

UNIVERSIDADE ESTADUAL PAULISTA

Instituto de Geociências e Ciências Exatas

Campus de Rio Claro

**UTILIZAÇÃO DA GEOESTATÍSTICA NA QUANTIFICAÇÃO DO
RISCO DE CONTAMINAÇÃO POR METAIS PESADOS, NA ÁREA
PORTUÁRIA DE SANTANA-AMAPÁ.**

Joaquim Carlos Barbosa Queiroz

Orientador: Prof. Dr. José Ricardo Sturaro

**Tese de Doutorado elaborada junto ao Curso de Pós-
Graduação em Geociências - Área de Concentração em
Geociências e Meio Ambiente para obtenção do título de
Doutor em Geociências**

Rio Claro (SP)

2003

Banca Examinadora

Prof. Dr. José Ricardo Sturaro (orientador)
IGCE/UNESP/Rio Claro (SP)

Prof. Dr. Paulo Milton Barbosa Landim
IGCE/UNESP/Rio Claro (SP)

Prof. Dr. Armando Zaupa Remarcere
IG/UNICAMP/Campinas (SP)

Prof. Dr. José Alberto Quintanilha
Escola Politécnica/USP/São Paulo (SP)

Prof. Dr. José Sílvio Golvone
IGCE/UNESP/Rio Claro (SP)

Joaquim Carlos Barbosa Queiroz

Rio Claro, 02 de dezembro de 2003

Resultado:

Este trabalho é dedicado à Verônica, meu amor,
esposa e companheira, pela presença constante
e serena e às filhinhas Verissa e Larissa, por
estarem conosco no caminho de volta.

*Por mais paradoxal que seja, todas as ciências
exatas se baseiam nas aproximações.*

Bertrand Russel

AGRADECIMENTOS

Ao Prof. Dr. José Ricardo Sturaro pela orientação, amizade, apoio, confiança e contribuições importantes na elaboração deste trabalho.

Aos professores da UNESP/Rio Claro, pelos ensinamentos transmitidos durante os cursos realizados, especialmente à Profa. Paulina Riedel, Prof. Ricardo Sturaro, Prof. Paulo M. Landim, Prof. Daniel Bonotto, Profa. Margarita Moreno, Prof. Chang Kiang e Prof. José H. Barcelos.

Aos Professores Dr. Paulo Milton Barbosa Landim, do Departamento de Geologia Aplicada do IGCE – UNESP Rio Claro e Profa. Dra. Dejanira de Angelis do Departamento de Bioquímica e Microbiologia da UNESP Rio Claro, pelas sugestões na qualificação.

À Profa. Dra. Elizabeth Yamada, do Departamento de Biologia da Universidade Federal do Pará, Coordenadora do projeto “Estudo sobre envolvimento de múltiplos fatores na gênese de doenças neuro-degenerativas: uma abordagem multidisciplinar para se investigar o papel de fatores ambientais (exposição a metais pesados e pesticidas) e genéticos na fisiopatogenia da doença de Parkinson”, que viabilizou a passagem, diárias e materiais necessários para levantamento de campo.

Ao Prof. Dr. Helyelson Paredes Moura, do Departamento de Ciências Exatas e Naturais da Fundação Universidade Federal do Amapá, pela amizade, apoio e incentivo.

Ao Eng. Químico Augusto César F. Saraiva, do Laboratório Central da Eletronorte – LACEN, pelas análises químicas.

A todos que auxiliaram nos trabalhos de campo, principalmente o meu primo Chico, meu irmão Rômulo, Nacélio e Adalberto, técnicos da Ampla Engenharia, Jussara Nunes, Amaro, Zé Maria, Lydia e Percy, Porfíria e Marco Aurélio.

À CAPES pelo apoio financeiro na forma de bolsa do PICDT.

A todos os colegas mestrandos e doutorandos da UNESP, pelo convívio saudável durante esses quatro anos.

À minha esposa Verônica, pelas sugestões e revisão do texto, e às minhas filhas Verissa e Larissa pela compreensão, apoio e estímulo durante a realização desse trabalho. Ao Toby Toy pela companhia e alegria.

Aos meus pais que sempre me apoiaram e acreditaram em mim.

Enfim, a todos que direta ou indiretamente contribuíram para a realização desta tese.

SUMÁRIO

LISTA DE FIGURAS	iii
LISTA DE TABELAS	ix
LISTA DE FOTOGRAFIAS	x
RESUMO	xi
ABSTRACT	xii
1. INTRODUÇÃO	01
2. REVISÃO BIBLIOGRÁFICA	05
2.1. Considerações gerais.....	05
2.2. Contaminação ambiental por metais pesados.....	07
3. MATERIAIS E MÉTODOS	11
3.1. Levantamento amostral.....	11
3.1.1. Planejamento de execução da amostragem	11
3.1.2. Procedimentos adotados na amostragem	12
3.2. Análise química dos dados	14
3.3. A Geoestatística.....	14
3.3.1. Introdução.....	14
3.3.2. Geoestatística em meio ambiente	16
3.3.3. Conceitos básicos de geoestatística.....	18
3.3.4. Descrição de padrões espaciais.....	22
3.3.5. Modelamento da variação espacial.....	26
3.3.6. Estimativa espacial.....	33
3.3.7. Krigagem.....	33
3.3.7.1. Krigagem Ordinária.....	34
3.3.7.2. Cokrigagem.....	37
3.3.7.3. Krigagem Fatorial	38
3.3.7.4. Krigagem Fatorial Multivariada.....	39
3.3.8. Análise da incerteza	44
3.3.8.1. Funções de perda e estimativas ótimas.....	47
3.3.8.2. Erros de classificação	54
3.3.8.3. Modelo de Decisão baseado em risco custo-benefício.....	56

3.3.9. Simulações estocásticas.....	59
3.3.9.1. Tipos de Simulações condicionais.....	61
3.3.9.2. Simulação Annealing.....	62
4. ESTUDO DE CASO	66
4.1. O problema estudado.....	66
4.1. Aspectos gerais da área em estudo.....	74
4.1.1. Localização	74
4.1.2. Geologia	74
4.1.3. Vegetação	76
4.1.4. Clima	77
4.1.5. Hidrografia	77
4.1.5. Uso e ocupação do solo	77
5. APLICAÇÃO DAS TÉCNICAS GEOESTATÍSTICAS E DISCUSSÃO DOS RESULTADOS	80
5.1. Análise exploratória dos dados.....	80
5.1.1. Distribuição de frequências	80
5.1.2. Descrição bivariada	86
5.2. Análise espacial do risco de contaminação.....	88
5.2.1. Mapas de localização dos poluentes	88
5.2.2. Descrição dos padrões espaciais	93
5.2.3. Modelamento da co-regionalização	97
5.2.3.1. Krigagem fatorial multivariada.....	98
5.2.4. Modelos de incerteza locais para o risco de contaminação.....	121
5.2.5. Utilização da simulação annealing na avaliação da concentração dos contaminantes na área de estudo	134
5.2.5.1 Avaliação da incerteza a partir das simulações	139
7. CONSIDERAÇÕES FINAIS	164
8. REFERÊNCIAS BIBLIOGRÁFICAS	166
ANEXOS	

LISTA DE FIGURAS

Figura 3.1 – Balanço Hídrico da Região de Macapá/AP, para o período de 1968 a 1990.....	13
Figura 3.2 – Área de amostragem em Santana/AP com os respectivos pontos de amostragem.....	13
Figura 3.3 – Diagrama de dispersão para os pares (x_i, y_i)	22
Figura 3.4 – Parâmetros do semivariograma.....	29
Figura 3.5 – Modelos semivariográficos do tipo transitivo (com patamar) mais usuais em Geoestatística.....	30
Figura 3.6 – Modelos potência para diferentes valores de w	30
Figura 3.7 – Tipos de anisotropia variográfica: a) geométrica b) zonal.....	31
Figura 3.8 – Funções de perda e estimativas ótimas, $z_L^*(\mathbf{u})$, para o critério dos mínimos quadrados e desvio médio absoluto.....	51
Figura 3.9 – Função de perda utilizada na avaliação do custo relativo para o manganês, com $w_1 = 1$, $w_2 = 2,5$ e $z_c = 0,1$ ppm.....	53
Figura 3.10 – Erros de classificação e classificações corretas das localizações na área de interesse.....	56
Figura 3.11 – Relação entre custo e risco; o conceito de risco ótimo.....	58
Figura 3.12 – Fluxograma com os passos utilizados na simulação <i>annealing</i>	65
Figura 4.1 – (a) Estado do Amapá no extremo norte do Brasil ; (b) e (c) municípios do estado do Amapá com destaque para Serra do Navio e Santana.....	75
Figura 4.2 – Seção geológica esquemática da área industrial de Santana.....	76
Figura 4.3 – Vista parcial do mapa urbano de Santana, mostrando o uso do solo em torno da área industrial da ICOMI.....	79
Figura 5.1 – Histogramas e distribuições acumuladas do arsênio, manganês, ferro e cádmio.....	83
Figura 5.2 – Histogramas e distribuições acumuladas do alumínio, selênio, chumbo e cobre.....	83
Figura 5.3 – Histogramas e distribuições acumuladas da prata, níquel, bário e estrôncio.....	84

Figura 5.4 – Histogramas e distribuições acumuladas do zinco, molibdênio, cobalto e cromo.....	84
Figura 5.5 – Diagrama de dispersão do coeficiente de correlação linear versus coeficiente de correlação <i>rank</i> para os pares de poluentes que excedem o limite máximo permitido.....	87
Figura 5.6 – Mapa de localização das amostras de alumínio.....	89
Figura 5.7 – Mapa de localização das amostras de selênio (× 0,01 ppm).....	89
Figura 5.8 – Mapa de localização das amostras de ferro.....	90
Figura 5.9 – Mapa de localização das amostras de manganês.....	90
Figura 5.10 – Mapa de localização das amostras de arsênio.....	91
Figura 5.11 – Mapa de localização das amostras de chumbo (× 0,01 ppm).....	91
Figura 5.12 – Mapa de localização das amostras de cobre (× 0,01 ppm).....	92
Figura 5.13 – Mapa de localização das amostras de cádmio (× 0,001 ppm).....	92
Figura 5.14 – Mapa potenciométrico da área industrial e residencial de Santana.....	94
Figura 5.15 – Semivariograma experimental omnidirecional para o alumínio.....	94
Figura 5.16 – Semivariograma experimental omnidirecional para o manganês.....	95
Figura 5.17 – Semivariograma experimental omnidirecional para o arsênio.....	95
Figura 5.18 – Semivariograma experimental omnidirecional para o cádmio.....	95
Figura 5.19 – Semivariograma experimental omnidirecional para o ferro.....	96
Figura 5.20 – Semivariograma experimental omnidirecional para o selênio.....	96
Figura 5.21 – Semivariograma experimental omnidirecional para o cobre.....	96
Figura 5.22 – Semivariograma experimental omnidirecional para o chumbo.....	97
Figura 5.23 – Gráficos das correlações entre as variáveis originais e as componentes principais antes e após a rotação ortogonal (varimax).....	102
Figura 5.24 – Semivariogramas experimentais das componentes principais e os semivariogramas ajustados pelo modelo linear de co-regionalização....	103
Figura 5.25 – Semivariogramas experimentais cruzados das componentes principais e os semivariogramas ajustados pelo modelo linear de co-regionalização.104	

Figura 5.26 – Mapa de contornos da primeira componente principal.....	105
Figura 5.27 – Mapa de contornos da segunda componente principal.....	105
Figura 5.28 – Mapa de contornos da terceira componente principal.....	106
Figura 5.29 – Semivariogramas experimentais do arsênio, manganês e alumínio e semivariogramas ajustados pelo modelo linear de co-regionalização....	107
Figura 5.29 – (Cont.) Semivariogramas experimentais do ferro, chumbo e selênio e semivariogramas ajustados pelo modelo linear de co-regionalização...	108
Figura 5.29 – (Cont.) Semivariogramas experimentais do cádmio e cobre e semivariogramas ajustados pelo modelo linear de co-regionalização....	109
Figura 5.30 – Gráficos das correlações entre as componentes espaciais $Z_i^s(\mathbf{u})$ e os fatores regionalizados $Y_k^s(\mathbf{u})$	112
Figura 5.31 – Mapa de contorno do primeiro fator regionalizado associado com o modelo esférico em curta escala.....	114
Figura 5.32 – Mapa de contorno do segundo fator regionalizado associado com o modelo esférico em curta escala.....	114
Figura 5.33 – Mapa de contorno do terceiro fator regionalizado associado com o modelo esférico em curta escala.....	115
Figura 5.34 – Mapa de contorno do primeiro fator regionalizado associado com o modelo esférico em longa escala.....	115
Figura 5.35 – Mapa de contorno do segundo fator regionalizado associado com o modelo esférico em longa escala.....	116
Figura 5.36 – Mapa de contorno do terceiro fator regionalizado associado com o modelo esférico em longa escala.....	116
Figura 5.37 – Mapa de contorno da componente espacial estimada para o arsênio associada com o modelo esférico em curta escala.....	117
Figura 5.38 – Mapa de contorno da componente espacial estimada para o arsênio associada com o modelo esférico em longa escala.....	117
Figura 5.39 – Mapa de contorno da componente espacial estimada para o manganês associada com o modelo esférico em curta escala.....	118
Figura 5.40 – Mapa de contorno da componente espacial estimada para o manganês associada com o modelo esférico em longa escala.....	118

Figura 5.41 – Mapa de contorno da componente espacial estimada para o ferro associada com o modelo esférico em curta escala.....	119
Figura 5.42 – Mapa de contorno da componente espacial estimada para o ferro associada com o modelo esférico em longa escala.....	119
Figura 5.43 – Mapa de contorno da componente espacial estimada para o cádmio associada com o modelo esférico em curta escala.....	120
Figura 5.44 – Mapa de contorno da componente espacial estimada para o cádmio associada com o modelo esférico em longa escala.....	120
Figura 5.45 – Perda associada a uma classificação incorreta de uma localização como segura, para a função de perda L_1 , $w_1 = 1,0$, para o manganês....	123
Figura 5.46 – Perda associada a uma classificação incorreta de uma localização como contaminada, para a função de perda L_2 , $w_2 = 0,5$, para o manganês....	124
Figura 5.47 – Perda associada a uma classificação incorreta de uma localização como contaminada, para a função de perda L_2 , $w_2 = 2,5$, para o manganês.....	124
Figura 5.48 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , $w_1 = 1,0$ e L_2 , $w_2 = 0,5$, para o manganês.....	125
Figura 5.49 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , $w_1 = 1,0$ e L_2 , $w_2 = 2,5$, para o manganês.....	125
Figura 5.50 – Perda associada a uma classificação incorreta de uma localização como segura, para a função de perda L_1 , $w_1 = 1,0$, para o arsênio.....	126
Figura 5.51 – Perda associada a uma classificação incorreta de uma localização como contaminada, para a função de perda L_2 , $w_2 = 0,5$, para o arsênio.....	126
Figura 5.52 – Perda associada a uma classificação incorreta de uma localização como contaminada, para a função de perda L_2 , $w_2 = 2,5$, para o arsênio.....	127
Figura 5.53 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , $w_1 = 1,0$ e L_2 , $w_2 = 0,5$, para o arsênio.....	127
Figura 5.54 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , $w_1 = 1,0$ e L_2 , $w_2 = 2,5$, para o arsênio.....	128
Figura 5.55 – Perda associada a uma classificação incorreta de uma localização como segura, para a função de perda L_1 , $w_1 = 1,0$, para o ferro.....	128
Figura 5.56 – Perda associada a uma classificação incorreta de uma localização como contaminada, para a função de perda L_2 , $w_2 = 0,5$, para o ferro.....	129
Figura 5.57 – Perda associada a uma classificação incorreta de uma localização como contaminada, para a função de perda L_2 , $w_2 = 2,5$, para o ferro.....	129

Figura 5.58 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , $w_1 = 1,0$ e L_2 , $w_2 = 0,5$, para o ferro.....	130
Figura 5.59 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , $w_1 = 1,0$ e L_2 , $w_2 = 2,5$, para o ferro.....	130
Figura 5.60 – Perda associada a uma classificação incorreta de uma localização como segura, para a função de perda L_1 , $w_1 = 1,0$, para o cádmio.....	131
Figura 5.61 – Perda associada a uma classificação incorreta de uma localização como contaminada, para a função de perda L_2 , $w_2 = 0,5$, para o cádmio.....	131
Figura 5.62 – Perda associada a uma classificação incorreta de uma localização como contaminada, para a função de perda L_2 , $w_2 = 2,5$, para o cádmio.....	132
Figura 5.63 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , $w_1 = 1,0$ e L_2 , $w_2 = 0,5$, para o cádmio.....	132
Figura 5.64 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , $w_1 = 1,0$ e L_2 , $w_2 = 2,5$, para o cádmio.....	133
Figura 5.65 – Gráficos de dispersão da média versus variância para janelas móveis de $0,5 \times 0,5$ km sobre a área de estudo para as variáveis em estudo.....	135
Figura 5.66 – Semivariogramas relativos experimentais e modelos ajustados para o arsênio, manganês, ferro e cádmio.....	136
Figura 5.67 – Q-Q plot de 10 simulações versus o conjunto de referência (amostra) dos dados de arsênio, manganês, ferro e cádmio.....	137
Figura 5.68 – Reprodução dos semivariogramas de 20 simulações para arsênio, manganês, ferro e cádmio e os modelos ajustados.....	137
Figura 5.69 – Mapas de probabilidade das concentrações de As, Mn, Fe e Cd que excederam o limite máximo permitido para cada metal.....	141
Figura 5.70 – Classificação das localizações com base na probabilidade das concentrações acima dos limites máximos permitidos serem maiores do que as probabilidades marginais de contaminação para o As, Mn, Fe e Cd.....	142
Figura 5.71 – Mapas da estimativa E-type das concentrações de As, Mn, Fe e Cd, derivados do pós-processamento das simulações.....	145
Figura 5.72 – Mapas da estimativa Mediana das concentrações de As, Mn, Fe e Cd, derivados do pós-processamento das simulações.....	146
Figura 5.73 – Mapas de classificação das localizações como contaminadas por As, Mn, Fe e Cd, baseados na estimativa E-type estar acima dos limites máximos permitidos para cada metal.....	147

Figura 5.74 – Mapas de classificação das localizações como contaminadas por As, Mn, Fe e Cd baseados na estimativa Mediana estar acima dos limites máximos permitidos para cada metal.....	148
Figura 5.75 – Mapas da estimativa Quantil das concentrações de As, Mn, Fe e Cd, derivados do pós-processamento das simulações.....	149
Figura 5.76 – Mapas de classificação das localizações como contaminadas por As, Mn, Fe e Cd baseados na estimativa Quantil estar acima dos limites máximos permitidos para cada metal.....	150
Figura 5.77– Mapas de probabilidades de risco α para As, Mn, Fe e Cd.....	151
Figura 5.78– Mapas de probabilidades de risco β para As, Mn, Fe e Cd	152
Figura 5.79 – Mapas da média das distribuições de custos de saúde obtidos pela aplicação da função de perda (3.91) aos valores simulados para As, Mn, Fe e Cd.....	155
Figura 5.80 – Mapas da variância das distribuições de custos de saúde obtidos pela aplicação da função de perda (3.91) aos valores simulados para As, Mn, Fe e Cd.....	156
Figura 5.81 – Mapas das simulações que produziram o custo mínimo e máximo na aplicação da função de perda (3.91) aos valores simulados para o As e Mn.....	157
Figura 5.82 – Mapas das simulações que produziram o custo mínimo e máximo na aplicação da função de perda (3.91) aos valores simulados para o Fe e Cd.....	158
Figura 5.83 – Distribuições dos custos globais do risco de se declarar a área segura em relação à contaminação, obtidas pela aplicação da função de perda (3.91) às simulações do arsênio, manganês, ferro e cádmio.....	159
Figura 5.84 – Curvas individuais de custos para o As (nível de corte de 0,05 ppm), Mn (nível de corte de 0,1 ppm), Fe (nível de corte de 0,3 ppm) e Cd (nível de corte de 0,001 ppm).....	161
Figura 5.85 – Mapas da distribuição espacial dos custos por parcela do As (C_{trat} e $E\{C_{falha}\}$) para o nível de probabilidade de 0,098 e nível de corte de 0,05 ppm e do Mn (C_{trat} e $E\{C_{falha}\}$) para o nível de probabilidade de 0,28 e nível de corte de 0,1 ppm.....	162
Figura 5.86 – Mapas da distribuição espacial dos custos por parcela do Fe (C_{trat} e $E\{C_{falha}\}$) para o nível de probabilidade de 0,37 e nível de corte de 0,3 ppm e do Cd (C_{trat} e $E\{C_{falha}\}$) para o nível de probabilidade de 0,07 e nível de corte de 0,001 ppm.....	163

LISTA DE TABELAS

TABELA 3.1 – Balanço Hídrico – Região de Macapá/AP, para o período de 1968 a 1990.....	12
TABELA 5.1 – Valores máximos permitidos (em mg/L) para águas doces.....	82
TABELA 5.2 – Estatísticas descritivas para o conjunto de dados (mg/L).....	85
TABELA 5.2 (Cont.) – Estatísticas descritivas para o conjunto de dados (mg/L).....	86
TABELA 5.3 – Matriz de coeficientes de correlação de Pearson e Rank.....	87
TABELA 5.4 – Matriz de correlação das variáveis.....	99
TABELA 5.5 – Cargas dos fatores com a percentagem da variância explicada antes e depois da rotação ortogonal das três primeiras Componentes Principais.....	101
TABELA 5.6 – Cargas dos três primeiros Fatores Regionalizados com a percentagem da variância explicada na pequena (\mathbf{B}^1) e longa (\mathbf{B}^2) escala.....	111
TABELA 5.7. Parâmetros (adimensionais) utilizados na programação <i>annealing</i>	136
TABELA 5.8 – Estatísticas obtidas para os custos globais de saúde (valores relativos) obtidos através da aplicação da função de perda às simulações.....	159

LISTA DE FOTOGRAFIAS

Foto 4.1 – Vista parcial da Bacia de Rejeitos onde eram depositados os resíduos de manganês provenientes das usinas de pelletização/sinterização.....	68
Foto 4.2 – Uma das pilhas de manganês comercializável (coberto com lonas) na área industrial da ICOMI.....	69
Foto 4.3 – Pilhas de rejeitos de manganês cobertos com lonas dentro da área da ICOMI.....	69
Foto 4.4 – Uma rua do bairro da Hospitalidade – bairro comercial de Santana, próximo ao Igarapé Boiero, aterrada com rejeitos de manganês.....	70
Foto 4.5 – Poço amazonas comunitário utilizado para consumo de água por moradores da Vila do Elesbão, próximo à área industrial da ICOMI.....	72
Foto 4.6 – Igarapé do Elesbão, localizado próximo à área industrial da ICOMI.....	72
Foto 4.7 – Vista aérea da Vila do Elesbão em Santana, Amapá, localizada próximo à área industrial da ICOMI.....	73

RESUMO

Neste trabalho é apresentada uma metodologia baseada na teoria geoestatística para a quantificação do risco de contaminação por metais pesados na área portuária de Santana/AP, onde eram desenvolvidas atividades relacionadas ao beneficiamento e comercialização do minério de manganês oriundo das minas de Serra do Navio/AP. Foram considerados 16 metais e 8 deles apresentaram concentração acima do estabelecido na legislação ambiental. Manganês, arsênio, ferro e cádmio apresentaram indícios de vínculos com atividades antrópicas; alumínio, selênio, cobre e chumbo parecem ter sua origem no próprio meio natural. Na classificação das áreas contaminadas, as estimativas foram baseadas em diferentes critérios de otimização. Devido a forte assimetria apresentada pelas variáveis, a classificação baseada em estimativas do tipo média e mediana mostrou-se inadequada. Estimativas quantis baseadas em funções de perda assimétricas apresentaram melhores resultados. A utilização de funções de perda específicas permitiu a avaliação do impacto econômico sobre custos de remediação de áreas contaminadas e custos de saúde, de modo a minimizar a perda esperada na classificação das áreas contaminadas. Os maiores custos globais de saúde foram observados para o ferro, seguido pelo manganês, arsênio e cádmio. Um modelo de decisão econômico baseado em riscos foi usado para avaliar os resultados. Esses resultados mostraram que os riscos ótimos, relacionados aos custos mínimos de remediação, foram aproximadamente iguais à proporção acima do limite máximo tolerável observada nas amostras (probabilidades marginais) de cada respectivo metal analisado.

ABSTRACT

In this study a methodology based on geostatistical theory was used to help quantify the risk of heavy metal contamination in the area surrounding Santana/AP. In this region activities related to the processing and commercialization of manganese ore from the Serra do Navio mine were carried out. We utilized results of 16 different metals (8 having concentrations above the limits established by environmental legislation). Manganese, arsenic, iron and cadmium showed evidence that their concentrations were due to anthropogenic activities. In contrast, the high aluminum, selenium, copper and lead concentrations appear to be related to the natural environment. For the classification of the contaminated areas, estimates were based on different optimization criteria. Due to high asymmetry presented by the variables the classification based in estimates of the average and median showed inadequate. Quantitative estimates based on functions of asymmetric loss yielded the best results. The utilization of functions of specific loss permitted the evaluation of economic impact of remediation costs and health costs in a manner to minimize the expected loss in the classification of such contaminated areas. The greatest overall health costs were observed for iron, followed by manganese, arsenic and cadmium. An economic decision model based on risks was used to evaluate the results. The results show that the optimal risk, related to the minimum cost of remediation, went approximately like the proportion above of the tolerable maximum observed in the samples (marginal probabilities) in each respective metal analyzed

1. INTRODUÇÃO

Em 1972, com a Conferência de Estocolmo, o mundo começou a despertar para a questão ambiental de uma forma mais intensa e global. Atualmente o meio ambiente é um tema importante tanto na política governamental quanto na preocupação dos cidadãos comuns em muitos países, uma vez que cada vez mais cresce a consciência de que a degradação ambiental ameaça a saúde e o bem estar social.

Os problemas relativos aos contaminantes ambientais estão, de alguma forma, relacionados ao crescente processo de industrialização verificado desde o final do século retrasado, onde, ao lado do desenvolvimento da pesquisa e da difusão de novas tecnologias, os processos de produção e seus produtos têm contribuído para colocar em perigo ou causar prejuízos à saúde do homem e dos ecossistemas. Esses contaminantes ambientais decorrentes das atividades desenvolvidas pelo homem podem ser controlados tanto na probabilidade de ocorrência quanto nas conseqüências. Das agências regulamentadoras espera-se, por exemplo, que os riscos de aparecimento de câncer e de outras doenças sejam controlados ou reduzidos a níveis suficientemente baixos. Para que isto seja conseguido devem ser respondidas questões fundamentais como o significado de risco ambiental, risco à saúde, risco à segurança, sua importância, assim como os procedimentos existentes para identificar, quantificar, analisar e avaliar esses riscos.

A resolução ou minimização de problemas ambientais exige o desenvolvimento e aplicação de políticas ambientais claras e eficientes. Neste sentido, o processo de avaliação de risco ambiental assume um papel importante para a execução de uma política ambiental, uma vez que pode auxiliar na gestão de risco e na tomada de decisões em questões relacionadas à saúde, remediação, econômicas, entre outras.

A avaliação de risco é reconhecida como a principal ferramenta científica e de custo-benefício para o estudo do tratamento de áreas contaminadas. Os procedimentos de avaliação de risco são geralmente baseados no modelo fonte-caminho-receptor e consistem na caracterização do local, no comportamento ambiental e toxicidade dos contaminantes, no caminho potencial de entrada dos contaminantes nos receptores (seres vivos), na exposição dos receptores aos contaminantes e sua resposta à dose. Na caracterização do risco é importante a identificação das fontes primária e secundária. A fonte primária é a causa real da contaminação e relaciona-se à natureza e local da descarga, aos mecanismos de transporte e processos ambientais ocorridos, enquanto a

fonte secundária é o ambiente impactado ao qual o receptor é exposto. É necessário o desenvolvimento de novos métodos para a quantificação, redução das incertezas e minimização de custos de investigação devido ao grande número de incertezas presentes durante o processo de avaliação de risco. A presença de incertezas ocorre principalmente por causa da heterogeneidade do solo e natureza freqüentemente acidental dos processos de contaminação.

Áreas contaminadas podem ser definidas como locais onde ocorre a presença de substâncias tóxicas distribuídas de forma não controlada nos diferentes compartimentos do ambiente, as quais, em função do nível de concentração existente, determinam riscos potenciais à saúde dos seres vivos ou prejuízos à qualidade dos recursos naturais.

Nos países industrializados e em desenvolvimento o aumento de áreas contaminadas ao longo das últimas décadas tem se tornado um dos problemas mais relevantes na área ambiental. Muitos desses países têm adotado políticas específicas para o gerenciamento e controle dessas áreas em virtude do grande impacto que causam na saúde pública, sobre os recursos hídricos (águas superficiais e subterrâneas), o solo e o patrimônio. Em geral, essas políticas estabelecem legislações específicas e inventários sobre os locais suspeitos de contaminação, promovem o desenvolvimento de procedimentos de avaliação, de tecnologias de remediação e a criação de fundos para o tratamento das áreas consideradas prioritárias.

A degradação das águas superficiais e subterrâneas, por problemas de contaminação e poluição, vem agravando as dificuldades para seu aproveitamento e intensificando sua escassez. A poluição das águas subterrâneas é muito mais preocupante do que a das águas superficiais, visto que estas rapidamente se renovam e se recuperam após cessar o lançamento de efluentes. Nos aquíferos, devido à lenta circulação das águas subterrâneas, capacidade de adsorção dos terrenos e pequeno tamanho dos canalículos, uma contaminação pode levar muito tempo até manifestar-se claramente. O notável poder de depuração dos aquíferos, em relação a muitos contaminantes, e o grande volume de água que armazenam, fazem com que as contaminações extensas se manifestem muito lentamente e as contaminações localizadas somente apareçam depois de algum tempo, e mesmo assim quando deslocadas para captações em exploração. Em outras palavras, os aquíferos são muito menos vulneráveis à poluição do que as águas superficiais. Mas, uma vez produzida a contaminação, a recuperação, dependendo do tipo de contaminante, pode levar muitos anos e até mesmo tornar-se economicamente inviável. De um modo geral, pode-se dizer

que as águas subterrâneas podem sofrer contaminação direta, sem diluição, quando o poluente atinge diretamente o aquífero através de poços negros ou poços abandonados e/ou com deficiências construtivas, e contaminação indireta, com diluição, quando o poluente atinge o aquífero depois de passar por alterações a partir do ponto de origem. A contaminação dessas águas pode se dar a partir das áreas de disposição de resíduos sólidos (urbanos e industriais); lagoas de tratamento de efluentes industriais; disposição de esgotos; disposição de resíduos radioativos; atividades agrícolas; vazamento de petróleo e derivados e rejeitos de atividades minerais (FEITOSA e MANOEL F., 1997).

A liberação de substâncias químicas muitas vezes de toxicidade desconhecida aos seres humanos e aos ecossistemas é um dos resultados do crescimento industrial desordenado. Uma vez no meio ambiente essas substâncias químicas experimentam uma série de transformações físicas e químicas, inclusive a combinação de outros produtos químicos que afetam sua toxicidade. Entre as substâncias químicas, vários metais utilizados em atividades industriais e liberados no meio ambiente são considerados tóxicos para o ser humano, onde os mais importantes são: arsênio (As), mercúrio (Hg), chumbo (Pb), manganês (Mn), titânio (Ti), níquel (Ni), selênio (Se), antimônio (Sb), berílio (Be) e cobalto (Co). Por causas das reações químicas que esses elementos sofrem no meio ambiente, muitas vezes eles adquirem as propriedades de bioacumulação e biomagnificação na cadeia alimentar, persistência no ambiente e distúrbios nos processos metabólicos dos seres vivos. A bioacumulação é definida por D'AMATO *et alli* (2002) como a absorção do elemento (ou composto) pelo organismo do meio abiótico ou biótico, podendo ou não a concentração exceder a da fonte. Na biomagnificação, a concentração do elemento aumenta ao longo da cadeia alimentar. A biomagnificação ocorre quando as concentrações de um poluente nos tecidos de um organismo excedem as concentrações do nível trófico adjacente inferior em mais de 100%. As bioacumulações e biomagnificações se encarregam de transformar concentrações consideradas normais em concentrações tóxicas para diferentes espécies da biota e para o homem. A persistência garante os efeitos ao longo do tempo ou longo prazo, mesmo depois de interrompidas as emissões.

Há uma demanda de produção crescente por muitos metais nos países desenvolvidos e em desenvolvimento. Tal fato sugere um aumento das probabilidades superpostas e sua dispersão e contato com os diversos compartimentos ambientais. Um metal pode ser disperso desde o momento da extração de seu minério, durante o tempo que se constitui o produto utilizável, até depois que o produto é descartado, reciclado ou

destruído. Em outras palavras, o homem está transformando parte dos metais imobilizados nos depósitos em metais mobilizados no ambiente. Esta situação se agrava nos países em desenvolvimento, para onde são exportadas fábricas mais poluidoras, dentro da lógica da divisão internacional do trabalho. Para agravar ainda mais o problema, nestes países, os cuidados e o controle de poluição industrial quase sempre não são adotados, equipamentos de abatimento de emissões nem sempre são instalados ou operados corretamente e a fiscalização é precária.

Investigações e/ou cuidadosa monitoração da qualidade das águas subterrâneas em situações selecionadas de campo são necessárias para promover um melhor entendimento e focalizar a atenção em problemas potenciais de contaminação das águas subterrâneas.

O presente trabalho desenvolve-se com o objetivo de contribuir para a quantificação e mapeamento dos riscos ambientais de contaminação por metais pesados, através da utilização de uma metodologia baseada na teoria geoestatística. Nesse sentido, foi realizada uma caracterização da área de estudo para se descobrir a origem dos contaminantes. Para os contaminantes provenientes de atividades antrópicas foram determinados magnitude, extensão e classificação das áreas contaminadas. Foram ainda avaliados e mapeados os custos de saúde e remediação na área de estudo e apresentado um modelo para tomada de decisão baseado na minimização dos custos. A utilização de uma abordagem probabilística permitiu que as incertezas associadas aos resultados obtidos fossem conhecidas.

Os procedimentos de análise foram aplicados em uma área com fortes indícios de contaminação ambiental por metais pesados, decorrentes da intensa atividade de mineração realizada durante várias décadas, localizada no estado do Amapá, ao norte do Brasil. Foram utilizados os resultados obtidos de análises químicas de águas superficiais e subterrâneas para 16 metais.

2. REVISÃO BIBLIOGRÁFICA

2.1. Considerações gerais

Existem várias definições e abordagens conceituais sobre poluição, meio ambiente, impacto ambiental e risco. Tais conceitos serão apresentados a seguir baseados, entre outros, nos trabalhos de BRILHANTE e CALDAS (1999); HANAN e BATALHA (1995); IBAMA (1995) e CONAMA (1986).

Do ponto de vista ecológico, a poluição é definida como qualquer alteração da composição e das características do meio, afetando de forma prejudicial os ecossistemas, ou ainda, como uma interferência danosa nos processos de transmissão de energia. A contaminação é um caso particular de poluição, caracterizada pela introdução no meio ambiente (água, ar, solo ou alimentos) de substâncias tóxicas ou radioativas e/ou organismos patogênicos em concentrações nocivas à saúde e vida dos seres humanos.

O meio ambiente pode ser definido como um meio físico suscetível de alteração pela atividade humana. Neste caso, ele é restrito à porção do espaço do sistema Terra-Atmosfera, caracterizado pela presença de dois fluidos – o ar e a água – cujas propriedades permitem a distribuição ao nível do solo, da energia do sol e, conseqüentemente, da repartição dos climas, que regula a circulação atmosférica e o ciclo da água. Dentre os vários problemas apresentados atualmente ao meio ambiente encontra-se a acumulação de substâncias dificilmente biodegradáveis como os metais, pesticidas e outros compostos orgânicos, constituindo uma ameaça para a qualidade do solo e dos lençóis freáticos.

Para o Conselho Nacional de Meio Ambiente (CONAMA, 1986), impacto ambiental é definido como qualquer alteração das propriedades físicas, químicas e biológicas do meio ambiente, causadas por qualquer forma de matéria ou energia resultante das atividades humanas que, direta ou indiretamente, afetam: a saúde, a segurança e o bem-estar da população; as atividades sociais e econômicas; a biota; as condições estéticas e sanitárias do meio ambiente; a qualidade dos recursos ambientais. Os impactos ambientais podem ser diretos ou indiretos, produzir-se a curto ou a longo prazo, ser de curta ou longa duração, ser cumulativos, reversíveis ou não, ser inevitáveis, locais, regionais, continentais, globais, naturais e antropogênicos.

Nos campos da saúde, segurança e ambiental, o risco é geralmente identificado como a probabilidade que indivíduos (ou uma população) estejam sujeitos à incidência

de efeitos adversos, como, danos, doenças ou morte causada por um agente químico em indivíduo a este exposto. O grau do risco é função do efeito adverso que pode resultar de uma ação particular. O efeito de um impacto pode ser positivo ou negativo, enquanto o efeito de um risco é sempre negativo. Entre os tipos de risco existentes, pode-se citar o risco ambiental, o risco à saúde e o risco econômico. Para MILLARD e NERCHAL (2001, p. 735), o processo de modelamento da exposição a um produto tóxico e predição de efeitos à saúde ou ecológicos é denominado avaliação de risco. A avaliação de risco é um processo multidisciplinar que envolve estudos ecológicos, químicos, meteorológicos, estatísticos, biológicos, políticos, entre outros. O modelo básico freqüentemente usado para a avaliação de risco ecológico e saúde humana é:

$$\text{Risco} = \text{Dose} \times \text{Prob (Efeito por unidade de dose)} \quad (1.1)$$

Isto é, o risco de dano (o efeito) a um indivíduo é igual à quantidade de tóxico que o indivíduo absorve (a dose) vezes a probabilidade de o efeito ocorrer para uma simples unidade do tóxico. Se o efeito é alguma forma de câncer, o segundo termo do lado direito da equação (1.1) é chamado de *fator potencial de câncer (Cancer Slope Factor – CLF)*. Esse termo é estimado a partir de um modelo, chamado curva dose-resposta, que relaciona a probabilidade do efeito para a dose recebida e é desenvolvido a partir de experimentos controlados em laboratórios com animais ou outros organismos e/ou a partir de estudos epidemiológicos de populações humanas. O primeiro termo do lado direito da equação (1.1), a dose, é estimada pela identificação das fontes do produto tóxico e pela quantificação dessa concentração. Segundo HALLENBECK (1994, p.1) e U.S.EPA (1995), a avaliação de risco envolve três passos principais:

- Identificação do perigo: trata-se da identificação do agente perigoso na sua essência, seus efeitos, as condições de exposição e a população-alvo;
- Avaliação da dose-resposta: descreve os dados, modelos e extrapolações usados para desenvolver a curva de dose-resposta para animais de laboratório, humanos e/ou espécies selvagens.
- Avaliação da exposição: identifica as fontes de exposição ambiental da(s) população (ões) em estudo. Descreve os principais caminhos, padrões e magnitudes da exposição, ou seja, refere-se à quantificação da concentração do agente nocivo em um meio, para um indivíduo ou grupo.

Dentro do contexto da avaliação de risco o presente trabalho está inserido, principalmente, na avaliação da exposição.

Uma vez que esses passos são completados, a U.S.EPA (1995) distingue outros processos identificados como a **caracterização do risco**, que é um resumo da avaliação de risco e integra toda a informação fornecida pela avaliação de risco, incluindo a análise de incerteza e sensibilidade e uma discussão sobre incerteza versus variabilidade; o **gerenciamento do risco**, que consiste no processo de tomada e justificativa de decisões a partir das informações da caracterização do risco, políticas de regulamentação e análises econômicas e sociais; e, finalmente, **comunicação do risco**, que é a troca de informações com o público.

LACAYO JR. (1994, p. 881) mostra que os estimadores de risco utilizados usualmente em trabalhos ambientais não são os melhores e que é necessário considerar mais de um procedimento nas estimativas de riscos ambientais. A análise de risco oferece um instrumento metodológico para a quantificação – espacialmente diferenciada – de danos causados pela atividade antrópica aos fatores naturais. O aspecto da quantificação é considerado o ponto mais positivo da análise de risco, se comparada a outros métodos utilizados no planejamento ambiental (por exemplo, análise do valor de uso). GILBERT (1994, p. 867) apresenta entre os desafios no planejamento e implementação de recuperação ambiental, o desenvolvimento de procedimentos estatísticos para conjunto de dados incompletos (censurados), dados com distribuições assimétricas e correlacionados no espaço e/ou tempo e o desenvolvimento e integração da geoestatística, sistemas de informações geográficas, análise de incerteza de fontes de contaminação, transporte e modelos de risco.

2.2. Contaminação ambiental por metais pesados

Todas as substâncias lançadas em um corpo d'água serão integradas a este ou a seu sedimento de fundo. O comportamento delas na água tem sido uma preocupação constante de diversas áreas do conhecimento. Isto por causa da importância da utilização da água para o consumo humano.

Segundo TEIXEIRA e FYFE (1994), a expressão *metais pesados* é geralmente utilizada para se referir aos íons dos elementos metálicos, tais como cobre, zinco, ferro, cromo e alumínio. É bastante útil a conceituação de que os metais pesados podem ser definidos com base na Tabela Periódica dos Elementos. Outro critério que tem sido usado, apresentado em BURREL, *apud* TEIXEIRA e FYFE (1994), é o que define metais pesados em termos de densidade, onde os metais pesados devem incluir os

elementos químicos que se apresentam com densidade superior a 4,5. Neste caso, a maioria desses metais (e suas respectivas densidades) é: Ti(4,51), **Se**(4,82), **As**(5,73), V(6), Sb(6,68), **Cr**(6,90), **Zn**(6,92), **Mn**(7,20), **Fe**(7,86), **Cd**(8,65), **Co**(8), **Ni**(8,8), **Cu**(8,95), **Mo**(10,20), **Ag**(10,50), **Pb**(11,35), Ta(11,85), Hg(13,60) e Au(19,3). Os elementos com destaque em negrito mais o alumínio, bário e estrôncio foram estudados no presente trabalho. Então, de acordo com este critério, selênio e arsênio, sendo metalóides (com características tanto de metais quanto de não-metais), foram incluídos no grupo dos metais pesados. Entretanto, a maioria dos autores mostra que o conceito de metal pesado baseado na densidade é ineficaz, uma vez que os lantanídeos, os actinídeos, bem como o Y e o Ra, seriam considerados metais pesados.

TEIXEIRA e FYFE (1994) mostram que o maior obstáculo encontrado para se definir, claramente, metais pesados é a existência de uma grande diversidade de elementos com propriedades tóxicas ameaçadoras a uma multiplicidade de formas de vida. Por isso, em termos de qualidade do ambiente, a tendência atual é enquadrar todos os elementos tóxicos na categoria de metais pesados. Assim, os termos elementos tóxicos, metais tóxicos e metais pesados são, com frequência, empregados indistintamente, além do que, deve-se considerar o fato de que todo elemento químico é tóxico quando absorvido além da capacidade de assimilação do organismo receptor. Assim, para cada metal pesado existe uma faixa de tolerância dentro da qual os diferentes efeitos se manifestam. A tolerância é estimada por meio da concentração a partir da qual um certo efeito biológico é observado nos organismos. Há uma concentração máxima considerada como tolerância crítica, a partir da qual os efeitos sobre os organismos são considerados tóxicos. Provavelmente a mais importante distinção entre os metais pesados e os demais elementos químicos seja o fato de os primeiros não serem biodegradáveis. Isto equivale a afirmar que, quando o metal pesado, assim caracterizado, entra no ambiente, sua toxidez não pode ser controlada por fatores biológicos. Segundo este critério, arsênio e selênio, sendo metalóides (com características tanto de metais quanto de não metais), podem entrar no grupo dos metais pesados.

Um grande desafio nas pesquisas é provar se existe ou não relação entre a ocorrência (natural ou devido a atividades humanas) de certos metais pesados e alguns tipos de doenças. A maior dificuldade é saber como os metais interagem entre si e no meio ambiente, quais as reações patológicas sobre os seres vivos e como suas concentrações podem ser modificadas pela atividade antrópica. A resposta a essas questões não é simples, uma vez que é grande a complexidade envolvida na interação

entre os elementos (poluentes) e o meio receptor. Após sua emissão por uma fonte qualquer, os poluentes percorrem diversos caminhos, em sua difusão no ambiente, até chegarem ao solo, ar e/ou água. Seu nível de concentração em cada ponto do percurso dependerá de diversos fatores: taxa de emissão, características de dispersão (em função das propriedades do poluente e do meio) e taxa de remoção do meio por agentes físicos, químicos e/ou biológicos ao longo de todo o percurso. Além disso, raramente esses poluentes ocorrem sozinhos. Na maioria dos casos, ocorrem em combinação com uma variedade de outros contaminantes e em proporções diferentes. Essas combinações podem exercer efeitos favoráveis ou adversos sobre os organismos, efeitos esses que variam muito entre as espécies e são manifestados em termos de tolerância.

LICHT (2001) apresentou vários relatos de estudos associando a presença de metais pesados com a saúde humana e animal. Diversas doenças são conhecidas como provocadas pela ingestão desequilibrada de muitos elementos, dentre as quais estão o excesso de cádmio, mercúrio e chumbo e as deficiências em ferro e zinco, relacionadas, principalmente, à exposição anormal resultante de certas atividades industriais ou a deficiências alimentares. Correlações aparentes sem relacionamento de causa e efeito são muito numerosas e controvertidas. Nesta categoria estão a dureza da água com moléstias cardiovasculares, o chumbo com esclerose múltipla, o cádmio com a hipertensão e arteriosclerose, o alumínio com o mal de Alzheimer, e uma gama de diversos elementos e diversos tipos de câncer. Por exemplo, BOKONBAEV *et alli*. (1996, *apud* LICHT, 2001) mostraram o impacto provocado por atividades minerais e industriais na região do complexo mineiro de Kemim, República do Kirgistão. Nessa investigação, que incluiu terras aráveis e concentrações urbanas, foram encontrados valores elevados de Be, Pb, Zn, Mo, Sn, Th, Cr, Co, V e Cd. Especialmente elevados foram os teores de Pb, Be, Cu, Zn e Cd. Como grande parte dos metais encontram-se na forma móvel, foram encontradas concentrações elevadas de Pb, Ni, Y e Mo em culturas de batata, cevada e aveia. Investigações médicas provaram que as concentrações de chumbo no plasma sanguíneo de mulheres e crianças superavam em 2 a 5 vezes os teores normais. O nível de morbidade em crianças e adolescentes foi também elevado, com ocorrências de moléstias crônicas como a pielonefrite, hepatite e anemia, ficando determinada uma clara dependência da morbidade humana com um ambiente geoquímico desfavorável. Também WEBB (1975, *apud* LICHT, 2001) mostrou que a deficiência e/ou excesso de metais como o cobre, zinco, molibdênio, cobalto, selênio, ferro, manganês e cromo podem levar ao fracasso do plantio ou à morte de animais.

Ainda em LICH (2001) encontra-se relatada a associação do câncer de pulmão em adultos masculinos na região de Córdoba, Argentina, com o arsênio nas águas, e os particulados aéreos ricos em arsênio provenientes de operações de indústrias químicas em Baltimore, USA, e de fundições de cobre na Suécia setentrional.

Diversos estudos têm mostrado que o arsênio inorgânico pode aumentar o risco de câncer no pulmão, câncer na pele, câncer na bexiga, câncer no fígado, câncer no rim e câncer na próstata. A Organização Mundial de Saúde (OMS, 1998) e a Agência de Proteção Ambiental dos Estados Unidos (U.S.EPA, 2000) determinaram que o arsênio inorgânico é um carcinogênico humano. O cromo VI (produzido por processos industriais) também é classificado nesse mesmo grupo, carcinógeno humano, pela OMS e Agência Para Substâncias Tóxicas e Registros de Doenças (ATSDR, 2001). Segundo a classificação do peso da evidência para carcinogenicidade, pertencem ao grupo D (não classificado como carcinógeno humano) a prata, o alumínio, o bário, o cádmio, o cobre, o selênio, o estrôncio e o zinco. O cobalto é classificado no grupo C (possível carcinógeno humano) e o chumbo no grupo B2, que indica evidência suficiente em animais e inadequada ou não evidência em humanos (U.S. EPA, 2000). O selênio sulfídico, o cádmio e compostos de cádmio, o níquel e alguns de seus compostos podem ser razoavelmente prevenidos como carcinógenos (ATSDR, 1997 e 1999).

As atividades de planejamento do uso e ocupação do solo urbano deveriam estar sempre subsidiadas por uma grande quantidade de informações acerca das características do meio físico. A legislação brasileira (por exemplo, a Resolução nº 20 do Conselho Nacional de Meio Ambiente, de 18/06/86) que define padrões de qualidade da água para consumo humano, estabelece limites para a presença e concentração de microorganismos e substâncias tóxicas utilizadas na agricultura e hidrocarbonetos. Segundo essa resolução, águas da Classe Especial destinam-se ao abastecimento doméstico, sem prévia ou com simples desinfecção, e águas da Classe 1 destinam-se ao abastecimento doméstico após tratamento simplificado. A água fornecida pelas empresas de tratamento e abastecimento público segue um controle rígido, no entanto, no que diz respeito às águas subterrâneas, a legislação raramente é seguida. A água de poços artesianos raramente recebe uma caracterização completa, conforme determina a legislação para uma água de Classe Especial. Isso ocorre por vários motivos, entre os quais a escassez de recursos laboratoriais capacitados, custos analíticos elevados se forem seguidas as especificações legais, e a crença popular de que a água subterrânea é totalmente isenta de impurezas e contaminantes.

3. MATERIAIS E MÉTODOS

3.1. Levantamento amostral

3.1.1. Planejamento de execução da Amostragem

As informações de temperatura e precipitação entre 1968 e 1990, foram obtidas de acordo com os dados climáticos da estação meteorológica de Macapá/AP, resultando na elaboração do Balanço Hídrico apresentado na Tabela 3.1 e Figura 3.1 (RELATÓRIO JPE,1998). Com base nesses resultados, o período mais apropriado para a coleta de amostras é entre os meses de julho a dezembro, período de déficit hídrico na região. A amostragem foi realizada nos dias 08, 09 e 14 de dezembro de 2001.

Devido principalmente à necessidade de recursos financeiros, apoio logístico e questões políticas locais, a amostragem não atendeu totalmente ao planejamento inicial, onde se previu levantar pelo menos 80 amostras de água e solo. No levantamento de campo foi possível realizar somente 42 amostras de água, sendo 26 amostras de poços amazonas (cacimbas), 10 amostras de poços de monitoramento, dentro da área industrial da ICOMI, e 06 amostras de águas superficiais de igarapés e rio. Além disso, foram ainda utilizados os resultados de 08 amostras de águas de poços de monitoramento da ICOMI obtidas do Relatório de Monitoramento Ambiental da área de Santana, realizado em setembro de 2001, pela AMPLA Engenharia. A Figura 3.2 mostra a área de amostragem com os respectivos pontos de coleta. Na coleta das amostras foi utilizado um amostrador de esfera apropriado, 90 frascos de polietileno (PET) de 1000 ml, sendo 45 para amostras de água utilizadas na análise de arsênio e 45 para as amostras de água utilizadas nas análises de manganês e ferro, uma vez que a preservação das amostras deveria ser feita com ácido clorídrico ou nítrico, dependendo do tipo de metal a ser analisado (KEITH, 1996, p. 274; SOUZA,1977, p. 3). Para a determinação da localização dos pontos de amostragem foi utilizado um Sistema de Posicionamento Global (GPS) do tipo *Magellan 2000*, série 24-600006-0000, do Curso de Pós-Graduação em Geociências e Meio Ambiente da UNESP–Rio Claro/SP.

3.1.2. Procedimentos adotados na Amostragem

De acordo com orientações do químico responsável pela análise das amostras e as referências consultadas (FEITOSA e FILHO, 1997, p.81; KEITH, 1996, p.377; SOUZA,1977, p.3), os procedimentos adotados na coleta das amostras foram os seguintes:

- Utilização de um marco estável (cota) na boca de cada poço;
- Estipulação de uma amostra de referência no campo;
- Lavagem dos frascos de polietileno e pipeta com ácido nítrico 1:1, sendo enxaguados com água deionizada (NBR 9898 da ABNT, 1987);
- Adição de ácido clorídrico (e nítrico) para preservação das amostras;
- Remoção da água estagnada do poço antes da coleta;
- Identificação e caracterização dos frascos contendo as amostras coletadas;
- Registro das informações de campo nas fichas de referência das amostras;
- Armazenamento das amostras em local fresco e escuro.

TABELA 3.1 – Balanço Hídrico – Região de Macapá/AP, para o período de 1968 a 1990 (RELATÓRIO JPE, 1998)

Mês	T °C	P Mm	ETP (th) mm	ETR mm	P – ETP(th) Mm	ALT mm	DEF mm	EXC mm
JAN	26	299.6	133.1	133.1	166.5	98	0	68.5
FEV	25.7	347	112.8	112.8	234.2	0	0	234.2
MAR	25.7	407.2	124.8	124.8	282.4	0	0	282.4
ABR	25.9	384.3	125.2	125.2	259.1	0	0	259.1
MAI	26.1	351.5	134.2	134.2	217.3	0	0	217.3
JUN	26.2	220.1	131.3	131.3	88.8	0	0	88.8
JUL	26.1	184.8	134.2	134.2	50.6	0	0	50.6
AGO	26.8	98	149.8	139	-51.8	-41	10.8	0
SET	27.5	42.6	157.6	83.6	-115	-41	74	0
OUT	27	35.5	151.8	47.5	-116.3	-12	104.3	0
NOV	27.7	58.4	161.6	62.4	-103.2	-4	99.2	0
DEZ	27	142.5	151.8	142.5	-9.3	0	9.3	0
ANUAL	26.5	2571.5	1668.2	1370.6	903.3	0.0	297.6	1200.9

NOTAS:

T : Temperatura média mensal do ar ambiente

P : Precipitação Pluviométrica Total média mensal

ETP(th): Evapotranspiração Potencial (corrigida conforme o método de Thornthwaite)

ETR : Evapotranspiração Real

ALT : Alteração

DEF : Déficit

EXC : Excesso.

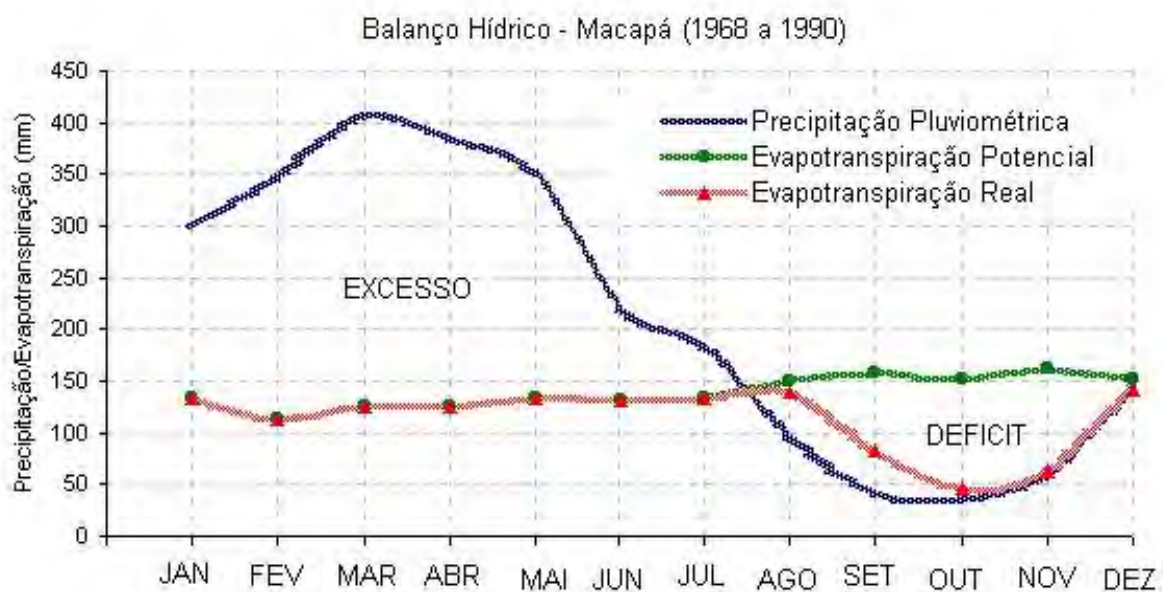


Figura 3.1 – Balanço Hídrico da Região de Macapá –AP, para o período de 1968 a 1990 (RELATÓRIO JPE, 1998).

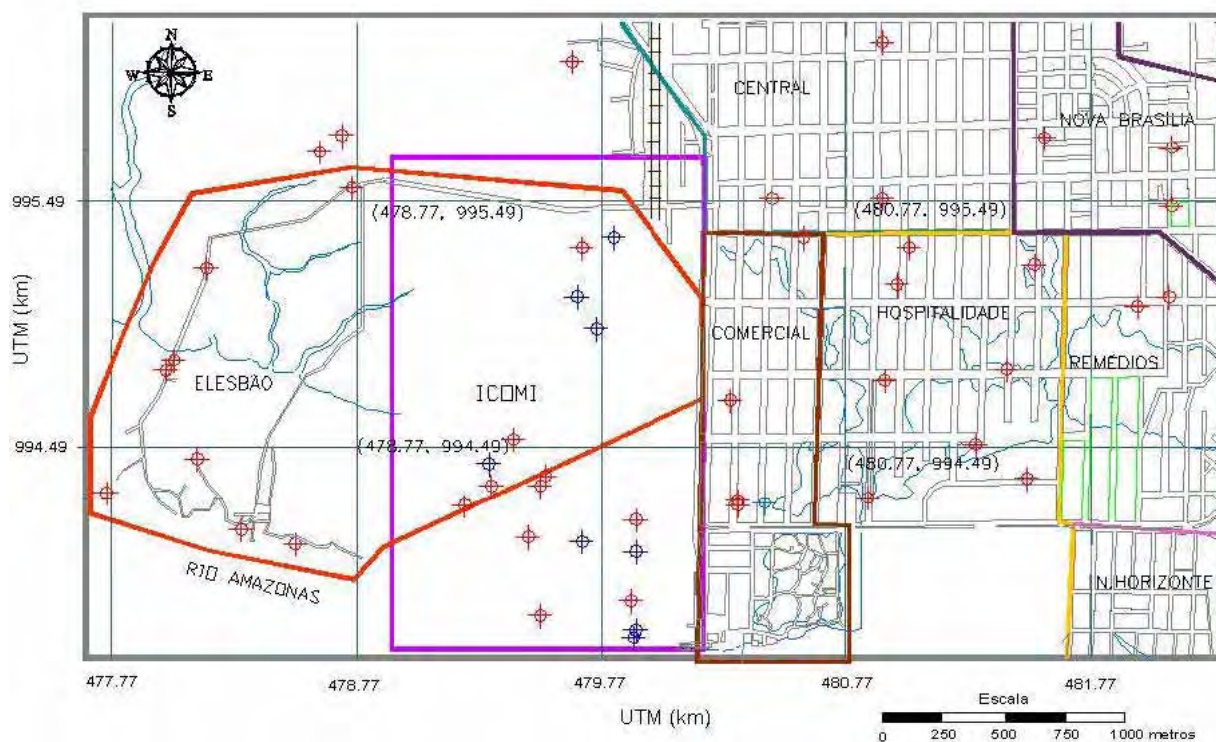


Figura 3.2 – Área de amostragem em Santana com os respectivos pontos de amostragem. As linhas em azul representam a rede de drenagem e pontos de amostragem em azul referem-se a amostras cedidas pela ICOMI.

3.2. Análise química dos dados

As análises químicas dos dados foram realizadas pelo Laboratório Central da Eletronorte-LACEN-, em Belém, Pará, onde foi utilizada a técnica de espectrometria de emissão atômica por plasma induzido. A espectrometria se baseia no princípio de que os átomos de um determinado elemento químico absorvem ou emitem energia eletromagnética dentro de um espectro bem definido, ou seja, os elementos químicos podem ser identificados pelo espectro eletromagnético de emissão ou absorção. Quando os átomos são excitados, alguns elétrons absorvem energia e saltam para um nível mais energético; quando a fonte de excitação é removida, os elétrons emitem energia e voltam ao nível energético anterior. Conhecendo-se o espectro de emissão ou absorção e medindo-se a intensidade de energia absorvida ou emitida, identificamos e quantificamos os elementos químicos presentes em uma amostra.

A técnica de espectrometria de emissão atômica por plasma induzido é utilizada para determinação multielementar simultânea. A amostra a ser analisada passa por um nebulizador e o aerosol produzido é injetado para uma tocha na temperatura de 10.000°C (região do plasma), onde ocorre a absorção de energia, excitando os átomos dos materiais presentes na amostra. Ao saírem da região do plasma, os átomos excitados emitem energia e voltam ao estado fundamental. A energia emitida é captada por um detector de estado sólido, que difere simultaneamente os sinais de energia para cada elemento químico, onde um microprocessador recebe os sinais e calcula a concentração, que é proporcional a energia emitida pelo elemento químico analisado.

O equipamento utilizado foi um Espectrofotômetro de Plasma Indutivamente acoplado **Varian** modelo *VISTA-PRO*. Os limites de detecção do equipamento e faixa linear de trabalho para os metais estão em anexo. As análises foram realizadas segundo normas e padrões utilizados pela Agência de Proteção Ambiental dos Estados Unidos (U.S.EPA, 1995).

3.3. A Geoestatística

3.3.1. Introdução

O termo *Geoestatística* teve origem, no início dos anos 60, com o engenheiro francês Georges Matheron, no Centre de Morphologie Mathématique, em Fontainebleau, França. O termo designa a metodologia desenvolvida com o objetivo de

estudar o comportamento das chamadas variáveis regionalizadas, utilizadas para o estudo da distribuição espacial de fenômenos naturais. O desenvolvimento dessa teoria por Matheron foi baseado nos trabalhos empíricos para estimativas de reservas minerais realizados pelo engenheiro de minas D.G. Krige e o estatístico H.S. Sichel, realizados na África do Sul, nos anos 50 (LANDIM, 1998, p.139).

É apresentada a seguir uma avaliação e perspectivas da Geoestatística sob o prisma da África do Sul nesses 50 anos, feita pelo próprio Krige, no 6^o Congresso Internacional de Geoestatística realizado em abril de 2000, na África do Sul, onde foram destacados os trabalhos de Matheron e Sichel, assim como a extraordinária velocidade na difusão e uso de computadores e programas geoestatísticos como maior fator para o rápido desenvolvimento da Geoestatística. Nos anos 50, seguindo os trabalhos de Sichel (SICHEL, 1947, 1952 *apud* KRIGE, 2000) sobre o modelo lognormal, tiveram implicações de longo alcance os trabalhos publicados por Krige (KRIGE, 1951, 1952 *apud* KRIGE, 2000). Esses trabalhos foram publicados em francês em 1955, quando Matheron iniciou o desenvolvimento da teoria de *Variáveis Regionalizadas*. Foi neste período que se usou, pela primeira vez, estimativas de regressão para blocos de reserva mineral. Nos anos 60, Matheron tornou essa técnica conhecida com o nome de *krigagem* e estabeleceu o primeiro modelo geoestatístico global para estimação de recuperação de reservas e muitos conceitos fundamentais em geoestatística, válidos até hoje. A partir dos anos 60, o conhecimento geoestatístico foi disseminado no mundo seguindo a formação da Escola de Fontainebleau na França e pelo trabalho de 4 de seus excelentes estudantes: Huijbrechts, David, Journel e Marechal, na Europa e mais tarde nos Estados Unidos e Canadá. Na década de 70, ocorreu grande crescimento não só da comunidade geoestatística como também de novos modelos como krigagem disjuntiva (MATHERON, 1975 *apud* KRIGE, 2000), funções de transferência (MATHERON, 1975^a *apud* KRIGE, 2000), krigagem universal (HUIJBRECHTS e MATHERON, 1977 *apud* KRIGE, 2000) e simulação condicional (JOURNEL, 1974 *apud* KRIGE, 2000). Em 1975 aconteceu o primeiro Congresso Internacional de Geoestatística em Roma. Dois Congressos Internacionais de Geoestatística ocorreram nos anos 80 (Tahoe e Avignon), período de rápida expansão da disponibilidade e velocidade de computadores e de programas adequados para krigagem, estimação variográfica e várias novas técnicas como a krigagem indicativa (JOURNEL, 1982 *apud* KRIGE, 2000), técnicas de simulação, krigagem disjuntiva isofatorial, entre outras. A década de 90 foi o período em que várias idéias foram introduzidas na África do Sul, como o uso de dados “soft”

(medidas indiretas da variável estimada/simulada) na aproximação Bayesiana (KRIGE *et alli.*, 1990 *apud* KRIGE,2000), co-krigagem de áreas virgens usando dados limitados de furos de sondagem e dados adjacentes regularizados de outras áreas exploradas (KRIGE, 1992a *apud* KRIGE,2000), técnicas de Componentes Principais para estimativas de estruturas espaciais anisotrópicas (KRIGE, 1999c *apud* KRIGE, 2000) e estudos práticos dos efeitos e implicações de tendências condicionais em estimativas de bloco. Fora da África do Sul ocorreu uma significativa expansão das técnicas geoestatísticas na área de petróleo e estudos ambientais (KRIGE,2000).

Uma marca registrada da geoestatística é sua praticidade, o que explica seu sucesso e aplicação em diversas áreas como: mineração (JOURNAL e HUIJBREGTS, 1978; RENDU, 1978), petróleo (DEUSTCH, 2002; SHRIVASTAVA, 1990), ciências do solo (GOOVAERTS, 1999; RIBEIRO JR.,1995), geotecnia (CROMER *et alli.*,1996; STURARO,1994), hidrogeologia (KITANIDIS, 1997), ecologia (ASSAD *et alli.*, 1999), e ciências ambientais (GOOVAERTS,1997; JOURNAL,1987).

3.3.2. Geoestatística aplicada ao meio ambiente

Os dados ambientais tais como concentrações de poluente são gerados por uma série de processos físicos (emissão, transporte, dispersão) que dão padrões nas distribuições de espaço/tempo desses dados. Raramente se pode deixar de considerar as correlações no espaço e tempo em dados ambientais. A descoberta e modelamento destes padrões de correlações espaciais permitem um melhor entendimento dos processos físicos subjacentes, além de auxiliar na interpolação espacial. Uma das características de aplicações ambientais é a escassez de dados detalhados, precisos e “brutos”. Essa escassez é freqüente e parcialmente compensada pela utilização de dados “soft”. Por isso, as ferramentas estatísticas utilizadas devem ser de natureza espacial, isto é, considerar e valorizar os padrões auto-espaciais e correlação cruzada entre variáveis, ser flexíveis para a escassez de dados, permitir tipos diferentes de dados, privilegiando os melhores dados. Em geral, deseja-se estimar a concentração de valores elevados de determinado elemento – e avaliar sua conectividade no espaço ou tempo – em vez dos valores médios espaciais e, por isso, geralmente as incertezas geradas precisam ser documentadas de modo quantitativo para ajudar na tomada de decisão.

A caracterização e avaliação de risco envolvem um grande número de incertezas, principalmente devido a heterogeneidade do solo e a natureza freqüentemente acidental

do processo de contaminação. A eliminação da(s) incerteza(s) não é possível, sendo necessárias análises de custo-benefício para estimar o risco e seleção de medidas apropriadas de remediação. A necessidade de quantificar e reduzir as incertezas e minimizar os custos de investigação tem encorajado fortemente o uso da geoestatística em estudos ambientais (FERGUSON, 1998; FERGUSON *et alli.*, 1998, *apud* CARLON *et alli.*, 2001).

Embora existam diversos trabalhos onde a geoestatística foi aplicada em problemas ambientais, na caracterização espacial (ou temporal) de fenômenos, sua prática ainda não pode ser considerada comum nessa área (CROMER, 1996). Entre os trabalhos relacionados à contaminação por metais pesados e outros contaminantes destacam-se os de CHANG *et alli.* (1999), que mapearam o conteúdo de arsênio no solo em Taiwan, usando a krigagem ordinária, cujos resultados mostraram contaminação do solo por arsênio a sudoeste de Taiwan. Esses resultados serão utilizados em pesquisas futuras para determinar a distribuição geográfica de padrões regionais de conteúdo de As em plantas e água subterrânea e a relação entre As e material de origem do solo, além da correlação com a doença do “pé-preto”. JUANG e LEE (1998) utilizaram a krigagem indicativa simples para estimar a probabilidade de delimitação incorreta de áreas perigosas em um local contaminado por metais pesados no solo (Cádmio e Chumbo), na cidade de Taoyuan, em Taiwan. GOOVAERTS (1997) apresenta em seu livro o uso de várias ferramentas geoestatísticas entre as quais a krigagem ordinária, krigagem simples, krigagem indicativa, cokrigagem e simulações condicionais na avaliação da contaminação do solo por metais pesados (Cd, Cu, Pb, Co, Cr, Ni e Zn), utilizando dados coletados pelo Instituto Federal de Tecnologia da Suíça, em Lausanne. ISTOK e RAUTMAN (1996) realizaram uma avaliação probabilística de contaminação de água subterrânea, onde foram definidas a extensão e severidade da contaminação de nitrato e Dacthal (um herbicida) em zonas saturadas e não-saturadas em uma área com 150 Km², próxima a Ontario, Oregon. Foi utilizada a simulação condicional sequencial Gaussiana para gerar os mapas probabilísticos. Os resultados mostraram que a caracterização da área na determinação da extensão e magnitude da contaminação dependerá do nível de incerteza que será tolerado na tomada de decisão. LEONTE e SCHOFIELD (1996) avaliaram a contaminação do solo por chumbo e critérios de limpeza em uma área escolhida para desenvolvimento residencial na Austrália, através do uso da krigagem indicativa no mapeamento da concentração de chumbo acima de níveis permitidos. JOURNEL (1987) apresentou os resultados mais evidentes de um

estudo de caso de poluição de chumbo no solo em Dallas, USA, onde foi utilizada a krigagem probabilística nas análises.

3.3.3. Conceitos básicos de geoestatística

Os conceitos apresentados a seguir foram baseados em livros textos de geoestatística como JOURNAL (1987, 1989), ISAAKS E SHIVASTAVA (1989), DEUSTCH e JOURNAL (1996), GOOVAERTS (1997) e DEUSTCH (2002). Como a maioria das análises geoestatísticas foi realizada utilizando-se os programas do “Geostatistical Software Library and User’s Guide” – GSLIB (DEUSTCH e JOURNAL, 1996) será usada neste trabalho, tanto quanto possível, a notação do guia do GSLIB.

A geoestatística trata com questões simples, mas importantes, de interpolação espacial e quantificação da incerteza para variáveis que apresentam continuidade espacial, ou seja, que podem ser medidas em qualquer ponto da área/região/zona em estudo. Ela está baseada em conceitos estatísticos tradicionais como:

- Variável Aleatória – é uma variável que pode apresentar uma série de resultados de acordo com alguma distribuição de probabilidades [frequências] ($p_i, i = 1, \dots, N$). As N probabilidades de ocorrências devem satisfazer as condições (1) $p_i \geq 0$, para todo $i = 1, \dots, N$ e (2) $\sum_{i=1}^N p_i = 1$. A variável aleatória (VA) é representada por uma letra maiúscula (Z) e os possíveis resultados pela letra minúscula correspondente, $z_i, i = 1, \dots, N$. Quando o número de resultados ou ocorrências é finito, a variável aleatória é denominada “discreta” ou categórica. Se o número de resultados possíveis é infinito, a variável aleatória é do tipo “contínua”. São exemplos de variáveis categóricas as propriedades geológicas como tipos de rochas, uso da terra, ou contagens de insetos, ou espécies de fósseis. Variáveis contínuas incluem propriedades petrofísicas como porosidade, permeabilidade, concentrações de metais ou poluentes. Em geral, na geoestatística, a VA está relacionada a alguma localização no espaço, e é representada pela notação $Z(\mathbf{u})$, onde \mathbf{u} é o vetor de coordenadas da localização. Vetores serão representados em negrito. A VA $Z(\mathbf{u})$ vista como uma função de \mathbf{u} é chamada uma função aleatória (FA), indicando um conjunto de VA’s definidos sobre algum campo de interesse, por exemplo, $\{Z(\mathbf{u}), \mathbf{u} \in \text{área de estudo}\}$. A VA $Z(\mathbf{u})$ também é dependente da informação no sentido que sua distribuição de

probabilidade muda conforme mais dados sobre o valor não amostrado $z(\mathbf{u})$ tornam-se disponíveis.

A distribuição de probabilidades de uma VA $Z(\mathbf{u})$ contínua pode ser totalmente caracterizada pela função de distribuição acumulada (FDA) definida como:

$$F(\mathbf{u};z) = Prob\{Z(\mathbf{u}) \leq z\}. \quad (3.1)$$

A equação (3.1) fornece a probabilidade que a VA Z seja menor ou igual a um determinado valor z , geralmente denominado valor de corte. A FDA apresenta as seguintes propriedades:

- a) $F(\mathbf{u};z)$ é não decrescente;
- b) $F(\mathbf{u};z) \in [0,1]$;
- c) $F(-\infty) = 0$ e $F(\infty) = 1$.

A probabilidade de Z ocorrer em um intervalo $[a,b]$, onde $b > a$, é dada pela diferença dos valores da FDA nos pontos a e b :

$$Prob\{Z(\mathbf{u}) \in [a,b]\} = F(b) - F(a). \quad (3.2)$$

Similarmente, a probabilidade de Z exceder um valor de corte z é escrita como:

$$Prob\{Z(\mathbf{u}) > z\} = 1 - F(\mathbf{u};z). \quad (3.3)$$

A função densidade de probabilidade (FDP) é a derivada da FDA, se esta for diferenciável, ou seja:

$$f(\mathbf{u};z) = F'(\mathbf{u};z) = \lim_{dz \rightarrow 0} \frac{F(\mathbf{u};z + dz) - F(\mathbf{u};z)}{dz}. \quad (3.4)$$

Quando a FDA é realizada para um conjunto específico de informações, por exemplo, (n) consistindo de n valores de dados vizinhos $Z(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha)$, $\alpha = 1, \dots, n$, a notação “condicional à n ” será utilizada, definindo a função de distribuição acumulada condicional (FDAC):

$$F(\mathbf{u};z|(n)) = Prob\{Z(\mathbf{u}) \leq z|(n)\}. \quad (3.5)$$

Uma notação similar é utilizada no caso da VA $Z(\mathbf{u})$ ser do tipo categórica tomada para qualquer valor K de resultados $k = 1, \dots, K$:

$$F(\mathbf{u};k|(n)) = Prob\{Z(\mathbf{u}) = k|(n)\}. \quad (3.6)$$

É importante observar que a FDAC (3.5) é uma função da localização \mathbf{u} , do tamanho da amostra (n), da configuração geométrica (a localização dos dados em \mathbf{u}_α , $\alpha = 1, \dots, n$) e dos valores amostrais (os n valores $z(\mathbf{u}_\alpha)$'s).

- Quantis – o quantil- p de uma distribuição $F(\mathbf{u}; z_p)$ é o valor z_p para o qual $F(\mathbf{u}; z_p) = \text{Prob}\{Z(\mathbf{u}) \leq z_p\} = p \in [0,1]$, ou seja, é o valor z_p da distribuição abaixo do qual estão uma proporção p de dados. Portanto, o quantil- p pode ser expresso em uma forma inversa da FDA, ou seja:

$$\text{O quantil-}p \ z_p = F^{-1}(\mathbf{u}; p) \text{ com } p \in [0,1]. \quad (3.7)$$

Os quantis mais comumente usados são:

- a) A mediana (M), equivalente ao quantil $z_{0.5}$

$$M = F^{-1}(\mathbf{u}; 0.5). \quad (3.8)$$

- b) Os quartis inferior e superior:

$$z_{0.25} = F^{-1}(\mathbf{u}; 0.25) \quad \text{e} \quad z_{0.75} = F^{-1}(\mathbf{u}; 0.75).$$

Uma forma de comparar duas distribuições é através de uma ferramenta gráfica conhecida como Q-Q plot, que consiste em cruzar graficamente os quantis das duas distribuições. Quando todos os pontos em um Q-Q plot caem sobre a linha de 45° , desenhada a partir da origem, as duas distribuições são exatamente iguais. Deslocamentos em relação à linha de 45° indicam diferenças nas duas distribuições, em particular:

1. Um deslocamento sistemático acima ou abaixo da linha de 45° indica que os *centros* ou médias das distribuições são diferentes. Um deslocamento acima da distribuição implica que os valores da distribuição Y são mais elevados do que os de X .
 2. Uma inclinação diferente de 45° indica que as dispersões ou variâncias das duas distribuições são diferentes. Uma inclinação maior do que 1 (ou 45°) indica que a variância de Y é maior do que a de X .
 3. Curvatura sobre o Q-Q plot indica que as duas distribuições têm formas diferentes.
- Valor esperado – é a média ponderada dos n possíveis resultados, onde cada resultado é ponderado por sua probabilidade de ocorrência:

– no caso discreto:

$$E\{Z\} = m = \sum_{i=1}^N p_i z_i = \frac{1}{N} \sum_{i=1}^N z_i. \quad (3.9)$$

– no caso contínuo, sob condições de existência das integrais:

$$E\{Z\} = m = \int_{-\infty}^{+\infty} z.dF(\mathbf{u}; z) = \int_{-\infty}^{+\infty} z.f(\mathbf{u}; z)dz . \quad (3.10)$$

$$\approx \sum_{k=1}^K z'_k [F(\mathbf{u}; z_{k+1}) - F(\mathbf{u}; z_k)] \quad \text{com } z'_k \in]z_k, z_{k+1}],$$

onde $F(\mathbf{u}; z)$ e $f(\mathbf{u}; z)$ são a FDA e FDP, respectivamente. A integral $\int_{-\infty}^{+\infty} z.dF(\mathbf{u}; z)$ foi aproximada por K classes com respectivas freqüências $[F(\mathbf{u}; z_{k+1}) - F(\mathbf{u}; z_k)]$ e z'_k é um valor dentro da k -ésima classe, por exemplo, o centro da classe.

A variância da VA Z é definida como o quadrado do desvio esperado de Z em relação à sua média:

$$\text{Var}\{Z\} = \sigma^2 = E\{[Z - m]^2\} = E[Z^2] - m^2 . \quad (3.11)$$

$$= \sum_{i=1}^N p_i (z_i - m)^2 , \text{ no caso discreto.}$$

$$= \int_{-\infty}^{+\infty} (z - m)^2 dF(\mathbf{u}; z) = \int_{-\infty}^{+\infty} (z - m)^2 f(\mathbf{u}; z)dz , \text{ no caso contínuo.}$$

A raiz quadrada da variância, σ , é chamada de desvio-padrão, e sua razão em relação à média, σ/m , para variáveis não negativas, é denominada coeficiente de variação (ou desvio-padrão relativo) o qual é adimensional. Coeficientes variação maiores do que 1 podem indicar a presença de valores altos (outliers) na distribuição.

Uma medida para avaliar a assimetria da distribuição é o coeficiente de assimetria, definido por:

$$\varphi = \frac{1}{n} \sum_{\alpha=1}^n \frac{[(z(\alpha) - m]^3}{\sigma^3} . \quad (3.12)$$

Distribuições simétricas apresentam $\varphi = 0$. Distribuições com assimetria positiva apresentam uma longa cauda devido a altos valores e φ é positivo e, por outro lado, distribuições com assimetria negativa apresentam uma longa cauda devido a valores baixos e φ é negativo. A média é chamada de primeiro momento. O segundo momento centrado na média é a variância, o terceiro momento é a assimetria, o quarto é a curtose.

A Curtose (K) mede o grau de achatamento da distribuição em relação à normal. Distribuições normais apresentam $K=3$ (mesocúrticas). Distribuições mais achatadas do que a normal (platicúrticas) têm valores de K menores do que 3 e distribuições mais

afiladas em relação à normal apresentam valores de K maiores do que 3. A curtose é calculada por:

$$K = \frac{1}{n} \sum_{\alpha=1}^n \frac{[(z(\alpha) - m)]^4}{\sigma^4}. \quad (3.13)$$

3.3.4. Descrição de padrões espaciais

- Distribuição bivariada

Em ciências da Terra é frequentemente importante conhecer o padrão de dependência de uma variável X em relação a outra Y , por exemplo, quando se deseja saber a relação entre pares de concentrações de metais medidas na mesma localização. A distribuição conjunta de resultados de um par de variáveis aleatórias X e Y é caracterizada pela FDA conjunta (ou bivariada) definida como:

$$F_{XY}(x,y) = Prob\{ X \leq x, e Y \leq y \} \quad (3.14)$$

estimada, na prática, pela proporção de pares de dados conjuntamente abaixo dos respectivos valores (valores de corte) x e y . Isto pode ser mostrado no diagrama de dispersão (Figura 3.3) onde cada par de dado (x_i, y_i) é plotado como um ponto.

O grau de dependência entre as duas variáveis X e Y pode ser caracterizado pela dispersão em torno da linha de 45° no diagrama de dispersão. A dependência perfeita ($X = Y$) corresponde a todos os pares experimentais (x_i, y_i) , $i = 1, \dots, N$ plotados sobre a linha de 45° .

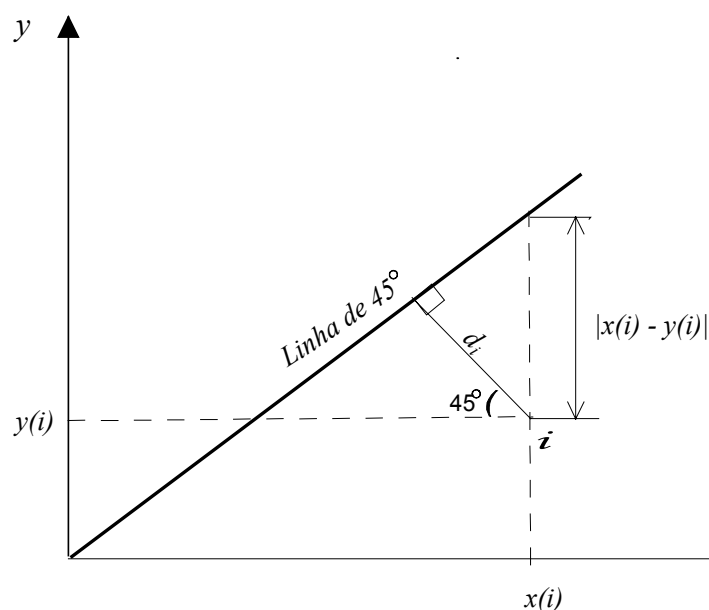


Figura 3.3 – Diagrama de dispersão para os pares (x_i, y_i)

O momento de inércia do diagrama de dispersão em torno da linha de 45° – chamado de “semivariograma” do conjunto de pares (x_i, y_i) – é definido como a metade da média das diferenças quadráticas dos teores entre as coordenadas de cada par, ou seja:

$$\gamma_{XY} = \frac{1}{N} \sum_{i=1}^N d_i^2 = \frac{1}{2N} \sum_{i=1}^N (x_i - y_i)^2 . \quad (3.15)$$

Quanto maior o valor do semivariograma, maior a dispersão e menos relacionadas são as duas variáveis X e Y .

Conforme definido anteriormente, o valor esperado de X é a média ponderada dos n resultados possíveis x_i , ponderados por suas respectivas probabilidades de ocorrência, isto é:

$$E\{X\} = m_X = \sum_{i=1}^N p_i x_i = \frac{1}{N} \sum_{i=1}^N x_i . \quad (3.16)$$

Analogamente à equação (3.10), o valor esperado do produto XY é definido como a média ponderada dos resultados conjuntos $XY = xy$, ponderados por suas respectivas probabilidades de ocorrência, ou seja,

$$E\{XY\} = \int \int_{-\infty}^{+\infty} xy \cdot d^2 F_{XY}(\mathbf{u}; x, y) = \int \int_{-\infty}^{+\infty} xy \cdot f_{XY}(\mathbf{u}; x, y) dx dy , \quad (3.17)$$

estimada na prática por: $\frac{1}{N} \sum_{i=1}^N x_i y_i$,

onde $f_{XY}(\mathbf{u}; x, y) = \frac{d^2 F_{XY}(\mathbf{u}; x, y)}{dx dy}$ é a função densidade de probabilidade (FDP) bivariada.

O momento bivariado (3.17) é chamado covariância não centrada das duas VA's X e Y . A covariância centrada (na média), ou simplesmente covariância, é definida por:

$$\begin{aligned} \text{Cov}\{X, Y\} &= \sigma_{XY} = E\{XY\} = E\{[X - m_X] \cdot [Y - m_Y]\} \\ &= E\{XY\} - m_X \cdot m_Y, \end{aligned} \quad (3.18)$$

estimada na prática por: $\frac{1}{N} \sum_{i=1}^N (x_i - m_X)(y_i - m_Y) = \frac{1}{N} \sum_{i=1}^N x_i y_i - m_X m_Y$.

Conforme a equação (3.11), a variância de X pode ser definida como a autocovariância de X sobre si mesma, isto é:

$$\sigma_X^2 = \text{Var}\{X\} = \text{Cov}\{X, X\} = E\{[X - m_X]^2\} = \left[\frac{1}{N} \sum_{i=1}^N x_i^2 - m_X^2 \right] \geq 0. \quad (3.19)$$

A covariância padronizada (adimensional) entre duas VA's X e Y é conhecida como coeficiente de correlação linear, ou seja:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{Cov}\{X, Y\}}{\sqrt{\text{Var}\{X\} \cdot \text{Var}\{Y\}}} \in [-1, +1]. \quad (3.20)$$

A relação experimental entre o semivariograma e a covariância pode ser obtida pelo desenvolvimento da equação (3.15), mostrada a seguir:

$$2\gamma_{XY} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2.$$

O termo ao quadrado é desenvolvido e é acrescentado o termo: $m_X^2 - m_X^2 + m_Y^2 - m_Y^2$

$$2\gamma_{XY} = \left[\frac{1}{N} \sum_i x_i^2 - m_X^2 \right] + \left[\frac{1}{N} \sum_i y_i^2 - m_Y^2 \right] - \frac{2}{N} \sum_i x_i y_i + m_X^2 + m_Y^2.$$

Os termos em colchetes são substituídos pela equação (3.19) e é acrescentado $2m_X m_Y - 2m_X m_Y$

$$2\gamma_{XY} = \sigma_X^2 + \sigma_Y^2 - 2 \left[\frac{1}{N} \sum_i x_i y_i - m_X m_Y \right] + m_X^2 + m_Y^2 - 2m_X m_Y.$$

$$2\gamma_{XY} = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} + (m_X - m_Y)^2 \geq 0. \quad (3.21)$$

A equação (3.21) mostra que um aumento no semivariograma $2\gamma_{XY}$ é acompanhado de um decréscimo na covariância σ_{XY} . Quanto maior for a dispersão dos pares (x_i, y_i) em torno da linha 45° no diagrama de dispersão, maior será o semivariograma $2\gamma_{XY}$ e menor a covariância σ_{XY} e o coeficiente de correlação ρ_{XY} , o que caracteriza o semivariograma como uma medida de variabilidade e a covariância e correlação como medidas de similaridade.

Para as variáveis padronizadas $X' = (X - m_X)/\sigma_X$ e $Y' = (Y - m_Y)/\sigma_Y$, cujas estatísticas correspondentes são $m_{X'} = m_{Y'} = 0$ e $\sigma_{X'}^2 = \sigma_{Y'}^2 = 1$, então:

$$\sigma_{X'Y'} = E \left\{ \frac{X - m_X}{\sigma_X} \cdot \frac{Y - m_Y}{\sigma_Y} \right\} = \rho_{XY} \in [-1, +1]. \quad (3.22)$$

A substituição de $m_{X'} = m_{Y'} = 0$ e $\sigma_{X'}^2 = \sigma_{Y'}^2 = 1$ na equação (3.21), resulta em:

$$\gamma_{X'Y'} = 1 - \rho_{XY} \in [0,2]. \quad (3.23)$$

Observa-se que, quando $\rho_{XY}=1 \Rightarrow \gamma_{X'Y'}=0$, portanto, todos os pares padronizados $\left(\frac{x_i - m_X}{\sigma_X}, \frac{y_i - m_Y}{\sigma_Y}\right)$ estão alinhados em torno da linha de 45° do diagrama de dispersão.

O coeficiente de correlação ρ_{XY} é invariante a qualquer transformação linear aplicada sobre as VA's X e/ou Y . O semivariograma γ_{XY} e covariância σ_{XY} , ao contrário, são dependentes de transformações lineares nas VA's X e Y , ou seja, de uma translação $X + b$, ou rescalonamento aX .

O coeficiente de correlação ρ_{XY} fornece uma medida da (cor)relação linear entre as duas variáveis X e Y . Se for interpretada de outro modo pode conduzir a erros. Tal como a variância, o coeficiente de correlação é fortemente afetado por valores extremos (muito altos ou muito baixos). Uma medida mais robusta é o coeficiente de correlação de postos, ρ_{XY}^R , onde as variáveis originais são ordenadas independentemente de modo crescente ou decrescente, determinando uma nova variável conhecida por *rank* utilizada para o cálculo de ρ_{XY}^R :

$$\rho_{XY}^R = \frac{1}{N} \frac{\sum_{i=1}^N [r(x_i) - m_{Rx}] \cdot [r(y_i) - m_{Ry}]}{\sigma_{Rx} \cdot \sigma_{Ry}}, \quad (3.24)$$

onde m_{Rx} e σ_{Rx} são as médias e desvio-padrão dos N *ranks* $r(x_i)$. Uma grande diferença entre ρ_{XY} e ρ_{XY}^R reflete ou uma relação não-linear entre as duas variáveis X e Y ou a presença de pares de valores extremos.

As VA's X e Y podem representar:

- (i) duas diferentes propriedades medidas em uma mesma localização, por exemplo, concentração (ou teor) de arsênio e manganês no solo;
- (ii) uma mesma propriedade medida em duas diferentes localizações do espaço, característica das variáveis regionalizadas, por exemplo, concentração de arsênio nas localizações x e $x + h$ separadas por um vetor h , sendo: $X = Z(x)$, $Y = Z(x + h)$;
- (iii) duas diferentes propriedades medidas em duas diferentes localizações, por exemplo, concentração de arsênio na localização x e concentração de manganês separada por um vetor h , sendo: $X = Z_{As}(x)$, $Y = Z_{Mn}(x + h)$.

Em todos os casos, o semivariograma γ_{XY} ou a correlação ρ_{XY} medirão o grau de variabilidade/similaridade entre as duas VA's X e Y .

O segundo caso (ii) é de particular interesse em problemas de interpolação espacial, onde se deseja inferir (mapear) uma determinada área para uma determinada propriedade, $Z(\mathbf{u})$, $\mathbf{u} \in$ área \mathcal{A} , a partir de uma amostra n de $Z(\mathbf{u})$. A Combinação de todos os $n(\mathbf{h})$ pares de dados de $Z(\mathbf{u})$ sobre a mesma área/zona/camada/população \mathcal{A} com tais pares separados aproximadamente pelo mesmo vetor \mathbf{h} (em comprimento e direção), permite estimar o semivariograma característico (ou experimental) da variabilidade espacial em \mathcal{A} :

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_{\alpha}) - z(\mathbf{u}_{\alpha} + \mathbf{h})]^2. \quad (3.25)$$

O semivariograma amostral funciona bem para conjunto de dados que não apresentem efeito proporcional (correlação entre a média e variância de combinações de valores que compõe o conjunto amostral) ou nos quais as localizações das amostras não são preferencialmente agrupadas. Nesse caso, uma alternativa consiste na utilização de semivariogramas relativos (ou padronizados) ao invés do semivariograma tradicional (ISAAKS e SRIVASTAVA, 1989). As medidas de variogramas relativos consideram a média dos dados em cada separação (*lag*) específica \mathbf{h} . Isso permite alguma correção dos agrupamentos preferenciais. Realmente, no semivariograma relativo geral, o semivariograma como definido em (3.25) é padronizado pelo quadrado médio dos dados usados em cada *lag*, ou seja, (DEUTSCH e JOURNAL, 1996, p.45)

$$\gamma_{RG}(\mathbf{h}) = \frac{\gamma(\mathbf{h})}{\left(\frac{m_{-\mathbf{h}} + m_{+\mathbf{h}}}{2} \right)^2}, \quad (3.26)$$

onde $m_{-\mathbf{h}}$ é a média dos valores $z(\mathbf{u}_{\alpha})$ e $m_{+\mathbf{h}}$ é a média dos valores $z(\mathbf{u}_{\alpha} + \mathbf{h})$.

3.3.5. Modelamento da variação espacial

O semivariograma experimental (3.25) é uma estimativa discreta de uma integral espacial bem determinada definindo uma média sobre \mathcal{A} :

$$\gamma_{\mathcal{A}}(\mathbf{h}) = \frac{1}{A(\mathbf{h})} \int_{\mathcal{A}} [z(\mathbf{u}) - z(\mathbf{u} + \mathbf{h})]^2 d\mathbf{u} \quad \text{para } \mathbf{u} \text{ e } \mathbf{u} + \mathbf{h} \in \mathcal{A}. \quad (3.27)$$

Tal como a VA $Z(\mathbf{u})$ e sua distribuição caracteriza a incerteza sobre o valor de certa propriedade na localização \mathbf{u} , a função aleatória $Z(\mathbf{u})$, $\mathbf{u} \in \mathcal{A}$, definida como um

conjunto de VA's dependentes caracterizará a incerteza espacial conjunta sobre \mathcal{A} . O semivariograma dessa função aleatória caracteriza o grau de dependência espacial entre duas variáveis aleatórias $Z(\mathbf{u})$ e $Z(\mathbf{u}+\mathbf{h})$ separados de um vetor \mathbf{h} .

Para o modelamento do semivariograma, realizado depois de construído o semivariograma experimental, é necessário que a hipótese de estacionariedade seja considerada. Esta hipótese estabelece, em resumo, que os dois primeiros momentos (média e variância) da diferença $[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]$ são independentes da localização \mathbf{u} e função somente do vetor \mathbf{h} . O segundo momento dessa diferença corresponde ao semivariograma, ou seja:

$$2\gamma(\mathbf{h}) = E\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2\} \text{ é independente de } \mathbf{u} \in \mathcal{A}. \quad (3.28)$$

Desenvolvendo da equação acima (acrescentando m^2 em todos os termos, por conveniência), obtemos:

$$2\gamma(\mathbf{h}) = \{E[Z^2(\mathbf{u})] - m^2\} + \{E[Z^2(\mathbf{u}+\mathbf{h})] - m^2\} - 2\{E[Z(\mathbf{u}).Z(\mathbf{u}+\mathbf{h})] - m^2\}.$$

$$2\gamma(\mathbf{h}) = \text{Var}\{Z(\mathbf{u})\} + \text{Var}\{Z(\mathbf{u}+\mathbf{h})\} - 2.\text{Cov}\{Z(\mathbf{u}),Z(\mathbf{u}+\mathbf{h})\}.$$

$$2\gamma(\mathbf{h}) = 2[C(0) - C(\mathbf{h})].$$

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}), \quad (3.29)$$

sendo que:

$$\text{Var}\{Z(\mathbf{u})\} = \text{Var}\{Z(\mathbf{u}+\mathbf{h})\} = \sigma^2 = C(0) \quad \text{para todo } \mathbf{u} \in \mathcal{A}.$$

$$\text{Cov}\{Z(\mathbf{u}),Z(\mathbf{u}+\mathbf{h})\} = C(\mathbf{h}) \quad \text{para todo } \mathbf{u} \in \mathcal{A}.$$

A relação (3.29) é utilizada então para a determinação do modelo semivariográfico. A variância $C(0)$ é chamada na geoestatística como patamar (ou sill). O semivariograma é a ferramenta preferida para inferência estatística porque oferece algumas vantagens sobre a covariância, entre as quais (CHRISTAKOS, 1984):

- (1) seu cálculo empírico está sujeito a erros menores;
- (2) oferece melhor caracterização da variabilidade espacial;
- (3) exige a chamada hipótese de estacionariedade intrínseca, ou seja, que $Z(\mathbf{u})$ é uma função aleatória com incrementos $Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})$ estacionários, mas não necessariamente ela própria estacionária.

CHRISTAKOS (1984) mostra ainda que, em geral, para uma função contínua escolhida como semivariograma é necessário e suficiente satisfazer a propriedade de

positivo definida, que garante que a variância de certas combinações lineares descritas por $Y = \sum_{i=1}^n \lambda_i Z(\mathbf{u}_i)$ sejam positivas. O modelo semivariográfico $\gamma(\mathbf{h})$ deve ser condicionalmente negativo definido, isto é:

$$-\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{h}) \geq 0 \quad \text{quando} \quad \sum_{i=1}^n \lambda_i = 0 \quad \forall n.$$

Outra condição a ser satisfeita pelo semivariograma é:

$$\lim_{|\mathbf{h}| \rightarrow \infty} \gamma(\mathbf{h}) / |\mathbf{h}|^2 = 0.$$

Uma vez que não é fácil verificar essas condições diretamente (CHRISTAKOS, 1984), a solução prática é usar combinações lineares de modelos básicos que são válidos, isto é, permissíveis. Os modelos básicos mais utilizados em geoestatística são apresentados a seguir:

1 – Modelo de efeito pepita

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{se } \mathbf{h} = 0 \\ C_0 & \text{caso contrário} \end{cases} \quad (3.30)$$

2 – Modelo Esférico

$$\gamma(\mathbf{h}) = \begin{cases} 0 & |\mathbf{h}| = 0 \\ C \left[\frac{3}{2} \left(\frac{|\mathbf{h}|}{a} \right) - \frac{1}{2} \left(\frac{|\mathbf{h}|}{a} \right)^3 \right] & 0 < |\mathbf{h}| \leq a \\ C & |\mathbf{h}| > a \end{cases} \quad (3.31)$$

3 – Modelo Exponencial

$$\gamma(\mathbf{h}) = \begin{cases} 0 & |\mathbf{h}| = 0 \\ C \left[1 - \exp\left(-\frac{3|\mathbf{h}|}{a}\right) \right] & |\mathbf{h}| \neq 0 \end{cases} \quad (3.32)$$

4 – Modelo Gaussiano

$$\gamma(\mathbf{h}) = \begin{cases} 0 & |\mathbf{h}| = 0 \\ C \left[1 - \exp\left(-\frac{3|\mathbf{h}|^2}{a^2}\right) \right] & |\mathbf{h}| \neq 0 \end{cases} \quad (3.33)$$

O modelo de efeito pepita (3.30) é usado para modelar a descontinuidade na origem causada por erros de amostragens, erros de medida ou microregionalizações da variável analisada. As componentes C e a são denominadas de patamar e alcance, respectivamente. O patamar, também conhecido como “sill”, representa o nível de variabilidade do semivariograma até a sua estabilização. O alcance (ou amplitude variográfica) é a distância observada até o nível onde a variabilidade se estabiliza. Indica a distância em que as amostras estão correlacionadas espacialmente (Fig. 3.4). A Figura 3.5 mostra exemplos de modelos semivariográficos esférico, exponencial e gaussiano. Esses modelos, que atingem um patamar, são chamados de transitivos.

5 – Modelo Potência

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , |\mathbf{h}| = 0 \\ c |\mathbf{h}|^w & , |\mathbf{h}| \neq 0 \end{cases} \quad (3.34)$$

onde

- $c \geq 0$ é o coeficiente de declividade;
- $0 \leq w \leq 2$ é o expoente.

Este modelo não é transitivo, ou seja, não atinge o patamar. É utilizado, em geral, para modelar fenômenos com capacidade infinita de dispersão. A Figura 3.6 mostra modelos potência para diferentes valores do parâmetro w . Quando $w = 1$, o modelo se chama linear. Os modelos transitivos possuem a covariância correspondente representada por $C(\mathbf{h}) = C(0) - \gamma(\mathbf{h})$. O modelo potência, ao contrário, por não possuir patamar, não possui a covariância correspondente.

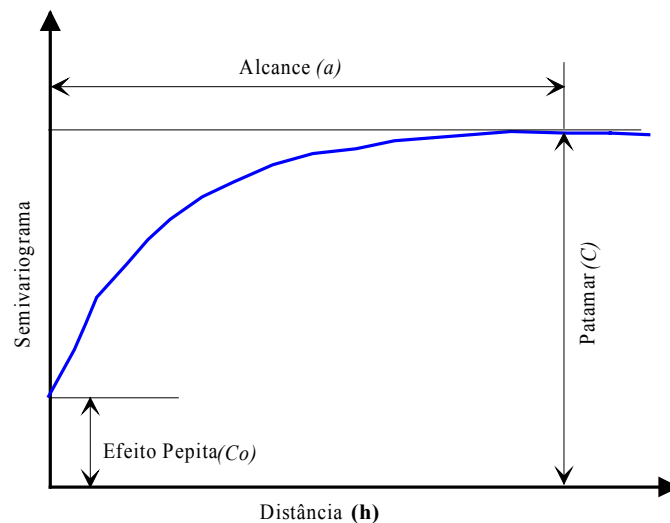


Figura 3.4 – Parâmetros do semivariograma

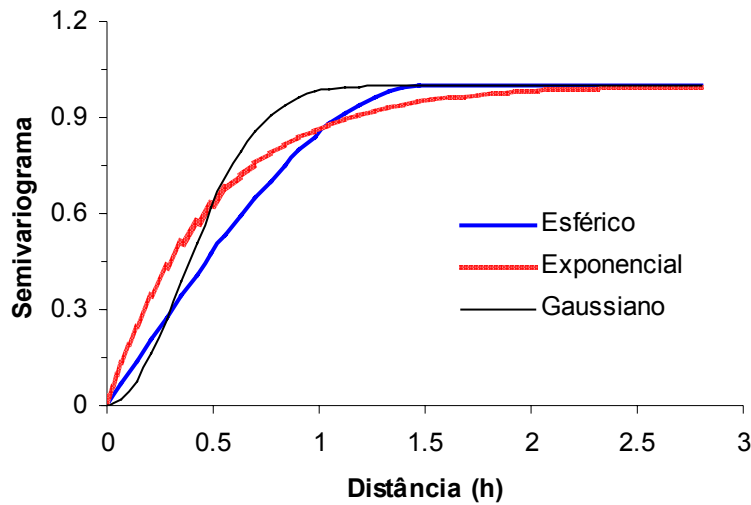


Figura 3.5 – Modelos semivariográficos do tipo transitivo (com patamar) mais usuais em geoestatística.

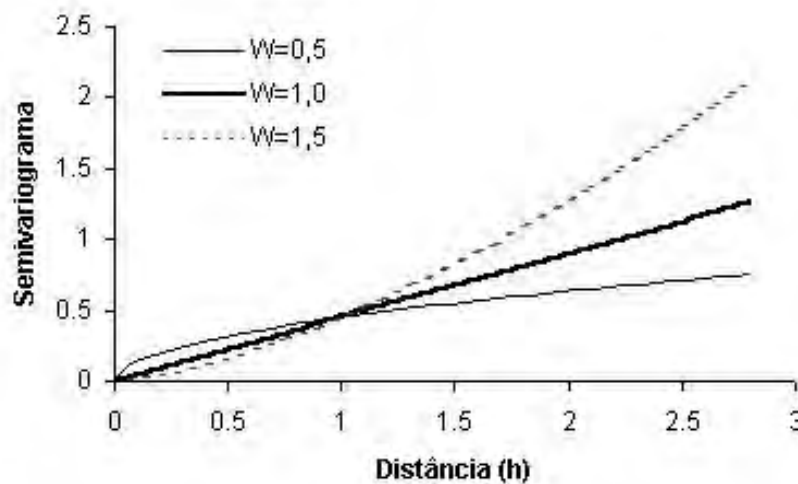


Figura 3.6 – Modelos potência para diferentes valores de w .

O semivariograma (e covariância) são funções do vetor \mathbf{h} (comprimento e direção). Quando a dependência é somente em relação ao comprimento (distância) do vetor \mathbf{h} , o modelo é chamado isotrópico. Se a dependência é também na direção do vetor \mathbf{h} , então o modelo é chamado anisotrópico. A anisotropia pode ser verificada através da observação dos semivariogramas experimentais em diferentes direções, por exemplo, 0° (Norte), 45° , 90° (Leste) e 135° . Caso os semivariogramas nessas direções sejam similares, o fenômeno é chamado *isotrópico* (ou omnidirecional), indicando que o padrão de variabilidade espacial não muda com a direção e um único modelo

(isotrópico) é suficiente para descrever a variabilidade espacial da variável em estudo. Por outro lado, se são observadas, entre os semivariogramas, diferenças nas direções, o fenômeno é denominado *anisotrópico*. Quando as diferenças são observadas somente em relação ao alcance, a anisotropia é chamada *geométrica*. Caso as diferenças sejam tanto no alcance quanto no patamar, a anisotropia é do tipo *zonal*. A Figura 3.7 ilustra esses dois casos.

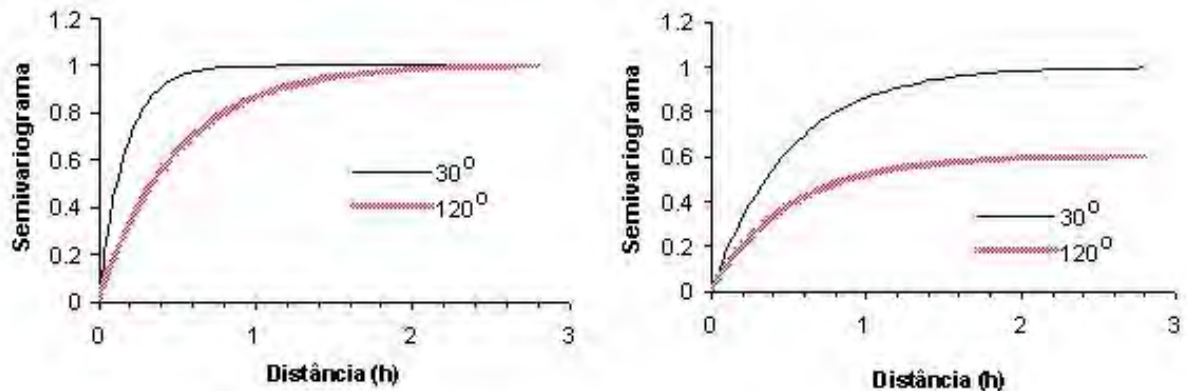


Figura 3.7 – Tipos de anisotropia variográfica; a) geométrica b) zonal

Em muitas situações dois ou mais modelos permissíveis devem ser combinados para ajustar a forma do semivariograma experimental. Isso pode ocorrer quando os dados estão disponíveis para uma extensão bastante longa e o semivariograma experimental ao longo daquela direção pode revelar diversas escalas de variabilidade. Essa sucessão aninhada (*nested*) de escalas (S) de variabilidade pode ser modelada por uma soma de semivariogramas:

$$\gamma(\mathbf{h}) = \sum_{s=0}^S b^s \gamma_s(\mathbf{h}) \quad \text{com } b^s \geq 0, \quad (3.35)$$

sendo b^s o patamar positivo do correspondente modelo semivariográfico $\gamma_s(\mathbf{h})$.

Cada estrutura $\gamma_s(\mathbf{h})$ caracteriza o padrão de variabilidade/continuidade espacial prevalecendo em uma escala particular. Por exemplo, as S estruturas podem ter alcances $a_1 \ll a_2 \ll \dots \ll a_S$.

O objetivo da análise estrutural (semivariográfica) é construir um modelo permissível que capture as mais importantes características espaciais da variável em estudo. Embora os valores do semivariograma experimental desempenhem um papel importante neste processo, a informação secundária, tal como a fornecida por um conhecimento físico da área e fenômeno pode ser de grande interesse.

GOOVAERTS (1999) apresenta o caso bivariado, onde o modelamento da co-regionalização entre duas variáveis aleatórias exige que sejam modelados os dois semivariogramas diretos $\gamma_{ii}(\mathbf{h})$ e $\gamma_{jj}(\mathbf{h})$ mais o semivariograma cruzado $\gamma_{ij}(\mathbf{h})$. Os três modelos devem ser construídos juntos para garantir que a matriz de modelos semivariográficos $\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]$ seja condicionalmente negativa semi-definida para todas as distâncias \mathbf{h} , o que requer que a seguinte desigualdade seja satisfeita:

$$|\gamma_{ij}(\mathbf{h})| \leq \sqrt{\gamma_{ii}(\mathbf{h})\gamma_{jj}(\mathbf{h})} \quad \forall \mathbf{h}, \quad (3.36)$$

A condição (3.35) é facilmente checada toda a vez que os três semivariogramas são modelados como combinações lineares do mesmo conjunto básico de modelos $\gamma_s(\mathbf{h})$:

$$\gamma_{ij}(\mathbf{h}) = \sum_{s=0}^S b_{ij}^s \gamma_s(\mathbf{h}) \quad \forall i, j, \quad (3.37)$$

ou, usando notação matricial

$$\Gamma(\mathbf{h}) = \sum_{s=0}^S \mathbf{B}^s \gamma_s(\mathbf{h}), \quad (3.38)$$

onde $\mathbf{B}^s = [b_{ij}^s]$ é referida como uma matriz de co-regionalização.

As condições suficientes para que o chamado modelo linear de co-regionalização (MLC) seja permissível são: (1) as funções $\gamma_s(\mathbf{h})$ são modelos semivariográficos permissíveis, e (2) cada matriz de co-regionalização é positivo-definida, o que implica que os coeficientes b_{ij}^s devem satisfazer as seguintes restrições:

$$b_{ii}^s \geq 0, \quad b_{jj}^s \geq 0 \quad \forall s, \quad (3.39)$$

$$|b_{ij}^s| \leq \sqrt{b_{ii}^s \cdot b_{jj}^s} \quad \forall s. \quad (3.40)$$

Na prática, o modelamento é realizado em dois passos:

1. primeiro, são modelados os dois semivariogramas diretos como combinações lineares de estruturas básicas selecionadas $\gamma_s(\mathbf{h})$;
2. em seguida, as mesmas estruturas básicas são ajustadas ao semivariograma cruzado sob a restrição (3.40).

3.3.6. Estimativa espacial

Depois que o modelo de dependência espacial entre as duas variáveis aleatórias $Z(\mathbf{u})$ e $Z(\mathbf{u}+\mathbf{h})$ é estabelecido, pode-se direcionar o problema da estimação de um valor desconhecido $z(\mathbf{u})$ a partir dos valores disponíveis. O objetivo não é somente encontrar uma estimativa $z^*(\mathbf{u})$ do valor desconhecido, mas também o modelamento da incerteza sobre essa estimativa. A incerteza é dependente da informação disponível, dos valores observados $z(\mathbf{u}_\alpha)$'s e do modelo de dependência espacial estabelecido.

O valor desconhecido $z(\mathbf{u})$ pode ser estimado a partir de uma combinação linear dos n valores observados mais um parâmetro λ_0 :

$$z^*(\mathbf{u}) = \lambda_0 + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})z(\mathbf{u}_\alpha), \quad (3.41)$$

onde $z^*(\mathbf{u})$ é chamada uma “estimativa” e $Z^*(\mathbf{u})$, a VA, é chamado “estimador”, isto é:

$$Z^*(\mathbf{u}) = \lambda_0 + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})Z(\mathbf{u}_\alpha). \quad (3.42)$$

Deve-se, portanto, determinar o conjunto de pesos $(\lambda_\alpha, \alpha = 1, \dots, n)$ junto com uma medida da incerteza associada, ou seja, uma medida da variância, $V[Z(\mathbf{u}) - Z^*(\mathbf{u})]$. Existem diversos métodos de estimação entre os quais a interpolação linear simples que atribui pesos todos iguais a $1/N$ ($N =$ número de amostras), o inverso do quadrado da distância, onde são usados pesos λ_α proporcionalmente inversos ao quadrado da distância (Euclidiana) que separa o valor estimado do valor observado. As principais falhas dos métodos tradicionais de estimação para utilização na geoestatística são: não considerar o padrão de dependência espacial da variável e não fornecer uma medida de erro para a estimativa.

3.3.7. Krigagem

Krigagem é a designação genérica adotada na geoestatística para uma família de algoritmos de regressão de mínimos-quadrados baseados no estimador de regressão linear $Z^*(\mathbf{u})$. O valor esperado do erro da VA é:

$$E\{Z(\mathbf{u}) - Z^*(\mathbf{u})\} = E\{Z(\mathbf{u})\} - E\{Z^*(\mathbf{u})\}. \quad (3.43)$$

Substituindo-se $Z^*(\mathbf{u})$ da 3.41, obtemos:

$$E\{Z(\mathbf{u})\} - \lambda_0 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u})E\{Z(\mathbf{u}_{\alpha})\} = m(\mathbf{u}) - \lambda_0 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}m(\mathbf{u}_{\alpha}).$$

Para que o estimador seja não tendencioso, o valor esperado do erro deve ser zero, então o parâmetro λ_0 é assumido como:

$$\lambda_0 = m(\mathbf{u}) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}m(\mathbf{u}_{\alpha}). \quad (3.44)$$

Substituindo o valor de λ_0 em 3.41, obtemos:

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u})[Z(\mathbf{u}_{\alpha}) - m(\mathbf{u}_{\alpha})], \quad (3.45)$$

onde $\lambda_{\alpha}(\mathbf{u})$ é o peso atribuído a cada valor observado $Z(\mathbf{u}_{\alpha})$ localizado dentro de determinada vizinhança $W(\mathbf{u})$ centrado em \mathbf{u} . Os $n(\mathbf{u})$ pesos são escolhidos de forma a minimizar a estimação ou erro da variância $\sigma_E^2(\mathbf{u}) = Var[Z^*(\mathbf{u}) - Z(\mathbf{u})]$ sob a condição de não-tendenciosidade do estimador. As diferenças entre os vários tipos de krigagem residem no modelo considerado para a tendência $m(\mathbf{u})$ na expressão (3.45). A krigagem simples (KS) considera a média $m(\mathbf{u})$ conhecida e constante por toda a área de estudo, enquanto a krigagem ordinária (KO) limita o domínio de estacionaridade da média à vizinhança local $W(\mathbf{u})$ e, ao contrário da KS, a média é desconhecida.

3.3.7.1. Krigagem Ordinária

A krigagem ordinária (KO) considera a variação local da média limitada ao domínio de estacionaridade da média à vizinhança local $W(\mathbf{u})$ centrada sobre a localização \mathbf{u} a ser estimada. Nesse caso, considera-se a média comum (estacionária) $m(\mathbf{u})$ na equação 3.45, ou seja:

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u})[Z(\mathbf{u}_{\alpha}) - m(\mathbf{u})]. \quad (3.46)$$

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u})Z(\mathbf{u}_{\alpha}) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u})\right]m(\mathbf{u}).$$

A média $m(\mathbf{u})$ desconhecida pode ser eliminada considerando-se a soma dos pesos ($\lambda_{\alpha}(\mathbf{u})$) da krigagem igual a 1. Deste modo:

$$Z_{KO}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{KO}(\mathbf{u})Z(\mathbf{u}_{\alpha}) \quad \text{com} \quad \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{KO}(\mathbf{u}) = 1. \quad (3.47)$$

A minimização da variância do erro ($Var[Z^*(\mathbf{u}) - Z(\mathbf{u})]$) sob a condição

$\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{KO}(\mathbf{u}) = 1$, permite a determinação dos pesos λ_{α} a partir do seguinte sistema de

equações chamado de sistema de krigagem ordinária (equações normais com restrições):

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta}^{KO}(\mathbf{u}) C(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha}) + \mu(\mathbf{u}) = C(\mathbf{u} - \mathbf{u}_{\alpha}), & \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta}^{KO}(\mathbf{u}) = 1 \end{cases}, \quad (3.48)$$

onde $C(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha})$ e $C(\mathbf{u} - \mathbf{u}_{\alpha})$ são, respectivamente, a covariância entre os pontos \mathbf{u}_{β} e \mathbf{u}_{α} e \mathbf{u} e \mathbf{u}_{α} e $\mu(\mathbf{u})$ é o parâmetro de Lagrange associado com a restrição

$$\sum_{\beta=1}^{KO} \lambda_{\beta}(\mathbf{u}) = 1.$$

O sistema de krigagem (3.48) apresenta uma e somente uma solução se:

- A função de covariância $C(\mathbf{h})$ é positivo-definida. Esta condição significa que a função $C(\mathbf{h})$ ou matriz de covariância $[C_{\alpha\beta}]$ deve garantir a positividade da variância para qualquer combinação de variáveis aleatórias $Z(\mathbf{u}_{\alpha})$, isto é:

$$Var\left\{\sum_{\alpha=1}^N \lambda_{\alpha} Z(\mathbf{u}_{\alpha})\right\} = \sum_{\alpha=1}^N \sum_{\beta=1}^N \lambda_{\alpha} \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) \geq 0.$$

- Não existem dois dados totalmente redundantes, isto é, $\mathbf{u}_{\alpha} \neq \mathbf{u}_{\beta}$ se $\alpha \neq \beta$.

A correspondente variância mínima do erro, chamada de variância da krigagem ordinária, é dada por:

$$\sigma_{KO}^2 = Var[Z(\mathbf{u}) - Z^*(\mathbf{u})] = C_0 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha} C(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha}) - \mu(\mathbf{u}). \quad (3.49)$$

onde $C_0 = Var Z(\mathbf{u}) = \sigma^2$.

Substituindo a covariância por sua expressão $C(\mathbf{h}) = C_0 - \gamma(\mathbf{h})$, o sistema (3.48) e a variância σ_{KO}^2 podem ser escritos em função do modelo semivariográfico $\gamma(\mathbf{h})$.

Assim, ao contrário dos estimadores lineares mais tradicionais, a krigagem utiliza um sistema de pesos que considera um modelo de correlação espacial específico à variável na área A em estudo. A krigagem fornece não somente uma estimativa de mínimos quadrados da variável em estudo, mas também o erro de variância associado. Entretanto, a variância da krigagem tem sido freqüentemente mal utilizada como uma medida de confiança da estimativa da krigagem, uma vez que a variância dos erros é independente dos valores reais dos dados e depende somente da configuração dos dados (JOURNAL e ROSSI, 1989; GOOVAERTS, 1999). Uma alternativa é modelar a

incerteza sobre o valor desconhecido $z(\mathbf{u})$ a partir da FDAC, $F(\mathbf{u};z|(n))$, daquela variável. O valor dessa função é determinado por uma série de valores de corte z_k que discretizam a faixa de variação de z :

$$F(\mathbf{u}; z_k | (n)) = \text{Pr ob}\{Z(\mathbf{u}) \leq z_k | (n)\}, \quad k = 1, \dots, K. \quad (3.50)$$

A estimação geoestatística não paramétrica dos valores da FDAC é baseada na interpretação da probabilidade condicional (equação 3.5) como o valor esperado condicional de uma variável indicativa $I(\mathbf{u}; z_k)$ com base nos n dados observados:

$$F(\mathbf{u}; z_k | (n)) = E\{I(\mathbf{u}) \leq z_k | (n)\}, \quad (3.51)$$

sendo $I(\mathbf{u}; z_k) = 1$ se $Z(\mathbf{u}) \leq z_k$ e zero se $Z(\mathbf{u}) > z_k$. Os valores da função de distribuição acumulada condicional podem então ser estimadas pela krigagem ordinária dos dados indicativos transformados.

Na implementação da aproximação indicativa deve-se escolher o conjunto de K valores de corte de modo que a faixa de valores da variável z seja dividida em $K + 1$ classes com frequências aproximadamente iguais. Para cada valor de corte z_k são computados os semivariogramas indicativos experimentais, definidos por:

$$\hat{\gamma}_I(\mathbf{h}; z_k) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{u}_\alpha; z_k) - i(\mathbf{u}_\alpha + \mathbf{h}; z_k)]^2, \quad (3.52)$$

onde $N(\mathbf{h})$ é o número de pares de dados dentro de determinada classe separados pelo vetor \mathbf{h} (distância e direção). Quanto maior for o valor de $\hat{\gamma}_I(\mathbf{h}; z_k)$, menos conectados no espaço são os pequenos ou grandes valores.

Uma vez que a incerteza sobre o valor não amostrado tenha sido modelada pela utilização da aproximação indicativa, uma estimativa para aquele valor desconhecido pode ser obtida a partir da função de distribuição acumulada condicional. A média (semelhante à equação 3.10) da distribuição acumulada condicional conhecida como estimativa E-type de $z(\mathbf{u})$ é definida como:

$$z(\mathbf{u})_E^* \approx \sum_{k=1}^{K+1} z_k [F(\mathbf{u}; z_k | (n)) - F(\mathbf{u}; z_{k-1} | (n))]. \quad (3.53)$$

E a variância condicional da distribuição acumulada condicional é obtida por:

$$\sigma^2(\mathbf{u}) \approx \sum_{k=1}^{K+1} [z_k - z(\mathbf{u})_E^*]^2 \cdot [F(\mathbf{u}; z_k | (n)) - F(\mathbf{u}; z_{k-1} | (n))], \quad (3.54)$$

onde:

- $z_k, k = 1, \dots, K$ são os valores de corte que discretizam a faixa variação dos valores z ;

- z'_k é a média da classe $(z_{k-1} - z_k]$, que depende do modelo de interpolação utilizado dentro das classes, por exemplo, para o modelo linear, $z'_k = (z_{k-1} + z_k)/2$.

A diferença sobre a krigagem ordinária é que a aproximação indicativa fornece uma medida de incerteza que considera os valores dos dados, enquanto a variância da krigagem depende somente da configuração dos dados e modelo semivariográfico. As estimativas E-type e krigagem são iguais somente quando os valores originais z são normalmente distribuídos e a FDAC é modelada por meio de uma aproximação multi-Gaussiana (RAUTMAN & ISTOK,1996).

3.3.7.2. Cokrigagem

A cokrigagem é uma extensão da krigagem no caso em que se deseja utilizar na estimativa de $Z(\mathbf{u})$ mais de uma variável, expressas como combinações lineares de variáveis primárias e secundárias. No caso de uma simples variável secundária (Y), o estimador de cokrigagem ordinária é escrito como:

$$Z_{COK}^*(\mathbf{u}) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1}(\mathbf{u})Z(\mathbf{u}_{\alpha_1}) + \sum_{\alpha_2=1}^{n_2} \lambda'_{\alpha_2}(\mathbf{u})Y(\mathbf{u}'_{\alpha_2}), \quad (3.55)$$

com as condições de não-tendenciosidade:

$$\sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1}(\mathbf{u}) = 1 \quad \text{e} \quad \sum_{\alpha_2=1}^{n_2} \lambda'_{\alpha_2} = 1,$$

onde λ_{α_1} , $\alpha_1 = 1, \dots, n_1$, são os pesos aplicados às n_1 amostras z e λ'_{α_2} , $\alpha_2 = 1, \dots, n_2$, são os pesos aplicados às n_2 amostras y .

A krigagem necessita de um modelo de covariância (ou semivariograma) para Z . A cokrigagem exige um modelo conjunto para a matriz de funções de covariâncias, incluindo a covariância de Z , $C_Z(\mathbf{h})$, a covariância de Y , $C_Y(\mathbf{h})$ e a covariância cruzada de Z - Y , $C_{ZY}(\mathbf{h}) = Cov\{Z(\mathbf{u}), Y(\mathbf{u}+\mathbf{h})\}$, dados por:

$$C_Z(\mathbf{h}) = E\{Z(\mathbf{u})Z(\mathbf{u}+\mathbf{h})\} - E\{Z(\mathbf{u})\} \{E\{Z(\mathbf{u}+\mathbf{h})\}\},$$

$$C_Y(\mathbf{h}) = E\{Y(\mathbf{u})Y(\mathbf{u}+\mathbf{h})\} - E\{Y(\mathbf{u})\} \{E\{Y(\mathbf{u}+\mathbf{h})\}\},$$

$$C_{ZY}(\mathbf{h}) = E\{Z(\mathbf{u})Y(\mathbf{u}+\mathbf{h})\} - E\{Z(\mathbf{u})\} \{E\{Y(\mathbf{u}+\mathbf{h})\}\},$$

ou $2\gamma_{ZY}(\mathbf{h}) = E\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})][Y(\mathbf{u}) - Y(\mathbf{u}+\mathbf{h})]\} = 2[C_{ZY}(0) - C_{ZY}(\mathbf{h})]$.

O correspondente sistema de cokrigagem para determinação dos n pesos λ_{α_1} e λ_{α_2} , semelhante ao sistema de krigagem, é composto de $(n_1 + n_2 + 2)$ equações com 2 parâmetros de Lagrange μ_1 e μ_2 associados às duas condições de não-tendenciosidade:

$$\begin{cases} \sum_{\beta=1}^{n_1} \lambda_{\beta}(\mathbf{u}) C_Z(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha}) + \sum_{\beta'=1}^{n_2} \lambda_{\beta'} C_{ZY}(\mathbf{u}_{\beta'} - \mathbf{u}_{\alpha}) + \mu_1 = C_Z(\mathbf{u} - \mathbf{u}_{\alpha}) & \text{para } \alpha = 1, \dots, n_1 \\ \sum_{\beta=1}^{n_1} \lambda_{\beta}(\mathbf{u}) C_{ZY}(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha'}) + \sum_{\beta'=1}^{n_2} \lambda_{\beta'} C_Y(\mathbf{u}_{\beta'} - \mathbf{u}_{\alpha'}) + \mu_2 = C_{ZY}(\mathbf{u} - \mathbf{u}_{\alpha'}) & \text{para } \alpha' = 1, \dots, n_2 \\ \sum_{\beta=1}^{n_1} \lambda_{\beta} = 1 \\ \sum_{\beta'}^{n'} \lambda_{\beta'} = 1 \end{cases} \quad (3.56)$$

Este sistema de cokrigagem apresenta uma e somente uma solução se:

(i) não há dois dados totalmente redundantes, isto é:

$$u_{\alpha} \neq u_{\beta} \quad \text{se: } \alpha \neq \beta = 1, \dots, n.$$

$$u_{\alpha'} \neq u_{\beta'} \quad \text{se: } \alpha' \neq \beta' = 1, \dots, n'.$$

(ii) a matriz de funções de covariância: $\begin{bmatrix} C_Z(\mathbf{h}) & C_{ZY}(\mathbf{h}) \\ C_{ZY}(\mathbf{h}) & C_Y(\mathbf{h}) \end{bmatrix}$ é positiva-definida.

Na estimativa de um elemento $z(\mathbf{u})$, a cokrigagem reúne não somente a correlação espacial dos dados $z(\mathbf{u}_{\alpha})$ do mesmo tipo, mas também a correlação espacial cruzada dos dados $y(\mathbf{u}_{\alpha'})$ de outro tipo. Isto é particularmente útil se a variável z a ser estimada é sub-amostrada em relação à outra variável y : $n \ll n'$. Em lugar de dois elementos diferentes z , y , digamos Pb e Cd , os dois conjuntos de dados podem corresponder a dois tipos diferentes de medida do mesmo elemento: por exemplo, z é uma medida precisa, mas cara da concentração de Pb , enquanto y é uma medida imprecisa, porém barata da mesma concentração de Pb , com geralmente $n' \gg n$. É provável que o padrão de auto-correlação $C_X(h)$ dos dados precisos será diferente daqueles mais erráticos, $C_Y(h)$, dos dados imprecisos. A cokrigagem permite agrupamento de dados de qualidade diferente, ainda privilegiando os dados de melhor qualidade.

3.3.7.3. Krigagem fatorial

O objetivo da krigagem fatorial, no caso de concentração de um metal, é descobrir qual a origem dessa concentração. Por exemplo, traços de metais no solo ou na água podem originar-se naturalmente de rochas ou podem ser resultado de atividades humanas, tais como mineração, resíduos industriais ou agricultura. Enquanto pouco

pode ser feito para corrigir ocorrências naturais de grandes concentrações de metais, medidas podem ser tomadas para prevenir o agravamento de poluições causadas por atividades antrópicas. Se as escalas nas quais os diferentes fatores (humanos, geológicos) operam são muito diferentes um do outro, então eles poderiam ser observados nos semivariogramas das concentrações dos metais. A análise estrutural realizada com modelos semivariográficos aninhados do tipo (3.34) podem revelar essas escalas. A função aleatória $Z(\mathbf{u})$ com um modelo semivariográfico aninhado $\gamma(\mathbf{h})$ pode ser interpretada como a soma de $(S + 1)$ funções aleatórias independentes $Z^s(\mathbf{u})$, com média zero e semivariograma $b^s \gamma_s(\mathbf{h})$:

$$Z(\mathbf{u}) = \sum_{s=0}^S Z^s(\mathbf{u}) + m(\mathbf{u}), \quad (3.57)$$

onde $m(\mathbf{u})$ é assumida localmente constante como na prática da krigagem ordinária.

Uma vez que as diversas escalas são identificadas no semivariograma, a correspondente componente espacial $Z^s(\mathbf{u})$ pode ser estimada e mapeada utilizando-se a krigagem ordinária onde cada componente espacial é associada com o respectivo semivariograma $\gamma_s(\mathbf{h})$. Os perfis ou mapas das estimativas das componentes espaciais podem auxiliar na separação de características locais e regionais do fenômeno em estudo. Em geral, o modelamento semivariográfico tem 3 escalas de variação espacial:

- (1) micro-escala – corresponde à componente do efeito pepita, onde o alcance é zero e as estimativas são nulas em qualquer localização não amostrada; pode estar relacionada à remobilização local da concentração do poluente e/ou erros de medidas;
- (2) escala local (pequeno alcance) – podem estar relacionadas às concentrações anômalas do metal – acima do limite máximo tolerável – e podem ser resultado de rochas naturalmente ricas de tal metal ou originar-se de atividades humanas, cujo impacto é temporariamente balanceado por pequenas concentrações naturais;
- (3) escala regional (longo alcance) – geralmente relacionada à influência da geologia regional.

3.3.7.4. Krigagem fatorial multivariada

A consideração da escala espacial no estudo das correlações entre variáveis pode ajudar no entendimento dos mecanismos físicos fundamentais que controlam os padrões espaciais do fenômeno em estudo. As relações observadas entre concentrações de diferentes tipos de metais poderiam estar associadas com a ocorrência de fontes comuns

de contaminação. A krigagem fatorial múltipla permite analisar as relações entre as variáveis $Z_i(\mathbf{u})$ nas escalas espaciais detectadas e modeladas pelos semivariogramas (aninhados) experimentais. Tal como a krigagem fatorial é baseada sobre o modelo linear de regionalização (3.35), a krigagem fatorial múltipla baseia-se no modelo linear de co-regionalização (equação 3.37) ajustado pelos semivariogramas diretos e cruzados. Sob tal modelo de co-regionalização cada função aleatória $Z_i(\mathbf{u})$ correlacionada pode ser interpretada como a soma de funções aleatórias não correlacionadas $\{Y_k^s(\mathbf{u}), k=1, \dots, p$ (número de variáveis); $s = 0, \dots, S$ (número de diferentes escalas espaciais)}, com coeficientes de transformação a_{ik}^s :

$$Z_i(\mathbf{u}) = \sum_{s=0}^S \sum_{k=1}^p a_{ik}^s Y_k^s + m(\mathbf{u}) \quad i = 1, \dots, p; \quad (3.58)$$

com $E\{Z_i(\mathbf{u})\} = m(\mathbf{u})$ e $E\{Y_k^s(\mathbf{u})\} = 0$,

onde $Y_k^s(\mathbf{u})$ são chamados *fatores regionalizados*, onde k indica os diferentes fatores em determinada escala espacial s . Para uma determinada escala fixada s os p fatores regionalizados têm a mesma função variográfica $\gamma_s(\mathbf{h})$. A equação (3.37) pode ser derivada da equação (3.58) considerando-se que os fatores regionalizados são, por construção, mutuamente ortogonais, ou seja, nesse caso pode-se escrever:

$$\begin{aligned} \frac{1}{2} E[\{Y_k^s(\mathbf{u}) - Y_k^s(\mathbf{m} + \mathbf{h})\} \{Y_{k'}^{s'}(\mathbf{u}) - Y_{k'}^{s'}(\mathbf{u} + \mathbf{h})\}] &= \gamma_s(\mathbf{h}), \text{ se } s = s' \text{ e } k = k'. \\ &= 0 \text{ caso contrário.} \end{aligned}$$

os variogramas simples e cruzados $\gamma_{ij}(\mathbf{h})$ podem ser expressos como combinação linear das funções semivariográficas básicas $\gamma_s(\mathbf{h})$ com coeficientes b_{ij}^s

$$\begin{aligned} \gamma_{ij}(\mathbf{h}) &= \frac{1}{2} E[\{Z_i(\mathbf{u}) - Z_i(\mathbf{u} + \mathbf{h})\} \{Z_j(\mathbf{u} + \mathbf{h}) - Z_j(\mathbf{u} + \mathbf{h})\}]; \\ &= \sum_{s=0}^S \sum_{s'=0}^S \sum_{k=1}^p \sum_{k'=1}^p a_{ik}^s a_{jk'}^{s'} \frac{1}{2} E[\{Y_k^s(\mathbf{u}) - Y_k^s(\mathbf{u} + \mathbf{h})\} \{Y_{k'}^{s'}(\mathbf{u}) - Y_{k'}^{s'}(\mathbf{u} + \mathbf{h})\}]; \\ &= \sum_{s=0}^S \sum_{k=1}^p a_{ik}^s a_{jk}^s \gamma_s(\mathbf{h}) = \sum_{s=0}^S b_{ij}^s \gamma_s(\mathbf{h}). \end{aligned} \quad (3.59)$$

onde
$$b_{ij}^s = \sum_{k=1}^p a_{ik}^s a_{jk}^s. \quad (3.60)$$

Pelo agrupamento dos fatores regionalizados $Y_k^s(\mathbf{u})$ com a mesma função variograma $\gamma_s(\mathbf{h})$, cada função aleatória $Z_i(\mathbf{u})$ pode ser expressa como uma soma de funções aleatórias ortogonais $Z_i^s(\mathbf{u})$ chamadas de *componentes espaciais*:

$$Z_i(\mathbf{u}) = \sum_{s=0}^S Z_i^s(\mathbf{u}) \quad \text{com} \quad Z_i^s(\mathbf{u}) = \sum_{k=1}^p a_{ik}^s Y_k^s(\mathbf{u}); \quad (3.61)$$

onde $Z_i^s(\mathbf{u})$ representa o comportamento da função aleatória $Z_i(\mathbf{u})$ em uma dada escala espacial s . A equação (3.59) torna-se:

$$\gamma_{ij}(\mathbf{h}) = \sum_{s=0}^S \gamma_{ij}^s(\mathbf{h}) \quad \text{com} \quad \gamma_{ij}^s(\mathbf{h}) = b_{ij}^s \gamma_s(\mathbf{h}); \quad (3.62)$$

onde $\gamma_{ij}^s(\mathbf{h})$ são os variogramas cruzados entre as componentes espaciais $Z_i^s(\mathbf{u})$ e $Z_j^s(\mathbf{u})$.

O modelo linear de co-regionalização (equação 3.58) e a equação (3.60) podem ser reescritos usando notação matricial:

$$\mathbf{Z}(\mathbf{u}) = \sum_{s=0}^S \mathbf{Y}^s(\mathbf{u})(\mathbf{A}^s)^T + \mathbf{m}. \quad (3.63)$$

$$\mathbf{B}^s = \mathbf{A}^s(\mathbf{A}^s)^T; \quad (3.64)$$

com $\mathbf{Z}(\mathbf{u}) = [Z_1(\mathbf{u}), \dots, Z_p(\mathbf{u})]$, $\mathbf{Y}^s(\mathbf{u}) = [Y_1^s(\mathbf{u}), \dots, Y_p^s(\mathbf{u})]$ e $\mathbf{A}^s(i,j) = a_{ij}^s$, $\mathbf{m} = [m_1, m_2, \dots, m_p]$ e \mathbf{B}^s é uma matriz positiva semi-definida de coeficientes b_{ij}^s chamada matriz de co-regionalização.

Assumindo uma estacionaridade de segunda ordem e comportamento assintótico da função variográfica básica $\gamma_s(\mathbf{h})$, isto é, $\gamma_s(\mathbf{h}) \rightarrow 1$ conforme $|\mathbf{h}| \rightarrow \infty$, a matriz de variância-covariância \mathbf{V} (da estatística clássica) é relacionada às matrizes \mathbf{B}^s (GOOVAERTS, 1992) por:

$$\mathbf{V} = \sum_{s=0}^S \mathbf{B}^s. \quad (3.65)$$

A equação (3.65) mostra que a estrutura de correlação descrita por \mathbf{V} é uma mistura de estruturas de correlação encontradas em diferentes escalas espaciais. Conseqüentemente \mathbf{V} não fornece uma idéia precisa das inter-relações entre as variáveis quando a estrutura de correlação muda com a escala espacial. Enquanto a análise multivariada clássica é realizada sobre a matriz \mathbf{V} de variância – covariância do conjunto de variáveis originais $Z_i(\mathbf{x})$, a krigagem fatorial multivariada trabalha com as matrizes de co-regionalização \mathbf{B}^s que descrevem as estruturas de correlações em diferentes escalas espaciais. GOOVAERTS (1992, 1993) mostra que componentes principais usados em métodos multivariados clássicos são combinações lineares dos

fatores regionalizados e, desta forma, são misturas das principais características dos dados para diferentes escalas espaciais.

A correlação linear entre quaisquer dos dois componentes espaciais – $Z_i^s(\mathbf{u})$ e $Z_j^s(\mathbf{u})$ – é medida pelo coeficiente de correlação estrutural definido como:

$$\rho_{ij}^s = \text{Corr}\{Z_i^s(\mathbf{u}), Z_j^s(\mathbf{u})\} = \frac{b_{ij}^s}{\sqrt{b_{ii}^s \cdot b_{jj}^s}}. \quad (3.66)$$

A matriz de coeficientes de correlação estrutural é indicada por $\mathbf{R}_s = [\rho_{ij}^s]$.

A krigagem fatorial multivariada estima a natureza regionalizada das variáveis pela análise de cada matriz de co-regionalização \mathbf{B}_s , ou pela matriz de coeficientes de correlação estrutural \mathbf{R}_s , separadamente. Deste modo, cada estrutura de correlação é distinguida pela filtragem das estruturas pertencentes a outras escalas de variação espacial.

Os seguintes procedimentos são utilizados na krigagem fatorial multivariada:

1 – Inicialmente deve-se fazer o modelamento da co-regionalização que envolve, na prática, os seguintes passos:

- i) computar os $p(p + 1)/2$ semivariogramas experimentais diretos e cruzados;
- ii) escolher o número (S) e as características (tipo e alcance) dos semivariogramas básicos $\gamma_s(\mathbf{h})$. WEBSTER *et alli.*, (1994) recomendam que, antes da tentativa de ajuste de um modelo de co-regionalização, se realize uma análise de componentes principais na matriz de variância-covariância \mathbf{V} (ou matriz de correlações \mathbf{R}) das variáveis originais e seja avaliada a estrutura de correlação espacial das componentes principais;
- iii) ajustar o modelo selecionado aos valores experimentais, ou seja, estimar os coeficientes b_{ij}^s sob a restrição de positivo semidefinida das matrizes de co-regionalização \mathbf{B}^s (PARDO-IQUIZGUIZA e DOWD, 2002).

2 – Aplicar métodos multivariados, geralmente análise de componentes principais, em cada matriz \mathbf{B}^s . Os elementos de cada matriz de co-regionalização \mathbf{B}^s representam as contribuições relativas de cada modelo básico $\gamma_s(\mathbf{h})$ dos semivariogramas diretos e cruzados. A aplicação da análise de componentes principais na matriz de co-regionalização \mathbf{B}^s conduz à seguinte decomposição espectral:

$$\mathbf{B}^s = \mathbf{Q}^s \mathbf{\Lambda}^s (\mathbf{Q}^s)^T = \mathbf{A}^s (\mathbf{A}^s)^T \quad \text{com} \quad \mathbf{A}^s = \mathbf{Q}^s (\mathbf{\Lambda}^s)^{1/2}. \quad (3.67)$$

Onde $\mathbf{Q}^s = [q_{ik}^s]$ é a matriz ortogonal de auto-vetores e $\mathbf{\Lambda}^s = \lambda_k^s$ é a matriz diagonal de auto-valores. O coeficiente de correlação entre a componente espacial $Z_i^s(\mathbf{u})$ e o fator regionalizado $Y_k^s(\mathbf{u})$ é dado por:

$$\rho_{ik}^s = \frac{q_{ki} \sqrt{\lambda_k^s}}{\sqrt{b_{ii}^s}}. \quad (3.68)$$

Onde b_{ii}^s é o *sill* do modelo básico $\gamma(\mathbf{h})$ do semivariograma direto. A representação gráfica das correlações mostra as inter-relações entre as variações dos metais em função da escala espacial.

3 – As variáveis $Y_k^s(\mathbf{u})$ (fatores regionalizados) são estimadas por cokrigagem ordinária como (PARDO-IQUIZGUIZA e DOWD, 2002):

$$\hat{Y}_k^s(\mathbf{u}) = \sum_{i=1}^p \sum_{\alpha=1}^n \lambda_{i\alpha}(\mathbf{u}) Z_i(\mathbf{u}_\alpha); \quad (3.69)$$

onde $\hat{Y}_k^s(\mathbf{u})$ é o estimador de cokrigagem ordinária do k -ésimo fator regionalizado na s -ésima escala espacial, p é o número de variáveis, n é o número de dados vizinhos à \mathbf{u}_α usados na estimativa e $\lambda_{i\alpha}(\mathbf{u})$ é o peso determinado para o valor da i -ésima variável na α -ésima localização, $Z_i(\mathbf{u}_\alpha)$. A minimização da estimativa da variância sob a restrição de não-tendenciosidade conduz a um sistema de $p(n+1)$ equações lineares, cuja solução fornece os pesos $\lambda_{i\alpha}(\mathbf{u})$. O sistema a ser resolvido é:

$$\sum_{i=1}^p \sum_{\beta=1}^n \lambda_{\beta i} \gamma_{ij}(\mathbf{u}_\alpha - \mathbf{u}_\beta) - \mu_i = a_{ik}^s \gamma_s(\mathbf{u}_\alpha - \mathbf{u}_0).$$

$$\sum_{\beta=1}^n \lambda_{\beta i} = 0 \quad \text{para } i = 1, \dots, p \quad \text{e } \alpha = 1, \dots, n.$$

O estimador é não tendencioso como uma consequência das condições impostas sobre os pesos. A condição de que a soma dos pesos seja zero significa que a média local de cada fator regionalizado é estabelecida como zero.

4 – Do mesmo modo, uma estimativa não tendenciosa de uma componente espacial $Z_i^s(\mathbf{u}_0)$ é obtida por co-krigagem como (PARDO-IQUIZGUIZA e DOWD, 2002):

$$\hat{Z}_i^s(\mathbf{u}_0) = \sum_{i=1}^p \sum_{\alpha=1}^n \lambda_{i\alpha}(\mathbf{u}) Z_i(\mathbf{u}_\alpha). \quad (3.70)$$

O sistema a ser resolvido é:

$$\sum_{i=1}^p \sum_{\beta=1}^n \lambda_{\beta i} \gamma_{ij}(\mathbf{u}_\alpha - \mathbf{u}_\beta) - \mu_i = b_{ik}^s \gamma_s(\mathbf{u}_\alpha - \mathbf{u}_0).$$

$$\sum_{\beta=1}^n \lambda_{\beta i} = 0 \quad \text{para } i = 1, \dots, p \quad \text{e } \alpha = 1, \dots, n;$$

onde $\gamma_s(\mathbf{u}_\alpha - \mathbf{u}_0)$ é o valor do s -ésimo semivariograma básico entre o α -ésimo ponto amostral e \mathbf{u}_0 , o ponto interpolado. O termo $b_{ik}^s \gamma_s(\mathbf{u}_\alpha - \mathbf{u}_0) = \gamma_{ik}^s(\mathbf{u}_\alpha - \mathbf{u}_0)$ é a semi-variância cruzada entre $Z_i(\mathbf{u}_\alpha)$ e $Z_i^s(\mathbf{u}_0)$. A condição de que a soma dos pesos seja zero significa que a média local de cada componente espacial é assumida como zero.

3.3.8. Análise da incerteza

A presença da incerteza torna a tomada de decisão uma tarefa difícil, mas inevitável em processos de remediação ambiental. A incerteza, entretanto, deve ser considerada em virtude do desconhecimento das reais concentrações de elementos tóxicos em localizações não amostradas. A quantificação da incerteza é feita por uma distribuição estatística que descreve as frequências de resultados na localização não amostrada. Aquela função de distribuição dependerá do número de dados na amostra, de sua configuração espacial, dos valores dos dados e do fenômeno espacial específico em estudo. Uma vez quantificada a distribuição, podem ser calculados os intervalos de probabilidade ou outras medidas de risco para determinadas aplicações (CAERS, 2001). A única forma de se reduzir o nível de incerteza é pela coleta de amostras adicionais e, mesmo assim, o nível de incerteza nunca será zero para um número finito de amostras (ISTOK e RAUTMAN, 1995). Pode-se, entretanto, minimizar o erro que se comete por causa de sua presença. A incorporação da incerteza é feita a partir da disponibilidade de um modelo de função de distribuição acumulada condicional, $F(\mathbf{u};z|(n))$, da(s) variável(eis) em estudo, e a minimização do erro pela utilização de funções de perda. A função $F(\mathbf{u};z|(n))$ pode ser obtida a partir dos valores amostrais e um modelo de dependência espacial (incerteza local) ou com base em um conjunto de simulações estocásticas para a(s) variável(eis) em estudo (incerteza espacial), onde cada simulação fornece um “mapa” de concentrações do contaminante na área de estudo, que é consistente com as concentrações conhecidas, com o histograma amostral e com os padrões de continuidade espacial exibidos pelos dados.

O modelamento da incerteza local refere-se à análise da incerteza sobre o valor da variável z , em cada localização não amostrada \mathbf{u} , quando realizada para as estimativas $z^*(\mathbf{u})$ obtidas pela krigagem e o respectivo erro de variância associado ou através da função de distribuição acumulada condicional, $F(\mathbf{u};z|(n))$ obtida com base em modelo semivariográfico. Nesse caso, o modelamento da incerteza sobre o valor desconhecido $z(\mathbf{u})$ pode ser feito através da FDAC, definida na equação (3.5), antes e independente da escolha de uma determinada estimativa. A FDAC modela completamente a incerteza na localização \mathbf{u} , uma vez que fornece a probabilidade do valor desconhecido ser menor (ou maior) do que um determinado valor de corte.

Um possível modelo *a priori* para a função de distribuição acumulada $Z(\mathbf{u})$ é a função de distribuição acumulada experimental (histograma acumulado) dos n valores observados:

$$F(\mathbf{u}, z) = \text{Pr ob}\{Z(\mathbf{u}) \leq z\} = \frac{1}{n} \sum_{\alpha=1}^n i(\mathbf{u}_{\alpha}; z) = \text{proporção de dados } z(\mathbf{u}_{\alpha}) \leq z; \quad (3.71)$$

onde a função indicativa i (do valor de corte z) é definida como:

$$i(z; z(\mathbf{u}_{\alpha})) = \begin{cases} 1, & \text{se } z(\mathbf{u}_{\alpha}) \leq z \\ 0, & \text{se } z(\mathbf{u}_{\alpha}) > z \end{cases} \quad (3.72)$$

O modelo *a priori* (3.71) considera pesos iguais ($1/n$) para a contribuição de cada dado e é independente da localização dos dados. Porém, se existe um modelo de dependência espacial (semivariograma) para a variável $Z(\mathbf{u})$, um modelo alternativo para (3.70) seria dado por uma média ponderada, isto é:

$$F(\mathbf{u}, z | (n) = \text{Pr ob}\{Z(\mathbf{u}) \leq z | (n)\} = \sum_{\alpha=1}^n \lambda(\mathbf{u}_{\alpha}) i(\mathbf{u}_{\alpha}; z) \in [0,1]. \quad (3.73)$$

O modelo (3.73) é chamado distribuição *posterior* ou função de distribuição acumulada condicional, onde a notação “ $| (n)$ ” indica o condicionamento à informação local, por exemplo, os $n(\mathbf{u})$ dados vizinhos a $z(\mathbf{u}_{\alpha})$. É posterior ao conhecimento da informação constituída pelos n valores de dados $z(\mathbf{u}_{\alpha})$ e ao modelo particular de dependência espacial considerado, por exemplo, o semivariograma indicativo $\gamma(\mathbf{h}; z_k)$ para o valor de corte z_k , como definido na equação (3.52).

Os pesos $\lambda(\mathbf{u}_{\alpha})$ podem ser calculados por um sistema de krigagem ordinária semelhante ao sistema dado em (3.48), ou seja:

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}^{KO}(\mathbf{u}) C_I(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha}; z_k) + \mu(\mathbf{u}; z_k) = C_I(\mathbf{u} - \mathbf{u}_{\alpha}; z_k), & \alpha = 1, \dots, n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}^{KO}(\mathbf{u}; z_k) = 1 \end{cases} \quad (3.74)$$

A função de distribuição $F(\mathbf{u}; z|n) = \text{Prob}\{Z(\mathbf{u}) \leq z|n\}$ feita condicional à informação disponível (n) modela totalmente a incerteza sobre $z(\mathbf{u})$ no sentido que intervalos de probabilidade podem ser derivados, tais como:

$$\text{Prob}\{Z(\mathbf{u}) \in (a, b]|n\} = F(\mathbf{u}; b|n) - F(\mathbf{u}; a|n). \quad (3.75)$$

ou a probabilidade de $Z(\mathbf{u})$ ultrapassar um determinado valor de corte, de grande importância em muitas aplicações ambientais, dada por:

$$\text{Prob}\{Z(\mathbf{u}) > b|n\} = 1 - F(\mathbf{u}; b|n). \quad (3.76)$$

Em muitos processos de tomada de decisão, mapas gerados pela probabilidade (3.76) são suficientes para delimitação de áreas onde medidas de remediação precisam ser tomadas.

Por outro lado, pode-se obter a função de distribuição $F(\mathbf{u}; z|n)$ a partir de simulações estocásticas que honram os valores amostrais e reproduzem estatísticas consideradas importantes, como o histograma amostral e modelos semivariográficos. Neste caso, em cada simulação é definida uma distribuição conjunta em todos os locais da malha tomados ao mesmo tempo. Isso garante que a incerteza espacial seja modelada a partir de múltiplas realizações daquelas distribuições conjuntas de valores no espaço. GOOVAERTS (1977, p. 369) mostra as diferenças entre as simulações estocásticas e estimativas obtidas por krigagem. A principal característica apresentada relaciona-se à suavização apresentada nos mapas de krigagem, onde valores de concentração pequenos são superestimados e valores altos são subestimados.

3.3.8.1. Funções de perda e estimativas ótimas

Apesar da avaliação probabilística da incerteza por meio da função de distribuição acumulada condicional representar um critério para uma tomada de decisão, as medidas de incerteza local, em geral, devem ser suplementadas por uma estimativa $z^*(\mathbf{u})$ do valor desconhecido, isso por que na tomada de decisão raramente se pensa somente em termos de probabilidade. A partir do modelo de incerteza pela função de distribuição acumulada condicional, pode-se também derivar diferentes estimativas ótimas considerando-se diferentes critérios, cada um configurado conforme as

necessidades do problema em estudo, ao invés de se considerar a estimativa usual de erro de mínimos quadrados.

Quando um valor desconhecido $z(\mathbf{u})$ é estimado por um simples valor $z^*(\mathbf{u})$ provavelmente ocorre um erro, $e(\mathbf{u}) = z^*(\mathbf{u}) - z(\mathbf{u})$, diferente de zero. Na estimativa de uma concentração tóxica, por exemplo, uma subestimativa daquela concentração poderá causar danos à saúde ou levar a reclamações de seguros e processos. Uma superestimativa da concentração poderá causar custos e limpezas desnecessárias. Ou seja, uma classificação errônea da concentração sempre ocasionará perdas. Pode-se, portanto, dizer que a perda é uma função do erro, isto é, da diferença entre o valor estimado e o valor verdadeiro da concentração. Para avaliar o impacto ou perda associada com qualquer erro é utilizada uma função $L(e(\mathbf{u}))$ conhecida, chamada função de perda. A função de perda é um conceito importante, pois proporciona um critério de minimização para a classificação. Uma classificação é dita ótima se a perda esperada for mínima. Embora o argumento $[z^*(\mathbf{u}) - z(\mathbf{u})]$ da função de perda seja desconhecido, a incerteza sobre $z(\mathbf{u})$ é modelada pela função de distribuição acumulada condicional $F(\mathbf{u}; z | (n))$, que é disponível. A idéia é usar este modelo de incerteza para determinar a perda esperada:

$$\varphi_L(z^*(\mathbf{u}) | (n)) = E\{L[z^*(\mathbf{u}) - Z(\mathbf{u})] | (n)\} = \int_{-\infty}^{\infty} L(z^*(\mathbf{u}) - z) dF(\mathbf{u}; z | (n)). \quad (3.77)$$

Na prática, a seguinte soma discreta aproximada é usada:

$$\varphi_L(z^*(\mathbf{u}) | (n)) \approx \sum_{k=1}^{K+1} L(z^*(\mathbf{u}) - \bar{z}_k) \cdot [F(\mathbf{u}; z_k | (n)) - F(\mathbf{u}; z_{k-1} | (n))]. \quad (3.78)$$

Onde

- $z_k, k = 1, \dots, K$ são K valores de corte que discretizam a faixa de variação dos valores z . Por convenção, $F(\mathbf{u}; z_0 | (n)) = 0$ e $F(\mathbf{u}; z_{K+1} | (n)) = 1$. Outros cortes z_k poderiam ser identificados com quantis p correspondendo a incrementos regularmente espaçados da função de distribuição acumulada condicional, $z_k = F^{-1}(\mathbf{u}; k / [K+1] | (n))$;
- \bar{z}_k é a média da classe $(z_{k-1}, z_k]$, que depende do modelo de interpolação dentro das classes; por exemplo, no caso mais simples do modelo linear, $\bar{z}_k = (z_{k-1} + z_k) / 2$.

A estimativa ótima para a função de perda L é o valor de $z^*(\mathbf{u})$ que minimiza a perda esperada, isto é,

$$z_L^*(\mathbf{u}) \text{ é o valor } z^*(\mathbf{u}) \text{ que minimiza } L(z^*(\mathbf{u}) - z(\mathbf{u})).$$

O procedimento para a determinação de uma estimativa ótima é dado a seguir:

1. Primeiramente é modelada a incerteza sobre o valor desconhecido $z(\mathbf{u})$ pela função de distribuição acumulada condicional $F(\mathbf{u};z|(n))$.
2. Em seguida, uma estimativa $z_L^*(\mathbf{u})$ é deduzida de acordo com um critério específico de otimização.

Deste modo, para um determinado modelo de incerteza, diferentes estimativas podem ser obtidas dependendo da função de perda escolhida. A qualidade da estimativa vai depender de sua utilização, não existindo uma estimativa ótima para todos os propósitos (JOURNEL, 1987; SRIVASTAVA, 1987).

JOURNEL (1984) mostra as soluções, isto é, as estimativas ótimas, para o critério do desvio médio quadrático ou dos mínimos quadrados, relacionado a uma função de perda simétrica quadrática e critério do desvio médio absoluto, relacionado a funções de perda lineares simétrica e assimétricas:

1. Para o critério dos mínimos quadrados, a função de perda assume a forma $L(e(\mathbf{u})) = [e(\mathbf{u})]^2$ e a estimativa ótima, $z_L^*(\mathbf{u})$, é o valor esperado (média) da função de distribuição acumulada condicional $F(\mathbf{u};z|(n))$, também conhecida como estimativa “E-type”, definida na eq. 3.53:

$$z_L^*(\mathbf{u}) = z_E^*(\mathbf{u}) \approx \sum_{k=1}^{K+1} \bar{z}_k \cdot [F(\mathbf{u}; z_x | (n)) - F(\mathbf{u}; z_{k-1} | (n))]; \quad (3.79)$$

sendo \bar{z}_k a média da FDAC $F(\mathbf{u};z|(n))$ dentro da classe $(z_{k-1} - z_k]$.

Por causa da elevação ao quadrado do erro de estimação, valores extremos (muito grandes ou muito pequenos) tendem a ter grande impacto sobre a perda esperada (Fig. 3.8 a).

2. Para o critério do desvio médio absoluto, sendo a função de perda linear e simétrica, isto é, $L(e(\mathbf{u})) = |e(\mathbf{u})|$, a estimativa ótima $z_L^*(\mathbf{u})$ é a mediana da FDAC $F(\mathbf{u};z|(n))$, que corresponde ao quantil 0,5 definido como:

$$z_L^*(\mathbf{u}) = q_{0,5}(\mathbf{u}) = F^{-1}(\mathbf{u};0,5|(n)). \quad (3.80)$$

A utilização da função linear em lugar da função quadrática reduz o impacto das caudas do modelo sobre a perda esperada (Fig. 3.8 b).

3. Para o critério do desvio médio absoluto e com uma função de perda que aumenta linearmente com o erro, mas com taxas diferentes que dependem da ocorrência de uma superestimativa ou subestimativa (Figura 3.8, gráficos *c* e *d*), isto é:

$$L[e(\mathbf{u})] = \begin{cases} w_1(z^*(\mathbf{u}) - z), & \text{para } z^*(\mathbf{u}) \geq z \quad (\text{superestimativa}) \\ w_2(z - z^*(\mathbf{u})), & \text{para } z^*(\mathbf{u}) \leq z \quad (\text{subestimativa}) \end{cases}; \quad (3.81)$$

onde os parâmetros positivos w_1 e w_2 são os impactos relativos vinculados à superestimativa e subestimativa, respectivamente. Neste caso, a estimativa ótima $z_L^*(\mathbf{u})$ é o quantil p da FDAC $F(\mathbf{u};z|(n))$:

$$z_L^*(\mathbf{u}) = q_p(\mathbf{u}) = F^{-1}(\mathbf{u};p|(n));$$

onde
$$p = \frac{w_2}{w_1 + w_2} \in [0,1].$$

Três casos podem ser considerados:

- quando $w_1 = w_2$: a função linear de perda (3.81) é simétrica, a estimativa ótima é o quantil 0.5, isto é, a mediana.
- quando $w_1 > w_2$, $p < 0.5$, a estimativa ótima $q_p = F^{-1}(\mathbf{u};p|(n))$ é menor do que a mediana. O impacto da superestimativa é maior do que da subestimativa de mesma magnitude. Em termos práticos, um custo mais alto de superestimativa poderia sugerir uma estimativa mais conservativa, e a melhor estimativa poderia conduzir à mediana. Neste caso, a superestimativa das concentrações tóxicas poderia levar a limpezas desnecessárias.
- $w_1 < w_2$: o impacto da subestimativa é maior do que da superestimativa de mesma magnitude. Portanto, $p > 0.5$, e a estimativa ótima é maior do que a mediana. Neste caso, a subestimativa de concentrações tóxicas acarretaria grande impacto na saúde da população por causa do risco de se deixar localizações contaminadas sem tratamento ou remediação.

Os gráficos *c* e *d* da Figura 3.8 mostram as funções de perda assimétricas, onde as estimativas ótimas são representadas pelo quantil p , q_p , da $F(\mathbf{u};p|(n))$, que pode ser interpretado como o valor q_p , que tem probabilidade p de ser superado pelo valor desconhecido:

$$\text{Prob}\{Z(\mathbf{u}) > q_p(\mathbf{u})|n\} = p.$$

No caso de $w_1 > w_2$, e, portanto, $p < 0,5$, uma estimativa alta de $q_p(\mathbf{u})$ indica que, provavelmente, o valor desconhecido é alto na localização \mathbf{u} . Quando $w_1 < w_2$, e, então $p > 0,5$, uma estimativa baixa de $q_p(\mathbf{u})$ indica que, provavelmente, o valor desconhecido é pequeno na localização \mathbf{u} .

O aumento do contraste entre os impactos relativos w_1 e w_2 produz estimativas quantis que correspondem a probabilidades acumuladas p muito baixas ou muito altas. O desafio é quantificar a atitude mais apropriada frente ao risco, isto é, a função de perda.

As distribuições espaciais de concentrações tóxicas em aplicações ambientais em geral são fortemente assimétricas. Por isso, funções assimétricas simples do tipo (3.81) deveriam ter preferência em lugar do critério usual de mínimos quadrados que penaliza igualmente a subestimativa e superestimativa. Por exemplo, em casos de decisão de limpeza de determinada área, o impacto de subestimativa da poluição (uma área contaminada não é limpa) é normalmente muito mais alto do que a superestimativa (uma área segura é limpa indevidamente). O impacto do erro de estimativa pode ainda variar de uma área a outra, por exemplo, a subestimativa de uma concentração tóxica é provavelmente mais prejudicial em áreas residenciais do que em pátios industriais. Pode-se definir uma função perda $L(e(\mathbf{u}))$ que mude localmente, dependendo se \mathbf{u} é uma área industrial ou residencial. Nesses casos, uma estimativa usual do tipo mínimos quadrados poderia ser imprópria. O conceito rigoroso de função de perda e estimativa ótima pode permitir maior consistência na tomada de decisão, uma vez que uma particular e possivelmente subjetiva função de perda tenha sido escolhida.

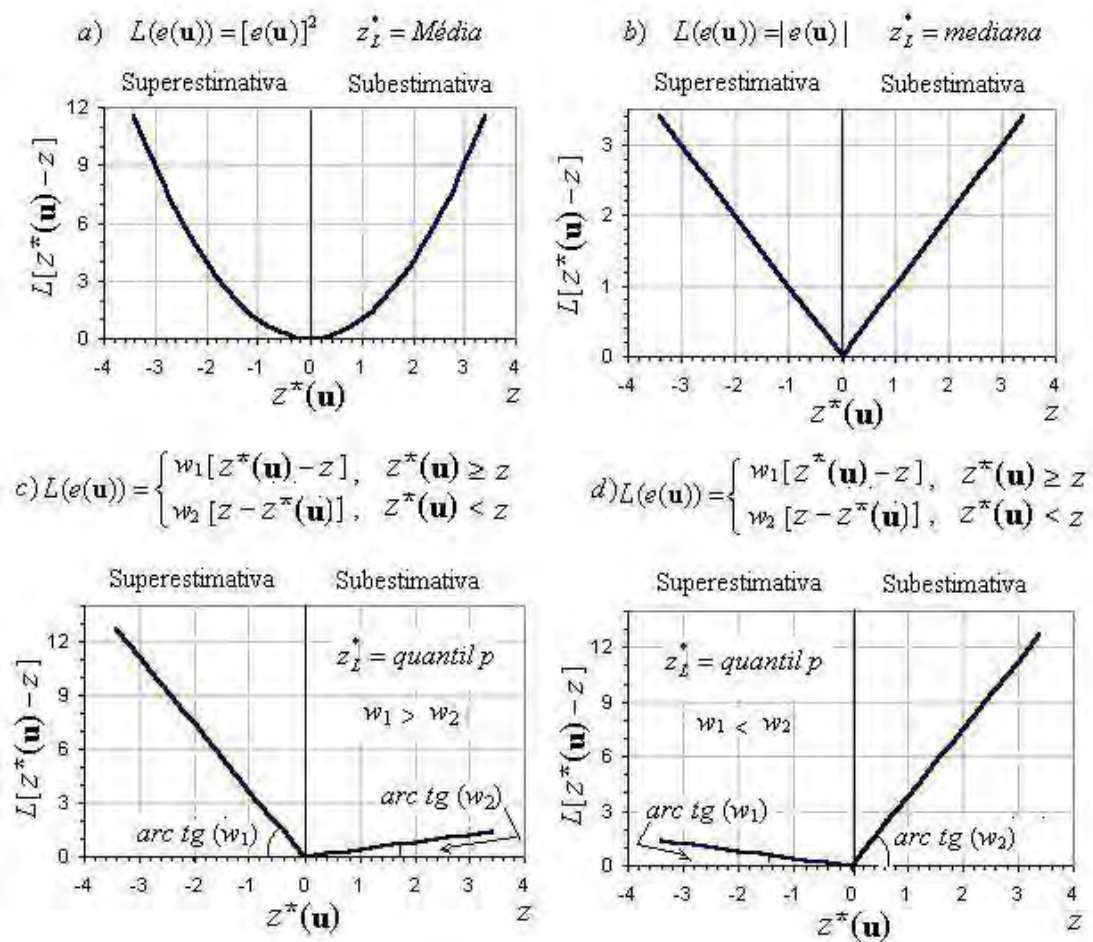


Figura 3.8 – Funções de perda e estimativas ótimas, $z_L^*(\mathbf{u})$, para o critério dos mínimos quadrados e desvio médio absoluto (GOOVAERTS, 1997, p.342).

Outro modo de considerar a incerteza no processo de tomada de decisão consiste em avaliar o impacto econômico de duas possíveis decisões usando o conceito de função de perda. Neste caso, a decisão é baseada sobre custos financeiros. A perda (custo) incorre somente no caso de uma decisão incorreta. Existem várias funções de perda que podem ser utilizadas nas quais são incorporados aspectos econômicos (GOOVAERTS, 1999 e 1998; MYERS, 1997, p.477, SRIVASTAVA, 1987). Em problemas de remediação ambiental, funções assimétricas são mais adequadas, mesmo que a escolha das funções a serem utilizadas contenha alguma subjetividade. As funções de perda assimétricas mostradas nos gráficos *c* e *d* da Figura 3.8 consideram que o custo da superestimativa, ou seja, de tratamento desnecessário, é linear aumentando a partir da origem. Em muitas situações, entretanto, o procedimento de limpeza (tratamento) consiste somente na remoção do solo classificado como contaminado, independente do valor da concentração do contaminante. Neste caso, pode-se usar uma função de perda

onde o custo de tratamento seja fixado e o custo de saúde seja proporcional à contaminação GOOVAERTS (1997, p.350), apresentado a seguir.

A perda associada com a classificação (errada) da localização \mathbf{u} como segura pode ser modelada como:

$$L_1(z(\mathbf{u})) = \begin{cases} 0 & \text{se } z(\mathbf{u}) \leq z_c \text{ (classificação correta)} \\ w_1[z(\mathbf{u}) - z_c] & \text{se } z(\mathbf{u}) > z_c \end{cases}; \quad (3.82)$$

onde w_1 é o custo relativo para o potencial de danos de saúde (w_1 pode indicar unidades monetárias/concentração, como dólar/ppm), e z_c é o limite de corte. Se a localização \mathbf{u} é realmente segura, $z(\mathbf{u}) \leq z_c$, então a classificação é correta e não há perda. Se na realidade a localização é contaminada, $z(\mathbf{u}) > z_c$, o custo de classificação errada é modelado como proporcional à real contaminação $[z(\mathbf{u}) - z_c]$.

A perda associada com a classificação (errada) de uma localização \mathbf{u} como contaminada pode ser modelada como

$$L_2(z(\mathbf{u})) = \begin{cases} 0 & \text{se } z(\mathbf{u}) > z_c \text{ (classificação correta)} \\ w_2 & \text{se } z(\mathbf{u}) \leq z_c \end{cases}; \quad (3.83)$$

Se a localização \mathbf{u} é realmente contaminada, $z(\mathbf{u}) > z_c$, então a classificação é correta e não há perda. Se a localização na realidade é segura, $z(\mathbf{u}) \leq z_c$, o custo de tratamento é modelado como um valor constante w_2 , que indica a penalidade imposta pelo tratamento desnecessário da localização. Por exemplo, se o tratamento consiste na remoção da camada superior de solo, o custo será independente da concentração real e w_2 é em unidades monetárias somente.

A Figura 3.9 mostra, como exemplo, a função de perda especificada no caso do manganês, para $w_1 = 1$ (equivalente a uma inclinação de 45° para L_1), $w_2 = 2,5$ e o limite máximo permitido, $z_c = 0,1$ ppm.

O modelo de $F(\mathbf{u}; z | (n))$ permite determinar a perda esperada vinculada aos dois tipos de classificação como:

$$\varphi_1(\mathbf{u}) = E[L_1(z(\mathbf{u})) | (n)] = \int_{-\infty}^{+\infty} L_1(z(\mathbf{u})) dF(\mathbf{u}; z | (n)).$$

$$\varphi_2(\mathbf{u}) = E[L_2(z(\mathbf{u})) | (n)] = \int_{-\infty}^{+\infty} L_2(z(\mathbf{u})) dF(\mathbf{u}; z | (n)).$$

Na prática, essas integrais são aproximadas como:

$$\varphi_1(\mathbf{u}) \cong \sum_{k=1}^{K+1} [L_1(\bar{z}_k)[F(\mathbf{u}; z_k | (n)) - F(\mathbf{u}; z_{k-1} | (n))]. \quad (3.84)$$

$$\varphi_2(\mathbf{u}) \cong \sum_{k=1}^{K+1} [L_2(\bar{z}_k)[F(\mathbf{u}; z_k | (n)) - F(\mathbf{u}; z_{k-1} | (n))]. \quad (3.85)$$

onde:

- $z_k, k = 1, \dots, K$ são K valores de corte discretizando a faixa de variação dos valores z . Por convenção, $F(\mathbf{u}; z_0 | (n)) = 0$ e $F(\mathbf{u}; z_{K+1} | (n)) = 1$. Outros valores de corte z_k poderiam ser identificados com quantis p correspondendo a incrementos regularmente espaçados de $z_k = F^{-1}(\mathbf{u}; k / [K+1] | (n))$.
- \bar{z}_k é a média da classe $(z_{k-1}, z_k]$, que depende do modelo de interpolação dentro das classes; por exemplo, para o modelo linear, $\bar{z}_k = (z_{k-1} + z_k)/2$.

A localização \mathbf{u} é então declarada segura ou contaminada de modo a minimizar a perda esperada resultante:

Se $\varphi_1(\mathbf{u}) > \varphi_2(\mathbf{u}) \Rightarrow \mathbf{u}$ é classificada como contaminada, pois $z(\mathbf{u}) > z_c$.

Se $\varphi_1(\mathbf{u}) < \varphi_2(\mathbf{u}) \Rightarrow \mathbf{u}$ é classificada como segura, uma vez que $z(\mathbf{u}) \leq z_c$.

Portanto, a decisão de limpeza de uma localização é tomada quando o custo esperado de saúde for maior do que o custo de tratamento desnecessário.

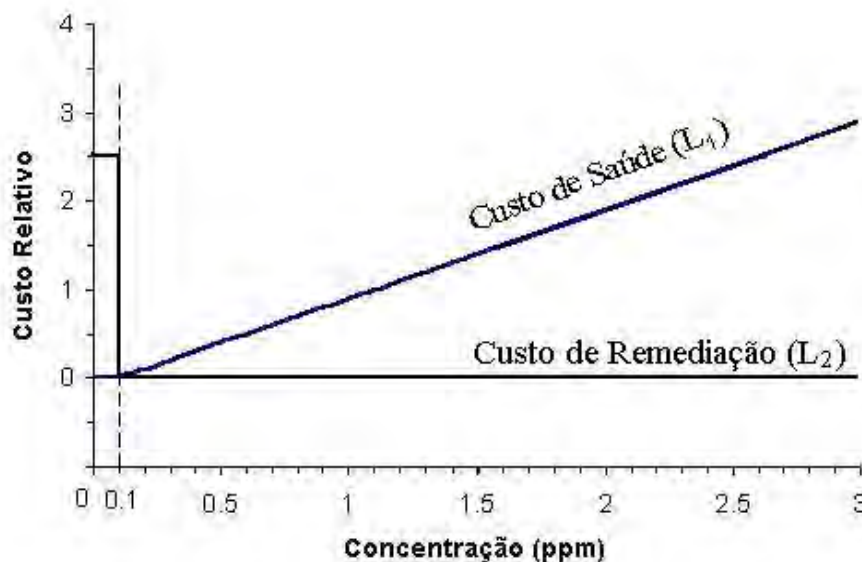


Figura 3.9 – Função de perda utilizada na avaliação do custo relativo para o manganês, com $w_1 = 1$, $w_2 = 2,5$ e $z_c = 0,1$ ppm.

3.3.8.2. Erros de classificação

A disponibilidade de modelos $F(\mathbf{u};z|(n))$ para cada localização \mathbf{u} dentro da área A em estudo permite que se obtenham estimativas da variável em localizações não amostradas. Entretanto, diferentes critérios de otimização e interpolação dos modelos da $F(\mathbf{u};z|(n))$ conduzem a estimativas diferentes, que podem ou não exceder um valor de corte z_c . Existe, portanto, um risco de se classificar como segura uma localização contaminada e, do mesmo modo, se classificar contaminada uma localização segura. Esses dois riscos ou erros de classificação podem ser avaliados a partir do modelo da $F(\mathbf{u};z|(n))$. Uma localização é classificada como segura quando a estimativa calculada para a mesma se encontra abaixo do limite máximo permitido (z_c) para o contaminante de interesse. Portanto, essa localização não estará sujeita a nenhum tratamento ou remediação. Caso contrário, a localização será classificada como contaminada e estará sujeita a tratamento.

O erro tipo I, também referido como risco $\alpha(\mathbf{u})$ ou falso positivo, ocorre quando a estimativa em uma localização segura \mathbf{u} ($Prob Z(\mathbf{u}) \leq z_c$) é superestimada ($z_L^*(\mathbf{u}) > z_c$), isto é, seu valor fica acima do limite máximo permitido e, conseqüentemente, a referida localização é classificada erroneamente como contaminada, acarretando limpezas desnecessárias, ou seja:

$$\begin{aligned} \alpha(\mathbf{u}) &= Prob\{Z(\mathbf{u}) \leq z_c \mid z_L^*(\mathbf{u}) > z_c, (n)\} \\ &= F(\mathbf{u}; z_c \mid (n)) \end{aligned} \quad (3.86)$$

para todas as localizações \mathbf{u} tal que a estimativa $z_L^*(\mathbf{u}) > z_c$.

O erro tipo II, referido como risco $\beta(\mathbf{u})$ ou falso negativo, ocorre se a estimativa em uma localização contaminada \mathbf{u} ($Prob Z(\mathbf{u}) > z_c$) é subestimada ($z_L^*(\mathbf{u}) \leq z_c$), isto é, seu valor fica abaixo do limite máximo permitido e, conseqüentemente, a referida localização é classificada erroneamente como segura, levando a riscos potenciais de saúde (de magnitude desconhecida), ou seja:

$$\begin{aligned} \beta(\mathbf{u}) &= Prob\{Z(\mathbf{u}) > z_c \mid z_L^*(\mathbf{u}) \leq z_c, (n)\} \\ &= 1 - F(\mathbf{u}; z_c \mid (n)) \end{aligned} \quad (3.87)$$

para todas as localizações \mathbf{u} tal que a estimativa $z_L^*(\mathbf{u}) \leq z_c$.

Pode-se observar que o risco $\beta(\mathbf{u})$ é definido somente nas localizações \mathbf{u} onde $z_L^*(\mathbf{u}) \leq z_c$, ou seja, onde o risco $\alpha(\mathbf{u})$ não é definido, e vice-versa. Em qualquer

localização particular \mathbf{u} , a *magnitude* do risco de classificação errada depende do modelo da $F(\mathbf{u};z|(n))$, não da estimativa particular considerada. O tipo de risco (α ou β) de classificação errada em \mathbf{u} depende, entretanto, da estimativa $z_L^*(\mathbf{u})$, isto é, do critério de otimização. Conforme a função de perda penaliza a subestimativa mais severamente, as estimativas tendem a ser maior (Figura 3.8d), limitando as áreas com risco $\beta(\mathbf{u})$ ao custo de maior risco potencial $\alpha(\mathbf{u})$.

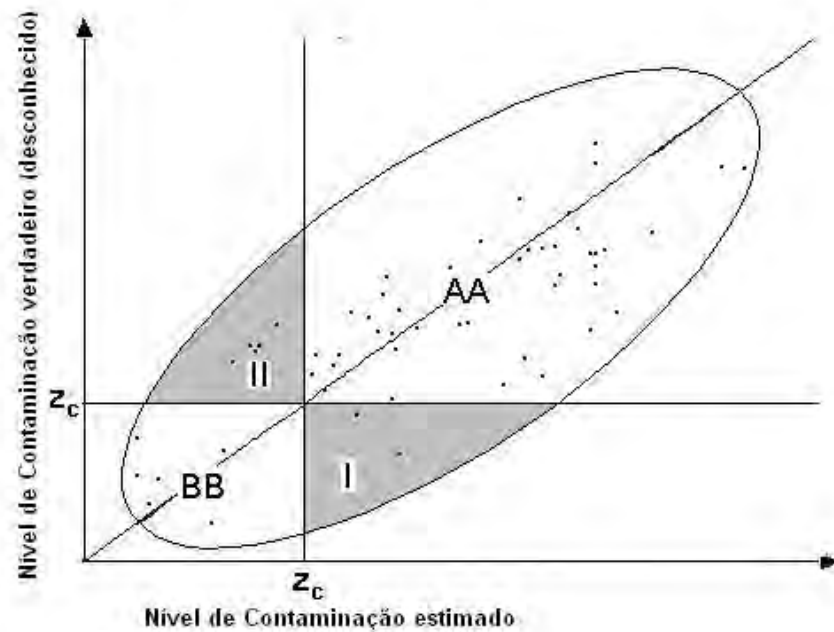
Os riscos de classificação errada podem ser usados para classificar localizações candidatas para amostragem adicional: localizações com mais altos riscos $\beta(\mathbf{u})$ são as primeiras candidatas a amostragem adicional.

A Figura 3.10 mostra uma representação visual das classificações corretas e incorretas mencionadas acima. Nesse gráfico, as concentrações estimadas são plotadas no eixo horizontal e as concentrações verdadeiras são plotadas no eixo vertical. Se os valores estimados fossem iguais aos valores verdadeiros, os pontos no gráfico ficariam sobre uma linha de 45° desenhada a partir da origem (0,0) do gráfico. Como os valores verdadeiros e estimados geralmente apresentam valores diferentes, o gráfico dos valores verdadeiros versus os valores estimados ficam dispersos em torno da linha de 45° . O valor de corte (limite máximo permitido), z_c , também é plotado perpendicularmente aos eixos dos valores verdadeiros e estimados. E pode-se observar 4 quadrantes de classificação, que correspondem aos tipos de erros e classificações corretas (Figura 3.10).

O quadrante inferior, denominado BB, contém as concentrações cujos valores verdadeiros e estimados encontram-se abaixo do valor de corte, z_c , de modo que, neste caso, as áreas foram corretamente classificadas como seguras. No quadrante superior, denominado AA, as concentrações apresentam valores verdadeiros e estimados acima do valor de corte, z_c , sendo, portanto, as áreas classificadas corretamente como contaminadas e sujeitas à remediação.

Os outros dois quadrantes, denominados I e II, representam os erros tipo I e II. O quadrante correspondente ao erro tipo I mostra as concentrações estimadas acima do valor de corte, z_c , embora os valores verdadeiros estejam abaixo de z_c . Como a tomada de decisão é baseada no valor estimado, as áreas correspondentes a essas concentrações serão classificadas erroneamente como contaminadas, pois na verdade estão limpas, acarretando custos caso haja tratamento (ou remediação) desnecessário. De outro modo, no quadrante relativo ao erro tipo II, as concentrações são estimadas abaixo do valor de

corde, z_c , enquanto os valores verdadeiros se encontram acima do mesmo. Ou seja, localizações contaminadas serão classificadas como limpas, colocando em risco a saúde da população e o meio ambiente.



- I = Erro Tipo I (Falso Positivo)
- II = Erro Tipo II (Falso Negativo)
- AA = Concentração Acima, Estimativa Acima (do valor de corte, z_c)
- BB = Concentração Abaixo, Estimativa Abaixo (do valor de corte, z_c)

Figura 3.10 – Erros de classificação e classificações corretas das localizações na área de interesse (MYERS, 1997, p.463).

3.3.8.3. Modelo de decisão baseado em risco custo-benefício

FREEZER *et al.* (1990) apresentam um modelo de decisão onde tentam quantificar vários tipos de custos, riscos e vantagens associadas com o planejamento e implementação de determinados projetos de engenharia em hidrogeologia ambiental. O objetivo da análise de decisão econômica é maximizar uma função, denominada função objetivo, que representa o retorno monetário geral associado com determinado projeto. Essa função objetivo é computada como o valor líquido de todos os rendimentos ou vantagens resultantes do projeto menos todos os custos ou despesas relevantes incluindo o custo esperado da incerteza da performance do determinado projeto. A função objetivo, neste caso, é definida como:

$$F_j = \sum_{t=1}^T \frac{1}{(1+i)^t} [B_j(t) - C_j(t) - R_j(t)]. \quad (3.88)$$

onde $B_j(t)$ são vantagens econômicas, $C_j(t)$ são custos econômicos e $R_j(t)$ são riscos econômicos (todos em unidades monetárias) associados com um determinado projeto, j , como uma função do tempo, t .

i é a taxa de interesse.

t é o tempo de duração do projeto, companhia, etc. (geralmente 5-7 anos).

$\frac{1}{(1+i)^t}$ é o termo taxa de desconto para a atualização do fluxo de caixa.

Os projetos de remediação ambiental são financiados em períodos curtos de tempo, em geral anualmente, de modo que o termo taxa de desconto pode ser abandonado na equação (3.88). Além disso, como em projetos de remediação ambiental raramente existem “vantagens econômicas” no sentido convencional de um fluxo de rendimentos, pode-se pensar na função objetivo em termos de minimização de custos (em vez de maximização de lucros), que pode ser estabelecida matematicamente conforme RAUTMAN (1996):

$$\text{Minimizar } \Phi = C_{total_i} = C_{carac_i} + C_{trat_i} + E\{C_{falha_i}\}, \quad (3.89)$$

onde C_{total_i} é custo total associado com a caracterização e remediação do local pelo uso de determinada tecnologia i . Esse custo compreende:

- Caracterização (C_{carac}): relacionada a custos com pessoal de campo, análises das amostras e interpretação de dados. Esses custos são fixados de modo que o custo de caracterização (C_{carac}) pode ser considerado constante;
- Tratamento (C_{trat}): são os custos envolvidos no tratamento ou remediação, como escavação, deslocamento e disposição do material contaminado;
- Falha ($E\{C_{falha}\}$): este termo representa os riscos econômicos associados com a probabilidade de falha tendo como conseqüências litígios, multas, penalidades e riscos à saúde. O custo esperado de falha, $E\{C_{falha}\}$, é definido como:

$$E\{C_{falha}\} = P_{falha} \cdot C_{falha}, \quad (3.90)$$

onde P_{falha} é a probabilidade de falha e C_{falha} é o custo incorrido se a falha realmente ocorre. No contexto da remediação ambiental os tipos de falha que podem ocorrer são a remoção (ou tratamento) de um solo não contaminado (falso positivo) ou a não remoção de um solo contaminado (falso negativo).

O custo total é uma função não linear dos dois tipos de custos, C_{trat_i} e $E\{C_{falha_i}\}$, de modo que o custo econômico de um projeto é minimizado pelo intercâmbio entre esses diferentes tipos de custos, um em relação ao outro. A Figura 3.11 ilustra a relação entre esses dois tipos de custos. A alternativa ótima é aquela que minimiza o custo total no ponto onde se cruzam os dois tipos de custo, e o risco associado com essa alternativa constitui o *risco ótimo* na perspectiva do responsável pelo tratamento – operador – (FREEZER *et al.*, 1994).

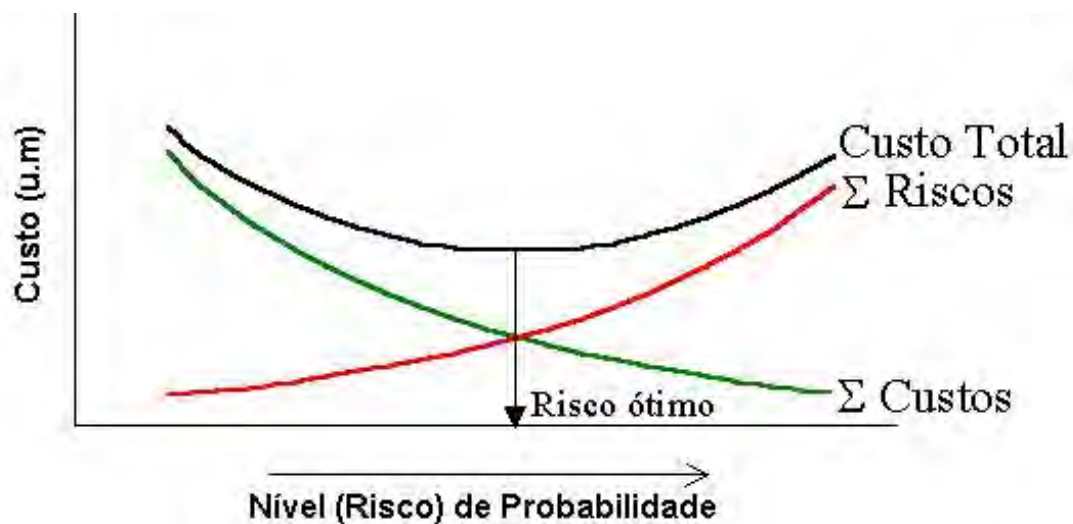


Figura 3.11 – Relação entre custo e risco; o conceito de risco ótimo.

Se o operador não quer aceitar qualquer risco de uma decisão de remediação incorreta, o nível de probabilidade, p , é igual a zero (Risco = 0) e não há alternativa a não ser tratar toda a área de estudo. Por exemplo, o solo contaminado e não contaminado seria inteiramente removido (se fosse este o tipo de tratamento) e, portanto, o risco de falha (risco à saúde) seria zero. Entretanto, o custo de tratamento seria muito alto. À medida que o nível de probabilidade aceitável aumenta (isto é, o operador torna-se menos relutante ao risco), o custo total diminui até atingir um valor mínimo (risco ótimo). Em seguida começa a aumentar até chegar ao nível de probabilidade igual a 1 (Risco = 100%) onde seriam tratados somente aqueles painéis com certeza absoluta de estarem contaminados, ou seja, nesse caso, somente os painéis relativos às amostras que excederem limite máximo permitido seriam removidos. Portanto, o risco de permanecerem painéis contaminados no local (relativos a pontos não amostrados) seria muito grande, levando a altos riscos à saúde.

Um *risco aceitável* é um conceito subjetivo. É estabelecido indiretamente em favor da sociedade pelas agências reguladoras. Se existe a percepção de que o risco ótimo na perspectiva do operador é maior do que o risco aceitável, então se deveria aumentar as penalidades de regulamentação, de modo que o risco ótimo ficasse mais próximo daquele ao qual a sociedade considera aceitável (FREEZER *et al.*, 1994).

3.3.9. Simulações estocásticas

A simulação estocástica representa uma metodologia para construção de resultados possíveis ou realizações alternativas da variável aleatória $Z(\mathbf{u})$ no modelamento da incerteza. As simulações são designadas como “simulações condicionais” se as realizações resultantes honram os valores das amostras em suas localizações específicas.

Embora a estimativa (krigagem) e a simulação geoestatística utilizem uma função estatística particular, o semivariograma, que mede a correlação espacial entre os dados, as simulações diferem da krigagem em muitos aspectos. A krigagem fornece a *melhor* estimativa linear não tendenciosa de uma variável regionalizada em uma localização não amostrada, calculada como uma combinação linear ponderada (média) de alguns dados medidos, localizados na vizinhança do dado a ser estimado, caracterizando a krigagem como uma técnica de suavização. A ênfase é sobre a precisão e acurácia (não tendenciosidade) local, ou proximidade da estimativa ao valor verdadeiro (desconhecido) sem considerar as propriedades estatísticas resultantes como um todo. As restrições para a minimização da variância do erro envolvem cada estimativa (cada nó da malha) separadamente, por isso chamadas *locais*. A simulação geoestatística é uma técnica alternativa de modelamento onde as estatísticas *globais* são mais importantes do que propriamente a acurácia local, porque envolvem todos os pontos da malha conjuntamente, como histograma e semivariograma dos valores modelados. A simulação tenta capturar as características essenciais dos dados, para reproduzi-las em um conjunto de imagens estocásticas alternativas, honrando os dados observados em suas localizações, chamados de mapas, simulações ou realizações da variável z , indicadas por $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in \text{área } A\}$, onde o l denota a l -ésima realização. Cada mapa é consistente com as concentrações de contaminantes mostradas pelo histograma amostral e com os padrões de continuidade espacial exibidos pelos dados através do semivariograma. A incerteza é incorporada através da análise do conjunto de

realizações geradas para o(s) contaminante(s) na área em estudo. Características espaciais, tais como seqüências de grandes valores, são consideradas certas se são observadas na maioria dos mapas simulados e consideradas incertas se, ao contrário, são vistas somente em poucos mapas simulados.

GOOVAERTS (1997) mostra que o uso de mapas suavizados produzidos por métodos de interpolação como krigagem é impróprio para aplicações sensíveis à presença de valores extremos (outliers) e seus padrões de continuidade, tais como a avaliação de fontes recuperáveis em depósitos minerais, modelamento de fluxos de fluidos em meios porosos ou delineamento de áreas contaminadas. A simulação estocástica tem sido crescentemente preferida à krigagem em aplicações onde a variação espacial deve ser preservada como no delineamento de áreas contaminadas.

Em muitos problemas práticos, as simulações estocásticas são pós-processadas, isto é, servem como entrada para funções de transferências, tais como simuladores de fluxo em engenharia de reservatórios ou processos de remediação em controle de poluição. A diferença entre funções de perda e funções de transferência é que essas consideram todas as localizações simultaneamente em vez de uma de cada vez. O processamento de cada realização de entrada produz um único valor para cada resposta, por exemplo, para o custo de remediação. A distribuição (histograma) dos L valores de resposta correspondendo às L realizações de entrada fornece uma medida de incerteza da resposta devido ao conhecimento imperfeito da distribuição espacial da variável z . Essa medida de incerteza pode então ser utilizada em análise de risco e tomada de decisão.

No problema da avaliação do impacto econômico de se declarar uma área segura em relação a determinado contaminante, em geral se deseja calcular o custo de não se tomar nenhuma medida de remediação. Esse custo pode ser computado a partir de qualquer realização $\{z^{(l)}(\mathbf{u}'_j), j = 1, \dots, N\}$ como a soma dos custos locais definidos nas N localizações \mathbf{u}'_j que discretizam a área de estudo A :

$$C(l) = \sum_{j=1}^N L_1(z^{(l)}(\mathbf{u}'_j)), \quad (3.91)$$

onde o custo associado L_1 de se declarar segura uma localização \mathbf{u}'_j pode ser dado pela expressão (3.76), isto é:

$$L_1(z^{(l)}(\mathbf{u}'_j)) = \begin{cases} 0 & \text{se } z^{(l)}(\mathbf{u}'_j) \leq z_c \\ w_1(\mathbf{u}'_j)[z^{(l)}(\mathbf{u}'_j) - z_c] & \text{se } z^{(l)}(\mathbf{u}'_j) > z_c \end{cases} \quad (3.92)$$

sendo z_c o valor de corte e $w_1(\mathbf{u}'_j)$ o custo relativo de subestimativa da concentração tóxica em \mathbf{u}'_j , ou seja, o custo associado com danos à saúde. Se o valor simulado $z^{(l)}(\mathbf{u}'_j)$ é menor do que o valor de corte z_c , a classificação é correta e não há custo. Caso contrário, o custo de classificação errônea será proporcional à contaminação $[z^{(l)}(\mathbf{u}'_j) - z_c]$. A aplicação da equação (3.92) nas simulações produzem mapas alternativos de custos de saúde que fornecem uma medida da incerteza sobre o impacto econômico de não se tomar nenhuma medida de remediação. A incerteza pode ser descrita pelo mapeamento da variância das distribuições dos custos em cada localização. O histograma dos custos, $C(l)$, $l=1, \dots, L$, computados para L realizações, fornece uma avaliação do risco envolvido na declaração de se considerar a área segura.

3.3.9.1. Tipos de Simulações condicionais

Em aplicações ambientais onde geralmente se trabalha com valores extremos, ou seja, com distribuições que apresentam fortes assimetrias, deve-se procurar fugir da exigência de hipóteses Gaussianas (distribuições normais) na análise dos dados. Entre os principais algoritmos de simulação que não são limitados ao formalismo Gaussiano temos a Simulação Sequencial Indicativa e a Simulação *Annealing*.

Existe um grupo de algoritmos de simulação conhecidos pelo nome genérico de algoritmos de *simulação seqüencial*. Uma função de distribuição condicional acumulada é modelada (através da krigagem) em cada ponto da malha, definidos em uma seqüência aleatória, utilizando-se os dados condicionantes, incluindo as amostras e os pontos previamente simulados. Deste modo, para garantir a reprodução do modelo de covariância (ou semivariograma), as funções de distribuição condicional acumulada de cada ponto são condicionais não somente aos n dados originais, mas também a todos os valores previamente simulados. As duas maiores classes de algoritmos de simulação seqüencial são classificadas de acordo com o modo com que suas séries de funções de distribuição condicional acumulada são determinadas. Quando usam o formalismo multi-Gaussiano são referidas como Simulação Sequencial Gaussiana (sgsim) e Simulação Sequencial Indicativa (sis) se empregam o formalismo indicativo. A

Simulação Sequencial Indicativa é uma técnica de simulação não-Gaussiana bastante utilizada em casos onde a análise estrutural (semivariográfica) ou informações qualitativas indicam que valores extremos poderiam mais correlacionados no espaço do que valores médios.

Na Simulação *Annealing* (SA), ao contrário de outros algoritmos de simulação, a criação de uma imagem estocástica é formulada como um problema de otimização sem referir-se a um modelo de função aleatória. Uma imagem inicial é gradualmente perturbada a fim de que se ajuste a restrições tais como a reprodução de um histograma e semivariograma (covariância) desejados, honrando os valores de dados em suas localizações. Diferentes realizações são obtidas a partir de diferentes imagens iniciais.

Pelas características dos dados a serem analisados, principalmente a presença de *outliers* e não normalidade das distribuições, optou-se, neste trabalho, pela utilização da simulação *annealing* no tratamento dos dados, devido à sua maior flexibilidade na incorporação de vários tipos de informação (como, padrões de continuidade espacial de classes específicas, dados *soft*, estatísticas de múltiplos-pontos) e na imposição de restrições adicionais para melhorar a reprodução das estatísticas desejadas.

3.3.9.2. Simulação *Annealing*

A simulação *annealing* é um nome genérico de uma família de algoritmos de otimização baseada no princípio de relaxação estocástica (GEMAN e GEMAN, 1984). Referências de aplicação dessa técnica em geoestatística podem ser vistas, por exemplo, em DEUTSCH e COCKERHAM (1994), DEUTSCH e JOURNAL (1996), GOOVAERTS (1997) e DEUTSCH (2002).

A idéia central da simulação *annealing* vem de uma analogia com a termodinâmica, especificamente ao modo como os líquidos congelam e cristalizam ou referente ao esfriamento e recozimento de metais. Em altas temperaturas as moléculas podem se mover livremente. Conforme a temperatura diminui lentamente, as moléculas se alinham em cristais que representam o estado de energia mínima do sistema. A relação entre temperatura (T) e energia (E) é mostrada na distribuição de probabilidade de Boltzman, $P\{E\} \sim \exp(-E / k_b T)$, onde se observa que mesmo em uma baixa temperatura existe uma probabilidade de que a energia seja bastante alta, ou seja, algumas vezes o sistema irá abandonar seu estado de baixa energia em favor de um estado de energia mais alta. Esse mesmo princípio é aplicado na simulação *annealing*:

um sistema mudará de uma configuração de energia O_1 para uma configuração de energia O_2 , com uma probabilidade $p = \exp(- (O_2 - O_1) / k_b T)$. O sistema mudará sempre que O_2 for menor do que O_1 , isto é, uma medida favorável sempre será tomada e algumas vezes ocorrerá uma medida desfavorável.

A Figura 3.12 ilustra o fluxograma geral para a simulação *annealing* aplicada a problemas geoestatísticos (DEUTSCH, 2002). A idéia básica da simulação *annealing* é perturbar continuamente uma imagem inicial até que ela se ajuste a alguma característica pré-especificada em uma função objetivo. Cada perturbação é aceita ou não, dependendo se ela conduz a imagem em direção à função objetivo.

No programa *sasim* do GSLIB (DEUSTCH e JOURNAL,1996), a imagem inicial é criada pelo deslocamento dos dados condicionantes aos nós mais próximos da malha e os valores dos nós restantes são extraídos aleatoriamente do histograma especificado pelo usuário. Uma função objetivo é definida como uma medida da diferença entre a característica espacial desejada e aquela gerada pela simulação, por exemplo, a diferença entre o semivariograma da simulação e o modelo semivariográfico. Qualquer combinação de cinco funções objetivo (histograma, semivariograma, semivariograma indicativo, correlação linear e distribuições condicionais) são permitidas no programa. Em geral, a função objetivo é montada sobre uma soma ponderada de N_c componentes O_i , cada uma medindo o desvio da simulação corrente (na i -ésima perturbação) das estatísticas desejadas:

$$O = \sum_{i=1}^{N_c} w_i O_i . \quad (3.93)$$

onde N_c é o número de componentes, w_i são pesos que controlam a importância relativa da i -ésima componente na função objetivo e O_i são as componentes da função objetivo. As componentes da função objetivo devem ser positivas e diminuir em direção a zero quando os dados ou a estatística espacial são reproduzidos pela simulação. Quando a função objetivo atingir um valor mínimo estabelecido, isto é, obtendo-se uma simulação aceitável próxima às estatísticas desejadas, o processo é concluído. Na prática, o processo também é concluído quando é excedido um tempo razoável de CPU (Unidade Central de Processamento). Caso contrário, a imagem é perturbada. Em geral, o número de perturbações exigido para alcançar a convergência é entre 10 e 1000 vezes o número de variáveis no sistema. Depois de cada perturbação cada componente da função objetivo é atualizada ou recalculada. O mecanismo de perturbação utilizado consiste da

troca de pares de valores z em duas localizações não amostradas u'_j e u'_k quaisquer, escolhidas aleatoriamente, ou seja, $z_{(0)}^{(l)}(\mathbf{u}'_j)$ torna-se $z_{(0)}^{(l)}(\mathbf{u}'_k)$ e vice-versa. Depois de cada troca, o valor da função objetivo (3.94) é atualizada. Foi utilizado o semivariograma para modelamento da continuidade espacial dos contaminantes, de modo que, a função objetivo incluiu um componente simples do tipo

$$O^{(i)} = \sum_{s=1}^S \frac{[\gamma(\mathbf{h}_s) - \hat{\gamma}_{(i)}(\mathbf{h}_s)]^2}{[\gamma(\mathbf{h}_s)]^2}. \quad (3.94)$$

onde $\gamma(\mathbf{h}_s)$ é o modelo de semivariograma desejado e $\hat{\gamma}_{(i)}(\mathbf{h}_s)$ é o valor do semivariograma no lag \mathbf{h} para a simulação na i -ésima perturbação, considerando-se um número especificado S de lags. A regra de decisão para aceitação de perturbações desfavoráveis foi estabelecida de acordo com uma distribuição de probabilidade exponencial negativa:

$$\Pr ob\{Aceitação\} = \begin{cases} 1 & se \ O_i \leq O_{i-1} \\ \exp\left(\frac{O_{i-1} - O_i}{t_i}\right) & se \ O_i > O_{i-1} \end{cases}. \quad (3.95)$$

Todas as perturbações favoráveis ($O_i \leq O_{i-1}$) são aceitas e as perturbações desfavoráveis são aceitas segundo uma distribuição exponencial de probabilidades. Quanto maior for o parâmetro t_i da distribuição de probabilidade, chamado *temperatura*, maior será a probabilidade de que uma perturbação desfavorável seja aceita na i -ésima interação. A temperatura é diminuída multiplicando-se a temperatura inicial t_0 pelo fator de redução λ quando bastantes perturbações tenham sido aceitas ou muitas perturbações tenham sido tentadas. O número máximo de perturbações aceita (K_{accept}) ou tentadas (K_{max}) é escolhido como um múltiplo de número N de nós da malha. DEUTSCH e JOURNAL (1996) sugerem que sejam utilizados em torno de 100 e 10 vezes os números de nós para K_{max} e K_{accept} , respectivamente. O critério de convergência para interromper o processo de otimização se dá quando a função objetiva alcançou um valor suficientemente pequeno (O_{min}) ou quando um número máximo de perturbações tentada na mesma temperatura alcançou um determinado números de vezes (S).

O esquema utilizado para reduzir a temperatura com o tempo é conhecido como *programação annealing*. A determinação de uma *programação annealing* adequada para diferentes problemas permanece um desafio importante na aplicação da simulação *annealing* em problemas geoestatísticos (DEUSTCH, 2002). Embora a simulação

annealing tenha sido bem adaptada a aplicações geoestatísticas, em algumas situações podem ocorrer efeitos indesejáveis conhecidos como efeitos de borda. Esses efeitos consistem de valores extremos preferencialmente localizados na borda da malha e são particularmente notáveis quando a temperatura é reduzida rapidamente ou quando a distribuição univariada é fortemente assimétrica. Na reprodução de semivariogramas de variáveis fortemente assimétricas, os valores extremos tendem a serem empurrados para as bordas porque eles contribuem somente uma vez para o cálculo do semivariograma. Pode-se tentar evitar tais efeitos ponderando-se as contribuições dos pares próximos às bordas da malha ou alternativamente reduzindo-se o número de *lags* (distâncias entre os pares de amostras) usados na função objetivo DEUSTCH e COCKERHAM (1994), CARLE (1997) e DEUSTCH (2002). As maiores dificuldades na implementação da simulação *annealing* estão em compor a função objetivo e escolher os muitos parâmetros inter-relacionados na programação *annealing*.

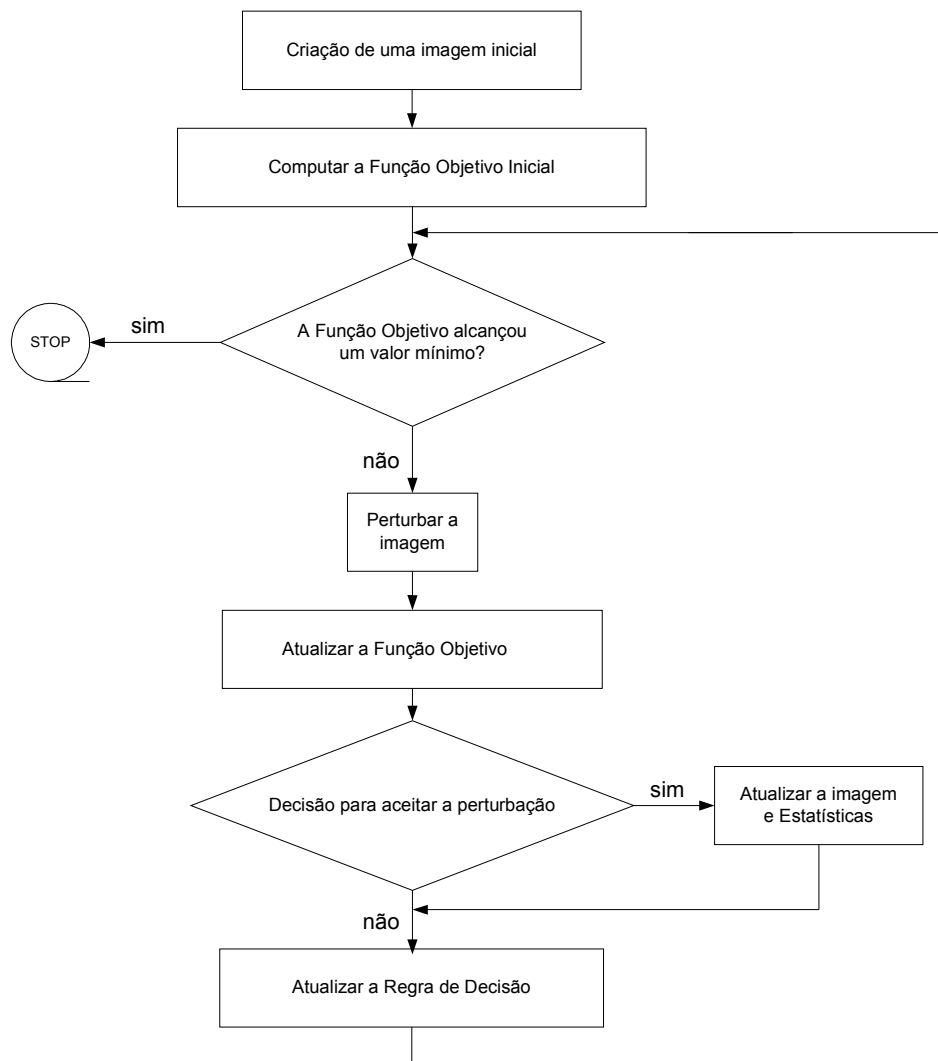


Figura 3.12 – Fluxograma com os passos utilizados na simulação *annealing* (DEUTSCH, 2002).

4. ESTUDO DE CASO

4.1. O problema estudado

Em 1946, foram descobertas ocorrências de minério de manganês em Serra do Navio, no então Território Federal do Amapá. Após constatar a importância das jazidas, o Governo Federal declarou-as reserva nacional, pelo Decreto-lei nº 9.858, de 13 de setembro de 1946, incumbindo a Administração Territorial de proceder ao imediato estudo do seu aproveitamento e autorizando-a a contratá-lo com entidades particulares ou de economia mista. Com base no mesmo Decreto, foi procedida pelo Conselho Nacional de Minas e Metalurgia concorrência para o referido aproveitamento, onde foi declarada vencedora, dentre três concorrentes, a Indústria e Comércio de Minérios S.A – ICOMI. As cláusulas contratuais que regeram os direitos e obrigações relacionadas com o aproveitamento das jazidas de manganês foram aprovadas pelo Decreto Federal nº 24.156, de 4 de dezembro de 1947 e revistas pelo de nº 28.162, de 31 de maio de 1950. Em 25 fevereiro de 1953, o Presidente da República, em despacho exarado em exposição de motivos do Ministro da Fazenda, a propósito da obtenção de financiamento para a execução do projeto de aproveitamento das mencionadas jazidas, declarou-o economicamente vantajoso do ponto de vista nacional, cuja execução seria iniciada em janeiro de 1954, inaugurando-se em janeiro de 1957, praticamente 10 anos após os primeiros estudos e trabalhos realizados pela ICOMI no Amapá. Dentre as diversas cláusulas que regem o contrato celebrado entre a ICOMI e o Governo do Território do Amapá, destacam-se o prazo de 50 anos para a vigência da concessão, com término no ano de 2003, a construção de uma estrada de ferro de aproximadamente 200 km de extensão, o pagamento de “royalties” diretamente ao Governo do Território e a reversão gratuita ao patrimônio da União de todas as instalações e benfeitorias existentes ao término do contrato.

As jazidas de manganês da Serra do Navio situavam-se ao longo de uma faixa de cerca de 8 km de comprimento, com direção noroeste que atravessa o rio Amapari, à latitude de 52°02’W, a cerca de 200 km da cidade de Macapá e de 500 Km da cidade de Belém do Pará. Para possibilitar a produção e escoamento desse minério, a ICOMI implantou no Território Federal do Amapá um parque industrial constituído basicamente de:

– instalações industriais para a mineração, beneficiamento e concentração do minério;

- uma estrada de ferro, ligando a mina ao porto;
- uma usina de pelotização no Porto de Santana;
- instalações portuárias especializadas em Santana;
- instalações e serviços de apoio e de infra-estrutura;
- duas vilas residenciais (Santana e Serra do Navio) para seus empregados e familiares.

Visando ao aproveitamento do minério de baixo teor de manganês, resultantes das operações de beneficiamento do minério na Serra do Navio, a ICOMI decidiu ampliar suas atividades industriais construindo uma usina de pelotização de minério de manganês. A usina teve caráter pioneiro, sendo, na verdade, a primeira unidade do seu gênero a entrar em operação industrial no mundo, cujas obras foram concluídas no final de 1971 e operou de 1972 até 1980. Essa usina de pelotização sofreu uma adaptação, passando a funcionar como uma usina de sinterização que operou de 1989 até 1997. Foram gerados nessa operação resíduos de manganês que eram estocados em uma Bacia de Rejeitos e em suas imediações, dentro da área industrial da ICOMI, em Santana. Nessa área também foi estocado o minério comercializável e não comercializável, além dos rejeitos de manganês (Fotos 4.1 a 4.3).

O minério de manganês tem sua aplicação principal na indústria siderúrgica, utilizado predominantemente na fabricação de numerosos tipos de aço, funcionando como agente desoxidante e removedor de impurezas, efeitos importantes para a melhoria da qualidade dos produtos. Cabe-lhe também assegurar as propriedades desejadas a certos tipos de aço e de ligas não ferrosas. Fora da indústria siderúrgica, o manganês é empregado, em pequena escala, na fabricação de baterias de células secas, como agente corante de cerâmica, na composição de vários produtos químicos e adubos.

Em novembro de 1997, após extrair cerca de 60 milhões de toneladas de minério, a ICOMI apresentou e enviou um relatório ao Departamento Nacional de Pesquisa Mineral – DNPM – onde demonstrava a exaustão da jazida e afirmava não existir mais minério de qualidade que justificasse o funcionamento da empresa. O DNPM local não aprovou o relatório e fez 7 exigências à empresa, entre as quais a recuperação do meio ambiente e das instalações industriais e residenciais que deveriam ser devolvidas ao poder público após o fim do contrato. Houve uma série de denúncias de problemas ambientais que poderiam comprometer a qualidade de vida dos moradores de Santana e Serra do Navio, municípios onde a empresa atuou. Entre os principais problemas ambientais denunciados destacam-se as contaminações do solo e das águas

superficiais e subterrâneas por arsênio, principalmente devido à estocagem de mais ou menos 150.000 toneladas de rejeitos de manganês contaminados na área industrial da ICOMI no município de Santana (Foto 4.3); e a utilização destes mesmos rejeitos de manganês para o aterro e pavimentação de ruas no mesmo município (Foto 4.4). Há a hipótese de que este problema surgiu nos anos 70 quando a usina de pelotização começou a entrar em funcionamento: o minério submetido a altas temperaturas teria provocado a evaporação do arsênio contido naturalmente no manganês. O arsênio retornaria então ao solo provocando possivelmente a contaminação do lençol freático. Se a contaminação realmente existir, o risco de afetar a saúde de parte da população pode ser significativo devido à utilização por grande parte dos habitantes locais de água de poços amazonas (cacimbas) e de riachos próximos à área afetada mostrados nas Fotos 4.5 e 4.6, respectivamente. O assunto, entretanto, tornou-se bastante polêmico pela sua enorme complexidade.



Foto 4.1 – Vista parcial da Bacia de Rejeitos onde eram depositados os resíduos de manganês provenientes das usinas de pelotização/sinterização. A Bacia foi limpa e aterrada pela ICOMI.



Foto 4.2 – Uma das pilhas de manganês comercializável (coberto com lonas) na área industrial da ICOMI. Ao fundo observa-se uma pilha de minério não comercializável.



Foto 4.3 – Pilhas de rejeitos de manganês cobertos com lonas dentro da área da ICOMI.



Foto 4.4 – Uma rua do bairro da Hospitalidade no bairro comercial de Santana e próximo ao Igarapé Boiero, aterrada com rejeitos de manganês. Uma fina camada de piçarra foi colocada por cima dos rejeitos.

Durante o ano de 1997 e primeiro quadrimestre de 1998, foram realizados serviços de auditoria ambiental na área portuária de Santana pela Jaakko Pöyry Engenharia – JPE. O relatório dessa auditoria confirmou a contaminação das águas subterrâneas por arsênio, salientando que essa contaminação se encontrava restrita à área da ICOMI. Entretanto, nesse mesmo relatório, foi constatada, na 3^a. Campanha de Amostragem, uma contaminação por arsênio e manganês das águas superficiais do Igarapé Elesbão 2 (Foto 4.6), que se encontra fora dos limites da propriedade da ICOMI (RELATÓRIO JPE, 1998).

Em amostras de peixes analisadas não foi detectada a bio-acumulação de arsênio, ocorrendo, no entanto, indícios da presença de Fe e Mn em algumas espécies (RELATÓRIO JPE, 1998). Durante os anos de 2000/2001, um trabalho realizado pelo Departamento de Química da Universidade Federal do Pará a pedido do Governo do Estado do Amapá, para avaliação da contaminação por arsênio na região delta do rio Amazonas e caracterização de arsênio em amostras de solo da cidade de Santana, Amapá, mostrou que a maioria (78%) das amostras de solos retiradas de 7 ruas aterradas com o rejeito de manganês apresentaram níveis de arsênio acima de 100 ppm ($\mu\text{g/g}$), indicando alta contaminação, sendo 57,89% com níveis de arsênio acima de 500 ppm.

As análises de água constataram apenas duas amostras no Canal Norte (Rio Amazonas) com níveis acima de 10 ppb (limite estabelecido pela OMS para águas de rio), sendo que a média nesse canal ficou em 8 ppb (RELATÓRIO TÉCNICO PARCIAL, UFPA, 2001). Outro estudo realizado para a avaliação da contaminação por arsênio no cabelo de moradores no município de Santana/AP mostrou que as análises do cabelo de 98 em 100 pessoas pesquisadas apresentaram índices de arsênio acima 1 µg/g, limite estabelecido pela OMS para amostras de cabelo (PARECER TÉCNICO, UFPA, 2000). Outro estudo realizado pelo Instituto Evandro Chagas de Belém/PA, no período de 6 a 23 de setembro de 2001, envolvendo 2.045 pessoas residentes na Vila do Elesbão, localizada próximo à área industrial da ICOMI (Foto 4.7) e redondezas, apresentou como resultado de As em cabelo de 1.986 pessoas a média de 0,56 µg/g, variando entre 0,06 e 5,85 µg/g. Em relação à presença de As no sangue de 1.927 pessoas, o resultado encontrado apresentou média de 5,95 µg/g, variando de 0,07 a 19,3 µg/g. O estudo concluiu que as patologias identificadas na comunidade do Elesbão em Santana/AP, através da análise clínico–epidemiológica–laboratorial não apresentam relação estatística significativa com as médias de As encontradas no sangue e cabelo da população estudada. As patologias estudadas foram parasitismo, geofasia, malária, hepatite pregressa e neoplasia (referente a câncer). Esse mesmo estudo mostrou que as análises físico-químicas realizadas no meio físico (águas superficiais e subterrâneas e sedimentos de fundo) apresentaram teores elevados de As e Mn, com médias de 131,4 e 0,80 µg/g, respectivamente, principalmente na área industrial da ICOMI e restritos à antiga bacia de rejeitos. Os teores de As nas águas dos Igarapés Elesbão 1 e 2, fora da área industrial da ICOMI, também foram elevados com médias de 124,7 e 75 µg/g, respectivamente (INSTITUTO EVANDRO CHAGAS, Belém, 2001). QUEIROZ *et al.* (2001) realizaram um estudo onde utilizaram simulação “annealing” para a localização, delimitação e extensão das fontes de contaminação restritas à área industrial da ICOMI. Os resultados mostraram que a extensão da área contaminada, com probabilidade acima de 80%, foi em torno de 150 m², 700 m² e 100 m² para o arsênio, manganês e ferro, respectivamente. Observa-se, portanto, que não existe consenso sobre as fontes de contaminação e sobre a relação entre os níveis de As e casos epidemiológicos na área, indicando, deste modo, a necessidade de mais estudos para elucidar essa polêmica questão.



Foto 4.5 – Poço amazonas comunitário utilizado para consumo de água por moradores da Vila do Elesbão, próximo à área industrial da ICOMI.



Foto 4.6 – Igarapé do Elesbão, localizado próximo à área industrial da ICOMI.

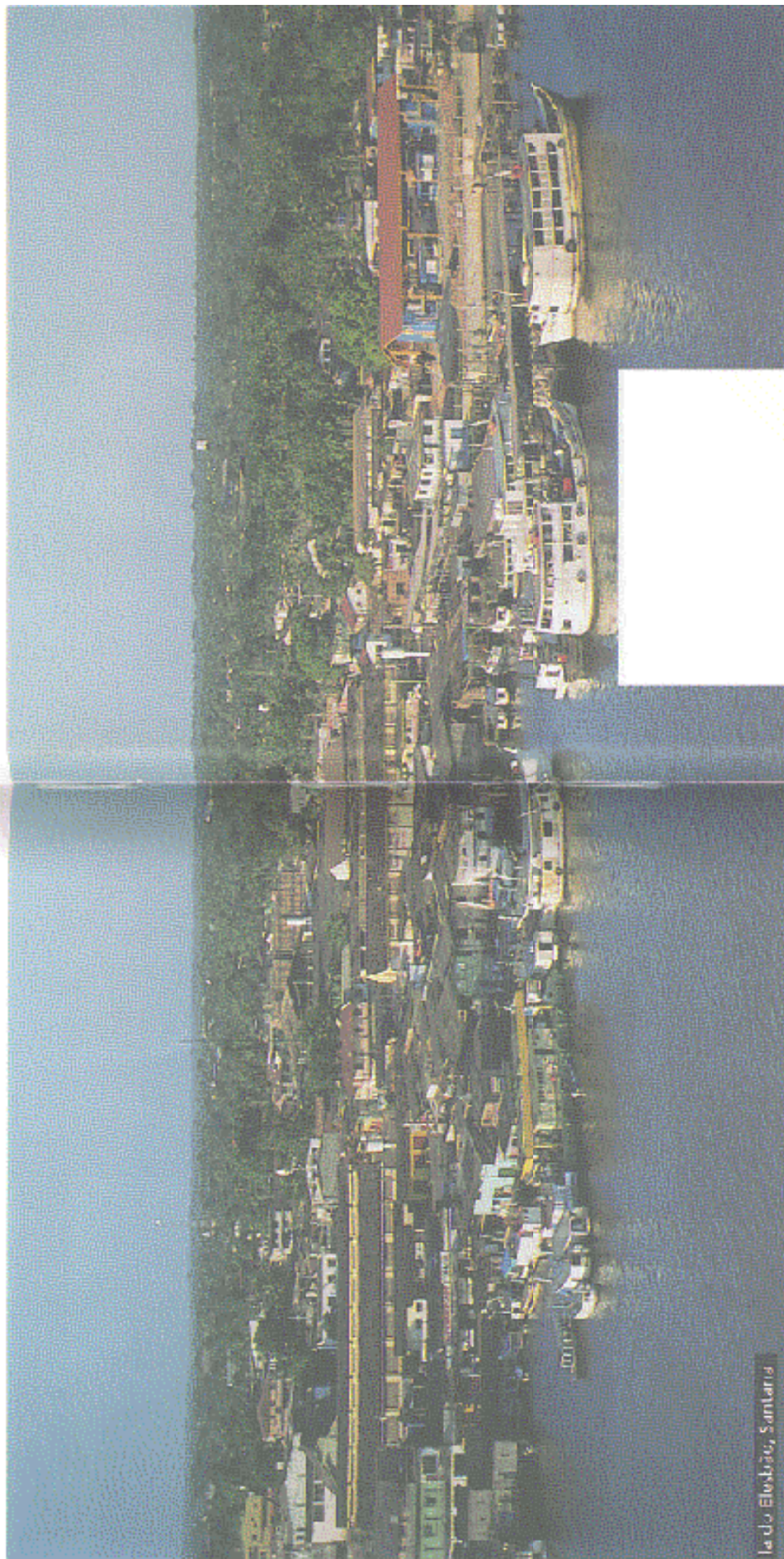


Foto 4.7 – Vista aérea da Vila do Elesbão em Santana, Amapá, localizada próximo à área industrial da ICOMI (Fonte: Amapá, um norte para o Brasil, 2000).

4.1. Aspectos gerais da área de estudo

4.1.1. Localização

O município de Santana está localizado no estado do Amapá, no extremo norte do Brasil, aproximadamente entre 50⁰ e 55⁰ W e 0⁰ a 5⁰ N. A cidade de Santana possui atualmente 80.439 habitantes, a segunda maior do referido estado (IBGE, Censo de 2000). Para a realização da exploração do manganês, a ICOMI construiu, além das instalações industriais, duas vilas residenciais com toda infraestrutura de saneamento e lazer, escolas, supermercado, hospital e hospedagens, para acolher as famílias e trabalhadores da empresa. Uma vila residencial foi construída próximo às minas de manganês no município de Serra do Navio e outra no município de Santana próxima à área industrial onde eram estocados os minérios de manganês e cromita, produtos e rejeitos. A ligação entre essas duas vilas residenciais ocorria através da estrada de ferro que transportava o minério de Serra do Navio até o porto de Santana, numa distância em torno de 200 km (Figura 4.1). Pôde-se observar que, ao longo da estrada de ferro e em torno das vilas residenciais da ICOMI, foram se formando pequenas comunidades de trabalhadores sem vínculo direto com a empresa, chegando algumas a alcançar um nível de desenvolvimento que deu origem a novos municípios, como Porto Grande e Pedra Branca. Esse processo, entretanto, se deu tal como em outras regiões do Brasil, onde as plantas industriais atraíram trabalhadores, mas simultaneamente contribuíram para a degradação ambiental e a expansão de assentamentos periféricos de baixa renda com condições urbanas mínimas em geral (ROLNIK, 1998).

4.1.2. Geologia

A área de estudo está assentada sobre sedimentos da Formação Barreiras, constituídos por argilas orgânicas siltosas, siltes argilosos e argilas duras com intercalações escassas de areias finas a grossas. A designação de “Barreiras”, consagrada pelo uso na nomenclatura brasileira, encontra uma variedade de sedimentos que ocorrem ao longo do litoral brasileiro, desde o Amapá até o Rio de Janeiro, e

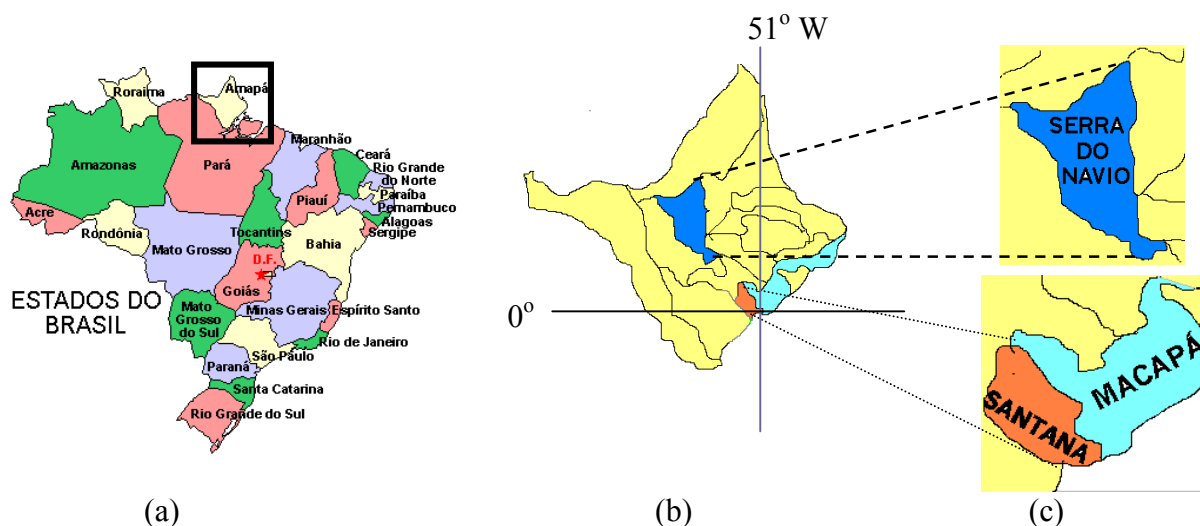


Figura 4.1 – (a) Estado do Amapá no extremo norte do Brasil ; (b) e (c) municípios do estado do Amapá com destaque para Serra do Navio e Santana, onde ocorreu a exploração (Serra do Navio), pelotização/sinterização e embarque do minério de manganês (Santana).

forma a maioria das “terras firmes” localizadas às margens do rio Amazonas e seus afluentes (DNPM/Projeto Radam, 1974). A Figura 4.2 apresenta uma seção geológica esquemática, onde mostra-se o arcabouço geológico local, através de um corte transversal, elaborado a partir das sondagens de reconhecimento. Observa-se nesse setor leste da área da ICOMI, onde estão concentradas diversas unidades e pátios de estocagem, que o perfil geológico às margens do rio Amazonas apresenta a seguinte seqüência do topo para a base: aluviões (argilas orgânicas siltosas) com extensão aproximada de 150m e espessuras de até 40m; horizonte de siltes argilosos (horizonte superior), com espessura contínua em torno de 6 a 8m e um horizonte de argilas duras (horizonte inferior). O nível de água (N.A) que separa o horizonte não saturado (acima do N.A), do saturado (abaixo do N.A), apresenta profundidades variando desde centímetros nas proximidades das margens do rio, até um valor máximo na ordem de 9,0 m, na porção norte. Essas profundidades oscilam ao longo do ano, devido às variações sazonais da superfície potenciométrica das águas subterrâneas (RELATÓRIO JPE, 1998).

Um aspecto interessante sobre os solos do Amapá, apresentado por GUERRA (1954), é a presença de ferro e alumínio acumulados após a lixiviação da cal. Esses tipos de solo são formados em regiões caracterizadas por precipitações pluviais abundantes e com períodos de secas bem marcadas como as observadas em toda a

região. O processo de laterização originando a formação de hidróxidos de ferro e alumínio constitui, segundo o Prof. Pierre Gorou, uma verdadeira *lepra pedológica* e é inteiramente desconhecido nas regiões temperadas (GUERRA, 1954).

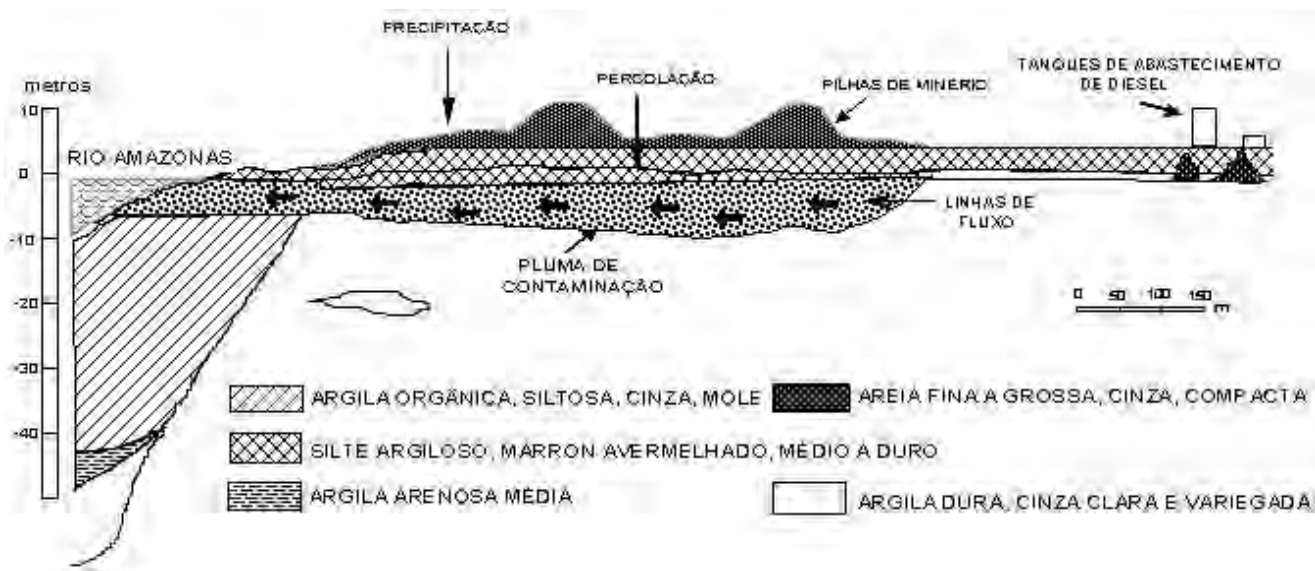


Figura 4.2 – Seção geológica esquemática da Área industrial de Santana (RELATÓRIO JPE, 1998).

4.1.3. Vegetação

O município de Santana ocupa uma área de 159.970 ha, onde a cobertura vegetal predominante é o cerrado com cerca de 48,95% (CHAGAS, 1998). Estudos do projeto RADAM apresentam o cerrado como uma formação predominantemente dos climas quentes úmidos, com chuvas torrenciais bem demarcadas pelo período seco e caracterizadas, sobretudo, por árvores tortuosas, de folhas raramente decíduas, bem como por formas biológicas adaptadas aos solos profundos, deficientes e aluminizados (DNPM, 1974). Em sua totalidade, o Cerrado do Amapá é marcado por características físico-naturais que guardam relativa uniformidade ao longo de sua distribuição espacial, ressalvadas as particularidades que se refletem em tipologias individualizadas, reflexo de um processo evolutivo ainda pouco estudado. Documento recente do BDT aponta para a necessidade de se considerar diferentes tipologias dominantes como conjuntos fisiográficos que podem servir de indicadores primários para definição e áreas prioritárias para uso sustentável e conservação da biodiversidade (CHAGAS, 1998).

4.1.4. Clima

Localizando-se na faixa equatorial, a área de estudo apresenta um clima do tipo quente, com pouca variação térmica, e superúmido. A Tabela 3.1 (pág.12) mostra que no período de 1968 a 1990, a variação de temperaturas em Macapá, a 25 km de Santana, foi de 25.7°C a 27.7°C, destacando-se os meses de setembro a dezembro com temperaturas acima de 27°C. Em relação a pluviosidade, observa-se, na mesma TABELA, grande variação, apresentando valores de 35,5 a 407,2 mm, alcançando um volume pluviométrico médio anual acima de 2500mm. As áreas mais pluviosas são justamente aquelas onde com mais frequência se dá a ocorrência de chuvas, correspondendo aos meses de janeiro a junho, com destaque para o mês de março com maior índice pluviométrico médio (IBGE, 1977).

4.1.5. Hidrografia

A rede hidrográfica constitui um elemento importante na paisagem física e na economia da região. O Porto de Santana fica em frente ao Canal Norte do Rio Amazonas, onde quase a totalidade de suas águas é lançada no Oceano Atlântico. Apesar do grande volume d'água, ocasionando uma descarga abundante em função do clima pluvioso na maior parte do ano, as cheias na região são moderadas, não se verificando grandes inundações, como em outras regiões do país. Observa-se na Fig. 2.3 que a área em estudo é cortada por diversos igarapés, que constituem geralmente um curso d'água estreito por vezes muito sinuosos, navegados apenas por pequenas embarcações (Foto 4.6). A região drenada pelos igarapés Elesbão 1 e Elesbão 2, que passam pela área industrial da ICOMI, é recoberta totalmente por sedimentos aluvionares de idade quaternária, argiloso e ricos em matéria orgânica, recoberta por camada de húmus com espessura entre 0,30 a 0,50 metros (PCA – ICOMI, 2001, p. 3.89)

4.1.6. Uso e ocupação do solo

A Figura 4.3 mostra a região em torno da área industrial da ICOMI no município de Santana caracterizado pelo solo predominantemente residencial, destacando-se à Sudeste (SE) a Vila Amazonas, construída especialmente para os trabalhadores e seus familiares que trabalhavam na ICOMI. Observa-se que a rede de drenagem (em azul) do

bairro do Elesbão, localizado a oeste, avança para dentro da área industrial da ICOMI. O Igarapé Elesbão 2, um dos que penetra na área da ICOMI, apresentou contaminação por arsênio e manganês (RELATÓRIO, JPE, 1998). Pequenos riachos também cortam a área residencial, principalmente o bairro da Hospitalidade, onde ocorreu o aterro e pavimentação de ruas com o rejeito de manganês proveniente das usinas de pelotização/sinterização da ICOMI.

Foram identificados no bairro do Elesbão, indústrias de médio e pequeno porte voltadas, sobretudo, para a comercialização de madeiras. A sudeste funciona uma indústria de grande porte, a Amazônia Celulose – AMCEL. A área comercial, onde fica o Porto de Santana com grande movimentação de embarcações de pequeno, médio e grande porte, localiza-se ao lado da área industrial da ICOMI.

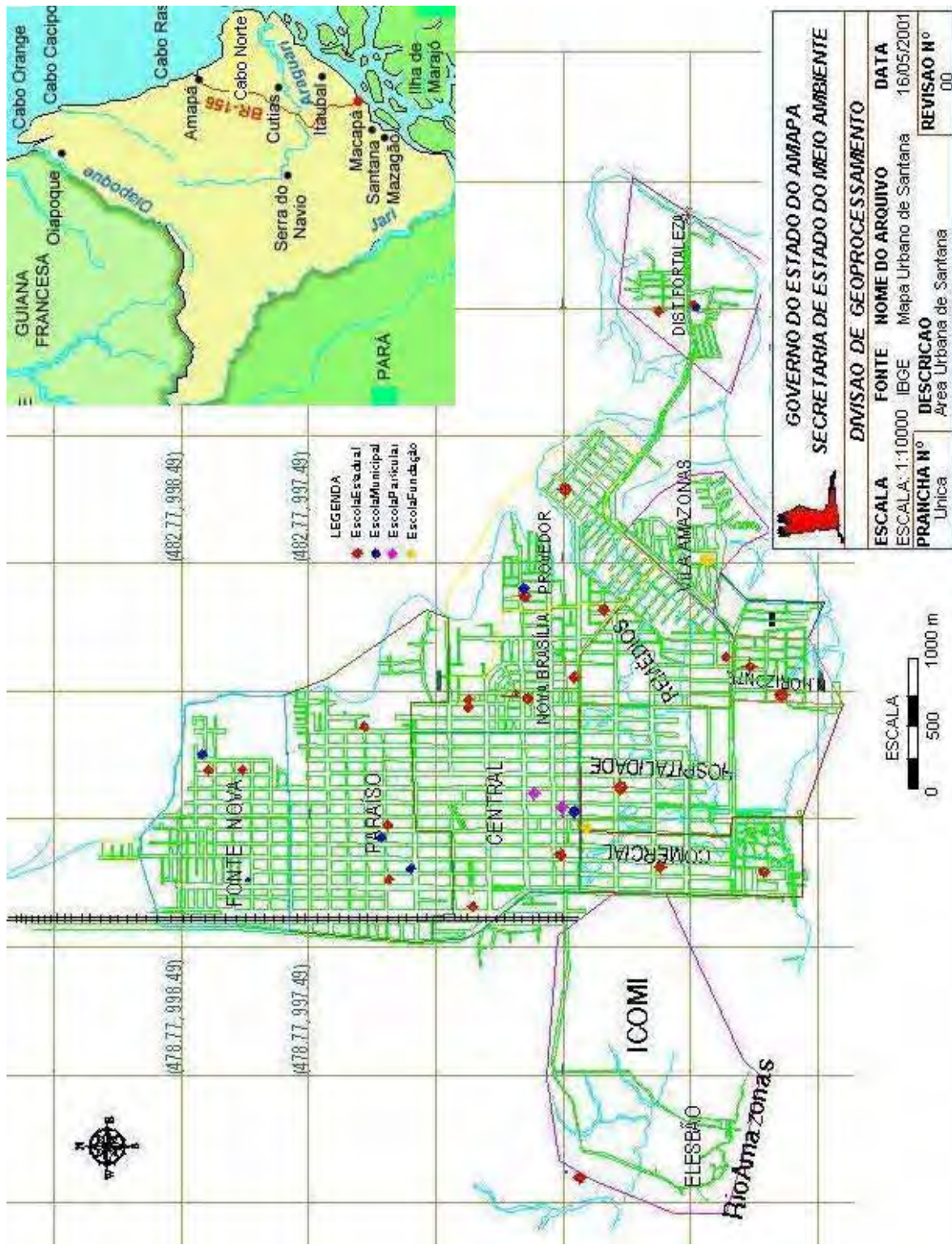


Figura 4.3 – Vista parcial do mapa urbano de Santana, mostrando o uso do solo em torno da área industrial da ICOMI

5. APLICAÇÃO DAS TÉCNICAS GEOESTATÍSTICAS E DISCUSSÃO DOS RESULTADOS

5.1 Análise exploratória dos dados

Nesta análise exploratória serão apresentadas as características mais relevantes do conjunto de dados da área de Santana. Os dados são constituídos por medidas de concentração (em ppm) na água de 16 elementos químicos. Esta parte puramente descritiva é um passo preliminar para a construção de modelos numéricos e probabilísticos para a incerteza na estimação espacial.

As distribuições univariadas (uma variável de cada vez) são descritas a seguir:

5.1.1 Distribuição de freqüências

O conjunto de medidas das variáveis será representado por $\{z(\alpha), \alpha = 1, \dots, n\}$, onde $z(\alpha)$ indica a concentração de cada elemento em n localizações α . Geralmente, a distribuição de variáveis contínuas é descrita por um histograma onde os valores das variáveis são discretizados em determinado número de classes de mesmo tamanho e a proporção relativa dos dados dentro de cada classe é indicada pela altura das colunas. Essas proporções relativas definem as freqüências das classes, portanto, o histograma descreve a distribuição de freqüências dos valores z para um determinado número de classes.

As Figuras 5.1 a 5.4 mostram os histogramas das concentrações dos poluentes expressos em mg/L (equivalente partes por milhão (ppm)). As caudas mais longas em praticamente todos os histogramas dos poluentes indicam a presença de pequeno número de grandes concentrações. Muitas vezes é importante conhecer a proporção de valores que excedem um limite máximo permitido para que se tenha, por exemplo, uma idéia preliminar da ação de remediação a ser realizada. A proporção de dados $z(\alpha)$ que não excede um determinado valor de corte (ou limite) z_k , denotada por $F(z_k)$, é dada por:

$$F(z_k) = \frac{1}{n} \sum_{\alpha=1}^n i(\alpha; z_k) \quad (5.1)$$

onde a transformação indicativa de um valor $z(\cdot)$ é definida como:

$$i(\alpha; z_k) = \begin{cases} 1 & \text{se } z(\alpha) \leq z_k \\ 0 & \text{caso contrário} \end{cases} \quad (5.2)$$

Essas proporções podem ser computadas como uma série de valores de corte z_k , resultando na função de distribuição acumulada $F(z_k)$, representada pela linha sólida nas Figuras 5.1 a 5.4 onde a linha vertical tracejada indica, para cada poluente, o valor máximo permitido para água doce, classes 1 e 2, de acordo com a resolução n^o. 20 do Conselho Nacional de Meio Ambiente (CONAMA), de 18 de junho de 1986. A Tabela 5.1 apresenta os valores máximos permitidos pela Organização Mundial de Saúde (OMS) e CONAMA. Pode-se verificar que não existe concordância nos limites máximos estabelecidos para vários poluentes. Para o molibdênio foi utilizado o limite fixado pela OMS. O alumínio apresentou maior percentual de poluição com 75,5%, seguido pelo selênio com 46,3%, ferro com 36,6%, manganês com 27%, arsênio com 9,8%, cobre e chumbo com 4,9% e cádmio com 7,3%. Os outros poluentes apresentaram valores abaixo dos limites máximos fixados.

As principais características de uma distribuição são descritas por seus valores de tendência central, medidas de dispersão, assimetria e curtose. A média aritmética e a mediana são as medidas mais usadas de tendência central. A variância, desvio-padrão, coeficiente de variação e quartis descrevem a dispersão da distribuição. O coeficiente de assimetria e a curtose indicam o grau de simetria ou a forma da distribuição em relação ao modelo de distribuição normal. Essas estatísticas foram descritas no capítulo 3.

A Tabela 5.2 apresenta, para cada elemento, os resultados obtidos para principais estatísticas. Pode-se observar que a maioria dos elementos possui acentuada assimetria positiva e valores de curtose elevados, ou seja, praticamente todas as variáveis apresentam distribuições não normais. Os altos valores das medidas de dispersão representadas pela variância, desvio-padrão e coeficiente de variação, indicam um alto grau de incerteza, sobretudo naqueles metais que apresentaram concentrações acima dos valores máximos permitidos (em negrito na referida Tabela) apontando para a necessidade de realização de amostras adicionais para que se possa obter melhores resultados.

TABELA 5.1 – Valores máximos permitidos (em mg/L) para águas doces.

Elemento			OMS ¹	CONAMA	Valores Medidos (mg/L)	
			mg/L	mg/L	Min	Max
1	Ferro	Fe		0,30	0,000	85,41
2	Manganês	Mn	0,50 (P)	0,10	0,002	51,44
3	Arsênio	As	0,01 (P) ²	0,05	0,000	22,92
4	Alumínio	Al	0,20	0,10	0,008	21,78
5	Selênio	Se	0,01	0,01	0,000	0,041
6	Chumbo	Pb	0,01	0,03	0,008	0,04
7	Cobre	Cu	2,00 (P) ³	0,02	0,000	0,03
8	Cádmio	Cd	0,003	0,001	0,000	0,003
9	Bário	Ba	0,70	1,00	0,004	0,989
10	Estrôncio	Sr			0,000	0,94
11	Molibdênio	Mo	0,07		0,000	0,88
12	Zinco	Zn		0,18	0,000	0,07
13	Cobalto	Co		0,20	0,000	0,03
14	Níquel	Ni	0,02 (P)	0,025	0,018	0,02
15	Cromo	Cr	0,05 (P)	0,05	0,000	0,002
16	Prata	Ag		0,01	0,000	0,001

(P): Valor regulador Provisório.

¹ Organização Mundial da Saúde, Extraído de “Guidelines for drinking-water quality”

² Para valores excedentes, o risco de câncer de pele é de 6×10^{-4} ,

³ Baseado em efeitos gastrointestinais agudos.

Os valores em negrito identificam concentrações acima do máximo permitido.

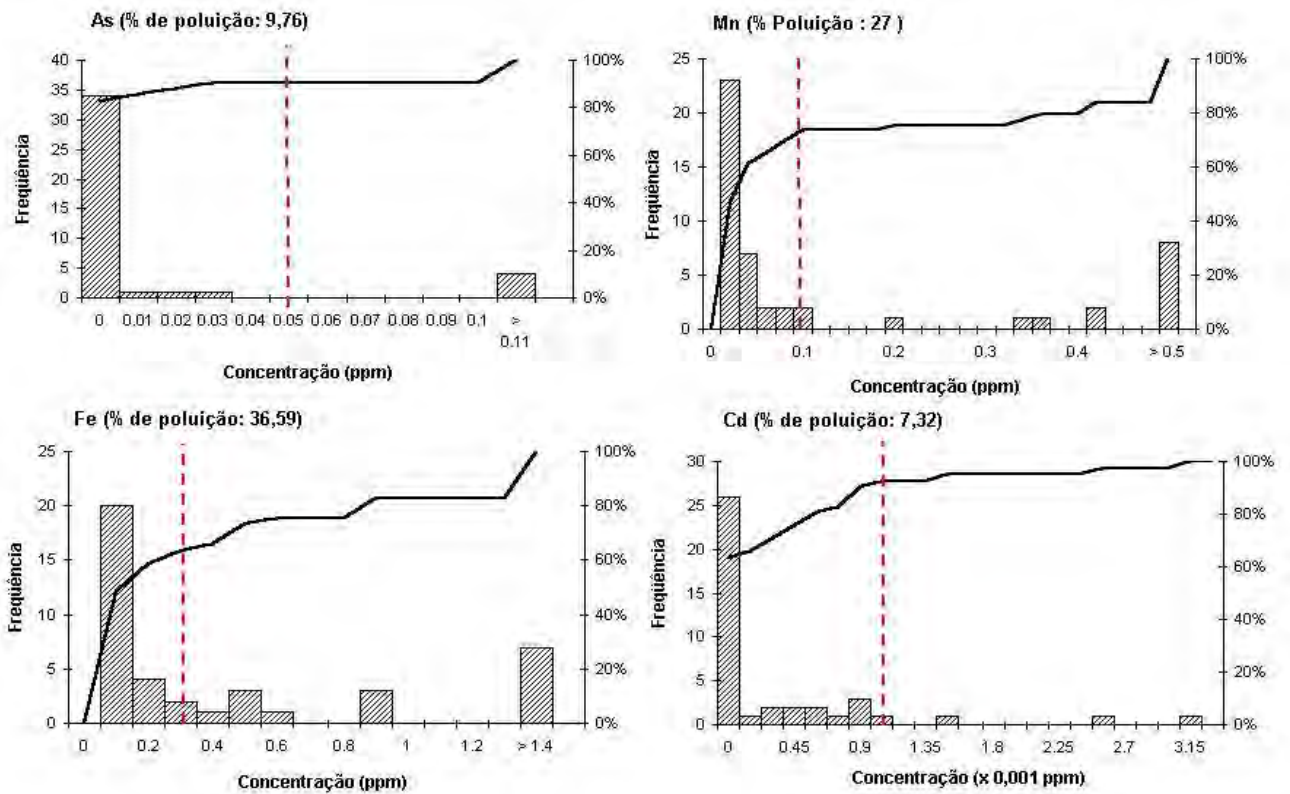


Figura 5.1 – Histogramas e distribuições acumuladas do arsênio (As), manganês (Mn), ferro (Fe) e cádmio (Cd). As linhas tracejadas indicam a tolerância máxima permitida (CONAMA).

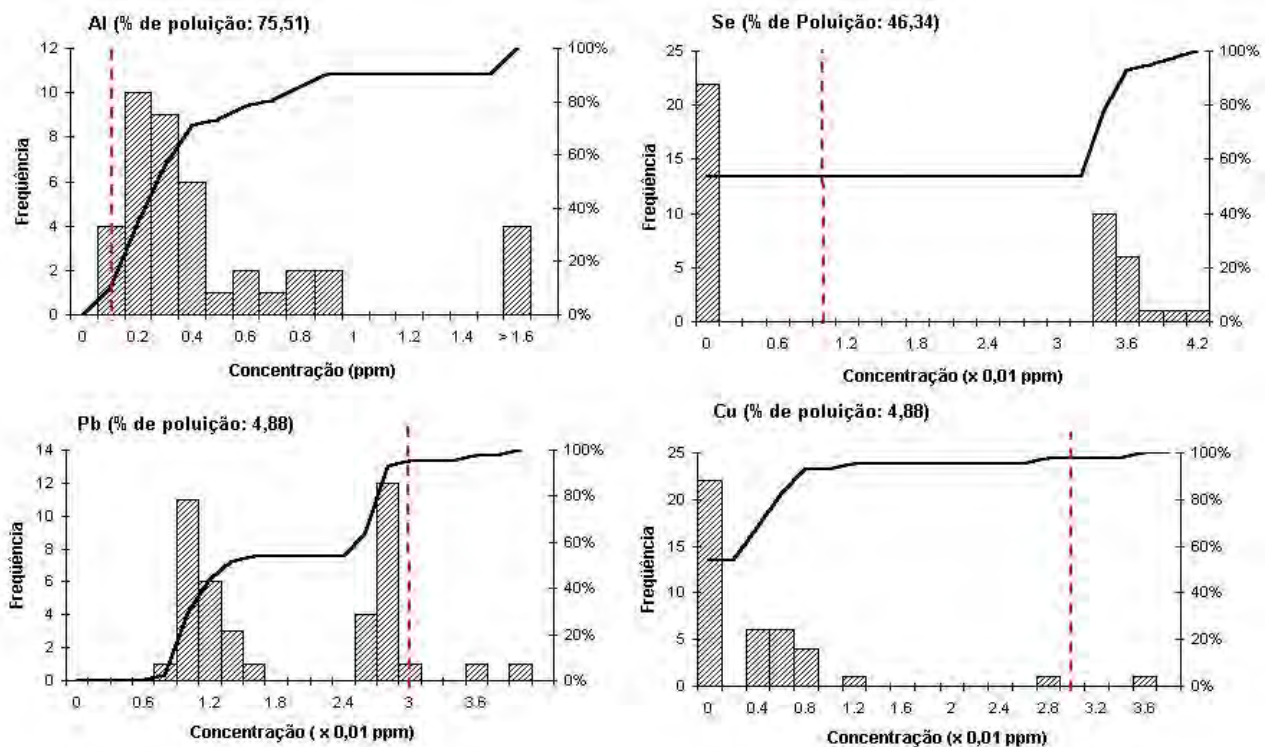


Figura 5.2 – Histogramas e distribuições acumuladas do alumínio (Al), selênio (Se), chumbo (Pb) e cobre (Cu). As linhas tracejadas indicam a tolerância máxima permitida (CONAMA).

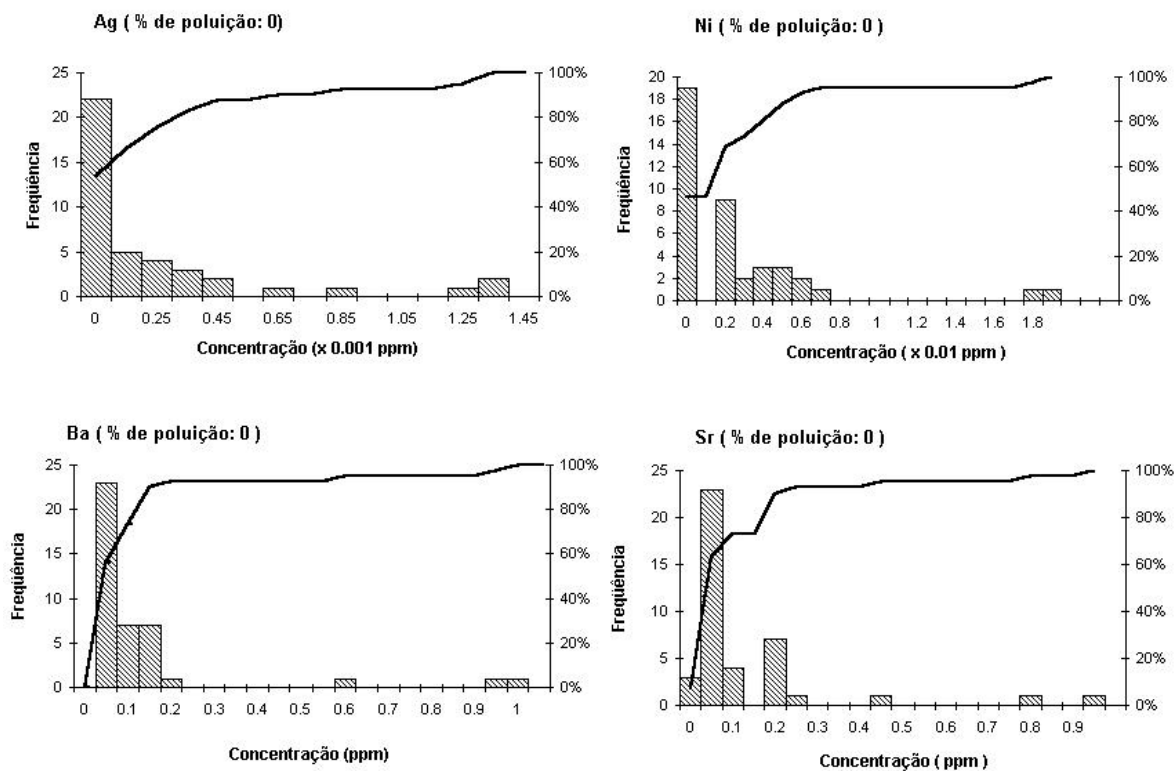


Figura 5.3 – Histogramas e distribuições acumuladas da prata (Ag), níquel (Ni), bário (Ba) e estrôncio (Sr).

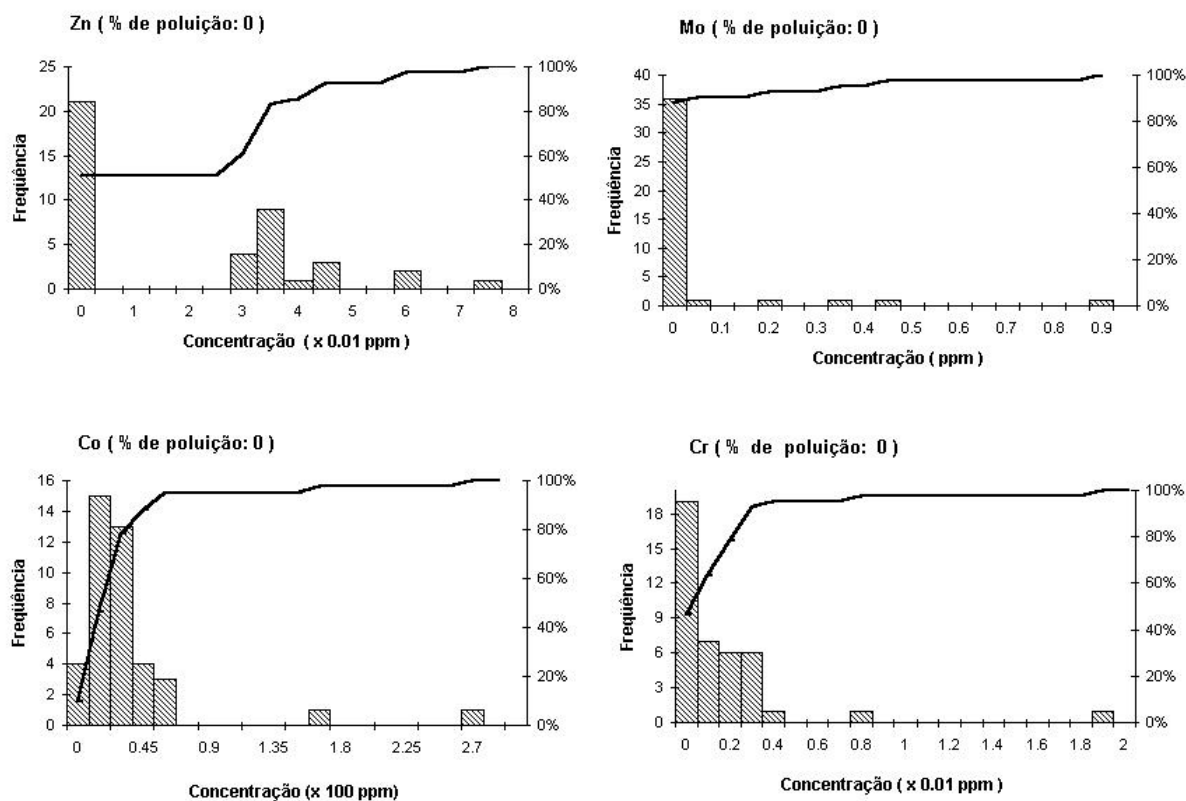


Figura 5.4 – Histogramas e distribuições acumuladas do zinco (Zn), molibdênio (Mo), cobalto (Co) e cromo (Cr)

TABELA 5.2 – Estatísticas descritivas para o conjunto de dados (mg/L)

	Ag ¹	Al	As	Ba	Cd ¹	Co ²
Média	0,20	1,45	1,28	0,11	0,32	0,29
Variância	0,12	17,5	22,55	0,05	0,43	0,21
Desvio-padrão	0,35	4,19	4,75	0,22	0,66	0,46
Coef. De Variação(%)	175,00	288,97	371,09	200,00	206,25	158,62
Mediana	0,00	0,21	0,00	0,04	0,00	0,17
Mínimo	0,00	0,01	0,00	0,00	0,00	0,00
Máximo	1,30	21,77	22,92	0,99	3,06	2,61
1 Quartil	0,00	0,10	0,00	0,02	0,00	0,12
3 Quartil	0,22	0,37	0,00	0,10	0,43	0,25
Assimetria	2,27	3,27	3,85	3,38	2,90	4,11
Curtose	4,54	10,42	14,35	11,09	9,15	18,59
N	41	49	41	41	41	41

¹: (x 0,001 mg/L)

²: (x 0,01 mg/L)

TABELA 5.2 (Cont.) – Estatísticas descritivas para o conjunto de dados (mg/L)

	Cr ²	Cu ²	Fe	Mn	Ni ²	Pb ²
Média	0,14	0,36	6,22	2,32	0,23	1,83
Variância	0,09	0,47	317,26	73,88	0,17	0,83
Desvio-padrão	0,31	0,68	17,81	8,59	0,41	0,91
Coef. de Variação(%)	221,43	188,89	286,33	370,26	178,26	49,73
Mediana	0,04	0,00	0,12	0,03	0,12	1,27
Mínimo	0,00	0,00	0,01	0,00	1,85	0,80
Máximo	1,82	3,42	85,41	51,44	0,02	3,93
1 Quartil	0,00	0,00	0,05	0,01	0,00	0,98
3 Quartil	0,17	0,52	0,59	0,19	0,32	2,63
Assimetria	4,40	3,35	3,67	5,05	2,99	0,38
Curtose	22,49	12,45	13,68	27,99	9,81	-1,29
N	41	41	49	49	41	41

²: (x 0,01 mg/L)

TABELA 5.2 (Cont.) – Estatísticas descritivas para o conjunto de dados (mg/L)

	Mo	Se ²	Sr	Zn
Média	0,04	1,60	0,10	0,02
Variância	0,03	3,07	0,04	0,00
Desvio-padrão	0,16	1,75	0,19	0,02
Coef. de Variação(%)	400,00	109,38	190,00	100,00
Mediana	0,00	0,00	0,03	0,00
Mínimo	0,00	0,00	0,00	0,00
Máximo	0,88	4,12	0,94	0,07
1 Quartil	0,00	0,00	0,01	0,00
3 Quartil	0,00	3,39	0,16	0,03
Assimetria	4,36	0,18	3,26	0,64
Curtose	20,64	-2,03	11,21	-0,64
N	41	41	41	41

5.1.2 Descrição bivariada

São apresentadas, na Tabela 5.3, as relações entre os contaminantes com níveis de concentração acima do limite máximo tolerável, ou seja, para o Al, Se, Fe, Mn, As, Cu, Cd e Pb. O coeficiente de correlação de Pearson (ρ) fornece uma medida da relação linear entre as variáveis e é complementado pelo coeficiente de Spearman (ρ_{rank}), que mede a intensidade da relação entre as variáveis considerando o posto (*rank*) de cada valor na distribuição. Ao contrário do coeficiente de correlação de Pearson, o coeficiente de Spearman não é afetado por valores extremos. Grandes diferenças entre os dois podem, portanto, indicar a presença de valores extremos ou a existência de relação não-linear entre as variáveis.

As correlações mais fortes foram apresentadas pelo par chumbo (Pb) e selênio (Se) e ferro (Fe) com cádmio (Cd), tanto para correlação linear quanto para correlação *rank* de Spearman. As amostras não conseguiram detectar correlação linear entre os pares ferro vs. manganês, arsênio vs. manganês, cobre vs. ferro e manganês e selênio, entretanto, o coeficiente de Spearman mostrou-se significativo ao nível de 1%, indicando associação entre valores elevados (ou baixos) desses elementos. Na Tabela 5.3 são apresentados os resultados para os demais pares. A Figura 5.5 do diagrama de

dispersão do coeficiente de correlação linear versus o coeficiente de correlação *rank* mostra que as medidas apresentam grandes diferenças, uma indicação da presença de pares extremos (outliers), que podem influir nos coeficientes de correlação linear. Observa-se que os coeficientes de correlação de Spearman são em sua grande maioria, maiores do que os respectivos coeficientes de correlação linear apontando para possíveis relações não lineares entre as variáveis (ISAACS, SHRIVASTAVA, 1989).

TABELA 5.3 – Matriz de coeficientes de correlação de Pearson e Spearman

		<i>Correlações de Spearman</i>							
		As	Mn	Al	Fe	Pb	Se	Cd	Cu
Correlações de Pearson	As	1	0.573	-0.281	0.380	0.340	0.290	0.408	0.130
	Mn	0.024	1	0.192	0.600	0.574	0.558	0.524	0.364
	Al	-0.094	-0.090	1	0.344	0.310	0.164	0.156	0.252
	Fe	0.070	0.166	0.369	1	0.561	0.590	0.676	0.575
	Pb	0.172	0.336	0.105	0.415	1	0.884	0.480	0.888
	Se	0.093	0.344	0.096	0.405	0.965	1	0.468	0.888
	Cd	0.085	0.196	0.349	0.934	0.475	0.428	1	0.402
	Cu	-0.046	-0.006	-0.026	0.050	0.525	0.559	0.047	1

Valores com destaque: significativos a menos de 5%

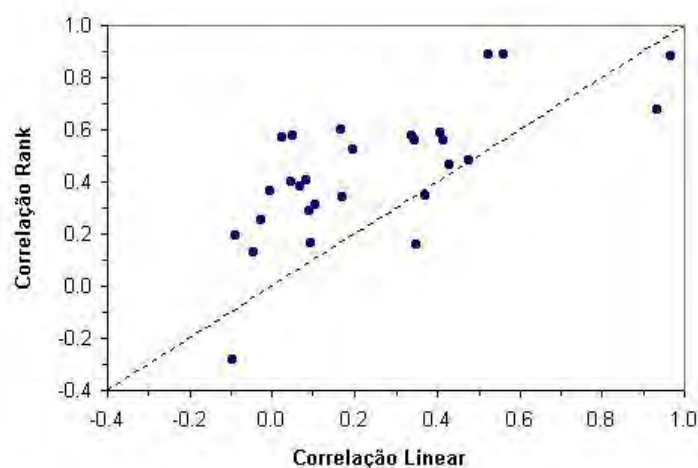


Figura 5.5 – Diagrama de dispersão do coeficiente de correlação linear versus coeficiente de correlação de Spearman para os 28 pares de poluentes que excedem o limite máximo permitido.

5.2 Análise espacial do risco de contaminação

Na quantificação e qualificação (classificação) do risco de contaminação na área de estudo serão utilizadas técnicas geoestatísticas que permitem a caracterização da área no sentido de se conhecer a origem dos contaminantes, estimativas da magnitude, extensão e localização das áreas contaminadas em localizações não amostradas e estimativas de custos de saúde e de remediação, com a inclusão das medidas de incertezas associadas.

5.2.1 Mapas de localização dos poluentes

São apresentados nas Figuras 5.6 a 5.13 os mapas de localização das amostras dos elementos com níveis de contaminação acima da máxima permitida pela legislação, ou seja, para o Al, Se, Fe, Mn, As, Cu, Pb e Cd. Todos os mapas apresentados neste trabalho foram construídos com auxílio do programa Surfer 7.00 (SURFER 7, 1999).

Para o alumínio, considerando-se o limite máximo tolerável de 0,1 ppm estabelecido pelo CONAMA, observa-se na Figura 5.6 a presença de amostras acima do referido limite espalhadas em praticamente toda a área de estudo. A maioria das amostras de selênio acima do limite estabelecido pelo CONAMA (0,01 ppm) encontram-se na área industrial da ICOMI e no bairro do Elesbão, como mostra a Figura 5.7. As maiores concentrações de valores acima dos limites estabelecidos de 0,3 ppm, 0,1 ppm e 0,05 ppm para amostras de ferro, manganês e arsênio, respectivamente, encontram-se dentro da área industrial da ICOMI, mostradas nas Figuras 5.8, 5.9 e 5.10, e podem refletir as fontes de contaminação em potencial. As amostras de chumbo encontram-se dispersas na área de estudo, observando-se, entretanto, um pequeno agrupamento de valores mais baixos na região NE. Os dois valores de concentração acima do limite estabelecido (0,03 ppm) foram medidos dentro da área industrial da ICOMI (Figura 5.11). As concentrações medidas para as amostras de cobre, mostradas na Figura 5.12, apresentam-se dispersas por toda a área de estudo, com predomínio de valores baixos de concentração na área residencial. O cádmio apresentou dois valores altos de concentração dentro da área industrial da ICOMI e outro no bairro da hospitalidade, próximo ao Igarapé Boiero (Figura 5.13).

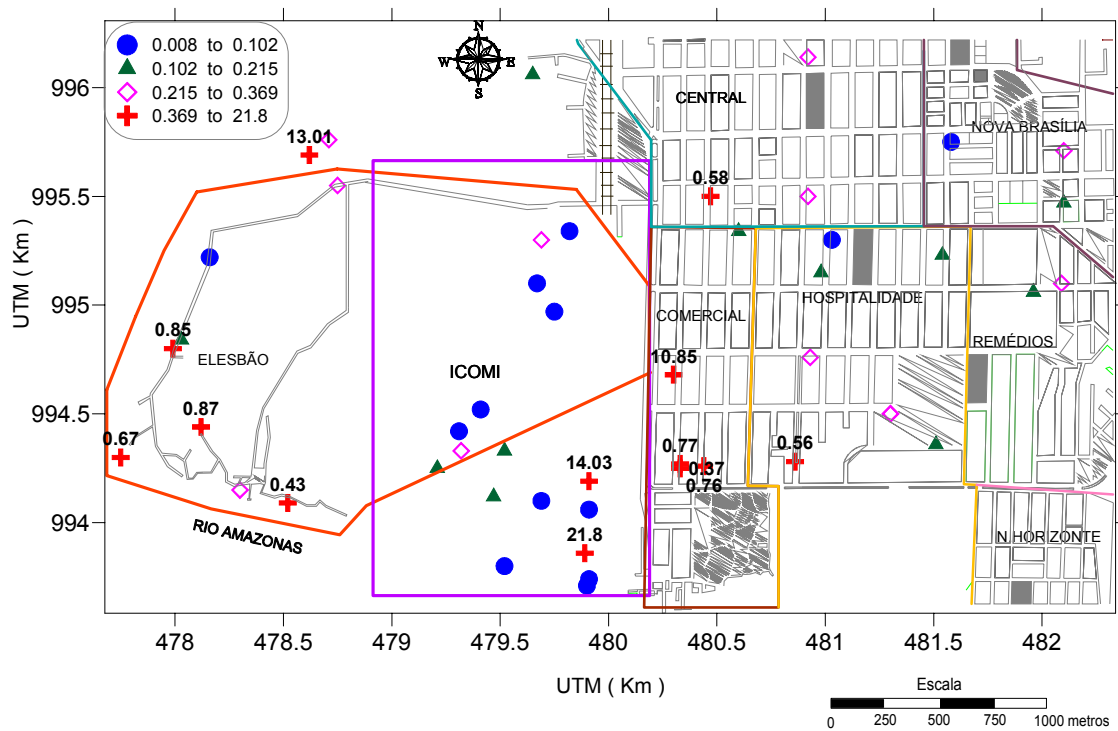


Figura 5.6 – Mapa de localização das amostras de alumínio. As 4 faixas de valores de concentração (em ppm) referem-se aos 1^o, 2^o, 3^o e 4^o quartis, respectivamente.

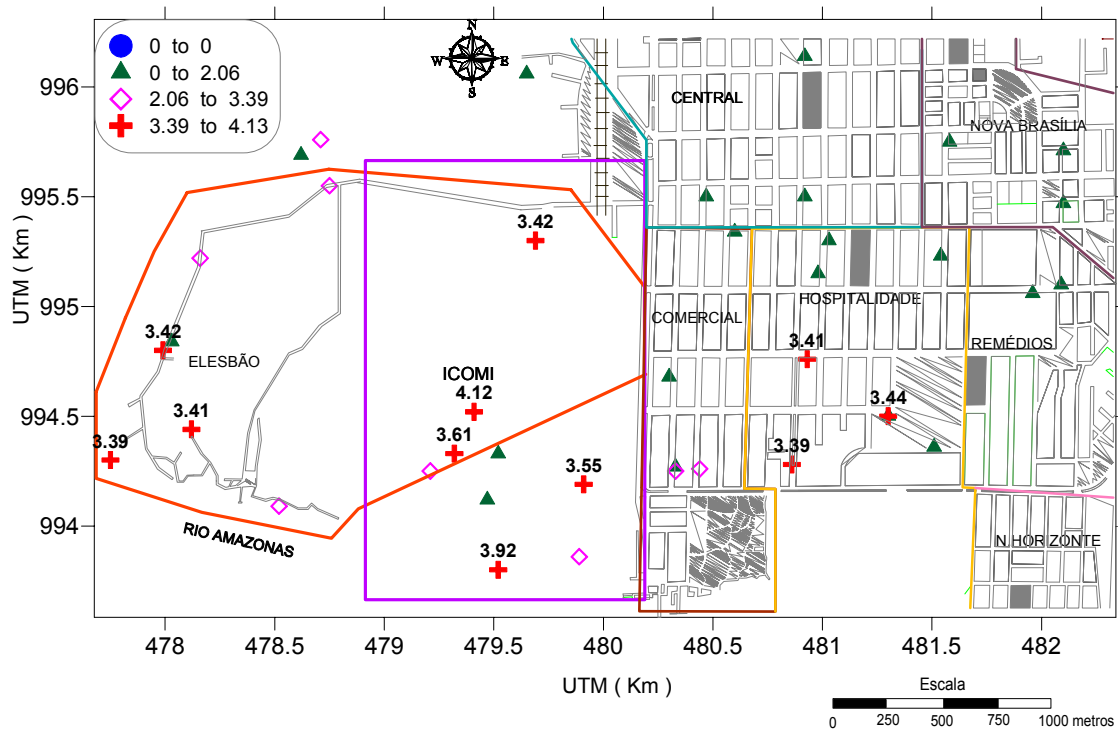


Figura 5.7 – Mapa de localização das amostras de selênio. As 4 faixas de valores de concentração (x 0,01 ppm) referem-se aos 1^o, 2^o, 3^o e 4^o quartis, respectivamente.

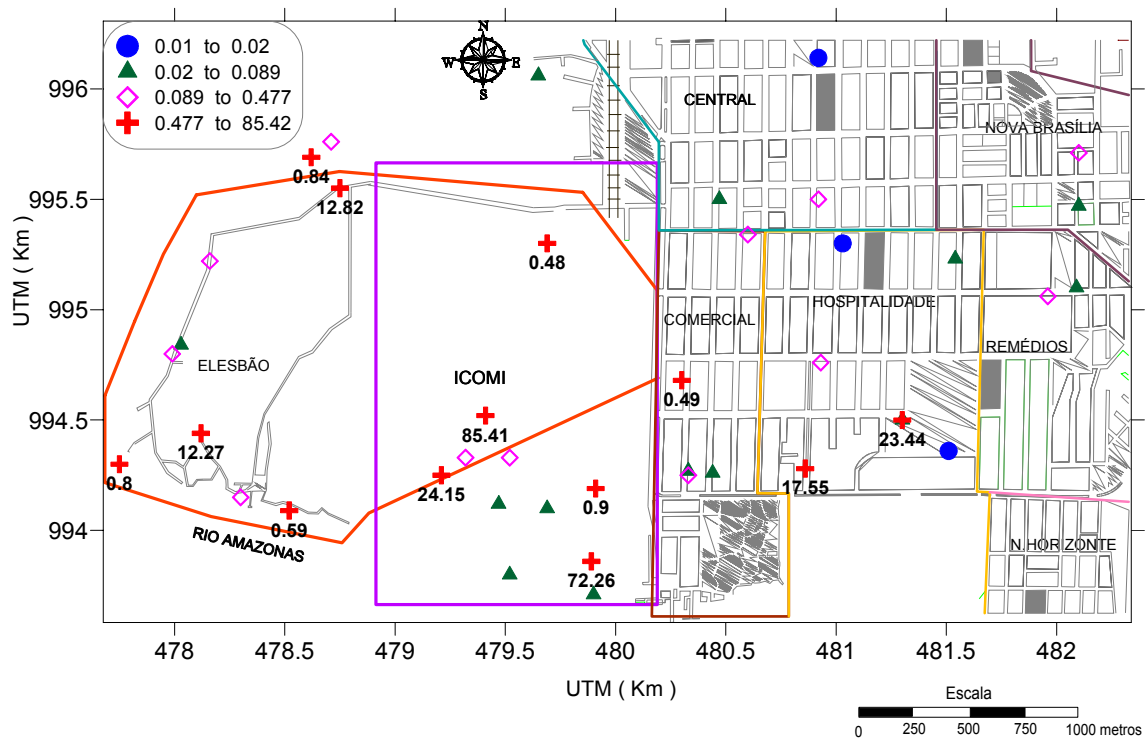


Figura 5.8 – Mapa de localização das amostras de ferro. As 4 faixas de valores de concentração (em ppm) referem-se aos 1^o, 2^o, 3^o e 4^o quartis, respectivamente.

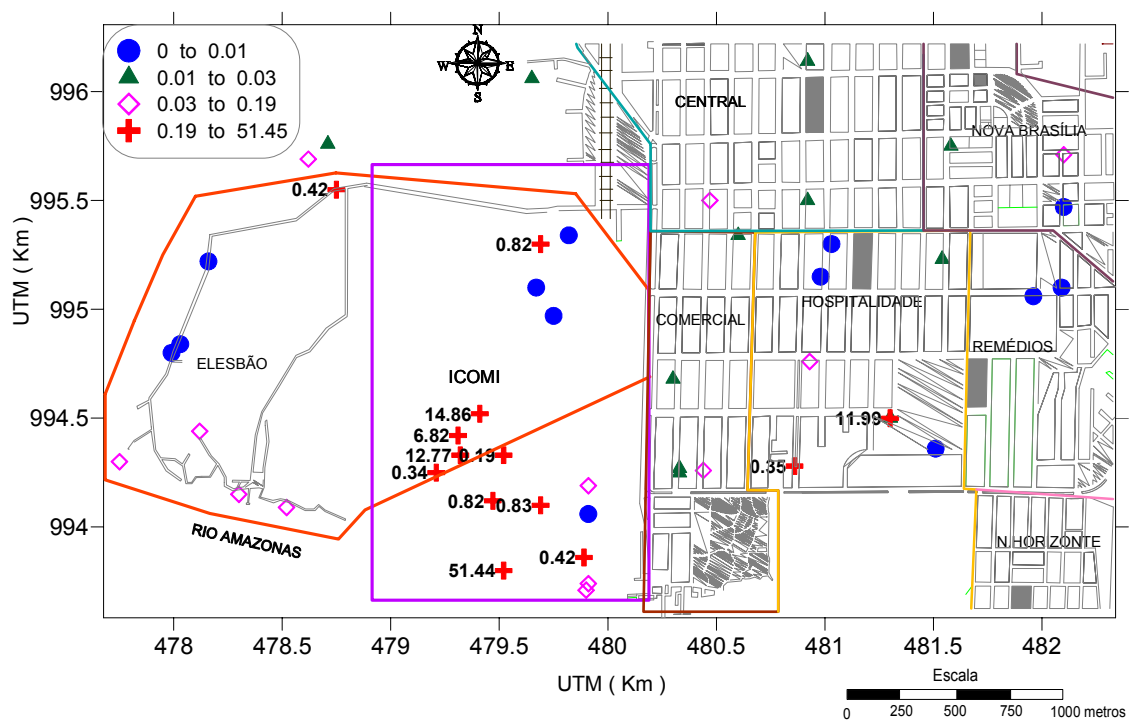


Figura 5.9 – Mapa de localização das amostras de manganês. As 4 faixas de valores de concentração (em ppm) referem-se aos 1^o, 2^o, 3^o e 4^o quartis, respectivamente.

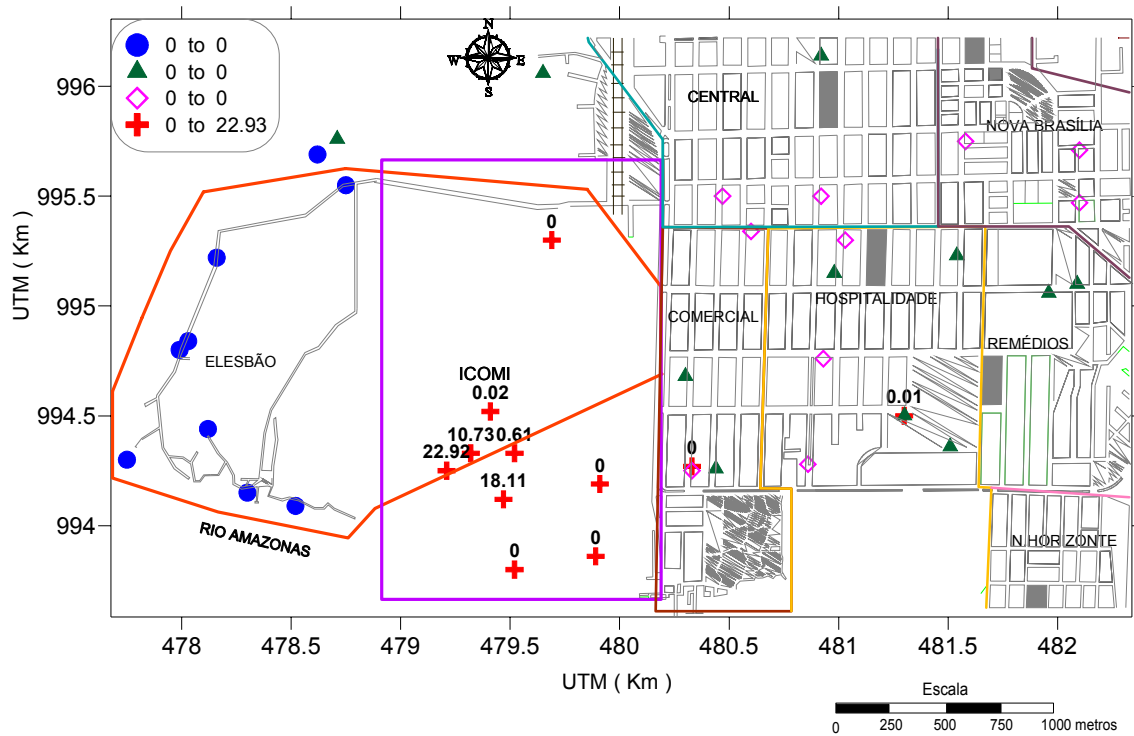


Figura 5.10 – Mapa de localização das amostras de arsênio. As 4 faixas de valores de concentração (em ppm) referem-se aos 1^o, 2^o, 3^o e 4^o quartis, respectivamente.

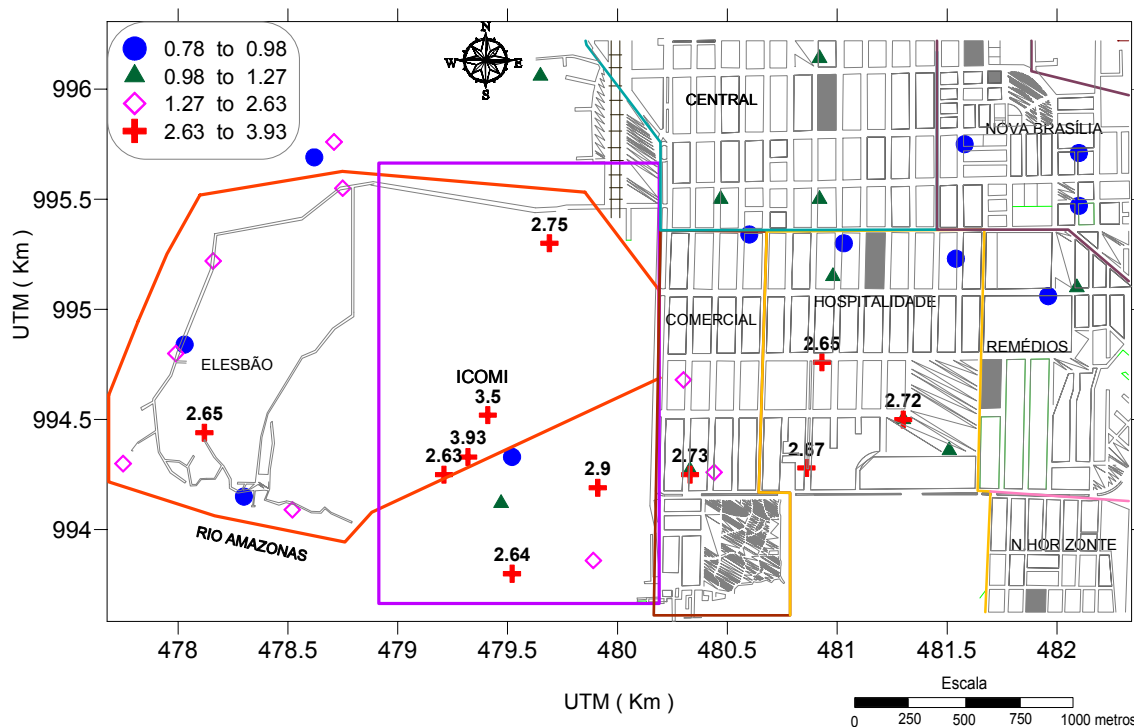


Figura 5.11 – Mapa de localização das amostras de chumbo. As 4 faixas de valores de concentração (× 0,01 ppm) referem-se aos 1^o, 2^o, 3^o e 4^o quartis, respectivamente.

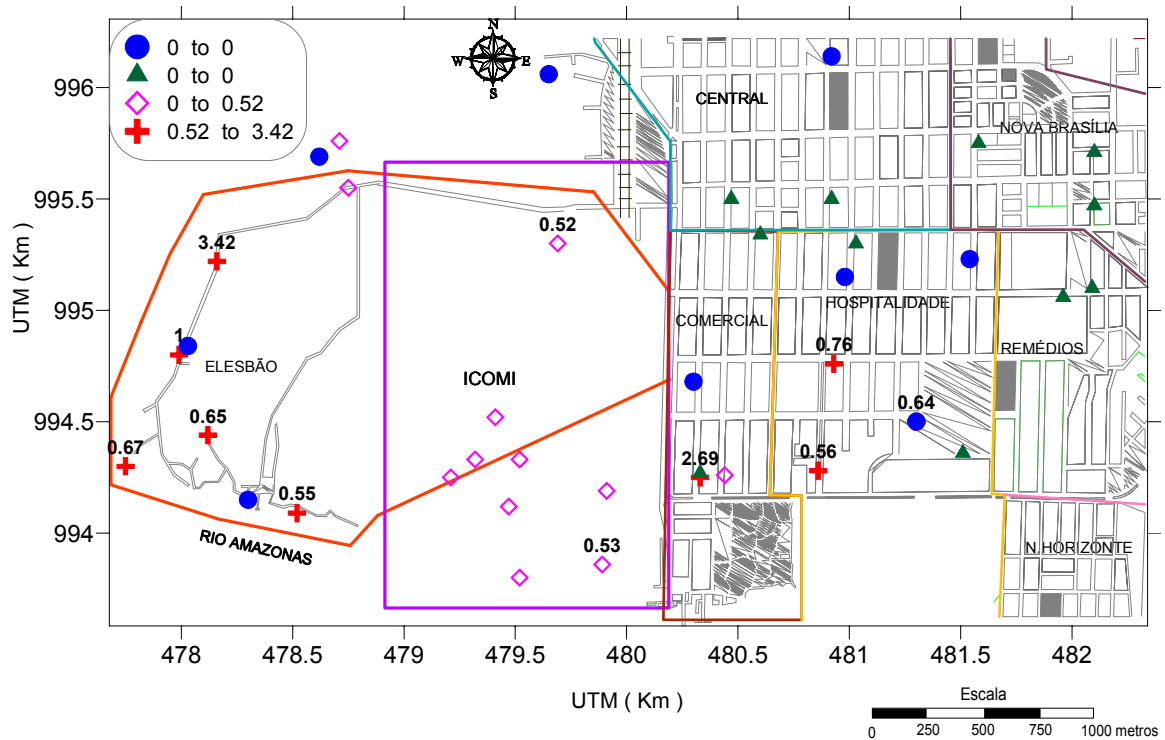


Figura 5.12 – Mapa de localização das amostras de cobre. As 4 faixas de valores de concentração ($\times 0,01$ ppm) referem-se aos 1^o, 2^o, 3^o e 4^o quartis, respectivamente.

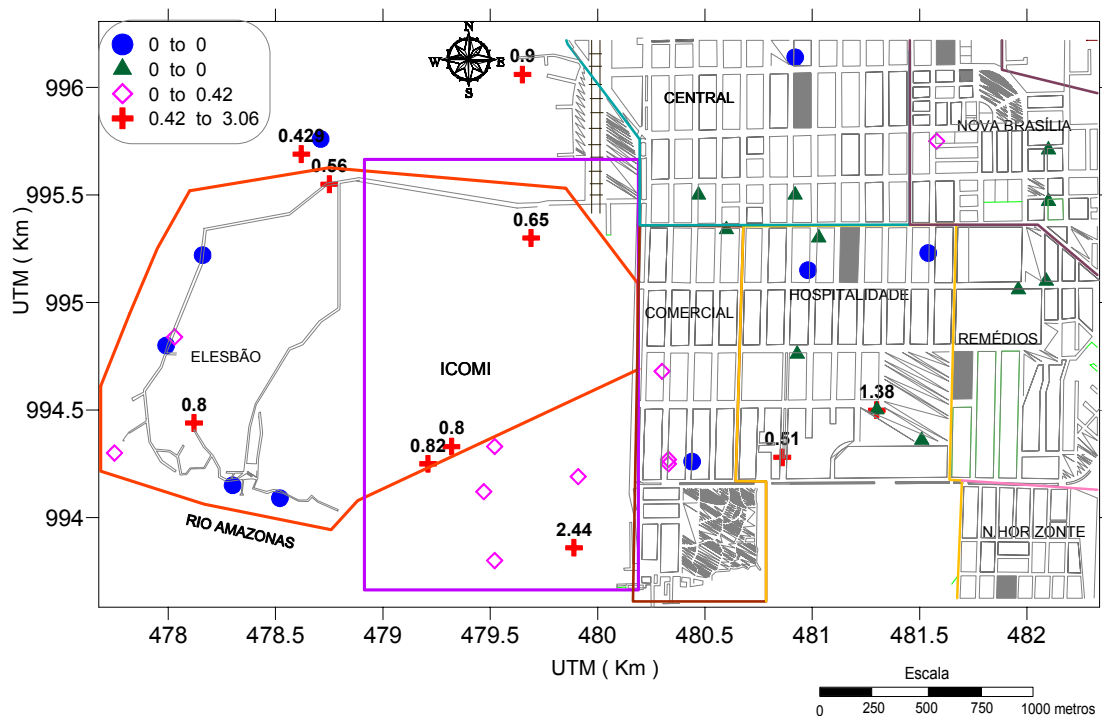


Figura 5.13 – Mapa de localização das amostras de cádmio. As 4 faixas de valores de concentração ($\times 0,001$ ppm) referem-se aos 1^o, 2^o, 3^o e 4^o quartis, respectivamente.

5.2.2 Descrição dos padrões espaciais

Em estudos de análises de risco ambientais sobre poluição/contaminação é necessário que seja realizada uma investigação para se conhecer a fonte, a distribuição, mobilidade e comportamento do poluente ou poluentes na área em estudo, entre outros pontos. A presença dos poluentes no ambiente pode ser originada naturalmente das rochas ou processos químicos naturais ou então de atividades humanas tais como mineração, resíduos industriais ou agricultura. Esses fatores (humanos ou naturais) podem ocorrer em diferentes escalas e deveriam ser percebidos na análise semivariográfica se escalas nas quais eles ocorrem são muito diferentes uma da outra. Deste modo, pretende-se compreender as origens dos poluentes a partir das diferentes escalas de ocorrências dos valores observados.

Inicialmente, a partir do conhecimento das profundidades dos níveis d'água (N.A) obtidas durante o levantamento amostral, foi construído o mapa potenciométrico (Figura 5.14) na área de estudo para se conhecer o sentido do fluxo subterrâneo. O nível d'água (N.A) que separa o horizonte não saturado (acima do N.A) do saturado (abaixo do N.A) apresenta profundidades que variam desde centímetros nas proximidades das margens do rio, até valores em torno de 30 m em poços amazonas localizados no bairro do elesbão. Essas profundidades oscilam ao longo do ano, devido às variações sazonais da superfície potenciométrica das águas subterrâneas. Deste modo, para as informações obtidas no período do levantamento (dezembro/2001), observa-se na Figura 5.14, um fluxo subterrâneo movimentando-se radialmente a partir do centro da área, região compreendida entre os poços 16 e 21, em direção ao rio Amazonas e demais áreas vizinhas.

Os padrões espaciais são, em geral, descritos utilizando-se os semivariogramas experimentais, conforme já apresentados no capítulo 3. Devido ao pequeno número de dados disponíveis e considerando que os mapas de localizações das amostras não apresentaram, em geral, qualquer indício forte de variação dependente de uma direção em particular, foram calculados somente os semivariogramas experimentais omnidirecionais para as variáveis em estudo.

Nas Figuras 5.15 a 5.22 são apresentados os semivariogramas experimentais omnidirecionais para as concentrações dos elementos em estudo. Foram calculados os semivariogramas dos logaritmos e plotados no mesmo gráfico, depois de se realizar o rescalonamento à variância da variável original, para as variáveis que apresentaram

todos os valores positivos e diferentes de zero. Nesses casos, praticamente todos os semivariogramas dos logaritmos se mostram mais suavizados. As diferenças, entretanto, não foram bastante acentuadas para justificar uma log-transformação nas variáveis.

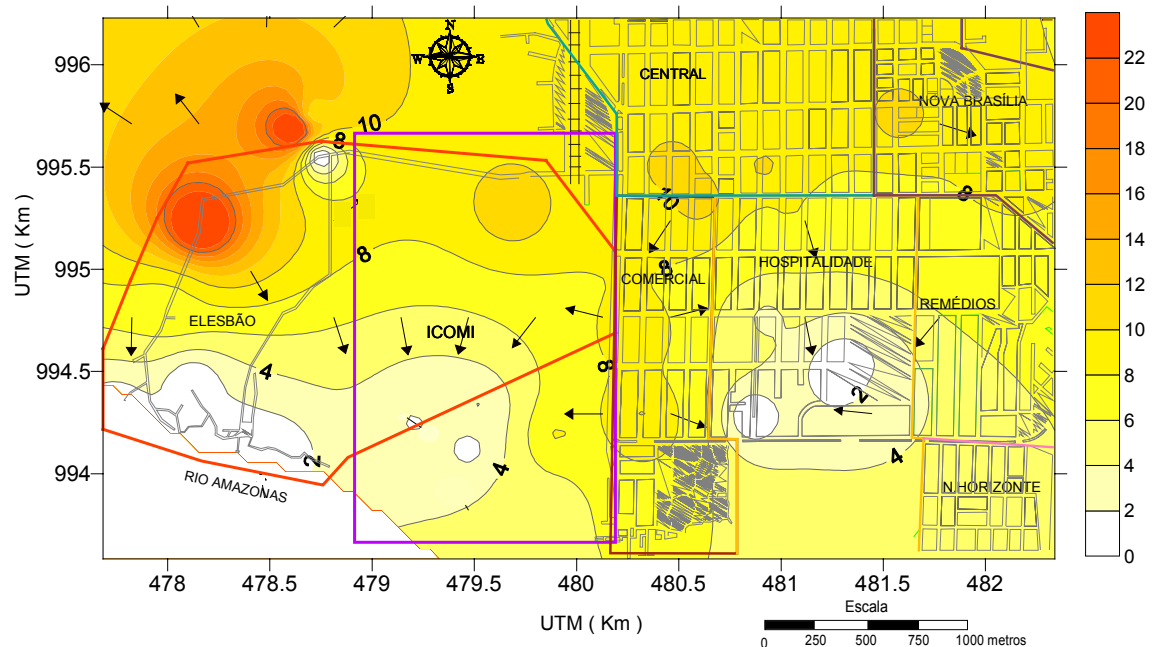


Figura 5.14 – Mapa potenciométrico da área industrial e residencial de Santana

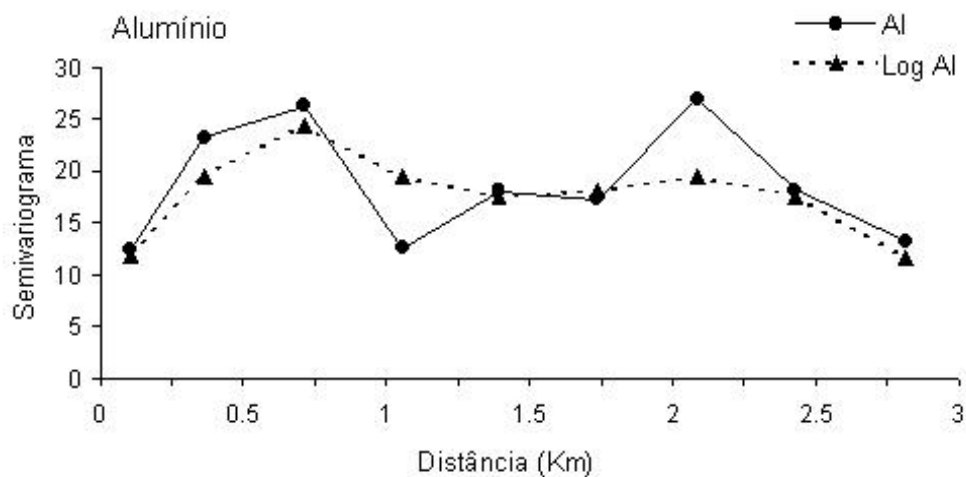


Figura 5.15 – Semivariograma experimental omnidirecional para o alumínio. O semivariograma dos logaritmos (linha pontilhada) foi rescalonado à variância original da variável.

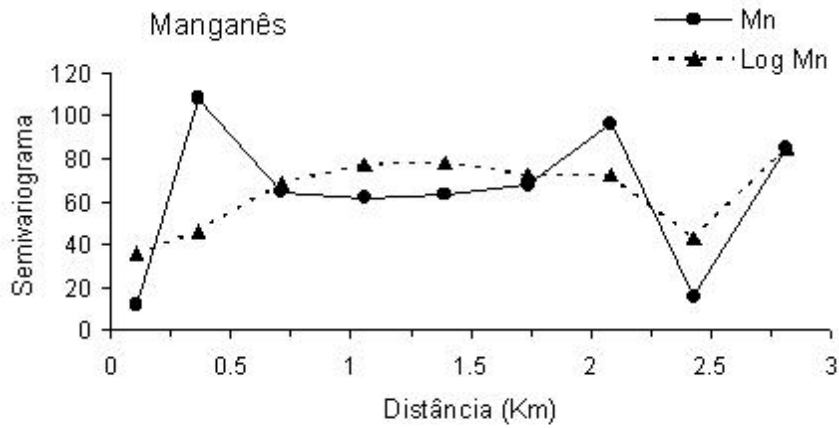


Figura 5.16 – Semivariograma experimental omnidirecional para o manganês. O semivariograma dos logaritmos foi rescalonado à variância original da variável.

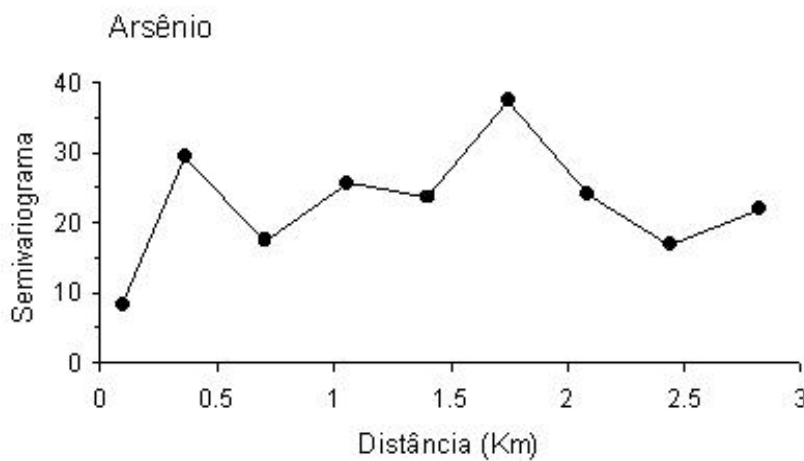


Figura 5.17 – Semivariograma experimental omnidirecional para o arsênio.

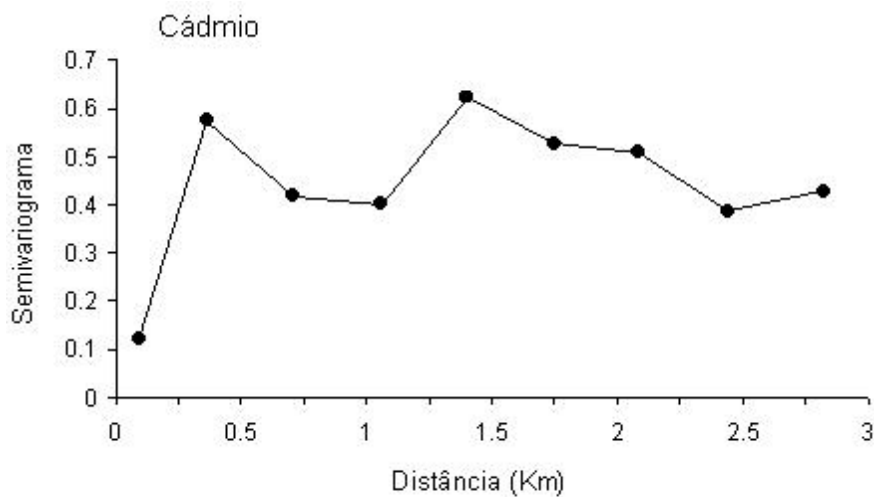


Figura 5.18 – Semivariograma experimental omnidirecional para o cádmio

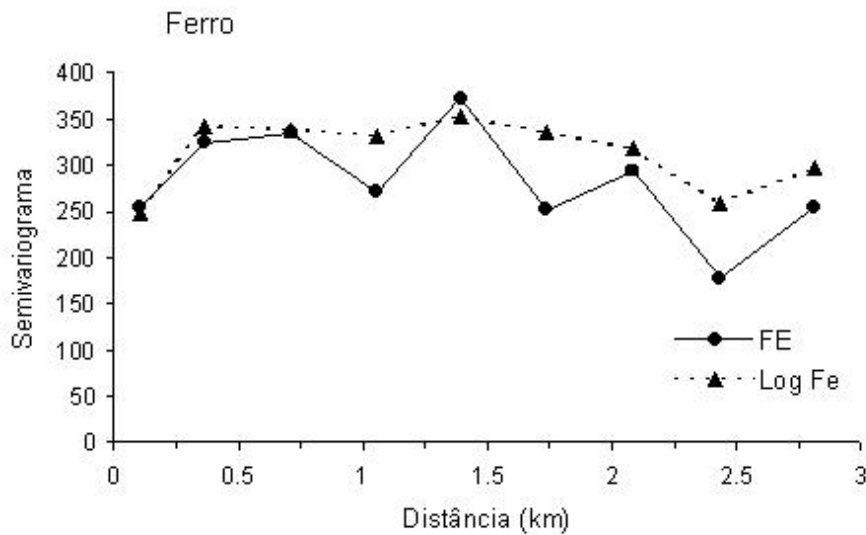


Figura 5.19 – Semivariograma experimental omnidirecional para o ferro. O semivariograma dos logaritmos (linha pontilhada) foi rescalonado à variância original da variável.

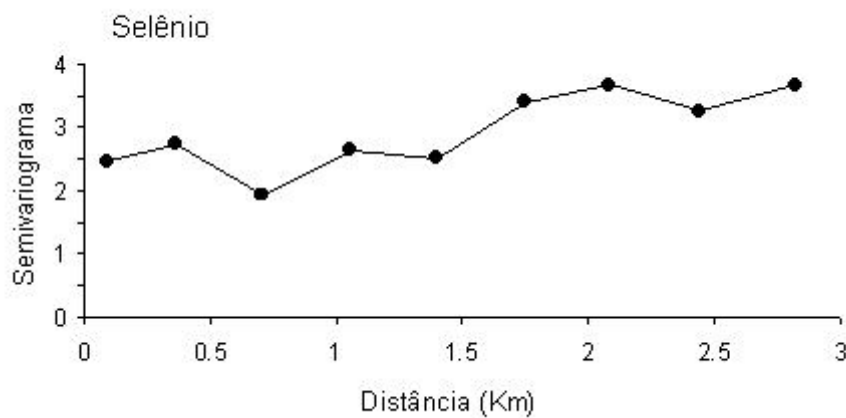


Figura 5.20 – Semivariograma experimental omnidirecional para o selênio.

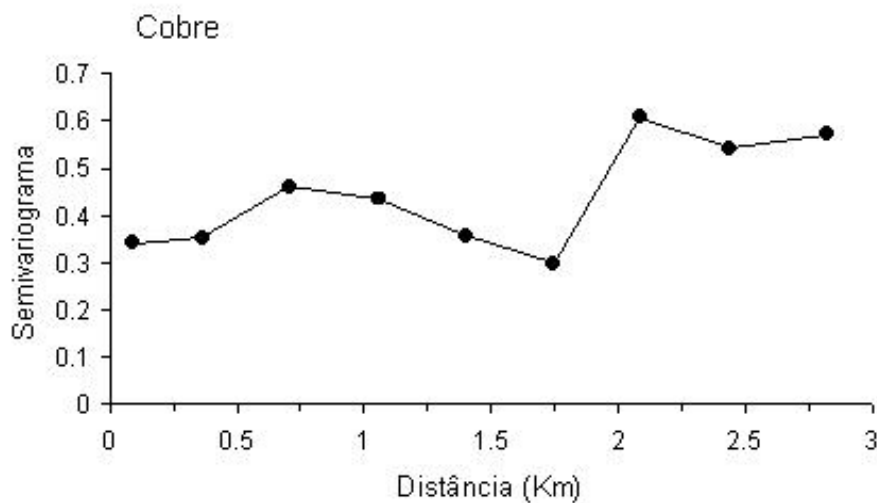


Figura 5.21 – Semivariograma experimental omnidirecional para o cobre.

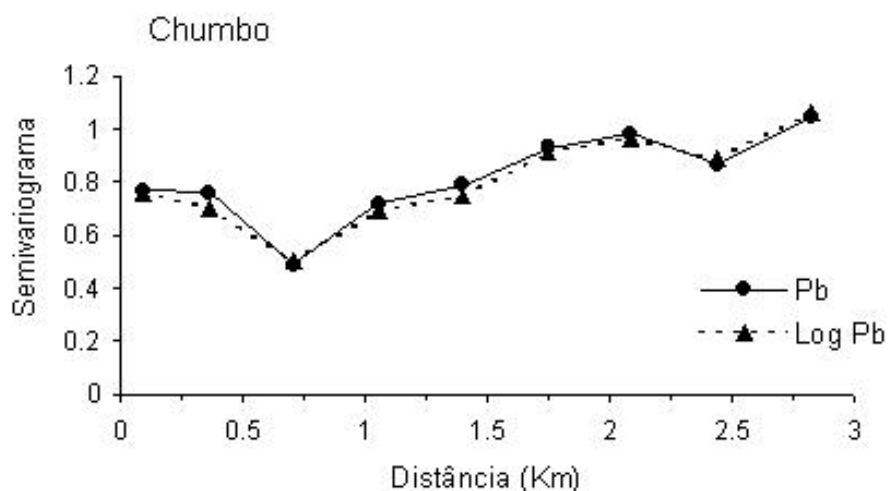


Figura 5.22 – Semivariograma experimental omnidirecional para o chumbo. O semivariograma dos logaritmos (linha pontilhada) foi rescalonado à variância original da variável.

As formas dos semivariogramas experimentais podem conduzir ao modelamento da variabilidade das variáveis como uma soma de diversas estruturas básicas (ou semivariogramas), sendo cada estrutura diretamente relacionada à escala espacial. Deste modo, a partir das inflexões das curvas dos semivariogramas experimentais pode-se distinguir, além das variações de micro-escala identificadas pelo efeito pepita, pelo menos duas escalas de variação espacial: uma variando de 0,25 a 1 km, que poderia ser considerada como uma escala local; e outra com variação entre 2 a 2,5 km, que poderia caracterizar uma escala regional. A estrutura de alcance mais longo, que identifica a escala regional, é mais visível na forma dos semivariogramas do selênio e chumbo, enquanto para o cádmio e ferro pode-se perceber melhor a estrutura de menor alcance, associada à escala local. Nos semivariogramas do arsênio, alumínio e do cobre pode-se perceber uma combinação das duas estruturas.

5.2.3 Modelamento da co-regionalização

Embora os semivariogramas experimentais forneçam informações sobre a continuidade espacial das variáveis, para se fazer inferências sobre a(s) variável(is) é necessário que se construa um modelo semivariográfico de modo que seja possível computar um valor do semivariograma em qualquer separação possível. Na realização da modelagem deve-se, portanto, ter disponível os valores do semivariograma (ou covariância) experimental, utilizar modelos semivariográficos permissíveis de acordo

com as restrições apresentadas no capítulo 3 e, sempre que for possível, recorrer a informações secundárias, tais como as fornecidas pelo conhecimento físico ou geologia da área e do fenômeno (variável) em estudo. Essas diferentes fontes de informações devem ser utilizadas para se construir um modelo permissível que consiga capturar as principais características espaciais da variável em estudo.

No modelamento da co-regionalização é necessário a inferência dos semivariogramas diretos e cruzados das variáveis em estudo. A maior dificuldade, neste caso, está na exigência de se realizar a modelagem dos variogramas conjuntamente, respeitando as restrições impostas aos modelos, descritas no capítulo 3. A utilização do modelo linear de co-regionalização proposto por GOULARD (1989) facilita o modelamento no caso de muitas variáveis, como no presente estudo. PARDO-IQUIZGUIZA e DOWD (2002) apresentam um programa (LCMFIT2) para o ajustamento simultâneo dos coeficientes (sill) dos semivariogramas, baseado no algoritmo de mínimos quadrados ponderados de GOULART e VOLTZ (1992), que avalia (e corrige) a condição de positiva-definida da matriz de coeficientes.

4.2.3.1 Krigagem fatorial multivariada

Neste estudo as informações são espaciais e multivariadas, uma vez que foram analisadas concentrações de diversos elementos químicos em cada amostra coletada. A partir das correlações observadas entre as variáveis, pretende-se avaliar os fatores básicos comuns relacionados a essas correlações, considerando-se a informação espacial e as diferentes escalas identificadas nas análises semivariográficas. Tal atenção à dependência da escala pode mostrar de modo mais detalhado a relação entre as variáveis e conduzir a um melhor entendimento dos mecanismos físicos básicos que controlam os padrões espaciais, que é de algum modo obscurecida em uma aproximação onde todas as diferentes fontes de informação são misturadas. Esta análise pode ser realizada por meio da krigagem fatorial multivariada que envolve o modelamento semivariográfico multivariado, análise de componentes principais e cokrigagem. Esse método permite a análise entre as diferentes estruturas de correlação espacial encontradas nas diversas escalas espaciais e produz um conjunto de fatores regionalizados que resumem as principais características do conjunto de dados em cada escala espacial.

Foram consideradas na análise multivariada somente as variáveis que apresentaram concentrações acima dos limites estabelecidos pelo CONAMA, ou seja,

arsênio (As), manganês (Mn), alumínio (Al), ferro (Fe), chumbo (Pb), selênio (Se), cádmio (Cd) e cobre (Cu).

A Tabela 5.4 apresenta a matriz de correlações entre as variáveis. Os valores em destaque indicam correlação significativa a um nível abaixo de 5%. Observa-se forte correlação entre Fe e Cd ($r = 0.934$) e com Pb com Se ($r = 0.965$). As relações entre as variáveis foram, primeiramente, estudadas de uma maneira clássica, com a aplicação do método de análise de componentes principais (ACP).

TABELA 5.4 – Matriz de correlação das variáveis.

	<i>As</i>	<i>Mn</i>	<i>Al</i>	<i>Fe</i>	<i>Pb</i>	<i>Se</i>	<i>Cd</i>	<i>Cu</i>
As	1.000							
Mn	0.024	1.000						
Al	-0.094	-0.090	1.000					
Fe	0.070	0.166	0.369	1.000				
Pb	0.172	0.336	0.105	0.415	1.000			
Se	0.093	0.344	0.096	0.405	0.965	1.000		
Cd	0.085	0.196	0.349	0.934	0.475	0.428	1.000	
Cu	-0.046	-0.006	-0.026	0.050	0.525	0.559	0.047	1.000

A idéia básica da ACP é criar novas variáveis ortogonais (componentes principais) que são combinações lineares das variáveis originais. Isso é feito através do cálculo dos autovalores (que medem a percentagem da variância total das medidas explicadas pela componente), e autovetores da matriz de correlações. As componentes principais são então ordenadas de acordo com sua contribuição à variância total dos dados. Em geral, as primeiras componentes explicam a maior parte da variância e, desta forma, são mais informativas. A estrutura global da correlação descrita pela matriz de correlação (ou variância-covariância) pode ser representada graficamente pela correlação (chamada de carga dos fatores ou componentes) entre as variáveis originais e as componentes principais dada por (JOHNSON e WICHERN, 1992):

$$\rho_{ik} = \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}^2}} \quad (5.3)$$

onde e_{ki} e λ_i são, respectivamente, os autovetores e autovalores da matriz de correlação (ou variância-covariância) e σ_{kk}^2 são as variâncias das variáveis originais. A análise de componentes principais foi realizada sobre a matriz de correlações entre as variáveis. As componentes principais ordenadas de acordo com sua contribuição à explicação da variância total dos dados são mostradas na Tabela 5.5. Essa Tabela apresenta as cargas dos fatores com a percentagem da variância explicada antes e depois da rotação ortogonal (varimax) das três primeiras componentes principais. Os respectivos gráficos das três componentes são mostrados na Figura 5.23. Foram escolhidos as três primeiras componentes principais que apresentaram autovalores maiores do que 1. Essas componentes explicam, juntas, 74% da variância total. A primeira componente é fortemente correlacionada com os elementos Se, Pb, Cd e Fe. Pode-se observar na Figura 5.23 (gráfico superior à esquerda) que a segunda componente separa esses elementos em dois grupos: Cu-Se-Pb-Mn-As e Al-Cd-Fe. A terceira componente separa os elementos em três grupos: Se-Pb-Fe-Cd, As-Mn e Al-Cu (Figura 5.23, gráfico superior à direita).

Os semivariogramas experimentais omnidirecionais diretos e cruzados obtidos a partir dos escores das três componentes principais são apresentados nas Figura 5.24 e 5.25. Um modelo linear de co-regionalização foi ajustado e utilizado no mapeamento das componentes, onde, cada componente foi estimada por cokrigagem. Foram ajustados dois modelos esféricos com alcances de 0,33 km e 2,0 km, respectivamente, além do efeito pepita presente em todos os casos. Observa-se uma pequena predominância da estrutura de longo alcance (2,0 km) na componente principal 1 e da estrutura de pequeno alcance (0,33 km) na componente principal 3. Embora nenhuma estrutura de longo ou pequeno alcance predomine fortemente em quaisquer das componentes, selênio e chumbo que apresentaram maior correlação com a primeira componente podem estar vinculadas à estrutura de longo alcance; e pode-se supor que arsênio e manganês, com maior correlação com a terceira componente, estejam mais ligadas à estrutura de pequeno alcance. Pode-se observar nos gráficos dos semivariogramas cruzados da Figura 5.25, que as três componentes apresentam correlações negativas entre si.

Uma vez que o espaçamento médio entre as amostras ficou em torno de 140 m, foi estabelecida uma malha de 0,05 Km x 0,05 Km para a elaboração dos mapas espaciais das componentes, de acordo com a regra prática referida em MYERS(1997, p. 492). Nas Figuras 5.26, 5.27 e 5.28 são apresentados os mapas cokrigados das três primeiras componentes, obtidos com auxílio do programa FACTOR2D (PARDO-IQUIZGUIZA e DOWD, 2002)

TABELA 5.5 – Cargas dos fatores com a percentagem da variância explicada das três primeiras Componentes Principais

Variáveis	Cargas das componentes principais		
	Cp 1	Cp 2	Cp 3
As	0.142	0.073	– 0.667
Mn	0.387	0.177	– 0.508
Al	0.303	– 0.584	0.425
Fe	0.753	– 0.560	– 0.059
Pb	0.888	0.356	0.005
Se	0.876	0.389	0.066
Cd	0.779	– 0.529	– 0.091
Cu	0.460	0.597	0.467
Autovalores	3.202	1.607	1.118
% da Variância	40.0	20.1	14.0
Acumulado	40.0	60.1	74.1

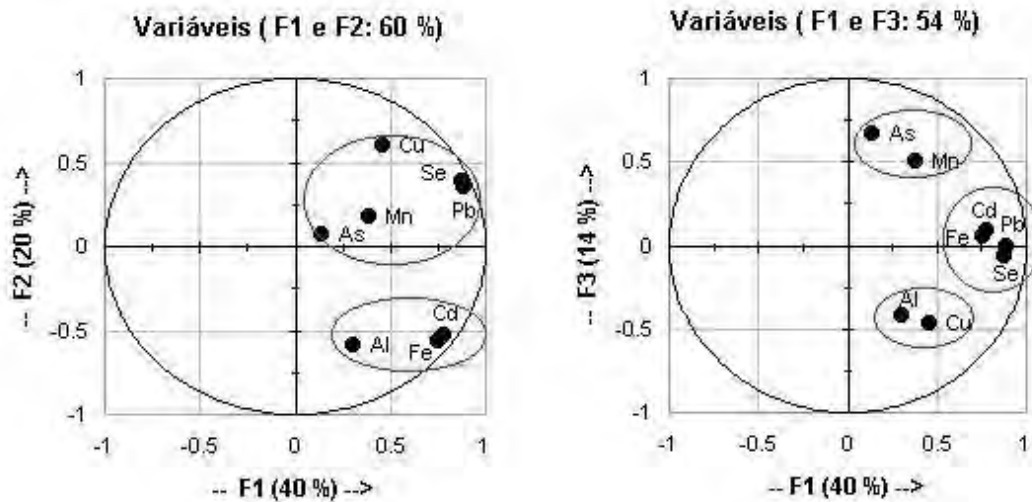


Figura 5.23 – Gráficos das correlações entre as variáveis originais e as componentes principais. Os eixos numerados são para a primeira, segunda e terceira componente principal, respectivamente. Os números entre parênteses são a percentagem da variância total explicada por cada componente.

Para a primeira componente, observam-se duas regiões pequenas com anomalias positivas dentro da área da ICOMI, que podem estar associadas com as variáveis mais correlacionadas com essa componente, ou seja, Se-Pb e Fe-Cd. No mapa da segunda componente, as duas regiões com anomalias negativas dentro da área da ICOMI podem estar associadas com as variáveis Al, Fe e Cd, que possuem razoável correlação negativa com essa componente (ver Tabela 5.5). No mapa da terceira componente, dentro dos limites da ICOMI, a anomalia negativa presente possivelmente relaciona-se ao arsênio e manganês, enquanto a anomalia positiva pode estar associada ao alumínio e cobre, que são as variáveis com maiores cargas negativa e positiva, respectivamente, com a terceira componente (ver Tabela 5.5). Observa-se ainda no bairro do Elesbão, duas pequenas regiões apresentando valores positivos relevantes. As anomalias observadas nos mapas das componentes podem estar associadas aos locais onde, possivelmente localizam-se as fontes de contaminação.

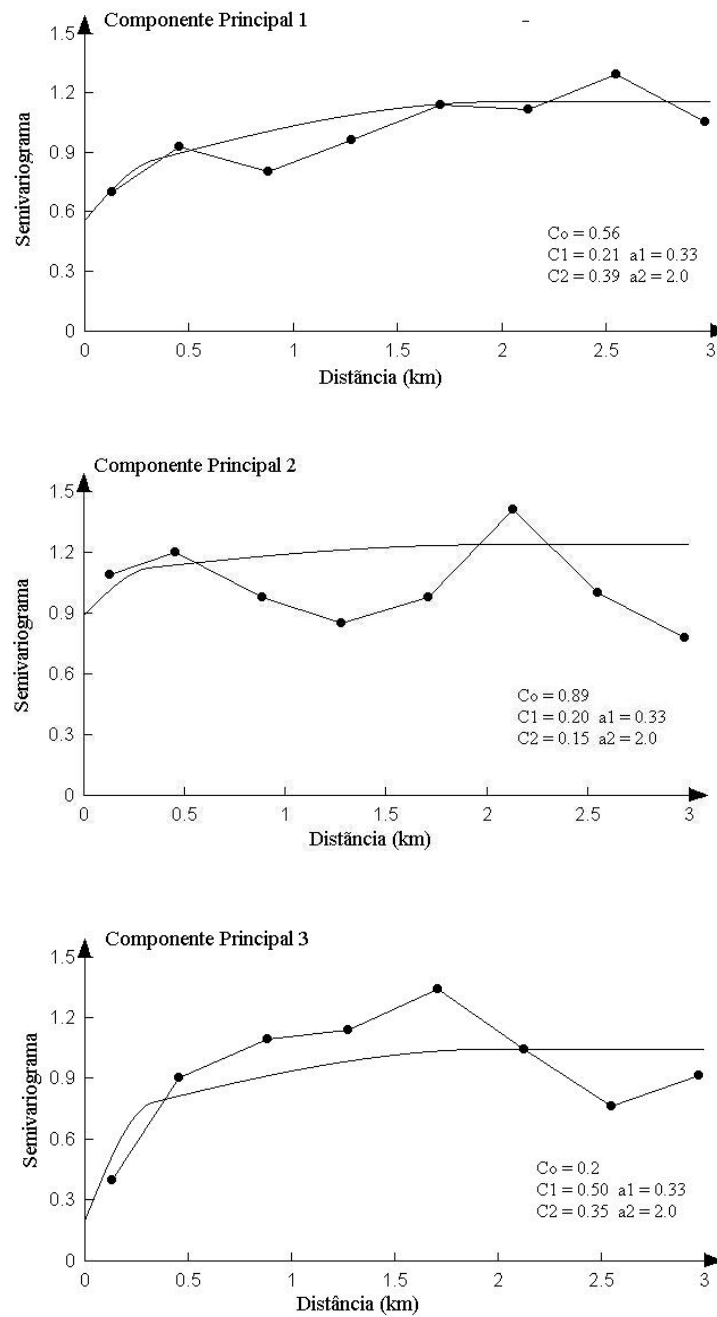


Figura 5.24 – Semivariogramas experimentais das três componentes principais (linhas ligadas por pontos) e os semivariogramas ajustados pelo modelo linear de co-regionalização (linhas contínuas)

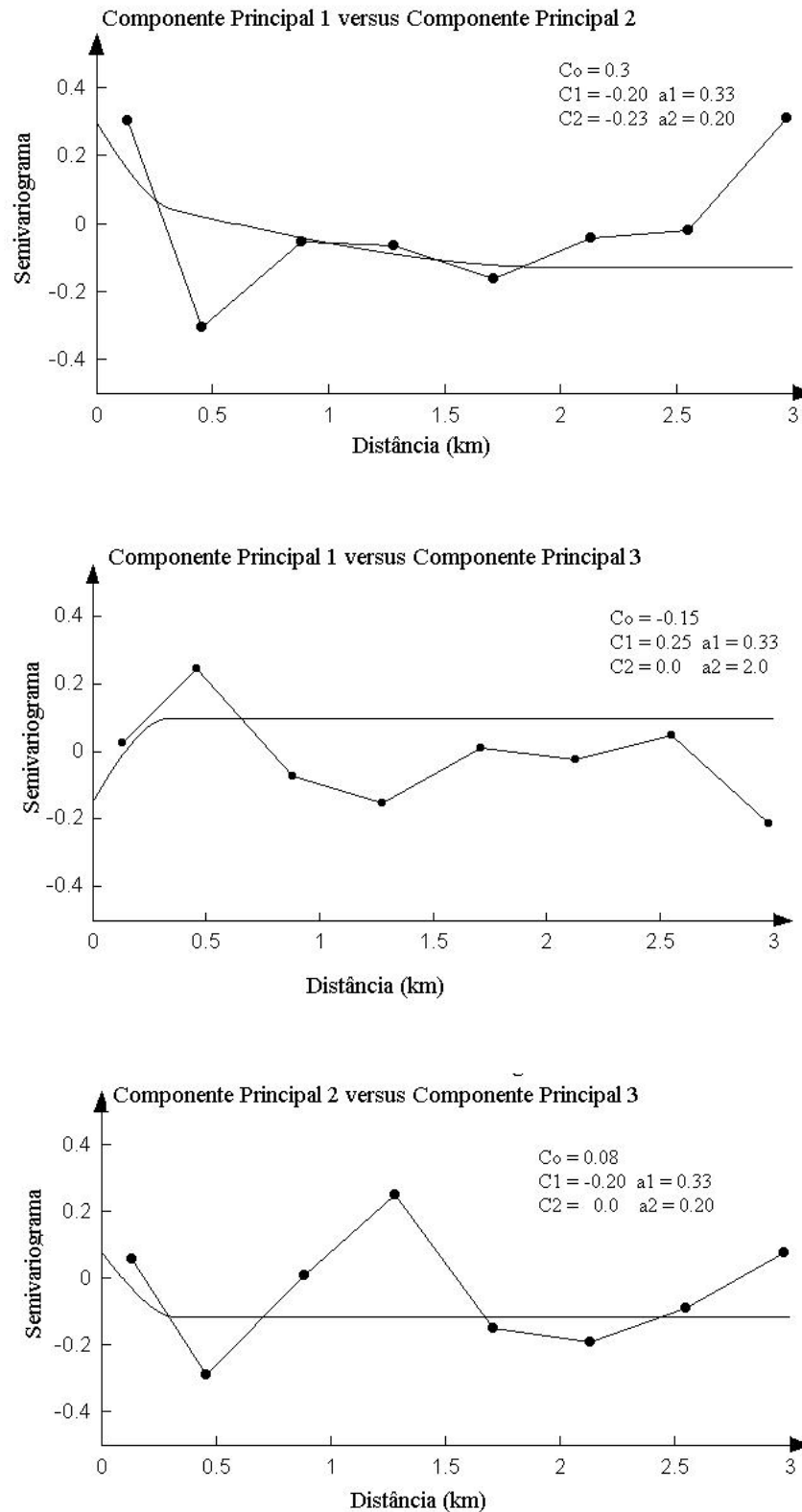


Figura 5.25 – Semivariogramas experimentais cruzados das três componentes principais (linhas ligadas por pontos) e os semivariogramas ajustados pelo modelo linear de co-regionalização (linhas contínuas)

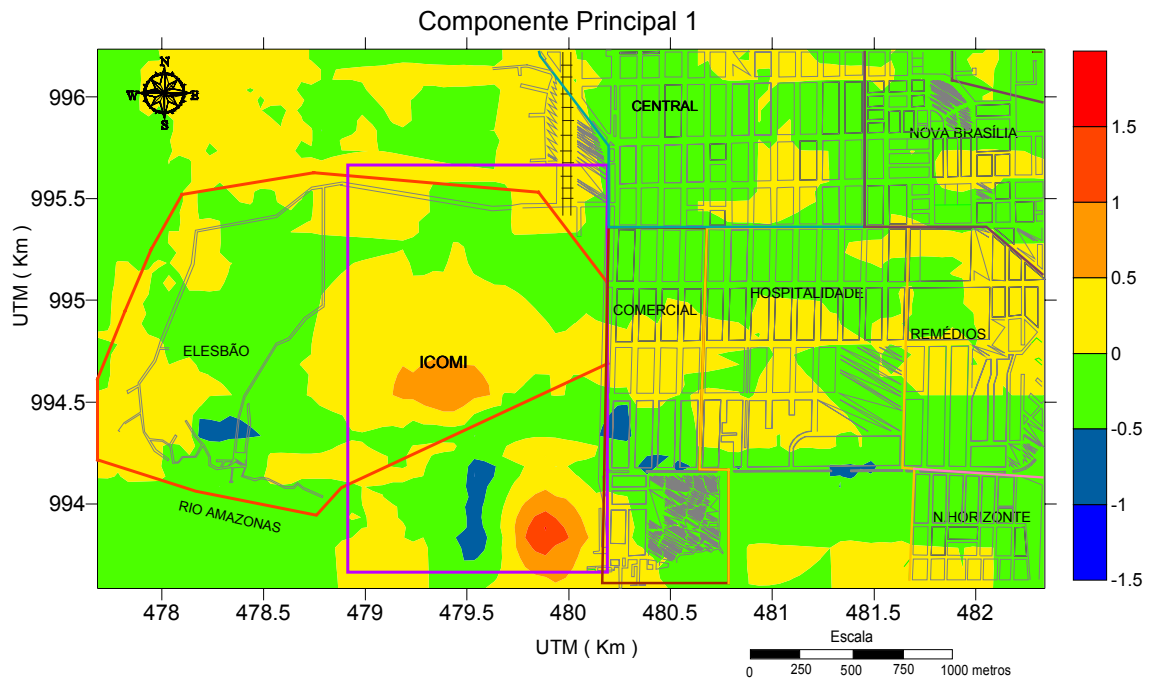


Figura 5.26 – Mapa de contornos da primeira componente principal.

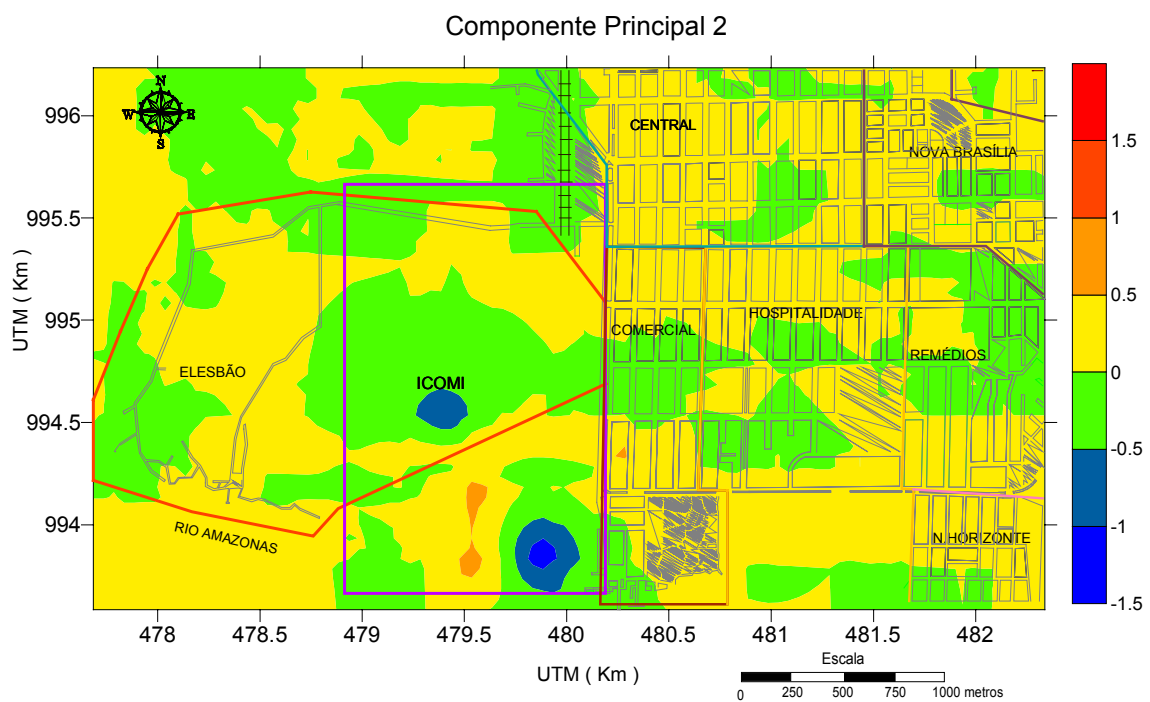


Figura 5.27 – Mapa de contornos da segunda componente principal.

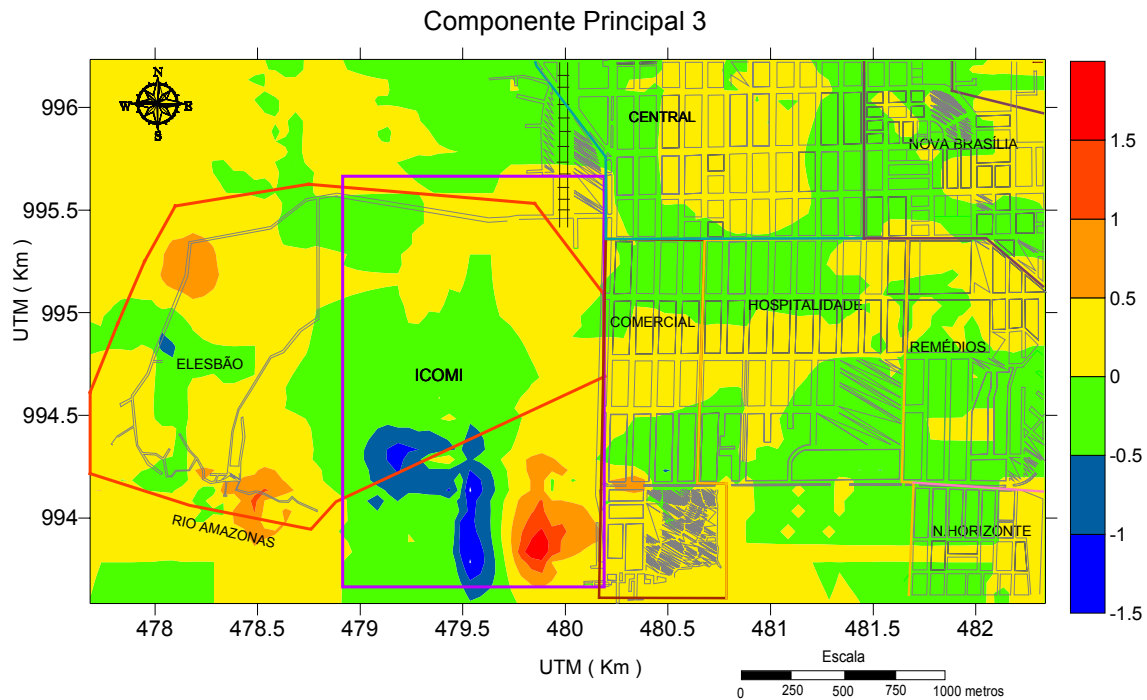


Figura 5.28 – Mapa de contornos da terceira componente principal.

Na aplicação da krigagem fatorial multivariada, antes da análise das relações entre as variáveis nas diferentes escalas espaciais, foi feita uma padronização nas mesmas por causa das diferenças consideráveis observadas na amplitude de cada uma (Tabela 5.2). Um modelo linear de co-regionalização foi ajustado a todos os semivariogramas diretos e cruzados das variáveis. Todos os 36 semivariogramas diretos e cruzados foram calculados e modelados como a soma de um efeito pepita e dois modelos esféricos com alcances 0,33 km e 2,0 km, escolhidos a partir dos modelos semivariográficos ajustados para as 3 primeiras componentes principais, que explicam a maior parte da variância. Os valores dos patamares (sill) dos três modelos, ou seja, os coeficientes b_{ij}^s , foram colocados em três matrizes de co-regionalização \mathbf{B}^0 , \mathbf{B}^1 e \mathbf{B}^2 . A Figura 5.29 mostra o modelo linear de co-regionalização ajustado aos semivariogramas experimentais diretos. Os semivariogramas cruzados com o respectivo modelo linear de co-regionalização ajustado são apresentados em anexo.

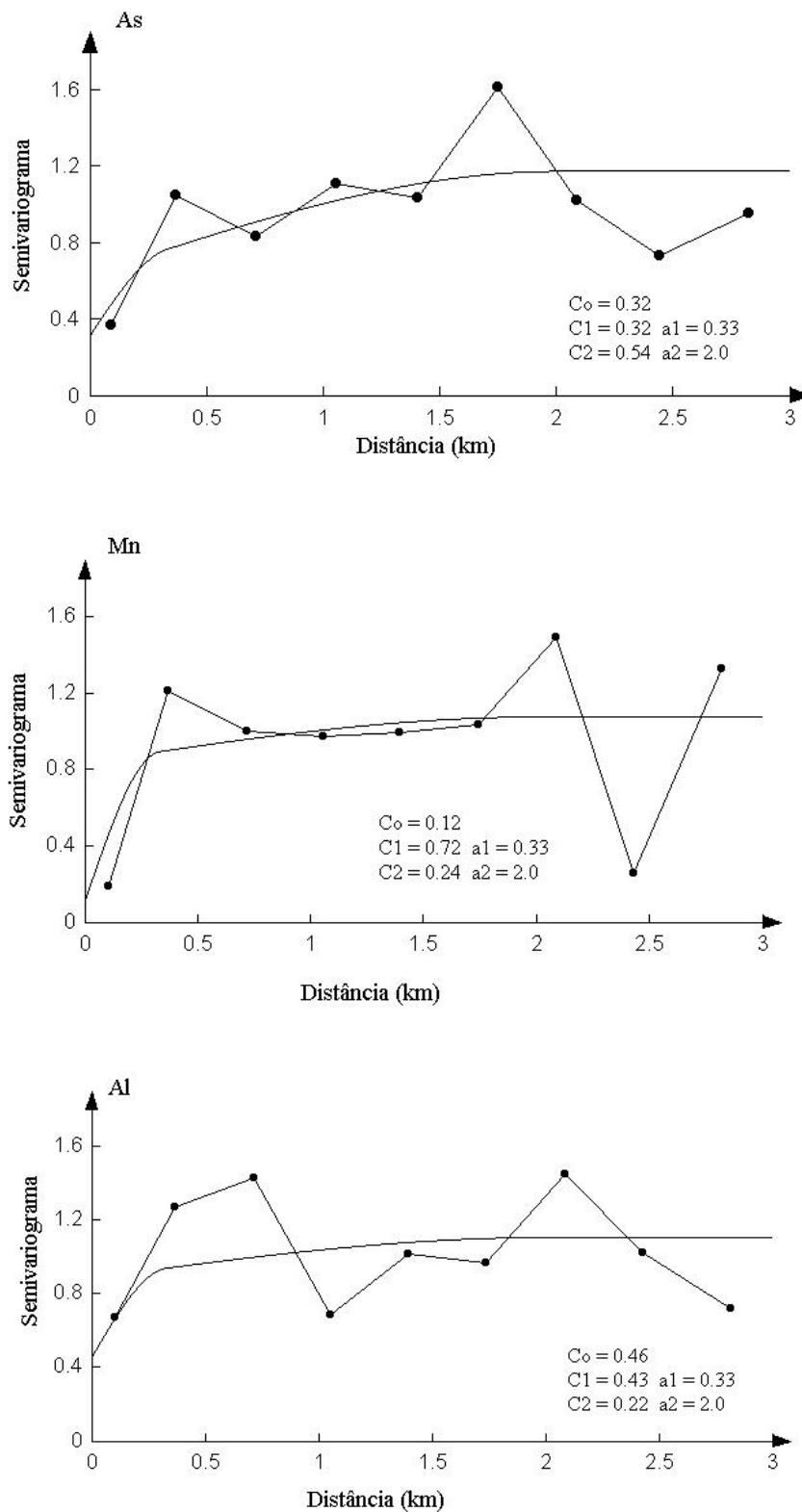


Figura 5.29 – Semivariogramas experimentais do arsênio, manganês e alumínio representados pelas linhas ligadas por pontos. As linhas contínuas são os semivariogramas ajustados pelo modelo linear de co-regionalização.

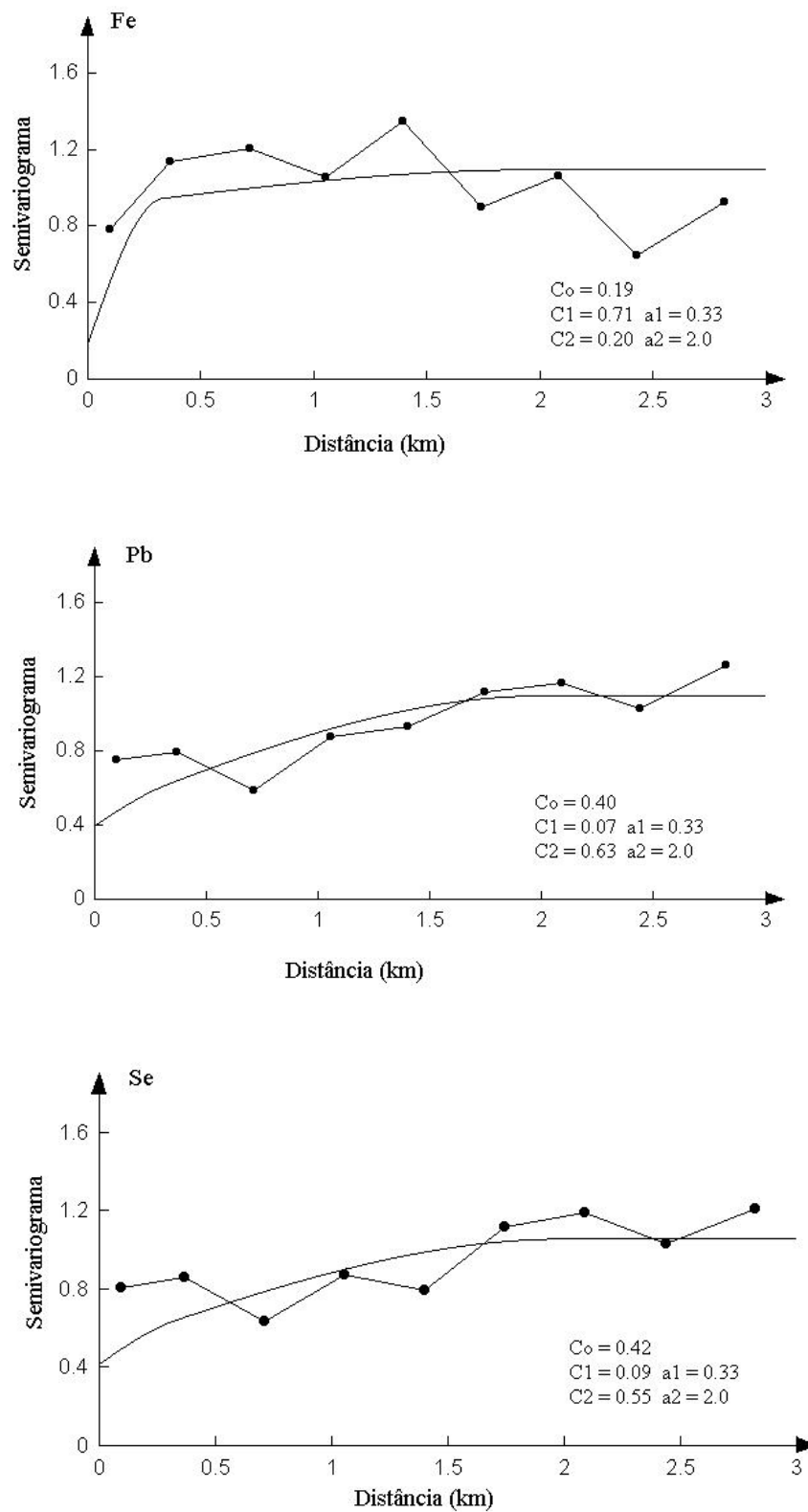


Figura 5.29 – (Cont.) Semivariogramas experimentais do ferro, chumbo e selênio representados pelas linhas ligadas por pontos. As linhas contínuas são os semivariogramas ajustados pelo modelo linear de co-regionalização.

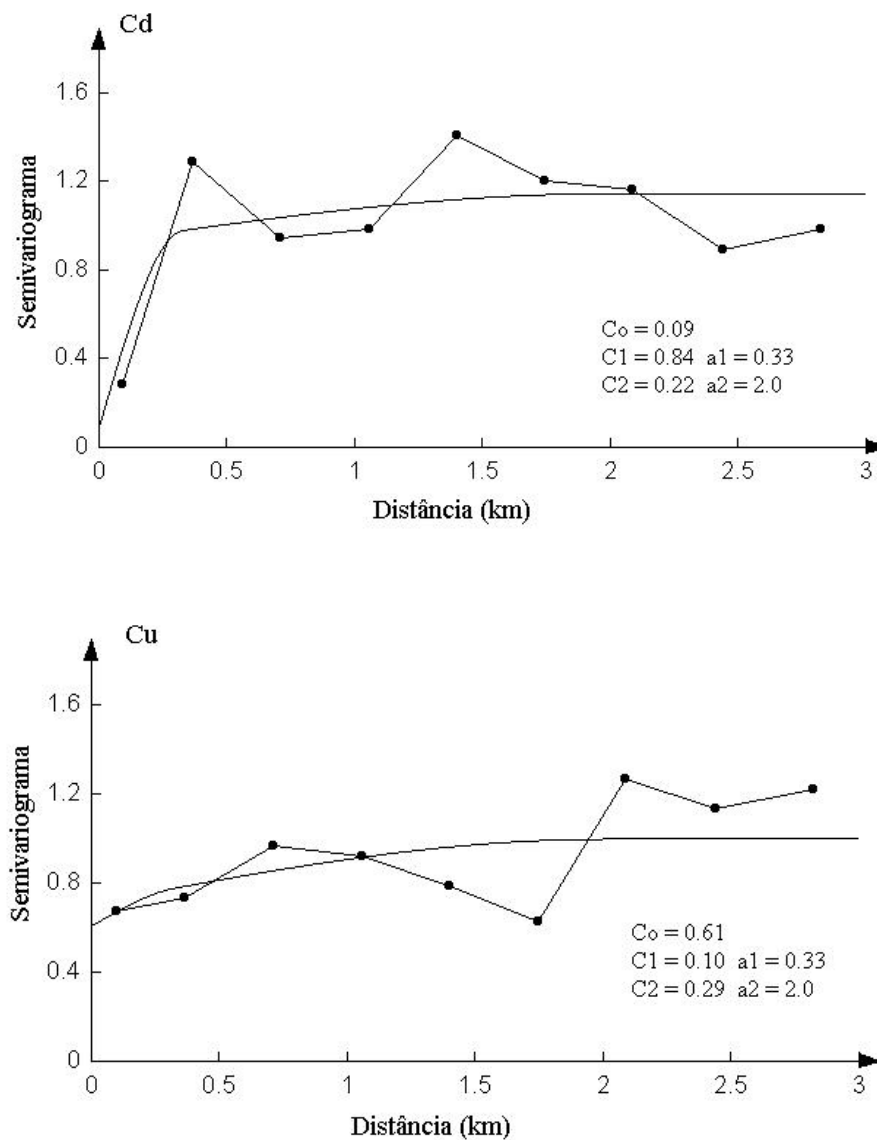


Figura 5.29 – (Cont.) Semivariogramas experimentais do cádmio e cobre representados pelas linhas ligadas por pontos. As linhas contínuas são os semivariogramas ajustados pelo modelo linear de co-regionalização.

Nos semivariogramas diretos, a importância relativa dos dois modelos esféricos muda de acordo com as escalas consideradas. A estrutura de longo alcance domina nos semivariogramas do chumbo, selênio e cobre enquanto a estrutura de curto alcance prevalece nos semivariogramas do manganês, ferro e cádmio. Nos semivariogramas do arsênio e alumínio ocorrem os tipos de estruturas sem forte predominância de uma sobre a outra. As contribuições relativas das duas estruturas nos semivariogramas diretos refletem o papel representado por atividades humanas ou fatores vinculados à longa escala como a geologia, no padrão espacial mostrado por cada variável. Os

semivariogramas cruzados ilustram como a importância relativa dos dois modelos esféricos muda de acordo com a correlação espacial entre determinadas variáveis.

As matrizes de co-regionalização \mathbf{B}^1 e \mathbf{B}^2 descrevem a estrutura de correlação das variáveis para pequena escala (0,33 km) e longa escala (2,0 km), respectivamente. Os resultados da análise de componentes principais dessas duas matrizes de co-regionalização fornecem os fatores regionalizados, $Y_k^s(\mathbf{u})$, que são as componentes principais de cada matriz na respectiva escala “s”, e as componentes espaciais, $Z_i^s(\mathbf{u})$. Analogamente à expressão (5.3), o coeficiente de correlação entre a componente espacial $Z_i^s(\mathbf{u})$ e o fator regionalizado $Y_k^s(\mathbf{u})$ é dado pela equação (3.68) ou seja:

$$\rho_{ik}^s = \frac{e_{ki}^s \sqrt{\lambda_i^s}}{\sqrt{b_{ii}^s}}, \quad s = 0, 1, 2 \quad i, k = 1, \dots, 8. \quad (5.4)$$

onde e_{ki}^s e λ_i^s são, respectivamente, autovetores e autovalores da matriz de co-regionalização \mathbf{B}^s ; e b_{ii}^s é o patamar (sill) da estrutura básica $\gamma_s(\mathbf{h})$ do modelo semivariográfico direto da variável Z_i . Os resultados da análise de componentes principais das duas matrizes \mathbf{B}^1 e \mathbf{B}^2 são apresentados na Tabela 5.6 e Figura 5.30.

Os três primeiros fatores regionalizados explicam 86,2 % e 99,5 % da variância total para as matrizes \mathbf{B}^1 e \mathbf{B}^2 , respectivamente. Para a matriz \mathbf{B}^1 (pequena escala espacial) observa-se forte correlação do ferro e cádmio com o primeiro fator regionalizado. O manganês é fortemente correlacionado com o segundo fator regionalizado e a contribuição do arsênio predomina fortemente em relação às outras variáveis no terceiro fator regionalizado. Isso sugere uma possível relação da contaminação desses elementos a fontes de contaminação vinculada a atividades humanas que, em geral, ocorrem em pequena escala espacial, neste caso, descrita pela matriz \mathbf{B}^1 . A matriz \mathbf{B}^1 não mostra correlação entre o arsênio e o manganês na escala considerada. Para a matriz \mathbf{B}^2 (longa escala espacial) o primeiro fator regionalizado explica a maior parte da variabilidade sendo que a contribuição de todas as variáveis, exceto o cobre, é relativamente alta com destaque para o chumbo e selênio. O cobre é mais fortemente correlacionado com o segundo fator e nenhuma variável apresenta contribuição significativa para o terceiro fator, que explica somente 3,2 % da variabilidade total. O alumínio apresenta correlação mais forte com o primeiro fator regionalizado na longa escala.

TABELA 5.6 – Cargas dos três primeiros Fatores Regionalizados com a percentagem da variância explicada na pequena (\mathbf{B}^1) e longa (\mathbf{B}^2) escala espacial.

Variáveis	Cargas dos Fatores Regionalizados – Matriz \mathbf{B}^1			Cargas dos Fatores Regionalizados – Matriz \mathbf{B}^2		
	Y_1^1	Y_2^1	Y_3^1	Y_1^2	Y_2^2	Y_3^2
As	0.000	-0.003	0.517	0.663	0.098	-0.116
Mn	0.001	0.817	0.006	0.447	-0.044	-0.144
Al	0.316	-0.007	-0.011	0.438	0.097	0.025
Fe	0.803	-0.002	0.002	0.417	0.077	0.038
Pb	0.223	0.005	-0.003	0.735	-0.147	0.092
Se	0.219	0.010	-0.009	0.685	-0.147	0.140
Cd	0.836	-0.001	0.005	0.415	0.078	-0.116
Cu	-0.006	-0.104	0.033	0.097	0.520	0.092
Autovalores	1.732	0.734	0.368	2.439	0.354	0.094
% da Variância	52,7	22.3	11.2	84.1	12.2	3.2
Acumulado	52.7	75.0	86.2	84.1	96.3	99.5

Na figura 5.30 pode-se observar a mudança da estrutura de correlação de uma escala espacial para outra. A análise das matrizes de co-regionalização permite a observação de relações entre as variáveis de acordo com a escala espacial que não poderiam ser detectadas na análise convencional da matriz \mathbf{R} de correlações (ou a matriz \mathbf{V} de variância-covariância). Por exemplo, a correlação entre ferro e cádmio é mais forte na pequena escala enquanto que selênio e chumbo apresentam maior correlação na longa escala. Os valores das cargas do primeiro fator regionalizado das matrizes \mathbf{B}^1 e \mathbf{B}^2 mostram que é mais provável que ocorra alguma correlação entre arsênio e manganês na longa do que na pequena escala espacial.

Uma comparação entre a Figura 5.23 e a Figura 5.30 pode mostrar que a estrutura de correlação descrita pela matriz \mathbf{R} (ou \mathbf{V}) é uma mistura de estruturas de correlações encontradas em diferentes escalas espaciais.

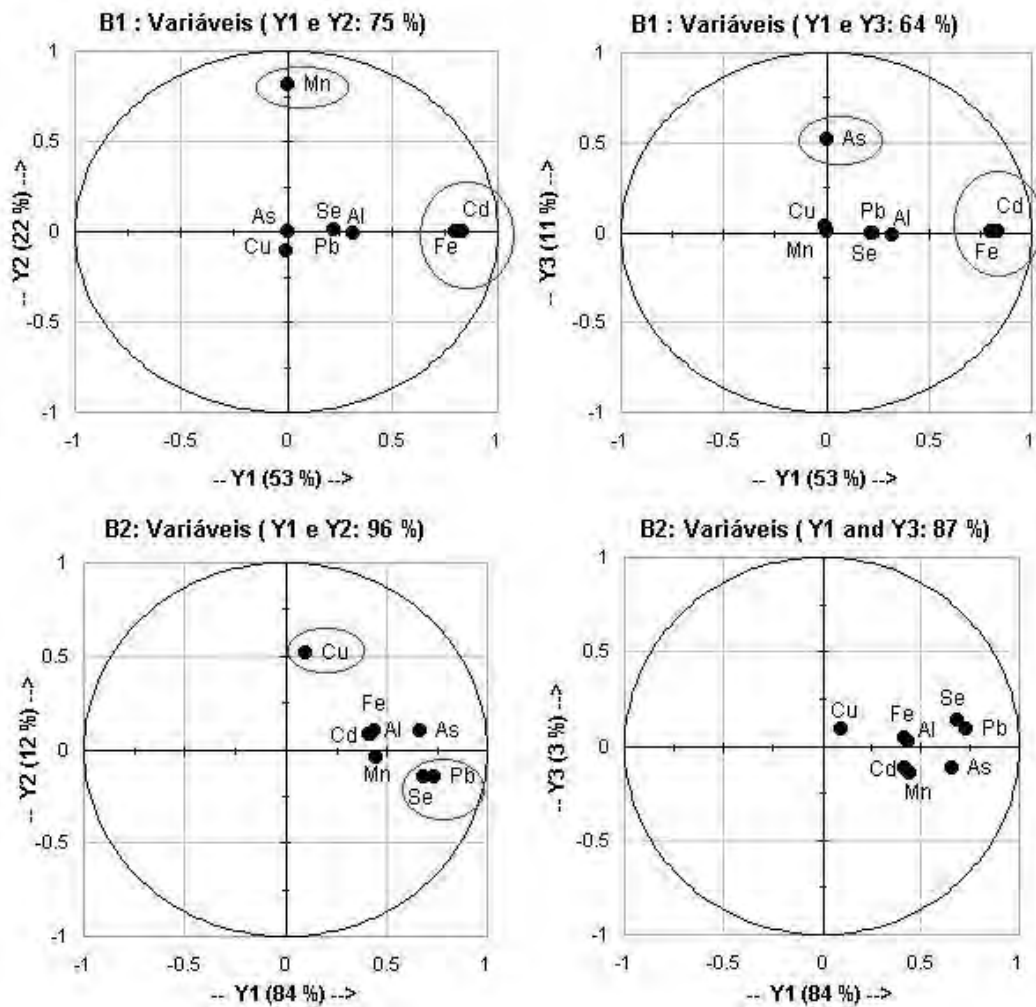


Figura 5.30 – Gráficos das correlações entre as componentes espaciais $Z_i^s(\mathbf{u})$ e os fatores regionalizados $Y_k^s(\mathbf{u})$ obtidos para as matrizes de co-regionalização \mathbf{B}^1 e \mathbf{B}^2 associadas com modelos esféricos de pequeno (0,33 km) e longo (2,0 km) alcance. Os eixos numerados são para o primeiro, segundo e terceiro fator regionalizado, respectivamente. Os números entre parênteses representam a percentagem da variância total explicada por cada fator.

As Figuras 5.31 a 5.36 mostram os mapas cokrigados dos três primeiros fatores regionalizados, $Y_k^s(\mathbf{u})$ $s = 1, 2$ e $k = 1, \dots, 8$, associados às escalas espaciais consideradas. O valor de qualquer fator regionalizado é igual a sua média local estabelecida como zero para qualquer distância acima de 0,33 km, no caso da pequena escala (Figuras. 5.31 a 5.33), e acima de 2,0 km no caso da longa escala (Figuras. 5.34 a 5.36), em relação aos pontos amostrais (ver seção teórica). Para a escala local (pequeno alcance), a presença de áreas que podem ser consideradas anômalas – com valores mais altos (ou baixos) – é observada principalmente no primeiro e segundo fatores regionalizados que,

juntos, explicam 75% da variação total. Essas regiões localizam-se quase totalmente dentro dos limites da ICOMI. A drenagem da área de estudo (em linha azul) foi plotada nos mapas relacionados à longa escala (Figuras 5.34 a 5.36). Pode-se observar uma boa concordância das linhas de drenagem com o padrão espacial apresentado, sobretudo pelo mapa do primeiro fator regionalizado associado com o modelo esférico de longa escala, que explica a maior parte da variabilidade (84,1%) com contribuição razoável de quase todas as variáveis, exceto o cobre.

Foi realizado o mapeamento do As, Mn, Fe e Cd, que mostraram indícios de estarem vinculados a uma escala local. As estimativas das componentes espaciais, $Z_i^s(\mathbf{u})$, $s = 1,2$ e $i = 1,\dots,8$, mostradas nas Figuras 5.37 a 5.44, foram calculadas por cokrigagem para a pequena ($s = 1$) e longa escala ($s = 2$). As linhas de drenagem são mostradas nos mapas relacionados à longa escala. Pode-se observar uma certa similaridade entre o padrão espacial apresentado pelos mapas de todos os elementos (As, Mn, Fe e Cd) com a drenagem da área de estudo. O mapa do arsênio na pequena escala apresenta duas pequenas áreas com anomalias negativas, uma dentro da área industrial da ICOMI e outra no bairro do Elesbão à margem do Rio Amazonas (Figura 5.37). No mapa desse mesmo elemento associado à longa escala observa-se regiões com altos valores positivos dentro da área da ICOMI, no bairro do Elesbão – à noroeste – e na área residencial, no bairro da hospitalidade, parecendo indicar uma disseminação do referido elemento na área de estudo (Figura 5.38). Os mapas relativos ao manganês mostram mais destaque desse elemento dentro da área da ICOMI, sendo observados altos valores positivos e negativos no mapa associado à pequena escala (Figura 5.39) e valores positivos no mapa associado à longa escala (Figura 5.40). Os mapas do ferro e cádmio são bastante similares, tanto na curta como na longa escala, provavelmente pela forte correlação entre esses elementos, apresentando altos valores positivos e negativos espalhados pela área de estudo, sendo preferencialmente localizados dentro da área da ICOMI, no caso da pequena escala, e valores positivos restritos à área da ICOMI em relação à longa escala (Figuras 5.41 a 5.44). A ocorrência de valores altos para os elementos analisados na área de estudo pode indicar a presença de fontes de contaminação possivelmente vinculadas às atividades industriais desenvolvidas pela ICOMI durante os últimos anos.

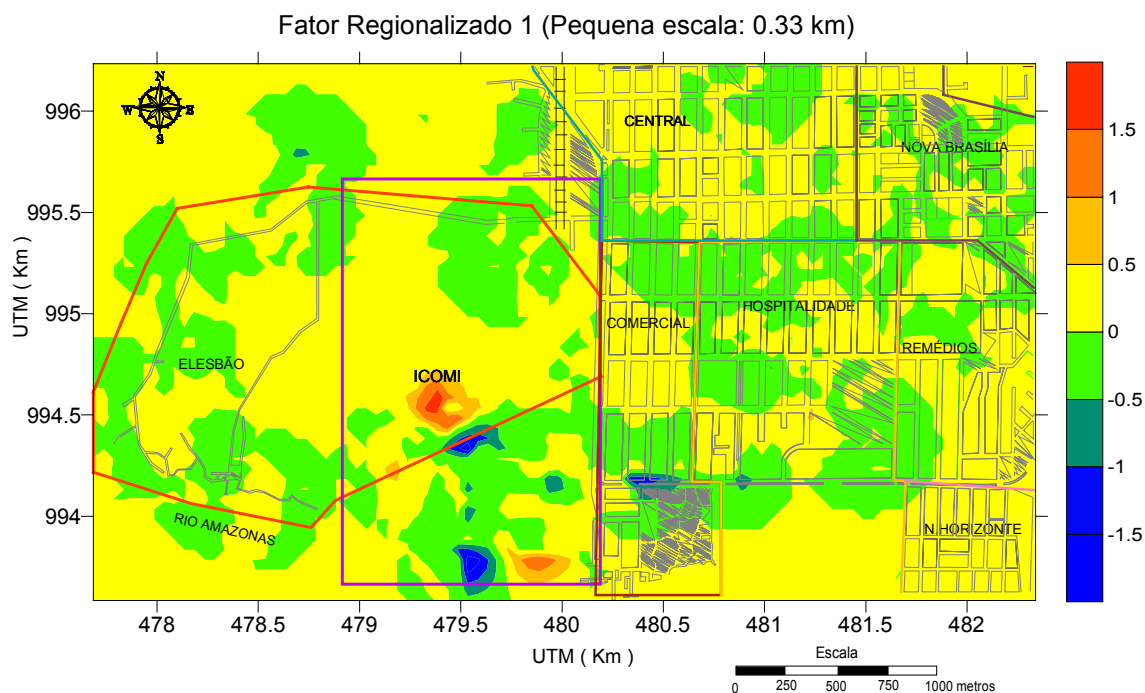


Figura 5.31 – Mapa de contorno do primeiro fator regionalizado associado com o modelo esférico em curta escala.

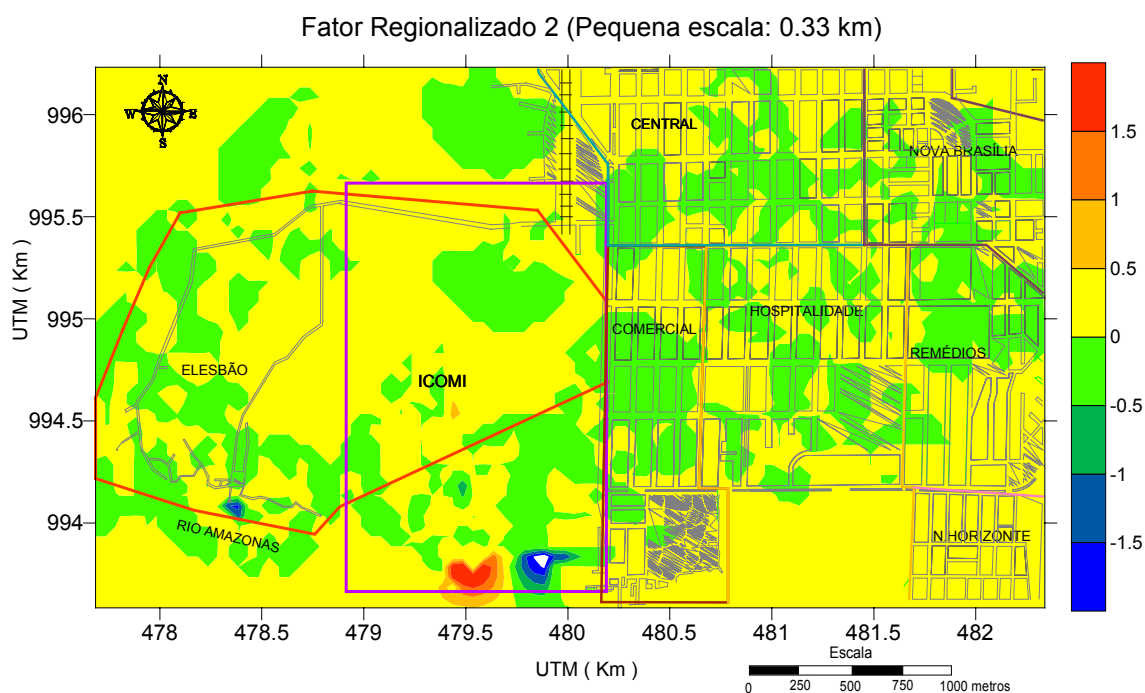


Figura 5.32 – Mapa de contorno do segundo fator regionalizado associado com o modelo esférico em curta escala.

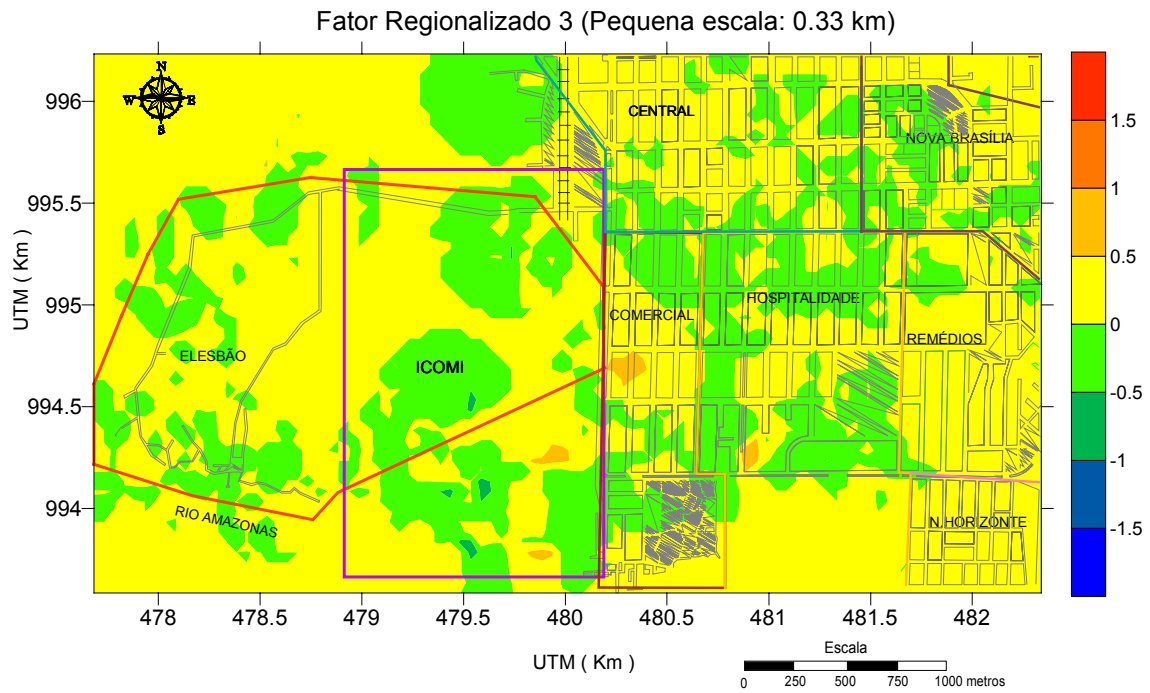


Figura 5.33 – Mapa de contorno do terceiro fator regionalizado associado com o modelo esférico em curta escala.

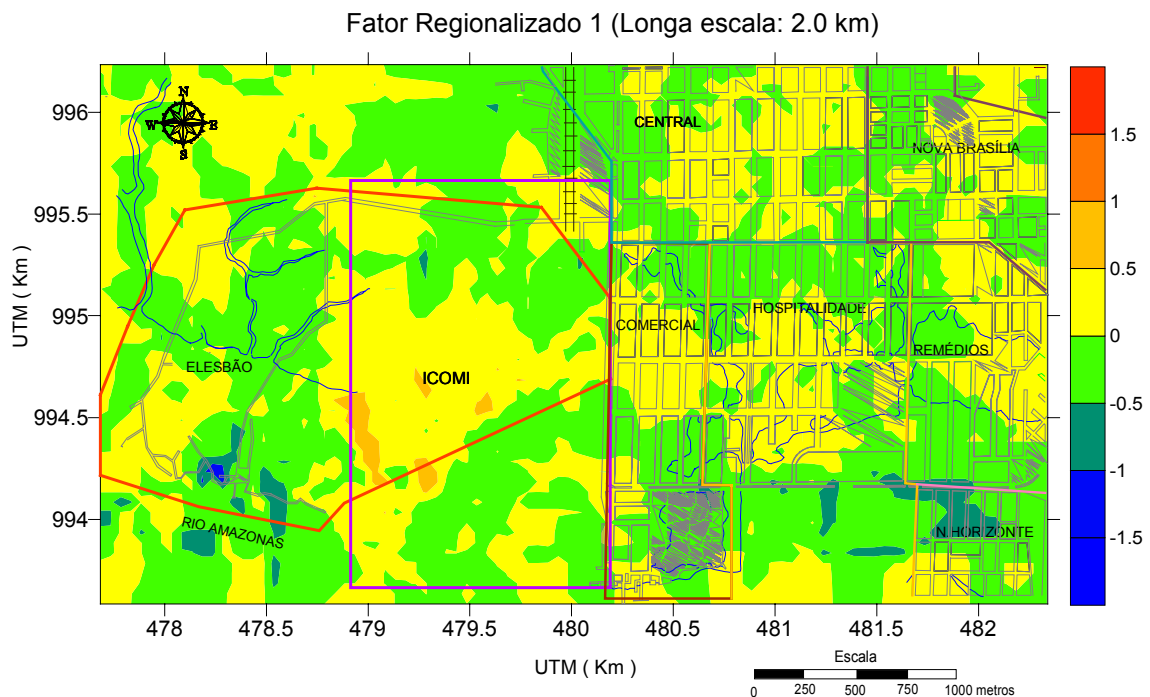


Figura 5.34 – Mapa de contorno do primeiro fator regionalizado associado com o modelo esférico em longa escala. As linhas em azul referem-se à drenagem da área.

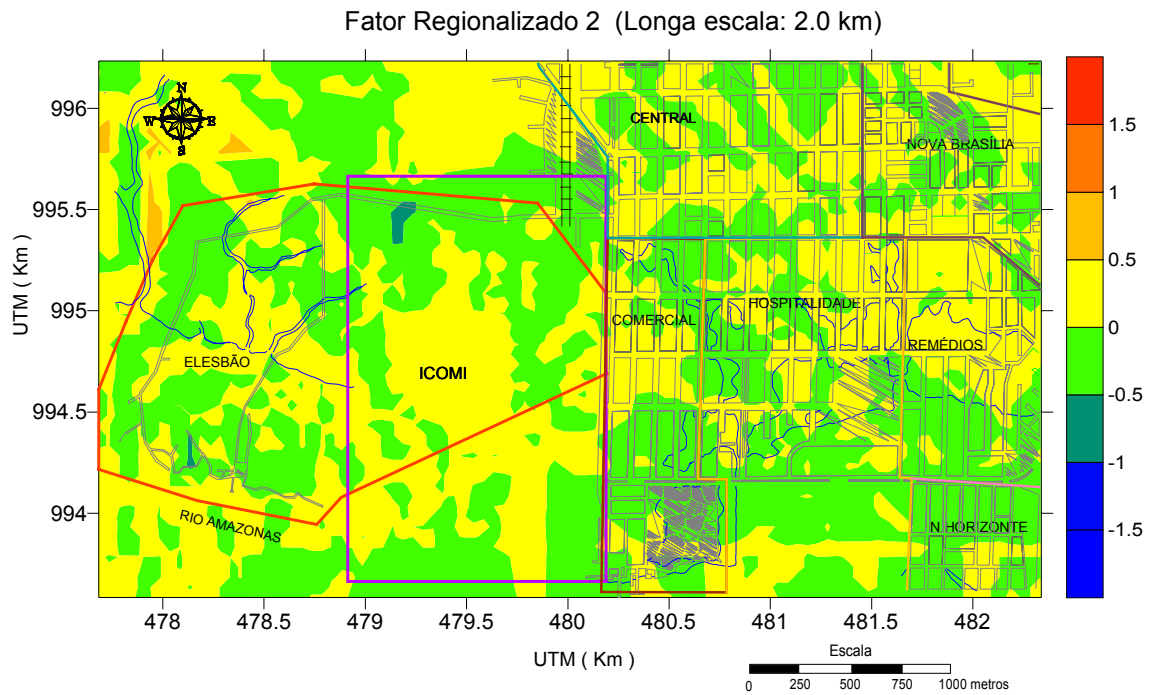


Figura 5.35 – Mapa de contorno do segundo fator regionalizado associado com o modelo esférico em longa escala. As linhas em azul referem-se à drenagem da área.

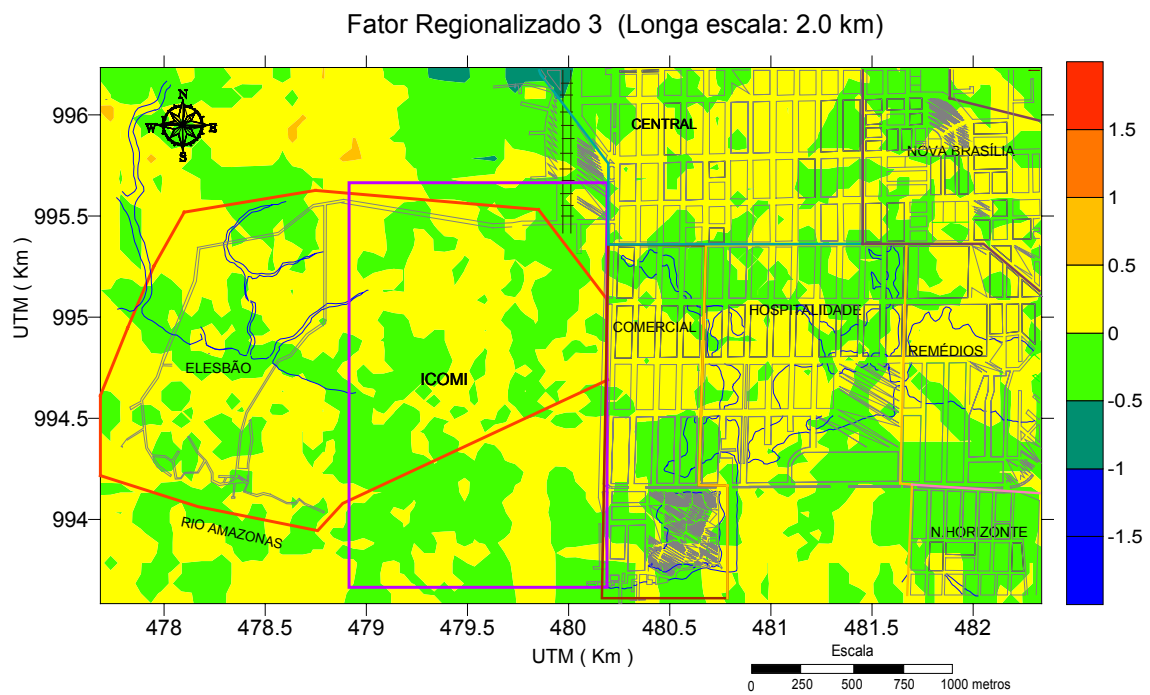


Figura 5.36 – Mapa de contorno do terceiro fator regionalizado associado com o modelo esférico em longa escala. As linhas em azul referem-se à drenagem da área.

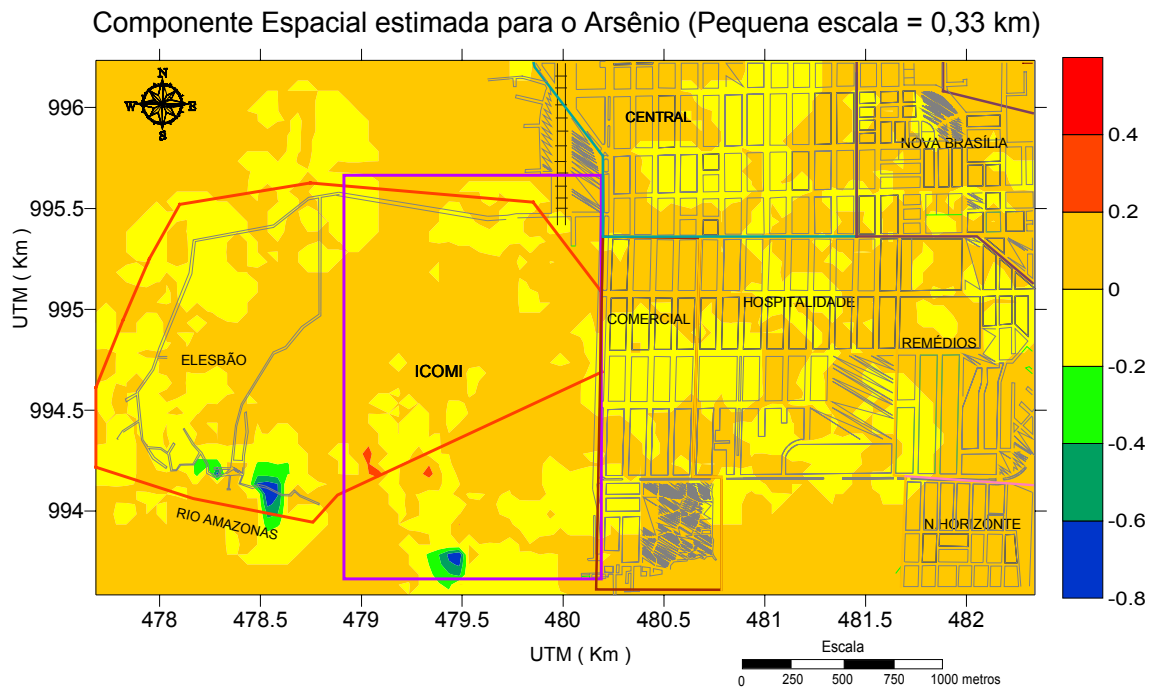


Figura 5.37 – Mapa de contorno da componente espacial estimada para o arsênio associada com o modelo esférico em curta escala.

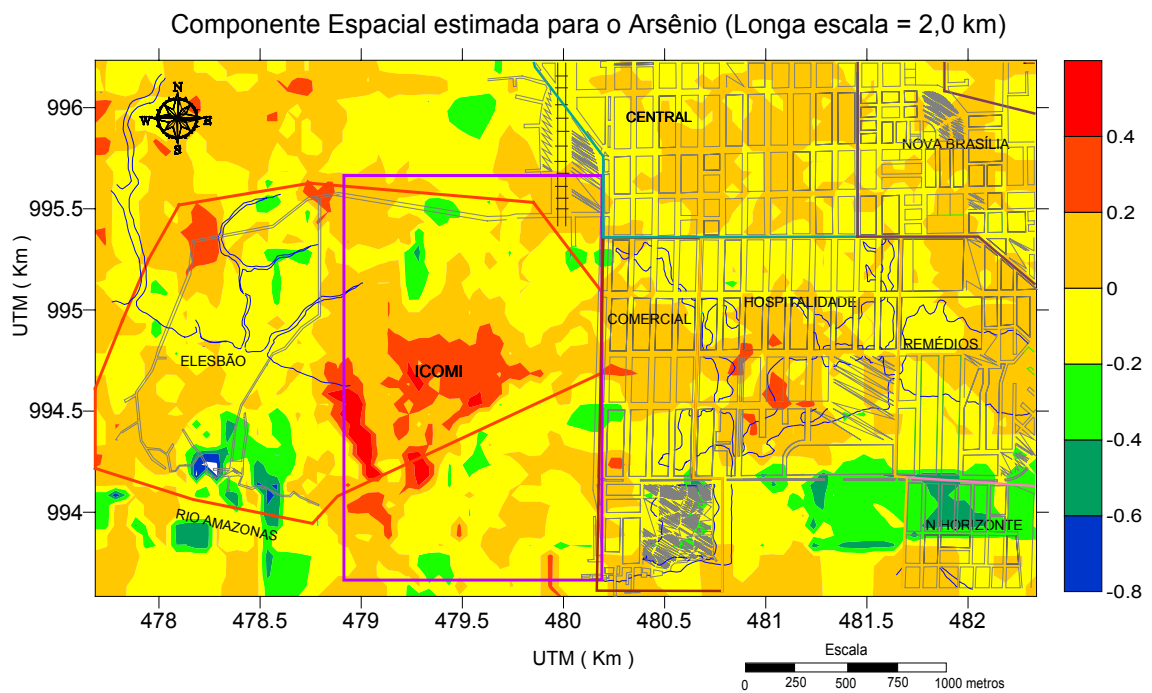


Figura 5.38 – Mapa de contorno da componente espacial estimada para o arsênio associada com o modelo esférico em longa escala. As linhas em azul referem-se à drenagem da área.

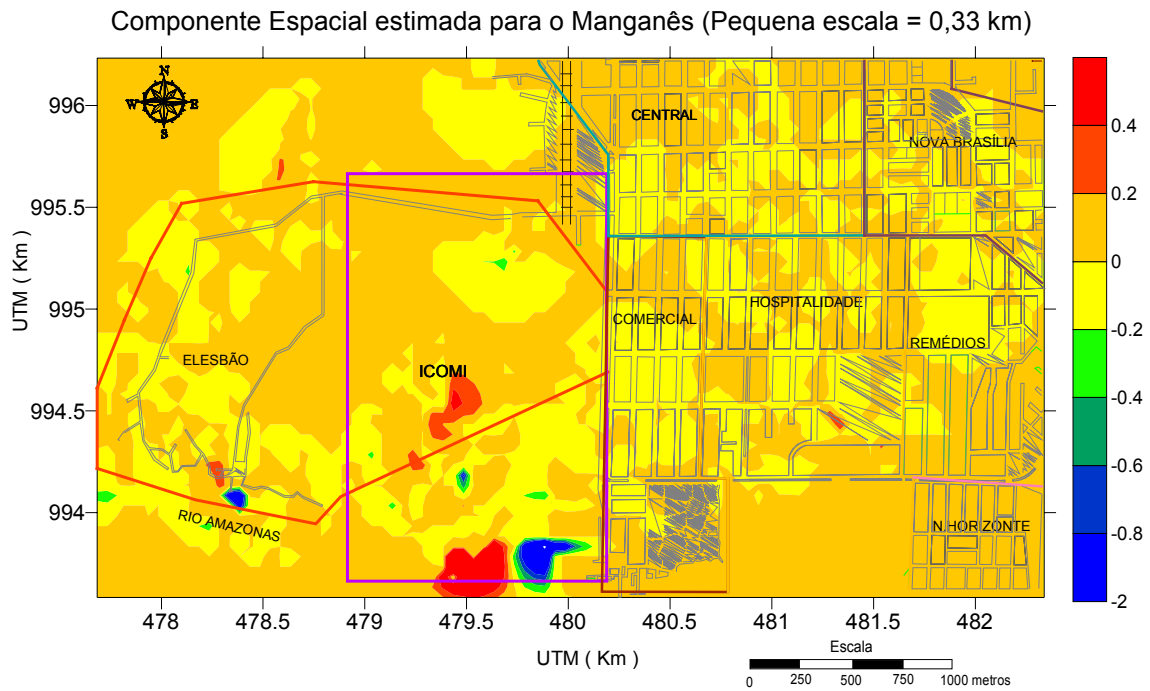


Figura 5.39 – Mapa de contorno da componente espacial estimada para o manganês associada com o modelo esférico em curta escala.

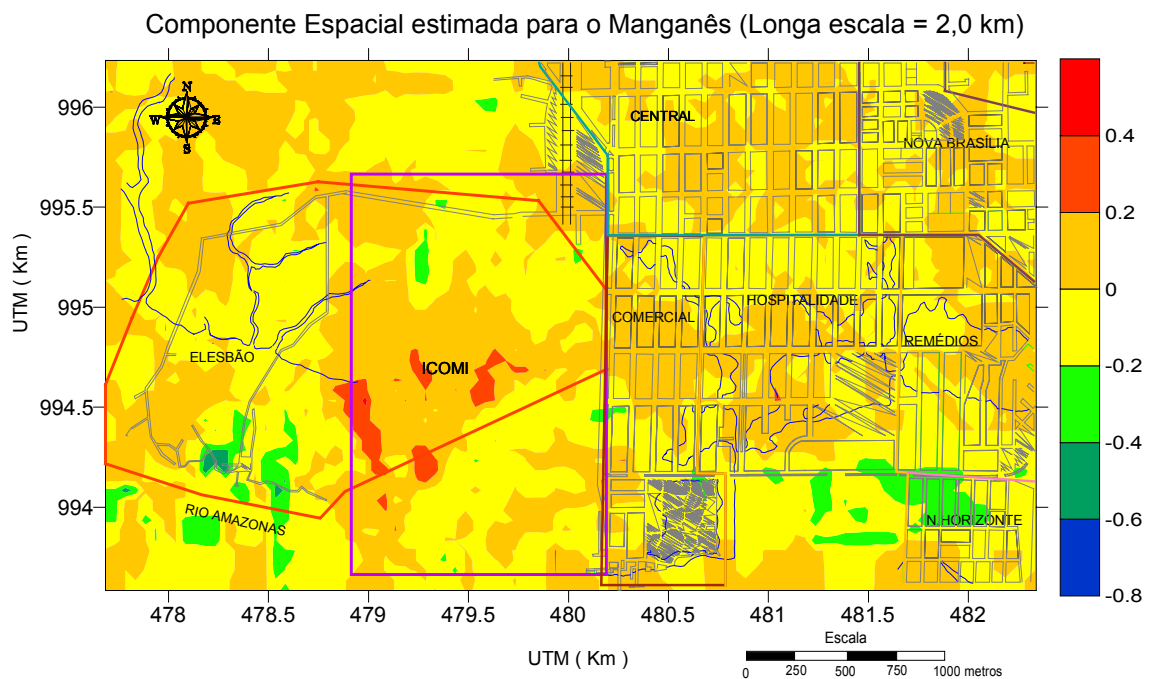


Figura 5.40 – Mapa de contorno da componente espacial estimada para o manganês associada com o modelo esférico em longa escala. As linhas em azul referem-se à drenagem da área.

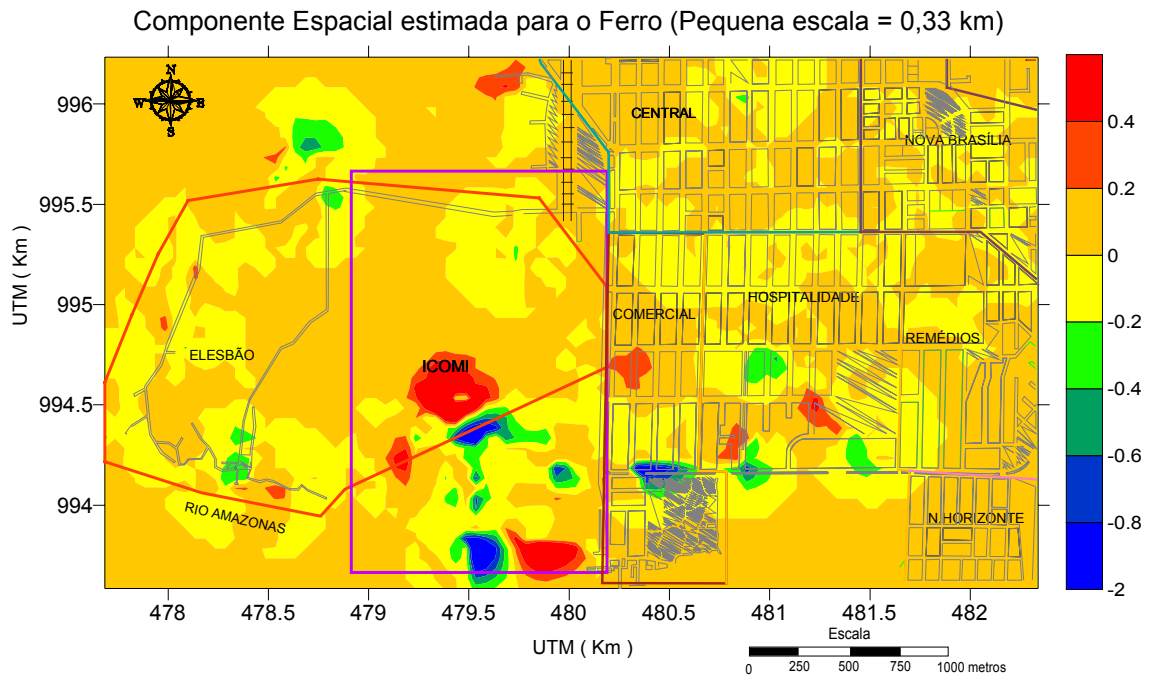


Figura 4.41 – Mapa de contorno da componente espacial estimada para o ferro associada com o modelo esférico em curta escala.

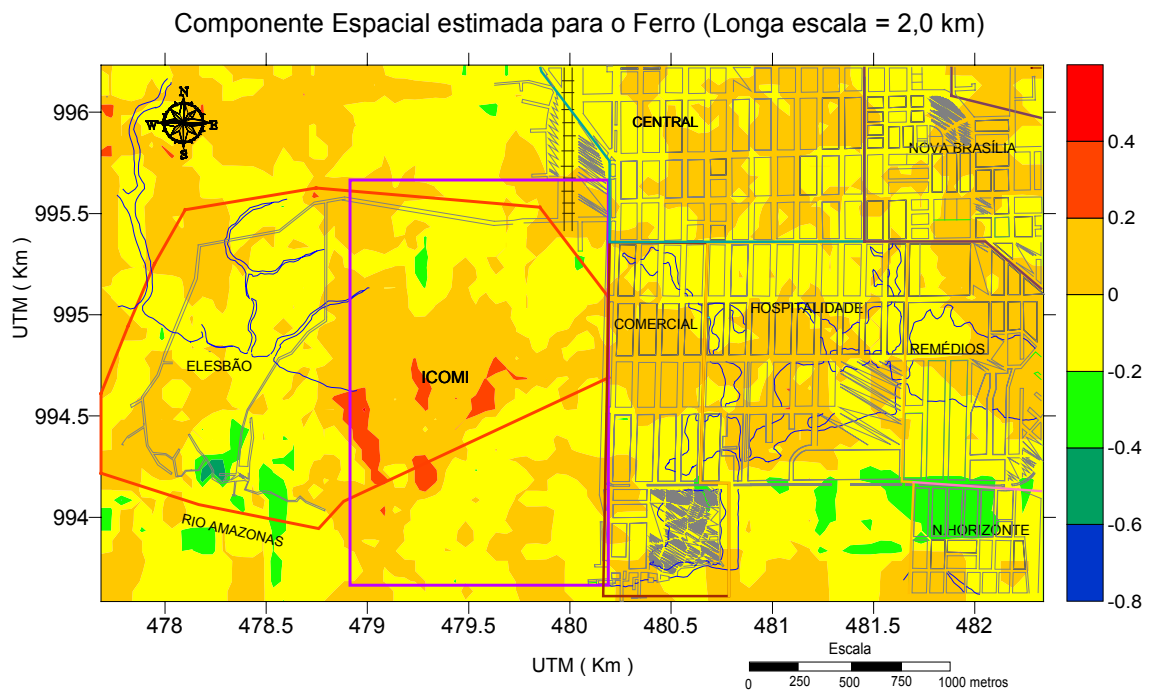


Figura 5.42 – Mapa de contorno da componente espacial estimada para o ferro associada com o modelo esférico em longa escala. As linhas em azul referem-se à drenagem da área.

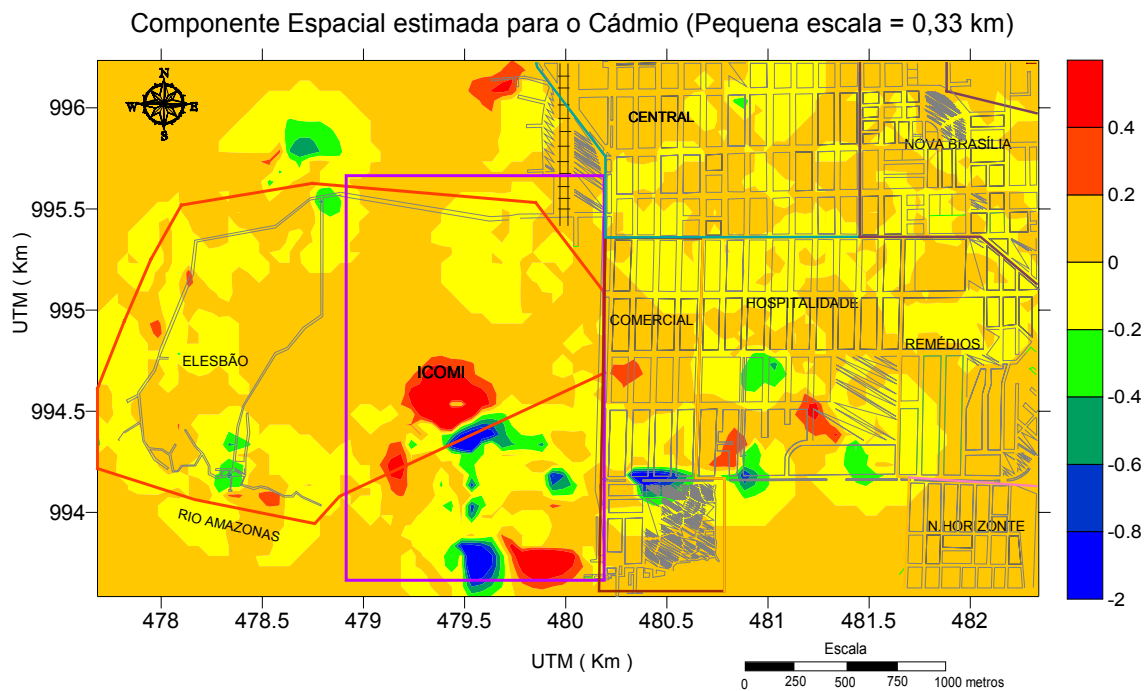


Figura 5.43 – Mapa de contorno da componente espacial estimada para o cádmio associada com o modelo esférico em curta escala.

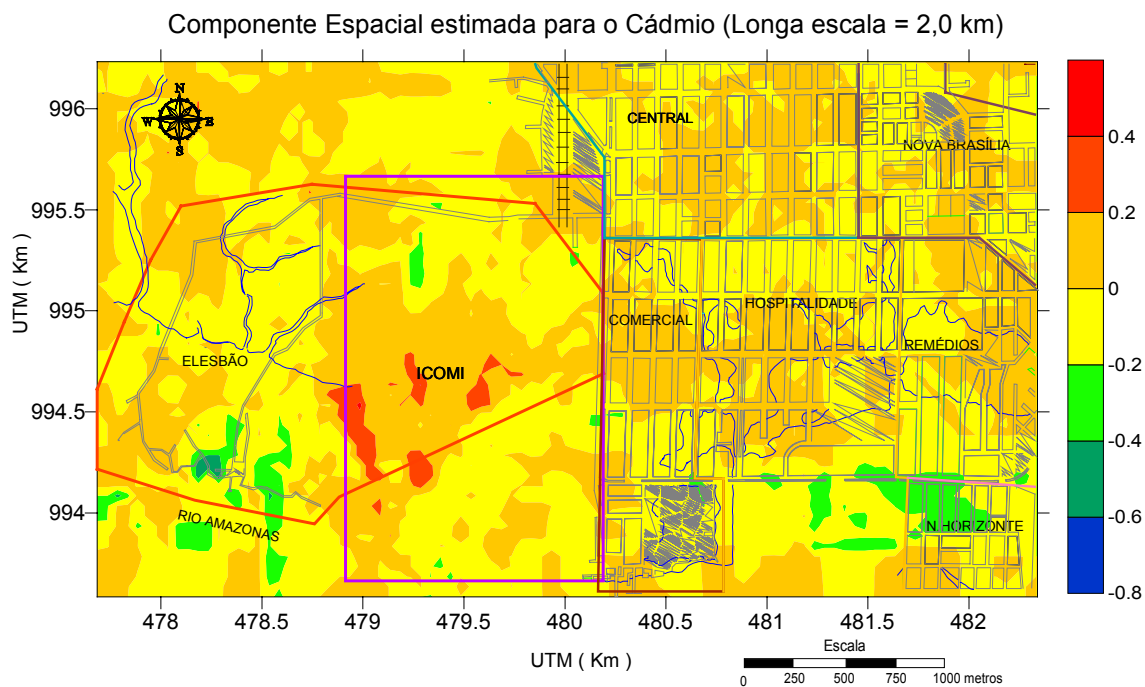


Figura 5.44 – Mapa de contorno da componente espacial estimada para o cádmio associada com o modelo esférico em longa escala. As linhas em azul referem-se à drenagem da área.

Os resultados obtidos pela krigagem fatorial multivariada mostraram que o arsênio, manganês, ferro e cádmio provavelmente estão vinculados a fontes de contaminações localizadas (pequena escala), ou seja, provenientes de atividades humanas, enquanto alumínio, selênio, chumbo e cobre apresentaram indícios de que suas concentrações são provenientes do próprio meio natural, por causa da correlação mais forte com variações de longa escala desses elementos.

5.2.4 Modelos de incerteza locais para o risco de contaminação

Uma localização não amostrada pode ser classificada com base na estimativa de seu verdadeiro valor ou, alternativamente, na probabilidade de que o ponto esteja abaixo (ou acima) de um determinado valor de corte. Estimativas envolvem erros. Erros estão sujeitos a custos. Se um ponto é incorretamente classificado como alto quando, na realidade, é baixo, há um custo de super-classificação (ou superestimativa). Ao contrário, se o ponto é incorretamente classificado como baixo quando, na verdade, é alto, há um custo de sub-classificação (ou subestimativa). No caso de contaminação ambiental, a superestimativa da extensão da contaminação aumentará o custo de remediação, mas a subestimativa da contaminação poderá resultar em perdas prematuras de vidas, ou alto custo de saúde. Em tal situação, o custo de subestimativa pode ser muito maior do que o custo de superestimativa. A unidade de custo de superestimativa e subestimativa depende de cada situação.

As funções de perda do tipo (3.82) e (3.83) foram utilizadas para a avaliação do impacto econômico para a tomada de decisões. Os pesos w_1 e w_2 , relacionados a custos de saúde e remediação, respectivamente, devem ser considerados relativos, uma vez que na prática esses valores variam bastante, dependendo, por exemplo, da alternativa tecnológica de atenuação, da empresa prestadora do serviço, preço de mercado. No caso em questão, a opção sugerida acerca da contaminação por arsênio e manganês na área industrial da ICOMI em Santana/AP, foi pela implantação de um aterro controlado, provavelmente, fora da área da ICOMI. (RELATÓRIO JAAKO POYRY, 1998, p. 38).

A perda esperada associada aos dois tipos de classificações modeladas pelas funções (3.82) e (3.83) foram computadas utilizando-se uma adaptação do programa *postik* da livraria GSLIB (em anexo) onde foram incorporadas as referidas funções com possibilidade de modificações para a inclusão de outros tipos de funções de perda.

Para a construção da função de distribuição acumulada recomenda-se que no mínimo 5 pontos de corte sejam utilizados (GOOVERTS, 1997, p. 285). Foram utilizados os valores de corte $z_k = 0,007, 0,013, 0,017, 0,03, 0,039, 0,084, 0,38$ e $2,03$ para o manganês e $z_k = 0,04, 0,089, 0,117, 0,292, 0,836, 17,554$ para o ferro. Esses valores foram utilizados em semivariogramas experimentais indicativos, modelados por modelos esféricos apresentados, em anexo. Por causa da forte assimetria apresentada pelas distribuições de arsênio e cádmio foi utilizado o semivariograma indicativo da mediana nos respectivos pontos de corte. Os modelos semivariográficos foram utilizados na krigagem indicativa para se obter o modelo da função de distribuição acumulada condicional, $F(\mathbf{u};z|(n))$. Com o modelo de $F(\mathbf{u};z|(n))$ foi possível determinar a perda esperada com base nas equações (3.84) e (3.85) e, em seguida, realizar a classificação da área em estudo. Os resultados de cada metal são apresentados a seguir.

Para a função de perda L_1 (3.82) foi considerado um custo relativo de saúde $w_1 = 1,0$ e para função de perda L_2 (3.83), custos relativos de remediação de $w_2 = 0,5$ e $w_2 = 2,5$. Os custos associados à classificação errônea de uma localização como segura, ocasionando danos de saúde para o Mn, As, Fe e Cd são mostrados, respectivamente, nos mapas das Figuras 5.45, 5.50, 5.55 e 5.60. Da mesma forma, os custos associados à classificação errônea de uma localização como contaminada, levando a limpezas desnecessárias nas Figuras 5.46 e 5.47 (Mn), Figuras 5.51 e 5.52 (As), Figuras 5.56 e 5.57 (Fe) e Figuras 5.61 e 5.62 (Cd). Os mapas de custos esperados de saúde mostram que dentro da área da ICOMI onde as concentrações são mais elevadas, provavelmente ultrapassando o máximo permitido, os custos são maiores em se declarar uma localização segura. Observa-se, para o manganês, duas regiões com altos custos esperados de saúde (Figura 5.55), localizadas próximas à bacia de rejeitos (Foto 4.1) e em locais utilizados para receber as pilhas de manganês comercializável e não comercializável (Foto 4.2) e uma pequena área com custos pouco significativos no bairro da hospitalidade, onde foi realizado o aterro de resíduos de manganês. Quanto ao arsênio, os maiores custos esperados de saúde são localizados na área que abrange a bacia de rejeitos onde eram depositados os resíduos de manganês (Figura 5.50). O mapa mostrado na Figura 5.55 de custos esperados de saúde para o ferro indica a maior dispersão desse metal na área de estudo, principalmente no distrito industrial da ICOMI e bairro do Elesbão e o cádmio apresenta duas regiões menores com custos maiores de saúde localizados dentro da área da ICOMI (Figura 4.60). Os mapas de custos esperados de remediação, ao contrário, mostram que os custos em se declarar uma região como

contaminada são maiores onde as concentrações são possivelmente mais baixas, com menores probabilidades de excederem o limite máximo estabelecido. Pode-se observar nos mapas de classificação das áreas contaminadas mostradas nas Figuras 5.48 e 5.49 (manganês), Figuras 5.53 e 5.54 (arsênio), Figuras 5.58 e 5.59 (ferro) e Figuras 5.63 e 5.64 (cádmio) que, conforme o custo de remediação w_2 aumenta, a perda esperada associada à limpeza desnecessária aumenta relativamente aos custos esperados de saúde, levando a mais localizações classificadas como seguras. Nesse caso, a classificação das áreas contaminadas resultantes da minimização da perda esperada para os pesos $w_1=1$ e $w_2=2,5$ mostram maior contaminação para o ferro, seguido do manganês, arsênio e cádmio. A estimativa do custo global de saúde, fornecida pela soma dos custos em todas as localizações, foi maior para o ferro (5128,6), que apresentou maior dispersão (Tabela 2) e maior probabilidade marginal de contaminação (0,366, ver Figura 5.1), entre as variáveis analisadas. O manganês, arsênio e cádmio apresentaram estimativas de custos globais de saúde na ordem de 1980, 1180 e 663, respectivamente.

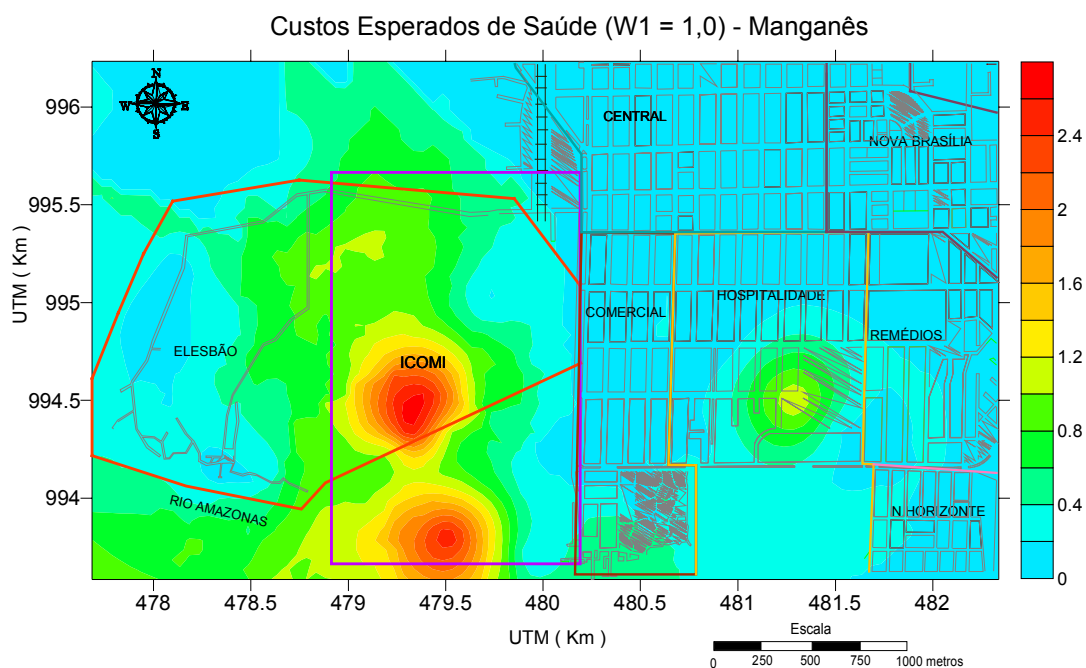


Figura 5.45 – Perda associada com uma classificação incorreta de uma localização como segura, para a função de perda L_1 , com $w_1 = 1,0$, para o manganês.

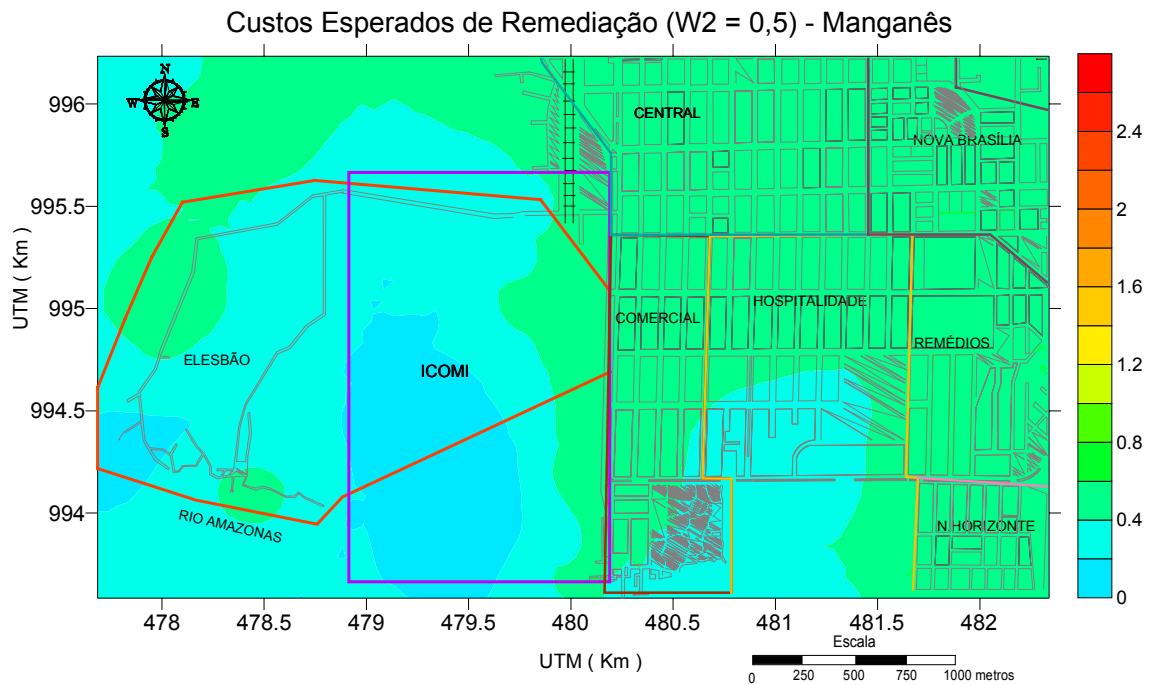


Figura 5.46 – Perda associada com uma classificação incorreta de uma localização como contaminada (limpeza desnecessária), para a função de perda L_2 , com $w_2 = 0,5$.

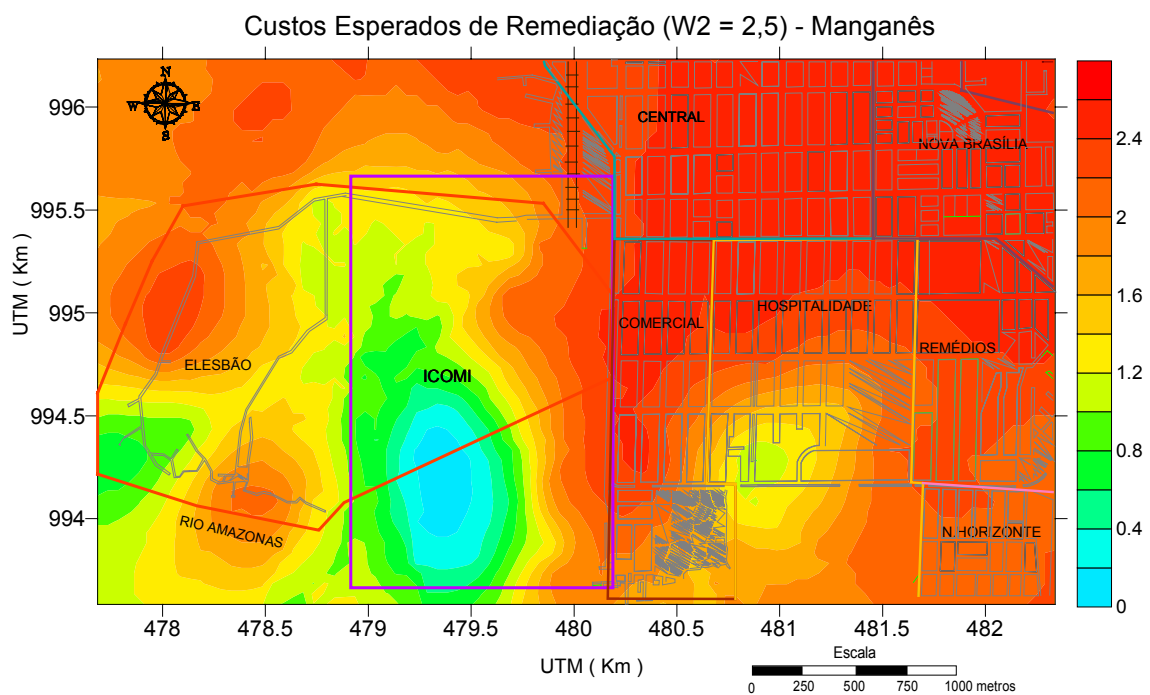


Figura 5.47 – Perda associada com uma classificação incorreta de uma localização como contaminada (limpeza desnecessária), para a função de perda L_2 , com $w_2 = 2,5$, para o manganês.

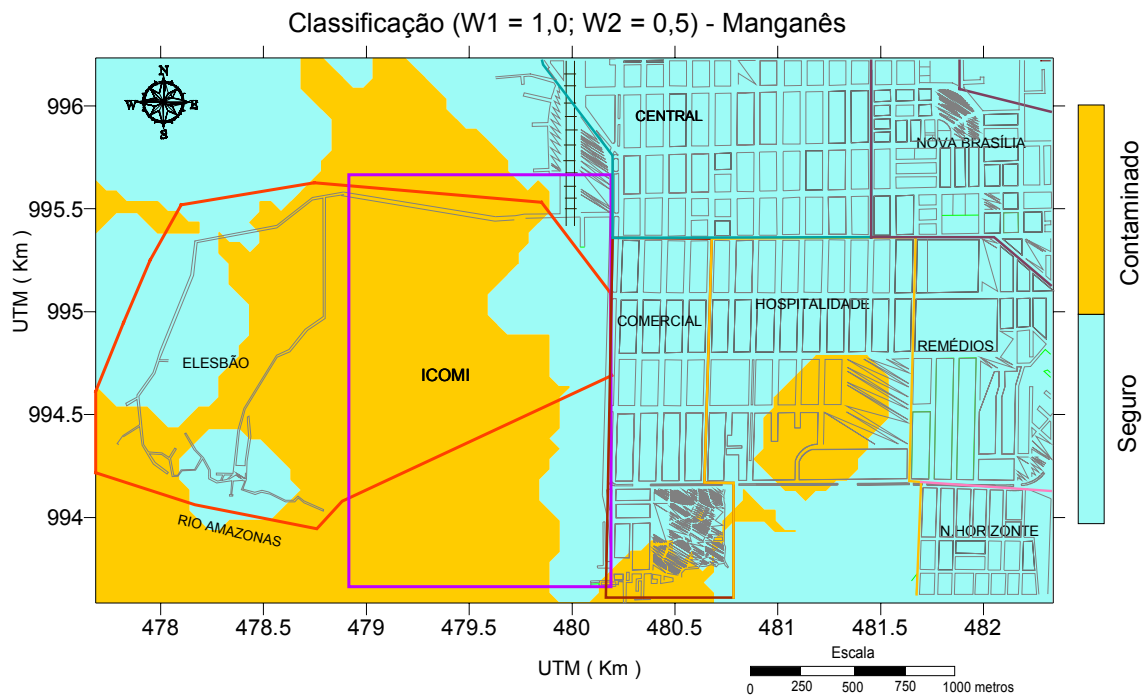


Figura 5.48 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , com $w_1 = 1,0$ e L_2 , com $w_2 = 0,5$, para o manganês.

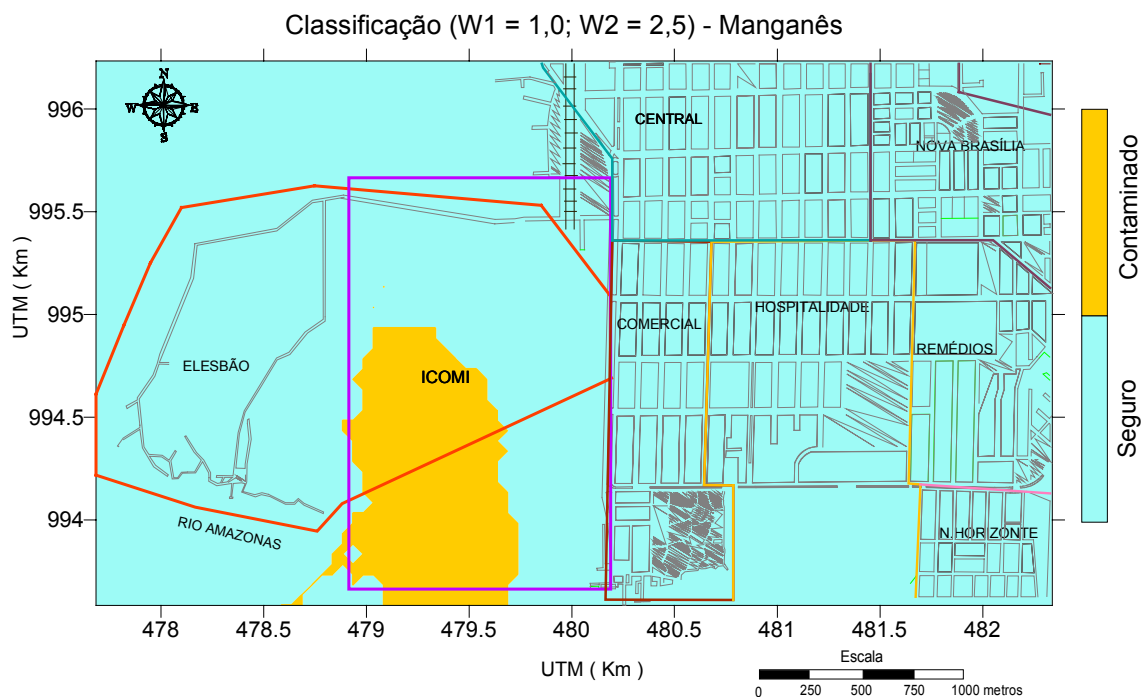


Figura 5.49 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , com $w_1 = 1,0$ e L_2 , com $w_2 = 2,5$, para o manganês.

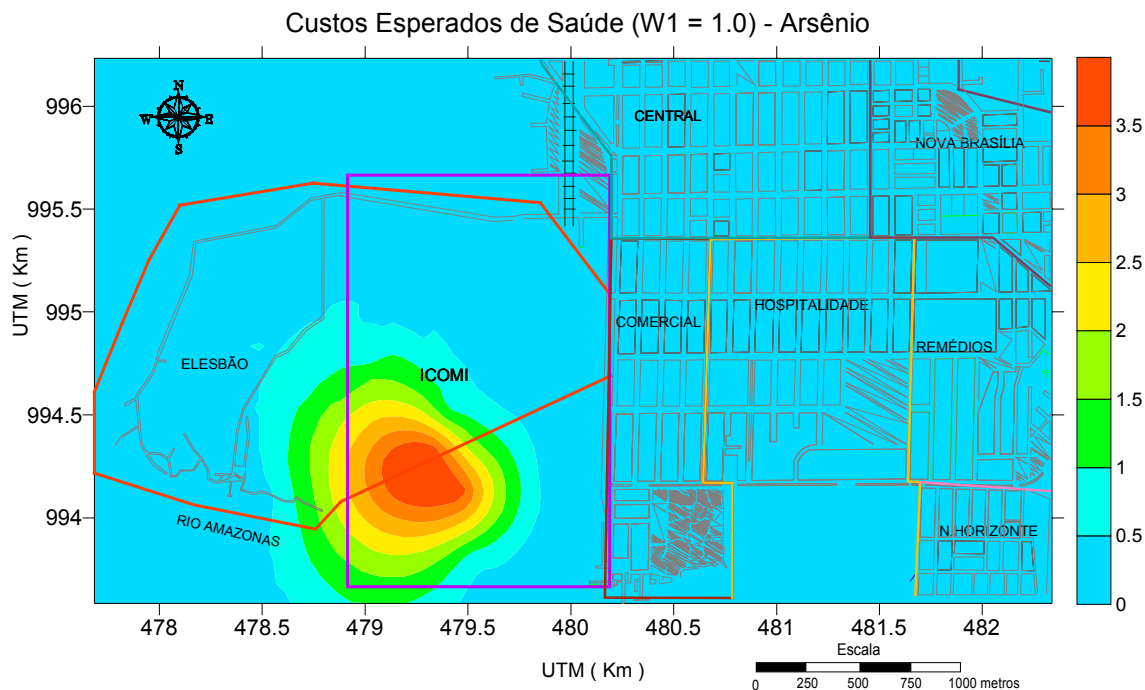


Figura 5.50 – Perda associada com uma classificação incorreta de uma localização como segura (danos á saúde), para a função de perda L_1 , com $w_1 = 1.0$, para o arsênio.

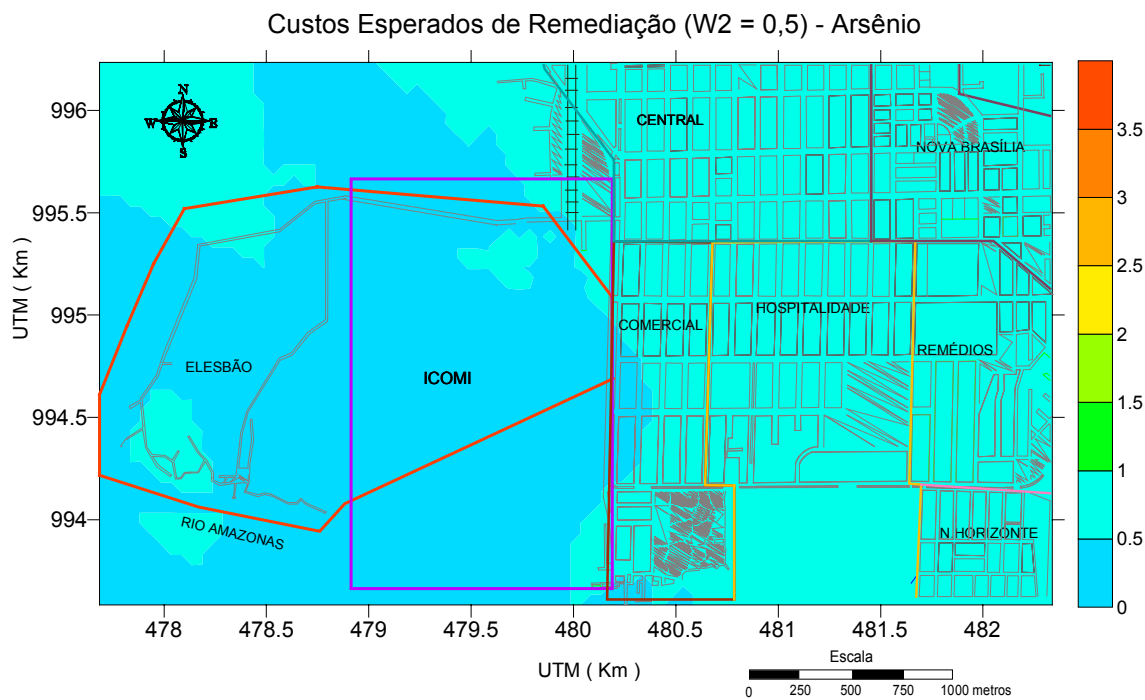


Figura 5.51 – Perda associada com uma classificação incorreta de uma localização como contaminada (limpeza desnecessária), para a função de perda L_2 , com $w_2 = 0,5$, para o arsênio.

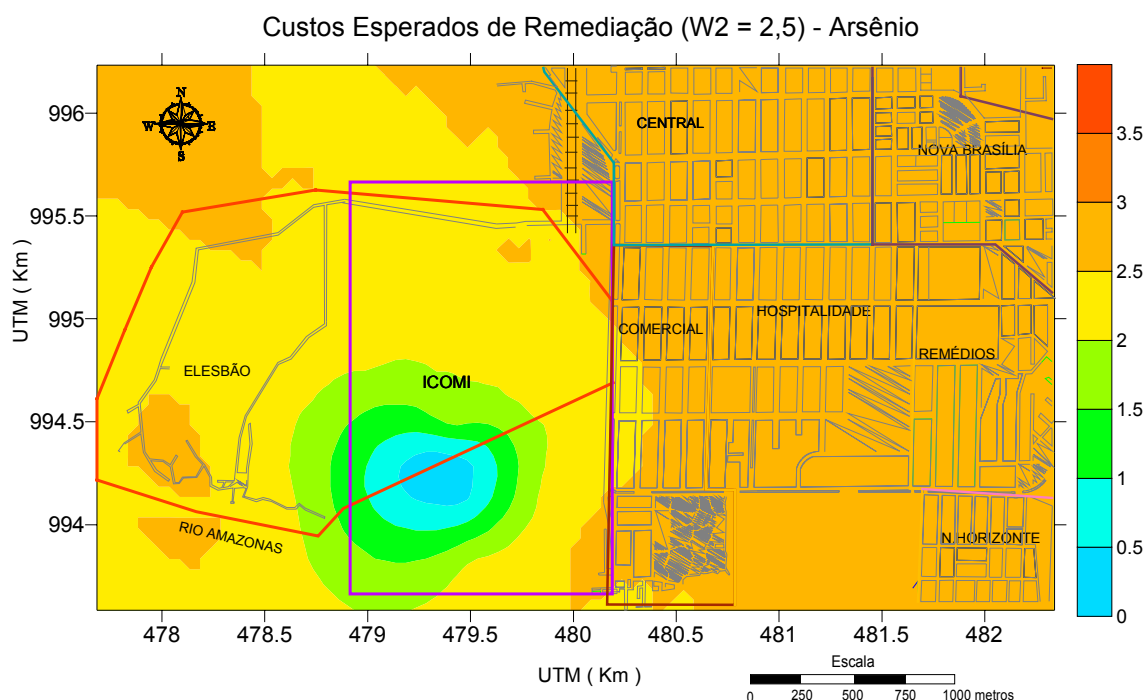


Figura 5.52 – Perda associada com uma classificação incorreta de uma localização como contaminada (limpeza desnecessária), para a função de perda L_2 , com $w_2 = 2,5$, para o arsênio.

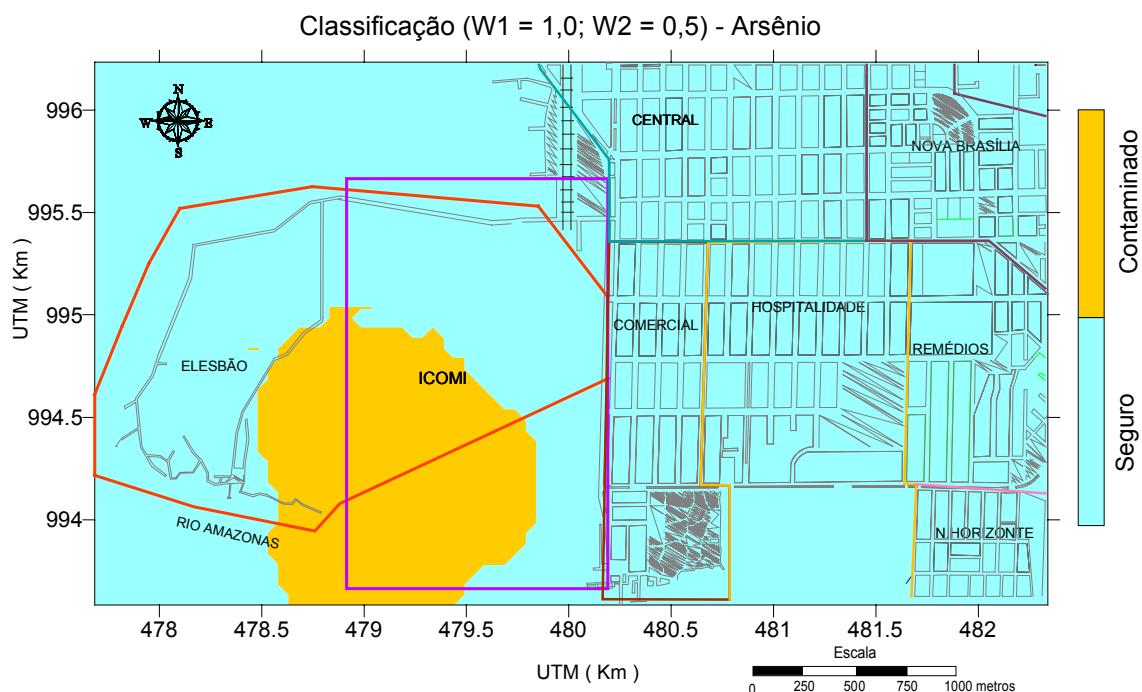


Figura 5.53 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , com $w_1 = 1,0$ e L_2 , com $w_2 = 0,5$, para o arsênio.

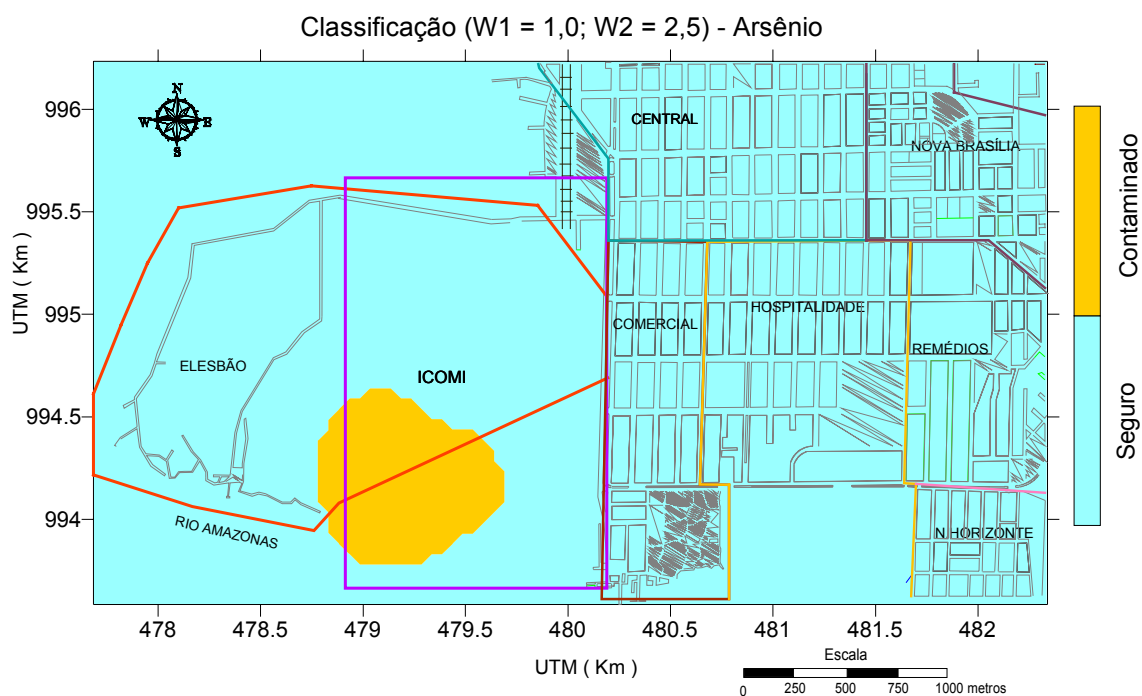


Figura 5.54 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , com $w_1 = 1,0$ e L_2 , com $w_2 = 2,5$, para o arsênio.

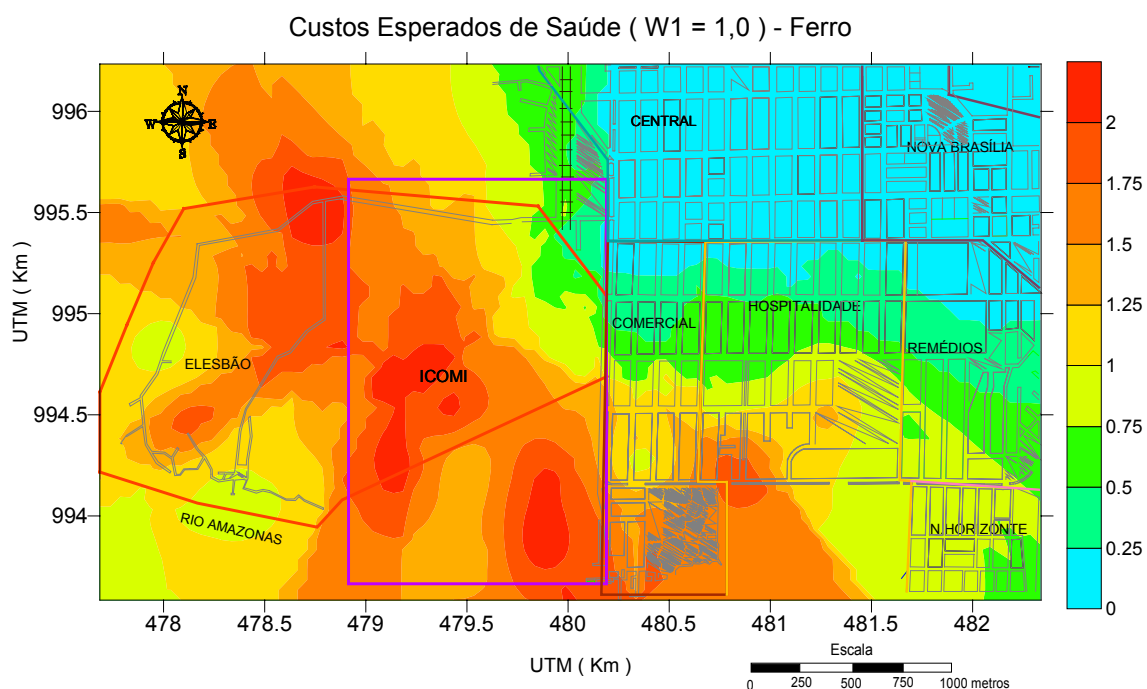


Figura 5.55 – Perda associada com uma classificação incorreta de uma localização como segura (danos à saúde), para a função de perda L_1 , com $w_1 = 1,0$, para o ferro.

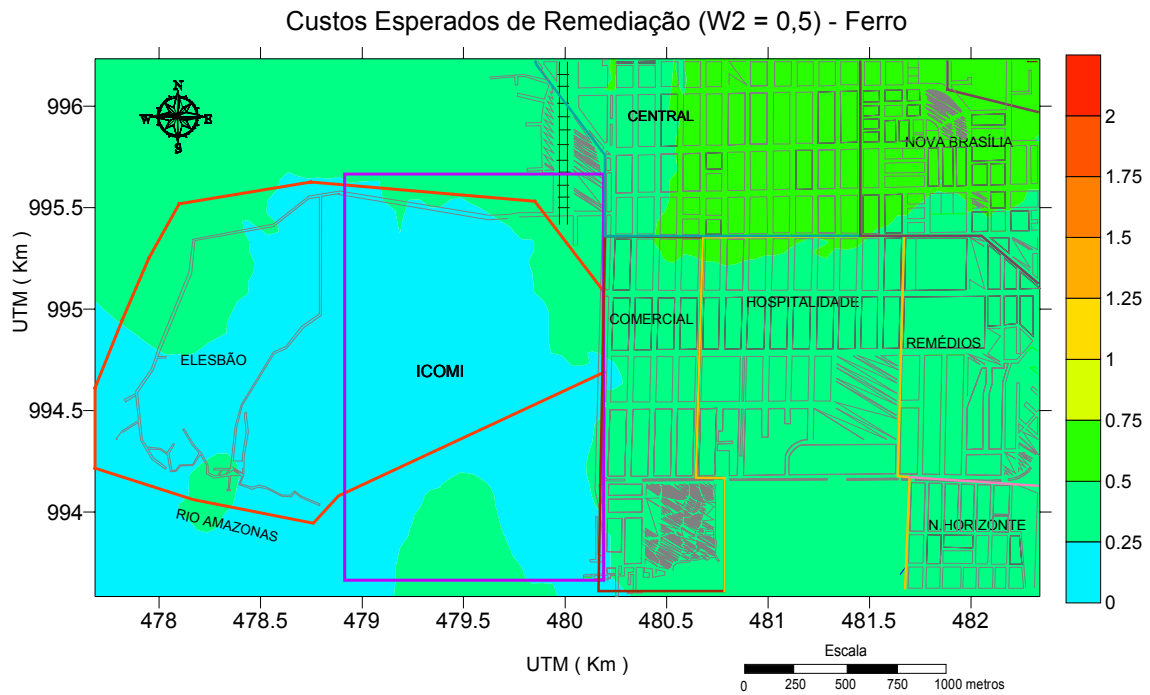


Figura 5.56 – Perda associada com uma classificação incorreta de uma localização como contaminada (limpeza desnecessária), para a função de perda L_2 , com $w_2 = 0,5$, para o ferro.

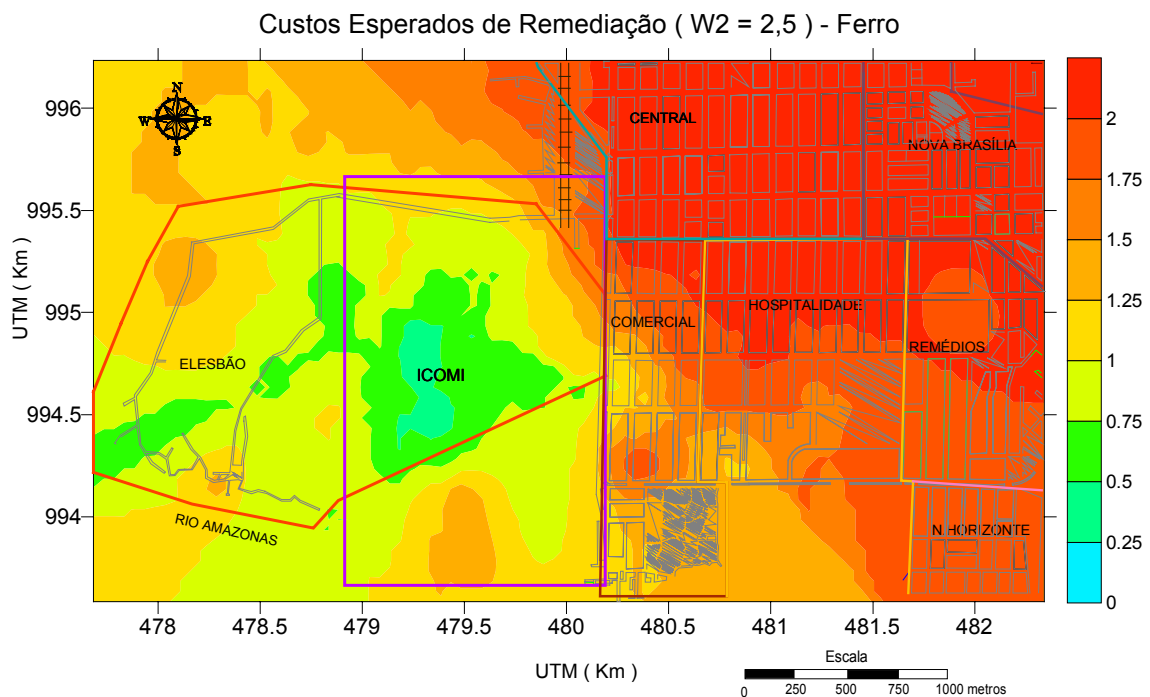


Figura 5.57 – Perda associada com uma classificação incorreta de uma localização como contaminada (limpeza desnecessária), para a função de perda L_2 , com $w_2 = 2,5$, para o ferro

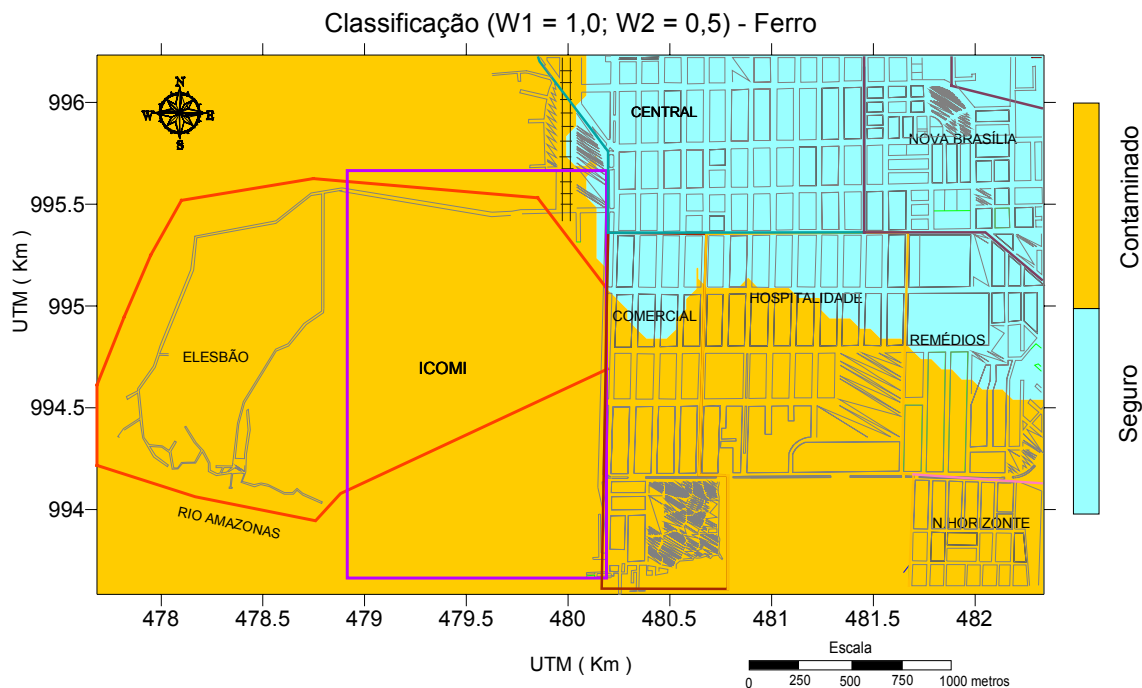


Figura 5.58 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , com $w_1 = 1,0$ e L_2 , com $w_2 = 0,5$, para o ferro.

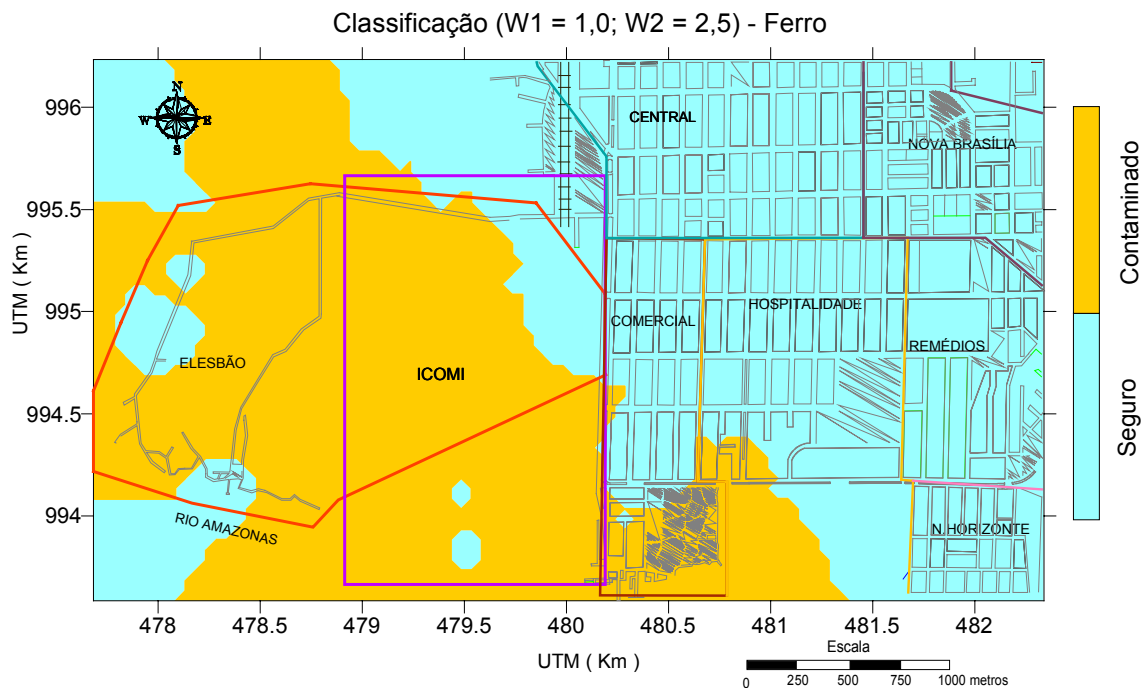


Figura 5.59 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , com $w_1 = 1,0$ e L_2 , com $w_2 = 2,5$, para o ferro.

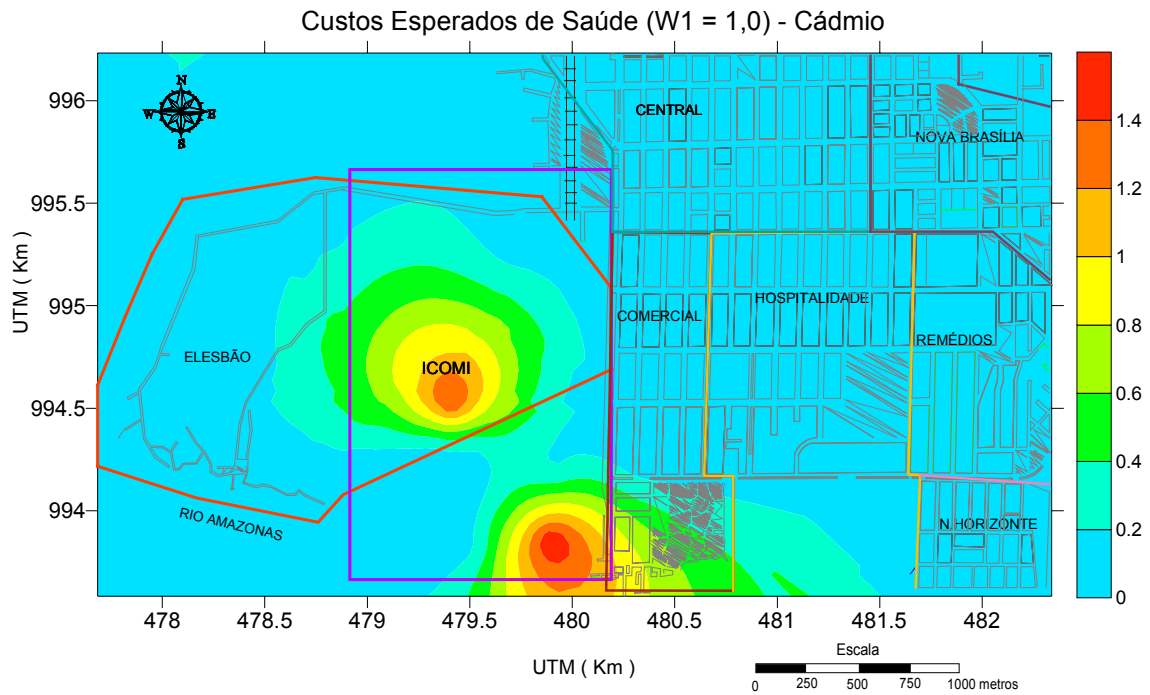


Figura 5.60 – Perda associada com uma classificação incorreta de uma localização como segura (danos à saúde), para a função de perda L_1 , com $w_1 = 1,0$, para o cádmio.

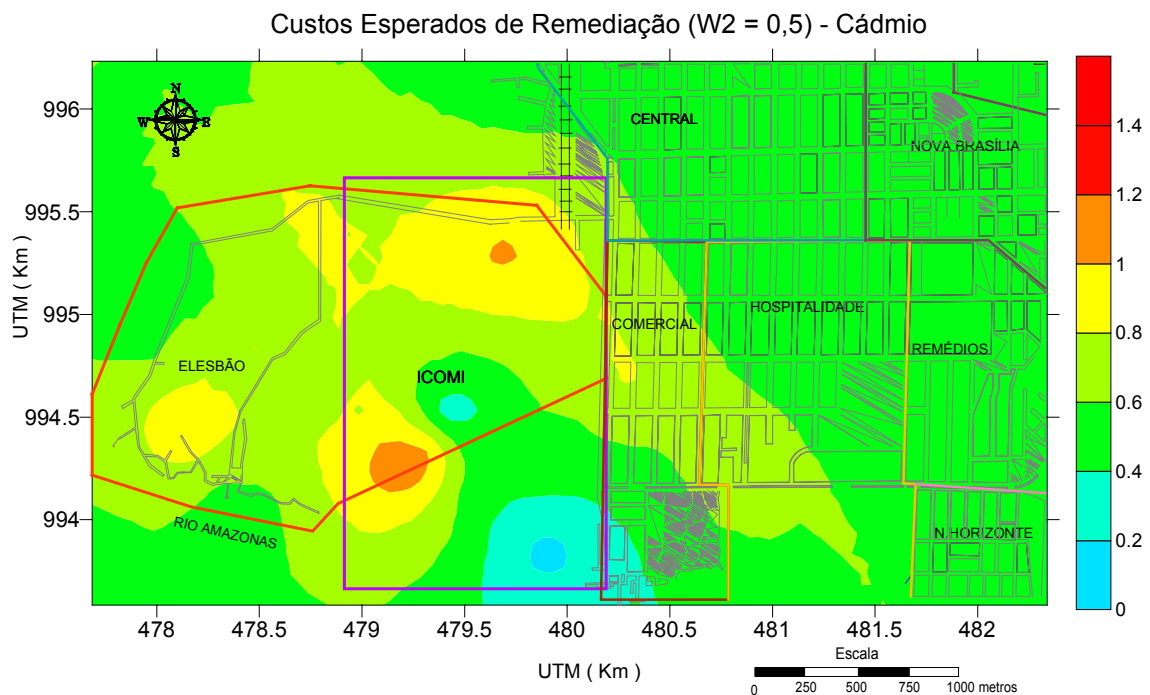


Figura 5.61 – Perda associada com uma classificação incorreta de uma localização como contaminada (limpeza desnecessária), para a função de perda L_2 , com $w_2 = 0,5$, para o cádmio.

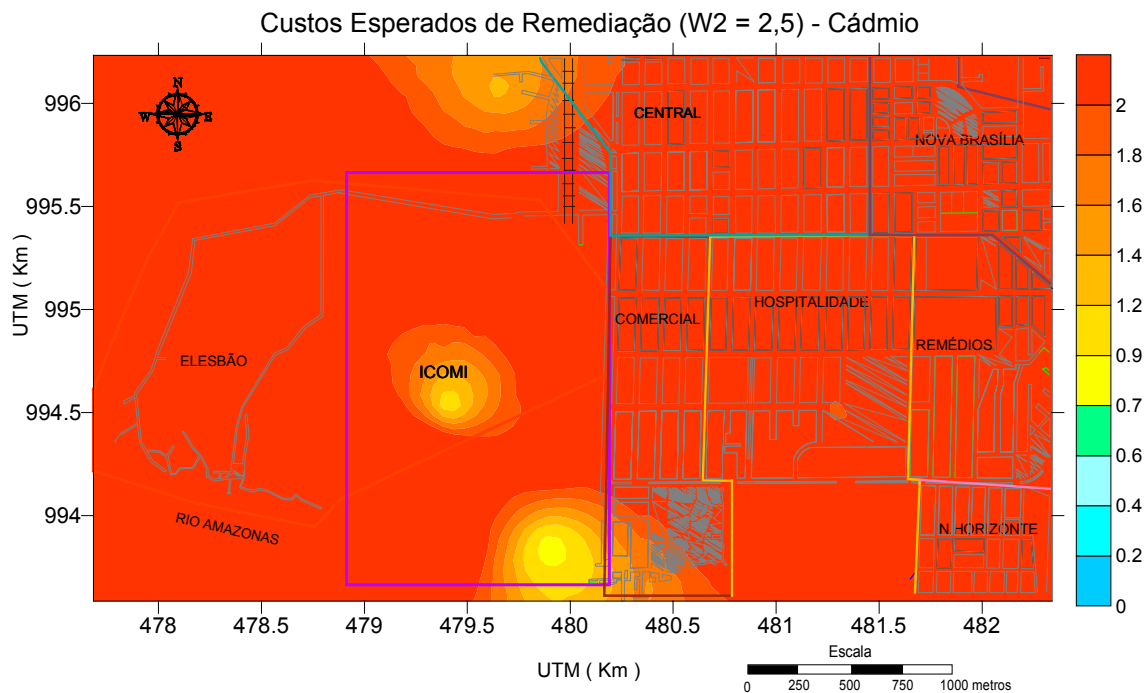


Figura 5.62 – Perda associada com uma classificação incorreta de uma localização como contaminada (limpeza desnecessária), para a função de perda L_2 , com $w_2 = 2,5$, para o cádmio.

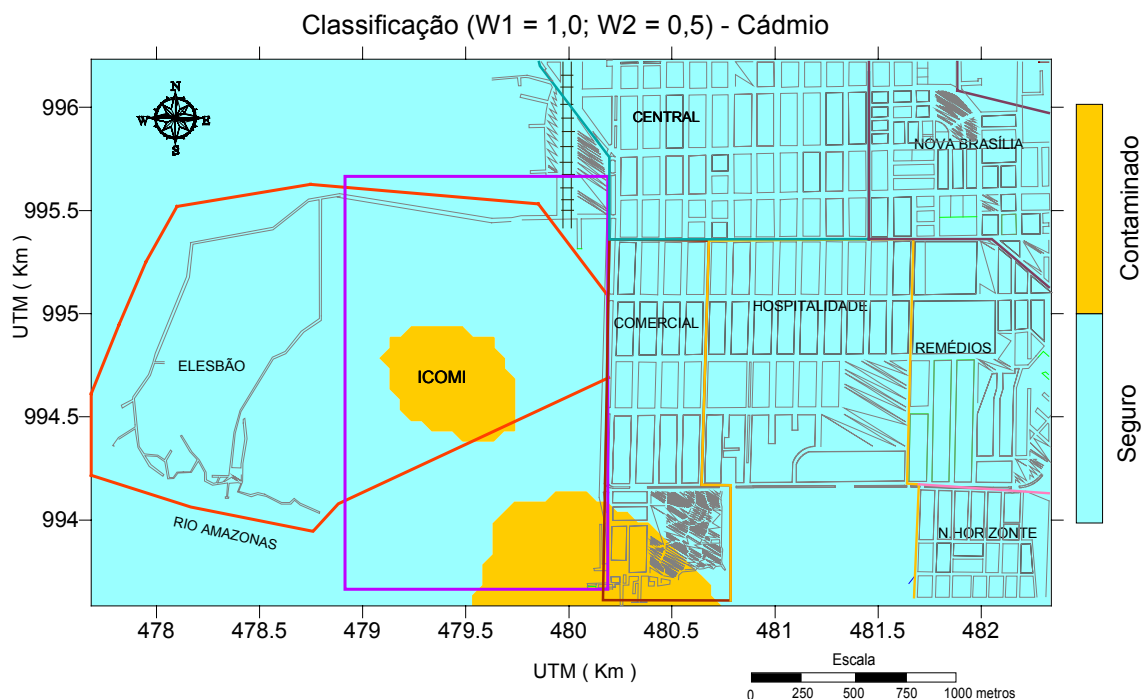


Figura 5.63 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , com $w_1 = 1,0$ e L_2 , com $w_2 = 0,5$, para o cádmio.

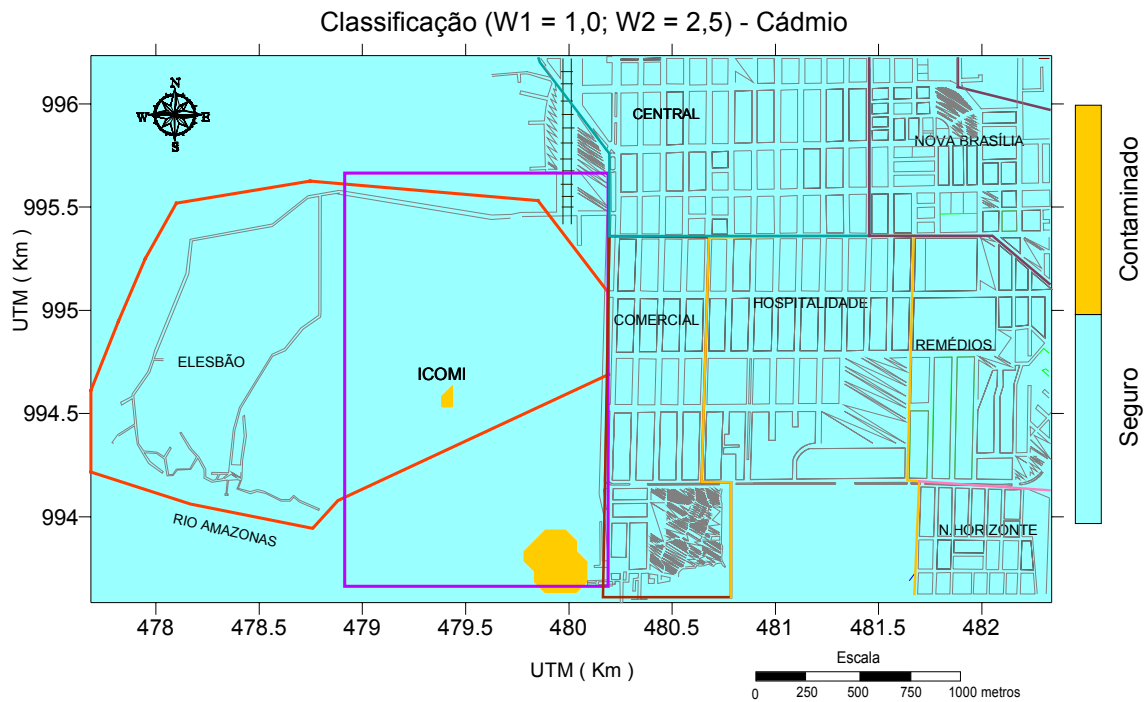


Figura 5.64 – Classificação resultante da minimização da perda esperada para as funções de perda L_1 , com $w_1 = 1,0$ e L_2 , com $w_2 = 2,5$, para o cádmio.

Pela utilização de modelos de incerteza local para a classificação das áreas contaminadas através de funções de perda relacionadas a custos econômicos, pôde-se delimitar as áreas contaminadas na região de estudo. Fortes indícios de contaminação por manganês e arsênio ocorrem principalmente dentro da área industrial da ICOMI. A grande dispersão apresentada pelo ferro levou à classificação de uma grande área contaminada por esse elemento, incluindo praticamente toda a área industrial da ICOMI, bairro do Elesbão e parte da área residencial. A contaminação por cádmio parece ser de magnitudes bem menores em relação aos outros elementos analisados.

5.2.5. Utilização da simulação annealing na avaliação da concentração dos contaminantes na área de estudo

GOOVAERTS (1998) mostra que a utilização de métodos de interpolação, como a krigagem, para o mapeamento da concentração de poluentes tendem a suavizar detalhes locais da variação espacial da variável em estudo, levando a superestimar valores mais baixos e subestimar valores mais altos. Essa suavização pode comprometer os resultados quando se deseja detectar grandes concentrações de poluentes. Por outro lado, a simulação estocástica permite a criação de mapas de concentrações de poluentes que reproduzem a variabilidade espacial dos dados sem o efeito de suavização. Além disso, pode-se gerar múltiplas simulações (realizações) que se ajustam razoavelmente às mesmas estatísticas amostrais. O conjunto de simulações alternativas são então pós-processados para se atingir diversos objetivos, como a construção de mapas probabilísticos para determinados níveis de corte, utilização de funções de perda para o delineamento de áreas para remediação e avaliação de impactos financeiros para determinados cenários.

As características apresentadas pelos conjuntos de dados em estudo, como presença de valores extremos, forte assimetria e poucos dados, conduziram à escolha do método da simulação annealing, descrita na parte 3 deste trabalho, que apresenta maior flexibilidade no tratamento de dados com as características acima (DEUTSCH, 2002, p. 275). As simulações foram realizadas para os metais manganês, arsênio, ferro e cádmio que apresentaram valores amostrais acima dos limites máximos permitidos por órgãos governamentais. Uma medida quantitativa e visual da incerteza espacial é fornecida pela geração de muitas simulações que se ajustem razoavelmente às mesmas amostras estatísticas (histograma e variograma) e dados condicionantes. Características especiais como grupos específicos de altos valores, que podem indicar a presença de contaminação, apresentam pouca incerteza quanto à sua ocorrência se são observadas na maioria das simulações. Por outro lado, a incerteza sobre determinada característica é alta se ela for observada somente em poucos mapas simulados.

Primeiramente, devido a forte assimetria apresentada pelas variáveis, foram elaborados gráficos de dispersão das médias versus variâncias calculados por meio de janelas móveis, para se avaliar a existência de efeito proporcional. Os resultados obtidos para cada metal são apresentados na Figura 5.65.

Pode-se observar a existência do efeito proporcional em todas as variáveis indicado pela forte correlação entre as médias e variâncias. O maior efeito proporcional foi apresentado pelo ferro (0,90 e 0,74 para a correlação linear e rank, respectivamente) e o manganês (0,72 e 0,51 para a correlação linear e rank, respectivamente) mostrou menor efeito proporcional. Por causa do forte efeito proporcional observado, foram utilizados semivariogramas relativos das variáveis (equação 3.26) na reprodução dos modelos semivariográficos nas simulações. A Figura 5.66 mostra os semivariogramas relativos das variáveis que, neste caso, se mostraram mais estáveis dos que os semivariogramas tradicionais (Figura 5.29). Uma comparação entre as Figuras 5.29 e 5.66 mostra que a suavização provocada pela padronização nos semivariogramas relativos minimizou significativamente os efeitos de escala observados nos semivariogramas tradicionais.

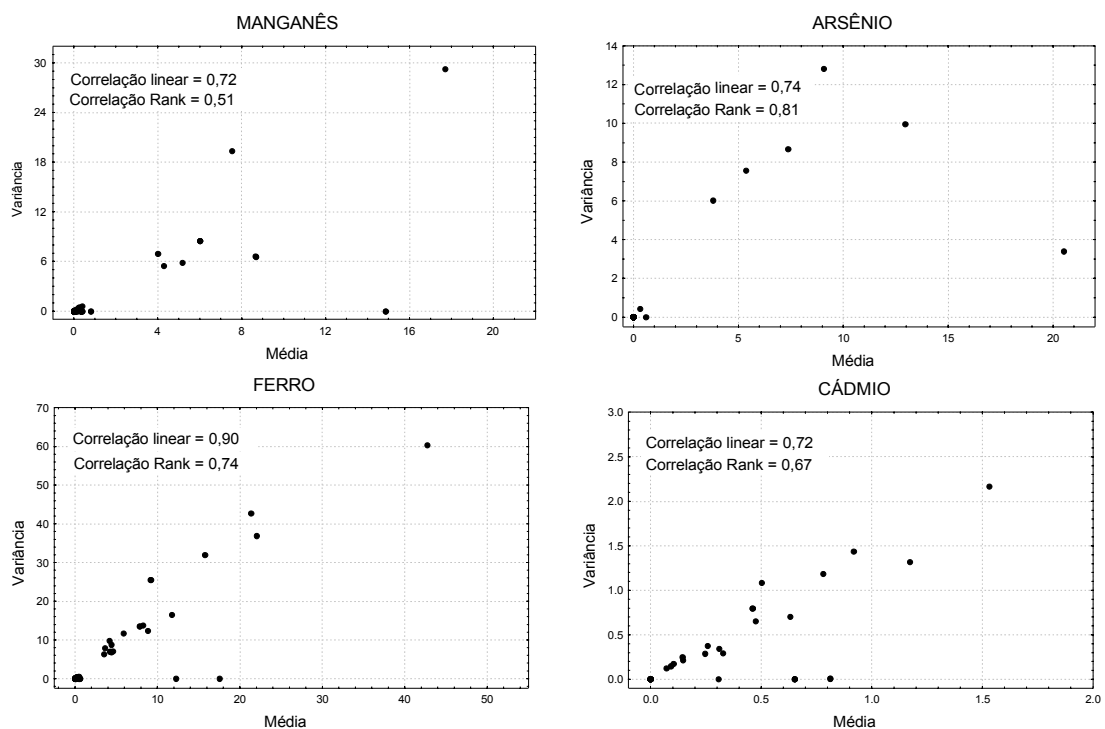


Figura 5.65. Gráficos de dispersão da média versus variância para janelas móveis de 0,5 x 0,5 km sobre a área de estudo para as variáveis em estudo.

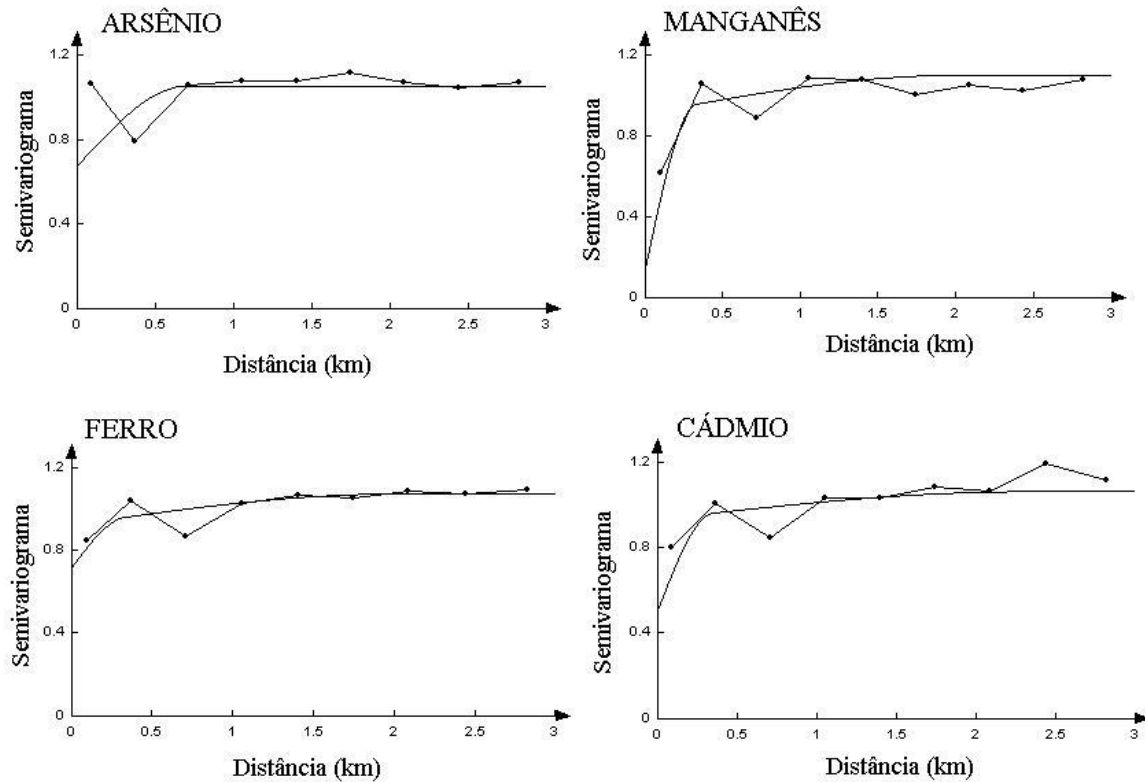


Figura 5.66 – Semivariogramas relativos experimentais (linhas ligados por pontos) e modelos ajustados (linhas contínuas) do arsênio, manganês, ferro e cádmio.

Foi utilizado o programa *sasim* da biblioteca GSLIB (DEUSCH E JOURNAL, 1998, p. 158) na geração das simulações e a Tabela 5.7 mostra os parâmetros utilizados na programação annealing para os metais em estudo. DEUTSCH e COKERHAM (1994) apresentam algumas sugestões para a escolha dos parâmetros, que auxiliaram bastante na execução das simulações. Foram geradas 100 simulações para cada variável sendo que, cada simulação, por construção, reproduz os dados amostrais nas localizações originais (isso é verdadeiro somente para as localizações dos dados que correspondem exatamente a um nó da malha simulada) e preserva as características estatísticas (histograma e semivariograma) dos dados amostrais. A reprodução do histograma amostral (referência) de cada variável pode ser visualizada pelo Q-Q plot entre os valores simulados e os valores de referência, mostrados na Figura 5.67. Foram plotados os quantis das 10 primeiras simulações de cada variável contra os mesmos quantis dos dados amostrais. Se os histogramas das simulações e dos dados de referências são similares, os pontos deveriam cair próximos à linha de 45°. Pode-se observar na Figura 5.67 que o manganês e o cádmio apresentaram melhores

resultados na reprodução dos histogramas amostrais. As simulações de arsênio e ferro apresentaram, em geral, valores um pouco acima dos valores de referência como indicam os pontos acima da linha de 45°. Mesmo assim, é possível considerar que não houve discrepâncias significativas nos histogramas simulados em relação aos histogramas amostrais.

Na Figura 5.68 são apresentados 20 semivariogramas obtidos de 20 (dentre as 100) simulações de cada variável. Os semivariogramas simulados apresentam mais variabilidade do que os histogramas, principalmente no caso do arsênio e ferro, na zona de maior correlação espacial (alcance). Entretanto, a qualidade da reprodução dos padrões de continuidade espacial pelos semivariogramas simulados pode ser considerada aceitável dada a disponibilidade de poucos dados amostrais, o que torna, de certa forma, mais difícil computar os semivariogramas experimentais e, conseqüentemente, modelos semivariográficos com melhores ajustes.

TABELA 5.7. Parâmetros (adimensionais) utilizados na programação *annealing*

Variável	T_0	λ	K_{\max}	K_{aceitos}	S	O_{\min}
Manganês	1,0	0,25	10	5	5	0,001
Arsênio	1,0	0,05	100	10	3	0,0001
Ferro	1,0	0,25	100	10	5	0,001
Cádmio	1,0	0,15	100	10	5	0,001

t_0 : “temperatura” inicial

λ : fator multiplicativo para redução da temperatura

K_{\max} : número máximo de perturbações tentadas

K_{aceitas} : número de perturbações aceitas

S : número de parada quando o número de perturbações atinge K_{\max}

O_{\min} : valor mínimo a ser alcançado pela função Objetivo

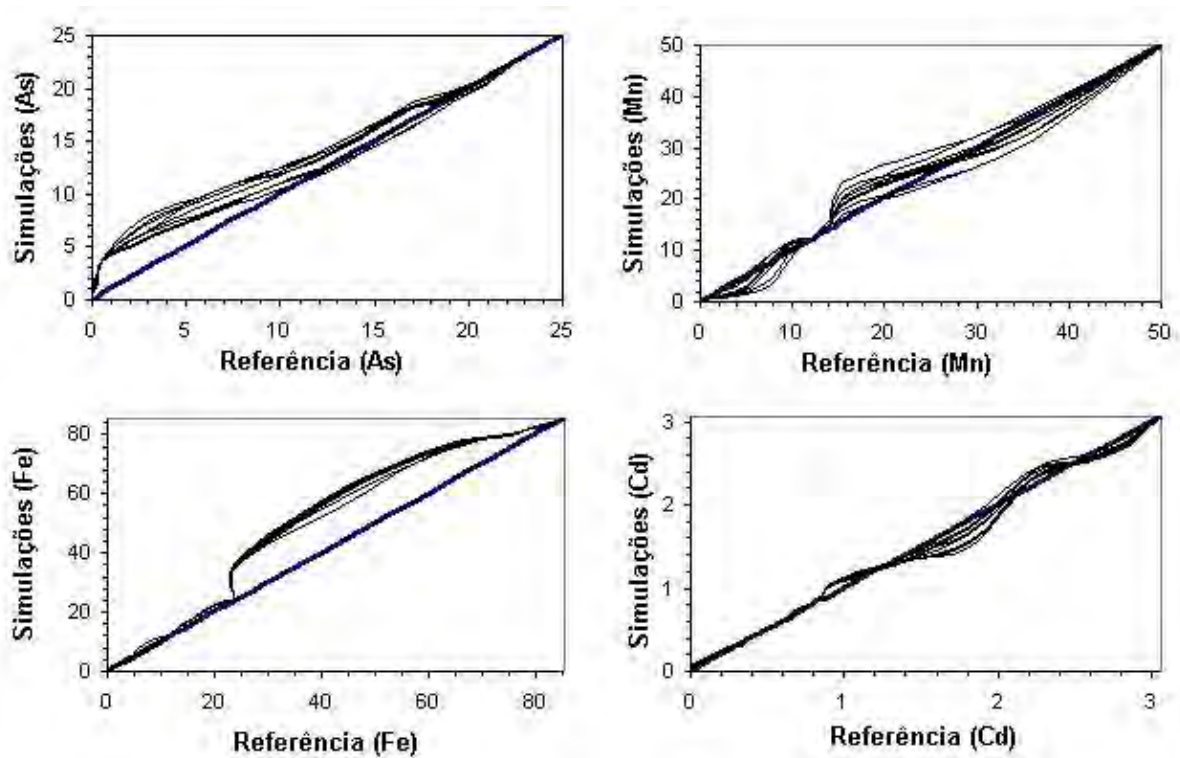


Figura 5.67 – Q-Q plot de 10 simulações versus o conjunto de referência (amostra) dos dados de arsênio, manganês, ferro e cádmio.

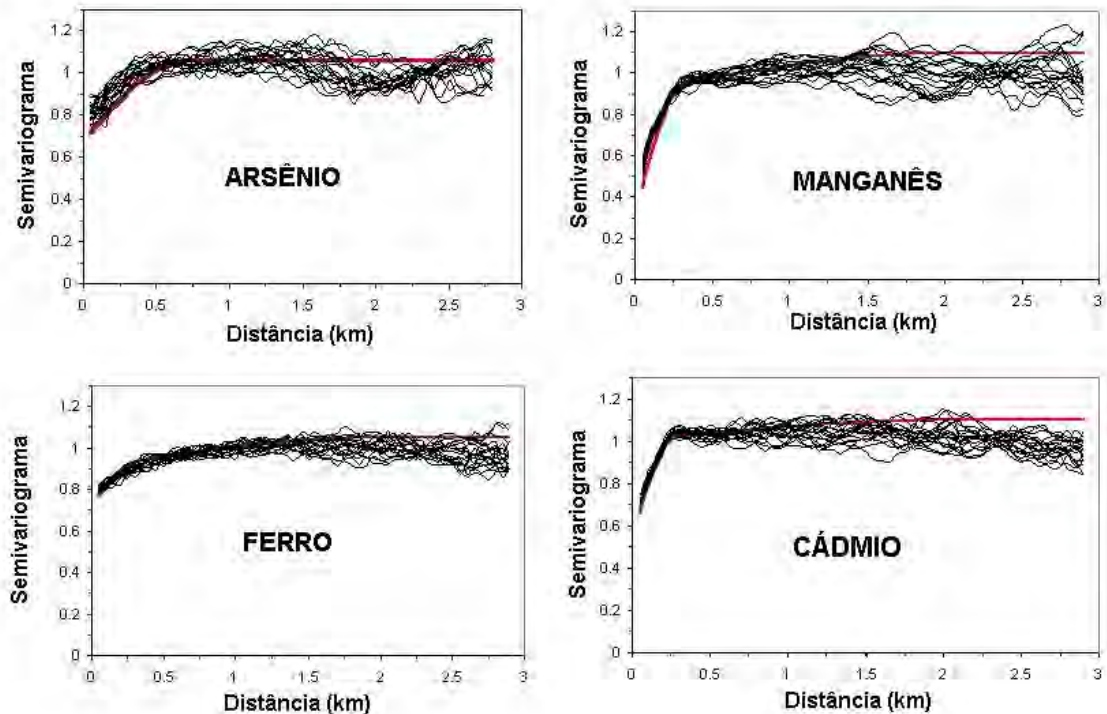


Figura 5.68 – Reprodução dos semivariogramas de 20 simulações para arsênio, manganês, ferro e cádmio. As linhas sólidas em vermelho representam os modelos ajustados.

5.2.6. Avaliação da incerteza a partir das simulações

Uma vez que as simulações reproduzem as características estatísticas das variáveis em estudo pode-se obter uma aproximação da função densidade de probabilidade em cada nó da malha, a partir dos valores simulados em um conjunto de imagens (mapas). Uma vez conhecida a distribuição de probabilidade em cada ponto da malha, os critérios para a delimitação de áreas perigosas (ou contaminadas) podem ser definidos.

Uma aproximação direta consiste em declarar contaminados todos os lugares onde a probabilidade de ultrapassar o limite máximo permitido, z_c , é maior do que uma determinada probabilidade de corte p_c . A escolha da probabilidade de corte a ser usada é determinada por diversos fatores, sendo importante considerar o impacto sobre a saúde e meio ambiente. Todavia, em geral, a probabilidade de corte é selecionada por negociação entre clientes/usuários e agências reguladoras (ISTOK e RAUTMAN, 1995). Como exemplo, neste trabalho, foi utilizada na delimitação das áreas contaminadas para cada metal a probabilidade marginal de contaminação, p_c , definida como a proporção de valores observados acima do limite máximo permitido para o respectivo metal. Na Figura 5.69 são apresentados os mapas de probabilidades para cada variável, onde foi considerado como valor de corte, z_c , o limite máximo permitido, ou seja, 0,05 ppm, 0,1 ppm, 0,3 ppm e 0,0001 ppm para o arsênio, manganês, ferro e cádmio, respectivamente. Esses mapas foram construídos a partir do conjunto de simulações pela contagem do número de pixels que excederam o limite de corte, z_c , estabelecido para cada metal, convertendo-se a soma em proporção e, em seguida, a probabilidade empírica apresentada em forma de mapa variando espacialmente. Considerando-se, por exemplo, uma probabilidade em torno de 40% ou 50%, esses mapas mostram que a maior dispersão de concentrações é apresentada pelo ferro, seguido pelo manganês, arsênio e cádmio. A maior probabilidade de ocorrência de altas concentrações localiza-se dentro da área industrial da ICOMI para todos os metais, com destaque para o manganês que apresenta maior área de contaminação. No bairro da hospitalidade, onde ocorreu o aterro de ruas com rejeitos de manganês, existem sinais de contaminação, principalmente de arsênio e manganês. A classificação das áreas contaminadas é apresentada na Figura 5.70 com base nos mapas de probabilidades acima, considerando-se como probabilidade de corte, as probabilidades marginais de

contaminação dos respectivos metais, ou seja, 0,098 para o arsênio, 0,27 para o manganês, 0,366 para o ferro e 0,073 para o cádmio.

Essa classificação não considera sub-regiões com diferentes sensibilidades à contaminação. Nesse caso, por exemplo, poderia ser tolerado um risco de 50% de contaminação para determinado metal dentro da área industrial, enquanto para áreas residenciais o mesmo risco ser considerado inaceitável. Nessas circunstâncias se poderia considerar uma probabilidade de corte que variasse de acordo com o uso da terra em cada lugar de interesse. Entretanto, a escolha de uma probabilidade de corte é um tanto subjetiva e, em tomada de decisões, muitas vezes é importante que mais critérios de classificação estejam disponíveis.

Outro critério para classificação consiste em se declarar contaminada todas as localizações cujas estimativas das concentrações do poluente se encontrem acima do limite máximo tolerável. Neste caso, é necessário estimar as concentrações em locais não amostrados, correndo-se o risco de se classificar incorretamente como contaminada uma localização segura e vice-versa. Como foi visto na parte 3 do presente trabalho, um critério de otimização para a minimização do erro pode ser através de funções de perda. Para cada metal foram avaliadas as estimativas ótimas baseadas nas funções de perda relativas ao desvio médio quadrático (média) e desvio médio absoluto (mediana e quantis) de acordo com as equações 3.79, 3.80 e 3.81, respectivamente. Os resultados são mostrados nas Figuras 5.71 a 5.78. A Figura 5.71 mostra os mapas de estimativas E-type, onde as maiores concentrações de todos os metais estudados aparecem dentro da área industrial da ICOMI, localizadas praticamente na mesma região. Fora da área da ICOMI, no bairro da hospitalidade, observa-se locais de altas concentrações de manganês, arsênio e ferro. As simulações resultantes para o ferro apresentaram efeitos indesejáveis, denominados efeitos de borda, que podem ser observados no mapa de estimativa E-type desse elemento. DEUTSCH e COCKERMAN (1994) e CARLE (1997) mostram a ocorrência de efeitos de borda na presença de distribuições com forte assimetria. Pode ser que o efeito proporcional altamente significativo apresentado pelo referido metal também tenha contribuído para a ocorrência dos referidos efeitos. Essa última hipótese, entretanto, necessita de estudos adicionais para a avaliação da relação entre efeitos de borda e efeito proporcional.

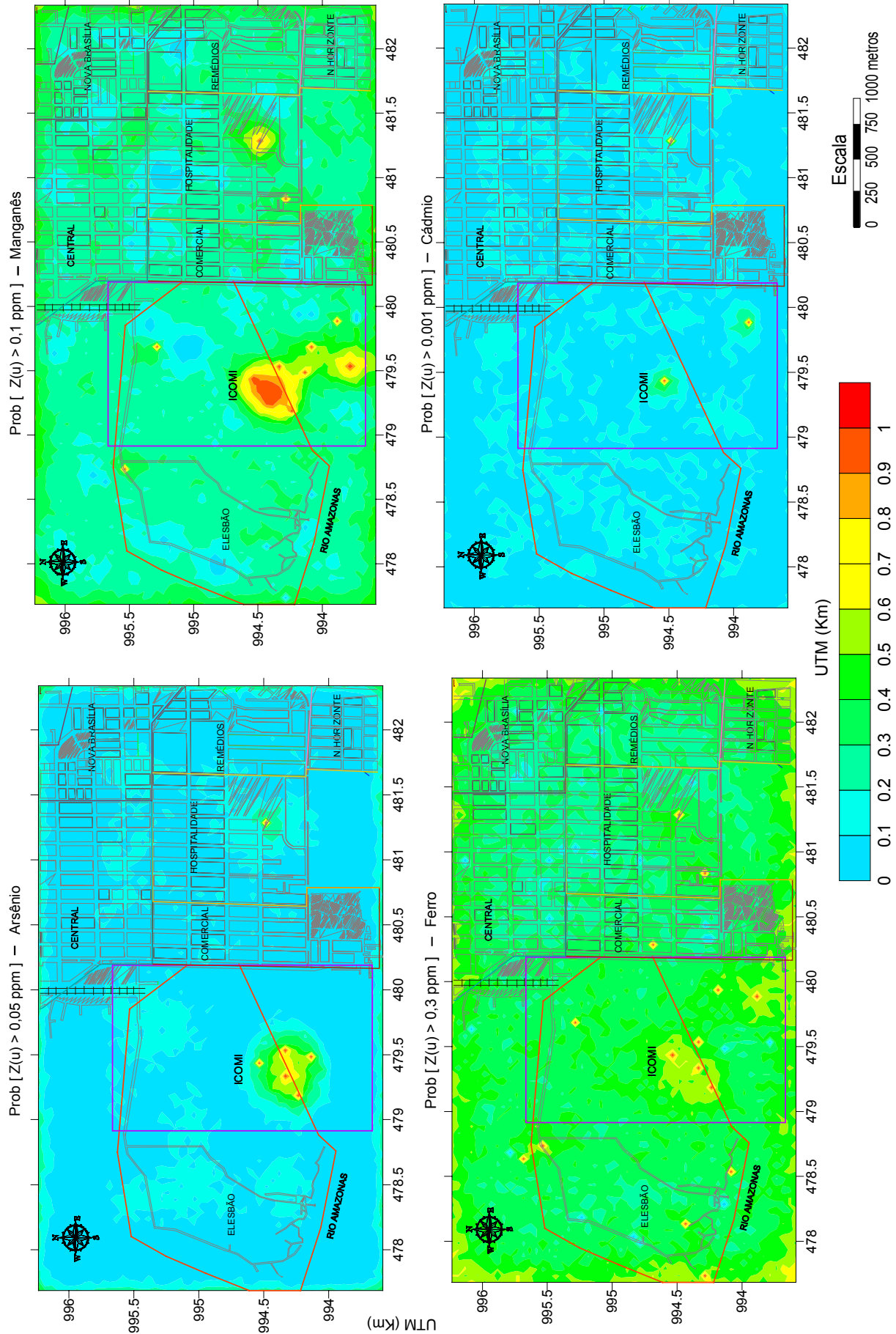


Figura 5.69 – Mapas de probabilidade das concentrações de As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita) excederem o limite máximo permitido de 0,05 ppm, 0,1 ppm, 0,3 ppm e 0,001 ppm, respectivamente.

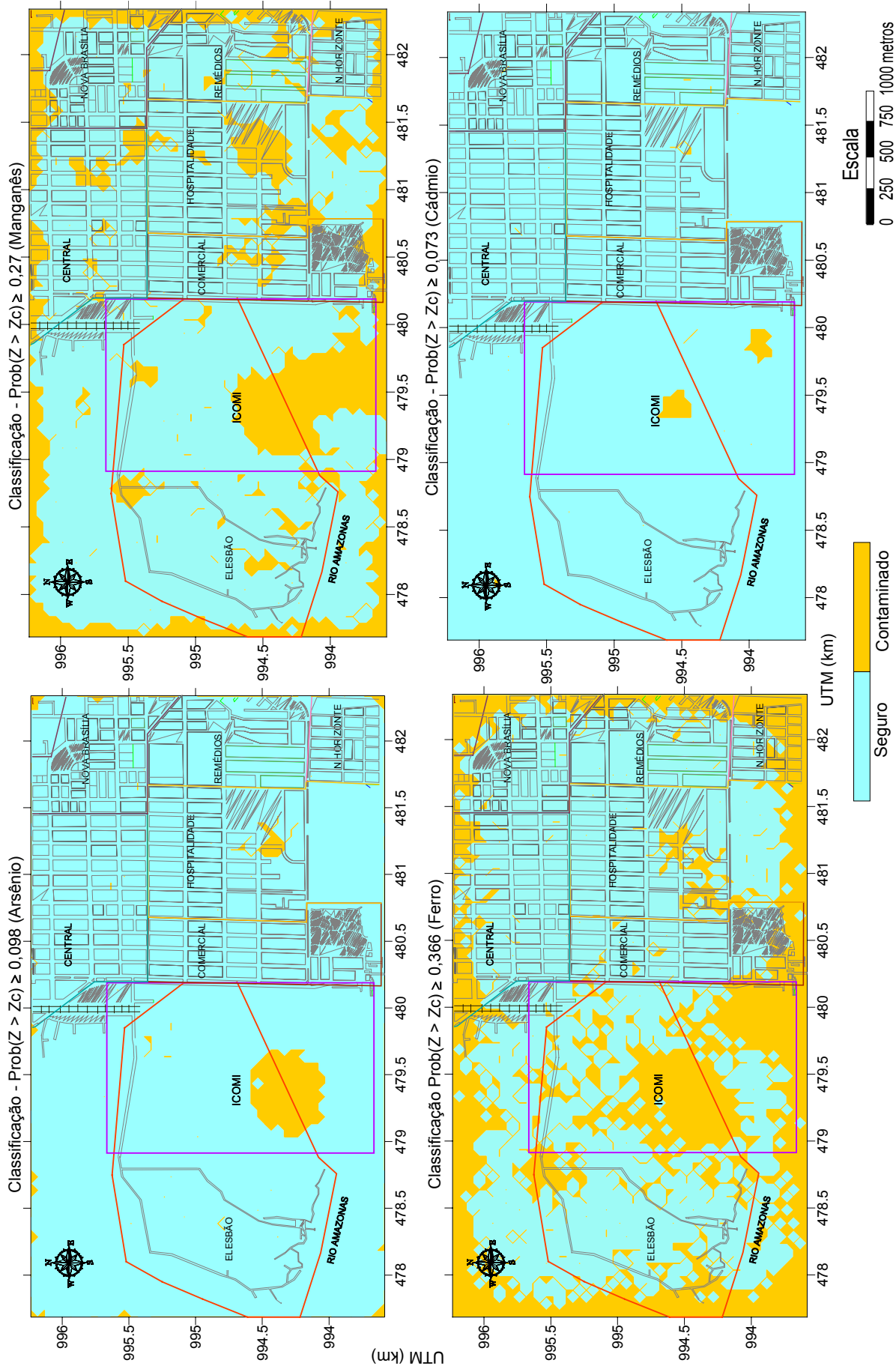


Figura 5.70 – Classificação das localizações com base na probabilidade das concentrações acima dos limites máximos permitidos serem maiores do que as probabilidades marginais de contaminação de 0,098, 0,27, 0,366 e 0,073 para As, Mn, Fe e Cd, respectivamente.

Os mapas de classificação apresentados nas Figuras 5.73 e 5.74 consideram as áreas como contaminadas para as estimativas da média (E-type) e mediana acima dos limites máximos permitidos para cada metal.

Os valores de concentrações apresentados nos mapas da mediana, na Figura 5.72, são muito menores do que os valores dos mapas da estimativa da média (E-type) (Figura 5.71). Isso ocorre porque, ao contrário da estimativa da média que tende a superestimar as concentrações fazendo com que mais localizações sejam consideradas contaminadas, a mediana não é afetada por valores extremos das distribuições assimétricas das variáveis em estudo. Deste modo, as classificações das áreas contaminadas baseadas em funções de perda simétricas como nos casos dos desvios médios quadráticos (média) e desvio médio absoluto (mediana) diferem bastante um do outro, e não parecem adequados para avaliar o impacto do erro de estimação em casos de contaminações, conforme se pode verificar nos mapas das Figuras 5.71 a 5.74. Enquanto os locais classificados como contaminados com base na estimativa da média (E-type) tomam quase a totalidade da área em estudo, exceto para o cádmio, as classificações baseadas na estimativa mediana mostram poucas ou quase nenhuma localização contaminada, como no caso de arsênio e cádmio.

Em situações onde decisões sobre a limpeza ou remediação de áreas contaminadas precisam ser tomadas, as consequências da subestimativa de uma concentração tóxica (acarretando em danos de saúde) pode ser muito maior do que uma superestimativa de mesma magnitude (limpeza desnecessária). Nesse caso, são mais adequadas funções de perda assimétricas como aquela baseada na expressão 3.80, onde a melhor estimativa corresponde ao quantil p . GOOVAERTS (1977, p. 366) mostrou que, quando se usa a estimativa quantil p para a classificação de localizações contaminadas, a proporção de localizações classificadas erroneamente é mínima quando a estimativa quantil p corresponde a uma probabilidade próxima a $p = F^*(z_c)$, onde $F^*(z_c)$ é a proporção de valores observados (amostra) abaixo do limite máximo tolerável, z_c , para o respectivo metal. Com base nesse critério, nas Figuras 5.75 a 5.78 são apresentados os mapas das estimativas quantil p com os respectivos riscos de erros α e β e as classificações das áreas contaminadas para cada variável, baseados na estimativa quantil p estar acima dos limites máximos permitidos para cada metal. O risco de erro α (ou falso positivo) é definido como a probabilidade de se declarar uma localização como perigosa (contaminada), quando na verdade ela é segura; e o risco de erro β (ou

falso negativo) é a probabilidade de se declarar uma localização como segura, quando na verdade ela é perigosa. Locais com altas concentrações apresentam menores riscos α , ou seja, a probabilidade de erro em declarar esses locais como contaminados é baixa. Assim, a distribuição espacial de probabilidades de riscos α poderia ser usada para auxiliar no delineamento de áreas perigosas para remediação baseada sobre um nível de probabilidade tolerável de delineamento incorreto. Os valores amostrais e simulados nas localizações possivelmente contaminadas por cádmio ficaram bem próximos do limite máximo permitido de 0,001 ppm, uma indicação de baixas concentrações. Isso explica porque o risco α de se declarar esses locais contaminados apresentou altas probabilidades (Figura 5.77 gráfico inferior à direita).

Os mapas de classificação das áreas contaminadas obtidos pelo uso das estimativas quantil p (Figura 5.75) e aqueles com base nas probabilidades marginais de contaminação (Figura 5.70) apresentaram forte semelhança entre si e boa aproximação com os mapas de classificação baseados na função de perda representada pelas equações (3.81) e (3.82), para os pesos relativos $w_I = 1,0$ e $w_I = 2,5$, utilizadas na minimização das perdas esperadas na avaliação da incerteza local e mostradas nas Figuras 5.49 (manganês), 5.54 (arsênio), 5.59 (ferro) e 5.64 (cádmio). Deve-se lembrar que as estimativas ótimas representadas pelos quantis relacionam-se a funções de perda do tipo assimétrica mostradas nos gráficos c e d da Figura 3.8, enquanto a função de perda constituída pelas equações (3.81) e (3.82) tem a forma mostrada na Figura 3.10.

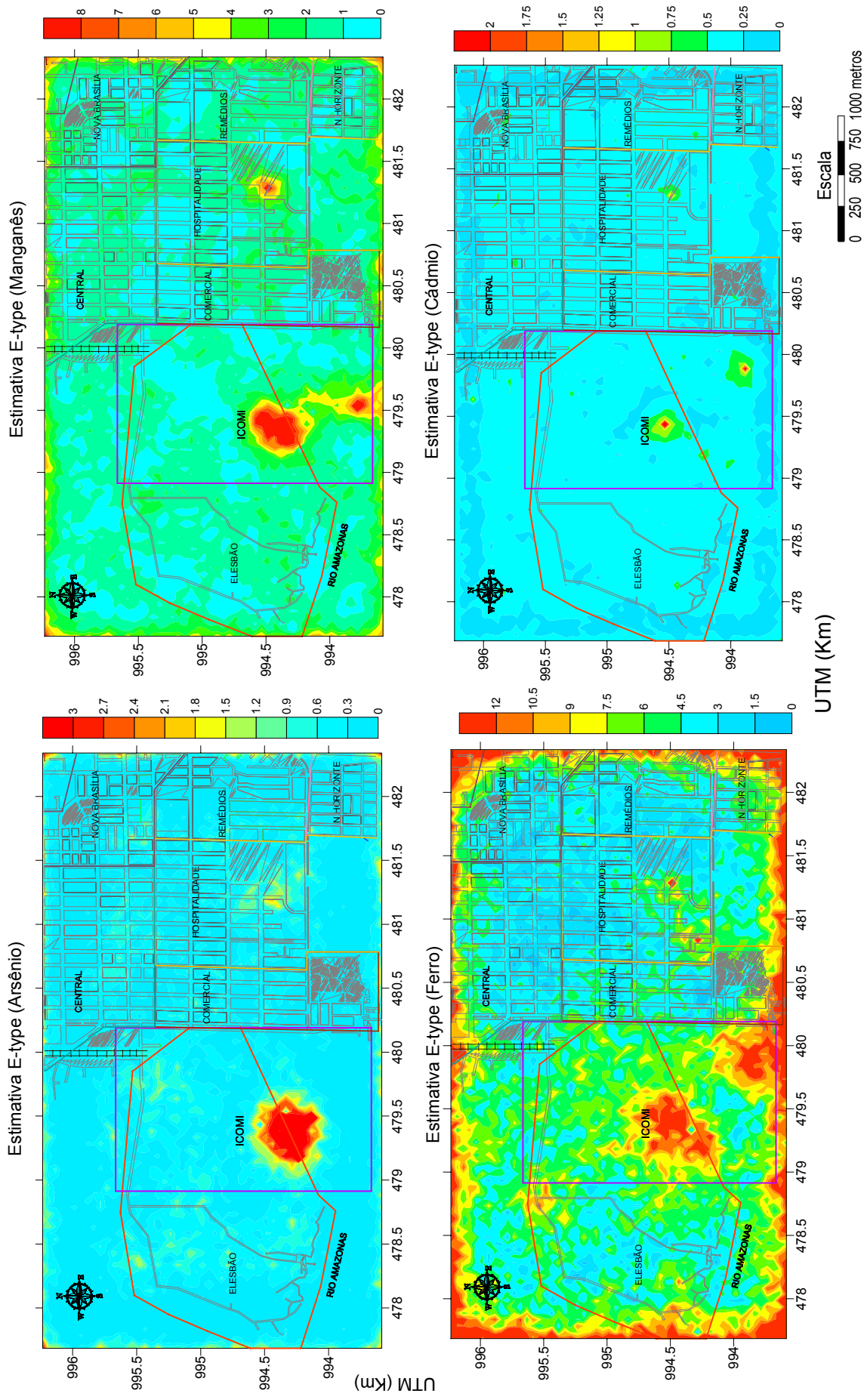


Figura 5.71 – Mapas das estimativas E-type (médias) das concentrações de As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita) derivados do pós-processamento das simulações.

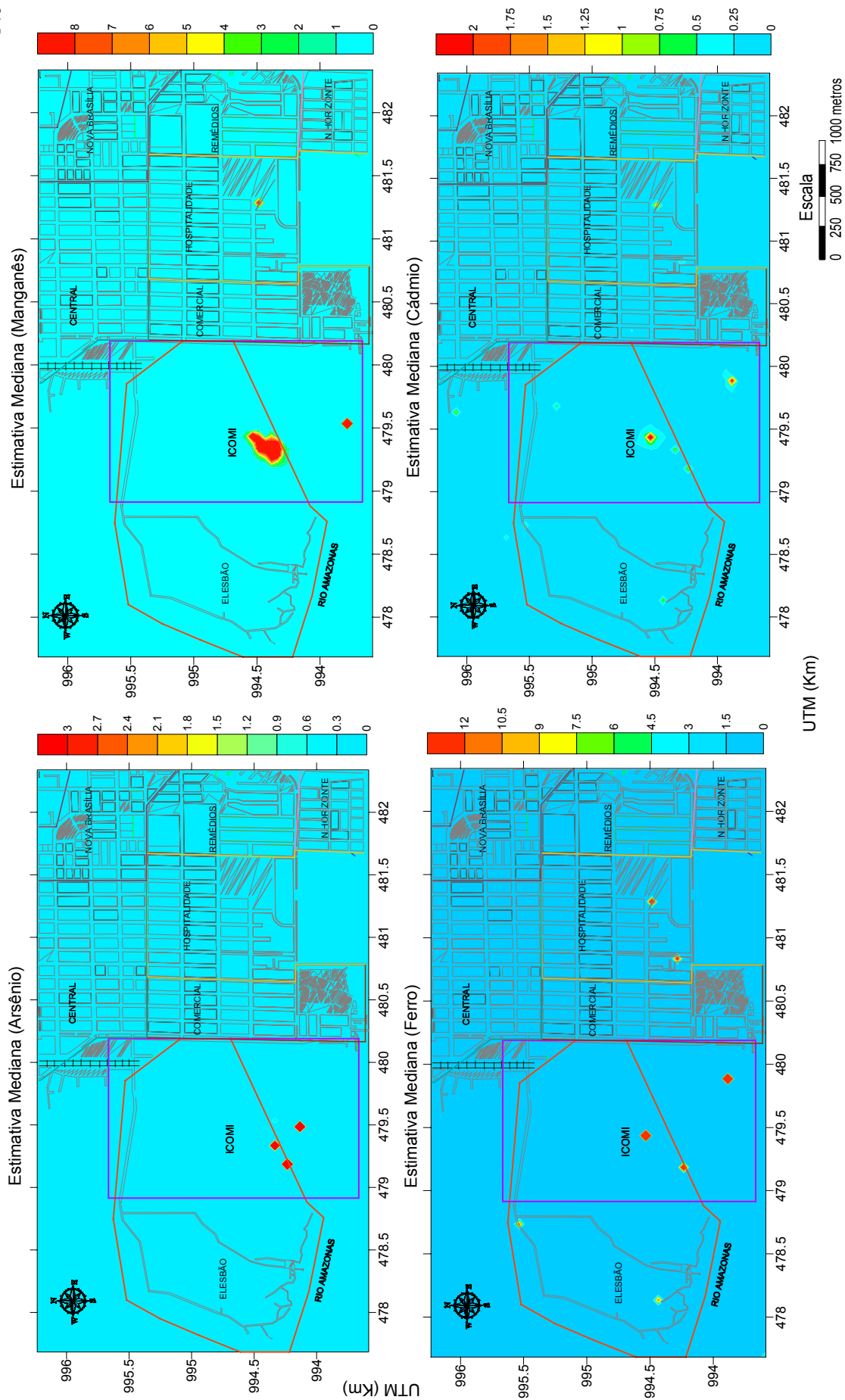


Figura 5.72 – Mapas das estimativas Médiana das concentrações de As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita), derivados do pós-processamento das simulações.

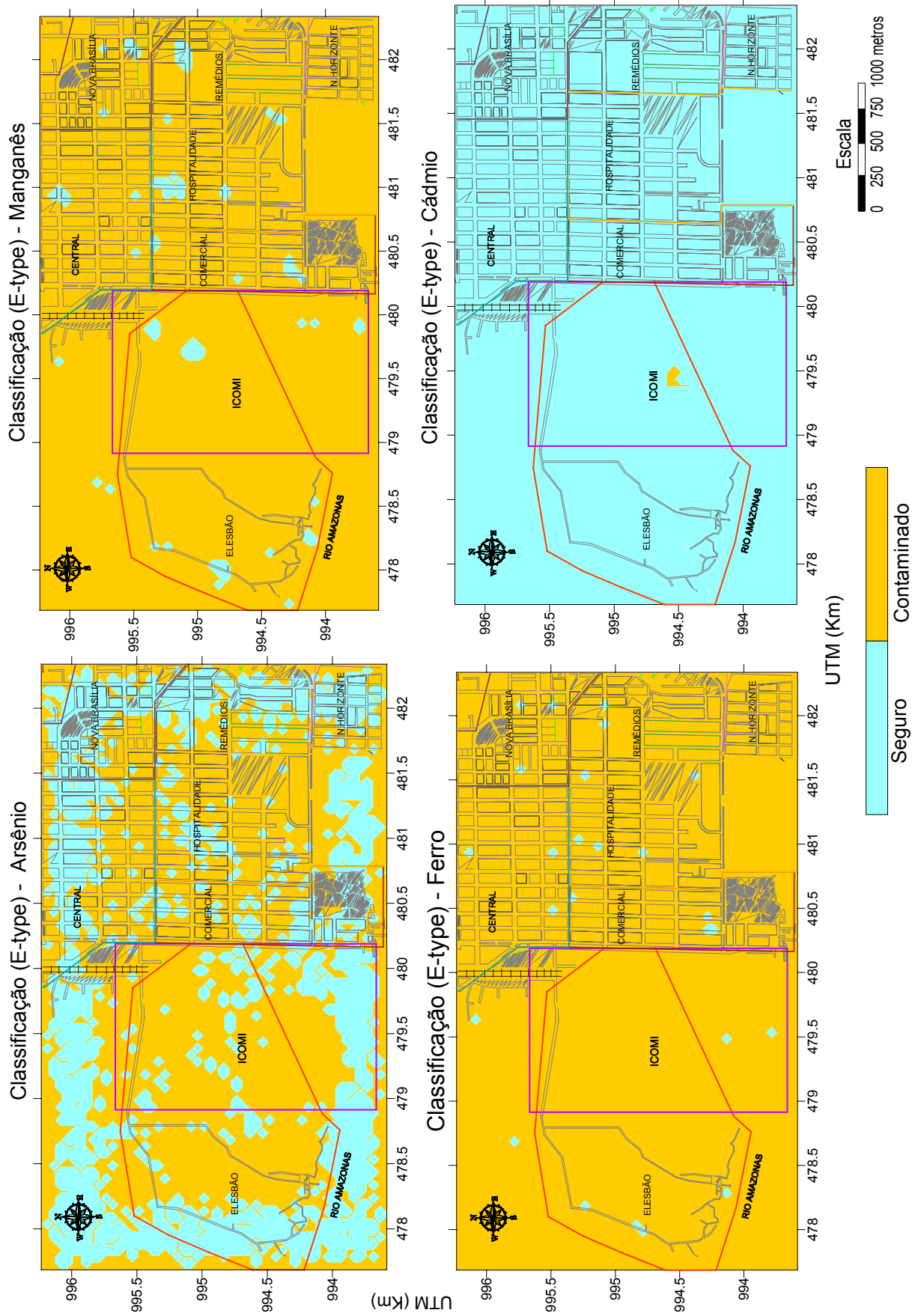


Figura 5.73 – Mapas de classificação das localizações como contaminadas por As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita), baseados na estimativa E-type estar acima dos limites máximos permitidos para cada metal.

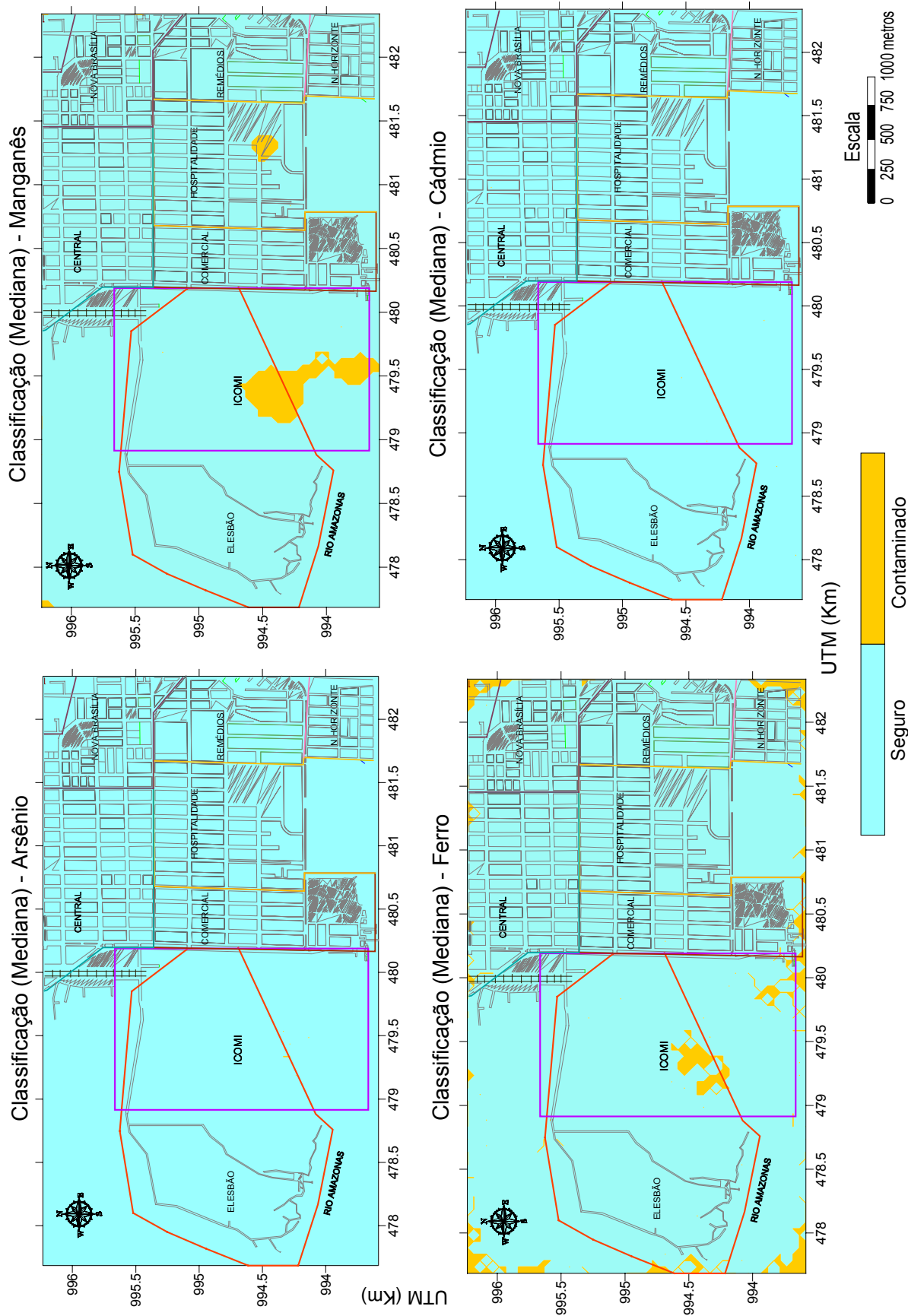


Figura 5.74 – Mapas de classificação das localizações como contaminadas por As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita) baseados na estimativa Mediana estar acima dos limites máximos permitidos para cada metal.

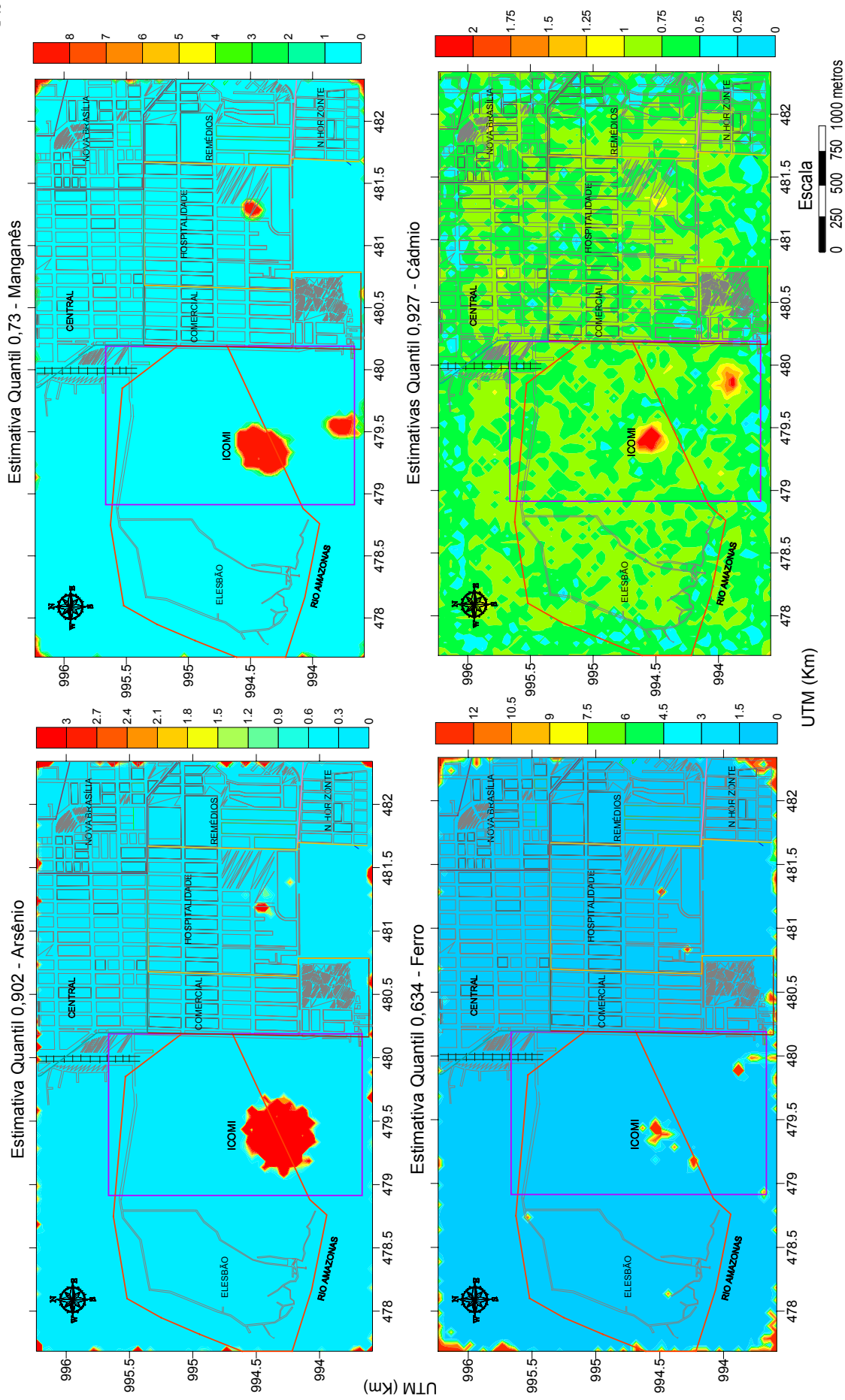
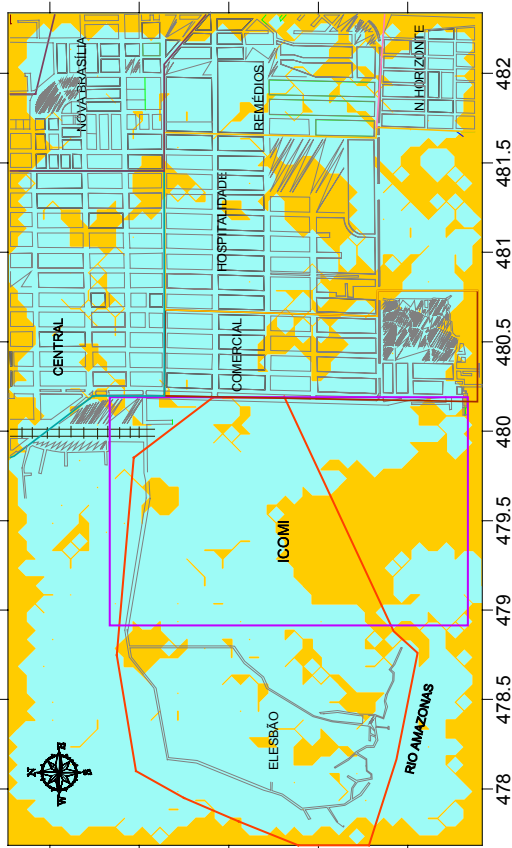
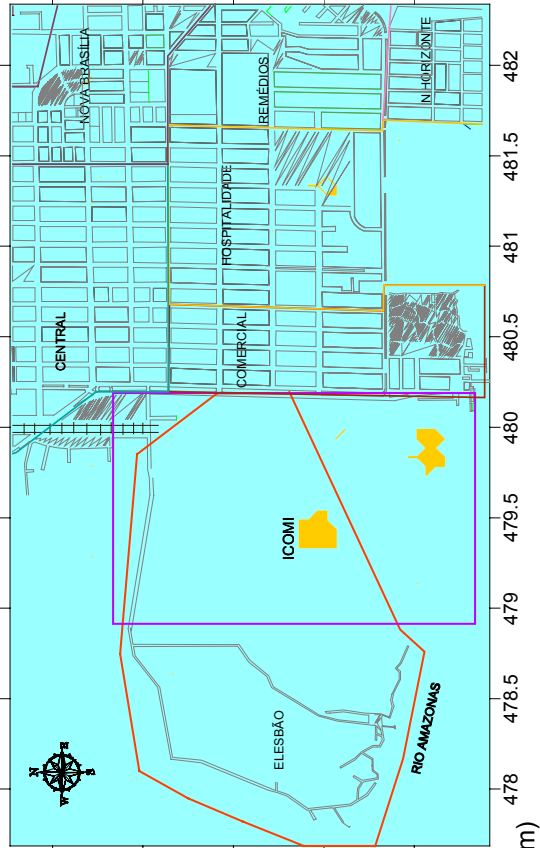


Figura 5.75 – Mapas das estimativas Quantis das concentrações de As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita), derivados do pós-processamento das simulações.

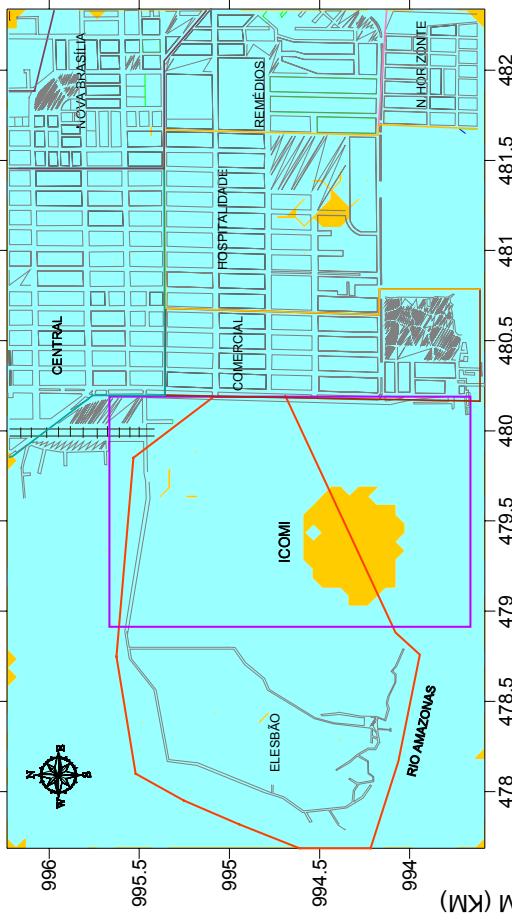
Classificação Quantil 0,73 - Manganês



Classificação Quantil 0,927 - Cádmio



Classificação Quantil 0,902 - Arsênio



Classificação Quantil 0,634 - Ferro

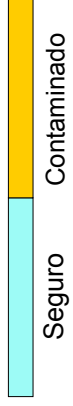
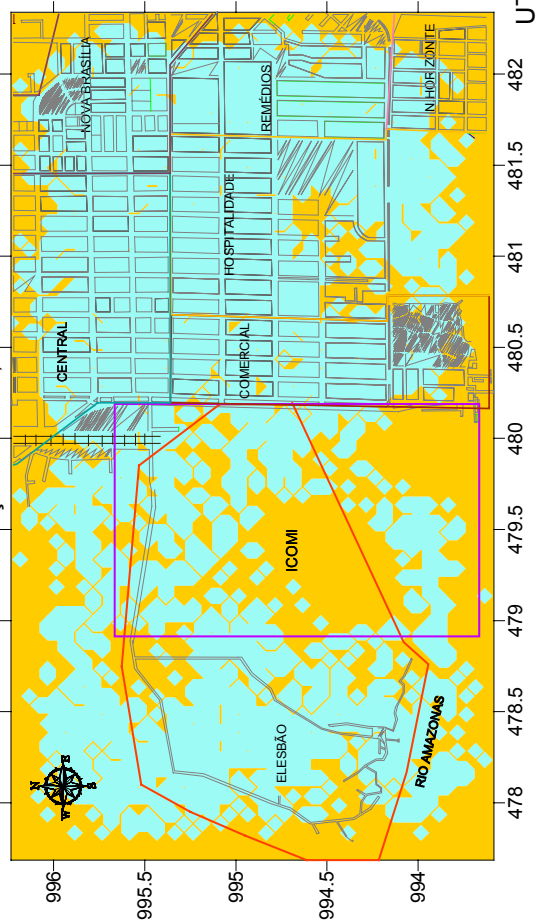


Figura 5.76— Mapas de classificação das localizações como contaminadas por As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita) baseados na estimativa Quantil acima dos limites máximos permitidos para cada metal.

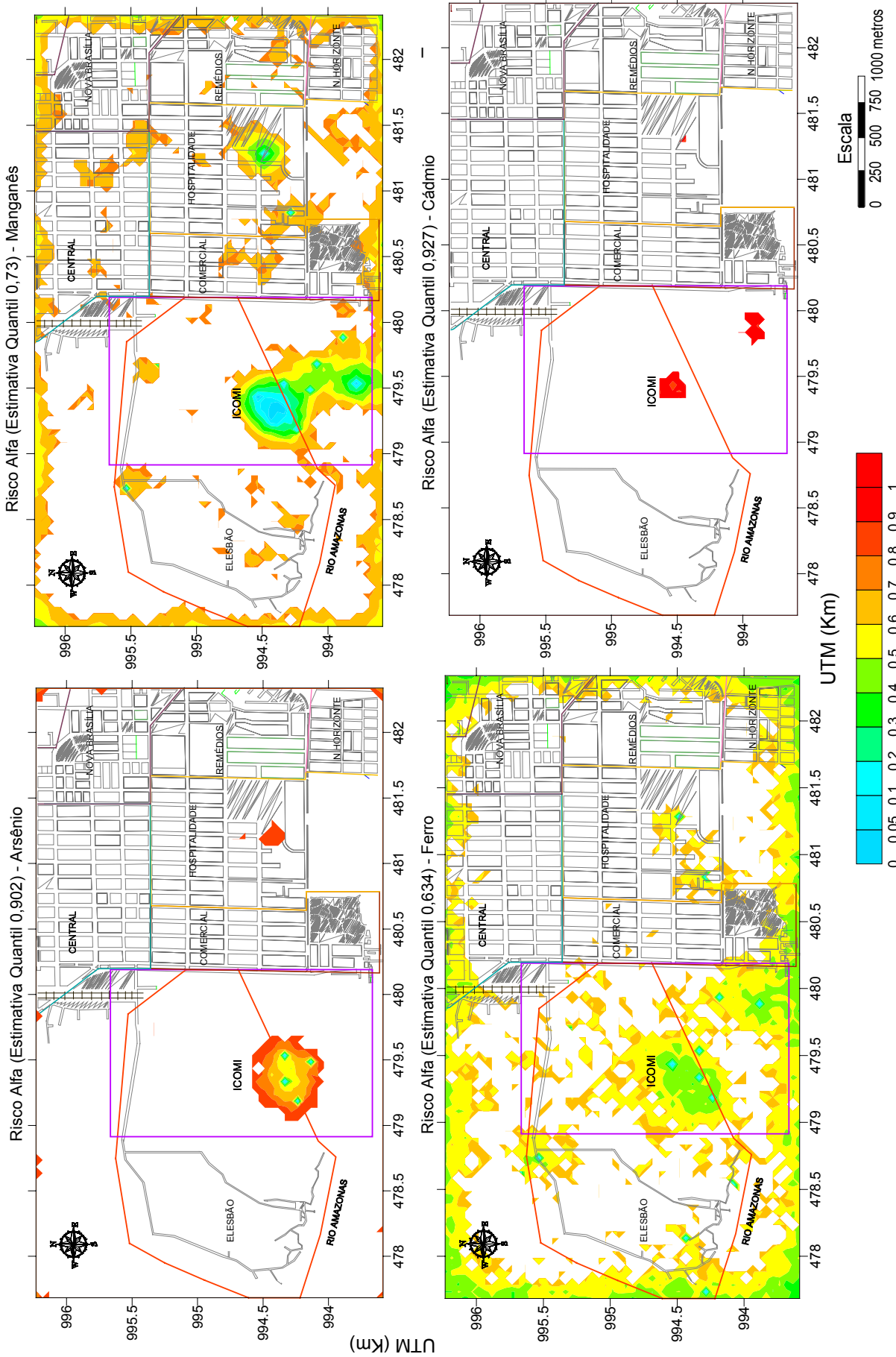


Figura 5.77 – Mapas de probabilidades de risco α (falso positivo), isto é, o risco de erro de se declarar uma localização contaminada, para As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita).

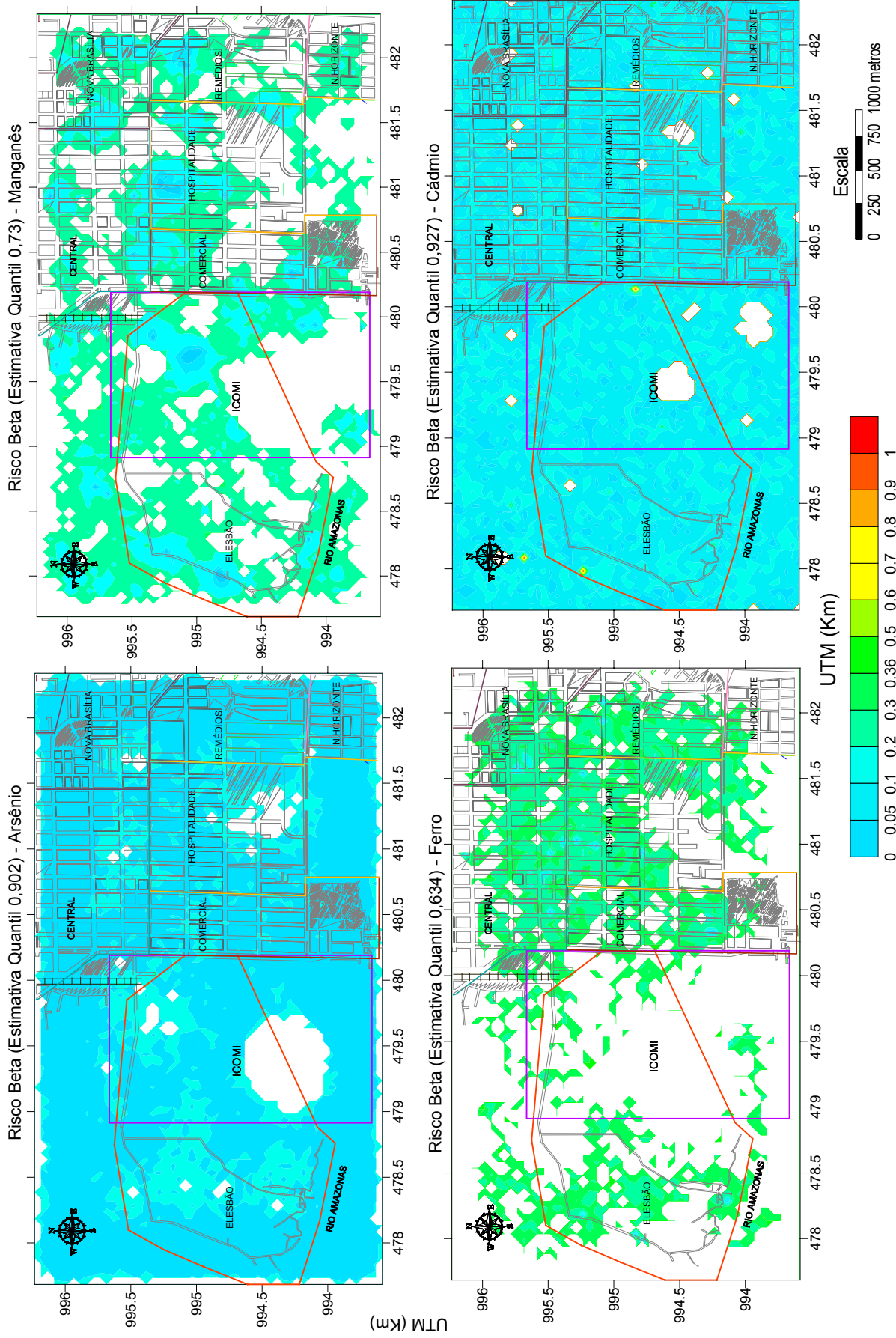


Figura 5.78– Mapas de probabilidades de risco β (falso negativo), isto é, o risco de erro de se declarar uma localização segura, para As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita).

As funções de perda podem ser utilizadas nas simulações de modo que sejam incorporados aspectos econômicos, onde a distribuição resultante das respostas permita que sejam conhecidas as incertezas sobre as conseqüências na tomada de decisão para um cenário particular. A função de perda (3.92), com $w_1(\mathbf{u}'_j)=1$, foi aplicada ao conjunto de 100 simulações geradas para cada variável e obteve-se 100 mapas alternativos de custos de saúde, que representam uma medida da incerteza sobre o impacto econômico de não se tomar nenhuma medida de remediação. O programa *postsim* da livraria GSLIB foi adaptado para a computação das funções de perda. O custo global em cada simulação é obtido como a soma dos custos locais de cada valor simulado pela utilização da equação (3.91). As simulações produzem custos globais que apresentam diferenças entre si por causa da falta de conhecimento total da distribuição espacial dos valores dos contaminantes na área de estudo. Essas diferenças refletem a incerteza sobre o impacto econômico de se declarar erradamente a localização segura. Deve-se considerar também a incerteza sobre a função de perda utilizada. A incerteza pode ser descrita pelo mapeamento da variância das distribuições dos custos de saúde em cada localização e os histogramas resultantes dos custos globais permitem uma avaliação probabilística das conseqüências financeiras para o cenário estudado. Nas Figuras 5.79 a 5.83 são apresentados os resultados obtidos.

As médias dos custos de saúde em cada localização obtidas pela aplicação da função de perda (3.92) ao conjunto de simulações para cada metal são mostradas na Figura 5.79. Os custos médios são maiores nas localizações que apresentam concentrações mais altas para todos os metais, e localizam-se principalmente dentro da área industrial da ICOMI onde provavelmente se encontra a principal fonte de contaminação. Observa-se um aumento nos custos médios em pequenas localizações no bairro da hospitalidade onde ocorreu o aterro de ruas com rejeitos de manganês. A Figura 4.80 mostra o mapa das variâncias onde a maior incerteza sobre os custos de saúde é observada nos locais de maiores concentrações. Maiores incertezas nos custos de saúde são verificadas para o ferro, em razão da grande variabilidade apresentada nas concentrações desse metal. Quanto maior a incerteza maior será o custo. Por isso os custos de saúde também foram bastante elevados para o ferro, uma vez que o custo de classificação errada é proporcional à contaminação $[z^{(l)}(\mathbf{u}'_j) - z_c]$. As Figuras 5.81 e 5.82 mostram as simulações que produziram o custo máximo (pior cenário) e o custo mínimo com a aplicação da função de perda (3.92). A simulação n.º 46 produziu o pior

cenário para arsênio com um custo máximo, C_{\max} , de 2336,83; e o custo mínimo, $C_{\min} = 1448,20$, foi obtido na simulação n.º 51 (Figura 5.81). A Tabela 5.8 mostra as estatísticas dos custos globais de saúde para os outros metais. Pode-se verificar que os maiores custos ocorreram para o ferro, seguido pelo manganês, arsênio e cádmio. Os histogramas dos custos globais (em valores relativos) mostrados na Figura 5.83, com as respectivas estatísticas na Tabela 5.8, permitem uma avaliação probabilística das conseqüências financeiras de se declarar erradamente a área de estudo segura em relação à contaminação. Por exemplo, há uma probabilidade acima de 75% (3º Quartil) de que o custo de saúde seja maior do que 2022,18 para o arsênio, 8139,00 para o manganês, 2948,77 para o ferro e 385,97 para o cádmio.

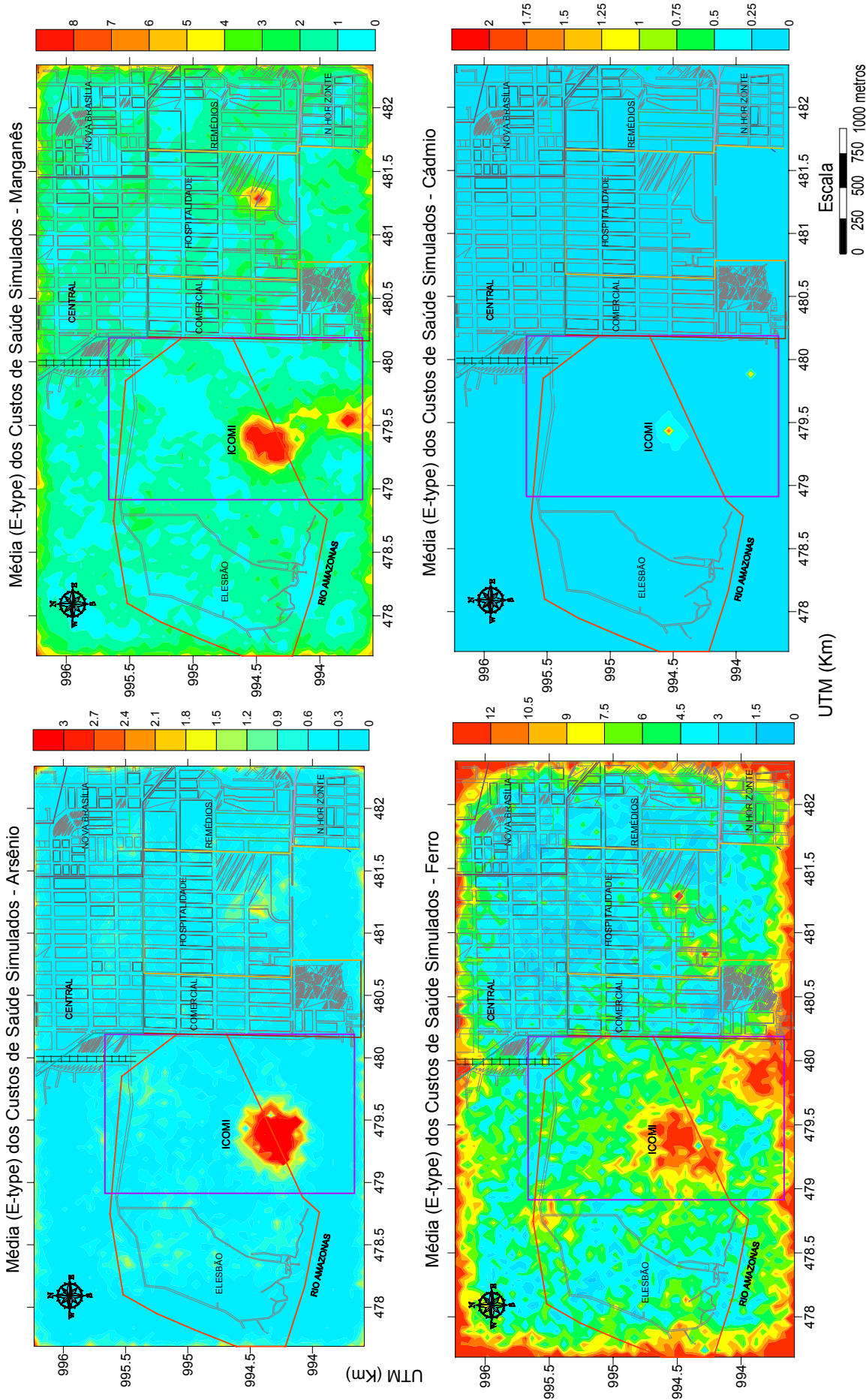


Figura 5.79 – Mapas da média da distribuição de custos de saúde obtidos pela aplicação da função de perda (3.92) aos valores simulados para As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita).

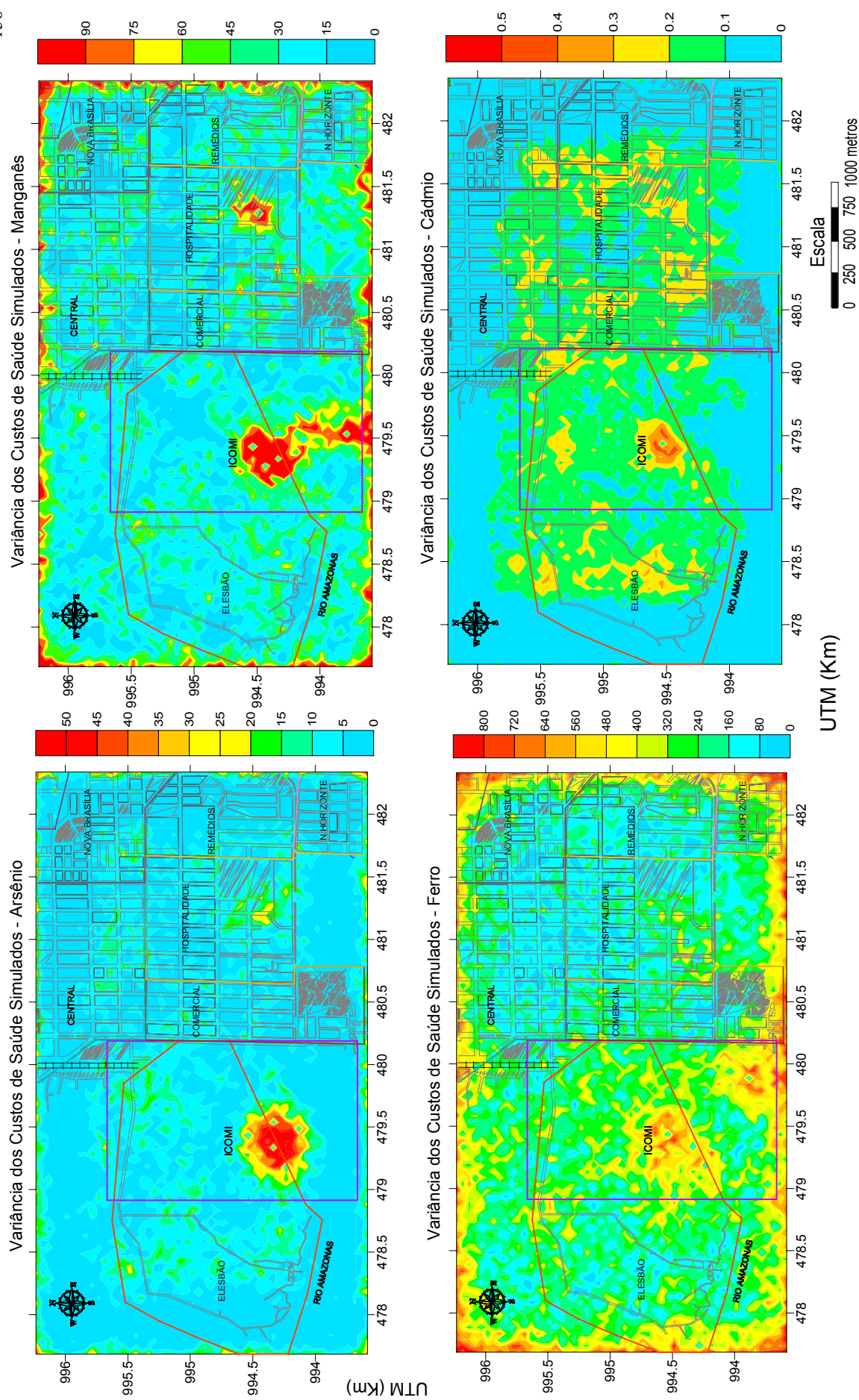


Figura 5.80 – Mapas da variância das distribuições de custos de saúde obtidos pela aplicação da função de perda (3.92) aos valores simulados para As (acima à esquerda), Mn (acima à direita), Fe (abaixo à esquerda) e Cd (abaixo à direita).

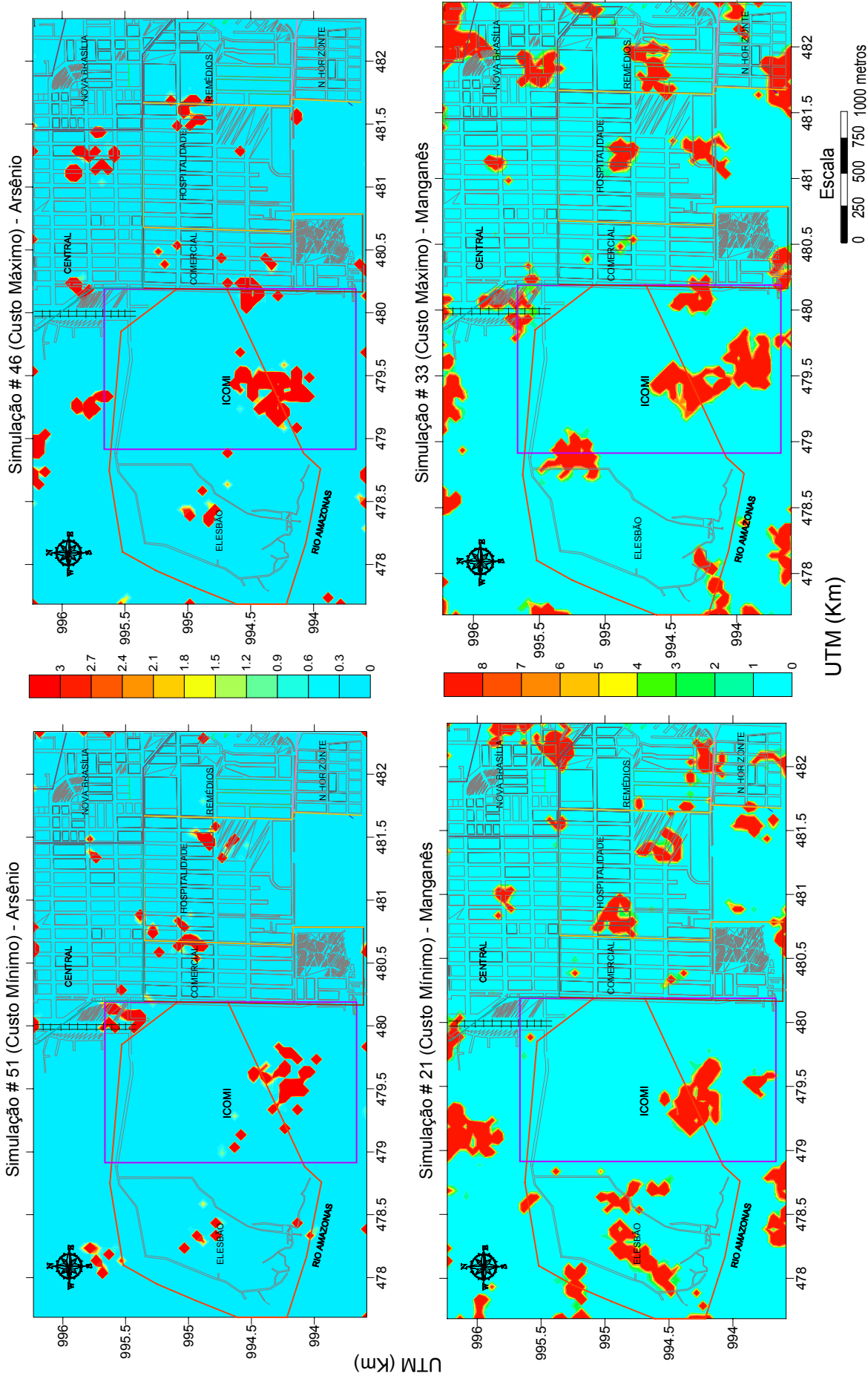
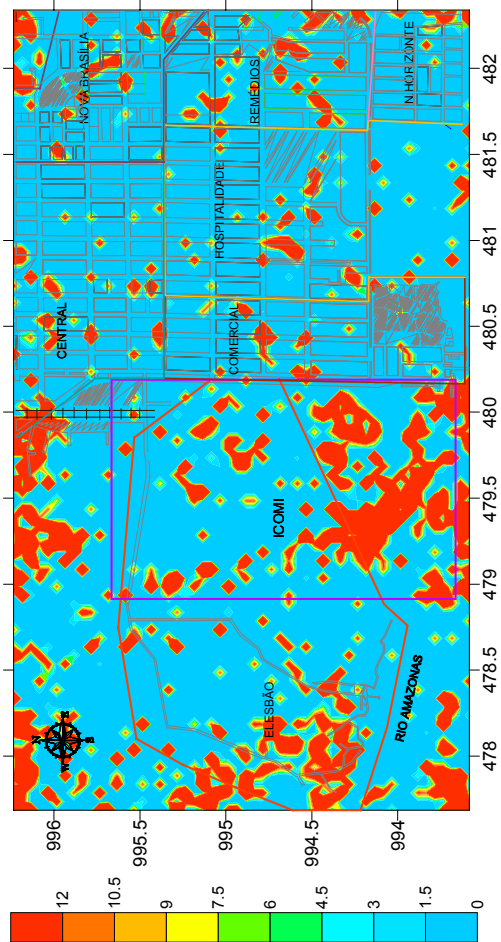
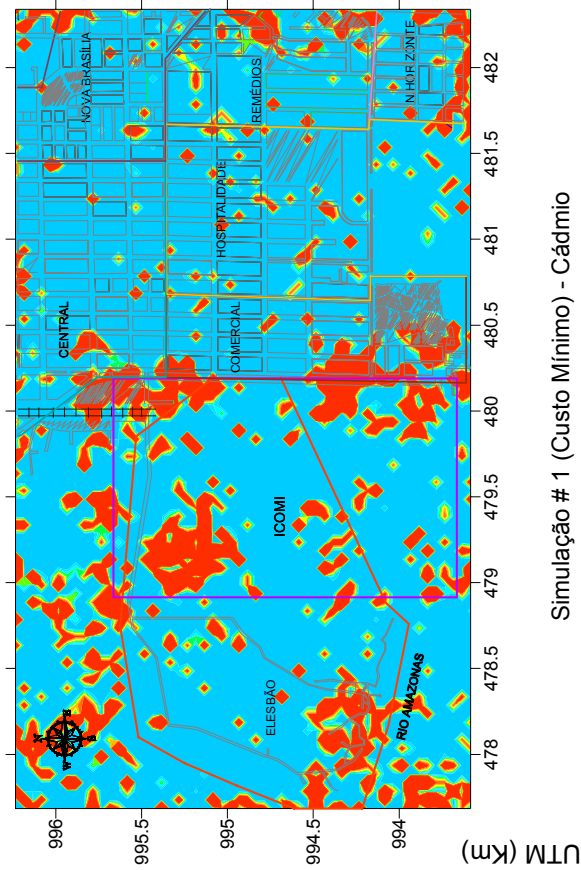


Figura 5.81 – Mapas das simulações que produziram o custo mínimo e máximo na aplicação da função de perda (3.92) aos valores simulados para o As (acima) e Mn (abaixo).

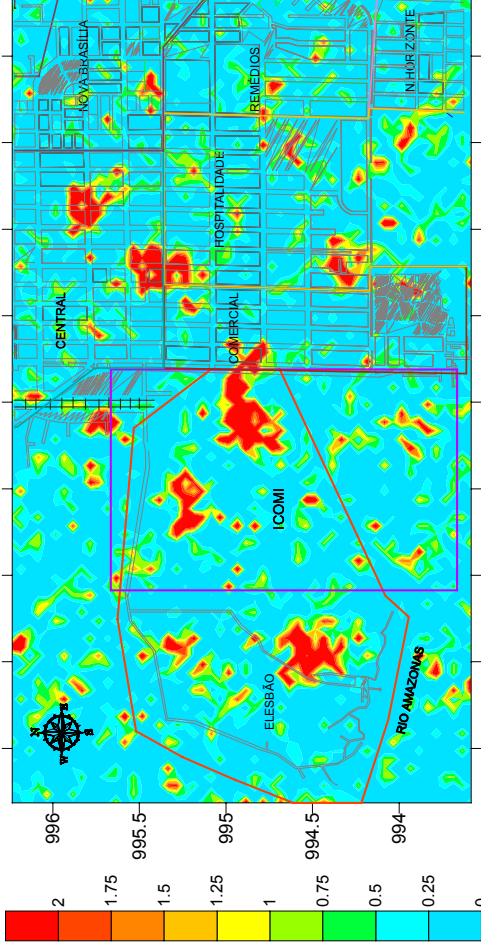
Simulação # 26 (Custo Máximo) - Ferro



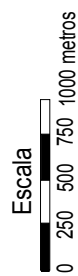
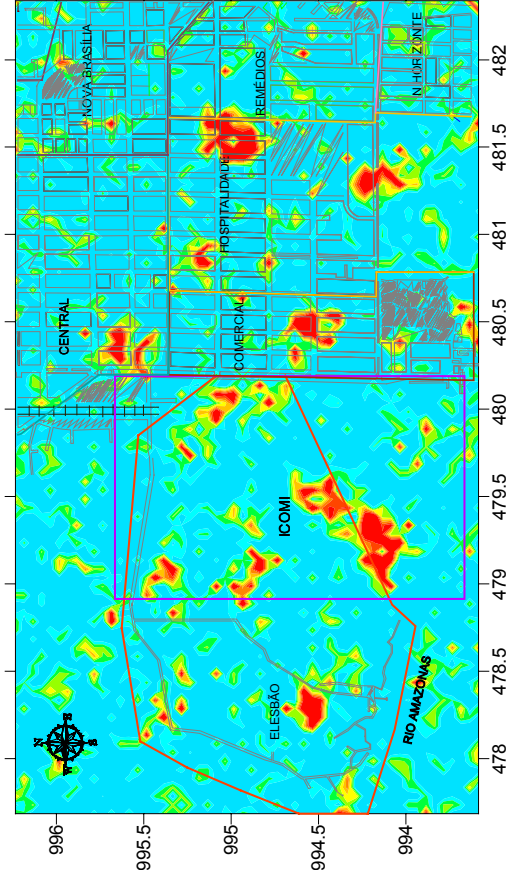
Simulação # 83 (Custo Mínimo) - Ferro



Simulação # 11 (Custo Máximo) - Cádmiu



Simulação # 1 (Custo Mínimo) - Cádmiu



UTM (Km)

Figura 5.82 – Mapas das simulações que produziram o custo mínimo e máximo na aplicação da função de perda (3.92) aos valores simulados para o Fe (acima) e Cd (abaixo).

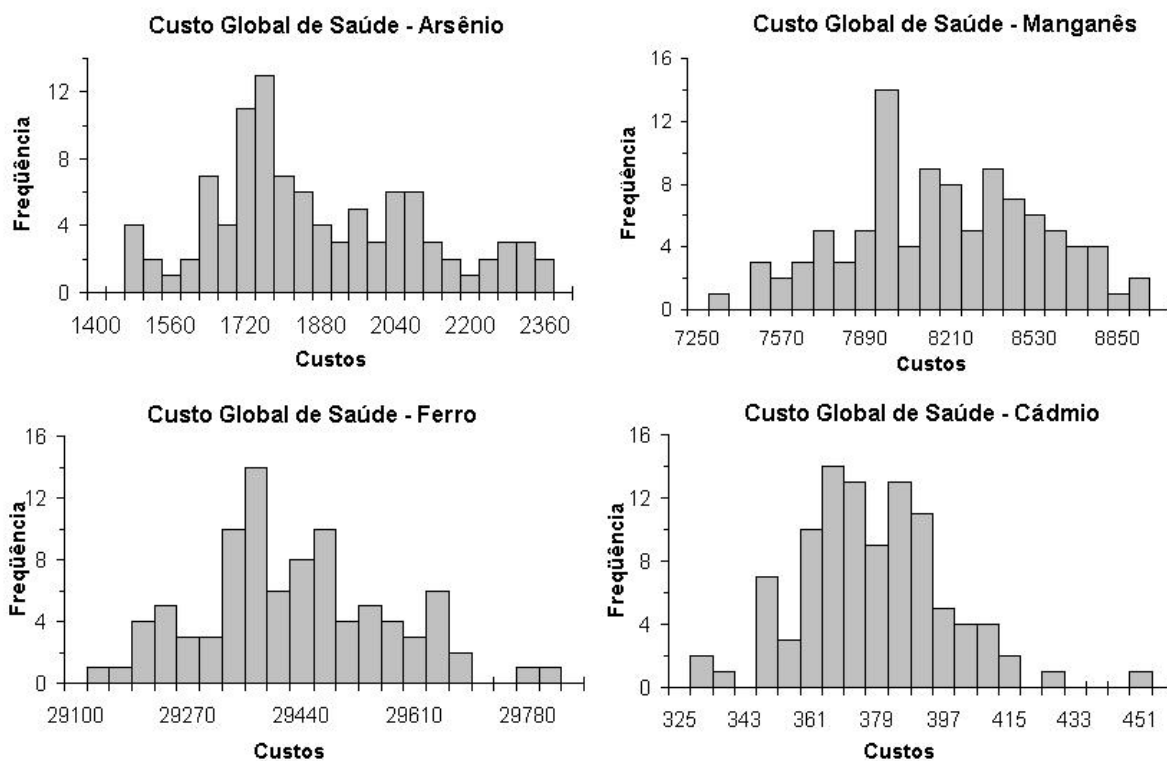


Figura 5.83 – Distribuições dos custos globais do risco de se declarar a área segura em relação à contaminação, obtidas pela aplicação da função de perda (3.92) às simulações do arsênio, manganês, ferro e cádmio.

TABELA 5.8 – Estatísticas obtidas para os custos globais de saúde (valores relativos) obtidos pela aplicação da função de perda (3.92) às simulações.

Custos	Arsênio	Manganês	Ferro	Cádmi
Média	1850,93	7868,49	29410,82	374,79
Mínimo	1448,20	7008,55	29124,94	327,85
1º Quartil	1712,24	7634,23	29318,87	362,25
Mediana	1792,24	7861,41	29394,75	374,12
3º Quartil	2022,18	8139,00	29487,77	385,97
Máximo	2336,83	8606,37	29786,27	447,26

O comportamento da função do custo total (equação 3.88) permite uma avaliação das conseqüências sobre as decisões tomadas na definição do plano de remediação. A Figura 5.84 mostra a relação entre os prováveis custos de tratamento e custos esperados de falha com as mudanças no nível de probabilidade. Pode-se observar que o risco ótimo para todos os metais fica próximo da probabilidade marginal, ou seja, 0,098 para o arsênio, 0,27 para o manganês, 0,366 para o ferro e 0,073 para o cádmio. Embora o risco ótimo para cada contaminante tenha sido aproximadamente igual à probabilidade marginal, não significa que necessariamente se deva utilizar esse nível de risco no tratamento das áreas contaminadas. O custo esperado total mínimo associado a esse risco beneficia o operador (responsável pelo tratamento). Pode ser que esses níveis de probabilidade sejam altos do ponto de vista da outra parte envolvida na questão, isto é, agência governamental e comunidade. O modelo de decisão aponta, nesse caso, para uma negociação entre as partes interessadas, de modo que possam ser avaliadas conseqüências relacionadas a fatores como saúde humana, ambiental, entre outros. Um procedimento mais conservativo que pode indicar mais contaminação do que realmente ocorre, poderia reduzir a probabilidade de falha, mas a um custo maior por causa do tratamento e remoção desnecessária do solo não contaminado. Por outro lado, um procedimento menos conservativo pode sugerir menos custo inicial de tratamento, mas deixa o operador exposto a grande responsabilidade potencial se resíduos de contaminação forem detectados mais tarde.

Os mapas de custos por parcelas são apresentados nas Figuras 5.85 e 5.86 para os níveis de probabilidades que apresentaram o custo mínimo em cada respectivo nível de corte para cada metal, ou seja, para o arsênio com nível de corte de 0,05 ppm e nível de probabilidade de 0,098; manganês com nível de corte de 0,1 ppm e nível de probabilidade de 0,27; ferro com nível de corte de 0,3 ppm e nível de probabilidade de 0,366; e cádmio com nível de corte de 0,001 ppm e nível de probabilidade de 0,073. A caracterização, tratamento e custos de falha variam espacialmente e os mapas desses custos podem fornecer informação útil em relação a levantamentos amostrais adicionais e à caracterização da área. O custo de tratamento é presumido constante para todos os contaminantes ao valor de 2,5 u.m (unidades monetárias) por parcela sendo que o custo de tratamento total *a priori* é equivalente a um valor de 12960 u.m (5076 painéis * 2,5). As áreas mais escuras nos mapas de custo de tratamento correspondem ao custo zero, ou seja, sem tratamento. Nos mapas de custos esperados de falha $E\{C_{falha}\}$, as áreas em vermelho são aquelas associadas com custos mais altos.

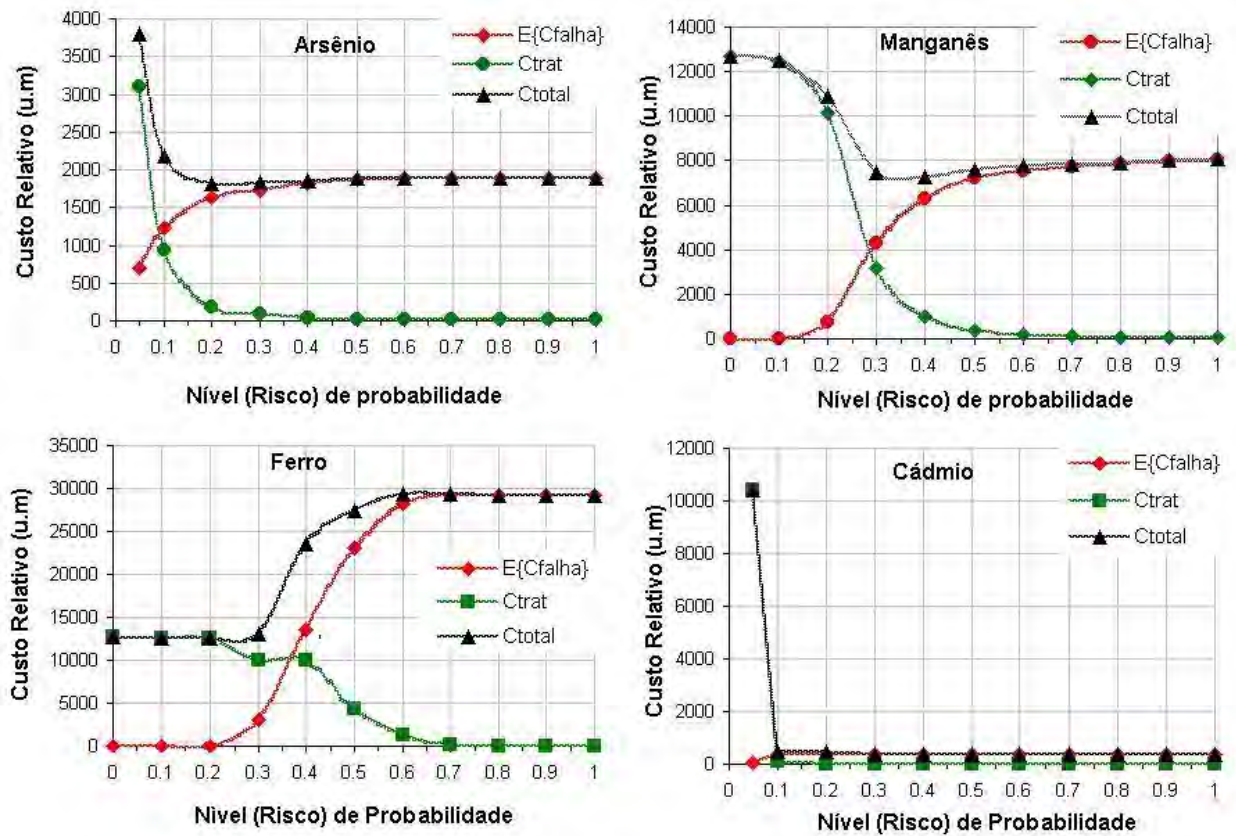


Figura 5.84 – Curvas individuais de custos para o arsênio com nível de corte de 0,05 ppm (acima à esquerda), manganês com nível de corte de 0,1 ppm (acima à direita), ferro com nível de corte de 0,3 ppm (abaixo à esquerda) e cádmio com nível de corte de 0,001 ppm (abaixo à direita).

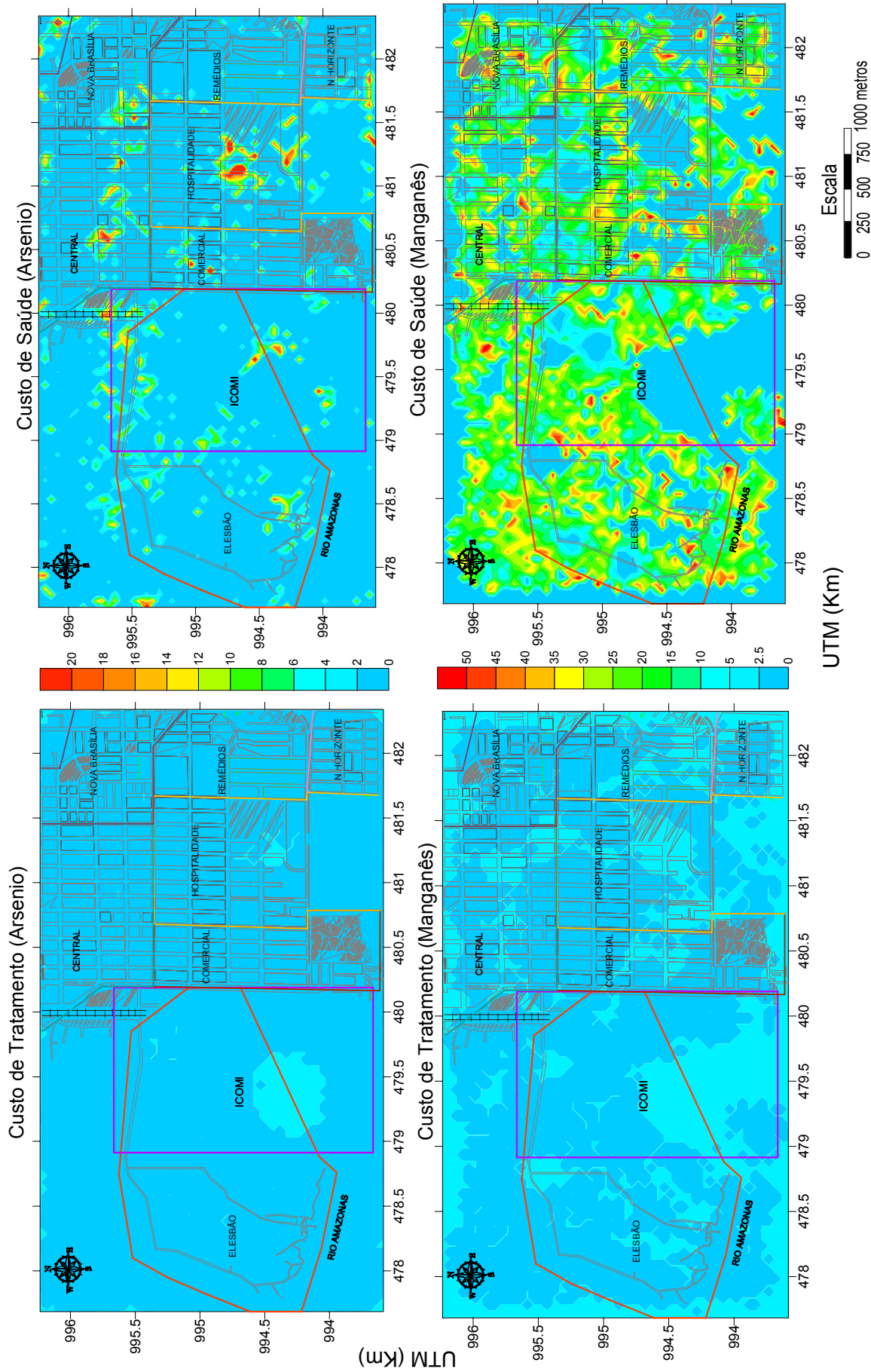


Figura 5.85 – Mapas da distribuição espacial dos custos por parcela do As (*Crat* acima à esquerda e *E{Cfalha}* acima à direita) para o nível de probabilidade de 0,098 e nível de corte de 0,05 ppm e do Mn (*Crat* abaixo à esquerda e *E{Cfalha}* abaixo à direita) para o nível de probabilidade de 0,28 e nível de corte de 0,1 ppm.

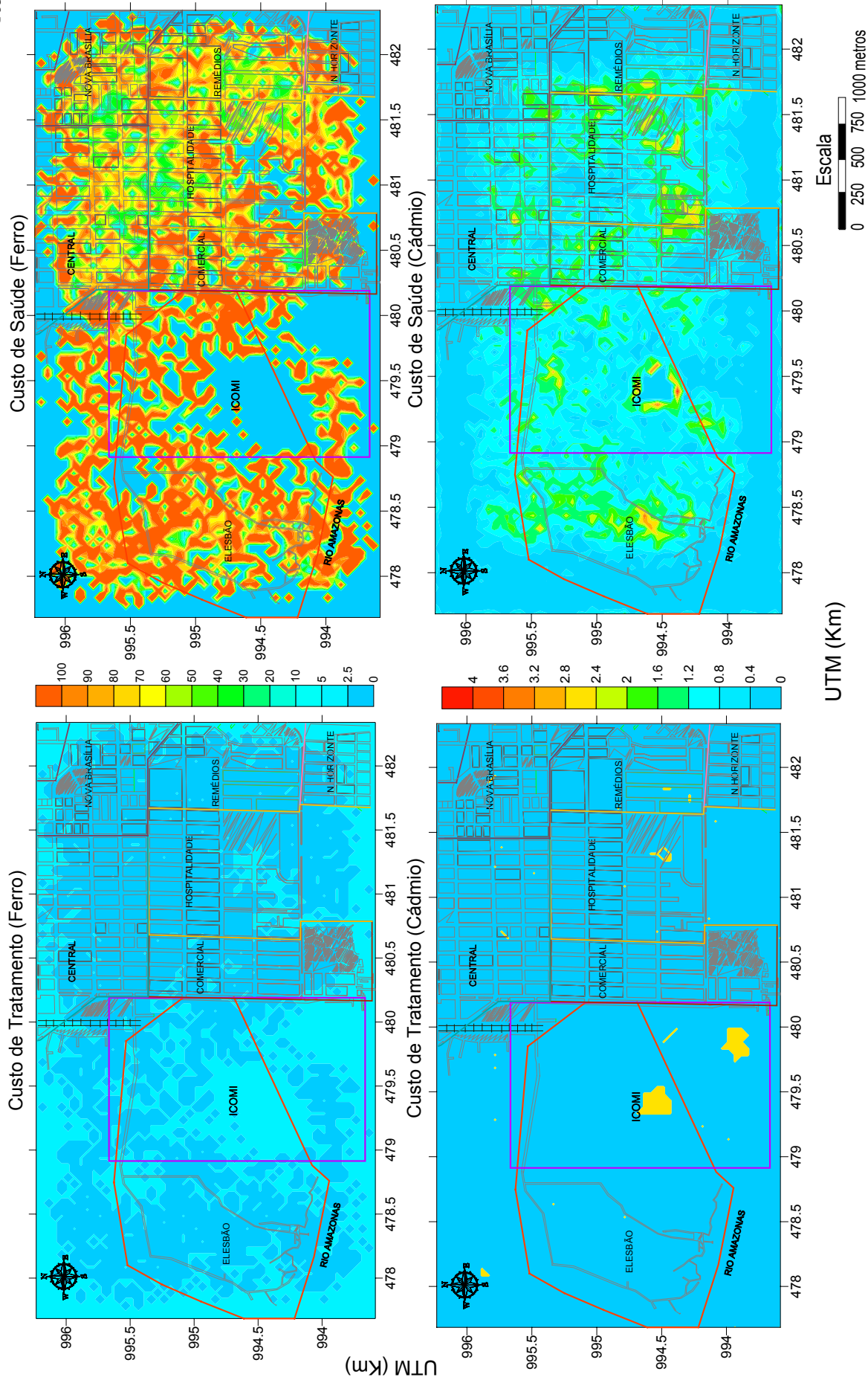


Figura 5.86 – Mapas da distribuição espacial dos custos por parcela do Fe (*Ctrat* acima à esquerda e *E{Cfalha}* acima à direita) para o nível de probabilidade de 0,37 e nível de corte de 0,3 ppm e do Cd (*Ctrat* abaixo à esquerda e *E{Cfalha}* abaixo à direita) para o nível de probabilidade de 0,07 e nível de corte de 0,001 ppm.

6. CONSIDERAÇÕES FINAIS

Mesmo com relativamente poucos dados, a metodologia empregada, baseada na teoria geoestatística, mostrou que pode contribuir eficientemente na avaliação ambiental de áreas com suspeitas de contaminação. A metodologia permite avaliar erros e incertezas em vários contextos como: erros de estimativas locais, erros de estimativas globais, erros de classificações local e global, probabilidades de um determinado limite de concentração ser ultrapassado e probabilidades de falso positivo e falso negativo. Em cada um desses procedimentos são associados perdas e custos. A utilização de estimativas baseadas em diferentes critérios de otimização, por exemplo, que não penalizam igualmente a subestimativa e superestimativa, conduz a uma classificação mais realista das áreas com suspeita de contaminação, além de minimizar os erros de classificação, sobretudo na presença de *outliers* e forte assimetria. Desta forma, a classificação de áreas contaminadas baseada em estimativas do tipo média e mediana mostrou-se inadequada, enquanto a classificação com base em estimativas do tipo quantil apresentou melhores resultados. A utilização de funções de perda associadas a custos econômicos mostrou resultados semelhantes em relação à classificação de áreas contaminadas obtidas pelas estimativas do tipo quantil e por aquelas baseadas nas probabilidades das concentrações estarem acima de determinado limite de corte.

A caracterização do local é fundamental na avaliação do risco. Nesse sentido a utilização do método multivariado geoestatístico assume um papel importante, pois auxilia na identificação das origens e comportamento dos contaminantes. No presente estudo a caracterização poderia, certamente, ser bastante enriquecida se houvesse a disponibilidade de informações adicionais como a geologia da área e análises das concentrações dos contaminantes no solo em níveis diferentes de profundidade. Mesmo assim, as atividades relacionadas a exploração e comercialização de manganês em Santana mostraram sinais de terem contribuído para a ocorrência das altas concentrações de manganês, arsênio, ferro e cádmio observadas na área de estudo. As altas concentrações de alumínio, selênio, chumbo e cobre parecem estar mais relacionadas ao próprio meio natural.

A utilização da simulação *annealing* (condicional) na avaliação da distribuição espacial das variáveis apresentou efeitos indesejáveis, denominados efeitos de borda, sobretudo na variável ferro, que apresentou maior dispersão (absoluta) e maior efeito proporcional entre os metais analisados. Uma possível relação entre esses efeitos e as

características apresentadas pela variável ferro necessita de mais estudos. As características, como a forte assimetria e poucos dados, apresentadas pelas distribuições das amostras dos metais analisados apontaram para a utilização da simulação *annealing*, entretanto, outros tipos de simulações estocásticas deveriam ser testados.

Embora sejam fortes as evidências de contaminação por metais pesados, sobretudo manganês, arsênio e ferro na área em estudo, os resultados obtidos podem servir como uma aproximação preliminar na questão da avaliação do risco de contaminação pelos referidos metais, uma vez que, devido a pouca quantidade de informações disponíveis, os níveis de incerteza mostrados nas medidas de dispersão foram bastante elevados. A redução dos níveis de incerteza através de um investimento em amostragens adicionais poderia reduzir os custos de remediação, uma vez que esses custos geralmente são maiores do que os custos de amostragem.

O modelo de decisão econômico baseado em riscos pode ser utilizado na elaboração de planos efetivos de custos para a remediação ambiental. O modelo envolve a minimização do custo total esperado da caracterização do local, da remediação (tratamento) e custos potenciais de falhas associados com saúde, segurança e responsabilidades legais. O risco ótimo, associado ao menor custo total, apresentou valores próximos da probabilidade marginal de cada metal, ou seja, da proporção das concentrações acima do limite máximo estabelecido pela legislação (CONAMA), obtida nas amostras. Isso mostra a importância de que a amostragem seja representativa da população em estudo. Entretanto, o risco ótimo do ponto de vista do operador (responsável pelo tratamento) pode não estar associado a um risco aceitável estabelecido pelas outras partes interessadas (agências governamentais e comunidade). Nesse caso, as penalidades poderiam ser aumentadas em caso de falhas na remediação (tratamento) da área contaminada, como uma forma de compensação entre os riscos associados com o custo mínimo e o risco aceitável.

Como a questão envolvendo os indícios de contaminação na área de Santana é complexa, uma vez que envolve diversas áreas de interesse, inclusive políticas, econômicas e sociais, a realização de estudos geológicos, hidrogeológicos, químicos, ecológicos e epidemiológicos, entre outros, são importantes com vistas a uma planificação e controle do problema. A incorporação da incerteza, inerente a qualquer processo natural relacionado à maioria das áreas de estudo acima, deve sempre ser considerada, pois permite maior controle na tomada de decisão.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- AMAPÁ, Um norte para o Brasil / org. Nilson Moulin. São Paulo. Ed. CORTEZ, 2000.
- ASSAD,M.L.L; FUKS,S.D;FARIA,C.D. Mapeamento de ocorrências de termiteiros por krigagem indicatriz. **Cadernos de Informações Georeferenciadas – CIG – Vol.1(3)**. Unicamp SP, 1999. Disp. em www.cpa.unicamp.br/revista/cigv1n3a2.html.
- ATSDR – Agency for Toxic Substances and Disease Registry. **Nickel**, Report n^o. 7440-02-0, sept. 1997. Disponível em www.atsdr.cdc.gov/toxfaq.html.
- ATSDR – Agency for Toxic Substances and Disease Registry. **Cadmium**, Report n^o. 7440-73-9, jun, 1999. Disponível em www.atsdr.cdc.gov/toxfaq.html.
- ATSDR – Agency for Toxic Substances and Disease Registry. **Arsenic**, Report n^o. 7440-38-2, jul, 2001. Disponível em www.atsdr.cdc.gov/toxfaq.html.
- BRILHANTE, O.M. E CALDAS,L.Q.A. **Gestão e Avaliação de Risco Ambiental**. Ed. FIOCRUZ, Rio de Janeiro, 1999. 155p.
- CAERS, JEF. Geostatistics: from pattern recognition to pattern reproduction. **Stanford University. Petroleum Engineering, Stanford, CA 94305-2220, USA. 2001.**
- CARLE,S.F. Implementation Schemes for Avoiding Artifact Discontinuities in Simulated Annealing. **Mathematical Geology**, 29(2):231-244. 1997.
- CARLON, C.; CRITTO, A.; MARCOMINI, A.; NATHANAIL, P. Risk based characterisation of contaminated industrial site using multivariate and geostatistical tools. **Environmental Pollution** 111, 417-427. 2001. Disp. www.elsevier.com/locate/envpol.
- CHAGAS, M.A. **Conservação do cerrado Amapá: contribuições ao debate**. Secretaria do Estado de Meio Ambiente, Ciência e Tecnologia. Macapá, 1998. 30 p.
- CHANG, T. K.;SHYU,G.S;LIN,Y.P;CHANG,N.C.Geostatistical analysis of soil arsenic content in Taiwan. **Journal Environmental Science Health A34(7):1485-1501. 1999.**
- CHRISTAKOS, G. On the problem of permissible covariance and variogram models. **Water Resources Research**, 20(2):251-265. 1984.
- CONAMA – CONSELHO NACIONAL DE MEIO AMBIENTE – **Resolução n^o. 20 de 18 de junho de 1986**. (D.O.U. de 30/07/86. pag. 11.356). Brasília.
- CROMER,M.V. Geostatistics for Environmental and Geotechnical Applications : A Technology Transferred. **Geostatistical for Environmental and Geotechnical Applications**, ASTM STP 1283. Eds. American Society for Testing and Materials, 3-12. 1996. 280 p.
- CROMER, M.V.; RAUTMAN, C. A.; ZELINSK, W. P. Geostatistical Simulation of Rock Quality Designation (RQD) to Support facilities design at Yucca Mountain,

Nevada. **Geostatistical for Environmental and Geotechnical Applications**. ASTM STP 1283. Eds. American Society for Testing and Materials, 1996. 218-235. 280 p.

D'AMATO, C.; TORRES, J.P.M.; MALM, O. DDT (Dicloro Difênil Tricloroetano): Toxicidade e Contaminação Ambiental—Uma Revisão. **Química Nova**, 25(6):995-1002. 2002.

DEUSTCH, C.V. **Geostatistical Reservoir Modeling**. Oxford University Press. New York, 2002. 376 p.

DEUSTCH, C.; COCKERHAM, P., Practical considerations in the application of simulated annealing to stochastic simulation. **Mathematical Geology**, 26(1):67-82. 1994

DEUSTCH, C. V.; JOURNEL, A. G. **GSLIB: Geostatistical Software Library and Uses's Guide**. Oxford University Press, New York, 1998. 369 p.

DNPM – Departamento Nacional da Produção Mineral – **Projeto RADAM**. Vol. 6. Folha NA/B8.22 – Macapá. Rio de Janeiro. 1974.

FEITOSA, F. A. C. e MANOEL FILHO, J. **Hidrogeologia. Conceitos e Aplicações**. Conselho de Produção e Recursos Minerais – CPRM –, Laboratório de Hidrogeologia – LABHID– Universidade Federal de Pernambuco. 1997. 412 p.

FREEZER, R.A.; MASSMAN, J.; SMITH, L.; SPERLING, T. JAMES, B. Hydrogeological decision analysis: 1. A Framework. **Ground Water**, 28(5):738-766. 1990.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions. Bayesian restoration of images: **IEEE Trans Pattern Anal. Machine Intell.** PAMI. V.6(6):721-741. 1984,

GILBERT, R.O. **An Overview of Statistical Issues Related to Environmental Cleanup**. Handbook of Statistics, Vol. 12. Elsevier Science B.V. 1994, 895 p.

GOOVAERTS, P. Factorial kriging Analysis: a useful tool for exploring the structure of multivariate spatial information. **Journal of Soil Science**, 43:597-619. 1992.

GOOVAERTS, P. Spatial Orthogonality of the Principal Components Computed from Coregionalized Variables. **Mathematical Geology**, 25(3):281-302. 1993

GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation**. Oxford University Press, New York. 1997. 483 p.

GOOVAERTS, P. Kriging vs. Stochastic simulation for risk analysis in soil contamination. **Université Catholique de Louvain (Unité BIOM) Place Croix de Sud, 2 Bte 16, 1348 Louvain-la-Neuve, Belgium**. 1998.

GOOVAERTS, P. Geostatistics in soil science: state-of-art and perspectives. Elsevier Science. **Geoderma**, 89:1-45. 1999.

GOULARD, M. Inference in a coregionalization model. In M. Armstrong, editor, **Geostatistics**, vol. 1:397-408. Kluwer, Dordrech. 1989.

GOULARD, M. e VOLTZ, M. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. **Mathematical Geology**, 24(3):269-286. 1992.

GUERRA, A. T. **Estudo geográfico do território do Amapá**. Biblioteca Gráfica Brasileira, Publicação n^o. 10. Série A. IBGE, Rio de Janeiro. 1954. 366 p.

HALLENBECK, W.H. **Quantitative Risk Assessment for Environmental and Occupational Health**. Second Edition. Lewis Publishers, Boca Raton. 1994. 224 p.

HANAN, S. A.; BATALHA, B. H. L. **Amazônia, Contradições no Paraíso Ecológico**. Cultura Editores Associados –São Paulo. 1995. 266 p.

IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis) **Avaliação de Impacto Ambiental: Agentes Sociais, Procedimentos e Ferramentas**. BRASÍLIA. 1995. 134p.

IBGE –Instituto Brasileiro de Geografia e Estatística – **Geografia do Brasil –Região Norte**. Vol. 1. Rio de Janeiro, 1977. 466 p.

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

ISTOK, J.D.; RAUTMAN, C.A. Probabilistic Assessment of Ground Water Contamination: 2. Results of case Study. **Ground Water**. Vol. 34, No. 6:1050-1064. 1996.

INSTITUTO EVANDRO CHAGAS – Ministério da Saúde – **Síntese dos Resultados da Investigação sobre Exposição do Arsênio na Comunidade do Elesbão, no Município de Santana/Amapá**. Seção de Meio Ambiente. Belém/PA, abril 2002.

JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**. Prentice Hall, Englewood Cliffs, New Jersey. 1992. 642p.

JOURNEL, A.G. Mad and conditional quantile estimators. In G. Verly *et al*, editors. **Geostatistical for Natural Resources Characterization**, Part 1:261-270. 1984.

JOURNEL, A.G. **Geostatistical for the environmental sciences**. EPA Projects No. CR 811893, Las Vegas, USA. 1987. 135 p.

JOURNEL, A.G. **Fundamentals of Geostatistics in Five Lessons**. Vol. 8. Short Course in Geology American Geophysical Union, Washington, D.C., 1989. 92p.

JOURNEL, A.G.; ROSSI, M.E. When Do We Need a Trend Model in Kriging? **Mathematical Geology**, 21(7):715-739. 1989. .

JOURNEL, A.G.; HUIJBREGTS, C.J. **Mining Geoestatics**. London Academic Press, 1978. 600 p.

JUANG, K.W.; LEE, D.Y. Simple Indicator Kriging for Estimating the Probability of Incorrectly Delineation hazardous Areas in a Contaminated Site. **Environmental Sciences & Technology**. Vol 32. No. 17: 2487-2493. 1998.

KEITH, L.H. **Principles of Environmental Sampling**. 2nd ed. American Chemical Society. Washington, USA. 1996. 848p.

KITANIDIS, P. K. **Introduction to Geostatistics: Applications to Hydrogeology**. Cambridge University Press, New York, 1997. 249 p.

KRIGE, D.G. Half a Century of Geostatistics From a South African Perspective. Keynote Address, **6th International Geostatistics Congress**, South African, 2000.

LACAYO JR., H. **Environmental Risk Estimation and Policy Decisions**. Handbook of Statistics, Vol. 12. Elsevier Science B.V. 1994, 895 p.

LANDIM, P. M. B. **Análise estatística de dados geológicos**. Fundação Editora da UNESP. São Paulo, 1998. 226 p.

LEONTE, D.; SHOFIELD, N. Evaluation of a soil contaminated site and clean-up criteria: a geostatistical approach. **Geostatistics for Env. and Geotechnical Applications**. ASTM 1283. American Society for Testing and Materials. 133-145. 1996. 280 p.

LICHT, O.A.B. **A Geoquímica Multielementar na Gestão Ambiental**. Tese de Doutorado. Universidade Federal do Paraná. Curitiba, 2001. 192 p.

MILLARD, S.P.; NERCHAL, M.K. **Environmental Statistics with S-Plus**. CRC Press LLC. USA, 2001. 830p.

MYERS, J.C. **Geostatistical Error Management: quantifying uncertainty for Environmental Sampling and Mapping**. Van Nostrand Reinhold. New York. USA. 1997. 571p.

OMS—Organização Mundial de Saúde—**Guidelines for drinking water quality. Health criteria and other supporting information** 2^a Ed. Vol. 2: 940-949. Genebra, 1998.

PARDO-IQUIZGUIZA, E.; DOWD, P.A. FACTOR2D: a computer program for factorial co-kriging. **Computers and Geosciences**, 28:857-875. 2002.

PARECER TÉCNICO – Universidade Federal do Pará. Laboratório de Química Analítica e Ambiental – LAQUANAM. **Avaliação da Contaminação por Arsênio no Cabelo de Moradores de Santana/AP**. Belém/PA. 2000. 46p.

PCA – PLANO DE CONTROLE AMBIENTAL. Relatório elaborado pela AMPLA ENGENHARIA/ICOMI – Indústria e Comércio de Minérios SA. 2001

QUEIROZ, J.C.B; STURARO, J.R.; RIEDEL, P.S. Geostatistic Mapping of Arsenic, Manganese and Iron Contamination Risk in the Port of Santana, Amapá, Brazil. **Conference International Association for Mathematical and Geology**, Cancun, Mexico, 2001. Disp. em www.kgs.ukans.edu/Conferences/IAMG/Sessions/G/g.html.

RAUTMAN,C.A. Geostatistics and cost-effective environmental remediation. Fifth Int. **Geostatistics Congress, University of Wollongong**, Wollongong, Austrália, 1996.

RAUTMAN,C.A.& ISTOK,J.D. Probabilistic Assessment of Ground-Water Contamination: 1 Geostatistic Framework.**Ground Water**,Vol.35, No.5:899-909. 1996.

RELATÓRIO JAAKKO POYRY ENGENHARIA (JPE): **Disposição Final dos Resíduos da Usina de Pelotização/Sinterização Estocados na Área Industrial da ICOMI em Santana, Amapá, Brasil**. Anteprojeto, Maio, 1998. 59p.

RELATÓRIO TÉCNICO PARCIAL – Universidade Federal do Pará. Laboratório de Química Analítica e Ambiental – LAQUANAM. **Caracterização de Arsênio em amostras de solo da cidade de Santana/AP. Avaliação da contaminação por metais Pesados no Delta do Rio Amazonas**. Belém/PA. 2001. 32p.

RENDU, J.M. **An Introduction to Geostatistical Methods of Mineral Evaluation**. South African Institute of Mining and Metallurgy. Series, Geostatistics 2, 1978.

RIBEIRO JR., P.J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. Tese de Mestrado. ESALQ-PIRACICABA, São Paulo,1995. 99p.

ROLNIK, R. Exclusão Territorial e Violência O caso de São Paulo, Brasil. São Paulo: **Instituto Polis**, novembro, 1998.

SHRIVASTAVA, R. M. Minimum variance or maximum profitability ? **CIM Bulletin**, 80(901):63-68. 1987.

SHRIVASTAVA, R. M. An Application of Geostatistical Methods for Risk Analysis in Reservoir Management. **Society of Petroleum Engineers**. SPE 20608. 1990

SOUZA,H.B. **Guia Técnico de Coleta de Amostras de Água**. CETESB,S.Paulo. 1977.

STURARO, J.R. **Mapeamento geoestatístico de propriedades geológico-geotécnicas obtidas de sondagens de simples reconhecimento**. Tese de Doutorado. São Carlos, 1994.183 p. Disponível em www.rc.unesp.br/igce/pos/indexpos.html.

SURFER 7.00. **User's guide. Contouring and 3D surface mapping for scientists and engineers**. Colorado:Golden Software, 1999. 619p.

TEIXEIRA, A. J. & FYFE, W. S. 1994. Notas preliminares sobre a toxidez potencial dos sedimentos do leito do Ribeirão Claro e da cinza hospitalar por metais pesados, em Rio Claro, SP. **Geociências**, São Paulo, 13(1): 243:261.

U.S.EPA – UNITED STATES ENVIRONMENTAL PROTECTION AGENCY. Risk Characterization Policy and Guidance. **Memorandum from Carol M. Brower**.1995.

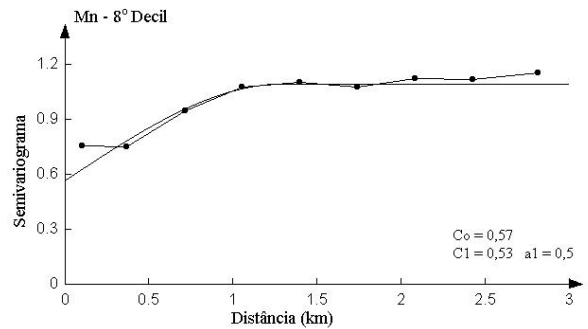
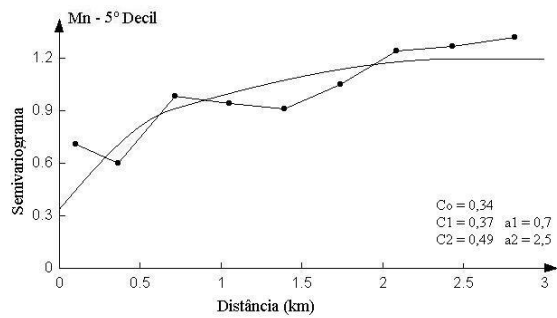
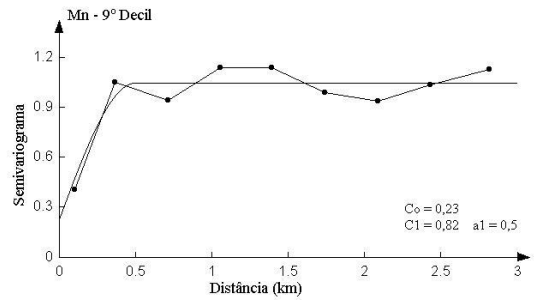
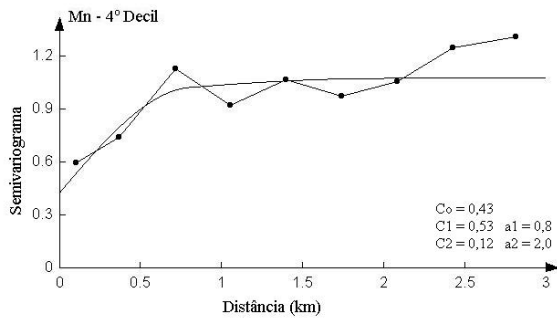
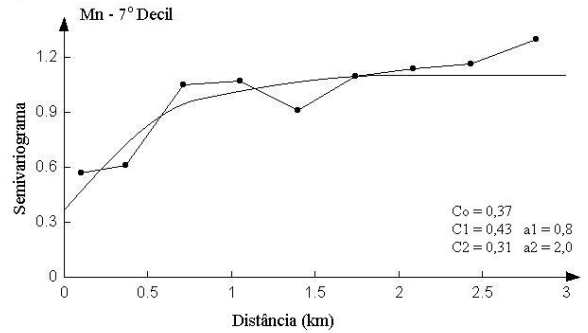
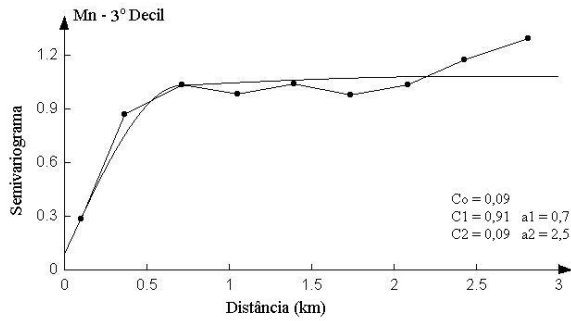
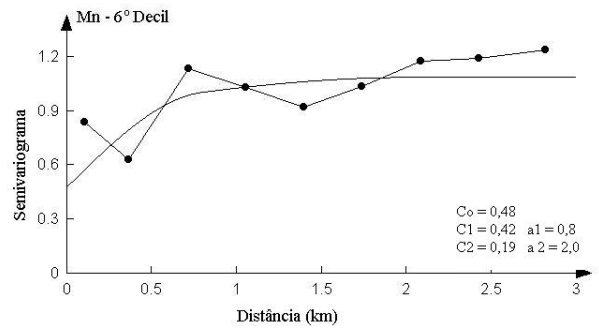
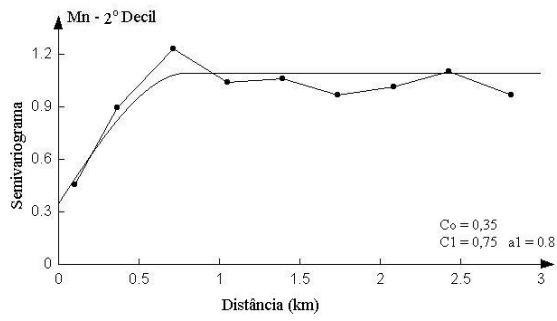
U.S.EPA – Drinking Water Standards and Health Advisories. **Office of Water 4304**, 2000. Disponível em www.epa.gov/OST.

WEBSTER,R.;ATTEIA,O.;DUBOIS,J.P. Coregionalization of trace metals in the soil in the Swiss Jura. **European Journal of Soil Science**, 45:205-218. 1994.

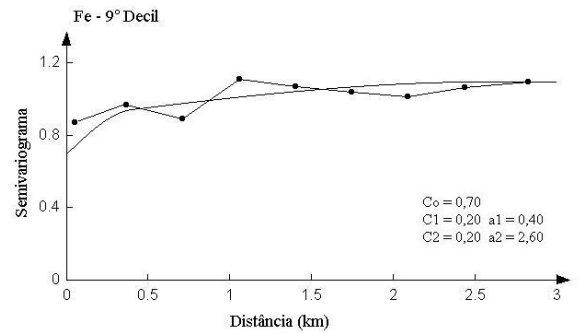
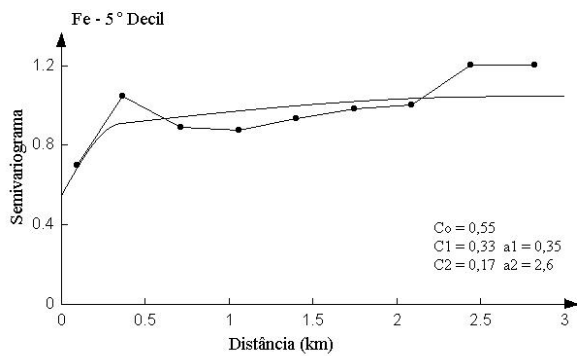
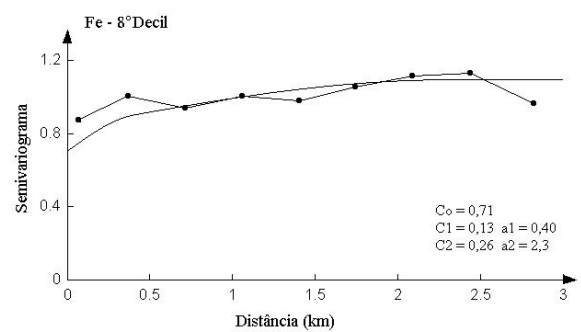
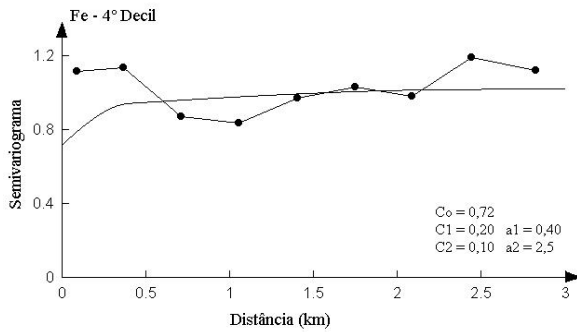
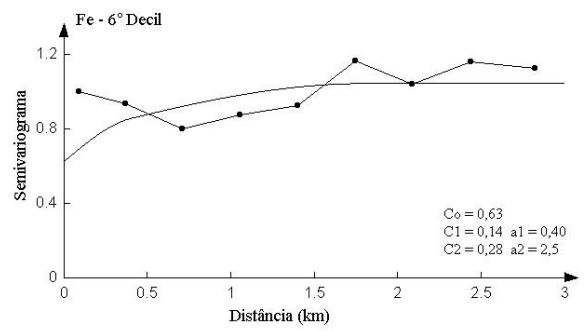
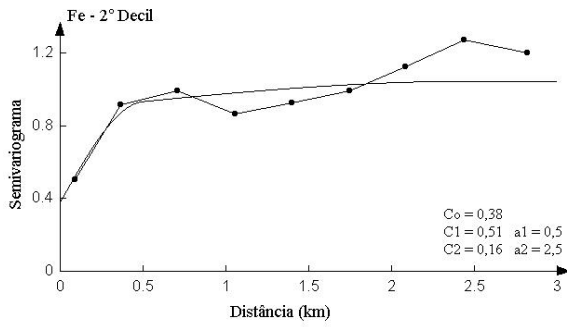
ANEXO 1

Semivariogramas indicativos experimentais (linhas ligadas por pontos) e modelos ajustados (linhas contínuas).

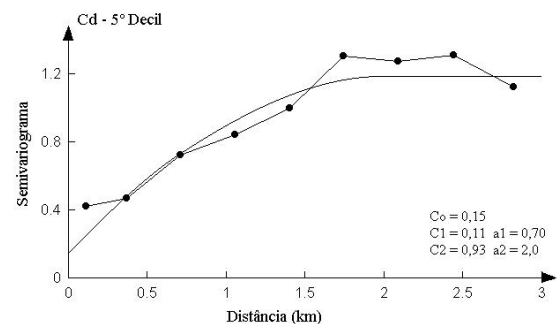
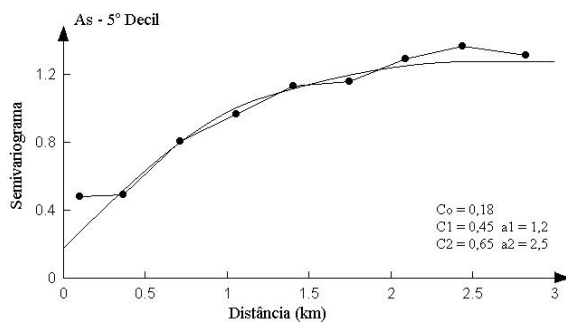
Semivariogramas experimentais cruzados das variáveis (linhas ligadas por pontos) e os semivariogramas ajustados pelo modelo linear de co-regionalização (linhas contínuas)



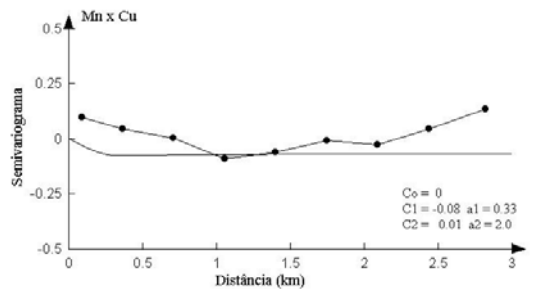
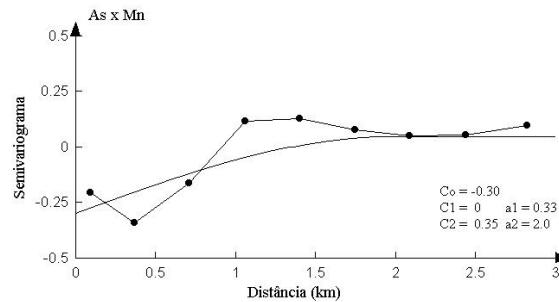
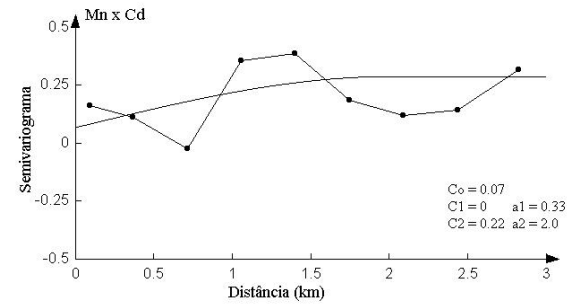
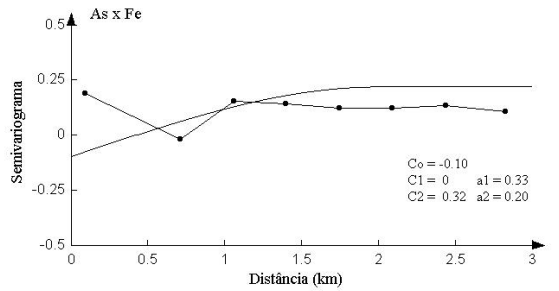
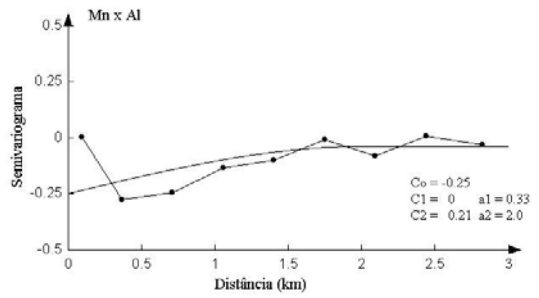
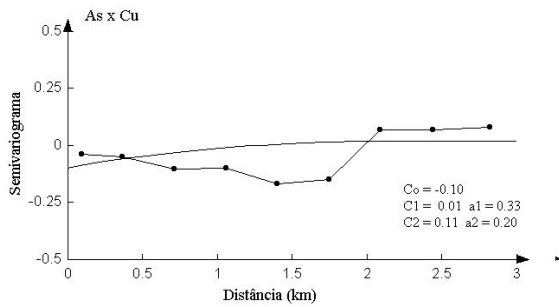
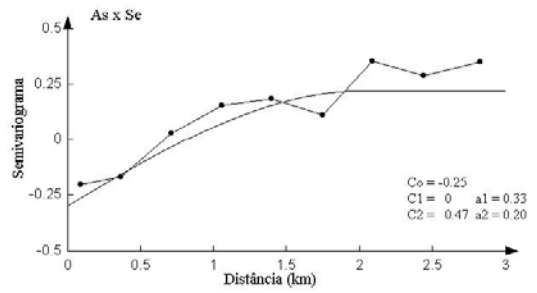
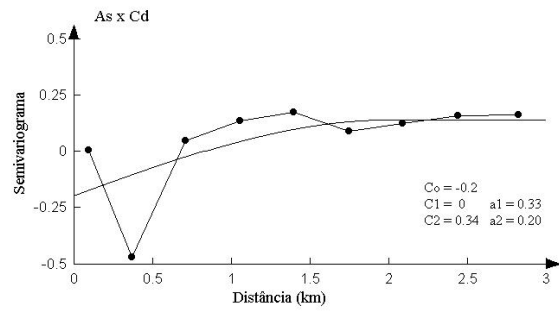
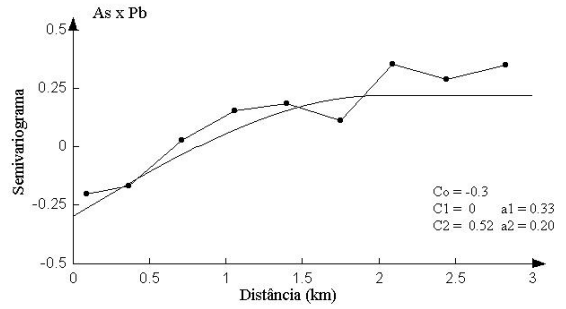
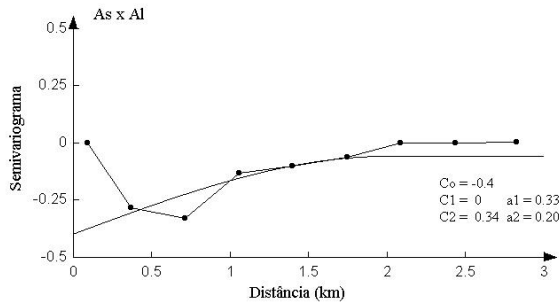
Semivariogramas indicativos experimentais (linhas ligados pos pontos) e modelos ajustados (linhas contínuas) para o manganês



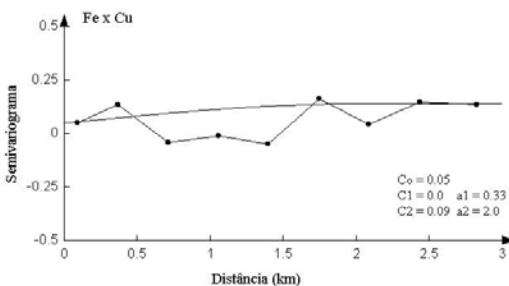
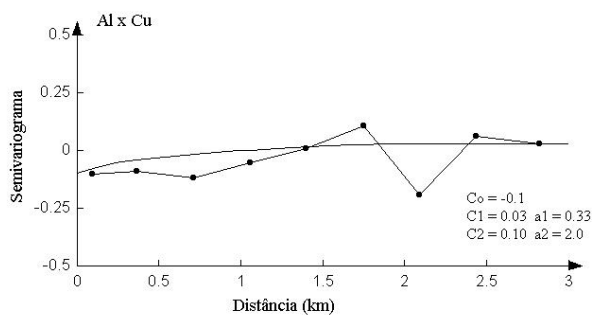
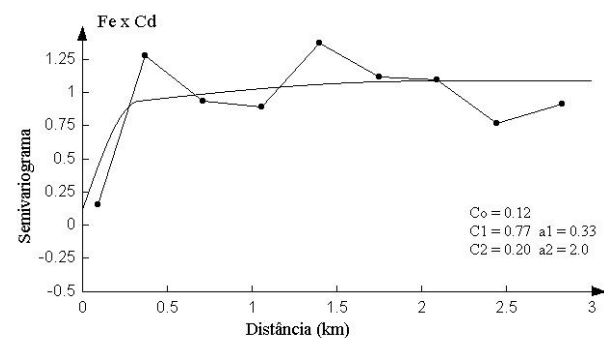
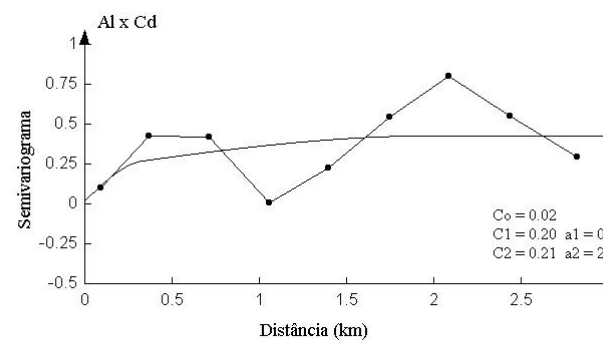
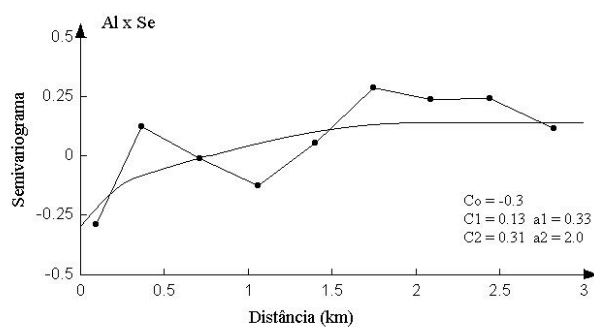
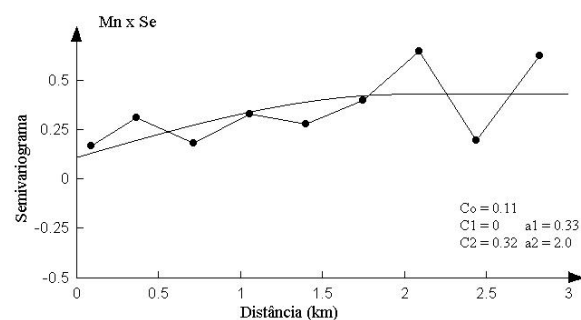
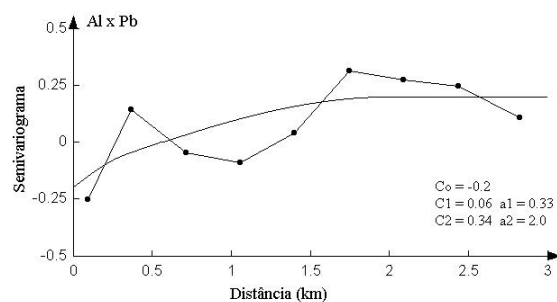
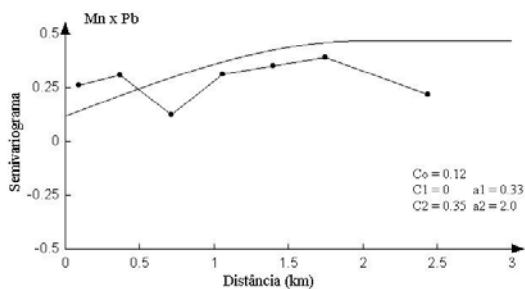
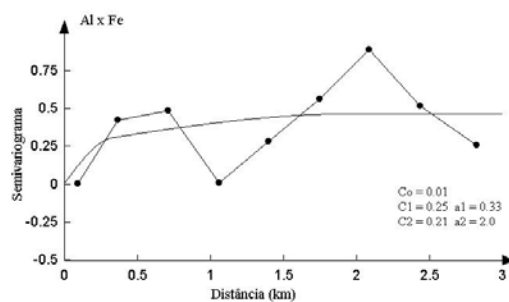
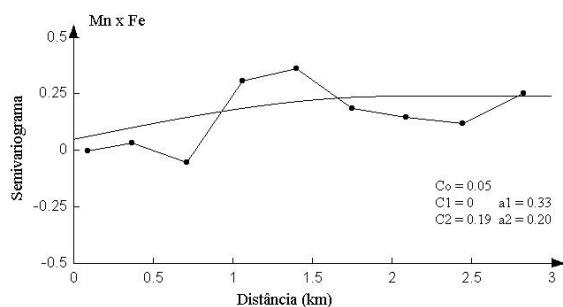
Semivariogramas indicativos experimentais (linhas ligados pos pontos) e modelos ajustados (linhas contínuas) para o ferro.



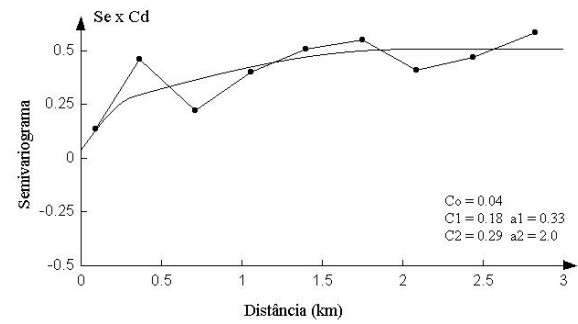
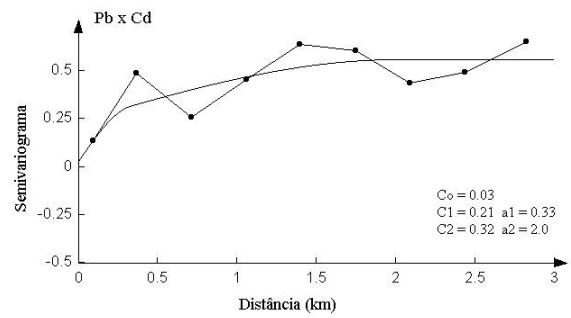
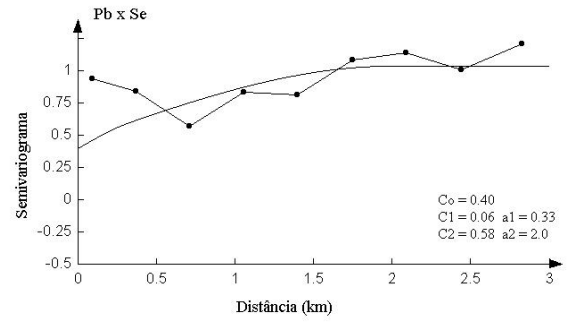
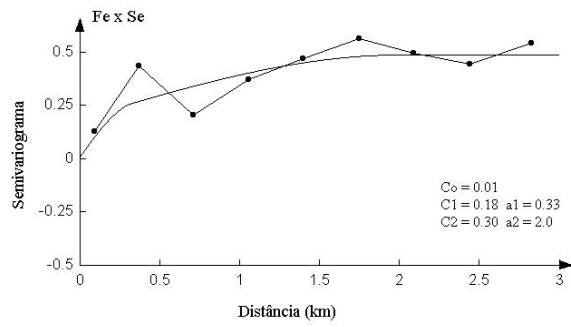
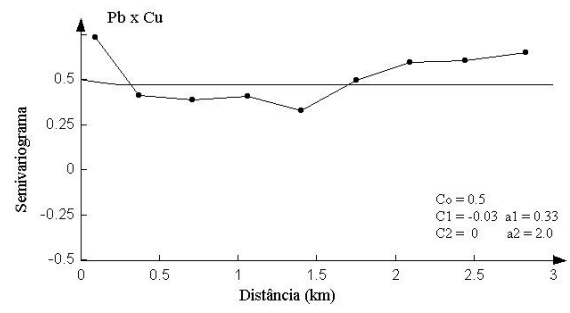
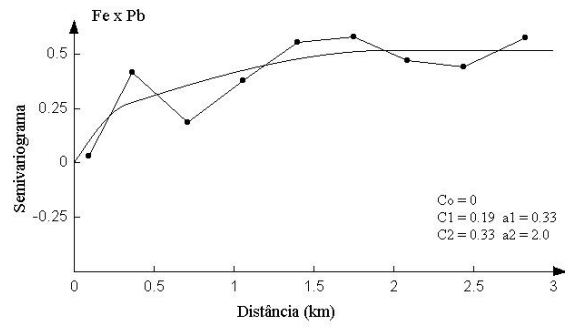
Semivariograma indicativo experimental (linhas ligados pos pontos) e modelo ajustado (linha contínua) para o arsênio (gráfico à esquerda) e cádmio (gráfico à direita)



Semivariogramas experimentais cruzados das variáveis (linhas ligadas por pontos) e os semivariogramas ajustados pelo modelo linear de co-regionalização (linhas contínuas)



Semivariogramas experimentais cruzados das variáveis (linhas ligadas por pontos) e os semivariogramas ajustados pelo modelo linear de co-regionalização (linhas contínuas)



Semivariogramas experimentais cruzados das variáveis (linhas ligadas por pontos) e os semivariogramas ajustados pelo modelo linear de co-regionalização (linhas contínuas)

ANEXO 2

Programa IKLOSS: adaptação do programa POSTIK da livreria GSLIB
para computação da perda esperada

Programa ikloss

O programa **ikloss** foi uma adaptação do programa **postik** da livreria GSLIB para o cálculo da perda esperada associada aos dois tipos de classificações modeladas pelas funções (3.81) e (3.82) na avaliação da incerteza local, para a classificação de áreas contaminadas. O programa utiliza a saída do programa **ik3d** que fornece o modelo da função de distribuição acumulada condicional, $F(\mathbf{u};z|(n))$. Com o modelo de $F(\mathbf{u};z|(n))$ pode-se determinar a perda esperada com base nas equações (3.83) e (3.84) para a classificação das áreaa contaminadas.

O programa executável **ikloss** com o respectivo código fonte IKLOSS.FOR, escrito em linguagem de programação FORTRAN 77 padrão ANSI, onde se pode modificá-lo facilmente para a inclusão de outros tipos de funções de perda, estão disponíveis em www.rc.unesp.br/igce/pos/indexpos.html.

Os parâmetros necessários para o programa **ikloss** são fornecidos por meio de um arquivo de parâmetros, do tipo ASC-II que pode ser criado em qualquer programa editor de textos. Um exemplo desse arquivo de parâmetros é apresentado na Figura A1.

```

                Parametros para IKLOSS
                *****

START OF PARAMETERS:
ikmn.out          \arquivo de saída de IK3D
ikl2w25.out       \arquivo de saída
2  2.5  0.10      \funcao de perda, custo(peso), cutoff
8                \numero de cortes
0.007 0.013 0.017 0.030 0.039 0.084 0.38 2.03 \os cortes
mn.dat           \arquivo com a distribuicao global
3  0    -1.0e21   1.0e21   \coluna da variavel, peso, tmin, tmax
0.0    5.0       \valor minimo e maximo de Z
1  1.0          \cauda inferior: opção, parametro
1  1.0          \meio           : opção, parametro
1  1.0          \cauda superior: opção, parametro
100            \discretizacao maxima

opcao para a funcao de perda = 1: Funcao de perda L1
                             = 2: Funcao de perda L2
```

Figura A1 –Arquivo de parâmetros para o programa IKLOSS

Os parâmetros de entrada são listados a seguir:

distin	distribuições de entrada (saída do programa IK3D)
outfl	arquivos de saída para as funções L_1 (equação 3.81) e L_2 (equação 3.82)
iout,iw,outpar	iout = 1: função de perda L_1 ; iout = 2: função de perda L_2 iw = custo relativo; outpar = valor de corte
nccut	numero de pontos de corte
ccut(i)	pontos de corte, $i=1, ncut$.
datafl	distribuição global de dados
ivr,iwt,tmin,tmax	parâmetros de entrada da distribuição global de dados ivr : coluna da variável iwt : peso (para desagrupamento) tmin e tmax : intervalo de valores (mínimo e máximo) que serão considerados. Qualquer valor fora desse intervalo será ignorado.
zmin,zmax	valores de dados de entrada mínimo e máximo permitidos.
ltail,ltpar	opção e parâmetro para a extrapolação da cauda inferior da distribuição acumulada. ltail =1 implementa a interpolação linear para o limite inferior zmin ; ltail =2 implementa o modelo potência de interpolação com $w = \text{ltpar}$ para o limite inferior zmin ; ltail =3 implementa a interpolação linear entre os quantis tabulados (DEUSCTH E JOURNAL, 1998, p. 134).
middle,midpar	opção e parâmetro para interpolação do centro da distribuição acumulada. middle =1 implementa a interpolação linear para o limite inferior zmin ; middle =2 implementa o modelo potência de interpolação com $w = \text{midpar}$ para o limite inferior zmin ; middle =3 implementa a interpolação linear entre os quantis tabulados (DEUSCTH E JOURNAL, 1998, p. 134).
utail,utpar	opção e parâmetro para extrapolação da cauda superior da distribuição acumulada. utail =1 implementa a interpolação linear para o limite inferior zmin ; utail =2 implementa o modelo potência de interpolação com $w = \text{utpar}$ para o limite superior zmax ; utail =3 implementa a interpolação linear entre os quantis tabulados; utail =4 implementa o modelo hiperbólico de extrapolação com $w = \text{utpar}$ (DEUSCTH E JOURNAL, 1998, p. 134).
maxdis	parâmetro de discretização máxima. O valor <i>default</i> é tipicamente fino; maior acurácia é obtido com um número maior e tempo de execução mais rápido é obtido com um número menor (DEUSCTH E JOURNAL, 1998, p. 238).

ANEXO 3

Limites de detecção do equipamento e faixa linear de trabalho para os metais.



**INSTRUÇÃO TÉCNICA DE ENSAIO
PARA A DETERMINAÇÃO DE METAIS
POR PLASMA INDUTIVAMENTE
ACOPLADO EM ÁGUA
ITICP**

**LABORATÓRIO
CENTRAL**

ANEXO

- Faixa de concentração – os limites de detecção do equipamento e faixa linear de trabalho para os metais

ELEMENTO	SIMBOLO	COMPRIMENTO DE ONDA (nm)	LIMITE DE DETECÇÃO(µg/L)
ALUMÍNIO	Al	308,215	19
ANTIMÔNIO	Sb	206,833	32
ARSÊNIO	As	193,696	5,9
BÁRIO	Ba	455,403	0,046
BERÍLIO	Be	313,042	0,3
BORO	B	249,773	5
CÁDMIO	Cd	226,502	0,82
CÁLCIO	Ca	317,933	1,2
CHUMBO	Pb	267,716	5,7
COBALTO	Co	228,616	1,7
COBRE	Cu	324,754	0,08
CROMO	Cr	259,940	1,2
FERRO	Fe	220,353	0,8
MAGNÉSIO	Mg	279,079	3,7
MANGANÊS	Mn	257,610	0,11
MOLIBIDÊNIO	Mo	202,030	3,7
NÍQUEL	Ni	231,604	2,7
SELÊNIO	Se	196,026	9,2
PRATA	Ag	328,068	0,87
POTÁSSIO	K		0,5
SÓDIO	Na	588,995	4,6
TÁLIO	Ta	190,864	40
TITÂNIO	Ti	334,941	3
VANÁDIO	V	292,402	8
ÍTRIO	Y	371,030	2,5
ZINCO	Zn	213,856	0,17
ESTANHO	Sn	189,989	30
ESTRÔNCIO	Sr		0,01

Edição controlada N^o : 001 - Não pode ser copiada

Rev.: 00

Data: 03/04/2000

Elaborado por:

Aprovado por:

Ctrl.: IT.019.012

Página: 83 / 6