

Número: 295



UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE GEOCIÊNCIAS

Pós-Graduação em Geociências

Área de Administração e Política de Recursos Minerais.

AUTOR: Moacir Américo Cornetti

**O impacto do uso de ponderadores de dados agrupados na
Geoestatística**

Dissertação apresentada ao Instituto de Geociências como parte dos requisitos para obtenção do título de Mestre em Geociências.

Este exemplar corresponde à redação final da tese defendida por Moacir Américo Cornetti e aprovada pelo Conselho Julgador em 21/08/2003

Orientador: Prof. Dr. Armando Zaupa Remacre

A handwritten signature in black ink, appearing to read "Armando Zaupa Remacre", written over a horizontal line. Below the line, the word "ORIENTADOR" is printed in capital letters.

CAMPINAS- SÃO PAULO

Agosto-2003

I

UNICAMP
BIBLIOTECA CENTRAL

UNICAMP
BIBLIOTECA CENTRAL
SEÇÃO CIRCULANTE

200331954

UNIDADE	BC
Nº CHAMADA	UNICAMP
	CBIS
V	EX
TOMBO BCI	56200
PROC.	16-124/03
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 14,00
DATA	10/20/03
Nº CPD	

FICHA CATALOGRÁFICA ELABORADA PELA

BIBLIOTECA DO IG - UNICAMP

CM00190378-9

Sib id 303185

Cornetti, Moacir Americo

O impacto do uso de ponderadores de dados agrupados na geoestatística /
Moacir Americo Cornetti.- Campinas,SP.: [s.n.], 2003.

Orientador: Armando Zaupa Remacre

**Dissertação (mestrado) Universidade Estadual de Campinas,
Instituto de Geociências.**

1. Geoestatística. I. Remacre, Armando Zaupa. II. Universidade
Estadual de Campinas, Instituto de Geociências. III. Título.

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE GEOCIÊNCIAS

Pós-Graduação em Geociências

Área de Administração e Política de Recursos Minerais.



UNICAMP

AUTOR: Moacir Américo Cornetti

ORIENTADOR: Prof. Dr. Armando Zaupa Remacre

Aprovada em: 21/08/03

EXAMINADORES:

Prof. Dr. Armando Zaupa Remacre -

Presidente

Prof. Dra. Elisabete Maria Pascholati

Prof. Dr. José Ricardo Sturaro

Campinas, 21 de agosto de 2003

Agradecimentos

Agradeço a minha família, por entender e compreender esta fase da minha vida que foi a confecção deste trabalho, roubando horas que poderíamos estar juntos.

Ao Paulo e Ricardo, da equipe de informática do Instituto de geociências, que compreenderam, as minhas “desaparecidas” e se esforçaram para que o Instituto não sentisse a minha falta.

Ao Prof Armando pela força e incentivo para eu iniciar e terminar este trabalho, ao Sérgio Sacani, pela ajuda e força, no desenvolvimento desta dissertação. A prof Elizabete e ao prof Rodrigo, pela ajuda na correção e em especial nos seus comentários.

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE GEOCIÊNCIAS

Pós-Graduação em Geociências

Área de Administração e Política de Recursos Minerais.



Resumo

Nesta dissertação são apresentadas algumas técnicas de como podemos, a partir de dados agrupados, calcular um conjunto de ponderadores associado a estes dados, que quando aplicados ao conjunto de amostras, possam nos indicar valores estatísticos mais próximos dos verdadeiros valores existentes na população, esta operação que em inglês é denominada de declustering, aqui será denominada de desagrupamento.

São abordados quatro métodos que se propõe a produzir estes ponderadores, e analisados os prós e contras de cada método. Estes métodos são Polígono de Influência, *Cell declustering*, Krigagem da média e Método de Richmod.

Aspectos de como devem ser modificados alguns parâmetros estatísticos de modo que os valores ponderados possam ser utilizados nestes cálculos, também são abordados.

São efetuados estudos de casos em sete populações conhecidas, para se analisar a influência dos ponderadores.

As populações que farão parte deste estudo são as seguintes: *declus.dat*, conjunto de dados que foi utilizado para o desenvolvimento do método *cell declustering*; o conjunto de dados utilizado em Isaaks (1989), denominado Walker Lake, e cinco populações simuladas especialmente para o desenvolvimento deste trabalho, utilizando simulação seqüencial gaussiana.

Após a obtenção dos variogramas ponderados, fez-se um estudo do impacto causado na krigagem e na simulação gaussiana, onde se apresenta um estudo de caso.

Após o trabalho podemos tirar conclusões sobre cada método, sendo que a princípio podemos abandonar os métodos do *cell declustering* e o polígono de influência, pois seus resultados não foram consistentes para a maioria das populações estudadas. O método de Richmod merece um estudo, de modo a diminuir seu poder de desagrupamento, sendo que o método que apresentou uma maior confiabilidade é o da krigagem da média.

UNIVERSIDADE ESTADUAL DE CAMPINAS



INSTITUTO DE GEOCIÊNCIAS

Pós-Graduação em Geociências

Área de Administração e Política de Recursos Minerais.

Abstract

This dissertation presents some techniques on calculating a weighted up set, from clustered data, that are associated to them. These data, when applied to the sample set, can indicate statistical values that are closer to the true ones than the previous. This operation is called declustering.

There are four described methods that intend to provide this weighted up set, through the analysis of the advantages and disadvantages of each method. These methods are: Area-of-Influence Polygon, Cell Declustering, Kriging of the Mean and Richmod Method.

The ways that some statistical parameters should be modified so that the considered values may be used in these calculations, are also described.

In order to analyze the influence of the weighted up set, there are seven known populations being studied.

The populations taking part in this study are the following: declus.dat, data set used for the development of the cell declustering method development; the data set used in Isaaks (1989), nominated Walker Lake, and five populations specifically simulated for this work, by using sequential Gaussian simulation.

After obtaining the weighted variogram, a study about the impact in the Kriging and the Gaussian simulation where was done at the location of the studied event.

From each method there are made conclusions, which lead to abandoning the Cell Declustering Method and the Influence Polygon, because their results are not consistent for the most studied populations. The Richmod Method deserves to be studied in order that his declustering power can be reduced.

The most reliable method presented in this dissertation is the Kriging of the Mean.

Índice

Resumo	V
Abstract	VI
Índice.....	VII
Lista de Figuras	IX
Lista de Tabelas	XIV
<hr/>	
Capítulo 1	1
Introdução.....	1
1.1 Exemplo.....	1
1.2 Métodos Estudados	2
1.3 Aplicação e exemplos.....	4
1.4 Organização da Dissertação.....	5
Capítulo 2	6
Revisão Metodológica.....	6
2.1 Polígono de Influência.....	6
2.2 Método – Cell Declustering (Declus).....	11
2.3 Método Krigagem da Média (KM)	15
2.3.1 Caso particular: Efeito de Pepita Puro.....	17

2.3.2 Influência dos parâmetros da variografia.....	19
2.4 Método Richmod.....	30
Capítulo 3	36
Impacto dos ponderadores nas estatísticas	36
Capítulo 4	45
Estudos de caso.....	45
4.1 Apresentação das populações	45
4.2 Estudo Sobre o Variograma.....	59
4.3 A simulação Gaussiana Condicional.....	65
Capítulo 5	74
Conclusões.....	74
Anexos	76
Anexo 1	76
Anexo 2	78
Referencias Bibliográficas.	83

Lista de Figuras

1.1 - Diagrama mostrando um exemplo de re-amostragem agrupada	2
2.1 - Exemplo de polígono de influência em dados não agrupados.	6
2.2 - Construção de uma área de influência de um ponto a partir de seus vizinhos. As linhas azuis indicam os segmento de reta perpendicular, colocado no ponto mediano das retas que une o ponto e seu vizinho, o polígono vermelho o polígono de influência formado.	8
2.3 - Exemplo de polígono de influência em dados agrupados. Os dados representados por cruces	9
2.4 - Exemplo de área de influência com agrupamentos distantes entre si.	10
2.5 - Comportamento dos ponderadores para cada amostra, para três modelos de variogramas, sem efeito de pepita, patamar igual a um.	20
2.6 - Comportamento dos ponderadores, com patamar igual a um, para três modelos de variogramas, e efeito de pepita igual a 0.15.	21
2.7 - Mostra os valores dos ponderadores, quando variamos a amplitude de 1,2, 3 e 4, sem modificar patamar, modelo e efeito de pepita.	21
2.8 - Representação de uma superfície e 14 pontos amostrados, com presença de agrupamentos em regiões que com altos valores para a altitude.	22
2.9 - Mapa base para as 14 amostras, cujos valores são cotas.	24
2.10 - Curvas da variação da média para os dados Cluster.dat, com as curvas representado efeito de pepita variando de zero a um. Para modelo esférico, patamar um, e o passo variando de 0 a 30.	25

2.11 - Curvas da variação da média para os dados Cluster.dat, sendo que cada curva representa um efeito de pepita que varia de zero a um. Com modelo gaussiano, patamar um, e o passo variando de 0 a 30	25
2.12 - Usando modelo exponencial, com curvas representando efeito de pepita de 0 a um, patamar um, e passo de 0 a 30,sobre o conjuntos de dados cluster.dat.	26
2.13 - Usando como dado os 14 pontos, curvas da variação da média, sendo que cada curva representa um efeito de pepita que varia de zero a um. Com modelo esférico, patamar um, e o passo variando de 0.5 a 4.3.	27
2.14 - Curvas da variação da média, cada curva representa um efeito de pepita que varia de zero a um. Com modelo exponencial, patamar um, e o passo variando de 0.5 a 4.3.....	27
2.15 - Representação da variabilidade da média, com passo de 0.5 a 4.3, patamar 1, em um modelo gaussiano, e cada curva representa efeito de pepita de 0 a 1.	27
2.16 - Variação dos pesos para diversos efeitos de pepita, utilizando modelo exponencial, patamar um, amplitude 3.5.	29
2.17 - Variação dos pesos para diversos efeitos de pepita que varia de zero a um, utilizando modelo esférico , patamar um, amplitude 3.	30
2.18 - Aplicação do método Declus sobre dados agrupados.	34
2.19 - Aplicação do método Richmod sobre dados agrupados.	34
2.20 - Aplicação do método Declus sobre dados agrupados, criados para comparação dos métodos Declus e Richmod.	34
2.21 - Aplicação do método Richmod sobre dados agrupados, criados para comparação dos métodos Declus e Richmod.	34
2.22 - Aplicação do método Declus sobre dados agrupados, criados para comparação dos métodos Declus e Richmod.	35

2.23 - Aplicação do método Richmod sobre dados agrupados, criados para comparação dos métodos Declus e Richmod.	35
3.1 - Figura representando o histograma acumulado crescente, com e sem ponderador, dados da tabela 3.2.	40
3.2 - Usando dados da Tabela 3.2, representação da curva quantidade de metal, para diversos teores de corte, em histogramas com e sem ponderador.	41
3.3 - A figura mostra a relação entre teor de corte e benefício para histogramas que usam de ponderador e sem ponderador, dados da tabela 3.2.	41
.....	
4.1 - Histograma da população de 2500, do arquivo de dados true.dat.	47
4.2 - Histograma da população de 140 dados de nome cluster.dat	47
4.3 - Histograma da população de Walker lake com 78000	47
4.4 - Histograma da de Walker lake, com 470 dados.	47
4.5 - Histograma sobre a população de 4000 dados.	48
4.6 - Histograma sobre a população de 228 dados.....	48
4.7 - Histograma sobre o conjunto de dados da população de 4000 pontos relativos a imagem 2.	48
.....	
4.8 - Histograma sobre o conjunto de 229 dados, retirados da população da imagem 2.	48
4.9 - Histograma da população da imagem 3 de 4000 dados.	49
4.10 - Histograma do conjunto de dados amostrados da população da imagem 3, 221 dados	49
4.11 - Histograma da população da imagem 3 de 4000 dados.	49
4.12 - Histograma do conjunto de dados amostrados da população da imagem 4, 218 dados	49

4.13 - Histograma sobre a população de 4000 dados.	50
4.14 - Histograma sobre a população de 221 dados.	50
4.15 - Histogramas referente ao conjunto de dados imagem1, na parte superior histograma normal, parte inferior histograma acumulado decrescente	51
4.16 - Histogramas referente ao conjunto de dados imagem2, na parte superior histograma normal, parte inferior histograma acumulado decrescente	52
4.17 - Histogramas referente ao conjunto de dados imagem 3, na parte superior histograma normal, parte inferior histograma acumulado decrescente	53
4.18 - Histogramas referente ao conjunto de dados imagem 4, na parte superior histograma normal, parte inferior histograma acumulado decrescente	54
4.19 - Histogramas referente ao conjunto de dados imagem 5, na parte superior histograma normal, parte inferior histograma acumulado decrescente	55
4.20 - Histogramas referente ao conjunto de dados cluster.dat, na parte superior histograma normal, parte inferior histograma acumulado decrescente	56
4.21- mostra o variograma dos dados da imagem 1 na direção azimute 90°, A curva marrom é o variograma sobre a imagem original 0°.	62
4.28 - A figura mostra os variogramas experimentais, relativos aos dados da imagem simulada 5 na direção azimute 90°.	62
4.29 - A figura mostra os variogramas experimentais, relativos aos dados da imagem simulada 5 na direção azimute 0°.	62
4.30 - A figura compara os resultados de cinco validações cruzadas, sobre o conjunto de dados cluster.dat e os dados originais, sendo que os variogramas foram calculados em quatro casos usando ponderador e um sem uso de ponderador.	63

4.31 - A figura mostra um pequeno espaço da Figura 4.30, mostrando onde há maior diferença entre os cálculos.	63
4.32 - Figura mostrando o resultado de quatro validações cruzadas nos 470 dados de Walker Lake, sendo que os variogramas utilizados foram ponderados pelos métodos aqui estudados.	64
4.33 - Médias das simulações por modelo, lembrando que a média da população real é de 278.0, na figura cada curva representa um ajuste de variograma, e cada conjunto de pontos na vertical, representa uma transformação gaussiana diferente sobre o histograma.	69
4.34 - Médias dos desvios padrão das simulações por modelo, lembrando que a média dos desvios padrão da população real é de 2.49.8, na figura cada curva representa um ajuste de variograma com determinado ponderador, e cada conjunto de pontos na vertical, representa uma transformação gaussiana diferentemente ponderada, aplicada sobre o histograma.	70
4.35 - Valores médios das simulações com a transformação gaussiana sem ponderador, com os variogramas ponderados pelos quatro métodos e o variograma sem ponderação.	71
4.36 - Valores médios das simulações, com transformação gaussiana ponderada usando o resultado fornecido pelo método Declus e com os variogramas ponderados pelos quatro métodos além do variograma sem ponderação.	71
4.37 - Apresenta as médias das simulações, cujo histograma foi aplicada transformação gaussiana, usando os ponderadores fornecidos pela KM, e sobre os variogramas ponderados pelos quatro métodos e também sem ponderação.	72
4.38 - Valores médios das simulações, com transformação gaussiana usando ponderador do polígono de influência, com os variogramas sem ponderador e também pelos quatro ponderadores.	72
4.39 - Valores médios das simulações, onde foi aplicada uma transformação gaussiana com ponderadores resultante de Richmod, variogramas com e sem ponderadores.	73

Lista de Tabelas

2.1 - Mostra para o mesmo conjunto de dados, diversos tamanhos de malha. Utilizado no método Declus para o calculo da malha ótima.....	13
2.1 - valores dos pesos para diversos efeitos de pepita, utilizando modelo esférico, patamar um, amplitude 3.5.....	29
3.1 - Dados para estudo.....	36
3.3 - Resumo das estatísticas sobre dados da tabela 3.2	43
4.1 - Dados sobre as cinco populações simuladas, utilizando simulação gaussiana não condicional no programa Isatis, A rotação foi especificada usando-se o referencial matemático. As anisotropias são relativas aos eixos x, y e z respectivamente	46
4.2 - dados relativos a população cluster .dat comparando a população real e a amostrada	47
4.3 - Apresenta os valores estatísticos dos dados de Walker Lake, população amostrada com 470 dados e dos 78000 dados.	47
4.4 - dados relativos a Imagem simulada 1, apresentando os dados da população completa e do conjunto de 228 amostras que foram retiradas.	48
4.5 - Apresenta as estatísticas do conjunto de dados relativos a imagem simulada 2. imagem completa 4000 pontos e os dados dela amostrados num total de 229 amostras.	48
4.6 - Conjunto de dados estatísticos, relativos a população da imagem 3 de 4000 pontos e a sua amostragem de 221 dados.	49
4.7 - Apresenta os valores estatísticos dos dados da Imagem 4, população amostrada com 218 dados e dos 4000 dados imagem onde foi feita a amostragem.	49
4.8 - Apresenta as estatísticas do conjunto de dados relativos a imagem simulada 5. imagem completa 4000 pontos e os dados dela amostrados num total de 221 amostras.	50
4.21 - A tabela a seguir mostra as diferenças das áreas da parte superior dos histogramas das sete populações estudadas, com diferentes ponderadores, quando comparadas com a população original	58

Capítulo 1

Introdução

Se dados são amostrados irregularmente no espaço, isto é, algumas amostras se concentram em uma pequena porção deste espaço, deve existir um conjunto de ponderadores que, quando aplicados de modo conveniente sobre o conjunto de amostras, anule a influência que estes agrupamentos provocam nas estatísticas.

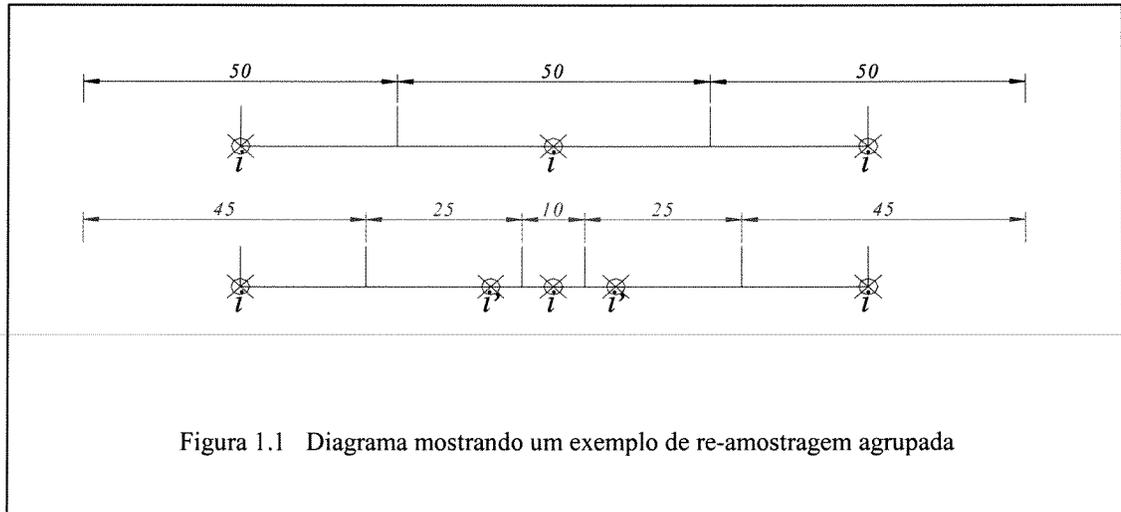
Quando vai-se fazer um estudo com variáveis espaciais, e estas variáveis representam valores geológicos, normalmente elas não são aleatoriamente nem regularmente distribuídas no espaço. Sendo assim, por várias razões, os dados podem ser amostrados (preferencialmente) com maior quantidade de pontos em determinadas regiões em detrimento de outras. O amostrador tem a tendência de concentrar um maior número de amostras em regiões de seu interesse, e assim, sem perceber, está criando um viés na sua amostragem. Se a amostragem é irregularmente distribuída no espaço, há normalmente uma concentração de amostras nas regiões de maior interesse. Esta tarefa de re-amostragem tem como objetivo melhorar o conhecimento da população a ser estudada, produzindo amostras na região de interesse, e criando, ao mesmo tempo, agrupamentos de amostras, com uma conseqüente influência nas estatísticas dos dados.

O que parece muito lógico do ponto de vista do empreendedor, causa uma diferença entre a estatísticas dos dados e os reais valores. Deve-se supor que as amostras, quando agrupadas em pequenas porções de espaço, não devem ter a mesma representatividade que amostras que estão representando um espaço maior.

1.1 Exemplo

Exemplificando-se este ponto de vista. Considere uma primeira amostragem com espaçamento de 50 m sobre uma linha reta, denominadas amostras i . Se posteriormente fizermos uma nova amostragem de duas novas amostras na vizinhança da amostra mais

rica, denominadas i' , estas distante de 5 m da amostra escolhida, podemos verificar que a representatividade fica prejudicada. Vide Figura 1.1



Do exposto, nota-se que, apesar de amostras representarem um espaço menor, quando se faz os estudos estatísticos sobre este conjunto de dados, não há uma preocupação com a distribuição espacial destas amostras, se está apenas interessado nos valores analisados em cada amostra. Se, como exemplificado, a re-amostragem foi feita na vizinhança das amostras mais ricas, a média, por exemplo, será deslocada na direção dos valores mais ricos. O mesmo acontece com o histograma, que apresentará esta mesma tendência.

Sendo assim, é válida a seguinte pergunta: O que poderíamos fazer para resolver ou minimizar este problema?

1.2 Métodos Estudados

Neste trabalho, analisa-se quatro métodos que se propõem gerar este conjunto de ponderadores, tendo como parâmetros sete conjuntos de dados obtidos de populações conhecidas. Os métodos são os seguintes: Polígono de influência, *Cell declustering* (Declus), Krigagem da média (KM) e Richmod

A técnica de desagrupamento consiste em se dar ponderadores p_i para amostras x_i de acordo com a irregularidade da amostragem. Este termo, desagrupamento, vem da tradução do termo em inglês “declustering” e talvez não é adequado, pois não reproduz a real intenção da proposta a ser seguida.

A palavra desagrupamento nos dá a impressão de desfacelamento do conjunto de amostras, deteriorando o grupo, que no caso não acontece, pois grupo continua existindo fisicamente, apenas o que se faz é atuar na influência que as amostras agrupadas fazem sobre as estatísticas.

Segundo Rivoirard (2000) “O desagrupamento é reconhecido como uma operação indispensável em geoestatística, quando ajuda a estimar o valor médio $z(V)$ de uma concentração sobre um domínio V ; esta operação, de aplicação bastante simples, depende de se ter os valores a que se usar como ponderadores.

A idéia, a princípio bastante simples, é de criar um conjunto de ponderadores que diminua a influência das amostras que estão próximas umas às outras. Neste estudo serão focadas quatro maneiras de se criar, um conjunto de ponderadores, utilizando as seguintes técnicas: Deutsch (1989), Richmod (2002), Krigagem da Média e Polígono de influência.

Das técnicas normalmente utilizadas, o de polígono de influência é a que mais seduz, pois é ela a que mais se aproxima da intuição de ponderação. É a imagem que uma amostra representa uma porção do espaço que a envolve e esta influência começa a diminuir a medida que se vai distanciando.

Uma segunda técnica já também bastante utilizada, é aquela proposta por Journel (1983) e desenvolvida por Deutsch (1989), que neste trabalho será chamada de Declus, que se propõe a dividir o espaço em pequenas porções, usando o inverso do número de amostras que se encontra em cada porção do espaço por ele definido, como o ponderador das amostras.

A técnica da krigagem da média utiliza-se dos ponderadores fornecidos pela krigagem quando esta é efetuada para se calcular a média das amostras; se está no caso,

interessado não é na média das amostras, mas sim no ponderador que cada amostra recebe para o cálculo da referida média.

O quarto método foi proposto por Richmod (2002) e pode ser entendido como um melhoramento do método Declus, pois suas porções de espaço não apresentam como naquele, uma distribuição uniforme, mas colocam-se no espaço de modo a acomodar os agrupamentos de dados.

1.3 Aplicação e exemplos

Neste trabalho, serão apresentados sete conjuntos de dados aos quais se efetuará tratamentos estatísticos utilizando-se sobre eles ponderadores fornecidos pelos quatro métodos de cálculo de ponderador.

Uma segunda parte do trabalho compreenderá o ajuste do variograma, utilizando-se os ponderadores sobre o variograma experimental e o estudo da introdução desta ferramenta nos resultados finais da krigagem e sobre a simulação gaussiana. Será utilizado o programa Anavar(1995) que possibilita o uso de ponderadores no cálculo do variograma, sendo que a rotina implementada no programa, segue a teoria desenvolvida por Rivoirard (2000). O programa foi desenvolvido no Instituto de Geociências da Unicamp, como ferramenta auxiliar no ensino da geoestatística e posteriormente foi acrescentado a esta ferramenta de cálculo de variograma a possibilidade de se utilizar ponderadores. Este mesmo programa foi sendo modificado à medida que o desenrolar da dissertação necessitava de um algoritmo que possibilitasse o uso de ponderadores; assim, foram introduzidos nos algoritmos de cálculo de histograma, média e variância.

Utilizando os variogramas experimentais calculados pelo Anavar, podemos importa-los para o Isatis, neste caso introduzindo os variogramas experimentais já calculados e efetuando o ajuste dos variogramas teóricos no Isatis, ou utilizando o programa Ajusta, também desenvolvido aqui no departamento de Geociências, para o ajuste dos variogramas teóricos e utilizando estes resultados. Utilizou-se do Programa Isatis, para efetuar a krigagem e simulação e como os dados foram retirados de uma população conhecida, pode-se efetuar uma comparação entre a população real e a

estimada e, efetuando uma krigagem com variogramas sem ponderador, comparar as duas populações e seu valor real.

Também se pode comparar os vários resultados fornecidos pela krigagem/simulação, quando se faz uso dos diferentes tipos de ponderadores aplicados ao variograma e/ou histograma. No estudo, efetuou-se apenas a simulação Gaussiana.

1.4 Organização da Dissertação

A dissertação está organizada segundo os seguintes capítulos:

- Capítulo 2, onde é realizada a revisão bibliográfica sobre os métodos abordados nesta dissertação.
- Capítulo 3, onde se apresenta procedimentos adotados, para a utilização dos ponderadores em diversos cálculos estatísticos.
- Capítulo 4, onde são apresentados estudos sobre a influência dos métodos de desagrupamento nas técnicas de krigagem e simulação gaussiana utilizando-se de diferentes conjuntos de dados.
- Capítulo 5, onde são relacionadas as principais conclusões e recomendações para a realização de trabalhos futuros.

Capítulo 2

Revisão Metodológica

Neste capítulo, realiza-se uma revisão da literatura sobre os métodos são abordados nesta dissertação, bem como uma descrição metodológica de seus respectivos conteúdos.

2.1 Polígono de Influência

O método do polígono de influência se baseia na intuição de que uma determinada amostra possui uma área de influência ao redor dela, ou melhor, ao redor de uma amostra existe uma população que possui relações com a amostra e esta área esta relacionada às distâncias entre o ponto em estudo e as amostras vizinhas mais próximas. O método utiliza-se das áreas formadas pelos polígonos, como base para a criação dos ponderadores. A Figura 2.1, exemplo do método do polígono de influência e como ele pode ser usado para desagrupar um conjunto de dados.

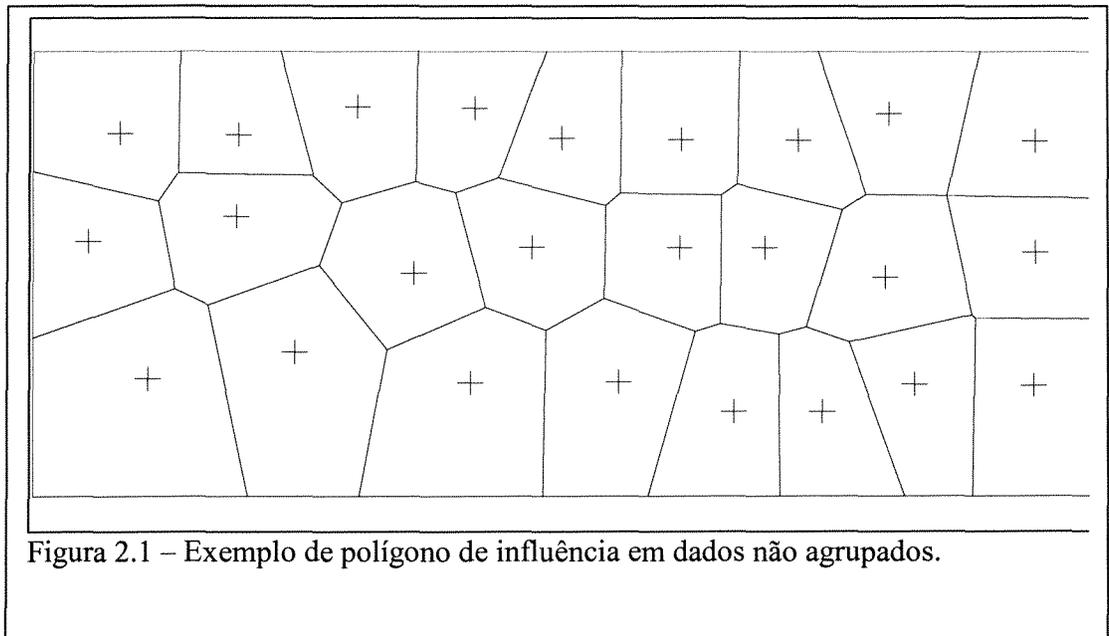


Figura 2.1 – Exemplo de polígono de influência em dados não agrupados.

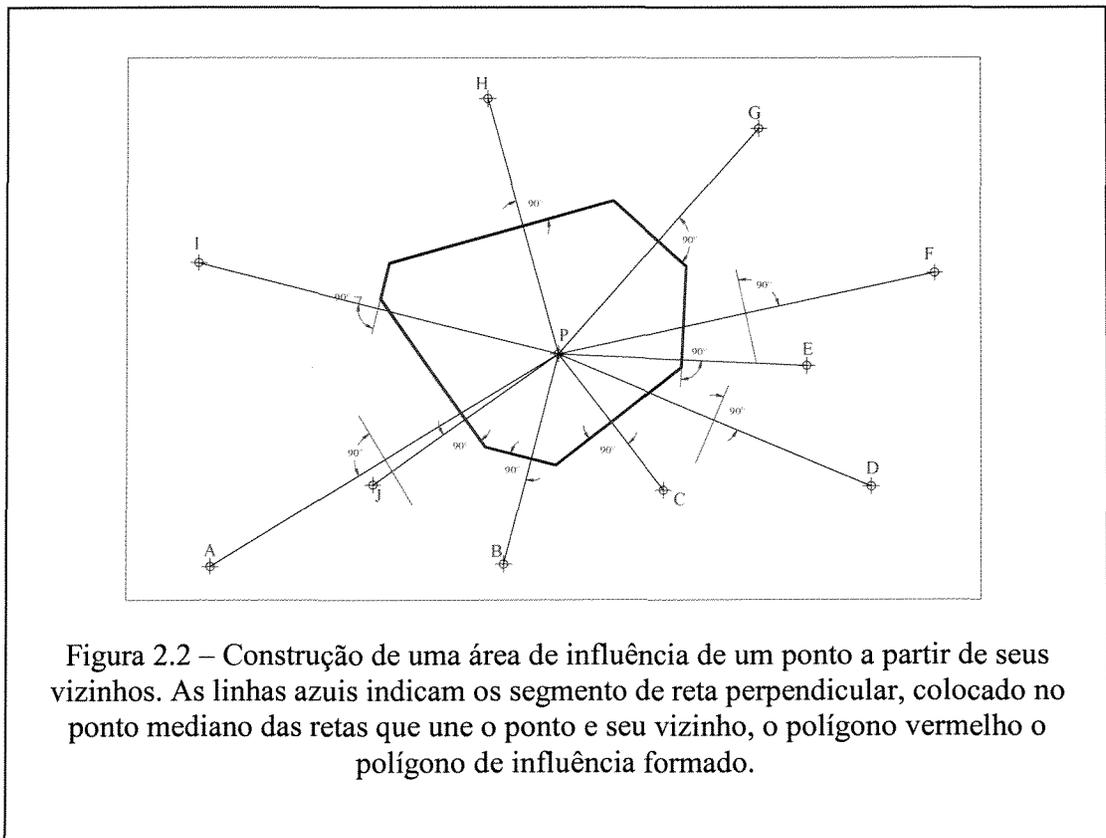
O uso do polígono de influência foi bastante utilizado na mineração para o cálculo de reservas minerais, partindo do pressuposto que em volta de um determinado ponto existem amostras com as mesmas características do ponto a ser estudado. Este fato não deixa de ser realidade, mas de uma amostra à outra deve existir uma função de variabilidade, e não um valor que se estende constante até encontrar a interseção com a outra amostra, como é o caso do polígono de influência. No caso de regiões com uma anisotropia bastante elevada, ou mesmo com uma anisotropia pequena, se o amostrador não levou em conta esta peculiaridade, fato bastante difícil quando ele ainda está na fase de amostragem, haverá um erro considerável. Somente na fase de estudo vai se chegar ou tentar chegar a anisotropia do corpo, se esta existir. Logicamente, no caso de mineração, um estudo geológico da área de trabalho ajuda consideravelmente nesta tarefa.

Como técnica ela é válida e induz a pensar em novas técnicas que pudessem resolver o problema aqui proposto, isto é, achar um conjunto de ponderadores. Este método também se apresenta útil quando os agrupamentos não são muito marcantes. Se estivermos em um campo, com um número mínimo de amostras, irregularmente espaçadas, podemos utilizar os ponderadores das amostras para os cálculos estatísticos, se não mais preciso, pelo menos mais realista.

Como exemplo, podemos citar um caso de um campo de petróleo, onde temos apenas cinco amostras distribuídas irregularmente. Podemos ponderar os valores representados nas amostras pela área de influência de cada amostra; este valor está mais próximo da realidade do que se calcularmos uma média, sem levarmos em conta a distribuição das amostras. Como veremos posteriormente este mesmo raciocínio pode ser usado no caso de se utilizar ponderadores obtidos pela KM. No caso dos dois outros métodos que serão estudados, quando existe uma pequena quantidade de amostras, dificilmente elas estarão em um mesmo agrupamento, o que irá conduzir a que todas as amostras possuam um mesmo ponderador, igual a $1/n$, onde n o número de amostras.

Outro fato neste método que merece ser abordado é que ele não considera as amostras próximas como um agrupamento, mas sim apenas a distribuição espacial das amostras que estão na sua vizinhança, grupos de amostras notadamente pertencentes a um mesmo agrupamento possuem valores de ponderadores totalmente diferentes, ligados exclusivamente à sua vizinhança.

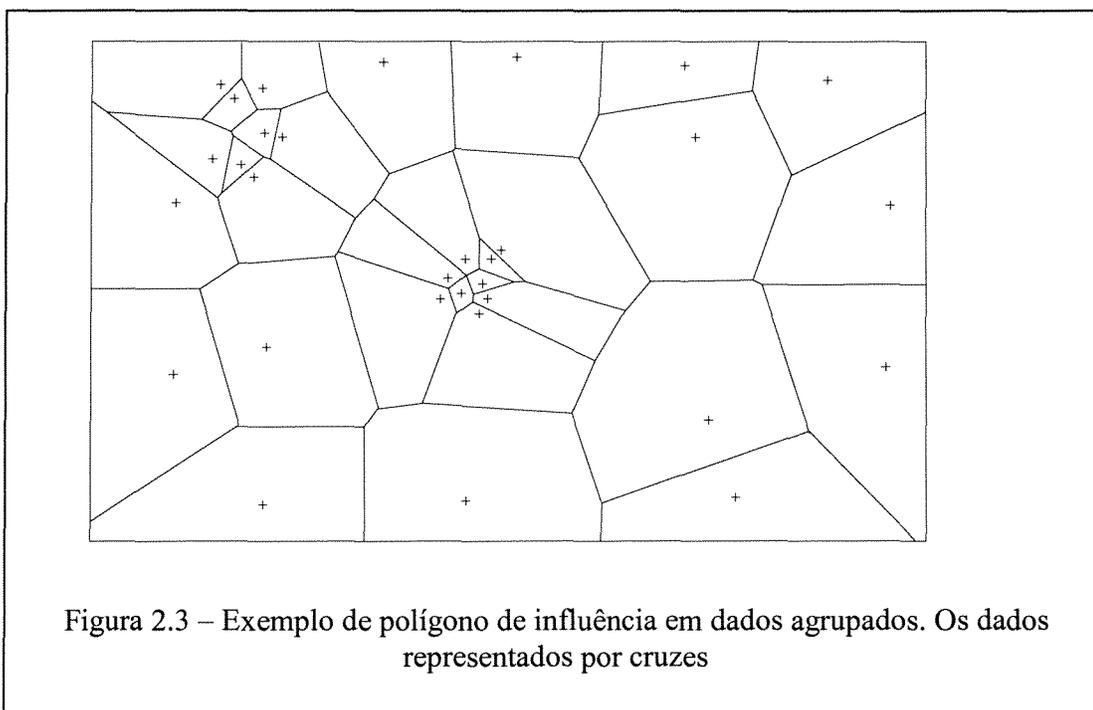
O uso do polígono de influência foi primeiramente descrito por volta de 1909, sendo usado comercialmente em 1920. Na década de 1970, a técnica foi utilizada por David (1977). No início, seu cálculo era efetuado utilizando a geometria clássica e o cálculo da área de influência era efetuado manualmente. No caso de grandes quantidades de dados, a sua construção era árdua e facilmente era possível cometer-se um erro, devido a grande quantidade de retas e intersecções, que devem ser feitas para se chegar ao resultado final. Com o advento da computação, o método foi então usado com o auxílio desta ferramenta, facilitando muito seu uso e eliminando os erros causados anteriormente pelo procedimento manual.



Tais programas de computador utilizam os polígonos de Voronoi ou triangulação de Delaunay, bastante difundidos na matemática, para o desenvolvimento do algoritmo do cálculo da área do polígono que se forma em torno do ponto. A Figura 2.2 ilustra a construção do polígono de influência do ponto P, com seus 10 vizinhos, sendo que os pontos A, B, D e F, estão mais distantes e não interferem na construção do polígono.

Este método apresenta dois inconvenientes principais: a construção dos polígonos em torno de pontos que se encontram nas bordas e a presença de agrupamentos em que a distância entre os pontos pertencentes ao agrupamento e os dados que os rodeiam varia muito.

Para resolver ou minimizar o primeiro problema, é necessário a criação de uma linha envoltória que delimita a área dos pontos que estão na periferia. Surge então a seguinte questão. Quão afastada esta linha deve passar dos pontos das bordas? A resposta



é que não existe uma regra para isto, o usuário decide o valor desta distância.

Este método é aplicável, quando os pontos têm uma distribuição relativamente uniforme no espaço e quando estas áreas são compatíveis com a distribuição dos pontos. No caso em estudo, quando ocorre um agrupamento de pontos, as áreas podem ficar com tamanhos diferentes para dados de um mesmo agrupamento. As Figura 2.3 e 2.4 mostram a deficiência do método quando aplicado a dados agrupados; observa-se que dados localizados em um mesmo agrupamento possuem ponderadores diferentes.

A Figura 2.4 mostra os problemas que o polígono de influência apresenta quando atua sobre dados que formam dois agrupamentos, que estão a uma distância relativamente grande quando comparada às distâncias entre os pontos contidos em cada agrupamento. Observa-se que apesar de estarem geometricamente definidas estas áreas não parecem representar a realidade sobre a área de influência dos dados.

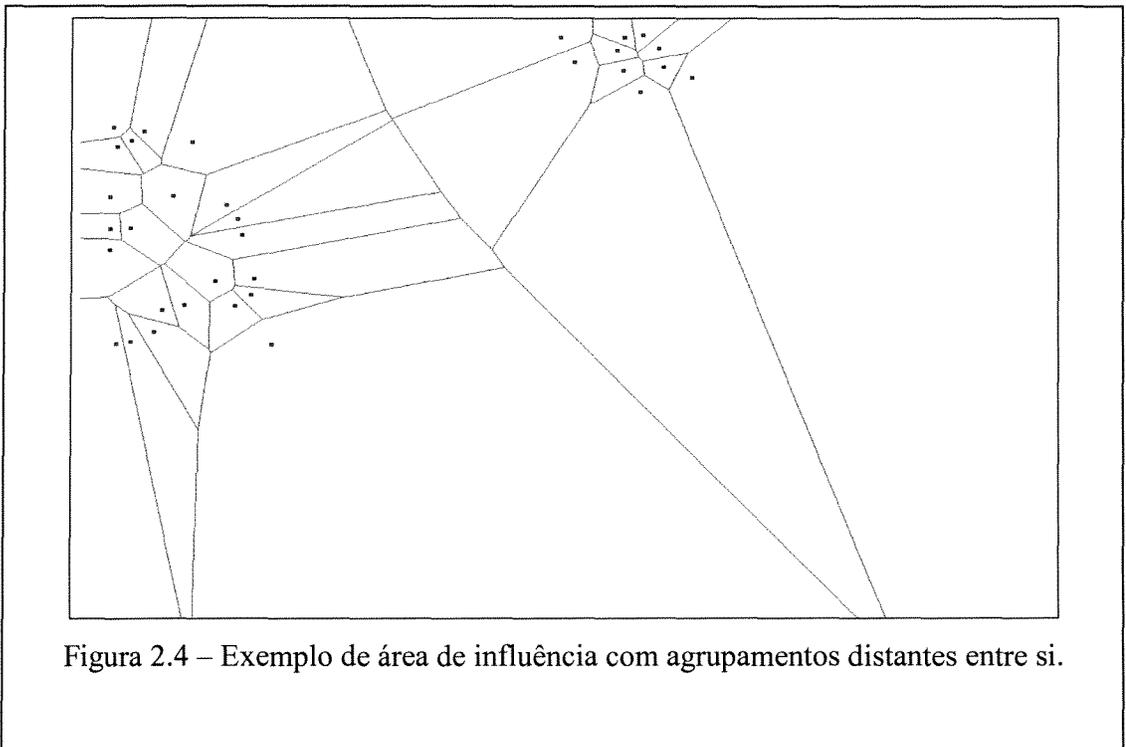


Figura 2.4 – Exemplo de área de influência com agrupamentos distantes entre si.

2.2 Método – Cell Declustering (Declus)

O método proposto inicialmente por Journel (1983), objetivava conseguir um conjunto de ponderadores com o intuito de diminuir a influência das amostras agrupadas. Este método também é conhecido como *Cell Declustering* Deutsch (1989), o qual implementou uma rotina em FORTRAN, que possibilitou o uso do método proposto por Journel. Esta faz parte das rotinas disponíveis na biblioteca de programas conhecidos como GSLIB (Geostatistical Software Library).

A proposta deste método é dividir o espaço em células de tamanho constante, de modo que cada amostra fique contida em uma célula, porém mais de uma amostra pode estar dentro da mesma célula, e neste fato se baseia a idéia do método, conseguir alocar dentro de uma mesma célula todas as amostras de um mesmo agrupamento e atribuindo a cada amostra um ponderador igual $1/n$, sendo n o número de amostras que estão contidas dentro da célula.

Imagine um papel transparente, todo quadriculado, com distância entre as linhas iguais entre si, e que as distâncias entre as colunas também são iguais entre si, mas não necessariamente a distância na horizontal é a mesma que na vertical.

Sobre o conjunto de dados, deve-se posicionar esta malha, rotacionando e transladando o papel, de modo que, cada agrupamento esteja dentro de uma mesma célula. Escolhido o ponto que o maior número de agrupamentos estejam dentro de uma única célula, faz-se uma contagem de quantos pontos têm em cada célula e efetua-se o cálculo da média, multiplicando-se cada valor por $1/n$, sendo n o número de pontos dentro da célula em questão. Modificam-se então os valores das dimensões na horizontal e vertical, e novamente encontra-se a posição ideal e um novo cálculo de média deve ser feito. Repete-se o procedimento diversas vezes. A posição e dimensões ideais, seriam aquelas que retornam o menor valor para a média, isto, caso os dados estejam com os agrupamentos colocados nas regiões ricas do corpo.

Se tivéssemos condição de colocar cada agrupamento em uma única célula, a média desta população seria mínima, se as amostras agrupadas estiverem concentradas

nas regiões ricas, pois os ponderadores destas amostras seriam menores, e sendo seus valores altos, a média tenderia a ir para um mínimo. Conseguir isto é uma tarefa árdua se o trabalho for executado manualmente. Foi pensando nisso que foi desenvolvido um programa computacional que procurasse a melhor malha, que quando colocada sobre os dados resulta em uma média mínima ou máxima, dependendo se os agrupamentos localizarem-se nas regiões ricas ou pobres, respectivamente.

O algoritmo para resolver este problema foi desenvolvido utilizando a seguinte linha de raciocínio:

1. Escolhe-se uma origem para a inserção da malha e divide-se o espaço em células de igual tamanho;

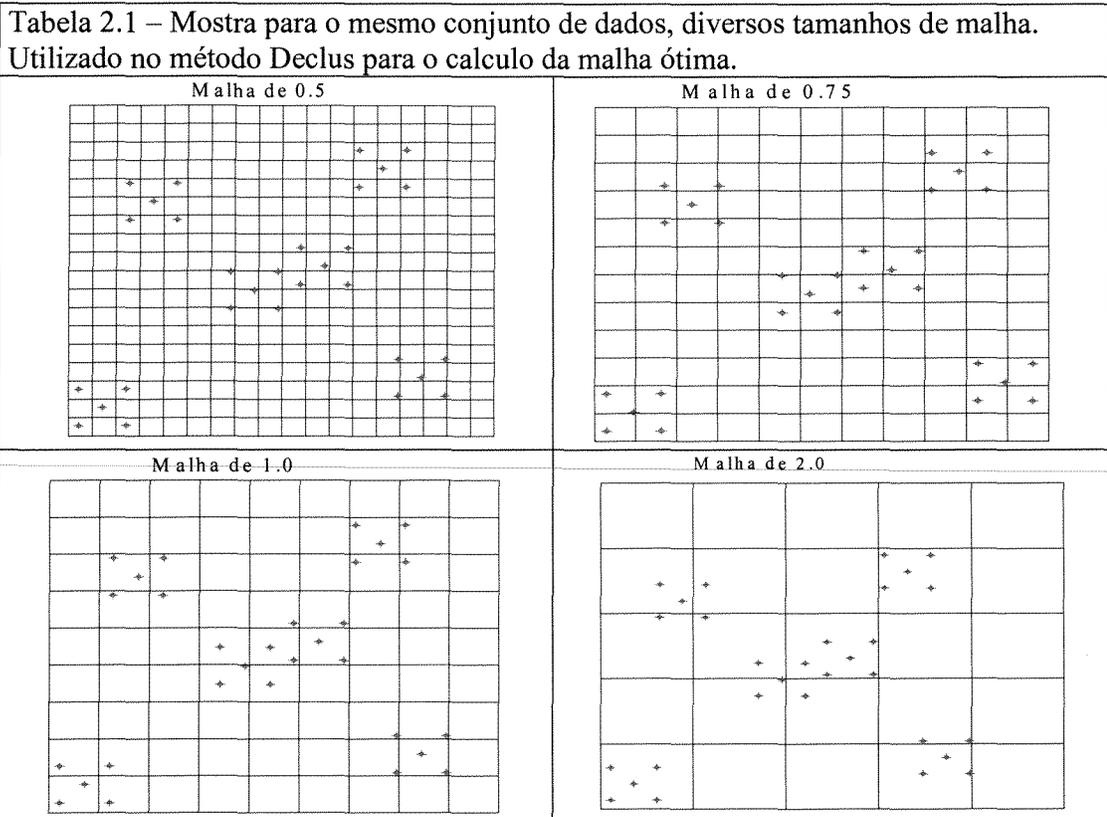
2. Tendo feito a divisão do espaço em células, percorrem-se todas as amostras e vai se alocando cada amostra em sua respectiva célula; ao final, ter-se-á o número de amostras que cada célula possui, e o inverso deste valor será utilizado como ponderador das respectivas amostras. Neste instante, deve calcular a média das amostras utilizando-se os ponderadores obtidos;

3. Agora se incrementa o tamanho da célula em um valor, e repete-se a tarefa anterior;

4. Faz se esta operação uma determinada quantidade de vezes, definida de antemão, sendo que a cada operação deve-se calcular a média e guardar este valor, assim como as respectivas dimensões das células;

5. Efetuando estes cálculos das médias, para cada configuração, cria-se uma função que relaciona o tamanho da célula versus a média, podendo-se então procurar o ponto de mínimo desta função em relação à média. É neste ponto, teoricamente, que um maior número de agrupamento ficaram alocados dentro de uma mesma célula.

A Tabela 2.1 apresenta 4 tamanhos diferentes de células, todas com o mesmo ponto de inserção.



Além das dimensões da célula, uma outra variável que é muito importante, é a origem da inserção das células; por isto, o método também apresenta uma faixa de aplicação do ponto de inserção.

Para se conseguir um efeito de anisotropia, é possível se colocar um fator dx , dy , que são fatores que multiplicam as dimensões da célula nas direções x e y , respectivamente.

Normalmente, as malhas são dispostas paralelas aos eixos das abscissas, mas é também possível atribuir-se um ângulo de inserção da malha, ângulo este que definirá a direção com que a malha em relação aos eixos das abscissas.

Do exposto, pode-se verificar a dificuldade em se alcançar a um valor ótimo, com esta possibilidade de mudar tantas variáveis. O valor ótimo é um valor praticamente inalcançável.

Este método apresenta certos inconvenientes. Um deles é que dificilmente as amostragens são feitas numa malha regular. Se por um acaso isto ocorrer, as reamostragens serão feitas nas partes ricas e conseqüentemente, seguindo o corpo a ser estudado, e não mais uma distribuição espacial bem definida; ao contrário do método que se distribui uniformemente sobre todo o espaço. Assim agrupamentos podem ficar separados devido à restrição do mesmo. Neste item devemos notar que para uma mesma malha, se modificarmos o ponto de inserção, o valor da média é modificado, e que agrupamentos reais podem ficar divididos.

O método se baseia na procura de uma média mínima/máxima, não se importando com a distribuição espacial dos dados, pois em nenhum momento se está interessado nos vizinhos do ponto, ou distância entre eles; os agrupamentos são encontrados de maneira indireta. Em amostragem sem uma tendência definida, ou se os agrupamentos estiverem divididos em regiões ricas e pobres, o método falha, já que seu intento é minimizar/maximizar a média.

Em um teste efetuado durante este estudo, apresentado na Figura 2.18, fez-se uma malha de amostragem e os agrupamentos foram definidos de modo que com uma distância dx para o tamanho da malha, os agrupamentos ficassem todos separados, isto é, existia uma configuração que englobaria cada agrupamento em uma única unidade da malha. Como o método procura a média mínima, esta distribuição espacial não foi respeitada e no final, o método resultou em um tamanho de malha que era 2.3 vezes maior que aquela desenhada.

Um outro ponto a ser levantado é sobre o ângulo a ser utilizado. Como ele é constante para as células, não se consegue ajustar agrupamentos que não possuem o mesmo ângulo.

Em conseqüência destas três variáveis, os agrupamentos poderão ser fracionados durante o processo, (se agrupando com determinado tamanho de célula) e depois ficarem divididos conforme se modifica os valores das dimensões das células durante o processo de pesquisa de minimizar/maximizar a média.

2.3 Método Krigagem da Média (KM)

Este método de cálculo de ponderadores, que utiliza a krigagem global, foi proposta inicialmente por Isaaks e Srivastava (1984) e Remacre (1996)

Na krigagem ordinária, o objetivo é estimar a função de uma variável regionalizada que é o valor de um bloco como uma média sobre um bloco. No caso da KM, a função é estimar o valor médio (desconhecido).

Quando se trabalha com valores espacialmente distribuídos, eles podem ou não ser tomados como independentes, pois geralmente possuem uma função de covariância associada entre eles. Se um modelo de covariância não for um efeito de pepita puro, esta média que se vai calcular supõe que as variáveis aleatórias em dois pontos distintos no espaço são correlatadas, e a sua formula é dada por:

$$m^* = \sum_{i=1}^n \lambda_{mi} Z(x_i), \quad (2.1)$$

onde: m^* é média estimada das amostras, n é o número de amostras, λ_m é o ponderador e $Z(x_i)$ é o valor da amostra.

Como calcular os ponderadores?

Seja $Z(x)$ Função aleatória de segunda ordem, ou seja, existe esperança

$$m = E[Z(x)]$$

para todo x , e existe covariância

$$C(h) = E[Z(x) - Z(x+h)] - m.$$

E queremos que

$$m^* - m = 0.$$

Este estimador deverá ser não enviesado e com variância mínima. Deste modo, o erro de estimativa deverá ser de zero.

Sendo assim:

$$E[m^* - m] = E\left[\sum_{i=1}^n \lambda_{mi} Z(x_i) - m\right] = 0 \quad (2.2)$$

$$= \sum_{i=1}^n \lambda_i E[Z(x_i)] - m \quad (2.3)$$

se $E[Z(x_i)] = m^*$, devemos ter que a soma dos ponderadores seja igual a um:

$$= \sum_{i=1}^M \lambda_i m - m = m \sum_{i=1}^M (\lambda_i - 1) \quad (2.4)$$

Como na krigagem ordinária, a variância é minimizada ao se forçar que seus ponderadores sejam mínimos, utilizando um multiplicador de Lagrange. O sistema de krigagem então fica:

$$\sum_{j=1}^M \lambda_{mj} C(x_i, x_j) = \mu_{mi} \quad i = 1, \dots, N \quad (2.5)$$

onde: μ_{mi} é o ponderador de lagrange e,

$$\sum_{i=1}^M \lambda_{mi} = 1 \quad (2.6)$$

A variância do erro de estimação é:

$$\text{var}[m^* - m] = \text{var}\left[\sum_{i=1}^n \lambda_{mi} Z(x_i) - m\right] = \sum_j \sum_i \lambda_{mi} \lambda_{mj} C(x_i, x_j) \quad (2.7)$$

onde $C(x_i, x_j)$ a covariância entre x_i e x_j .

Em notação matricial, o lado esquerdo do sistema é idêntico a krigagem ordinária e o lado direito é uma coluna de zeros, exceto, o último termo, o qual é igual a um, correspondendo à obrigação dos ponderadores terem soma igual a um.

$$[k]= \begin{pmatrix} \bar{C}(x_1, x_1) & \dots & \bar{C}(x_1, x_\beta) & \dots & \bar{C}(x_1, x_n) & 1 \\ \vdots & & \vdots & & \vdots & \vdots \\ \bar{C}(x_\beta, x_1) & \dots & \bar{C}(x_\beta, x_\beta) & \dots & \bar{C}(x_\beta, x_n) & 1 \\ \vdots & & \vdots & & \vdots & \vdots \\ \bar{C}(x_n, x_1) & \dots & \bar{C}(x_n, x_\beta) & \dots & \bar{C}(x_n, x_n) & 1 \\ 1 & & 1 & & 1 & 0 \end{pmatrix} \begin{bmatrix} \lambda_{n1} \\ \vdots \\ \lambda_n \\ 0 \\ -\mu_0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (2.8)$$

Do exposto, nota-se que o cálculo da média através da KM, leva em conta a disposição dos pontos no espaço, o valor do ponderador será diferente para cada valor dependendo da disposição espacial, e do modelo de correlação que se assume existir entre os dados. Dados da borda terão ponderadores diferentes daqueles no centro e dados agrupados terão um valor de ponderador diferente daqueles isolados.

2.3.1 Caso particular: Efeito de Pepita Puro

Um caso interessante estudado é quando o modelo de regionalização for efeito de pepita puro. Neste caso teremos:

$$C_{nug}(x_i - x_j) = \begin{cases} \sigma^2 & \text{se } x_i = x_j \\ 0 & \text{se } x_i \neq x_j \end{cases} \quad (2.9)$$

sendo C_{nug} , a covariância para o efeito de pepita puro.

Neste caso o sistema de krigagem se simplifica e torna-se:

$$\sum_{j=1} \lambda_{mi}(x_i, x_j) = \mu_{KM} \quad i = 1, \dots, n \quad (2.10)$$

$$\sum_{i=1}^n \lambda_{mi} = 1 \quad (2.11)$$

Todos os ponderadores são iguais e a soma é igual a um, portanto:

$\lambda_{ni} = 1/n$, ou seja o estimador krigagem da média é equivalente a média aritmética.

Do exposto, chega-se a conclusão que não devemos utilizar este método de cálculo de ponderadores quando os variogramas apresentarem um comportamento de efeito de pepita puro, visto que os ponderadores neste caso são todos iguais a $1/n$, sendo n o número de dados amostrados.

Uma incoerência do método, é que para se conseguir os ponderadores, deve-se partir de um variograma e sendo os dados agrupados, porque não utilizar ponderadores sobre os variogramas? Como pode-se fazer uma krigagem se não temos um variograma teórico? Para resolver este impasse, com os dados que se dispõe, deve-se calcular um variograma e utilizando-se deste, auferir um variograma teórico, que vai-se utilizar para fazer a krigagem e desde modo, conseguir os ponderadores, que serão utilizados para se chegar ao variograma que primeiramente era desejado.

No caso do uso da krigagem da média, como um método de cálculo de ponderadores, deve-se notar que ao contrário dos outros métodos aqui estudados, que são todos baseados somente na distribuição espacial dos dados, isto não acontece, pois, além da distribuição espacial tem-se também a covariância entre os dados. Se estivermos trabalhado com duas variáveis diferentes, os ponderadores, como são definidos pela covariância entre as amostras, podem mudar de variável para variável. Também neste caso, existe a possibilidade de que uma variável não possua a mesma anisotropia da outra, os ponderadores para uma mesma amostra variam de variável para variável.

2.3.2 Influência dos parâmetros da variografia

Pelo fato do método KM, se utiliza da função de covariância, é interessante verificar o comportamento dos ponderadores quando se modifica os parâmetros de ajuste de variograma.

Estes parâmetros a serem estudados são os seguintes: patamar, modelo de variografia, efeito de pepita, amplitude e anisotropia. No caso, o efeito de pepita, que não deixa de ser um modelo de variograma, poderia ser estudado junto àqueles, mas dada a sua grande influência, vai-se desenvolver um estudo a parte.

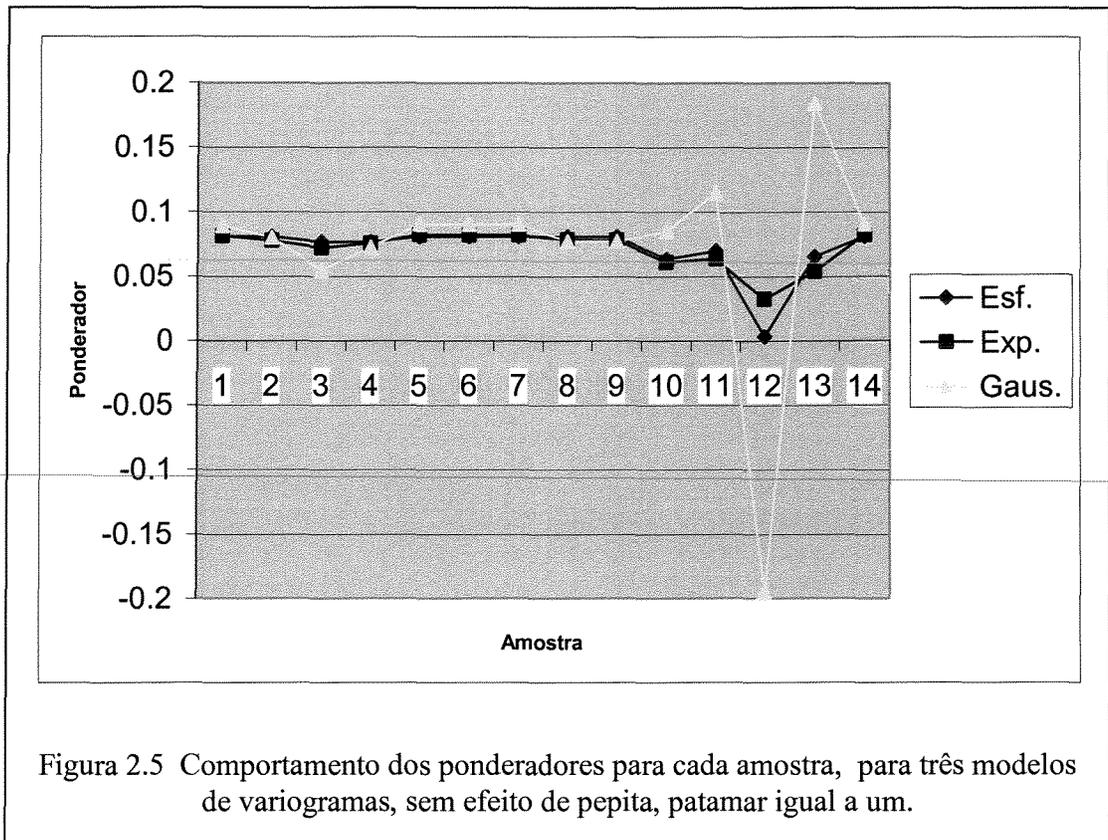
Neste trabalho são apresentados apenas alguns exemplos. Estudos mais aprofundados sobre as influências nos ponderadores, podem ser vistos em livros de Geoestatística, visto que os parâmetros aqui estudados são os mesmos fornecidos aos dados durante um estudo geoestatístico. Em especial, cita-se o livro de Isaaks e Srivastava (1989, pg 300-315).

2.3.2.1 O patamar

De acordo com Isaaks e Srivastava (1989) o patamar não modifica os valores dos ponderadores e este é um dado importante, tendo em vista que, normalmente, o uso de ponderador é seguido de uma diminuição da variância dos dados e, conseqüentemente, do patamar.

2.3.2.2 O modelo de variograma

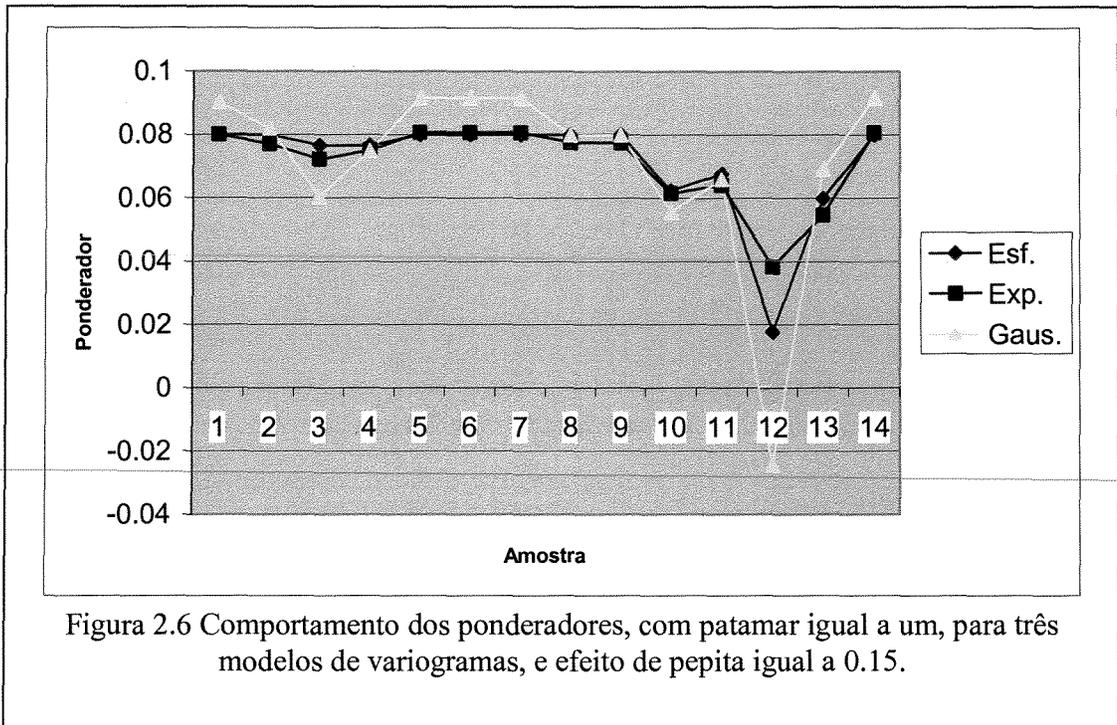
Este caso, é de vital importância, dado que os valores dos ponderados são extremamente afetados pelo modelo usado, é importante levar-se em conta a continuidade na origem. Modelos como o gaussiano, que possui grande continuidade na origem, diverge substancialmente de modelos como o esférico, que apresenta um comportamento mais linear na origem.



As figuras 2.5 e 2.6 mostram o comportamento dos ponderadores, quando modificamos o modelo do variograma, com e sem efeito de pepita, respectivamente.

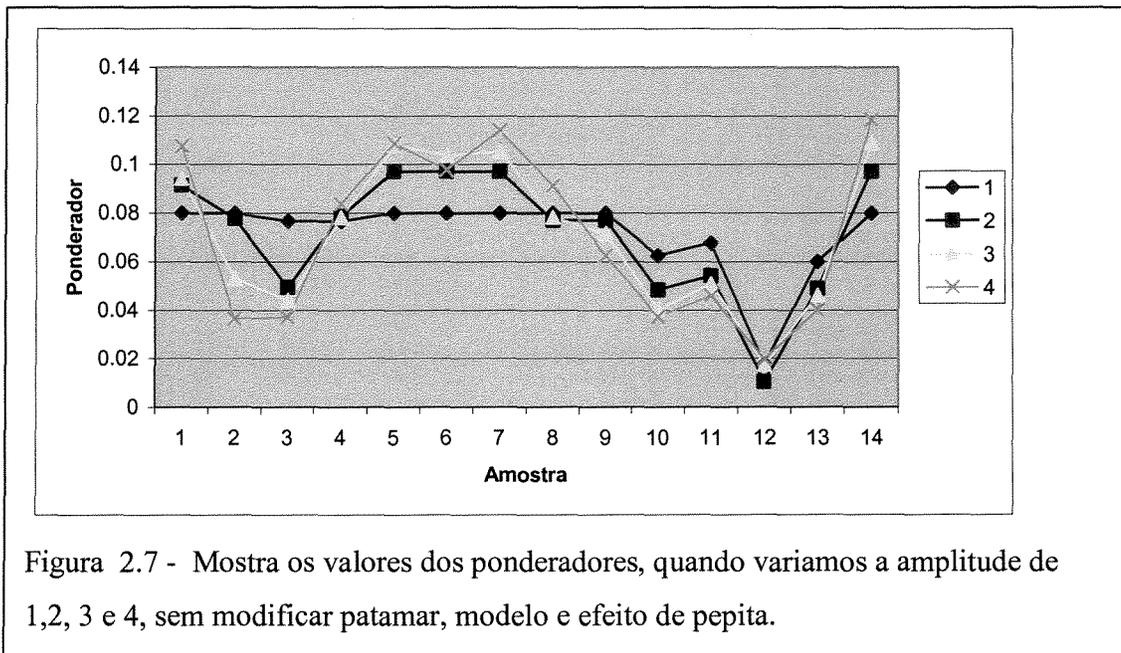
2.3.2.3 O Efeito de Pepita

O efeito de pepita é um dos parâmetros de maior importância, fato que originou um estudo de caso na subseção 2.3.1, quando se comenta o uso da KM como gerador de ponderador. Um alto valor, do efeito de pepita, próximo ao do patamar, implica que os dados terão seus ponderadores o mais próximo do valor $1/n$; quando o valor do efeito de pepita tende ao patamar, os ponderadores tendem a $1/n$.



2.3.2.4 A amplitude

Quando modificamos a amplitude de um variograma, os ponderadores não sofrem



grande variações, são necessárias grandes variações para que um pequeno valor seja

acrescido ao ponderador, sendo que este pequeno valor pode ser positivo ou negativo,

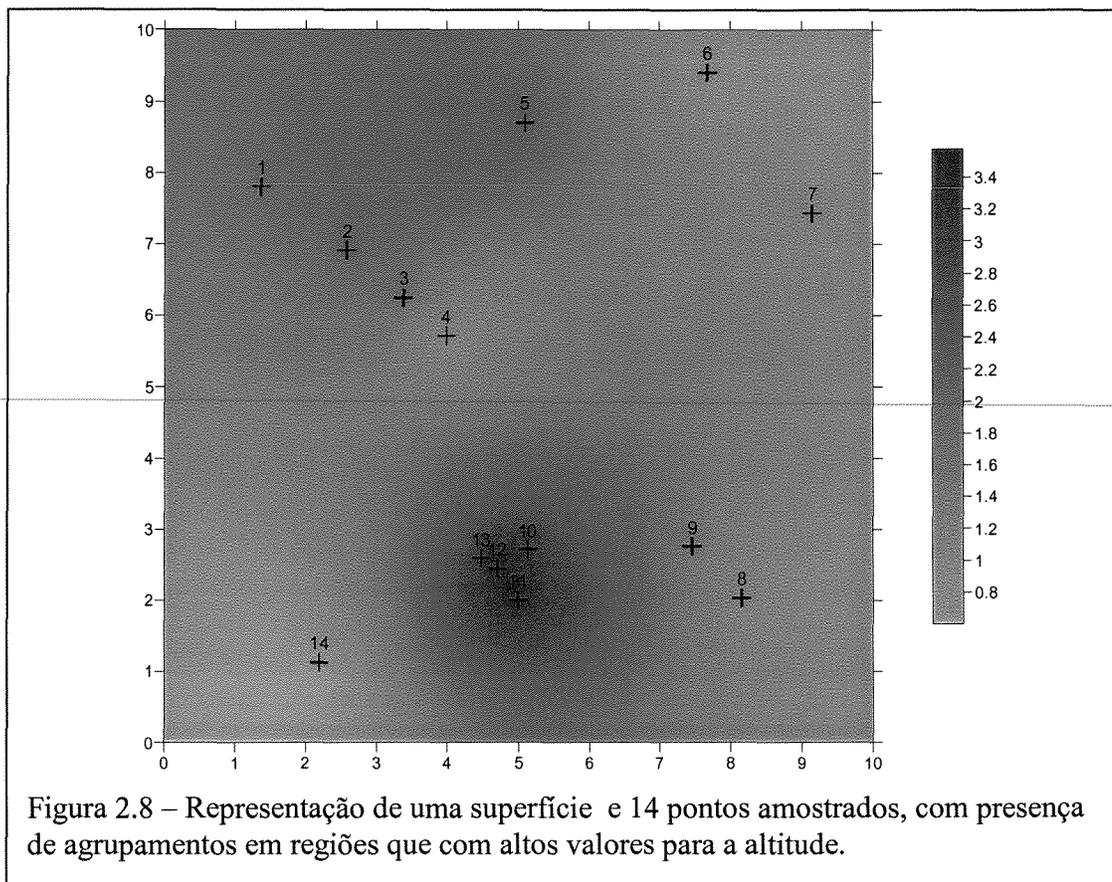


Figura 2.8 – Representação de uma superfície e 14 pontos amostrados, com presença de agrupamentos em regiões que com altos valores para a altitude.

como mostra a Figura 2.7, onde foi modificada a amplitude de quatro vezes e a maior variação nos pesos foi de 50 %

2.3.2.5 A anisotropia

Um fator de anisotropia pode causar grandes modificações nos ponderadores. Em Srivastava (1989), são discutidos vários exemplos que ilustram a influência que este parâmetro tem sobre os ponderadores, grosseiramente, uma anisotropia em uma direção que seja o dobro da outra, implica que os ponderadores terão o dobro do valor na direção de menor amplitude.

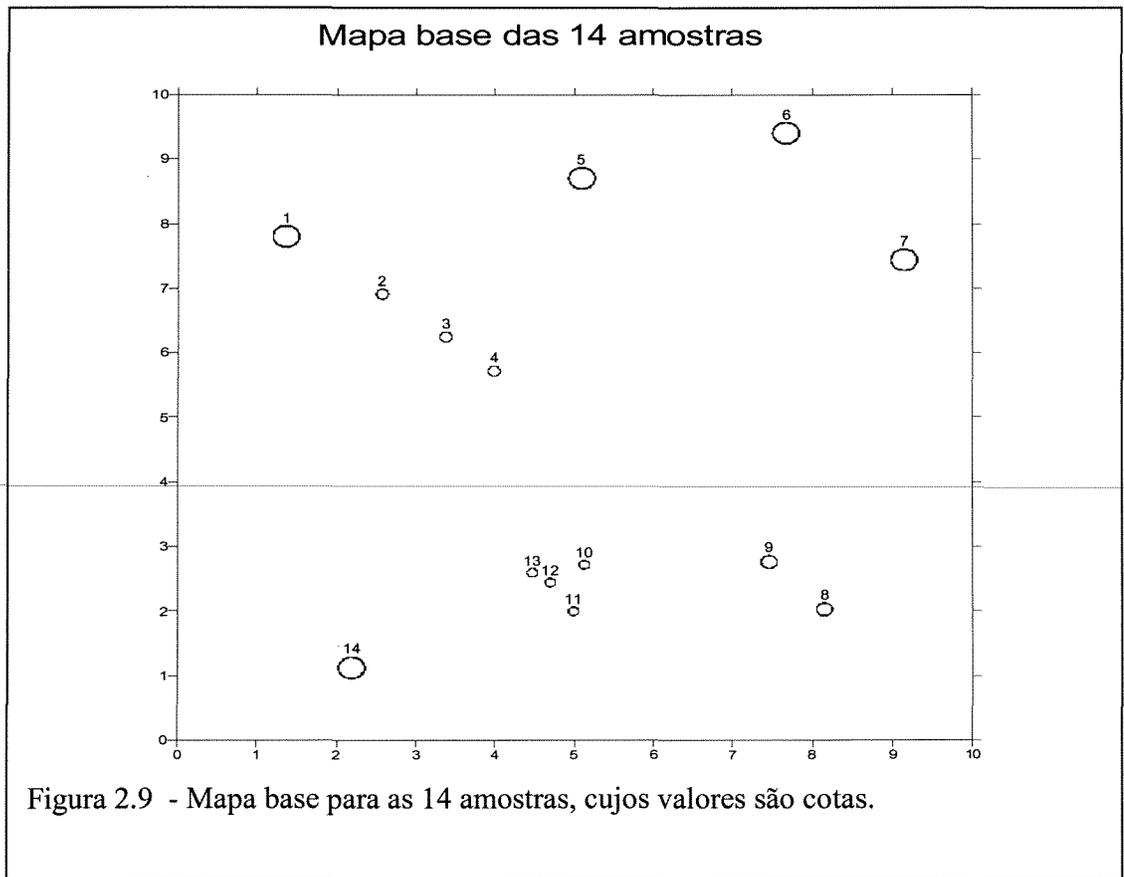
Do exposto, podemos notar que quando se utiliza da KM para o cálculo dos ponderadores, tem-se uma infinidade de opções, pode-se conseguir qualquer valor para os

ponderadores, mas, um ajuste compatível; e um mínimo conhecimento sobre os dados é necessário, para não incorrerem em erros grosseiros

A Figura 2.8 representa a distribuição espacial de 14 amostras, sobre a superfície da qual elas foram amostradas, que serão utilizadas para um estudo da variabilidade dos ponderadores para **KM** como ferramenta de cálculo dos ponderadores. Na Figura 2.9, apenas como ilustração, estão colocadas as amostras identificadas por círculos.

No caso da pesquisa de ponderadores utilizando a KM, deve-se investigar qual é o modelo de variograma a ser utilizado para se conseguir a menor média, ou o melhor conjunto de ponderadores, que quando aplicados aos dados forneça uma previsão o mais real possível da população da qual eles foram subtraídos.

Deste modo, dar-se-á uma visão geral do comportamento dos ponderadores, quando vai se modificando os parâmetros de entrada da rotina de cálculo dos ponderadores. O pequeno número de amostras, 14, é proposital, para que não nos percamos em uma grande quantidade de curvas e nos atentemos para a variação dos ponderadores.



Antes de se estudar o comportamento dos ponderadores vamos verificar o comportamento da média sobre o conjunto de dados *cluster.dat*.

Primeiramente, vamos visualizar as variabilidades da “média”, quando se altera os três parâmetros possíveis de serem modificados. O primeiro deles é relativo ao efeito de pepita; nas figuras abaixo, mostramos as curvas de variabilidade da média, quando o efeito de pepita vai de zero, curva mais embaixo da família, até a curva superior, que representa efeito de pepita puro, quando todos os ponderadores passam a apresentar o mesmo valor de $1/n$.

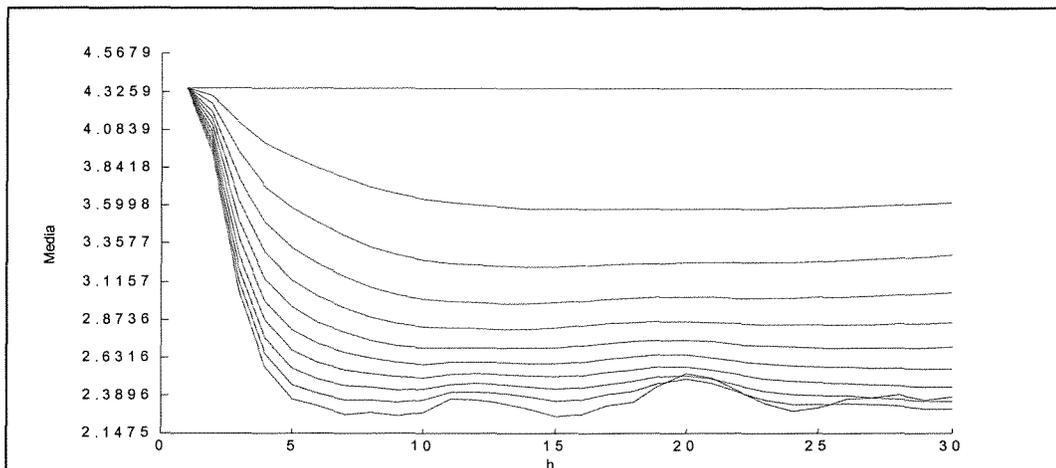


Figura 2.10 - Curvas da variação da média para os dados Cluster.dat, com as curvas representado efeito de pepita variando de zero a um. Para modelo esférico, patamar um, e o passo variando de 0 a 30.

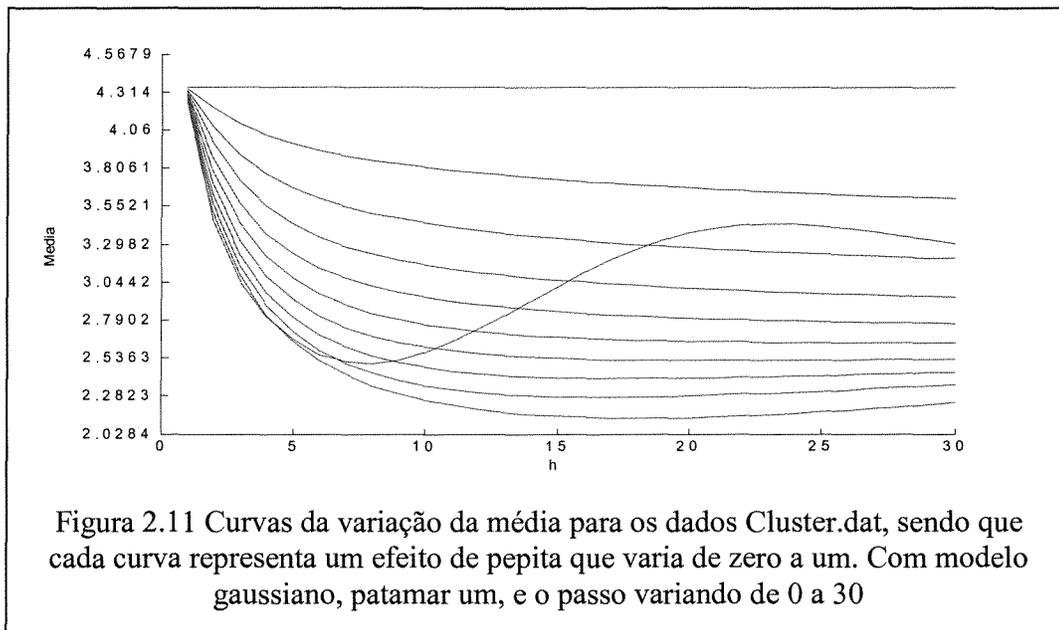


Figura 2.11 Curvas da variação da média para os dados Cluster.dat, sendo que cada curva representa um efeito de pepita que varia de zero a um. Com modelo gaussiano, patamar um, e o passo variando de 0 a 30

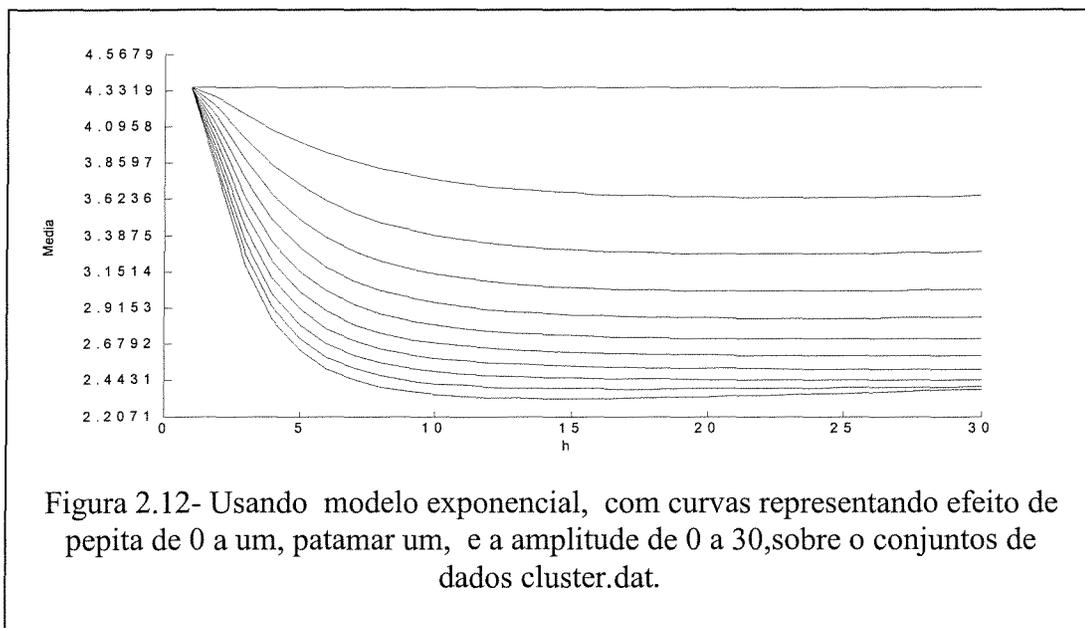
No caso da Figura 2.11 é mostrada a variabilidade utilizando um modelo gaussiano, com efeito de pepita zero ocorre um tipo de descontinuidade que deve ocorrer graças a problemas de estabilidade do modelo, devido a existência de dados que estão muito próximos entre si.

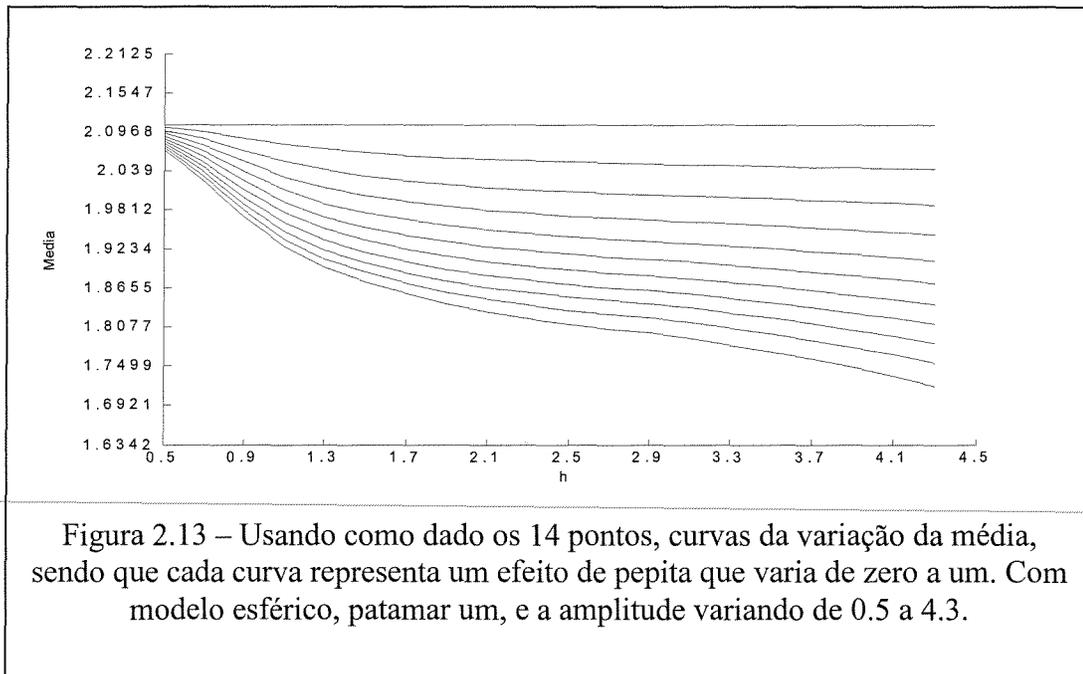
No estudo dos dados *cluster.dat*, visto acima, o valor que se espera da média é igual a 2.58, e quando estes dados foram ponderados utilizando modelos com efeitos de pepita pequeno, a média ficou abaixo do esperado, assim como quando utilizamos altos valores, a média sobe para o valor da média dos dados, já que os ponderadores vão se igualando, e se aproximando do valor $1/n$.

Passemos agora para o estudo das 14 amostras e sua população. A população da qual os dados foram retirados tem uma média de 1.74; os dados sem ponderador apresentam a média igual a 2.11, um acréscimo de 21% sobre a média real.

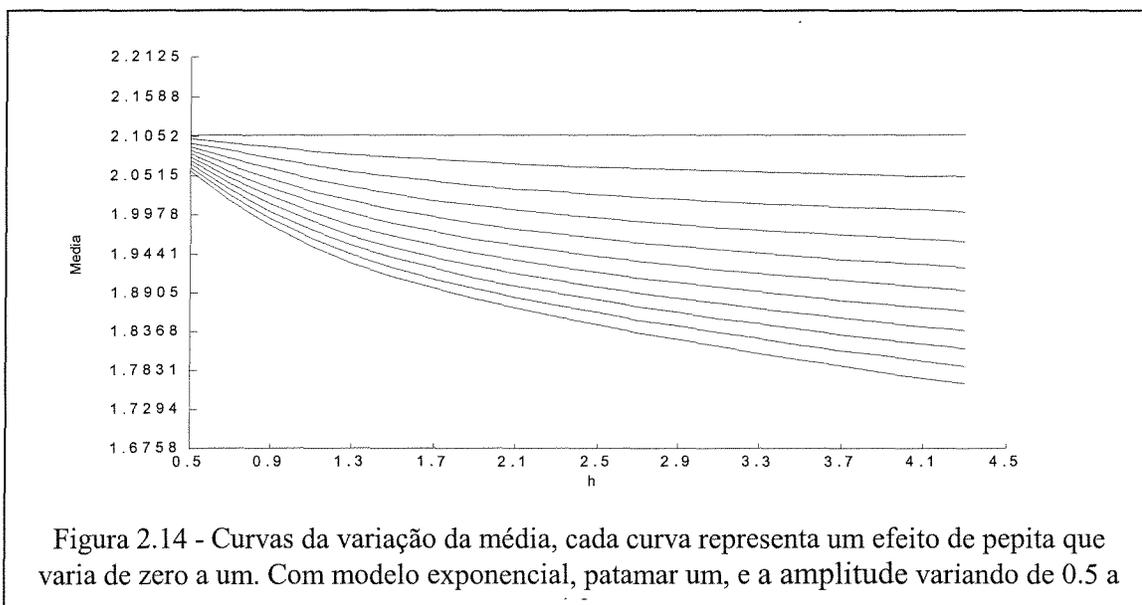
Primeiramente como no caso anterior, mostra-se a variabilidade das médias, relativas à mudança de efeito de pepita e tipo de modelo de variograma a ser utilizado.

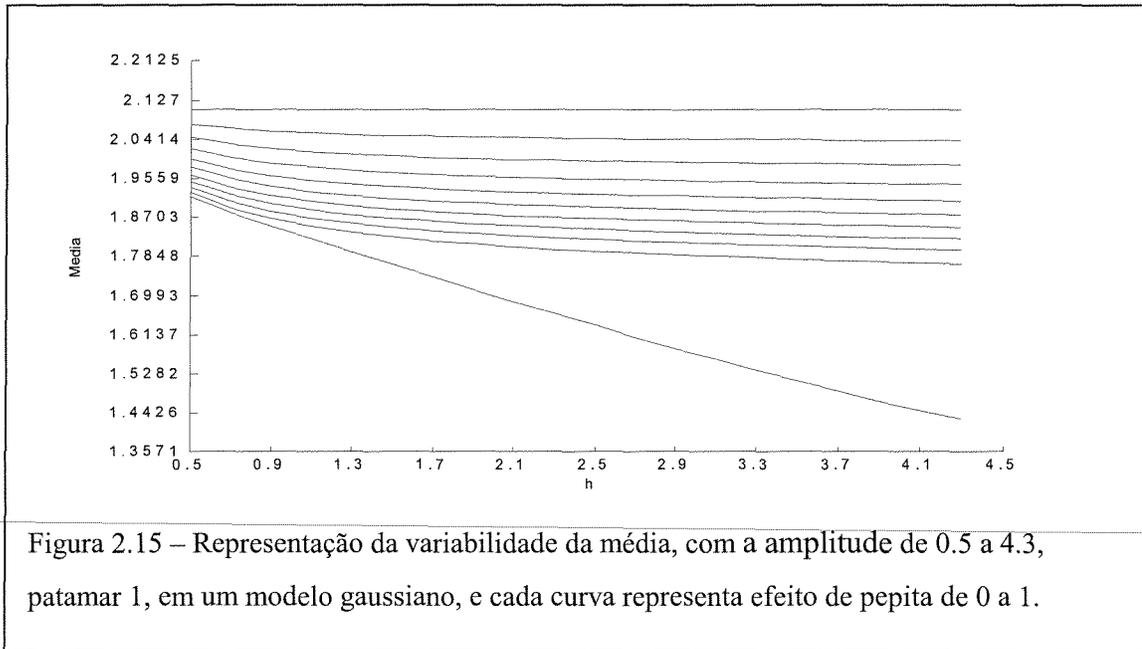
A Figura 2.13 representa a família de curvas geradas do cálculo da média do conjunto de pontos fixando o valor do efeito de pepita e variando o valor de h , para este caso utilizou-se o modelo esférico. Cada uma das curvas representa o efeito de pepita variando de zero (curva inferior) a um (reta horizontal superior). A reta da média passa a ser representada pela reta horizontal superior, quando o efeito de pepita é igual a um, pois, qualquer que seja o valor de h , o valor do ponderador para as amostras é sempre igual a $1/n$. Sendo h o valor do passo do variograma e n o número de amostras





A Figura 2.14, apresenta as mesmas características da Figura 2.13, só que agora é mostrado a variação da média quando se varia o efeito de pepita, usando um modelo exponencial.





E na Figura 2.15 com o modelo gaussiano. Neste também há uma inflexão na curva de efeito de pepita zero, tanto é que a partir de um certo valor de distância, a média se torna muito maior que a média com efeito de pepita um, subindo muito rapidamente, o que nos leva a conclusão da possibilidade de que o formato da curva mais se deve à instabilidade do modelo, que necessariamente ao conjunto dos dados.

Uma outra forma de analisar o comportamento dos ponderadores é verificar a variação dos ponderadores, quando da variação do efeito de pepita sobre as amostras.

A tabela abaixo apresenta os valores dos ponderadores das 14 amostras, com o efeito de pepita variando de zero a um, utilizando um modelo esférico.

Efeito de Pepita			0.000	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000	
x	y	elevação												
1	1.361	7.802	2.100	0.348	0.240	0.197	0.169	0.147	0.130	0.115	0.102	0.091	0.081	0.071
2	2.572	6.914	2.200	-0.046	0.025	0.041	0.049	0.055	0.059	0.063	0.066	0.068	0.071	0.071
3	3.371	6.249	2.000	-0.045	-0.033	-0.013	0.003	0.017	0.028	0.039	0.048	0.057	0.065	0.071
4	3.981	5.720	1.200	-0.132	-0.074	-0.043	-0.020	-0.002	0.014	0.028	0.041	0.052	0.063	0.071
5	5.100	8.706	2.500	0.005	0.043	0.056	0.065	0.070	0.074	0.076	0.078	0.077	0.076	0.071
6	7.674	9.410	1.200	0.232	0.182	0.161	0.146	0.135	0.125	0.115	0.105	0.096	0.084	0.071
7	9.145	7.435	1.300	0.155	0.145	0.138	0.131	0.124	0.118	0.111	0.103	0.095	0.084	0.071
8	8.149	2.029	1.200	0.331	0.198	0.161	0.139	0.123	0.111	0.101	0.092	0.085	0.077	0.071
9	7.460	2.764	1.800	-0.112	0.012	0.038	0.050	0.058	0.063	0.066	0.068	0.070	0.071	0.071
10	5.130	2.718	3.200	-0.092	-0.037	-0.014	0.001	0.014	0.025	0.034	0.043	0.052	0.061	0.071
11	4.993	1.999	3.500	-0.086	0.049	-0.045	-0.045	-0.047	-0.049	-0.052	-0.055	-0.059	-0.064	0.071
12	4.701	2.443	3.600	-0.005	-0.004	0.008	0.017	0.025	0.033	0.040	0.046	0.054	0.062	0.071
13	4.472	2.596	3.100	-0.079	-0.016	0.002	0.013	0.023	0.031	0.039	0.046	0.054	0.062	0.071
14	2.188	1.126	0.600	0.354	0.271	0.225	0.191	0.164	0.141	0.122	0.105	0.092	0.080	0.071

Para entender a tabela 2.1, na linha da amostra 1, temos x, y, o valor da cota da amostra e nos onze valores que se seguem, vemos o ponderador que é atribuído a esta amostra quando a média é mínima no intervalo de h fornecido, mantendo o mesmo modelo de variograma, e variando o efeito de pepita.

Ou graficamente, como mostra a Figura 2.16:

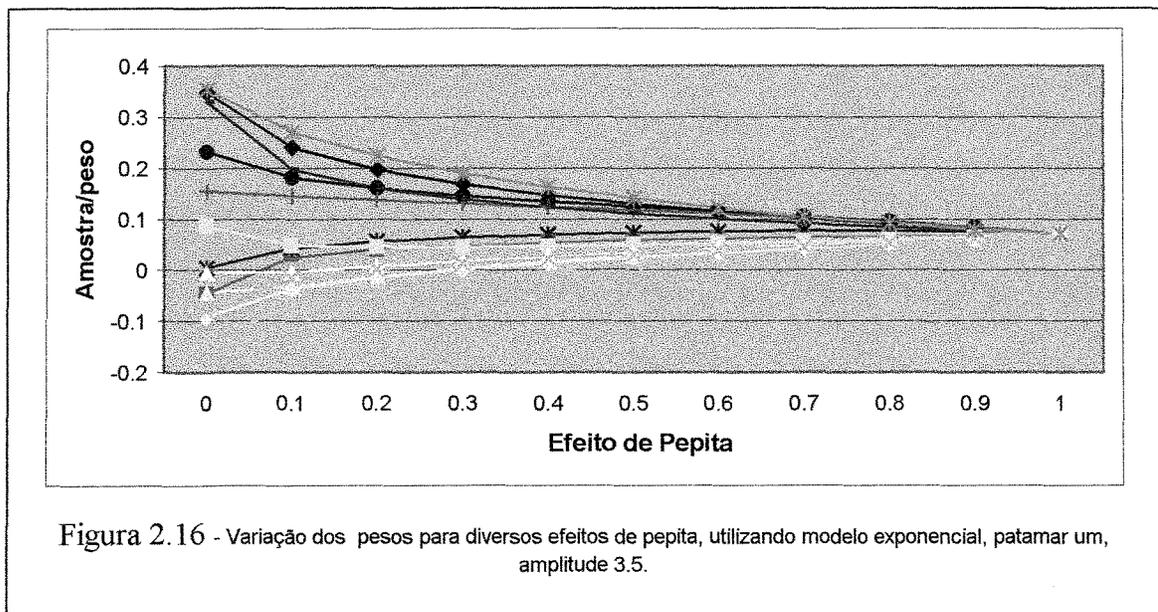
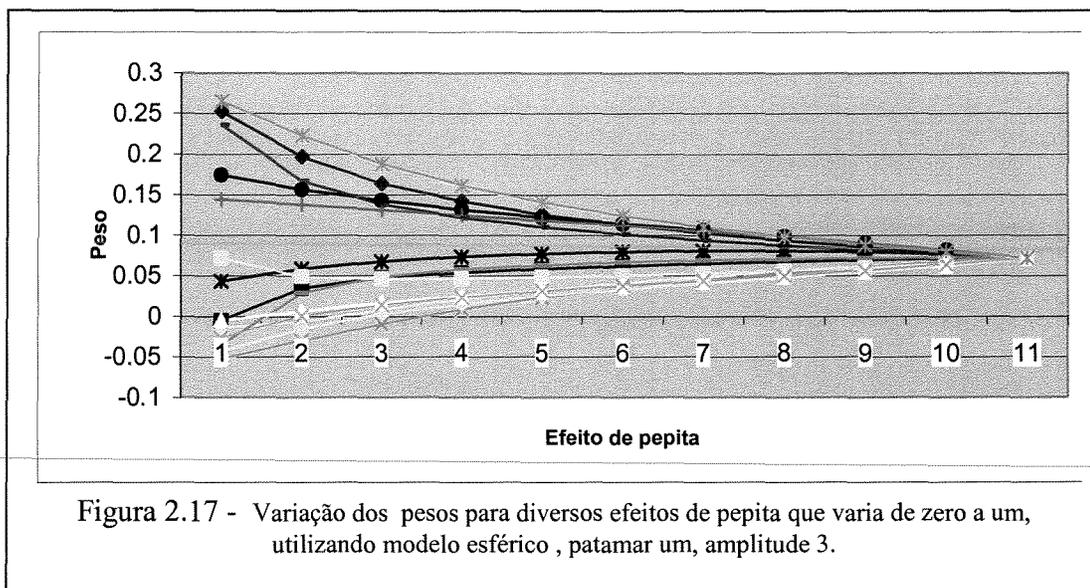


Figura 2.16 - Variação dos pesos para diversos efeitos de pepita, utilizando modelo exponencial, patamar um, amplitude 3.5.



O gráfico mostra a variação que o ponderador de cada amostra sofre, quando esta é submetida a um diferente valor do efeito de pepita. As curvas representam as 14 amostras usadas no estudo.

No caso da Figura 2.16 foi utilizado um modelo exponencial, com amplitudes variando de 0 a 30, e escolhendo neste intervalo aquela posição que nos fornecia a menor média. Analisando as Figuras 2.16 e 2.17, percebe-se que elas têm o mesmo comportamento uma em relação as outras, conforme o efeito de pepita vai variando, apesar do modelo não ser o mesmo.

2.4 Método Richmod

O método proposto por Richmod (2002), é bem intuitivo, pois para agrupar amostras que estejam próximas umas das outras, não leva em consideração, dimensões pré-estabelecidas e ângulos. A única informação que deve ser fornecida é a distância máxima **D** que um ponto deve estar do outro, para que ambos passem a formar um grupo.

Teoricamente, basta fazer em cada ponto um círculo de raio **D**, pesquisar círculos que possuam mais de um ponto interno a ele, entre todos os círculos criados. Poderão se

formar vários conjuntos que possuam mais de um ponto no seu interior, e outros tantos apenas com o ponto que deu origem ao círculo. Conta-se o número de amostras de cada conjunto formado N_i , e a cada amostra do conjunto recebe então um ponderador igual a $1/N_i$.

O interessante deste método, é que não importa a direção, tamanho, ou anisotropia do agrupamento, ele será identificado se todos os seus pontos estiverem a uma distância menor que **D**.

No caso de existir uma anisotropia de busca, pode-se efetuar operações matemáticas de modo que se deforme a geometria espacial dos dados, tornando-os isotrópicos. Estas operações se resumem, na sua essência, a apenas uma transformação de coordenadas, que pode ser realizada com uma multiplicação matricial sobre o conjunto de coordenadas.

Este método, em dados que se apresentam em conjuntos (um poço, uma amostragem em canaleta), este conjunto pode vir a se tornar um único agrupamento. Deve-se ter bastante atenção quando os dados estão nestas condições. Pois se, a distância entre duas amostras consecutivas for menor que a distância máxima D , elas iriam se agrupando, formando um único grupo.

Esta metodologia alcança 100% dos agrupamentos, o que em caso de amostragem de agrupamentos na zona rica, diminui de forma significativa a média e a variância dos dados.

Do exposto, pode-se notar que o método de Richmod é uma evolução do método Declus (anteriormente visto 2.2), visto que ele se amolda aos dados e não apresenta os inconvenientes que estão presentes neste método. Pode-se criar um conjunto de amostras, de modo que os dois métodos forneçam os mesmos resultados, isto é, os agrupamentos formados serão os mesmos em ambos os métodos. Para isto, basta que os agrupamentos sejam todos do mesmo tamanho e forma, e que sejam amostrados em pontos que levem o algoritmo do Declus a passar pela “solução ótima”, nesta configuração espacial.

No caso do método agora tratado, foi necessário o desenvolvimento de uma rotina de computador que gerasse o conjunto de ponderadores quando fornecido o conjunto de dados e a distância máxima entre as amostras.

O seguinte algoritmo foi idealizado:

1 - parte-se da premissa que cada amostra forma um único grupo G , portanto coloca-se cada amostra como sendo um grupo.

2 - cria-se um vetor V de amostras já agrupadas, colocando-se falso para todas as amostras. Então, partindo da primeira amostra, faz-se $V(1)$ igual à verdadeiro, e pesquisa-se no universo de amostras cujo valor V seja falso, procurando por amostras que estejam a uma distância menor que aquela estipulada como raio de pesquisa máximo, se alguma amostra for encontrada, coloca-se o valor de $G(1)$ em $G(j)$. Para cada amostra que é adicionada ao grupo é necessário percorrer todo o universo de amostras ainda não agrupadas, usando-a como base de pesquisa. Esta operação garante que um grupo não será dividido em dois.

3 - após todas as amostras possíveis de serem adicionadas ao grupo da amostra 1, serem identificadas como tal, inicia-se o processo novamente, tendo como semente a primeira amostra que não foi colocada no grupo da primeira amostra. Este procedimento deverá ser executado até que todas as amostras pertençam a um grupo.

O método apresenta as seguintes qualidades:

1 - quando se aumenta a distância, o grupo cresce, e um grupo definido em uma distância menor não é dividido com o aumento da distância, ocorrendo a preservação do grupo desde que este seja formado.

2 - não existe ponto de inserção, pois os grupos se originam a partir das amostras, e não de um ponto específico.

3 - não existe ângulo, pois a pesquisa é efetuada em todas as direções.

No caso de corpos que apresentem anisotropias, estas devem ser eliminadas utilizando cálculos matemáticos simples sobre as amostras.

Como inconveniente do método deve se notar que quando se aumenta em demasia a distância máxima, as amostras podem se unir em imensos grupos que tomam todo o corpo. Exemplo é o caso de poços de petróleo, no qual se as distâncias entre as amostras subsequentes no furo não forem superior à distância máxima, o poço será tomado como um único grupo.

Devido a sua grande capacidade de aglutinação dos dados, este método pode ser usado para se calcular uma média/variância mínima que o corpo pode ter, se os agrupamentos foram feitos em regiões ricas do campo de amostragem. Nos estudos que foram feitos neste trabalho, usando-se os vários métodos de desagrupamento, este é o que sempre forneceu o menor valor da média e variância para os dados em estudo. Com isto podemos usa-lo como parâmetro, dado que a população em estudo dificilmente terá uma média e variância menor que a fornecida quando se emprega este método.

Uma modificação que poderia ser introduzida ao método para minimizar esta grande capacidade de aglutinação, seria a introdução de área, que se baseia no seguinte: cada ponto teria uma área que corresponderia à área do círculo de raio D usado no algoritmo de pesquisa de vizinhança; no caso de agrupamentos seria utilizada a área formada pela união destas varias áreas.

Mostra-se, nas figuras seguintes, a formação de grupos: na Figura 2.18 utilizando-se do método proposto por Deutsch(1989), e na Figura 22.19 pelo método Richmod, utilizando a distância máxima de pesquisa 1.1.

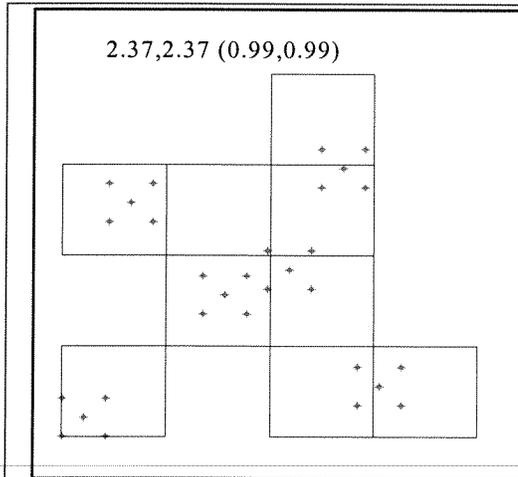


Figura 2.18 – Aplicação do método Declus sobre dados agrupados.

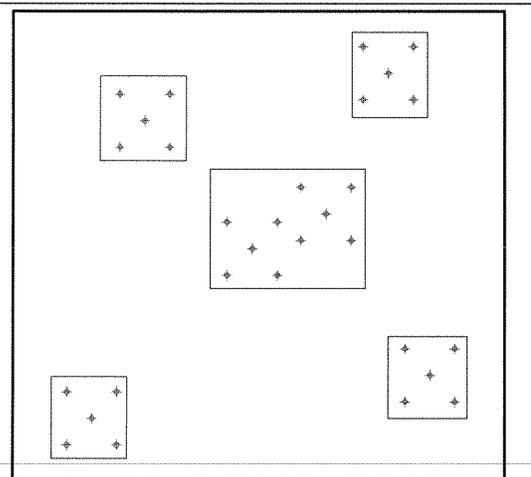


Figura 2.19 - Aplicação do método Richmod sobre dados agrupados.

Nas Figuras 2.20 e 2.21 pode-se ver o modo de trabalho dos métodos, no caso do método proposto por Deutsch (1989), a média mínima é conseguida com a configuração da Figura 2.20, e no método proposto por Richmod(2002), a configuração aparece na Figura 2.21, nota-se diferença entre os dois resultados. No primeiro, há formação de seis agrupamento com cinco amostras cada; portanto cada amostra receberá um ponderador de $1/5$. No caso do método de Richmod (2002), haverá a formação de somente quatro agrupamentos, a, b, c se reduziram a apenas um, as amostras terão neste agrupamento ponderador de $1/15$, os demais terão amostras com ponderador igual a $1/5$.

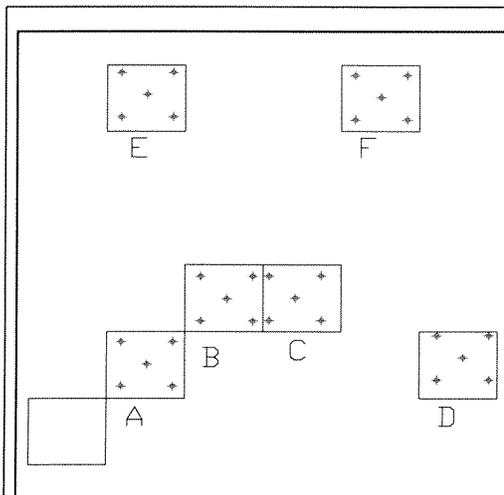


Figura 2.20 - Aplicação do método Declus sobre dados agrupados, criados para comparação dos métodos Declus e Richmod.

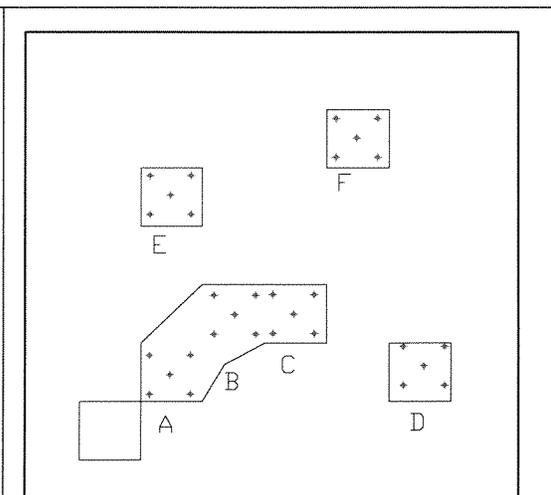


Figura 2.21 - Aplicação do método Richmod sobre dados agrupados, criados para comparação dos métodos Declus e Richmod.

No caso das Figuras 2.22 e 2.23, quando se aplica os métodos Declus e Richmod sobre este conjunto de amostras, os dois apresentaram os mesmos valores para os ponderadores.

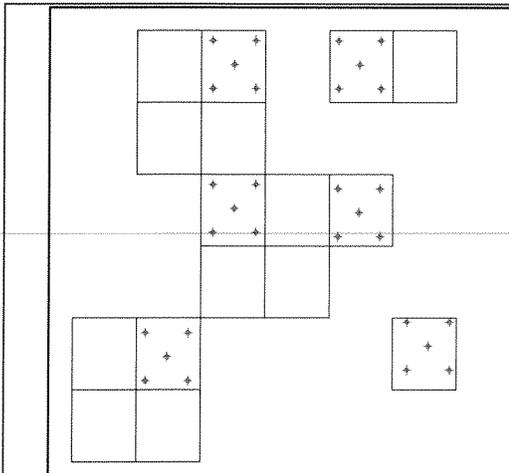


Figura 2.22 - Aplicação do método Declus sobre dados agrupados, criados para comparação dos métodos Declus e Richmod.

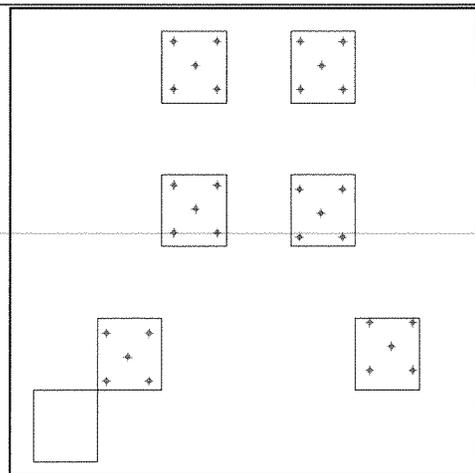


Figura 2.23 - Aplicação do método Richmod sobre dados agrupados, criados para comparação dos métodos Declus e Richmod.

Capítulo 3

Impacto dos ponderadores nas estatísticas

Neste capítulo, apresenta-se os procedimentos que devem ser adotados, para que se possa fazer uso dos ponderadores em diversos cálculos estatísticos. O uso não é de direto; é necessário que algumas modificações sejam efetuadas nas formulações, para que seja possível a introdução do conjunto de ponderadores nos cálculos.

Neste estudo, as seguintes ferramentas estatísticas serão usadas: média, variância, mediana, histograma e variograma e correlação entre duas variáveis.

Os seguintes valores serão usados para exemplificar os procedimentos que serão apresentados:

N. Amostra	1	2	3	4	5	6	7	8	9	10	11	12	13	14
valores	1	1	1	2	1	1	1	3	4	8	5	6	6	2
Pesos	0.13	0.04	0.04	0.04	0.13	0.13	0.13	0.06	0.06	0.03	0.03	0.03	0.03	0.13

Para o cálculo dos pesos foi utilizado o método de Richmod, sobre os dados apresentados na Figura 2.8.

Abaixo seguem os procedimentos utilizados para a introdução de ponderadores nas diversas rotinas de cálculo.

Média

O valor da média aritmética para um conjunto de dados z é definido como

$$\frac{\sum_{i=1}^n z_i}{n}, \quad (3.1)$$

no caso da média ponderada, este valor passa a ser

$$\frac{\sum_1^n z_i * p_i}{\sum_1^n p_i}, \quad (3.2)$$

sendo **p** o vetor que contém os ponderadores para o valor de **z** correspondente.

Deste modo, nota-se que a média de dados que possuem um fator de ponderação é simplesmente o valor da variável multiplicada pelo ponderador correspondente.

Na maneira clássica:

$$1+1+1+2+1+1+1+3+4+8+5+6+6+2 = 42 / 14 = 3$$

no caso da média ponderada seria:

$$0.13*1+0.04*1+0.04*1+0.04*2+0.13*1+0.13*1+0.13*1+0.06*3+0.06*4+0.03*8+0.03*5+0.03*6+0.03*6+0.13*2 = 2.11 / 1. = 2.11$$

Mediana

O cálculo da mediana, da maneira convencional é feito ordenando-se os dados usando-se o valor mais central desta seqüência como mediana. Agora, como o ponderador de cada amostra não é o mesmo, deve-se modificar o procedimento para este cálculo. Procede-se à ordenação dos dados como no caso anterior e depois se acumulam os ponderadores correspondentes até que se consiga o valor de 50% do total da soma dos ponderadores, neste ponto encontra-se a mediana dos dados. Nota-se que em valores absolutos, o número de dados à direita e à esquerda da mediana, dificilmente serão iguais. A soma dos ponderadores necessariamente tem que ser igual a um.

Neste caso usando os dados da Tabela 3.1 e ordenando-os com seus respectivos pesos teríamos:

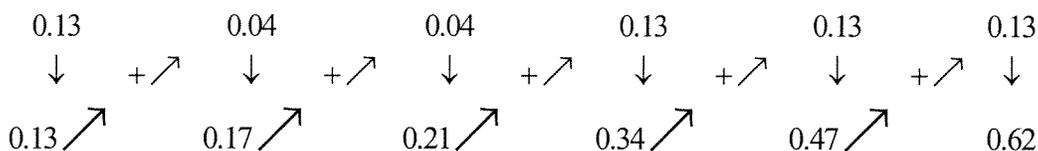
Tabela 3.2 – Tabela associando o valor do dado com o ponderador fornecido pelo método Richmod.

valores	1	1	1	1	1	1	2	2	3	4	5	6	6	8
Pesos	0.13	0.04	0.04	0.13	0.13	0.13	0.04	0.13	0.06	0.06	0.03	0.03	0.03	0.03

Da maneira clássica, a mediana seria o sexto valor mais o sétimo, divididos por dois.

$$(1+2)/2 = 1.5 \text{ (dados em destaque na tabela 3.2).}$$

se utilizarmos os ponderadores temos:



Desta maneira, tem-se o histograma acumulado crescente. Este histograma difere do clássico, pois são somados os pesos respectivos de cada valor ao invés do valor um. Ao se atingir a meta de 50%, volta-se à seqüência de dados e lê-se o valor correspondente, portanto a mediana no caso é 1.

Na Figura 3.1, que representa o histograma acumulado crescente, pode-se verificar a seqüência descrita acima. A partir da curva do histograma, partindo da probabilidade de 50%, chega-se ao valor da mediana.

Variância

O valor da variância é calculado classicamente da seguinte maneira:

$$\frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2, \tag{3.4}$$

Ao se introduzir os ponderadores para cada amostra, este cálculo deve ser modificado para que os ponderadores façam parte da formação matemática, de acordo com a equação dada abaixo:

$$s^2 = \sum_{i=1}^n p_i (z_i - \bar{z})^2 \quad (3.5)$$

ou, de maneira simplificada

$$\frac{\sum_{i=1}^n p_i * z_i^2}{\sum_{i=1}^n p_i} - \left(\frac{\sum_{i=1}^n p_i * z_i}{\sum_{i=1}^n p_i} \right)^2 \quad (3.6)$$

Usando os dados da Tabela 3.1, o valor da variância clássica é de 5.29, enquanto que se for utilizado o conjunto de ponderadores, este valor cai para 3.29.

Histograma

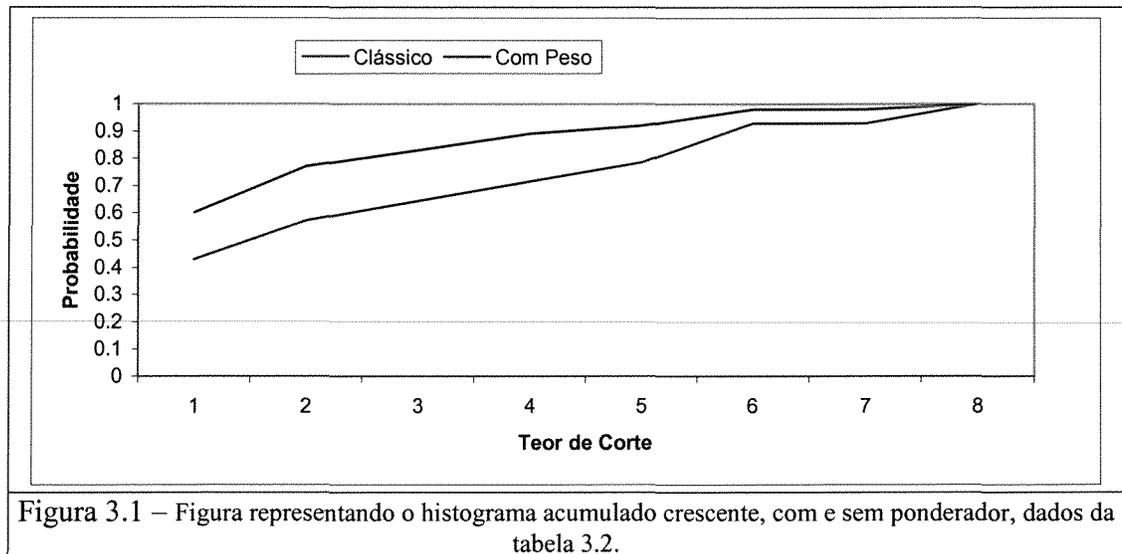
O histograma é, sem dúvida a ferramenta estatística que mais sofre influência do uso de ponderadores. A sua forma pode ser modificada substancialmente, pois com a introdução de pesos, a probabilidade da classe é bastante alterada.

Quando se faz o cálculo do histograma clássico, a cada amostra que pertence a uma classe, o valor de uma unidade é acrescido a esta.

No caso do histograma ponderado o valor a ser incrementado na classe a qual o valor da variável faz parte, deve ser acrescida do ponderador correspondente. Assim no final dos cálculos a probabilidade de cada classe no histograma absoluto que era um valor inteiro, agora não necessariamente o será.

Nota-se, com este procedimento, que não há alteração nos valores de máximo e mínimo do histograma. Todas as classes que existiam no histograma normal continuaram a existir no histograma com ponderador, apenas o valor da classe poderá mudar.

Para mostrar esta influência, mostra-se o histograma acumulado crescente, Figura 3.1.



Na Figura 3.1, observa-se que para um dado teor de corte, existe uma grande diferença quanto à probabilidade encontrada. Esta diferença é realçada nas amostras de valor de teor mais baixo, pois o agrupamento, nesta amostragem foi feito em uma região rica.

Partindo desta constatação, pode-se mostrar alguns gráficos que na mineração tem grande influência na reserva e no retorno financeiro. Vide anexo 1 para maiores informações sobre estas funções de análise.

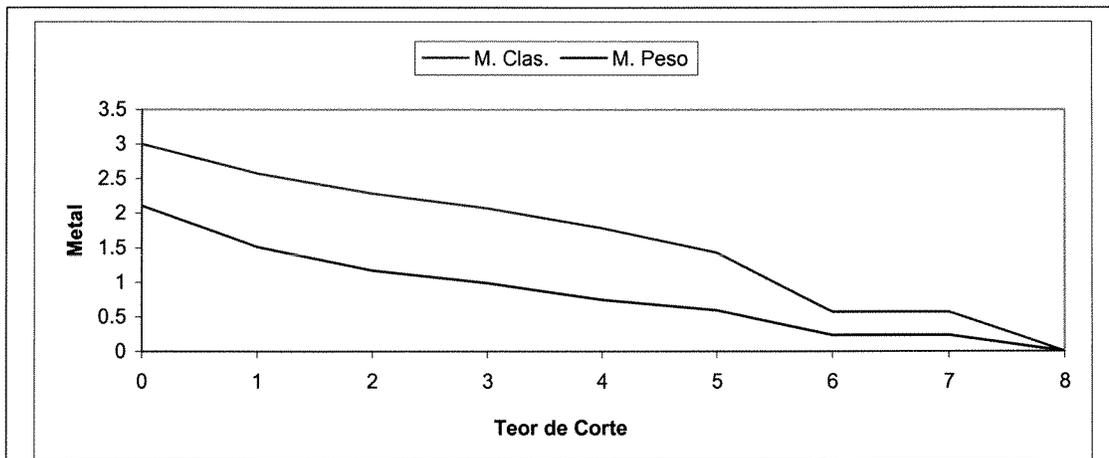


Figura 3.2 – Usando dados da Tabela 3.2, representação da curva quantidade de metal, para diversos teores de corte, em histogramas com e sem ponderador.

A Figura 3.2 mostra as curvas teor/metal contido. Examinando a figura e tomando como base um teor de corte igual a 2, teríamos uma quantidade de metal igual a 2.28, usando o histograma clássico e de 1.17 com a introdução dos ponderadores.

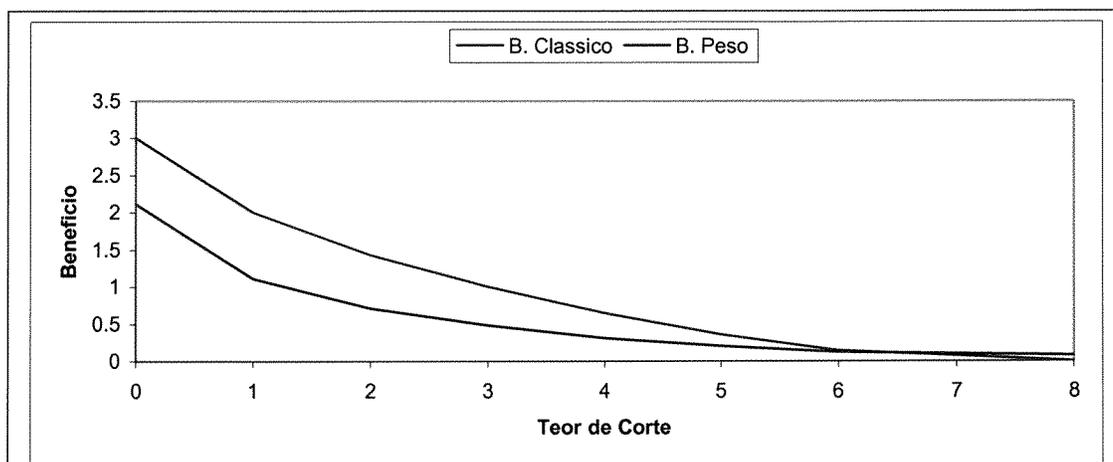


Figura 3.3 - A figura mostra a relação entre teor de corte e benefício para histogramas que usam de ponderador e sem ponderador, dados da tabela 3.2.

Na Figura 3.3, que mostra a relação entre teor de corte e benefício, observa-se novamente a grande diferença entre as curvas.

Correlação

No caso da correlação e regressão entre duas variáveis, analisadas em uma mesma amostra, existe um fato que se deve aludir. No caso dos ponderadores provirem de um cálculo utilizando a krigagem da média, pode-se ter um ponderador para cada variável, dependendo da variabilidade destas no espaço. Efetuando um estudo com duas variáveis, como uma co-krigagem, co-simulação, etc, e nos cálculos preliminares, não utilizarmos de ponderadores, e estes dados estiverem agrupados, pode-se chegar a conclusão que os dados não apresentam nenhuma correlação, o que pode ser falso, quando se fizer a ponderação do dado.

No caso da utilização de um dos outros métodos de cálculo de ponderador, não há modificação no cálculo da correlação.

Variograma

A idéia de variograma com a utilização de ponderadores foi descrita por Jacques Rivoirard (2000), que desenvolveu a sua formulação matemática, bem como a técnica para o cálculo do variograma com ponderador.

Seja $z(x)$ uma variável regionalizada, a qual é amostrada nos pontos x_i ($i=1, 2, 3, \dots, n$) dentro do domínio e é representada pela função intrínseca $Z(x)$ com variograma $\gamma(h)$, classicamente descrito como:

$$\gamma^*(h) = 0.5 \frac{1}{N(h)} \sum_{i=1} [Z(x_i) - Z(x_i + h)]^2 ; \quad (3.7)$$

onde a cada par de pontos (x_i, x_j) formado dentro de uma classe de distâncias, atribuímos um mesmo ponderador (igual a 1), dentro desta classe de distância.

No variograma convencional o ponderador p_i de cada ponto é fixo e tem valor 1 (um), assim para cada par de pontos é acrescido o valor da diferença elevada ao quadrado; ao final, este valor deve ser dividido pelo número de pares de pontos formados

multiplicado por dois. Com a introdução do ponderador sobre o conjunto de dados, a cada par de pontos deve ser acrescido ao somatório, o valor da diferença ao quadrado, multiplicado pelo produto dos ponderadores relativos a cada ponto que fazem parte do par formado. Neste caso, o valor da somatória deverá ser dividido pela somatória da multiplicação dos dois ponderadores relativos ao par de ponto formado multiplicado por dois

Em um universo de dados, em que a todo ponto existe um ponderador, cada amostra x_i possui um ponderador p_i , cada par de pontos (x_i, x_j) recebe um ponderador igual ao produto de seus ponderadores $P_{ij} = p_i \cdot p_j$. Assim o variograma passa a ser descrito da seguinte maneira:

$$\frac{1}{2 \sum (p_i \cdot p_{i+h})} \sum_{i=1}^n [z(x_i + h) - z(x_i)]^2 \cdot (p_i \cdot p_{i+h}) \quad (3.8)$$

Nota-se que o variograma na origem, isto é o valor do variograma para os passos, menores que o agrupamento, não apresenta diferença entre os valores do variograma clássico e ponderado, notadamente, quando nenhum par de pontos é formado entre um valor interno a um agrupamento e a um ponto externo a este.

Resumo

Analisando os resultados obtidos dos dados da tabela 3.1

Tabela 3.3 – Resumo das estatísticas sobre dados da tabela 3.2		
	Clássico	Ponderado
Moda	1	1
Mediana	2	1
Média	3	2.14
Variância	5.29	3.29

Nota-se , Tabela 3.3, que o uso de ponderados, nos cálculos estatísticos tem grande influência, para dados agrupados.

Nos histogramas como mostrado anteriormente, este artifício também modifica os resultados. Assim, numa tomada de decisão, levarmos em conta os resultados obtidos, na estatística, chega-se à conclusão que o uso de ponderados é de extrema importância em estudos que apresentam as características de dados agrupados.

Capítulo 4

Estudos de caso

Neste capítulo são apresentados estudos utilizando-se as técnicas descritas de desagrupamento sobre sete conjuntos de dados. Destas populações, cinco delas foram simuladas e amostradas especificamente para este trabalho.

A amostragem nestas populações foi feita de modo que houvesse uma maior concentração de dados nas regiões ricas, mas para que o viés não fosse muito grande, utilizou-se da seguinte metodologia: a partir do histograma da imagem, é retirada uma quantidade de amostras que honra as estatísticas originais, isto é, a distribuição dos dados tem a mesma forma e aparência que a original; após esta etapa, é feita uma nova amostragem, e para tal, utiliza-se do histograma, selecionando as classes mais ricas e sorteando uma amostra pertencente a ela; então, resgatam-se quatro dos seus vizinhos na imagem original. Deste modo, tem-se um falso enriquecimento da população.

4.1 Apresentação das populações

A primeira população estudada, é o conjunto de dados apresentado por Deutsch(1989), conhecido como *cluster.dat*. Esta população consiste de 2500 dados dos quais foram amostrados 140.

A segunda população é o conjunto de dados apresentados em Isaaks & Srivastava(1989), conhecido com Walker Lake. É aqui utilizada a variável V, que consiste de uma população de 78000 pontos, dos quais são amostrados 470 pontos.

As outras cinco populações simuladas, foram geradas por meio de simulações seqüências gaussianas não condicionais, com as características apresentadas na Tabela 4.1

Tabela 4.1 Dados sobre as cinco populações simuladas, utilizando simulação gaussiana não condicional no programa Isatis, A rotação foi especificada usando-se o referencial matemático. As anisotropias são relativas aos eixos x, y e z respectivamente

Imagem	modelo	Efeito pepita	amplitude	patamar	Anisotropia	Rotação
1-	Exponencial	0.2	10	1		
2-	Esférico	0.3	16	1		
3-	Gaussiano	0.2	10	1	10/25/10m	
4-	Cúbico	0.0	10	1	10/15/10m	
	Esférico		9.3	1		
5-	Sincardinal	0.0	20	1	20/30/20m	76°
	Esférico	0.0	10	1		

Dependendo dos ponderadores atribuídos, várias serão os resultados das estatísticas que obteremos com o mesmo conjunto de amostras, por isto um método quando aplicado a um conjunto de dados pode-nos levar a um resultado ótimo, mas no conjunto de resultados este método pode apresentar-se não confiável. Nos casos em estudo, o conhecimento das populações reais, nos informa qual é o resultado que deveríamos encontrar.

Abaixo são mostradas as estatísticas das populações e das amostragens relativas a estas populações e seus histogramas, com diferentes tipos de ponderadores.

Tabela 4.2 – dados relativos a população cluster .dat comparando a população real e a amostrada

	N. de dados	média	Mediana	Desvio Padrão
População	2500	2.580	0.960	5.151
Sem ponderador	140	4.350	2.195	6.703
Polígono de Infl.		2.812	1.270	5.121
Declus		2.501	1.10	4.615
KM		2.753	1.210	4.614
Richmod		1.844	0.990	3.248

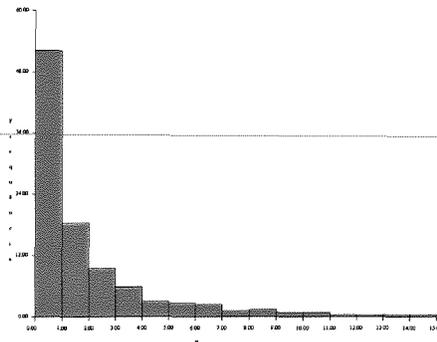


Figura 4.1 – Histograma da população de 2500, do arquivo de dados true.dat.

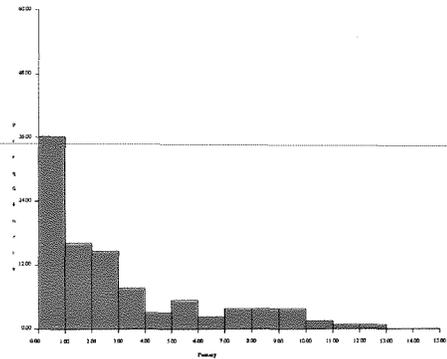


Figura 4.2 – Histograma da população de 140 dados de nome cluster.dat

Tabela 4.3 – Apresenta os valores estatísticos dos dados de Walker Lake, população amostrada com 470 dados e dos 78000 dados.

	N. de dados	média	Mediana	Desvio Padrão
População	78000	277.979	221.215	249.834
Sem ponderador	470	436.457	425.25	299.85
Declus		289.791	234.400	254.137
Km		272.847	224.400	239.980
Polígono		349.759	308.800	282.744
Richmod		274.302	226.100	228.196

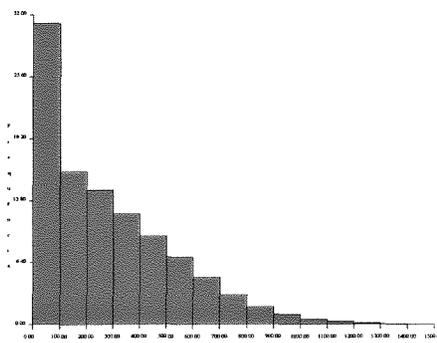


Figura 4.3 – Histograma da população de Walker lake com 78000

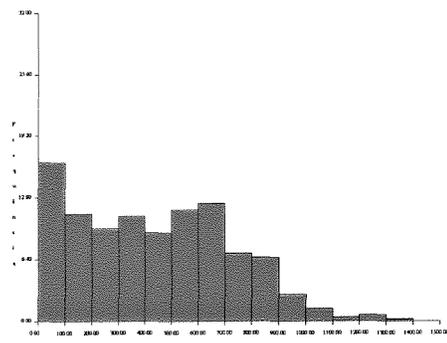


Figura 4.4 – Histograma da de Walker lake, com 470 dados.

Tabela 4.4 – dados relativos a Imagem simulada 1, apresentando os dados da população completa e do conjunto de 228 amostras que foram retiradas.

	N. de dados	média	Mediana	Desvio Padrão
População	4000	20.358	16.810	15.889
Sem ponderador	228	25.571	22.015	18.095
Declus		20.194	17.790	16.125
Km		20.455	17.810	16.295
Polígono		20.743	18.060	16.363
Richmod		18.369	16.800	14.952

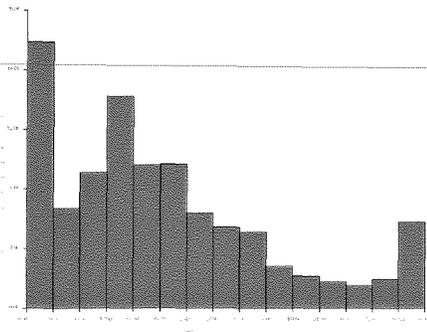


Figura 4.5 – Histograma sobre a população de 4000 dados.

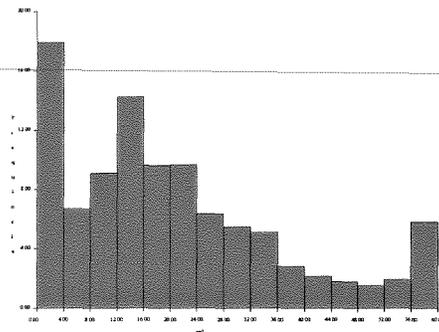


Figura 4.6 – Histograma sobre a população de 228 dados.

Tabela 4.5 – Apresenta as estatísticas do conjunto de dados relativos a imagem simulada 2. imagem completa 4000 pontos e os dados dela amostrados num total de 229 amostras.

	N. de dados	média	Mediana	Desvio Padrão
População	4000	88.468	75.755	51.703
Sem ponderador	229	105.394	125.530	51.352
Declus		90.578	91.040	49.847
Km		86.457	77.040	47.817
polígono		92.261	96.030	49.615
Richmod		84.339	75.560	48.820

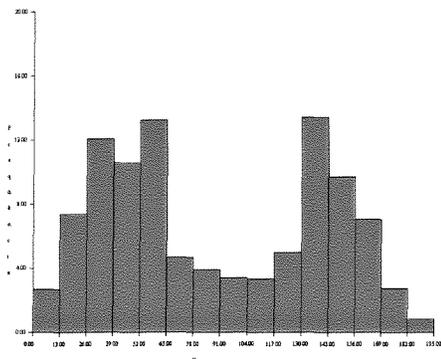


Figura 4.7 – Histograma sobre o conjunto de dados da população de 4000 pontos relativos a imagem 2.

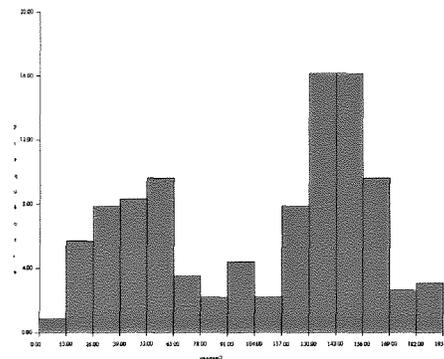


Figura 4.8- Histograma sobre o conjunto de 229 dados, retirados da população da imagem 2.

Tabela 4.6 – Conjunto de dados estatísticos, relativos a população da imagem 3 de 4000 pontos e a sua amostragem de 221 dados.

	N. de dados	média	Mediana	Desvio Padrão
População	4000	92.739	85.540	57.601
Sem ponderador	221	117.025	132.670	54.296
Declus		93.953	86.230	54.329
Km		92.092	89.430	55.578
Polígono		99.712	105.090	55.320
Richmod		84.703	79.740	52.831

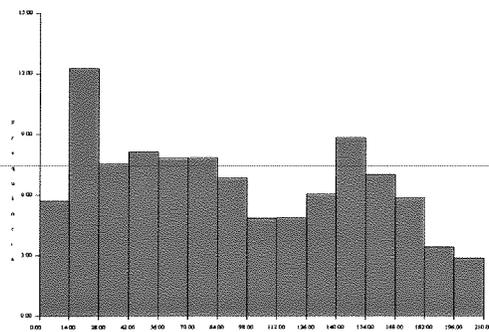


Figura 4.9 – Histograma da população da imagem 3 de 4000 dados.

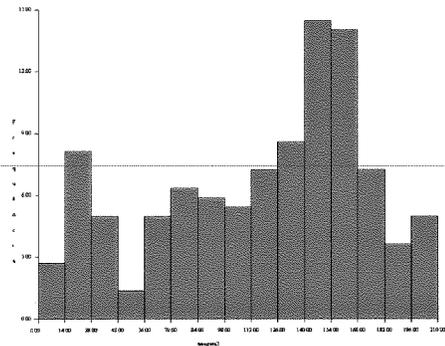


Figura 4.10 – Histograma do conjunto de dados amostrados da população da imagem 3, 221 dados.

Tabela 4.7 – Apresenta os valores estatísticos dos dados da Imagem 4, população amostrada com 218 dados e dos 4000 dados imagem onde foi feita a amostragem.

	N. de dados	média	Mediana	Desvio Padrão
População	4000	2553.288	2298.26	1898.28
Sem ponderador	218	3646.556	4562.84	1988.175
Declus		2807.507	2907.390	1989.897
Km		2755.115	2829.070	1987.792
Polígono		2974.837	3123.910	2006.081
Richmod		2397.043	1999.340	1851.277

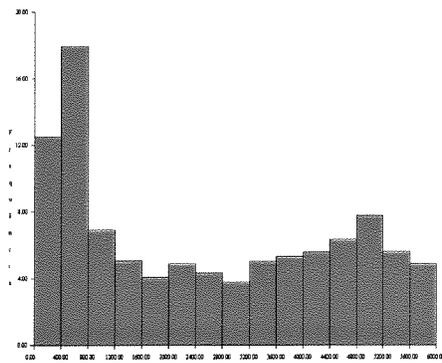


Figura 4.11 - Histograma da população da imagem 3 de 4000 dados.

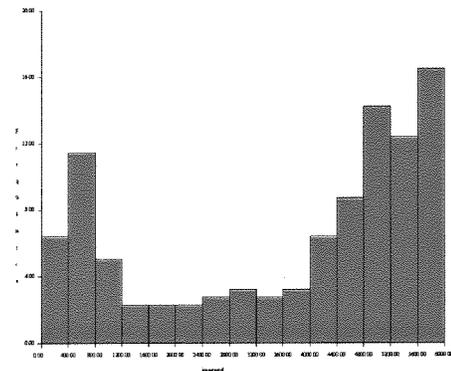


Figura 4.12 – Histograma do conjunto de dados amostrados da população da imagem 4, 218 dados

Tabela 4.8 – Apresenta as estatísticas do conjunto de dados relativos a imagem simulada 5. imagem completa 4000 pontos e os dados dela amostrados num total de 221 amostras.

	N. de dados	média	Mediana	Desvio Padrão
População	4000	749.454	823.045	503.778
Sem ponderador	221	986.82	1188.59	493.55
Declus		779.334	784.490	502.063
Km		741.672	765.780	487.025
Polígono		815.682	904.040	504.437
Richmod		688.386	711.070	474.951

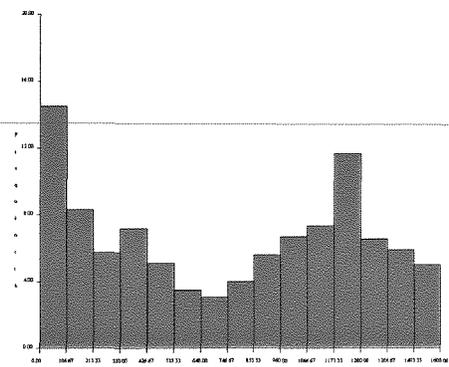


Figura 4.13 – Histograma sobre a população de 4000 dados.

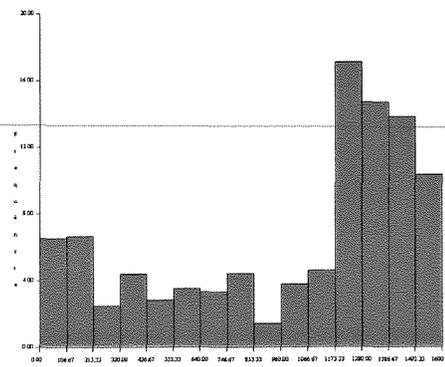


Figura 4.14 - - Histograma sobre a população de 221 dados.

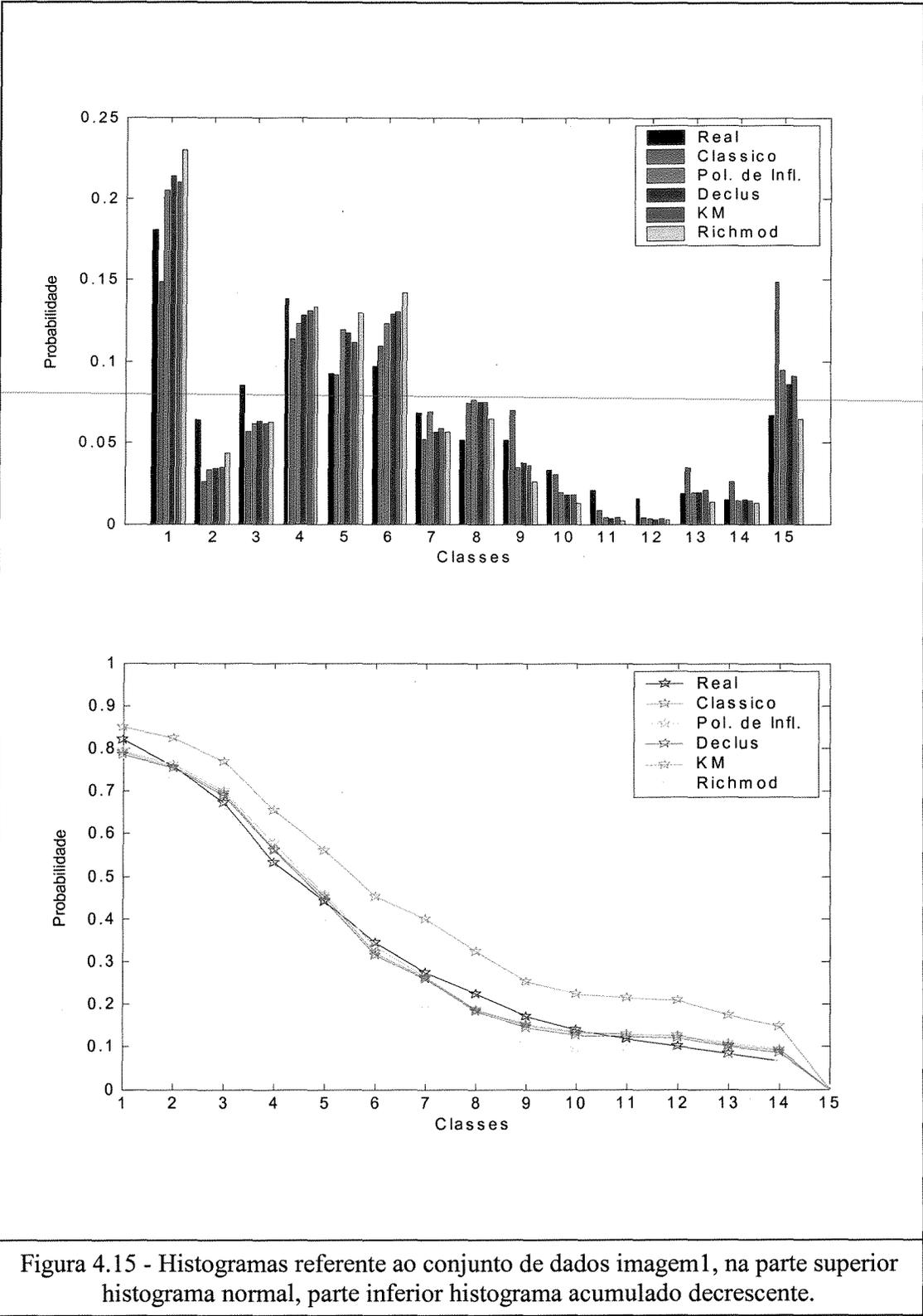


Figura 4.15 - Histogramas referente ao conjunto de dados imagem1, na parte superior histograma normal, parte inferior histograma acumulado decrescente.

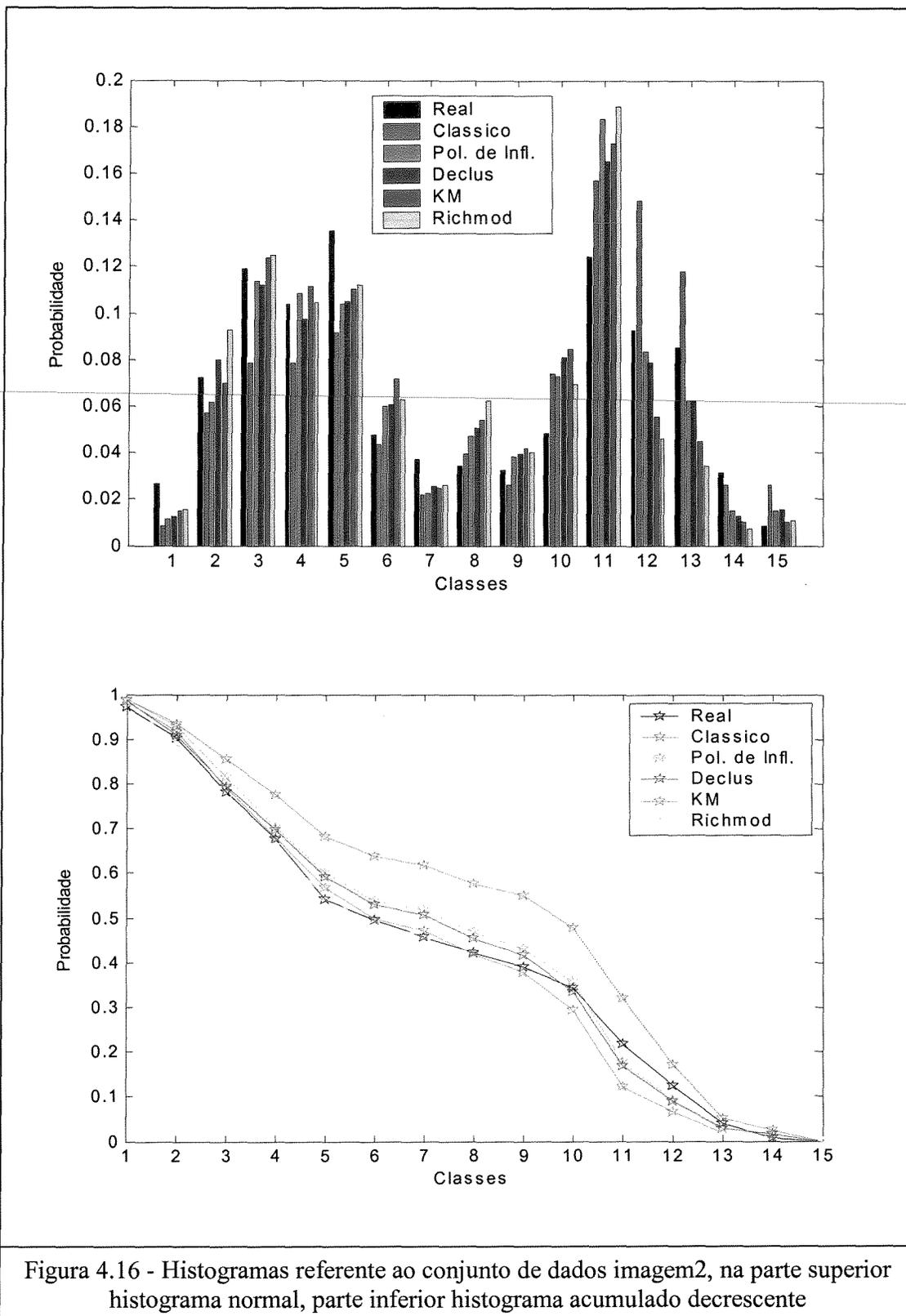


Figura 4.16 - Histogramas referente ao conjunto de dados imagem2, na parte superior histograma normal, parte inferior histograma acumulado decrescente

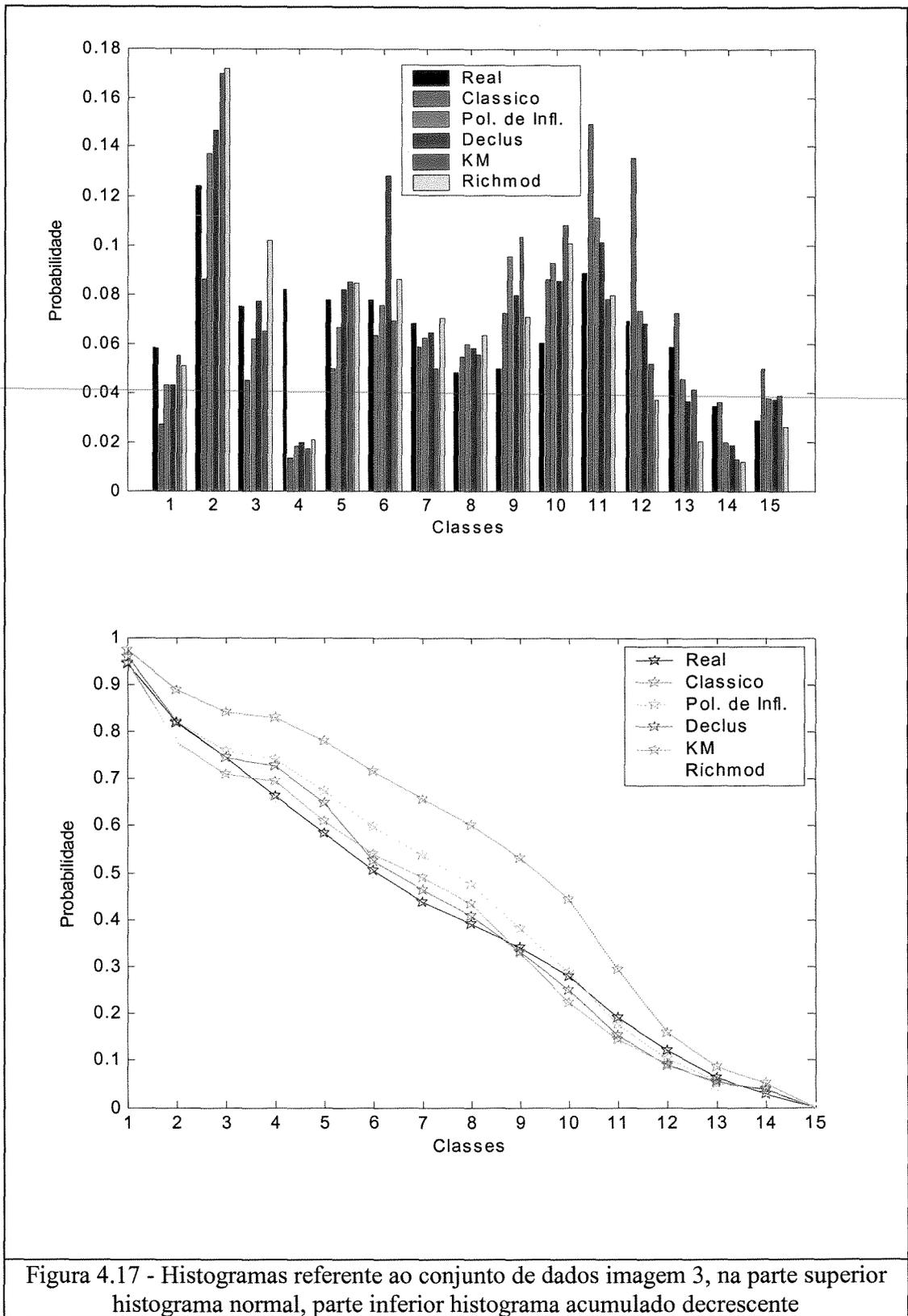


Figura 4.17 - Histogramas referente ao conjunto de dados imagem 3, na parte superior histograma normal, parte inferior histograma acumulado decrescente

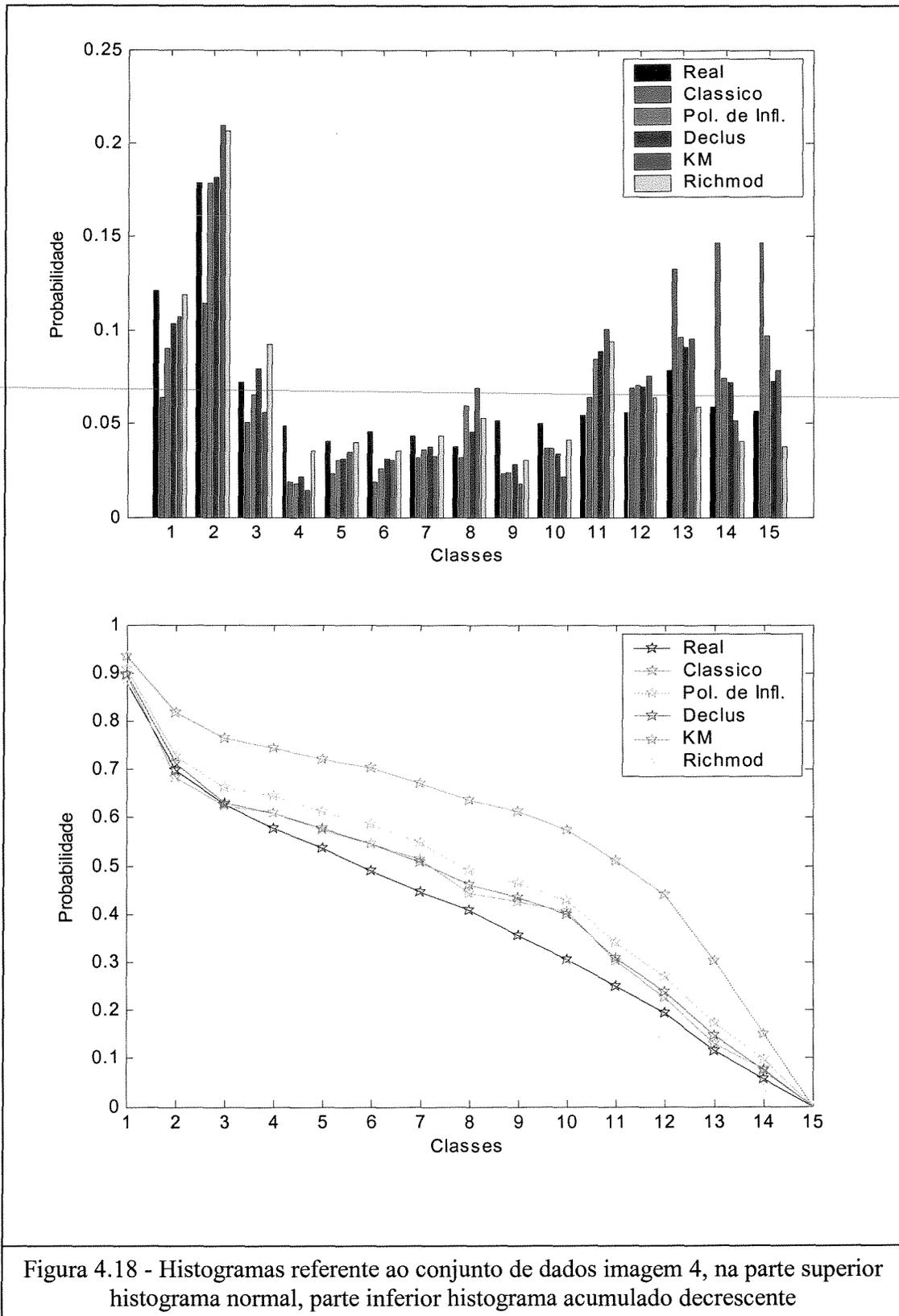


Figura 4.18 - Histogramas referente ao conjunto de dados imagem 4, na parte superior histograma normal, parte inferior histograma acumulado decrescente

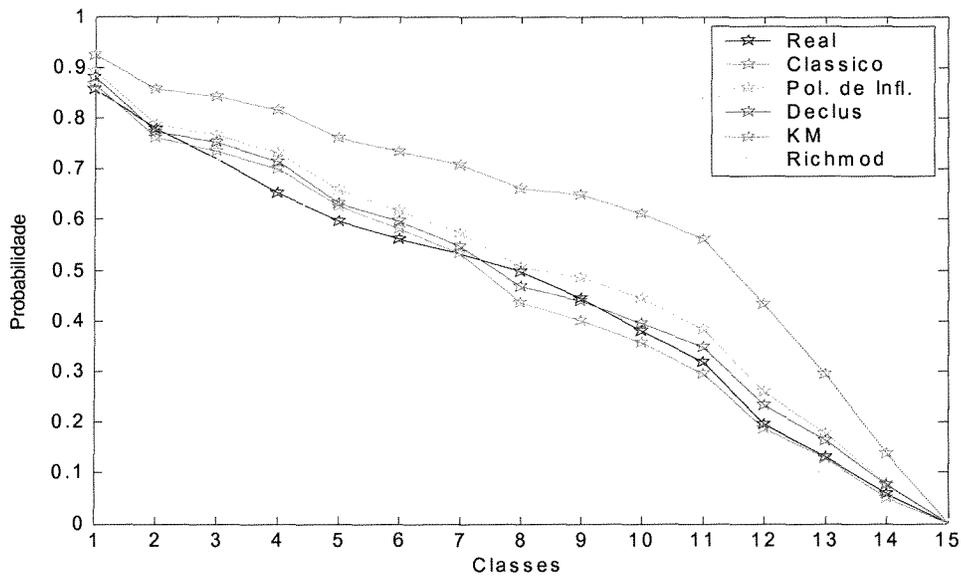
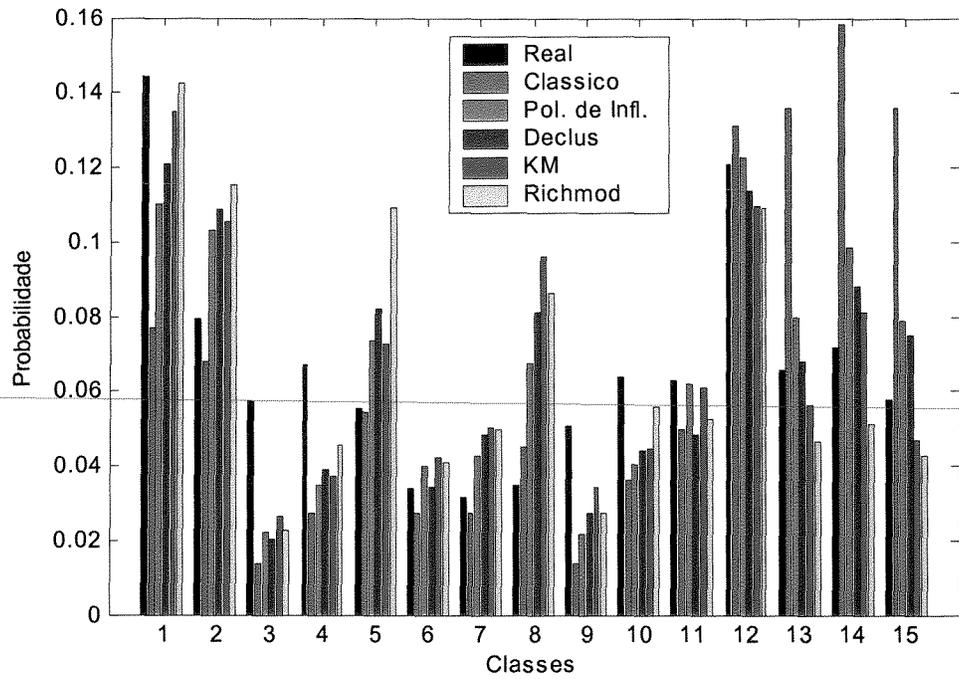


Figura 4.19 - Histogramas referente ao conjunto de dados imagem 5, na parte superior histograma normal, parte inferior histograma acumulado decrescente

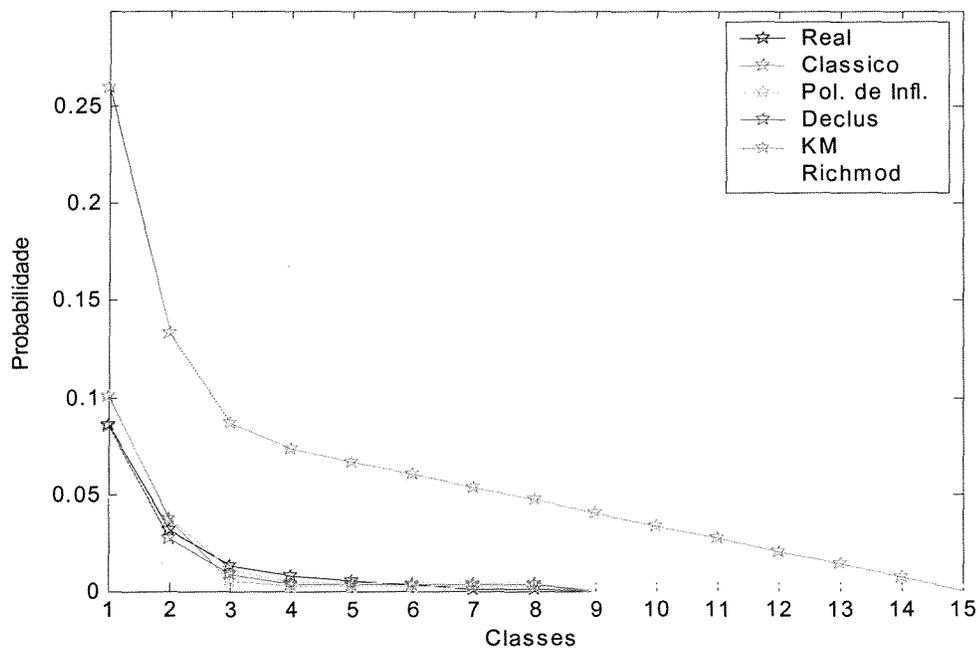
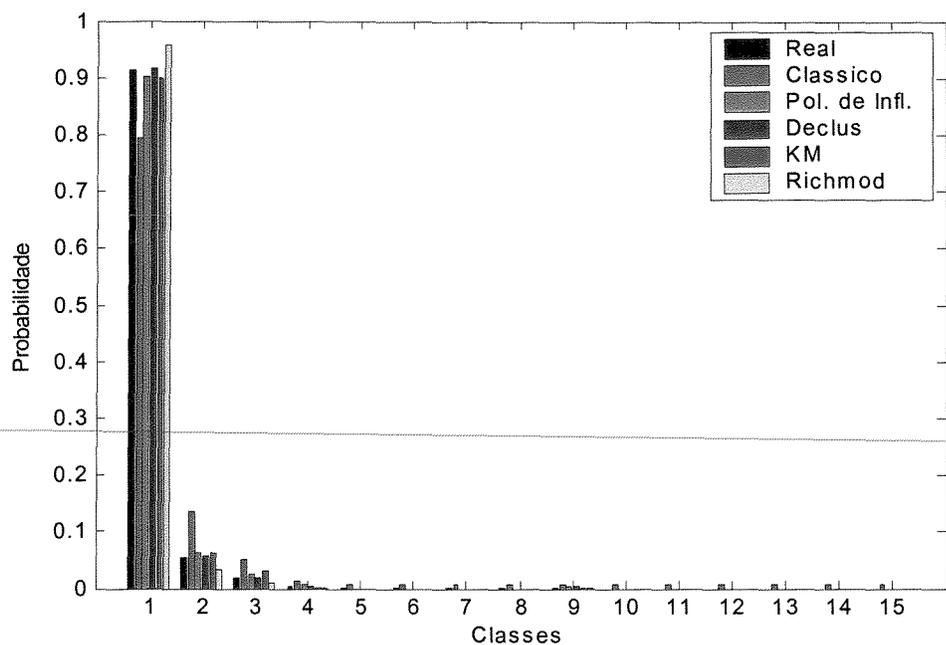


Figura 4.20 - Histogramas referente ao conjunto de dados cluster.dat, na parte superior histograma normal, parte inferior histograma acumulado decrescente

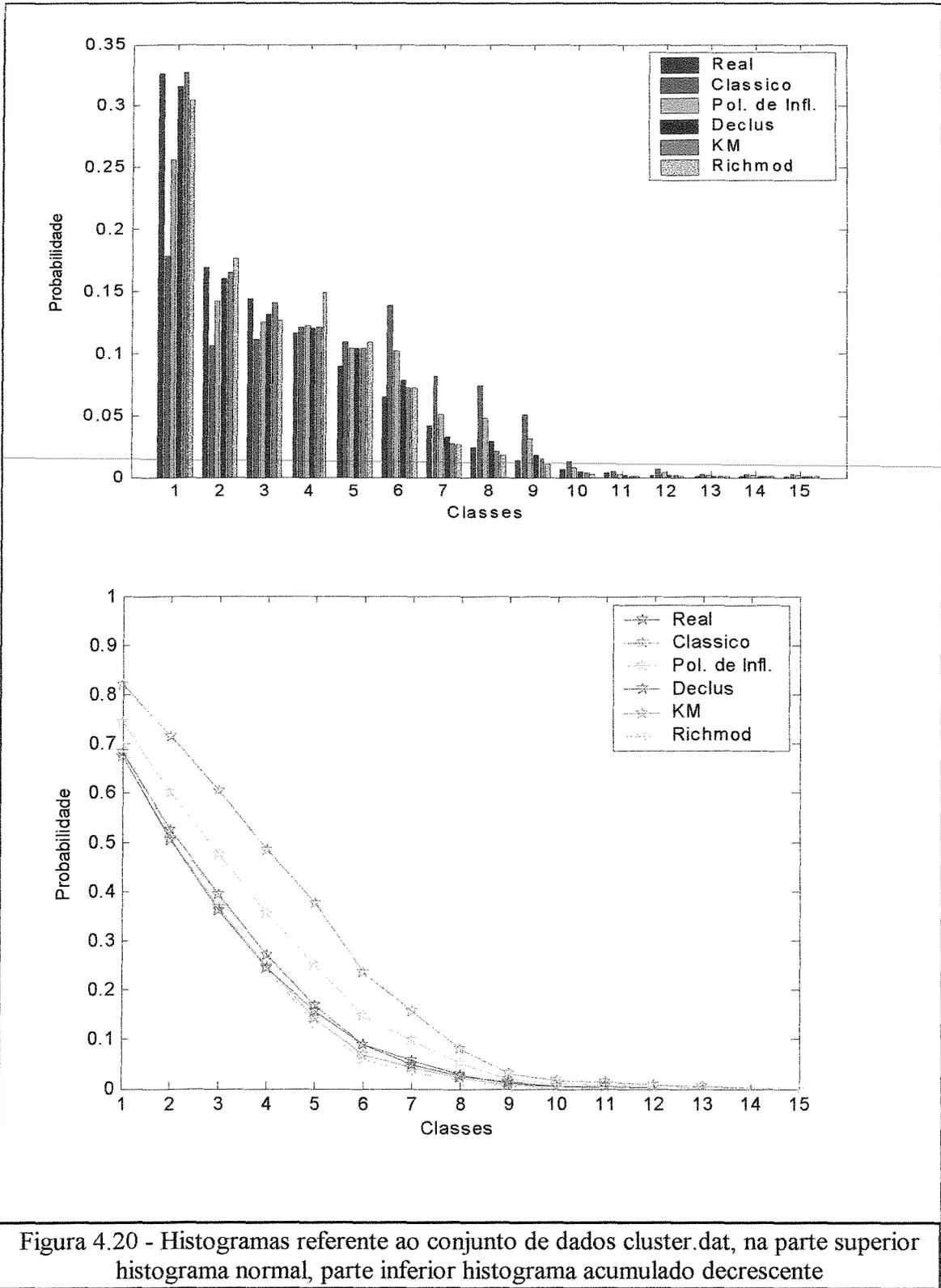


Figura 4.20 - Histogramas referente ao conjunto de dados cluster.dat, na parte superior histograma normal, parte inferior histograma acumulado decrescente

A Tabela 4.8 representa as médias dos erros ocorridos nas classes da metade superior do histograma para as várias populações e para os diferentes tipos ponderadores que foram aplicados sobre os dados.

Na primeira linha de dados, tem-se os resultados das médias das diferenças entre o histograma da população real com o histograma clássico. As demais linhas têm as médias das diferenças dos histogramas ponderados pelos diversos métodos.

Tabela 4.8 – A tabela a seguir mostra as diferenças das áreas da parte superior dos histogramas das sete populações estudadas, com diferentes ponderadores, quando comparadas com a população original

	imagem1	imagem2	imagem3	imagem4	imagem5	cluster	Walker Leak
Clássico	-121.672	-159.227	-232.702	-217.485	-192.257	-4.8254	-161.75608
Pol._Infl.	7.95556	-50.2444	-103.211	-100.856	-61.4889	-1.22222	-63.83362
Declus	31.36667	-40.5333	-49.3111	-58.6667	-38.6889	-0.25556	1.64416
Km	24.34444	-4.58889	-37.2444	-60.5667	-23.5111	0.51111	22.9886
Richmod	97.27778	11.05556	28	28.16667	42.11111	1.88889	35.21083

O ideal seria que os valores fossem próximos de zero. Não foram utilizadas todas as classes do histograma, pois em todas as populações amostradas foram feitas re-amostragens para nas regiões mais ricas da população, o que acarreta um maior número de dados nas classes superiores. Levando em conta que a área de um histograma é necessariamente igual a um, usou-se as classes superiores para diminuir o efeito de compensação que as inferiores teriam sobre a outra parte. Se o histograma fosse totalmente reproduzido, em todas as classes a diferença seria zero, assim a média seria também zero.

Nos cálculos das diferenças utilizou-se o valor da classe real menos o valor da classe do histograma a ser comparado; um valor médio maior que zero, indica que o histograma analisado tem em média valores inferiores ao do real, e ao contrário, valores negativos, implicam em um valor maior para o histograma analisado; isto é de fácil visualização quando atentamos para o histograma clássico que possui um valor negativo bastante expressivo. No caso do histograma ponderado pelos valores fornecidos pelo método de Richmod (2002), acontece o contrário; este método tem a característica de

sub-estimar os resultados estatísticos, que ocorre também no caso do histograma, pode-se dizer que no caso de um estudo sobre um conjunto de dados, o resultado apresentado por este método seria o mais conservador dos métodos. Vê se que em todas as populações estudadas ele sempre esteve abaixo dos dados reais.

4.2 Estudo Sobre o Variograma

Este tópico tem como objetivo mostrar o impacto sobre o variograma, quando seu cálculo faz-se com uso de ponderadores, bem como a influência que o ajuste acarreta nos cálculos finais de krigagem.

O primeiro impacto que se nota, decorrente da diminuição da variância, que reflete na queda do patamar, quando este existir.

Comportamento na origem: se os agrupamentos tiverem amostras que formem par com amostras não pertencentes a ele, esta operação acarretará um menor efeito de pepita

Forma serrilhada: este fator serrilhado do variograma se dá quando uma amostra rica faz par com uma pobre, para um determinado passo, e no passo seguinte estas combinações entre amostras de alto e baixo valor deixam de existir. Os agrupamentos não resolvem este problema completamente, pois se amostras não agrupadas tiverem valores ricos e pobres, ou se dentro do agrupamento houverem amostras ricas e pobres. Neste caso, os pares de pontos continuarão a apresentar uma grande diferença entre si.

No caso dos variogramas com dados agrupados, existe a possibilidade de se diminuir este comportamento, pois se os valores ricos estiverem dentro do agrupamento, e estes dados fizerem par com um ponto fora do agrupamento e a diferença entre os valores que estão sendo analisados for grande, o ponderador da amostra agrupada diminui o fator com que este par entra na soma do passo.

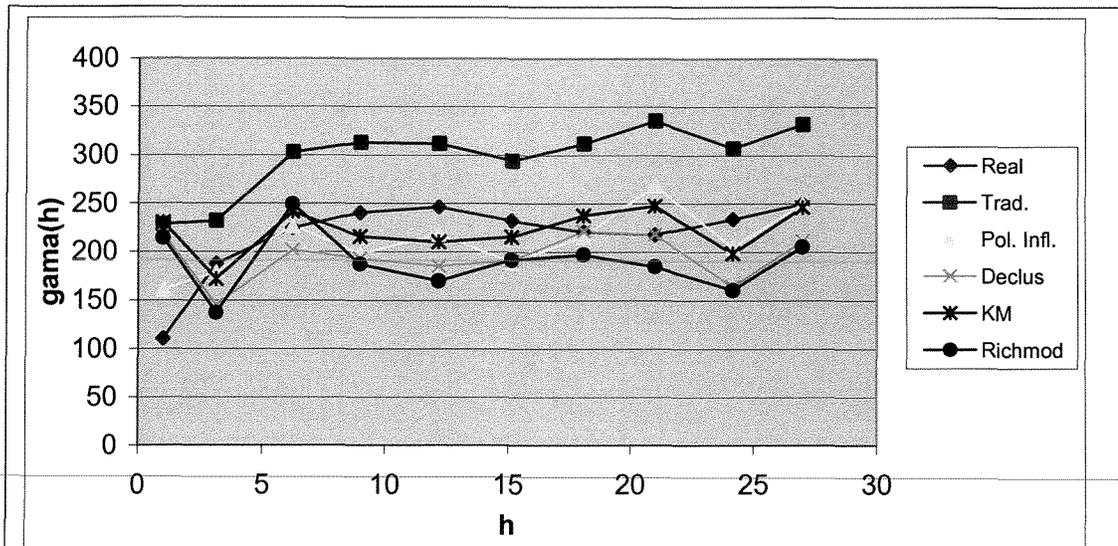


Figura 4.21- mostra o variograma dos dados da imagem 1 na direção azimute 90°, A curva marrom é o variograma sobre a imagem original

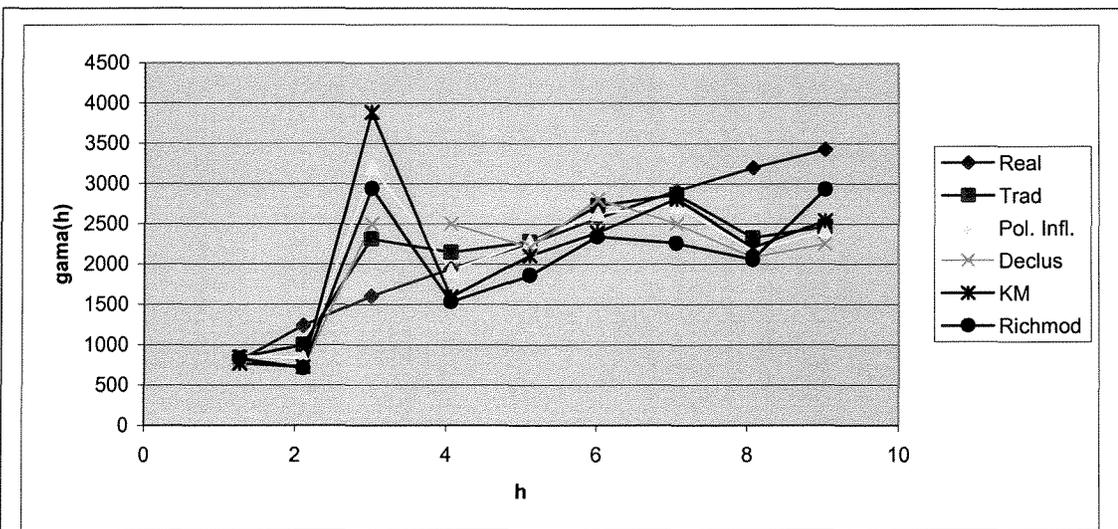


Figura 4.22 - Variogramas da imagem 3 na direção azimute 90°, mostrando os os variogramas experimentais, ponderados pelos quatro métodos, mais o sem ponderação e também o variograma sobre a imagem completa.

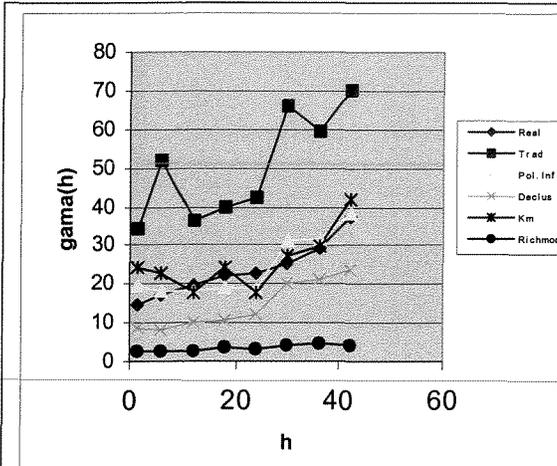


Figura 4.23- Variogramas do o conjunto de dados declus na direção azimute 30°

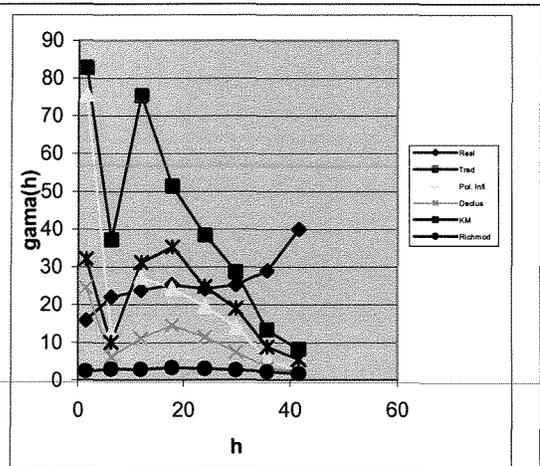


Figura 4.24 - Variogramas do o conjunto de dados declus na direção azimute 120°

Em ambos os casos, aparecem os variogramas ponderados, e também o variograma sem ponderador (trad) e o variograma sobre o conjunto de dados exhaustivos (real).

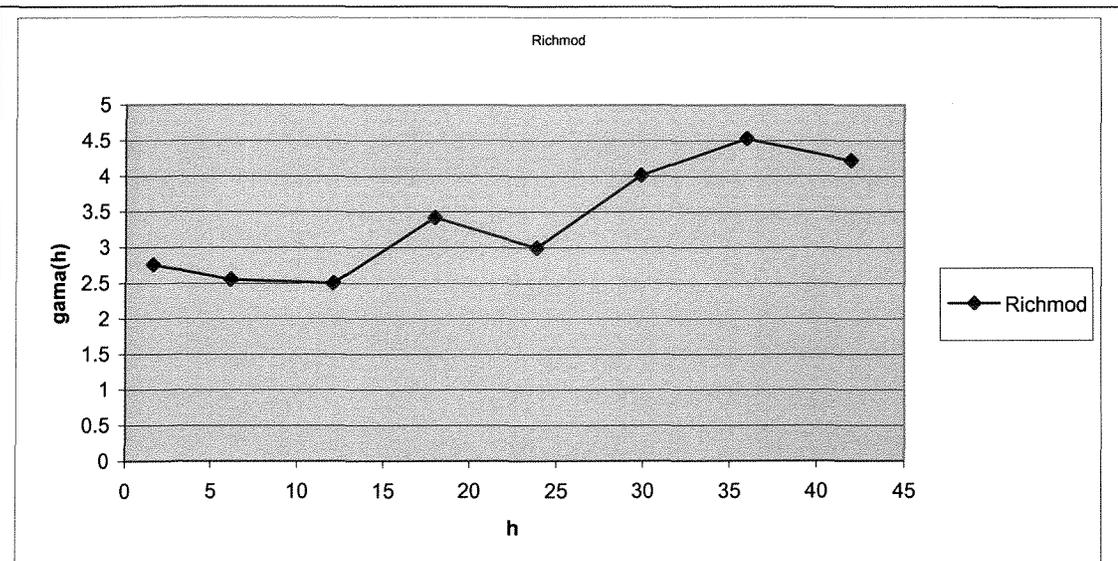


Figura 4.25 – Mostra o variograma ponderado por Richmod na direção azimute 30°, relativo ao conjunto de dados cluster.dat.

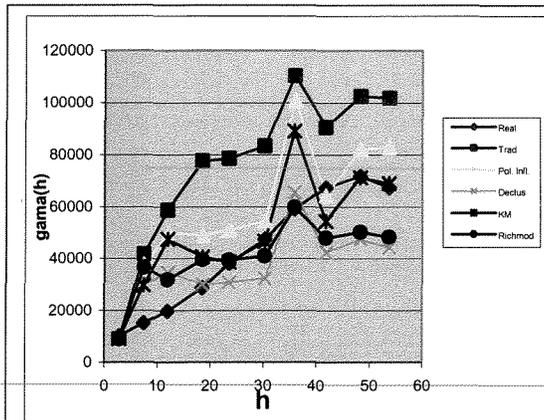


Figura 4.26 – figuras dos variogramas na direção azimute 90°.

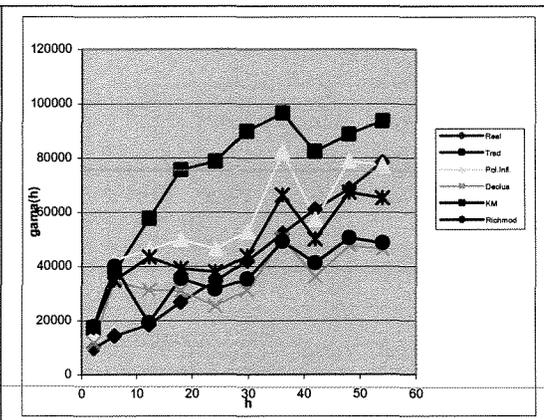


Figura 4.27 - figuras dos variogramas na direção 0°.

Ambas figuras mostram os variogramas experimentais relativos aos dados de Walker Lake, apresentando os quatro variogramas ponderados, variograma sem ponderação e o variograma sobre a imagem sem amostragem.

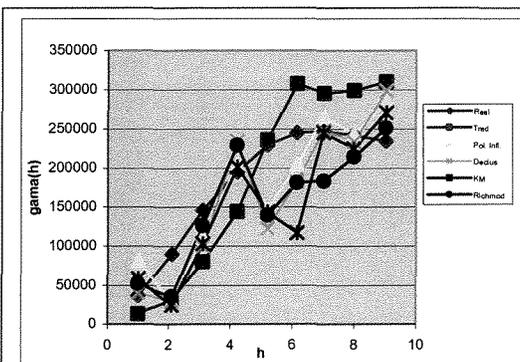


Figura 4.28 – A figura mostra os variogramas experimentais, relativos aos dados da imagem simulada 5 na direção azimute 90°.

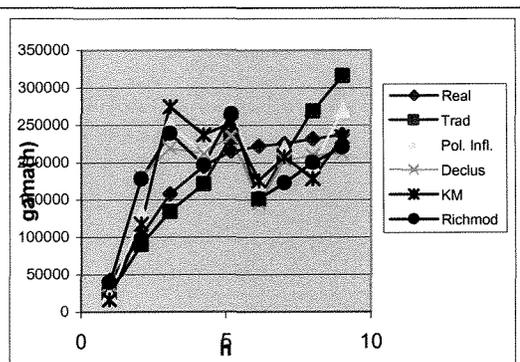
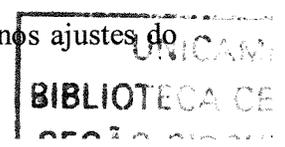


Figura 4.29 - A figura mostra os variogramas experimentais, relativos aos dados da imagem simulada 5 na direção azimute 0°.

Ambas as figuras referem se aos variogramas experimentais do conjunto de dados da imagem 5, e apresenta os variogramas ponderados pelos quatro métodos, mais o variograma sobre a imagem e também o variograma sem ponderar

Pelas figuras anteriores, pode-se chegar a conclusão que o uso de ponderadores no caso da variografia, reflete basicamente a diminuição do patamar; como este fato altera somente a variância de krigagem, os resultados são pequenas diferenças nos ajustes do



variograma e portanto o resultado final do cálculo da krigagem sofrerá modificações somente por diferenças sutis.

Para que se possa observar este fato vai-se aplicar uma rotina de “cross validation” nestes dados, com os vários tipos de variogramas, podendo-se então calcular a quantidade de “metal” que cada resultado fornece. Isto feito, vê-se que as curvas são bastante coincidentes, com apenas uma pequena diferença nos valores mais ricos. A curva marrom nas figuras 4.30, 4.31 e 4.32, representa o dado em si, e as outras os dados estimados

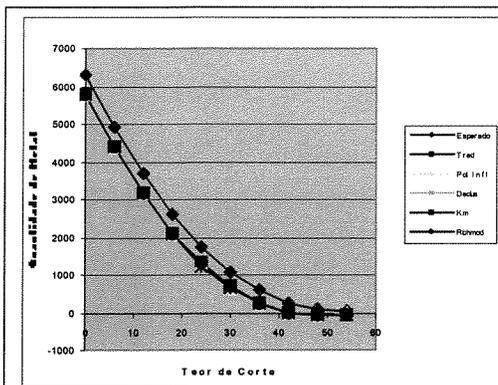


Figura 4.30 – A figura compara os resultados de cinco validações cruzadas, sobre o conjunto de dados cluster.dat e os dados originais, sendo que os variogramas foram calculados em quatro casos usando ponderador e um sem uso de ponderador.

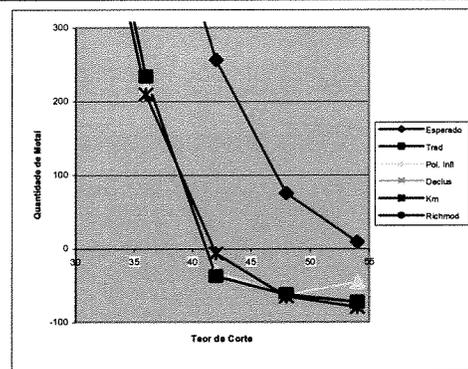


Figura 4.31 – A figura mostra um pequeno espaço da Figura 4.30, mostrando onde há maior diferença entre os cálculos.

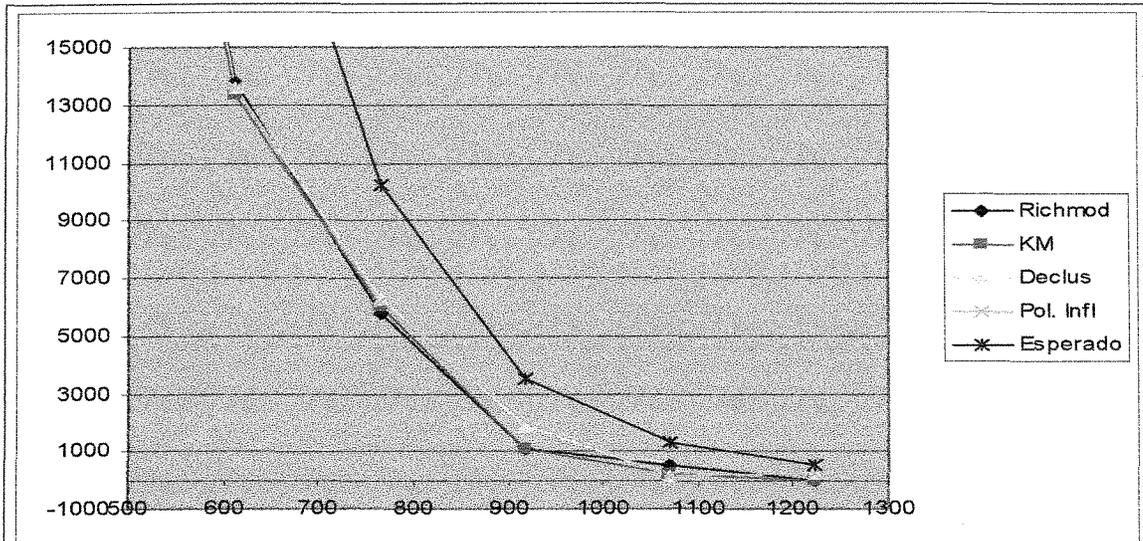


Figura 4.32 – Figura mostrando o resultado de quatro validações cruzadas nos 470 dados de Walker Lake, sendo que os variogramas utilizados foram ponderados pelos métodos aqui estudados.

No caso de krigagem, quando se efetua um estudo, a diferença entre os variogramas ponderados e o sem ponderação é muito pequena, o ganho que se tem é bastante reduzido, depende mais do ajuste; como a diferença é sutil, um determinado ajuste ou escolhendo um modelo diferente para o ajuste, causa mais impacto no resultado final que o ajuste em cima da deformação causada pelo uso de ponderador. Em algumas circunstâncias pode ocorrer que o efeito de pepita modifique o modelo usado.

Deve-se notar que no caso da krigagem, entra apenas nos cálculos o variograma ajustado e os dados, dados estes que sofrerão por conta da própria krigagem, um efeito de ponderação. A krigagem quando calcula os ponderadores leva em consideração a distribuição espacial dos dados e a covariância entre estas amostras, isto implica que um agrupamento, quando entrar no cálculo de um ponto, será por conta da proximidade entre os dados, sofrer um efeito de ponderação, podendo amostras nestes caso receberem ponderador negativo.

Tudo isto ocorre, se o ponto a ser krigado está entre uma amostra e o agrupamento. Se o ponto está no interior do agrupamento, ele se comporta normalmente, distribuindo os ponderadores pelos pontos pertencentes ao agrupamento.

4.3 A simulação Gaussiana Condicional

A simulação seqüencial gaussiana é a aplicação dos princípios da simulação seqüencial para a função aleatória com distribuição gaussiana. As funções de distribuição, condicionadas aos valores vizinhos dos pontos a serem simulados, serão gaussianas se seus parâmetros forem determinados por meio do sistema de krigagem simples. Isaaks (1990), Journel (1993) e Deutsch & Journel (1998) apresentam o algoritmo básico desse método de simulação seqüencial (Souza, 2002)

A simulação seqüencial pode ser aplicada para qualquer função de distribuição condicional, desde que seus parâmetros possam ser determinados pela krigagem simples. Os parâmetros da função de distribuição condicional são:

- 1- média condicional, que é o valor da variável estimada por krigagem simples;
- 2- a variância condicional, que é a variância de estimativa da krigagem simples.

Para que qualquer tipo de função aleatória possa ser aplicada na simulação seqüencial gaussiana, é necessário fazer a transformação dessa função em uma função gaussiana multivariada estacionária. Essa transformação é válida desde que preserve o modelo de covariância. A função aleatória gaussiana multivariada $Y(x)$ se caracteriza por (Deutsch & Journel, 1992):

- 1- seus subconjuntos são gaussianas multivariados;
- 2- a combinação linear de variáveis aleatórias gaussianas é gaussiana;

3- a covariância nula entre as variáveis aleatórias garante que elas são independentes;

4- as distribuições condicionais de variáveis aleatórias gaussianas são gaussianas;

Uma vez satisfeitas as condições acima, a função aleatória $Y(x)$ pode ser simulada por meio da simulação seqüencial gaussiana da seguinte maneira:

1- transformação dos z-dados da distribuição das amostras disponíveis para uma distribuição gaussiana multivariada e modelamento do variograma desses dados normalizados;

2- definição de um caminho randômico que passe por todos os pontos a serem simulados;

3-em cada ponto, construir a função de distribuição condicional para o ponto, baseada nos pontos vizinhos com amostras originais e previamente simuladas. A função de distribuição cumulativa condicional é caracterizada pela média e pela variância de estimativa calculada pela krigagem simples;

4-sorteio de um valor simulado dessa função de distribuição;

5-adição do valor simulado ao conjunto de amostras, mais os pontos simulados anteriormente;

6-repetição do procedimento para todos os pontos.

7-transporte os valores simulados sistematicamente de volta ao espaço original, aplicando de maneira inversa a mesma função utilizada em (1), permitindo confrontar o conjunto de realizações com o modelo de histograma dos dados.

Esse procedimento será repetido tantas vezes quantas forem as realizações geradas.

Transformação gaussiana dos dados

Conforme salientado anteriormente, a técnica de simulação seqüencial gaussiana exige que os dados obedeam a uma distribuição normal. Os dados, em geral, não obedecem essa condição. Portanto deve-se transformar os dados originais (condicionantes) para o espaço normal (Deutsch & Journel, 1998). Novamente, devido à amostragem preferencial apresentada pelos dados amostrais, é mais uma vez necessário utilizar ponderadores para obter a transformação adequada.

Se, a distribuição e média da população em estudo, estiverem enviesadas pela existência de um grande número de dados nas populações mais ricas, em forma de agrupamentos, esta anomalia do histograma/média implicará que a função transformação não operará de maneira satisfatória; a média, por exemplo terá um valor mais alto que o real e a forma do histograma também não será a mesma que a forma dos dados reais, ambas seguiram a forma e média dos dados amostrados.

Deste exposto fica claro à necessidade do uso de ponderadores no cálculo das simulações. Este fato é tão real, que no programa ISATIS, é a unicamp rotina que é possível se atribuir uma variável de ponderação. No caso do ISATIS, os valores dos ponderadores devem ser importados, pois o cálculo dentro programa ainda é bastante pobre de recursos; ele aplica o método do Declus, mas não apresenta sua potencialidade no diz respeito à procura por uma boa disposição das células.

Todos os cálculos e ajustes foram feitos no Isatis, como este não possibilita que variogramas sejam calculados com ponderador, foi feita uma pequena interface, que escrevia diretamente os variogramas calculados pelo Anavar, como um arquivo de parâmetros do Isatis; assim, é minimizada a possibilidade de erros que se teria, quando fosse efetuado um ajuste no Ajusta e este fosse “transportado” para o Isatis. Este programa foi escrito em QBasic, que apenas lê as três colunas de dados que compõem um variograma, a saber:- passo, número de pares de pontos e o valor do semi variograma, e escreve estes em um arquivo no padrão de arquivo de parâmetro de variograma experimental.

Para melhor compreender esta necessidade do uso de ponderador para o cálculo da simulação Condicional Gaussiana, serão apresentadas algumas simulações feitas, utilizando o ISATIS.

No caso da simulação, criou-se novas populações gaussianas, num total de cinco, que compreendem as populações da variável dos dados de Walker Lake, quando estes são transformados em gaussianas, a primeira população foi transformada, sem a utilização de ponderador e outras quatro foram geradas usando um dos ponderadores em estudo.

A partir destas cinco populações, foram feitos vinte e cinco ajustes de variogramas, cada população com e sem. De posse destes vinte e cinco ajustes, foram geradas 250 simulações, 10 simulações para cada modelo de variograma.

Estas simulações podem ser analisadas de diversas maneiras. Em uma primeira instância, vai se verificar as médias e desvios padrões para as populações estudadas; cada população é na Figura 4.33, representada por uma linha, e nesta linha cada ponto é representa um modelo de variografia utilizado.

Só para lembrar, os dados possuem as seguintes médias:

- Sem ponderação : 436.5
- Com ponderador gerado pelo método de Declus: 289.8
- Com ponderador gerado pelo método da krigagem da Média: 272.8
- Com ponderador gerado pelo método do Polígono de influência: 349.8
- Com ponderador gerado pelo método de Richmod: 274.3

O eixo das abscissas, está na posição do valor da média real .

O ponto 1º representa a média das populações com ajuste de variogramas clássicos ou sem ponderador, o 2º, ajuste com ponderador fornecido pelo método de Declus, 3º, utilizando ponderadores fornecidos pela KM, 4º, fornecidos pelo método de Richmod (2002) e finalmente o 5º utilizando ponderador fornecido pelo polígono de influência.

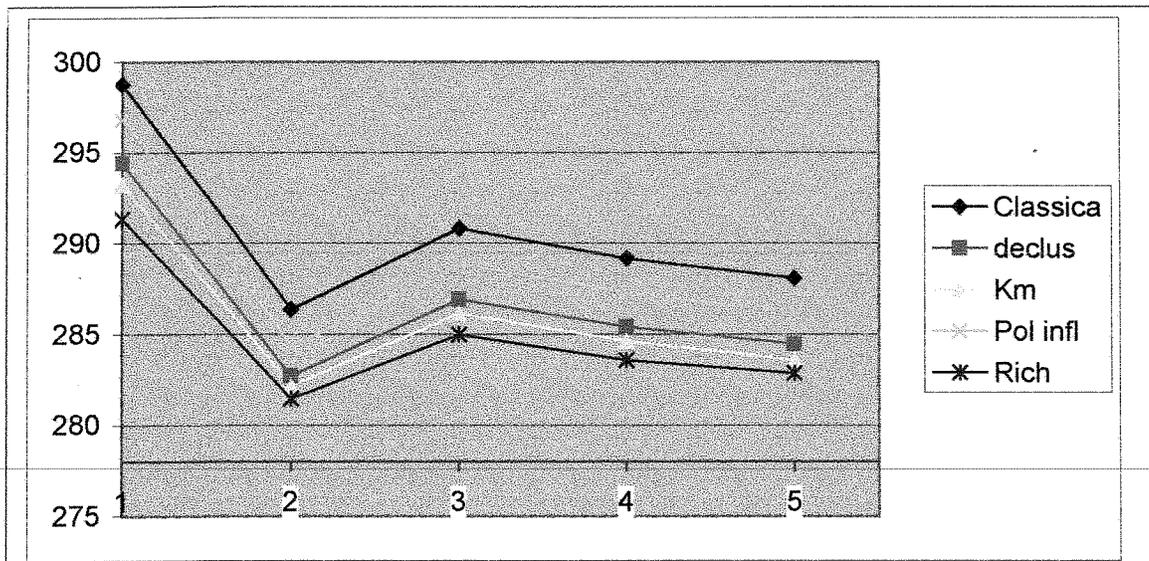


Figura 4.33 - Médias das simulações por modelo, lembrando que a média da população real é de 278.0, na figura cada curva representa um ajuste de variograma, e cada conjunto de pontos na vertical, representa uma transformação gaussiana diferente sobre o histograma.

Pode-se notar a grande influência que o método de ponderação tem sobre as simulações; nota-se que no caso da não utilização de ponderador, as curvas ficaram bem acima do valor esperado.

Pode-se também observar o comportamento dos desvios padrão, das mesmas simulações, Figura 4.34.

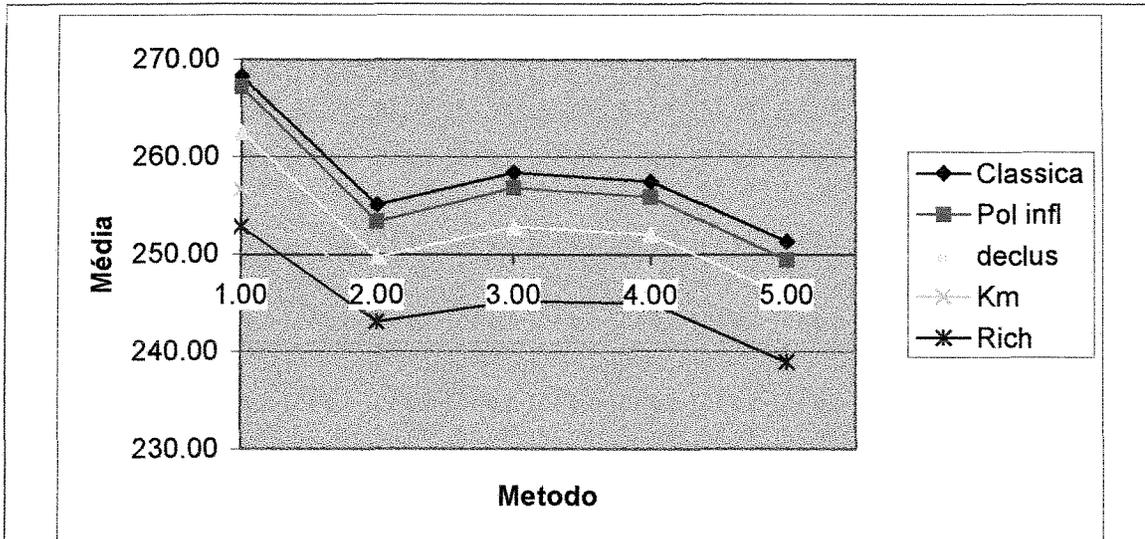


Figura 4.34 - Médias dos desvios padrão das simulações por modelo, lembrando que a média dos desvios padrão da população real é de 2.49.8, na figura cada curva representa um ajuste de variograma com determinado ponderador, e cada conjunto de pontos na vertical, representa uma transformação gaussiana diferentemente ponderada, aplicada sobre o histograma.

Outra vez, as curvas das populações que não se utilizou ponderadores e no caso do polígono de influência apresentaram um valor bastante superior ao esperado.

Pode-se também verificar individualmente todas as simulações e verificar seu comportamento quando aplicada a determinado modelo variográfico.

A figura abaixo mostra as médias das dez simulações, sobre os dados que não tiveram ponderadores, mas com variogramas com e sem ponderação.

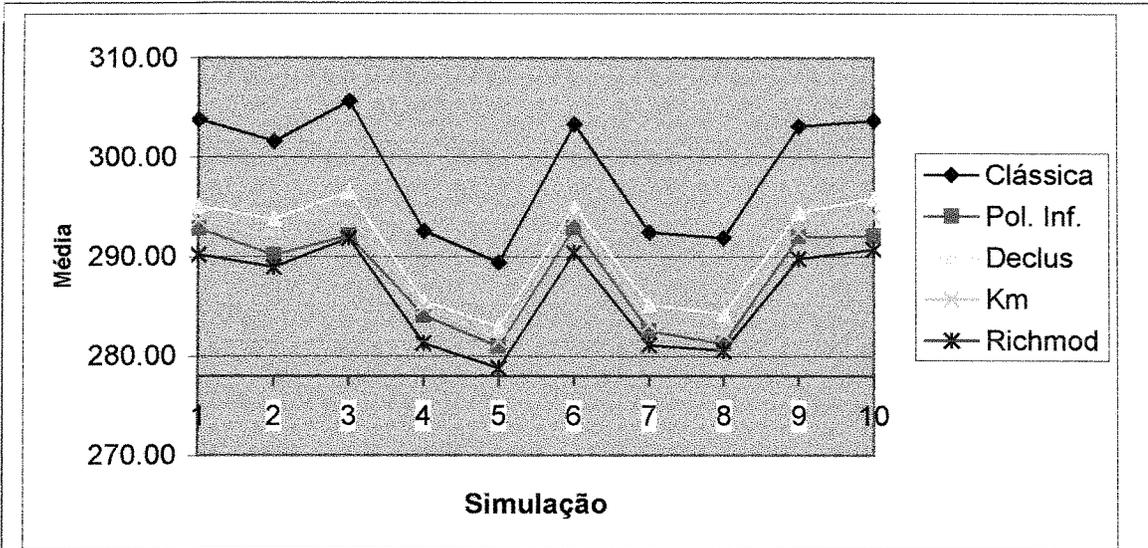


Figura 4.35 - Valores médios das simulações com a transformação gaussiana sem ponderador, com os variogramas ponderados pelos quatro métodos e o variograma sem ponderação.

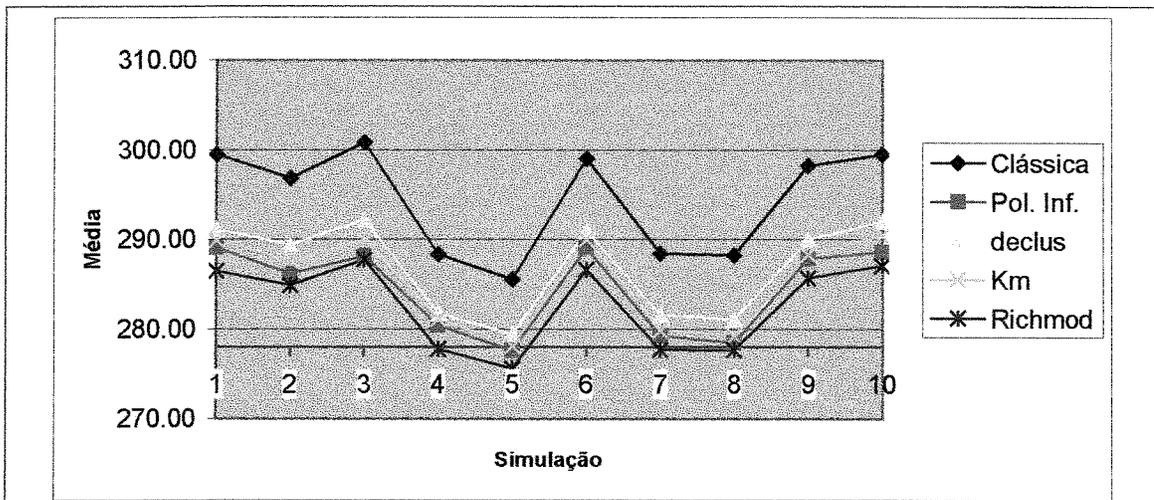


Figura 4.36 - Valores médios das simulações, com transformação gaussiana ponderada usando o resultado fornecido pelo método Declus e com os variogramas ponderados pelos quatro métodos além do variograma sem ponderação.

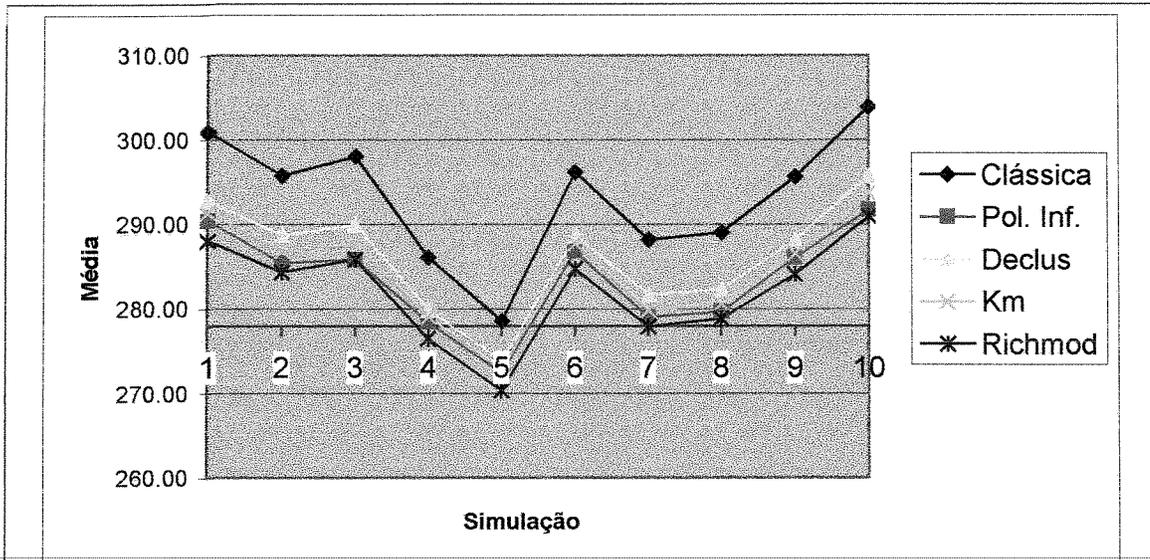


Figura 4.37 – Apresenta as médias das simulações, cujo histograma foi aplicada transformação gaussiana, usando os ponderadores fornecidos pela KM, e sobre os variogramas ponderados pelos quatro métodos e também sem ponderação.

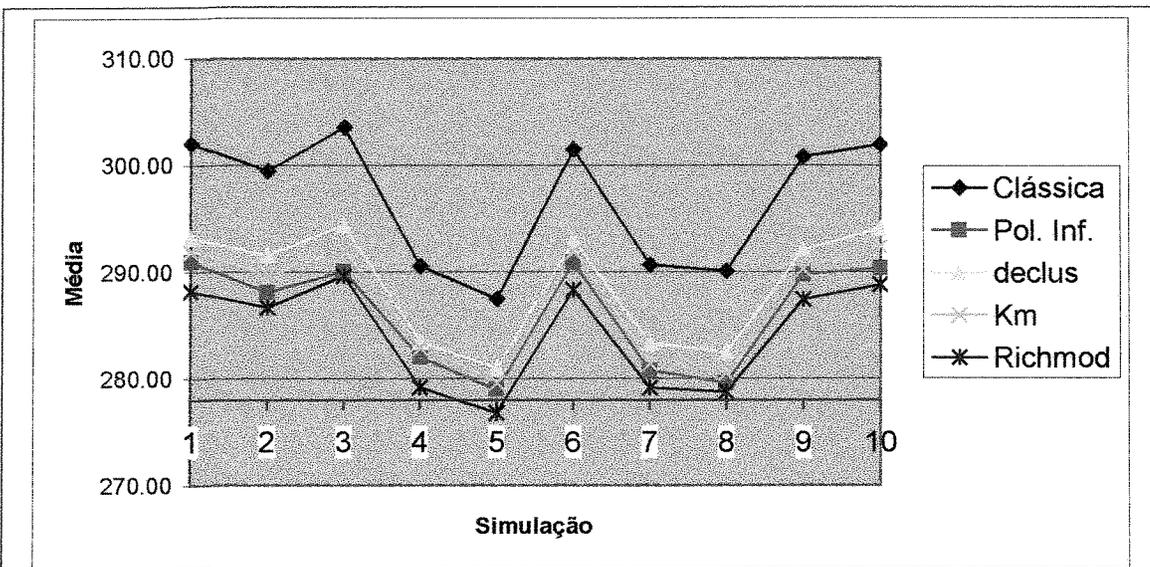


Figura 4.38 – Valores médios das simulações, com transformação gaussiana usando ponderador do polígono de influência, com os variogramas sem ponderador e também pelos quatro ponderadores.

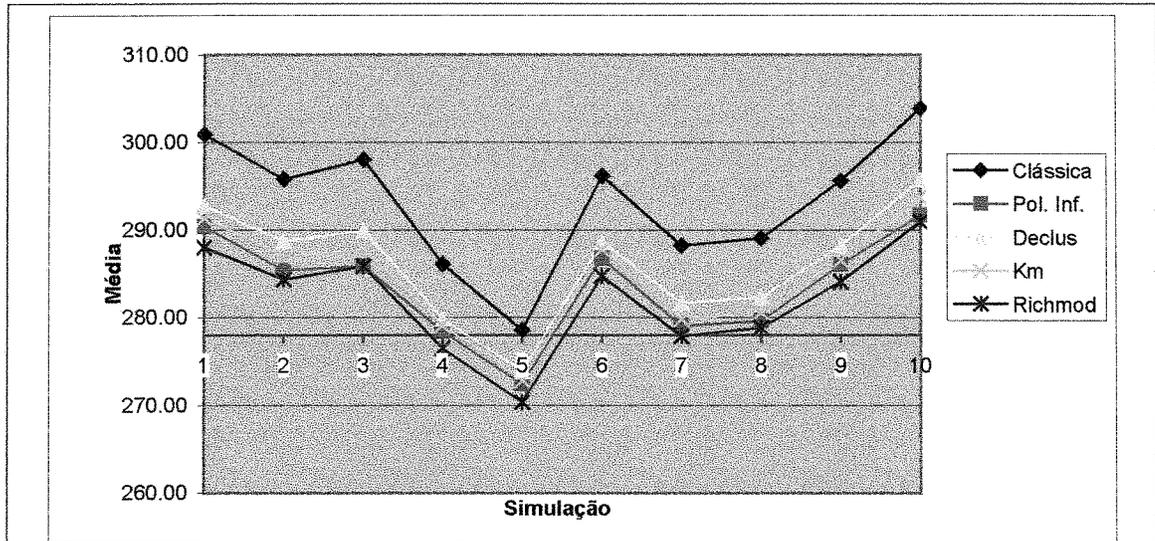


Figura 4.39 - Valores médios das simulações, onde foi aplicada uma transformação gaussiana com ponderadores resultante de Richmod, variogramas com e sem ponderadores.

Como pode-se verificar nos gráficos anteriormente mostrados, a necessidade de uso de ponderar no caso da simulação é marcante, os métodos fornecem resultados diferentes, tanto para o uso de variogramas com ponderador, como principalmente quando utilizamos ponderador para o cálculo da transformação gaussiana.

Neste caso, as transformações efetuadas com o auxílio dos ponderadores fornecidos pela KM e pelo método de Richmod, forneceram melhores resultados, pois as médias das simulações estão mais próximo da média esperada, que é a da população real.

Capítulo 5

Conclusões

Observando os estudos feitos durante este trabalho, fica difícil dizer qual método é melhor, qual apresenta o melhor resultado. Tudo vai depender da qualidade dos dados. Os ponderadores cada um apresenta uma média diferente, e também uma variância diferente, nos casos estudados aqui, que se possui os dados reais. Um fato que se nota, é que quando o fenômeno de agrupamento estiver caracterizado, um método de ponderador deve ser utilizado, pois, a não utilização de um ponderador, pode levar a resultados não confiáveis, com médias e variâncias bastante distintas das que realmente ocorrem no corpo em estudo.

Sempre vai ficar a pergunta qual o método a ser utilizado para gerar ponderador. Todos os métodos apresentam prós e contras, o método proposto por Deutsch (1989), é de fácil aplicação, em dois ou três minutos se pode ter um conjunto de ponderadores. No caso do polígono de influência, não é fácil de se conseguir programas que nos forneçam os valores procurados, isto no caso de 2D, no caso de 3D, fica ainda mais difícil.

No caso de método aqui chamando Declus, ele só é válido quando os agrupamentos foram efetuados em regiões ricas ou somente em regiões pobres, quando então se busca uma média mínima ou máxima respectivamente, caso contrário, quando os grupos estão espalhados nas regiões ricas e pobres, o método é completamente inválido.

Quando esta configuração de grupos em regiões ricas e pobres for utilizada para a produção de ponderadores utilizando-se da KM, devemos abandonar a proposta feita por Remacre(1996), que no seu artigo propõe a busca pela média mínima; deve-se sim efetuar o ajustamento do variograma, apenas visando um melhor ajuste visual e baseado em conhecimentos sobre a população em estudo.

Um programa que calcule a Krigagem da média, não é difícil de encontrar; é possível também se modificar algum já existente para que nos forneça os ponderadores que desejamos.

Os ponderadores advindos de algoritmo proposto por Richmod (2002), podem ser conseguidos implementado-se o algoritmo em uma linguagem qualquer. Neste trabalho, ele foi feito em Qbasic, e depois, mais com intuito de visualização, foi ajustado para rodar em Vbasic; o algoritmo é muito rápido, e com poucas opções. Fornecido no adendo, deve-se explicar que este programa apresenta apenas o corpo principal do algoritmo, pecando bastante no quesito de uso; os arquivos de saída podem ser modificados, o resultado gráfico, não possui saída, mas estas rotinas podem ser facilmente idealizadas por qualquer programador. Por sua simplicidade ele pode ser facilmente convertido para outra linguagem que o usuário esteja acostumado a trabalhar.

Do exposto, embasado nos resultados obtidos, nas sete populações estudadas, na facilidade de uso, principalmente para quem já está trabalhando em um estudo geoestatístico, o método de geração de ponderadores da KM, foi o que melhor respondeu as expectativas. Gerando resultados confiáveis, e dentro dos limites esperados.

Anexos

Anexo 1

Funções de recuperação e parametrização de reservas

O cálculo da reserva recuperável constitui-se no problema de estabelecer relações entre tonelagem e teor, as quais podem ser representadas sob diversos tipos de curvas (Rivoirard-1994), tais como mostra a Tabela 1.

O conjunto das funções da Tabela 1 é chamado de **Funções de Recuperação (FR)** e seu estabelecimento, de **Parametrização de Reservas (PR)**.

Com tais funções é possível responder questões importantes para a implantação de projetos de mineração e para o planejamento da lavra, tais como: Qual a quantidade de minério da reserva mineral cujos blocos possuem teores maiores que o *cut-off* z ? Qual a quantidade de metal contido na porção da reserva, cujos blocos possuem teores maiores que z ? Qual o teor médio dos blocos com teores maiores que z ? E a questão mais importante: Qual o benefício (lucro financeiro) gerado da lavra e tratamento de minério dos blocos que possuem teores maiores que o *cut-off* z ?

Tabela 1: Funções de Recuperação

Função	Variável Contínua	Variável Discreta
Tonelagem versus z : $T(z)$	$T(z) = P(Z(v) > z) = \int_z^{\infty} f(z) dz$ $T(z) = 1 - F(z)$	$T(z) = \frac{1}{n} \sum_{i=1}^n I_i(z)$

		$T(z) = \sum_{i=1}^n v_i \cdot d_i \cdot I_i(z)$
Metal versus z : $Q(z)$	$Q(z) = \int_z^{\infty} x \cdot f(x) dx$	$Q(z) = \frac{1}{n} \sum_{i=1}^n Z(v_i) \cdot I_i(z)$ $Q(z) = \sum_{i=1}^n v_i \cdot d_i \cdot Z(v_i) \cdot I_i(z)$
Teor médio versus z : $m(z)$		$m(z) = \frac{Q(z)}{T(z)}$
Custo versus z : $C(z)$		$C(z) = z \cdot T(z)$
Benefício versus z : $B(z)$		$B(z) = Q(z) - z \cdot T(z)$
Metal versus tonelagem: $Q(T)$		$Q(T) = \int_0^{+\infty} \min(T, T(u)) du$
<p>$f(z)$ é a probabilidade de cada valor z (função densidade de probabilidade, d a densidade e I a indicatriz de cada valor z, valendo 1 se $Z \geq z$ e 0 caso contrário, e n no número de blocos.</p>		

As FR podem ser obtidas usando funções de densidade de probabilidade (fdp) ajustadas aos histogramas experimentais, aos dados (contínuo), ou ainda, trabalhando diretamente sobre os dados estimados (discreto).

Anexo 2

Programa em visual basic para o cálculo dos ponderadores utilizando o método de Richmod

```
Attribute VB_Name = "Module1"

Global cluster(1500) As Integer
Global clustera(1500) As Integer
Global nd As Integer, nvar As Integer
Global xmi!, ymi, xma!, yma!
Global xmin(10), ymin(10), xmax(10), ymax(10)
Global nomevar(10) As String
Global test
Global dados(10, 1500) As Single
Global ix As Integer, iy As Integer, iv As Integer
```

```
Function existe(arq As String)
' funcao que retorna
' 1 caso arq exista
' 0 caso arq nao exista
```

```
On Error GoTo 333
  Open arq For Input As 99
  existe = 1
  Close 99
  Exit Function
333 Resume 334
334 existe = 0
End Function
```

```
Sub lelinha(i$, ni(), nn)
On Error GoTo ele:
llen = Len(i$)
ini = 0
kll = 0
For i = 1 To llen
  If Asc(Mid$(i$, i, 1)) = 9 Then
    Mid$(i$, i, 1) = ""
  End If
Next i
While ini <= llen
  i$ = ""
  While i$ = ""
    ini = ini + 1
    i$ = Mid$(i$, ini, 1)
  Wend
  ppos = InStr(ini + 1, i$, " ")
  If ppos = 0 Then ppos = llen
  kll = kll + 1
  ni(kll) = Val(Mid$(i$, ini, ppos - ini + 1))
  If kll >= nn Then
    GoTo lfim:
  End If
  ini = ppos
Wend
ele:
'erro$ = "Lelinha"
'leu = ""
if Err = 13 Then
  Beep
  ni(kll) = test
  ierro = MsgBox(" linha de dados nao numerica " & Chr$(13) & i$, 1, "")
```

```

    If ierro = 2 Then
        Resume 111
111 erro$ = "2"
        Exit Sub
    End If
    Resume Next
End If
Ifim:
End Sub

Sub cal_cluster(dma)
Dim ia(30000) As Integer, ja(30000) As Integer
dmax = dma ^ 2
k = 0
ix = Principal.cvar(0).Index + 1
iy = Principal.cvar(1).Index + 1
iv = Principal.cvar(2).Index + 1
For i = 1 To nd - 1
    For j = i + 1 To nd
        dx = dados(ix, i) - dados(ix, j)
        dy = dados(iy, i) - dados(iy, j)
        dd = dx * dx + dy * dy
        If dd = 0 Then
            MsgBox ("Pto duplicado " + Format(i, "#### &") + Format(j, "####"))
        End If
        If dd < dmax Then
            k = k + 1
            ia(k) = i
            ja(k) = j
        End If
    Next j
Next i
np = k
ReDim sos(np) As Integer
For i = 1 To np - 1
    ReDim p(nd) As Integer, pa(nd) As Integer
    If sos(i) <> 0 Then GoTo fimi
    p(1) = ia(i)
    p(2) = ja(i)
    pa(ia(i)) = 1
    pa(ja(i)) = 1
    sos(i) = i
    ip = 2
1 flag = 0
    For j = i + 1 To np
        If sos(j) <> 0 Then GoTo firmj
        For k = 1 To ip
            If ia(j) = p(k) Or ja(j) = p(k) Or ia(j) = p(k) Or ja(j) = p(k) Then
                If pa(ia(j)) <> 1 Then
                    ip = ip + 1
                    p(ip) = ia(j)
                    pa(ia(j)) = 1
                    sos(j) = i
                    flag = 1
                End If
                If pa(ja(j)) <> 1 Then
                    ip = ip + 1
                    p(ip) = ja(j)
                    pa(ja(j)) = 1
                    sos(j) = i
                    flag = 1
                End If
                sos(j) = i
            End If
        End If
    Next k
    firmj:
    Next j
    If flag = 1 Then GoTo 1
    ic = ic + 1

```

```

For j = 1 To ip
    cluster(p(j)) = ic
    clustera(p(j)) = ip
Next j
fimi:
Next i
media = 0
s = 0
ss = 0
Dim rv(1 To 1500)
Dim rx(1 To 1500)
Dim ry(1 To 1500)
For i = 1 To ic
    For j = 1 To np
        If cluster(j) = i Then
            rv(i) = rv(i) + dados(iv, j) / clustera(j)
            rx(i) = rx(i) + dados(ix, j) / clustera(j)
            ry(i) = ry(i) + dados(iy, j) / clustera(j)
        End If
    Next j
Next i

For i = 1 To nd
    Print Format(x(i), "####.##"); Format(y(i), "####.##"); Format(cluster(i), "###"); Format(clustera(i), "###")
    Print #2, Format(dados(ix, i), "000.## "); Format(dados(iv, i), "000.## "); Format(cluster(i), "00 "); clustera(i)
    If clustera(i) = 0 Then clustera(i) = 1
    ss = ss + 1 / clustera(i)
    s = s + dados(iv, i) * 1 / clustera(i)
    media = media + dados(iv, i) / nd
Next i
Open "d:\temp\td.dat" For Output As #2
Print #2, "teste"
Print #2, "3"
Print #2, "x"
Print #2, "y"
Print #2, "v"
For i = 1 To nd
    If cluster(i) = 0 Then
        Print #2, dados(ix, i); dados(iy, i); dados(iv, i)
    End If
Next i
For i = 1 To ic
    Print #2, rx(i); ry(i); rv(i)
Next i
Print USING; "###.#### ###.###"; Sqr(dmax); s / ss
Principal.media.Caption = Format(s / ss, "#####.###")
Principal.Combo1.AddItem (Format(Val(Principal.tdmax.Text), "0000.00 ") + Format(s / ss, "#####.###"))
For i = 0 To Principal.Combo1.ListCount - 2
    Principal.Combo1.ListIndex = i
    a1$ = Principal.Combo1.Text
    Principal.Combo1.ListIndex = i + 1
    a2$ = Principal.Combo1.Text
    If a1$ = a2$ Then
        Principal.Combo1.RemoveItem (i + 1)
    End If
Exit For
End If
Next i
Close #2
Principal.Combo1.ListIndex = Principal.Combo1.ListCount - 1
Principal.variancia.Caption = Format(media, "#####.###")
End Sub

Sub eixo(xmi, xma, ymi, yma, ntickx, nticky, tx$, ty$, p As PictureBox)
Dim resx, resy As Single
p.FontName = "times new roman"
If tamanho_fonte <= 0 Then tamanho_fonte = 10
p.FontSize = tamanho_fonte
p.FontBold = False
'eixo x
ticc = p.Height / 10000

```

```

p.Scale (-15, 110)-(115, -13)
p.Line (0, -ticc)-(0, 0), RGB(0, 0, 0)
intx = (xma - xmi) / ntickx
dx = 100 / ntickx
ep$ = eprecx
If Abs(xmi) > 10000 Then ep$ = "0.00E+00"
r$ = Format(xmi, ep$)
la = Len(r$)
reth = p.Width / 10000
retv = p.Height / 10000
p.CurrentX = -la / 2
p.CurrentY = -ticc * 2
' p.PSet (-la, -3), RGB(255, 255, 255)
p.Print r$
For i = 0 To 99.9 Step dx
  If grideixox = -1 Then
    p.Line (i, 0)-(i, 100), RGB(0, 0, 0)
  End If
  p.Line (i, 0)-(i + dx, 0), RGB(0, 0, 0)
  p.Line -(i + dx, -ticc), RGB(0, 0, 0)
  i = i + 1
  resx = xmi + intx * i
  r$ = Format(resx, ep$)
  la = p.TextWidth(r$)
  p.CurrentX = dx + i - la / 2
  p.CurrentY = -ticc * 2
  p.Print r$

```

Next i

' eixo y

```

p.Line (-1, 0)-(0, 0), RGB(0, 0, 0)
dy = 100 / nticky
inty = (yma - ymi) / nticky
ep$ = eprecy
If Abs(ymi) > 10000 Then ep$ = "0.00E+00"
r$ = Format(ymi, ep$)
la = p.TextWidth(r$)
p.CurrentX = -la - reth + ddt
p.CurrentY = retv
p.Print r$
i = 0
For i = 0 To 99.99 Step dy
  If grideixoy = -1 Then
    p.Line (0, i)-(100, i), RGB(0, 0, 0)
  End If
  p.Line (0, i)-(0, i + dy), RGB(0, 0, 0)
  p.Line -(-1, i + dy), RGB(0, 0, 0)
  i = i + 1
  resy = ymi + inty * i
  r$ = Format(resy, ep$)
  la = p.TextWidth(r$)
  p.CurrentX = -la - reth + ddt
  p.CurrentY = i + dy + retv
  p.Print r$

```

Next i

```

p.FontName = "times new roman"
p.FontBold = True
la = (100 - p.TextWidth(titulop)) / 2
p.CurrentX = la
p.CurrentY = 109
p.Print titulop
p.FontBold = False
la = Len(tx$)
p.CurrentX = 50 - la
p.CurrentY = -7
'p.PSet (50 - la, -12), RGB(255, 255, 255)
p.Print tx$
la = Len(ty$)

```

```
ini = 50 + la * 2
For i = 1 To la
  p.CurrentX = -p.TextWidth(r$) - reth * 2
  p.CurrentY = ini - i * 5
  'p.PSet (-13, ini - i * 5), RGB(255, 255, 255)
  p.Print Mid$(ty$, i, 1)
Next i
End Sub
```

Referencias Bibliográficas.

BLEINER, C.; PERSEVAL, P.; RAMBERT, F.; RENARD, D.; TOUFFAIT, Y. 2000. Isatis software manual V.4.05. Geostatistics & Ecole des Mines de Paris, March 2000, Paris, France. 585p.

CORNETTI, M.A. & REMACRE A.Z, ANAVAR - Análise de variabilidade, Laboratório de Informática Geológica, Campinas, UNICAMP, 1995.

CRUZ JR, A. C.; REMACRE, A. Z. Impacto dos efeitos de suporte e informação sobre a recuperação de reservas Revista da Escola de Minas de Ouro Preto, Ouro Preto MG, v. 54, p. 1-11, 2001

DAVIS, J.C. Statistics and data analysis in geology 2^a Ed. New York, John Wiley & Sons, 1973. 646p.

DEUTSCH, C.V., Declus: a fortran 77 program to determine optimal spatial declustering weights. Computers & Geosciences 15 (3), 1989, 325-332

ISAAKS, E. H. ; SRIVASTAVA, R. M. An Introduction to applied geostatistics, New York, Oxford University Press. 1989. 561p.

HAYES, W. B. , KOCH, G. S. Constructing and analyzing area-of-influence polygons by Computer. Computers & Geosciences 10 (3), 1984, 411-430

JOURNEL, A. G. HUIJBREGTS, CH, J. Mining geostatistics (Fort Edition) London, Academic Press. 1989. 600p.

REMACRE, A.Z. Krigagem da média. Cadernos de Geociências (São Paulo), V15 N.1, 1996, P9-16

RICHMOND, ANDREW, Two points declustering for weighting data pairs in experimental variogram calculations, *Computer & Geosciences* V. 28 (2002) 231-241

RIVOIRARD (2000), J, Weighted variograms, *Kleingled. W.J. Krige, D.C. (Eds), Proceedings of the 6th International Geostatistics Congress, Cape Town, South Africa, 2000, 11p.*

SOUZA, LUIZ E (2002), Estimativa de incertezas e sua aplicação na classificação de recursos minerais Porto Alegre: PPGEM, paginas 81 a 83. Tese de mestrado
