frac{F_{DCCA}^2(n)}{F_{DFA}\{x_i\}(n)F_{DFA}\{y_i\}(n)} \quad (3.22)$$

Sendo adimensional, o campo de variação do ρ_{DCCA} está compreendido entre -1 e 1, com $\rho_{DCCA} = 1$ correlação cruzada perfeita e $\rho_{DCCA} = -1$ anticorrelação cruzada perfeita. A ausência de correlação cruzada ocorre quando $\rho_{DCCA} = 0$. A tabela 3.2 mostra as condições para análise de correlação cruzada tendenciosa.

Tabela 3.2: Níveis de correlação cruzada

ρ_{DCCA}	Condição
1	Correlação cruzada perfeita
0	Correlação cruzada inexistente
-1	Anticorrelação cruzada perfeita

Fonte: Zebende (2011)

A robustez do ρ_{DCCA} foi provada teoricamente por [Podobnik et al. \(2011\)](#), por meio de teste estatístico tendo como base a *Desigualdade de Cauchy*. O teste mostrou que, para janelas sobrepostas, para o intervalo $[-1, 1]$ as correlação tornam-se estatisticamente significativas, levando em consideração o tamanho N da série e o tamanho n do *box*. Nesse sentido, podemos enumerar algumas vantagens do ρ_{DCCA} em relação a outros coeficientes de correlação cruzada (a exemplo do coeficiente de correlação cruzada de Pearson) como a possibilidade de mensurar correlação entre dois sinais em escalas diferentes de tempo ([WANG; XIE, 2013](#)), a análise da correlação entre as séries retirando a tendência a fim de não mascarar as verdadeiras correlações e ainda, a capacidade de analisar séries não lineares ([SILVA; CASTRO; FILHO, 2014](#)).

Desta forma, diversos campos do conhecimento já foram modelados por meio do coeficiente de correlação cruzada, como por exemplo estudo de acidentes de táxi ([SILVA, 2009](#)), de fenômenos climáticos ([VASSOLER; ZEBENDE, 2012](#)), do mercado financeiro ([SILVA et al., 2019](#)), ([REBOREDO; RIVERA-CASTRO; ZEBENDE, 2014](#)), ([SILVA et al., 2016](#)) ([SILVA, 2016](#)), de indicadores de criminalidade ([SAMPAIO et al., 2008](#)), do comportamento da produção industrial ([GUEDES, 2014](#)), da dinâmica temporal dos casos de dengue ([FILHO et al., 2015](#)), sinais biológicos ([FILHO, 2019](#)) dentre outros.

3.4 Percurso metodológico

A presente pesquisa é caracterizada por um estudo exploratório e descritivo, visto que tem como objetivo principal descrever as características dos acidentes de trânsito nas rodovias federais da Bahia e analisar as correlações entre as variáveis envolvidas. No que tange ao levantamento do dados, trata-se de uma pesquisa documental ([LAKATOS; MARCONE, 2003](#)), abordagem quali-quantitativa, pois que pretende descrever e explicar a dinâmica dos acidentes de maneira a observar seus efeitos.

O desenvolvimento da pesquisa ocorreu em várias etapas, conforme mostra a Figura 3.12. A partir da identificação do problema, o estabelecimento dos objetivos e a delimitação da área de estudo, seguimos para a obtenção dos dados a fim de compor as séries de acidentes e construir a base de dados. Para a construção da base de dados, inicialmente, extraímos do portal virtual da PRF as estatísticas anuais de acidentes registrados nos municípios do Brasil em formato *csv*. Em seguida, no software RStudio filtramos os dados mantendo apenas as estatísticas referentes ao estado da Bahia e os exportamos para o Excel formando um único arquivo com os registros do período 2007-2021. A etapa seguinte abrangeu organização e tratamento dos dados, foi realizada com o suporte da linguagem R e dos *softwares* Excel e Qgis 3.10 para análise exploratória das séries temporais e elaboração de gráficos para a análise espacial dos acidentes. Para compor as bases em SIG, utilizamos também a malha dos municípios da Bahia disponibilizada

pelo IBGE e o *shape* das rodovias extraído do site do Ministério da Infraestrutura. A terceira etapa contemplou o processamento dos dados para análise de existência (ou não) de correlação cruzada de longo alcance entre as variáveis envolvidas no estudo. Esse passo foi desenvolvido por meio do aplicativo ρ_{DCCA} .exe e do software Origin 16.0.

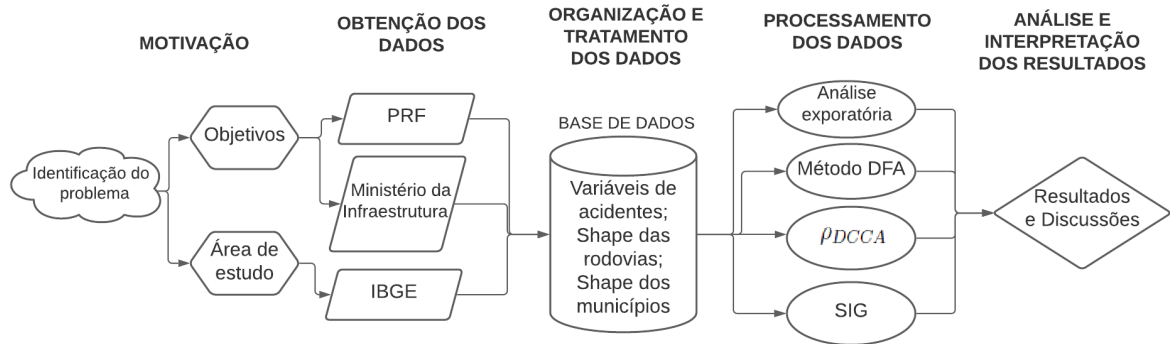


Figura 3.12: Esquema do processo metodológico empregado.

Resultados e Conclusões

Neste capítulo apresentamos os resultados alcançados na pesquisa, bem como o método empregado a partir dos objetivos propostos.

4.1 Caracterização dos acidentes

A cada dia, a Bahia registra em média, aproximadamente, 2 mortes e 9 acidentes nas rodovias federais. Somente em 2021, foram 3.306 acidentes, sendo 2.499 com vítimas, o que resultou em 234 mortes no ano, somente nas rodovias federais do estado. Nos quinze anos analisados (2007 a 2021), a Bahia teve 104.999 acidentes sendo 46.720 (44,4%) com vítimas (feridos leves + feridos graves) e 7.620 (7,3%) com mortos. Desse total de acidentes com 247.340 pessoas envolvidas, 59,1% são ilesos, 31,9% são feridos, sendo 23,7% feridos leves e 8,2% feridos graves e 3,9% contabilizaram mortos no local do acidente, ou seja, 9.721 vítimas fatais (Figura 4.1). Assim, em 2021 as vítimas em acidentes atingiram a marca de 35,8% com mortas ou feridas.

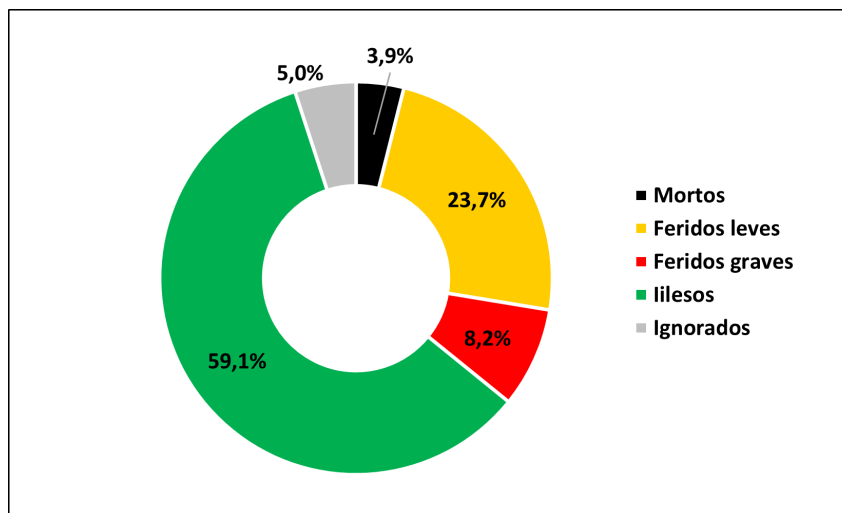


Figura 4.1: Bahia: Percentuais de pessoas envolvidas nos acidentes, segundo a situação da vítima, 2007-2021.

A evolução do total de acidentes apresentou acréscimo no decorrer dos anos 2007 a 2011 de cerca de 10%, em média, por ano. No período de 2012 a 2014, esses totais se mantiveram bem próximos, registrando uma média de 10.378 acidentes por ano, conforme retrata a Figura 4.1. No entanto, a partir de 2015 houve uma diminuição acentuada no

número de acidentes registrados pela PRF, sofrendo uma queda de 31% em relação ao ano imediatamente anterior, saindo de 10.407 para 7.152 acidentes.

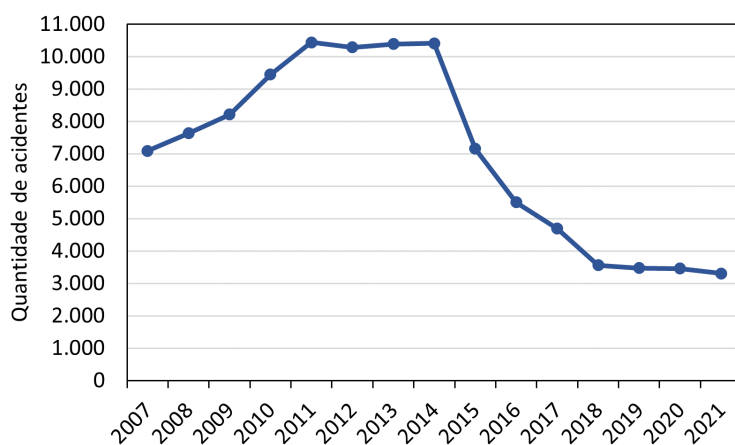


Figura 4.2: Bahia: Evolução dos acidentes de trânsito registrados nas rodovias federais do estado da Bahia, 2007-2021. Fonte: PRF. Nota: Dados processados pelo autor.

Importante destacar que a partir junho de 2015, o registro dos acidentes sem vítimas (ilesos) nas rodovias federais passou a ser de responsabilidade do condutor e não mais de um agente da PRF, por meio do formulário eletrônico (e-DAT). Sendo a PRF responsável por registrar apenas os acidentes que envolvem pessoas feridas ou mortas. Desta forma, somente os acidentes com vítima tinham a garantia de serem, de fato, reportados visto que, neste caso, a ocorrência é realizada pessoalmente por policiais rodoviários federais. O eDAT vale somente para acidentes que não tenham feridos ou mortos, sem dano ao meio ambiente ou patrimônio público, envolvendo menos que cinco veículos e que nenhum deles esteja transportando produtos perigosos. Com essa ferramenta, o registro de acidentes sem vítimas nas rodovias federais passou a ser realizado exclusivamente via formulário eletrônico, passando a ser de responsabilidade dos envolvidos em acidentes sem vítimas nas rodovias a iniciativa de informar as ocorrências. Segundo a PRF, a mudança no procedimento que se refere apenas ao registro da declaração visa agilizar e otimizar o atendimento aos cidadãos (EBC, 2015). Diante disso, vale ressaltar que não se pode tomar dados de notificação de acidentes como se fossem o universo dos acidentes devido à subnotificação dos dados, ou seja, os registros policiais de acidentes nas rodovias federais apenas são uma estimativa subestimada dos acidentes. Segundo, (KAHN, 2005) o fenômeno de subnotificação ainda que possa variar em grau de país para país, é algo que atinge a todos.

A Figura 4.1 mostra que desde de 2015, após a mudança na forma de registro dos acidentes, o número de ilesos sofreu queda acentuada (38,6%). Ao mesmo tempo, a quantidade de feridos leves também apresentou queda em seus totais por ano a partir de 2015, porém numa proporção menor (13,0%). Mortos e feridos graves mantiveram-se, praticamente, no

mesmo patamar, não apresentando mudança significativa após a alteração da normativa ao longo dos anos.

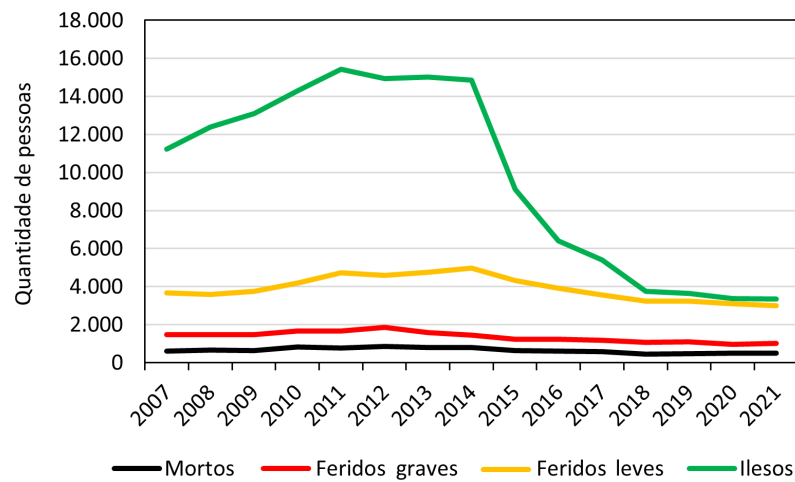
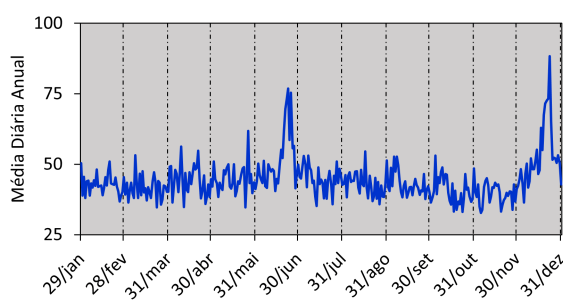
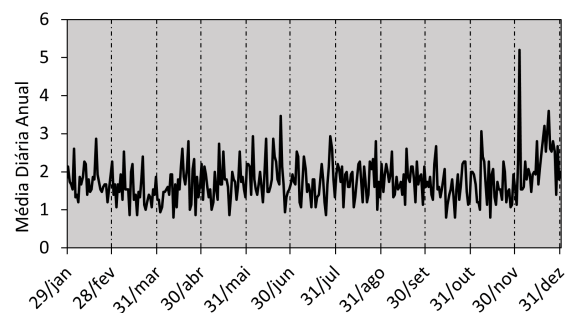


Figura 4.3: Bahia: Evolução da quantidade de pessoas envolvidas em acidentes de trânsito segundo a situação de gravidade, 2007-2021.

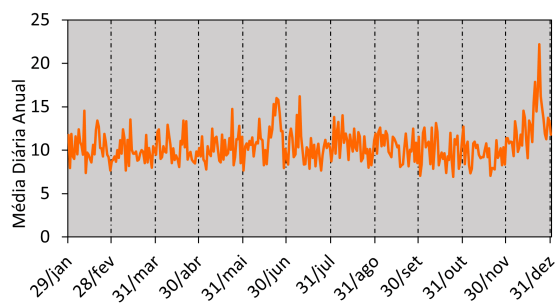
As médias diárias anuais revelam um comportamento sazonal em alguns períodos do ano. De acordo com a Figura 4.4, os meses de junho e dezembro, quando ocorrem as festas juninas e natalina, respectivamente, registraram as médias mais elevadas de *Feridos leves* e *Mortos*, sobretudo nos últimos dias do ano. Nesses períodos também notamos um aumento expressivo na média de *Pessoas* e *Veículos* envolvidos nos acidentes.



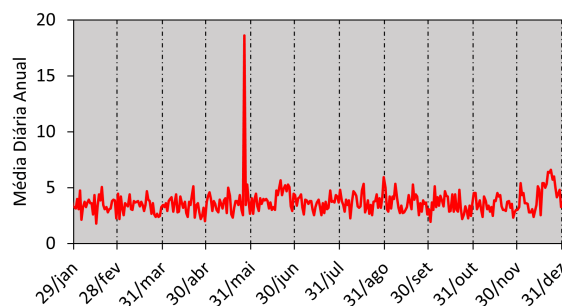
(a) *Pessoas*



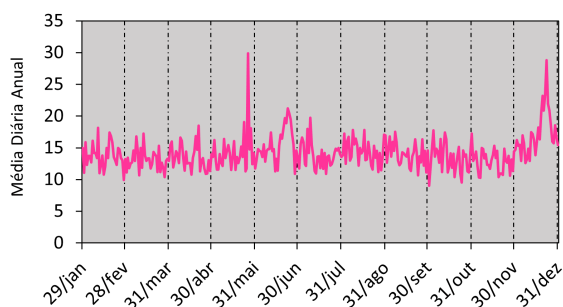
(b) *Mortos*



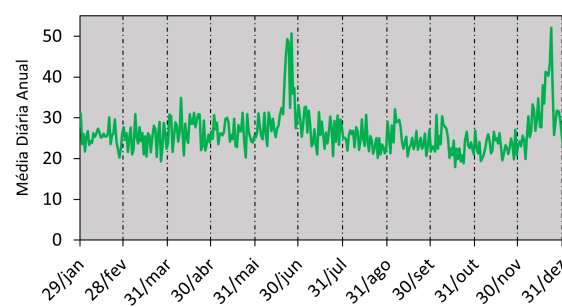
(c) *Feridos leves*



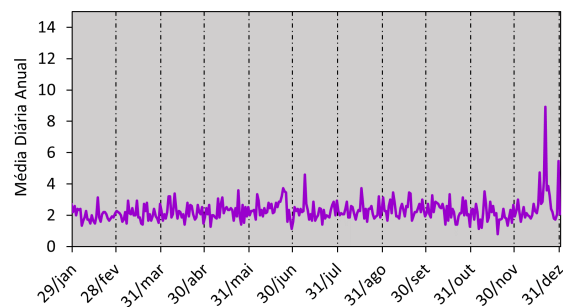
(d) *Feridos graves*



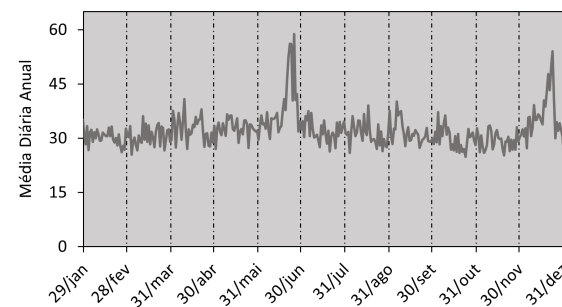
(e) *Feridos*



(f) *Ilesos*



(g) *Ignorados*



(h) *Veículos*

Figura 4.4: Bahia: Média diária anual das séries de acidentes, 2007-2021.

Segundo dados da PRF, de janeiro a março de 2020, antes da implementação das medidas de isolamento no país em virtude da pandemia da COVID-19, a Bahia registrou 839 acidentes de trânsito nas rodovias federais. Houve um aumento de 3,4% dos acidentes registrados no mesmo período de 2019, isto é, 811 acidentes. Durante o decreto de isolamento social (abril a dezembro de 2020), período de menor circulação da população, houve uma inversão dos dados registrando-se 2.660 acidentes em 2019 e, no mesmo período de 2020, foram 2.615, ou seja, houve uma redução de 1,7% em relação ao mesmo período do ano anterior. Apesar dessa redução, no que tange à quantidade de mortes nesses períodos,

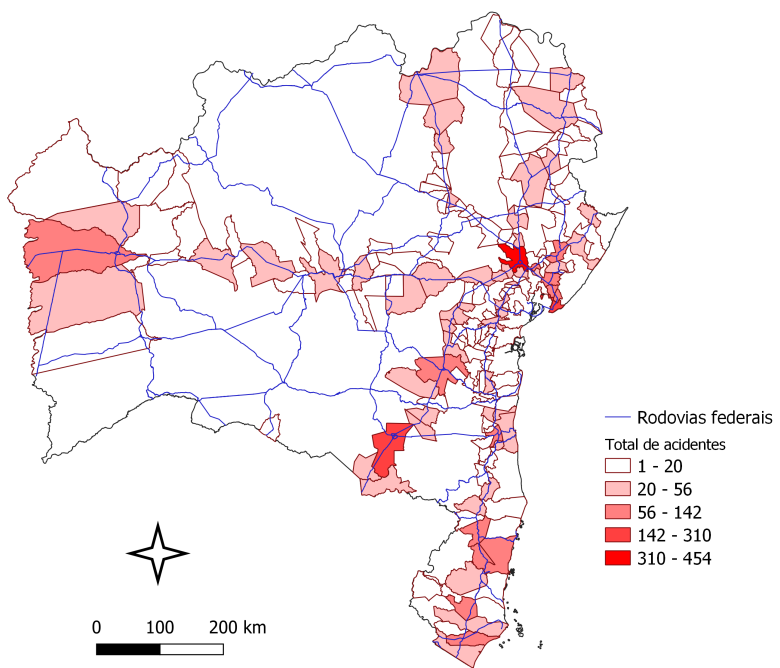
observamos um aumento de 50% das vítimas fatais no período pré pandemia (janeiro a março de 2020) e queda de apenas 3,1% no total de mortos período pandêmico de 2020 (abril a dezembro).

Do total de 104.999 acidentes registrados nas rodovias federais da Bahia entre 2017-2021, mais de 3/4 deles (75,6%) ocorreram nas três rodovias principais do estado, BR 101, BR 324 e na BR 116. Quase metade do total das ocorrências (58,9%) foi caracterizada pelos tipos colisão traseira (22,6%), saída de pista (13,7%) e colisões lateral (13,1%) e transversal (9,5%), tendência que se repete naquelas três principais rodovias da Bahia. O principal tipo de acidente causando morte foi colisão frontal, responsável por 33,3% dos acidentes com óbitos.

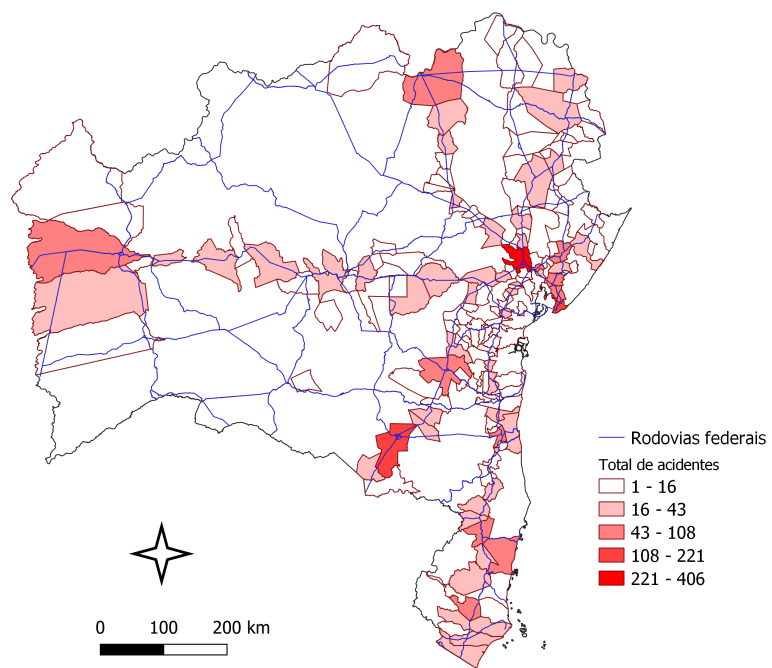
As causas dos acidentes com morte são as mais diversas. Entretanto, pelo menos 50,5% deles foram causados devido a uma postura indevida do condutor: falta de atenção (17,3%), velocidade incompatível (11,4%), ultrapassem indevida (11%), ingestão de álcool (4,5%), dormir ao volante (4,7%), desobediência à sinalização e/ou às normas de trânsito pelo condutor (2,6%). Apenas 1,8% dos acidentes foram causados por defeito no veículo, por motivo elétrico, mecânico ou de suspensão. Por outro lado, as 27.176 ocorrências em que tiveram as pessoas envolvidas tiveram apenas ferimentos leves ou ficaram ilesas tiveram 72,4% dos acidentes do tipo colisão (traseira 29,1%, lateral 17,6%, transversal 16,6% e frontal 9,1%). Nesses casos, 30,7% dos acidentes foram causados por falta de atenção do condutor. Do total de 104.999 acidentes, 32,4% deles e 22,9% das 27.176 de pessoas ilesas ou com ferimentos leves envolvidas, representam a categoria *outras causas*, reforçando a existência do fenômeno de subnotificação dos dados.

Fazendo uma análise espacial dos acidentes, por município ao longo dos anos 2017-2020¹, observamos que a maioria das ocorrências se dão, praticamente, nas mesmas localidades (Figura 4.1). Os municípios Feira de Santana, Salvador, Vitória da Conquista, Barreiras, Eunápolis, São Sebastião do Passé e Simões Filho se destacam negativamente com os maiores volumes de acidentes nesse período, acima de 2 900 acidentes. O número de ocorrências com morte foram maiores também em alguns desses municípios: Feira de Santana (521), Vitória da Conquista (385), Barreiras (293), Jequié (238) e Salvador (205).

¹Os dados abertos de acidentes disponíveis no portal virtual da PRF só foram georreferenciados a partir de 2017. No entanto, os arquivos referentes ao ano de 2021 apresentam aparente inconsistência no formato das coordenadas geográficas (latitude e longitude) impossibilitando estabelecer comparação com os anos anteriores.



(a) 2017



(b) 2018

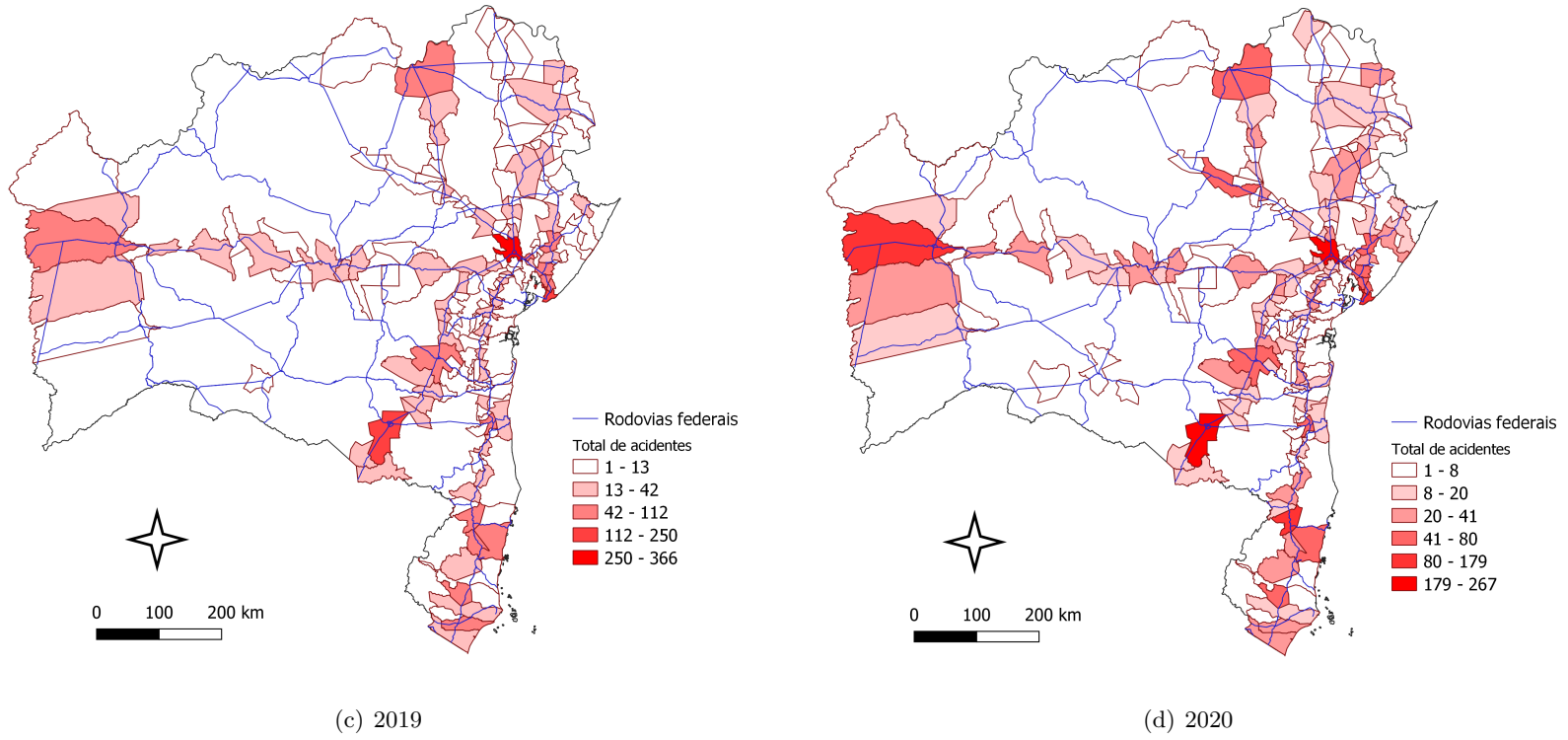


Figura 4.5: Evolução dos acidentes por municípios da Bahia, 2017-2020.

Apesar de apresentar com precisão as informações ano a ano, estes mapas não mostram onde está localizada a maior concentração dos acidentes. Nesse sentido, para melhor análise e visualização deste evento, usaremos o mapa de calor (*Heatmap*) que possibilita identificar a concentração dos locais críticos dos acidentes com vítimas fatais. Enquanto o mapa de graduação do volume de acidentes permite visualizar em quais locais concentram-se as rodovias que registraram maior frequência de acidentes no período, o mapa de calor revela a densidade de ocorrência dos acidentes por área em diferentes tons. As tonalidades mais escuras representam áreas de maior intensidade e as mais claras, a de menor intensidade. Aqui, a cor vermelha indica maior intensidade, representando agrupamento dos acidentes em regiões pontuais. Uma concentração um pouco menor do fenômeno está sendo representada pela cor laranja e amarela. As regiões em azul no interior do mapa representam uma intensidade ainda menor.

A partir da análise dos acidentes com morte no período global 2017-2020, é possível identificar algumas áreas de maior concentração se repetem assim como na análise anual dos dados. É possível notar que os pontos de maior intensidade aparecem destacando negativamente os municípios de Feira de Santana e Vitória da Conquista como áreas de maior raio de concentração de mortos (Figura 4.6).

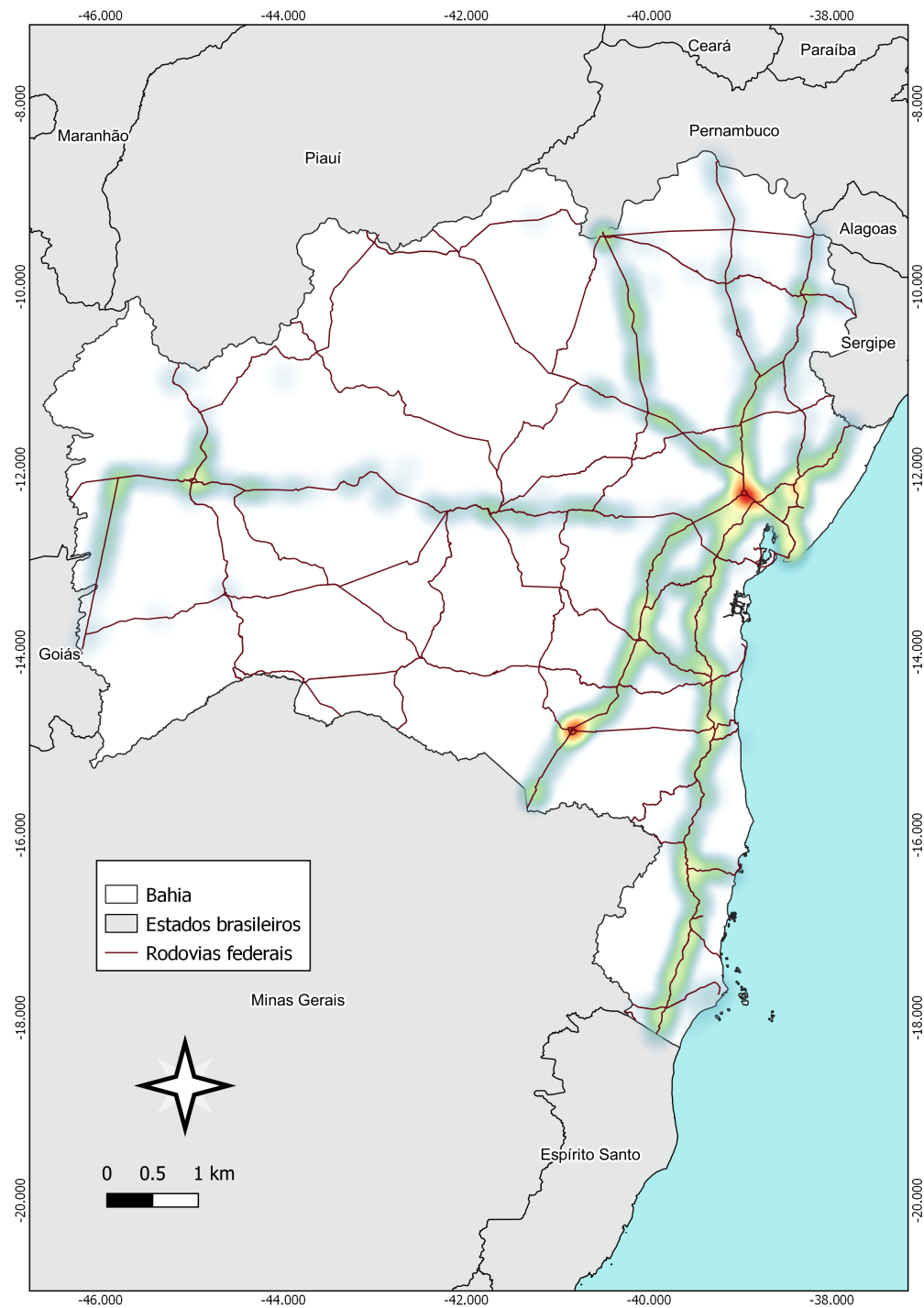
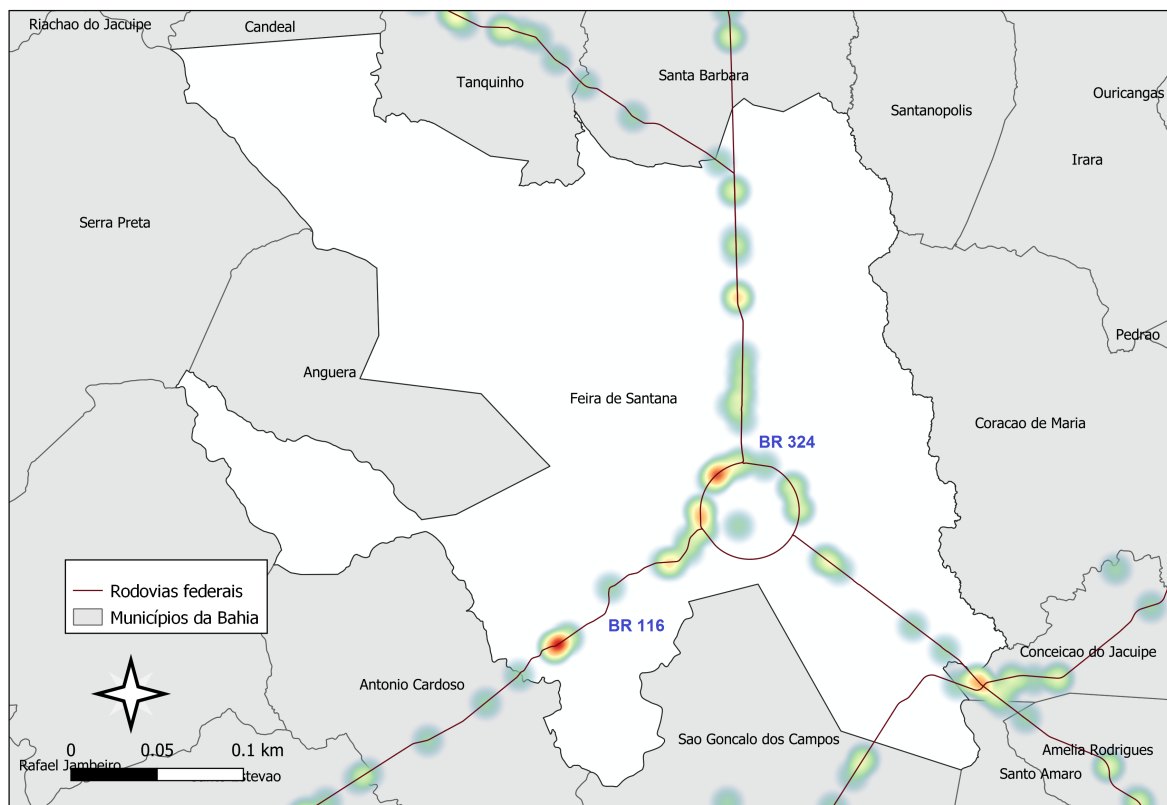


Figura 4.6: Concentração de mortos nos municípios com maior volume de ocorrências registradas em 2017-2020.

Sendo colisão frontal o tipo de acidente com morte com maior participação em 2017-2020, seguem os mapas de calor das duas cidades com maior concentração deste fenômeno. Feira de Santana foi município baiano com maior densidade de óbitos no local do acidente. O fenômeno se concentra no Anel de Contorno (Figura 4.7(a)), BR 324 e BR 116, representando um ponto estratégico da cidade e de grande relevância para o município,

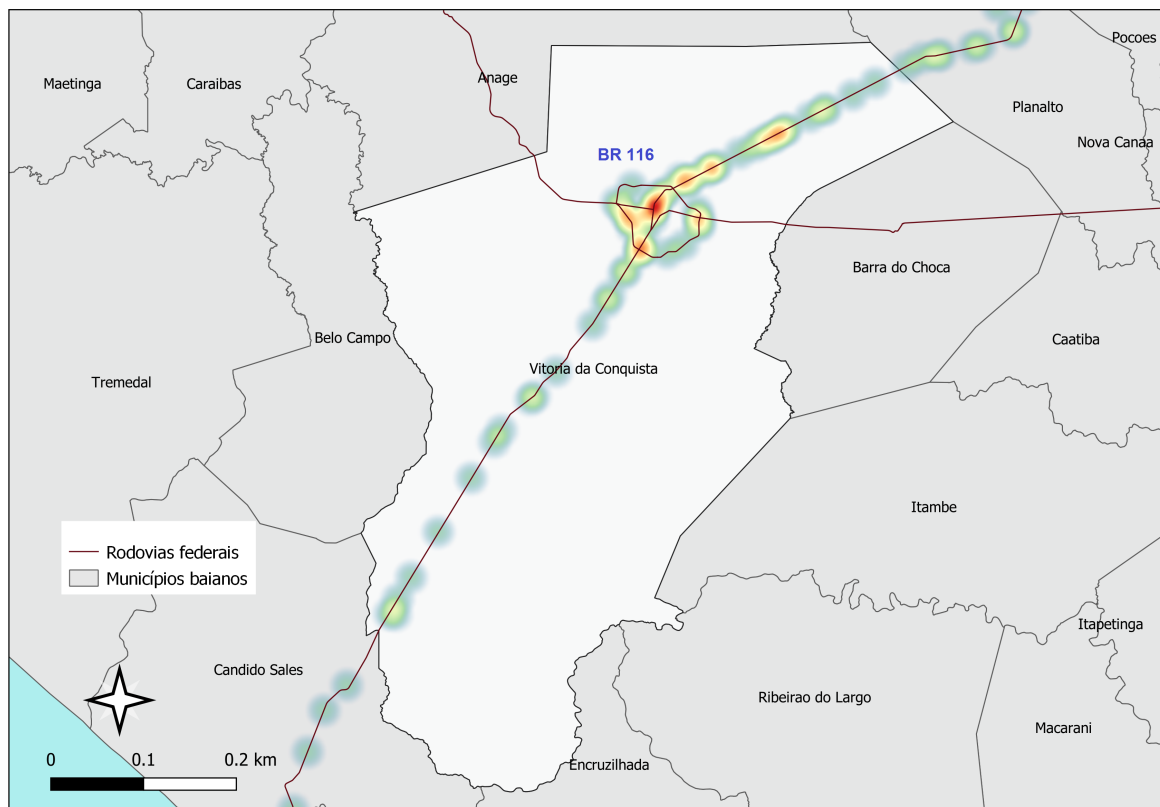
dado o alto fluxo de veículos e por onde dá acesso à cidade. Importante ressaltar que até 2018, quando foi realizada a primeira duplicação do Anel de Contorno com o objetivo de minimizar a frequência de acidentes no local, a média de acidentes nos anos 2016-2017 era de 227 por ano ². No período 2018-2021 essa média caiu para 148 acidentes por ano. Isso pode ser explicado não apenas pela subnotificação dos registros após 2015. A partir de 2020 a fiscalização, autuação e aplicação de medidas administrativas cabíveis concernentes às infrações de trânsito na Avenida Eduardo Fróis Mota, o Anel de Contorno, passaram a ser de responsabilidade da Superintendência Municipal de Tráfego (SMT). Assim, o levantamento dos locais de acidentes e dos serviços de atendimento, socorro e salvamento de vítimas nos trechos de rodovias federais: BR-116 (entre o Complexo Viário da Cidade Nova e o viaduto próximo a Pousada da Feira) e na BR-324 entre os quilômetros 512,6 e o 519,5 - viaduto Portal do Sertão, deixaram de ser realizados pela PRF. Desta forma, esses registros deixaram de fazer parte das estatísticas de acidentes reportados pela PRF (SMT, 2020).



(a)

Figura 4.7: Feira de Santana, Bahia: Concentração de mortos nos municípios com maior volume de ocorrências registradas em 2017-2020.

²Até 2015 a média era 533 no entanto, não foi considerada no cálculo para evitar imprimir viés na análise proposta, devido à mudança na metodologia de registro de acidentes eDat



(a)

Figura 4.8: Vitória da Conquista, Bahia: Concentração de mortos nos municípios com maior volume de ocorrências registradas em 2017-2020.

Notamos que os municípios que apresentaram maior incidência de acidentes fatais, também são as duas das cidades mais populosas do estado depois da capital baiana: Feira de Santana 624,1 mil hab. e Vitória da Conquista 343,6 mil hab.. Vale destacar que, de acordo com o Ministério da Saúde, desde 2006, a violência no trânsito no Brasil está se deslocando das maiores para as menores cidades. Essa tendência se repete também em algumas cidades da Bahia entre 2017 e 2021. A taxa média de mortes nas rodovias federais nos municípios com até 5 mil hab., que registraram acidentes com óbito em 2021, corresponde a 33,3 por cem mil residentes enquanto naqueles com população acima de 50 mil pessoas a média foi de 6,0 por cem mil moradores. O ministério aponta dois fatores que contribuíram para este cenário: a população nas pequenas cidades cresceu menos do que no resto do país, enquanto a violência no trânsito cresceu mais (PINHO, 2008).

No que tange ao volume de acidentes registrados em 2021, a Bahia contabilizou 8.375 pessoas envolvidas. Desse total, 6.095 (72,8%) eram do sexo masculino e 1.698 (20,3%) eram do sexo feminino, 582 (6,9%) não foram identificadas. Estatísticas semelhantes a essas aparecem em outras pesquisas de acidentes de trânsito: (BARROS et al., 2003),

(JUNIOR; BERTHO; VEIGA, 2019). Ao estudar o perfil epidemiológico de vítimas e motoristas infratores em Campinas/SP, Mauro (2001) verificou que 76,1% do total de condutores nas ocorrências de acidentes de trânsito dos quatro distritos policiais estudados em Campinas no ano de 1996 eram do sexo masculino, assim como 64,3% do total de pessoas envolvidas nos acidentes. Por outro lado, 16,3% de mulheres autora e condutora estiveram envolvidas nos acidentes e do total de envolvidas, condutoras e não condutoras, 22,4% eram mulheres (13,3% não informaram). Nesta pesquisa, verificamos que, do total de 8.375 pessoas envolvidas, 5.447 (65,0%) eram condutores, 2.679 (32,0%) passageiros, 170 (2,0%) pedestres, 76 (0,9%) testemunhas e 3 cavaleiros. Sendo que 4.724 condutores (86,7%) eram do sexo masculino e 309 (5,7%) do sexo feminino, 414 (7,6%) não foram identificados. Considerando os condutores mortos segundo o sexo, o total foi de 301, sendo 96% homens e 4% mulheres. No que tange à faixa etária desses condutores que foram a óbito no local do acidente, 13,7% tinham de 16 a 25 anos, 78,1% deles 26 a 59 anos e 8,2% tinham 60 anos ou mais.

4.2 Séries temporais dos acidentes

4.2.1 Análise exploratória

A análise exploratória das séries de acidentes foi realizada de acordo com as seguintes etapas: mensuração das estatísticas descritivas, suposição de normalidade, avaliação da estacionariedade e decomposição das séries.

Tabela 4.1: Análise descritiva das séries temporais diárias das variáveis de acidentes na Bahia, de 1º de janeiro de 2007 a 31 de dezembro de 2021

	Pessoas	Mortos	Feridos leves	Feridos graves	Feridos	Ilesos	Ignorados	Veículos
Média	45,14	1,77	10,68	3,72	14,40	26,69	2,65	32,33
Desvio padrão	24,91	1,95	6,49	4,37	8,96	18,23	2,63	17,51
CV*	55,20	109,75	60,77	117,53	62,23	68,32	116,57	54,18
Soma	247.340	9.721	58.530	20.239	78.919	146.235	12.460	177.144
Assimetria	1,36	2,68	1,73	25,87	6,15	1,17	10,46	0,84
Curtose	9,35	25,33	9,67	1.267,90	155,16	5,08	293,96	4,08
Mín (Máx)	1 (354)	0 (36)	0 (67)	0 (230)	0 (297)	0 (143)	0 (94)	1 (140)

*Coeficiente de variação (%).

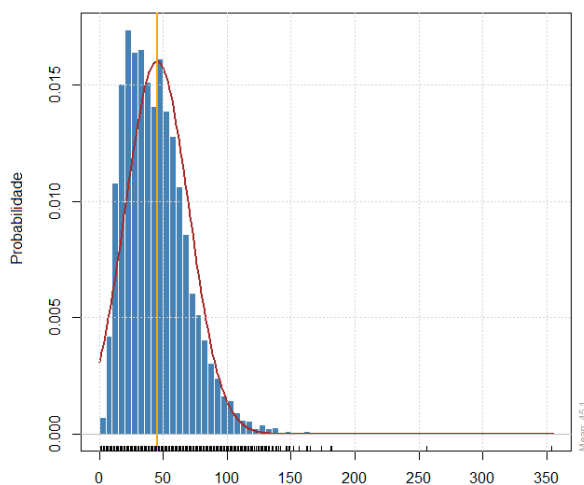
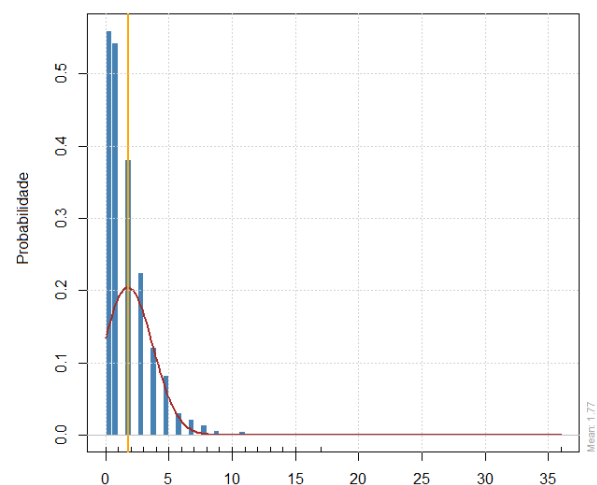
As estatísticas descritivas permitem conhecer o comportamento dos acidentes. Analisando a Tabela 4.2, o $CV > 15\%$ indicou que as séries possuem distribuição não-homogêneas

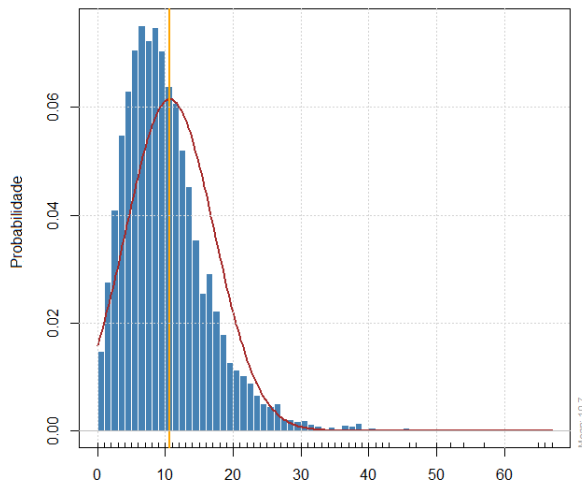
dos dados. A série *Feridos graves* apresenta maior variabilidade relativa em torno da média, expressando maior dispersão dos dados (Figura 4.9(d)). Nesse sentido, com base na distribuição de frequências, observamos que as oito séries apresentam assimetria à esquerda ($As > 0$). Sendo que, a série *Feridos graves* mostra-se com maior assimetria, denotando concentração de valores abaixo da média do período. Além disso, a análise da curtose indica que as séries apresentam uma distribuição leptocúrtica (afunilada). Esses resultados sinalizam não normalidade dos dados.

Tabela 4.2: Resultados do Teste Jarque-Bera, Bahia, 2007-2021

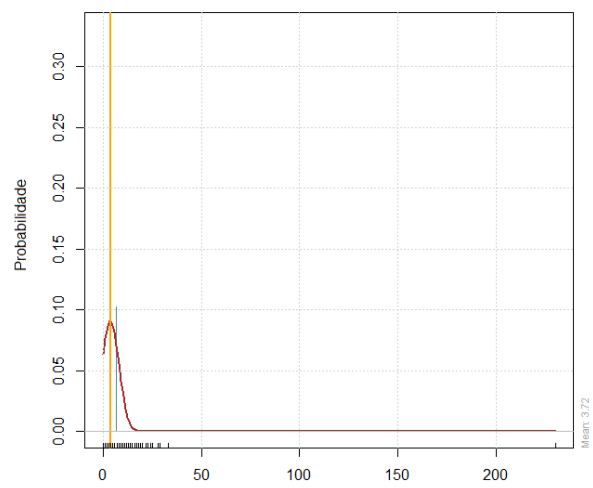
Variável	Estatística χ^2	p-valor
Pessoas	$1,0880 \times 10^4$	$2,0 \times 10^{-16}$
Mortos	$1,1630 \times 10^5$	$2,0 \times 10^{-16}$
Feridos leves	$1,3033 \times 10^4$	$2,0 \times 10^{-16}$
Feridos graves	$3,9198 \times 10^8$	$2,0 \times 10^{-16}$
Feridos	$5,3202 \times 10^6$	$2,0 \times 10^{-16}$
Ilesos	$2,1286 \times 10^3$	$2,0 \times 10^{-16}$
Ignorados	$1,9426 \times 10^7$	$2,0 \times 10^{-16}$
Veículos	$9,1736 \times 10^2$	$2,0 \times 10^{-16}$

Os resultados do teste Jarque-Bera mostram que os dados das oito séries não estão distribuídos normalmente, visto que a hipótese nula com o nível de 5% de significância ($p\text{-valor} < 0,05$) foi rejeitada. A não normalidade dos dados pode ser confirmada também por meio do histograma das séries ao serem comparados com a curva normal sendo esboçada nos gráficos que seguem:

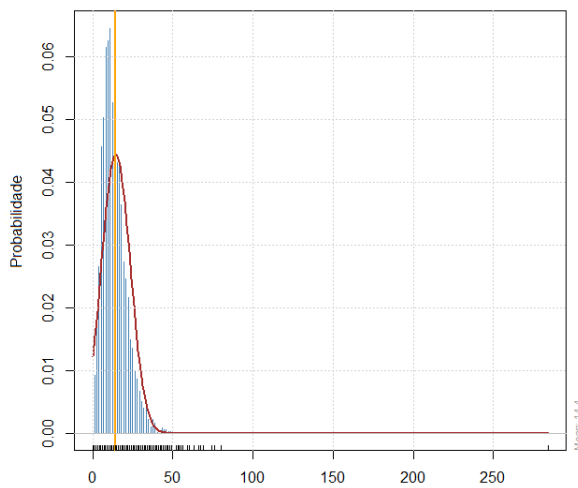
(a) *Pessoas*(b) *Mortos*



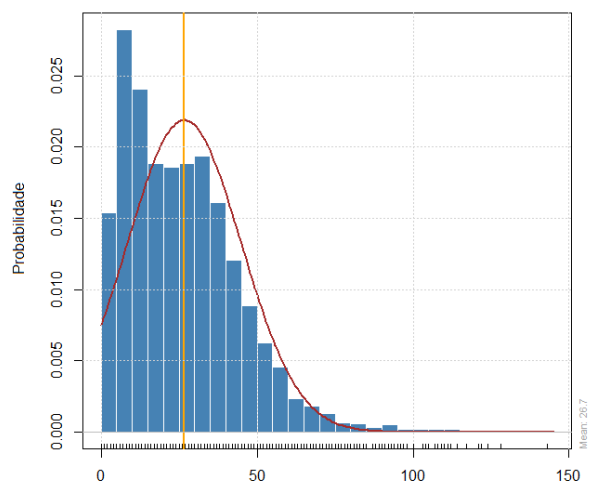
(c) *Feridos leves*



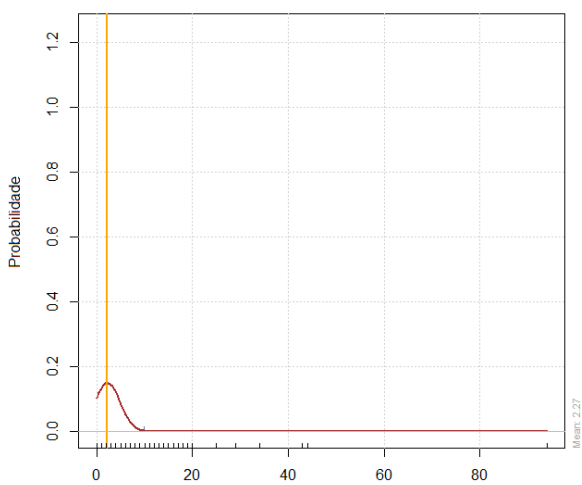
(d) *Feridos graves*



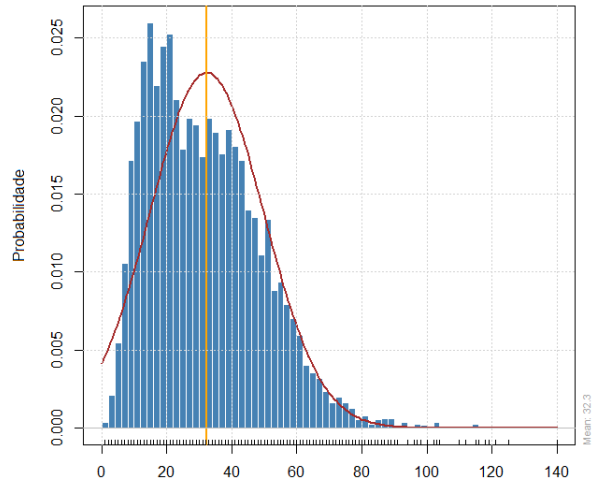
(e) *Feridos*



(f) *Ilesos*



(g) *Ignorados*



(h) *Veículos*

Figura 4.9: Histogramas das séries temporais de acidentes, Bahia, 2007-2021.

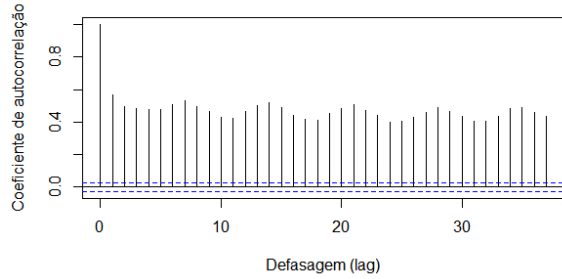
Para investigar a suposição de estacionariedade calculamos o teste de Dickey-Fuller Aumentado (ADF) a função de autocorrelação, assim é possível mensurar a interdependência das séries temporais (GUEDES, 2014). Nos resultados do ADF, rejeita-se a hipótese de raiz unitária ao nível de significância de 5% ($p\text{-valor} < 0,05$), demonstrando que as oito séries apresentam comportamento não estacionário, conforme mostra a Tabela 4.3:

Tabela 4.3: Resultados do ADF, Bahia, 2007-2021

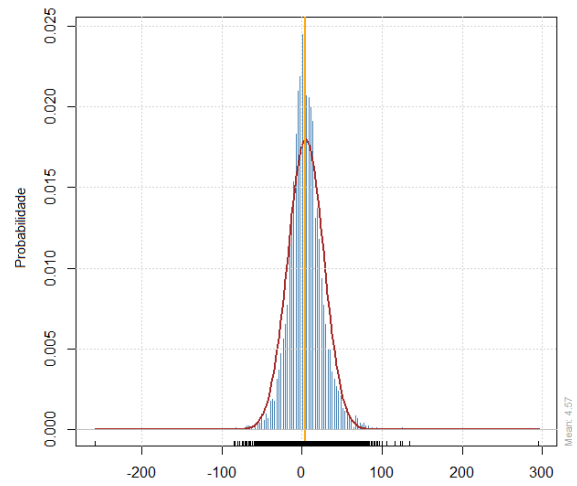
Variável	Estatística- t	p-valor
Pessoas	-1,704 x 10	2,0 x 10 ⁻¹⁶
Mortos	-4,245 x 10	2,0 x 10 ⁻¹⁶
Feridos leves	-2,626 x 10	2,0 x 10 ⁻¹⁶
Feridos graves	-4,541 x 10	2,0 x 10 ⁻¹⁶
Feridos	-2,688 x 10	2,0 x 10 ⁻¹⁶
Ilesos	-1,911 x 10	2,0 x 10 ⁻¹⁶
Ignorados	-3,893 x 10	2,0 x 10 ⁻¹⁶
Veículos	-1,435 x 10	2,0 x 10 ⁻¹⁶

A Figura 4.10 exibe comportamento não estacionário para todas as oito séries temporais. Conforme mostram os correlogramas, os valores decaem lentamente para zero. Isso decorre da tendência na qual uma observação de um lado da média tende a ser seguida por um grande número de observações do mesmo lado (EHLERS, 2007). No entanto, para Filho (2009), a FAC é adequada para detectar memória de séries estacionárias. Nesse sentido, utilizaremos a análise de resíduos para verificar se os coeficientes r_k são estatisticamente significativos. O histograma dos resíduos mostra se seguem (ou não) uma distribuição normal, sendo diferentes de zero (3.8), concluiremos que a série temporal é não estacionária.

Nos correlogramas, o eixo vertical indica a autocorrelação e o horizontal a defasagem. A linha tracejada azul indica onde é significativamente diferente de zero. Como mostra a Figura 4.10, praticamente todos os valores ACF estão fora do limite da linha tracejada azul indicando uma autocorrelação diferente de zero, ou seja, ausência de estacionariedade nas séries.

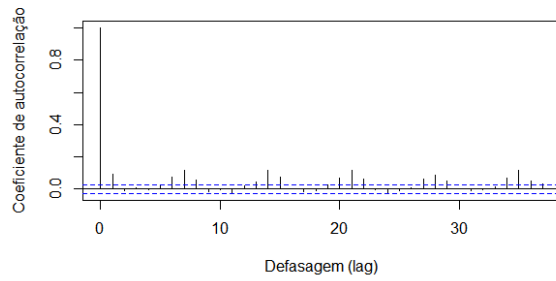


(a) FAC estimada *Pessoas*

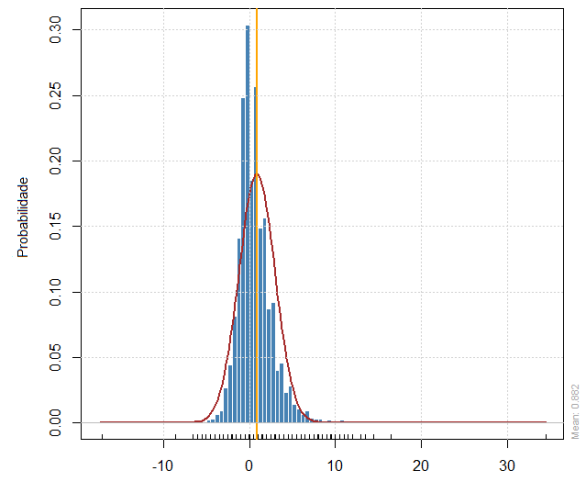


(b) Histograma dos resíduos *Pessoas*

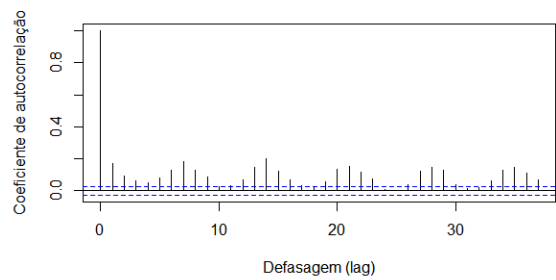
Figura 4.10: Correlogramas das séries temporais e histogramas dos resíduos das séries de acidentes, Bahia, 2007-2021.



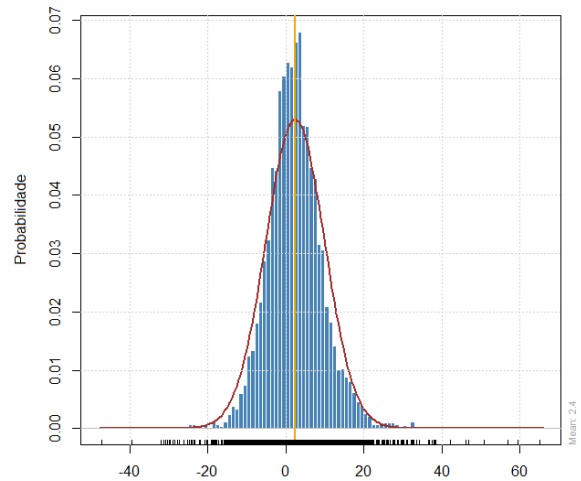
(a) FAC estimada *Mortos*



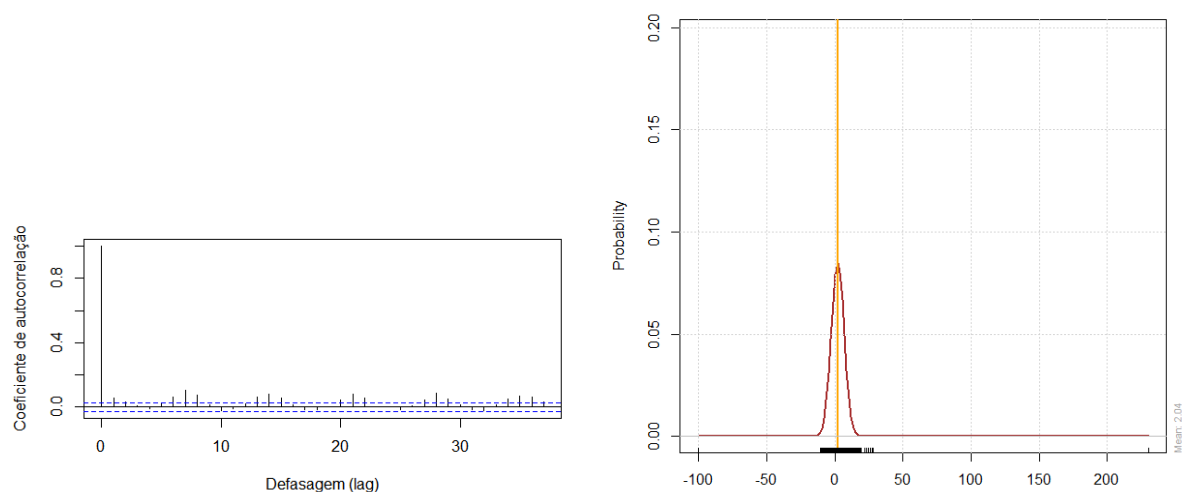
(b) Histograma dos resíduos *Mortos*



(c) FAC estimada *Feridos leves*



(d) Histograma dos resíduos *Feridos leves*

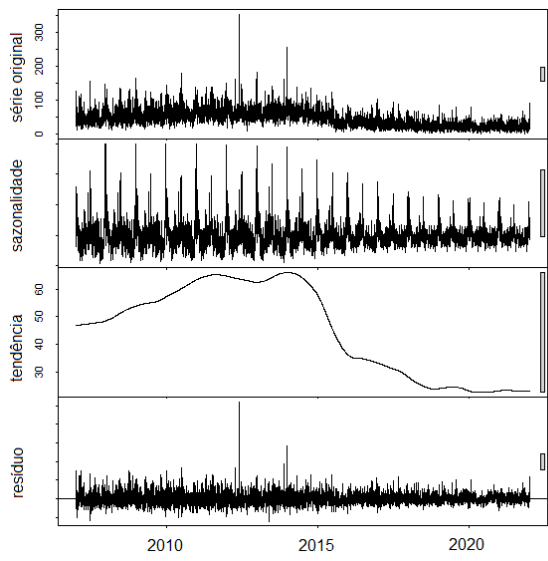
(e) FAC estimada *Feridos leves*(f) Histograma dos resíduos *Feridos graves*

A análise de resíduos exibiu valores para r_k conforme a Tabela 4.4. Neste caso, o p-valor $< 0,05$ implica dizer que as oito séries de acidentes são não estacionárias.

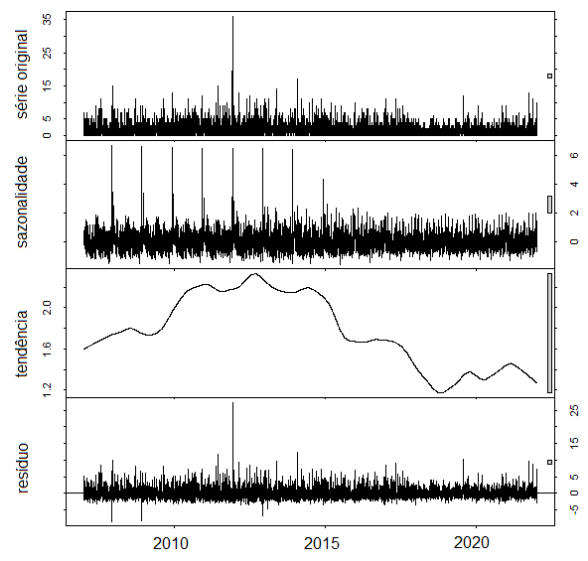
Tabela 4.4: Resultados do Teste Ljung-Box, Bahia, 2007-2021

Variável	Estatística χ^2	p-valor
Pessoas	$8,8515 \times 10^2$	$2,0 \times 10^{-16}$
Mortos	$5,9393 \times 10^2$	$2,0 \times 10^{-16}$
Feridos leves	$1,0123 \times 10^3$	$2,0 \times 10^{-16}$
Feridos graves	$6,6512 \times 10^2$	$2,0 \times 10^{-16}$
Feridos	$1,0026 \times 10^3$	$2,0 \times 10^{-16}$
Ilesos	$1,0309 \times 10^3$	$2,0 \times 10^{-16}$
Ignorados	$5,1668 \times 10^2$	$2,0 \times 10^{-16}$
Veículos	$8,8865 \times 10^2$	$2,0 \times 10^{-16}$

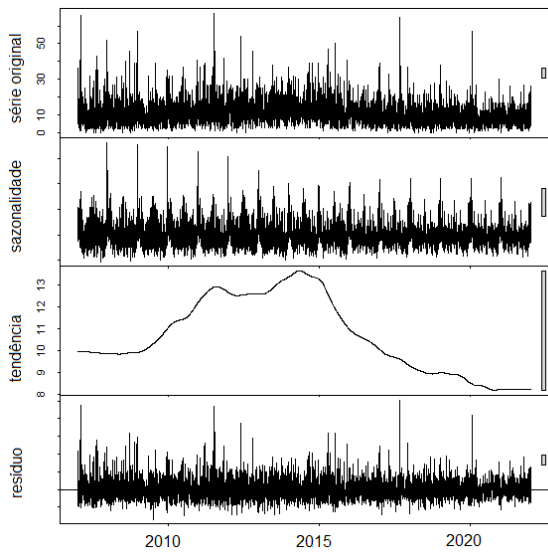
A decomposição das séries exibe os quatro movimentos principais característicos das séries temporais: Tendência, Ciclo, Sazonalidade, e Aleatoriedade. A Figura 4.11 mostra a decomposição das séries através do método multiplicativo (3.2), visto que os resíduos são correlacionados, construída no *software* RStudio. A componente periódica corresponde a um intervalo de 12 meses. De modo geral, os primeiros dias de janeiro concentram o maior volume de acidentes. Nota-se também, para todas as séries, uma queda expressiva da tendência a partir de 2015, exceto para *Feridos graves* e *Feridos leves*, cujo comportamento pode ser observado a partir de 2012, depois da obrigatoriedade do uso do sistema *ABS* a partir de 2011 e do uso de *airbag*, além da implementação compulsória desse equipamento em veículos novos desde 2009.



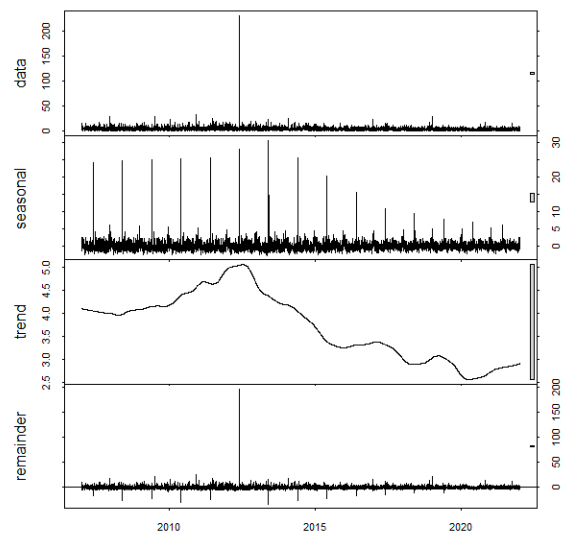
(g) Série *Pessoas*



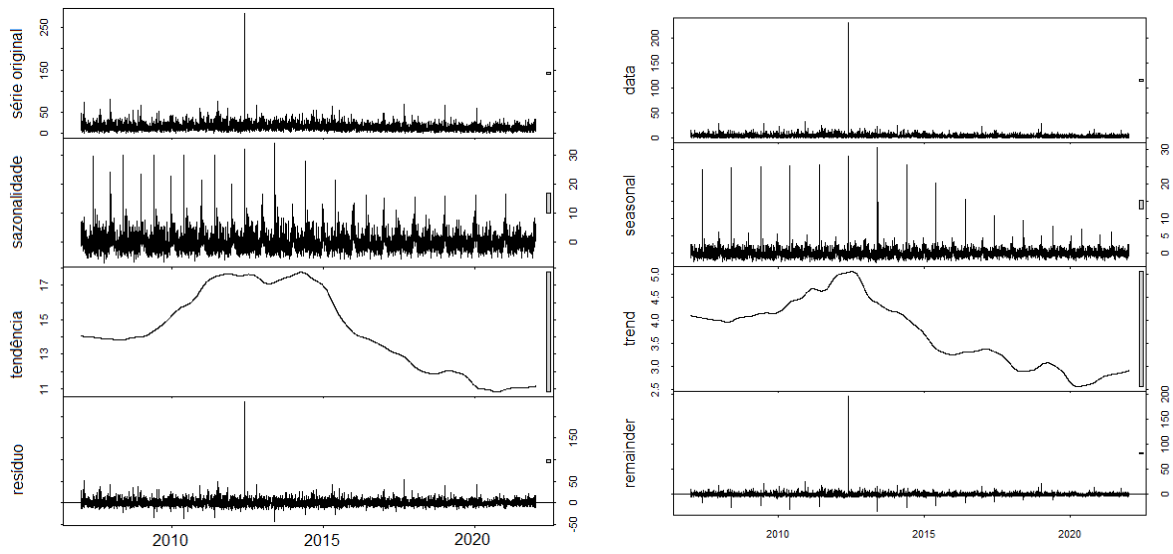
(h) Série *Mortos*



(i) Série *Feridos leves*

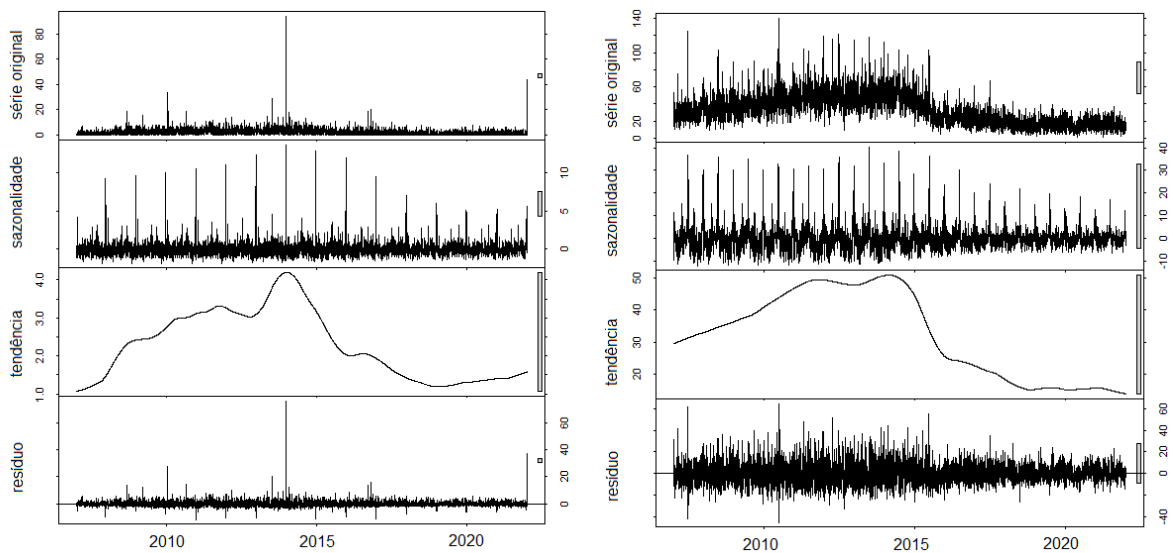


(j) Série *Feridos graves*



(k) Série *Feridos*

(l) Série *Ilesos*



(m) Série *Ignorados*

(n) Série *Veículos*

Figura 4.11: Decomposição das séries temporais de acidentes

As séries originais estão representadas como *data* na Figura 4.11. As séries exibem eventos específicos, grande volume de vítimas fatais em um único acidente e redução dos volumes a partir de 2015. Nota-se que, em 2015 as séries *Pessoas*, *Ilesos* e *Veículos* apresentam queda expressiva, de mais de 30% cada uma.

Diversos trabalhos que analisam séries temporais de acidentes de trânsito apresentam aspectos relevantes sobre essa temática no entanto, na maioria deles os autores se restringem a fazer uma análise descritiva dos dados, ou, no máximo, estudar a presença (ou não) de memória de longo alcance nos dados de acidentes. Citamos como exemplo, o trabalho de

Jianjun e Huijun (2014), eles investigaram a complexidade do tráfego misto com bloqueio induzido por acidente aplicando o DFA e verificaram que de acordo com o expoente de escala, o fluxo de tráfego pode exibir características de correlação de longo alcance, correlação de curto alcance ou ausência de relação de lei de potência. Nas características de correlação de longo alcance, que é fortemente dependente da probabilidade de entrada do veículo, a relação de velocidade lenta do veículo e o tempo de duração do bloqueio. Outro trabalho interessante com aplicação do DFA foi realizado por Remy (2011) ao monitorar a autossimilaridade estatística de canais fisiológicos, ambientais e veiculares, por meio da aplicação do DFA. O pesquisador concluiu que através da inspeção das propriedades de auto-similaridade, é possível gerar uma estratégia de previsão viável.

Aqui, aplicamos o DFA para verificar a autocorrelação das séries de acidentes e o ρ_{DCCA} para mensurar o nível de correlação de longo alcance existente entre elas. As análises foram construídas a partir dos resultados acumulados (2007-2021), dos resultados ano-a-ano e, finalmente, comparando a variação relativa desses resultados entre os anos. Em seguida, analisamos a correlação cruzada entre as variáveis oito variáveis estudadas.

4.2.2 Resultados 2007-2021

4.2.2.1 Autocorrelação das séries

O expoente obtido através do método DFA permite 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 (FILHO, 2009). Os resultados estão exibidos na Tabela 4.5 e nas Figuras 4.12 e 4.19.

A interpretação do expoente DFA para as oito séries estudadas considera que:

- se $\alpha_{DFA} < 0,5$, os acidentes exibem um comportamento antipersistentes (altas seguidas de baixas ou baixas seguidas de altas);
- se $\alpha_{DFA} = 0,5$, os acidentes apresentam um comportamento sem memória;
- se $\alpha_{DFA} > 0,5$, os acidentes de maneira persistentes (altas seguidas de altas ou baixas seguidas de baixas).

De modo geral, observamos um comportamento persistente dos dados nas oito séries estudadas. O total de *Veículos*, *Ilesos* e *Pessoas* exibiram α_{DFA} mais elevados, 0,84, 0,81

e 0,78, respectivamente. Em seguida, *Ignorados*, *Feridos leves* e *Feridos* registraram 0,67, 0,62 e 0,60, respectivamente. Os menores valores de α_{DFA} foram de *Mortos*, 0,55 e *Feridos graves* 0,52.

Tabela 4.5: Expoente de autocorrelação, Bahia, 2007-2021

Variável	α_{DFA}	Erro
Pessoas	0,78	0,02
Mortos	0,55	0,01
Feridos leves	0,62	0,01
Feridos graves	0,52	0,01
Feridos	0,60	0,01
Ilesos	0,81	0,02
Ignorados	0,67	0,02
Veículos	0,84	0,02

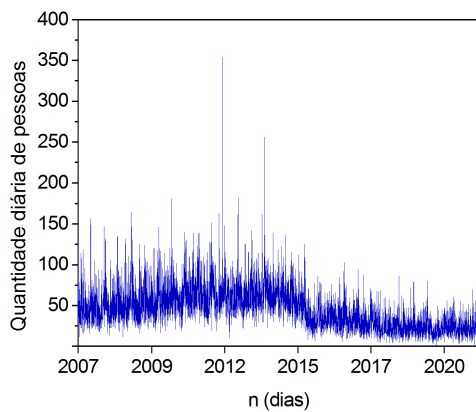
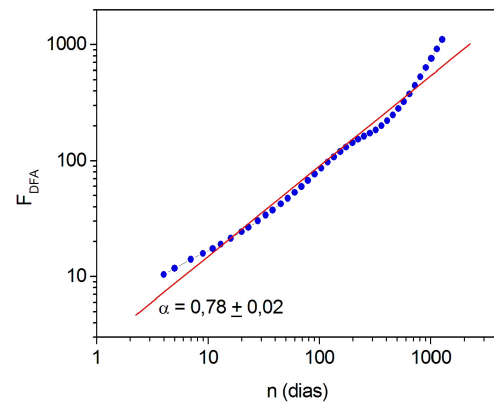
(a) Série *Pessoas*(b) *DFA Pessoas*

Figura 4.12: Bahia: Quantidade de pessoas envolvidas em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de pessoas envolvidas em acidentes, 2007-2021 (b)

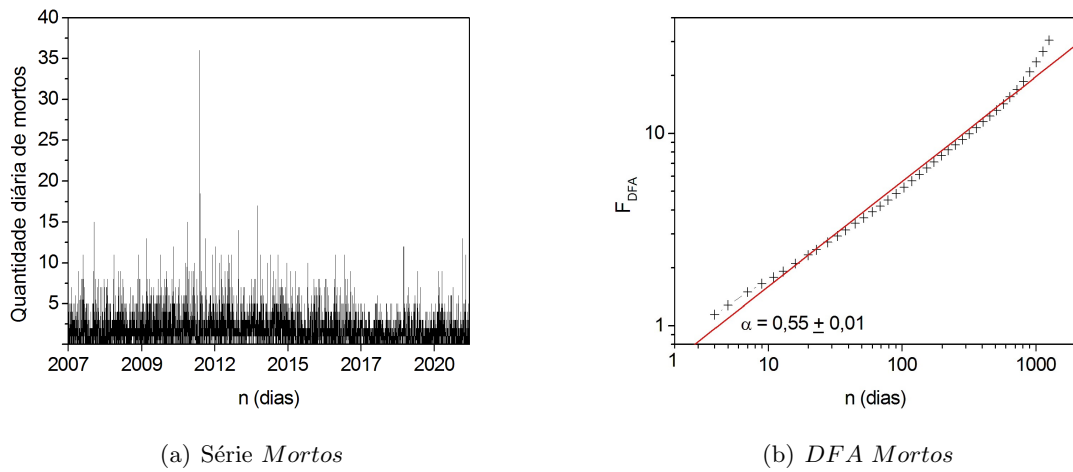


Figura 4.13: Bahia: Quantidade de vítimas fatais em acidentes, 2007-2021 (a). F_{DFA} em função de n no total de vítimas fatais em acidentes, 2007-2021 (b)

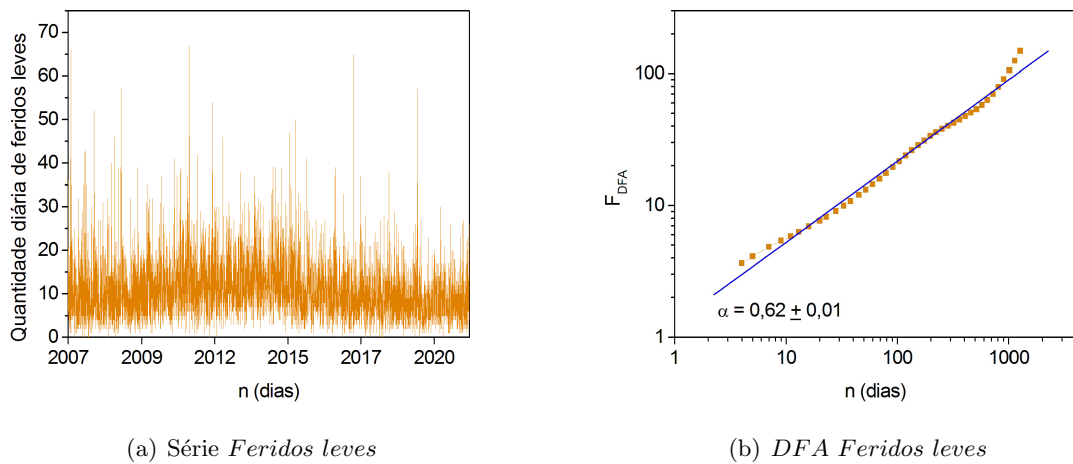


Figura 4.14: Bahia: Quantidade de feridos leves em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de feridos leves em acidentes, 2007-2021 (b)

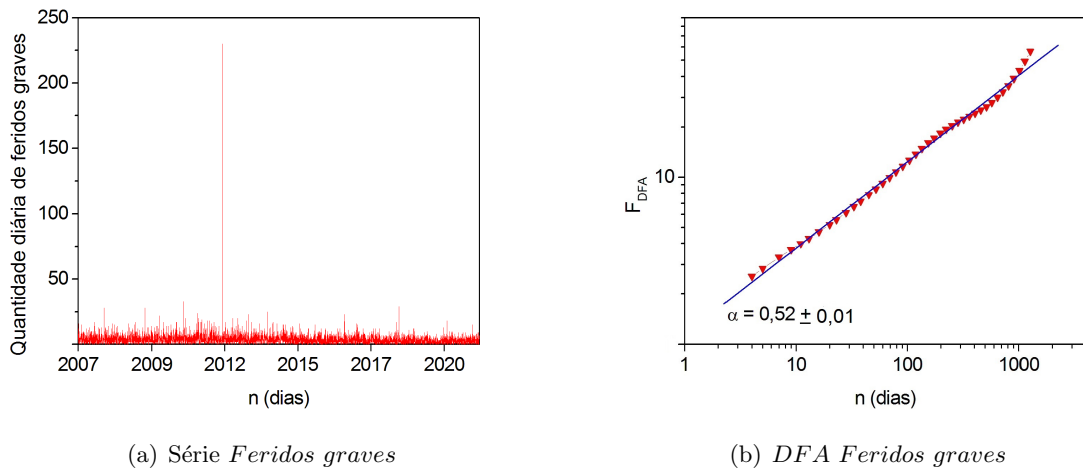


Figura 4.15: Bahia: Quantidade de feridos graves em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de feridos graves em acidentes, 2007-2021 (b)

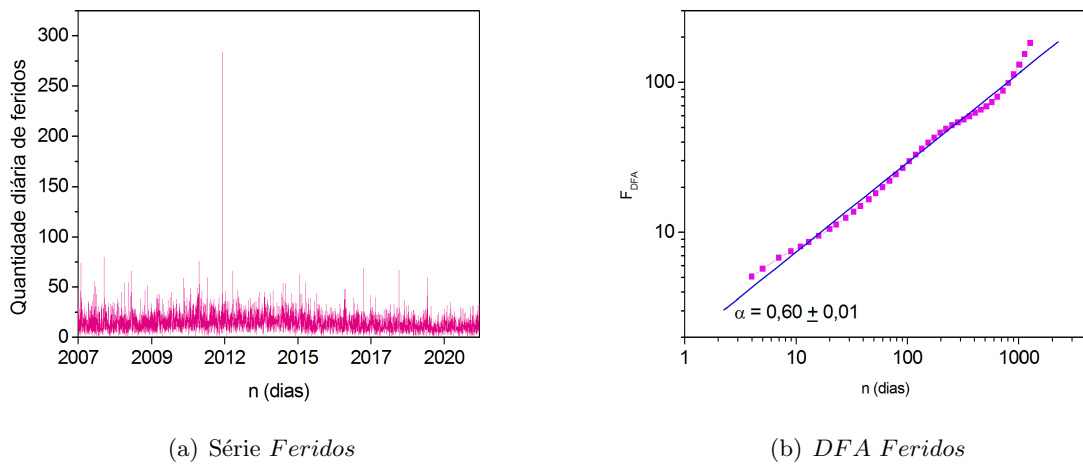


Figura 4.16: Bahia: Quantidade diária de feridos em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de feridos em acidentes, 2007-2021 (b)

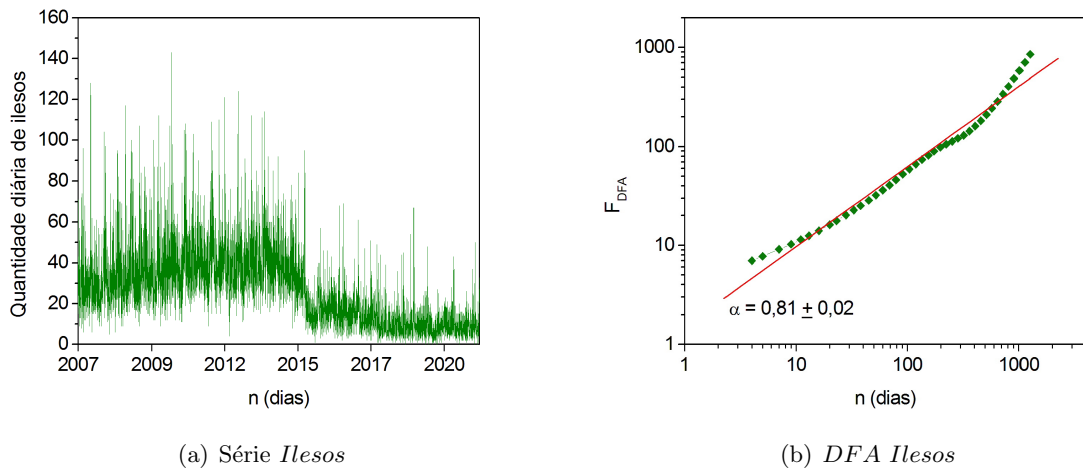


Figura 4.17: Bahia: Quantidade diária de ilesos em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de ilesos em acidentes, 2007-2021 (b)

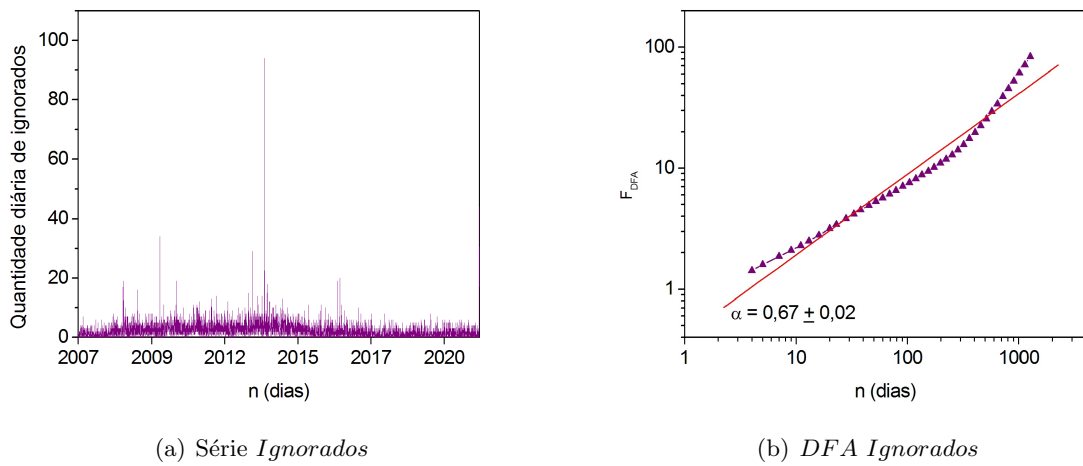


Figura 4.18: Bahia: Quantidade diária de ignorados em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de ignorados em acidentes, 2007-2021 (b)

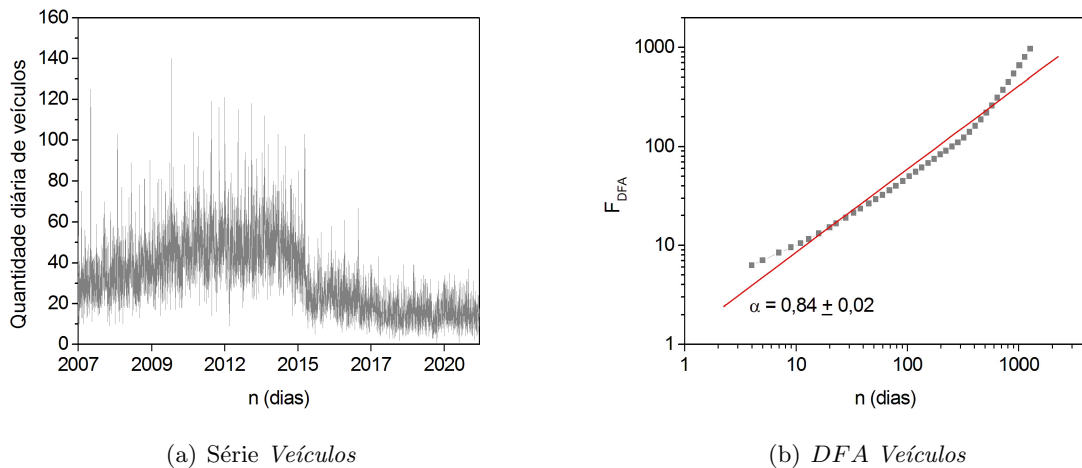


Figura 4.19: Bahia: Quantidade diária de veículos envolvidos em acidentes, 2007-2021 (a); F_{DFA} em função de n no total de veículos envolvidos em acidentes, 2007-2021 (b)

A aplicação do DFA evidencia que as flutuações na ocorrência dos acidentes analisados são persistentes ($\alpha_{DFA} > 0,5$) para quase todas as séries. Assim, os valores altos tendem a serem seguidos por valores altos, e valores baixos tendem a serem seguidos por valores baixos.

Entretanto, o método DCCA e o coeficiente ρ_{DCCA} possibilitarão verificar se as variações são correlacionadas. A seguir apresentaremos o coeficiente de correlação cruzada que dá conta de mensurar o nível de correlação cruzada entre as séries.

4.2.2.2 Correlação cruzada entre as séries

Na análise da correlação cruzada entre as séries, investigamos a intensidade da correlação entre a variável *Mortos* e as demais variáveis. Isso porque, compreender a correlação a longo prazo dessas variáveis pode ser de grande importância para minimizar os impactos da violência dos acidentes, sobretudo da quantidade de mortos, prioridade essencial da PRF.

Para os dados acumulados, o coeficiente ρ_{DCCA} mostrou-se positivo para todas as séries, indicando a existência de correlação cruzada entre *mortos* e as demais variáveis, conforme mostra a Figura 4.2.2.2.

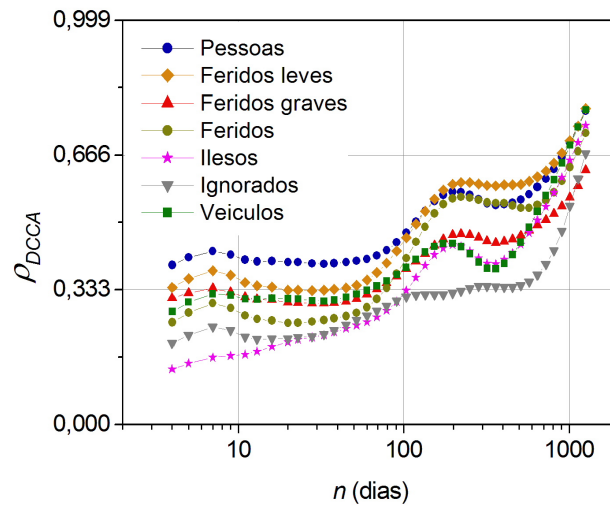


Figura 4.20: Correlação entre *Mortos* e as demais séries de acidentes reportados pela PRF, Bahia, 2007 a 2021.

O nível de correlação cruzada do total de *Mortos* com as demais séries, denota que em pequenas escalas de tempo ($n \leq 100$) quase todas as variáveis analisadas, exceto *Pessoas* e *Feridos*, apresentam níveis mais baixos de correlação ($0,000 < \rho_{DCCA} \leq 0,333$). Este resultado indica que as oscilações na quantidade de *Feridos leves*, *Feridos graves*, *Ilesos*, *Ignorados* e *Veículos* afetam com baixa intensidade a quantidade de *Mortos* até, aproximadamente, 100 dias. Por outro lado, os níveis de correlação aumentam em grandes escalas de tempo para todas as variáveis (exceto *Ignorados*), seguindo uma correlação média ($0,333 < \rho_{DCCA} \leq 0,666$) em uma escala de $100 < n \leq 1000$ dias, para alta ($0,666 < \rho_{DCCA} \leq 1,000$), $n \geq 1000$, sobretudo para *Pessoas* e *Feridos leves* e *Feridos*.

A Figura 4.21 apresenta graficamente a correlação cruzada da variável *Mortos* com cada uma das variáveis separadamente.

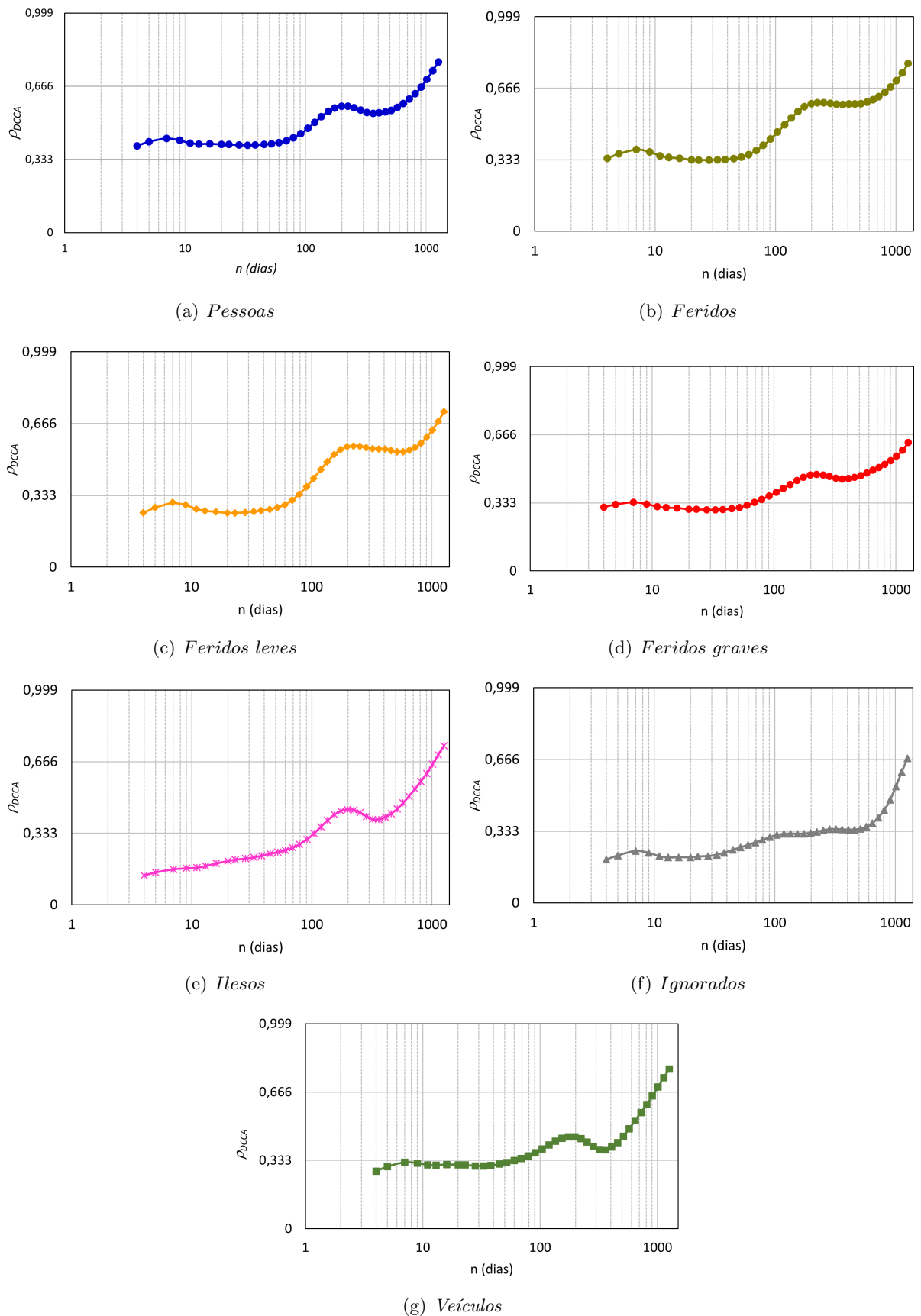


Figura 4.21: Correlação cruzada de *Mortos* com *Pessoas*, *Feridos leves*, *Feridos graves*, *Feridos*, *Ignorados* e *Veículos*, Bahia, 2007-2021.

De maneira mais específica, a Tabela 4.6 exibe os valores médios do ρ_{DCCA} ao longo de 2007-2021 nas correlações de *Mortos* e as demais variáveis para diferentes escalas.

Tabela 4.6: Mortos: valores médios do coeficiente ρ_{DCCA}

	$n \leq 30$	$30 < n \leq 90$	$90 < n \leq 365$	$n > 365$
Pessoas	0,41	0,40	0,54	0,63
Feridos leves	0,27	0,29	0,51	0,59
Feridos graves	0,31	0,32	0,44	0,52
Feridos	0,35	0,35	0,55	0,65
Ilesos	0,18	0,25	0,40	0,55
Ignorados	0,22	0,26	0,33	0,44
Veículos	0,31	0,33	0,41	0,56

Em síntese, podemos afirmar que as variáveis *Feridos graves*, *Ilesos*, *Ignorados* e *Veículos* apresentaram menor intensidade ($\rho_{DCCA} < 0,50$) nas correlações para pequenas escalas ($n \leq 90$). Enquanto *Pessoas*, *Feridos leves* e *Feridos* exibiram maior grau de intensidade nas correlações em grandes escalas de tempo ($n > 365$).

4.2.3 Resultados ano a ano

Com o intuito de analisarmos a correlação a partir da perspectiva anual, o mesmo intervalo temporal comumente adotado pela PRF para análise das ocorrências de acidentes, aplicamos o DFA nas oito séries, para $N = 365$ dias, a fim de identificar a existência (ou não) de estacionariedade nas séries de acidentes em períodos anuais.

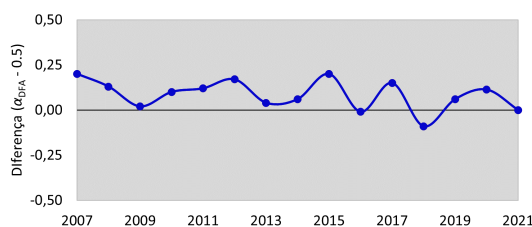
4.2.3.1 Autocorrelação das séries

Para uma melhor compreensão, apresentamos a Tabela 4.7 onde subtraímos 0,5 do expoente DFA de cada uma das séries, ano a ano. Assim, para os valores $(\alpha - 0,5) < 0$ as células aparecem em vermelho indicando antipersistência da série. Por outro lado, as células em azul, $(\alpha - 0,5) > 0$, representam um comportamento persistente da série. De modo geral, há mudança de sentido nas correlações, passando de correlacionado para anti-correlacionado e vice-versa, ou seja, nem sempre a série é persistente para um período de 365 dias ($N = 365$). Os valores de $(\alpha - 0,5) = 0$, denotam ausência de memória da série.

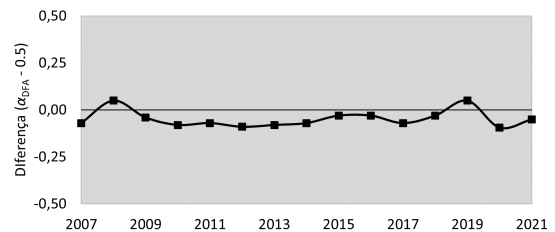
Tabela 4.7: Expoentes α_{DFA} subtraído de 0,5 ($\alpha_{DFA}-0,5$), 2007-2021

Ano	Pessoas	Mortos	Feridos leves	Feridos graves	Feridos	Ilesos	Ignorados	Veículos
2007	0,20	-0,07	-0,05	-0,07	-0,04	0,25	-0,07	0,25
2008	0,13	0,05	-0,05	0,04	-0,03	0,16	0,07	0,16
2009	0,02	-0,04	-0,06	-0,04	-0,07	0,06	-0,14	0,06
2010	0,10	-0,08	0,00	0,03	0,01	0,05	-0,07	0,15
2011	0,12	-0,07	-0,03	0,01	-0,02	0,16	0,07	0,13
2012	0,17	-0,09	0,08	0,03	0,07	0,26	-0,03	0,26
2013	0,04	-0,08	-0,01	-0,08	-0,03	0,10	-0,21	0,10
2014	0,06	-0,07	-0,02	0,08	0,00	0,05	-0,07	0,11
2015	0,20	-0,03	-0,02	-0,03	-0,03	0,29	0,02	0,28
2016	-0,01	-0,03	-0,01	-0,06	-0,01	-0,01	-0,02	0,08
2017	0,15	-0,07	0,02	-0,03	0,01	0,23	0,01	0,17
2018	-0,09	-0,03	-0,06	-0,16	-0,10	-0,08	0,04	0,01
2019	0,06	0,05	-0,03	0,03	0,01	0,07	-0,02	0,02
2020	0,11	-0,09	0,05	-0,08	0,02	0,08	0,02	0,14
2021	0,00	0,05	0,01	0,08	0,00	0,03	0,20	0,06

Este cenário pode ser representado graficamente pela Figura 4.22. Os valores acima de 0,5 estão representados pelos pontos localizados na parte superior do gráfico, nesses pontos a série apresenta comportamento persistente. Os gráficos mostram que apenas a série *Veiculos* é persistente em praticamente todo o período 2007-2021, assim como *Pessoas* e *Ilesos*. Por outro lado, *Feridos leves*, *Feridos graves*, *Feridos* e *Ignorados* exibem mudança de sentido nas correlações, passando de correlacionado para anti-correlacionado e vice-versa. A série *Mortos*, por sua vez, mostra-se quase sempre anti-persistente.



(a) *Pessoas*



(b) *Mortos*

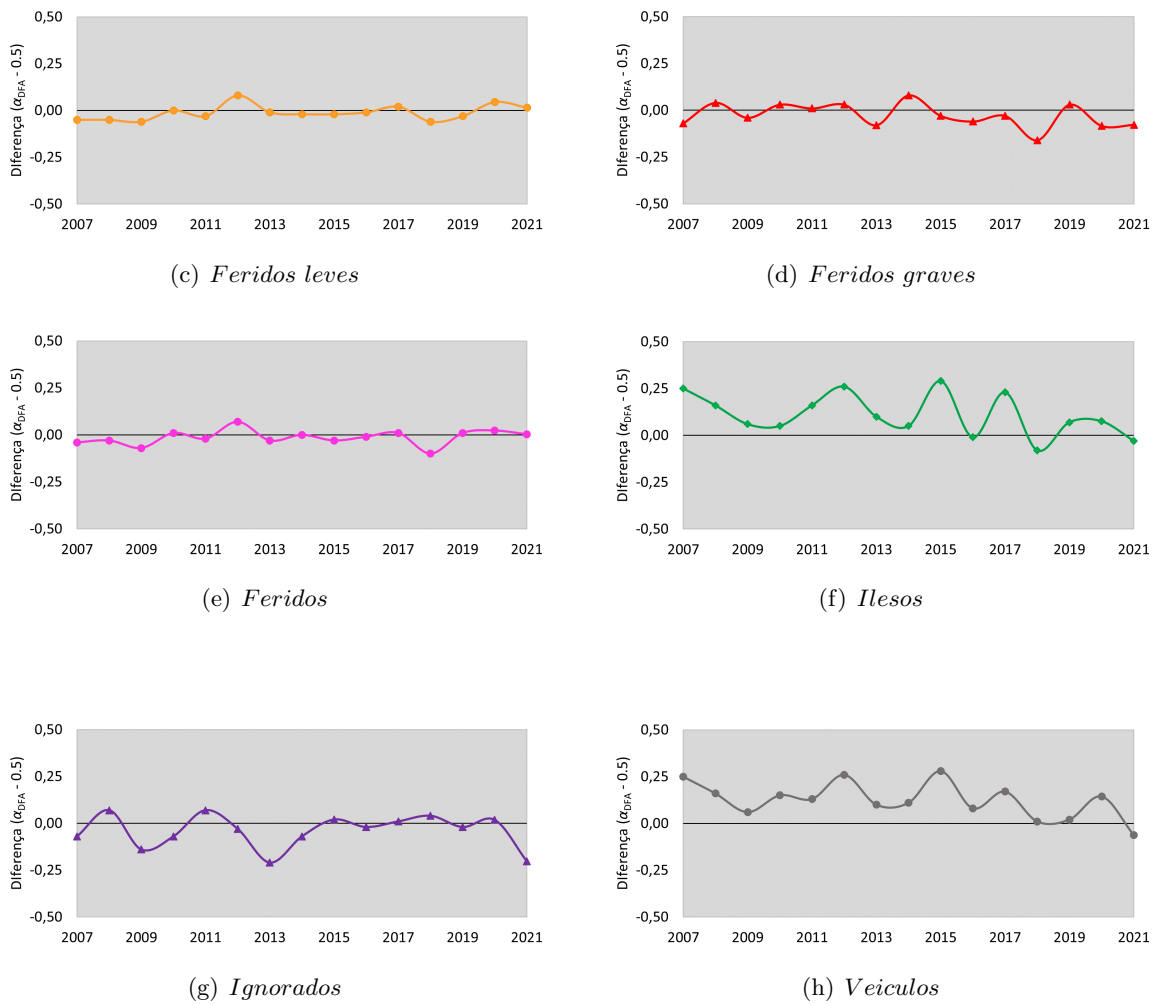


Figura 4.22: Variação anual do coeficiente de autocorrelação das séries temporais de acidentes, 2007-2021.

Nota-se, através da Figura 4.22, que as flutuações anuais nas séries de acidentes nem sempre apresentam persistência ($\alpha > 0,5$), isto é, não se pode afirmar que valores altos tendem a serem seguidos por valores altos, nem que valores baixos tendem a ser seguidos por valores baixos, em um período de 12 meses.

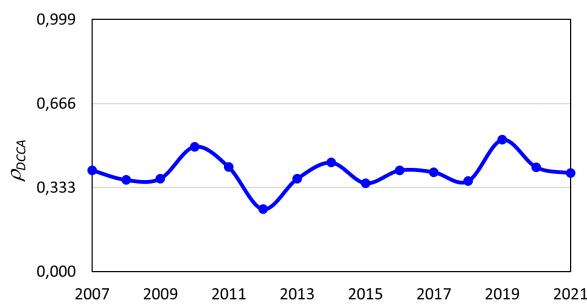
4.2.3.2 Correlação cruzada entre as séries

Na análise ano a ano, os valores médios do coeficiente ρ_{DCCA} apresentaram-se positivos para todos os anos, quando verificada a correlação entre *Mortos* e as demais séries (Tabela 4.8).

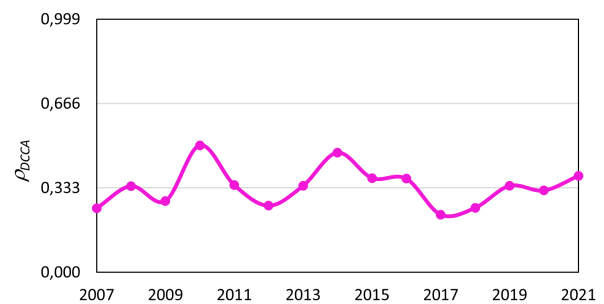
Tabela 4.8: Mortos: Valor médio por ano do ρ_{DCCA} , Bahia, 2007-2021

Ano	Pessoas	Feridos leves	Feridos graves	Feridos	Ilesos	Ignorados	Veículos
2007	0,40	0,18	0,32	0,25	0,29	0,10	0,36
2008	0,36	0,22	0,42	0,34	0,19	0,23	0,34
2009	0,37	0,23	0,24	0,28	0,23	0,14	0,29
2010	0,49	0,38	0,50	0,50	0,26	0,31	0,40
2011	0,41	0,24	0,39	0,34	0,20	0,35	0,32
2012	0,25	0,18	0,26	0,26	0,03	0,14	0,08
2013	0,37	0,22	0,43	0,34	0,21	0,21	0,26
2014	0,43	0,33	0,52	0,47	0,14	0,27	0,30
2015	0,35	0,34	0,29	0,37	0,14	0,27	0,30
2016	0,40	0,29	0,37	0,37	0,10	0,27	0,40
2017	0,39	0,18	0,20	0,23	0,23	0,27	0,31
2018	0,36	0,20	0,23	0,25	0,14	0,16	0,32
2019	0,52	0,24	0,39	0,34	0,36	0,34	0,44
2020	0,41	0,28	0,25	0,32	0,20	0,19	0,31
2021	0,58	0,29	0,41	0,38	0,38	0,30	0,45
Média	0,39	0,25	0,34	0,33	0,19	0,23	0,32

Apesar do ρ_{DCCA} apresentar valores menores ou iguais a 0,5 para praticamente todas as variáveis, nota-se uma correlação cruzada mais intensa da variável *Mortos* com *Pessoas*, *Feridos graves*, *Feridos* e *Veículos*, cujos valores médios em alguns anos chegaram a 0,39, 0,34, 0,33 e 0,32, respectivamente, ou seja, há uma correlação cruzada no nível médio ($0,333 < \rho_{DCCA} < 0,666$). Isto significa que, quantidades grandes (ou pequenas) de pessoas, feridos, feridos graves e veículos envolvidos nos acidentes, tendem a serem seguidas por quantidades grandes (ou pequenas, respectivamente) de vítimas fatais em acidentes de trânsito. No entanto, nesse panorama anual, o ρ_{DCCA} exibiu mudança nos níveis das correlações, conforme mostra a Figura 4.23.



(a) *Pessoas*



(b) *Feridos*

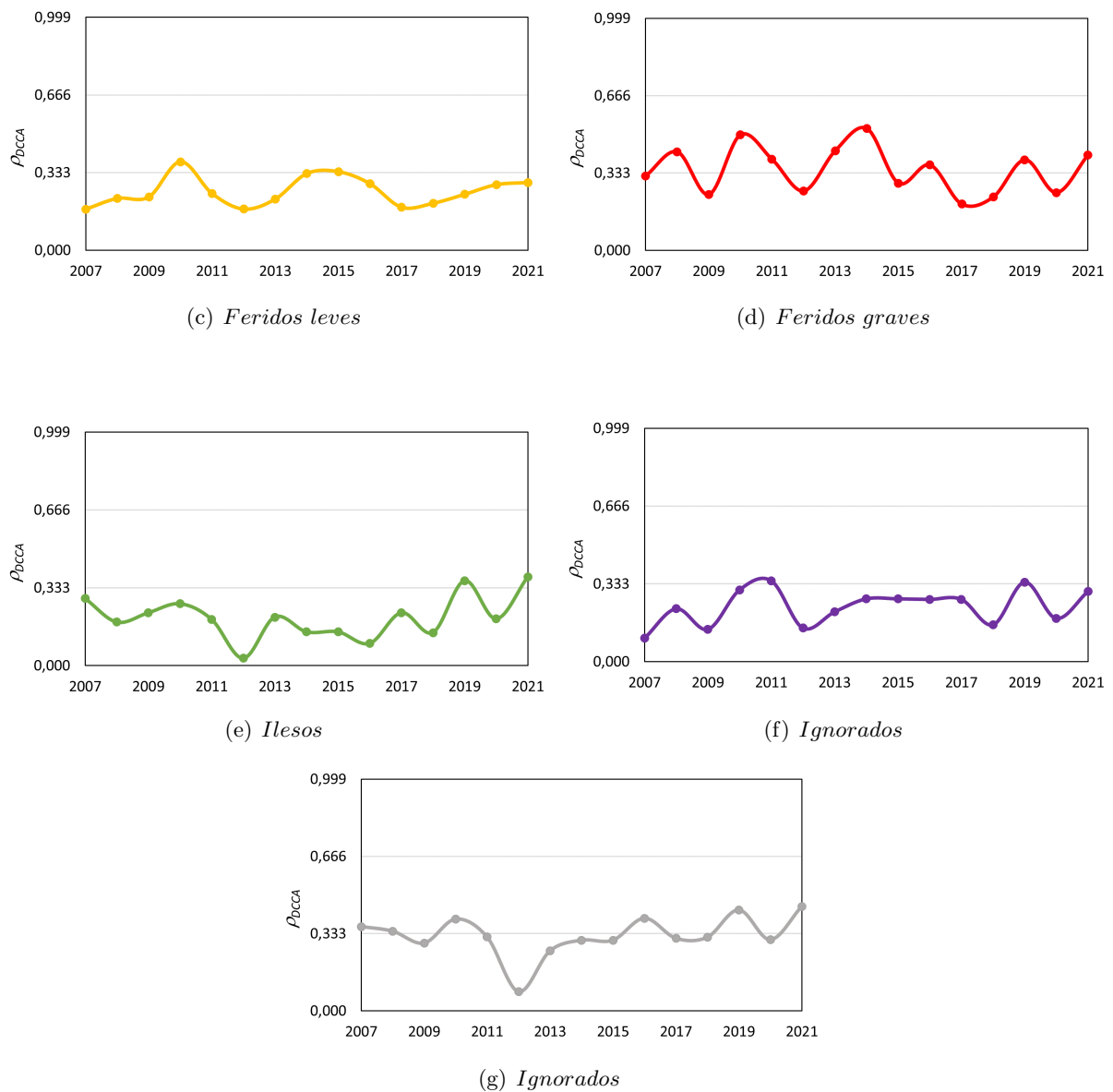
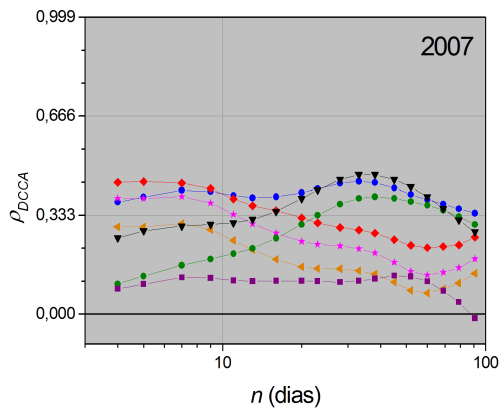


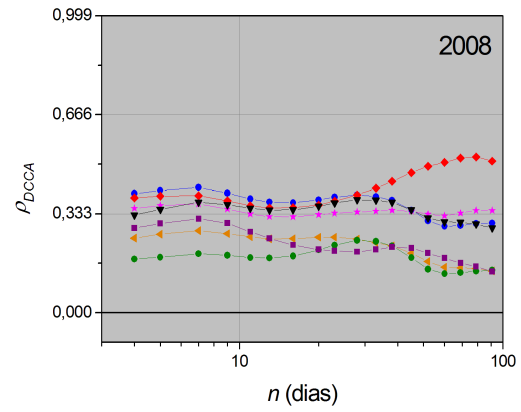
Figura 4.23: Variação anual do coeficiente de autocorrelação das séries temporais de acidentes, 2007-2021.

Com base na análise ano-a-ano da correlação cruzada entre *Mortos* e as demais séries, exibidos na Figura 4.24, verificam-se que as variáveis *Pessoas* e *Veículos* apresentam comportamentos semelhantes em todos os anos. Ambas exibem em suas flutuações uma variação no nível de correlação de média ($0,333 < \rho_{DCCA} \leq 0,666$) para fraca ($0,000 < \rho_{DCCA} \leq 0,333$) em quase todos os anos entre 2007 e 2016, com exceção de 2010, 2012 e 2014, que seguem nível médio para alto ($0,666 < \rho_{DCCA} < 1,000$) para $n > 45$, bem como no período de 2017 até 2020. Importante destacar que nos anos 2013 e 2015, *Pessoas* e *Veículos* sofreram uma intensa declinação no valor do ρ_{DCCA} . Já em 2012, esse expoente exibiu flutuação com forte declinação desde $n = 10$, voltando a crescer novamente a partir de $n = 50$. Neste mesmo ano, *Veículos* e *Ilesos* evidenciaram uma anticorrelação, para $25 < n < 80$. A variável *Feridos graves* exibiu mudança nos níveis de correlação, saindo de

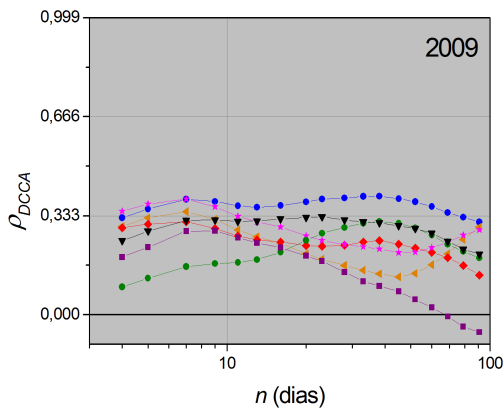
média para forte em 2014 a partir de $n > 30$, no entanto, na maioria dos anos, oscilou entre os níveis de correlação médio e fraco. As variáveis *Feridos leves* e *Ilesos*, se mantiveram com uma correlação fraca durante quase todos os períodos. Sendo que em 2010, *Ilesos* apresentou oscilação para um nível médio de correlação para $n > 40$. *Ignorados*, por sua vez, exibiu correlação fraca em quase todos os anos analisados.



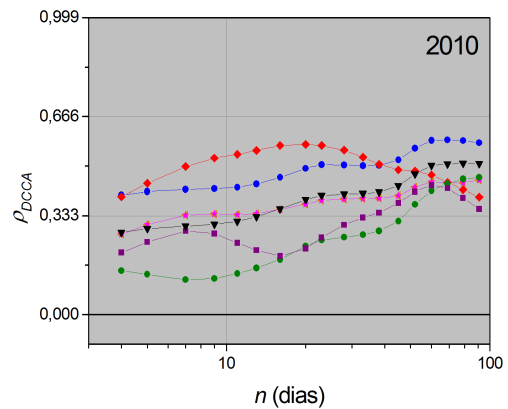
(a)



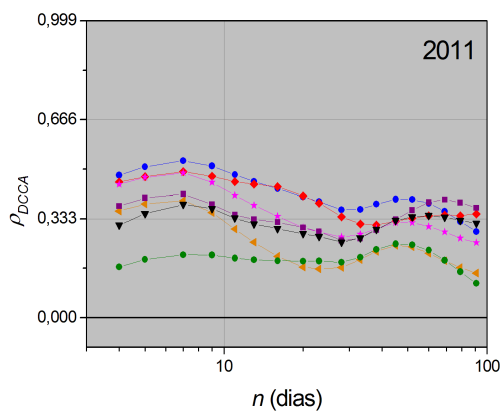
(b)



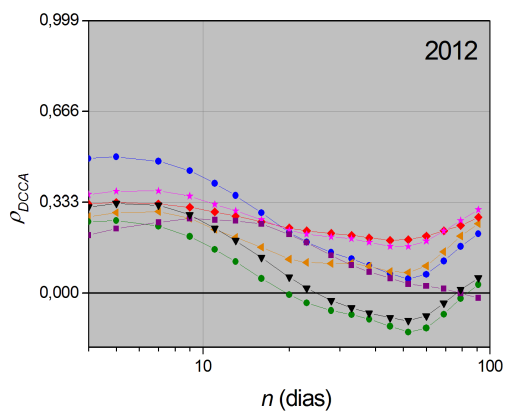
(c)



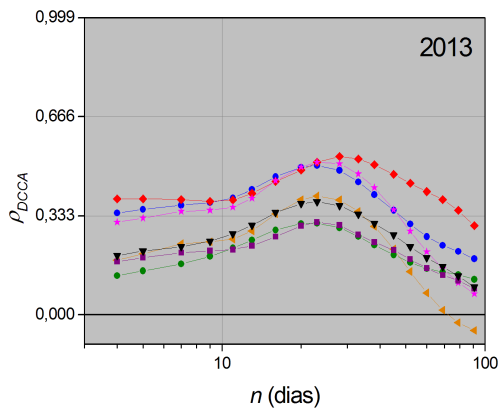
(d)



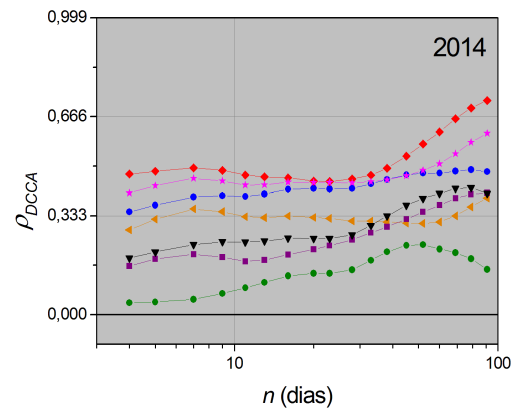
(e)



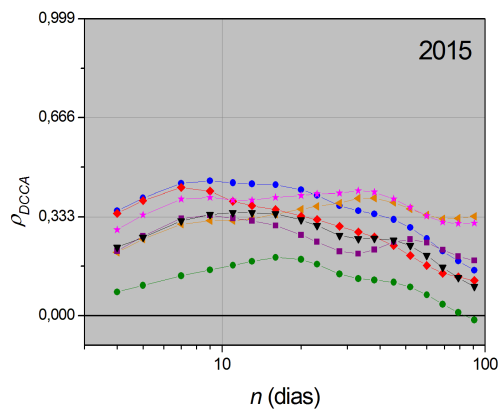
(f)



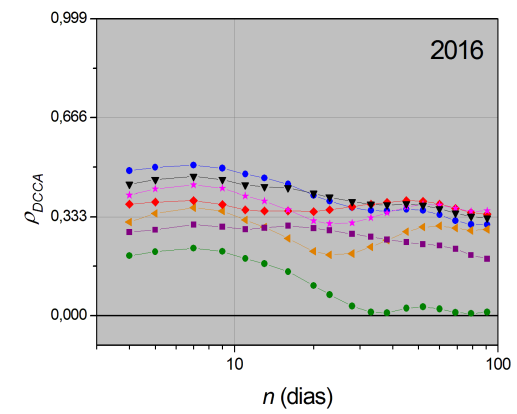
(g)



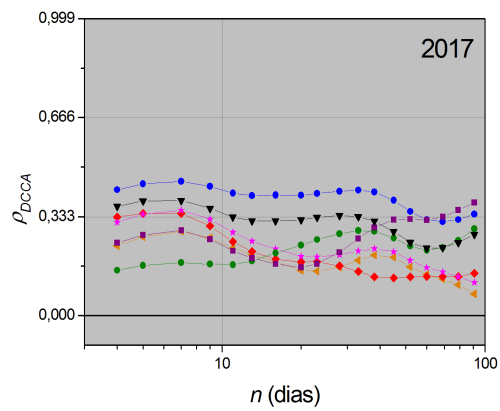
(h)



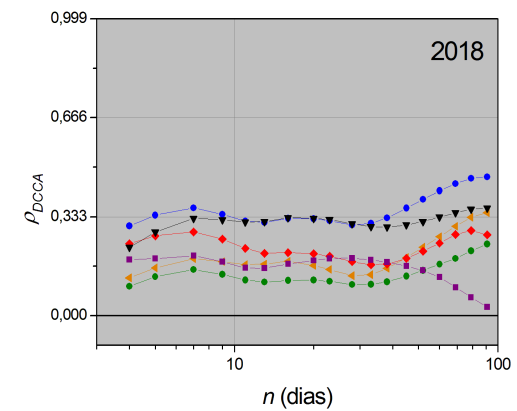
(i)



(j)



(k)



(l)

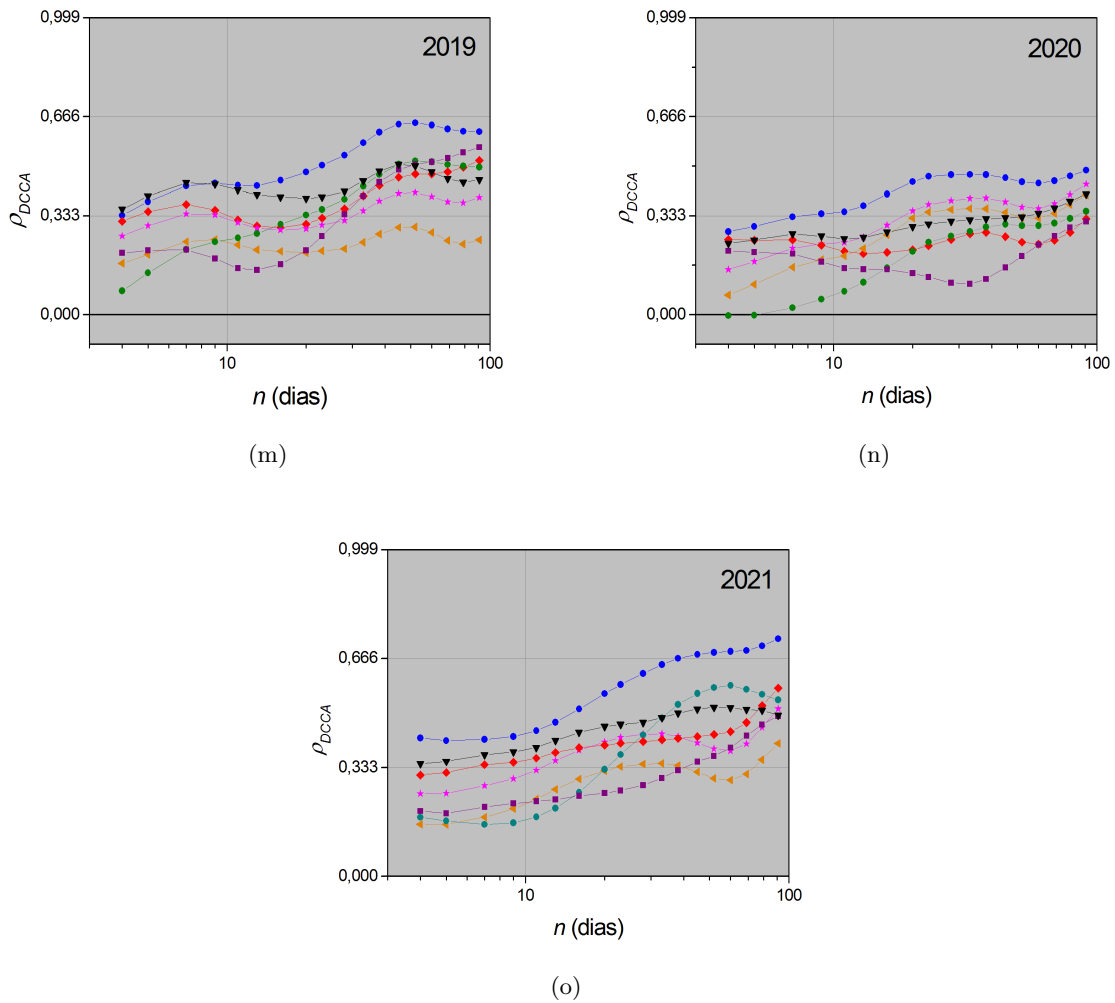


Figura 4.24: Correlação cruzada de Mortos com *Pessoas* (●), *Feridos leves* (◀), *Feridos graves* (◆), *Feridos* (★), *Ilesos* (●), *Ignorados* (■) e *Veiculos* (▼), 2007 a 2021.

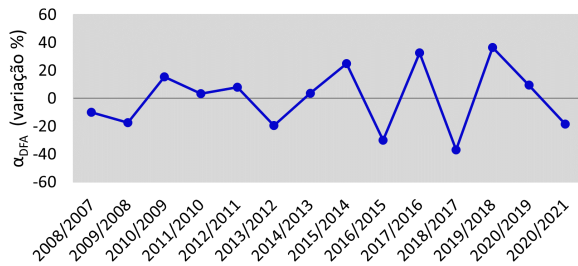
4.2.4 Variação relativa entre os anos

As análises neste cenário foram realizadas com o intuito de avaliar a autocorrelação e correlação cruzada de determinado ano com relação a do ano imediatamente anterior, possibilitando visualizar a variação dessas medidas entre -60% e 60%.

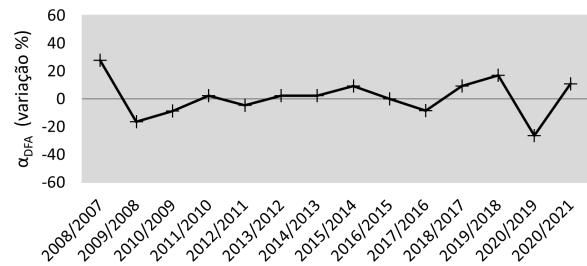
4.2.4.1 Autocorrelação das séries

Ao compararmos os resultados de cada ano com os do ano respectivamente anterior, verificamos que ao longo do período 2007-2021 o expoente DFA apresenta variação relativa para todas as séries, exibindo mudança de sentido ao passar de correlacionado para anti-

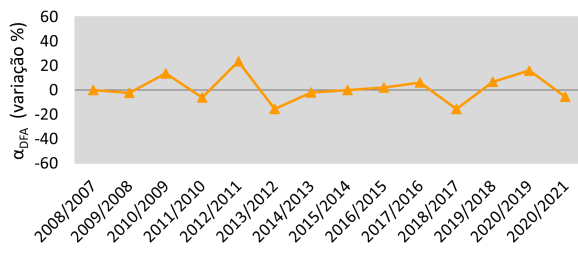
correlacionado e vice-versa. Na Figura 4.25 os que valores acima do zero indicam períodos em que a correlação é positiva, ao passo que valores abaixo do zero a correlação é negativa (anticorrelação). Isto significa que analisar a autocorrelação das séries de acidente em um período de 365 dias ou menos, pode não refletir claramente se há ou não existência de memória longa, inviabilizando comparações com outros períodos anuais.



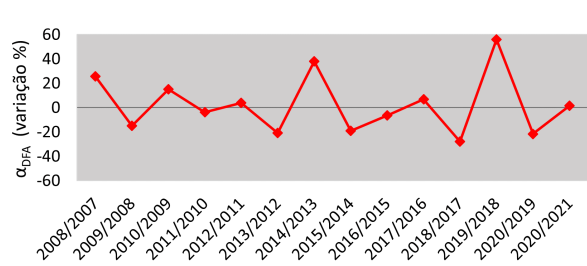
(a) *Pessoas*



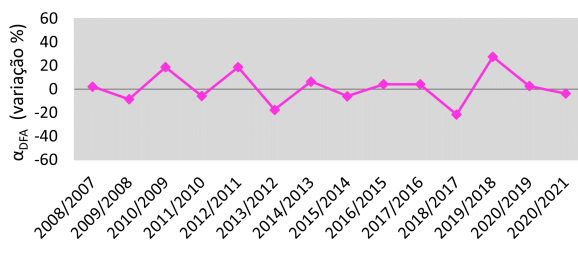
(b) *Mortos*



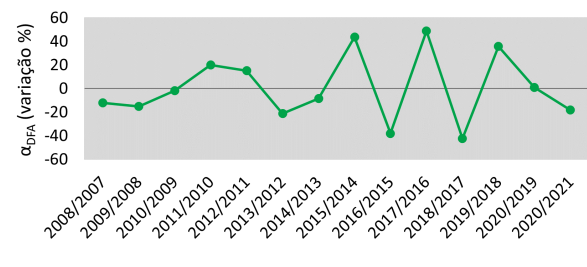
(c) *Feridos leves*



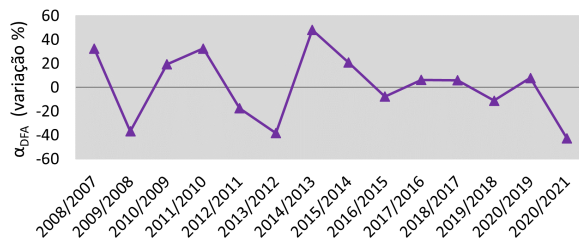
(d) *Feridos graves*



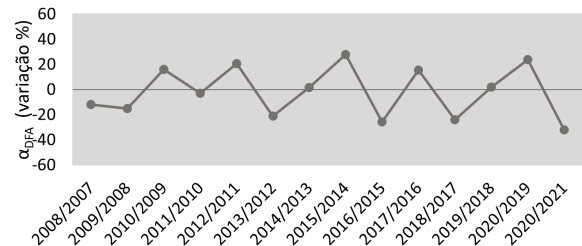
(e) *Feridos*



(f) *Ilesos*



(g) Ignorados

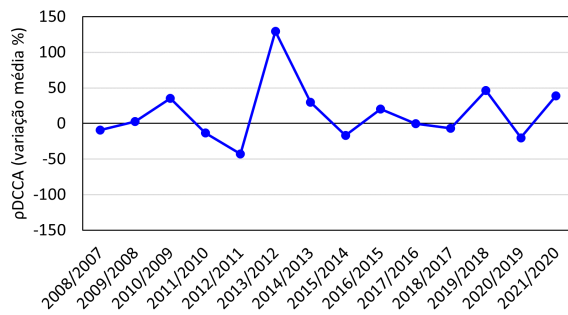


(h) Veículos

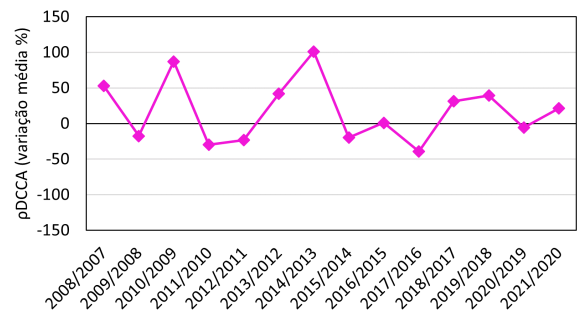
Figura 4.25: Expoente DFA: variação relativa entre os anos, Bahia, 2007-2021

4.2.4.2 Correlação cruzada entre as séries

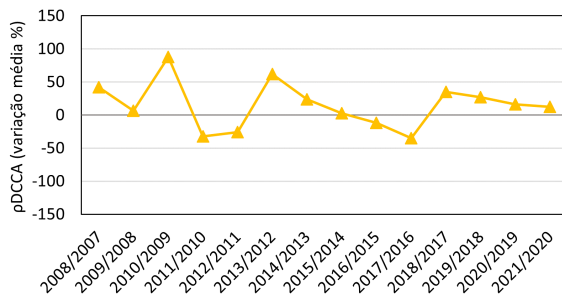
Considerando ainda a variação relativa entre os anos, notamos que o ρ DCCA se mostrou variando no período 2007-2021 para todas as séries, ora apresentando correlação, ora anti-correlação entre *Mortos* e as demais séries, conforme pode ser visto na Figura 4.26.



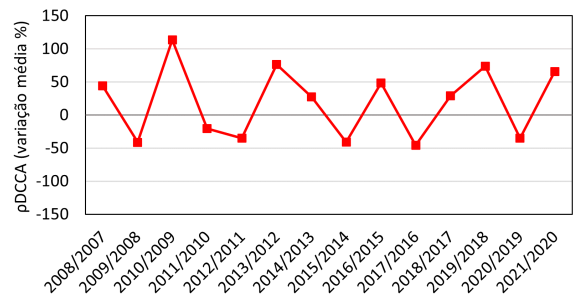
(a) Pessoas x Mortos



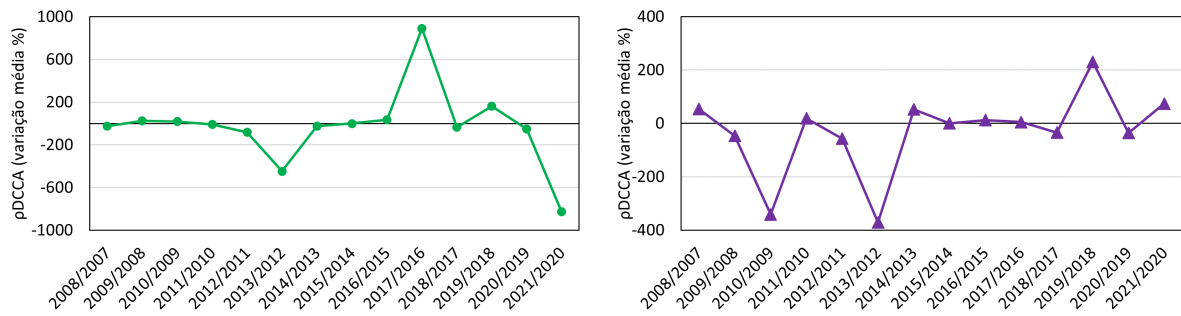
(b) Feridos x Mortos



(c) Feridos leves x Mortos

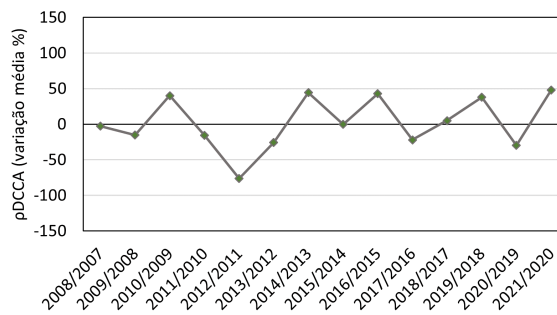


(d) Feridos graves x Mortos



(e) *Ilesos x Mortos*

(f) *Ignorados x Mortos*



(g) *Veículos x Mortos*

Figura 4.26: *Mortos*: correlação cruzada - variação relativa anual do coeficiente de correlação cruzada, Bahia, 2007-2021.

4.3 Conclusões

Nesta dissertação, estudamos as autocorrelações de lei de potência e correlações cruzadas em séries de acidentes de trânsito nas rodovias federais da Bahia, usando análise descritiva, os métodos DFA e DCCA e o coeficiente de correlação cruzada ρ_{DCCA} . As técnicas foram implementadas nas séries temporais de acidentes das variáveis *Pessoas*, *Mortos*, *Feridos*, *Feridos leves*, *Feridos graves*, *Feridos*, *Ilesos*, *Ignorados* e *Veículos*.

Precisamente, nossos resultados principais foram:

No que tange à correlação das variáveis acidentes identificamos

- i) Para todas as séries temporais de acidentes encontramos autocorrelações de lei de potência no período 2007-2021 com um comportamento persistente, sendo que *Ilesos* e *Veículos* apresentaram maior persistência e *Mortos* e *Feridos graves* e *Mortos* exibiram menor;

- ii) A autocorrelação das séries quando analisadas num período de 365 dias ou menos pode não refletir claramente a existência (ou não) de memória longa, inviabilizando comparações com outros períodos anuais;
- iii) Os resultados do ρ_{DCCA} indicaram correlação cruzada em todos os cruzamentos realizados, o que mostra que as oscilações nas quantidades de *pessoas*, *feridos leves*, *feridos graves*, *feridos*, *ilesos*, *ignorados* e *veículos*, afetam o quantidade de *mortos* segundo uma correlação cruzada positiva (altas seguidas de altas ou baixas seguidas de baixas);
- iv) A variável *Mortos* com *Pessoas*, *Feridos leves* e *Feridos* exibe uma lei de potência com níveis mais altos de correlação cruzada para grandes escalas de tempo ($n > 60$). *Ilesos* apresentou correlação cruzada mais fraca do que as demais variáveis em todas as escalas estudadas;
- v) O coeficiente de correlação cruzada se mostrou menor ou igual a 0,5 para praticamente todas variáveis quando analisado em um intervalo de 365 dias, sendo mais intensa entre *Mortos* e *Pessoas*, *Feridos graves*, *Feridos* e *Veículos*;
- vi) A variação relativa de correlação cruzada entre *Mortos* e as demais séries mostrou mudança de sentido passando de correlacionado para anti-correlacionado e vice-versa quando analisada ano-a-ano.

A caracterização dos acidentes mostrou que:

- vii) Os municípios de Feira de Santana de Vitória da Conquista, cidades com maior quantidade de habitantes, depois da capital baiana, apresentaram a maior densidade de acidentes com morte por colisão frontal, tipo de acidente com maior frequência de vítimas fatais;
- viii) O Anel de Contorno de Feira de Santana é o local de maior incidência de acidentes deste município;
- ix) A grande maioria (72,8%) das pessoas envolvidas nos acidentes entre 2017-2020 nas rodovias federais da Bahia são do sexo masculino, cenário já relatado em outras pesquisas.

Destacamos que a partir da análise do ρ_{DCCA} esclarecemos a relação entre a variável *Mortos* e as demais variáveis envolvidas no estudo. Nesse sentido, se duas variáveis exibem correlação cruzada, cada uma delas, a qualquer momento, depende não apenas de seus próprios valores, mas também de valores passados da outra variável. Tais resultados podem ser úteis no planejamento de políticas públicas e ações que visem a mitigação do volume e gravidade dos acidentes das rodovias federais avaliadas por esta pesquisa. Para tanto, a difusão dessas informações devem ser feitas de forma acessível e didática com o

intuito de alcançar compreensão da metodologia e efetividade das ações a serem tomadas para a mitigação do problema.

Os resultados apresentados neste trabalho também revelam que na aplicação do DFA, DCCA e ρ_{DCCA} o volume das variáveis não é o mais importante. Ou seja, mesmo com a mudança na forma de registro dos acidentes com vítimas a partir de junho de 2015, os métodos de correlação das séries não mudam em suas análises e resultados. Isto porque o mais importante nas análises desses métodos é a flutuação das séries não estacionárias.

Finalmente, esses resultados podem contribuir com estudos futuros expandindo as análises para as demais regiões do país, além de oferecer subsídios para modelagem de dados de acidentes. Nessa perspectiva, recomendamos analisar os acidentes nos dias que antecedem e que sucedem feriados importantes em que um grande número de pessoas dirigem por horas para aproveitar feriados os prolongados (festas juninas, natalinas, carnaval, páscoa). Recomendamos também um estudo geoestatístico com base nos métodos para séries não estacionárias, identificando a correlação de variáveis espaciais. Esse tipo de pesquisa tem sido uma ferramenta poderosa estudar diversos fenômenos sociais.

No Apêndice desta dissertação apresentamos dois artigos publicados pelo Grupo de Pesquisa em Modelagem e Estudos de Sistemas Complexos - UEFS com a minha participação. O primeiro artigo analisa a correlação de variáveis meteorológicas em cidades da Bahia ([ZEBENDE et al., 2021](#)). O segundo, apresenta correlações de longo prazo nas ações do G-20 ([GUEDES et al., 2022](#)).

Artigos publicados

Trabalhos publicados pelo Grupo de Pesquisas em Modelagem e Estudos de Sistemas Complexos - UEFS

Artigo 1: Uma visão hora a hora da autocorrelação em dados de temperatura e umidade relativa do ar na Bahia.

Artigo 2: *Efficiency and long-range correlation in G-20 stock indexes: a sliding windows approach.*





DOI: 10.5380/abclima

UMA VISÃO HORA A HORA DA AUTOCORRELAÇÃO EM DADOS DE TEMPERATURA E UMIDADE RELATIVA DO AR NA BAHIA

*AN HOUR-BY-HOUR VIEW OF AUTOCORRELATION IN
TEMPERATURE AND RELATIVE HUMIDITY DATA IN BAHIA*

*UNA VISIÓN HORA A HORA DE LA AUTOCORRELACIÓN EN LOS
DATOS DE TEMPERATURA Y HUMEDAD RELATIVA EN BAHÍA*

Gilney Figueira Zebende  

Universidade Estadual de Feira de Santana
gfzebende@uefs.br

Andréa de Almeida Brito  

Instituto Federal de Educação Ciência e Tecnologia da Bahia
britodea@gmail.com

Roberta Pimenta Cunha dos Santos  

Universidade Estadual de Feira de Santana
rpcs5577@gmail.com

Aloísio Machado da Silva Filho  

Universidade Estadual de Feira de Santana
aloisioestatistico@uefs.br

Resumo: Aqui, propomos analisar empiricamente as autocorrelações em dados meteorológicos de temperatura e umidade relativa do ar medidos hora a hora. Tal motivação se justifica, pois quando se fala/estuda o tema do aquecimento global e a ação antrópica, as referidas variáveis meteorológicas são muito importantes e utilizadas a séculos. Aqui, tal análise será implementada pelo método DFA, largamente empregado na análise de séries temporais não estacionárias. Como nosso estudo de caso, empregamos dados obtidos de quatro estações meteorológicas e automáticas, entre o litoral, recôncavo, oeste e norte do Estado da Bahia. Portanto, com as séries temporais das variáveis envolvidas em cada estação, tendo estas aproximadamente um total de uma década de medida (hora a hora), calculamos a função de flutuação do método DFA. Por não existir uma única lei de potência, os resultados da função de flutuação identificam claramente componentes sazonais entre as variáveis, assim como para todas as estações. Com tais resultados foi possível medir a persistência das séries temporais, e obter três escalas temporais, ou seja, $4 \leq n \leq 30$, $31 \leq n \leq 365$ e $n > 365$ dias. Finalmente, para cada uma destas escalas temporais e hora do dia, calculamos o valor do expoente de autocorrelação, α_{DFA} . Com o valor de α_{DFA} como função da hora, mostramos que há um determinado padrão de

persistência a depender da escala temporal envolvida, isto tanto para a temperatura quanto para a umidade relativa do ar, e que pequenas alterações ocorrem de uma estação meteorológica para outra. Desta forma, este artigo dá uma nova visão das autocorrelações de variáveis meteorológicas, via método DFA, contribuindo para pesquisas relacionadas ao tema climatologia.

Palavras-chave: Variáveis Meteorológicas. Séries Temporais. Autocorrelação.

Abstract: Here, we propose to empirically analyze the autocorrelations in meteorological data of temperature and relative humidity measured hourly. Such motivation is justified, because when one speaks/studies the subject of global warming and anthropic action, these meteorological variables are very important and have been used for centuries. Here, such analysis will be implemented by the DFA method, widely employed in the analysis of non-stationary time series. As our case study, we employ data obtained from four meteorological and automatic stations, between the coast, recôncavo, west and north of Bahia State. Therefore, with the time series of the variables involved in each station, having these approximately a total of a decade of measurement (hour by hour), we calculated the fluctuation function of the DFA method. Because there is no single power law, the results of the fluctuation function clearly identify seasonal components among the variables, as well as for all stations. With such results it was possible to measure the persistence of the time series, and to obtain three time scales, i.e. $4 \leq n \leq 30$, $31 \leq n \leq 365$ and $n > 365$ days. Finally, for each of these time scales and time of day, we calculated the value of the autocorrelation exponent, α_{DFA} . With the value of α_{DFA} as a function of time, we show that there is a certain pattern of persistence depending on the time scale involved, this for both temperature and relative humidity, and that small changes occur from one weather station to another. Thus, this paper gives a new view of autocorrelations of meteorological variables, via the DFA method, contributing to research related to the subject of climatology.

Keywords: Meteorological Variables. Time Series. Auto-correlation.

Resumen: Aquí proponemos analizar empíricamente las autocorrelaciones en los datos meteorológicos de temperatura y humedad relativa medidos cada hora. Esta motivación se justifica porque cuando se habla/estudia el tema del calentamiento global y la acción antrópica, estas variables meteorológicas son muy importantes y se utilizan desde hace siglos. Aquí, dicho análisis se implementará mediante el método DFA, ampliamente empleado en el análisis de series temporales no estacionarias. Como caso de estudio, empleamos datos obtenidos de cuatro estaciones meteorológicas y automáticas, entre el litoral, el recôncavo, el oeste y el norte del estado de Bahía. Por tanto, con las series temporales de las variables implicadas en cada estación, teniendo éstas aproximadamente un total de una década de medición (hora a hora), calculamos la función de fluctuación del método DFA. Al no existir una única ley de potencia, los resultados de la función de fluctuación identifican claramente los componentes estacionales entre las variables, así como para todas las estaciones. Con estos resultados fue posible medir la persistencia de las series temporales y obtener tres escalas de tiempo, es decir, $4 \leq n \leq 30$, $31 \leq n \leq 365$ y $n > 365$ días. Por último, para cada una de estas escalas de tiempo y hora del día, calculamos el valor del exponente de autocorrelación, α_{DFA} . Con el valor de α_{DFA} en función del tiempo, mostramos que existe un cierto patrón de persistencia en función de la escala temporal implicada, esto tanto para la temperatura como para la humedad relativa, y que se producen pequeños cambios de una estación meteorológica a otra. Así, este artículo ofrece una nueva visión de las autocorrelaciones de las variables meteorológicas, a través del método DFA, contribuyendo a la investigación relacionada con la climatología del tema.

Palabras-clave: Variables Meteorológicas. Series de tiempo. Autocorrelación.

Submetido em: 23/04/2020

Aceito para publicação em: 26/07/2021

Publicado em: 22/09/2021

INTRODUÇÃO

As componentes climáticas nunca estão em equilíbrio, variam constantemente com o tempo. Assim, pequenas alterações nas suas componentes iniciais podem levar o sistema a eventos totalmente inesperados no futuro, por exemplo, podem resultar em grandes desastres naturais (Geographic, 2021). Estas premissas caracterizam o clima como um sistema complexo, podendo o mesmo ser estudado com novos modelos matemáticos ou ferramentas estatísticas.

Do ponto de vista da climatologia, os modelos matemáticos são importantes, pois podem ser aplicados para estudar a dinâmica do clima, projeções futuras, mudanças na temperatura do ar, entre outras aplicações. Dados meteorológicos acumulados ao longo de muitos anos estão sendo analisados por técnicas estatísticas e por diversos modelos matemáticos. Como se sabe, a maioria desses modelos matemáticos usa sistemas de equações diferenciais baseadas nas leis da Física. Portanto, para estudar este sistema complexo, devemos definir adequadamente as variáveis fundamentais, tais como: temperatura do ar, pressão atmosférica, direção e velocidade do vento, umidade relativa do ar entre outras, de acordo com o capítulo cinco em WMO (1989).

Especificamente, há um grande número de situações, seja no mundo real ou não, em que uma determinada variável $u(i)$ é medida em intervalos de tempo iguais e sucessivos, formando uma série temporal $\{u(i)\}$. Algumas estratégias para o estudo de tais séries temporais no tempo já foram desenvolvidas, como por exemplo utilizar a estatística descritiva para descrever e sumarizar o conjunto dos dados com algumas medidas, tais como, as medidas de tendência central (média, mediana e moda) e as medidas de dispersão (desvio padrão, variância, o valor máximo e mínimo e curtose) e a representação do comportamento da série por meio de gráficos. Porém, se quisermos analisar o efeito de memória (de curto e longo prazo) em séries temporais, outras técnicas estatísticas devem ser empregadas, como por exemplo a análise R/S (Hurst, 1951) ou mesmo o método **DFA** (Peng *et al.*, 1994), com mais de 4000 citações.

O método *Detrended Fluctuation Analysis* (Análise de Flutuações Retirando Tendências) (**DFA**) tem-se mostrado robusto ao longo destes anos, no que tange principalmente a análise de autocorrelação de longo prazo. Variações do método DFA também já foram implementadas, entre elas o método *Detrended Cross-Correlation Analysis*



(Análise de Correlações Cruzadas Retirando Tendências) (**DCCA**) por Podobnik e Stanley (2008), que trata da correlação cruzada entre distintas séries temporais e que deu origem ao coeficiente de correlação cruzada proposta por Zebende (2011), assim como o coeficiente de correlação cruzada múltiplo por Zebende e da Silva Filho (2018a). Muitas aplicações destes métodos baseados no método **DFA** podem ser encontrados, como por exemplo em: Vassoler e Zebende (2012), Zebende *et al.* (2013, 2018b), de Almeida Brito *et al.* (2019), Zebende *et al.* (2020), entre muitos outros.

Portanto, dentro da área de climatologia (caracterizada por variáveis complexas), é importante o emprego de novas técnicas estatísticas robustas, como é o caso do método **DFA**, para a obtenção de novos e relevantes estudos. Assim, neste artigo nos motivamos a analisar as autocorrelações em dados de temperatura e umidade relativa do ar medidos hora a hora, pelo método **DFA**. Com nosso estudo de caso, empregamos dados fornecidos pelo Instituto Nacional de Meteorologia e obtidos de quatro estações automáticas, como veremos na próxima seção.

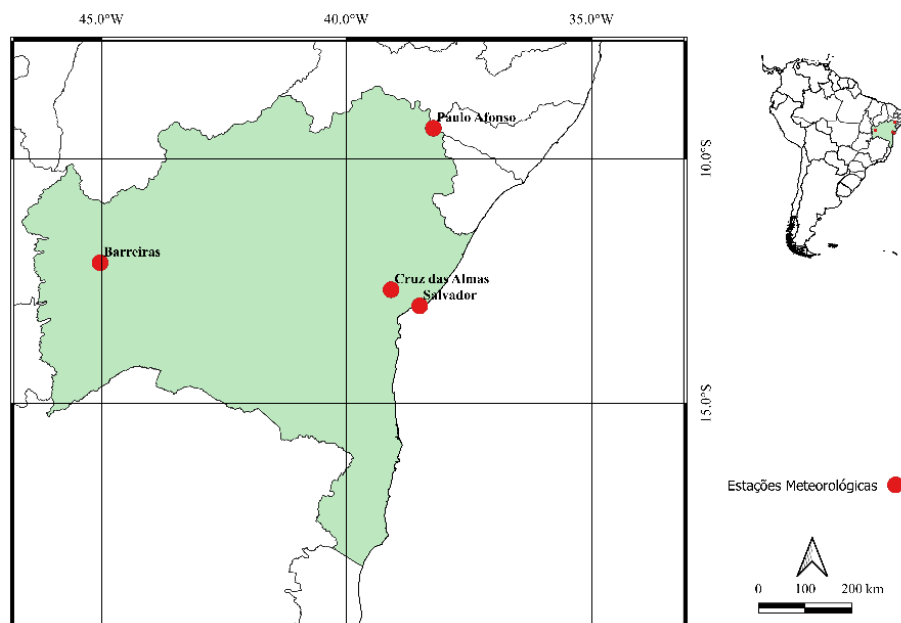
Logo, para cumprir com nossos objetivos, dividimos este artigo em seções, a contar da Introdução aqui, seguimos com Material e Métodos, Resultados e Discussão e por fim as Conclusões.

MATERIAL E MÉTODOS

Dados

Aqui vamos analisar a autocorrelação das variáveis temperatura do ar (medida em °C) e umidade relativa do ar (em %), obtidas em quatro estações meteorológica de superfície automática localizadas no Estado da Bahia, com dados horários e administradas pelo Instituto Nacional de Meteorologia, INMET (2021). Para isto, abaixo na Figura 1, apresentamos localização para as estações e na Tabela 1 a descrição dos dados.

Figura 1 - Estado da Bahia e as quatro estações localizadas nas cidades de Barreiras, Cruz das Almas, Paulo Afonso e Salvador.



Fonte: Elaborado pelos autores (2020).

Tabela 1 - Informação geográfica e descrição dos dados.

Estação	Início	Fim	N
Barreiras	22/12/2001	30/09/2010	2047
Cruz das Almas	26/01/2003	30/09/2010	1919
Paulo Afonso	18/06/2003	30/09/2010	2065
Salvador	07/10/2000	30/09/2010	2616

Fonte: Elaborado pelos autores (2020).

A importância dessas estações (cidades), decorre do fato de que as mesmas possuem os melhores bancos de dados e também pelas suas cidades possuírem as seguintes características:

- **Barreiras:** É a cidade mais populosa e o mais importante centro agrícola da região oeste da Bahia, destacando-se no agronegócio nacional (internacional) como grande produtor de algodão e soja.
- **Cruz das Almas:** Considerada um importante centro no Recôncavo baiano, localizada ao redor da baía de Todos os Santos, cobrindo não apenas a costa, mas também toda a região do interior que a circunda. Possui alguns centros de pesquisa agropecuária, como a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA).

- **Paulo Afonso:** É a cidade onde está localizado o complexo da usina hidrelétrica do Vale do São Francisco (CHESF). Possui quatro plantas, PA I, PA II, PA III e PA IV. A principal característica da usina hidrelétrica é ter sido a primeira usina subterrânea instalada no Brasil, porque suas turbinas estão a mais de 80 metros abaixo do nível do rio São Francisco.
- **Salvador:** É o centro econômico e a capital da Bahia. Notável por sua comida, música e arquitetura, com fortes influências africanas e portuguesas. Com mais de 2,9 milhões de pessoas, é a cidade mais populosa do Nordeste e a terceira mais populosa do Brasil.

Para a análise de séries temporais, existem alguns caminhos bem conhecidos a serem seguidos, como em Huber (1981) e Wilcox (2012). Porém, se as séries temporais exibirem comportamentos complexos, como autoafinidade, podemos aplicar novas estratégias para sua análise, como em Peng *et al.* (1992), Zebende *et al.* (1998), Jiang e Zhou (2011), Podobnik *et al.* (2011), Kristoufek (2014), Wang *et al.* (2013), Zebende *et al.* (2018a), entre outros. Desta forma, neste artigo adotaremos o método **DFA**, método este introduzido pelos autores para a análise de autocorrelação em séries temporais não estacionárias, e que aqui será descrito logo na próxima sub-seção.

Método DFA (Peng *et al.*, 1994)

Este método fornece uma relação entre a função de flutuação, F_{DFA} , e a escala temporal n . O método **DFA** tem sido muito eficiente na detecção de autocorrelações de longo alcance com caudas longas, do tipo lei de potência. A importância do método pode ser comprovada pelo grande número de aplicações e citações em todas as áreas do conhecimento, como em Király e János (2005), Koscielny-Bunde *et al.* (1998), Fraedrich e Blender (2003), Guo *et al.* (2016). Nesta direção, abaixo brevemente faremos uma introdução do algoritmo do método **DFA**, que envolve os seguintes passos:

Passo 1- Considere um determinado sinal $u(i)$, onde $i = 1, \dots, N$, sendo N o tamanho (número de pontos) da série temporal. O sinal $u(i)$ é integrado, e assim obtém-se $x(k) = \sum_{i=1}^k [u(i) - \langle u \rangle]$, com $\langle u \rangle$ sendo o valor médio de u em toda série temporal e com $k = 1, \dots, N$;

Passo 2- O sinal integrado $x(k)$ é dividido em caixas de igual tamanho n (a escala temporal);

Passo 3- Para cada caixa de tamanho n , um ajuste polinomial de ordem l (geralmente é usado $l = 1$) é feito em $x(k)$, que será a tendência do sinal dentro da caixa. A coordenada y do ajuste dentro de cada caixa será definida por $x_n(k, l)$;

Passo 4- O sinal integrado $x(k)$ é “destendenciado”, isto pela subtração de $x_n(k, l)$ em cada caixa (de tamanho n);

Passo 5- Para caixas sobrepostas, com $(N - n)$ valores, calcula-se a função de covariância em todos as caixas, ou seja,

$$f_{DFA}^2(n, i) = \frac{1}{(n + 1)} \sum_{k=i}^{i+n} [x(k) - x_n(k, l)]^2$$

E, finalmente calcula-se a sua função de flutuação, definida por

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

Passo 6 - O cálculo, dos passos anteriores, é repetido para as escalas temporais que compreendem, $4 \leq n \leq N/4$, e assim tem-se a relação entre F_{DFA} e a escala temporal n .

Se o sinal for caracterizado por uma lei de potência, então $F_{DFA} \sim n^{\alpha_{DFA}}$. Neste sentido, α_{DFA} é definido como sendo o expoente de autocorrelação de longo-alcance, com as seguintes propriedades, ver Walleczek (2000) e Zebende *et al.* (2017), (Quadro 1).

Quadro 1 - Informação geográfica e descrição dos dados.

expoente	tipo de série temporal
$\alpha_{DFA} < 0,5$	anti-persistente
$\alpha_{DFA} \cong 0,5$	descorrelacionada, “ruído branco”
$\alpha_{DFA} > 0,5$	persistente de longo alcance
$\alpha_{DFA} \cong 1,0$	ruído $1/f$
$\alpha_{DFA} > 1,0$	não estacionária
$\alpha_{DFA} \cong 3/2$	ruído Browniano

Fonte: Elaborado pelos autores (2020).

O expoente α_{DFA} é semelhante ao expoente de Hurst (1951), porém α_{DFA} pode ser aplicado a sinais não estacionários e é mais robusto do que Hurst, ver Hu *et al.* (2001) e Chen *et al.* (2002), onde efeitos de tendência e não estacionaridades foram testados. Heneghan e McDarby (2000) estabeleceram a relação entre o método **DFA** e a análise espectral, e concluíram que o **DFA** e a análise espectral fornecem caracterizações equivalentes de sinais estocásticos com autocorrelação de longo-prazo. Porém, nem sempre há uma caracterização de lei de potência única, assim tendências e sazonalidades podem estar presentes no sinal.

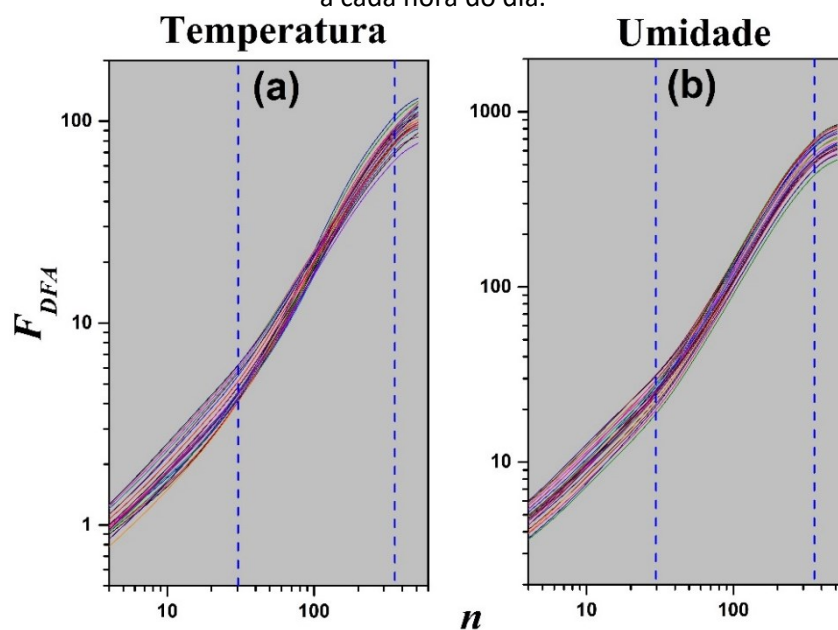
Nestes casos o método **DFA** também pode ser aplicado para a análise destes efeitos, como pode ser visto em Zebende *et al.* (2017).

Desta forma, com nossos dados tabulados e com a introdução do método **DFA**, a seguir na próxima seção, apresentamos os resultados as séries temporais de temperatura do ar e da umidade relativa do ar, onde identificamos tendências e sazonalidades, bem como outros padrões em função da hora do dia e da estação meteorológica empregada.

RESULTADOS E DISCUSSÃO

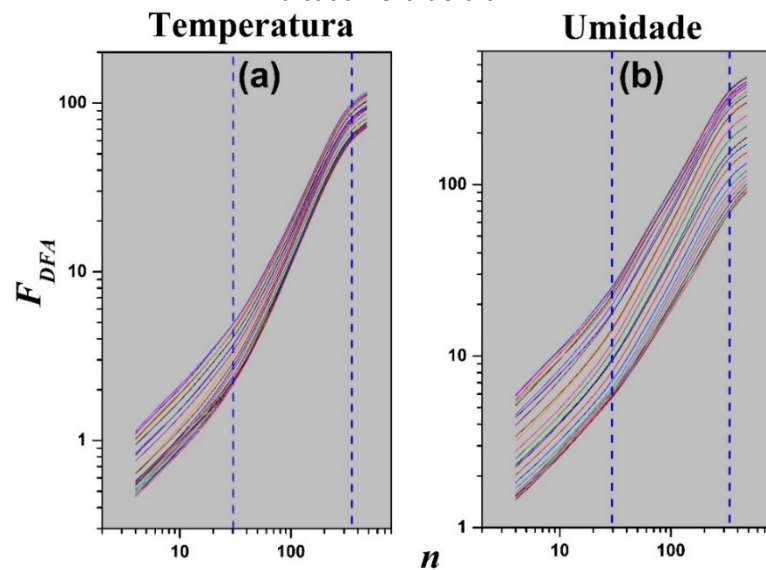
Como temos dados horários para a temperatura do ar e para a umidade relativa do ar, vamos calcular a função de flutuação, $F_{DFA}(n)$, para cada hora local do dia (dada em UTC) e tentar ver a influência do horário e da localização das estações nos resultados. Portanto, a aplicação do método **DFA** para as estações localizadas nas cidades de Barreiras, Cruz das Almas, Paulo Afonso e Salvador, levando-se em consideração os valores horários da temperatura do ar e da umidade relativa do ar encontram-se nas Figuras 2, 3, 4 e 5.

Figura 2 - $F_{DFA}(n)$ como função da escala temporal n , isto para a temperatura do ar (a) e para a umidade relativa do ar (b), localizada na estação de Barreiras. As linhas verticais representam a escala temporal $n = 30$ e $n = 365$ dias respectivamente. Nesta figura temos as 24 curvas referentes a cada hora do dia.



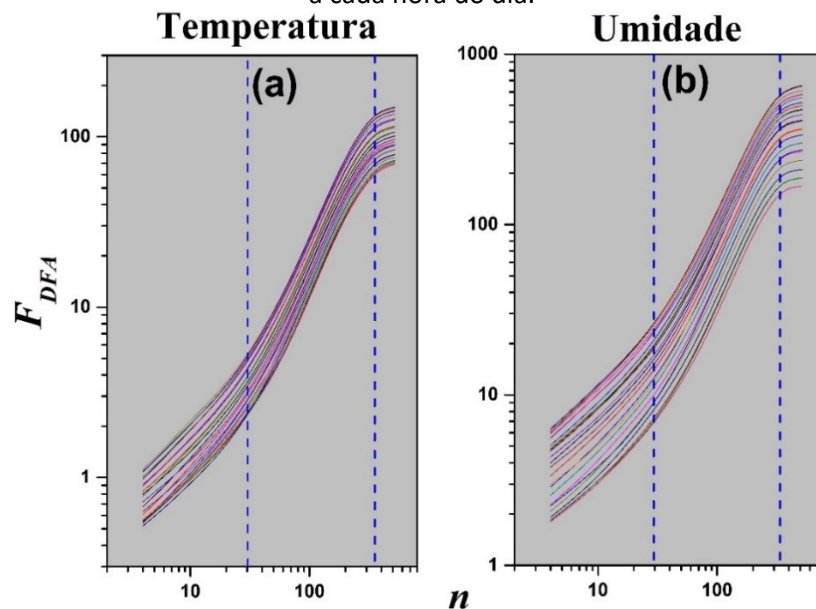
Fonte: Elaborado pelos autores (2020).

Figura 3 - $F_{DFA}(n)$ como função da escala temporal n , isto para a temperatura do ar (a) e para a umidade relativa do ar (b), localizada na estação de Cruz das Almas. As linhas verticais representam a escala temporal $n = 30$ e $n = 365$ dias respectivamente. Nesta figura temos as 24 curvas referentes a cada hora do dia.



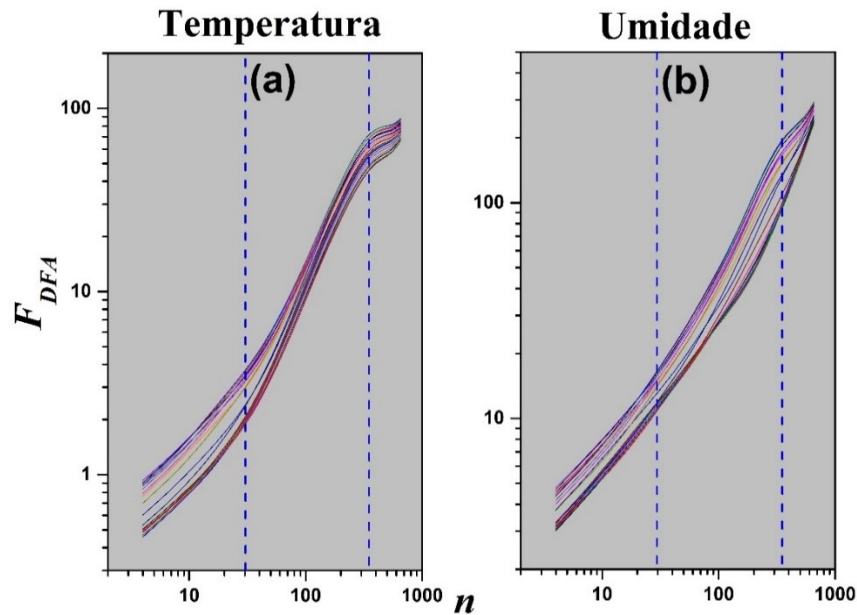
Fonte: Elaborado pelos autores (2020).

Figura 4 - $F_{DFA}(n)$ como função da escala temporal n , isto para a temperatura do ar (a) e para a umidade relativa do ar (b), localizada na estação de Paulo Afonso. As linhas verticais representam a escala temporal $n = 30$ e $n = 365$ dias respectivamente. Nesta figura temos as 24 curvas referentes a cada hora do dia.



Fonte: Elaborado pelos autores (2020).

Figura 5 - $F_{DFA}(n)$ como função da escala temporal n , isto para a temperatura do ar (a) e para a umidade relativa do ar (b), localizada na estação de Salvador. As linhas verticais representam a escala temporal $n = 30$ e $n = 365$ dias, respectivamente. Nesta figura temos as 24 curvas referentes a cada hora do dia.



Fonte: Elaborado pelos autores (2020).

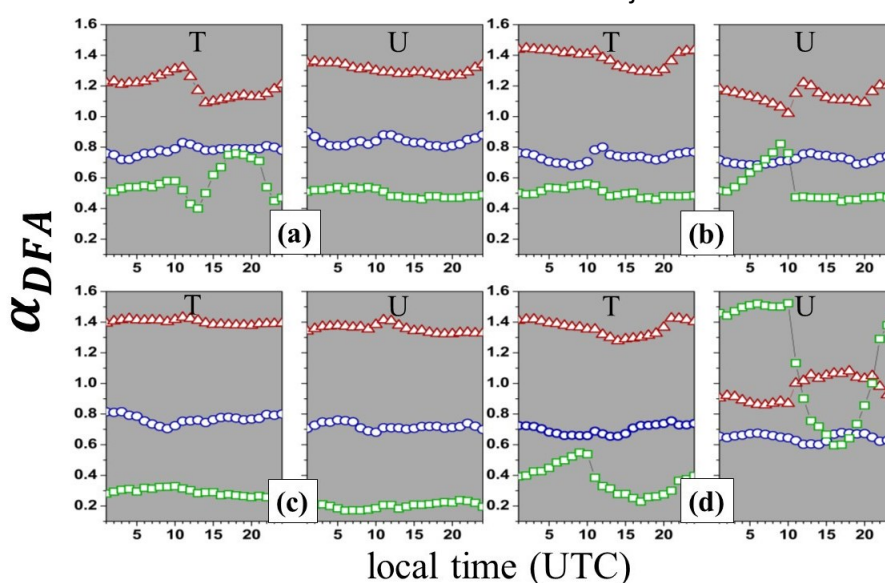
Pelas quatro figuras apresentadas acima e considerando toda a escala temporal envolvida, percebe-se não haver uma única lei de escala para $F_{DFA}(n)$, isto tanto para a temperatura do ar quanto para a umidade relativa do ar, e independentemente do horário e da estação (cada linha na figura representa uma hora local). Especificamente, para a estação de Barreiras os dados estão mais ajustados, enquanto que para a estação de Paulo Afonso não. De maneira clara, é possível identificar três transições para a função de flutuação, $F_{DFA}(n)$, que estão associadas a componentes sazonais.

De conhecimento desta peculiaridade nos resultados, demarcamos três escalas temporais para uma análise mais profunda, identificadas pelas duas linhas verticais nas figuras, ou seja, uma em $n = 30$ e a outra em $n = 365$ dias (linhas verticais). Com tal marcação, assim como realizado em Zebende *et al.* (2017), calculamos os valores dos expoentes α_{DFA} dentro de cada escala temporal, ou seja:

- de $n = 4$ até $n = 30$ dias, tendo α_{DFA-1} (círculo azul);
- de $n = 31$ até $n = 365$ dias, tendo α_{DFA-2} (triângulo vermelho);
- e com $n > 365$ dias, tendo α_{DFA-3} (quadrado verde).

Também teremos este valor para cada hora do dia (linhas coloridas nos resultados), no intuito de facilitarmos a visualização do efeito do passar do dia. Tal realização encontra-se na Figura 6, que nos dá o valor do expoente α_{DFA} como função da hora do dia e dentro de cada escala temporal específica.

Figura 6 - Expoente de autocorrelação em função da hora local, aferido para as estações de: (a) Barreiras, (b) Cruz das Almas, (c) Paulo Afonso e (d) Salvador. Onde T é o valor de α_{DFA} para a temperatura do ar e U α_{DFA} para a umidade relativa do ar. Círculo representa a escala temporal $4 \leq n \leq 30$, triângulo $31 \leq n \leq 365$ e quadrado $n > 365$. Cada valor do expoente de autocorrelação calculado teve um $R^2 \geq 0.98$ no seu ajuste linear.



Fonte: Elaborado pelos autores (2020).

Na figura 6 fica claro que de fato há três valores para α_{DFA} a depender da escala temporal empregada, identificados pelos símbolos em azul, vermelho e verde. Neste figura vemos que para a escala temporal de $4 \leq n \leq 30$, tanto para a temperatura do ar quanto para a umidade relativa do ar, independentemente da estação meteorológica, $\alpha_{DFA} \cong 0,80$, ou seja, tem comportamento persistente. Porém, se olharmos as outras escalas temporais observa-se outros padrões, isto é, para $31 \leq n \leq 365$ dias, α_{DFA} geralmente é maior do que 1, evidenciando um comportamento persistente não estacionária (exceto para Salvador e vendo a Umidade Relativa do Ar). Finalmente, para as escalas temporais de $n > 365$ dias, α_{DFA} evidencia um comportamento de antipersistência, com valores menores do que 0.5 (exceto novamente para Salvador e vendo a Umidade Relativa do Ar). Também, de uma forma geral, o valor do expoente α_{DFA} tem um comportamento qualitativo semelhante tanto para a



temperatura do ar com para a umidade relativa do ar, mostrando de forma indireta a associação entre as variáveis.

CONCLUSÕES

Neste trabalho propomos a análise de autocorrelação das séries temporais de temperatura do ar e umidade relativa do ar, medidas hora a hora em quatro estações meteorológicas localizadas no estado da Bahia. Tal estudo foi realizado pelo método **DFA**. Com tal análise foi possível identificar pela função de flutuação, $F_{DFA}(n)$, três componentes sazonais, aqui divididas nas seguintes escalas temporais: **período I** ($4 \leq n \leq 30$), **período II** ($31 \leq n \leq 365$) e **período III** ($n > 365$) dias. Assim, para cada escala temporal e para cada hora do dia, criamos um novo diagrama, representado aqui neste artigo. Com este diagrama construído, foi possível observar a variação na persistência de cada série temporal, para a temperatura do ar e também para a umidade relativa do ar, a assim comparar uma estação meteorológica com a outra. Finalmente, percebemos que há uma maior variabilidade para os resultados referentes a umidade relativa do ar, principalmente para as estações localizadas próximas ao litoral e períodos II e III. Nossos achados servirão como proposta metodológica para avaliar as autocorrelações de séries temporais de dados relativos ao clima no tempo e espaço, contribuindo assim no avanço de pesquisas relacionadas ao tema climatologia.

AGRADECIMENTOS

Os autores agradecem ao Instituto Nacional de Meteorologia (INMET) pelos dados fornecidos para a pesquisa. A. A. Brito agradece a Fundação de Amparo à Pesquisa na Bahia (FAPESB), pela bolsa de estudo BOL 0262/2017. G. F. Zebende agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa de pesquisa 310136/2020-2.

REFERÊNCIAS

CHEN, Zhi; IVANOV, Plamen Ch.; HU, Kun; STANLEY, H. Eugene. Effect of nonstationarities on detrended fluctuation analysis. **Phys. Rev. E**, v. 65, n. 4, p. 041107, 2002.

DE ALMEIDA BRITO, A.; de ARAÚJO, H. A.; ZEBENDE, G. F. Detrended multiple cross-correlation coefficient applied to solar radiation, air temperature and relative humidity. **Scientific Reports**, v. 9, n. 1, p. 1-10, 2019.

FRAEDRICH, K.; BLENDER, R. Scaling of atmosphere and ocean temperature correlations in observations and climate models. **Phys. Rev. Lett.**, v. 90, n. 10, p. 108501, 2003.

GEOGRAPHIC, N. **Natural Disasters**. set. de 2021. Disponível em: <https://www.nationalgeographic.com/environment/topic/natural-disasters-weather>. Acesso em: set. de 2021.

GUO, E.; ZHANG, J.; Si, H.; DONG, Z.; CAO, T.; LAN, W. 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**, v.130, p. 597-607, 2016.

HENEGHAN, C.; MCDARBY, G. Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. **Phys. Rev. E**, v. 62, n. 5, 6103-6110, 2000.

INMET, Instituto Nacional de Meteorologia. **Dados de estação meteorológica de observação de superfície automática**. set. de 2021. Disponível em: <https://portal.inmet.gov.br>. Acesso em: set. de 2021.

HU, Kun; IVANOV, Plamen Ch.; CHEN, Zhi; CARPENA, Pedro; EUGENE Stanley, H. Effect of trends on detrended fluctuation analysis. **Phys. Rev. E**, v. 64, n. 1, p. 011114, 2001.

HUBER, P. J. **Robust Statistics**. John Wiley and Sons, New York, 1981.

HURST, H. E. Long-term storage capacity of reservoirs. **Transactions of the American Society of Civil Engineers**, v. 116, n. 1, p. 770-799, 1951.

JIANG, Z.-Q.; ZHOU, W.-X. Multifractal detrending moving-average cross-correlation analysis. **Phys. Rev. E**, v. 84, n. 1, p. 016106, 2011.

KIRÁLY, A.; JÁNOSI, I. M. Detrended Fluctuation Analysis of daily temperature records: Geographic dependence over Australia. **Meteorology and Atmospheric Physics**, v. 88, p. 119-128, 2005.

KOSCIELNY-BUNDE, Eva; BUNDE, Armin; HAVLIN, Shlomo; ROMAN, H. Eduardo; GOLDREICH, Yair; SCHELLNHUBER, Hans-Joachim. Indication of a universal persistence law governing atmospheric variability. **Phys. Rev. Lett.**, v. 81. n. 3, p. 729-732, 1998.

KRISTOUFEK, L. Measuring cross-correlation between non-stationary series with DCCA coefficient. **Physica A: Statistical Mechanics and its Applications**, v. 402, p. 291-298, 2014.

PENG, C.-K.; BULDYREV, S. V.; GOLDBERGER, A. L.; HAVLIN, S.; SCIORTINO, F.; SIMONS, M.; STANLEY, H. E. Long-range correlations in nucleotide sequences. **Nature**, v. 356, p. 168-170, 1992.

PENG, C.-K.; BULDYREV, S. V.; HAVLIN, S.; SIMONS, M.; STANLEY, H. E.; GOLDBERGER, A. L. Mosaic organization of DNA nucleotides. **Phys. Rev. E**, v. 49, n. 2, p. 1685-1689, 1994.



PODOBNIK, B.; JIANG, Z.-Q.; ZHOU, W.-X.; STANLEY, H. E. Statistical tests for power-law cross-correlated processes. **Phys. Rev. E**, v. 84, p. 066118, 2011.

PODOBNIK, B.; STANLEY, H. E. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. **Phys. Rev. Lett.**, v. 100, n. 8, p. 084102, 2008.

VASSOLER, R.; ZEBENDE, G. F. DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity. **Physica A: Statistical Mechanics and its Applications**, v. 391, n.7, p. 2438-2443, 2012.

WALLECZEK, J. **Self-organized biological dynamics and nonlinear control: toward understanding complexity, chaos and emergent function in living systems**. Cambridge University Press, 2000, p. 421-428.

WANG, G.-J.; XIE, C.; CHEN, S.; YANG, J.-J.; YANG, M.-Y. 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**, v. 392, n. 17, p. 3715-3730, 2013.

WILCOX, R. **Introduction to Robust Estimation and Hypothesis Testing (Third Edition)**. Academic Press, Boston, 2012.

WMO, World Meteorological Organization. **Calculation of Monthly and Annual 30-year Standard Normals**, WCDP (Series), Edição 10, WMO, Geneva, 1989.

ZEBENDE, G. F.; de OLIVEIRA, P. M. C.; PENNA, T. J. P. Long-range correlations in computer diskettes. **Phys. Rev. E**, v. 57, n.3, p. 3311-3314, 1998.

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

ZEBENDE, G. F.; SILVA, M. F.; FILHO, A. M. DCCA cross-correlation coefficient differentiation: Theoretical and practical approaches. **Physica A: Statistical Mechanics and its Applications**, v. 392, n.8, p. 1756-1761, 2013.

ZEBENDE, G. F.; FERNANDEZ, B. F.; PEREIRA, M. G. Analysis of the variability in the sdb star KIC 10670103: DFA approach. **Monthly Notices of the Royal Astronomical Society**, v. 464, p. 2638-2642, 2017.

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

ZEBENDE, G. F.; BRITO, A. A.; SILVA FILHO, A. M.; CASTRO, A. P. pDCCA applied between air temperature and relative humidity: An hour/hour view. **Physica A: Statistical Mechanics and its Applications**, v.494, p. 17-26, 2018b.

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

Efficiency and long-range correlation in G-20 stock indexes: a sliding windows approach

E. F. Guedes^{a,b}, R. P. C. Santos^e, L. H. R. Figueredo^c, P.A. da Silva^c, R. M. T. S. Dias^{d,e}, G. F. Zebende^e

^a*Brazilian Hospital Services Company - Clímério de Oliveira Maternity Hospital, Health Regulation and Evaluation Sector, Salvador, Bahia, Brazil*

^b*Federal University of Recôncavo Baiano, Exact and Technological Sciences Center, Cruz das Almas, Bahia, Brazil*

^c*Applied Statistics Program, Jorge Amado University Center, Bahia, Brazil*

^d*Polytechnic Institute of Setúbal, Setúbal, Portugal*

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

Abstract

This article aims to analyze whether the financial crises of the past 20 years have reduced efficiency, in its weak form, in 19 stock markets belonging to the 20 most developed economies (G-20). The sample period comprises the period from January 2, 2000 to February 5, 2021 with the respective financial crises, namely: Dot-com, Argentina, Subprime, Sovereign debt, China stock market crash (2015-2016), UK's withdrawal from the European Union, and the global pandemic of 2020. The results highlight that most markets show signs of (in)efficiency in each sliding window (1000 days), that is, they show asymmetries and non-Gaussian distributions, and $\alpha_{DFA} \neq 0.5$. These findings suggest that the random walk hypothesis is rejected in certain markets, which has implications for investors, since some returns may be expected, creating arbitrage and abnormal profit opportunities, contrary to the random walk and informational efficiency hypotheses. The results found also open room for market regulators to take steps to ensure better informational data across international financial markets.

Key words: Autocorrelation, DFA, efficiency, sliding window, times series.

Email address: efgestatistico@gmail.com (E. F. Guedes)

1. Introduction

The efficient markets hypothesis (EMH) developed by Fama [1, 2] has been the subject of intense discussion in the specialized literature [3–7]. The EMH's weak form holds that the market is efficient at reflecting all available public information. Market returns are independent and therefore do not help predict future returns. This means that no investor can obtain abnormal returns only through the analysis of historical series [8].

Although the EMH is supported by assumptions of rationality of economic agents and market unpredictability, empirical evidence signals the inefficiency in the behavior of stock prices and stock indexes, mainly on issues such as information asymmetry [9, 10], anomalies (calendar effect, momentum effect, firm size effect, etc) [11–17], long memory [18–22], and the occurrence of black swans by the market dynamics itself [23–28].

In order to analyze whether financial crises decrease efficiency (weak form) in stock markets, in this paper the sliding windows approach is implemented to investigate the EMH hypothesis in 19 stock markets that make up the G-20 (largest economies in the world). With this approach it is possible to take successive samples of a time series and thus obtain, for example, time series of descriptive statistics such as mean, standard deviation and the skewness and kurtosis coefficients, as well as the autocorrelation exponent α_{DFA} . The α_{DFA} is an exponent of scale, a self-affinity parameter that represents the long-range power law correlation of a signal. The great advantage of DFA over many methods is that it allows the detection of long-range correlations embedded in seemingly non-stationary time series and avoids the spurious detection of apparent long-range correlations, which are indicative of nonstationarity [29].

In recent years, the sliding windows approach has been used in analyses in many different areas of knowledge. In finance, for example, it has been widely employed in univariate analyses on efficiency (weak form) of stock price or stock market indexes [30–45], and recently in bivariate analyses on markets dependence [46, 47].

To perform this analysis, this article is structured in sections, namely: Brief Literature Review (Section 2); Methodology (Section 3); Results (Section 4) and finally Conclusions (Section 5).

2. Brief literature review

The literature on this topic is so large that a complete review is impossible and will not be attempted here. Briefly, previous studies that have investigated EMH are presented.

The first concept of market efficiency was presented by Gibson [48]. The author considered that the stock prices had complete information. Later, the French mathematician Bachelier [49] found that the behavior of asset prices fluctuated randomly, that is, they are independent of previous fluctuations, thus formulating the hypothesis of random walk. Pearson [50] introduced the term random walk to describe the path taken by a drunkard, who staggers in an unpredictable and random pattern, which indicated the impossibility of predictions, by investors, based on historical prices, failing to beat the market [51, 52].

Kendall and Hill [53] examined stock prices and found that fluctuations exhibited a random walk behavior with almost zero autocorrelation in price changes. Similarly, Working [54] and Roberts [55] found that the movements in stock returns had the properties of random walk. Later Cowles [56] re-examined the results of the study by Cowles and Jones [57] and still found mixed evidence of serial correlation, even after correcting an error caused by the mean.

Fama[58] shows that an efficient market is made up of a group of rational agents that compete for predicting the behavior of asset prices. It is assumed that all relevant information is already available to all market agents, that is, the arrival of new information is quickly reflected in prices, thus preventing agents from being able to predict price fluctuation and thus obtain anomalous returns without incur additional risk.

The topic of the efficient market hypothesis (EMH) defending the current price of assets reflects all the information available at a given time, and the price adjusts quickly, as new and unforeseen information comes into the market. The mean reversion hypothesis, also known as negative series correlation, has been interpreted as an efficient correction mechanism in developed markets and a sign of a speculative bubble in emerging financial markets [59, 60].

Different studies have addressed the issue of market efficiency, by analyzing the hypothesis of predictability of returns through the analysis of mean-reversion patterns of stock prices, inspired by the seminal works of Poterba and Summers [61] and also Fama and French [60], who documented

mean-reversion in stock markets over time horizons longer than one year. According to Malafeyev *et al.* [62], when the random walk and informational efficiency hypotheses are rejected, they cause extreme commotions in stock prices. The occurrence of these phenomena may eventually decrease the implementation of efficient portfolio diversification strategies.

Hamid *et al.* [63] analyzed the financial markets of Pakistan, India, Sri Lanka, China, Korea, Hong Kong, Indonesia, Malaysia, Philippines, Singapore, Thailand, Taiwan, Japan and Australia, from January 2004 to December 2009. The authors suggest that prices do not follow the random walk hypothesis for countries in Asia-Pacific region.

Nisar and Hanif [64] examined major stock markets in South Asia, including India, Pakistan, Bangladesh and Sri Lanka, showing that the South Asian stock markets reject the random walk hypothesis, evidencing market (in)efficiency. Mehla and Goyal [65] highlight that the Indian stock market does not show random walk features and, as such, is not efficient in the weak form.

Sensoy and Tabak [42] show that the 2008 financial crisis has had a detrimental effect, mostly on the EU stock markets. However, the eurozone sovereign debt crisis had a significant adverse effect only on the markets in France, Spain and Greece. Shirvani and Delcoure [66] analyzed 16 OECD markets, showing that they are efficient, as the random walk hypothesis was not rejected.

Ngene *et al.* [67] examined 18 emerging markets in the presence of multiple, unforeseen and successive structural breaks. When single break tests are used, the random walk hypothesis is rejected. However, when using double-level change tests in the average, and due tolerance to various structural breaks is made, the results are consistent with the hypothesis of random walk in most markets.

Mensi *et al.* [68] show that the MF-DFA analysis shows efficiency for sectorial stock markets. In addition, the results show a high efficiency in the long term, but moderate in the short term, and that these markets became less efficient after the beginning of the global financial crisis.

Aggarwal [69] analyzed the Korean stock price index from July 1997 to September 2016. The author shows that the series do not follow a random walk model. The results have significant implications for investors, since market inefficiency can affect the domestic and international flows of an economy.

Rehman *et al.* [70] analyzed the KSE 100, S&P BSE 500 and CSE All Share indexes. Results suggest that the random walk process and the hy-

potheses of efficiency in its weak form are not verified. The concept of random walk was examined considering only stock markets, ignoring the other financial markets.

Malafeyev *et al.* [62] studied the stock markets of China and India, analyzing the impact of the 2008 global financial crisis and the recent 2015 Chinese crisis. Data for the last 20 years from the Bombay Stock Exchange (BSE200) and the Shanghai Stock Exchange Composite Index were selected and divided into four sub-periods, i.e., before the financial crisis period (period I), during the 2008 financial crisis (period II), after the subprime crisis, before the Chinese crisis (period III), and from the onset of the Chinese crisis to date (period IV). The evidence confirms that the Indian and Chinese stock market do not show market efficiency in their weak form.

Pernagallo and Torrisi [71] analyzed whether the behavior of daily returns on the stock market indexes of 12 emerging economies corroborates the hypothesis of fat tails and whether these series show long memory. Hurst's exponents oscillate between 0.51 and 0.62 highlighting significant long memories, showing similarities between the stock markets of emerging economies and developed countries.

Bariviera *et al.* [36] and Sensoy and Tabak [37, 38, 43] analyzed the efficiency of the European Union and Asian markets in the face of the Subprime and debt crises using R/S Analysis and DFA. The authors concluded that European stock markets showed different degrees of efficiency and that the Subprime crisis had an adverse effect on almost all European Union countries. However, the Eurozone debt crisis had a significant adverse effect only on the markets of France, Spain and Greece.

Other studies, such as Alvarez-Ramirez *et al.* [32], analyzed the efficiency of the Dow Jones, S&P500 and Nasdaq indexes using the DFA method. The authors found a trend in the long-term behavior of the exponent α_{DFA} . The Nasdaq index showed a significant change in persistence or antipersistence that was not observed in the Dow Jones and S&P500.

Cajueiro and Tabak [31, 34] investigated the long-term dependence on returns and volatility of 9 indexes using Hurst's exponent (R/S Analysis). The results found suggest that Hurst's exponent varies over time due to changes in the return dynamics. In addition, findings provide evidence of long memory mainly in interest rates, which should be considered in asset pricing models.

Alvarez-Ramirez *et al.* [33] and Sensoy and Hacıhasanoglu [41] examined the efficiency in the weak form in future energy markets using the DFA and

R/S Analysis respectively). In both studies, the authors concluded that future oil markets outperform others in terms of efficiency. In addition, [41] signaled that gasoline, oil and gas futures markets exhibit weak efficient behavior.

Kang *et al.* [40] examined the efficiency of exchange rates in Asian-Pacific markets using the Hurst exponent obtained from the R/S Analysis method. The results indicated that the dynamics of returns and market volatility had a long memory. In addition, the exponent Hurst showed an upward trend during the 1997 Asian currency crisis and the Subprime crisis, signaling that currency markets have become more inefficient and predictable in recent years during the financial crisis.

Ferreira *et al.* [44] analyzed the evolution of 63 European Union bank shares before and after the Subprime crisis. Through the α_{DFA} exponent. The authors concluded that the crisis had an impact on the efficiency of bank stocks in Europe. Guedes *et al.* [45] analyzed the effect of the Brexit referendum in the 28 countries of the European Union (EU). In this study, they used the α_{DFA} exponent to investigate EMH in all EU countries before and after the referendum. The authors concluded that the referendum did not change much the pattern efficiency of stock exchange indexes, except Bulgaria and Austria.

3. Data and Methodology

3.1. Data

Considering that our objective is to analyze the behavior of the returns of the stock market indexes over time, regarding descriptive statistics (mean, standard deviation and asymmetry and kurtosis coefficients) and the weak form of EMH, we take the closed index (S_t) of the 19 richest countries in the world (see Table 1 and Fig.1), with the series grouped into: *i*) G8 (richer economies) and *ii*) Emerging (emerging economies).

Table 1: G-20 indexes

Country	Group	Index	Start	End
CAN	G8	S&P TSX	Jan/02/2000	Feb/05/2021
FRA	G8	CAC 40	Jan/02/2000	Feb/05/2021
GER	G8	DAX	Jan/02/2000	Feb/05/2021
ITA	G8	FTSE MIB	Jun/04/2003	Feb/05/2021
JPN	G8	NIKKEI 225	Jan/02/2000	Feb/05/2021
RUS	G8	MOEX	Jan/02/2000	Feb/05/2021
UK	G8	FTSE 100	Jan/02/2000	Feb/05/2021
USA	G8	NYSE	Jan/02/2000	Feb/05/2021
ARG	Emerging	S&P Merval	Jan/02/2000	Feb/05/2021
AUS	Emerging	S&P ASX 200	Jan/02/2000	Feb/05/2021
BRA	Emerging	IBOVESPA	Jan/02/2000	Feb/05/2021
CHN	Emerging	SSE	Jan/02/2000	Feb/05/2021
IND	Emerging	S&P BSE SENSEX	Jan/02/2000	Feb/05/2021
INA	Emerging	JKSE	Jan/02/2000	Feb/05/2021
MEX	Emerging	S&P BMV IPC	Jan/02/2000	Feb/05/2021
KSA	Emerging	TASI	Jan/02/2000	Feb/05/2021
RSA	Emerging	FTSE JSE	Jul/04/2000	Feb/05/2021
KOR	Emerging	KOSPI	Jan/02/2000	Feb/05/2021
TUR	Emerging	BIST 100	Jan/02/2000	Feb/05/2021

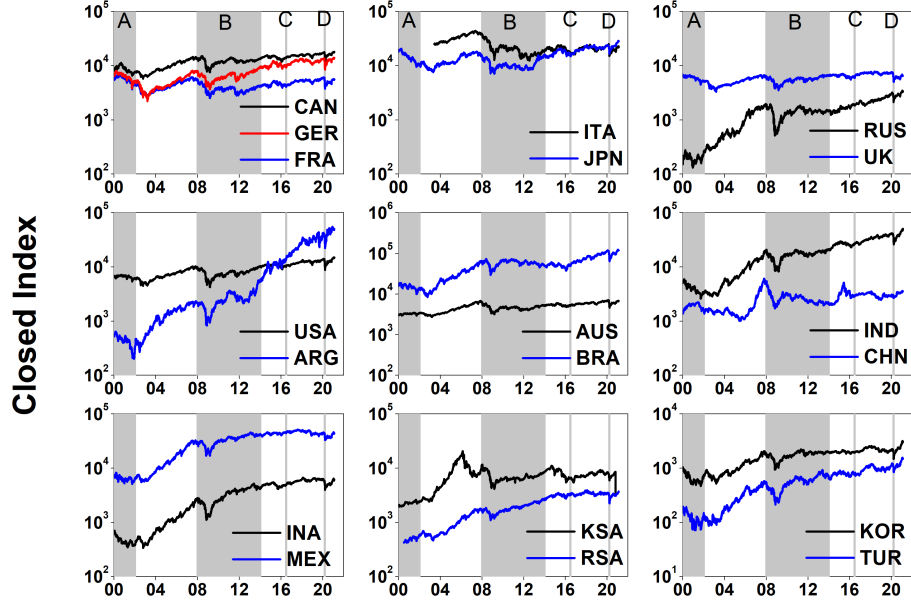


Figure 1: (Color on-line) Closed index from Jan/02/2000 to Feb/05/2021. Vertical bars in gray represent some financial crises.

From these time series, we can analyze the behavior of the indices in different crises [72], such as:

- A - Argentina triple crisis in 2001 (bank, currency and debt crises) and Dot-com crash in 2001 (bubble in telecommunications stocks and internet in retail (OTC) in the United States between 1995-2000);
- B - Subprime Crisis in 2008 (bubble in the real estate market and mortgage-backed securities) and European Sovereign Debt Crisis 2009/2013 (public debt crisis in several European Union member countries);
- C - BREXIT referendum (Jun / 23/2016) - withdrawal of the United Kingdom from the European Union;
- D - COVID-19 (Mar / 11/2020) - declaration of COVID-19 as a pandemic by the World Health Organization.

All these crises were identified in Fig. 1 by the Vertical bars in gray area.

In order to standardize the indexes for better comparison, we calculated the return of values given by

$$r_t = \ln \left(\frac{S_t}{S_{t-1}} \right) \quad (1)$$

With the return, we will obtain the descriptive statistics and the α_{DFA} exponent. For this purpose, we present a brief theoretical description of the Detrended Fluctuation Analysis (DFA) method.

3.2. Detrended Fluctuation Analysis (DFA)

Peng *et al.* [73] developed DFA to analyze the existence of serial dependence, with the advantage of being possible to be used in non-stationary data as well. Its main advantage is to avoid spurious detection of long-range dependence due to nonstationary data. For a given r_i signal, the algorithm is described as follows:

1. Define the cumulative series of original signal fluctuations about its mean, \bar{r} with N points, as:

$$X_k = \sum_{i=1}^k (r_i - \bar{r}) \text{ with } k = 1, 2, \dots, N; \quad (2)$$

2. Divide X_k into boxes of equal length n (time scale);
3. In each box, fit the local trend of X_k by a polynomial function $X_{n,i}$, which represents the local trend of the box;
4. For the given n box size, compute the root-mean-squared detrended fluctuation of the signal X_k as:

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

5. The last step is repeated for each box n ($4 \leq n \leq \frac{N}{4}$) (time scale) to obtain F_{DFA} as a function of n .
6. If long-range autocorrelation appear in this time series, then

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

where:

- $\alpha_{DFA} < 0.5$ identifies long range anti-persistent behavior;
- $\alpha_{DFA} > 0.5$ identifies long range persistent behavior;
- $\alpha_{DFA} = 0.5$ is the signature of a uncorrelated signal;

The advantage of the DFA over other methods lies in allowing the detection of long-range autocorrelations embedded in non-stationary time series, and in avoiding spurious detection of apparent long-range autocorrelations, which are an artifact of nonstationarity [45, 74, 75].

3.3. Detrended Fluctuation Analysis with sliding windows approach

To analyze the dynamics of α_{DFA} , we applied the sliding windows approach with 1000 points of window of length (w) [44–46], see Figure 2 for this representation. According to the weak form of the Efficient Market Hypothesis, it is expected that stock prices are random, so return rates should behave like a white noise ($\alpha_{DFA} = 0.5$).

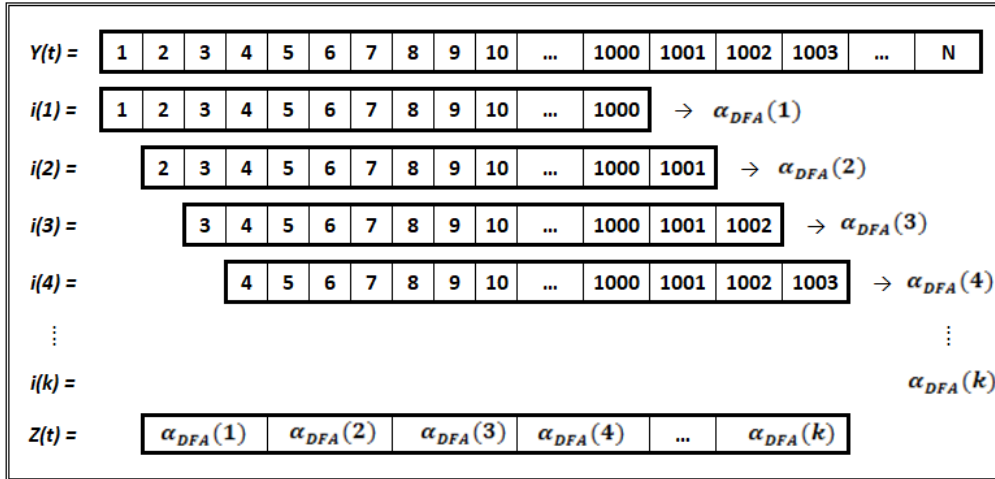


Figure 2: Sliding windows procedure with $w = 1000$.

In addition we computed the mean, standard deviation and the asymmetry and kurtosis coefficients (descriptive statistics), both estimated by the maximum likelihood method, with the asymmetry and kurtosis coefficients denoted by

$$Skewness = \frac{1}{w} \sum_{i=1}^w \left(\frac{r_i - \bar{r}}{\sigma} \right)^3 \quad (5)$$

$$Kurtosis = \frac{1}{w} \sum_{i=1}^w \left(\frac{r_i - \bar{r}}{\sigma} \right)^4 \quad (6)$$

The function that implements the aforementioned descriptive statistics and the DFA method are available in R language in the SlidingWindows package [76]. In the next section, results of our analysis are presented.

4. Results

4.1. G8 stock indexes

Initially we started the analysis using descriptive statistics, as shown in Fig.3. In this figure we can see mean (A), standard deviation (B), skewness (C), and kurtosis (D). We can see for gray area a more volatile behavior, mainly in the pandemic case.

During the Subprime and sovereign debt crises, the trend was reversed, going from negative to positive (from the point of view of the average value of returns). For example, RUS showed greater prominence in relation to the average values before and during the Subprime and sovereign debt crises. From 2014 to 2020, in general, indexes were stabilized. With the COVID-19 pandemic, indices fell again and, soon after, stabilized at the same level of the 2014 to 2020 period.

Fig.3 (B) shows that the standard deviation (risk) of returns was high in 2008 and 2014, having dropped in 2014 to 2020, that leads us to conclude that its volatility has decreased.

We can see that the asymmetry in returns is more evident in 2020, in the pandemic of COVID-19, mainly in CAN (Fig.3 (C)). The kurtosis of returns is always leptocurtics ($k > 3$), mainly in the Subprime crises in 2008 and the COVID-19 (Fig.3 (D)). These results pointed out by the descriptive statistics show us non-normality of returns and still provide empirical evidence about long tail distributions.

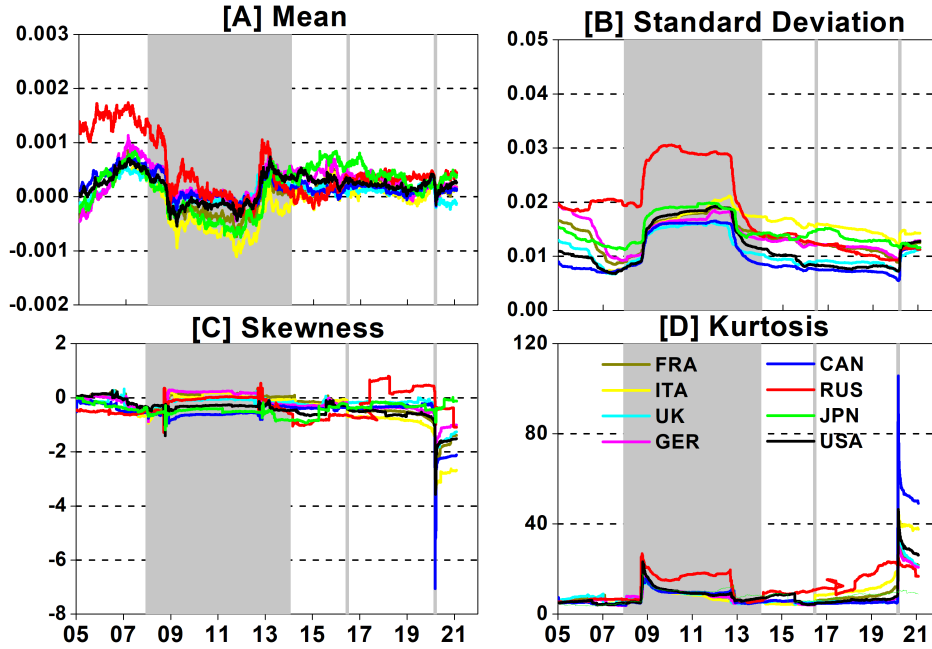


Figure 3: (Color on-line) Descriptive statistics with sliding windows length $w = 1000$. The shaded area represents the time between crises.

From the point of view of autocorrelation by DFA method, Fig. 4 show the α_{DFA} exponent as a function of time. Therefore, considering the weak efficiency hypothesis, in general we found that most indexes exhibited an α_{DFA} exponent below 0.5, excluding the pandemic period with $\alpha_{DFA} > 0.5$. Specifically, if we analyze from 2005 to 2008, we noticed that, the value of α_{DFA} is predominantly greater than 0.5 for RUS and CAN. For 2021, there is a strong tendency for exponents above 0.5.

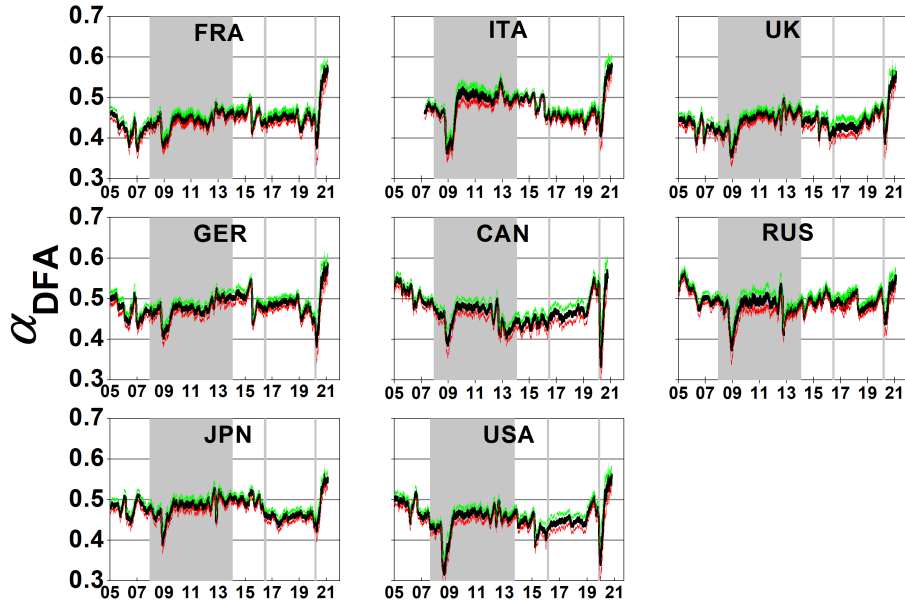


Figure 4: (Color on-line) α_{DFA} as a function of time with sliding windows length $w = 1000$. The shaded area represents the time between crises. Green and red lines represent 95% of confidence interval.

4.2. Emerging stock indexes

In the case of other markets (emerging markets), it is possible to observe that before the Subprime and Public Debt crisis the mean of return Fig.4 (A) was positive, except for the CHN index. After the Subprime and Sovereign Debt crises, all indexes showed positive trend or were closer to zero, except ARG index.

Regarding the standard deviation Fig.5 (B), we found that the volatility was increased in the crisis of 2008 and 2014, having dropped after this time period, which leads us to conclude that its volatility has decreased, except for the case of ARG index.

From the point of view of skewness Fig.5 (C), the returns is more evident before 2008 in the CHN index and in the IND index in the Subprime,

Sovereign Debt crises and the COVID-19 pandemic. Fig.5 (D) shows that the kurtosis of returns is always leptocurtic ($k > 3$), mainly in the Subprime, Sovereign Debt crises and the COVID-19 pandemic crisis, in the IND and KSA indexes. The results presented by descriptive statistics also point to a non-normal returns and still provide empirical evidence about long tail distributions.

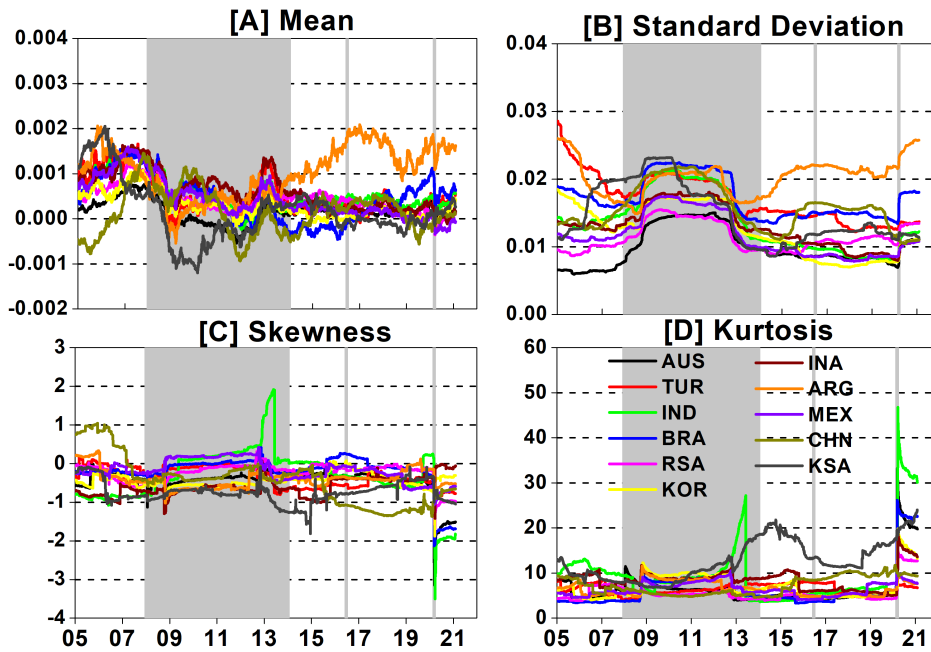


Figure 5: (Color on-line) Descriptive statistics with sliding windows length $w = 1000$. The shaded area represents the time between crises.

Now, analyzing the autocorrelation with DFA method and α_{DFA} exponent in general it is noticed that in relation to the EMH hypothesis, all indexes had exponents above 0.5 around 2005, see Fig.6.

Between 2008 and 2014, the value of α_{DFA} can be greater than, less than, or approximately 0.5, depending on the index. For example, ARG, TUR, INA, KSA and KOR indexes a persistent behavior with $\alpha_{DFA} > 0.5$. However, AUS, BRA, RSA and MEX indexes showed antipersistent behavior

with exponent $\alpha_{DFA} < 0.5$. On the other hand, IND and CHN indexes exhibited an mean value of α_{DFA} closer to 0.5.

In general, the case of the pandemic crisis, emerging markets performed similarly to G8 indexes, except for ARG and CHN indexes.

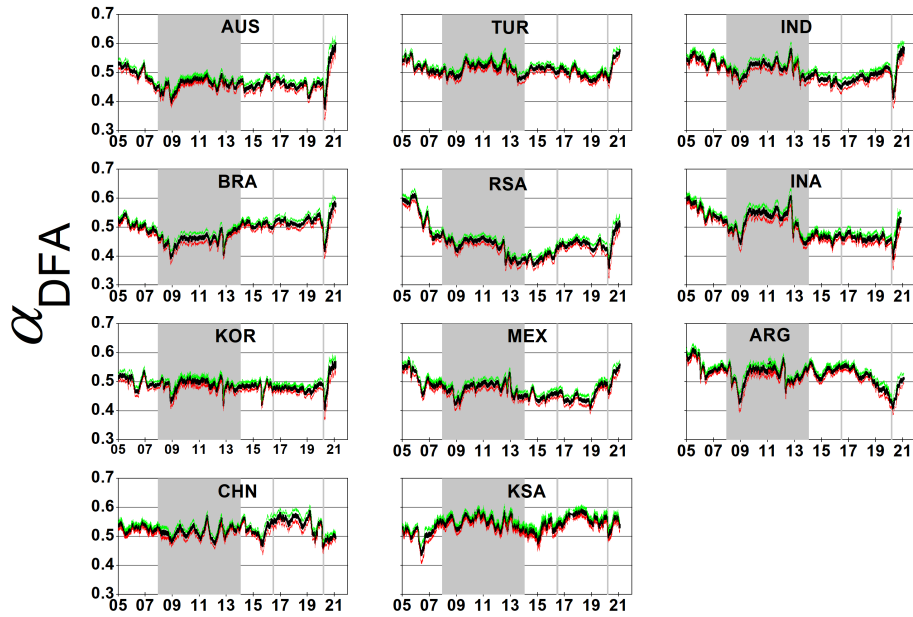


Figure 6: (Color on-line) α_{DFA} as a function of time with sliding windows length $w = 1000$. The shaded area represents the time between crises. Green and red lines represent 95% of confidence interval.

5. Conclusions

We have tested whether the financial crises of the past 20 years have reduced efficiency, in its weak form, in the 19 stock markets belonging to the 20 most developed economies (G-20). For this purpose, we applied the Detrended Fluctuation Analysis (DFA) methodology, with sliding windows, to check if the markets under analysis are persistent, antipersistent or random.

In general, we observed that the analyzed stock markets show non-Gaussian asymmetries and distributions (long tails), with small differences between periods of crisis (extreme volatility) and non-crisis periods. Most DFA exponents have compound tendencies, that is, long memories (persistence) or short-term memories (anti-persistence).

Regarding GER, JPN, TUR and KOR when comparing them with other markets, we find that these stock markets have an average tendency towards equilibrium. These findings are not beneficial for investors, since arbitrage operations are significantly reduced, that is, investors will have difficulties in obtaining returns above the market average without incurring additional risk.

The general conclusion to be retained and sustained by the results obtained, through tests carried out with statistical and econophysical models, demonstrates that financial crises, in general, have significant impact on the memory properties of the analyzed markets. The α_{DFA} exponents suggest that the random walk hypothesis is mostly rejected, except for GER, JPN, TUR and KOR stock indexes, which show average tendencies for equilibrium. The presence of persistence in returns may be beneficial for investors because, if they adjust their trading strategies with the necessary lags, they will be able to obtain returns above the market average without incurring additional risk. However, it was not the object of this study to analyze anomalous returns. These conclusions also open space for market regulators to take steps to ensure better information among international financial markets. We believe that investors should diversify their portfolios and invest in less risky markets, in order to mitigate risk and improve the efficiency of their portfolios.

6. References

References

- [1] E. Fama, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25 (2) (1970) 383–417.
- [2] E. F. Fama, Efficient capital markets: II, *The Journal of Finance* 46 (5) (1991) 1575 – 1617.
- [3] J. Dowd, G. Gorton, Stock market efficiency and economic efficiency: Is there a connection?, *The Journal of Finance* 52 (3) (1997) 4–6.

- [4] B. G. Malkiel, The efficient market hypothesis and its critics, *Journal of Economic Perspectives* 17 (1) (2003) 59–82.
- [5] R. J. Shiller, From efficient markets theory to behavioral finance, *Journal of Economic Perspectives* 17 (1) (2003) 83–104.
- [6] M. Sewell, The efficient market hypothesis: Empirical evidence, *International Journal of Statistics and Probability* 1 (2) (2012) 164.
- [7] A. G. Titan, The efficient market hypothesis: Review of specialized literature and empirical research, *Procedia Economics and Finance* 32 (2015) 442–449.
- [8] J. Liu, Q. Song, Y. Qi, S. Rahman, , S. Sriboonchitta, Measurement of systemic risk in global financial markets and its application in forecasting trading decisions, *Sustainability* 12 (10) (2020) 4000.
- [9] S. G. Fifield, J. Jetty, Further evidence on the efficiency of the Chinese stock markets: A note, *Research in International Business and Finance* 22 (3) (2008) 351 – 361.
- [10] A. Beltratti, B. Bortolotti, M. Caccavaio, Stock market efficiency in China: Evidence from the split-share reform, *The Quarterly Review of Economics and Finance* 60 (2016) 125 – 137.
- [11] K. R. French, Stock returns and the weekend effect, *Journal of Financial Economics* 8 (1) (1980) 55 – 69.
- [12] R. H. Thaler, Anomalies: The january effect, *Economic Perspectives* 1 (1) (1987) 197–201.
- [13] R. A. Ariel, High stock returns before holidays: Existence and evidence on possible causes, *The Journal of Finance* 45 (5) (1990) 1611–1626.
- [14] A. Agrawal, K. Tandon, Anomalies or illusions? evidence from stock markets in eighteen countries, *Journal of International Money and Finance* 13 (1) (1994) 83 – 106.
- [15] F. Guidi, R. Gupta, S. Maheshwari, Weak-form market efficiency and calendar anomalies for Eastern Europe equity markets, *Journal of Emerging Market Finance* 10 (3) (2011) 337–389.

- [16] T. Verheyden, L. De Moor, F. Van Den Bossche, A tale of market efficiency, *Review of Business and Economic Literature* 58 (2013) 140 – 158.
- [17] A. Todea, D. Lazar, Global crisis and relative efficiency: empirical evidence from Central and Eastern European stock markets, *The Review of Finance and Banking* 4 (1) (2012) 045 – 053.
- [18] A. W. Lo, Long-term memory in stock market prices, *Econometrica* 59 (5) (1991) 1279–1313.
- [19] Y.-W. Cheung, K. S. Lai, A search for long memory in international stock market returns, *Journal of International Money and Finance* 14 (4) (1995) 597 – 615.
- [20] I. N. Lobato, C. Velasco, Long memory in stock-market trading volume, *Journal of Business & Economic Statistics* 18 (4) (2000) 410–427.
- [21] S. Sadique, P. Silvapulle, Long-term memory in stock market returns: international evidence, *International Journal of Finance & Economics* 6 (1) (2001) 59–67.
- [22] F. Lillo, J. D. Farmer, The long memory of the efficient market, *Studies in Nonlinear Dynamics & Econometrics* 8 (3) (2004) 1–35.
- [23] D. Sornette, *Why Stock Markets Crash Critical Events in Complex Financial Systems*, Princeton University Press, 2003.
- [24] T. Kaizoji, D. Sornette, Market bubbles and crashes, *Munich Personal RePEc Archive* (40798) (2008) 1 – 25.
- [25] W. A. Risso, The informational efficiency and the financial crashes, *Research in International Business and Finance* 22 (3) (2008) 396 – 408.
- [26] S. Galam, Market efficiency, anticipation and the formation of bubbles-crashes, *arXiv preprint arXiv:1106.1577* (2011).
- [27] Z.-Q. Jiang, W.-J. Xie, W.-X. Zhou, Testing the weak-form efficiency of the WTI crude oil futures market, *Physica A* 405 (2014) 235–244.
- [28] Z.-Q. Jiang, W.-J. Xie, W.-X. Zhou, D. Sornette, Multifractal analysis of financial markets: a review 82 (12) (2019) 125901.

- [29] J. Sukpitak, V. Hengpunya, Efficiency of Thai stock markets: Detrended fluctuation analysis, *Physica A* 458 (2016) 204 – 209.
- [30] A. Carbone, G. Castelli, H. Stanley, Time-dependent Hurst exponent in financial time series, *Physica A* 344 (2004) 267 – 271.
- [31] D. O. Cajueiro, B. M. Tabak, Testing for time-varying long-range dependence in volatility for emerging markets, *Physica A* 346 (3) (2005) 577–588.
- [32] J. Alvarez-Ramirez, J. Alvarez, E. Rodriguez, G. Fernandez-Anaya, Time-varying Hurst exponent for US stock markets, *Physica A* 387 (2008) 6159 – 6169.
- [33] J. Alvarez-Ramirez, J. Alvarez, E. Rodriguez, Short-term predictability of crude oil markets: A detrended fluctuation analysis approach, *Energy Economics* 30 (5) (2008) 2645 – 2656.
- [34] D. O. Cajueiro, B. M. Tabak, Testing for time-varying long-range dependence in real state equity returns, *Chaos, Solitons & Fractals* 38 (1) (2008) 293–307.
- [35] D. O. Cajueiro, P. Gogas, B. M. Tabak, Does financial market liberalization increase the degree of market efficiency? the case of the Athens stock exchange, *International Review of Financial Analysis* 18 (1 – 2) (2009) 50 – 57.
- [36] A. Bariviera, M. B. Guercio, L. B. Martinez, A comparative analysis of the informational efficiency of the fixed income market in seven European countries, *Economics Letters* 116 (3) (2012) 426 – 428.
- [37] A. Sensoy, Time-varying long range dependence in market returns of FEAS members, *Chaos, Solitons & Fractals* 53 (2013) 39–45.
- [38] A. Sensoy, B. M. Tabak, How much random does European Union walk? a time-varying long memory analysis, *Working Paper Series* (342) (2013) 1–31.
- [39] C. Guangxi, H. Yan, C. Weijun, Time-varying long memories of the Chinese currency and stock markets based on the Hurst exponent, *Fluctuation and Noise Letters* 13 (01) (2014) 1450007.

- [40] S. H. Kang, R. McIver, S.-Y. Park, S.-M. Yoon, Long memory features evolve in the time-varying process in Asia-Pacific foreign exchange markets, *Procedia Economics and Finance* 14 (2014) 286–294.
- [41] A. Sensoy, E. Hacihasanoglu, Time-varying long range dependence in energy futures markets, *Energy Economics* 46 (2014) 318–327.
- [42] A. Sensoy, B. M. Tabak, Time-varying long term memory in the European Union stock markets, *Physica A* 436 (2015) 147–158.
- [43] A. Sensoy, B. M. Tabak, Dynamic efficiency of stock markets and exchange rates, *International Review of Financial Analysis* 47 (2016) 353 – 371.
- [44] P. Ferreira, A. Dionisio, E. F. Guedes, G. F. Zebende, A sliding windows approach to analyse the evolution of bank shares in the European Union, *Physica A* 490 (2018) 1355 – 1367.
- [45] E. Guedes, P. Ferreira, A. Dionísio, G. Zebende, An econophysics approach to study the effect of BREXIT referendum on European Union stock markets, *Physica A* 523 (2019) 1175 – 1182.
- [46] E. Guedes, G. Zebende, DCCA cross-correlation coefficient with sliding windows approach, *Physica A* 527 (2019) 121286.
- [47] O. Tilfani, P. Ferreira, M. Y. E. Boukfaoui, Dynamic cross-correlation and dynamic contagion of stock markets: a sliding windows approach with the DCCA correlation coefficient, *Empirical Economics* 60 (2019) 1127 – 1156.
- [48] G. Gibson, *The Stock Markets of London, Paris and New York*, G.P. Putnam’s Sons., New York, 1889.
- [49] L. Bachelier, *Théorie de la spéculation* [ph.d. thesis in mathematics], *Annales Scientifiques de l’École Normale Supérieure* 17 (1900) 21–86.
- [50] K. Pearson, The problem of the random walk, *Nature* 72 (1905) 318.
- [51] A. Cowles, Can stock market forecasters forecast?, *Econometrica* 1 (3) (1933) 309 – 324.

- [52] H. Working, The investigation of economic expectations, *The American Economic Review* 39 (3) (1949) 150–166.
- [53] M. G. Kendall, A. B. Hill, The analysis of economic time-series-part i: Prices, *Journal of the Royal Statistical Society. Series A (General)* 116 (1) (1953) 11–34.
- [54] H. Working, A random-difference series for use in the analysis of time series, *Journal of the American Statistical Association* 29 (185) (1934) 11–24.
- [55] H. V. Roberts, Stock-market "patterns" and financial analysis: Methodological suggestions, *The Journal of Finance* 14 (1) (1959) 1–10.
- [56] A. Cowles, A revision of previous conclusions regarding stock price behavior, *Econometrica* 28 (4) (1960) 909–915.
- [57] A. Cowles, H. E. Jones, Some a posteriori probabilities in stock market action, *Econometrica* 5 (3) (1937) 280–294.
- [58] E. F. Fama, Random walks in stock market prices, *Financial Analysts Journal* 21 (5) (1965) 55–59.
- [59] L. H. Summers, Does the stock market rationally reflect fundamental values?, *The Journal of Finance* 41 (1986) 591–601.
- [60] E. F. Fama, K. R. French, Dividend yields and expected stock returns, *Journal of Financial Economics* 22 (1) (1988) 3–25.
- [61] J. M. Poterba, L. H. Summers, Mean reversion in stock prices: Evidence and implications, *Journal of Financial Economics* 22 (1) (1988) 27 – 59.
- [62] O. Malafeyev, A. Awasthi, K. Kambekar, A. Kupinskaya, Random walks and market efficiency in Chinese and Indian equity markets, *Statistics, Optimization & Information Computing* 7 (2019) 1 – 25.
- [63] K. Hamid, M. T. Suleman, S. Z. A. Shah, R. S. I. Akash, Testing the weak form of efficient market hypothesis: Empirical evidence from Asia-Pacific markets, *International Research Journal of Finance and Economics* (2010) 121 – 133.

- [64] S. Nisar, M. Hanif, Testing weak form of efficient market hypothesis: Empirical evidence from South Asia, *World Applied Sciences Journal* 17 (4) (2011) 414 – 427.
- [65] S. Mehla, S. K. Goyal, Empirical evidence on weak form of efficiency in Indian stock market, *Asia-Pacific Journal of Management Research and Innovation* 8 (1) (2012) 59 – 68.
- [66] H. Shirvani, N. V. Delcoure, The random walk in the stock prices of 18 OECD countries: Some robust panel-based integration and cointegration tests, *Journal of Economic Studies* 43 (4) (2016) 598 – 608.
- [67] G. Ngene, K. A. Tah, A. F. Darrat, The random-walk hypothesis revisited: new evidence on multiple structural breaks in emerging markets, *Macroeconomics and Finance in Emerging Market Economies* 10 (1) (2017) 88–106.
- [68] W. Mensi, A. K. Tiwari, S.-M. Yoon, Global financial crisis and weak-form efficiency of islamic sectoral stock markets: An MF-DFA analysis, *Physica A* 471 (2017) 135 – 146.
- [69] D. Aggarwal, Random walk model and asymmetric effect in Korean composite stock price index, *Afro-Asian Journal of Finance and Accounting* 8 (1) (2018) 85.
- [70] R. S. S., D. P. Pompeo, The Impact of Brexit on EU27 on Trade, Investments and Financial Services, *Global Economy Journal* 18 (1) (2018) 1–17.
- [71] G. Pernagallo, B. Torrisi, An empirical analysis on the degree of gaussianity and long memory of financial returns in emerging economies, *Physica A* 527 (2019) 121296.
- [72] G. Zanalda, Financial crises, history of, in: J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition), second edition Edition, Elsevier, Oxford, 2015, pp. 183 – 190.
- [73] 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.

- [74] 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.
- [75] G. F. Zebende, F. M. Oliveira-Filho, J. Leyva-Cruz, Auto-correlation in the motor/imaginary human EEG signals: A vision about the F_{DFA} fluctuations, PLOS ONE 12 (2017) e0183121.
- [76] E. F. Guedes, I. C. da Cunha Lima, G. F. Zebende, A. M. Silva-Filho, SlidingWindows: Methods for Time Series Analysis, R package version 0.1.4 (2020).
URL <https://CRAN.R-project.org/package=SlidingWindows>

Referências Bibliográficas

- ABNT. *Pesquisa de sinistros de trânsito - Terminologia*. Terceira edição. [S.l.], 2020. [2.1](#)
- ALKHADOUR, W. et al. Traffic accidents detection using geographic information systems (gis). *Traffic*, v. 12, n. 4, 2021. [3.3.2](#)
- ANDRADE, F. R. de; ANTUNES, J. L. F. Tendência do número de vítimas em acidentes de trânsito nas rodovias federais brasileiras antes e depois da década de ação pela segurança no trânsito. *Cadernos de Saúde Pública*, SciELO Public Health, v. 35, p. e00250218, 2019. [2.3.1](#)
- ANDRADE, F. R. de; ANTUNES, J. L. F. Distracted driving of a motor vehicle on Brazilian federal highways as a cause of traffic accidents. *Revista Brasileira de Epidemiologia*, SciELO Brasil, v. 23, 2020. [2.3.1](#)
- ANDRADE, S. S. C. d. A.; MELLO-JORGE, M. H. P. de. Mortalidade e anos potenciais de vida perdidos por acidentes de transporte no brasil, 2013. *Revista de Saúde Pública*, SciELO Brasil, v. 50, 2016. [2.3.1](#)
- ARSAC, L. M.; DESCHODT-ARSAC, V. Detrended fluctuation analysis in a simple spreadsheet as a tool for teaching fractal physiology. *Advances in physiology education*, American Physiological Society Bethesda, MD, v. 42, n. 3, p. 493–499, 2018. [3.3.4](#)
- AZEVEDO, S.; SABA, H.; MORET, M. Auto-afinidade na série temporal da dengue. SENAI/CIMATEC, 2014. [3.3.4](#)
- BACCHIERI, G.; BARROS, A. Acidentes de trânsito no Brasil de 1998 a 2010: muitas mudanças e poucos resultados. *Revista de Saúde Pública*, SciELO Brasil, v. 45, p. 949–963, 2011. [2.3.1](#)
- BAHADORIMONFARED, A. et al. Trends of fatal road traffic injuries in iran (2004–2011). *PloS one*, Public Library of Science, v. 8, n. 5, p. e65198, 2013. [1](#), [2.3.1](#)
- BAR-YAM, Y. General features of complex systems. *Encyclopedia of Life Support Systems (EOLSS)*, UNESCO, EOLSS Publishers, Oxford, UK, v. 1, 2002. [2.3.1](#)
- BARROS, A. J. et al. Acidentes de trânsito com vítimas: sub-registro, caracterização e letalidade. *Cadernos de Saúde Pública*, SciELO Public Health, v. 19, p. 979–986, 2003. [2.3.1](#), [4.1](#)
- BASTOS, A. d. S. C.; OLIVEIRA, D. d. M.; BARRETO, T. G. Mortalidade por acidentes de trânsito no estado da bahia. *Rev. Eletrônica Atualiza Saúde— Salvador*, v. 3, n. 3, p. 15–23, 2016. [2.3.1](#)
- BETGAMASCHI, R. B. *SIG aplicado a segurança no Trânsito - estudo de caso no município de Vitória-ES*. Dissertação (Mestrado) — Departamento de Geografia do Centro de Ciências Humanas e Naturais da UFES, 2010. [3.3.2](#)
- BLACK, R.; HURST, H.; SIMAIKA, Y. *Long-term storage: an experimental study*. [S.l.]: Constable, 1965. [3.3.3](#)

- BLAZQUEZ, C. A.; CELIS, M. S. A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. *Accident Analysis & Prevention*, Elsevier, v. 50, p. 304–311, 2013. [3.3.2](#)
- BOATENG, F. G. Why africa cannot prosecute (or even educate) its way out of road accidents: insights from ghana. *Humanities and Social Sciences Communications*, Palgrave, v. 8, n. 1, p. 1–11, 2021. [1](#)
- BRITO, A. de A. et al. Métodos estatísticos aplicados a temperatura ambiente e umidade relativa do ar na cidade de porto seguro (ba). *Métodos Estatísticos Aplicados a Temperatura Ambiente e Umidade relativa do Ar na Cidade de Porto Seguro (BA)*, Atena Editora, p. 1–388. [3.3.5](#)
- CARMO, r. A.; NERY, A. A.; ROCHA, R. M. Repercussões dos acidentes de trânsito: uma revisão integrativa. *Revista de Pesquisa: Cuidado é Fundamental*, p. 732–738, 2019. [1](#)
- CARVALHO, R. C. de. Acidentes de trânsito envolvendo adolescentes: o registro da situação de Feira de Santana, Bahia, em 2001. *Revista Baiana de Saúde Pública*, v. 29, n. 1, p. 22–22, 2005. [1](#), [2.3.1](#)
- CÂMARA, G.; DAVIS, C.; MONTEIRO, A. M. V. Introdução à ciência da geoinformação. 2001. [3.3.2](#)
- CNT. *Pesquisa CNT de Rodovias*. [S.l.], 2019. [1](#)
- COMMANDEUR, J. J. et al. On statistical inference in time series analysis of the evolution of road safety. *Accident Analysis & Prevention*, Elsevier, v. 60, p. 424–434, 2013. [1](#)
- COSTA, D. C. de S.; MORAES, R. M. D. Análise dos acidentes de trânsito utilizando técnicas e análise espacial. 2009. [3.3.2](#)
- DAI, D. Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a gis environment. *Journal of Transport Geography*, Elsevier, v. 24, p. 206–214, 2012. [3.3.2](#)
- DIAS, R. et al. Long memory in stock returns: Evidence from the eastern european markets. In: EDP SCIENCES. *SHS Web of Conferences*. [S.l.], 2021. v. 91. [3.3.4](#), [3.1](#)
- DIESEL, L. E. *SIG na prevenção em acidentes de trânsito*. Dissertação (Mestrado) — Universidade Federal de Santa Catarina UFPR, 2005. [3.3.2](#)
- EBC. *PRF lança documento para simplificar registro de acidentes sem vítimas*. 2015. <https://agenciabrasil.ebc.com.br/geral/noticia/2015-06/prf-lanca-documento-para-simplificar-registro-de-acidentes-sem-vitimas>. Acessado em 18/08/2022. [4.1](#)
- EHLERS, R. S. *Análise de Séries Temporais*. [S.l.]: Universidade do Paraná, 2007. [3.3.1](#), [3.3.1](#), [3.3.1](#), [4.2.1](#)
- FEBRES, J. D. et al. Influence of seat-belt use on the severity of injury in traffic accidents. *European transport research review*, Springer, v. 12, n. 1, p. 9, 2020. [1](#)
- FEDER, J. *Fractals (Physics of solids and liquids)*. [S.l.]: Library of Congress Cataloging in Publication Data, 1988. ISBN 978-1-4899-2126-0. [3.3.3](#)

- FEI, G. et al. Effectiveness of implementing the criminal administrative punishment law of drunk driving in china: An interrupted time series analysis, 2004-2017. *Accident Analysis & Prevention*, Elsevier, v. 144, p. 105670, 2020. [2.3.1](#)
- FENG, M.; WANG, X.; QUDDUS, M. Developing multivariate time series models to examine the interrelations between police enforcement, traffic violations, and traffic crashes. *Analytic methods in accident research*, Elsevier, v. 28, p. 100139, 2020. [2.3.1](#)
- FERREIRA, L. N. *Acidente de trânsito e seus efeitos na qualidade de vida e funcionalidade de condutores de veículos motorizados terrestres - estudo longitudinal de base populacional. Tese (Doutorado) apresentada ao Programa de Pós-graduação / Orientador: Dr. Eduardo Mota.* Tese (Doutorado) — Universidade Federal da Bahia, 2016. [2.3.1](#)
- FERREIRA, P. C. P. *Uma proposta de tratamento dos dados sobre acidentnes nas rodovias federais brasileiras com vistas a atuar preventivamente na redução de sua ocorrência. Tese (Doutorado) apresentada ao Programa de Pós-graduação em Engenharia de Transportes / Orientador: Carlos David Nassi.* Rio de Janeiro: UFRJ. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2017. [2.3.1](#)
- FIGUEIREDO, B. C. L. de. *Análise multifractal da velocidade do vento em Pernambuco.* Dissertação (Mestrado) — Universidade Federal Rural de Pernambuco, 2014. [3.3.4](#)
- FILHO, A. M. *Correlação de longo alcance em indicadores de criminalidade de Salvador-BA: DFA e DCCA. Dissertação (Mestrado) apresentada ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial / Orientador: Dr. Gilney Figueira Zebende.* Salvador: SENAI CIMATEC. Dissertação (Mestrado) — Faculdade de Tecnologia SENAI CIMATEC, 2009. [3.3.1](#), [3.3.4](#), [4.2.1](#), [4.2.2.1](#)
- FILHO, A. M. d. S. et al. Correlação cruzada da dengue nos três biomas da Bahia. *Congresso Latinoamericano de Biomatemática*, 2015. [3.3.6](#)
- FILHO, A. M. S. *Autocorrelação e correlação cruzada: teoria e aplicações. Tese (Doutorado) apresentada ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial / Orientador: Dr. Gilney Figueira Zebende.* Salvador: SENAI CIMATEC. Tese (Doutorado) — SENAI CIMATEC, 2014. [3.3.4](#)
- FILHO, O. *Amplitude da flutuação e correlação cruzada em sinais eletroencefalográficos: uma modelagem com a função de flutuação rms e o coeficiente de correlação cruzada rhoDCCA.* Dissertação (Mestrado) — CENTRO UNIVERSITÁRIO SENAI CIMATEC, 2019. [3.3.6](#)
- FRANÇA, A. M. *Diagnóstico dos acidentes de trânsito nas rodovias estaduais de Santa Catarina utilizando um sistema de informação geográfica.* Dissertação (Mestrado) — UFSC, 2008. [3.3.2](#)
- FULLER, W. A. *Introduction to statistical time series.* [S.l.]: John Wiley & Sons, 2009. v. 428. [3.3.1](#)
- GARCIA, L. Sistemas complexos: uma nova abordagem para a promoção da atividade física em populações. *Revista Brasileira de Atividade Física & Saúde*, v. 22, n. 6, p. 499–500, 2017. [1](#)
- GREENE, W. H. *Econometric Analysis.* [S.l.]: Prentice Hall, New Jersey, 2002. [3.3.1](#)

- GUEDES, E. et al. Efficiency and long-range correlation in g-20 stock indexes: A sliding windows approach. *Fluctuation and Noise Letters*, World Scientific, p. 2250033, 2022. [3.3.4](#), [4.3](#)
- GUEDES, E. F. *Estudo do Comportamento da Produção da Indústria de Transformação. Dissertação (Mestrado) apresentada ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial / Orientador: Dr. Gilney Figueira Zebende. Salvador: Faculdade SENAI CIMATEC. Dissertação (Mestrado) — SENAI CIMATEC, 2014.* [3.3.1](#), [3.3.1](#), [3.3.1](#), [3.3.6](#), [4.2.1](#)
- GUJARATI, D. N.; PORTER, D. C. *Econometria Básica*. [S.l.]: 5 ed. São Paulo, 2011. [3.3.1](#)
- HU, K. et al. Effect of trends on detrended fluctuation analysis. *Physical Review E*, APS, v. 64, n. 1, p. 011114, 2001. [3.3.4](#)
- IBGE. *Gerência de Estudos e Análises da Dinâmica Demográfica: Projeção da População por Sexo e Idade para o Período 2000 - 2030 (Revisão 2018)*. 2020. [Www.ibge.gov.br](http://www.ibge.gov.br). Acessado em 18/08/2022. [3.2](#)
- IVAJSNŠIČ, D. et al. Revealing the spatial pattern of weather-related road traffic crashes in slovenia. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 11, n. 14, p. 6506, 2021. [2.3.1](#)
- IVAJSNŠIČ, D. et al. A spatial decision support system for traffic accident prevention in different weather conditions. *Acta geographica Slovenica*, v. 61, n. 1, p. 75–92, 2021. [3.3.2](#)
- JARQUE, C. M.; BERA, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, Elsevier, v. 6, n. 3, p. 255–259, 1980. [3.3.1](#)
- JIANJUN, W.; HUIJUN, S. Detrended fluctuation analysis of mixed traffic flow time series with a blockage. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 43, n. 3, p. 487–494, 2014. [2.3.1](#), [4.2.1](#)
- JIN, L. et al. Analysis on the influencing factors of driving behaviours based on theory of planned behaviour. *Advances in Civil Engineering*, Hindawi, v. 2021, 2021. [1](#)
- JORGE, M. H. P. d. M.; LATORRE, M. R. Acidentes de trânsito no brasil: dados e tendências. *Cadernos de Saúde Pública*, SciELO Public Health, v. 10, p. S19–S44, 1994. [1](#)
- JUNIOR, G. T. B.; BERTHO, A. C. S.; VEIGA, A. d. C. A letalidade dos acidentes de trânsito nas rodovias federais brasileiras em 2016. *Revista Brasileira de Estudos de População*, SciELO Brasil, v. 36, 2019. [4.1](#)
- KAHN, T. *Estatística de criminalidade: manual de interpretação*. São Paulo, 2005. [4.1](#)
- LACZKOWSKI, I. M. L. M. S. Aplicacao do metodo dfa na identificacao de correlacoes e fractalidades em series temporais sonoras. *VIII Encontro de Producao Cientifica e Tecnologica*, 2013. [3.3.4](#)
- LAKATOS, E. M.; MARCONE, M. de A. *Fundamentos da metodologia científica*. [S.l.]: Camara Brasileira do Livro, SP, Brasil, 2003. [3.4](#)

LI, X.; WU, L.; YANG, X. Exploring the impact of social economic variables on traffic safety performance in hong kong: A time series analysis. *Safety science*, Elsevier, v. 109, p. 67–75, 2018. [2.3.1](#)

LIMA, N. et al. Correlação de longo alcance temporal da velocidade do vento nos municípios de ceará-mirim e natal no rio grande do norte. *HOLOS*, Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte, v. 8, p. 56–64, 2017. [3.3.4](#)

LIN, A.; SHANG, P.; ZHAO, X. The cross-correlations of stock markets based on dcca and time-delay dcca. *Nonlinear Dynamics*, Springer, v. 67, n. 1, p. 425–435, 2012. [3.3.5](#)

LJUBISAVLJEVIC, S. B. S. M. D. S. M. Detrended fluctuation analysis of time series of a ring fusimotor neuron. *Physica A*, 1999. [3.3.4](#)

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, v. 65, n. 2, p. 297–303, 1978. [3.3.1](#)

LOBATO, C. *Análise dos acidentes nas rodovias federais concedidas no Brasil. Dissertação (Mestrado) apresentada ao Curso de Mestrado em Geotecnia e Transportes / Orientador: Antônio Artur de Souza, Ph.D. Belo Horizonte: UFMG. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, 2018.* [2.3.1](#)

LUCAMBIO, F. Diferentes testes para verificar normalidade de uma amostra aleatória. *Statistic Research of Paraná. e*, v. 1, p. 1–12, 2008. [3.3.1](#)

MALIN, F.; SILLA, A.; MLADENOVIC, M. N. Prevalence and factors associated with pedestrian fatalities and serious injuries: case finland. *European transport research review*, Springer, v. 12, p. 1–17, 2020. [2.3.1](#)

MALYSHKINA, N. V.; MANNERING, F. L. Markov switching multinomial logit model: An application to accident-injury severities. *Accident Analysis & Prevention*, Elsevier, v. 41, n. 4, p. 829–838, 2009. [2.3.1](#)

MANDELBROT, B. B. States of randomness from mild to wild, and concentration from the short to the long run. In: *Fractals and Scaling in Finance*. [S.l.]: Springer, 1997. p. 117–145. [3.3.3](#)

MARÍN, L.; QUEIROZ, M. S. A atualidade dos acidentes de trânsito na era da velocidade: uma visão geral. *Cadernos de Saúde Pública*, SciELO Public Health, v. 16, p. 7–21, 2000. [1](#), [2.2](#)

MARÍN-LEÓN, L. et al. Tendência dos acidentes de trânsito em campinas, são paulo, brasil: importância crescente dos motociclistas. *Cadernos de Saúde Pública*, SciELO Public Health, v. 28, p. 39–51, 2012. [1](#)

MARTINS, M. A. *Modelo de Decisão para redução de risco de acidentes de trânsito nas rodovias federais de Pernambuco: desenvolvimento de um Sistema de Apoio à Decisão. Dissertação (Mestrado) apresentada ao Programa de Pós-graduação em Engenharia de Produção / Orientador: Dr. Thalles Vitelli Garcez. Caruaru: UFPE, 2020. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2020.* [2.3.1](#)

MAURO, M. L. F. *ACIDENTES DE TRÂNSITO: PERFIL EPIDEMIOLÓGICO DEVÍTIMAS E CARACTERIZAÇÃO DE ALGUNS TRAÇOS DE PERSONALIDADE DE MOTORISTAS INFRATORES EMCAMPINAS, SÃO PAULO. Dissertação (Mestrado) — Universidade Estadual de Campinas, 2001.* [2.1](#), [4.1](#)

- MILONE, G. *Estatística: geral e aplicada*. [S.l.]: Câmara Brasileira do Livro, São Paulo, 2006. [3.3.1](#)
- MINERVI, N. A. *Análise fatorial e de séries temporais aplicadas às séries dos acidentes de trânsito nas rodovias federais brasileiras. Dissertação (Mestrado) apresentada ao Programa de Pós-graduação / Orientador: Dr. Anselmo Chaves Neto. Curitiba: UFPR. Dissertação (Mestrado) — UFPR, 2013. [2.3.1](#)*
- MOLAN, G.; MOLAN, M. Theoretical model for accident prevention based on root cause analysis with graph theory. *Safety and health at work*, Elsevier, v. 12, n. 1, p. 42–50, 2021. [2.3.1](#)
- MONTEGOMERY, D. C. *Introducion to Time Series Analysis and Forecasting*. [S.l.]: Wiley-Interscience, 2008. [1](#), [2.3.1](#), [3.3.1](#)
- MONTEIRO, A. M. V. et al. Análise espacial de dados geográficos. *Brasília: Embrapa*, 2004. [3.3.2](#)
- MORETIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais. Modelos Lineares univariados*. [S.l.]: ABE - Projeto Fisher, 2008. [2.3.1](#), [3.3.1](#)
- N.SODERLUND; ZWI, A. B. *Mortalidad por accidentes de tránsito en países industrializados I y en desarrollo*. 1995. <https://iris.paho.org/handle/10665.2/15522>. [1](#)
- OLIVEIRA, J. C. M. de et al. Avaliação de impacto das concessões na redução de acidentes, feridos e mortos em rodovias federais do distrito federal e entorno. *Revista Meta: Avaliação*, v. 11, n. 31, p. 251–273, 2019. [2.3.1](#)
- OLIVEIRA, R. P. d.; ACHCAR, J. A. Compositional statistical models under a bayesian approach: An application to traffic accident data in federal highways in brazil. *Pesquisa Operacional*, SciELO Brasil, v. 40, 2020. [2.3.1](#)
- ONSV. *Trilha histórica: criação do Movimento Maio Amarelo e campanhas publicitárias em prol de um trânsito mais seguro*. 2021. <https://www.onsv.org.br/>. Acessado em 18/08/2022. [1](#), [2.2](#)
- OPAS. *Segurança no trânsito nas Américas*. [S.l.], 2016. [2.2](#)
- OPAS. *Estado de la seguridad vial en lá Region de las Américas*. [S.l.], 2019. [1](#), [2.1](#)
- ORGANIZATION, W. H. *The top 10 causes of death*. 2020. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Acessado em 18/08/2022. [1](#)
- PENG, C.-K. et al. Mosaic organization of dna nucleotides. *Physical review e*, APS, v. 49, n. 2, p. 1685, 1994. [3.3.3](#), [3.3.4](#)
- PENG, C.-K. et al. Long-range anticorrelations and non-gaussian behavior of the heartbeat. *Physical review letters*, APS, v. 70, n. 9, p. 1343, 1993. [3.3.4](#)
- PEREYRA, K. B. E. D.-R. A. T. S. de la P. M. A. *The Costs os Road Injuries in Latin American*. <https://publications.iadb.org/publications/english/document/The-costs-of-road-injuries-in-Latin-America-2013.pdf>, 2013. [1](#)
- PINHO Ângela. *Violência no trânsito cresce nas pequenas cidades*. 2008. <https://www1.folha.uol.com.br/fsp/cotidian/ff0412200805.htm>. Acessado em 28/08/2022. [4.1](#)

- PNUD. *Programa das Nações Unidas para o Desenvolvimento: Declaração do Milênio*. 2016. <https://www.br.undp.org/content/brazil/pt/home/library/ods/declaracao-do-milenio.html>. Acessado em 18/08/2022. [2.2](#)
- PODOBNIK, B. et al. Statistical tests for power-law cross-correlated processes. *Physical Review E*, APS, v. 84, n. 6, p. 066118, 2011. [3.3.6](#)
- PODOBNIK, B.; STANLEY, H. E. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. *Physical review letters*, APS, v. 100, n. 8, p. 084102, 2008. [3.3.5](#), [3.3.5](#)
- POST, T. Irish. *Irish scientist Mary Ward - the first person in the world to be killed by car in 1869*. 2016. <https://www.irishpost.com/news/mary-ward-irish-scientist-became-worlds-first-car-death-day-1869-99542>. Acessado em 29/01/2021. [2.1](#)
- QUDDUS, M. A. Time series count data models: an empirical application to traffic accidents. *Accident Analysis & Prevention*, Elsevier, v. 40, n. 5, p. 1732–1741, 2008. [2.3.1](#)
- RAMSTEDT, M. Alcohol and fatal accidents in the united states a time series analysis for 1950–2002. *Accident Analysis & Prevention*, Elsevier, v. 40, n. 4, p. 1273–1281, 2008. [2.3.1](#)
- REBOREDO, J. C.; RIVERA-CASTRO, M. A.; ZEBENDE, G. F. Oil and us dollar exchange rate dependence: A detrended cross-correlation approach. *Energy Economics*, Elsevier, v. 42, p. 132–139, 2014. [3.3.6](#)
- REMY, S. L. Exploring the relationship between degrees of self similarity and altered driving states. In: IEEE. *The 2011 International Joint Conference on Neural Networks*. [S.l.], 2011. p. 3170–3177. [2.3.1](#), [4.2.1](#)
- RIBEIRO, L. de A. et al. Análise das causas dos acidentes automobilísticos nas rodovias federais da bahia entre 2014 e 2017. *Revista de Medicina*, v. 99, n. 1, p. 27–34, 2020. [1](#), [2.3.1](#)
- RIOS, P. A. A. *Mortalidade por acidentes de trânsito: evolução em anos recentes e diferenças regionais na Bahia. Dissertação (Mestrado) apresentada ao Programa de Pós-graduação em Saúde Coletiva / Orientador: Dr. Eduardo Luiz Andrade Mota*. Dissertação (Mestrado) — Universidade Federal da Bahia, 2010. [2.3.1](#)
- RIOS, P. A. A. et al. Fatores associados a acidentes de trânsito entre condutores de veículos: achados de um estudo de base populacional. *Ciência e Saúde Coletiva*, SciELO Brasil, v. 25, p. 943–955, 2020. [2.3.1](#)
- ROCHA, L. S.; FERNANDES, V. d. O.; LUSTOSA, P. B. Análise espacial através do estimador de intensidade kernel para as áreas sujeitas a riscos hidrológicos no município de salvador–ba. In: *Curitiba. Anais do XXV Congresso Brasileiro de Cartografia*. [S.l.: s.n.], 2011. [3.3.2](#)
- ROCHA, M.; NASSI, C. Modelagem estatística dos acidentes de trânsito na cidade do rio de janeiro com emprego de sistema de informações geográficas. *Rede Ibero-Americana de Estudo em Pólos Geradores de Viagens*, 2012. [2.3.1](#)

RODRIGUEZ, H. *Primeiro acidente de carro do Brasil foi a 4 km/h e envolveu Olavo Bilac*. 2020. <https://quatorrodas.abril.com.br/noticias/primeiro-acidente-de-carro-do-brasil-foi-a-4-km-h-e-envolveu-olavo-bilac/>. Acessado em 18/08/2022. 2.1

ROZESTRATEN, R. J. A. *Ambiente, Trânsito e Psicologia. Comportamento Humano no Trânsito*. (org.) Hoffman, M. H.; Cruz, R. M.; Alchieri,. [S.l.]: Casa do Psicólogo, 2003. 2.1

SAMPAIO, J. et al. Uma análise estatística dos indicadores de criminalidade em salvador. 2008. 3.3.6

SEBEGO, M. et al. The impact of alcohol and road traffic policies on crash rates in botswana, 2004–2011: a time-series analysis. *Accident Analysis & Prevention*, Elsevier, v. 70, p. 33–39, 2014. 2.3.1

SILVA, L. et al. $\rho_{x,y}$ between open-close stock markets. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 534, p. 122152, 2019. 3.3.6

SILVA, M. da; CASTRO, A. de; FILHO, A. da S. Quantificando a influência do mercado de câmbio nos preços do milho e da soja no município de barreiras. *Conjuntura e Planejamento*, v. 182, p. 45–51, 2014. 3.3.6

SILVA, M. F. da. *Uma proposta para difusão do conhecimento em correlações cruzadas de séries temporais econômicas*. Tese (Doutorado) — Universidade Federal da Bahia, 2016. 3.3.6

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*, Elsevier, v. 453, p. 1–8, 2016. 3.3.6

SILVA, P. *Indicadores Sociais e Econômicos em uma cooperativa de táxis (Salvador-BA) por meio de Análise das Séries Temporais*. Dissertação (Mestrado) apresentada ao Programa de Pós-Graduação em Modelagem Computacional e Tecnologia Industrial / Orientador: Prof. Dr. Gilney Zebende. Salvador: SENAI CIMATEC. Dissertação (Mestrado) — SENAI CIMATEC, 2009. 3.3.6

SMT. *Anel de Contorno passa a ser fiscalizado pela SMT (Superintendência Municipal Trânsito)*. 2020.

<https://www.feiradesantana.ba.gov.br/servicos.asp?titulo=Anel%20de%20Contorno%20passa%20a%20ser%20fiscalizado%20pela%20SMT>
Acessado em 18/08/2022. 4.1

SOUZA, S. R.; TABAK, B. M.; CAJUEIRO, D. O. Investigação da memória de longo prazo na taxa de câmbio no brasil. *Revista Brasileira de Economia*, SciELO Brasil, v. 60, n. 2, p. 193–209, 2006. 3.3.1, 3.3.3

SOUZA, V. d. R. de; CAVENAGHI, S.; ALVES, J. E. D. Mapeamento dos óbitos por local de ocorrência dos acidentes de trânsito na cidade do rio de janeiro. *Anais*, p. 1–22, 2016. 3.3.2

SUKHAI, A. et al. Temporal variations in road traffic fatalities in south africa. *Accident Analysis & Prevention*, Elsevier, v. 43, n. 1, p. 421–428, 2011. 2.3.1

TAQQU, M. S.; TEVEROVSKY, V.; WILLINGER, W. Estimators for long-range dependence: an empirical study. *Fractals*, World Scientific, v. 3, n. 04, p. 785–798, 1995. 3.3.3

- VASSOLER, R.; ZEBENDE, G. Dcca cross-correlation coefficient apply in time series of air temperature and air relative humidity. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 391, n. 7, p. 2438–2443, 2012. [3.3.6](#), [3.3.6](#)
- VISWANATHAN, G. M. et al. Deviations from uniform power law scaling in nonstationary time series. *Physical Review E*, APS, v. 55, n. 1, p. 845, 1997. [3.3.4](#)
- VLAHOGIANNI, E.; KARLAFTIS, M. Temporal aggregation in traffic data: implications for statistical characteristics and model choice. *Transportation Letters*, Taylor & Francis, v. 3, n. 1, p. 37–49, 2011. [2.3.1](#)
- WASELFISZ, J. J. *Mapa da violência 2013*. [S.l.]: UNESCO Brasil, 2013. [1](#)
- WANG, G.-J.; XIE, C. Cross-correlations between renminbi and four major currencies in the renminbi currency basket. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 392, n. 6, p. 1418–1428, 2013. [3.3.6](#)
- WEST, A. C. I. B. Acidente de trânsito: Um recorte do oeste baiano. *Humanidades*, v. 5, n. 2, 2016. [2.3.1](#)
- WHO. *Global Status Report on Road Safety*. [S.l.], 2018. [1](#), [2.2](#)
- WHO. *Plano Global Década de Ação pela segurança no trânsito, 2021-2030*. [S.l.], 2021. [2.2](#)
- XIONG, Y.; TOBIAS, J. L.; MANNERING, F. L. The analysis of vehicle crash injury-severity data: A markov switching approach with road-segment heterogeneity. *Transportation research part B: methodological*, Elsevier, v. 67, p. 109–128, 2014. [2.3.1](#)
- ZEBENDE, G.; FILHO, A. M. Cross-correlation between time series of vehicles and passengers. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 388, n. 23, p. 4863–4866, 2009. [3.3.5](#), [3.3.5](#)
- ZEBENDE, G.; SILVA, M. da; FILHO, A. M. DCCA cross-correlation coefficient differentiation: Theoretical and practical approaches. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 392, n. 8, p. 1756–1761, 2013. [3.3.5](#)
- ZEBENDE, G. F. DCCA cross-correlation coefficient: Quantifying level of cross-correlation. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 390, n. 4, p. 614–618, 2011. [1](#), [3.3.6](#), [3.2](#)
- ZEBENDE, G. F. et al. Uma visão hora a hora da autocorrelação em dados de temperatura e umidade relativa do ar na bahia. *Revista Brasileira de Climatologia*, v. 29, 2021. [3.3.4](#), [4.3](#)
- ZEBENDE, G. F.; FILHO, F. M. O.; CRUZ, J. A. L. Auto-correlation in the motor/imaginary human eeg signals: A vision about the fdFA fluctuations. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 9, p. e0183121, 2017. [3.3.4](#)
- ZHAO, Y. *R and data mining: Examples and case studies*. [S.l.]: Academic Press, 2012. [3.3.5](#)

Análise das correlações em séries temporais de acidentes de trânsito e de outras séries

ROBERTA PIMENTA CUNHA DOS SANTOS

Feira de Santana, Agosto de 2022.