

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-graduação em Estatística Aplicada

Felipe Augusto Nascimento de Jesus

**Análise de agrupamentos aplicada à delimitação de zonas sismogênicas: um
caso de estudo para o estado de Minas Gerais**

Belo Horizonte
2023

Felipe Augusto Nascimento de Jesus

Análise de agrupamentos aplicada à delimitação de zonas sismogênicas: um caso de estudo para o estado de Minas Gerais

Monografia apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Especialista em Estatística Aplicada.

Orientadora: Profa. Dra. Sueli Aparecida Mingoti

Belo Horizonte
2023

2023, Felipe Augusto Nascimento de Jesus.
Todos os direitos reservados.

Jesus, Felipe Augusto Nascimento de.

J58a Análise de agrupamentos aplicada à delimitação de zonas
sismogênicos [recurso eletrônico]: um caso de estudo para o
estado de Minas Gerais / Felipe Augusto Nascimento de Jesus
— 2023.

1 recurso online (55 f. il, color).

Orientadora: Sueli Aparecida Mingoti
Monografia (especialização) - Universidade Federal de Minas
Gerais, Instituto de Ciências Exatas, Departamento de Estatística
Referências: 54-55.

1. Estatística. 2. Clustering 3. Análise multivariada. 4.
Geoprocessamento I. Mingoti, Sueli Aparecida. II. Universidade
Federal de Minas Gerais. I. Instituto de Ciências Exatas,
Departamento de Estatística. III. Título.

CDU 519.2 (043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa
CRB 6/1510 - Universidade Federal de Minas Gerais – ICEx



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
P Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

ATA DO 302ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE FELIPE AUGUSTO NASCIMENTO DE JESUS.

Aos quatorze dias do mês de julho de 2023, às 10:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno **Felipe Augusto Nascimento de Jesus**, intitulado: “*Análise de agrupamentos aplicada à delimitação de zonas sísmogênicas: um caso de estudo para o estado de Minas Gerais*”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, a Presidente da Comissão, Professora Sueli Aparecida Mingoti – Orientadora, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 14 de julho de 2023.

Documento assinado digitalmente



SUELI APARECIDA MINGOTI
Data: 18/08/2023 13:05:15-0300
Verifique em <https://validar.it.gov.br>

Prof.^a Sueli Aparecida Mingoti (Orientadora)
Departamento de Estatística / UFMG

Documento assinado digitalmente



ELA MERCEDES MEDRANO DE TOSCANO
Data: 17/08/2023 17:57:14-0300
Verifique em <https://validar.it.gov.br>

Prof.^a Ela Mercedes Medrano de Toscano
Departamento de Estatística / UFMG

Roberto da Costa
Quinino:80871291720

Assinado de forma digital por
Roberto da Costa
Quinino:80871291720
Dados: 2023.08.17 15:39:43 -03'00'

Prof. Roberto da Costa Quinino
Departamento de Estatística / UFMG



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

DECLARAÇÃO DE CUMPRIMENTO DE REQUISITOS PARA CONCLUSÃO DO CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA.

Declaro para os devidos fins que Felipe Augusto Nascimento de Jesus, número de registro 2021681666, cumpriu todos os requisitos necessários para conclusão do curso de Especialização em Estatística e que entregou para sua orientadora, a professora Sueli Aparecida Mingoti, o trabalho, que aprovou a versão final. O trabalho foi apresentado no dia 14 de julho de 2023 com o título “Análise de agrupamentos aplicada à delimitação de zonas sismogênicas: um caso de estudo para o estado de Minas Gerais”.

Belo Horizonte, 21 de agosto de 2023

Roberto da Costa
Quinino:80871291720

Assinado de forma digital por
Roberto da Costa
Quinino:80871291720
Dados: 2023.08.21 14:48:16 -0300

Prof. Roberto da Costa Quinino
Coordenador do curso de
Especialização em Estatística
Departamento de Estatística / UFMG

AGRADECIMENTOS

Os meus sinceros agradecimentos a todas as pessoas, instituições de pesquisa e empresas que, de alguma forma, contribuíram para o desenvolvimento desta pesquisa, e em especial agradeço:

Ao Programa de Pós-Graduação do Instituto de Ciências Exatas (ICEx) da UFMG, pela oportunidade e apoio para realizar essa pesquisa.

À Profa. Dra. Sueli Aparecida Mingoti (UFMG), pela sua orientação e apoio os quais foram fundamentais para a conclusão dessa pesquisa.

Aos amigos da turma de Estatística Aplicada 2021, pelas discussões técnicas e não-técnicas, ajuda na escolha do tema desenvolvido, incentivos e risadas.

Aos amigos Daniel Coelho e Victor Salles pelas discussões técnicas e não-técnicas relacionadas ao desenvolvimento desse trabalho.

Aos meus irmãos, Laisa, Lucas e Ana, que sempre me apoiaram e pelos constantes incentivos que me conduziram à buscar a excelência no desenvolvimento dessa pesquisa e na vida.

Ao meu avô Milton e minha dinda Anita, que sempre me apoiaram, por serem exemplos e pelos incentivos que me conduziram à buscar a excelência no desenvolvimento dessa pesquisa e na vida.

À minha noiva, Bruna da Luz Farias, que compreendeu a importância deste desafio acadêmico e suportou meus nervosismos e ausências, pelo apoio, incentivo e equilíbrio espiritual, fundamentais para a realização dessa pesquisa.

Finalmente, agradeço aos meus pais, Maria das Graças do Nascimento e Christian de Jesus, pelo apoio incondicional e incentivo essencial para a realização dessa pesquisa e por serem meus exemplos e guias em tudo que busco realizar ao longo da minha vida, incluindo essa pesquisa.

"Conhecer-se é dominar-se, dominar-se é triunfar.."

- Jigoro Kano

RESUMO

O objetivo desse trabalho foi sugerir e avaliar a aplicabilidade de uma metodologia estruturada, baseada em técnicas estatísticas de agrupamentos (*clusters*) e geoprocessamento de dados espaciais, para o mapeamento de possíveis zonas sismogênicas em Minas Gerais. Os resultados apresentaram correlação com domínios tectônicos presentes na área de interesse e com zonas sismogênicas obtidas por outros autores, demonstrando que a combinação das técnicas de estatística multivariada (análise de *clusters*), e de geoprocessamento é uma metodologia objetiva para o mapeamento de zonas sismogênicas. Entretanto, os resultados também mostram que essa abordagem ainda carece de revisão embasada na opinião de sismólogos especializados na temática. Sugere-se que, em trabalhos futuros, avalie-se a incorporação de mais parâmetros da região de interesse, tais como as características geológicas e estruturais, de forma que essas novas variáveis possam agregar mais diferenciação ou similaridade para os grupos resultantes, além do uso de mais técnicas de agrupamentos que possam incorporar variáveis categóricas.

Palavras-chave: Técnicas de agrupamento; zonas sismogênicas; estatística multivariada; geoprocessamento.

ABSTRACT

The goal of this work was to suggest and evaluate the applicability of a structured methodology, based on statistical clustering techniques (*clusters*) and geoprocessing of spatial data, for mapping possible seismogenic zones in Minas Gerais. The results correlated with tectonic domains present in the area of interest and with seismogenic zones obtained by other authors, demonstrating that the combination of multivariate statistical techniques (*clusters* analysis) and geoprocessing is an objective methodology for mapping of seismogenic zones. However, the results also show that this approach still needs to be revised based on the opinion of seismologists specialized in the subject. It is suggested that, in future work, the incorporation of more parameters of the region of interest be evaluated, such as geological and structural characteristics, so that these new variables can add more differentiation or similarity to the resulting groups, in addition to the use of more grouping methods that can incorporate categorical variables.

Keywords: Clustering techniques; seismogenic zones; multivariate statistics; geoprocessing.

LISTA DE FIGURAS

Figura 2.1 - Mapa de localização da área de interesse, com coordenadas em WGS84	16
Figura 3.1 - Zonas sismogênicas do Brasil, com destaque para a área de interesse (retângulo vermelho tracejado), adaptado de Schweig (2013).	18
Figura 3.2 - Esquema de armazenamento de distâncias em uma matriz de distância 4x4, extraída de Mingoti (2007, p. 146).	19
Figura 3.3 - Ilustração gráfica do método de ligação simples, adaptada de Mingoti (2007, p. 153).	20
Figura 3.4 - Dendograma (à direita) para 8 pontos de exemplo (à esquerda), extraído de Härdle, Simar et al. (2015, p. 395-396).	21
Figura 3.5 - Exemplo de localização de um centroide a partir de múltiplos pontos, extraído de (PAMULA, 2023).	25
Figura 3.6 - Exemplo de polígonos de Voronoi, extraído de Weisstein (2008).	26
Figura 4.1 - Catálogo Sísmico Brasileiro, retirado de RSBR (2022).	27
Figura 4.2 - Sismos selecionados para o processamento (946 elementos amostrais) após a etapa de filtragem (coordenadas em WGS84).	29
Figura 4.3 - Domínios tectônicos na área de interesse.	30
Figura 5.1 - Histograma da variável latitude do conjunto total de sismos.	33
Figura 5.2 - Histograma da variável longitude do conjunto total de sismos.	34
Figura 5.3 - Histograma da variável magnitude do conjunto total de sismos.	34
Figura 5.4 - Histograma da variável profundidade do conjunto total de sismos.	35
Figura 5.5 - Relação de disponibilidade de informação de profundidade dos sismos. Nessa figura, o ponto (“.”) é utilizado como notação para decimal.	36
Figura 5.6 - Nível de fusão em relação ao número de grupos, com destaque para $g = 8$ (linha vermelha).	37
Figura 5.7 - Coeficiente R^2 e parâmetro Pseudo-F em relação ao número de grupos, com destaque para $g = 8$ (linha vermelha).	38
Figura 5.8 - Diagramas de caixa obtidos para cada grupo ($g = 8$).	40
Figura 5.9 - Comparação do conjunto de sismos agrupados por meio do método de Ward (à esquerda) e pelo método K-Médias (à direita).	42
Figura 5.10 - Comparação entre diagramas de caixa obtidos a partir dos grupos resultantes de cada método para $g = 8$	44

Figura 5.11 - Associação entre sismos agrupados pelo método de Ward e os tipos de domínios tectônicos compreendidos pela área de interesse.	46
Figura 5.12 - Histograma da contagem de sismos, separados em função dos grupos, e associados aos tipos de domínios tectônicos compreendidos pela área de interesse.	47
Figura 5.13 - Histograma da contagem de sismos, separados em função dos grupos, e associados aos domínios tectônicos específicos compreendidos pela área de interesse.	47
Figura 5.14 - Comparação da classificação dos sismos com zonas sismogênicas do Brasil extraídas de Schweig (2013), previamente apresentadas na Figura 3.1.	48
Figura 5.15 - Associação entre os grupos de sismos com zonas sismogênicas do Brasil extraídas de Schweig (2013). A legenda NA significa “Não aplicável” e representa eventos que não foram associados com nenhuma zona sismogênica.	49
Figura 5.16 - Potenciais zonas sismogênicas delimitadas a partir dos grupos obtidos pelo método de Ward.	50
Figura 5.17 - Comparação das potenciais zonas sismogênicas delimitadas nesse estudo (em verde) com as zonas sismogênicas extraídas do trabalho de Schweig (2013).	51

LISTA DE TABELAS

Tabela 3.1 - Algoritmo das técnicas hierárquicas aglomerativas, segundo Härdle, Simar et al. (2015).	20
Tabela 3.2 - Histórico de fusão dos 8 pontos de exemplo.	21
Tabela 4.1 - Variáveis utilizadas no projeto.	28
Tabela 5.1 - Estatísticas descritivas dos parâmetros de interesse.	35
Tabela 5.2 - Estatísticas descritivas dos parâmetros de interesse padronizados, exceto da profundidade.	37
Tabela 5.3 - Destaque dos parâmetros utilizados no critério de seleção do número de grupos g , com ênfase em valores obtidos em $g = 8$, destacado em vermelho, selecionado com número de grupos a ser utilizado nesse trabalho. “D. Ward” corresponde à distância de Ward.	39
Tabela 5.4 - Estatística descritiva dos grupos obtidos para solução de $g = 8$	41
Tabela 5.5 - Concordância e discordância na comparação entre os grupos resultantes de cada método.	43
Tabela 5.6 - Comparação entre as estatísticas descritivas dos grupos obtidos para solução de $g = 8$ para o método de Ward (WD) e K-Médias (KM).	45
Tabela 5.7 - Síntese da associação entre grupos obtidos pelo método de Ward e as zonas sismogênicas extraídas do trabalho de Schweig (2013).	49

SUMÁRIO

1. INTRODUÇÃO	14
2. OBJETIVO	16
3. FUNDAMENTAÇÃO TEÓRICA	17
3.1. Atividade sísmica no Brasil	17
3.2. Análise de agrupamentos (<i>Cluster</i>)	18
3.2.1 Método de Ligação Simples	20
3.2.2 Método de Ward	22
3.2.3 K-Médias	24
3.3. Análise espacial	25
3.3.1 Centroide	25
3.3.2 Diagrama de Voronoi	25
4. METODOLOGIA APLICADA	26
4.1. Base de dados	26
4.2. Análise exploratória descritiva	30
4.3. Análise de agrupamento	31
4.4. Associação aos domínios tectônicos	31
4.5. Análise espacial e delimitação das potenciais zonas sismogênicas	32
5. RESULTADOS E DISCUSSÃO	33
5.1. Análise exploratória	33
5.2. Número de grupos	37
5.3. Grupos resultantes	42
5.4. Associação com com domínios tectônicos	46
5.5. Comparação com estudos anteriores	48
5.6. Delimitação das zonas sismogênicas	50
6. CONSIDERAÇÕES FINAIS	52
REFERÊNCIAS BIBLIOGRÁFICAS	54

1. INTRODUÇÃO

A atividade sísmica no Brasil é reconhecidamente baixa (ASSUMPÇÃO et al., 2016). Ainda assim, segundo Assumpção et al. (2016), instalações críticas, como usinas e reatores nucleares, e barragens hidrelétricas, têm feito uso sistemático de análises sismológicas específicas, e.g., análises de risco sísmico. O risco sísmico representa a relação entre o perigo sísmico (ou ameaça sísmica), o qual depende de características regionais, e a vulnerabilidade de estruturas expostas ao perigo. Lopes e Nunes (2011) ressaltam que a avaliação do perigo sísmico no Brasil vem ganhando mais relevância para grandes obras de engenharia, como usinas hidrelétricas e barragens de rejeito. Embora o nível de atividade sísmica no Brasil seja relativamente baixo comparado aos países vizinhos (ASSUMPÇÃO et al., 2016), é importante que grandes obras de engenharia incorporem-no em seu mapa de risco.

Conforme destacam Lopes e Nunes (2011) e Assumpção et al. (2016), o mapa preliminar e regional da ameaça sísmica no Brasil, presente na NBR 15.421 (ABNT, 2006), é pouco representativo para o risco sísmico, sendo necessário a realização de estudos mais atuais e com maior detalhamento acerca da ocorrência de sismos, ainda incipientes no Brasil, para melhorar o conhecimento dos efeitos sísmicos em diferentes regiões e para a construção de cenários sísmicos mais realistas.

Um dos principais parâmetros para a análise de ameaça sísmica é o conhecimento de zonas sismogênicas ou zonas sísmicas, as quais consistem em delimitações das potenciais áreas-fontes sísmicas de uma certa região. Segundo o Serviço Geológico Norte-americano, o termo zona sismogênica é usado para descrever uma área onde os terremotos tendem a se concentrar (USGS, 2019). Entretanto, em muitos casos, não há um procedimento claro e consistente para delimitação dessas zonas (WEATHERILL; BURTON, 2009), sendo necessário a opinião de sismólogos especialistas para construção de cenários nos projetos de avaliação do perigo sísmico. Tal procedimento, de certa forma, implica em resultados subjetivos. Assumpção et al. (2016) reforça acerca dessa limitação, apontando que o conhecimento da sismicidade intraplaca, para delimitação de zonas sismogênicas, como é o caso do Brasil, ainda é incompleto, e para lidar com essas incertezas associadas à natureza do fenômeno a estratégia adotada é o uso de vários modelos alternativos criados por especialistas independentes, cada um com um peso num esquema de árvore lógica, ou cenários possíveis.

Uma abordagem objetiva mencionada por Assumpção et al. (2016) é o uso de modelos de ocorrência dos sismos ponto a ponto, por meio da densidade de distribuição dos epicentros em raios pré-definidos dependentes da magnitude dos eventos. Já Garcia, Wald e Hearne (2012) apresentaram uma abordagem de determinação automática do regime tectônico e o domínio sismotectônico de um evento sísmico, seu tipo

de fonte mais provável e a classe correspondente da Equação de Previsão do Movimento do Terreno (*Ground-Motion Prediction Equation - GMPE*). Uma das principais etapas do trabalho foi a definição de vários domínios sismotectônicos, que poderiam ser tratados de forma análoga às zonas sismogênicas, utilizando critérios de divisão relacionados a características tectônicas e sismológicas derivadas de catálogos sísmicos globais e regionais, por exemplo: latitude, longitude, profundidade hipocentral, mecanismo focal. Enquanto no trabalho de Rehman, Burton e Weatherill (2014), os autores exploraram técnicas de agrupamento para delimitar zonas sismogênicas no Paquistão, região conhecida pela grande atividade sísmica, por meio do método de K-médias, com ênfase nos parâmetros de localização (latitude, longitude e profundidade) dos eventos sísmicos.

Portanto, conforme discutido, o desenvolvimento de técnicas objetivas para definição de zonas sismogênicas é um tema que fomenta debates e o desenvolvimento de novos trabalhos na comunidade técnico-científica. Nesse sentido, a estatística multivariada, com ênfase nas técnicas de agrupamento, pode contribuir com o entendimento acerca de padrões e similaridades entre eventos sísmicos já registrados no Brasil, por meio dos atributos diretamente e indiretamente relacionados a cada um dos eventos, e.g., localização, magnitude, profundidade, domínio tectônico, proporcionando resultados que poderão fomentar, de forma objetiva, estratégias para delimitação de zonas sismogênicas.

2. OBJETIVO

O objetivo do presente trabalho é avaliar a aplicabilidade de técnicas estatísticas de agrupamento, por meio do método da Distância de Ward (WARD, 1963), em dados históricos da atividade sísmológica no estado de Minas Gerais (e regiões adjacentes), conforme apresenta a Figura 2.1, para a delimitação de possíveis zonas sísmogênicas na área de interesse, apresentando tal procedimento como uma abordagem objetiva para se alcançar tais informações.

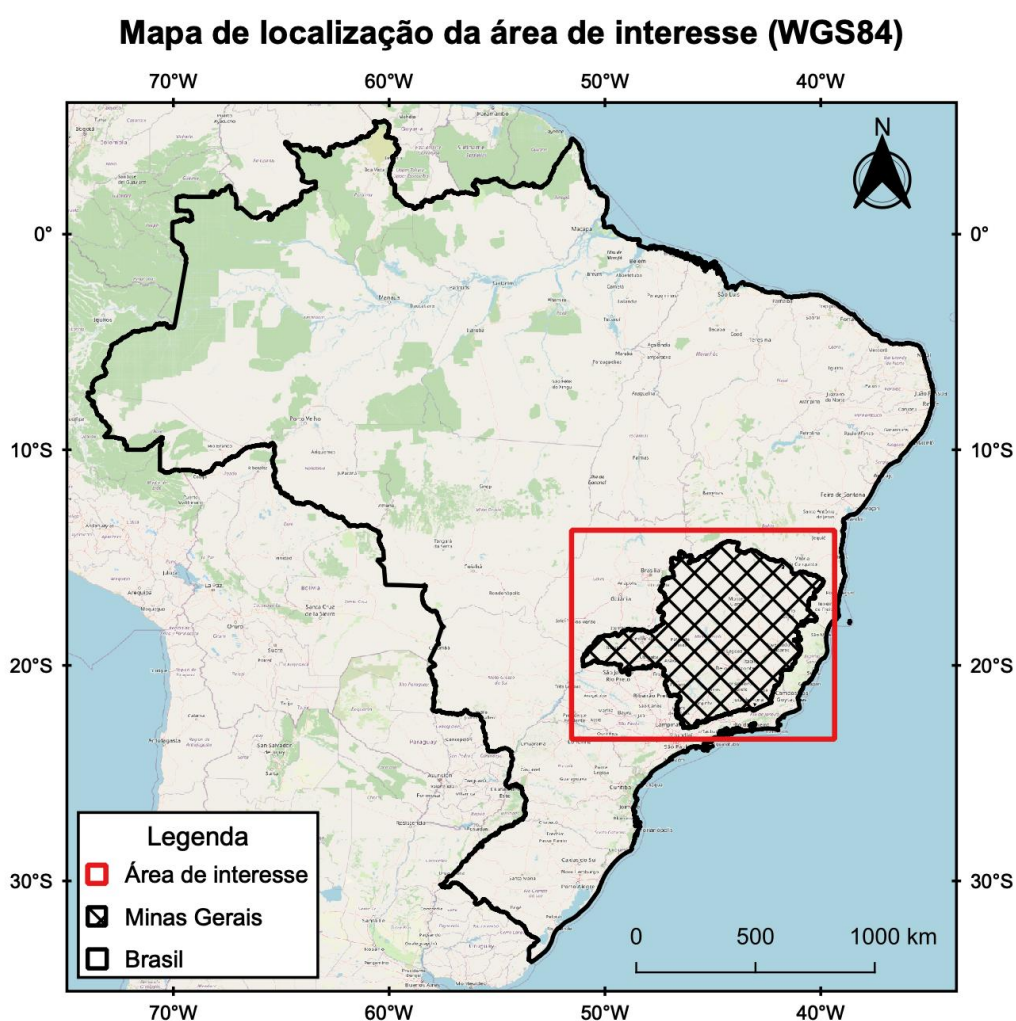


Figura 2.1: Mapa de localização da área de interesse, com coordenadas em WGS84¹.

¹Sistema de referência geodésico global estabelecido pelo U.S. Department of Defense desde 1960 com o objetivo de fornecer o posicionamento e navegação em qualquer parte do mundo, através de informações espaciais

3. FUNDAMENTAÇÃO TEÓRICA

Nessa seção, são apresentados os principais fundamentos teóricos necessários para o desenvolvimento dessa pesquisa.

3.1. Atividade sísmica no Brasil

A maioria dos sismos que ocorrem no território brasileiro são decorrentes da sismicidade intraplaca, uma vez que o Brasil está situado no interior da Placa Sul-Americana, distante de suas bordas oeste e leste, respectivamente representadas pela zona de subducção andina e pela Cadeia Meso-Atlântica. A exceção ao estado do Acre e o extremo oeste do Amazonas, em que a origem dos sismos está relacionada à subducção da Placa de Nazca em relação à Placa Sul-Americana.

Berrocal (2016) descreve dois principais tipos de atividade sísmica: (i) não ligada à tectônica e (ii) ligada à tectônica. No primeiro caso, o autor explica que alguns sismos podem ser causados por efeitos diversos, sem relação com a geotectônica da região, e.g., deslizamentos de terra, detonações de explosivos em minerações ou explosões em testes nucleares. Já no segundo caso, ligado à geotectônica, o autor explica que a grande maioria dos sismos que ocorrem no Brasil está associada a algum tipo de falha geológica que é ativada, ou reativada, através da liberação de energia acumulada por algum processo tectônico, e.g., residual de processos tectônicos de eras geológicas antigas ou pelo atual deslocamento da placa Sul Americana.

Entretanto, diversos pesquisadores buscam estabelecer relações entre informações sismológicas, geológicas e geofísicas para explicar a origem da atividade sísmica. No trabalho de Schweig (2013), a autora trás um capítulo sobre os conceitos de sismicidade e discute sobre as características da sismicidade brasileira. A Figura 3.1, extraída do trabalho de Schweig (2013), apresenta as zonas sismogênicas do Brasil, adaptadas do trabalho de Mito (1993).

Na Figura 3.1, destaca-se a área de interesse desse estudo (linha tracejada em vermelho), na qual é possível observar que ocorrem ao menos dez (10) zonas sismogênicas: M - Porangatu (parcialmente); N - Passos; O - Paraguaçu (parcialmente); P - Jequitinhonha; Q - Paraopeba; S - Jequitáí; T - Ribeirão Preto; W - Pinhal; Y - Cunha; e Z - Santos (parcialmente). Ainda é possível que, por conta da proximidade ao limite da área, exista influência de mais duas zonas: U - Presidente Prudente (porção sudoeste); e X - Cananéia (porção sul). Entretanto, vale destacar que uma grande parcela da área não pode ser diferenciada em termos de zonas sismogênicas, isso ocorre pelo fato de as zonas apresentadas não apresentarem continuidade em toda a região de interesse.

De acordo com Schweig (2013), Mito (1993) faz uma análise sismotectônica no sentido de estabelecer os limites das áreas de instabilidade do território brasileiro, através da integração visual de informações sismológicas, geológicas e geofísicas.

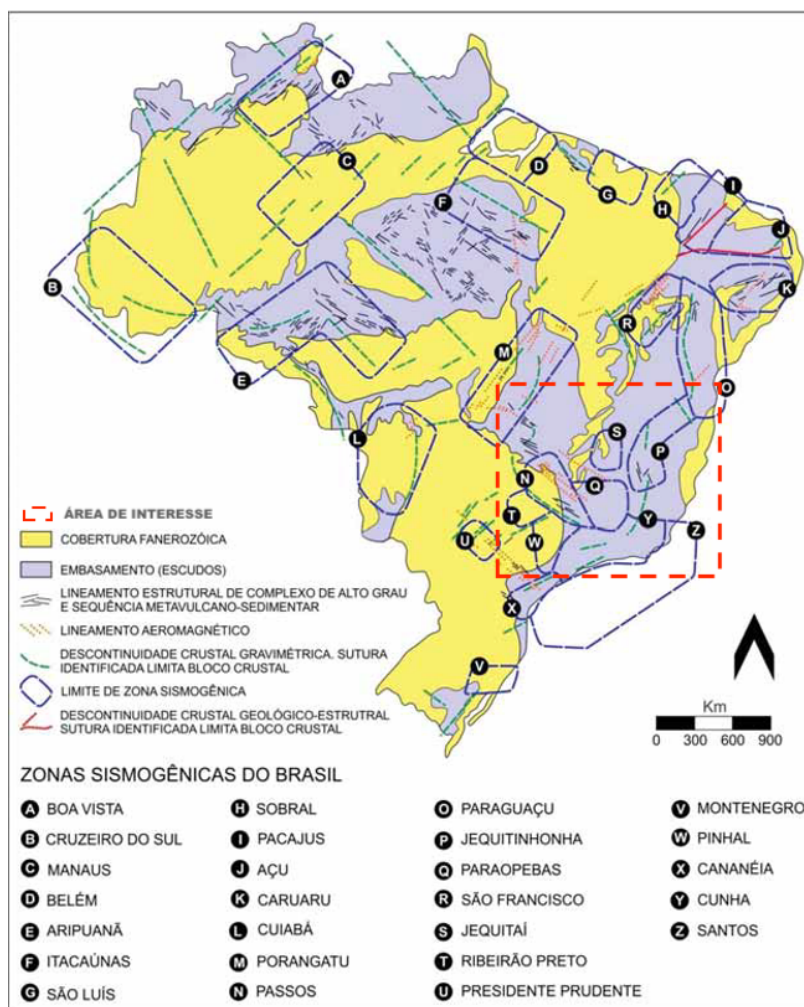


Figura 3.1: Zonas sismogênicas do Brasil, com destaque para a área de interesse (retângulo vermelho tracejado), adaptado de Schweig (2013).

3.2. Análise de agrupamentos (*Cluster*)

Segundo Mingoti (2007), a análise de agrupamentos tem como objetivo dividir os elementos de um amostra, ou população, em grupos, cujos elementos sejam similares entre si com respeito às variáveis que neles foram medidas, e que os elementos pertencentes a grupos diferentes sejam heterogêneos em relação as mesmas características.

A comparação entre os elementos amostrais pode ser feita por meio de métricas, e.g., medidas de distâncias, entre informações de p-variáveis medidas nos respectivos elementos. Essa comparação é traduzida por meio das medidas de similaridade e dissimilaridade.

Dentre as medidas mais comuns, e apropriadas para comparação entre variáveis numéricas, pode-se citar a Distância Euclidiana, definida na Equação 1:

$$d(X_l, X_k) = \sqrt{\sum_{m=1}^p (X_{ml} - X_{mk})^2} \quad (1)$$

Nesse caso, dois elementos amostrais X_l e X_k , com $l \neq j$, são comparados nas p variáveis pertencentes a um vetor de observações, sendo que nos termos X_{ml} e X_{mk} , m denota a variável correspondente, $m = 1, 2, \dots, p$. Essa métrica, assim como outras medidas de distâncias, corresponde a uma medida de dissimilaridade, i.e., quanto menor os seus valores, mais similares serão os elementos comparados. As distâncias entre os diversos elementos amostrais são armazenadas numa matriz de dimensão $n \times n$, chamada de matriz de distâncias, conforme exemplo da Figura 3.2, sendo n o número total de elementos amostrais. No caso do exemplo, tem-se $n = 4$.

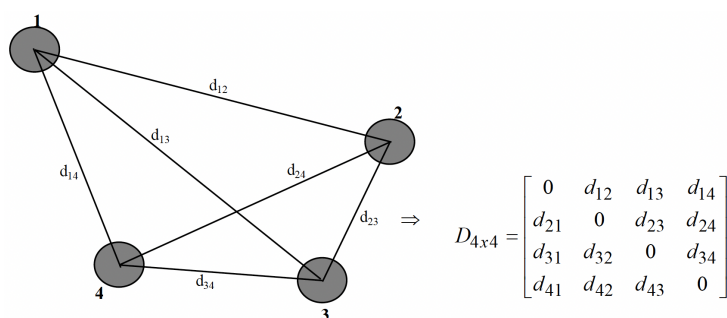


Figura 3.2: Esquema de armazenamento de distâncias em uma matriz de distância 4x4, extraída de Mingoti (2007, p. 146).

Com relação a construção de agrupamentos, também chamados de conglomerados, Mingoti (2007) e Härdle, Simar et al. (2015) destacam que as técnicas de agrupamento são frequentemente classificadas em hierárquicas (subdividida em aglomerativa e divisiva) e não hierárquicas. Nas técnicas hierárquicas, são realizadas análises exploratórias para definição do número de grupos g . Já nas técnicas não hierárquicas, o número de grupos g já é pré-estabelecido.

Os algoritmos hierárquicos aglomerativos começam com a partição mais fina (mais grosseira, no caso dos divisivos) possível e unem os grupos (dividem os grupos, no caso dos divisivos) passo a passo. Esse processo é repetido até que todos os elementos estejam unidos em um único grupo (ou dividido no número máximo de grupos, i.e., um único elemento em cada grupo). A Tabela 3.1 apresenta o algoritmo das técnicas hierárquicas aglomerativas, segundo Härdle, Simar et al. (2015).

Tabela 3.1: Algoritmo das técnicas hierárquicas aglomerativas, segundo Härdle, Simar et al. (2015).

Algoritmo
Construir a partição de maior detalhe, i.e., com todos os elementos em separado
Computar a matriz de distâncias
Encontrar dois grupos com a distância mais próxima
Juntar esses dois grupos em um grupo
Calcular a distância entre os novos grupos e obter matriz de distância reduzida
Repetir até todos os grupos estejam aglomerados em um só

Segundo Mingoti (2007), não existe uma resposta exata para escolha do número final de grupos g . De forma geral, essa escolha tem como propósito encontrar o número g que esteja associado a uma “divisão natural” dos elementos. Em literatura são descritos diversos métodos de agrupamentos hierárquicos, os quais diferem na forma do cálculo da distância (dissimilaridade) entre os agrupamentos.

Entretanto, para o propósito desse estudo, a seguir serão apresentados os métodos de Ligação Simples, a título de exemplificação, e Ward, a título de descrição do instrumento utilizado nas análises.

3.2.1. Método de Ligação Simples

É um método reconhecido na literatura internacional como *Single Linkage*, no qual calcula-se as distâncias entre os elementos dos dois grupos, e a distância entre os dois grupos é definida pelo valor dos elementos mais próximos entre si. A Figura 3.3 ilustra o método de ligação simples.

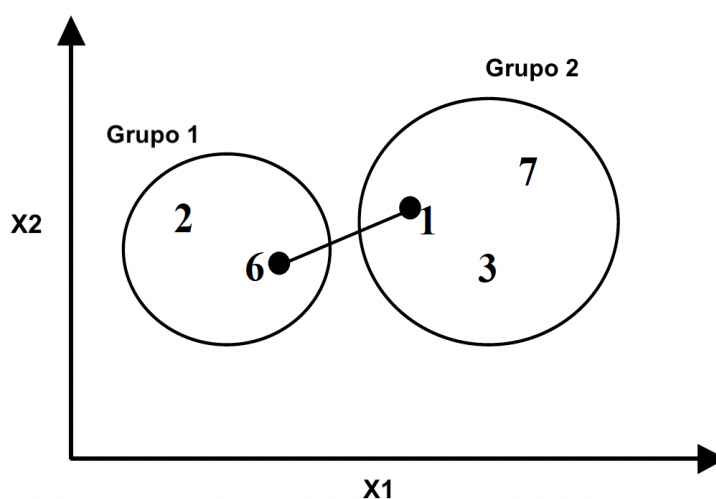


Figura 3.3: Ilustração gráfica do método de ligação simples, adaptada de Mingoti (2007, p. 153).

Segundo Härdle, Simar et al. (2015), esse método também é chamado de algoritmo do vizinho mais próximo. Como consequência de sua construção, a ligação única tende a formar grandes grupos. Grupos que diferem, mas não estão bem separados, podem ser classificados em um grupo, desde que tenham dois pontos próximos. A Figura 3.4 apresenta um exemplo de um dendograma² obtido a partir de 8 pontos, com o histórico de fusão na Tabela 3.2 .

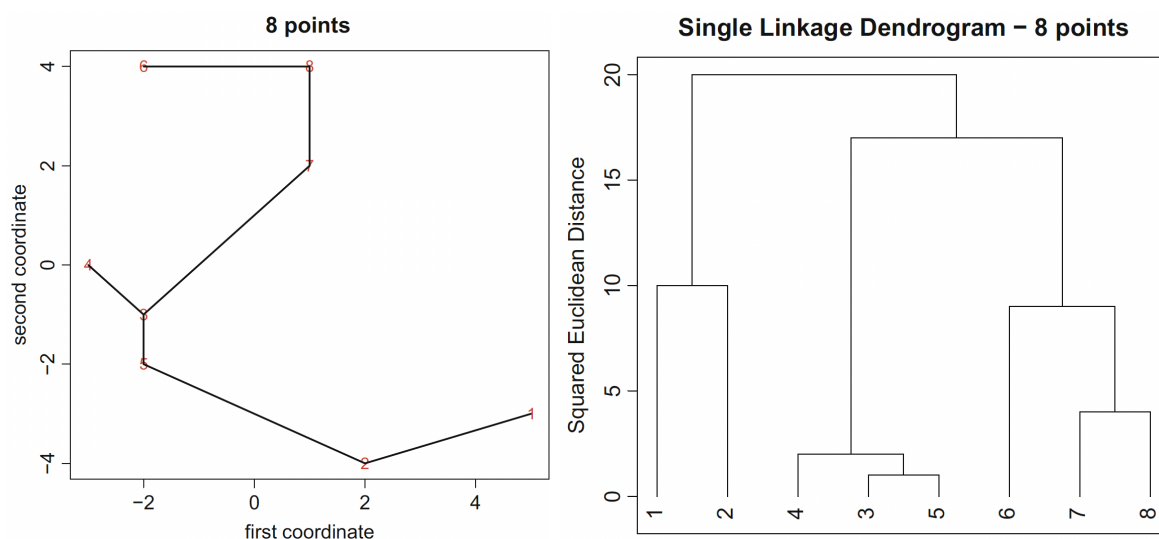


Figura 3.4: Dendograma (à direita) para 8 pontos de exemplo (à esquerda), extraído de Härdle, Simar et al. (2015, p. 395-396).

Tabela 3.2: Histórico de fusão dos 8 pontos de exemplo.

Passo	Grupos (g)	Fusão
1	7	{3} e {5}
2	6	{3, 5} e {4}
3	5	{7} e {8}
4	4	{6} e {7, 8}
5	3	{1} e {2}
6	2	{3, 4, 5} e {6, 7, 8}
7	1	{1, 2} e {3, 4, 5, 6, 7, 8}

²Diagrama de árvore que exhibe os grupos formados por agrupamento de observações em cada passo e em seus níveis de similaridade. O nível de similaridade é medido ao longo do eixo vertical e as diferentes observações são listadas ao longo do eixo horizontal (Minitab, LLC, 2021).

3.2.2. Método de Ward

Como pode ser observado na Figura 3.4, no método de Ligação Simples, quando se passa do passo k para o $k + 1$, ou seja, se passa do número de grupos $n - k$ para $n - k - 1$, a qualidade da partição decresce com o aumento do nível de fusão, implicando na diminuição do nível de similaridade. Nesse caso, a variação entre os grupos diminui e dentro dos respectivos grupos aumenta, o que implica no cenário indesejado para a análise de agrupamentos. Em 1963, Ward propôs um método de agrupamento fundamentado nas variações entre grupos e dentro dos respectivos grupos a cada passo do algoritmo.

O Método de Ward (WARD, 1963), também chamado de mínima variância, tem como princípios:

1. No primeiro passo, cada elemento é um agrupamento; e
2. A cada passo do algoritmo, calcula-se a soma de quadrados da distância Euclidiana (SS_i) entre cada elemento e o vetor de médias do conglomerado, e posteriormente a distância entre os demais grupos.

A soma de quadrados da distância Euclidiana (SS_i) entre cada elemento e o vetor de médias de um mesmo grupo é calculada a partir da Equação 2.

$$SS_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})'(X_{ij} - \bar{X}_{i.}) \quad (2)$$

onde n_i é o número de elementos no conglomerado C_i quando se está no passo k , X_{ij} é o vetor de observações do j -ésimo elemento amostral que pertence ao i -ésimo agrupamento, $\bar{X}_{i.}$ corresponde ao centroide do agrupamento C_i . No passo k , a soma de quadrados total dentro dos grupos é calculada como:

$$SSR = \sum_{i=1}^{g_k} SS_i \quad (3)$$

onde g_k é o número de grupos quando se está no passo k . A distância entre agrupamentos, chamada de soma de quadrados entre agrupamentos, é calculada por meio da Equação 4:

$$d(C_l, C_i) = \left[\frac{n_l n_i}{n_l + n_i} \right] (\bar{X}_l - \bar{X}_i)'(\bar{X}_l - \bar{X}_i) \quad (4)$$

onde C_l e C_i são dois grupos, n_i é o tamanho do C_i e n_l é o tamanho do C_l . Nesse, a cada passo do algoritmo, os agrupamentos que minimizam a distância da Equação 4 são combinados, resultando também em menores valores da SSR . Entretanto, o fator $\frac{n_l n_i}{n_l + n_i}$ atua como um ponderador em função do tamanho dos conglomerados.

Por fim, escolhe-se o número de agrupamentos (g) com base em critérios exploratórios. A seguir, são descritos alguns dos critérios que podem ser utilizados para a determinação do valor do número de grupos, e que foram utilizados nessa monografia.

- Análise do nível de fusão (distância): avalia-se a curva de decrescimento do nível de similaridade entre os agrupamentos que estão sendo combinados a cada passo do algoritmo. Durante a análise da curva, de forma geral, são observados “pontos de salto”, que se destacam no gráfico, os quais indicam o momento ideal de interrupção do algoritmo (MINGOTI, 2007), em termos de número de grupos g , bem como a composição final dos grupos, que pode ser representada por meio de dendogramas, mas para grandes valores de n torna-se difícil ver os resultados nesse tipo de gráfico;
- Análise da soma de quadrados entre grupos (coeficiente R^2):

$$R^2 = SSB/SST_c \quad (5)$$

onde SST_c corresponde à Soma de Quadrados Total, corrigida para a média global de cada variável

$$SST_c = \sum_{i=1}^{g_k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})'(X_{ij} - \bar{X}) \quad (6)$$

e SSB corresponde à Soma de Quadrados Total entre os g^k grupos da partição (Equação 7)

$$SSB = \sum_{i=1}^{g_k} n_i (X_{i.} - \bar{X})'(X_{i.} - \bar{X}) \quad (7)$$

Nesse caso, avalia-se a curva de crescimento do coeficiente R^2 em função do número de grupos formados a cada passo do algoritmo. Assim como na análise de fusão, nesse gráfico também podem ser procurados “pontos de salto”; e

- Estatística Pseudo-F, calculada a partir da equação:

$$F = \left(\frac{n - g^k}{g^k - 1} \right) \left(\frac{R^2}{1 - R^2} \right) \quad (8)$$

onde g^k é o número de grupos relacionados com a partição do respectivo passo k do algoritmo de agrupamento. Segundo Harabasz e Karoński (1974), se o gráfico de Pseudo-F em relação ao número de grupos não for monotonicamente crescente, i.e., apresentar um valor de máximo, esse representa a partição ideal do conjunto de observações, equivalente à maior heterogeneidade entre os grupos particionados.

Entretanto, conforme destacado no trabalho de Junior e Cezário (2018), como muitos algoritmos de agrupamento requerem que o número de grupos seja pré-fixado (métodos não hierárquicos), ou realizam agrupamentos entre n -grupos até um único grupo (métodos hierárquicos), existe a necessidade de validação dos resultados obtidos a partir dessas análises. Nesse a abordagem de validação adotada nesse trabalho se baseia na comparação de resultados obtidos a partir de dois métodos distintos: Ward (hierárquico) e K-Médias (não hierárquico).

3.2.3. K-Médias

Adicionalmente, são apresentados conceitos do método de agrupamento não hierárquico K-Médias (HARTIGAN; WONG, 1979). Esse método foi escolhido como critério complementar de avaliação dos resultados obtidos nessa monografia a partir da análise de agrupamentos feita pelo método de Ward.

Diferente dos métodos hierárquicos, apresentados anteriormente, nessa abordagem é necessário especificar o número g de grupos (que no método K-Médias é denotado por k). A partir desse número, os algoritmos particionam os elementos amostrais buscando semelhança interna (dentro do mesmo grupo) e separação dos demais grupos.

O método K-Médias é composto por quatro etapas:

1. Escolha dos g centroides, também chamados de sementes, para iniciar a partição;
2. Comparação entre cada elemento amostral e os centroides iniciais, por meio de distância Euclidiana (em geral), e posterior alocação no agrupamento de maior similaridade, ou seja, de menor distância;
3. Cálculo dos novos centroides considerando o agrupamento resultante do passo anterior; e
4. Repetição dos passos 2 e 3 até que não seja mais necessária a realocação dos elementos amostrais. Essa interrupção ocorre quando não há melhoria acentuada entre o passo 2 e o 3, ou seja, não há aumento expressivo de similaridade.

3.3. Análise espacial

3.3.1. Centroide

Segundo a definição apresentada por Weisstein (2002), um centroide, também chamado de centro geométrico ou centro de gravidade, corresponde ao centro de massa de um objeto de densidade uniforme, e.g., um plano bidimensional ou um sólido tridimensional. De forma ilustrativa, pode ser interpretado como o ponto no qual uma lâmina se equilibraria ao ser colocada sobre uma agulha.

Para pontos múltiplos, polilinhas ou polígonos divididos em várias partes, ele é calculado usando o centro médio ponderado de todas as partes da feição de interesse. O peso de uma feição pontual é 1, de uma feição de linha é o seu comprimento e de feições de polígono corresponde a sua área. A Figura 3.5 ilustra a localização de um centroide **G** entre 4 pontos distintos.

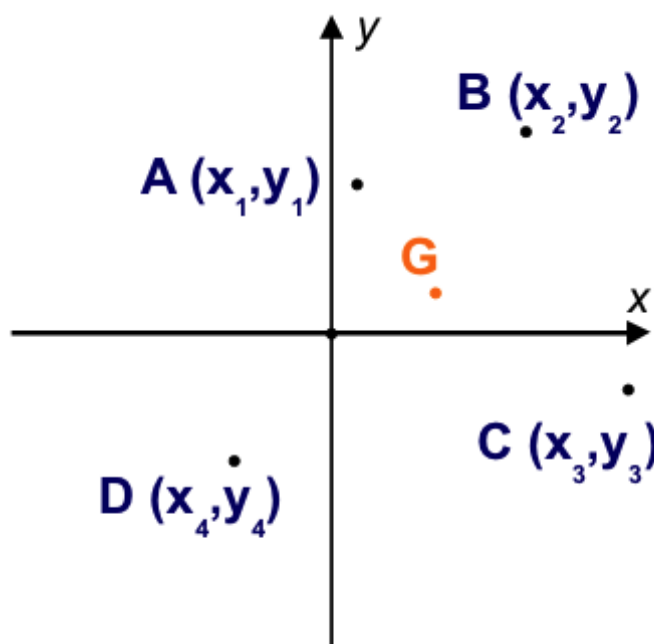


Figura 3.5: Exemplo de localização de um centroide a partir de múltiplos pontos, extraído de (PAMULA, 2023).

3.3.2. Diagrama de Voronoi

Os Diagramas de Voronoi representam partições (polígonos) de um plano com n pontos em polígonos convexos de modo que cada polígono contenha exatamente um ponto gerador e cada ponto em um determinado polígono esteja mais próximo de seu ponto gerador do que de qualquer outro, conforme exemplo da Figura 3.6.

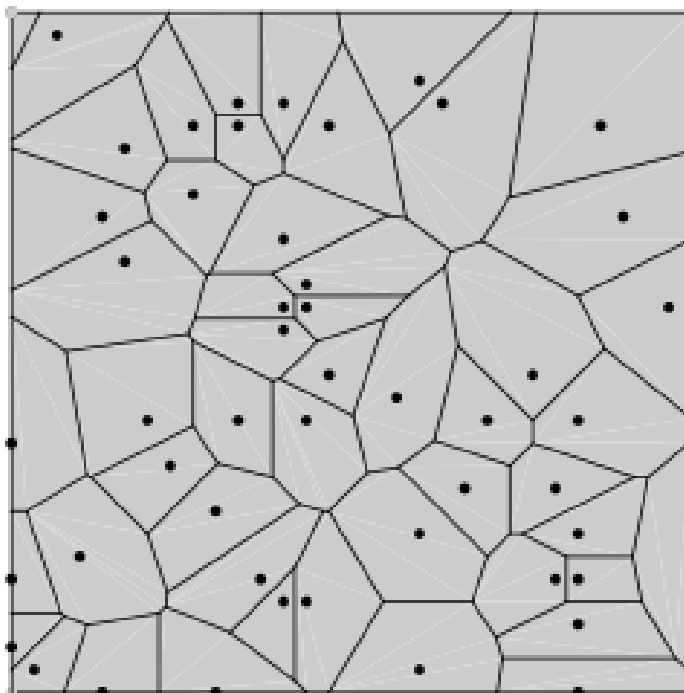


Figura 3.6: Exemplo de polígonos de Voronoi, extraído de Weisstein (2008).

Representam uma ferramenta com diversas aplicações em áreas como computação gráfica, epidemiologia, geofísica e meteorologia, além de serem utilizados também no contexto de análises de agrupamentos (AURENHAMMER, 1991). Mais detalhes acerca dos Diagramas de Voronoi podem ser encontrados no trabalho de Fortune (1995) e sobre as diversas formas de aplicação no trabalho de Aurenhammer (1991).

4. METODOLOGIA APLICADA

Nessa seção são apresentadas as principais etapas metodológicas e ferramentas utilizadas no desenvolvimento desta pesquisa. As etapas são apresentadas nos tópicos a seguir.

4.1. Base de dados

Para o desenvolvimento deste trabalho, foram utilizados dados do Catálogo Sísmico Brasileiro, do Centro de Sismologia da USP, publicado em 27 de janeiro de 2022. Esse catálogo contém os sismos brasileiros registrados entre 1560 e 2020, composto por: catálogo original de Berrocal J. et al. (1984), diversas atualizações feitas nas últimas décadas em colaboração com vários grupos de Sismologia do Brasil, correção de sismos históricos, correções de erros em catálogos internacionais, revisão de sismos históricos e atualização de epicentros e magnitudes de sismos estudados com redes locais.

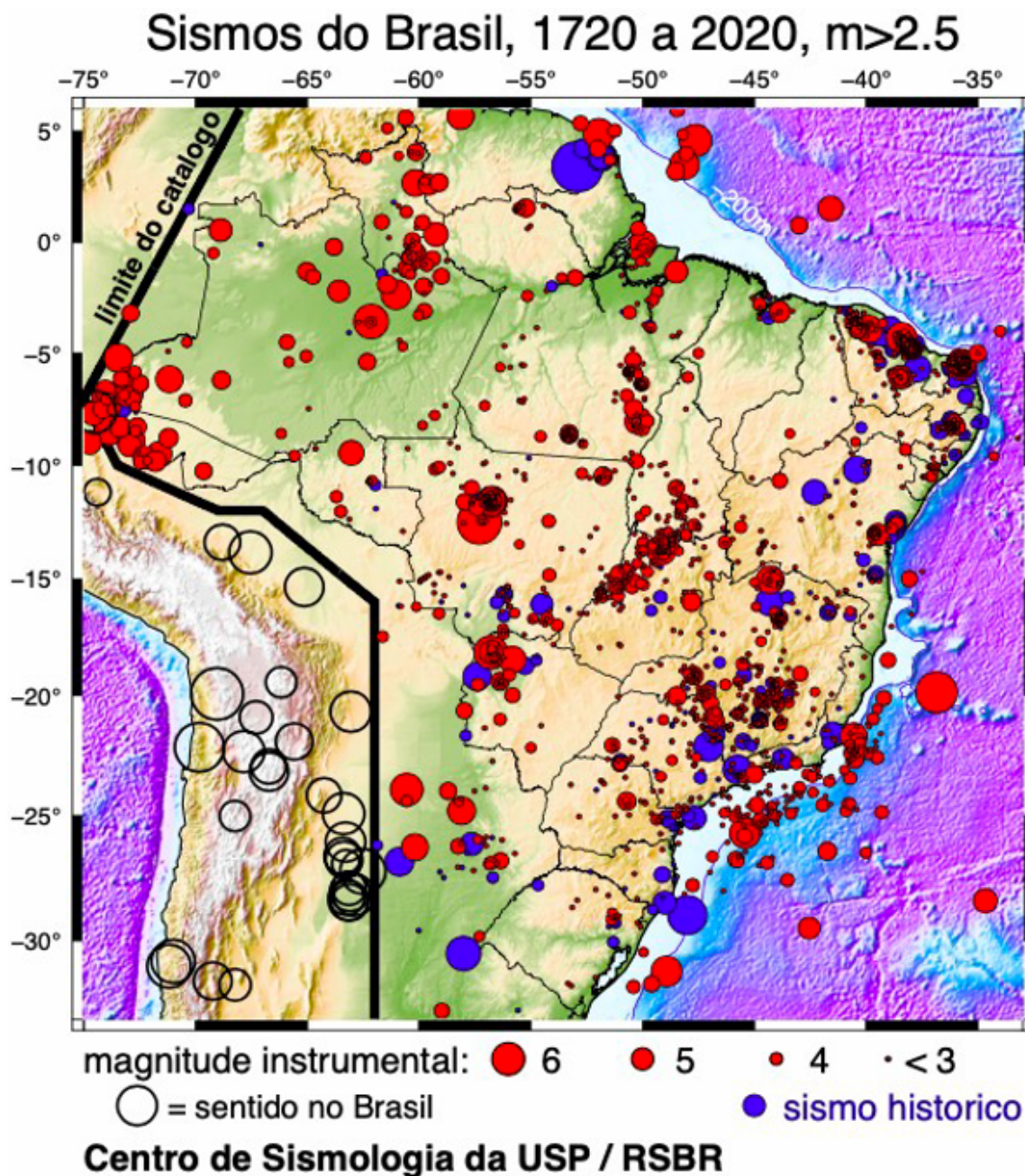


Figura 4.1: Catálogo Sísmico Brasileiro, retirado de RSBR (2022).

No Catálogo Sísmico adotado, que contempla 4574 eventos, cada registro de evento sísmico possui as seguintes informações: localização do sismo (latitude, longitude, estado da união e cidades mais próximas), data e hora de sua ocorrência, incluindo o fuso horário utilizado para referência, profundidade, magnitude, área afetada e intensidade máxima no epicentro. Contudo, para a análise proposta nesse trabalho, foram excluídas algumas variáveis na etapa de preparação dos dados, reduzindo a quantidade de dados utilizados. A base de dados analisada contemplou as seguintes variáveis descritas na Tabela 4.1.

Tabela 4.1: Variáveis utilizadas no projeto.

Campo	Significado
latit	Latitude do epicentro ($^{\circ}$)
longit	Longitude do epicentro ($^{\circ}$)
depth	Profundidade do hipocentro (km)
mag	Magnitude na escala Richter regional (mR)

Destaca-se que a escala Richter é a mais conhecida para determinar qual a intensidade de um terremoto, que foi desenvolvida por Charles F. Richter em 1935, no Instituto de Tecnologia da Califórnia, a partir do estudo de cerca de 200 terremotos ao ano. O termo regional faz referência a abordagem proposta por (ASSUMPÇÃO, 1983), que traz uma metodologia de cálculo mais adequada para sismos no Brasil.

Antes das etapas de processamento, os dados passaram por três etapas de filtragem:

1. Optou-se pelo uso do catálogo “Clean”, no qual sismos com erros de localização acima de 100 km, eventos duvidosos, falsos e repetidos foram removidos pela própria equipe que trabalha para a Rede Sismográfica Brasileira (RSBR);
2. Sismos que tiveram sua magnitude calculada em magnitudes diferentes da Richter regional (mR) foram removidos; e
3. Sismos fora da área de interesse foram removidos.

Cabe ressaltar que os eventos sísmicos filtrados para a área de interesse foram cuidadosamente processados e avaliados pela equipe de especialistas que opera a Rede Sismográfica Brasileira (RSBR) durante a construção do catálogo sísmico. Entretanto, Minas Gerais é uma região de intensa atividade minerária, na qual, em muitos casos, são utilizados desmontes com explosivos no processo de extração mineral. Tais desmontes podem gerar de sinais sísmicos com certos aspectos similares ao de eventos sísmicos de baixa magnitude e causar a contaminação de catálogos sismológicos, conforme descrito no trabalho de Wiemer e Baer (2000). Ainda que existam algoritmos capazes de eliminar essa contaminação do catálogo, não é possível descartar a possibilidade de que algumas observações correspondam a esses eventos de natureza antrópica e que, de alguma forma, podem influenciar em análises dessa base de dados.

A Figura 4.2 apresenta os 946 sismos, categorizados em função da magnitude, utilizados para a análise de agrupamentos, que restaram após as três etapas de filtragem mencionadas.

Sismos restantes após etapa de filtragem (WGS84)

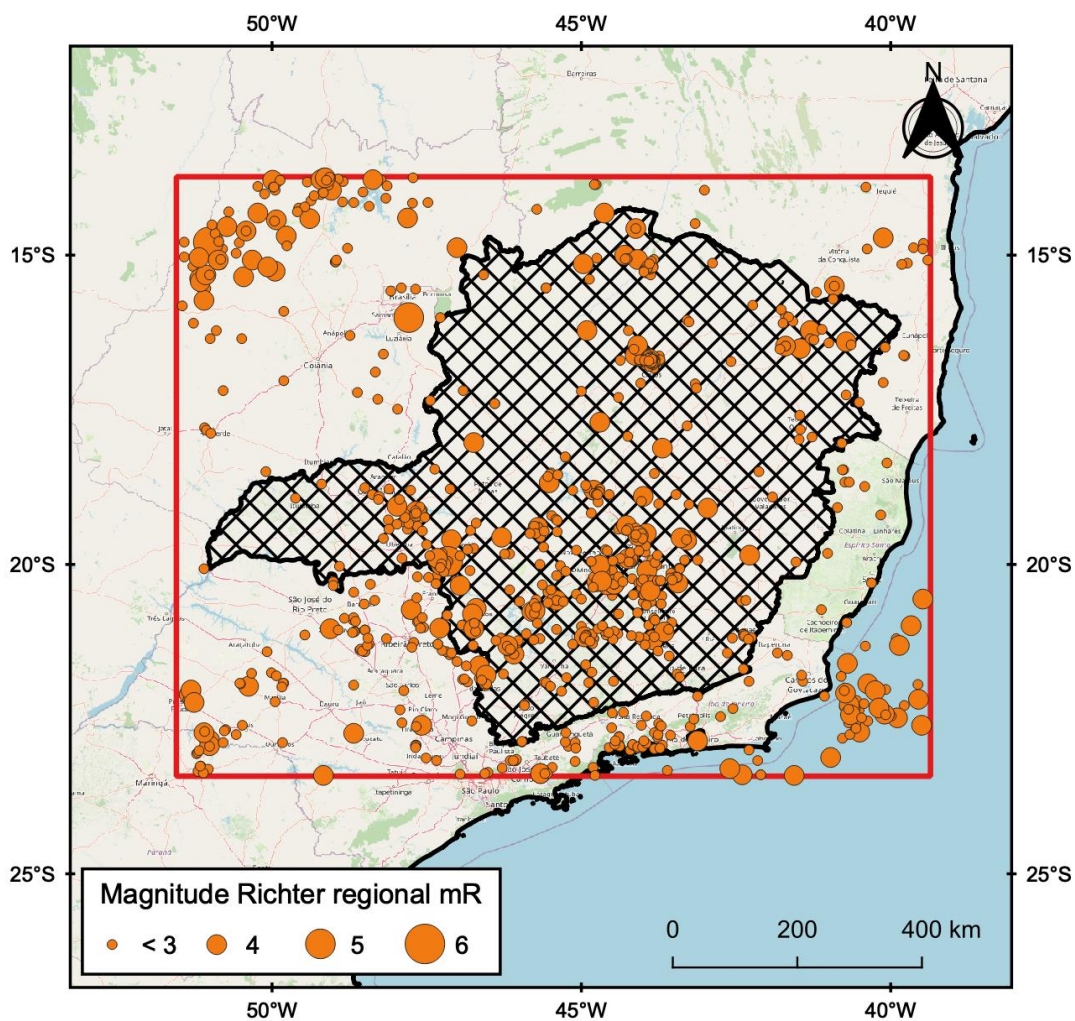


Figura 4.2: Sismos selecionados para o processamento (946 elementos amostrais) após a etapa de filtragem (coordenadas em WGS84).

Além dos dados sismológicos, e assim como no trabalho de (SCHWEIG, 2013), foram utilizados dados sobre os limites dos domínios tectônicos do Brasil, com ênfase na área de interesse desse estudo. A Figura 4.3 apresenta os domínios tectônicos presentes na área de interesse.

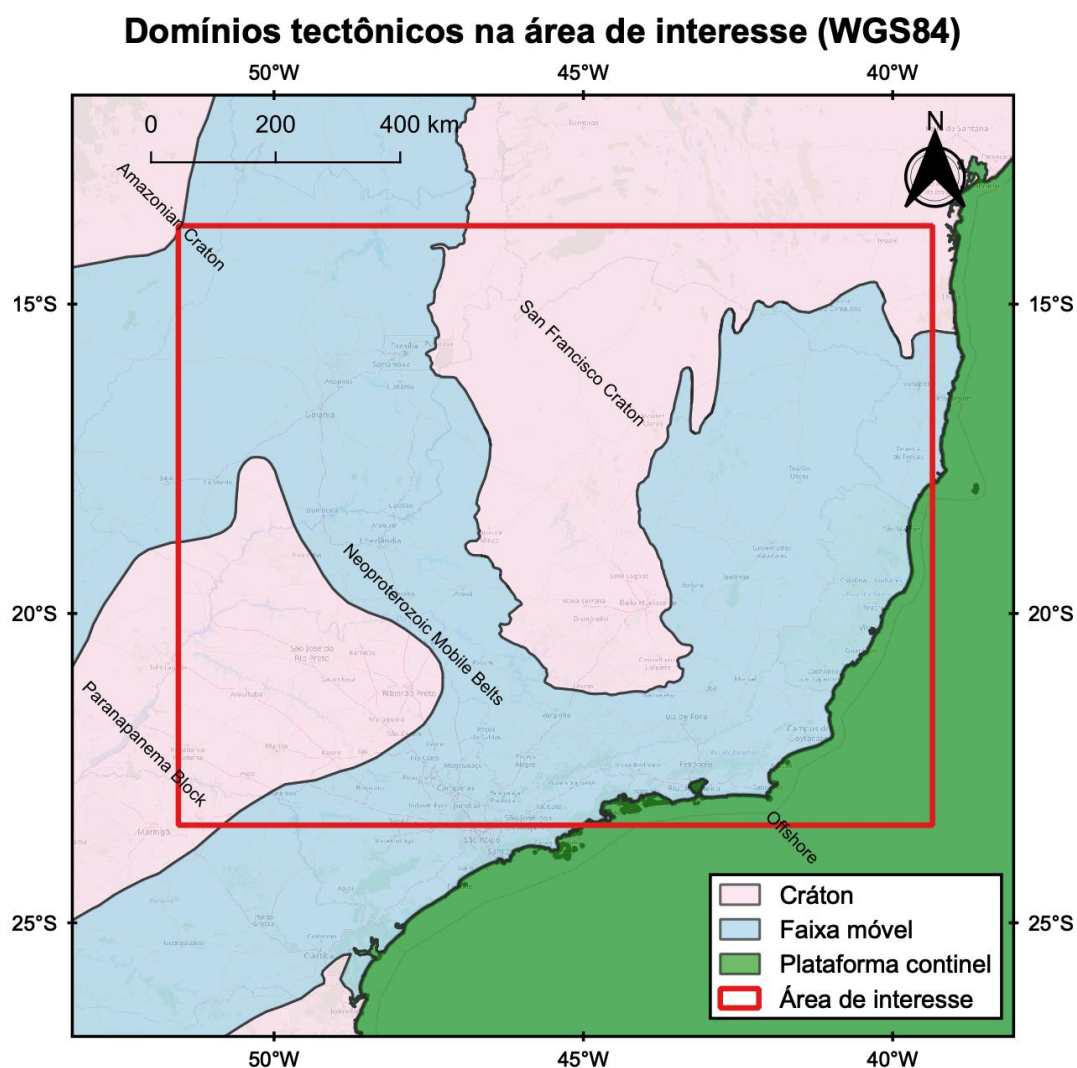


Figura 4.3: Domínios tectônicos na área de interesse.

Os domínios são representados por zonas estáveis e instáveis. A distinção entre as zonas estáveis (crátons) e as zonas instáveis (faixas móveis) é a taxa de atividade tectônica num determinado intervalo de tempo (ou ciclo). É notável que ocorra mais atividade tectônica nas zonas instáveis quando comparadas às zonas estáveis, e.g., orogênese³.

4.2. Análise exploratória descritiva

Para análise exploratória descritiva dos dados optou-se por utilizar os gráficos de histogramas e parâmetros básicos da estatística descritiva clássica. A seguir, apresenta-se a lista de etapas e as ferramentas utilizadas na geração de gráficos e tabelas explorados nas análises:

³Conjunto de processos que levam à formação ou rejuvenescimento de montanhas ou cadeias de montanhas produzido principalmente pelo diastrofismo, ou seja, pela deformação compressiva da litosfera continental

- Foram utilizadas como ferramentas gráficas nativas do programa R, versão 4.2.1 (R Core Team, 2022), dentro da plataforma RStudio (RStudio Team, 2020) para geração de histogramas das variáveis de interesse (latitude, longitude, magnitude e profundidade); e
- Na mesma plataforma do RStudio foram calculadas estatísticas amostrais (e.g., média, mediana, quartis etc.) das variáveis de interesse e das suas respectivas padronizações⁴;

4.3. Análise de agrupamento

Para análise de agrupamento dos dados foi utilizado o método de Ward. A seguir, apresenta-se a lista de procedimentos e as ferramentas utilizadas nessa etapa:

- Foram utilizadas rotinas na plataforma RStudio baseadas nas bibliotecas *Nb-Clust* (CHARRAD et al., 2014) e *cluster* (MAECHLER et al., 2022) para as análises dos níveis de difusão, do coeficiente R^2 e do parâmetro Pseudo-F para escolha do número de grupos g ;
- Os gráficos de análises exploratórias dessa etapa foram gerados a partir de ferramentas gráficas nativas do RStudio;
- Após a escolha do número grupos g , foram gerados grupos a partir do método de Ward e do K-Médias, para comparação e validação dos resultados; e
- A comparação dos grupos resultantes foi feita por meio de análise visual em gráficos de pontos categorizados em função dos grupos.

Ressalta-se que a implementação do método de K-Médias foi feita com base nos resultados obtidos pelo método de Ward. As sementes iniciais do algoritmo foram definidas a partir dos vetor de médias de cada grupo resultante da solução obtida pelo método de Ward.

4.4. Associação aos domínios tectônicos

Para a verificação de associação entre os grupos resultantes e os diferentes domínios tectônicos presentes na área de interesse, foi utilizada uma técnica de análise espacial baseada na união de atributos de entidades espaciais sobrepostas, chamada de “*Spatial Join*” (união espacial). A seguir, apresenta-se a lista de etapas e as ferramentas utilizadas nessa etapa:

- Foi utilizado o algoritmo de união espacial do QGIS para unir os atributos de “Tipo” e “Nome” da base de dados de domínios tectônicos da área de interesse (Figura 4.3) aos eventos sísmicos agrupados; e
- Após a união foram utilizadas rotinas no programa Spyder baseadas nas bibliotecas *Pandas* (CHARRAD et al., 2014) e *Matplotlib* (MAECHLER et al., 2022)

⁴Processo de colocar diferentes variáveis na mesma escala subtraindo a média e dividindo pelo desvio padrão do conjunto.

para a geração de gráficos de barras; e

- A verificação de correlação entre eventos agrupados e domínios tectônicos foi feita por meio de análise visual dos gráficos gerados na etapa anterior.

4.5. Análise espacial e delimitação das potenciais zonas sismogênicas

A análise para delimitação das zonas sismogênicas, por meio de uma metodologia objetiva, teve como base o trabalho de Weatherill e Burton (2009). Nesse trabalho, os autores utilizaram polígonos de Voronoi para traçar zonas sismogênicas na região do Egeu a partir dos centroides obtidos a partir de uma análise de agrupamento de eventos sísmicos. A seguir, apresenta-se a lista de etapas e as ferramentas utilizadas nessa etapa:

- Foram utilizadas rotinas no programa QGIS para calcular a posição dos centroides dos grupos obtidos;
- A partir dos centroides, foi utilizado o algoritmo "Diagramas de Voronoi" do programa QGIS para traçar os polígonos de Voronoi e definir as partições da área de estudo nas potenciais zonas sismogênicas; e
- Os grupos resultantes, centroides e zonas sismogênicas calculadas foram comparados com os limites de zonas sismogênicas extraídos do trabalho de Schweig (2013) na região de interesse.

5. RESULTADOS E DISCUSSÃO

Nessa seção são apresentados os resultados obtidos nessa pesquisa em conjunto com comentários e análises críticas dos produtos gerados ao longo de cada etapa.

5.1. Análise exploratória

A Figura 5.1 apresenta o histograma das observações de latitude do epicentro, na qual observa-se assimetria positiva com um pico em -22 e outro (secundário) em -16.

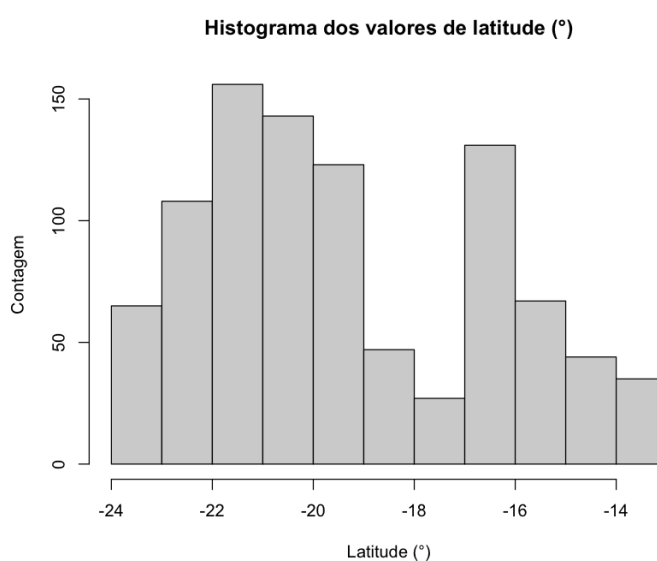


Figura 5.1: Histograma da variável latitude do conjunto total de sismos.

A Figura 5.2 apresenta o histograma das observações de longitude do epicentro, na qual observa-se assimetria negativa com um pico em -44.

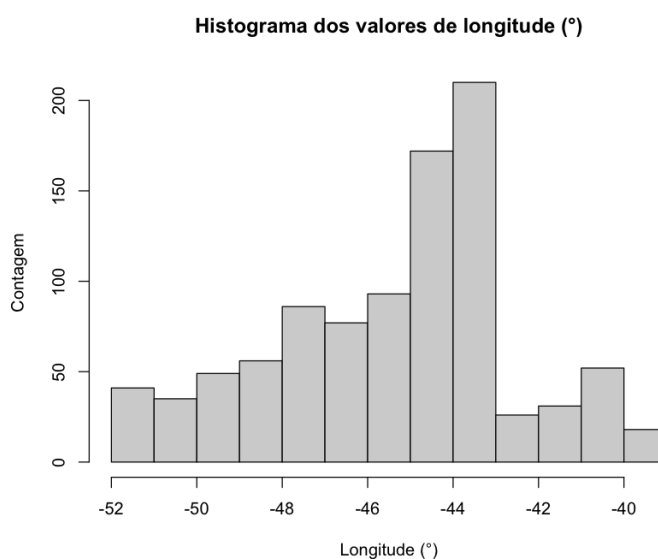


Figura 5.2: Histograma da variável longitude do conjunto total de sismos.

A Figura 5.3 apresenta o histograma das observações de magnitude na escala Richter regional (mR) do epicentro, na qual observa-se maior simetria, quando comparado aos conjuntos de coordenadas dos epicentros, com valores centrados entre 2 e 3.

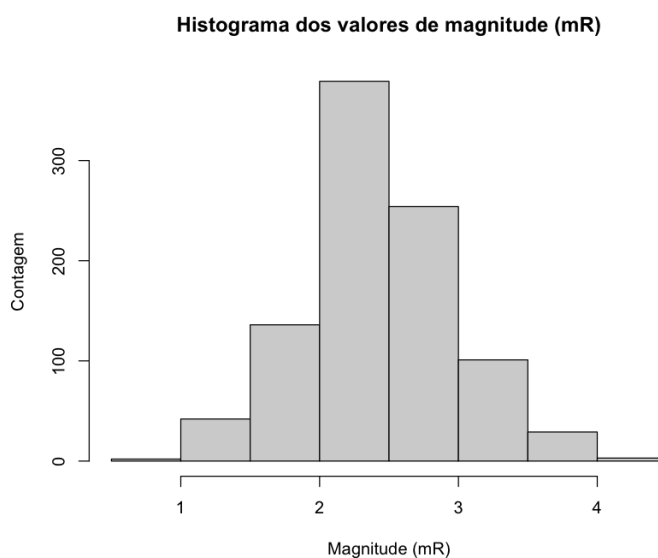


Figura 5.3: Histograma da variável magnitude do conjunto total de sismos.

Entretanto, ressalta-se que o comportamento simétrico, com aparência de uma distribuição “Normal”, não reflete a natureza física dos eventos sísmicos. A distribuição teórica de eventos sísmicos, com a relação de frequência e magnitude, é descrita pela Lei de Gutenberg e Richter (GUTENBERG; RICHTER, 1944):

$$\log(N) = a - bM \quad (9)$$

na qual N é o número total de eventos sísmicos e M é o valor da magnitude. Nesse caso, o formato de distribuição observado é possivelmente uma consequência da limitação de amostragem da instrumentação para registrar eventos de magnitudes menor que 2 (eventos de baixa energia).

A Figura 5.4 apresenta o histograma das observações de profundidade do hipocentro, na qual observa-se assimetria positiva com um pico dominante em valores da ordem de 0.

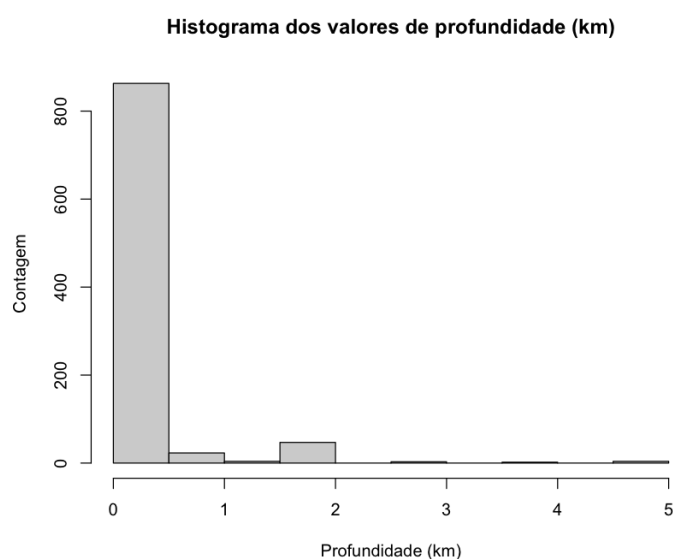


Figura 5.4: Histograma da variável profundidade do conjunto total de sismos.

A Tabela 5.1 apresenta algumas estatísticas descritivas obtidas a partir da base de dados filtrada de sismos.

Tabela 5.1: Estatísticas descritivas dos parâmetros de interesse.

Estatística	Latitude (°)	Longitude (°)	Profundidade (km)	Magnitude (mR)
Média	-19.34	-45.39	0.17	2.48
Desvio Padrão	2.79	2.81	0.61	0.54
Mínimo	-23.41	-51.45	0.00	0.80
25% (Q1)	-21.36	-47.37	0.00	2.10
Mediana (Q2)	-19.98	-44.74	0.00	2.40
75% (Q3)	-16.70	-43.87	0.00	2.80
Máximo	-13.74	-39.40	5.00	4.20

Nessa tabela, o ponto (".") é utilizado como notação para decimal.

Na Tabela 5.1 é possível observar que existe uma diferença expressiva entre na ordem de grandeza das coordenadas dos eventos sísmicos em comparação ao parâmetro de magnitude. Adicionalmente, é possível observar a dominância do valor de 0.00 para as profundidades dos eventos sísmicos.

A Figura 5.5 apresenta um gráfico de setores com a relação de disponibilidade de informação de profundidade dos sismos.

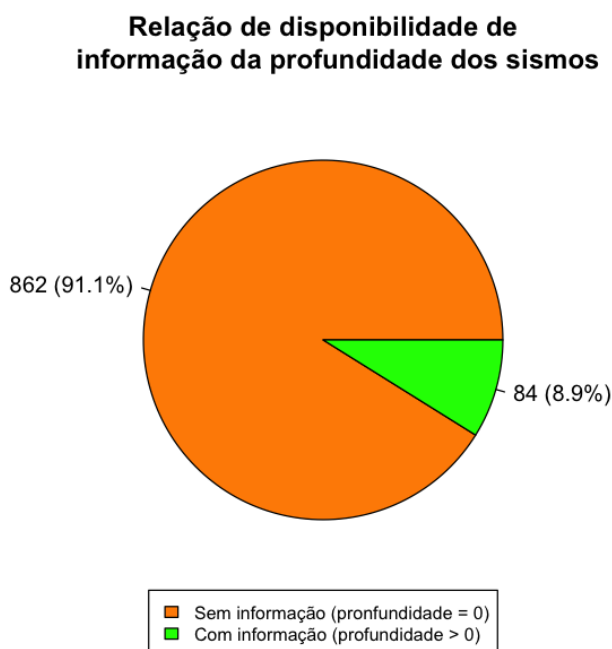


Figura 5.5: Relação de disponibilidade de informação de profundidade dos sismos. Nessa figura, o ponto (“.”) é utilizado como notação para decimal.

Na Figura 5.5, nota-se que mais de 90% dos eventos apresentam profundidade do hipocentro de 0.00 km. Entretanto, segundo a RSBR (2022), o valor de 0 (zero) para profundidade de hipocentro significa ausência de informação. Nesse caso, opta-se pela desconsideração do parâmetro de profundidade do hipocentro para as análises de agrupamentos dessa base de dados.

Ainda com relação à discrepância entre as grandezas das observações de coordenadas e de magnitude dos sismos, a Tabela 5.2 apresenta algumas estatísticas descritivas acerca das padronizações dessas variáveis.

Tabela 5.2: Estatísticas descritivas dos parâmetros de interesse padronizados, exceto da profundidade.

Estatística	Latitude (°)	Longitude (°)	Magnitude (mR)
Média	0.00	0.00	0.00
Desvio Padrão	1.00	1.00	1.00
Mínimo	-1.46	-2.16	-3.10
25% (Q1)	-0.73	-0.70	-0.70
Mediana (Q2)	-0.23	0.24	-0.14
75% (Q3)	0.95	0.54	0.59
Máximo	2.01	2.13	3.17

Nessa tabela, o ponto (“.”) é utilizado como notação para decimal.

Na Tabela 5.2, é possível observar que a diferença entre as variáveis passou para a mesma ordem de grandeza, o que implica em maior ponderação para as variáveis no processo de agrupamento. Nesse caso, além da retirada do parâmetro de profundidade, opta-se pelo uso das variáveis padronizadas nas análises de agrupamentos dessa base de dados.

5.2. Número de grupos

A Figura 5.6 apresenta os gráficos de comportamento do nível de fusão, em relação à distância de Ward dos eventos e à raiz quadrada da distância Ward. Ressalta-se que os resultados foram obtidos a partir do método Ward.D2, implementado no software R e que corresponde ao método de Mínima Variância de Ward, considerando a distância Euclidiana ao quadrado.

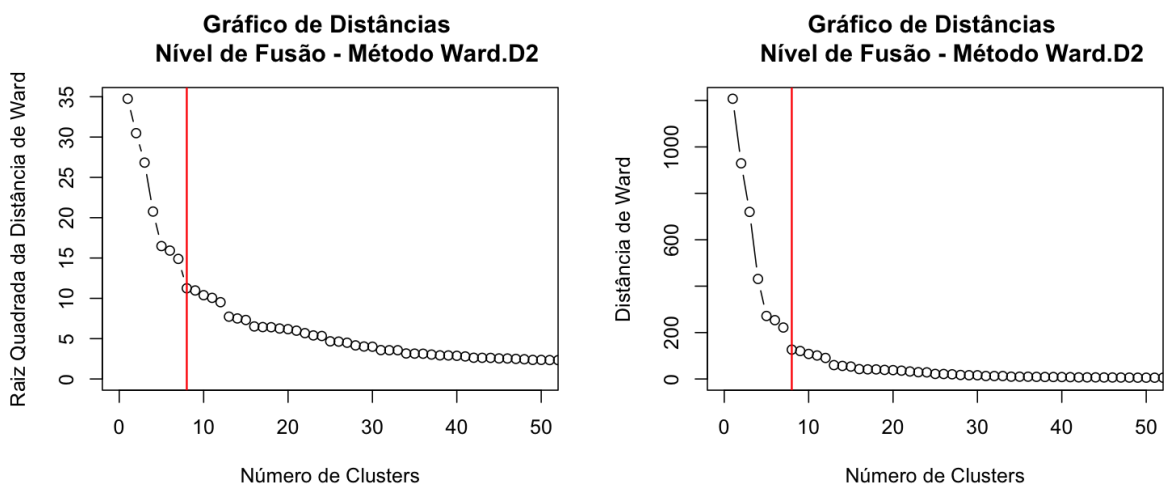


Figura 5.6: Nível de fusão em relação ao número de grupos, com destaque para $g = 8$ (linha vermelha).

Ao considerar o critério de identificação de pontos de salto acentuado no nível de fusão dos grupos, na Figura 5.6, nota-se que existem quebras expressivas entre 5 e 13, com destaque para um salto que ocorre em $g = 8$ (linha vermelha).

A Figura 5.7 apresenta os gráficos do coeficiente R^2 e da estatística Pseudo-F, ambos em função do número de grupos obtidos ao longo dos passos de agrupamento.

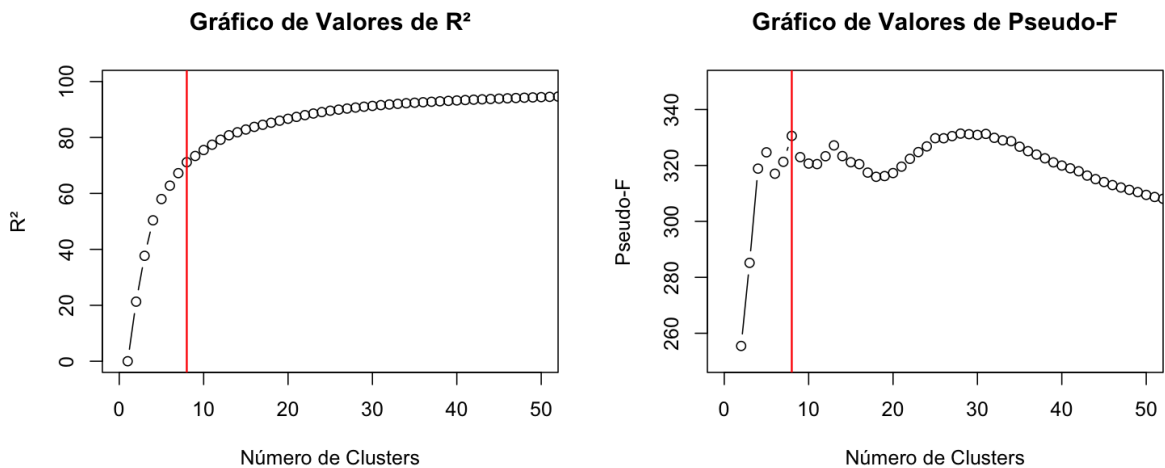


Figura 5.7: Coeficiente R^2 e parâmetro Pseudo-F em relação ao número de grupos, com destaque para $g = 8$ (linha vermelha).

Com relação ao coeficiente de determinação R^2 , nota-se que a partir de 8 grupos (destacado na linha vermelha) o crescimento passou a ser lento, com valores superiores a 70%, portanto, representando o início de uma região de comportamento estável. Já a estatística Pseudo-F apresenta um pico em $g = 8$, com valor superado apenas para número de grupos superior a 20.

Tabela 5.3: Destaque dos parâmetros utilizados no critério de seleção do número de grupos g , com ênfase em valores obtidos em $g = 8$, destacado em vermelho, selecionado com número de grupos a ser utilizado nesse trabalho. “D. Ward” corresponde à distância de Ward.

g	N. Fusão (D. Ward)	N Fusão (D. Ward ^{1/2})	Coef. R^2 (%)	Pseudo-F
4	431.04	20.76	50.38	318.87
5	271.55	16.48	57.99	324.69
6	253.42	15.92	62.78	317.05
7	221.72	14.89	67.25	321.29
8	126.47	11.25	71.16	330.56
9	120.43	10.97	73.39	322.97
10	107.96	10.39	75.51	320.67
11	101.21	10.06	77.41	320.48
12	90.66	9.52	79.20	323.30

Nessa tabela, o ponto (“.”) é utilizado como notação para decimal.

A Figura 5.8 apresenta os diagramas de caixa obtidos para cada grupo considerando a solução de $g = 8$. Já Tabela 5.4 apresenta, em números, a estatística descritiva dos grupos obtidos para a mesma solução.

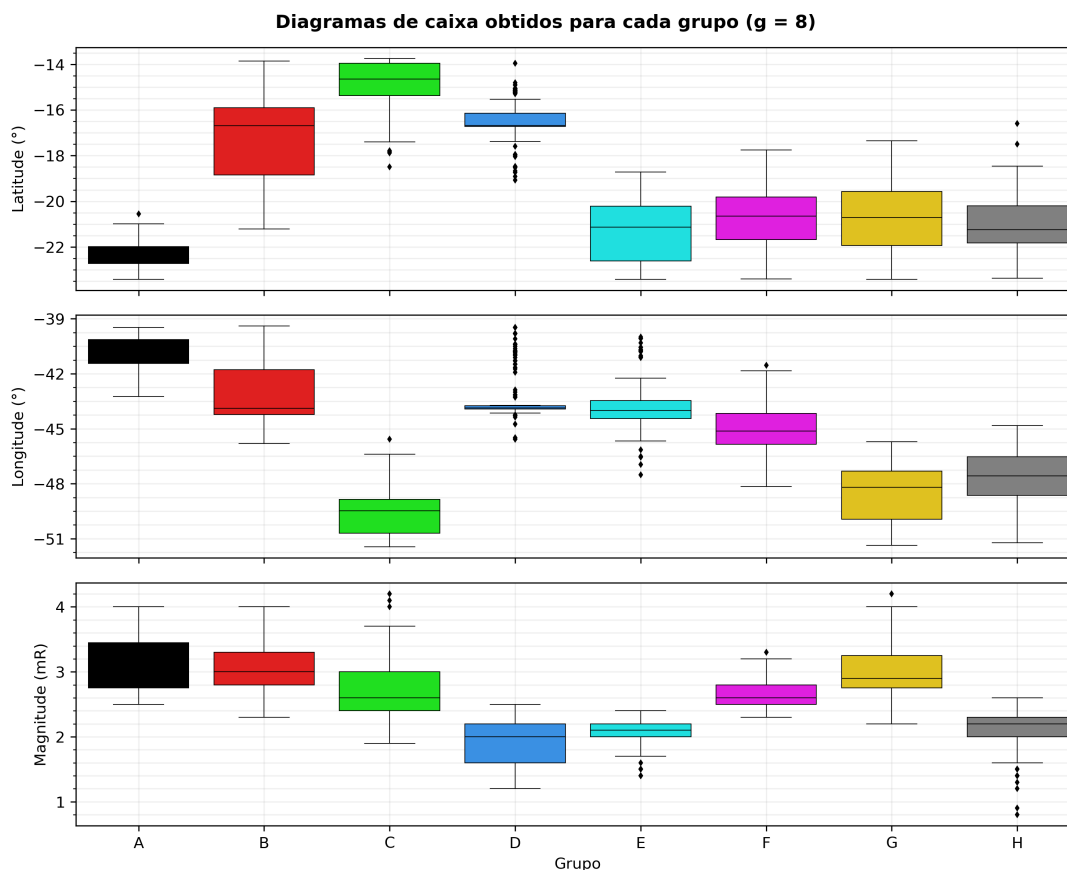


Figura 5.8: Diagramas de caixa obtidos para cada grupo ($g = 8$).

Na Figura 5.8, é possível notar a diferenciação dos três parâmetros em relação aos grupos obtidos para a solução de $g = 8$. Entretanto, destaca-se a existência de observações discrepantes (*outliers*) em termos dos latitude e longitude nos grupos C e D, e em termos de magnitude no grupo C, E e H.

No caso dos parâmetros de latitude e longitude, as discrepâncias podem correr em função da localização mais afastada de alguns eventos sísmicos que foram englobados pelos grupos C e D.

Já com relação à magnitude, a maioria dos *outliers* ocorre para magnitudes baixas (inferiores a 2), possivelmente em função da baixa quantidade de eventos dessa magnitude no catálogo, devido às limitações de instrumentação para detectar eventos de baixíssima energia como esses. Para os casos de *outliers* em magnitudes superiores a 4, possivelmente está relacionado à baixa frequência de ocorrência de eventos dessa magnitude no Brasil e, por consequência, a baixa quantidade de observações no catálogo.

Tabela 5.4: Estatística descritiva dos grupos obtidos para solução de $g = 8$.

Latitude (°)								
Grupo	n	Média	D. Pad.	Mín.	P. 25%	P. 50%	P. 75%	Máx.
A	43	-22.35	0.67	-23.41	-22.73	-22.42	-21.99	-20.56
B	102	-17.18	1.94	-21.22	-18.85	-16.70	-15.91	-13.86
C	109	-14.90	1.14	-18.50	-15.38	-14.65	-13.95	-13.74
D	127	-16.47	0.90	-19.08	-16.72	-16.68	-16.15	-13.95
E	156	-21.28	1.31	-23.41	-22.62	-21.13	-20.21	-18.72
F	216	-20.88	1.44	-23.40	-21.68	-20.65	-19.81	-17.76
G	75	-20.68	1.45	-23.41	-21.94	-20.71	-19.57	-17.35
H	118	-21.05	1.36	-23.37	-21.82	-21.25	-20.20	-16.60
Longitude (°)								
Grupo	n	Média	D. Pad.	Mín.	P. 25%	P. 50%	P. 75%	Máx.
A	43	-40.81	0.99	-43.24	-41.43	-40.57	-40.15	-39.48
B	102	-43.24	1.59	-45.80	-44.22	-43.89	-41.78	-39.40
C	109	-49.52	1.27	-51.45	-50.70	-49.47	-48.85	-45.57
D	127	-43.33	1.32	-45.57	-43.92	-43.87	-43.74	-39.47
E	156	-43.84	1.31	-47.51	-44.46	-44.01	-43.45	-39.99
F	216	-45.15	1.25	-48.15	-45.85	-45.12	-44.16	-41.53
G	75	-48.58	1.55	-51.36	-49.94	-48.20	-47.32	-45.70
H	118	-47.82	1.73	-51.22	-48.63	-47.57	-46.53	-44.83
Magnitude (mR)								
Grupo	n	Média	D. Pad.	Mín.	P. 25%	P. 50%	P. 75%	Máx.
A	43	3.09	0.43	2.50	2.75	3.00	3.45	4.00
B	102	3.05	0.38	2.30	2.80	3.00	3.30	4.00
C	109	2.73	0.51	1.90	2.40	2.60	3.00	4.20
D	127	1.90	0.37	1.20	1.60	2.00	2.20	2.50
E	156	2.06	0.20	1.40	2.00	2.10	2.20	2.40
F	216	2.64	0.21	2.30	2.50	2.60	2.80	3.30
G	75	3.01	0.41	2.20	2.75	2.90	3.25	4.20
H	118	2.09	0.32	0.80	2.00	2.20	2.30	2.60

Nessa tabela, o ponto (“.”) é utilizado como notação para decimal.

5.3. Grupos resultantes

A Figura 5.9 apresenta o resultado dos conglomerados de sismos adotando 8 grupos ($g = 8$) pelo método de Ward (à esquerda) e pelo método de K-Médias (à direita).

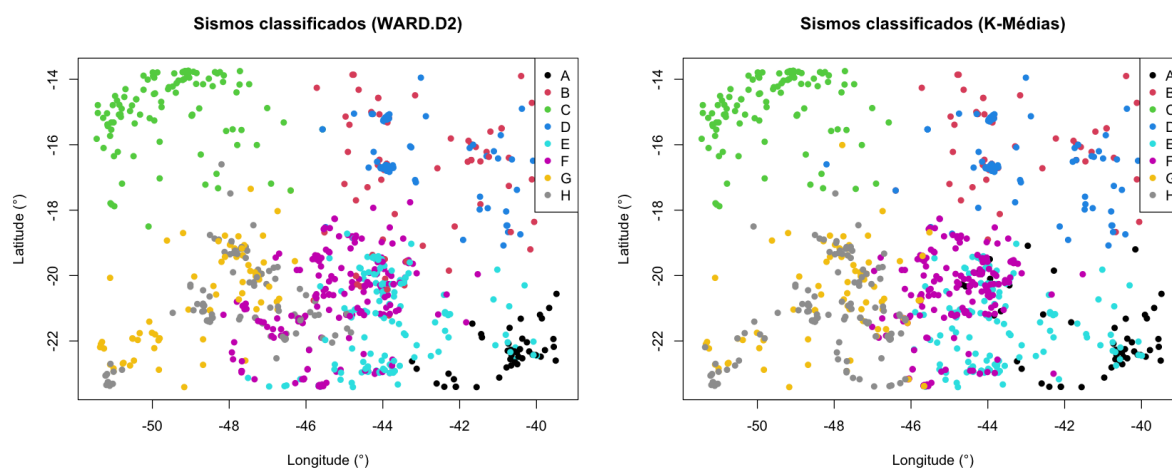


Figura 5.9: Comparação do conjunto de sismos agrupados por meio do método de Ward (à esquerda) e pelo método K-Médias (à direita).

Na Figura 5.9, em ambos os resultados, nota-se que os grupos apresentam uma distribuição espacial aproximadamente homogênea entre si, i.e., eventos classificados dentro de um mesmo grupo tendem a ficar espacialmente distribuídos próximos entre si. Em alguns casos, é possível notar alguns eventos com localização discrepante aos aglomerados formados pela maioria, com maior destaque para os grupos D e E, assim como observado no diagrama de caixa da Figura 5.8. Em outros casos, visualmente, nota-se alguns grupos encontram-se “misturados” em algumas porções, por exemplo, o os grupos B (vermelho) e D (azul), que estão concentrados na região nordeste da figura. Adicionalmente, nota-se similaridade nas concentrações de eventos dos mesmos grupos obtidos pelas diferentes metodologias.

A seguir, são apresentadas informações de comparação entre os eventos agrupados pelos dois métodos. Também são apresentados os valores de R^2 e para o Pseudo-F, obtidos a partir da solução de K-Médias.

- Número de eventos classificados nos mesmos grupos por ambas as metodologias: 781 (83%);
- Número de eventos classificados em grupos diferentes por ambas as metodologias: 165 (17%);
- Coeficiente R^2 obtido pela solução do método K-Médias: 73,50%
- Pseudo-F obtido pela solução do método K-Médias: 371,65

A Tabela 5.5 detalha os percentuais de concordância e discordância entre os métodos obtidos para cada grupo.

Tabela 5.5: Concordância e discordância na comparação entre os grupos resultantes de cada método.

Grupo	Concordante (n)	Concordante (%)	Discordante (n)	Discordante (%)
A	36	84	7	16
B	65	64	37	36
C	105	96	4	4
D	118	93	9	7
E	132	85	24	15
F	165	76	51	24
G	64	85	11	15
H	96	81	22	19

Nota-se que a maioria dos grupos teve uma concordância superior a 80%, com maior destaque para os grupos C e D, com valores superiores a 90%. Entretanto, nota-se que o grupo B teve a menor discordância de todos com 64% dos eventos coincidentes.

Ao comparar os parâmetros R^2 e Pseudo-F entre as duas soluções (Ward: $R^2 = 71,16\%$ e Pseudo-F = 330,56), verifica-se que a solução obtida pelo método K-Médias apresenta um pouco mais de similaridade dentro dos grupos e maior heterogeneidade entre os grupos, ainda que a diferença entre os parâmetros obtidos em ambas as soluções não seja grande.

Adicionalmente, a Figura 5.10 apresenta uma comparação entre diagramas de caixas obtidos a partir dos grupos resultantes de cada método para $g = 8$. Já Tabela 5.6 apresenta, em números, a comparação entre as estatísticas descritivas dos grupos obtidos para as mesmas soluções.

Na Figura 5.10 e na Tabela 5.6, nota-se que os grupos apresentam distribuições parecidas entre si. Observa-se que a maior diferença ocorre em termos de contagem, com destaque para os grupos A e B. Também é possível notar que as diferenças são mais acentuadas nos extremos, i.e., máximos e mínimos, enquanto as médias e os percentis apresentam valores próximos entre si.

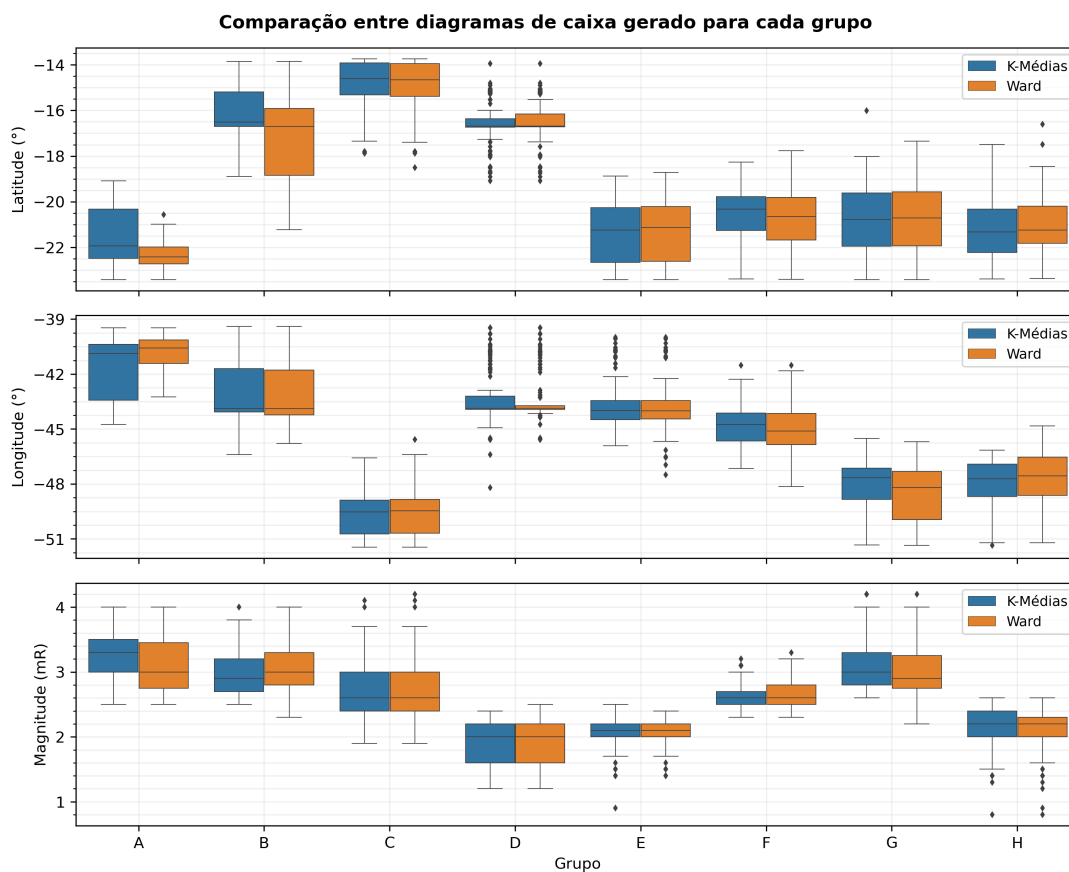


Figura 5.10: Comparação entre diagramas de caixa obtidos a partir dos grupos resultantes de cada método para $g = 8$.

Na Figura 5.10, além da similaridade entre as distribuições dos grupos, é possível notar similaridade nos *outliers*, reforçando as suposições relacionadas à localização dos eventos nos grupos C e D, e sobre as limitações relacionados aos extremos de magnitude para os grupos C, E e H.

Tabela 5.6: Comparação entre as estatísticas descritivas dos grupos obtidos para solução de $g = 8$ para o método de Ward (WD) e K-Médias (KM).

Latitude (°)																
Grupo	n		Média		D. Pad.		Mín.		P. 25%		P. 50%		P. 75%		Máx.	
Método	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM
A	43.00	58.00	-22.35	-21.52	0.67	1.26	-23.41	-23.41	-22.73	-22.48	-22.42	-21.94	-21.99	-20.32	-20.56	-19.09
B	102.00	75.00	-17.18	-16.15	1.94	1.08	-21.22	-18.90	-18.85	-16.70	-16.70	-16.51	-15.91	-15.20	-13.86	-13.86
C	109.00	106.00	-14.90	-14.85	1.14	1.09	-18.50	-17.88	-15.38	-15.32	-14.65	-14.61	-13.95	-13.91	-13.74	-13.74
D	127.00	135.00	-16.47	-16.55	0.90	0.93	-19.08	-19.08	-16.72	-16.73	-16.68	-16.68	-16.15	-16.36	-13.95	-13.95
E	156.00	156.00	-21.28	-21.42	1.31	1.26	-23.41	-23.41	-22.62	-22.65	-21.13	-21.25	-20.21	-20.26	-18.72	-18.87
F	216.00	202.00	-20.88	-20.65	1.44	1.32	-23.40	-23.38	-21.68	-21.26	-20.65	-20.32	-19.81	-19.78	-17.76	-18.26
G	75.00	82.00	-20.68	-20.79	1.45	1.49	-23.41	-23.41	-21.94	-21.95	-20.71	-20.78	-19.57	-19.62	-17.35	-16.01
H	118.00	132.00	-21.05	-21.29	1.36	1.38	-23.37	-23.39	-21.82	-22.23	-21.25	-21.33	-20.20	-20.33	-16.60	-17.49
Longitude (°)																
Grupo	n		Média		D. Pad.		Mín.		P. 25%		P. 50%		P. 75%		Máx.	
Método	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM
A	43.00	58.00	-40.81	-41.78	0.99	1.76	-43.24	-44.75	-41.43	-43.43	-40.57	-40.88	-40.15	-40.37	-39.48	-39.48
B	102.00	75.00	-43.24	-43.16	1.59	1.64	-45.80	-46.40	-44.22	-44.07	-43.89	-43.89	-41.78	-41.70	-39.40	-39.40
C	109.00	106.00	-49.52	-49.58	1.27	1.19	-51.45	-51.45	-50.70	-50.72	-49.47	-49.52	-48.85	-48.88	-45.57	-46.58
D	127.00	135.00	-43.33	-43.32	1.32	1.44	-45.57	-48.20	-43.92	-43.92	-43.87	-43.87	-43.74	-43.21	-39.47	-39.47
E	156.00	156.00	-43.84	-43.72	1.31	1.35	-47.51	-45.92	-44.46	-44.48	-44.01	-43.98	-43.45	-43.44	-39.99	-39.99
F	216.00	202.00	-45.15	-44.85	1.25	0.97	-48.15	-47.16	-45.85	-45.66	-45.12	-44.75	-44.16	-44.12	-41.53	-41.53
G	75.00	82.00	-48.58	-48.05	1.55	1.55	-51.36	-51.34	-49.94	-48.85	-48.20	-47.67	-47.32	-47.13	-45.70	-45.52
H	118.00	132.00	-47.82	-48.15	1.73	1.59	-51.22	-51.36	-48.63	-48.69	-47.57	-47.73	-46.53	-46.93	-44.83	-46.15
Magnitude (mR)																
Grupo	n		Média		D. Pad.		Mín.		P. 25%		P. 50%		P. 75%		Máx.	
Método	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM	WD	KM
A	43.00	58.00	3.09	3.25	0.43	0.35	2.50	2.50	2.75	3.00	3.00	3.30	3.45	3.50	4.00	4.00
B	102.00	75.00	3.05	2.96	0.38	0.35	2.30	2.50	2.80	2.70	3.00	2.90	3.30	3.20	4.00	4.00
C	109.00	106.00	2.73	2.73	0.51	0.49	1.90	1.90	2.40	2.40	2.60	2.60	3.00	3.00	4.20	4.10
D	127.00	135.00	1.90	1.91	0.37	0.37	1.20	1.20	1.60	1.60	2.00	2.00	2.20	2.20	2.50	2.40
E	156.00	156.00	2.06	2.06	0.20	0.23	1.40	0.90	2.00	2.00	2.10	2.10	2.20	2.20	2.40	2.50
F	216.00	202.00	2.64	2.61	0.21	0.21	2.30	2.30	2.50	2.50	2.60	2.60	2.80	2.70	3.30	3.20
G	75.00	82.00	3.01	3.10	0.41	0.37	2.20	2.60	2.75	2.80	2.90	3.00	3.25	3.30	4.20	4.20
H	118.00	132.00	2.09	2.15	0.32	0.31	0.80	0.80	2.00	2.00	2.20	2.20	2.30	2.40	2.60	2.60

Nessa tabela, o ponto (“.”) é utilizado como notação para decimal.

5.4. Associação com com domínios tectônicos

A Figura 5.11 apresenta o resultado dos conglomerados de sismos, para a solução de 8 grupos ($g = 8$) pelo método de Ward, sobreposto aos domínios tectônicos compreendidos pela área de interesse.

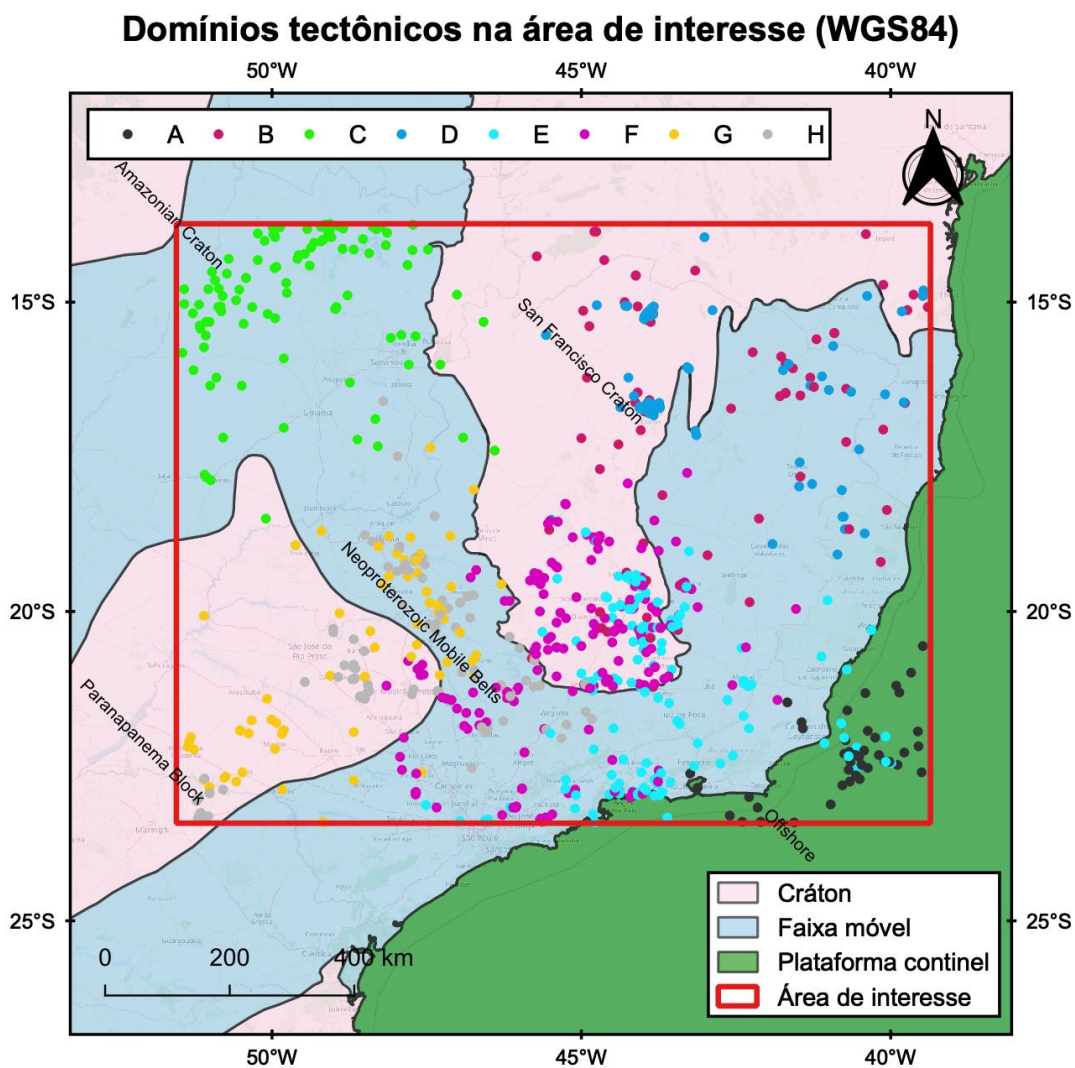


Figura 5.11: Associação entre sismos agrupados pelo método de Ward e os tipos de domínios tectônicos compreendidos pela área de interesse.

A Figura 5.12 apresenta o resultado da associação espacial entre os conglomerados de sismos, obtidos pelo método de Ward, com os tipos de domínios tectônicos compreendidos pela área de interesse. Os tipos de domínios tectônicos correspondem as entidades tectônicas categorizadas em função da sua natureza tectônica, conforme discutido na Subseção 4.1.

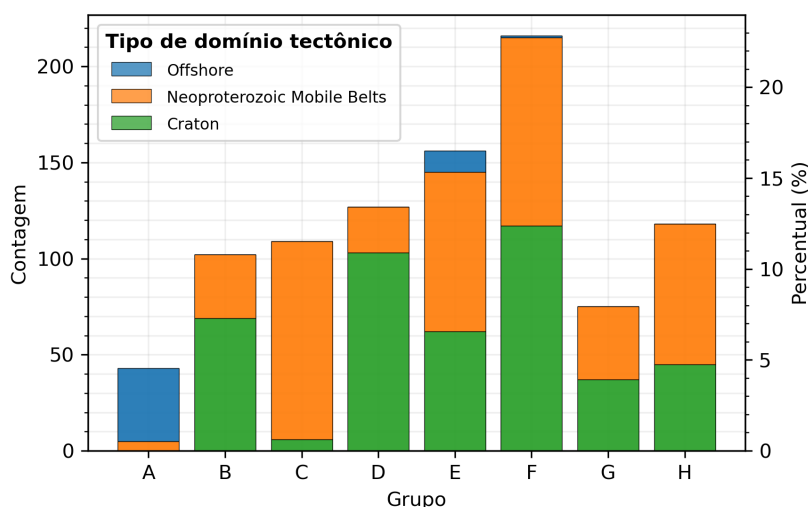


Figura 5.12: Histograma da contagem de sismos, separados em função dos grupos, e associados aos tipos de domínios tectônicos compreendidos pela área de interesse.

Na Figura 5.12, nota-se que não há nenhum grupo que tenha associado-se a apenas um único tipo de domínio tectônico. Entretanto, é possível verificar que os grupos A, C e D apresentam um predomínio marcante dos domínios Offshore, Faixas Móveis do Proterozóico e Cráton, respectivamente.

A Figura 5.13 apresenta o resultado da associação espacial entre os conglomerados de sismos pelo método de Ward, com os domínios tectônicos específicos compreendidos pela área de interesse. Os domínios tectônicos específicos correspondem as entidades tectônicas individualizadas em função de suas respectivas ocorrências.

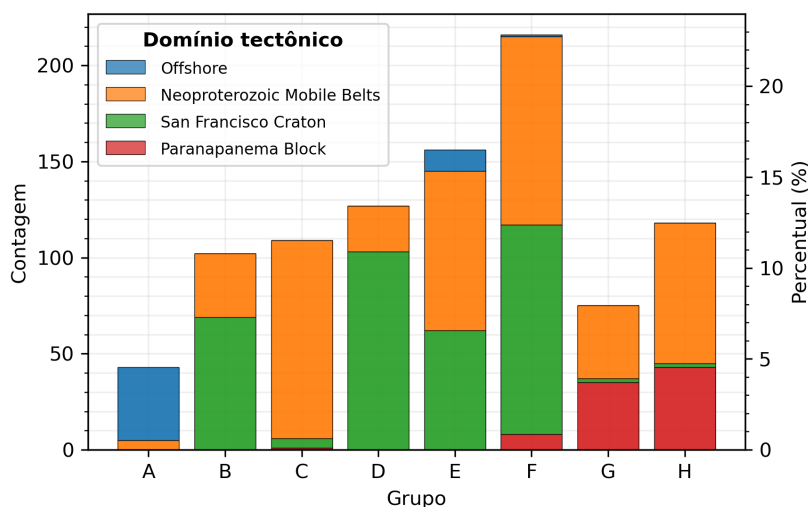


Figura 5.13: Histograma da contagem de sismos, separados em função dos grupos, e associados aos domínios tectônicos específicos compreendidos pela área de interesse.

Na Figura 5.13, assim como observado na análise dos tipos de domínios, nota-se que não há nenhum grupo que tenha associado-se a apenas um único domínio tectônico específico. Entretanto, é possível observar a diferença em relação aos terrenos mais estáveis (crátons), representados pelo Crátón São Francisco (predominante nos grupos B, C, D, E e F) e pelo Bloco Paranapanema (predominante nos grupos G e H).

5.5. Comparação com estudos anteriores

Para uma avaliação complementar da classificação de sismos, utilizou-se o trabalho de Schweig (2013) como referência. A Figura 5.14, apresenta as zonas sismogênicas do Brasil, previamente apresentadas na Figura 3.1, extraídas do trabalho Schweig (2013) como uma adaptação do trabalho de Mito (1993). Ressalta-se que essas zonas sismogênicas foram obtidas através da integração visual de informações sismológicas, geológicas e geofísicas.

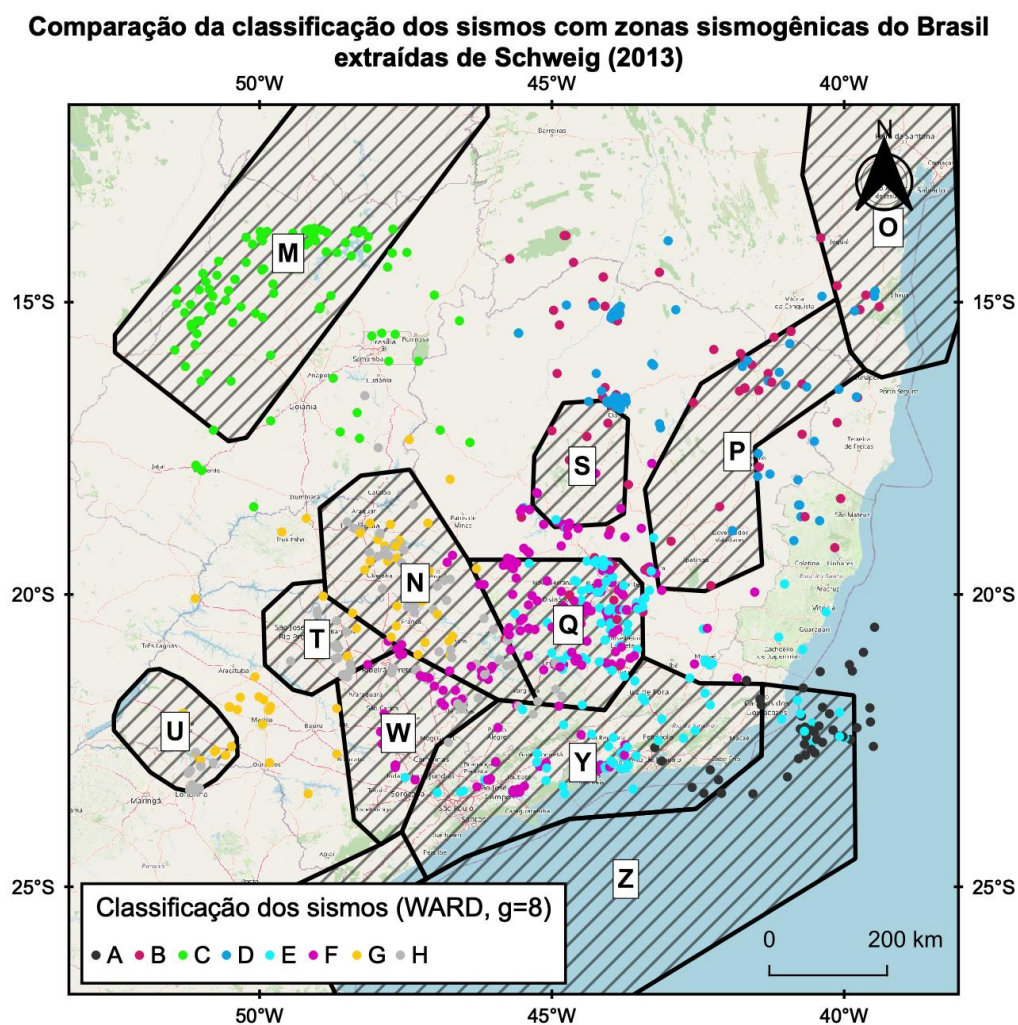


Figura 5.14: Comparação da classificação dos sismos com zonas sismogênicas do Brasil extraídas de Schweig (2013), previamente apresentadas na Figura 3.1.

A Figura 5.15 apresenta o resultado da associação espacial entre os grupos de sismos pelo método de Ward e as zonas sismogênicas do Brasil, extraídas de Schweig (2013), compreendidas pela área de interesse.

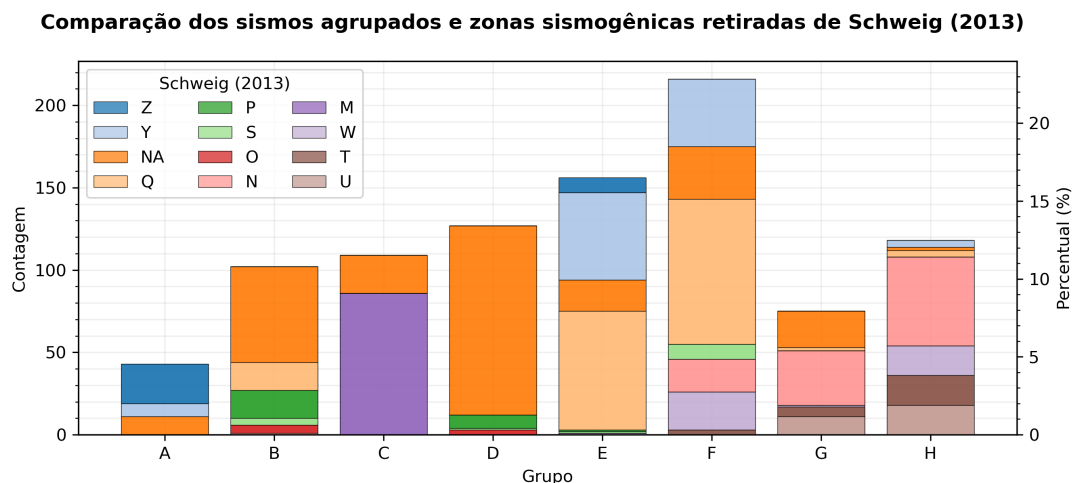


Figura 5.15: Associação entre os grupos de sismos com zonas sismogênicas do Brasil extraídas de Schweig (2013). A legenda NA significa “Não aplicável” e representa eventos que não foram associados com nenhuma zona sismogênica.

Adicionalmente, a Tabela 5.7 apresenta uma síntese da associação entre os conglomerados obtidos pelo método de Ward e as zonas sismogênicas extraída do trabalho de Schweig (2013).

Tabela 5.7: Síntese da associação entre grupos obtidos pelo método de Ward e as zonas sismogênicas extraídas do trabalho de Schweig (2013).

Grupo	Zona(s) sismogênica(s) (SCHWEIG, 2013)
A	Z, Y
B	N, O, P, S, Q
C	M
D	O, S, P
E	W, S, P, Q, Y
F	T, W, N, S, Q, Y
G	U, T, W, N, Q
H	U, T, W, N, Q, Y

Na Figura 5.15 e na Tabela 5.7, nota-se que somente o grupo C associou-se a apenas a uma única zona sismogênica (M - Porangatu). Destaca-se que nenhum grupo apresentou a mesma relação de zonas sismogênicas. Entretanto, é possível verificar que aproximadamente 30% dos sismos não foram associados à nenhuma zona

sismogênicas (NA), ou seja, existem muitos sismos que podem não ser representados por meio das características gerais dessas zonas sismogênicas.

5.6. Delimitação das zonas sismogênicas

A Figura 5.16 apresenta os centroides dos 8 grupos obtidos pelo método de Ward os respectivos polígonos de Voronoi ajustados para os centroides, correspondentes aos potenciais limites das zonas sismogênicas estimados nesse estudo.

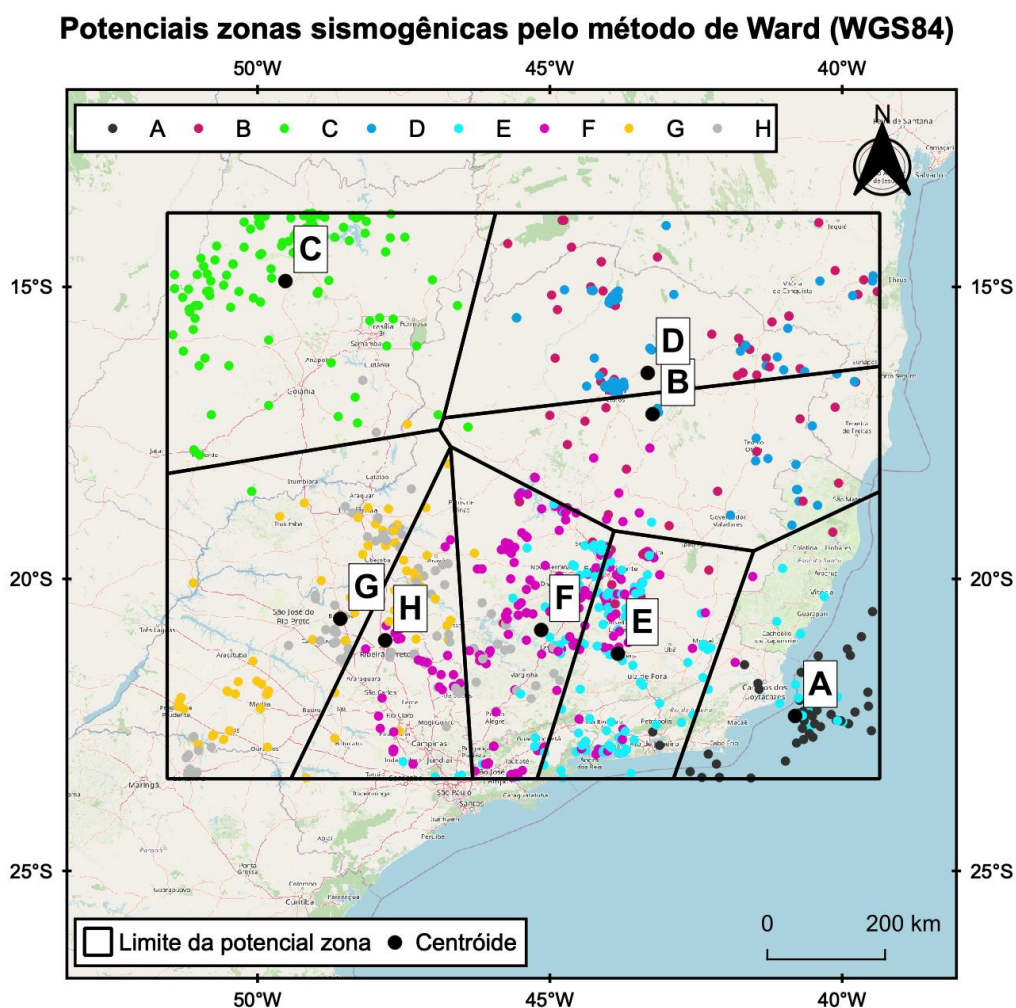


Figura 5.16: Potenciais zonas sismogênicas delimitadas a partir dos grupos obtidos pelo método de Ward.

Na Figura 5.16, nota-se que nenhuma das zonas compreendem sismos de apenas um único grupo. O que se observa é que há uma predominância de sismos correspondentes ao grupo que deu origem à respectiva zona sismogênica. Possivelmente, esse comportamento ocorre por conta do uso do centróide como o ponto de referência para o cálculo dos polígonos de Voronoi. Nesse caso, o centróide é influenciado por todos os pontos pertencentes ao mesmo grupo, inclusive pontos que geograficamente não ficam aglomerados (comportamento esperado para os grupos).

A Figura 5.17 apresenta um comparativo entre as potenciais zonas sismogênicas delimitadas nesse estudo as zonas sismogênicas extraídas do trabalho de Schweig (2013).

Na Figura 5.17, é possível observar certa concordância entre a zona C, obtida nesse trabalho, e a zona M, extraída do trabalho de Schweig (2013). Também se observa concordância entre as zona B e as zonas S e P, extraídas do trabalho de Schweig (2013). Nos demais casos, nota-se que que não há uma correspondência visual clara entre as zonas. Nota-se que os polígonos tracejados (zonas sismogênicas obtidas nesse trabalho) compreendem porções de mais de uma zona sismogênica do trabalho de Schweig (2013), como por exemplo, a zona G (obtida nesse trabalho) e as zonas U, T e N do trabalho de Schweig (2013).

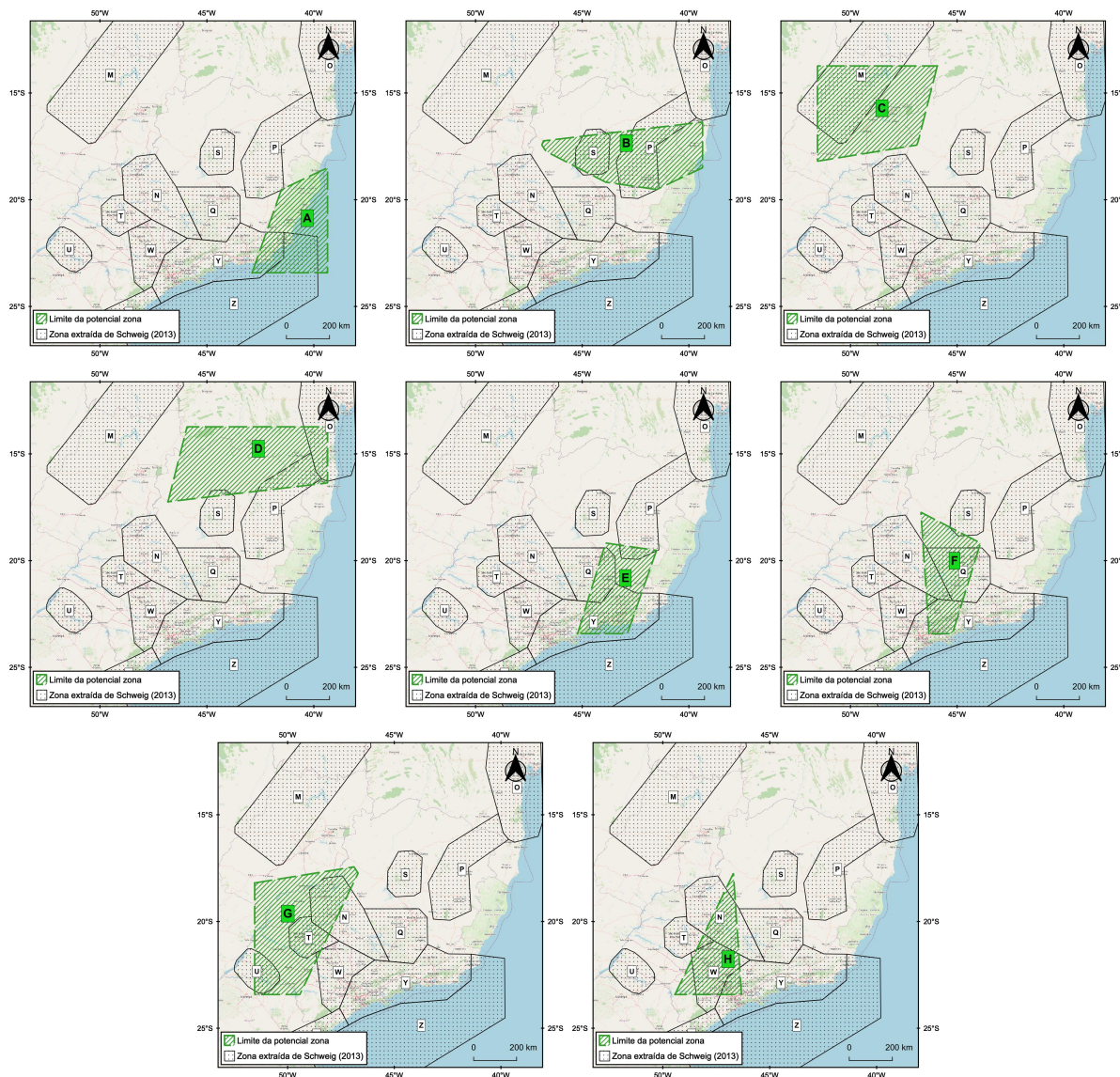


Figura 5.17: Comparação das potenciais zonas sismogênicas delimitadas nesse estudo (em verde) com as zonas sismogênicas extraídas do trabalho de Schweig (2013).

Possivelmente as diferenças observadas na Figura 5.17 ocorrem devido à: (i) diferença de escala de mapeamento; (ii) restrição espacial de dados (limitado à área de interesse do estudo); e ao (iii) uso de uma base de dados mais atualizada - que, por consequência, incorpora um número maior de eventos quando comparado aos estudos de outros autores.

Adicionalmente, ressalta-se que os resultados obtidos nesse estudo apresentam continuidade espacial ao longo de toda a área de interesse, ou seja, toda a região de interesse pode ser diferenciada em termos de zonas sismogênicas, evitando a não associação de sismos que ocorrem em regiões fora dos limites das zonas sismogênicas de acordo com outros autores anteriormente mencionados.

6. CONSIDERAÇÕES FINAIS

A partir dos resultados apresentados e das análises realizadas nesse estudo, foi possível concluir que:

- A técnica de agrupamentos, através do método de Ward, produziu 8 agrupamentos que fomentaram a delimitação de zonas sismogênicas de Minas Gerais e regiões adjacentes;
- A exclusão do parâmetro de profundidade foi necessária por conta da grande quantidade de dados com ausência desse parâmetro. Adicionalmente, a padronização das demais variáveis foi importante para evitar o maior peso das coordenadas em relação ao parâmetro de magnitude;
- Foi observada uma alta convergência entre os resultados obtidos pelos métodos de Ward e K-Médias. Ainda que a solução obtida por K-Médias tenha apresentado mais de similaridade nos grupos e maior heterogeneidade entre os grupos, o método de Ward ainda se mostra como uma escolha coerente por permitir uma análise objetiva para a escolha do número de grupos;
- Na avaliação das distribuições dos parâmetros nos grupos resultantes, foram observados alguns elementos discrepantes, com maior destaque para os grupos C, D, E e H. Tais discrepâncias foram relacionadas ao maior afastamento de alguns eventos nos grupos C e D, em relação aos aglomerados formados pela maioria dos eventos, e também sobre as limitações relacionados aos extremos de magnitude nos grupos C, E e H, tanto pela baixa detecção de eventos de magnitude inferior a 2, quanto na rara ocorrência de eventos com magnitude superior a 4;
- Os sismos classificados apresentaram associações com domínios tectônicos compreendidos pela região de interesse e com zonas sismogênicas retiradas de estudos prévios, com maior destaque para os grupos A (plataforma continental) e C (Faixas móveis do Neoproterozóico);

- As zonas sismogênicas delimitadas nesse estudo, diferente das zonas extraídas de outros trabalhos, como por exemplo Schweig (2013), apresentam continuidade espacial, i.e., toda a região de interesse pode ser diferenciada em termos de distintas zonas sismogênicas;
- A combinação das técnicas de estatística multivariada e geoprocessamento se demonstrou uma metodologia objetiva viável para o mapeamento de zonas sismogênicas;

A metodologia utilizada neste trabalho para delimitação das zonas sismogênicas se baseou em um algoritmo matemático e poderia ser reavaliada com base na opinião de sismólogos especializados no tema e alteração nos parâmetros do modelo (variáveis de entrada, número de grupos etc.). As divergências observadas entre as zonas sismogênicas auferidas e as obtidas de estudos prévios possivelmente ocorreram devido à diferença de escala de mapeamento, limites espaciais da área de interesse, e ao uso de uma base de dados mais atualizada que, por consequência, incorpora um número maior de eventos quando comparado aos estudos de outros autores. Também existem questões que podem estar relacionadas à subjetividade da opinião dos sismólogos que trabalharam nesses estudos.

Adicionalmente, sugere-se que, em trabalhos futuros, a avaliação dos resultados incorpore a ótica da caracterização geológica e geofísica da região de interesse, incluindo novas variáveis que poderão agregar mais diferenciação ou similaridade para os grupos (como foi o caso dos domínios tectônicos), além do teste de outras técnicas de agrupamentos, e.g., *K-Prototypes* (HUANG, 1998), que possam incluir mais variáveis categóricas que poderão ser associadas aos eventos sísmicos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABNT. Projeto de estruturas resistentes a sismos. *Norma brasileira ABNT NBR 15.421*, 2006.
- ASSUMPÇÃO, M. A regional magnitude scale for brazil. *Bulletin of the Seismological Society of America*, The Seismological Society of America, v. 73, n. 1, p. 237–246, 1983.
- ASSUMPÇÃO, M. d. S. et al. Terremotos no brasil: preparando-se para eventos raros. *Boletim SBGf*, n. 96, p. 25–29, 2016.
- AURENHAMMER, F. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 23, n. 3, p. 345–405, 1991.
- BERROCAL, J. Atividade sísmica e sua ocorrência. *Bull. Braz. Geophys. Soc.*, p. 30–31, 2016.
- BERROCAL J., M. et al. Sismicidade do brasil. *Rede Sismográfica Brasileira*, Editado por IAG/CNEN, 1984.
- CHARRAD, M. et al. NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, v. 61, n. 6, p. 1–36, 2014. Disponível em: <<https://www.jstatsoft.org/v61/i06/>>.
- FORTUNE, S. Voronoi diagrams and delaunay triangulations. *Computing in Euclidean geometry*, World Scientific, p. 225–265, 1995.
- GARCIA, D.; WALD, D. J.; HEARNE, M. A global earthquake discrimination scheme to optimize ground-motion prediction equation selection. *Bulletin of the Seismological Society of America*, Seismological Society of America, v. 102, n. 1, p. 185–203, 2012.
- GUTENBERG, B.; RICHTER, C. F. Frequency of earthquakes in california. *Bulletin of the Seismological society of America*, Seismological Society of America, v. 34, n. 4, p. 185–188, 1944.
- HARABASZ, C. T.; KAROŃSKI, M. A dendrite method for cluster analysis. In: *Communications in Statistics*. [S.l.: s.n.], 1974. v. 3, n. 1, p. 1–27.
- HÄRDLE, W.; SIMAR, L. et al. *Applied multivariate statistical analysis*. 4. ed. [S.l.]: Springer, 2015.
- HARTIGAN, J. A.; WONG, M. A. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, JSTOR, v. 28, n. 1, p. 100–108, 1979.
- HUANG, Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, Springer, v. 2, n. 3, p. 283–304, 1998.
- JUNIOR, M.; CEZÁRIO, J. Comparação de critérios para determinação do número de clusters. 2018.
- LOPES, A. E. d. V.; NUNES, L. C. Intensidades sísmicas de terremotos: formulação de cenário sísmico no brasil. *Revista USP*, n. 91, p. 90–103, 2011.
- MAECHLER, M. et al. *cluster: Cluster Analysis Basics and Extensions*. [S.l.], 2022. R

package version 2.1.4 — For new features, see the 'Changelog' file (in the package source). Disponível em: <<https://CRAN.R-project.org/package=cluster>>.

MINGOTI, S. A. Análise de dados através de métodos estatística multivariada: uma abordagem aplicada. In: *Análise de dados através de métodos estatística multivariada: uma abordagem aplicada*. [S.l.: s.n.], 2007. p. 295–295.

Minitab, LLC. *Dendrograma*. 2021. Disponível em: <<https://support.minitab.com/pt-br/minitab/20/help-and-how-to/statistical-modeling/multivariate/how-to/cluster-observations/interpret-the-results/all-statistics-and-graphs/dendrogram/>>.

MIOTO, J. Sismicidade e zonas sismogênicas do brasil. *Tese de Doutorado*, v. 2, 1993.

PAMULA, H. *Centroid calculator*. Omni Calculator, 2023. Disponível em: <<https://www.omnicalculator.com/math/centroid#what-is-centroid>>.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. R version 4.2.1. Disponível em: <<https://www.R-project.org/>>.

REHMAN, K.; BURTON, P. W.; WEATHERILL, G. A. K-means cluster analysis and seismicity partitioning for pakistan. *Journal of seismology*, Springer, v. 18, n. 3, p. 401–419, 2014.

RSBR, R. S. B. *Catálogo Sísmico Brasileiro*. 2022.

RStudio Team. *RStudio: Integrated Development Environment for R*. Boston, MA, 2020. Disponível em: <<http://www.rstudio.com/>>.

SCHWEIG, C. Caracterização tectonoestratigráfica da porção adjacente ao epicentro do sismo são vicente na bacia de santos. *Dissertação de Mestrado*, Universidade Estadual Paulista (UNESP), 2013.

USGS. *What is a seismic zone, or seismic hazard zone?* United States Geological Survey, 2019. Disponível em: <<https://www.usgs.gov/faqs/what-seismic-zone-or-seismic-hazard-zone>>.

WARD, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, Taylor & Francis, v. 58, n. 301, p. 236–244, 1963.

WEATHERILL, G.; BURTON, P. W. Delineation of shallow seismic source zones using k-means cluster analysis, with application to the aegean region. *Geophysical Journal International*, Blackwell Publishing Ltd Oxford, UK, v. 176, n. 2, p. 565–588, 2009.

WEISSTEIN, E. W. *Geometric centorid*. A Wolfram Web Resource, 2002. Disponível em: <<https://mathworld.wolfram.com/GeometricCentroid.html>>.

WEISSTEIN, E. W. *Voronoi diagram*. A Wolfram Web Resource, 2008. Disponível em: <<https://mathworld.wolfram.com/VoronoiDiagram.html>>.

WIEMER, S.; BAER, M. Mapping and removing quarry blast events from seismicity catalogs. *Bulletin of the Seismological Society of America*, Seismological Society of America, v. 90, n. 2, p. 525–530, 2000.