

GRUPO DE ESTUDO DE PLANEJAMENTO DE SISTEMAS ELÉTRICOS - GPL

METODOLOGIA GEORREFERENCIA PARA AGRUPAMENTO DE USINAS EÓLICAS E FOTOVOLTAICAS

**FELIPE MOREIRA GONÇALVES(1)
EMPRESA DE PESQUISA ENERGÉTICA(1)**

RESUMO

O agrupamento de usinas eólicas e solares no setor elétrico brasileiro é fundamental para a representação das incertezas dessas fontes nos modelos do setor elétrico. A metodologia Georreferenciada tem como objetivo definir áreas que internamente possuam recursos eólicos ou fotovoltaicos semelhantes, o que permite o agrupamento das usinas. Para isso foram utilizados como dados de entrada as usinas eólicas do AMA e dados solarimétricos simulados no SAM, que foram divididos em grupos por ferramenta de SIG com base em restrição geográfica e no perfil de potência mensal e horário. Esses grupos foram posteriormente ajustados por meio de testes estatísticos e análise de gráficos.

PALAVRAS-CHAVE

Agrupamento, Clusters, Usinas Fotovoltaicas, Usinas Eólicas, Dados AMA

1.0 INTRODUÇÃO

Dado o crescimento da capacidade instalada das usinas eólicas e solares no setor elétrico brasileiro, a clusterização dessas usinas é uma etapa cada vez mais importante para a representação das incertezas dessas fontes nos modelos utilizados no setor.

A principal razão para agrupar usinas é a melhoria da performance computacional ao rodar os modelos. Além disso, a visualização dos resultados das usinas de forma agrupada também facilita as análises, pois permite uma representação mais clara da operação, bem como a análise dos recursos eólico e solar.

A metodologia Georreferenciada tem como objetivo definir áreas que internamente possuam recursos eólicos ou fotovoltaicos semelhantes, o que permite o agrupamento das usinas. A metodologia é realizada em duas etapas, considerando dados de entrada mensais e horários, e utiliza ferramentas de softwares de Sistema de Informação Geográfica (SIG) para realizar uma divisão inicial de grupos considerando restrições geográficas e dados de geração de usinas. Por meio de análise de gráficos do perfil das usinas e testes estatísticos a divisão dos grupos é ajustada.

A metodologia tem como produto final polígonos criados para cada grupo utilizando ferramentas de SIG permitindo a designação de grupo para qualquer usina, mesmo que não esteja presente nos dados de entrada. Também foi desenvolvida uma ferramenta em linguagem R que permite a definição de grupo para usinas de forma automática a partir dos resultados da aplicação da metodologia.

2.0 DADOS DE ENTRADA

O objetivo da metodologia Georreferenciada é o agrupamento de usinas com base no recurso, independentemente da fonte. Dessa forma, na aplicação da metodologia se buscou utilizar dados de entrada que sejam possíveis de serem comparados do ponto de vista do recurso, ou seja, usinas, simuladas ou reais, com configurações e equipamentos semelhantes de modo a evitar que as diferenças significativas de tecnologia possam interferir na divisão de grupos. Os dados de entradas selecionados para aplicação da metodologia definem, de forma aproximada, os limites geográficos da região que a metodologia poderá ser aplicada.

O AMA consiste em dados meteorológicos medidos pelos donos de parques eólicos e enviados para EPE por força de contrato de energia. Para aplicação da metodologia Georreferenciada nos dados eólicos, os dados de geração foram obtidos por simulação da geração dos parques a partir dos dados do AMA. Ao todo foram utilizados 461 parques eólicos, sendo 384 no Nordeste e 77 no Rio Grande do Sul, a localização desses parques é apresentada pela Figura 1.

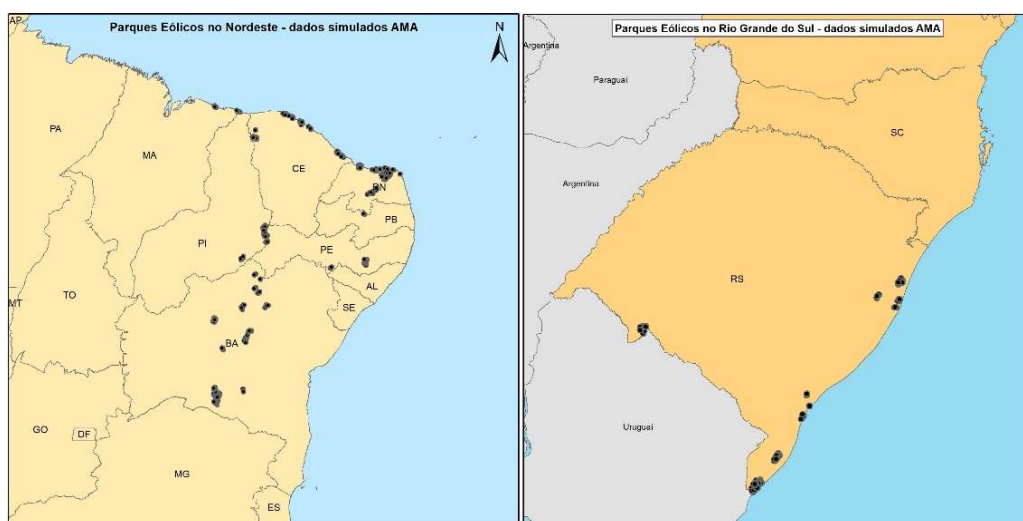


Figura 1 – Dados de entrada eólicas no Nordeste

Os dados utilizados do AMA são provenientes de simulações que consideram modelos de aerogeradores da base de dados de projetos do AEGE, que deverão ser instalados nos próximos anos, de acordo com as tendências do mercado. Dessa forma, o agrupamento é realizado comparando configurações de usinas do mesmo período sem grandes discrepâncias, evitando um viés tecnológico na definição dos grupos

A partir dos dados do AMA e das curvas de potência das turbinas, foi calculada a energia de cada parque a cada 10 minutos, em três séries de geração horária entre os anos de 2017 e 2019. A metodologia detalhada da obtenção desses dados é descrita em Nota Técnica da EPE (2021) (1).

Os dados originalmente em MWh foram normalizados para Potência Unitária (PU) dividindo pelo valor de geração máximo de cada usina. Para obtenção dos dados mensais, inicialmente foi obtido um valor médio de cada usina para cada um dos 36 meses, em seguida foi realizada a divisão do valor de cada mês pelo maior valor de cada usina resultando em valores de 0 a 1 em PU. A partir desse resultado foi realizado a média dos três anos para cada mês para que, enfim, obtenha-se os dados mensais. Para obtenção dos dados horário, foi calculado a média para cada hora do dia de cada usina e em seguida foi feito a divisão pelo seu maior valor, obtendo assim valores em PU.

Para aplicação da metodologia utilizando dados do recurso fotovoltaico, foram simulados no software System Advisor Model (SAM) dados horários de irradiação global horizontal, proveniente do BRASIL-SR, que é fornecida pelo Laboratório de Modelagem e Estudo de Recursos Renováveis de Energia (LABREN) do INPE, e de temperatura, proveniente do MERRA 2. A Figura 2 apresenta a localização das usinas simuladas.

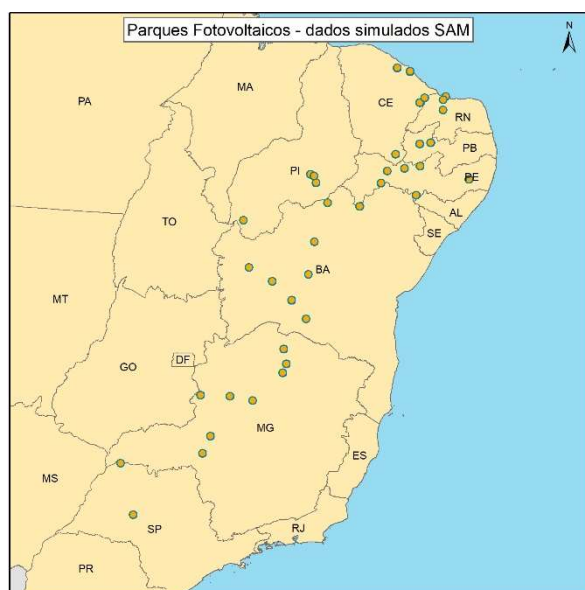


Figura 2 – Dados de entrada Usinas Fotovoltaicas

Ao todo foram simuladas 38 usinas fotovoltaicas nos municípios que tiveram usinas contratadas nos leilões de energia entre 2014 e 2019 e estudos prospectivos. As simulações foram realizadas considerando equipamentos e arranjo idênticos para todas as usinas e tiveram como resultado séries horárias com doze anos já em fator de capacidade, de forma que não é necessário normalizar. A metodologia detalhada da simulação pode ser consultada em Nota Técnica da EPE (2021) (1).

Os dados simulados de séries horárias com 12 anos foram convertidos em dados mensais e horários para utilização na metodologia Georreferenciada por meio do cálculo da média aritmética dos seus respectivos meses e horas.

3.0 METODOLOGIA GEORREFERENCIADA

A primeira etapa da metodologia Georreferenciada é realizada no software ArcGIS adicionando as usinas escolhidas como dado de entrada, em formato georreferenciado com os dados em potência unitária ou fator de capacidade do recurso mensal e horário presentes em sua tabela de atributos. A partir dessas informações, a ferramenta Grouping Analysis, presente no ArcGIS, separa os pontos em um número de grupos determinado pelo usuário, com base nos seus atributos e considerando uma restrição geográfica. Na metodologia Georreferenciada foi definida como restrição geográfica a triangulação de Delauney, que determina que para fazer parte de um grupo é necessário que um ponto tenha pelo menos um vizinho em comum com outro ponto do grupo.

A ferramenta Grouping Analysis separa os grupos utilizando o método Spatial "K"luster Analysis by Tree Edge Removal (Skater). O primeiro passo desse método é construir grafos de conectividade que representam os relacionamentos de vizinhança das feições. A partir desses grafos, uma árvore de extensão mínima é criada que resume a relação espacial das feições e a similaridade dos dados, sendo que feições se tornam nós conectados por arestas ponderadas com valor proporcional à similaridade dos objetos que ela conecta. Para criar dois grupos, uma aresta é podada de modo a minimizar as diferenças nos grupos resultantes, esse último processo se repete até atingir o número de grupos desejados. A aresta a ser cortada é escolhida de modo que minimize as diferenças nos grupos resultantes e evite, quando possível, a criação de grupos com apenas um elemento.

A ferramenta Grouping Analysis também faz uma análise do número ideal de grupos baseado nos atributos selecionados como dados de entrada. Essa análise é feita por meio da estatística Pseudo-F desenvolvida por Calinski e Harabasz.

A metodologia Georreferenciada utiliza o maior valor obtido no teste Pseudo-F para definir, de modo inicial, o número de grupos. Posteriormente, essa escolha do número de grupos é verificada e otimizada por análise de gráfico e testes estatísticos de modo a identificar se existem grupos com perfis muito parecidos que podem ser agregados ou usinas com perfis muito diferentes dentro de um grupo justificando a divisão do conjunto em dois ou mais grupos.

A análise por testes estatístico T de Student verifica se os valores médios de dois grupos são estatisticamente iguais considerando um nível de confiança. Na metodologia apresentada, considerou-se 95% como nível de confiança e consequentemente 0.05 como p-valor mínimo para aprovar hipótese de os dados pertencerem ao mesmo grupo. Da mesma forma da análise por gráfico, o teste estatístico T permite comparar se diferentes grupos podem ser agregados e se é necessário dividir um grupo em dois ou mais novos conjuntos.

O resultado da aplicação da ferramenta Grouping Analysis pode gerar algumas dúvidas pontuais de designação de grupo de uma ou mais usinas considerando a localização geográfica delas. As análises para definição de número de grupos por gráficos e testes estatísticos também podem ser aplicados neste tipo de situação com a diferença que para a estatísticas T, ao invés de buscar um p-valor de 0.05 é verificado qual p-valor é maior na comparação das usinas com diferentes grupos.

Foram gerados polígonos para cada grupo com o objetivo de designar a qual grupo pertencem as usinas que não fizeram parte dos dados de entrada, seja por não terem dados de geração no momento da aplicação da metodologia ou por não serem adequados aos objetivos da metodologia, sendo necessário apenas inserir suas coordenadas geográficas em mapa contendo a camada dos polígonos criados.

A definição dos polígonos foi realizada por funções disponíveis em qualquer software de SIG. Primeiramente foi aplicado o diagrama de Voronoi nas usinas utilizadas como dado de entrada na metodologia resultando em polígonos no qual qualquer ponto em seu interior tenha como menor distância uma das usinas. Em sequência é realizada a função Dissolve para juntar todos os polígonos cujas usinas pertencem ao mesmo grupo, de modo a se obter um polígono por grupo.

Também foi desenvolvida uma ferramenta em linguagem R, que utiliza como insumo os polígonos gerados e a partir de uma tabela com coordenadas de múltiplas usinas fornecida pelo usuário, indica em quais grupo cada usina pertence de acordo com os resultados da metodologia Georreferenciada. Essa ferramenta permite a aplicação do

resultado da metodologia de forma automática, sem ser necessária a utilização de um software de SIG para a definição do grupo da usina.

4.0 – APLICAÇÃO METODOLOGIA

4.1 – Usinas Eólicas

As estações do AMA foram inseridas no ArcGIS, em duas camadas, uma do Nordeste e outra do Rio Grande do Sul, em formato shapefile com tabela de atributos contendo os dados da geração simulada mensal e horária. A partir desses dados foi aplicada a metodologia Georreferenciada de forma separada para o Nordeste e o Rio Grande do Sul com o objetivo de dividir as usinas em diferentes regiões com base no recurso eólico.

4.1.1 – Nordeste

A aplicação da ferramenta do ArcGIS, Grouping Analysis, nos dados mensais do conjunto de 384 parques eólicos proveniente do AMA no Nordeste, teve como resultado teste estatísticos Pseudo-F indicando valores maiores para 2, 5 ou 6 grupos, conforme gráfico apresentado pela Figura 3.

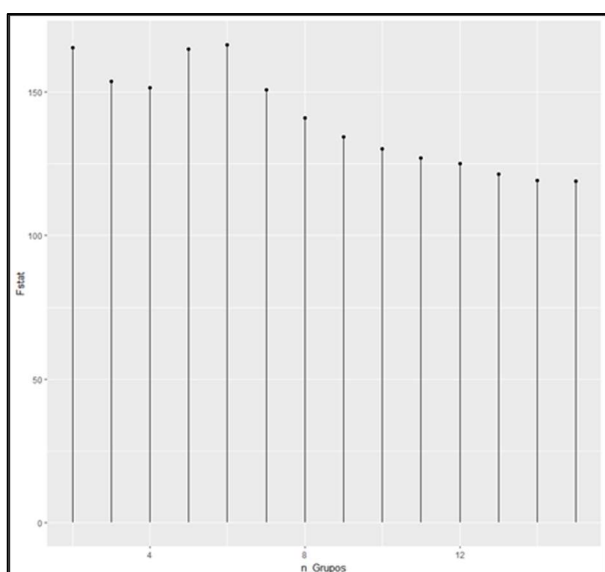


Figura 3 - Resultado teste estatístico Pseudo-F para dados simulados AMA no Nordeste

O perfil mensal das usinas eólicas do Nordeste dividida em dois grupos demonstrou uma diferença clara entre os grupos 1 e 2, porém cada grupo ser muito heterogêneo, o que indica que dividir o Nordeste em apenas duas regiões de recurso eólico é insuficiente, sendo necessário explorar a divisão em mais grupos.

O mapa abaixo apresenta as divisões resultantes do Grouping Analysis para cinco e seis grupos de eólicas no Nordeste.

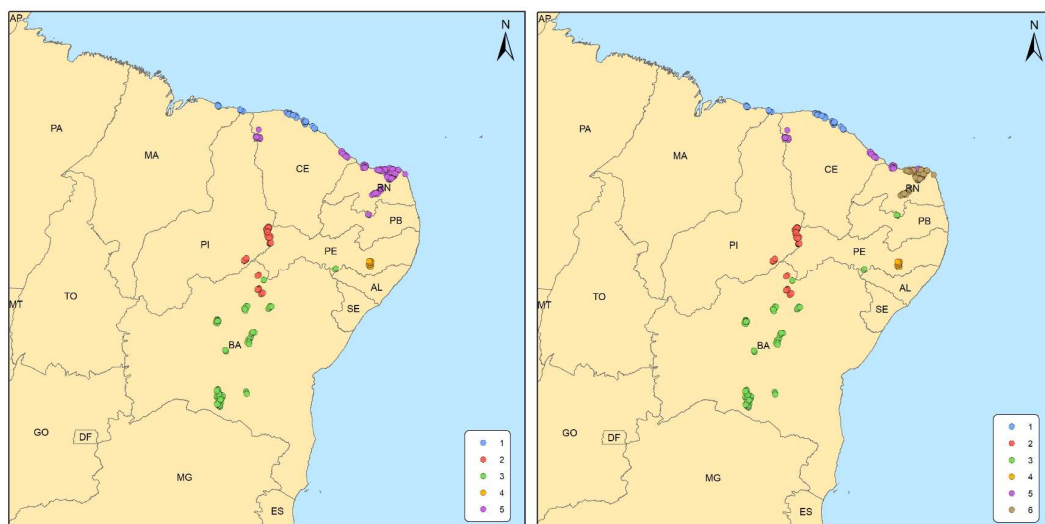


Figura 4 – Divisão inicial de eólicas no Nordeste em 5 e 6 grupos

A divisão em seis grupos apresenta como novidade a existência de dois grupos na região do litoral potiguar, interior cearense e fronteira dos dois estados.

Optou-se pela divisão de usinas eólicas no Nordeste em seis grupos considerando que a estatística T indicou que os grupos gerados na nova divisão são realmente distintos com p-valor de $5,6 \cdot 10^{-15}$ e que a estatística pseudo-F resultou em valor maior para seis grupos do que cinco. Além disso, foi realizada uma análise da homogeneidade dos meses de maior e menor produção dentro dos grupos e se verificou uma maior coerência na divisão em seis grupos.

Com a definição dos grupos mensais, foi realizada a inserção dos dados de produção horários no ArcGIS e para cada um dos seis grupos se executou o Grouping Analysis com o objetivo de verificar internamente a existência de diferentes perfis com base nos dados horários.

Todas as usinas dos grupos dois e quatro apresentam perfis horários similares, de modo que não se justifica uma nova divisão interna. A realização de testes estatísticos e análises gráficas para diferentes números de divisão nos grupos um, três, cinco e seis resultou na identificação diferentes perfis horários cujos perfis podem ser identificados na Figura 5.

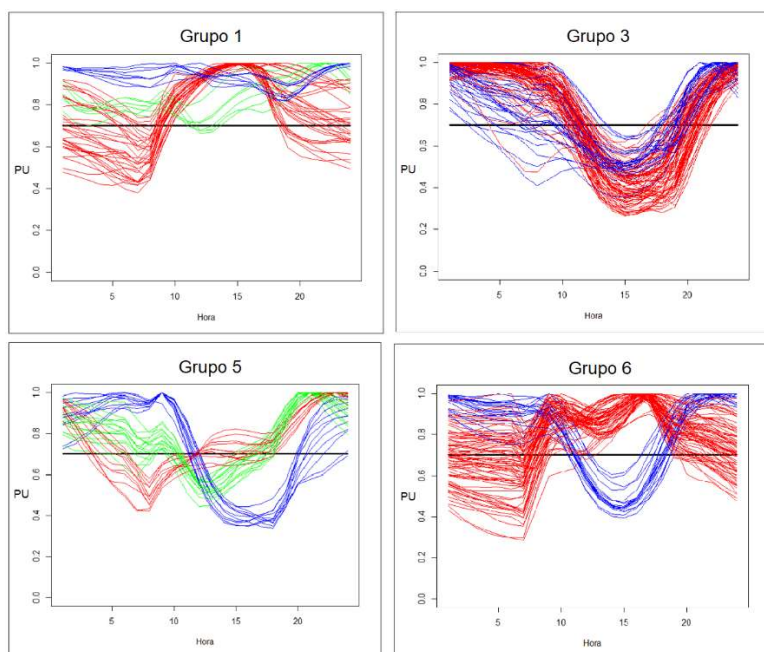


Figura 5 – Diferentes perfis horários grupos mensais um, três, cinco e seis

A figura 6 apresenta o mapa com o resultado do agrupamento do recurso eólico no Nordeste já incluindo os polígonos que indicam quais regiões tem o mesmo recurso eólico.

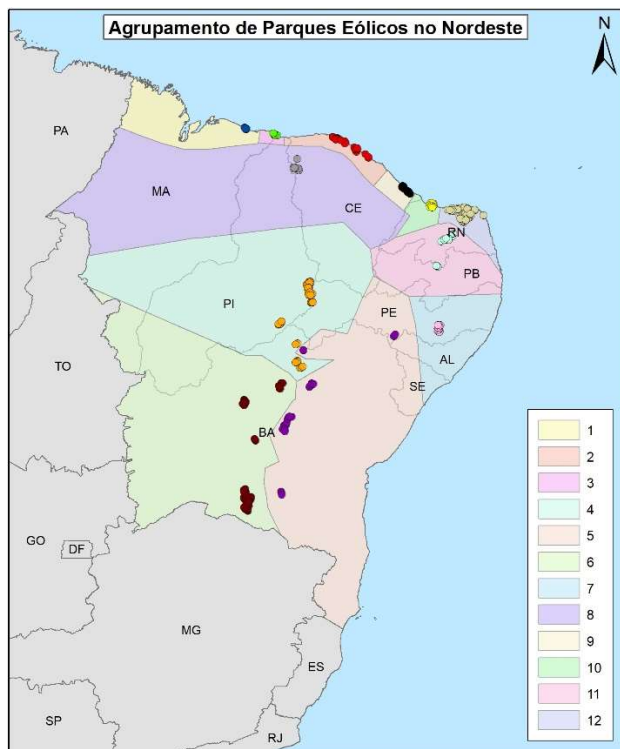


Figura 6 – Resultado agrupamento eólicas no Nordeste

4.1.2 – Rio Grande do Sul

A aplicação da ferramenta Grouping Analysis nos dados mensais das 77 usinas eólicas do Rio Grande do Sul presentes no AMA indicou uma divisão ótima de três grupos conforme mostra Figura 7.

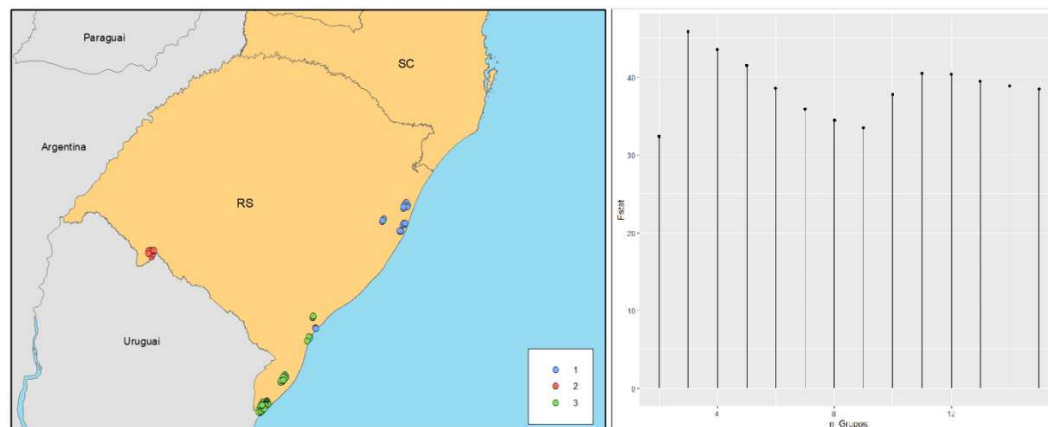


Figura 7 – Resultado inicial divisão de grupos das eólicas do AMA no Rio Grande do Sul

É possível observar no mapa da Figura 7 que existem quatro regiões de usinas no Rio Grande do Sul, no litoral norte, na região sudoeste, no extremo sul e no litoral entre o extremo sul e o litoral norte, sendo que esta última região é a única que ficou indefinida com usinas nos grupos um e três. A comparação do p-valor do teste estatístico T indicam que cada grupo de eólicas na região de dúvida foi definido corretamente para o grupo de maior proximidade estatística. No entanto, a comparação gráfica dos perfis mensais dessas usinas com os perfis das demais usinas desses grupos indica uma maior aderência com grupo três, conforme pode ser visto na Figura 8.

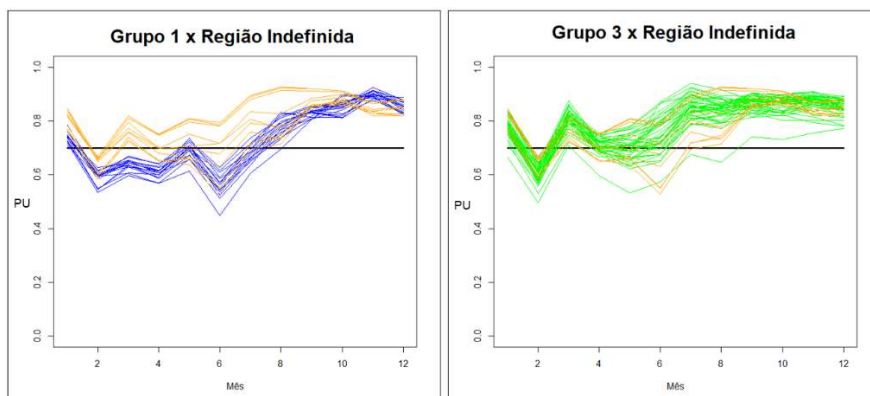


Figura 8 – Comparação gráfica entre região indefinida e grupos um e três

Além da comparação gráfica e da proximidade geográfica, verificou-se que a alocação das usinas da região de dúvida para o grupo três resulta em um desvio padrão médio dos grupos menor (0,0324 MWh versus 0,0391 MWh).

Foi realizada a execução da ferramenta Grouping Analysis com os dados mensais das usinas para cada um dos três grupos mensais com o objetivo de verificar a existência de diferentes perfis horários justificando a divisão interna em algum dos grupos. O grupo três foi o único em que se identificou mais de um perfil horário, com as usinas no litoral ganhando produção após às 10h enquanto as usinas no interior perdem produção no mesmo horário. A Figura 9 apresenta a nova divisão do grupo três e os perfis de produção horário das usinas.

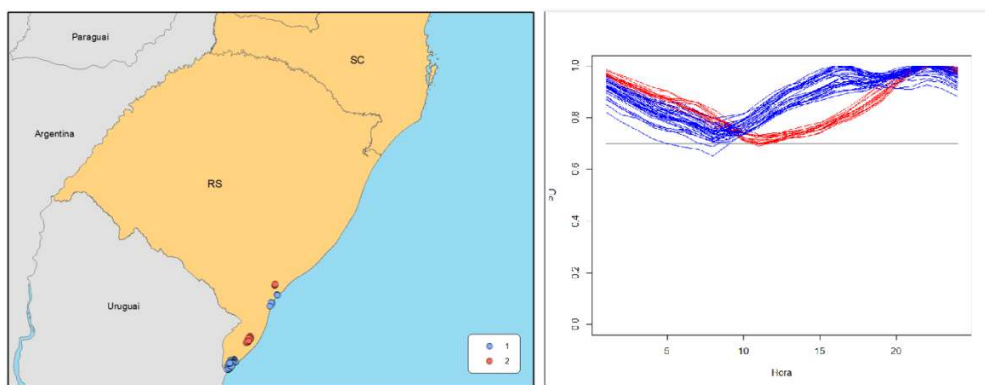


Figura 9 – Os Diferente perfis de produção horário presentes no grupo mensal três

O resultado do agrupamento das eólicas no Rio Grande do Sul com os polígonos gerados, que indicam regiões de mesmo recurso eólico, é apresentado no mapa abaixo, denominado Figura 10. Os números dos grupos nesse mapa já foram alterados para que haja uma continuidade com os grupos eólicos do Nordeste.

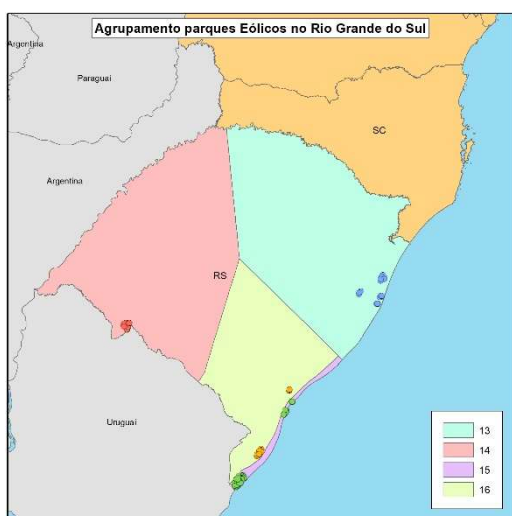


Figura 10 – Resultado agrupamento eólicas no Rio Grande do Sul

4.2 – Usinas Fotovoltaicas

A execução do Grouping Analysis utilizando os dados solatimétricos mensais resultou em valores da estatística pseudo F semelhantes entre oito e quinze grupos. Optou-se por utilizar inicialmente a divisão realizada pela ferramenta em onze grupos, pois a partir desse número, atinge-se um patamar e o valor da pseudo F não aumenta de forma significativa. A análise dos gráficos e de testes estatísticos indicou que os grupos resultantes são bastante coesos, não sendo necessário novas divisões. No entanto, ao comparar todos os grupos entre si pelo teste estatístico T de Student, em alguns casos, a hipótese de pertencerem ao mesmo grupo foi aceita, sendo assim possível fundir grupos. A Figura 11 apresenta o p-valor do teste estatístico entre todos os grupos com pelo menos duas usinas, foi considerado nível de confiança de 95%, então valores superiores a 0.05 indicam a possibilidade de unir os grupos.

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	1.00000000	0.09037339	0.013094738	0.0126895057	0.02345488	0.075394045	0.0147789378	0.083690984	0.018016933	0.0516943898
[2,]	0.09037339	1.00000000	0.051803765	0.0164314750	0.13320170	0.062583462	0.0140965571	0.090587677	0.095799492	0.2473016360
[3,]	0.01309474	0.05180377	1.000000000	0.0384515501	0.05796002	0.005616811	0.0044897485	0.041134656	0.060874680	0.0104579384
[4,]	0.01268951	0.01643148	0.038451550	1.000000000	0.01103302	0.020325565	0.0006395031	0.038253423	0.006617570	0.0026015746
[5,]	0.02345488	0.13320170	0.057960020	0.0110330204	1.00000000	0.018068992	0.0111034828	0.133660991	0.043627978	0.0137593145
[6,]	0.07539405	0.06258346	0.005616811	0.0203255647	0.01806899	1.000000000	0.0018865056	0.025134678	0.005385315	0.0208058463
[7,]	0.01477894	0.01409656	0.004489749	0.0006395031	0.01110348	0.001886506	1.000000000	0.002048280	0.010301093	0.0005146695
[8,]	0.08369098	0.09058768	0.041134656	0.0382534234	0.13366099	0.025134678	0.0020482803	1.000000000	0.008578119	0.0909853458
[9,]	0.01801693	0.09579949	0.060874680	0.0066175697	0.04362798	0.005385315	0.0103010929	0.008578119	1.000000000	0.0071350808
[10,]	0.05169439	0.24730164	0.010457938	0.0026015746	0.01375931	0.020805846	0.0005146695	0.090985346	0.007135081	1.0000000000

Figura 11 - P-valor do teste T entre os grupos mensais solares

Apesar de haver mais combinações de grupos com p-valor superior a 0.05, o agrupamento mensal foi reduzido a apenas oito grupos, pois nem todos os grupos similares são espacialmente próximos. A Figura 12 apresenta mapa e perfil dos grupos.

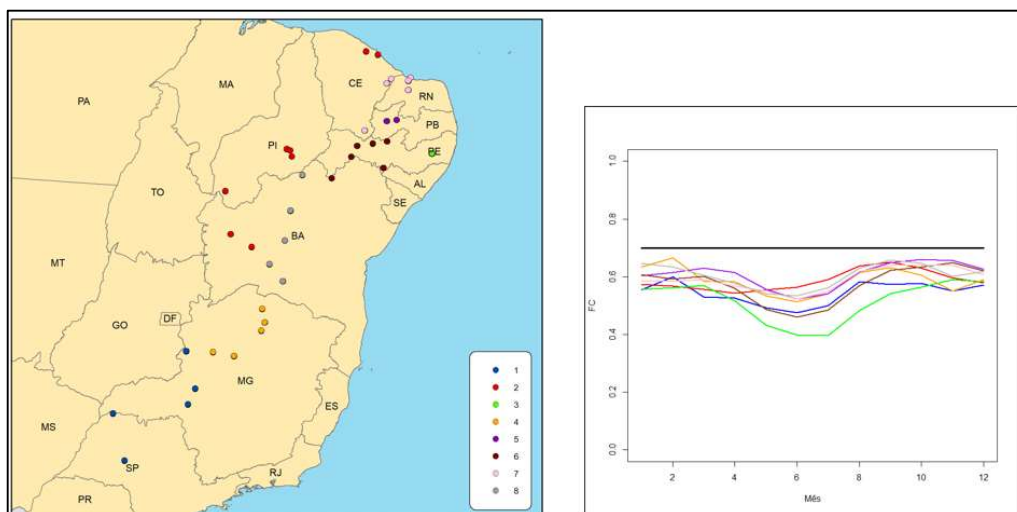


Figura 12 - Mapa e perfil dos oito grupos mensais das fotovoltaicas

Para cada grupo mensal foi executado a ferramenta Grouping Analysis com os dados horários, além da realização de análises de gráficos, como o exemplo apresentado na Figura 13, e testes estatísticos para determinar se existem diferentes perfis horários que tornam necessária a criação de novos grupos.

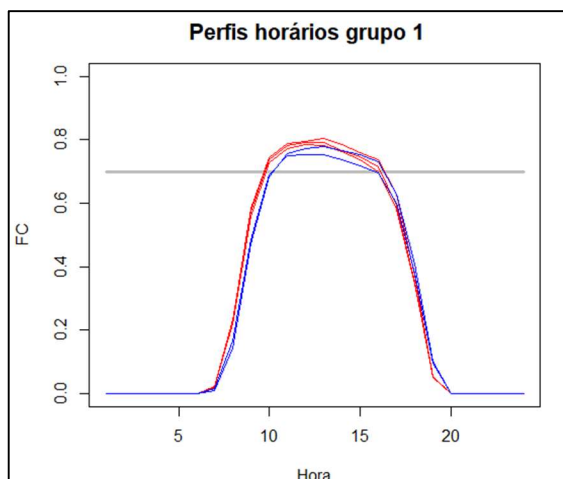


Figura 13 - Comparação entre dados horários de um grupo com mesmo perfil mensal

Foram identificados diferentes perfis horários nos grupos um, quatro, seis e sete resultando num total de doze grupos. A Figura 14 apresenta o mapa final do agrupamento com os polígonos das regiões com mesmo perfil solar.

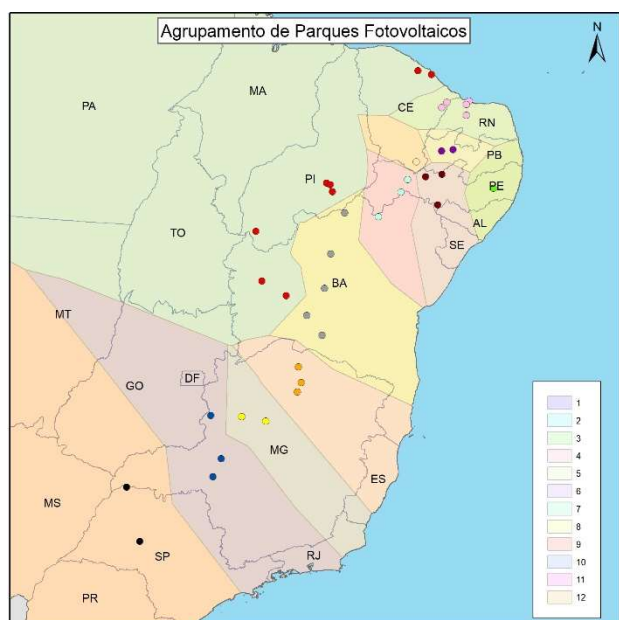


Figura 14 – Mapa com o resultado do agrupamento solar

5.0 – Conclusão

A metodologia Georreferenciada teve como resultado a divisão do recurso eólico em doze regiões no Nordeste e quatro regiões no Rio Grande do Sul e a divisão do recurso solar em doze grupos na área de interesse.

A geração de polígonos das regiões com recurso eólico ou solar semelhantes permite que um planejador do setor elétrico possa definir o agrupamento de usinas que tenha interesse em modelar mesmo que não exista nenhum dado de produção delas permitindo assim que os resultados da metodologia Georreferenciada sejam utilizados sem a necessidade de refazer todos os passos da metodologia. No entanto, havendo mais dados confiáveis, e em diferentes locais, de produção de usinas eólicas ou fotovoltaicas, é recomendável rodar a metodologia Georreferenciada novamente com o intuito de atualizar seus resultados.

A ferramenta desenvolvida na linguagem computacional R possibilita a designação automática de grupo para qualquer usina inserindo apenas as suas coordenadas geográficas e gera um mapa interativo, permitindo que o usuário visualize as áreas dos grupos juntamente com as usinas inseridas por ele.

6.0 - REFERÊNCIAS BIBLIOGRÁFICAS

- (1) EPE – Dados de entrada para modelos elétricos e energéticos: metodologias e premissas (EPE/DEE/NT/011/2021-R1), 2021
- (2) EPE. Sistema AMA. Disponível em: <https://www.epe.gov.br/pt/acesso-restrito/sistema-ama>. Acesso em: 19 de jun. de 2023.
- (3) NREL. System Advisor Model. Disponível em: <https://sam.nrel.gov/>. Acesso em: 19 de jun. de 2023.

7.0 DADOS BIOGRÁFICOS



Engenheiro Cartógrafo e Agrimensor formado pela UFPR em 2012. Analista de Pesquisa Energética da EPE desde 2014 trabalhando com cartografia, georreferenciamento, tratamento de dados geoespaciais e planejamento energético.