

Universidade Federal Fluminense

CAROLINE DE OLIVEIRA COSTA SOUZA ROSA

Estudo de complementaridade entre as energias
hidrelétrica, eólica e fotovoltaica nas regiões Sudeste
e Centro-Oeste

VOLTA REDONDA

2019

CAROLINE DE OLIVEIRA COSTA SOUZA ROSA

**Estudo de complementaridade entre as energias
hidrelétrica, eólica e fotovoltaica nas regiões Sudeste
e Centro-Oeste**

Dissertação apresentada ao Programa de Pós-graduação em Modelagem Computacional em Ciência e Tecnologia da Universidade Federal Fluminense, como requisito parcial para obtenção do título de Mestre em Modelagem Computacional em Ciência e Tecnologia. Área de Concentração: Modelagem Computacional.

Orientador:

Profa. Dra. Eliane da Silva Christo

Coorientador:

Profa. Dra. Kelly Alonso Costa

UNIVERSIDADE FEDERAL FLUMINENSE

VOLTA REDONDA

2019

Ficha catalográfica automática - SDC/BEM
Gerada com informações fornecidas pelo autor

R788e Rosa, Caroline de Oliveira Costa Souza
Estudo de complementaridade entre as energias hidrelétrica,
eólica e fotovoltaica nas regiões Sudeste e Centro-Oeste /
Caroline de Oliveira Costa Souza Rosa ; Eliane da Silva
Christo, orientadora ; Kelly Alonso Costa, coorientadora.
Volta Redonda, 2019.
156 f. : il.

Dissertação (mestrado)-Universidade Federal Fluminense,
Volta Redonda, 2019.

DOI: <http://dx.doi.org/10.22409/PPG-MCCT.2019.m.13332433750>

1. Energia renovável. 2. Complementaridade. 3.
Otimização. 4. Produção intelectual. I. Christo, Eliane da
Silva, orientadora. II. Costa, Kelly Alonso, coorientadora.
III. Universidade Federal Fluminense. Escola de Engenharia
Industrial e Metalúrgica de Volta Redonda. IV. Título.

CDD -

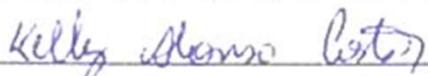
Estudo de complementaridade entre as energias hidrelétrica, eólica e
fotovoltaica nas regiões Sudeste e Centro-Oeste

Caroline de Oliveira Costa Souza Rosa

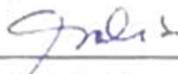
Aprovada por:



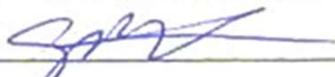
Profa. Eliane da Silva Christo, D.Sc. / UFF



Profa. Kelly Alonso Costa, D.Sc. / UFF



Profa. Cecilia Toledo Hernández, D.Sc. / UFF



Prof. Gustavo Benitez Alvarez, D.Sc. / UFF



Prof. Wesley Luiz da Silva Assis, D.Sc. / UFF



Prof. Eduardo Pestana de Aguiar, D.Sc. / UFJF

Volta Redonda, 25 de fevereiro de 2019.

Para meus pais, Véra e Cilas

Agradecimentos

Agradeço primeiramente a Deus pela vida, por Sua constante direção e companhia, e pela oportunidade de cursar e concluir este curso de mestrado.

Aos meus pais, Cilas e Véra, pelo amor, pelo suporte incessante, pelo amparo nos momentos difíceis e por partilharem as minhas alegrias.

Às minhas orientadoras, Kelly e Eliane, que me apresentaram à área acadêmica e que têm me acompanhado e aconselhado desde a graduação.

À Universidade Federal Fluminense e toda a sua equipe de funcionários, pela disponibilidade do curso e dos recursos para o nosso aprendizado.

Aos meus amigos, pelo incentivo e pelos momentos de descontração. Em especial, à Luiza, ao Caio e ao Douglas, que além disso, construíram diretamente na elaboração deste trabalho.

Aos meus avós, tias e tios e toda a família, pelo carinho em cada etapa percorrida.

A cada professor, que durante toda a minha vida, contribuiu para a minha formação.

“Knowledge comes by taking things apart: analysis.

But wisdom comes by putting things together.”

John Alexander Morrison

Resumo

A inserção de fontes renováveis intermitentes nos sistemas elétricos tradicionais deve ser planejada, já que a variabilidade dos recursos naturais pode afetar a confiabilidade e o custo da eletricidade gerada. Este trabalho avalia a complementaridade dos recursos eólico, fotovoltaico e hídrico nas regiões Sudeste e Centro-Oeste e identifica de que forma devem ser combinados, de modo que a característica intermitente das fontes seja suavizada. Para isso, foram coletados dados de radiação solar global, velocidade do vento e vazões fluviais de estações localizadas nessas regiões. Buscou-se avaliar as séries de cada uma dessas fontes através da Análise Fatorial para compreender a similaridade de um mesmo recurso em localidades diferentes e agrupar as séries com comportamento semelhante. Pôde-se agrupar satisfatoriamente as séries de vazões. O recurso solar, por sua distribuição uniforme nas regiões, não respondeu bem a grupos com mais de dois fatores e as estações foram agrupadas em um grande grupo. Por outro lado, o recurso eólico possui variabilidade geográfica alta e não pôde ser agrupado, de modo que as estações permaneceram isoladas. Para avaliar a complementaridade entre os grupos, foram utilizados o coeficiente de correlação linear de Pearson e a análise gráfica. Constatou-se que fontes diferentes têm, em geral, correlações baixas, o que favorece a complementaridade delas. O *mix* ótimo de fontes intermitentes foi obtido através da construção e resolução de dois modelos de otimização, ambos classificados como modelos de programação quadrática. Apesar de terem objetivos diferentes, os dois modelos apontaram que o *mix* renovável intermitente ideal deve ter aproximadamente 50% de participação fotovoltaica. Estes *mixes* ótimo poderão guiar a expansão da parcela intermitente da geração elétrica nas regiões consideradas.

Abstract

The use of intermittent renewable energy sources by traditional energy systems must be well planned because the variability of natural resources can affect the reliability and the costs of the generated electricity. This work studies the complementarity between wind, photovoltaic and small hydro generation in Southeastern and Central-Western Brazil, and identifies how these sources should be combined so that their intermittency is smoothed. In order to achieve this, global solar radiation, wind speed and river inflow data measured at the region was used. These data series were assessed using Factor Analysis, with the aim of understanding how similar the behaviour of a resource is when measured in different locations. The analysis has also been used to group the variables, according to their similarities. The river inflow series were grouped successfully. Solar radiation is evenly distributed in the region, so models with two or more factors did not suit it. Therefore, all stations were grouped together. On the other hand, wind speeds vary significantly in the region, and the series could not be grouped and remained isolated. The complementarity between the groups was assessed using the Pearson correlation coefficient and graphic analysis. In general, different sources had low correlation between each other, what increases their complementarity potential. The optimal intermittent renewable mix was achieved by developing and solving two optimization models. Despite having different objectives, both models suggested that the photovoltaic share in ideal renewable intermittent mix should be around 50%. These optimal mixes can guide the expansion of the intermittent share of electricity generation in the region.

Palavras-chave

1. Integração de fontes renováveis
2. Complementaridade
3. Otimização
4. Energia eólica
5. Energia fotovoltaica
6. Pequenas hidrelétricas

Sumário

| | |
|---|------------|
| Lista de Figuras | xii |
| Lista de Tabelas | xvi |
| 1 Introdução | 17 |
| 1.1 Objetivos | 19 |
| 1.1.1 Objetivo Geral | 19 |
| 1.1.2 Objetivos Específicos | 19 |
| 1.2 Justificativa e Relevância | 20 |
| 1.3 Estrutura do Trabalho | 20 |
| 1.4 Limitações | 21 |
| 2 Fundamentação Teórica | 22 |
| 2.1 Panorama Atual da Eletricidade | 22 |
| 2.1.1 A eletricidade no mundo e no Brasil | 22 |
| 2.1.2 A eletricidade nas regiões Sudeste e Centro-Oeste | 25 |
| 2.2 Energias Renováveis | 30 |
| 2.2.1 Energia hidráulica | 30 |
| 2.2.2 Energia eólica | 34 |
| 2.2.3 Energia solar | 40 |
| 2.3 Inserção de Fontes Renováveis no Sistema Elétrico | 44 |
| 2.4 Correlação Linear | 46 |
| 2.5 Análise de Fatores | 51 |

| | | |
|----------|---|-----------|
| 2.5.1 | Pré-análise | 52 |
| 2.5.2 | Formulação de uma Análise de Fatores | 56 |
| 2.5.3 | Extração de Fatores | 58 |
| 2.5.4 | Escolha do número de fatores | 61 |
| 2.5.5 | Avaliação das cargas | 62 |
| 2.5.6 | Rotação de fatores | 63 |
| 2.6 | Otimização | 65 |
| 3 | Trabalhos Relacionados | 67 |
| 3.1 | Contextualização | 67 |
| 3.2 | Estudos de Complementaridade | 69 |
| 3.3 | Estudos de Complementaridade com Otimização | 72 |
| 4 | Metodologia | 74 |
| 4.1 | Aquisição e mineração de dados | 75 |
| 4.1.1 | Base de dados | 75 |
| 4.1.2 | Avaliação de Dados | 79 |
| 4.2 | Cálculo das Séries de Potência | 88 |
| 4.2.1 | Geração Fotovoltaica | 88 |
| 4.2.2 | Geração Eólica | 89 |
| 4.2.3 | Geração Hidrelétrica | 91 |
| 4.3 | Avaliação de Correlações Lineares | 91 |
| 4.4 | Análise de Fatores | 92 |
| 4.4.1 | Adequação das Séries | 92 |
| 4.4.2 | Análise de Fatores | 94 |
| 4.5 | Avaliação de complementaridade | 96 |
| 4.6 | Otimização | 98 |

| | | |
|----------|--|------------|
| 4.6.1 | Formulação I | 99 |
| 4.6.2 | Formulação II | 101 |
| 4.6.3 | Modelagem | 101 |
| 5 | Resultados | 104 |
| 5.1 | Correlações Lineares | 104 |
| 5.2 | Pré-análise | 109 |
| 5.3 | Análise de Fatores | 110 |
| 5.3.1 | Vazões da ANA | 110 |
| 5.3.2 | Vazões do ONS | 112 |
| 5.3.3 | Dados de Radiação Solar | 113 |
| 5.3.4 | Dados de Velocidade do Vento | 114 |
| 5.4 | Avaliação de Complementaridade | 116 |
| 5.4.1 | Complementaridade Horária | 116 |
| 5.4.2 | Complementaridade Diária | 120 |
| 5.5 | Otimização | 125 |
| 5.5.1 | Formulação I | 125 |
| 5.5.2 | Formulação II | 126 |
| 5.5.3 | Comparação dos Resultados | 128 |
| 6 | Conclusões e Trabalhos Futuros | 131 |
| 6.1 | Conclusões | 131 |
| 6.2 | Trabalhos Futuros | 132 |
| | Anexo A | 134 |
| | Apêndice A | 136 |
| | Apêndice B | 141 |

| | |
|--------------------|------------|
| Sumário | xi |
| Apêndice C | 142 |
| Apêndice D | 144 |
| Apêndice E | 145 |
| Referências | 148 |

Lista de Figuras

| | | |
|------|--|----|
| 2.1 | Evolução da capacidade instalada de geração de eletricidade no Brasil. | 24 |
| 2.2 | Participação de cada tipo de usina na potência fiscalizada total dos estados das regiões SE e CO. No centro de cada gráfico se encontra a capacidade instalada naquele estado. | 25 |
| 2.3 | Combustíveis usados em usinas termelétricas à biomassa nas regiões Sudeste e Centro-Oeste. | 26 |
| 2.4 | Mini e micro geradores nas regiões Sudeste e Centro-Oeste | 27 |
| 2.5 | Expansão esperada na matriz elétrica das regiões Sudeste e Centro-Oeste | 28 |
| 2.6 | Custos Fixos e Variáveis para Geração de Eletricidade no Brasil | 29 |
| 2.7 | Diagrama básico de funcionamento de uma usina hidrelétrica. | 32 |
| 2.8 | Usinas hidrelétricas a fio d'água (à esquerda) e de desvio (à direita). | 33 |
| 2.9 | Usina hidrelétrica com barragem. | 33 |
| 2.10 | Usina hidrelétrica reversível. | 34 |
| 2.11 | Velocidade média do vento durante as quatro estações do ano no Brasil. | 35 |
| 2.12 | Capacidade instalada acumulada de mini e micro geradores eólicos no Brasil | 37 |
| 2.13 | Diagrama de bloco de um sistema eólico. | 37 |
| 2.14 | Alguns modelos de turbinas eólicas. (a), (b) e (c): modelos com eixo horizontal. (d) e (e): modelos com eixo vertical. | 38 |
| 2.15 | Geração de potência de um aerogerador. | 39 |
| 2.16 | Geração elétrica líquida da usina eólica de Gargaú. | 39 |
| 2.17 | Radiação solar global média para cada estação no Brasil. | 41 |
| 2.18 | Radiação direta, difusa e refletida. | 42 |
| 2.19 | Diagrama de bloco de um sistema de geração elétrica fotovoltaico. | 43 |

| | | |
|------|---|-----|
| 2.20 | Gráficos de dispersão de variáveis (a) com correlação linear positiva; (b) com correlação linear negativa; (c) sem correlação linear | 47 |
| 2.21 | Exemplos de séries com correlação linear positiva, correlação linear negativa e sem correlação linear com respeito a uma variável base (acima); e as mesmas séries plotadas como diagramas de dispersão | 48 |
| 2.22 | Esquema gráfico de uma análise de fatores com quatro variáveis observáveis (X_i , representadas por quadrados) e seis não observáveis (círculos), sendo duas delas fatores (F_j) comuns às variáveis, e as demais, componentes específicos de cada variável (ε_i). Os valores c_{ij} representam a influência do fator j na variável i | 52 |
| 2.23 | Relação entre o valor-p, o parâmetro nulo testado e o parâmetro calculado com base na amostra. | 54 |
| 2.24 | Exemplo de gráfico <i>scree</i> . Neste caso, a recomendação seria usar 4 fatores. | 62 |
| 2.25 | Exemplo de matriz de cargas | 63 |
| 2.26 | Exemplo de rotação em um modelo de análise fatorial com dois fatores. Os pontos representam as variáveis e são plotados de acordo com a carga relacionada a cada fator. (a) Eixos originais. (b) Eixos rotacionados | 64 |
| 3.1 | Algumas áreas de enfoque de trabalhos que avaliam a inserção de fontes de intermitentes em sistemas elétricos. | 68 |
| 4.1 | Diagrama com o resumo dos métodos adotados no estudo | 74 |
| 4.2 | Etapas do controle de qualidade de dados meteorológicos | 79 |
| 4.3 | Localização geográfica das estações eólicas e solares consideradas no estudo | 84 |
| 4.4 | Localização geográfica das estações do ONS | 85 |
| 4.5 | Localização geográfica das estações da ANA | 87 |
| 4.6 | Diagrama esquemático do algoritmo de seleção de fatores. | 95 |
| 4.7 | Comparação entre o comportamento original da série e (a) a série com padronização <i>z-score</i> e (b) a série dividida por seu valor médio. | 97 |
| 5.1 | Correlações entre as estações da ANA | 104 |
| 5.2 | Correlações entre as estações do ONS | 105 |

| | | |
|------|---|-----|
| 5.3 | Correlações entre as estações solares do INPE | 106 |
| 5.4 | Correlações entre as estações solares do INPE durante a noite (a) e o dia (b) | 107 |
| 5.5 | Correlações entre as estações eólicas do INPE | 108 |
| 5.6 | Resultado da análise de autovalores para os dados da Ana (cada índice representa um autovalor). (a) Autovalor por ordem de extração. (b) Variância acumulada ao longo dos autovalores. | 111 |
| 5.7 | Análise com cinco fatores para os dados da ANA. | 112 |
| 5.8 | Resultado da análise de autovalores para os dados de vazões do ONS (cada índice representa um autovalor). (a) Autovalor por ordem de extração. (b) Variância acumulada ao longo dos autovalores. | 113 |
| 5.9 | Análise com seis fatores para os dados de vazões do ONS. | 114 |
| 5.10 | Resultado da análise de autovalores para os dados de radiação solar (cada índice representa um autovalor). (a) Autovalor por ordem de extração. (b) Variância acumulada ao longo dos autovalores. | 115 |
| 5.11 | Resultado da análise de autovalores para os dados de velocidade do vento (cada índice representa um autovalor). (a) Autovalor por ordem de extração. (b) Variância acumulada ao longo dos autovalores. | 116 |
| 5.12 | Correlação linear horária entre o grupo fotovoltaico e a geração eólica de cada estação. | 117 |
| 5.13 | Correlação linear horária entre o grupo fotovoltaico e a geração eólica de cada estação nos períodos noturno (a) e diurno (b). | 118 |
| 5.14 | Comportamento diário médio das séries de acordo com a estação do ano. A linha preta tracejada representa o grupo fotovoltaico e as demais são as séries eólicas. Os gráficos estão na mesma escala (eixos idênticos). | 119 |
| 5.15 | Correlação linear diária entre o grupo fotovoltaico, os grupos hidrelétricos e as gerações eólicas. | 120 |
| 5.16 | Correlações lineares diárias entre o grupo fotovoltaico, os grupos hidrelétricos e as séries eólicas, durante a primavera (a), o verão (b), o outono (c) e o inverno (d). | 122 |

| | | |
|------|--|-----|
| 5.17 | Comportamento anual médio das séries. Obs.: os gráficos têm escalas de eixo vertical distintas. | 124 |
| 5.18 | Participação dos grupos de fontes na potência total do <i>mix</i> renovável, de acordo com a resolução do modelo da Formulação I. As fontes não apresentadas tiveram participação nula. | 125 |
| 5.19 | Comportamento anual médio do grupo 2 da ANA. | 126 |
| 5.20 | Participação dos grupos de fontes na potência total do <i>mix</i> renovável, de acordo com a resolução do modelo da Formulação II. As fontes não apresentadas tiveram participação nula. | 127 |
| 5.21 | (a) Comparação entre a série histórica de demanda elétrica e a série do <i>mix</i> intermitente otimizada de acordo com a Formulação II. (b) Diferença percentual entre geração do <i>mix</i> otimizado e a demanda durante o período de estudo. | 128 |
| A.1 | Código da função <i>solve.QP()</i> para resolução de problemas de programação quadrática. | 135 |

Lista de Tabelas

| | | |
|-----|--|-----|
| 2.1 | Potência total e participação de cada tipo de usina no <i>mix</i> elétrico das regiões SE/CO. | 26 |
| 2.2 | Geração hidrelétrica e fração correspondente à geração elétrica total nos estados das regiões Sudeste e Centro-Oeste, no Brasil e em todo o mundo. . | 31 |
| 2.3 | Potencial Eólico para o Brasil e para as regiões Sudeste e Centro-Oeste, a 50m de altura. | 36 |
| 2.4 | Potencial de geração fotovoltaica residencial. | 42 |
| 5.1 | Observações versus Número de Variáveis | 109 |
| 5.2 | Resultados do teste de esfericidade de Bartlett | 109 |
| 5.3 | Valores do KMO para cada fonte | 109 |
| 5.4 | Resumo dos valores de MSA de cada fonte | 110 |
| 5.5 | Valor da Função Objetivo da Formulação I para os cenários original, sem geração fotovoltaica e sem o grupo 2 da ANA | 126 |
| 5.6 | Valor da Função Objetivo da Formulação II para os cenários original, sem geração fotovoltaica e sem o grupo 2 da ANA | 127 |
| 5.7 | <i>Mix</i> renovável intermitente instalado nas regiões SE e CO | 129 |

Capítulo 1

Introdução

O domínio da geração de eletricidade, que ocorreu por volta da segunda metade do século XIX, revolucionou a vida humana, modificando desde processos industriais e agrícolas até a rotina da sociedade. Ainda hoje, a demanda por eletricidade permanece crescente, mas a geração passa por uma espécie de revolução interna, movida pelo objetivo de se ter um suprimento confiável, com custos acessíveis e alinhado com metas ambientais internacionais [70].

Essa revolução tem sido marcada pelo aumento da participação de fontes renováveis, especialmente não-hídricas nas matrizes elétricas. Isso porque estudos têm mostrado que é possível utilizar porcentagens elevadas de fontes renováveis, frente ao risco dos combustíveis fósseis chegarem ao limite de sua disponibilidade aliado às metas ambientais internacionais, especialmente no que diz respeito à diminuição das emissões de gases do efeito estufa [44].

Há, no entanto, alguns obstáculos à inserção em larga escala de fontes renováveis nos sistemas elétricos. Algumas tecnologias ainda são pouco eficientes e outras ainda têm custo muito elevados, como a geração undielétrica e as células combustíveis. As fontes cujas tecnologias de exploração se encontram mais desenvolvidas enfrentam outros desafios.

Dentre as fontes renováveis consolidadas, há aquelas cujo suprimento é aproximadamente constante, como a energia geotérmica, hidrelétrica com regulação de reservatórios e biomassa. No entanto, há outras que dependem de variáveis meteorológicas, e por isso, têm variabilidade alta e são pouco previsíveis. É o caso da geração eólica, solar e hidrelétrica (quando desprovida de reservatório para armazenagem). Essas fontes são chamadas de intermitentes.

A expansão da parcela renovável intermitente de uma matriz elétrica deve ser acompanhada com cautela devido à variabilidade dos recursos naturais, como o vento, a radiação solar e a precipitação, que podem variar diária, mensal e anualmente [88]. Essas variações geram a necessidade de se ter um sistema elétrico flexível, para que não haja demanda não atendida ou elevação de custos para suprir esta inconstância.

Inicialmente, pensava-se que, para isso, o suporte por geradores termoelétricos era inevitável, mas atualmente vários estudos apontam para a viabilidade de se utilizar fontes renováveis que variem de forma complementar no tempo ou no espaço geográfico, a fim de diminuir a variância global da geração [18].

Em 2015, o Brasil se comprometeu, junto à Organização das Nações Unidas, a aumentar para 23% a participação de fontes renováveis não-hídricas na geração de eletricidade até 2030, destacando a energia eólica, solar e biomassa [21]. Em 2016, a participação dessas fontes na produção de energia elétrica foi de 16,1% [49].

Dentre as regiões brasileiras, Sudeste e Centro-Oeste são responsáveis por 41,87% da geração elétrica e por 57,4% do consumo [48]. Nessas regiões, a geração de eletricidade por fontes renováveis não hídricas é semelhante ao total nacional: 15,1% no Sudeste e 15,9% no Centro-Oeste [49]. No entanto, essa geração baseia-se fortemente em biomassa, especialmente bagaço de cana e lixo. A participação de fontes intermitentes ainda é muito baixa nessas regiões e, portanto, a análise de complementaridade é bem-vinda para orientar a expansão do parque gerador. Nesse caso, o estudo não só deve avaliar como os recursos variam nas regiões, mas também calcular o *mix* que maximize os benefícios da complementaridade, seja minimizando a variância do grupo, ou minimizando o desvio entre demanda e geração.

Apesar das metas tratarem de fontes renováveis não hídricas, várias pequenas hidrelétricas com geração intermitente já estão instaladas nas regiões e devem ser consideradas. Mesmo uma expansão futura desse tipo de geração hídrica pode suavizar ainda mais a variabilidade da geração renovável.

Estudos de complementaridade utilizam métodos diferentes de acordo com o objetivo relacionado [68]. Alguns avaliam apenas o comportamento dos recursos naturais para diferentes períodos de tempo e estações do ano, fazendo uso principalmente de cálculos de correlações lineares. Há também trabalhos que avaliam correlações, mas vão além: buscam otimizar o conjunto das fontes estudadas, minimizando a variância do grupo ou minimizando custos de armazenagem ou demanda não atendida, por exemplo.

Já foram feitos estudos de complementaridade entre fontes intermitentes no Brasil, como o [106] [23] [37] [18] [33], mas não foram encontrados registros de trabalhos que tenham realizado o processo de otimização considerando mais de uma opção para um recurso, ou seja, que tenha considerado diversos grupos de geração hidrelétrica, fotovoltaica e eólica, o que seria razoável pela extensão territorial considerada. Além disso, também não foram encontrados estudos de complementaridade específicos para as regiões Sudeste e Centro-Oeste.

Para criar grupos dentro de cada tipo de fonte, deve-se optar por algum método de agrupamento de variáveis. Um método bem estabelecido e que pode ser usado é a Análise de Fatores. Outros trabalhos já realizaram agrupamentos por fonte [63][23], porém eles não incluíam etapas de otimização. No trabalho [55], utilizaram-se dois grupos distintos de geração hidrelétrica, mas só havia mais uma variável (potência fotovoltaica), de forma que não foi necessário elaborar um modelo de otimização formal, pois o problema, que precisa definir apenas duas variáveis, pôde ser resolvido testando-se pares de valores. Não foi encontrado um estudo de complementaridade com otimização que apresente um modelo que maximize mais do que três variáveis.

1.1 Objetivos

1.1.1 Objetivo Geral

Como objetivo principal, buscou-se avaliar a complementaridade dos recursos eólico, fotovoltaico e hídrico nas regiões Sudeste e Centro-Oeste e identificar de que forma devem ser combinados, de modo que a característica intermitente das fontes seja suavizada.

1.1.2 Objetivos Específicos

1. Compreender a matriz elétrica atual das regiões Sudeste e Centro-Oeste, através de pesquisa bibliográfica;
2. Coletar e tratar dados meteorológicos para obter séries de potência elétrica teórica;
3. Agrupar geograficamente cada geração (eólica, fotovoltaica e hidrelétrica), de acordo com suas similaridades, utilizando Análise de Fatores;
4. Avaliar a complementaridade entre os grupos de fontes;

5. Elaborar um modelo de otimização de *mix* renovável que contemple mais do que três variáveis;
6. Obter a combinação das fontes intermitentes em questão que minimize a variância do grupo e a combinação que minimize a demanda não atendida e a geração excedente.

1.2 Justificativa e Relevância

A inserção de fontes renováveis intermitentes nos sistemas elétricos deve ser feita de forma planejada, a fim de evitar possíveis falhas de suprimento.

Análises de complementaridade têm sido realizadas para diversas fontes e em diversas regiões do mundo e, aliadando-as a algum método de otimização, tem-se um método quantitativo para direcionar a tomada de decisão em direção a um parque gerador mais limpo.

1.3 Estrutura do Trabalho

Esta dissertação se subdivide em seis capítulos: introdução, fundamentação teórica, trabalhos relacionados, metodologia, resultados e conclusões e trabalhos futuros.

Este primeiro capítulo contextualiza a pesquisa, apresenta o objetivo geral do trabalho e os objetivos específicos, bem como a justificativa do estudo e a estrutura e as limitações do trabalho. O segundo capítulo apresenta a matriz elétrica atual das regiões Sudeste e Centro-Oeste, fundamentos teóricos de geração de eletricidade por fontes intermitentes e seu potencial nas regiões estudadas, a problemática de inserir fontes intermitentes em sistemas elétricos e a base teórica das ferramentas que são utilizadas: correlação linear e análise de fatores. O terceiro capítulo apresenta uma revisão dos trabalhos relacionados a este. O quarto capítulo descreve os métodos aplicados, incluindo informações sobre as bases de dados usadas. O quinto capítulo traz os resultados do estudo e a discussão deles. O último capítulo apresenta as conclusões tiradas e sugestões de trabalhos futuros que possam contribuir para o desenvolvimento do tema.

1.4 Limitações

O trabalho concerne somente às regiões Sudeste e Centro-Oeste e o período estudado foi de janeiro de 2007 a dezembro de 2011, devido às restrições dos dados disponíveis. Apenas as séries de potência eólica e solar possuíam resolução horária (oito medições ao longo do dia), logo a otimização foi feita a nível diário, para se considerar também as séries de potência hidrelétrica.

Capítulo 2

Fundamentação Teórica

Neste capítulo, aborda-se as características básicas de consumo e geração de eletricidade no Brasil, no mundo e nas regiões Sudeste e Centro-Oeste. Além disso, são apresentados fundamentos de geração de eletricidade a partir das fontes intermitentes consideradas neste trabalho. Discorre-se também da inserção de fontes intermitentes nos sistemas elétricos tradicionais. Em seguida, são apresentadas as bases teóricas das ferramentas incluídas na metodologia do trabalho: correlação linear, análise de fatores e otimização.

2.1 Panorama Atual da Eletricidade

Esta seção trata da geração e do consumo de eletricidade, partindo de uma análise geral do Brasil e do mundo e, depois, avaliando especificamente as regiões Sudeste e Centro-Oeste.

2.1.1 A eletricidade no mundo e no Brasil

A demanda mundial por eletricidade tem crescido em média 3% ao ano desde o ano 2000, representando 19% do consumo total de energia em 2017 [70]. Esse crescimento, no entanto, está diretamente ligado a países emergentes, já que o consumo dos países mais ricos se manteve aproximadamente estável nesse período (2000-2017) [70]. A ampliação da participação de fontes renováveis na geração elétrica, por outro lado, é uma realidade global. A expectativa é que cresçam a uma taxa de 2,9% ao ano, entre 2012 e 2040, com destaque para as fontes renováveis não-hídricas, cuja perspectiva de crescimento é de 5,7% ao ano [41].

A demanda por fontes alternativas está associada não só a esforços globais de diminuir a emissão de gases de efeito estufa, mas também à preocupação com as finitas reservas de combustíveis fósseis [67], aos desastres ecológicos ligados ao suprimento de energia [38] e à busca por fontes de baixo custo de geração elétrica, que levam a uma competitividade econômica maior [111].

O Brasil, como economia em desenvolvimento, segue a tendência mundial, tanto de aumento no consumo elétrico, como na reformulação de seu *mix* de fontes energéticas para geração de eletricidade. Atualmente o país conta com os seguintes tipos de usinas de geração [14]:

- Termelétrica (UTE), movidas a combustíveis fósseis ou biomassa;
- Termonuclear (UTN);
- Eólica (EOL);
- Fotovoltaica (UFV);
- Hidrelétrica, que de acordo com sua potência instalada e área de reservatório (como será apresentado na Seção 2.2), é subdividida em:
 - centrais geradoras hidrelétricas (CGHs)
 - pequenas centrais hidrelétricas (PCHs)
 - usinas hidrelétricas (UHEs)
- Undielétrica (uma usina de baixa capacidade implementada no porto de Pecém-CE)

A evolução da capacidade instalada nacional de cada tipo de usina é apresentada na Figura 2.1.

Percebe-se que a matriz brasileira é significativamente renovável, porém isso se deve à participação de grandes hidrelétricas. A busca pela diversificação das fontes renováveis ainda está se iniciando. Por um lado, a geração eólica cresceu 26,5% entre 2016 e 2017, mas por outro, a participação de fontes não renováveis na geração de eletricidade aumentou de 2016 para 2017 (de 19,6% para 20,8% do total) [49].

Outra característica importante da matriz elétrica brasileira é a sua ampla conexão, através do Sistema Interligado Nacional (SIN), que reúne usinas, linhas de transmissão

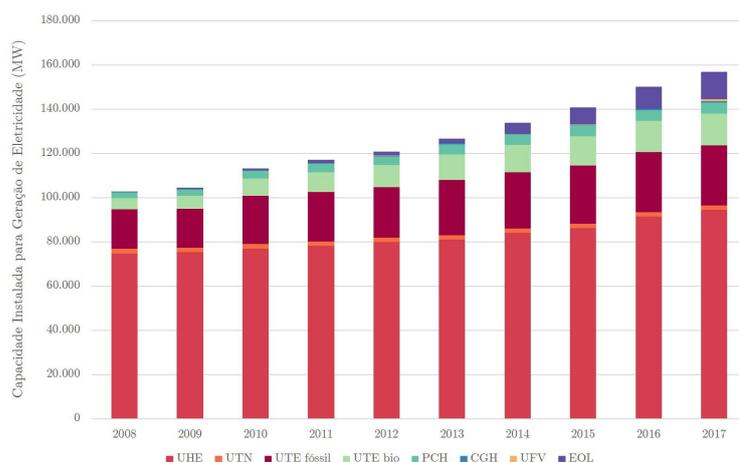


Figura 2.1: Evolução da capacidade instalada de geração de eletricidade no Brasil.
Fonte: Elaboração própria com dados de [49]

e ativos de distribuição [10]. Além dele, devido à vegetação e a aspectos geográficos de algumas áreas, existem também sistemas isolados [10]. No entanto, a quantidade de sistemas isolados vem diminuindo. Em 2008, o SIN cobria 96,6% da capacidade elétrica nacional [10] e em 2016, a cobertura passou a ser de 98,3% [90]. Hoje, a demanda dos sistemas isolados não chega a 1% do total nacional e inclui algumas localidades na região norte, o extremo norte de Mato Grosso e a ilha de Fernando de Noronha [94].

A grande dimensão do SIN é muito importante, pois permite a importação e exportação de energia entre os estados, aumentando a confiabilidade do sistema. O SIN se subdivide em quatro grandes subsistemas: Sul, Norte, Nordeste e Sudeste/Centro-Oeste (SE/CO). Este último foi responsável por 58,2% do consumo elétrico brasileiro em 2016 [48].

Apesar da grande participação no SIN, o subsistema SE/CO não tem sido o grande impulsionador da renovação do *mix* elétrico brasileiro no que diz respeito ao uso de fontes intermitentes. Por outro lado, a biomassa tem apresentado boa participação. A próxima subseção é dedicada à compreensão da matriz elétrica desses estados.

Ressalta-se que recentemente parte dos estados de Rondônia e Acre foram incorporados ao subsistema SE/CO. No entanto, ainda há muitos sistemas isolados nos dois estados, o que dificulta a limitação geográfica do estudo. Além disso o consumo de eletricidade dos estados das regiões sudeste e centro-oeste correspondem a 98,7% do consumo do subsistema SE/CO [48]. Logo, serão considerados apenas os estados das regiões sudeste e centro-oeste daqui em diante.

2.1.2 A eletricidade nas regiões Sudeste e Centro-Oeste

As regiões Sudeste e Centro-Oeste compreendem os estados Rio de Janeiro (RJ), Espírito Santo (ES), Minas Gerais (MG), São Paulo (SP), Mato Grosso (MT), Mato Grosso do Sul (MS) e Goiás (GO), além do Distrito Federal (DF).

Neles existem os mesmos tipos de usinas que as apresentadas para o Brasil, com exceção da usina undielétrica. A participação de cada usina na potência fiscalizada dos estados é apresentada na Figura 2.2.

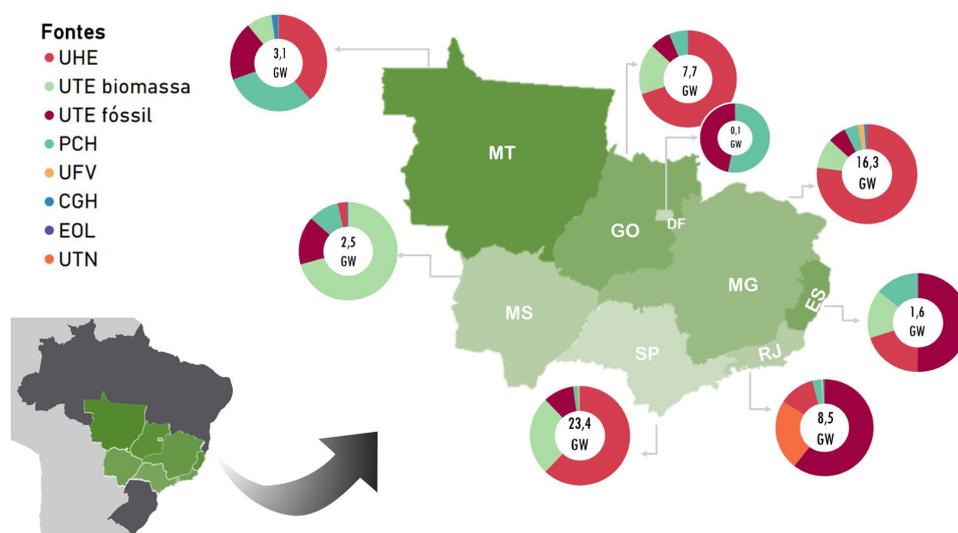


Figura 2.2: Participação de cada tipo de usina na potência fiscalizada total dos estados das regiões SE e CO. No centro de cada gráfico se encontra a capacidade instalada naquele estado.

Fonte: Elaboração própria com dados de [14]

Pode-se perceber que mais da metade da capacidade instalada das regiões provém dos estados de MG e SP e que estes dependem fortemente de grandes hidrelétricas. Os estados do Rio de Janeiro, Espírito Santo, e o Distrito Federal têm elevada participação de usinas termelétricas que utilizam combustíveis fósseis. É interessante notar que o estado de Mato Grosso do Sul tem quase 75% de sua matriz composta por fontes renováveis não hídricas. A potência por fonte é apresentada na Tabela 2.1.

As usinas apresentadas na Tabela 2.1 podem ser classificadas, quanto ao destino da eletricidade gerada, em serviço público de distribuição, autoprodução de energia (indús-

| Usina | Potência Fiscalizada (MW) | Participação |
|--------------|---------------------------|--------------|
| UHE | 35.098,86 | 55,6% |
| UTE biomassa | 11.239,52 | 17,81% |
| UTE fóssil | 10.761,61 | 17,05% |
| PCH | 3.225,92 | 5,11% |
| UTN | 1.990,00 | 3,15% |
| UFV | 478,77 | 0,76% |
| CGH | 302,7 | 0,48% |
| EOL | 28,21 | 0,04% |

Tabela 2.1: Potência total e participação de cada tipo de usina no *mix* elétrico das regiões SE/CO.

Fonte: Elaboração própria com dados de [14]

trias que geram energia para consumo próprio podendo, ou não, comercializar parte da energia), produção independente de energia, quando os produtores vendem, por sua conta e risco, a energia a distribuidoras ou consumidores independentes [9] [47] , ou ainda, se a central geradora tem até 5.000kW de potência, sua implantação é apenas comunicada a Aneel e é classificada como registro [12].

Da Tabela 2.1 pode-se extrair duas informações importantes. A primeira é a participação positiva de biomassa na malha elétrica, o que desperta o interesse em compreender que tipos de combustíveis são utilizados por elas. A segunda é que as fontes renováveis intermitentes (eólica, solar e pequenas hidrelétricas - PCHs e CGHs) têm participação ainda pouco expressiva, não chegando a 7% da capacidade instalada.

Quanto às termelétricas à biomassa, a participação dos combustíveis utilizados por elas são apresentados na Figura 2.3.

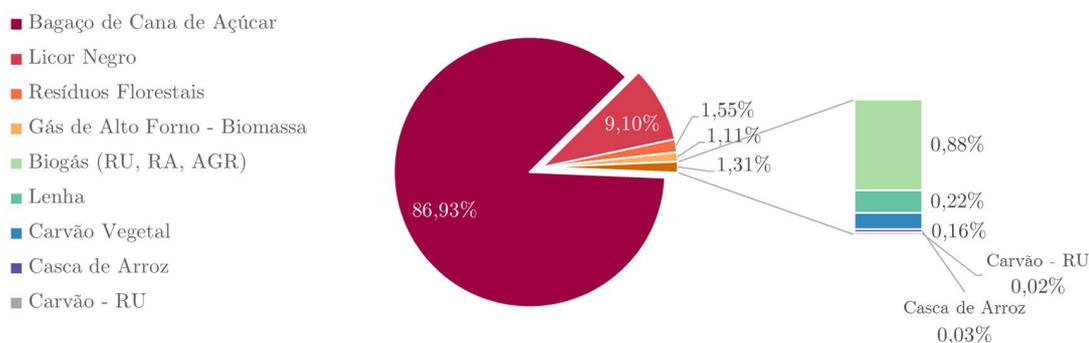


Figura 2.3: Combustíveis usados em usinas termelétricas à biomassa nas regiões Sudeste e Centro-Oeste.

Fonte: Elaboração própria com dados de [14]

A grande utilização do bagaço de cana de açúcar como combustível se deve à forte indústria sucroalcooleira brasileira. Essa fonte não só é interessante como recurso renovável de geração de eletricidade, como também mostra que aquilo que poderia ser visto como resíduo da produção de álcool e açúcar, é na verdade, um subproduto, o que aumenta a ecoeficiência da indústria em questão. De forma semelhante, o segundo combustível mais utilizado, o licor negro (lixívia), é um subproduto da indústria papelreira. No grupo de biogás, RU representa resíduos sólidos urbanos, RA representa resíduos animais e AGR representa resíduos agroindustriais.

Ainda no quesito capacidade instalada, deve-se destacar a existência de mini e microgeradores, não contabilizados acima e que se caracterizam pela produção de eletricidade “a partir de pequenas centrais geradoras que utilizam fontes com base em energia hidráulica, solar, eólica, biomassa ou cogeração qualificada, conectadas à rede de distribuição por meio de instalações de unidades consumidoras” [11]. Neste grupo se encontram, por exemplo, os painéis fotovoltaicos residenciais que enviam energia excedente para rede e em troca recebem créditos da empresa distribuidora, abatendo o valor de sua conta de energia.

Apresenta-se a potência instalada dos mini e micro geradores do SE/CO na Figura 2.4.

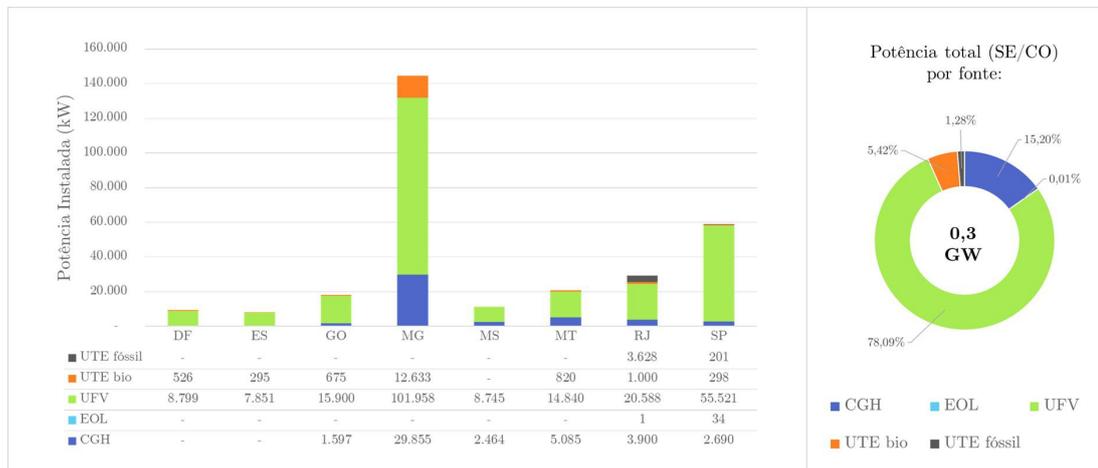


Figura 2.4: Mini e micro geradores nas regiões Sudeste e Centro-Oeste
Fonte: Elaboração própria com dados de [15]

Os grandes destaques em mini/micro geração são os painéis fotovoltaicos e a grande participação do estado de Minas Gerais. No entanto, se a potência total de todos os mini/micro geradores fosse contabilizada junto à capacidade instalada da Tabela 2.1, a participação deles não chegaria a 0,5% do total.

Além de compreender a matriz elétrica atual, deve-se avaliar qual é a expectativa de expansão dela. Atualmente, há empreendimentos em operação, mas também existem empreendimentos sendo construídos e outros aprovados, mas que ainda não tiveram as obras iniciadas. Mostra-se o cenário esperado quando as obras se concluírem na Figura 2.5 (a potência contabilizada das usinas que ainda não estão em operação é a outorgada).

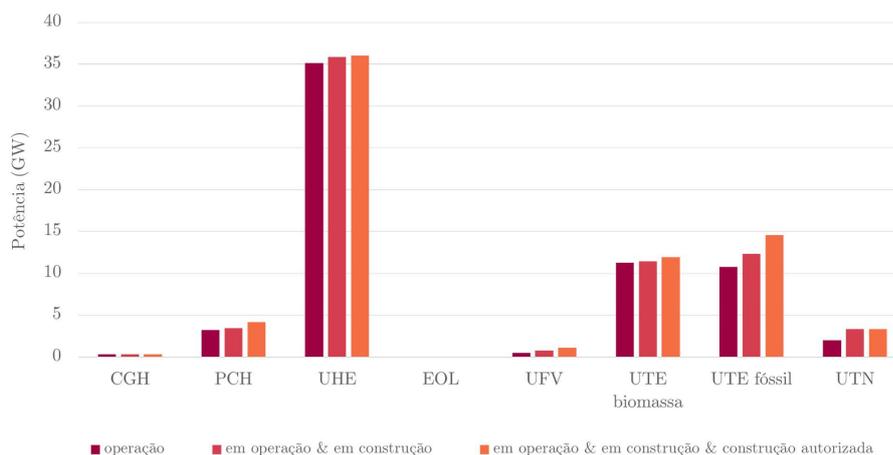


Figura 2.5: Expansão esperada na matriz elétrica das regiões Sudeste e Centro-Oeste
Fonte: Elaboração própria com dados de [14]

A expectativa de crescimento das usinas fotovoltaicas é um fato positivo, ainda que a variação seja pequena. No entanto, o que mais chama a atenção na Figura 2.5 é o crescimento esperado das termelétricas de fontes fósseis. Por um lado, a matriz atual tem como ponto forte o fato de que as termelétricas que utilizam biomassa têm participação parelha com as de combustível fóssil, com potência instalada até mesmo superior. Por outro lado, a tendência é de crescimento mais forte das térmicas fósseis, o que vai na contramão da tendência mundial de substituição das fontes fósseis.

Das novas termelétricas convencionais, 99% da potência esperada virá de usinas que utilizam gás natural como combustível. A ampla utilização do gás natural na geração termelétrica do estado merece atenção por conta de três aspectos fundamentais – a finitude do recurso, seu custo e seus impactos ambientais, ligados principalmente à emissão de gases causadores de efeito estufa, por ser um combustível de origem fóssil.

No que tange ao custo, [29] realizaram um levantamento do preço médio de venda de eletricidade proveniente de diversas fontes no Brasil, considerando os custos fixos médios entre 2005 e 2015 da venda em leilões e os custos variáveis, que englobam o custo do combustível para geração flexível de energia e demais custos variáveis [45], para o mês de março de 2015. Os dados apresentados pelo trabalho em questão são mostrados de forma

gráfica por meio da Figura 2.6.

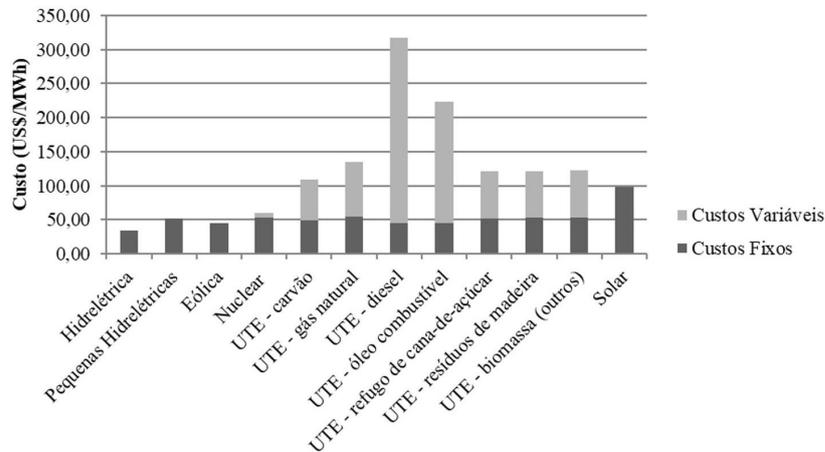


Figura 2.6: Custos Fixos e Variáveis para Geração de Eletricidade no Brasil
Fonte: Elaboração própria com dados de [29]

Pode-se observar que as usinas termelétricas apresentam custos variáveis significativos, que levam ao aumento do custo da energia como um todo. Ademais, além destes custos variáveis e dos custos fixos, os sistemas termelétricos possuem custos de partida e, devido aos processos de aquecimento e resfriamento necessários para o seu acionamento/desligamento, após a partida a unidade deve continuar em funcionamento por certo tempo antes de ser desligada e, de igual modo, após o desligamento, a unidade deve permanecer desativada por algumas horas [28].

O Brasil tem como meta ampliar a participação de fontes renováveis não-hídricas para 23% do total até 2030, destacando as energias eólica, solar e proveniente de biomassa [21]. Apesar das metas serem nacionais, não é irrazoável esperar que os subsistemas contribuam de forma homogênea para isso. Entretanto, a expectativa de expansão apresentada na Figura 2.5 não condiz com as metas estipuladas. A participação de biomassa é significativa, porém aparenta estar estagnada. Há oportunidade de inserção de fontes intermitentes (eólica e solar, especialmente fotovoltaica), que podem complementar as PCHs e CGHs instaladas e também potencialmente intermitentes.

2.2 Energias Renováveis

Esta Seção aborda características e princípios básicos de geração de eletricidade a partir de cursos hídricos, do vento e do sol, além de informações sobre o potencial de geração nas regiões estudadas.

2.2.1 Energia hidráulica

Nesta subseção, refere-se à hidroeletricidade. Apresenta-se os mecanismos básicos de aproveitamento da energia hidráulica e o potencial de geração hidrelétrica no Brasil e no mundo.

2.2.1.1 O Recurso e seu Potencial

De acordo com a publicação *The World Factbook* [26], o Brasil é o país com a maior disponibilidade hídrica em longo prazo, com aproximadamente 8.233 km^3 de recursos hídricos renováveis, contabilizando precipitação, águas subterrâneas e entradas de água através de países vizinhos.

A reserva total de água no mundo se encontra nas formas de vapor atmosférico, rios, lagos, água subterrânea, oceanos, gelo e geleiras. Ao receber a radiação solar, parte da água no estado líquido evapora e fica sob a forma de vapor na troposfera, camada mais baixa da atmosfera, onde circula graças ao movimento do vento, e eventualmente, se condensa e precipita novamente sobre a superfície terrestre. A água precipitada pode ser absorvida pelo solo, alimentar geleiras, evaporar (de forma instantânea, a partir de reservatórios, ou pela evapotranspiração da vegetação) ou ainda escoar pela terra, alcançando rios, lagos e oceanos. A parcela da precipitação que seguirá cada um desses destinos depende de diversos fatores, como a vegetação, permeabilidade e topografia da região em questão. Dessa forma, a vazão dos rios depende do ciclo hidrológico, mas não unicamente do montante de precipitação. [79][30][38].

A energia da água que flui através de rios pode se tornar aproveitável à sociedade através de aproveitamentos hidráulicos, que são aqueles que usufruem da energia mecânica dos cursos de água para realização de trabalho ou geração de eletricidade, ou seja, utilizam tanto a energia potencial existente graças à altitude do recurso em relação ao nível do mar, quanto à energia cinética dada a vazão do mesmo [62].

A utilização da energia hidráulica para geração de trabalho útil através de rodas d'água

remonta ao primeiro século a.C., graças aos gregos, e, desde então, foi muito utilizada pela humanidade. O advento do gerador elétrico no século XIX, no entanto, revolucionou a utilização da força das águas e permitiu o desenvolvimento de usinas hidrelétricas, capazes de converter a energia mecânica em energia elétrica [67].

As usinas hidrelétricas são muito importantes nos dias de hoje. A geração dessas usinas, tanto em valores absolutos quanto em percentual do montante total de eletricidade gerada, para os estados das regiões Sudeste e Centro-Oeste, além do Brasil e de todo o mundo, é apresentada na Tabela 2.2.

| | Geração Hidrelétrica (TWh) | | | | Fração da Produção Elétrica Total Proveniente de Hidrelétricas | | | |
|---------------|----------------------------|---------|---------|--------|--|-------|-------|-------|
| | 2008 | 2010 | 2012 | 2014 | 2008 | 2010 | 2012 | 2014 |
| RJ | 6,3 | 7,51 | 7,95 | 5,19 | 14,9% | 17,4% | 15,4% | 8,6% |
| SP | 74,4 | 78,25 | 78,49 | 56,65 | 91,5% | 91,7% | 90,9% | 87,5% |
| MG | 57,81 | 62,85 | 70,45 | 45,24 | 92,5% | 95,8% | 95% | 86,3% |
| ES | 1,44 | 2,1 | 2,14 | 1,75 | 24% | 34% | 31% | 20% |
| GO | 24,33 | 28 | 38,41 | 26,04 | 96,4% | 95% | 94,7% | 83% |
| MT | 7,64 | 7,3 | 9,67 | 10,91 | 89,4% | 94,2% | 82,5% | 75,9% |
| MS | 21,64 | 21,79 | 21 | 15,84 | 100% | 96,8% | 91,7% | 73,2% |
| Brasil | 369,62 | 403,36 | 415,41 | 373,51 | 79,8% | 78,2% | 75,2% | 63,3% |
| Mundo | 3180,52 | 3408,45 | 3630,46 | 3855,7 | 16,6% | 16,7% | 16,8% | 17% |

Tabela 2.2: Geração hidrelétrica e fração correspondente à geração elétrica total nos estados das regiões Sudeste e Centro-Oeste, no Brasil e em todo o mundo.

Fonte: Balanços energéticos estaduais e [49] e [42]

Pode-se perceber que, entre os anos de 2008 e 2014, a participação das hidrelétricas aumentou discretamente a nível global, porém caiu em todos os estados apresentados e no Brasil, mesmo que o montante gerado tenha aumentado em alguns casos. Esse comportamento coincide com as recentes restrições à construção de grandes hidrelétricas, devido a seus grandes impactos ambientais e sociais [10].

Em contrapartida, empreendimentos hidrelétricos de menor porte vêm recebendo muitos investimentos e incentivos nos últimos anos [38]. Dentre tais empreendimentos, encontram-se as PCHs e as CGHs. A Aneel classifica como PCHs as usinas hidrelétricas com potência instalada entre 3.000 e 30.000 kW, reservatórios de até $13km^2$ e produção independente ou autoprodução de energia [12]. Empreendimentos hidrelétricos com potência inferior a 3.000 kW são consideradas CGHs e estão isentas de registro pela Aneel [12]. As grandes usinas hidrelétricas (UHEs) compreendem os demais casos.

O Brasil possui um potencial instalado de geração hidrelétrica muito alto, como

mostrou-se na Figura 2.1. Entretanto, estima-se que os aproveitamentos em operação não chega à metade do potencial hidráulico nacional (44%) [49]. Na região Sudeste, o aproveitamento atual é de 59% e, na região Centro-Oeste, é de apenas 32% [49].

2.2.1.2 Tecnologias de Aproveitamento

O esquema básico da geração de eletricidade em uma usina hidrelétrica envolve a liberação da água do curso hídrico, partindo ou não de um reservatório, para a casa de máquinas, onde movimentará uma turbina ligada a um gerador elétrico, que transformará a energia mecânica em elétrica, como apresentado pelo diagrama da Figura 2.7.

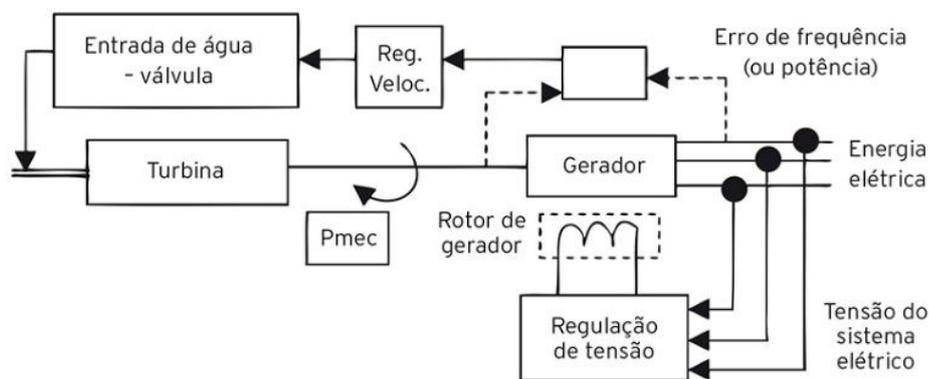


Figura 2.7: Diagrama básico de funcionamento de uma usina hidrelétrica.
Fonte: [38]

Basicamente, as usinas hidrelétricas podem ser do tipo fio d'água, ao pé da represa ou de desvio [50]. A escolha do tipo de usina dependerá, dentre outros fatores, da queda líquida e da vazão disponíveis. Os empreendimentos do tipo “fio d'água”, em sua maioria, não possuem reservatórios, aproveitando a vazão natural do rio. As usinas de desvio possuem casa de máquina localizada fora do leito do rio, sendo que a água chega até ela através de um canal que capta a água em um ponto do rio, normalmente com o auxílio de uma represa [79]. São ilustradas uma usina a fio d'água e uma de desvio na Figura 2.8.

As usinas ao pé da represa são caracterizadas pela presença de reservatório e casa de máquinas próxima ao mesmo, como mostrado na Figura 2.9.

Vêm sendo utilizadas também, usinas reversíveis, que funcionam como mecanismo de armazenamento de energia para segurança do sistema elétrico. Esses empreendimentos são formados por dois reservatórios em níveis diferentes e uma casa de máquinas capaz de bombear água do reservatório inferior para o superior, quando não há requerimento de

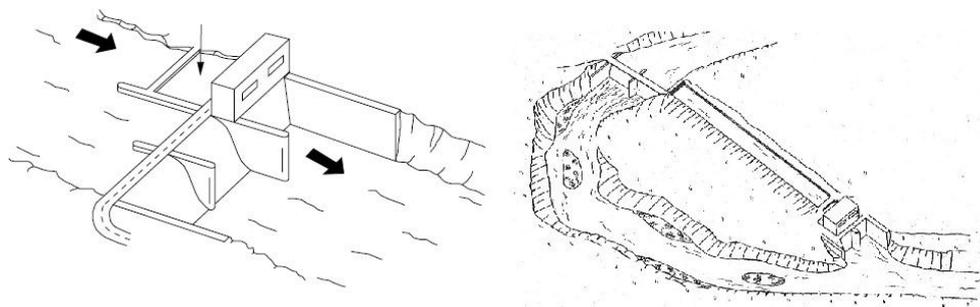


Figura 2.8: Usinas hidrelétricas a fio d'água (à esquerda) e de desvio (à direita).
Fonte: [79]

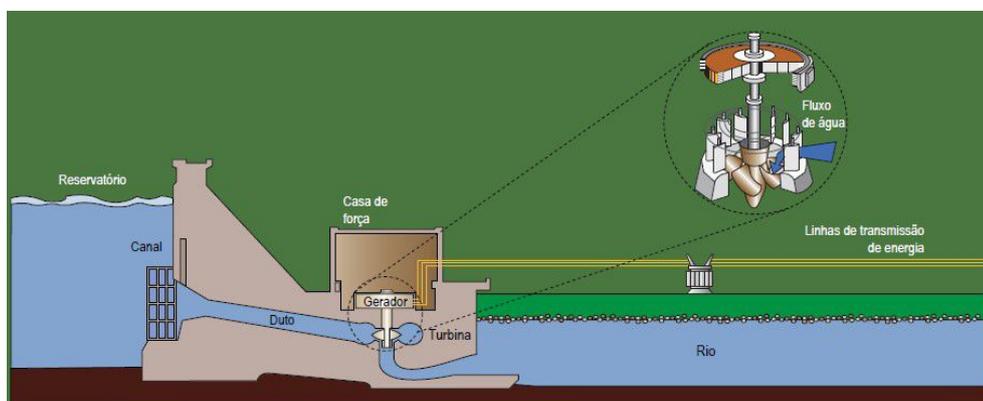


Figura 2.9: Usina hidrelétrica com barragem.
Fonte: [79]

eletricidade, e de turbinar a água do superior, lançando-o no inferior, gerando eletricidade para atender à demanda, quando necessário [34]. Ilustra-se este caso na Figura 2.10.

Pelo fato da geração hidrelétrica utilizar tanto a energia potencial como a cinética da água, sua aplicação se ajusta a casos diversos, desde quedas altas e vazões pequenas a quedas baixas e vazões altas. Para tanto, além da escolha do tipo de usina, deve-se optar por uma turbina adequada. As turbinas modernas podem ser de impulso ou reação, dependendo da forma com que a água chega até as mesmas – as a reação giram em decorrência da Terceira Lei de Newton, enquanto as de impulso recebem jatos de água [67].

A potência teórica entre dois pontos de um curso de água é deduzida a partir da Equação de Bernoulli e é dada pela Equação 2.1 [79].

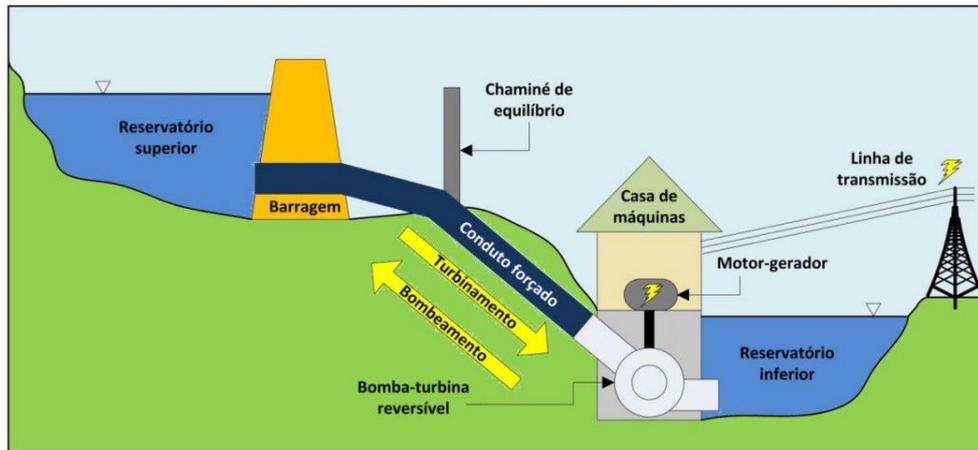


Figura 2.10: Usina hidrelétrica reversível.

Fonte: [22]

$$P = \rho g Q h \quad (2.1)$$

em que:

 P = potência hidráulica [W] ρ = densidade da água [kg/m^3] g = aceleração da gravidade [m/s^2] Q = vazão volumétrica [m^3/s] h = diferença de altura entre os dois pontos considerados [m]

A potência de uma usina hidrelétrica depende ainda da multiplicação da potência apresentada na Equação 2.1 pelo rendimento total da usina, que considera o rendimento do sistema hidráulico, da turbina e do gerador [38].

2.2.2 Energia eólica

Nesta subseção, refere-se à geração eólica. Apresenta-se os mecanismos básicos de aproveitamento da energia dos ventos, o potencial de geração eólica no Brasil e alguns dos impactos associados a essa fonte de energia.

2.2.2.1 O Recurso e seu Potencial

A geração eólica no Brasil passou de 93GWh em 2005 para 42.373GWh em 2017, ou seja, ela cresceu mais de 400 vezes em doze anos [47][49]. O país assume posição de

liderança na América Latina em potência eólica instalada, que hoje passa de 12GW, e se constitui no mercado eólico *onshore* mais promissor da região [119].

Cerca de 2,5% da radiação solar incidente sobre a camada externa da atmosfera terrestre é utilizada para movimentar massas de ar na atmosfera do planeta [79]. Quando diferenças de pressão de ar são criadas (por variação na temperatura superficial do planeta, por exemplo), massas de ar fluem das áreas de alta pressão para as de baixa pressão [79]. Fatores geográficos e ambientais também influenciam na velocidade e nas características do vento em cada localidade [114].

Mostra-se a variação da velocidade média do vento, a 50m de altura, para cada estação do ano no Brasil na Figura 2.11.

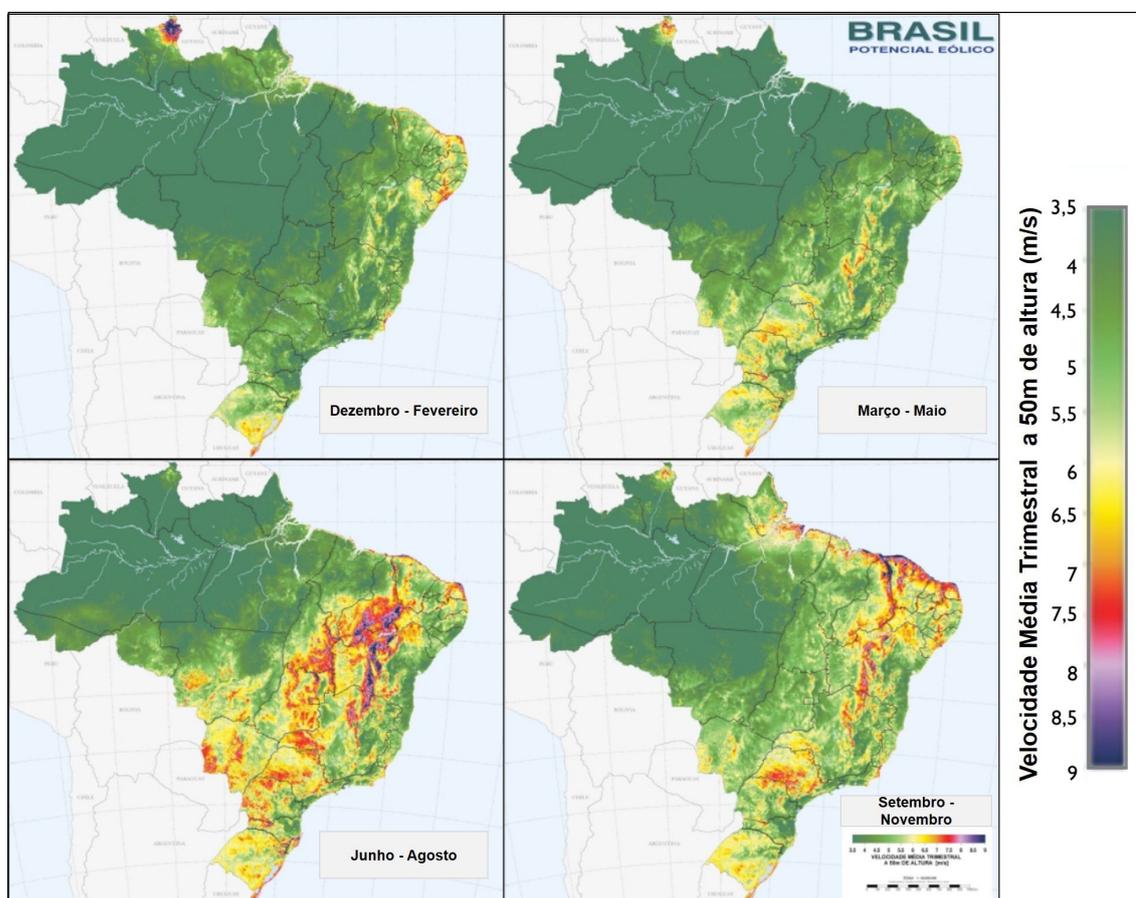


Figura 2.11: Velocidade média do vento durante as quatro estações do ano no Brasil.

Fonte: [36]

Empreendimentos que utilizam energia eólica são aqueles que convertem parte da energia cinética na atmosfera em outra forma útil de energia, como a energia mecânica [108]. Tais empreendimentos já são utilizados há muitos anos, como exemplo, pode-se citar os moinhos de vento utilizados na moagem de grãos e no bombeamento de água,

que se espalharam pela Europa a partir do século X e que, na Holanda do século XVIII chegaram a cerca de 20.000 unidades [38]. Atualmente, a utilização mais comum da energia dos ventos é feita pelos aerogeradores, que a convertem em eletricidade [62].

O Atlas do Potencial Eólico Brasileiro [36] estimou, considerando turbinas a 50m de altura e faixas de velocidade do vento acima de 7m/s, um potencial de 143GW para o Brasil, cuja potência fiscalizada instalada atualmente é de 13,5GW, além dos 2,2GW de usinas em construção e dos 2,6GW de empreendimentos cuja construção ainda não foi inicializada [14]. Na Tabela 2.3, sumariza-se os potenciais estimados para diversas faixas de velocidade do vento, considerando turbinas a 50m de altura, para o Brasil e para as regiões Sudeste e Centro-Oeste.

| Velocidades do vento acima de | Potencial Eólico (GW) | | |
|-------------------------------|-----------------------|---------|--------------|
| | Brasil | Sudeste | Centro-Oeste |
| 6 m/s | 1334,78 | 351,72 | 101,5 |
| 6,5 m/s | 463,49 | 122,34 | 19,28 |
| 7 m/s | 143,47 | 29,74 | 3,08 |
| 7,5 m/s | 43,35 | 6,65 | 0,29 |
| 8 m/s | 13,36 | 1,78 | 0,01 |
| 8,5 m/s | 3,55 | 0,59 | 0,00 |

Tabela 2.3: Potencial Eólico para o Brasil e para as regiões Sudeste e Centro-Oeste, a 50m de altura.

Fonte: [36]

Além do potencial para grandes e médias usinas eólicas, como apresentado acima, deve-se destacar ainda o crescimento das mini e micro usinas eólicas (o limite de potência para cada categoria varia de acordo com o país; a Aneel considera microgeração até 75kW e minigeração entre 75kW e 5MW [13]). Em 2014, a capacidade mundial instalada desses pequenos aerogeradores alcançou 830.832kW, contra 443.260kW em 2010, tendo a China, os Estados Unidos e o Reino Unido como os líderes desse mercado [27].

O Brasil registrou crescimento na potência instalada de mini e microgeração eólica nos últimos anos, sobretudo nas regiões nordeste e sul. Ilustra-se esse crescimento por meio da Figura 2.12.

2.2.2.2 Tecnologias de aproveitamento

Em geral, os aerogeradores seguem o esquema ilustrado na Figura 2.13. O vento faz girar o rotor conectado a um eixo, ligado ao gerador (conversor) elétrico por meio de engrenagens, enquanto o sistema de controle visa o funcionamento seguro do sistema, e

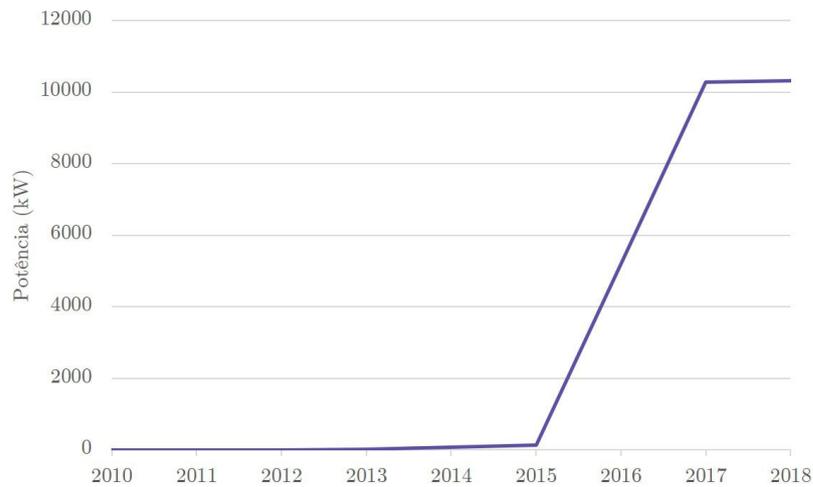


Figura 2.12: Capacidade instalada acumulada de mini e micro geradores eólicos no Brasil
Fonte: [15]

possui diversos sensores [38].

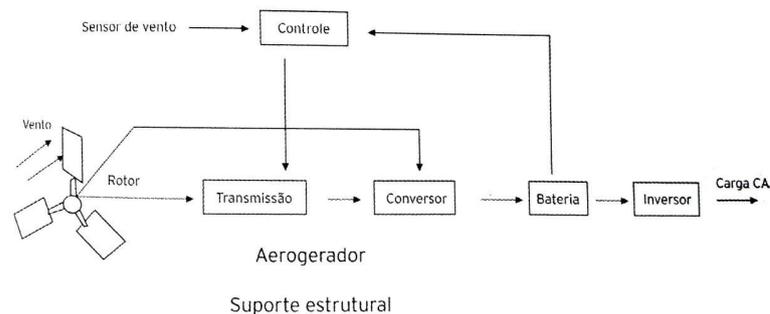


Figura 2.13: Diagrama de bloco de um sistema eólico.
Fonte: [38]

O eixo das turbinas eólicas pode ser vertical ou horizontal. As mais utilizadas são as de eixo horizontal, como o propulsor de duas ou três pás, considerados os aerogeradores mais eficientes [67]. Este tipo de rotor deve estar associado a mecanismos que façam com que a área varrida pelas pás seja sempre perpendicular à direção dos ventos [38]. Neste ponto, os aerogeradores de eixo vertical apresentam como vantagem o fato de que não precisam se adaptar à direção do vento, porém a instalação em torres altas é difícil para estes modelos, e assim, não conseguem aproveitar velocidades de vento mais elevadas [67]. Exemplos de turbinas com eixos vertical e horizontal são apresentados na Figura 2.14.

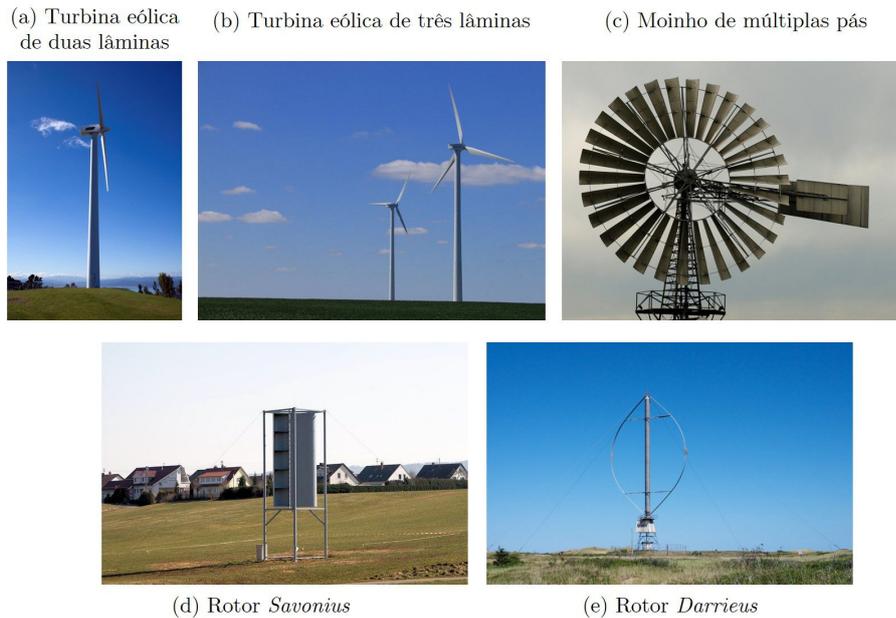


Figura 2.14: Alguns modelos de turbinas eólicas. (a), (b) e (c): modelos com eixo horizontal. (d) e (e): modelos com eixo vertical.

Fonte: (a) - [59], (b) - [58], (c) - [57], (d) - [53], (e) - [54].

A potência máxima extraída de uma corrente de vento é calculada por meio da Equação 2.2 [114].

$$P = \frac{1}{2} C_p A \rho u_0^3 \quad (2.2)$$

em que:

P = potência eólica [W]

C_p = fator de eficiência que depende do equipamento

A = área pela qual passa a coluna de vento [m^2]

ρ = densidade do vento [kg/m^3]

u_0 = velocidade do vento [m/s]

O valor máximo de eficiência esperado para C_p é 0,59 e é conhecido como “limite de Betz”. É interessante notar que, caso um aerogerador ultrapassasse esse limite e alcançasse eficiência máxima (100%), então este extrairia toda a energia cinética e o vento, após passar pelas lâminas, teria velocidade zero, fazendo com que o ar se acumulasse [67].

Na prática, os aerogeradores possuem as chamadas velocidades de *cut-in* e de *cut-out*. A primeira indica a qual velocidade do vento a turbina é, de fato, capaz de iniciar sua operação e a segunda marca o limite de velocidade do vento para qual o aerogerador se mantém em funcionamento, já que ventos muito fortes podem danificar os equipamentos, e assim, o sistema deve se proteger de avarias [38]. Esse comportamento é ilustrado na Figura 2.15, em que V_m é a velocidade mínima (*cut-in*), V_n é a velocidade nominal (velocidade máxima aproveitável, de acordo com as características do aerogerador) e V_c é a velocidade de corte (*cut-out*).



Figura 2.15: Geração de potência de um aerogerador.
Fonte: [2]

Considerando a sazonalidade do regime de ventos, como ilustrado na Figura 2.11, e a dependência cúbica que a potência de um aerogerador possui com relação à velocidade dos ventos (Equação 2.2), tem-se uma geração eólica extremamente variável ao longo do ano. Na Figura 2.16, a seguir, apresenta-se a geração elétrica da usina eólica de Gargaú, no litoral norte do estado no período de novembro de 2010 a agosto de 2016.

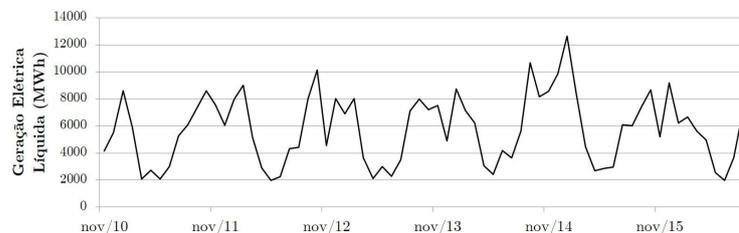


Figura 2.16: Geração elétrica líquida da usina eólica de Gargaú.
Fonte: [43]

2.2.2.3 Impactos associados à geração eólica.

De acordo com [38], os impactos de parques eólicos são muito menores comparados ao de termelétricas ou grandes hidrelétricas, e, segundo ele, os principais são

- ruído mecânico: tem origem em equipamentos do aerogerador, mas pode ser contornado através de utilização de engrenagens silenciosas, estruturas mais resistentes e proteção acústica;
- ruído aerodinâmico: decorrente de interação entre o vento e as pás do rotor. Pode ser minimizado fazendo-se com que as turbinas girem com velocidade de rotação menor para velocidades de vento baixas;
- interferência eletromagnética: causada pelo desvio de ondas eletromagnéticas (como de rádio e televisão) pelas pás do rotor. Pode ser mitigado através da escolha do material das pás e do formato da torre de acordo com o contexto de comunicações do local de instalação;
- colisão com pássaros: apesar de não ser tão comum, alerta para a construção de parques eólicos fora da rota de migração das espécies;
- impacto visual nas paisagens: depende das características da turbina e do contexto sociocultural de sua instalação.

2.2.3 Energia solar

Nesta subseção, refere-se à geração solar, com enfoque na geração fotovoltaica. Apresenta-se os mecanismos básicos dessa forma de geração de eletricidade, o potencial de geração fotovoltaica no Brasil e nos estados do SE/CO e alguns dos impactos associados a essa fonte.

2.2.3.1 O recurso e seu potencial

As reações de fusão nuclear que ocorrem no interior do Sol fazem com que parte de sua massa se converta em energia, aumentando sua temperatura e emitindo radiação, que após alcançar a superfície da estrela, se propaga pelo vácuo com distribuição espectral razoavelmente contínua, alcançando outros corpos celestes a sua volta [114] [62].

A radiação solar recebida pela atmosfera terrestre é influenciada, dentre outros fatores, pela órbita elíptica da Terra em torno do Sol, que faz com que a radiação varie ao longo

do ano, e também, pelo ângulo de inclinação do eixo da Terra com relação a essa órbita, que leva a uma variação da radiação recebida em cada região do planeta [79].

Ilustra-se a variação da radiação solar global (soma das radiações direta e difusa, que serão apresentadas à frente) entre as estações do ano no Brasil na Figura 2.17.

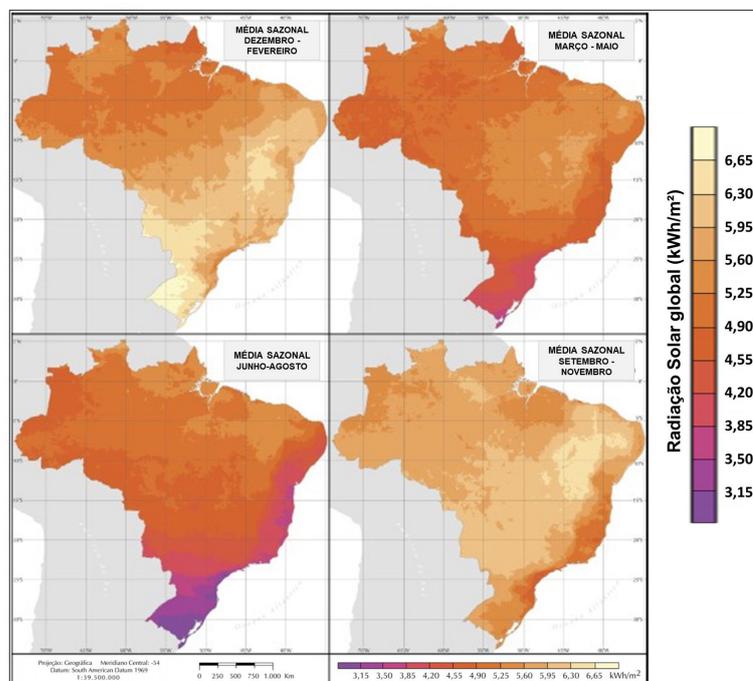


Figura 2.17: Radiação solar global média para cada estação no Brasil.
Fonte: [97]

Da radiação que alcança a atmosfera da Terra, aproximadamente 31% são refletidos ainda nas camadas mais externas da atmosfera, 17,4% são absorvidos pela atmosfera, 4,2% são refletidos próximos à superfície do planeta e, dessa forma, cerca de 47,4% alcançam realmente a superfície de continentes e oceanos [79]. Esta parcela que chega à superfície se divide em radiação direta, difusa e refletida, como verifica-se na Figura 2.18. A radiação direta é aquela que veio diretamente do sol, sem ter tido sua direção influenciada; radiação difusa é a que chega de forma indireta, sofrendo difusão na atmosfera, por partículas sólidas ou pelas nuvens, por exemplo; a radiação pode ainda ser refletida por superfícies, especialmente em áreas próximas ao mar e à neve [79] [62].

Para se converter a radiação solar em eletricidade, as tecnologias mais utilizadas atualmente são os sistemas heliotérmicos (termossolares) e os fotovoltaicos, que podem integrar grandes usinas ou podem ser usados em menor escala, como painéis instalados em telhados residenciais [38].

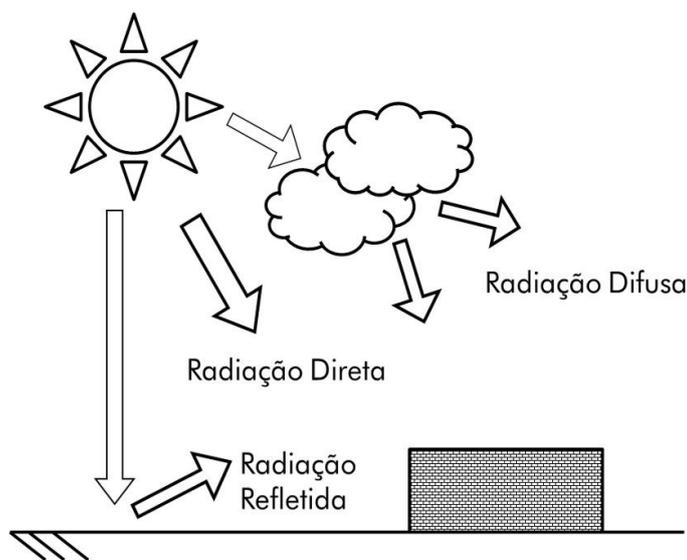


Figura 2.18: Radiação direta, difusa e refletida.

Fonte: Autora

Um estudo da Empresa de Pesquisa Energética (EPE) buscou estimar o potencial fotovoltaico residencial brasileiro, considerando a irradiação solar incidente e a área de telhados em cada região [46]. Os resultados indicaram que, em todos os estados, há potencial para suprir completamente a demanda residencial (baseado no consumo de 2013). As estimativas para o Brasil e para os estados das regiões Sudeste e Centro-Oeste são apresentados na Tabela 2.4.

| UF | Potencial Fotovoltaico Residencial (MW médios) | Potencial de Geração Consumo Residencial |
|--------|---|---|
| SP | 7.100 | 160% |
| MG | 3.675 | 318% |
| RJ | 2.685 | 183% |
| ES | 595 | 236% |
| GO | 1.220 | 270% |
| MT | 570 | 229% |
| MS | 505 | 282% |
| DF | 410 | 164% |
| Brasil | 32.820 | 230% |

Tabela 2.4: Potencial de geração fotovoltaica residencial.

Fonte: [46]

2.2.3.2 Tecnologias de Aproveitamento

Os sistemas heliotérmicos concentram a irradiação solar, convertendo-a em calor e, posteriormente, este calor é convertido em eletricidade, através de uma termelétrica ou máquina a vapor [38][86]. Para o bom aproveitamento desses sistemas, é necessária alta incidência de radiação direta, ou seja, poucas nuvens e baixos níveis pluviométricos [10]. Os sistemas fotovoltaicos convertem diretamente a radiação solar em eletricidade através do chamado efeito fotovoltaico, ocorrido quando alguns materiais semicondutores geram corrente elétrica ao incidir sobre eles radiação solar [62]. Apesar da geração de eletricidade aumentar conforme a intensidade da luz, os painéis fotovoltaicos também convertem energia em dias nublados [10].

Este trabalho foca na geração fotovoltaica, pois sua tecnologia de aproveitamento se encontra mais difundida no Brasil e também devido à facilidade da instalação de painéis fotovoltaicos em edifícios, não exigindo áreas desocupadas.

O esquema básico de funcionamento de um sistema fotovoltaico é apresentado por meio da Figura 2.19.

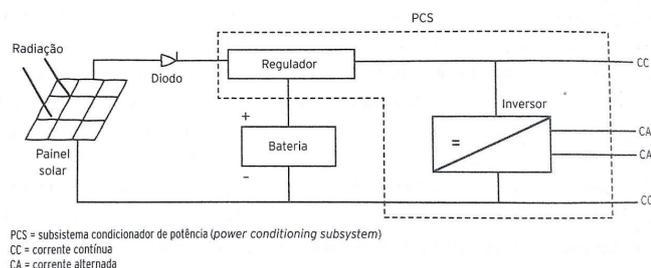


Figura 2.19: Diagrama de bloco de um sistema de geração elétrica fotovoltaico.

Fonte: [38]

Células fotovoltaicas são formadas por duas camadas de semicondutores (uma positiva e outra negativa), pois quando um material semicondutor, como o silício, absorve fótons, seus portadores de carga negativos e positivos se separam e, dado que a junção das camadas de semicondutores gera um campo elétrico, a presença deste faz com que uma corrente elétrica seja gerada e possa ser coletada [79] [114] [86].

As células, para obter-se a potência desejada, podem ser associadas em módulos (painéis), que serão sustentados por uma estrutura de madeira, aço ou alumínio e que pode contar com um dispositivo que rastreie o movimento do sol, reorientando o sistema de acordo com a incidência dos raios solares [38].

A potência de saída de um painel fotovoltaico depende da radiação incidente e da temperatura da célula, pois estas variáveis alteram a curva corrente-tensão do dispositivo [79]. A influência da temperatura na eficiência de uma célula fotovoltaica depende do material de composição das células e da montagem de todo o sistema [107]. De forma simplificada, a potência de saída de um sistema fotovoltaico é calculada por meio da Equação 2.3 [38][68]

$$P = \eta AR_S \quad (2.3)$$

em que:

P = potência gerada [W]

η = rendimento total do sistema

A = área do painel [m^2]

R_S = radiação solar incidente sobre o painel [W/m^2]

2.2.3.3 Impactos associados à geração solar fotovoltaica

Na fase de operação, sistemas fotovoltaicos têm impactos quase nulos, já que não têm emissões de poluentes, nem geram ruído [114]. Na fase de fabricação dos módulos, os impactos dependem das tecnologias utilizadas. Placas de silício cristalino têm impactos baixos, além de consumirem poucos recursos considerados escassos; as que utilizam cádmio e telúrio trazem alerta quanto à grande quantidade de componentes tóxicos; placas CIS (cádmio-índio-gálio) podem gerar gases tóxicos, como o seleneto de hidrogênio [79]. Quanto ao consumo de energia na fase de produção comparado à operação, este demora de 3 a 4 anos para ser compensado, dependendo das condições de operação, e tal *energy payback* é satisfatório, já que a vida útil estimada para um painel fotovoltaico é em torno de 20 anos [114].

Após esta breve introdução sobre as gerações hidrelétrica, eólica e fotovoltaica, a seção seguinte tratará de que forma essas fontes são inseridas em sistemas elétricos.

2.3 Inserção de Fontes Renováveis no Sistema Elétrico

O Sistema Interligado Nacional (SIN) cuida da produção e transmissão de quase toda energia elétrica no Brasil. Ele busca aumentar a segurança do sistema elétrico, permitindo

envio de energia de uma região à outra, a fim de suprir geração insuficiente em um local e/ou problemas na rede de transmissão. Dessa forma, aumenta-se a probabilidade da demanda nacional ser atendida, ou seja, eleva-se a confiabilidade do sistema como um todo.

Como dificuldades do SIN, podem-se citar atrasos na expansão de linhas de transmissão, que fazem com que algumas usinas, especialmente eólicas, fiquem desconectadas da rede e, ademais, existem muitas usinas termelétricas obsoletas, instaladas nas décadas de 60 e 70, sendo que algumas chegam a operar com eficiências abaixo de 20% [29].

Por muitos anos, o SIN se caracterizou como um sistema hidrotérmico, com predominância de geração hidrelétrica. No entanto, de acordo com [91], novas usinas hidrelétricas com reservatórios que permitem regularização plurianual não ingressam no sistema desde a década de 90, fazendo com que a geração termelétrica tenha aumentado significativamente nos últimos anos. Além disso, o documento também manifesta preocupação com o início do funcionamento de usinas hidrelétricas de grande porte a fio d'água (geração com forte sazonalidade) na bacia Amazônica e com a expectativa de um aumento considerável na geração eólica pelos próximos cinco anos, destacando sua alta intermitência e a baixa previsibilidade [91].

O autor de [31] expõe que a intermitência e imprevisibilidade da geração de parques eólicos torna necessário que o restante do sistema atue com maior flexibilidade, já que o custo baixo da geração eólica torna o seu despacho prioritário e, assim, as demais usinas podem se submeter a mais partidas e paradas e, eventualmente, podem ser utilizadas com eficiência não otimizada. O trabalho cita um caso ocorrido na Espanha, onde a expansão mal programada do parque eólico reduziu o fator de capacidade de usinas tradicionais que, ao operarem fora do ponto ótimo, levaram a um aumento da tarifa da eletricidade [31]. Desse modo, muitos estudos vêm sendo realizados para avaliar qual grau de penetração eólica mantém a segurança do sistema e minimiza os custos, como será tratado no Capítulo 3.

Os autores de [88] apresentam diversas estratégias para a utilização de energia eólica em larga escala, como interligação com outras redes, geração distribuída, armazenamento de energia e uso de complementaridade entre fontes renováveis.

Quanto à interligação, o SIN brasileiro é considerado único dado seu tamanho e suas características [90] o que aumenta sua capacidade em mitigar os efeitos da intermitência da geração eólica [31] e das demais fontes renováveis. Atualmente para compensar as usinas eólicas, são utilizadas usinas hidrelétricas do Controle Automático de Geração, mas, com o

aumento da participação de fontes intermitentes, o ONS estima que, além destas, podem ser utilizadas usinas térmicas de rápida tomada de carga, ainda não incorporadas ao sistema elétrico brasileiro [91].

No que concerne ao armazenamento de energia, [34] realizou um estudo comparativo entre doze tecnologias existentes, considerando o impacto ambiental, a vida útil, o custo de investimento, a eficiência, o tempo de resposta e a capacidade de cada uma. A análise contemplou as perspectivas ambiental, econômica e técnica da sustentabilidade e as tecnologias mais bem avaliadas foram as de bombeamento de água (usinas reversíveis), super capacitor e super condutividade, respectivamente. O trabalho ainda sugeriu a utilização de reservatórios já existentes de hidrelétricas como armazenamento como nas usinas reversíveis, mas com a vantagem de não haver gasto de energia para bombear, e apontou que a porcentagem de volume útil e de energia gastos por usinas hidrelétricas para compensar a variabilidade eólica não são expressivos [34].

A variação temporal dos recursos renováveis se dá de maneira ímpar para cada fonte e para cada localidade. Desse modo, o planejamento de operação das usinas poderá ser prejudicado se a sazonalidade de geração das usinas que utilizam as fontes intermitentes não for bem compreendida. Em contrapartida, se estudos indicarem que vários recursos, ou ainda o mesmo recurso em locais diferentes, se complementam ao longo do dia ou do ano, então o planejamento energético poderá aproveitar essa complementaridade e considerar a operação conjunta de geradores a fim de diminuir a variabilidade da geração total. A análise de complementaridade é tema deste trabalho e será detalhada no Capítulo 3.

Dois métodos que serão utilizados para o estudo de complementaridade deste trabalho são a correlação linear e a análise de fatores. As próximas seções apresentam uma breve fundamentação teórica desses temas.

2.4 Correlação Linear

Diz-se que duas variáveis têm correlação perfeita quando todos os valores delas satisfazem uma equação que as relaciona, de modo que, ao observar as variáveis em um gráfico de dispersão, percebe-se que estas se aproximam da curva expressa por tal equação [109]. Nos casos em que há correlação linear, a curva que relaciona as variáveis é uma reta ($Y = aX + b$, com a e b constantes). Se a correlação for positiva, a reta terá inclinação positiva e, se a correlação for negativa, a reta terá inclinação negativa. Exemplos de gráficos de dispersão de variáveis com correlação linear positiva, correlação linear negativa e

nenhuma correlação são apresentados na Figura 2.20.

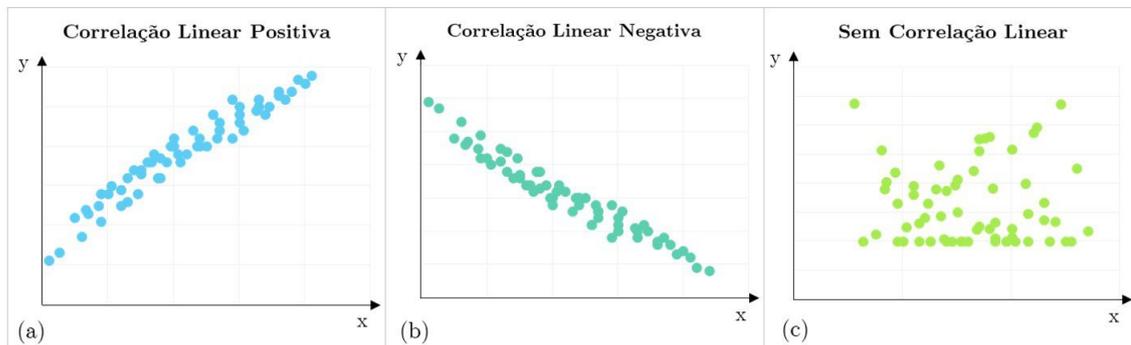


Figura 2.20: Gráficos de dispersão de variáveis (a) com correlação linear positiva; (b) com correlação linear negativa; (c) sem correlação linear

Fonte: Autora

No caso de séries de dados como as deste estudo, pode-se observar melhor o conceito de correlação linear por meio da Figura 2.21, que mostra as variáveis em ordem cronológica e os mesmos pontos apresentados em três diagramas de dispersão, cada um relacionando uma série à variável base.

Foi dito que existe correlação perfeita quando todos os valores das variáveis se ajustam a uma curva, porém quando nem todos os valores seguem a equação, surgem variações no valor de uma das variáveis que não podem ser explicadas por variações da outra. Quando o interesse em duas variáveis não reside em usar uma para prever a outra, mas apenas em saber se há ou não relação linear entre elas e se esta relação é fraca ou forte, recorre-se ao cálculo do coeficiente de correlação linear ou coeficiente de correlação de Pearson (ρ) para uma população, calculado por meio da Equação 2.4.

$$\rho = \frac{Cov(X, Y)}{\sigma_x \sigma_y} \quad (2.4)$$

em que:

$Cov(X, Y)$ = covariância entre as variáveis aleatórias X e Y

σ_x = desvio-padrão de X

σ_y = desvio-padrão de Y

Quando não se tem acesso aos dados de todos os indivíduos de cada população, mas sim, amostras dessas populações, a Equação 2.4 pode ser adaptada partindo do cálculo da covariância e dos desvios amostrais. A covariância de uma amostra ($cov(X, Y)$) é calculada por meio da Equação 2.5.

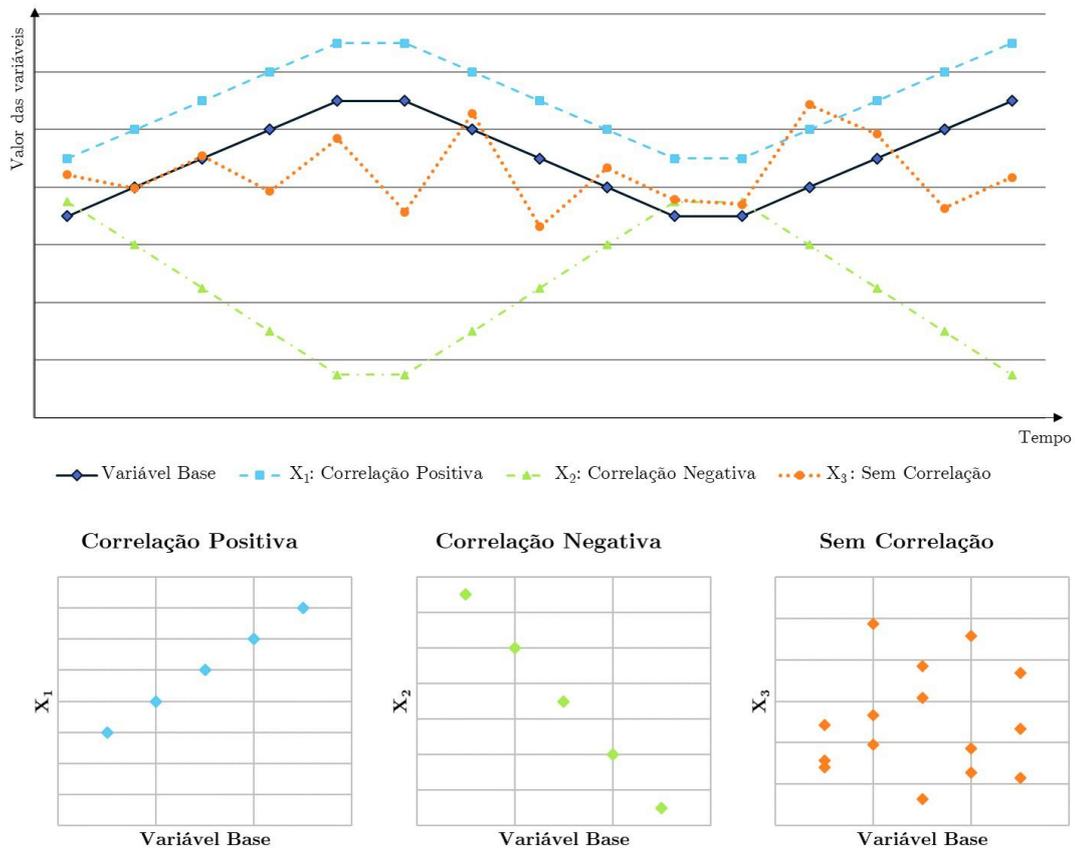


Figura 2.21: Exemplos de séries com correlação linear positiva, correlação linear negativa e sem correlação linear com respeito a uma variável base (acima); e as mesmas séries plotadas como diagramas de dispersão

Fonte: Adaptado de [100]

$$cov(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \tag{2.5}$$

em que:

- x_i = observações da primeira amostra
- y_i = observações da segunda amostra
- \bar{x} = média das observações da primeira amostra
- \bar{y} = média das observações da segunda amostra
- n = número de observações nas duas amostras

Já o desvio padrão de uma amostra é calculado por meio da Equação 2.6.

$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \tag{2.6}$$

Dessa forma, substituindo as Equações 2.5 e 2.6 na Equação 2.4, obtemos a forma de cálculo do coeficiente de correlação amostral (r), como expresso na Equação 2.7.

$$\begin{aligned}
 r &= \frac{\text{cov}(X, Y)}{s_x s_y} \\
 &= \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}} \\
 &= \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}{n-1}} \\
 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \\
 \therefore r &= \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}} \tag{2.7}
 \end{aligned}$$

em que:

$$\begin{aligned}
 S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\
 S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2 \\
 S_{yy} &= \sum_{i=1}^n (y_i - \bar{y})^2
 \end{aligned}$$

Diz-se que duas variáveis têm relação positiva quando valores elevados de uma delas coincidem com valores elevados da outra. Por outro lado, nas variáveis relacionadas de forma negativa, quando uma delas apresenta valores mais elevados, a outra assume valores mais baixos e vice-versa. A formulação do coeficiente de correlação amostral é muito útil para a compreensão da maneira como as relações entre as variáveis são absorvidas na análise.

Observando a quantidade S_{xy} , que corresponde ao numerador da Equação 2.7, percebe-se que quando a relação entre x e y é positiva, se x_k é maior do que \bar{x} , y_k também tende a ser maior que \bar{y} . Da mesma forma, se x_k é menor do que \bar{x} , y_k provavelmente será menor que \bar{y} . Assim, o produto $(x_i - \bar{x})(y_i - \bar{y})$ tende a ser positivo para a maior parte dos pares (x_i, y_i) , pois se o fator relacionado a x for positivo, o de y também deverá ser, resultando em um produto positivo. Por outro lado, se o fator de x for negativo, o de y

provavelmente também o será e, novamente, o produto será positivo.

De forma análoga, conclui-se que, se a relação entre x e y for negativa, o produto $(x_i - \bar{x})(y_i - \bar{y})$ tem propensão a ser sempre negativo, resultando em um S_{xy} negativo.

No entanto, se x e y não estão correlacionadas, o produto $(x_i - \bar{x})(y_i - \bar{y})$ ora será positivo, ora negativo, de tal forma que no somatório, um produto tenderá a anular outro e o resultado de S_{xy} será próximo de zero.

Há de se ressaltar, porém, que S_{xy} é extremamente sensível às unidades usadas. Então, a fim de favorecer a interpretação e a comparação da medida de correlação de duas amostras, o termo do denominador da Equação 2.7 é inserido.

Algumas propriedades importantes de r são [35]:

1. Havendo duas amostras, a escolha de uma ou outra como x (ou y) não afeta o valor de r ;
2. O valor de r independe das unidades de x e y , ou seja, ele não é alterado se todas as observações das variáveis forem multiplicadas por alguma constante ou se tiverem uma constante acrescida ao seu valor;
3. O valor absoluto de r não poderá ser maior do que uma unidade ($-1 \leq r \leq 1$).

Da terceira propriedade de r (ρ) e da interpretação de S_{xy} , pode-se compreender que quando duas variáveis têm correlação perfeitamente positiva, r é igual a 1 e, quando a correlação negativa é perfeita, r é igual a -1 .

Correlações perfeitas (estatisticamente) são esperadas entre variáveis relacionadas de forma determinística, ou seja, que têm relação de dependência tão clara que o conhecimento de uma das variáveis possibilita o cálculo do valor exato da outra, como na relação entre deslocamento e tempo, em problemas de movimento uniforme na Física, ou entre o resultado de vendas e receita, na área financeira. No entanto, tratando-se de variáveis relacionadas de maneira não determinística, pode haver uma relação de dependência, porém a proporção que relaciona as variáveis nem sempre é igual, pois outros fatores as influenciam.

Não há consenso entre os autores no que diz respeito a como mensurar o quão fortemente duas variáveis com relação não determinística estão ligadas. Neste estudo, basearemos a análise na interpretação proposta por [35], expressa no Quadro 2.1.

| Valor absoluto de r (ρ) | Interpretação |
|----------------------------------|----------------------------|
| $0 \leq r \leq 0,5$ | Correlação Linear Fraca |
| $0,5 < r < 0,8$ | Correlação Linear Moderada |
| $ r \geq 0,8$ | Correlação Linear Forte |

Quadro 2.1: Interpretação dos valores do coeficiente de correlação linear

É interessante notar que r^2 representa a porcentagem da variabilidade de uma variável explicada pela relação linear entre elas. Isto é, se $r = 0,5$, então $r^2 = 0,25$ e 25% da variabilidade de uma variável é explicada pela correlação linear com a outra. Com $r = 0,8$, a porcentagem sobe para 64% [35]. Outra ressalva importante é que a inexistência de correlação linear ($r \approx 0$), não implica que as variáveis não estão correlacionadas. É possível que a curva que as relaciona não seja uma reta, mas uma função diferente que poderá ser investigada.

2.5 Análise de Fatores

A Análise de Fatores, também conhecida como Análise Fatorial, é um método mais aprofundado para avaliar como algumas variáveis se relacionam entre si ¹. O método não só é usado para compreender a estrutura de variância de um conjunto de variáveis, mas também pode ser aplicado para reduzir o conjunto original de variáveis, para evitar redundância, prover uma análise mais parcimoniosa ou simplesmente viabilizar a realização de alguns estudos posteriores.

Essa análise é conduzida quando acredita-se que as variáveis consideradas dependem linearmente de outras variáveis que não puderam ser observadas, que são denominadas *fatores* [112]. Dessa forma, as variáveis que dependem fortemente de um fator específico, tendem a ter uma parcela significativa de variância em comum, inerente ao fator. Logo, busca-se encontrar variáveis que tenham alta correlação entre si e baixa com as demais [75].

Um esquema de análise fatorial com dois fatores e quatro variáveis é apresentado na Figura 2.22.

¹Há dois tipos de análises de fatores: a análise R, que agrupa variáveis, e a análise Q que agrupa casos (respondentes), de forma similar a técnicas de agrupamento [52]. A análise do tipo R é a de interesse deste estudo e, daqui por diante, a utilização do termo "análise de fatores", se referirá à análise do tipo R.

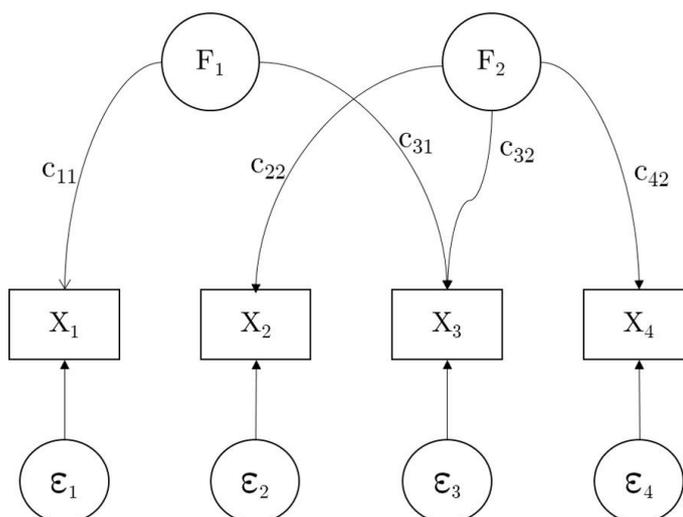


Figura 2.22: Esquema gráfico de uma análise de fatores com quatro variáveis observáveis (X_i , representadas por quadrados) e seis não observáveis (círculos), sendo duas delas fatores (F_j) comuns às variáveis, e as demais, componentes específicos de cada variável (ϵ_i). Os valores c_{ij} representam a influência do fator j na variável i .

Fonte: Adaptado de [105]

A essência da análise de fatores reside em encontrar padrões entre as variáveis e, para isso, é necessário que haja correlação suficiente entre elas [52]. Assim, alguns testes devem ser feitos antes de se dar início à análise, para saber se há de fato evidências da existência de fatores subjacentes ao conjunto de variáveis.

2.5.1 Pré-análise

Antes de se conduzir a análise de fatores, algumas considerações devem ser feitas. Segundo [64], para a análise de fatores, o número de observações não deve ser menor do que cinco vezes o número de variáveis, mas o ideal é que haja pelo menos dez vezes mais observações que variáveis. Um bom balanceamento entre o número de variáveis e o de observações ajuda a evitar que o resultado da análise fique superajustado à amostra e não corresponda à realidade populacional [64].

Além de avaliar a qualidade da amostra em termos do número de observações, deve-se avaliar também as correlações entre as variáveis. Como a análise de fatores buscará encontrar grupos de variáveis bem correlacionadas entre si, devem existir correlações significativas pelo menos entre algumas das variáveis. Para isso, pode-se realizar os seguintes procedimentos [64][8][116][25][1]:

2.5.1.1 Avaliar as correlações lineares (e/ou correlações parciais) entre as variáveis.

Um número substancial de correlações acima de 0,3 é indício de que a análise fatorial pode ser apropriada, apesar de não ser condição suficiente por si só, pois é um método um tanto subjetivo, envolvendo inspeção visual.

2.5.1.2 Realizar o teste de esfericidade de Bartlett.

O teste parte da hipótese nula de que a matriz de covariância é a matriz identidade, ou seja, de que não há correlações significantes entre as variáveis (elementos fora da diagonal principal). A hipótese alternativa (H_a) é a de que a matriz de covariância não é a matriz identidade. Para a análise de fatores ser considerada apropriada, a hipótese nula deve ser rejeitada. A estatística de teste foi definida por Bartlett como na Equação 2.8.

$$\chi^2 = -(n - 1 - \frac{2p + 5}{6}) \ln|R| \quad (2.8)$$

em que:

χ^2 = estatística do teste, que segue uma distribuição qui-quadrado com $\frac{p(p-1)}{2}$ graus de liberdade

n = número de observações

p = número de variáveis

R = matriz de correlação das p variáveis

Em um teste de hipóteses, o valor da estatística de teste é calculado a partir dos dados que estão sendo avaliados (parâmetro calculado : θ) e comparado com o valor gerado considerando a hipótese nula verdadeira (parâmetro nulo : θ_0). No caso em que H_0 é verdadeira, a matriz de correlação é uma matriz identidade e, conseqüentemente, seu determinante ($|R|$) é 1. Desse modo, o parâmetro nulo testado será igual a zero, pois o $\ln|R| = \ln(1) = 0$. Por outro lado, conforme as correlações entre as variáveis vão se tornando mais fortes, o determinante da matriz de correlação se aproxima de zero, fazendo com que o valor do parâmetro calculado aumente exponencialmente, pois $\lim_{x \rightarrow 0} \ln(x) = -\infty$. Ou seja, quanto mais bem correlacionada forem as variáveis, mais o parâmetro calculado se distanciará do parâmetro nulo (zero).

Logo, para o teste de esfericidade de Bartlett, as hipóteses são:

$$H_0 : \theta = \theta_0 = 0$$

$$H_a : \theta \neq 0$$

Geralmente realiza-se o teste de hipótese calculando o valor-p ao invés de realizá-lo para um nível de confiança $(1-\alpha)$ pré-determinado². O valor-p (ou *p-value*, em inglês) é a probabilidade de se obter uma estatística de teste tão grande ou maior do que a calculada, caso H_0 seja verdadeira. Ou seja, se p for grande, a estatística calculada não estaria tão distante do valor do parâmetro assumido na hipótese nula. Por outro lado, valores pequenos de p, favorecem H_a em detrimento a H_0 . Essa relação é ilustrada na Figura 2.23. Percebe-se que para diminuir o valor-p, o parâmetro calculado se desloca para a direita, se afastando do parâmetro nulo.

Comparando o valor-p com a abordagem do nível de significância, aceita-se H_0 se o valor-p é maior do que o nível de significância e rejeita-se H_0 , se o valor-p for menor do que o nível de significância.

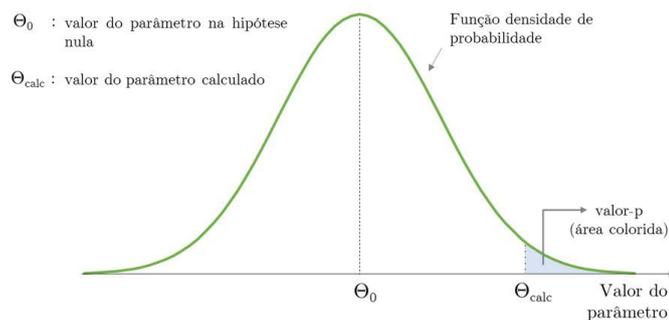


Figura 2.23: Relação entre o valor-p, o parâmetro nulo testado e o parâmetro calculado com base na amostra.

Fonte: Adaptado de [35]

Um valor-p menor ou igual a 0,05 é geralmente considerado suficiente para o prosseguimento da análise.

2.5.1.3 Calcular a medida de adequação da amostra (MSA) e o coeficiente de Kaiser-Meier-Olkin (KMO)

O KMO é um indicador geral, que se referencia a todo o conjunto de variáveis, enquanto a MSA gera valores para cada uma das variáveis.

A MSA é calculada por meio da Equação 2.9.

²Neste caso, α é chamado "nível de significância" e corresponde à probabilidade de se rejeitar H_0 sendo ela verdadeira.

$$MSA_j = \frac{\sum_{\substack{k \\ k \neq j}} r_{jk}^2}{\sum_{\substack{k \\ k \neq j}} r_{jk}^2 + \sum_{\substack{k \\ k \neq j}} q_{jk}^2} \quad (2.9)$$

em que:

r_{jk} = correlação linear entre as variáveis j e k

q_{jk} = correlação parcial entre as variáveis j e k

A correlação parcial busca identificar se um par de variáveis está mesmo correlacionado ou se a aparente correlação indicada pelo coeficiente de Pearson é fruto da influência de outra variável.

A correlação parcial entre duas variáveis é calculada por meio da Equação 2.10.

$$q_{jk} = \frac{-b_{jk}}{\sqrt{b_{jj}b_{kk}}} \quad (2.10)$$

em que:

b_{jk} = elemento (j,k) da inversa da matriz de correlação

b_{jj} e b_{kk} = elementos da diagonal principal da matriz de correlação inversa

Para calcular a matriz de correlações parciais, deve-se dividir cada elemento da matriz inversa à de correlação pela raiz quadrada dos elementos da matriz formada pelo produto entre o vetor ($d_{n \times 1}$) formado pelos elementos da diagonal principal da matriz inversa e sua versão transposta ($d_{1 \times n}^T$): $\mathbf{d}\mathbf{d}^T$.

Pode-se perceber que as diagonais das matrizes de correlação e de correlação parcial não são contabilizadas nos somatórios do cálculo do indicador. Deve-se avaliar a MSA de cada variável e eliminar aquelas que estão abaixo do aceitável (0,5) uma por vez e repetir o cálculo do indicador até que todas as variáveis tenham valores aceitáveis [52].

O KMO é uma generalização do MSA que engloba todas as variáveis, como mostrado na Equação 2.11.

$$KMO = \frac{\sum_{\substack{j \\ j \neq k}} \sum_{\substack{k \\ j \neq k}} r_{jk}^2}{\sum_{\substack{j \\ j \neq k}} \sum_{\substack{k \\ j \neq k}} r_{jk}^2 + \sum_{\substack{j \\ j \neq k}} \sum_{\substack{k \\ j \neq k}} q_{jk}^2} \quad (2.11)$$

A interpretação dos valores do KMO proposta por [78] é apresentada no Quadro 2.2.

| Valor do KMO | Interpretação |
|----------------------|---------------|
| $KMO < 0,5$ | Inaceitável |
| $0,5 \leq KMO < 0,6$ | Mísero |
| $0,6 \leq KMO < 0,7$ | Medíocre |
| $0,7 \leq KMO < 0,8$ | Moderado |
| $0,8 \leq KMO < 0,9$ | Admirável |
| $KMO > 0,9$ | Ótimo |

Quadro 2.2: Interpretação dos valores do índice KMO.

2.5.2 Formulação de uma Análise de Fatores

Como introduzido anteriormente, a análise de fatores descreve cada variável X_i ($i = 1, \dots, n$) como uma combinação linear de m variáveis não observadas, denominadas fatores (F_j), como o modelo apresentado na Equação 2.12 [75].

$$X_i - \mu_i = \sum_{j=1}^m c_{ij} F_j + \varepsilon_i \quad (2.12)$$

em que:

X_i = variáveis observadas

μ_i = média da variável i

F_j = fatores

c_{ij} = carga da variável i associada ao j -ésimo fator

ε_i = erro ligado à variável i

n = número de variáveis

m = número de fatores

É necessário à análise fatorial encontrar o valor de todas cargas. A técnica de regressão linear não pode ser utilizada porque os fatores F_j não são observáveis [112]. Deve-se buscar um método diferente e, para isso, algumas condições precisam ser estabelecidas.

Inicialmente, os fatores F_j devem ser considerados independentes (modelo ortogonal), porém modelos avançados admitem fatores correlacionados. Além disso, por conveniência matemática, define-se que a média de cada fator é nula e a variância é unitária [112]. De forma semelhante, os erros têm média nula e a variância de cada ε_i é representada por φ_i . Os erros e os fatores também são independentes, ou seja, $Cov(\varepsilon, F) = 0$ [75].

Logo, a variância de cada X_i ($V(X_i)$) pode ser calculada por meio da Equação 2.13.

$$V(X_i) = V\left(\mu_i + \sum_{j=1}^m c_{ij}F_j + \varepsilon_i\right) \quad (2.13)$$

Como na Equação 2.13 tem-se a variância de uma combinação linear de variáveis independentes, ela pode ser reescrita como a Equação 2.14.

$$V(X_i) = V(\mu_i) + V\left(\sum_{j=1}^m c_{ij}F_j\right) + V(\varepsilon_i) \quad (2.14)$$

E, de modo semelhante, a Equação 2.14 é equivalente à Equação 2.15.

$$V(X_i) = V(\mu_i) + V(c_{i1}F_1) + V(c_{i2}F_2) + \dots + V(c_{im}F_m) + V(\varepsilon_i) \quad (2.15)$$

No entanto, μ_i e c_{ij} são termos constantes, e sabe-se que a variância de uma constante é nula e a variância do produto entre uma contante e uma variável é igual à constante ao quadrado multiplicada pela variância da variável, então reescreve-se a Equação 2.14 como a Equação 2.16.

$$V(X_i) = 0 + c_{i1}^2 V(F_1) + c_{i2}^2 V(F_2) + \dots + c_{im}^2 V(F_m) + V(\varepsilon_i) \quad (2.16)$$

Definiu-se $V(F_j) = 1 \forall j \in [1, m]$ e $V(\varepsilon_i) = \varphi_i$. Daí, vem a Equação 2.17.

$$V(X_i) = 0 + c_{i1}^2 + c_{i2}^2 + \dots + c_{im}^2 + \varphi_i \quad (2.17)$$

Dessa forma, a variância de X_i é expressa por meio da Equação 2.18.

$$V(X_i) = \sum_{j=1}^m c_{ij}^2 + \varphi_i \quad (2.18)$$

Percebe-se que a variância foi decomposta em dois termos. O primeiro ($\sum_{j=1}^m c_{ij}^2$) é a porção da variância que é explicada pelos fatores e compartilhada com outras variáveis. Esse primeiro termo é conhecido como comunalidade. Já o segundo termo é a parcela de variância exclusiva da variável i . Chamar-se-á este termo de variância específica, apesar de que alguns autores ainda segmentam φ_i em variância específica e de erro (gerada por erros na coleta de dados ou incerteza das medidas) [64].

Pode-se provar ainda que a covariância entre X_i e X_k quaisquer ($\forall i, k \in [1, n]$) é calculada por meio da Equação 2.19.

$$Cov(X_i, X_k) = \sum_{j=1}^m c_{ij}c_{kj} \quad (2.19)$$

E que a covariância entre uma variável X_i e um fator F_j é tal como apresenta na Equação 2.20.

$$Cov(X_i, F_j) = c_{ij} \quad (2.20)$$

Os fatores já têm, por definição, média nula e desvio-padrão unitário, mas caso as variáveis X_i sejam padronizadas de igual modo, as cargas fatoriais c_{ij} representarão o coeficiente de correlação linear entre a variável i e o fator j , e, conseqüentemente, o valor de cada carga ao quadrado representa a porcentagem de variância da variável i explicada pelo fator j [64].

A Equação 2.12 pode ser reescrita de forma matricial como na Equação 2.21.

$$(\mathbf{X} - \boldsymbol{\mu})_{n \times 1} = \mathbf{C}_{n \times m} \mathbf{F}_{m \times 1} + \mathbf{E}_{n \times 1} \quad (2.21)$$

em que $\mathbf{X} - \boldsymbol{\mu}$ é um vetor coluna com as diferenças $(X_i - \mu_i)$; \mathbf{C} é a matriz de cargas; \mathbf{F} é um vetor coluna com os fatores; e \mathbf{E} é um vetor coluna com os termos ε_i .

Unindo-se as Equações 2.18 e 2.19, pode-se definir a matriz de covariância de \mathbf{X} , expressa na Equação 2.22.

$$Cov(\mathbf{X})_{n \times n} = \mathbf{C}_{n \times m} \mathbf{C}_{m \times n}^T + \boldsymbol{\Phi}_{n \times n} \quad (2.22)$$

em que \mathbf{C}^T é a matriz de cargas transposta e $\boldsymbol{\Phi}$ é uma matriz diagonal com as variâncias específicas φ_i na diagonal principal.

2.5.3 Extração de Fatores

Existem basicamente duas técnicas de extração de fatores: o método do componente principal (MCP) e o método de fatores comuns (MFC). Segundo [64], o MCP é indicado quando o objetivo da análise é obter um número mínimo de fatores a fim de reduzir os

dados, enquanto o MFC teria melhor capacidade de interpretar a estrutura subjacente às variáveis. No entanto, não há consenso sobre qual método é o mais adequado. Alguns autores defendem que o MFC tem fundamentação teórica mais sólida, porém o método não gera solução única (um único modelo pode gerar escores fatoriais diferentes) e ocasionalmente não consegue calcular comunalidades válidas para as variáveis, o que faz com que outros autores defendam fortemente o MCP [64].

Apesar desse debate, estudos têm mostrado que os resultados dos métodos são muitas vezes análogos [64] e, inclusive, quando uma análise é realizada pelos dois métodos e há concordância na estrutura fatorial resultante, tem-se um indicativo de que a análise foi feita de forma apropriada [75].

2.5.3.1 Análise de Fatores pelo Método do Componente Principal

O MCP assume inicialmente que o número de fatores (m) é igual ao número de variáveis (n) e, conseqüentemente, as variâncias específicas (φ_i) são consideradas nulas. Posteriormente, os fatores cuja participação é desprezível serão eliminados e poder-se-á admitir as variâncias específicas novamente. É costume centralizar os dados, ou, se as variáveis não houverem sido medidas nas mesmas unidades, recomenda-se padronizá-los.

De acordo com essa formulação inicial, a Equação 2.19 pode ser reescrita simplesmente como a Equação 2.23.

$$\mathbf{Cov}(\mathbf{X})_{n \times n} = \mathbf{C}_{n \times m} \mathbf{C}_{m \times n}^{\top} \quad (2.23)$$

Uma matriz de covariância é simétrica, pois $\mathit{Cov}(X_i, X_k) = \mathit{Cov}(X_k, X_i)$.

O Teorema Espectral da álgebra linear afirma que uma matriz quadrada real "será simétrica se, e somente se, for ortogonalmente diagonalizável" [98]. Em decorrência desse teorema, pode-se provar que uma matriz simétrica real \mathbf{A} pode ser escrita como na Equação 2.24, através de um processo chamado de decomposição espectral [98].

$$\mathbf{A} = \lambda_1 q_1 q_1^{\top} + \lambda_2 q_2 q_2^{\top} + \dots + \lambda_n q_n q_n^{\top} = \sum_{i=1}^n \lambda_i q_i q_i^{\top} \quad (2.24)$$

λ_i = autovalores de \mathbf{A}

q_i = autovetores de \mathbf{A}

q_i^{\top} = autovetores transpostos de \mathbf{A}

Logo, a matriz de covariância ($Cov(X)$) também pode ser descomposta dessa forma, como na Equação 2.25.

$$Cov(X) = \sum_{i=1}^n \lambda_i q_i q_i^\top = \begin{bmatrix} \sqrt{\lambda_1} \vec{q}_1 & \sqrt{\lambda_2} \vec{q}_2 & \dots & \sqrt{\lambda_n} \vec{q}_n \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \vec{q}_1 \\ \sqrt{\lambda_2} \vec{q}_2 \\ \vdots \\ \sqrt{\lambda_n} \vec{q}_n \end{bmatrix} = \mathbf{C} \mathbf{C}^\top \quad (2.25)$$

Sendo assim, cada fator j é representado por um autovalor e pode-se determinar o valor das cargas. Organizam-se os λ_i de tal modo que $\lambda_1 > \lambda_2 > \dots > \lambda_n$.

O modelo, até o momento, considera o número de fatores (m) igual ao número de variáveis (n). Entretanto, como os últimos autovalores são pequenos e explicam uma parcela pouco expressiva da variância, cria-se uma matriz de cargas aproximadas \widetilde{C}_m excluindo-se os $(n - m)$ fatores finais, como na Equação 2.26, e, conseqüentemente, o modelo passa a ter m fatores. A escolha do valor de m será discutida na Seção 2.5.4.

$$\widetilde{C}_m = \begin{bmatrix} \sqrt{\lambda_1} \vec{q}_1 & \sqrt{\lambda_2} \vec{q}_2 & \dots & \sqrt{\lambda_m} \vec{q}_m \end{bmatrix} \quad (2.26)$$

Feito isso, as variâncias específicas são readmitidas e parte da variância deixa de ser explicada pelo modelo. As variâncias específicas do novo modelo são calculadas de acordo com a Equação 2.27.

$$\tilde{\varphi}_i = V(X_i) - \sum_{j=1}^m \tilde{c}_{ij}^2 = V(X_i) - \tilde{h}_i \quad (2.27)$$

em que $\tilde{h}_i = \sum_{j=1}^m \tilde{c}_{ij}^2$ é a comunalidade das variáveis relativa ao modelo de m fatores.

A variância não explicada pelo modelo pode ser calculada através de uma matriz residual (\mathbf{R}), como mostrado na Equação 2.28.

$$\mathbf{R} = Cov(\mathbf{X}) - (\widetilde{C} \widetilde{C}^\top + \widetilde{\Phi}) \quad (2.28)$$

Já a proporção da variância total da amostra explicada pelo j -ésimo fator pode ser calculada de acordo com a Equação 2.29.

$$p_j = \frac{\sum_{i=1}^n \widetilde{c}_{ij}^2}{\sum_{i=1}^n V(X_i)} \quad (2.29)$$

Ou seja, a variância explicada pelo fator j (em porcentagem) é dada dividindo-se a soma das cargas ao quadrado de todas as variáveis associadas ao fator j pela variância total amostral (soma das variâncias amostrais de cada uma das variáveis) [75]. É interessante notar que o numerador da Equação 2.29 ($\sum_{i=1}^n \widetilde{c}_{ij}^2$) corresponde ao autovalor λ_j [75].

2.5.3.2 Método dos Fatores Comuns

O Método da Máxima Verossimilhança (MVV) é uma das abordagens mais usadas de fatores comuns e será o método discutido nesta seção. No MVV, assume-se que F_j e ε_j seguem distribuição normal conjunta e, por conta disso, as observações X_j são normalmente distribuídas.

Utiliza-se, então, a função de verossimilhança tendo a média e a covariância como parâmetros. Os estimadores das cargas e das variâncias específicas são calculados a partir da maximização da função de verossimilhança. É importante destacar que, diferentemente do método de componente principal, neste método não se assume valores iniciais para Φ , mas sim, esses valores são estimados com base na função de verossimilhança.

2.5.4 Escolha do número de fatores

Segundo [52], alguns dos critérios a serem observados para definir o número de fatores serão apresentados a seguir.

1. **Critério da Raiz Latente.** Também chamado de critério de Guttman-Kaiser. Avalia-se os autovalores correspondentes a cada fator (raízes latentes). Apesar de tratar de autovalores, a análise pode ser conduzida na análise por componente principal ou por outro método e é mais recomendado quando o número de variáveis está entre 20 e 50 [52]. Apenas os fatores com autovalor maior do que 1 são considerados significativos. Uma explicação mais intuitiva e menos formal para isso é que, quando os dados são padronizados, cada variável contribui com uma unidade à soma dos autovalores, e assim, autovalores menores do que 1 não conseguiriam explicar nem ao menos a variância de uma das variáveis [120][75].
2. **Critério a priori.** Para determinar o número de fatores a serem extraídos, deve-se avaliar o alicerce teórico, ou seja: examinando o caso, quantos fatores são esperados

e quantos podem ser sustentados? Por mais que não haja uma resposta exata para esses questionamentos, deve-se buscar interpretar o caso com base nas informações disponíveis.

- 3. Critério da Variância Explicada.** A porcentagem mínima de variância explicada desejada também pode ser um indicador muito bom do número de fatores necessários. A porcentagem mínima aceitável depende da área de aplicação. Geralmente o nível aceitável é mais baixo na área de ciências sociais, e mais alto para ciências naturais [52].
- 4. Critério do teste *scree*.** À medida que os autovalores vão sendo extraídos, diminui-se a parcela de variância comum e a variância específica começa a dominar os autovalores. Dessa forma, é de interesse selecionar os fatores que tenham mais variância comum do que específica. O gráfico *scree* plota os autovalores e o índice de cada um, de acordo com a ordem de extração. Busca-se o ponto em que há uma “dobra” no gráfico e a partir do qual a curva fique aproximadamente paralela ao eixo horizontal [75][52]. Um exemplo de gráfico *scree* é apresentado na Figura 2.24.

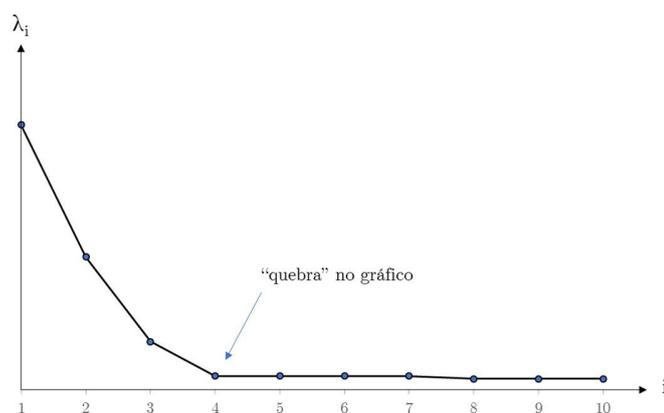


Figura 2.24: Exemplo de gráfico *scree*. Neste caso, a recomendação seria usar 4 fatores.
Fonte: Autora

Geralmente, utiliza-se mais de um método para a definição do número de fatores a ser adotado [52][75].

2.5.5 Avaliação das cargas

A matriz de cargas, também chamada de matriz fatorial, é tal que as linhas representam variáveis e as colunas, fatores. Os elementos da matriz são as cargas (c_{ij}) da variável

i no fator j , como introduzidas na Equação 2.12. Um exemplo de matriz de cargas é apresentado na Figura 2.25.

$$\begin{array}{c} X_1 \\ X_2 \\ \vdots \\ X_n \end{array} \begin{bmatrix} F_1 & F_2 & \dots & F_m \\ c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{bmatrix}$$

Figura 2.25: Exemplo de matriz de cargas
Fonte: Autora

As cargas representam o quão fortemente a variável i depende do fator j e deve-se avaliar o valor delas para interpretar o modelo. Os autores de [52] propõe os critérios apresentados no Quadro 2.3 para estudos que usam variáveis padronizadas.

| Valor absoluto de c_{ij} | Interpretação |
|----------------------------|--|
| $0,3 \leq c \leq 0,4$ | Nível mínimo para interpretação da estrutura |
| $0,5 < c \leq 0,7$ | Cargas praticamente significantes |
| $ c > 0,7$ | Estrutura bem definida (desejável) |

Quadro 2.3: Interpretação dos valores das cargas da matriz fatorial

Comumente, a matriz de cargas apresenta cargas concentradas em um único fator ou valores predominantemente baixos, o que dificulta a sua interpretação. Nesses casos, recomenda-se realizar o processo de rotação.

Variáveis com cargas significantes em mais de um fator não são bem vistas. Na presença delas, deve-se considerar tentar outro método de rotação, ou até mesmo eliminar a variável, exceto quando há justificção teórica para o resultado [52].

2.5.6 Rotação de fatores

Os métodos de extração de fatores os selecionam por ordem de maior variância, o que faz com que a maior parte da variância fique concentrada nos primeiros fatores [52]. É desejável redistribuir a variância entre os demais fatores a fim de que o modelo seja mais consistente em termos teóricos, favorecendo sua interpretação. Para isso recorre-se à técnica de rotação.

Em Álgebra Linear, rotacionar uma matriz significa multiplicá-la por uma matriz ortogonal, gerando uma rotação rígida dos eixos coordenados (o ângulo entre os eixos permanece inalterado) [75]. Pode-se provar que ao rotacionar a matriz de cargas encontrada, a nova matriz de cargas não irá afetar a matriz covariância do modelo, o que possibilita a utilização desta técnica em Análise de Fatores [112]. Os eixos que sofrem a rotação correspondem aos fatores. O papel da rotação na interpretação dos fatores é ilustrado na Figura 2.26.

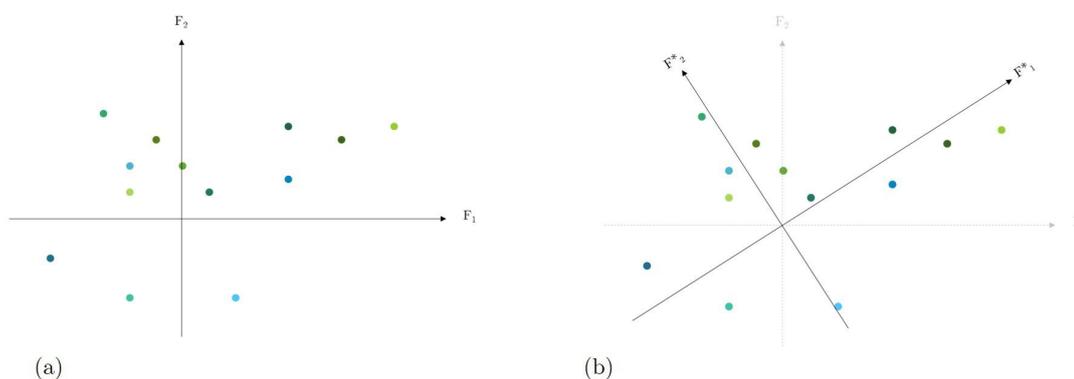


Figura 2.26: Exemplo de rotação em um modelo de análise fatorial com dois fatores. Os pontos representam as variáveis e são plotados de acordo com a carga relacionada a cada fator. (a) Eixos originais. (b) Eixos rotacionados

Fonte: Adaptado de [112].

Pode-se perceber que originalmente era difícil discernir de que forma as variáveis se agrupariam, pois as cargas estavam razoavelmente equilibradas entre os fatores. No entanto, após a rotação, fica mais claro quais variáveis estão fortemente ligadas a um fator ou a outro. É importante notar que não só a matriz de variância não é alterada, como também a matriz residual e as comunalidades [75]. Além disso, a técnica de rotação pode ser aplicada tanto após a análise por componente principal quanto após a análise de fatores comuns.

Existem, no entanto, muitas rotações possíveis e, para selecionar aquela que trará os resultados desejados, foram criados alguns critérios de rotação [112][52]. As principais abordagens de rotação ortogonal são apresentadas a seguir. Ressalta-se que existem métodos de rotação oblíqua (o ângulo dos eixos pode variar após a rotação), que permitem que os fatores sejam correlacionados, o que, inclusive, aumenta o risco da análise se superajustar à amostra e diminuir sua capacidade de generalização [52].

Os métodos mais comuns de rotação ortogonal são:

1. *Quartimax*: busca simplificar as linhas da matriz fatorial (matriz de cargas), ou seja, cada variável deve ter carga alta em um fator e baixa nos demais. Tem como ponto negativo o fato de que tende a agrupar todas as variáveis em um fator só, o que geralmente já ocorre antes da rotação, então nem sempre há um benefício adicional;
2. *Varimax*: busca simplificar as colunas, buscando fatores com participações altas de algumas variáveis e baixa das demais, deixando a análise mais clara;
3. *Equimax*: consiste em um meio termo entre os métodos anteriores, porém não é bem aceito e, conseqüentemente, é pouco usado [52].

2.6 Otimização

A busca por métodos científicos para solucionar problemas de alocação de recursos caracteriza a Pesquisa Operacional (PO), cuja origem remonta à Segunda Guerra Mundial [110][66]. Em problemas de PO, deve-se identificar quais variáveis podem ser alteradas a fim de alcançar determinado objetivo, frequentemente a maximização ou minimização (otimização) de uma função, a chamada “função objetivo” [110]. Deve-se observar ainda a quais restrições as variáveis estão sujeitas, como os limites físicos de determinado recurso. Essa etapa de criação do modelo é de suma importância para a análise posterior e deve-se garantir que o modelo matemático descreve satisfatoriamente as condições reais.

A escolha de algoritmo para resolução do problema elaborado depende das características do modelo e das funções envolvidas. Por exemplo, quando a função objetivo (FO) e as restrições são lineares, tem-se um problema de programação linear, geralmente solucionado pelo algoritmo SIMPLEX; já quando as variáveis de decisão só podem assumir valores inteiros, o problema é de programação inteira [110]. Nesta seção, destacar-se-á os problemas de programação quadrática, que correspondem aos modelos a serem elaborados na Seção 4.6.

Os problemas de programação quadrática fazem parte da área de programação não-linear e são caracterizados por terem função objetivo na forma de um polinômio de segundo grau e restrições lineares, como na Equação 2.30. Destaca-se que na Equação 2.30, ao invés de “Minimizar”, poderia-se definir o modelo com “Maximizar”, de acordo com a aplicação.

$$FO : \text{Minimizar } z = f(x) \tag{2.30}$$

sujeito a

$$h(x) = 0$$

$$g(x) \geq 0 \qquad \text{ou " } \leq \text{ "}$$

em que $f(x)$ é uma função quadrática das x variáveis de decisão; $h(x)$ é uma função linear e representa uma restrição de igualdade; $g(x)$ também é uma função linear e representa uma restrição com inequação.

A resolução de problemas desse tipo geralmente segue ou um método Simplex modificado e calculado a partir de determinadas condições para um ponto ser considerado ótimo de acordo com o problema, conhecidas como condições de Karush-Kuhn-Tucker; ou através de métodos de projeção do gradiente [61].

Capítulo 3

Trabalhos Relacionados

Neste capítulo será apresentada uma revisão de trabalhos publicados relativos à questão de inserção de fontes intermitentes em sistemas elétricos, incluindo aqueles que utilizam o conceito de "complementaridade".

3.1 Contextualização

Fontes renováveis de energia são aquelas cuja taxa de renovação do recurso na natureza é superior à de utilização (e.g. sol, vento, água), ou cuja disponibilização do recurso depende de ação humana (e.g. biomassa, resíduos urbanos e agrícolas, etc) [38]. As fontes renováveis se subdividem em dois grandes grupos: as fontes intermitentes e as não intermitentes.

As fontes intermitentes estão geralmente associadas a variáveis climáticas e têm comportamento variável e que não pode ser previsto com exatidão. A irradiação solar, os ventos, as vazões fluviais e as marés são exemplos de fontes intermitentes. Por outro lado, biomassa, energia geotérmica e usinas hidrelétricas com reservatórios são consideradas fontes não intermitentes [44].

Como apresentado na Seção 2.3, os sistemas elétricos tradicionais, projetados para trabalhar com suprimento constante e confiável de energia, encontram dificuldade em absorver um volume grande de fontes intermitentes. Nesse sentido, vários trabalhos relacionados à inserção de fontes intermitentes nas matrizes elétricas têm sido publicados.

Os métodos de estudo e as resoluções de dados utilizados ainda não são padronizados [44]. Mesmo o enfoque dos trabalhos varia aproximadamente por três grandes grupos, com fronteiras tênues e eventuais interseções: penetração máxima, complementaridade e

formulação de *mix ótimo*, como na Figura 3.1.

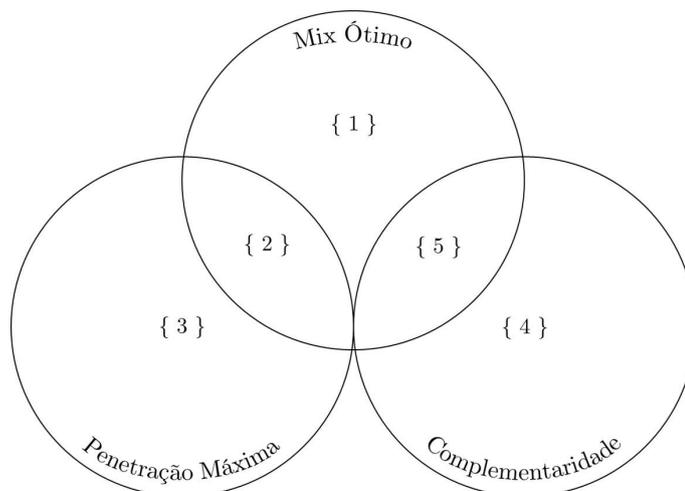


Figura 3.1: Algumas áreas de enfoque de trabalhos que avaliam a inserção de fontes de intermitentes em sistemas elétricos.

Fonte: Autora

Os estudos de penetração (área 3 da Figura 3.1) têm como objetivo principal avaliar qual a porcentagem máxima de inserção de uma determinada fonte renovável no sistema elétrico, sem que a energia gerada e não aproveitada não ultrapasse determinado limite ou que a demanda não seja suprida. Podem avaliar também se determinada meta de inserção (e.g. 40%) é viável ou não.

Pode-se citar, por exemplo, o trabalho de [32] que investigou a máxima penetração eólica viável para o nordeste brasileiro, através da análise horária da carga (demanda) e do fator de capacidade das usinas eólicas já instaladas. O objetivo foi obter um vertimento (energia gerada que não pode ser absorvida pelo sistema) tolerável (menos que 1% ao ano), já que valores pequenos são aceitáveis para fontes intermitentes e permitem uma maior participação das mesmas, mas valores altos implicam em prejuízos no retorno financeiro já que as usinas operarão a um fator de capacidade muito baixo [32]. Considerando o restante do sistema completamente flexível, o estudo estimou uma penetração máxima de 65% para a região.

Outros estudos avaliam a inserção máxima de mais de uma fonte intermitente, de forma que o interesse não é apenas calcular a penetração máxima, mas também a combinação mais eficiente dessas fontes, utilizando alguma técnica de otimização. Neste caso, os trabalhos se encaixariam na área 2 da Figura 3.1, ou seja, na interseção entre penetração máxima e *mix ótimo*.

Cita-se por exemplo o artigo [40], no qual os autores buscam avaliar se usinas eólicas,

solares e geotérmicas poderiam substituir as usinas nucleares e termelétricas fósseis da Suíça. As grandes hidrelétricas com reservatórios já instaladas serviriam de apoio para as fontes intermitentes. Para isso, tomaram como base a geração não renovável de 2014 e averiguaram se o *mix* proposto poderia tê-la substituído. Como a geração geotérmica não é intermitente, um determinado valor foi estipulado para ela e o restante deveria ser suprido por uma combinação ótima de geração eólica e fotovoltaica. Foram utilizados dados de radiação solar provenientes de satélites, dados de estações de medição de vento e a série histórica de demanda. Constatou-se que o ideal é que a participação da energia fotovoltaica represente entre 30% e 60% da combinação eólico-solar.

Há ainda estudos que buscam otimizar toda a matriz elétrica, considerando fontes intermitentes, outras fontes renováveis e usinas térmicas (área 1 da Figura 3.1).

É o caso, por exemplo, do estudo [102] que buscou obter um *mix* ótimo de geração termelétrica, fotovoltaica, eólica e hidrelétrica (intermitente e não intermitente) no Brasil. Foi construído um modelo de otimização com dependência temporal cujo objetivo era minimizar a utilização de termelétricas. Foram utilizados dados diários e utilizou-se dados meteorológicos de velocidade do vento, vazões e irradiação. Concluiu-se que as fontes eólica e fotovoltaica ajudam a estabilizar o sistema termo-hidrelétrico. O *mix* ótimo seria composto de aproximadamente 37% de geração fotovoltaica, 9% de eólica e 50% de hidrelétrica, ou seja, a necessidade de geração termelétrica seria de apenas 2%.

Outro trabalho na área de *mix* ótimo foi conduzido em [77]. Nele os autores criaram um modelo de otimização buscando utilizar energia fotovoltaica e hidrelétrica proveniente de usinas a fio d'água. Essas usinas teriam um pequeno reservatório que serviria como armazenamento de energia também (usinas reversíveis). O modelo busca maximizar a utilização das energias renováveis respeitando a restrição de que o deficit/excesso de energia não ultrapasse 5% da demanda.

3.2 Estudos de Complementaridade

Os estudos na área de complementaridade pura (área 4 da Figura 3.1) avaliam se duas ou mais fontes intermitentes tem variabilidade não correlacionada ou correlacionada negativamente, de modo que a utilização de ambas diminui as variabilidades individuais. Alternativamente, pode-se avaliar o mesmo recurso, mas em locais diferentes. As técnicas mais utilizadas nesses estudos são análise gráfica e cálculo de correlações.

Os autores de [24], por exemplo, avaliaram a geração eólica, fotovoltaica e de pequenas

hidrelétricas em Portugal, partindo de séries de geração real. Avaliaram o comportamento anual, mensal, diário e horário das fontes através de gráficos. Calcularam os coeficientes de correlação linear entre as séries eólicas e solares e constataram correlação negativa em todos os horizontes temporais que vai se intensificando conforme o intervalo de tempo aumenta (correlação anual mais forte que a mensal, e assim sucessivamente).

Outro artigo da área é [23], um trabalho brasileiro que propôs o conceito de "mapas de correlação hídrico-eólicos" cujo objetivo é facilitar a visualização das correlações de determinada região. Os autores utilizam os chamados diagramas de Voronoi para agrupar estações de medições de vazões, definem a usina hidrelétrica mais representativa de cada bacia hidrográfica, calculam as correlações entre as estações eólicas e depois as correlações entre as estações eólicas e as bacias hidrográficas, apresentando os resultados em mapas. Utilizaram-se os coeficientes de Pearson e de Spearman para o cálculo das correlações. Para o estudo, foram utilizados dados mensais de vazões e de velocidade do vento. A nível mensal, os autores constataram correlações positivas entre os regimes eólicos ao longo do país. Constatou-se complementaridade entre os recursos eólico e hídrico no Nordeste, na maior parte do Sudeste e em parte do Centro-Oeste.

O trabalho de [63] avaliou a complementaridade geográfica da irradiação solar na Espanha. Para isso, reduziu-se o conjunto inicial de dados, utilizando análise de componente principal e, em seguida, agrupou-se as séries utilizando *clusters*. Em seguida, avaliaram a variabilidade do recurso graficamente e calcularam a complementaridade (coeficiente de correlação). Consideraram-se diferentes tecnologias e configurações para a estimativa da energia gerada com base nos dados de radiação solar.

Em [76], os autores avaliam a complementaridade eólico-solar em pequenos sistemas híbridos isolados e propõem um modelo de cálculo da confiabilidade do sistema, baseando-se nos fatores de capacidade das fontes, coeficientes de correlação entre elas em diferentes escalas temporais e na capacidade de armazenamento. Foram usados dados de 86 estações na Polônia, medidos entre os anos 2010 e 2016, e avaliaram-se escalas de 15 em 15 minutos, horária, diária e mensal. Constatou-se que a confiabilidade entra-horária depende fortemente do armazenamento disponível, mas a níveis diário e mensal, a complementaridade tem papel importante no aumento da confiabilidade.

Os autores de [87] buscaram avaliar a complementaridade eólica e hidrelétrica por todo o Brasil, considerando séries de 1948 a 2010 de geração hidrelétrica de todos os subsistemas do SIN e dados de geração eólica teórica para treze pontos no país – dez na região nordeste e três na região sul. O coeficiente de correlação linear foi calculado,

indicando complementaridade entre as fontes. No entanto o estudo ainda discriminou a produção em parcela previsível e imprevisível (mais custosa para o sistema elétrico) e buscou verificar a existência de complementaridade da parcela imprevisível. Neste caso, os autores consideraram que o coeficiente de correlação linear não captaria suficientemente as relações não lineares e utilizaram, além deste, a aproximação por cópulas. Concluiu-se que as parcelas de geração não explicáveis pela sazonalidade não apresentaram correlação na maioria dos casos.

Em [37], os autores avaliaram a correlação entre os recursos eólico e solar medidos na ilha de Fernando de Noronha, pertencente ao estado de Pernambuco, na região Nordeste. Foram utilizados dados de velocidade do vento e radiação solar acumulada, de 2004 a 2013. As séries passaram pelo processo de normalização e foram analisadas através das análises DFA (*detrended fluctuation analysis*) e DCCA (*detrended cross-correlation method*). A primeira verifica como cada série apresenta dependência de seus valores passados, enquanto a segunda avalia a dependência de uma série aos valores históricos dela própria e aos de outra série. Os autores concluíram que ambas as séries apresentam dependência a seus valores passados e aos valores da outra série.

Os autores de [106], partindo da ideia de que o sistema elétrico brasileiro depende fortemente da geração hidrelétrica, avaliaram se a geração eólica *offshore* poderia complementar o regime hidráulico, enquanto os grandes reservatórios de usinas hidrelétricas ajudariam a regular a variabilidade horária e diária dos ventos. O estudo considerou dados de satélites para ventos em doze pontos na costa brasileira, além de dados de precipitação e temperatura para doze bacias hidrográficas no país, entre 1989 e 2009. As vinte e quatro séries foram analisadas através do cálculo do coeficiente de correlação linear (coeficiente de Pearson), da análise de coerência (que analisa correlação no domínio da frequência) e da análise de *cluster*, que agrupa pontos com comportamentos semelhantes. Os autores concluíram que o recurso eólico *offshore* complementa a si mesmo e aos recursos hídricos, constituindo-se, assim, uma alternativa para complementar a geração elétrica nacional.

Outro trabalho nacional é [18], que avaliou as gerações hidrelétrica e fotovoltaica para o estado do Rio Grande do Sul, no sul do país, além de propor um índice adimensional para estudos de complementaridade. O índice em questão avalia a complementaridade no tempo, de média e de amplitude e o produto destes três itens, que variam de 0 a 1, gera o valor final do índice.

Os autores de [101] tiveram como objetivo propôr um novo método de visualização de complementaridade entre fontes localizadas em regiões diferentes, de modo que pudessem

avaliar a complementaridade de uma região com várias outras, sem tomar uma região como referência para comparação. Para isso, foi proposto que o domínio geográfico seja discretizado em várias células hexagonais e que uma “rosa de complementaridade” seja inserida dentro de cada uma. Cada rosa é composta por um conjunto de retas, representando a relação da célula onde a rosa está com as demais. Cada reta aponta na direção de outra célula, sendo que o comprimento da reta representa a distância entre elas e a cor representa a força da complementaridade.

3.3 Estudos de Complementaridade com Otimização

Alguns estudos de complementaridade, no entanto, não se limitam a avaliar o comportamento dos recursos, mas também visam encontrar um *mix* de fontes intermitentes que (i) minimize a variância, ou (ii) que minimize a demanda não atendida e a geração excedente, ou ainda, (iii) que minimize necessidade de armazenamento. Alguns trabalhos eventualmente mesclam mais de um desses objetivos. Dessa forma, a análise de complementaridade utiliza métodos de otimização e gera artigos na região 5 da Figura 3.1.

Nessa área, pode-se citar [68], que avaliou a complementaridade eólico-solar na província de Ontário, Canadá. Foram utilizados dados de velocidade do vento e irradiância global horizontal entre 2003 e 2005 de quatro localidades para as análises. A potência total do *mix* foi definida como a soma de uma constante multiplicando a radiação solar mais outra constante multiplicando a velocidade do vento ao cubo. As constantes foram determinadas de acordo com o montante de potência total desejado. Nenhum pressuposto de tecnologias de aproveitamento foi assumido. Realizou-se análise gráfica e testou-se vários cenários de participação de fontes e localidades. Concluiu-se que utilizar as duas fontes diminui a variabilidade das fontes isoladas e que dispersá-las geograficamente diminui ainda mais a variabilidade.

Outro trabalho da área é [19]. Nele, os autores avaliam a complementaridade entre as fontes eólica e solar nas ilhas Britânicas. Utilizaram-se dados de velocidade do vento e de irradiação acumulada. Inicialmente, avaliou-se um conjunto de gráficos (distribuição conjunta, desvio médio mensal e correlação média mensal). Em seguida, verificou-se o coeficiente de variação de cada região. Por fim, constatou-se que o cenário que minimiza a variabilidade da potência do *mix* solar-eólico é composto por 70% de energia solar e 30% de energia eólica.

O estudo [88] conduzido em Portugal, buscou otimizar um *mix* de fontes eólica, solar e hídrica. Foram utilizadas séries de 50 anos de radiação solar global, velocidade do vento e vazão medida em represas. Definiu-se que a série solar seguia uma distribuição normal e definiu-se um modelo matemático (seguindo distribuição de Weibull) para a série eólica que dependia da solar. Por fim, o modelo de vazões foi feito com dependência dos dois modelos anteriores. O trabalho propôs um modelo de otimização multi-objetivo, que envolve minimizar a intermitência entre anos e meses, além de minimizar custos e carga não garantida.

Os autores de [65] avaliaram a complementaridade entre geração eólica e solar por toda a Europa. Foram utilizados dados horários em série de oito anos (entre 2000 e 2007). Compararam o comportamento das duas fontes com a demanda e buscaram dois métodos de otimização: (i) minimizar a energia não atendida; (ii) minimizar a necessidade de armazenamento. Testam para uma matriz composta só por geração eólica e solar, e depois, permitem a utilização de fontes fósseis. Os resultados indicam um *mix* ótimo de 55% de energia eólica e 45% de solar, em uma matriz 100% renovável. Se outras fontes não intermitentes forem inseridas, a participação eólica tende a aumentar e, conseqüentemente, a solar diminui.

O trabalho [55] buscou obter o melhor *mix* entre geração fotovoltaica e hídrica na Itália, sendo que foram considerados dois grupos de vazões com comportamentos distintos: aquelas influenciadas fortemente por degelo e as que sofrem influência de chuvas. Foi realizada uma otimização pensando em minimizar o desvio entre geração e demanda e outra buscando minimizar a necessidade de armazenagem.

O estudo [33] avaliou a complementaridade entre as fontes eólica, solar e proveniente de pequenas hidrelétricas no estado do Rio de Janeiro e considerou o *mix* ótimo como aquele de menor variância. O resultado obtido foi de 62% para a geração fotovoltaica, 21% para a eólica e 17% para a hidrelétrica. O *mix* encontrado apresentou melhora de 61% no desvio-padrão, comparado ao desvio do pior cenário.

Esta dissertação se enquadra nesta área de complementaridade com otimização. A metodologia adotada no estudo será apresentada a seguir.

Capítulo 4

Metodologia

Este capítulo descreverá os métodos utilizados neste estudo. Um resumo deles se encontra na Figura 4.1. Cada etapa será descrita detalhadamente nas próximas seções. Grande parte das análises foram feitas através do software gratuito e de código aberto RStudio que se aplica à linguagem de programação R.

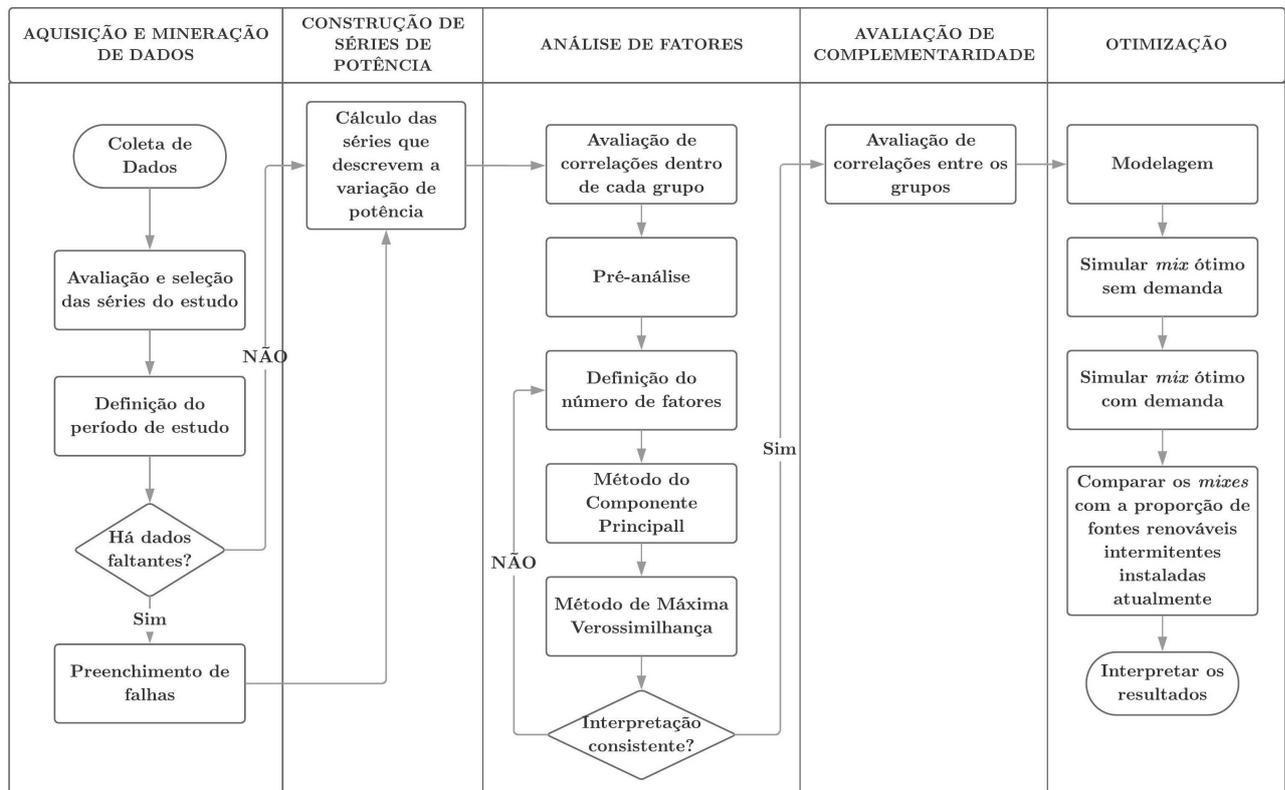


Figura 4.1: Diagrama com o resumo dos métodos adotados no estudo
Fonte: Autora

4.1 Aquisição e mineração de dados

4.1.1 Base de dados

Para a realização do estudo, dever-se-ia optar por séries de geração de eletricidade ou por dados ambientais, como velocidade do vento, radiação solar e vazões fluviais. A primeira opção poderia apresentar resultados mais fidedignos, já que as séries representariam gerações reais de usinas existentes. No entanto, muitas vezes essas séries não estão disponíveis ou, quando estão, abrangem um curto período de tempo ou apresentam dados sintetizados, que dificultam uma análise apurada. Dessa forma, muitos estudos adotam a segunda opção. Como há poucas unidades geradoras dos tipos UFV e EOL nas regiões SE/CO e as PCHs e CGHs não disponibilizam dados de geração diária, optou-se pela estimativa de potência através de dados ambientais.

Para a base de dados ambientais, poder-se-ia utilizar dados oriundos de plataformas de coleta de dados localizadas em solo ou informações provenientes de satélites. Cada opção tem vantagens e desvantagens. Enquanto estações em solo proveem boa resolução temporal de dados, os dados de satélite têm, em geral, uma resolução mais esparsa [82]. Por outro lado, as estações em solo representam bem a localidade onde estão instaladas, mas como nem todos os municípios as têm, é necessário uma rede com boa quantidade delas para que a análise represente satisfatoriamente toda a região estudada. Neste quesito, os dados de satélite podem ter um desempenho melhor.

Em [85], por exemplo, um estudo foi feito visando eleger o método mais apropriado para estimar a irradiação solar em localidades sem dados próprios. Para isso, realizaram estimativas a partir da utilização de dados de PCDs e compararam os resultados a estimativas de radiação baseadas em outras variáveis, como irradiação no topo da atmosfera e cobertura de nuvens, provenientes de satélites geoestacionários. Os resultados do estudo apontaram que para localidades até 60km dos pontos de medições, os métodos tiveram resultados similares, mas que conforme essa distância aumentava, a estimativa a partir de satélites apresentava erros menores.

Em estudos de complementaridade de fontes intermitentes, a resolução temporal é especialmente importante. Logo, optou-se pela utilização de dados medidos diretamente em estações em solo. O período de estudo foi determinado de tal forma que todos os estados das regiões sudeste e centro-oeste tivessem ao menos uma estação para representá-los.

Os dados de radiação solar e velocidade do vento selecionados são provenientes de Plataformas de coleta de dados (PCDs) gerenciadas pelo Instituto Nacional de Pesquisas Espaciais (INPE) e ligadas ao Sistema Integrado de Dados Ambientais (SINDA). Duas fontes de dados de vazões fluviais foram utilizadas: as estações do Hidroweb, sistema da Agência Nacional de Águas (ANA), e as estações de responsabilidade do Operador Nacional do Sistema Elétrico (ONS). A série histórica de demanda elétrica do subsistema SE/CO é também disponibilizada pelo ONS.

As próximas subseções descrevem cada uma dessas fontes de dados.

4.1.1.1 Plataformas de coleta de dados do INPE

As PCDs do sistema SINDA do INPE são estações meteorológicas automáticas dispersas pelo território nacional. As PCDs começaram a ser instaladas no início da década de 90. De acordo com [39], o software do SINDA cadastra PCDs, coleta e transmite os dados enviados por elas, avalia a consistência dos dados, os converte para as unidades padrão e disponibiliza os dados no website [73]. No entanto, não foi encontrado um detalhamento da etapa de avaliação de consistência de dados.

Algumas estações transmitem dados de hora em hora, enquanto outras o fazem a cada três horas. Nem todas são dotadas dos mesmos equipamentos, dependendo de sua finalidade, pois podem prover dados para análises ambientais, hidrológicas, meteorológicas ou agrometeorológicas [72]. As plataformas de interesse deste trabalho são aquelas dotadas de sensores ultra-sônicos de vento e/ou sensores de radiação solar global.

Os sensores que computam a velocidade do vento, são dotados de três transdutores ultrassônicos, de forma que o valor obtido independe da altitude, temperatura ou umidade do local. Além disso, uma técnica de processamento de sinais remove valores possivelmente equivocados, ocasionados por chuva forte ou granizo, por exemplo [84]. O valor transmitido é a velocidade média do vento, calculada a partir de 200 medições instantâneas captadas durante um período de 10 minutos [83]. Já os sensores que captam a radiação solar global, realizam 1080 medições, com um intervalo de 10 segundos entre elas, durante 3 horas e fornecem a integração desses valores [83]. É importante destacar que os dados das PCDs têm seus horários segundo a hora universal GMT e não ao horário de Brasília [83].

4.1.1.2 Estações Convencionais da ANA

As estações hidrométricas convencionais da ANA, que enviam dados ao sistema Hidroweb [3] para formação de séries históricas, são gerenciadas por diversas entidades, porém a ANA requer o envio de relatórios anuais acerca da consistência dos dados medidos. A agência provê diretrizes, além de alguns softwares, para auxiliar neste processo, mas a metodologia de análise fica a critério dos administradores das estações [7]. Os dados divulgados podem ser classificados da seguinte forma:

1. Quanto à consistência:
 - Bruto
 - Consistido
2. Quanto ao método de obtenção de vazões:
 - Curva de descarga
 - Transferência de vazões
 - Soma de vazões
 - Método Acústico Doppler (ADCP)
3. Quanto ao status de cada medição:
 - Branco
 - Real
 - Duvidoso
 - Régua Seca

4.1.1.3 Estações do SIN

Os dados do Sistema de Acompanhamento de Reservatórios(SAR) [5] do SIN, apesar de hospedados em um site da ANA, são de responsabilidade do ONS que acompanha estações fluviométricas, geralmente associadas a grandes reservatórios, para geração de eletricidade. Os dados de vazões recebem um tratamento especial para que as vazões

resultantes reflitam aquelas que seriam observadas caso não houvesse retirada de água para uso consuntivo, evaporação intensificada devido às represas e reservatórios à montante [92][6]. Os métodos de cálculo de vazões naturais podem ser revistos quando metodologias mais apuradas são desenvolvidas. As vazões naturais são especialmente importantes para estudos que visam a instalação de novos aproveitamentos hídricos, incluindo a geração de eletricidade, pois o planejamento destes depende do conhecimento de características inerentes às bacias, sem interferência antrópica [6].

4.1.1.4 Estações da ANA *versus* Estações do SIN

A motivação para o uso de duas fontes de dados de vazões, neste caso, não decorre unicamente da disponibilidade de ambas. Como as séries de vazões naturais, como disponibilizadas pelo SIN, são as indicadas para os cálculos de viabilidade de futuros empreendimentos hidrelétricos, elas foram selecionadas para representar o comportamento esperado de novas PCHs ou CGHs que poderão vir a ser instaladas.

Algumas estações da ANA, por outro lado, estão localizadas próximas a PCHs e CGHs, e as vazões por elas medidas são similares àquelas que passam pelas usinas. Assim, selecionamos estações que estivessem próximas a PCHs e CGHs existentes, desativadas ou em fase de construção (iniciada ou não) para representar a geração atual proveniente dessas usinas. Para isso, comparamos as coordenadas das estações da ANA com as coordenadas de PCHs e CGHs e selecionamos as estações que estivessem a um raio de até 5km de alguma dessas usinas e no mesmo rio que elas.

4.1.1.5 Série de Demanda Elétrica

O Operador Nacional do Sistema Elétrico disponibiliza em seu site [93] séries horárias de demanda por eletricidade (carga elétrica) para cada subsistema. Como apresentado na Seção 2.1.1, o subsistema SE/CO passou a incluir parte de dois estados da região norte, mas como 98,7% do consumo do subsistema vem de fato das regiões Sudeste e Centro-Oeste, considerou-se que a demanda histórica do subsistema representa satisfatoriamente o consumo nas regiões do estudo.

4.1.2 Avaliação de Dados

4.1.2.1 Controle de Qualidade e Preenchimento de Falhas

A qualidade dos dados depende da prevenção de falhas. Porém, mesmo com atenção à prevenção, erros podem ocorrer e daí vem a importância da validação, para verificar imprecisão, inconsistência ou implausibilidade dos dados, e da eventual correção deles [118]. A Organização Meteorológica Mundial (OMM) prevê diretrizes para o controle de qualidade de dados desde a coleta nas estações. Sugere-se um pré-processamento de dados, etapa na qual deve-se checar duplicatas, o código da estação e data e hora das medições, por exemplo [117]. A OMM recomenda que a primeira parte do controle de qualidade seja feita ainda na estação, partindo de dados brutos, e a segunda, na estação ou no centro meteorológico nacional [118]. As etapas sugeridas são similares para as duas partes e são resumidas na Figura 4.2.

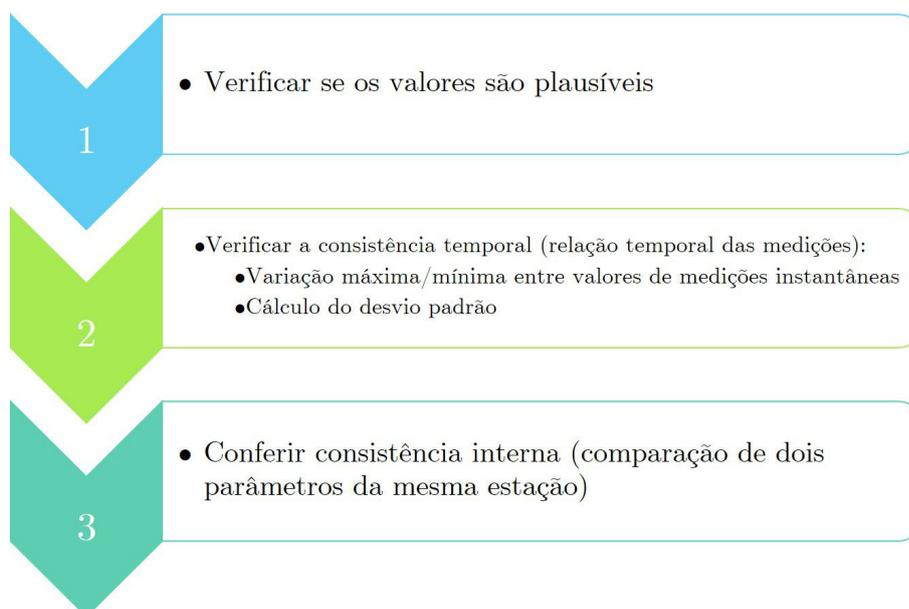


Figura 4.2: Etapas do controle de qualidade de dados meteorológicos
Fonte: Adaptado de [118]

As decisões no controle de qualidade, sejam manuais ou automáticas, devem ser evidenciadas através de indicadores de qualidade de cada sistema [117]. Inicialmente, os dados são apenas classificados como corretos ou errados, pois são comparados à faixa de valores possíveis da variável medida. Depois, são comparados ao que se espera em uma determinada região, estação do ano ou mês e então são classificados como corretos, inconsistentes, duvidosos, errôneos ou faltantes [118] [117]. Lidar com valores considerados suspeitos não é uma tarefa trivial, pois é necessário ponderar o risco de aceitar valores

errados contra o risco de rejeitar valores corretos[117].

De maneira similar, trabalhar com valores faltantes também não é simples. Por um lado, descartar variáveis com dados faltantes representa um mau uso de informação disponível [103]. E, por outro, ignorar a existência de valores faltantes pode inviabilizar determinadas análises, que não têm suas condições de aplicação satisfeitas nessas condições, ou levar a interpretações equivocadas acerca das variáveis. Isso porque os estimadores calculados a partir do conjunto de dados disponíveis poderão ser viciados, gerando resultados tendenciosos, já que é difícil afirmar que todos os erros são aleatórios, ou seja, independem do valor real que deveria ser medido [103] [118].

Sendo assim, uma solução plausível é substituir os dados faltantes ou inválidos por algum valor razoável. A técnica a ser utilizada depende, dentre outros fatores, da área de aplicação e da extensão dos valores faltantes [115]. Uma das soluções mais simples e abundantes na literatura é a interpolação linear, que é indicada para sequências curtas de valores faltantes [115] [60] [99]. Muitos autores utilizam técnicas de estimativa de valores faltantes a partir de medições de estações próximas [81] [99] [104] ou de outros parâmetros da mesma estação [113]. Outros métodos também usados são algoritmo de maximização de expectativa [103], matriz de transição de Markov [74], avaliação de tendência linear ou regressão linear (simples e múltipla) [96] [99].

Por este estudo avaliar justamente as relações entre estações, não seria ideal utilizar dados de algumas para preencher falhas de outras. Logo, optou-se por escolher um método de substituição de dados faltantes que levasse em conta apenas a estação em questão. Por simplicidade, adotamos o método de interpolação linear e escolhemos séries com até 10% de valores faltantes ou duvidosos.

Das estações selecionadas, apenas as do Hidroweb fornecem algum tipo de indicador de qualidade de dados, identificando as medições como reais, duvidosas ou de momentos de régua seca. Algumas medições, no entanto, não têm o indicador preenchido ou o campo é identificado como "Branco".

As vazões do ONS, apesar de serem divulgadas sem indicador, passam por um controle de qualidade cauteloso, de acordo com as etapas seguintes [89]:

1. Consistência em tempo real: identificação de valores suspeitos e correção de erros grosseiros;
2. Uso de médias móveis e agrupamento de sub-bacias para eliminar vazões incrementais negativas e eliminação de erros identificados na etapa anterior ou de erros

identificados já na segunda etapa;

3. Análise das séries hidrológicas levando em conta características das bacias.

Já os dados do SINDA/INPE não disponibilizam indicadores de qualidade dos dados e é desconhecido aos autores o processo de validação deles. Na realidade, o INPE tem outro sistema de fornecimento de dados ambientais - o Sistema Nacional de Organização de Dados Ambientais (SONDA) - que tem dados coletados a cada minuto ou de 10 em 10 minutos e cujo controle de qualidade é explícito e segue as diretrizes da OMM. No entanto, ainda há poucas estações disponíveis, o que limita a cobertura geográfica, e o período de medições em comum não é longo. Dessa forma, as estações do SONDA não foram selecionadas para este estudo.

Dito isto, o tratamento de cada fonte de dados é detalhado nas subseções seguintes.

4.1.2.2 Tratamento dos dados de PCDs do INPE

Os dados das PCDs são disponibilizados no site do sistema SINDA [73]. Como não foi encontrado um relatório da validação dos dados por parte do INPE, buscou-se reproduzir as etapas da Figura 4.2, o quanto fosse possível. Seguiu-se estes passos:

1. **Coletou-se os dados.** Foi feito o download dos dados para todos os anos disponíveis de cada estação que tivesse registros de velocidade do vento e/ou radiação solar, totalizando 125 estações (121 com dados de radiação solar e 124 com dados de velocidade do vento). Os anos de funcionamento de cada estação não são necessariamente os mesmos. No geral, havia dados disponíveis de 1997 a 2017.
2. **Verificou-se a existência de datas e horários duplicados.** No entanto, não foram encontradas duplicatas em nenhuma das estações;
3. **Verificou-se a omissão de algum horário.** Caso positivo, inseriu-se o horário na linha do tempo e o termo "sem dados", no campo de medição correspondente;
4. **Verificou-se a existência de valores fisicamente impossíveis.** Os pesquisadores do sistema SONDA adotam o limite superior de 25m/s para a velocidade do vento a 10m de altura [71]. Este valor máximo se refere à velocidade instantânea, porém adotamos o mesmo para a velocidade média. Nas estações cujos sensores de vento estão a 3m ou 5m de altura, utilizou-se a Equação 2.2 e um fator de rugosidade de 0,26 (terreno com muitas árvores, cerca viva ou poucas edificações [38]),

para calcular velocidades máximas aceitáveis de 18,3m/s e 20,9m/s respectivamente. Destaca-se que valores extremos (*outliers*) não foram removidos, a não ser que se enquadrassem na categoria de dados fisicamente impossíveis.

5. **Verificou-se a existência de valores suspeitos.** Como a velocidade do vento fornecida é média e a radiação solar, integrada durante as três horas, não foi possível verificar a variação temporal de acordo com os padrões da OMM, como na segunda fase da Figura 4.2. No entanto, percebemos que determinados valores se repetiam em horários seguidos por um período relativamente longo. Verificou-se então, tanto para as séries eólicas quanto solares, a existência de 11 medições repetidas em sequência ou repetição de um valor ao menos 7 vezes durante as últimas 11 medições. Os valores identificados não foram excluídos ou substituídos, mas sim identificados como “suspeitos” e não foram contabilizados como dados disponíveis para escolha das estações mais completas.
6. **Calculou-se a disponibilidade de dados de cada série.** Todas as séries possuíam valores faltantes e algumas também tinham dados suspeitos. Calculou-se, então, a porcentagem de dados disponíveis por ano de cada série.
7. **Seleção das séries mais completas.** Neste estudo são utilizadas séries com pelo menos 90% de dados disponíveis. Sabia-se de antemão que as séries do INPE têm limitação geográfica e temporal maior que as séries de vazões, logo determinou-se o período de estudo a partir das estações do INPE. Para tanto, verificou-se que o ano com maior número de estações com 90% ou mais de dados disponíveis foi 2008. Em seguida, excluiu-se as séries que não tinham essa disponibilidade em 2008 e selecionou-se o ano próximo (2007 ou 2009) que tivesse a maior quantidade de estações que atendessem os requisitos.

O processo de seleção de ano e exclusão de séries com muitos dados faltantes no período selecionado deveria ser repetido até que se conseguisse um período de anos considerado satisfatório, ou de tal forma que, se um ano a mais fosse escolhido, algum dos estados deixaria de ser representado. De fato, nossa seleção prosseguiu até o período de 2007 a 2011, pois a única estação restante de Mato Grosso não tinha boa disponibilidade de dados nem em 2006, nem em 2012. Assim, foram selecionadas 34 séries de radiação solar e 26 de velocidade do vento.

8. **Avaliação da sequência de dados faltantes.** Não só se verificou a porcentagem de dados faltantes, como também a forma como esses *gaps* se distribuem nas séries.

Longas sequências de dados faltantes não foram aceitas. Nesta etapa, foram excluídas três séries de radiação solar e uma de velocidade do vento, todas com mais de 110 dados faltantes em sequência. A maior sequência das séries restantes foi de 23 valores faltantes consecutivos.

9. **Preenchimento de falhas.** O processo de substituição de valores faltantes foi o mesmo para as séries de radiação solar e as de velocidade do vento. Adotou-se o seguinte:

- Preferencialmente, os valores faltantes foram substituídos pela média entre os valores observados naquele mesmo horário no dia anterior e no seguinte;
- Quando isso não foi possível, calculou-se a média entre a observação anterior e a seguinte;
- Se nenhum dos procedimentos anteriores pôde ser aplicado, substituiu-se o *gap* pela média mensal do horário.

Apresenta-se na Figura 4.3, a distribuição geográfica das estações do INPE utilizadas neste estudo, além dos climas do Brasil. Mais informações sobre as estações selecionadas e as abreviações usadas são apresentadas no Apêndice A.

4.1.2.3 Tratamento dos dados do ONS

Os dados de vazões naturais médias diárias gerenciados pelo ONS são disponibilizados no site do sistema SAR [5]. Com relação à coleta e ao tratamento deles, seguiu-se estas etapas:

1. **Coletou-se os dados.** Foi feito o download dos dados de 2007 a 2011 de todas as estações da região considerada. Ao todo, 93 estações tinham dados no período solicitado.
2. **Verificou-se o período disponível real das séries.** Duas estações selecionadas na etapa anterior possuíam dados apenas a partir de dezembro de 2008 ou início de 2009 e, logo, foram desconsideradas. Além delas, a estação da usina de Espora teve medições nulas por onze meses consecutivos. Foi feita uma breve pesquisa e pôde-se constatar que a barragem da usina se rompeu no fim de janeiro de 2008, quando as medições se tornaram zero. A estação foi, então, desconsiderada.

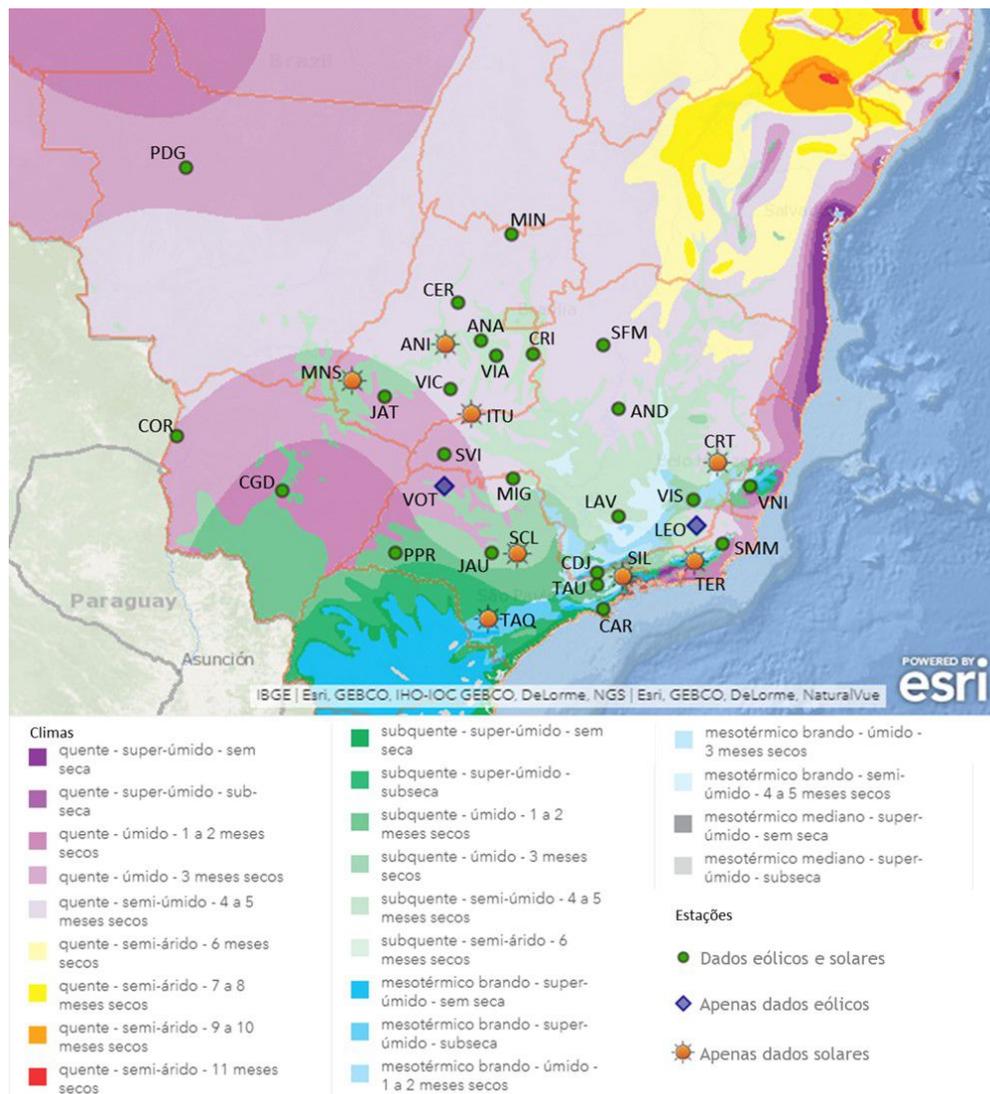
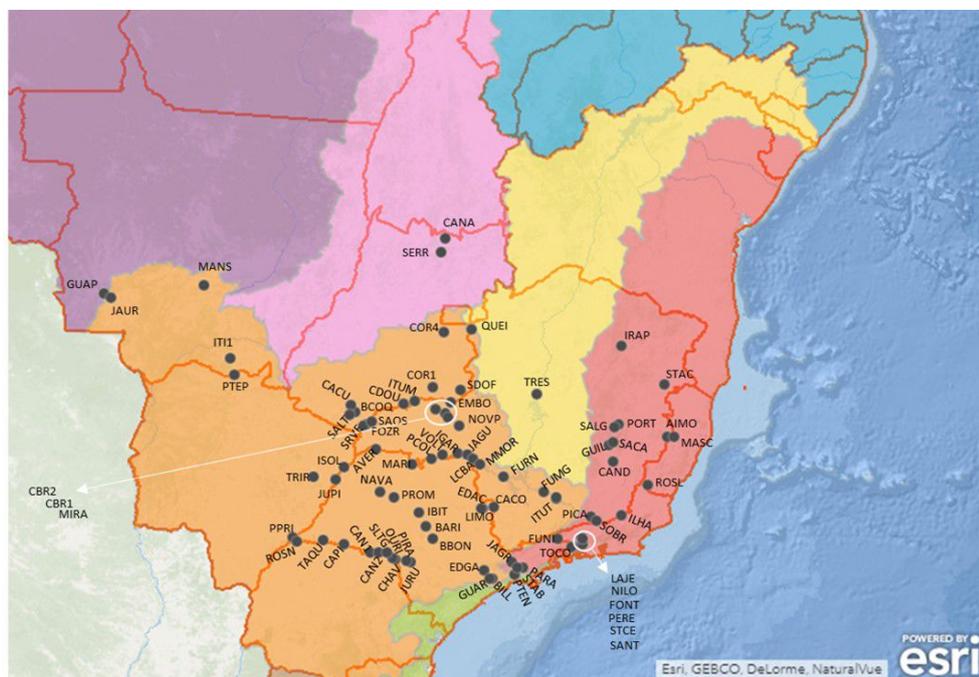


Figura 4.3: Localização geográfica das estações eólicas e solares consideradas no estudo
 Fonte: Elaboração própria através do ESRI ArcGIS online [51], com camada de climas de [69]

3. **Verificou-se a existência de datas duplicadas ou faltantes.** No entanto, não foram encontradas duplicatas. Quanto a datas faltantes, apenas uma estação não estava completa, e de tal forma que mais de 5% do total de dados não estavam disponíveis. Como ela foi a única a se enquadrar neste caso e o número de estações do ONS era expressiva, decidiu-se descartar a estação em questão.
4. **Verificou-se a existência de valores fisicamente impossíveis.** No caso de vazões, buscou-se valores negativos nas séries. Duas delas os possuíam e foram descartadas. Destaca-se que valores extremos (*outliers*) não foram removidos, a não ser que se enquadrassem na categoria de dados fisicamente impossíveis.

5. **Calculou-se a disponibilidade dos dados.** Todas as 87 estações restantes estavam completas, sem nenhum valor faltante. Então, não foi necessário tomar nenhuma medida de preenchimento de falhas. *Observação:* Durante análises posteriores, surgiram evidências de colinearidades na matriz de dados de vazões do ONS. Verificou-se que, de fato, duas séries eram perfeitamente iguais. Excluiu-se uma delas e a análise foi conduzida para 86 variáveis.

Apresenta-se na Figura 4.4, a distribuição geográfica das estações do ONS utilizadas neste estudo, além das bacias hidrográficas brasileiras. Mais informações sobre as estações selecionadas e as abreviações usadas são apresentadas no Apêndice A.



Bacias Hidrográficas - Departamento Nacional de Águas e Energia Elétrica(DNAEE) :

- ATLÂNTICO, TRECHO LESTE
- ATLÂNTICO, TRECHO NORTE/NORDESTE
- ATLÂNTICO, TRECHO SUDESTE
- RIO AMAZONAS
- RIO PARANÁ
- RIO SÃO FRANCISCO
- RIO TOCANTINS
- RIO URUGUAI
- Estações do ONS

Figura 4.4: Localização geográfica das estações do ONS

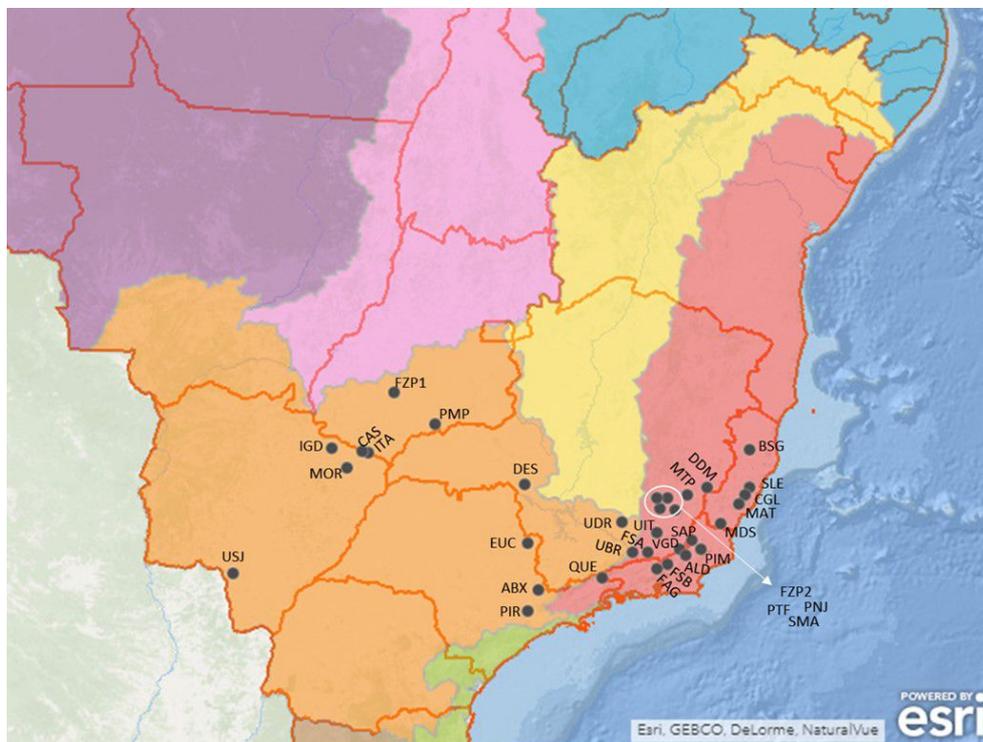
Fonte: Elaboração própria através do ESRI ArcGIS online [51], com camada de regiões hidrográficas de [4]

4.1.2.4 Tratamento dos dados da ANA

Os dados das estações convencionais da ANA são disponibilizados no site do sistema Hidroweb [3]. A coleta e tratamento dos dados seguiu os seguintes passos:

1. **Coletou-se os dados.** Como não havia opção de seleção de período, foi feito o download das séries históricas completas de vazões médias diárias de todas as estações fluviométricas localizadas nas regiões Sudeste e Centro-Oeste. Ao todo, foram coletados dados de 4879 estações.
2. **Verificou-se quais estações tinham dados no período de 2007 a 2011.** Nesta etapa também verificou-se se havia datas duplicadas ou algum mês faltante. Não foram encontradas datas duplicadas para a mesma classificação de consistência e as estações com meses faltantes foram descartadas. Após essa etapa, 340 estações foram selecionadas.
3. **Calculou-se a porcentagem de dados disponíveis.** Além da avaliação de dados faltantes, calculou-se a porcentagem de dados identificados como "reais", duvidosos", "régua seca" ou "branco".
4. **Selecionou-se as séries apropriadas.** As séries consideradas adequadas foram aquelas cuja soma de valores faltantes e de dados com classificação diferente de "real" não ultrapassou 10% do total. Além disso, apenas séries com dados consistentes foram usadas. Após essa seleção, restaram 208 estações.
5. **Verificação de valores fisicamente impossíveis.** Assim como no caso das estações do ONS, buscou-se valores negativos de vazões, mas nenhum foi encontrado. Destaca-se que valores extremos (*outliers*) não foram removidos.
6. **Identificou-se as estações mais próximas de PCHs ou CGHs.** Como o intuito da utilização das séries da ANA é representar a geração de PCHs e CGHs já existentes, comparou-se as coordenadas das estações com as coordenadas de PCHs e CGHs. Selecionou-se as estações que estavam a um raio de até 5km das usinas. Verificou-se também se as estações próximas dos aproveitamentos hidrelétricos, também estavam instaladas no mesmo rio. Ao fim desta etapa, 33 estações foram escolhidas. Todas as estações selecionadas não tinham valores faltantes, então a etapa de preenchimento de falhas não foi necessária.

As estações selecionadas são apresentadas na Figura 4.5. Mais informações sobre as estações selecionadas e as abreviações usadas são apresentadas no Apêndice A.



Bacias Hidrográficas - Departamento Nacional de Águas e Energia Elétrica(DNAEE) :

- | | |
|--|---|
| ■ ATLÂNTICO, TRECHO LESTE | ■ RIO PARANÁ |
| ■ ATLÂNTICO, TRECHO NORTE/NORDESTE | ■ RIO SÃO FRANCISCO |
| ■ ATLÂNTICO, TRECHO SUDESTE | ■ RIO TOCANTINS |
| ■ RIO AMAZONAS | ■ RIO URUGUAI |
- Estações da ANA

Figura 4.5: Localização geográfica das estações da ANA

Fonte: Elaboração própria através do ESRI ArcGIS online [51], com camada de regiões hidrográficas de [4]

4.1.2.5 Tratamento dos Dados de Demanda Elétrica

Os dados de demanda elétrica são disponibilizados no site do ONS [93]. A coleta e tratamento dos dados seguiu os seguintes passos:

- **Coletou-se os dados.** Foi feito o download dos dados para o período desejado. A

série inclui todo o sistema SE/CO, não há como refinar por estado.

- **Verificou-se a disponibilidade dos dados** Foi avaliado se havia dados faltantes e constatou-se que havia um valor faltante em cada ano, sempre no mês de outubro e à meia-noite. Provavelmente essa falha tem relação com o início do horário de verão, apesar de que não há nenhuma compensação dessa falta no mês de fevereiro. Como as séries de demanda só serão utilizadas à nível diário posteriormente (média das medições de cada dia), essas falhas horárias não foram preenchidas.
- **Verificou-se a existência de duplicatas.** No entanto, nenhum valor duplicado foi encontrado.

4.2 Cálculo das Séries de Potência

Como citado no início da Seção 4.1.1, este trabalho, como outros na literatura, não pôde utilizar dados de geração real de usinas e, desta forma, deverá estimar o comportamento da potência de saída, de acordo com as equações de potência teórica de cada tipo de usina, apresentadas pelas Equações 2.1, 2.2 e 2.3.

Para isso, podem-se considerar características específicas para cada forma de geração de eletricidade – através da escolha da potência nominal e do arranjo das usinas fictícias – e optar também por determinadas tecnologias, a partir da escolha do modelo de painel fotovoltaico e das turbinas eólicas e hídricas e assim, estimar os parâmetros das equações com base nessas seleções. Entretanto, este tipo de abordagem condiciona os resultados a tais tecnologias e cenários, o que pode ser indesejável em alguns casos. As regiões Sudeste e Centro-Oeste, objetos deste estudo, possuem um parque de energias renováveis intermitentes ainda pouco explorado e que deve ser expandido nos próximos anos. Assim, para apoiar decisões de novos empreendimentos, optou-se por não delimitar a fronteira tecnológica do trabalho, na tentativa de apresentar resultados não tendenciosos nesse sentido.

4.2.1 Geração Fotovoltaica

Considerando a potência de um painel fotovoltaico, expressa por meio da Equação 2.3, para um determinado empreendimento, sabe-se que a área das placas é uma constante, já que os painéis têm medidas fixas. A eficiência, como exposto na Seção 2.2.3, varia conforme a temperatura das placas, segundo cada equipamento. No entanto, de acordo

com [68] , pode-se assumir que a eficiência varia pouco o suficiente para ser considerada constante. Consequentemente, a única variável que contribui de forma significativa para a variação da potência gerada por um painel fotovoltaico é a radiação solar incidente sobre o mesmo. Assim, para a análise da geração fotovoltaica, utilizaram-se séries de radiação solar global, com dados de janeiro de 2007 a dezembro de 2011, em concordância com o período disponível das outras fontes.

Ressalta-se que o ideal seria diferenciar a radiação solar direta e difusa incidente sobre o painel, mas como esta informação não é disponibilizada pelas estações de medições, a radiação solar global será utilizada como um todo, conforme [37]. A variação do ângulo de incidência da radiação direta devido ao movimento de rotação da Terra, não foi considerada já que a utilização de sistemas de rastreamento permite que se tenha incidência perpendicular da radiação sobre as placas durante grande parte do tempo.

A Equação 2.3 seria, pois, reescrita como a Equação 4.1.

$$P_{FV} = K_{FV} R_S \quad (4.1)$$

em que:

P_{FV} = potência do sistema fotovoltaico

K_{FV} = produto da eficiência do painel pela área das placas, assumido como constante

R_S = radiação solar incidente sobre as placas

4.2.2 Geração Eólica

No que tange à potência de saída de um aerogerador, expressa por meio da Equação 2.2, segundo [68] , a densidade do ar, a área pela qual passa a coluna de vento (que depende das dimensões do equipamento) e a eficiência são valores que podem ser considerados aproximadamente constantes. Em geral, a velocidade medida em estações de coletas de dados não corresponde à velocidade na altura de instalação do rotor, fato este que pode ser contornado aproximando-se a velocidade na altura desejada por meio da Equação 4.2 [38].

$$V = V_0 \left(\frac{H}{H_0} \right)^n \quad (4.2)$$

em que:

V = velocidade do vento na altura que se pretende calcular

V_0 = velocidade do vento conhecida

H = altura desejada

H_0 = altura de onde se conhece a velocidade do vento

n = fator de rugosidade do terreno (depende do tipo de terreno e das construções e vegetação em volta)

De fato, a altura desejada e a altura da medição são valores constantes, assim como a rugosidade do terreno para um determinado local, ou seja, o termo $\left(\frac{H}{H_0}\right)^n$ pode ser entendido como constante, de modo que a relação entre V e V_0 é linear [68].

Logo, a Equação 2.2 pode ser reescrita como a Equação 4.3.

$$\begin{aligned}
 P_E &= K_1 u^3 \\
 &= K_1 \left(\left(\frac{H}{H_0} \right)^n u_0 \right)^3 \\
 &= K_1 \left(\frac{H}{H_0} \right)^{3n} u_0^3 \\
 &= K_1 K_2 u_0^3 \\
 \therefore P_E &= K_E u_0^3
 \end{aligned} \tag{4.3}$$

em que:

P_E = potência do aerogerador

K_1 = termo $\frac{1}{2}C_p A \rho$ da Equação 2.2, considerado constante

$K_2 = \left(\frac{H}{H_0}\right)^{3n}$, termo também assumido como constante

K_E = constante igual ao produto $K_1 K_2$

Observa-se que a variação da potência depende basicamente da variação da velocidade do vento ao cubo na altura conhecida.

Portanto, para a análise de geração eólica do presente trabalho, converteram-se as duas séries de velocidade do vento para velocidade do vento ao cubo, no período de janeiro de 2007 a dezembro de 2011.

4.2.3 Geração Hidrelétrica

Quanto à geração hidrelétrica, é expressa na Equação 2.1 a potência de saída de um empreendimento desta natureza. Analisando as variáveis envolvidas, o produto ρg , em que ρ é a densidade da água e g é a aceleração da gravidade, é o peso específico da água, cujo valor em determinado local varia, sobretudo, devido a variações de temperatura da água.

No contexto da geração hidrelétrica, a gama de temperaturas que um curso hídrico geralmente apresenta permite uma variação muito pequena do peso específico da água, de modo que este pode ser assumido como constante [95]. A altura h considerada para o cálculo da potência é a queda líquida, dada pela subtração das perdas hidráulicas ao longo do circuito adutor do valor da queda bruta (diferença de altura entre a tomada d'água e o canal de descarga) [20][62]. Apesar do valor da queda poder apresentar variações, com base em [56] [23], e [80], ele pode ser considerado constante neste tipo de estudo, especialmente por envolver PCHs, que apresentam pouca ou nenhuma capacidade de armazenagem de água. Desta forma, a Equação 2.1 (1) é reduzida à Equação 4.4.

$$P_H = K_H Q \quad (4.4)$$

em que:

P_H = potência hidrelétrica

K_H = constante

Q = vazão volumétrica da água

Em conclusão, para estudar a geração hidrelétrica a partir de PCHs/CGHs foram utilizadas as séries de vazões disponíveis, abrangendo o período de janeiro de 2007 a dezembro de 2011.

4.3 Avaliação de Correlações Lineares

Após a seleção e o preparo das séries para o estudo, avaliou-se a estrutura de correlação entre séries do mesmo grupo (fotovoltaico, eólico, hidrelétrico - ONS e hidrelétrico - ANA). As matrizes de correlação linear (coeficientes de Pearson) foram calculadas através do comando `cor()` do pacote `stats` do R.

No caso das séries eólicas e solares, com resolução horária e 14608 dados no total, foi feito o cálculo expresso na Equação 4.5 para todos os pares (x, y) de estações, de acordo

com a Equação 2.7.

$$r_{xy} = \frac{\sum_{i=1}^{14608} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{14608} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{14608} (y_i - \bar{y})^2}} \quad (4.5)$$

com a média calculada ao longo de toda a série, como na Equação 4.6.

$$\bar{x} = \frac{\sum_{i=1}^{14608} (x_i)}{14608} \quad (4.6)$$

Para as séries hidrelétricas, com resolução diária e 1826 dados, o cálculo foi feito como nas Equações 4.5 e 4.6, porém com o limite 1826 ao invés de 14608.

A fim de facilitar a visualização e a interpretação dos resultados, os valores foram plotados em um correlograma obtido pelo comando **corrplot** do pacote homônimo *corrplot*. Neste caso, ao invés de apresentar o valor da correlação entre duas séries x e y , o correlograma apresentará um círculo cujo raio dependerá do valor absoluto da correlação (quanto maior a correlação, maior o círculo) e cuja cor dependerá do valor: valores próximos de 1 são azulados, valores próximos de -1 são avermelhados e valores próximos de zero são brancos.

4.4 Análise de Fatores

4.4.1 Adequação das Séries

Para verificar se séries eram adequadas para a análise de fatores, realizaram-se as seguintes avaliações, além da análise de correlação linear descrita no item anterior.

- Comparação do número de variáveis ao número de observações
- Teste de Esfericidade de Bartlett
- Cálculo do KMO

Inicialmente, verificou-se quão grande era o número de observações em relação à quantidade de variáveis de cada série. Para serem consideradas aceitáveis, as séries deveriam ter no mínimo dez vezes mais observações do que variáveis [52].

Em seguida, o teste de esfericidade de Bartlett foi feito no RStudio através da implementação do código expresso no Algoritmo 1, de autoria do professor Shigenobu Aoki

[17].

No Algoritmo 1, procede-se da seguinte forma: Primeiramente, define-se o nome do teste, que será exibido junto aos resultados e lê-se o rótulo dos dados. Em seguida, a matriz de dados é atualizada retirando-se eventuais valores faltantes. Define-se o número de observações (n) através do número de linhas, e o número de variáveis (p), corresponde ao número de colunas. Feito isso, calcula-se a estatística de teste qui-quadrado, de acordo com a Equação 2.8. Os graus de liberdade são calculados de acordo com p e n . Em seguida, calcula-se o valor-p de acordo com a estatística calculada e seus respectivos graus de liberdade (a configuração “lower.tail=FALSE” indica que a probabilidade estimada é a de que o parâmetro calculado seja maior do que o estabelecido na hipótese nula). Por último, nomeia-se a estatística calculada e os graus de liberdade e retorna-se os resultados obtidos e as respectivas identificações.

Algoritmo 1 Bartlett.sphericity.test()

```

1: method <- "Teste de esfericidade de Bartlett"
2: data.name <- deparse(substitute(x))
3: x <- subset(x, complete.cases(x))
4: n <- nrow(x)
5: p <- ncol(x)
6: chisq <- (1-n+(2*p+5)/6)*log(det(cor(x)))
7: df <- p*(p-1)/2
8: p.value <- pchisq(chisq, df, lower.tail=FALSE)
9: names(chisq) <- "X-squared"
10: names(df) <- "df"
11: return(structure(list(statistic=chisq,parameter=df,p.value=p.value,
    method=method, data.name=data.name), class="hstest"))

```

De forma similar, o critério de adequação da amostra KMO foi implementado no R a partir do código disponibilizado pelo professor Shigenobu Aoki em seu site [16]. A função usada é apresentada no Algoritmo 2.

No Algoritmo 2, procede-se da seguinte maneira: Inicialmente, a matriz de dados é atualizada, retirando-se eventuais valores faltantes. Em seguida, calcula-se a matriz de correlações lineares entre as variáveis. Depois, cria-se uma matriz com os coeficientes de correlação ao quadrado. Para o cálculo das correlações parciais, calcula-se a matriz inversa à de correlação e extraem-se os elementos da diagonal principal dessa matriz inversa, adicionando-os a um vetor. Calcula-se, então, as correlações parciais ao quadrado, seguindo a Equação 2.10. A fim de não utilizar as diagonais principais das matrizes de correlação e de correlação parcial nos cálculos seguintes, anulam-se esses valores. Calcula-se o KMO, como na Equação 2.11, e o índice MSA de cada variável, como na Equação

2.9. Por fim, retornam-se os valores dos indicadores.

Algoritmo 2 `kmo()`

```

1: x <- subset(x, complete.cases(x))
2: r <- cor(x)
3: r2 <- r ^ 2
4: i <- solve(r)
5: d <- diag(i)
6: p2 <- (-i/sqrt(outer(d, d))) ^ 2
7: diag(r2) <- diag(p2) <- 0
8: KMO <- sum(r2)/(sum(r2)+sum(p2))
9: MSA <- colSums(r2)/(colSums(r2)+colSums(p2))
10: return(list(KMO=KMO, MSA=MSA))

```

4.4.2 Análise de Fatores

Constatada a adequação das amostras, partiu-se para a Análise de Fatores em si. Optou-se por realizá-la pelo método do componente principal e pelo método da Máxima Verossimilhança, a fim de comparar os resultados, já que a coincidência deles é um indicador da consistência da análise.

Para esta etapa, as séries foram padronizadas (normalização *z-score*), já que no caso das vazões, os dados provêm de rios de portes diferentes e, no caso dos dados eólicos, as estações de medição se encontram em alturas diferentes. A priori, os dados solares não precisariam ser padronizados, mas como todas as demais séries foram padronizadas, optou-se por realizar o mesmo procedimento nos dados solares para que a metodologia fosse a mesma em todos os casos. A padronização foi feita através da função `scale()` do R.

4.4.2.1 Definição do Número de Fatores.

Antes de conduzir a análise de fatores em si, os critérios de escolha do número de fatores apresentados na Seção 2.5.4 foram seguidos. Para isso, criou-se um algoritmo implementado de forma simples no RStudio. O algoritmo é apresentado no Apêndice B, e um diagrama esquemático com os passos seguidos e as funções utilizadas é apresentado na Figura 4.6.

A partir das informações geradas, pode-se avaliar quantos autovalores são maiores ou iguais a uma unidade (critério da raiz latente), verificar o número de fatores sugerido

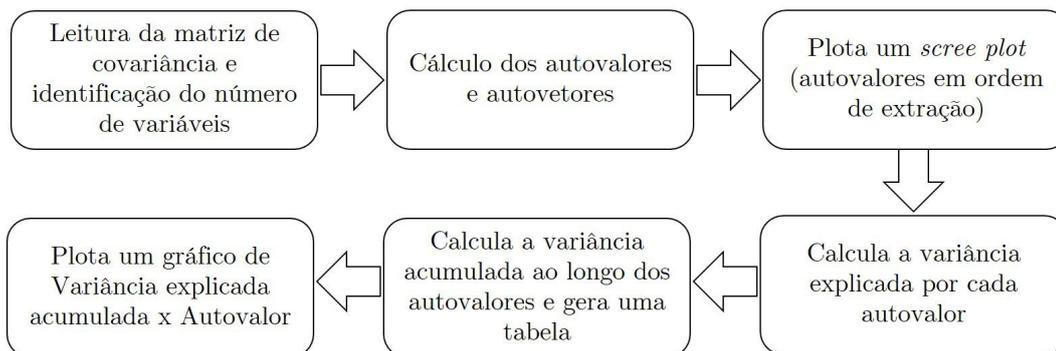


Figura 4.6: Diagrama esquemático do algoritmo de seleção de fatores.

Fonte: Autora

pelo teste *scree* e, a partir desses dois métodos, verificar se há consistência teórica e se a variância explicada a ser obtida será satisfatória.

Definido o número de fatores mais adequado, realiza-se então as análises fatoriais. É importante destacar que após as análises pode ser necessário testar outros números de fatores plausíveis.

4.4.2.2 Análise de Fatores por Componente Principal

A Análise pelo MCP foi feita através da implementação de um algoritmo no RStudio. O Algoritmo implementado se encontra no Apêndice C. As etapas principais do algoritmo, cujas entradas são a matriz de covariância dos dados e o número de fatores escolhido, são as seguintes:

1. Calcular os autovalores e autovetores e armazená-los;
2. Calcular a variância explicada e a variância explicada acumulada. A variância explicada por cada autovalor é dada pela razão entre o autovalor e o traço da matriz de covariância (soma dos elementos da diagonal principal);
3. Calcular as cargas. Para isso, cria-se um vetor com as raízes dos autovalores e então, cada elemento desse vetor multiplicará a coluna correspondente da matriz de autovetores, gerando a matriz de cargas, como na Equação 2.26;
4. Selecionar as cargas correspondentes aos m fatores. A redução da dimensão poderia ter sido feita antes mesmo do item 3, porém esta etapa foi realizada neste momento por conveniência;

5. Calcular as comunalidades. Cria-se uma nova matriz com as cargas ao quadrado e a comunalidade de cada variável é dada pela soma dos elementos da linha correspondente;
6. Calcular as variâncias específicas. Para isso, subtrai-se as comunalidades da diagonal principal da matriz de covariância (Equação 2.27);
7. Calcular a matriz residual. Primeiramente, calcula-se o produto da matriz de cargas pela sua transposta. Em seguida, soma-se essa nova matriz à matriz diagonal de variâncias específicas. A matriz residual é calculada subtraindo esta última matriz da matriz de covariâncias;
8. A última etapa consiste em calcular o resíduo máximo da matriz residual.

O tipo de rotação escolhido foi o Varimax. A rotação para o MCP, foi feita utilizando a função `varimax()` do pacote `stats` do R.

4.4.2.3 Análise de Fatores pelo Método da Máxima Verossimilhança

Para o MMV, utilizou-se a função `factanal()`, também do pacote `stats` do R. Para utilizar a função, deve-se inserir ou a matriz de covariância, ou os próprios dados, além do número de fatores desejados. Os resultados apresentados já são rotacionados.

4.5 Avaliação de complementaridade

Após a definição dos grupos de fontes, a série histórica representativa de cada grupo foi calculada realizando-se a média entre as potências elétricas das estações formadoras do grupo.

Durante a análise de fatores, as séries foram padronizadas pelo método *z-score*, isto é, foi subtraído de cada potência calculada, o valor médio da série correspondente e, em seguida, dividiu-se esse valor resultante pelo desvio-padrão da série. Assim, todas as séries passaram a ter valor médio zero e desvio-padrão de uma unidade. Isso era importante para garantir que nenhuma série fosse “privilegiada” durante a análise de fatores.

Nesta etapa, no entanto, o importante é que o comportamento das séries de potência, ainda que adimensional, seja correspondente ao das séries originais. Isso porque, na prática, as constantes K_H , K_E e K_{FV} , definidas nas Equações 4.4, 4.3 e 4.1 tornam

possível a comparação entre a potência de cada fonte (P_H , P_E e P_{FV}), pois assim, todas as séries seguem a mesma unidade de medida [kW]. No entanto, se apenas o comportamento ao longo do tempo for de interesse do estudo e não a magnitude das séries em si¹, pode-se optar por algum método de normalização que deixe as séries na mesma escala sem alterar seu perfil de variação, de forma semelhante ao que aconteceria atribuindo-se valores a K_H , K_E e K_{FV} .

Para isso, seguiu-se o procedimento adotado pelos autores de [65] e pelos de [56], e normalizou-se as séries originais dividindo cada valor de potência pela média da série correspondente (valor médio de toda a série). Um exemplo que compara a padronização *z-score* e a normalização pela média à série original é apresentado na Figura 4.7.

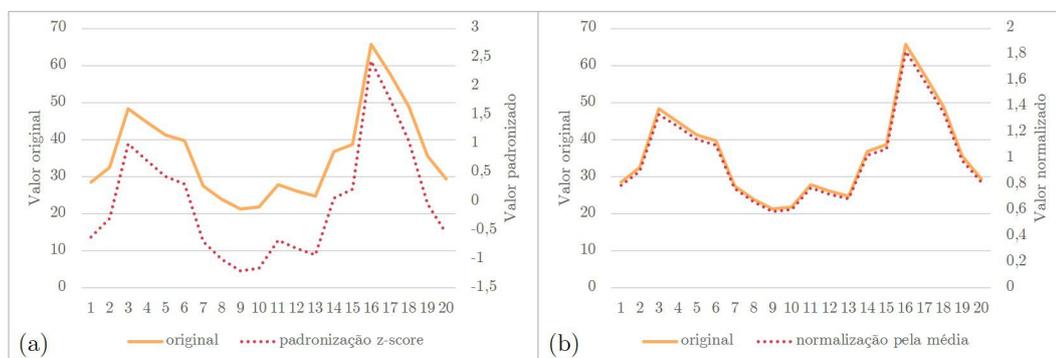


Figura 4.7: Comparação entre o comportamento original da série e (a) a série com padronização *z-score* e (b) a série dividida por seu valor médio.

Fonte: Autora

Logo, as séries históricas dos grupos foram calculadas com base na normalização pela média (as séries foram normalizadas antes de serem agrupadas).

A primeira etapa da avaliação de complementaridade consiste em avaliar os recursos eólico e fotovoltaico a nível intra-diário. A geração hidrelétrica não é incluída nesta etapa, pois as séries são disponibilizadas em escala diária. Os passos seguidos nesta etapa são os seguintes:

1. Cálculo da correlação horária entre o grupo fotovoltaico e as estações eólicas
2. Avaliação da correlação durante o dia e a noite
3. Avaliação do comportamento médio ao longo do dia: calculou-se o valor médio de cada série para cada hora disponível

¹Neste trabalho o interesse não é definir valores absolutos (kW) de participação de cada fonte, mas sim, participações relativas de cada grupo (%)

Todas as matrizes de correlação foram calculadas através do comando `cor()` do pacote `stats` do R e apresentadas em forma de correlograma gerado pelo comando `corrplot` do pacote `corrplot`.

A avaliação a nível diário engloba as três fontes renováveis. Para isso, calculou-se a média dos valores horários para cada dia, no caso eólico e fotovoltaico, para criar as séries diárias. Em seguida, seguiu-se as seguintes etapas:

1. Cálculo da correlação entre o grupo fotovoltaico, as estações eólicas e os grupos hidrelétricos
2. Avaliação da correlação de acordo com as estações do ano
3. Avaliação do comportamento médio ao longo do ano: calculou-se o valor médio de cada série para cada dia do ano, com base nos cinco valores disponíveis (as séries se estendem por cinco anos)

O cálculo e a visualização das matrizes de correlação foram feitos da mesma forma que na análise horária.

4.6 Otimização

Feita a avaliação das correlações entre os grupos de fontes, buscou-se encontrar a participação ótima de cada um no *mix* de fontes renováveis. A potência total do *mix* renovável no instante j é calculada por meio da Equação 4.7.

$$P_j = \sum_{i=1}^n w_i p_{ij} = w_1 p_{1j} + w_2 p_{2j} + \dots + w_n p_{nj} \quad (4.7)$$

em que:

P_j = potência do *mix* renovável no instante j

p_{ij} = potência da grupo i no instante j

w_i = participação (%) da fonte i na potência total do *mix* renovável

n = número de grupos de fontes renováveis

m = número de instantes de tempo considerados

As variáveis de decisão são os w_i . No entanto, a definição da função objetivo depende se a demanda por energia elétrica será considerada ou não. Para isso, definiu-se dois

problemas: um que não levava em conta a série histórica de demanda (Formulação I) e outro que sim (Formulação II).

4.6.1 Formulação I

4.6.1.1 Modelagem

No caso em que não se considera a demanda, o objetivo é ter um *mix* renovável estável, ou seja, menos intermitente o possível. Para isso, a função objetivo deverá buscar minimizar a variância da série de potência elétrica do *mix* renovável.

Observando a Equação 4.7, percebe-se que a potência do *mix* renovável é uma combinação linear da participação dos grupos (w_i) e da potência do grupo correspondente (p_i). Como apenas a potência do grupo varia no tempo, a variância da potência renovável total é calculada por meio da Equação 4.8 [35].

$$V(P) = \sum_{i=1}^n \sum_{k=1}^n w_i w_k Cov(p_i, p_k) \quad (4.8)$$

em que:

- P = potência do *mix* renovável
- p_i = potência da grupo i
- w_i = participação (%) do grupo i na potência total do *mix* renovável
- $Cov(p_i, p_k)$ = covariância entre as séries de potência dos grupos i e k

Dessa forma, o problema é modelado de acordo com a Equação 4.9.

$$FO : \text{Minimizar } z = \sum_{i=1}^n \sum_{k=1}^n w_i w_k Cov(p_i, p_k) \quad (4.9)$$

sujeito a

$$\sum_{i=1}^n w_i = 1$$

$$w_i \geq 0, \forall i \in [1, n]$$

Na linha da função objetivo, busca-se minimizar a variância da série histórica de

potência elétrica do *mix* renovável, como apresentado na Equação 4.7; a primeira restrição assegura que a soma dos pesos percentuais seja igual a uma unidade (100%); por fim, tem-se as restrições de não-negatividade dos pesos.

De fato, os pesos não podem ser maiores do que uma unidade, mas essa restrição seria redundante, pois já foi estabelecido que a soma deles deve ser igual a uma unidade e que são não-negativos.

4.6.1.2 Resolução

Observando a função objetivo, pode-se perceber que o produto entre as variáveis de decisão ($w_i w_k$) torna-a uma função quadrática. Já as restrições são lineares. Consequentemente, o modelo é classificado como um problema de programação quadrática. O modelo deverá ser resolvido pela função **solve.QP()** do pacote *quadprog* do R, que implementa um método de projeção de gradiente para resolução do problema de programação quadrática, de acordo com o apresentado em [61]. O código da função **solve.QP()** é apresentado no Anexo A.

A função lê os dados do problema escrito de forma matricial, dessa forma pode-se reescrever o problema de acordo com a Equação 4.10.

$$FO : \text{Minimizar } z = \mathbf{w}^\top \Sigma \mathbf{w} \quad (4.10)$$

sujeito a

$$\begin{aligned} \mathbf{R}_{1 \times n} \mathbf{w}_{n \times 1} &= 1 \\ \mathbf{I}_{n \times n}^\top \mathbf{w}_{n \times 1} &\geq \mathbf{0}_{n \times 1} \end{aligned}$$

em que:

\mathbf{w} : vetor coluna com as variáveis w_i

Σ : matriz de covariância entre as séries históricas dos grupos

\mathbf{R} : vetor linha composto por “1” em cada posição

\mathbf{I} : matriz identidade de dimensão n

$\mathbf{0}$: vetor nulo

Ressalta-se que é necessário verificar se o problema é do tipo convexo antes de implementar o método citado, o que significa que a matriz envolvida no termo quadrático deve ser positiva definida, ou seja, seus autovalores devem ser estritamente positivos [61][98]. Essa condição foi verificada e o problema se adéqua a este método de solução.

O Apêndice D demonstra que a função objetivo na forma matricial apresentada é equivalente à apresentada no modelo da Seção 4.6.1.1.

4.6.2 Formulação II

4.6.3 Modelagem

Quando a demanda elétrica é considerada, busca-se um *mix* renovável cuja geração elétrica se aproxime ao máximo do comportamento da demanda, ou seja, deseja-se minimizar a diferença entre demanda e geração.

Seja c_j a demanda por eletricidade (carga elétrica) no dia j e P_j a potência do *mix* renovável definida na Equação 4.7. Busca-se minimizar a soma das diferenças entre elas, $\sum_j (c_j - P_j)$. No entanto, o termo $(c_j - P_j)$ será ora positivo ora negativo, de forma que diferenças elevadas com sinais opostos serão compensadas e a minimização não gerará um resultado coerente. Por isso, na definição do modelo, considerar-se-á o erro quadrático, ou seja o somatório das diferenças ao quadrado: $\sum_j (c_j - P_j)^2$. Logo, apresenta-se a função objetivo como a Equação 4.11.

$$\begin{aligned}
 z &= \sum_{j=1}^m (c_j - P_j)^2 \\
 &= \sum_{j=1}^m \left(c_j - \sum_{i=1}^n w_i p_{ij} \right)^2 \\
 &= \sum_{j=1}^m \left[c_j^2 - 2c_j \sum_{i=1}^n (w_i p_{ij}) + \sum_{i=1}^n (w_i p_{ij})^2 \right] \\
 \therefore z &= \underbrace{\sum_{j=1}^m c_j^2}_{(a)} - 2 \underbrace{\sum_{j=1}^m \sum_{i=1}^n (c_j w_i p_{ij})}_{(b)} + \underbrace{\sum_{j=1}^m \left(\sum_{i=1}^n w_i p_{ij} \right)^2}_{(c)} \quad (4.11)
 \end{aligned}$$

O primeiro termo da Equação 4.11 (a) representa um termo constante, já que não

depende das variáveis de decisão w_i . Já o segundo termo (b), depende linearmente dos w_i . Por fim, o terceiro termo (c) depende das variáveis de decisão ao quadrado. Dessa forma, a FO é uma função de segundo grau.

O modelo é, então, expresso de acordo com a Equação 4.12.

$$FO : \text{Minimizar } z = \sum_{j=1}^m c_j^2 - 2 \sum_{j=1}^m \sum_{i=1}^n (c_j w_i p_{ij}) + \sum_{j=1}^m \left(\sum_{i=1}^n w_i p_{ij} \right)^2 \quad (4.12)$$

sujeito a

$$\sum_{i=1}^n w_i = 1$$

$$w_i \geq 0, \forall i \in [1, n]$$

As restrições nesta formulação são as mesmas da Formulação I: deve-se garantir que os pesos sejam não negativos (segunda restrição) e que a soma deles seja de uma unidade, representando 100% da potência do *mix* renovável (primeira restrição).

4.6.3.1 Resolução

A Formulação II também se caracteriza como um problema de programação quadrática, por ter função objetivo quadrática e restrições lineares. Assim como na Formulação I, o problema será resolvido pela função `solve.QP()` do pacote *quadprog* do R, cujo código é apresentado no Anexo A.

O modelo da Formulação II é reescrito na forma matricial na Equação 4.13.

$$FO : \text{Minimizar } z = C - 2\mathbf{c}^\top \mathbf{A}\mathbf{w} + \mathbf{w}^\top \mathbf{B}\mathbf{w} \quad (4.13)$$

sujeito a

$$\mathbf{R}_{1 \times n} \mathbf{w}_{n \times 1} = 1$$

$$\mathbf{I}_{n \times n}^\top \mathbf{w}_{n \times 1} \geq \mathbf{0}_{n \times 1}$$

em que:

C : soma das cargas ao quadrado ao longo do período considerado (escalar)

w : vetor coluna com as variáveis w_i

c : vetor com os valores de demanda ao longo do tempo

A : matriz com as séries históricas de cada grupo formando suas colunas

B : matriz simétrica. Na diagonal principal, encontram-se as somas no tempo da potência elétrica ao quadrado de cada grupo i ($\sum_j p_{ij}^2$). Os demais elementos são a soma no tempo dos produtos entre o grupo i e o grupo k ($\sum_j p_{ij} p_{kj}$).

R : vetor linha composto por “1” em cada posição

I : matriz identidade de dimensão n

$\mathbf{0}$: vetor nulo

A matriz \mathbf{B} é positiva definida, o que possibilita a utilização do método utilizado pela função `solve.QP()`.

O Apêndice E demonstra que a função objetivo na forma matricial apresentada é equivalente à apresentada no modelo da Seção 4.6.3.

Capítulo 5

Resultados

Neste capítulo são apresentados os resultados dos métodos descritos no Capítulo 4 e a interpretação deles.

5.1 Correlações Lineares

As correlações lineares entre as séries de vazões da ANA são apresentadas na Figura 5.1.

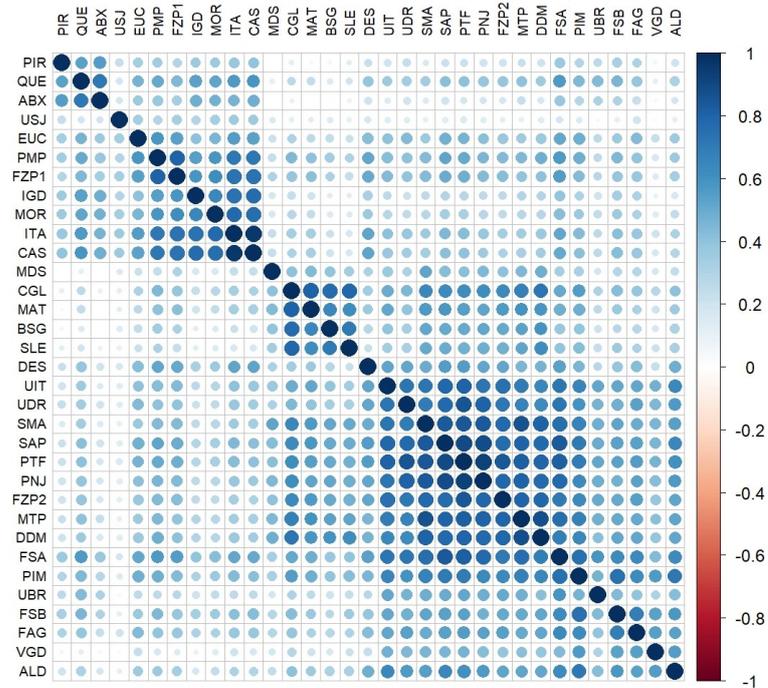


Figura 5.1: Correlações entre as estações da ANA
Fonte: Autora

Percebe-se um equilíbrio entre o número de correlações positivas fracas e as positivas moderadas e altas. Conseqüentemente, pode-se verificar a formação de grupos de estações mais bem correlacionadas (“quadrados” em torno da diagonal principal). O grande número de correlações maiores que 0,3 sugerem que as séries podem ser adequadas à análise de fatores.

As correlações lineares entre as séries de vazões do ONS são apresentadas na Figura 5.2.

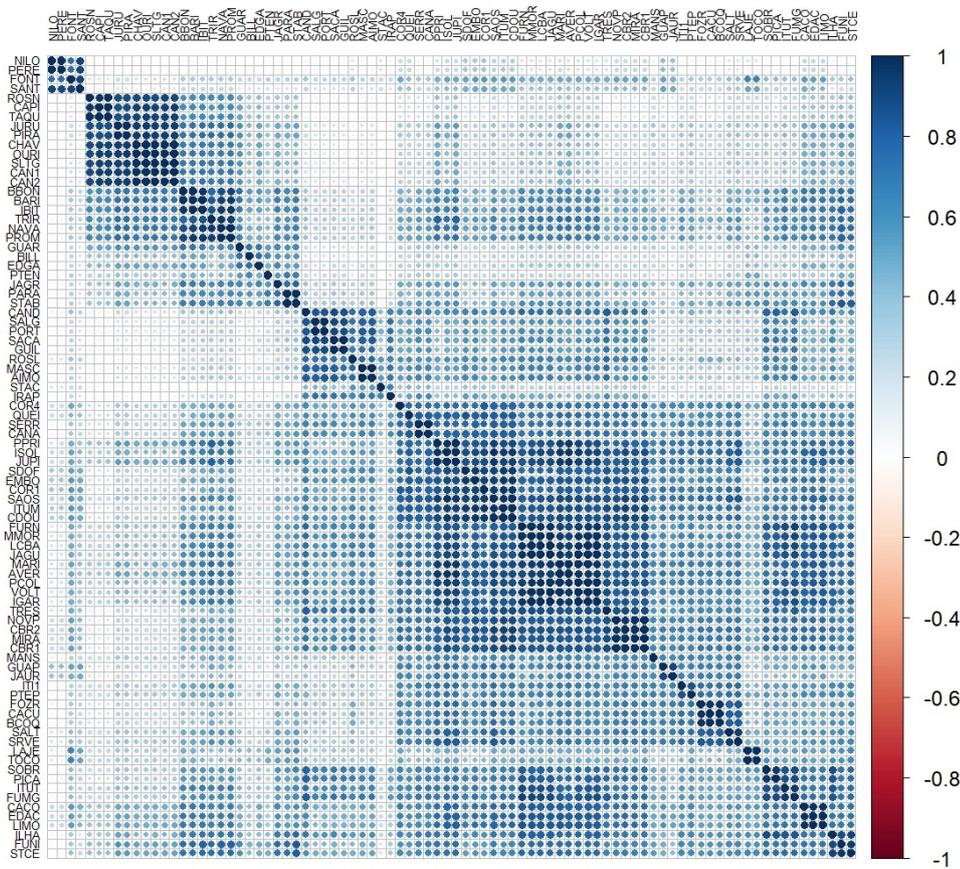


Figura 5.2: Correlações entre as estações do ONS

Fonte: Autora

As correlações entre as séries do ONS têm distribuição parecida com as da ANA. Há pouquíssimos casos de variáveis não correlacionadas e o equilíbrio entre as correlações mais fracas e mais fortes gera grupos mais bem correlacionados. Aparentemente, os dados são aptos à análise de fatores.

As correlações lineares entre as séries solares são apresentadas na Figura 5.3.

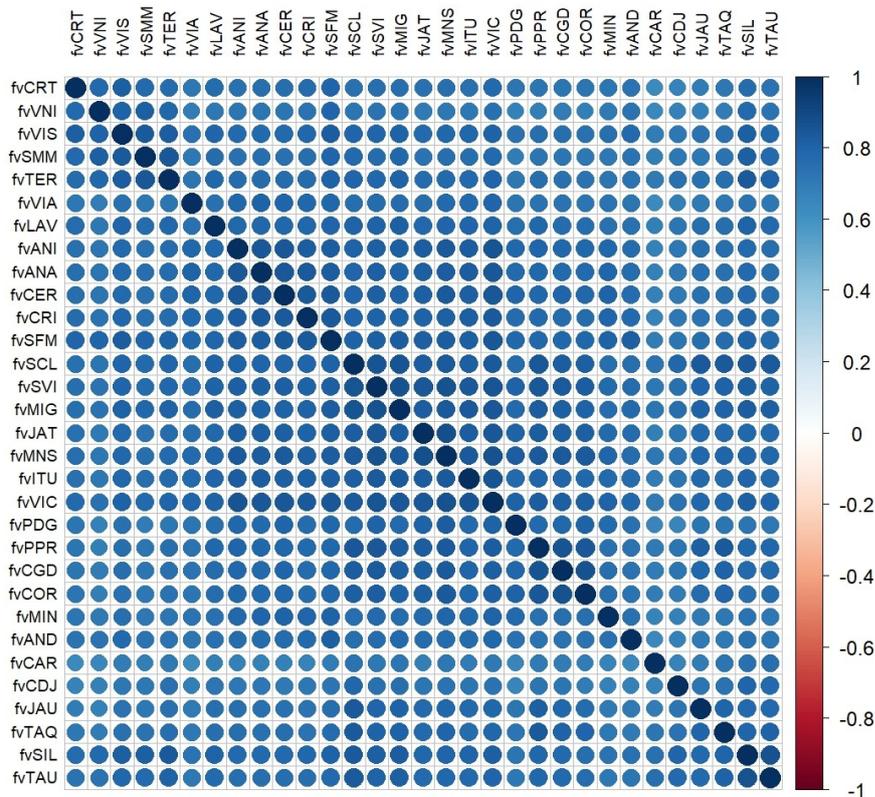


Figura 5.3: Correlações entre as estações solares do INPE

Fonte: Autora

Diferentemente dos casos de vazões, as correlações das estações solares são majoritariamente positivas moderadas ou fortes e se distribuem de forma razoavelmente homogênea, de modo que a existência de grupos não é facilmente identificada. A grande quantidade de correlações fortes não contraindica a análise fatorial.

Como é sabido que a radiação solar tem ciclo diário característico, segmentou-se a análise para os períodos da noite e do dia. Os resultados são apresentados na Figura 5.4.

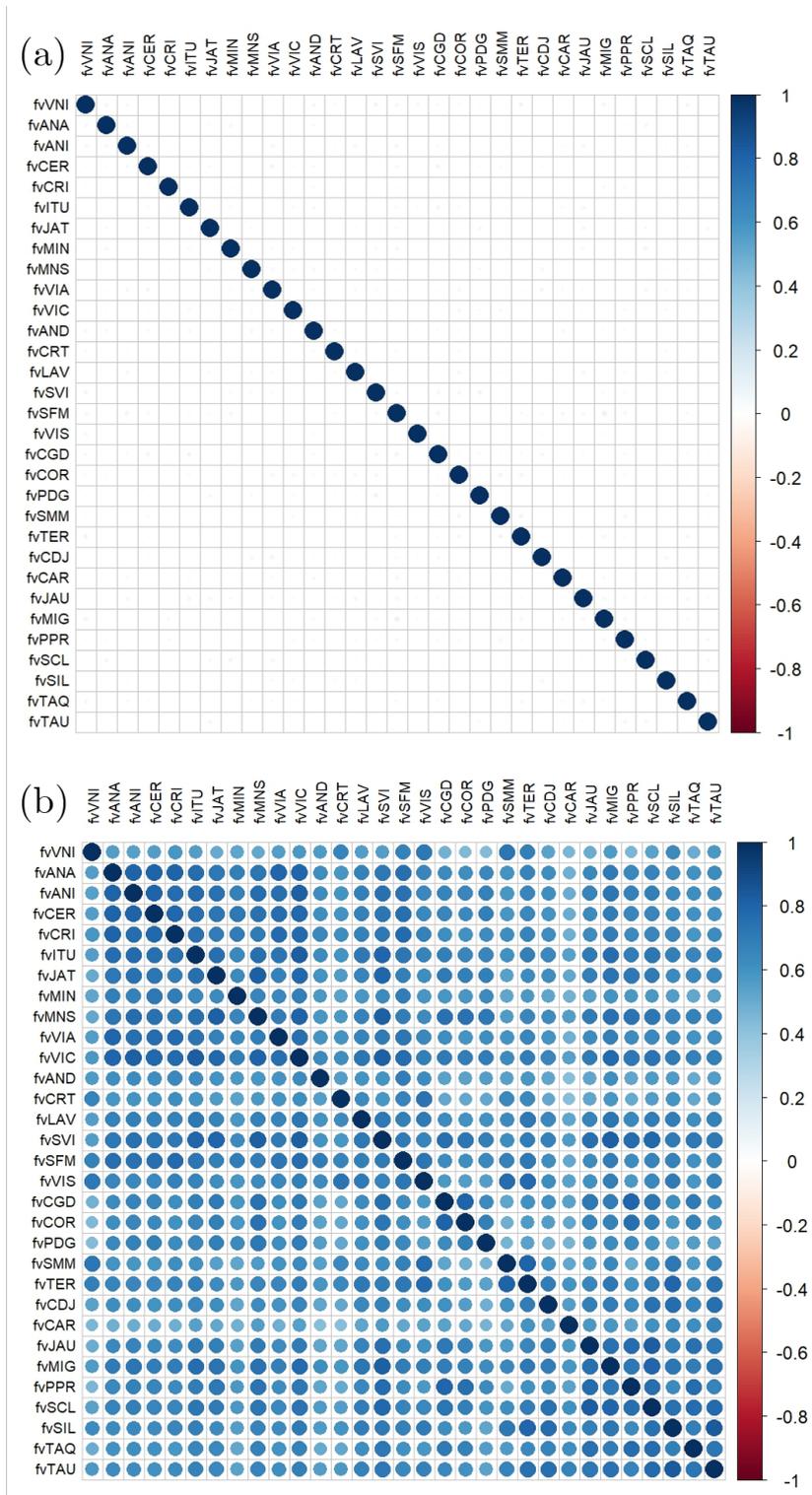


Figura 5.4: Correlações entre as estações solares do INPE durante a noite (a) e o dia (b)

Fonte: Autora

Como a variação de valores é pequena durante a noite, qualquer pequena alteração, tem um peso importante, o que justifica a matriz sem correlações significativas no período noturno. Já a matriz diurna, apesar de não tão forte como a total da Figura 5.3, ainda assim tem muitas correlações moderadas e fortes o que reforça a conclusão de que o recurso solar tem comportamento uniforme ao longo das regiões consideradas, mesmo com a eventual influência de nuvens, precipitação e altitude.

As correlações lineares entre as séries eólicas são apresentadas na Figura 5.5.

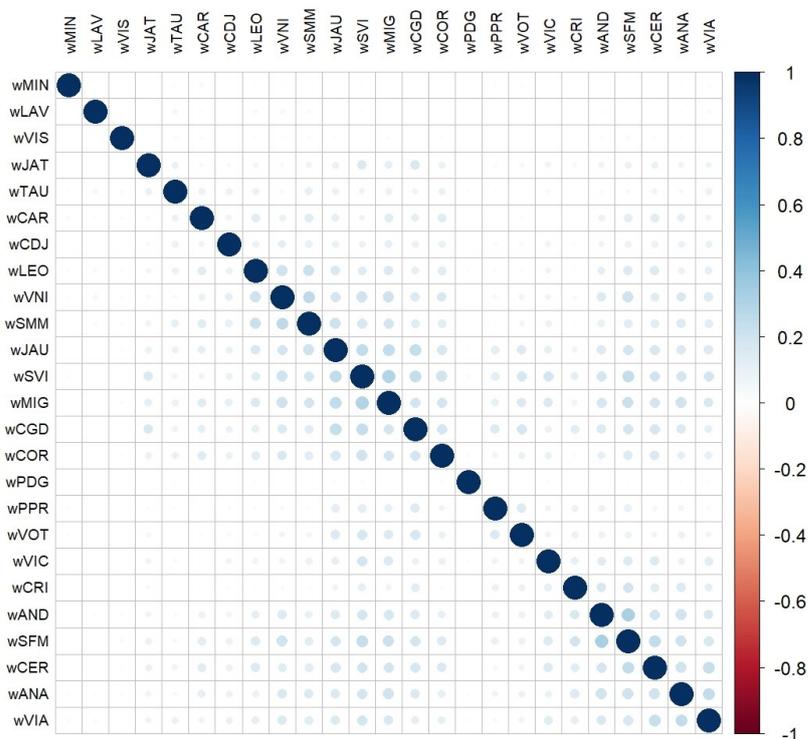


Figura 5.5: Correlações entre as estações eólicas do INPE

Fonte: Autora

Percebe-se que as estações eólicas têm correlações positivas fracas entre si ou mesmo não estão correlacionadas. O resultado ilustrado na Figura 5.5 indica que existe uma forte variabilidade geográfica do recurso eólico. Por um lado, isso pode dificultar o agrupamento das séries, mas por outro, é um indicador de que o recurso eólico tem potencial para complementar a si mesmo quando explorado em localidades diferentes na região.

5.2 Pré-análise

Com relação à análise de número de observações e quantidade de variáveis, o grupo com a menor proporção foi o de vazões do ONS, com 21 vezes mais observações do que variáveis. Dessa forma, todos os grupos excedem significativamente o mínimo de 10 vezes mais observações proposto por [64]. As proporções de cada grupo são apresentadas na Tabela 5.1.

| | ANA | ONS | Solar | Eólico |
|-----------------------|-------|-------|--------|--------|
| Variáveis | 33 | 86 | 31 | 25 |
| Observações | 1826 | 1826 | 14608 | 14608 |
| Observações/Variáveis | 55,33 | 21,23 | 471,23 | 584,32 |

Tabela 5.1: Observações versus Número de Variáveis

As informações dos testes de esfericidade de Bartlett são mostradas na Tabela 5.2. Estão incluídos os valores das estatísticas de cada teste (χ^2) com seus respectivos graus de liberdade. O valor-p é calculado de acordo com eles e valores pequenos indicam que a hipótese nula deve ser rejeitada. Em todos os casos, as hipóteses nulas de que as matrizes de correlação não têm correlações significantes foram rejeitadas para qualquer nível de significância ($1-\alpha$) razoável. Deste modo, todos os grupos foram aprovados para a análise fatorial pelo teste.

| | ANA | ONS | Solar | Eólico |
|--------------------|----------|----------|----------|----------|
| χ^2 | 61.593 | 581.700 | 65.6310 | 26.536 |
| graus de liberdade | 528 | 3655 | 465 | 300 |
| valor-p | <2,2e-16 | <2,2e-16 | <2,2e-16 | <2,2e-16 |

Tabela 5.2: Resultados do teste de esfericidade de Bartlett

Ao calcular as medidas de adequação de amostra, o código não conseguiu obter uma solução para os dados de vazões do ONS e informou que a matriz era singular. Como essa singularidade pode ser fruto de dependência linear entre as variáveis, avaliou-se mais detalhadamente as séries e constatou-se que duas delas eram perfeitamente iguais - as estações de Camargos (CAMA) e Itutinga (ITUT). Removeu-se a série Camargos e manteve-se a de Itutinga. Após a alteração, as análises foram refeitas¹ e as medidas de adequação da amostra foram calculadas. Os valores de KMO para cada grupo são apresentados na Tabela 5.3.

| | ANA | ONS | Solar | Eólico |
|-----|------|------|-------|--------|
| KMO | 0,95 | 0,96 | 0,99 | 0,87 |

Tabela 5.3: Valores do KMO para cada fonte

¹Os resultados como apresentados anteriormente na Figura 5.2 e na Tabela 5.2 foram calculados após a remoção de uma das séries

Apesar do recurso eólico apresentar um KMO mais baixo que os demais, ainda assim é um valor bastante satisfatório. Os resultados do índice MSA não serão apresentados aqui, por serem muitos valores. No entanto, a Tabela 5.4 apresenta um resumo dos resultados.

| | ANA | ONS | Solar | Eólico |
|--------|-------|-------|-------|--------|
| Mínimo | 0,885 | 0,776 | 0,989 | 0,555 |
| Médio | 0,947 | 0,95 | 0,994 | 0,817 |
| Máximo | 0,974 | 0,994 | 0,997 | 0,898 |

Tabela 5.4: Resumo dos valores de MSA de cada fonte

As séries eólicas, assim como na análise de correlação linear e no cálculo do KMO, foram as que apresentaram pior desempenho. No entanto, nenhuma série chegou a apresentar MSA menor que 0,5. Dessa forma, todos os quatro grupos foram aprovados para a análise de fatores pelo teste de Bartlett e pela medida de adequação da amostra.

5.3 Análise de Fatores

5.3.1 Vazões da ANA

A primeira etapa da análise de fatores resultou no valor dos autovalores e na variância explicada por eles. Os gráficos gerados para os dados da ANA são mostrados na Figura 5.6. O gráfico superior é o *scree plot* e mostra os autovalores por ordem de extração, enquanto o gráfico inferior mostra o acúmulo da variância ao longo dos autovalores.

Percebe-se que a variância explicada por cada autovalor é significativamente diferente entre primeiros autovalores e depois começa a ter variações pequenas por volta do sétimo autovalor. Os cinco primeiros autovalores têm valor acima de uma unidade. Pode-se notar também que 80% da variância é explicada pouco antes da marca de 10 fatores. De fato, do cálculo da variância acumulada vem que 8 fatores explicam 79,6% da variância, 9 explicam 81,4% e 10 explicam 83,1%. Após essas observações, optou-se por prosseguir a análise com 9 fatores, por conta da variância explicada.

Calculou-se a matriz de cargas e pôde-se perceber a existência de um fator geral, o que sugeriu a necessidade de se rotacionar a matriz. Após a rotação, as cargas se distribuíram melhor e os valores aceitáveis (maiores que 0,4) foram destacados. Procedeu-se, então a análise de fatores pelo método da máxima verossimilhança e as cargas significativas também foram destacadas. No entanto, ao comparar os métodos, percebeu-se que dois fatores não seriam utilizados e que duas variáveis ficariam isoladas (MDS e USJ).

Decidiu-se então tentar uma análise com sete fatores. Nesse caso, apesar de todos os

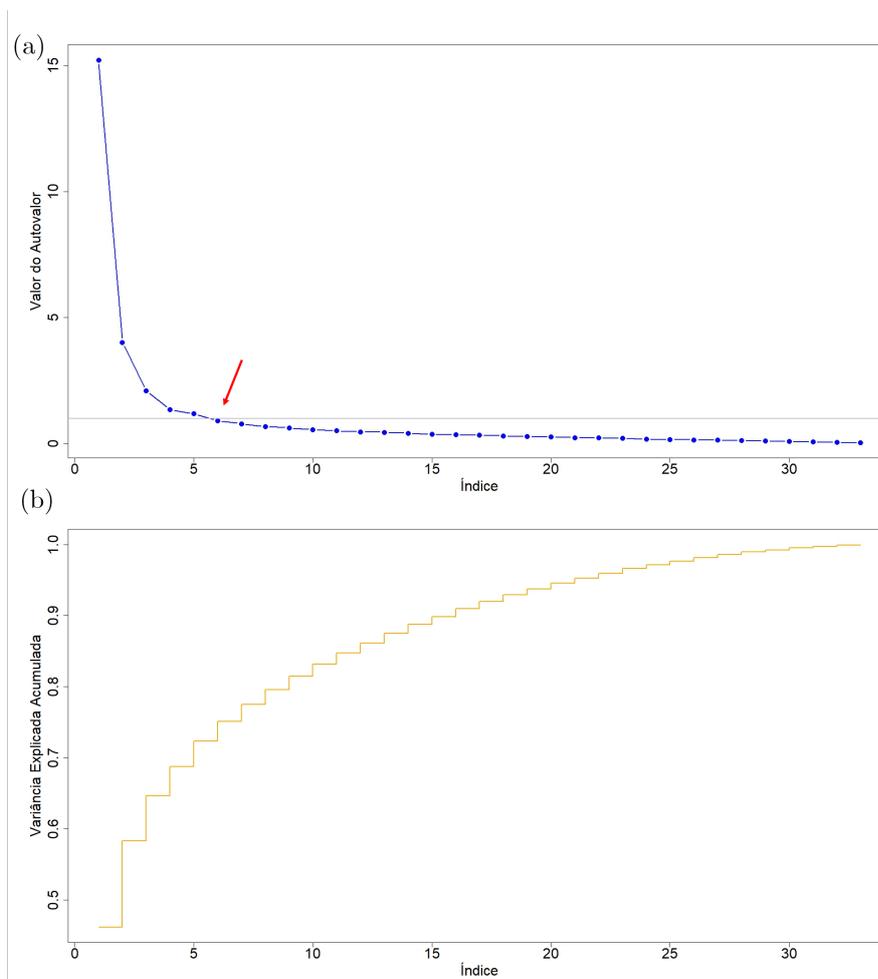


Figura 5.6: Resultado da análise de autovalores para os dados da Ana (cada índice representa um autovalor). (a) Autovalor por ordem de extração. (b) Variância acumulada ao longo dos autovalores.

Fonte: Autora

fatores terem cargas aceitáveis, dois deles se dedicaram a uma estação cada (novamente MDS e USJ), por ambos os métodos. Dessa forma, testou-se a análise com cinco fatores, como indicado pelo teste da raiz latente.

Na análise com cinco fatores, o MMV manteve as duas estações isoladas, porém o MCP as uniu a outros grupos. As demais estações coincidiram pelos dois métodos. Os cinco fatores foram, então, mantidos e optou-se por seguir o MCP para as estações USJ e MDS.

O gráfico que destaca como as estações foram agrupadas é apresentado na Figura 5.7.

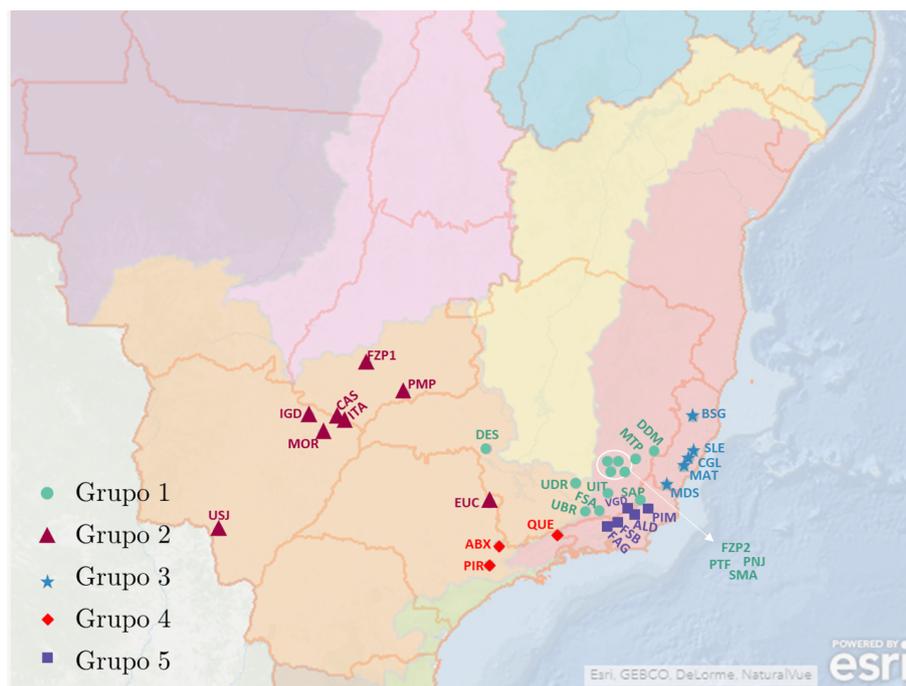


Figura 5.7: Análise com cinco fatores para os dados da ANA.

5.3.2 Vazões do ONS

Para definir o número de fatores, avaliaram-se os autovalores, o *scree plot* e a variância explicado. Essas informações são apresentadas na Figura 5.8.

Quanto à raiz latente, os oito primeiros autovalores são maiores do que uma unidade. A "quebra" do *scree plot* ocorre também entre o oitavo e o décimo autovalores. A variância explicada acumulada passa de 90% com 11 autovalores e precisa de cinco para explicar mais de 80% da variância.

Testou-se inicialmente a análise com oito fatores, mas houve uma quantidade significativa de discrepâncias entre os métodos e várias variáveis tinham cargas significantes em mais de um fator. Os resultados para seis fatores foram mais satisfatórios. Os casos de variáveis com cargas significantes em mais de um fator diminuíram consideravelmente e quando ocorriam em um método, o mesmo não ocorria pelo outro, o que facilitava o "desempate". Ressalta-se que, no entanto, o MMV apresentou a tendência de utilizar apenas cinco fatores, enquanto o MCP utilizou claramente seis. Testou-se a análise com 5 fatores, mas nesse caso houve mais discrepância entre os métodos. Manteve-se, então, a análise com seis fatores, dando preferência ao MCP para a definição dos fatores.

O gráfico que destaca como as estações foram agrupadas é apresentado na Figura 5.9.

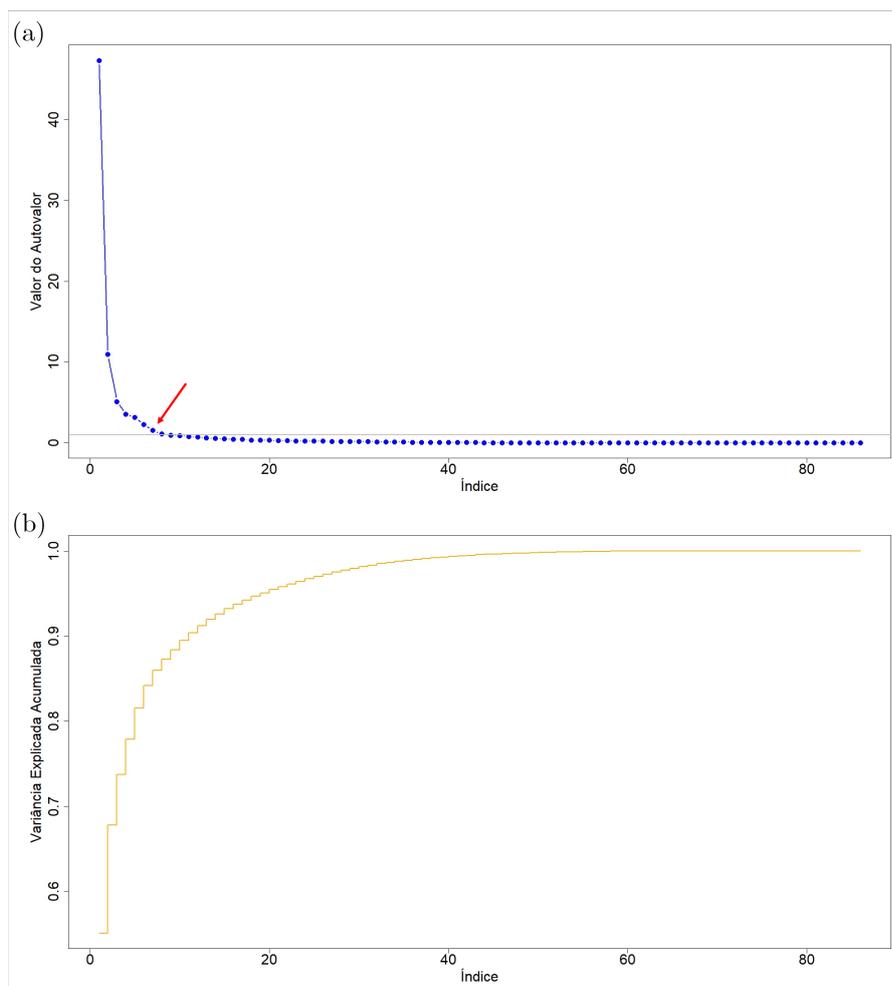


Figura 5.8: Resultado da análise de autovalores para os dados de vazões do ONS (cada índice representa um autovalor). (a) Autovalor por ordem de extração. (b) Variância acumulada ao longo dos autovalores.

Fonte: Autora

5.3.3 Dados de Radiação Solar

Novamente a análise dos autovalores e da variância explicada foi usada para se definir o número provável de fatores. Apresenta-se, por meio da Figura 5.10, o *scree plot* e o gráfico de variância acumulada para as séries de radiação.

Pode-se perceber que o *scree plot* apresenta um gráfico bem distinto dos de vazões. Apenas o primeiro autovalor é maior do que uma unidade e os demais são pequenos em relação a ele. A "quebra" no gráfico ocorre no segundo autovalor, então, por esse critério, três fatores poderiam ser usados. Isso se reflete no gráfico de variância também. São necessários apenas dois autovalores para ultrapassar 80% de variância explicada.

Inicialmente, procedeu-se a análise com três fatores, porém ocorreram muitos casos

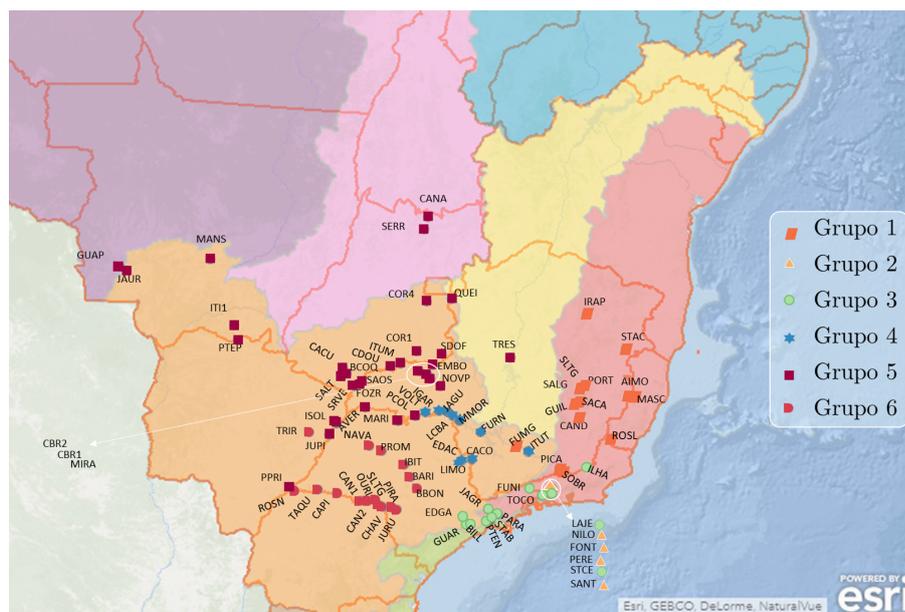


Figura 5.9: Análise com seis fatores para os dados de vazões do ONS.

de cargas significativas em mais de um fator para uma única variável. Isso se repetiu na análise com dois fatores (67% das variáveis apresentaram cargas maiores que 0,5 nos dois fatores pelos dois métodos). Por fim, recorreu-se ao número de fatores sugerido pelo método da raiz latente: apenas um.

Dessa forma, optou-se por agrupar todas as séries de radiação em um único grupo. Do ponto de vista teórico, esse resultado é aceitável, pois praticamente todos os estados das regiões SE/CO se encontram na zona tropical, ao sul da linha do Equador e à norte do trópico de Capricórnio (com exceção de parte do estado de São Paulo e o extremo sul de Mato Grosso do Sul). Por isso, o movimento aparente diário do sol e o tempo de insolação são semelhantes nesses estados.

5.3.4 Dados de Velocidade do Vento

Procedeu-se a análise de autovalores e variância explicada para os dados de velocidade do vento e os resultados são apresentados na Figura 5.11.

Neste caso, é difícil identificar a "quebra" do *scree plot*, já que os valores não chegam a ter comportamento paralelo ao eixo horizontal. Da análise dos autovalores, os seis primeiros autovalores são maiores do que uma unidade e o valor do sétimo é aproximadamente um. Por outro lado, a variância acumulada cresce lentamente. São necessários 15 autovalores para alcançar 70% de variância explicada e 18 para explicar 80%.

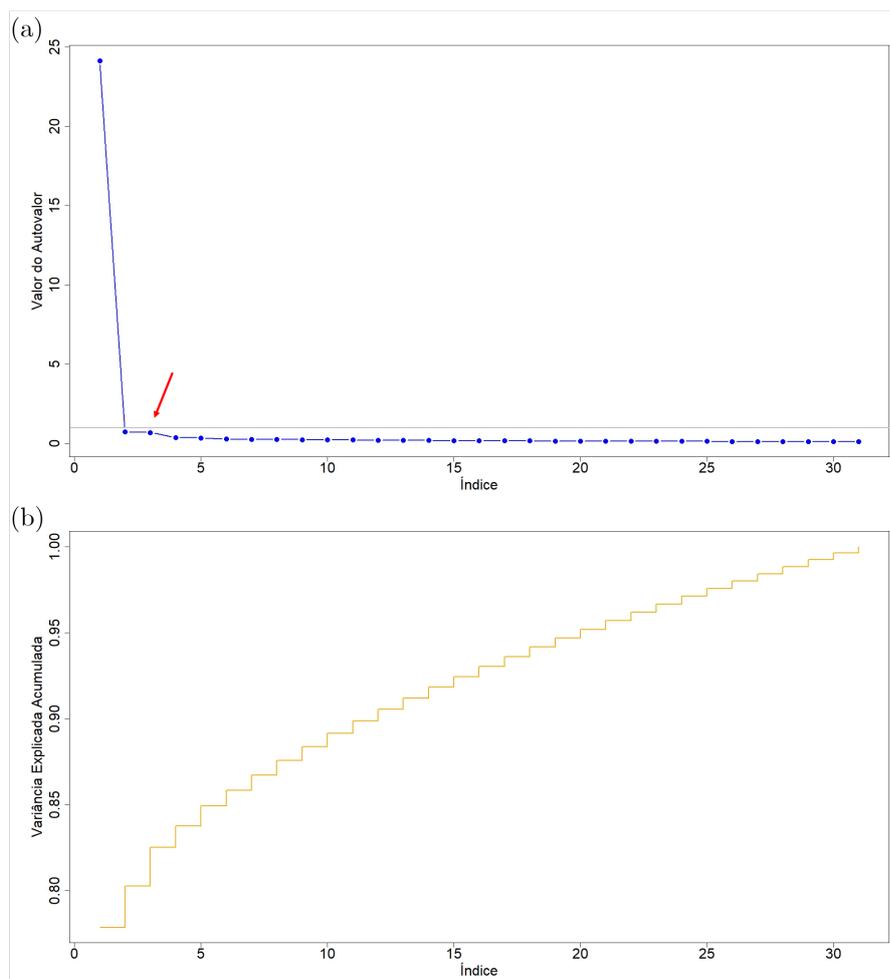


Figura 5.10: Resultado da análise de autovalores para os dados de radiação solar (cada índice representa um autovalor). (a) Autovalor por ordem de extração. (b) Variância acumulada ao longo dos autovalores.

Inicialmente, tentou-se utilizar 15 fatores, por conta da variância explicada. No entanto, a função do MMV retornou uma mensagem de erro afirmando que era "incapaz de otimizar a partir desse valor inicial". Buscou-se alterar alguns limites, como o valor mínimo de variância específica, porém sem sucesso. Pelo MCP, 9 das 25 séries ficaram isoladas. Buscou-se utilizar 13 fatores, mas novamente o MMV não pôde gerar resultados. Pelo MCP, sete estações ainda ficaram isoladas. Abaixo de 13 autovalores, a variância explicada não ultrapassaria 60%, o que é um valor bastante baixo para uma análise envolvendo ciências naturais. Por isso, optou-se por manter as séries de vento isoladas. Esse resultado está de acordo com as correlações lineares apresentadas na Figura 5.5 e indica que há forte variabilidade geográfica do regime eólico horário no Brasil.

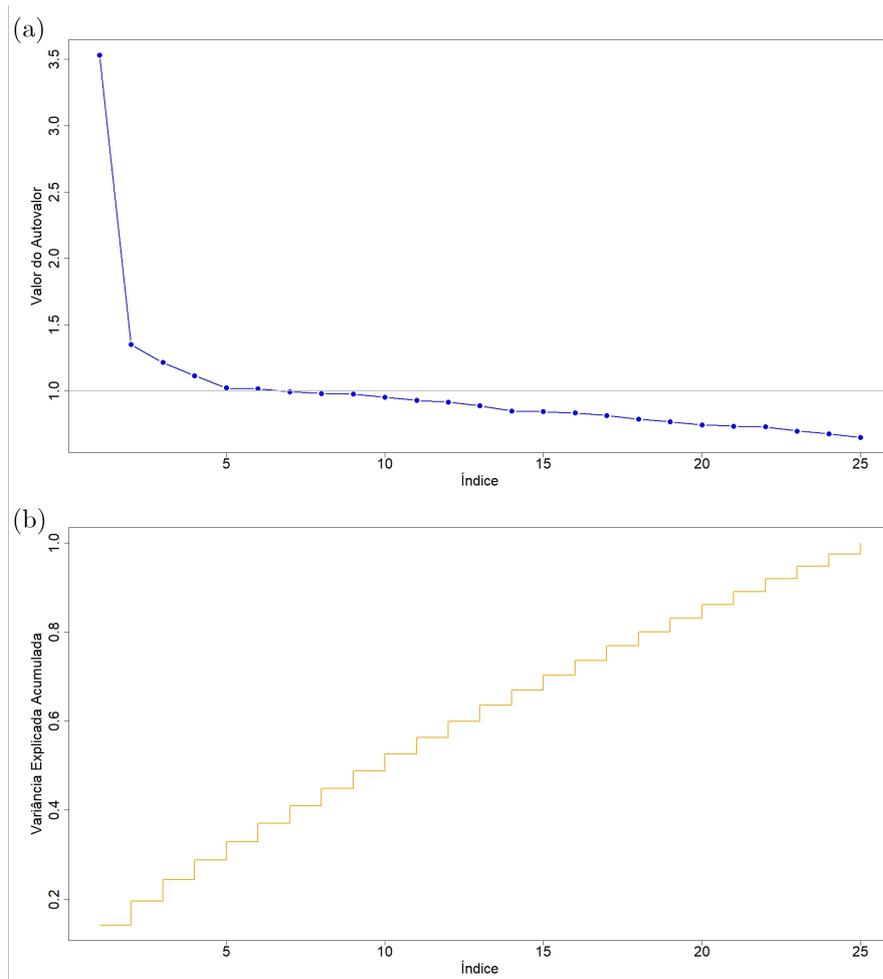


Figura 5.11: Resultado da análise de autovalores para os dados de velocidade do vento (cada índice representa um autovalor). (a) Autovalor por ordem de extração. (b) Variância acumulada ao longo dos autovalores.

5.4 Avaliação de Complementaridade

A avaliação de complementaridade se subdividiu na avaliação horária entre os grupo solar e as séries eólicas e na análise diária, considerando também a geração hidrelétrica. Primeiramente, será discutida a complementaridade horária.

5.4.1 Complementaridade Horária

Os coeficientes de correlação linear entre os dados horários de radiação solar e velocidade do vento ao cubo são apresentados na Figura 5.12.

Percebe-se que o grupo solar tem correlações positivas moderadas com algumas estações e correlações mais fracas com as demais. Esse comportamento sugere que há potencial

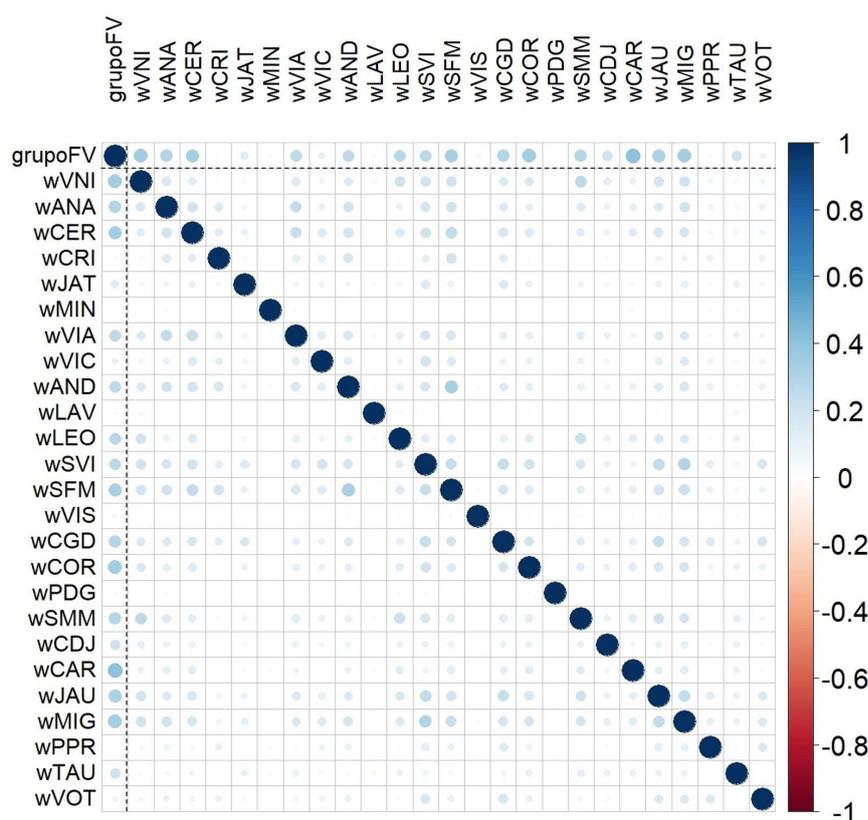


Figura 5.12: Correlação linear horária entre o grupo fotovoltaico e a geração eólica de cada estação.

Fonte: Autora

para que a geração fotovoltaica complemente a eólica, porém isso depende da localização geográfica dos parques eólicos (a complementaridade é favorecida quando a correlação é baixa).

A fim de compreender melhor a relação entre essas fontes, também foram calculadas as correlações para os períodos do dia e da noite. Considerando as séries originais (dados a cada três horas), agruparam-se os horários 21h, 0h, 3h e 6h no grupo “noite” e 9h, 12h, 15h e 18h no grupo “dia”.

Percebe-se que as correlações se tornam mais fortes durante o dia, tanto entre as estações eólicas como entre geração fotovoltaica e eólica. Ainda sim, em ambos períodos, a predominância é de correlações positivas fracas e de grupos não correlacionados.

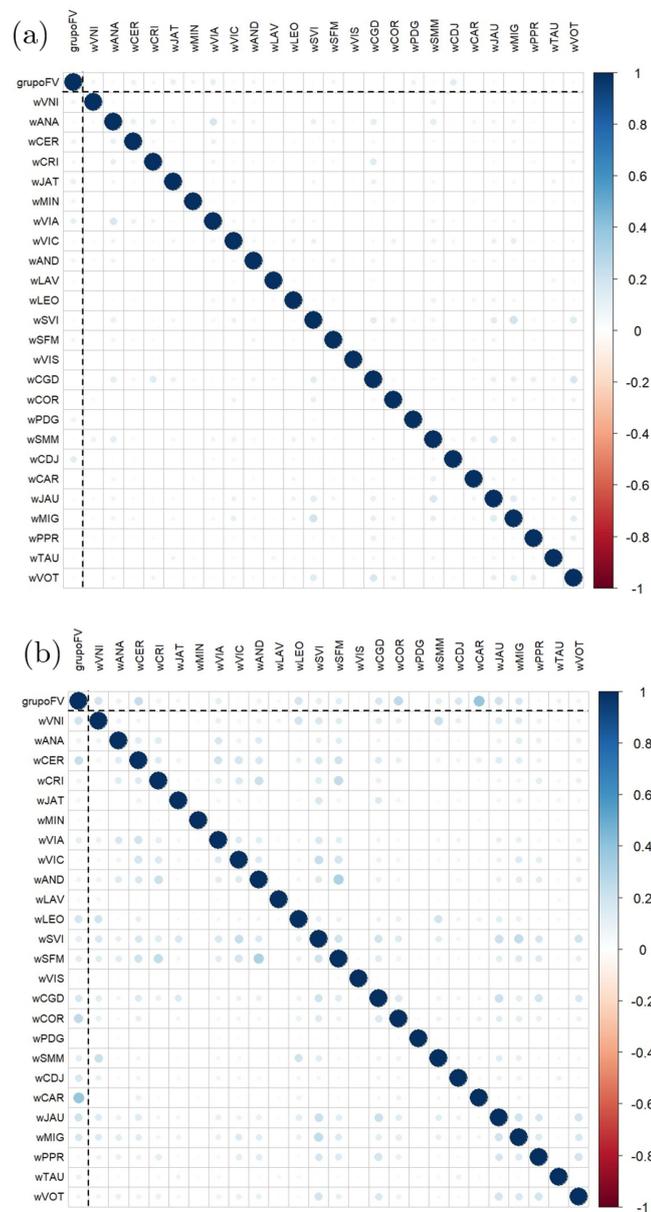


Figura 5.13: Correlação linear horária entre o grupo fotovoltaico e a geração eólica de cada estação nos períodos noturno (a) e diurno (b).

Fonte: Autora

Também avaliou-se o comportamento médio de cada série ao longo do dia e de acordo com a estação do ano, como apresentado na Figura 5.14. Os meses de outubro, novembro e dezembro representam a primavera; janeiro, fevereiro e março representam o verão; abril, maio e junho representam o outono; e julho, agosto e setembro representam o inverno.

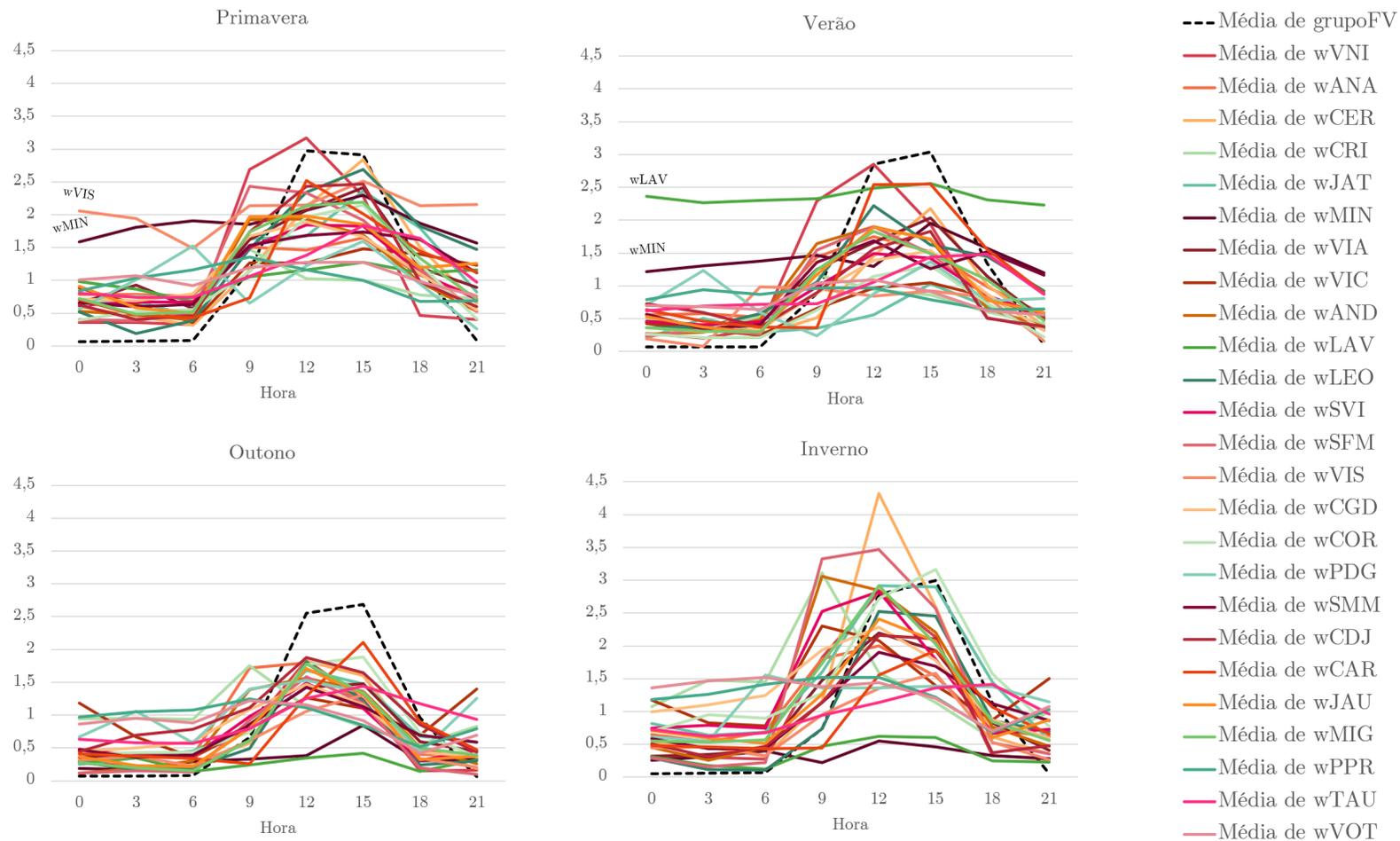


Figura 5.14: Comportamento diário médio das séries de acordo com a estação do ano. A linha preta tracejada representa o grupo fotovoltaico e as demais são as séries eólicas. Os gráficos estão na mesma escala (eixos idênticos).

Fonte: Autora

É notável que a sazonalidade diária da radiação varia pouco ao longo do ano em termos de magnitude, enquanto que as séries eólicas não têm um comportamento diário bem definido e a magnitude de potência eólica também varia ao longo do ano. Algumas estações eólicas têm comportamento semelhante ao da radiação, pois a potência gerada é maior durante o dia, mas diferenciando-se por valores não tão próximos de zero durante a noite e a madrugada. Durante o outono, a potência eólica apresenta seus menores valores, enquanto que no inverno, estação seguinte, surgem séries com os picos mais elevados de todo ano.

5.4.2 Complementaridade Diária

Os coeficientes de correlação linear entre os dados diários de potência fotovoltaica, hidrelétrica e eólica são apresentados na Figura 5.15.

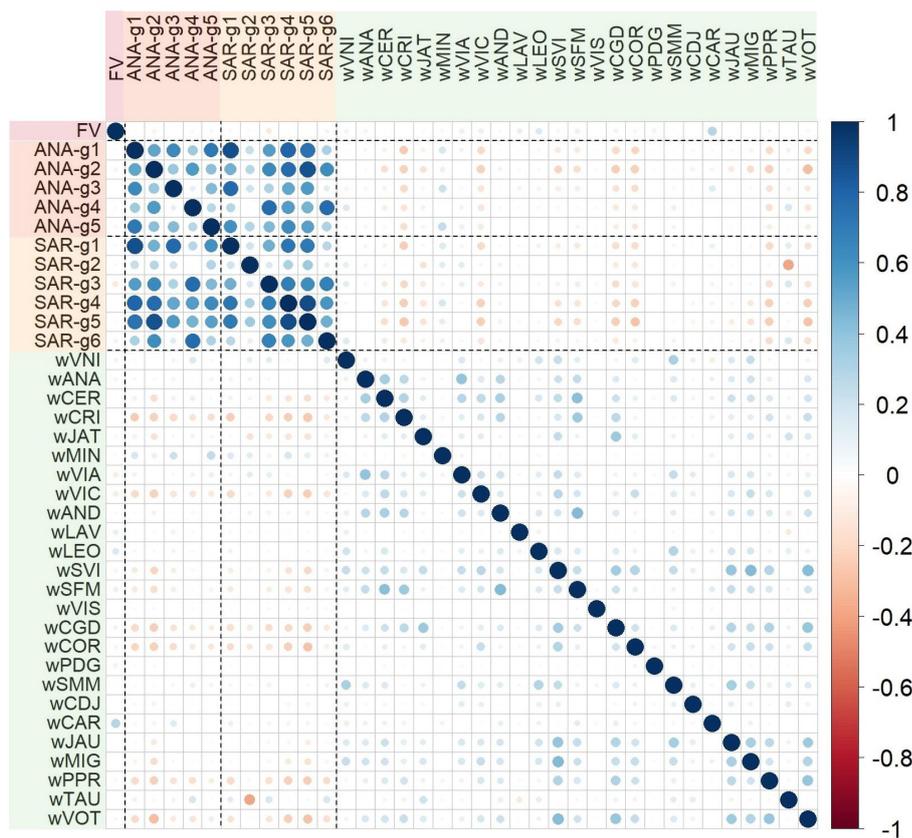


Figura 5.15: Correlação linear diária entre o grupo fotovoltaico, os grupos hidrelétricos e as gerações eólicas.

Fonte: Autora

Percebe-se que, em geral, as séries hidrelétricas têm correlações positivas moderadas

e fortes entre si, independentemente se são estações da ANA ou do ONS. Entre as séries eólicas, surgem algumas correlações mais fortes que no caso horário, mas mantém-se o equilíbrio entre séries correlacionadas e não correlacionadas. É notável que o grupo fotovoltaico tem correlações fracas (positivas e negativas) com todas as demais séries, com exceção de uma estação eólica. Deve-se ressaltar também que algumas estações eólicas e hidrelétricas têm correlações negativas entre si.

A análise de correlação entre os diferentes tipos de fonte, com predomínio de séries fracamente correlacionadas indica que elas têm comportamento distinto ao longo dos dias e que uma combinação ideal pode minimizar os efeitos negativos da intermitência individual.

Para uma avaliação mais completa, também calculou-se as correlações de acordo com as estações do ano, como apresentado na Figura 5.16.

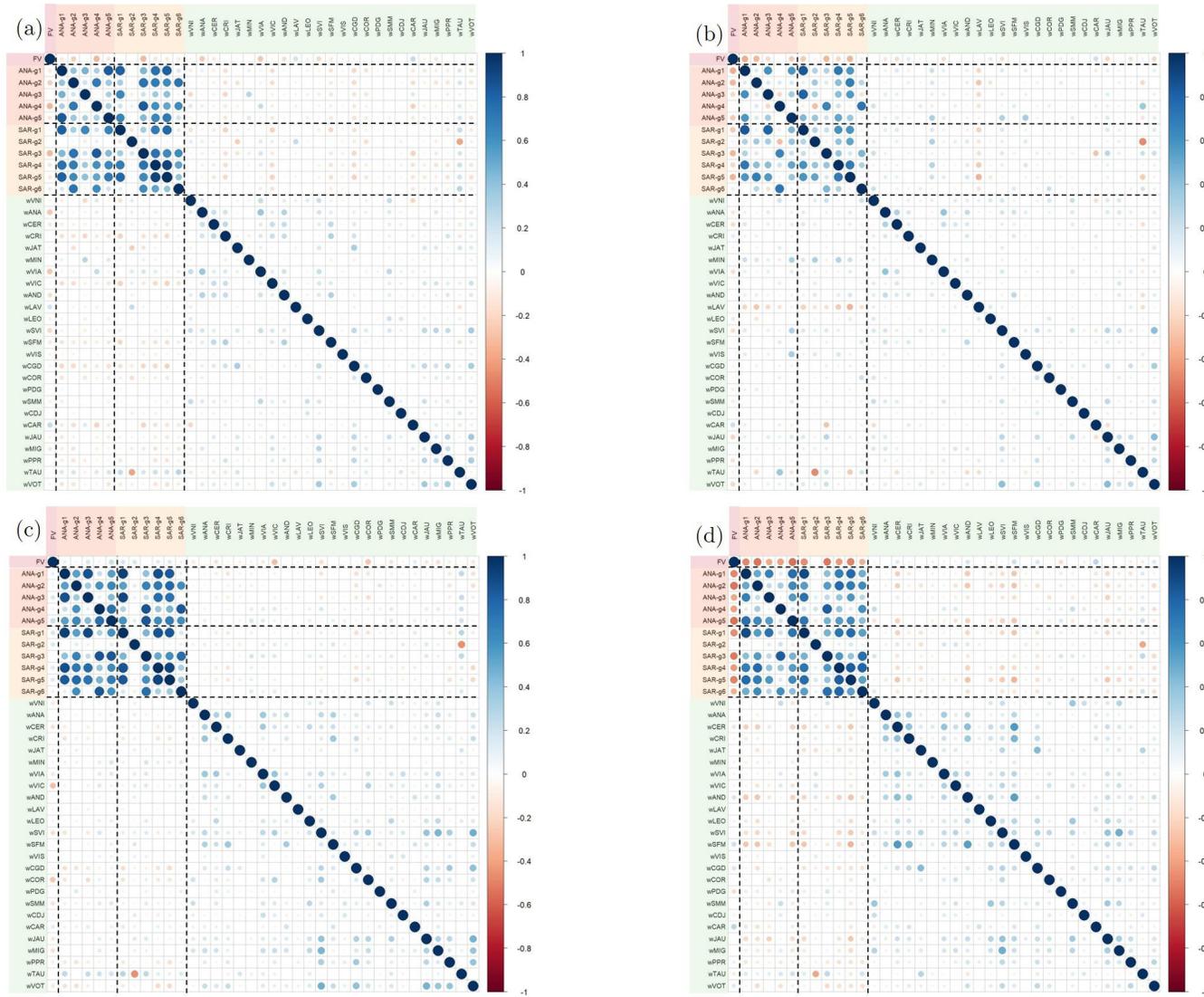


Figura 5.16: Correlações lineares diárias entre o grupo fotovoltaico, os grupos hidrelétricos e as séries eólicas, durante a primavera (a), o verão (b), o outono (c) e o inverno (d).

Fonte: Autora

A análise segmentada pelas estações do ano traz a informação de que, principalmente durante o verão e o inverno, diversas séries hidrelétricas têm correlações negativas moderadas com o grupo fotovoltaico. Percebe-se também que as correlações entre as séries hidrelétricas diminuí durante o verão.

O comportamento anual médio de cada série é comparado ao da demanda por eletricidade no subsistema SE/CO do SIN, como é apresentado na Figura 5.17. Ressalta-se que quatro séries eólicas (wMIN;wLAV;wVIS;wPDG) não são exibidas, pois possuem pontos extremos que dificultam a visualização, devido ao aumento na escala do gráfico para incluí-las.

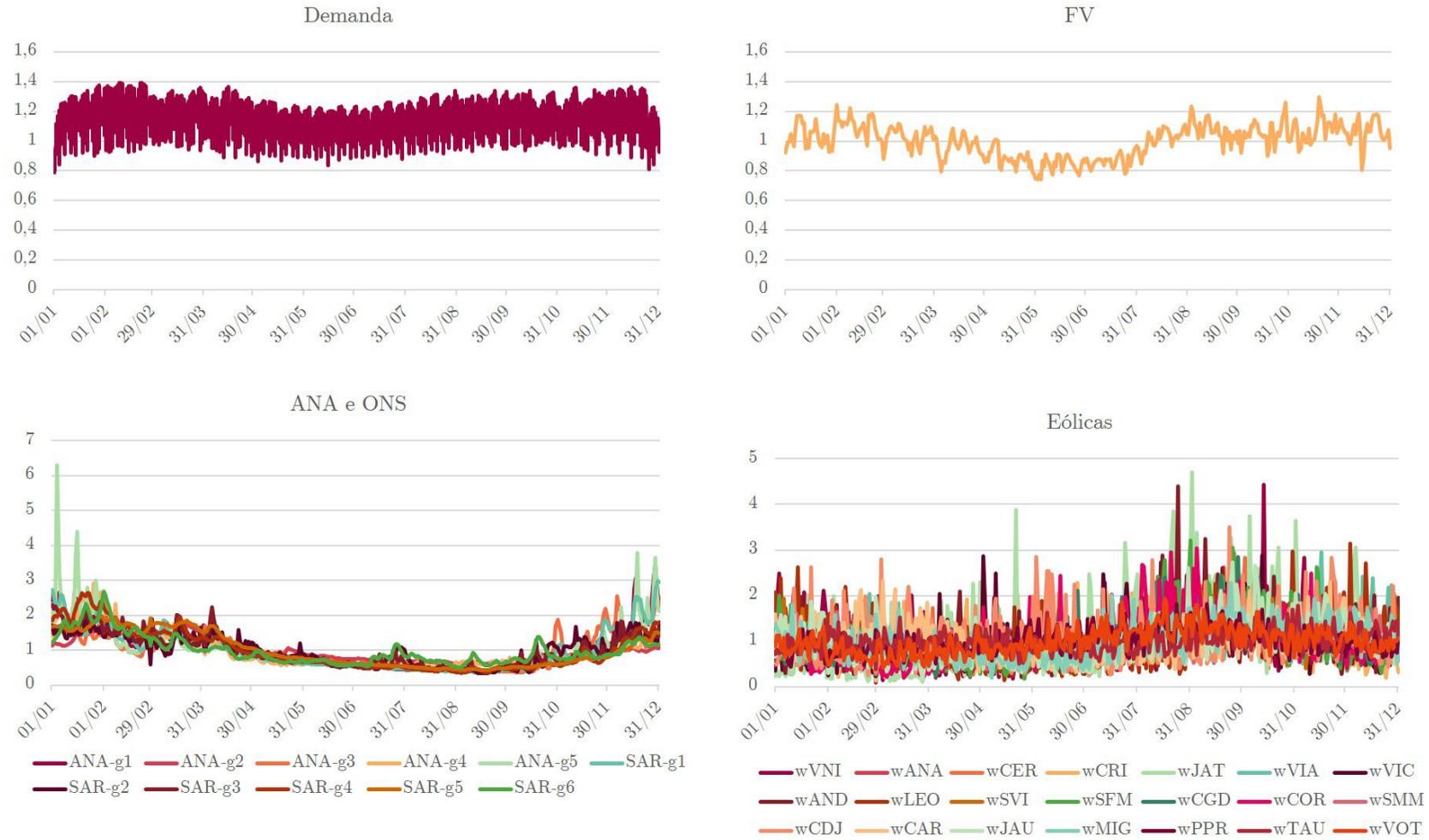


Figura 5.17: Comportamento anual médio das séries. Obs.: os gráficos têm escalas de eixo vertical distintas.

Fonte: Autora

A demanda por eletricidade diminui entre os meses de maio e agosto (outono/inverno), possivelmente porque diminui o consumo elétrico para refrigeração de ambientes. Essa variação é pequena comparada à sua sazonalidade semanal (o consumo elétrico é geralmente menor durante os finais de semana). Dentre as fontes de energia, a fotovoltaica apresentou comportamento mais estável ao longo do ano, apesar da geração menor entre abril e julho. A potência hidrelétrica é menor durante os meses centrais, o que é coerente com a menor precipitação nessa época do ano. Mais uma vez verifica-se a forte variabilidade da geração eólica, porém percebe-se um aumento na potência a partir de agosto aproximadamente.

5.5 Otimização

5.5.1 Formulação I

A resolução do modelo de otimização de acordo com a Formulação I (minimização da variância) resultou nas participações percentuais apresentadas na Figura 5.18.

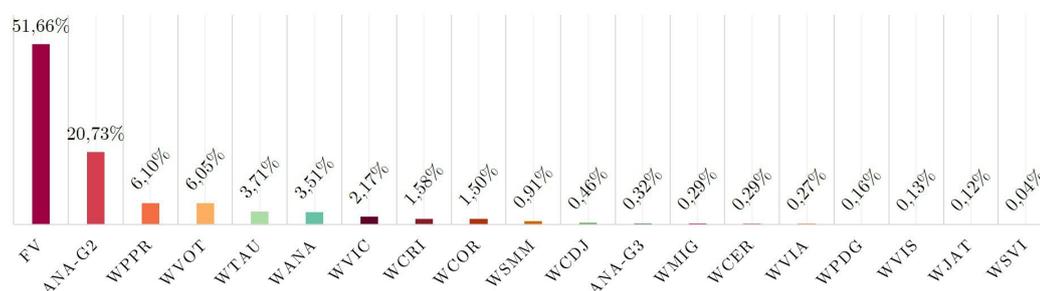


Figura 5.18: Participação dos grupos de fontes na potência total do *mix* renovável, de acordo com a resolução do modelo da Formulação I. As fontes não apresentadas tiveram participação nula.

Fonte: Autora

Deve-se destacar que, neste modelo de otimização, a geração fotovoltaica tem um papel importante, representando mais da metade da geração renovável intermitente. Depois dela, tem-se o grupo 2 de geração hidrelétrica da ANA. Esse grupo corresponde às estações da região central do Brasil. O perfil anual médio desse grupo é apresentado na Figura 5.19.

Nota-se que, quando comparado ao ano médio dos demais grupos hidrelétricos, como apresentado na Figura 5.17, o perfil de variação do grupo 2 da ANA tem amplitude menor do que a média, o que favorece este grupo durante a otimização.



Figura 5.19: Comportamento anual médio do grupo 2 da ANA.

Fonte: Autora

Além da otimização com todas as séries, repetiu-se a resolução do modelo (i) excluindo a série fotovoltaica e mantendo-se as demais e (ii) excluindo a série 2 da ANA e mantendo-se as demais, inclusive a fotovoltaica. Comparou-se, então, o resultado da função objetivo (variância da potência do *mix* intermitente) do caso original a esses dois cenários, como apresentado na Tabela 5.5.

| | Otimização 1 | Variação |
|------------|--------------|----------|
| FO geral | 0.0162 | - |
| FO sem FV | 0.0361 | 122.9% |
| FO sem ANA | 0.0189 | 16.6% |

Tabela 5.5: Valor da Função Objetivo da Formulação I para os cenários original, sem geração fotovoltaica e sem o grupo 2 da ANA

Fonte: Autora

No cenário sem o grupo 2 da ANA, a participação fotovoltaica aumentou para 58% e o restante ficou dividido entre os grupos 3 e 5 do ONS e 14 estações eólicas. A piora no valor da função objetivo é significativo, mas não tão grande quando comparado ao caso sem geração fotovoltaica, onde a variância mais do que dobrou. Nesse cenário, a participação do grupo 2 da ANA aumentou para 38% e o restante se dividiu entre algumas estações eólicas e dois outros grupos da ANA.

5.5.2 Formulação II

A resolução do modelo de otimização de acordo com a Formulação II (minimização da diferença entre geração e demanda) resultou nas participações percentuais apresentadas na Figura 5.20.

Nota-se que novamente a geração fotovoltaica representa mais da metade da potência

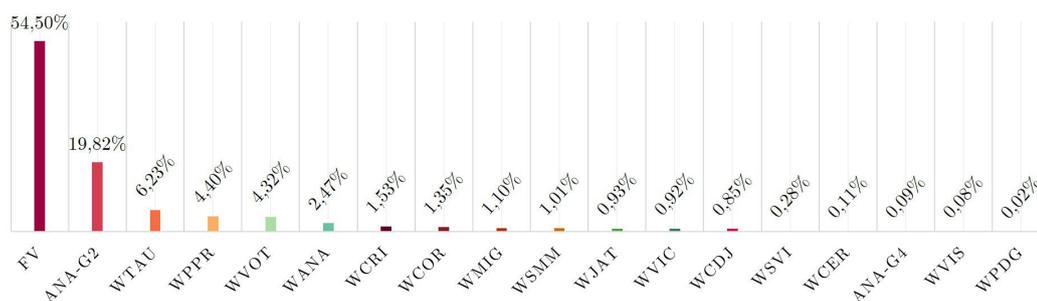


Figura 5.20: Participação dos grupos de fontes na potência total do *mix* renovável, de acordo com a resolução do modelo da Formulação II. As fontes não apresentadas tiveram participação nula.

Fonte: Autora

do *mix* intermitente, com um percentual de participação um pouco maior neste modelo. O segundo grupo da ANA também permanece como um componente importante do *mix*. As estações eólicas wTAU, wPPR, wVOT e wANA reaparecem nas posições seguintes.

Também nesta Formulação, repetiu-se a resolução do modelo (i) excluindo a série fotovoltaica e mantendo-se as demais e (ii) excluindo a série 2 da ANA e mantendo-se as demais, inclusive a fotovoltaica. Comparou-se, então, o resultado da função objetivo (soma dos desvios ao quadrado entre geração e demanda elétrica) do caso original a esses dois cenários, como apresentado na Tabela 5.6.

| | Otimização 2 | Variação |
|------------|--------------|----------|
| FO geral | 76.481 | - |
| FO sem FV | 117.29 | 53.4% |
| FO sem ANA | 81.675 | 6.8% |

Tabela 5.6: Valor da Função Objetivo da Formulação II para os cenários original, sem geração fotovoltaica e sem o grupo 2 da ANA

Fonte: Autora

No cenário sem o grupo 2 da ANA, a participação fotovoltaica aumentou para 62% e o restante ficou dividido entre os grupos 3 e 5 do ONS, o grupo 4 da ANA e 14 estações eólicas. Assim como no caso da Formulação I, a piora no valor da função objetivo é mais significativa no caso sem geração fotovoltaica, do que sem o grupo 2 da ANA. Sem a geração solar, a participação do grupo 2 da ANA aumentou para 40% e o restante se dividiu apenas entre estações eólicas, com destaque para wTAU, wVOT e wPPR, com mais de 10% de participação cada.

Comparou-se a série do *mix* intermitente otimizado conforme a Formulação II à série histórica de demanda. Também calculou-se a diferença percentual entre a geração de

energia e a demanda para cada dia j , ou seja, $\frac{\text{geração}_j - \text{demanda}_j}{\text{demanda}_j}$. Como resultado, obteve-se a Figura 5.21.



Figura 5.21: (a) Comparação entre a série histórica de demanda elétrica e a série do *mix* intermitente otimizada de acordo com a Formulação II. (b) Diferença percentual entre geração do *mix* otimizado e a demanda durante o período de estudo.

Fonte: Autora

Percebe-se que o *mix* otimizado tem mais dificuldade em suprir a demanda durante os meses mais frios, entre abril e agosto. De fato, tanto a geração fotovoltaica como a hidrelétrica apresentam menor potência nesses meses, como foi apresentado na Figura 5.17. Dessa forma, um *mix* fortemente dependente dessas formas de geração tende a ter produção menor nesse período. Esse comportamento pode ser compensado permitindo maior geração não aproveitada do *mix* (geração maior do que a demanda), ou utilizando-se outras fontes de energia não intermitentes, como as hidrelétricas com capacidade de regulação de reservatória e as usinas à biomassa.

5.5.3 Comparação dos Resultados

Apesar dos objetivos distintos, os dois modelos de otimização tiveram resultados semelhantes. Em ambos, a participação fotovoltaica representa mais de 50% da potência intermitente total, seguida pelo grupo 2 da ANA e pelas estações eólicas wTAU, wPPR,

wVOT e wANA.

O comportamento mais estável da geração fotovoltaica ao longo do ano, como visto na Seção 5.4, provavelmente foi o que mais contribuiu para essa forte participação do recurso solar nos *mixes* renováveis intermitentes ótimos. O grupo 2 da ANA foi favorecido por ter uma amplitude menor de variação ao longo do ano, como apresentado na Figura 5.19. Deve-se ressaltar que, como limitação deste estudo, considerou-se que as usinas absorveriam todos os valores medidos, o que não ocorre ao se definir a potência máxima de uma usina. Essa informação é relevante, pois a amplitude real dos demais grupos hidrelétricos pode ser definida de forma que o impacto do aumento de vazões nos meses de verão não chegue a valores tão elevados, comparados à média anual, como visto na Figura 5.17. Isso pode favorecer outras regiões hidrelétricas e não apenas o grupo ANA-g2, mas deve ser estudado em trabalhos futuros.

Revisando a Tabela 2.1 e a Figura 2.4, pode-se obter uma estimativa para o *mix* renovável intermitente instalado atualmente nas regiões, como apresentado na Tabela 5.7.

| Fonte | Potência Instalada (kW) | Porcentagem |
|----------------------|-------------------------|-------------|
| Setor Elétrico | | |
| PCH | 3.225.917,17 | 74,75% |
| UFV | 478.772,45 | 11,09% |
| CGH | 302.698,55 | 7,01% |
| EOL | 28.208,24 | 0,65% |
| Mini e Micro Geração | | |
| CGH | 45.590,58 | 1,06% |
| EOL | 35,4 | 0% |
| UFV | 234.201,24 | 5,43% |
| Total | | |
| Total | 4.315.423,63 | 100% |

Tabela 5.7: *Mix* renovável intermitente instalado nas regiões SE e CO
Fonte: Elaboração própria com dados de [14] e [15].

Atualmente a proporção é de 83% de hidrelétricas para 17% de fotovoltaicas. De acordo com os resultados deste estudo, percebe-se que pode haver crescimento da geração eólica, mas sobretudo, deve-se aumentar a participação da geração fotovoltaica, para que se aproxime da porcentagem otimizada (entre 50 e 55%), além de se expandir o *mix* como um todo.

Outros estudos também apontam para grande participação da geração fotovoltaica na geração de eletricidade. Por exemplo, ao otimizar o sistema elétrico brasileiro como um todo, os autores de [102] obtiveram um *mix* elétrico com as seguintes proporções:

37% de geração fotovoltaica, 9% de eólica e 50% de hidrelétrica e o restante balanceado por reservatórios de hidrelétricas e por termelétricas. Para o estado do Rio de Janeiro, o trabalho [33] apontou para um *mix* renovável com 62% de geração fotovoltaica.

Capítulo 6

Conclusões e Trabalhos Futuros

6.1 Conclusões

O estudo apresentou uma revisão da matriz elétrica das regiões Sudeste e Centro-Oeste e pôde-se constatar que a participação de fontes renováveis na geração de eletricidade baseia-se em grandes usinas hidrelétricas e na geração termelétrica a partir de biomassa, especialmente bagaço de cana-de-açúcar. A participação de fontes intermitentes renováveis ainda é pequena (menos do que 7% da capacidade instalada), mesmo que sejam bem-vindas do ponto de vista ambiental e tenham custos operacionais baixos.

A fim de direcionar a expansão da parcela intermitente da matriz elétrica, foram coletados dados de radiação solar global, velocidade do vento e vazões fluviais de estações espalhadas pelas regiões consideradas. Foram calculadas séries que descrevem o comportamento da potência elétrica que poderia ser extraída nesses locais, sem realizar suposições quanto às tecnologias a serem utilizadas.

Essas séries foram avaliadas e agrupadas através da Análise Fatorial. Pôde-se agrupar satisfatoriamente as séries de vazões. O recurso solar, por sua distribuição uniforme nas regiões, não respondeu bem a grupos com mais de dois fatores e as estações foram agrupadas em um grande grupo. Por outro lado, o recurso eólico possui variabilidade geográfica alta e não pôde ser agrupado, de modo que as estações permaneceram isoladas.

Após o agrupamento, avaliou-se a complementaridade entre os grupos através de correlações lineares e análise de geração média ao longo do tempo. A nível horário, constatou-se que a geração fotovoltaica pode complementar a eólica em algumas localidades, já que têm correlações próximas de zero. De igual modo, a geração eólica pode complementar a si mesma, quando bem distribuída geograficamente. Ressalta-se que tanto a geração

eólica como a solar têm produção menor durante a noite. Já a nível diário, os resultados apontam que pode haver complementaridade entre os grupos de fontes distintas, com destaque para correlações negativas entre os grupos hidrelétricos e eólicos.

Após avaliação de complementaridade, otimizou-se o *mix* renovável intermitente por dois critérios diferentes: minimizar a variância da potência do *mix* e minimizar a diferença entre a energia gerada e a demanda por eletricidade. Em ambos, obteve-se como resultado uma participação fotovoltaica maior do que 50%, seguida por cerca de 20% de um dos grupos hidrelétricos e por algumas estações eólicas. Esses resultados apontam para a necessidade de se expandir a participação fotovoltaica na matriz, seja através de usinas fotovoltaicas ou incentivos à utilização de painéis fotovoltaicos em residências e edifícios.

Devem ser ressaltados, como limitações do estudo, os seguintes fatores: a disponibilidade geográfica das estações, pois apesar de ter-se buscado a melhor cobertura geográfica que não comprometesse a abrangência temporal, alguns estados, sobretudo na região Centro-Oeste, tiveram poucas estações para representá-los; a existência de valores faltantes nas bases de dados utilizadas, que foram preenchidos; a resolução diária para etapa de otimização, ao invés de horária, devido às características das bases de dados de vazões; e o fato de que supôs-se que toda energia proveniente dos dados ambientais seria absorvida pelos sistemas de geração, o que não ocorre na realidade, pois os equipamentos dependem de valores mínimos, seja de radiação, velocidade do vento ou vazão, para funcionarem e o limite máximo de geração é restrito pela potência nominal instalada dos geradores correspondentes.

6.2 Trabalhos Futuros

Como sugestão a trabalhos futuros, sugere-se realizar a análise excluindo determinada porcentagem de valores muito baixos ou muito altos nas séries de potência elétrica a fim de que os dados considerados fiquem mais próximos dos que seriam de fato absorvidos pelas usinas.

Outra sugestão é otimizar com o objetivo de minimizar a ocorrência de geração menor que a demanda, e então, comparar com os resultados deste trabalho. Pode-se ainda subdividir as séries e calcular o *mix* ótimo para cada amostra, a fim de constatar a estabilidade da solução.

A eficiência de outros métodos de otimização, como redes neurais e programação genética, também pode ser verificada para a solução dos modelos propostos. E, com

respeito ao agrupamento das variáveis, técnicas de análise de *cluster* e redes de Kohonen podem ser consideradas.

ANEXO A

Na Figura A.1, apresenta-se o código da função *solve.QP()* do pacote **quadprog** do R, utilizada para a resolução dos problemas de programação quadrática deste trabalho.

```

### Copyright (C) 1995-2010 Berwin A. Turlach <Berwin.Turlach@gmail.com>
### This program is free software; you can redistribute it and/or modify it under the
terms of the GNU General Public License as published by the Free Software Foundation;
either version 2 of the License, or at your option) any later version.
### This program is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
GNU General Public License for more details.
### You should have received a copy of the GNU General Public License
along with this program; if not, write to the Free Software
Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111- 1307,USA.
###
solve.QP <- function(Dmat, dvec, Amat, bvec, meq=0, factorized=FALSE){

  n    <- nrow(Dmat)
  q    <- ncol(Amat)
  if( missing(bvec) )
    bvec <- rep(0,q)

  if( n != ncol(Dmat) )
    stop("Dmat is not symmetric!")
  if( n != length(dvec) )
    stop("Dmat and dvec are incompatible!")
  if( n != nrow(Amat) )
    stop("Amat and dvec are incompatible!")
  if( q != length(bvec) )
    stop("Amat and bvec are incompatible!")
  if( (meq > q) || (meq < 0) )
    stop("Value of meq is invalid!")

  iact <- rep(0,q)
  nact <- 0
  r    <- min(n,q)
  sol  <- rep(0,n)
  lagr <- rep(0,q)
  crval <- 0
  work <- rep(0,2*n+r*(r+5)/2+2*q+1)
  iter <- rep(0,2)

  res1 <- .Fortran(.QP_qpgen2,
                  as.double(Dmat), dvec=as.double(dvec),
                  as.integer(n), as.integer(n),
                  sol=as.double(sol), lagr=as.double(lagr),
                  crval=as.double(crval),
                  as.double(Amat), as.double(bvec), as.integer(n),
                  as.integer(q), as.integer(meq),
                  iact=as.integer(iact), nact=as.integer(nact),
                  iter=as.integer(iter), work=as.double(work),
                  ierr=as.integer(factorized))

  if( res1$ierr == 1)
    stop("constraints are inconsistent, no solution!")
  else if( res1$ierr == 2)
    stop("matrix D in quadratic function is not positive definite!")

  list(solution=res1$sol,
        value=res1$crval,
        unconstrained.solution=res1$dvec,
        iterations=res1$iter,
        Lagrangian = res1$lagr,
        iact=res1$iact[1:res1$nact])
}

```

Figura A.1: Código da função $solve.QP()$ para resolução de problemas de programação quadrática.

APÊNDICE A

Este Apêndice apresenta informações sobre as estações de coleta de dados usadas neste estudo. Primeiramente, são apresentadas no Quadro A.1 as estações de radiação solar do INPE.

| Código da PCD | Estação | Estado | Variável Medida | Abreviatura usada |
|---------------|----------------------|--------|---------------------------------|-------------------|
| 31959 | Venda Nova Imigrante | ES | RadSolAcum (MJ/m ²) | VNI |
| 32360 | Anápolis | GO | RadSolAcum (MJ/m ²) | ANA |
| 31873 | Anicuns | GO | RadSolAcum (MJ/m ²) | ANI |
| 32350 | Ceres | GO | RadSolAcum (MJ/m ²) | CER |
| 32481 | Cristalina | GO | RadSolAcum (MJ/m ²) | CRI |
| 32349 | Itumbiara | GO | RadSolAcum (MJ/m ²) | ITU |
| 32484 | Jataí | GO | RadSolAcum (MJ/m ²) | JAT |
| 32482 | Minaçu | GO | RadSolAcum (MJ/m ²) | MIN |
| 32345 | Mineiros | GO | RadSolAcum (MJ/m ²) | MNS |
| 32342 | Vianópolis | GO | RadSolAcum (MJ/m ²) | VIA |
| 32341 | Vicentinópolis | GO | RadSolAcum (MJ/m ²) | VIC |
| 32506 | Andrequicé | MG | RadSolAcum (MJ/m ²) | AND |
| 32508 | Caratinga | MG | RadSolAcum (MJ/m ²) | CRT |
| 32511 | Lavras | MG | RadSolAcum (MJ/m ²) | LAV |
| 31929 | Santa Vitória | MG | RadSolAcum (MJ/m ²) | SVI |
| 31930 | Santa Fé de Minas | MG | RadSolAcum (MJ/m ²) | SFM |
| 32509 | Viçosa | MG | RadSolAcum (MJ/m ²) | VIS |
| 31950 | Campo Grande | MS | RadSolAcum (MJ/m ²) | CGD |
| 31949 | Corumbá | MS | RadSolAcum (MJ/m ²) | COR |
| 32595 | Porto dos Gaúchos | MT | RadSolAcum (MJ/m ²) | PDG |
| 31956 | Santa Maria Madalena | RJ | RadSolAcum (MJ/m ²) | SMM |
| 31954 | Teresópolis | RJ | RadSolAcum (MJ/m ²) | TER |
| 32527 | Campos do Jordão | SP | RadSolAcum (MJ/m ²) | CDJ |
| 32521 | Caraguatatuba | SP | RadSolAcum (MJ/m ²) | CAR |
| 31978 | Jaú | SP | RadSolAcum (MJ/m ²) | JAU |
| 31975 | Miguelópolis | SP | RadSolAcum (MJ/m ²) | MIG |
| 31980 | Presidente Prudente | SP | RadSolAcum (MJ/m ²) | PPR |
| 31973 | São Carlos | SP | RadSolAcum (MJ/m ²) | SCL |
| 32533 | Silveiras | SP | RadSolAcum (MJ/m ²) | SIL |
| 31981 | Taquarivaí | SP | RadSolAcum (MJ/m ²) | TAQ |
| 32531 | Taubaté | SP | RadSolAcum (MJ/m ²) | TAU |

Quadro A.1: Estações de Radiação Solar Seleccionadas

Já as estações que medem velocidade do vento são apresentadas no Quadro A.2.

| Código da PCD | Estação | Estado | Variável Medida | Abreviatura usada |
|---------------|------------------------|--------|--------------------|-------------------|
| 31959 | Venda Nova Imigrante | ES | VelVento10m (m/s) | VNI |
| 32360 | Anápolis | GO | VelVento5m (m/s) | ANA |
| 32350 | Ceres | GO | VelVento5m (m/s) | CER |
| 32481 | Cristalina | GO | VelVento10m (m/s) | CRI |
| 32484 | Jataí | GO | VelVento10m (m/s) | JAT |
| 32482 | Minacu | GO | VelVento10m (m/s) | MIN |
| 32342 | Vianópolis | GO | VelVento5m (m/s) | VIA |
| 32341 | Vicentinópolis | GO | VelVento5m (m/s) | VIC |
| 32506 | Andrequice | MG | VelVento10m (m/s) | AND |
| 32511 | Lavras | MG | VelVento10m (m/s) | LAV |
| 32510 | Leopoldina | MG | VelVento10m (m/s) | LEO |
| 32503 | Montes Claros de Goiás | MG | VelVento10m (m/s) | MCG |
| 31929 | Santa Vitória | MG | VelVento10m (m/s) | SVI |
| 31930 | Santa Fé de Minas | MG | VelVento10m (m/s) | SFM |
| 32509 | Viçosa | MG | VelVento10m (m/s) | VIS |
| 31950 | Campo Grande | MS | VelVento10m (m/s) | CGD |
| 31949 | Corumbá | MS | VelVento10m (m/s) | COR |
| 32595 | Porto dos Gaúchos | MT | VelVento3m (m/s) | PDG |
| 31956 | Santa Maria Madalena | RJ | VelVento10m (m/s) | SMM |
| 32527 | Campos do Jordão | SP | VelVento10m (m/s) | CDJ |
| 32521 | Caraguatatuba | SP | VelVento10m (m/s) | CAR |
| 31978 | Jaú | SP | VelVento10m (m/s) | JAU |
| 31975 | Miguelópolis | SP | VelVento10m (m/s) | MIG |
| 31980 | Presidente Prudente | SP | VelVento10m (m/s) | PPR |
| 32531 | Taubaté | SP | VelVento10m (m/s) | TAU |
| 31976 | Votuporanga | SP | VelVento10m (m/s) | VOT |

Quadro A.2: Estações de Velocidade do Vento Seleccionadas

As estações do ONS são mostradas no Quadro A.3.

| Estação | UF | Abreviação |
|----------------------------|-----------|-------------------|
| A. VERMELHA | MG / SP | AVER |
| AIMORES | MG | AIMO |
| B. BONITA | SP | BBON |
| B.COQUEIROS | GO | BCOQ |
| BARIRI | SP | BARI |
| BILLINGS | SP | BILL |
| C.BRANCO-1 | MG | CBR1 |
| C.BRANCO-2 | MG | CBR2 |
| C. DOURADA | MG / GO | CDOU |
| CACONDE | SP | CACO |
| CACU | GO | CACU |
| CAMARGOS ¹ | MG | CAMA |
| CANA BRAVA | GO | CANA |
| CANDONGA | MG | CAND |
| CANOAS I | SP | CAN1 |
| CANOAS II | SP | CAN2 |
| CAPIVARA | SP | CAPI |
| CHAVANTES | SP | CHAV |
| CORUMBA | GO | COR1 |
| CORUMBA-4 | GO | COR4 |
| E. DA CUNHA | SP | EDAC |
| EDGARD SOUZA | SP | EDGA |
| EMBORCAÇÃO | MG / GO | EMBO |
| FONTES | RJ | FONT |
| FOZ R.CLARO | GO | FOZR |
| FUNIL | RJ | FUNI |
| FUNIL-MG | MG | FUMG |
| FURNAS | MG | FURN |
| GUAPORE | MT | GUAP |
| GUARAPIRANGA | SP | GUAR |
| GUILM. AMORIM | MG | GUIL |
| I. SOLTEIRA | SP / MS | ISOL |
| IBITINGA | SP | IBIT |
| IGARAPAVA | MG / SP | IGAR |
| ILHA POMBOS | MG / RJ | ILHA |
| IRAPE | MG | IRAP |
| ITIQUEIRA I | MT | ITI1 |
| ITUMBIARA | MG / GO | ITUM |
| ITUTINGA | MG | ITUT |
| JAGUARA | MG | JAGU |
| JAGUARI | SP | JAGR |
| JAURU | MT | JAUR |
| JUPIA | SP / MS | JUPI |
| JURUMIRIM | SP | JURU |
| L. C. BARRETO | MG / SP | LCBA |
| LAJES | RJ | LAJE |
| LIMOEIRO | SP | LIMO |
| M. MORAES | MG | MMOR |
| MANSO | MT | MANS |
| Continua na próxima página | | |

¹Excluída posteriormente por ser idêntica à série ITUT

Quadro A.3 – continuação

| Estação | UF | Abreviação |
|-----------------|-----------|-------------------|
| MARIMBONDO | MG / SP | MARI |
| Mascarenhas | ES | MASC |
| MIRANDA | MG | MIRA |
| N. AVANHANDAVA | SP | NAVA |
| NILO PEÇANHA | RJ | NILO |
| NOVA PONTE | MG | NOVP |
| OURINHOS | SP | OURI |
| P. COLOMBIA | MG / SP | PCOL |
| PARAIBUNA | SP | PARA |
| PEREIRA PASSOS | RJ | PERE |
| PICADA | MG | PICA |
| PIRAJU | SP | PIRA |
| PONTE DE PEDRA | MT / MS | PTEP |
| PONTE NOVA | SP | PTEN |
| PORTO ESTRELA | MG | PORT |
| PORTO PRIMAVERA | SP / MS | PPRI |
| PROMISSÃO | SP | PROM |
| QUEIMADO | MG / GO | QUEI |
| ROSAL | RJ / ES | ROSL |
| ROSANA | SP | ROSN |
| S.DO FACÃO | GO | SDOF |
| S.R.VERDINHO | GO | SRVE |
| SA CARVALHO | MG | SACA |
| SALTO | GO | SALT |
| SALTO GRANDE CM | MG | SALG |
| SALTO GRANDE CS | SP | SLTG |
| SANTA BRANCA | SP | STAB |
| SANTA CECILIA | RJ | STCE |
| SANTANA | RJ | SANT |
| SÃO SIMÃO | MG / GO | SAOS |
| Serra da Mesa | GO | SERR |
| SOBRAGI | MG | SOBR |
| STA.CLARA-MG | MG | STAC |
| TAQUARUÇU | SP | TAQU |
| TOCOS | RJ | TOCO |
| TRÊS IRMÃOS | SP | TRIR |
| TRÊS MARIAS | MG | TRES |
| VOLTA GRANDE | MG / SP | VOLT |

Quadro A.3: Estações de Vazões Naturais Diárias do ONS

Por fim, as estações da ANA são apresentadas no Quadro A.4.

| Código | Estação | UF | % Dados reais | % Dados estimados | Abreviação |
|----------|----------------------------------|----|---------------|-------------------|------------|
| 56998400 | BARRA DE SÃO GABRIEL | ES | 92.55% | 7.45% | BSG |
| 57130000 | SANTA LEOPOLDINA | ES | 99.73% | 0.27% | SLE |
| 57170000 | CÓRREGO DO GALO | ES | 99.73% | 0.27% | CGL |
| 57250000 | MATILDE | ES | 100% | 0% | MAT |
| 57880000 | MIMOSO DO SUL | ES | 96.77% | 3.23% | MDS |
| 60680000 | PONTE MEIA PONTE | GO | 93.15% | 6.85% | PMP |
| 60785005 | FAZENDA PARAÍSO | GO | 97.04% | 2.96% | FZP1 |
| 60970000 | ITAJÁ | GO | 98.69% | 1.31% | ITA |
| 56075000 | PORTO FIRME | MG | 99.51% | 0.49% | PTF |
| 56110005 | PONTE NOVA JUSANTE | MG | 99.67% | 0.33% | PNJ |
| 56240000 | FAZENDA PARAÍSO | MG | 100% | 0% | FZP2 |
| 56385000 | SÃO MIGUEL DO ANTA | MG | 98.25% | 1.75% | SMA |
| 56460000 | MATIPÓ | MG | 100% | 0% | MTP |
| 56983000 | DORES DE MANHUMIRIM | MG | 94.96% | 5.04% | DDM |
| 58500000 | USINA BRUMADO | MG | 100% | 0% | UBR |
| 58516500 | FAZENDA SANTO ANTÔNIO | MG | 99.84% | 0.16% | FSA |
| 58658000 | VOLTA GRANDE | MG | 100% | 0% | VGD |
| 58710000 | USINA ITUERÉ | MG | 99.95% | 0.05% | UIT |
| 60220000 | DESEMBOQUE | MG | 90.09% | 9.91% | DES |
| 61115000 | USINA SÃO JOÃO DEL REI | MG | 91.18% | 8.82% | UDR |
| 60968000 | CASSILÂNDIA | MS | 93.37% | 6.63% | CAS |
| 63001600 | INDÁIA GRANDE | MS | 99.89% | 0.11% | IGD |
| 63001750 | MORANGAS | MS | 99.51% | 0.49% | MOR |
| 64608000 | USINA SÃO JOÃO II | MS | 98.58% | 1.42% | USJ |
| 58235100 | QUELUZ | SP | 100% | 0% | QUE |
| 61818080 | UHE EUCLIDES DA CUNHA BARRAMENTO | SP | 100% | 0% | EUC |
| 62335000 | PCH PIRAPORA MONTANTE | SP | 100% | 0% | PIR |
| 62600000 | RIO ABAIXO (FAZ. CACHOEIRA) | SP | 93.26% | 6.74% | ABX |
| 58420000 | FAZENDA SOBRADINHO | RJ | 99.95% | 0.05% | FSB |
| 58434000 | FAGUNDES | RJ | 99.95% | 0.05% | FAG |
| 58790002 | SANTO ANTÔNIO DE PÁDUA II | RJ | 96.82% | 3.18% | SAP |
| 58850000 | PIMENTEL | RJ | 98.19% | 1.81% | PIM |
| 58857000 | ALDEIA | RJ | 96.33% | 3.67% | ALD |

Quadro A.4: Estações de Vazões Médias Diárias da ANA

APÊNDICE B

Algoritmo 3 AF-partel(cov)

```

1: nomes<-colnames(cov)
2: p<-ncol(cov)
3: indice=1:p
4: r<-eigen(cov)
5: vet<-r$vectors
6: lamb<-r$values
7: Autov <- paste( c("lamb"), 1:p, sep=)
8: traco<-sum(diag(cov))
9: varexpl<-lamb/traco
10: varexacum<-acumular(varexpl)
11: plot (indice,varexpl,type="b",col="green",xlab="Índice",ylab="Variância Explicada",
    pch=16,cex=0.2,lwd=2)
12: plot(indice,lamb,type="b",col="blue",xlab="Índice",ylab="Valor do Autovalor",
    pch=16,cex=0.2,lwd=2)
13: plot(indice,varexacum,type="s",col="orange",xlab="Índice",ylab="Variância Ex-
    plicada Acumulada",lwd=2)
14: MatVarExplicada<-data.frame(Autov,varexpl,varexacum)
    colnames(MatVarExplicada)<-c("Autovalor","Var. explicada","Var. expl. acum.")
15: resultados<-list(número_de_variáveis=p,autovalores=lamb,autovetores=vet,
    var_explicada=MatVarExplicada)
16: return(resultados)

```

APÊNDICE C

Algoritmo 4 AF-componente-principal(cov,m)

```

1: #Definição dos nomes dos fatores
   FatLabs <- paste( c("F"), 1:m, sep=) #Cria uma lista de F1, F2,..., Fm
2: #Nomes das variáveis
   nomes<-colnames(cov)
3: #Cálculo dos autovalores e autovetores
   r<-eigen(cov)
   vet<-r$vector
   lamb<-r$values
4: #Variância explicada por cada autovalor
   traco<-sum(diag(cov))
   varexpl<-lamb/traco
   varexplsel<-varexpl[c(1:m)]
   varexplacum<-acumular(varexplsel)
5: #Cálculo das cargas
   lamraiz<-sqrt(lamb)
   cargas<-sweep(vet,2,lamraiz, "*")
6: #Cálculo da Comunalidade
   cargas2<-cargas2
   cargas2sel<-cargas2,[c(1:m)]
   comunalidade<-rowSums(cargas2sel)
7: #Variâncias Específicas
   varesp<-diag(cov)-comunalidade
8: #Matriz residual
   cargassel<-cargas[,c(1:m)]
   ll<-cargassel%*%t(cargassel)
   residual<-cov-(ll+diag(varesp))
   residuomaximo<-max(residual)
9: #Resultados
   #1) variâncias específicas MatVarEsp<-data.frame(nomes,varesp)
   #2) cargas MatCargas<-data.frame(nomes,cargassel); colnames(MatCargas)<-
   c("var",FatLabs);
   #3) comunalidades MatComunalidades<-data.frame(nomes,comunalidade)
   #4) variância explicada MatVarExplicada<-data.frame(FatLabs,varexplsel,varexplacum);
   colnames(MatVarExplicada)<-c("Fatores","Var. explicada","Var. expl. acum.")
10: #Lista com resultados
   resul<-list(fatores=FatLabs,variáveis=nomes,autovetores=vet,autovalores=lamb,
   traço=traco, var_expl=varexpl,var_expl_m=varexplsel,var_expl_acum=varexplacum,
   raiz_autovalores=lamraiz, cargas=cargas, cargas_quad=cargas2,cargas_sel=cargassel,
   cargas_quad_sel=cargas2sel,comunalidades=comunalidade,var_específica=varesp,
   LL=ll,matriz_residual=residual,matvex=MatVarExplicada,matco=MatComunalidades,
   matves=MatVarEsp, matca=MatCargas,residuo_max=residuomaximo)
   return(resul); invisible(resul)
11: #Exibir
   print("residuo máximo: ");residuomaximo
   print("Variâncias específicas"); MatVarEsp
   print("Cargas"); MatCargas
   print("Comunalidades"); MatComunalidades
   print("Variância Explicada"); MatVarExplicada

```

APÊNDICE D

Pode-se verificar que as formulações da função objetivo apresentadas nas Seções 4.6.1.1 e 4.6.1.2 são equivalentes, como apresentado a seguir.

$$w^\top \Sigma w = \begin{bmatrix} w_1 & w_2 & \dots & w_n \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

$$= \begin{bmatrix} (w_1\sigma_{11} + w_2\sigma_{21} + \dots + w_n\sigma_{n1}) \\ (w_1\sigma_{12} + w_2\sigma_{22} + \dots + w_n\sigma_{n2}) \\ \dots \\ (w_1\sigma_{1n} + w_2\sigma_{2n} + \dots + w_n\sigma_{nn}) \end{bmatrix}^\top \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix}$$

$$= w_1^2\sigma_{11} + w_1w_2\sigma_{21} + \dots + w_1w_n\sigma_{n1} + w_1w_2\sigma_{12} + w_2^2\sigma_{22} + \dots + w_2w_n\sigma_{n2} + \dots + w_1w_n\sigma_{1n} + w_2w_n\sigma_{2n} + \dots + w_n^2\sigma_{nn}$$

$$= \sum_{k=1}^n w_1w_k\sigma_{k1} + \sum_{k=1}^n w_2w_k\sigma_{k2} + \dots + \sum_{k=1}^n w_nw_k\sigma_{kn}$$

$$= \sum_{i=1}^n \sum_{k=1}^n w_iw_k\sigma_{ki}$$

APÊNDICE E

Pode-se verificar que as formulações da função objetivo apresentadas nas Seções 4.6.3 e 4.6.3.1 são equivalentes, como apresentado a seguir.

a) Termo constante

$$C = \sum_{j=1}^m c_j^2$$

b) Termo que depende linearmente dos w_i

$$z = \mathbf{c}^\top \mathbf{A} \mathbf{w}$$

$$= \begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix} \begin{bmatrix} p_{11} & p_{21} & \dots & p_{n1} \\ p_{12} & p_{22} & \dots & p_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m} & p_{2m} & \dots & p_{nm} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

$$= \begin{bmatrix} (c_1 p_{11} + c_2 p_{12} + \dots + c_m p_{1m}) \\ (c_1 p_{21} + c_2 p_{22} + \dots + c_m p_{2m}) \\ \dots \\ (c_1 p_{n1} + c_2 p_{n2} + \dots + c_m p_{nm}) \end{bmatrix}^\top \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

$$= c_1 w_1 p_{11} + c_2 w_1 p_{12} + \dots + c_m w_1 p_{1m} + c_1 w_2 p_{21} + c_2 w_2 p_{22} + \dots + c_m w_2 p_{2m} + (\dots) + c_1 w_n p_{n1} + c_2 w_n p_{n2} + \dots + c_m w_n p_{nm}$$

$$= c_1 (w_1 p_{11} + w_2 p_{21} + \dots + w_n p_{n1}) + c_2 (w_1 p_{12} + w_2 p_{22} + \dots + w_n p_{n2}) + \dots +$$

$$\begin{aligned}
& + c_m(w_1 p_{1m} + w_2 p_{2m} + \dots + w_n p_{nm}) \\
& = c_1 \left(\sum_{i=1}^n w_i p_{i1} \right) + c_2 \left(\sum_{i=1}^n w_i p_{i2} \right) + \dots + c_m \left(\sum_{i=1}^n w_i p_{im} \right) \\
& = \sum_{j=1}^m c_j \left(\sum_{i=1}^n w_i p_{ij} \right)
\end{aligned}$$

c) Termo quadrático

$$z = \mathbf{w}^\top \mathbf{B} \mathbf{w}$$

$$= \begin{bmatrix} w_1 & w_2 & \dots & w_n \end{bmatrix} \begin{bmatrix} \sum_{j=1}^m p_{1j}^2 & \sum_{j=1}^m p_{1j} p_{2j} & \dots & \sum_{j=1}^m p_{1j} p_{nj} \\ \sum_{j=1}^m p_{1j} p_{2j} & \sum_{j=1}^m p_{2j}^2 & \dots & \sum_{j=1}^m p_{2j} p_{nj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^m p_{1j} p_{nj} & \sum_{j=1}^m p_{nj} p_{2j} & \dots & \sum_{j=1}^m p_{nj}^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

$$= \begin{bmatrix} (w_1 \sum_{j=1}^m p_{1j}^2 + w_2 \sum_{j=1}^m p_{1j} p_{2j} + \dots + w_n \sum_{j=1}^m p_{1j} p_{nj}) \\ (w_1 \sum_{j=1}^m p_{1j} p_{2j} + w_2 \sum_{j=1}^m p_{2j}^2 + \dots + w_n \sum_{j=1}^m p_{nj} p_{2j}) \\ \dots \\ (w_1 \sum_{j=1}^m p_{1j} p_{nj} + w_2 \sum_{j=1}^m p_{nj} p_{2j} + \dots + w_n \sum_{j=1}^m p_{nj}^2) \end{bmatrix}^\top \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix}$$

$$\begin{aligned}
& = w_1^2 \sum_{j=1}^m p_{1j}^2 + w_1 w_2 \sum_{j=1}^m p_{1j} p_{2j} + \dots + w_1 w_n \sum_{j=1}^m p_{1j} p_{nj} + w_1 w_2 \sum_{j=1}^m p_{1j} p_{2j} + w_2^2 \sum_{j=1}^m p_{2j}^2 + \dots + \\
& + w_2 w_n \sum_{j=1}^m p_{nj} p_{2j} + \dots + w_1 w_n \sum_{j=1}^m p_{1j} p_{nj} + w_2 w_n \sum_{j=1}^m p_{2j} p_{nj} + \dots + w_n^2 \sum_{j=1}^m p_{nj}^2 \\
& = \sum_{j=1}^m [w_1^2 p_{1j}^2 + w_1 w_2 p_{1j} p_{2j} + \dots + w_1 w_n p_{1j} p_{nj} + w_1 w_2 p_{1j} p_{2j} + w_2^2 p_{2j}^2 + \dots + w_2 w_n p_{nj} p_{2j} + \dots + \\
& + w_1 w_n p_{1j} p_{nj} + w_2 w_n p_{2j} p_{nj} + \dots + w_n^2 p_{nj}^2] \\
& = \sum_{j=1}^m [(w_1 p_{1j} + w_2 p_{2j} + \dots + w_n p_{nj})(w_1 p_{1j} + w_2 p_{2j} + \dots + w_n p_{nj})]
\end{aligned}$$

$$= \sum_{j=1}^m \left[(w_i p_{ij})^2 \right]$$

Referências

- [1] ADAMS, T., BEZNER, J., STEINHARDT, M. The conceptualization and measurement of perceived wellness: Integrating balance across and within dimensions. *American Journal of health promotion* 11, 3 (1997), 208–218.
- [2] AMARAL, B. M. Modelos varx para geração de cenários de vento e vazão aplicados à comercialização de energia. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2011.
- [3] ANA. HIDROWEB - Sistema de Informações Hidrológicas. Disponível em <http://www.snirh.gov.br/hidroweb/publico/medicoes_historicas_abas.jsf>.
- [4] ANA. Portal de metadados geospaciais da ana. Disponível em: <<http://metadados.ana.gov.br/geonetwork/srv/pt/main.home>>. Acesso em 26 nov. 2018.
- [5] ANA. SAR - Sistema de Acompanhamento de Reservatórios. Disponível em: <Disponível em: <<http://sar.ana.gov.br/MedicaoSin>>. Acesso em 18 ago. 2018.
- [6] ANA. Qualificação de Dados Hidrológicos e Reconstituição de Vazões Naturais no País. Relatório Técnico, ANA, Brasília, 2011.
- [7] ANA. Orientações para consistência de dados fluviométricos. Relatório Técnico, ANA, SGH, Brasília, 2012.
- [8] ANDERSON, P., CRITCHLEY, J., CHAN, J., COCKRAM, C., LEE, Z., THOMAS, G., TOMLINSON, B. Factor analysis of the metabolic syndrome: obesity vs insulin resistance as the central abnormality. *International Journal of Obesity* 25, 12 (2001), 1782.
- [9] ANEEL. *Atlas de energia elétrica do Brasil*, 2 ed. ANEEL, Brasília, 2005.
- [10] ANEEL. *Atlas de energia elétrica do Brasil*, 3 ed. ANEEL, Brasília, Brasil, 2008.
- [11] ANEEL. *Micro e minigeração distribuída: sistema de compensação de energia elétrica*, 1 ed. ANEEL, Brasília, 2014.
- [12] ANEEL. Resolução normativa nº 673, de 4 de agosto de 2015. estabelece os requisitos e procedimentos para a obtenção de outorga de autorização para exploração de aproveitamento de potencial hidráulico com características de pequena central hidrelétrica – pch. *Diário Oficial da União* (ago. 2015), 79.
- [13] ANEEL. Resolução normativa nº 687, de 24 de novembro de 2015. altera a resolução normativa nº 482, de 17 de abril de 2012, e os módulos 1 e 3 dos procedimentos de distribuição – prodist. *Diário Oficial da União* (nov. 2015).

- [14] ANEEL. Banco de informações de geração. Disponível em: <<http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>>. Acesso em 7 nov. 2018, 2018.
- [15] ANEEL. Micro e minigeração distribuída (ren 482/2012). Disponível em: <http://www.aneel.gov.br/outorgas/geracao/-/asset_publisher/mJhnKIi7qcJG/content/registro-de-central-geradora-de-capacidade-reduzida/655808?inheritRedirect=false>. Acesso em 18 nov. 2018, 2018.
- [16] AOKI, S. Critérios de adequação da amostragem de kaiser, orkin e mayer. Disponível em: <<http://aoki2.si.gunma-u.ac.jp/R/kmo.html>>. Acesso em 21 out. 2018.
- [17] AOKI, S. Teste de esfericidade de bartlett. Disponível em: <<http://aoki2.si.gunma-u.ac.jp/R/Bartlett.sphericity.test.html>>. Acesso em 21 out. 2018.
- [18] BELUCO, A., DE SOUZA, P. K., KRENZINGER, A. A dimensionless index evaluating the time complementarity between solar and hydraulic energies. *Renewable Energy* 33, 10 (2008), 2157–2165.
- [19] BETT, P. E., THORNTON, H. E. The climatological relationships between wind and solar energy supply in britain. *Renewable Energy* 87 (2016), 96–110.
- [20] BRASIL. Portaria mme nº 861, de 18 de outubro de 2010. *Diário Oficial da União* (outubro 2010).
- [21] BRASIL. Pretendida contribuição nacionalmente determinada para consecução do objetivo da convenção-quadro das nações unidas sobre mudança do clima. Relatório Técnico, Brasil, 2016.
- [22] CANALES, F. A., BELUCO, A., MENDES, C. A. B. Usinas hidrelétricas reversíveis no brasil e no mundo: aplicação e perspectivas. *Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental* 19, 2 (2015), 1230–1249.
- [23] CANTÃO, M. P., BESSA, M. R., BETTEGA, R., DETZEL, D. H., LIMA, J. M. Evaluation of hydro-wind complementarity in the brazilian territory by means of correlation maps. *Renewable Energy* 101 (2017), 1215–1225.
- [24] CASTRO, R., CRISPIM, J. Variability and correlation of renewable energy sources in the portuguese electrical system. *Energy for Sustainable Development* 42 (2018), 64–76.
- [25] CHEN, M.-Y., WANG, E. K., YANG, R.-J., LIU, Y.-M. Adolescent health promotion scale: development and psychometric testing. *Public Health Nursing* 20, 2 (2003), 104–110.
- [26] CIA. Cia world factbook. Disponível em: <<https://www.cia.gov/library/publications/the-world-factbook/rankorder/2201rank.html>>. Acesso em 23 ago. 2016., 2016.
- [27] COUNCIL, G. . G. W. E. Global wind energy outlook 2016. Disponível em: <, 2016.

- [28] DA SILVA, E. L. *Formação de preços em mercados de energia elétrica*. Sagra Luzzatto, 2001.
- [29] DA SILVA, R. C., DE MARCHI NETO, I., SEIFERT, S. S. Electricity supply security and the future role of renewable energy sources in brazil. *Renewable and Sustainable Energy Reviews* 59 (2016), 328–341.
- [30] DAVIE, T. *Fundamentals of hydrology*, 2 ed. Routledge, 2008.
- [31] DE AZEVEDO CAVADOS, G. Análise do impacto da introdução das fontes intermitentes no setor elétrico brasileiro: estudo de caso da região nordeste. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.
- [32] DE JONG, P., KIPERSTOK, A., SÁNCHEZ, A. S., DARGAVILLE, R., TORRES, E. A. Integrating large scale wind power into the electricity grid in the northeast of brazil. *Energy* 100 (2016), 401–415.
- [33] DE OLIVEIRA COSTA SOUZA ROSA, C., COSTA, K. A., DA SILVA CHRISTO, E., BERTAHONE, P. B. Complementarity of hydro, photovoltaic, and wind power in rio de janeiro state. *Sustainability* 9, 7 (2017), 1130.
- [34] DESTER, M. Reliability of electricity supply regarding the integration of intermittent sources in brazil’s power mix. *IEEE Latin America Transactions* 14, 3 (2016), 1302–1307.
- [35] DEVORE, J. L. *Probabilidade e Estatística para Engenharia e Ciências*, 8 ed. Cengage Learning, 2014.
- [36] DO AMARANTE, O. A. C., BROWER, M., ZACK, J., DE SÁ, A. L. *Atlas do potencial eólico brasileiro*. Ministério de Minas e Energia, Brasília, Brasil, 2001.
- [37] DOS ANJOS, P. S., DA SILVA, A. S. A., STOŠIĆ, B., STOŠIĆ, T. Long-term correlations and cross-correlations in wind speed and solar radiation temporal series from fernando de noronha island, brazil. *Physica A: Statistical Mechanics and its Applications* 424 (2015), 90–96.
- [38] DOS REIS, L. B. *Geração de Energia Elétrica*, 2 ed. Manole, 2011.
- [39] DOS SANTOS, M. A. F., DE FÁTIMA MATTIELLO FRANCISCO, M., YAMAGUTI, W. O sistema nacional de dados ambientais e a coleta de dados por satélite. Em *Anais XVI Simpósio Brasileiro de Sensoriamento Remoto* (2013), p. 9116–9123.
- [40] DUJARDIN, J., KAHL, A., KRUYT, B., BARTLETT, S., LEHNING, M. Interplay between photovoltaic, wind energy and storage hydropower in a fully renewable switzerland. *Energy* 135 (2017), 513–525.
- [41] EIA. International Energy Outlook. Relatório Técnico, U.S. Energy Information Administration, Washington, 2016.
- [42] EIA. International energy statistics. Disponível em: <<https://www.eia.gov/beta/international/data/browser>>, 2018.

- [43] ELETROBRAS. Programas – programa de incentivo às fontes alternativas de energia elétrica (proinfra). Disponível em: <<http://www.eletrobras.com/elb/Proinfra/data/Pages/LUMISABB61D26PTBRIE.htm>>. Acesso em 17 jul. 2016, 2016.
- [44] ENGELAND, K., BORGA, M., CREUTIN, J.-D., FRANÇOIS, B., RAMOS, M.-H., VIDAL, J.-P. Space-time variability of climate variables and intermittent renewable electricity production—a review. *Renewable and Sustainable Energy Reviews* 79 (2017), 600–617.
- [45] EPE. Informe técnico epe-dee-it-56/2013. leilão a-5 de 2013. preços médios de referência dos combustíveis vinculados ao cvu de usinas termelétricas. 11 de junho de 2013. Relatório Técnico, EPE, 2013.
- [46] EPE. Nota técnica dea 19/14: Inserção da geração fotovoltaica distribuída no brasil – condicionantes e impactos. Relatório Técnico, EPE, Rio de Janeiro, 2014.
- [47] EPE. Balanço energético nacional 2015 – ano base 2014. Relatório Técnico, EPE, Rio de Janeiro, 2015.
- [48] EPE. *Anuário Estatístico de Energia Elétrica*. EPE, Brasília, 2017.
- [49] EPE. Balanço Energético Nacional: ano base 2017. Relatório Técnico, Empresa de Pesquisa Energética, Rio de Janeiro, 2018.
- [50] ESHA. *Guide on How to Develop a Small Hydropower Plant*, 2004.
- [51] ESRI. Arcgis online. Disponível em: <<https://www.arcgis.com/home/webmap/viewer.html>>. Acesso em 22 ago. 2018, 2018.
- [52] F.HAIR, J., C.BLACK, W., J.BABIN, B., ANDERSON, R. E., TATHAM, R. L. *Análise Multivariada de Dados*, 6 ed. Bookman, 2009.
- [53] FLICKR. Savonius rotor windrad in rainau buch. Disponível em: <<https://www.flickr.com/photos/thomaskappel/16892899462/in/photolist-xK856-R3A9KZ-hFvYLt-xK7Yj-akKL27-rJLwth>>. Acesso em 16 jan. 2019.
- [54] FLICKR. Unique éolienne des îles, proche de la lagune du havre-aux-maisons. Disponível em: <<https://www.flickr.com/photos/ebbtide/7770591064/in/photolist-zmrWKC-eqbia7-fMYGZ-4Dcrp2-bfCfst-PPGHxD-7tpQ43-akKL27-XV6jBL-2tsyp>>. Acesso em 16 jan. 2019.
- [55] FRANÇOIS, B., ZOCCATELLI, D., BORGA, M. Assessing small hydro/solar power complementarity in ungauged mountainous areas: A crash test study for hydrological prediction methods. *Energy* 127 (2017), 716–729.
- [56] FRANÇOIS, B., BORGA, M., CREUTIN, J., HINGRAY, B., RAYNAUD, D., SAUTERLEUTE, J. Complementarity between solar and hydro power: Sensitivity study to climate characteristics in northern-italy. *Renewable Energy* 86 (2016), 543 – 553.
- [57] FREEIMAGES. Free moinho de vento americano stock photo. Disponível em: <<https://pt.freeimages.com/photo/american-windmill-1355153>>. Acesso em 16 jan. 2019.

- [58] FREEIMAGES. Free moinhos de vento stock photo. Disponível em: <<https://pt.freeimages.com/photo/windmills-1315947>>. Acesso em 16 jan. 2019.
- [59] FREEIMAGES. Free turbina eólica 1 stock photo. Disponível em: <<https://pt.freeimages.com/photo/wind-turbine-1-1210048>>. Acesso em 16 jan. 2019.
- [60] GLASBEY, C. A. Imputation of missing values in spatio-temporal solar radiation data. *Environmetrics* 6, 4 (7 1995), 363–371.
- [61] GOLDFARB, D., IDNANI, A. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming* 27 (1983), 1–33.
- [62] GONZÁLEZ, J. A. C., PÉREZ, R. C., SANTOS, A. C., GIL, M. A. C. *Centrales de energías renovables: generación eléctrica con energías renovables*, 2 ed. Pearson Educación, 2009.
- [63] GUTIERREZ, C., GAERTNER, M. Á., PERPIÑÁN, O., GALLARDO, C., SÁNCHEZ, E. A multi-step scheme for spatial analysis of solar and photovoltaic production variability and complementarity. *Solar Energy* 158 (2017), 100–116.
- [64] HAIR, J. F., BLACK, W. C., BABIN, B. J., ANDERSON, R. E., TATHAM, R. L. *Análise Multivariada de Dados*, 6 ed. Bookman, 2009.
- [65] HEIDE, D., BREMEN, L. V., GREINER, M., HOFFMANN, C., SPECKMANN, M., BOFINGER, S. Seasonal optimal mix of wind and solar power in a future, highly renewable europe. *Renewable Energy* 35, 11 (2010), 2483–2489.
- [66] HILLIER, F. S., LIEBERMAN, G. J. *Introdução à Pesquisa Operacional*, 8 ed. McGraw Hill, 2006.
- [67] HINRICHS, R. A., KLEINBACH, M., REIS, L. B. *Energia e Meio Ambiente*, 4 ed. Cengage Learning, 2011.
- [68] HOICKA, C. E., ROWLANDS, I. H. Solar and wind resource complementarity: Advancing options for renewable electricity integration in ontario, canada. *Renewable Energy* 36, 1 (2011), 97–107.
- [69] IBGE. Serviços do arcgis. Disponível em: <<https://mapas.ibge.gov.br/interativos/servicos/wms-do-arcgis.html>>. Acesso em 7 nov. 2018, 2018.
- [70] IEA. World Energy Outlook 2018. Relatório Técnico, International Energy Agency, 2018.
- [71] INPE. Sonda - validação dos dados. Disponível em: <<http://sonda.ccst.inpe.br/infos/validacao.html>>.
- [72] INPE/CPTEC. MARSP. Disponível em: <<http://marsp.cptec.inpe.br/pcd.shtml>>. Acesso em 19 nov. 2018.
- [73] INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS INPE. Dados Históricos de Plataformas de Coletas de Dados. Disponível em: <<http://sinda.crn.inpe.br/PCD/SITE/novo/site/historico/index.php>>. Acesso em 10 dez. 2017.

- [74] JACK KLINE, MICHAEL MILLIGAN. An Evaluation of Hourly Average Wind-Speed Estimation Techniques. Relatório Técnico, National Renewable Energy Laboratory, Golden, Colorado, 1998.
- [75] JOHNSON, R. A., WICHERN, D. W. *Applied Multivariate Statistical Analysis*, 6 ed. Prentice Hall, 2007.
- [76] JURASZ, J., BELUCO, A., CANALES, F. A. The impact of complementarity on power supply reliability of small scale hybrid energy systems. *Energy* 161 (2018), 737–743.
- [77] JURASZ, J., CIAPAŁA, B. Integrating photovoltaics into energy systems by using a run-off-river power plant with pondage to smooth energy exchange with the power grid. *Applied energy* 198 (2017), 21–35.
- [78] KAISER, H. F., RICE, J. Little jiffy, mark iv. *Educational and psychological measurement* 34, 1 (1974), 111–117.
- [79] KALTSCHMITT, M., STREICHER, W., WIESE, A. *Renewable Energy: Technology, Economics and Environment*, 1 ed. Springer, 2007.
- [80] KOUGIAS, I., SZABÓ, S., MONFORTI-FERRARIO, F., HULD, T., BÓDIS, K. A methodology for optimization of the complementarity between small-hydropower plants and solar pv systems. *Renewable Energy* 87 (2016), 1023–1030.
- [81] LEMOS FILHO, G., OLIVEIRA, E., LYRA, G., CRUZ, S. Assimilação, controle de qualidade e análise de dados de meteorológicos apoiados por proveniência. Em *VII Brazilian E-science Workshop* (2013).
- [82] LUIZ, E. W., MARTINS, F. R., GONÇALVES, A. R., PEREIRA, E. B. Analysis of intra-day solar irradiance variability in different Brazilian climate zones. *Solar Energy* 167 (6 2018), 210–219.
- [83] MAGINA, F. C. MARSP - Monitoramento Ambiental Eixo Rio de Janeiro e São Paulo - Amostragem. Disponível em: <<http://marsp.cptec.inpe.br/amostragem.shtml>>. Acesso em 19 nov. 2018.
- [84] MAGINA, F. C. MARSP - Monitoramento Ambiental Eixo Rio de Janeiro e São Paulo - Sensores de Plataformas de Coleta de Dados. Disponível em: <<http://marsp.cptec.inpe.br/sensores.shtml>>. Acesso em 19 nov. 2018.
- [85] MARTINS, F. R., PEREIRA, E. B. Estudo comparativo da confiabilidade de estimativas de irradiação solar para o sudeste brasileiro obtidas a partir de dados de satélite e por interpolação/extrapolação de dados de superfície. *Revista Brasileira de Geofísica* 29 (6 2011), 265–276.
- [86] MME. Plano nacional de energia 2030. Relatório Técnico, MME, Brasília, 2007.
- [87] MOLNÁR, P., CAMARGO, L. S., RAMOS, D. Applying copulas functions for wind and hydro complementarity evaluation: A brazilian case. Em *2015 12th International Conference on the European Energy Market (EEM)* (2015), IEEE.

- [88] MOURA, P. S., DE ALMEIDA, A. T. Multi-objective optimization of a mixed renewable system with demand-side management. *Renewable and Sustainable Energy Reviews* 14, 5 (2010), 1461–1468.
- [89] ONS. Procedimentos de REde - Submódulo 10.22 : RO-AO.BR.02. Relatório Técnico, ONS, 2015.
- [90] ONS. O que é o sin - sistema interligado nacional. Disponível em: <http://www.ons.org.br/conheca_sistema/o_que_e_sin.aspx>, 2016.
- [91] ONS. Plano da operação energética 2016/2020 – pen 2016, sumário executivo. Relatório Técnico, ONS, Rio de Janeiro, 2016.
- [92] ONS. Submódulo 9.2 - Acompanhamento, análise e tratamento dos dados hidroenergéticos do Sistema Interligado Nacional. Relatório Técnico, ONS, 2016.
- [93] ONS. Curva de carga horária. Disponível em: <http://ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/curva_carga_horaria.aspx>. Acesso em 20 nov. 2017, 2017.
- [94] ONS. O sistema interligado nacional. Disponível em: <<http://ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>>. Acesso em 18 nov. 2018, 2018.
- [95] ONUDI. Pequenas centrais hidrelétricas: Aplicação ao seu desenvolvimento na América Latina e no Caribe. Relatório Técnico, ONUDI, 2013.
- [96] PERDOMO, R., BANGUERO, E., GORDILLO, G. Statistical modeling for global solar radiation forecasting in Bogotá. Em *2010 35th IEEE Photovoltaic Specialists Conference* (6 2010), IEEE, p. 002374–002379.
- [97] PEREIRA, E. B., MARTINS, F. R., DE ABREU, S. L., RÜTHER, R. *Atlas brasileiro de energia solar*, 1 ed. INPE, São José dos Campos, 2006.
- [98] POOLE, D. *Linear Algebra: A Modern Introduction*, 4 ed. Cengage Learning, 2015.
- [99] RAMOS-CALZADO, P., GOMEZ-CAMACHO, J., PEREZ-BERNAL, F., PITA-LOPEZ, M. F. A novel approach to precipitation series completion in climatological datasets: application to Andalusia. *International Journal of Climatology* 28 (2008), 1525–1534.
- [100] RAMÍREZ, J. Merra-based study of the wind/solar resources and their complementarity to the hydro resource for power generation in Colombia. Dissertação de Mestrado, Carl von Ossietzky Universität Oldenburg, Alemanha, 2015.
- [101] RISSO, A., BELUCO, A., ALVES, R. D. C. M. Complementarity roses evaluating spatial complementarity in time between energy resources. *Energies* 11 (2018), 1–14.
- [102] SCHMIDT, J., CANCELLA, R., JR, A. O. P. The role of wind power and solar pv in reducing risks in the Brazilian hydro-thermal power system. *Energy* 115 (2016), 1748–1757.

- [103] SCHNEIDER, T. Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values. *Journal of Climate* 14, 5 (3 2001), 853–871.
- [104] SCHWANDT, M., CHHATBAR, K., MEYER, R., FROSS, K., MITRA, I., VASHISTHA, R., GIRIDHAR, G., GOMATHINAYAGAM, S., KUMAR, A. Development and Test of Gap Filling Procedures for Solar Radiation Data of the Indian SRRA Measurement Network. *Energy Procedia* 57 (1 2014), 1100–1109.
- [105] SHALIZI, C. R. Advanced data analysis from an elementary point of view. Disponível em: <<http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/>>. Acesso em 16 nov. 2018, 2018.
- [106] SILVA, A. R., PIMENTA, F. M., ASSIREU, A. T., SPYRIDES, M. H. C. Complementarity of brazil's hydro and offshore wind power. *Renewable and Sustainable Energy Reviews* 56 (2016), 413–427.
- [107] SKOPLAKI, E., PALYVOS, J. A. On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. *Solar energy* 83, 5 (2009), 614–624.
- [108] SORENSEN, B. *Renewable Energy: Physics, Engineering, Environmental Impacts, Economics & Planning*, 3 ed. Elsevier, 2004.
- [109] SPIEGEL, M. R. *Estatística*. Pearson Makron Books, 1993.
- [110] TAHA, H. A. *Pesquisa Operacional: uma visão geral*, 8 ed. Pearson Prentice Hall, 2008.
- [111] TOLMASQUIM, M. T., GUERREIRO, A., GORINI, R. Matriz energética brasileira: uma prospectiva. *Novos estudos-CEBRAP*, 79 (2007), 47–69.
- [112] TRYFOS, P. *Methods for business analysis and forecasting: text and cases*. Wiley, 1997, cap. Factor analysis.
- [113] TURRADO, C. C., DEL CARMEN MEIZOSO LÓPEZ, M., LASHERAS, F. S., GÓMEZ, B. A. R., ROLLÉ, J. L. C., DE COS JUEZ, F. J. Missing data imputation of solar radiation data under different atmospheric conditions. *Sensors (Basel, Switzerland)* 14, 11 (10 2014), 20382–99.
- [114] TWIDELL, J., WEIR, T. *Renewable energy Resources*, 2 ed. Taylor & Francis, 2006.
- [115] WEBMET. WebMET - The Meteorological Resource Center. Disponível em: <http://www.webmet.com/met_monitoring/681.html>.
- [116] WILLIAMS, B., ONSMAN, A., BROWN, T. Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine* 8, 3 (2010).
- [117] WMO. Guide on the global data-processing system. Relatório Técnico, World Meteorological Organisation, Geneva, Switzerland, 1993.
- [118] WMO. Guide to the global observing system. Relatório Técnico, World Meteorological Organisation, Geneva, Switzerland, 2017.

-
- [119] WWEA. Small wind world report – 2016 summary. Relatório Técnico, World Wind Energy Association, Bonn, Germany, 2016.
- [120] YEOMANS, K. A., GOLDER, P. A. The guttman-kaiser criterion as a predictor of the number of common factors. *The Statistician* (1982), 221–229.