

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE GEOCIÊNCIAS

Monografia do Trabalho de Formatura

**APLICAÇÃO DE ALGORITMOS DE AGRUPAMENTO NA DEFINIÇÃO DE
DOMÍNIOS PARA MODELAGEM E ESTIMATIVA DE TEORES NA MINA DE
FERRO DE CAPITÃO DO MATO NOVA LIMA-MG**

(TF –22/25)

Ivan Silva Carvalho

Orientador: Prof. Dr. Marcelo Monteiro da Rocha

Coorientadora: Geóloga Giulia Marina Cerqueira Dias

SÃO PAULO

2022

Agradecimentos

Dedico estes agradecimentos primeiramente a minha família por todo carinho, apoio e compreensão, meus pais Rosa e Geraldo meus maiores exemplos, ao meu irmão Ruan e sua esposa Dri que estiveram sempre ao meu lado, e as minhas sobrinhas Laura e Luísa que me lembram todos os dias do motivo de continuar.

A Vale S.A. que cedeu a base de estudos utilizada neste trabalho, a Geovariances pela liberação gratuita do programa Isatis® e Isatis.neo®, e a Datamine pela liberação do programa Studio RM®.

Agradeço também aos meus orientadores Marcelo Rocha e Giulia Cerqueira, que me guiaram neste trabalho de formatura, sempre dispostos a me ouvir e ajudar com meus problemas, despertando ainda mais meu gosto pela geoestatística, também a minha antiga orientadora Adriana Alves e todos os membros do projeto PMP, que me ensinaram como realizar um verdadeiro trabalho científico.

A todos os funcionários do IGc, que fazem o instituto funcionar, especialmente ao Antônio (LIG) e Camila (Litoteca), e a todos os professores do IGc, em especial ao Boggi, Claudio, Valdecir, Gallas, Lucelene e Fabio pela dedicação com as aulas e com os alunos.

Aos amigos da graduação, principalmente aos da Turma 60 que estive mais próximo nestes últimos anos, Marquinhos, Tampax, Michelle, Frida, Portera, Monange, Levi, Manu, Mijoleta, Jonas, Meska, Jason, Beisço, Gouvs, Marmita, Boi, Muri, Tey, Moranguete, Chocks etc...(me desculpem aos que posso ter esquecido), sem vocês para conversar, rir, abraçar, reclamar e fofocar, não seria possível chegar até aqui, a graduação valeu a pena pelo simples fato de ter conhecido vocês.

Deixo um agradecimento mais que especial, aos meus dois melhores amigos que conheci na USP, Bags e Pitu, estivemos juntos em nossos melhores e piores momentos desde que nos conhecemos, ao Rodrigo meu companheiro do E209 que trocamos histórias, conversas, conselhos e piadas sempre com um bom café.

A Bia, que acreditou em mim quando nem eu mesmo acreditava, por todo carinho, paciência, disposição, companheirismo, por você tenho total admiração e só posso agradecer, ao seu lado os dias ruins passam rápido e os bons parecem durar para sempre.

Por fim, agradeço às todas as instituições públicas que passei, pelo ensino de qualidade, estrutura e oportunidades que me trouxeram até aqui.

Sumário

RESUMO	5
ABSTRACT	6
1. INTRODUÇÃO	7
2. OBJETIVOS.....	7
3. FUNDAMENTAÇÃO TEÓRICA.....	8
3.1. Estatística	8
3.1.1. Estatística Descritiva.....	8
3.1.2. <i>Estatística Bivariada.....</i>	<i>8</i>
3.2. Geoestatística	8
3.3. Krigagem Ordinária.....	10
3.4. Domínios de estimativa	10
3.4.1. <i>K-Médias (K-Means)</i>	<i>11</i>
3.4.2. <i>Aglomerativo Hierárquico.....</i>	<i>13</i>
3.4.3. <i>Agrupamento Geoestatístico Hierárquico</i>	<i>14</i>
3.5. Depósitos de Ferro	15
3.6. Minérios de Ferro	16
3.7. Área de estudo	16
4. MATERIAIS E MÉTODOS.....	17
4.1. Materiais	17
4.1.1. <i>Banco de dados</i>	<i>17</i>
4.2. Métodos	18
4.2.1. <i>Revisão bibliográfica</i>	<i>18</i>
4.2.2. <i>Validação, análise exploratória e tratamento dos dados</i>	<i>19</i>
4.2.3. <i>Definições de subdomínios</i>	<i>19</i>
4.2.4. <i>Estimativa.....</i>	<i>19</i>
4.2.5. <i>Modelo de Blocos.....</i>	<i>20</i>
4.2.6. <i>Validação e comparação entre resultados.....</i>	<i>20</i>
5. RESULTADOS OBTIDOS	20

5.1. Estatística Geral	20
5.1.1 Teores	20
5.1.2 Litotipos	21
5.1.3 Estatística bivariada	22
5.2. Agrupamentos	25
5.2.1 Agrupamento Manual	26
5.2.2 K-médias	27
5.2.3 Aglomerativo Hierárquico	28
5.2.4 Aglomerativo Hierárquico Geoestatístico	28
5.3. Classificação e Distribuição dos Domínios	29
5.4. Geoestatística	33
5.5. Estimativa	33
5.5.1 Domínio Único	34
5.5.2 Domínios Pobre e Rico	35
5.6. Estatística e Validação das Estimativas	37
6. INTERPRETAÇÕES DOS RESULTADOS	39
6.1. Agrupamentos	39
6.2. Geoestatística	41
6.3. Validação e comparação dos resultados	41
7. CONCLUSÕES	42
8. REFERÊNCIAS BIBLIOGRÁFICAS	44
APÊNDICE-I-BOXPLOTS E HISTOGRAMAS POR VARIÁVEL	48
APÊNDICE-II- HISTOGRAMAS POR VARIÁVEL	49
APÊNDICE-III-ESTATÍSTICA DESCRITIVA POR LITOTIPO	50
APÊNDICE-IV-ESTATÍSTICA DESCRITIVA POR DOMÍNIO	53
APÊNDICE-V-AJUSTE DOS MODELOS TEÓRICOS DE VARIOGRAMA	57
APÊNDICE-VI-GRÁFICOS QUANTIL-QUANTIL	66

RESUMO

Esse trabalho foi realizado na mina de Capitão do Mato (CMT) localizada na Província Mineral do Quadrilátero Ferrífero (QF). Aplica estatística multivariada, por meio de técnicas de agrupamento, para a definição de domínios de estimativa, definidos manualmente ou por aprendizado de máquina, para os teores de Fe, buscando compreender essas aplicações quando comparada a estimativas sem o uso de agrupamentos. Para a realização do agrupamento, destacam-se os algoritmos K-médias (KM), o aglomerativo hierárquico (HC), métodos da estatística clássica elaborados para dados não regionalizados, e o agrupamento geoestatístico hierárquico (GHC) uma modificação do HC, que considera a localização espacial das amostras, buscando respeitar a continuidade espacial nos domínios de estimativa. No presente trabalho foram utilizadas 9502 amostras, validadas, submetidas a quatro agrupamentos distintos (Manual, KM, HC, GHC), que as classificou em minério rico minério pobre e rejeito. Foram gerados onze cenários distintos, incluindo o cenário não agrupado, a análise geoestatística nesses cenários evidenciou o controle estrutural-litológico na direção N110° e a influência das amostras de itabiritos na direção de *downhole*. Os cenários de agrupamento apresentam melhores resultados estatísticos na estimativa quando comparados a estimativa de domínio único, mas com menor continuidade espacial, destacam-se a estimativa do domínio pobre gerado pelo algoritmo KM, que apresenta melhor resultado de validação cruzada e maior estruturação do variograma, e as estimativas de domínio manual que apresentam estimativas menos suavizadas. A menor continuidade espacial dos agrupamentos não inviabiliza os resultados, mas a comparação dos resultados do agrupamento manual com os feitos por algoritmo não apresenta diferenças estatísticas ou de precisão de estimativa, relevantes, portanto, não justificando o tempo empenhado para geração de domínios pelos métodos de agrupamento, assim recomenda-se a estimativa de teores com domínios manualmente gerados para a mina de Capitão do Mato.

ABSTRACT

This work was carried out at the Capitão do Mato mine (CMT) located in the Mineral Province of the Ferriferic Quadrangle (QF). It applies multivariate statistics, by means of clustering techniques, to the definition of estimation domains, defined manually or by machine learning, for Fe grade, seeking to understand these applications when compared to estimates without the use of clusters. For the realization of the grouping, the K-media algorithms (KM), the hierarchical agglomerative (HC), classical statistics methods elaborated for non-regionalized data, and the hierarchical geostatistical clustering (GHC) a modification of the HC, which considers the spatial location of the samples, seeking to respect the spatial continuity in the estimation domains. In the present work, 9502 validated samples were used, submitted to four distinct clusters (Manual, KM, HC, GHC), which classified them as rich ore, poor ore, and tailing. Eleven distinct scenarios were generated, including the no grouped scenario, geostatistical analysis in these scenarios showed structural-lithological control in the N110° direction and the influence of itabirite samples in the downhole direction. The cluster scenarios present better statistical results in the estimation when comparing the single domain estimate, but with lower spatial continuity, we highlight the estimation of the poor domain generated by the KM algorithm, which presents better cross-validation result and greater variogram structuring, and manual domain estimates that have less smoothed estimates. The smaller spatial continuity of the clusters does not make the results unfeasible, but the results of the manual clustering with those made by algorithm do not present significant statistical or estimation precision differences, therefore, not justifying the time spent to generate domains by the clustering methods, so it is recommended to estimate grades with manually generated domains for the Capitão do Mato mine.

1. INTRODUÇÃO

Os recursos minerais devem ser quantificados ao longo de todas as etapas de um projeto de mineração, em toda a indústria mineral as técnicas de estimativa geoestatística são amplamente disseminadas, porque permitem a inferência e utilização da variabilidade natural dos depósitos, através da função variograma e permitem a estimativa de recursos e reservas minerais com maior precisão do que aquelas obtidas com a aplicação de interpoladores convencionais. (Borradaile, 2003; Yamamoto e Landim, 2013; Revuelta, 2017).

Nas últimas décadas o avanço tecnológico, computacional e analítico, potencializou a quantificação de diferentes dados geológicos, incluindo dados regionalizados, razão pela qual os tratamentos de dados multivariados regionalizados são cada vez mais difundidos nas geociências, em sua maioria originadas no estudo de estatística multivariada (Fouedjio, 2016). A integração de diversos tipos de dados permite a melhor compreensão dos fenômenos geológicos ligados a diversos fatores condicionantes (Landim, 2011).

Neste trabalho de Formatura (TF) aplicou-se as técnicas de aprendizado de máquina K-médias, aglomerativo hierárquico e Aglomerativo hierárquico geoestatístico, além do agrupamento manual para a definição de domínios e das estimativa de teores em cada cenário, a comparação das estatísticas descritivas entre o conjunto amostral e os modelos estimados, para uma base de dados de análises químicas de sondagens realizadas na mina de Capitão do Mato, localizada em Nova Lima (MG), explotada pela Vale S.A.

2. OBJETIVOS

O projeto teve como objetivo principal estimar os teores de ferro da mina de Capitão do Mato, localizada em Nova Lima-MG, utilizando de técnicas de estatística multivariada, para a definição de domínios e subdomínios de estimativa. A definição de domínios, foi feita manualmente por meio da litoquímica como condicionantes e por algoritmos de agrupamento não supervisionados, K-média, agrupamento hierárquico e agrupamento hierárquico geoestatístico. Os diversos modelos gerados pelas diferentes técnicas foram comparados entre si, em termos de precisão, continuidade espacial e estatística.

3. FUNDAMENTAÇÃO TEÓRICA

3.1. Estatística

3.1.1. Estatística Descritiva

Segundo Yamamoto, Bettencourt e Montanheiro (2001), Borradaile (2003), Revuelta (2017), a análise estatística das amostras é essencial, visando a compreensão do comportamento das variáveis de interesse dentro do corpo de minério. O cálculo das estatísticas descritivas deve ser realizado na etapa inicial do trabalho.

As principais estatísticas descritivas utilizadas são, segundo Yamamoto, Bettencourt e Montanheiro (2001), Borradaile (2003) e Landim (2011), as medidas de tendência central, com a média sendo a principal, as medidas de dispersão, representadas pela variância e seus derivados, desvio padrão e coeficiente de variação e por fim as medidas de forma, sendo principal a assimetria.

3.1.2. Estatística Bivariada

A análise bivariada tem como objetivo compreender a relação entre duas variáveis, descrevendo o grau de relação, seja por um único parâmetro, como o coeficiente de correlação de *Pearson* para relações lineares, ou por uma equação obtida pela análise de regressão Trauth (2007).

3.2. Geoestatística

A etapa de análise Geoestatística, lida com dados cuja localização é especialmente definida e que sejam auto correlacionados, denominadas variáveis regionalizadas $Z(X)$, cujo conceito foi apresentado por Matheron (1963) que as define como sendo realizações de funções aleatórias dependentes da localização, ou seja, uma função real com um valor definido em cada ponto do espaço.

As variáveis regionalizadas diferenciam-se das variáveis aleatórias dos métodos estatísticos tradicionais por reconhecer as características qualitativas: localização, suporte, continuidade e anisotropia, cujas descrições podem ser encontradas em Matheron (1963), Matheron (1971), Yamamoto (2001), entre outros.

A Teoria das Variáveis Regionalizadas tem como base a hipótese intrínseca, que garante que uma função, a função intrínseca, descreve o comportamento espacial da variável

regionalizada dentro do espaço e que essa função é uma característica intrínseca da regionalização (IPT, 1989).

A função variograma ($2\gamma(h)$), fundamental na Geoestatística, por descrever o comportamento espacial da variável regionalizada, é calculada como (Matheron, 1963; Yamamoto, 2001; Deutsch, 2003; Rossi e Deutsch, 2014; entre outros):

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

Onde $2\gamma(h)$ é a função variograma; n é o número de pares de pontos separados por uma distância h ; $Z(x_i)$ é o valor da variável regionalizada no ponto (x_i) ; $Z(x_i + h)$ é o valor da variável regionalizada no ponto $(x_i + h)$

A função variograma é graficamente representada pelo variograma, que quando estruturado apresenta, amplitude, patamar, efeito pepita e variância espacial (Armstrong, 1998; Yamamoto, 2001) como pode ser observado na Figura 1. Os campos estruturado e aleatório representam, respectivamente, a região do variograma onde existe dependência espacial entre pares de pontos e a região no variograma onde os pares de pontos são especialmente independentes.

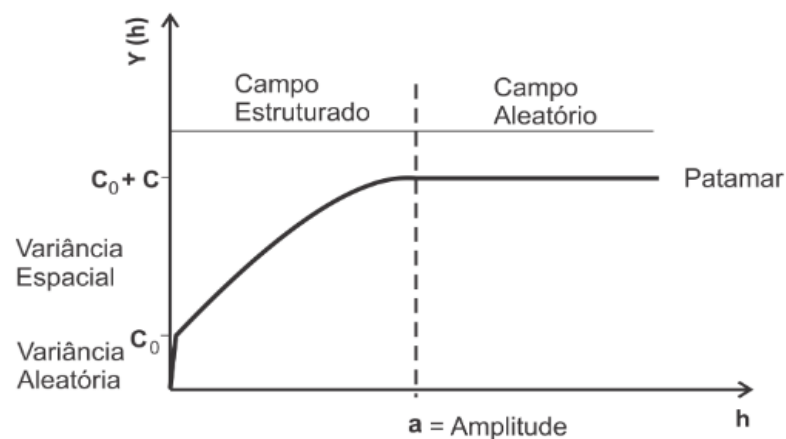


Figura 1 Variograma típico e suas propriedades (extraído de Yamamoto, 2001).

O significado das propriedades do Variograma são:

Amplitude: distância a partir da qual as amostras passam a ser espacialmente independentes, reflete o grau de continuidade das amostras.

Patamar: valor de variância no qual o variograma estabiliza-se, representado por $C+C_0$ sendo C a variância espacial e C_0 o valor do efeito pepita, reflete o grau de heterogeneidade das amostras.

Efeito pepita: o valor da função variograma próximo à origem.

Variância espacial: Diferença entre o patamar e o efeito pepita, indica como uma amostra se diferencia da outra em função da distância.

3.3. Krigagem Ordinária

Segundo Yamamoto e Landim (2013) krigagem é o nome genérico para um conjunto de técnicas que busca determinar o melhor estimador linear não enviesado, com minimização da variância do erro (Revuelta, 2017). A aplicação da krigagem é realizada após a análise geoestatística, pois há a necessidade de conhecer o modelo teórico de variograma, que é uma função matemática pré-definida que foi ajustada ao variograma experimental (Yamamoto, 2001).

A aplicação da krigagem ordinária é realizada em dados que apresentam distribuição normal ou apresentam distribuição assimétrica negativa, sem que seja necessário qualquer tipo de transformação, ou seja aplica-se diretamente sobre os dados originais (Yamamoto e Landim, 2013). A krigagem ordinária não necessita que a média populacional seja conhecida, pois usa apenas médias locais, porém para garantir a minimização da variância a soma dos pesos deve ser estrita a 1, esta restrição é chamada de condição de restrição da krigagem ordinária (Yamamoto, 2001; Yamamoto e Landim, 2013).

O estimador da krigagem ordinária, é apresentado por diversos autores entre eles Yamamoto (2001), Deutsch (2003), Yamamoto e Landim (2013), Rossi e Deutsch (2014), sendo o trabalho de Matheron (1963), seminal, é calculado como:

$$z_{ko}^*(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \text{ com } \sum_{i=1}^n \lambda_i = 1$$

Onde $z_{ko}^*(x_0)$ é o valor da variável de interesse em um ponto não amostrado, calculado a partir da combinação linear dos pontos de dados vizinhos $z(x_i)$, sendo λ_i os pesos que cada amostra receberá.

3.4. Domínios de estimativa

A definição de domínios de estimativa é comumente utilizada para contornar a baixa probabilidade de um depósito natural poder ser considerado um domínio estacionário, desta forma o depósito deve ser dividido em domínios que satisfaçam a hipótese intrínseca para determinada(s) variável(is) (Abzalov, 2016), a partir dos domínios de estimativa os volumes

mineralizados são determinados. Por consequência, são um importante fator na tonelagem estimada acima do teor de corte (Rossi e Deutsch, 2014).

Há uma variedade de métodos, para estabelecer a definição de domínios de estimativa, conforme apresentado em Wackernagel (2003), Landim (2011) e Fouedjio (2020). Essa definição pode ser feita manualmente, ou por meio de aprendizado de máquina (*machine learning*), onde se destaca a análise de agrupamentos (*cluster analysis*), um método não supervisionado, sem resultados pré-definidos que sirvam de exemplo para o modelo definir os domínios (Romary *et al.*, 2012). Para a realização do agrupamento (*clustering*), destacam-se os algoritmos K-médias, em que é obrigatório definir o número de domínios a serem gerados no início do processo de agrupamento, o aglomerativo hierárquico em que a definição do número de domínios a serem gerados não é necessário (Romary *et al.*, 2012; Fouedjio, 2016; Faraj e Ortiz, 2021) e o agrupamento geoestatístico hierárquico, proposto por Romary *et al.* (2015), como uma modificação do aglomerativo hierárquico, em que a união de domínios pelo algoritmo é condicionada a estruturação dos dados em um gráfico de acordo com sua localização, buscando respeitar a continuidade espacial nos domínios de estimativa.

3.4.1. K-médias (K-Means)

O Algoritmo K-médias, apresentado por Macqueen (1967), também chamado de algoritmo de Lloyd, devido ao clássico trabalho na área de Engenharia elétrica de Lloyd (1982), é o principal algoritmo de agrupamento, para as mais diversas áreas, devido a sua fácil aplicação (McCool, Robinson e Reinders, 2012).

O agrupamento tem início com a distribuição das amostras de forma aleatória em K-agrupamentos, sendo *K* um número de grupos pré-estabelecidos pelo usuário, que são atualizados, de acordo com (Macqueen, 1967; McCool, Robinson e Reinders, 2012; Wilcox, 2012):

- I. Cálculo do centroide de cada agrupamento, definido como o valor *p*-dimensional de médias dos elementos dos K-agrupamentos.
- II. Atribuir cada ponto de observação ao agrupamento de centroide mais próximo, isto é, cuja distância euclidiana seja menor.

Estes passos são iterados até que os agrupamentos não se alterem mais, como o resultado é condicionado a distribuição aleatória da primeira iteração cada execução do algoritmo terá um resultado único, sendo recomendado executá-lo até ser observado um padrão de agrupamento comum (Romary *et al.*, 2012; Wilcox, 2012; James *et al.*, 2013). A Figura 2 apresenta uma síntese do algoritmo.

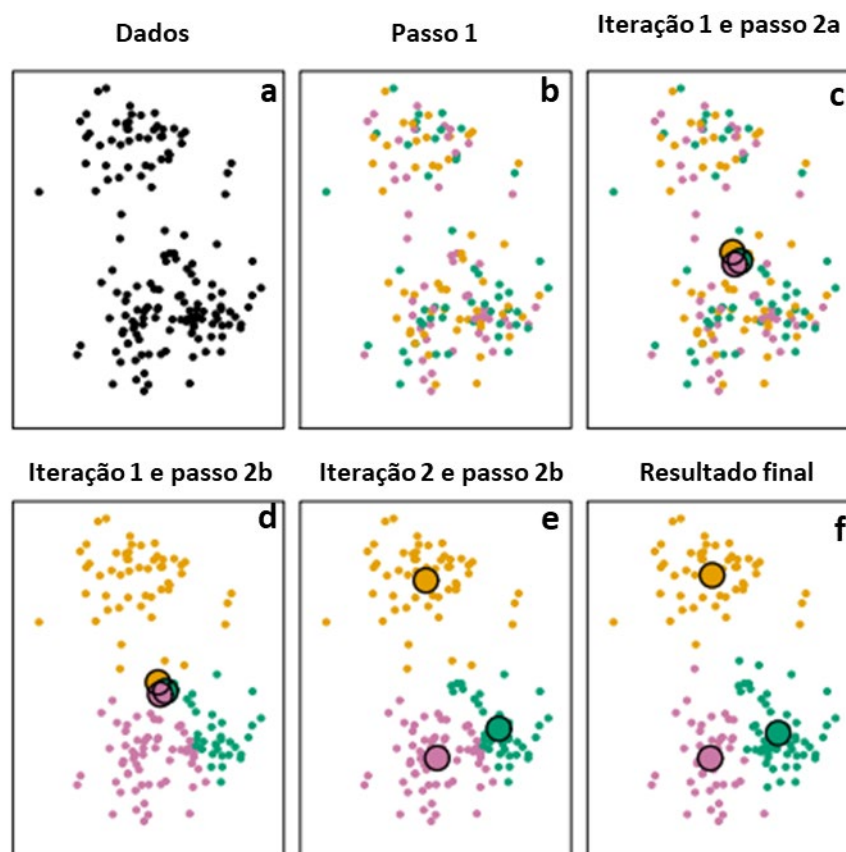


Figura 2 Apresentação de um agrupamento por meio da iteração do algoritmo K-médias, com $k=3$, onde cada cor representa um grupo, com suas respectivas amostras e centroides, (a) Apresentação do espaço amostral, (b) Passo 1, Distribuição aleatória das amostras em 3 grupos, (c) Passo 2a, cálculo dos centroides dos agrupamentos e primeira iteração, (d) Passo 2b, redistribuição das amostras de acordo com a posição dos centroides, (e) iteração 2, os centroides são reposicionados de acordo com as amostras a eles atribuídas na primeira iteração, (f) o processo se repete n vezes até não haver mais mudanças nos centroides. (James *et al.*, 2013).

A escolha do número de agrupamentos comumente é feita pelo método do cotovelo (*elbow method*), chamado assim pelo formato característico do gráfico (Fig. 3), cuja ideia básica é usar o quadrado da distância entre os pontos de amostra em cada *cluster* e seu centroide para fornecer uma série de números de grupos para a realização do agrupamento (Kassambara, 2017; Yuan e Yang, 2019).

Realizando a soma dos erros ao quadrado (SEQ ou inércia) obtém-se um indicador de desempenho, quando o número de grupos se aproxima do número de agrupamentos ideal, o SEQ mostra um rápido declínio, já quando o número de grupos excede o número ideal, a SEQ diminui de forma suave (Yuan e Yang, 2019), portando o ponto de inflexão (*elbow point*), será o número recomendado de grupos visto que, evita tanto o agrupamento de amostras muito distintas em poucos grupos, quanto o agrupamento de poucas amostras em muitos grupos (Kassambara, 2017; Yuan e Yang, 2019).

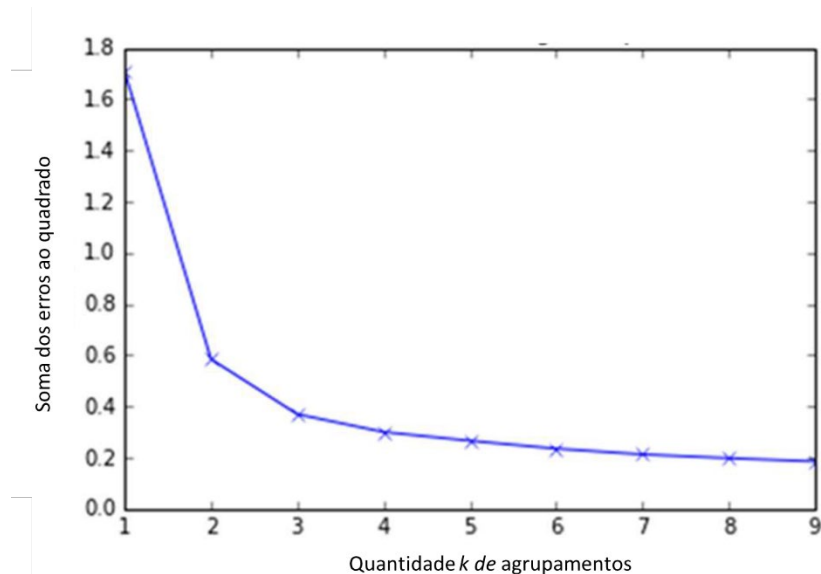


Figura 3 A quantidade de agrupamentos pode ser melhor determinada traçando a curva K-SEQ e encontrando a inflexão, neste exemplo o ponto de inflexão ocorre quando $k = 2$, portanto o efeito de agrupamento do conjunto de dados é o mais próximo ao real quando o número de grupos escolhidos é 2. (Yuan e Yang, 2019).

3.4.2. Aglomerativo Hierárquico

O agrupamento hierárquico, é um conjunto de algoritmos que contornam as principais dificuldades do K-médias, visto que não é necessário definir o número de grupos para iniciar o algoritmo e seus resultados são sempre iguais para um mesmo conjunto de dados (James *et al.*, 2013; Kassambara, 2017).

O agrupamento hierárquico, pode ser subdividido nos tipos aglomerativo e divisivo, que se utilizam de dendrogramas (Fig.4a e 4b), ou seja, estruturam os dados em uma hierarquia multinível, sendo o primeiro, o mais utilizado, nele cada dado é tratado como um grupo onde cada grupo é sucessivamente unido a outro, com base em semelhanças, até que haja um único grupo (Fig. 4a), já o segundo tipo faz o caminho inverso, os dados são agrupados em um único grupo e são divididos sucessivamente, é pouco utilizado por ser muito mais complexo do que o método aglomerativo (Hexmoor, 2015)

A semelhança entre grupos pode ser definida de diversas maneiras (Teichgraeber e Brandt, 2018), tradicionalmente, para conjuntos amostrais grandes ($n > 100$), utiliza-se o método de Ward, onde cada grupo é interpretado como um centroide e os pares são agrupados por meio da menor distância euclidiana entre eles (Fig. 4a e 4b). (Hexmoor, 2015; Teichgraeber e Brandt, 2018)

A obtenção de k grupos é definida a partir da escolha do nível hierárquico correspondente, o nível hierárquico dos dendrogramas é relativo à medida de proximidade entre os grupos ou à distância euclidiana entre os pares no espaço multivariado (James *et al.*, 2013; Kassambara, 2017).

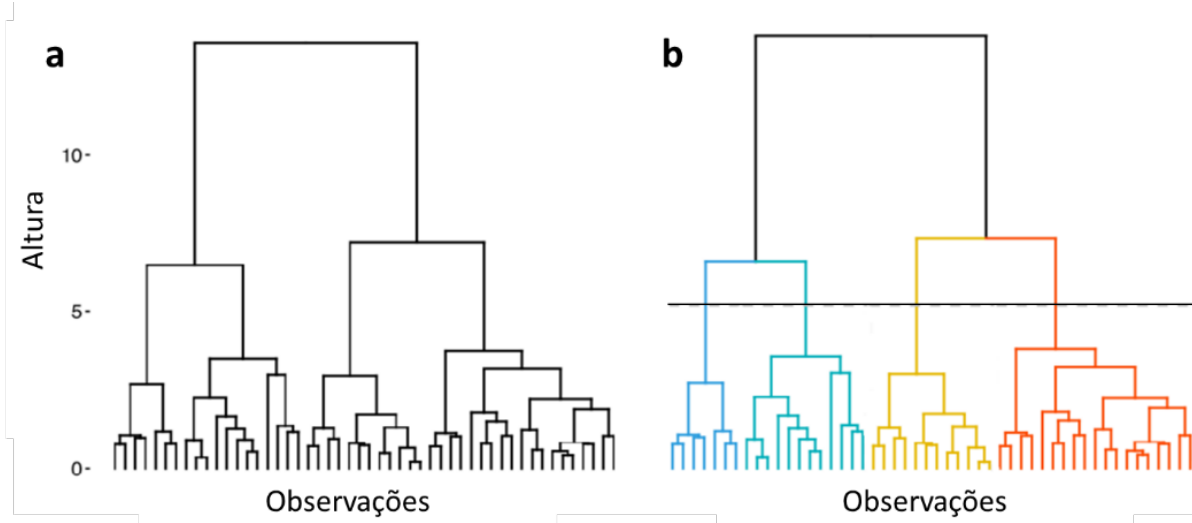


Figura 4: **a)** Dados em uma hierarquia multinível (Dendrograma), **b)** A obtenção de 4 grupos a partir da escolha do nível hierárquico correspondente (distância euclidiana entre os pares no espaço multivariado) (Yuan e Yang, 2019).

3.4.3. Agrupamento Geoestatístico Hierárquico

Ambos os algoritmos anteriormente apresentados, podem levar em consideração as coordenadas das amostras como parte das variáveis, mas não podem ser considerados métodos espaciais, visto que tratam as coordenadas e os atributos da mesma forma, gerando eventualmente, resultados que não refletem a continuidade espacial das variáveis ou a caracterização estatística dos domínios, impactando na coerência desses domínios com o fenômeno real (Fouedjio, 2016; Faraj e Ortiz, 2021), alguns autores propõem métodos espaciais, desenvolvidos a partir dos métodos não espaciais anteriormente citados, tais como o agrupamento geoestatístico hierárquico apresentado por Romary *et al.* (2015), uma modificação do aglomerativo hierárquico, em que a união dos grupos pelo algoritmo apresenta uma restrição, condicionada a distribuição espacial dos dados, dois grupos só podem ser agrupados se apresentarem conexão, definida por meio da triangulação de Delaunay ou por uma matriz de conectividade, calculada a partir da delimitação gráfica C , duas amostras δ_i e δ_j , $i \neq j$ - são conectadas se houver uma aresta em C ligando δ_i e δ_j . Isto é representado por $\delta_i \leftrightarrow \delta_j$, a matriz é expressa da seguinte forma:

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix},$$

Se $C_{ij} = 1$, as amostras i e j são conectadas, caso $C_{ij} = 0$, não (Bubnova *et al.*, 2020). Dois *clusters* são conectados (Fig. 4) se existir conexão de eventuais pares de pontos no entre os dois grupos Romary *et al.* (2015).

3.5. Depósitos de Ferro

Sistemas de mineralização de ferro hospedado em formações ferríferas bandadas (*Banded Iron Formation* - BIF), representam os maiores depósitos de minério de ferro do mundo (Fig. 5), ocorrendo em todo Pré-cambriano (Hagemann *et al.*, 2016).

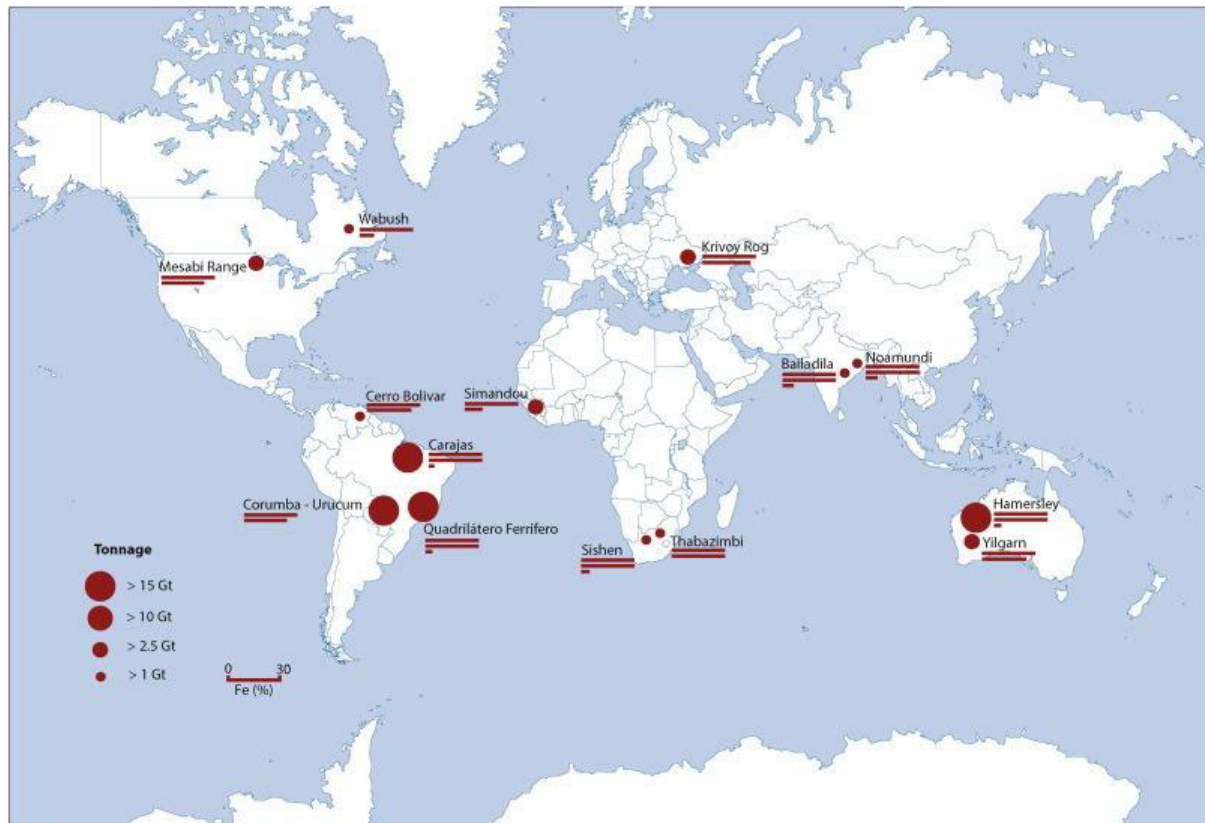


Figura 5: Localização de distritos e depósitos significativos de minério de ferro hospedados em BIF. O comprimento cumulativo dos gráficos de barras abaixo de cada localidade indica o teor de Fe (extraído de Hagemann *et al.*, 2016).

Os depósitos BIF apresentam classificações quanto a seus ambientes deposicionais (Hagemann *et al.*, 2016), sendo classificados em tipo Algoma (e. g. o distrito de minério de ferro Serra Norte na Província Mineral de Carajás), BIF tipo Lago superior Proterozoico (e.g. QF) e BIF do tipo Rapitan-Urucum neoproterozoico (e. g. Urucum), a configuração geotectônica destes ambientes é esquematizada na Figura 6.

Nos três ambientes deposicionais, o fluido hidrotermal tem fluxo controlado por cinturões de *greenstones belts* ou estruturas em escala de bacia, como *strike-slip*, falhas normais ou *grabens*, os fluidos hidrotermais ascendem ao longo destas zonas de alta

permeabilidade, durante a orogênese, causando alteração hipogênica e mineralização (Hagemann *et al.*, 2016).

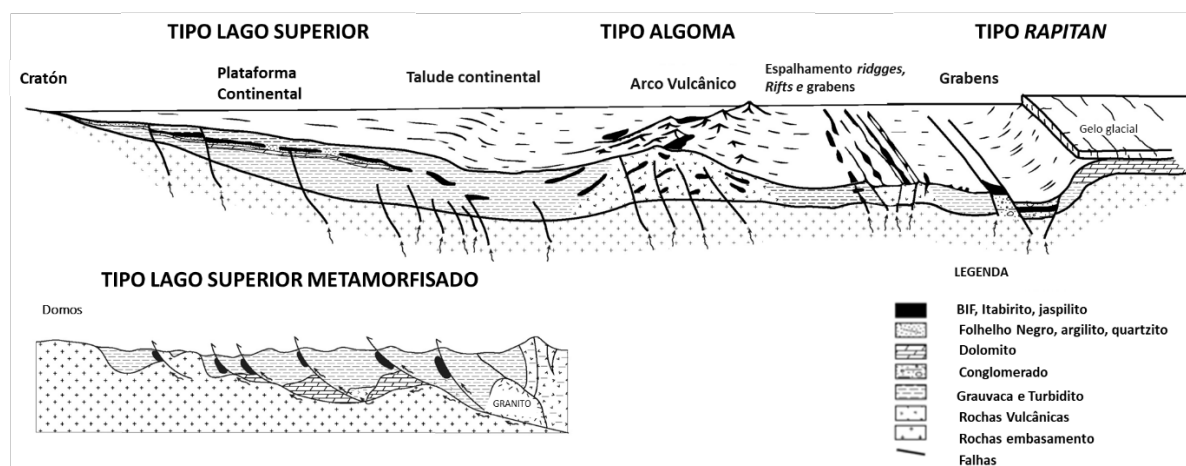


Figura 6: Configuração geotectônica de BIFs do tipo Algoma, Lago Superior, *Rapitan* e configurações do tipo Itabirito do Lago Superior metamorfisados. Esses BIFs e itabiritos são considerados o protólito das rochas mineralizadas em ferro de baixo e alto teor. (extraído de Hagemann *et al.*, 2016).

Hagemann *et al.* (2016), descrevem resumidamente que as formações do tipo Lago Superior proterozoicas, como as do QF, em depósitos extensos, formados em sucessões sedimentares também proterozoicas de margem passiva, sem associação evidente com vulcanismo; as fácies ferruginosas são essencialmente oxidadas e os itabiritos tiveram maior desenvolvimento no paleoproterozoico.

3.6. Minérios de Ferro

Os minérios de ferro são comumente descritos em termos de diminuição da dureza física relativa como: duro a médio, friável e, finalmente, poeira ou pó. Os minérios de ferro no Quadrilátero Ferrífero muitas vezes variam de friável a em pó, com granulação areia e os grãos de hematita são tipicamente muito densas com pouco ou nenhum micro poro (Holmes, Lu e Lu, 2022).

A dureza e características relacionadas ao minério de ferro são fatores importantes em sua moagem e beneficiamento bem como na quantidade de minério granulado (*lump*), obtido (Holmes, Lu e Lu, 2022). A definição dos tipos de minérios é dada pelo tamanho da partícula, sendo do tipo *lump* quando as partículas têm entre 31,5 mm a 6,3 mm, do tipo *sinter feed* quando os grãos tem diâmetro entre 6,3 mm e 0,15 mm e, por fim, *pellet feed* (pelotas), com partículas menores do que 0,15 mm (Clout e Manuel, 2022).

3.7. Área de estudo

O Quadrilátero Ferrífero (QF), hospeda depósitos de minério de ferro de classe mundial (Holmes, Lu e Lu, 2022), localizado no sul do Cráton São Francisco em rochas de idades arqueanas à neoproterozoicas, cercado por cinturões orogenéticos vergentes para seu interior (Rosière e Chemale Jr., 2000). O QF está localizado na zona externa de dois orógenos superpostos – a zona de *foreland* de um orógeno Riachão-Orosiriano e o substrato da margem leste, retrabalhada, do cráton São Francisco (Alkmim e Texeira, 2017).

Localizada no QF, a mina de Capitão do Mato (Fig. 7), encontra-se no lado contrário do flanco da sinclinal Moeda, no Supergrupo Minas, em rochas das Formações Moeda e Batatal e em rochas dos Grupos Caraça e Itabira (Dorr, 1969), sendo este último, subdividido nas formações Cauê e Gandarela, que se intercalam horizontal e verticalmente (Rosière e Rolim, 2016), a mina desenvolveu-se predominantemente na direção noroeste-sudeste, dentro da Formação Cauê predominantemente, com dobras sinclinais isoformais, de direção NW/SE, controlando a morfologia do minério, até seus limites leste, que ocorrem frequentemente como lentes, estiradas ou não, em charneiras de dobras e em zonas de cisalhamento (Suckau *et al.*, 2004)

Em 31 de dezembro de 2021, as reservas medidas totais de minério de ferro do complexo minerário Vargem Grande (no qual a mina Capitão do Mato está inserida) correspondiam a 1.346,6 milhões de toneladas métricas a um teor médio provado de 46,2%, incluindo umidade e teor de ferro em base seca (Vale, 2021), o minério bruto (*run-of-mine*) é processado por meio de etapas padrão de britagem, classificação e concentração, produzindo *sinter feed*, granulado e *pellet feed* nas plantas de beneficiamento localizadas nos complexos de mineração, com a produção de 31,3 milhões de toneladas métricas de minério de ferro com recuperação de beneficiamento de 85,6% (Vale, 2022).

4. MATERIAIS E MÉTODOS

4.1. Materiais

4.1.1. Banco de dados

O banco de dados utilizado neste trabalho é composto por informações de 856 sondagens realizadas na mina de Capitão do Mato, composto por 14515 amostras, incluindo rochas mineralizadas e rochas estéreis. A base de dados é composta pelas informações da boca dos furos de sonda (*header*), de azimutes e mergulhos (*survey*) e *assays* compostas pelo comprimento das amostras ao longo dos testemunhos e da análise química para Fe, Si, P, Al e P (globais e por faixas granulométricas), e classificação dos litotipos de cada amostra.

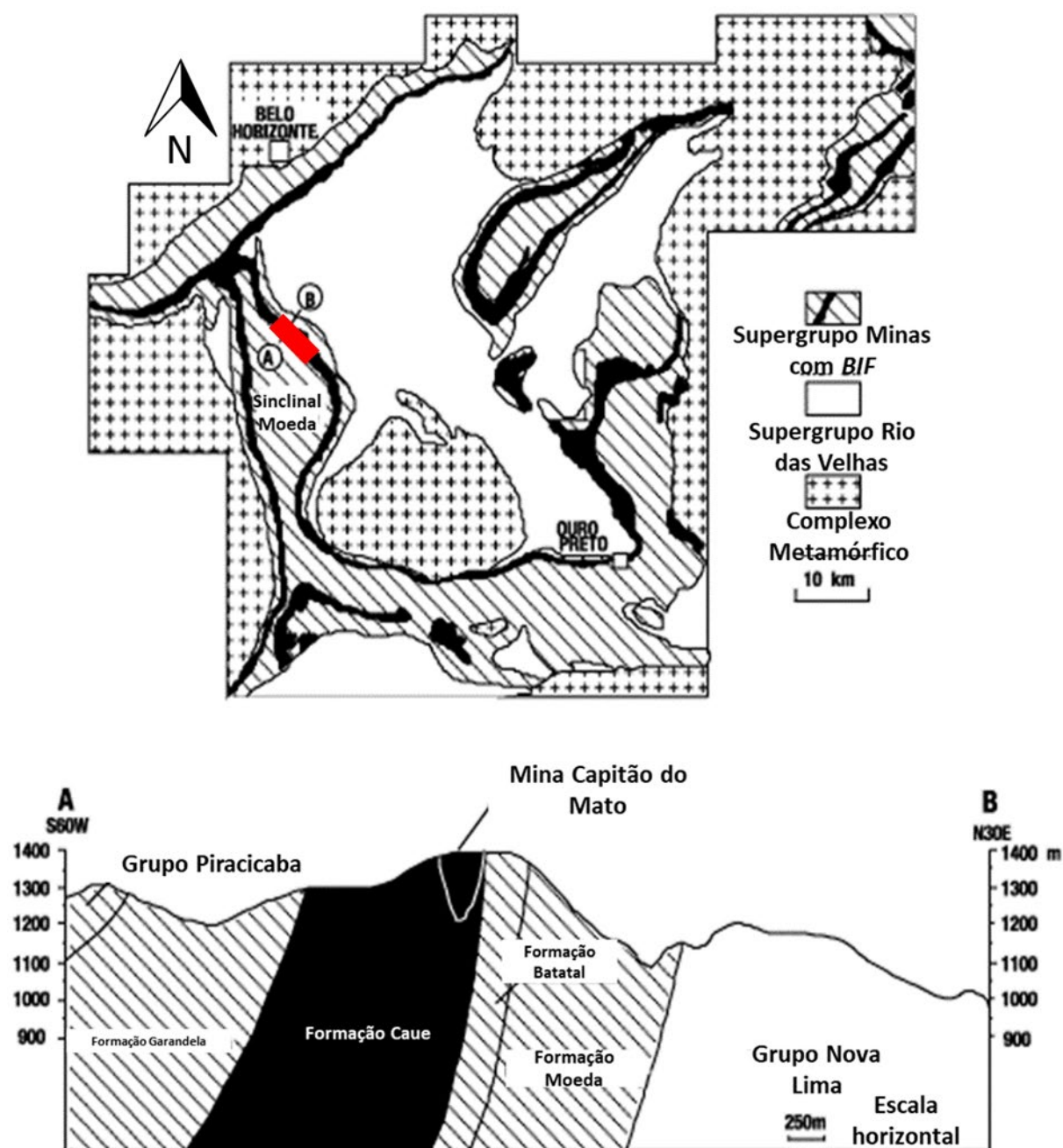


Figura 7: – Mapa geológico simplificado e seção geológica no perfil A-B da mina de Capitão do Mato, destacada em vermelho no mapa, Quadrilátero ferrífero, Minas Gerais, Brasil (Extraído de Varajão *et al.*, 2002).

4.2. Métodos

4.2.1. Revisão bibliográfica

A revisão bibliográfica foi realizada durante todo o desenvolvimento do trabalho, para tal, a Biblioteca do IGc-USP foi utilizada além bases de dados de artigos científicos disponíveis na internet às quais a USP tem acesso, permitindo uma melhor compreensão da área de estudo, técnicas empregadas e resultados obtidos.

4.2.2. Validação, análise exploratória e tratamento dos dados

O projeto teve início com a validação do banco de dados utilizado, buscando averiguar e porventura corrigir eventuais inconsistências existentes e identificar as configurações da malha de sondagem tais como, distância entre sondagens, comprimento total e orientação espacial da sondagem e o tamanho das amostras. Nesta etapa foi realizada a análise estatística preliminar das variáveis, com auxílio da biblioteca *seaborn* (Waskom, 2021) da linguagem *python 3* (Rossum e Drake, 2009).

As amostras foram regularizadas, por meio do cálculo do teor composto, para que o suporte amostral seja o mesmo em todas as amostras e adequar as amostras à escala de trabalho.

A validação e preparação do banco de dados original para a estimativa foi feita nas seguintes etapas, seguindo os procedimentos comumente adotados nos relatórios de estimativa da Vale S.A.:

- Exclusão das amostras sem dados de química.
- Verificação de *overlap* de intervalo.
- Exclusão das amostras sem dados de granulometria.
- Exclusão das amostras com somatória química ou granulométrica inadequados (<95% ou >105%).
- Regularização das amostras em 10 metros.

Após estes procedimentos, restaram 9502 amostras.

4.2.3. Definições de subdomínios

As definições de domínios foram realizadas a partir das amostras dos testemunhos de sondagem. As análises de agrupamento para a definição não supervisionada dos subdomínios foram feitas utilizando os algoritmos de K-médias, aglomerativo hierárquico, por meio da biblioteca- *scikit-learn* (Pedregosa *et al.*, 2011) e o aglomerativo hierárquico geoestatístico, disponível no *software isatis.neo*.

4.2.4. Estimativa

A técnica de estimativa utilizada será a krigagem ordinária (KO), por meio do *software isatis.neo*, seguindo o fluxo de trabalho apresentado na Figura 8.

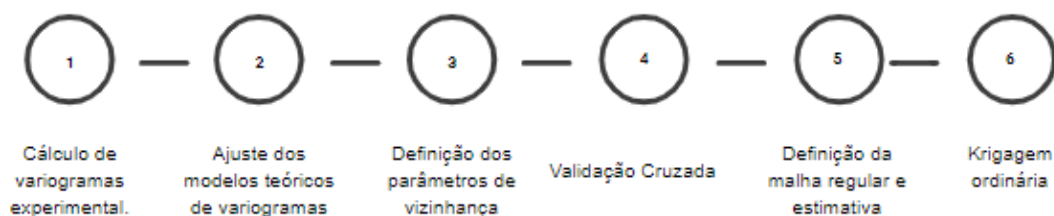


Figura 8: Fluxo geral de trabalho para a realização de estimativa por krigagem (fonte: o autor).

4.2.5 Modelo de Blocos

O ajuste do modelo de blocos, foi rotacionado para N110 visando adequá-lo à linha base da amostragem. Como a malha de amostragem é irregular, as dimensões dos blocos de estimativa foram definidas como a metade da mediana da distância entre amostras por domínio, cada domínio estabelecido conta com uma distribuição amostral única, optou-se por utilizar a fronteira convexa evitar extrapolações da região a ser estimada.

4.2.6. Validação e comparação entre resultados

Os métodos de validação e comparação globais e locais para as estimativas foram:

- I. Validação Cruzada
- II. Comparação das estatísticas descritivas do conjunto amostral e dos modelos estimados
- III. Comparação dos domínios definidos com a distribuição dos litotipos de minério
- IV. Comparação dos resultados de estimativa obtidos para os diferentes agrupamentos com a estimativa feita para o domínio não agrupado.

5. RESULTADOS OBTIDOS

5.1. Estatística Geral

5.1.1 Teores

As distribuições dos teores das amostras foram analisadas por meio de histogramas para a identificação do tipo de assimetria (negativa, positiva ou neutra), em especial da variável Fe (%) que apresenta distribuição assimétrica negativa característica, conforme observado no histograma (Fig. 9a), para a identificação dos *outliers* construíram-se *box-plots* (Fig. 9b), os demais gráficos, encontram-se no apêndice I e II.

As estatísticas descritivas dos teores químicos globais das amostras são apresentadas na Tabela 1, a estatística descritiva por litologia se encontra no apêndice III.

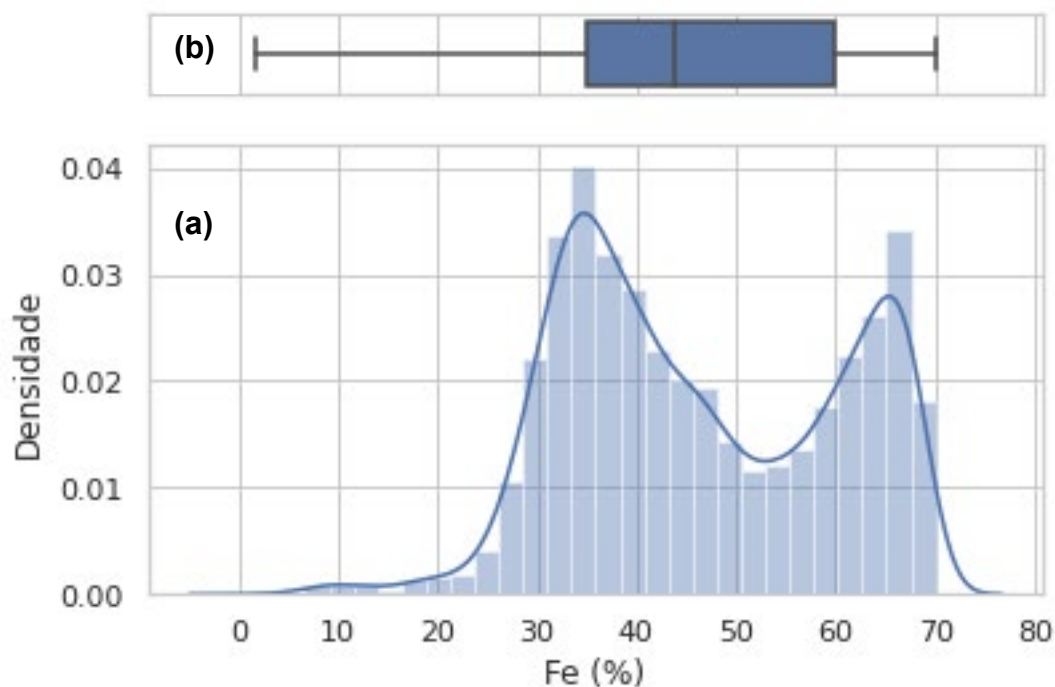


Figura 9: a) Histograma de Fe (%), com distribuição assimétrica negativa, b) e *Box-plot* da variável Fe (%), apresenta duas populações, identificável pela presença dos dois picos de frequência (fonte: o autor).

Tabela 1 – Estatística descritiva dos teores globais das amostras, Q25, Q50 e Q75, correspondem respectivamente ao primeiro, segundo e terceiro quartil (fonte: o autor).

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
				Padrão			
<i>Al</i>	0,012	31,919	1,260	1,550	0,400	0,780	1,570
<i>Pf</i>	0,001	37,376	2,001	2,568	0,480	1,140	2,530
<i>P</i>	0,006	1,092	0,052	0,053	0,020	0,040	0,060
<i>Si</i>	0,100	93,456	29,036	20,567	5,410	33,360	47,070
<i>Fe</i>	1,810	69,970	46,820	13,930	35,090	44,350	60,860

5.1.2 Litotipos

Apresenta-se na Tabela 2 são apresentados os critérios de classificação dos litotipos de minério para a mina de Capitão do Mato, conforme o utilizado na Vale S.A.. Nesta classificação não estão discriminados os litotipos estéreis, quartzito ferruginoso (QF) e itabirito dolomítico (IDO) e Canga (CG).

Tabela 2 – Chave de classificação litológica para formação ferrífera da mina de Capitão do Mato, sendo G1, a porcentagem da amostra presente na primeira faixa granulométrica(fonte: o autor).

Código	Litotipos	Fe %	Mn %	Al₂O₃ %	Pf %	G1
HC	Hematita Compacta	≥ 62	< 1,00	< 3,00	< 3,50	≥ 50
HF	Hematita Friável	≥ 62	< 1,00	< 3,00	< 3,50	< 50
HCT	Hematita Contaminada	≥ 60	Mn ≥ 1,00 e/ ou Al ≥ 3,00 e/ ou Pf ≥ 3,50			
IFR	Itabirito Friável Rico	≥ 52 e < 62	< 1,00	< 3,00	< 3,50	< 50
IF	Itabirito Friável	≥ 20 e < 52	< 1,00	< 3,00	< 3,50	< 50
IC	Itabirito Compacto	≥ 20 e < 62	< 1,00	< 3,00	< 3,50	≥ 50
IAL	Itabirito Aluminoso	≥ 20 e < 60	-	≥ 3,00	-	-
IGO	Itabirito Goethítico	≥ 20 e < 60	-	< 3,00	≥ 3,50	-
IMN	Itabirito Manganêsífero	≥ 20 e < 60	≥ 1,00	-	-	-

Pode-se observar na Figura 10 a proporção de cada litotipo na base de dados, é possível notar que há um percentual maior de litotipos friáveis, cerca de 40% e litotipos compactos, cerca de 35%, as amostras contaminadas correspondem a 20% das amostras e o restante é composto por amostras de rochas estéreis (IDO e QF).

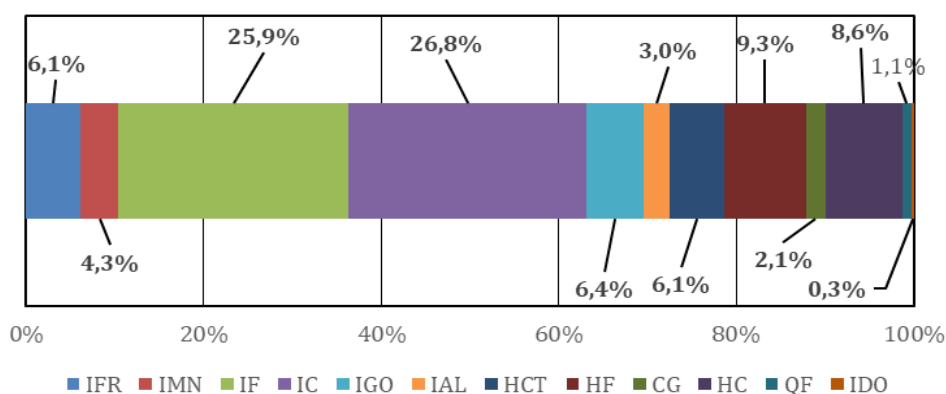


Figura 10: Proporção de litologias no banco de dado, Hematita Compacta (HC), Hematita Friável (HF), Hematita Contaminada (HCT), Itabirito Friável Rico (IFR), Itabirito Friável (IF), Itabirito Compacto (IC), Itabirito Aluminoso (IAL), Itabirito Goethítico (IGO), Itabirito Manganêsífero (IMN), Quartzito Ferruginoso (QF) e Utibirito Dolomítico (IDO) e Canga (CG). (fonte: o autor).

5.1.3 Estatística bivariada

A análise estatística bivariada foi realizada para definir a melhor abordagem para as análises agrupamentos e buscar identificar variáveis com coeficiente de correlação de

pearson elevadas, cujo módulo seja maior que 0,9. A Figura 11 contém a matriz de correlação entre as variáveis de interesse, observa-se que o ferro apresenta uma forte correlação negativa com a sílica, como esperado para amostras de depósitos do tipo BIF.

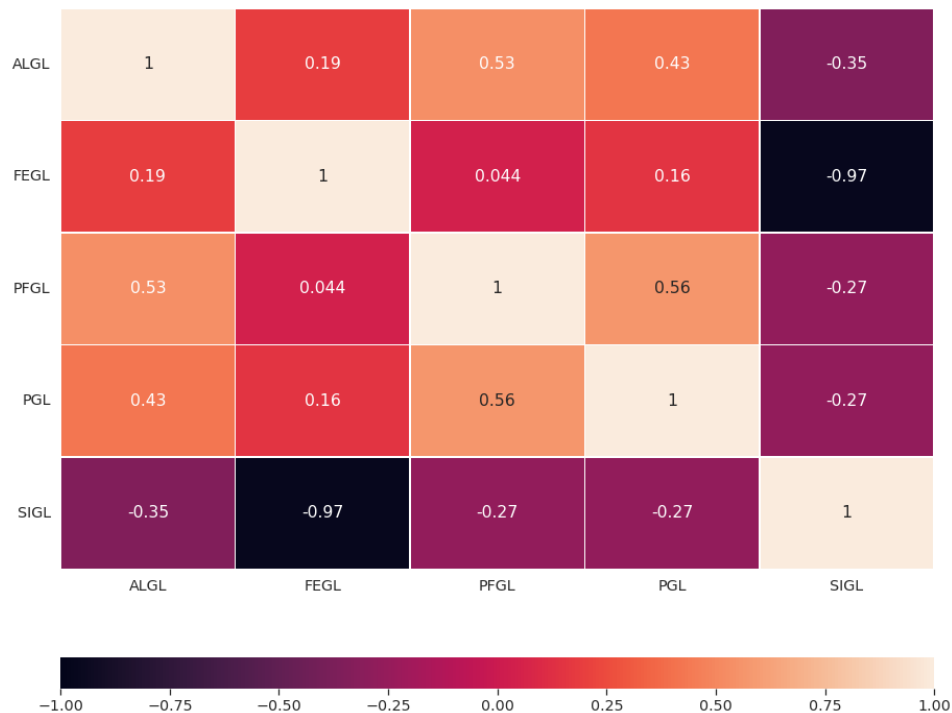


Figura 11: Matriz de correlação entre amostras sem agrupamento, GL=Teor global do parâmetro na amostra (fonte: o autor).

A correlação entre Si e Fe, quando analisada por litologia, apresenta um comportamento variado, alta correlação negativa nas amostras de Itabirito e menor correlação nas amostras de hematita e CG, com correlação positiva nas amostras de IDO, como observado nos diagramas de dispersão da Figura 12.

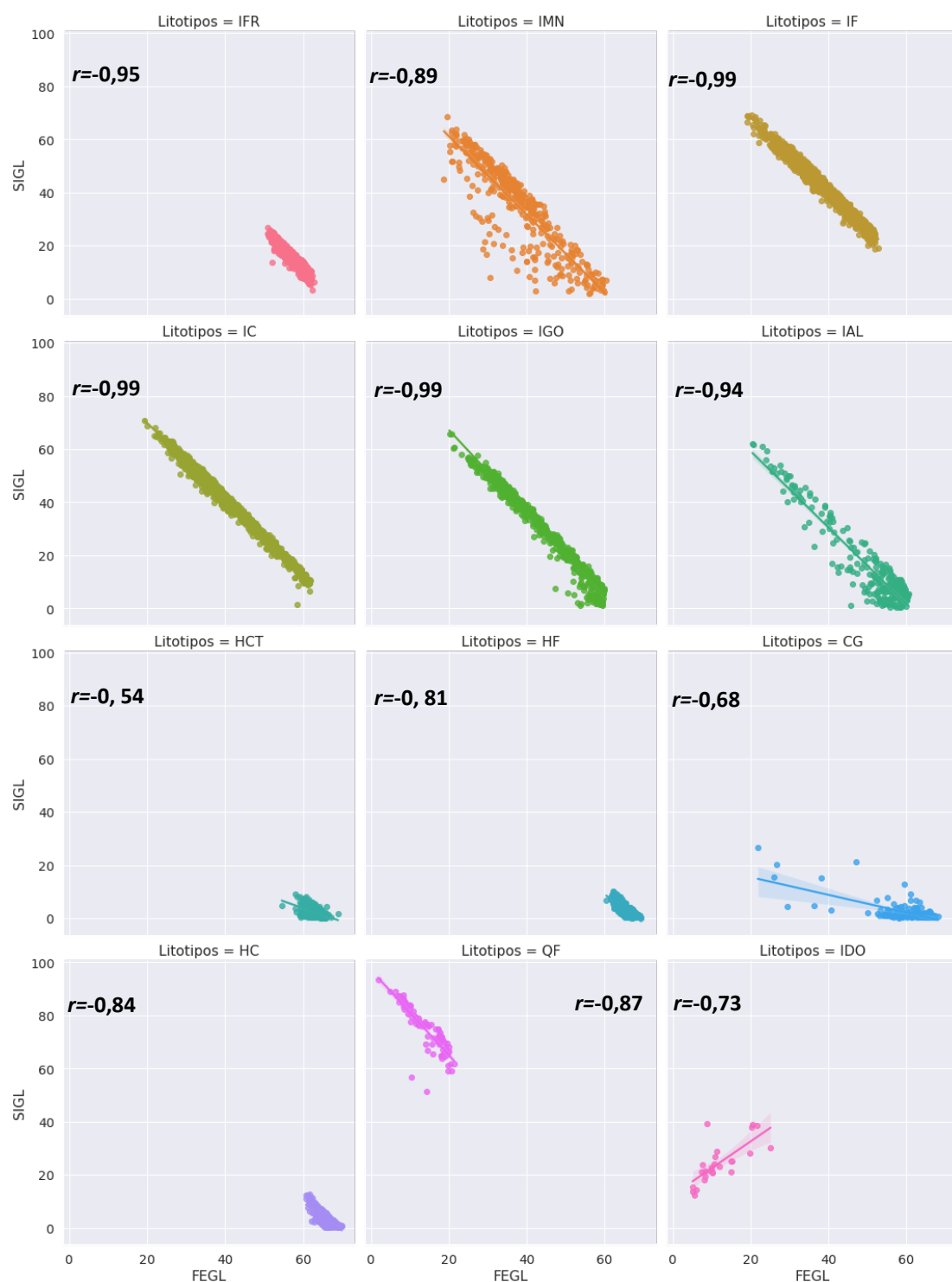


Figura 12: Gráficos de dispersão entre os teores globais de Si e Fe, com respectivos coeficientes de correlação separados por litotipo

5.2. Agrupamentos

O primeiro agrupamento foi manualmente realizado, classificaram-se as amostras em rejeito, minério rico e minério pobre, de acordo com a chave de classificação da Tabela 3, sendo o rejeito o grupo das amostras de teor de $\text{Fe}(\%) \leq 45\%$ e/ou alto teor de contaminantes, o minério rico, as amostras de Hematita (HC, HF, HCT) e CG com $\text{Fe}(\%) \geq 60\%$ e, por fim, o minério pobre, amostras de Itabirito (IF, IC, IAL, IGO, IMN) e CG com teor de $\text{Fe}(\%) \leq 60\%$

Tabela 3 – Chave de classificação de minério de Ferro para formação ferrífera da mina de Capitão do Mato, as amostras de Hematita não apresentam valores mínimos pois por definição tem teores de Fe acima de 60% e baixo nível de contaminantes sendo, portanto, sempre classificados como minério de rico. (fonte: o autor).

Código	Litotipos	Fe %	Si %	Al %	Pf %
HC	Hematita Compacta	-	-	-	-
HF	Hematita Friável	-	-	-	-
HCT	Hematita Contaminada	-	-	-	-
IFR	Itabirito Friável Rico	≥ 45	≤ 52	-	-
IF	Itabirito Friável	≥ 45	-	-	-
IC	Itabirito Compacto	≥ 45	-	-	-
IAL	Itabirito Aluminoso	≥ 45	-	≤ 8	-
IGO	Itabirito Goethítico	≥ 45	-	-	≤ 8
IMN	Itabirito Manganésífero	≥ 45	≤ 52	-	-
CG	Canga	≥ 45	-	≤ 8	-

Os outros agrupamentos foram obtidos com os algoritmos KM, HC e GHC a partir das variáveis-critério (ferro, alumínio, fósforo, silício e perda ao fogo), considerando-se as correlações sílica-ferro observadas na análise estatística bivariada, o uso da variável Si (%) foi dispensado para evitar ruídos e a geração de grupos redundantes. A aplicação dos algoritmos, ocorreu com as amostras já regularizadas para evitar que as diferentes escalas enviassem os agrupamentos.

De acordo com o método do cotovelo (Fig. 13), o agrupamento mais eficiente para K-médias, por meio da aplicação da biblioteca *scikit-learn* (Pedregosa *et al.*, 2011), ocorre com $k=3$, visto que é o ponto onde ocorre a inflexão na soma dos erros aos quadrados (SEQ), conforme a execução do *elbow method* presente na biblioteca *Scikit-learn*, a escolha de um

número menor de grupos resultaria em agrupamentos com amostras muito heterogêneas entre si e um número maior em grupos pouco representativos devido ao baixo número de amostras, o algoritmo foi executado a predefinição de três agrupamentos até que foi possível determinar padrão de agrupamento mais comum ao final das execuções.

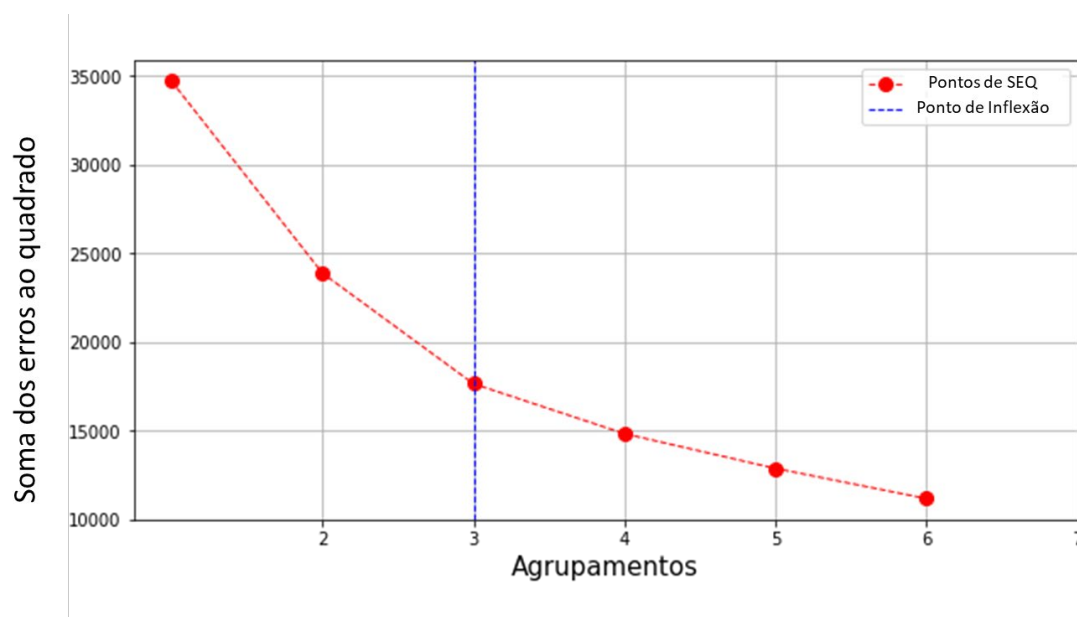


Figura 13: O valor K pode ser melhor determinado traçando a curva K-SEQ, utilizando a biblioteca *scikit-learn*, encontrando o ponto de inflexão, neste caso o ponto de inflexão ocorre quando $k = 3$ (fonte: o autor).

A aplicação do método aglomerativo hierárquico, por meio da biblioteca *Scikit-learn* (Pedregosa *et al.*, 2011), e do aglomerativo hierárquico geoestatístico (*software Isatis.neo*), foram realizadas de forma a gerarem três grupos de agrupamento, isto é, escolheram-se os níveis hierárquicos que resultavam em três agrupamentos, utilizando-se do método de semelhança de proximidade de Ward para ambos os algoritmos.

A interface do aglomerativo hierárquico geoestatístico no *software Isatis.neo*, permitiu a personalização dos pesos das variáveis, atribuindo-se um peso 2 para a variável Fe e incluindo a variável categórica de classificação litológica das amostras, rotacionou-se a direção de agrupamento para N110, de acordo com a malha de amostragem e por fim, definiu-se que os agrupamentos formados deveriam ter ao menos 500 amostras para evitar a geração de grupos muito pequenos.

5.2.1 Agrupamento Manual

As estatísticas descritivas dos teores químicos das amostras agrupadas manualmente são apresentadas no Apêndice III e os histogramas de Fe dos domínios na Figura 14. Cada domínio apresenta uma distribuição de frequência semelhante a forma de um intervalo do histograma de Fe (%), apresentado na Figura 14, o domínio de rejeito apresenta uma distribuição próxima a normal sendo equivalente ao primeiro pico de

frequência, o domínio pobre equivale ao intervalo entre picos e o domínio rico ao intervalo do segundo pico de frequência.

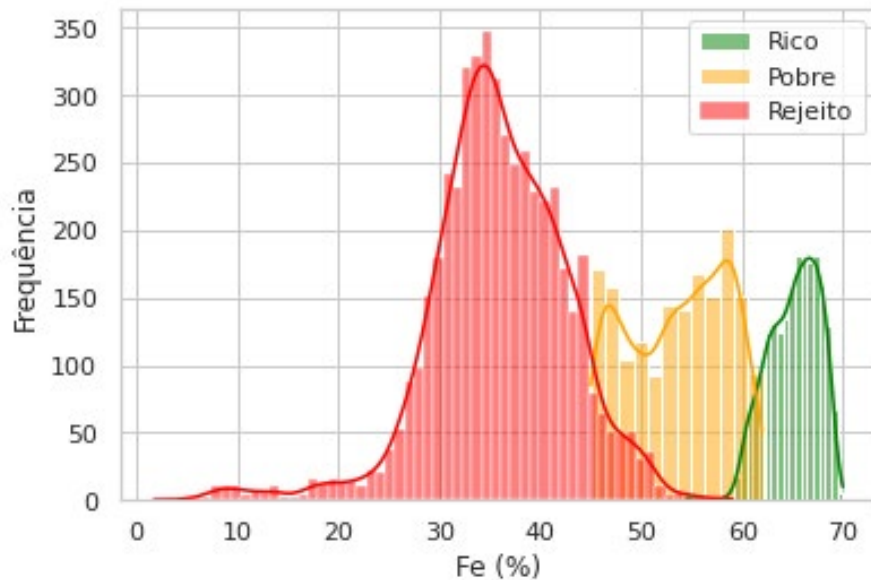


Figura 14: Histograma de Fe (%) com distribuição assimétrica negativa, para as amostras manualmente classificadas como minério rico, pobre e rejeito (fonte: o autor).

5.2.2 K-médias

As estatísticas descritivas dos teores das amostras agrupadas por K-médias são apresentadas no apêndice III e os histogramas de Fe dos domínios na Figura 15. O domínio rejeito apresenta uma distribuição próxima a normal, enquanto os domínios de minério apresentam forte assimetria assimétrica negativa.

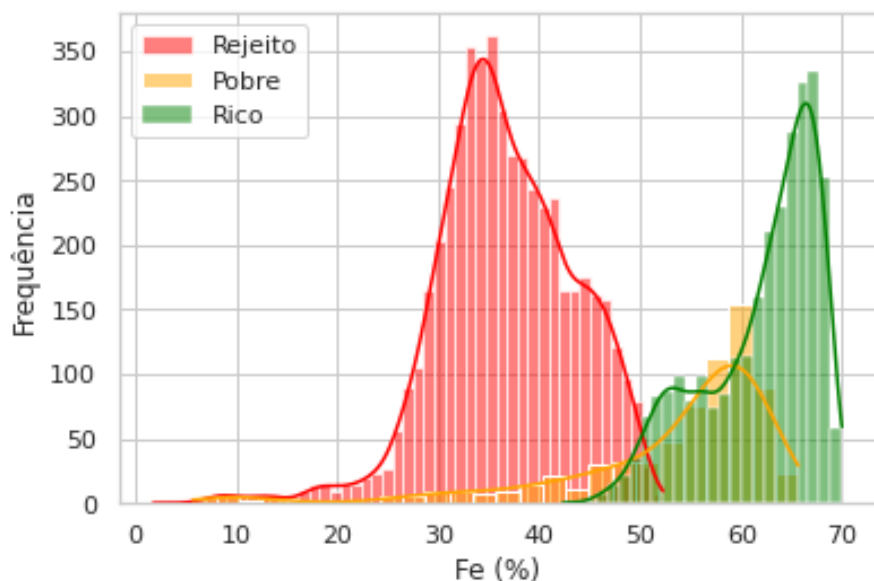


Figura 15: Histograma de Fe (%), para as amostras agrupadas por meio do algoritmo K-médias (fonte: o autor).

5.2.3 Aglomerativo Hierárquico

As estatísticas descritivas dos teores das amostras agrupadas por aglomerativo hierárquico são apresentadas no apêndice III e os histogramas de Fe dos domínios na Figura 16, a distribuição de frequência dos domínios ocorre de forma semelhante àquelas obtidas pelo algoritmo K-médias.

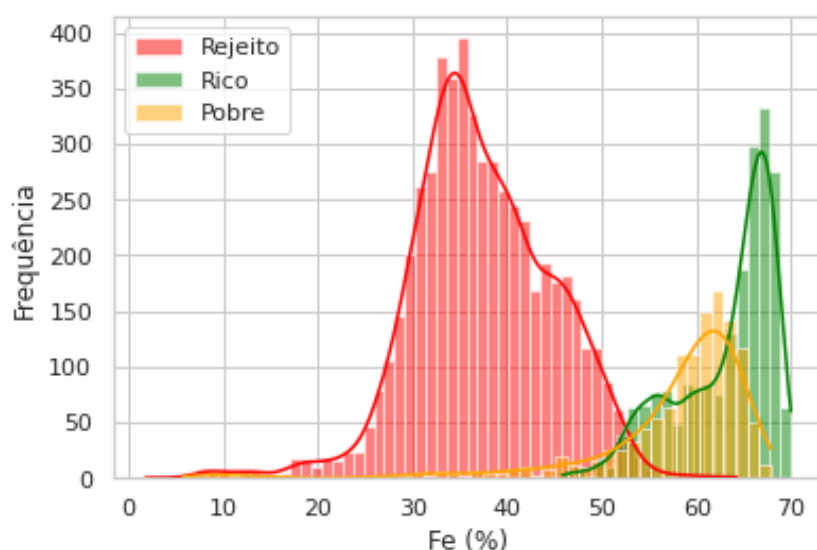


Figura 16: Histograma de Fe (%), para as amostras agrupadas por meio do algoritmo *aglomerativo hierárquico* (fonte: o autor).

5.2.4 Aglomerativo Hierárquico Geoestatístico

As estatísticas descritivas das análises químicas das amostras agrupadas por aglomerativo hierárquico geoestatístico são apresentadas no apêndice III e os histogramas de Fe dos domínios na Figura 17., a distribuição dos domínios pobre e rico são semelhantes, havendo grande sobreposição dos teores destes grupos.

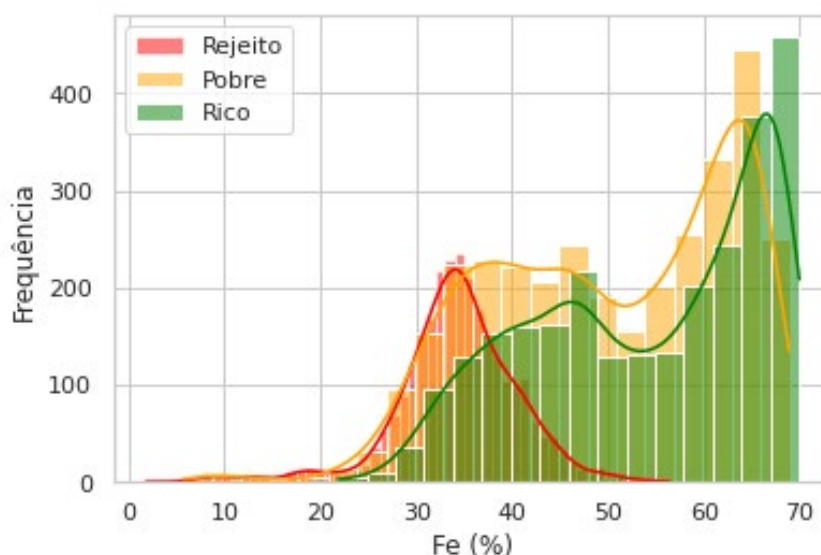


Figura 17: Histograma de Fe (%), para as amostras agrupadas por meio do algoritmo *aglomerativo hierárquico geoestatístico* (fonte: o autor).

5.3. Classificação e Distribuição dos Domínios

Com a análise estatística de cada cluster apresentada no apêndice III e sintetizada na Tabela 4, verificou-se a tendência de todos os algoritmos distinguirem: um grupo de alto teor de ferro e baixa concentração de contaminantes (domínio rico), um grupo com teor de ferro menor e maior concentração média de contaminantes, Al e Pf (domínio pobre) e um grupo com teor de ferro média menor que 37% e concentração intermediária de contaminantes (domínio rejeito).

Tabela 4 – Teores médios das variáveis critérios em todos os métodos de agrupamento e das amostras não agrupadas. (fonte: o autor).

Domínio	Cluster	Al (%)	Fe (%)	Pf (%)	P (%)
KM	<i>0 (Rejeito)</i>	0,77	36,67	1,39	0,04
	<i>1 (Pobre)</i>	4,40	52,90	8,01	0,15
	<i>2 (Rico)</i>	1,47	61,99	1,81	0,05
HC	<i>0 (Pobre)</i>	3,65	57,58	5,98	0,12
	<i>1 (Rejeito)</i>	0,79	37,35	1,51	0,04
	<i>2 (Rico)</i>	1,17	63,14	1,16	0,04
GHC	<i>0 (Rejeito)</i>	0,84	34,26	1,78	0,05
	<i>1 (Pobre)</i>	1,81	50,05	2,33	0,06
	<i>2 (Rico)</i>	1,14	54,20	2,06	0,05
Manual	<i>Rejeito</i>	0,96	35,93	1,71	0,04
	<i>Pobre</i>	1,89	54,14	2,95	0,07
	<i>Rico</i>	1,61	65,17	2,18	0,06
Único	-	1,30	46,29	2,08	0,05

A distribuição espacial dos agrupamentos é apresentada na Figura 18, os resultados dos algoritmos K-médias e aglomerativo hierárquico são semelhantes entre si e ao agrupamento manual e o aglomerativo geoestatístico hierárquico apresenta uma divisão bem marcada entre a porção NW e SE, com a primeira sendo predominantemente composta pelo *cluster Rico* e a segunda porção pelo *cluster pobre*.

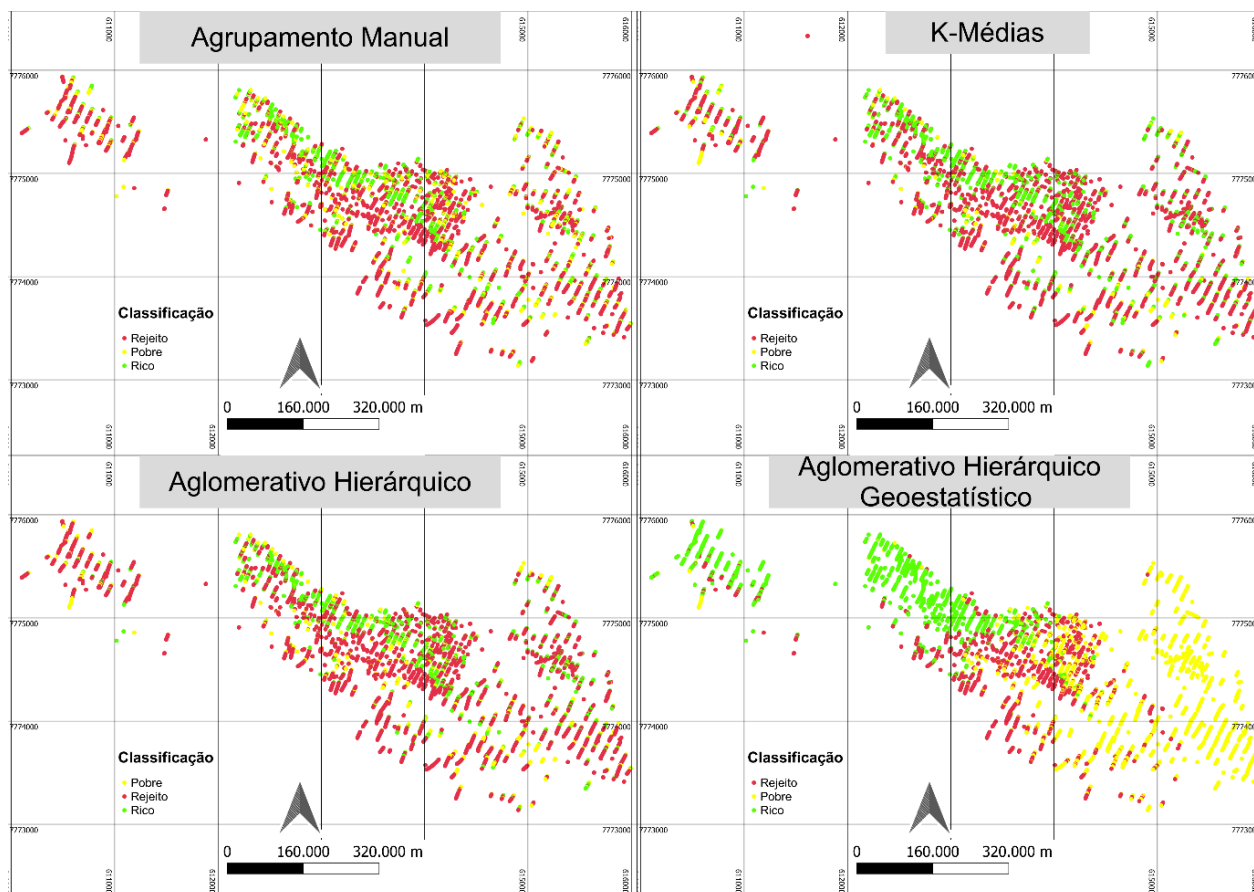


Figura 18: Distribuição espacial das amostras com a legenda de clusters por algoritmo de análise de agrupamento (fonte: o autor).

Para melhor compreensão desses agrupamentos realizaram-se os *box-plots* das variáveis critérios (Fig. 19), em comparação ao domínio manual, verifica-se que possuem distribuições estatísticas semelhantes, evidenciando que as coincidências entre métodos não estão restritas aos valores médios. As diferenças entre os domínios definidos pelo manual e os formados por algoritmos são:

A remoção de *outliers*, para as variáveis P, Pf e Al, nos domínios ricos gerado pelos métodos KM e KC, em relação, ao domínio rico manualmente definido.

Os domínios gerados pelo GHC diferenciam-se dos formados pelo método manual pela menor concentração de contaminantes no domínio Rico.

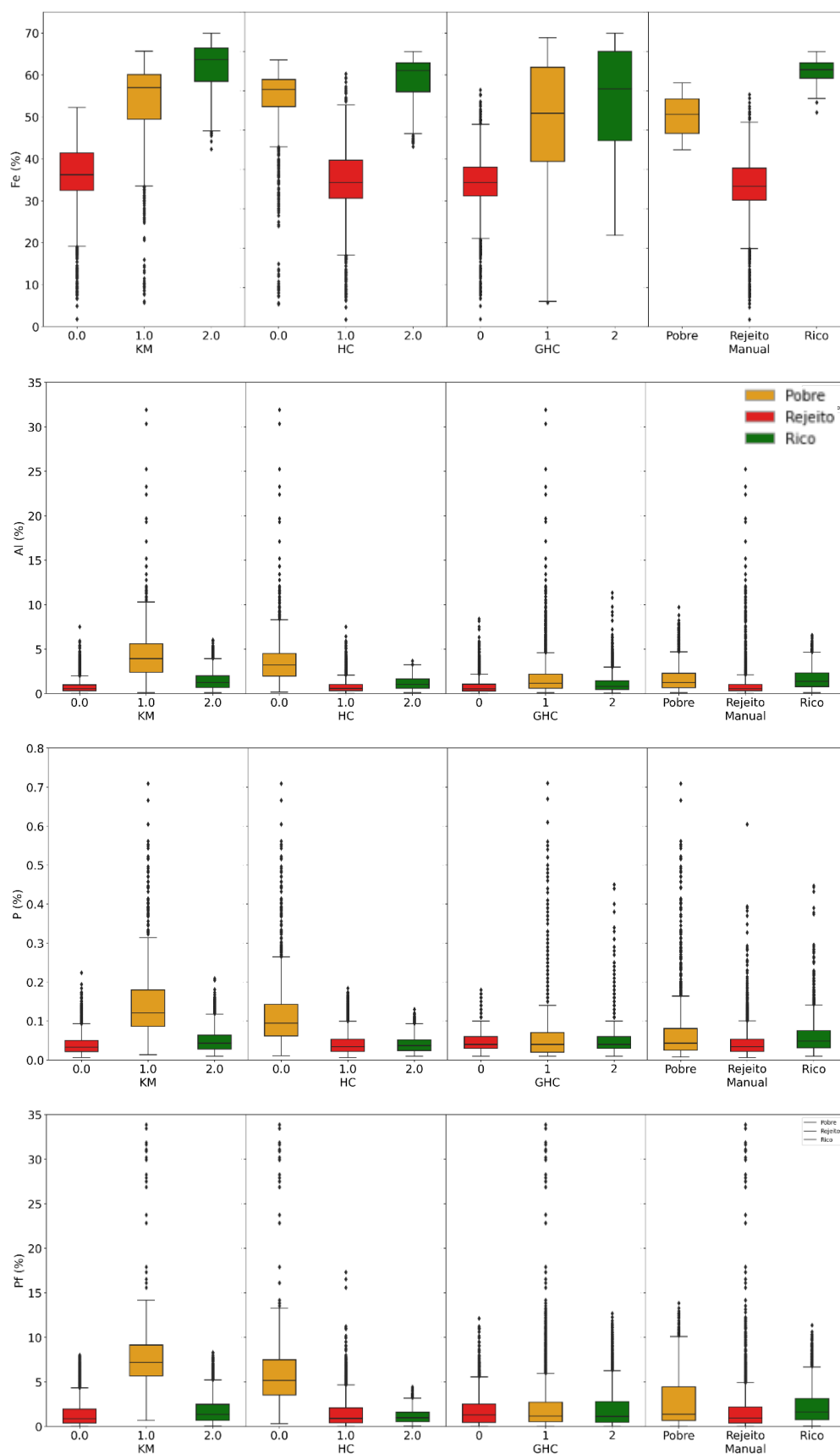


Figura 19: Box-plots dos teores de Fe (%), Al (%), P (%) e Pf (%) em cada agrupamento utilizado (K-médias (KM), Aglomerativo Hierárquico (HC) e Aglomerativo Hierárquico Geoestatístico (GHC) e manual) (fonte: o autor).

A distribuição dos litotipos em cada domínio (Fig. 20), é marcada por grandes diferenças entre os cenários de agrupamento:

- Hematitas: Classificação de algumas amostras de HCT pelos algoritmos como parte do domínio pobre. O método GHC agrupou outras significativas quantidades de amostras de HF como pertencentes ao domínio pobre.
- Itabiritos não contaminados: Os algoritmos classificaram algumas amostras como parte do domínio rico, as amostras de IFR são consideradas inteiramente parte do domínio rico pelo algoritmo HC. O agrupamento GHC, reduz a quantidade de amostras de IF e IC, classificadas como rejeito, as distribuindo para o domínio pobre.
- Itabiritos contaminados e encaixantes: Os algoritmos, apresentam distribuição semelhante ao agrupamento manual das amostras contaminadas, com variação das amostras de IMN. Os métodos HC e GHC classificam algumas amostras de IMN como parte do domínio rico, as amostras de QF são identificadas por todos os métodos como parte do rejeito, enquanto as de IDO são reconhecidas como parte do domínio pobre por todos os algoritmos.

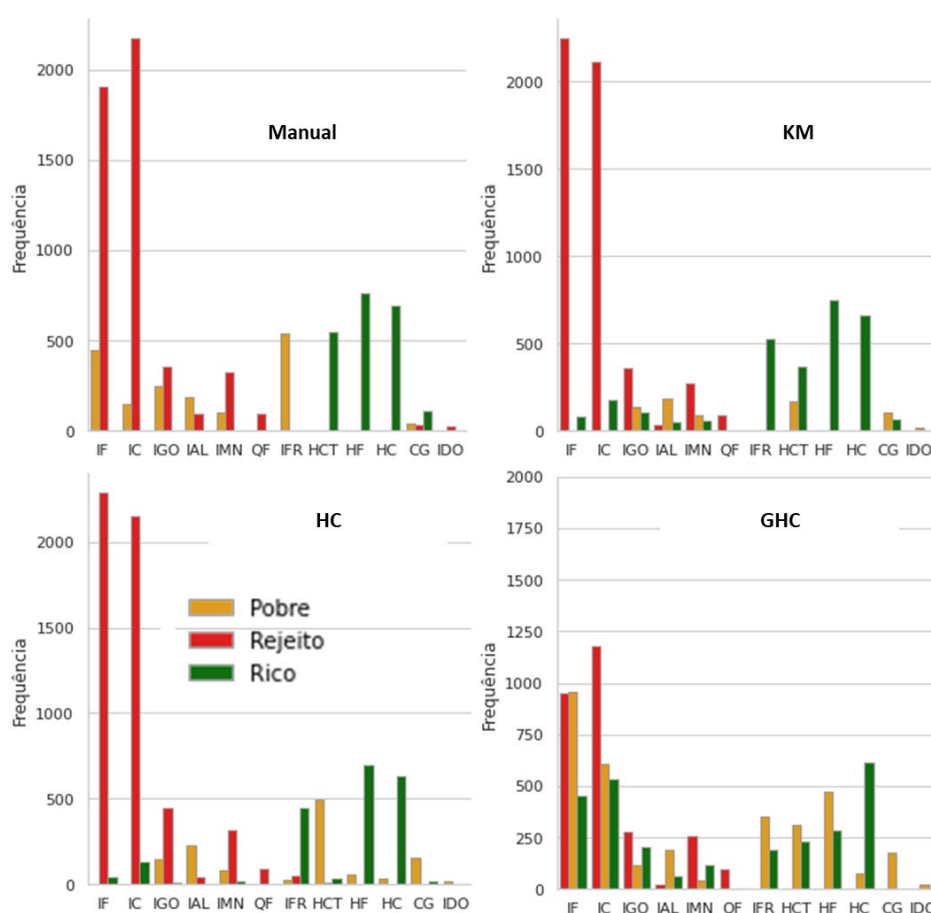


Figura 20: Distribuição dos litotipos em cada *cluster*, por agrupamento manual e pelos algoritmos de análise de agrupamento. K-médias (KM), Aglomerativo Hierárquico (HC) e Aglomerativo Hierárquico Geoestatístico (GHC) (fonte: o autor).

5.4. Geoestatística

O cálculo dos variogramas experimentais da variável ferro foi realizado para onze diferentes domínios, os 10 obtidos pelos métodos de agrupamento, para minério pobre e rico, além do domínio único, sempre com tamanho do passo igual a mediana das distâncias entre amostras nas direções horizontais e 10 metros para a direção normal ao plano de referencial (plano horizontal).

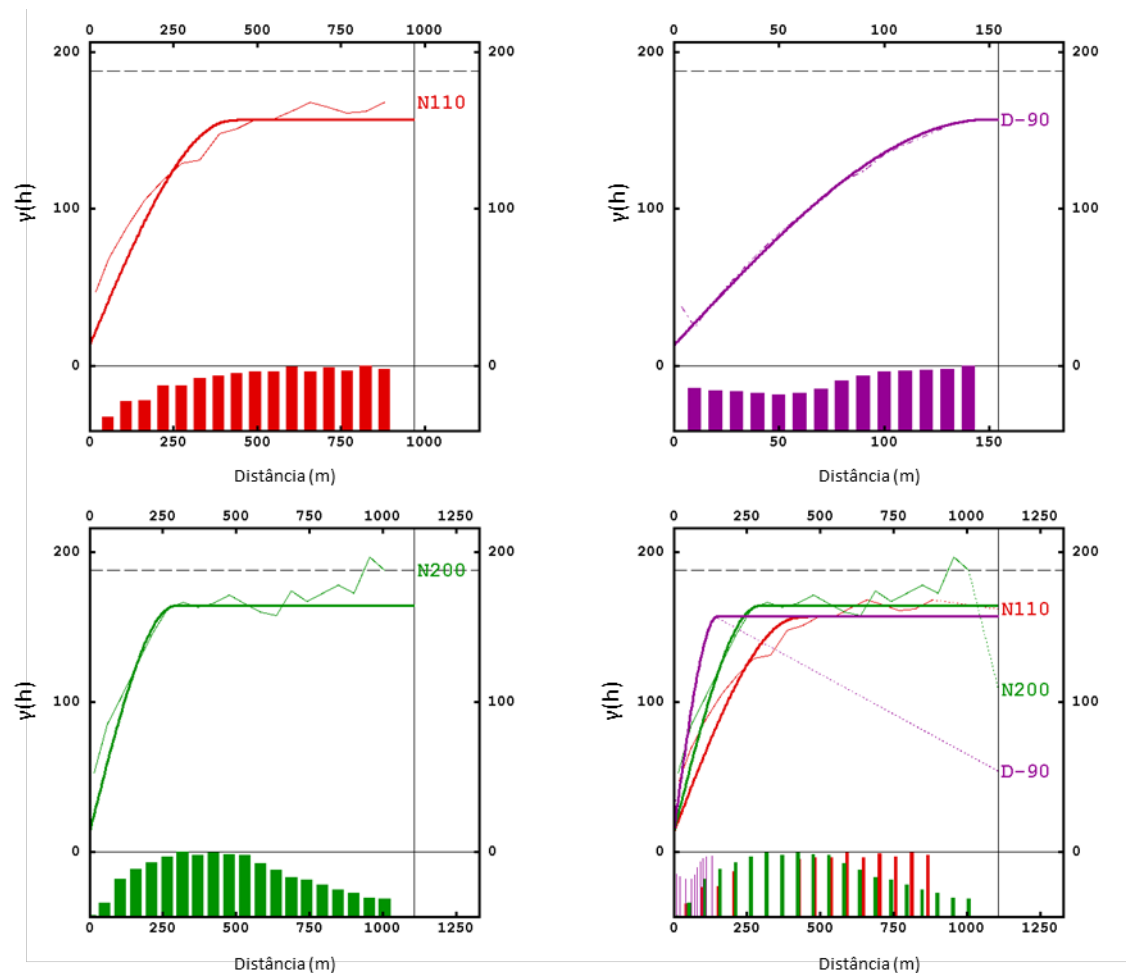
Em todos os cálculos, utilizou-se a direção de maior continuidade do domínio único, N110°, e a direção ortogonal a ela N200°, além da direção de *downhole*. Os efeitos pepita utilizados foram definidos a partir do variograma experimental do *downhole* para melhor qualidade do ajuste nesta direção.

Apresenta-se na Figura 21 o ajuste dos modelos teóricos de variogramas do domínio único, que são bem estruturados nas direções *downhole* e N110°, com patamar abaixo da variância a priori. As barras na parte inferior dos variogramas indicam os números de pares de pontos obtidos em cada passo calculado. Os ajustes dos outros modelos teóricos de variograma, os outros cenários podem ser observados no Apêndice V.

5.5. Estimativa de Teores de Ferro

Os raios de busca foram configurados com as dimensões e atitudes das elipsoides de continuidade. Variou-se os parâmetros de vizinhança, setores em quadrante ou octante e variação de amostras de 4 a 12 por setores, buscando-se o melhor valor de coeficiente de correlação de *Pearson* na validação cruzada. Os blocos foram discretizados para 2x2x1, conforme o praticado na indústria para a estimativa de recursos minerais.

Em cada cenário, blocos do subdomínio minério de Fe pobre foram estimados apenas com amostras do domínio pobre, utilizando-se a mesma lógica para o subdomínio minério rico e todas as amostras foram utilizadas na estimativa sem agrupamento.



Direção	Estrutura 1	Estrutura 2	Estrutura 3
Efeito Pepita		13	
U	415	500	∞
V	290	300	300
W	150	∞	∞
Patamar	120	24	7

Figura 21: Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio Único, respectivamente N110°, Downhole, N200° e as 3 direções juntas (fonte: o autor).

5.5.1 Domínio Único

O modelo de blocos estimado no domínio único permite o reconhecimento de 3 grandes domínios de teores de Fe (Fig. 22), o primeiro na porção NE com concentrações elevadas, com Fe (%) ultrapassando 66%, outra em SE com teores de Fe de até 63%, e por fim uma grande concentração de baixos teores na porção SE-NW, concentrações pontuais de Fe (%) com variação de 45% até 60%, ocorrem dispersas neste último domínio.

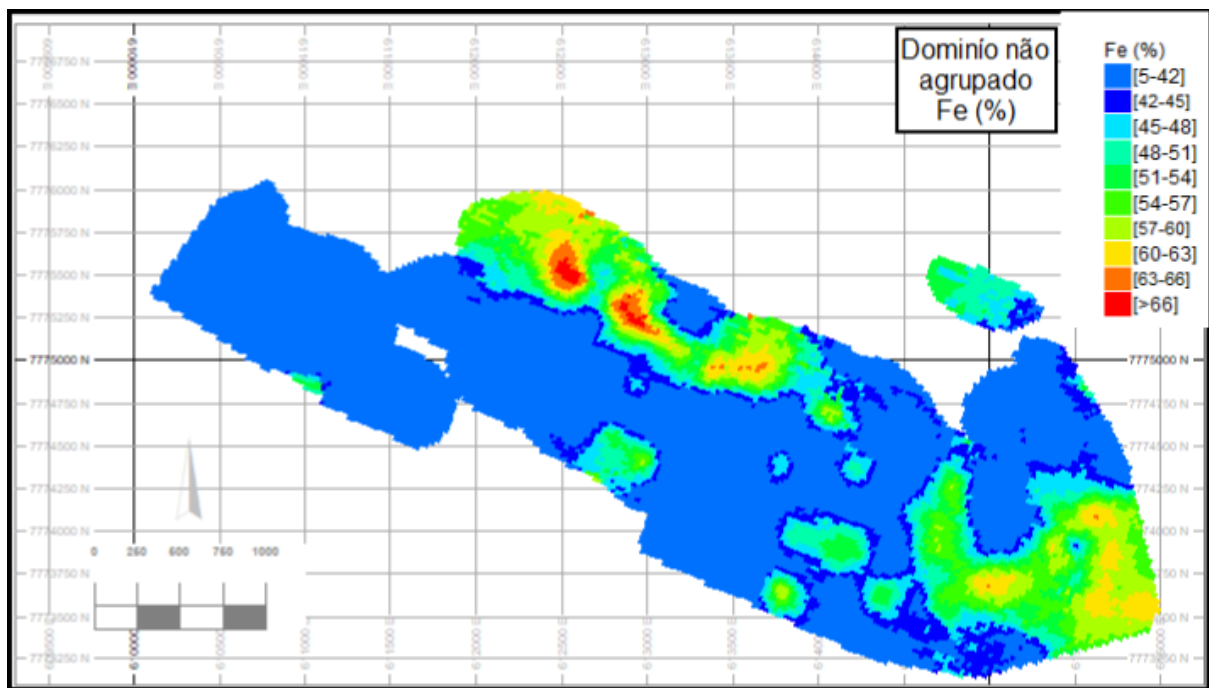


Figura 22: Vista em planta do modelo de blocos com teores estimados de ferro no cenário não agrupado (fonte: o autor).

5.5.2 Domínios Pobre e Rico

Os modelos de blocos estimados nos domínios pobres e ricos, são apresentados nas Figuras 23 e 24, respectivamente, os blocos estimados correspondem, espacialmente, as concentrações de baixo e alto teor apresentadas no domínio Único.

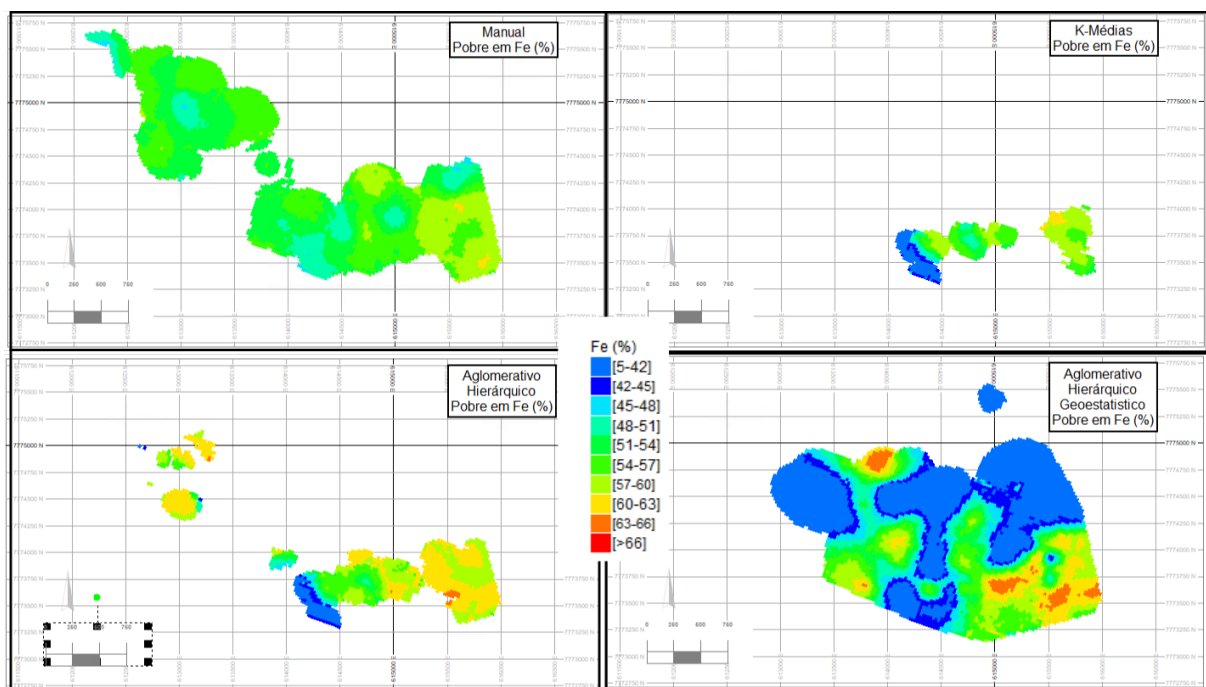


Figura 23: Vista em planta dos modelos de blocos com teores estimados de ferro para os agrupamentos de minério pobre (fonte: o autor).

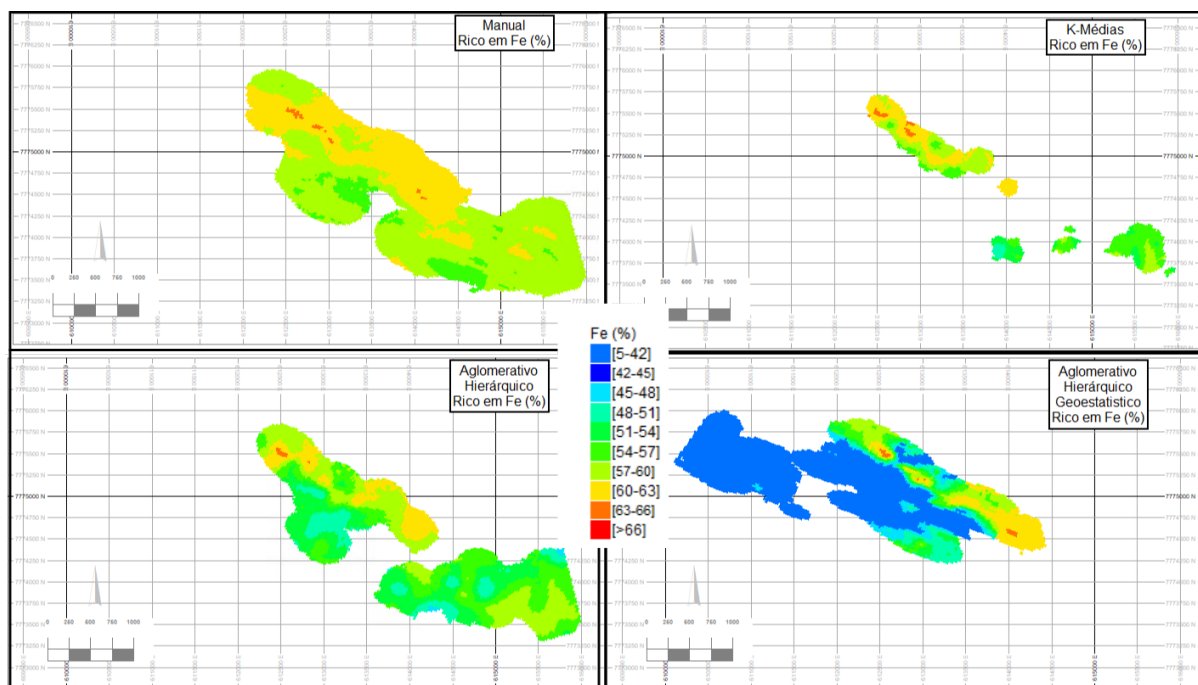


Figura 24: Vista em planta dos modelos de blocos com teores estimados de ferro para os agrupamentos de minério rico (fonte: o autor).

5.6. Estatística e Validação das Estimativas

Os resultados de validação cruzada para cada cenário são apresentados na Tabela 5, por meio do coeficiente de *Pearson*, as estimativas de GHC apresentam maior coeficiente médio entre seus 2 domínios, enquanto o domínio pobre KM apresenta o maior coeficiente de *Pearson* entre todos os cenários.

Tabela 5 – Coeficiente de Pearson da validação cruzada, K-médias (KM), Aglomerativo Hierárquico (HC) e Aglomerativo Hierárquico Geoestatístico (GHC) (fonte: o autor).

<i>Método</i>	<i>Domínio</i>	<i>Pearson</i>
<i>KM</i>	<i>Pobre</i>	0,913
	<i>Rico</i>	0,655
<i>HC</i>	<i>Pobre</i>	0,878
	<i>Rico</i>	0,574
<i>GHC</i>	<i>Pobre</i>	0,889
	<i>Rico</i>	0,876
<i>Manual</i>	<i>Pobre</i>	0,409
	<i>Rico</i>	0,760
<i>Único</i>	-	0,910

Observando a Tabela 6, verifica-se que todos os agrupamentos resultaram em estimativas com médias mais elevadas do que à média amostral e o teor médio de Fe (%) estimado no domínio único, com exceção do domínio pobre Aglomerativo Hierárquico Geoestatístico (GHC), o oposto ocorre com o desvio padrão, devido a uma distribuição com menor variabilidade das amostras nos agrupamentos.

Tabela 6 – Estatísticas descritivas dos teores de ferro nas amostras e nos modelos estimados em cada cenário, todos os valores estão em (%), K-médias (KM), Aglomerativo Hierárquico (HC) e Aglomerativo Hierárquico Geoestatístico (GHC). (fonte: o autor).

Método	Domínio	Média	Mínimo	Máximo	Desvio Padrão
<i>KM</i>	<i>Pobre</i>	52,94	6,75	64,97	0,19
	<i>Rico</i>	60,92	46,49	70,33	0,06
<i>HC</i>	<i>Pobre</i>	57,80	6,47	67,02	0,13
	<i>Rico</i>	62,95	51,61	69,59	0,05
<i>GHC</i>	<i>Pobre</i>	48,41	6,61	70,41	0,26
	<i>Rico</i>	53,22	24,90	74,36	0,20
<i>Manual</i>	<i>Pobre</i>	54,25	45,46	66,00	0,06
	<i>Rico</i>	64,39	57,36	69,82	0,03
<i>Único</i>	-	49,18	4,49	75,17	0,21

Pela análise dos gráficos quantil-quantil (Apêndice VI), observa-se que as estimativas se afastam da distribuição amostral nas duas extremidades, devido ao efeito de suavização de estimativa. Os cenários que menos se afastam são os do agrupamento manual, apresentando uma suavização muito baixa quando comparada a estimativa do cenário não agrupado (Fig. 25).

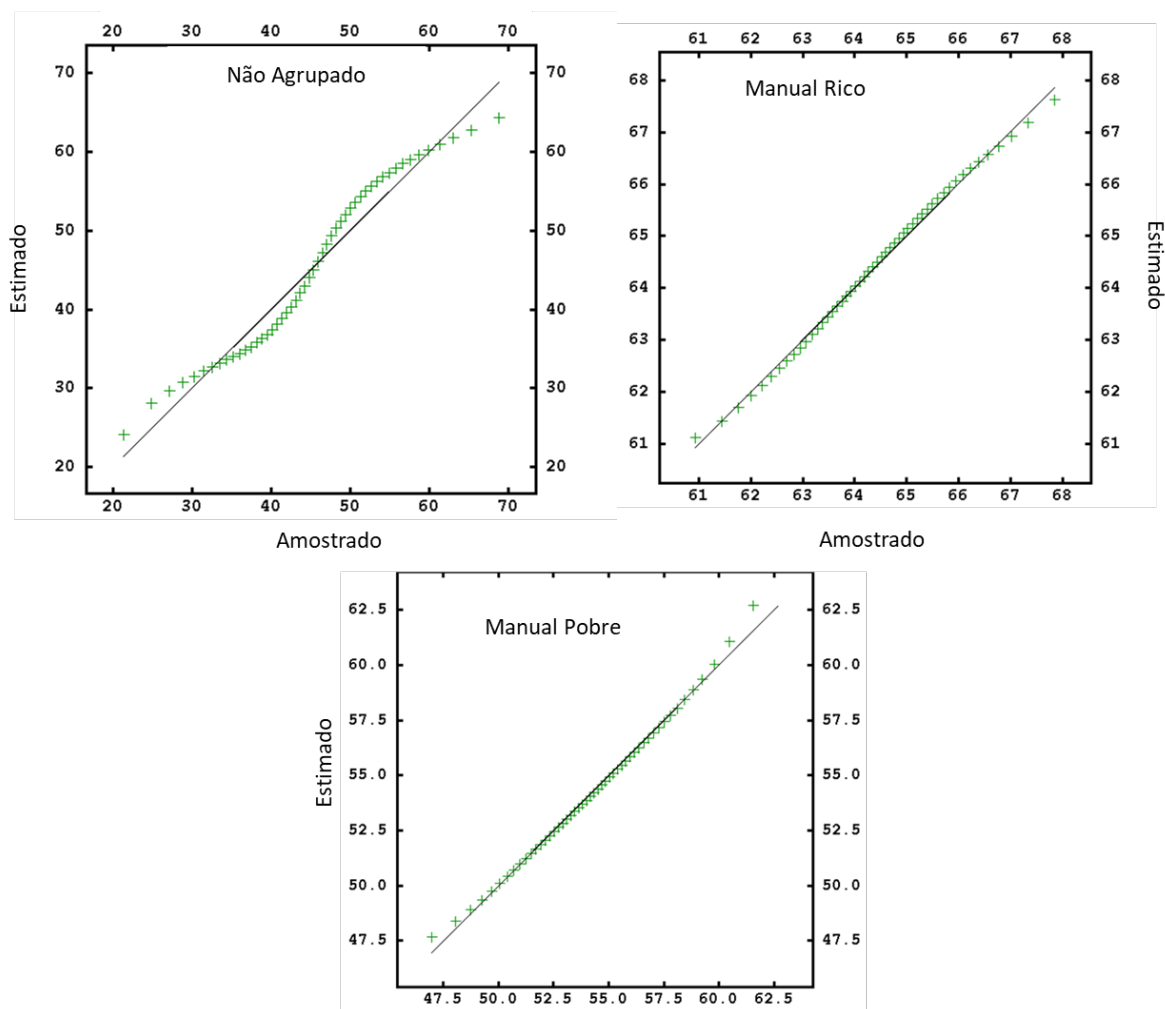


Figura 25: Gráficos quantil-quantil entre os teores de ferro do conjunto amostral e as estimativas dos cenários com domínio único e com agrupamento manual, para o domínio pobre e rico (fonte: o autor).

6. INTERPRETAÇÕES DOS RESULTADOS

6.1. Agrupamentos

O dendrograma da Figura 26, apresenta a classificação dos agrupamentos realizados pelo HC, nele é possível identificar que o maior nível hierárquico diferencia os teores de Al, gerando um grupo que posteriormente foi interpretado como domínio pobre, com alta concentração de Al e um grupo que precede a formação dos grupos “rico” e “rejeito”, que é formado com as amostras com concentração de Al menores que 3%, o nível hierárquico seguinte, distingue os grupos de Alto teor e baixo teor de ferro, gerando assim a divisão entre o domínio de minério “rico” e “rejeito”, desta forma gerando os 3 domínios utilizados neste trabalho.

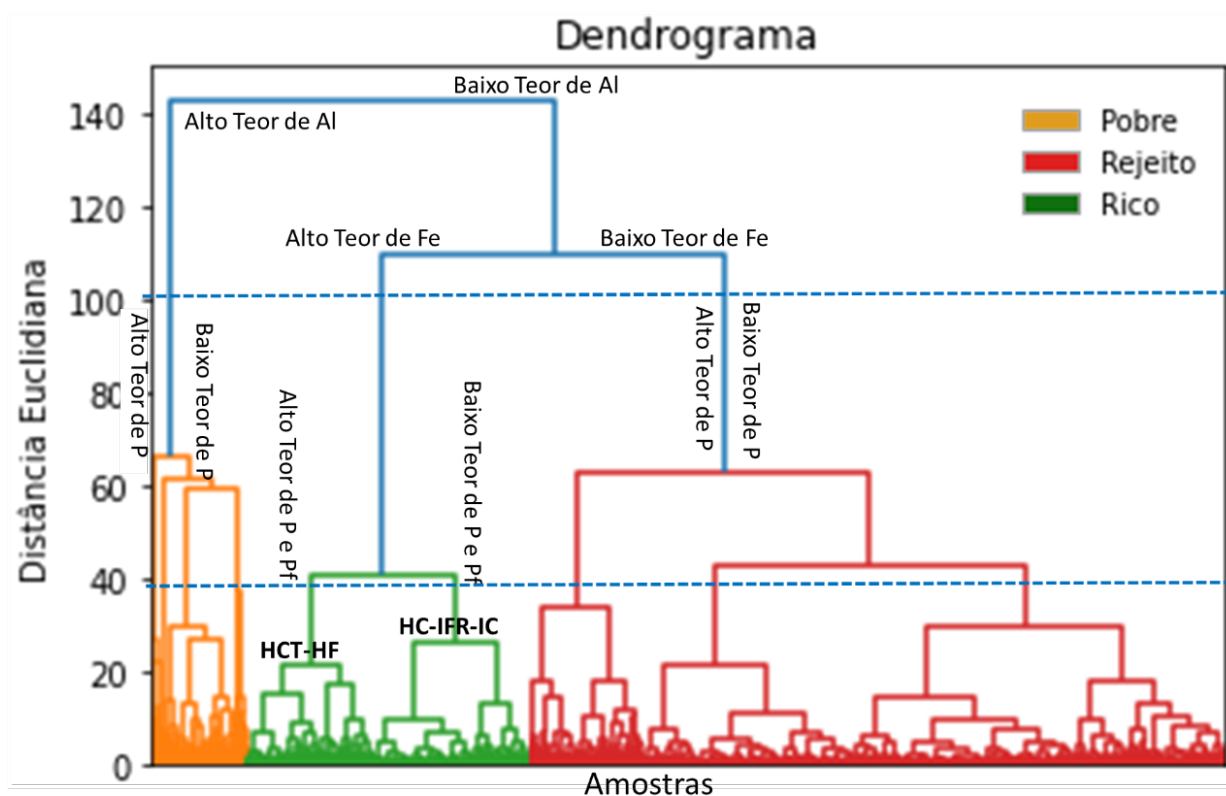


Figura 26: Dendrograma do agrupamento algoritmo Aglomerativo Hierárquico, realizado pelo método de Ward, indicando a distância euclidiana utilizada neste trabalho (~100), bem como a discutida no texto de cerca de 40, apresenta ainda as divisões feitas em cada nível hierárquico e eventual separação das hematitas no domínio rico (fonte: o autor).

A análise do dendrograma (Fig. 26) reforça a ideia de 3 domínios, indicado pelo método do cotovelo do algoritmo K-médias, utilizados neste trabalho, visto que caso 2 grupos fossem estabelecidos não haveria a diferenciação entre amostras com alto e baixo teor de ferro, já que a divisão seria feita apenas com base no teor de Al.

Já a escolha de grupos maiores que 3, por exemplo quando é feita a escolha do nível hierárquico próximo de 40, gerariam agrupamentos com poucas amostras, divisão ainda maior das amostras de Hematitas, devido as diferenças de concentrações de P e Pf, perda considerável da continuidade espacial e uma maior dificuldade na classificação do que é rejeito ou não.

6.2. Geoestatística

A direção N110 foi a de maior continuidade em todos os domínios estimados, evidenciando um controle estrutural-litológico da formação minério da mina de CMT, ligados a sinclinal Moeda e a evolução do QF.

A análise geoestatística em agrupamentos, apresentou menor variância em todos os métodos em relação ao domínio único, devido a menor variabilidade das amostras. Os agrupamentos de domínios pobres apresentam, em geral, pior estruturação dos seus variogramas experimentais (Apêndice V), o que ocorre ligado a redução no número de amostras.

A direção *downhole*, apresenta o comportamento mais heterogêneo entre os agrupamentos, entretanto, quando existe um grupo que concentra as amostras de itabiritos, não há a formação de um patamar dentro do campo geométrico, como é o caso do agrupamento manual, mas em casos de distribuição dos itabiritos entre os domínios, como ocorre no método KM, os variogramas experimentais ficam melhor estruturados (Fig. 27).

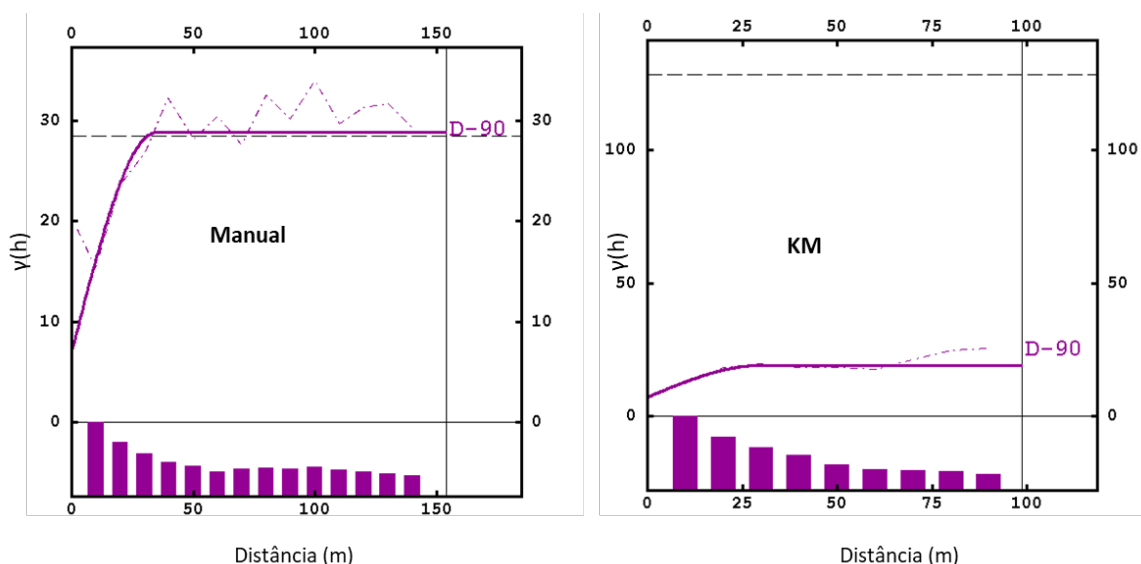


Figura 27: Ajuste dos modelos de Variograma para o agrupamento pobre manualmente definido e pelo algoritmo KM.

6.3. Validação e comparação dos resultados

Os métodos de agrupamento apresentam distribuição estatística de menor variabilidade e maiores teores médios quando comparados a estimativa de domínio único, mas com menor continuidade espacial, principalmente pelos métodos K-médias e

Aglomerativo Hierárquico. Os resultados de estimativa, continuidade espacial e suavização mais semelhantes ao domínio único foram obtidos a partir do método aglomerativo hierárquico geoestatístico.

Destacam-se as estimativas, do domínio pobre gerado pelo método K-médias, por apresentar o variograma experimental mais bem estruturado, resultado de validação cruzada mais próximos a 1 e dos domínios manualmente definidos que apresentam as menores suavizações de estimativa.

A menor continuidade espacial das estimativas de agrupamentos, não inviabiliza seus resultados, visto que os domínios estimados podem ser reconhecidos na estimativa de domínio único.

As estimativas feitas com base no domínio único e pelo algoritmo aglomerativo hierárquico geoestatístico, aparentam ser mais adequadas nas estimativas de planejamento a longo prazo, visto que é importante conhecer a variação espacial dos teores médios e a continuidade espacial do depósito. Já as estimativas com definição de subdomínios por aglomerativo Hierárquico, K-médias e manualmente definidos, podem ser aplicados no planejamento de curto prazo visto que reduzem os valores de suavização, e distribuem melhor os teores dos contaminantes, o que impacta diretamente no avanço da frente de lavra, no beneficiamento do minério e no controle de teores nas pilhas de rejeito,

7. CONCLUSÕES

O uso do método do cotovelo no algoritmo K-médias, para estabelecer qual o número de agrupamentos ideal, mostrou-se válido também para o aglomerativo hierárquico e aglomerativo hierárquico geoestatístico, com o valor $K=3$ correspondentes ao número de grupos manualmente definidos.

Distinguem-se em todos os cenários três domínios, o primeiro rico em ferro, o segundo pobre em ferro e com maiores teores de Al e por fim um terceiro grupo de rejeito, sendo composto de amostras abaixo do teor de beneficiamento de ferro e/ou alto teor de contaminantes que inviabilizam o beneficiamento.

A análise geoestatística evidencia o forte controle estrutural geológico da mineração da mina de Capitão do Mato, visto que em todos os cenários a direção de N110 foi a de maior continuidade no plano horizontal, ainda foi possível reconhecer o impacto de um agrupamento composto predominantemente de itabiritos. nos modelos experimentais de variograma, sem a presença de amostras de hematita não há a estabilização dos valores de variância dentro do campo geométrico.

Os métodos de agrupamento apresentam distribuição estatística com menor variabilidade e maiores teores médios quando comparados a estimativa de domínio único,

mas com menor continuidade espacial. Destacam-se a estimativa do domínio pobre gerado pelo algoritmo K-médias, que apresenta o melhor resultado de validação cruzada e melhor estruturação do variograma e as estimativas dos domínios manuais que apresentam efeito de suavização menos pronunciados.

A aplicação do aglomerativo hierárquico geoestatístico apresenta semelhança com os resultados de domínio único, possivelmente ligada a atribuição de um maior peso à variável ferro e, principalmente, pela restrição, condicionada a distribuição espacial dos dados. Sua aplicação é fácil e intuitiva quando incorporada a um *software* (no caso o *Isatis.neo® mining*), devido a fácil personalização dos pesos, números de agrupamentos e a incorporação de variáveis categóricas, o que auxilia na compreensão da base de dados e dispensando o conhecimento de linguagens de programação. A incorporação de eventuais novas amostras em agrupamentos pré-definidos é feita facilmente, sem a necessidade de reagrupar todas as amostras ou utilizar métodos extensos e complexos como ocorre com a análise Discriminante Múltipla para os algoritmos K-médias e hierárquico aglomerativo.

A menor continuidade espacial dos agrupamentos não inviabiliza os seus resultados, mas a comparação dos resultados do agrupamento manual com os realizados por algoritmo, não apresentam diferenças estatísticas (ou de precisão das estimativas) significativas, portanto, não se justifica o tempo para geração de domínios pelo algoritmo, sendo assim recomenda-se a estimativa com domínios manualmente gerados para a mina de Capitão do Mato.

A utilização dos algoritmos, podem ser aplicadas para amostragens que ainda não apresentam agrupamento manual definido, visto que o *elbow method*, apresenta o número ótimo de grupos igual ao operacionalmente definido na mina de Capitão do Mato.

Os métodos de agrupamento não supervisionados devem ter suas aplicações testadas em depósitos de substâncias com alta variabilidade, como ouro, em malhas de amostragem menos densas, ou em cenários de planejamento de lavra em curto prazo, visto que nesta etapa é preciso definir melhor o contraste entre minério rico, pobre e rejeito.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- Abzalov, M., 2016. Modern Approaches in Solid Earth Sciences: Applied Mining Geology, v. 12, New York, Springer International Publishing Switzerland, 443 p.
- Alkmim, F. F. e Teixeira, W. 2017. São Francisco Craton, Eastern Brazil. Regional Geology Reviews. Springer International Publishing Switzerland, 326 p.
- Armstrong, M., 1998. The Variogram in Basic Linear Geostatistics p. 25–46.
- Borradaile, G., 2003. Statistics Of Earth Science Data. 353 p.
- Bubnova, A., Ors, F., Rivoirard, J., Cojan I., Romary T., Automatic Determination of Sedimentary Units from Well Data. in Math Geosci V52, p. 213–231.
- Clout, J.M.F. e Manuel, J.R., 2022. Chapter 2 - Mineralogical, chemical, and physical metallurgical characteristics of iron ore, Editor(s): Liming Lu, In Woodhead Publishing Series in Metals and Surface Engineering, Iron Ore, Woodhead Publishing, v.2 p. 59-108.
- Dorr, J. V. N., 1969. Physiographic, Stratigraphic and Structural Development of the Quadrilátero Ferrífero Minas Gerais, Brazil. Geological Survey Professional Paper 641-A.
- Deutsch, C. V., 2003. Geostatistics, in Encyclopedia of Physical Science and Technology, Academic Press, v.2, p. 697-707.
- Faraj, F. e Ortiz, J., 2021. Mining, Metallurgy & Exploration, v. 38, New York, Springer International Publishing Switzerland, p. 1609-1623
- Fouedjio, F., 2016. A hierarchical clustering method for multivariate geostatistical data. in Spatial Statistics. v.18 p. 333-351.
- Fouedjio, F., 2020. Clustering of multivariate geostatistical data. WIREs Computational Statistics v.12 p. 1510-1512.
- Hagemann, S.G., Angerer, T., Duuring, P., Rosière, C.A., Figueiredo e Silva, R.C., Lobato, L., Hensler, A.S., Walde, D.H.G., 2016. BIF-hosted iron mineral system: A review, Ore Geology Reviews, v. 76, p.317-359.
- Hexmoor H., 2015. Chapter 6 - Diffusion and Contagion, In Emerging Trends in Computer Science and Applied Computing, Computational Network Science, Morgan Kaufmann, 45-64p.

Holmes, R. J., Lu, Y., Lu, L., 2022. Chapter 1 - Introduction: Overview of the global iron ore industry, Editor(s): Liming Lu, In Woodhead Publishing Series in Metals and Surface Engineering, Iron Ore Woodhead Publishing, v.2, p. 1-56.

IPT-Instituto de Pesquisas Tecnológicas do Est. de São Paulo, 1989. Estudo e implementação de novas técnicas e desenvolvimento de programas em geoestatística: aplicações em geologia. São Paulo, IPT. 98p. (IPT Relatório 27.799).

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.

Landim, P.M.B., 2011. Análise estatística de dados geológicos multivariados. São Paulo, Editora Oficina De Textos, 208 p.

Lloyd, S.P., 1982. Least squares quantization in PCMIEEE Transactions on Information Theory, v.28, n.2., 129-137p.

McCool M., Robison A. D., Reinders J., 2012. K-Means Clustering, 279-289p.

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.

Matheron, G., 1971. The Theory Of Regionalized Variables And Its Applications. Paris, Les Cahiers Du Centre De Morphologie Mathematique De Fontainebleu. 211 p.

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

Revuelta, M. B., 2017. Mineral Resources: From Exploration to Sustainability Assessment. Springer. 653 p.

Romary, T., Rivoirard, J., Deraisme, J., Quinones, C., Freulonm, X., 2012. Domainning by clustering multivariate geostatistical data. Ninth International Geostatistics Congress, Norway, France. p. 455-466.

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

Rosière, C. A. e Chemale Jr., F., 2000. Itabiritos E Minérios De Ferro De Alto 30 Teor Do Quadrilátero Ferrífero – Uma Visão Geral E Discussão. Geonomos, Instituto de Geociências, Universidade Federal de Minas Gerais, Belo Horizonte - MG, Brasil v.8 n.2 p. 155.

Rosière, C. A. e Rolim, V. K., 2016. Formações ferríferas e minério de alto teor associado, O minério de ferro no Brasil –geologia, metalogênese e economia. Recursos Minerais no Brasil: problemas e desafios. Rio de Janeiro: Academia Brasileira de Ciências, p. 32-45.

Rossi, M. E., Deutsch C. V., 2014. Mineral Resource Estimation. Springer, 332p.

Rossum, G. V. e Drake, F. L. 2009. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Trauth, M.H., 2007. Estatísticas bivariais. In: Matlab® Recipes for Earth Sciences. Springer, Berlim, Heidelberg. p.61-82.

Teichgraeber H e Brandt, A.R., 2018. Systematic Comparison of Aggregation Methods for Input Data Time Series Aggregation of Energy Systems Optimization Problems, Editor(s): Mario R. Eden, Marianthi G. Ierapetritou, Gavin P. Towler, Computer Aided Chemical Engineering, Elsevier, v. 44, 955-960p.

Siegel A. F. 2012. Correlation and Regression: Measuring and Predicting Relationships, in: Practical Business Statistics, v.6 p. 291-345

Suckau, V.E., Costa, T.A.V., Suita, M.T.F., Oliveira, D.M., Ferreira, F.A., 2004. Aspectos estruturais, petrológicos e metalogenéticos da mina de ferro do Tamanduá, MBR, Quadrilátero Ferrífero, MG. XLII Congr. Brasil. de Geol., Anais, Araxá, Outubro 2004; 45-64p.

Vale, 2022. Relatório Anual da Administração, 227p. Disponível em: vale.com/brasil/PT/investors/information-market/annualreports/20f/Paginas/default.aspx (acessado em 08/2022).

Vale, 2021, Relatório de Referência, v.4, 750p. Disponível em: vale.com/brasil/PT/investors/information-market/annual-reports/reference-form (acessado em 01/2022).

Varajão, C.A.C., Bruanderick, A., Ramanaidou Robert, R., Gilkes, J., 2002. Microporosity of BIF hosted massive hematite ore, Iron Quadrangle, Brazil. Anais da Academia Brasileira de Ciências [online]. 2002, v. 74, n. 1, p. 113-126.

Wackernagel H., 2003. Multivariate Geostatistics An Introduction with Applications. Berlin Springer. 387 p.

Waskom, M. L., 2021. Seaborn: statistical data visualization, Journal of Open Source Software, v.6, n.60, p.3021.

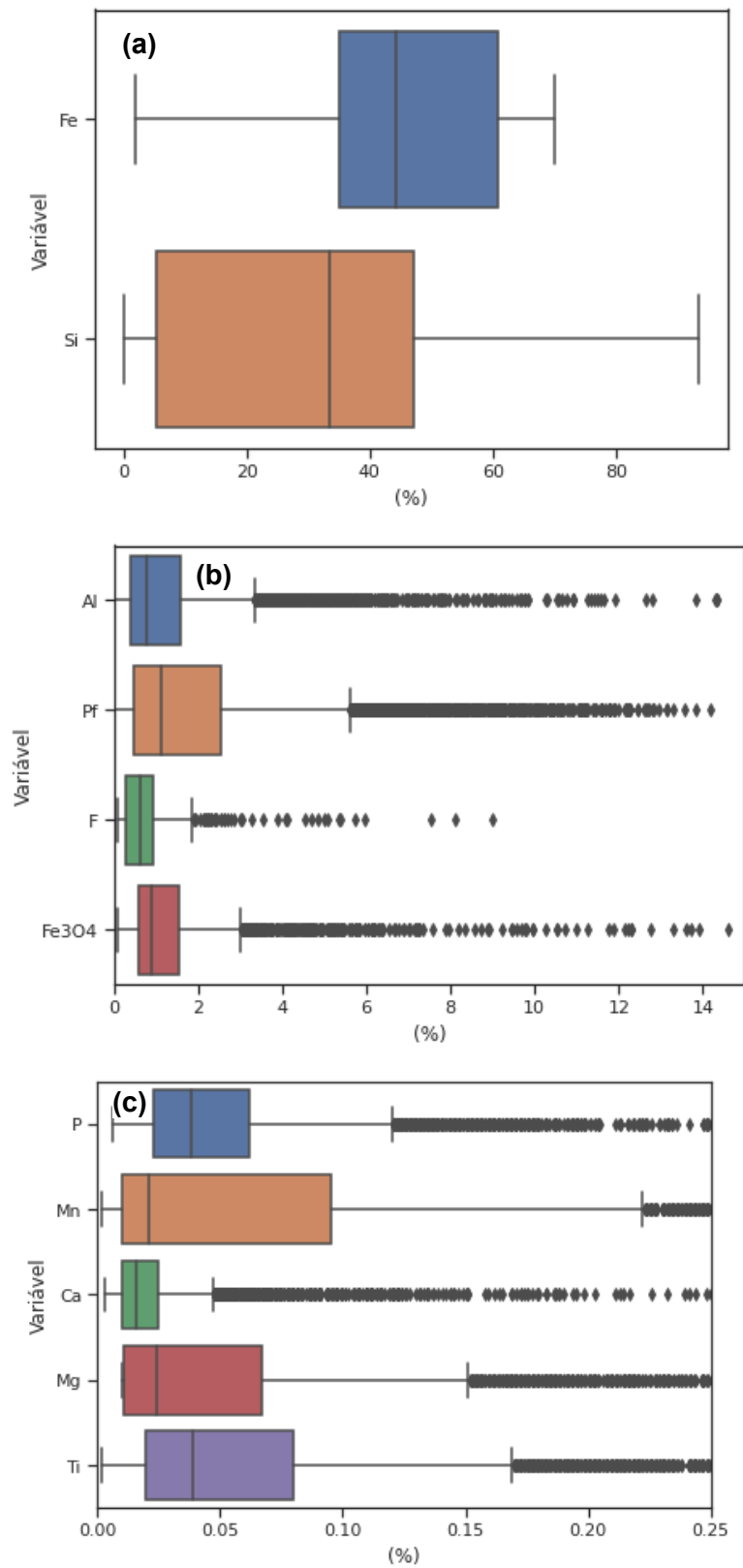
Yamamoto, J. K. 2001. Avaliação E Classificação De Reservas Minerais. São Paulo, Editora Da Universidade De São Paulo, 226 p.

Yamamoto, J. K., Bettencourt, J. S. e Montanheiro, T. J., 2001. Análise estatística. in Avaliação e classificação de reservas minerais. p. 49-68.

Yamamoto, J.K. e Landim, P.M.B., 2013. Geoestatística. Conceitos E Aplicações. São Paulo, Editora Oficina De Textos, 215 p.

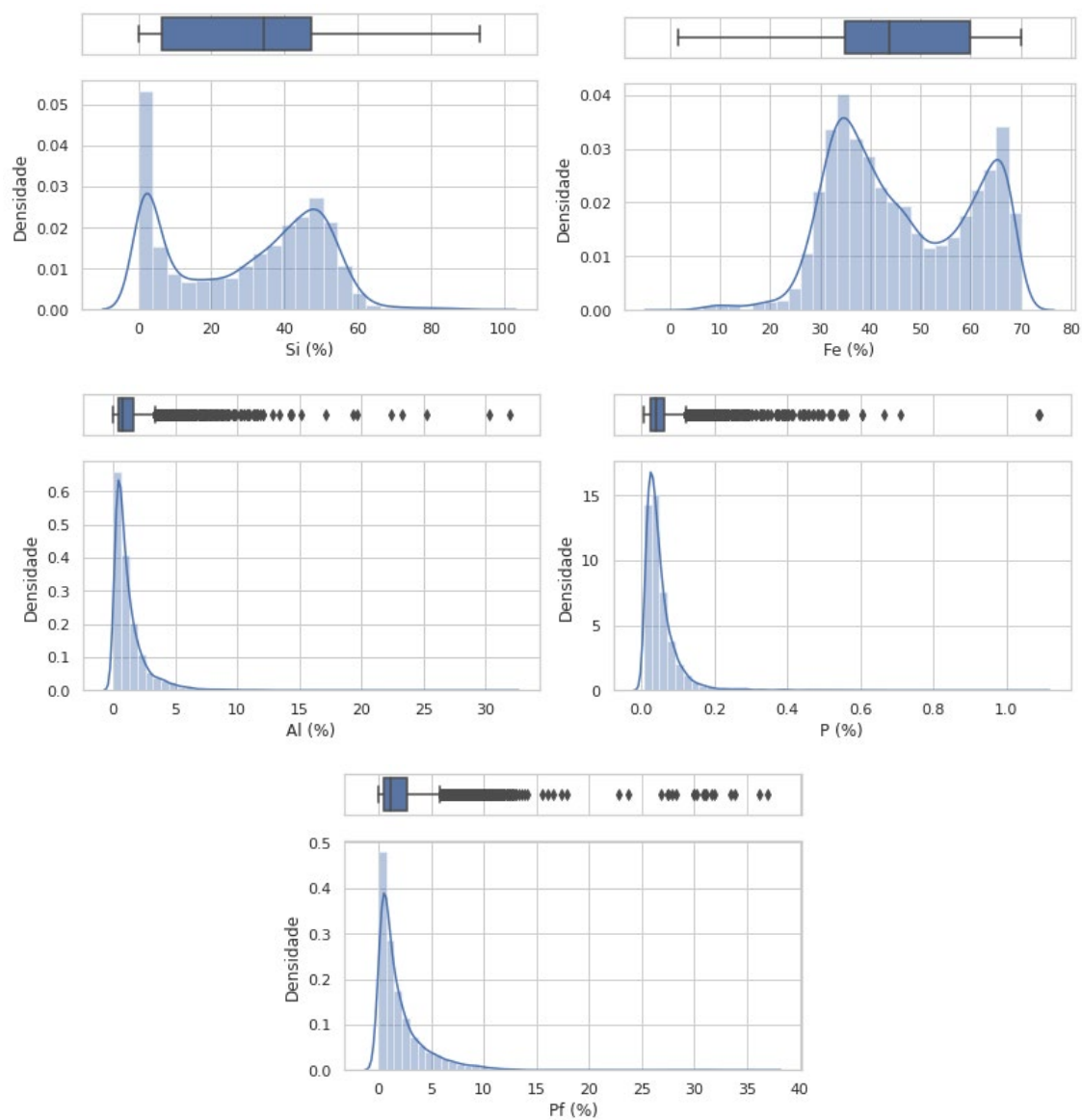
Yuan, C.e Yang, H., 2019. Research on K-value selection method of K-means clustering algorithm: J Multidisciplinary Scientific Journal, v. 2, p. 266-235.

APÊNDICE-I-BOXPLOTS E HISTOGRAMAS POR VARIÁVEL



Apêndice I: Boxplots Teores globais de (a) Fe e Si (b) Al, Pf, F, Fe₃O₄ (c) P, Mn, Ca, Mg e Ti.

APÊNDICE-II- HISTOGRAMAS POR VARIÁVEL



Apêndice II: Histogramas com *Boxplots* para melhor compreensão dos outliers dos Teores globais de Si, Fe, Al, P e Pf.

APÊNDICE-III-ESTATÍSTICA DESCRITIVA POR LITOTIPO

Apêndice III a – Estatística descritiva dos teores globais (%) das amostras para IFR.

	Nº de amostras	Média	Desvio Padrão	Coefficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	540	15,61	4,85	0,3104	5,41	26,71	11,37	15,67	19,50
Fe	540	56,96	3,06	0,05365	50,82	61,99	54,17	56,83	59,68
Pf	540	1,24	0,78	0,6256	0,11	3,47	0,64	1,01	1,60
P	540	0,04	0,03	0,6932	0,01	0,28	0,02	0,03	0,05
Al	540	1,27	0,68	0,5340	0,15	4,65	0,69	1,13	1,77

Apêndice III b – Estatística descritiva dos teores globais (%) das amostras para IMN.

	Nº de amostras	Média	Desvio Padrão	Coefficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	424	34,47	16,13	0,4678	1,80	65,08	20,42	37,77	47,32
Fe	424	38,03	9,59	0,2521	18,69	60,00	30,72	36,75	43,98
Pf	424	4,24	2,52	0,5948	0,47	12,13	2,19	3,80	5,62
P	424	0,08	0,07	0,8473	0,01	0,52	0,04	0,06	0,10
Al	424	1,97	1,77	0,9015	0,09	13,42	0,81	1,43	2,49

Apêndice III c – Estatística descritiva dos teores globais (%) das amostras para IF.

	Nº de amostras	Média	Desvio Padrão	Coefficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	2355	42,94	9,31	0,2169	18,84	69,01	35,73	43,52	50,14
Fe	2355	38,09	6,65	0,1745	19,06	51,95	32,97	37,75	43,41
Pf	2355	1,27	0,81	0,6370	0,00	3,49	0,63	1,07	1,79
P	2355	0,04	0,02	0,5659	0,01	0,18	0,02	0,03	0,05
Al	2355	0,89	0,58	0,6510	0,09	5,11	0,48	0,75	1,16

Apêndice III d – Estatística descritiva dos teores globais (%) das amostras para IC.

	Nº de amostras	Média	Desvio Padrão	Coefficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	2320	43,24	10,77	0,2491	1,36	70,76	37,07	45,55	50,85
Fe	2320	38,63	7,53	0,1950	19,21	61,92	33,28	37,00	43,05
F	2320	1,02	1,03	1010	0,10	6,28	0,52	0,80	1,11
P	2320	0,03	0,02	0,6868	0,01	0,44	0,02	0,03	0,04
Al	2320	0,45	0,37	0,8192	0,02	4,28	0,22	0,35	0,54

Apêndice III e – Estatística descritiva dos teores globais (%) das amostras para IGO

	Nº de amostras	Média	Desvio Padrão	Coefficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	597	31,12	16,14	0,5186	1,00	65,71	17,47	35,21	45,09
Fe	597	42,87	10,10	0,2355	20,11	59,92	34,41	41,04	52,04
Pf	597	5,39	1,87	0,3475	1,81	13,30	3,98	4,87	6,27
P	597	0,10	0,07	0,7531	0,02	1,09	0,06	0,09	0,11
Al	597	1,49	1,24	0,8306	0,11	9,73	0,64	1,20	1,97

Apêndice III f – Estatística descritiva dos teores globais (%) das amostras para IAL.

	Nº de amostras	Média	Desvio Padrão	Coeficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	280	15,52	15,06	0,9706	0,43	61,85	4,78	9,02	19,85
Fe	280	50,30	9,76	0,1940	20,49	59,98	46,07	53,91	57,69
Pf	280	5,90	2,91	0,4931	1,43	13,84	3,54	5,40	7,75
P	280	0,14	0,13	0,9339	0,02	1,09	0,06	0,09	0,15
Al	280	5,45	2,41	0,4430	2,47	19,32	3,78	4,72	6,27

Apêndice III g– Estatística descritiva dos teores globais (%) das amostras para HCT.

	Nº de amostras	Média	Desvio Padrão	Coeficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	544	2,58	1,61	0,6253	0,11	9,06	1,29	2,24	3,70
Fe	544	62,41	1,70	0,02718	54,50	68,97	61,16	62,38	63,64
Pf	544	4,77	2,07	0,4340	0,45	11,36	3,18	4,33	5,99
P	544	0,09	0,06	0,6759	0,01	0,45	0,05	0,07	0,11
Al	544	2,73	1,27	0,4664	0,28	6,39	1,62	2,82	3,65

Apêndice III h – Estatística descritiva dos teores globais (%) das amostras para HF.

	Nº de amostras	Média	Desvio Padrão	Coeficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	761	2,93	2,14	0,7309	0,10	10,01	1,25	2,32	4,07
Fe	761	65,53	1,57	0,02394	61,51	69,28	64,39	65,64	66,77
Pf	761	1,53	0,70	0,4604	0,10	3,45	0,99	1,45	1,99
P	761	0,05	0,03	0,5246	0,01	0,20	0,03	0,04	0,06
Al	761	1,50	0,66	0,4411	0,11	3,07	0,98	1,47	1,94

Apêndice III i – Estatística descritiva dos teores globais (%) das amostras para CG.

	Nº de amostras	Média	Desvio Padrão	Coeficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	183	2,08	3,34	1,603	0,15	21,28	0,75	1,05	1,79
Fe	183	59,93	6,75	0,1127	25,95	68,20	58,10	61,35	64,09
Pf	183	6,26	2,99	0,4775	1,37	17,91	3,95	5,69	8,25
P	183	0,13	0,09	0,6951	0,03	0,55	0,07	0,10	0,15
Al	183	4,97	4,79	0,9623	0,21	31,92	2,17	4,07	5,94

Apêndice III j – Estatística descritiva dos teores globais (%) das amostras para HC.

	Nº de amostras	Média	Desvio Padrão	Coeficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	690	2,17	2,11	0,9725	0,33	12,38	0,89	1,39	2,41
Fe	690	67,00	1,71	0,02552	61,23	69,97	66,20	67,41	68,27
Pf	690	0,84	0,67	0,7976	0,02	3,46	0,37	0,62	1,12
P	690	0,04	0,03	0,7062	0,01	0,21	0,02	0,04	0,05
Al	690	0,84	0,56	0,6638	0,13	2,92	0,42	0,70	1,11

Apêndice III I – Estatística descritiva dos teores globais (%) das amostras para QF.

	Nº de amostras	Média	Desvio Padrão	Coeficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	95	74,43	8,18	0,1100	51,25	93,46	69,26	74,70	82,02
Fe	95	14,10	4,40	0,3124	1,81	21,41	10,21	14,20	18,03
Pf	95	1,91	1,16	0,6060	0,15	6,51	1,07	1,54	2,67
P	95	0,04	0,03	0,6119	0,01	0,16	0,03	0,04	0,06
Al	95	2,55	1,88	0,7350	0,21	14,34	1,39	2,20	3,20

Apêndice III I – Estatística descritiva dos teores globais (%) das amostras para IDO. NA= Não Analisado

	Nº de amostras	Média	Desvio Padrão	Coeficiente de variação	Mínimo	Máximo	Q25%	Mediana	Q75%
Si	22	23,91	6,64	0,2779	14,47	38,80	20,20	22,25	26,71
Fe	22	11,71	4,93	0,4213	5,74	25,14	8,57	10,07	13,40
Pf	22	28,23	5,73	0,2030	15,59	36,93	26,87	29,93	31,66
P	22	0,02	0,01	0,4338	0,01	0,05	0,02	0,02	0,02
Al	22	0,87	0,53	0,6089	0,10	2,41	0,52	0,72	1,31

APÊNDICE-IV-ESTATÍSTICA DESCRITIVA POR DOMÍNIO

Apêndice IV a – Estatística descritiva dos teores globais das amostras agrupadas como minério de alto teor. Número de amostras: 2058

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
				Padrão			
<i>Al</i>	0,105	6,572	1,674	1,183	0,764	1,372	2,319
Fe	54,498	69,970	65,081	2,460	63,200	65,373	67,087
<i>PF</i>	0,025	11,364	2,306	2,076	0,753	1,637	3,145
<i>P</i>	0,010	0,447	0,061	0,046	0,031	0,048	0,075
<i>Si</i>	0,100	12,382	2,511	2,013	1,014	1,868	3,492

Apêndice IV b – Estatística descritiva dos teores globais das amostras agrupadas como baixo teor de minério de Fe. Número de amostras: 1697

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
				Padrão			
<i>Al</i>	0,100	9,728	1,829	1,660	0,667	1,242	2,285
Fe	45,008	61,994	53,554	4,925	49,139	53,990	57,940
<i>PF</i>	0,001	13,838	2,844	2,961	0,664	1,387	4,470
<i>P</i>	0,008	1,092	0,070	0,082	0,025	0,043	0,081
<i>Si</i>	0,152	35,378	17,924	9,167	10,447	17,879	25,410

Apêndice IV c – Estatística descritiva dos teores globais das amostras agrupadas como rejeito. Número de amostras: 4926

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
				Padrão			
<i>Al</i>	0,020	31,919	0,956	1,607	0,307	0,555	1,034
Fe	1,806	59,035	35,930	6,829	32,172	35,741	40,359
<i>PF</i>	0,001	36,931	1,714	2,554	0,385	0,958	2,201
<i>P</i>	0,006	1,086	0,043	0,038	0,022	0,034	0,053
<i>Si</i>	0,447	93,456	45,070	10,155	39,261	45,806	51,196

Apêndice IV d – Estatística descritiva dos teores globais das amostras agrupadas como domino rejeito pelo algoritmo KM.

Número de amostras: 5125

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
Padrão							
<i>Al</i>	0,020	7,530	0,766	0,683	0,310	0,550	0,980
Fe	1,810	52,270	36,671	6,877	32,490	36,240	41,430
<i>PF</i>	0,000	8,010	1,390	1,367	0,380	0,880	1,970
<i>P</i>	0,010	0,220	0,040	0,026	0,020	0,030	0,050
<i>Si</i>	24,700	93,460	44,813	9,483	38,190	45,340	50,950

Apêndice IV e – Estatística descritiva dos teores globais das amostras agrupadas como domino rico pelo algoritmo KM Número de amostras: 2834

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
Padrão							
<i>Al</i>	0,100	6,030	1,472	1,020	0,690	1,230	1,998
Fe	42,360	69,970	61,990	5,594	58,453	63,685	66,430
<i>PF</i>	0,020	8,320	1,815	1,519	0,670	1,350	2,498
<i>P</i>	0,010	0,210	0,049	0,030	0,030	0,040	0,060
<i>Si</i>	0,100	27,620	7,545	7,745	1,420	3,770	12,918

Apêndice IV f – Estatística descritiva dos teores globais das amostras agrupadas como domino pobre pelo algoritmo KM. Número de amostras: 716

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
Padrão							
<i>Al</i>	0,100	31,920	4,405	3,309	2,408	3,930	5,593
Fe	5,740	65,670	52,895	11,343	49,450	56,985	60,115
<i>PF</i>	0,680	36,930	8,012	4,373	5,660	7,195	9,135
<i>P</i>	0,010	1,090	0,155	0,116	0,090	0,120	0,180
<i>Si</i>	0,150	56,900	8,978	10,343	1,810	4,390	13,123

Apêndice IV g – Estatística descritiva dos teores globais das amostras agrupadas como domino pobre pelo algoritmo HC. Número de amostras: 1256

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
				Padrão			
<i>Al</i>	0,180	31,920	3,655	2,752	1,960	3,210	4,513
Fe	5,740	67,860	57,585	9,326	55,988	60,340	62,838
<i>PF</i>	0,300	36,930	5,981	4,026	3,538	5,165	7,503
<i>P</i>	0,010	1,090	0,120	0,099	0,060	0,090	0,140
<i>Si</i>	0,100	56,900	6,196	7,857	1,500	3,430	7,315

Apêndice IV h – Estatística descritiva dos teores globais das amostras agrupadas como domino de rejeito pelo algoritmo HC. Número de amostras: 5400

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
				Padrão			
<i>Al</i>	0,020	7,530	0,795	0,718	0,320	0,570	1,030
Fe	1,810	64,270	37,348	7,444	32,650	36,685	42,330
<i>PF</i>	0,000	17,320	1,510	1,580	0,400	0,920	2,100
<i>P</i>	0,010	0,180	0,041	0,027	0,020	0,030	0,050
<i>Si</i>	1,630	93,460	43,633	10,618	36,598	44,700	50,653

Apêndice IV i – Estatística descritiva dos teores globais das amostras agrupadas como domino rico pelo algoritmo HC. Número de amostras: 2019

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
				Padrão			
<i>Al</i>	0,100	3,660	1,173	0,686	0,595	1,040	1,650
Fe	45,810	69,970	63,144	5,127	59,710	65,180	67,095
<i>PF</i>	0,020	4,440	1,158	0,779	0,540	0,980	1,600
<i>P</i>	0,010	0,130	0,040	0,021	0,020	0,040	0,050
<i>Si</i>	0,100	31,300	6,971	7,212	1,360	3,370	11,480

Apêndice IV g – Estatística descritiva dos teores globais das amostras agrupadas como domino pobre pelo algoritmo GHC. Número de amostras: 1256

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
Padrão							
<i>Al</i>	0,020	8,410	0,840	0,890	0,290	0,530	1,050
Fe	1,810	56,430	34,260	6,480	31,150	34,410	38,040
<i>PF</i>	0,000	12,130	1,780	1,690	0,460	1,300	2,510
<i>P</i>	0,010	0,180	0,050	0,030	0,030	0,040	0,060
<i>Si</i>	2,940	93,460	47,510	9,690	42,310	48,080	52,490

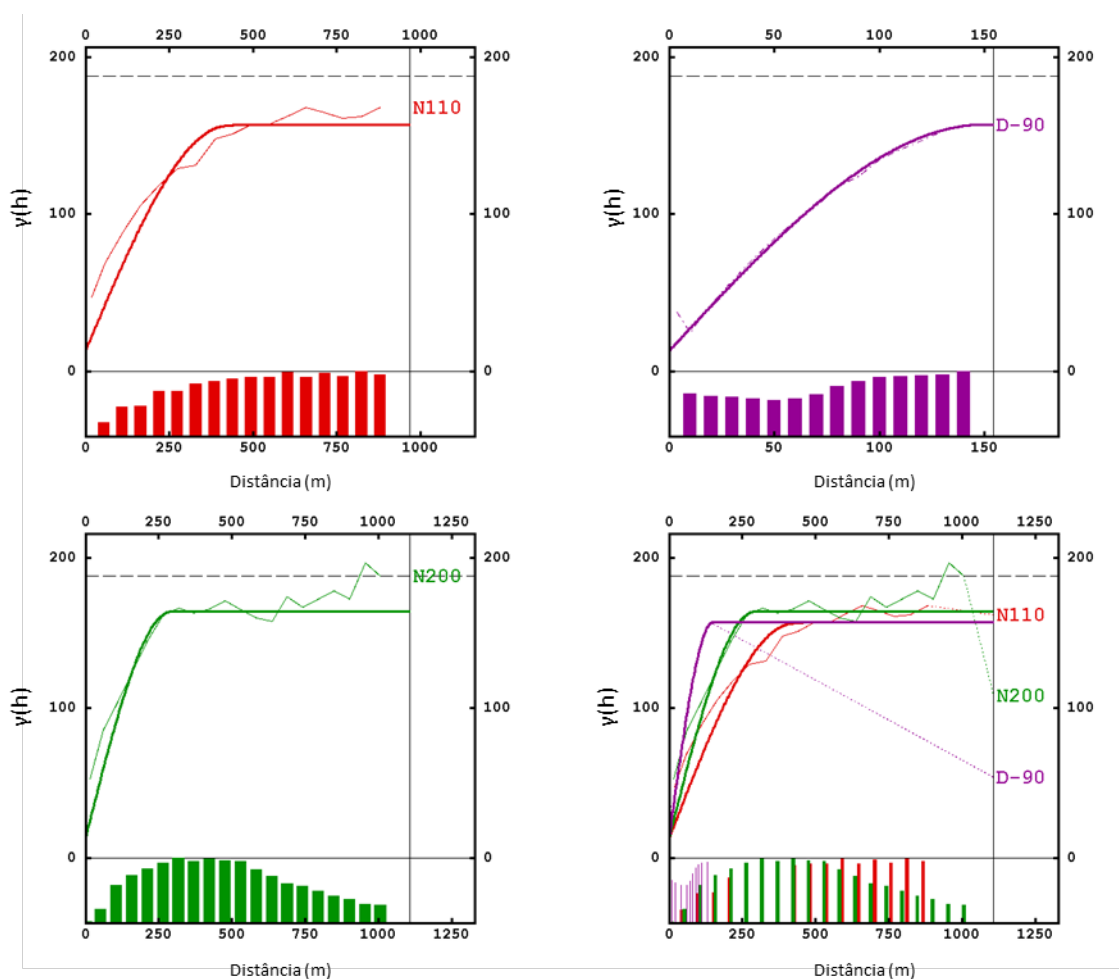
Apêndice IV h – Estatística descritiva dos teores globais das amostras agrupadas como domino de rejeito pelo algoritmo GHC. Número de amostras: 5400

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
Padrão							
<i>Al</i>	0,100	31,920	1,810	2,130	0,590	1,170	2,200
Fe	5,740	68,920	50,050	12,620	39,430	50,910	61,850
<i>PF</i>	0,000	36,930	2,330	3,300	0,550	1,170	2,720
<i>P</i>	0,010	1,090	0,060	0,070	0,020	0,040	0,070
<i>Si</i>	0,130	70,760	23,530	19,110	3,840	21,970	40,750

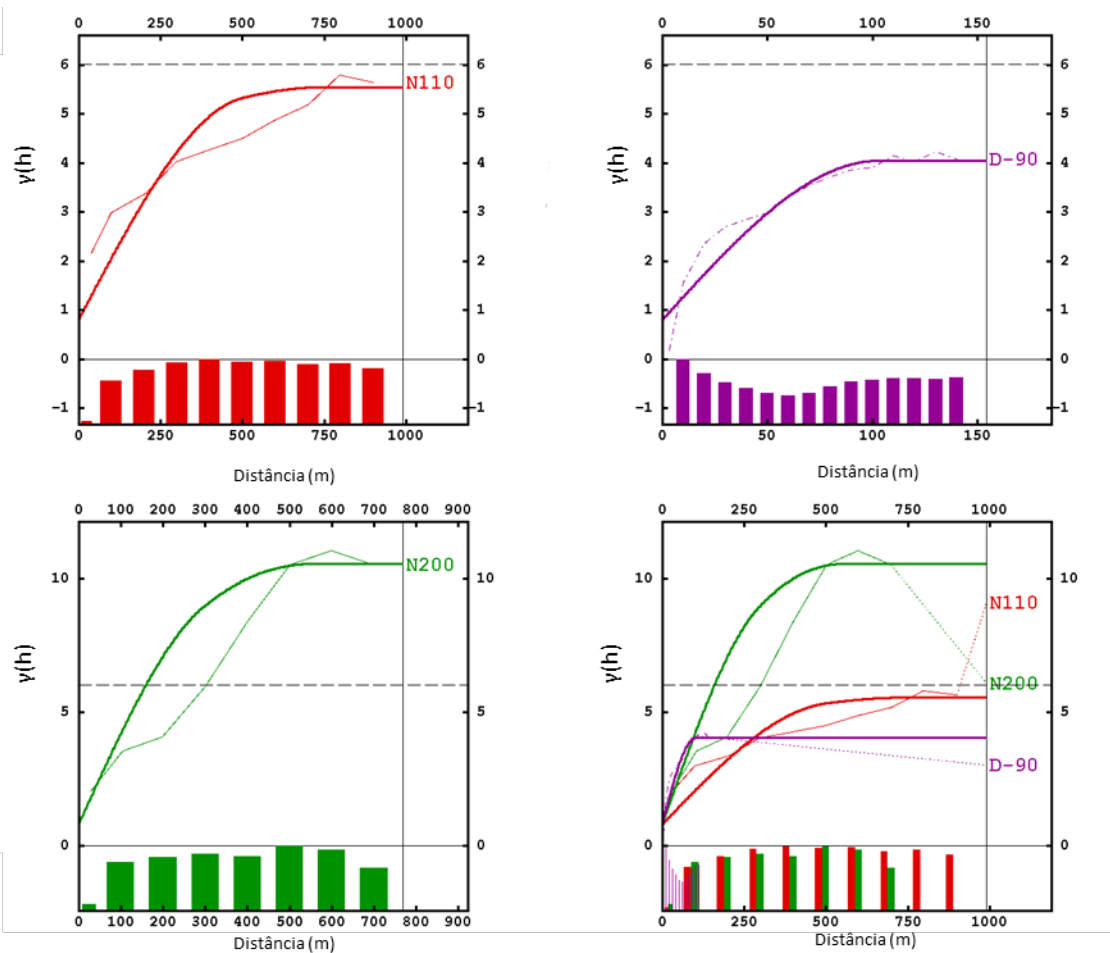
Apêndice IV i – Estatística descritiva dos teores globais das amostras agrupadas como domino rico pelo algoritmo GHC. Número de amostras: 2019

Parâmetro	Mínimo	Máximo	Média	Desvio	Q25	Q50	Q75
Padrão							
<i>Al</i>	0,050	11,340	1,140	1,080	0,440	0,800	1,460
Fe	21,790	69,970	54,200	11,970	44,320	56,680	65,650
<i>PF</i>	0,010	12,690	2,060	2,270	0,480	1,130	2,800
<i>P</i>	0,010	0,450	0,050	0,040	0,030	0,040	0,060
<i>Si</i>	0,100	62,530	18,620	17,380	2,060	12,850	33,960

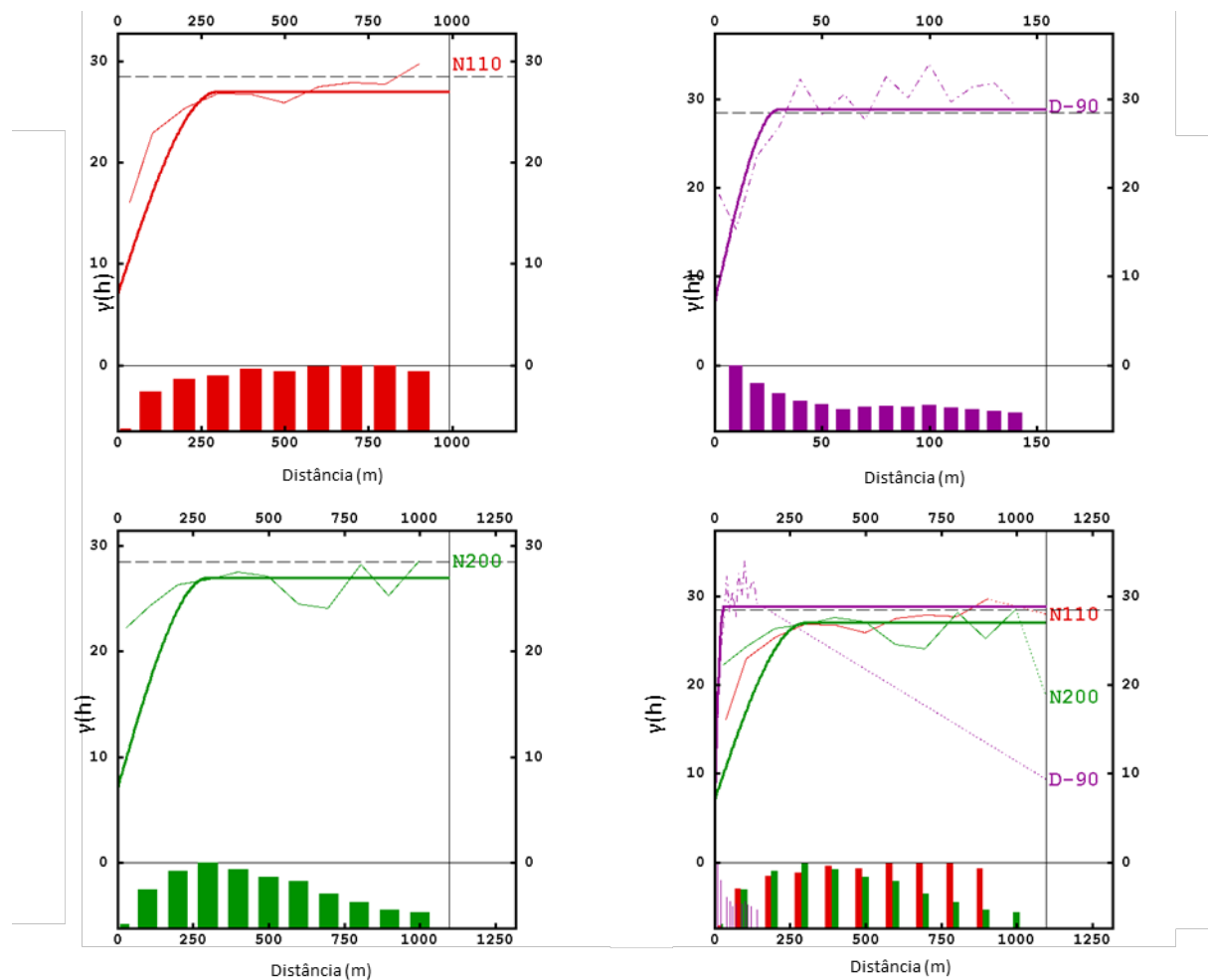
APÊNDICE-V-AJUSTE DOS MODELOS TEÓRICOS DE VARIOGRAMA



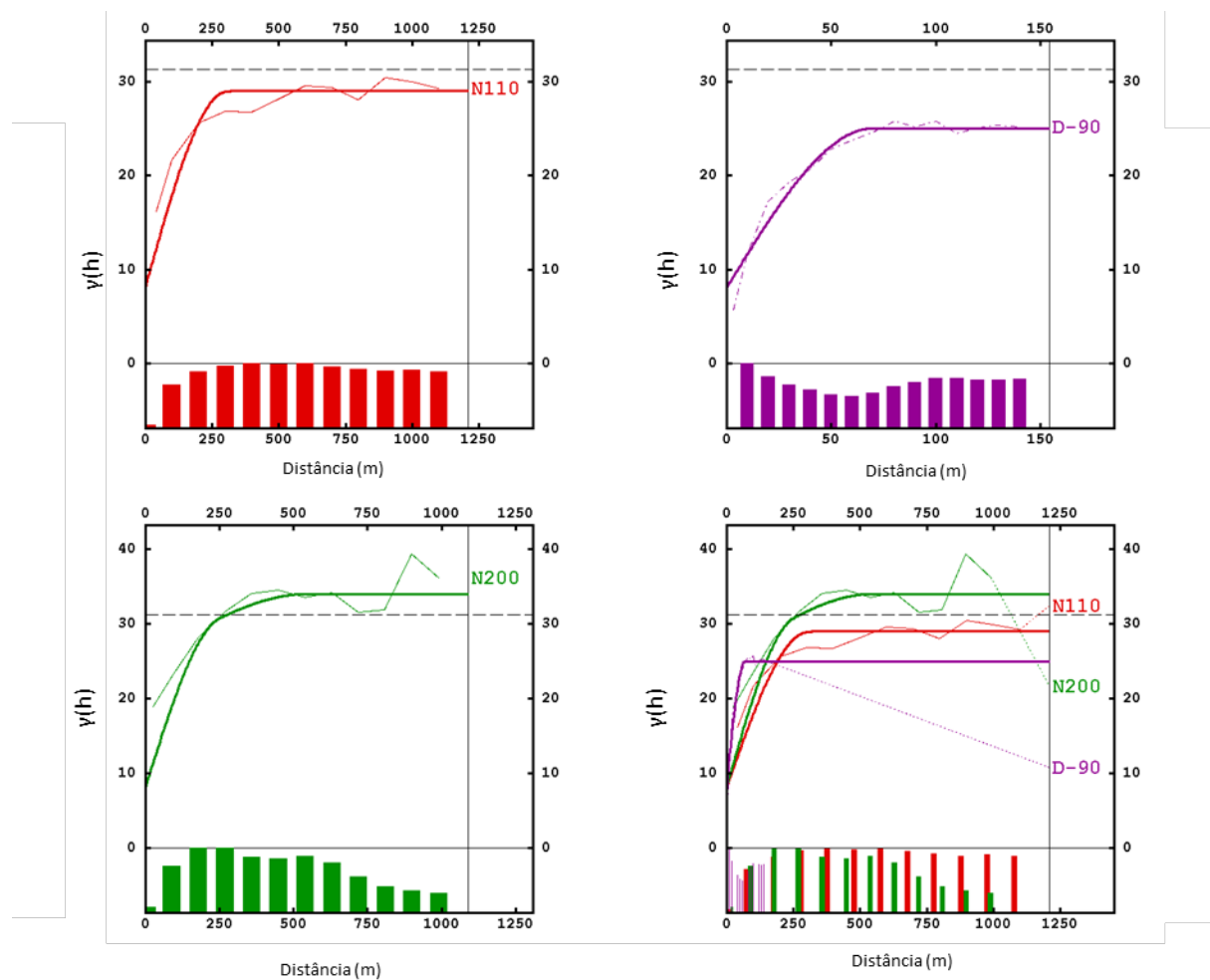
Apêndice Va) Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio Único, respectivamente N110°, *Downhole*, N200° e sobreposição das 3 direções.



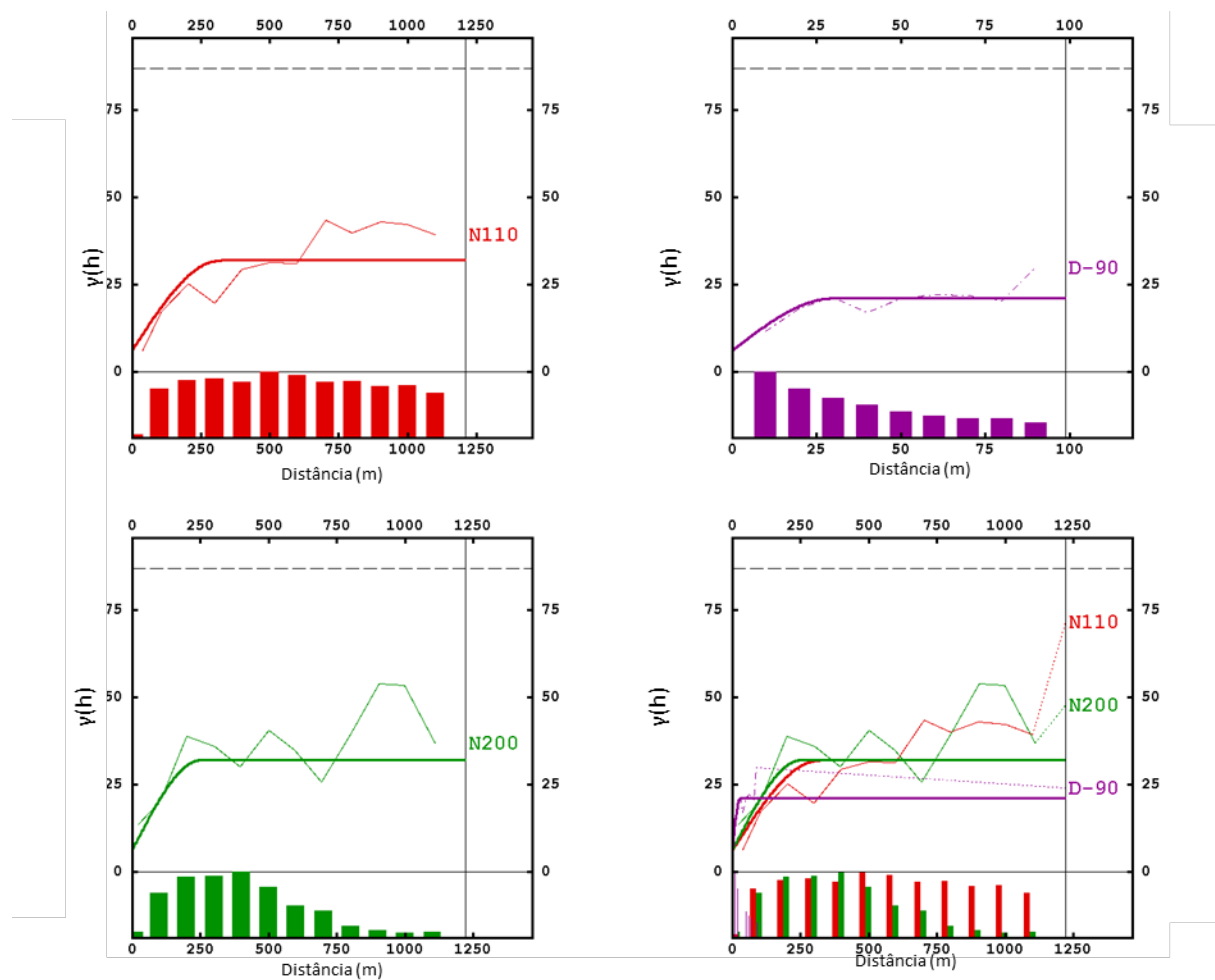
Apêndice Vb) Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio manual Rico, respectivamente N110°, *Downhole*, N200° e sobreposição das 3 direções.



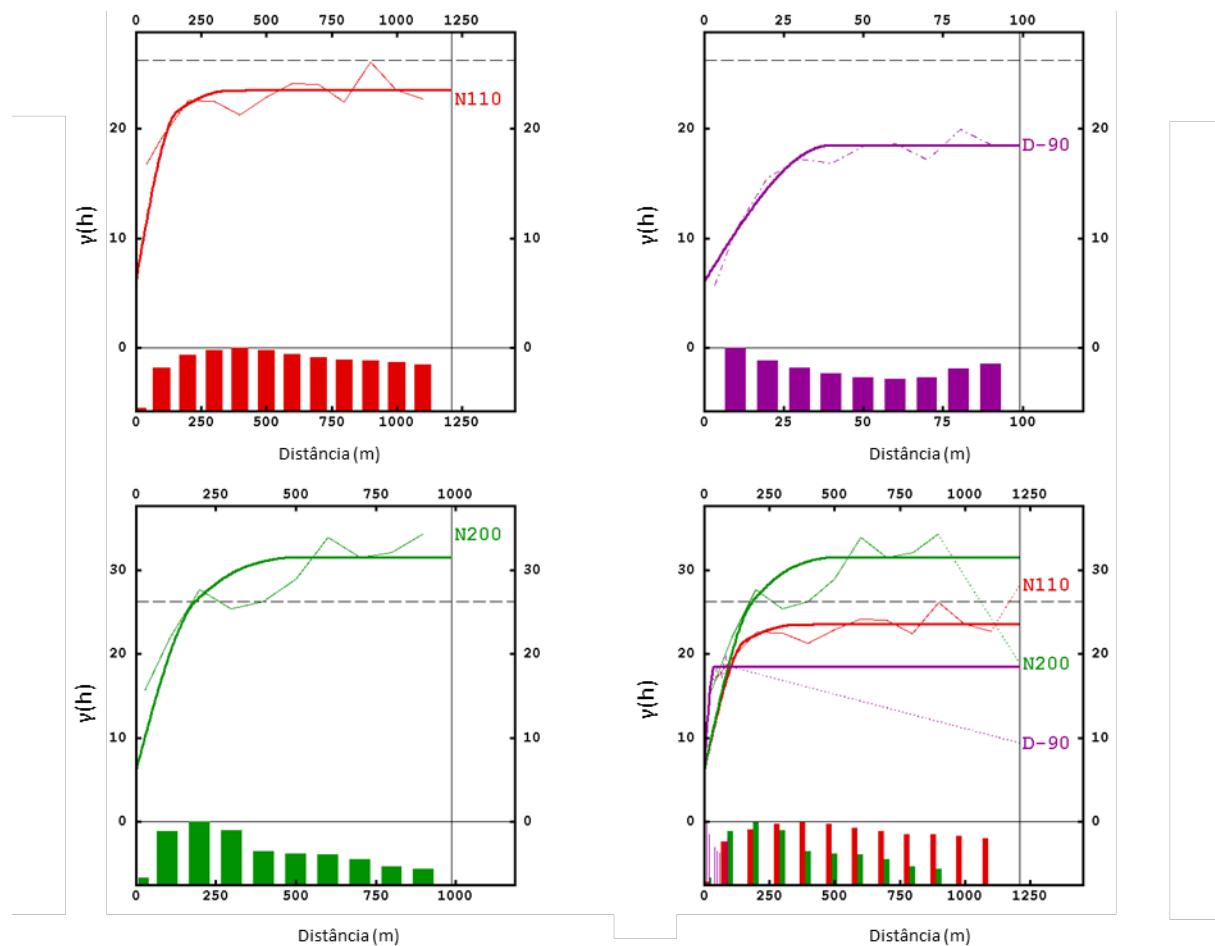
Apêndice Vc) Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio manual Pobre, respectivamente N110°, *Downhole*, N200° e sobreposição das 3 direções.



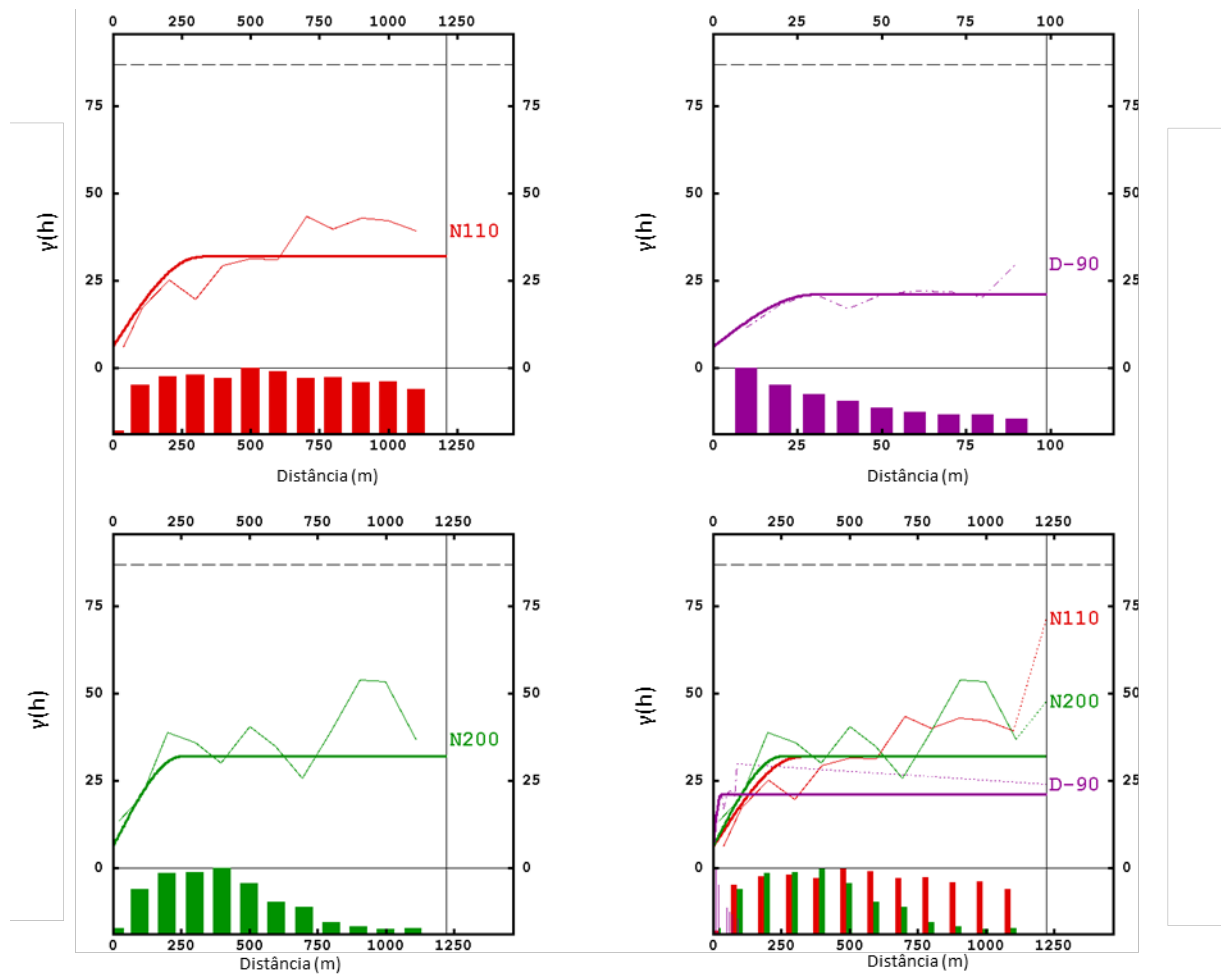
Apêndice Vd) Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio KC Rico, respectivamente N110°, *Downhole*, N200° e sobreposição das 3 direções.



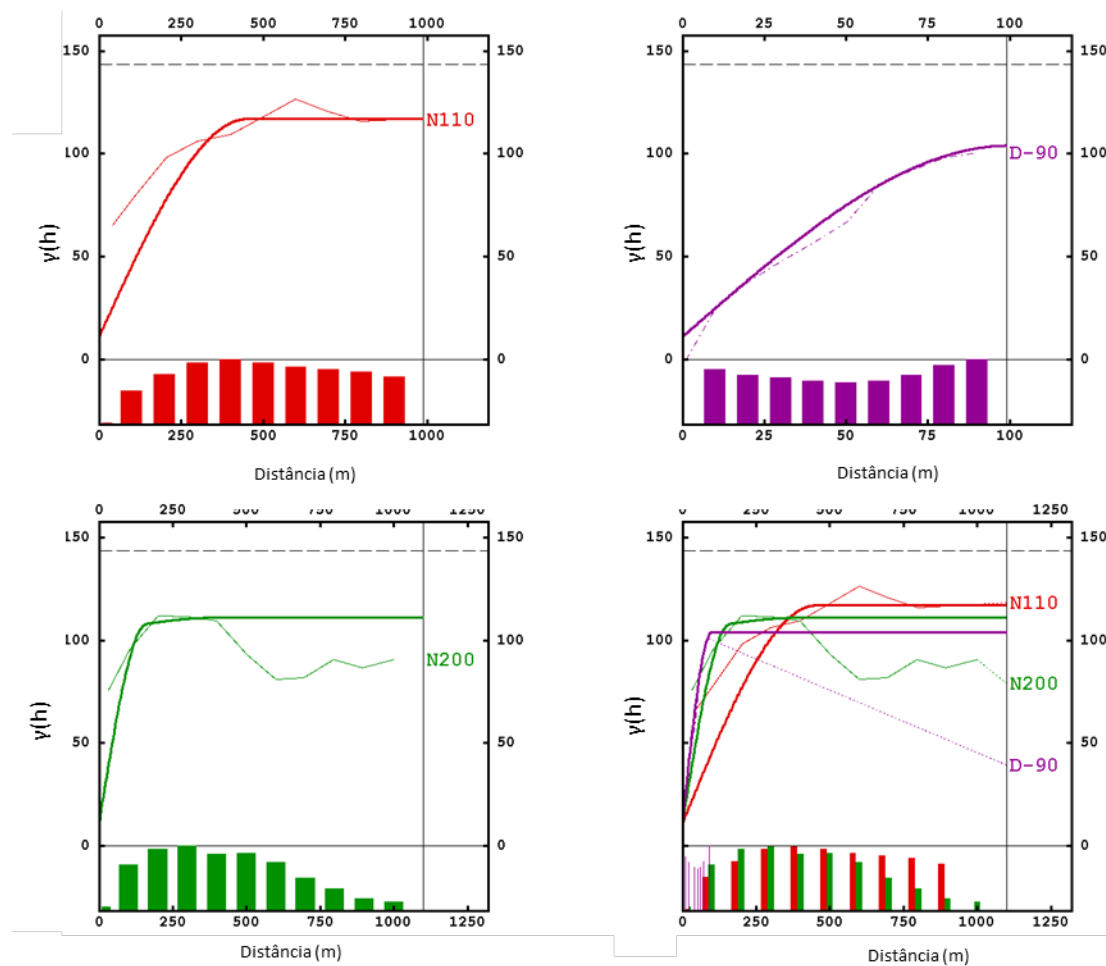
Apêndice Ve) Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio KC Pobre, respectivamente N110°, *Downhole*, N200° e sobreposição das 3 direções.



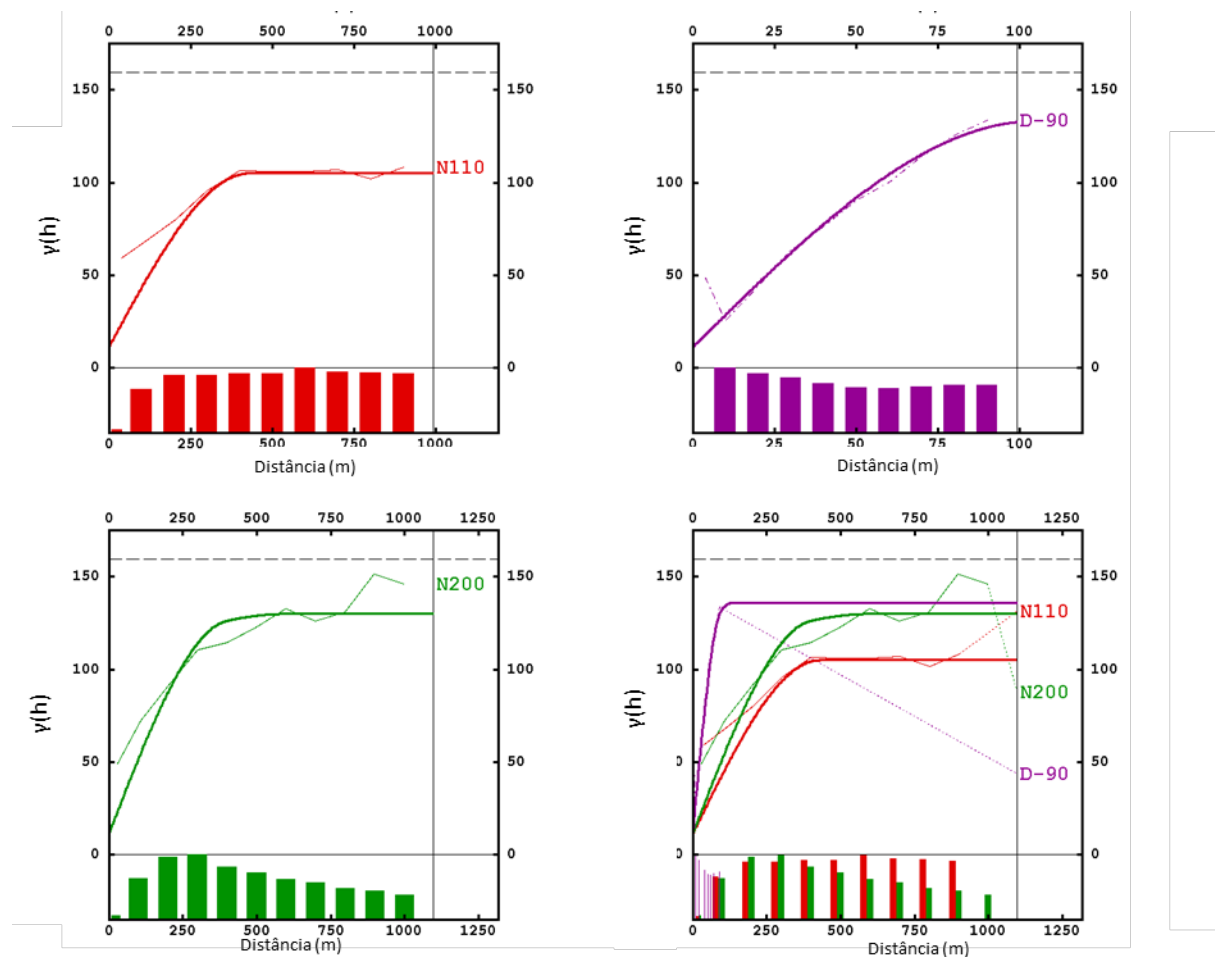
Apêndice Vf) Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio HC Rico, respectivamente N110°, *Downhole*, N200° e sobreposição das 3 direções.



Apêndice Vg) Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio HC Pobre, respectivamente N110°, *Downhole*, N200° e sobreposição das 3 direções.

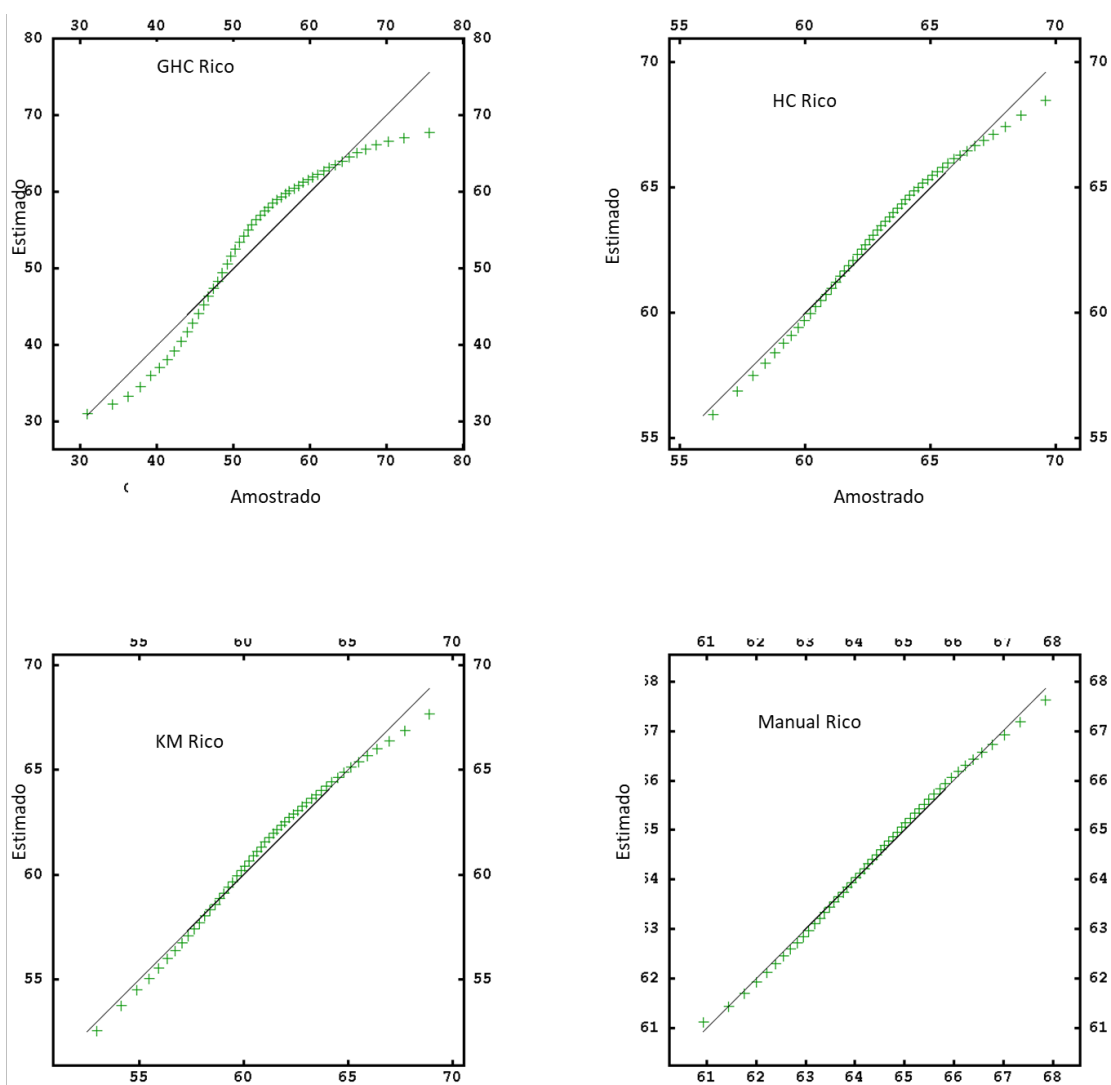


Apêndice Vh) Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio GHC Rico, respectivamente N110°, *Downhole*, N200° e sobreposição das 3 direções.

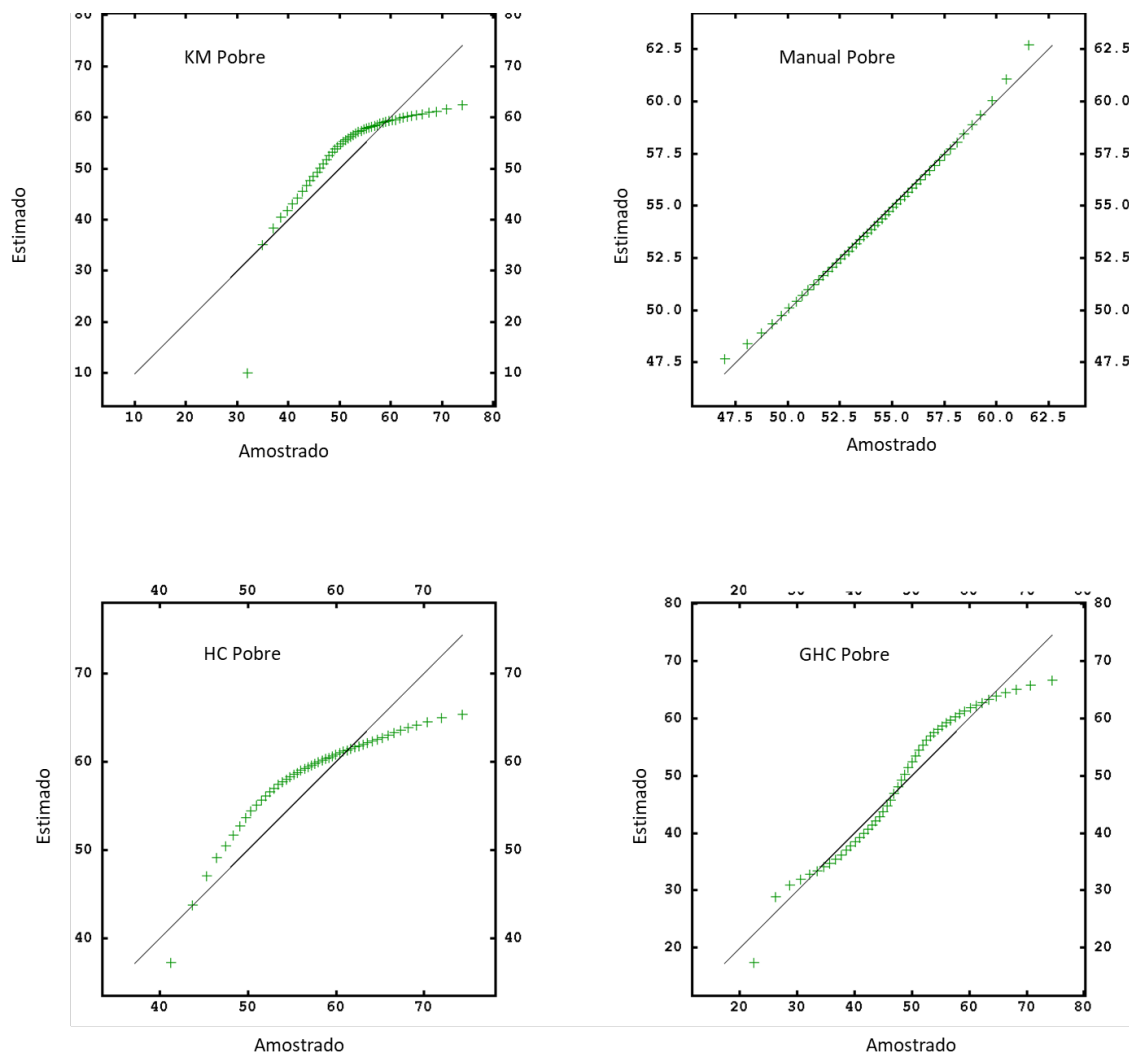


Apêndice Vi) Variogramas experimentais, com histograma de pares de pontos e respectivos modelos teóricos ajustados do ferro no cenário de domínio GHC Pobre, respectivamente N110°, *Downhole*, N200° e sobreposição das 3 direções.

APÊNDICE-VI-GRÁFICOS QUANTIL-QUANTIL



Apêndice VIa: Gráficos quantil-quantil entre os teores de ferro do conjunto amostral e as estimativas dos cenários ricos



Apêndice VIb: Gráficos quantil-quantil entre os teores de ferro do conjunto amostral e as estimativas dos cenários pobres