



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE CIÊNCIAS EXATAS E DA TERRA  
DEPARTAMENTO DE CIÊNCIAS ATMOSFÉRICAS E CLIMÁTICAS  
GRADUAÇÃO EM METEOROLOGIA

**PAULA ANDRESSA ALVES DE ARAÚJO**

**Imputação de dados horários de velocidade do vento no território  
continental do Brasil**

NATAL/RN

MARÇO-2023

# **Imputação de dados horários de velocidade do vento no território continental do Brasil**

PAULA ANDRESSA ALVES DE ARAÚJO

Trabalho de Conclusão de Curso  
apresentado ao Curso de Bacharelado  
em Meteorologia da Universidade Federal  
do Rio Grande do Norte como requisito  
parcial para a obtenção do título de  
Bacharel em Meteorologia.

Orientador(a): Prof. Dr. Cláudio Moisés  
Santos e Silva

NATAL/RN

MARÇO-2023

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE CIÊNCIAS EXATAS E DA TERRA  
DEPARTAMENTO DE CIÊNCIAS ATMOSFÉRICAS E CLIMÁTICAS  
GRADUAÇÃO EM METEOROLOGIA

A Monografia “**Imputação de dados horários de velocidade do vento no território continental do Brasil**”

elaborado por Paula Andressa Alves de Araújo

e aprovada por todos os membros da Banca Examinadora foi aceita pelo Colegiado do Curso de Meteorologia e homologada pelos membros da banca, como requisito parcial à obtenção do título de **BACHAREL EM METEOROLOGIA**.

Natal, 13 de março de 2023

BANCA EXAMINADORA

---

Prof Dr.Cláudio Moisés Santos e Silva (Orientador - UFRN)

---

Profª Dra.Daniele Tôrres Rodrigues (UFPI)

---

Dra.Keila Rego Mendes (INCT-ONDACBC)

---

Dra.Samira de Azevedo Santos Emiliavaca (ISI-ER)

Universidade Federal do Rio Grande do Norte - UFRN  
Sistema de Bibliotecas - SISBI  
Catalogação de Publicação na Fonte. UFRN - Biblioteca Setorial Prof. Ronaldo Xavier de Arruda - CCET

Araújo, Paula Andressa Alves de.

Imputação de dados horários de velocidade do vento no território continental do Brasil / Paula Andressa Alves de Araújo. - 2023.

58f.: il.

Monografia (Graduação em Meteorologia) - Universidade Federal do Rio Grande do Norte, Centro de Ciências Exatas e da Terra, Departamento de Ciências Atmosféricas e Climáticas. Natal, RN, 2023.

Orientação: Prof. Dr. Cláudio Moisés Santos e Silva.

1. Energia eólica - Monografia. 2. Outliers - Monografia. 3. Análise subdiária - Monografia. I. Silva, Cláudio Moisés Santos e. II. Título.

RN/UF/CCET

CDU 621.548

## DEDICATÓRIA

*À minha família, meu mundo, DEDICO!*

## AGRADECIMENTOS

Confesso a vontade de possuir os dons da escrita, para assim poder transmitir toda a minha gratidão de forma clara e bela. Todavia, apesar das dificuldades em me expressar através das palavras, reconheço ser fundamental recordar e agradecer a todos os apoios nesta empreitada e etapa da vida. Assim, reservo este espaço para realizar meus sinceros agradecimentos, um tanto desajeitados e sob o risco de esquecimentos (peço que me perdoem), porém muito sinceros, a todos aqueles que direta ou indiretamente contribuíram para que eu nunca desistisse do meu sonho.

Agradeço, primeira e grandiosamente, a Deus, que com imensa bondade me deu força, saúde e perseverança para todas as experiências vividas. Ele quem esteve comigo em todos os momentos, e que sabe intimamente tudo que vivi e o que eu de viver, com a permissão dEle. Minha companhia mais fiel e o único capaz de saciar meu coração.

A meus pais, Cícero e Francisca Araújo e aos meus irmãos, Douglas e Elias Araújo, que sempre confiaram e me ajudaram, antes, durante e após o processo de escrita deste trabalho, fornecendo-me amor e atenção muito além do merecido. Aproveito o espaço para demonstrar minha gratidão aos meus avós: Altamiro e Francisca Alves, e Elias Araújo (in memoriam), por todo amor e exemplo em vida. Da mesma maneira, agradeço à minha avó, Maria de Lourdes Araújo, por todo carinho e cuidado. Gratidão a meu namorado, Walter Junior, pela calma necessária, carinho e incentivos durante esse percurso.

No ambiente acadêmico, destaco os incentivos do professor Dr. Cláudio Santos e Silva, pela orientação deste estudo, dedicação e todo o conhecimento transmitido: obrigada! Não poderia deixar de agradecer a todo o corpo docente da universidade por transmitirem seus conhecimentos nas aulas que tive o prazer de me matricular durante a minha formação. Estendo os agradecimentos às minhas orientadoras de estágio e pesquisa: Dr. Danielle Rodrigues, Dr. Keila Mendes e Dra. Samira Emiliavaca, dentre as outras mulheres merecedoras de todo reconhecimento por suas lutas diárias, e com as quais tive o privilégio de conviver na comunidade acadêmica.

Agradeço aos incontáveis amigos que essa graduação me propiciou conhecer, em destaque, ao grupo que sempre esteve presente e apto, às vezes nem tanto, para os desafios provindos dessa corrida: Everton Souza, José Augusto Neto, Moisés Costa e Thiago Silva. Foram inúmeras coisas vividas juntos nesses anos, sempre com o lema: “Vai dar certo, confia!”.

Aproveito ainda para relembrar os bons momentos e todo o aprendizado proporcionados por toda a equipe do ISI-ER, local de onde carrego inesquecíveis recordações e uma grande carga de aprendizado. Agradeço também aos amigos externos à UFRN que também me auxiliaram durante a graduação e a finalização desta pesquisa, dentre eles, o Prof. Emerson Gadelha. Enfim, pessoas fantásticas e que estiveram comigo nessa caminhada.

## RESUMO

O vento é uma importante variável meteorológica para planejamentos urbanos e rurais de atividades econômicas, como a produção de energia. Entretanto, há poucos estudos com abrangência nacional que visem fornecer e avaliar estatisticamente bancos de dados na escala horária, em parte, pela ausência de séries de dados observados sem falhas, ao longo da distribuição continental do Brasil. Assim, o objetivo da presente pesquisa é integralizar um banco de dados de velocidade do vento, com amostragem horária sem falhas para todo o território continental do Brasil. Utilizou-se o método de imputação múltipla para um conjunto de dados oriundos, inicialmente, de 449 estações automáticas de superfície, posteriormente reduzido para 421 estações. Os dados possuem amostragem horária e cobrem o período de 01 de janeiro de 2010 a 31 de dezembro de 2020. Tanto a série temporal original quanto a imputada, apresentaram grandes quantidades de *outliers*, afetando análises quantitativas. Foram imputados um total de 5.560.709 valores e, após a imputação, os valores do desvio padrão e da variância foram menores, com diferenças de 0,06 e 0,19, respectivamente, indicando uma diminuição da variabilidade dos dados. Verificou-se que os maiores valores foram observados nas regiões Sul e Sudeste. Os menores valores horários foram observados na região Nordeste. Alguns anos, como 2010, 2016, 2020 e 2021, apresentaram aumento no valor da média. O mês de setembro apresentou a maior variação dos dados antes e após a imputação. Os dados imputados apresentaram um valor mais elevado no primeiro quartil, Q1, em grande parte dos meses, quando comparado ao Q1 dos dados não imputados. Na análise horária, houve maior variação entre os dados originais e imputados entre 11:00h e 20:00h. Porém, o *outlier* de maior valor ocorreu às 09:00h. Conclui-se que as diferenças entre as estatísticas descritivas da série incompleta e da após a imputação não foram tão destoantes, fato que implica que os dados imputados foram gerados como valores próximos à mediana e sem impactar as características da série original.

**Palavras-chave:** Energia eólica, *outliers*, análise subdiária.

## ABSTRACT

Wind is an important meteorological variable for both rural and urban economic planning, such as energy production. However, there is a shortage of Nationwide studies aiming statistical database provision and evaluation on the time scale, which is partially due to the lack of flawless series of observed data across the Brazilian mainland. Thus, the objective of this research is to fulfill a database of wind speed, having time sampling without failures for the entire continental area of Brazil. Initially, the dataset was retrieved from 449 automatic surface stations, a number then reduced to 421, against which the multiple input method has been executed. Data have time sampling and coverage of the period between January 1<sup>st</sup>, 2010 and December 31<sup>st</sup>, 2020. Both the original time series and the inputs present a significant amount of *outliers*, which affect the quantitative analysis. A total of 5,560,709 values have been input and, after that, values of standard deviation and variance have been lower, with differences of 0.06 and 0.19, respectively, indicating a decrease of the data variability. Highest values have been observed in Southern and Southeastern regions. Lowest time values have been observed in the Northeastern region. Some years (2010, 2016, 2020 and 2021) present an average value increase. September presented the highest value variation of data before and after the input. Data input presented a higher value in the first quarter, Q1, in most of the months, when compared with the Q1 of the data not input. In the temporal analysis, there was a large variation between the original data and the input data between 11:00 AM e 08:00 PM. Yet, the *outlier* of highest level occurred at 09:00 AM. The study reveals that differences between descriptive statistics of incomplete series and the series after inputting have not been dissonant, which indicates that data input have been generated as values near the median, with no impact to the original series characteristics.

**Keywords:** Wind power, *outliers*, sub-daily analysis.

## SUMÁRIO

<b>RESUMO</b>	<b>7</b>
<b>ABSTRACT</b>	<b>8</b>
<b>SUMÁRIO</b>	<b>9</b>
<b>LISTA DE FIGURAS</b>	<b>11</b>
<b>LISTA DE TABELAS</b>	<b>12</b>
<b>LISTA DE ABREVIATURAS</b>	<b>13</b>
1 INTRODUÇÃO	14
2 OBJETIVOS	17
2.1 Objetivo geral	17
2.2 Objetivos específicos	17
3 FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA	18
3.1 Aspectos gerais de tempo e clima	18
3.2 Vento e circulação geral da atmosfera	21
3.3 Estudos sobre a variabilidade do vento no Brasil	24
3.4 Aplicações de métodos de imputação de dados	25
4. MATERIAL E MÉTODOS	27
4.1 Características gerais da variabilidade climática no Brasil por regiões	27
4.1.1 Região Norte (NB)	28
4.1.2 Região Nordeste (NEB)	28
4.1.3 Região Centro-Oeste (COB)	29
4.1.4 Região Sudeste (SEB)	30
4.1.5 Região Sul (SB)	31
4.2 Dados	32
4.3 Imputação múltipla de dados	34
4.4 Análise estatística e gráfica	36
5 RESULTADOS	39
5.1 Aspectos gerais	39
5.2 Análise por região geográfica do Brasil	40
5.3 Variação Interanual	43
5.4 Variação Mensal	44
5.5 Ciclo diurno	45

6 CONCLUSÕES E SUGESTÕES DE TRABALHOS FUTUROS

46

Referências

47

## LISTA DE FIGURAS

- Figura 1** - Distribuição da radiação solar incidente em relação à latitude e aos meses do ano. 19
- Figura 2** - Distribuição do vento e dos sistemas de pressão atmosférica para uma Terra em rotação ideal em macroescala. 23
- Figura 3** - Mapa ilustrativo da distribuição geográfica das estações meteorológicas mantidas separadas regionalmente. 33
- Figura 4** - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) considerando o conjunto total de dados. 40
- Figura 5** - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) separados por região geográfica do Brasil. 42
- Figura 6** - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) separados por ano. 43
- Figura 7** - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) separados por mês. 45
- Figura 8** - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) por hora. 46

## LISTA DE TABELAS

<b>Tabela 1</b> - Distribuição do número de estações considerando as regiões geográficas do Brasil.	33
<b>Tabela 2</b> - Valores das medidas de tendências centrais, e medidas de dispersão dos bancos de dados incompletos e imputados, considerando o conjunto total de dados.	40
<b>Tabela 3</b> - Valores das medidas de tendências centrais, e medidas de dispersão dos bancos de dados incompletos e imputados, horariamente.	42

## LISTA DE ABREVIATURAS

<b>BDMEP</b>	Banco de Dados Meteorológicos para Ensino e Pesquisa
<b>CO</b>	Centro-Oeste Brasileiro
<b>IM</b>	Imputação Múltipla
<b>INMET</b>	Instituto Nacional de Meteorologia
<b>MAR</b>	<i>Missing at Random</i>
<b>MCAR</b>	<i>Missing Completely At Random</i>
<b>NMAR</b>	<i>Not Missing At Random</i>
<b>NEB</b>	Nordeste Brasileiro
<b>NA</b>	<i>Not Available</i>
<b>NB</b>	Norte Brasileiro
<b>SD</b>	Desvio Padrão
<b>SEB</b>	Sudeste Brasileiro
<b>SB</b>	Sul Brasileiro
<b>TSM</b>	Temperatura da Superfície do Mar
<b>u2</b>	Velocidade do vento horizontal
<b>VAR</b>	Variância
<b>WRF</b>	<i>Weather Research and Forecasting Model</i>
<b>ZCIT</b>	Zona de Convergência Intertropical

## 1 INTRODUÇÃO

Sob o contexto do atual cenário mundial, onde consumo sustentável é um dos tópicos mais importantes no desenvolvimento econômico, o vento é uma das variáveis meteorológicas mais importantes, pois é crescente a produção de energia a partir desta fonte renovável. A importância da realização de estudos sobre vento e suas influências para diferentes atividades da sociedade são demonstradas em diversos estudos e por diferentes linhas de pesquisa.

Dessa forma, para enfrentar grandes desafios atuais, a caracterização da velocidade e direção do vento surge como uma solução para contribuir na melhoria dos planejamentos de atividades agrícolas, elaboração e execução de projetos de engenharia, na verificação da viabilidade e potencial de geração de energia eólica, na orientação para navegação aérea e marítima, bem como na análise do acamamento de plantas, deformação de paisagem e no planejamento ambiental (BAENA et al., 2005; COSTA, 2009; DE MORAIS et al., 2014; SANTOS; CARVALHO; REBOITA, 2016; ROCKETT et al., 2017; CASTELHANO, 2022; GOBO et al. (2017).

Em concordância com o avanço da tecnologia e o aumento da busca por fontes renováveis de energia, muitos países têm se preocupado em desenvolver estudos com ênfase na dinâmica da velocidade do vento, uma vez que o custo-benefício da geração de energia por fonte eólica apresenta desempenho satisfatório em comparação a outras fontes de energia renováveis, visto que já há, na cadeia produtiva da energia eólica, uma série de tecnologias bem avançadas, com custos relativamente baixos para a manutenção de equipamentos (PE, 2018). Portanto, o crescimento da produção de energia eólica se justifica uma vez que a demanda de energia elétrica de um país está correlacionada com a sua atividade econômica, que está diretamente relacionada ao Produto Interno Bruto do país (BRONZATTI; IAROSINSKI NETO, 2008).

O Brasil, com um vasto litoral e regiões com elevada topografia, tem obtido destaque devido à grande disponibilidade de recurso eólico em diversas partes de seu território continental e *offshore*. Neste cenário, a região Nordeste do Brasil (NEB) apresenta o maior potencial para a geração de energia eólica, com a variabilidade do vento apresentando elevadas frequências tanto de rajadas quanto

de valores de velocidade média (SANTOS; SANTOS E SILVA, 2013; LOPES et al., 2017; GILLILAND; KEIM, 2018b). Portanto, o NEB é a região com maior potencial para o aproveitamento do vento na geração de energia elétrica renovável (ANJOS, 2017; SOBRAL, 2009).

Neste contexto, o estudo do vento e suas características em diferentes escalas, incluindo as subdiárias torna-se fundamental para gestão de políticas públicas associadas à produção de energia, bem como e para o fomento e desenvolvimento de pesquisas científicas. Todavia, no Brasil, apesar do vento ser uma das mais importantes variáveis meteorológicas e, embora exista a necessidade do conhecimento prévio de suas características para projetos na área de energia, existe uma falta de cobertura espacial das estações meteorológicas em algumas regiões do país, além da escassez de dados em resolução subdiária e sub-horária.

Em função disto, observa-se que pesquisas referentes à dinâmica dos ventos com amostragem temporal de alta frequência, a despeito de sua importância demonstrada, ainda são incipientes. No âmbito das ciências climáticas são comumente encontradas séries temporais climatológicas provenientes de estações meteorológicas com dados ausentes em função de diversas razões, como por exemplo, interrupções e mau funcionamento ou manutenção dos instrumentos de medição.

Desta forma, um dos fatores de maior dificuldade relacionados às pesquisas em escala horária é a ausência e a qualidade das séries utilizadas, as quais em sua maior parte apresentam falhas, o que inviabiliza a caracterização da variabilidade de alta frequência do vento em algumas localidades. Esta problemática análoga à qualidade das séries históricas é habitual não apenas do Brasil, mas em várias partes da América do Sul. Ou seja, esses fatores constituem uma grande barreira para o desenvolvimento das pesquisas sobre eventos meteorológicos em escala subdiária, principalmente em nível nacional.

Portanto, a motivação do presente estudo é a necessidade de se realizar um processo de imputação dos dados ausentes, com vias a produzir uma base de dados em alta frequência (amostragem horária) completa de velocidade do vento no Brasil. O intuito deste estudo é no sentido de que os dados horários de vento podem ser usados para diversos setores socioeconômicos, como o exemplo citado da geração de energia eólica, para a escolha de locais adequados à instalação de um

sistema com capacidade de aproveitamento, são necessários estudos da frequência e da intensidade em que os ventos ocorrem (ZAHARIM et al., 2009).

## **2 OBJETIVOS**

### **2.1 Objetivo geral**

Gerar uma base de dados de velocidade do vento na escala horária para o território continental do Brasil a partir do método de imputação múltipla de dados.

### **2.2 Objetivos específicos**

- Comparar, por meio de estatística descritiva, as bases de dados com falhas e após a imputação dos dados ausentes;
- Caracterizar a variação diurna, mensal e anual dos dados originais e preenchidos, observando prováveis similaridades ou diferenças;

### 3 FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA

#### 3.1 Aspectos gerais de tempo e clima

Tempo é o estado da atmosfera em um dado momento e lugar, que pode sofrer mudanças abruptas (ou não) em um curto espaço de tempo ou manter-se com características pouco variáveis em determinados lugares ao longo das horas, dias e/ou semanas (SARTORI, 2003). Segundo Mendonça e Danni-Oliveira (2017), o tempo atmosférico é o estado momentâneo da atmosfera em um dado instante e lugar, e entende-se por estado da atmosfera o conjunto de atributos que a caracterizam naquele momento, tais como radiação, temperatura, umidade, precipitação, nebulosidade, pressão e ventos.

Por outro lado, o clima também corresponde às características da atmosfera, porém compreendendo os aspectos estatísticos de uma série de observações contínuas durante um longo período. Mendonça e Danni-Oliveira (2017), ressaltam que os conceitos clássicos de clima revelam a preocupação com a apreensão do que seja a característica do clima em termos do comportamento médio dos elementos atmosféricos, tais como a média da temperatura, precipitação, pressão atmosférica, componentes da radiação e, dentre outros, direção e velocidade do vento.

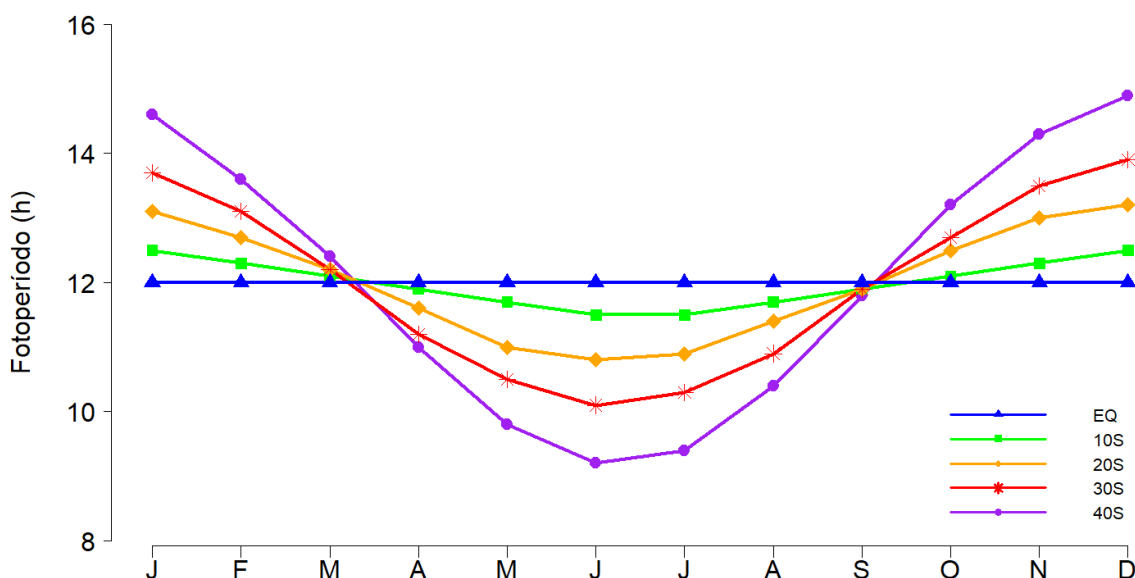
Dessa forma, o clima é afetado pelos diversos fatores climáticos, os quais são agentes causais externos ou internos que condicionam os elementos que variam no tempo e espaço e que compõem o clima (temperatura, umidade, vento, pressão atmosférica, etc.), comunicando assim, ao meio atmosférico, as suas propriedades, características e peculiaridades (VIEIRA; PICULLI, 2009). Esses fatores físicos podem ser subdivididos em três categorias: macro, meso e micro-escala. Alguns desses fatores climáticos são: a latitude, a distribuição dos oceanos e continentes, bem como a topografia da superfície (GUIMARÃES, 2015; VIEIRA; PICULLI, 2009).

Sob a perspectiva global, o tempo e clima e, portanto, as condições do vento em determinada região, são modulados pela circulação geral da atmosfera. Além disso, há fatores geográficos que influenciam diretamente na variabilidade climática de cada região. Como primeiro fator climático, há a latitude, a qual é um marco

geográfico, comumente utilizado em cálculos esféricos a partir da trigonometria esférica (DE OLIVEIRA; SARAIVA, 2015), e que geograficamente proporciona uma medida, em graus, da distância de um ponto sobre a superfície terrestre até a Linha Equatorial, responsável por subdividir o globo terrestre em Hemisférios: Norte e Sul. Ambos os hemisférios variam de  $0^\circ$  a  $90^\circ$ , de modo que o valor mínimo corresponde ao Equador e o máximo aos polos.

Associada a inclinação axial da Terra em relação à órbita elíptica ( $\sim 23^\circ 26'$ ) e com o movimento de translação da Terra, a latitude tem influência sobre a interação do sistema terra-sol. Portanto, os valores diários da radiação solar variam em função da latitude e do movimento de translação, fazendo com que os dias do ano e diferentes regiões tenham variações nos valores de incidência solar sobre a superfície e no fotoperíodo (Figura 1). Utilizando os parâmetros empregados sob a ótica geográfica da latitude, a região entre os trópicos ( $45^\circ\text{N}$  e  $45^\circ\text{S}$ ) recebe a radiação solar durante todo o ano na mesma proporção por latitude, enquanto que nas latitudes médias e altas, de  $45^\circ\text{N}$  a  $90^\circ\text{N}$  ou  $45^\circ\text{S}$  a  $90^\circ\text{S}$ , apresentam variações da radiação solar sazonalmente.

Figura 1 – Distribuição da radiação solar incidente em relação à latitude e aos meses do ano.



Fonte: Adaptado de Sentelhas e Angelocci (2012).

Dessa forma, quanto maior a latitude, maior a variação da radiação solar incidente durante o ano, assim como também será maior o ângulo de inclinação da radiação solar que incide na superfície (VIANELLO; ALVES, 2006). Ou seja, quanto maior a distância entre uma região e a Linha Equatorial maior será a sua variação da irradiação solar ao longo do ano, sendo um fator muito importante na definição do clima de cada região da Terra.

Essas diferenças de incidência solar ocasionam diferentes aquecimentos da superfície, gerando regiões com saldos de energia positivos, como é o caso da zona entre os trópicos, a qual recebe muita energia no período diurno, e outras com menores saldos de radiação, como as médias e altas latitudes. Assim, esse aquecimento diferencial do globo, juntamente com a distribuição assimétrica de oceanos e continentes e das características topográficas continentais, colaboram na geração da circulação geral da atmosfera, através do transporte de calor localizados nas áreas com maiores saldos de energia para regiões com menores valores, redistribuindo calor, umidade e quantidade de movimento de maneira heterogênea na Terra (FERREIRA; DA SILVA MELO, 2005).

Porém, ao analisarmos localmente a temperatura, além da latitude, devemos levar em conta a altitude e o relevo do local/região (planícies, platôs e planaltos). Neste sentido, a altitude ou elevação pode ser definida como sendo a distância vertical de um ponto em relação ao nível médio do mar (GROTZINGER; JORDAN, 2013). Em associação com o relevo e a latitude, a altitude é um fator condicionante dos elementos que compõem o clima como a precipitação pluvial, o fluxo de vento, e a temperatura de uma determinada região.

Fritzens et al. (2008) descreve que nas regiões tropicais e subtropicais, uma diferença de altitude de algumas centenas de metros provoca sensíveis mudanças no clima, no solo, na vegetação e em consequência na adaptação de animais e vegetais e no uso da terra. Milanesi (2007) descreve que os efeitos orográficos do clima são quaisquer controle que exerce o relevo sobre os atributos do clima em uma determinada localidade, gerando perturbações na dinâmica atmosférica em diferentes escalas, a depender da morfologia do relevo.

Vianello e Alves (2006) explica que ocorre uma taxa de arrefecimento do ar em decorrência da ascensão (movimento vertical para cima) das massas de ar, as quais ao ascenderem são expostas a uma menor pressão, o que faz com que haja expansão e diminuição da sua temperatura. Esse resfriamento ocorre gradualmente,

de modo que o gradiente vertical médio da troposfera é de aproximadamente  $0,6^{\circ}\text{C}/100\text{m}$  (MENDONÇA; DANNI-OLIVEIRA, 2007), podendo esse valor variar de acordo com a concentração de vapor de água na atmosfera.

Quanto à correlação entre a altitude, a precipitação e o fluxo de vento, há como exemplo a ocorrência de chuvas orográficas, que são oriundas da ação física do relevo atuando como uma barreira à advecção livre do ar. Uma vez que um fluxo de ar saturado é obrigado a elevar-se devido a um obstáculo de relevo, há o seu resfriamento em maiores altitudes sem haver a troca de calor entre ele e o meio, condensando-se e gerando nebulosidade (MILANESI e GALVANI, 2009), posteriormente podendo se converter em precipitação na encosta ativa do processo, a barlavento. Enquanto que o fluxo ao transpor a linha limite do relevo, agora com direção em sentido descendente, aquece e torna a atmosfera a sotavento relativamente seca.

### **3.2 Vento e circulação geral da atmosfera**

Considerando o aquecimento diferencial do globo pela radiação solar, a distribuição assimétrica de oceanos e continentes, além da topografia dos continentes, esse conjunto de Fatores Climáticos ocasionam, como consequência, um aquecimento diferencial das regiões da Terra e geram densidades diferentes no globo (AGNOLIN; MURARA, 2021; SILVA, 2018).

A diferença de densidade em torno do globo terrestre, a qual é ocasionada em função de fatores térmicos e/ou mecânicos, fazendo com que regiões que recebem maior quantidade de radiação solar, tenham o seu ar mais aquecido e, por consequência, seja menos denso, promovendo a ascensão do ar e formando os centros de baixa pressão. Por outro lado, nas regiões mais frias, o inverso ocorre, de modo que o ar mais frio tende a permanecer próximo à superfície, gerando centros de alta pressão.

Seguindo esse raciocínio, nota-se a existência de movimentos do ar ocorrendo, originadas nas áreas de alta pressão em direção às de baixa pressão, como resposta ao vazio deixado pelo ar aquecido que tende a ascender, de modo que esse espaço é sempre preenchido pelo ar vindo dos centros de alta pressão, causando o deslocamento horizontal do ar até que se estabeleça um equilíbrio barométrico (MARIN et al., 2008).

Desse modo, esses desequilíbrios termodinâmicos são os geradores da circulação geral da atmosfera, composta por grandes células de circulação, as quais são as principais responsáveis por promoverem as tentativas de retorno do sistema atmosférico ao seu estado de equilíbrio termodinâmico (CURRY; WEBSTER, 2005).

Para a compreensão climática da circulação atmosférica é preciso utilizar-se das escalas de análise: micro, meso e macro (AMBRIZZI, 2003). Quando trata-se de circulação em grande escala, refere-se a movimentos da atmosfera em proporções globais. Dessa forma, trata-se de um modelo de circulação composto por grandes células, em pares espelhados em cada hemisfério.

A primeira célula é a de Hadley, que possui ramos ascendentes sobre a linha equatorial e ramos descendentes sobre  $30^\circ$ , promovendo o movimento dos ventos em direção ao equador em superfície, onde ascendem direcionados às médias latitudes em altitude, transportando calor localizados na região equatorial com maiores saldos de energia para os trópicos, redistribuindo calor, umidade e quantidade de movimento (FERREIRA; DA SILVA MELO, 2005).

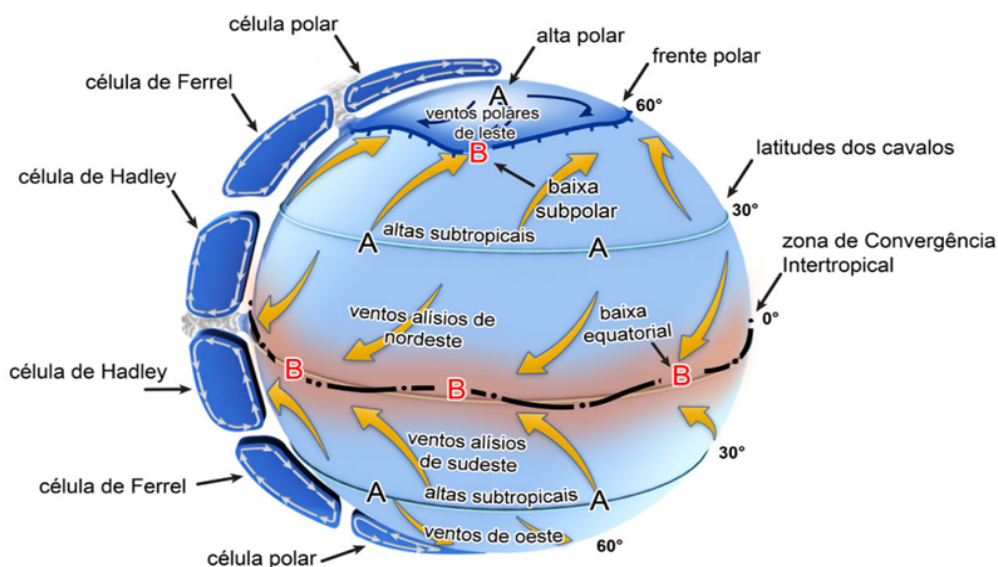
As células de Ferrel apresentam ramos descendentes em latitudes médias de  $30^\circ$  e ascendentes sobre  $60^\circ$ , promovendo o movimento dos ventos em direção aos pólos em superfície e em  $60^\circ$  ascendem direcionados ao Equador em altitude. Foi apresentada por William Ferrel (1856), reformulando o modelo de Hadley e gerando o modelo tricelular, no qual os ventos possuem um comportamento definido que foi sistematizado nas classificações da célula de Hadley (1735), célula de Ferrel e a célula Polar (TEIXEIRA, 2016).

Finalmente, as células polares possuem os ramos descendentes sobre os pólos e os ascendentes em  $60^\circ$ , promovendo o movimento do ar em direção aos pólos em superfície, os quais ascendem direcionados às médias latitudes em altitude.

Sendo assim, os centros de alta pressão estão localizados sobre  $30^\circ$  (Altas subtropicais) e  $90^\circ$  (Altas polares). Por outro lado, as regiões de baixa pressão sobre o equador e sobre os trópicos de capricórnio ( $45^\circ\text{S}$ ) e câncer ( $45^\circ\text{N}$ ), denominados de Baixa Equatorial e Subpolar, nessa ordem. A circulação geral da atmosfera controla o vento em grande escala e está diretamente associada à energia cinética gerada pelo movimento dessas massas de ar (LIMA, 2018). Sem o balanço térmico promovido pela circulação geral, a tendência seria de que os trópicos fossem ainda mais quentes, e de modo análogo, os pólos muito mais frios.

Ou seja, o vento tem como principal função distribuir o calor e massa na atmosfera. Além disso, as circulações oceânicas também desempenham papel fundamental na distribuição de energia no globo.

Figura 2 - Distribuição do vento e dos sistemas de pressão atmosférica para uma Terra em rotação ideal em macroescala.



Fonte: Ferreira (2020).

Dessa forma, os ventos se deslocam das altas para as baixas pressões, e sofrem influências da rotação da terra, da força de Coriolis, de forças gravitacionais, e da interação com a superfície terrestre (MUNHOZ, 2008). É devido à rotação da terra que é formado, em torno do equador, um regime de vento conhecido como ventos alísios, tais que ocorrem durante todo o ano na região tropical, produzindo a Zona de Convergência Intertropical (ZCIT).

Além da ZCIT, há diversos outros sistemas que ocorrem em função da circulação geral da atmosfera. Como exemplo, o Anticiclone Subtropical do Atlântico Sul, e as Correntes de Jatos, associados à célula de Hadley (LIMA, 2019).

O comportamento e a dinâmica dos ventos, por se tratar de um dos elementos que variam no tempo e espaço e que compõem o clima, são influenciados sob diferentes escalas. Como a velocidade do vento está diretamente relacionada à diferença de pressão entre duas regiões, quanto maior for essa

diferença, maior será a velocidade do deslocamento do vento. Em micro escala, o vento é influenciado, por exemplo, pelas características topográficas, rugosidade do solo, altura e possíveis obstáculos existentes na região (CRESESB, 2011), podendo apresentar variações significativas em distâncias curtas, como centenas de metros, em consequência da atuação dos fatores de influência no regime dos ventos.

Estudos demonstram a importância do conhecimento sobre a velocidade do vento, tais como Gobo et al. (2017), onde são associados a sensação térmica de desconforto com o efeito do chamado vento norte na cidade de Santa Maria, interior do Rio Grande do Sul. Alonso e Godinho (1992) demonstraram o efeito dos ventos sobre a poluição atmosférica para Cubatão, no ano de 1980. No âmbito da agricultura há o estudo realizado por Caramori et al. (1986), onde é feita uma comparação do desenvolvimento de mudas de cafeeiro sob efeito de diferentes fluxos de ventos, apresentando que a velocidade do vento influencia nas características das plantas, como taxa de transpiração, área foliar, diâmetro do caule e altura.

### **3.3 Estudos sobre a variabilidade do vento no Brasil**

Santos e Santos e Silva (2013) utilizaram dados de velocidade do vento derivados de anemógrafos Universal Fuess, a 10 m de altura, de 47 estações meteorológicas, para investigar aspectos climatológicos e tendências de velocidade do vento, entre 1986 e 2011. Foram aplicados métodos para preenchimento de dados faltantes, cálculos estatísticos quantitativos, *boxplots*, análise de cluster e análise de tendência usando o método estatístico Mann-Kendall. Concluíram que a variabilidade sazonal das médias da velocidade do vento de cada grupo apresentou os maiores valores durante o inverno e a primavera, enquanto que os menores durante o verão e outono. Também demonstraram que alguns grupos possuem uma grande amplitude na variabilidade sazonal da variável estudada, as quais foram associadas à ocorrência de eventos de El Niño e La Niña.

Em diversas literaturas a variável velocidade do vento é apresentada como estudo central. Gilliland e Keim (2018a) analisaram as características do vento de superfície do Brasil com base na localização do centro de alta pressão máxima no Atlântico Sul, conhecido como Anticiclone do Atlântico Sul, utilizando três conjuntos de dados de reanálise para o período de 1980 a 2014. Eles apresentaram a

existência de uma mudança latitudinal do sistema em direção aos pólos durante todas as estações do ano, além da existência de influências do Anticiclone sobre a velocidade do vento e a pressão ao nível do mar do norte do Brasil, variando as condições de acordo com a posição latitudinal do sistema.

Gilliland e Keim (2018b) também examinaram as características espaciais e temporais da velocidade do vento na superfície (10 m) do Brasil entre 1980 e 2014, utilizando dois conjuntos de dados *in situ* e três de reanálise climática, e utilizando a regressão quantílica para demonstrar as mudanças na velocidade do vento em diferentes percentis, apontando que os maiores valores positivos e negativos de tendências espaciais da velocidade do vento ocorrem nos percentis superiores entre 75 e 95%.

A fim de avaliar a variabilidade do vento e a densidade de potência simulados pelo *Weather Research and Forecasting Model* (WRF) para a região Sul do Brasil, Tuchtenhagen et al. (2020) comparou os resultados das simulações de produtos que quantificam a velocidade do vento sobre o oceano para um período de 5 anos, equivalentes ao intervalo entre 2006 e 2010. Foi demonstrado que o modelo WRF pode ser usado como uma ferramenta para avaliar o potencial de geração de energia eólica na região Sul do Brasil.

Santos et al. (2016), utilizando dados coletados *in situ* de torres anemométricas e uma simulação em mesoescala do WRF, quantificaram os recursos eólicos em duas localidades com diferentes topografias do NEB. Foi demonstrado que ambas as localidades apresentam potencial eólico anual favorável para implantação de empreendimentos eólicos em alturas de 60 e 50 m acima do solo.

### **3.4 Aplicações de métodos de imputação de dados**

Segundo Mcknight et al. (2007), de um modo geral, o termo “dados faltantes” significa a ausência de algum tipo de informação sobre o fenômeno em que se está interessado. Esses dados faltantes, geralmente, tornam-se um incômodo comum em análises científicas, causando imprecisões devido à incompletude (FARHANGFAR et al., 2004). Como exemplo disto, as ciências climáticas comumente apresentam dados faltantes (*missing data*) em decorrência de diversos fatores, entre eles: entrada de dados manualmente (FARHANGFAR et al., 2007),

equipamentos e instrumentos de medições com falhas operacionais (COLANTONIO et al., 2010) e manutenção dos instrumentos de medição.

A prevenção da ocorrência de dados faltantes sempre será a melhor solução. Todavia, quando a precaução não é acessível ou viável, o tratamento dessas bases de dados é a melhor solução para se obter resultados confiáveis. Portanto, é de suma importância o uso de estratégias de tratamento dos dados faltantes nas variáveis, sendo uma dessas soluções a abordagem dos dados faltantes utilizando técnicas adequadas desenvolvidas para contornar esse problema (VAN DER HEIJDEN, 2006).

Assim, torna-se imprescindível determinar uma abordagem de análise que seja adequada para um banco de dados incompleto, o que se torna uma das dificuldades conceituais e computacionais levantadas nesse processamento (SCHAFER & GRAHAM, 2002), uma vez que métodos inadequados podem levar a conclusões erradas sobre uma série de dados (RUBIN, 1996; LITTLE; RUBIN, 2002; MCKNIGHT et al., 2007; COLANTONIO et al., 2010).

Diante dessa problemática, diversos métodos estatísticos, que envolvem o preenchimento dessas lacunas utilizando valores admissíveis, têm sido desenvolvidos a fim de solucionar esses problemas de dados ausentes, e tem sido uma área de pesquisa muito ativa nas últimas décadas (RUBIN, 1987; SCHAFER, 1999; ZHANG, 2003; VAN DER HEIJDEN, 2006; HAREL; ZHOU, 2007; KENWARD; CARPENTER, 2007). Com a ideia de se resolver essas questões foi desenvolvida (RUBIN, 1987) a técnica de Imputação Múltipla (IM).

O emprego da Imputação implica no preenchimento de dados faltantes em uma dada série temporal, utilizando uma estimativa de possíveis valores para compor essas lacunas. Destacam-se os métodos de Imputação Simples, o qual consiste, basicamente, em substituir todos os dados faltantes na série com um único valor, e o método de Imputação Múltipla, no qual serão preenchidos por  $n$  valores, com  $n \geq 2$  (RODRIGUES, 2019). A partir desse segundo, são formadas  $n$  bases de dados, completas, e que podem ser analisadas através de procedimentos convencionais (RUBIN, 1987; VERONEZE, 2011).

Sabe-se que a ausência de dados afeta diretamente os resultados finais de uma determinada análise, ocasionando a perda de eficiência e impedindo a utilização de alguns métodos estatísticos paramétricos (MOLENBERGHS et al., 2004; VERONEZE, 2011). Da mesma forma que o não tratamento dos dados

faltantes, a aplicação de métodos de imputação de modo inadequado também pode ocasionar em interpretações equivocadas e afetar a generalização dos resultados das análises (RUBIN, 1996; LITTLE; RUBIN, 2002; MCKNIGHT et al., 2007; COLANTONIO et al., 2010).

Dessa forma, o método de IM, quando corretamente aplicado, produz estimativas não viciadas mesmo ocorrendo a perda completamente aleatória dos dados. Por esse motivo, a IM tem se tornado um método cada vez mais aplicado no tratamento de dados faltantes, sendo um método flexível quanto à aplicabilidade, se bem usada, pode lidar com dados faltantes de todos os tipos (quantitativos, categóricos ordinais, nominais, etc) (NUNES, 2007; WHITE, 2007).

Na literatura é possível encontrar em trabalhos como Firat et al. (2010), a utilização da imputação de dados como solução para observações faltantes de séries de precipitação de estações meteorológicas localizadas na Turquia. Há também trabalhos como o de Xavier (2016; 2022), utilizando a imputação como um método para a geração de um novo banco de dados para o Brasil. Rodrigues (2019) também aplicou métodos de imputação para preencher lacunas ocasionadas por dados faltantes de séries de precipitação provenientes de estações meteorológicas no nordeste do Brasil. Como exemplo da utilização de métodos de imputação como tratamento de dados de velocidade do vento, Wesonga (2015) realizou o preenchimento de estações localizadas em Uganda.

## **4. MATERIAL E MÉTODOS**

### **4.1 Características gerais da variabilidade climática no Brasil por regiões**

O Brasil é o quinto país do mundo com a maior expansão territorial, com aproximadamente 8.510.345,5 km<sup>2</sup> (IBGE, 2021), correspondendo a quase 50% de todo o território Sul-Americano. É constituído por 5 regiões: Centro-Oeste (COB), Nordeste (NEB), Norte (NB), Sudeste (SEB) e Sul (SB). Como consequência dessa imensa extensão, possui uma grande variabilidade climática, devido aos diversos fatores que determinam o clima do país (MENDONÇA; DANNI-OLIVEIRA, 2007).

Cerca de 94% de todo o território brasileiro encontra-se com predominância de climas quentes e com pouca amplitude térmica, subdividindo-se em cerca de 55% na zona equatorial e 39% tropical, enquanto que os demais 6% equivalem a

um território encontrado na faixa climática subtropical, apresentando temperaturas médias mais amenas e grandes amplitudes térmicas durante o ano (MENDONÇA; DANNI-OLIVEIRA, 2007). Porém, essa região subtropical, devido à proximidade e às características topográficas da América do Sul, finda sendo influenciada pela região tropical, especialmente pela umidade oriunda da Amazônia.

As zonas climáticas estão distribuídas pelo território brasileiro em função também do padrão de distribuição da radiação solar, a qual é um dos fatores que determinam o clima de um lugar (MENDONÇA; DANNI-OLIVEIRA, 2007). Dessa forma, é necessário entender a distribuição solarimétrica do Brasil. Serão apresentados as caracterizações de cada região brasileira.

#### **4.1.1 Região Norte (NB)**

Sendo a maior região do Brasil em termos de extensão territorial, corresponde a cerca de 3.870.000 km<sup>2</sup> (IBGE, 2021). Possui o total de 7 unidades federativas: Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e Tocantins. A região da Floresta Amazônica pertencente ao território brasileiro se encontra na NB, sendo a floresta mais importante e rica do mundo. Em termos de clima predominante, a região é compreendida como equatorial, o que indica altas temperaturas, muita chuva, alta umidade e baixa amplitude térmica.

A vegetação presente em grande parte da Floresta Amazônica é composta por árvores altas, largas e juntas. Isso torna o norte uma região com muitas florestas e uma grande variedade de fauna e flora. Além disso, há o extenso litoral norte do estado do Pará e o litoral do Amapá. As menores temperaturas dessa região se dão nas regiões serranas, enquanto que as mais elevadas se encontram ao longo do vale do Rio Amazonas, e devido à umidade atmosférica e nebulosidade elevada, possui homogeneidade térmica significativa (MENDONÇA e DANNI-OLIVEIRA, 2007). Além de ser uma região que possui a ZCIT como um importante vetor na definição da dinâmica da atmosfera.

#### **4.1.2 Região Nordeste (NEB)**

A região Nordeste é a terceira maior do Brasil em área, correspondendo a 1.558.000 km<sup>2</sup> (IBGE, 2021), cerca de 18,31% do território brasileiro, e sendo

identificados 3.338 km de costa litorânea. É composta por 9 unidades federativas: Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe. É uma região que possui diferentes topografias, como a Chapada Diamantina na Bahia, dunas, falésias, regiões serranas e o sertão, com vegetação típica da Caatinga.

Há também o destaque para o estado do Maranhão, que apresenta características e vegetação de três diferentes biomas, de modo que a leste há os Lençóis Maranhenses, localizado às margens do Oceano Atlântico, a oeste há a selva amazônica, e ao sul as chapadas do Cerrado. O litoral do estado do Piauí apresenta o Delta do Parnaíba, que é um conjunto natural de ilhas e ecossistemas, com variados mangues, dunas e igarapés.

O NEB apresenta atributos climáticos extremos quando comparados às demais regiões brasileiras; como exemplo, em algumas zonas dessa região, destacam-se baixos níveis de umidade, escassez de chuvas anuais, irregularidade nos índices de precipitações ao longo dos anos, períodos de estiagem longos, ausência de rios perenes e irregulares precipitações pluviais extremamente limitadas a um curto período do ano (AB'SABER, 1974; REIS, 1976). Sendo assim, as adversidades climáticas estão associadas a essas variabilidades tanto espacial quanto temporal, que caracterizam a região como semiárida. (AZEVEDO; SILVA; RODRIGUES, 1998). Porém, segundo Kayano e Andreoli (2009), na região são reconhecidos três tipos climáticos: Clima Litorâneo Úmido, Clima Tropical e Clima Tropical Semiárido.

A Zona de Convergência Intertropical (ZCIT) é um dos fatores de maior influência sobre a dinâmica atmosférica do NEB (MENDONÇA e DANNI-OLIVEIRA, 2007; MARENGO, 2001), além da Temperatura da Superfície do Mar (TSM) do Atlântico Sul e Atlântico Norte, e outros sistemas de tempo que possuem escalas espaciais e temporais distintas (MARENGO; VALVERDE, 2007; MARENGO et al., 2011; FERREIRA; REBOITA; ROCHA, 2019).

#### **4.1.3 Região Centro-Oeste (COB)**

Formado por 3 estados e o Distrito Federal: Goiás, Mato Grosso e Mato Grosso do Sul. Corresponde a cerca de 19% do território nacional, equivalentes a 1.612.000 km<sup>2</sup> (IBGE, 2021). Uma das características mais marcantes dessa região

é a vasta fauna e flora, e abundância das águas. Dessa forma, o clima predominante do COB é o tropical semi úmido, apresentando apenas duas estações definidas pela ocorrência das chuvas, uma denominada de “verão” que é a estação chuvosa e que ocorre entre novembro e abril, e uma denominada “inverno”, entre maio e agosto, com temperaturas acima de 30°C, com mínimas entre 10°C e 15°C.

Quanto às características geográficas do COB, é destacada pela composição do Pantanal, o qual possui a sua porção norte como a maior planície inundável do planeta. Toda a região apresenta uma grande biodiversidade, cerca de 2000 espécies de plantas que também possuem características de outros biomas vizinhos, como da Amazônia, ao norte; Cerrado, a leste; Mata Atlântica, ao centro-sul; e Chaco da Bolívia e do Paraguai, a oeste (ALHO, 2005; POTT et al., 2011).

O estado do Mato Grosso, é composto por 3 biomas distintos: o Pantanal, a Floresta Amazônica e o Planalto Central do Brasil. Além do Pantanal, o COB é destacado pela Chapada dos Veadeiros, com muitas serras, matas, cachoeiras e cavernas (PRZYBYSZEWSKI, 2017). Mais ao sul da região, também há a Serra da Bodoquena, onde há lagos, cachoeiras, grutas, cavernas, nascentes, riachos, aquírios naturais e rios de águas transparentes (PRZYBYSZEWSKI, 2017). Por outro lado, Brasília, capital do Brasil, apresenta baixa umidade relativa do ar e alta temperatura média anual, por ser uma região de baixa latitude como as planícies e planaltos.

#### **4.1.4 Região Sudeste (SEB)**

Considerada a região mais populosa do país, é composta por 4 estados: Espírito Santo, Minas Gerais, Rio de Janeiro e São Paulo. Correspondente a cerca de 924.511 km<sup>2</sup> (IBGE, 2021), aproximadamente 10,86% do território brasileiro. A região apresenta um clima variado. A variação climática é subdividida como: um clima Tropical sobre o litoral do Rio de Janeiro e Espírito Santo, o norte de Minas Gerais, e o oeste paulista; um clima Tropical de Altitude nas regiões com maiores altitudes como a Serra do Mar, no Rio de Janeiro; Semiárido, presente no estado de Minas Gerais, com aspecto de temperaturas mais altas, e menores umidades relativas; e o Subtropical, ao sul de São Paulo, com chuva bem distribuída e grande variação térmica (PRZYBYSZEWSKI, 2017).

O relevo do SEB também é um grande destaque por não apresentar singularidade, apresentando quatro tipos: planícies e terras baixas costeiras, serras e planaltos do leste e do sudeste, planalto meridional e a depressão periférica (PRZYBYSZEWSKI, 2017). No SEB há uma topografia onde se sobressaem serras, chapadas, como Mantiqueira, Canastra, Serra do Mar, que favorecem a ocorrência de precipitação mais acentuada. Os litorais do Rio de Janeiro, Espírito Santo e São Paulo também são destaque quanto à classificação climática da região. Na capital capixaba há também quase 20 áreas de preservação ambiental, como o manguezal urbano e grandes áreas de preservação de fauna e flora locais.

#### **4.1.5 Região Sul (SB)**

Sendo a menor região do Brasil em extensão territorial, o SB ocupa uma área de 576.774,31 km<sup>2</sup> (IBGE, 2021), e é composta por três estados: Paraná, Rio Grande do Sul e Santa Catarina. Localizada na zona subtropical, ou seja, abaixo do trópico de capricórnio, apresenta um clima muito intrínseco dessa região em comparação com as demais, pertencendo a uma zona temperada com temperaturas médias anuais entre 14°C e 22°C. Essa região também apresenta a ocorrência de geadas em determinados períodos do ano em função de sua posição geográfica e relevo.

A região Sul possui um relevo muito variada quanto à sua composição, apresentando deste um extenso litoral que decorre em todos os estados da região, planaltos, planícies, montanhas, baías, serras, cânions, além das Cataratas do Iguaçu, localizadas no Parque Nacional do Iguaçu que é considerado Patrimônio Natural da Humanidade desde 1986. Grande parte do SB é ocupada pelo Planalto Meridional. As condições de geadas acontecem em sua maioria sobre os relevos de grandes altitudes localizadas nos estados do Rio Grande do Sul e de Santa Catarina, podendo em casos extremos a ocorrência de neve.

A região catarinense possui paisagens excêntricas, cercadas por uma floresta de Araucárias, vales, campos, rios e cânions, sendo destaque, durante os períodos de temperaturas baixas, a Serra do Rio do Rastro, o Morro da Igreja e a Serra do Corvo Branco (PRZYBYSZEWSKI, 2017). Na costa litorânea do Paraná, encontra-se a Ilha do Mel, tombada como Patrimônio Ecológico da Humanidade e cercada por ecossistemas de Restinga e Floresta Atlântica.

Portanto, entende-se que as temperaturas, e conseqüentemente as zonas climáticas, são subdivididas seguindo o padrão de distribuição da radiação solar (MENDONÇA; DANNI-OLIVEIRA, 2007), de modo que as temperaturas médias anuais diminuem gradativamente de norte ao sul, reforçando essa correlação com a radiação solar. Sabendo disso, as maiores temperaturas médias anuais ocorrem no NB e NEB do país (podendo superar 26°C), as temperaturas intermediárias (24,1 °C a 26 °C) ocorrem na maior parte do país e no extremo sul do país são verificadas as menores temperaturas médias anuais abaixo de 18°C (MENDONÇA; DANNI-OLIVEIRA, 2007).

O Brasil, por possuir uma grande extensão territorial, tende a apresentar variados fenômenos climáticos. A exemplo disso há o contraste entre uma Amazônia quente e úmida e o NEB semiárido. Martins (2022) apresenta que a distribuição espacial das médias anuais destaca os dois grandes contrastes pluviométricos do país: a Região Norte, com as médias mensais mais elevadas (superiores a 2.800mm) e o sertão nordestino (média anual entre 125mm e 1200mm).

#### 4.2 Dados

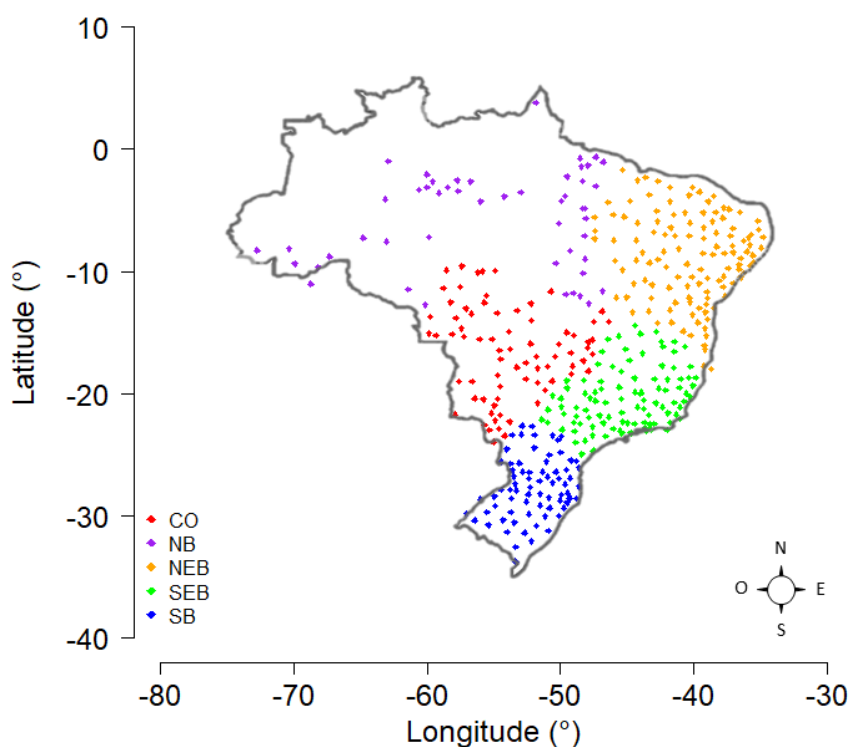
Os dados utilizados foram provenientes de estações meteorológicas pertencentes ao Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) do Instituto Nacional de Meteorologia (INMET) - disponível em: <www.inmet.gov.br>. Referem-se à série de velocidade do vento (m/s) horária, coletadas de 449 estações meteorológicas de superfície distribuídas pelas 5 regiões geográficas do Brasil, no período entre 01/01/2010 e 31/12/2021.

Para a análise dos dados horários, foram aplicados procedimentos a fim de evitar falhas instrumentais. A metodologia utilizada tomou como base a utilizada por Xavier et al. (2022); porém consistindo apenas de sua primeira etapa, de modo que foram excluídos os dados que ultrapassam os limites inferiores e superiores pré-especificados, removendo *outliers* “óbvios”, mantendo apenas os valores da velocidade do vento ( $u_2$ ) com valor superior ou igual a 0 m/s e inferior a 100 m/s:

$$0 \text{ m/s} \leq u_2 < 100 \text{ m/s} \quad (1)$$

Na segunda etapa, foram descartados os dados derivados de estações com dados faltantes superiores a 40% de sua composição total, totalizando 28 estações descartadas e 421 mantidas (Figura 3). Na tabela 1 são apresentadas as distribuições espaciais das estações conservadas.

Figura 3 - Mapa ilustrativo da distribuição geográfica das estações meteorológicas mantidas separadas regionalmente.



Fonte: da autora

Tabela 1 - Distribuição do número de estações considerando as regiões geográficas do Brasil.

<b>Região geográfica</b>	<b>Quantidade de estações mantidas</b>
COB	75
NEB	121
NB	50
SEB	102
SB	73

Fonte: da autora

### 4.3 Imputação múltipla de dados

Seguindo o que foi proposto por Rubin (1987), a fim de resolver o problema de ausência de dados em pesquisas e incluir a incerteza da imputação nos resultados, corrigindo o maior problema associado à imputação única, a Interpolação Múltipla, basicamente, inclui 3 etapas:

- 1° etapa: é utilizado um banco de dados, composto por dados ausentes e dados observados, para ser aplicado sobre essa série temporal um método de IM, gerando  $n$  bancos de dados completos. Uma vez realizada a imputação, os dados observados mantêm os mesmos valores originários. Porém, cada banco de dados terá a variação dos  $n$  valores imputados, e é a magnitude dessa assimetria entre os dados imputados que traduz as incertezas dos valores gerados. Durante essa etapa, é necessário identificar o mecanismo de ausência e o padrão dos dados faltantes, uma vez que a validade dos resultados dos diferentes métodos, aplicados a dados incompletos, dependem do mecanismo associado às ausências. Rubin (1976) apresentou três mecanismos teóricos gerais que aplicam um padrão sobre a ocorrência da ausência de dados: perdas completamente ao acaso (*missing completely at random* – MCAR), perdas ao acaso (*missing at random* – MAR) e perdas não-aleatórias (*not missing at random* – NMAR) (CAMARGOS, 2011; VERONEZE, 2011).
- 2° etapa: Consiste na aplicação de métodos estatísticos tradicionais de análise padrão, próprios para dados completos, tratando os dados gerados através da imputação como se realmente fossem conjuntos completos de dados.
- 3° etapa: Nessa etapa os resultados das  $n$  análises devem ser combinadas a fim de produzir estimativas que consideram a incerteza dos valores imputados, dessa forma é obtida a inferência da imputação utilizada (VERONEZE, 2011). Para essa etapa são aplicadas as regras propostas por Rubin (1987), que são comumente encontradas e utilizadas amplamente nas literaturas referentes à IM.

O método de imputação é aplicado com o objetivo de que sejam geradas estimativas dos dados faltantes ( $X_{faltantes}$ ). De acordo com Schafer (1999),  $X$  pode ser a média, a correlação, o coeficiente de regressão ou a razão de chances. Ou seja, qualquer medida escalar a qual se deseje estimar. Neste estudo, a estimativa foi a média das estimativas individuais.

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n \hat{X}_j \quad (2)$$

O método de IM escolhido durante o estudo foi baseado no estudo feito por Rodrigues (2019), onde foi utilizada a reamostragem *bootstrap* via algoritmo *Expectation Maximization* (EM), designado BootEM.

Durante a caracterização da primeira etapa, o método de IM BootEM, segundo Rodrigues (2019), é caracterizado por imputar os valores faltantes de uma série temporal por estimativas geradas pelos algoritmos EM implementados em várias amostras *bootstrap* dos dados originais (RUBIN, 1994; HONAKER KING; BLACKWELL, 2011). A simulação dos dados se deu através do algoritmo BootEM, implementado no pacote Amélia II (HONAKER; KING; BLACKWELL, 2011) do *software* R (R. Core Team 2015), o qual possui um algoritmo voltado para a imputação múltipla utilizando dados *cross-section* e séries temporais, e que apresenta cinco opções de funções de transformação da série a fim de melhorar a qualidade dos valores imputados (DE SOUZA, 2014).

Utilizando o algoritmo BootEM, são criados  $n$  bancos de dados completos através da imputação de  $n$  valores para cada dado faltante da série temporal, de modo que todos os bancos de dados gerados mantenham os valores prévios dos dados observados, gerando variações apenas aos valores imputados.

O modelo de imputação através do pacote Amélia necessita que o conjunto de dados  $X$  (composto por observações e falhas) possua uma distribuição normal multivariada com vetor de médias  $\mu$  e matriz de covariâncias  $\Sigma$  (AMARAL, 2014), ou seja,

$$X \sim N(\mu, \Sigma) \quad (3)$$

É assumido que as falhas são aleatórias do tipo MAR, uma vez que não há padrões referentes às falhas, e que constituem uma matriz  $X_{faltantes}$  de falha que depende apenas de  $X_{observados}$ . Como o conjunto  $X$ , a princípio, é constituído por  $X_{observados}$  e por  $X_{faltantes}$ , a imputação múltipla irá considerar os parâmetros dos dados completos,  $\theta = (\mu, \Sigma)$ . Assim, ao ser considerado que as falhas são do tipo MAR e que a matriz  $X_{faltantes}$  é independente dos parâmetros  $\mu$  e  $\Sigma$  dos dados completos, é definida a verossimilhança. Esse método é descrito em Amaral (2014).

Segundo Dantas et al. (2016), o algoritmo BootEM seguramente converge para o estimador de máxima verossimilhança, baseando-se na ideia de substituir uma difícil maximização por uma sequência de maximização mais fáceis, envolvendo dois passos: o passo “E” que calcula o valor esperado do logaritmo da verossimilhança completo; e o passo “M”, que encontra seu máximo. Os passos são repetidos até se atingir a convergência.

Durante a reamostragem foram observados valores negativos. Uma vez que a velocidade do vento não apresenta variações negativas, esses valores foram descartados da análise, de modo que foram aplicadas médias aritméticas sobre os demais dados imputados para cada dado faltante que não sofreram esse truncamento, resultando em bancos de dados sem falhas (Q) para cada estação meteorológica, através do valor médio da imputação múltipla. Segundo Amaral (2014), também é possível aplicar métodos como o coeficiente de regressão ou uma primeira diferença, bem como sortear o valor de Q dentre todos os valores  $n$  imputados para cada observação perdida.

#### 4.4 Análise estatística e gráfica

Com o objetivo de demonstrar graficamente as diferenças estatísticas entre os bancos de dados incompletos e os bancos de dados imputados, foram gerados 4 composições de *boxplot* com orientação vertical, divididos em escala temporal e espacial: anual, horário, mensal e por região.

O *boxplot* é um modelo gráfico, útil para sumarizar e analisar dados quantitativos, que possui um formato de caixa e hastes, subdividido por quartis (Q1, Q2 e Q3), que exhibe medidas estatísticas dos dados representados. A estrutura

básica é construída por: uma caixa, que assume comumente o formato retangular; a mediana, representada por uma linha localizada dentro da caixa e correspondente ao segundo quartil (Q2); por hastes que se assemelha à letra “T”, representando os valores compreendidos entre a caixa e os valores limites inferior e superior do conjunto de dados, de modo que a extremidade é comumente denominada de fence; e em caso de dados discrepantes, atípicos ou extremos, o *boxplot* pode apontar a presença de *outliers*, que são habitualmente representados por símbolo excedentes aos hastes.

Dessa forma, esse tipo de gráfico exibe a tendência central não-paramétrica (mediana), a dispersão dos dados (através dos quartis), a simetria ou distribuição da amostra (máximos e mínimos) e possíveis valores atípicos e/ou extremos (*outliers*). Referente à estrutura do modelo de *boxplot* utilizado no atual trabalho, no eixo vertical estão representados os valores numéricos referentes à velocidade do vento provenientes dos bancos de dados analisados, e no eixo horizontal são apresentadas as subdivisões relativas às escalas aplicadas citadas (temporal e espacial), chamadas de fatores de interesse. Em resumo, segundo Neto (2017), os quartis são divididos em 4 medidas:

- Primeiro quartil (Q1): são localizados 25% dos valores do conjunto de dados abaixo de sua estrutura, de modo que também é chamado de quartil inferior ou 25º percentil. Representado pela linha limite inferior da caixa;
- Mediana ou segundo quartil (Q2): divide a caixa de modo a manter a parte superior e inferior à sua estrutura com 50% dos valores em ambas. É o 50º percentil. Representada pela linha dentro da caixa;
- Terceiro quartil (Q3): onde se localiza 75% dos valores. Também chamado de quartil superior ou 75º percentil. Representado pela linha limite superior da caixa.
- A amplitude entre o primeiro e o terceiro quartil determina o intervalo interquartílicos. No gráfico, sua dimensão é identificada pela extensão da caixa, e estende-se do Q1 a Q3, ou seja, corresponde ao intervalo dos 50% dos dados em torno da mediana. Com relação aos limites superiores e inferiores, desconsiderando os *outliers*, compreendem os dados que tenham valores máximos e mínimos da série, respectivamente, e que não ultrapassem a proporção de uma vez e meia o valor do intervalo interquartílico. Os dados que ultrapassem essa medida serão definidos e

apresentados como *outliers*. Os valores extremos ultrapassam a medida de duas vezes e meia o intervalo interquartilico (NETO, 2017). A aproximação entre a mediana e o valor da média (que pode ser adicionada a gráficos de *boxplot*), aponta uma posição simétrica da mediana dentro da caixa; e forma antagônica, a posição assimétrica aponta uma aproximação com dados não-paramétricos, os quais podem estar mais próximos do quartil inferior (Q1) ou do quartil superior (Q3).

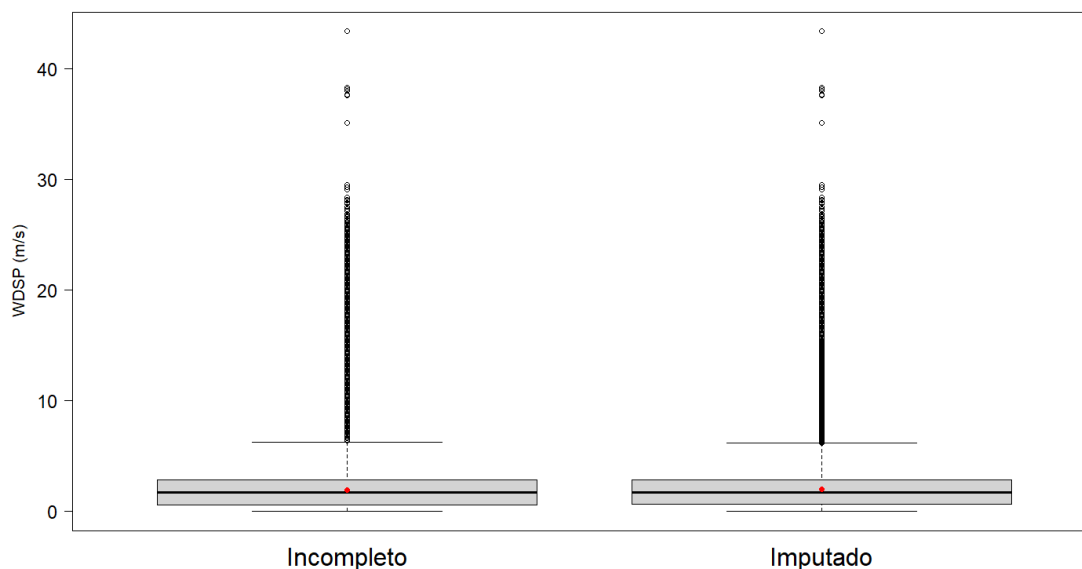
## 5 RESULTADOS

### 5.1 Aspectos gerais

Apresenta-se na Figura 4a e 4b a distribuição do banco de dados incompletos e imputados, respectivamente. Nota-se a existência de uma grande quantidade de *outliers* em ambos bancos de dados. No entanto, mesmo com a presença de muitos *outliers*, a média, que é uma medida influenciada por esses dados atípicos, apresenta uma aproximação da mediana (a qual não sofre influência de valores discrepantes ou anômalos), antes e depois da imputação. Como é possível verificar através da Tabela 2, o intervalo interquartilico tornou-se maior após a imputação, o que indica um aumento na quantidade de dados em torno da mediana, ou seja, dentro da caixa. O intervalo entre a haste inferior e o Q1 tem proporção muito menor que a quantidade de dados acima do Q3, o que indica que os dados com valores acima da média e da mediana apresentam mais variabilidade do que os 25% inferiores.

A quantidade de dados imputados no total, representado na Tabela 2 por NA (*not available*), correspondeu a 5.560.709 valores horários, o que representa cerca de 14% da base de dados. Após a imputação, o Desvio padrão (SD), que expressa o grau de dispersão de um conjunto de dados (quanto maior, mais dispersos são os dados), apresentou um valor menor, o que consente com o valor da Variância (Var), a qual determina o afastamento entre os dados e a média, e que também diminuiu após a IM. Apesar da diferença de 0,3 m/s entre as duas medianas, as médias se mantiveram iguais.

Figura 4 - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) considerando o conjunto total de dados.



Fonte: da autora

Tabela 2 - Valores das medidas de tendências centrais, e medidas de dispersão dos bancos de dados incompletos e imputados, considerando o conjunto total de dados.

	MÍN	Q1	Q2	MÉD	Q3	MAX	SD	VAR	NA
INCOMPLETO	0,00	1,00	2,00	2,00	3,00	43,00	1,74	3,02	5560709
IMPUTADO	0,00	0,70	1,70	2,00	2,90	43,40	1,68	2,83	-

Fonte: da autora

## 5.2 Análise por região geográfica do Brasil

Na Figura 5 são apresentados os *boxplots* dos dados incompletos e imputados, separados por região geográfica do Brasil. Os dados imputados apresentaram uma aproximação maior entre a mediana e a média, principalmente para COB, NB e SEB. Ainda com relação às medianas, todas as regiões apresentaram aumentos consideráveis dessa medida, enquanto que a média não apresentou grandes variações, tendo o COB apresentado uma média menor após a imputação.

Ao comparar as variâncias regionalmente, é possível observar que, apesar de SB apresentar a segunda menor quantidade de estações entre as regiões, os

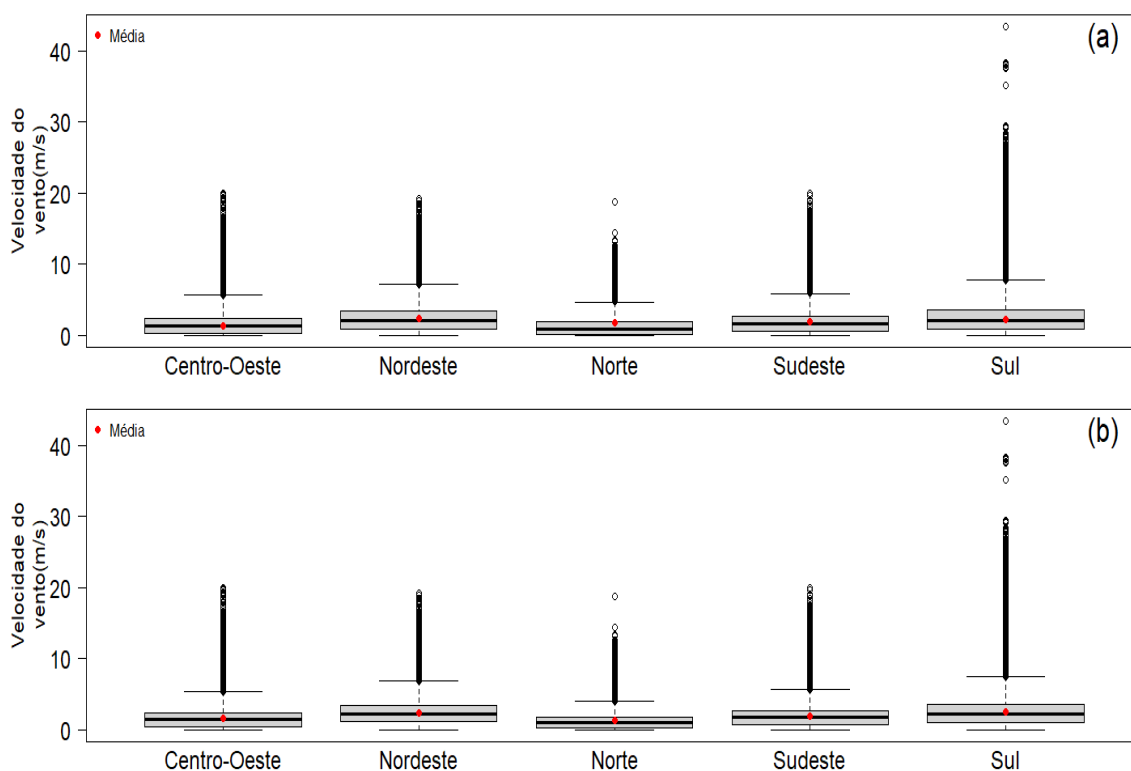
valores de Var e SD foram os maiores, antes e depois da imputação, o que pode estar associado à grande variabilidade da velocidade do vento na região ou em decorrência da grande quantidade de *outliers* presente. Em função disso, observa-se através da largura ou altura das caixas, que apesar da pouca variabilidade dos dados entre os 25% inferiores, o SB apresenta a caixa mais alongada entre todas as regiões, bem como um intervalo relativamente grande entre o Q3 e a haste superior, os maiores valores de velocidade do vento da série.

O NB, apresentando a menor caixa e a haste superior com a altura inferior às demais regiões, apresentou a menor quantidade de dados ausentes, vinculado ao fato de possuir a menor quantidade de estações, apresentou os menores valores de Var e SD, e a menor diferença entre a média e a mediana, antes e após a IM. O grande intervalo entre Q3 e a haste superior aponta que grande parte das observações encontram-se acima da média e da mediana, porém foi a região que apresentou o menor valor máximo de velocidade do vento, o qual se manteve o mesmo antes e depois da IM, bem como foi a única região com mudança negativa do valor de Q3.

O NEB, por apresentar a maior quantidade de estações, foi a região que possui o maior número de dados faltantes, o que possibilita um impacto mais significativo da imputação sobre as medidas estatísticas. Todavia, a região não apresentou variações discrepantes das demais, o que pode ser um indicativo da boa qualidade da imputação. Em concordância com o apontado, os valores de Var e de SD apresentaram reduções após a IM, bem como houve uma aproximação entre a média e a mediana, as quais apresentaram um aumento dos seus valores, fator que indica que a imputação gerou estimativas com valores acima da média e da mediana.

Todas as regiões mantiveram os valores de máximos sem muita variação, bem como Q3, com exceção do NB. As estimativas promoveram uma diminuição da VAR e do SD em todas as estações, como também apresentou um aumento no valor de Q1, apontando que os valores imputados encontram-se mais próximos à média e à mediana, as quais apresentaram aumento após a IM, com exceção SEB, que manteve a mesma média.

Figura 5 - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) separados por região geográfica do Brasil.



Fonte: da autora

Tabela 3 - Valores das medidas de tendências centrais, e medidas de dispersão dos bancos de dados incompletos e imputados, regionalmente.

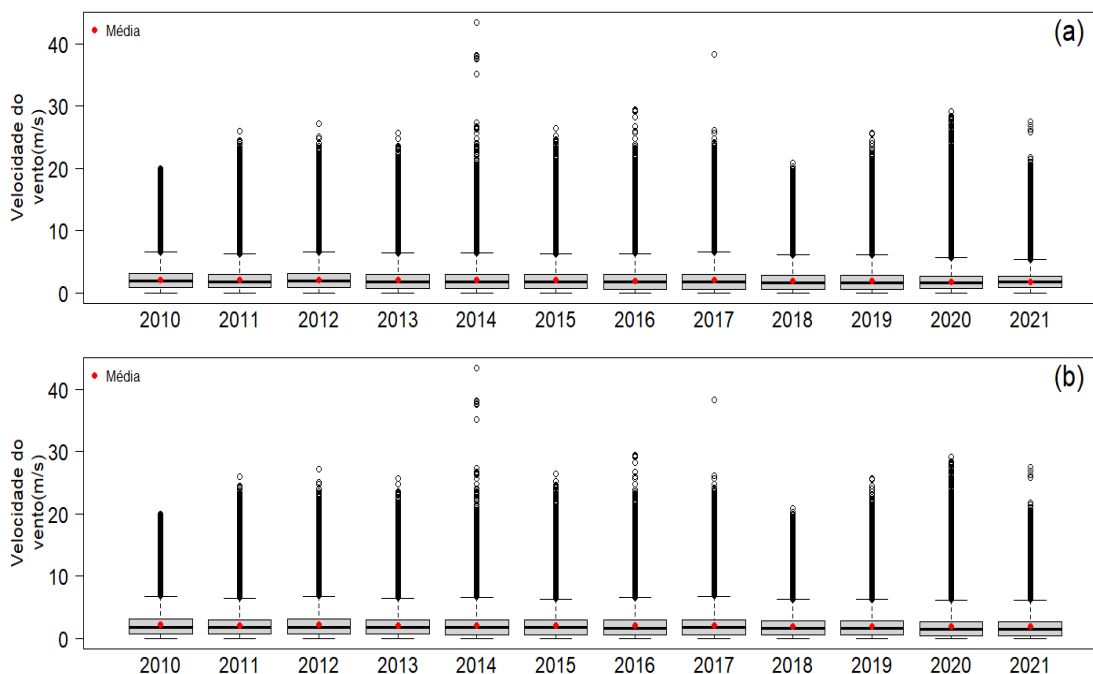
		MÍN	Q1	Q2	MÉD	Q3	MAX	SD	VAR	NA
INC.	COB	0,00	0,20	1,30	1,60	2,40	20,00	1,45	2,10	1201721
INC.	NEB	0,00	0,90	2,10	2,30	3,40	19,20	1,70	2,90	1750674
INC.	NB	0,00	0,10	0,90	1,20	1,90	18,80	1,20	1,43	1017531
INC.	SEB	0,00	0,60	1,60	1,90	2,70	20,00	1,60	2,55	820461
INC.	SB	0,00	0,80	2,10	2,50	3,60	43,00	2,18	4,74	786383
IMP.	COB	0,00	0,40	1,40	1,58	2,40	20,00	1,37	1,87	-
IMP.	NEB	0,00	1,10	2,20	2,34	3,40	19,20	1,63	2,67	-
IMP.	NB	0,00	0,30	1,02	1,22	1,80	18,80	1,11	1,24	-
IMP.	SEB	0,00	0,70	1,70	1,90	2,70	20,00	1,56	2,45	-
IMP.	SB	0,00	1,00	2,20	2,55	3,60	43,40	2,14	4,59	-

Fonte: da autora

### 5.3 Variação Interanual

Nas Figuras 6a e 6b são apresentados os *boxplots* referentes aos dados incompletos e imputados, respectivamente, com separação anual. Graficamente não são apresentadas mudanças intensas entre as séries. Porém, é notável os valores das médias mais elevados em parte dos anos, como 2010, 2016, 2020 e 2021. Isso pode ser sugestivo da maior ausência de dados ter ocorrido nesses anos, uma vez que foram os que sofreram maior mudança nos valores da média. Porém, esse fato também pode estar associado à presença dos maiores valores e frequências de *outliers*, como em 2014, apresentando o valor máximo de ambas as séries. É possível apontar que os anos de 2011 e 2021 apresentaram um maior distanciamento entre Q3 e a haste superior (de modo que o valor máximo da haste aumentou) após a IM, indicando uma maior variabilidade dos dados com a inserção das estimativas geradas.

Figura 6 - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) separados por ano.



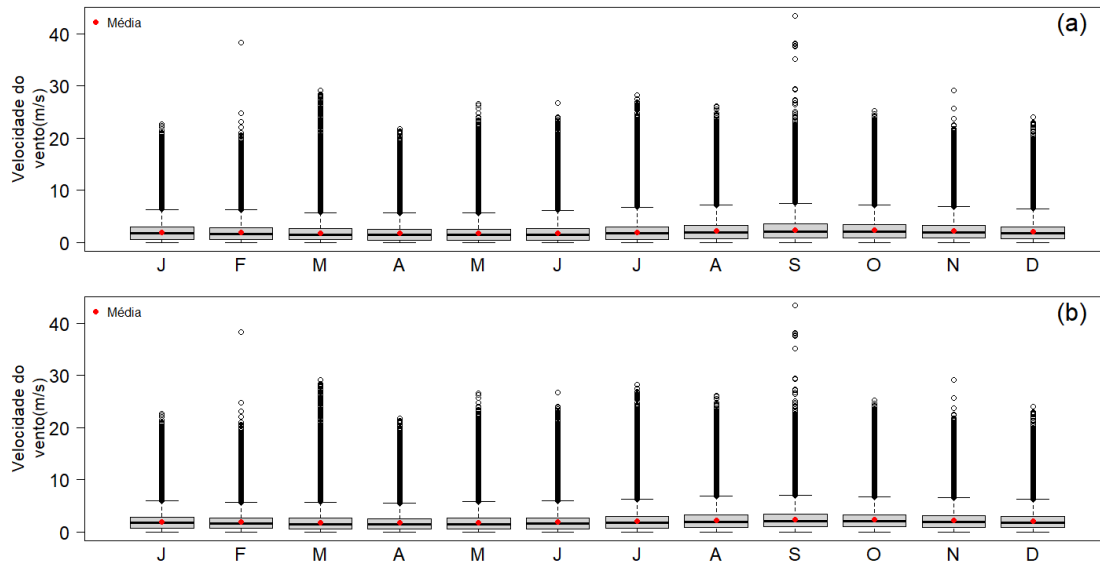
Fonte: da autora

#### 5.4 Variação Mensal

Na estatística exploratória via *boxplot* mensal, apresentada na Figura 7a e 7b, correspondentes a dados incompletos e dados imputados, respectivamente, é notável que após a IM não ocorreram muitas mudanças visuais na estrutura das caixas e no comportamento da média, mediana, SD, VAR, e quartis. Isso demonstra que a imputação não apresentou impactos significativos no comportamento da variável. Tal característica poderia ser mais explorada na ausência ou diminuição dos valores atípicos. Os maiores efeitos da inserção das estimativas encontram-se em um sucinto aumento da distância entre haste inferior e Q1 em grande parte dos meses, demonstrando o aumento da quantidade de valores mais próximos à mediana e média, ou seja, que as estimativas geradas se aproximam da média, e não se concentram nos menores valores de velocidade do vento e com uma menor variabilidade dos dados próximos à mediana.

O mês de Setembro apresentou a maior variação dos dados antes e após a IM, apesar de ter uma diminuição da estrutura da sua caixa após a inclusão dos valores imputados, além de ser o mês possuinte do maior valor observado, identificado como ocorrido no SB na seção anterior.

Figura 7 - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) separados por mês.



Fonte: da autora

## 5.5 Ciclo diurno

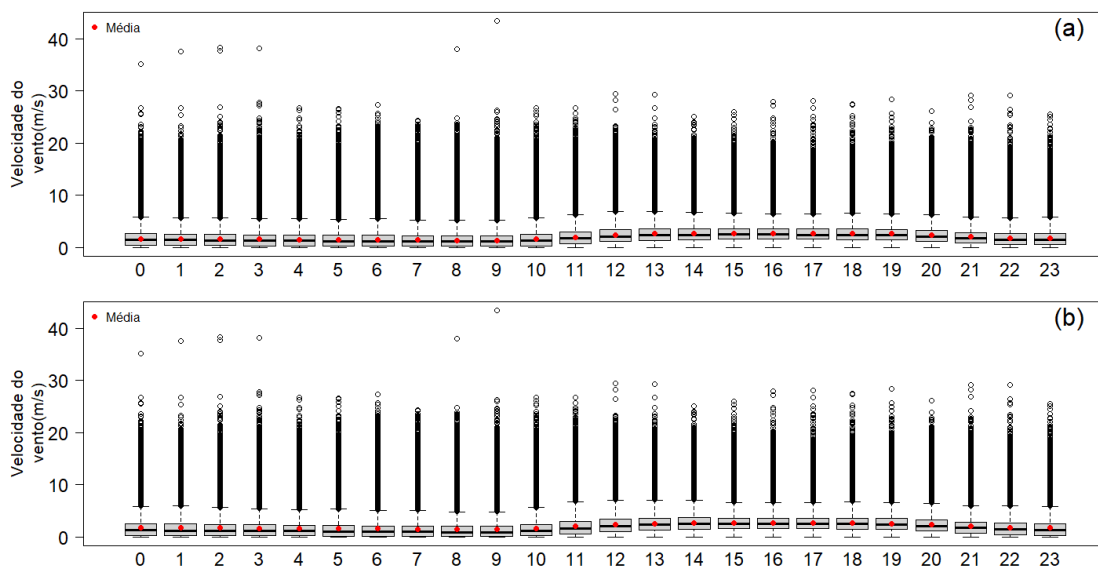
Nas Figuras 8a e 8b são apresentados os *boxplots* dados incompletos e imputados, nessa ordem, na partição horária. É notável um afastamento entre a mediana e a média durante as primeiras horas, de modo a apresentar um aumento no valor da média, a qual sofre influência de *outliers*, os quais são apontados com os maiores valores nesse intervalo, bem como às 08:00h e 09:00h, as quais apresentam as mesmas características após a IM. Os demais horários não destacam muitas variações quanto a essas medidas. É perceptível também que há uma diminuição entre a haste inferior e o Q1 durante as primeiras horas, e um alongamento dessa distância entre 12:00h e 20:00h, ainda que não muito perceptível.

Todos os horários apresentaram um grande intervalo entre a haste superior e o valor de Q3, demonstrando ainda mais variabilidade durante as maiores velocidades.

Com essa divisão horária é possível distinguir os horários com maior variação da velocidade do vento, através da estrutura do gráfico. Assim, é apontado que o

intervalo entre 11:00h e 20:00h apresentam as maiores variações e uma frequência maior da ocorrência das maiores velocidades do vento, quando desconsiderados os dados atípicos. Porém, foi às 09:00h a ocorrência da maior observação do banco de dados coletado.

Figura 8 - *Boxplot* para a velocidade do vento dos dados incompletos (a) e imputados (b) por hora.



Fonte: da autora

## 6 CONCLUSÕES E SUGESTÕES DE TRABALHOS FUTUROS

Este trabalho apresentou o preenchimento de um banco de dados de velocidade do vento com amostragem horária para todo o território continental do Brasil, para isso foi aplicado o método de Imputação Múltipla. O objetivo foi comparar, através de estatística quantitativa, as bases de dados com falhas e com os valores imputados e caracterizar a variação diurna, mensal e anual dos dados originais e preenchidos. Tal elaboração torna-se importante devido à incipiência de pesquisas referentes à dinâmica dos ventos com alta frequência temporal como a horária, apesar da grande importância.

Isto posto, foi observado após a realização da IM que a série temporal original, após a aplicação dos limites pré-estabelecidos, apresentou uma grande

quantidade de *outliers*. A ocorrência desse fato afeta a análise quantitativa, uma vez que esses dados atípicos têm grande influência sobre medidas como a média e o desvio padrão. Apesar disso, em aspectos gerais, foi notado que o intervalo interquartilício tornou-se maior após a imputação, indicativo de aumento da variabilidade dos dados em torno da mediana.

Foram imputados um total de 5.560.709 dados, equivalente a, aproximadamente, 14% dos dados da série completa, e após o preenchimento dessas lacunas, os valores do SD e VAR diminuíram, indicando uma diminuição da variabilidade dos dados. Foi notado que o maior valor observado da velocidade do vento ocorreu na região Sul, durante o mês de setembro. O SB também é considerado a região com a maior variabilidade, enquanto que o NB a menor. Alguns anos apresentaram o aumento do valor da média, distanciando-a da mediana, o que pode ser um indicativo da influência dos *outliers* sobre a dinâmica da imputação.

O mês de Setembro apresentou a maior variação dos dados antes e após a IM. Foi observado, com a inserção das estimativas, um sucinto aumento da distância entre haste inferior e Q1 em grande parte dos meses. Na análise horária, foram distinguidos os horários com maior variação da velocidade do vento: entre 11:00h e 20:00h. Porém, o *outlier* de maior valor ocorreu às 09:00h. Por fim, as diferenças entre as estatísticas descritivas da série incompleta e da após a imputação não foram tão destoantes, fato que implica que os dados imputados foram gerados como valores próximos à mediana.

Espera-se que este trabalho sirva como subsídio para diversos outros estudos do âmbito das ciências climáticas e demais, e para futuros trabalhos é sugerido a diminuição dos valores limites pré-estabelecidos, a fim de diminuir a quantidade de *outliers*, que se mostraram como grandes atuantes na dinâmica da IM. O aumento da série temporal também é sugestão para a prática da análise da variabilidade das características espaciais e temporais da velocidade do vento.

## Referências

AB'SABER, Aziz Nacib. O domínio morfoclimático semi-árido das caatingas brasileiras. **Geomorfologia**, n. 43, p. 1-39, 1974.

AGNOLIN, Eduarda Regina; MURARA, Pedro. OS PRINCIPAIS SISTEMAS ATMOSFÉRICOS ATUANTES SOBRE O NORTE DO RIO GRANDE DO SUL.

ALHO, C. J. R.; GONÇALVES, H. C. Biodiversidade do Pantanal: Ecologia & Conservação, 135p. **Campo Grande, Brazil: Editora Uniderp, 2005.**

ALONSO, Claudio D.; GODINHO, Roberto. A evolução da qualidade do ar em Cubatão. **Química Nova**, v. 15, n. 2, p. 126-136, 1992.

AMARAL, Marcus Vinicius Silva Gurgel do. **Ajuste de modelos e comparação de séries temporais para dados de vazão específica em microbacias pareadas.** 2014. Tese de Doutorado. Universidade de São Paulo.

AMBRIZZI, Tércio. El nino/oscilações sul e teleconexões atmosféricas no hemisfério austral. 2003. Tese (Livre Docência) – Universidade de São Paulo, São Paulo, 2003. Acesso em: 07 de março de 2023.

ANJOS, Max. Orientações climáticas para o planejamento urbano numa cidade costeira do nordeste do Brasil: Aracaju-SE. 2017.

**Área territorial brasileira 2020. Rio de Janeiro: IBGE, 2021.**

AZEVEDO, PV de; SILVA, BB da; RODRIGUES, M. F. G. Previsão estatística das chuvas de outono no Estado do Ceará. **Revista Brasileira de Meteorologia**, v. 13, n. 1, p. 19-30, 1998.

BAENA, Luiz Gustavo Nascentes *et al.* Aplicação de modelo para geração de séries sintéticas de dados climáticos: Parte II-velocidade do vento e umidade relativa. **Rev Bras Agrometeorologia**, v. 13, n. 2, p. 451-457, 2005.

BRONZATTI, Fabricio Luiz; IAROZINSKI NETO, Alfredo. Matrizes energéticas no Brasil: cenário 2010-2030. **Encontro Nacional de Engenharia de Produção**, v. 28, p. 13-16, 2008.

CAMARGOS, Vitor Passos et al. Imputação múltipla e análise de casos completos em modelos de regressão logística: uma avaliação prática do impacto das perdas em covariáveis. **Cadernos de Saúde Pública**, v. 27, n. 12, p. 2299-2313, 2011.

CASTELHANO, Francisco. ANÁLISE DA DINÂMICA DOS VENTOS EM ARACAJU/SERGIPE. **Revista OKARA: Geografia em debate**, v. 16, n. 1, p. 125-135, 2022.

CARAMORI, Paulo Henrique et al. Efeitos do vento sobre mudas de cafeeiro Mundo Novo e Catuaí Vermelho. **Pesquisa Agropecuária Brasileira**, v. 21, n. 11, p. 1113-1118, 1986.

COLANTONIO, Alessandro et al. ABBA: Adaptive bicluster-based approach to impute missing values in binary matrices. In: **Proceedings of the 2010 ACM Symposium on Applied Computing**. 2010. p. 1026-1033.

COSTA, Gabriel Brito et al. Análise espacial e temporal do vento no estado de Alagoas. 2009.

CRESESB. Energia eólica: estado atual e perspectivas. Disponível em: <[http://www.agg.ufba.br/cier\\_eolica.pdf](http://www.agg.ufba.br/cier_eolica.pdf)> Acesso em: 07 de março de 2023.

CURRY, Judith A.; WEBSTER, Peter J. **Thermodynamics of atmospheres and oceans**. Elsevier, 1998.

DANTAS, Leydson Galvêncio et al. Aplicação de modelos para séries temporais e pluviométricas no Estado da Paraíba. 2016.

DE MORAIS, Giuliana Mairana *et al.* Caracterização da velocidade e direção do vento em Mossoró/RN. **Revista Brasileira de Geografia Física**, v. 7, n. 04, p. 746-754, 2014.

DE OLIVEIRA, Marcelo Tuler; SARAIVA, Sérgio Luiz Costa. **Fundamentos de Geodésia e Cartografia: Série Tekne**. Bookman Editora, 2015.

DE SOUZA, MARCOS FERNANDO FERREIRA; DE EFICIÊNCIA, PROPOSTA DE METODOLOGIA; DE SERVIDORES, ATUANDO NA MATRIZ. INSTITUTO DE TECNOLOGIA PARA O DESENVOLVIMENTO, 2014.

FARHANGFAR, Alireza; KURGAN, Lukasz A.; PEDRYCZ, Witold. Experimental analysis of methods for imputation of missing values in databases. In: **Intelligent Computing: Theory and Applications II**. SPIE, 2004. p. 172-182.

FARHANGFAR, Alireza; KURGAN, Lukasz A.; PEDRYCZ, Witold. A novel framework for imputation of missing values in databases. **IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans**, v. 37, n. 5, p. 692-709, 2007.

FERREIRA, Antonio Geraldo; DA SILVA MELLO, Namir Giovanni. Principais sistemas atmosféricos atuantes sobre a região Nordeste do Brasil e a influência dos oceanos Pacífico e Atlântico no clima da região. **Revista brasileira de climatologia**, v. 1, n. 1, 2005.

FERREIRA, Glauber Willian de Souza; REBOITA, Michelle Simões; DA ROCHA, Rosmeri Porfírio. Vórtices ciclônicos de altos níveis nas cercanias do nordeste do Brasil: Climatologia e análise da vorticidade potencial isentrópica. **Anuário do Instituto de Geociências**, v. 42, n. 3, p. 568-585, 2020.

FERREIRA, Moniki Dara de Melo. **Estudo da velocidade do vento através de downscaling dinâmico em alta resolução sobre terreno complexo no Nordeste do Brasil**. 2020. Dissertação de Mestrado. Universidade Federal do Rio Grande do Norte.

FIRAT, Mahmut et al. Missing data analysis and homogeneity test for Turkish precipitation series. **Sadhana**, v. 35, p. 707-720, 2010.

FRITZSONS, Elenice; MANTOVANI, Luiz Eduardo; DE AGUIAR, Ananda Virgínia. Relação entre altitude e temperatura: uma contribuição ao zoneamento climático no estado do Paraná. **Revista de estudos ambientais**, v. 10, n. 1, p. 49-64, 2008.

GILLILAND, Joshua M.; KEIM, Barry D. Position of the South Atlantic Anticyclone and its impact on surface conditions across Brazil. **Journal of Applied Meteorology and Climatology**, v. 57, n. 3, p. 535-553, 2018a.

GILLILAND, Joshua M.; KEIM, Barry D. Surface wind speed: trend and climatology of Brazil from 1980–2014. **International Journal of Climatology**, v. 38, n. 2, p. 1060-1073, 2018b.

GOBO, J.P.A., ALVES, R.R., SILVEIRA, T.S., ONÇA, D.S., MONTEIRO, L.M., WOLLMANN, C.A., GALVANI, E. A Influência Do Vento Regional Na Sensação Térmica De Pedestres Em Espaços Urbanos Abertos: Estudo De Caso Do Vento Norte Em Santa Maria-RS, **Revista Ra'ega**, v. 40, p. 110-129, 2017.

GROTZINGER, John; JORDAN, Tom. **Para Entender a Terra-6**. Bookman Editora, 2013.

GUIMARÃES, Sullyandro Oliveira et al. Projeções de Mudanças Climáticas sobre o Nordeste Brasileiro dos Modelos do CMIP5 e do CORDEX. **Revista Brasileira de Meteorologia**, v. 31, p. 337-365, 2016.

HAREL, Ofer; ZHOU, Xiao-Hua. Multiple imputation: review of theory, implementation and software. **Statistics in medicine**, v. 26, n. 16, p. 3057-3077, 2007.

HONAKER, James; KING, Gary; BLACKWELL, Matthew. Amelia II: A program for missing data. **Journal of statistical software**, v. 45, p. 1-47, 2011.

KAYANO, Mary Toshio; ANDREOLI, Rita Valéria. Clima da região Nordeste do Brasil. **Tempo e clima no Brasil**, v. 1, p. 213-233, 2009.

KENWARD, Michael G.; CARPENTER, James. Multiple imputation: current perspectives. **Statistical methods in medical research**, v. 16, n. 3, p. 199-218, 2007.

LIMA, Isael Defferson Morais da Costa et al. Elaboração e montagem de um protótipo de baixo custo para o bombeamento de água utilizando energia eólica. 2018.

LIMA, Mírian de Paula et al. Circulação atmosférica associada aos extremos secos e chuvosos em Uberlândia-MG. 2019.

LITTLE, R. J.; RUBIN, D. B. Statistical inference with missing data. 2002.

LOPES, Iug et al. Potencial de geração de energia eólica no Submédio São Francisco. **Journal of Environmental Analysis and Progress**, p. 330-340, 2017.v

MARENGO, J. A. Mudanças Climáticas Globais e Regionais: Avaliação do Clima Atual do Brasil e Projeções de Cenários Climáticos do Futuro. **Revista Brasileira de Meteorologia**. v. 16, n. 1, p. 01-18, 2001.

MARENGO, J. A.; VALVERDE, M. C. Caracterização do clima no Século XX e Cenário de Mudanças de clima para o Brasil no Século XXI usando os modelos do IPCC-AR4. **Revista Multiciência**, v. 8, p. 5–28, 2007.

MARENGO, José A. et al. Variabilidade e mudanças climáticas no semiárido brasileiro. **Recursos hídricos em regiões áridas e semiáridas**, v. 1, p. 385-422, 2011.

MARIN, Fábio R.; ASSAD, Eduardo D.; PILAU, F. A. Clima e ambiente: introdução à climatologia para ciências ambientais. 2008.

MCKNIGHT, Patrick E. et al. **Missing data: A gentle introduction**. Guilford Press, 2007.

MENDONÇA, Francisco; DANNI-OLIVEIRA, Inês Moresco. **Climatologia: noções básicas e climas do Brasil**. Oficina de textos, 2017.

MILANESI, Marcos Alexandre. **Avaliação do efeito orográfico na pluviometria de vertentes opostas da Ilha de São Sebastião (Ilhabela-SP)**. 2007. Tese de Doutorado. Universidade de São Paulo.

MILANESI, M. A.; GALVANI, E. A Chuva Orográfica no Parque Estadual de Ilhabela (PEIB-SP)-Estrada De Castelhanos. **Anais do XIII Simpósio Brasileiro de Climatologia Geográfica. Viçosa**, 2009.

MOLENBERGHS, Geert et al. Analyzing incomplete longitudinal clinical trial data. **Biostatistics**, v. 5, n. 3, p. 445-464, 2004.

MUNHOZ, Fabrício César; GARCIA, Anice. Caracterização da velocidade e direção predominante dos ventos para a localidade de Ituverava-SP. **Revista Brasileira de Meteorologia**, v. 23, p. 30-34, 2008.

NETO, José Valladares et al. Boxplot: um recurso gráfico para a análise e interpretação de dados quantitativos. **Revista Odontológica do Brasil Central**, v. 26, n. 76, 2017.

NUNES, Luciana Neves. Métodos de imputação de dados aplicados na área da saúde. 2007.

PE. 2018. Portal energia – energia renováveis. Disponível em: <https://www.portal-energia.com/vantagens-desvantagens-da-energia-eolica/>. Acesso em: 07 de março de 2023.

POTT, A. et al. Diversidade de plantas do Pantanal. **Brazilian Journal of Biology**, v. 71, p. 265-273, 2011.

PRZYBYSZEWSKI, Juliana. **A competitividade turística entre as regiões brasileiras**. 2017. Tese de Doutorado. Instituto Politecnico de Braganca (Portugal).

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2015.

REIS, A.C.. Clima da caatinga. **Anais da Academia Brasileira de Ciências**, Rio de Janeiro, v. 48, n. 325-335, 1976.

ROCKETT, Gabriela C. *et al.* Análise espaço-temporal dos ventos no extremo norte da Planície Costeira do Rio Grande do Sul, Brasil. **Pesquisas em Geociências**, v. 44, n. 2, p. 203-219, 2017.

RODRIGUES, Daniele Tôres. Análise de eventos extremos de precipitação no Nordeste do Brasil. 2019.

RUBIN, Donald B. Inference and missing data. **Biometrika**, v. 63, n. 3, p. 581-592, 1976.

RUBIN, Donald B. Multiple imputation after 18+ years. **Journal of the American statistical Association**, v. 91, n. 434, p. 473-489, 1996.

RUBIN, Donald B. Multiple imputation for survey nonresponse. 1987.

SANTOS, Alexandre Torres et al. Assessment of wind resources in two parts of Northeast Brazil with the use of numerical models. **Meteorological Applications**, v. 23, n. 4, p. 563-573, 2016.

SANTOS, Alexandre Torres; SANTOS E SILVA, Cláudio Moisés. Seasonality, interannual variability, and linear tendency of wind speeds in the Northeast Brazil from 1986 to 2011. **The Scientific World Journal**, v. 2013, 2013.

SANTOS, Tailine Corrêa; CARVALHO, Vanessa Silveira Barreto; REBOITA, Michelle Simões. Avaliação da influência das condições meteorológicas em dias com altas

concentrações de material particulado na Região Metropolitana do Rio de Janeiro. **Engenharia Sanitária e Ambiental**, v. 21, p. 307-313, 2016.

SARTORI, Maria da Graça Barros. A dinâmica do clima do Rio Grande do Sul: indução empírica e conhecimento científico. **Terra livre**, n. 20, p. 27-50, 2003.

SCHAFER, Joseph L. Multiple imputation: a primer. **Statistical methods in medical research**, v. 8, n. 1, p. 3-15, 1999.

SCHAFER, Joseph L.; GRAHAM, John W. Missing data: our view of the state of the art. **Psychological methods**, v. 7, n. 2, p. 147, 2002.

SENTELHAS, Paulo Cesar; ANGELOCCI, Luiz Roberto. **Radiação Solar Balanço de Energia: Meteorologia Agrícola**. ESALQ/USP. 2012. Apresentação de Powerpoint. 47 slides. color. Disponível em: [http://www.leb.esalq.usp.br/leb/aulas/lce306/Aula5\\_2012.pdf](http://www.leb.esalq.usp.br/leb/aulas/lce306/Aula5_2012.pdf). Acesso em: 12 de março de 2023.

SILVA, Isamara de Mendonça. **Propriedades termodinâmicas da atmosfera e estudo da entropia próximo à superfície nas capitais do Nordeste do Brasil**. 2018. Trabalho de Conclusão de Curso. Universidade Federal do Rio Grande do Norte.

SOBRAL, Fábio Stefano Batista et al. Avaliação do potencial eólico para geração de energia na zona rural do Estado de Sergipe. 2009.

TEIXEIRA, R.A.M.S. Projeto de protótipo de turbina eólica de eixo horizontal. 2016. Tese de Doutorado. Universidade Federal do Rio de Janeiro.

TUCHTENHAGEN, Patrícia et al. WRF model assessment for wind intensity and power density simulation in the southern coast of Brazil. **Energy**, v. 190, p. 116341, 2020.

VAN DER HEIJDEN, Geert JMG et al. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. **Journal of clinical epidemiology**, v. 59, n. 10, p. 1102-1109, 2006.

VERONEZE, Rosana. **Tratamento de dados faltantes empregando biclusterização com imputação múltipla**. 2011. Tese de Doutorado. UNIVERSIDADE ESTADUAL DE CAMPINAS.

VIANELLO, R. L.; ALVES. A. R. Climatologia. **Meteorologia básica e aplicada**. Viçosa: UFV, p. 377-444, 2006.

VIEIRA, Luciano; PICULLI, Francisco José. Meteorologia e climatologia agrícola. **Cidade Gaúcha–Pr: Universidade Estadual de Maringá**, 2009.

WESONGA, Ronald. On multivariate imputation and forecasting of decadal wind speed missing data. **SpringerPlus**, v. 4, n. 1, p. 1-8, 2015.

WHITE, Ian R.; WOOD, Angela; ROYSTON, Patrick. Multiple imputation in practice. **Statistical Methods in Medical Research**, v. 16, n. 3, p. 195-198, 2007.

XAVIER, Alexandre C. et al. New improved Brazilian daily weather gridded data (1961–2020). **International Journal of Climatology**, 2022.

XAVIER, Alexandre C.; KING, Carey W.; SCANLON, Bridget R. Daily gridded meteorological variables in Brazil (1980–2013). **International Journal of Climatology**, v. 36, n. 6, p. 2644-2659, 2016.

ZAHARIM, Azami et al. Fitting of statistical distributions to wind speed data in Malaysia. **European Journal of Scientific Research**, v. 26, n. 1, p. 6-12, 2009.

ZHANG, Shichao; ZHANG, Chengqi; YANG, Qiang. Data preparation for data mining. **Applied artificial intelligence**, v. 17, n. 5-6, p. 375-381, 2003.