



COPPE/UFRJ

**ANÁLISE ESTATÍSTICA E GEOESTATÍSTICA DE DADOS GEOQUÍMICOS
DE SUPERFÍCIE APLICADA A EXPLORAÇÃO DE HIDROCARBONETOS**

Michel Ângelo Pérez Bedregal

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Civil, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Civil.

Orientador: Luiz Landau

Rio de Janeiro
Dezembro de 2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

ANÁLISE ESTATÍSTICA E GEOESTATÍSTICA DE DADOS GEOQUÍMICOS DE
SUPERFÍCIE APLICADA A EXPLORAÇÃO DE HIDROCARBONETOS

Michel Ângelo Pérez Bedregal

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM
CIÊNCIAS EM ENGENHARIA CIVIL.

Aprovada por:

Prof. Luiz Landau, D.Sc.

Prof. Nelson Francisco Favilla Ebecken, D.Sc.

Dr. Felix Thadeu Teixeira Gonçalves, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
DEZEMBRO DE 2008

Pérez Bedregal, Michel Ângelo

Análise Estatística e Geoestatística de Dados Geoquímicos de Superfície Aplicada a Exploração de Hidrocarbonetos/ Michel Ângelo Pérez Bedregal. - Rio de Janeiro: UFRJ/COPPE, 2008.

XVII, 139 p.: il.; 29,7 cm

Orientador: Luiz Landau

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia Civil, 2008.

Referências Bibliográficas: p. 127-139.

1. Análise estatística. 2. Determinação de limiar. 3. Geoquímica de gases. 4. Aplicação geoestatística. I. Landau, Luiz. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

À minha família, com muito orgulho...

AGRADECIMENTOS

À Deus por me iluminar e me dar força, paz e sabedoria durante este trabalho.

Em especial aos meus queridos pais Ricardo e Goretti pelos inúmeros conselhos, pelo eterno incentivo, pelo ombro amigo cheio de conforto, amor e carinho. Por serem responsáveis por toda a minha formação, pela confiança em mim depositada e por serem exemplos na minha vida, obrigado sempre.

À Olga, minha esposa, pelo amor, carinho e paciência despendidos nos momentos de preocupação e pelos conselhos úteis. Por me proporcionar momentos de lazer e descontração fundamentais para o desencadear deste trabalho.

À toda a minha família, irmãos, sobrinhos, e principalmente ao Bedregal pelo incentivo, apoio moral e pelo exemplo de profissional e irmão, e ao meu sobrinho Victor pelos momentos de pura alegria.

À Luís Landau, pelo apoio financeiro, técnico e administrativo e por demonstrar de forma ímpar que somos capazes, se realmente estivermos dispostos a crescer.

Ao meu amigo e co-orientador Felix Gonçalves, pelas palavras de incentivo, atenção e dedicação a minha pesquisa.

Ao CNPq e ANP pelo apoio financeiro concedido permitindo realizar esta pesquisa.

À todos os meus amigos da universidade, Jean, Sylvia, Kátia, Marco, Fábio, Márcia, Patrícia, Ana Paula, Juliana, Renata e Carolina, por me proporcionarem felicidade, companheirismo e força, dividindo comigo as alegrias e dificuldades.

Aos professores do mestrado, em especial ao Bacocoli, Bandeira e Pellon, pela transmissão de conhecimentos e experiências, contribuindo fortemente para a minha formação profissional.

E à todos os funcionários que indiretamente me ajudaram na realização deste trabalho, principalmente aos do Lab2M e LAMCE, em especial a Verônica, Mônica, Magda (ex-funcionária), Allan (ex-funcionário), Marco e Serginho.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ANÁLISE ESTATÍSTICA E GEOESTATÍSTICA DE DADOS GEOQUÍMICOS DE SUPERFÍCIE APLICADA A EXPLORAÇÃO DE HIDROCARBONETOS

Michel Ângelo Pérez Bedregal

Dezembro/2008

Orientador: Luiz Landau

Programa: Engenharia Civil

Esta dissertação analisou a distribuição de dados geoquímicos de superfície (hidrocarbonetos gasosos), identificou dados geoquímicos *outliers* por diversos métodos e comparou métodos de interpolação determinísticos (curvatura mínima e inverso do quadrado da distância) e geoestatístico (krigagem). As medidas obtidas em campo foram analisadas através da estatística descrita clássica e espacial. As análises estatísticas mostraram que os dados geoquímicos de hidrocarbonetos gasosos não seguem, de maneira geral, uma distribuição lognormal ou normal. Métodos para identificar dados *outliers* nos dados geoquímicos foram revisados e testados. A regra [média (\bar{x}) \pm 2 desvios padrão (dp)] utilizada para estimar valores de limiar dividindo dados *background* de anômalos, até então utilizada desde sua introdução na década de 60, mostrou resultados diferentes do proposto pelo método, no qual deveria identificar aproximadamente 2^{1/2}% de dados em cada extremo. A mediana (med) \pm 2 desvios mediano absoluto (dma), o *boxplot* e 98º percentil, pareceram ser mais apropriados na estimação de limiar e na determinação da extensão de dados *background*. Porém, cada um pode resultar em diferentes valores de limiar. A interpolação desses dados para geração de mapas de superfície pelo método geoestatístico, que tem como base para interpolação a krigagem, leva em consideração as características espaciais de autocorrelação de variáveis regionalizadas que apresentaram melhores resultados do que pelos métodos determinísticos.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

STATISTICAL ANALYSIS AND GEOSTATISTIC OF SURFACE GEOCHEMICAL DATA APPLIED TO THE HYDROCARBONS EXPLORATION

Michel Ângelo Pérez Bedregal

December/2008

Advisor: Luiz landau

Department: Civil Engineering

This dissertation examined the distribution of geochemical data from surface (gaseous hydrocarbons), geochemical outliers data, identified by various methods, and compared deterministic interpolation methods (minimum curvature and the inverse square of the distance) and geostatistical (kriging). The measurements obtained in the field were analyzed by classical descriptive statistics and spacial. Statistical analysis showed that the geochemical data of hydrocarbon gases do not follow in general, a lognormal distribution or a normal one. Methods to identify outliers in the geochemical data were reviewed and tested. The rule [mean (\bar{x}) \pm 2 standard deviations (sdev)] used to estimate the threshold values dividing background of anomalous data, hitherto used since its introduction in the 60s, showed different results of the proposed method in which it should identify approximately 2¹/₂% of data on each end. The median (med) \pm 2 median absolute deviations (mad), the boxplot and 98 th percentile, appeared to be more appropriate in the estimates of threshold and the determination of the extent of background data. However, each can result in different values of threshold. The interpolation of these data in order to generate surface maps through geostatistical method, which is based on the kriging interpolation, takes into account the characteristics of spatial autocorrelation of regional variables, which showed better results than the deterministic methods.

SUMÁRIO

DEDICATÓRIA.....	iv
AGRADECIMENTOS.....	v
RESUMO.....	vi
ABSTRACT.....	vii
SUMÁRIO.....	viii
LISTA DE FIGURAS.....	xi
LISTA DE TABELAS.....	xvi
CAPÍTULO 1 – INTRODUÇÃO.....	1
1.1 – Motivação científica para a realização do trabalho.....	1
1.2 – Objetivos.....	1
1.3 – Organização da dissertação.....	2
CAPÍTULO 2 – GEOQUÍMICA DE SUPERFÍCIE.....	4
2.1 – Introdução.....	4
2.2 – Definição e princípios da exploração.....	5
2.3 – Processos de Migração e exsudação.....	6
2.3.1 – Migração por efusão.....	8
2.3.2 – Migração por difusão.....	8
2.3.3 – Migração por movimento.....	10
2.4 – Objetivos da geoquímica de superfície.....	10
2.4.1 – Exploração.....	10
2.4.2 – Reconhecimento.....	10
2.4.3 – Potencial exploratório.....	11
2.5 – Métodos de prospecção de geoquímica de superfície.....	11
2.5.1 – Métodos diretos.....	11
2.5.2 – Métodos indiretos.....	12
2.5.3 – Métodos não convencionais.....	12
2.6 – Reconhecimento de anomalias.....	13
2.7 – Métodos de determinação de <i>background</i> e anomalias.....	14
2.8 – Origens dos hidrocarbonetos gasosos.....	20
2.8.1 – Tipos de gás natural e sua importância.....	20
2.9 – Benefícios da geoquímica de exploração.....	22

CAPÍTULO 3 – MÉTODOS DE INTERPOLAÇÃO E GEOESTATÍSTICA.....	23
3.1 – Métodos de interpolação.....	23
3.2 – Teoria da amostragem clássica.....	27
3.3 – Teoria geoestatística.....	28
3.4 – Variáveis regionalizadas.....	29
3.5 – Modelagem probabilística e estacionaridade.....	32
3.6 – Número de vizinhos das estimativas.....	37
3.7 – Descrição e modelagem da estrutura de variabilidade espacial.....	39
3.7.1 – Semivariância e semivariograma.....	40
3.7.2 – Modelos teóricos de semivariograma.....	44
3.8 – Métodos de estimação de semivariograma.....	48
3.9 – A estimação geoestatística.....	49
3.9.1 – Krigagem ordinária.....	50
3.9.2 – Krigagem universal.....	53
3.10 – Validação cruzada.....	53
CAPÍTULO 4 – MATERIAIS E MÉTODOS.....	55
4.1 – Localização e delineamento amostral.....	55
4.2 – Análise estatística.....	56
4.2.1 – Análise exploratória descritiva.....	57
4.3 – Determinação de <i>background</i> e anomalia.....	60
4.4 – Análise exploratória espacial.....	60
4.5 – Análise geoestatística.....	61
4.6 – Comparação entre os métodos de interpolação.....	63
CAPÍTULO 5 – RESULTADOS E DISCUSSÕES.....	64
5.1 – Análise exploratória descritiva.....	64
5.1.1 – Dados originais.....	64
5.1.2 – Dados transformados.....	73
5.2 – Análise de <i>background</i> e anomalia.....	80
5.3 – Análise exploratória espacial.....	92
5.4 – Análise geoestatística.....	102
5.4.1 – Análise semivariográfica.....	102
5.4.2 – Interpolação por krigagem, curvatura mínima e inverso do quadrado da distância.....	107
5.4.3 – Validação cruzada.....	115

5.4.4 – Comparação entre métodos de interpolação.....	124
CAPÍTULO 6 – CONCLUSÕES E RECOMENDAÇÕES.....	126
REFERÊNCIAS BIBLIOGRÁFICAS.....	127

LISTA DE FIGURAS

Figura 2.1 – Fluxo dos caminhos e direções da difusão de hidrocarbonetos leves em folhelhos geradores. A) em direção à rocha carreadora; B) em direção à zona de fratura; C) em direção à falha; D) em direção ao final da camada de siltito (modificado de LEYTHAEUSER <i>et al.</i> , 1982).....	9
Figura 2.2 – <i>Boxplot</i> , destacando suas principais características.....	19
Figura 3.3 - Configuração de pontos de amostragem para estimativa no ponto A.....	26
Figura 3.4 - Inter-relações entre amostras e ponto a ser estimado para cálculo das equações de krigagem.....	26
Figura 3.5 – Efeito pepita (modificado VALENTE, 1989).....	35
Figura 3.6 - Comportamento na vizinhança da origem de quatro tipos de semivariograma.....	42
Figura 3.7 - Semivariograma com efeito pepita.....	43
Figura 3.8 – Principais modelos semivariográficos teóricos.....	45
Figura 3.9 - Aspecto do semivariograma com total independência entre as amostras.....	48
Figura 4.10 – Distribuição espacial das amostras do levantamento geoquímico, com a localização amostral dos 220 pontos.....	56
Figura 5.11 – (a) Histograma de frequência, (b) gráfico <i>boxplot</i> e (c) de probabilidade para os dados de metano (ppm).....	65
Figura 5.12 – (a) Histograma de frequência, (b) gráfico <i>boxplot</i> e (c) de probabilidade para a variável C_2^-/C_2	68
Figura 5.13a: Histograma de frequência para as variáveis C_2 , C_2^- , C_3 , C_3^- , $i-C_4$, C_4 , C_4^- ,	

i-C ₅ , C ₅ , C ₆ ⁺ , C ₂ C ₅ e C ₁ /C ₂ C ₅	70
Figura 5.13b: <i>Boxplot</i> para as variáveis C ₂ , C ₂ ⁻ , C ₃ , C ₃ ⁻ , i-C ₄ , C ₄ , C ₄ ⁻ , i-C ₅ , C ₅ , C ₆ ⁺ , C ₂ C ₅ e C ₁ /C ₂ C ₅	71
Figura 5.13c: Gráfico de probabilidade para as variáveis C ₂ , C ₂ ⁻ , C ₃ , C ₃ ⁻ , i-C ₄ , C ₄ , C ₄ ⁻ , i-C ₅ , C ₅ , C ₆ ⁺ , C ₂ C ₅ e C ₁ /C ₂ C ₅	72
Figura 5.14a: Histograma de frequência para as variáveis transformadas (ln) C ₂ , C ₂ ⁻ , C ₃ , C ₃ ⁻ , i-C ₄ , C ₄ , i-C ₅ , C ₅ , C ₆ ⁺ , C ₂ C ₅ e C ₁ /C ₂ C ₅	76
.	
Figura 5.14b: <i>Boxplot</i> para as variáveis transformadas (ln) C ₂ , C ₂ ⁻ , C ₃ , C ₃ ⁻ , i-C ₄ , C ₄ , i-C ₅ , C ₅ , C ₆ ⁺ , C ₂ C ₅ e C ₁ /C ₂ C ₅	77
Figura 5.14c: Gráfico de probabilidade para as variáveis transformadas (ln) C ₂ , C ₂ ⁻ , C ₃ , C ₃ ⁻ , i-C ₄ , C ₄ , i-C ₅ , C ₅ , C ₆ ⁺ , C ₂ C ₅ e C ₁ /C ₂ C ₅	78
Figura 5.15a – Histograma de frequência, gráfico <i>boxplot</i> e de probabilidade para a variável lnC ₁	79
Figura 5.15b – Histograma de frequência, gráfico <i>boxplot</i> e de probabilidade para as variável lnC ₄ ⁻	79
Figura 5.16 – Postplot dos valores de background (azul claro) e outliers (azul escuro) das variáveis C ₁ (a), C ₂ (b), C ₂ ⁻ (c), C ₃ (d), C ₃ ⁻ (e), i-C ₄ (f), C ₄ (g), C ₄ ⁻ (h), i-C ₅ (i), C ₅ (j), C ₆ ⁺ (k), C ₂ ⁻ / C ₂ (l), C ₂ C ₅ (m) e C ₁ /C ₂ C ₅ (n), utilizando o limiar superior via: [média + 2 desvios padrão], [mediana + 2 desvios mediano absoluto], <i>boxplot</i> (upper whisker) e 98 ^o percentil para os valores originais e transformados respectivamente.....	91
Figura 5.17 – Superfície de valores (a) e dispersão das medidas do gás metano (ln) amostradas nos sentidos Oeste-Leste (b) e Sul-Norte (c).....	93
Figura 5.18 – Gráficos de dispersão das médias e variâncias dos valores de metano (ln) amostradas nos sentidos das colunas (a) e linhas (b).....	94
Figura 5.19 – Análise descritiva espacial pelo “ <i>Postplot</i> ” da variável transformada	

metano.....	95
Figura 5.20 – Superfície de valores (a) e dispersão das medidas de C_2C_5 (ln) amostradas nos sentidos Oeste-Leste (b) e Sul-Norte (c).....	97
Figura 5.21 – Gráficos de dispersão das médias e variâncias dos valores de C_2C_5 (ln) amostradas nos sentidos das colunas (a) e linhas (b).....	98
Figura 5.22 – Análise descritiva espacial pelo “ <i>Postplot</i> ” da variável $\ln C_2C_5$	98
Figura 5.23 – Superfície de valores (a) e dispersão das medidas da variável eteno ($C_2^=$) amostradas nos sentidos Oeste-Leste (b) e Sul-Norte (c).....	100
Figura 5.24 – Gráficos de dispersão das médias e variâncias dos valores de eteno ($C_2^=$) amostradas nos sentidos das colunas (a) e linhas (b).....	101
Figura 5.25 – Análise descritiva espacial pelo “ <i>Postplot</i> ” da Variável Eteno ($C_2^=$).....	101
Figura 5.26 - Semivariograma experimental na direção 0° das ordenadas.....	102
Figura 5.27 – Semivariograma para a variável $\ln C_1$, ajustado ao modelo esférico com os seguintes parâmetros: $C_0= 0,0786$; $a= 3025$ m e $C = 0,1012$	103
Figura 5.28 - Semivariograma experimental omnidirecional para a variável $\ln C_2C_5$	104
Figura 5.29 – Semivariograma para a variável $\ln C_2C_5$ ajustado ao modelo exponencial com os seguintes parâmetros: $C_0=0,1728$; $a=2802$ m, $C=0,1520$ e $C+C_0=0,3247$	105
Figura 5.30 - Semivariograma experimental na direção 0° das ordenadas.....	106
Figura 5.31 – Semivariograma para a variável eteno, ajustado ao modelo esférico com os seguintes parâmetros: $C_0= 16,9218$; $a=2981$ m e $C=7,3537$	106
Figura 5.32 - Mapas de contorno interpolada para o metano (ppm) pelo método de Krigagem ordinária considerando o modelo esférico (a), pelo método de curvatura	

mínima (b) e inverso do quadrado da distância (c).....	109
Figura 5.33 - Mapas de superfície interpolados para o metano (ppm) pelo método de Krigagem ordinária (a), método de curvatura mínima (b) e inverso do quadrado da distância (c).....	110
Figura 5.34 - Mapas de contorno interpolados para o C ₂ C ₅ (ppm) pelo método de Krigagem ordinária (a), método de curvatura mínima (b) e inverso do quadrado da distância (c).....	111
Figura 5.35 - Mapas de superfície interpolados para o C ₂ C ₅ (ppm) pelo método de Krigagem ordinária (a), método de curvatura mínima (b) e inverso do quadrado da distância (c).....	112
Figura 5.36 - Mapas de contorno interpolados para eteno (C ₂ ⁺) (ppm) pelo método de Krigagem ordinária (a), método de curvatura mínima (b) e inverso do quadrado da distância (c).....	113
Figura 5.37 - Mapas de superfície interpolados para eteno (C ₂ ⁻) (ppm) pelo método de Krigagem ordinária (a), método de curvatura mínima (b) e inverso do quadrado da distância (c).....	114
Figura 5.38 - Figura 5.38 - Mapa de erros dos valores estimados em relação aos observados da variável metano ajustado ao modelo esférico.....	116
Figura 5.39 – Gráficos de dispersão das estimativas de krigagem em relação aos valores observados na validação cruzada para a variável metano no modelo esférico.....	117
Figura 5.40 – Histograma de frequência dos erros (a) e dos erros padronizados da variável metano (b), ajustado ao modelo esférico.....	117
Figura 5.41 - Figura 5.41 - Mapa de erros dos valores estimados em relação aos observados da variável C ₂ C ₅ ajustado ao modelo exponencial.....	119
Figura 5.42 – Gráficos de dispersão das estimativas de krigagem em relação aos valores observados na validação cruzada para a variável C ₂ C ₅ no modelo	

exponencial..... 120

Figura 5.43 – Histograma de frequência dos erros (a) e dos erros padronizados da variável C_2C_5 (b), ajustado ao modelo exponencial..... 120

Figura 5.44 - Mapa de erros dos valores estimados da variável eteno em relação aos observados ajustado ao modelo esférico..... 122

Figura 5.45 – Gráficos de dispersão das estimativas de krigagem em relação aos valores observados na validação cruzada para a variável eteno no modelo esférico..... 123

Figura 5.46 – Histograma de frequência dos erros (a) e dos erros padronizados da variável eteno (b), ajustado ao modelo esférico..... 123

LISTA DE TABELAS

Tabela 5.1 – Análise estatística descritiva das amostras originais (ppm).....	66
Tabela 5.2 – Testes formais de normalidade Shapiro-Wilk (S-W) e Kolmogorov-Smirnov (K-S) para os dados originais.....	68
Tabela 5.3 – Análise estatística descritiva das amostras transformadas (ln).....	74
Tabela 5.4 – Testes formais de normalidade Shapiro-Wilk (S-W) e Kolmogorov-Smirnov (K-S) para os dados transformados.....	75
Tabela 5.5 - Testes formais de normalidade Shapiro-Wilk (S-W) e Kolmogorov-Smirnov (K-S) para dados originais e transformados (ln, log10 e Sqrt).....	80
Tabela 5.6 – Média, desvio padrão (DP), mediana, desvio mediano absoluto (DMA), para os dados originais e transformados (ln) e resultados da definição do limiar superior via: [média+2 desvios padrão (dp)], [mediana (med) + 2 desvios medianos absolutos (dma)], o <i>boxplot</i> e [98ºpercentil].....	83
Tabela 5.7 – Quantificação de valores outliers amostrados utilizando o limiar superior via: [média+2 desvios padrão (dp)], [mediana (med) + 2 desvios medianos absolutos (dma)], o <i>boxplot</i> [upper whisker] e [98º percentil].....	84
Tabela 5.8 - Parâmetros e modelos dos semivariogramas para a variável metano (ln).....	103
Tabela 5.9 - Parâmetros e modelos dos semivariogramas para a variável lnC ₂ C ₅	105
Tabela 5.10 - Parâmetros e modelos dos semivariogramas para a variável eteno.....	107
Tabela 5.11 – Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável metano.....	115

Tabela 5.12 – Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável C_2C_5 118

Tabela 5.13 – Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável eteno..... 121

Tabela 5.14 – Valores do quadrado médio do erro para interpoladores usados na distribuição espacial das variáveis C_1 , C_2C_5 e $C_2^=$ 125

CAPÍTULO 1 - INTRODUÇÃO

1.1 Motivação científica para a realização do trabalho

A pesquisa petrolífera é uma atividade complexa, tanto do ponto de vista científico quanto tecnológico, que exige equipes multidisciplinares e o emprego de diferentes tecnologias. Em áreas de fronteira exploratória, raramente uma descoberta de óleo ou gás pode ser creditada ao uso de uma única ferramenta, mas resulta de diferentes contribuições que se complementam e se aperfeiçoam continuamente.

Atividades e estudos que visam determinar o potencial petrolífero de uma determinada região, isto é, a sua capacidade produtiva e a valoração de suas reservas de óleo e gás é um fator de grande interesse da indústria petrolífera.

Dentre as técnicas, o método de prospecção geoquímica de superfície tem sido empregado com sucesso como ferramenta de suporte a exploração de petróleo, seja proporcionando uma melhor avaliação do potencial petrolífero de áreas ou bacias sedimentares, ou permitindo uma análise mais criteriosa do risco geológico de *leads* e prospectos. Este método de prospecção consiste basicamente na detecção de microexsudações de hidrocarbonetos gasosos em amostras de solo.

Na última década, houve uma grande evolução na sensibilidade dos equipamentos laboratoriais propiciando a detecção de concentrações diminutas de hidrocarbonetos. Essa tecnologia permitiu também o aprimoramento das técnicas de amostragem que tornou mais fácil e simplificada a coleta (de quantidades relativamente pequenas de gases à baixa pressão), o transporte (com menor perigo de acidentes) e pronta análise dos gases. Contudo, pouco tem sido feito no sentido do desenvolvimento das técnicas interpretativas, tendo se limitado a aplicação de técnicas estatísticas básicas, desconsiderando a dependência espacial das variáveis.

1.2 Objetivos

Este trabalho tem o objetivo de testar a premissa que os dados geoquímicos de gases, seguem uma distribuição lognormal ou normal conforme descrito na literatura.

Testar diferentes abordagens para definição de dados de *background* e anômalos, sendo uma delas a regra [média \pm 2 desvios padrão] descrita por Hawkes e Webb (1962) e tradicionalmente utilizada até os dias de hoje.

Aplicar a metodologia geoestatística para identificar a magnitude da variabilidade espacial dos dados de hidrocarbonetos gasosos, as características qualitativas ligadas à estrutura do fenômeno natural que elas representam, tais como: a localização, a continuidade espacial, a anisotropia e a estrutura de correlação existente entre valores tomados em dois pontos adjacentes no espaço. No caso de dependência espacial, comparar as interpolações pelos métodos: krigagem geoestatística, inverso do quadrado da distância e mínima curvatura.

1.3 Organização da dissertação

Esta dissertação está inicialmente constituída de três capítulos teóricos, um capítulo de resultados e um último capítulo com a conclusão do trabalho de pesquisa. A organização deste trabalho de pesquisa, com a descrição do conteúdo de cada capítulo, está descrita resumidamente abaixo:

- O Capítulo 1 descreve uma introdução da tese. A ênfase está na descrição dos objetivos e na motivação encontrada para a escolha do trabalho de tese. E por último, de que forma está constituída a dissertação, que corresponde a esta seção.
- O Capítulo 2 descreve a geoquímica de superfície, mostrando inicialmente um histórico, suas principais definições, princípios, processos que envolvem a migração dos gases, objetivos, os métodos de prospecção diretos e indiretos mais usados na análise geoquímica e suas principais características, métodos de determinação de limiar, no qual separa dados *backgrounds* de anômalos, tendo uma importância relevante por se tratar de parte do objetivo desta dissertação de mestrado. E, finalmente, alguns benefícios que esta pode trazer para a indústria de petróleo e gás.
- O Capítulo 3 aborda os principais métodos de interpolação determinísticos e geoestatístico, bem como a metodologia geoestatística; contempla as teorias de amostragem clássica e geoestatística; define as variáveis regionalizadas, a modelagem probabilística e estacionaridade, os modelos teóricos de semivariograma mais utilizados, a estimação geoestatística e, por último, descreve dois tipos de krigagem e a validação cruzada.

□ O Capítulo 4 descreve os materiais e métodos, mostrando a localização e a distribuição das amostras em campo, os parâmetros da análise exploratória descritiva e espacial, os métodos de determinação de limiares, cada qual resultando em diferentes limiares, e, por fim, a análise geoestatística e a comparação entre os métodos de interpolação abordados.

□ O Capítulo 5 mostra os resultados e discussões, envolvendo as fases abordadas no capítulo 4. No estudo dos resultados, faz-se uma descrição do banco de dados, avaliam-se os resultados das aplicações das metodologias, utilizando-se figuras, gráficos, tabelas e mapas, no final, discutem-se os resultados.

□ O Capítulo 6 corresponde as conclusões do trabalho, incluindo as principais dificuldades encontradas para a realização do presente trabalho e algumas direções futuras a serem tomadas.

CAPÍTULO 2 - GEOQUÍMICA DE SUPERFÍCIE

2.1 Introdução

Desde os primórdios da exploração geoquímica de petróleo, exsudações (*seeps*) de óleo e gás têm sido usadas como excelentes indicadores de acumulações de hidrocarbonetos em subsuperfície. A primeira experiência a relatar a concentração de gás de hidrocarbonetos no solo devido a acumulação de gás e óleo em subsuperfície foi realizado em 1929 na Alemanha por Laubmeyer (1933). Avaliando uma acumulação de óleo conhecida, ele coletou amostras de gás no solo de perfurações sistematicamente localizadas um a dois metros de profundidade. Usando equipamento analítico portátil, ele demonstrou que amostras acima das acumulações estavam enriquecidas em metano (JONES III *et al.*, 1999).

Investigações da presença de gás no solo foram iniciadas logo após na então União Soviética por Sokolov (1933, 1971), que verificou os resultados de Laubmeyer (KARTSEV *et al.*, 1959), mas medindo tanto metano como hidrocarbonetos mais pesados. Ao apresentar uma metodologia de exploração geoquímica terrestre, na academia de ciências da Rússia, em 1929, determinou o marco inicial do emprego da geoquímica de petróleo como ferramenta auxiliar na pesquisa de hidrocarbonetos.

A primeira publicação sobre prospecção geoquímica de petróleo nos Estados Unidos foi publicada por Teplitz e Rogers (1935). Sendo seguida por publicações de Rosaire (1938, 1940), Horvitz (1939, 1969), Ransone (1947), Pirson (1960) e muito outros (JONES & DROZD, 1983).

No Brasil, em 1969, o geólogo Waldir Resende, de acordo com Babinski e Mello (1984) realizou um projeto pioneiro de geoquímica de superfície na área de Juruá (Alto Amazonas), obtendo excelentes resultados.

Contudo, somente nas últimas décadas tem havido um interesse renovado na geoquímica de superfície, devido ao desenvolvimento de novos métodos analíticos e de interpretação, que vem produzindo um novo corpo de dados e *insight* sobre a geoquímica de exploração (ROCHA, 2005).

A aplicação da geoquímica de superfície na exploração de petróleo tem sido amplamente baseada na detecção direta de hidrocarbonetos leves correspondente a

observações visíveis de exsudações de óleo e gás, denominadas de macroexsudações (*macroseepage*) ou por medidas da reação de produtos de hidrocarbonetos próximos à superfície resultando em microexsudações (*microseepage*) (KLUSMAN, 2002).

A exsudação gasosa de hidrocarbonetos em superfície é um processo já documentado em várias partes do mundo. Os reservatórios de petróleo não são completamente selados, permitindo escapes de fluidos, principalmente gases que, por processos de migração, emanam na superfície formando exsudações gasosas (ALMEIDA-FILHO, 2001).

De acordo com Van der Meer (2002), o processo de escape de hidrocarbonetos ocorre, principalmente, por efusão na forma livre resultante de altas pressões diferenciais nos reservatórios. A migração ocorre por várias formas, através de falhas, fraturas, planos de acamamentos e de rochas com altas porosidades, propiciando o alcance dos fluidos até a superfície.

Indicações em superfície de exsudações de óleo e gás têm sido observadas por milhares de anos e têm liderado a descoberta de importantes áreas que produzem petróleo. Embora a descoberta de uma anomalia de geoquímica de superfície não garanta a descoberta comercialmente significativa de petróleo, esta anomalia estabelece a presença de hidrocarbonetos na área de interesse. As exsudações de hidrocarbonetos em superfície representam o final do caminho de migração (SCHUMACHER). Segundo Davidson (2004), de acordo com a evolução da geoquímica de superfície de exploração pode-se inferir que o melhor uso da geoquímica de superfície é determinar onde não perfurar um poço.

Estas anomalias podem representar concentrações de hidrocarbonetos presentes nos sedimentos e águas, anomalias microbiológicas e botânicas, mudanças mineralógicas e alterações elétricas, magnéticas e propriedades sísmicas próximas à superfície, bem como sedimentos deposicionais (MELLO, 1984).

2.2 Definição e princípios da exploração

A exploração através da geoquímica de superfície investiga a presença de hidrocarbonetos quimicamente identificáveis e seus produtos de alteração em

superfície ou próximo à superfície dos quais os mesmos são indicadores para a localização de acumulações de gás e óleo.

Estudos geoquímicos de fato documentam que exsudações de hidrocarbonetos originados a partir de acumulações de óleo e gás seguem alguns princípios básicos como:

- (1) Os hidrocarbonetos migram em direção à superfície oriunda de acumulações em subsuperfície ou de rochas geradoras que reconhecidamente podem ser detectáveis próximas à superfície;
- (2) são dinâmicos, têm os selos imperfeitos e respondem as mudanças do reservatório;
- (3) podem ser visíveis (macroexsudações) em superfície ou detectadas analiticamente (microexsudações);
- (4) os hidrocarbonetos podem se mover através dos estratos em um tempo relativamente curto;
- (5) a migração é principalmente vertical, mas também pode ser lateral, cobrindo grandes distâncias;
- (6) a relação entre anomalias superficiais e as acumulações em subsuperfície variam desde as mais simples as mais complexas, com relação direta com a complexidade geológica estrutural;
- (7) a presença de macroexsudações ou microexsudações de hidrocarbonetos é uma evidência direta de geração de hidrocarbonetos em subsuperfície (SCHUMACHER, 2000).

2.3 Processos de migração e exsudação

Parte-se do princípio que acumulações significativas de gás e/ou óleo em subsuperfície podem seguir em direção à superfície, por meio de falhas, fraturas, planos de acamamento ou por efusão (diferencial de pressão) e difusão (diferencial de concentração), onde ocorre a emanção de hidrocarbonetos (BABINSKI & MELLO, 1984). As exsudações de gás e óleo, na maioria das vezes, não ocorrem diretamente acima de prospectos petrolíferos. São encontradas no final de rotas de migração, deslocadas dezenas ou até centenas de quilômetros do posicionamento geográfico identificado para a acumulação. Logo, para a utilização das exsudações de óleo e gás na análise de bacias devem-se avaliar as possíveis rotas de migração lateral e vertical, no intuito de correlacioná-las com acumulações em subsuperfície (THRASHER *et al.*,

1996), todavia não são suficientes para que se possa classificar uma área como economicamente viável ou não (SCHUMACHER, 2002).

Diversos tipos de exsudações ocorrem em bacias que geram hidrocarbonetos ativamente. Exsudações ativas são facilmente detectáveis pela maioria dos métodos de amostragem geoquímica. As áreas onde os hidrocarbonetos em subsuperfície não formam exsudações ativas são caracterizadas por exsudações passivas. Estas ocorrem em bacias onde a geração de hidrocarbonetos é restrita ou a migração é esporádica ou inibida por uma barreira de migração.

Segundo Link (1952), uma exsudação de hidrocarbonetos é definida como um local na atual superfície terrestre onde hidrocarbonetos líquidos ou gasosos emanam, ou já emanaram, naturalmente e podem ser prontamente observados. De acordo com Abrams *et al.* (2001), exsudações com essa característica são definidas como macroexsudações por apresentarem hidrocarbonetos visíveis e em grandes concentrações. Os mesmos autores propõem o termo microexsudações para as emanações de hidrocarbonetos não identificadas visualmente, mas detectadas mediante análises geoquímicas ou efeitos induzidos em solos e sedimentos.

As taxas de microexsudações e as concentrações de hidrocarbonetos em superfície podem variar significativamente com o tempo. Estas exsudações de hidrocarbonetos em superfície e as anomalias de geoquímica no solo aparecem e desaparecem em curtos espaços de tempo, em semanas, meses e anos. Observações empíricas e simulações computacionais sugerem que o mecanismo de microexsudações é flutuante e o fluxo de gás de fase contínua passa através dos poros das rochas e das fraturas.

As microexsudações de hidrocarbonetos, imperceptíveis aos nossos sentidos, podem ser registradas em análises especiais e são basicamente compostas por metano (CH_4), etano (C_2H_6), propano (C_3H_8), butano (C_4H_{10}) e pentano (C_5H_{12}), interação com o material da coluna estratigráfica e geram em superfície uma ampla variedade de alterações de natureza física, química, mineralógica, botânica e microbiológica. A percepção de anomalias de geoquímica de superfície tem sido usada como uma ferramenta de exploração para identificar novos reservatórios de hidrocarbonetos (LEIFER *et al.*, 2003).

Assim, os métodos de exploração de superfície assumem que os hidrocarbonetos migram em uma direção preferencialmente vertical a partir das rochas geradoras e reservatórios até a superfície e são controlados por pelo menos três mecanismos de exsudação (MATTHEWS, 1996a), descritos a seguir:

2.3.1 Migração por efusão

Acredita-se que o transporte por efusão seja o modo dominante na movimentação hidrocarbonetos para o reservatório e para o ambiente próximo à superfície. Características naturalmente localizadas em muitas anomalias associadas com microexsudações e macroexsudações são mais consistentes com um modelo de efusão do que com um modelo de difusão. Segundo Duchschereer (1980), o processo de efusão está mais associado com macroexsudações e o processo migratório por difusão com microexsudações. Conhecimentos adquiridos pela monitoração de emanações sobre reservatórios de armazenamento de gás, onde o gás sob pressão em subsuperfície era típico de reservatórios verdadeiros, mostrou taxas de transporte vertical de vários metros (dezenas) por dia, claramente maior que as distâncias de migração ditadas somente pelo mecanismo de difusão (JONES & THUNE, 1982).

A característica e freqüentemente a natureza linear das anomalias sugerem que falhas e fraturas são partes importantes no movimento destes gases. De acordo com Zorkin *et al.* (1977, apud JONES & BURTELL), a magnitude dos valores de gás em superfície próximos as falhas aumenta drasticamente logo após um terremoto em que um movimento de falhas foi envolvido. Um extensivo estudo, envolvendo a monitoração de 105 de poços, de 3 metros a 5 metros de profundidade, foi realizado no campo de óleo de Mulchto, nordeste de Salkhalin, na Rússia. Foram analisadas 3.700 amostras, que foram coletadas por um período de quatro meses nos poços mais ativos. Os resultados deste estudo forneceram uma forte evidência da relação tectônica com fluxo de gás nas emanações. Este estudo não deixa nenhuma dúvida que falhas e fraturas fornecem o controle principal na efusão de gases da subsuperfície.

2.3.2 Migração por difusão

Baseia-se no movimento das moléculas de hidrocarbonetos, independentemente do fluxo de água dos poros das rochas, ocorrendo de regiões de alta concentração para regiões de mais baixa concentração ou de regiões de alto potencial químico para as

regiões de baixo potencial. Geralmente é um fenômeno espontâneo e irreversível. O fluxo dos componentes do petróleo por difusão é proporcional ao seu gradiente de concentração. O tamanho das moléculas também é fator determinante no transporte por difusão, quanto menor as moléculas, maior a difusão (TISSOT & WELT, 1984; HUNT, 1996).

Portanto, para que ocorra o transporte por difusão deve-se ter um gradiente de concentração. Tais condições existem dentro da rocha geradora, próximo à superfície de contato com a rocha carreadora, acima ou abaixo desta, ao redor das fraturas e falhas (Figura 2.1). Essa forma de migração produz microexsudações (TISSOT & WELT, 1984).

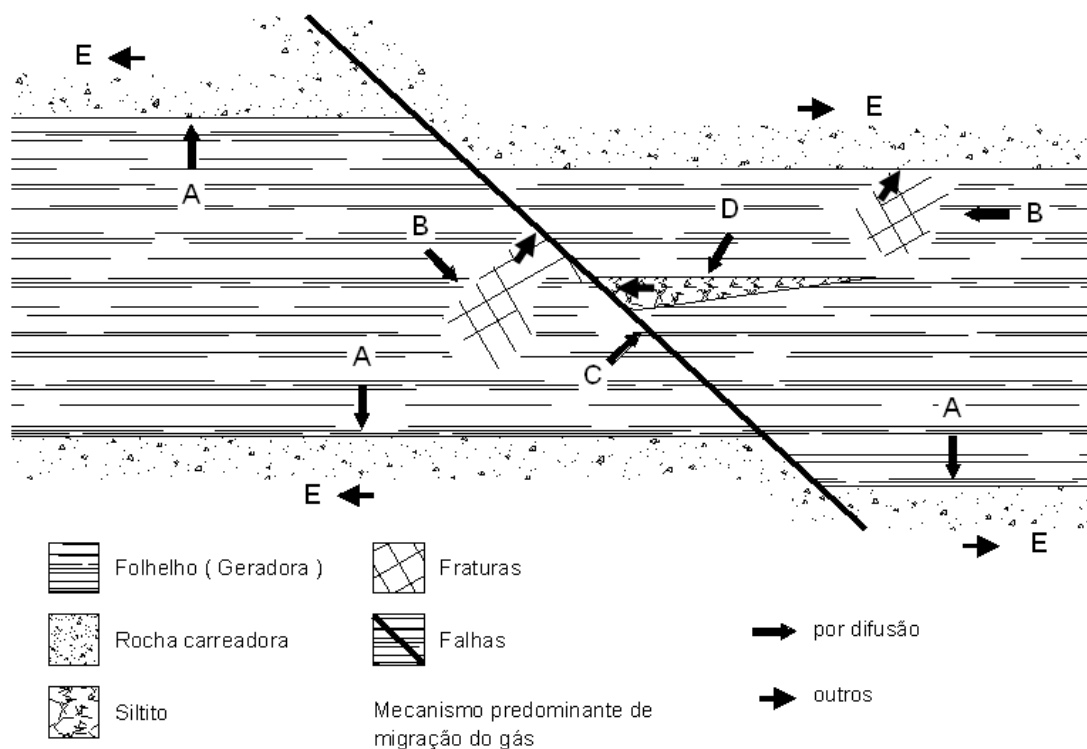


Figura 2.1 – Fluxo dos caminhos e direções da difusão de hidrocarbonetos leves em folhelhos geradores. A) em direção à rocha carreadora; B) em direção à zona de fratura; C) em direção à falha; D) em direção ao final da camada de siltito (modificado de LEYTHAEUSER *et al.*, 1982).

2.3.3 Migração por movimento

A movimentação vertical de hidrocarbonetos leves (baixo peso molecular) ocorre através da dissolução em água e através da rocha selante como resultado do potencial hidrodinâmico ou químico (DUCHSCHERER, 1980). Matthews (1996a), concluiu que a hipótese mais razoável que envolve este processo de migração vertical de microbolhas (tamanho coloidal) de gás é que a migração ocorre por uma rede interconectada, através de microfraturas preenchida com água subterrânea. As bolhas flutuantes de gás de tamanho coloidal são deslocadas prontamente para cima em taxas de até vários milímetros por segundo. Esta rápida migração explica o desenvolvimento rápido de anomalias de hidrocarbonetos leves na superfície diretamente sobre os reservatórios formados e também seu rápido desaparecimento com o esgotamento deste reservatório. Geralmente as microexsudações são invisíveis e em tais níveis de concentração que não podem ser distinguidos sem métodos analíticos modernos.

Hidrocarbonetos pesados (alto peso molecular) podem migrar por processos tal como água de lavagem (quando dissolvidas em água subterrânea) ou por compressão causada pelo colapso dos espaços dos poros como resultado da compactação (MATTHEWS, 1996a).

2.4 Objetivos da geoquímica de superfície

2.4.1 Exploração

O principal objetivo de um programa de exploração geoquímica é estabelecer a presença e distribuição de hidrocarbonetos quimicamente identificáveis e seus produtos de alteração em superfície suficientemente acima do normal para ser chamada de anomalia, sendo os mesmos indicadores para a localização de acumulações de gás e óleo. Sendo mais importante determinar o provável potencial de hidrocarbonetos, bem como favorecer a compreensão a respeito da qualidade, maturidade e rotas de migração.

2.4.2 Reconhecimento

A presença de macroexsudações ou microexsudações de hidrocarbonetos provém de uma evidência direta da geração de hidrocarbonetos termogênicos, tornando evidente

a presença de um sistema petrolífero ativo e sendo possível identificar as porções da bacia que são mais prospectivas. Adicionalmente, a composição química destas exsudações pode determinar quais áreas são mais propensas para a geração de gás ou óleo (SCHUMACHER, 2000).

2.4.3 Potencial exploratório

Os resultados de um programa geoquímico de superfície podem nos levar a avaliar melhor o risco exploratório, identificando aqueles prospectos associados com fortes anomalias geoquímicas e ressaltando os prospectos com base em sua possível carga de hidrocarbonetos.

2.5 Métodos de prospecção de geoquímica de superfície

Nos últimos anos cada vez mais tem sido valorizado o estudo da geoquímica de superfície em grandes campanhas. Os métodos de prospecção na geoquímica de superfície baseiam-se na detecção das causas e efeitos físicos, químicos ou biológicos decorrentes da presença dos hidrocarbonetos na superfície. Os métodos de prospecção de geoquímica de superfície podem ser subdivididos em diretos, indiretos e métodos não convencionais.

2.5.1 Métodos diretos

Tais métodos permitem a detecção e quantificação de hidrocarbonetos que estão presentes na superfície ou muito perto da superfície dos solos, sedimentos e rochas. Baseado no tipo de hidrocarbonetos e na sua ocorrência, eles podem ser classificados em quatro tipos principais:

(1) Gases de hidrocarbonetos livres dentro do solo, rocha ou sedimento. O método inclui coleta e concentração direta.

(2) Gases de hidrocarbonetos absorvidos, ligados ao solo, rocha ou sedimento. O método inclui amostragem direta e processos mecânicos e/ou químicos para ocorrer liberação induzida dos gases antes da concentração.

(3) Gases de hidrocarbonetos absorvidos e retidos em cristais e partículas minerais nos solos, sedimentos e rochas. O método inclui amostragem direta e

processos mecânicos e químicos (extração ácida), induzindo liberação antes da concentração.

(4) Hidrocarbonetos sólidos, gasosos e líquidos retidos nos solos, sedimentos e rochas ou dissolvidos em corpos d'água. O método inclui amostragem direta, extração com solventes e hidroquímicos, métodos para induzir liberação antes de concentração.

2.5.2 Métodos indiretos

São métodos que detectam distúrbios químicos, físicos e biológicos causados por concentrações de hidrocarbonetos na superfície. Eles incluem detecção e quantificação de elementos inorgânicos e / ou compostos. Os mais importantes são:

- Potenciais redox;
- Radiação de superfície gama, com base em levantamento radiométrico;
- Gases nobres no solo, baseados em concentração direta;
- Elementos-traço do solo e mineralogia, baseados em conjuntos minerais e concentrações;
- Hidroquímicos inorgânicos, baseados em sais dissolvidos na água.

2.5.3 Métodos não convencionais

Consistem na detecção de hidrocarbonetos através da associação indireta, utilizando métodos como:

- Análises microbiológicas baseadas na distribuição da população bacteriana;
- Método de sensoriamento remoto baseado no comportamento espectral do solo ou vegetação;
- Método geobotânico baseado na diversidade e concentração de espécimes vegetais.

As identificações de respostas secundárias geradas por emanações de hidrocarbonetos na superfície têm grande importância na prospecção da geoquímica e têm sido reportadas por vários pesquisadores. Estas incluem: (1) o estudo da

microbiologia no solo (SOLI, 1954; 1957; KARTSEV *et al.*, 1959; SEALEY, 1974a, 1974b); (2) efeito redutor (PIRSON *et al.*, 1969; DONOVAN, 1974; FERGUSON, 1975); (3) isótopos de carbono e oxigênio (DONOVAN *et al.*, 1974) e muitos outros efeitos conforme levantados por Matthews (1985) (JONES III *et al.*, 1999).

O método de prospecção geoquímica ideal é o que combina, pelo menos, duas técnicas. A utilização dessa abordagem aumenta muito o valor da prospecção de geoquímica de superfície.

No entanto, os métodos que diretamente medem hidrocarbonetos no solo ou na atmosfera do solo têm maior aceitação (KARTSEV *et al.*, 1959).

2.6 Reconhecimentos de anomalias

Microexsudações de hidrocarbonetos são dados inerentemente ruidosos e exige uma densidade de amostragem adequada para distinguir entre uma área anômala e *background*. As principais causas de ambigüidade e falhas de interpretação envolvendo estudos de geoquímica de superfície são atribuídas à forma de amostragem e/ou seleção de um método de pesquisa inadequado.

Para otimizar o reconhecimento de uma anomalia, o padrão de amostragem e o número de amostras devem levar em consideração os objetivos da pesquisa, o tamanho e a forma esperada da anomalia (ou objetivo geológico), a variação natural prevista nas medidas em superfície, e uma provável relação índice / ruído (MATTHEWS, 1996b). Definir a população de *background* é uma parte essencial para o reconhecimento e delineamento da anomalia. Matthews (1996b) sugere que para a avaliação de um prospecto, 80% das amostras coletadas deveriam ser obtidas fora da área de interesse.

A população amostrada ou amostra estatística engloba amostras individuais ou medidas realizadas no campo. A população amostrada será então muito menor que a população total, requerendo que a primeira seja representativa da última.

Para pesquisas corretamente projetadas, e sob condições geológicas ideais, uma extensa área de anomalias geoquímicas pode aproximar-se muito dos limites produtivos do reservatório em profundidade. Uma anomalia não necessariamente implica em uma acumulação econômica de petróleo.

Nos levantamentos de geoquímica de superfície, a separação de amostras com anomalias de amostras *background* é uma das mais críticas. Não existe uma proporção de amostras com anomalias a serem distinguidas a partir de amostras *background*. A prática comum de considerar amostras maiores que um, dois ou três desvios padrão da média como anomalias existenciais, será abordada neste trabalho, bem como outros métodos adotados para determinar o limiar. Este limiar ou fronteira entre anomalias e *background*, deve ser determinado objetivamente usando os dados disponíveis (ROCHA, 2005) e conhecimentos interdisciplinares.

A abundância normal de um elemento na terra é comumente referida como *background*. Para qualquer elemento em particular, o valor de *background* é variável de acordo com a natureza do material e região em qual ocorre. Portanto, a distribuição de um elemento em qualquer classe particular de material é raramente uniforme. Deste modo, é usualmente mais realístico mostrar *background* com uma variação em lugar de um valor absoluto. Isso é verdadeiro até em um ambiente relativamente uniforme. A natureza do próprio ambiente, portanto, pode também ter notadamente influência na distribuição, que debaixo de certas condições, alguns elementos podem ser enriquecidos e outros empobrecidos. Conseqüentemente, qualquer material da terra é envolvido (HAWKES & WEBB, 1962).

2.7 Métodos de determinação de *background* e Anomalias

A maioria dos métodos estatísticos é largamente fundamentada na suposição que os dados geoquímicos mostram uma distribuição lognormal ou normal. A fundamentação dessa distribuição iniciou através de Ahrens (1953, 1954a, 1954b, 1957, apud REIMANN & FILZMOSER, 1999) e Shaw (1961, apud LICHT & FRIZZO, 2007), que propôs que a distribuição de elementos-traço em materiais naturais tende ser mais próxima de uma distribuição lognormal que normal. Deste modo, os logaritmos dos dados tendem a ser distribuídos simetricamente sobre a média geométrica, já em uma distribuição simétrica de valores naturais (brutos) são distribuídos simetricamente sobre a média aritmética. Assim temos na verdade dois tipos de distribuição derivadas da curva de Gauss: a normal, quando os dados originais têm essa conformação e a lognormal, quando a conformação é apresentada pelos dados transformados em logaritmos. Esta suposição foi vastamente criticada na época (e.g: AUBREY 1954, 1956; CHAYES 1954; MILLER & GOLDBERG 1955; VISTELIUS 1960), e segundo Reimann e Filmoser (1999), o dano estava feito. Estava postulada a lognormalidade como a primeira lei da geoquímica (REIMANN & FILZMOSER, 1999).

Na literatura observam-se diferentes metodologias (e.g., HUBER, 1981; HAMPLE *et al.*, 1986; ROUSSEUW & LEROY, 1987; BARNETT & LEWIS, 1994; DUTTER *et al.*, 2003; SINCLAIR, 1974, 1976, 1991; GOVETT *et al.*, 1975; GARRETT, 1989; MIESCH, 1981; STANLEY, 1988; CHEN *et al.*, 1996, 1999) utilizadas para identificar o valor limite em dados geoquímicos que separa os dados de *background* de anômalos.

Tradicionalmente na geoquímica, baixos valores, ou baixos *outliers*, não têm sido visto como tão importantes quanto altos valores. Segundo Reimann *et al.* (2005), esse conceito é errôneo, os baixos valores *outliers* podem ser muito importantes. Na exploração geoquímica eles podem indicar zonas alteradas (depleção de certos elementos) relacionadas a acumulações próximas.

Neste trabalho abordaremos os métodos que são, ou têm sido, rotineiramente usados em geoquímica como valores limites que separam os dados de *background* de anômalos: a regra [média (\bar{x}) \pm 2 desvios padrão (dp)], *boxplot*, [mediana \pm 2 desvios medianos absolutos (dma)] e 98º percentil.

1) [média (\bar{x}) \pm 2 desvios padrão (dp)]

A definição de dados *background* e anômalos é uma questão fundamental e de bastante controvérsia na exploração geoquímica. Anomalia geoquímica é definida como uma região onde as concentrações dos elementos de interesses são maiores que certo valor limite. Sendo este por essa metodologia determinado baseado em parâmetros estatísticos, tais como média e desvio padrão (LI *et al.*, 2002).

No início da década de 60, Hawkes e Webb (1962) realizaram estudos que demonstraram que os dados geoquímicos coletados geralmente eram distribuídos lognormalmente, conforme estudos de Ahrens (1953, 1954a, 1954b, 1957), sugerindo para estimar o possível limiar em distribuições gaussianas como [média+2 desvios padrão (dp)]. Os dados com valores abaixo desse limiar são considerados de *background* e dados acima do limiar, de anômalos.

O problema com a aplicação desta regra [média (\bar{x}) \pm 2 desvios padrão (dp)] é que só é válida para os parâmetros de uma população verdadeira. O cálculo da média e desvio padrão da amostra é usado para estimar a média e desvio padrão da população. Mas estes cálculos estimativos são extremamente influenciados por

valores extremos, derivados da distribuição dos dados ou de uma segunda distribuição (contaminação). Segundo Reimann *et al.* (2005), a implicação é que a regra [média $(\bar{x}) \pm 2$ desvios padrão (dp)] não é, e nunca foi válida. Para evitar este problema extremo, isto é, “óbvios” *outliers* são muitas vezes removidos dos dados previamente antes do cálculo, no qual introduzem subjetividade. Do ponto de vista estatístico, estas “*outliers*” deveriam ser removidas, no entanto, do ponto de vista geoquímico estes valores extremos podem ter significado, constituindo anomalias geoquímicas. Outro método utilizado é a logtransformação dos dados (ln ou log10) para minimizar a influência dos dados *outliers* (REIMANN & FILZMOSER, 1999) e então calculada novamente a média e desvio padrão, depois esses valores são antilogados e utilizados na regra [média geométrica ± 2 desvios padrão geométrico].

Segundo Reimann e Filzmoser (1999), a transformação dos dados não aproxima de fato de uma distribuição lognormal e que quase nunca é testado essa normalidade nos dados. Quando testado, é comumente encontrado que os dados não seguem uma distribuição lognormal. Negligenciar que os dados geoquímicos não mostram nem uma distribuição normal nem lognormal levará a um resultado tendencioso ou errôneo quando utilizamos [média $(\bar{x}) \pm 2$ desvios padrão (dp)].

A regra [média $(\bar{x}) \pm 2$ desvios padrão (dp)] para estimar os valores de limiar que divide os dados *background* de anômalos, ainda continua a ser usado quase 50 anos após sua introdução. Vários trabalhos na literatura fazem uso dessa metodologia (e.g., FRIZZO & LICHT, 2007; GLÓRIA, 2002; RICHERS *et al.*, 1982; SECHMAN & DZIENIEWICZ, 2006; TOBIAS *et al.* 1996), no qual consideram que dados geoquímicos seguem uma distribuição lognormal ou normal. Segundo Li *et al.* (2002), essa regra é de uso corrente na exploração geoquímica na China.

A proposta de Hawkes e Webb (1962) para determinar o limiar em distribuições gaussianas tem variação na forma original proposta dependendo do pesquisador.

Segundo Licht (1998), é quase um dogma em prospecção geoquímica a seleção de dois limiares para a definição de anomalias: média mais 2 e 3 desvios padrão. Ao adotarmos esses critérios, estaremos selecionando sempre 2,5% e 0,1% dos teores de uma campanha de amostragem. A adoção indiscriminada desses limiares aconteceu a partir de uma leitura precipitada e da generalização apressada de alguns conceitos de Hawkes e Webb (1962).

Segundo Maranhão (1983, apud PEREIRA, 2003), o valor limiar é definido sendo a [média geométrica + 2 dp (aritmético)], onde as anomalias de 1ª ordem corresponderiam aos valores acima da média geométrica + 3 dp (aritmético) da população; as de 2ª ordem compreenderiam os valores entre o limiar e a soma da média geométrica + 3 dp (aritmético) e as anomalias de 3ª ordem seriam os valores entre *background* e o valor limiar.

Enquanto Litch e Oliveira (1989) consideraram como anomalias de 3ª ordem os teores maiores ou iguais à média geométrica multiplicada pelo desvio geométrico; as de 2ª ordem, os valores iguais ou maiores do que a média geométrica multiplicada pelo desvio geométrico ao quadrado; e os de 1ª ordem, os teores iguais ou maiores do que a média geométrica multiplicada pelo desvio geométrico elevado ao cubo.

A tradicional utilização da média e do desvio padrão na definição dos limiares de anomalias geoquímicas é uma simplificação perigosa, que pode produzir o “apagamento” de anomalias sutis, ou até mesmo criar anomalias não significativas. A esse respeito, Garret (1989) escreveu: “Eu gostaria de saber quantos autores realmente leram e entenderam Hawkes e Webb”.

Na geoquímica, a presença de valores *outliers* ou extremos são característicos nos dados de campo, ocasionando uma distribuição fortemente assimétrica, sendo a média e desvio padrão conseqüentemente aumentados, ocasionando assim um aumento exagerado da extensão de dados de *background* (DÍEZ *et al.*, 2007).

Vários métodos têm sido usados para tratar este problema, incluindo: a logtransformação dos dados para minimizar a influência dos dados *outliers* (AHRENS, 1954; HAWKES & WEBB 1962); a identificação de valores *outliers* utilizando o *boxplot* (TUKEY, 1977); a substituição da média com o valor da mediana (med), a qual não é afetada por valores *outliers* se estes são menores que a metade do conjunto de dados; a substituição do desvio padrão (dp) com o desvio mediano absoluto (dma) e o cálculo do 98º percentil.

2) *Boxplot*

Outra solução para identificar extremos valores é o *boxplot* (TUKEY, 1977). O *boxplot* é um gráfico que possibilita representar a distribuição de um conjunto de dados com

base em alguns de seus parâmetros descritivos: a mediana (q_2), o quartil inferior (q_1), o quartil superior (q_3) e da amplitude interquartil ($AIQ = q_3 - q_1$).

O *boxplot* permite avaliar a simetria dos dados, sua dispersão e a existência ou não de *outliers* nos mesmos. A linha central da caixa marca a mediana do conjunto de dados. A parte inferior da caixa é delimitada pelo quartil inferior (q_1) e a parte superior pelo quartil superior (q_3). A Figura 2.2 apresenta o *boxplot*, destacando suas principais características.

As cercas inferiores (*lower inner fence* - LIF) e superiores (*upper inner fence* - UIF) são definidas como $[LIF = q_1 - 1.5 \cdot AIQ]$ e $[UIF = q_3 + 1.5 \cdot AIQ]$, respectivamente. A haste inferior (*lower whisker*) e superior (*upper whisker*) se estende, respectivamente, do quartil inferior até o menor valor dos dados não inferior a $q_1 - 1.5 \cdot AIQ$ $\{lower\ whisker = \min [x(x < LIF)]\}$ e do quartil superior até o maior valor dos dados não superior a $q_3 + 1.5 \cdot AIQ$ $\{upper\ whisker = \max [x(x < UIF)]\}$. Os dados com valores inferiores a “*lower whisker*” e superiores a “*upper whisker*” são representados individualmente no gráfico sendo estes dados caracterizados como *outliers* ou atípicos, que são assinalados com um círculo e/ou a letra x, podem representar erros de introdução de dados, portanto devem ser corrigidos ou eliminados, ou caso façam parte do fenômeno em estudo (característica de dados geoquímicos), devem ser mantidos, assinalando-se a sua existência (REIMANN *et al.*, 2005).

No caso de dados logtransformados, os “*fences*” são calculados, e então antilogados para calcular os “*whisker*”. Neste trabalho utilizaremos $\{upper\ whisker = \max [x(x < UIF)]\}$ como valor limiar, onde valores abaixo desse valor são considerados *background* e valores superiores, anômalos. Segundo Reimann *et al.* (2005), o *boxplot* é mais informativo se o verdadeiro número de *outliers* estiver abaixo de 10%.

c) [mediana \pm 2 desvios medianos absolutos (dma)]

A maioria dos dados geoquímicos não é normalmente distribuída. Conseqüentemente, os dados são melhores representados pela mediana (med) e desvio mediano absoluto (dma). Segundo Reimann *et al.* (2005), Hampel (1974) e Díez *et al.* (2007), a mediana e desvio mediano absoluto são estimadores robustos contra valores extremos (*outliers*). A presença de *outliers* não altera o valor do desvio mediano absoluto (dma), diferentemente do desvio padrão (dp) que é muito sensível a presença de valores *outliers*.

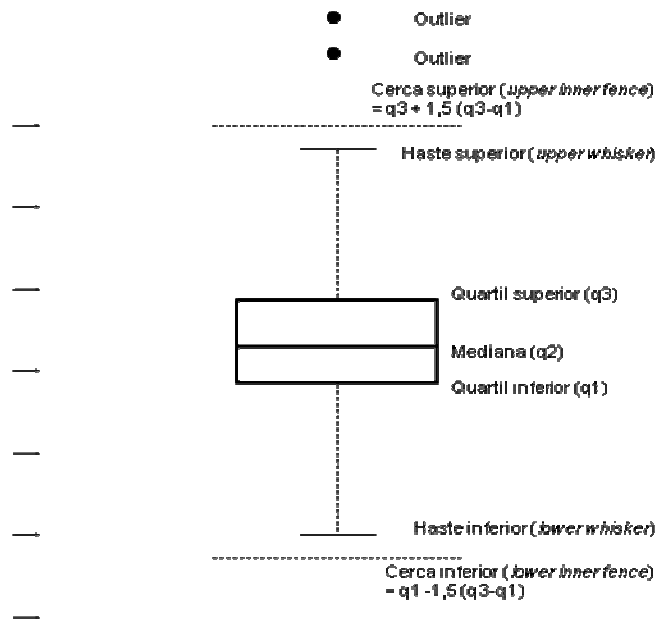


Figura 2.2 – *Boxplot* destacando suas principais características.

A mediana é o resultado que representa o ponto médio de uma série de dados ordenados, isto é, metade dos dados está acima e outra metade abaixo. Para um determinado conjunto de dados (n) ímpar, é dado por $[med = x(n+1) / 2]$, para n par $\{med = \{[x(n/2) + x(n/2) + 1] / 2\}$.

O desvio mediano absoluto (dma) é definido como a mediana (med) dos desvios absolutos da mediana de todos os dados:

$$dma = mediana [x_i - med],$$

Segundo Reimann *et al.* (2005), dentre os procedimentos abordados neste trabalho, a $[med + 2 \text{ desvios mediano absoluto } (dma)]$ é mais informativo se aplicado com a proporção de valores *outliers* acima de 15%. Segundo Díez *et al.* (2007), a substituição da média (\bar{x}) e desvio padrão (dp) pelos valores da mediana (med) e desvio mediano absoluto (dma), respectivamente, não são afetados por *outliers* se esses são menores que 50% do conjunto de dados.

d) 98º percentil

O cálculo do 98º percentil resulta na mesma percentagem de *outliers* para todas as variáveis. No 98º Percentil, 98% dos elementos estarão abaixo dele e 2% estarão acima.

Segundo Reimann *et al.* (2005), essa aproximação não é necessariamente válida, porque a real percentagem de *outliers* pode ser muito diferente. Em alguns casos, órgãos ambientais do Canadá têm usado 98º percentil como valor mais sensível para determinar o limiar ao invés da regra [média + 2dp] (e.g., Ontario Ministry of Environment and Energy, 1993).

2.8 Origens dos hidrocarbonetos gasosos

Reconhecer se uma amostra de gás natural é um gás biogênico ou termogênico pode ter implicações críticas para avaliar a presença de óleo e/ou gás em uma bacia. Como descrito a seguir, análises geoquímicas podem revelar essa origem, assim como a presença ou não de um sistema petrolífero efetivo em uma bacia.

2.8.1 Tipos de gás natural e sua importância

Hidrocarbonetos gasosos são produzidos por dois processos distintos: respiração bacteriana devido à degradação da matéria orgânica (biogênico) e alteração termal de precursores orgânicos líquidos ou sólidos (termogênico).

Gases biogênicos originam-se em regimes imaturos (baixa profundidade, temperatura e pressão) dos processos geológicos atuantes na matéria orgânica após a sua deposição, como resultado do metabolismo de bactérias anaeróbicas. O gás biogênico é muito seco, isto é, consiste quase na sua totalidade de metano (CH₄).

Na literatura, alguns autores sugerem que os compostos C₂-C₄ podem ter origem biogênica, entretanto, sólidas afirmações só são associadas à geração de metano (CH₄) e eteno (C₂H₄) como principais produtos das bactérias (MACKENNA & KALLIO, 1965 apud JONES III *et al.*, 1999; RICE, 1975).

Em contrapartida, o gás termogênico pode ser seco, ou pode conter concentrações significantes de gás úmido [etano (C₂H₆), propano (C₃H₈), butano (C₄H₁₀)] e condensado (C₅⁺). Os gases termogênicos são formados em maiores profundidades, temperaturas e pressões e podem ter sua origem ligada a geração em dois processos:

- (1) Craqueamento termal da material orgânica sedimentar em hidrocarbonetos líquidos e gasosos, onde o gás é formado durante ou imediatamente a geração de óleo (condensados), e chamado de gás termogênico “primário”;

(2) craqueamento termal do óleo a altas temperaturas em gás (gás termogênico “secundário”) e pirobetume.

Como descrito acima, gás biogênico não é relacionado com processos que formam petróleo. Como resultado, se uma amostra de gás é de origem bacteriológica, então a presença desse gás não diz nada sobre a probabilidade de existir um efetivo sistema petrolífero na bacia. Caso houvesse presença de hidrocarbonetos líquidos na mesma área, a presença seria um caso fortuito e não relacionado à origem biogênica. Em contrapartida, se uma amostra de gás for determinada como termogênica, então há possibilidade que um efetivo sistema petrolífero em subsuperfície exista (OILTRACERS).

Os gases naturais podem ser caracterizados geneticamente através da concentração e composição isotópica de carbono e hidrogênio no metano e etano. Considerando que o etano e alguns outros hidrocarbonetos gasosos mais pesados possam ser gerados através de processos mediados por bactérias, esses gases somente ocorrem em quantidades muito pequenas, pois a quantidade da fração C_{2+} em gases biogênicos é menor que 0,05 (SCHOELL, 1980), ou no caso de uso das razões C_1/C_2 quando superior a 500 (DAVIS & SQUIRES, 1954; REITSEMA *et al.*, 1978; COLEMAN, 1979) e $C_1/(C_2+C_3)$ quando superior a 1000 (BERNARD *et al.*, 1976; FRANK *et al.*, 1970; SACKETT, 1977) seriam suficientes para delinear um processo de geração de gás não termogênico.

Amostras com elevadas concentrações de gases, sendo a fração de gás úmido maior que 5,0% [$(\Sigma C_2-C_3/\Sigma C_1-C_3) \times 100$], são prováveis indicadores de um processo de geração de gás termogênico. Deste modo, a razão [$(\Sigma C_2-C_5/\Sigma C_1-C_5) \times 100$], onde contém também butano e pentano, é um parâmetro que contribui na avaliação dos processos termogênico *versus* biogênico.

Eteno, propeno e buteno pertencem à família dos hidrocarbonetos conhecidos como olefinas. Eles contêm uma ligação dupla entre dois carbonos, tal ligação significa que são hidrocarbonetos insaturados. Estes compostos são derivados principalmente da atividade bacteriológica, sendo as razões eteno / etano e propeno / propano utilizadas na avaliação de hidrocarbonetos insaturados (biogênico) *versus* saturados (termogênico) (ABRAMS, 2005)

2.9 Benefícios da geoquímica de exploração

Dentre as vantagens no emprego dos métodos geoquímicos para a pesquisa de hidrocarbonetos podemos citar a pequena profundidade de coleta de amostras e também a rapidez e a economia nas análises em comparação aos métodos geofísicos (BABINSKI, 1986). Em outras palavras, quando usada corretamente, a geoquímica de superfície é uma técnica excepcionalmente confiável em relação aos outros métodos de exploração. Nos últimos anos, os altos custos exploratórios e a visualização do grande potencial dos métodos não convencionais derrubaram as últimas barreiras em relação à geoquímica de superfície, que se tornou ferramenta exploratória de grande importância mesmo para as grandes companhias petrolíferas.

O potencial de benefícios da geoquímica de exploração de sucesso são muitos e incluem o seguinte:

- Detectar diretamente hidrocarbonetos e/ou mudanças induzidas por hidrocarbonetos em ambientes próximos a superfície ou em superfície *onshore* e *offshore*.
- Documentar a presença de sistemas petrolíferos ativos na área de interesse (rocha geradora ativa, geração de hidrocarbonetos e migração).
- Permitir uma melhor avaliação das bacias, *plays* e/ou prospectos antes de adquirir arrendamentos ou de conduzir inspeções sísmicas detalhadas. Os custos de avaliações geoquímicas são relativamente mais baixos que as inspeções sísmicas.
- Usar os dados geoquímicos para inferir parâmetros e predição de valores em locais não amostrados.
- Avaliar áreas onde as inspeções sísmicas não são práticas ou ineficazes devido a fatores geológicos ou ambientais. A maioria dos métodos de exploração geoquímica causa leves alterações no ambiente.
- Os métodos são aplicáveis tanto para trapas estratigráficas, como para trapas estruturais com a capacidade de identificar trapas invisíveis ou mal imageados com dados sísmicos (SCHUMACHER).

CAPÍTULO 3 – MÉTODOS DE INTERPOLAÇÃO E GEOESTATÍSTICA

3.1 Métodos de interpolação

Interpolar é predizer (ou estimar) o valor da variável em estudo num ponto não amostrado (LANDIM, 2002). Através deste procedimento estimam-se valores de uma variável em área interior aos pontos de amostragem disponíveis, permitindo representar em mapa, de forma contínua, o comportamento de variáveis amostradas pontualmente.

Burrough (1986) comenta que as funções de interpolação podem ser definidas como globais ou locais. A função global é dependente de todos os pontos amostrados, sendo que a alteração de um valor de entrada afeta o mapa inteiro. As funções locais são definidas para pequenas porções dos mapas, sendo aplicadas sucessivamente até cobrir toda a área do mesmo. Uma alteração num valor de entrada afeta apenas o resultado de um subconjunto. Essas regiões podem ser definidas por raios de influência, ou por quantidade de amostras vizinhas.

As funções de interpolação globais e locais podem ser classificadas em dois (2) tipos básicos: interpoladores determinísticos (e.g., vizinho mais próximo, curvatura mínima, triangulação e inverso da distância ao quadrado) e interpoladores estocásticos (e.g., Krigagem). Os valores inferidos pelos interpoladores determinísticos são tratados como dados sem erros, ou seja, os valores obtidos não estão contaminados por erros nos dados de entrada ou pelo algoritmo de interpolação. Esta é a grande desvantagem desses interpoladores em relação aos estocásticos, os quais utilizam as ferramentas da geoestatística para inferências de valores de atributos com estimativas de incertezas (FELGUEIRAS, 1999).

Os softwares mais comumente utilizados para a geração de mapas geralmente disponibilizam vários métodos de geração de *grids* para os usuários, sendo a escolha do método uma etapa importante no processo de geração de mapas. Os métodos mais comuns de interpolação local presentes nos *softwares* são:

- Inverso do quadrado da distância: É um interpolador de médias ponderadas que não é exato. Quando o ponto de uma malha é calculado, os pesos atribuídos aos pontos são fracionários, porém sua soma é um. Se qualquer observação é coincidente com um ponto da malha, a distância entre essa

observação e o ponto é zero com peso um, enquanto para as demais observações o peso é zero. É muito usado em diversas situações por ser um método rápido de interpolação (ISAACS & SRIVASTAVA, 1989). Sua equação é dada por:

$$\bar{Z} = \frac{\sum_{i=1}^n \left(\frac{1}{d_i^2} Z_i \right)}{\sum_{i=1}^n \left(\frac{1}{d_i^2} \right)}$$

Onde:

d é a distância euclidiana entre o i -ésimo ponto de vizinhança e o ponto amostrado; \bar{Z} , a variável interpolada; Z_i , valor da variável da i -ésima localidade vizinha.

Os resultados são variáveis, desde altamente tendenciosos a favor dos pontos mais próximos a resultados onde o peso é praticamente o mesmo para todos os pontos próximos.

- Vizinho mais próximo: Método que objetiva converter dados regularmente espaçados para arquivos *grid*, respondendo bem também nos casos em que dispõe de malhas regulares com poucas ausências de pontos de observação. É atribuído o valor do ponto mais próximo para cada nó do *grid*.
- Triangulação com interpolação linear: Utiliza a triangulação de Delaunay, muito utilizada para modelamento de superfície de terreno em topografia. Esse método apresenta resultados fortemente dependentes da rede de pontos de amostragem trabalhada, ou seja, uma mesma área retratada por meio de diferentes malhas de amostragem mostra padrões de distribuição diferentes para a variável estudada.
- Mínima curvatura: Esta técnica gera superfícies suavizadas. Sua superfície lembra um plano fino, linearmente elástico, que passa por cada um dos dados disponíveis com uma curvatura mínima, gerando a superfície mais suavizada possível, reproduzindo os valores da variável. Mas isso não ocorre sempre, pelo qual não pode ser considerado um interpolador exato. Produz um *grid* repetidamente pela aplicação de uma equação sobre o *grid* já existente, buscando suavizá-lo, e cada passo é considerado uma interação.

São efetuados cálculos de derivação repetidamente até que seja alcançada uma diferença (convergência ou tolerância) entre os valores amostrados e os estimados, especificada pelo usuário ou até que um número máximo de interações seja alcançado. Através desse método, artefatos (formas estranhas) são gerados especialmente no interior das células onde o valor inicial de Z é calculado utilizando a média global dos valores amostrados; ocorre quando não há valores amostrados dentro ou próximo da célula cujo nó deve ser estimado. Outra fonte geradora de artefatos é o número insuficiente de interações (ANDRIOTTI, 2003). Esse método é muito usado em ciências da terra.

- Krigagem: É um método muito recomendável para a geração de *grids*, sendo um dos métodos mais flexíveis e úteis para *grids* de qualquer tipo de dados. Segundo Landim (1998), a Krigagem é um processo de estimação de valores de variáveis distribuídas espacialmente, a partir de valores adjacentes considerados como interdependentes. Pode ser entendida como uma série de técnicas de análise de regressão que procura minimizar a variância estimada a partir de um modelo prévio que leva em conta a dependência estocástica entre os dados distribuídos nos espaço.

Os métodos de interpolação, cujos resultados apoiarão a interpretação do fenômeno estudado, exigem, para sua aplicação apropriada, um bom conhecimento da variável estudada, onde a interdisciplinaridade é de vital importância.

A diferença entre os diversos métodos está em como, a partir de valores conhecidos, podem ser estimados os valores não amostrados. A krigagem se destaca por levar em consideração a variabilidade espacial dos dados.

Com objetivo de evidenciar a diferença básica entre interpoladores determinísticos e geoestatísticos, vejamos uma situação hipotética (Figura 3.3) em que 5 pontos amostrados são utilizados para calcular o ponto A.

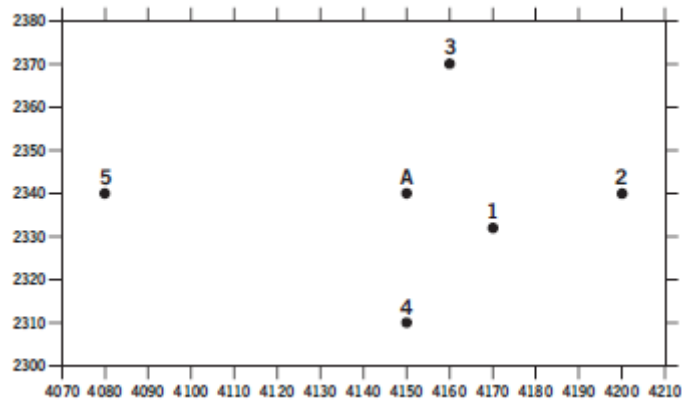


Figura 3.3 - Configuração de pontos de amostragem para estimativa no ponto A.

Neste caso, para estimar um valor desconhecido a partir de valores conhecidos pelo método inverso do quadrado da distância (determinístico) importa apenas as distâncias euclidianas para a atribuição dos pesos e isso é válido para todos os outros algoritmos, com exceção da krigagem (geoestatístico). Na krigagem há necessidade de encontrar, preliminarmente, a variabilidade espacial dos dados o que é conseguido por uma análise variográfica. Em outras palavras, antes da aplicação da krigagem é preciso calcular um semivariograma experimental e em seguida modelá-lo.

Na krigagem para a estimativa de A são necessários não apenas os valores das distâncias de cada um dos 5 pontos em relação a A, mas também as distâncias entre os pontos. Não são as distâncias euclidianas, porém, que são utilizadas, mas sim os valores resultantes da análise espacial pelo variograma modelado. A Figura 3.4 mostra as relações entre amostras (em preto) e entre as amostras e o ponto a ser estimado (em vermelho).

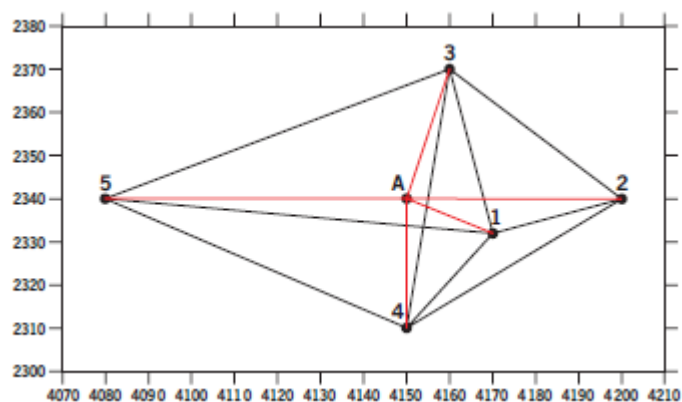


Figura 3.4 - Inter-relações entre amostras e ponto a ser estimado para cálculo das equações de krigagem.

3.2 Teoria da amostragem clássica

A teoria da amostragem clássica, diz que o valor médio de uma dada característica, em uma região amostrada, é igual ao valor da esperança desta característica em qualquer ponto dentro da região, com um erro de estimativa correspondente à variância dos dados amostrados (TRANGMAR *et al.*, 1987). Um dos princípios que a teoria de amostragem clássica se baseia é a casualização, que visa conferir a modelagem estatística um contexto conceitual de independência dos erros, minimizando assim, os efeitos da correlação espacial, possivelmente existente na área de experimentação utilizada, permitindo então a utilização da inferência clássica na análise de dados de experimento. Contudo, principalmente em experimentos que utilizam maiores extensões de campo, a casualização pode ter eficiência reduzida e duvidosa, e modelagens adicionais podem ser introduzidas. Como, por exemplo, a adoção de um método de análise que utilize a informação da posição da parcela no experimento, neste caso a geoestatística. Segundo Yates, citado por Duarte (2000) cabe a este princípio neutralizar os efeitos da correlação espacial. Assim, as posições das unidades amostrais no campo podem ser ignoradas nesta abordagem. A pressuposição desses métodos é que as variações numa dada característica, de um local para outro, são aleatórias (independentes). A teoria da amostragem clássica consiste em dois componentes, conforme a equação apresentada a seguir:

$$Z(s_i) = T(s_i) + \varepsilon(s_i)$$

Onde, $Z(s_i)$ é a variável avaliada em s_i , $T(s_i)$ é a função determinística que descreve a componente estrutural de Z em s_i , e $\varepsilon(s_i)$ é o erro aleatório. A parte $\varepsilon(s_i)$ é tão importante quanto $T(s_i)$, e a análise destes resíduos são fundamentais em todo trabalho estatístico. O que se deseja é que a parte $\varepsilon(s_i)$ não contenha nenhuma “suavidade” ou comportamento estruturado. É fundamental que ela seja aleatória. É importante ressaltar que a teoria de amostragem clássica desperdiça a estrutura de dependência espacial quando ela está presente na característica avaliada. Deve-se observar que a formulação desta teoria discerne somente o tamanho da variabilidade, através do desvio padrão, e não discerne a estrutura da variabilidade.

3.3 Teoria geoestatística

Geoestatística é um ramo da matemática que une o conceito de variáveis aleatórias com o conceito de variáveis regionalizadas, gerando um novo conceito de funções aleatórias, que são posteriormente processadas por aplicativos computacionais.

Os métodos geoestatísticos foram desenvolvidos devido aos estudos do engenheiro de minas Georges Matheron na França no final da década de 50 e início da década de 60. Estes métodos estão fundamentados na teoria das variáveis regionalizadas, que foi formalizada por Matheron a partir de estudos práticos desenvolvidos por Daniel G. Krige (1951), que ao trabalhar no cálculo de reservas nas minas de ouro do Rand na África do Sul, concluiu que somente a informação dada pela variância era insuficiente para explicar o fenômeno em estudo. Para tal, seria necessário levar em consideração a distância entre as observações. A partir daí surge o conceito de geoestatística, que leva em consideração a localização geográfica e a dependência espacial.

A geoestatística é a aplicação de métodos matemáticos e estatísticos a problemas das ciências da terra e pode ser utilizada em diversas áreas como: geologia pura (e.g., análise de elementos-traço em rochas metamórficas); em pesquisa mineral (e.g., análise de dados geoquímicos de pesquisa); em pesquisa petrolífera (e.g., caracterização de reservatórios) delimitando áreas potenciais; tradução da geologia para modelos usados em engenharia; determinação do tipo de amostragem; estimativa espacial e temporal; simulação bem como, a análise de riscos e quantificação de reservas.

As técnicas da estatística clássica assumem que todas as amostras são aleatórias e independentes de uma distribuição de probabilidade simples - esta suposição é chamada estacionaridade. Sua aplicação não envolve qualquer conhecimento da posição atual das amostras ou do relacionamento entre amostras. Já a geoestatística assume que a distribuição das diferenças de variáveis entre dois pontos amostrados é a mesma para todo o depósito, e que isto depende somente da distância entre eles e da orientação dos pontos. Em outras palavras, diferenças nas variáveis devem ser consistentes, não constantes, para o depósito - hipótese intrínseca ou quase-estacionariedade (CLARK, 2000).

Um exemplo muito simples mostra a diferença entre a estatística e a geoestatística considerando os seguintes valores:

- Amostra 1: 2 – 7 – 3 – 6 – 2 – 9 – 4 – 8 – 5
- Amostra 2: 2 – 3 – 5 – 7 – 9 – 8 – 6 – 4 – 2

Sob a ótica da estatística clássica, o valor médio e a variância são idênticos para as duas amostragens. Entretanto, segundo a avaliação no espaço, a primeira amostra possui um comportamento muito errático, enquanto a segunda amostra apresenta uma uniformidade espacial.

3.4 Variáveis regionalizadas

A teoria das variáveis regionalizadas, também denominada geoestatística por sua aplicação nos estudos de Geologia e Mineralogia, foi um importante passo no conhecimento humano. Segundo Andriotti (2003), entendem-se como variáveis regionalizadas, V.R., uma função que varia de um lugar a outro no espaço com certa aparência de continuidade. São variáveis cujos valores estão relacionados de algum modo com a posição espacial que ocupam.

A continuidade atribuída às variáveis regionalizadas se manifesta, sobretudo pela tendência de tomarem valores mais próximos, quanto menos afastados estejam dois pontos amostrados. As peculiaridades dessas variáveis ferem o pressuposto básico da técnica de estatística clássica, ou seja, a independência entre as observações. Uma medida tomada em um ponto guarda relações de dependência com medidas tomadas em pontos adjacentes sugerindo uma estrutura de correlação.

Essas variáveis portam consigo um duplo aspecto contraditório, por um lado revelam uma característica aleatória que se manifesta na irregularidade e variação imprevisível de um ponto para outro e por outro lado, manifestam-se nas variáveis regionalizadas, características estruturantes ou espaciais que refletem ligações existentes entre os pontos do espaço, ou seja, os valores numéricos observados não são inteiramente independentes.

As variáveis regionalizadas também não podem ser tratadas pela estatística clássica visto que: (i) nesta as amostras podem ser indefinidamente repetidas; (ii) as amostras são independentes, logo o resultado de uma amostra não pode ser influenciado pelo resultado daquelas amostras precedentes.

Estas imposições da estatística clássica, em geral, não são respeitadas pelas

variáveis regionalizadas, pois quando se tira uma amostra num determinado ponto (s) o teor da referida amostra é um valor único, fisicamente determinado, sendo muito difícil a repetição dessa experiência. Se fosse retirada outra amostra num ponto muitíssimo próximo do (s) considerado, poder-se-ia dizer que a condição (i) estaria satisfeita, porém neste caso não se estaria respeitando o condicionante (ii).

Segundo Guerra (1988), as variáveis regionalizadas são representadas, na prática, por certa quantidade de dados numéricos brutos disponíveis, a partir dos quais são obtidas informações sobre as características do fenômeno natural em estudo. Tais características são:

- **Localização:** Uma variável regionalizada é numericamente definida por um valor, o qual está associado a uma amostra de tamanho, forma e orientação específicos. Essas características geométricas da amostra são denominadas suporte geométrico. O suporte geométrico não necessariamente compreende volumes, podendo se referir também a áreas e linhas. Vale salientar que somente no espaço geométrico onde a variável é susceptível de tomar valores definidos e no interior do qual sua variação será estudada as variáveis regionalizadas tomam seus valores. Este espaço é denominado de campo geométrico.
- **Continuidade:** Dependendo do fenômeno observado, a variação espacial de uma variável regionalizada pode ser grande ou pequena. A existência de uma continuidade mais ou menos estável na variação de uma variável regionalizada pode ser expressa por meio de uma flutuação mais ou menos importante entre os valores de amostras vizinhas. Tal flutuação reflete o grau de dependência ou independência que existe entre um valor e outro. Quando essa continuidade é pouco definida e não pode ser confirmada, diz-se que há a presença do efeito de pepita. Quando os valores representativos das características do atributo estudado na área são totalmente independentes, trata-se de uma variável aleatória, considerada como um caso particular de variável regionalizada.
- **Anisotropia ou zonalidade:** Fenômeno que indica se os valores da variável regionalizada não apresentam variações significativas ao longo de uma direção privilegiada, apresentando por variações rápidas ou irregulares em outra(s) direção(ões).

Considerando ainda que os fenômenos naturais não possam ser tratados como fenômenos aleatórios, faz-se necessário utilizar ferramentas matemáticas que permitam estudar as duas características essenciais das variáveis regionalizadas, ou seja, aspecto aleatório e aspecto espacial.

Assim, a geoestatística se propõe estudar dois objetivos principais: ser capaz de extrair da aparente desordem dos dados disponíveis uma imagem da variabilidade dos mesmos e uma medida da correlação existente entre os valores tomados em dois pontos do espaço, separados por certo vetor h em que $|h| = h$, os quais são objetivos da análise estrutural, feita por meio do semivariograma; ser capaz de medir a precisão de toda predição ou estimativa através de dados fragmentados, ou seja, há necessidade de uma teoria de estimativa que é obtida através da Krigagem.

O estudo da dependência espacial de atributos presentes no solo, através da teoria das variáveis regionalizadas ou geoestatística, permite a interpretação e a projeção dos resultados com base na estrutura da sua variabilidade natural.

A teoria das variáveis regionalizadas pressupõe que cada dado $Z(s_i)$ é modelado como uma variável aleatória que pode ser expressa pela soma de três componentes: um componente estrutural, associado a um valor médio constante ou a uma tendência constante; um componente aleatório, espacialmente correlacionado, e um ruído aleatório ou erro residual, isto é,

$$Z(s_i) = T(s_i) + W(s_i) + \varepsilon(s_i)$$

Onde:

$T(s_i)$ é uma função determinística que descreve a componente estrutural Z em s_i ;

$W(s_i)$ é um termo estocástico, que varia localmente e depende espacialmente de $T(s_i)$;

$\varepsilon(s_i)$ é um ruído aleatório não correlacionado normalmente distribuído e com média zero e variância σ^2 .

A geoestatística atua no termo estocástico $W(s_i)$. A teoria das variáveis regionalizadas pode ser vista como um grande ramo da estatística espacial, que estuda a relação de uma variável regionalizada com ela mesma numa outra posição (CRESSIE, 1993).

Na prática, não é possível conhecer as funções de distribuição de $Z(s_i)$ em detalhe, por esse motivo, utilizam-se alguns pressupostos sobre o comportamento das funções. Normalmente, assume-se que as variáveis são descritas por funções conhecidas, do tipo gaussiano e lognormal, distribuições essas consideradas como limite de todas as funções de distribuições contínuas.

Não existem repetições das amostras nos pontos amostrados. Tais repetições não são mesmo possíveis para variáveis obtidas a partir de métodos extrativos. Assim, não é possível obter a função de distribuição no ponto e nem os momentos das variáveis $Z(s_i)$.

Uma forma de solucionar este problema é utilizar os dados em todo o campo para estimar a função de distribuição das variáveis aleatórias, isto é, assumir que o comportamento da função de distribuição local é idêntico ao da função de distribuição global. Essa abordagem nada mais é do que a aplicação da hipótese de ergodicidade.

Ergodicidade é uma propriedade que permite uma estimação consistente a partir de amostras retiradas de uma única realização para um fenômeno contínuo gerado aleatoriamente. É assumida a suposição de que uma única realização é igual à média de todas as outras possíveis realizações do processo estocástico. Isto permite usar os dados amostrais para modelar a função aleatória e realizar estimação de parâmetros não tendenciosos bem como predição de valores para a característica estudada em locais não observados da superfície (CRESSIE, 1993).

A base para a adoção dos procedimentos geoestatísticos e sua justificativa é o entendimento e o tratamento da variável em questão. Para aplicação desses procedimentos é fundamental a compreensão dos instrumentos e conceitos utilizados, tais como: modelagem probabilística e estacionaridade, descrição e modelagem da estrutura de variabilidade espacial e verificação da distribuição espacial para cada uma das variáveis obtendo mapas que representam os valores das propriedades em estudo na área.

3.5 Modelagem probabilística e estacionaridade

Considerando que em certos pontos de uma região qualquer sejam realizadas amostras de uma variável mencionada, tem-se um conjunto de dados espacialmente

distribuídos, ou seja, medidas de um atributo acompanhadas de suas respectivas coordenadas. Com a informação da localização de cada amostra é possível calcular a distância euclidiana entre os pontos observados.

Segundo Andriotti (2003), para cada ponto amostrado, posição s_i , pode-se avaliar uma determinada característica $Z(s_i)$, realização da variável aleatória Z no ponto s_i . O conjunto destas variáveis aleatórias é considerado um processo estocástico, descrito da forma:

$$Z(s_i): s_i \in D \subset \mathfrak{R}^d$$

onde:

$Z(s_i)$, a variável aleatória que varia continuamente em D ;

s_i , a posição da variável, considerada fixa;

D é a região em estudo;

\mathfrak{R}^d é o espaço d -dimensional ($d=1, 2, 3$ ou 4).

Para $d=1$ os dados estão em uma transeção, com $d=2$ em um plano, $d=3$ em um volume que se somado a variável tempo tem-se a dimensão 4.

O conjunto de valores $Z(s_i)$ para todos os pontos s_i no domínio D , isto é, a variável regionalizada $Z(s_i)$ pode ser considerada uma particular realização do conjunto de variáveis aleatórias $Z(s_i): s_i \in D \subset \mathfrak{R}^d$. Este conjunto de variáveis aleatórias é chamado função aleatória e pode ser denotado por $Z(s)$ (ANDRIOTTI, 2003).

Esta definição de função aleatória expressa os aspectos de estrutura e aleatoriedade da variável regionalizada, pois: (i) localmente, no ponto s_i , $Z(s_i)$ é uma variável aleatória; (ii) $Z(s)$ é também uma função aleatória no sentido de que, para cada par de pontos s_i e s_{i+h} as correspondentes variáveis aleatórias $Z(s_i)$ e $Z(s_{i+h})$ não são, em geral independentes, mas são relatadas pela correlação, expressando a estrutura espacial da variável regionalizada inicial $Z(s_i)$.

Portanto, um tratamento para variáveis ditas regionalizadas que apresentem aspectos aleatórios e espaciais, ditas regionalizadas deve em uma análise estrutural, avaliar a

variabilidade e correlação entre pontos fornecendo estimativas associadas a uma avaliação de sua qualidade.

A teoria das variáveis regionalizadas é uma proposta para problemas dessa natureza, onde os aspectos de localização, continuidade e anisotropia da variável regionalizada são levados em consideração.

A variável regionalizada é definida dentro de uma determinada região no espaço onde a variável é suscetível de tomar valores definidos e no interior do qual se estuda sua variação. Segundo Valente (1989), na prática a variável regionalizada não é medida em pontos do espaço, conforme conceituados pela geometria euclidiana, mas sim em suportes de dimensões finitas e perfeitamente determinadas. Por exemplo, para um teor o suporte é o volume da amostra retirada e analisada.

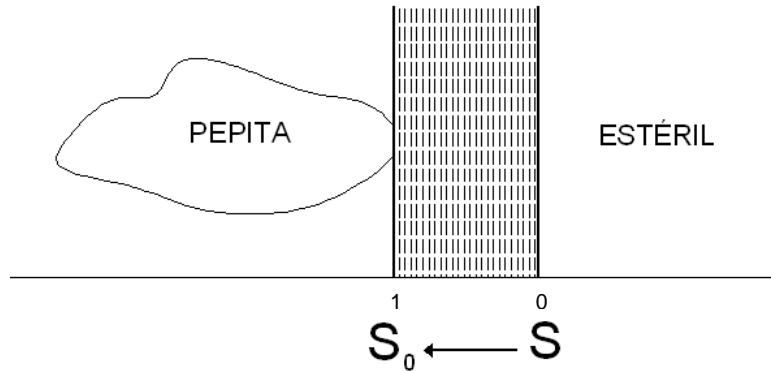
É de fundamental importância compreender que caso mude o suporte obtém-se outra regionalização diferente da primeira, até sob o ponto de vista da estatística clássica.

A variável regionalizada que se apresenta com uma continuidade mais ou menos estável em sua variação espacial pode ser expressa através de certa flutuação entre os valores de amostras vizinhas. Esta flutuação reflete o grau de dependência que existe entre um valor e outro. Apesar da complexidade das flutuações verificadas para as diversas variáveis regionalizadas, em geral, existe uma continuidade em média.

$$\lim_{s \rightarrow s_0} E[Z(s) - Z(s_0)]^2 = 0$$

Sendo a variância zero, a homogeneidade é total, e há continuidade. Algumas variáveis com um caráter geométrico marcado dependente de uma formação, refletem um alto grau de dependência entre um valor e seu adjacente, como é o caso da espessura de uma camada de carvão (VALENTE, 1989). Existem circunstâncias onde esta continuidade é pouco definida e não pode ser confirmada, tal como acontece com minerais erráticos. Então neste caso surge o denominado efeito pepita ($C_0 \neq 0$), definido da forma abaixo (Figura 3.5).

EFEITO DE PEPITA



$$\lim_{s \rightarrow s_0} E \{ [Z(s) - Z(s_0)]^2 \} = C_0 \neq 0$$

Figura 3.5 – Efeito pepita (modificado VALENTE, 1989).

Quanto menor a distância entre os pontos, maior a igualdade e a continuidade, indicando o quão são parecidas as amostras. Por último, existem casos onde os valores representativos não possuem dependência entre eles, ou seja, são totalmente independentes tratando-se neste caso de uma variável aleatória (VALENTE, 1989). O estudo dessa continuidade nos leva a considerar a variável aleatória como um caso particular de variável regionalizada.

E por fim, o aspecto de anisotropia que conforme Kitanidis (1997) refere-se à existência de uma direção privilegiada ao longo da qual os valores não variam de forma significativa, este fenômeno é também denominado de zonalidade. Já que as variações da variável regionalizada ocorrem de maneira mais acentuada segundo determinadas direções, o estudo anisotrópico para algumas características pode assumir um valor econômico, pois se pode planejar de maneira mais eficaz a implantação de malhas de amostragens.

O duplo aspecto aleatório-estruturado das variáveis regionalizadas pode ser interpretado com apoio na teoria das funções aleatórias, F.A. Uma função aleatória nada mais é que a variável aleatória a uma infinidade de componentes.

No entanto, duas características principais limitam a modelagem de fenômenos naturais, espacialmente distribuídos, por variáveis aleatórias. Uma característica é a

impossibilidade de repetir indefinidamente o experimento e a outra se refere à dependência das observações. Visto que o resultado da amostragem para cada variável aleatória é composto de uma única realização em cada ponto e, portanto de cada variável, tem-se a impossibilidade de fazer qualquer tipo de inferência sobre este processo.

Por este motivo, o estudo das variáveis regionalizadas exige que algum tipo de estacionaridade condizente com o problema em questão deva ser assumida de forma que possibilite a estimação de pelo menos os dois primeiros momentos da distribuição da variável aleatória, que em geral estão relacionados com parâmetros de interesse como a média, a correlação, a covariância e a semivariância.

A função aleatória é estacionária quando sua lei de distribuição de probabilidades é invariante por translação, isto é, a F.A. $Z(s_1), Z(s_2), \dots, Z(s_n)$ possui a mesma lei de distribuição de probabilidades que $Z(s_{1+h}), Z(s_{2+h}), \dots, Z(s_{n+h})$. Nestas condições tem-se:

$$E[Z(s)] = m(s) = E[Z(s+h)] = m = cte$$

Pois sendo $Z(s)$ uma F.A. estacionária, a lei é invariante por translação e temos então $m(s)$ para qualquer vetor h , sendo que $m(s)$ uma constante com m independente do ponto de apoio s (VALENTE, 1989).

Pode-se então, neste caso, supor $m(s) = 0$ para o que basta considerar a F.A. $Z(s) - m$, dita centrada, em vez de $Z(s)$. Segundo Isaaks & Srisvastava (1989) e Cressie (1991), a variável regionalizada, $Z(s)$, deve satisfazer as hipóteses de estacionaridade. A forma de estacionaridade usualmente assumida na análise geoestatística é a chamada hipótese intrínseca ou estacionaridade fraca e é expressa totalmente em termos das diferenças, $[Z(s_{i+h}) - Z(s_i)]$, das variáveis regionalizadas:

- (i) $E[Z(s_{i+h}) - Z(s_i)] = m(h) = 0$, para qualquer posição no espaço s_i e vetor separador h , estabelecendo que $Z(s_i)$ é constante.
- (ii) $E\{[Z(s_{i+h}) - Z(s_i)]^2\} = 2\gamma(h)$, para qualquer s_i e h .

A expressão (ii) define o valor da semivariância, $\gamma(h)$, que deve ser independente da posição dos pontos, sendo função apenas da distância entre eles. Esta hipótese

pode ser considerada um tipo de estacionaridade dos incrementos, que é formulada sob a variável aleatória

$$G(h) = [Z(s_{i+h}) - Z(s_i)],$$

a qual representa diferenças entre as variáveis separadas pela distância h . A hipótese de estacionaridade de segunda ordem ou estacionaridade forte requer:

(i) $E\{[Z(s_i)]\} = m_s$, para qualquer s , isto é, uma função aleatória independente da posição s_i , $Z(s_i)$ é constante como na hipótese intrínseca.

(ii) $E\{[Z(s_{i+h}) - m_s].[Z(s_i) - m_s]\} = \text{cov}(h)$, para qualquer s_i .

Existe uma função covariância para cada par de valores $Z(s_{i+h})$ e $Z(s_i)$ que depende apenas do vetor separador h o que implica que a variância, σ^2 , de $Z(s_i)$ não depende de s .

As ferramentas geoestatísticas utilizadas no estudo da variabilidade espacial, tais como semivariograma e autocorrelograma, são funções principalmente de distâncias especificadas. Desta maneira a vizinhança usada na estimativa torna-se um ponto importante.

3.6 Número de vizinhos das estimativas

A vizinhança usada na estimativa torna-se um ponto de extrema importância na krigagem. Vários são os métodos que podem ser utilizados para a determinação do número de vizinhos na estimativa, cada um com vantagens e desvantagens como será discutido em seguida. Qualquer que seja o critério usado para a escolha do método, deve-se levar em conta o ganho de precisão em relação ao aumento de tempo de computação (CLARK, 2000). A seguir, são apresentados os métodos de vizinhança com algumas de suas vantagens e desvantagens.

a) Vizinhança única

Quando o tamanho do conjunto de dados em termos de número de amostras disponíveis tiver tamanho razoável relativo à quantidade de memória e tempo de processamentos disponíveis no computador, pode-se usar o procedimento chamado vizinhança única. Neste procedimento todos os valores medidos são considerados

vizinhos e serão utilizados na estimativa. Deve-se sempre lembrar que a decisão de se usar vizinhança única baseia-se em sua praticidade relativa ao tamanho do conjunto de dados e não na precisão obtida na estimativa. A razão para tanto reside no alcance do semivariograma, pois os pesos associados a vizinhos separados por distâncias maiores do que o alcance não deve ter contribuição significativa no valor estimado. Outro ponto importante é que para se usar vizinhança única é necessário que o semivariograma seja definido até a maior distância existente no espaço.

A vantagem deste método reside no fato que uma vez invertida a matriz de coeficientes, então as estimativas podem ser feitas para qualquer espaçamento com um pequeno consumo de tempo de processamento de computador. Desse modo, a matriz invertida pode ser arquivada no computador e usada quantas vezes for necessário, desde que não mude o modelo do semivariograma e distribuição dos pontos amostrados no espaço.

Existem algumas variáveis que, embora mudem as magnitudes de variação, preservam entre si a maneira como variam no espaço, apresentando semivariogramas que podem ser agrupados em um único, quando divididos individualmente, pelas respectivas variâncias. Como exemplo pode-se incluir umidade do solo amostrada em pequenos espaços de tempo. Nesse caso então, as vantagens da vizinhança única aumentam porque modelos de semivariograma escalonados podem ser usados para obter pesos comuns a todas as variáveis.

b) Distância constante

O método de distância constante seleciona para cada ponto estimado uma vizinhança dentro de um círculo de raio especificado. Por consequência, nos cantos de uma área retangular tem-se um quarto de círculo com um quarto do número de vizinhos. Uma vantagem imediata na utilização deste método é o conhecimento exato que se tem da distância na qual os vizinhos para estimativa são procurados. Entretanto, o número de vizinhos pode mudar ao longo do campo fazendo com que o tamanho do sistema matricial seja variável. Isto pode ser um problema se exceder o valor usado na dimensão das matrizes do programa computacional utilizado.

c) Vizinhos constantes

O número de vizinhos constantes é um método bastante utilizado, pois mantém constante o número de vizinhos em qualquer posição do campo. Os vizinhos são procurados, inicialmente, dentro de um raio inicial. Se o número de vizinhos encontrados for menor que o limite especificado, a distância é incrementada e o processo se reinicia. Caso contrário, se o número de vizinhos encontrados for maior que o limite, apenas o número especificado mais próximo será utilizado. Conseqüentemente, a distância sobre a qual se procuram vizinhos varia sobre o campo.

Obviamente, as vantagens deste método são as desvantagens do anterior (distância constante) e vice-versa. Porém, em situações em que a amostragem foi efetuada em espaçamentos regulares, a distância de procura por vizinhos não muda muito e as desvantagens deste método são minimizadas. Devido ao fato de as amostragens regulares serem as mais usadas e as facilidades inerentes deste método, o mesmo é o mais comumente usado.

d) Procedimento de quadrantes

O procedimento de quadrantes consiste em utilizar um número especificado de vizinhos em cada quadrante ao redor do valor a ser estimado, fazendo com que a estimativa receba contribuição semelhante em número de todas as direções. As desvantagens deste método estão nas estimativas dos pontos pertencentes aos cantos e extremidades do campo e também no desconhecimento da distância em que se localizam os vizinhos. Por estas razões, o método dos quadrantes apresenta mais desvantagens que vantagens.

3.7 Descrição e modelagem da estrutura de variabilidade espacial

Assumida a estacionaridade, hipótese intrínseca, e considerando que a associação das variáveis em pontos distintos é maior à medida que mais próximos estes pontos estejam um do outro, a etapa seguinte consiste em descrever e modelar estas relações entre distâncias e associação espacial.

Várias medidas se prestam para a descrição desta relação, tais como a autocovariância e autocorrelação, usuais na análise de séries temporais. Em

geoestatística a medida normalmente utilizada é a semivariância. Ao contrário da covariância e correlação, a semivariância é uma medida de dissimilaridade, ou seja, o valor da semivariância é maior à medida que as variáveis estão menos associadas.

3.7.1 Semivariância e semivariograma

A semivariância exige uma hipótese de estacionaridade menos restritiva em relação às outras medidas como a covariância, que exige estacionaridade de segunda ordem. Esta exigência menor no modelo de função aleatória não tem conseqüências na maioria dos casos práticos (DEUTSCH & JOURNAL, 1992). Por este motivo, a semivariância pode ser utilizada em um maior número de situações.

A função semivariância é definida a partir dos desvios das medidas (d) em relação à reta 1:1 do gráfico de dispersão de $Z(s_i)$ em relação à $Z(s_{i+h})$.

$$[Z(s_{i+h}) - Z(s_i)]^2 = d^2 + d^2$$

$$d^2 = \frac{1}{2} [Z(s_{i+h}) - Z(s_i)]^2 \quad (1)$$

Considerando:

$$\bar{Z} = \frac{\sum_{i=1}^n Z(s_i)}{n}$$

O valor médio da variável em estudo e o erro médio em relação à média, como:

$$\bar{\varepsilon}(s_i) = \frac{\sum_{i=1}^n [Z(s_i) - \bar{Z}]}{n}$$

Onde n corresponde ao número de realizações da variável regionalizada $Z(s_i)$.

Entretanto, como o valor de $\bar{\varepsilon}(s_i)$ é zero, considere o quadrado do erro

$$[\varepsilon(s_i)]^2 = [Z(s_i) - \bar{Z}]^2.$$

Então o quadrado médio do desvio (erro, resíduo) deve ser:

$$QMR = \frac{\sum_{i=1}^n d^2}{N(h)}, \quad (2)$$

em que $N(h)$ corresponde ao número de pares de observações separadas pelo mesmo h . Substituindo o valor da equação 1 em 2 tem-se segundo Journel & Huijbregts (1978):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^n [Z(s_{i+h}) - Z(s_i)]^2, \quad (3)$$

quando a hipótese intrínseca é mantida, o valor de (3) é um estimador, pelo método dos momentos, sem tendências de $\gamma(h)$.

Com o semivariograma espera-se que valores mais próximos sejam mais similares do que aqueles mais afastados e que a diferença entre dois valores dependa apenas da sua posição relativa.

Na forma gráfica, a função semivariância ou o semivariograma, apresenta em condições ideais, o seguinte comportamento:

- (i) para um $h = 0$, a função semivariância é zero ou não mostra variância;
- (ii) com o aumento de h , $\gamma(h)$ cresce gradualmente e aproxima-se de uma constante C , atingindo o patamar ou "sill".

A dependência das observações é avaliada pela estrutura do semivariograma. A distância na qual estas se tornam independentes é chamada de alcance ou raio de influência a .

O crescimento mais ou menos rápido do semivariograma representa, com efeito, a maneira mais ou menos rápida com a qual é diminuída a influência de uma amostra sobre as zonas mais afastadas. A função $G(h)$ pode representar as características estruturais específicas de uma variável regionalizada analisando o comportamento na origem, a anisotropia e o fenômeno de transição.

A continuidade de uma V.R. está ligada ao comportamento do semivariograma nas vizinhanças da origem. Segundo Guerra (1988), podem-se distinguir quatro tipos de semivariograma (Figura 3.6).

Uma variável que tenha continuidade estrita ou alta continuidade, por exemplo, uma variável com caráter puramente geométrico, graficamente apresenta um

comportamento parabólico na origem (Figura 3.6a). Uma variável que possui continuidade “média” apresenta um semivariograma contínuo na origem, porém apresentando uma tangente oblíqua (Figura 3.6b). Outra variável que apresenta uma tangente vertical na origem, ou seja, aquela em que $G(h) = C_0$, quando h tende a zero, terá um $G(h)$ como na Figura 3.6c. Este tipo de descontinuidade recebe o nome de efeito pepita (“*nugget effect*”, “*random variance*”). O efeito pepita, C_0 , representa as variações locais ou a pequena escala, como erros de análise, amostragem, etc. Finalmente uma variável que apresenta efeito pepita puro, ou seja, um fenômeno totalmente aleatório (Figura 3.6d).

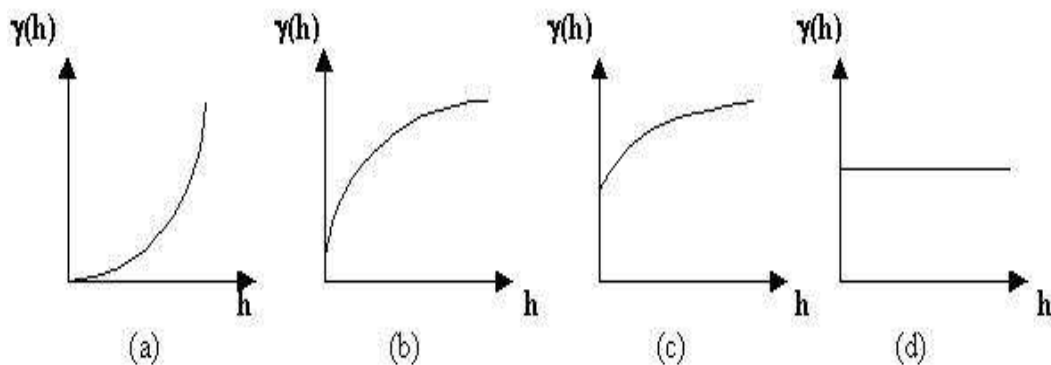


Figura 3.6 - Comportamento na vizinhança da origem de quatro tipos de semivariograma.

Dada a relação dependência da função semivariância com a distância h e sendo h um vetor, o semivariograma deve ser construído segundo distintas direções no espaço, visando estudar como se deforma a função semivariograma ao mudar a direção de h , de modo a colocar em evidência a anisotropia existente no fenômeno.

A anisotropia refere-se à existência de uma direção privilegiada ao longo da qual os valores da variável regionalizada não variam de forma significativa. Segundo Vieira (2000), a maioria das variáveis em ciência do solo pode ter um comportamento anisotrópico, aconselhando examinar semivariogramas para várias direções, tais como: 0° na direção das ordenadas; 90° na direção das abscissas; 45° e -45° nas duas diagonais. Pode suceder, contudo, que $\gamma(h)$ não dependa da direção, isto é, calculado em diferentes direções, forneça o mesmo semivariograma experimental. Neste caso, diz-se que o fenômeno regionalizado é isotrópico.

Os diferentes tipos de erros aleatórios ou sistemáticos, na obtenção das amostras e medidas das variáveis em intervalos inferiores ao de amostragem, geram uma descontinuidade na origem, denominada de efeito pepita (“*nugget effect*”) e indicada por C_0 (ANDRIOTTI, 2003). O patamar, então, é $C + C_0$ e um exemplo pode ser visto na Figura 3.7.

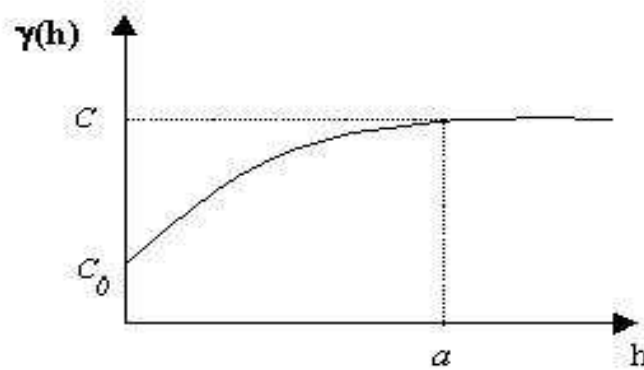


Figura 3.7 - Semivariograma com efeito pepita.

O valor α é denominado alcance e corresponde à idéia de zona de influência de uma amostra. Ele marca a distância a partir da qual um ponto da variável em estudo não tem mais influência sobre o ponto vizinho, marcando assim, o início da zona de pura aleatoriedade.

Este tipo de situação refere-se ao fenômeno de transição que se caracteriza por semivariograma cujo crescimento tende até um patamar $C + C_0$. O patamar (“*sill*”, “*palier*”, soleira) $C + C_0$, marca o valor da variância total da variável em estudo, C é denominada variância de dispersão espacial e representa as diferenças espaciais entre os valores de uma variável tomada em dois pontos separados por distâncias cada vez maiores.

Segundo Andriotti (2003), existem três tipos de semivariogramas, a saber: semivariograma observado ou experimental, aquele obtido a partir da amostragem realizada, portanto o único conhecido; o semivariograma verdadeiro é o semivariograma real do fenômeno em estudo e é sempre desconhecido. E o semivariograma teórico, que é um semivariograma de referência.

Um modelo que é ajustado por tentativas, bem como seus parâmetros sobrepostos aos pontos estimados, não tem rigor e deve ser utilizado algum procedimento que

possibilite encontrar, com boa aproximação, informações sobre a amostra, como será abordado no item 3.10 (validação cruzada).

3.7.2 Modelos teóricos de semivariograma

O gráfico do semivariograma empírico estimado por $\gamma(h)$ é formado por uma série de valores, sobre os quais se objetiva ajustar uma função. O modelo de ajuste deve representar o melhor possível o comportamento de $\gamma(h)$.

Dos modelos teóricos existentes, ressaltam-se: o esférico, o exponencial e o de Gauss ou gaussiano, os quais são modelos com patamar. Como exemplo de modelos sem patamar temos o linear, potência e o logarítmico ou de De Wijs. Uma vez escolhido um semivariograma teórico serão utilizadas suas propriedades para a análise semivariográfica.

Modelo esférico

É o esquema mais comumente utilizado nas variáveis estudadas em geociências e representa um fenômeno de transição. Apresenta um crescimento rápido na origem, onde tem comportamento linear. Sua expressão matemática é:

$$\gamma(h) = \begin{cases} C \left[\frac{3|h|}{2a} - \frac{1}{2} \frac{|h|^3}{a^3} \right], & 0 < |h| \leq a \\ C, & |h| > a \end{cases}$$

Inicia-se na origem e vai crescendo à medida que aumenta a separação das amostras, mostrando maior similaridade de valores mais próximos. Quando não há a presença do efeito pepita (C_0) tem-se $C = \sigma^2$ (variância *a priori* finita ou patamar) e a amplitude ou alcance.

Para $h > a$ não há correlação entre $Z(s_i)$ e $Z(s_{i+h})$ e $\gamma(h)$ alcança o patamar. A distância a corresponderá à distância de influência de cada amostra. A distância a , chamada de alcance, tende a se estabilizar por volta do patamar (Figura 3.8A), indicando uma total aleatoriedade ou independência das amostras. Isto sugere que para distâncias inferiores ao alcance adotam-se os procedimentos geoestatísticos e, em caso contrário, as técnicas da estatística clássica.

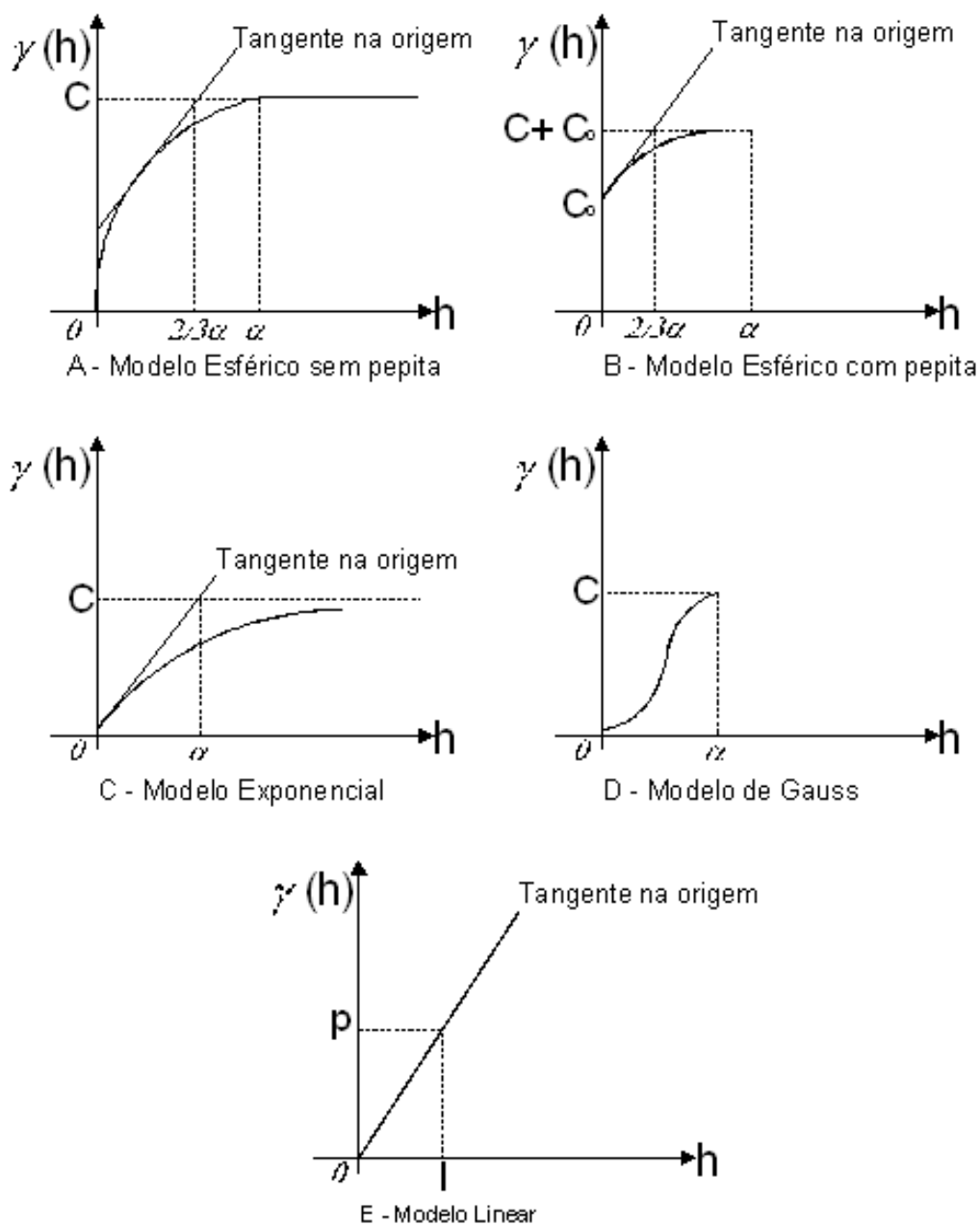


Figura 3.8 – Principais modelos semivariográficos teóricos.

Ocorrendo efeito pepita, tem-se:

$$\gamma(h) = \begin{cases} C_0 + C \left[\frac{3|h|}{2a} - \frac{1}{2} \frac{|h|^3}{a^3} \right], & 0 < |h| \leq a \\ C_0 + C, & |h| > a \end{cases}$$

Onde a representação genérica está na Figura 3.8B, sendo a a amplitude e $C + C_0$ o patamar. Observando a tangente na origem o modelo assume uma posição particular tratando de um modelo semivariográfico com comportamento linear na origem e o ponto onde a referida tangente corta o patamar representa $\frac{2}{3}a$ facilitando o cálculo automático.

Modelo exponencial

Também apresenta comportamento linear à origem (Figura 3.8C). Quando não há efeito pepita sua equação é:

$$\gamma(h) = \begin{cases} C(1 - e^{-\frac{h}{a}}), & 0 < |h| \leq a \\ C, & |h| > a \end{cases}$$

sendo a amplitude a determinada pelo encontro da tangente na origem com o patamar C igual à variância *a priori*, variância finita. Quando existe efeito pepita C_0 a sua equação fica:

$$\gamma(h) = \begin{cases} C_0 + C(1 - e^{-\frac{h}{a}}), & 0 < |h| \leq a \\ C_0 + C, & |h| > a \end{cases}$$

e o patamar passa a ser $C + C_0$. O modelo exponencial é um modelo de fácil aplicação devido a sua simples equação matemática.

O alcance, nesse modelo, tem significado puramente analítico, sendo o patamar só alcançado pela curva de forma assintótica, ou seja, teoricamente quando $h = \infty$. Na prática, usa-se um alcance igual a $3a$.

Modelo Gaussiano

Segundo Landim (1998), um modelo semivariográfico que também representa um fenômeno de transição, a curva é parabólica junto à origem e não linear como os modelos anteriores, é o modelo de Gauss. A tangente nesse ponto é horizontal o que indica pequena variabilidade para curtas distâncias, genericamente representado na Figura 3.8D. Quando não há efeito pepita a sua equação é:

$$\gamma(h) = \begin{cases} C(1 - e^{-\frac{h^2}{a^2}}), & 0 < |h| \leq a \\ C, & |h| > a \end{cases}$$

tratando-se de um modelo semivariográfico mais regular possível, visto admitir tangente horizontal na origem. O seu patamar C e a sua amplitude a são determinados como no semivariograma esférico e têm exatamente o mesmo significado. O alcance neste modelo também tem significado puramente analítico, sendo o alcance prático tomado como sendo $\sqrt{3}.a$ (ANDRIOTTI, 2003). Quando existe efeito pepita (C_0) a sua equação fica:

$$\gamma(h) = \begin{cases} C_0 + C(1 - e^{-\frac{h^2}{a^2}}), & 0 < |h| \leq a \\ C_0 + C, & |h| > a \end{cases}$$

em que o patamar passa a ser $C_0 + C$. A existência do efeito pepita neste modelo só admite-se pela presença de erros humanos na amostragem, uma vez que os fenômenos representados por este modelo são os mais regulares possíveis.

Modelo Linear

Um exemplo de modelo sem patamar é o modelo linear e tem por expressão geral

$$\gamma(h) = C_0 + p|h|^\alpha,$$

onde p é a inclinação da reta representativa do semivariograma, sendo utilizado para valores de $0 < \alpha < 2$; sua variância *a priori* é infinita. O esquema linear recebe esse nome quando $\alpha = 1$ (Kitanidis, 1997). Genericamente representa-se conforme Figura 3.8E e pode-se dizer que é a reta que se obtém unindo o ponto experimental obtido para $h = 1$ com a origem $h = 0$.

A principal observação para este modelo refere-se ao seu crescimento monótono e o fenômeno é não estacionário. Algumas situações que não apresentam patamar definido são consideradas de infinita capacidade de dispersão, ajustando-se aos modelos que consideram este fato, como por exemplo, um modelo linear. No caso de total independência das amostras, tem-se o chamado efeito aleatório puro, e um semivariograma da Figura 3.9.

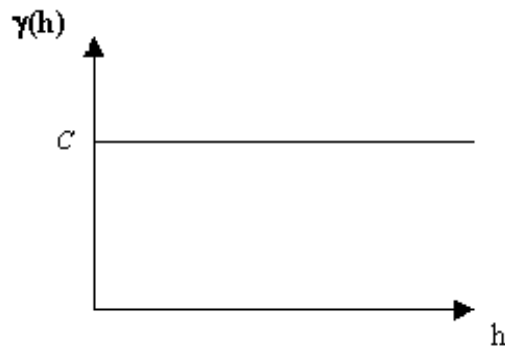


Figura 3.9 - Aspecto do semivariograma com total independência entre as amostras.

3.8 Métodos de estimação do semivariograma

Existem diversos métodos de ajuste de modelos semivariográficos, McBratney & Webster (1986) comentam o ajuste de semivariogramas abordando os métodos de mínimos quadrados (ordinários, ponderados e generalizados) e de máxima verossimilhança. Segundo os autores, o método de mínimos quadrados ponderados tem melhor realismo e eficiência computacional, sendo o melhor que se adequou a este trabalho.

O ajuste por mínimos quadrados ordinários e ajuste empírico também chamado de ajuste “a sentimento” são bastante comentados na literatura (JOURNEL & HUIJBREGTS, 1978; ISAAKS & SRIVASTAVA, 1989; CRESSIE, 1991).

O método de mínimos quadrados ordinário é citado por Ribeiro Júnior (1995) como inadequado porque os pontos da nuvem de dados correspondem às estimativas provenientes de diferentes pares de pontos e estas estimativas não possuem a mesma precisão e não são independentes.

No método “a sentimento” escolhe-se um modelo e seus parâmetros e este é sobreposto aos pontos estimados, a partir de uma nuvem de pontos obtém-se a

estimação dos parâmetros do semivariograma através de uma avaliação visual. Alguns autores como McBratney & Webster (1986) e Ribeiro Júnior (1995) consideram o método pouco rigoroso e não fidedigno.

Apesar das críticas, diversos textos geoestatísticos clássicos, como o de Journel & Huijbregts (1979), adotam este método fazendo ressalva ao método de ajuste automático. Ribeiro Júnior (1995) recomenda a utilização do ajuste empírico como partida para estimar os parâmetros do modelo de interesse e dar continuidade com outros métodos de melhor embasamento estatístico, os quais não são de fácil implementação computacional.

3.9 A estimação geoestatística

Muitas vezes, além de um modelo de dependência espacial, existe o interesse em se obter valores em pontos não amostrados, ou obter uma malha de pontos interpolados que possibilite a visualização do comportamento da variável na região através de mapas ou gráficos de superfície.

O interesse de se detalhar uma área sugere a idéia de um método de interpolação que forneça estimativas de valores em pontos ou médias em sub-regiões e que os estimadores utilizados sejam não viesados e de variância mínima. Segundo Ribeiro Jr. (1995), a acurácia de uma estimativa depende dentre vários fatores de:

- (i) número de pontos avaliados e a qualidade da informação obtida de cada um deles referem-se ao método de medida, sua precisão e homogeneidade;
- (ii) posição dos pontos no campo deve ser arranjada de maneira que abranja a maior área possível evitando concentrações em sub-regiões;
- (iii) continuidade espacial de variáveis interpoladas, considera que quantidades com suaves variações podem ser melhor interpoladas que quantidades com flutuações mais erráticas.

A proposta geoestatística de interpolação é conhecida como Krigagem e foi desenvolvida por Matheron em homenagem ao engenheiro de minas D.G.Krige, o qual em trabalhos com reserva de ouro na África do Sul introduziu o uso do processo de médias móveis para estimação em depósitos minerais. A Krigagem é um processo que estima valores de variáveis espacialmente distribuídas a partir de valores adjacentes

interdependentes considerando todos os principais fatores necessários para se ter uma boa acurácia. O estimador de krigagem equação (6)

$$\bar{Z}(s_0) = \sum_{i=1}^{N(h)} \lambda_i Z(s_i), \tag{6}$$

é um interpolador BLUE (“*Best Linear Unbiase Estimator*”) desenvolvido a partir de dados referentes a uma variável de natureza regionalizada. O estimador é dito linear, pois a estimação é obtida combinando os valores amostrados e ponderando os pontos amostrais para que se obtenha o menor erro quadrático médio.

Pode ser aplicado para previsão do valor de uma variável em um ponto específico dentro de um campo geométrico, krigagem pontual; para o cálculo do valor médio de uma variável regionalizada em uma determinada região, krigagem em blocos, e na estimação de “*drift*”, tendência local de uma variável de interesse de modo similar a análise de superfície de tendência (VALENTE, 1989).

Existem diversos tipos de krigagem: a simples, ordinária, universal, indicadora, probabilística e outras. Neste trabalho, nos restringiremos a krigagem do tipo ordinária e universal, com estimativas pontuais.

3.9.1 Krigagem ordinária

O estimador de krigagem é uma combinação linear dos pontos conhecidos, em que os ponderadores são proporcionais às distâncias euclidianas entre o ponto a ser estimado e os demais pontos conhecidos, incorporando a estrutura de variabilidade na região de estimação.

As distâncias consideradas não são somente as distâncias euclidianas, mas também as distâncias entre vizinhos. Considere um fenômeno isotrópico e a estimação de um ponto qualquer em que os vizinhos a este ponto estão igualmente espaçados e não agrupados, nesta situação os pesos destes vizinhos são semelhantes.

Agora considere a situação em que os vizinhos ao ponto a ser estimado estão distanciados igualmente, existindo um agrupamento de alguns valores. Nesta situação é razoável que os pesos de cada valor agrupado sejam menores que do valor não agrupado, mesmo que a distância entre estes pontos e o ponto estimado seja a

mesma, pois os dados agrupados trazem informações redundantes de uma mesma região.

Esta característica do estimador é denominada de “*declustering*”, pois o preditor considera uma medida de associação entre os pontos e suas respectivas vizinhanças. Geralmente, a medida de associação utilizada é a semivariância, por isso a necessidade do semivariograma e do modelo ajustado.

Se o modelo ajustado para o semivariograma é correto e não apresenta erros de medidas devem-se, então, determinar os valores dos pesos, λ_i , que assegurem as propriedades de mínima variância e não tendenciosidade.

Assumida a hipótese intrínseca, para a não tendenciosidade ser assegurada deve-se ter:

$$\begin{aligned} E\{\bar{Z}(s_0) - Z(s_0)\} &= 0 \\ E\{\bar{Z}(s_0) - Z(s_0)\} &= E\left\{\sum_{i=1}^{N(h)} \lambda_i Z(s_i) - Z(s_0)\right\} = \\ \sum_{i=1}^{N(h)} \lambda_i E\{Z(s_i)\} - E\{Z(s_0)\} &= 0 \end{aligned}$$

Assumindo que $E\{Z(s_i)\}$ é constante para todo ponto s_i , igual a um valor m , tem-se:

$$m \sum_{i=1}^{N(h)} \lambda_i - m = 0 \Rightarrow \sum_{i=1}^{N(h)} \lambda_i = 1$$

Esta restrição garante a não tendenciosidade do estimador. Os coeficientes de estimação são obtidos minimizando o erro de estimação $[\bar{Z}(s_0) - Z(s_0)]$, em que a variância de estimação é dada por:

$$Var[\bar{Z}(s_0) - Z(s_0)] = E\{[\bar{Z}(s_0) - Z(s_0)]^2\} - \{E[\bar{Z}(s_0) - Z(s_0)]\}^2,$$

em que o primeiro e o segundo termos correspondem, respectivamente, ao erro quadrático médio e ao viés. Para o estimador não viciado, o erro quadrático médio é igual à variância, ou seja, o viés é nulo. Substituindo a equação do estimador de krigagem na expressão da variância tem-se:

$$Var[\bar{Z}(s_0) - Z(s_0)] = -\sum_i \sum_j \lambda_i \lambda_j \gamma(s_i, s_j) + 2 \sum_i \lambda_i \gamma(s_i, s_0)$$

Conforme Calvete & Ramirez (1990) e Ribeiro Júnior (1995), inserindo o multiplicador de Lagrange que permite solução do sistema de equações de krigagem:

$$Var[\bar{Z}(s_0) - Z(s_0)] = -\sum_i \sum_j \lambda_i \lambda_j \gamma(s_i, s_j) + 2 \sum_i \lambda_i \gamma(s_i, s_0) - 2\mu(\sum_i \lambda_i)$$

A variância quando minimizada, por meio do cálculo das derivadas de primeira ordem em relação aos parâmetros, sujeita a restrição $\sum_{i=1}^{N(h)} \lambda_i = 1$ e igualada a zero, resulta em um sistema de equações de krigagem do tipo:

$$\left\{ \begin{array}{l} \sum_{i=1}^{N(h)} \lambda_i \gamma(s_i, s_0) + \mu = \gamma(s_0, s_0) \\ \sum_{i=1}^{N(h)} \lambda_i = 1 \end{array} \right.$$

Resolvendo o sistema tem-se a estimativa:

$$\bar{Z}(s_0) = \sum_{i=1}^{N(h)} \lambda_i Z(s_i),$$

em que o estimador $\hat{Z}(s_0)$ é *BLUE* (*Best Linear Unbiased Estimator* - Melhor Estimador Linear não-Viciado).

Para realizar uma predição de um processo intrinsecamente estacionário é conveniente modelizar o variograma mediante uma função semidefinida negativa. Habitualmente se emprega a metade do variograma, que se denomina semivariograma (QUÍLEZ, 2006).

Através da krigagem ordinária tem-se $\bar{Z}(s_0)$, que é uma estimativa do valor esperado de $Z(s_0)$ ou $E[Z(s_0)]$. Por estimar uma média, o processo de krigagem implica numa suavização dos valores preditos para a região em estudo não fornecendo uma estimativa da dispersão destas variáveis, pois a variância de krigagem avalia apenas a configuração da vizinhança.

Repetindo o processo de krigagem ordinária em vários pontos de modo a formar uma malha fina é possível obter um mapa das estimativas na região estudada facilitando a interpretação quanto ao comportamento espacial da variável.

3.9.2 Krigagem universal

Para a obtenção de um variograma, é suposto que a variável regionalizada tenha um comportamento fracamente estacionário, onde os valores esperados, assim como sua covariância espacial, sejam os mesmos por uma determinada área. Assume-se, desse modo, que os valores dentro da área de interesse não apresentem tendência que possam afetar os resultados.

Isso nem sempre acontece, pois existem situações em que a variável exibe uma variação sistemática e torna-se necessário, então, metodologia específica para a devida correção. Para tanto, o método da “krigagem universal” é utilizado para resolver um problema desse tipo (DEUTSCH & JOURNEL, 1992). Trata-se de uma metodologia geoestatística para situações não estacionárias.

Nesse tipo de krigagem o tratamento matemático leva em consideração ao mesmo tempo a tendência e a estimativa dos valores e, como consequência, os erros de estimativa incluem os erros associados à estimativa da superfície de tendência (CLARK & HARPER, 2000).

3.10 Validação cruzada

O preditor de krigagem, como visto anteriormente, considera que o modelo do semivariograma é correto e que não há erros de medida. A partir do pressuposto ele é obtido de forma a atender duas qualidades desejáveis em um preditor a não tendenciosidade e a variância mínima. Mas como avaliar se o modelo escolhido está correto?

A validação cruzada apresenta-se como uma ferramenta adequada, pois a idéia é: se o fenômeno foi satisfatoriamente modelado é possível reproduzir, com boa aproximação, informações da amostra (VIEIRA, 1996).

O procedimento da validação cruzada consiste em retirar um ponto do conjunto original de dados e estimá-lo a partir do restante. Utilizando a krigagem, é também estimada a variância da estimativa retornando o ponto ao conjunto estima-se o

próximo ponto repetindo o processo para todos os pontos amostrados (RIBEIRO JR, 1995).

Em cada ponto tem-se o erro de estimação e padronizando-o pelo desvio padrão da estimativa espera-se, conforme Gonçalves (1997), que os erros deduzidos de krigagem tenham média nula e variância unitária. Os valores das estatísticas obtidas a partir dos erros e dos gráficos mostram o comportamento das estimativas e são utilizados na avaliação.

Interessante observar que a validação cruzada não avalia simplesmente o modelo de semivariograma escolhido, mas toda a modelagem do processo estocástico em questão, isto é, são colocados à prova a decisão de estacionaridade os estimadores utilizados, o tratamento dos dados atípicos, o modelo de semivariograma adotado e a decisão relativa à anisotropia.

Nem sempre é fácil identificar a causa dos problemas apresentados nos resultados da validação cruzada, mas com a união da experiência, conhecimento dos problemas em questão e análise minuciosa dos diversos recursos descritivos, em especial os gráficos, pode-se identificar as origens dos problemas.

CAPÍTULO 4 – MATERIAIS E MÉTODOS

Os dados são provenientes de um conjunto de medidas de concentrações de gases (ppm) metano (CH_4), etano (C_2H_6), eteno (C_2H_4), propano (C_3H_8), propeno (C_3H_6), isobutano ($i\text{-C}_4\text{H}_{10}$), butano (C_4H_{10}), buteno (C_4H_8), isopentano ($i\text{-C}_5\text{H}_{10}$), pentano (C_5H_{10}) e hidrocarbonetos maiores que pentano, também representados como C_1 , C_2 , C_2^- , C_3 , C_3^- , $i\text{-C}_4$, C_4 , C_4^- , $i\text{-C}_5$, C_5 e C_6^+ , e das razões eteno/etano (C_2^-/C_2), metano/(etano+propano+isobutano+butano+isopentano+pentano) ($\text{C}_1/\text{C}_2\text{C}_5$) e do somatório etano, propano, isobutano, butano, isopentano e pentano (C_2C_5).

Os programas computacionais utilizados são: SPSS (11.0.1), Idrise32 e Surfer (8.01) para os cálculos estatísticos descritivos, obtenção de semivariogramas e dos mapas de contorno e superfície, respectivamente.

4.1 Localização e delineamento amostral

Foi realizado um extensivo levantamento bibliográfico das metodologias aplicadas para determinação de *background* e anomalias de dados geoquímicos e ambientais. Praticamente todos os trabalhos encontrados na literatura estão voltados para as áreas de mineração e agricultura.

Os dados são provenientes de um bloco exploratório, localizado numa bacia sedimentar brasileira tipo rifte (*onshore*). As amostras foram coletadas com espaçamento regular de 500m, com exceção de algumas amostras, resultando em uma malha de pontos de 26 linhas (12,93 km) e 9 colunas (4,32 km), perfazendo uma área aproximada de 55,87 km² com uma extensão diagonal aproximada 13,63 km (Figura 4.10). Demais informações sobre o bloco exploratório, localização, método de coleta, análise químicas, etc., foram subtraídas deste trabalho por ser tratar de dados confidenciais.

Em cada ponto foram amostradas medidas das variáveis em ppm: metano (C_1), etano (C_2), eteno (C_2^-), propano (C_3), propeno (C_3^-), isobutano ($i\text{-C}_4$), butano (C_4), buteno (C_4^-), isopentano ($i\text{-C}_5$), pentano (C_5), hexano+ (C_6^+).

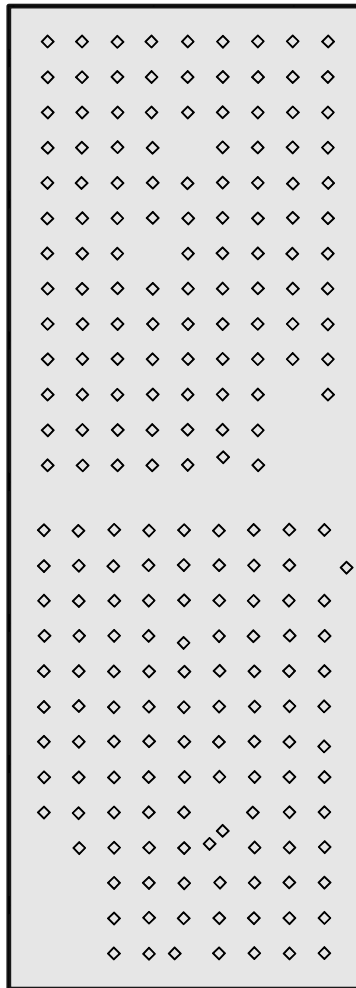


Figura 4.10 – Distribuição espacial das amostras do levantamento geoquímico com a localização amostral dos 220 pontos.

4.2 Análise estatística

As análises estatísticas têm o objetivo de descrever e testar algumas hipóteses a respeito de uma população da qual um conjunto de dados foi coletado (POTVIN, 1993).

O ponto de partida para a análise dos dados espacialmente distribuídos foi a divisão da análise em quatro etapas: a primeira consistiu em uma descrição geral dos dados, onde foram utilizadas as análises exploratórias descritivas e gráficas; a segunda parte foi analisada diversos métodos para identificação de dados *outliers*; na terceira parte considerou as posições das amostras, em que as análises semivariográficas amostral e modelada foram aplicadas para estudar a magnitude e a estrutura da variabilidade espacial das variáveis com avaliação do ajuste do modelo de semivariograma por meio da validação cruzada. Sendo assim, executado interpolações de pontos não

amostrados através da krigagem e confecção de mapa de superfície de valores; na quarta parte foram efetuadas interpolações de pontos não amostrados sem considerar as posições das amostras pelos métodos determinísticos de curvatura mínima e inverso do quadrado da distância.

4.2.1 Análise exploratória descritiva

A análise exploratória dos dados é uma etapa de fundamental importância e obrigatória dentro de qualquer estudo estatístico. É uma etapa precursora introduzida para averiguar as propriedades estatísticas e matemáticas dos dados (BURROUGH *et al.*, 1996). Emprega grande variedade de técnicas gráficas e quantitativas, visando maximização na obtenção de informações ocultas na sua estrutura, descoberta de variáveis importantes nas suas tendências, e não apenas na sua variação, detecção de comportamentos anômalos do fenômeno, tomada de decisões com relação à retirada ou não de dados ou na remoção de tendências, teste de validação das hipóteses assumidas, escolha de modelos e obter o melhor resultado das variáveis.

A decisão quanto ao tipo de estacionaridade que deve ser assumida é uma afirmação ou suposição a respeito do comportamento das variáveis aleatórias e depende do entendimento e conhecimento do fenômeno em estudo. Apesar da hipótese de estacionaridade não ser testada, a decisão de assumir algum tipo de estacionaridade pode ser orientada por uma análise exploratória adequada dos dados. A falta de estacionaridade pode ser tanto na média quanto na variância. Isto justifica a elaboração dos gráficos que trazem informações não só de medidas de posição, mas também de dispersão.

As técnicas descritivas exploratórias devem abordar todas as diversas fases da análise, e não serem utilizadas apenas inicialmente. Análise exploratória através de gráficos de dispersão, *boxplot*, histogramas e gráfico de probabilidade normal foram utilizados para identificação de tendências, possíveis pontos discrepantes, forma de distribuição, escolha da análise a ser aplicada, decisão do tipo de estacionaridade que pode ser assumida e caracterização da variabilidade.

O histograma consiste numa representação gráfica da distribuição de freqüências de uma massa de medições, normalmente um gráfico de barras verticais. É um gráfico composto por retângulos justapostos em que a base de cada um deles corresponde ao intervalo de classe e a sua altura à respectiva freqüência. Quando o número de dados

aumenta indefinidamente e o intervalo de classe tende a zero, a distribuição de freqüência passa para uma distribuição de densidade de probabilidades.

A construção de histogramas tem caráter preliminar em qualquer estudo e é um importante indicador da distribuição de dados. Podem indicar se uma distribuição aproxima-se de uma função normal, como pode indicar mistura de populações quando se apresentam bimodais. O histograma também é usado para determinar a posição da mediana, pois esta está na posição em que passando uma linha vertical por esse ponto o histograma fica dividido em duas partes com áreas iguais, no entanto não é uma medida quantitativa precisa de assimetria e curtose.

A mediana é uma estatística útil quando se quer dividir o conjunto de dados em dois subgrupos. Quando o conjunto de dados é dividido em quatro subgrupos iguais obtendo com isso mais dois valores, em que 25% das observações estão abaixo de um deles e 75% estão abaixo do outro, estes valores são denominados de primeiro quartil e terceiro quartil, respectivamente.

A representação gráfica envolvendo os quartis é o *boxplot* (Item 2.7), no qual informa a variabilidade, simetria dos dados e identifica descritivamente diferenças nos comportamentos de grupos de variáveis. Sendo assim, através do *boxplot* pode-se ter idéia da configuração da distribuição de freqüência dos dados.

A normalidade dos dados pode ser avaliada através dos métodos gráficos de probabilidade normal e testes estatísticos de aderência, como os de Kolmogorov-Smirnov (K-S), que é uma técnica não paramétrica e de Shapiro-Wilk (S-W), técnica paramétrica.

O gráfico de probabilidade (*Q-Q plot*) apresenta os valores observados no eixo horizontal, e os valores que se esperava obter se a normalidade se verificasse no eixo vertical. Assim, se a distribuição for normal, os pontos devem situar-se aleatoriamente sobre a reta do gráfico; qualquer desvio dos pontos da reta representa um desvio da normalidade.

Os testes de aderência Shapiro-Wilk (S-W) e Kolmogorov-Smirnov (K-S) servem para comparar as probabilidades empíricas de uma variável com as probabilidades teóricas estimadas pela função de distribuição em teste, verificando se os valores da amostra podem razoavelmente ser considerados como provenientes de uma população com

aquela distribuição teórica. Nos testes de aderência, a hipótese nula (H_0) admite que a distribuição seja a especificada (normal, lognormal, e outras), com os seus parâmetros estimados com base nos dados amostrais.

As tabelas de K-S para testar a validade do modelo normal com estimação de parâmetros foram corrigidas por *Lillifors* e é essa correção que aparece na determinação do p-valor. O p-valor é o menor valor do nível de significância que conduz à rejeição da hipótese nula. Assim, valores “grandes” do p-valor conduzem à “aceitação” da hipótese nula e valores pequenos do p-valor à sua rejeição. Quando avaliamos os resultados de uma experiência, os descrevemos em função da sua significância. A significância refere-se o quão seguro podemos estar de que os nossos resultados são realmente diferentes. Diferentes níveis de significância podem ser adotados, sendo adotado neste trabalho o nível de significância do teste de 0,05.

Outros parâmetros da distribuição normal são a assimetria, que expressa o grau de desvio, ou afastamento da simetria. Uma curva normal apresenta assimetria igual à “0”, daí a utilidade desta medida para comparar outras distribuições com uma gaussiana ou normal, onde um valor negativo indica dados deslocados para a direita (cauda negativa longa) e valores positivos indicam dados deslocados para a esquerda (cauda positiva longa) em relação à média.

A curtose, grau de achatamento de uma distribuição, pode ser platicúrtica (achatada, valores menores de 3) ou leptocúrtica (alongada, valores maiores de 3), sendo 3 o valor para a curva normal (SNEDECOR & COCHRAN, 1967; SPIEGEL, 1985), a curtose é definida também pelo excesso de curtose, que é o valor da curtose menos 3, positiva para uma distribuição leptocúrtica, negativa para uma platicúrtica e nula para uma normal. Pela curtose, é possível se obter uma indicação da intensidade das freqüências das observações nas vizinhanças dos valores centrais, ou seja, da mediana – que divide a série de dados em 50% à sua esquerda e 50% à sua direita – e da média aritmética, caso esta última seja representativa dos valores centrais.

Já o coeficiente de variação (CV), é a medida de dispersão que se presta para a comparação entre diferentes conjuntos de dados. Segundo Koch e Link (1971), para valores de coeficiente de variação menores que 0,3 ou 30%, a distribuição é normal ou muito aproximadamente normal. Para valores de CV maiores que 0,8 ou 80%, a distribuição é fortemente assimétrica ou mesmo lognormal. Há, no entanto, uma faixa de valores definida por CV em 0,4 (40%) e 0,7 (70%) que não permite qualquer conclusão. Segundo Frizzo e Licht (2007), para valores de coeficientes de variação

maiores que 0,5 ou 50%, indicam conveniência da transformação dos dados originais para seu logaritmo.

Como forma de classificar a variabilidade dessas variáveis, utilizou-se como parâmetro o comportamento das variáveis do solo que, segundo Warrick e Nielsen (1980), classifica em baixa ($CV < 0,12$ ou 12%), média ($0,12 < CV < 0,62$) e alta ($CV > 0,62$ ou 62%).

4.3 Determinação de *background* e anomalia

Existem diversos critérios para definir se uma observação pode ser considerada discrepante ou não; quatro desses métodos são utilizados neste trabalho a título de comparação de métodos: a regra [média (\bar{x}) + 2 desvios padrão (dp)]; *boxplot* [*upper whisker*], [mediana (med) + 2 desvios medianos absolutos (dma)] e 98º percentil como valores limites que separam os dados de *background* de anômalos.

4.4 Análise exploratória espacial

Esta análise é mais detalhada, pois ao se estudar o comportamento da variável as posições das amostras foram consideradas em algum momento da análise. O “*postplot*”, gráfico de superfície de valores e outras ferramentas, que consideram as posições das amostras, foram técnicas utilizadas na análise exploratória espacial.

O “*postplot*” é um gráfico construído pelo programa Surfer 8.01 em que os dados são representados nos pontos onde foram amostrados. Para cada valor foi associado um símbolo indicando a qual classe ele pertence. Este gráfico constituiu-se em um instrumento de verificação de tendências associadas às posições espaciais. Concentrações de símbolos iguais podem indicar sub-regiões e variações gradativas; ao longo de alguma direção indicam a existência de tendências.

Para atender a hipótese intrínseca não devem haver padrões neste gráfico. O “*postplot*” fornece uma idéia inicial da continuidade do fenômeno, identificando regiões com concentração de amostras de elevados ou baixos valores que sugerem a divisão da área em sub-regiões.

Outra ferramenta é o gráfico de superfície de valores que apresenta a superfície das medidas da variável definida pela amostragem. A análise da superfície é grosseira e

fornece apenas uma idéia do comportamento da variável na região. Uma superfície mais detalhada é obtida após a construção por interpolações de uma malha mais fina que a amostrada.

Gráficos dos valores amostrais em relação às linhas ou colunas são métodos alternativos para identificar tendência ou sub-regiões. Outra ferramenta para a investigação de dados espacialmente atípicos é a análise do efeito proporcional, que é o aumento da variância de acordo com o aumento da média. Na natureza este efeito proporcional é observado em fenômenos com distribuições assimétricas, lognormal, por exemplo.

A análise do efeito proporcional consiste em construir gráficos de dispersões da média e da variância no sentido das linhas e colunas. Em uma situação que não apresenta tendência, tem-se uma expectativa dos pontos em torno da média. Para o caso de dispersão que varia no espaço, a expectativa dos pontos afasta-se da média gradualmente.

4.5 Análise geoestatística

Na geoestatística, a análise do semivariograma é uma etapa importante, pois o modelo de semivariograma escolhido é a interpretação da estrutura de correlação espacial a ser utilizada nos procedimentos inferenciais da krigagem. A análise completa do semivariograma compreende os seguintes passos:

- Levantamento do semivariograma experimental: o cálculo do valor do semivariograma em cada distância é realizado utilizando os dados amostrais da variável regionalizada $Z(s_i)$ na equação (3), que é o estimador do semivariograma. O semivariograma experimental, quando apresenta dependência espacial, é uma curva irregular com flutuações que crescem com os valores de h . Os últimos pontos deste gráfico têm menor significância estatística, pois envolvem poucos pares de pontos. A curva experimental obtida é na realidade um estimador do verdadeiro semivariograma desconhecido, este possui propriedades matemáticas precisas. Portanto, fazendo-se necessário o ajuste de um modelo teórico que sirva de base para os cálculos posteriores;

- Ajuste a uma família de modelos de semivariogramas: o modelo ajustado ao semivariograma experimental está relacionado com o comportamento do semivariograma na origem e a existência ou não de um patamar. Ao modelo une-se o efeito pepita por extrapolação da curva. A adequação de um modelo teórico é fundamental, pois a partir deste modelo serão feitas inferências em relação ao semivariograma verdadeiro;
- Validação do modelo a ser utilizado nos procedimentos da krigagem: a adequada modelagem da estrutura de dependência espacial valida qualquer outro procedimento de inferência e interpolação, além de ser um forte subsídio para decisões práticas sobre o fenômeno em estudo. Conhecido o semivariograma da variável em estudo e havendo dependência espacial entre as amostras, o próximo passo é obter informações de pontos não amostrados no campo através do método de interpolação denominado de krigagem.

Os métodos de krigagem são métodos de interpolação que procuram minimizar o erro da estimação; na realidade, o erro médio de estimação é nulo. O problema que se coloca normalmente é o de estimar o valor de uma variável em locais não amostrados, $Z(s_0)$, a partir dos valores de locais amostrados, $Z(s_i)$. O estimador de krigagem $\bar{Z}(s_0)$ é também um estimador linear que considera a organização espacial da variável:

$$\bar{Z}(s_0) = \sum_{i=1}^{N(h)} \lambda_i Z(s_i),$$

em que λ_i é o ponderador da distância e $\sum_{i=1}^{N(h)} \lambda_i = 1$.

A krigagem é um método exato e não viesado, isto é, os valores nos locais amostrados são reproduzidos e o erro médio de estimação é nulo.

A análise geoestatística, em algumas situações, exige certo grau de subjetividade. Sendo assim, para avaliar se o fenômeno foi modelado de forma adequada, alguns autores têm utilizado o método da validação cruzada.

Na validação cruzada, assume-se que uma determinada amostra não tenha sido coletada, ou seja, elimina-se o seu valor e se estima a partir dos dados circundantes; após essa estimação, o valor real dessa amostra é reintroduzido no sistema e se

repete o processo para todas as outras amostras, de forma que para cada ponto é possível obter o erro de estimação. Uma estimação terá sido sem *bias* se o erro médio for zero, ou seja, se os valores estimados tiverem uma diferença média em relação aos valores experimentais igual a zero, e a variância estiver em torno de um.

É importante ressaltar que a análise geoestatística não se trata de um método contínuo em uma direção. A análise geoestatística consiste em ir e voltar, refazer e comparar antes de qualquer decisão definitiva. Várias decisões são tomadas ao longo do processo, ou seja, os resultados intermediários e finais não são obtidos de forma única, e as técnicas descritivas e exploratórias devem estar em todas as fases do processo de análise e não somente na fase inicial.

4.6 Comparação entre os métodos de interpolação

A Krigagem permite que se faça uma validação cruzada para checagem dos dados, ou pelo menos uma comparação entre os erros. Por este método, o ideal seria ter um erro médio padronizado dos valores preditos próximo de zero, um quadrado médio do erro o mais baixo possível e um quadrado médio do erro padronizado próximo de um.

No caso dos interpoladores determinísticos (e.g., curvatura mínima e o inverso quadrado da distância), somente fornecem o quadrado médio do erro, este tem que ser o mais baixo possível. O quadrado médio do erro (QME) é dado por:

$$QME = \left\{ \left[\sum_{i=1}^{n(i)} (\bar{Z}_i - Z_i)^2 \right] / n(i) \right\}$$

Os resultados obtidos para o quadrado médio do erro serão comparados, e o método que apresentar o menor quadrado médio residual será o método considerado como o mais eficiente, isto para que os valores dos pontos no mapa interpolado sejam o mais parecido possível com os valores recolhidos nesses pontos.

CAPÍTULO 5 – RESULTADOS E DISCUSSÕES

5.1 Análise exploratória descritiva

5.1.1 Dados originais

As estatísticas descritivas para os dados originais de concentrações de gases (ppm) C_1 , C_2 , C_2^- , C_3 , C_3^- , $i-C_4$, C_4 , C_4^- , $i-C_5$, C_5 e C_6^+ , e das razões C_2^-/C_2 , C_2C_5 e C_1/C_2C_5 estão apresentados na Tabela 5.1.

Para os dados C_1 , o excesso de curtose e assimetria foram superiores a um, levando-se em conta que a curtose de uma curva normal é de 3, a curtose dessa série de dados indica uma curva 35% além do valor, tendo a função de distribuição um pico mais elevado e denominado leptocúrtica.

Existe uma estreita relação entre o valor das medidas de tendência central e o comportamento da assimetria, tendo a média apresentado maior que a mediana, demonstrando a assimetria positiva dos dados. Um expressivo desvio-padrão (8,90) implica numa menor representatividade estatística da média dos dados observados.

O coeficiente de variação de 0,43 mostra essa menor representatividade estatística da média dos dados e não permite nenhuma conclusão da distribuição dos dados de acordo com Koch e Link (1971), e segundo Frizzo e Licht (2007), não indica conveniência da transformação dos dados brutos (originais) para seu logaritmo.

Segundo Warrick e Nielsen (1980), pode-se classificar a variabilidade da variável como média ($0,12 < CV < 0,62$). Podemos observar graficamente os valores da variável C_1 (Figura 5.11).

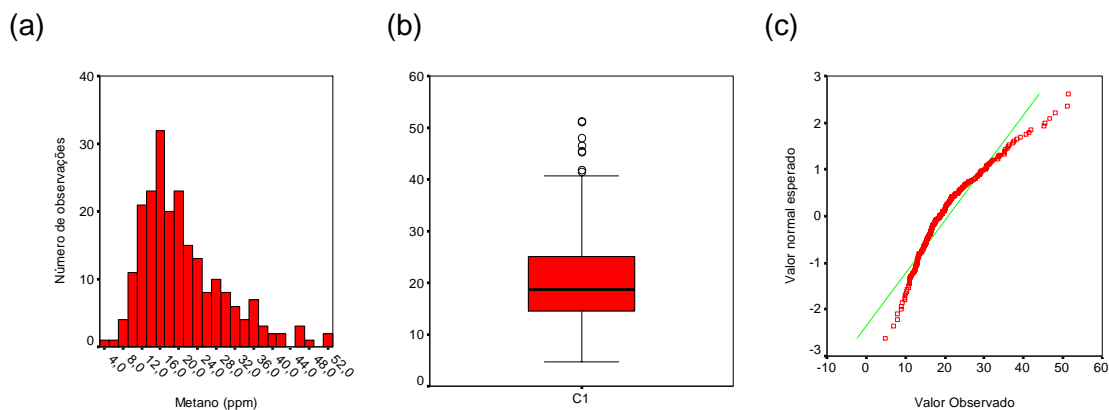


Figura 5.11 – (a) Histograma de freqüência, (b) gráfico *boxplot* e (c) de probabilidade para os dados de metano (ppm).

O histograma de freqüência e o gráfico *boxplot* apresentaram valores referentes à média e mediana diferente, e foram deslocados do centro da distribuição, indicando uma assimetria positiva. O gráfico de probabilidade normal para os dados de metano demonstrou que grande parte dos dados não está situada aleatoriamente sobre a reta do gráfico, representando um provável desvio da normalidade. No gráfico *boxplot*, observamos oito amostras com valores fora do intervalo dado pelo limite superior $[Q3+(1,5)*AIQ]$, sugerindo possíveis candidatos a valores discrepantes.

As estatísticas de assimetria e curtose apontam para um distanciamento da hipótese de normalidade. Estas suspeitas são confirmadas pelos testes formais de normalidade (Tabela 5.2), onde os valores de p-valor tanto para o teste de Shapiro-Wilk e Kolmogorov-Smirnov foram abaixo de 5%. Em ambos os casos, rejeita-se a hipótese nula de que a distribuição é normal.

Tabela 5.1 – Análise estatística descritiva das amostras originais (ppm).

ESTATÍSTICAS	C_1	C_2	C_2^-	C_3	C_3^-	$i-C_4$	C_4	C_4^-	$i-C_5$	C_5	C_6^+	C_2^-/C_2	C_2C_5	C_1/C_2C_5
N.º de observações	220	220	220	220	220	220	220	220	220	220	220	220	220	220
N.º de observações válidas	220	220	220	220	220	220	220	220	220	220	220	220	220	220
Média	20,79	2,44	8,51	1,21	5,37	0,17	0,56	0,92	0,30	0,53	1,43	3,82	5,21	4,52
Mediana	18,70	1,98	7,63	1,01	2,90	0,15	0,44	0,76	0,15	0,26	0,54	3,89	4,18	4,46
Valor mínimo	4,74	0,49	0,66	0,35	0,46	0,03	0,17	0,18	0,03	0,03	0,03	0,30	1,15	0,64
Valor máximo	51,33	8,81	31,53	4,08	73,19	1,01	1,91	3,39	4,42	6,49	38,61	11,60	20,03	8,62
Amplitude	46,59	8,32	30,87	3,73	72,73	0,99	1,74	3,21	4,39	6,47	38,59	11,20	18,88	7,98
Quartil Inferior	14,61	1,41	4,83	0,71	1,88	0,12	0,32	0,56	0,11	0,17	0,32	2,37	2,87	3,79
Quartil Superior	25,19	3,06	10,63	1,49	5,24	0,22	0,73	1,16	0,25	0,44	0,93	5,06	6,55	5,24
Amplitude Interquartilica	10,58	1,64	5,80	0,77	3,36	0,10	0,41	0,60	0,14	0,27	0,61	2,69	3,70	1,45
Variância	79,13	2,06	26,50	0,48	77,92	0,015	0,12	0,28	0,31	0,90	19,64	3,11	10,73	1,66
Desvio Padrão	8,90	1,43	5,15	0,69	8,83	0,12	0,34	0,53	0,56	0,95	4,43	1,76	3,28	1,29
Assimetria	1,10	1,64	1,24	1,56	4,85	2,45	1,61	1,55	4,80	4,22	6,92	0,42	1,71	0,16
Excesso de Curtose	1,06	2,95	1,96	2,56	27,58	11,31	2,57	2,83	25,46	18,50	51,66	0,83	3,18	0,94
Limite Inferior ($Q_1-1,5XAIQ$)	-1,26	-1,05	-3,87	-0,45	-3,16	-0,03	-0,30	-0,34	-0,10	-0,24	-0,60	-1,67	-2,68	1,62
Limite Superior ($Q_3+1,5XAIQ$)	41,06	5,52	19,33	2,65	10,28	0,37	1,35	2,06	0,46	0,85	1,85	9,10	12,10	7,42
CV	0,43	0,59	0,61	0,57	1,64	0,71	0,61	0,58	1,87	1,79	3,10	0,46	0,63	0,29

De acordo com a Tabela 5.1, para as demais variáveis, a média e a mediana também não apresentaram entre si valores próximos, apresentando uma variação mínima para as razões C_1/C_2C_5 e $C_2^{\bar{}}/C_2$, em que a variação ficou abaixo de 2%, sendo seguida por $C_2^{\bar{}}$ (11%) e máxima para C_6^+ (164%). Os valores de assimetria e excesso de curtose apresentaram valores positivos indicando uma distribuição assimétrica positiva e um comportamento leptocúrtico da curva de distribuição, tendo a variável C_6^+ apresentado o maior valor para ambos e as razões $C_2^{\bar{}}/C_2$ e C_1/C_2C_5 apresentando os menores valores para o excesso de curtose e assimetria respectivamente, indicando uma possível distribuição gaussiana.

Observando os valores do coeficiente de variação (CV), os menores valores também são das razões C_1/C_2C_5 (0,29) e $C_2^{\bar{}}/C_2$ (0,46) e a máxima para C_6^+ (3,10). Segundo Koch e Link (1971), para valores de coeficiente de variação menores que 0,3 ou 30%, a distribuição é normal ou muito aproximadamente normal, tendo a variável C_1/C_2C_5 apresentado valor de CV dentro dessa faixa. As variáveis ($C_3^{\bar{}}$, $i-C_5$, C_5 e C_6^+) apresentaram uma distribuição fortemente assimétrica ou mesmo lognormal, com coeficiente de variação 1,64, 1,87, 1,79 e 3,10 respectivamente. Para essas mesmas variáveis ($C_3^{\bar{}}$, $i-C_5$, C_5 e C_6^+) foram observados desvios padrão maiores do que a média. Quando isto acontece, o uso da distribuição normal fica inviabilizado (MACKAY & PATERSON, 1984).

As demais não permitem qualquer conclusão segundo Koch e Link (1971), já que ficaram com valores de CV entre 40% e 70%. Segundo Warrick e Nielsen (1980), pode-se classificar a variabilidade das variáveis como média ($0,12 < CV < 0,62$) para as variáveis C_1 , C_2 , $C_2^{\bar{}}$, C_3 , C_4 , $C_4^{\bar{}}$, $C_2^{\bar{}}/C_2$, C_2C_5 e C_1/C_2C_5 , e como alta ($CV > 0,62$) para as demais ($C_3^{\bar{}}$, $i-C_4$, $i-C_5$, C_5 e C_6^+).

Os *boxplots* mostram valores superiores ao limite superior para todas as variáveis, sendo a variável $C_2^{\bar{}}/C_2$ a que apresenta a menor quantidade de valores *outliers* (dois). A variável C_1/C_2C_5 é a única que apresenta tanto valores maiores como menores que o limite superior e inferior respectivamente, sugerindo possíveis candidatos a valores discrepantes.

As estatísticas para as variáveis C_1/C_2C_5 e $C_2^{\bar{}}/C_2$ não apontam claramente um distanciamento da hipótese de normalidade quanto às demais. Analisando estas suspeitas pelo testes formais de normalidade (Tabela 5.2), podemos observar que

somente para C_2^-/C_2 não se rejeita a hipótese nula de que a distribuição é normal de acordo com K-S (valor em negrito), para as demais variáveis rejeita-se a hipótese nula de que a distribuição é normal.

Tabela 5.2 – Testes formais de normalidade Shapiro-Wilk (S-W) e Kolmogorov-Smirnov (K-S) para os dados originais.

#	C_1	C_2	C_2^-	C_3	C_3^-	$i-C_4$	C_4	C_4^-	$i-C_5$	C_5	C_6^+	C_2^-/C_2	C_2C_5	C_1/C_2C_5
S-W (*)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,002
K-S (*)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,079	0,000	0,004

*p-valor

Podemos observar graficamente a variável C_2^-/C_2 (Figura 5.12), e as demais variáveis que como o metano (Figura 5.11) tiveram rejeitada a hipótese nula de que os dados seguem uma distribuição normal (Figuras 5.13a, 5.13b e 5.13c).

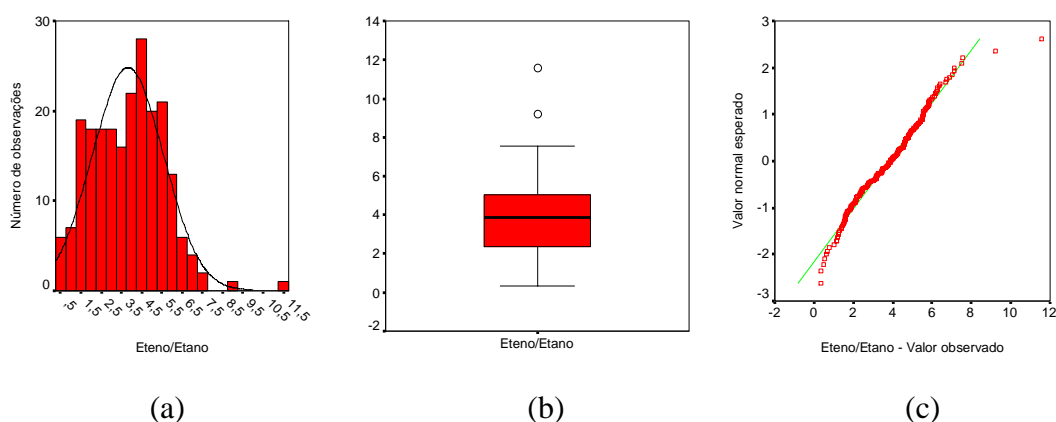


Figura 5.12 – (a) Histograma de freqüência, (b) gráfico *boxplot* e (c) de probabilidade para a variável C_2^-/C_2 .

Como podemos verificar através dos histogramas, *boxplots* e gráficos de probabilidade, a maioria das variáveis não apresenta um ajuste ao comportamento normal. Estas observações sugerem a existência significativa de *outliers* para estas variáveis e talvez um melhor ajuste a distribuição lognormal como descrito por Ahrens (1953, 1954a, 1954b, 1957).

Do ponto de vista estatístico, estas *outliers* deveriam ser removidas, no entanto, do ponto de vista geoquímico, estes valores extremos podem ter significado, constituindo anomalias geoquímicas. Segundo Reimann *et. al.* (2005), a detecção de dados *outliers* é uma das principais tarefas da análise estatística de dados de geoquímica. Para minorar eventuais efeitos nocivos provocados por esses fenômenos (existência de “*outliers*” e comportamento lognormal), foi usada a transformação (ln, log10 e raiz quadrada (*Sqrt*)) dos dados, evitando assim a remoção de *outliers* e verificando se os dados geoquímicos melhor se ajustam ao comportamento lognormal.

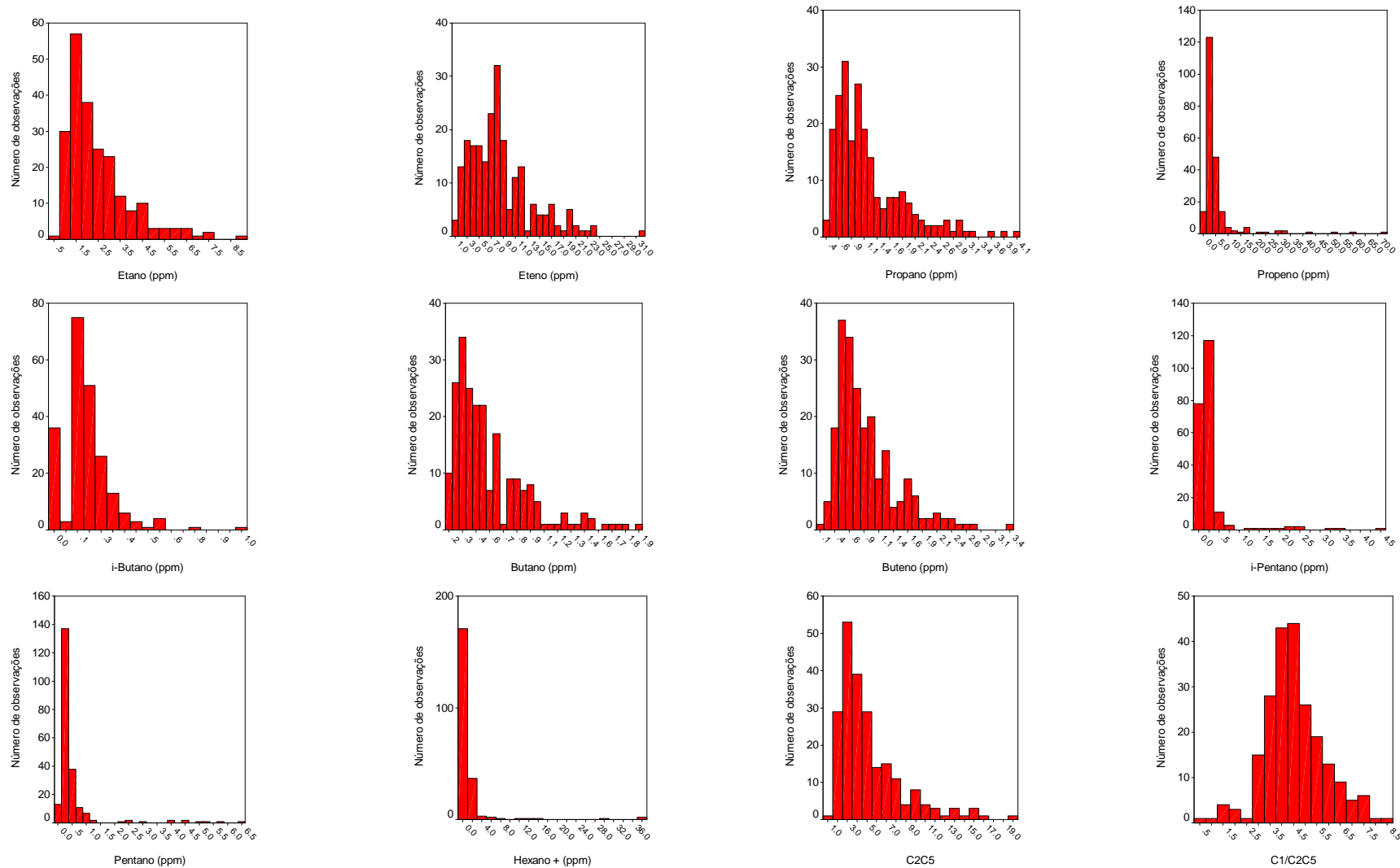


Figura 5.13a - Histograma de frequência para as variáveis C_2 , C_2^- , C_3 , C_3^- , $i-C_4$, C_4 , C_4^- , $i-C_5$, C_5 , C_6^+ , C_2C_5 e C_1/C_2C_5 .

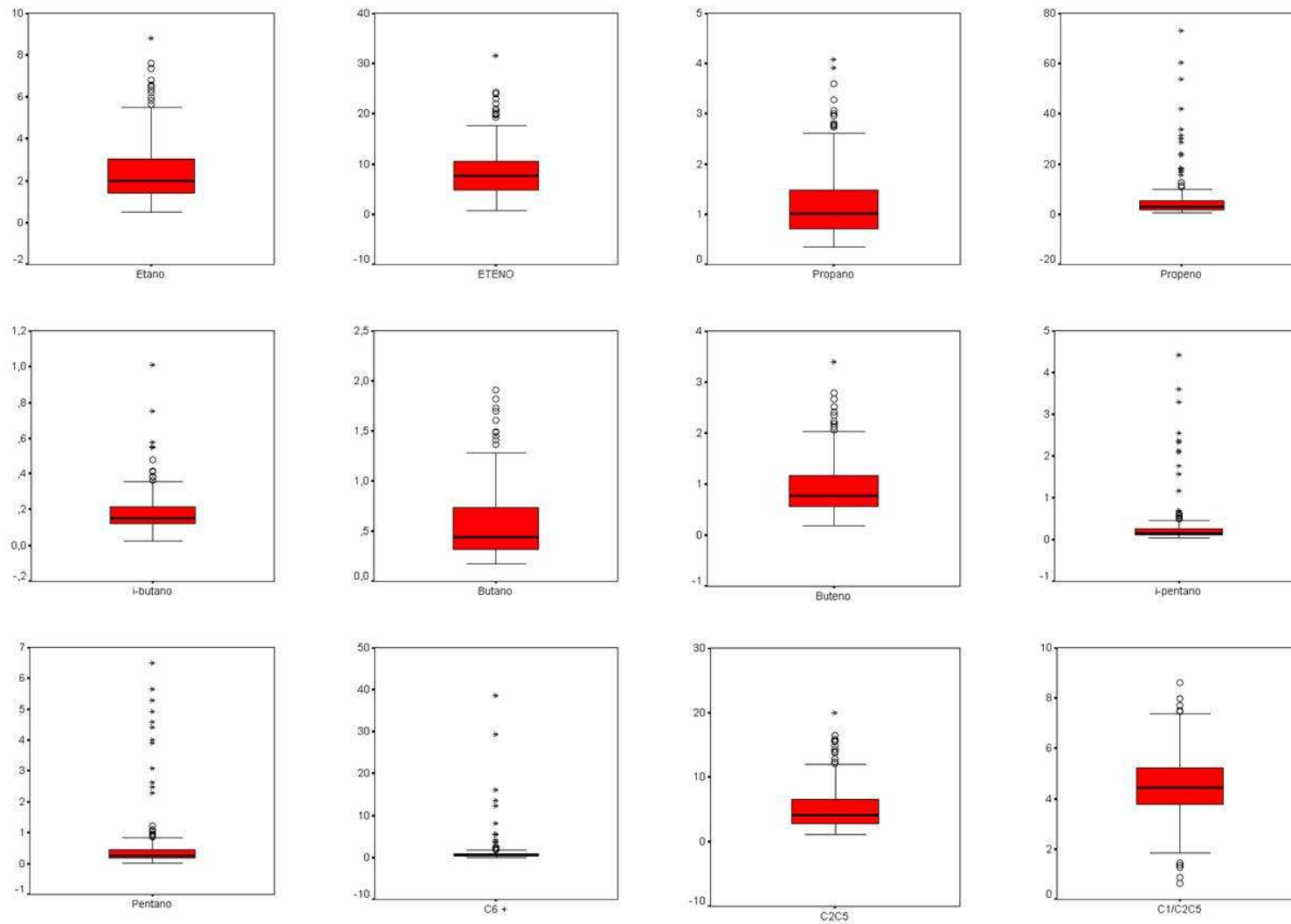


Figura 5.13b - *Boxplot* para as variáveis C_2 , C_2^- , C_3 , C_3^- , $i-C_4$, C_4 , C_4^- , $i-C_5$, C_5 , C_6^+ , C_2C_5 e C_1/C_2C_5 .

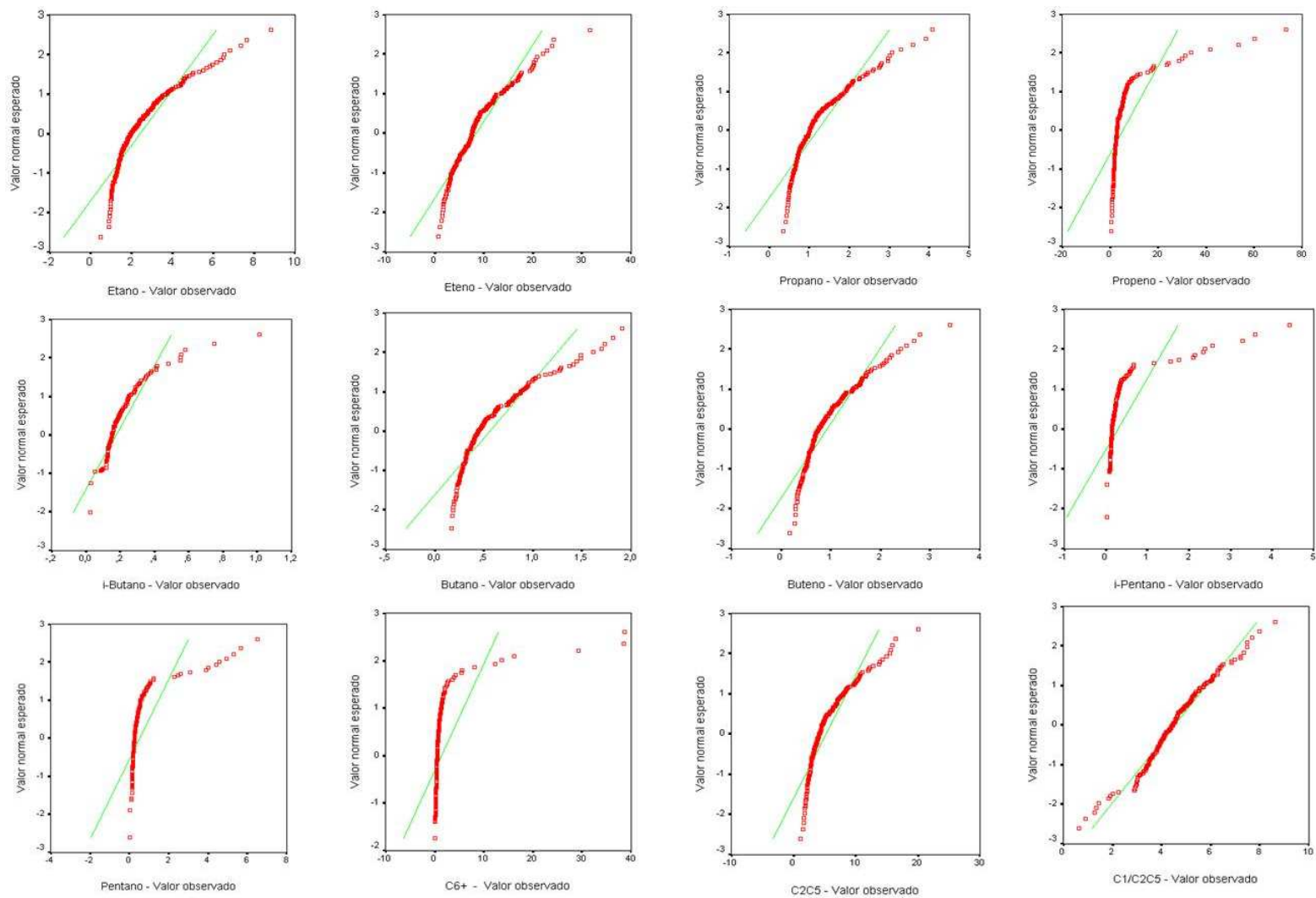


Figura 5.13c - Gráfico de probabilidade para as variáveis C_2 , C_2^- , C_3 , C_3^- , $i-C_4$, C_4 , C_4^- , $i-C_5$, C_5 , C_6^+ , C_2C_5 e C_1/C_2C_5 .

5.1.2 Dados transformados (ln)

As estatísticas descritivas para os dados transformados (ln) das variáveis C_1 , C_2 , C_2^- , C_3 , C_3^- , $i-C_4$, C_4 , C_4^- , $i-C_5$, C_5 e C_6^+ , e das razões C_2C_5 e C_1/C_2C_5 estão apresentados na Tabela 5.3. A variável C_2^-/C_2 não foi transformada devido ao fato de os dados seguirem uma distribuição normal como demonstrado no item anterior.

A Tabela 5.3 mostra que a variação entre a média e mediana de forma geral ficou menor após a transformação, mas a transformação dos dados não aproximou de fato numa distribuição lognormal.

Os valores de assimetria e excesso de curtose apresentaram também valores mais próximos da normalidade, obtendo menores valores de assimetria (0,04) e excesso de curtose (0,01) para a variável $\ln C_1$, e maiores valores para a variável $\ln C_1/C_2C_5$ (-1,86 e 7,16 respectivamente), onde no item anterior ficou próxima a normalidade. As variáveis $\ln C_2^-$, $\ln i-C_4$, $\ln C_6^+$, e $\ln C_1/C_2C_5$ tiveram seus valores de média menores que a mediana, apresentando assimetria negativa, ou seja, tiveram seus dados levemente concentrados à direita, para as demais variáveis, a assimetria é positiva, com maior acúmulo de dados à esquerda.

A curtose dessa série de dados indica uma menor curva para a variável $\ln C_1$, seguida das variáveis $\ln C_4^-$ e $\ln i-C_4$, que distanciam do valor de uma curva normal em 0,3%, -8,3% e 9,3% respectivamente; as demais variáveis têm seus dados variando de -10,0% ($\ln C_2$) a 238,0% ($\ln C_1/C_2C_5$). As variáveis $\ln C_2$, $\ln C_3$, $\ln C_4$, $\ln C_4^-$ e $\ln C_2C_5$ indicam valores de curvas 10,0%, 14,0%, 15,0%, 8,3% e 10,3% aquém do valor da curtose de uma curva normal, indicando curvas mais achatadas que a normal e, por esse motivo, classificadas como platicúrtica. Tal perfil indica certo afastamento dos dados em relação ao valor da média, provocando um alongamento horizontal da curva e seu achatamento vertical. Já as variáveis $\ln C_1$, $\ln C_2^-$, $\ln C_3^-$, $\ln i-C_4$, $\ln i-C_5$, $\ln C_5$, $\ln C_6^+$ e $\ln C_1/C_2C_5$ apresentam valores para curtose além do valor para uma curva normal (0,3%; 16,7%; 65,3%; 9,3%; 55,0%; 73,7%; 73,0% e 238,7% respectivamente), o que significa que as distribuições são leptocúrticas e, por conseguinte, evidenciam um excesso de curtose em relação à distribuição Gaussiana. Somente para a variável $\ln C_1$, a assimetria (0,04) e excesso de curtose (0,01) obtiveram valores próximos à zero, sendo a variável que mais se assemelha com uma distribuição normal.

Tabela 5.3 – Análise estatística descritiva das amostras transformadas (ln).

ESTATÍSTICAS	$\ln C_1$	$\ln C_2$	$\ln C_2^=$	$\ln C_3$	$\ln C_3^=$	$\ln i-C_4$	$\ln C_4$	$\ln C_4^=$	$\ln i-C_5$	$\ln C_5$	$\ln C_6^+$	$\ln C_2 C_5$	$\ln C_1 / C_2 C_5$
N.º de observações	220	220	220	220	220	220	220	220	220	220	220	220	220
N.º de observações válidas	220	220	220	220	220	220	220	220	220	220	220	220	220
Média	2,95	0,75	1,95	0,05	1,20	-2,00	-0,74	-0,22	-1,83	-1,23	-0,67	1,49	1,46
Mediana	2,93	0,66	2,03	0,01	1,07	-1,90	-0,82	-0,27	-1,90	-1,34	-0,62	1,43	1,49
Valor mínimo	1,56	-0,71	-0,42	-1,05	-0,80	-3,70	-1,77	-1,71	-3,69	-3,69	-3,69	0,14	-0,45
Valor máximo	3,94	2,18	3,45	1,40	4,3	0,00	0,65	1,22	1,49	1,87	3,65	3,00	2,15
Amplitude	2,38	2,89	3,87	2,45	5,1	3,70	2,42	2,93	5,18	5,56	7,34	2,86	2,60
Quartil Inferior	2,68	0,35	1,58	-0,34	0,63	-2,11	-1,14	-0,58	-2,18	-1,76	-1,14	1,05	1,13
Quartil Superior	3,23	1,12	2,37	0,40	1,66	-1,52	-0,31	0,15	-1,39	-0,82	-0,08	1,88	1,65
Amplitude Interquartílica	0,55	0,77	0,79	0,74	1,03	0,59	0,83	0,73	0,79	0,94	1,07	0,83	0,32
Variância	0,17	0,27	0,42	0,26	0,70	0,63	0,30	0,28	0,96	0,92	1,67	0,30	0,12
Desvio Padrão	0,41	0,52	0,65	0,51	0,84	0,79	0,54	0,53	0,98	0,96	1,29	0,54	0,34
Assimetria	0,04	0,38	-0,55	0,39	1,06	-0,81	0,41	0,19	0,50	0,54	-0,11	0,42	-1,86
Excesso de Curtose	0,01	-0,30	0,50	-0,42	1,96	0,28	-0,45	-0,25	1,65	2,21	2,19	-0,31	7,16
Limite Inferior ($Q_1 - 1,5XAIQ$)	1,86	-0,81	0,40	-1,45	-0,92	-3,00	-2,39	-1,68	-3,37	-3,17	-2,75	-0,20	0,65
Limite Superior ($Q_3 + 1,5XAIQ$)	4,06	2,28	3,56	1,51	3,21	-0,64	0,94	1,25	-0,21	0,59	1,53	3,13	2,13
CV	0,14	0,69	0,33	10,20	0,70	-0,40	-0,73	-2,41	-0,54	-0,78	-1,93	0,36	0,23

Os valores máximos para as variáveis $\ln C_3^-$, $\ln i-C_4$, $\ln i-C_5$, $\ln C_5$, $\ln C_6^+$ e $\ln C_1/C_2C_5$ estão fora do intervalo dado pelo limite superior e as variáveis $\ln C_1$, $\ln C_2^-$, $\ln C_2^-$, $\ln C_3$, $\ln C_4$, $\ln C_4^-$ e $\ln C_2C_5$ estão fora do intervalo dado pelo limite inferior, sugerindo possíveis candidatos a valores discrepantes.

Segundo Warrick e Nielsen (1980), pode-se classificar a variabilidade como média ($0,12 < CV < 0,62$) para as variáveis $\ln C_1$, $\ln C_2^-$, $\ln i-C_4$, $\ln i-C_5$, $\ln C_2^-/C_2$, $\ln C_2C_5$ e $\ln C_1/C_2C_5$, e como alta ($CV > 0,62$) para as demais ($\ln C_2$, $\ln C_3$, $\ln C_3^-$, $\ln C_4$, $\ln C_4^-$, $\ln C_5$ e $\ln C_6^+$), tendo apresentado o valor mais baixo para $\ln C_1$ (0,14) e valor mais alto para $\ln C_3$ (10,2).

As estatísticas para as variáveis apontam um distanciamento da hipótese de normalidade, sendo a variável $\ln C_1$ a que mais se assemelha com uma distribuição normal. Analisando os dados pelo testes formais de normalidade (Tabela 5.4), podemos confirmar que não se rejeita a hipótese nula de que a distribuição é normal para a variável $\ln C_1$ pelos testes S-W e K-S e para a variável $\ln C_4^-$ pelo teste S-W (valores em negrito), já para as demais variáveis rejeita-se a hipótese nula de que a distribuição é lognormal.

Tabela 5.4 – Testes formais de normalidade Shapiro-Wilk (S-W) e Kolmogorov-Smirnov (K-S) para os dados transformados.

#	$\ln C_1$	$\ln C_2$	$\ln C_2^-$	$\ln C_3$	$\ln C_3^-$	$\ln i-C_4$	$\ln C_4$	$\ln C_4^-$	$\ln i-C_5$	$\ln C_5$	$\ln C_6^+$	$\ln C_2C_5$	$\ln C_1/C_2C_5$
S-W (*)	0,429	0,003	0,001	0,003	0,000	0,000	0,001	0,248	0,000	0,000	0,000	0,004	0,000
K-S (*)	0,200	0,022	0,000	0,013	0,000	0,000	0,025	0,020	0,000	0,000	0,000	0,018	0,000

*p-valor

Podemos verificar melhor a dispersão dos dados através dos gráficos *boxplots*, histogramas de frequência e de probabilidade para as variáveis transformadas $\ln C_2$, $\ln C_2^-$, $\ln C_3$, $\ln C_3^-$, $\ln i-C_4$, $\ln C_4$, $\ln i-C_5$, $\ln C_5$, $\ln C_6^+$, $\ln C_2C_5$ e $\ln C_1/C_2C_5$ (Figura 5.14a, 5.14b e 5.14c, respectivamente), as quais tiveram rejeitada a hipótese nula de que a distribuição é lognormal.

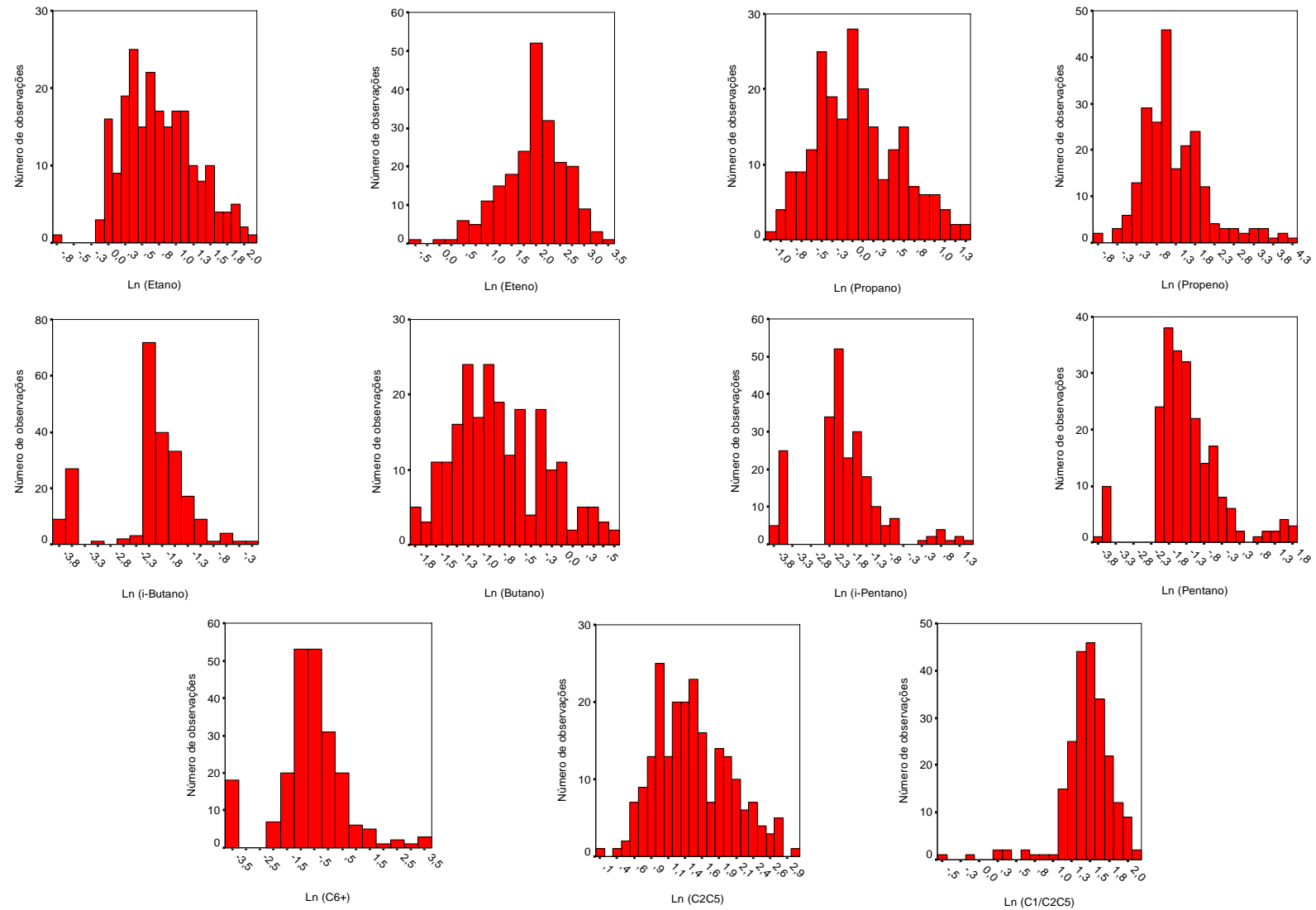


Figura 5.14a - Histograma de frequência para as variáveis transformadas (ln) C_2 , C_2^- , C_3 , C_3^- , $i-C_4$, C_4 , $i-C_5$, C_5 , C_6^+ , C_2C_5 e C_1/C_2C_5 .

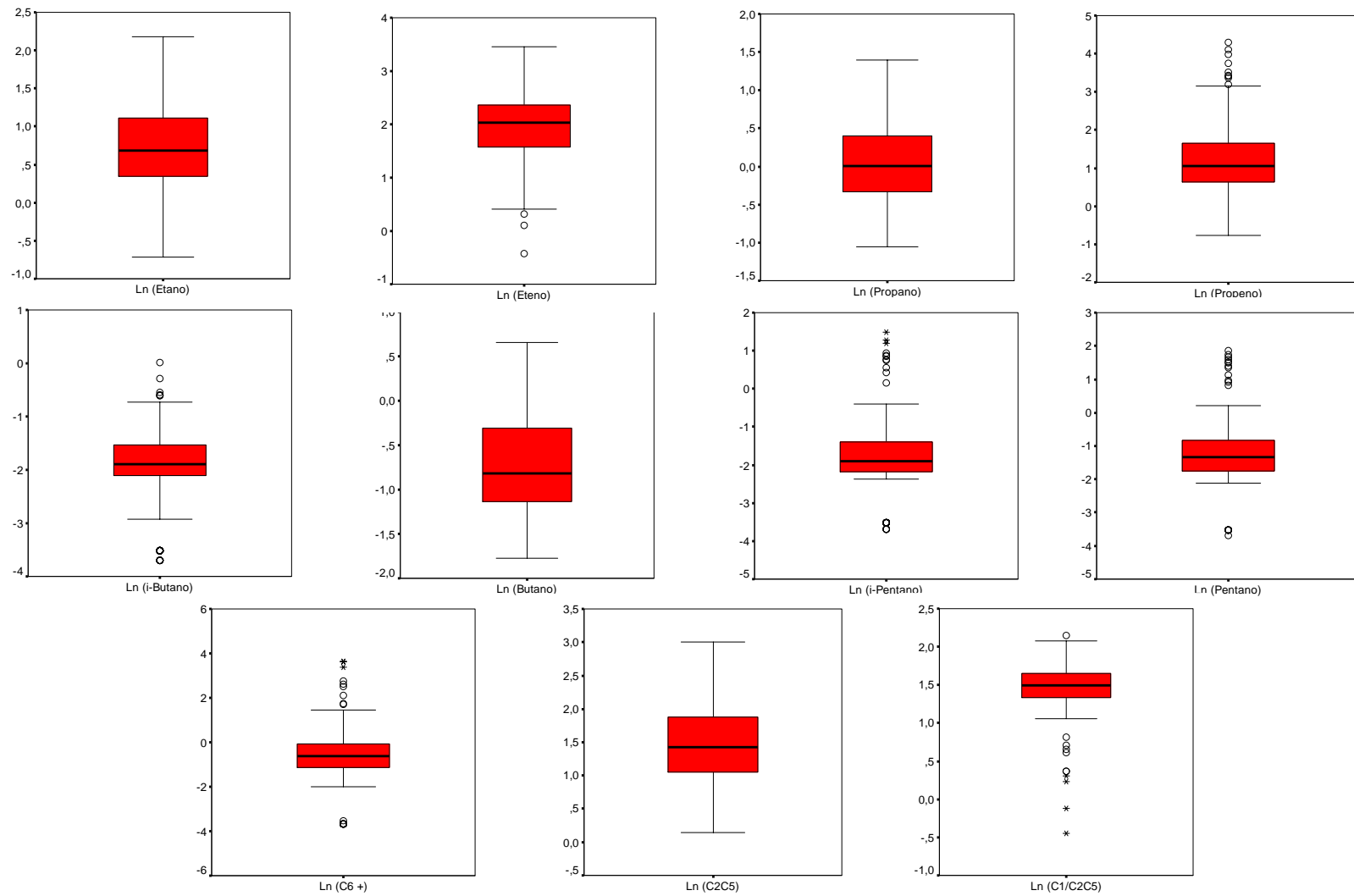


Figura 5.14b - Boxplot para as variáveis transformadas (ln) C_2 , C_2^- , C_3 , C_3^- , $i-C_4$, C_4 , $i-C_5$, C_5 , C_6^+ , C_2C_5 e C_1/C_2C_5 .

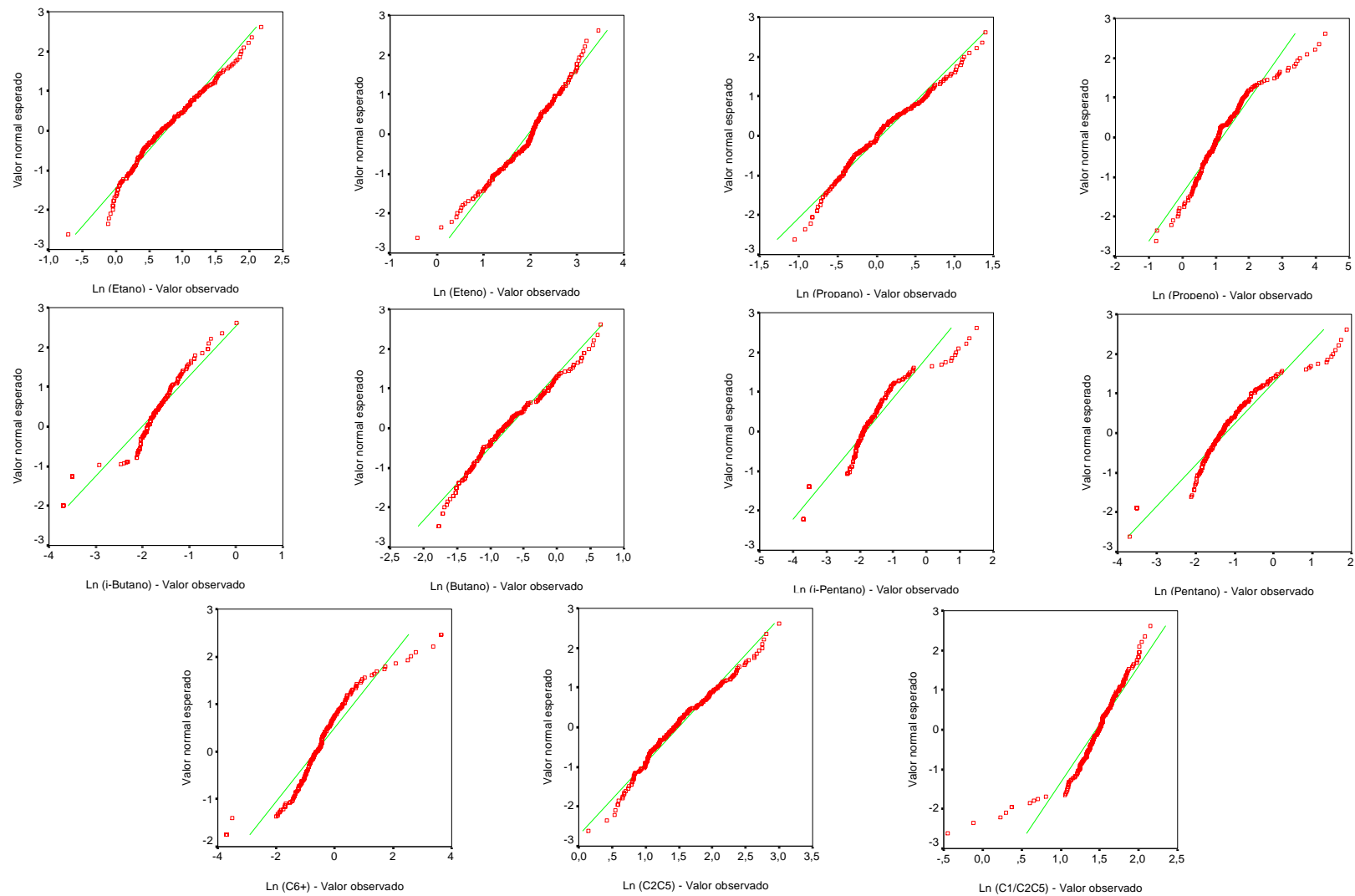


Figura 5.14c - Gráfico de probabilidade para as variáveis transformadas (ln) C_2 , C_2^- , C_3 , C_3^- , $i-C_4$, C_4 , $i-C_5$, C_5 , C_6^+ , C_2C_5 e C_1/C_2C_5 .

Os gráficos para a distribuição dos dados das variáveis $\ln C_1$ (Figura 5.15a) e $\ln C_4$ (Figura 5.15b), que de acordo com os testes de normalidade não se rejeita a hipótese nula que a distribuição é lognormal, podem ser observados abaixo. Observa-se uma leve assimetria positiva dos dados, uma boa distribuição dos dados no *boxplot* e em torno da reta de normalidade, já para as demais variáveis os pontos não situam aleatoriamente sobre a reta do gráfico, esses desvios dos pontos da reta representam um desvio da normalidade.

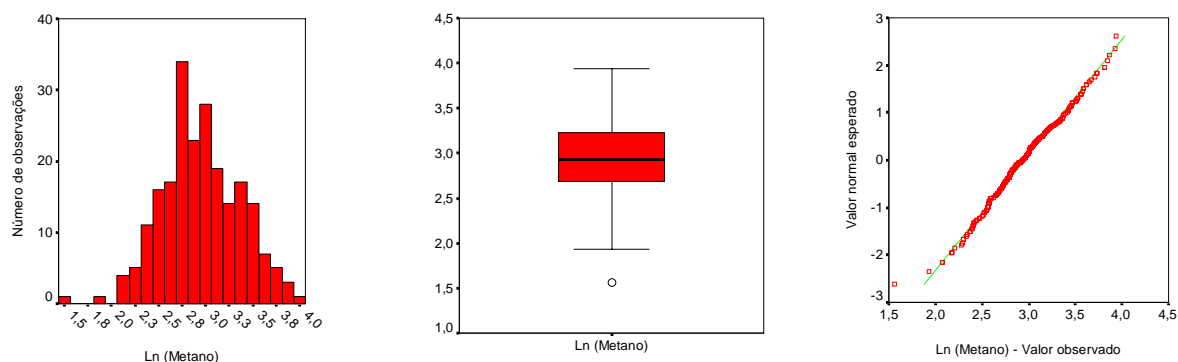


Figura 5.15a – Histograma de freqüência, gráfico *boxplot* e de probabilidade para a variável $\ln C_1$.

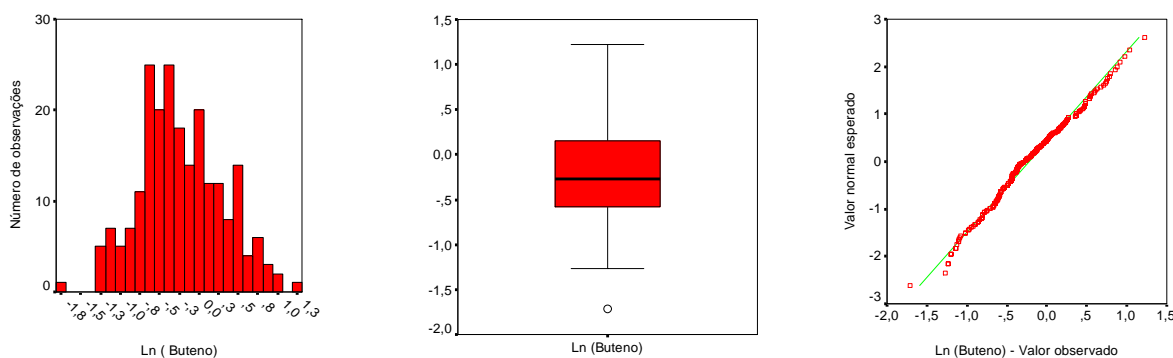


Figura 5.15b – Histograma de freqüência, gráfico *boxplot* e de probabilidade para a variável $\ln C_4$.

Fazendo a transformação dos dados por outros métodos (Tabela 5.5), podemos verificar que os dados transformados (\log_{10}) tiveram os mesmos resultados dos testes de normalidade para as mesmas variáveis ($\ln C_1$ e $\ln C_4$). Já para os dados transformados pela raiz quadrada (*Sqrt*), o teste de normalidade mostra que não se rejeita a hipótese nula de que a distribuição é normal somente para a variável $\ln C_1/C_2C_5$ (valores em negrito).

Tabela 5.5 - Testes formais de normalidade Shapiro-Wilk (S-W) e Kolmogorov-Smirnov (K-S) para dados originais e transformados (ln, log10 e Sqrt).

	Orig.		Ln		Log10		Sqrt	
	S-W	K-S	S-W	K-S	S-W	K-S	S-W	K-S
C1	0.000	0.000	0.429	0.200	0.429	0.200	0.000	0.003
C2	0.000	0.000	0.003	0.022	0.003	0.036	0.000	0.000
C2 ⁼	0.000	0.000	0.001	0.000	0.002	0.000	0.015	0.000
C3	0.000	0.000	0.003	0.013	0.003	0.023	0.000	0.000
C3 ⁼	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IC4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C4	0.000	0.000	0.001	0.025	0.001	0.032	0.000	0.000
C4 ⁼	0.000	0.000	0.248	0.020	0.250	0.015	0.000	0.000
IC5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C6 ⁺	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ete/Eta	0.001	0.079	0.000	0.000	0.000	0.000	0.002	0.004
C2C5	0.000	0.000	0.004	0.018	0.003	0.047	0.000	0.000
C1/C2C5	0.002	0.004	0.000	0.000	0.000	0.000	0.000	0.099

5.2 Análise de *background* e anomalia

Na definição do limiar, separando os dados de *background* de anômalos, foram utilizados quatro métodos: A regra [média+2 desvios padrão (dp)]; o *boxplot* [*upper whisker*], [mediana (med) + 2 desvios mediano absoluto (dma)] e 98º percentil.

Quando analisamos os dados originais e estimamos os limites de variação de *background* usando diferentes procedimentos (Tabela 5.6), o método [mediana (med) + 2 desvios mediano absoluto (dma)] entrega os mais baixos limiares e identifica, conseqüentemente, os mais altos números de *outliers* (uma média de 19%), seguidos pelo *boxplot* [*upper whisker*] (média de 5,9%). A razão pela qual apresentam uma larga proporção de extremos valores identificados é que os estimadores robustos de posição (mediana) e escala (desvio mediano absoluto (med) ou amplitude interquartil (AIQ)) são relativamente não influenciados por extremos valores de uma distribuição de dados lognormal ou fortemente assimétrica. Tanto a posição quanto a escala são

relativamente baixas em relação à média e desvio padrão, resultando em valores mais baixos de limiares e maiores porcentagens de valores extremos identificados (Tabela 5.7).

O limiar definido pelo *boxplot* [*upper whisker*] é em todos os casos mais perto, porém mais baixo que o limite obtido por dados logtransformados usando a regra [média+2 desvios padrão (dp)].

A Tabela 5.7 mostra que o clássico método, [média+2 desvios padrão (dp)], detectou para os dados transformados que segue uma distribuição lognormal (C_1 e C_4^-), um percentual de 2,7% de extremos valores, próximo do valor de 2,5% como esperado da teoria. Reimann *et al.* (2005), demonstra que para uma grande quantidade de amostras com distribuição normal, a quantidade de extremos valores (*outliers*) é igual à esperada pela teoria descrita por Hawkes e Webb (1962).

A regra [média+2 desvios padrão (dp)], na maioria dos casos (8 variáveis), tem valores de limiar mais baixos que aqueles obtidos pelo método *boxplot* [*upper whisker*] (Tabela 5.7). Na Tabela 5.5, podemos verificar que a maioria dos dados originais apresenta baixos valores do limiar em relação aos dados logtransformados, alterando assim a quantidade de valores anômalos. A determinação de limiar por 98º percentil não alterou com a transformação dos dados, mostrando a mesma quantidade de valores *outliers* independente da distribuição dos dados. Os limiares dos dados transformados através do *boxplot* (*upper whisker*) apresentam na maioria das vezes valores iguais ou maiores que o valor máximo dos dados indicando ausência de dados *outliers*.

Os dados originais das variáveis C_3^- , $i-C_5$, C_5 e C_6^+ apresentaram uma distribuição fortemente assimétrica, com coeficiente de variação 1,64, 1,87, 1,79 e 3,10 respectivamente (Tabela 5.1). Esses valores de CV mostram desvios padrão maiores do que a média tendo gerado valores de limiar pela regra [média+2 desvios padrão (dp)] maiores em relação aos dados logtransformados.

Os exemplos mostrados demonstram a grande variedade de limiares que pode ser estimada dependendo do método e da transformação escolhidos. Contudo, fica a questão: Qual desses métodos de estimação de limiar é o melhor? Se é que algum deles serve para esta função. Segundo Reimann *et al.* (2005), esta questão não pode ser respondida, porque desconhecemos a verdadeira distribuição estatística e espacial dos dados reais (originais). Não obstante, a questão é bastante significativa, porque

deduções por meio de estatística podem ser questionáveis. Entretanto, elas servem para orientar na exploração geoquímica e estabelecer critérios. No entanto, a determinação do limiar entre anomalias e *background* deve ser definida levando em consideração os objetivos da pesquisa, o tamanho e a forma esperados da anomalia, a variação natural prevista nas medidas de superfície utilizando-se de todos os dados disponíveis da região, conhecimentos adquiridos e interdisciplinares.

Na Figura 5.16 através do *postplot*, pode-se observar de maneira mais clara como diferentes metodologias para a determinação de limiar impactam na quantidade de valores *outliers*. Os dados de *background* estão representados na cor azul clara e dados *outliers* na cor azul escura.

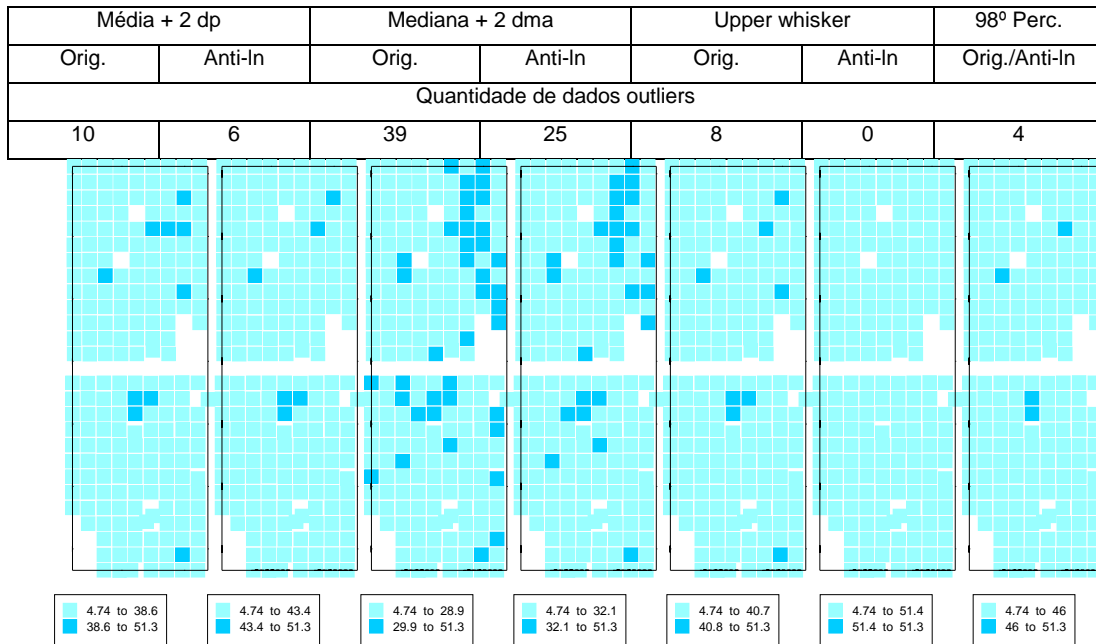
Tabela 5.6 – Média, desvio padrão (DP), mediana, desvio mediano absoluto (DMA), para os dados originais e transformados (ln) e resultados da definição do limiar superior via: [média+2 desvios padrão (dp)], [mediana (med) + 2 desvios medianos absolutos (dma)], o *boxplot* [upper whisker] e [98º percentil].

	Média		DP		Mediana		DMA		[Média+2DP]		[Med+2DMA]		Upper whisker		98º Percentil	
	Orig.	Ln	Orig.	Ln	Orig.	Ln	Orig.	Ln	Orig.	Anti-ln	Orig.	Anti-ln	Orig.	Anti-ln	Orig.	Anti-ln
C1	20.79	2.95	8.90	0.41	18.70	2.93	5.11	0.27	38.59	43.38	28.92	32.14	40.75	51.41	46.05	
C2	2.44	0.75	1.43	0.52	1.98	0.66	0.67	0.37	5.30	5.99	3.32	4.06	5.49	8.85	6.71	
C2 ⁺	8.51	1.95	5.15	0.65	7.63	2.03	2.92	0.39	18.81	25.79	13.47	16.61	17.63	31.81	22.52	
C3	1.21	0.05	0.69	0.51	1.01	0.01	0.33	0.36	2.59	2.92	1.67	2.08	2.61	4.09	3.19	
C3 ⁺	5.37	1.20	8.83	0.84	2.90	1.07	1.22	0.48	23.03	17.81	5.34	7.61	9.90	23.57	38.44	
IC4	0.17	-2.00	0.12	0.79	0.15	-1.90	0.04	0.26	0.41	0.66	0.23	0.25	0.35	0.48	0.55	
C4	0.56	-0.74	0.34	0.54	0.44	-0.82	0.16	0.36	1.24	1.40	0.76	0.90	1.28	1.92	1.66	
C4 ⁺	0.92	-0.22	0.53	0.53	0.76	-0.27	0.24	0.34	1.98	2.32	1.24	1.51	2.04	3.39	2.47	
IC5	0.30	-1.83	0.56	0.98	0.15	-1.90	0.05	0.39	1.42	1.14	0.25	0.33	0.44	0.68	2.48	
C5	0.53	-1.23	0.95	0.96	0.26	-1.34	0.11	0.46	2.43	1.99	0.48	0.66	0.84	1.23	4.78	
C6 ⁺	1.43	-0.67	4.43	1.29	0.54	-0.62	0.26	0.53	10.29	6.75	1.06	1.55	1.79	4.22	15.05	
Ete/Eta	3.82	1.20	1.76	0.58	3.89	1.36	1.36	0.34	7.34	10.59	6.61	7.69	7.56	11.60	7.37	
C2C5	5.21	1.49	3.28	0.54	4.18	1.43	1.44	0.40	11.77	13.07	7.06	9.30	11.96	20.08	15.58	
C1/C2C5	4.52	1.46	1.29	0.34	4.46	1.49	0.73	0.16	7.10	8.50	5.92	6.11	7.36	8.00	7.48	

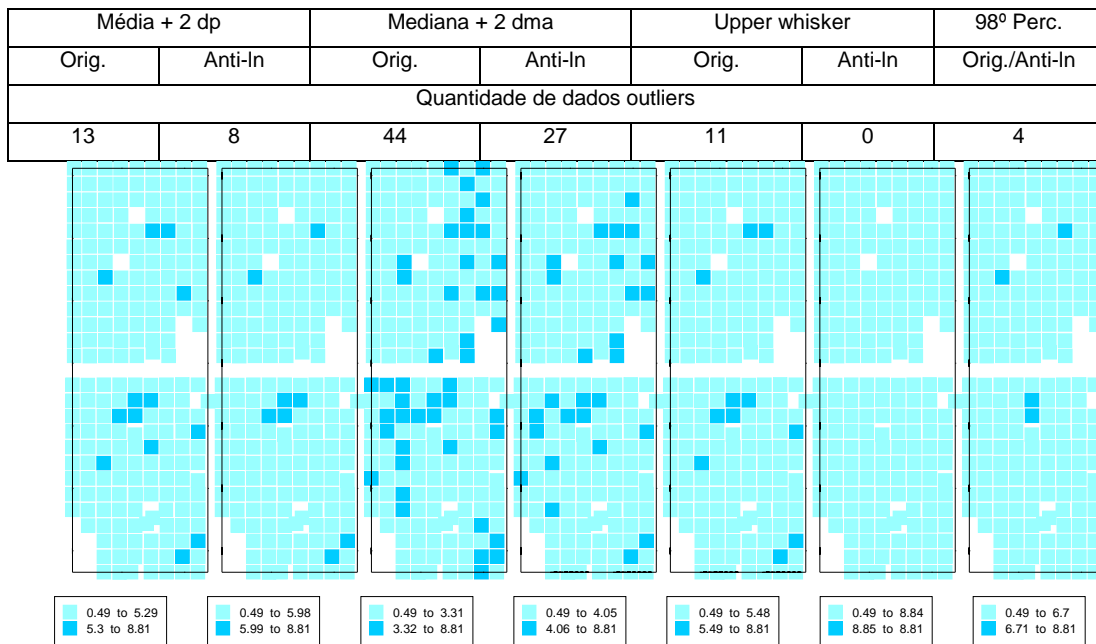
Tabela 5.7 – Quantificação de valores *outliers* amostrados utilizando o limiar superior via: [média+2 desvios padrão (dp)], [mediana (med) + 2 desvios medianos absolutos (dma)], o *boxplot* [*upper whisker*] e [98º percentil].

	Outliers							Outliers (%)								
	[Média+2DP]		[Med+2DMA]		Upper whisker		98º Percentil		[Média+2DP]		[Med+2DMA]		Upper whisker		98º Percentil	
	Orig.	Anti-Ln	Orig.	Anti-Ln	Orig.	Anti-Ln	Orig.	Anti-Ln	Orig.	Anti-Ln	Orig.	Anti-Ln	Orig.	Anti-Ln	Orig.	Anti-Ln
C1	10	6	39	25	8	0	4	4.55	2.73	17.73	11.36	3.64	0.00	1.82		
C2	13	8	44	27	11	0	4	5.91	3.64	20.00	12.27	5.00	0.00	1.82		
C2 ⁼	13	1	35	20	13	0	4	5.91	0.45	15.91	9.09	5.91	0.00	1.82		
C3	13	8	46	24	12	0	4	5.91	3.64	20.91	10.91	5.45	0.00	1.82		
C3 ⁼	10	12	51	25	18	9	4	4.55	5.45	23.18	11.36	8.18	4.09	1.82		
IC4	7	2	47	34	13	6	4	3.18	0.91	21.36	15.45	5.91	2.73	1.82		
C4	14	10	49	31	11	0	4	6.36	4.55	22.27	14.09	5.00	0.00	1.82		
C4 ⁼	12	6	47	32	11	0	4	5.45	2.73	21.36	14.55	5.00	0.00	1.82		
IC5	10	11	51	32	21	11	4	4.55	5.00	23.18	14.55	9.55	5.00	1.82		
C5	11	12	48	29	23	12	4	5.00	5.45	21.82	13.18	10.45	5.45	1.82		
C6 ⁺	6	7	46	25	21	9	4	2.73	3.18	20.91	11.36	9.55	4.09	1.82		
Ete/Eta	4	1	10	2	2	0	4	1.82	0.45	4.55	0.91	0.91	0.00	1.82		
C2C5	13	9	46	25	12	0	4	5.91	4.09	20.91	11.36	5.45	0.00	1.82		
C1/C2C5	10	1	30	23	6	1	4	4.55	0.45	13.64	10.45	2.73	0.45	1.82		

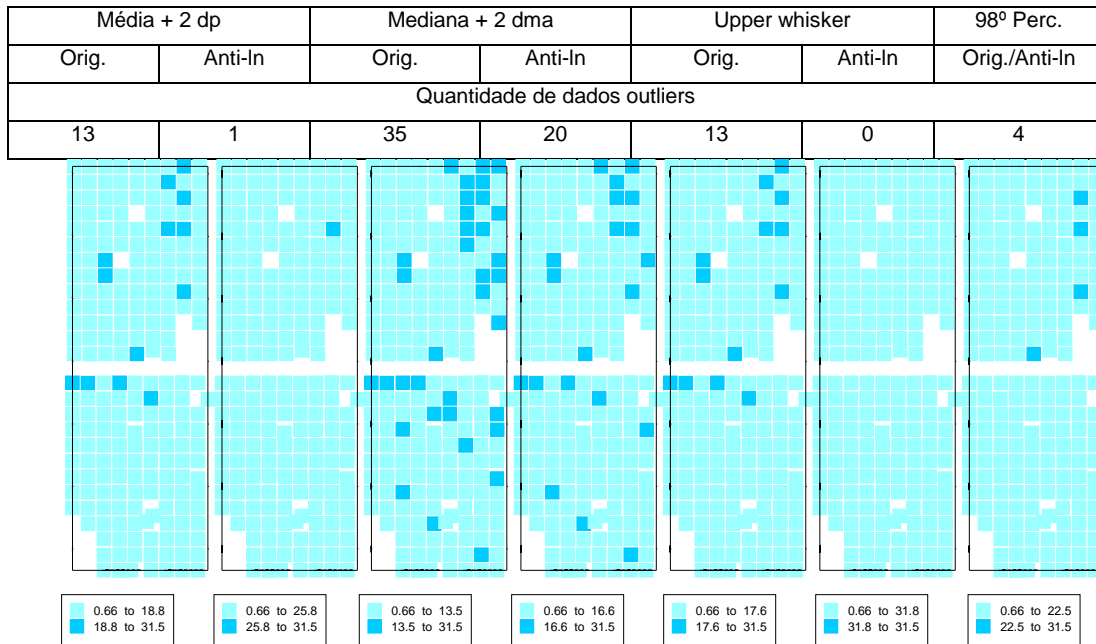
a) Metano



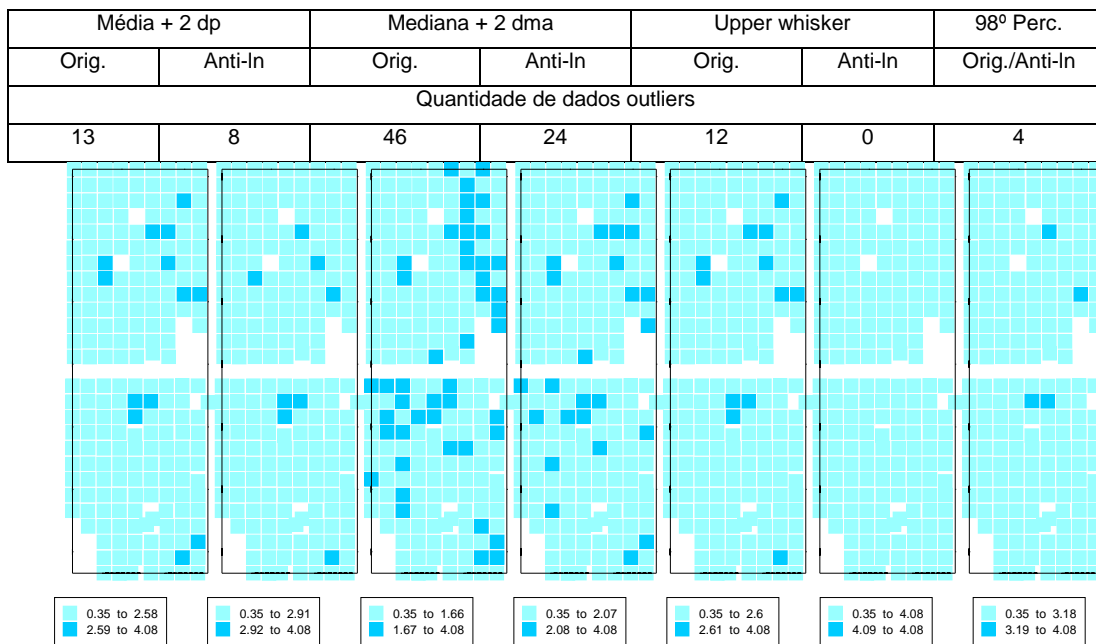
b) Etano



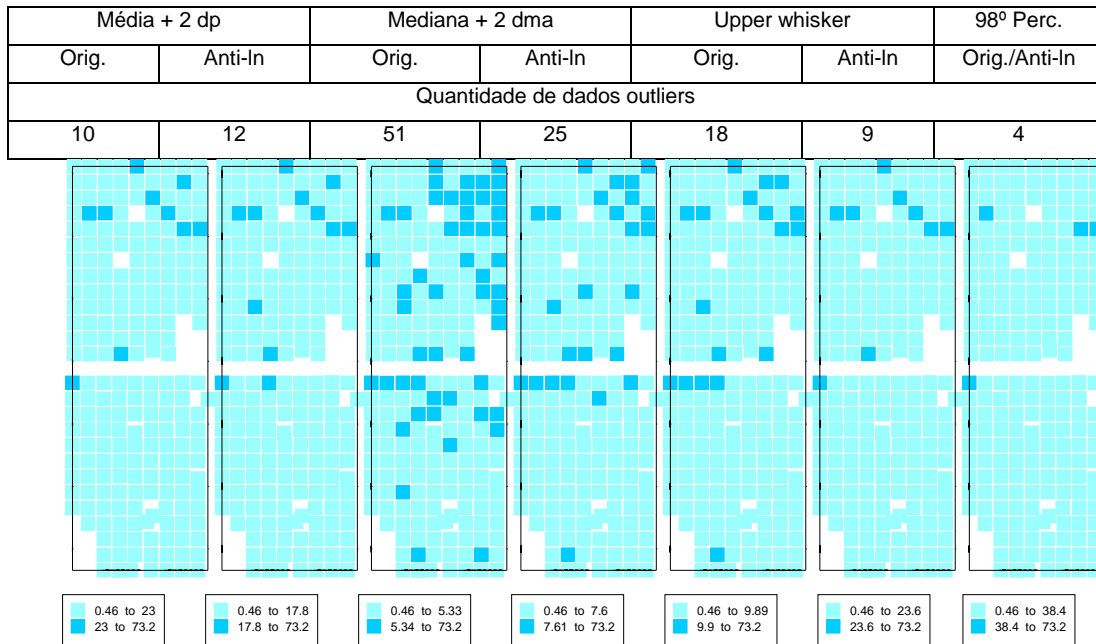
c) Eteno



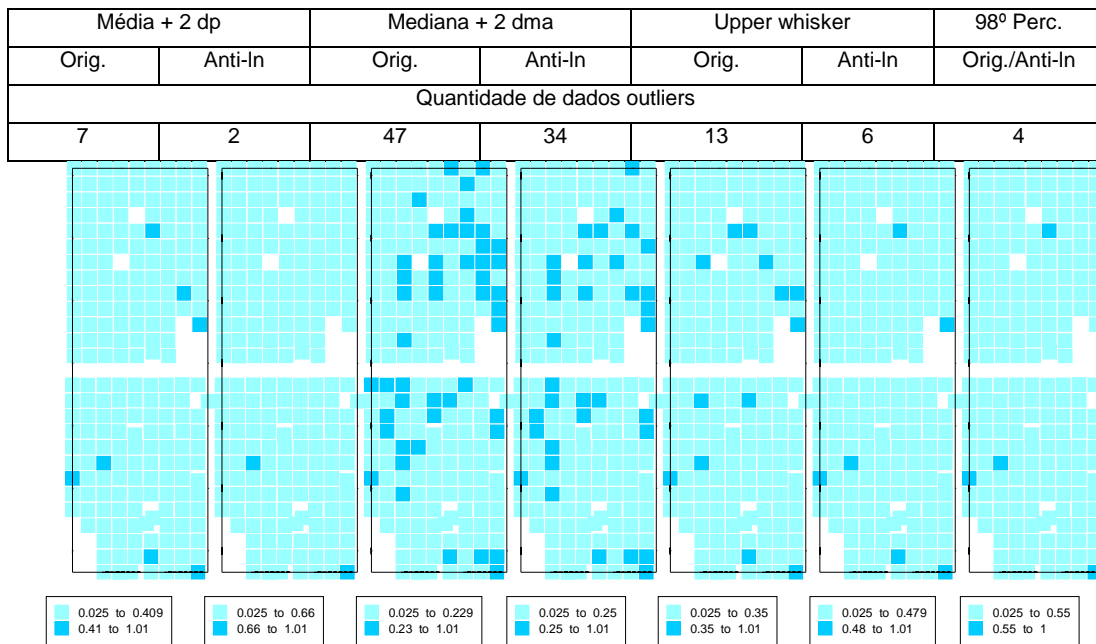
d) Propano



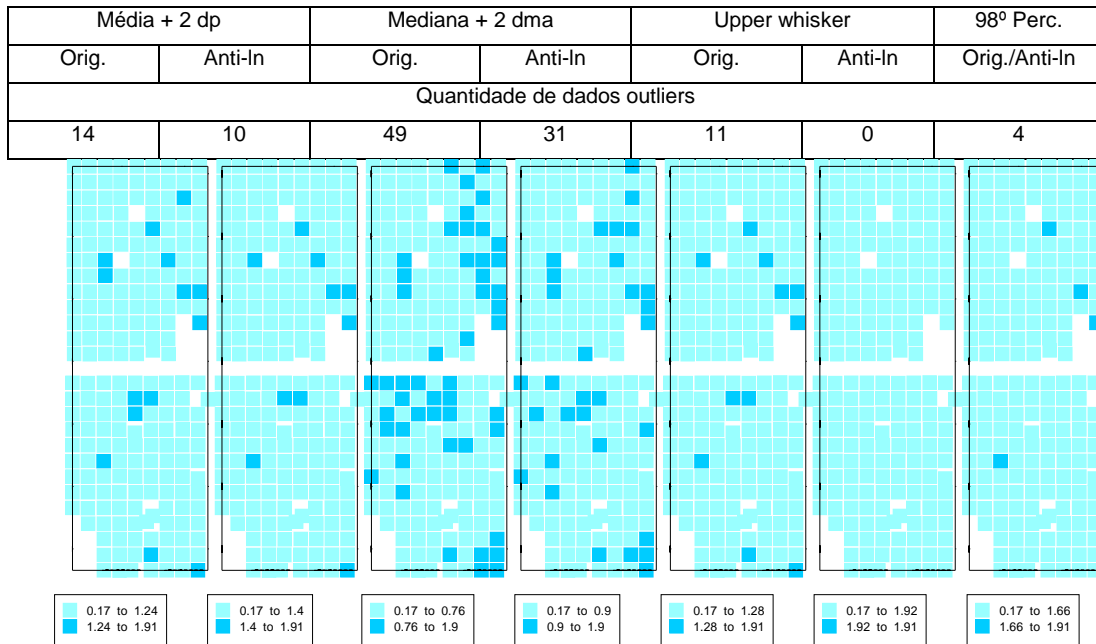
e) Propeno



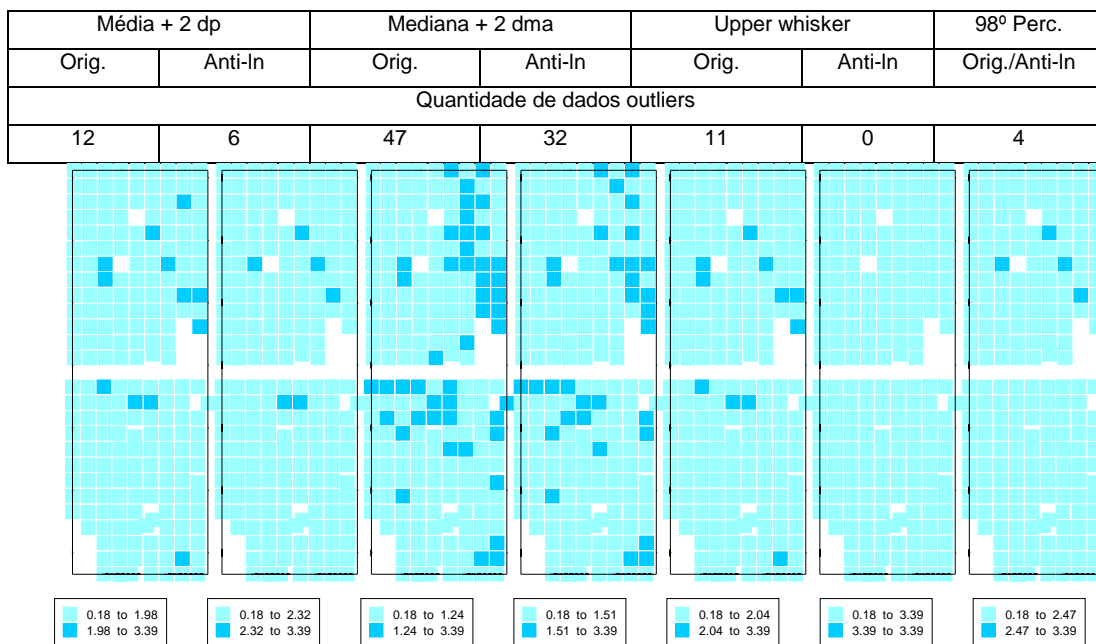
f) Isobutano



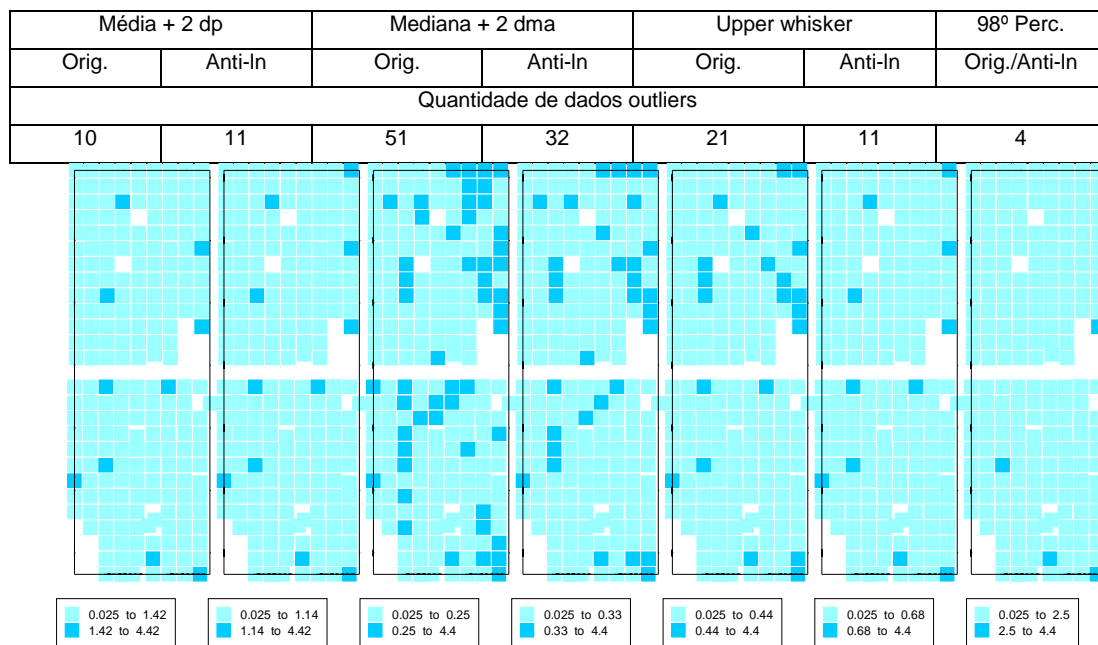
g) Butano



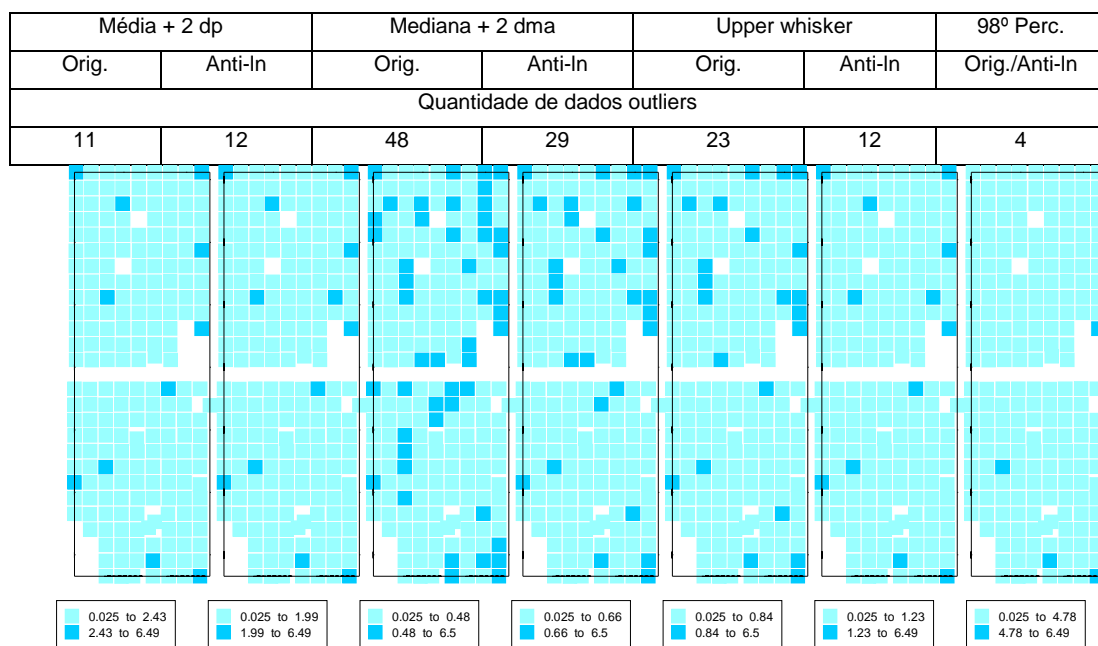
h) Buteno



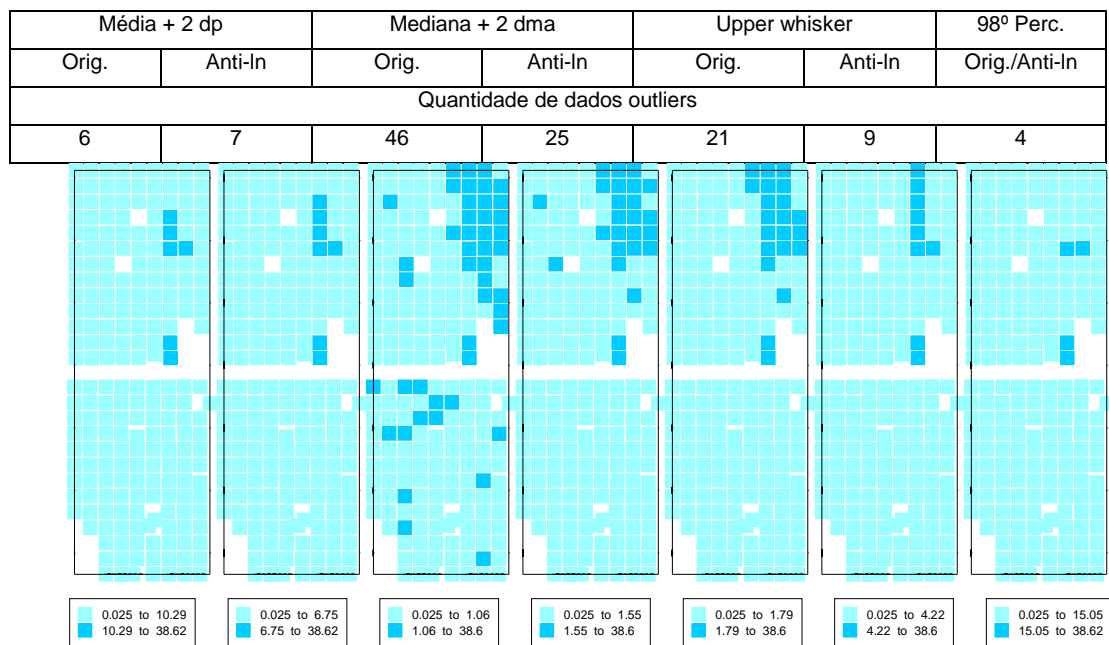
i) Isopentano



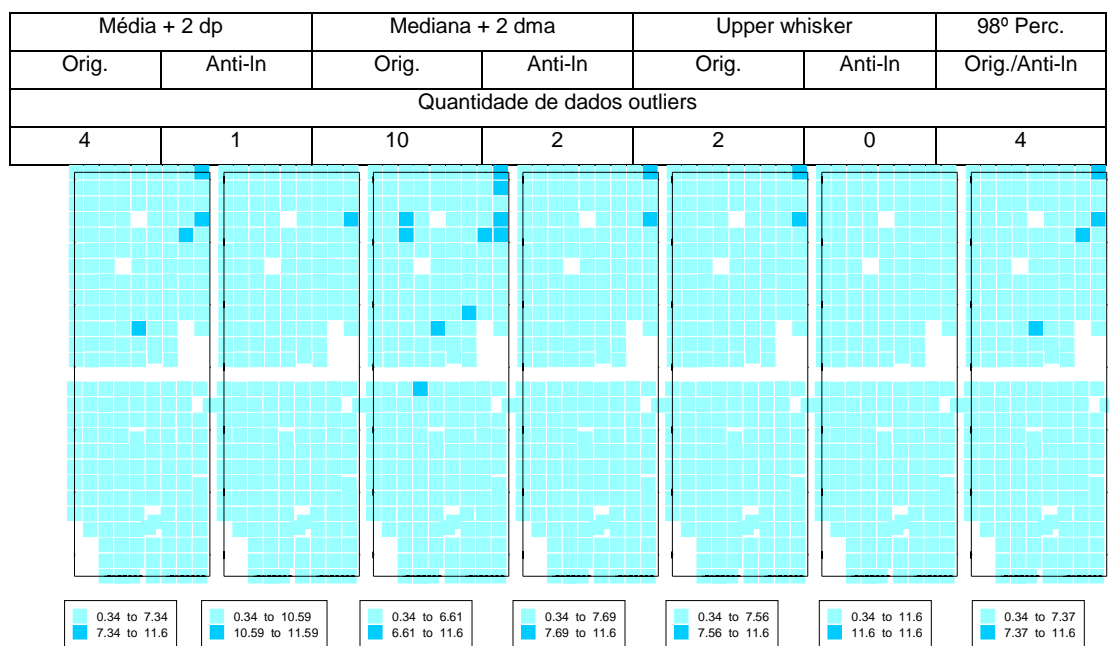
j) Pentano



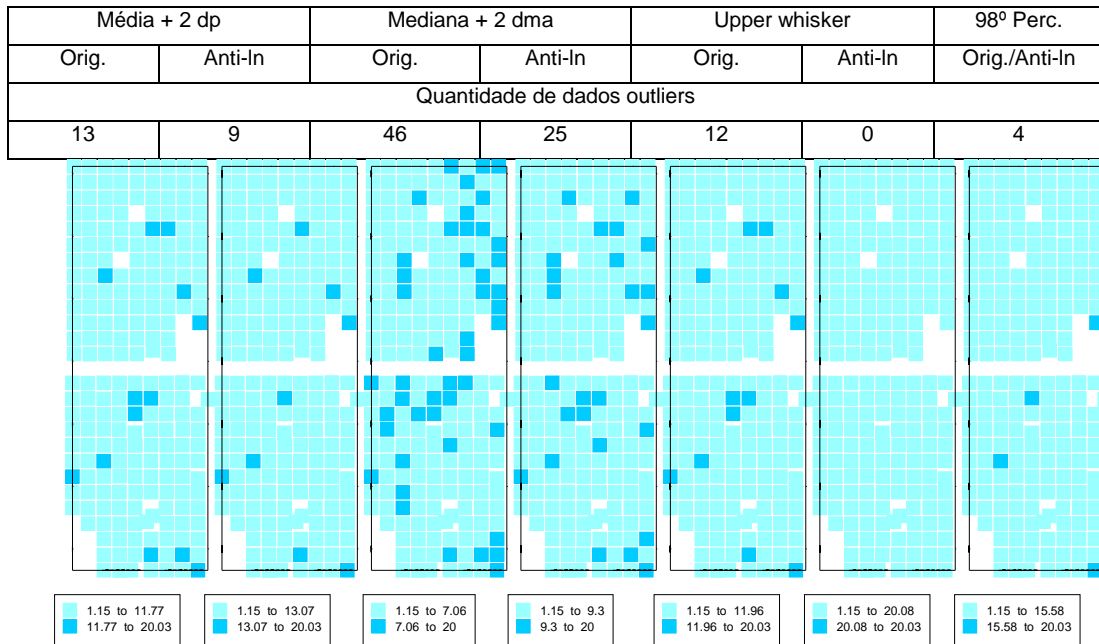
k) Hexano⁺



l) Eteno/Etano



m) C_2C_5



n) C_1/C_2C_5

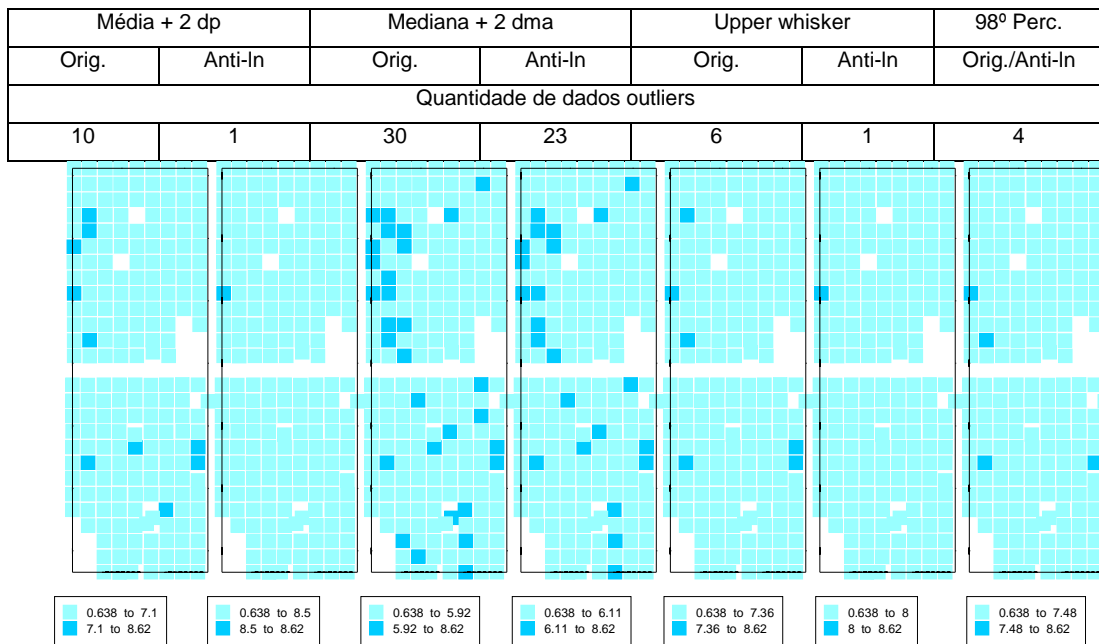


Figura 5.16 – *Postplot* dos valores de *background* (azul claro) e *outliers* (azul escuro) das variáveis C_1 (a), C_2 (b), C_2^- (c), C_3 (d), C_3^- (e), $i-C_4$ (f), C_4 (g), C_4^- (h), $i-C_5$ (i), C_5 (j), C_6^+ (k), C_2^-/C_2 (l), C_2C_5 (m) e C_1/C_2C_5 (n), utilizando o limiar superior via: [média + 2 desvios padrão], [mediana + 2 desvios mediano absoluto], *boxplot* (*upper whisker*) e 98° percentil para os valores originais e transformados, respectivamente.

5.3 Análise exploratória espacial

A análise exploratória espacial foi realizada para três variáveis (C_1 , C_2 e C_3), e para melhor entendimento, os resultados são discutidos separadamente.

- **Metano (C_1)**

Como visto anteriormente, os valores transformados (\ln) se ajustaram melhor a uma distribuição normal, adequando-se melhor ao tratamento geoestatístico. Observando a superfície dos valores (Figura 5.17a) da variável transformada metano, verificam-se aglomerações das medidas para valores próximos ao valor dado pelo limite superior [$q_3 + 1,5 \times AIQ$] em duas regiões distintas, uma localizada à direita superior do gráfico e a outra próxima ao centro.

Este fato pode estar associado a processos de migração por difusão, podendo esse gradiente de concentração estar acontecendo em subsuperfície logo abaixo desta, ou ao redor de fraturas e falhas, levando a valores maiores de metano. No gráfico de dispersão das medidas em relação às linhas (Figura 5.17b) e colunas (Figura 5.17c), notam-se pequenas concentrações de valores nas últimas linhas e últimas colunas e próximo ao centro do gráfico, mas não maiores que o valor dado pelo limite superior.

A dispersão nas linhas e colunas das medidas de metano (Figuras 5.17b e 5.17c) concordou com a análise gráfica da superfície de valores da Figura 5.17a, onde os valores maiores próximos ao limite superior são encontrados nas últimas linhas das ordenadas e últimas colunas das abscissas e próximas ao centro do gráfico, confirmando áreas de maior concentração do gás metano.

Apesar de a análise estatística descritiva apontar uma distribuição normal para os dados transformados (\ln) e segundo Isaaks e Srivastava (1989), mais importante que a normalidade dos dados é a ocorrência ou não do chamado efeito proporcional, em que a média e a variabilidade dos dados sejam constantes no estudo. No entanto, em se tratando de dados obtidos na natureza, o ajuste de uma distribuição teórica é apenas aproximado (WARRICK e NIELSEN, 1980). A não existência do efeito proporcional identifica a estacionaridade dos dados amostrais. Segundo Gonçalves *et al.* (2001), se a variação dos dados em torno da média é proporcional à magnitude desta, a estacionaridade mínima necessária ao uso da geoestatística pode estar comprometida.

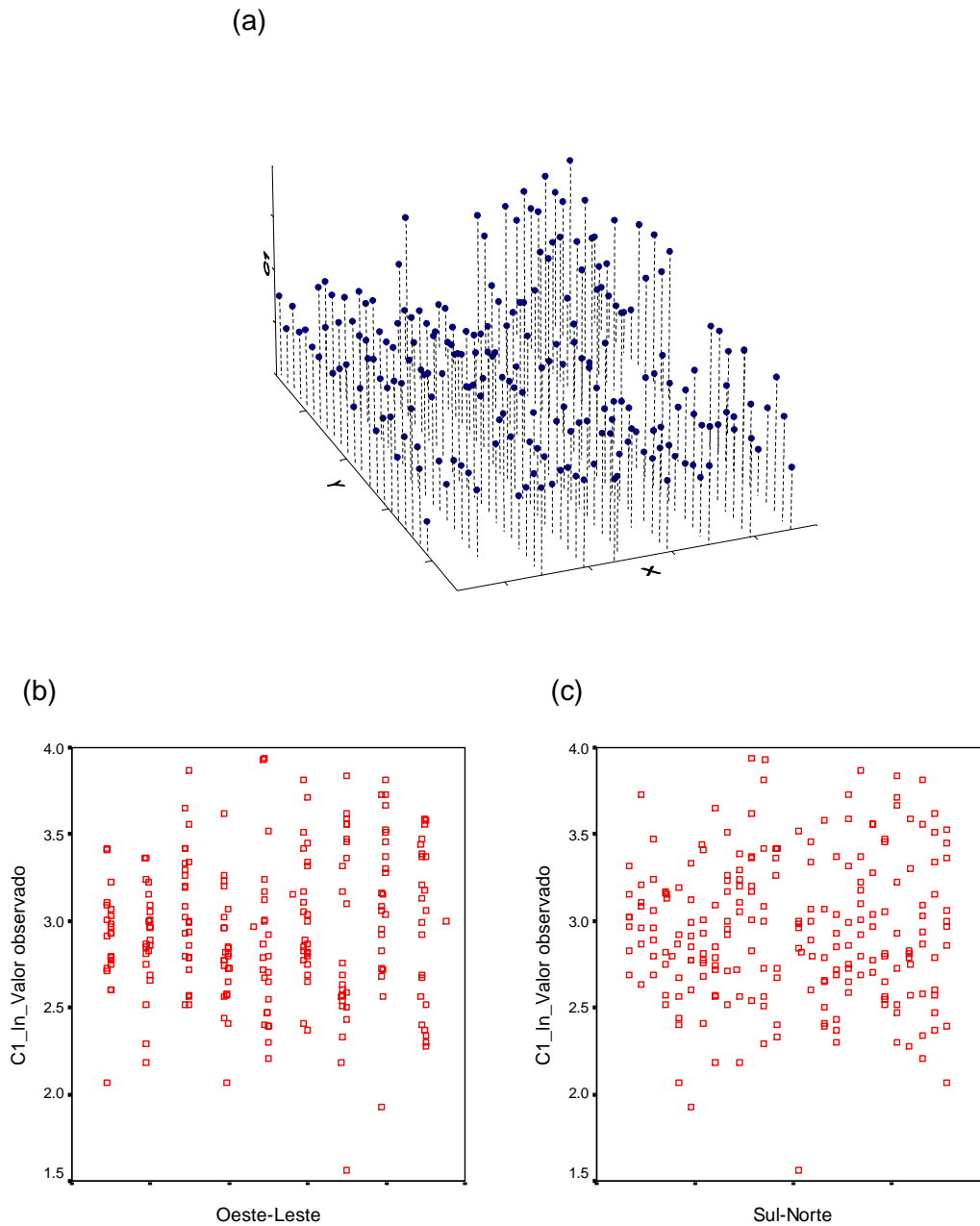
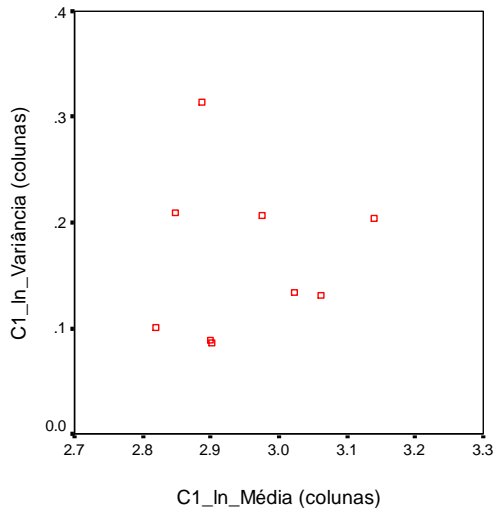


Figura 5.17 – Superfície de valores (a) e dispersão das medidas do gás metano (ln) amostradas nos sentidos Oeste-Leste (b) e Sul-Norte (c).

Assim, a existência do efeito proporcional é investigada por meio da Figura 5.18, em que as medidas de posição menores corresponderam, ainda que em poucos pontos, as medidas de dispersão mais elevadas no sentido da largura do bloco (coluna), enquanto no sentido do comprimento (linha) as medidas de posição médias corresponderam, ainda que em poucos pontos, as medidas de dispersão mais elevadas. A Figura 5.18 não apresentou efeito proporcional nítido, podendo assumir

um comportamento estabilizado em torno de uma constante com alguns pontos mais afastados.

(a)



(b)

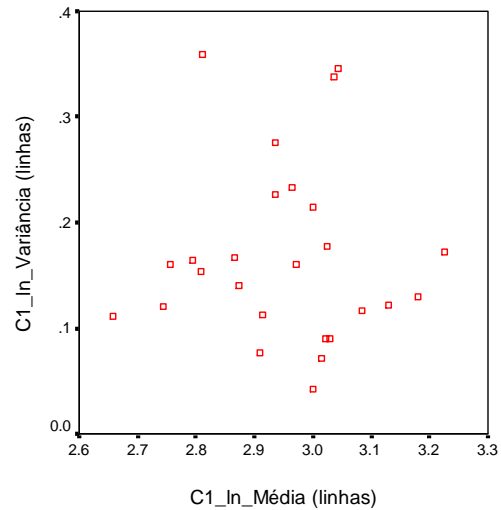


Figura 5.18 – Gráficos de dispersão das médias e variâncias dos valores de metano (In) amostradas nos sentidos das colunas (a) e linhas (b).

O “*postplot*” dos valores de metano (Figura 5.19), exceto pelas concentrações que ocorrem à direita superior do gráfico e próximo ao centro, o que já havia sido comentada nas figuras anteriores, pareceu estar de acordo com a hipótese intrínseca não apresentando tendências. A existência de sub-regiões (símbolos iguais de forma concentrada) e/ou tendências (variações gradativas em alguma direção) são situações incompatíveis com a hipótese intrínseca assumida na análise geoestatística (RIBEIRO JUNIOR, 1995).

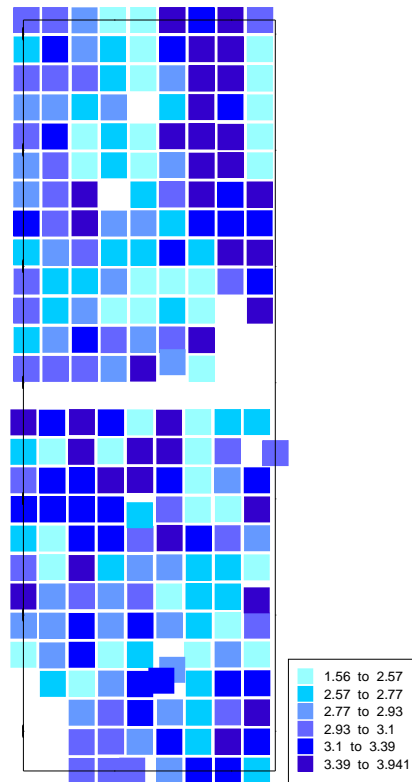


Figura 5.19 – Análise descritiva espacial pelo “*Postplot*” da variável transformada metano.

- **Variável (C_2C_5)**

Os valores transformados (ln) da variável se ajustaram próximo a uma distribuição normal, adequando-se melhor ao tratamento geoestatístico. Observando a superfície dos valores (Figura 5.20a) da variável, verificam-se aglomerações das medidas para valores próximos ao valor dado pelo limite superior (3,13) em três regiões distintas, uma localizada à direita superior do gráfico, a outra próxima ao centro e a última localizada à direita inferior do gráfico.

Este fato também pode estar associado a processos de migração por difusão, podendo esse gradiente de concentração estar acontecendo em subsuperfície logo abaixo desta, ou ao redor de fraturas e falhas, levando a valores maiores de C_2C_5 . No gráfico de dispersão das medidas em relação às linhas (Figura 5.20b) e colunas (Figura 5.20c), notam-se pequenas concentrações de valores no início, meio e fim das linhas e nas redondezas da terceira e últimas colunas, mas não maiores que o valor dado pelo limite superior.

A dispersão nas linhas e colunas das medidas de C_2C_5 concordou com a análise gráfica da superfície de valores.

Como a análise estatística descritiva da variável $\ln C_2C_5$ aponta uma proximidade com distribuição normal para os dados transformados (\ln), sendo assim, a existência do efeito proporcional é investigada por meio da Figura 5.21, as medidas de posição menores corresponderam, ainda que em poucos pontos, as medidas de dispersão mais elevadas no sentido da largura do bloco (coluna). Enquanto que no sentido do comprimento (linha) as medidas de posição baixa e média corresponderam, ainda que em poucos pontos, as medidas de dispersão mais elevadas.

No geral verifica-se certa estabilização da variância em relação ao leve deslocamento da média. A Figura 5.21, apesar de aparentar uma leve tendência no sentido das colunas, não apresenta efeito proporcional nítido, podendo assumir um comportamento estabilizado em torno de uma constante com alguns pontos mais afastados.

O “*postplot*” dos valores de C_2C_5 (\ln), exceto pela existência de pequenas sub-regiões que ocorrem à direita superior e inferior do gráfico e próximo ao centro, a qual já havia sido comentada nas figuras anteriores, pareceu estar de acordo com a hipótese intrínseca não apresentando tendências (Figura 5.22).

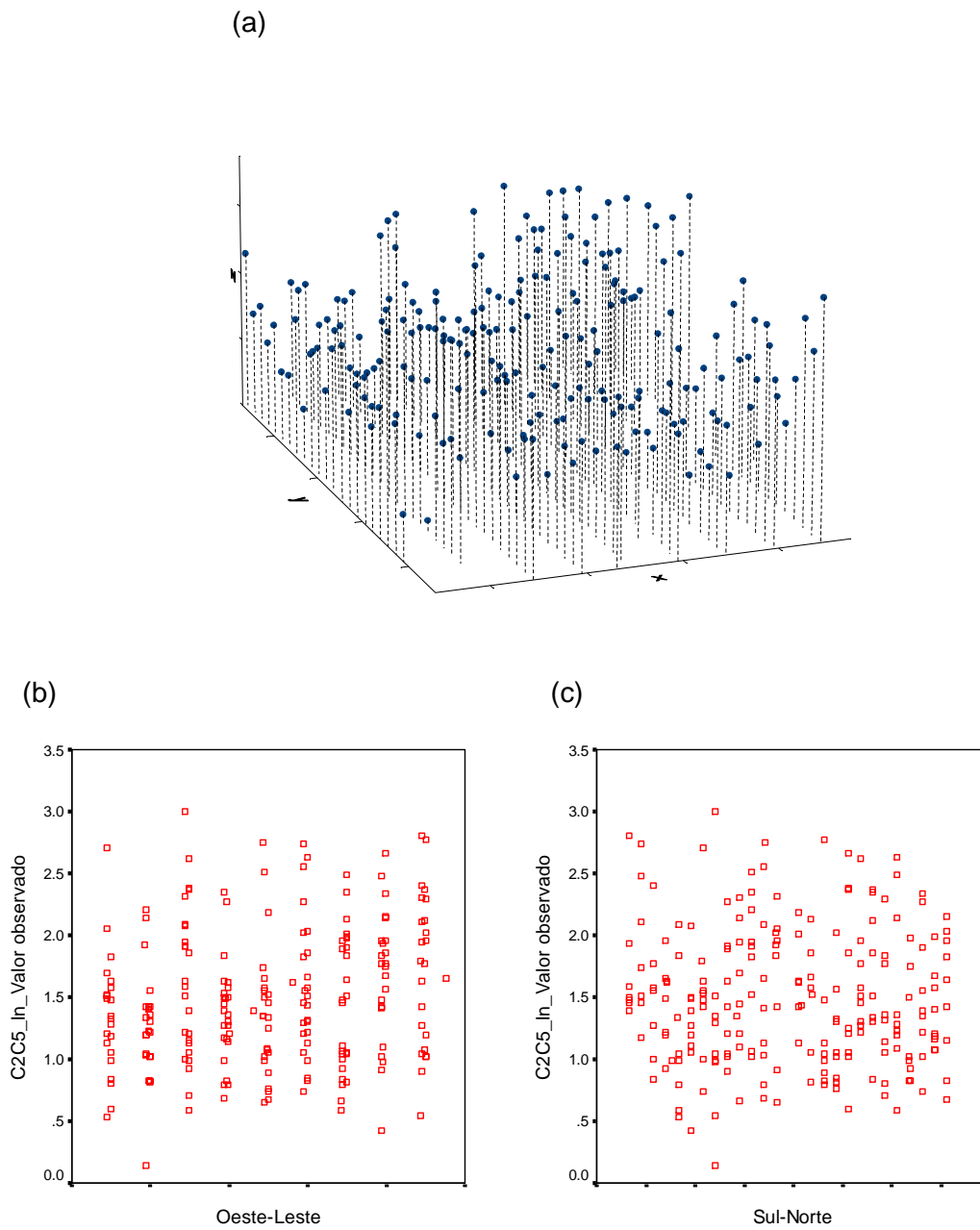


Figura 5.20 – Superfície de valores (a) e dispersão das medidas de C_2C_5 (ln) amostradas nos sentidos Oeste-Leste (b) e Sul-Norte (c).

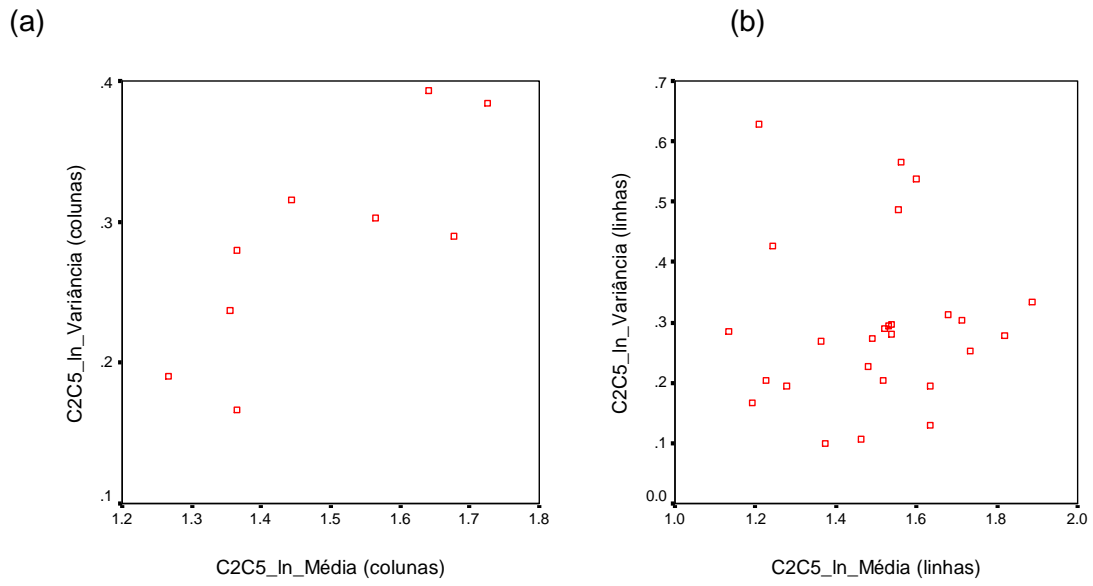


Figura 5.21 – Gráficos de dispersão das médias e variâncias dos valores de C_2C_5 (In) amostradas nos sentidos das colunas (a) e linhas (b).

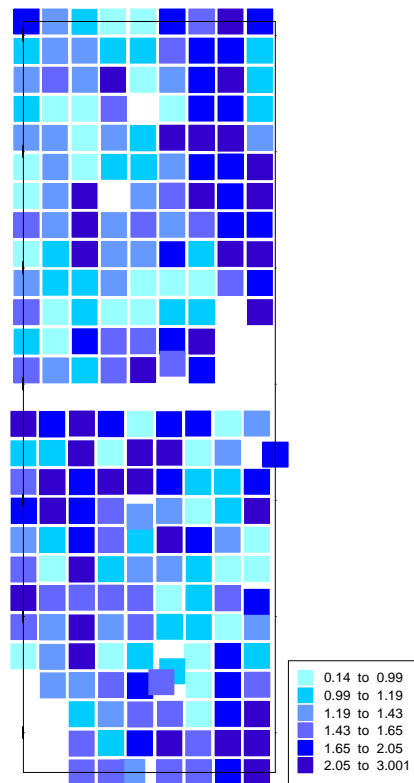


Figura 5.22 – Análise descritiva espacial pelo “*Postplot*” da variável InC_2C_5 .

- **Variável eteno (C_2^-)**

Observando a superfície dos valores (Figura 5.23a) da variável eteno, verificam-se aglomerações das medidas para valores próximos ao valor dado pelo limite superior (3,13) em três regiões distintas, uma localizada à direita superior do gráfico, a outra próxima ao centro e a última localizada à direita inferior do gráfico.

Este fato pode estar associado a processos biogênicos acontecendo em subsuperfície logo abaixo desta, levando o aparecimento de anomalias de hidrocarbonetos insaturados. No gráfico de dispersão das medidas em relação às linhas (Figura 5.23b) e colunas (Figura 5.23c), notam-se valores levemente maiores que o valor dado pelo limite superior (19,33) em praticamente todas as colunas, valores mais altos em pequena concentração nas últimas colunas, e se estendendo do centro até as últimas linhas. A dispersão nas linhas e colunas das medidas de C_2^- concordou com a análise gráfica da superfície de valores.

Como os dados não seguem nem uma distribuição normal ou lognormal, a existência do efeito proporcional é investigada por meio da Figura 5.24. As medidas de posição menores corresponderam, ainda que em poucos pontos, as medidas de dispersão mais elevadas no sentido da largura do bloco (coluna), enquanto no sentido do comprimento (linha) as medidas de posição médias corresponderam, ainda que em poucos pontos, as medidas de dispersão mais elevadas. A Figura 5.24, não apresenta efeito proporcional nítido, podendo assumir um comportamento estabilizado em torno de uma constante com alguns pontos mais afastados.

O “*postplot*” dos valores da variável eteno (C_2^-), exceto pela existência de pequenas aglomerações que ocorrem à direita superior do gráfico e próximo ao centro, a qual já havia sido comentada nas figuras anteriores, pareceu estar de acordo com a hipótese intrínseca não apresentando tendências (Figura 5.25).

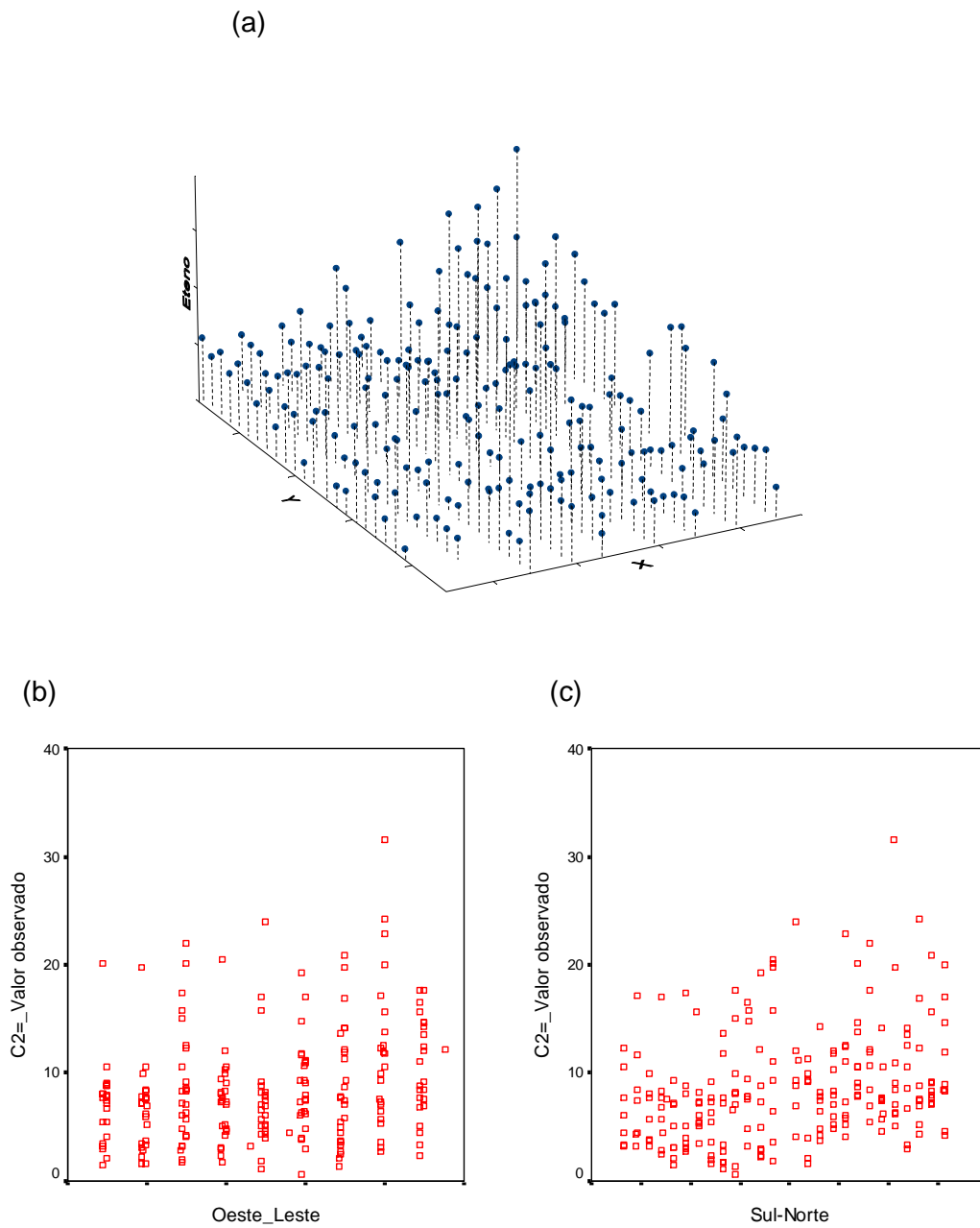


Figura 5.23 – Superfície de valores (a) e dispersão das medidas da variável eteno ($C_2=$) amostradas nos sentidos Oeste-Leste (b) e Sul-Norte (c).

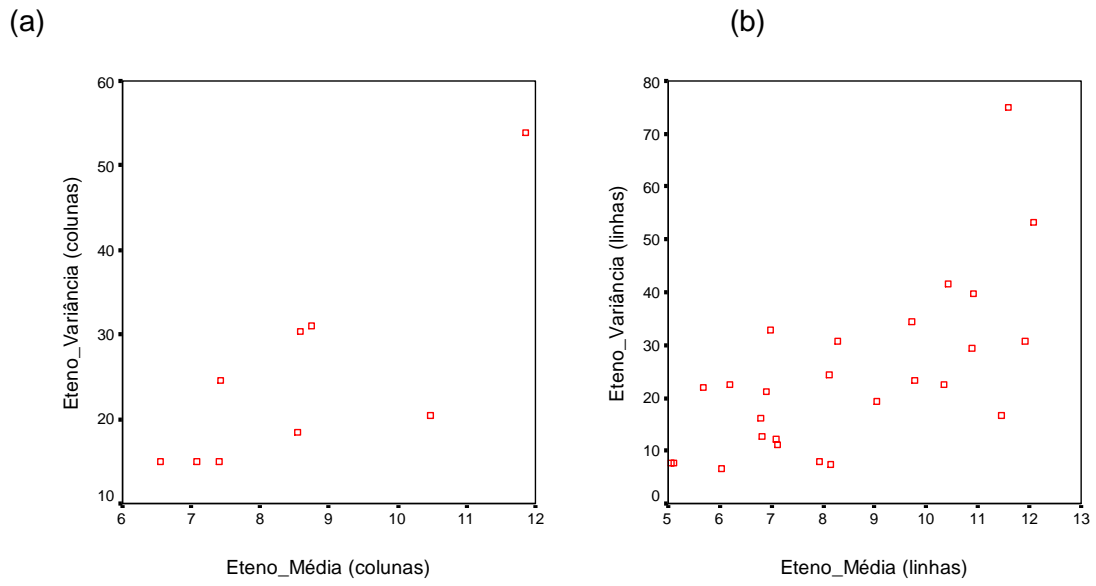


Figura 5.24 – Gráficos de dispersão das médias e variâncias dos valores de eteno (C_2^-) amostradas nos sentidos das colunas (a) e linhas (b).

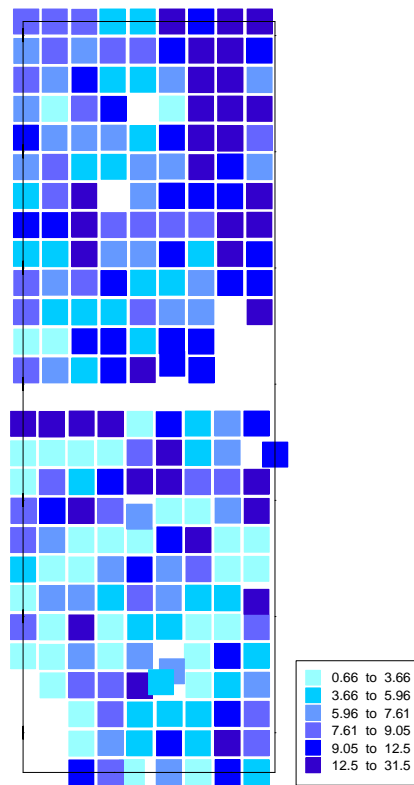


Figura 5.25 – Análise descritiva espacial pelo “*Postplot*” da Variável Eteno (C_2^-).

5.4 Análise Geoestatística

5.4.1 Análise semivariográfica

Após o cálculo do semivariograma experimental, pode-se tentar um ajuste sobre a nuvem de pontos, obtendo-se um semivariograma modelado deve-se observar que cada ponto do semivariograma representa a medida de dessemelhança entre os vizinhos próximos.

Visto que variáveis de hidrocarbonetos presentes no solo podem apresentar um comportamento anisotrópico ou isotrópico, foram efetuados semivariogramas direcionais (anisotropia) e omnidirecional (isotropia), de modo a perceber como varia a distância entre os pares em função do $V(x)$, segundo diferentes direções.

- **Metano (C_1)**

Para os dados transformados de metano ($\ln C_1$) o semivariograma experimental que melhor se ajustou foi na direção 0° das ordenadas (Figura 5.26) com tolerância angular de $22,5^\circ$.

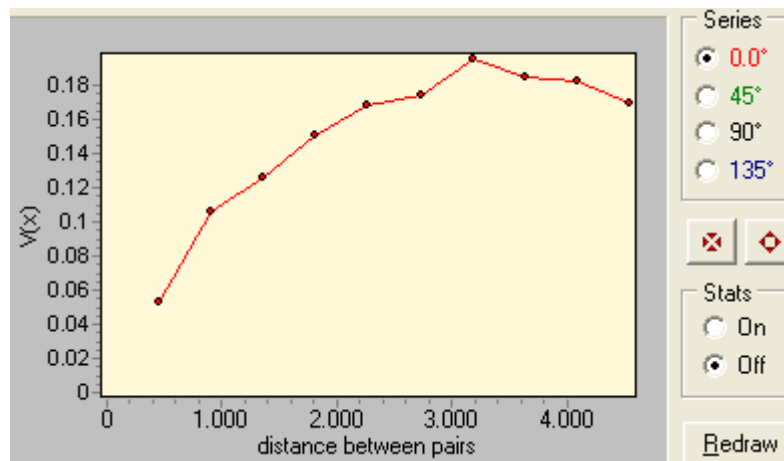


Figura 5.26 - Semivariograma experimental na direção 0° das ordenadas.

Foi efetuado um ajustamento em relação ao efeito pepita, patamar e alcance do semivariograma experimental pelo método dos mínimos quadrados ponderados e o modelo matemático teórico escolhido foi o esférico (Figura 5.27), o qual descreveu melhor o padrão da variabilidade espacial indicando que os dados não apresentam, como esperado, uma tendência, pois possuem um patamar (LANDIM *et al.*, 2002).

Os parâmetros do modelo foram efeito pepita ($C_0 = 0,07865$), patamar ($C + C_0 = 0,1798$) e alcance (a) de 3025 metros. Ou seja, existe uma estrutura de dependência espacial entre as amostras até 3025 metros na direção de 0° das ordenadas. Desta forma, as estimativas devem ser realizadas considerando-se o modelo de semivariograma.

Segundo Trangmar *et al.* (1995), o alcance estabelece o limite de dependência espacial entre amostras, isto é, para distâncias iguais ou menores que o alcance, diz-se que os valores vizinhos de uma variável estão espacialmente correlacionados e podem ser utilizados na estimativa de valores em qualquer ponto entre eles. Portanto, a estatística clássica pode ser aplicada a partir de amostras coletadas acima desta distância, ou seja, o domínio da geoestatística é de 3025 m. Outros modelos ajustados com seus respectivos parâmetros podem ser observados na Tabela 5.8.

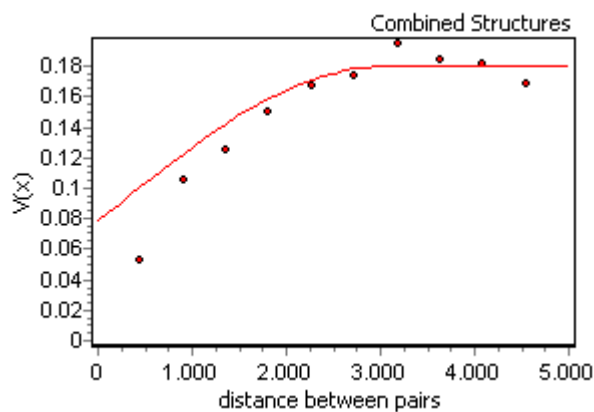


Figura 5.27 – Semivariograma para a variável $\ln C_1$, ajustado ao modelo esférico com os seguintes parâmetros: $C_0 = 0,0786$; $a = 3025$ m e $C = 0,1012$.

Tabela 5.8 - Parâmetros e modelos dos semivariogramas para a variável metano (ln).

Variável	Modelo	C_0	C	$C + C_0$	a
Metano	Esférico	0,07865	0,10124	0,17989	3025
	Exponencial	0,03131	0,14881	0,18012	2529
	Gaussiano	0,08867	0,09032	0,17899	2147

O menor valor do efeito pepita, $C_0 = 0,0313$, foi observado para o modelo exponencial, ou seja, foi o modelo que melhor explicou a variabilidade espacial. O modelo que descreveu a distância de maior dependência entre as amostras foi o esférico com alcance de 3025 metros.

- C_2C_5

Foi utilizada na análise semivariográfica a variável transformada de C_2C_5 por apresentar uma distribuição próxima da normal. Para os dados transformados de C_2C_5 ($\ln C_2C_5$), o semivariograma experimental que melhor se ajustou foi o omnidirecional (isotropia) (Figura 5.28).

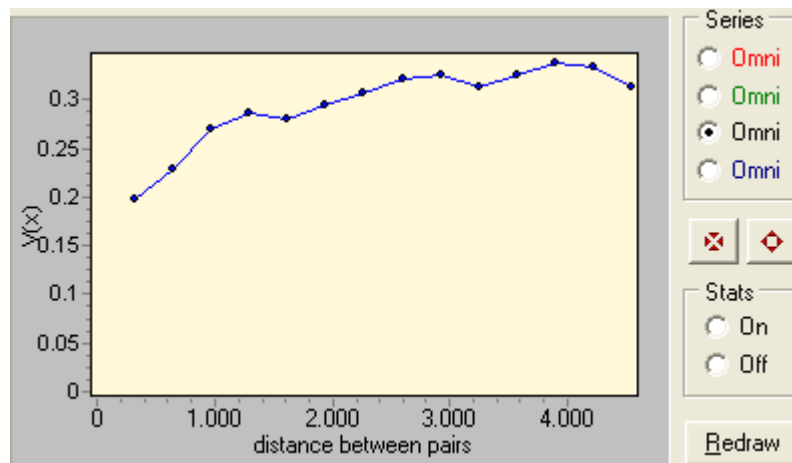


Figura 5.28 - Semivariograma experimental omnidirecional para a variável $\ln C_2C_5$.

Foi efetuado um ajustamento em relação ao efeito pepita, patamar e alcance do semivariograma experimental pelo método dos mínimos quadrados ponderados e o modelo matemático teórico escolhido foi o exponencial (Figura 5.29), o qual descreveu melhor o padrão da variabilidade espacial.

Os parâmetros do modelo foram efeito pepita ($C_0 = 0,1728$), patamar ($C + C_0 = 0,3247$) e alcance (a) de 2802 metros. Isto é, existe uma estrutura de dependência espacial entre as amostras até 2802 metros em qualquer direção. Desta forma, as estimativas devem ser realizadas considerando-se o modelo de semivariograma. A estatística clássica pode ser aplicada a partir de amostras coletadas acima desta distância, ou seja, o domínio da geoestatística é de 2802 m. Outros modelos ajustados com seus respectivos parâmetros podem ser observados na Tabela 5.9.

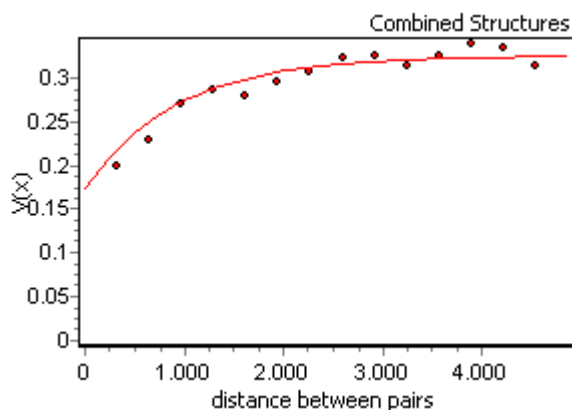


Figura 5.29 – Semivariograma para a variável $\ln C_2 C_5$ ajustado ao modelo exponencial com os seguintes parâmetros: $C_0 = 0,1728$; $a = 2802$ m, $C = 0,1520$ e $C + C_0 = 0,3247$.

Tabela 5.9 - Parâmetros e modelos dos semivariogramas para a variável $\ln C_2 C_5$.

Variável	Modelo	C_0	C	$C + C_0$	a
Metano	Esférico	0,2151	0,1066	0,3218	2924
	Exponencial	0,1728	0,1520	0,3247	2802
	Gaussiano	0,2320	0,0881	0,03248	2355

O menor valor do efeito pepita, $C_0 = 0,1728$, foi observado para o modelo exponencial, ou seja, foi o modelo que melhor explicou a variabilidade espacial. O modelo que descreveu a distância de maior dependência entre as amostras foi o esférico com alcance de 2924 metros.

- **Eteno (C_2^-)**

Para os dados da variável eteno, o semivariograma experimental que melhor se ajustou foi na direção 0° das ordenadas (Figura 5.30) com tolerância angular de $22,5^\circ$.

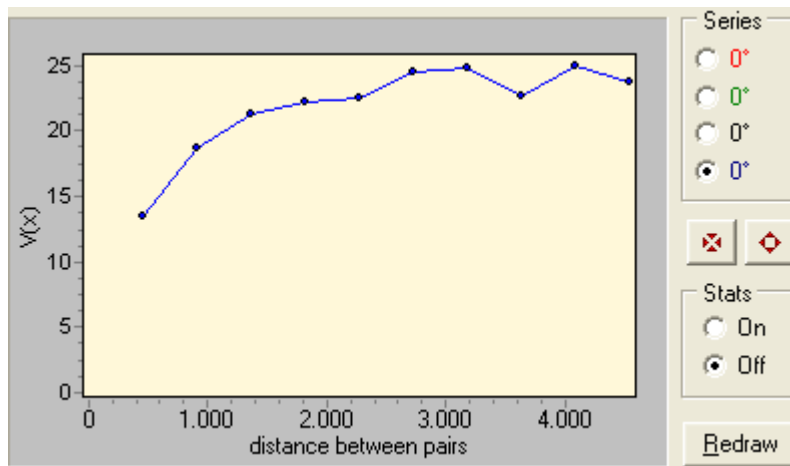


Figura 5.30 - Semivariograma experimental na direção 0° das ordenadas.

Foi efetuado um ajustamento em relação ao efeito pepita, patamar e alcance do semivariograma experimental pelo método dos mínimos quadrados ponderados e o modelo matemático teórico escolhido foi o esférico (Figura 5.31), o qual descreveu melhor o padrão da variabilidade espacial.

Os parâmetros do modelo foram efeito pepita ($C_0 = 16,9218$), patamar ($C + C_0 = 24,2755$) e alcance (a) de 2981 metros. Existe uma estrutura de dependência espacial entre as amostras até 2981 metros na direção de 0° das ordenadas. Desta forma, as estimativas devem ser realizadas considerando-se o modelo de semivariograma. A estatística clássica pode ser aplicada a partir de amostras coletadas acima desta distância, ou seja, o domínio da geoestatística é de 2981 m. Outros modelos ajustados com seus respectivos parâmetros podem ser observados na Tabela 5.10.

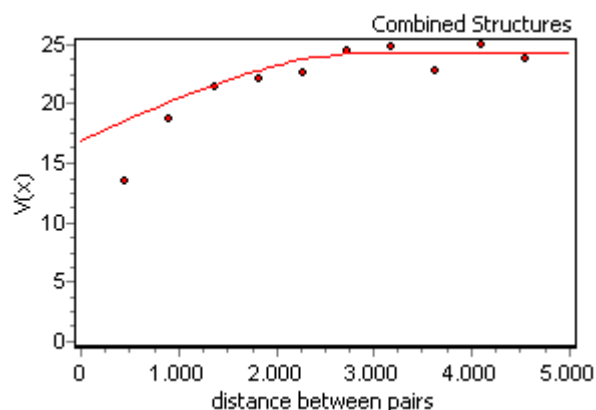


Figura 5.31 – Semivariograma para a variável eteno, ajustado ao modelo esférico com os seguintes parâmetros: $C_0 = 16,9218$; $a = 2981$ m e $C = 7,3537$.

Tabela 5.10 - Parâmetros e modelos dos semivariogramas para a variável eteno.

Variável	Modelo	C_0	C	$C + C_0$	a
Metano	Esférico	16,9218	7,3537	24,2755	2981
	Exponencial	14,7491	9,7737	24,5228	2962
	Gaussiano	18,5976	5,6782	24,2758	2750

O menor valor do efeito pepita, $C_0=14,7491$, foi observado para o modelo exponencial, isto é, foi o modelo que melhor explicou a variabilidade espacial. O modelo que descreveu a distância de maior dependência entre as amostras foi o esférico com alcance de 2981 metros.

5.4.2 Interpolação por krigagem ordinária, curvatura mínima e inverso do quadrado da distância.

A Krigagem para cada variável em estudo considerou o modelo ajustado para o semivariograma, já os métodos da curvatura mínima e inverso do quadrado da distância foram ajustados conforme o padrão do software Surfer 8.01.

Com o objetivo de comparar graficamente os interpoladores, os mapas de contorno e superfície destas interpolações são mostrados. Foi utilizado o método [mediana + 2 dma] para a determinação de limiares, onde a cor verde corresponde aos valores de *background* e as cores de laranja a vermelha aos valores anômalos. Nos mapas de contorno, os pontos marcados com o sinal “+” correspondem aos valores originais da amostra, quanto maior o tamanho do sinal, maior é o valor da amostra.

- **Metano (C_1), C_2C_5 e Eteno ($C_2^=$)**

Os mapas de contorno e superfície dos valores de C_1 (Figuras 5.32a e 5.33a), C_2C_5 (Figuras 5.34a e 5.35a) e $C_2^=$ (Figuras 5.36a e 5.37a), estimados pelo método de krigagem, apresentam as mesmas sub-regiões descritas nas análises da superfície de valores. Entretanto, foi apresentado um efeito suavizador das diferenças, visto que o interpolador de krigagem superestima para baixas medidas e subestima para elevados valores, ou seja, tende à média dos dados.

Esse resultado é obtido principalmente pelo fato desse interpolador ser não-viciado, com variância mínima, pois é um interpolador ótimo. A região dita anômala aparece bem delineada.

A interpolação pelo método de curvatura mínima para valores de C_1 (Figuras 5.32b e 5.33b), C_2C_5 (Figuras 5.34b e 5.35b) e C_2^- (Figuras 5.36b e 5.37b) apresentou outras sub-regiões não descritas na análise da superfície de valores como podemos observar, apresentando uma distribuição espacial menos homogênea que a apresentada pelo método de Krigagem. Podemos verificar também uma extrapolação de valores, principalmente nas bordas do bloco.

Já a interpolação pelo método do inverso do quadrado da distância (IQD), os mapas de contorno e superfície para os valores de C_1 (Figuras 5.32c e 5.33c), C_2C_5 (Figuras 5.34c e 5.35c) e C_2^- (Figuras 5.36c e 5.37c) apresentaram uma distribuição espacial mais homogênea do que pelo método de curvatura mínima.

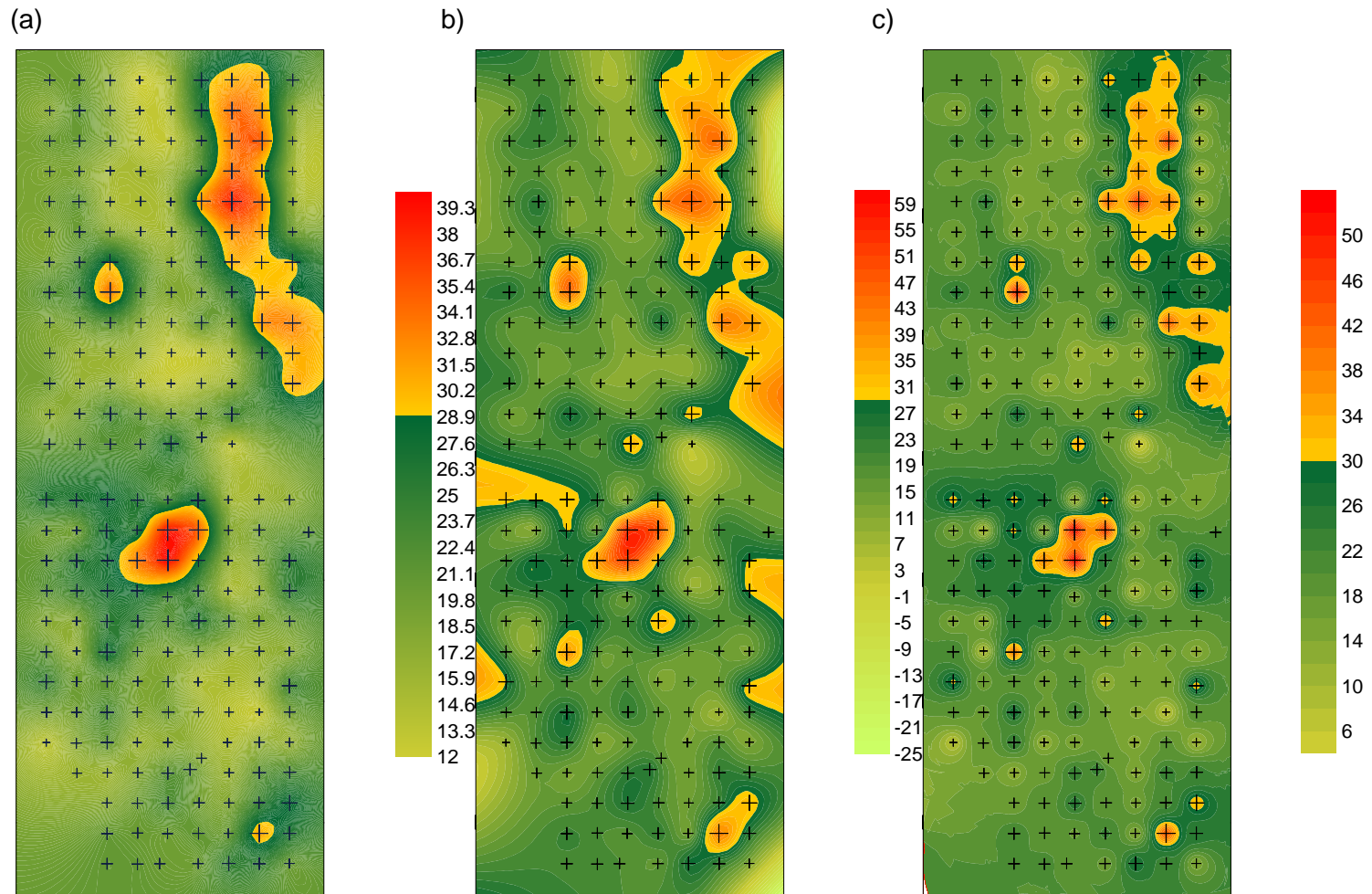


Figura 5.32 - Mapas de contorno interpolado para o metano (ppm) pelo método de Krigagem ordinária considerando o modelo esférico (a), pelo método de curvatura mínima (b) e inverso do quadrado da distância (c).

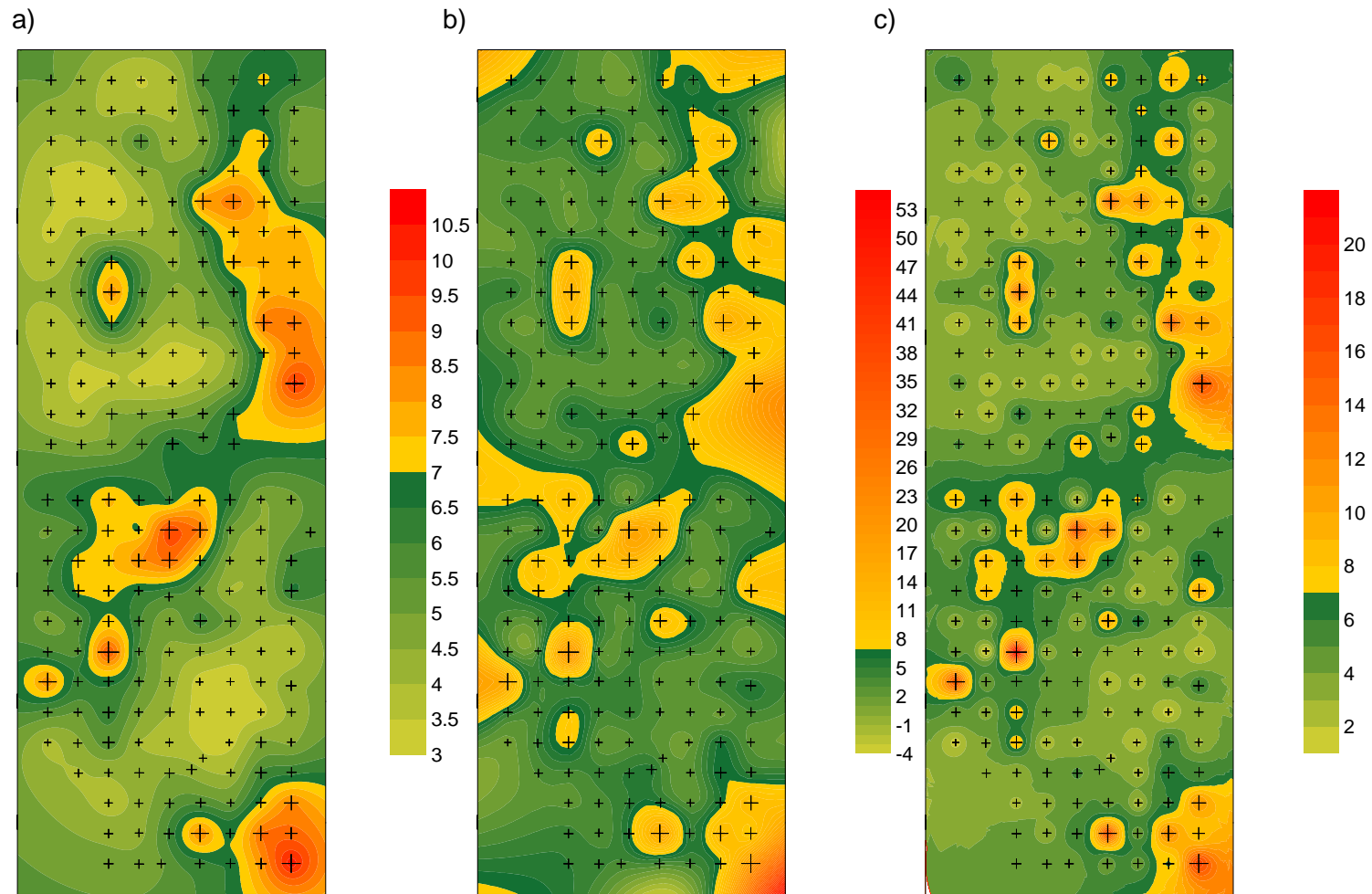


Figura 5.34 - Mapas de contornos interpolados para o C_2C_5 (ppm) pelo método de Krigagem ordinária (a), método de curvatura mínima (b) e inverso do quadrado da distância (c).

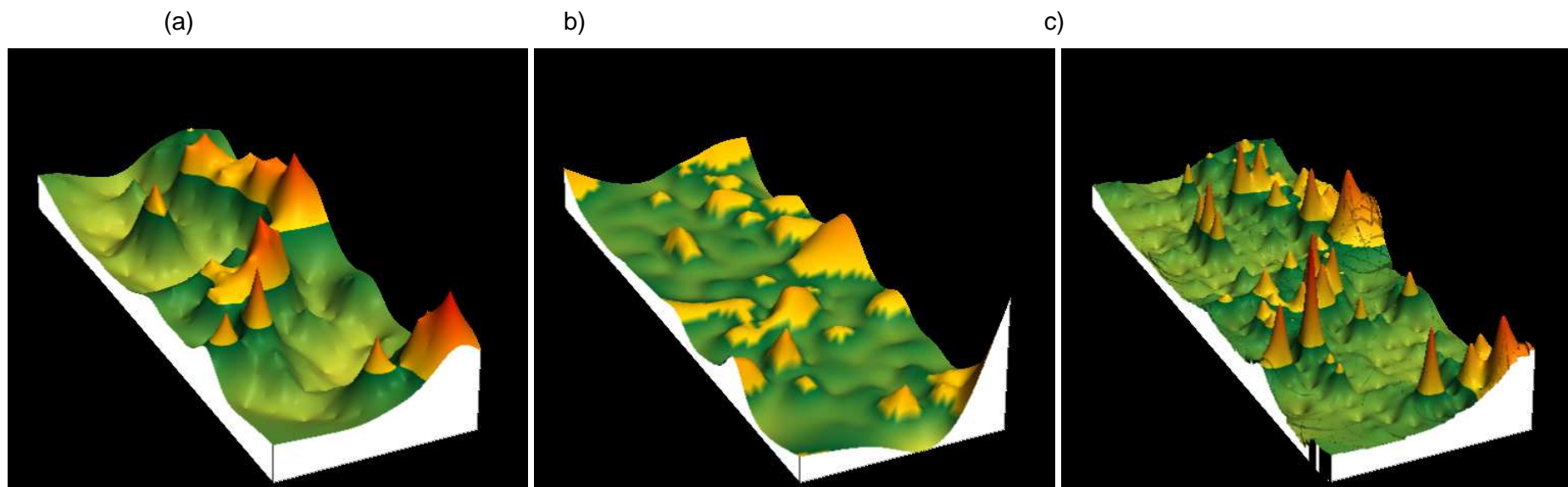


Figura 5.35 - Mapas de superfície interpolada para o C_2C_5 (ppm) pelo método de Krigagem ordinária (a), método de curvatura mínima (b) e inverso do quadrado da distância (c).

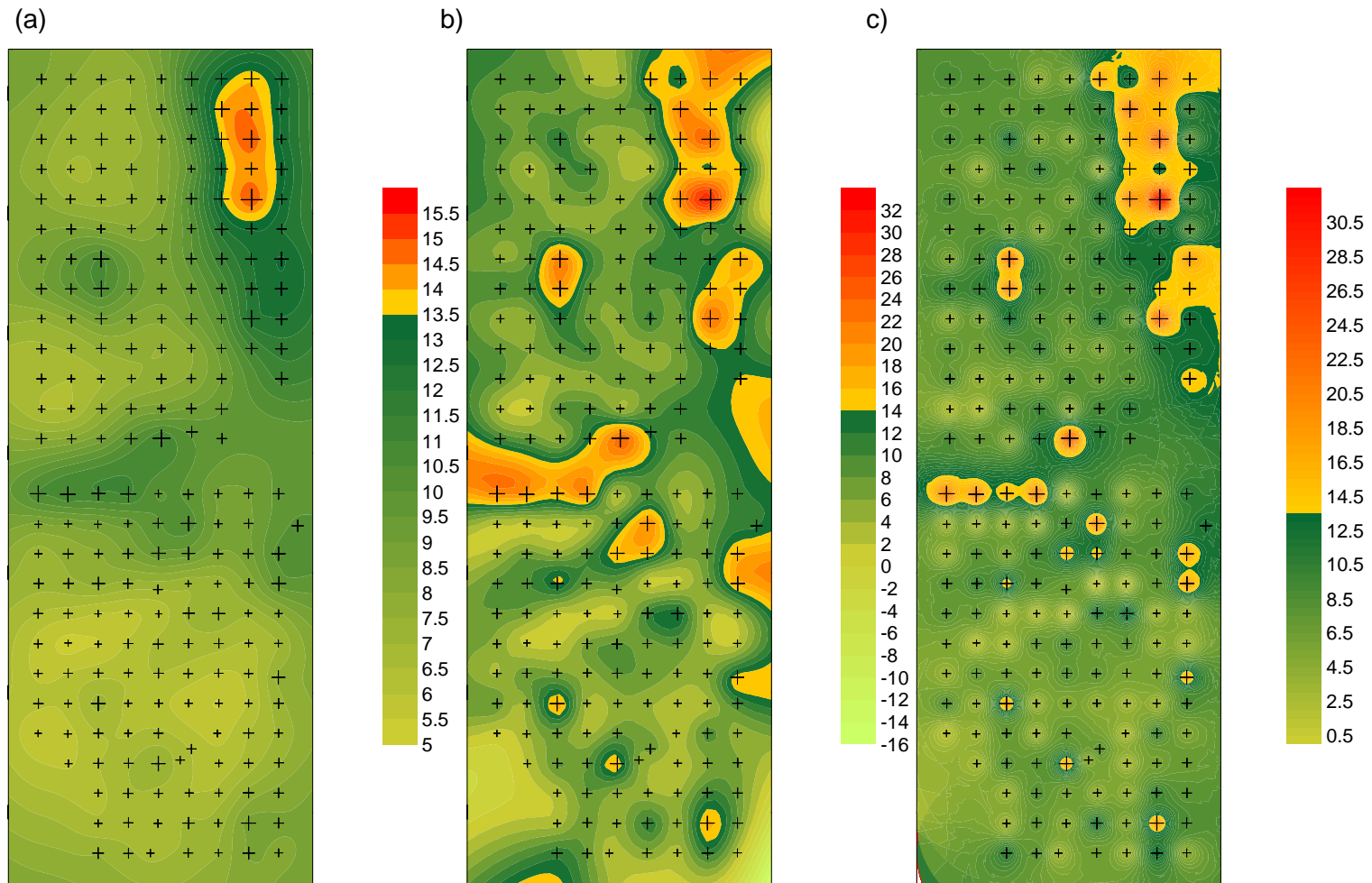


Figura 5.36 - Mapas de contornos interpolados para eteno (C_2) (ppm) pelo método de Krigagem ordinária (a), método de curvatura mínima (b) e inverso do quadrado da distância (c).

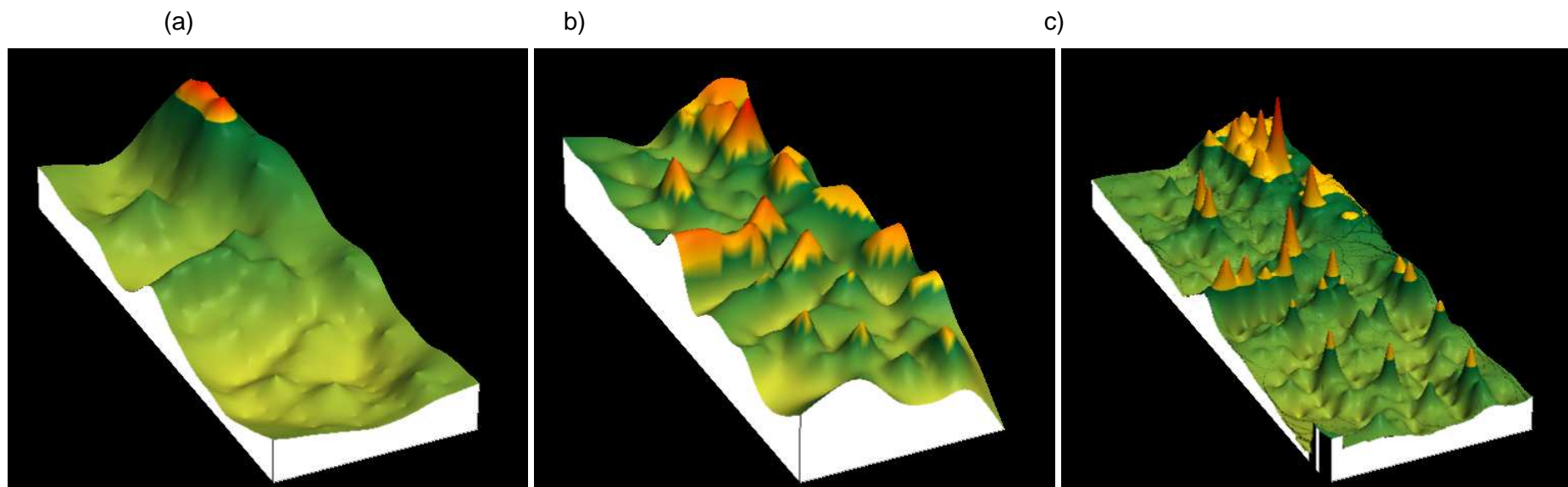


Figura 5.37 - Mapas de superfície interpolada para eteno ($C_2^=$) (ppm) pelo método de Krigagem ordinária (a), método de curvatura mínima (b) e inverso do quadrado da distância (c).

5.4.3 Validação cruzada

Na análise da Krigagem, um modelo com ajuste adequado foi aquele que apresentou, na validação cruzada, seus erros padronizados com distribuição próxima da normal padrão.

- **Metano (C₁)**

Na validação cruzada da variável metano, três modelos foram testados e suas estatísticas estão na Tabela 5.11. Observou-se que o modelo mais adequado para descrever o fenômeno de dependência espacial da variável metano foi o modelo esférico, apesar de o modelo exponencial ter apresentado menor valor do efeito pepita, $C_0 = 0,0313$.

Tabela 5.11 – Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável metano.

	Modelo	Esférico	Exponencial	Gaussiano
Média (ppm)		0,0040	0,0044	0,0042
Variância (ppm) ²		1,00	1,18	1,20

A figura 5.38 apresenta dispersão dos erros de Krigagem com aparente uniformidade e sem tendência nos valores calculados. No mapa dos erros de krigagem os pontos marcados com o sinal “+” correspondem aos valores superestimados e “x” aos valores subestimados, quanto maior o tamanho do sinal, maior é a descontinuidade espacial dos dados.

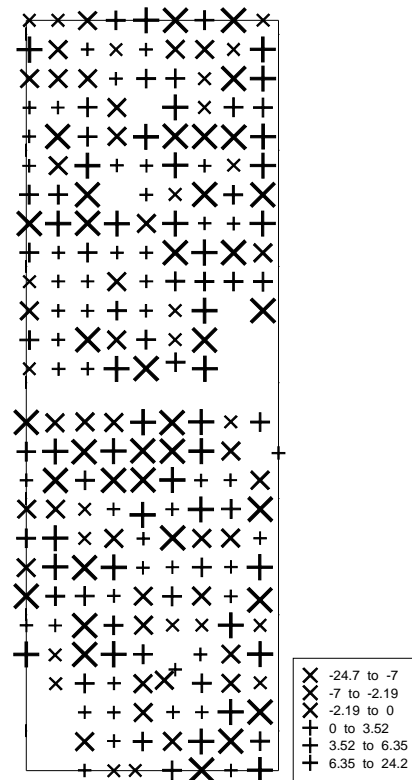


Figura 5.38 - Mapa de erros dos valores estimados em relação aos observados da variável metano ajustado ao modelo esférico.

Podemos verificar que os valores subestimados estão correlacionados com os valores mais altos dos dados observados da variável, bem como os valores superestimados estão correlacionados com os valores mais baixos. O desvio padrão e média dos erros padronizados da validação cruzada para o modelo esférico foram 1,00 e 0,004, respectivamente.

A dispersão dos erros positivos e negativos pela krigagem foi semelhante (Figura 5.39), observando maior afastamento da reta 1:1 para os erros negativos (quando o valor da diferença entre o valor estimado e original é menor que zero) nas maiores medidas de metano. Quanto aos erros positivos, as superestimativas da variável foram observadas nos menores valores de metano e os demais se equilibraram entre os valores superestimados e os subestimados, fato que concorda com a característica de não viés do estimador de Krigagem.

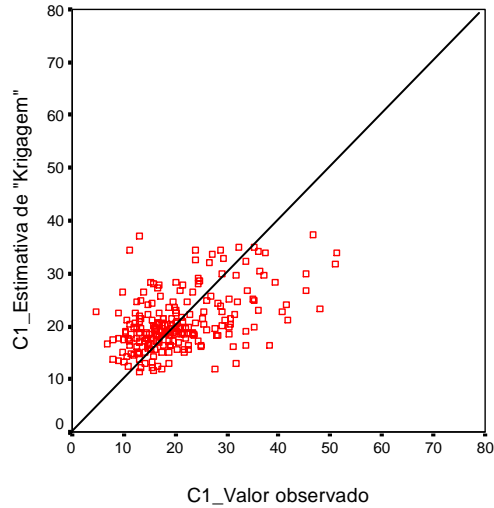


Figura 5.39 – Gráficos de dispersão das estimativas de krigagem em relação aos valores observados na validação cruzada para a variável metano no modelo esférico.

Na Figura 5.40 podemos observar os histogramas de frequência dos erros (resíduos) e dos erros padronizados da variável metano, que apresentou dados conforme Tabela 5.11. As diferenças padronizadas devem ter, em uma situação ideal, média nula, variância unitária e ser normalmente distribuídas. As estatísticas assinalaram valores próximos do ideal (Figura 5.40b). A padronização foi realizada a partir do desvio padrão de krigagem obtido para cada um dos pontos da malha.

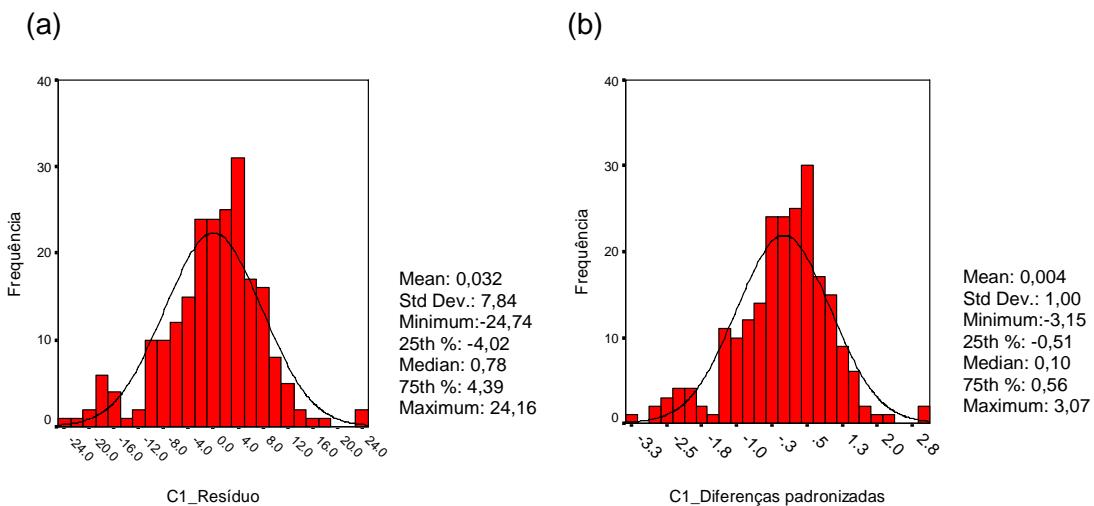


Figura 5.40 – Histograma de frequência dos erros (a) e dos erros padronizados da variável metano (b), ajustado ao modelo esférico.

A distribuição dos erros não se afastou da normal padrão indicando um ajuste muito bom do modelo esférico, ou seja, o modelo ajustado descreve adequadamente a dependência espacial dos dados de metano.

- C_2C_5

Três modelos foram testados na validação cruzada e suas estatísticas estão na Tabela 5.12. Observou-se que o modelo mais adequado para descrever o fenômeno de dependência espacial da variável foi o modelo exponencial, sendo também o modelo que apresentou menor valor do efeito pepita, $C_0 = 0,17283$.

A dispersão dos erros de krigagem (Figura 5.41) apresenta com aparente uniformidade e sem tendência nos valores calculados. Podemos verificar que os valores subestimados estão correlacionados com os valores mais altos dos dados observados, bem como os valores superestimados estão correlacionados com os valores mais baixos.

Tabela 5.12 – Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável C_2C_5 .

Modelo	Esférico	Exponencial	Gaussiano
Média (ppm)	-0,004	-0,003	-0,004
Variância (ppm) ²	0,9351	0,9974	0,9555

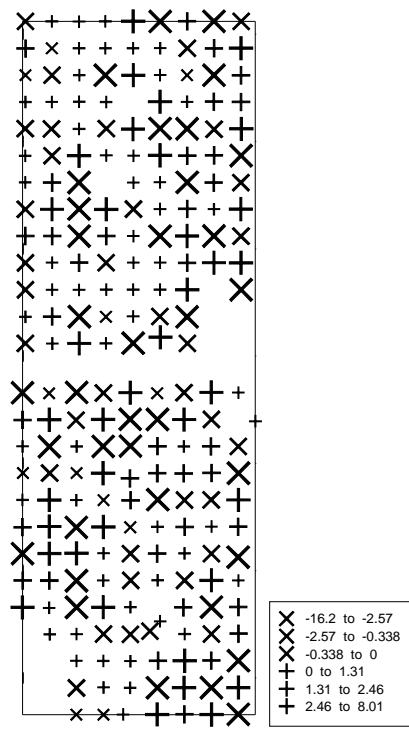


Figura 5.41 - Mapa de erros dos valores estimados em relação aos observados da variável C_2C_5 ajustado ao modelo exponencial.

A dispersão dos erros positivos e negativos pela krigagem é mostrada na Figura 5.42. Observa-se um maior afastamento da reta 1:1 para os erros negativos (quando o valor da diferença entre o valor estimado e original é menor que zero) nas maiores medidas de C_2C_5 . Quanto aos erros positivos, as superestimativas da variável foram observadas nos menores valores de metano e os demais se equilibraram entre os valores superestimados e os subestimados, fato que concorda com a característica de não viés do estimador de Krigagem.

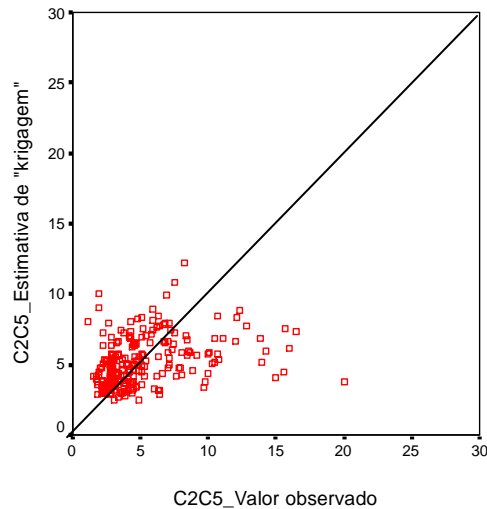


Figura 5.42 – Gráficos de dispersão das estimativas de krigagem em relação aos valores observados na validação cruzada para a variável C_2C_5 no modelo exponencial.

Na Figura 5.43 podemos observar os histogramas de frequência dos erros (resíduos) e dos erros padronizados da variável C_2C_5 . As estatísticas assinalaram valores próximos do ideal (Figura 5.43b). A padronização foi realizada a partir do desvio padrão de krigagem obtido para cada um dos pontos da malha.

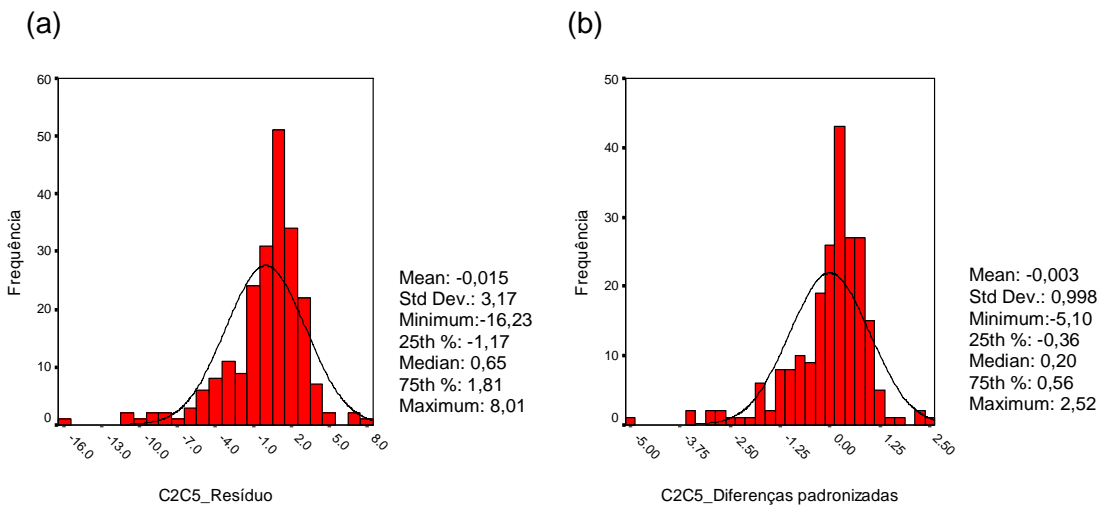


Figura 5.43 – Histograma de frequência dos erros (a) e dos erros padronizados da variável C_2C_5 (b), ajustado ao modelo exponencial.

O histograma de frequência de erros apresentou uma pequena assimetria negativa, refletindo uma leve tendência de ocorrências de subestimativas. A distribuição dos erros não se afastou da normal padrão indicando um ajuste muito bom do modelo

exponencial, ou seja, o modelo ajustado descreve adequadamente a dependência espacial dos dados de C_2C_5 .

- **Eteno (C_2)**

Na validação cruzada da variável eteno também foram testados três modelos e suas estatísticas estão na Tabela 5.13. Observou-se que o modelo mais adequado para descrever o fenômeno de dependência espacial da variável eteno foi o esférico, tendo o modelo exponencial apresentado menor valor do efeito pepita, $C_0 = 14,7491$.

Tabela 5.13 – Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável eteno.

Modelo	Esférico	Exponencial	Gaussiano
Média	-0,002	-0,004	-0,003
Variância	1,003	0,994	1,014

A figura 5.44 apresenta aparente uniformidade e sem tendência nos valores calculados. Lembrando que no mapa dos erros de krigagem os pontos marcados com o sinal “+” correspondem aos valores superestimados e “x” aos valores subestimados, quanto maior o tamanho do sinal, maior é a descontinuidade espacial dos dados.

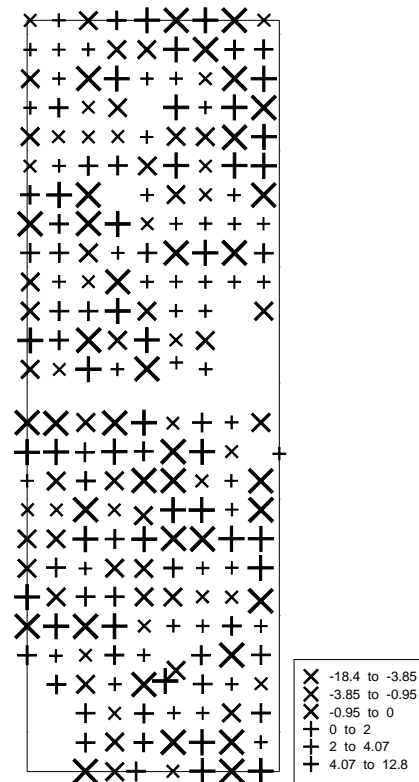


Figura 5.44 - Mapa de erros dos valores estimados da variável eteno em relação aos observados ajustado ao modelo esférico.

Podemos verificar que os valores subestimados estão correlacionados com os valores mais altos dos dados observados da variável, bem como os valores superestimados estão correlacionados com os valores mais baixos. O desvio padrão e média dos erros padronizados da validação cruzada para o modelo esférico foram 1,003 e -0,002, respectivamente.

O gráfico de dispersão dos erros positivos e negativos pela krigagem foi semelhante (Figura 5.45), observando maior afastamento da reta 1:1 para os erros negativos nas maiores medidas de eteno. Quanto aos erros positivos, as superestimativas da variável foram observadas nos menores valores e os demais se equilibraram entre os valores superestimados e os subestimados.

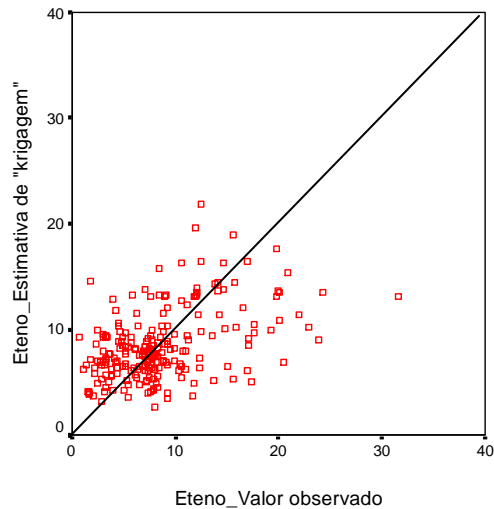


Figura 5.45 – Gráficos de dispersão das estimativas de krigagem em relação aos valores observados na validação cruzada para a variável eteno no modelo esférico.

Os histogramas de frequência dos erros (resíduos) e dos erros padronizados da variável eteno são mostrados na Figura 5.46. As estatísticas assinalaram valores próximos do ideal (Figura 5.46b).

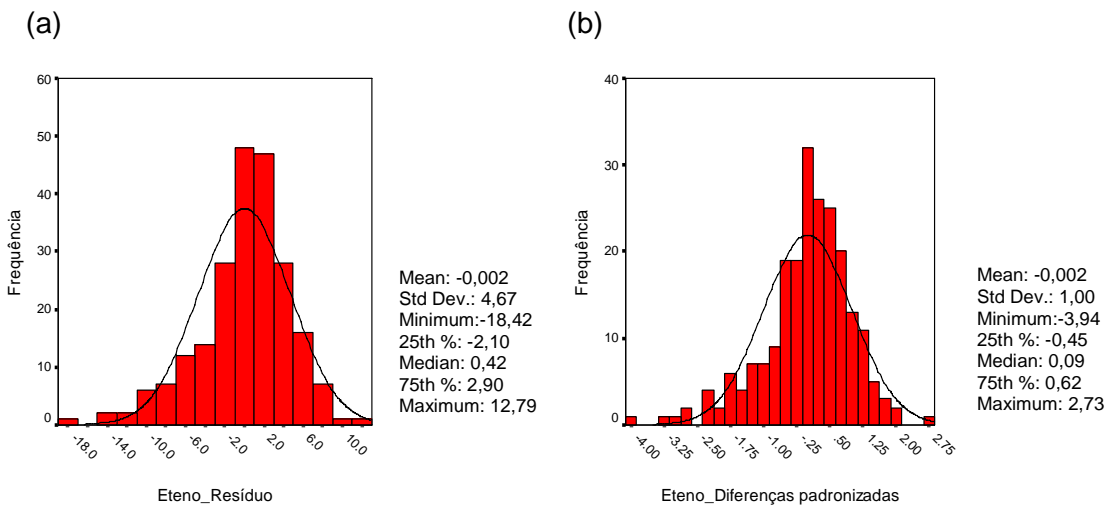


Figura 5.46 – Histograma de frequência dos erros (a) e dos erros padronizados da variável eteno (b), ajustado ao modelo esférico.

O histograma de frequência de erros apresentou para a variável eteno uma distribuição bem próxima da normal padrão e com boa simetria.

5.4.4 Comparação entre métodos de interpolação

- **Metano (C_1), C_2C_5 e Eteno ($C_2^=$)**

Pela comparação dos mapas obtidos, verifica-se que o método de krigagem ordinária apresenta distribuição espacial muito mais homogênea do que os demais. Esse resultado é obtido principalmente pelo fato desse interpolador ser não-viciado, com variância mínima.

O desempenho desses interpoladores foi obtido e comparado usando o critério do quadrado médio do erro (QME). Para os valores estimados pelos métodos do inverso do quadrado da distância e curvatura mínima, esse critério pode ser aplicado diretamente, pois não são interpoladores ótimos, e a diferença dos valores estimados e os observados não são nulos. Entretanto, para o interpolador de krigagem ordinária, por ser ótimo, essa diferença é nula, logo, para ter os valores estimados no ponto observado, o método de validação cruzada (VIEIRA, 2000; ISAAKS & SRIVASTAVA, 1989) foi usado. Esse método envolve a estimativa de cada ponto medido, "fazendo de conta" que ele não existe durante a sua estimativa. A razão é que krigagem ordinária é um interpolador exato, passando exatamente pelo ponto observado, quando esse é usado no cálculo.

Na Tabela 5.14, apresenta-se os resultados obtidos para o critério de comparação (QME) para as variáveis C_1 , C_2C_5 e $C_2^=$. O valor do critério deve ser próximo de zero se o algoritmo é preciso. Altos erros estimados são obtidos para os três interpoladores. Para os dados das três variáveis, o interpolador que apresentou resultado mais acurado, foi o de krigagem ordinária, seguido pelo o inverso do quadrado da distância e, por último, curvatura mínima.

Tabela 5.14 – Valores do quadrado médio do erro para interpoladores usados na distribuição espacial das variáveis C_1 , C_2C_5 e $C_2^=$.

Variável	Krigagem Ordinária	Inverso do Quadrado da Distância	Curvatura Mínima
	QME		
C_1	61,29	65,5	97,35
C_2C_5	9,55	10,31	12,73
$C_2^=$	20,31	21,74	34,6

CAPÍTULO 6 – CONCLUSÕES E RECOMENDAÇÕES

Os dados geoquímicos de hidrocarbonetos gasosos não seguem de maneira geral uma distribuição lognormal ou normal, sendo uma exceção e não uma regra. Até quando outros diferentes métodos de transformação são usados, os dados não se aproximaram de uma distribuição normal. Portanto, não existe uma boa razão para continuar a usar a regra [média \pm 2 desvios padrão (dp)].

A regra [mediana \pm 2 desvios mediano absoluto], seguido do *boxplot* parecem ser mais adequados para determinar o valor limiar quando utilizamos dados geoquímicos, que são caracterizados por forte assimetria.

O 98º percentil parece não ser adequado para estimar valores de limiar, pois entregou a mesma quantidade de valores anômalos para qualquer tipo de distribuição.

O interpolador geoestatístico de krigagem ordinária apresenta melhor resultado que os demais interpoladores que ignoram a dependência espacial entre observações. Sendo possível indicar com maior precisão os locais de distribuição das variáveis analisadas, dando subsídios para delimitar áreas com possível potencial petrolífero e novas pesquisas com custo reduzido.

Como recomendações, sugerimos uma avaliação geoquímica de superfície mais abrangente, aplicando também a análise geoestatística para as demais variáveis.

Analisar possíveis influências de fatores ambientais (características do solo, tais como umidade, granulometria, temperatura, ph, condutividade, etc.) sobre as concentrações de gases medidas nas amostras.

As principais dificuldades encontradas foram a falta de trabalhos aplicados a geoquímica de superfície de hidrocarbonetos gasosos aplicando ferramentas geoestatísticas, o aprendizado estatístico e geoestatístico e a ausência de maiores informações interdisciplinares sobre a região estudada que poderiam contribuir para uma melhor análise dos dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAMS, M. A., 2005, "Significance of Hydrocarbon Seepage Relative to Petroleum Generation and Entrapment". *Marine and Petroleum Geology*, 22: 457-477.
- ABRAMS, M. A., SEGALL, M. P., BURTELL, S. G., 2001, *Best Practices For Detecting, Identifying and Characterizing Near-Surface Migration of Hydrocarbons Within Marine Sediments*, In: Proceedings of The 2001 Offshore Technology Conference, OTC 13039, Houston, Apr-May. pp. 1-14.
- AHRENS, L. H., 1953, "A Fundamental Law of Geochemistry". *Nature*, 172:1148.
- AHRENS, L. H., 1954a, "The Lognormal Distribution of The Elements" (A Fundamental Law of Geochemistry and Its Subsidiary). *Geochimica et Cosmochimica Acta*, 5: 49-74.
- AHRENS, L. H., 1954b, "The Lognormal Distribution of the Elements II". *Geochimica et Cosmochimica Acta*, 6: 121-132.
- AHRENS, L. H., 1957, "Lognormal-Type Distribution III". *Geochimica et Cosmochimica Acta*, 11: 205-213.
- ALMEIDA-FILHO, R. 2001, *Processamento Digital de Imagens Landsat-TM na Detecção de Áreas de Microexsudação de Hidrocarbonetos*, Região da Serra de Tonã, Bahia. In: SBSR, Workshops, Fox do Iguaçu, X, INPE, pp. 235- 242.
- ANDRIOTTI, J. L. S., 2003, *Fundamentos de Estatística e Geoestatística*. São Leopoldo: Unisinos. 165 p.
- AUBREY, K. V., 1954, "Frequency Distribution of the Concentrations of Elements in Rocks". *Nature*, 174: 141-142.
- AUBREY, K. V., 1956, "Frequency distribution of elements in igneous rocks". *Geochimica et Cosmochimica Acta*, 6: 83-90.
- BABINSKI, N. A., MELLO, M. R., 1984, *Geoquímica de Superfície*. In: Geoquímica do Petróleo. CENPES, PETROBRAS.

- BARNETT V., LEWIS T., 1994, *Outliers in Statistical Data*. 3 ed. New York: John Wiley & Sons, Inc.
- BERNARD, B. B., BROOKS, J. M., SACKET, W. M., 1976, "Natural Gas Seepage in the Gulf of Mexico". *Earth and Planetary Science Letters*, Amsterdam, v. 31, n. 1, pp. 48-54.
- BURROUGH, P. A., 1986, *Principles of Geographical Information System for Land Resources Assessment*. New York: Oxford University Press. 220p.
- BURROUGH, P.A., VAN RIJN, R., RIKKEN, M., 1996, *Spatial Data Quality and Error Analysis Issues: GIS Functions and Environmental Modeling*. In: GIS and Environmental Modeling: Progress and Research Issues. Edited by Michael F. Goodchild, Louis T. Steyaert, Bradley O. Parks, Carol Johnston, David Maidment, Michael Crane, and Sandi Gendinning, pp. 29-34.
- CHAYES, F., 1954, "The Lognormal Distribution of Elements: A Discussion". *Geochimica et Cosmochimica Acta*, 6: 119-121.
- CHENG, Q., 1999, "A Spatial and Scaling Modeling for Geochemical Anomaly Separation". *J. Geochem. Explor.* 65: 175-194.
- CHENG, Q., AGTERBERG, F. P., BONHAM-CARTER, G. F., 1996, A Spatial Analysis Method for Geochemical Anomaly Separation. *J. Geochem. Explor.* 56: 183–195.
- CLARK, I. & HARPER, W. V., 2000, *Practical geostatistics*. Ecosse North America Llc, Colombus, Ohio, USA. 442p.
- COLEMAN, D. D., 1979, *The Origin of Drift-Gas Deposits as Determined by Radiocarbon Dating of Methane*. In: R. Berger and H.E. Seuss (Editors), Radiocarbon Dating, Proceedings of the Ninth International Radiocarbon Dating Conference, 1976. University of California Press, Berkley, pp. 365-387.
- CRESSIE, N. A. C., 1991, *Statistics for Spatial Data*. New York: John Wiley & Sons, Inc. 900 p.

- CRESSIE, N. A. C., 1993, *Statistics for Spatial Data*. Revised Edition. John Wiley & Sons, Inc. 928p.
- DAVIDSON, M. J., 2004, "Evolution of Scientific Geochemical Exploration". *Oil & Gas Journal*, v. 102.23, June 21, pp. 35-80.
- DAVIS, J. B., SQUIRES, R. M., 1954, "Detection of Microbially Produced Hydrocarbons Other Than Methane". *Science*, 119: 381–382.
- DEUTSCH, C. V., JOURNEL, A. G., 1992, *GSLIB: Geostatistical Software Library and User's Guide*. New York: Oxford University Press. 336 p.
- DÍEZ, M., SIMÓN, M., DORRONSORO, C., GARCÍA, I., MATÍN, F., 2007, "Background Arsenic Concentrations in Southeastern Spanish Soils". *Science of the Total Environment*, 378: 5-12.
- DONOVAN, T. J., FRIEDMAN, J., GLEASON, J. D., 1974, "Recognition of Petroleum-Bearing Traps by Unusual Isotopic Compositions of Carbonate-Cemented Rocks". *Geology*, v. 2, pp. 351–354.
- DONOVAN, T.J., 1974, "Petroleum Microseepage at Cement, Oklahoma: Evidence and Mechanism". *Am. Assoc. Pet. Geol. Bull.* 58-3, 429–446.
- DUARTE, J. B., 2000, *Sobre o Emprego e a Análise Estatística do Delineamento em Blocos Aumentados no Melhoramento Genético Vegetal*. Tese (Doutorado) – Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, Universidade de São Paulo. 292 p.
- DUCHSCHERER, W. JR., 1980, "Geochemical Methods of Prospecting for Hydrocarbons". *Oil & Gas Journal*, v. 1, pp. 194-208.
- DUTTER R., FILZMOSE P., GATHER U., ROUSSEEUW, P., 2003, *Developments in Robust Statistics. International Conference on Robust Statistics, 2001*. Heidelberg: Physik-Verlag.

- FELGUEIRAS, C. A., 1999, *Modelagem Ambiental com Tratamento de Incertezas em Sistemas de Informação Geográfica: O Paradigma Geoestatístico por Indicação*. São José dos Campos: INPE. 212 p.
- FERGUSON, J. D., 1975, *Surface Alteration and Mineralization of Permian Red Beds Overlying Several Oil Fields in Southern Oklahoma*. Master's Thesis, Oklahoma State University, Stillwater, Oklahoma, 95 p.
- FRIZZO, S. J., LICHT, O. A. B., 2007, "Estatísticas Uni e Bivariadas Aplicadas à Prospecção Geoquímica". In: LICHT, O. A. B., MELLO, C. S. B., Silva, C. R.(Eds), *Prospecção Geoquímica de Depósitos Minerais Metálicos, Não-Metálicos, Óleo e Gás*. SBGq - Sociedade Brasileira de Geoquímica / CPRM – Serviço Geológico do Brasil, Rio de Janeiro. Cap. 19: 595-618.
- GARRET, R. G., 1989, "A Cry from the Heart. Explore: , *The Assoc. of Exploration Geochemists*. n. 66: 18-20.
- GLÓRIA, K. S., 2002, *Metodologias Alternativas para Detecção e Localização de Vazamentos*. Boletim Técnico PETROBRAS, Rio de Janeiro, 45 (2).
- GONÇALVES, A. C. A., FOLEGATTI, M. V., MATA, J. D. V., 2005, "Análises Exploratória e Geoestatística da Variabilidade de Propriedades Físicas de um Argissolo Vermelho. Maringá". *Acta Scientiarum*. v.23, n.5, pp.1149-1157.
- GOVETT, G. J. S., GOODFELLOW, W. D., CHAPMAN, A., CHORK, C. Y., 1975, "Exploration Geochemistry Distribution of Elements and Recognition of Anomalies". *Math. Geol.* 7: 415–446.
- GUERRA, P. A. G., 1988, *Geoestatística Operacional*. Brasília: Departamento Nacional da Produção Mineral. 145 p.
- HAMPEL F. R, RONCHETTI E. M, ROUSSEEUW P. J., STAHEL W., 1986, *Robust Statistics. the Approach Based on Influence Functions*. New York: John Wiley & Sons. 528p.
- HAMPEL, F., 1974, "The Influence Curve and Its Role in Robust Estimation". *J. Am. Statist. Assoc.* 69: 383–393.

- HAWKES, H. E., WEBB, J. S., 1962, *Geochemistry in Mineral Exploration*. Harper & Row, Publishers, New York. 409 p.
- HORVITZ, L., 1939, "On Geochemical Prospecting". *Geophysics*, 4: 210-228.
- HORVITZ, L., 1969, *Hydrocarbon Geochemical Prospecting After Thirty Years*. In: Heroy, W. B. (Editor), *Unconventional Methods in Exploration for Petroleum and Natural Gas*. Southern Methodist University Press, Dallas, pp. 205-218.
- HUBER, P. J., 1981, *Robust Statistics*. New York: John Wiley & Sons.
- HUNT, J. M., 1996, *Petroleum Geochemistry and Geology*. W. H. Freeman Co.
- ISAAKS, E. H., SRIVASTAVA, R. M., 1989, *An Introduction to Applied Geoestistics*. New York: Oxford University Press. 561 p.
- JONES III, V. T., MATTHEWS, M. D., RICHERS, D. M., 1999, *Light Hydrocarbons for Petroleum and Gas Prospecting. Geochemical Remote of the Subsurface*. Handbook of Exploration Geochemical, V. 7, Elsevier Science B. V. Disponível em: <<http://www.eti-geochemistry.com/elsevier/>>. Acesso: 29/08/2007.
- JONES, V. T., DROZD, R. J., 1983, "Predictions of Oil or Gas Potential by Near Surface Geochemistry". *The American Association of Petroleum Geologists. Bulletin* v. 67, n. 6 (June), pp. 932-952.
- JONES, V. T., THUNE, H.W., 1982, *Surface Detection of Retort Gases from an Underground Coal Gasification Reactor in Steeply Dipping Beds Near Rawlins, Wyoming*. Soc. Petroleum Engineers, SPE 11050, 24 p.
- JONES, V. T., BURTELL, S. G. Hydrocarbon Flux Variations in Natural and Anthropogenic Seeps. Disponível em: <<http://www.eti-geochemistry.com/flux/index.html/>>. Acesso: 29/08/2007.
- JOURNEL, A. G., HUIJBREGTS, C. J., 1978, *Mining Geostatistics*. London: Academic Press. 600 p.

- KARTSEV, A. A., TABARSARANSKII, Z. A., SUBBOTA, M. I., MOGILEVSKII, G. A., 1959, *Geochemical Methods of Prospecting and Exploration for Petroleum and Natural Gas*: Berkeley, Los Angeles, University of California Press, 347 p.
- KITANIDIS, P. K., 1997, *Introduction to Geostatistics: Applications to Hydrogeology*. Cambridge: Cambridge University Press. 249 p.
- KLUSMAN, R. W., 2002, "Interpretation and Display of Surface Geochemical Data". In: Schumacher, D. E Leshack, L. A, (Eds). *Surface Exploration Case Histories: Applications of Geochemistry, Magnetism and Remote Sensing*. AAPG Studies/SEG Geophysical References Series, n. 11, pp.1-24.
- KOCH Jr., G. S., LINK, R. F., 1971, *Statistical Analysis of Geological Data*. New York: John Willey & Sons Inc.
- LANDIM, P. M. B., 1998, *Análise Estatística de Dados Geológicos*. Rio Claro: UNESP.
- LANDIM, P. M. B., 2000, *Introdução aos Métodos de Estimção Espacial para Confecção de Mapas*. Rio Claro: UNESP. 20 p.
- LANDIM, P. M. B., 2002, *Introdução à Confecção de Mapas pelo Software Surfer*. Rio Claro: UNESP. 20 p.
- LANDIM, P. M. B., STURARO, J. R., MONTEIRO, R. C., 2002, *Krigagem Ordinária para Situações com Tendência Regionalizada*. DGA, IGCE, UNESP: Rio Claro. Lab. Geomatématica, Texto Didático n. 6, 12p. Disponível em <<http://www.rc.unesp.br/igce/aplicada/textodi.html>>. Acesso em: 17/10/2008.
- LAUBMEYER, G., 1933, "A New Geophysical Prospecting Method": *Zeitschrift for Petroleum*, v. 29, n. 18, pp. 14.
- LEIFER, I., CLARK, J.F., LUYENDYK, B., VALENTINE, D., 2003, "Identifying future directions for subsurface hydrocarbon migration research", *EOS (American Geophysical Union Transactions)* 84(37): 364-371.

- LEYTHAEUSER, D., SHAEFER, R. G., YÜKLER, A., 1982, "Role of Diffusion in Primary Migration of Hydrocarbons". *American Association of Petroleum Geologists. Bulletin*, n. 66, pp. 408-429.
- LI, C., MA, T., SHI, J., 2002, "Application of a Fractal Method Relating Concentrations and Distances for Separation of Geochemical Anomalies from Background". *J. Geochem. Explor.*, 77: 167–175.
- LICHT, O. A. B., 1998, *Prospecção Geoquímica: Princípios, Técnicas e Métodos*. Rio de Janeiro: CPRM. 236 p.
- LINK, W. K., 1952, "Significance of Oil and Gas Seeps in World Oil Exploration". *American Association of Petroleum Geologists. Bulletin*, v. 36, n. 8 (Aug), pp. 1505-1538.
- LITCH, O. A. B., OLIVEIRA, L. M., 1989, "Prospecção Geoquímica de Mineralizações Estaníferas da Região de Capivari-Pardo, Paraná, Brasil". In: *Congresso Brasileiro De Geoquímica, 2*, Rio de Janeiro. Anais...Sbgq, pp. 99-120.
- MACKAY, D., PATERSON, S., 1984, "Spatial Concentration Distribution". *Env.Sci. & Technology*, Washington, v.18, n.1, pp. 207A-214A.
- MARANHÃO, R. J. L., 1983, *Introdução à Pesquisa Mineral*. Banco do Nordeste do Brasil; ETE-NE. Monografia, 10:680.
- MARTÍNEZ, J., LLAMAS, J., DE MIGUEL, E., REY, J., HIDALGO, M.C., 2006, "Determination of the Geochemical Background in a Metal Mining Site: Example of the Mining District of Linares (Spain)". *Journal of Geochemical Explor.*, 94: 19-29.
- MATTHEWS, M. D., 1996a, "Migration – "A View from the Top". In: *Hydrocarbon Migration and its Near-Surface Expression* (Schumacher D. and Abrams, M. A., eds): AAPG Memoir 66, pp. 139-155.
- MATTHEWS, M. D., 1996b, "Importance of sampling design and density in target recognition". In: *Hydrocarbon Migration and its Near-Surface Expression* (Schumacher D. and Abrams, M. A., eds): AAPG Memoir 66, pp. 243-253.

- MATTHEWS, M.D., 1985, "Effects of Hydrocarbon Leakage on Earth Surface Materials". In: M.J. Davidson (ed.), *Unconventional Methods in Exploration for Petroleum and Natural Gas*, Southern Methodist University, Dallas, pp. 27-44.
- MCKENNA, E.J., KALLIO, R.E., 1965, The Biology of Hydrocarbons. *Annual Reviews of Microbiology*, 19: 182-208.
- MEAD, R., CURNOW, R. N., HASTED, A. M., 1993, *Statics Methods in Agriculture and Experimental Biology*. London: Chapman & Hall. 335 p.
- MELLO, M. R., 1984, *Geoquímica do Petróleo*. Rio de Janeiro, CENPES, PETROBRAS.
- MIESCH, A.T., 1981, "Estimation of The Geochemical Threshold and its Statistical Significance". *J. Geochem. Explor.* 16, 49–76.
- MILLER, R. L., GOLDBERG, E. D., 1955, "The normal distribution in geochemistry". *Geochimica et Cosmochimica Acta*, 8: 53-62.
- OILTRACERS L.L.C. *Determining the Origin of Hydrocarbon Gas Shows and Gas Seeps*. Disponível em: <<http://www.gaschem.com/determ.html>>. Acesso: 26/08/2007.
- ONTARIO MINISTRY OF ENVIRONMENT AND ENERGY AT., 1993, *Ontario Typical Range of Chemical Parameters in Soil, Vegetation, Moss Bags and Snow*. Toronto: Ontario Ministry of Environment and Energy. Disponível em: <<http://www.ene.gov.on.ca/>>. Acesso: 5/11/2008.
- PEREIRA, R. M., 2003, *Fundamentos de Prospecção Mineral*. Rio de Janeiro: Interciência.
- PIRSON, S. J., 1960, "How to Make Geochemical Exploration Succeed". *World Oil*, 150 (5): 93-6.
- PIRSON, S.J., 1969, Geological, "Geophysical and Chemical Modifications of Sediments in The Environment of Oil Fields". In: W. B. Heroy (Ed),

Unconventional Methods In Exploration For Petroleum And Natural Gas.
Southern Methodist University Press, Dallas, pp. 159-186.

POTVIN, C., 1993, ANOVA: "Experiments in Controlled Environments". In: SCHEINER, S.M. & GUREVITCH, J. (Eds.). *Design and Analysis of Ecological Experiments*. London: Chapman & Hall. pp. 46-68.

QUÍLEZ, A. L., 2006, *Estatística Espacial en Epidemiología y Medio Ambiente*. Tese de Doutorado. Universidade de Valência, Espanha. Disponível em: <<http://www.uv.es/antonio/EEEMA/EEEMA.pdf>>. Acesso: 10/09/2008.

RANSONE, W.R., 1947, "Geochemical History of the Hardy Oil Field, Jones County, Texas": *Geophysics*, v. 12, pp. 384-392.

REIMANN, C., FILZMOSER, P., 1999, "Normal and Lognormal Data Distribution in Geochemistry: Death of a Myth. Consequences for the Statistical Treatment of Geochemical and Environmental Data". *Environmental Geology*, 39 (9) July. Springer-Verlag. pp. 1001-1014.

REIMANN, C., FILZMOSER, P., GARRET, R., 2005, "Background and Threshold: Critical Comparison of Methods of Determination". *Sci Total Environ*; 346: 1–16.

REITSEMA, R. H., LINDBERG, F. A., KALTENBACK, A. J., 1978, "Light Hydrocarbons in Gulf of Mexico Water: Sources and Relation to Structural Highs". *J. Geochem. Explor.* 10: 139-51

RIBEIRO JÚNIOR, P. J., 1995, *Métodos Geoestatísticos no Estudo da Variabilidade Espacial dos Parâmetros do Solo*. Dissertação de Mestrado - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo. Piracicaba, 99p.

RICE, D. D., 1975, "Origin of the Conditions for Shallow Accumulations of Natural Gas": *Wyoming Geol. Assoc.* 27th Ann. Field Trip Guidebook.

RICHERS, D. M., REED, R. J., HORSTMAN, K. C., MICHELS, G. D., BAKER, R. N., LUNDELL, L., MARRS, R. W., 1982, "Landsat and Soil-Gas Geochemical Study of Patrick Draw Oil Field, Sweetwater County, Wyoming". *AAPG Memoir*, 66, N. 7, pp. 903-922.

- ROCHA, C. L., 2005, *Análise de Fronteiras de Reservatório de Petróleo Através de Geoquímica de Superfície e Mineração de Dados*. Tese de Mestrado. Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.
- ROSAIRE, E. E., 1938, "Shallow Stratigraphic Variations Over Gulf Coast Structures". *Geophysics*, v. 3, n. 2, pp. 96-115.
- ROSAIRE, E. E., 1940, "Geochemical Prospecting for Petroleum". In: *Symposium on Geochemical Exploration*. American Association of Petroleum Geologists Bulletin, V. 24, pp. 1400-1433.
- ROUSSEEUW P. J., LEROY A. M., 1987, *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.
- SACKETT, W. M. 1977, "Use of Hydrocarbon Sniffing in Offshore Exploration". *J. Geochem Explor.* 7, pp. 243-54.
- SCHOELL, M., 1983, "Genetic Characterization of Natural Gases". *American association of petroleum geologist*. Bulletin 67, pp. 2225-2238.
- SCHUMACHER, D. *Surface geochemical exploration for petroleum*. Disponível em : [http:// www.gmtgeochem.com/pdf/ch18.pdf](http://www.gmtgeochem.com/pdf/ch18.pdf). Acesso: 24/08/2007.
- SCHUMACHER, D., 2000, "Surface Geochemical Exploration for Oil and Gas: New Life for an Old Technology". *The Leading Edge*, pp. 258-261.
- SCHUMACHER, D., 2002, "Hydrocarbons Geochemical Exploration for Petroleum". In: D. Schumacher e M.A. Abrams (eds). *Hydrocarbon Migration and its Near-Surface Expression*. AAPG, Memoir 66, pp.184-204.
- SEALEY, J. Q., 1974a, "A Geomicrobiological Method of Prospecting for Petroleum". Part 1. *Oil and Gas Journal*, April 8, pp. 142-146.
- SEALEY, J. Q., 1974b, "A Geomicrobiological Method of Prospecting for Petroleum". Part 2. *Oil and Gas Journal*, April 15, pp. 98-102.

- SECHMAN, H., DZIENIEWICZ, M., 2007, "Influence of Oil Moisture on The Results of Surface Geochemical Survey Applied to Petroleum Exploration". *J. Geochem. Explor.* 56: 267-282.
- SHAW, D. M., 1961, "Element Distribution Laws of Geochemistry". *Geochimica et Cosmochimica Acta*, 23: 116-134.
- SIEGEL, S., 1975, *Estatística Não-Paramétrica para as Ciências do Comportamento*. São Paulo: McGraw-Hill.
- SINCLAIR, A. J., 1974, "Selection of Thresholds in Geochemical Data Using Probability Graphs". *J. Geochem. Explor.* 3: 129–149.
- SINCLAIR, A. J., 1976, "Application of Probability Graphs in Mineral Exploration". *Assoc. Explor. Geochem. Spec.* Vol. 4, 95 p.
- SINCLAIR, A. J., 1991, "A Fundamental Approach to Threshold Estimation in Exploration Geochemistry; Probability Plots Revisited". *J. Geochem. Explor.* 41: 1–22.
- SNEDECOR, G. W., COCHRAN, W. G., 1967, *Statistical Methods*. 6. ed. Ames, Iowa: Iowa State University. 593 p.
- SOARES, A., 2000, *Geoestatística Aplicada a Ciência da Terra e do Ambiente*. IST Press, 206 p.
- SOKOLOV, V. A., 1933, "The Gas Survey as a Method of Prospecting for Oil and Gas Formations": *Technika, NGRI*, n. 1 (February).
- SOKOLOV, V. A., 1971, "The Theoretical Foundations of Geochemical Prospecting for Petroleum and Natural Gas and the Tendencies of it's Development": *CIM Special*, v. 11, pp. 544-549.
- SOLI, G., 1954, "Geomicrobiological Prospecting". *American Association Of Petroleum Geologists. Bulletin*, v. 38, pp. 2255-2558.

- SOLI, G., 1957, "Microorganisms and Geochemical Methods of Oil Prospecting". *American Association of Petroleum Geologists. Bulletin*, v. 41, pp. 134-140.
- SPIEGEL, M. R., 1985, *Estatística*. 2 ed. São Paulo: McGraw Hill. 454 p.
- STANLEY, C. R., 1988, *Comparison of Data Classification Procedures in Applied Geochemistry Using Monte Carlo Simulation*. Ph.D. Thesis, Univ. of British Columbia, Vancouver, 256 p.
- TEPLITZ, A. J., RODGERS, J. K., 1935, *Research Project DA-51-3(1)*. Gulf Research and Development Company, Archives Location 71-B23, Box 14D-34.
- THRASHER, J. *et al.*, 1996, "Understanding Geology as The Key to Using Seepage in Exploration: The Spectrum of Seepage Styles". In: Schumacher, D. & Abrams, M. (Eds.). *Hydrocarbon Migration and it's Near-Surface Expression*. Tulsa: AAPG Memoir 66. Pp. 223-241.
- TISSOT, B. P., WELTE, D. H., 1984, *Petroleum Formation and Occurrence*. 2 ed., Berlin, Springer-Verlag.
- TOBIAS, F. J., BECH, J., SÁNCHEZ ALGARRA, P., 1997, *Statistical Approach to Discriminate Background and Anthropogenic of Trace Elements in Soils of Catalonia, Spain*. *Water, Air, Soil Pollution* 100:63-78. Kluwer Academic Publishers, Netherlands.
- TRANGMAR, B. B., YOST, R. S., UEHARA, G., 1985, "Application of Geostatistics to Spatial Studies of Soil Properties". *Advances In Agronomy*, v. 38, pp.45-94.
- TUKEY, J. W., 1977, *Exploratory Data Analysis*. Reading: Addison-Wesley. 688 p.
- VALENTE, J. M. G. P., 1989, *Geomatématica: Lições de Geoestatística*. 2. ed. Ouro Preto: Fundação Gorceix. v. 3, pp. 511-713.
- VAN DER MEER, F., VAN DIJK, P., VAN DER WERF, H., YANG, H., 2002, "Remote Sensing and Petroleum Seepage: A Review and Case Study". *Terra Nova*. 14 (1), pp. 1-17.

VIEIRA, S.R. "Geoestatística em Estudos de Variabilidade Espacial do Solo". In: NOVAIS, R.F., ALVAREZ, V.H., SCHAEFER, C.E.G.R. (Ed.). *Tópicos em Ciência do Solo*. Viçosa: Sociedade Brasileira de Ciência do Solo, 2000. v. 1, pp.1-54.

VISTELIUS, A. B., 1960, "The Skew Frequency Distribution and the Fundamental Law of the Geochemistry Process". *J. Geol*, 68: 1-22.

WARRICK, A. W., NIELSEN, D. R., 1980, "Spatial Variability of Some Physical Properties of the Soil". In: Hill, D. ed. *Applications of Soil Physics*, New York: Academic Press, Cap. 13, pp. 319-344.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)