

UNIVERSIDADE DE SÃO PAULO

Instituto de Geociências

Monografia do Trabalho de Formatura

**APLICAÇÃO DE ALGORITMOS DE AGRUPAMENTO NA DEFINIÇÃO DE SUBDOMÍNIOS
PARA ESTIMATIVA DOS TEORES DE FERRO DA MINA CAPANEMA, SANTA
BÁRBARA (MG)**

TF 21/27

Lúcio Dias Cruz Massari

Orientador: Professor Doutor Marcelo Monteiro da Rocha

Coorientador: Geólogo Lucas Bassan

São Paulo

2022

UNIVERSIDADE DE SÃO PAULO

Instituto de Geociências

Monografia do Trabalho de Formatura

**APLICAÇÃO DE ALGORITMOS DE AGRUPAMENTO NA DEFINIÇÃO DE SUBDOMÍNIOS
PARA ESTIMATIVA DOS TEORES DE FERRO DA MINA CAPANEMA, SANTA
BÁRBARA (MG)**

TF 21/27

Lúcio Dias Cruz Massari

Orientador: Professor Doutor Marcelo Monteiro da Rocha

Coorientador: Geólogo Lucas Bassan

São Paulo

2022

Agradecimentos

Dedico meus agradecimentos a todos que enriqueceram a jornada que este trabalho encerra. Desses, gostaria de destacar a minha família pelo carinho, apoio e aprendizados recebidos, por tornarem o início desta jornada possível e por serem, cada um à sua maneira, grandes exemplos de vida. Em especial à minha mãe, meu pai, minhas tias, meus irmãos e irmãs, ao Ju e à Pri, por estarem sempre presentes e me incentivarem. Dos que não estão mais presentes fisicamente, minha gratidão à vó Jô pelo carinho bruto e ao vó Toninho por todas as botas de campo engraxadas, além do amor incondicional de ambos.

Agradeço também a todos professores do Instituto de Geociências da USP, sobretudo ao meu orientador prof. Marcelo Rocha pela forma de ensinar e toda a atenção dispensada ao longo da realização deste trabalho. Aos professores Caetano, Lena, Rafa, Moreschi e Torrinha por despertarem em mim o gosto pela geologia econômica e o setor mineral, e aos professores Boggiani, Giannini, Mário Campos, Ninguém, Valdecir, Vinicius e Verônica (MAE-USP) também pela maneira ímpar de ensinar e pelo companheirismo.

Ao amigo e coorientador Lucas Bassan por me acompanhar, ensinar e ajudar em tantas etapas dos últimos anos.

A todos funcionários do Instituto de Geociências, especialmente ao casal mais forrozeiro do IGc (Miriam e Ideval) e ao restante da equipe do Museu de Geociências por tantas experiências incríveis.

Aos muitos amigos da graduação e da pós graduação, em especial aos da Turma 59 e do GGeo e aos integrantes dos seletos grupos The N, Beija & Grita e Gogeo. Pelo apoio mútuo e ótimas lembranças, agradeço aos queridos da geologia Nit, Bibifa, Brisão, Calypso, Chicó, Dani, Dino, Geovana, Kiwi, Lilo, Miçanga, Miss, Ocotô, Pamona, Pinguim, Provisório, Sizu, Tom, Trevas e Waze; e aos queridíssimos de outras bandas Marcela, Fernanda, Renata, Cadol, Isa, Camila e Tales.

Ao Serviço Geológico do Brasil e à empresa Datamine pelas oportunidades de estágio e os conhecimentos adquiridos, e também aos colegas do Programa Datamine Universidades Franco, André, Maciel, Bénédicte e Raylline.

Aos meus gatinhos Boris e Tigresa pelos abraços ronronantes que aliviaram cada estresse de 2021.

E por fim, agradeço às universidades públicas pela produção, preservação e divulgação científica e cultural, pelo ensino de qualidade e seu retorno imensurável para a nossa sociedade.

Sumário

RESUMO	8
ABSTRACT	9
1. INTRODUÇÃO.....	10
2. METAS E OBJETIVOS.....	11
3. LEVANTAMENTO BIBLIOGRÁFICO	11
3.1. Contexto Geológico Regional	11
3.1.1. <i>Litoestratigrafia do Quadrilátero Ferrífero</i>	12
3.1.2. <i>Sinclinal Ouro Fino</i>	14
3.1.3. <i>Aspectos econômicos da mina Capanema</i>	15
3.2. Análise de Agrupamento (<i>Cluster Analysis</i>)	16
3.2.1. <i>Algoritmo K-means</i>	16
3.2.2. <i>Algoritmo Aglomerativo Hierárquico.....</i>	18
3.3. Estimativa e Validação dos Resultados	19
3.3.1. <i>Geoestatística.....</i>	19
3.3.2. <i>Krigagem ordinária (KO).....</i>	21
3.3.3. <i>Métodos de validação de estimativa</i>	22
4. MATERIAIS E MÉTODOS	24
4.1. Materiais	24
4.1.1. <i>Banco de dados.....</i>	24
4.1.2. <i>Infraestrutura de informática.....</i>	26
4.2. Métodos	26
4.2.1. <i>Levantamento bibliográfico.....</i>	26
4.2.2. <i>Validação e análise exploratória dos dados.....</i>	26
4.2.3. <i>Definições de subdomínios</i>	26
4.2.4. <i>Modelo de blocos, análise geoestatística e estimativa</i>	26
4.2.5. <i>Validação da estimativa e comparação entre os cenários.....</i>	26
5. RESULTADOS OBTIDOS	27
5.1. Análise Exploratória dos Dados.....	27
5.1.1. <i>Litotipos da mina Capanema.....</i>	27
5.1.2. <i>Validação e tratamento dos dados.....</i>	28
5.1.3. <i>Análise estatística.....</i>	28
5.1.1. <i>Análise visual das variáveis no espaço</i>	31
5.2. Agrupamento das Amostras	32
5.3. Modelo de Blocos.....	36
5.3.1. <i>Definição dos Subdomínios.....</i>	37
5.4. Análise Geoestatística	38

5.5. Estimativas	39
5.6. Validação e Avaliação dos Efeitos do Agrupamento nas Estimativas	40
6. INTERPRETAÇÃO E DISCUSSÃO DOS RESULTADOS	43
6.1. Agrupamentos	43
6.2. Análise Geoestatística	45
6.3. Validação e Avaliação do Efeito do Agrupamento nas Estimativas	46
7. CONCLUSÕES.....	47
8. CONSIDERAÇÕES FINAIS.....	49
9. REFERÊNCIAS BIBLIOGRÁFICAS.....	49
APÊNDICE I	53
APÊNDICE II	56
APÊNDICE III	57
APÊNDICE IV	59
APÊNDICE V	64
APÊNDICE VI.....	66
APÊNDICE VII.....	68
APÊNDICE VIII.....	69
APÊNDICE IX.....	70
APÊNDICE X.....	72

Índice de Figuras

Figura 1: Mapa geológico do Quadrilátero Ferrífero e localização da mina Capanema. Modificado do banco de dados da CPRM (2004).....	11
Figura 2: Coluna estratigráfica do Quadrilátero Ferrífero. (1) Complexos granito-gnáissicos do embasamento, (2) Rochas ultramáficas, (3) Rochas maficas, (4) Unidade químico-clástica, (5) Unidade clástica, (6) Tonalitos, (7) Conglomerados e quartzitos da Fm. Moeda, (8) Filitos carbonosos da Fm. Batatal, (9) Itabiritos da Fm. Cauê, (10) Dolomitos da Fm. Gandarela, (11) Quartzitos e filitos, (12) Conglomerados e filitos, (13) Conglomerados e quartzitos, (14) <i>Sills</i> e diques de diabásio. (Dardenne & Schobbenhaus, 2001).....	13
Figura 3: Mapa geológico do Sinclinal Ouro Fino e perfil geológico A-B seccionando a cava atual da mina Capanema. (Franco & Endo, 2004)	15
Figura 4: Ilustração da primeira iteração do algoritmo <i>K-means</i> . (a) Distribuição aleatória das amostras em 3 grupos, (b) cálculo dos centroides dos agrupamentos e (c) redistribuição das amostras de acordo com a posição dos centroides. (James et al., 2013)	17
Figura 5: Ilustração de mais iterações do algoritmo <i>K-means</i> . (a) Os centroides são repositionados de acordo com as amostras a eles atribuídas na primeira iteração e (b) o processo se repete <i>n</i> vezes até atingir a estabilização dos centroides. (James et al., 2013) 17	17
Figura 6: Exemplo de gráfico da inércia total pela quantidade de <i>clusters</i> . (Yuan & Yang, 2019)	18
Figura 7: Representação de (a) dados estruturados em uma hierarquia multiníveis (dendrograma) e (b) seleção de um nível hierárquico para gerar quatro <i>clusters</i> distintos. (Kassambara, 2017)	19
Figura 8: Propriedades de um modelo de semivariograma. (Sinclair & Blackwell, 2004)....	20
Figura 9: Histogramas (a) dos teores das amostras, (b) dos teores estimados na validação cruzada e (c) das diferenças entre os valores reais e estimados; e (d) gráfico de dispersão entre valores reais e estimados. (Sinclair & Blackwell, 2004).....	22
Figura 10: <i>Swath plot</i> ao longo da direção E-W, com fatias N-S de espessura igual a 50 metros. (Rossi & Deutsch, 2014)	23
Figura 11: Malha dos furos de perfuratriz em planta.	25
Figura 12: Seção vertical SW-NE dos furos de perfuratriz.	25
Figura 13: Malha dos furos de perfuratriz em perspectiva com legenda de litotipo.....	27
Figura 14: <i>Boxplots</i> dos teores de (a) ferro e (b) alumina por litotipo do banco de dados validado e tratado.	29
Figura 15: Histogramas gerais dos teores de (a) ferro, (b) alumina, (c) fósforo em escala de log-probabilidade aritmética e (d) sílica; e (e) da perda por calcinação dos dados validados e tratados.....	30
Figura 16: Gráficos de dispersão entre (a) ferro e sílica e (b) entre os <i>outliers</i> de fósforo e a perda por calcinação do banco de dados validado e tratado com legenda de litotipos.....	31

Figura 17: Malha em perspectiva com legenda de teores de fósforo.....	32
Figura 18: Perfil vertical SW-NE da malha com legenda de teores de ferro.	32
Figura 19: Gráfico da inércia total (soma total dos quadrados intra- <i>cluster</i>) pela quantidade de <i>clusters</i> com o algoritmo de análise de agrupamento <i>K-means</i>	33
Figura 20: Malha dos furos de perfuratriz com a legenda de <i>clusters</i> por algoritmo de análise de agrupamento. Linhas tracejadas demonstram em quais porções da malha o resultado do <i>K-means</i> se assemelha mais com o HC e o GHC.	34
Figura 21: Boxplots dos teores de ferro, alumina e fósforo em cada <i>cluster</i> por algoritmo de análise de agrupamento. <i>K-means</i> (KM), Aglomerativo Hierárquico (HC) e Aglomerativo Hierárquico Geoestatístico (GHC).....	35
Figura 22: Modelo de blocos sobre a malha dos furos de perfuratriz em planta.	36
Figura 23: Esquematização da relação entre os subdomínios das amostras (estéril e minério) e a definição de subdomínios no modelo de blocos (estéril, transição e minério).....	37
Figura 24: Vista em planta do modelo de blocos em cada um dos quatro cenários de definição de subdomínios para estimativa.....	38
Figura 25: Variogramas experimentais e respectivos modelos teóricos ajustados do ferro no cenário de domínio único. (a) Omnidirecional, (b) direção 1, (c) direção 2 e (d) direção 3. Eixo vertical: variância relativa à variância a priori (<i>gamma</i>).	39
Figura 26: Vista em planta do modelo de blocos com teores estimados de ferro em cada cenário.	40
Figura 27: Histogramas dos teores de ferro (a) nas amostras e (b) resultantes da validação cruzada; (c) histograma dos erros e (d) dispersão entre os teores reais e as estimativas pontuais para o cenário do <i>K-means</i>	41
Figura 28: Gráficos Quantil-Quantil entre os teores de ferro do conjunto amostral e as estimativas dos cenários com domínio único e com agrupamento por <i>K-means</i> e GHC.....	42
Figura 29: Médias móveis do ferro por fatias de orientação E-W e com espessura igual a 100 metros.....	43
Figura 30: Dendrograma do agrupamento realizado pelo método Ward do algoritmo Aglomerativo Hierárquico. Altura = Distância euclidiana entre os pares no espaço multivariado.	44
Figura 31: Análises de contato entre os subdomínios estéril (0) e minério (1) para (a) o algoritmo HC e (b) o algoritmo <i>K-means</i>	45
Figura 32: Elipsoide de continuidade do ferro no domínio único.	45
Figura 33: Variogramas experimentais e respectivos modelos teóricos na direção de maior continuidade do ferro para (a) o domínio único e (b) o subdomínio de minério do HC. Eixo vertical: variância relativa à variância a priori (<i>gamma</i>).	46
Figura 34: Comparação entre os teores das amostras e do modelo de blocos nos diferentes cenários de agrupamento no setor norte do depósito.	47

Índice de Tabelas

Tabela 1: Teores de corte máximos e mínimos para o minério e para o ROM da mina Capanema por faixa granulométrica. (Rocha, 1999)	16
Tabela 2: Síntese das informações do banco de dados.	24
Tabela 3: Estatísticas descritivas gerais do banco de dados validado e tratado.....	29
Tabela 4: Matriz de correlação linear entre as variáveis numéricas do banco de dados validado e tratado, destacadas as maiores correlações.	31
Tabela 5: Teores médios das variáveis critério em cada agrupamento por algoritmo de análise de agrupamento, destaca as médias que caracterizam cada grupo.	34
Tabela 6: Propriedades das células parentais do modelo de blocos.....	36
Tabela 7: Os quatro cenários de definição dos subdomínios estéril, minério e de transição.	37
Tabela 8: Estatísticas descritivas dos teores de ferro nas amostras e nos modelos estimados em cada cenário, destacados os valores mínimo e máximo, média e variância mais próximos do conjunto amostral.....	41

RESUMO

Algoritmos de análise de agrupamento têm sido testados e continuamente adaptados para realizar a definição de subdomínios para estimativa de recursos e reservas minerais com critérios multivariados, como o *K-means* (KM) e o Aglomerativo Hierárquico (HC), elaborados para variáveis não regionalizadas; e o Aglomerativo Hierárquico Geoestatístico (GHC), que considera as relações geoestatísticas entre as amostras. No presente estudo, 6.719 furos de perfuratriz do tipo *rockdrill* da mina de ferro Capanema, localizada na porção centro-leste do Quadrilátero Ferrífero (Minas Gerais), foram submetidos a quatro diferentes métodos de agrupamento para avaliar os efeitos da definição de subdomínios na estimativa dos teores de ferro. Os agrupamentos foram realizados manualmente (primeiro cenário) com base nos litotipos da mina; e pela aplicação dos algoritmos KM, HC e GHC (segundo a quarto cenários) com base nos teores de ferro, alumina e fósforo dos furos. A análise geoestatística das amostras nos cinco cenários (incluso o de domínio único) revelou forte controle estrutural-litológico na distribuição dos teores de ferro, pois os planos principais das elipsoides de continuidade são paralelos à xistosidade S_2 do Sinforme Córrego das Flechas, que abrange o depósito. O agrupamento das amostras conferiu, no geral, maior acurácia e precisão para as estimativas com krigagem ordinária em relação ao cenário de domínio único, embora com pequena diferença. Destaca-se o agrupamento do algoritmo *K-means* como o cenário que apresentou os melhores resultados de estimativa, verificados na validação cruzada (maior coeficiente de correlação linear e menores média e variância dos erros) e na comparação de suas estatísticas descritivas com as do conjunto amostral (menor alteração da média e da variância dos teores de ferro nas amostras). A estimativa sem agrupamento ocasionou o maior efeito de suavização pela superestimativa e subestimativa de teores baixos e altos, respectivamente, o que resultou no modelo com a distribuição mais homogênea de teores. Para a mina Capanema, ainda que os pequenos ganhos em acurácia e precisão de estimativa conferidos pela definição de subdomínios não justifiquem os efeitos negativos observados e o tempo empregado, os métodos de agrupamento podem ter útil aplicação nos contextos em que há necessidade de maior detalhamento dos contrastes locais de teores, como na avaliação de substâncias heterogêneas e modelos de curto prazo.

ABSTRACT

Cluster analysis algorithms have been tested and continually adapted to define subdomains for estimating mineral resources and reserves based on multivariate criteria, such as K-means (KM) and Agglomerative Hierarchical Clustering (HC), developed for non-regionalized variables; and the Geostatistical Hierarchical Clustering (GHC), which considers the geostatistical relationships between samples. In the present study, 6,719 rockdrill-type samples from the Capanema iron mine, located in the east-central portion of Quadrilátero Ferrífero (Minas Gerais), were submitted to four different subdomains definition methods to assess its effects on the estimates for iron. The subdomains definitions were performed manually (first scenario) based on the mine's lithotypes, and by cluster analysis with KM, HC and GHC algorithms (second to fourth scenarios) based on the iron, phosphorus and alumina sample contents. The geostatistical analysis of the samples in all five scenarios (single-domain included) revealed a strong structural-lithological control in the distribution of iron contents, as the main planes of the continuity ellipsoids are parallel to the S_2 schistosity of Córrego das Flechas synform, which comprises the deposit. The clustering of samples provided, in general, slightly higher accuracy and precision for ordinary kriging estimates in relation to the single-domain scenario, but. The clustering of K-means algorithm stands-out as the scenario that presented the best estimation results, verified by cross validation (higher linear correlation coefficient and smaller mean and variance of errors) and by comparison of its descriptive statistics with those of the dataset (smallest change in the iron grades mean and variance). The single-domain estimate caused the greatest smoothing effect by the overestimation and underestimation of low and high grades, respectively, which resulted in the model with the most homogeneous distribution of grades. For the Capanema mine, even though the small gains in accuracy and precision of estimation conferred by the definition of subdomains do not justify the time taken, the grouping methods can be usefully applied in contexts in which there is a need for greater detailing of the local contrasts of contents, as in the evaluation of heterogeneous substances and short-term models.

1. INTRODUÇÃO

A avaliação de recursos e reservas minerais é composta pela quantificação de volume, massa e teores de bens minerais, bem como na identificação das geometrias e extensões dos corpos mineralizados no espaço tridimensional (Yamamoto, 2001). Para isso, a avaliação se baseia na teoria de amostragem, que se refere ao ato de selecionar amostras de forma que estas representem o fenômeno espacial estudado, permitindo não só sua compreensão, como também a inferência de informações nos pontos não amostrados, o que é possibilitado pelo conceito de variável regionalizada (Matheron, 1963). Os teores de uma substância de interesse são tidos como uma variável regionalizada, pois apresentam simultaneamente uma propriedade estruturada, variando em função de sua posição no espaço; e uma propriedade aleatória, possuindo alguma imprevisibilidade natural (Sinclair & Blackwell, 2004).

Uma variável regionalizada pode ser considerada estacionária quando sua distribuição de frequência é constante em todo o domínio, condição raramente observada na natureza (Goovaerts, 1997). Alternativamente, para o contexto de recursos e reservas minerais, utiliza-se a decisão ou escolha de estacionariedade, também denominada como hipótese intrínseca ou de quasi-estacionariedade, que requer apenas que a variação espacial seja estacionária de segunda ordem em todo o domínio (Goovaerts, 1997; Sinclair & Blackwell, 2004). A definição de subdomínios pode ser fundamentada em um ou mais critérios, entre os quais se destacam os geológicos, estatísticos e técnico-operacionais, a depender das características da ocorrência mineral e as necessidades do projeto. Para realizar estimativas com métodos geoestatísticos, por exemplo, a divisão do domínio tem como principal finalidade satisfazer a hipótese intrínseca. Nesse contexto, uma série de algoritmos de análise de agrupamento (*cluster analysis*) tem sido testada e continuamente adaptada para realizar a definição de subdomínios com critérios multivariados (Faraj & Ortiz, 2021). Enquanto a maioria dos algoritmos foi elaborada para variáveis não regionalizadas, como o *K-means* e o Aglomerativo Hierárquico; também há aqueles que consideram as relações geoestatísticas entre as amostras, como o Aglomerativo Hierárquico Geoestatístico (Romary et al., 2015; Fouedjio, 2016; Moreira et al., 2020; Faraj & Ortiz, 2021).

Neste estudo, diferentes métodos de definição de subdomínios foram utilizados para a estimativa dos teores de ferro da mina Capanema, localizada entre os municípios de Santa Bárbara, Itabirito e Ouro Preto. Operada de 1982 a 2003 pela Minas da Serra Geral S.A. (MSG), a campanha utilizada consiste em 8.345 furos de perfuratriz do tipo *rockdrill* obtidos pela empresa, com classificação de litotipos e teores de ferro, sílica, fósforo e alumina, além da porcentagem de perda por calcinação (PPC). Os furos foram submetidos a um agrupamento tradicional, baseado nos litotipos, e à análise de agrupamento com os algoritmos *K-means*, Aglomerativo Hierárquico e Aglomerativo Hierárquico Geoestatístico; para comparar os efeitos da definição de subdomínios nas estimativas dos diferentes cenários.

2. METAS E OBJETIVOS

Este trabalho teve como objetivo a estimativa dos teores de ferro da mina Capanema, no Quadrilátero Ferrífero, com diferentes métodos de definição de subdomínios para estimativa, visando comparar e avaliar os efeitos de cada método nos resultados. O estudo também teve como metas comparar a eficiência dos algoritmos de análise de agrupamento (*cluster analysis*) para o banco de dados da mina e discutir seus potenciais usos.

3. LEVANTAMENTO BIBLIOGRÁFICO

3.1. Contexto Geológico Regional

A mina de ferro Capanema está localizada na porção centro-leste do Quadrilátero Ferrífero (QF) (Figura 1), Minas Gerais, borda sul do Cráton São Francisco.

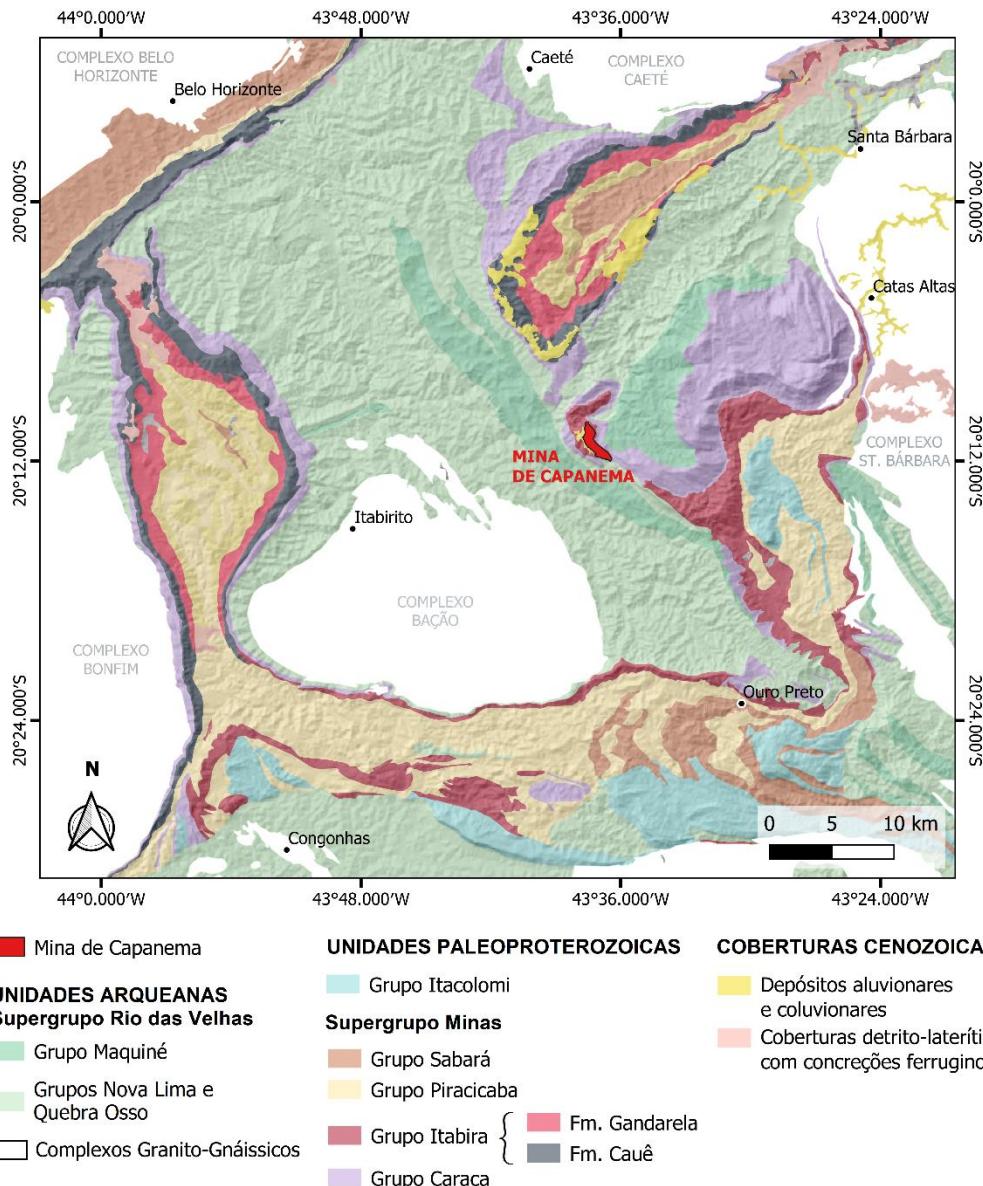


Figura 1: Mapa geológico do Quadrilátero Ferrífero e localização da mina Capanema. Modificado do banco de dados da CPRM (2004).

3.1.1. Litoestratigrafia do Quadrilátero Ferrífero

De acordo com Noce (1995) e Dardenne & Schobbenhaus (2001), a evolução do QF pode ser sintetizada em três eventos: o ciclo paleoarqueano Pré-Rio das Velhas, o ciclo neoarqueano Rio das Velhas e o ciclo paleoproterozoico Transamazônico.

Ciclo Paleoarqueano pré-Rio das Velhas

O primeiro ciclo compreende o retrabalhamento de crosta mais antiga em sucessivos eventos tectono-termais entre 2,85 e 3,30 Ga, o qual resultou na formação dos complexos tonalíticos-trondjemíticos-granodioríticos (TTG) do embasamento (Figura 2), expostos em estruturas dômicas (Noce, 1995).

Ciclo Neoarqueano Rio das Velhas

O ciclo neoarqueano se relaciona à formação do Supergrupo Rio das Velhas entre 2,78 e 2,70 Ga, uma sequência metavulcanossedimentar do tipo *greenstone belt* (Noce, 1995; Dardenne & Schobbenhaus, 2001). O Supergrupo Rio das Velhas é composto pelos grupos Quebra Osso, Nova Lima e Maquiné (Figura 2), resultantes da aglutinação de blocos continentais e fechamento oceânico. O Grupo Quebra Osso, unidade basal, é formado por rochas metakomatiáticas de fácies xisto verde com estruturas do tipo *pillow lava* e texturas *spinifex*, intercaladas com metassedimentos químicos e serpentinitos. O Grupo Nova Lima é constituído por rochas metassedimentares clásticas, químicas e vulcanoclásticas, além de metavulcânicas félsicas e máficas, que se associam a Formações Ferríferas Bandadas (FFBs) do tipo Algoma. Os filitos grafítosos deste grupo hospedam os principais depósitos auríferos do QF, do tipo orogênico, como os depósitos de classe mundial Morro Velho e Cuiabá (Dardenne & Schobbenhaus, 2001). O Grupo Maquiné, porção superior do *greenstone belt*, corresponde a uma sequência metassedimentar de conglomerados, quartzitos, filitos e xistos.

Ciclo Transamazônico

O Ciclo Transamazônico, de 2,6 a 2,0 Ga, consiste de uma série de processos sedimentares, tectônicos e magmáticos que resultaram na formação do Supergrupo Minas, em discordância com o Supergrupo Rio das Velhas e os complexos TTG (Noce, 1995). Da base para o topo, o Supergrupo Minas é composto pelas sequências fluviais deltaicas e marinhas plataformais dos grupos Caraça, Itabira e Piracicaba, pelo grupo marinho imaturo Sabará (Rosière & Chemale Jr, 2000), conforme indicado na coluna estratigráfica da Figura 2. Além do início da sedimentação da Bacia Minas, o Grupo Caraça registra a gradação dos metaconglomerados e quartzitos da Formação Moeda para os metapelitos da Formação Batatal (Dardenne & Schobbenhaus, 2001). Em contato transicional com o Grupo Caraça, o Grupo Itabira (2,52 a 2,42 Ga) corresponde ao maior horizonte de sedimentos químicos do

Supergrupo Minas e é subdividido nas formações Cauê e Gandarela, que se intercalam horizontal e verticalmente (Rosière & Chemale Jr, 2000).

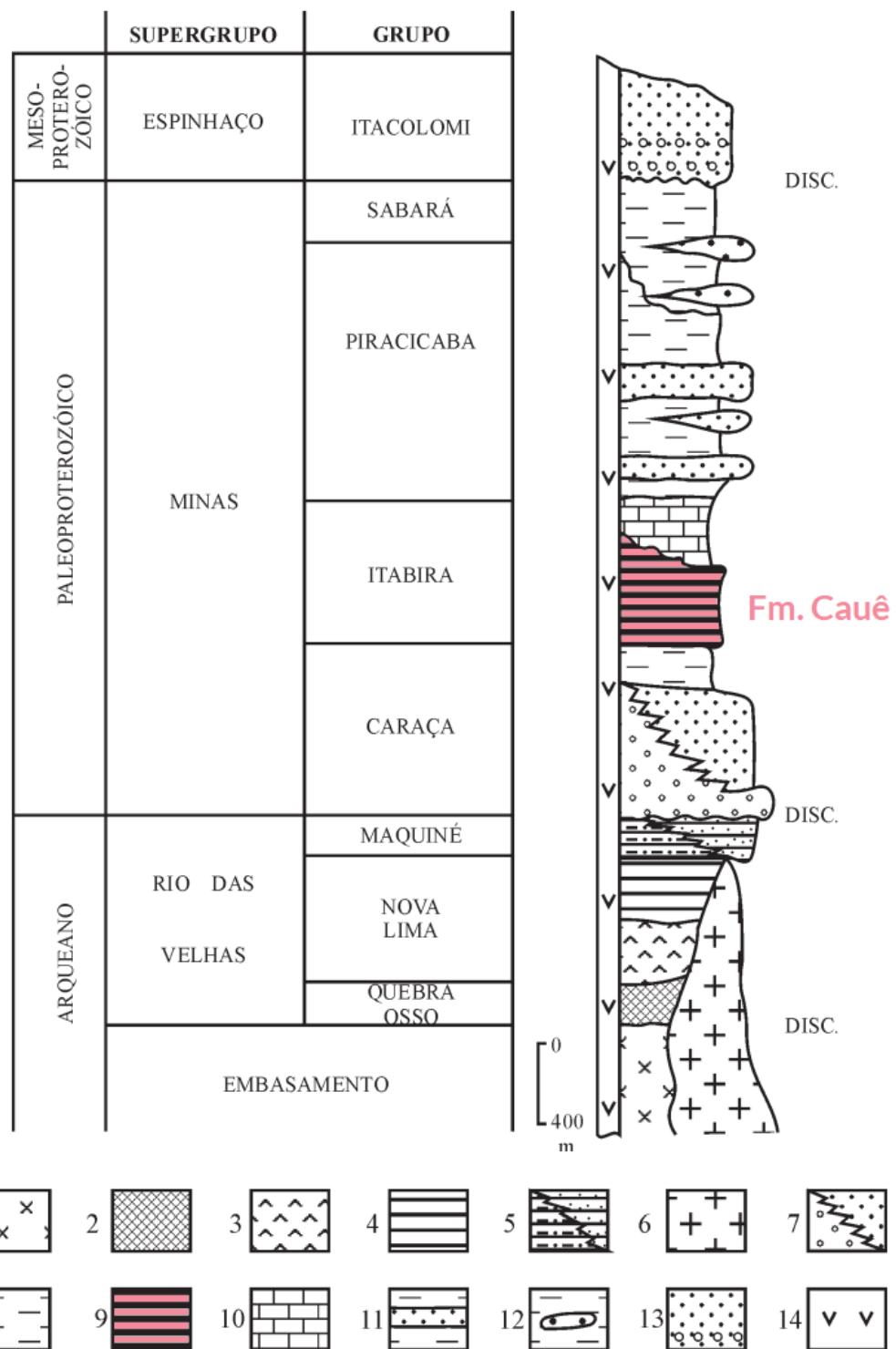


Figura 2: Coluna estratigráfica do Quadrilátero Ferrífero. (1) Complexos granito-gnáissicos do embasamento, (2) Rochas ultramáficas, (3) Rochas máficas, (4) Unidade químico-clástica, (5) Unidade clástica, (6) Tonalitos, (7) Conglomerados e quartzitos da Fm. Moeda, (8) Filitos carbonosos da Fm. Batatal, (9) Itabiritos da Fm. Cauê, (10) Dolomitos da Fm. Gandarela, (11) Quartzitos e filitos, (12) Conglomerados e filitos, (13) Conglomerados e quartzitos, (14) Sills e diques de diabásio. (Dardenne & Schobbenhaus, 2001)

A Formação Cauê é caracterizada por FFBs metamórficas, deformadas e oxidadas denominadas itabiritos, além de filitos e dolomitos ferruginosos. De origem sedimentar (tipo

Lago Superior), os itabiritos contêm de 250 a 300 m de espessura e compõem os principais depósitos de ferro de alto teor do Quadrilátero Ferrífero, como as minas Águas Claras, Alegria e Capanema (Dardenne & Schobbenhaus, 2001). Rosière & Chemale Jr (2000) classificam os itabiritos em três tipos compostionais, todos contendo hematita e martita como fases dominantes e com kenomagnetita subordinada. Os tipos são **(a)** quartzo-itabirito, o mais comum do QF, que possui o quartzo como um de seus componentes mineralógicos principais, tanto nos níveis claros do microbandamento quanto em veios sacaroidais; **(b)** itabiritos dolomíticos, os quais apresentam bandas de Fe-dolomita, quartzo e óxidos de ferro, com ou sem calcita, clorita e anfibólio como acessórios; e **(c)** os itabiritos anfibolíticos, que apresentam anfibólios de composição variada, como grunerita-cummingtonita, tremolita e actinolita.

A Formação Gandarela é composta por metadolomitos com itabiritos e filitos subordinados e tem seu topo marcado por discordância erosiva com o Grupo Piracicaba (Figura 2), predominantemente clástico. Os sedimentos do Grupo Piracicaba representam a alternância de ciclos transgressivos em um período de expansão oceânica (Noce, 1995; Dardenne & Schobbenhaus, 2001) e também apresentam discordância regional com a unidade sobreposta, o Grupo Sabará. Topo do Supergrupo Minas, o Grupo Sabará marca a colisão com o arco Transamazônico, tratando-se de uma sequência do tipo *flysch* (pré a sin-orogênica) de 2,12 a 2,03 Ga (Noce, 1995).

Após um período de estabilidade tectônica, inicia-se uma fase extensional por volta de 1,75 Ga (rifte Espinhaço). O Grupo Itacolomi e o Supergrupo Espinhaço correspondem a depósitos continentais correlatos a vulcanismo ácido continental e intrusões graníticas anorogênicas (Dardenne & Schobbenhaus, 2001).

3.1.2. Sinclinal Ouro Fino

A mina Capanema encontra-se na porção sul do Sinclinal Ouro Fino (Figura 3), cujo modelo mais recente foi proposto por Franco & Endo (2004). De acordo com os autores, o sinclinal corresponde a uma dobra antiformal, reclinada, de eixo WNW-ESE sub-horizontal e vergência para sul, desconectada do Sinclinal Gandarela pela Falha do Fundão.

O Sinforme Córrego das Flechas apontado no perfil da Figura 3 é a estrutura dominante da mina e corresponde à segunda fase deformacional descrita pelos autores, na qual foram geradas dobras de xistosidade S2 de direção NW-SE e cimento moderado para NE, posteriormente sobrepostas pelos eventos D3 e D4.

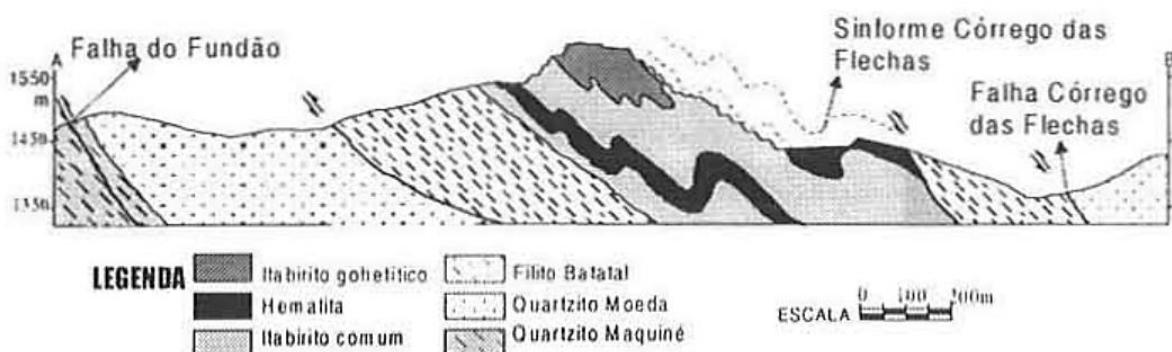
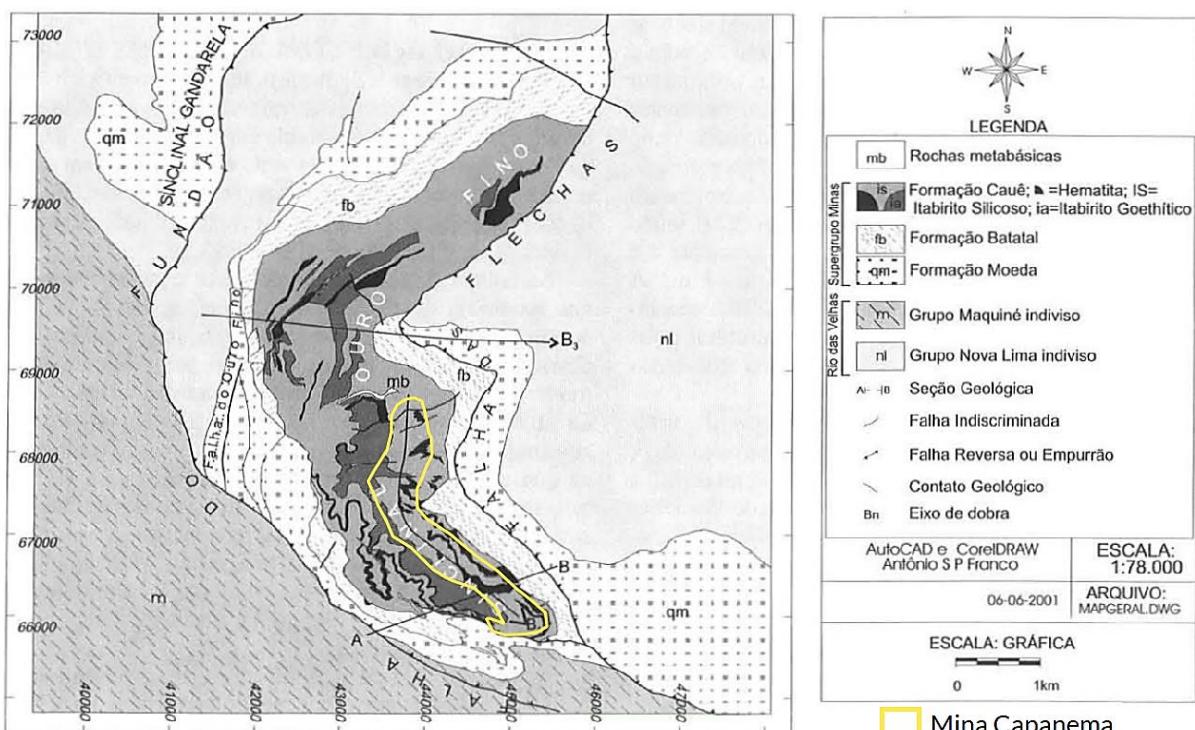


Figura 3: Mapa geológico do Sinclinal Ouro Fino e perfil geológico A-B seccionando a cava atual da mina Capanema. (Franco & Endo, 2004)

3.1.3. Aspectos econômicos da mina Capanema

Após sua descoberta em meados da década de 60, a mina Capanema foi operada de 1982 a 2003 pela Minas da Serra Geral S.A. (MSG), uma *joint venture* da Companhia Vale do Rio Doce (CVRD) em associação com empresas japonesas, como a Kawasaki Steel Corporation. A maior parte da produção da MSG (cerca de 10,5 Mtpa) era adquirida pela CVRD, a qual realizava o seu beneficiamento na mina de Timopeba, a 11 km de distância, e a comercialização do minério (Vale, 2012).

Em dezembro de 2020, a CVRD obteve as licenças para retomar as operações da mina com o Projeto Capanema Umidade Natural e, em novembro do mesmo ano, começou a comercialização do minério que havia sido extraído e ainda se encontrava estocado na mina desde sua paralisação em 2003 (Notícias Vale, 2021). Estima-se que a produção se inicie no segundo semestre de 2023, com capacidade média anual de 18 Mt e vida útil de 7 anos. Nos primeiros 5 anos, no entanto, essa capacidade será de 14 Mtpa, pois a operação se

concentrará no reaproveitamento da pilha de estéril denominada WH, de hematita alterada enriquecida em alumina e fósforo, sem a geração de rejeito (Formulário de Referência Vale, 2021).

Em 31 de dezembro de 2020, as reservas totais de minério de ferro do Complexo Mariana (no qual a mina Capanema está inserida) correspondiam a 1.810 Mt a um teor médio de 46,7% (reservas provadas de 727Mt @ 45% e prováveis de 1.083 Mt @ 47,9%) (Relatório Anual Vale, 2020). Os recursos obtidos por Rocha (1999) para a mina Capanema correspondem a cerca de 139 Mt a um teor médio de 62,30%, levando-se em consideração os teores médios de corte da mina. Na Tabela 1, são apresentados os teores de corte para o minério e para o *Run of Mine* (ROM) por faixa granulométrica. O ROM do Complexo Mariana é processado por britagem padrão, classificação e concentração para produção de *sinter feed*, minério granulado e *pellet feed* nas usinas de beneficiamento (Relatório Anual Vale, 2020).

Tabela 1: Teores de corte máximos e mínimos para o minério e para o ROM da mina Capanema por faixa granulométrica. (Rocha, 1999)

	Faixa granulométrica	Fe (%)	Al ₂ O ₃	SiO ₂ (%)	P (%)	PPC (%)
	<0,15 mm	62,50	*	*	*	*
Min.	>0,15 mm e <6,3	60,20	*	*	*	*
	>6,3	56,00	*	*	*	*
	<0,15 mm	*	2,90	3,30	0,10	*
Máx.	>0,15 mm e <6,3	*	1,50	8,40	0,09	*
	>6,3	*	1,20	16,0	0,055	*
	Min.	59,80	1,10	7,80	0,065	3,00
ROM						
	Máx.	60,50	1,90	9,70	0,085	3,80

3.2. Análise de Agrupamento (*Cluster Analysis*)

Análise de agrupamento consiste num conjunto de técnicas que visam distribuir as observações em grupos (*clusters*) de modo que a semelhança entre elas seja a maior possível dentro de cada grupo e a menor entre grupos. Entre os métodos mais utilizados, podem-se citar os algoritmos não supervisionados *K-means* (MacQueen, 1967) e Aglomerativo Hierárquico (Sokal & Sneath, 1963). Para ambos os métodos, a escolha das variáveis é decisiva e deve-se favorecer as que tenham alto poder de discriminação e sejam relevantes ao problema (James et al., 2013).

3.2.1. Algoritmo K-means

Na primeira iteração, o algoritmo distribui as amostras de maneira aleatória entre uma quantidade *k* de grupos predefinida pelo usuário (Figura 4a) e calcula as médias (centroïdes)

de cada grupo (Figura 4b). As amostras são então redistribuídas de modo que a soma do quadrado das distâncias entre amostras e centroides seja mínima (*within-clusters-sum-of-squares*) (Figura 4c) (James et al., 2013; Kassambara, 2017).

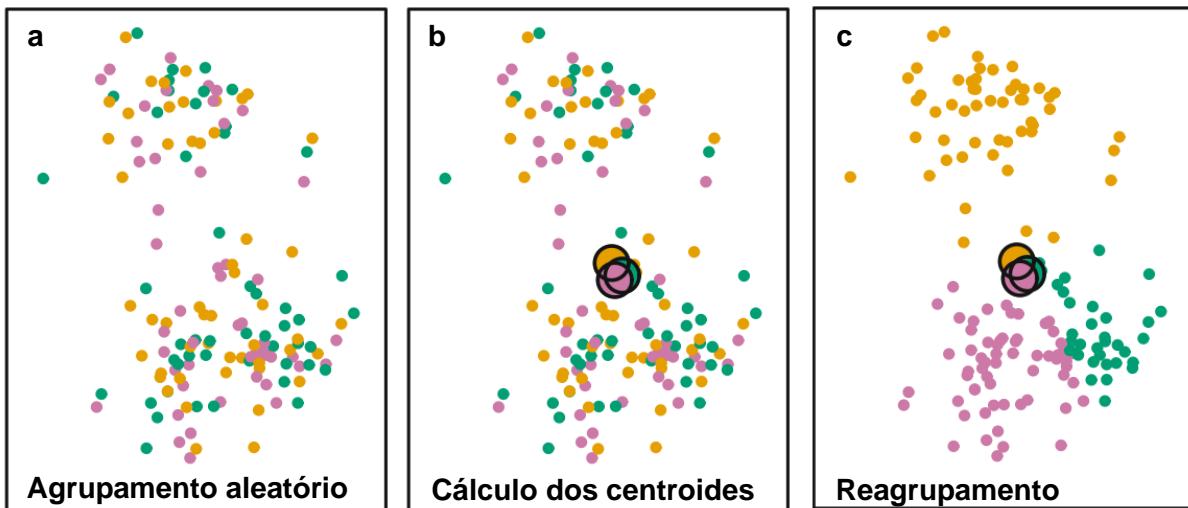


Figura 4: Ilustração da primeira iteração do algoritmo *K-means*. (a) Distribuição aleatória das amostras em 3 grupos, (b) cálculo dos centroides dos agrupamentos e (c) redistribuição das amostras de acordo com a posição dos centroides. (James et al., 2013)

Em seguida, a posição dos centroides é recalculada (Figura 5a) e as amostras são novamente redistribuídas para que pertençam ao *cluster* de centroide mais próximo. Essa nova inclusão e remoção de amostras nos *clusters* alteram suas médias, alterando também a posição dos centroides. Este processo de cálculo dos centroides e redistribuição das amostras se repete até que as médias dos *clusters* se estabilizem (Figura 5b).

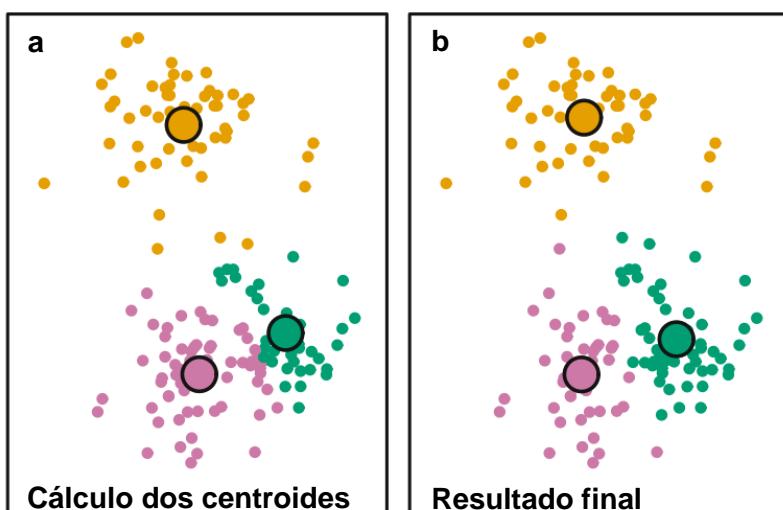


Figura 5: Ilustração de mais iterações do algoritmo *K-means*. (a) Os centroides são reposicionados de acordo com as amostras a eles atribuídas na primeira iteração e (b) o processo se repete n vezes até atingir a estabilização dos centroides. (James et al., 2013)

Além da quantidade k predefinida de *clusters*, o agrupamento realizado pelo *K-means* também depende da distribuição aleatória da primeira iteração (James et al., 2013; Kassambara, 2017). A cada execução do algoritmo, portanto, o resultado final obtido é

diferente. Na literatura, recomenda-se executá-lo mais de uma vez e observar qual padrão de agrupamento é mais frequente (James et al., 2013) e, ainda, há soluções que buscam otimizar os agrupamentos e retornar resultados mais consistentes, como o *K-means++* de Arthur & Vassilvitskii (2007).

Um dos métodos para a escolha da quantidade de agrupamentos é denominado método do cotovelo e consiste na análise da soma total do quadrado das distâncias (inércia) pela quantidade de *clusters* (Figura 6). Os pontos de inflexão (*elbow points*) do gráfico podem ser considerados como quantidades adequadas de *clusters*, pois evitam tanto o agrupamento de amostras muito distintas em poucos grupos, quanto o agrupamento de poucas amostras em muitos grupos (Kassambara, 2017; Yuan & Yang, 2019).

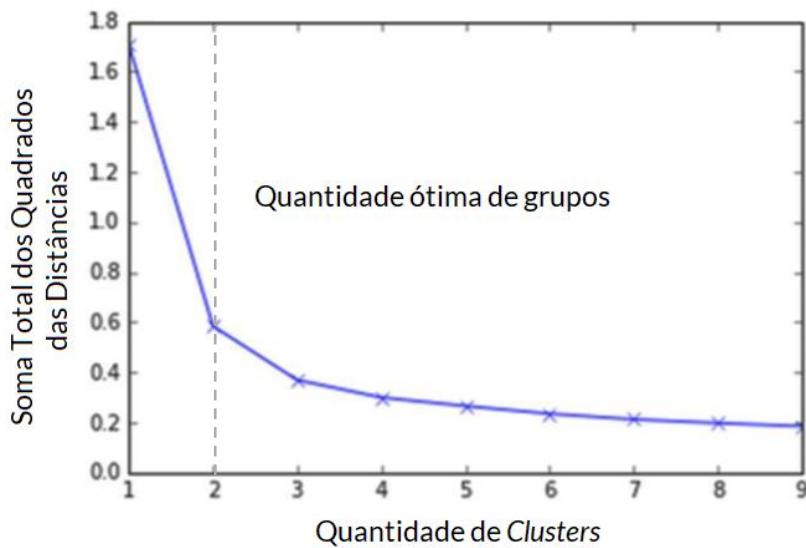


Figura 6: Exemplo de gráfico da inércia total pela quantidade de *clusters*. (Yuan & Yang, 2019)

3.2.2. Algoritmo Aglomerativo Hierárquico

O agrupamento hierárquico pode ser subdividido nos tipos aglomerativo e divisivo, que se caracterizam pela estruturação dos dados em uma hierarquia multiníveis, geralmente representada por dendrogramas (Figura 7) (Kassambara, 2017). Para isso, cada observação é tratada como um *cluster* e os *clusters* são sucessivamente unidos em pares até que haja um único grupo (raiz). Os algoritmos aglomerativos possuem diferentes métodos para determinar a semelhança entre os pares de *clusters*, como o método Ward, um dos mais utilizados. O método Ward assume que cada grupo é representado por um centroide, realizando a união dos pares com o objetivo de minimizar a variância desses centroides, de modo semelhante ao *K-means* (Kassambara, 2017).

Duas vantagens dos métodos hierárquicos sobre o *K-means* são: dispensar a predefinição da quantidade de *clusters* e retornar o mesmo resultado sempre que executado para determinados dados. Para obter k grupos, o usuário deve escolher o nível hierárquico

correspondente, conforme ilustrado na Figura 7b. Geralmente, o nível hierárquico (ou altura) dos dendrogramas corresponde a uma medida relativa da proximidade entre os grupos ou, como no caso da Figura 7b, à distância euclidiana entre os pares no espaço multivariado.

O *K-means*, por sua vez, é menos sensível a *outliers* e sua evolução é dinâmica, ou seja, um ponto inicialmente inserido em um *cluster* pode ser atribuído a outro ao longo das n iterações (James et al., 2013; Kassambara, 2017).

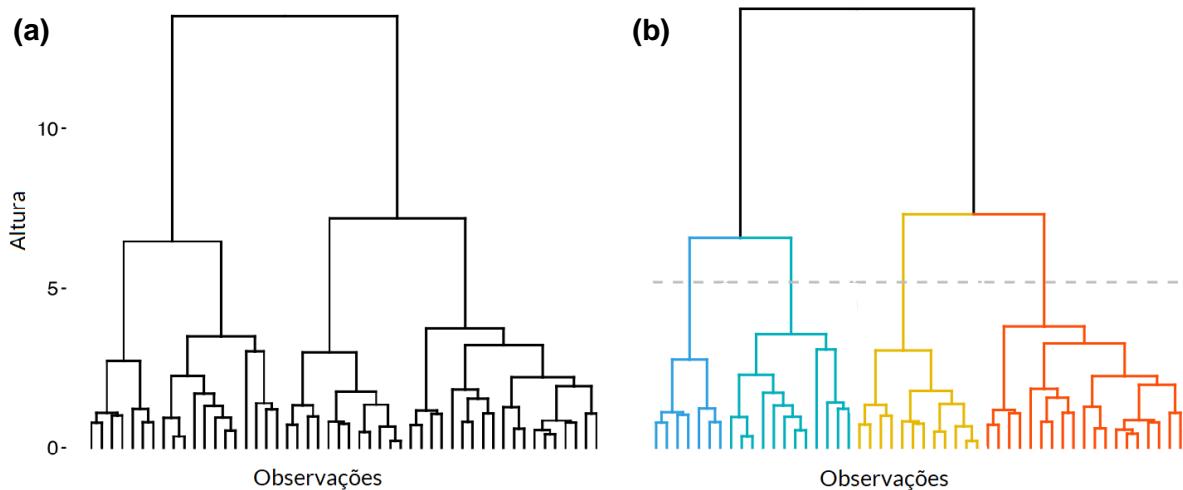


Figura 7: Representação de (a) dados estruturados em uma hierarquia multiníveis (dendrograma) e (b) seleção de um nível hierárquico para gerar quatro *clusters* distintos. (Kassambara, 2017)

Para que os algoritmos *K-means* e Aglomerativo Hierárquico considerem a posição espacial de amostras regionalizadas, é possível inserir as coordenadas geográficas como variáveis adicionais. Entretanto, neste caso, os algoritmos não distinguem o espaço geográfico do multivariado e, por isso, frequentemente desconsideram ou a continuidade espacial ou a consistência estatística de cada domínio, honrando apenas um dos dois aspectos (Romary et al., 2015; Fouedjio, 2016). Autores como Romary et al. (2015) e Fouedjio (2016) propuseram metodologias para incluir a dependência espacial de variáveis regionalizadas no processo de agrupamento com algoritmos não supervisionados. O método Aglomerativo Hierárquico Geoestatístico, elaborado por Romary et al. (2015), é do tipo espacial por considerar, além da proximidade multivariada entre os pontos, a proximidade geográfica pelo uso de gráficos de amostragem.

3.3. Estimativa e Validação dos Resultados

3.3.1. Geoestatística

A hipótese intrínseca implica que a distribuição espacial de uma variável regionalizada dentro do campo geométrico pode ser descrita por uma função (1) que depende apenas da distância e orientação entre os pontos (Goovaerts, 1997):

$$2\gamma(h) = \frac{1}{n} \sum_{i=1}^n [Z(x + h) - Z(x)]^2 \quad (1)$$

sendo $2\gamma(h)$ a função variograma; n a quantidade de pares de pontos separados por uma distância h , $Z(x)$ o valor da variável regionalizada no ponto x e $Z(x+h)$ o valor da mesma variável no ponto $(x+h)$ (Yamamoto & Landim, 2013). Os resultados do cálculo da função variograma, que é uma função discreta, podem ser observados em um gráfico de dispersão que relaciona a variância espacial entre pares de pontos com as distâncias de separação entre eles, denominado variograma experimental. Os variogramas experimentais não são utilizados nas etapas subsequentes, pois a função variograma deve ser contínua, com valores de variância espacial para todas as direções e distâncias. Por isso, deve-se lançar mão do ajuste de um modelo teórico (modelo teórico de variograma) com funções positivas e monotônicas crescentes, como o esférico, o exponencial e o gaussiano, que são os mais comuns (Rossi & Deutsch, 2014). As propriedades de um modelo teórico de variograma descrevem o comportamento da variável regionalizada em determinada orientação e são (Figura 8):

- i. Alcance ou amplitude (a): distância a partir da qual as amostras passam a ser independentes e separam o campo estruturado (em que há correlação espacial) do campo aleatório (em que as amostras são espacialmente independentes);
- ii. Efeito pepita (C_0): valor da função variograma próximo a origem, comumente diferente de zero devido à amostragem, aos processos analíticos e a variabilidade natural da mineralização;
- iii. Patamar ($C_0 + C_1$): valor da variância em que o variograma se estabiliza ou perde a estrutura (variância espacial máxima), que caracteriza o campo aleatório;
- iv. Variância espacial (C_1): patamar subtraído o efeito pepita.

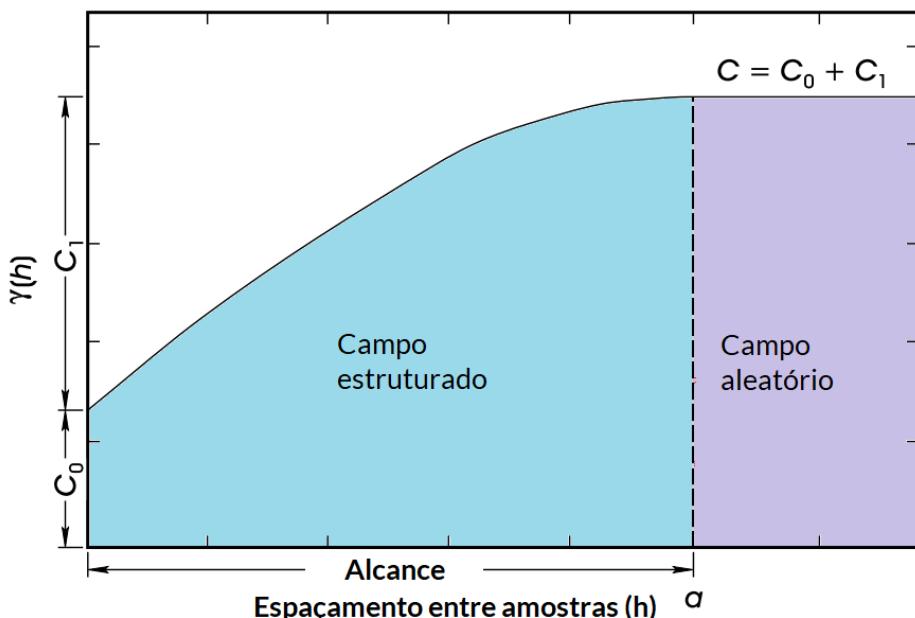


Figura 8: Propriedades de um modelo de semivariograma. (Sinclair & Blackwell, 2004)

Tratando-se de uma função direcional, as propriedades de um variograma podem mudar de acordo com a direção analisada e, neste caso, configuram um fenômeno anisotrópico. A anisotropia pode ser classificada como geométrica (diferentes amplitudes), zonal (diferentes patamares) ou mista (diferentes amplitudes e patamares); e representada como uma elipsoide de continuidade a partir da análise estrutural de três direções: a de maior alcance (D1), a de menor alcance ortogonal a D1 (D2) e a ortogonal a D1 e D2 (Yamamoto & Landim, 2013).

3.3.2. Krigagem ordinária (KO)

A krigagem ordinária é considerada um melhor estimador linear não enviesado (*best linear unbiased estimator* - BLUE), pois a esperança matemática dos erros é nula (não enviesado) e a determinação dos pesos de estimativa é realizada para minimizar a variância desses erros (Yamamoto & Landim, 2013). A krigagem ordinária também é denominada como um algoritmo não-estacionário, pois não assume que a média global é invariante para todo o depósito como a krigagem simples, e considera que as médias (desconhecidas) são invariantes apenas localmente. Para isso, e para que a ponderação não seja enviesada (soma dos pesos igual a um), o método utiliza o multiplicador de Lagrange (μ) como incógnita (Rossi & Deutsch, 2014). Os pesos são então determinados pela resolução de um sistema com $n+1$ equações e $n+1$ pesos, que pode ser apresentado na forma matricial (Sinclair & Blackwell, 2004; Yamamoto & Landim, 2013):

$$\begin{bmatrix} \gamma(s_1, s_1) & \gamma(s_1, s_2) & \gamma(s_1, s_n) & 1 \\ \gamma(s_2, s_1) & \gamma(s_2, s_2) & \gamma(s_2, s_n) & 1 \\ \gamma(s_n, s_1) & \gamma(s_n, s_2) & \gamma(s_n, s_n) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(s_1, s_0) \\ \gamma(s_2, s_0) \\ \gamma(s_n, s_0) \\ 1 \end{bmatrix} \quad (2)$$

Matriz C *Matriz D*

em que $\gamma(s_i, s_j)$ são as variâncias espaciais entre as amostras estimadoras, w_i são os pesos a serem determinados e $\gamma(s_i, s_0)$ são as variâncias espaciais entre cada ponto amostrado e o ponto estimado s_0 . Desse modo, a Matriz D é considerada a matriz de distância estatística, pois quanto menor for a variância espacial entre a amostra estimadora e o ponto estimado, maior será o peso atribuído a ela (Pyrcz & Deutsch, 2014). Por sua vez, a Matriz C é denominada matriz de desagrupamento, pois quanto maior a proximidade estatística entre duas amostras estimadoras, menor é a variância e, consequentemente, menores os pesos atribuídos a elas. Esta propriedade é capaz de gerar pesos negativos (efeito tela), que minimizam os efeitos de redundância dos dados, gerada pelo adensamento da malha e o alinhamento de amostras em uma mesma direção. Entretanto, pesos negativos também podem apresentar efeitos indesejáveis, como (Sinclair & Blackwell, 2004):

- i. Superestimativa, quando uma amostra de baixo teor recebe peso negativo;
- ii. Subestimativa, quando uma amostra de alto teor recebe peso negativo;
- iii. Estimativa de teores fisicamente impossíveis (negativos), quando um *outlier*

recebe peso negativo, por menor que seja o peso.

Na ocorrência desses efeitos, sobretudo da estimativa de teores fisicamente impossíveis, é possível utilizar diferentes métodos de correção dos pesos, como igualar todos os pesos negativos a zero e reajustar os positivos proporcionalmente, mantendo-se o não enviesamento.

3.3.3. Métodos de validação de estimativa

Dentre os métodos amplamente utilizados pode-se citar a comparação entre as estatísticas descritivas das amostras e do modelo estimado, a validação cruzada e a análise de deriva.

A comparação entre as estatísticas descritivas consiste em um método de validação global das estimativas e deve considerar o efeito de suavização, próprio de estimativas baseadas em médias ponderadas, conforme mostra a Figura 9b com a redução da variância. De modo geral, busca-se observar quão diferentes são a distribuição amostral e a de teores estimados através da comparação dos histogramas e das medidas de tendência central e dispersão, além da elaboração de gráficos de dispersão como Quantil-Quantil ou Percentil-Percentil (Abzalov, 2016).

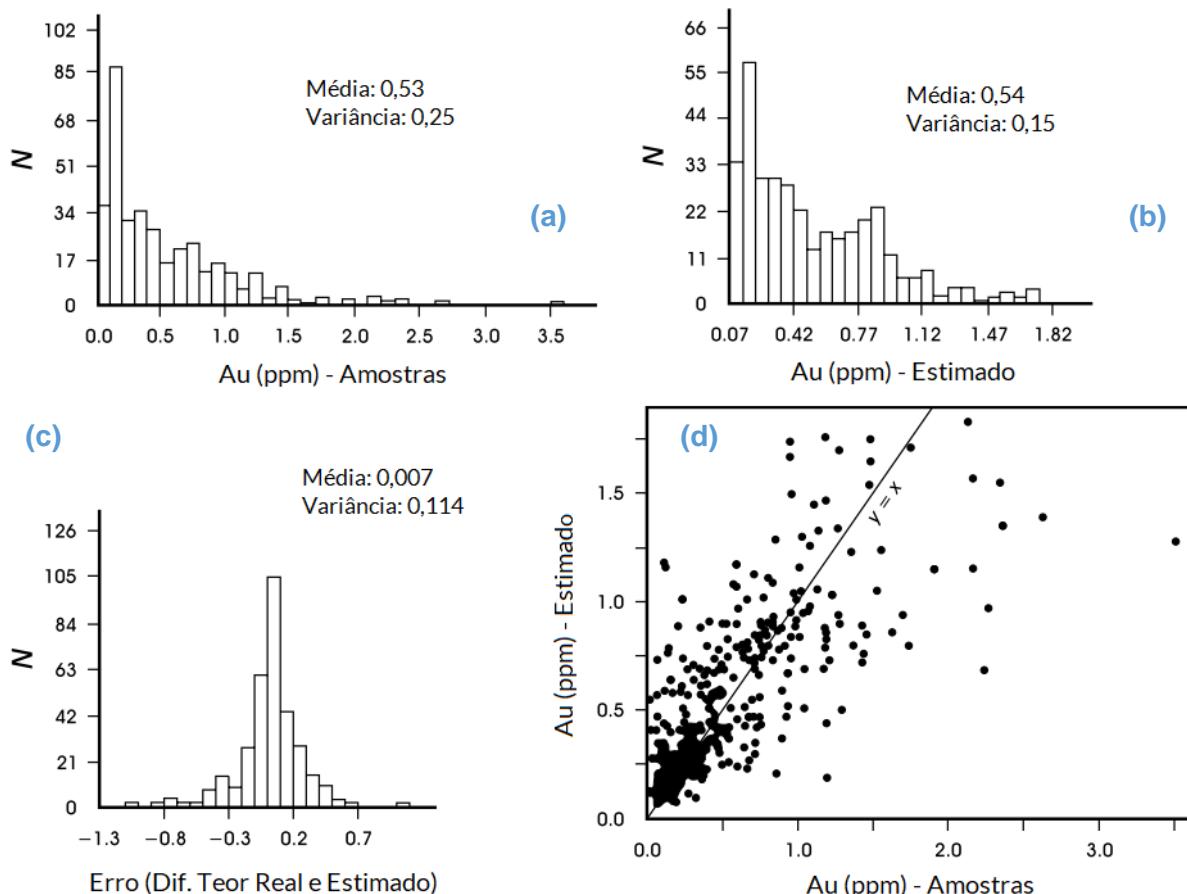


Figura 9: Histogramas **(a)** dos teores das amostras, **(b)** dos teores estimados na validação cruzada e **(c)** das diferenças entre os valores reais e estimados; e **(d)** gráfico de dispersão entre valores reais e estimados. (Sinclair & Blackwell, 2004)

Na validação cruzada, um método de validação local, executa-se a extração de uma amostra do banco de dados e sua estimativa pontual pela interpolação das amostras remanescentes na vizinhança. Este processo é repetido sucessivamente para as n amostras do banco de dados com determinados parâmetros de vizinhança e, para os métodos geoestatísticos, modelos teóricos de variograma; o que permite o confronto entre os valores reais das amostras extraídas e suas estimativas pontuais. Quanto maior a correlação linear positiva entre o cenário real e o cenário estimado, melhor é considerada a eficácia dos parâmetros de vizinhança e modelos teóricos de variograma utilizados (Figura 9d). Além da dispersão entre os valores reais e as estimativas pontuais, observa-se também a distribuição do erro (diferença entre eles), cuja média desejada é zero. O deslocamento da média do erro para a direita ou para a esquerda indica que houve superestimativa sistemática ou subestimativa sistemática, respectivamente (Figura 9c), refletindo a acurácia da estimativa, enquanto a variância do erro representa a sua precisão (Sinclair & Blackwell, 2004; Abzalov, 2016).

Análises de deriva consistem na subdivisão do depósito em fatias de determinada espessura ao longo de um eixo de referência para a visualização gráfica das médias móveis das amostras e do modelo estimado (Abzalov, 2016). Através do gráfico, denominado gráfico em fatias, *swath plot* ou gráfico em teia (Figura 10), observa-se a aderência entre as médias móveis, que reflete a acurácia da interpolação realizada e aponta em quais fatias do depósito houve subestimativa ou superestimativa, tratando-se de um método de validação local.

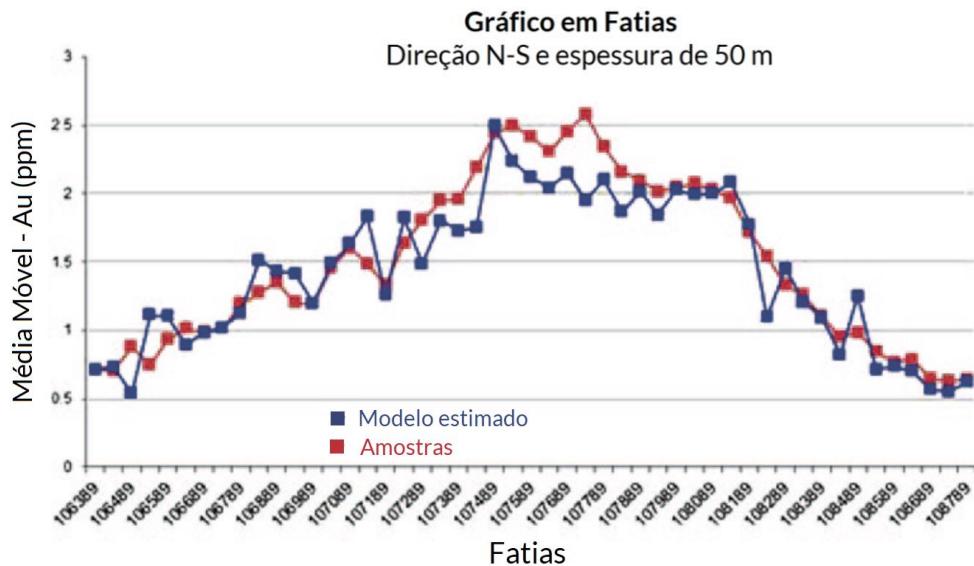


Figura 10: *Swath plot* ao longo da direção E-W, com fatias N-S de espessura igual a 50 metros. (Rossi & Deutsch, 2014)

4. MATERIAIS E MÉTODOS

4.1. Materiais

4.1.1. Banco de dados

O banco de dados consiste em uma única planilha que abrange um conjunto de 8.345 furos verticais de perfuratriz tipo *rockdrill*. Como cada furo corresponde a uma amostra de 13 metros, são no total 108.485 metros amostrados com informações de litotipos, PPC e teores de ferro, alumina, fósforo e sílica (Tabela 2).

Tabela 2: Síntese das informações do banco de dados.

Campo	Informação
<i>FILENAME</i>	Nome do arquivo de origem (não acessado)
X	Coordenada em X
Y	Coordenada em Y
Z	Coordenada em Z (centro do furo)
<i>BHID</i>	Identificação dos furos
<i>FROM</i>	Início do intervalo (m)
<i>TO</i>	Final do intervalo (m)
<i>BLOCO</i>	Identificação do bloco de lavra
<i>LITH</i>	Código de litotipo
<i>Fe</i>	Teor de ferro (%): Fe ₂ O ₃
<i>SiO₂</i>	Teor de sílica (%)
<i>Al₂O₃</i>	Teor de alumina (%)
<i>P</i>	Teor de fósforo (%)
<i>PPC</i>	Perda por calcinação (%)
<i>LENGTH</i>	Comprimento das amostras (m)
<i>RADIUS</i>	(Campo formacional para softwares)
<i>A0</i>	Azimute dos furos
<i>B0</i>	Mergulho dos furos
<i>C0</i>	Desvio dos furos

Os furos foram obtidos de maneira sistemática, ou seja, a malha estudada é regular. Horizontalmente, a malha é rotacionada para a direção de 135°. A distância entre amostras é de 28,28 metros nas direções NE-SW e NW-SE para todos os pontos, que definem uma área amostral de aproximadamente 1,5 km² (Figura 11). Verticalmente, as amostras estão distribuídas entre 18 bancadas, que distam 13 metros uma da outra (Figura 12).

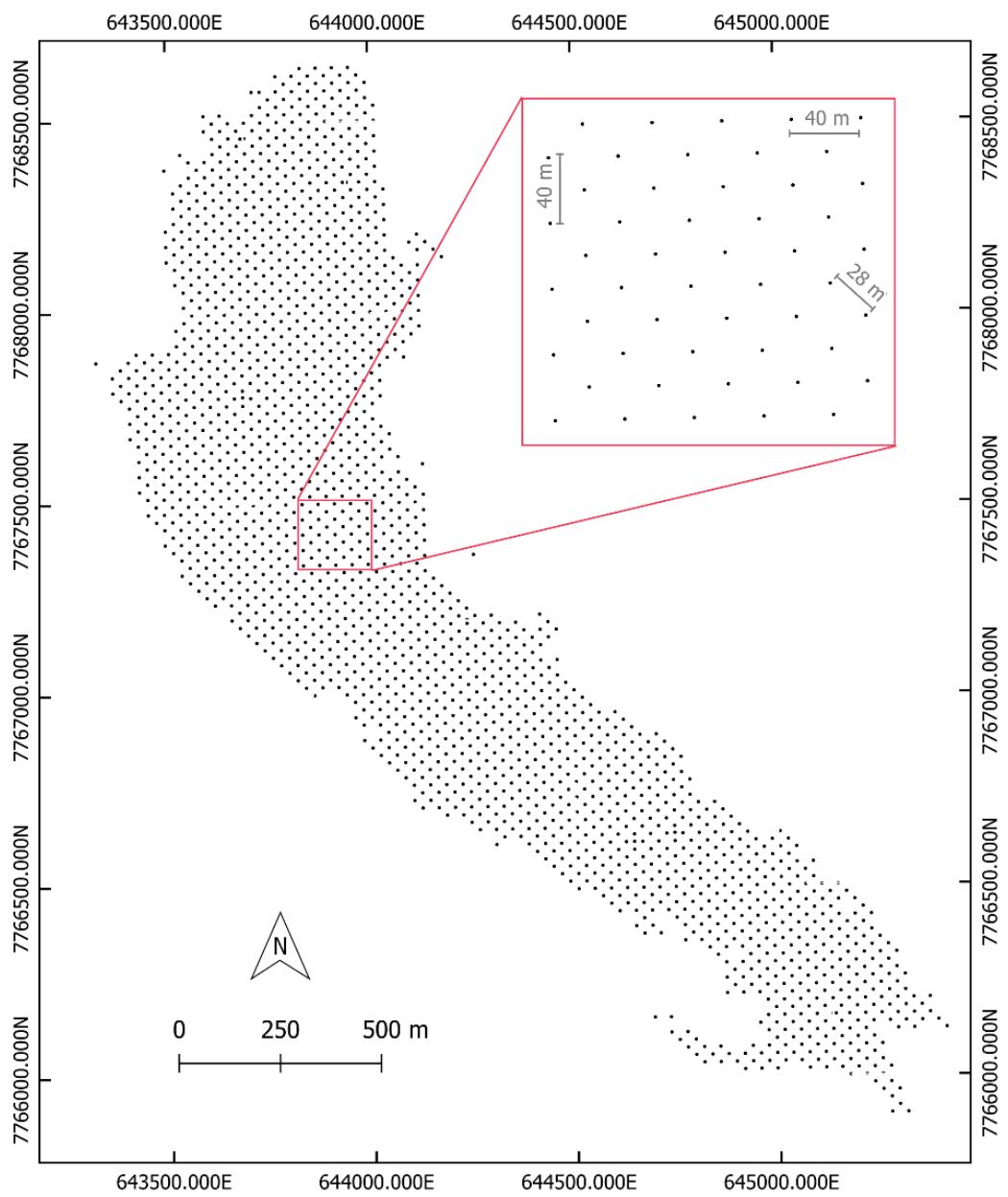


Figura 11: Malha dos furos de perfuratriz em planta.

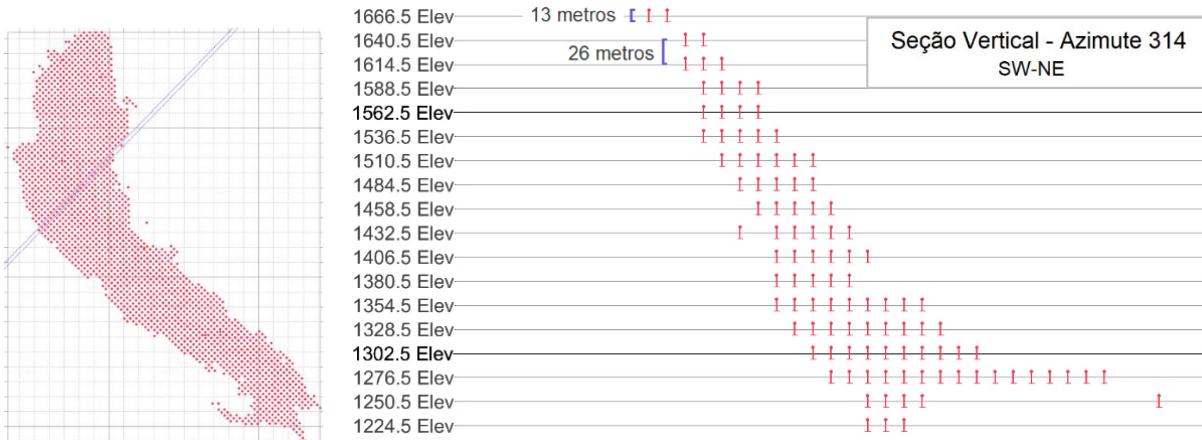


Figura 12: Seção vertical SW-NE dos furos de perfuratriz.

4.1.2. Infraestrutura de informática

Foram utilizados os softwares comerciais Studio RM® (Datamine Software), Supervisor® (Datamine Software) e Isatis.neo Mining® (Geovariances), bem como programação em linguagem Python com o uso do software livre Jupyter Notebook e as bibliotecas Sklearn (Pedregosa et al., 2011) e Scipy (Virtanen et al., 2020).

4.2. Métodos

4.2.1. Levantamento bibliográfico

O levantamento e o estudo de dissertações, teses, artigos, livros e demais obras acerca dos temas pertinentes permitiram a maior compreensão da área de estudo, técnicas empregadas e resultados obtidos.

4.2.2. Validação e análise exploratória dos dados

A validação do banco de dados consistiu na busca por valores incongruentes e sua correção, quando necessária. Como parte da análise exploratória dos dados (AED), foram realizadas as análises estatísticas uni e bivariada dos dados.

4.2.3. Definições de subdomínios

As amostras foram submetidas à definição de subdomínios com base na classificação original dos litotipos da mina em minério ou estéril e a análise de agrupamento com os algoritmos *K-means* (KM), Aglomerativo Hierárquico (HC) e Aglomerativo Hierárquico Geoestatístico (GHC).

4.2.4. Modelo de blocos, análise geoestatística e estimativa

Os blocos do modelo foram classificados em estéril, minério ou de transição, a depender da classificação das amostras no entorno, para realizar as estimativas por subdomínios. Após a análise geoestatística de todos os cenários, as estimativas foram realizadas por krigagem ordinária (KO).

4.2.5. Validação da estimativa e comparação entre os cenários

Para avaliar e comparar os agrupamentos, foram utilizadas legendas de teores e análises de contato, além da obtenção de *boxplots* por subdomínio em cada cenário. Métodos de validação locais e globais de estimativa foram aplicados também para avaliar os efeitos de cada cenário nas interpolações. Entre eles: a comparação das estatísticas descritivas do conjunto amostral e dos modelos estimados, a validação cruzada para obtenção dos erros e coeficientes de correlação linear em cada cenário, a análise de deriva e a geração de gráficos quantil-quantil (Q-Q plots).

5. RESULTADOS OBTIDOS

5.1. Análise Exploratória dos Dados

5.1.1. Litotipos da mina Capanema

Os litotipos presentes no campo LITH são HA, HB, SI, WH, SIA e SIP (Figura 13), conforme classificação realizada pela MSG e registrada por Massahud (1996 *apud* Rocha, 1999) em relatório interno, a partir dos critérios: cor, textura, compactação e teores de ferro, alumina, fósforo e sílica. Desses, são considerados litotipos de minério HA, HB e SI, em contraste aos demais litotipos, com baixo teor de ferro (~50%) e/ou alta concentração de contaminantes.

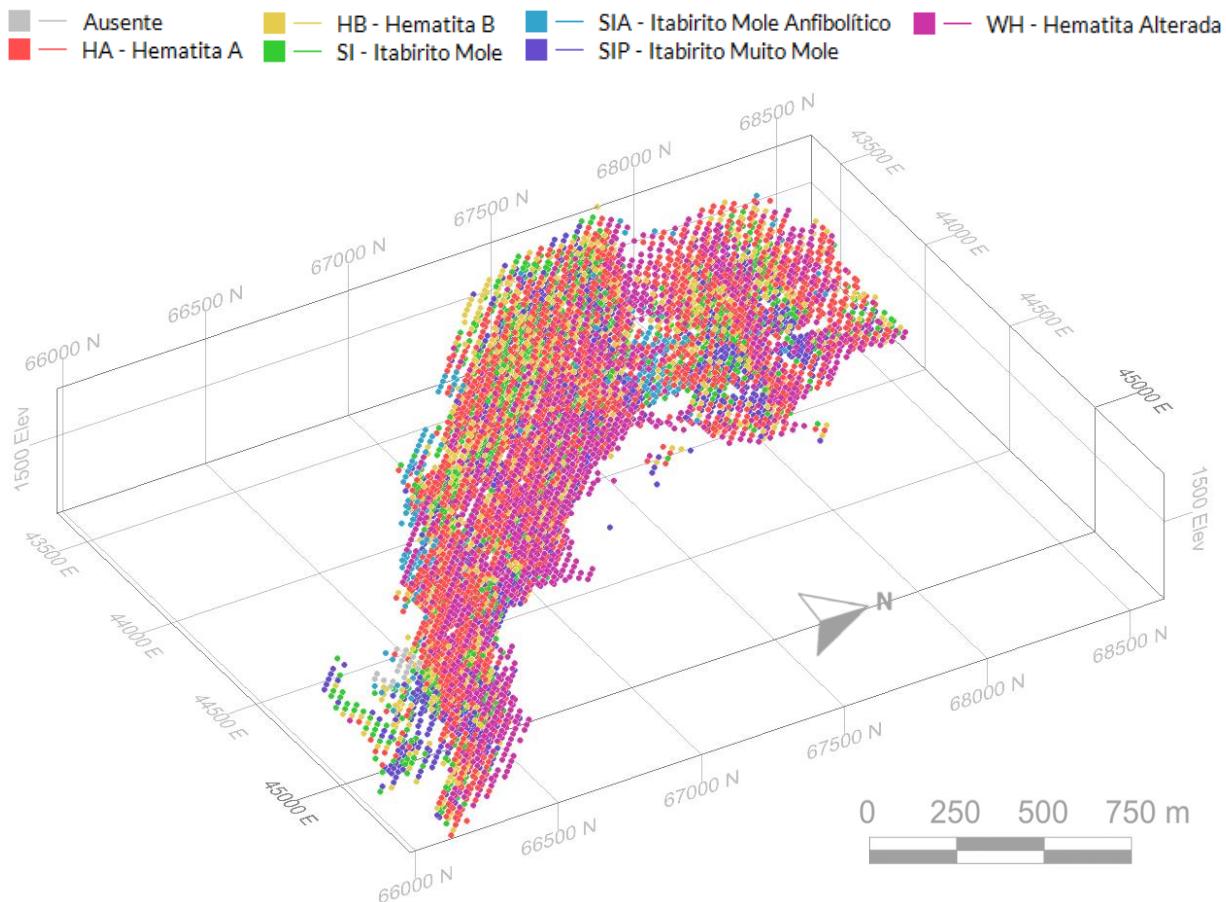


Figura 13: Malha dos furos de perfuratriz em perspectiva com legenda de litotipo.

Conforme Massahud (1996 *apud* Rocha, 1999) e Fonseca (2014), os litotipos presentes se caracterizam como itabiríticos e hematíticos:

Itabiríticos

Compostos sobretudo por hematita (Fe_2O_3), com baixa concentração de magnetita (Fe_3O_4) e conteúdo variável de goethita (FeOH_n).

- i. Itabirito Mole (SI): intercalação de bandas de sílica e hematita não consolidadas com 50% a 60% de ferro. Considerado um itabirito rico.

- ii. Itabirito Muito Mole (SIP): intercalação de bandas de sílica e hematita não consolidadas, menos compacto que o anterior, com teores de ferro inferiores a ~50%. Considerado um itabirito pobre.
- iii. Itabirito Mole Anfibolítico (SIA): intercalação de bandas de sílica e hematita não consolidadas com anfibólito bastante alterado, em que o óxido de ferro predominante é a goethita e os teores variam de 55 a 60%.

Hematíticos

Descritos como camadas contínuas com 10 a 30 metros de espessura com teores de ferro acima de 62%. Ainda de acordo com Fonseca (2014), quanto mais exposto e alterado o hematitito, mais empobrecido em ferro e enriquecido em contaminantes (alumina e fósforo).

- i. Hematita A (HA): laminada e bandada, friável a pulverulenta, de baixa dureza e com poucos leitos de sílica livre.
- ii. Hematita B (HB): laminada e bandada, com maior frequência de leitos silicosos.
- iii. Hematita Alterada (WH): laminada e bandada, rica em minerais hidratados (hidróxidos de ferro, hidróxidos de alumínio e silicatos hidratados de alumínio), com teores em torno de 60% de ferro, baixa sílica e rica em contaminantes.

5.1.2. Validação e tratamento dos dados

Foram identificados 1.365 registros cujos valores de ferro, alumina, sílica e fósforo eram simultaneamente iguais a zero, além de possuírem PPC ausente. Estes registros, distribuídos por quatro bancadas e não correspondentes a litotipos específicos, foram interpretados como valores ausentes e desconsiderados no fluxo de trabalho, restando 6.980 amostras válidas. Ainda, visando a definição dos subdomínios, foram removidas 261 amostras que possuíam teor ausente de fósforo, pois os algoritmos de agrupamento não admitem o uso de observações ausentes, restando 6.719 amostras para o fluxo de trabalho.

Tendo em vista a definição de subdomínios com base nos litotipos, realizou-se a classificação das 18 amostras com litotipo ausente através de Análise Discriminante Múltipla, a qual utiliza n variáveis métricas (ferro, alumina e fósforo) agrupadas em i grupos discretos (os seis litotipos da mina) para definir, com base em amostras conhecidas, a qual grupo pertence uma amostra indefinida (Pinches, 1980). Como resultado, treze das amostras foram classificadas como HB, três amostras como HA e duas como SIA, com índice de acerto de aproximadamente 81%.

5.1.3. Análise estatística

As medidas de tendência central, dispersão e forma de todas as variáveis foram obtidas tanto para as amostras de maneira geral, apresentadas na Tabela 3, quanto por litotipo

(Apêndice I). Destacam-se na Tabela 3 a alta variabilidade do fósforo ($CV = 4,62$) e a baixa variabilidade do ferro ($CV = 0,09$), que chega a ser inferior a 0,06 na maioria dos litotipos. Destaca-se também o alto teor médio do ferro para o depósito como um todo (59,89%).

Tabela 3: Estatísticas descritivas gerais do banco de dados validado e tratado.

Variável	n	Mínimo	Máximo	Média	Variância	s	CV
Fe	6.719	3,08	67,67	59,89	28,62	5,35	0,09
Al_2O_3	6.719	0,00	20,20	1,49	1,69	1,30	0,87
P	6.719	0,00	10,61	0,12	0,33	0,58	4,62
SiO_2	6.719	0,00	93,30	8,57	79,22	8,90	1,04
PPC	6.719	0,01	14,37	3,95	5,43	2,33	0,59

n = número de amostras; s = desvio padrão, CV = coeficiente de variação

Conforme esperado, os litotipos pobres em fósforo e alumina e ricos em ferro são HA (~65% Fe), HB (~62% Fe) e SI (~58% Fe), conforme exemplificado na Figura 14. As amostras dos demais litotipos apresentam ou baixo teor de ferro (SIP) ou alta PPC e concentração de contaminantes (SIA e WH), como é ilustrado na Figura 14 e nos boxplots do Apêndice II.

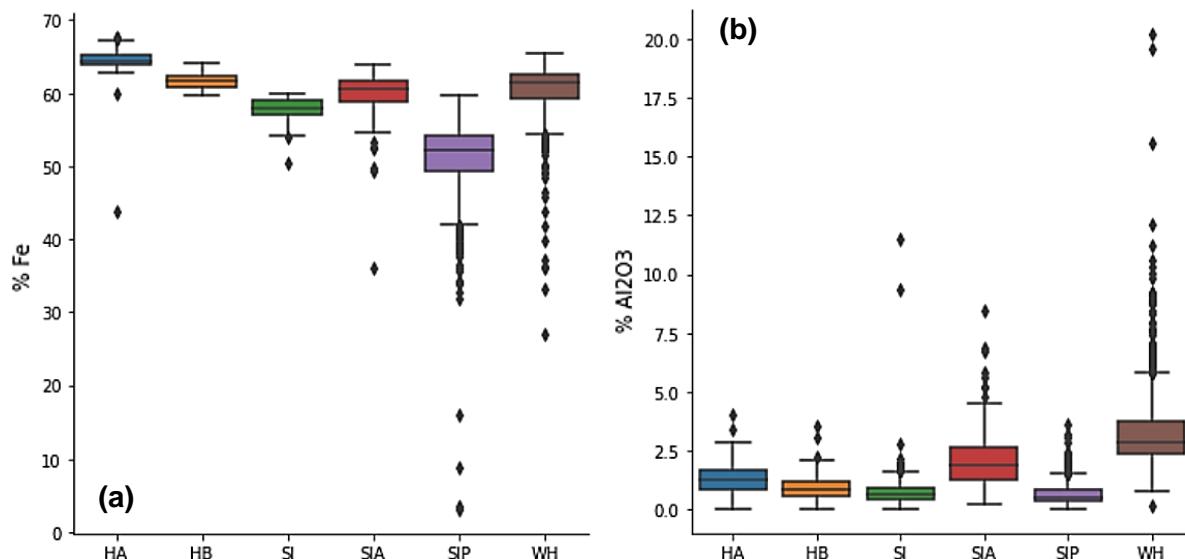


Figura 14: Boxplots dos teores de (a) ferro e (b) alumina por litotipo do banco de dados validado e tratado.

Nos histogramas apresentados na Figura 15, verifica-se: a distribuição assimétrica positiva dos conteúdos de sílica, alumina e fósforo e da PPC, a distribuição assimétrica negativa do conteúdo de ferro e a presença de *outliers*. Observa-se a ocorrência de *outliers* de ferro, sílica, alumina e fósforo, sendo que os valores extremos do fósforo podem ser considerados uma segunda população com teores entre 1-10%, aproximadamente (Figura 15c). Os histogramas para os dados originais e para os dados validados e tratados podem ser observados no Apêndice III.

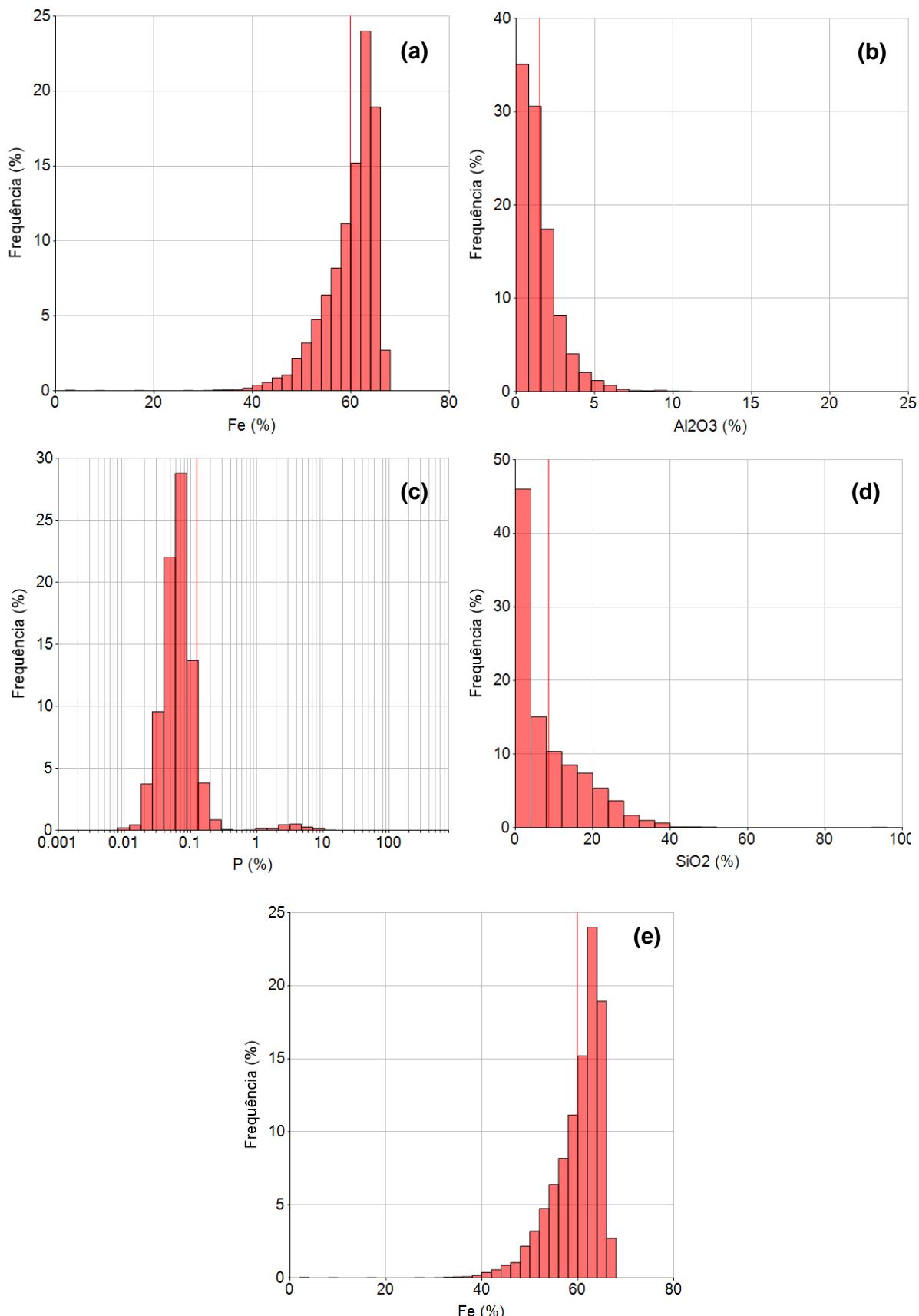


Figura 15: Histogramas gerais dos teores de **(a)** ferro, **(b)** alumina, **(c)** fósforo em escala de log-probabilidade aritmética e **(d)** sílica; e **(e)** da perda por calcinação dos dados validados e tratados.

Na análise bivariada, os coeficientes de correlação linear de Pearson (Tabela 4) apontam correlação linear superior a $|0,7|$ apenas entre as variáveis sílica e ferro ($r = -0,91$) e moderada entre PPC e alumina ($r = 0,78$).

Tabela 4: Matriz de correlação linear entre as variáveis numéricas do banco de dados validado e tratado, destacadas as maiores correlações.

	<i>Al₂O₃</i>	<i>Fe</i>	<i>P</i>	<i>PPC</i>	<i>SiO₂</i>
<i>Al₂O₃</i>	-	0,088	0,050	0,780	-0,430
<i>Fe</i>	0,088	-	0,020	0,210	-0,910
<i>P</i>	0,050	0,020	-	0,094	-0,048
<i>PPC</i>	0,780	0,210	0,094	-	-0,560
<i>SiO₂</i>	-0,430	-0,910	-0,048	-0,560	-

Os gráficos de dispersão evidenciam como os teores de ferro e sílica foram critérios preponderantes na classificação dos litotipos HA, HB, SI e SIP, como ilustrado na Figura 16a. Curiosamente, quando observados separadamente, os *outliers* do fósforo (teores 1-10%, conforme o histograma) possuem correlação linear muito elevada com a alumina e a PPC (Figura 16b). Os diagramas de dispersão de todos os pares de variáveis podem ser observados no Apêndice V.

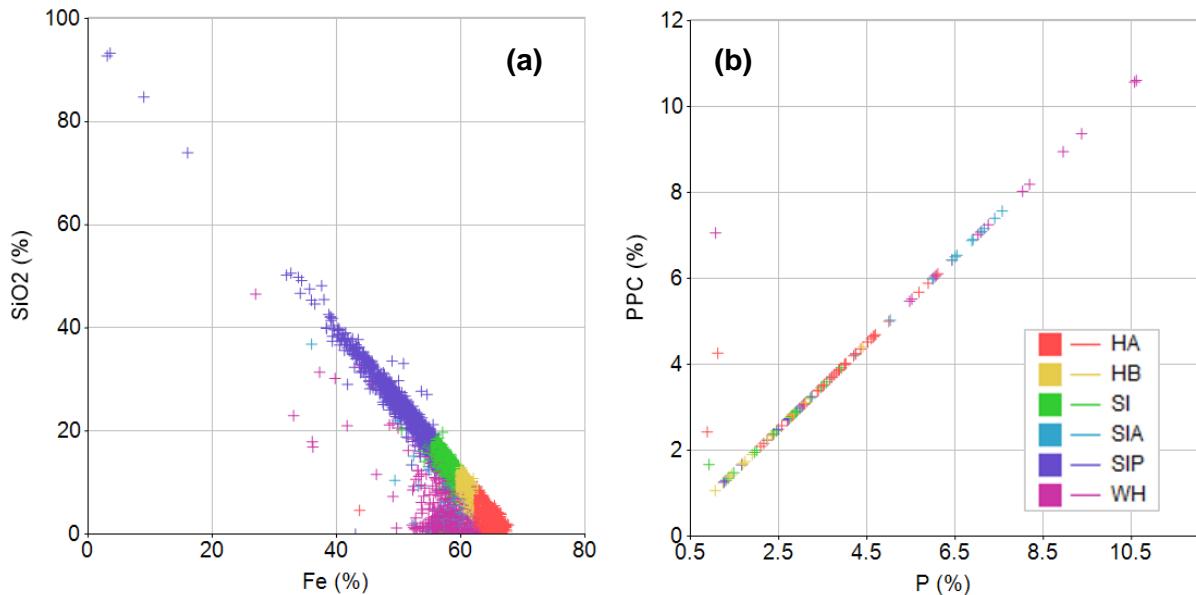


Figura 16: Gráficos de dispersão entre (a) ferro e sílica e (b) entre os *outliers* de fósforo e a perda por calcinação do banco de dados validado e tratado com legenda de litotipos.

5.1.1. Análise visual das variáveis no espaço

O uso de legendas de teores evidenciou, visualmente, alguns padrões na distribuição espacial de teores dos elementos analisados, como a forte variação vertical do fósforo (Figura 17). Verifica-se também menores teores de ferro e alumina e menores porcentagens de PPC nas amostras mais distantes da superfície topográfica, conforme ilustrado no perfil da Figura

18, devido à geometria da Sinforme Córrego das Flechas e o forte controle litológico (Figura 3). O Apêndice IV é composto pelos mapas de pontos amostrais, com legenda de teores, de todas as variáveis.

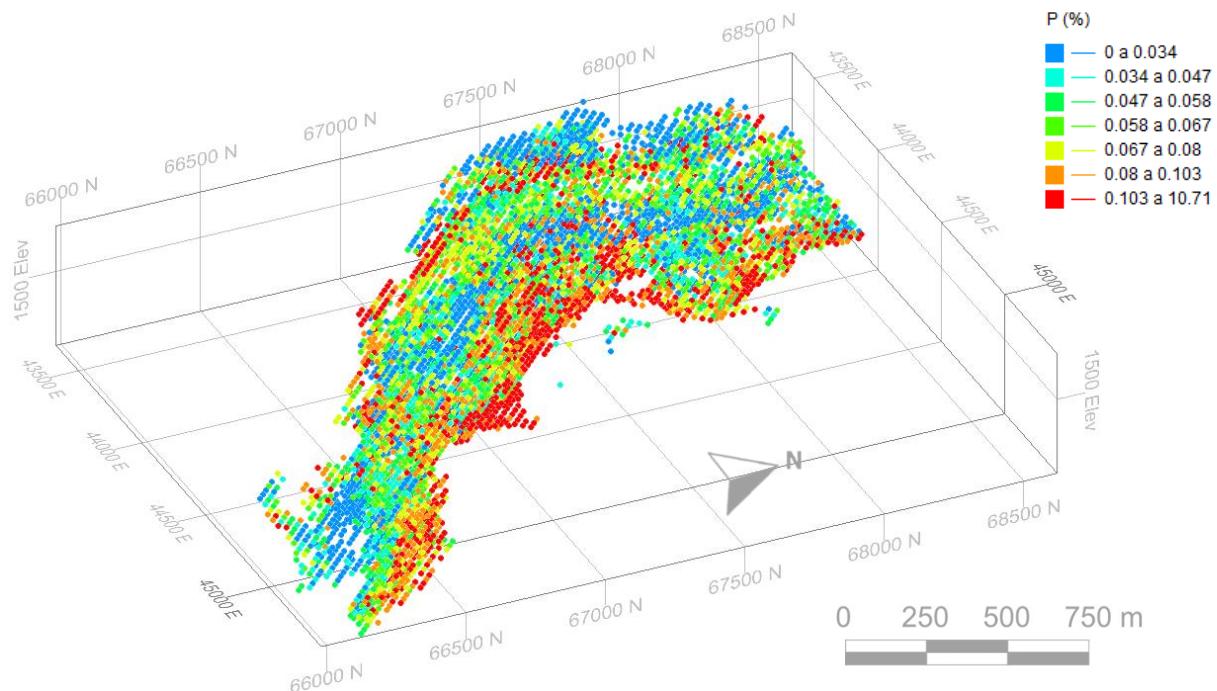


Figura 17: Malha em perspectiva com legenda de teores de fósforo.



Figura 18: Perfil vertical SW-NE da malha com legenda de teores de ferro.

5.2. Agrupamento das Amostras

A primeira definição realizada se baseou somente nos litotipos. As amostras de HA, HB e SI foram classificadas manualmente como minério (1) e as amostras de SIA, SIP e WH, como estéril (0). Neste estudo, “estéril” refere-se a todas as observações que não apresentam simultaneamente alto teor de ferro (>55% para a mina Capanema) e baixa concentração de contaminantes, de acordo com a Tabela 1 de referência.

Os demais cenários de agrupamento foram obtidos com os algoritmos de análise de agrupamento KM, HC e GHC a partir das variáveis-critério ferro, alumina e fósforo. Considerando-se as correlações sílica-ferro e PPC-alumina observadas na análise

exploratória, o uso da sílica e da alumina foi dispensado para evitar ruídos e a geração de grupos redundantes. Antes da aplicação dos algoritmos, as variáveis-critério foram padronizadas para evitar que as diferentes escalas enviesassem os agrupamentos.

De acordo com o método do cotovelo (Figura 19), o agrupamento mais eficiente para o banco de dados com as variáveis-critério escolhidas corresponde à definição de quatro subdomínios. Na aplicação do *K-means* com a biblioteca Scikit-learn (Pedregosa et al., 2011), o algoritmo foi executado dez vezes com a predefinição de quatro *clusters* e escolheu-se um dos resultados que apresentavam o padrão de agrupamento mais frequente.

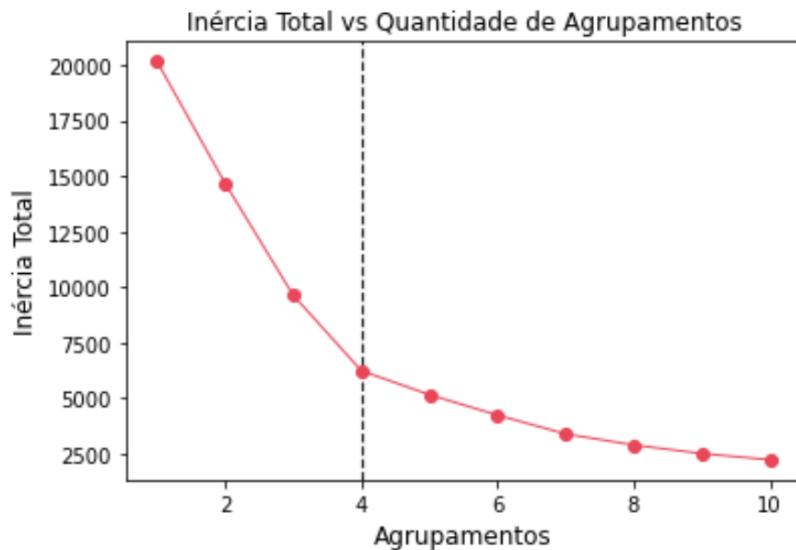


Figura 19: Gráfico da inércia total (soma total dos quadrados intra-cluster) pela quantidade de *clusters* com o algoritmo de análise de agrupamento *K-means*.

Na aplicação do HC (biblioteca Scikit-learn) e do GHC (*software Isatis.neo*), os níveis hierárquicos escolhidos foram os que resultavam também em quatro agrupamentos e o critério de proximidade utilizado foi o Ward para ambos os algoritmos. Embora a interface do GHC no *software Isatis.neo* seja bastante personalizável, optou-se por utilizar a configuração *default* da ferramenta, que atribui o mesmo peso a todas as variáveis-critérios e assume um cenário isotrópico de amplitude igual a 1.000 m.

Com a análise estatística de cada *cluster*, verificou-se a tendência de todos os algoritmos distinguirem: um grupo de alto teor de ferro e baixa concentração de contaminantes (*cluster 0*), dois grupos com alto teor de ferro e alta concentração de cada contaminante (*clusters 1 e 2*), e um quarto grupo de baixo teor de ferro (*cluster 4*), conforme pode ser observado na Tabela 5. Para padronizar todos os cenários, os *clusters* foram reagrupados em minério (1) e estéril (0) conforme apresenta-se na Tabela 5, utilizando-se como critério as estatísticas de ferro e de contaminantes em cada um. Esta reclassificação binária também serviu para aumentar a quantidade de observações em cada grupo e assim não restringir demais a quantidade de amostras estimadoras disponíveis para cada bloco, além de facilitar a classificação dos próprios blocos com base nas amostras adjacentes, excluindo-se

situações em que um único bloco poderia ser circundado por quatro grupos diferentes, por exemplo.

Tabela 5: Teores médios das variáveis critério em cada agrupamento por algoritmo de análise de agrupamento, destaca-se as médias que caracterizam cada grupo.

Algoritmo	Cluster	Teor médio			Subdomínio
		% Fe	% Al ₂ O ₃	% P	
<i>K-means</i>	0	62,91	1,31	0,07	Minério
	1	60,83	1,60	4,66	Estéril
	2	59,50	4,11	0,09	Estéril
	3	52,75	0,67	0,05	Estéril
Aglomerativo	0	62,12	0,96	0,05	Minério
	1	60,81	1,37	3,97	Estéril
	2	61,05	3,10	0,08	Estéril
	3	50,71	0,69	0,05	Estéril
Hierárquico	0	61,90	1,09	0,12	Minério
	1	61,29	1,56	5,08	Estéril
	2	61,19	3,13	0,08	Estéril
	3	52,52	0,70	0,05	Estéril
Geoestatístico	0	61,90	1,09	0,12	Minério
	1	61,29	1,56	5,08	Estéril
	2	61,19	3,13	0,08	Estéril
	3	52,52	0,70	0,05	Estéril

Espacialmente, os resultados dos algoritmos HC e GHC são muito semelhantes e se diferenciam do *K-means* sobretudo em relação aos clusters 0 e 2 (Figura 21), pois cerca de 50% das amostras do cluster 2 nos algoritmos do tipo hierárquico foram atribuídas ao cluster 0 pelo *K-means*.

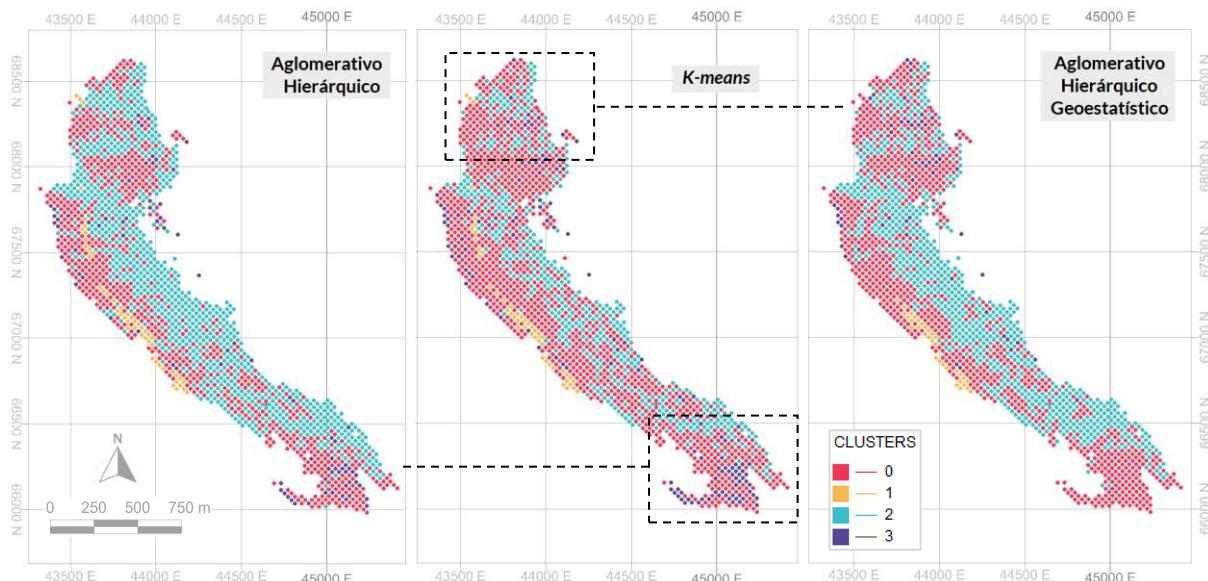


Figura 20: Malha dos furos de perfuratriz com a legenda de clusters por algoritmo de análise de agrupamento. Linhas tracejadas demonstram em quais porções da malha o resultado do *K-means* se assemelha mais com o HC e o GHC.

Na Figura 21, verifica-se que a semelhança entre os clusters identificados por cada algoritmo não se restringe às médias e todos possuem distribuições estatísticas parecidas. Observa-se também a inclusão de outliers indesejados no cluster 0, de maior interesse, pelos algoritmos *K-means* (para a variável fósforo) e GHC (para as variáveis ferro, alumina e

fósforo).

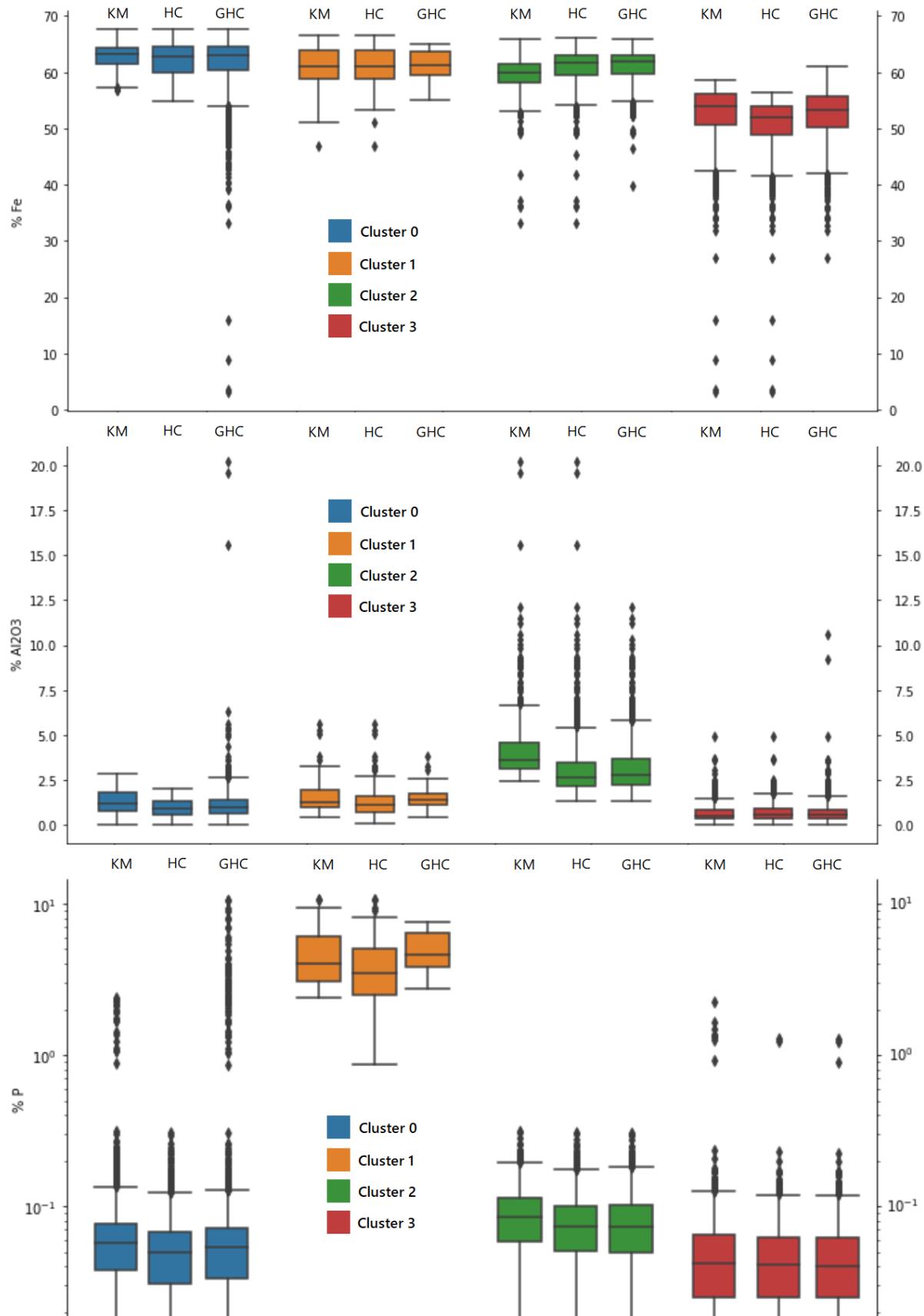


Figura 21: Boxplots dos teores de ferro, alumina e fósforo em cada cluster por algoritmo de análise de agrupamento. K-means (KM), Aglomerativo Hierárquico (HC) e Aglomerativo Hierárquico Geoestatístico (GHC).

5.3. Modelo de Blocos

As configurações utilizadas para a geração do modelo de blocos visaram a colocação das amostras nas arestas verticais das células parentais, conforme a Figura 22. Para isso, as células parentais têm as configurações da Tabela 6. Como o modelo de blocos foi rotacionado, refere-se aos eixos X, Y e Z após rotação como eixos U, V e W, respectivamente. São, no total, 47 células no eixo U (-46°) do modelo, 110 células no eixo V (44°) e 18 no eixo W (vertical).

Tabela 6: Propriedades das células parentais do modelo de blocos.

Rotação	Dimensão em U	Dimensão em V	Dimensão em W
-46°	28,28 m	28,28 m	26 m

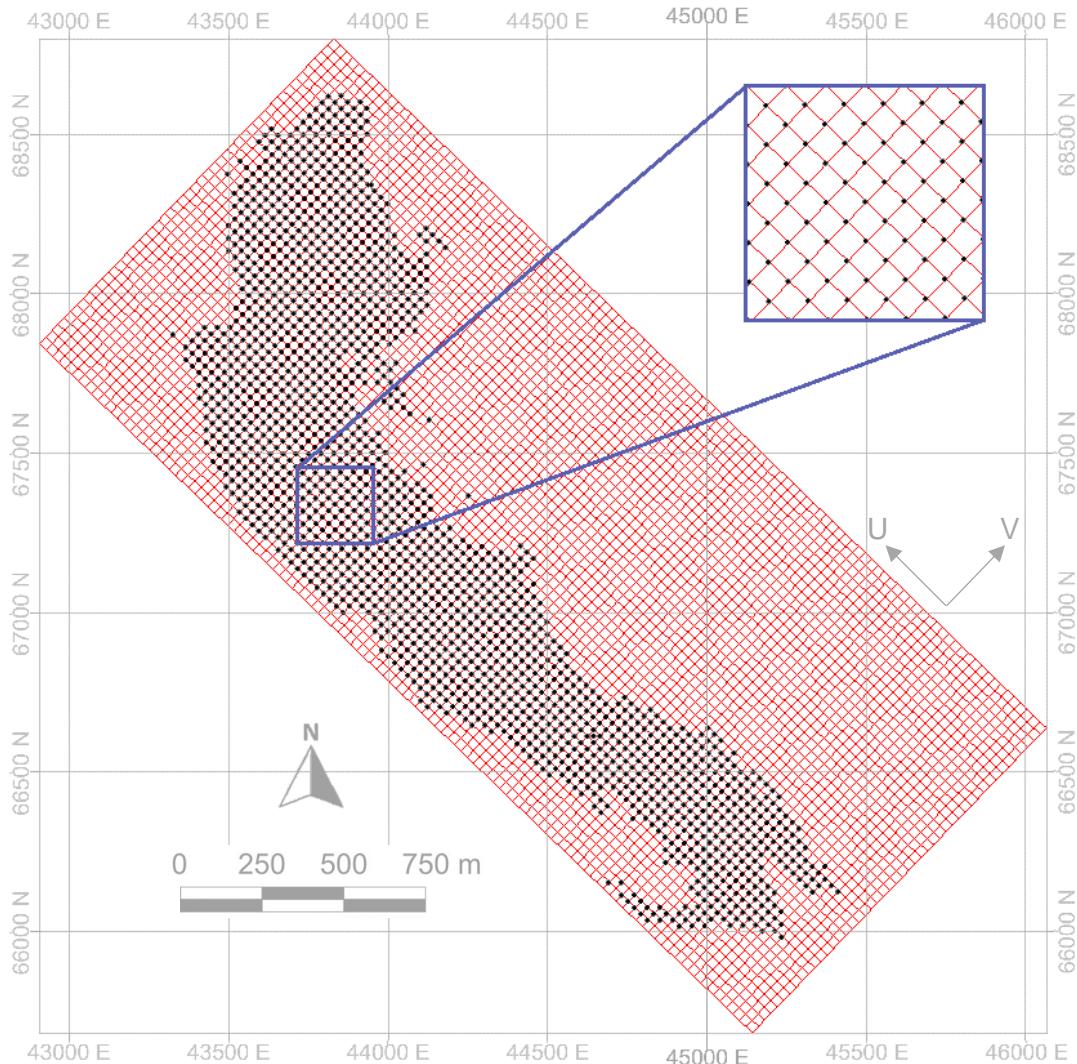


Figura 22: Modelo de blocos sobre a malha dos furos de perfuratriz em planta.

5.3.1. Definição dos Subdomínios

A partir da reclassificação binária (Tabela 5), os cenários foram simplificados sem prejuízo no agrupamento das amostras de maior interesse (minério), o que facilitou a agrupamento das células do modelo de blocos nos mesmos subdomínios de suas amostras adjacentes. Para isso, as células foram consideradas minério quando possuíam mais amostras dessa natureza em suas arestas verticais, e consideradas estéril utilizando-se a mesma lógica. Ainda, as células que compartilhavam a mesma quantidade de amostras de estéril e de minério foram consideradas um subdomínio de transição, conforme esquematizado na Figura 23



Figura 23: Esquematização da relação entre os subdomínios das amostras (estéril e minério) e a definição de subdomínios no modelo de blocos (estéril, transição e minério).

Por fim, as amostras e o modelo de blocos receberam quatro novos campos, relacionados aos quatro cenários de definição de subdomínios para estimativa, conforme indicado na Tabela 7 e ilustrado na Figura 24.

Tabela 7: Os quatro cenários de definição dos subdomínios estéril, minério e de transição.

Campo	Informação
<i>ESTMIN_LITH</i>	Amostras classificadas como 0 (estéril) ou 1 (minério) com base nos litotipos; e blocos classificados como 0, 1 ou 2 (transição).
<i>ESTMIN_KM</i>	Amostras classificadas como 0 (estéril) ou 1 (minério) com base nos clusters do algoritmo <i>K-means</i> ; e blocos classificados como 0, 1 ou 2 (transição).
<i>ESTMIN_HC</i>	Amostras classificadas como 0 (estéril) ou 1 (minério) com base nos clusters do algoritmo Aglomerativo Hierárquico; e blocos classificados como 0, 1 ou 2 (transição).
<i>ESTMIN_GHC</i>	Amostras classificadas como 0 (estéril) ou 1 (minério) com base nos clusters do algoritmo Aglomerativo Hierárquico Geoestatístico; e blocos classificados como 0, 1 ou 2 (transição).

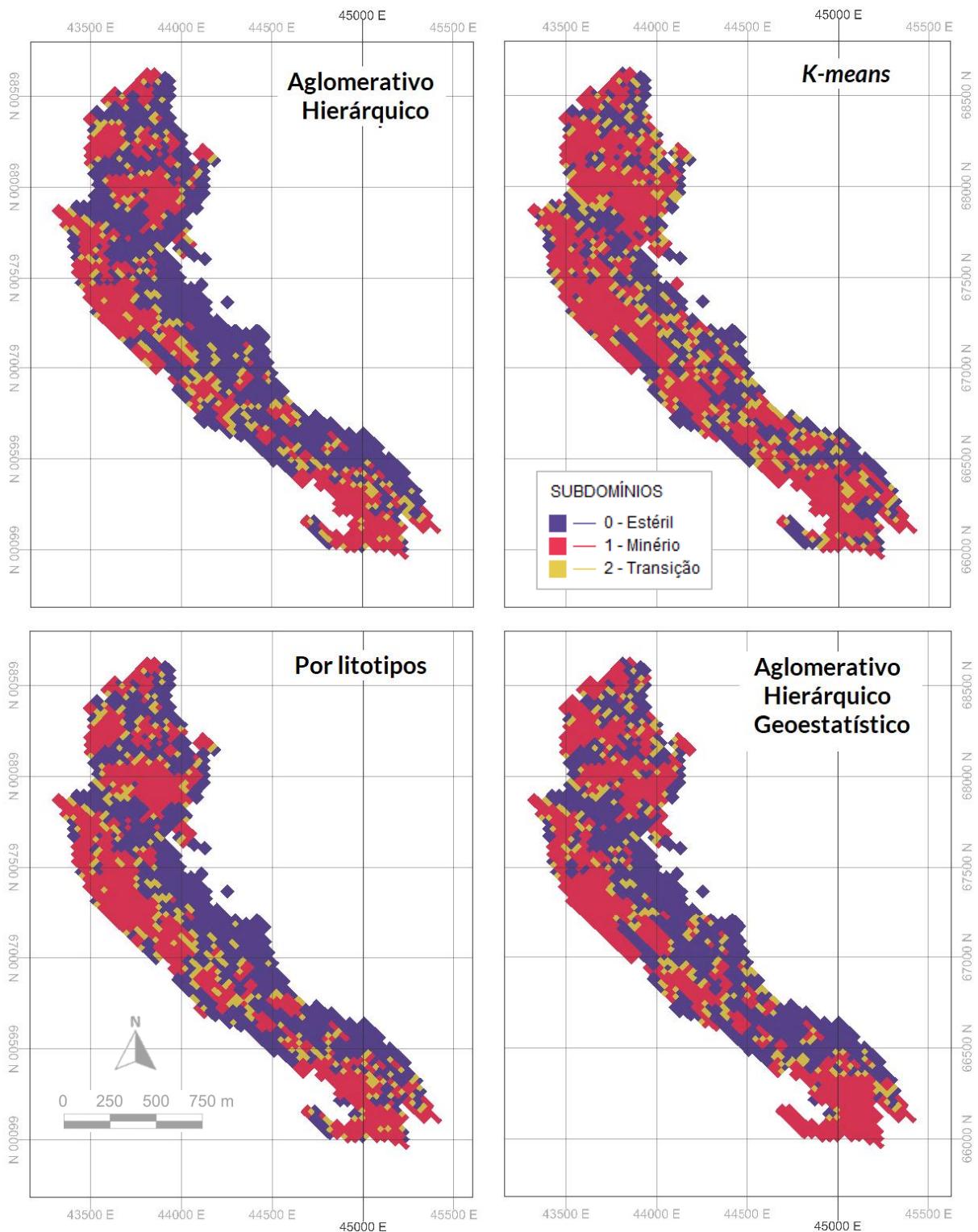


Figura 24: Vista em planta do modelo de blocos em cada um dos quatro cenários de definição de subdomínios para estimativa.

5.4. Análise Geoestatística

O cálculo dos variogramas experimentais foi realizado para o ferro nos cinco diferentes cenários, sempre com tamanho do passo igual a 40 metros. Em todos os cálculos, utilizaram-se 18 direções (variação de 10° na direção angular) e tolerâncias angulares de 10° sem largura máxima. Visto que cada furo é composto de uma única amostra, os variogramas *downhole* não foram calculados e os efeitos pepita utilizados foram obtidos dos variogramas

omnidirecionais. No cenário de domínio único, os eixos D1 e D2 possuem patamares menores que a variância a priori e alcances próximos de 200 metros (Figura 25). O terceiro eixo (perpendicular à xistosidade) tem continuidade inferior (<120 metros) e patamar acima da variância a priori. Os modelos teóricos de variograma e as elipsoides de continuidade de cada subdomínio por cenário podem ser observados nos Apêndice VI e Apêndice VII, respectivamente.

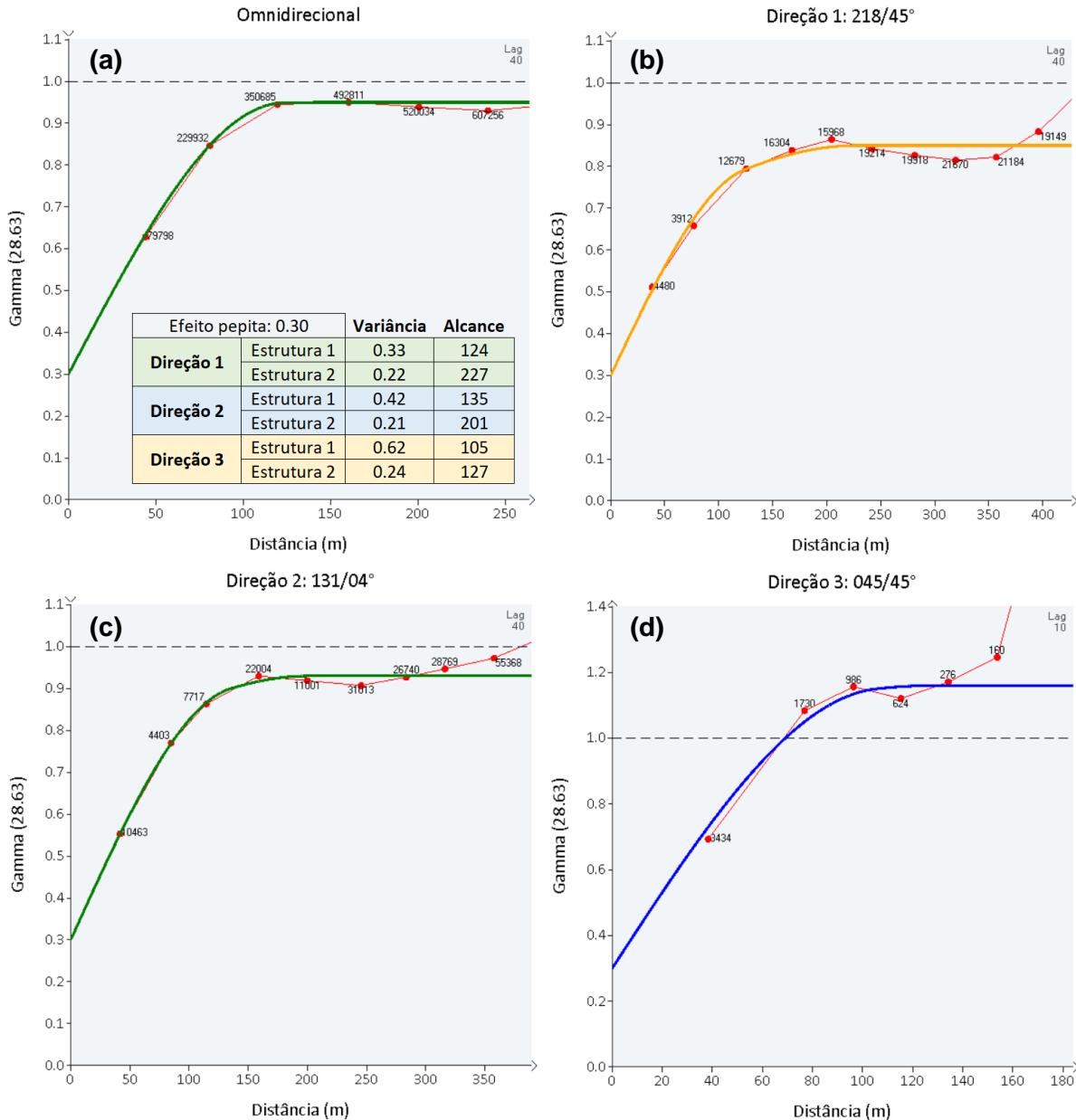


Figura 25: Variogramas experimentais e respectivos modelos teóricos ajustados do ferro no cenário de domínio único. (a) Omnidirecional, (b) direção 1, (c) direção 2 e (d) direção 3. Eixo vertical: variância relativa à variância a priori (γ).

5.5. Estimativas

Os volumes de busca foram configurados com as mesmas dimensões e altitudes das elipsoides de continuidade. Utilizou-se também mínimo de 4 e máximo de 12 amostras para a estimativa. Os blocos também foram discretizados três vezes em cada eixo.

O alinhamento das amostras na malha favoreceu a geração de muitos pesos negativos e a estimativa de teores negativos de ferro. Portanto, os pesos negativos foram corrigidos igualando-os a zero e com o reajuste dos pesos positivos. Em cada cenário, blocos do subdomínio estéril foram estimados apenas com amostras do domínio estéril, utilizando-se a mesma lógica para o subdomínio minério (Figura 27). Na estimativa dos blocos de transição, foram interpoladas tanto as amostras do tipo estéril quanto as do tipo minério, e utilizados os variogramas do cenário de domínio único.

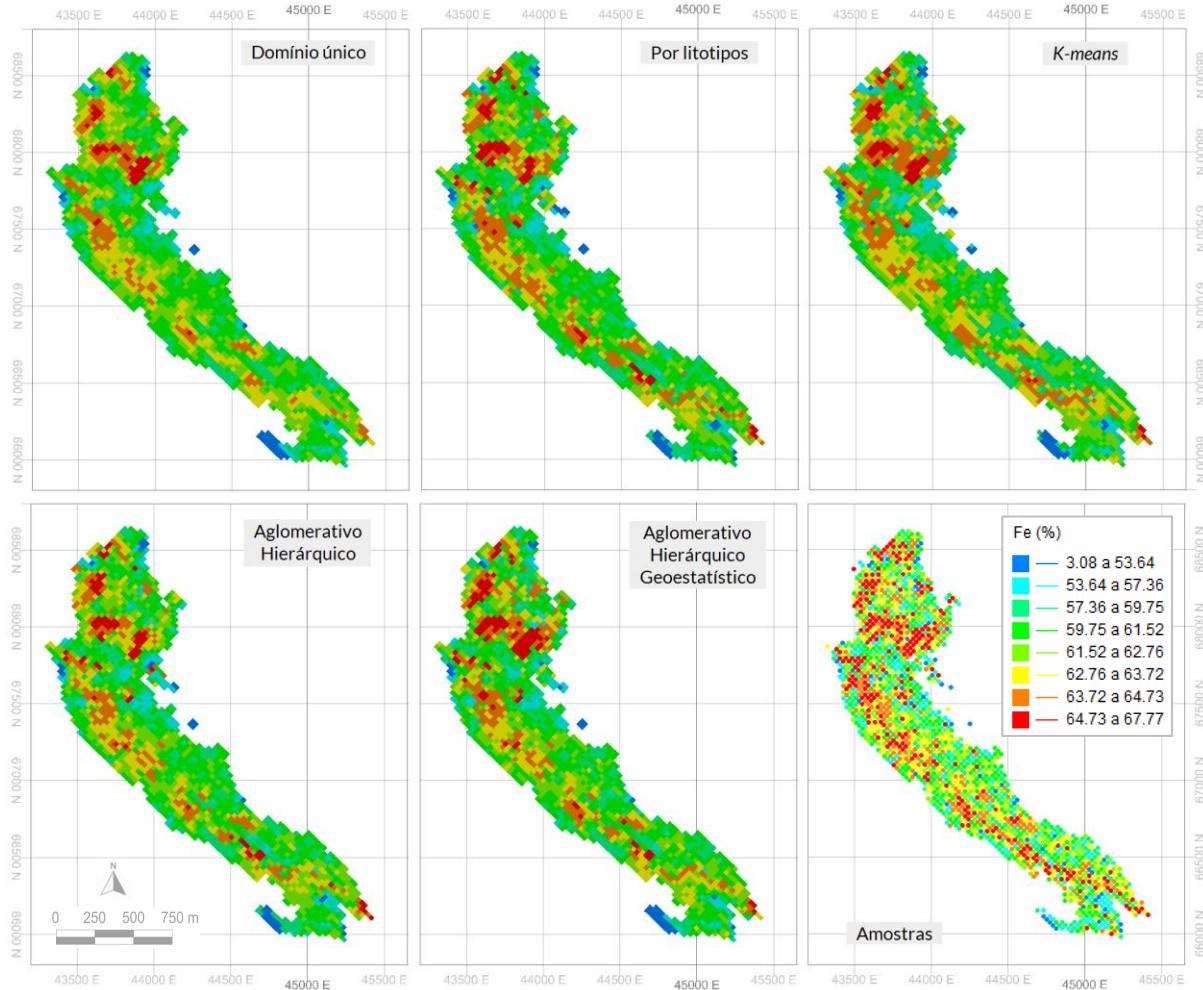


Figura 26: Vista em planta do modelo de blocos com teores estimados de ferro em cada cenário.

5.6. Validação e Avaliação dos Efeitos do Agrupamento nas Estimativas

Na Tabela 8, verifica-se que todos os agrupamentos resultaram em estimativas com menor diferença entre a média amostral e estimada, mas que esta diferença permanece muito próxima da gerada pelo uso de um domínio único. Pelos histogramas de teores estimados (Apêndice VIII), verifica-se que todas as definições de subdomínios resultaram na estimativa de duas populações estatísticas de ferro (histogramas bimodais), enquanto os cenários de domínio único e agrupamento pelo algoritmo GHC melhor preservaram a forma unimodal da distribuição amostral.

Tabela 8: Estatísticas descritivas dos teores de ferro nas amostras e nos modelos estimados em cada cenário, destacados os valores mínimo e máximo, média e variância mais próximos do conjunto amostral.

	Amostras	Domínio único	Litotipos	KM	HC	GHC
<i>Mínimo</i>	3,08	18,29	12,75	23,12	9,74	12,29
<i>Máximo</i>	67,67	66,25	66,06	66,12	66,21	66,59
<i>Média</i>	59,89	59,60	59,71	59,74	59,66	59,65
<i>Variância</i>	28,62	13,77	17,97	19,29	17,88	18,41

Os erros obtidos com a validação cruzada são muito próximos a zero (sem viés) em todos os cenários, mas foi o cenário de agrupamento com o *K-means* que retornou o menor erro médio e a menor variância (Figura 28 e Apêndice IX). Consequentemente, o coeficiente de correlação linear de Pearson deste cenário também é o maior observado ($r = 0,836$). Nota-se que os coeficientes de correlação para os agrupamentos são maiores devido ao efeito de “afunilamento” (Apêndice IX-B) gerado pelas distribuições bimodais do ferro estimado.

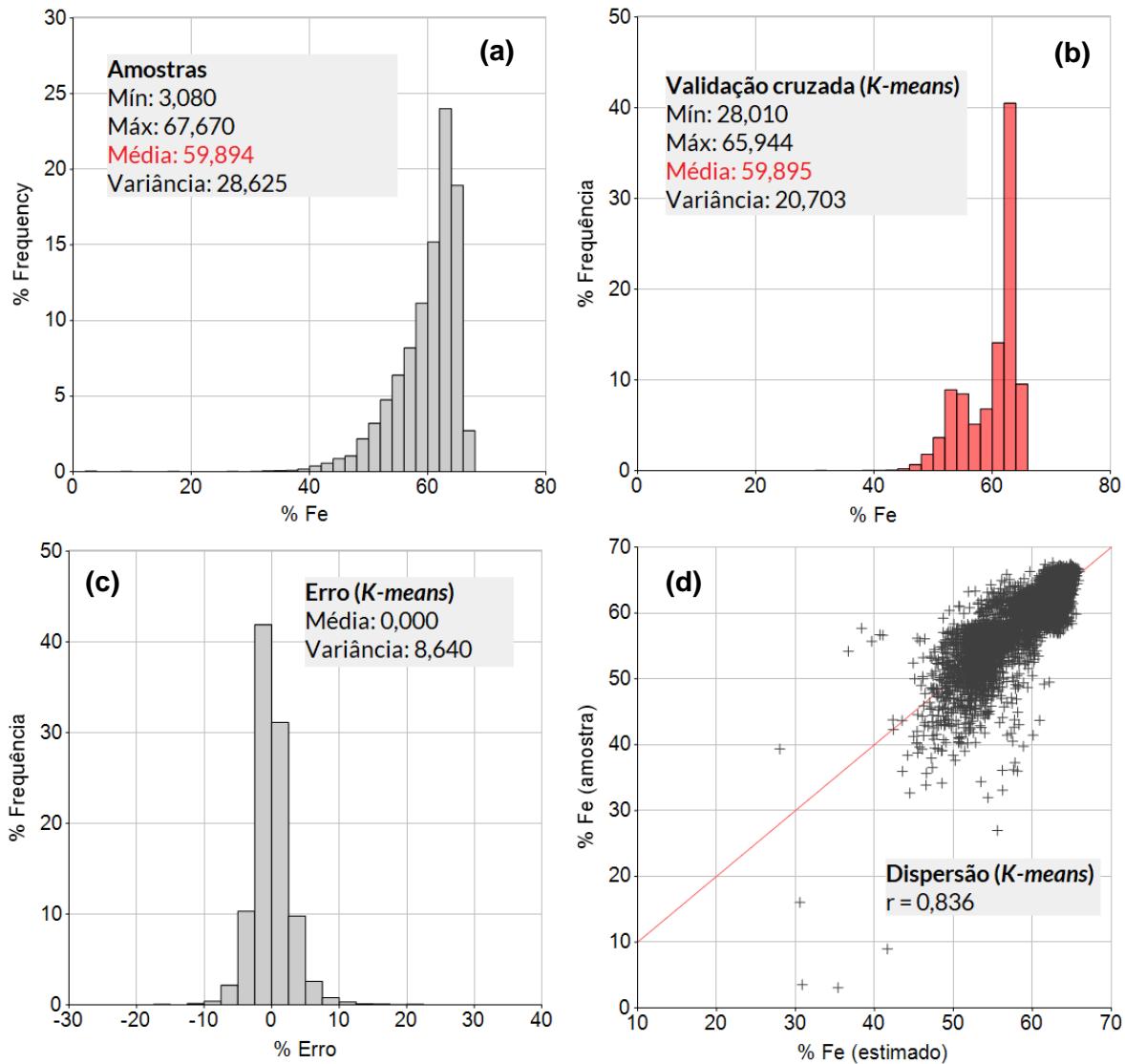


Figura 27: Histogramas dos teores de ferro **(a)** nas amostras e **(b)** resultantes da validação cruzada; **(c)** histograma dos erros e **(d)** dispersão entre os teores reais e as estimativas pontuais para o cenário do *K-means*.

Os cenários de agrupamento por litotipos, HC e GHC também apresentaram menor variância de erros e maiores coeficientes de correlação em relação ao domínio único, embora as diferenças sejam pouco significativas. Verifica-se também que os maiores erros se concentram nos subdomínios de estéril, devido à mistura de amostras de baixo teor de ferro com amostras de alto teor de ferro e alta concentração de contaminantes. Para o *K-means*, por exemplo, a variância do erro no subdomínio minério é igual a 3,429 (40% menor que a variância geral).

Pela análise dos gráficos quantil-quantil (Apêndice X), observa-se que todos os métodos se afastam da distribuição amostral nas duas extremidades, conforme o esperado pelo efeito de suavização de estimativa. O cenário de domínio único é o que mais se afasta e o cenário do *K-means*, o que mais se aproxima (Figura 29). Como consequência das formas dos histogramas, os cenários de domínio único e agrupamento com GHC são os que mais se aproximam de uma distribuição normal, enquanto os demais apresentam comportamento bimodal. Na Figura 28, estão destacados os intervalos de teores de maior semelhança.

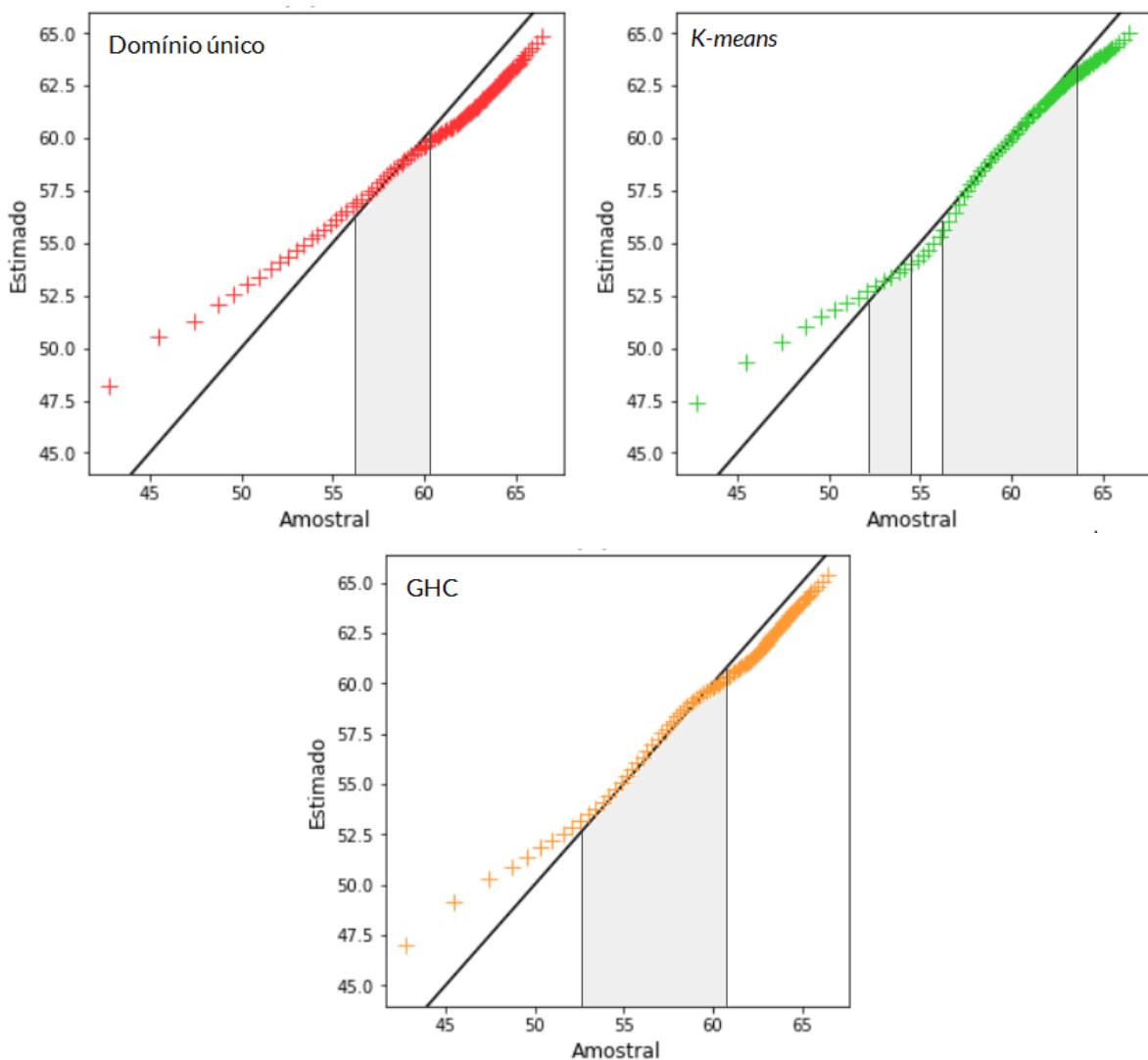


Figura 28: Gráficos Quantil-Quantil entre os teores de ferro do conjunto amostral e as estimativas dos cenários com domínio único e com agrupamento por *K-means* e GHC.

Na análise de deriva da Figura 29, pode-se ver como as acuráncias das estimativas nos diferentes cenários são altas e muito próximas entre si, conforme indicado pelos erros observados na validação cruzada, apresentando boa aderência com as médias móveis da malha amostral. Ainda, a curva do GHC é a que melhor se adere à curva de estimativa com domínio único em praticamente todas as fatias, o que provavelmente é um efeito da sua propriedade de incluir a posição geográfica dos pontos no critério de agrupamento.

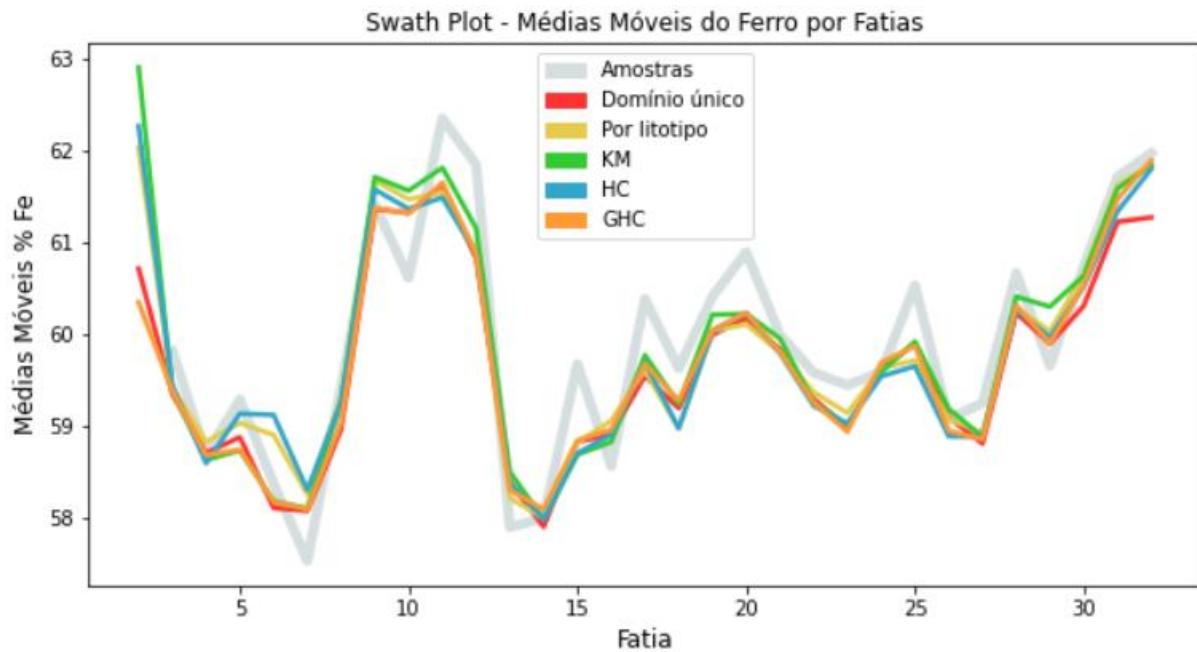


Figura 29: Médias móveis do ferro por fatias de orientação E-W e com espessura igual a 100 metros.

6. INTERPRETAÇÃO E DISCUSSÃO DOS RESULTADOS

6.1. Agrupamentos

O dendrograma da Figura 30 permite identificar que o maior nível hierárquico do HC distingue as duas populações de fósforo observadas na AED, tal como o maior nível do GHC, devido ao coeficiente de variação muito superior ao das demais variáveis. Portanto, para os objetivos do estudo, a utilização dos quatro *clusters* apontados pelo método do cotovelo do algoritmo *K-means* também para os métodos hierárquicos foi essencial, visto que o uso de dois *clusters* discretizaria as amostras apenas pelo teor de fósforo, mascarando a importância do ferro e não permitindo a sua reclassificação em minério ou estéril. Até mesmo o agrupamento em três *clusters* não atenderia aos objetivos do estudo, visto que neste caso o *cluster* rico em ferro agruparia também as amostras ricas em alumina.

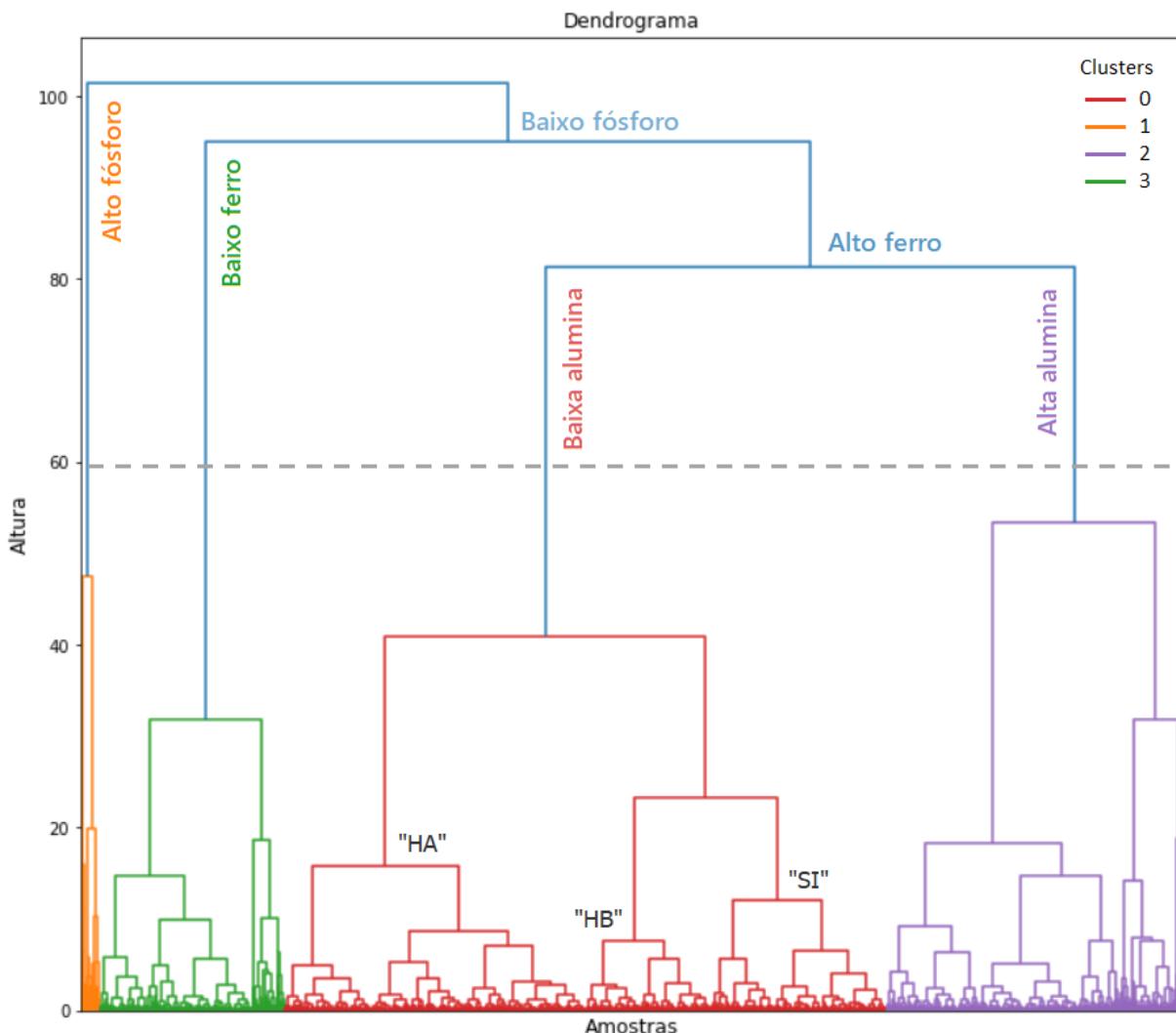


Figura 30: Dendrograma do agrupamento realizado pelo método Ward do algoritmo Aglomerativo Hierárquico. Altura = Distância euclidiana entre os pares no espaço multivariado.

O cluster 0 é constituído por dois subgrupos na altura ~30 (Figura 30), os quais se relacionam com os litotipos de minério. Enquanto o subgrupo denominado como "HA" na Figura 30 é composto majoritariamente por amostras de Hematita A com significativa presença de HB e menor frequência de Itabirito Mole Anfibolítico e Hematita Alterada, o outro subgrupo é composto em sua maioria por Hematita B e Itabirito Mole (SI) com a inclusão menos expressiva de amostras dos demais litotipos. Conforme verificado na análise exploratória, os teores de HB são todos intermediários aos de HA e SI (Apêndice I e Apêndice II).

Sob análise de contato, o contato entre os subdomínios estéril e minério gerados com o *K-means* foi calculado como abrupto (*hard boundary*), enquanto os demais foram consideradas transicionais (*soft boundary*) (Figura 31). Nota-se no gráfico de análise de contato que os subdomínios do *K-means* apresentam médias móveis inferiores às do HC para o estéril (~55%) e semelhantes às do HC (~62%) para o minério, porém com menor variância.

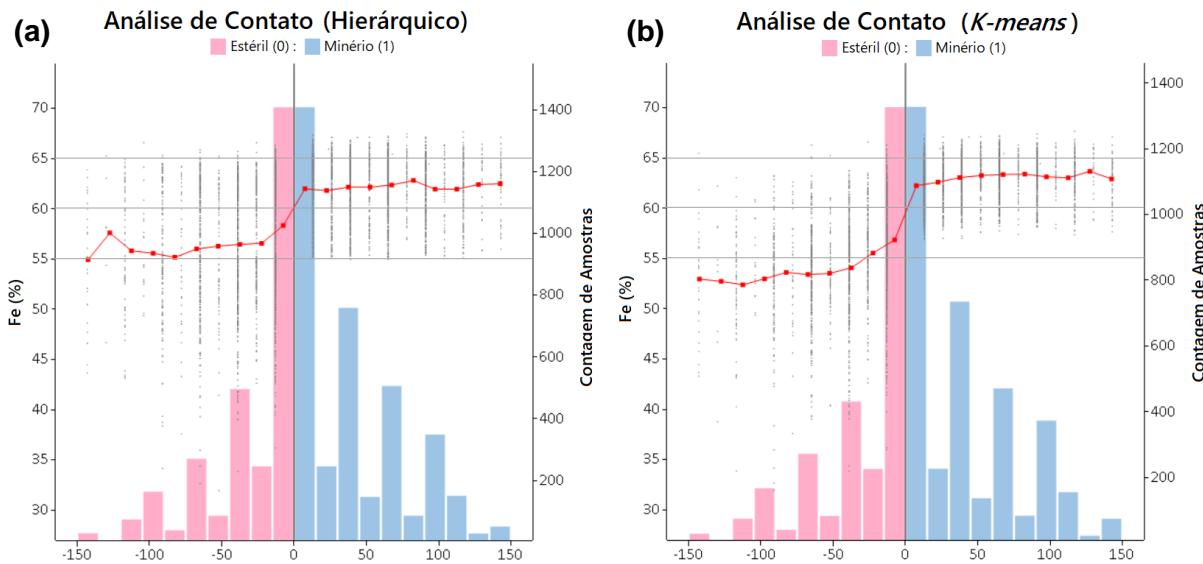


Figura 31: Análises de contato entre os subdomínios estéril (0) e minério (1) para **(a)** o algoritmo HC e **(b)** o algoritmo *K-means*.

6.2. Análise Geoestatística

O principal plano da elipsoide de continuidade do ferro, definido pelos eixos D1 218/45º e D2 131/04º no cenário de domínio único, é paralelo à xistosidade S₂ do Sinforme Córrego das Flechas (Figura 32), flanco leste do Sinclinal Ouro Fino.

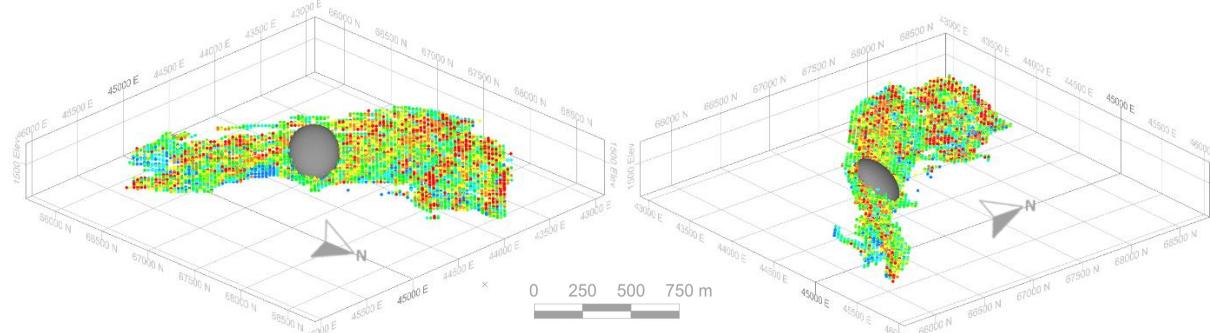


Figura 32: Elipsoide de continuidade do ferro no domínio único.

A análise geoestatística para os cenários de agrupamento resultou em variogramas de direções próximas às verificadas no domínio único (Apêndice VI) e, em geral, os eixos D1 e D2 apresentam maiores alcances (Figura 33). Os variogramas experimentais, entretanto, são pior estruturados, fato que pode ser atribuído à redução da quantidade de amostras disponíveis para análise. Embora todos os cenários preservem o principal plano de continuidade, pode-se observar no Apêndice VII que apenas o *K-means* apresenta D1 na direção NE-SW (como no domínio único).

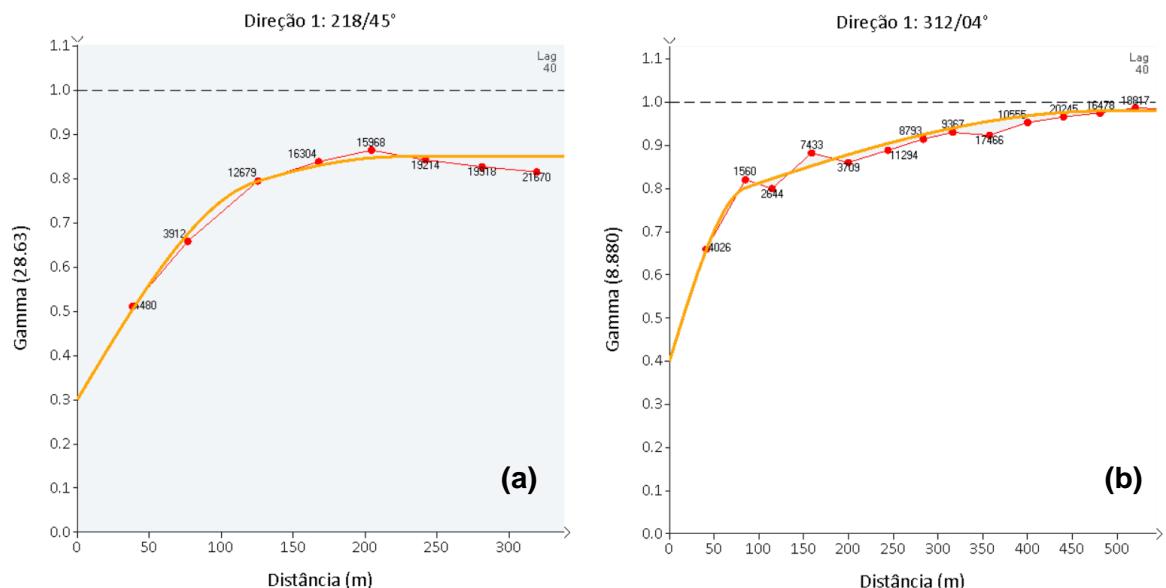


Figura 33: Variogramas experimentais e respectivos modelos teóricos na direção de maior continuidade do ferro para (a) o domínio único e (b) o subdomínio de minério do HC. Eixo vertical: variância relativa à variância a priori (γ).

6.3. Validação e Avaliação do Efeito do Agrupamento nas Estimativas

Embora os cenários de agrupamento tenham apresentado resultados melhores que o cenário de domínio único na validação cruzada, ressalta-se que o método não reflete exatamente a acurácia e precisão das estimativas dos blocos. Entre os motivos, destacam-se a diferença de suporte (estimativa pontual *versus* estimativa de bloco discretizado), a posição relativa das amostras ao centroide de cada célula e entre as amostras; e o agrupamento dos blocos em três subdomínios e das amostras em dois. Portanto, a validação cruzada não considera o subdomínio de transição, de modo que os erros e os coeficientes de correlação para os cenários de agrupamento podem ser ainda mais semelhantes aos do cenário de domínio único do que o observado.

Embora o *K-means* e os demais métodos de agrupamentos apresentem melhores resultados estatísticos para as estimativas, é o cenário de domínio único que retorna o resultado espacial de menor variabilidade e a distribuição de teores estimados com maior continuidade (Figura 34). Os resultados do cenário de agrupamento do GHC são os que mais se assemelham ao uso de domínio único em todos os métodos de validação utilizados (validação cruzada, análise de deriva e comparação das estatísticas descritivas), além destes cenários apresentarem as elipsóides de continuidade mais parecidas.

Portanto, a estimativa com domínio único pode ser mais adequada para o planejamento a longo prazo, no qual se faz importante a compreensão dos teores médios e sua distribuição geral. Já as estimativas com definição de subdomínios devem ter aplicação mais útil no planejamento de curto prazo por reduzir o efeito de suavização, o que teria implicação na

otimização do avanço da frente de lavra e no controle de teores da pilha de minério de acordo com os estabelecidos para o *Run of Mine* da mina Capanema (Tabela 1).

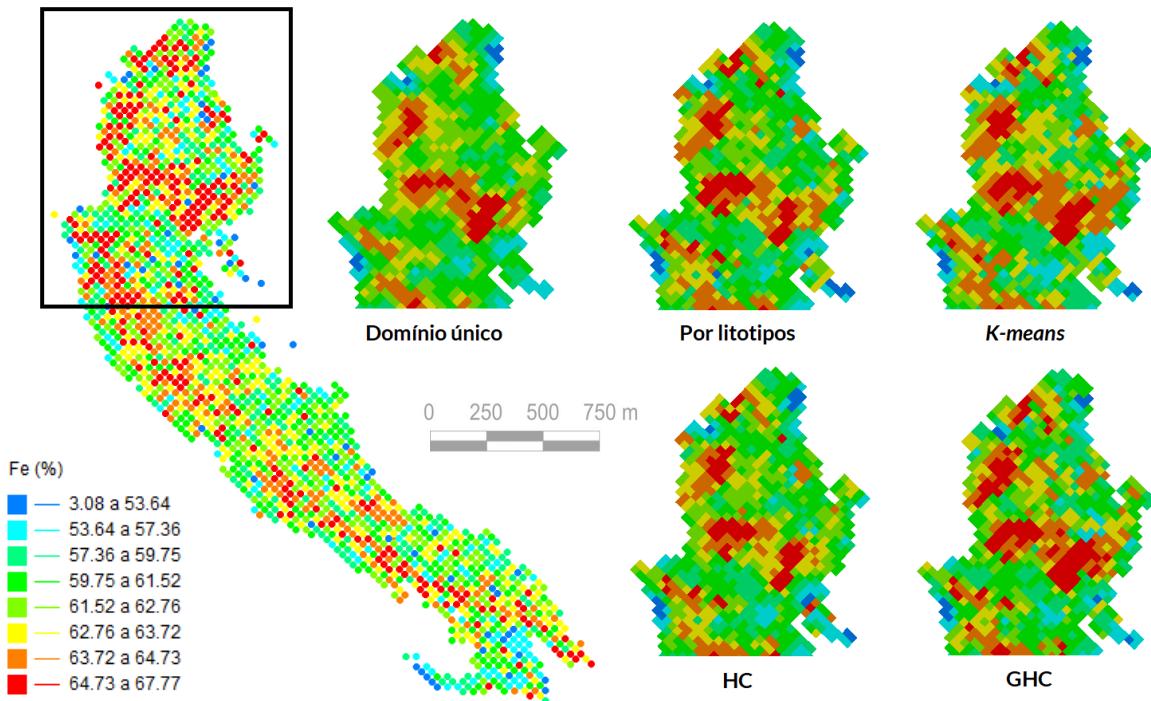


Figura 34: Comparação entre os teores das amostras e do modelo de blocos nos diferentes cenários de agrupamento no setor norte do depósito.

7. CONCLUSÕES

Pode-se concluir que a análise exploratória de dados foi essencial para compreender a malha utilizada e a distribuição dos teores no depósito e como todas as variáveis se correlacionam, o que permitiu a tomada de decisões nas demais etapas, como: a dimensão, rotação e posição do modelo de blocos, a escolha das variáveis-critério para o agrupamento com análise de agrupamento, a realização do próprio agrupamento baseado nos litotipos e a definição dos parâmetros de vizinhança para estimativa.

Ainda, o uso do método do cotovelo para escolha da quantidade ótima de agrupamentos com o algoritmo *K-means* foi satisfatória para os objetivos do estudo e adequada também para os algoritmos aglomerativos hierárquicos. Os resultados de agrupamento de todos se assemelharam, tanto estatística quanto espacialmente, e distinguiram: um *cluster* com alto teor de ferro e baixos teores de alumina e fósforo, dois *clusters* com alta concentração de contaminantes e um *cluster* pobre em ferro. Após a reclassificação dos *clusters* nos subdomínios minério ou estéril, verificou-se que o agrupamento manual dos litotipos também retornou resultados semelhantes aos dos algoritmos.

A análise geoestatística revelou forte controle estrutural-litológico na distribuição dos teores de ferro, cujo plano de maior continuidade é paralelo à xistosidade S_2 do Sinforme

Córrego das Flechas, que corresponde ao corpo de minério. Este resultado é coerente com o modelo genético de FFBs do tipo Lago Superior e com a evolução geológica do QF. A redução da quantidade de amostras disponíveis para análise estrutural nos cenários de subdomínio resultou em variogramas com maiores alcances, porém pior estruturados.

O agrupamento das amostras conferiu, no geral, maior acurácia e precisão para as estimativas de todos os cenários de agrupamento em relação ao cenário de domínio único, que resultou no modelo de menor variabilidade pela suavização mais expressiva dos teores inferiores e superiores a ~57%. Destaca-se o *K-means* como método que retornou os melhores resultados de validação, podendo ser considerado um método de alta eficácia na discriminação estatística de *clusters*. Os ganhos em acurácia e precisão, no entanto, foram pequenos, o que se deve tanto pela baixa variabilidade natural do ferro, quanto pela alta eficácia da krigagem ordinária. Provavelmente, ganhos mais significativos seriam observados com o uso de métodos que apresentam maior efeito de suavização, como os não geoestatísticos ou mesmo a krigagem simples, nos casos em que não houver estacionariedade de primeira ordem e/ou a média global estiver enviesada.

O resultado do algoritmo GHC foi o que mais se aproximou do uso de domínio único, provavelmente por considerar as posições geográficas das amostras e não apenas as relações estatísticas entre as substâncias escolhidas. Sua inserção em um *software* apresenta grandes benefícios para os usuários, tanto por dispensar o conhecimento de linguagens de programação para realizar agrupamentos complexos e altamente configuráveis, quanto por reunir a análise de agrupamento e as demais etapas da avaliação de recursos e reservas minerais em uma única interface. Ressalta-se ainda a vantagem que o algoritmo oferece de atribuir novas amostras a agrupamentos previamente definidos, sem a necessidade de reclassificar todo o conjunto amostral ou recorrer a métodos externos como a Análise Discriminante Múltipla, como aconteceria para os algoritmos *K-means* e Aglomerativo Hierárquico.

Para a mina Capanema, ainda que os ganhos em acurácia e precisão de estimativa não justifiquem o tempo empregado para realizar a definição dos subdomínios, conclui-se que os métodos de agrupamento são eficazes e podem ter útil aplicação em outros contextos de avaliação de recursos e reservas minerais, o que deve ser melhor explorado, como: na avaliação de recursos e reservas com malhas menos densas e/ou não regulares, na avaliação de substâncias com alta variabilidade, em que se esperam contatos abruptos; e para modelos de curto prazo, em que há maior necessidade de detalhamento dos contrastes locais nas estimativas de teores altos e baixos.

8. CONSIDERAÇÕES FINAIS

Com base nos resultados obtidos e tomadas de decisão realizadas ao longo do estudo, recomenda-se para trabalhos futuros:

- i. A investigação dos métodos de agrupamento tradicionais (manuais) e por análise de agrupamento em outros contextos, como em malhas pouco densas e/ou não regulares, para depósitos de substâncias heterogêneas ou para os próprios contaminantes do ferro, de maior variabilidade;
- ii. Verificar se a transformação das variáveis-critério é mais adequada por normalização ou por padronização para aplicar os algoritmos de agrupamento;
- iii. A remoção dos *outliers* inseridos de maneira indesejada nos *clusters* de maior interesse pelos algoritmos de agrupamento, atribuindo-os a outros grupos;
- iv. Investigar as diferenças entre os algoritmos de agrupamento KM, HC e GHC para furos que contenham mais de uma amostra, visto que o GHC considera as distâncias espaciais e, neste caso, deverá apresentar grupos mais contínuos;
- v. Realizar o agrupamento com o algoritmo GHC com outras configurações que não a padrão do software Isatis.neo®, como a atribuição de diferentes pesos para as variáveis-critério escolhidas, a inserção dos litotipos como critério categórico e principalmente a configuração de uma elipsoide de continuidade que corresponda aos modelos teóricos de variograma do ferro e o campo geométrico em questão;
- vi. Tendo em vista a geometria do Sinforme Córrego das Flechas e suas dobras parasitas, investigar a relevância de se considerar anisotropias dinâmicas para realizar as estimativas.

9. REFERÊNCIAS BIBLIOGRÁFICAS

Abzalov, M., 2016, Modern Approaches in Solid Earth Sciences: Applied Mining Geology, v. 12, New York, Springer International Publishing Switzerland, doi:10.1007/978-3-319-39264-6.

Arthur, D., Vassilvitskii, S., 2007, k-means++: The advantages of careful seeding: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics Philadelphia, USA, p. 1027-1035.

Dardenne, M., Schobbenhaus, C., 2001, Metalogênese do Brasil, Brasília, Editora Universidade de Brasília, 392 p.

Faraj, F., Ortiz, J., 2021, A Simple Unsupervised Classification Workflow for Defining Geological Domains Using Multivariate Data: Mining, Metallurgy & Exploration, v. 38, New York, Springer International Publishing Switzerland, doi:10.1007/s42461-021-00428-5.

Fonseca, L., 2014, Avaliação de Métodos de Perfilagem Geofísica na Pesquisa de Minério de Ferro - Estudo de Caso: Definição de Contatos Litológicos na Mina de Capanema, MG [Dissertação de Mestrado]: Ouro Preto, Universidade Federal de Ouro Preto, Escola de Minas, 125 p.

Franco, A., Endo, I., 2004, Sinclinal Ouro Fino revisitado, quadrilátero ferrífero, Minas Gerais: uma hipótese sobre a sua origem e evolução: Revista Brasileira de Geociências, v. 34, n. 2, p. 167-174.

Fouedjio, F., 2016, A hierarchical clustering method for multivariate geostatistical data: Spatial Statistics, v. 18, p. 333–351, doi:10.1016/j.spasta.2016.07.003.

Goovaerts, P., 1997, Geostatistics for natural resources evaluation: New York, Oxford University Press, 483 p.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013, An Introduction to Statistical Learning with Applications in R: New York, Springer Texts in Statistics, 436 p.

Kassambara, A., 2017, Practical guide to cluster analysis in R: unsupervised machine learning: Journal of Computational and Graphical Statistics, Sthda, v. 1, 187 p.

MacQueen, J., 1967, Some methods for classification and analysis of multivariate observations: Fifth Berkeley Symposium on Mathematical Statistics and Probability, v. 1, n. 14, p. 281-297.

Matheron, G., 1963, Principles of geostatistics: Economic geology, v. 58, p. 1246-1266, doi.org/10.2113/gsecongeo.58.8.1246.

Moreira, G.C., Costa, J.F.C.L, Marques, D.M., 2020, Defining geologic domains using cluster analysis and indicator correlograms: a phosphate-titanium case study: Applied Earth Science: Transactions of the Institute of Mining and Metallurgy, v. 129, p. 176–190, doi:10.1080/25726838.2020.1814483.

Noce, C., 1995, Geocronologia dos Eventos Magmáticos, Sedimentares e Metamórficos na Região do Quadrilátero Ferrífero, Minas Gerais [Tese de Doutorado]: São Paulo, Universidade de São Paulo, Instituto de Geociências, 128 p.

Pedregosa, F. et al., 2011, Scikit-learn: Machine Learning in Python: Journal of Machine Learning Research 12, p. 2825-2830.

Pinches, G., 1980, Factors influencing classification results from multiple discriminant analysis: Journal of Business Research, v. 8, p. 429-456, doi: 10.1016/0148-2963(80)90017-X.

Pyrcz, M.J., Deutsch, C.V., 2014, Geostatistical reservoir modeling: New York, Oxford University Press, 433 p.

Rocha, M.M., 1999, Estudo Comparativo entre os Métodos Computacionais de Avaliação de Reservas e Controle de Teores da Mina Capanema, Santa Bárbara, MG [Dissertação de Mestrado]: São Paulo, Universidade de São Paulo, Instituto de Geociências, 134 p.

Romary, T., Ors, F., Rivoirard, J., Deraisme, J., 2015, Unsupervised classification of multivariate geostatistical data: Two algorithms: Computers & Geosciences, v. 85, p. 96–103, doi:10.1016/j.cageo.2015.05.019.

Rosière, C.A., Chemale Jr, F., 2000, Itabiritos E Minérios De Ferro De Alto Teor Do Quadrilátero Ferrífero – Uma Visão Geral E Discussão: Geonomos, v. 8, p. 27–43, doi:10.18285/geonomos.v8i2.155.

Rossi, M.E., Deutsch, C.V., 2014, Mineral Resource Estimation: New York, Springer International Publishing Switzerland, 332 p., doi:10.1007/978-1-4020-5717-5.

Serviço Geológico do Brasil, 2004, Serviço de mapas geológicos da CPRM na escala 1:1.000.000, GeoSGB. Disponível em: geoportal.cprm.gov.br/geosgb (acessado em 04/2021).

Sinclair, J.A., Blackwell, H.G., 2004, Applied Mineral Inventory Estimation: Cambridge, United Kingdom, Cambridge University Press, 381 p.

Sokal, R.R., Sneath, P.H.A., 1963, Principles of numerical taxonomy: Principles of numerical taxonomy.

Vale, 2012, Nossa História: Rio de Janeiro, Verso Brasil, 420 p.

Vale, 2020, Relatório Anual da Administração, 56 p. Disponível em: <http://www.vale.com/brasil/PT/investors/information-market/annual-reports> (acessado em 01/2022).

Vale, 2021, Relatório de Referência, v. 14, 747 p. Disponível em: vale.com/brasil/PT/investors/information-market/annual-reports/reference-form (acessado em 01/2022).

Vale Notícias, 2021, Ouro Preto e Mariana Novembro/Dezembro 2021, 2 p. Disponível em: vale.com/brasil/PT/sustainability/vale-nas-comunidades/minas-gerais (acessado em 01/2022).

Virtanen, P. et al., 2020, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python: Nature Methods, 17(3), p. 261-272.

Yamamoto, J.K., 2001, Avaliação e Classificação de Reservas Minerais: São Paulo, Edusp, v. 38, 226 p.

Yamamoto, J.K., Landim, P.M.B., 2013, Geoestatística, conceitos e aplicações: São Paulo, SP, Brasil, Oficina de Textos, 215 p.

Yuan, C., Yang, H., 2019, Research on K-value selection method of K-means clustering algorithm: J Multidisciplinary Scientific Journal, v. 2, p. 266-235.

APÊNDICE I - ESTATÍSTICAS DESCRIPTIVAS POR LITOTIPO

Apêndice I-A) Estatísticas descritivas de ferro nos dados validados e tratados.

Medida	HA	HB	SI	SIA	SIP	WH
Quantidade	2065	926	898	170	1202	1458
Mínimo	43,70	59,65	50,47	35,98	3,08	26,97
Máximo	67,67	64,10	59,99	63,90	59,68	65,57
Média	64,55	61,67	57,97	59,97	51,11	60,59
Mediana	64,46	61,78	58,01	60,64	52,23	61,44
Variância	1,13	0,80	1,59	9,37	22,56	9,97
Desvio Padrão	1,06	0,89	1,26	3,06	4,75	3,16
Coef. de Variação	0,02	0,01	0,02	0,05	0,09	0,05
Assimetria	-3,38	-0,20	-0,43	-3,55	-3,39	-3,33
Curtose	70,57	-1,11	0,46	21,75	24,27	22,01

Apêndice I-B) Estatísticas descritivas de alumina nos dados validados e tratados.

Medida	HA	HB	SI	SIA	SIP	WH
Quantidade	2065	926	898	170	1202	1458
Mínimo	0,02	0,03	0,00	0,21	0,04	0,15
Máximo	4,05	3,57	11,51	8,46	3,62	20,20
Média	1,28	0,91	0,73	2,18	0,65	3,25
Mediana	1,27	0,82	0,61	1,89	0,51	2,86
Variância	0,28	0,20	0,35	1,83	0,18	2,31
Desvio Padrão	0,53	0,44	0,60	1,35	0,42	1,52
Coef. de Variação	0,42	0,49	0,81	0,62	0,65	0,47
Assimetria	0,24	0,84	10,25	1,45	1,85	3,52
Curtose	-0,27	1,18	166,52	3,03	5,25	25,23

Apêndice I-C) Estatísticas descritivas de fósforo nos dados validados e tratados.

Medida	HA	HB	SI	SIA	SIP	WH
<i>Quantidade</i>	2065	926	898	170	1202	1458
<i>Mínimo</i>	0,00	0,00	0,00	0,00	0,00	0,00
<i>Máximo</i>	6,43	4,41	4,49	7,57	3,25	10,61
<i>Média</i>	0,12	0,10	0,10	0,59	0,06	0,15
<i>Mediana</i>	0,06	0,05	0,05	0,09	0,04	0,08
<i>Variância</i>	0,27	0,13	0,16	2,98	0,05	0,55
<i>Desvio Padrão</i>	0,52	0,35	0,40	1,72	0,22	0,74
<i>Coef. de Variação</i>	4,26	3,60	4,00	2,91	3,47	4,78
<i>Assimetria</i>	8,21	7,96	7,73	3,19	11,62	10,98
<i>Curtose</i>	70,41	67,50	62,36	8,50	140,85	125,10

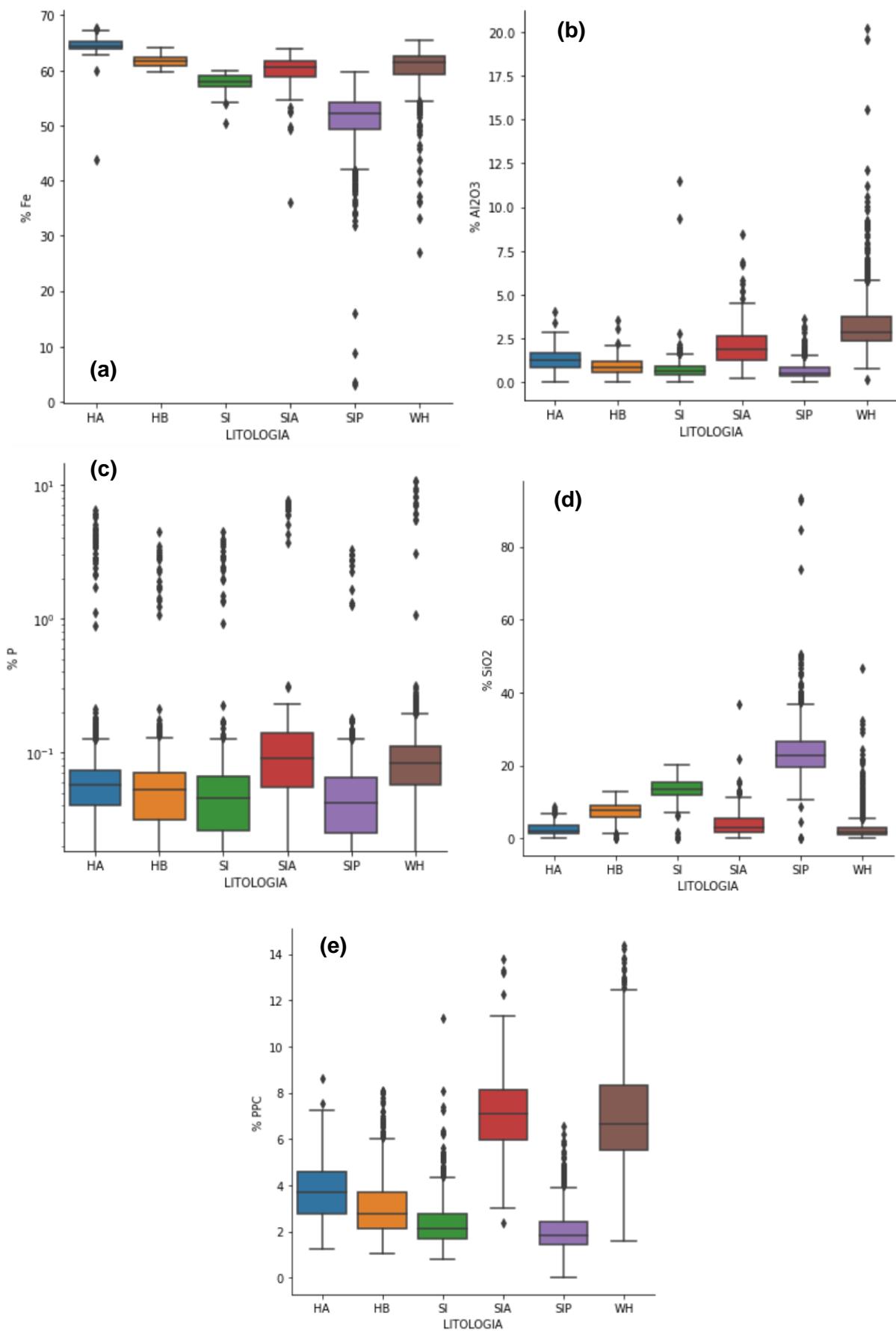
Apêndice I-D) Estatísticas descritivas de sílica nos dados validados e tratados.

Medida	HA	HB	SI	SIA	SIP	WH
<i>Quantidade</i>	2065	926	898	170	1202	1458
<i>Mínimo</i>	0,00	0,00	0,00	0,00	0,00	0,00
<i>Máximo</i>	8,53	12,90	20,23	36,85	93,30	46,57
<i>Média</i>	2,49	7,43	10,51	4,42	23,88	2,75
<i>Mediana</i>	2,01	7,65	13,60	3,05	22,58	1,59
<i>Variância</i>	2,66	5,63	7,38	18,74	49,96	12,28
<i>Desvio Padrão</i>	1,63	2,37	2,72	4,33	7,07	3,50
<i>Coef. de Variação</i>	0,66	0,32	0,20	0,98	0,30	1,28
<i>Assimetria</i>	0,99	-0,42	-1,48	3,34	2,89	4,63
<i>Curtose</i>	0,19	-0,10	5,84	18,38	21,85	33,01

Apêndice I-E) Estatísticas descritivas de PPC nos dados validados e tratados.

Medida	HA	HB	SI	SIA	SIP	WH
<i>Quantidade</i>	2065	926	898	170	1202	1458
<i>Mínimo</i>	1,22	1,06	0,80	2,35	0,00	1,57
<i>Máximo</i>	8,60	8,10	11,24	13,79	6,53	14,37
<i>Média</i>	3,73	3,08	2,36	7,18	2,03	6,97
<i>Mediana</i>	3,71	2,78	2,13	7,10	1,85	6,67
<i>Variância</i>	1,44	1,71	0,96	3,88	0,69	4,80
<i>Desvio Padrão</i>	1,20	1,31	0,98	1,97	0,83	2,19
<i>Coef. de Variação</i>	0,32	0,42	0,41	0,27	0,41	0,31
<i>Assimetria</i>	0,25	1,17	2,26	0,59	1,43	0,40
<i>Curtose</i>	-0,54	1,22	10,64	0,87	3,25	0,08

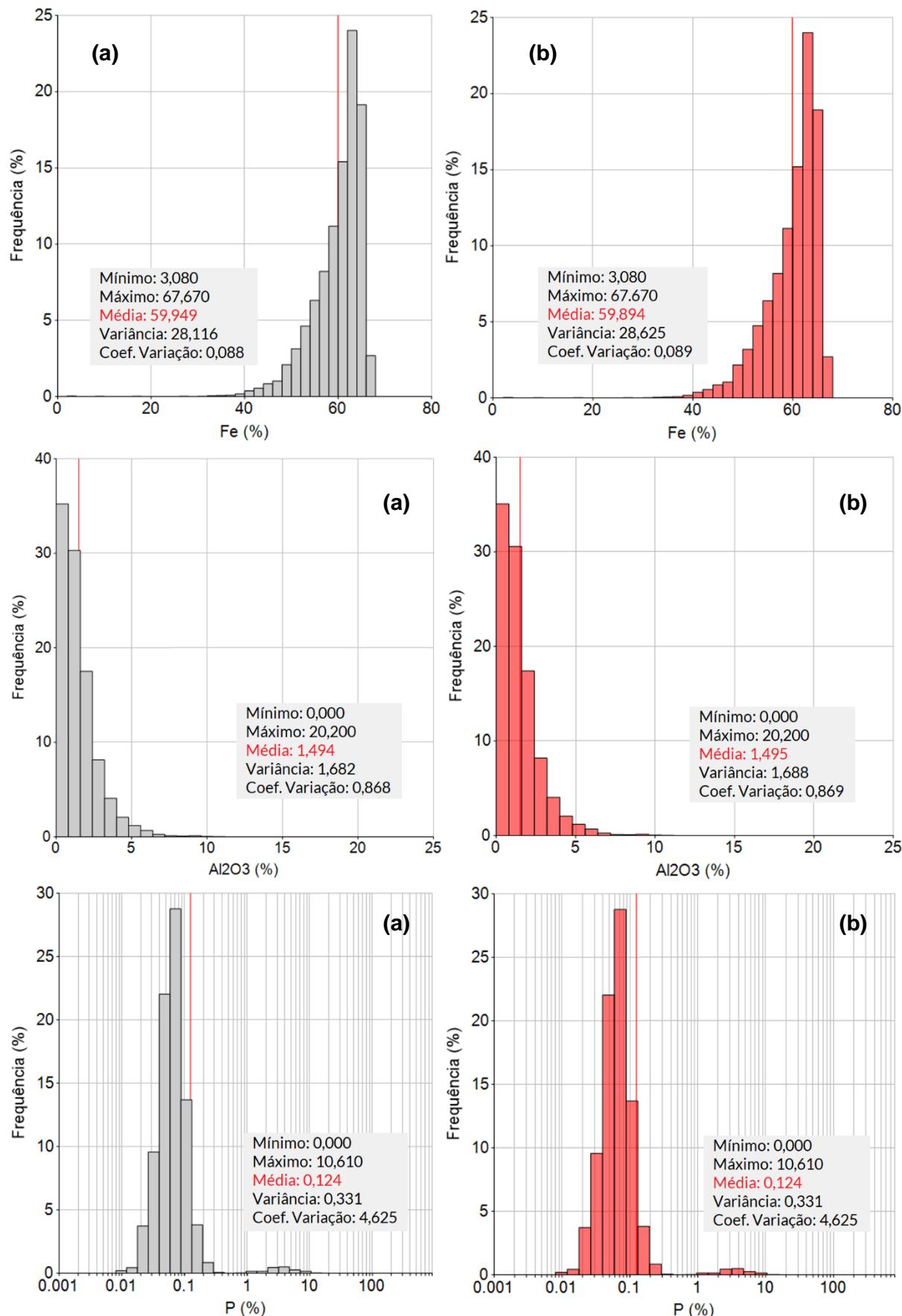
APÊNDICE II – BOXPLOTS POR LITO TIPO

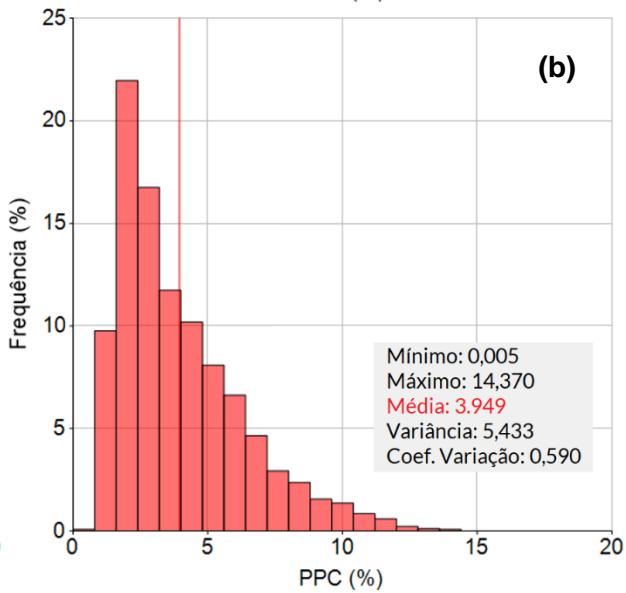
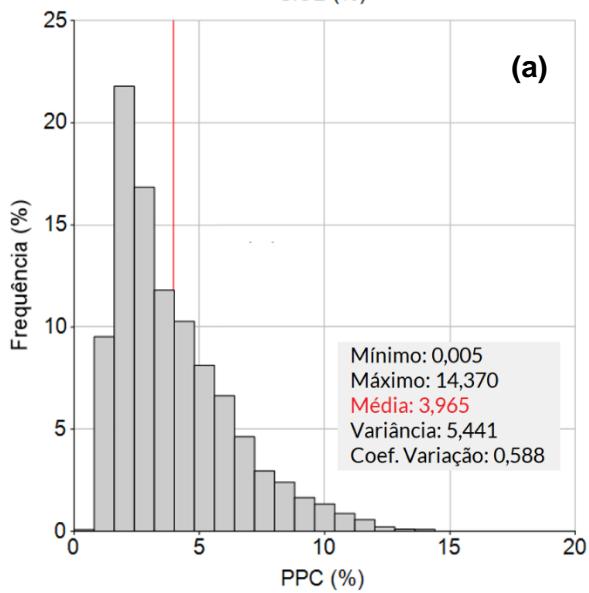
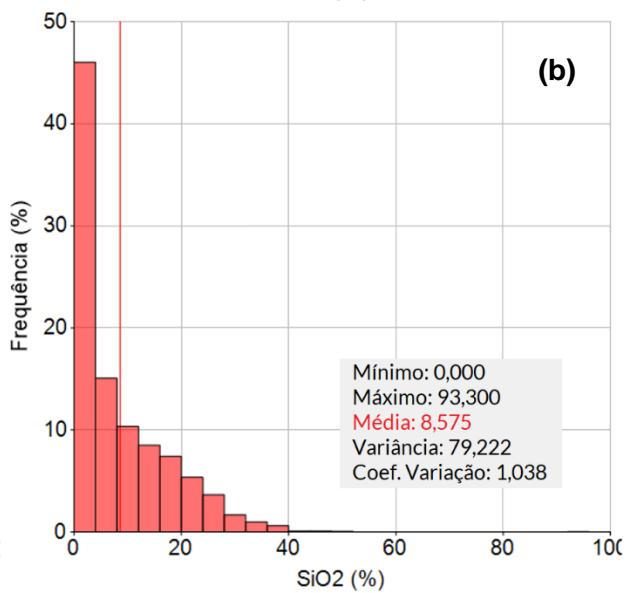
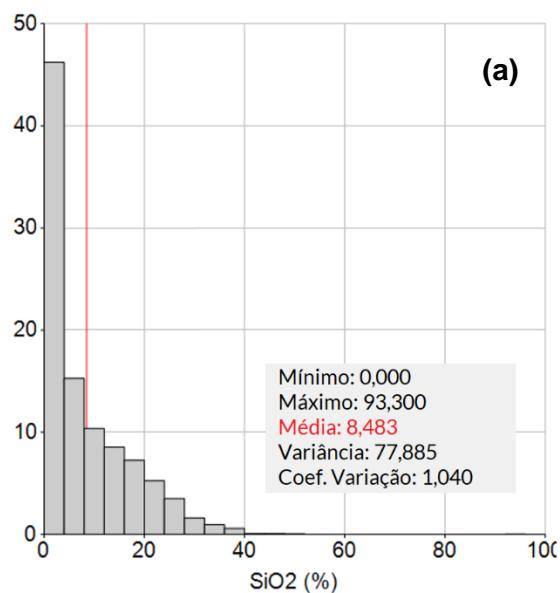


Apêndice II) Boxplots dos teores de **(a)** ferro, **(b)** alumina, **(c)** fósforo e **(d)** sílica e **(e)** da perda por calcinação do banco de dados validado e tratado.

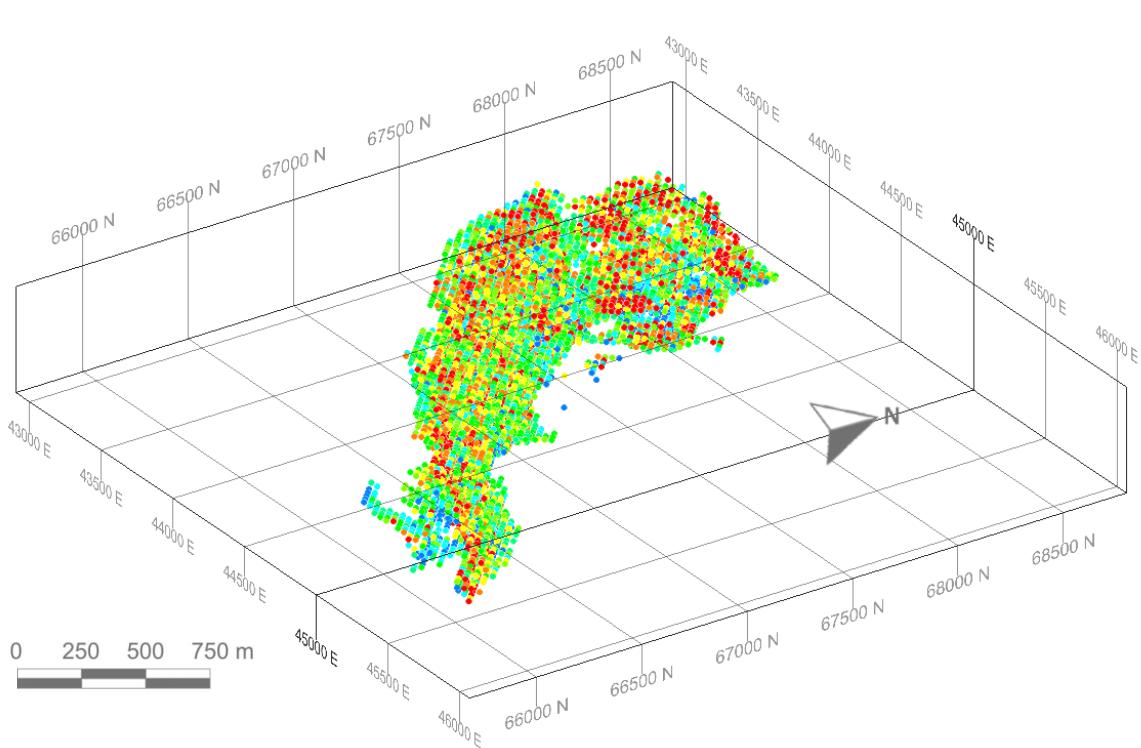
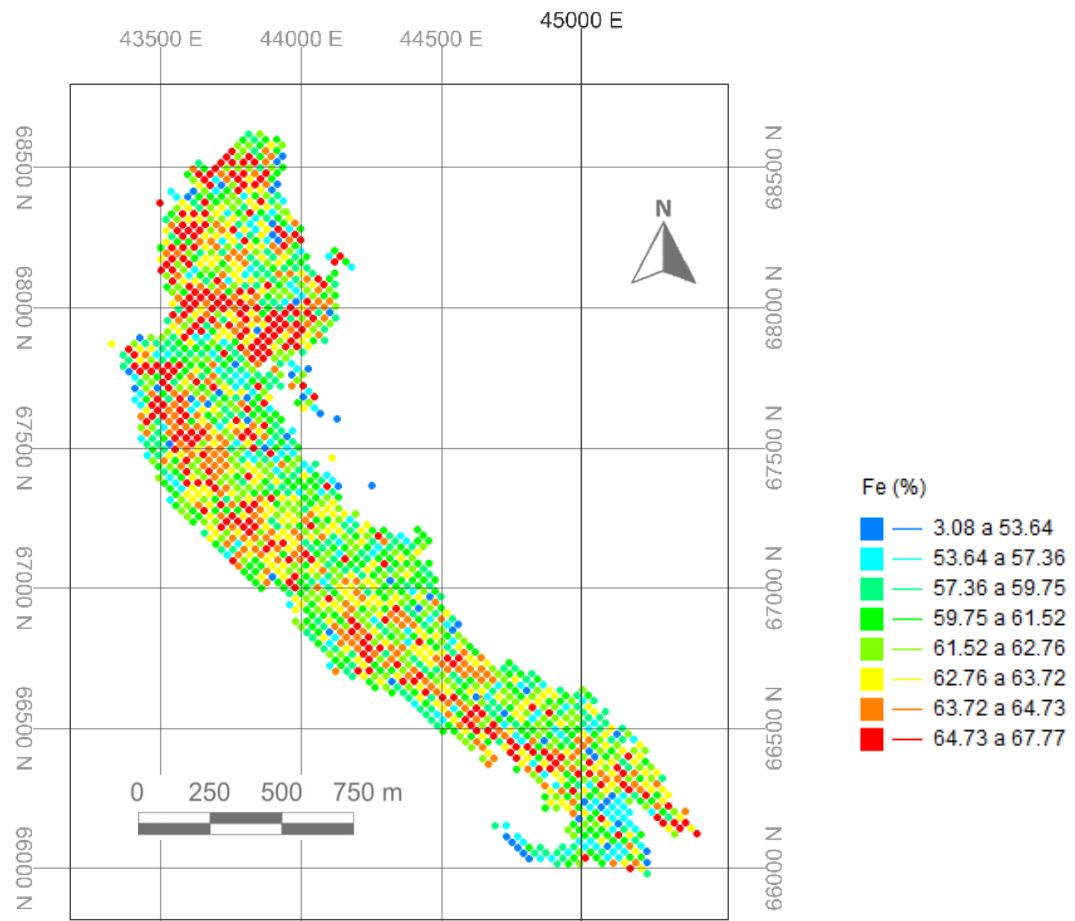
APÊNDICE III – HISTOGRAMAS

Apêndice III) Histograma dos teores de ferro, alumina, fósforo, sílica e da porcentagem de PPC para **(a)** o banco de dados validado e **(b)** o banco de dados validado e tratado.

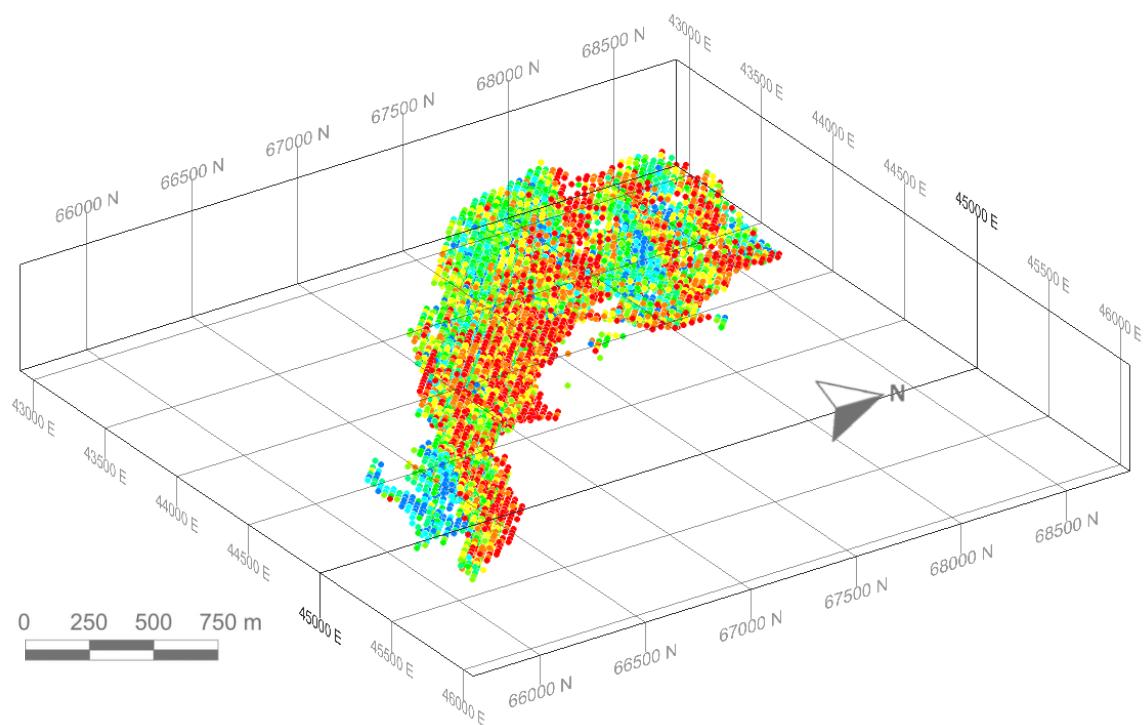
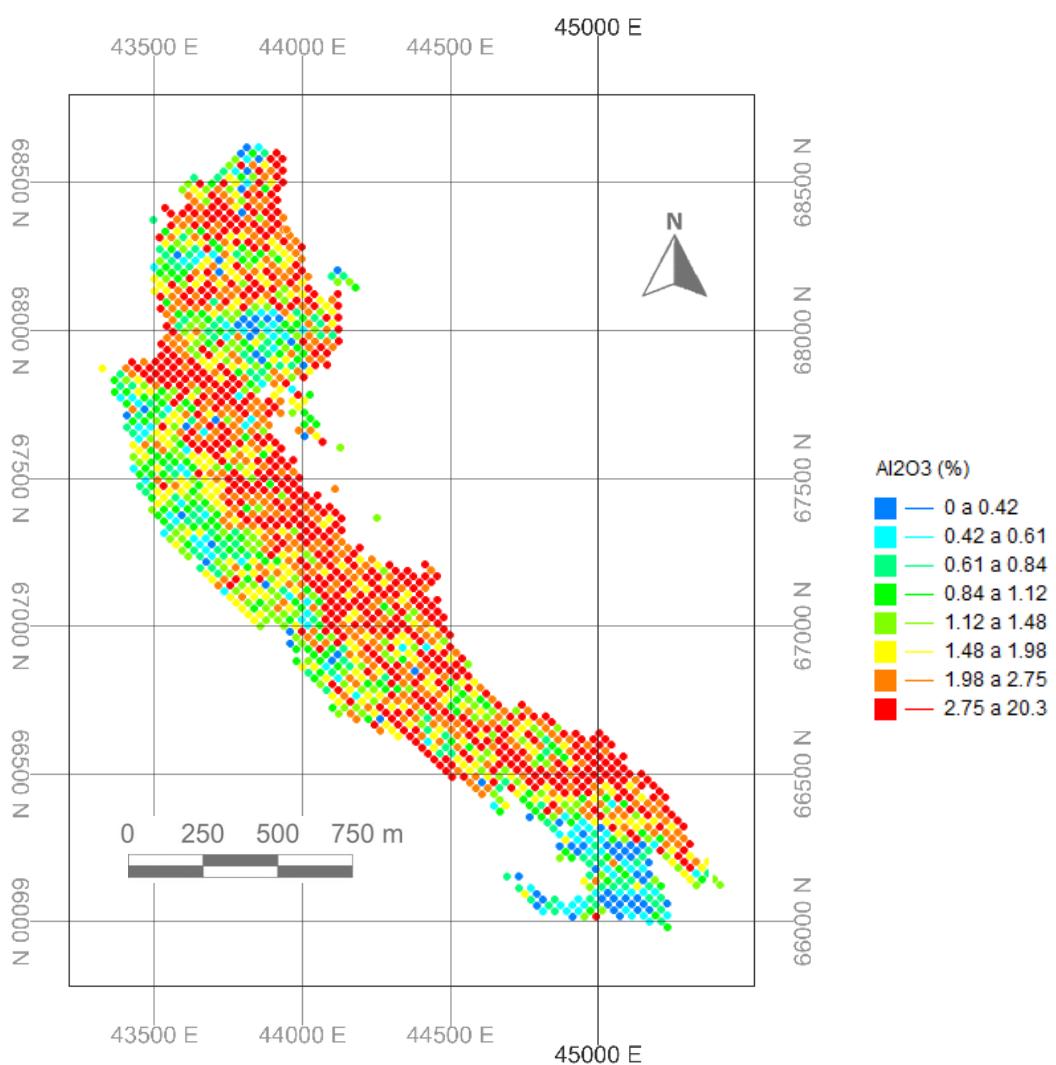




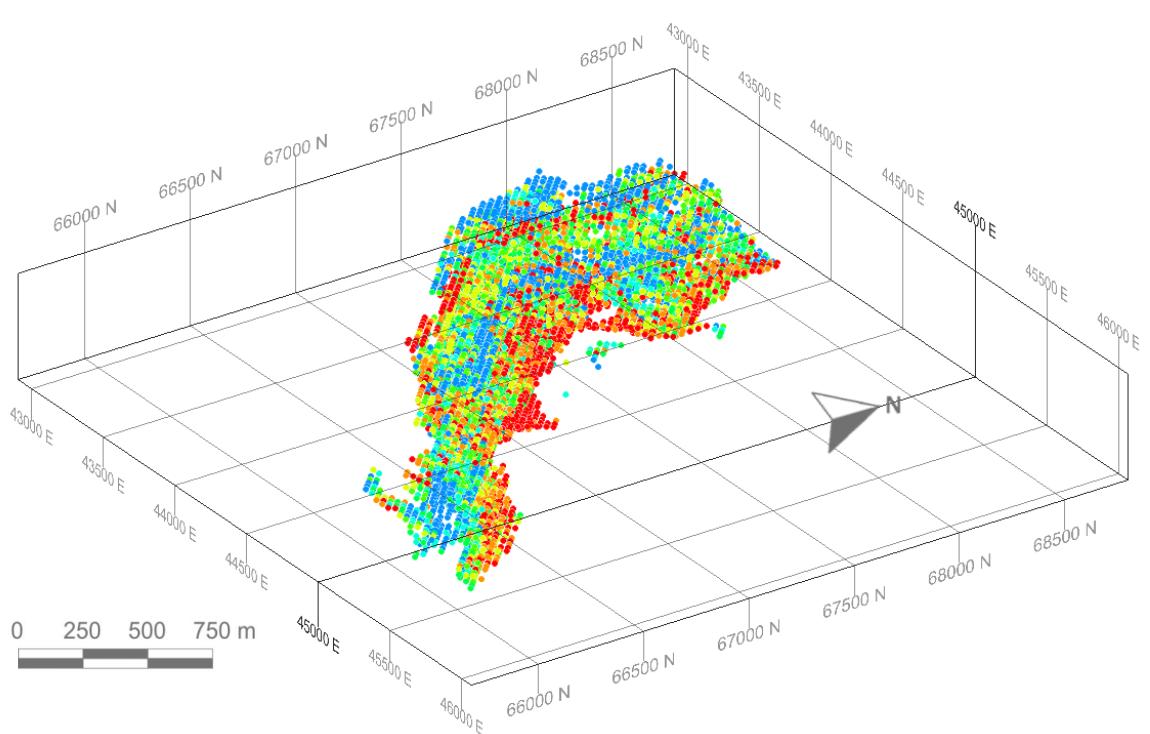
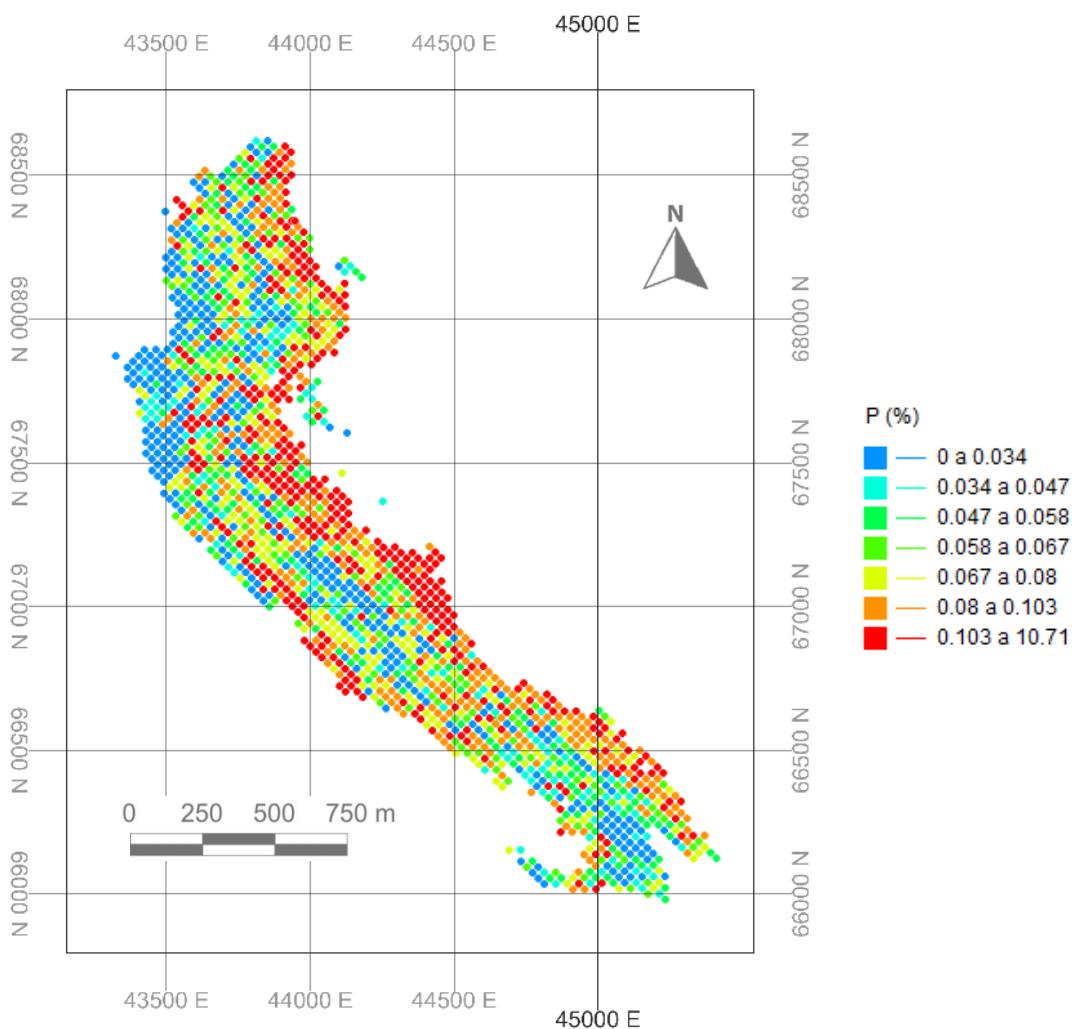
APÊNDICE IV – MALHA COM LEGENDAS DE TEORES E PPC



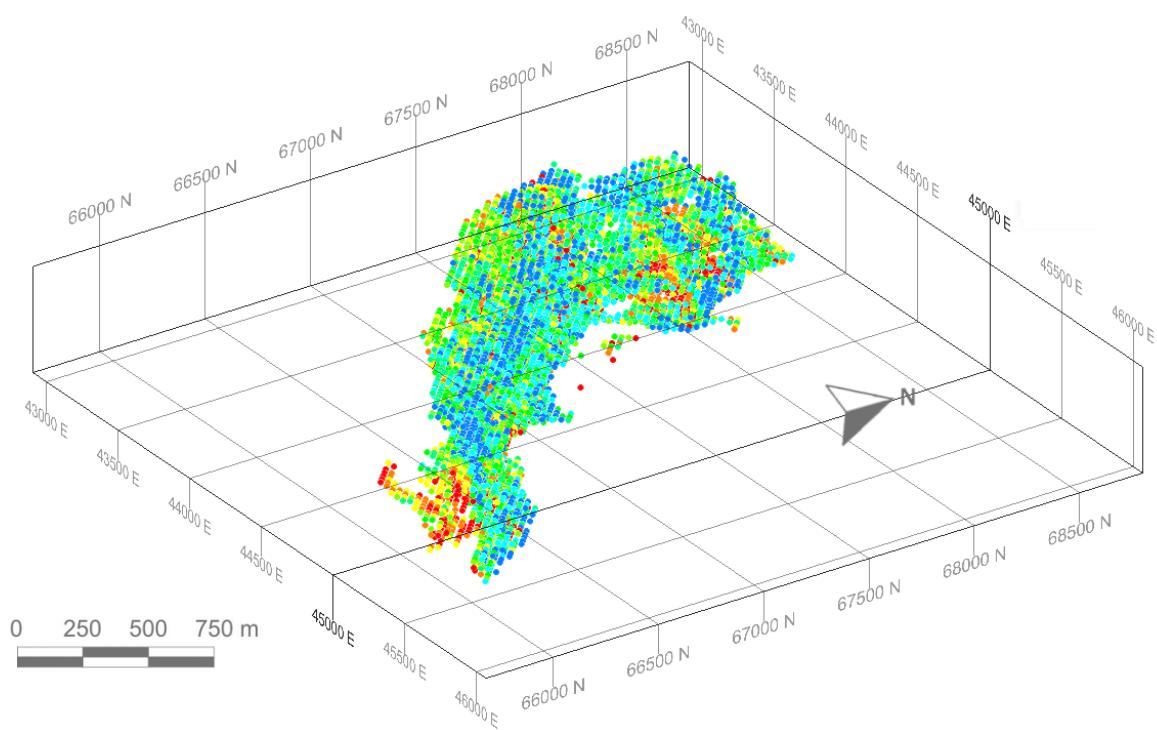
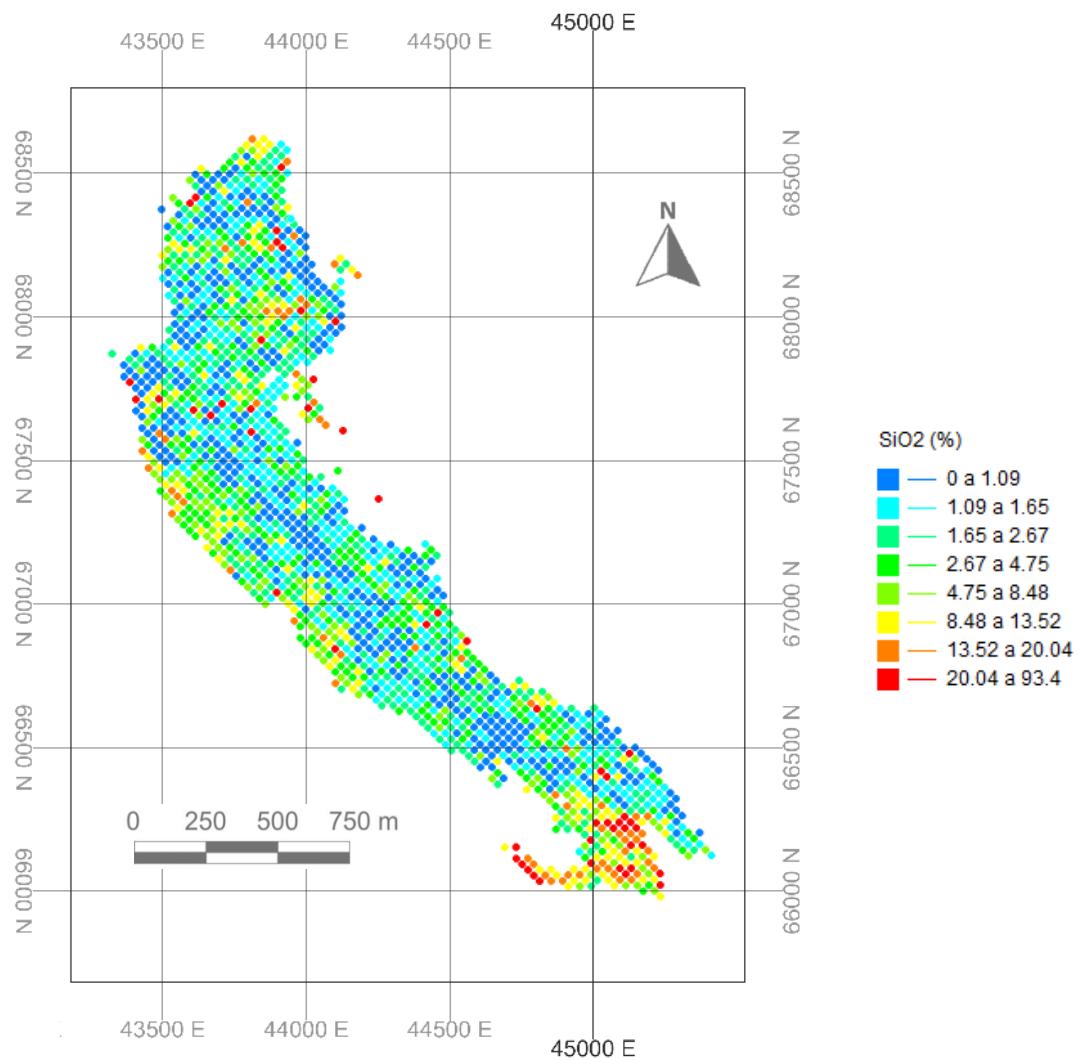
Apêndice IV-A) Malha com legenda dos teores de ferro em planta e em perspectiva.



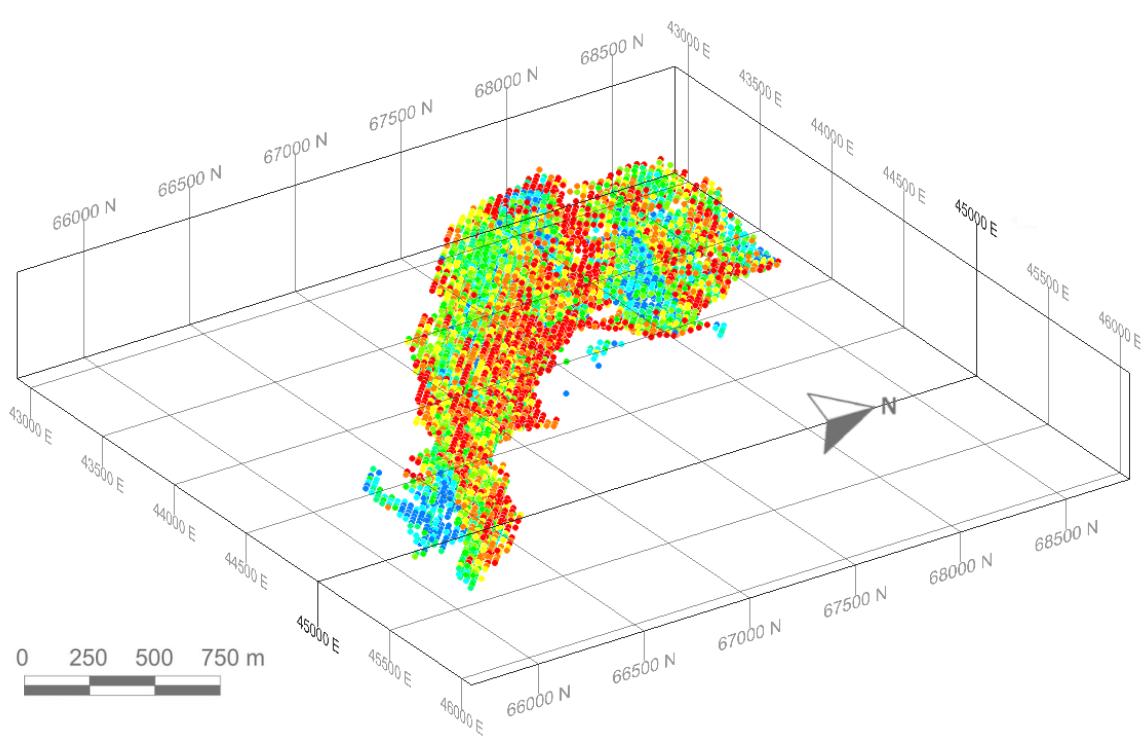
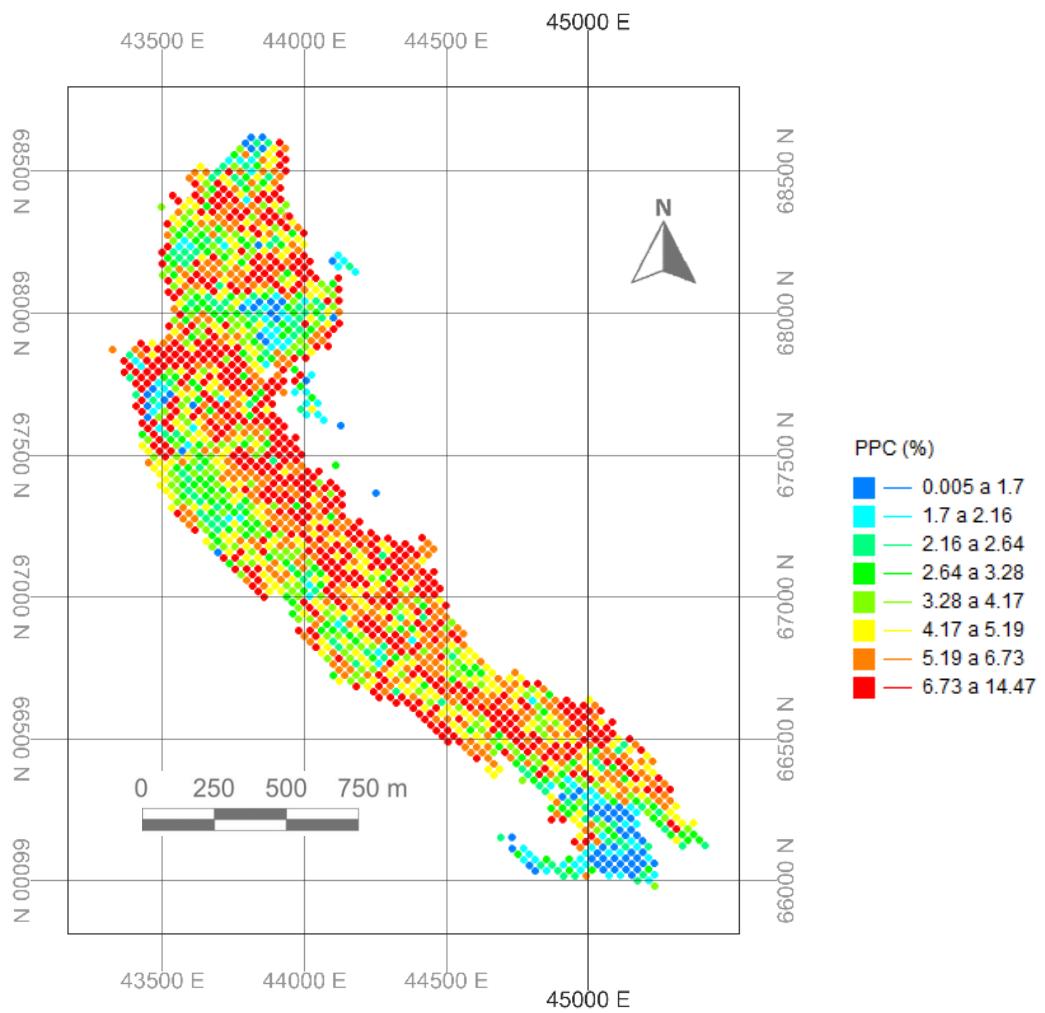
Apêndice IV-B) Malha com legenda de teor de alumina em planta e em perspectiva.



Apêndice IV-C) Malha com legenda de teor de fósforo em planta e em perspectiva.



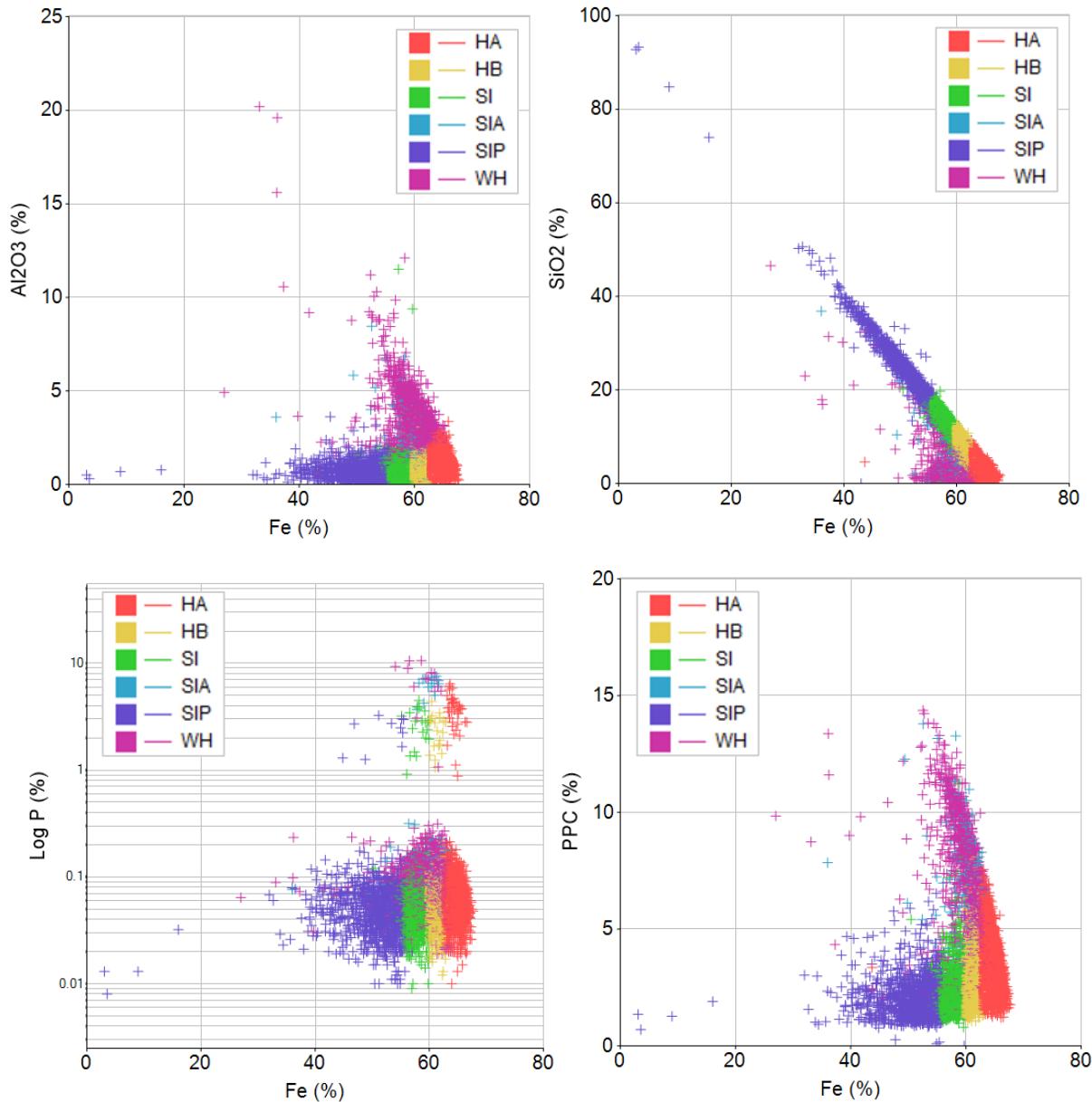
Apêndice IV-D) Malha com legenda de teor de sílica em planta e em perspectiva.

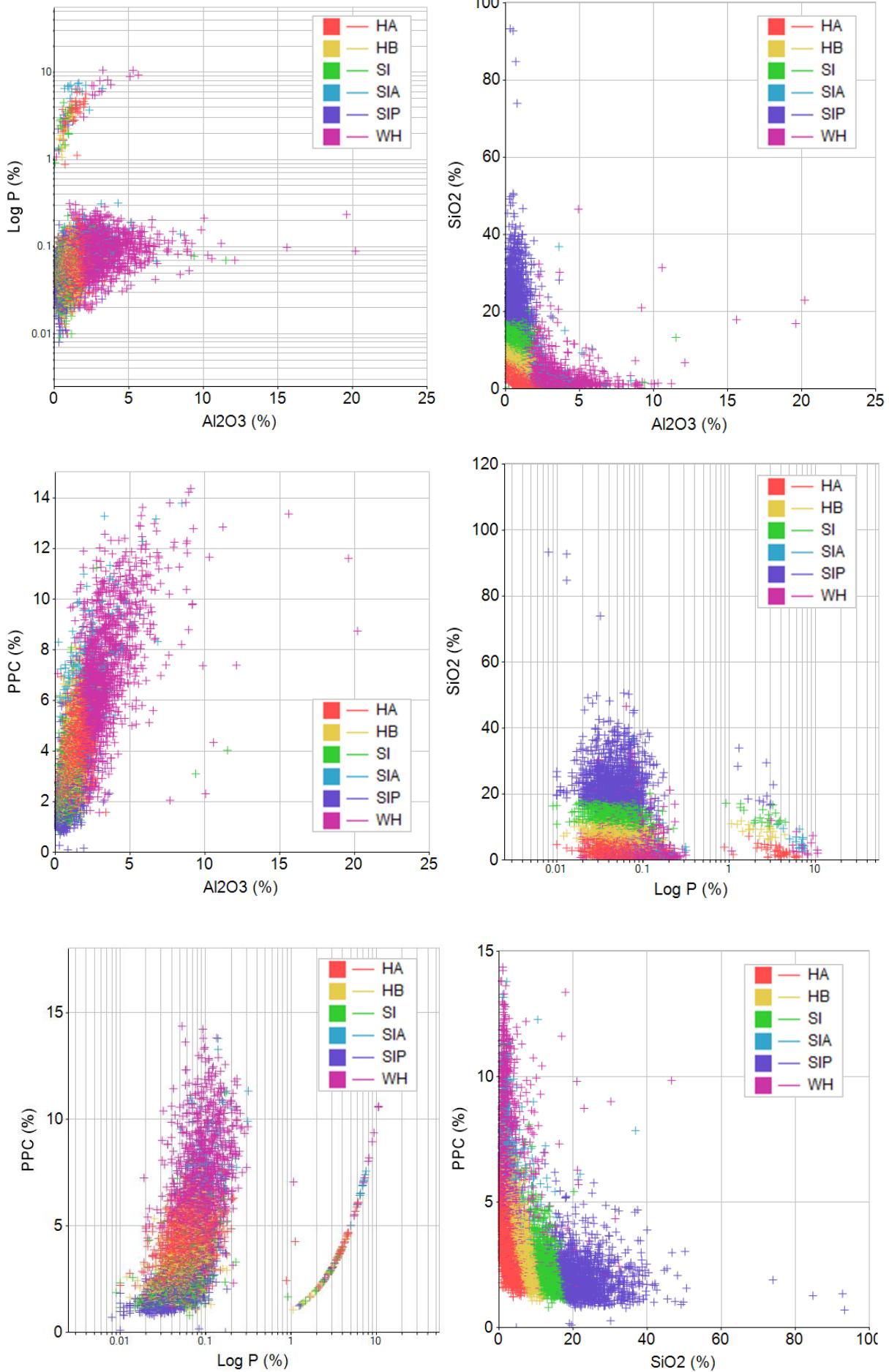


Apêndice IV-E) Malha com legenda de PPC em planta e em perspectiva.

APÊNDICE V – GRÁFICOS DE DISPERSÃO

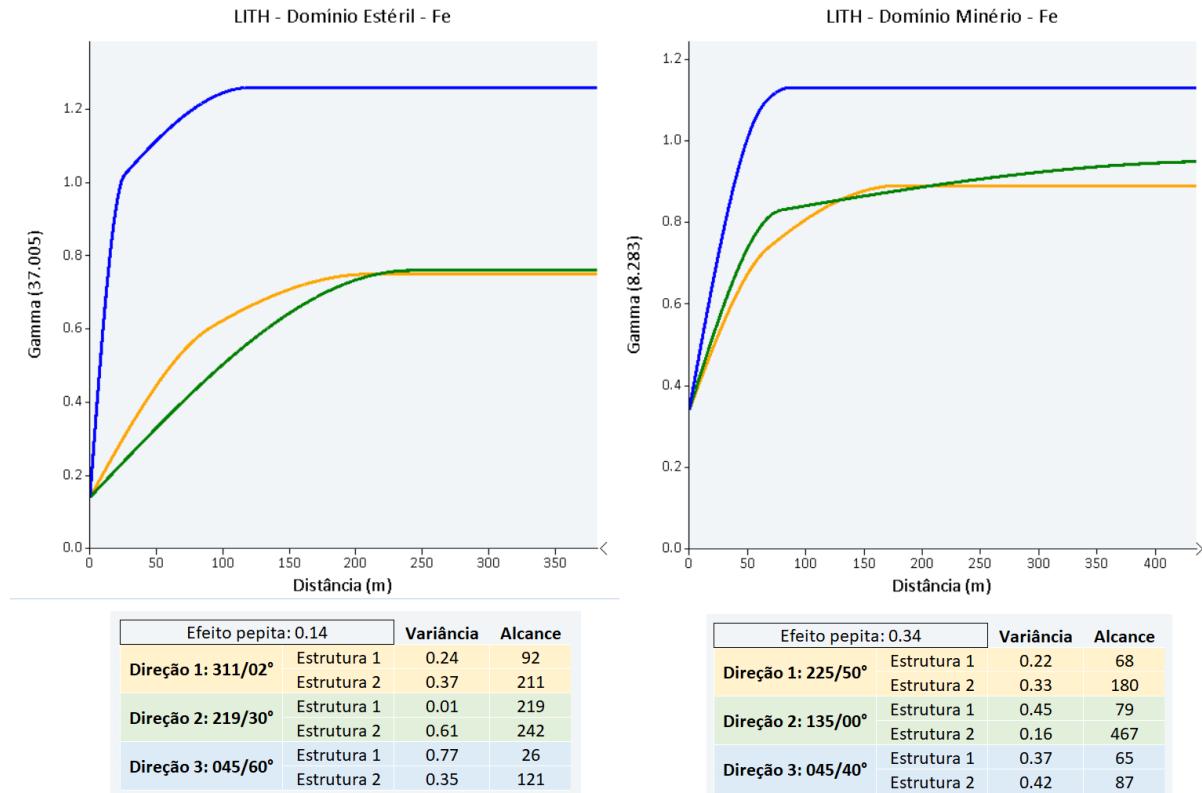
Apêndice V) Gráficos de dispersão entre todas as variáveis numéricas (Fe, Al₂O₃, P, SiO₂ e PPC) com legenda de litotipos.



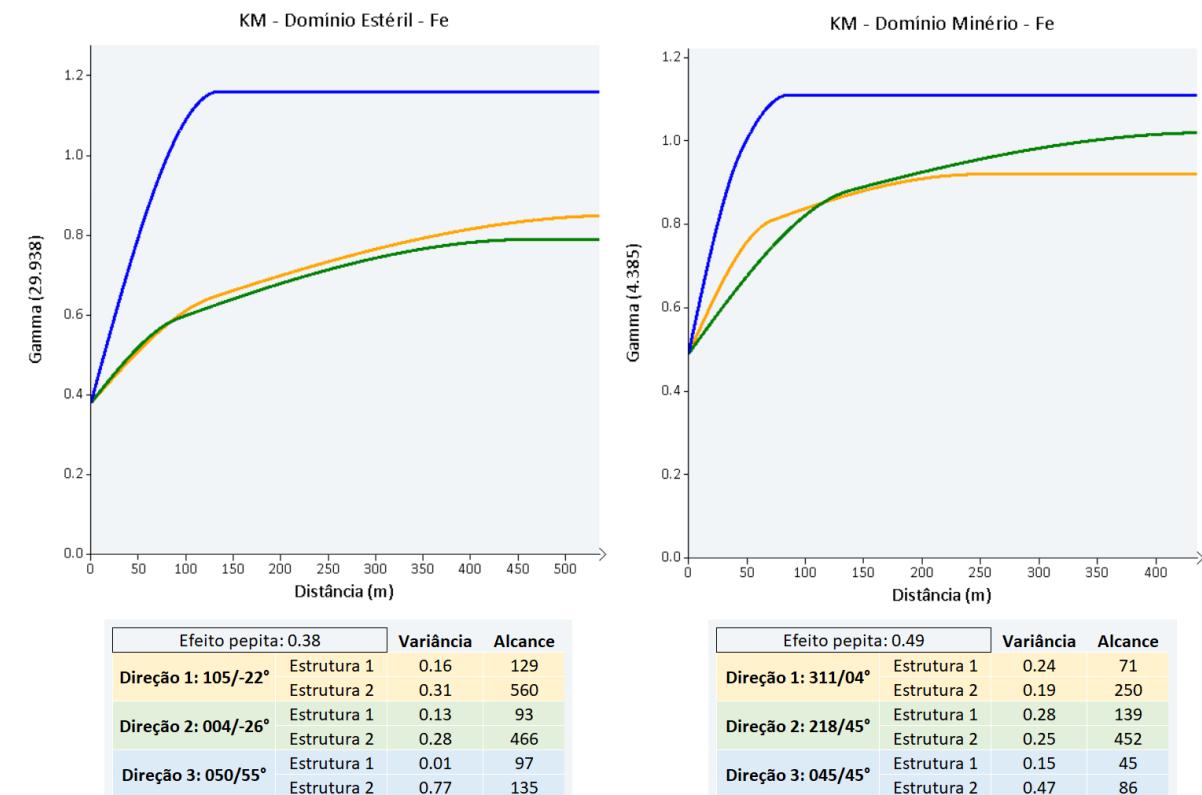


APÊNDICE VI – MODELOS DE VARIOGRAMA DO FERRO

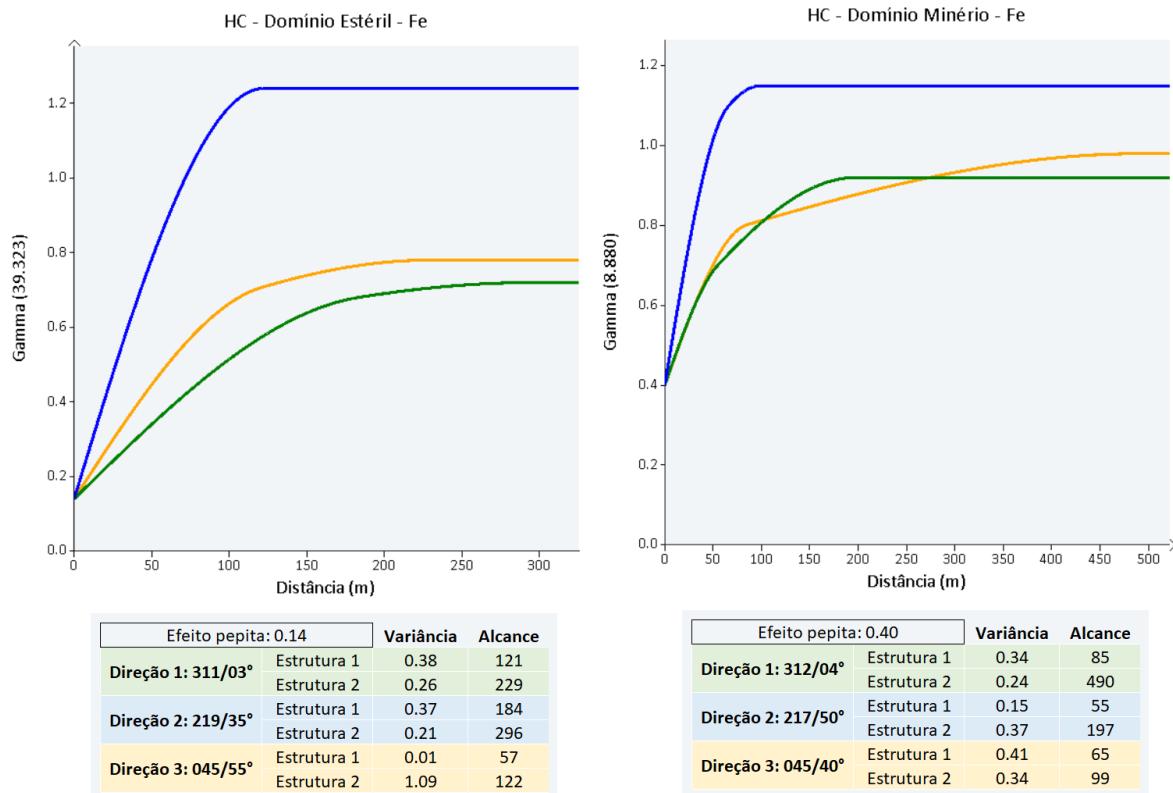
Apêndice VI-A) Modelos teóricos de variograma dos subdomínios estéril e minério no cenário de agrupamento por litotipos.



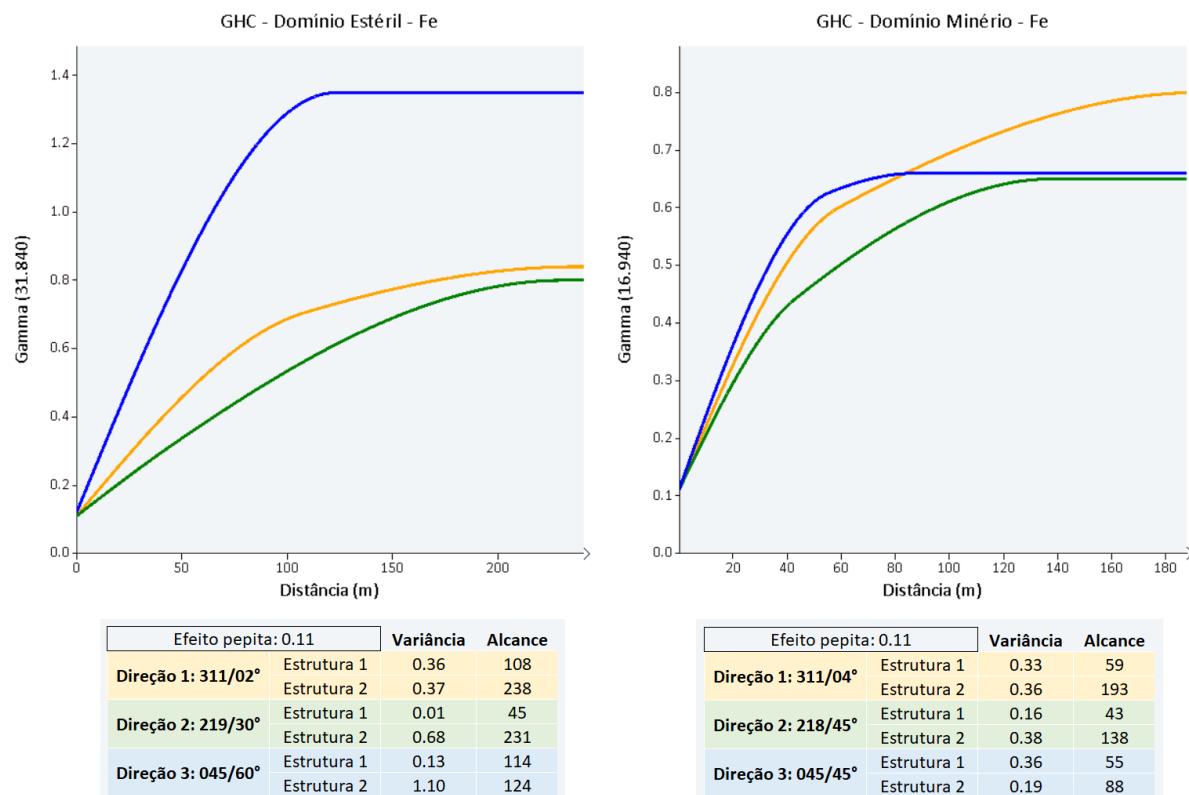
Apêndice VI-B) Modelos teóricos de variograma dos subdomínios estéril e minério no cenário de agrupamento do K-means.



Apêndice VI-C) Modelos teóricos de variograma dos subdomínios estéril e minério no cenário de agrupamento do Aglomerativo Hierárquico.

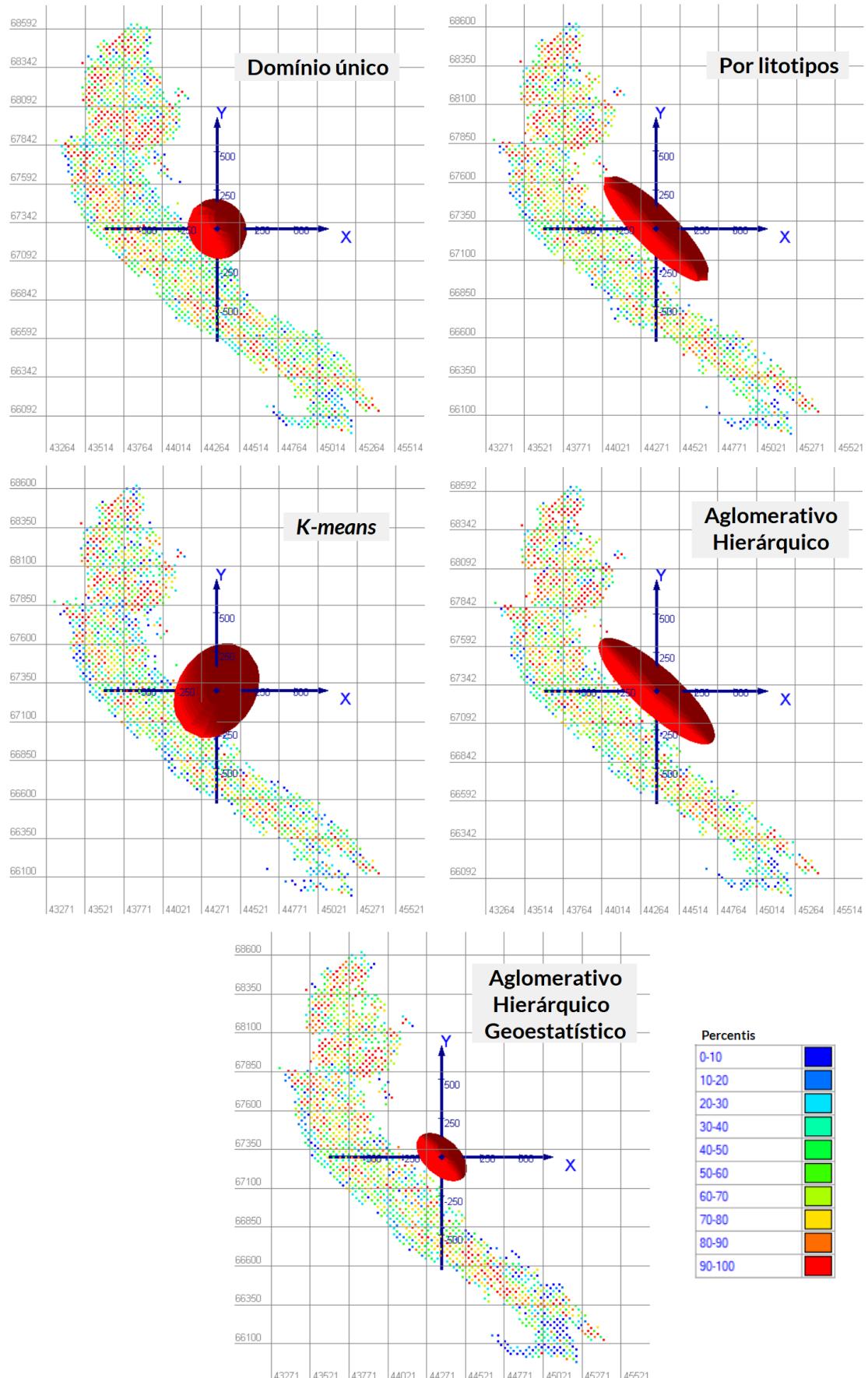


Apêndice VI-D) Modelos teóricos de variograma dos subdomínios estéril e minério no cenário de agrupamento do Aglomerativo Hierárquico Geoestatístico.



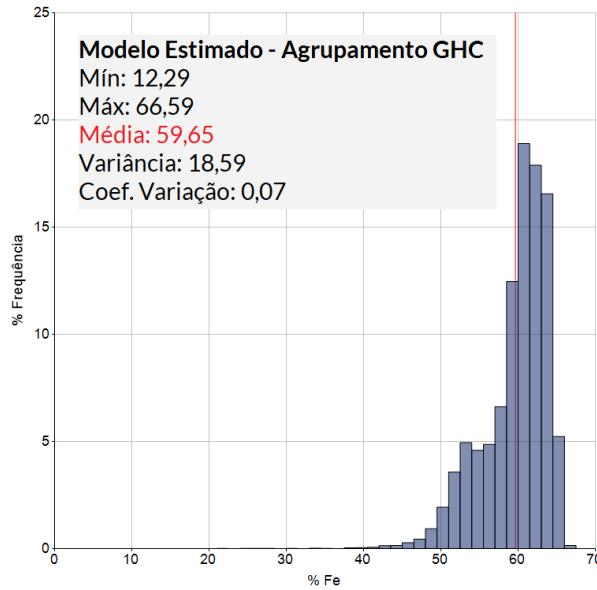
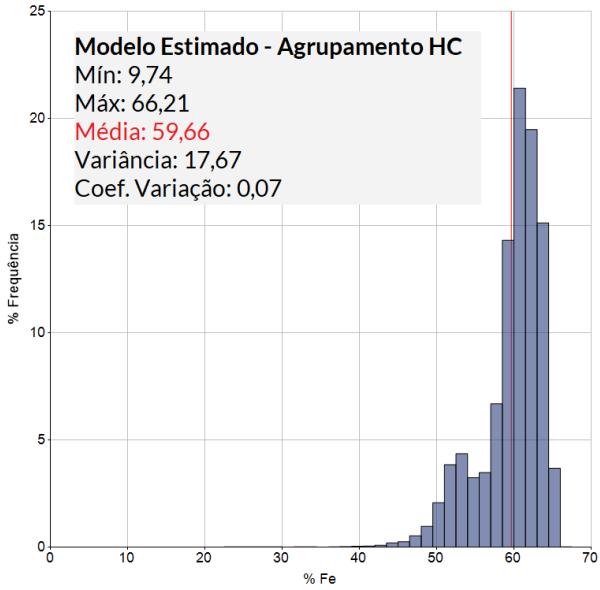
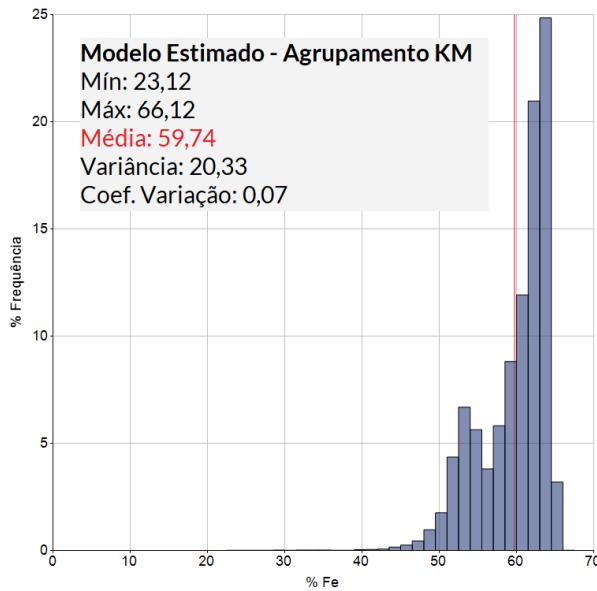
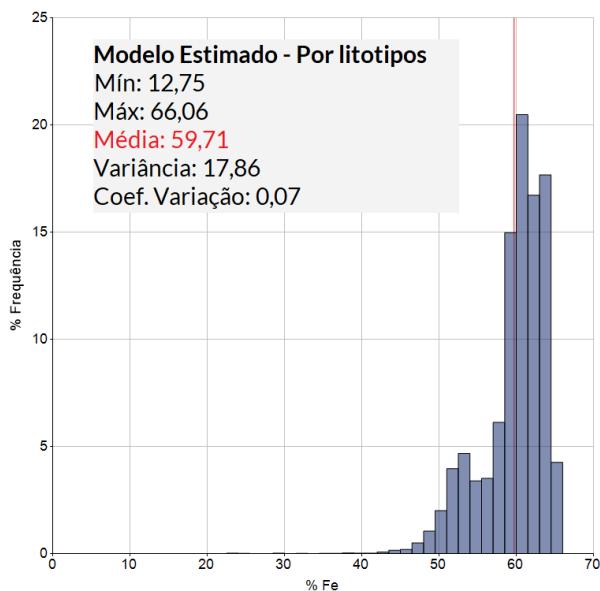
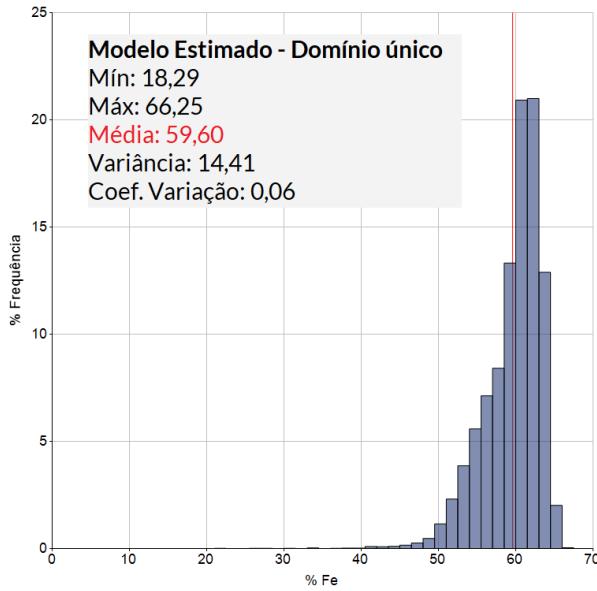
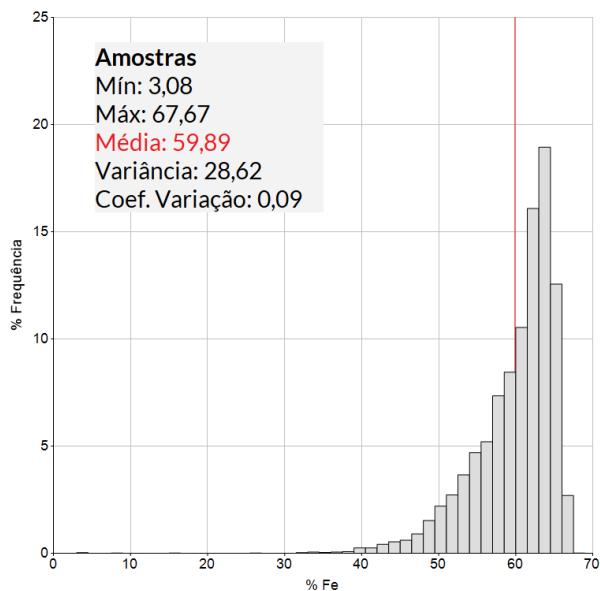
APÊNDICE VII – ELIPSOIDES DE CONTINUIDADE

Apêndice VII) Elipsoides de continuidade do ferro no domínio único e nos subdomínios de minério dos cenários de agrupamento.



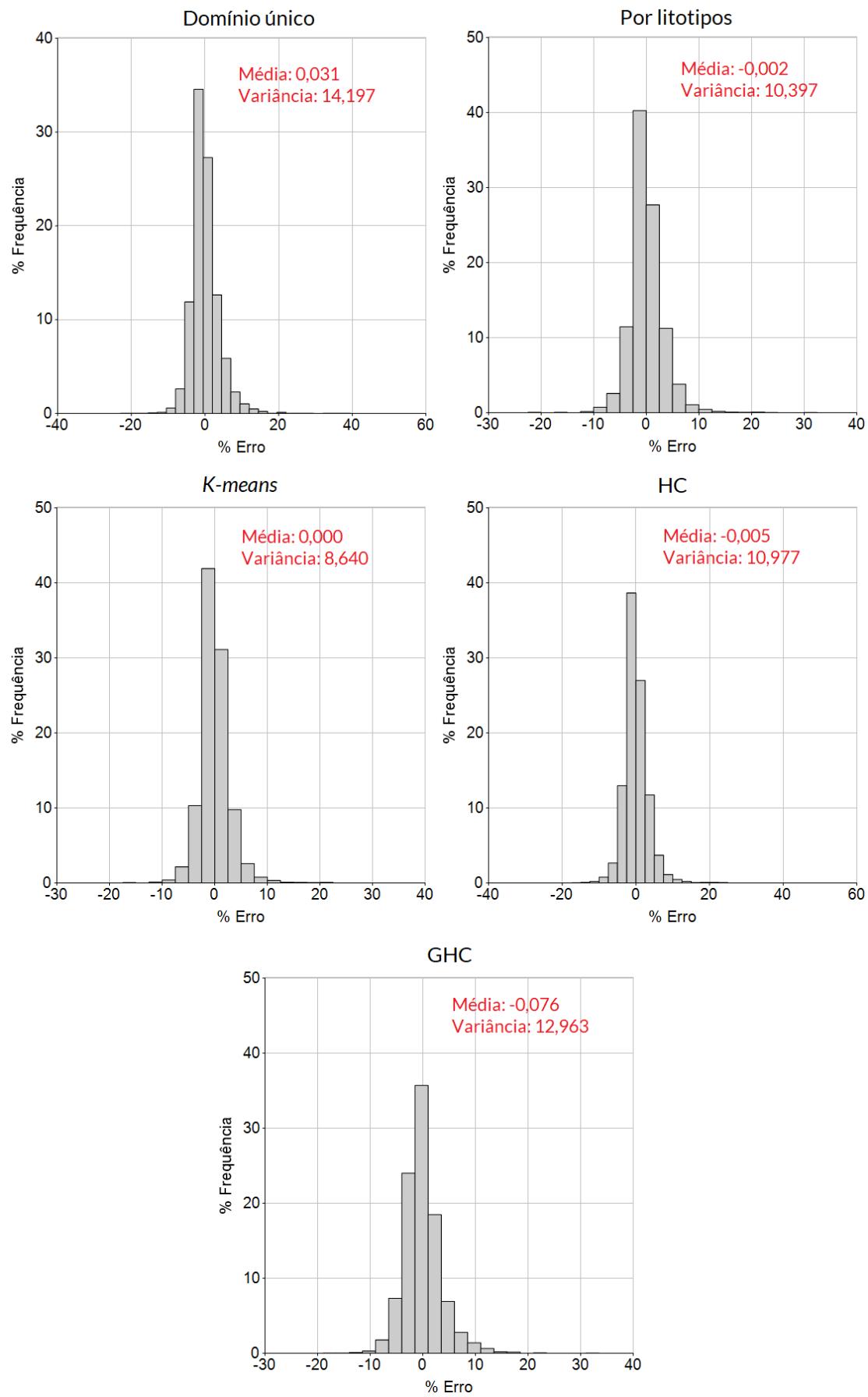
APÊNDICE VIII – HISTOGRAMAS DOS TEORES ESTIMADOS DE FERRO

Apêndice VIII) Histogramas dos teores de ferro das amostras e estimados em todos os cenários.

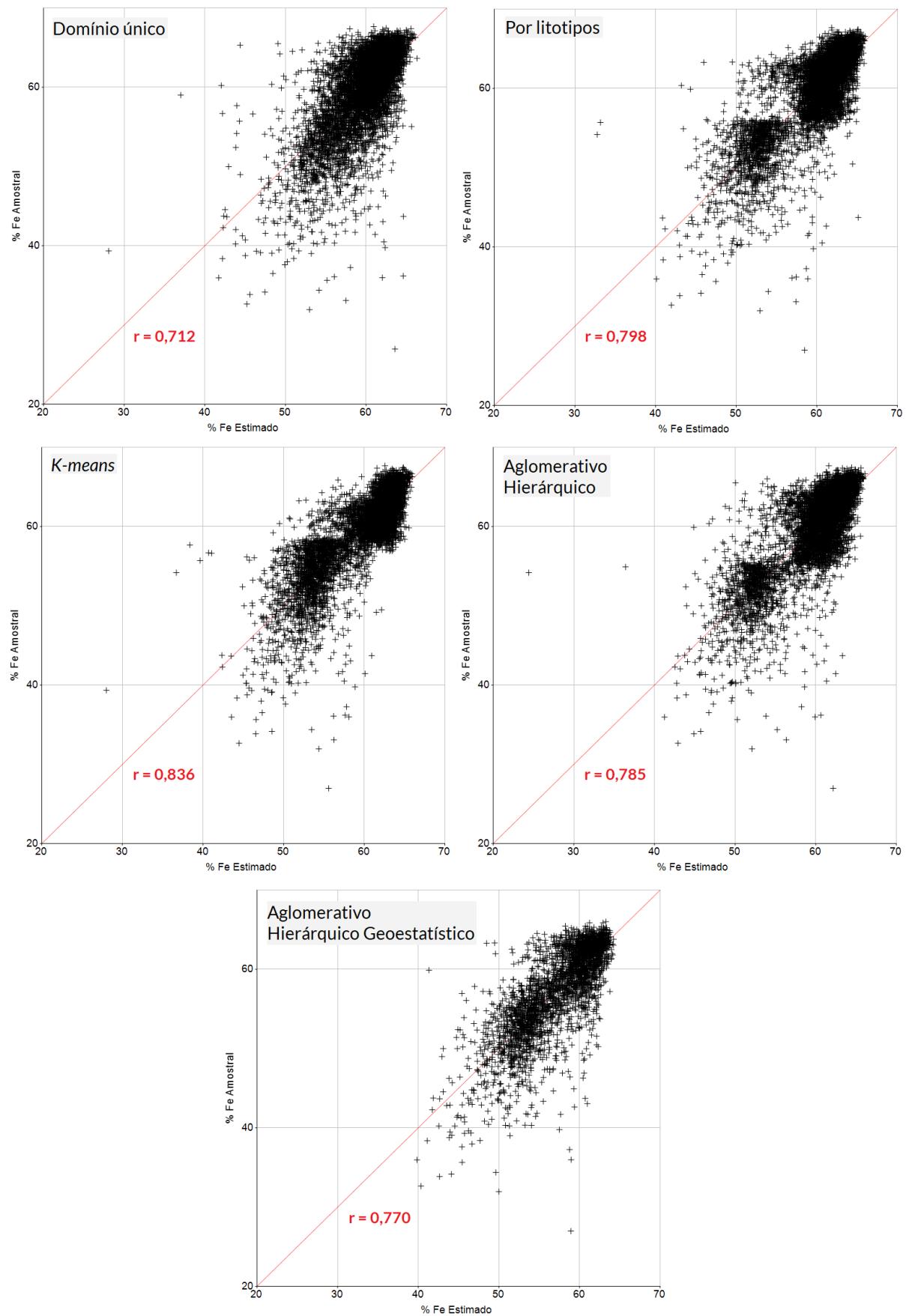


APÊNDICE IX – VALIDAÇÕES CRUZADAS

Apêndice IX-A) Histogramas das diferenças entre os teores reais e as estimativas pontuais (erro).



Apêndice IX-B) Gráfico de dispersão entre os teores reais e as estimativas pontuais.



APÊNDICE X – GRÁFICOS QUANTIL-QUANTIL

Apêndice X) Gráficos Quantil-Quantil entre o conjunto amostral e os teores de ferro estimados em cada cenário.

