UNIVERSIDADE DE SÃO PAULO INSTITUTO DE ENERGIA E AMBIENTE PROGRAMA DE PÓS-GRADUAÇÃO EM ENERGIA

JOSÉ VITOR PEREIRA MIGUEL

AVALIAÇÃO DA GERAÇÃO DE ENERGIA ELÉTRICA NO BRASIL EM CONDIÇÕES DE ESCASSEZ DE RECURSOS EÓLICOS E HÍDRICOS

SÃO PAULO 2021

JOSÉ VITOR PEREIRA MIGUEL

AVALIAÇÃO DA GERAÇÃO DE ENERGIA ELÉTRICA NO BRASIL EM CONDIÇÕES DE ESCASSEZ DE RECURSOS EÓLICOS E HÍDRICOS

Tese apresentada ao Programa de Pós-Graduação em Energia do Instituto de Energia e Ambiente da Universidade de São Paulo para a obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Ildo Luís Sauer

Versão Corrigida

SÃO PAULO 2021

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

FICHA CATALOGRÁFICA

Miguel, José Vitor Pereira. Avaliação da geração de energia elétrica no Brasil em condições de escassez de recursos eólicos e hídricos. / José Vitor Pereira Miguel; orientador: Ildo Luís Sauer. -- São Paulo, 2021. 224 f.: il.; 30 cm.

Tese (Doutorado em Ciências) – Programa de Pós-Graduação em Energia – Instituto de Energia e Ambiente da Universidade de São Paulo.

1. Recursos hídricos. 2. Geração de energia elétrica - Brasil. 3. Planejamento energético. I. Título.

Elaborado por Maria Penha da Silva Oliveira CRB-8/6961

Nome: MIGUEL, José Vitor Pereira

Título: Avaliação da geração de energia elétrica no brasil em condições de escassez de recursos eólicos e hídricos

Banca Examinadora

Tese apresentada ao Programa de Pós-Graduação em Energia do Instituto de Energia e Ambiente da Universidade de São Paulo para a obtenção do título de Doutor em Ciências.

Aprovado em:

Prof. Dr.	Instituição:
Julgamento:	Assinatura:
Prof. Dr	Instituição:
Julgamento:	Assinatura:
Prof. Dr	Instituição:
Julgamento:	Assinatura:
Prof. Dr	Instituição:
Julgamento:	Assinatura:
Prof. Dr	Instituição:
Julgamento:	Assinatura:
Prof. Dr	Instituição:
Presidente	Assinatura:

À sociedade brasileira.

AGRADECIMENTOS

Aos meus pais serei eternamente grato pelos valores e ensinamentos transmitidos; pela educação e formação pessoal e profissional que puderam me proporcionar; pelo suporte e voto de confiança que me são concedidos em cada uma de minhas escolhas; pelo apoio incondicional.

Agradeço ao professor Ildo, que desde a primeira aula de "Usos Finais" eu pude notar sua dedicação e compromisso com a ciência e com a transformação da sociedade; reconheci genuína disposição em compartilhar seu conhecimento – não apenas técnico ou teórico, mas também sobre coisas banais, pavões, culinária etc. Obrigado pelos desafios impostos, pela confiança e pela liberdade; foram elementos fundamentais na execução de minha pesquisa.

Aos demais professores e mestres da USP, sou grato por todo o aprendizado. Sinto orgulho de ter feito parte deste que considero o ambiente mais rico que já botei meus pés no Brasil. Foi um privilégio, sem dúvidas. Sinto-me endividado com a sociedade e para com o contribuinte. Pretendo fazer o possível para retribuir.

Aos colegas do grupo de pesquisa do CPLEN; somos de origens e áreas distintas, mas falamos uma mesma língua, temos objetivos comuns. Agradeço por terem contribuído com minha pesquisa e por terem proporcionado um cotidiano sempre agradável no ambiente do IEE.

À minha irmã Maria Laura, agradeço o porto seguro concedido na capital durante maior parte desta empreitada. Vejo em você um lar, seja onde for. Conte comigo sempre!

Aos amigos de faculdade, de república, irmãos e primos, meus conterrâneos: sou grato por terem dividido comigo as minhas memórias mais marcantes. Parte de mim foi construída por tudo o que, juntos, pudemos vivenciar.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Agradeço à CAPES pela importância e papel que ainda busca desempenhar pela ciência no Brasil.

"Nature, to be commanded, must be obeyed." (Francis Bacon)

RESUMO

MIGUEL, José Vitor Pereira. **Avaliação da geração de energia elétrica no Brasil em condições de escassez de recursos eólicos e hídricos**. 2021. 224 f. Tese (Doutorado em Ciências) – Programa de Pós-Graduação em Energia da Universidade de São Paulo, São Paulo, 2021.

A intensificação em curso da exploração de recursos eólicos para geração de energia no Brasil é acompanhada por um rearranjo das participações de diferentes fontes energéticas na matriz elétrica, sendo plausível que o atendimento à carga futura seja majoritariamente sustentado por recursos eólicos e hídricos. Na medida em que estes recursos manifestam variabilidade interanual em suas disponibilidades, aumenta-se a complexidade do planejamento do setor elétrico e do dimensionamento da capacidade instalada de geração no longo prazo, haja vista que esta variabilidade pode acarretar ocorrências de eventos extremos de escassez eólica, hídrica, ou até mesmo períodos de escassez conjunta destes recursos. Entende-se que eventos desta sorte debilitariam, em última instância, a geração anual de energia. Nesse sentido, este trabalho pretendeu compreender características dos eventos de escassez eólica e verificar se fenômenos climáticos de larga escala e baixa frequência possam ter influenciado suas ocorrências. Ademais, o objetivo principal da pesquisa foi avaliar o comportamento da geração combinada de energia pelas fontes eólica e hídrica em condições de escassez conjunta destes recursos. Para tanto, um extenso ferramental analítico foi aplicado sobre séries temporais constituídas de dados eólicos que representam o NE do Brasil e de dados hídricos que representam a região SE/CO. Mediante técnicas de clusterização, foi possível identificar subregiões cujo vento se comporta similarmente e, por meio de análises empíricas, permitiu-se o reconhecimento histórico de ocorrência de eventos extremos que remontam escassez destes recursos. Os resultados apontam para a ocorrência de eventos de escassez conjunta nos anos de 1964 e 1971. Isoladamente, eventos de escassez eólica em 1973-1974 e 1984-1985 parecem ter sido influenciados pelas fases negativas concomitantes dos modos de variabilidade climática El Niño Oscilação Sul e padrão dipolo do Oceano Atlântico em sua porção tropical. Na última etapa, cenários com base em diferentes composições da matriz elétrica foram conceituados para simular a geração por recursos eólicos no NE combinada com a geração por recursos hídricos no SE/CO e representar o atendimento à carga de 2019 bem como a carga projetada para 2050. Foi revelado que, em condições de escassez, a geração combinada por estes recursos pode ser reduzida em cerca de 80 TWh se comparada ao que fora verificado no ano de 2019. Em um cenário que considera 190 GW de capacidade instalada de geração pela fonte eólica, o atendimento à carga futura pode sofrer redução de cerca de 200 TWh por ano em eventos de escassez conjunta de recursos eólicos e hídricos. Estes números reforçam a tese de que condições como estas devem conduzir tomadas de decisão no que tange o dimensionamento e a composição do parque gerador a nível nacional, a fim de que o setor elétrico brasileiro seja robusto a ponto de tolerar a variabilidade interanual intrínseca aos recursos eólicos e hídricos.

Palavras-chave: Recursos eólicos. Recursos hídricos. Variabilidade interanual. Modos de variabilidade climática. Escassez. Geração de energia.

ABSTRACT

MIGUEL, José Vitor Pereira. Assessment of Brazil's electric energy generation under wind and hydro resources' scarcity. 2021. 224 f. Thesis (Doctorate in Sciences) – Graduate Program in Energy, University of São Paulo, 2021.

The ongoing intensification of wind resources' exploitation towards energy generation in Brazil is followed by a rearrangement on the shares of different energy sources in the electricity mix, seeming likely that future load to be mainly sustained by wind and hydro resources. As far as these resources exhibit interannual variability on their availability, planning the power sector and dimensioning the long-term future installed power capacity can become complex tasks, provided that this variability might entail the occurrence of extreme events such as wind scarcity, hydro droughts or, even worse, both happening concurrently. It is understood that, ultimately, such events would undermine the annual power supply. In that sense, this research intended to comprehend some characteristics of wind scarcity events and to examine if large scale and low frequency climate phenomena could have influenced incidents like these before. Moreover, this research's main objective was to evaluate the combined energy generation behavior from both wind and hydro resources under extreme scarcity events. For this purpose, an extensive analytical toolkit was implemented over time series of wind data representing the Northeast of Brazil and hydro data representative to the Southeast/Midwest region. Upon clustering techniques, it was possible to identify subregions where the wind behaves similarly and, through empirical analysis, the historical occurrences of both wind and hydro resources' scarcity was recognized. Results point to extreme events like these back in 1964 and 1971. In isolation, wind droughts in 1973-1974 and 1984-1985 seemed to have been influenced by the negative phases of El Niño South Oscillation and the dipole pattern on the tropical Atlantic Ocean happening simultaneously. Subsequently, scenarios based on different electricity mix compositions were conceptualized to simulate energy generation from wind resources in the Northeast combined with hydro resources in the Southeast/Midwest subsystem to represent the load following for the year of 2019 as well as for the 2050 projected load. It was revealed that, under diminished availability due to scarcity, the conjoint generation from these resources could have been reduced in about 80 TWh when compared to the verified generation for 2019. As for a future scenario considering 190 GW of installed wind power capacity, the load meeting could suffer the reduction of 200 TWh in a single year due to scarcity. These findings reinforce the thesis that conditions like these must conduct decision making when it comes to dimensioning and figuring the composition of the generation mix to a national level, so that Brazil's power sector can be robust up to the point of tolerating the interannual variability intrinsic to wind and hydro resources.

Keywords: Wind resources. Hydro resources. Interannual variability. Climate modes of variability. Scarcity. Energy generation.

LISTA DE FIGURAS

Figura 1 – A evolução da capacidade instalada de energia eólica no Brasil27
Figura 2 – Matriz da capacidade instalada de energia elétrica no Brasil ao final de 202028
Figura 3 – Previsão da participação das fontes na matriz de energia elétrica de 2019 a 2029.29
Figura 4 – Fatores e escalas temporais associadas à variabilidade de recursos naturais30
Figura 5 – Gestão da capacidade e da geração de energia em um sistema elétrico31
Figura 6 – Histórico da geração de energia pelas fontes hídrica e eólica em todo o SIN34
Figura 7 – Empreendimentos de geração elétrica em operação no Brasil
Figura 8 – Subsistemas do SIN40
Figura 9 – Histórico da geração de energia pela fonte hídrica em cada subsistema do SIN40
Figura 10 – Histórico da geração de energia pela fonte eólica em cada subsistema do SIN41
Figura 11 – Modelo conceitual da circulação global atmosférica42
Figura 12 – Posicionamento da ZCIT em (a) anos secos e (b) anos chuvosos no NE43
Figura 13 – Posicionamento variável da ZCAS45
Figura 14 – Aproveitamento hidrelétrico na região hidrográfica do Paraná
Figura 15 – Quantidade média anual de passagens de frentes frias de 1979 a 200547
Figura $16 - a$) anomalias da velocidade de vento em superfície para o primeiro trimestre de
2015, expressas em desvios padrão de distância da média de 1979-2014 e b) série temporal de
desvios padrão para o período de janeiro-março na região delimitada pelo retângulo52
Figura 17 – Anomalias de precipitação relativas ao período de 1982-2015: anos hidrológicos
de 1995-201455
Figura 18 – As crises hídricas de 2012 a 2017 em suas respectivas regiões
Figura 19 – Sistemas atmosféricos atuantes na baixa troposfera da América do Sul59
Figura 20 – ENOS em suas três fases: a) neutra, b) quente (El Niño) e c) fria (La Niña) 62
Figura 21 – Regiões NINO63
Figura 22 - a) anomalias de TSM mensais retificadas para o período 1900-2008, b) série
temporal para o índice Niño-3.4, c) análise espectral da série temporal e d) desvio padrão
mensal do índice Niño-3.464
Figura 23 – a) anomalias de TSM mensais para o período 1900-2008, b) série temporal para o
índice de PDO, com curva suavizada por médias móveis quinquenais67
Figura 24 – a) anomalias de TSM mensais para o período 1870-2008, b) série temporal para o
índice de AMO, com curva estimativa para a componente natural do índice filtrado69

Figura 25 - Anomalias de TSM, precipitação e vento em superfície para a) AMM durante o
outono austral e b) AEM durante o inverno austral
Figura 26 – Médias mensais de AMM (campo de TSM) de 1948-2018 71
Figura 27 – Representação esquemática da estrutura em grade cartesiana utilizada em Modelos
Gerais de Circulação
Figura 28 – Exemplos de séries temporais contínua (a) e discreta (b)
Figura 29 – Elementos de um <i>boxplot</i>
Figura 30 – Exemplos de um histograma de frequência com curva suavizada representativa de
sua distribuição (à esquerda) e de uma função de distribuição acumulada (à direita)
Figura 31 – Distribuições probabilísticas mais comuns
Figura 32 - PDF hipotética para uma variável aleatória e o cálculo, por integração, da
probabilidade de ocorrência de um intervalo de valores específicos de <i>x</i>
Figura 33 – Correlogramas comuns: processo puramente aleatório (a); processo alternado (b);
processo com tendência (c) e processo com sazonalidade (d)
Figura 34 – Exemplo de uma mesma série temporal tratada como estacionária no curto prazo
(a), não estacionária no curto prazo (b) e estacionária no longo prazo (c)
Figura 35 – Exemplo de decomposição de uma série temporal em outras três: tendência, ciclo
sazonal e resíduos
Figura 36 – Exemplo de regressão linear
Figura 37 – Exemplo da aplicação da filtragem por médias móveis
Figura 38 – Médias de temperatura de cada mês (a) e a identificação de uma função sinusoidal
com amplitude regulada e defasagem para representar o ciclo anual (linha grossa em b) 86
Figura 39 – Transformação de uma série temporal para o domínio da frequência pela análise de
Fourier
Figura 40 – Exemplo de <i>aliasing</i> , onde um fenômeno (linha pontilhada) que apresenta variações
em frequências maiores do que a frequência com que os dados (pontos vazios) foram registrados
(linha sólida)
Figura 41 – Exemplo de um periodograma com todas as frequências de 0 a 0,5 (à esquerda) e
ampliado na região das frequências de 0 a 0,03 (à direita)
Figura 42 - Periodograma suavizado que apresenta vazamentos (leakage indicado pela linha
tracejada) e após aplicação de <i>tapering</i> (linha sólida)
Figura 43 – Comportamento errático de duas diferentes amostras de mesmo comprimento n
para estimativa do espectro de um mesmo fenômeno

Figura 44 - Rotação de eixos tridimensionais para a direção de maior variabilidade de um
conjunto de dados96
Figura 45 – Exemplo de um <i>scree plot</i>
Figura 46 – Exemplo de agrupamento ou clusterização de séries temporais
Figura 47 – Tipos de clusterização de séries temporais integrais100
Figura 48 – A distância entre séries temporais por distância Euclidiana (a) e por DTW (b).102
Figura 49 - Nove séries temporais em 3 padrões diferentes P1, P2 e P3 (a); dendrogramas
resultantes da clusterização baseada na forma (distância Euclidiana) (b) e com dissimilaridade
baseada na estrutura (índice adaptativo com $q = 2$) (c)105
Figura 50 – Abordagens aglomerante e divisora da clusterização hierárquica106
Figura 51 - Representação de algumas diferentes definições de dissimilaridade entre dois
clusters (de cima para baixo: average linkage; single linkage; complete linkage)108
Figura 52 – Região delimitada para avaliação de recursos eólicos116
Figura 53 – Distribuição espacial das séries temporais de base de dados NOAA (a), ERA (b) e
MERRA (c)
Figura 54 - Reservatórios de UHEs e a área com bordas em vermelho que representa 70% da
capacidade total de armazenamento do Brasil119
Figura 55 – UHEs mais representativas de cada bacia hidrográfica brasileira
Figura 56 – Esquema dos procedimentos metodológicos implementados
Figura 57 – Dendrograma das séries NOAA após clusterização hierárquica divisora com matriz
dissimilaridade baseada na distância Euclidiana128
Figura 58 - Dendrograma das séries ERA após clusterização hierárquica aglomerante com
método de aglomeração average linkage e com matriz dissimilaridade baseada na correlação
de Pearson129
Figura 59 – Dendrograma das séries MERRA após clusterização hierárquica aglomerante com
método de aglomeração average linkage e com matriz dissimilaridade baseada no índice
adaptativo com $q = 1,7$
Figura 60 – Resultados da clusterização das séries temporais de: NOAA por repartição k-Means
(a); ERA hierárquica aglomerante (b); MERRA hierárquica aglomerante (c)131
Figura 61 – Localização geográfica das séries temporais representativas
Figura 62 – Séries temporais NOAA, seus histogramas e boxplots
Figura 63 – Séries temporais ERA, seus histogramas e <i>boxplots</i> 137
Figura 64 – Séries temporais MERRA, seus histogramas e <i>boxplots</i>
Figura 65 – Diferentes abordagens para definição de eventos de vazão reduzida

Figura 66 – Função de distribuição empírica acumulada	. 141
Figura 67 – Histograma e função densidade suavizada	. 143
Figura 69 – Análise fragmentada mês a mês da série N55	. 145
Figura 69 - Série temporal N55 contrastada com um limiar harmônico móvel com ba	se na
média e desvio padrão de cada mês e a identificação de períodos de escassez	. 146
Figura 70 – Séries temporais normalizadas e a ocorrência de eventos de escassez	. 148
Figura 71 – Séries plotadas em anos com forte sinal de escassez de recursos eólicos	. 157
Figura 72 – Exemplo de remoção de tendência e ciclo anual na série M139	. 161
Figura 73 – Mapeamento das duas primeiras PCs de cada base de dados	. 162
Figura 74 – PCs com aplicação de <i>taper</i>	. 164
Figura 76 – Correlogramas das PCs	. 165
Figura 76 – Tabela qui-quadrado e as regiões interpoladas	. 166
Figura 77 – Periodogramas das PCs de cada base de dados	. 166
Figura 78 – Índices e estações do ano com sinal da correlação acima de 0,4 para as PC1	s dos
clusters	. 175
Figura 79 – Séries temporais dos índices NINO 3.4 e MGI de 1950 a 2014	. 177
Figura 80 – Consumo final por eletricidade: projeções de 2013 a 2050	. 180
Figura 81 – Carga de energia total: projeções de 2013 a 2050	. 180
Figura 82 – Curva de carga diária em sistemas predominantemente hídricos: sem energia e	ólica
(a) e com participação da energia eólica (b)	. 181
Figura 83 – Curva de carga anual com dados mensais para o período de 2014 a 2019	. 182
Figura 84 – Atendimento à carga do sistema elétrico do Brasil em 2019	. 182
Figura 85 – Atendimento à carga em 2019 pelas fontes eólica (NE) e hídrica (SE/CO)	. 184
Figura 86 – Curva de potência Vestas V112 3.0 MW	. 188
Figura 87 – Curva de potência Enercon E126 7.5 MW	. 189
Figura 88 – Perfis teóricos da velocidade do vento para diferentes índices de rugosidade	. 190
Figura 89 – ENA e geração pela fonte hídrica no subsistema SE/CO de 2015 a 2019	. 193
Figura 90 – Correção dos valores de ENA	. 193
Figura 91 – Resultados do cenário II: escassez conjunta de 1964	. 198
Figura 92 – Resultados do cenário III: escassez conjunta de 1971	. 198
Figura 93 – Resultados do cenário IV	. 201
Figura 94 – Resultados do cenário V	. 202
Figura 95 – Resultados do cenário VI	. 202
Figura 96 – Resultados do cenário VI: geração a partir dos ciclos médios anuais	. 203

LISTA DE TABELAS

Tabela 1 – Crises do setor elétrico brasileiro e suas principais causas
Tabela 2 – Algumas bases de dados de reanálise disponíveis74
Tabela 3 – Contribuição do comportamento e da distância entre séries conforme valor de $q104$
Tabela 4 – Características das bases de dados de reanálise116
Tabela 5 – Características dos modos de variabilidade e seus índices
Tabela 6 – Características dos dados representativos de recursos hídricos121
Tabela 7 – Características dos dados de carga e geração de energia elétrica121
Tabela 8 - Coeficiente de correlação entre a matriz dissimilaridade e a matriz de distância
cofenética resultante do processo de clusterização hierárquica127
Tabela 9 – Coeficiente de Dunn para os resultados de cada processo de clusterização131
Tabela 10 – Seleção de séries temporais representativas de cada <i>cluster</i>
Tabela 11 – Dados estatísticos gerais das séries selecionadas
Tabela 12 - Média anual de velocidade do vento [m/s] associada a diferentes intervalos de
tempo de retorno e probabilidades de não excedência e a porcentagem relativa à média histórica
da série original142
Tabela 13 - Ocorrências mês a mês de média de velocidade do vento abaixo dos limiares de
referência149
Tabela 14 – Ocorrências por ano de meses com média de velocidade do vento abaixo do limiar
de 1,5
Tabela 15 - Classificação das médias anuais de velocidade do vento [m/s] e relações com as
médias totais
Tabela 16 – Classificação das médias anuais de vazão natural [m³/s] e relações com as médias
totais
Tabela 17 – Porcentagem da variância total explicada pelas PCs de 1 até 7161
Tabela 18 – Porcentagem da variância total explicada pelas PCs de 1 até 7 por <i>cluster</i> 170
Tabela 19 - Grau de associação entre os MV e as PC1s e PC2s de cada cluster para a estação
verão171
Tabela 20 - Grau de associação entre os MV e as PC1s e PC2s de cada cluster para a estação
outono
Tabela 21 – Grau de associação entre os MV e as PC1s e PC2s de cada cluster para a estação
inverno173

Tabela 22 - Grau de associação entre os MV e as PC1s e PC2s de cada cluster para a estação
primavera 174
Tabela 23 – Teste de significância estatística da tabela t para valores da correlação de Pearson
<i>r</i>
Tabela 24 – Sinal dos índices NINO 3.4 e MGI em anos de escassez eólica 178
Tabela 25 – Características dos cenários conceituados 186
Tabela 26 - Potência e coeficiente de potência em função da velocidade do vento para as
turbinas eólicas
Tabela 27 – Valores típicos de rugosidade para diferentes condições 191
Tabela 28 – Características das UHEs consideradas 192
Tabela 29 – Velocidade mensal média do vento das séries representativas [m/s] 195
Tabela 30 – ENA para o subsistema SE/CO [GWh] 196
Tabela 31 – Resultados das simulações de geração nos cenários II e III 197
Tabela 32 – Distribuição da carga anual por entre os 12 meses 199
Tabela 33 – Geração eólica por uma única unidade geradora

LISTA DE ABREVIATURAS E SIGLAS

ABEEólica	Associação Brasileira de Energia Eólica
ACP	Análise de Componentes Principais
AEM	Atlantic Equatorial Mode - Modo Equatorial do Atlântico
AGNES	Agglomerative Nesting
AMM	Atlantic Meridional Mode - Modo Meridional do Atlântico
AMO	Atlantic Multidecadal Oscillation - Oscilação Multidecadal do Atlântico
ANA	Agência Nacional de Águas
ANEEL	Agência Nacional de Energia Elétrica
ASAN	Anticiclone Subtropical do Atlântico Norte
ASAS	Anticiclone Subtropical do Atlântico Sul
C20r	Twentieth Century Reanalysis Project
CA	Ciclo Anual
CDF	Cumulative Distribution Function - Função de Distribuição Acumulada
CFSR	Climate Forecast System Reanalysis
CO	Centro-Oeste
DIANA	Divisive Analysis
DJF	Dezembro - Janeiro - Fevereiro
DTW	Dynamic Time Warping
EMCWF	European Centre for Medium-Range Weather Forecasts
ENA	Energia Natural Afluente
ENOS	El Niño-Oscilação Sul
EPE	Empresa de Pesquisa Energética
ESQ	Erro da Soma dos Quadrados
GEE	Gases do Efeito Estufa
GWP	Global Water Partnership
HN	Hemisfério Norte
HS	Hemisfério Sul
IQR	Intervalo Interquartil
JJA	Junho - Julho - Agosto
MAM	Março - Abril - Maio
MERRA	Modern Era Retrospective-analysis for Research and Applications
MGI	Gradiente do Atlântico Tropical
MJO	Madden-Julian Oscillation - Oscilação de Madden-Julian
MV	Modos de Variabilidade
Ν	Norte
NAO	North Atlantic Oscillation - Oscilação do Atlântico Norte
NASA	National Aeronautics and Space Administration
NCAR	National Centre for Atmospheric Research

NCEP	National Centers for Environmental Prediction
NE	Nordeste
NOAA	National Oceanic and Atmospheric Administration
ONS	Operador Nacional do Sistema
PAM	Partitioning Around Medoids - Repartição em Torno de Medoides
PC	Componente Principal
PCH	Pequenas Centrais Hidrelétricas
PDF	Probability Density Function - Função densidade de probabilidade
PDO	Pacific Decadal Oscillation - Oscilação Decadal do Pacífico
PROINFA	Programa de Incentivo às Fontes Alternativas
S	Sul
SAM	South Annular Mode - Modo Anular Sul
SDI	Standardized Drought Index
SE	Sudeste
SIN	Sistema Interligado Nacional
SON	Setembro - Outubro - Novembro
SPI	Standardized Precipitation Index
SRSI	Standardized Reservoir Supply Index
SSFI	Standardized Streamflow Index
TSM	Temperatura da Superfície do Mar
UHE	Usina Hidrelétrica
WMO	World Meteorological Organization
ZCAS	Zona de Convergência do Atlântico Sul
ZCIT	Zona de Convergência Intertropical

SUMÁRIO

1	INTRODUÇÃO	
1.1	CONTEXTUALIZAÇÃO	27
1.2	OBJETIVOS	35
1.3	JUSTIFICATIVA	
1.4	HIPÓTESE	
1.5	ESTRUTURA CAPITULAR DA TESE	
2	REVISÃO BIBLIOGRÁFICA	
2.1	RECURSOS EÓLICOS E HÍDRICOS NO BRASIL	
2.1	.1 Recursos eólicos que alimentam o subsistema NE do SIN	41
2.1	.2 Recursos hídricos que alimentam o subsistema SE/CO do SIN	44
2.1	.3 Eventos de escassez de recursos eólicos e hídricos	48
2.2	MODOS DE VARIABILIDADE CLIMÁTICA DE BAIXA FREQUÊNCIA	58
2.2	.1 El Niño-Oscilação Sul (ENOS)	61
2.2	.2 Oscilação Decadal do Pacífico (PDO)	65
2.2	.3 Oscilação Multidecadal do Atlântico (AMO)	68
2.2	.4 Modos de variabilidade do Atlântico Tropical	69
2.3	BASES DE DADOS DE REANÁLISE	72
2.4	ANÁLISE DE SÉRIES TEMPORAIS	75
2.4	.1 Conceitos gerais de séries temporais	75
2.4	.2 Análise de Fourier	87
2.4	.3 Análise Espectral	89
2.4	.4 Análise de Componentes Principais	94
2.5	CLUSTERIZAÇÃO DE SÉRIES TEMPORAIS	99
2.5	.1 Dissimilaridade entre séries temporais	101
2.5	.2 Clusterização hierárquica	106

2.5	.3	Clusterização por repartição 1	109
2.5	.4	Clusterização híbrida 1	111
2.5	.5	Validação interna da clusterização 1	112
3	Μ	IATERIAIS E MÉTODOS DE PESQUISA 1	115
3.1	D	ADOS UTILIZADOS 1	115
3.1	.1	Dados representativos de recursos eólicos 1	115
3.1	.2	Dados dos modos de variabilidade 1	117
3.1	.3	Dados representativos de recursos hídricos1	118
3.1	.4	Dados de carga e geração de energia elétrica 1	121
3.2	PI	ROCEDIMENTOS METODOLÓGICOS 1	121
4	C	LUSTERIZAÇÃO DE SÉRIES TEMPORAIS DE DADOS EÓLICOS 1	125
4.1	C	LUSTERIZAÇÃO HIERÁRQUICA 1	126
4.2	C	LUSTERIZAÇÃO POR REPARTIÇÃO <i>K-MEANS</i> E HÍBRIDA 1	130
4.3	V	ALIDAÇÃO INTERNA DOS RESULTADOS DE CLUSTERIZAÇÃO 1	130
5	A	NÁLISE DE SÉRIES TEMPORAIS DE DADOS EÓLICOS 1	135
5.1	E	STATÍSTICA DESCRITIVA 1	135
5.2	E	VENTOS EXTREMOS: ESCASSEZ DE RECURSOS EÓLICOS 1	139
5.2	.1	Análise de frequência empírica 1	139
5.2	.2	Análise mês a mês de ocorrência de eventos extremos 1	144
5.2	.3	Análise ano a ano de ocorrência de eventos extremos 1	150
5.2	.4	Ocorrências de escassez conjunta de recursos eólicos e hídricos 1	158
5.3	V	ARIAÇÕES PERIÓDICAS DE RECURSOS EÓLICOS 1	160
5.3	.1	Análise de Componentes Principais das séries temporais de dados eólicos 1	160
5.3	.2	Análise Espectral das componentes principais 1	163
6	F	ATORES DE INFLUÊNCIA SOBRE A VARIABILIDADE DOS RECURS	OS
EÓLI	CC	DS 1	169

6.2	OS FATORES DE INFLUÊNCIA DURANTE EVENTOS DE E	SCASSEZ DE
REC	URSOS EÓLICOS	177
7	A GERAÇÃO DE ENERGIA EM EVENTOS DE ESCASSEZ DI	E RECURSOS
EÓLI	COS E HÍDRICOS	179
7.1	CARGA DE ENERGIA NO BRASIL	179
7.2	CONCEPÇÃO DE CENÁRIOS	
7.3	CONVERSÃO ENERGÉTICA	
7.3	.1 Energia proveniente de recursos eólicos	
7.3	.2 Energia proveniente de recursos hídricos	191
7.4	SIMULAÇÕES DA GERAÇÃO DE ENERGIA ELÉTRICA	194
7.4	.1 Resultados dos cenários II e III	196
7.4	.2 Resultados dos cenários IV, V e VI	
8	CONCLUSÕES E CONSIDERAÇÕES FINAIS	
8.1	LIMITAÇÕES E SUGESTÕES PARA TRABALHOS FUTUROS	
REFE	ERÊNCIAS BIBLIOGRÁFICAS	211

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

A inserção da energia eólica na matriz elétrica do Brasil foi pautada por razões políticoeconômicas. Com o lançamento do Programa de Incentivo às Fontes Alternativas (PROINFA) em 2002, fora dado o primeiro passo para o desenvolvimento da indústria eólica no Brasil, com a contratação de 1,4 GW de capacidade a ser instalada ao longo da primeira década do século XXI. A partir de 2009, a introdução de uma sistemática de leilões de contratação promoveu a competição e a redução dos preços de geração por diferentes fontes de energia. Tal fato somado aos ganhos de expertise dos atores envolvidos (como projetistas, investidores e fabricantes de equipamentos e aerogeradores), contribuíram para o estabelecimento da indústria eólica brasileira (BAYER, 2018).

Se em 2009 a capacidade de geração eólica não alcançava a marca de 1 GW, ao final do ano de 2020 a potência instalada chegou a quase 17 GW – resultado de um total de 660 parques eólicos que agregam cerca de 8 mil aerogeradores em operação (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL), 2020; ASSOCIAÇÃO BRASILEIRA DE ENERGIA EÓLICA (ABEEÓLICA), 2020). A previsão é de aumento da capacidade instalada para 27 GW até 2024, como ilustra a Figura 1.



Figura 1 – A evolução da capacidade instalada de energia eólica no Brasil

Fonte: adaptado de Associação Brasileira de Energia Eólica (2020)

O aproveitamento dos recursos hídricos para geração de energia elétrica percorreu trajetória diferente daquela dos recursos eólicos. Com exceção das usinas de Belo-Monte, Santo Antônio e Jirau, a maioria dos grandes empreendimentos de geração hidrelétrica (usinas hidrelétricas (UHEs) com potência instalada acima de 1 GW) foi construída até o final da década de 90, principalmente nas bacias hidrográficas localizadas nas regiões do Sul (S) e Sudeste (SE) do Brasil (ANEEL, 2020). Por outro lado, as chamadas Pequenas Centrais Hidrelétricas (PCHs – usinas com capacidade instalada máxima de 30 MW) cresceram em número de unidades em operação, sobretudo a partir dos anos 2000, e concentram-se também nas regiões S e SE (FERREIRA et al., 2016).

Esta reconstrução histórica faz-se necessária para ilustrar a situação atual, em que a geração de energia elétrica no Brasil encontra-se majoritariamente sustentada por recursos naturais como os hídricos e eólicos, além de outros de origem fóssil e à base de biomassa (Figura 2). Ademais, uma maior diversificação da matriz da capacidade instalada de energia elétrica está sendo desenhada para os próximos anos, onde a participação proporcional da geração pela fonte hídrica se reduzirá ainda mais, ao passo que outras fontes renováveis de energia serão cada vez mais representativas proporcionalmente na matriz nacional, como é o caso das fontes eólica, solar e biomassa (Figura 3) (EMPRESA DE PESQUISA ENERGÉTICA (EPE), 2020a).



Figura 2 – Matriz da capacidade instalada de energia elétrica no Brasil ao final de 2020

Fonte: ANEEL (2020) e ABEEólica (2020)



Figura 3 – Previsão da participação das fontes na matriz de energia elétrica de 2019 a 2029

O crescimento da participação de fontes renováveis na geração de energia pode ser considerado vantajoso sob os prismas da sustentabilidade de tais recursos – contribuindo para a manutenção de uma matriz elétrica de baixo carbono – e do baixo custo marginal de operação (custo por unidade de energia produzida para atender a um acréscimo no sistema). Do ponto de vista da disponibilidade, entretanto, o uso de tais recursos torna-se, de certo modo, incerto.

Isto se deve à estocasticidade intrínseca aos recursos eólicos, denotando uma variabilidade natural da velocidade do vento que se manifesta em diferentes escalas temporais. No curto prazo, rápidas flutuações na velocidade do vento ocorrem aleatoriamente em questão de segundos, minutos, horas e/ou dias, que pode perturbar o equilíbrio instantâneo entre oferta e demanda por eletricidade; é a chamada intermitência (JABIR et al., 2017). Nesta escala de tempo, alguns fenômenos atmosféricos como a passagem de frentes frias e quentes, ciclones e anticiclones, ondas atmosféricas, tempestades, entre outros, podem agir sobre os recursos eólicos, intensificando ou enfraquecendo-os (DIAS e SILVA, 2009).

Por esta ótica, a incorporação de sistemas de geração com base em recursos renováveis estocásticos e intermitentes à rede elétrica fornece desafios para gerir uma oferta de energia que seja estável e segura. Todavia, complicações deste teor estão sendo sanadas mediante a adoção de tecnologias de armazenamento de energia, assim preservando a regularidade do atendimento à carga e evitando problemas técnicos em termos de proteção, qualidade e controle de despacho

da energia (BARANES, JACQMIN e POUDOU, 2017; NOTTON et al., 2018; MEZA et al. 2020).

Ocorre que os recursos eólicos, assim como os hídricos, apresentam variabilidade também em maiores escalas temporais, que vão desde semanas, meses, anos, décadas e até séculos, sendo influenciados por uma série de fenômenos climáticos. Considerando a variabilidade em escala intrassazonal fenômenos como a Oscilação de Madden-Julian podem influenciar nos padrões de circulação e precipitação sobre a América do Sul (KAYANO, JONES e DIAS, 2009). Acima de um ano, a variabilidade climática interanual tem como principal fator de influência o fenômeno El Niño-Oscilação Sul (ENOS), impactando também na precipitação sobre várias regiões do Brasil (GRIMM, 2009).

Já em escalas temporais superiores a décadas, existem modos de variabilidade decenal ou multidecenal que ocorrem no Oceano Pacífico e Atlântico e que afetam o clima sobre a América do Sul (KAYANO e ANDREOLI, 2009a). Há, ainda, mudanças climáticas em escalas de tempo seculares associadas à composição atmosférica – notadamente a concentração de gases do efeito estufa (GEE) como o dióxido de carbono – e que podem impactar sobre padrões de circulação em nível global (MARENGO, 2009).

Em suma, a variabilidade de recursos como os hídricos e eólicos ocorre em múltiplas escalas temporais e associam-se a inúmeros fatores e fenômenos climáticos que ocorrem com frequência variável (Figura 4).



Figura 4 – Fatores e escalas temporais associadas à variabilidade de recursos naturais

Somando participação conjunta de quase 70% na composição da matriz elétrica, as fontes hídrica e eólica merecem devida ênfase por estarem interligadas aos padrões de

Fonte: Pulwarty e Sivakumar (2014)

circulação climática e, desta forma, sofrem influência direta de fenômenos climáticos em suas variadas frequências de ocorrência. Existe, nesse sentido, uma conexão entre a variabilidade da disponibilidade destes dois recursos e o planejamento e gestão do setor elétrico em diferentes escalas temporais. Se em escalas de curto prazo determinados fenômenos climáticos devem ser levados em consideração para a operação do sistema de maneira a evitar distúrbios no equilíbrio entre oferta e demanda por energia, em escalas de longo prazo a ocorrência de fenômenos climáticos de baixa frequência requerem atenção no que toca o planejamento geral da capacidade e expansão do sistema elétrico como um todo (Figura 5).





Fonte: adaptado Diagne et al. (2013) e Notton et al. (2018)

Escalas temporais acima de um ano são interessantes já que normalmente compreendem a ocorrência de alguns eventos climáticos extremos – como períodos de escassez de recursos hídricos usualmente referidos por períodos de seca hídrica. O setor elétrico no Brasil já sofreu múltiplas crises energéticas de diferentes durações e escopos geográficos, tendo uma frequência de ocorrência que figura entre 10 e 15 anos. Na maioria dos casos estas crises associaram-se a condições climáticas extremas que de alguma forma alteraram o volume médio de precipitação por um dado período temporal em determinadas regiões. Se historicamente o Brasil sempre foi dependente de hidroeletricidade, a vulnerabilidade às condições climáticas requer adoção de medidas estratégicas para melhor preparo às possíveis futuras crises (HUNT, STILPEN e FREITAS, 2018).

O planejamento da expansão do sistema elétrico é crucial para o desenvolvimento de um parque gerador que seja capaz de evitar eventuais interrupções de abastecimento de energia elétrica no longo prazo. Já fora constatado que para a crise do setor elétrico de 2001, mesmo que em condições hidrológicas desfavoráveis, teria sido possível resistir à variabilidade natural dos recursos hídricos caso a expansão da capacidade instalada não tivesse ficado aquém da evolução do consumo de energia elétrica (BRANCO et al., 2002).

Diferentemente de 2001, o cenário atual é relativamente mais complexo, muito em função da diversidade das fontes energéticas sendo exploradas, da pluralidade de atores envolvidos no contexto do sistema elétrico nacional (que muitas vezes apresentam interesses divergentes) e pelas próprias características básicas de funcionamento da geração, transmissão e distribuição de energia elétrica no Brasil que entraram em vigência após as reformas do setor elétrico. Logo, a compreensão, aprimoramento e resolução desta questão central do planejamento da expansão do parque gerador deve ser buscada de modo multidisciplinar.

Custos de operação e segurança operativa devem inexoravelmente ser considerados no que tange o planejamento da expansão do sistema de geração no Brasil. O balanço entre estes dois atributos será consequência direta da participação relativa de cada uma das fontes que futuramente comporão a matriz de energia elétrica. Por um lado, garante-se um baixo custo de operação mediante exploração das fontes hídrica, eólica, biomassa e solar – em detrimento da flexibilidade operacional. Por outro, garante-se a disponibilidade e flexibilidade por meio da exploração de recursos fósseis (de maior custo operacional), além das tecnologias de armazenamento, a exemplo de baterias e usinas hidrelétricas reversíveis (EPE, 2020a).

Em todo caso, a gestão da expansão de um sistema elétrico deve minimamente levar em conta a projeção da disponibilidade de recursos associados ao clima. É particularmente relevante compreender o passado climático e as condições sobre as quais se deram as disponibilidades de recursos hídricos e eólicos em escalas temporais de longo prazo. O entendimento das variabilidades climáticas interanual, decenal e multidecenal e seus efeitos sobre o Brasil representa um ativo para a realização da previsão das disponibilidades futuras destes recursos e, por conseguinte, permite introduzir elementos de suporte à tomada de decisão quanto ao planejamento de longo prazo da capacidade do setor elétrico brasileiro.

Tal qual existem decréscimos em disponibilidade de recursos hídricos interanualmente, existem também períodos em que a variabilidade interanual de recursos eólicos culmina em redução da velocidade do vento em determinadas regiões. Em 2015, por exemplo, verificou-se um abrangente e extenso período de baixas velocidades de vento de superfície nos Estados Unidos que impactou fortemente a geração de energia pela fonte eólica. Foi constatado que altas temperaturas de superfície na porção oeste-tropical do Oceano Pacífico contribuíram para o estabelecimento e manutenção destas anomalias de velocidade do vento. Desta forma, concluiu-se que a variabilidade interanual dos recursos eólicos é influenciada não somente pelo fenômeno ENOS em diferentes graus de intensidade, como também por outros modos de variabilidade do Pacífico Norte, indicando que fatores climáticos podem ser utilizados para antecipar futuros períodos de baixas velocidades de vento (LLEDÓ et al., 2018; MOHAMMADI e GOUDARZI, 2018).

Adicionalmente, é possível que a somatória dos potenciais de geração por ambas as fontes apresente um comportamento diferente daquele observado individualmente. Isto sugere a necessidade de exploração de séries temporais representativas para avaliar as probabilidades de ocorrência de uma seca conjunta de tais recursos – o que representaria um cenário alarmante em termos de geração de energia elétrica a nível nacional – viabilizando, por conseguinte, a revelação de informações pertinentes para o planejamento de longo prazo do parque gerador do setor elétrico brasileiro.

Certamente que uma capacidade de geração pelas fontes hídrica e eólica que seja dimensionada para suportar cenários de escassez conjunta destes recursos seria de interesse nacional a fim de garantir o suprimento em qualquer cenário adverso. Mesmo porque com a penetração da fonte eólica na matriz elétrica, em breve o sistema elétrico brasileiro aumentará em complexidade, com provável inclusão de peculiaridades que demandarão soluções para o planejamento e operação (CANTÃO et al., 2017).

Neste contexto, a compreensão da complementaridade entre recursos hídricos e eólicos em suas diferentes escalas temporais ganha relevância. Sabe-se que, intraanualmente (sazonalmente) as regiões NE e SE apresentam comportamento complementar para recursos hídricos e eólicos. Durante os meses do período chuvoso no SE (de novembro a abril) a intensidade do vento decresce no NE, ao passo que de maio a outubro o padrão é invertido – a intensidade do vento aumenta enquanto o potencial de geração pela fonte hídrica decresce em função do período seco para tais regiões (PALFI e ZAMBON, 2013; EIFLER NETO, RISSO e BELUCO, 2014).

Observa-se que, de fato, para todo o SIN a geração de energia elétrica pela fonte eólica e a Energia Natural Afluente (ENA – informações relativas ao potencial energético contido nas vazões de rios em condições específicas, vide seção 3.1.3) obedecem um padrão sazonal; quando em meses de menor disponibilidade de recursos hídricos o sistema é compensado pelo aumento em disponibilidade de recursos eólicos – como evidenciado pelos picos de geração eólica coincidindo com os vales de ENA (Figura 6).



Figura 6 - Histórico da geração de energia pelas fontes hídrica e eólica em todo o SIN

Fonte: Operador Nacional do Sistema Elétrico (ONS, 2020a) - acesso em 16/12/2020

Já em caráter multianual, a complementaridade apresenta importância não apenas pela variação ano a ano que ambos os recursos apresentam, mas também porque os reservatórios do parque gerador hidrelétrico são gerenciados multianualmente. Assim sendo, verifica-se, ainda que em análise preliminar que não se serviu de testes de significância estatística, potencial complementaridade em períodos plurianuais, favorecendo a opção de compor a matriz elétrica a preços competitivos apenas com recursos renováveis (CHADE RICOSTI e SAUER, 2013).

Estudos indicam que a variabilidade interanual de recursos eólicos aparenta ser menor em comparação à mesma para recursos hídricos, indicando que não se pode afirmar que existe um relacionamento interanual consistente entre tais recursos. A correlação entre séries temporais de recursos hídricos e eólicos com remoção do ciclo anual não mostrou-se significativa. Ainda assim, integrar a geração eólica ao SIN pode permitir a redução da variabilidade interanual conjunta entre ambos os recursos (SCHMIDT, CANCELLA e PEREIRA JR., 2016a).

Frente à expansão de novos empreendimentos hidrelétricos na região Norte, aumentar a participação da fonte eólica na matriz elétrica do Brasil parece fazer mais sentido. A primeira opção não é capaz de reduzir a necessidade de instalação de capacidade de geração por térmicas a combustíveis fósseis. A fonte eólica, por sua vez, pode fazer reduzir a variância total da geração anual, sobretudo se acompanhada da adoção de outras tecnologias como geração distribuída por solar fotovoltaica e armazenamento eletroquímico de curto prazo e se for
assessorada pela flexibilidade de armazenamento dos reservatórios das UHEs (SCHMIDT, CANCELLA e PEREIRA JR., 2016b; SCHMIDT, CANCELLA e PEREIRA JR., 2016c).

Diante do que fora exibido, parece que a expansão do setor elétrico terá a fonte eólica como principal direcionamento. E, segundo a EPE (2015), a carga por energia pode apresentar crescimento anual superior a 3%, totalizando 1340 TWh no ano de 2050. Portanto, o norteamento desta pesquisa envolve o dimensionamento e composição do parque gerador nacional que seja capaz de atender a carga futura em condições de escassez de recursos que denotem mínima disponibilidade para geração de energia elétrica.

Considerando o planejamento para a expansão do sistema elétrico brasileiro, fica evidente a necessidade de compreender e estimar os potenciais de cada fonte estocástica com um razoável grau de acurácia. Para tanto, é imprescindível que se reconheça os principais fatores e fenômenos climáticos que influenciam a variabilidade interanual de recursos eólicos e hídricos e que podem, consequentemente, modular para a ocorrência de períodos de escassez de cada um deles ou de uma escassez conjunta.

1.2 OBJETIVOS

O foco desta pesquisa é contribuir com o desenvolvimento e planejamento da expansão de longo prazo do sistema elétrico brasileiro partindo de uma extensiva exploração de informações sobre disponibilidade e variabilidade interanual de recursos eólicos e hídricos. Deste modo, o objetivo principal da pesquisa, que se relaciona com a possibilidade de reunir elementos que auxiliem processos de tomada de decisão acerca da composição da matriz elétrica a nível nacional no longo prazo, é:

 Avaliar o comportamento da geração combinada de energia pelas fontes eólica e hídrica em diferentes cenários de composição da matriz elétrica brasileira e em condições de escassez conjunta de recursos eólicos e hídricos.

Para tanto, a pesquisa transita pelos seguintes objetivos específicos:

 Identificar sub-regiões de uma extensa área do NE do Brasil que apresentem padrões de similaridade entre as características estatísticas dos recursos eólicos;

- Analisar séries temporais históricas que representem o passado da disponibilidade de recursos eólicos e hídricos, de maneira a investigar a ocorrência prévia de eventos extremos que remontem escassez destes recursos;
- Calcular as probabilidades de ocorrência de eventos extremos de escassez eólica e verificar a existência de periodicidades relevantes;
- Examinar índices representativos dos alguns modos de variabilidade climática em busca de relações de influência sobre a variabilidade interanual de recursos eólicos no Brasil;
- Simular a geração de energia elétrica pelas fontes eólica e hídrica com base em variadas composições da matriz elétrica e em condições de escassez destes recursos.

1.3 JUSTIFICATIVA

Em decorrência da participação crescente da fonte eólica e da participação expressiva da fonte hídrica na matriz elétrica brasileira, é pertinente explorar a interação da variabilidade da disponibilidade destes recursos do ponto de vista da geração de energia. Particularmente, a ocorrência de períodos de escassez de recursos hídricos é tema recorrente na literatura acadêmica, visto que a exploração energética destes é historicamente mais antiga. Por outro lado, períodos de escassez de recursos eólicos tornam-se cada vez mais relevantes à medida que a capacidade instalada desta fonte se expande. Apesar disso, poucos trabalhos abordam este tema do ponto de vista da geração de energia elétrica considerando o caso brasileiro, fazendo desta proposta de estudo algo de valor original.

Outrossim, a variabilidade da somatória do potencial de geração destes dois recursos pode apresentar um comportamento conjunto diferente de suas variabilidades individuais, revelando a possibilidade de ocorrência de uma escassez conjunta que denote disponibilidade mínima. Tal eventualidade tampouco foi avaliada cientificamente, no sentido de mapear ocorrências passadas e determinar probabilisticamente as chances de ocorrência futura – adicionando originalidade ao projeto.

Ademais, em se tratando de recursos cujas disponibilidades estão atreladas à variabilidade climática natural, a análise da influência de fenômenos climáticos de larga escala e com baixa frequência de ocorrência sobre eventos de escassez hídrica, eólica e conjunta, merece escrutínio. A busca por elementos que auxiliem na explicação da incidência destes

eventos faz-se relevante, visto que podem proporcionar implicações na previsão climática de longo prazo e em tomadas de decisão a nível societário no que toca o planejamento e dimensionamento do parque gerador brasileiro, sobretudo para que este seja capaz de assegurar o suprimento de energia elétrica à demanda futura.

Espera-se que os resultados desta pesquisa possam não apenas contribuir com o planejamento de longo prazo da capacidade instalada de geração de energia elétrica no Brasil, mas também alertar para a necessidade de prever a disponibilidade de recursos energéticos naturais em diferentes condições climatológicas de outras regiões geográficas.

1.4 HIPÓTESE

Esta pesquisa parte do pressuposto de que, em função da variabilidade interanual de recursos eólicos e hídricos, períodos de escassez conjunta destes são passíveis de ocorrência. Tal cenário representaria mínima disponibilidade para geração de energia, ameaçando, por conseguinte, a segurança do fornecimento energético a nível nacional.

Ademais, configura-se a hipótese de que fenômenos climáticos de grande escala e baixa frequência podem modular a variabilidade de recursos eólicos, influenciando na ocorrência de condições extremas de mínima disponibilidade para geração de energia.

Ainda, acredita-se ser possível quantificar a composição futura da matriz elétrica do Brasil com base em capacidade instalada de geração por estas duas fontes, de maneira que o parque gerador seja robusto o suficiente para atender à demanda por energia elétrica em períodos de escassez hídrica, eólica ou conjunta.

1.5 ESTRUTURA CAPITULAR DA TESE

A presente tese foi segmentada em oito capítulos. No primeiro, INTRODUÇÃO, um sucinto e atual panorama da geração de energia elétrica por recursos de natureza estocástica no Brasil é apresentado, bem como a problemática que justifica o estudo e os objetivos pretendidos por este trabalho. O capítulo 2 reúne toda a REVISÃO BIBLIOGRÁFICA dos conceitos reconhecidos como fundamentais à realização deste trabalho. Boa parte destes são novamente

abordados no capítulo 3, MATERIAIS E MÉTODOS, haja vista que, junto aos materiais e ferramentas analíticas, fazem parte dos procedimentos metodológicos adotados.

Os capítulos de 4 a 7 são dedicados cada um a uma etapa do desenvolvimento da pesquisa. O capítulo 4 envolve uma análise de clusterização de séries temporais de dados eólicos, a fim de realizar um pré-processamento e promover a redução da dimensionalidade do conjunto de dados. O capítulo 5 relata a execução de uma série de ferramentas da grande área da análise de séries temporais, contemplando análises de distribuições de frequência, ocorrências de eventos extremos, funções ortogonais empíricas e análise espectral. O capítulo 6 abarca a avaliação de alguns modos de variabilidade climática como possíveis moduladores de ocorrências de baixa disponibilidade de recursos eólicos.

No capítulo 7, buscou-se conectar os resultados obtidos nos capítulos anteriores com o tema central da tese de geração de energia. Nele há a descrição da simulação e avaliação da geração combinada de energia pelas fontes eólica e hídrica com base em variadas composições da matriz elétrica e em condições de escassez destes recursos. Por fim, o capítulo 8, CONCLUSÕES E CONSIDERAÇÕES FINAIS, traz o fechamento da tese, seguido das REFERÊNCIAS BIBLIOGRÁFICAS consultadas ao longo da pesquisa.

2 REVISÃO BIBLIOGRÁFICA

2.1 RECURSOS EÓLICOS E HÍDRICOS NO BRASIL

A distribuição geográfica de recursos hídricos e eólicos para a geração de energia elétrica no Brasil é dependente de regimes climáticos locais que naturalmente proverão o abastecimento dos empreendimentos de geração em operação. Nesse sentido, dadas as localizações dos empreendimentos hidrelétricos e eólioelétricos em operação (Figura 7), é possível observar determinada concentração dos primeiros sobre as regiões Sul (S), Sudeste (SE) e Centro-Oeste (CO), tal qual existe um aglutinamento dos segundos sobre as regiões Nordeste (NE) e S do Brasil.



Figura 7 – Empreendimentos de geração elétrica em operação no Brasil

Fonte: Agência Nacional de Águas (2017)

Embora as razões pelas quais se deram estas concentrações sejam de relevância histórica, não se faz necessário o aprofundamento destas questões perante o contexto da pesquisa. O que é salientado relaciona-se com as regiões de maior importância em termos de quantidade de energia gerada pelas fontes hídrica e eólica. Nesse sentido, tomando o SIN e seus quatro subsistemas (N, NE, SE/CO e S – Figura 8), verifica-se que a expressividade da geração

é maior no subsistema Sudeste/Centro-Oeste para a fonte hídrica (Figura 9), responsável por aproximadamente 65% da geração hídrica total do SIN e no subsistema Nordeste para a fonte eólica (Figura 10), responsável por cerca de 85% da geração eólica total do SIN (ONS, 2020a).



Figura 8 – Subsistemas do SIN

Fonte: adaptado de ONS (2017)





Fonte: ONS (2020a) - acesso em 16/12/2020



Figura 10 – Histórico da geração de energia pela fonte eólica em cada subsistema do SIN

2.1.1 Recursos eólicos que alimentam o subsistema NE do SIN

Em escala global, a formação dos ventos deve-se -se ao aquecimento desigual da Terra, dado que a intensidade de radiação solar é maior em regiões equatoriais e menor em regiões polares. Gradientes de pressão atmosférica resultam desta não-uniformidade, fazendo com que massas de ar se movimentem de regiões de alta pressão para outras de baixa pressão. O NE brasileiro é atingido pelos ventos alísios de sudeste, que consistem em movimentos contínuos próximos à superfície em direção às menores pressões da faixa equatorial, desviados na direção contrária à rotação da Terra. O dado fenômeno é resultado da atuação da força de Coriolis, pois a rotação da Terra tende a desviar massas de ar que se movimentam do polo ao Equador, provocando desvio perpendicular ao movimento e no sentido oposto ao da rotação terrestre. Esse desvio resulta em baixos níveis nos ventos alísios de sudeste no Hemisfério Sul e de nordeste no Hemisfério Norte (Figura 11). Em altitudes mais altas, o vento segue em direção ao polo até aproximadamente 30 graus de latitude, num padrão cíclico conhecido como célula de Hadley ou célula tropical (CUSTÓDIO, 2013). A Figura 11 ilustra conceitualmente a circulação atmosférica em escala global, evidenciando as células de Hadley e os ventos alísios.



Figura 11 - Modelo conceitual da circulação global atmosférica

Fonte: Martins, Guarnieri e Pereira (2008)

Dada a previamente citada complementaridade hidroeólica, na qual em períodos de seca hídrica os recursos eólicos parecem aumentar em disponibilidade para geração de energia (Figura 7), é possível que os sistemas climáticos que influenciam na disponibilidade de recursos hídricos na região influenciem também na disponibilidade de recursos eólicos de maneira inversamente proporcional. Nesse sentido, grandes sistemas climáticos como os Anticiclones Subtropicais do Atlântico Sul (ASAS) e do Atlântico Norte (ASAN) e a Zona de Convergência Intertropical (ZCIT), comumente associados à precipitação no NE, associam-se também à intensificação ou enfraquecimento dos ventos alísios que sopram ali.

Localizada sobre o eixo do cavado equatorial (região de baixa pressão sobre o Equador, vide Figura 11), a ZCIT é uma extensa região de convergência dos ventos alísios de nordeste no Hemisfério Norte (HN) e de sudeste no Hemisfério Sul (HS) que se caracteriza por movimentos ascendentes, baixas pressões, nebulosidade e chuvas abundantes. Sua posição é variável, migrando sazonalmente de sua posição mais ao norte (em torno de 14N), durante agosto-setembro, para sua posição mais ao sul (em torno de 2S), durante março-abril. Migração esta que, associada a fatores que podem fortalecer ou enfraquecer os ventos alísios, possui papel importante na determinação da estação chuvosa do NE do Brasil (MOLION e BERNARDO, 2002; KAYANO e ANDREOLI, 2006; MELO, CAVALCANTI e SOUZA, 2009).

A ZCIT pode ser afetada por um conjunto de sistemas meteorológicos que atuam sobre a faixa equatorial dos oceanos, a exemplo da região de convergência dos alísios de sudeste e nordeste e das áreas de máxima Temperatura da Superfície do Mar (TSM), conectando-se com a atuação do ASAS e do ASAN. Desta forma, estações chuvosas deficientes ou abundantes no NE podem depender do deslocamento do ASAS e ASAN, tendo em vista que em anos de seca o ASAS se expande em direção ao Equador com intensificação dos alísios de sudeste sobre o NE, enquanto o ASAN se retrai a latitudes mais altas no HN. O padrão é invertido em anos chuvosos (MOLION e BERNARDO, 2002; KAYANO e ANDREOLI, 2006; MELO, CAVALCANTI e SOUZA, 2009).

Outrossim, a variabilidade interanual do deslocamento norte-sul da ZCIT é associada ao modo de variabilidade do Oceano Atlântico, em que anomalias assimétricas de TSM oscilam entre as bacias norte e sul do Atlântico Tropical, gerando um gradiente de temperatura interhemisférico sobre o Atlântico Equatorial. Exemplo disso é que em anos de precipitação intensa no NE, o gradiente de TSM é mais forte no sentido de norte para sul e a ZCIT se desloca para posições mais ao sul (5S, Figura 12b). Já em anos de seca hídrica ela permanece bloqueada mais ao norte, influenciada por um gradiente de TSM mais forte no sentido de sul para norte (Figura 12^a) (MOLION e BERNARDO, 2002; KAYANO e ANDREOLI, 2006; MELO, CAVALCANTI e SOUZA, 2009).

A Figura 12 ilustra este efeito sobre a ZCIT, onde os sistemas de alta pressão de ambos os hemisférios é indicado pela letra A e os ventos alísios pelas setas adjacentes.



Figura 12 – Posicionamento da ZCIT em (a) anos secos e (b) anos chuvosos no NE

Fonte: Melo, Cavalcanti e Souza (2009)

Outro fator possivelmente relacionado tanto com a variabilidade da TSM sobre o Atlântico Tropical quanto com o posicionamento da ZCIT é o fenômeno ENOS. Padrões de teleconexões atmosféricas indicariam que as anomalias de TSM sobre o Pacífico Equatorial induziriam a formação de anomalias do gradiente meridional de TSM sobre o Atlântico Equatorial. No setor norte do NE verificou-se anomalias negativas de precipitação associados a episódios de El Niño. Em 2000, ano de evento La Niña, houve intensificação da ZCIT,

inclusive associando-se a um aumento da passagem de frentes frias que penetraram até latitudes equatoriais sobre o Atlântico. Ainda assim, é a variabilidade de TSM do Atlântico Tropical que hoje é tida como determinante para as anomalias de precipitação e intensidade dos alísios sobre o NE, enquanto ocasionalmente o ENOS pode reforçá-las ou enfraquecê-las (MOLION e BERNARDO, 2002; KAYANO e ANDREOLI, 2006; MELO, CAVALCANTI e SOUZA, 2009).

A variabilidade interanual do clima sobre o NE pode, ainda, ser modulada por fenômenos de baixa frequência, como a Oscilação Decadal do Pacífico (PDO). A interação complexa entre o ENOS e a PDO pode resultar em diferentes anomalias de precipitação e de intensidade dos ventos alísios no NE, a depender das fases de cada modo de variabilidade que atuam sobre o Pacífico. Ainda assim, a variabilidade do clima do NE depende mais do Atlântico que do Pacífico (KAYANO e ANDREOLI, 2007).

2.1.2 Recursos hídricos que alimentam o subsistema SE/CO do SIN

Os recursos hídricos que abastecem as UHEs do subsistema SE/CO são dependentes dos regimes de chuvas a que regiões em diferentes escalas espaciais estão sujeitas. Em boa parte destas macrorregiões do SE e CO do Brasil a estação chuvosa tem início em meados de outubro, com ocorrência de chuvas mais intensas e frequentes entre dezembro e fevereiro e enfraquecimento do volume de precipitação após março/abril (CARVALHO e JONES, 2009).

Além do padrão do ciclo anual de precipitação, os totais anuais de precipitação se assemelham em quase toda a área do SE e CO, excedendo 1500 mm/ano (AGÊNCIA NACIONAL DE ÁGUAS - SISTEMA NACIONAL DE INFORMAÇÕES SOBRE RECURSOS HÍDRICOS, 2016). Todavia, cada área sofre atuação de sistemas climáticos diferentes. Os sistemas atmosféricos que atuam no CO são tanto de origem tropical quanto extratropical. O setor mais ao norte sofre influência de sistemas que atuam na Amazônia, já a região mais ao sul sofre influência de sistemas extratropicais, como frentes frias e linhas de instabilidade. Para o SE, eventos de chuva ocorrem quando os sistemas frontais e ciclones subtropicais e extratropicais se intensificam (REBOITA et al. 2010).

Este ciclo anual da chuva sobre o SE e CO é marcado pela presença, no verão austral, de uma banda de nebulosidade e atividade chuvosa com orientação noroeste-sudeste, se

estendendo desde a Amazônia, passando pelo SE e por vezes chegando até o Oceano Atlântico em sua porção subtropical (Figura 13). Esta faixa associa-se a um escoamento convergente de umidade na baixa troposfera e é denominada Zona de Convergência do Atlântico Sul (ZCAS) (CARVALHO, JONES e LIEBMANN, 2004).





Fonte: adaptado de Pegorim (2017)

Existe uma certa regularidade da atuação da ZCAS, no sentido em que todos os verões as suas características são observáveis. Contudo, variações quanto ao posicionamento e extensão da ZCAS determinam variações na precipitação macrorregional que, por sua vez, permitem a ocorrência de eventos extremos, como secas e enchentes em diferentes porções do território brasileiro. Em anos de eventos intensos de ZCAS, máximos de precipitação ocorrem na sua região de domínio e mínimos sobre os subtrópicos (DRUMOND e AMBRIZZI, 2006; CARVALHO e JONES, 2009)

A variabilidade da ZCAS também é marcada em diferentes escalas temporais, desde intrasazonais (10-100 dias), interanuais ou até mesmo decenais. Particularmente, em escalas interanuais o fenômeno ENOS parece ter efeito modulador sobre o padrão oscilatório do regime de chuvas sobre a América do Sul (DRUMOND e AMBRIZZI, 2006).

Similarmente, um ciclo de 15 anos para as temperaturas de superfície do Atlântico parece coerente com vazões do rio Paraná – que passa pela região hidrográfica do Paraná e alimenta várias UHEs do subsistema SE/CO, dentre elas, Itaipu, sendo talvez a bacia mais

relevante para o SIN em termos de geração pela fonte hídrica (Figura 14) (ROBERTSON e MECHOSO, 2000).



Figura 14 – Aproveitamento hidrelétrico na região hidrográfica do Paraná

Fonte: adaptado de Agência Nacional de Águas (2015)

As frentes frias contribuem também para a intensificação dos regimes de precipitação sobre as regiões SE e CO. Quando avançam em direção ao equador, as frentes frias interagem com o ar úmido e quente tropical, formando sistemas convectivos de fortes e excessivas chuvas sobre o continente – por vezes provocando enchentes e deslizamentos – principalmente durante o verão austral e na condição de encontro com a ZCAS (CAVALCANTI e KOUSKY, 2009).

A passagem de frentes frias possui padrão sazonal, com maior frequência de maio a setembro, durante o outono e inverno e menor frequência durante o verão austral. Observa-se um gradiente sobre regiões S, CO e costa leste do Brasil (Figura 15), sendo que a frequência de ocorrência decresce à medida que decresce também a latitude (CAVALCANTI e KOUSKY, 2009).



Figura 15 – Quantidade média anual de passagens de frentes frias de 1979 a 2005

Fonte: Cavalcanti e Kousky (2009)

Durante a primavera austral, foi constatado que a variabilidade interanual da frequência de ocorrência da passagem de frentes frias sobre o SE parece ser modulada pelo fenômeno ENOS, o que indica que o regime de chuvas é também afetado nesta estação do ano (BLÁZQUEZ e SOLMAN, 2017). O número de 51 frentes frias que atingiu a região S no ano de 2001; foi o mínimo do período de 1980-2002. Por outro lado, o máximo deste período na mesma região foi de 80 ocorrências no ano de 1984, quando se iniciava o fenômeno La Niña (ANDRADE *apud* CAVALCANTI e KOUSKY, 2009). Desta forma, a frequência da passagem de frentes frias torna-se um indicativo de eventos de precipitação máxima e mínima sobre regiões de interesse para a geração pela fonte hídrica.

Todos esses e outros fatores e fenômenos climáticos atuam simultaneamente em interações complexas para determinar a quantidade de chuvas que pode ocorrer em um dado ano em cada região do território brasileiro. Consequentemente as vazões dos rios abastecidos por esse regime de chuvas são indiretamente afetadas para que, por fim, o potencial de geração de energia pela fonte hídrica seja também associado às mesmas interações climáticas. A variabilidade interanual dos fenômenos climáticos é, assim, ligada à variabilidade interanual das vazões de rios que alimentam o parque gerador hidrelétrico.

2.1.3 Eventos de escassez de recursos eólicos e hídricos

A variabilidade da disponibilidade de recursos hídricos e eólicos em escalas temporais interanual, decadal e/ou superior permite a ocorrência de períodos de escassez com duração relativamente prolongada (meses, estações, anos). O interesse em torno do tema de secas hídricas e eólicas é pertinente na medida em que a baixa disponibilidade destes recursos impacta diretamente na geração de energia elétrica a nível nacional.

Os períodos de seca constituem desvios das médias climáticas de longo prazo, todavia são fenômenos que fazem parte da variabilidade climática natural. Em particular, as secas hídricas podem ocorrer em qualquer zona climática – em regiões de alto ou baixo volume anual médio de precipitação – e estão majoritariamente relacionadas com a redução da precipitação recebida ao longo de um período superior a uma estação ou um ano. As razões para sua ocorrência são complexas, não apenas pela dependência de fatores atmosféricos, mas também de processos hidrológicos que alimentam umidade à atmosfera. Temperaturas, ventos de alta magnitude, baixa umidade relativa, momento de ocorrência e características das chuvas (incluindo distribuição de dias chuvosos, intensidade e duração da chuva) contribuem significativamente para a ocorrência de secas hídricas. Uma vez estabelecidas condições hidrológicas para a seca, um mecanismo de retroalimentação é ativado, pelo qual o esgotamento da umidade de camadas superiores do solo reduz as taxas de evapotranspiração das plantas, reduzindo, por fim, a umidade relativa da atmosfera e a probabilidade de formação de chuvas - dado que será mais difícil alcançar condições de saturação num sistema de baixa pressão atuante sobre a região em questão. Apenas distúrbios que podem carregar suficiente umidade de fora da região seca para dentro dela serão capazes de produzir precipitação o bastante para encerrar as condições de seca (MISHRA e SINGH, 2010; WORLD METEOROLOGICAL ORGANIZATION (WMO) e GLOBAL WATER PARTNERSHIP (GWP), 2016; MARENGO, TORRES e ALVES, 2017).

Em contraste com aridez, cuja particularidade climática permanente é restrita a áreas de baixa pluviometria, uma seca hídrica é uma aberração temporária. Outrossim, a seca hídrica diferencia-se de um outro fenômeno conhecido como onda de calor pela escala temporal de sua duração: enquanto a primeira pode persistir por meses ou anos, a segunda apresenta duração típica de uma semana. A combinação de ambos os fenômenos é possível e passível de desastrosas consequências socioeconômicas (MISHRA e SINGH, 2010; MARENGO, TORRES e ALVES, 2017).

Em função da dependência da geração pela fonte hídrica, o setor elétrico no Brasil já sofreu múltiplas crises energéticas de diferentes durações e escopos geográficos, tendo uma frequência de ocorrência que figura entre 10 e 15 anos. Na maioria dos casos estas crises associaram-se a condições climáticas extremas que de alguma forma alteraram o volume médio de precipitação por um dado período temporal em determinadas regiões, vide Tabela 1 (HUNT, STILPEN e FREITAS, 2018).

Período	Estados/Regiões	Principais causas
1924-1925	São Paulo	Seca no rio Tietê e seus afluentes
1938-1947	São Paulo	Dificuldade de importação de equipamentos em função da 2a Guerra Mundial
1950-1957	Sudeste	Seca e aumento da demanda
1951-1964	Rio de Janeiro	Seca no rio Paraíba e falta de capacidade geradora
1963-1964	São Paulo e Rio de Janeiro	Seca: drástica redução na vazão dos rios Paraíba e Piraí
1967	Rio de Janeiro	Alagamento da usina Nilo Peçanha
1986	Sul	Seca na região Sul
1987-1988	Nordeste	Seca na região Nordeste
1995-1999	Manaus	Falta de investimento em função de regulamentação deficiente
2001-2002	Escopo nacional (exceto Sul)	Seca nas regiões Sudeste, Nordeste e Norte e atraso na expansão de geração térmica e linhas de transmissão
2014-2015	Escopo nacional	Seca nas regiões Sudeste, Nordeste, Sul e Norte
2012-2017	Nordeste	Seca na região Nordeste

Tabela 1 - Crises do setor elétrico brasileiro e suas principais causas

Fonte: adaptado de Hunt, Stilpen e Freitas (2018)

A determinação exata do início, fim e duração de uma seca é de difícil realização, muito em função da não existência de uma definição precisa e universal de seca hídrica – uma vez que as variáveis hidrometeorológicas e a natureza estocástica dos recursos hídricos são distintas para diferentes regiões do mundo. As definições existentes para seca hídrica dependem das variáveis utilizadas para descrevê-la, podendo ser classificada em cinco categorias: a) seca meteorológica, definida como escassez de precipitação sobre uma região por um período temporal; b) seca hidrológica, relacionada a um período de inadequação entre os recursos hídricos de superfície e/ou subterrâneos e os usos múltiplos da água de um dado sistema de gerenciamento hídrico; c) seca agrícola, referente a um período de declínio da umidade no solo e consequente quebra da produção agrícola sem nenhuma relação com os recursos hídricos de superfície; d) seca socioeconômica, associada com a falha de sistemas de recursos hídricos para atender à demanda por água pela sociedade; e) seca de água subterrânea, quando há decréscimo no nível e na recarga de reservatórios subterrâneos de água (MISHRA e SINGH, 2010). Comumente, um índice de seca hídrica é uma ótima opção para avaliar tanto quantitativamente quanto qualitativamente o seu efeito e para definir diferentes parâmetros a ela associados, incluindo intensidade, duração, severidade e extensão espacial. Os índices são representações numéricas computadas a partir de variáveis/indicadores climáticos ou hidrometeorológicos como precipitação, temperatura, vazão de um rio, nível de um reservatório, umidade do solo e camada de neve. O uso de índices é interessante no que tange o monitoramento do clima em diferentes escalas temporais, dado que permite identificação de períodos úmidos de curto prazo dentro de secas prolongadas ou períodos secos de curto prazo dentro de períodos úmidos de longo prazo (WMO e GWP, 2016).

Salienta-se que a avaliação da seca requer que a variável ou índice seja capaz de quantificá-la em diferentes escalas de tempo, portanto uma série temporal com dados anuais, mensais ou diários é de suma importância. A identificação de um período seco deve ser pautada pela definição de um truncamento ou limiar, podendo ser uma constante ou função variando no tempo (MISHRA e SINGH, 2010).

Inúmeros índices são propostos na literatura científica, fazendo com que a seleção de um deles seja resultado de um processo reflexivo. As variáveis e indicadores considerados no cálculo do índice, o tipo de clima do local a ser monitorado, bem como o tipo de seca que será avaliado são fatores que devem ser levados em consideração para uma seleção apropriada. Tomando a produção de energia pela fonte hídrica, faz sentido a escolha de um índice que inclua variáveis como vazões de rios, níveis de reservatórios e precipitação em seu cálculo.

Desta forma, alguns índices desenvolvidos para monitoramento da hidrologia de uma região podem ser de grande valia. Dois deles requerem apenas dados de médias mensais de vazões de rios: *Standardized Streamflow Index* (SSFI) e *Streamflow Drought Index* (SDI). Outro requer dados de médias mensais de níveis de reservatórios: *Stantardized Reservoir Supply Index* (SRSI). Estes três são baseados num outro índice que avalia a precipitação e que é amplamente utilizado no monitoramento de secas do ponto de vista meteorológico e foi, inclusive, recomendado pela WMO como o principal índice a ser adotado para avaliação de secas meteorológicas: *Standardized Precipitation Index* (SPI). Assim, SSFI, SDI e SRSI utilizam a mesma modelagem do SPI para analisar séries temporais e desenvolver distribuições probabilísticas para diferentes escalas de tempo (WMO e GWP, 2016).

O cálculo de índices de seca hídrica para uma dada região deve ser realizado com cautela. Por exemplo, o índice SPI é um produto definido com base em dados de precipitação,

então se houver frequência alta de registros numéricos significativos de zero precipitação (comum em regiões com regime climático seco), a distribuição probabilística poderá ser enviesada. Limitações e lacunas nas séries temporais, simular dados de precipitação utilizando uma distribuição probabilística pode causar erros no cálculo do índice caso a distribuição não for adequadamente selecionada. Vale ressaltar que a precipitação muda ao longo do tempo, portanto é importante averiguar os parâmetros da distribuição em diferentes escalas temporais. Para uma eficaz categorização da severidade da seca é importante definir limiares adequados para diferentes regiões hidroclimáticas (MISHRA e SINGH, 2010). Eventos de seca hidrológica normalmente são avaliados perante um limiar preestabelecido que pode variar em função do tempo, sendo estimado conforme uma janela sazonal móvel e observações deste período. Não obstante, para a comparação de regimes de seca e severidade obtidos de diferentes fontes de energia e/ou diferentes regiões, os limiares não devem ser invariáveis no tempo, pois os eventos de seca devem ser definidos per se, e não como eventos incomuns quando comparados com um calendário climatológico (RAYNAUD et al., 2018). Por estas razões, optou-se pela não utilização destes índices para representar eventos de seca, sendo os próprios dados de vazão de rios diretamente analisados para indicar disponibilidade de recursos hídricos (vide seção 5.2.3).

Entende-se que os mesmos princípios e premissas adotadas para a avaliação, caracterização e monitoramento de secas hídricas podem ser utilizados também para recursos eólicos. O fato é que o fenômeno da escassez eólica, embora apresente similaridades com a seca hídrica do ponto de vista climático, ganhou relevância apenas recentemente, quando foi constatada uma diminuição em disponibilidade de recursos eólicos nos Estados Unidos no primeiro e segundo trimestres de 2015, ano em que o aproveitamento energético desta fonte era já significativo e bem estabelecido neste e em outros países do mundo.

Na ocasião, a redução da velocidade do vento em superfície – em especial nos estados do Texas, Oklahoma e Kansas, onde estão concentrados os maiores parques eólicos – não foi prevista pela indústria eólica, trazendo problemas financeiros a empresas do setor pela baixa produção e baixa receita no período de escassez. Estima-se que as anomalias normalizadas da variável velocidade de vento alcançaram três desvios padrão de distância da média de longo prazo (de 1979-2014) por uma vasta área, revelando ser um fenômeno infrequente. O evento ocorrido foi inicialmente associado ao ENOS em sua fase El Niño que, entretanto, não se encontrava fortemente estabelecido, apesar de historicamente modular a variabilidade interanual da velocidade do vento sobre os Estados Unidos. Assim, constatou-se que a seca eólica durante o primeiro trimestre de 2015 pode ter sido influenciada pelo modo de

variabilidade do Pacífico Norte, que se encontrava em sua fase quente com anomalias positivas de TSM sobre as regiões norte e tropical oeste (LLEDÓ et al., 2018).

A Figura 16 registra o enfoque na região retangular laranja, onde os recursos eólicos tornaram-se menos disponíveis para geração de energia, com acentuação da coloração azul conforme escala utilizada. A velocidade média do vento para o primeiro trimestre de 2015 chegou a ser reduzida em 3 unidades de desvio padrão (considerando as séries temporais normalizadas). Embora não exista um limiar consensual para indicar que trata-se de um evento extremo de escassez de recursos eólicos, esta abordagem é interessante no sentido em que compara o desempenho atual frente ao próprio histórico do local







O desenvolvimento de índices de vento para o monitoramento e avaliação de possíveis secas eólicas atreladas à geração de energia, tal qual existem para secas hídricas, ganha relevância na medida em que se intensifica o aproveitamento energético mundial de recursos eólicos. Um índice de vento pode prover uma noção da magnitude da velocidade do vento para períodos de interesse em relação à sua média calculada para o longo prazo. A variabilidade interanual do vento deve ser compreendida e, assim, derivar estimativas da produção energética de parques eólicos, não apenas do ponto de vista da estrutura financeira acerca de seus desenvolvimentos, mas sobretudo da perspectiva de projeções climáticas acuradas – a ponto de influenciar a elaboração de políticas relacionadas ao fornecimento de energia em dimensões nacionais (PRYOR, BARTHELMIE e SCHOOF, 2006; WATSON, KRITHARAS e HODGSON, 2015).

Índices de vento podem ser divididos em quatro amplas categorias: a) derivados diretamente de observações de velocidade de vento em superfície; b) provenientes de observações diretas da geração de energia em parques eólicos ou seleção de aerogeradores; c) procedentes de gradientes de pressão na forma de ventos geostróficos triangulados de observações de pressão; d) resultados de modelos numéricos de previsão do tempo e assimilação de observações, a exemplo dos dados de reanálise.

A primeira categoria possui a vantagem de indicar variações no clima eólico local, porém pode ser sensível a efeitos muito regionalizados. A segunda possui vantagens similares à primeira, no entanto pode fornecer séries temporais com falta de homogeneidade devido a mudanças no portfólio de aerogeradores além da indisponibilidade de máquinas por interrupções variadas. A terceira e quarta categorias possuem a vantagem de filtrar a maioria das anomalias em escala local que surgem pelo microclima, orografia, mudanças no entorno, mudanças na instrumentação etc. Por outro lado, usar dados suavizados espacialmente pode mascarar a variabilidade regional, desprezando as características locais e subestimando a variabilidade interanual do vento (WATSON, KRITHARAS e HODGSON, 2015).

A ocorrência de secas eólicas é tema relativamente novo. De fato, o único artigo científico que especificamente menciona o termo seca eólica foi o supracitado de Lledó et al. (2018) sobre o caso norte-americano. Foi também relatado que em 2010 a região noroeste da Europa passou por período de condições eólicas muito fracas. Medições na velocidade do vento na Holanda, Alemanha e Dinamarca apontaram para a mais baixa disponibilidade de recursos eólicos desde que os registros sistemáticos do vento começaram a ser realizados já na década de 1980 (BAKKER, VAN DEN HURK e COELINGH, 2013).

Em artigo explorando duas regiões diferentes da Espanha, a implementação de redução de escala em séries temporais de velocidade do vento para o período de 1871-2009 permitiu avaliar a variabilidade de longo prazo em conformidade com o impacto de padrões prevalecentes de circulação atmosférica de larga escala. Embora o texto não mencione a ocorrência de períodos de escassez de vento, é possível analisar que as séries temporais apresentam mínimos que remetem a períodos de baixa disponibilidade de vento para geração de energia (KIRCHNER-BOSSI et al., 2015).

Considerando dados de resolução espacial não tão alta para o mundo como um todo, verificou-se que a variabilidade interanual média do recurso eólico é de 11% (para mais ou para menos variando mês a mês) – sendo mais intensa sobre áreas oceânicas do que continentais. Entretanto, o fato dos fenômenos ENOS e PDO serem moduladores de longo prazo fazem com que a magnitude da variação interanual da velocidade do vento decresça quando a duração do período avaliado é aumentada (KRAKAUER e COHAN, 2017).

No Brasil, um estudo apontou para anomalias negativas na velocidade do vento na região NE nos anos de eventos La Niña, mas os resultados não foram traduzidos em termos de produção de energia e tampouco é possível afirmar se configuram anos de baixa intensidade da velocidade do vento (SANTOS e SILVA, 2013). Similarmente, recursos eólicos em regiões do Chile e Argentina são influenciados pelo fenômeno ENOS em escalas mensais e sazonais (principalmente durante a primavera austral), indicando que pode ser utilizado como elemento prognosticador da produção de energia eólica (BIANCHI, SOLARTE e GUOZDEN, 2017).

Outro artigo aborda a complementaridade hidroeólica interanual no longo prazo para o Brasil, revelando que adicionar capacidade de geração pela fonte eólica pode promover a redução da variabilidade interanual das duas fontes somadas. Contudo, tal estudo não enfatiza os períodos de escassez destes recursos energéticos e nem avaliam se há influência de fenômenos climáticos de baixa frequência como moduladores da variabilidade (SCHMIDT, CANCELLA e PEREIRA JR., 2016a).

Relatos acerca da ocorrência de secas eólicas e do impacto destas na geração de energia no Brasil não são comuns, reforçando a necessidade da realização desta pesquisa. Em contrapartida, a ocorrência de secas hídricas no Brasil é muito bem documentada. Tomando a região de interesse da bacia hidrográfica do Paraná (considerando geração pela fonte hídrica), durante o período de 1995 a 2014 os anos hidrológicos (setembro-agosto) de 2001 e 2014 foram os mais críticos em termos de anomalias negativas de precipitação em relação à média

climatológica de 1982-2015 (Figura 17). Esta seca categorizada como meteorológica se transforma em seca hidrológica dado o reflexo da redução de chuvas sobre as vazões de rios da bacia do Paraná (MELO et al., 2016).



Figura 17 – Anomalias de precipitação relativas ao período de 1982-2015: anos hidrológicos de 1995-2014

Não por acaso, estes dois anos de seca hídrica proporcionaram redução no armazenamento de água nos reservatórios das UHEs e, por conseguinte, queda na produção de energia. Em conjunto com a seca de 2001, durante a década de 1990 o aumento no consumo por eletricidade se deu em taxa maior que o aumento da capacidade instalada de geração, contribuindo para a ocorrência de blecautes e culminando com a exigência, por parte do Governo, de uma política de racionamento de energia com redução no consumo em cerca de 20%. Tal episódio provocou consequências socioeconômicas para a população, que passou a

exigir maior grau de confiabilidade no fornecimento energético. Foi então que o país passou por reforços na infraestrutura de geração, transmissão e distribuição de energia, inclusive com aumento na capacidade instalada e diversificação da matriz elétrica para aliviar a dependência pela fonte hídrica. Não obstante, o cenário se repetiu nos anos de 2014 e 2015, quando uma estiagem na região SE provocou vazões abaixo do normal em diversos rios dos estados Minas Gerais, São Paulo e Rio de Janeiro, ocasionando diminuição do armazenamento na maior parte dos reservatórios do SIN. O nível médio dos reservatórios atingiu o mínimo das últimas duas décadas. Desta vez não houve racionamento; todavia, as usinas térmicas foram acionadas para operar na base em detrimento das vazões turbinadas de várias UHEs, o que causou um aumento nas tarifas de eletricidade (ANA, 2017; HUNT, STILPEN e FREITAS, 2018).

As anomalias negativas de precipitação durante a seca de 2014 mostraram-se resultado de um conjunto de condições atmosféricas e oceânicas que fizeram da ZCAS menos expressiva no período. Tais condições incluem anomalias positivas de TSM sobre o Pacífico Equatorial Oeste e sobre o Atlântico Sudoeste (próximo à costa do Brasil), além de um centro de alta pressão atuando sobre esta mesma região no Atlântico e se estendendo à região SE. Tal anticiclone transportava ar mais seco que o normal do Atlântico até o SE e desviava o fluxo úmido da Amazônia em direção à região S (COELHO et al., 2016).

Além do SE, a região do NE sofre regularmente, com secas reportadas desde o século XVI. Relatos apontam para a ocorrência de escassez hídrica nos anos de 1583, 1603, 1624, 1692, 1711, 1723-1724, 1744-1746, 1754, 1760, 1772, 1766-1767, 1777-1780, 1784, 1790-1794, 1804, 1809, 1810, 1816-1817, 1824-1825, 1827, 1830-1833, 1845, 1877-1879, 1888-1889, 1891, 1898, 1900, 1902-1903, 1907, 1915, 1919, 1932-1933, 1936, 1941-1944, 1951-1953, 1958, 1966, 1970, 1976, 1979-1981, 1982-1983, 1992-1993, 1997-1998, 2001- 2002, 2005, 2007, 2010 e 2012-2017. Ademais, projeções relacionadas às mudanças climáticas de origem antropogênica e às mudanças no uso da terra indicam que a região do NE do Brasil será cada vez mais suscetível à ocorrência de secas que, em conjunto com outros fatores como desmatamento, aumento nas temperaturas e na frequência de incêndios e superexploração de água subterrânea, podem desencadear um processo de desertificação, transformando o regime climático de semiárido para árido (MARENGO, TORRES e ALVES, 2017).

Secas severas foram observadas no NE, além de outras áreas, nos anos de 2012-2017 (Figura 18), resultado de baixas taxas de precipitação. Vale ressaltar que a probabilidade de

ocorrência de totais anuais de chuva e de vazão média anual dessa magnitude (muito baixos) ocorrerem em tais locais era menor do que 1%, ou 100 anos de tempo de retorno¹ (ANA, 2017).



Figura 18 – As crises hídricas de 2012 a 2017 em suas respectivas regiões

Fonte: Agência Nacional de Águas (2017)

Para o período de 2012-2017, as causas dos baixos índices de precipitação e da dinâmica tendencial interanual da chuva incluem fatores meteorológicos, hidrológicos, demográficos e até mesmo institucionais em diferentes proporções. Em alguns estados do NE como Ceará, Rio Grande do Norte, Paraíba e Pernambuco, baixos índices de precipitação, a irregularidade do seu regime, reduzida disponibilidade de águas subterrâneas, temperaturas elevadas durante todo o ano, baixas amplitudes térmicas, forte insolação e altas taxas de evapotranspiração, que normalmente superam os totais pluviométricos, contribuíram para o balanço hídrico desfavorável que se arrastou de 2012 a 2017, na mais severa e longa seca observada desde o início do século XX (ANA, 2017).

Como visto, anomalias de TSM sobre o Pacífico ou Atlântico podem influenciar anomalias de precipitação sobre regiões no Brasil. Nesse sentido, torna-se atrativo correlacionar a ocorrência de períodos de escassez de recursos hídricos ou eólicos com os índices para

¹ Tempo de retorno é uma expressão comumente utilizada em hidrologia e corresponde ao inverso da probabilidade. Assim, se um evento possui um tempo de retorno de 100 anos significa dizer que há 1 chance em 100 de que esse evento aconteça num ano qualquer (ANA, 2017).

diferentes modos de variabilidade climática, partindo do pressuposto que mudanças nos padrões de circulação atmosférica de larga escala podem afetar regiões distantes pelas denominadas teleconexões (MISHRA e SINGH, 2010).

No entanto, fenômenos como ENOS explicam apenas parte da variabilidade da pluviometria. Para o período de 1849-1992, de 46 eventos El Niño, somente 21 deles associaram-se com secas no NE brasileiro (KANE, 1997). Nas secas recentes de 1992, 1998, 2002, 2010 e de 2012-2017, apenas em 1998, 2002 e 2015-2016 ocorreram durante anos ENOS. Outros períodos de seca no NE são atribuídos ao posicionamento anômalo da ZCIT na região do Atlântico, em função de sua porção norte estar relativamente mais aquecida que a sul. É o caso das condições de seca hídrica que começaram a se intensificar em 2012, concomitante a anomalias positivas de TSM no Atlântico Norte que favoreceu o deslocamento da ZCIT para o HN e a redução de precipitação no NE. Em 2015-2016 o ENOS em sua fase quente (El Niño) agravou as condições e contribuiu para a manutenção da seca na região (CUNHA et al., 2018; MARENGO et al., 2018).

O interesse nas secas hídricas do NE se justifica mais pela geração eólica do que pela hídrica. Pois, como previamente abordado, a complementaridade hidroeólica é muito bem marcada numa escala de tempo sazonal, o que significa que em boa parte do Brasil, maior (menor) volume de precipitação ocorre em período de menor (maior) intensidade de ventos dentro de um mesmo ano (CANTÃO et al., 2017). Contudo, isto pode não se verificar em longos períodos de seca hídrica.

Desta forma – e considerando as principais regiões brasileiras em termos de geração de energia por cada uma dessas fontes – o que se busca é o entendimento da complementaridade da variabilidade interanual destes recursos, de modo a verificar se um período de escassez de um deles é passível de ser concomitante à escassez de outro. Também, procura-se avaliar se uma escassez conjunta destes recursos pode representar um cenário de potencial de geração ainda mais baixo do que os cenários de seca de apenas um deles.

2.2 MODOS DE VARIABILIDADE CLIMÁTICA DE BAIXA FREQUÊNCIA

Como previamente discutido, alguns sistemas climáticos possuem atuação marcante sobre as regiões mais relevantes para a geração de energia pelas fontes hídrica e eólica, a exemplo da ZCAS e frentes frias no SE/CO e ZCIT, ASAS e ASAN para o NE. Entretanto, existe uma interação complexa entre estes e outros sistemas² (Figura 19) que ocorre na baixa troposfera da América do Sul como um todo, além de regiões adjacentes e em outros níveis de altitude, e que de certo modo determina a variabilidade e disponibilidade de recursos hídricos e eólicos – evidentemente que as escalas espaciais e temporais desta influência são relativamente menores do que outros fenômenos climáticos abordados em seguida.





Fonte: adaptado de Reboita et al. (2010)

Estes sistemas atmosféricos e suas interações podem, por sua vez, serem regidos por ciclos oscilatórios denominados modos de variabilidade climática. Os oceanos possuem importante papel de atuação no sistema climático global em função de suas grandes capacidades de armazenamento de calor – aproximadamente 3,5 m de água contém energia equivalente a uma coluna atmosférica inteira. Embora seja comum a afirmação de que a atmosfera possui "memória curta" enquanto os oceanos, em função de sua grande inércia térmica, fornecem a

 $^{^2}$ Boa parta dos sistemas que atuam na América do Sul podem ser vistos na revisão elaborada e apresentada em Reboita et al. (2010).

memória para variações climáticas, é a colaboração interativa entre eles e outras partes do sistema climático que ativam variações climáticas ao redor da Terra. A inércia térmica dos oceanos é comunicada à atmosfera via transferência turbulenta de energia radiativa em suas superfícies. Tais fluxos de energia, por sua vez, dependem da TSM e de vários outros parâmetros atmosféricos, incluindo velocidade do vento, temperatura do ar, umidade e nebulosidade. Particularmente, variações vagarosas na TSM fornecem uma fonte de potencial previsibilidade para flutuações climáticas em diferentes escalas temporais (TRENBERTH e HURRELL, 1994; DESER et al., 2010).

Nesse sentido, a atmosfera da Terra, oceanos, criosfera e hidrologia continental estão em constante transferência de massa, momento e energia. Consequentemente, parâmetros climáticos globais ou regionais – como temperatura atmosférica, TSM, pressão em superfície, precipitação, velocidade do vento, entre outros – flutuam com aproximada regularidade, de modo quase-periódico.

Os mais conhecidos modos de variabilidade se estendem por grandes áreas do globo, incluindo o El Niño-Oscilação Sul (ENOS), Oscilação do Atlântico Norte (NAO – acrônimo para o termo original em inglês *North Atlantic Oscillation*), Oscilação Decadal do Pacífico (PDO – acrônimo para o termo original em inglês *Pacific Decadal Oscillation*), Oscilação de Madden-Julian (MJO – acrônimo para o termo original em inglês *Madden-Julian Oscillation*), Oscilação Multidecadal do Atlântico (AMO – acrônimo para o termo original em inglês *Atlantic Multidecadal Oscillation*). Outros atuam mais regionalmente, como é o caso dos Modos Anulares dos polos Norte e Sul. O estado destas flutuações ou modos de variabilidade é monitorado pela utilização de índices climáticos escalares (VIRON, DICKEY e GHIL, 2013).

O clima da América do Sul e, por conseguinte, o clima do Brasil, é fortemente marcado pela atuação de alguns modos de variabilidade globais. É o caso do ENOS (este sendo o mais influente de todos os modos de variabilidade), da PDO e da AMO. Além desses, alguns outros atuam em menor escala espacial e que, não obstante, podem delinear a variabilidade climática do continente sul-americano, como o Modo Meridional do Atlântico (AMM – acrônimo para o termo original em inglês *Atlantic Meridional Mode*) e o Modo Anular Sul (SAM – acrônimo para o termo original em inglês *Southern Annular Mode*) (ZHOU e LAU, 2001; GRIMM, 2009; KAYANO e ANDREOLI, 2009b; LOIKITH et al., 2017).

De particular interesse para este trabalho, estes modos de variabilidade com influência sobre a variação climática em escalas interanual, decadal e multidecadal e que impactam na disponibilidade de recursos hídricos e eólicos nas regiões SE e NE do Brasil são aprofundados nas seções subsequentes.

2.2.1 El Niño-Oscilação Sul (ENOS)

Uma das mais notáveis fontes de variabilidade interanual em tempo e clima globais é o fenômeno ENOS. Isolado, o termo El Niño era inicialmente utilizado para indicar um aquecimento anormal das águas que banham a costa do Peru e do Equador. Posteriormente foi associado a um aquecimento oceânico anômalo de maiores proporções que apresentava ligações com outros padrões anômalos climáticos globais. A componente atmosférica amarrada ao El Niño é a Oscilação Sul (OS). O padrão OS é bem estabelecido como um padrão gangorra (onda estacionária) centrado em latitudes tropicais e subtropicais, com centros de ação localizados sobre a Indonésia e o Pacífico Tropical Sul. Assim, apenas ao final da década de 60 que a OS foi conectada ao El Niño para descrever o fenômeno em que o Oceano Pacífico e a atmosfera colaboram conjuntamente, sendo chamado cientificamente de ENOS. O El Niño se refere à sua fase quente; a sua fase oposta (fria) é chamada de La Niña, consistindo em um resfriamento de grandes proporções das águas do Pacífico Tropical (TRENBERTH, 1997; TRENBERTH e CARON, 2000).

O ENOS pode ser descrito como uma perturbação climática ao redor do estado médio do sistema acoplado oceano-atmosfera sobre o Pacífico Tropical. Este estado médio (Figura 20a) consiste em um gradiente de TSM apontando de leste para oeste com sobreposição de ventos alísios na atmosfera tropical que sustentam um *feedback* dinâmico positivo: a diferença de temperatura ao longo do equador reforça a intensidade dos ventos alísios ao favorecer aquecimento ascendente de larga escala sobre o Pacífico Equatorial Oeste e resfriamento descendente sobre o Pacífico Equatorial Leste. Os ventos alísios que atuam na superfície oceânica causam a ascendência da termoclina e afloramento (ressurgência) de águas frias subsuperficiais à leste (Figura 20b). Os ventos alísios e afloramentos equatoriais associados mantêm esta distribuição climatológica na TSM do Pacífico Tropical. Devido ao *feedback* positivo, uma mudança modesta tanto na TSM equatorial quanto nos ventos alísios pode desencadear uma reação em cadeia no sistema acoplado oceano-atmosfera. Caso ocorra um enfraquecimento nos alísios equatoriais, a ressurgência equatorial decrescerá, provocando relaxamento da inclinação oeste-leste da termoclina e redução no contraste oeste-leste da TSM.

Como a circulação de Walker é mantida pelo gradiente de TSM, o enfraquecimento neste causará também o enfraquecimento dos alísios, por fim causando aquecimento no Pacífico Equatorial em sua porção leste. Este mecanismo de retroalimentação é o elemento vital para o desenvolvimento de episódios quentes e frios do ENOS (Figura 20c) (CHANG et al. 2006).





Fonte: National Oceanic and Atmospheric Administration - National Climatic Data Center (2018) O *feedback* negativo dominante envolve um ajuste dinâmico da termoclina aos campos de vento sobrepostos com um atraso que eventualmente culmina com o fim de um evento. O padrão espacial associado com a sua fase quente (El Niño) consiste em anomalias positivas de TSM ao longo do Pacífico Equatorial à leste e ao longo de toda a costa americana, rodeado por anomalias negativas mais fracas sobre o Pacífico Tropical (DESER et al., 2010).

Várias regiões do Pacífico Tropical são consideradas importantes para o monitoramento e identificação de um evento El Niño ou La Niña – são as chamadas regiões NINO (Figura 21).



Figura 21 – Regiões NINO

Fonte: Australian Government - Bureau of Meteorology (2018)

O índice mais comumente utilizado para representar a variabilidade de TSM associada ao ENOS é a média de TSM por área na região compreendida por 5N-5S e 170W-120W (retângulo – Figura 22a), referido como índice NINO 3.4. Acredita-se que este indicador possui representatividade de TSM equatoriais médias ao longo do Pacífico. A climatologia de longo prazo da área em questão deve ser subtraída dos valores de TSM médios por área observados, obtendo, assim, as anomalias. Tipicamente utiliza-se medias móveis de 5 meses normalizadas pelo desvio padrão do período da climatologia. Um evento El Niño ou La Niña são definidos quando o valor de NINO 3.4 excede +/- 0,4°C para um período de 6 meses consecutivos ou mais. A série temporal deste índice exibe uma irregularidade de eventos quentes e frios, com durações típicas de 1 a 1,5 anos, com frequência de ocorrência aproximadamente entre 3 e 8 anos. Os eventos em fase quente mais intensos ocorreram em 1982-1983 e 1997-1998, com amplitudes de 2°C-2,5°C, enquanto eventos moderados em sua fase quente apresentaram amplitudes entre 1°C e 2°C (Figura 22b). Eventos em sua fase fria (La Niña) tendem a ser menos intensos, porém com duração relativamente maior. Análise da densidade espectral da série temporal do índice NINO 3.4 revela um alcance dominante de períodos entre 2,5-8 anos, com sensibilidade ao período da amostra registrada (Figura 22c). O ENOS é sazonalmente dependente como indicado pelo ciclo sazonal do desvio padrão da anomalia da série temporal para NINO 3.4 (Figura 22d), que apresenta valores mínimos (~0,55°C) durante abril-junho e valores máximos (~0,95°C) durante novembro-janeiro (DESER et al., 2010).



Figura 22 – a) anomalias de TSM mensais retificadas para o período 1900-2008, b) série temporal para o índice Niño-3.4, c) análise espectral da série temporal e d) desvio padrão mensal do índice Niño-3.4

Fonte: Deser et al. (2010)

A maior fonte de variabilidade da precipitação no Brasil em 3 das 4 estações climáticas anuais (outono, inverno e primavera) é o fenômeno ENOS. No verão, ainda que outros fatores locais como a interação entre a superfície continental e atmosfera exerçam maior influência sobre a precipitação, a forçante remota ENOS apresenta grande contribuição (GRIMM, 2011). O ENOS interfere tanto no posicionamento da ZCIT quanto da ZCAS, logo possui atuação indireta sobre a disponibilidade de recursos hídricos e eólicos em regiões de interesse no Brasil (DRUMOND e AMBRIZZI, 2006; SCHNEIDER, BISCHOFF e HAUG, 2014; ROSSO et al., 2018). As variações na circulação de Walker provocam seca no NE do Brasil em eventos El Niño, ao passo que períodos anomalamente úmidos ocorrem associados a eventos La Niña (KAYANO et al., 2016). Temperatura é outro parâmetro climático que sofre influência do modo de variabilidade em questão, com extremos de temperatura associados ao fenômeno sobretudo nas regiões N, NE e SE do Brasil (LOIKITH et al., 2017).

Além disso, a variabilidade da força das teleconexões resultantes do ENOS relacionase com outros modos de variabilidade, o que possibilita a atuação construtiva (sinais iguais) ou destrutiva (sinais opostos), a depender da fase oscilatória de cada um deles. Isto significa que o monitoramento climático pode ser aprimorado caso inclua informações a respeito de modos de variabilidade distintos, visto que diferentes cenários climáticos resultariam da interação destes padrões oscilatórios (KAYANO e ANDREOLI, 2007)

Tais fatos justificam e reforçam o interesse sobre o ENOS como um potencial indicador da variabilidade interanual de recursos hídricos e eólicos, fazendo do índice que o valora uma informação relevante para associar com o histórico de vazões e ventos nas regiões de interesse para geração de energia no Brasil.

2.2.2 Oscilação Decadal do Pacífico (PDO)

Tal qual o fenômeno ENOS, a Oscilação Decadal do Pacífico (PDO) também se associa com anomalias mensais de TSM ao longo do Oceano Pacífico. Inicialmente, o corpo científico havia sugerido três principais características para distinguir a PDO do ENOS: primeiro, eventos PDO no século XX persistiram por 20 a 30 anos, enquanto típicos eventos ENOS persistiram por 6 a 18 meses; segundo, as principais evidências climáticas da PDO foram visualmente identificadas nos extratrópicos, sobretudo no setor Pacífico Norte/América do Norte, enquanto que evidências secundárias foram percebidas nos trópicos (fatores opostos aos de ENOS);

terceiro, os mecanismos causadores da variabilidade PDO eram desconhecidos, ao contrário das causas para a variabilidade de ENOS, que eram relativamente bem compreendidas (MANTUA e HARE, 2002).

Atualmente ficou claro que a PDO possui padrão espacial de anomalias de TSM diferente daquele para o ENOS: o gradiente de TSM na PDO é orientado na direção norte-sul, restrito à região do Pacífico Norte (entre 20N e 70N – Figura 27a). Ainda, a amplitude das anomalias é consideravelmente menor daquelas para o ENOS. Com respeito à variabilidade da PDO, diferentes mecanismos foram propostos como interferentes, incluindo o fluxo estocástico de calor de superfície, teleconexões dos trópicos ao Pacífico Norte, interações oceano-atmosfera, flutuações aleatórias e outras induzidas pelo próprio ENOS. A inércia térmica e dinâmica do oceano contribui para a característica temporal de baixa frequência da PDO, enquanto processos atmosféricos são responsáveis pela coerência espacial de larga escala da PDO. Uma implicação fundamental destas análises é que, diferentemente do ENOS, a PDO é tida não como um único modo físico, e sim como uma combinação de processos atmosféricos e oceânicos que varrem os trópicos e extratrópicos – o que faz da sua previsibilidade algo muito complexo (NEWMAN et al., 2003; DESER et al., 2010; NEWMAN et al., 2016).

O índice que afere a PDO é a componente principal da análise por funções ortogonais empíricas de anomalias médias mensais de TSM na região do Pacífico Norte (retângulo – Figura 27a) após remoção da anomalia média global de TSM. Em sua fase positiva, o padrão PDO consiste em anomalias negativas de TSM nas regiões oeste e central do Pacífico Norte, e anomalias positivas de TSM em sua porção leste. O oposto é válido para a sua fase negativa. Anomalias de TSM em outras regiões oceânicas são associadas ao padrão da PDO, sobretudo na região leste do Pacífico Tropical (Figura 23a) (MANTUA et al., 1997; MANTUA e HARE, 2002; DESER et al., 2010).

A série temporal para a PDO (Figura 23b) exibe considerável variabilidade decadal, por isso seu nome. Mesmo anomalias mensais predominantemente apresentam períodos em que um sinal tem duração de duas décadas ou mais (a exemplo dos períodos de 1908-945, 1947-1976 e 1977-1998). As transições decadais de TSM, também identificadas como a inversão da polaridade da PDO, são acompanhadas por mudanças abrangentes na atmosfera, oceano e ecossistemas marinhos (MANTUA et al., 1997; DESER et al., 2010).

No que se refere ao clima em superfície continental, a fase quente da PDO (anomalias positivas de TSM) é correlacionada com períodos de seca no leste australiano, Japão, extremo

leste da Rússia e em boa parte dos Estados Unidos, porém tende a coincidir, também, com períodos de anomalias positivas de precipitação no oeste australiano, México e até mesmo o SE brasileiro. Outro ponto interessante é que a PDO atua conjuntamente com o ENOS e, estando em fases iguais, padrões de teleconexões ENOS se fortalecem, afetando a variabilidade de precipitação sobre a América do Sul (MANTUA e HARE, 2002; KAYANO e ANDREOLI, 2007).

Figura 23 – a) anomalias de TSM mensais para o período 1900-2008, b) série temporal para o índice de PDO, com curva suavizada por médias móveis quinquenais



Fonte: Deser et al. (2010)

2.2.3 Oscilação Multidecadal do Atlântico (AMO)

As TSMs no Atlântico Norte passam por lentas variações com periodicidade da ordem de 65 a 80 anos. Este fenômeno é denominado Oscilação Multidecadal do Atlântico (AMO). Ele é considerado um modo de oscilação natural da circulação termoalina do Oceano Atlântico, sendo intrínseco ao oceano e estocasticamente forçado por fluxos atmosféricos dinâmicos. A amplitude deste modo de variabilidade climática é aumentada em função de interações do acoplamento oceano-atmosfera (DELWORTH e MANN, 2000; ENFIELD, MESTAS-NUÑEZ e TRIMBLE, 2001; DESER et al., 2010).

O índice para AMO pode ser definido como a média de anomalias de TSM por área da região do Atlântico Norte (0-70N), com remoção da anomalia média global de TSM. Fases quentes da AMO ocorreram aproximadamente entre 1920-1970, depois novamente a partir da década de 1990; fases frias ocorreram no período de 1900-1920 e depois de 1970-1990 (Figura 24a). O padrão espacial associado a essa série temporal, obtido pela regressão linear das anomalias de TSM em cada localidade no índice de AMO, exibe valores positivos sobre todo o Atlântico Norte, como maiores magnitudes (0,5°C) ao sul da Groenlândia (DESER et al., 2010).

A AMO é associada com alterações de precipitação em grande escala, sobretudo sobre o Sahel, sudeste dos Estados Unidos e Brasil, além de ser correlacionada com a frequência de furacões intensos no Atlântico Norte (ENFIELD, MESTAS-NUÑEZ e TRIMBLE, 2001; TING et al., 2009). No Brasil, a região NE é sensível à fase da AMO, que pode atuar conjuntamente com outros modos de variabilidade do Oceano Atlântico (vide seção 2.2.4) e com o ENOS: um período seco é passível de ocorrer quando a AMO está em sua fase quente combinada com eventos El Niño e/ou dipolo positivo de TSM no Atlântico Tropical da mesma maneira que um período anomalamente úmido pode ocorrer se a AMO em sua fase fria for combinada com eventos La Niña e/ou dipolo negativo de TSM no Atlântico Tropical (KAYANO et al., 2016).

O fato da possibilidade de que efeitos antropogênicos possam estar contribuindo para a fase recente positiva da AMO – em função do aumento da concentração de GEE na atmosfera – tem provocado debates consideráveis pela comunidade científica. Como o índice da AMO é baseado em registros de temperatura, é provável que ele integre o sinal antropogênico do aquecimento global e, por conseguinte, confunda o real estado da AMO. Ao remover este sinal desta componente de variabilidade é possível estimar o componente natural da AMO, como mostra a curva verde (Figura 24b), que indica que a AMO está em uma fase quente modesta, e

não em uma fase quente extrema como seria inferido puramente pelo índice de TSM (TING et al., 2009).



Figura 24 – a) anomalias de TSM mensais para o período 1870-2008, b) série temporal para o índice de AMO, com curva estimativa para a componente natural do índice filtrado

Fonte: Ting et al. (2009) e Deser et al. (2010)

2.2.4 Modos de variabilidade do Atlântico Tropical

No Oceano Atlântico Tropical existem duas principais formas de variabilidade no acoplamento oceano-atmosfera: o Modo Equatorial do Atlântico (AEM, acrônimo para o termo original em inglês *Atlantic Equatorial Mode*), também conhecido como Niño do Atlântico, e o Modo Meridional do Atlântico (AMM). Estes dois padrões distintos interferem na variabilidade interanual da migração anual da ZCIT e se manifestam nas estações do outono austral (março-

abril, quando a ZCIT alcança máxima latitude ao sul) e do inverno austral (junho-agosto, quando a ZCIT alcança máxima latitude ao norte) (SERVAIN et al., 1999; CHANG et al., 2006; SMIRNOV e VIMONT, 2011)

O AMM é o modo estatístico dominante de variabilidade interannual-decadal no acoplamento oceano-atmosfera sobre o Atlântico Tropical exceto para a estação do inverno austral, quando o AEM se torna ligeiramente mais proeminente, portanto, ênfase é dada ao primeiro modo. O AMM é caracterizado por um gradiente meridional de TSM, no qual ventos na camada limite sopram em direção à região de águas anomalamente quentes e desviam para a esquerda (direita) no HS (HN) conforme a força de Coriolis, provocando um deslocamento meridional da ZCIT em direção ao hemisfério mais quente (CHIANG e VIMONT, 2004; VIMONT e KOSSIN, 2007).

No AMM, durante o outono austral TSM mais quentes aparecem nos trópicos e máxima precipitação sazonal se desloca para o Atlântico Equatorial à oeste e em direção ao território tropical da América do Sul. É no decorrer desta estação que a anomalia de chuvas é caracterizada por um padrão dipolo ao longo do Equador. Um gradiente meridional anômalo para TSM e ventos correlacionam-se com esta anomalia de precipitação. Estas correlações refletem uma situação dinamicamente consistente, onde um gradiente sul-norte de TSM mais forte que o normal orienta os ventos em direção ao norte, com ventos alísios mais fracos que o normal ao norte e mais fortes que o normal ao sul da ZCIT. Precipitação e TSM ficam abaixo do normal ao sul da ZCIT e acima do normal ao norte. Este padrão de circulação implica no enfraquecimento e deslocamento da ZCIT rumo ao hemisfério mais quente (Figura 25a) (CHANG et al., 2006).

Durante o inverno austral, um pronunciamento frio de TSM se desenvolve sobre o Oceano Atlântico Equatorial em sua porção leste, enquanto a ZCIT se desloca para seu ponto mais latitudinalmente ao norte, estendendo sobre território adjacente do oeste do continente africano. Nesta estação, um padrão anômalo de precipitação alcança o seu máximo ao longo da costa norte do Golfo da Guiné. O padrão anômalo correspondente de TSM é máximo à leste da bacia com padrão convergente de ventos alísios equatoriais (como ilustra a Figura 25b) (CHANG et al., 2006).


Figura 25 – Anomalias de TSM, precipitação e vento em superfície para a) AMM durante o outono austral e b) AEM durante o inverno austral



Numericamente, uma série temporal para o AMM é projetada a partir de dados de TSM e das componentes zonal e meridional do vento a 10m para a região delimitada em 21S-32N, 74W-15E, considerando o período de 1950-2005. Os dados são suavizados utilizando médias móveis de 3 meses e têm sazonalidade e tendência removidas. Um índice que representa a variabilidade ENOS é removido dos dados mediante regressão linear e a análise de máxima covariância é aplicada à matriz de covariância temporal entre os campos de TSM e de vento. Uma série temporal é então obtida, projetando anomalias de TSM ou de vento (Figura 26) (CHIANG e VIMONT, 2004).



Figura 26 – Médias mensais de AMM (campo de TSM) de 1948-2018

Fonte: National Oceanic and Atmospheric Administration - Earth System Research Laboratory's Physical Sciences Division (2018)

Um outro índice, conhecido por MGI, pode ser utilizado para avaliar o relacionamento do tipo dipolo na porção tropical do Atlântico. Ele deriva da diferença entre anomalias das médias mensais de TSM em uma área acima do Equador (5N-20N e 30W-60W) e anomalias das médias mensais de TSM de uma área abaixo do Equador (0S-20S e 10E- 30W), considerando a climatologia de 1981 a 2010. Este índice afere, portanto, o gradiente de TSM sobre as porções norte e sul do Oceano Atlântico tropical (ENFIELD et al., 1999).

Flutuações na precipitação do NE do Brasil são correlacionadas com anomalias de TSM e vento, no sentido em que, sazonalmente, a hidrologia é mais intensa nos meses de março a maio, quando a ZCIT está posicionada no ponto mais austral (XIE e CARTON, 2004). Nesse sentido, associar anos mais secos ou mais úmidos com valores representativos de AMM e da diferença entre TSM no Atlântico Tropical Norte e no Atlântico Tropical Sul parece ser interessante no que toca o entendimento da variabilidade da disponibilidade de recursos hídricos e eólicos na região supracitada.

2.3 BASES DE DADOS DE REANÁLISE

As bases de dados de reanálise constituem um esforço global para rodar modelos de última geração com dados históricos meteorológicos (GIEBEL, 2000). É pertinente indicar que duas amplas fontes de informação sobre o Sistema Terra são possíveis: umas provenientes de medições diretas (observações) e outras que englobam o entendimento da evolução espaço-temporal, tipicamente incorporadas em modelos matemáticos que descrevem a relação entre muitas variáveis e parâmetros em conhecidos processos físicos que remetem à circulação e interações gerais (LAHOZ e SCHNEIDER, 2014). A Figura 27 traz um esquema de como a Terra é segmentada horizontalmente (em dimensões de latitude e longitude) e verticalmente (dimensão da altitude), formando colunas de volumes em formato cúbico empilhados que interações.

Bases de dados de reanálise assimilam observações passadas de diferentes sistemas observacionais de medição com um modelo de previsão numérica meteorológica para proporcionar uma imagem instantânea fisicamente consistente das condições da evolução do tempo e clima terrestres em intervalos regulares no curso de longos períodos temporais (PARKER, 2016).





Fonte: Edwards (2011)

Os métodos de observação do Sistema Terra a partir de instrumentação incluem observações *in situ* de estações de superfície, boias e aeronaves e observações de satélites em órbitas baixas e altas (satélites geoestacionários). Coletivamente, estas plataformas observacionais constituem um sistema de observação global. As plataformas satelitais apresentam cobertura substancialmente global, embora em baixa resolução espacial e temporal. Já as plataformas *in situ* de sensoriamento remoto possuem alta resolução espaço-temporal, porém não têm abrangência global, sendo comumente utilizadas para avaliar e calibrar os dados de satélites (LAHOZ e SCHNEIDER, 2014).

As diferenças entre observações e dados de reanálise vão além da discutível origem das informações, dado que ambas as fontes de informação envolvem inferências com cálculos baseados em teorias físicas. Distintamente das observações, dados de reanálise são sustentados por previsões. Não obstante, tal distinção não é necessariamente significante. O que importa de fato é se os resultados das previsões são suficientemente precisos. A incerteza dos resultados modelados é, assim, menos compreendida do que aquela para as observações. Ainda assim, a solução pode ser satisfatória se for acurada (PARKER, 2016).

Determinados desenvolvedores de modelos de previsão do clima disponibilizam bancos de dados climáticos compreendendo informações que exploram o passado climático terrestre (Tabela 2). Cada um destes conjuntos de dados de reanálise possui suas características específicas, a exemplo das resoluções temporal e espacial, além do período total englobado pelas séries temporais. Algumas destas bases de dados de reanálise são geradas pelo National Centres for Environmental Prediction (NCEP) e pelo National Centre for Atmospheric Research (NCAR) – ambos nos Estados Unidos – como é o caso da NCEP/NCAR R1, NCEP-DOE R2 e também da Climate Forecast System Reanalysis (CFSR). A base de dados Twentieth Century Reanalysis Project (C20r) é gerada pelo National Oceanic and Atmospheric Administration (NOAA), também nos EUA, e compreende mais de 100 anos de dados. Os projetos de dados de reanálise chamados ERA-20C e ERA-Interim são desenvolvidos pelo European Centre for Medium-Range Weather Forecasts (ECMWF) - uma organização intergovernamental apoiada por 34 nações europeias. A National Aeronautics and Space Administration (NASA) desenvolve nos EUA sua própria base de dados: Modern Era Retrospective- analysis for Research and Applications, que está em sua segunda versão (MERRA-2) (DEE et al., 2016).

Reanálises	Resolução espacial	Resolução temporal	Período
NCEP/NCAR R1	2,5° x 2,5°	6 horas	1948 - presente
NCEP-DOE R2	2,5° x 2,5°	6 horas	1979 - presente
CFSR (NCEP)	0,5° x 0,5°	6 horas	1979 - 2011
C20r V2 (NOAA)	2,0° x 2,0°	6 horas	1850 - 2014
ERA-20C	0,5° x 0,5°	3 e 6 horas	1900 - 2010
ERA-Interim	0,75° x 0,75°	6 horas	1989 - presente
MERRA-2 (NASA)	0,5° x 0,625°	1 hora	1980 - presente

Tabela 2 – Algumas bases de dados de reanálise disponíveis

Fonte: adaptado de Dee et al. (2016)

Os produtos destes estudos são registros confiáveis e adequados que permitem investigar a climatologia, calcular anomalias, conduzir estudos empíricos e diagnósticos quantitativos, explorar o entendimento aprimorado do processo climático e realizar validações. São, portanto, essenciais para estudar fenômenos sazonais a multianuais (TRENBERTH, 2010).

Contudo, apesar das vantagens de cobrir longos períodos temporais em locais que observações históricas não são disponíveis (como é o caso de vários países do HS), dados de reanálise para velocidade de vento podem apresentar níveis significativos de incerteza e tendência quando comparados com observações diretas (ROSE e APT, 2016). Embora os resultados destes modelos que reconstroem mais de um século de informação climática sejam

compatíveis com muitos dados observacionais, uma consistência estatística rigorosa ainda não foi alcançada, sugerindo que a incerteza inerente aos dados é, de certa forma, subestimada (COMPO et al., 2013).

É importante ressaltar que tais modelos são alterados na medida em que deficiências são identificadas ou quando novos grupos de informação são assimilados ou mesmo quando sensores de medição são substituídos. O uso de tais dados requer cuidado e o acompanhamento dos relatórios que informam sobre tais mudanças é vital para que alterações não se confundam com possíveis falsos padrões a serem identificados. Consequentemente, a comparação de conjuntos de dados entre diferentes projetos de reanálise pode ser útil já que, ao menos à primeira vista, a divergência entre os resultados indica um limite inferior de incerteza (PARKER, 2016). Nesse sentido, é relevante que as análises de séries temporais de dados de vento sejam realizadas com diferentes projetos de reanálise, comparando os resultados e verificando e/ou corrigindo tendências sempre que possível.

Todos estes conjuntos de dados são disponibilizados por cada órgão administrador de seus respectivos projetos. O acesso a essas fontes se dá por *download* direto nos portais especializados do NCAR, NOAA, NASA e ECMWF. Comumente, estes mesmos portais, em especial o NCAR e NOAA, disponibilizam dados também para os modos de variabilidade de interesse desta pesquisa (ENOS, PDO, AMO, AMM e MGI). Como visto nas seções 2.2.1-2.2.4, cada qual possui o seu devido índice representativo, com séries temporais variando no tempo total de extensão, mas usualmente contemplando médias mensais.

2.4 ANÁLISE DE SÉRIES TEMPORAIS

2.4.1 Conceitos gerais de séries temporais

Uma série temporal é definida como qualquer conjunto de observações realizadas e ordenadas ao longo do tempo. As observações podem ser realizadas continuamente, como em um sismógrafo onde a agulha não para de registrar (Figura 28a), caracterizando a série temporal como contínua, ou em intervalos de tempo específicos, normalmente espaçados equivalentemente, caracterizando-a como discreta, como em um gráfico de índice de preços anuais de ativos financeiros (Figura 28b) (CHATFIELD e XING, 2019).

Figura 28 – Exemplos de séries temporais contínua (a) e discreta (b)



Fonte: Chatfield e Xing (2019)

Por sua vez, uma série temporal discreta pode ser derivada de uma contínua, caso a leitura ou registro de informações se dê em intervalos periódicos, tornando-a amostral. Similarmente, uma série temporal pode ter seus valores (contínuos ou não) agregados em intervalos de tempo equivalentes, na forma de médias mensais ou anuais por exemplo, tornando-a também discreta (CHATFIELD e XING, 2019).

Muitos objetivos circundam a análise de séries temporais, dentre os quais destacam-se:

- a) Descrição: ao plotar as observações num gráfico em função do tempo, as principais propriedades da série são evidenciadas, como máximos (picos), mínimos (vales), tendências, ciclos sazonais e outras periodicidades, observações aberrantes não consistentes com o restante dos dados (também conhecidas como *outliers*);
- b) Explicação: quando duas ou mais variáveis são plotadas, é possível utilizar a variação no tempo de uma delas para explicar a variação de outra(s);
- c) Previsão: valores futuros podem ser previstos para uma dada série temporal;
- d) Controle: um sistema físico pode ter suas observações monitoradas para que se aprimore o controle sobre ele (ou sobre as implicações das observações).

Uma série temporal univariada S_T – referente a uma única variável de interesse – pode ser definida como um vetor de ordem $n \times 1$ onde cada observação x se dá num instante t de tempo e num determinado espaço multidimensional (com coordenadas de latitude, longitude e altura, por exemplo). Evidentemente que num mesmo intervalo de tempo outras variáveis podem ser registradas e contidas na mesma série temporal, tornando-a multivariada, como a temperatura e pressão de uma localidade pontual no oceano (MORETTIN e TOLOI, 2018).

$$S_T = \{x_{t=1}, x_{t=2}, x_{t=3}, \dots, x_{t=n}\}$$
(1)

A partir de uma dada sequência de valores, é adequado realizar a análise exploratória de dados, de maneira a compreender a natureza do processo descrito pela série. Além da média \bar{x} dos valores x_t , os quantis ou percentis q_p refletem os valores esperados para exceder, com uma probabilidade p, um valor qualquer pertencente ao conjunto de dados. Alguns percentis são frequentemente utilizados, a exemplo dos percentis de 25%, de 50% e de 75%. O de 50% ($q_{0,5}$) trata-se da mediana, que ocupa a posição central, localizada exatamente na metade do conjunto n de observações ordenadas no tempo. Se n for ímpar, a mediana será o valor central, do contrário, convencionalmente se extrai a média dos dois valores centrais. Já os percentis $q_{0,25}$ e $q_{0,75}$, também conhecidos como quartil inferior e quartil superior, localizam-se nos pontos equidistantes dos extremos da série à mediana e são calculados de maneira similar ao $q_{0,5}$. A diferença entre estes quartis, ou intervalo interquartil (IQR), é um indicativo de quão dispersas estão 50% das observações centrais. É propício comparar o IQR com o desvio padrão amostral σ do conjunto de dados. Elevado ao quadrado, o desvio padrão amostral retorna a variância amostral, medida que também ilustra quão espalhados uns dos outros estão os valores registrados na série temporal (WILKS, 2019).

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t \tag{2}$$

$$q_{0,5} = \begin{cases} x_{(n+1)/2}, & \text{se n impar} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2}, & \text{se n par} \end{cases}$$
(3)

$$IQR = q_{0,75} - q_{0,25} \tag{4}$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n} (x_t - \bar{x})^2}$$
(5)

$$var(S_T) = \sigma^2 = \frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x})^2$$
 (6)

Uma maneira visual de informar rapidamente boa parte das informações descritivas acerca de uma série temporal é por meio de diagramas denominados *boxplots* e que são capazes de informar os mínimos e máximos, os quartis inferiores, superiores e as medianas. Em sua

versão completa (Figura 29), o *boxplot* reflete até mesmo o grau de excepcionalidade ao indicar a ocorrência de *outliers* e valores extremos em função da magnitude de *IQR* (WILKS, 2019).



Figura 29 – Elementos de um boxplot



Outro diagrama de interesse a ser elaborado é o histograma (Figura 30a) – este é capaz de representar a distribuição de frequência de um conjunto de dados. O alcance total (amplitude) dos valores é segmentado em intervalos de classe de largura equivalente e o número de valores que se enquadra em cada intervalo é contabilizado. Quanto mais observações contabilizadas em um mesmo intervalo, maior será a barra que cresce verticalmente, indicando alta frequência de ocorrência – que pode ser descrita pelo eixo vertical em numeração de contagem ou proporção do número total de observações. É válido apontar para a necessidade de se determinar um número apropriado para a largura dos intervalos de classe, dado que se forem muito largos, podem resultar num histograma muito suavizado, mascarando informações de relevância; se forem muito estreitos, a plotagem pode tornar-se irregular e de difícil interpretação. Uma boa largura (*bin*) deve levar em consideração o número *n* de observações, o intervalo interquartil *IQR* e uma constante *c* que pode assumir, não estritamente, um valor entre 2 e 2,6. Esta mesma distribuição empírica de frequência pode ser visualizada pela curva densidade suavizada de *kernel* não paramétrica (WILKS, 2019).

$$bin \approx \frac{c * IQR}{n^{1/3}} \tag{7}$$

Relacionada ao histograma, há também a função de distribuição acumulada (CDF – do inglês *cumulative distribution function* – Figura 30b). Trata-se de um gráfico cujo eixo vertical apresenta as estimativas de probabilidade cumulativa associadas aos valores no eixo horizontal – forma prática de visualizar a frequência estimada para a probabilidade de que uma nova observação arbitrária que venha a ocorrer no futuro não exceda o valor correspondente ao valor

no eixo horizontal. A construção da função de distribuição acumulada deve estimar as probabilidades de ocorrência de cada valor *x* classificado em ordem estatística (WILKS, 2019).





Fonte: adaptado de Wilks (2019)

As informações que o histograma ilustra de maneira empírica podem ser descritas por meio de funções matemáticas específicas que também representam distribuições e variações dos dados de uma série temporal, porém idealizada por meio de parâmetros. Tais distribuições paramétricas são abstrações que buscam se aproximar das informações originais. Os parâmetros da distribuição são características que representam propriedades de uma população de dados, diferentemente do cálculo estatístico direto de uma amostra desta população (WILKS, 2019).

As distribuições podem ser discretas quando os dados atribuem valores predeterminados, finitos – como o resultado de um lançamento de um dado. Distribuições discretas mais comuns são a binomial, geométrica e *Poisson* (Figura 31). Em contrapartida, as distribuições podem ser contínuas, quando a variável pode atribuir qualquer valor dentro de uma gama de possíveis resultados, a exemplo de todos os números reais não negativos ou até mesmo toda a linha de números reais. Trata-se do caso de variáveis atmosféricas como temperatura, pressão, velocidade do vento etc. Embora a natureza do sistema de registro destas variáveis possa se apropriar de arredondamentos para valores discretos, o conjunto de possíveis dados é grande o suficiente que permitem o tratamento como quantidades contínuas. Este tipo de distribuição abrange curvas bem conhecidas como a Gaussiana (Normal), log-normal, gama, exponencial e *Weibull* (Figura 31) (WILKS, 2019).



Figura 31 – Distribuições probabilísticas mais comuns

Fonte: Owen (2018)

O cálculo de probabilidades de ocorrência de uma variável aleatória contínua envolve a integração de funções contínuas chamadas funções densidade de probabilidade (PDFs – do inglês *probability density functions* – Figura 32) ou simplesmente referenciadas pelo termo densidade. A PDF para uma variável aleatória x é denotado por f(x) e a soma de todos os possíveis valores da quantidade aleatória deve se equivaler a 1 – condição que deve ser satisfeita por qualquer PDF. A probabilidade é proporcional à área abaixo da curva, devendo se estabelecer o raciocínio em torno de uma faixa de valores possíveis em detrimento da probabilidade de ocorrência de um valor exato. É similar ao conceito da CDF, que busca especificar a probabilidade que uma quantidade x qualquer não excederá um outro valor qualquer. No caso da PDF, estipula-se um intervalo de valores possíveis (inferior e superior) dentro do qual existe uma probabilidade da quantidade x ocorrer, sendo calculada com base na área da curva da densidade, como ilustra a Figura 35 (WILKS, 2019).

$$\int_{x} f(x) \, dx = 1 \tag{8}$$

$$\Pr[x_{inferior} \le x \le x_{superior}] = \int_{x_{inferior}}^{x_{superior}} f(x) \, dx \tag{9}$$

Figura 32 – PDF hipotética para uma variável aleatória e o cálculo, por integração, da probabilidade de ocorrência de um intervalo de valores específicos de *x*



Contrastar duas séries temporais pode dizer muito a respeito do grau de associação entre elas. Utiliza-se o coeficiente de correlação de Pearson (r) para quantificar este relacionamento entre as observações x_t de uma série com as observações y_t de outra série. Este coeficiente resulta entre [- 1, 1], medindo a força de associação linear entre as duas variáveis – valores positivos (negativos) de r indicam que o comportamento das séries temporais é similar (oposto) e se uma variável aumenta, outra tende a aumentar (diminuir) e valor de r igual a 0 indica total independência entre as séries (SHUMWAY e STOFFER, 2016; CHATFIELD e XING, 2019).

Esta mesma métrica pode ser utilizada num formato simplificado para comparar uma série temporal com ela mesma, desde que exista uma defasagem temporal (*lag*) entre os pares de pontos alinhados, do contrário o valor de *r* será 1. Trata-se do cálculo da autocorrelação temporal, que busca identificar se o processo descrito pela série temporal apresenta memória, ou seja, se existe persistência de certa similaridade por sucessivos intervalos de tempo. A função autocorrelação é determinada pelo cálculo do coeficiente r_{lag} para diferentes valores de *lag*. Plota-se os resultados num gráfico conhecido por correlograma para interpretar a natureza do fenômeno: se é uma série completamente aleatória (Figura 33a); se se alterna em torno da média (Figura 33b); se contém tendência (não estacionariedade) (Figura 33c) ou se apresenta sazonalidade (Figura 34d) (SHUMWAY e STOFFER, 2016; CHATFIELD e XING, 2019).

$$COR\left(S_{1_{T}}, S_{2_{T}}\right) = r = \frac{Cov(S_{1_{T}}, S_{2_{T}})}{\sigma_{1} * \sigma_{2}} = \frac{\sum_{t=1}^{n} (x_{t} - \bar{x}) (y_{t} - \bar{y})}{\sqrt{\sum_{t=1}^{n} (x_{t} - \bar{x})^{2}} \sqrt{\sum_{t=1}^{n} (y_{t} - \bar{y})^{2}}}$$
(10)

$$COR\left(S_{1_{T}}, S_{1_{T}}\right) = r_{lag} = \frac{\sum_{t=1}^{n-lag} (x_{t} - \bar{x}) \left(x_{t+lag} - \bar{x}\right)}{\sum_{t=1}^{n} (x_{t} - \bar{x})^{2}}, \qquad lag = 0, 1, 2, \dots$$
(11)



Figura 33 – Correlogramas comuns: processo puramente aleatório (a); processo alternado (b); processo com tendência (c) e processo com sazonalidade (d)

Fonte: Chatfield e Xing (2019)

Se a série temporal descreve um processo puramente aleatório, isto significa que os valores sequenciados são independentes e identicamente distribuídos, possuindo média igual a zero e distribuição normal com variância constante. A série não apresenta autocorrelação, independentemente do *lag*. Um processo com essas características é referido como ruído branco. Entretanto, particularidade de muitas séries temporais é o fato de que as observações sucessivas não são mutuamente independentes, fazendo com que a ordem temporal com que se deram deva ser levada em consideração. Desta forma, valores futuros podem ser previstos de observações dependentes já ocorridas no passado (CHATFIELD e XING, 2019).

Boa parte das séries temporais são referidas como processos estocásticos, no sentido em que observações passadas determinam o futuro apenas parcialmente e os valores futuros possuem uma probabilidade de distribuição condicionada ao histórico conhecido de valores passados. Do contrário, se a série temporal for previsível com 100% de acurácia, ela é referida como um processo determinístico. Em termos gerais, a maioria dos processos físicos reais envolvem um elemento aleatório em sua estrutura, fazendo deles processos estocásticos que podem ser descritos como fenômenos estatísticos que evoluem no tempo de acordo com leis probabilísticas (MORETTIN e TOLOI, 2018; CHATFIELD e XING, 2019).

De maneira a possibilitar e validar algumas análises de séries temporais, propriedade comumente assumida para tais é a da estacionariedade, em que se assume que suas características estatísticas são razoavelmente preservadas ao longo do tempo. Isto significa que os valores de uma série se desenvolvem no tempo de maneira aleatória, porém estável – em torno de uma média constante (Figura 34a). Evidentemente que séries temporais costumeiramente apresentam algumas formas de não estacionariedade, a exemplo de processos

atmosféricos que exibem ciclos diurnos ou anuais. Por exemplo, a temperatura de um local fixo: espera-se que durante o dia os valores sejam maiores do que os valores observados durante a noite em função da incidência dos raios solares e da rotação da Terra, de mesma maneira que no verão os valores superam aqueles registrados no inverno tendo em vista o eixo inclinado da Terra em seu movimento de translação ao redor do sol que trazem consigo as diferentes estações do ano (MORETTIN e TOLOI, 2019; CHATFIELD e XING, 2019; WILKS, 2019).

Certos processos, a depender da fatia de tempo sendo analisada, podem apresentar formas de não estacionariedade, como mudança de regime (Figura 34b) ou, ainda, estacionariedade no longo prazo com períodos menores também estacionários (Figura 34c). Fato é que a maioria dos métodos de análise de séries temporais requerem que se assuma estacionariedade dos dados – o que para alguns processos naturais como os atmosféricos isto se verifica quando a série temporal é longa o suficiente para revelar a estacionariedade (KOUTSOYIANNIS, 2011; WILKS, 2019).

Figura 34 – Exemplo de uma mesma série temporal tratada como estacionária no curto prazo (a), não estacionária no curto prazo (b) e estacionária no longo prazo (c)



Fonte: Koutsoyiannis (2011)

Existem duas abordagens para lidar com séries não estacionárias, ambos visando processar os dados de um jeito que permita que a estacionariedade seja razoavelmente aceita. A primeira implica em transformar a série mediante subtração da média dos dados, fazendo com que os valores representem anomalias x'_t em torno de uma média constante. Também pode ser interessante normalizar estas anomalias x'_t ao dividi-las pelo desvio padrão σ , obtendo as variáveis normalizadas z_t , também conhecidas por *z*-scores. Transformações deste escopo possibilitam estabilizar a variância e fazer da distribuição um processo normal. A segunda abordagem é conduzida pela estratificação da série temporal, realizando a análise em menores trechos isolados, curtos o suficiente para serem considerados quase estacionários (CHATFIELD e XING, 2019; WILKS, 2019).

$$z_t = \frac{x_t - \bar{x}}{\sigma} = \frac{x'_t}{\sigma} \tag{12}$$

Variações típicas de séries temporais, como tendência, sazonalidade e correlação entre observações sucessivas, merecem atenção, sendo possível remover estes efeitos pela decomposição da série temporal para evidenciar os chamados resíduos, que podem ou não ser aleatórios. A remoção destes efeitos pode ser observada na Figura 35, onde a série original e seus dados observados são decompostos em 3 diferentes séries: tendência (*trend*), ciclo sazonal (*seasonal*) e resíduos (no caso, considerados aleatórios – *random*). Se estes três efeitos forem somados, a série original é reconstruída.





Fonte: Chatfield e Xing (2019)

A tendência pode ser definida como uma mudança de longo prazo no nível médio da série. Maneira de se remover tendência é subtrair, de cada observação x_t , os valores y_t determinados pela reta resultante do cálculo da regressão linear entre os valores da série temporal e os valores dos intervalos de tempo *t* correspondentes. A reta da regressão linear é aquela que passa pela nuvem de pontos, se aproximando destes pelo método dos mínimos quadrados (Figura 36) (SHUMWAY e STOFFER, 2016; CHATFIELD e XING, 2019).

$$y_t = a + b.t \tag{13}$$

$$b = \frac{n\sum_{t=1}^{n} t \cdot x_t - \sum_{t=1}^{n} t \sum_{t=1}^{n} x_t}{n\sum_{t=1}^{n} t^2 - (\sum_{t=1}^{n} t)^2}$$
(14)

$$a = \frac{1}{n} \sum_{t=1}^{n} x_t - \frac{b}{n} \sum_{t=1}^{n} t$$
(15)

Figura 36 – Exemplo de regressão linear



Outra maneira de lidar com tendências é a aplicação de filtros lineares, que convertem uma série temporal em outra por meio da multiplicação das observações por um conjunto de pesos w_r usualmente simétricos ou equivalentes. Tal técnica é capaz de suavizar flutuações e estimar a média local, revelando um valor de x_t que seja suavizado, Sm (x_t) . Esta operação é conhecida por médias móveis (exemplo na Figura 37 – curva azul representa a série filtrada) e a determinação do número de pesos pode ser arbitrariamente realizada, de maneira a calibrar a suavização da série temporal a ser filtrada (SHUMWAY e STOFFER, 2016).

$$m_t = \sum_{r=-q}^{+q} w_r x_{t+r}$$
, onde $\sum w_r = 1$ (16)

$$Sm(x_t) = \frac{1}{2q+1} \sum_{r=-q}^{+q} x_{t+r}$$
(17)



Quanto às variações sazonais, se referem aos ciclos ou padrões e flutuações recorrentes que ocorrem com periodicidade fixa, a exemplo dos ciclos diário ou anual. Comumente são padrões conhecidos, em especial para variáveis atmosféricas, e que não serão objetos de análise, fazendo sentido removê-los. Por exemplo, se for desejável compreender a variação diária da temperatura de maneira isolada, o ciclo anual pode atrapalhar, pois ele também é manifestado diariamente em estações diferentes – o ciclo anual acaba aparecendo em variações de menor ou maior frequência. Ao remover este efeito anual, o padrão diário fica enaltecido, facilitando a análise de um dia em contraste com outro. Alternativa para tal feito envolve o cálculo de médias para cada mês (ex: média de todos os meses de Janeiro, de todos os meses de Fevereiro etc.) para subtrair dos dados originais ou, ainda, utilizar estas médias para calcular harmônicos resultantes da transformada de Fourier para representar o ciclo anual e posteriormente removê-los dos dados originais, a exemplo da Figura 38 (CHATFIELD e XING, 2019; WILKS, 2019).

Figura 38 – Médias de temperatura de cada mês (a) e a identificação de uma função sinusoidal com amplitude regulada e defasagem para representar o ciclo anual (linha grossa em b)



2.4.2 Análise de Fourier

A análise de Fourier ou análise de harmônicos de uma série temporal consiste na decomposição desta em uma somatória de componentes senoidais cujos coeficientes são as suas transformadas discretas de Fourier e que a representarão no domínio da frequência (Figura 39). Trata-se de um procedimento de análise de dados que busca descrever e mensurar as variações de uma série temporal em diferentes escalas de tempo ao comparar tais flutuações com termos sinusoidais (funções de seno e cosseno). Estas funções trigonométricas são ditas harmônicas por possuírem frequências que exibem múltiplos inteiros da frequência fundamental determinada pelo tamanho da amostra que, no caso, será o comprimento da série temporal analisada (BLOOMFIELD, 2000; WILKS, 2019).





Fonte: Trekhleb (2018)

Qualquer série temporal S_T composta por um número par de *n* pontos pode ser representada por uma somatória da média de seus valores com funções harmônicas que dependem de uma frequência ω_k , amplitude C_k e ângulo de fase Φ_k , onde *k* é um número inteiro que indica o número do harmônico e a amplitude C_k relaciona-se com dois preditores A_k e B_k que podem ser determinados mediante regressão múltipla. À medida que se aumenta o número de harmônicos a serem somados para representar a série, melhor o resultado. O número máximo de funções harmônicas é igual a n/2, quando se atinge a representação exata da série temporal (BLOOMFIELD, 2000; WILKS, 2019).

$$S_T = \bar{S} + \sum_{k=1}^{n/2} \{ C_k \cos[\omega_k t - \Phi_k] \} = \bar{S} + \sum_{k=1}^{n/2} \{ A_k \cos[\omega_k t] + B_k \sin[\omega_k t] \}$$
(18)

$$\omega_k = \frac{2\pi k}{n} \tag{19}$$

$$\Phi_{k} = \begin{cases} \tan^{-1}(B_{k}/A_{k}) &, & A_{k} > 0\\ \tan^{-1}(B_{k}/A_{k}) \pm \pi, & A_{k} < 0\\ \pi/2 &, & A_{k} = 0 \end{cases}$$
(20)

$$A_k = \frac{2}{n} \sum_{t=1}^n S_t \cos[\omega_k t]$$
(21)

$$B_k = \frac{2}{n} \sum_{t=1}^n S_t \sin[\omega_k t]$$
(22)

$$C_k = [A_k^2 + B_k^2]^{1/2}$$
(23)

Tomando o segundo harmônico como exemplo (k = 2), ele será descrito pela função cosseno que completa exatos dois ciclos por completo ao longo dos n pontos da série temporal original. Ele terá sua própria amplitude C_2 e anglo de fase Φ_2 . O ângulo ω_2 executa um ciclo completo quando t vai de 1 até n/2 e depois executa um segundo ciclo completo entre t = n/2e t = n. Porém no primeiro harmônico, ω_1 varia por um único ciclo completo de 0 a 2 π à medida que t aumenta de 1 até n.

A frequência ω_1 é a mais baixa possível para cada série temporal, denominada frequência fundamental. No extremo oposto, quando k = n/2, a maior frequência $\omega_{n/2}$ assume o valor de π . É a chamada frequência *Nyquist*, onde um ciclo completo é executado em apenas dois intervalos de tempo e um número total de n/2 ciclos é executado ao longo de todo o comprimento da série. A frequência *Nyquist* não depende de *n*, mas sim da frequência de amostragem – resolução temporal com que os dados da série original foram registrados – impondo limitação no tocante às informações disponíveis para a análise espectral (BLOOMFIELD, 2000; CHATFIELD e XING, 2019; WILKS, 2019).

Evidentemente que o fenômeno descrito pela série temporal pode manifestar variações que ocorrem em frequência mais alta do que a própria frequência *Nyquist*. Trata-se do risco de *aliasing* (Figura 40), onde nem toda informação acerca do fenômeno estudado fora captada e registrada na série temporal. Por conseguinte, qualquer variação atribuível a frequências maiores que a *Nyquist* será distribuída por entre as outras n/2 frequências representadas pelos harmônicos de Fourier, ocultando a verdadeira frequência em que ocorreu (BLOOMFIELD, 2000; CHATFIELD e XING, 2019; WILKS, 2019).

Figura 40 – Exemplo de *aliasing*, onde um fenômeno (linha pontilhada) que apresenta variações em frequências maiores do que a frequência com que os dados (pontos vazios) foram registrados (linha sólida)



A partir da análise de Fourier, percebe-se que a determinação de um harmônico depende apenas do valor de k, mas não de k + 1, k - 1 nem de nenhum outro índice. Isto implica que os coeficientes A_k e B_k de cada harmônico podem ser computados de maneira independente de qualquer outro harmônico. Logo, propriedade das funções harmônicas é que, se possuírem frequências que exibem múltiplos inteiros da frequência fundamental, elas não são correlacionadas – consequência da ortogonalidade entre as funções seno e cosseno – e que permite explorar cada harmônico de maneira isolada (BLOOMFIELD, 2000; WILKS, 2019).

O cálculo dos harmônicos de uma série temporal é facilitado por um algoritmo denominado *Fast Fourier Transform* (FFT), tendo em vista que alguns coeficientes A_k e B_k são previsivelmente equivalentes para diferentes valores de k, poupando alguns cálculos repetitivos e aumentado a eficiência da computação. O armazenamento dos resultados é, contudo, convenientemente feito em termos da notação exponencial complexa de Euler, fazendo com que algumas manipulações sejam menos laboriosas (WILKS, 2019).

$$e^{i\omega_k t} = \cos[\omega_k t] + i\sin[\omega_k t]$$
(24)

$$S_T = \bar{S} + \sum_{k=1}^{n/2} \{A_k + iB_k\} e^{i\omega_k t}$$
(25)

2.4.3 Análise Espectral

Analisar uma série temporal como uma coleção de coeficientes A_k e B_k que são funções da frequência ω_k ao invés de analisá-la como uma coleção de informações registradas em função do tempo pode ser interessante para sua interpretação. Vantagem direta desta perspectiva é a possibilidade de examinar as contribuições independentes resultantes de processos que variam com diferentes velocidades e que operam num espectro de diferentes frequências. Assim, a análise espectral propõe estimar uma função densidade espectral para examinar graficamente o espectro estimado de energia da série temporal que resulta da plotagem do quadrado das amplitudes C_k em função das frequências. Este estimador é conhecido por periodograma (Figura 41), plotado num gráfico em que o eixo das ordenadas é usualmente passado para escala logarítmica. É, portanto, permissível verificar a proporção da variação da série original que ocorre em cada frequência harmônica (WILKS, 2019; PERCIVAL e WALDEN, 2020).

Figura 41 – Exemplo de um periodograma com todas as frequências de 0 a 0,5 (à esquerda) e ampliado na região das frequências de 0 a 0,03 (à direita)





Como o relacionamento entre os preditores de um harmônico e a série temporal não depende de outras funções harmônicas sendo calculadas para representar a série, a proporção de variação desta série associada a cada harmônico também é fixa, sendo que a força da relação entre um dado harmônico k e os dados da série temporal pode ser expressa integralmente em termos da amplitude C_k . Pelo teorema de Parseval, a variância total de uma série temporal é subdividida entre todos os seus n/2 harmônicos, cada qual representando uma variação de diferente escala temporal. Pelo periodograma, a área total abaixo da curva equivale à variância total naquela frequência que corresponde ao pico (CHATFIELD e XING, 2019; WILKS, 2019).

$$\sum_{t=1}^{n} (S_t - \bar{S})^2 = n \sum_{k=1}^{(n/2)-1} \left\{ \frac{C_k^2}{2} + n A_{n/2}^2 \right\}$$
(26)

$$I(\omega_k) = \frac{n}{4\pi} C_k^2 \tag{27}$$

Anteriormente ao cálculo do periodograma, é mister eliminar a tendência e a sazonalidade da série temporal, haja vista que uma tendência pode introduzir componentes de frequência extremamente baixa (um pico quando a frequência for zero, por exemplo) enquanto a variação sazonal pode produzir picos na frequência sazonal e em múltiplos inteiros da frequência sazonal (a exemplo do ciclo anual). Caso estes efeitos não sejam eliminados, a análise espectral será muito provavelmente dominada pelos mesmos, obscurecendo o periodograma (SHUMWAY e STOFFER, 2016; CHATFIELD e XING, 2019).

Após remoção de tendência e sazonalidade e ainda antes de proceder com a análise espectral, é apropriado realizar o procedimento de *tapering*, que consiste na multiplicação nos extremos da série temporal por uma função h_{taper} (como a janela cosseno), com o intuito de enaltecer os dados no centro da série em detrimento daqueles nas extremidades, evitando descontinuidades. A porcentagem das extremidades a ser modificada pelo *taper* é definida à *priori*, usualmente optando por 5% ou 10% (*taper* de 10% ou 20%) mas não mais que 25% (*taper* de 50%). Moderadamente, processo de *tapering* reduz possíveis "vazamentos" no espectro suavizado, quando variações em uma frequência acabam sendo transferidas para frequências vizinhas (Figura 42) (BLOOMFIELD, 2000; SHUMWAY e STOFFER, 2016).

$$h_{taper} = \begin{cases} 0.5 \left[1 - \cos\left(\frac{2\pi t}{taper * n}\right) \right], & 1 \le t \le (taper * n/2) \\ 0.5 \left[1 - \cos\left(\frac{2\pi (n - t)}{taper * n}\right) \right], & n - (taper * n/2) + 1 \le t \le n \end{cases}$$
(28)
$$0\% < taper \le 50\% \quad e \quad t = 1, 2, 3, ..., n$$

Figura 42 – Periodograma suavizado que apresenta vazamentos (*leakage* indicado pela linha tracejada) e após aplicação de *tapering* (linha sólida)



Fonte: Shumway e Stoffer (2016)

Uma vez realizada a análise espectral, seu periodograma resultante pode apresentar bastante ruído (Figura 41), sendo pertinente suavizá-lo para aprimorar a visualização de picos no espectro – a Figura 42 ilustra um periodograma já suavizado. Maneira conveniente para tal feito é utilizar um modelo de médias móveis ou janela de Daniell, onde um número comumente ímpar de *m* pontos é selecionado para calcular médias dos valores de $I(\omega_k)$. Todavia, o valor de *m* não deve ser alto, do contrário corre-se o risco de suavizar em excesso e reduzir os picos, provocando viés na análise. A prudência gira em função do número *n* de pontos da série: quanto maior o valor de *n*, permite-se aumentar o valor de *m* também (CHATFIELD e XING, 2019).

$$\hat{I}(\omega_k) = \frac{1}{m} \sum_k I(\omega_k)$$
(29)

Embora de fácil computação, o periodograma pode apresentar duas deficiências estatísticas. Primeiro, seu valor estimado não é garantido de estar em concordância razoável com a real distribuição espectral do fenômeno analisado. Quando isto acontece, o periodograma é tido como um estimador enviesado – sendo interessante aplicar *tapering* para modificar o periodograma e reduzir este viés. Segundo, para muitas aplicações o seu grau de variabilidade é inaceitável, por não ser considerado um estimador estatisticamente consistente (PERCIVAL e WALDEN, 2020).

Ao tomar amostras diferentes de uma mesma população de dados, assume-se que diferentes coeficientes de Fourier serão computados, resultando em espectros também distintos. Parte-se da premissa de que cada valor de C_k^2 seja um fiel estimador da densidade espectral, o que significa que calcular a média de seus valores correspondentes a amostras diferentes aproximaria ao real espectro da população equivalente. Entretanto, infelizmente a distribuição amostral para cada C_k^2 é bastante ampla – particularmente aproxima-se da distribuição quiquadrado χ^2 com alguns graus de liberdade. Sendo assim, as estimativas para o periodograma são positivamente assimétricas, o que faz com que valores individuais de C_k^2 representem precariamente o real espectro do fenômeno de interesse, vide Figura 43 (WILKS, 2019)

Ocorre que os níveis de confiança para as estimativas de espectros com base em amostras diferentes de uma mesma população de dados são também relativamente amplos, com o agravante de que aumentar o tamanho da amostra não aumenta a precisão da estimativa do espectro – já que faria aumentar o número de frequências analisadas. A distribuição χ^2 , por conseguinte, não depende de *n*. Em contrapartida, o número *dof* de graus de liberdade depende

do *taper* (porcentagem modificada pelo processo de *tapering*) bem como do número *m* da janela de Daniell (MADDEN e JULIAN, 1971; CHATFIELD e XING, 2019; WILKS, 2019).





Fonte: Wilks (2019)

Solução proposta é investigar se altos valores de C_k^2 (picos no periodograma) calculados para diferentes amostras foram resultantes de processos puramente aleatórios – num processo de ruído branco, onde todas as amplitudes espectrais são equivalentes – ou se refletem uma periodicidade real que fora parcialmente escondida. Por conseguinte, a hipótese nula de que um dado C_k^2 para a frequência f_k seja significantemente maior do que o espectro de ruído vermelho àquela frequência $S(f_k)$ será rejeitada para um nível de confiança α desde que a frequência f_k tenha sido definida a partir da amostra, ou seja, uma informação predeterminada sem relação de dependência com os dados utilizados para calcular C_k^2 . Para testar a hipótese nula, a função de densidade espectral $S(f_k)$ assume a forma de um processo autorregressivo de primeira ordem (lag = 1), onde nem todas as amplitudes espectrais são equivalentes (WILKS, 2019).

$$S(f_k) = \frac{\left(1 - r_{lag=1}^2\right) * \frac{2}{n} \sum_{k=1}^{n/2} \hat{I}(\omega_k)}{1 + r_{lag=1}^2 - 2r_{lag=1} * \cos\left(2\pi f_k\right)} , \qquad 0 \le f_k \le 1/2$$
(30)

$$f_k = \frac{k}{n} = \frac{\omega_k}{2\pi} \tag{31}$$

$$\frac{dof * C_k^2}{S(f_k)} \sim \chi_{dof}^2 \tag{32}$$

$$dof = 2m \left[1 - \frac{4 * taper}{\pi} \right] \tag{33}$$

$$\Pr\left[\frac{dof * C_k^2}{\chi_{dof}^2 \left(1 - \frac{\alpha}{2}\right)} < S(f_k) \le \frac{dof * C_k^2}{\chi_{dof}^2 \left(\frac{\alpha}{2}\right)}\right] = 1 - \alpha$$
(34)

$$C_k^2 \ge \frac{S(f_k)}{dof} \chi_{dof}^2 (1 - \alpha) \tag{35}$$

Por fim, uma vez calculados o periodograma suavizado pelo método da janela de Daniell, o processo de ruído vermelho e seus respectivos limites superior e inferior para um dado nível de confiança α , a interpretação deve levar em consideração a largura da banda (*bwd*). Grosso modo, a largura da banda é a largura da janela espectral – é delimitada pelo número de pontos *n* da amostra e diretamente proporcional ao tamanho *m* da janela de Daniell e à frequência *Nyquist*. Os picos da amplitude espectral observados no periodograma suavizado devem estar minimamente separados por esta largura da banda, do contrário é possível que estejam relacionados à mesma frequência (BLOOMFIELD, 2000; SHUMWAY e STOFFER, 2016; CHATFIELD e XING, 2019).

$$bwd = \frac{2m * f_{n/2}}{n} \tag{36}$$

2.4.4 Análise de Componentes Principais

A Análise de Componentes Principais (ACP) – também conhecida por Análise de Funções Ortogonais Empíricas – é um artifício matemático capaz de reduzir uma base de dados composta por muitas variáveis em uma nova base de dados composta por um menor número de variáveis. Estas novas variáveis serão combinações lineares das variáveis originais, não correlacionadas entre si e selecionadas para representar a maior fração possível da variabilidade contida na base de dados original. Esta situação se dá quando existem correlações substanciais entre variáveis dispostas em séries temporais, indicando que elas contêm informação redundante (WILKS, 2019).

Dados multivariados atmosféricos e de outros campos da geofísica geralmente exibem correlações entre as variáveis, fazendo com que os resultados de uma ACP representem suas variações de maneira mais compacta, com potencial para elucidar acerca de variações não apenas temporais, mas também espaciais. A ACP permite, portanto, interpretar os dados originais de uma maneira mais efetiva, tendo em vista o formato comprimido dos dados e em

função da matriz covariância – objeto estatístico que descreve rigorosamente as relações que conectam variáveis distintas (NAVARRA e SIMONCINI, 2010; WILKS, 2019).

Para uma matriz X de dimensões $n \times j$, ou tempo × espaço, contendo múltiplos vetores dispostos verticalmente e que representam um número j de séries temporais S_{i_T} de mesma dimensionalidade n, a ACP permite identificar uma outra matriz U composta por novos vetores u_i cujos elementos são combinações lineares dos elementos das séries S_{i_T} , carregando quase toda a informação relevante da coleção original contida em X. As novas variáveis ou componentes principais – elementos de u_i que vão responder sucessivamente pela maior quantidade da variabilidade conjunta dos elementos x_{i_t} das séries S_{i_T} – são definidas pelos autovetores da matriz covariância S_X entre as séries, projetando-as sobre os respectivos autovetores e_i . Cada um dos j autovetores contém um elemento pertencente a cada uma das séries temporais e suas variáveis. Similarmente, cada componente principal é composto, como em uma média ponderada, por frações dos valores x_{i_t} pertencentes à sua respectiva série temporal. Assim, se as séries originais possuírem n observações, as componentes principais (PCs) serão também séries temporais compostas de n valores u_{i_t} . As PCs podem ser analisadas com as mesmas ferramentas matemáticas destinadas às séries temporais, desde análise exploratória de dados até análise espectral (NAVARRA e SIMONCINI, 2010; WILKS, 2019).

$$X = \begin{bmatrix} S_{1_T} & S_{2_T} & \cdots & S_{j_T} \end{bmatrix} = \begin{bmatrix} x_{1_1} & x_{2_1} & \cdots & x_{j_1} \\ x_{1_2} & x_{2_2} & \cdots & x_{j_2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1_n} & x_{2_n} & \cdots & x_{j_n} \end{bmatrix}$$
(37)

$$S_{X} = X * [X]^{T} = \begin{bmatrix} \sigma_{1}^{2} & cov(S_{1_{T}}, S_{2_{T}}) & \cdots & cov(S_{1_{T}}, S_{j_{T}}) \\ cov(S_{2_{T}}, S_{1_{T}}) & \sigma_{2}^{2} & \cdots & cov(S_{2_{T}}, S_{j_{T}}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(S_{j_{T}}, S_{1_{T}}) & cov(S_{j_{T}}, S_{2_{T}}) & \cdots & \sigma_{j}^{2} \end{bmatrix}$$
(38)

$$S_X = U * [E]^2 * [U]^T$$
(39)

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_j \end{bmatrix} = \begin{bmatrix} u_{1_1} & u_{2_1} & \cdots & u_{j_1} \\ u_{1_2} & u_{2_2} & \cdots & u_{j_2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1_n} & u_{2_n} & \cdots & u_{j_n} \end{bmatrix}$$
(40)

$$E = \begin{bmatrix} e_1 & e_2 & \cdots & e_j \end{bmatrix}$$
(41)

$$u_i = [E]^T * X \tag{42}$$

$$u_{i_t} = \sum_{t=1}^{n} e_{t,i} x_{i_t}, \quad i = 1, 2, \dots, j$$
(43)

Geometricamente, o primeiro autovetor e_1 apontará para a direção em que os vetores de cada série temporal exibem maior variabilidade, associando-se, desta forma, ao maior autovalor λ_1 . O segundo autovetor e_2 estará associado ao segundo maior autovalor λ_2 e assim sucessivamente, de maneira que os subsequentes autovetores serão numerados de acordo com magnitudes posicionando-se as decrescentes de seus autovalores associados, perpendicularmente a todos os autovetores anteriores. Esta condição de ortogonalidade define, a partir destes autovetores e suas direções, um novo sistema de coordenadas pelo qual os dados serão analisados. A matriz ortogonal E que contém os autovetores define a rotação rígida para que este novo sistema de coordenadas seja orientado de maneira que cada eixo consecutivo esteja alinhado com a direção da máxima variabilidade conjunta dos dados em X, preservando a condição de ortogonalidade com o eixo anterior. A dependência do conjunto de dados implica no empirismo associado à determinação de funções ortogonais (WILKS, 2019).

Propriedade notável das PCs é que elas são séries temporais não são correlacionadas entre si, o que implica em covariâncias nulas entre os seus pares de elementos. Cada PC representa uma fração da variação total contida em X, sendo esta fração proporcional aos autovalores λ_i . Logo, o conjunto de todos as PCs representa a variância total inerente à X (WILKS, 2019). A Figura 44 representa o esquema da ACP.





Fonte: Bellemans et al. (2018)

A ACP pode também ser executada partindo de uma matriz correlação ao invés de uma matriz covariância. Enquanto na matriz covariância as variáveis originais de cada série temporal devem ser centradas (a média dos valores de uma série temporal é subtraída de cada valor individualmente), na matriz correlação as variáveis devem ser normalizadas (além da subtração da média, os valores individuais devem ser divididos pelo desvio padrão). A diferença entre estas duas alternativas reside na ênfase dada às variáveis responsáveis pela maior variância do conjunto de dados. Pela matriz covariância, os componentes principais serão determinados por estas variáveis de maior peso na variância, ao passo que pela matriz correlação as variáveis normalizadas terão pesos equivalentes, já que apresentam variância unitária. A escolha prévia de qual rota prosseguir com a ACP deve levar em consideração as unidades do sistema internacional que caracterizam as variáveis. Se as séries temporais são compostas de valores mensurados em unidades distintas, ou seja, representam variáveis diferentes, é sensato utilizar a matriz correlação, evitando, por conseguinte, resultados de variâncias baseados em magnitudes relativas arbitrárias (WILKS, 2019).

Caso a análise envolva séries temporais de mesma variável, mas há interesse no comportamento genérico de todo o conjunto de dados, novamente a matriz correlação pode ser a alternativa mais interessante. Do contrário, a matriz covariância permitirá identificação dos maiores centros de variância – o que também pode ser interessante. De qualquer forma, os padrões mais importantes permanecem apenas parcialmente alterados ao utilizar matriz covariância ou correlação, o que faz com que esta seleção não seja um problema de grande relevância (NAVARRA e SIMONCINI, 2010).

A maioria das aplicações de ACP para dados atmosféricos envolve análises de campos espaciais de um conjunto de séries de uma determinada variável (ex: temperatura, pressão, velocidade do vento etc.). O autovetor resultante de cada PC é composto de elementos correspondentes a cada ponto geográfico do campo espacial de séries temporais e que podem ser plotados num mapa que representa a área coberta pelos dados atmosféricos. Mapas como este indicam as localizações geográficas que mais contribuem para sua respectiva PC, indicando a distribuição espacial das anomalias simultâneas representada pela PC. Estas disposições espaciais dos autovetores são passíveis de interpretação como os modos não correlacionados de variabilidade dos campos das quais as PCs foram extraídas (WILKS, 2019).

O exame dos mapas de coeficientes das funções ortogonais empíricas deve ser feito com cautela para não interpretar estruturas espaciais em termos de propriedades dinâmicas ou cinemáticas do processo físico sendo descrito. Há casos em que este tipo de interpretação pode ser razoável para a primeira PC, cujo autovetor é orientado na direção da máxima variância do conjunta dos dados. Os autovetores subsequentes representarão uma menor proporção da variância total e, sendo obrigatoriamente ortogonais aos anteriores, apresentam correlação nula com eles independentemente da natureza do processo físico representado pelos dados. Todavia, não devem ser analisados como modos de variabilidade independentes uns dos outros. Embora a primeira PC possa representar um importante padrão de variabilidade do processo físico em questão, ela pode muito bem incluir aspectos de outros modos correlatos. Uma mesma PC pode ser resultado de influências de padrões distintos (WIKLE, ZAMMIT-MANGION e CRESSIE, 2019; WILKS, 2019).

É válido ressaltar que o número de PCs eleitas para análise (de plotagem de mapas e outras técnicas de análise de séries temporais) deve ser menor que o número de séries temporais do conjunto inicial de dados, do contrário a ACP não cumpre seu propósito de redução de dimensionalidade. Esta determinação pode ser pautada pela proporção desejada da variância total a ser explicada de maneira cumulativa (soma-se a variância de cada PC u_i) e dispor os resultados da variância explicada por cada PC num diagrama denominado *scree plot* pode ser útil, com objetivo de localizar o ponto em que há uma separação do declive mais acentuado de um outro trecho cuja inclinação é mais suave em decaimento para a direita. A Figura 45 ilustra um caso em que, somadas, as três primeiras PCs explicam cerca de 60% da variância total. Fundamentalmente, a decisão é tomada por critérios subjetivos, sendo desejável ao menos trabalhar com PCs que, somando, explicam entre 70% e 90% da variância total (WIKLE, ZAMMIT-MANGION e CRESSIE, 2019; WILKS, 2019).

$$var(u_i) = \frac{\lambda_i}{\sum_{i=1}^j \lambda_i} * 100\%$$
(44)





Fonte: adaptado de Wilks (2019)

2.5 CLUSTERIZAÇÃO DE SÉRIES TEMPORAIS

A clusterização é uma técnica de mineração de dados em que informações similares são agrupadas em grupos ou *clusters* homogêneos cujos objetos que os integram apresentam máxima similaridade entre si e mínima similaridade com objetos de outros *clusters*, tornandose em uma abordagem útil para a análise exploratória de dados que seja capaz de organizar e pré-processar informações de interesse (AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).

De mesma forma, séries temporais podem ser agrupadas por meio desta técnica de clusterização pois, embora possam ser constituídas de muitos dados, cada série temporal em sua integralidade é considerada como um único objeto. A classificação e clusterização destes objetos parte da identificação de séries temporais cujos elementos apresentem máxima similaridade entre si, alocando-as em diferentes conjuntos tal qual o grau de dissimilaridade entre estes *clusters* de séries temporais seja também maximizado (LIAO, 2005; WANG, SMITH e HYNDMAN, 2006; AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).

A utilidade deste processo reside na identificação de características estruturais comuns em um conjunto de séries temporais, permitindo a descoberta de padrões informacionais. A Figura 39 ilustra um exemplo da clusterização de séries temporais, onde é possível visualizar como alguns aspectos (formas, amplitudes, variações, ciclicidades, magnitudes das observações, tendências) das séries temporais influenciam no processo de clusterização – as séries temporais da Figura 46, previamente ordenadas de 1 a 14, foram sequenciadas de maneira que suas similaridades fossem enaltecidas, facilitando o agrupamento em *clusters* que poderiam conter pares ou trios de séries temporais.



Figura 46 - Exemplo de agrupamento ou clusterização de séries temporais

Fonte: Wang, Smith e Hyndman (2006)

Na clusterização de séries temporais, deve-se, primeiramente, considerar se os dados são discretos ou contínuos, se foram uniformemente registrados ou não, se são univariados ou multivariados e até mesmo se as séries a serem agrupadas possuem mesma dimensionalidade e comprimento. É desejável que os dados tenham sido uniformemente registrados, do contrário alguns tratamentos devem ser realizados com antecedência, convertendo os dados brutos contidos nas séries temporais para formatos adequados à clusterização, seja por discretização, seja mediante extração de vetores de menores dimensionalidades ou pela extração de parâmetros e modelagem das séries originais (LIAO, 2005).

A depender da natureza dos dados, do contexto e da problemática, a clusterização de séries temporais integrais pode ser realizada em função de suas formas (onde um algoritmo é aplicado diretamente aos dados brutos das séries temporais), suas características ou seus modelos representativos e, ainda, baseada tanto nas formas quanto nas características das séries temporais – como mostra Figura 47 (AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).





Fonte: adaptado de Aghabozorgi, Shirkhorshidi e Wah (2015)

Nas abordagens com base em característica e/ou modelos, é necessário reduzir as séries temporais a informações apropriadas antes de aplicar um algoritmo. As características globais das séries temporais abrangem tendência, sazonalidade, periodicidade, assimetria, curtose, entre outras (WANG, SMITH e HYNDMAN, 2006). Já os modelos usualmente envolvem a determinação de parâmetros de processos ARIMA ajustados às séries temporais ou

transformação discreta de Fourier, para posterior aplicação de um algoritmo de clusterização (MONTERO e VILAR, 2014; AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).

De maneira geral, a clusterização com base nas formas das séries temporais presta-se bem quando estas são relativamente curtas ou se a ocorrência dos padrões não se dá no mesmo instante de tempo – por exemplo se há defasagem entre a ocorrência de picos e/ou vales. No entanto, tais medidas são propensas a erros quando lidando com longas sequências numéricas que apresentam elevada quantidade de ruído ou dados anômalos. Para estas situações em que a dimensionalidade das séries temporais é elevada, medidas de dissimilaridade baseadas nas estruturas das séries temporais podem ser mais apropriadas (MONTERO e VILAR, 2014; AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015). Métodos baseados em modelagem transformam séries temporais em modelos paramétricos para cada série temporal, onde os parâmetros extraídos são analisados e agrupados em *clusters*, não raro apresentando, contudo, problemas de escalabilidade e desempenho reduzido quando os *clusters* estão próximos uns aos outros (AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).

2.5.1 Dissimilaridade entre séries temporais

Algoritmos de clusterização se apoiam em medidas de dissimilaridade entre séries temporais. Um primeiro ponto é decidir se a clusterização será governada por medidas de dissimilaridade baseadas na forma ou baseadas na estrutura. As primeiras visam comparar o perfil geométrico das séries (ou, alternativamente, representações delas projetadas para reduzir suas dimensionalidades). Assim, a dissimilaridade baseada na forma é principalmente dominada por comparações locais. Por outro lado, as segundas visam comparar estruturas dinâmicas que descrevem o desempenho global das séries (MONTERO e VILAR, 2014).

Para calcular a dissimilaridade com base no perfil geométrico de duas séries temporais genéricas S_{i_T} e S_{j_T} de mesma dimensionalidade *n*, a distância de Minkowski é uma forma generalizada para tal:

$$d_{L_p}(S_{i_T}, S_{j_T}) = \left(\sum_{t=1}^n (S_{i_t} - S_{j_t})^p\right)^{1/p}$$
(45)

Esta distância de Minkowski possui ordem p, sendo p um número inteiro positivo que, se igual a 2, indicará a distância Euclidiana; se igual a 1, indicará a distância Manhattan. Embora esta métrica seja sensível a transformações de sinal, a noção de distância baseia-se na proximidade dos valores observados em instantes correspondentes, de modo que as observações sejam tratadas como se fossem independentes, o que significa que a distância não varia se houver permutações ao longo do tempo (MONTERO e VILAR, 2014). A Figura 48a ilustra o cálculo desta distância de Minkowski para p = 2 (Euclidiana).

O cálculo da distância entre séries temporais pelo método *Dynamic Time Warping* (DTW) busca por um mapeamento *r* entre as séries para que uma distância específica entre observações de duas séries temporais acopladas (S_{i_t}, S_{j_t}) seja minimizada. Esta distância ignora a estrutura temporal dos valores, permitindo acoplar observações que ocorreram em instantes diferentes (MONTERO e VILAR, 2014).

$$|r| = \sum_{t=1,\dots,m} |S_{i_t} - S_{j_t}|$$
(46)

$$d_{DTW}\left(S_{i_T}, S_{j_T}\right) = \min_{r \in M} |r| \tag{47}$$

A diferença entre as distâncias Euclidiana e DTW pode ser observada na Figura 48, sendo que as distâncias DTW não são medidas em pares de pontos no mesmo instante. Já o cálculo da distância Euclidiana requer correspondência temporal entre pares de valores.





Fonte: Zhang et al. (2014)

Ao tratar séries temporais mais curtas, univariadas e com ocorrência de similaridade em suas formas sem defasagem no tempo, é comum o cálculo da distância Euclidiana entre os dados brutos de duas séries para cada instante. Caso a similaridade seja evidente nas formas das séries temporais independentemente do instante em que ela ocorre (apresentando, neste caso, defasagem no tempo), métodos elásticos como o DTW são mais apropriados (AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).

Outro critério para calcular dissimilaridade entre duas séries temporais S_{i_T} e S_{j_T} recorre ao coeficiente de correlação de Pearson. Este método é, diferentemente da distância Euclidiana ou distância DTW, baseado na estrutura de duas séries (MONTERO e VILAR, 2014):

$$COR\left(S_{1_{T}}, S_{2_{T}}\right) = \frac{\sum_{t=1}^{n} \left(S_{i_{t}} - \bar{S}_{i_{T}}\right) \left(S_{j_{t}} - \bar{S}_{j_{T}}\right)}{\sqrt{\sum_{t=1}^{n} \left(S_{i_{t}} - \bar{S}_{i_{T}}\right)^{2}} \sqrt{\sum_{t=1}^{n} \left(S_{j_{t}} - \bar{S}_{j_{T}}\right)^{2}}}$$
(48)

$$d_{COR}(S_{i_T}, S_{j_T}) = \sqrt{2\left(1 - COR(S_{i_T}, S_{j_T})\right)}$$
(49)

Existe, ainda, um método adaptativo para cálculo de dissimilaridade que considera tanto medidas de proximidade entre séries temporais com base em suas formas quanto com base em suas estruturas. A distância entre séries será determinada por uma medida de dissimilaridade com base nas formas (podendo ser distância Euclidiana ou DTW) e à correlação temporal de primeira ordem entre as séries (MONTERO e VILAR, 2014):

$$CORT\left(S_{i_{T}}, S_{j_{T}}\right) = \frac{\sum_{t=1}^{n-1} \left(S_{i_{t+1}} - S_{i_{t}}\right) \left(S_{j_{t+1}} - S_{j_{t}}\right)}{\sqrt{\sum_{t=1}^{n-1} \left(S_{i_{t+1}} - S_{i_{t}}\right)^{2}} \sqrt{\sum_{t=1}^{n-1} \left(S_{j_{t+1}} - S_{j_{t}}\right)^{2}}}$$
(50)

$$d_{CORT}(S_{i_T}, S_{j_T}) = \varphi_q[CORT(S_{i_T}, S_{j_T})] \cdot d(S_{i_T}, S_{j_T})$$
(51)

$$\varphi_q \left[CORT\left(S_{i_T}, S_{j_T}\right) \right] = \frac{2}{1 + \exp\left(q \cdot CORT\left(S_{i_T}, S_{j_T}\right)\right)}, q \ge 0$$
(52)

Neste método, se *CORT* (S_{1_T}, S_{2_T}) se aproxima de 1 então ambas as séries temporais apresentam um comportamento dinâmico similar (suas tendências, em qualquer instante do tempo, são similares em direção e em taxa de crescimento/decrescimento). Se próximo de - 1, implica que as tendências apresentam mesma taxa de crescimento/decrescimento, mas opostas em direção. Se igual a 0, indica que não há monotonocidade entre S_{i_T} e S_{j_T} e suas taxas de crescimento/decrescimento são estocasticamente e linearmente independentes. O índice de dissimilaridade modulado pela proximidade entre os dados brutos dependerá de uma função adaptativa de equalização φ_q para automaticamente modular a distância convencional entre os dados brutos $d(S_{i_T}, S_{j_T})$ – que pode se dar por d_{L_p} ou d_{DTW} – conforme a correlação temporal. O peso da dissimilaridade aumenta (decresce) à medida em que a correlação temporal decresce de 0 a – 1 (aumenta de 0 até + 1). Adicionalmente, $d_{CORT}(S_{i_T}, S_{j_T})$ deve se aproximar à distância entre os dados brutos à medida em que a correlação temporal *CORT* (S_{i_T}, S_{j_T}) chega a zero. O valor de q, por sua vez, é arbitrário e modula as contribuições da proximidade entre os valores brutos e do comportamento das séries para o índice de dissimilaridade $d_{CORT}(S_{i_T}, S_{j_T})$. Quanto maior (menor) o valor de q, maior (menor) a contribuição de *CORT* (S_{i_T}, S_{j_T}) e menor (maior) a contribuição de $d(S_{i_T}, S_{j_T})$. Se q for igual a 0 (Tabela 3), haverá contribuição de 100% de $d(S_{i_T}, S_{j_T})$, que pode ser a própria distância Euclidiana ou DTW (CHOUAKRIA e NAGABHUSHAN, 2007; MONTERO e VILAR, 2014).

Valor de q	Contribuição do comportamento $CORT \left(S_{i_T}, S_{j_T}\right) [\%]$	Contribuição da distância entre dados brutos $d(S_{i_T}, S_{j_T})$ [%]
q = 0	0	100
q = 1	46,2	53,7
q=2	76,2	23,8
<i>q</i> = 3	90,5	9,4
$q \ge 5$	≈ 100	pprox 0

Tabela 3 – Contribuição do comportamento e da distância entre séries conforme valor de q

Fonte: Chouakria e Nagabhushan (2007)

Para ilustrar diferenças no tocante aos conceitos de dissimilaridade baseados na forma ou na estrutura, tem-se a Figura 49a, que apresenta nove séries temporais em três padrões (P1, P2 e P3). Visualiza-se que as séries de padrão P1estão em fase oposta às de P2 – no instante em que umas atingem o pico, as outras atingem o vale – e em mesma fase às de P3, com picos e vales coincidentes.

Contudo, séries de P1 estão mais próximas às de P2, são mais similares em termos de forma, o que implica que num eventual processo de clusterização em 2 *clusters* com base na distância Euclidiana, então estas séries estariam agrupadas num mesmo *cluster* (Figura 49b). Séries de P1 e P3, embora distantes umas das outras, apresentam tendências de crescimento/decrescimento similares, o que implica que num eventual processo de clusterização em 2 *clusters* com base na estrutura (como correlação de Pearson ou índice adaptativo com q = 2), estas séries pertenceriam a um mesmo *cluster* (Figura 49c) (MONTERO e VILAR, 2014).





Fonte: adaptado de Montero e Vilar (2014)

Existem outros métodos matemáticos para calcular distância entre séries temporais para quando elas apresentarem alta dimensionalidade ou quando não possuírem observações discretas, sendo necessário extrair características estatísticas ou partir para modelos representativos. Nestes casos, recorre-se à comparação de sequências de valores extraídos das séries originais – como autocorrelação, correlação cruzada, valores espectrais, coeficientes *wavelets* – que permitem representar a estrutura dinâmica de cada série por um vetor de menor dimensionalidade (MONTERO e VILAR, 2014).

Ao se calcular as distâncias entre pares de séries temporais, armazenam-se os valores em uma estrutura matricial. Trata-se da matriz dissimilaridade $M(d(S_{i_T}, S_{j_T}))$, cujos elementos são simétricos em relação à diagonal principal – sendo que esta é composta por zeros, já que a dissimilaridade entre um objeto e ele mesmo é zero (KAUFMAN e ROUSSEEUW, 2005).

$$M(d(S_{i_{T}}, S_{j_{T}})) = \begin{bmatrix} d(S_{1_{T}}, S_{1_{T}}) = 0 & d(S_{1_{T}}, S_{2_{T}}) & \cdots & d(S_{1_{T}}, S_{n_{T}}) \\ d(S_{1_{T}}, S_{2_{T}}) & d(S_{2_{T}}, S_{2_{T}}) = 0 & \cdots & d(S_{2_{T}}, S_{n_{T}}) \\ \vdots & \vdots & \ddots & \vdots \\ d(S_{1_{T}}, S_{n_{T}}) & d(S_{2_{T}}, S_{n_{T}}) & \cdots & d(S_{n_{T}}, S_{n_{T}}) = 0 \end{bmatrix}$$
(53)

A matriz dissimilaridade é dado de entrada de algoritmos que buscam minimizar a distância entre séries de um mesmo *cluster* enquanto buscam, simultaneamente, maximizar a distância entre diferentes *clusters*. Existem diferentes amplos grupos de algoritmos de

(b)

clusterização, sendo hierárquica e por repartição os grupos mais comuns, porém havendo outros baseados em modelos e até um grupo de algoritmos que combina diferentes métodos, desenvolvendo uma clusterização híbrida em múltiplas etapas (AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).

2.5.2 Clusterização hierárquica

O grupo de algoritmos de clusterização hierárquica envolve a organização de uma hierarquia de *clusters* partindo de uma abordagem aglomerante ou divisora, que diferem entre si pela direção oposta com a qual constroem essa hierarquia (Figura 50). A aglomerante (também conhecida por Agglomerative Nesting – AGNES – ou bottom-up) inicialmente considera cada série temporal como um *cluster* individual, são as ramificações, para que gradualmente sejam aglomerados em clusters maiores (os nós) até que se tenha todas as séries num único cluster raiz. A abordagem divisora (também conhecida como Divisive Analysis -DIANA – ou top-down), em contraste, inicia com todos os objetos em um único grande cluster (a raiz) e gradualmente os divide até que cada série se torne seu próprio *cluster* (KAUFMAN e SHIRKHORSHIDI ROUSSEEUW, 2005; AGHABOZORGI, WAH, 2015: e KASSAMBARA, 2017).



Figura 50 - Abordagens aglomerante e divisora da clusterização hierárquica

O interessante deste grupo de algoritmos é a representação gráfica da árvore hierárquica resultante – o dendrograma (a exemplo da Figura 49) – que evidencia a distância entre uma série e outra, de modo a facilitar a decisão acerca do agrupamento. Isto se deve ao fato de que na clusterização hierárquica não é necessário especificar o número final de *clusters* a ser

Fonte: adaptado de Kassambara (2017)
produzido antes da aplicação do algoritmo. Portanto, a partir do dendrograma, pode-se decidir em que nível a hierarquia será segmentada para gerar os *clusters* em numeração adequada aos dados (AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015; KASSAMBARA, 2017).

Todavia, a clusterização hierárquica sofre com o defeito de que nunca poder-se-á reparar o que foi feito em etapas anteriores. De fato, uma vez o algoritmo aglomerante uniu dois objetos num mesmo *cluster*, eles não mais serão separados – de mesma forma que no algoritmo divisor, quando dois objetos são divididos em *clusters* distintos eles não poderão ser reunidos novamente. Esta rigidez com a qual algoritmos hierárquicos operam é apontada como sua grande vantagem (por reduzir tempo de computação) como sua principal desvantagem (a inabilidade de corrigir decisões errôneas) (KAUFMAN e ROUSSEEUW, 2005).

Na abordagem divisora, a análise se inicia com todos os objetos dentro de um mesmo cluster Q_k e procede pela divisão sucessiva de cada cluster em dois clusters menores, até que cada série temporal esteja dentro de seu próprio cluster. Este algoritmo requer uma matriz dissimilaridade pois, em cada estágio, os clusters são vasculhados e aquele que apresentar maior dissimilaridade (representada pelo diâmetro $D(Q_k)$) entre quaisquer duas de suas séries temporais $S_{i_T} e S_{j_T}$ será selecionado e varrido em busca da série temporal que apresentar maior dissimilaridade média com todas as outras séries deste mesmo cluster. Ela será segregada em um novo cluster. As séries mais próximas daquela segregada serão, então, transferidas do cluster de origem para este mesmo grupo até que dois clusters distintos se formem e sejam analisados no próximo estágio.

$$D(Q_k) = \max_{\substack{S_{i_T} \in Q_k \\ S_{j_T} \in Q_k}} d(S_{i_T}, S_{j_T})$$
(54)

Já na abordagem aglomerante, além da matriz dissimilaridade, os algoritmos requerem valores de dissimilaridade entre *clusters*, e existem maneiras distintas de se definir uma distância entre *clusters*. É possível que a dissimilaridade entre um par de *clusters* assuma o valor médio das distâncias entre pares de séries temporais $d(S_{i_T}, S_{j_T})$, onde cada série habita um destes *clusters* (*average linkage*). Pode, também, assumir o valor da menor distância dentre todas as dissimilaridades de cada par de séries temporais que habitam *clusters* diferentes (*single linkage*). Ou, ainda, pode assumir o valor da maior distância dentre todas as dissimilaridades de cada par de séries temporais que habitam *clusters* diferentes (*single linkage*). Gu, ainda, pode assumir o valor da maior distância dentre todas as dissimilaridades de cada par de séries temporais que habitam *clusters* diferentes (*single linkage*) (Figura 51) (KAUFMAN e ROUSSEEUW, 2005).

Figura 51 – Representação de algumas diferentes definições de dissimilaridade entre dois *clusters* (de cima para baixo: *average linkage*; *single linkage*; *complete linkage*)



Fonte: Kaufman e Rousseeuw, 2005

Sejam dois *clusters* genéricos Q_1 e Q_2 , ambos compostos por um conjunto de séries temporais que apresentam dissimilaridades $d(S_{i_T}, S_{j_T})$ entre si (podendo estas terem sido calculadas pelo método da distância de Minkowski, DTW, índice adaptativo, correlação de Pearson ou outra), a distância entre *clusters* poderá ser calculada por:

$$d_{sl}(Q_1, Q_2) = \min_{\substack{S_{i_T} \in Q_1 \\ S_{j_T} \in Q_2}} d(S_{i_T}, S_{j_T})$$
(55)

$$d_{cl}(Q_1, Q_2) = \max_{\substack{S_{i_T} \in Q_1 \\ S_{j_T} \in Q_2}} d(S_{i_T}, S_{j_T})$$
(56)

$$d_{avg}(Q_1, Q_2) = \frac{1}{|Q_1||Q_2|} \sum_{\substack{S_{i_T} \in Q_1 \\ S_{j_T} \in Q_2}} d(S_{i_T}, S_{j_T})$$
(57)

Onde as dissimilaridades entre clusters d_{sl} , d_{cl} e d_{avg} representam os métodos *single linkage*, *complete linkage* e *average linkage*, respectivamente e $|Q_1|$ e $|Q_2|$ são os números de objetos que os habitam.

Existem outros métodos para desenvolver a aglomeração e que envolvem a distância Euclidiana entre os centroides $\bar{c}(Q_1) \in \bar{c}(Q_2)$ de dois *clusters* genéricos $Q_1 \in Q_2$, podendo ser propícios para dados intervalados não-ausentes. São os métodos *centroid*, Ward.D e Ward.D2, cujas distâncias entre $Q_1 \in Q_2$ são dadas por d_c , $d_{w1} \in d_{w2}$, respectivamente:

$$d_c(Q_1, Q_2) = \|\bar{c}(Q_1) - \bar{c}(Q_2)\|$$
(58)

$$d_{w1}(Q_1, Q_2) = \frac{|Q_1||Q_2|}{|Q_1| + |Q_2|} \|\bar{c}(Q_1) - \bar{c}(Q_2)\|^2$$
(59)

$$d_{w2}(Q_1, Q_2) = \sqrt[2]{\frac{|Q_1||Q_2|}{|Q_1| + |Q_2|}} \|\bar{c}(Q_1) - \bar{c}(Q_2)\|^2$$
(60)

$$ESQ (Q_1 \cup Q_2) = \sum_{i \in (Q_1 \cup Q_2)} \left\| X_{T_i} - \bar{c}(Q_1 \cup Q_2) \right\|^2$$
(61)

Os métodos Ward.D e Ward.D2 utilizam o mesmo critério de otimização, que perpassa pela minimização da variância ou erro da soma de quadrados (*ESQ*) entre cada objeto e o centroide resultante da fusão de dois *clusters* Q_1 e Q_2 . Isto se dá pela busca exponencial no espaço de soluções possíveis e, por ser uma busca exaustiva, garante soluções próximas do ótimo de maneira satisfatória (KAUFMAN e ROUSSEEUW, 2005; MURTAGH e LEGENDRE, 2014).

2.5.3 Clusterização por repartição

Um método de repartição construirá um número k de *clusters* de maneira que duas condições sejam satisfeitas: i) cada grupo deve conter ao menos um objeto (o que implica que o número k deve ser menor ou igual ao número de objetos) e ii) cada objeto deve pertencer exatamente a um grupo (o que implica que dois *clusters* diferentes não podem ter objetos em comum). Tais condições garantem que todos os objetos sejam agrupados e pertençam a um dos k grupos. Vale ressaltar que k é predeterminado pelo usuário, de maneira que um algoritmo de repartição repartirá e alocará os objetos em quantos *clusters* forem desejados, porém buscando fazê-lo de maneira que objetos de um mesmo *cluster* sejam similares, enquanto objetos de diferentes *clusters* sejam bastante diferentes (KAUFMAN e ROUSSEEUW, 2005).

Um dos algoritmos mais utilizados para a clusterização por repartição é o *k-Means*, onde cada *cluster* possui um protótipo que equivale ao valor médio de todos os objetos (ou seja, será um objeto artificial), de modo a minimizar a distância total entre todos os objetos pertencentes a um mesmo *cluster* e seu protótipo. Em outros algoritmos de repartição a definição dos protótipos assume outras diretrizes, a exemplo do *Partitioning Around Medoids* (PAM –

repartição em torno de medoides, em tradução livre) ou *k-Medoids*, onde um número *k* de objetos será selecionado conforme maior proximidade ao centro de seus *clusters*, tornando-se seus protótipos representativos (KAUFMAN e ROUSSEEUW, 2005; AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).

Possivelmente que, em se tratando de séries temporais como os objetos de análise, a clusterização por repartição seja desafiadora e nada trivial. Delimitar o número *k* de *clusters* pode ser impraticável quando deseja-se obter resultados de clusterização que sejam naturais e que revelem uma estrutura já presente nos dados. Assim, bons resultados partirão de uma seleção apropriada de protótipos e que preferencialmente as séries temporais tenham dimensionalidades equivalentes (AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).

No algoritmo *k-Means* proposto por Hartigan e Wong (1979), define-se a variação total de um *cluster* como a soma dos quadrados das distâncias Euclidianas entre cada objeto e o objeto médio deste *cluster*. No caso, cada série temporal S_{i_T} será alocada no *cluster* que minimiza a soma dos quadrados das distâncias entre esta série S_{i_T} e uma série temporal média μ_k , que funcionará como protótipo de centroide daquele *cluster* (KASSAMBARA, 2017).

$$SQD(Q_k) = \sum_{S_{i_T} \in Q_k} (S_{i_T} - \mu_k)^2$$
(62)

Desta forma, ao indicar o número k de *clusters* a serem gerados pelo método k-Means, o algoritmo passa a selecionar aleatoriamente um mesmo número k de séries temporais que servirão de protótipos provisórios para os k *clusters*. As outras séries temporais serão alocadas nos *clusters* cujos protótipos provisórios apresentam menor distância Euclidiana entre elas. Os protótipos de cada *cluster* são então atualizados e assumem valores μ_k , sendo utilizados para reavaliar se as séries temporais devem ser realocadas para um *cluster* cujo protótipo apresenta menor distância com a série em questão. Este processo deve ser repetido de maneira iterativa, até que se minimize o valor *SQD* (Q_k) de cada *cluster* (KASSAMBARA, 2017).

O resultado desta clusterização por repartição *k-Means* é sensível à seleção aleatória dos protótipos provisórios. O que significa que a cada diferente execução do algoritmo, diferentes protótipos serão escolhidos, acarretando diferentes *clusters*. Caso a ordem dos objetos seja rearranjada, possivelmente o algoritmo chegará em uma solução distinta também, fazendo com que a determinação de uma clusterização ótima seja um processo moroso.

Algoritmo de clusterização por repartição alternativo ao *k-Means* é o PAM, que apresenta menor sensibilidade a *outliers* e produz *clusters* visualmente mais compactos, em formatos esféricos. A seleção de protótipos representativos, denominados medoides neste algoritmo, não é aleatória, mas respeita uma certa centralidade nos *clusters* que os envolve. Nesse sentido, é possível que os protótipos sejam de fato representativos de todos os objetos sendo clusterizados, carregando suas características estruturais. Tal recurso permite que novas análises sejam realizadas diretamente com os medoides, partindo da premissa de que eles representam os outros objetos de seus *clusters* (KAUFMAN e ROUSSEEUW, 2005).

O medoide é definido como o objeto que apresenta a mínima dissimilaridade média em relação a todos os outros objetos de um mesmo *cluster*, ele é o objeto mais centralizado entre os outros. Para um número k predeterminado de *clusters*, os primeiros k medoides são aqueles cujas somas de dissimilaridades com todos os outros objetos sejam minimizadas. Subsequentemente em cada estágio, outro objeto qualquer é selecionado, que no caso pode ser uma série temporal S_{i_T} , tal qual a diferença entre a dissimilaridade entre S_{i_T} e a série medoide S_{k_T} de seu *cluster* e a dissimilaridade entre S_{i_T} e uma outra série qualquer S_{j_T} seja positiva. Realiza-se este processo de maneira iterativa e soma-se estas diferenças positivas até que se maximize tal soma e até que todas as séries medoides S_{k_T} sejam encontradas:

$$\max \sum_{j} \max(d(S_{i_T}, S_{k_T}) - (S_{i_T}, S_{j_T}), 0)$$
(63)

Após esta etapa, é possível aprimorar os medoides, trocando-os por outros que ainda não foram selecionados. A função objetivo é minimizar a soma das dissimilaridades entre todos os objetos e o seu medoide mais próximo, e as trocas se sucedem até alcançar o mínimo e os medoides mais representativos sejam identificados (KAUFMAN e ROUSSEEUW, 2005).

2.5.4 Clusterização híbrida

Alguns algoritmos de clusterização operam partindo de múltiplas abordagens. Em específico, o algoritmo de clusterização hierárquica por repartição *k-Means* visa aprimorar a deficiência do algoritmo de repartição *k-Means*, que determina os centroides de cada *cluster* de maneira aleatória e iterativa. Desta forma, no algoritmo híbrido, a primeira etapa requer a aplicação da clusterização hierárquica para determinar o dendrograma que será segmentado em

um número *k* de *clusters*. Depois estes *clusters* terão seus centroides calculados e serão utilizados para determinar novos *clusters* na etapa de clusterização por repartição *k-Means*. As duas etapas são realizadas conforme explicado anteriormente (KASSAMBARA, 2017).

2.5.5 Validação interna da clusterização

Como já mencionado, a escolha por um algoritmo de clusterização dependerá tanto do tipo de dados disponíveis quanto do objetivo da análise. Por vezes, vários algoritmos diferentes são aplicáveis, em função de que argumentos *à priori* podem não ser suficientes para reduzir a seleção a um único método. Nestes casos e, partindo do pressuposto que a análise de clusterização é normalmente utilizada como uma ferramenta descritiva ou exploratória (sem intenções de se chegar a inferências ou comprovar/rejeitar hipóteses), pode ser interessante aplicar vários algoritmos no mesmo conjunto de dados (KAUFMAN e ROUSSEEUW, 2005).

Partindo da definição básica de que os objetos em um mesmo *cluster* devem apresentar máxima similaridade entre si e que objetos de *clusters* distintos devem apresentar máxima dissimilaridade entre si, é desejável que distância média entre os objetos de mesmo *cluster* seja minimizada – denotando compactação – e a distância média entre *clusters* seja maximizada – denotando separação. Nesse sentido, medidas de validação interna normalmente abordam estes dois termos (HALKIDI, BATISTAKIS e VAZIRGIANNIS, 2001; KASSAMBARA, 2017).

O coeficiente de Dunn é uma destas medidas. Nesta métrica calcula-se a média da análise combinatória das distâncias entre pares de *clusters* pelo método do vizinho mais próximo (*single linkage*) e calcula-se a média de maiores diâmetros de cada *cluster*, que se dá pela maior distância entre pares de objetos. Para que indique boa compactação dos *clusters* individuais e boa separação média de um *cluster* a outro, o coeficiente de Dunn deve ser apresentar maiores valores positivos:

$$Dunn = \frac{\left(\frac{k!}{2! (k-2)!}\right)^{-1} \sum_{1}^{k-1} \min_{\substack{S_{i_T} \in Q_k \\ S_{j_T} \in Q_{k+1}}} d(S_{i_T}, S_{j_T})}{\frac{1}{k} \sum_{1}^{k} \max_{\substack{S_{i_T} \in Q_k \\ S_{j_T} \in Q_k}} d(S_{i_T}, S_{j_T})}$$
(64)

Para a abordagem hierárquica, em específico, é possível partir de diferentes métodos para cálculo de distância tanto entre séries temporais como entre *clusters*, produzindo resultados

diferentes. Logo, a avaliação acerca de qual combinação de métodos de cálculo de dissimilaridade pode ser mais interessante para prosseguir com a análise de clusterização é uma etapa significativa.

Ao realizar a clusterização hierárquica (aglomerante ou divisora), existe uma maneira de medir o quão bem o dendrograma gerado após o cálculo de dissimilaridade entre *clusters* reflete a matriz dissimilaridade calculada para as séries temporais. Este dendrograma – $H(h(S_{i_T}, S_{j_T}))$ – assume formato similar à matriz dissimilaridade $M(d(S_{i_T}, S_{j_T}))$, onde para um conjunto de *n* séries temporais existirá uma chamada distância cofenética entre as ramificações e os nós da árvore hierárquica que indica o nível de proximidade $h(S_{i_T}, S_{j_T})$ com o qual duas séries temporais se aglomeram num dado *cluster* pela primeira vez. Deve-se computar a correlação entre a matriz dissimilaridade e a matriz de distâncias cofenéticas (HALKIDI, BATISTAKIS e VAZIRGIANNIS, 2002).

$$Coph = COR(M(d(S_{i_T}, S_{j_T})), H(h(S_{i_T}, S_{j_T})))$$
(65)

$$Coph = \frac{\frac{1}{N}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d(S_{i_{T}}, S_{j_{T}}) \cdot h(S_{i_{T}}, S_{j_{T}}) - \mu_{M} \cdot \mu_{H}}{\sqrt{\frac{1}{N}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d(S_{i_{T}}, S_{j_{T}})^{2} - \mu_{M}^{2}} \sqrt{\frac{1}{N}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} h(S_{i_{T}}, S_{j_{T}})^{2} - \mu_{H}^{2}}}$$
(66)

$$N = n \cdot (n-1) \tag{67}$$

$$\mu_M = \frac{1}{N} \sum_{i=1}^{n-1} \sum_{j=i+1}^n M(d(S_{i_T}, S_{j_T}))$$
(68)

$$\mu_{H} = \frac{1}{N} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} H(h(S_{i_{T}}, S_{j_{T}}))$$
(69)

Se a clusterização for válida, a ligação de series temporais no dendrograma deve apresentar forte correlação com a distâncias armazenadas na matriz dissimilaridade – se aproximando de 1, preferencialmente acima de 0,75 (HALKIDI, BATISTAKIS e VAZIRGIANNIS, 2002; KASSAMBARA, 2017).

3 MATERIAIS E MÉTODOS DE PESQUISA

A presente pesquisa partiu de uma abordagem quantitativa, pela qual definiu-se um conjunto de variáveis mensuráveis de maneira que, uma vez operacionalizadas, permitiram a compreensão de informações de relevância ao tema da pesquisa. Além do trabalho com variáveis mensuráveis e operacionalizáveis, a causalidade é uma outra característica da abordagem quantitativa, pela qual busca-se o estabelecimento de relação entre variáveis dependentes (efeito) e independentes (causas). O princípio da reprodutibilidade buscou ser respeitado e é até mesmo incentivado, pois entende-se que a reprodução da pesquisa com outra população de dados pode corroborar com os resultados, além de permitir a verificação da existência de viés e tendências (MIGUEL et al., 2012).

3.1 DADOS UTILIZADOS

3.1.1 Dados representativos de recursos eólicos

Partindo da conjectura (e necessidade) de que é importante mapear o maior número possível de ocorrência de períodos de escassez do recurso eólico determinados pela sua variabilidade interanual, fez sentido explorar séries temporais mais longas, que superam um século de extensão, com registros mensais de velocidade do vento (em m/s). Dados eólicos foram extraídos de bases de dados de reanálise de maior abrangência temporal, a exemplo das bases C20r V2 e ERA-20C, bem como da base MERRA-2, esta de menor abrangência, com propósito de validação em função do teor de incerteza de seu modelo quantitativo ser relativamente menor. Todos estes dados são de domínio público, sendo disponibilizados nos portais de cada um dos núcleos responsáveis pela modelagem de reanálise.

Os dados eólicos provenientes de projetos de reanálise abrangem uma área de tamanho flexível. Na intenção de representar o NE do Brasil por ser a região mais significativa em termos de geração pela fonte eólica, a delimitação das fronteiras da área a ser analisada foi feita de maneira a englobar o território em que se concentram a maior parte dos parques eólicos. Desta forma, tomou-se a região delimitada pelas coordenadas de latitude 18S-0 e coordenadas de longitude de 48W-33W (retângulo na Figura 52). Esta região envolve, também, a costa do NE

brasileiro, permitindo o estudo de séries temporais de velocidade de vento para localidades offshore – em alto mar.



Figura 52 - Região delimitada para avaliação de recursos eólicos

Fonte: elaboração própria do autor

No que se refere à resolução espacial, estas três bases de dados diferem entre si, o que determinou o número de pontos (definidos por coordenadas de latitude e longitude) e número de séries temporais que preenchem a região predeterminada, como mostra a Tabela 4 e a Figura 53 a seguir. Doravante as bases de dados serão referenciadas no texto como NOAA, ERA e MERRA.

Base de dados	Resolução espacial	Resolução e cobertura temporal	Altura de registro da magnitude da velocidade do vento	Link de acesso
NOAA 20th Century Reanalysis V2c (COMPO et al., 2011)	2,0° x 2,0° 72 séries temporais	dados mensais 1851 - 2014	10 m acima do solo	https://psl.noaa.gov/data/ gridded/data.20thC_Rean V2c.html
ECMWF ERA-20C (POLI et al., 2016)	1,125° x 1,125° 221 séries temporais	dados mensais 1900 - 2010	100 m acima do solo	https://ecmwf.int/en/fore casts/datasets/reanalysis- datasets/era-20c
MERRA 2 (GELARO et al., 2017)	0,5° x 0,667° 925 séries temporais	dados mensais 1980 - 2017	100 m acima do solo	https://gmao.gsfc.nasa.go v/reanalysis/MERRA-2/

Fonte: elaboração própria do autor



Figura 53 – Distribuição espacial das séries temporais de base de dados NOAA (a), ERA (b) e MERRA (c)

Fonte: elaboração própria do autor

3.1.2 Dados dos modos de variabilidade

O universo de modos de variabilidade é bastante extenso, contemplando fenômenos climáticos que ocorrem em diferentes porções da Terra, principalmente sobre os oceanos Atlântico, Pacífico e Índico. Tais fenômenos que correm em um sistema geofísico são comumente caracterizados, diagnosticados e monitorados por meio de índices climáticos quantitativos.

A escolha de alguns destes índices representativos dos modos de variabilidade foi pautada pela região de interesse dos recursos eólicos, fazendo sentido que estes índices se refiram a fenômenos com um grau de proximidade espacial relativamente alto. Assim, partiuse do pressuposto que a variabilidade de TSM sobre o oceano Atlântico possa se associar mais fortemente com a disponibilidade de recursos eólicos no NE do Brasil. Ademais, o corpo de conhecimento existente aponta para a existência de relações entre modos de variabilidade do Pacífico (notadamente o ENOS) e variações em diferentes variáveis atmosféricas sobre o território brasileiro.

Nesse sentido, optou-se por trabalhar com 5 índices de diferentes modos de variabilidade, sendo 3 deles representativos de fenômenos no Oceano Atlântico (AMM, AMO e MGI) e outros 2 representativos de fenômenos no Oceano Pacífico (NINO 3.4 e PDO). Assim como os dados eólicos, estes também podem ser acessados gratuitamente pelos portais dos

centros de pesquisa que os preparam. As características gerais destes índices – que já foram descritos detalhadamente na seção 2.2 – podem ser conferidas na Tabela 5.

Modo de variabilidade e seu índice	Breve descrição da variável climática representada pelo índice	Resolução e cobertura temporal	<i>Link</i> de acesso
Modo Meridional do Atlântico - AMM	Padrão espacial definido pela máxima covariância entre TSM e componentes meridionais e zonais do vento na região 21S-32N, 74W-15E do Oceano Atlântico (CHIANG e VIMONT, 2004)	dados mensais 1948 - 2018	https://psl.noaa.go v/data/timeseries/ monthly/AMM/a mmsst.data
Oscilação Multidecadal do Atlântico - AMO	Baseado em médias mensais de anomalias de TSM na bacia Norte do Oceano Atlântico, tendo médias mensais globais de TSM removidas (ENFIELD, MESTAS-NUNEZ e TRIMBLE, 2001)	dados mensais 1948 - 2018	https://psl.noaa.go v/data/correlation/ amon.us.data
El Niño Oscilação Sul - NINO 3.4	Baseado em médias mensais de anomalias de TSM na região equatorial 5N-5S e 170W-120W do Oceano Pacífico com remoção da média climatológica (RAYNER et al., 2003)	dados mensais 1870 - 2018	https://psl.noaa.go v/gcos_wgsp/Tim eseries/Data/nino 34.long.anom.data
Oscilação Decadal do Pacífico - PDO	Principal componente da variância conjunta de anomalias mensais de TSM no Norte do Oceano Pacífico. Médias mensais globais de TSM foram removidas (MANTUA et al., 1997)	dados mensais 1900 - 2018	http://research.jisa o.washington.edu/ pdo/PDO.latest
Gradiente do Atlântico Tropical - MGI	Diferença entre anomalias de TSM médias da região 5N-20N, 60W-30W e da região 20S-0, 30W-10E (ENFIELD et al., 1999)	dados mensais 1950 - 2018	https://cpc.ncep.n oaa.gov/products/ GODAS/

Tabela 5 - Características dos modos de variabilidade e seus índices

Fonte: elaboração própria do autor

3.1.3 Dados representativos de recursos hídricos

Para representar os recursos hídricos em sua região mais relevante (do subsistema SE/CO), duas opções foram elencadas. A primeira se refere ao uso de dados de vazões naturais no local de aproveitamento (em uma UHE, por exemplo), pelos quais se informa o fluxo local de água do rio que corre ali – volume de água num dado instante (em m³/s). Outra opção é a utilização de dados de Energia Natural Afluente (ENA), que consistem em informações do conteúdo energético potencial que as vazões de rios podem prover. Trata-se de uma quantidade de energia que poderia ser gerada a partir de condições preestabelecidas para o estado atual do sistema, porém que não foi de fato verificada em termos de geração: dados históricos de ENA remetem ao potencial de geração do passado dadas as condições atuais do sistema.

A ENA é calculada a partir das vazões naturais, porém levando em consideração a produtibilidade equivalente do conjunto turbina-gerador do aproveitamento hidrelétrico relativa

ao armazenamento de 65% do volume útil dos reservatórios no sistema presente (referente à queda obtida pela diferença entre o nível de montante, correspondente a um armazenamento de 65% do volume útil, e o nível médio do canal de fuga). Interessante para o uso de dados de ENA é a unidade de medição, já convertida para energia, em MW médios ou MWh. A ENA pode ser determinada em base diária, semanal, mensal ou anual e, também, por bacia e por subsistema, de acordo com os sistemas de aproveitamentos hidroelétricos existentes nas configurações de bacias hidrográficas e de subsistemas elétricos (ONS, 2020a).

A escolha do subsistema SE/CO como objeto de estudo se justifica por conter um grande número de UHEs em operação (Figura 7, seção 2.1.1) e também pelo seu grande volume de armazenamento nos reservatórios, sobretudo na bacia hidrográfica do Paraná. Porção à nordeste desta bacia concentra a maioria das UHEs de uma região interbacias que acumula 70% da capacidade total de armazenamento do Brasil como um todo, a exemplo de Furnas e Emborcação (Figura 54). E a relevância se intensifica na medida em que a energia armazenável não cresceu (ao contrário da demanda); cheios, os reservatórios desta região acumulavam, em 1970, capacidade de prover energia para um período de 2 a 3 anos, ao passo que em 2018 somente para um período de 4 a 5 meses (HUNT, STILPEN e FREITAS, 2018).

Figura 54 – Reservatórios de UHEs e a área com bordas em vermelho que representa 70% da capacidade total de armazenamento do Brasil



Fonte: Hunt, Stilpen e Freitas (2018)

Para a bacia hidrográfica do Paraná, a UHE de Marimbondo foi apontada como aquela de maior representatividade ao se avaliar pares de séries temporais de vazões mensais de cada UHE desta bacia. O maior valor do coeficiente de correlação par a par e a vazão média mais significativa foram os critérios para seleção da UHE em questão (Figura 55). Ainda, constatouse que em escala de tempo sazonal, a bacia do Paraná (sendo representada pela UHE de Marimbondo) é bem correlacionada de maneira inversa com uma porção da região do NE com alto potencial de geração pela fonte eólica (CANTÃO et al., 2017).





Fonte: Cantão et al. (2017)

Levando em conta o alto volume de geração e o fato de situar-se ao final de toda uma cascata de UHEs que lhe concede o privilégio de receber água que percorre a maioria dos rios da região hidrográfica do Paraná, Itaipu foi outra UHE elegida para análise. Somadas às estas duas UHEs (Itaipu e Marimbondo), dados de ENA para todo o subsistema SE/CO podem agregar ao conjunto informacional de dados hídricos. o que pode fazer disto um agregado informacional de relevância. Todos os dados em questão acumulam um histórico que percorrem o período de 1931 até o presente, o que pareceu razoável para avaliar a variabilidade interanual de recursos hídricos em conjunto com os eólicos, vide Tabela 6.

Entende-se que, em função da participação expressiva de geração de eletricidade pela fonte hídrica no Brasil, o sistema elétrico brasileiro é gerenciado e operacionalizado conforme critérios que ponderam o volume dos reservatórios das UHEs, vazões dos rios, dentre outros fatores hidrológicos. Os dados hídricos explorados nesta pesquisa foram selecionados para

representar, de maneira bastante simplificada, o potencial de geração pela fonte hídrica em condições específicas atreladas à variabilidade interanual.

Variável	Unidade de medição	Resolução e cobertura temporal	Link de acesso				
Vazão natural do ponto de aproveitamento	[m³/s]	dados mensais	https://sintegre.ons.				
da UHE de Itaipu		1931 - 2019	org.br/				
Vazão natural do ponto de aproveitamento	[m³/s]	dados mensais	https://sintegre.ons.				
da UHE de Marimbondo		1931 - 2019	org.br/				
Energia Natural Afluente (ENA) do subsistema SE/CO	[MW médios]	dados mensais	https://sintegre.ons.				
	ou [MWh]	1931 - 2019	org.br/				
Fonta: alaboração própria do autor							

Tabela 6 - Características dos dados representativos de recursos hídricos

Fonte: elaboração própria do autor

3.1.4 Dados de carga e geração de energia elétrica

Necessários à ilustração e às simulações do balanço entre oferta e demanda de energia elétrica no Brasil, alguns dados de demanda, carga e geração de foram levantados junto ao histórico de operação do ONS. São informações de domínio público, com acesso facilitado pelo portal do ONS (Tabela 7).

Variável	Unidade de	Resolução e	Link de acesso
	medição	cobertura temporal	
Carga de energia elétrica	[GWh]	dados mensais 2014 - 2019	http://www.ons.org.br/paginas/resultad os-da-operacao/historico-da-operacao
Geração de energia elétrica pela fonte eólica	[GWh]	dados mensais 2014 - 2019	http://www.ons.org.br/paginas/resultad os-da-operacao/historico-da-operacao
Geração de energia elétrica pela fonte hídrica	[GWh]	dados mensais 2014 - 2019	http://www.ons.org.br/paginas/resultad os-da-operacao/historico-da-operacao
	Б (11 ~ / ! 1	

Tabela 7 – Características dos dados de carga e geração de energia elétrica

Fonte: elaboração própria do autor

3.2 PROCEDIMENTOS METODOLÓGICOS

A pesquisa em questão utilizou de um extenso ferramental matemático abarcado pelo campo da Análise de Séries Temporais – previamente introduzido e conceituado nas seções 2.4 e 2.5. A implementação destas técnicas e análises propostas foi auxiliada pela adoção de alguns *softwares* de interesse habilitados à manipulação de dados e séries temporais e que disponibilizam ampla gama de funções. A principal plataforma de operação foi o RStudio, que dispõe de uma interface associada à linguagem computacional R e que é repleto de funções

capazes de solucionar tarefas mais complexas. Vantagem do R é a existência de uma comunidade ativa de usuários que incrementam novos pacotes de funções em código aberto para uso gratuito por outros usuários, fazendo dele um *kit* bastante completo e de alto teor analítico. Outro *software* extensivamente utilizado foi o Microsoft Excel, programa de fácil utilização e com razoável praticidade para representações visuais e execução de tarefas relativamente mais simples que demandam baixo poder computacional.

A maior parte da pesquisa foi tomada pela análise de séries temporais de dados eólicos. Todas as bases de dados de reanálise eram constituídas de um grande volume de informações – registros de velocidade do vento ao longo do tempo. Como etapa de pré-processamento, iniciou-se com a aplicação de técnicas de clusterização para agrupar séries temporais de comportamento semelhante em *clusters* comuns, processo este descrito no Capítulo 4. Destes *clusters*, algumas séries temporais foram eleitas como suas representantes, por apresentarem maior correlação média com todas as outras séries vizinhas de mesmo agrupamento. Esta etapa fora enquadrada em linha tracejada de coloração verde na Figura 56.

Estas séries representativas passaram por uma sequência de análises com vistas para descrever suas características gerais (médias, quartis, *IQR*, *boxplots*, distribuições de frequência empírica, histograma) bem como a ocorrência de eventos extremos de escassez de recursos eólicos (probabilidades de ocorrência, tempo de retorno) e escassez conjunta com recursos hídricos. Já a identificação de possíveis padrões cíclicos, ou periodicidades, foi facilitada pela Análise de Componentes Principais (ACP) e pela Análise Espectral, tendo como alvo todas as séries temporais originais – e não apenas as séries representativas dos *clusters*. Todos estes procedimentos estão relatados no Capítulo 5 e foram destacados pelos dois enquadramentos tracejados de coloração cinza na Figura 56.

A etapa subsequente, descrita no Capítulo 6, consistiu na exploração de fenômenos climáticos com conhecida atuação sobre sistemas atmosféricos brasileiros para verificar se os modos de variabilidade exercem influência sobre a ocorrência de eventos de escassez eólica. Para tanto, a ACP foi utilizada novamente, desta vez sobre as séries temporais de dados eólicos de mesmo *cluster* a fim de extrair componentes principais que representam boa parte da variância conjunta de cada *cluster*. Estas componentes foram analisadas por correlação de Pearson com os índices dos modos de variabilidade descritos pelos fenômenos climáticos explorados – enquadramento laranja da Figura 56.

O último capítulo aborda a concepção de cenários com base em potencial instalável de recursos eólicos e hídricos e com base na carga que possa vir a se concretizar futuramente. Uma série de premissas foram adotadas para estabelecer critérios de simulações de geração com base em recursos eólicos e hídricos em anos de baixa disponibilidade em cada um dos cenários conceituados. Foi necessário converter dados eólicos em dados de energia gerada e, portanto, um modelo de conversão energética foi adotado.

Nesta etapa, dados hídricos de ENA e dados de carga são considerados também, assim como boa parte dos resultados obtidos em etapas prévias reaparecem para auxiliar no delineamento dos cenários e das premissas adotadas. Entende-se que estas últimas análises e resultados finais (quadro tracejado de coloração roxa na Figura 56) constituem a principal contribuição desta tese, conectando todos os outros apuramentos.



Figura 56 - Esquema dos procedimentos metodológicos implementados

Fonte: elaboração própria do autor

4 CLUSTERIZAÇÃO DE SÉRIES TEMPORAIS DE DADOS EÓLICOS

As séries temporais que compõem as bases de dados NOAA e ERA são mais longas do que as séries da base MERRA, o que significa que são compostas por sequências numéricas de um maior número de valores. Em todo caso, para a análise de clusterização não houve necessidade de recorrer a modelos de representação de séries temporais para promover a redução da dimensionalidade. Este artifício teria como foco a aceleração do processo de agrupamento pela redução do tempo computacional de execução, já que os vetores representativos seriam obrigatoriamente mais curtos por serem compostos de um menor número de elementos (AGHABOZORGI, SHIRKHORSHIDI e WAH, 2015).

Entende-se que mesmo a base do NOAA com séries temporais mais longas (de quase 2 mil observações cada, vide Tabela 4) não implica em demanda computacional exacerbada. As séries temporais das três bases de dados já são resultantes de modelos computacionais climáticos de reanálise, o que as torna menos propensas a distorções e *outliers* e seria possível que estudá-las por meio de outros modelos aumentaria ainda mais o nível de incerteza acerca da variável. Ademais, as observações nelas contidas representam valores de médias mensais – são, portanto, discretas. Desta forma, as séries temporais com seus dados brutos foram analisadas por meio de algoritmos de clusterização que levam em consideração a dissimilaridade entre todos os valores de uma série e todos os valores de outra série temporal.

O cálculo da dissimilaridade entre as séries temporais, como visto na seção 2.5.1, pode ser realizado com base na forma ou na estrutura. Para as séries pesquisadas, inspecionar visualmente se há defasagem temporal ou se as tendências são equivalentes seria laborioso, de maneira que realizar a clusterização com base nos dois conceitos de dissimilaridade e posterior análise de validação interna pareceu uma rota mais sensata. Por conseguinte, optou-se por calcular a matriz dissimilaridade utilizando o método da distância Euclidiana, da correlação de Pearson e o do índice adaptativo: respectivamente as funções diss("EUCL"), diss("COR") e diss("CORT") do pacote de funções "TSclust" (MONTERO e VILAR, 2014) no RStudio.

Com respeito aos algoritmos de clusterização, também não havia um que fizesse mais sentido para prosseguir com o agrupamento. Assim, algoritmos de clusterização hierárquica, por repartição e até mesmo um algoritmo híbrido foram utilizados, partindo da premissa de que a análise de validação interna permitiria a seleção dos resultados de clusterização mais aderentes a cada uma das bases de dados.

4.1 CLUSTERIZAÇÃO HIERÁRQUICA

Para a realização da clusterização hierárquica, algoritmos com abordagens aglomerante e divisora foram explorados. Para tal, o pacote "stats" (R CORE TEAM, 2020) no RStudio dispõe as funções diana(), de abordagem divisora, e hclust(), de abordagem aglomerante e que requer a indicação de um dos métodos de aglomeração (vide seção 2.5.2), permitindo escolher entre os métodos Ward.D2, *complete linkage, average linkage*, Ward.D, *centroid* e *single linkage*.

Todos estes métodos e abordagens foram aplicadas diretamente às três diferentes matrizes dissimilaridade (distância Euclidiana, correlação de Pearson e índice adaptativo – funções diss("EUCL"), diss("COR") diss("CORT")) de cada base de dados. Vale ressaltar que para o cálculo da matriz dissimilaridade com a função diss("CORT") foi necessário indicar valores arbitrários para o termo q do índice adaptativo. Isto se deu de maneira sistemática, saindo de q = 0 até q = 5, acrescendo de 0,1 em 0,1.

Após executadas as funções supracitadas, calculou-se o coeficiente de correlação entre as matrizes dissimilaridade e as matrizes de distâncias cofenéticas, como medida de validação interna, em busca de uma combinação de cálculo de matriz dissimilaridade e método de clusterização que fosse mais aderente aos dados NOAA, ERA e MERRA. No caso, valores mais altos dos resultados desta correlação indicam métodos mais interessantes para prosseguir com o agrupamento das séries temporais. Para cada combinação de cada base de dados um valor diferente foi encontrado, o que mostra quão sensível é a análise de clusterização aos parâmetros do algoritmo e à matriz dissimilaridade que os alimenta. Os resultados foram dispostos na Tabela 8.

Para NOAA, a abordagem divisora (DIANA) com matriz dissimilaridade baseada na distância Euclidiana foi superior, chegando a 0,8261. Para os dados ERA, a abordagem aglomerante com o método de clusterização hierárquica *average linkage* e dissimilaridade resolvida por correlação de Pearson atingiu o valor mais interessante, de 0,8136. Os dados MERRA foram mais bem arranjados em *clusters* pelo método do índice adaptativo com o valor de q = 1,7 e retornando 0,7693 – o que mostra que as estruturas das séries temporais de ERA e MERRA possuem um pouco mais de relevância do que seus valores dos elementos que as compõem.

Abordagem da Método da		Dados NOAA – Método do cálculo da matriz dissimilaridade				
clusterização hierárquica	clusterização hierárquica	Distância Euclidiana	Correlação de Pearson	Índice Adaptativo $q = 0,2$		
Aglomerante	Ward.D2	0,8160	0,7048	0,8167		
Aglomerante	complete linkage	0,7930	0,7317	0,7895		
Aglomerante	average linkage	0,8075	0,7727	0,8054		
Aglomerante	Ward.D	0,8137	0,6727	0,8104		
Aglomerante	centroid	0,8049	0,6982	0,8238		
Aglomerante	single linkage	0,7282	0,6085	0,7425		
Divisora	Diana	0,8261	0,7279	0,8138		
Abordagem da	Método da	Dados ERA – Métod	o do cálculo da mat	riz dissimilaridade		
clusterização clusterização hierárquica hierárquica		Distância Euclidiana	Correlação de Pearson	Índice Adaptativo $q = 0,1$		
Aglomerante	Ward.D2	0,7144	0,6991	0,7121		
Aglomerante	complete linkage	0,7230	0,7503	0,7113		
Aglomerante	average linkage	0,7742	0,8136	0,7725		
Aglomerante	Ward.D	0,7011	0,5941	0,6996		
Aglomerante	centroid	0,7665	0,6951	0,7640		
Aglomerante	single linkage	0,6824	0,6556	0,6811		
Divisora	Diana	0,7266	0,8125	0,7291		
Abordagem da	Método da	Dados MERRA – Méte	odo do cálculo da m	atriz dissimilaridade		
clusterização hierárquica	clusterização hierárquica	Distância Euclidiana	Correlação de Pearson	Índice Adaptativo $q = 1,7$		
Aglomerante	Ward.D2	0,6961	0,6808	0,6382		
Aglomerante	complete linkage	0,7050	0,6794	0,7074		
Aglomerante	average linkage	0,6802	0,7542	0,7693		
Aglomerante	Ward.D	0,6456	0,6282	0,6241		
Aglomerante	centroid	0,7296	0,6801	0,7566		
Aglomerante	single linkage	0,4787	0,3892	0,5369		
Divisora	Diana	0,7099	0,7107	0,6690		

Tabela 8 – Coeficiente de correlação entre a matriz dissimilaridade e a matriz de distância cofenética resultante do processo de clusterização hierárquica

Fonte: elaboração própria do autor

Resultado dos processos de clusterização hierárquica envolve a apresentação da hierarquia em um dendrograma, podendo este ser segmentado em uma determinada altura para delimitar o número final de *clusters*. Isto foi feito com o melhor resultado de correlação entre as matrizes dissimilaridade e de distância cofenética para cada base de dados, optando pelo respectivo método de cálculo da matriz dissimilaridade para prosseguir com a segmentação de 5 *clusters*.

Tal escolha do número de *clusters* levou em consideração o tamanho da região analisada, visando segregar geograficamente as séries temporais que ficam localizadas em alto mar das séries continentais, possivelmente delineando uma região costeira de transição. Ademais, faz sentido também que séries de latitudes mais baixas sejam diferentes de séries com latitudes mais altas. Consequentemente, a opção do número de *clusters* igual a 5 pareceu sensata, capaz de revelar padrões em 5 sub-regiões da macro região delimitada para o NE do Brasil. Vale ressaltar que a escolha do número de *clusters* deve de fato ser arbitrária e predeterminada pelo usuário conforme o que se espera encontrar. Este mesmo número foi utilizado na clusterização por repartição *k-Means* também (a ser apresentada na próxima seção 4.2).

Os dendrogramas de cada base de dados estão representados nas Figuras 57, 58 e 59, onde pode-se observar que cada um dos 5 *clusters* é colorido com uma cor distinta. As 72 séries da base NOAA são mais facilmente identificadas, diferentemente das séries ERA e MERRA, que são muitas – logo, seus dendrogramas não ficam totalmente legíveis na resolução das dimensões de uma folha A4. Em todo caso, o propósito é meramente evidenciar que a segregação se dá em determinada altura da árvore hierárquica, proporcionando uma noção da quantidade de séries temporais de cada base de dados que foram agrupadas em cada um dos 5 *clusters*.





Fonte: elaboração própria do autor



Figura 58 – Dendrograma das séries ERA após clusterização hierárquica aglomerante com método de aglomeração *average linkage* e com matriz dissimilaridade baseada na correlação de Pearson

Fonte: elaboração própria do autor

Figura 59 – Dendrograma das séries MERRA após clusterização hierárquica aglomerante com método de aglomeração *average linkage* e com matriz dissimilaridade baseada no índice adaptativo com q = 1,7



Fonte: elaboração própria do autor

4.2 CLUSTERIZAÇÃO POR REPARTIÇÃO K-MEANS E HÍBRIDA

A clusterização por repartição com o algoritmo *k-Means* foi executada por meio da função kmeans() do pacote de funções "stats" (R CORE TEAM, 2020) no RStudio. Como visto na seção 2.5.3, este método é um pouco mais direto e não demanda cálculo de outros tipos de dissimilaridade entre séries temporais diferentes de distâncias Euclidianas. É requerido, em contrapartida, que o usuário determine, *a priori*, o número *k* de *clusters*, e o algoritmo se encarrega de selecionar, aleatoriamente, centroides artificiais que buscarão se aproximar das séries temporais de mesmo *cluster* em um número predefinido de iterações, até que uma solução satisfatória seja encontrada.

Na clusterização híbrida, o processo é similar à por repartição *k-Means*, contudo, inicialmente é realizado um processo de clusterização hierárquica aglomerante com base na distância Euclidiana entre as séries temporais para a criação de um número *k* de *clusters* provisórios, cujos centroides são determinados pela média das séries temporais de cada *cluster*. Estes centroides alimentarão a segunda etapa de repartição *k-Means*. Todo este processo é simplificado pela função hkmeans() do pacote "factoextra" no RStudio (KASSAMBARA e MUNDT, 2020), sendo necessário indicar o número *k* de *clusters* e um método para a etapa da clusterização hierárquica, que deve ser exclusivamente aglomerante.

Novamente o número de *clusters* ficou predefinido como 5, pelas mesmas razões explanadas na seção anterior. Quanto ao método da etapa de clusterização hierárquica aglomerante, recorreu-se aos resultados da Tabela 5, onde é revelado, na coluna da distância Euclidiana, qual deles melhor se adere à cada base de dados. Para NOAA, ERA e MERRA observa-se que os melhores resultados foram com os métodos Ward.D2 (0,8160), *average linkage* (0,7742) e *centroid* (0,7296), respectivamente.

4.3 VALIDAÇÃO INTERNA DOS RESULTADOS DE CLUSTERIZAÇÃO

Três processos de clusterização foram realizados para os conjuntos de séries temporais de cada base de dados: hierárquica, por repartição *k-Means* e híbrida, cada qual com suas especificidades, a exemplo do cálculo da matriz dissimilaridade e método de aglomeração. Os resultados de cada um deles envolveram o agrupamento de subconjuntos de séries temporais

sempre em 5 *clusters*. Para avaliar qual dos três processos foi mais aderente aos dados, utilizouse o coeficiente de Dunn (seção 2.5.5) como métrica de comparação que, como já mencionado, deve ser o maior possível para denotar boa compactação de cada *cluster* individual e maior distância entre *clusters* distintos. Os resultados podem ser observados na Tabela 9.

Base de dados	Método de clusterização	Coeficiente de Dunn				
	Distância Euclidiana + DIANA	1,1487				
NOAA	k-Means	1,5421				
	Distância Euclidiana + Ward.D2 + k-Means	1,5421				
	Correlação de Pearson + average linkage	1,2545				
ERA	k-Means	1,2050				
	Distância Euclidiana + average linkage + k-Means	1,2050				
	Índice Adaptativo $q = 1,7 + average \ linkage$	1,6961				
MERRA	k-Means	1,4837				
	Distância Euclidiana + centroid + k-Means	1,4901				
Fonte: elaboração própria do autor						

Tabela 9 - Coeficiente de Dunn para os resultados de cada processo de clusterização

Para as séries temporais da NOAA tanto a clusterização por repartição *k-Means* quanto a híbrida retornaram o valor de Dunn equivalente a 1,5421, indicando que as duas soluções tenham possivelmente convergido para o ótimo. Quanto às séries de ERA e MERRA, ambas apresentaram maior Dunn (1,2545 e 1,6961 – respectivamente) na clusterização hierárquica aglomerante. Os resultados são insuficientes para apontar um algoritmo que seja ideal – este não é o objetivo da análise. Ressalta-se a importância de testar mais de um método em busca daquele que desempenhe de maneira mais satisfatória em relação aos demais. A visualização dos 5 *clusters* de cada base de dados pode ser observada na Figura 60.

Figura 60 – Resultados da clusterização das séries temporais de: NOAA por repartição *k-Means* (a); ERA hierárquica aglomerante (b); MERRA hierárquica aglomerante (c)



Fonte: elaboração própria do autor

O marcador das coordenadas de cada série temporal foi colorido de maneira a distinguir os *clusters*, mas sem intenção de classificá-los. Fica evidente que, genericamente, os agrupamentos resultantes possuem similaridades, permitindo extrair algumas conclusões acerca da disposição geográfica:

- Existe uma tendência à diferenciação dos pontos em alto mar daqueles continentais, constituindo *clusters* distintos;
- Os *clusters* apresentaram diferenciação entre regiões mais ao Norte ou ao Sul;
- Há a presença de um grande *cluster* continental à esquerda, que varre quase toda a amplitude latitudinal (pontos roxos em (a), laranjas em (b) e verdes em (c));
- Houve uma delimitação pertinente da área do NE em que se concentram a maior parte dos parques eólicos em operação no Brasil (pontos verdes em (a) e, embora compartilhando características de pontos em alto mar, pontos verdes em (b) e pontos vermelhos de (c)).

Não é sensato ponderar sobre o comportamento do vento em cada uma das regiões delimitadas, dado que as informações contidas em cada série temporal não foram ainda esmiuçadas. É prudente reconhecer, por ora, apenas que séries temporais de mesmo *cluster* compartilham de mesmas formas e/ou características.

Retomando o objetivo principal desta etapa de clusterização, foi necessário extrair de cada *cluster* uma série temporal representativa ao seu subconjunto, de maneira a reduzir o volume informacional para análise. Uma possibilidade seria utilizar o próprio centroide de cada *cluster*, porém os centroides resultantes da clusterização por repartição *k-Means* são artificiais – séries temporais geradas a partir de suas vizinhas de mesmo *cluster*. Alternativa seria extrair a média da região, sintetizando uma nova série temporal média, contendo traços de todas as outras. No entanto esta também seria artificial, com potencial de suavizar variações de interesse.

Optou-se, então, por calcular correlações de Pearson entre pares de séries temporais de mesmo *cluster* e aquelas que apresentaram maior média foram selecionadas para representar suas respectivas sub-regiões e prosseguir com a análise de séries temporais. Os resultados foram dispostos na Tabela 10. A numeração das séries faz referência à identificação proposta pelo próprio RStudio que, ao realizar a leitura dos dados, armazenou as séries temporais numeradas em uma tabela. Portanto a interpretação de cada item da Tabela 10 segue o exemplo da série de número E193 da base de dados ERA, que obteve alto grau de correlação média (0,9329) com todos os seus pares de mesmo *cluster* de número Q_{E4} e que fora colorido de roxo na Figura 60.

As localizações destas séries temporais foram plotadas e podem ser conferidas na Figura 61, coloridas conforme as cores dos seus respectivos *clusters*. Tais séries foram descritas em uma análise exploratória de dados no próximo capítulo.

Base de	Numeração das séries	Clu	ster a que	Média das correlações de Pearson em relação às outras	Localização	
dados re	representativas	entativas pertence		séries de mesmo <i>cluster</i>	Latitude	Longitude
	N55	$Q_{\rm N1}$	laranja	0,7936	- 12,381	- 35,625
	N38	Q_{N2}	verde	0,7228	- 8,571	- 37,5
NOAA	N15	$Q_{\rm N3}$	azul	0,7679	- 2,857	- 35,625
	N42	$Q_{\rm N4}$	roxo	0,8736	- 10,476	- 45
	N29	$Q_{\rm N5}$	vermelho	0,8663	- 6,667	- 39,375
ERA	E33	$Q_{\rm E1}$	verde	0,8981	- 2,25	- 40,5
	E135	Q_{E2}	laranja	0,8725	- 11,25	- 42,75
	E190	$Q_{\rm E3}$	azul	0,7920	- 15,75	- 39,375
	E193	$Q_{\rm E4}$	roxo	0,9329	- 15,75	- 36
	E212	Q_{E5}	vermelho	0,8769	- 18	- 43,875
	M139	Q_{M1}	azul	0,9248	- 2,5	- 40
	M211	Q_{M2}	laranja	0,7789	- 4	- 41,875
MERRA	M179	Q _{M3}	roxo	0,8747	- 3,5	- 46,25
	M581	Q_{M4}	verde	0,8333	- 11,5	- 45
	M622	Q _{M5}	vermelho	0,7734	- 12	- 35

Tabela 10 – Seleção de séries temporais representativas de cada *cluster*

Fonte: elaboração própria do autor

Figura 61 –	Localização	geográfica da	s séries	temporais	representativas
-------------	-------------	---------------	----------	-----------	-----------------



Fonte: elaboração própria do autor

5 ANÁLISE DE SÉRIES TEMPORAIS DE DADOS EÓLICOS

5.1 ESTATÍSTICA DESCRITIVA

Provenientes de modelos, os dados explorados pertencentes às séries temporais possuem incertezas e podem não refletir a realidade com fidelidade. Qualquer análise futura utilizando estes dados como entrada poderia ser enviesada, já que as incertezas são propagadas e até mesmo multiplicadas para outros resultados. Desta forma, cabe redobrar cautela ao inferir sobre os resultados obtidos dos dados analisados.

Estatísticas gerais (Tabela 11) e algumas representações das séries selecionadas foram preparadas para uma breve inspeção visual dos dados brutos mensais de velocidade do vento: amplitudes de variações, tendências e padrões aparentes, ciclos e frequências de ocorrências. Observa-se a plotagem das séries temporais e seus respectivos histogramas e diagramas de *boxplots* com variações mensais para as séries da NOAA, ERA e MERRA nas Figuras 62, 63 e 64, respectivamente.

Base de dados	Série temporal	Mínimo [m/s]	Máximo [m/s]	Média [m/s]	q _{0.25} [m/s]	q _{0.5} [m/s]	q _{0.75} [m/s]	<i>IQR</i> [m/s]
	N55	4,946	8,708	6,863	6,459	6,881	7,257	0,798
	N38	2,080	5,695	4,235	3,746	4,316	4,767	1,021
NOAA	N15	3,948	9,874	7,230	6,434	7,278	8,124	1,690
	N42	1,023	3,763	2,145	1,636	2,063	2,591	0,955
	N29	1,560	5,311	3,583	2,893	3,687	4,242	1,349
	E33	1,394	11,714	7,952	6,471	8,519	9,669	3,198
	E135	0,984	6,483	3,862	2,926	4,022	4,836	1,910
ERA	E190	0,992	4,750	2,906	2,485	2,920	3,319	0,834
	E193	1,735	8,618	4,990	4,259	4,945	5,691	1,432
	E212	0,323	4,519	2,485	2,004	2,442	2,926	0,922
	M139	4,559	11,895	8,562	7,040	8,736	10,078	3,038
MERRA	M211	2,862	6,629	4,744	4,066	4,685	5,458	1,392
	M179	1,639	3,451	2,397	2,052	2,382	2,750	0,698
	M581	3,030	7,348	4,797	3,963	4,663	5,687	1,724
	M622	4,974	9,973	7,122	6,647	7,051	7,559	0,912

Tabela 11 – Dados estatísticos gerais das séries selecionadas

Fonte: elaboração própria do autor

Os dados estatísticos da Tabela 11 apontam para algumas caracterizações gerais. Eles confirmam que pontos continentais (N38, N42, N29, E135, E190, E212, M211, M179 e M581) apresentam, em média, menores velocidades do vento em comparação com os pontos em alto

mar (N55, N15, E33, M139 e M622), com exceção do ponto E193 que, embora *offshore*, não apresenta média muito alta. Os quartis e valores de *IQR* apontam para algumas séries com alta variabilidade, a exemplo de E33 e M139 – ambas localizadas quase de maneira sobreposta (Figura 61), indício de que os modelos de reanálise ERA e MERRA representam o vento de maneira semelhante. O mesmo pode ser afirmado para os modelos NOAA e MERRA, vide séries N55 e M622 que estão localizadas bem próximas uma da outra e apresentam valores de quartis e *IQR* bem similares.



Figura 62 – Séries temporais NOAA, seus histogramas e boxplots

Fonte: elaboração própria do autor

As séries da NOAA são mais longas, fazendo com que suas plotagens sejam relativamente sobrecarregadas de informação e de difícil interpretação. Porém é possível visualizar flutuações diferentes daquelas mês a mês e uma leve tendência descendente para abaixo da média de longo prazo (linhas vermelhas sólidas e verdes tracejadas, respectivamente). Quanto à magnitude da velocidade do vento, as séries N15 e N55 apresentam maiores valores, ambas com médias ao redor de 7 m/s, fato justificado pela localização em alto mar. N38 apresenta valores maiores que N29, mas se diferenciam por pouco e ficam abaixo de N15 e

N55. A série N42 apresenta valores inferiores em geral, dada sua localização mais continental, onde os ventos alísios já perderam força.

Já os histogramas revelam que as distribuições das frequências de ocorrências são diferentes para cada série e não convergem para um mesmo formato, algumas com assimetrias positivas, outras negativas. Os *boxplots* indicam que todas as séries possuem um ciclo anual bem definido, com maiores valores nos meses de agosto, setembro e outubro, e menores valores nos meses de fevereiro até maio. As séries N38 e N29, localizadas mais próximas dos parques eólicos no Brasil, possuem menores *IQR* mensais justamente nos meses de maiores velocidades do vento, algo que pode ser benéfico para menores variações na geração pela fonte eólica.



Figura 63 – Séries temporais ERA, seus histogramas e boxplots

Fonte: elaboração própria do autor

As plotagens das séries ERA revelam convergência para a ocorrência de tendências positivas (linha vermelha sólida superando a linha verde tracejada). Também é possível verificar visualmente que existem flutuações em escalas temporais superiores a um ano. Alguns histogramas (E190, E193 e E212) convergem para uma forma comum, se assemelhando a uma

curva Gaussiana. Ao estudar os *boxplots*, percebe-se que nem todas as séries possuem o ciclo anual tão marcante como ocorre com as séries NOAA. Das 5 séries ERA, 3 estão localizadas em latitudes mais a sul, denotando baixas velocidades médias do vento e ciclo anual atenuado. Vale ressaltar que as localizações das séries representativas dos *clusters* resultantes do modelo ERA não coincidiram com as localizações de maior concentração de parques eólicos em operação no NE do Brasil.



Figura 64 – Séries temporais MERRA, seus histogramas e boxplots

Fonte: elaboração própria do autor

As séries do MERRA M139 e M622, ambas localizadas em alto mar (*offshore*), apontam para maiores valores de velocidade do vento, ambas com tendência (linha sólida vermelha) levemente ascendente em relação à média (linha verde tracejada). Todas as séries apresentam amplitude de variação vertical consistente, sendo possível observar o ciclo anual nas próprias plotagens, algo também aparente nos diagramas de *boxplots* e que confirmam maiores magnitudes no segundo semestre em detrimento dos meses do primeiro semestre, similar aos *boxplots* das séries da NOAA. Todavia os padrões não são tão convergentes, já que as séries M581 e M622 possuem maiores valores no mês de julho, enquanto as outras três nos meses de outubro ou setembro. Novamente os histogramas não convergem para um padrão similar, o que sugere que os modelos captam variações regionais da velocidade do vento.

5.2 EVENTOS EXTREMOS: ESCASSEZ DE RECURSOS EÓLICOS

5.2.1 Análise de frequência empírica

Na hidrologia, os efeitos de períodos de vazão reduzida e/ou seca (Figura 65) podem ser significativos, capazes de assolar a sociedade e o meio ambiente de diversas formas, sendo, portanto, pertinente estudá-los para melhor planejamento de reservatórios, redes de abastecimento etc. Analogamente, a análise de valores extremos em séries temporais de dados eólicos com vistas para a ocorrência de mínimos em escalas temporais que abrangem meses até anos é objeto de relevância para dimensionamento da capacidade de geração pela fonte eólica a nível nacional.



Figura 65 – Diferentes abordagens para definição de eventos de vazão reduzida

Fonte: Salas et al. (2019)

Uma possível definição para o termo "vazão reduzida" se refere à vazão mínima de um rio ao longo de um número consecutivo de dias (intervalo) que pode ocorrer num dado período unitário maior. A vazão reduzida depende, por conseguinte, de um intervalo e de um período unitário específicos e predefinidos (Figura 65a), com valores típicos de 7 dias para o intervalo e 1 ano para o período unitário. Outra abordagem preocupa-se com a duração da vazão reduzida, que pode ser definida como o número de períodos consecutivos (ex: dias) em que a vazão de um rio figura abaixo de um certo patamar ou valor de referência, associando-se com déficit cumulativo de volume de água. Neste caso, a intensidade do evento de vazão reduzida será

definida pela razão entre o volume acumulado e a duração do evento (Figura 65b). Devido à natureza aleatória de fenômenos deste tipo, os métodos para análise são baseados em conceitos probabilísticos e estocásticos (SALAS et al., 2019).

Metodologia similar já foi explorada para identificação de períodos de diferentes durações em que a intensidade da velocidade do vento fica abaixo de um determinado patamar. As escalas temporais exploradas abrangeram, contudo, durações de horas até dias, mas não meses ou anos, com intenção de auxiliar a operação de curto prazo de um sistema elétrico alimentado por energia eólica que apresenta variabilidade no curto prazo (LEAHY e MCKEOGH, 2013; PATLAKAS et al., 2017).

A presente pesquisa buscou desenvolver um caminho dedicado à análise de valores extremos levando em consideração que os dados da NOAA, ERA e MERRA utilizados são mensais e requerem adaptações às metodologias supracitadas. Assim, as séries temporais originais foram analisadas com uma janela temporal móvel de comprimento igual a 3, extraindo o valor mínimo dentre os 3 valores compreendidos pela janela, segregando um total de 10 valores por ano para cálculo de médias anuais. Este procedimento gera uma nova série temporal S_{min} composta por valores médios anuais y_j , com comprimento equivalente ao número de anos compreendidos pela série original S_T .

$$S_T = \{x_{t=1}, x_{t=2}, x_{t=3}, \dots, x_{t=n}\}$$
(70)

$$y_{j_{t}} = \min[x_{t-1}, x_{t}, x_{t+1}] \begin{cases} t = 2, 3, ..., 11 & quando \ j = 1 \\ t = 13, 14, ..., 23 & quando \ j = 2 \\ \vdots & \\ t = n - 10, n - 9, ..., n - 1 & quando \ j = \frac{n}{12} \in \mathbb{Z} \end{cases}$$
(71)

$$y_j = \frac{y_{j_t} + y_{j_{t+1}} + \dots + y_{j_{t+9}}}{10}, \quad onde \ t = 2, 14, 26, \dots, n-10$$
(72)

$$S_{min} = \left\{ y_{j=1}, y_{j=2}, \dots, y_{j=\frac{n}{12}} \right\}$$
(73)

Em posse das séries de mínimos, uma maneira de se assessorar estatisticamente a probabilidade de ocorrência de determinados patamares de velocidade do vento para um número de anos $N = \frac{n}{12}$, consiste em calcular a distribuição empírica de frequência (ou função de distribuição acumulada CDF) dos valores das séries S_{min} sortidos em ordem crescente. A probabilidade de não excedência F(j) de um determinado valor y_j pode ser utilizada para calcular o seu período esperado de retorno T_r :

$$F(j) = Pr\left[Y \le y_j\right] = \frac{j}{(N+1)}, \quad j = 1, 2, ..., N$$
(74)

$$T_r(y_j) = \frac{1}{\Pr\left[Y \le y_j\right]} \tag{75}$$

Esta sequência de etapas foi executada nas 5 séries temporais de cada base de dados, derivando séries S_{min} que compreendem médias anuais considerando os valores mínimos. A função de distribuição empírica acumulada CDF foi traçada (Figura 66) e intervalos de 10, 20, 30, 40, 50 e 100 anos foram considerados como valores esperados de tempo de retorno para eventos cuja velocidade média anual do vento não supere o valor encontrado pela análise. Foram determinados patamares de velocidade média do vento intrínsecos a estas condições, comparando-os com as médias históricas de cada série temporal original, cujos valores já foram dispostos na Tabela 11.



Figura 66 - Função de distribuição empírica acumulada

Fonte: elaboração própria do autor

Basa da	Sária	Tempo de retorno e probabilidade de não excedência					
dados	temporal	10 anos	20 anos	30 anos	40 anos	50 anos	100 anos
dudos	temporta	10 %	5 %	3,33 %	2,5 %	2 %	1 %
	N55	6,253 ↓9%	6,142 <mark>↓11%</mark>	6,119 _{↓11%}	6,088 <mark>↓11%</mark>	5,981 _{13%}	5,892 _{↓14%}
	N38	3,518 _{↓17%}	3,433 _{19%}	3,297 _{122%}	3,225 _{↓24%}	3,213 _{↓24%}	3,187 _{125%}
NOAA	N15	6,444 _{↓11%}	6,307 _{↓13%}	6,186 _{↓14%}	6,152 _{↓15%}	6,146 _{15%}	6,046 <mark>↓16%</mark>
	N42	1,677 _{↓22%}	1,645 _{↓23%}	1,621 _{↓24%}	1,600 _{↓25%}	1,593 _{↓26%}	1,557 _{↓27%}
	N29	2,857 _{↓20%}	2,743 _{↓23%}	2,706 _{↓24%}	2,702 _{125%}	2,690 _{↓26%}	2,629 _{↓27%}
	E33	5,914 _{↓18%}	5,667 _{121%}	5,447 _{↓24%}	5,398 _{125%}	5,360 _{126%}	5,240 _{127%}
	E135	2,957 _{↓42%}	2,912 _{↓43%}	2,881 _{↓43%}	2,871 _{↓43%}	2,860 _{↓44%}	2,723 _{146%}
ERA	E190	2,070 _{↓41%}	2,022 _{↓42%}	1,992 _{↓43%}	1,972 _{↓44%}	1,969 _{↓44%}	1,923 _{145%}
	E193	3,652 _{138%}	3,461 _{141%}	3,422 _{↓42%}	3,417 _{↓42%}	3,412 _{↓42%}	3,342 _{↓43%}
	E212	1,675 _{↓64%}	1,646 _{165%}	1,619 _{↓66%}	1,597 _{↓66%}	1,568 _{167%}	1,527 _{168%}
	M139	6,941 _{19%}	6,778 <mark>↓21%</mark>	6,741 _{121%}	6,711 _{↓22%}	6,669 _{↓22%}	6,585 _{123%}
	M211	3,986 _{↓16%}	3,906 <mark>_{18%}</mark>	3,875 _{18%}	3,857 _{19%}	3,840 _{19%}	3,805 _{120%}
MERRA	M179	2,033 _{15%}	2,019 <mark>_{16%}</mark>	2,017 _{16%}	2,014 <mark>_{16%}</mark>	2,007 _{16%}	1,994 _{↓17%}
	M581	4,224 <mark>↓12%</mark>	4,179 <mark>↓13%</mark>	4,138 <mark>,14%</mark>	4,113 _{↓14%}	4,088 <mark>↓15%</mark>	4,039 _{16%}
	M622	6,457 _{↓9%}	6,347 _{↓11%}	6,328 _{↓11%}	6,312 _{↓11%}	6,292 _{↓12%}	6,252 _{↓12%}

Tabela 12 – Média anual de velocidade do vento [m/s] associada a diferentes intervalos de tempo de retorno e probabilidades de não excedência e a porcentagem relativa à média histórica da série original

Fonte: elaboração própria do autor

Os valores dispostos na Tabela 12 e as curvas CDF da Figura 66 apresentam a mesma informação de maneiras distintas. O objetivo não é comparar uma série contra outra, mas sim esclarecer que eventos extremos de baixa velocidade anual média possuem maior tempo de retorno, o que significa que podem ocorrer com uma frequência mais baixa. Cabe comparação com as médias totais calculadas para as séries temporais originais (vide Tabela 11), pois os valores associados aos tempos de retorno são apenas uma fração do que é indicativo para a velocidade média histórica esperada naquela localização.

Estas porcentagens em vermelho são importantes no sentido de estender os resultados destas séries temporais à região delimitada pelos *clusters* a que pertencem e posteriormente desenvolver cenários de geração com base nestes valores. Tomando o tempo de retorno de 20 anos como exemplo, a série N38 apresentaria probabilidade de 5% de não exceder o valor anual médio de 3,433 – valor ~19% mais baixo do que a média total calculada para o período de 1851 a 2014.

Caso este desempenho aconteça a cada 20 anos e caso seja verificado nas demais séries de mesmo *cluster* Q_{N2} , a geração nos parques eólicos localizados nesta região seria drasticamente afetada. De mesma forma, para o tempo de retorno de 50 anos a série M622 teria 2% de probabilidade de ter seu valor médio anual abaixo de 6,292 – valor ~12% inferior à
média total calculada para o período de 1980 a 2017. Este valor de referência, se estendido ao seu *cluster* Q_{M5}, indicaria uma geração por recursos eólicos um pouco menos comprometedora (comparada ao caso da série N38), ainda assim são aspectos a serem levados em consideração em se tratando do planejamento da capacidade de geração a nível nacional.

Outro caminho envolveria o ajuste de uma PDF adequada para cada série S_{min} de maneira a explorar as distribuições probabilísticas dos valores compreendidos a fim de calcular a probabilidade de ocorrência de um determinado patamar de velocidade do vento. Em se tratando de dados com longa cobertura temporal, inicialmente foi necessário plotar os histogramas destas séries e suas curvas de densidade suavizada para rápida inspeção visual em busca de uma PDF facilmente reconhecível (Figura 67).



Figura 67 - Histograma e função densidade suavizada

Fonte: elaboração própria do autor

Comumente, estudos hidrológicos se valem de modelos paramétricos como Weibull, log-Pearson, log-normal ou distribuição generalizada de valores extremos, com dois ou três parâmetros, para representar as distribuições probabilísticas de séries de vazão (SALAS et al., 2019). Contudo, estes modelos são alimentados por dados diários ou em outras escalas temporais menores do que mensal para séries de vazão reduzida.

Para velocidade do vento, a PDF mais usualmente utilizada é a Weibull de dois parâmetros (fator de escala, relacionado à velocidade média do vento no local, e fator de forma, relacionado à variância da velocidade do vento em torno da velocidade média). Dados horários são ideais para este ajuste. Há de se considerar que são modelos propícios para uso quando não se conhece o comportamento histórico do vento no local (CUSTÓDIO, 2013).

Pela Figura 67, fica evidente que não é trivial identificar uma única função bem conhecida que possa se ajustar à todas as diferentes distribuições descritas. Alguns histogramas e suas respectivas curvas de densidade probabilística até se assemelham ao modelo paramétrico Gaussiano, outras se aproximam de uma curva de Weibull. Porém não constituem formatos característicos, não são livres de assimetrias (positiva ou negativa) e julgou-se que o ajuste de uma PDF traria mais incerteza à interpretação das informações sobre recursos eólicos

5.2.2 Análise mês a mês de ocorrência de eventos extremos

Foi pertinente inspecionar as séries representativas de cada *cluster* em busca de eventos considerados extremos, que possam denotar uma escassez da disponibilidade de recursos eólicos. Questão ainda não sedimentada é o limiar ou referência abaixo do qual remeteria uma disponibilidade reduzida para geração. A importância do estabelecimento de um limiar vai na direção de facilitar o monitoramento mês a mês do desempenho eólico. É sabido que os aerogeradores possuem uma velocidade mínima de entrada, ou *cut-in*, para dar partida no giro das pás e, consequentemente, gerar energia elétrica. Todavia, este valor faz mais sentido se considerado para o curto prazo, quando se requer explorar a intermitência do vento. Alguns trabalhos caminham nesse sentido, estudando eventos de escassez que duram entre horas e alguns dias (LEAHY e MCKEOGH, 2013; PATLAKAS et al., 2017).

Eleger um valor único do tipo limiar universal para definir com exatidão a entrada em um período de escassez de recursos eólicos seria incurial. Os resultados da análise exploratória de dados apresentados na seção 5.1 abordam justamente o fato de que em localidades diferentes o vento se comporta de maneira específica e tem uma variabilidade inerente, podendo apresentar uma maior ou menor amplitude de médias mensais ao longo do ano. Tal fato sugere que a definição de um limiar deve respeitar a especificidade local.

Nesse sentido, a possibilidade de utilizar a média de longo prazo da própria série temporal como referência para indicar se o vento está acima ou abaixo deste valor tampouco é uma estratégia adequada. Como visto nas seções anteriores, o vento que sopra no NE possui um ciclo anual bem definido, com médias mensais mais altas no segundo semestre em relação ao primeiro. Sendo o caso, provavelmente o suposto indicador de escassez eólica acusaria a ocorrência de um período de baixa disponibilidade por volta dos meses de março, abril e maio. Ademais, as variabilidades de cada mês (representadas pelos valores de *IQR*) não são homogêneas, restringindo à opção de considerar limiares distintos para cada mês de cada diferente série temporal.

No entanto, analisar mês a mês a variação e ocorrência de períodos de escassez de maneira fragmentada (com limiares e valores médios de cada mês específico, como ilustra a Figura 68) pode ser impraticável, no sentido em que não se verificaria uma sequência de meses em que a velocidade do vento ficou abaixo (ou acima) do esperado. As informações devem ser posteriormente recompostas para determinar as durações dos eventos extremos, por exemplo.



Figura 68 – Análise fragmentada mês a mês da série N55

Fonte: elaboração própria do autor

Alternativa a esta inspeção desagregada envolve desenvolver um limiar móvel que percorra toda a série temporal, respeitando tanto a média quanto a variabilidade de cada mês, formando, ao invés de uma reta simples, uma curva harmônica (aproximadamente senoidal e similar ao ciclo anual), como exemplifica a Figura 69.





Fonte: elaboração própria do autor

O posicionamento deste limiar harmônico ao longo do plano cartesiano permite guiar a identificação dos meses considerados de baixa disponibilidade de recursos eólicos. Detalhe fundamental é que o seu deslocamento para cima ou para baixo em relação ao eixo das abscissas determina quais valores de referência para cada mês serão de fato utilizados para caracterizá-lo como escasso ou não. No caso da Figura 69, a definição e posicionamento do limiar harmônico se deu pelo cálculo do ciclo anual médio – determinado pelas médias de cada mês (média de todos os janeiros, de todos os fevereiros etc.) – com subtração de duas unidades dos valores de desvio padrão do mês correspondente (desvio padrão de todos os janeiros, de todos os fevereiros etc.). Este padrão se repete harmonicamente ao longo de toda a série.

De certo modo, a escolha da subtração de duas unidades de desvio padrão foi arbitrária, já que não existe ainda um consenso sobre o que configura um período de escassez eólica. Ela foi minimamente pautada pelo estudo da ocorrência do evento extremo de escassez eólica nos EUA que durou alguns meses de 2015, com médias de velocidade do vento de duas a três unidades de desvio padrão abaixo da média histórica. A geração pela fonte eólica foi severamente prejudicada na ocasião (LLEDÓ et al, 2018).

Como realizado neste estudo de Lledó et al. (2018), maneira relativamente mais simples de verificar o histórico de períodos de escassez é transformar as séries temporais pelo procedimento de normalização, com vistas para a remoção das características de não estacionariedade, como tendência e ciclo anual. Desta forma, os valores se transformam em anomalias centradas ao redor de zero, revelando os intervalos de tempo em que eles ficaram anomalamente acima ou abaixo do esperado, em unidades de desvio padrão. O limiar é então representado como uma reta paralela ao eixo das abscissas, tal qual é feito nos estudos de hidrologia, com aspecto visual simplificado que também permita monitoramento do desempenho do vento mês a mês.

Por esta rota, o essencial é registrar os intervalos de tempo em que houve escassez eólica para posteriormente revisitar a série original nos mesmos instantes e analisar a sequência de períodos escassos e seus respectivos valores. Para esta pesquisa, foram considerados três diferentes valores de limiares: - 1,5 unidades de desvio padrão, denotando baixa disponibilidade de recursos eólicos em intensidade branda; - 2 unidades de desvio padrão, para eventos de escassez moderada; - 2,5 unidades de desvio padrão, para indicar eventos de escassez severa. Uma rápida inspeção visual na plotagem das séries temporais normalizadas (contendo as anomalias de velocidade do vento em função do tempo) para os dados NOAA, ERA e MERRA e os três limiares predefinidos, é suficiente para observar a ocorrência histórica de eventos nestas três escalas de intensidade (Figura 70).

Pela Figura 70, observa-se que todas as séries tiveram o limiar de 1,5 rompido por diversas vezes, mesmo sendo as séries do MERRA mais curtas em relação às demais. O limiar de 2,0 foi ultrapassado em todas as séries, com exceção de M179 e M581, obviamente em uma quantidade de vezes menor do que o limiar de 1,5. Já o limiar de 2,5 acusou ocorrência de escassez severa em um menor número de vezes, configurando de fato um evento mais raro. Com exceção da série E135, todas as séries da NOAA e ERA apresentaram a superação desta referência de 2,5 unidades de desvio padrão. Em contrapartida, considerando as séries MERRA, apenas a de número M622 teve esse limiar rompido, já que o recorte temporal de 1980 a 2017 das séries do MERRA pode ser considerado muito curto para contemplar ocorrência de eventos dessa severidade. Em razão da baixa resolução da Figura 70, contudo, não é cabível indicar em quais meses de quais anos se deram estes rompimentos dos limiares, sendo esta informação de grande valia para a análise de ocorrências de períodos de escassez de recursos eólicos no NE do Brasil.



Figura 70 - Séries temporais normalizadas e a ocorrência de eventos de escassez



Prosseguiu-se, desta forma, para a tabulação (Tabela 13) dos instantes em que ocorreram estes eventos identificados visualmente. À medida em que se aumenta o limiar de 1,5 para 2 e depois para 2,5, reduz-se o número de ocorrências de seus rompimentos. Verifica-se pela escala de intensificação da coloração vermelha que, geralmente, a ocorrência é muito frequente nos meses de março, maio e, principalmente, abril. Mesmo para o limiar de 1,5 unidades de desvio padrão, quase não houve ocorrência destas reduções na intensidade do vento nos principais meses para a geração eólica (entre julho e outubro). Tal fato implica que as chances de ocorrer escassez de recursos eólicos no primeiro semestre são maiores do que no segundo semestre de cada ano – informação importante para o setor eólico.

Base de	Série		Núm	ero de	ocorrêi	ncias al	baixo	do lin	niar de	1,5 un	idades	do des	svio pad	lrão
dados	temporal	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
	N55	1	10	21	51	32	9	1	0	0	0	0	0	125
	N38	2	21	50	57	32	3	0	0	0	0	0	0	165
NOAA	N15	4	22	53	64	24	0	0	0	0	0	0	0	167
	N42	0	0	7	17	31	0	0	0	0	0	0	0	55
	N29	2	24	56	65	15	0	0	0	0	0	0	0	162
	E33	2	29	45	53	14	0	0	0	0	0	0	0	143
	E135	18	31	36	4	0	0	0	0	0	2	7	21	119
ERA	E190	1	1	20	23	17	11	4	1	0	2	2	3	85
	E193	1	5	20	27	7	2	0	2	1	6	9	3	83
	E212	6	7	16	5	13	9	0	0	0	1	6	8	71
	M139	0	4	11	14	6	0	0	0	0	0	0	0	35
	M211	0	1	1	12	10	1	0	0	0	0	0	0	25
MERRA	M179	0	0	3	14	10	0	0	0	0	0	0	0	27
	M581	2	2	4	1	0	0	0	0	0	0	0	0	9
	M622	0	3	10	11	3	0	0	0	0	0	0	0	27
Base de	Série		Nún	nero de	ocorré	èncias a	abaixo	do lii	niar de	2 uni	dades	do desv	vio padı	ão
dados	temporal	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
	N55	0	4	7	26	12	2	0	0	0	0	0	0	51
	N38	0	4	16	25	10	0	0	0	0	0	0	0	55
NOAA	N15	0	2	7	23	8	0	0	0	0	0	0	0	40
	N42	0	0	0	0	0	0	0	0	0	0	0	0	0
	N29	0	0	5	20	1	0	0	0	0	0	0	0	26
	E33	0	7	17	28	7	0	0	0	0	0	0	0	59
	E135	1	6	7	1	0	0	0	0	0	0	2	1	18
ERA	E190	0	0	9	12	7	4	0	0	0	0	0	0	32
	E193	0	0	3	6	2	1	0	1	0	1	2	2	18
	E212	2	0	1	1	3	2	0	0	0	0	1	0	10
	M139	0	0	2	6	2	0	0	0	0	0	0	0	10
	M211	0	0	0	0	1	0	0	0	0	0	0	0	1
MERRA	M179	0	0	0	0	0	0	0	0	0	0	0	0	0
	M581	0	0	0	0	0	0	0	0	0	0	0	0	0
	M622	0	0	3	2	0	0	0	0	0	0	0	0	5
Base de	Série		Núm	ero de	ocorrêi	ncias al	oaixo	do lin	niar de l	2,5 un	idades	do des	svio pač	lrão
dados	temporal	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
	N55	0	1	1	7	0	0	0	0	0	0	0	0	9
	N38	0	0	2	10	0	0	0	0	0	0	0	0	12
NOAA	N15	0	0	2	2	0	0	0	0	0	0	0	0	4
	N42	0	0	0	0	0	0	0	0	0	0	0	0	0
	N29	0	0	0	0	0	0	0	0	0	0	0	0	0
	E33	0	0	1	4	3	0	0	0	0	0	0	0	8
	E135	0	0	0	0	0	0	0	0	0	0	0	0	0
EKA	E190	0	0	2	1	1	0	0	0	U	0	0	0	4
	E193	0	0	0	4	0	0	0	0	0	0	0	1	5
	E212	1	0	1	0	1	1	0	0	0	0	0	0	4
	M139	U	0	0	0	0	U	U	0	U	U	U	0	U
	M170	U	0	0	0	0	U	U	0	U	U	U	0	U
MEKKA	M1/9	U	0	0	0	0	U	U	0	U	U	U	0	0
	M222	U	0	1	1	0	U	0	U	U	0	0	U	0
	IVIO22	U	U	1	1	U	U	U	U	U	0	0	U	2

Tabela 13 - Ocorrências mês a mês de média de velocidade do vento abaixo dos limiares de referência

Fonte: elaboração própria do autor

As séries temporais da base de dados NOAA, em relação às da ERA e MERRA, apresentaram um número superior de ocorrências de eventos em que a média de velocidade do vento ficou abaixo dos limiares em unidades do desvio padrão, o que pode ser justificado pelos seus comprimentos também superiores, que percorrem a linha do tempo de maneira retrospectiva até o ano de 1851. O que se nota é uma concentração deste rompimento dos limiares nos meses do primeiro semestre (de janeiro a junho – sobretudo nos meses de fevereiro, março, abril e maio), com praticamente nenhuma ocorrência nos meses de julho a dezembro.

Os resultados para as séries ERA também apresentam esta concentração nos meses do primeiro semestre, no entanto algumas séries, sobretudo E135, E193 e E212, obtiveram variações abaixo dos limiares de 1,5 e 2 unidades de desvio padrão nos meses de outubro, novembro e dezembro. Séries do MERRA, por sua vez, apresentaram um menor número de eventos abaixo dos três limiares considerados. São séries mais curtas que as do NOAA e ERA, contemplando apenas de 1980 até 2017. Por exemplo apenas por duas vezes o limiar de 2,5 fora superado, ambas para a séries M622. Mas para os outros limiares, a proporção de ocorrências segue compatível com a proporcionalidade dos comprimentos relativos das séries temporais analisadas.

De maneira geral, estes resultados corroboram com a análise prévia dos *boxplots* mensais de cada série temporal, que apontam para uma maior variabilidade da velocidade do vento nos meses de janeiro a junho, independentemente da localização geográfica específica dentro da área nordestina estudada. Consequentemente, cabe inferir que, caso eventos de escassez de recursos eólicos venham a ocorrer no futuro, as chances de que ocorram nos meses do primeiro semestre são maiores

5.2.3 Análise ano a ano de ocorrência de eventos extremos

Seguindo o objetivo de explorar períodos de escassez de recursos eólicos associados à variabilidade interanual, a etapa posterior consistiu no mapeamento do número de ocorrências abaixo dos limiares num mesmo ano. Para tal, adaptou-se a abordagem utilizada nos estudos de hidrologia em que as durações dos eventos de escassez ganham mais atenção – em função das implicações, não apenas energéticas, mas também sociais, que uma seca hídrica prolongada pode acarretar. No caso da hidrologia, a metodologia supracitada considera períodos consecutivos como um único evento de seca. Se porventura este evento teve uma interrupção

com apenas um período em que a vazão ficou acima do limiar após o qual retornou para abaixo do mesmo limiar, a abordagem entenderá como eventos distintos de seca, mesmo que tenham ocorrido num mesmo ano hidrológico.

Diferentemente, a presente pesquisa preocupou-se apenas com o número total de períodos (meses) abaixo do limiar que se deram num mesmo ano. Entende-se que a geração pela fonte eólica pode ser impactada de qualquer forma, sejam os períodos com desempenho abaixo do limiar consecutivos ou não.

Apenas o limiar de 1,5 unidades de desvio padrão foi utilizado neste momento, já que qualquer evento de baixa disponibilidade de recursos eólicos tabulado para os outros limiares de 2 e de 2,5 também estarão contemplados nesta seleção para avaliação. A intenção é evidenciar os anos que mais apresentaram eventos de superação do limiar. Outro critério adotado foi que os anos em que o número de meses abaixo do limiar de 1,5 ficou entre 0 e 2 foram ignorados – partindo da premissa de que é um número insuficiente para denotar um desempenho anual abaixo do esperado, principalmente porque nos principais meses da geração eólica (entre julho e outubro) a frequência de ocorrência foi bastante reduzida. Do contrário, caso a frequência de ocorrência fosse mais alta no mês de setembro, por exemplo, seria perfeitamente plausível que apenas um mês que ficasse muito abaixo de um desses limiares de referência faria com que a geração pela fonte eólica fosse impactada negativamente.

Os resultados foram alocados na Tabela 14. Reforça-se que nem todos os anos foram elencados – apenas anos a partir de 1900 que apresentaram ocorrência de 3 ou mais meses abaixo do limiar por ao menos uma das séries analisadas, sendo que não necessariamente estas 3 ocorrências foram consecutivas. As células da tabela em que este número é equivalente ou superior a 3 foram coloridas de vermelho em escala crescente; quanto maior o número de ocorrências, mais forte a tonalidade do vermelho. As células em branco não foram removidas por servirem o propósito de contrastar com os resultados de interesse.

Algumas séries quase não indicam ocorrência de mais de dois meses abaixo do limiar de 1,5 (ex: N42, E190, E193, E212, M581 e M622). Pelo princípio da precaução, não cabe inferir que, de modo geral, os *clusters* a que estas séries fazem parte não passaram por eventos de escassez, apenas é constatável que as localizações específicas destas séries podem de fato não ter apresentado variabilidade suficiente para denotar anos de escassez. Outrossim, as séries que tiveram 3 ou mais meses abaixo do limiar em anos recorrentes podem ter passado por diversos eventos de escassez – resultado tampouco certeiro para seus *clusters*.

A			NOAA					ERA					MERRA	A	
Ano	N55	N38	N15	N42	N29	E33	E135	E190	E193	E212	M139	M211	M179	M581	M622
1907	0	0	0	0	0	0	0	1	4	0	-	-	-	-	-
1910	0	0	0	0	0	3	2	1	1	1	-	-	-	-	-
1911	0	0	0	0	0	1	0	1	3	0	-	-	-	-	-
1913	1	0	3	0	1	4	0	2	4	3	-	-	-	-	-
1916	1	3	5	2	4	3	4	0	0	0	-	-	-	-	-
1917	0	3	4	0	4	4	2	0	0	0	-	-	-	-	-
1919	0	0	2	0	3	0	0	2	1	1	-	-	-	-	-
1921	3	4	4	0	4	4	2	0	0	0	-	-	-	-	-
1922	1	2	0	0	1	1	3	1	2	2	-	-	-	-	-
1923	2	1	2	0	2	3	2	2	2	4	-	-	-	-	-
1924	2	4	4	1	4	4	4	3	2	0	-	-	-	-	-
1926	l	3	2	l	2	3	2	2	1	1	-	-	-	-	-
1927	0	0	0	0	0	2	4	2	2	2	-	-	-	-	-
1928	1	0	1	0	1	2	1	3	1	1	-	-	-	-	-
1929	2	1	3 1	2	0	3 2	3	0	2	0	-	-	-	-	-
1933	3	1	3	0	2	1	4	0	1	0	-	-	-	-	-
1935	1	2	4	1	2	- - 	3	1	2	1	_	_	_	_	_
1937	0	0	1	0	1	3	2	2	1	1	_	_	_	-	_
1940	1	3	2	1	2	0	3	0	0	0	-	-	-	-	-
1943	0	0	0	0	0	0	1	3	1	0	-	-	-	-	-
1944	1	0	0	3	0	3	0	0	0	0	-	-	-	-	-
1945	1	3	4	0	3	2	2	0	1	0	-	-	-	-	-
1949	0	0	0	0	0	0	1	1	2	3	-	-	-	-	-
1961	0	1	3	0	2	3	3	0	0	2	-	-	-	-	-
1963	0	2	3	1	3	3	2	2	2	0	-	-	-	-	-
1964	1	4	4	0	3	2	0	0	0	1	-	-	-	-	-
1965	3	2	2	0	1	2	3	2	2	2	-	-	-	-	-
1967	3	3	1	2	2	3	2	0	0	0	-	-	-	-	-
1968	3	3	2	1	3	3	0	1	0	0	-	-	-	-	-
1971	5	4	3	1	3	3	4	2	1	1	-	-	-	-	-
1972	1	2	1	1	1	0	0	5	1	2	-	-	-	-	-
1975	2 1	4	4	2	3	2	1	1	2 2	4	-	-	-	-	-
1974	4	4	3	1	4		2	0	2 1	1	-	-	-	-	-
1984	5	3	2	0	2	3	0	0	1	0	3	2	3	0	- 1
1985	2	4	3	2	3	4	2	0	0	1	4	3	0	0	0
1986	1	4	3	2	3	3	0	0	0	0	3	1	1	0	0
1988	2	3	1	1	3	2	1	2	1	1	3	2	2	0	3
1989	1	3	3	1	2	2	2	0	0	0	3	2	1	0	1
1995	2	3	2	3	4	1	2	0	0	0	2	1	1	0	1
1999	0	0	0	1	0	0	0	0	0	0	0	0	3	0	1
2000	0	4	1	1	3	3	0	0	0	1	1	0	1	1	0
2004	1	3	0	0	1	0	1	0	0	1	0	1	0	0	0
2008	1	2	2	1	2	2	2	0	0	1	2	3	3	1	1
2009	1	2	2	3	2	3	2	1	1	1	2	1	2	1	2

Tabela 14 – Ocorrências por ano de meses com média de velocidade do vento abaixo do limiar de 1,5

Fonte: elaboração própria do autor

A presença de várias células avermelhadas em uma mesma coluna sinaliza que a localização geográfica da série temporal em questão pode ser propensa a eventos de escassez (caso das séries N38, N25, N29, E33, E135 e possivelmente M139) – conclusão que cautelosamente pode ser estendida às demais séries temporais vizinhas de mesmo *cluster*. Três séries (N55, N15 e E33) apresentaram o número máximo de 5 meses abaixo do limiar em um mesmo ano. Várias outras séries apresentaram este número igual a 4, indicando que os anos em que ocorreram podem ter tido desempenho bem aquém do esperado. Já as linhas da Tabela 14 com várias células avermelhadas sugerem convergência entre os modelos de reanálise e apontam para anos abaixo do limiar em mais de uma sub-região do NE, como é o caso dos anos 1921, 1924, 1974, 1984 e 1985.

O período de 1971, 1973, 1974 e 1975 chama bastante a atenção, com forte sinal de redução da intensidade do vento em várias regiões com duração multianual prolongada. A existência de anos de uma mesma década próximos uns dos outros com coloração vermelha bem como a ausência de anos de determinadas décadas remetem à presença de variabilidade decadal nos recursos eólicos – anos das décadas de 1950 sequer aparecem na Tabela 14, por exemplo.

Para avaliar se esta prática com base na normalização de séries temporais e eleição de alguns limiares medidos em unidades do desvio padrão é eficaz no que tange a monitoração da disponibilidade de recursos eólicos, as médias anuais de cada série foram calculadas e ordenadas. As 20 piores médias anuais de cada uma das 15 séries temporais sendo exploradas foram dispostas na Tabela 15 em classificação em ordem crescente. As células coloridas em vermelho indicam os anos em consonância com os resultados da Tabela 14: anos que apresentaram 3, 4 ou 5 ocorrências de meses abaixo do limiar de 1,5 foram coloridas em escala crescente.

Como exemplo, o ano de 1973, classificado como a pior média anual para a série N55 e que apresentou média anual 9,1% abaixo da média total de longo prazo (que considera, para as séries NOAA, o período total que contempla de 1851 a 2014), teve sua célula colorida de vermelho em mesma tonalidade que na Tabela 14. Isto se deve às 5 ocorrências de meses abaixo do limiar de 1,5 unidades de desvio padrão.

	N55		N38		N15		N42		N29
Média	a total: 6,863	Média	a total: 4,235	Média	a total: 7,230	Média	total: 2,145	Média	total: 3,583
Ano	Média anual	Ano	Média anual	Ano	Média anual	Ano	Média anual	Ano	Média anual
1973	6,241 _{19,1%}	1964	3,521 _{16,9%}	1973	6,259 _{↓13,4%}	1973	1,724 _{↓19,6%}	1973	2,892 _{↓19,3%}
1984	6,359 _{17,3%}	1973	3,578 _{15,5%}	1917	6,296 _{↓12,9%}	2009	1,730 _{↓19,4%}	1964	3,007 _{16,1%}
1974	6,390 _{16,9%}	1985	3,604 _{14,9%}	1964	6,342 _{12,3%}	1989	1,795 _{↓16,3%}	1985	3,039 _{15,2%}
1953	6,462 _{15,8%}	1974	3,616 _{14,6%}	1985	6,425 _{↓11,1%}	1974	1,814 _{↓15,4%}	1916	3,045 _{↓15,0%}
1971	6,482 _{↓5,6%}	1989	3,645 _{13,9%}	1921	6,458 _{↓10,7%}	1986	1,845 _{↓14,0%}	1917	3,062 _{↓14,5%}
1952	6,487 _{↓5,6%}	1971	3,723 _{12,1%}	1963	6,525 _{19,7%}	2006	1,859 _{↓13,3%}	1851	3,105 _{↓13,3%}
1951	6,507 _{15,2%}	1975	3,760 _{↓11,2%}	1875	6,527 _{19,7%}	1971	1,872 _{12,7%}	1921	3,106 _{↓13,3%}
1920	6,508 _{15,2%}	2000	3,763 _{↓11,2%}	1974	6,540 _{19,5%}	2003	1,885 _{↓12,1%}	1974	3,109 _{13,2%}
1965	6,532 _{14,8%}	1921	3,825 _{19,7%}	1920	6,579 _{↓9,0%}	1994	1,887 _{↓12,0%}	1989	3,124 _{12,8%}
1969	6,560 _{↓4,4%}	1956	3,828 _{↓9,6%}	1989	6,581 _{↓9,0%}	2005	1,893 _{↓11,8%}	1971	3,156 _{↓11,9%}
1954	6,563 _{14,4%}	1988	3,832 _{19,5%}	1924	6,616 _{18,5%}	1984	1,894 _{↓11,7%}	1963	3,163 _{↓11,7%}
1972	6,566 _{14,3%}	1972	3,841 _{19,3%}	1923	6,627 _{18,3%}	2004	1,896 _{11,6%}	1920	3,168 _{↓11,6%}
1914	6,596 _{13,9%}	1873	3,843 _{19,3%}	1972	6,630 _{18,3%}	1921	1,902 _{↓11,3%}	1956	3,205 _{10,5%}
1934	6,624 _{13,5%}	1872	3,875 _{18,5%}	1975	6,646 _{↓8,1%}	1917	1,903 _{↓11,3%}	2000	3,214 _{10,3%}
1993	6,633 _{13,4%}	1918	3,880 _{↓8,4%}	1872	6,647 _{18,1%}	1975	1,903 _{↓11,3%}	1975	3,224
1955	6,644 _{13,2%}	1858	3,892 _{18,1%}	1984	6,684 _{↓7,6%}	1987	1,903 _{↓11,3%}	1988	3,238 <mark>↓9,6%</mark>
1946	6,654 _{13,0%}	1924	3,896 _{↓8,0%}	1935	6,690 _{17,5%}	1993	1,911 _{↓10,9%}	1873	3,244 _{↓9,4%}
1923	6,669 _{12,8%}	1963	3,923 _{↓7,4%}	1916	6,697 _{17,4%}	1945	1,911 _{↓10,9%}	1972	3,245 <mark>↓9,4%</mark>
1930	6,672 _{12,8%}	1967	3,932 _{↓7,2%}	1882	6,735 _{↓6,8%}	1920	1,915 _{10,7%}	1872	3,251 <mark>↓9,3%</mark>
1956	6,678 _{12,7%}	1965	3,938 _{↓7,0%}	1873	6,738 _{↓6,8%}	1929	1,917↓ <mark>10,6%</mark>	1918	3,273 <mark>↓8,6%</mark>
	T 22		T105		-				
	E33		E135		E190		E193		E212
Média	E33 a total: 7,952	Média	E135 a total: 3,862	Média	E190 a total: 2,906	Média	E193 1 total: 4,990	Média	E212 a total: 2,485
Média Ano	E33 a total: 7,952 Média anual	Média Ano	E135 a total: 3,862 Média anual	Média Ano	E190 a total: 2,906 Média anual	Média Ano	E193 a total: 4,990 Média anual	Média Ano	E212 a total: 2,485 Média anual
Média Ano 1924	E33 a total: 7,952 Média anual 6,365 _{120,0%}	Média Ano 1971	E135 a total: 3,862 Média anual 3,137 _{18,8%}	Média Ano 1928	E190 a total: 2,906 Média anual 2,344 _{19,3%}	Média Ano 1913	E193 a total: 4,990 Média anual 3,738 _{125,1%}	Média Ano 1923	E212 a total: 2,485 Média anual 1,944 _{↓21,8%}
Média Ano 1924 1974	E33 a total: 7,952 Média anual 6,365↓20,0% 6,398↓19,5%	Média Ano 1971 1962	E135 a total: 3,862 Média anual 3,137 _{18,8%} 3,265 _{15,4%}	Média Ano 1928 1923	E190 a total: 2,906 Média anual 2,344 _{19,3%} 2,513 _{13,5%}	Média Ano 1913 1911	E193 total: 4,990 Média anual 3,738 _{25,1%} 4,011 _{19,6%}	Média Ano 1923 1972	E212 a total: 2,485 Média anual 1,944 _{121,8%} 1,974 _{120,6%}
Média Ano 1924 1974 1917	E33 a total: 7,952 Média anual 6,365↓20,0% 6,398↓19,5% 6,449↓18,9%	Média Ano 1971 1962 1924	E135 a total: 3,862 Média anual 3,137 _{↓18,8%} 3,265 _{↓15,4%} 3,334 _{↓13,7%}	Média Ano 1928 1923 1965	E190 a total: 2,906 Média anual 2,344 _{19,3%} 2,513 _{13,5%} 2,544 _{12,4%}	Média Ano 1913 1911 1928	E193 total: 4,990 Média anual 3,738 _{125,1%} 4,011 _{19,6%} 4,026 _{19,3%}	Média Ano 1923 1972 1971	E212 a total: 2,485 Média anual 1,944 _{121,8%} 1,974 _{120,6%} 2,034 _{118,2%}
Média Ano 1924 1974 1917 1985	E33 a total: 7,952 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$	Média Ano 1971 1962 1924 1985	E135 a total: 3,862 Média anual 3,137 _{18,8%} 3,265 _{15,4%} 3,334 _{13,7%} 3,407 _{11,8%}	Média Ano 1928 1923 1965 1911	E190 total: 2,906 Média anual 2,344 _{19,3%} 2,513 _{13,5%} 2,544 _{12,4%} 2,597 _{10,6%}	Média Ano 1913 1911 1928 1923	E193 Média anual 3,738 _{25,1%} 4,011 _{19,6%} 4,026 _{19,3%} 4,116 _{17,5%}	Média Ano 1923 1972 1971 1907	E212 total: 2,485 Média anual 1,944 _{121,8%} 1,974 _{120,6%} 2,034 _{18,2%} 2,061 _{117,1%}
Média Ano 1924 1974 1917 1985 1921	E33 a total: 7,952 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$	Média Ano 1971 1962 1924 1985 1922	E135 a total: 3,862 Média anual 3,137 _{↓18,8%} 3,265 _{↓15,4%} 3,334 _{↓13,7%} 3,407 _{↓11,8%}	Média Ano 1928 1923 1965 1911 1972	E190 total: 2,906 Média anual 2,344 _{↓19,3%} 2,513 _{↓13,5%} 2,544 _{↓12,4%} 2,597 _{↓10,6%} 2,600 _{↓10,5%}	Média Ano 1913 1911 1928 1923 1927	E193 total: 4,990 Média anual 3,738 _{25,1%} 4,011 _{19,6%} 4,026 _{19,3%} 4,116 _{17,5%} 4,198 _{15,9%}	Média Ano 1923 1972 1971 1907 1913	E212 total: 2,485 Média anual 1,944 _{121,8%} 1,974 _{120,6%} 2,034 _{118,2%} 2,061 _{117,1%} 2,099 _{115,5%}
Média Ano 1924 1974 1917 1985 1921 1935	E33 A total: 7,952 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$ $6,602_{\downarrow 17,0\%}$	Média Ano 1971 1962 1924 1985 1922 2009	E135 a total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 11,6\%}$	Média Ano 1928 1923 1965 1911 1972 1925	E190 A total: 2,906 Média anual $2,344_{\downarrow 19,3\%}$ $2,513_{\downarrow 13,5\%}$ $2,544_{\downarrow 12,4\%}$ $2,597_{\downarrow 10,6\%}$ $2,600_{\downarrow 10,5\%}$	Média Ano 1913 1911 1928 1923 1927 1907	E193 Média anual $3,738_{\downarrow 25,1\%}$ $4,011_{\downarrow 19,6\%}$ $4,026_{\downarrow 19,3\%}$ $4,116_{\downarrow 17,5\%}$ $4,198_{\downarrow 15,9\%}$ $4,201_{\downarrow 15,8\%}$	Média Ano 1923 1972 1971 1907 1913 1911	E212 Média anual 1,944 _{21,8%} 1,974 _{20,6%} 2,034 _{18,2%} 2,061 _{17,1%} 2,099 _{15,5%} 2,103 _{15,4%}
Média Ano 1924 1974 1917 1985 1921 1935 1973	E33 A total: 7,952 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$ $6,602_{\downarrow 17,0\%}$ $6,708_{\downarrow 15,6\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972	E135 a total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 11,6\%}$ $3,441_{\downarrow 10,9\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924	E190 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,601 \downarrow 10,5%	Média Ano 1913 1911 1928 1923 1927 1907 1912	E193 Média anual 3,738 _{25,1%} 4,011 _{19,6%} 4,026 _{19,3%} 4,116 _{17,5%} 4,198 _{15,9%} 4,201 _{15,8%}	Média Ano 1923 1972 1971 1907 1913 1911 1965	E212 Média anual 1,944 _{121,8%} 1,974 _{120,6%} 2,034 _{118,2%} 2,061 _{117,1%} 2,099 _{115,5%} 2,103 _{115,4%} 2,110 _{115,1%}
Média Ano 1924 1974 1917 1985 1921 1935 1973 1975	E33 A total: 7,952 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$ $6,602_{\downarrow 17,0\%}$ $6,708_{\downarrow 15,6\%}$ $6,808_{\downarrow 14,4\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923	E135 a total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 11,6\%}$ $3,446_{\downarrow 10,8\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971	E190 A total: 2,906 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,602 \downarrow 10,5% 2,602 \downarrow 10,5% 2,605 \downarrow 10,3%	Média Ano 1913 1911 1928 1923 1927 1907 1912 1929	E193 total: 4,990 Média anual 3,738 _{25,1%} 4,011 _{19,6%} 4,026 _{19,3%} 4,116 _{17,5%} 4,198 _{15,9%} 4,201 _{15,8%} 4,212 _{15,6%}	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973	E212 total: 2,485 Média anual 1,944 _{121,8%} 1,974 _{120,6%} 2,034 _{118,2%} 2,061 _{117,1%} 2,099 _{115,5%} 2,103 _{115,4%} 2,110 _{115,1%}
Média Ano 1924 1974 1917 1985 1921 1935 1973 1975 1934	E33 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$ $6,602_{\downarrow 17,0\%}$ $6,708_{\downarrow 15,6\%}$ $6,808_{\downarrow 14,4\%}$ $6,828_{\downarrow 14,1\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923 1965	E135 A total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 11,6\%}$ $3,446_{\downarrow 10,8\%}$ $3,446_{\downarrow 10,8\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973	E190 Média anual $2,344_{\downarrow 19,3\%}$ $2,513_{\downarrow 13,5\%}$ $2,544_{\downarrow 12,4\%}$ $2,597_{\downarrow 10,6\%}$ $2,600_{\downarrow 10,5\%}$ $2,602_{\downarrow 10,5\%}$ $2,605_{\downarrow 10,3\%}$ $2,617_{\downarrow 9,9\%}$	Média Ano 1913 1911 1928 1923 1927 1907 1912 1929 1925	E193 Média anual $3,738_{\downarrow 25,1\%}$ $4,011_{\downarrow 19,6\%}$ $4,026_{\downarrow 19,3\%}$ $4,116_{\downarrow 17,5\%}$ $4,198_{\downarrow 15,9\%}$ $4,210_{\downarrow 15,6\%}$ $4,240_{\downarrow 15,0\%}$ $4,287_{\downarrow 14,1\%}$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904	E212 Média anual 1,944 _{121,8%} 1,974 _{120,6%} 2,034 _{118,2%} 2,061 _{117,1%} 2,099 _{115,5%} 2,103 _{115,4%} 2,110 _{115,1%} 2,153 _{113,4%}
Média Ano 1924 1974 1917 1985 1921 1935 1973 1975 1934 1913	E33 A total: 7,952 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$ $6,602_{\downarrow 17,0\%}$ $6,708_{\downarrow 15,6\%}$ $6,808_{\downarrow 14,4\%}$ $6,828_{\downarrow 14,1\%}$ $6,929_{\downarrow 12,9\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923 1965 1974	E135 A total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 11,6\%}$ $3,441_{\downarrow 10,9\%}$ $3,446_{\downarrow 10,8\%}$ $3,455_{\downarrow 10,5\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936	E190 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,600 \downarrow 10,5% 2,602 \downarrow 10,5% 2,605 \downarrow 10,3% 2,617 \downarrow 9,9% 2,625 \downarrow 9,7%	Média Ano 1913 1911 1928 1923 1927 1907 1912 1929 1925 1904	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,287\downarrow14,1\%$ $4,302\downarrow13,8\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927	E212 Média anual 1,944 _{121,8%} 1,974 _{120,6%} 2,034 _{118,2%} 2,061 _{117,1%} 2,099 _{115,5%} 2,103 _{115,4%} 2,110 _{115,1%} 2,153 _{113,4%} 2,160 _{113,1%}
Média Ano 1924 1974 1917 1985 1921 1935 1973 1973 1975 1934 1913 1920	E33 A total: 7,952 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$ $6,602_{\downarrow 17,0\%}$ $6,708_{\downarrow 15,6\%}$ $6,808_{\downarrow 14,4\%}$ $6,828_{\downarrow 14,1\%}$ $6,929_{\downarrow 12,9\%}$ $6,987_{\downarrow 12,1\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1972 1923 1965 1974 1933	E135 A total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 10,9\%}$ $3,446_{\downarrow 10,8\%}$ $3,446_{\downarrow 10,8\%}$ $3,455_{\downarrow 10,5\%}$ $3,477_{\downarrow 10,0\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936 1963	E190 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,602 \downarrow 10,5% 2,602 \downarrow 10,5% 2,605 \downarrow 10,3% 2,617 \downarrow 9,9% 2,625 \downarrow 9,7% 2,626 \downarrow 9,6%	Média Ano 1913 1911 1928 1923 1927 1907 1912 1929 1925 1904 1937	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,287\downarrow14,1\%$ $4,302\downarrow13,8\%$ $4,330\downarrow13,2\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927 1962	E212 Média anual $1,944_{\downarrow 21,8\%}$ $1,974_{\downarrow 20,6\%}$ $2,034_{\downarrow 18,2\%}$ $2,061_{\downarrow 17,1\%}$ $2,099_{\downarrow 15,5\%}$ $2,103_{\downarrow 15,4\%}$ $2,110_{\downarrow 15,1\%}$ $2,153_{\downarrow 13,4\%}$ $2,160_{\downarrow 13,1\%}$ $2,172_{\downarrow 12,6\%}$ $2,182_{\downarrow 12,2\%}$
Média Ano 1924 1974 1917 1985 1921 1935 1973 1973 1975 1934 1913 1920 1989	E33 Média anual $6,365\downarrow20,0\%$ $6,398\downarrow19,5\%$ $6,449\downarrow18,9\%$ $6,474\downarrow18,6\%$ $6,567\downarrow17,4\%$ $6,602\downarrow17,0\%$ $6,708\downarrow15,6\%$ $6,808\downarrow14,4\%$ $6,828\downarrow14,1\%$ $6,929\downarrow12,9\%$ $6,987\downarrow12,1\%$ $7,013\downarrow11,8\%$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923 1965 1974 1933 1916	E135 A total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 11,6\%}$ $3,446_{\downarrow 10,8\%}$ $3,446_{\downarrow 10,8\%}$ $3,445_{\downarrow 10,5\%}$ $3,477_{\downarrow 10,0\%}$ $3,487_{\downarrow 9,7\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936 1963 1937	E190 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,600 \downarrow 10,5% 2,602 \downarrow 10,5% 2,605 \downarrow 10,3% 2,617 \downarrow 9,9% 2,625 \downarrow 9,7% 2,626 \downarrow 9,6% 2,635 \downarrow 9,3%	Média Ano 1913 1911 1928 1923 1927 1907 1912 1929 1925 1904 1937 1965	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,240\downarrow15,0\%$ $4,302\downarrow13,8\%$ $4,330\downarrow13,2\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927 1962 1919	E212 Média anual $1,944_{\downarrow 21,8\%}$ $1,974_{\downarrow 20,6\%}$ $2,034_{\downarrow 18,2\%}$ $2,061_{\downarrow 17,1\%}$ $2,099_{\downarrow 15,5\%}$ $2,103_{\downarrow 15,4\%}$ $2,110_{\downarrow 15,1\%}$ $2,153_{\downarrow 13,4\%}$ $2,172_{\downarrow 12,6\%}$ $2,182_{\downarrow 12,2\%}$ $2,193_{\downarrow 11,8\%}$
Média Ano 1924 1974 1917 1985 1921 1935 1973 1975 1934 1913 1920 1989 1933	E33 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$ $6,602_{\downarrow 17,0\%}$ $6,708_{\downarrow 15,6\%}$ $6,808_{\downarrow 14,4\%}$ $6,828_{\downarrow 14,1\%}$ $6,929_{\downarrow 12,9\%}$ $6,987_{\downarrow 12,1\%}$ $7,013_{\downarrow 11,8\%}$ $7,047_{\downarrow 11,4\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923 1965 1974 1933 1916 1963	E135 A total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 11,6\%}$ $3,446_{\downarrow 10,8\%}$ $3,446_{\downarrow 10,8\%}$ $3,445_{\downarrow 10,5\%}$ $3,477_{\downarrow 10,0\%}$ $3,487_{\downarrow 9,7\%}$ $3,502_{\downarrow 9,3\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936 1963 1937 1926	E190 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,602 \downarrow 10,5% 2,602 \downarrow 10,5% 2,602 \downarrow 10,5% 2,602 \downarrow 10,5% 2,602 \downarrow 10,5% 2,602 \downarrow 10,5% 2,625 \downarrow 9,7% 2,625 \downarrow 9,7% 2,625 \downarrow 9,3% 2,639 \downarrow 9,2%	Média Ano 1913 1911 1928 1923 1927 1927 1907 1912 1929 1925 1904 1937 1965 1938	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,287\downarrow14,1\%$ $4,302\downarrow13,8\%$ $4,330\downarrow13,2\%$ $4,343\downarrow13,0\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927 1962 1919 2009	E212 Média anual 1,944 \downarrow 21,8% 1,974 \downarrow 20,6% 2,034 \downarrow 18,2% 2,061 \downarrow 17,1% 2,099 \downarrow 15,5% 2,103 \downarrow 15,4% 2,110 \downarrow 15,1% 2,153 \downarrow 13,4% 2,160 \downarrow 13,1% 2,172 \downarrow 12,6% 2,182 \downarrow 12,2% 2,193 \downarrow 11,8% 2,207 \downarrow 11,2%
Média Ano 1924 1974 1917 1985 1921 1935 1973 1973 1975 1934 1913 1920 1989 1933 1937	E33 Média anual $6,365\downarrow_{20,0\%}$ $6,398\downarrow_{19,5\%}$ $6,449\downarrow_{18,9\%}$ $6,474\downarrow_{18,6\%}$ $6,567\downarrow_{17,4\%}$ $6,602\downarrow_{17,0\%}$ $6,708\downarrow_{15,6\%}$ $6,808\downarrow_{14,4\%}$ $6,828\downarrow_{14,1\%}$ $6,929\downarrow_{12,9\%}$ $6,987\downarrow_{12,1\%}$ $7,013\downarrow_{11,8\%}$ $7,047\downarrow_{11,4\%}$ $7,118\downarrow_{10,5\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923 1972 1963 1974 1933 1916 1963 1929	E135 A total: 3,862 Média anual $3,137\downarrow_{18,8\%}$ $3,265\downarrow_{15,4\%}$ $3,334\downarrow_{13,7\%}$ $3,407\downarrow_{11,8\%}$ $3,407\downarrow_{11,8\%}$ $3,414\downarrow_{10,9\%}$ $3,446\downarrow_{10,8\%}$ $3,446\downarrow_{10,8\%}$ $3,446\downarrow_{10,8\%}$ $3,4455\downarrow_{10,5\%}$ $3,477\downarrow_{10,0\%}$ $3,487\downarrow_{9,7\%}$ $3,502\downarrow_{9,3\%}$ $3,547\downarrow_{8,1\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936 1963 1937 1926 1976	E190 Média anual $2,344_{\downarrow 19,3\%}$ $2,513_{\downarrow 13,5\%}$ $2,544_{\downarrow 12,4\%}$ $2,597_{\downarrow 10,6\%}$ $2,600_{\downarrow 10,5\%}$ $2,602_{\downarrow 10,5\%}$ $2,605_{\downarrow 10,3\%}$ $2,625_{\downarrow 9,7\%}$ $2,625_{\downarrow 9,7\%}$ $2,625_{\downarrow 9,3\%}$ $2,635_{\downarrow 9,3\%}$ $2,639_{\downarrow 9,2\%}$ $2,648_{\downarrow 8,9\%}$	Média Ano 1913 1911 1928 1923 1927 1927 1907 1912 1929 1925 1904 1937 1965 1938 1973	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,287\downarrow14,1\%$ $4,302\downarrow13,8\%$ $4,330\downarrow13,2\%$ $4,343\downarrow13,0\%$ $4,352\downarrow12,8\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927 1962 1919 2009 1930	E212 Média anual $1,944_{\downarrow 21,8\%}$ $1,974_{\downarrow 20,6\%}$ $2,034_{\downarrow 18,2\%}$ $2,061_{\downarrow 17,1\%}$ $2,099_{\downarrow 15,5\%}$ $2,103_{\downarrow 15,4\%}$ $2,110_{\downarrow 15,1\%}$ $2,153_{\downarrow 13,4\%}$ $2,160_{\downarrow 13,1\%}$ $2,172_{\downarrow 12,6\%}$ $2,193_{\downarrow 11,8\%}$ $2,207_{\downarrow 11,2\%}$ $2,239_{\downarrow 9,9\%}$
Média Ano 1924 1974 1917 1985 1921 1935 1973 1975 1934 1913 1920 1989 1933 1937 1910	E33 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$ $6,602_{\downarrow 17,0\%}$ $6,708_{\downarrow 15,6\%}$ $6,808_{\downarrow 14,4\%}$ $6,828_{\downarrow 14,1\%}$ $6,929_{\downarrow 12,9\%}$ $6,987_{\downarrow 12,1\%}$ $7,013_{\downarrow 11,8\%}$ $7,047_{\downarrow 11,4\%}$ $7,118_{\downarrow 10,5\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923 1965 1974 1933 1916 1963 1916 1963 1929	E135 A total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 11,6\%}$ $3,446_{\downarrow 10,8\%}$ $3,446_{\downarrow 10,8\%}$ $3,446_{\downarrow 10,8\%}$ $3,4455_{\downarrow 10,5\%}$ $3,477_{\downarrow 10,0\%}$ $3,487_{\downarrow 9,7\%}$ $3,502_{\downarrow 9,3\%}$ $3,555_{\downarrow 8,0\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936 1963 1937 1926 1976 1976	E190 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,600 \downarrow 10,5% 2,602 \downarrow 10,5% 2,605 \downarrow 10,3% 2,605 \downarrow 10,3% 2,617 \downarrow 9,9% 2,625 \downarrow 9,7% 2,626 \downarrow 9,6% 2,635 \downarrow 9,3% 2,639 \downarrow 9,2% 2,648 \downarrow 8,9% 2,675 \downarrow 7,9%	Média Ano 1913 1911 1928 1923 1927 1927 1907 1912 1929 1925 1904 1937 1965 1938 1973	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,287\downarrow14,1\%$ $4,302\downarrow13,8\%$ $4,330\downarrow13,2\%$ $4,336\downarrow13,1\%$ $4,343\downarrow13,0\%$ $4,376\downarrow12,3\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927 1962 1919 2009 1930 1928	E212 Média anual $1,944_{\downarrow 21,8\%}$ $1,974_{\downarrow 20,6\%}$ $2,034_{\downarrow 18,2\%}$ $2,061_{\downarrow 17,1\%}$ $2,099_{\downarrow 15,5\%}$ $2,103_{\downarrow 15,4\%}$ $2,110_{\downarrow 15,1\%}$ $2,153_{\downarrow 13,4\%}$ $2,160_{\downarrow 13,1\%}$ $2,182_{\downarrow 12,2\%}$ $2,182_{\downarrow 12,2\%}$ $2,207_{\downarrow 11,2\%}$ $2,239_{\downarrow 9,9\%}$ $2,250_{\downarrow 9,4\%}$
Média Ano 1924 1974 1917 1985 1921 1935 1973 1973 1973 1975 1934 1913 1920 1989 1933 1937 1910 1929	E33 Média anual $6,365\downarrow_{20,0\%}$ $6,398\downarrow_{19,5\%}$ $6,449\downarrow_{18,9\%}$ $6,474\downarrow_{18,6\%}$ $6,567\downarrow_{17,4\%}$ $6,602\downarrow_{17,0\%}$ $6,708\downarrow_{15,6\%}$ $6,808\downarrow_{14,4\%}$ $6,828\downarrow_{14,1\%}$ $6,929\downarrow_{12,9\%}$ $6,987\downarrow_{12,1\%}$ $7,013\downarrow_{11,8\%}$ $7,047\downarrow_{11,4\%}$ $7,118\downarrow_{10,5\%}$ $7,120\downarrow_{10,5\%}$ $7,177\downarrow_{9,8\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923 1965 1974 1933 1916 1963 1929 1928	E135 Média anual $3,137\downarrow18,8\%$ $3,265\downarrow15,4\%$ $3,334\downarrow13,7\%$ $3,407\downarrow11,8\%$ $3,407\downarrow11,8\%$ $3,414\downarrow10,9\%$ $3,446\downarrow10,8\%$ $3,446\downarrow10,8\%$ $3,446\downarrow10,8\%$ $3,4455\downarrow10,5\%$ $3,447\downarrow10,0\%$ $3,487\downarrow9,7\%$ $3,502\downarrow9,3\%$ $3,555\downarrow8,0\%$ $3,556\downarrow7,9\%$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936 1963 1937 1926 1978 1978 1927	E190 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,602 \downarrow 10,5% 2,603 \downarrow 10,3% 2,625 \downarrow 9,7% 2,62619,6% 2,639 \downarrow 9,2% 2,648 \downarrow 8,9% 2,675 \downarrow 7,9% 2,687 \downarrow 7,5%	Média Ano 1913 1911 1928 1923 1927 1907 1912 1929 1925 1904 1937 1965 1938 1938 1973	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,287\downarrow14,1\%$ $4,302\downarrow13,8\%$ $4,336\downarrow13,1\%$ $4,343\downarrow13,0\%$ $4,376\downarrow12,3\%$ $4,378\downarrow12,3\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927 1962 1919 2009 1930 1928 1922	E212 Média anual $1,944_{\downarrow 21,8\%}$ $1,974_{\downarrow 20,6\%}$ $2,034_{\downarrow 18,2\%}$ $2,061_{\downarrow 17,1\%}$ $2,099_{\downarrow 15,5\%}$ $2,103_{\downarrow 15,4\%}$ $2,110_{\downarrow 15,1\%}$ $2,153_{\downarrow 13,4\%}$ $2,160_{\downarrow 13,1\%}$ $2,172_{\downarrow 12,6\%}$ $2,182_{\downarrow 12,2\%}$ $2,193_{\downarrow 11,8\%}$ $2,207_{\downarrow 11,2\%}$ $2,250_{\downarrow 9,4\%}$ $2,269_{\downarrow 8,7\%}$
Média Ano 1924 1974 1917 1985 1921 1935 1973 1975 1934 1913 1920 1989 1933 1937 1910 1929 1923	E33 Média anual $6,365\downarrow 20,0\%$ $6,398\downarrow 19,5\%$ $6,449\downarrow 18,9\%$ $6,474\downarrow 18,6\%$ $6,567\downarrow 17,4\%$ $6,602\downarrow 17,0\%$ $6,708\downarrow 15,6\%$ $6,808\downarrow 14,4\%$ $6,828\downarrow 14,1\%$ $6,929\downarrow 12,9\%$ $6,987\downarrow 12,1\%$ $7,013\downarrow 11,8\%$ $7,047\downarrow 11,4\%$ $7,118\downarrow 10,5\%$ $7,181\downarrow 9,7\%$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923 1965 1974 1933 1916 1963 1929 1928 1927 1927	E135 A total: 3,862 Média anual $3,137\downarrow_{18,8\%}$ $3,265\downarrow_{15,4\%}$ $3,334\downarrow_{13,7\%}$ $3,407\downarrow_{11,8\%}$ $3,407\downarrow_{11,8\%}$ $3,414\downarrow_{10,9\%}$ $3,446\downarrow_{10,8\%}$ $3,446\downarrow_{10,8\%}$ $3,446\downarrow_{10,8\%}$ $3,446\downarrow_{10,8\%}$ $3,447\downarrow_{10,0\%}$ $3,447\downarrow_{10,0\%}$ $3,477\downarrow_{10,0\%}$ $3,555\downarrow_{10,5\%}$ $3,555\downarrow_{18,0\%}$ $3,556\downarrow_{7,9\%}$ $3,562\downarrow_{7,8\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936 1963 1937 1926 1976 1978 1927 1927	E190 Média anual $2,344_{\downarrow 19,3\%}$ $2,513_{\downarrow 13,5\%}$ $2,544_{\downarrow 12,4\%}$ $2,597_{\downarrow 10,6\%}$ $2,600_{\downarrow 10,5\%}$ $2,602_{\downarrow 10,5\%}$ $2,605_{\downarrow 10,3\%}$ $2,625_{\downarrow 9,7\%}$ $2,625_{\downarrow 9,7\%}$ $2,625_{\downarrow 9,7\%}$ $2,635_{\downarrow 9,3\%}$ $2,639_{\downarrow 9,2\%}$ $2,648_{\downarrow 8,9\%}$ $2,675_{\downarrow 7,9\%}$ $2,693_{\downarrow 7,3\%}$	Média Ano 1913 1911 1928 1923 1927 1927 1907 1912 1929 1925 1904 1937 1965 1938 1973 1930 1963 1972	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,240\downarrow15,0\%$ $4,302\downarrow13,8\%$ $4,330\downarrow13,2\%$ $4,336\downarrow13,1\%$ $4,343\downarrow13,0\%$ $4,352\downarrow12,8\%$ $4,376\downarrow12,3\%$ $4,383\downarrow12,2\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927 1962 1919 2009 1930 1928 1922 1945	E212 Média anual $1,944_{\downarrow 21,8\%}$ $1,974_{\downarrow 20,6\%}$ $2,034_{\downarrow 18,2\%}$ $2,061_{\downarrow 17,1\%}$ $2,099_{\downarrow 15,5\%}$ $2,103_{\downarrow 15,4\%}$ $2,110_{\downarrow 15,1\%}$ $2,153_{\downarrow 13,4\%}$ $2,160_{\downarrow 13,1\%}$ $2,172_{\downarrow 12,6\%}$ $2,182_{\downarrow 12,2\%}$ $2,193_{\downarrow 11,8\%}$ $2,207_{\downarrow 11,2\%}$ $2,239_{\downarrow 9,9\%}$ $2,250_{\downarrow 9,4\%}$ $2,275_{\downarrow 8,5\%}$
Média Ano 1924 1974 1917 1985 1921 1935 1973 1975 1934 1913 1920 1989 1933 1937 1910 1929 1923 1916	E33 Média anual $6,365_{\downarrow 20,0\%}$ $6,398_{\downarrow 19,5\%}$ $6,449_{\downarrow 18,9\%}$ $6,474_{\downarrow 18,6\%}$ $6,567_{\downarrow 17,4\%}$ $6,602_{\downarrow 17,0\%}$ $6,708_{\downarrow 15,6\%}$ $6,808_{\downarrow 14,4\%}$ $6,828_{\downarrow 14,1\%}$ $6,929_{\downarrow 12,9\%}$ $6,987_{\downarrow 12,1\%}$ $7,013_{\downarrow 11,8\%}$ $7,047_{\downarrow 11,4\%}$ $7,118_{\downarrow 10,5\%}$ $7,120_{\downarrow 10,5\%}$ $7,181_{\downarrow 9,7\%}$ $7,187_{\downarrow 9,6\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1923 1965 1974 1933 1916 1963 1929 1928 1927 1967 1940	E135 A total: 3,862 Média anual $3,137_{\downarrow 18,8\%}$ $3,265_{\downarrow 15,4\%}$ $3,334_{\downarrow 13,7\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,407_{\downarrow 11,8\%}$ $3,414_{\downarrow 10,9\%}$ $3,446_{\downarrow 10,8\%}$ $3,446_{\downarrow 10,8\%}$ $3,446_{\downarrow 10,8\%}$ $3,4455_{\downarrow 10,5\%}$ $3,477_{\downarrow 10,0\%}$ $3,487_{\downarrow 9,7\%}$ $3,502_{\downarrow 9,3\%}$ $3,555_{\downarrow 8,0\%}$ $3,556_{\downarrow 7,9\%}$ $3,566_{\downarrow 7,7\%}$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936 1963 1937 1926 1937 1926 1978 1927 1974 1974	E190 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,601 \downarrow 10,5% 2,602 \downarrow 10,5% 2,602 \downarrow 10,5% 2,605 \downarrow 10,3% 2,605 \downarrow 10,3% 2,617 \downarrow 9,9% 2,625 \downarrow 9,7% 2,62619,6% 2,635 \downarrow 9,3% 2,639 \downarrow 9,2% 2,648 \downarrow 8,9% 2,675 \downarrow 7,9% 2,693 \downarrow 7,3% 2,694 \downarrow 7,3%	Média Ano 1913 1911 1928 1923 1927 1927 1907 1912 1929 1925 1904 1937 1965 1938 1930 1930 1963 1930 1963	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,240\downarrow15,0\%$ $4,287\downarrow14,1\%$ $4,302\downarrow13,8\%$ $4,330\downarrow13,2\%$ $4,336\downarrow13,1\%$ $4,343\downarrow13,0\%$ $4,352\downarrow12,8\%$ $4,376\downarrow12,3\%$ $4,383\downarrow12,2\%$ $4,394\downarrow12,0\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927 1962 1919 2009 1930 1928 1922 1945 1924	E212 Média anual $1,944_{\downarrow 21,8\%}$ $1,974_{\downarrow 20,6\%}$ $2,034_{\downarrow 18,2\%}$ $2,061_{\downarrow 17,1\%}$ $2,099_{\downarrow 15,5\%}$ $2,103_{\downarrow 15,4\%}$ $2,110_{\downarrow 15,1\%}$ $2,153_{\downarrow 13,4\%}$ $2,160_{\downarrow 13,1\%}$ $2,172_{\downarrow 12,6\%}$ $2,193_{\downarrow 11,2\%}$ $2,207_{\downarrow 11,2\%}$ $2,239_{\downarrow 9,9\%}$ $2,250_{\downarrow 9,4\%}$ $2,275_{\downarrow 8,5\%}$ $2,275_{\downarrow 8,4\%}$
Média Ano 1924 1974 1917 1985 1921 1935 1973 1973 1973 1973 1973 1973 1934 1913 1920 1933 1937 1910 1929 1923 1916 1925	E33 A total: 7,952 Média anual $6,365\downarrow_{20,0\%}$ $6,398\downarrow_{19,5\%}$ $6,449\downarrow_{18,9\%}$ $6,474\downarrow_{18,6\%}$ $6,567\downarrow_{17,4\%}$ $6,602\downarrow_{17,0\%}$ $6,708\downarrow_{15,6\%}$ $6,808\downarrow_{14,4\%}$ $6,828\downarrow_{14,1\%}$ $6,929\downarrow_{12,9\%}$ $6,987\downarrow_{12,1\%}$ $7,013\downarrow_{11,8\%}$ $7,047\downarrow_{11,4\%}$ $7,118\downarrow_{10,5\%}$ $7,120\downarrow_{10,5\%}$ $7,187\downarrow_{9,6\%}$ $7,200\downarrow_{9,5\%}$	Média Ano 1971 1962 1924 1985 1922 2009 1972 1972 1973 1963 1974 1933 1916 1963 1929 1928 1927 1928 1927 1967 1940	E135 A total: 3,862 Média anual $3,137\downarrow18,8\%$ $3,265\downarrow15,4\%$ $3,334\downarrow13,7\%$ $3,407\downarrow11,8\%$ $3,407\downarrow11,8\%$ $3,414\downarrow10,9\%$ $3,446\downarrow10,8\%$ $3,446\downarrow10,8\%$ $3,446\downarrow10,8\%$ $3,446\downarrow10,8\%$ $3,4455\downarrow10,5\%$ $3,4475\downarrow10,0\%$ $3,477\downarrow10,0\%$ $3,487\downarrow9,7\%$ $3,502\downarrow9,3\%$ $3,555\downarrow8,0\%$ $3,556\downarrow7,9\%$ $3,566\downarrow7,7\%$ $3,583\downarrow7,2\%$	Média Ano 1928 1923 1965 1911 1972 1925 1924 1971 1973 1936 1963 1937 1926 1976 1976 1978 1927 1974 1980 1966	E190 Média anual 2,344 \downarrow 19,3% 2,513 \downarrow 13,5% 2,544 \downarrow 12,4% 2,597 \downarrow 10,6% 2,600 \downarrow 10,5% 2,602 \downarrow 10,5% 2,603 \downarrow 10,5% 2,603 \downarrow 10,5% 2,639 \downarrow 10,5% 2,648 \downarrow 8,9% 2,693 \downarrow 7,5% 2,694 \downarrow 7,3% 2,697 \downarrow 7,2%	Média Ano 1913 1911 1928 1923 1927 1907 1912 1925 1904 1937 1965 1938 1973 1930 1963 1930 1963 1972 1905 1936	E193 Média anual $3,738\downarrow25,1\%$ $4,011\downarrow19,6\%$ $4,026\downarrow19,3\%$ $4,116\downarrow17,5\%$ $4,198\downarrow15,9\%$ $4,201\downarrow15,8\%$ $4,212\downarrow15,6\%$ $4,240\downarrow15,0\%$ $4,240\downarrow15,0\%$ $4,237\downarrow14,1\%$ $4,302\downarrow13,8\%$ $4,336\downarrow13,1\%$ $4,336\downarrow13,1\%$ $4,352\downarrow12,8\%$ $4,376\downarrow12,3\%$ $4,376\downarrow12,3\%$ $4,383\downarrow12,2\%$ $4,394\downarrow12,0\%$ $4,414\downarrow11,6\%$	Média Ano 1923 1972 1971 1907 1913 1911 1965 1973 1904 1927 1962 1919 2009 1930 1928 1922 1945 1924 1924	E212 Média anual $1,944_{\downarrow 21,8\%}$ $1,974_{\downarrow 20,6\%}$ $2,034_{\downarrow 18,2\%}$ $2,061_{\downarrow 17,1\%}$ $2,099_{\downarrow 15,5\%}$ $2,103_{\downarrow 15,4\%}$ $2,110_{\downarrow 15,1\%}$ $2,153_{\downarrow 13,4\%}$ $2,160_{\downarrow 13,1\%}$ $2,172_{\downarrow 12,6\%}$ $2,193_{\downarrow 11,8\%}$ $2,207_{\downarrow 11,2\%}$ $2,239_{\downarrow 9,9\%}$ $2,250_{\downarrow 9,4\%}$ $2,275_{\downarrow 8,5\%}$ $2,275_{\downarrow 8,4\%}$ $2,279_{\downarrow 8,3\%}$

Tabela 15 - Classificação das médias anuais de velocidade do vento [m/s] e relações com as médias totais

	M139		M211		M179		M581		M622
Média	a total: 8,562	Média	a total: 4,744	Média	1 total: 2,397	Média	1 total: 4,797	Média	1 total: 7,122
Ano	Média anual	Ano	Média anual	Ano	Média anual	Ano	Média anual	Ano	Média anual
1985	7,260 _{↓15,2%}	1985	4,078 <mark>↓14,0%</mark>	1984	2,153 _{10,2%}	2009	4,449 _{↓7,3%}	1984	6,768 _{15,0%}
1984	7,599 _{↓11,2%}	1999	4,373 _{↓7,8%}	2006	2,159 _{19,9%}	2000	4,510 _{↓6,0%}	1987	6,855 _{13,7%}
1989	7,706 _{↓10,0%}	1986	4,415 _{↓6,9%}	1999	2,198 _{↓8,3%}	1997	4,544 _{↓5,3%}	1993	6,862 _{↓3,7%}
2009	7,850 _{↓8,3%}	1988	4,418 _{↓6,9%}	1996	2,212 _{↓7,7%}	1989	4,615 _{↓3,8%}	1990	6,909 _{↓3,0%}
1986	7,882 _{↓7,9%}	2009	4,423 _{↓6,8%}	1988	2,218 _{↓7,5%}	2006	4,616 _{↓3,8%}	2015	6,919 _{↓2,8%}
1988	8,018 <mark>↓6,4%</mark>	2011	4,428 _{↓6,7%}	1985	2,219 _{17,4%}	2011	4,624 _{↓3,6%}	1983	6,925 <mark>↓2,8%</mark>
2008	8,147 _{↓4,8%}	2008	4,439 _{↓6,4%}	2008	2,234 _{↓6,8%}	1993	4,631 _{↓3,5%}	1988	6,927 _{12,7%}
1995	8,180 _{↓4,5%}	1989	4,444 _{↓6,3%}	1980	2,238 _{↓6,6%}	1995	4,641 _{↓3,3%}	1980	6,932 _{↓2,7%}
1994	8,205 _{↓4,2%}	1996	4,459 _{↓6,0%}	2000	2,269 _{↓5,3%}	1987	4,679 _{↓2,5%}	2009	6,989 _{↓1,9%}
2011	8,207 _{↓4,1%}	1984	4,539 _{↓4,3%}	1989	2,275 _{↓5,1%}	2001	4,679 _{↓2,4%}	2003	6,999 _{↓1,7%}
1996	8,261 _{13,5%}	2001	4,540 _{↓4,3%}	1986	2,285 _{14,7%}	1985	4,688 _{↓2,3%}	2016	7,018 _{↓1,5%}
2000	8,358 _{↓2,4%}	2004	4,547 _{↓4,1%}	2011	2,297 _{↓4,2%}	2008	4,697 _{↓2,1%}	1996	7,018 _{↓1,5%}
2006	8,379 _{↓2,1%}	2003	4,561 _{↓3,9%}	2001	2,321 _{13,2%}	1994	4,715 _{↓1,7%}	1997	7,039 _{↓1,2%}
1982	8,433↓1,5%	2000	4,580 _{↓3,5%}	1995	2,354 _{↓1,8%}	1996	4,726 _{↓1,5%}	1994	7,052 _{↓1,0%}
1999	8,466 <mark>↓1,1%</mark>	2010	4,611 _{↓2,8%}	2009	2,356 _{↓1,7%}	2016	4,734 _{↓1,3%}	2010	7,087 <mark>↓0,5%</mark>
2003	8,522 _{↓0,5%}	1980	4,614 _{↓2,7%}	1981	2,366 _{↓1,3%}	1986	4,757 _{↓0,8%}	2014	7,106 _{↓0,2%}
1980	8,554 _{↓0,1%}	1995	4,628 _{↓2,4%}	2004	2,366 _{↓1,3%}	1998	4,762 _{↓0,7%}	1985	7,112 _{↓0,1%}
2002	8,572 _{10,1%}	2002	4,657 _{↓1,8%}	1994	2,374 _{↓1,0%}	2004	4,772 _{↓0,5%}	1989	7,115 _{10,1%}
2001	8,580 _{10,2%}	2006	4,677 _{↓1,4%}	2002	2,379 _{↓0,8%}	2003	4,772 _{↓0,5%}	1986	7,118 _{10,1%}
1991	8,669 _{1,2%}	1997	4,770 _{↑0,5%}	2010	2,382 _{↓0,6%}	1999	4,772 _{↓0,5%}	1995	7,126 _{↑0,1%}

Fonte: elaboração própria do autor

Os valores da Tabela 15 revelam que nem sempre anos que apresentam 3 ou mais meses abaixo do limiar terão médias anuais muito abaixo da média de longo prazo (considerando a série temporal integralmente). Tomando as séries da NOAA como exemplo, alguns anos da década de 1950 e o ano de 1972 não cumpriram com os critérios para entrar na Tabela 14 e ainda assim aparecem na Tabela 15 com médias anuais bem abaixo da média total. Alguns casos extremos, como o ano de 1973 para a série N42, que indica apenas um valor abaixo do limiar e apesar disso foi um ano com média anual 19,6% abaixo da média total.

O mesmo ocorre para séries da base de dados ERA, a exemplo de E193, que apresenta médias anuais muito abaixo da média de longo prazo em anos que constam apenas uma ocorrência de meses anômalos para baixo do limiar, como é o caso de 1928, cuja média calculada foi 19,3% abaixo da média histórica. Para séries MERRA as inconsistências tampouco passam despercebidas, vide série M211 que embora tenha apresentado média anual 7,8% no ano de 1999, não houve sequer um mês que tenha rompido o limiar neste mesmo.

Para algumas séries temporais, por outro lado, diversos anos com médias notoriamente abaixo das médias totais correspondentes foram acusados com meses abaixo do limiar num mesmo ano e que figuram nas primeiras posições da classificação trazida pela Tabela 15, como 1973, 1974, 1984, 1985 etc. Isto demonstra que o método é, sim, capaz de revelar os piores anos de disponibilidade de recursos eólicos. Entretanto, há anos como o de 1916 na série N15 que, apesar do alto número de ocorrências, não refletem como sendo o pior cenário já registrado frente aos outros anos.

Algumas conclusões podem ser extraídas dos resultados tabelados. A primeira delas é que o monitoramento dos recursos eólicos embasado por limiares que consideram anomalias dos valores das séries temporais não é a melhor alternativa como índice do desempenho do vento. Inconsistências como as elencadas anteriormente são cabíveis e podem fazer com que a percepção sobre o desempenho do vento seja enviesada. A segunda envolve o grau de associação dos piores anos classificados em ordem decrescente com o número de ocorrências de meses abaixo do limiar. Embora exista uma correlação entre estes valores, essa não é linearmente perfeita, o que significa que mesmo que o número de ocorrências seja alto, a média anual não necessariamente figurará entre as primeiras posições da ordenação classificatória. A partir disso, a terceira conclusão é formulada: alguns meses cujas médias mensais de velocidade do vento são, sem pormenorizar, historicamente maiores (como agosto, setembro e outubro – tema já discutido a partir da análise dos *boxplots* das Figuras 61, 62 e 63), eles muito provavelmente ponderam mais fortemente a favor a redução da média anual do vento mesmo que seus valores não estejam abaixo do limiar adotado.

Entende-se que o desenvolvimento de um índice para acompanhamento mensal dos recursos eólicos considerando a variabilidade de longo prazo é uma ferramenta imprescindível para apoiar decisões que circundam o dimensionamento da capacidade de geração a nível nacional. Apesar do ângulo aqui abordado para tratar esta questão não ter retornado um produto consistente que garanta níveis ideais de confiabilidade, as conclusões que puderam ser extraídas encaminham para o que se acredita ser a abordagem mais propícia. A sugestão de solução permeia a ponderação prévia de cada mês – ou até mesmo de cada estação – para definir o limiar a ser considerado, de maneira que meses em que historicamente há maiores médias de velocidade do vento possuam pesos maiores do que os demais, equalizando e personalizando conforme especificidades locais.

À parte esta temática de monitoramento, a consulta dos valores de médias anuais ordenados e a relação que estes possuem com as médias de longo prazo escancara que cada série possui uma variabilidade diferente em termos de média anual. Os piores registros para séries ERA apresentam valores anuais mais de 20% abaixo da média total, ao passo que algumas séries MERRA ficam apenas 5% abaixo da respectiva média total. Alguns períodos históricos chamam mais atenção por serem recorrentes em praticamente todas as séries temporais analisadas. Os resultados para os anos de 1973, 1974, 1984 e 1985 sinalizam convergência entre os modelos NOAA, ERA e MERRA. A plotagem dos valores específicos pare estes anos em contraste com o ciclo anual e média de longo prazo podem ser observadas na Figura 71.



Figura 71 – Séries plotadas em anos com forte sinal de escassez de recursos eólicos

Fonte: elaboração própria do autor

Estes anos plotados servem o propósito de fornecer uma ideia visual de como se dá o comportamento do vento ao longo desses anos em que os dados indicam redução na intensidade e escassez de recursos eólicos. A série N55 apresenta dados mensais abaixo dos valores médios (ciclo anual) em todos os meses, com poucas exceções. Outras séries NOAA indicam que o vento ficou abaixo do ciclo anual praticamente ao longo de todos os anos ilustrados na Figura 71. As demais séries não possuem um padrão marcante – os valores mensais ora ficam abaixo ora acima do ciclo anual, requerendo cautela para afirmar a caracterização de períodos de escassez generalizada por toda a região NE.

Por fim, os dados ordenados da Tabela 15, quando cruzados com as informações da Tabela 12 sobre os valores anuais médios associados a uma probabilidade de não excedência e um tempo de retorno, desvelam que cada série temporal variou ao longo de seu período de abrangência e atingiu diferentes patamares de baixa disponibilidade – conforme sua própria variabilidade intrínseca. Ao tomar os piores anos de cada série, poucos alcançaram os valores do tempo de retorno de 10 anos (N55, N38, N29 e E33), apenas um chegou no valor do tempo de retorno de 20 anos (N15), enquanto o restante não variou a ponto de atingir os patamares dispostos na Tabela 12.

As razões pelas quais isso se deu circundam a abordagem pessimista de lidar com os mínimos das séries temporais pela janela temporal móvel de 3 valores. Há de se compreender que, embora pessimista, é uma perspectiva que busca superestimar a ocorrência de eventos de escassez para dimensionar o setor elétrico com níveis robustos de confiabilidade, com intenções genuínas de precaução frente ao pior cenário estatisticamente plausível.

5.2.4 Ocorrências de escassez conjunta de recursos eólicos e hídricos

No Capítulo 1 foi discutida a questão da complementaridade intraanual entre recursos eólicos e hídricos no Brasil, tema já bem explorado cientificamente. O que merece escrutínio é se há registro de condições de escassez simultânea (num mesmo ano) destes recursos. Para tanto, para representar a região do subsistema SE/CO bem como a região hidrográfica do Paraná, duas séries temporais contendo dados mensais de vazão de rios (em m³/s) próximas às UHEs de Itaipu e Marimbondo e uma terceira série temporal contendo valores de ENA (em MWmédios) para todo o subsistema SE/CO foram analisadas de maneira similar à seção anterior; médias anuais de cada série foram calculadas e ordenadas de maneira crescente. As 20 piores médias anuais de vazões de rio e de ENA foram dispostas na Tabela 16 em contraste com as médias históricas de cada série temporal em sua totalidade.

Ao contrário dos dados de velocidade do vento, é perceptível que os valores anuais de vazões de rios e de ENA podem ser bem mais inferiores à média total. De modo geral, os piores anos de vazões ficam na faixa de 30 a 50% abaixo da média histórica, chegando a superar 60% no ano de 2014 para Marimbondo, enquanto os piores anos de ENA ficam até mais de 30% abaixo da média.

Ι	taipu [m³/s]	Mar	imbondo [m³/s]	ENA SI	E/CO [MWmédios]
Média	a total: 10373,84	Médi	a total: 1821,07	Médi	a total: 41791,71
Ano	Média anual	Ano	Média anual	Ano	Média anual
1955	5823,08 _{↓43,9%}	2014	671,67 _{↓63,1%}	1971	28115,44 132,7%
1934	5974,33 _{↓42,4%}	1953	975,83 _{↓46,4%}	1955	28550,54 131,7%
1953	6013,33 _{↓42,0%}	1971	1004,92 _{↓44,8%}	1953	29491,58 _{129,4%}
1944	6046,25 141,7%	1955	1017,92 _{↓44,1%}	1954	29926,97 _{128,4%}
1954	6529,75 _{↓37,1%}	1934	1026,58 _{↓43,6%}	1969	30238,12 _{127,6%}
1969	6817,42 _{J34,3%}	2015	1031,25 143,4%	1934	31892,29 _{123,7%}
1949	7088,42 131,7%	1954	1035,42 _{↓43,1%}	1936	32924,53 121,2%
1936	7140,75 _{J31,2%}	2017	1045,92 _{↓42,6%}	1963	33265,76 _{120,4%}
1964	7185,00 _{130,7%}	1933	1059,67 _{↓41,8%}	1944	33380,36 120,1%
1952	7304,58 _{129,6%}	2001	1119,17 _{J38,5%}	2017	33673,11 19,4%
1971	7430,33 _{J28,4%}	1969	1237,33 _{J32,1%}	2019	33746,49 19,3%
1968	7464,67 _{↓28,0%}	1956	1261,00 _{↓30,8%}	1968	33841,74 _{↓19,0%}
1938	7817,92 _{↓24,6%}	1944	1351,25 125,8%	1964	33895,98 _{18,9%}
1933	7851,17 _{↓24,3%}	1968	1372,83 _{124,6%}	2014	34123,77 _{↓18,3%}
1941	7854,58 _{124,3%}	1963	1395,08 _{123,4%}	1939	34279,35 _{18,0%}
1939	7983,17 _{123,0%}	1964	1402,08 123,0%	2001	34841,26 _{↓16,6%}
1959	8033,08 122,6%	1941	1440,33 _{120,9%}	1956	34884,10 _{↓16,5%}
1956	8173,58 _{121,2%}	1990	1445,17 _{120,6%}	1938	35506,41 15,0%
1940	8235,75 _{120,6%}	1936	1463,42 _{19,6%}	1975	36683,90 12,2%
1945	8310,58 19,9%	1975	1553,67 14,7%	1941	37153,78 🚛 11,1%

Tabela 16 – Classificação das médias anuais de vazão natural [m³/s] e relações com as médias totais

Fonte: elaboração própria do autor

Quando cruzadas com a Tabela 15, as informações desvelam os anos em que tanto a disponibilidade de recursos eólicos quanto a de hídricos mostraram-se baixas. É o caso de 1971, presente em praticamente todos as séries temporais NOAA e ERA e que fora considerado o 1°, 11° e o 3° pior ano de recursos hídricos do subsistema SE/CO como um todo, da UHE de Itaipu e de Marimbondo, respectivamente. Há também o ano de 1964, classificado como o 1°, 2° e 3° piores anos das séries N38, N29 e N15, respectivamente; figurando na décima terceira posição classificatória do subsistema SE/CO, nona de Itaipu e décima sexta de Marimbondo.

Individualmente, as séries temporais apresentaram condições piores, mas 1971 e 1964 podem ser considerados os piores anos já registrados, denotando escassez conjunta de recursos eólicos no NE e hídricos no SE. Os dados destes anos foram, portanto, utilizados no Capítulo 7 para simular o atendimento à carga, extrapolando os valores de velocidade do vento de séries temporais para toda a sub-região do NE onde se concentram os parques eólicos. Semelhantemente, os valores de vazão de rio foram extrapolados para o subsistema SE/CO como um todo, representando a geração pela fonte hídrica.

5.3 VARIAÇÕES PERIÓDICAS DE RECURSOS EÓLICOS

Ao longo dos anos compreendidos pelas séries temporais, oscilações regulares em escalas temporais superiores a um ano são passíveis de existência, delineando periodicidades que possuem relevância para projeções futuras de disponibilidade de recursos eólicos. A investigação sobre este tema considerando a região NE como um todo foi facilitada por uma Análise de Componentes Principais seguida de uma Análise Espectral.

5.3.1 Análise de Componentes Principais das séries temporais de dados eólicos

A ACP foi realizada considerando, para cada base de dados, a matriz que contempla todas as séries temporais. Neste primeiro momento, optou-se por desconsiderar os resultados de clusterização de modo a analisar espacialmente toda a região compreendida pelas séries temporais. Todavia, as séries não foram analisadas com seus dados brutos; passaram pelo procedimento de remoção de tendência e ciclo anual (mas não foram normalizadas, já que são univariadas).

A remoção da tendência foi feita mediante cálculo da regressão linear entre os valores de velocidade do vento $(x_{t=1}, x_{t=2}, ..., x_{t=n})$ e os índices dos intervalos de tempo (t = 1, 2..., n). De cada valor x_t foi subtraído o termo do coeficiente angular *b* multiplicado pelo seu respectivo instante *t*. O ciclo anual foi removido pelo cálculo das médias de todos os meses (média de todos os meses de janeiro, de todos os meses de fevereiro etc.), transformado a série temporal (já sem tendência) pela análise de Fourier e subtraindo a somatória dos dois primeiros harmônicos que representam o ciclo anual.

A Figura 72 exemplifica a plotagem da série temporal original M139 sendo decomposta em outras três séries: a tendência, a componente do ciclo anual estimado pela somatória dos dois primeiros harmônicos de Fourier e a série já com tendência e ciclo anual removidos, onde os valores agora parecem estar variando em torno zero, como em um processo puramente aleatório.



Figura 72 - Exemplo de remoção de tendência e ciclo anual na série M139

Fonte: elaboração própria do autor

A ACP foi executada pela função prcomp() do pacote "stats" (R CORE TEAM, 2020) no Rstudio para cada matriz de séries temporais (sem tendência e sem ciclo anual). A configuração utilizada foi pela rota da matriz covariância e não da matriz correlação. Esta função gera automaticamente um número de componentes principais (PCs) equivalente ao número inicial de séries temporais. Entretanto, cada uma das PCs carrega uma porcentagem da variância total em ordem decrescente. Por exemplo, a ACP gerou, para a matriz de 72 séries da NOAA, 72 PCs, sendo que da primeira até a última há uma quantidade de variância explicada que decai quase que exponencialmente. Esta etapa foi feita também com as matrizes da ERA e MERRA, e as proporções da variância total explicada por cada PC estão dispostas na Tabela 17.

Base de		Porcentagem da variância explicada [%]											
dados	PC1	PC2	PC3	PC4	PC5	PC6	PC7						
NOAA	53,6	16,6	9,6	6,2	4,1	2,9	1,6						
ERA	49,6	20,0	7,2	5,6	4,5	2,8	1,8						
MERRA	43,6	19,6	9,2	6,4	3,8	3,1	2,1						

Tabela 17 – Porcentagem da variância total explicada pelas PCs de 1 até 7

Fonte: elaboração própria do autor

Como esperado, a primeira componente principal (PC1) traz consigo a maior parte da variância total intrínseca ao conjunto original de dados. Por exemplo, para resultados da ACP nos dados da NOAA, 53,6% da variância total é explicada pela PC1. Somando os valores da PC1 e PC2, fica evidente que para NOAA e ERA, a variância acumulada se aproxima de 70% e os resultados das PC1 e PC2 para dados MERRA ficam em 63,2% da variância total explicada.

Estes resultados são interessantes para mapear o padrão espacial desta variância. Plotagem dos coeficientes ou elementos dos autovetores das primeiras e segundas PCs de cada base de dados em suas devidas coordenadas de latitude e longitude revelam o padrão espacial que o conjunto de séries apresenta (Figura 73).





Fonte: elaboração própria do autor

Pela Figura 73, observa-se que as 3 bases de dados possuem padrões espaciais bastante semelhantes. A plotagem dos coeficientes dos autovetores da PC1, que explica maior parte da variância total dos dados, aponta para uma convergência do comportamento do vento por toda a região do NE. As manchas vermelhas mais escuras indicam que as séries temporais ali localizadas (tanto em alto mar, bem próximo à costa e entre 0S e 5S de latitude, quanto no interior do continente no entorno da latitude 10S e longitude entre 40W e 45W) têm maior influência na variação total da região colorida e atuam no mesmo sentido, ou seja, se a intensidade do vento aumenta num destes pontos, aumenta também em todos os outros pontos.

O mapeamento dos coeficientes dos autovetores da PC2, que explica uma porção relativamente menor da variância total, já revela um padrão de oposição. Neste modo, percebese manchas vermelhas mais fortes (em alto mar e no interior do continente, ambas em latitudes abaixo de 10S) mas também manchas azuladas no norte da área delimitada, próximo à linha do Equador, regiões estas que carregam consigo maior influência na variância total deste modo espacial. No entanto, atuam em sentidos opostos, ou seja, se a intensidade do vento aumenta na região avermelhada, ela diminui na região azulada, e vice-versa.

A intepretação dos resultados permite afirmar que, na maior parte do tempo, predominam os modos de variação da velocidade do vento em toda a região analisada como descrito pelos mapas das PC1, ou seja, o vento tem um comportamento uniforme na variação de sua magnitude. O segundo modo (mapas das PC2) descreve um comportamento presumivelmente justificado pelo posicionamento da ZCIT, já que esta apresenta deslocamento latitudinal para o norte e para o sul da linha do Equador e interfere na intensidade dos ventos alísios de sudeste que alimentam a região.

5.3.2 Análise Espectral das componentes principais

Além dos autovetores e seus coeficientes, os elementos de cada PC – também conhecidos por *scores* – são também produtos da ACP e podem ser analisados de mesma maneira que uma série temporal qualquer, com o diferencial de explicarem boa parte da variância conjunta sem necessidade de analisar todas as séries temporais originais. De fato, esta redução de dimensionalidade é a principal motivação para executar uma ACP.

No caso das matrizes de séries da NOAA, ERA e MERRA, este procedimento simplifica a investigação de sinais ocultos que indiquem padrões de repetição ou ciclicidades por meio da Análise Espectral (AE). Este recurso matemático, ao invés de ser implementado em todas as séries temporais da região, foi utilizado apenas nas duas primeiras PCs resultantes da ACP de cada base de dados, vide seção anterior. Tal análise foi desenvolvida pela função spec.pgram(), também disponível no pacote de funções "stats" (R CORE TEAM, 2020) para o RStudio.

As PCs, assim como qualquer outra série temporal, passaram por alguns procedimentos: normalização e aplicação de *taper* de 10%, tendo suas extremidades alteradas de modo a aprimorar o resultado da AE. A plotagem das PCs com e sem *taper* pode ser verificada na Figura 74. As PCs, por serem resultantes de matrizes de séries temporais que tiveram seu ciclo anual e tendência removidas, também aparentam descrever processos aleatórios, com valores variando ao redor de zero, assim como se vê no exemplo da Figura 72.



Figura 74 – PCs com aplicação de taper



Entretanto, fossem processos puramente aleatórios, as séries descritas pelas PCs não apresentariam memória, como evidenciado pelos correlogramas da Figura 75. As autocorrelações com diferentes *lags* retornam valores não próximos de zero, o que indica que os processos descritos podem ser autorregressivos. Tal fato possui importância na medida em que as amplitudes espectrais dispostas no periodograma resultante da Análise Espectral devem acompanhar o processo autorregressivo de primeira ordem do tipo ruído vermelho, que guiam os limites de confiança inferior de 5% e superior de 95% e auxiliam na interpretação do periodograma.





Fonte: elaboração própria do autor

Cada PC passou do domínio do tempo para o domínio da frequência pela aplicação da transformada de Fourier, calculando todas as funções harmônicas representativas das PCs e que carregam consigo porções da variação total de cada série. A AE visa identificar determinadas frequências harmônicas em que esta variação é proporcionalmente maior, plotando as amplitudes espectrais num periodograma que representa todo o espectro de frequências, desde a mais baixa (frequência fundamental) até a mais alta (frequência *Nyquist*). Previamente à plotagem, os valores das amplitudes espectrais $I(\omega_k)$ foram suavizados com um filtro de médias móveis do tipo janela de Daniell de 5 valores, deixando a aparência menos ruidosa a fim de enaltecer os picos.

O número *dof* de graus de liberdade – dependente da porcentagem do *taper* e do número de pontos da janela de Daniell – foi calculado em 8,72676. Este valor do *dof* foi utilizado para buscar, em uma tabela qui-quadrado, valores correspondentes ao intervalo de confiança de 95% e de 5% para transformá-los em limites superior e inferior, respectivamente, à curva de ruído vermelho que fora determinada pelo processo autorregressivo de primeira ordem.

Tabelas qui-quadrado comumente trazem números inteiros para graus de liberdade, sendo necessário a realização de uma interpolação simples nas colunas de 0,95 e 0,05 correspondentes aos valores de *dof* equivalentes a 8 e a 9, como ilustra a Figura 76. Após a interpolação, obteve-se os valores de referência de 3,162 (para o nível de confiança de 95%) e 16,533 (para o nível de confiança de 5%), que se situam aproximadamente entre os valores predeterminados para *dof* equivalente a 8 e 9.

dof	0.995	0.975	0.95	0.05	0.025	0.005
1	3.93E-05	0.000982	0.00393	3.841	5.024	7.879
2	0.01	0.0506	0.103	5.991	7.378	10.597
3	0.0717	0.216	0.352	7.815	9.348	12.838
4	0.207	0.484	0.711	9.488	11.143	14.86
5	0.412	0.831	1.145	11.07	12.833	16.75
6	0.676	1.237	1.635	12.592	14.449	18.548
7	0.989	1.69	2.167	14.067	16.013	20.278
8	1.344	2.18	2.733	15.507	17.535	21.955
9	1.735	2.7	3.325	16.919	19.023	23.589
10	2.156	3.247	3.94	18.307	20.483	25.188

Figura 76 - Tabela qui-quadrado e as regiões interpoladas

Fonte: adaptado de Chisquaretable (2020)

Por fim, para cada uma das duas PCs referentes aos dados NOAA, ERA e MERRA, quatro curvas são plotadas em escala logarítmica num periodograma: as amplitudes espectrais suavizadas, o processo de ruído vermelho e os intervalos de confiança com limite superior de 95% e inferior de 5%. No eixo das abcissas estão os valores do inverso da frequência, ou períodos, em número de meses, que correspondem a flutuações. Resultados foram dispostos na Figura 77.



Figura 77 – Periodogramas das PCs de cada base de dados

Fonte: elaboração própria do autor

Todos os periodogramas produzidos pela Análise Espectral foram examinados em busca de alguns picos que superam a curva do limite superior de 95% e que são indicativos de sinais fortes com uma frequência e um período correspondentes.

Vale recordar que o ciclo anual fora previamente removido, do contrário seria certo de encontrar um pico de amplitude espectral no período correspondente de 12 meses. De maneira geral, todos apontam para uma forte variação com período abaixo de 12 meses, dado o pico inicial, porém esta variação se enquadra numa escala intraanual, fugindo do foco da busca por periodicidades que superam a escala temporal anual.

Para dados da NOAA, a PC1 não apresenta picos muito altos, sendo aquele que ocorre aproximadamente com período de 150 meses talvez o mais forte, sugerindo ciclicidade a cada 12,5 anos. Os picos ao final, além de serem relativamente não tão fortes, sugerem variações com períodos muito altos (acima de 900 meses ou 75 anos) e frequência muito baixa, o que não interessa tanto para a indústria eólica, já que o planejamento é realizado numa escala temporal menor. Os picos da PC2 reforçam a ocorrência da ciclicidade de 12,5 anos e três outras periodicidades mais relevantes, uma de aproximadamente 330 meses (~ 27 anos), outra de quase 500 meses (~ 42 anos) e mais uma de cerca de 650 meses (~ 54 anos).

A PC1 dos dados da ERA apresentou um comportamento diferente, com apenas um pico relevante, indicando sinais recorrentes com período de aproximadamente 330 meses – ciclicidade de cerca de 27 anos. Outro pico não tão alto é visível em torno do período de 110 meses. A PC2 reforça a periodicidade de 27 anos, além de sugerir um novo pico em torno do período de 440 meses (~ 37 anos). Em relação aos periodogramas das PCs da NOAA, destacase o pico de 330, reforçando a suposição de ciclos com período de 27 anos.

As duas PCs dos dados MERRA convergem para a existência de dois picos, um com período de 170 e outro de quase 200 meses, sugerindo ciclos de aproximadamente 14 e16 anos. Dois outros consideráveis picos, um de quase 25 meses (~ 2 anos) e outro de quase 50 meses (~ 4 anos) na PC1 e PC2, respectivamente, sugerem variações de escalas temporais menores do que os outros revelados nos demais periodogramas.

Com exceção das possíveis flutuações com ciclos de 27 anos (como evidenciado pelos periodogramas PC2 – NOAA, PC1 – ERA e PC2 – ERA), os resultados não convergem para ciclos comuns. A maioria indica variações de escala temporal que superam aquelas de baixa frequência decadal. Ressalvas à parte, especialmente no tocante à incerteza dos dados de reanálise, a Análise Espectral das PCs de cada base de dados indica que variações podem

ocorrer com baixa frequência, em diferentes períodos cíclicos acima de 10 anos. Estes resultados permitem afirmar que variações na velocidade do vento são passíveis de recorrência em cada um dos períodos revelados nos periodogramas, sugerindo que, no longo prazo, é perfeitamente possível que, de maneira geral e a depender dos sinais cíclicos sobrepostos, a região do NE do Brasil venha a passar por intervalos temporais de baixa disponibilidade de recursos eólicos.

6 FATORES DE INFLUÊNCIA SOBRE A VARIABILIDADE DOS RECURSOS EÓLICOS

A busca por fatores de influência que possam estar associados à variabilidade dos recursos eólicos é etapa descrita neste Capítulo 6. Alguns fenômenos climáticos de padrão oscilatório que ocorrem em regiões oceânicas são bem compreendidos e considerados como modos de variabilidade sobre alterações em variáveis atmosféricas de menor escala espacial. Assim, a identificação de qualquer um destes fenômenos que apresente correlação significativa com os dados eólicos pode ser um potencial dado de entrada no tocante à realização de projeções futuras de disponibilidade de recursos eólicos na região do NE do Brasil.

6.1 CORRELAÇÕES ENTRE COMPONENTES PRINCIPAIS DOS *CLUSTERS* E MODOS DE VARIABILIDADE

Acessória ao objetivo deste capítulo, a ACP foi novamente executada, desta vez sobre o conjunto de séries temporais de cada um dos *clusters* formados no Capítulo 4. O intuito foi extrair as PCs de cada *cluster* que contabilizam pela maior parte da variância total da correspondente matriz de séries temporais. Ao invés de prosseguir com as próprias séries temporais originais, entende-se que a investigação em torno das PCs enaltece a variabilidade de cada *cluster* – variabilidade esta que pode estar associada a alguns fenômenos climáticos e outras variáveis de interesse. Os *clusters* foram organizados em matrizes de séries temporais que tiveram tendência e ciclo anual removidos pela mesma estratégia tal qual previamente explanado na seção 5.3.1.

A função prcomp(), pertencente ao pacote de funções "stats" (R CORE TEAM, 2020) no RStudio, foi então executada pela rota da matriz covariância, sendo estas matrizes de séries temporais (sem tendência e sem ciclo anual) os dados de entrada. Como dados de saída, as PCs de cada *cluster* foram geradas automaticamente em número equivalente ao número inicial de séries temporais da matriz. Tais PCs carregam frações da variância total que as sub-regiões apresentaram ao longo de todo o comprimento das séries temporais.

Os resultados das proporções da variância total explicada pelas PCs numeradas de 1 até 7 estão dispostas na Tabela 18.

Base de	Cluster -		Po	orcentagem o	da variância	explicada [9	%]	
dados	Clusier	PC1	PC2	PC3	PC4	PC5	PC6	PC7
	Q_{N1}	47,9	30,3	11,2	5,9	2,5	0,8	0,6
	Q_{N2}	75,8	14,3	5,0	3,2	0,8	0,3	0,3
NOAA	Q_{N3}	77,5	13,1	5,6	2,7	0,5	0,3	0,2
	Q_{N4}	57,8	19,1	12,5	3,6	1,8	1,5	1,1
	Q _{N5}	73,5	15,5	7,0	2,1	1,3	0,4	0,1
	Q_{E1}	75,2	10,7	5,9	3,0	1,5	0,9	0,7
	Q_{E2}	65,9	15,5	7,8	2,7	2,5	1,5	0,9
ERA	Q_{E3}	75,6	13,1	5,3	3,6	0,8	0,6	0,4
	$Q_{\rm E4}$	90,0	4,8	3,8	0,5	0,4	0,2	0,1
	Q_{E5}	83,4	12,0	2,3	0,9	0,6	0,4	0,2
	Q _{M1}	72,2	15,3	5,0	2,1	1,5	1,1	0,6
	Q _{M2}	56,1	21,6	10,5	3,5	2,6	1,8	1,3
MERRA	Q _{M3}	69,6	10,2	5,7	3,0	1,9	1,9	1,9
	Q_{M4}	62,1	12,4	6,7	4,3	2,1	1,6	1,2
	Qm5	56,9	25,0	5,5	3,5	2,3	1,5	1,0

Tabela 18 – Porcentagem da variância total explicada pelas PCs de 1 até 7 por cluster

Fonte: elaboração própria do autor

Como esperado, a proporção da variância explicada pelas primeiras PCs é bastante superior às demais, chegando a 90% (Q_{E4}) mas também ficando abaixo de 50% (Q_{N1}). Julgouse que utilizar as PC1 e PC2 de cada *cluster* seria suficiente para, genericamente, representar uma variância acumulada de ao menos 75%.

As PCs passaram pelo procedimento de normalização, de modo a trazer as anomalias para uma escala comum antes de rodar a função de cálculo de correlação. Outro aspecto importante foi a segmentação das PCs por estação do ano. Como elas são, na prática, séries temporais, os valores que as compõem são mensais e correspondentes à linha do tempo das séries originais. Por conseguinte, dividiu-se cada série temporal em 4 subséries com os meses de cada estação do ano: dezembro, janeiro e fevereiro (DJF) para o verão; março, abril e maio (MAM) para o outono; junho, julho e agosto (JJA) para o inverno; setembro, outubro e novembro (SON) para a primavera.

Nesta etapa os dados dos modos de variabilidade (MV – vide seção 3.1.2) são utilizados para aferir o grau de associação entre os fenômenos climáticos descritos pelos seus índices quantitativos e as duas primeiras PCs de cada *cluster*. Isto é feito por meio cálculo do coeficiente de correlação de Pearson. A maioria das séries temporais dos índices levantados (AMM, AMO, PDO, NINO 3.4 e MGI) são um pouco mais curtas, com dados de 1950 em diante (o mesmo se vale para as PCs do MERRA, que se iniciam em 1980), portanto o recorte temporal do cálculo

das correlações respeitou a série de menor comprimento. As séries temporais dos índices também foram segmentadas em subséries respectivas às 4 estações do ano, portanto foram calculadas as correlações entre cada PC e cada índice para cada estação do ano: verão, outono, inverno e primavera vide Tabelas 19, 20, 21 e 22, respectivamente. As células das tabelas contendo os valores foram coloridas em escala de tons de vermelho (para correlações negativas) e tons de azul (para correlações positivas); quanto mais distante de 0, mais forte a coloração.

Ao examinar as Tabelas 19, 20, 21 e 22, logo nota-se uma alta dose de informação que merece uma filtragem em busca do que é realmente relevante. Os valores coloridos em tons mais fortes, seja de vermelho, seja de azul, que denotam, para cada estação do ano, alguns padrões de associação entre as regiões representadas pelos *clusters* e os MV. Os valores nas porções à esquerda destas tabelas possuem mais importância, dado que fazem referências às PC1s de cada *cluster* que representam maior porção de sua variância total.

<i>Cluster</i> - variância	Corre	lação do	s MV co	m as PC1 1950 - 20	s das	<i>Cluster</i> - variância	Corre	lação do	s MV co	m as PC2 1950 - 20	es das
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{N1} - 47,9	0,29	0,08	0,17	0,04	0,32	Q _{N1} - 30,3	0,06	0,11	0,12	0,06	0,24
Q _{N2} - 75,8	0,18	0,18	0,12	0,04	0,39	Q _{N2} - 14,3	0,04	-0,11	0,12	0,12	0,01
Q _{N3} - 77,5	0,26	0,05	0,15	-0,06	0,46	Q _{N3} - 13,1	0,34	0,11	0,09	-0,07	0,31
Q _{N4} - 57,8	0,20	0,29	0,06	0,09	0,45	Q _{N4} - 19,1	-0,08	0,05	0,15	0,29	0,03
Q _{N5} - 73,5	0,17	0,25	0,11	0,13	0,45	Q _{N5} - 15,5	-0,04	-0,03	0,16	0,25	0,05
<i>Cluster -</i> variância	Corre séri	lação do es ERA	s MV co (verões 1	m as PC1 950 - 20	s das 10)	<i>Cluster -</i> variância	Corre séri	lação do les ERA	s MV co (verões 1	m as PC2 1950 - 20	es das 10)
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{E1} - 75,2	0,23	0,21	0,10	0,20	0,41	Q _{E1} -10,7	0,21	0,12	-0,10	-0,32	0,11
Q _{E2} - 65,9	0,05	-0,02	0,21	0,19	0,11	Q _{E2} - 15,5	-0,02	0,13	-0,03	0,06	0,10
Q _{E3} - 75,6	-0,08	-0,19	-0,12	-0,11	-0,12	Q _{E3} - 13,1	-0,05	-0,12	-0,07	-0,02	-0,13
Q _{E4} - 90,0	-0,23	-0,08	-0,10	0,04	-0,22	Q _{E4} - 4,8	-0,03	0,02	0,18	0,16	-0,06
Q _{E5} - 83,4	-0,07	0,05	-0,17	-0,04	-0,11	Q _{E5} - 12,0	0,02	0,07	-0,06	-0,05	0,01
<i>Cluster -</i> variância	Corre séries	lação do MERRA	s MV co A (verões	m as PC1 s 1980 - 2	s das 2017)	<i>Cluster -</i> variância	Corre séries	lação do S MERRA	s MV co A (verões	m as PC2 s 1980 - 2	es das 2017)
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{M1} - 72,2	0,10	0,06	-0,04	0,16	0,37	Q _{м1} - 15,2	0,14	0,05	-0,06	-0,23	0,23
Q _{M2} - 56,1	-0,18	-0,03	0,19	0,51	0,07	Q _{M2} - 21,6	-0,05	0,02	0,12	0,07	-0,15
Q _{M3} - 69,6	-0,31	-0,09	0,25	0,66	-0,11	Q _{м3} - 10,2	-0,18	-0,26	0,06	-0,04	-0,23
Q _{M4} - 62,1	0,00	-0,11	0,23	0,23	0,15	Q _{M4} -12,4	0,09	-0,04	-0,12	-0,40	0,11
Q _{M5} - 56,9	0,09	-0,01	-0,05	0,04	0,33	Q _{M5} - 25,0	-0,11	-0,01	0,30	0,14	-0,03

Tabela 19 - Grau de associação entre os MV e as PC1s e PC2s de cada cluster para a estação verão

Fonte: elaboração própria do autor

No verão, os índices AMM, AMO e PDO não apresentam uma relação muito significativa com as PCs independentemente do *cluster* e da base de dados a que pertencem.

Um sinal um pouco mais forte, mas não o suficiente para afirmar a existência de um alto grau de associação entre as variáveis, é detectado para o MGI em todos *clusters* NOAA, em um da base de dados ERA e em dois MERRA, com correlações positivas entre 0,35 e 0,45. O NINO 3.4 apresentou correlações superiores positivas com apenas dois dos *clusters* MERRA (Q_{M2} e Q_{M3}), resultado relativamente insuficiente para afirmar que o fenômeno ENOS em sua fase fria (quente) pode influenciar o no decréscimo (aumento) da velocidade do vento nestas regiões.

Cluster -	Corre	lação do	s MV co	m as PC1	s das	Cluster -	Corre	lação do	s MV co	m as PC2	s das
variância	séries	s NOAA	(outonos	s 1950 - 2	2014)	variância	série	s NOAA	(outonos	s 1950 - 2	014)
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{N1} - 47,9	0,24	0,26	0,09	0,24	0,40	Q _{N1} - 30,3	0,23	0,26	0,14	0,28	0,37
Q _{N2} - 75,8	0,36	0,36	0,08	0,33	0,56	Q _{N2} - 14,3	0,19	0,19	0,20	0,31	0,29
Q _{N3} - 77,5	0,33	0,26	0,15	0,33	0,53	Q _{N3} - 13,1	0,10	0,11	0,01	0,07	0,17
Q _{N4} - 57,8	0,23	0,19	0,03	0,12	0,35	Q _{N4} - 19,1	0,11	0,15	0,08	0,26	0,21
Q _{N5} - 73,5	0,38	0,30	0,03	0,28	0,56	Q _{N5} - 15,5	0,08	0,15	0,06	0,25	0,23
Cluster -	Corre	lação do	s MV co	m as PC1	s das	Cluster -	Corre	lação do	s MV co	m as PC2	s das
variância	série	es ERA (outonos	<u>1950 - 20</u>)10)	variância	série	es ERA (outonos	<u>1950 - 20</u>)10)
PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{E1} - 75,2	0,50	0,35	0,29	0,46	0,67	Q _{E1} -10,7	0,21	0,36	-0,07	0,07	0,18
Q _{E2} - 65,9	0,15	0,15	0,26	0,26	0,21	Q _{E2} - 15,5	0,19	0,31	-0,10	0,13	0,26
Q _{E3} - 75,6	0,07	-0,01	0,06	0,06	0,00	Q _{E3} - 13,1	0,32	0,39	0,07	0,14	0,44
Q _{E4} - 90,0	0,09	0,06	-0,05	0,09	0,02	Q _{E4} - 4,8	0,17	0,29	0,03	0,02	0,25
Q _{E5} - 83,4	0,10	0,07	-0,14	-0,06	0,08	Q _{E5} - 12,0	-0,10	-0,03	-0,06	-0,04	-0,06
Cluster -	Corre	lação do	s MV co	m as PC1	s das	Cluster -	Corre	lação do	s MV co	m as PC2	s das
variância	séries	MERRA	(outono	os 1980 -	2017)	variância	séries	MERRA	(outono	os 1980 -	2017)
PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{M1} - 72,2	0,34	0,12	0,18	0,52	0,60	Q _{м1} - 15,2	0,32	0,35	0,06	0,13	0,29
Q _{M2} - 56,1	0,16	-0,03	0,20	0,55	0,44	Q _{M2} - 21,6	0,14	0,12	-0,02	0,11	0,18
Q _{M3} - 69,6	0,09	-0,08	0,23	0,57	0,34	Q _{M3} - 10,2	0,09	-0,06	0,00	-0,14	0,04
Q _{M4} - 62,1	0,27	0,02	0,17	0,20	0,32	Q _{M4} -12,4	-0,40	-0,06	-0,19	-0,32	-0,45
Q _{M5} - 56,9	0,13	-0,09	0,03	0,17	0,24	Q _{м5} - 25,0	0,14	0,19	0,24	0,38	0,29

Tabela 20 - Grau de associação entre os MV e as PC1s e PC2s de cada cluster para a estação outono

Fonte: elaboração própria do autor

Os resultados para o outono são os que mais convergem para a existência de influência do ENOS (índice NINO 3.4) e do gradiente do Atlântico Tropical (índice MGI), com sinal positivo de moderado (r > 0.4) a forte (r > 0.6) em boa parte dos *clusters* de todas as bases de dados nos primeiros modos da ACP (PC1s) e, com sinal um pouco mais fraco, porém não insignificante, com os segundos modos da ACP (PC2s) – a região do *cluster* Q_{M4} sendo a única exceção. Isto sugere que o ENOS, quando em sua fase quente (fria), e quando as TSMs da porção tropical norte do Atlântico aumentam (diminuem) ao passo que as da porção tropical sul do Atlântico diminuem (aumentam), a velocidade do vento no NE como um todo intensifica

(reduz). Os índices AMO e PDO permaneceram com sinal fraco e o sinal da AMM apareceu apenas em Q_{E1} na sua primeira PC, indicativo de que aquela região pode ser suscetível à variação do modo meridional do Atlântico. No entanto, este resultado não converge com as regiões similares representadas pelos *clusters* Q_{N3} ou Q_{M1} .

<i>Cluster</i> - variância	Corre séries	lação do NOAA	s MV co (inverno	m as PC1 s 1950 - 2	s das 2014)	<i>Cluster</i> - variância	Corre séries	lação do NOAA	s MV co (inverno	m as PC2 s 1950 - 2	ls das 2014)
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{N1} - 47,9	0,16	0,24	-0,09	-0,15	0,10	Q _{N1} - 30,3	0,15	0,20	0,11	0,22	0,29
Q _{N2} - 75,8	0,23	0,38	0,07	0,08	0,32	Q _{N2} - 14,3	0,20	0,18	0,11	0,13	0,29
Q _{N3} - 77,5	0,25	0,33	0,05	0,14	0,38	Q _{N3} - 13,1	-0,01	0,08	-0,15	-0,18	-0,09
Q _{N4} - 57,8	0,16	0,16	0,02	-0,03	0,22	Q _{N4} - 19,1	0,16	0,15	0,02	0,07	0,19
Q _{N5} - 73,5	0,27	0,33	0,04	0,00	0,32	Q _{N5} - 15,5	0,08	0,05	0,10	0,15	0,18
<i>Cluster -</i> variância	Corre série	lação do s ERA (i	s MV co invernos	m as PC1 1950 - 20	s das 010)	<i>Cluster -</i> variância	Corre série	lação do s ERA (i	s MV co invernos	m as PC2 1950 - 20	2s das 010)
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{E1} - 75,2	0,37	0,34	0,22	0,28	0,53	Q _{E1} -10,7	-0,02	0,22	-0,23	-0,41	-0,24
Q _{E2} - 65,9	0,25	0,36	-0,04	-0,07	0,19	Q _{E2} - 15,5	-0,09	-0,07	-0,06	-0,15	-0,11
Q _{E3} - 75,6	0,17	0,30	-0,08	-0,20	0,02	Q _{E3} - 13,1	-0,13	-0,25	0,07	0,04	-0,07
Q _{E4} - 90,0	-0,13	-0,23	0,12	0,19	-0,05	Q _{E4} - 4,8	0,00	0,16	-0,02	0,01	-0,06
Q _{E5} - 83,4	-0,17	-0,25	0,12	0,15	-0,12	Q _{E5} - 12,0	-0,13	-0,06	0,02	0,01	-0,14
<i>Cluster -</i> variância	Corre séries	lação do MERRA	s MV co (inverno	m as PC1 os 1980 -	s das 2017)	<i>Cluster -</i> variância	Corre séries	lação do MERRA	s MV co (inverno	m as PC2 os 1980 -	es das 2017)
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{M1} - 72,2	0,12	-0,02	0,31	0,26	0,38	Q _{M1} - 15,2	-0,05	0,16	-0,40	-0,52	-0,37
Q _{M2} - 56,1	-0,01	-0,13	0,44	0,45	0,34	Q _{M2} - 21,6	-0,09	-0,17	0,16	0,33	0,08
Q _{M3} - 69,6	-0,03	-0,11	0,51	0,48	0,32	Q _{M3} - 10,2	0,19	0,34	-0,14	-0,23	0,07
Q _{M4} - 62,1	0,11	0,17	-0,13	-0,13	0,15	Q _{M4} -12,4	-0,10	-0,13	-0,26	-0,26	-0,38
Q _{M5} - 56,9	0,08	0,18	-0,22	-0,20	0,05	Q _{M5} - 25,0	-0,10	-0,11	0,06	0,20	0,05

Tabela 21 - Grau de associação entre os MV e as PC1s e PC2s de cada cluster para a estação inverno

Fonte: elaboração própria do autor

Embora mais fraco do que no outono, o sinal de MGI permanece em alguns *clusters* NOAA, ERA e MERRA também no inverno, na faixa de 0,35, alcançando 0,53 em Q_{E1} . Os sinais dos dois modos do Pacífico, ENOS e PDO, parecem ganhar relevância no inverno, porém apenas nos *clusters* Q_{M2} e Q_{M3} onde os valores das correlações de Pearson ficam entre 0,45 e 0,50 para as primeiras PCs. Na PC2 de Q_{E1} e Q_{M1} o sinal da PDO e NINO 3.4 são invertidos, sugerindo que o resfriamento das águas do Pacífico na região NINO 3.4 fortaleceria a intensidade do vento nesta sub-região do NE em uma porcentagem menor do tempo (já que a variância explicada pela PC2 é inferior).

Cluster -	Corre séries l	Correlação dos MV com as PC1s das							m as PC2 as 1950 -	28 das 2014)	
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{N1} - 47,9	0,37	0,18	-0,07	-0,13	0,24	Q _{N1} - 30,3	0,19	0,23	0,14	0,22	0,27
Q _{N2} - 75,8	0,29	0,35	0,07	0,07	0,23	Q _{N2} - 14,3	0,12	-0,09	0,04	0,08	0,09
Q _{N3} - 77,5	0,34	0,21	-0,02	0,01	0,37	Q _{N3} - 13,1	0,24	0,09	-0,11	-0,23	0,12
Q _{N4} - 57,8	0,18	0,40	-0,05	-0,05	0,26	Q _{N4} - 19,1	0,08	-0,06	0,09	0,16	0,17
Q _{N5} - 73,5	0,22	0,39	-0,02	0,00	0,27	Q _{N5} - 15,5	0,12	-0,04	0,15	0,20	0,11
<i>Cluster -</i> variância	Corre séries	lação do ERA (pi	s MV co rimavera	m as PC1 s 1950 - 2	s das 2010)	<i>Cluster</i> - variância	Corre séries	lação do ERA (pi	s MV co rimavera	m as PC2 s 1950 - 2	2s das 2010)
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{E1} - 75,2	0,31	0,37	-0,09	0,02	0,38	Q _{E1} -10,7	0,21	0,27	-0,10	-0,15	-0,12
Q _{E2} - 65,9	0,12	0,26	-0,12	-0,05	0,16	Q _{E2} - 15,5	-0,01	0,01	0,11	0,10	-0,10
Q _{E3} - 75,6	0,19	0,31	-0,06	0,05	0,25	Q _{E3} - 13,1	-0,25	-0,12	0,03	0,13	-0,13
Q _{E4} - 90,0	0,13	0,10	0,05	-0,01	0,12	Q _{E4} - 4,8	0,38	0,28	-0,14	-0,15	0,21
Q _{E5} - 83,4	-0,06	-0,10	0,11	0,06	-0,07	Q _{E5} - 12,0	0,08	0,23	-0,07	0,05	0,12
<i>Cluster -</i> variância	Corre séries N	lação do IERRA (s MV co (primave	m as PC1 ras 1980	s das - 2017)	<i>Cluster -</i> variância	Corre séries N	lação do IERRA (s MV co (primave	m as PC2 ras 1980	2s das - 2017)
explicada PC1 [%]	AMM	AMO	PDO	NINO 3.4	MGI	explicada PC2 [%]	AMM	AMO	PDO	NINO 3.4	MGI
Q _{м1} - 72,2	0,09	0,05	-0,09	-0,14	0,23	Q _{M1} - 15,2	-0,16	-0,03	-0,19	-0,31	-0,29
Q _{M2} - 56,1	-0,04	0,05	0,18	0,11	0,22	Q _{M2} - 21,6	0,00	-0,10	0,19	-0,03	-0,13
Q _{M3} - 69,6	-0,09	-0,03	0,26	0,27	0,16	Q _{M3} - 10,2	0,01	0,05	-0,19	-0,10	0,03
Q _{M4} - 62,1	0,04	0,15	-0,11	-0,13	0,10	Q _{M4} -12,4	-0,07	-0,15	-0,04	-0,02	-0,11
Q _{M5} - 56,9	0,07	0,03	-0,22	-0,25	0,06	Q _{M5} - 25,0	-0,17	-0,02	0,08	0,00	-0,06

Tabela 22 - Grau de associação entre os MV e as PC1s e PC2s de cada cluster para a estação primavera

Fonte: elaboração própria do autor

O sinal da MGI continuou positivo e convergindo para todos as PC1s dos *clusters* na primavera, mas tornou-se enfraquecido. Alguns modos do Atlântico (AMM e AMO) mostramse fracamente correlacionados com as séries NOAA e ERA, também insuficiente para afirmar influência destes MV sobre o NE nesta estação.

É pertinente ressaltar que valores de correlação de Pearson, ainda que altos (próximos de +1 ou -1), não implicam causalidade, o que significa que a afirmação de que os fenômenos climáticos descritos pelos índices avaliados provocam o aumento/decréscimo da intensidade da velocidade do vento (e vice-versa) seria equivocada. O grau de associação resultante apenas sinaliza para um comportamento minimamente regulado: para correlações positivas, em resposta ao aumento (decréscimo) das TSMs na área oceânica em questão, quando se aumenta (diminui) um dos índices, a velocidade média do vento nas regiões correspondentes aos *clusters* também aumenta (diminui). Existe ainda a possibilidade de que as correlações sejam espúrias, decorrentes de valores das séries temporais meramente coincidentes. Por esta razão, o trabalho

com diferentes bases de dados faz-se necessário; é mais difícil que uma coincidência que implique em correlações espúrias seja replicada por conjuntos de dados distintos.

Por conseguinte, tomou-se apenas os resultados das PC1s de cada *cluster* que ficaram acima de 0,4 para o preparo da Figura 78, evidenciando quais sinais foram mais expressivos e em que estação do ano para, por fim, identificar sub-regiões cujos padrões de correlação são similares. Observa-se que, mesmo que as regiões dos *clusters* não sejam perfeitamente coincidentes, alguns resultados são convergentes: parece haver influência do ENOS e do gradiente do Atlântico tropical na porção que vai do centro ao nordeste da área retangular colorida (onde ficam os *clusters* Q_{N3}, Q_{E1}, Q_{M1} e partes de Q_{M1} e de Q_{M2}) principalmente durante o outono.

NINO 3.4 e MGI no outono MGI no verão NINO 3.4 no verão e no outono AMM. NINO 3.4 e MGI no outono NINO 3.4 e PDO no inverno MGI no verão e no outono MGI no inverno NOAA MERRA ERA Q_{N3} **Q**E1 **Q**_{M1} Q_{M3} Q_{N5} ON 105 105 **O**E2 Ом5 Q_{N4} Q_{N1} Ом QE4 O ЭM 205 MGI no verão MGI no outono NINO 3.4 no verão AMO no inverno NINO 3.4 e MGI no outono NINO 3.4 e PDO no inverno

Figura 78 – Índices e estações do ano com sinal da correlação acima de 0,4 para as PC1s dos clusters

Fonte: elaboração própria do autor

De modo a avaliar se estes valores de correlação são estatisticamente significantes, o cálculo da razão t_{ratio} foi realizado considerando o número N de pares de dados utilizados em cada cálculo do coeficiente r de correlação de Pearson. Para NOAA, ERA e MERRA, N foi equivalente a 195, 183 e 114, respectivamente, fazendo referência ao número de meses em cada uma das quatro estações (verão, outono, inverno e primavera). O número *dof* de graus de liberdade, que depende apenas de N, permitiu guiar pela análise da tabela de valores críticos de t_{ratio} para extrair os valores de referência que consideram o nível de confiabilidade de $\alpha = 0,001$ (TTABLE, 2020). Os resultados podem ser conferidos na Tabela 23, deixando evidente que os resultados obtidos de t_{ratio} superam os valores críticos tabelados de t_{ratio} e, portanto, estão dentro do nível de confiabilidade de 0,001 - são estatisticamente significantes.

$$t_{ratio} = r \frac{\sqrt{dof}}{\sqrt{1 - r^2}} \tag{76}$$

$$dof = N - 2 \tag{77}$$

Tabela 23 – Teste de significância estatística da tabela t para valores da correlação de Pearson r

Base de dados	N e <i>dof</i>	<i>Cluster</i> - variância explicada PC1 [%]	Modo de variabilidade	Estação do ano	Correlação de Pearson r	t _{ratio}	Valor t_{ratio} crítico para $\alpha = 0.001$
NOAA	195 e 193	Q _{N3} - 77,5	MGI	Verão	0,46	7,150	3,342
		Q _{N4} - 57,8	MGI	Verão	0,45	6,963	
		Q _{N5} - 73,5	MGI	Verão	0,45	7,079	
		Q _{N1} - 47,9	MGI	Outono	0,40	6,065	
		Q _{N2} - 75,8	MGI	Outono	0,56	9,352	
		Q _{N3} - 77,5	MGI	Outono	0,53	8,605	
		Q _{N5} - 73,5	MGI	Outono	0,56	9,454	
		Q _{N4} - 57,8	AMO	Primavera	0,40	5,977	
ERA	183 e 181	Q _{E1} - 75,2	MGI	Verão	0,41	6,048	3,345
		Q _{E1} - 75,2	AMM	Outono	0,50	7,767	
		Q _{E1} - 75,2	NINO 3.4	Outono	0,46	6,970	
		Q _{E1} - 75,2	MGI	Outono	0,67	12,142	
		Q _{E1} - 75,2	MGI	Inverno	0,53	8,409	
MERRA	114 e 112	Q _{M2} - 56,1	NINO 3.4	Verão	0,51	6,218	3,380
		Q _{M3} - 69,6	NINO 3.4	Verão	0,66	9,351	
		Q _{M1} - 72,2	NINO 3.4	Outono	0,52	6,416	
		Q _{M1} - 72,2	MGI	Outono	0,60	7,925	
		Q _{M2} - 56,1	NINO 3.4	Outono	0,55	6,950	
		Q _{M2} - 56,1	MGI	Outono	0,44	5,220	
		Q _{M3} - 69,6	NINO 3.4	Outono	0,57	7,313	
		Q _{M2} - 56,1	PDO	Inverno	0,44	5,213	
		Q _{M2} - 56,1	NINO 3.4	Inverno	0,45	5,407	
		Q _{M3} - 69,6	PDO	Inverno	0,51	6,287	
		Q _{M3} - 69,6	NINO 3.4	Inverno	0,48	5,800	

Fonte: elaboração própria do autor

Na medida em que os fenômenos climáticos explorados apresentam padrões oscilatórios com frequências interanuais, as informações levantadas nesta seção são de suma importância para o planejamento do dimensionamento e operação do sistema elétrico brasileiro. Se os MV são fatores de influência sobre a intensidade dos recursos eólicos nas regiões apontadas, a previsão de suas ocorrências pode, então, ser incorporada à previsão de recursos eólicos em escala interanual.

6.2 OS FATORES DE INFLUÊNCIA DURANTE EVENTOS DE ESCASSEZ DE RECURSOS EÓLICOS

Seguindo o foco da pesquisa de explorar os eventos de escassez de recursos eólicos, buscou-se verificar quais eram os sinais dos fenômenos descritos pelos índices quando tais eventos (já elencados no Capítulo 5) ocorreram. Esta análise foi a maneira encontrada para reforçar evidências de influência (ou não) de alguns dos MV sobre a disponibilidade de recursos eólicos, em especial no que tange a ocorrência de eventos extremos.

Ao plotar as séries temporais dos índices NINO 3.4 e MGI (Figura 79), basta uma rápida inspeção visual para identificar que os comportamentos não são regulares, apresentam flutuações distintas: ora são opostos – com sinais positivos para um e negativos para outro (ex: por volta de 1955 e de 2010) – ora são coincidentes – com sinais equivalentes positivos (ex: 1958) ou negativos (ex: 1974 e 1985).





Uma vez revisitada a Tabela 15 (Capítulo 5) e tendo suas informações cruzadas com os sinais dos índices NINO 3.4 e MGI, elaborou-se a Tabela 24 com vistas para a identificação dos sinais destes MV nos anos cuja intensidade do vento ficou abaixo de sua média histórica nas sub-regiões representadas pelos *clusters* de séries temporais. Alguns dos anos indicados na Tabela 15 foram recorrentes para diferentes séries temporais de diferentes bases de dados, o que denota convergência de períodos de baixa disponibilidade em boa parte do NE do Brasil.

Estes anos que apareceram por repetidas vezes na Tabela 15 foram listados na Tabela 24, juntamente aos sinais do NINO 3.4 e do MGI. Se positivo (célula azul), significa que a região representada pelo índice em questão apresentou anomalias positivas de TSM naquele dado ano, portanto as águas oceânicas ficaram com temperaturas acima da média histórica. Se o sinal for negativo (célula vermelha), significa que a região oceânica apresentou anomalias negativas de TSM naquele ano e a superfície do mar estava mais fria do que a média histórica.

Ano de escassez de recursos eólicos	Sinal NINO 3.4	Sinal MGI	
1963	+1	+1	
1971	- 1	-1	
1972	+1	-1	
1973	-1	-1	
1974	-1	-1	
1975	-1	-1	
1984	-1	-1	
1985	-1	-1	
1989	-1	-1	
1994	+1	-1	
2000	-1	-1	
2009	+1	-1	

Tabela 24 - Sinal dos índices NINO 3.4 e MGI em anos de escassez eólica

Fonte: elaboração própria do autor

O padrão que transparece é o de influência do fenômeno ENOS e do gradiente do tipo dipolo na porção tropical do Oceano Atlântico – ambos, quando em fase fria, parecem modular a redução da intensidade da velocidade do vento na região NE do Brasil. De maneira genérica, pode-se dizer que a partir da década de 1970 condições como estas aconteceram ao menos uma vez por década. Todavia, a periodicidade de ocorrência desta concomitância das fases frias e o decorrente decréscimo na disponibilidade de recursos eólicos não é nada regular, como deflagram as flutuações na Figura 79.

Nem mesmo a intensidade de um evento como este possui regularidade. Se nos eventos de 1974 a 1975 e de 1984 a 1985 as condições foram um pouco mais intensas e persistentes, se associando aos piores anos de escassez eólica (vide Tabela 15), o mesmo não pode ser dito para os anos de 1989 e 2000. Há também o fato de que alguns anos de baixa disponibilidade de recursos eólicos não necessariamente ocorreram em concomitância com sinais negativos das fases frias dos MV, a exemplo de 1963, 1994 e 2009.

Como visto, velocidade do vento no NE e TSM sobre o Pacífico equatorial e sobre o Atlântico tropical são variáveis correlacionadas positivamente, em especial durante os meses do outono. A simples análise ano a ano dos MV e dos eventos de escassez eólica corrobora com esta afirmação. Portanto, é prudente concluir cautelosamente que o ENOS e o padrão dipolo do Atlântico Tropical, quando em fases simultaneamente negativas, podem fazer intensificar condições preexistentes de escassez eólica na região NE. Estes resultados são relevantes ao grande tema desta pesquisa, que tangencia a geração de energia, haja vista que um possível futuro evento de escassez eólica pode prejudicar a oferta de energia no Brasil.
7 A GERAÇÃO DE ENERGIA EM EVENTOS DE ESCASSEZ DE RECURSOS EÓLICOS E HÍDRICOS

A última etapa da pesquisa consistiu em converter todo o conjunto informacional descrito nos capítulos anteriores para uma linguagem energética. Partiu-se do pressuposto que a demanda por energia elétrica apresentará incrementos ano a ano, de maneira que estressar as condições de escassez de recursos em cenários de geração de energia para examinar o comportamento do atendimento à carga anual no Brasil pode proporcionar subsídios à tomada de decisão relativa ao planejamento e dimensionamento de longo prazo do parque gerador nacional. Nesta fase, não somente recursos eólicos, mas também os hídricos foram explorados, haja vista a associação destes com fenômenos e variáveis atmosféricas (de maneira similar ao que ocorre com os recursos eólicos) e em função da preponderância de hidrelétricas na geração de energia elétrica do sistema elétrico brasileiro, como previamente mencionado no capítulo introdutório.

7.1 CARGA DE ENERGIA NO BRASIL

O crescimento na demanda por energia elétrica no Brasil pode ser dado como certo. Presumivelmente, tal situação se materializará em bases mais eficientes em todos os setores consumidores (residencial, industrial, comercial), com ganhos em eficiência energética da ordem de 5 a 10% ao longo das décadas de 2020 e 2030 e com potencial de representar uma redução de 300 TWh de energia elétrica até o final da década de 2030 (EPE, 2015; DANTAS et al., 2017).

Considerando um alto crescimento de 3,1% ao ano, a demanda prevista pode superar a faixa de 1600 TWh por ano para a década de 2050, aumento a ser observado tanto pela transmissão e distribuição convencional quanto mediante autoprodução de energia *in loco* e geração distribuída (geração para suprimento e consumo na própria unidade consumidora com possível utilização da rede elétrica de concessionárias de distribuição e/ou de transmissão para exportar o excedente energético) – vide Figura 80. Assim, é permissível afirmar que, ainda que as medidas de eficiência tenham seu valor e importância, elas não eliminarão a necessidade de expansão do parque gerador do sistema elétrico brasileiro, provendo nova capacidade instalada capaz de atender a demanda crescente (EPE, 2015; DANTAS et al., 2017).



Figura 80 - Consumo final por eletricidade: projeções de 2013 a 2050

Por sua vez, a carga de energia corresponde à energia total a ser produzida por unidades geradoras a fim de atender a demanda dos consumidores finais e suprir inevitáveis perdas elétricas decorrentes da transmissão e distribuição. Garantir o atendimento ao consumo final e à carga projetada para 2050 (Figura 81) irá requerer investimentos em novas unidades geradoras com capacidade instalada suficiente (EPE, 2015; DANTAS et al., 2017).





Na maior parte dos países que possuem geração pela fonte hídrica, esta é considerada como complementar à geração por fontes térmicas. O acompanhamento instantâneo da carga é conduzido pelas grandes unidades térmicas geradoras, com UHEs, em sua maioria, operando a fio d'água ou com baixa capacidade de armazenamento para atendimento dos picos na demanda e para reduzir o estresse das plantas térmicas, buscando reduzir também o custo marginal de geração. No Brasil o oposto é verificado: a fonte hídrica é preponderante sobre as térmicas. O acompanhamento da carga se dá pelas grandes UHEs com reservatórios de alto volume de armazenamento, enquanto boa parte das térmicas é acionada na base, sustentando o balanço

energético entre oferta e demanda (Figura 82a). No caso, a água é um bem gratuito, logo o custo marginal de operação é determinado apenas pelas térmicas (ZURN et al., 2017).

Dado que o padrão da carga para um dado dia (ou mesmo um ano) é relativamente previsível, a intensificação da inserção da geração pela fonte eólica pode causar um desequilíbrio em função da sua natureza estocástica de curto prazo. Não obstante, em sistemas elétricos com predominância hídrica, a energia eólica, por ser não-despachável, designa que a geração instantânea seja injetada no SIN automaticamente e, simplificadamente³, ela pode ser modelada como uma "carga negativa", fazendo reduzir a carga efetiva que é massivamente atendida pelas UHEs, poupando água em seus reservatórios (Figura 82b). (ZURN et al., 2017).

Figura 82 – Curva de carga diária em sistemas predominantemente hídricos: sem energia eólica (a) e com participação da energia eólica (b)



Fonte: adaptado de Zurn et al. (2017)

Esta intepretação tem sua utilidade, sendo interessante subtrair os dados de geração eólica dos dados da carga em ordem cronológica, obtendo a carga efetiva (ou líquida) cujo abastecimento seria realizado pelas outras fontes. Tomando a demanda futura para um dado ano, diferentes desdobramentos decorrentes de cenários de geração eólica em conjunto com a hídrica podem ser levantados para análise.

Ao observar o perfil anual da curva de carga com dados mensais para os anos mais recentes no Brasil (de 2014 a 2019, vide Figura 83), é nítido o perfil sazonal, com picos nos meses de janeiro, março e outubro e vales nos meses de junho, julho e agosto. Ademais, a demanda por energia aumenta ano após ano, ainda que irregularmente. A expansão da capacidade instalada do sistema elétrico deverá minimamente acompanhar esta evolução com ritmo equivalente – caso o objetivo seja manter o nível de segurança energética.

³ Tal simplificação parte de uma negligência às possíveis perdas e congestões das linhas de transmissão por contingências e/ou sobrecarga, já que os centros de carga são normalmente distantes dos principais pontos de geração (ZURN et al., 2017).



Figura 83 - Curva de carga anual com dados mensais para o período de 2014 a 2019



O detalhamento do atendimento à carga foi exemplificado para o ano de 2019 (Figura 84). É notável a grande participação da fonte hídrica, que, embora segregada no gráfico, inclui o desempenho da UHE de Itaipu, responsável por atender cerca de 10% da carga anual. A faixa verde representa o que pode ser considerada a "carga negativa", que é atendida de maneira nãodespachável pela geração eólica. O restante fora atendido por todas as outras fontes em conjunto (contemplando térmicas a combustíveis fósseis, biomassa ou nuclear, fotovoltaica etc.).



Figura 84 - Atendimento à carga do sistema elétrico do Brasil em 2019

Fonte: ONS (2020a)

O propósito dos cenários conceituados nesta pesquisa envolve a ilustração da sensibilidade do sistema elétrico à variabilidade interanual dos recursos eólicos e hídricos. Nesse sentido, o objetivo específico requer o exame do atendimento da carga futura diante de múltiplos cenários que se distinguem entre si por composições variadas de capacidades instaladas de geração pelas fontes eólica e hídrica, configurando participações variadas de cada uma delas na matriz elétrica nacional. Estes cenários contemplam os plausíveis eventos de escassez de recursos energéticos predeterminados pela primeira etapa da pesquisa, entendidos como um fator limitante no que toca a geração de energia e atendimento da carga futura.

7.2 CONCEPÇÃO DE CENÁRIOS

O panorama futuro do setor elétrico num horizonte de tempo de longo prazo requer considerações acerca do montante populacional e do nível de atividade econômica do país. Padrões tecnológicos, hábitos culturais, restrições impostas pelo meio ambiente e a disponibilidade de recursos possuem também suas doses de importância. A análise torna-se dificultosa na medida em que cada uma destas dimensões possui um razoável nível de incerteza, por vezes de tipo não quantificável. As características inerentes ao setor elétrico do Brasil, como padrões de consumo energético, diretrizes de expansão da oferta por eletricidade, ritmo do desenvolvimento e da disseminação de *smart grids* ou veículos elétricos, aspectos regulatórios e comerciais, ambiente de negócios empresariais etc., deverão coevoluir ao longo do tempo de modo nada determinístico. A presença destas e de outras variáveis incertas acentuam a complexidade do sistema e da tarefa de estabelecer previsões (DANTAS et al., 2017).

A concepção de cenários é, nesse sentido, uma opção interessante para examinar a interação de algumas destas variáveis, constituindo uma importante ferramenta analítica para auxílio à tomada de decisões a nível governamental com vistas ao desenvolvimento de estratégias de negócios e de alocação de recursos. Cenários podem ser compreendidos como trajetórias comportamentais de sistemas que apresentam variáveis interdependentes não livres de incerteza. A ideação de cenários requer uma delimitação apropriada do problema, do estado inicial das variáveis e definição de hipóteses, critérios e premissas consistentes, com vistas à obtenção de cenários relevantes e plausíveis que permitam uma compreensão realística dos comportamentos possíveis do sistema em questão ao longo do tempo. A técnica requer esforço para que se reduza ao mínimo o número possível de trajetórias, haja vista que o número de

cenários cresce à medida em que se incorpora um maior número de variáveis consideradas incertas, fazendo com que simplificações sejam, por vezes, necessárias (DANTAS et al., 2017).

Os cenários idealizados nesta pesquisa foram atrelados à disponibilidade interanual de recursos eólicos e hídricos que, grosso modo, representam, em conjunto, uma elevada fração da oferta e geração de energia elétrica no Brasil. Assim, variáveis como padrões de consumo, nível de atividade econômica, tecnologias em uso, tarifas de energia elétrica e outras variáveis associadas à demanda por eletricidade não foram modeladas. O objetivo foi apenas testar a sensibilidade da oferta de energia em condições específicas de escassez de recursos eólicos e hídricos. Para tal, os resultados da seção 5.2.3 foram resgatados; são informações sobre os anos em que as séries temporais de velocidade de vento e de vazão de rio ficaram concomitantemente abaixo das médias históricas. Dois anos em específico foram explorados: 1964 e 1971.

O cenário I, considerado como de referência, foi o do ano base de 2019. Neste ano a capacidade instalada por todas as fontes somadas era de 170 GW. O consumo por eletricidade chegou a 482 TWh, sendo a carga equivalente a 594 TWh. Pelo exame da Figura 85, percebese que em 2019 toda a geração hídrica do subsistema SE/CO atendeu aproximadamente 40% da carga (sendo 10% de Itaipu e 30% de todas as outras UHEs). Por sua vez, a geração eólica do subsistema NE atendeu cerca de 5% no primeiro semestre e 10% no segundo semestre, tendo capacidade instalada de 15 GW ao longo de 2019, dos quais 12,6 GW concentrados apenas no NE. Combinadas, as fontes eólica no subsistema NE e hídrica no SE/CO atenderam praticamente metade da carga de 2019 (ABEEÓLICA, 2020; ONS, 2020a).



Figura 85 – Atendimento à carga em 2019 pelas fontes eólica (NE) e hídrica (SE/CO)

Fonte: ONS (2020a)

O ano base serviu para efeitos comparativos com os cenários II e III. Estes envolveram a simulação da geração pelas fontes eólica (subsistema NE) e hídrica (subsistema SE/CO) utilizando os dados de velocidade do vento e de vazão de rio dos anos de 1964 e 1971, respectivamente. Especificamente, visou-se compreender como seria o atendimento à carga verificada para o ano de 2019 utilizando os mesmos valores de capacidade instalada destas fontes daquele ano caso a disponibilidade de recursos eólicos e hídricos fosse similar ao que indicam os dados de 1964 e 1971.

Três cenários futuros (IV, V e VI) foram conceituados tomando por base a carga futura a ser atendida pela rede do SIN (barras amarelas na Figura 80, desconsiderando autoprodução de energia elétrica) que fora modelada em estudo para o ano de 2050: 1340 TWh (EPE, 2015). Se em 2050 a carga pode alcançar valores equivalentes à carga de 2019 multiplicada por um fator de ~ 2,25 o planejamento da expansão deve considerar diversas fontes para compor o parque gerador nacional objetivando reduções de investimentos, de custos operacionais, de emissões de GEE, além de prover segurança energética. Evidentemente que tecnologias distintas das existentes podem emergir ao longo dos anos e o universo de soluções viáveis pode tornar-se ainda maior (geração de energia por fusão nuclear controlada como exemplo extraordinário). Entretanto este exercício de considerar as fontes já exploradas é relevante do ponto de vista do alto potencial instalável e disponível no Brasil, em especial para a fonte eólica.

O Brasil não possui, até o final do ano de 2020, capacidade instalada de aproveitamento de recursos eólicos em alto mar (*offshore*) – apenas em terra firme (*onshore*). Acredita-se que futuramente a exploração *offshore* se materialize, tendo em vista custos decrescentes, curva de aprendizado, novas tecnologias materiais, infraestrutura e o alto potencial instalável, que pode superar de 5 TW de capacidade em alturas superiores a 100 m apenas no NE (EPE, 2020b). Embora em menor escala, o potencial de aproveitamento *onshore* também é significativo: difundidos pelas regiões NE e S, há cerca de 250 GW a 100 m ou 440 GW a 150 m (EPE, 2018).

O Plano Nacional de Energia 2050 apresenta seus próprios cenários futuros de composição da matriz elétrica brasileira em diferentes condições que levam em conta aspectos regulatórios, econômicos, de estrutura financeira e de dependência da exploração ou não da fonte hídrica em larga escala. A capacidade instalada de energia eólica total pode ficar na faixa de 110 a 195 GW, aumentando a participação desta fonte para entre 22 e 33% do total em potência instalada. Pode até mesmo superar os 200 GW, caso não se permita a expansão da geração hídrica, principalmente na região Norte. Não obstante, o quanto do montante de eólicas remeteria às *offshore* é um valor bastante incerto (EPE, 2020c).

Os cenários futuros se pautaram pela proposição de que o potencial eólico pode ser mais ou menos explorado, com presença mais ou menos expressiva de geração eólica *offshore*. Assumiu-se que, caso a energia eólica apresente uma trajetória "moderada", o total instalado ficaria em 110 GW, sendo 100 GW *onshore* e 10 GW *offshore*. Em um cenário intermediário, o total ficaria em 150 GW dos quais apenas 30 GW seriam *offshore*. Em uma trajetória mais intensificada, a potência total alcançaria 190 GW, com 140 GW *onshore* e 50 GW *offshore*.

Quanto ao potencial hidrelétrico, este já foi bem explorado no Brasil. Ainda assim, restam cerca de 52 GW disponíveis para projetos de UHEs com capacidade instalada superior a 30 MW e 16 GW para projetos de potência inferior a 30 MW. Entretanto, a expansão da exploração de recursos hídricos possui limitações relacionadas aos impactos socioambientais decorrentes de projetos hidrelétricos de grande porte, haja vista que 40 GW deste potencial situam-se em unidades de conservação, territórios indígenas e quilombolas, sobretudo na região Norte (EPE, 2018). Na região SE/CO existe quantidade razoável de potencial instalável que eventualmente pode vir a ser explorada no horizonte temporal até 2050. Sem embargo, isto foi considerado improvável. Desta forma, os cenários futuros fixaram a capacidade instalada de geração por recursos hídricos na região SE/CO em 63,2 GW. É uma simplificação para melhor trabalhar as comparações entre 2050 e 2019, já que a representatividade dos dados históricos de ENA estão atreladas ao dimensionamento e a aspectos técnicos do parque gerador de 2019.

A Tabela 25 reúne as características dos 6 cenários distinguidos nos parágrafos anteriores. Ressalta-se que não faz parte desta pesquisa estabelecer a composição total da matriz elétrica brasileira, o que justifica a agregação das capacidades das demais fontes e subsistemas num valor só, bem como a indefinição destas e da capacidade instalada total do SIN nos cenários IV, V e VI. O exercício limitou-se a avaliar a geração pelas fontes eólica e hídrica nos seus respectivos subsistemas preponderantes em condições de escassez conjunta destes recursos.

			Capacidade instalada [GW]						
Cenário		Carga [TWh]	Total	Fonte eólica subsistema NE		Fonte hídrica	Demais fontes e		
				onshore	offshore	subsistema SE/CO	subsistemas		
Ι	Base - 2019	594	167	12,6	0	63,2	91,2		
II	Escassez 1964	594	167	12,6	0	63,2	91,2		
III	Escassez 1971	594	167	12,6	0	63,2	91,2		
IV	Futuro - 2050a	1340	?	100	10	63,2	?		
V	Futuro - 2050b	1340	?	120	30	63,2	?		
VI	Futuro – 2050c	1340	?	140	50	63,2	?		

Tabela 25 - Características dos cenários conceituados

Fonte: elaboração própria do autor

7.3 CONVERSÃO ENERGÉTICA

7.3.1 Energia proveniente de recursos eólicos

Modelos de conversão de uma variável atmosférica para energia podem se apoiar em sistemas genéricos de produção de energia. Tal generalização pode ser uma fonte de incerteza, já que há uma grande variedade de soluções tecnológicas com suas particularidades, sobretudo quanto ao relacionamento entre a variável atmosférica em questão e sua transformação em energia elétrica. São relacionamentos tradicionalmente não-lineares e variam de uma planta geradora para outra. A título de exemplificação, aerogeradores que compõem parques eólicos se diferenciam entre si conforme limiares de *cut-in* e *cut-off*, curva de potência e altura do rotor, enquanto hidrelétricas possuem tipos diferentes de turbinas, altura da barragem e da queda e tipo de reservatório. Consequentemente, modelos devem levar cada especificidade em consideração para representar a realidade (RAYNAUD et al., 2018).

Para uma turbina eólica, a produção média de energia E_{wind} por um tempo T de operação depende de sua curva de potência – característica específica que prevê a geração de energia para cada valor de velocidade do vento – e de valores $v_{in} e v_{out}$, para partida (*cut-in*) e parada (*cut-off*) no giro das pás. A potência de saída da turbina varia com o cubo da velocidade do vento v que sopra à altura do rotor num dado instante t. Ainda, seu cálculo leva em conta a área circular A que é varrida pelas pás de comprimento equivalente ao raio do círculo, depende de um coeficiente de potência⁴ C_p e da densidade do ar ρ_{ar} no local. Assumindo 100% de disponibilidade da turbina neste tempo T de operação, as equações abaixo descrevem este comportamento:

$$E_{wind} = T \int_{v_{in}}^{v_{out}} f(v) P(v)$$
(78)

$$P(v) = \frac{1}{2} \cdot C_p \cdot \rho_{ar} \cdot A \cdot v(t)^3$$
(79)

Os valores de velocidade do vento podem ser assumidos conforme uma PDF f(v) que, como anteriormente mencionado, comumente assume uma curva de Weibull que bem

⁴ A potência do vento não é totalmente convertida em potência útil. Ela varia com a velocidade do vento conforme características aerodinâmicas das pás e de outros elementos mecânicos da turbina (CUSTÓDIO, 2013).

representa a distribuição de frequência de dados horários de velocidade do vento. No entanto, em se tratando de dados de médias mensais, estes foram inseridos diretamente nas equações para definir produções também mensais médias. O tempo *T* de operação foi equivalente ao número de horas de cada mês do ano (744 horas para meses com 31 dias, 720 horas para meses com 30 dias e 672 para fevereiro). A densidade do ar ρ_{ar} considerada foi de 1,225 kg/m³ conforme indica a *International Standard Atmosphere* para as condições de 15° Celsius a nível do mar (GRAUL e POPPINGA, 2018).

Dentre as opções disponíveis no banco de dados do pacote "bReeze" (GRAUL e POPPINGA, 2018) do RStudio, a turbina eólica Vestas V112 3.0 MW foi eleita como modelo para conversão de velocidade do vento em energia para dados *onshore*. Tomou-se como base um valor de potência nominal compatível com boa parte dos aerogeradores em operação no Brasil e que seja adequada à altura de 100 m dos valores de velocidade do vento. Este aerogerador possui potência nominal de 3,075 MW, velocidade de *cut-in* de 3 m/s e *cut-out* de 25 m/s, sendo capaz de atingir sua potência nominal com velocidade de 12 m/s (Figura 86 e Tabela 26). Seu conjunto de 3 pás conferem diâmetro de 112 metros e área de varredura de 9852 m² (WIND TURBINE MODELS, 2015a).



Figura 86 - Curva de potência Vestas V112 3.0 MW



Para converter dados representativos de regiões *offshore*, a turbina Enercon E126 7.5 MW foi eleita também por estar disponível no banco de dados do pacote "bReeze" (GRAUL e POPPINGA, 2018) do RStudio. Seguindo as características de turbinas *offshore*, sua potência nominal é elevada, ainda que apresente os mesmos valores de *cut-in* e *cut-out* da Vestas V112. O diâmetro da área circular varrida pelas pás supera 126 metros, com área equivalente a 12668 m². O rotor pode ser disposto em uma torre também mais alta, de 135 metros de altura (WIND TURBINE MODELS, 2015b). Os detalhes das curvas de potência desta turbina são dispostos na Figura 87 e Tabela 26.



Figura 87 – Curva de potência Enercon E126 7.5 MW

Fonte: Wind Turbine Models (2015b)

Tabela 26 - Potência e coeficiente de potência em função da velocidade do vento para as turbinas eólicas

Velocidade do	Vestas V112 3.0	MW	Enercon E126 7.5	Enercon E126 7.5 MW	
vento [m/s]	Potência [kW]	C_p	Potência [kW]	C_p	
3,0	23	0,16	55	0,19	
4,0	130	0,34	175	0,35	
5,0	301	0,40	410	0,42	
6,0	557	0,43	760	0,45	
7,0	912	0,44	1250	0,47	
8,0	1377	0,45	1900	0,48	
9,0	1954	0,45	2700	0,48	
10,0	2572	0,43	3750	0,48	
11,0	2988	0,37	4850	0,47	
12,0	3065	0,29	5750	0,43	
13,0	3075	0,23	6500	0,38	
14,0	3075	0,19	7000	0,33	
15,0	3075	0,15	7350	0,28	
16,0	3075	0,12	7500	0,24	
17,0	3075	0,10	7500	0,20	
18,0	3075	0,09	7500	0,17	
19,0	3075	0,07	7500	0,14	
20,0	3075	0,06	7500	0,12	
21,0	3075	0,06	7500	0,11	
22,0	3075	0,05	7500	0,09	
23,0	3075	0,04	7500	0,08	
24,0	3075	0,04	7500	0,07	
25,0	3075	0,03	7500	0,06	

Fonte: adaptado de Wind Turbine Models (2015a) e Wind Turbine Models (2015b)

A conversão para unidades energéticas requer que se admita o vento à altura aproximada em que fica posicionado o rotor de um aerogerador. Assim, faz-se necessário promover a extrapolação vertical dos dados, corrigindo os valores de velocidade do vento pela transformação da altura representativa de medição para a altura do rotor em metros em relação ao solo. Razão para este procedimento perpassa pela compreensão do perfil vertical do vento que descreve o aumento da velocidade do vento com o aumento da altura – o perfil do vento no local depende de um fator de atrito que indica a resistência ao fluxo do vento que o entorno exerce (Figura 88).





Fonte: Brower et al. (2012)

Esta característica adimensional pode ser expressada pela razão entre a mudança da componente horizontal da velocidade do vento com a alteração da altura. Valores v_1 da velocidade do vento à altura $h_1 = 10$ m foram passados para velocidade v_2 , incidente na altura do rotor $h_2 = 100$ m. Curiosamente, esta equação não é embasada por teorias meteorológicas, porém provou-se muito eficaz para representar o perfil vertical em uma grande variedade de casos (BROWER et al., 2012).

$$\frac{\nu_2}{\nu_1} = \left(\frac{h_2}{h_1}\right)^{\alpha} \tag{80}$$

O fator de atrito α é o expoente de uma lei de potência que representa a rugosidade local, podendo ser definido conforme o tipo de terreno e vegetação existentes ao redor do aerogerador ou do local de medição. O conceito por trás desta quantidade está relacionado à resistência imposta pela superfície ao redor do local de aproveitamento energético (ou aferição do vento). Se a vegetação for relativamente alta ou se houver edificações, valores de α tendem a ser maiores. De mesma forma, se o terreno for dito complexo (repleto de sulcos, rochedos, ranhuras) há tendência de maiores valores de rugosidade α . Em contrapartida, superfícies aquáticas apresentam menor resistência ao fluxo do vento (Tabela 27) (BROWER et al., 2012).

Tipo de terreno	Cobertura terrestre e vegetação	Gama aproximada para a rugosidade
Plano	Baixa ou moderada vegetação	0,12 - 0,25
Plano	Floresta	0,25 - 0,40
Complexo (vales protegidos)	Variada	0,25 - 0,60
Complexo (vales falhos)	Variada	0,10 - 0,20
Complexo (cumes de montanha)	Baixa ou moderada vegetação	0,15 - 0,25
Complexo (cumes de montanha)	Floresta	0,20 - 0,35
Offshore em clima temperado	Água	0,10 - 0,15
Offshore em clima tropical	Água	0,07 - 0,10
	Γ_{1} (0.10)	

Tabela 27 – Valores típicos de rugosidade para diferentes condições

Fonte: Brower et al. (2012)

7.3.2 Energia proveniente de recursos hídricos

A potência entregue por uma UHE é derivada da energia da água em queda ao longo de um rio. Dados de descarga fluvial são convertidos em dados de energia aplicando uma função clássica de conversão. A potência entregue pelas unidades geradoras (turbinas hidráulicas) depende do fluxo de água que passa por elas (equivalente à vazão do rio $q_{nat}(t)$ a ser turbinada em um dado instante t), além de sua eficiência η_H , da aceleração gravitacional g, da densidade da água $\rho_{água}$ e da altura h da queda líquida nominal, que utiliza a diferença entre o nível de água médio e o nível normal à jusante, descontando perdas de carga que ocorrem nos circuitos hidráulicos. A energia produzida por uma UHE requer a multiplicação desta potência por um número T de horas de operação (FRANÇOIS et al., 2016; RAYNAUD et al., 2018).

$$E_{hydro} = T.P(q) \tag{81}$$

$$P(q) = \eta_H. g. h. \rho_{\acute{a}gua}. q_{nat}(t)$$
(82)

Tal qual na conversão do recurso eólico, o tempo *T* de operação é equivalente ao número de horas de cada mês do ano (744 horas para meses com 31 dias, 720 horas para meses com 30 dias e 672 para fevereiro). A aceleração gravitacional e a densidade da água $\rho_{água}$ são equivalentes a 9,81 m/s² e 1 kg/m³, respectivamente. Os demais parâmetros são específicos para cada uma das UHEs consideradas. A Tabela 28 elucida para as UHEs de Itaipu e Marimbondo.

UHE	Eficiência média [%]	Potência nominal [MW]	Queda líquida nominal [m]	Área do reservatório [km ²]
Itaipu	95,92	7000	117	823555
Marimbondo	93,28	1488	59,2	118515
	F	11 ~ (

Tabela 28 – Características das UHEs consideradas

O cálculo da ENA para um subsistema envolve a somatória da multiplicação das vazões naturais $q_{nat}(i, t)$ que ocorrem num dado instante t num dado local i de aproveitamento (dentre os n locais de aproveitamento existentes no subsistema considerado) pelas produtibilidades médias prod(i) dos conjuntos turbinas-geradores de aproveitamento hidrelétrico. A vazão natural afluente a um reservatório deve ser equivalente à vazão natural do reservatório localizado à montante, somada à vazão natural incremental consistida que ocorreu entre estes dois locais de aproveitamento. A produtibilidade, por sua vez, é referente à queda obtida pela diferença entre o nível de montante, correspondente à altura relativa a um armazenamento de 65% do volume útil do reservatório, e o nível médio do canal de fuga, ponto à jusante da barragem onde o rio segue o seu curso (ONS, 2020b):

$$ENA_{subsistema}(t) = \sum_{i=1}^{n} q_{nat}(i, t). prod(i)$$
(83)

É imprescindível reforçar que valores de ENA não representam o que de fato é gerado nos locais de aproveitamento. São apenas estimativas simplificadas do potencial energético contido nas vazões e que foram utilizadas como maneira atenuada de indicar, minimamente, as ordens de grandeza envolvidas na redução de geração frente a períodos de escassez conjunta. Equitativamente, a conversão energética dos dados de vazões naturais de cada UHE e suas respectivas produtibilidades instantâneas, refletiria apenas uma aproximação da geração. Razões para esta discrepância têm suas raízes no SIN, que tem seus reservatórios gerenciados com base em vários critérios de operação que decidem se a água será armazenada, turbinada ou até mesmo vertida. Isto faz sentido tendo em vista o ciclo de chuvas no SE/CO, com precipitação intensificada principalmente no verão austral, em detrimento do inverno. Diferentemente de Itaipu, a UHE de Marimbondo, não é operada a fio d'água, o que faz com que as decisões de geração sejam diferentes a fim de balancear ao longo do ano.

Para fins ilustrativos, a Figura 89 retrata o quão distantes ficam os valores de ENA SE/CO em relação ao que fora gerado pela fonte hídrica neste mesmo subsistema nos anos de 2015 a 2019. As curvas sólidas representam a geração e as pontilhadas a ENA, sendo que anos

Fonte: elaboração própria do autor

comuns foram coloridos por cores iguais. É notório que os valores de ENA respeitam a disponibilidade hídrica maior (menor) que ocorre nos meses do verão (inverno) austral, num padrão sazonal bem delineado.



Figura 89 – ENA e geração pela fonte hídrica no subsistema SE/CO de 2015 a 2019

Evidentemente que o uso dos dados brutos de ENA como matéria prima para análises de geração de energia abriria margem para uma noção distorcida do desempenho da fonte hídrica. Desta forma, foi necessário "corrigir" os dados mensais de ENA para médias anuais multiplicadas por um fator de 0,8. Julgou-se necessário realizar esta operação, já que os valores de ENA destes anos superestimavam a geração em cerca de 20%. Resultados na Figura 90.



Figura 90 - Correção dos valores de ENA

Fonte: ONS (2020a)

Fonte: elaboração própria do autor

É visível que, após esta correção, os valores de ENA (na Figura 90 identificados como ENAm) ficaram mais próximos das gerações verificadas pelo ONS nos anos de 2015 a 2019, já que as médias buscam representar o balanceamento realizado na operação do SIN de modo elucidativo, ainda que simplificado. Por conseguinte, optou-se por utilizar os valores médios de ENA de 1964 e 1971 corrigidos por este fator de 0,8 para alimentar os cálculos dos cenários de geração da próxima seção.

7.4 SIMULAÇÕES DA GERAÇÃO DE ENERGIA ELÉTRICA

Nesta última etapa, os dados coletados e analisados são finalmente utilizados para conversão energética, simulando cenários de geração em condições específicas conforme descrito anteriormente. Para os dados eólicos, ao invés de utilizar todas as séries temporais representativas (definidas no Capítulo 4), foi sensato utilizar apenas aquelas que representam os *clusters* localizados sobre as porções *onshore* e *offshore* do NE. A porção *onshore* condiz com o local onde hoje se concentram os parques eólicos, ao passo que a porção *offshore* foi definida como uma área que pode vir a ser explorada futuramente – sub-região onde a intensidade da velocidade do vento é superior em relação às demais.

Da Figura 78, verifica-se que os *clusters* melhor localizados são: Q_{N2}, Q_{N3}, Q_{E1}, Q_{M1} e Q_{M5}. O *cluster* Q_{E1} é relativamente vasto, abrangendo tanto a porção *onshore* quanto *offshore*, sendo que sua série temporal representativa (E33) está localizada em alto mar. Somado ao fato dos anos de 1964 e 1971 não figurarem entre os seus piores registros (Tabela 15), optou-se por não prosseguir com simulações de geração com estes dados, já que não seria possível simular geração *onshore* de maneira isolada. O mesmo ocorre com o *cluster* Q_{M5}. Contudo, a sua série M622 está localizada em latitudes mais a sul, o que faz reduzir a intensidade da velocidade do vento e, por conseguinte, faz com que seus dados sejam propícios para as simulações desejadas.

Diferentemente das séries temporais ERA e MERRA, os dados eólicos das séries NOAA representam a velocidade do vento para uma altura de 10 metros em relação ao solo. Por conseguinte, as duas séries temporais originais da base NOAA foram extrapoladas verticalmente: N38 para a altura de 100 metros (em função da turbina Vestas) e N15 para a altura de 135 metros (em função da turbina eólica Enercon). N38 é pertencente ao *cluster* Q_{N2} e foi utilizada para representar a sub-região do NE onde se concentra a maioria dos parques eólicos em operação no Brasil. Em se tratando de área relativamente vasta, fica um pouco

incerto reduzi-la a características básicas de terreno e vegetação. Alguns parques eólicos ficam localizados na costa, outros no interior do continente, situados próximos a serras de terreno complexo. Portanto, o valor de rugosidade α foi fixado em 0,25 como uma média abrangente.

A série N15 (pertencente do *cluster* Q_{N3}) representa uma área *offshore* onde, embora não exista ainda aproveitamento de recursos eólicos para geração de energia elétrica, é uma sub-região candidata a receber empreendimentos do tipo tendo em vista a alta velocidade do vento ali verificada. A área é marítima, rodeada por água, simplifica a decisão acerca do valor de α : foi definido em 0,1 para esta série temporal. O mesmo procedimento de extrapolação foi necessário para a série temporal M139 (*cluster* Q_{M1}) por representar a área *offshore* da base de dados MERRA. Todavia, neste caso a transformação da altura foi de 100 para 135 metros em relação ao solo, utilizando o mesmo valor 0,1 de rugosidade.

Ponto importante é que as séries do MERRA possuem menor cobertura temporal (de 1980 em diante), portanto não foram consideradas nos cenários II e III. Os dados MERRA foram eleitos para os cenários IV, V e VI. Os dados desta base são entendidos como mais recentes e contemplam apenas resultados de modelos que utilizam registros satelitais mais bem distribuídos globalmente, com melhores resolução temporal e espacial. Assim, para representar cenário de escassez de recursos, dados do ano de 1984 foram separados para prosseguimento. Este foi o pior ano de disponibilidade de recursos eólicos da série M622 e segundo pior ano de M139. Para contraste, o ciclo anual (CA – médias de todos os janeiros, fevereiros etc.) de ambas estas séries foram extraídos para representar a geração em condições normais.

	N3	8 *a	N1:	5 * ^b	Me	522	M13	9 *°
Mês	onsl	hore	offsl	hore	onsl	hore	offsl	hore
	1964	1971	1964	1971	CA	1984	CA	1984
1	5,985	7,497	7,051	9,364	6,833	6,080	7,640	6,840
2	5,701	6,026	7,401	7,771	6,612	6,126	7,202	6,515
3	4,775	4,832	6,381	6,901	6,268	5,065	6,477	4,762
4	4,480	4,038	5,714	6,446	6,292	6,439	6,344	4,869
5	4,680	5,031	6,662	7,175	6,832	7,224	7,439	5,340
6	6,121	5,451	7,884	8,750	7,744	6,989	8,937	8,519
7	6,310	6,703	8,401	9,325	8,133	8,034	10,057	9,151
8	7,076	7,810	9,833	10,100	7,759	7,546	11,110	10,173
9	7,397	8,001	10,308	10,419	7,354	7,212	11,203	9,936
10	7,537	8,079	9,882	10,237	7,257	6,923	10,605	9,879
11	7,513	7,949	9,572	9,601	7,333	7,040	9,925	9,630
12	7,553	8,042	9,637	9,720	7,043	6,533	8,931	8,350

Tabela 29 – Velocidade mensal média do vento das séries representativas [m/s]

Fonte: elaboração própria do autor

Para dados hídricos, optou-se por não utilizar os dados de vazões naturais de Itaipu e Marimbondo, já que modelar a produção de apenas duas UHEs para representar toda a região possivelmente acarretaria num incremento considerável no grau de incerteza das análises. Assim, apenas os valores de 1964 e 1971 de ENA para o subsistema SE/CO foram elencados (e corrigidos, como previamente explicado) para ilustrar a geração em condições de escassez conjunta de maneira bastante simplificada, e os valores médios de cada mês (ciclo anual) foram adotados para fins comparativos. A compilação destes dados pode ser conferida na Tabela 30.

Mâs	l	Dados brutos				Médias anuais corrigidas			
Mes	CA	1964	1971		CA	1964	1971		
1	48861,6	34161,7	28250,8	-	24314,4	19667,0	16424,6		
2	47371,7	47175,9	21520,5		24314,4	19667,0	16424,6		
3	51282,4	36319,6	27811,4		24314,4	19667,0	16424,6		
4	39403,2	28260,0	22305,9		24314,4	19667,0	16424,6		
5	29729,3	23243,3	18285,3		24314,4	19667,0	16424,6		
6	23484,2	16936,6	16740,8		24314,4	19667,0	16424,6		
7	19115,8	14791,8	15296,3		24314,4	19667,0	16424,6		
8	15311,0	11815,9	11157,4		24314,4	19667,0	16424,6		
9	14203,2	9259,3	10285,5		24314,4	19667,0	16424,6		
10	17599,4	15856,9	15796,1		24314,4	19667,0	16424,6		
11	22636,6	22440,4	20861,0		24314,4	19667,0	16424,6		
12	35718,2	34744,2	38058,1		24314,4	19667,0	16424,6		

Tabela 30 - ENA para o subsistema SE/CO [GWh]

Fonte: elaboração própria do autor

7.4.1 Resultados dos cenários II e III

Dos 15 GW de capacidade instalada de geração eólica no Brasil em 2019, aproximadamente 12,6 GW estavam localizados no NE, compondo um conjunto de inúmeras turbinas eólicas *onshore* com especificações técnicas distintas entre si. Entretanto, partiu-se da premissa de que esses 12,6 GW seriam constituídos de um total de 4100 aerogeradores *onshore* Vestas V112 3.0 MW. É uma aproximação simplificada, porém minimamente capaz de indicar a ordem de grandeza que a geração eólica pode alcançar dadas as condições de escassez. Apenas os valores da série N38 para os anos de 1964 e 1971 foram convertidos para energia gerada utilizando das equações dispostas anteriormente e das especificações técnicas de potência e coeficiente de potência da turbina supracitada.

Já com relação aos 63,2 GW de capacidade instalada de geração hidrelétrica no subsistema SE/CO, a geração total desta região será representada pelos dados de ENA já convertidos para unidades de energia, considerando os valores dos anos de 1964 e 1971. A geração combinada por estes dois recursos foi simulada para então conjecturar como seria o atendimento à carga de 2019 em condições de baixa disponibilidade. Os resultados dos cenários II e III estão dispostos na Tabela 31 e ilustrados nas Figuras 91 e 92.

	Geraç	ão eólica subsiste	ema NE	Geração hidrelétrica subsistema SE/CO			
Mês	Ocorrido 2019	Simulado 1964	Simulado 1971	Ocorrido 2019	Simulado 1964	Simulado 1971	
1	3759,4	1692,8 _{155,0%}	3412,1 _{19,2%}	26512,4	19667,0 _{↓25,8%}	16424,6 _{↓38,0%}	
2	1954,1	1294,0 _{↓33,8%}	1564,2 _{↓20,0%}	22590,2	19667,0 _{↓12,9%}	16424,6 _{127,3%}	
3	2142,0	781,6 _{163,5%}	810,0 _{162,2%}	24206,0	19667,0 _{↓18,8%}	16424,6 _{↓32,1%}	
4	2254,6	600,7 _{↓73,4%}	398,6 _{182,3%}	23387,5	19667,0 _{↓15,9%}	16424,6 _{↓29,8%}	
5	3445,1	732,1 _{178,7%}	937,4 _{172,8%}	21403,0	19667,0 _{18,10%}	16424,6 _{↓23,3%}	
6	4718,1	1757,0 _{162,8%}	1182,6 _{↓74,9%}	18599,3	19667,0 ^{†5,70%}	16424,611,7%	
7	4596,0	1988,5 _{156,7%}	2411,0 _{↓47,5%}	18184,7	19667,0 _{18,20%}	16424,6 _{↓9,70%}	
8	5840,5	2869,3 _{150,9%}	3901,6 _{133,2%}	18243,1	19667,0 _{17,80%}	16424,6 _{↓10,0%}	
9	5425,4	3172,1 _{141,5%}	4106,1 _{124,3%}	20400,8	19667,0 _{↓3,60%}	16424,6 _{↓19,5%}	
10	5411,5	3467,2 _{135,9%}	4367,2 _{↓19,3%}	22086,7	19667,0 _{↓11,0%}	16424,6 _{↓25,6%}	
11	4536,1	3323,9 _{126,7%}	4007,9 _{↓11,6%}	19960,8	19667,0 _{↓1,50%}	16424,6117,7%	
12	4500,4	3489,8 _{↓22,5%}	4308,8 _{↓4,3%}	24731,1	19667,0 _{↓20,5%}	16424,6 _{↓33,6%}	

Tabela 31 – Resultados das simulações de geração nos cenários II e III

Fonte: elaboração própria do autor

Estes resultados escancaram que mesmo que a variância interanual dos recursos hídricos seja superior à dos eólicos, em se tratando de geração energética, uma pequena variação na velocidade do vento pode provocar profundas alterações na geração por esta fonte. Em alguns meses a geração simulada ficou 70% abaixo da verificada em 2019. Isto relaciona-se com a conversão energética numa turbina eólica, que leva em consideração o cubo da velocidade do vento.

O exame dos resultados do cenário II aponta para uma geração combinada pelas fontes eólica (subsistema NE) e hídrica (subsistema SE/CO) de aproximadamente 40% da carga de 2019 durante os meses de janeiro até maio e de cerca de 45% da carga de junho até dezembro. Caso um desempenho desta magnitude venha a ocorrer em função da escassez conjunta dos recursos como indicam os dados analisados de 1964, a carga anual remanescente (~56%) deveria ser atendida pelas demais fontes e subsistemas. Se no ano base de 2019 as fontes e outros subsistemas foram responsáveis por ~ 48%, revela-se um decréscimo de geração pela geração eólica e hídrica em conjunto da ordem de 47,5 TWh.



Figura 91 - Resultados do cenário II: escassez conjunta de 1964

Fonte: elaboração própria do autor

No que se refere ao cenário III, as condições mostraram-se ainda pior que o cenário II. A geração combinada dos objetos de pesquisa ficou na faixa de 35 a 37% da carga nos meses de janeiro a maio, após o qual sobe para a faixa de 40% a 43%. A comparação com o ano do cenário base traz à tona que outras fontes e subsistemas deveriam suprir 61,5% da carga nas condições severas da escassez de 1971 identificada pelos dados de recursos hídricos e eólicos. Isto representa um acréscimo de aproximadamente 80 TWh, quantia nada negligenciável em se tratando de energia elétrica.



Figura 92 - Resultados do cenário III: escassez conjunta de 1971

Fonte: elaboração própria do autor

7.4.2 Resultados dos cenários IV, V e VI

A carga anual prevista para 2050 pela EPE (2015) pode alcançar 1340 TWh em condições tidas como favoráveis – e desafiadoras – para a expansão do setor elétrico brasileiro. Este valor, contudo, não fora especificado para cada mês do ano; assumir uma média equitativamente distribuída por entre os meses desmancharia o típico perfil anual apresentado pela carga (como apresentado na Figura 83). Foi pertinente distribuir esta carga total seguindo o padrão aproximado verificado nos anos de 2014 a 2019, em que os meses são naturalmente ponderados, muito em função da quantidade total de dias, dias úteis, finais de semana e feriados. Portanto a carga de 2050 seria aproximadamente distribuída como mostra a Tabela 32 a partir da multiplicação por um fator de ponderação mensal média.

Mâa	_		Ponderação	Carga [GWh]				
Mes	2014	2015	2016	2017	2018	2019	mensal [%]	2050
1	48674,1	51647,2	47755,4	50685,5	51008,3	54417,8	8,9	119009,3
2	45215,5	45634,0	47499,0	46971,9	46012,6	48342,1	8,2	109418,9
3	47115,0	49870,2	50028,6	51353,5	52572,3	51467,4	8,8	118312,4
4	43845,1	46461,7	49118,7	46500,0	48582,6	49480,2	8,3	111106,3
5	44627,3	46165,4	46680,2	47548,5	47499,4	49830,6	8,2	110465,9
6	41993,8	44138,1	44685,5	45361,5	45775,3	45897,0	7,8	104792,9
7	43507,0	45440,8	45838,6	45526,2	47299,9	46938,3	8,0	107414,1
8	43772,0	46270,0	47015,7	47064,7	47951,2	47622,7	8,2	109427,2
9	43902,3	46145,3	45829,1	46992,7	46973,2	48090,4	8,1	108737,3
10	46406,8	49021,9	47922,7	49432,0	50179,8	51707,4	8,6	115285,7
11	44647,2	47046,1	46739,7	47521,8	48737,0	49783,6	8,3	111296,9
12	45846,5	48275,8	48444,5	49566,4	50465,3	50659,7	8,6	114733,1
Total	539552,5	566116,5	567557,6	574524,7	583056,8	594237,1	100	1340000,0

Tabela 32 – Distribuição da carga anual por entre os 12 meses

Fonte: elaboração própria do autor

O acompanhamento de uma carga desta ordem de grandeza é sem precedentes. Para os cenários IV, V e VI, apenas as séries MERRA foram utilizadas: M622 para representar geração *onshore* e M139 para a geração *offshore*, extraindo o ano de 1984 de ambas as séries temporais para encenar condições de escassez eólica, porque este ano foi recorrente na classificação de piores médias anuais das séries MERRA. Simular a geração pela fonte eólica requer parâmetros comparativos e os seus ciclos anuais foram utilizados para converter as informações de velocidade do vento das médias mensais em energia, equacionando os aspectos técnicos das duas turbinas elegidas para conversão. A Tabela 33 retrata o quanto o potencial de geração *onshore* e *offshore* poderia ser reduzido se as condições de 1984 se repetissem, relativizando a

partir da geração que seria obtida se o ciclo anual médio histórico de cada uma das séries prevalecesse. Como visto previamente, a série M139 possui variabilidade maior que M622, tanto que, caso o desempenho de 1984 seja verificado, as possíveis reduções de geração *offshore* superam 60% para vários meses, ao passo que para *onshore* são, em média, menos expressivas.

	Vestas V112 3.0 MW			Enercon E126 7.5 MW		
Mês	M622 onshore	M622 onshore		M139 offshore	M139 offshore	
	CA	1984		CA	1984	
1	627,5	434,0 _{↓30,8%}		1226,4	781,2 _{↓36,3%}	
2	506,4	400,8 _{↓20,9%}		919,2	601,5 _{↓34,6%}	
3	475,4	234,5 _{150,7%}		720,7	224,1 _{168,9%}	
4	465,4	498,8 ^{†7,20%}		651,7	236,2 _{463,7%}	
5	627,2	744,8 ^{18,8} %		1127,2	341,8 _{169,7%}	
6	897,7	651,0 _{↓27,5%}		1914,0	$1515,1_{\downarrow 20,8\%}$	
7	1086,8	1047,7 _{↓3,60%}		2815,2	1940,7 _{↓31,1%}	
8	933,1	850,8 _{↓8,80%}		3685,6	2665,7 _{127,7%}	
9	760,4	717,1 _{15,70%}		3628,2	2404,1 _{133,7%}	
10	754,9	652,5 _{↓13,6%}		3263,3	2441,5 _{125,2%}	
11	753,7	$667, 1_{\downarrow 11,5\%}$		2621,7	2188,8 _{↓16,5%}	
12	690,2	538,3 _{↓22,0%}		1973,8	1474,5 _{↓25,3%}	
	-	~				

Tabela 33 – Geração eólica por uma única unidade geradora

Fonte: elaboração própria do autor

O atendimento à carga de 2050 pela fonte eólica no subsistema NE requereu uma multiplicação destes valores da Tabela 33 por um número equivalente de turbinas (*onshore* e *offshore*) que se aproxima da quantia de capacidade instalada atribuída a cada um dos cenários IV, V e VI. Para o cenário IV, 100 GW de potência instalada *onshore* equivale a 32520 turbinas Vestas V112 3.0 MW, enquanto 1333 turbinas Enercon E126 7.5 MW são consideradas para representar os 10 GW de geração *offshore*. O cenário V que assume capacidade instalada intermediária de energia eólica e é simulado por meio de 39024 turbinas Vestas e 4000 turbinas Enercon. Por fim, o cenário VI que favorece o incremento de potência instalada *onshore* e *offshore*.

Por um lado, é válido reiterar que, independentemente do número de turbinas, a redução do potencial de geração relativa à média histórica do ciclo anual indicada na Tabela 33 seria verificada se as condições de 1984 fossem reproduzidas futuramente. Por outro lado, a quantidade de energia entregue ao SIN dependeria da proporção *onshore-offshore*. Os valores sugerem que uma participação maior de parques *offshore* acarretaria flutuações de maior impacto, tendo em vista a maior variabilidade apresentada pelos dados de velocidade do vento da série M139.

Partindo da pressuposição de que a potência instalada total de geração pela fonte hídrica não será modificada ao longo dos próximos anos no subsistema SE/CO até 2050, os valores de ENA de 1971 foram novamente adotados para representar condições de escassez hídrica. Mesmo que não sejam correspondentes temporais dos dados MERRA de 1984, o objetivo abordará a mínima disponibilidade possível destes recursos, como se num mesmo ano ocorresse escassez verificada em 1971 para recursos hídricos e a escassez verificada em 1981 para recursos eólicos.

Ponto de relevância é que é presumível que as regras e restrições de operação do SIN sejam alteradas com o tempo e com a mudança da composição da matriz elétrica. O aumento da participação da fonte eólica forneceria uma quantidade de energia razoavelmente superior, principalmente no segundo semestre, sendo interessante que o potencial hídrico seja reequilibrado para suprir mais energia no primeiro semestre, ao invés de buscar prover uma geração constante como considerado anteriormente. Tal presunção vai de acordo com o aproveitamento da complementaridade intraanual natural existente entre recursos eólicos e hídricos. Logo, valores de ENA de 1971 não tiveram sua média calculada para preservar a sazonalidade existente nos dados brutos. Os dados brutos foram, entretanto, corrigidos pelo mesmo fator de 0,8 para não superestimar a geração pela fonte hídrica. Os resultados das simulações dos cenários IV, V e VI podem ser visualizados nas Figuras 93, 94 e 95.



Figura 93 - Resultados do cenário IV

A inspeção do gráfico referente ao cenário IV evidencia que o substancial aumento da carga passível de ocorrência nas próximas três décadas poderia fazer com que a participação

Fonte: elaboração própria do autor

conjunta das fontes hídrica no subsistema SE/CO e eólica no NE deixe de ter a capacidade de atendimento de cerca de 40% da carga em anos de escassez como verificado nos cenários II e III. Pode-se afirmar que, embora significativa, a multiplicação da capacidade instalada de geração eólica de 2019 por um fator de quase 9 – passando de 12,6 GW para 110 GW em 2050 – deveria ser complementada por uma capacidade de geração bastante alta por qualquer outra fonte em qualquer outro subsistema para suprir 65% da carga em anos de escassez conjunta, ou uma quantia energética da ordem de 870 TWh; valor superior à própria carga de 2019.



Figura 94 - Resultados do cenário V

Fonte: elaboração própria do autor



Figura 95 - Resultados do cenário VI

Fonte: elaboração própria do autor

O exame dos gráficos relativos aos outros dois cenários V e VI desvela que, à medida em que a capacidade instalada de geração eólica no NE ganha expressividade na matriz elétrica, o atendimento à futura carga de 2050 terá participação ativa essencial pela fonte eólica, com ganhos substanciais pela capacidade de geração *offshore*. Entretanto, parece acentuar-se um comportamento sazonal, muito em função da maior disponibilidade de recursos eólicos no segundo semestre. Os valores dos primeiros semestres sugerem que a geração *offshore* fica bem aquém da geração nos meses de julho, agosto e setembro. Em números absolutos, o cenário V aponta para a necessidade de suprimento de outras fontes e outros subsistemas de aproximadamente 440 TWh no primeiro semestre e 350 TWh no segundo. Tais valores análogos e correspondentes para o cenário VI são de 400 TWh e 290 TWh.

Estas observações corroboram com as análises do Capítulo 5, em que os meses de fevereiro, março e abril foram os que apresentaram maior número de ocorrências de velocidade média do vento abaixo de limiares de 1,5 ou 2 ou 2,5 unidades do desvio padrão. Portanto, sendo as variabilidades individuais destes meses maiores do que os demais, em anos de escassez conjunta é possível esperar que a sazonalidade se acentue, já que a conversão energética pondera muito a favor do valor de velocidade do vento.

A fins comparativos, o mesmo cenário VI foi simulado, porém os dados de entrada foram aqueles dos ciclos médios anuais tanto dos recursos eólicos *onshore* e *offshore* no NE quanto de recursos hídricos de ENA no SE/CO, representando a geração em anos em que a disponibilidade equivale às médias históricas que indicam as séries temporais originais.



Figura 96 - Resultados do cenário VI: geração a partir dos ciclos médios anuais

Fonte: elaboração própria do autor

Comparado ao cenário VI de escassez conjunta, os resultados da Figura 96 apontam para uma redução média de 14,7% (equivalente a cerca de 200 TWh por ano) na necessidade de atendimento à carga por outras fontes e subsistemas caso as disponibilidades de recursos eólicos e hídricos apresentem-se equivalentes às suas médias históricas. À parte a obviedade de que a geração combinada aumenta consideravelmente a participação conjunta de atendimento à carga – com valores sempre acima de 50% mesmo nos piores meses e superando 70% em julho e agosto – enaltece-se também que a sazonalidade é suavizada ao longo do ano. Em anos em que a disponibilidade destes recursos se aproxima da média histórica, a participação da fonte hídrica é enaltecida no primeiro semestre com concomitante aumento da geração eólica *onshore* e *offshore* nestes mesmos meses; aumento este em maior proporção daquele para o segundo semestre.

Mesmo com a complementaridade hidroeólica intraanual que se verifica naturalmente no Brasil entre os subsistemas SE/CO e NE, haverá uma carência de aumento em participação de outras fontes e tecnologias que sejam capazes de preencher a lacuna que se aprofunda nos meses do primeiro semestre de anos de escassez conjunta de recursos eólicos e hídricos. Adicionalmente, acredita-se que a inserção de capacidade de geração pela fonte eólica pode requisitar e forçar uma mudança nas regras e restrições de operação de UHEs com reservatório, de maneira que todo o sistema elétrico seja gerido em função desta fonte não despachável. Em todo caso, entende-se que os benefícios da adoção em massa da geração eólica no Brasil superam as desvantagens, sendo imprescindível que efeitos colaterais como os identificados na presente pesquisa sejam melhor compreendidos. Na medida do possível, é factível que sejam até mesmo evitados, bastando a realização do dimensionamento e planejamento de longo prazo do parque gerador nacional sob à luz da possibilidade de ocorrência de escassez de recursos como os eólicos e hídricos.

8 CONCLUSÕES E CONSIDERAÇÕES FINAIS

Esta pesquisa tem a pretensão de contribuir com o planejamento da expansão de longo prazo do sistema elétrico brasileiro a partir do desenvolvimento de elementos informacionais que possam apoiar decisões envolvendo o dimensionamento e a composição do parque gerador a nível nacional. Haja vista que a demanda futura por eletricidade pode mais do que dobrar até o ano de 2050, existe, em última instância, uma preocupação relacionada ao patamar de segurança energética a ser estabelecido ou sustentado.

Acredita-se que, se por um lado tal patamar for elevado, há o risco de alocar recursos humanos, materiais e financeiros em demasia, acarretando o superdimensionamento de capacidade de geração – recursos estes que poderiam ser utilizados em outras áreas societárias. Por outro lado, se o patamar é diminuto, existe o risco de enfrentar crises no abastecimento de energia elétrica pela baixa disponibilidade de recursos. A capacidade de geração e a definição das fontes energéticas que suprirão o atendimento à carga são, consequentemente, objetos relevantes e que requerem sopesar vantagens e desvantagens dos recursos disponíveis.

É plausível que decisões acerca das fontes e tecnologias de energia que futuramente serão adotadas e que terão estabelecimento de infraestrutura adequada aos seus aproveitamentos energéticos levarão em consideração critérios econômicos, buscando reduzir custos de geração e de operação. No entanto, a presente pesquisa buscou reforçar a importância da compreensão da disponibilidade de recursos estocásticos que, por sua vez, fazem parte de sistemas climáticos dotados de intrínseca variabilidade. Entende-se que o suprimento de energia elétrica deve ser garantido, mesmo em condições de escassez destes recursos.

Nesse sentido, um extenso ferramental matemático e estatístico foi empregado sobre conjuntos de informações meteorológicas e climáticas para estimular a compreensão da variabilidade interanual de recursos eólicos no NE e hídricos no SE/CO. Bases de dados de séries temporais contendo reconstruções históricas de velocidade do vento e vazões naturais de rios foram esmiuçadas para caracterizar a variabilidade interanual destas variáveis e analisar as ocorrências de eventos extremos que denotam mínima disponibilidade de geração.

Tamanha a quantidade de informação contida nas bases de dados eólicos, foi conveniente aplicar técnicas de clusterização para promover a divisão do NE do Brasil em cinco sub-regiões que contemplam séries temporais de recursos eólicos com alta similaridade. A análise destes agrupamentos ou *clusters* foi facilitada pela extração de séries temporais

representativas de cada um deles, para então prosseguir com a investigação da ocorrência prévia de eventos extremos que remontem escassez de recursos eólicos. E os resultados indicaram que, ao longo da cobertura temporal dos dados explorados, por vários anos o vento mostrou-se bem abaixo de sua média histórica. Na maioria dos piores anos das séries temporais estudadas, esta redução ficou entre 10 e 20% abaixo da média de longo prazo. São valores menores do que os verificados para as séries de recursos hídricos, que pela mesma análise verificou-se que as reduções de suas disponibilidades costumam ficar entre 30 e 45% abaixo da média histórica.

Em se tratando de variável atmosférica, a velocidade do vento, assim como outras variáveis representativas de recursos hídricos, está suscetível à influência de fenômenos climáticos de maior escala geográfica e de baixa frequência de ocorrência, com possibilidade de sofrer modulação de sua variabilidade interanual. O exame dos índices representativos dos alguns modos de variabilidade climática confirmou a suscetibilidade da intensificação de períodos de escassez eólica quando os fenômenos El Niño Oscilação Sul e o dipolo do Atlântico Tropical convergem para suas fases negativas, principalmente durante os meses do outono. Este resultado era esperado, corroborando com o conhecimento preexistente de que nestas condições específicas a ZCIT pode se deslocar para a porção mais a sul da linha do Equador, intensificando a precipitação e reduzindo a magnitude do vento no NE.

Seria imprudente concluir sobre o tempo de retorno de condições como estas e sobre a existência abstrusa de padrões periódicos diferentes do ciclo anual nas séries temporais de velocidade do vento. De qualquer maneira, como os fenômenos climáticos de larga escala explorados são bem conhecidos e têm um monitoramento ativo nas regiões em que se desenvolvem, é possível utilizar destas informações para realizar previsões de períodos de escassez eólica e/ou hídrica no NE do Brasil.

Hipótese trazida na introdução do trabalho e que versava sobre a eventualidade de ocorrência de períodos de escassez conjunta de recursos hídricos e eólicos foi confirmada pela simples constatação de que nos anos de 1964 e 1971 a disponibilidade da velocidade do vento no NE e as vazões naturais em rios da região SE/CO mostraram-se muito abaixo de suas médias históricas. Do ponto de vista energético, condições como estas são preocupantes para assegurar o atendimento à carga em escala de tempo anual. A partir destes indícios, cenários foram construídos para simular a geração de energia elétrica pelas fontes eólica no NE e hídrica no SE/CO com base em variadas composições da matriz elétrica e em condições de escassez destes recursos. Os produtos destas simulações constituíram o que se entende como principal contribuição desta tese.

Foi possível quantificar o quanto de energia elétrica deve ser gerado por outras fontes diferentes da eólica no subsistema NE e da hídrica no subsistema SE/CO caso as condições de escassez destes recursos verificadas em 1964 e 1971 venham a ocorrer novamente. Frente à geração verificada em 2019, os períodos de escassez conjunta poderiam fazer suprimir a geração anual em cerca de 45 a 80 TWh num único ano. Já em cenários futuros, onde a capacidade instalada de geração pela fonte eólica foi livremente alterada para simular o atendimento à carga de 2050, o aumento da participação de geração tanto *onshore* quanto *offshore* parece induzir o surgimento de forte padrão sazonal na geração, principalmente em anos de escassez eólica. Ademais, as conversões das séries temporais de velocidade do vento para quantidades energéticas fazem com que a variabilidade interanual dos recursos eólicos, ainda que menor do que as de recursos hídricos, impliquem em drásticas reduções de geração.

Estes resultados enfatizam a imprescindibilidade de que se dimensione o parque gerador para um mínimo de potência instalada de geração que sustente um patamar de segurança energética e que resista à variabilidade interanual inerente aos recursos como eólicos e hídricos. O parque gerador também deve ser adequado – mediante alterações nas regras e restrições de operações do SIN, sobretudo no que tange as UHEs com reservatório – ao padrão associado de atendimento à carga que pode ser modificado em função do aumento de capacidade instalada pela fonte eólica. Eventuais períodos de escassez devem servir como embasamento para a seleção de fontes e tecnologias de energia que se complementem e que mitiguem o risco de o sistema elétrico brasileiro não atender à carga em condições desfavoráveis.

Por esta ótica, não houve ambição de idealizar uma composição de diferentes fontes e tecnologias energéticas para o atendimento à carga futura. Reconhece-se que são questões que envolvem uma complexidade para além do escopo do trabalho realizado e que ainda há muito a ser desenvolvido. Não obstante, este trabalho advoga a favor da instalação massiva de potencial de geração pela fonte eólica, com a ressalva de que se alcance um entendimento das particularidades que reservam os recursos eólicos.

8.1 LIMITAÇÕES E SUGESTÕES PARA TRABALHOS FUTUROS

O exercício das simulações de geração com base nos cenários propostos foi baseado em uma série de premissas e escolhas arbitrárias que, ao reproduzir um modelo consideravelmente simplificado, possivelmente tenha potencializado o acúmulo de incertezas nos resultados obtidos. Embora não tenha feito parte do escopo da pesquisa a quantificação destas incertezas que permeiam as análises, considerou-se pertinente identificá-las para exposição de pontos a serem aprimorados caso este trabalho tenha continuidade.

Talvez a maior fonte de incerteza sejam os próprios dados de reanálise que foram utilizados para representar os recursos eólicos. Por serem resultados de modelos que buscam reconstruir o histórico de variáveis meteorológicas em uma grade global, a fidelidade das informações ali contidas pode ser questionável. No entanto, o Brasil carece de registros eólicos consistentes e abrangentes, tanto espacialmente quanto temporalmente, sendo impraticável mapear o comportamento do vento em toda a região nordestina por um período que supere um século de comprimento. Ainda assim, muitos dos resultados obtidos na seção de análise de séries temporais convergiram, em especial aqueles de ocorrência de períodos de escassez eólica.

Outra potencial deficiência dos dados eólicos envolve a subestimação dos valores de velocidade do vento, já que as reanálises possuem limitações na resolução espacial que acabam por suavizar potenciais pontos de aceleração eólica que dependem da orografia local. Para sanar esta questão, uma extensa análise de validação contra dados de medições sensoriais poderia ser conduzida a fim de verificar se há subestimação (ou até mesmo superestimação) para então definir um fator de correção, por exemplo. Em caso de subestimação da velocidade do vento, os cálculos de geração seriam também subestimados, podendo ser visto como um efeito colateral que se traduziria em medidas conservadoras para o setor elétrico.

Ressalta-se que as conversões de dados eólicos em dados de energia assumiram uma série de simplificações, desde a utilização de uma única série temporal para representar toda a região do NE onde se concentra a geração *onshore* ou *offshore* até a escolha das turbinas eólicas e alguns de seus parâmetros associados. Similarmente, a conversão de dados de vazões de rios em dados de ENA e a extrapolação destes para a geração total do subsistema SE/CO também partiu de simplificações e premissas que podem acumular incertezas. Adicionalmente, o enfoque sobre a variabilidade interanual negligenciou variabilidades em outras escalas temporais de curto ou médio prazo, como segundos e minutos até dias e semanas. Foi assumido que o SIN estivesse habilitado para lidar com quaisquer flutuações, independentemente dos desafios associados à intermitência dos recursos eólicos.

Ainda assim, o objetivo nunca fora modelar o sistema elétrico com fidelidade – apenas buscou-se analisar a sensibilidade da geração à disponibilidade limitada de recursos eólicos e hídricos com base apenas no potencial instalável, desconsiderando disponibilidade das máquinas de conversão, linhas de transmissão e qualquer outra infraestrutura.

Para fins de aprimoramento e como sugestão de trabalhos futuros, é altamente recomendável que pesquisas similares façam uso dos procedimentos metodológicos propostos e que eles sejam aplicados a diferentes populações de dados e a diferentes regiões do Brasil. Ademais, espera-se que o tema ganhe importância e que suscite o desenvolvimento de modelagem mais fidedigna e refinada do sistema elétrico, de maneira a aprimorar a simulação de cenários de geração com base em períodos de escassez de recursos energéticos em geral, contemplando também a penetração de geração distribuída (ex: unidades de geração solar fotovoltaica) e o uso de tecnologias de armazenamento (ex: bancos de baterias eletroquímicas).

É mister que se avance na modelagem para executar simulações que porventura contemplem todas as fontes energéticas desagregadas por regiões e subsistemas, tratando as unidades geradoras individualmente e respeitando restrições e critérios de operação do SIN. Pressupõe-se que, desta maneira, será factível atingir uma representação veraz, viabilizando resultados que guiem tomadas de decisão mais assertivas quanto à composição e dimensionamento da matriz elétrica do Brasil.

Por fim, reitera-se o desejo de que as contribuições alcançadas não se limitem à produção de conhecimento acadêmico. Apesar de todas as limitações, a intenção de colaborar com o planejamento de longo prazo do sistema elétrico brasileiro é genuína. Partiu-se de motivações e aspirações de promover uma melhor compreensão dos recursos naturais e suas variabilidades intrínsecas para que, em última instância, se aprimore o aproveitamento do potencial energético que eles têm a oferecer à sociedade.

REFERÊNCIAS BIBLIOGRÁFICAS

AGÊNCIA NACIONAL DE ÁGUAS. Conjuntura dos recursos hídricos no Brasil 2017: relatório pleno / Agência Nacional de Águas. -- Brasília: ANA, 2017. 169p.: iI.

AGÊNCIA NACIONAL DE ÁGUAS. Conjuntura dos recursos hídricos no Brasil: regiões hidrográficas brasileiras – Edição Especial. -- Brasília: ANA, 2015. 163 p.: il. ISBN: 978-85-8210-027-1

AGÊNCIA NACIONAL DE ÁGUAS. Quantidade de água: precipitação média anual e tempo de retorno, 2016. Sistema Nacional de Informações sobre Recursos Hídricos. Disponível em: <u>http://www.snirh.gov.br/snirh/snirh-1/acesso-tematico/quantidade-de-agua</u>. Acesso em 21/12/2018.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. Sistema de Informações de Geração da ANEEL – SIGA, 2020. Disponível em: <u>https://www.aneel.gov.br/siga</u>. Acesso em: 17/12/2020.

AGHABOZORGI, S.; SHIRKHORSHIDI, A. S.; WAH, T. Y. Time-series clustering – A decade review. **Information Systems**, v. 53, p.16–38, 2015.

ANDRADE, K. M. *Climatologia e comportamento dos sistemas frontais sobre a América do Sul.* 2005. 185 f. (INPE-14056-TDI/1067). Dissertação (Mestrado em Meteorologia) – INPE, São José dos Campos, 2005.

AUSTRALIAN GOVERNMENT - BUREAU OF METEOROLOGY. ENSO Wrap-Up: Current state of the Pacific and Indian Oceans. Disponível em: http://www.bom.gov.au/climate/enso/. Acesso em 26/12/2018.

ASSOCIAÇÃO BRASILEIRA DE ENERGIA EÓLICA. InfoVento 18 – Versão Português, 2020. Disponível em: <u>http://abeeolica.org.br/dados-abeeolica/</u>. Acesso em 17/12/2020.

BAKKER, A. M. R.; VAN DEN, B. J. J. M.; COELINGH, J. P. Decomposition of the windiness index in the Netherlands for the assessment of future long-term wind supply. **Wind Energy**, v. 16, p. 927–938, 2013.

BARANES, E.; JACQMIN, J.; POUDOU, J. C. Non-renewable and intermittent renewable energy sources: Friends and foes? **Energy Policy**, v. 111, n. April, p. 58–67, 2017.

BAYER, B. Experience with auctions for wind power in Brazil. **Renewable and Sustainable Energy Reviews**, v. 81, n. November 2016, p. 2644–2658, 2018.

BELLEMANS, A. et al. Feature extraction from principal component analysis based reducedorder models using orthogonal rotation. Disponível em: <u>https://www.researchgate.net/publication/324331758_Title_Feature_extraction_from_princip</u> <u>al_component_analysis_based_reduced-order_models_using_orthogonal_rotation</u>. Acesso em 09/12/2020.

BIANCHI, E.; SOLARTE, A.; GUOZDEN, T. M. Large scale climate drivers for wind resource in Southern South America. **Renewable Energy**, v. 114, p. 708–715, 2017.

BLÁZQUEZ, J.; SOLMAN, S. A. Interannual variability of the frontal activity in the Southern Hemisphere: relationship with atmospheric circulation and precipitation over southern South America. **Climate Dynamics**, v. 48, n. 7–8, p. 2569–2579, 2017.

BLOOMFIELD, P. Fourier Analysis of Time Series: An Introduction. 2 ed, John Wiley & Sons, Inc, New York, 2000.

BROWER, M. et al. Wind Resource Assessment: A Practical Guide to Developing a Wind Project. 1 ed. Hoboken, New Jersey: John Wiley & Sons, Inc., p. 290, 2012.

CANTÃO, M. P. et al. Evaluation of hydro-wind complementarity in the Brazilian territory by means of correlation maps. **Renewable Energy**, v. 101, p. 1215–1225, 2017.

CARVALHO, L. M. V.; JONES, C. Zona de Convergência do Atlântico Sul. In: CAVALCANTI, I. F. A. et al. **Tempo e Clima no Brasil**. 1 ed. São Paulo: Oficina de Textos, 2009. cap. 6, p. 95-109.

CARVALHO, L. M. V.; JONES, C., LIEBMANN, B. The South Atlantic Convergence Zone: Intensity, Form, Persistence, and Relationshipswith Intraseasonal to Interannual Activity and Extreme Rainfall. **Journal of Climate**. v. 17, issue 1, p. 88-108, 2004

CAVALCANTI, I. F. A.; KOUSKY, V. E. Frentes Frias Sobre o Brasil. In: CAVALCANTI, I. F. A. et al. **Tempo e Clima no Brasil**. 1 ed. São Paulo: Oficina de Textos, 2009. cap. 9, p. 135-147.

CHADE RICOSTI, J. F.; SAUER, I. L. An assessment of wind power prospects in the Brazilian hydrothermal system. **Renewable and Sustainable Energy Reviews**, v. 19, p. 742–753, 2013.

CHANG, P. et al. Climate fluctuations of tropical coupled systems - The role of ocean dynamics. **Journal of Climate**, v. 19, n. 20, p. 5122–5174, 2006.

CHATFIELD, C.; XING, H. **The Analysis of Time Series: an introduction with R**. 7. ed. [S.I.]: Chapman and Hall/CRC, 414 p., 2019.

CHISQUARETABLE. Statistical Tables for All. Chi Square Table, 2020. Disponível em: <u>http://chisquaretable.net</u>. Acesso em: 12/11/2020.

CHIANG, J. C. H.; VIMONT, D. J. Analogous Pacific and Atlantic meridional modes of tropical atmosphere-ocean variability. **Journal of Climate**, v. 17, n. 21, p. 4143–4158, 2004.

CHOUAKRIA, A. D.; NAGABHUSHAN, P. N. Adaptive dissimilarity index for measuring time series proximity. Advances in Data Analysis and Classification, v. 1, p. 1-5, 2007.

COELHO, C. A. S. et al. The 2014 southeast Brazil austral summer drought: regional scale mechanisms and teleconnections. **Climate Dynamics**, v. 46, n. 11–12, p. 3737–3752, 2016.

COLES, S. An Introduction to Statistical Modeling of Extreme Values. 1. ed. London: Springer, 2001, 221 p.

COMPO, G. P. et al. The Twentieth Century Reanalysis Project. **Quarterly Journal of the Royal Meteorological Society**. v.137, issue 654, p. 1-28, 2011.

COMPO, G. P. et al. Independent confirmation of global land warming without the use of station temperatures. **Geophysical Research Letters**, v. 40, n. 12, p. 3170–3174, 2013.

CUNHA, A. P. M. A. et al. Changes in the spatial-temporal patterns of droughts in the Brazilian Northeast. **Atmospheric Science Letters**, v. 19, n. 10, p. 1–8, 2018.

DANTAS, G. A. et al. Prospects for the Brazilian electricity sector in the 2030s: Scenarios and guidelines for its transformation. **Renewable and Sustainable Energy Reviews**. v. 68, p. 997–1007, 2017.

DEE et al. The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables, 2016. Disponível em: <u>https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables/</u>. Acesso em 07/01/2019.

DELWORTH, T. L.; MANN, M. E. Observed and simulated multidecadal variability in the Northern Hemisphere. **Climate Dynamics**, v. 16, n. 9, p. 661–676, 2000.

DESER, C. et al. Sea Surface Temperature Variability: Patterns and Mechanisms. Annual Review of Marine Science, v. 2, p. 115–143, 2010.

DIAGNE, M. et al. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. **Renewable and Sustainable Energy Reviews**, v. 27, p. 65–76, 2013.

DIAS, M. A. F. S.; SILVA, M. G. A. J. Para entender tempo e clima. In: CAVALCANTI, I. F. A. et al. **Tempo e Clima no Brasil**. 1 ed. São Paulo: Oficina de Textos, 2009. cap. 1, p. 15-21.

DRUMOND, A. R. M.; AMBRIZZI, T. Inter ENSO variability and its influence over the South American monsoon system. Advances in Geosciences, v. 6, n. 1988, p. 167–171, 2006.

EDWARDS, P. N. History of climate modeling. Wiley Interdisciplinary Reviews: Climate Change, v. 2, n. 1, p. 128–139, 2011.

EIFLER NETO, E.; RISSO, A.; BELUCO, A. Complementarity in Time between Wind and Water Resources in Northeastern Brazil. International Journal of Environmental Engineering and Natural resources, v. 1, n. 1, p. 13–18, 2014.

EMPRESA DE PESQUISA ENERGÉTICA. Plano decenal de expansão de energia 2029. Brasília: MME/EPE, 2020a. Disponível em: <u>https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/plano-decenal-de-expansao-de-energia-2029</u>. Acesso em: 09/01/20<u>21</u>.

EMPRESA DE PESQUISA ENERGÉTICA. Plano Nacional de Energia 2050. Brasília: MME/EPE, 2020b. Disponível em: <u>https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/Plano-Nacional-de-Energia-2050</u>. Acesso em: 09/01/20<u>21</u>.

EMPRESA DE PESQUISA ENERGÉTICA. Nota técnica DEA 13/15: Demanda de Energia 2050. Rio de Janeiro, 2015. Disponível em: <u>https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/Plano-Nacional-de-Energia-2050</u>. Acesso em: 21/01/2021.

EMPRESA DE PESQUISA ENERGÉTICA. Roadmap Eólica *Offshore* Brasil – Perspectivas e caminhos para a energia eólica marítima. NT-EPE-PR-001/2020-r2, Rio de Janeiro, 2020c.
EMPRESA DE PESQUISA ENERGÉTICA. Potencial dos Recursos Energéticos no Horizonte 2050. Série Recursos Energéticos. Nota Técnica PR 04/18. Rio de Janeiro, 2018.

ENFIELD, D. B. et al. How ubiquitous is the dipole relationship in tropical Atlantic sea surface temperatures? **Journal of Geophysical Research**, v. 104, n.C4, p. 7841–7848, 1999.

ENFIELD, D. B., MESTAS-NUNEZ, A. M., TRIMBLE, P. J. The Atlantic Multidecadal Oscillation and its relationship to rainfall and river flows in the continental U.S. **Geophysical Research Letters**, 28: p. 2077-2080, 2001.

ENGELAND, K. et al. Space-time variability of climate variables and intermittent renewable electricity production – A review. **Renewable and Sustainable Energy Reviews**, v. 79, n. February, p. 600–617, 2017.

FERREIRA, J. H. I. et al. Assessment of the potential of small hydropower development in Brazil. **Renewable and Sustainable Energy Reviews**, v. 56, p. 380–387, 2016.

FRANÇOIS et al. Increasing climate-related-energy penetration by integrating run-of-the river hydropower to wind/solar mix. **Renewable Energy**. v. 87, p. 686-696, 2016.

GELARO, R. et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). Journal of Climate. v. 30, issue 14, p. 5419–5454, 2017.

GIEBEL, G. Equalizing Effects of the Wind Energy Production in Northern Europe Determined from Reanalysis Data - Riso technical Report. v. 1182, n. May, p. 20, 2000.

GRAUL e POPPINGA. bReeze: Functions for Wind Resource Assessment. R package version 0.4-3, 2018. Disponível em: <u>https://CRAN.R-project.org/package=bReeze</u>. Acesso em 25/01/2021.

GRIMM, A. M. Interannual climate variability in South America: Impacts on seasonal precipitation, extreme events, and possible effects of climate change. **Stochastic Environmental Research and Risk Assessment**, v. 25, n. 4, p. 537–554, 2011.

GRIMM, A. M. Variabilidade Interanual do Clima no Brasil. In: CAVALCANTI, I. F. A. et al. **Tempo e Clima no Brasil**. 1 ed. São Paulo: Oficina de Textos, 2009. cap. 22, p. 353-374.

HARTIGAN, J. A.; WONG, M. A. A K-Means Clustering Algorithm. Applied Statistics, v. 28, issue 1, p. 100-108, 1979.

HALKIDI, M.; BATISTAKIS, Y.; VAZIRGIANNIS, M. Clustering Validity Checking Methods: Part II. **SIGMOD Rec.**, v. 31, n. 3, p. 19–27, 2002.

HUNT, J. D.; STILPEN, D.; FREITAS, M. A. V. A review of the causes, impacts and solutions for electricity supply crises in Brazil. **Renewable and Sustainable Energy Reviews**, v. 88, n. January 2017, p. 208–222, 2018.

JABIR, M. et al. Intermittent smoothing approaches for wind power output: A review. **Energies**, v. 10, n. 10, 2017.

KANE, R. P. Prediction of droughts in north-east Brazil: Role of ENSO and use of periodicities. **International Journal of Climatology**, v. 17, n. 6, p. 655–665, 1997.

KASSAMBARA, A. Practical Guide to Cluster Analysis in R. 1 ed, STHDA, 2017.

KASSAMBARA, A.; MUNDT, F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7, 2020. Disponível em: <u>https://CRAN.R-project.org/package=factoextra</u>. Acesso em 25/01/2021.

KAUFMAN, L. ROUSSEEUW, P. J. Finding Groups in Data: An Introduction to Cluster Analysis. 2 ed. John Wiley & Sons, Inc, New Jersey, 2005.

KAYANO, M. T.; ANDREOLI, R. V. Relationships between rainfall anomalies over northeastern Brazil and the El Niño-Southern Oscillation. Journal of Geophysical Research Atmospheres, v. 111, n. 13, p. 1–11, 2006.

KAYANO, M. T.; ANDREOLI, R. V. Relations of South American summer rainfall interannual variations with the Pacific Decadal Oscillation. **International Journal of Climatology**, v. 27, p. 431-540, 2007.

KAYANO, M. T.; ANDREOLI, R. V. Variabilidade Decenal a Multidecenal. In: CAVALCANTI, I. F. A. et al. **Tempo e Clima no Brasil**. 1 ed. São Paulo: Oficina de Textos, 2009a. cap. 23, p. 375-383.

KAYANO, M. T.; ANDREOLI, R. V. Clima da Região Nordeste do Brasil. In: CAVALCANTI, I. F. A. et al. **Tempo e Clima no Brasil**. 1 ed. São Paulo: Oficina de Textos, 2009b. cap. 14, p. 213-233.

KAYANO, M. T. et al. A further analysis of the tropical Atlantic SST modes and their relations to north-eastern Brazil rainfall during different phases of Atlantic Multidecadal Oscillation. **International Journal of Climatology**, v. 36, n. 12, p. 4006–4018, 2016.

KAYANO, M. T.; JONES, C.; DIAS, P. L. S. Variabilidade Intrassazonal. In: CAVALCANTI, I. F. A. et al. **Tempo e Clima no Brasil**. 1 ed. São Paulo: Oficina de Textos, 2009. cap. 21, p. 337-351.

KIRCHNER-BOSSI, N. et al. A Long-Term perspective of Wind Power Output Variability. **International Journal of Climatology**, v. 35, p. 2635–1646, 2015.

KOUTSOYIANNIS, D. Hurst-Kolmogorov Dynamics and Uncertainty. Journal of the American Water Resources Association, v. 47, n.3, p. 481–495, 2011.

KRAKAUER, N.; COHAN, D. Interannual Variability and Seasonal Predictability of Wind and Solar Resources. **Resources**, v. 6, n. 3, p. 29, 2017.

KUSHNIR, Y. et al. The Physical Basis for Predicting Atlantic Sector Seasonal-to-Interannual Climate Variability. **Journal of Climate**, v. 19, p. 5949-5970, 2006.

LAHOZ, W. A.; SCHNEIDER, P. Data assimilation: making sense of Earth Observation. **Frontiers in Environmental Science**, v. 2, n. May, p. 1–28, 2014.

LEAHY, P, G.; MCKEOGH, E. J. Persistence of low wind speed conditions and implications for wind power variability. **Wind Energy**. v. 16, p. 575–586, 2013.

LIAO, T. W. Clustering of time series data – a survey. **Pattern Recognition**, v. 38, p.1857–1874, 2005.

LLEDÓ, L. et al. Investigating the Effects of Pacific Sea Surface Temperatures on the Wind Drought of 2015 over the United States. **Journal of Geophysical Research: Atmospheres**, v. 123, n. 10, p. 4837–4849, 2018.

LOIKITH, P. C. et al. The Influence of Recurrent Modes of Climate Variability on the Occurrence of Monthly Temperature Extremes Over South America. Journal of Geophysical Research: Atmospheres, v. 122, n. 19, p. 10297–10311, 2017.

MADDEN, R. A.; JULIAN, P. R. Detection of a 40 – 50 Day Oscillation in the Zonal Pacific Wind in the Tropical Pacific. **Journal of the Atmospheric Sciences**. v. 28, p. 702-708, 1971.

MANTUA, N.J. et al. A Pacific interdecadal climate oscillation with impacts on salmon production. **Bulletin of the American Meteorological Society**, 78, p. 1069-1079, 1997.

MANTUA, N. J.; HARE, S. R. The Pacific Decadal Oscillation. Journal of Oceanography, v. 58, p. 35-44, 2002.

MARENGO, J. A. Mudanças Climáticas: detecção e cenários futuros para o Brasil até o final do Século XXI. In: CAVALCANTI, I. F. A. et al. **Tempo e Clima no Brasil**. 1 ed. São Paulo: Oficina de Textos, 2009. cap. 25, p. 407-424.

MARENGO, J. A. et al. Climatic characteristics of the 2010-2016 drought in the semiarid northeast Brazil region. Anais da Academia Brasileira de Ciencias, v. 90, n. 2, p. 1973–1985, 2018.

MARENGO, J. A.; TORRES, R. R.; ALVES, L. M. Drought in Northeast Brazil—past, present, and future. **Theoretical and Applied Climatology**, v. 129, n. 3–4, p. 1189–1200, 2017.

MARTINS, F. R.; GUARNIERI, R. A.; PEREIRA, E. B. O aproveitamento da energia eólica. **Revista Brasileira de Ensino de Física**, v. 30, n. 1, p. 1304.1-1304.13, 2008.

MELO, A. B. C.; CAVALCANTI, I. F. A.; SOUZA, P. P. Zona de Convergência Intertropical do Atlântico. In: CAVALCANTI, I. F. A. et al. **Tempo e Clima no Brasil**. 1 ed. São Paulo: Oficina de Textos, 2009. cap. 2, p. 25-41.

MELO, D. D. C. D. et al. Reservoir storage and hydrologic responses to droughts in the Paraná River basin, south-eastern Brazil. **Hydrology and Earth System Sciences**, v. 20, n. 11, p. 4673–4688, 2016.

MEZA, C. G. et al. **Estado da arte de Armazenamento de Energia e Inserção de Fontes Renováveis Intermitentes**. Relatório do Seminário Internacional de Armazenamento de Energia. Instituto de Energia e Ambiente, Universidade de São Paulo, Brasil, 2020. MIGUEL, P. A. C. Metodologia de pesquisa para engenharia de produção e gestão de operações [recurso eletrônico] / Paulo Augusto Cauchick Miguel (organizador). – Rio de Janeiro: Elsevier: ABEPRO, 2012.

MISHRA, A. K.; SINGH, V. P. A review of drought concepts. **Journal of Hydrology**, v. 391, n. 1–2, p. 202–216, 2010.

MOHAMMADI, K.; GOUDARZI, N. Study of inter-correlations of solar radiation, wind speed and precipitation under the influence of El Niño Southern Oscillation (ENSO) in California. **Renewable Energy**, v. 120, p. 190–200, 2018.

MOLION, L. C. B.; BERNARDO, S. O. Uma revisão da dinâmica das chuvas no Nordeste brasileiro. **Revista Brasileira de Meteorologia**, v. 17, n. 1, p. 1–10, 2002.

MONTERO, P; VILAR, J. A. TSclust: An R Package for Time Series Clustering. Journal of Statistical Software, 62(1), 1-43, 2014. Disponível em: <u>http://www.jstatsoft.org/v62/i01/</u>. Acesso em 25/01/2021.

MURTAGH, F.; LEGENDRE, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? **Journal of Classification**, v. 31, p.274-295, 2014.

MORETTIN, P. A.; TOLOI, C. M. C. Análise de Séries Temporais. 3 ed. São Paulo: Blucher, 2018, 474p.

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION - EARTH SYSTEM RESEARCH LABORATORY'S PHYSICAL SCIENCES DIVISION. Monthly Climate Timeseries: Atlantic Meridional Mode (AMM) SST Index, 2018. Disponível em: https://www.esrl.noaa.gov/psd/data/timeseries/monthly/AMM/. Acesso em 25/12/2018

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION - NATIONAL CLIMATIC DATA CENTER. The Walker Cell and ENSO, 2018. Disponível em: <u>https://www.ncdc.noaa.gov/teleconnections/enso/</u>. Acesso em: 25/12/2018

NAVARRA, A.; SIMONCINI, V. A Guide to Empirical Orthogonal Functions for Climate Data Analysis. 1 ed. Springer Science+Business Media B.V., 2010

NEWMAN, M. et al. The Pacific decadal oscillation, revisited. **Journal of Climate**, v. 29, n. 12, p. 4399–4427, 2016.

NEWMAN, M.; COMPO, G. P.; ALEXANDER, M. A. Letters of the Pacific Decadal Oscillation. Journal of Climate, v. 16, n. 23, p. 3853–3857, 2003.

NOTTON, G. et al. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. **Renewable and Sustainable Energy Reviews**, v. 87, n. December 2016, p. 96–105, 2018.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO. Procedimentos de rede - Submódulo 2.4: Critérios para estudos energéticos e hidrológicos. Revisão 2020.12, 2020b.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO. Plano da Operação Energética 2017/2021, 2017. Disponível em: <u>http://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/RE-3-0108-2017_PEN2017_SumarioExecutivo.pdf</u>. Acesso em 23/10/2020.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO. Sobre o SIN: o que é o SIN. Disponível em: <u>http://ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin</u>. Acesso em 13/12/2018.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO. Resultados da Operação: histórico da operação, 2020a. Disponível em: <u>http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/</u>. Acesso em 16/12/2020.

OWEN, S. Common Probability Distributions: The Data Scientist's Crib Sheet, 2018.Disponívelem:<u>https://medium.com/@srowen/common-probability-distributions-</u><u>347e6b945ce4</u>. Acesso em 26/12/2020.

PALFI, G. C.; ZAMBON, R. C. Hydro and Wind Power Complementarity and Scenarization in Brazil. **World Environmental and Water Resources Congress 2013**, p. 2414–2424, 2013.

PARKER, W. S. Reanalyses and Observations: What's the Difference? Bulletin of the American Meteorological Society, v. 97, n. 9, p. 1565–1572, 2016.

PATLAKAS et al. Low wind speed events: persistence and frequency. **Wind Energy**, v. 20, p. 1033–1047, 2017.

PEGORIM, J. Como a ZCAS se forma? 2017. Disponível em: <u>https://www.climatempo.com.br/noticia/2017/11/17/como-a-zcas-se-forma--2964</u>. Acesso em 20/12/2018.

PENG, R. D. A Very Short Course on Time Series Analysis. Disponível em: <u>https://bookdown.org/rdpeng/timeseriesbook/</u>. Acesso em: 16/12/2020.

PERCIVAL, D. B.; WALDEN, A. T. **Spectral Analysis for Univariate Time Series**. 1 ed. Cambridge University Press, New York, 2020.

PRADO JR., F. A. et al. How much is enough? An integrated examination of energy security, economic growth and climate change related to hydropower expansion in Brazil. **Renewable and Sustainable Energy Reviews**. v. 53, p. 1132–1136, 2016

PRYOR, S. C.; BARTHELMIE, R. J.; SCHOOF, J. T. Inter-annual variability of wind indices across Europe. **Wind Energy**, v. 9, n. 1–2, p. 27–38, 2006.

POLI, P. et al. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. Journal of Climate. v. 29, issue 11, p. 4083–4097, 2016.

PULWARTY, R. S.; SIVAKUMAR, M. V. K. Information systems in a changing climate: Early warnings and drought risk management. Weather and Climate Extremes, v. 3, p. 14–21, 2014.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2020. Disponível em: <u>https://www.R-project.org/</u>. Acesso em 25/01/2021.

RAYNAUD, D. et al. Energy droughts from variable renewable energy sources in European climates. **Renewable Energy**, v. 125, p. 578–589, 2018.

RAYNER N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. **J. Geophys. Res.**, 108 (D14), 4407, 2003.

REBOITA, M. S. et al. Precipitation regimes in South America: a bibliography review. **Revista Brasileira de Meteorologia**, v. 25, n. 2, p. 185–204, 2010.

ROBERTSON, A. W.; MECHOSO, C. R. Interannual and Interdecadal Variability of the South Atlantic Convergence Zone. **Monthly Weather Review**, v. 128, n. 8, p. 2947–2957, 2000.

ROSE, S.; APT, J. Quantifying sources of uncertainty in reanalysis derived wind speed. **Renewable Energy**, v. 94, p. 157–165, 2016.

ROSSO, F. et al. Influence of the Antarctic Oscillation on the South Atlantic Convergence Zone. **Atmosphere**, v. 9, n. 11, p. 431, 2018.

SANTOS, A. T. S.; E SILVA, C. M. S. Seasonality, Interannual Variability, and Linear Tendency of Wind Speeds in the Northeast Brazil from 1986 to 2011. **The Scientific World Journal**, v. 2013, p. 1–10, 2013.

SCHMIDT, J.; CANCELLA, R.; JUNIOR, A. O. P. The effect of windpower on long-term variability of combined hydro-wind resources: The case of Brazil. **Renewable and Sustainable Energy Reviews**, v. 55, p. 131–141, 2016a.

SCHMIDT, J.; CANCELLA, R.; PEREIRA JR., A. O. The role of wind power and solar PV in reducing risks in the Brazilian hydro-thermal power system. **Energy**, v. 115, p.1748-1757, 2016b.

SCHMIDT, J.; CANCELLA, R.; PEREIRA JR., A. O. An optimal mix of solar PV, wind and hydro power for a low-carbon electricity supply in Brazil. **Renewable Energy**, v. 85, p.137-147, 2016c.

SCHNEIDER, T.; BISCHOFF, T.; HAUG, G. H. Migrations and dynamics of the intertropical convergence zone. **Nature**, v. 513, n. 7516, p. 45–53, 2014.

SERVAIN, J. et al. Relationship between the equatorial and meridional modes of climatic variability in the tropical Atlantic. **Geophysical Research Letters**, v. 26, n. 4, p. 485–488, 1999.

SHUMWAY, R. H.; STOFFER, D. S. Time Series Analysis and Its Applications With R Examples. 4 ed. Springer, 2016.

SMIRNOV, D.; VIMONT, D. J. Variability of the Atlantic Meridional Mode during the Atlantic hurricane season. **Journal of Climate**, v. 24, n. 5, p. 1409–1424, 2011.

THOMPSON, D. W. J.; SOLOMON, S. Interpretation of recent Southern Hemisphere climate change. **Science**, v. 296, n. 5569, p. 895–899, 2002.

THOMPSON, D. W. J. et al. Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. **Nature Geoscience**, v. 4, n. 11, p. 741–749, 2011.

THOMPSON, D. W. J.; WALLACE, J. M. Annular modes in the extratropical circulation. Part II: Trends. **Journal of Climate**, v. 13, n. 5, p. 1018–1036, 2000.

THOMPSON, D. W. J.; WALLACE, J. M. The Arctic oscillation signature in the wintertime geopotential height and temperature fields. **Geophysical Research Letters**, v. 25, n. 9, p. 1297–1300, 1998.

TING, M. et al. Forced and internal twentieth-century SST trends in the North Atlantic. **Journal of Climate**, v. 22, n. 6, p. 1469–1481, 2009.

TREKHLEB, O. Playing with Discrete Fourier Transform Algorithm in JavaScript, 2018. Disponível em: <u>https://dev.to/trekhleb/playing-with-discrete-fourier-transform-algorithm-in-javascript-53n5</u>. Acesso em 26/11/2020.

TRENBERTH, K. E. Atmospheric reanalyses: A major resource for climate services. **GEO** News, 2010.

TRENBERTH, K. E. The definition of El Niño - ProQuest. Bulletin of the American Meteorological Society, v. 78, n. 12, p. 2771–2777, 1997.

TRENBERTH, K. E.; CARON, J. M. The southern oscillation revisited: Sea level pressures, surface temperatures, and precipitation. **Journal of Climate**, v. 13, n. 24, p. 4358–4365, 2000.

TRENBERTH, K. E.; HURRELL, J. W. Decadal atmosphere-ocean variations in the Pacific. **Climate Dynamics**, v. 9, n. 6, p. 303–319, 1994.

TTABLE. T Value Table. Disponível em: <u>http://www.ttable.org/</u>. Acesso em: 11/12/2020.

VIMONT, D. J.; KOSSIN, J. P. The Atlantic Meridional Mode and hurricane activity. **Geophysical Research Letters**, v. 34, n. 7, p. 1–5, 2007.

VIRON, O.; DICKEY, J. O.; GHIL, M. Global modes of climate variability. **Geophysical Research Letters**, v. 40, n. 9, p. 1832–1837, 2013.

WATSON, S. J.; KRITHARAS, P.; HODGSON, G. J. Wind speed variability across the UK between 1957 and 2011. **Wind Energy**, v. 18, p. 21–42, 2015.

WANG, X.; SMITH, K.; HYNDMAN, R. Characteristic-Based Clustering for Time Series Data. **Data Mining and Knowledge Discovery**, v. 13, p. 335–364, 2006.

WIKLE, C. K.; ZAMMIT-MANGION, A.; CRESSIE, N. **Spatio-Temporal Statistics with R**. 1 ed. Chapman & Hall/CRC, p. 396, 2019

WILKS, D. **Statistical Methods in the Atmospheric Sciences**. 4 ed. Amsterdam: Elsevier, p. 818, 2019.

WIND TURBINE MODELS. Vestas V112 Onshore, 2015a. Disponível em: <u>https://en.wind-turbine-models.com/turbines/7-vestas-v112-onshore</u>. Acesso em 23/01/2021.

WIND TURBINE MODELS. Enercon E-126 7.580, 2015b. Disponível em: <u>https://en.wind-turbine-models.com/turbines/14-enercon-e-126-7.580</u>. Acesso em: 23/01/2021.

WORLD METEOROLOGICAL ORGANIZATION; GLOBAL WATER PARTNERSHIP. Handbook of Drought Indicators and Indices (M. Svoboda and B.A. Fuchs). Integrated Drought Management Programme (IDMP), Integrated Drought Management Tools and Guidelines Series 2. Geneva, 2016.

XIE, S. P.; CARTON, J. A. Tropical atlantic variability: Patterns, mechanisms, and impacts. **Geophysical Monograph Series**, v. 147, p. 121–142, 2004.

ZHANG, Z. et al. MODIS NDVI time series clustering under dynamic time warping. **International Journal of Wavelets, Multiresolution and Information Processing**, v. 12, n. 5, 14 p., 2014.

ZHOU, J.; LAU, K. M. Principal modes of interannual and decadal variability of summer rainfall over South America. **International Journal of Climatology**, v. 21, n. 13, p. 1623–1644, 2001.

ZURN, H. H. et al. Electrical energy demand efficiency efforts in Brazil, past, lessons learned, present and future: A critical review. **Renewable and Sustainable Energy Reviews**, v. 67, p. 1081–1086, 2017.