

**ALEX DA SILVA SANTOS**

**GEOESTATÍSTICA APLICADA NA AVALIAÇÃO DA QUALIDADE DE  
DADOS GEOESPACIAIS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Civil, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2013

Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV

T

S237g Santos, Alex da Silva, 1978-  
2013 Geoestatística aplicada na avaliação da qualidade de dados  
geoespaciais / Alex da Silva Santos. – Viçosa, MG, 2013.  
xiv, 148 f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.

Orientador: Nilcilene das Graças Medeiros.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 137-144.

1. Análise espacial (Estatística). 2. Cartografia.  
3. Metadados. I. Universidade Federal de Viçosa. Departamento  
de Engenharia Civil. Programa de Pós-Graduação em Engenharia  
Civil. II. Título.

CDD 22 ed. 519.5

**ALEX DA SILVA SANTOS**

**GEOESTATÍSTICA APLICADA NA AVALIAÇÃO DA QUALIDADE DE  
DADOS GEOESPACIAIS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Civil, para obtenção do título de Magister Scientiae.

APROVADA: 7 de outubro de 2013.

---

Prof. Mauricio Galo

---

Prof. Jugurta Lisboa Filho  
(Coorientador)

---

Prof. Gérson Rodrigues dos Santos  
(Coorientador)

---

Prof. Nilcilene das Graças Medeiros  
(Orientadora)

*À minha mãe Ernestina Maria da Silva Santos (in memoriam),  
ao meu pai Claudio Ferreira dos Santos (in memoriam), à  
minha esposa Marcia Vaz, ao nosso filho Pedro Vaz e a Deus.*

*“Nenhum homem realmente produtivo pensa como se estivesse  
escrevendo uma dissertação.”  
(Albert Einstein)*

## AGRADECIMENTOS

Primeiramente a Deus, por tudo e por estar presente em todos os momentos de nossas vidas! À minha esposa Marcia Vaz Barbosa e ao nosso filho Pedro Vaz Barbosa Santos, fontes de incentivo desde o início até a conclusão deste trabalho. Aos meus pais Ernestina Maria da Silva Santos (*in memoriam*) e Cláudio Ferreira dos Santos (*in memoriam*), com os quais tenho buscado forças nos momentos de fraqueza, à minhas irmãs Vânia e Vanessa (gêmeas), aos meus sobrinhos Guilherme Mazzei e Sarah Mazzei (primos irmãos) e seus avós paternos Eleide e Rubens Mazzei. A minha tia Maria e meus primos: Jane, Ney e Robson Real.

Aos meus cunhados "tios" Gabriel, Judith, Francisco, Rogério, Marly, Dilene e Helenice, inclusive a Jaqueline, e a toda família Barbosa.

Ao amigo e compadre Salomão Soares e família (agora amigo de IBGE).

Um agradecimento especial à minha amiga Anna Lúcia Barreto de Freitas, do IBGE, com quem tanto aprendi e um exemplo a seguir em minha carreira profissional.

Aos meus eternos orientadores Nilcilene das Graças Medeiros, Gérson Rodrigues dos Santos e Jugurta Lisboa Filho da UFV, e ao meu orientador de graduação Jorge Luís Nunes e Silva Brito da UERJ, verdadeiros inspiradores na arte de aprender e ensinar.

À Maria Cristina Zennaro (UERJ), por ter feito parte de uma fase importante de minha vida e grande incentivadora dos meus estudos.

Aos amigos e companheiros do IBGE, em especial a Alessandra Gouveia, Azarias Aquino, Aline Lopes, Célia Vianna, Claudio João, Denise Torres, Edilce Burity, Eduardo Abraão, Graciosa Rainha, Guiomar Regato, Hamilton Ramos, Hebert Azevedo, Iracema Alice, Irenil Leocadio, Jaime Pitaluga, João Bosco, Leila Freitas, Luiz Xavier, Marcelo Maranhão, Márcia Gago, Márcia Mathias, Maria Amélia, Marilea Melo, Maurício Krumbiegel, Moema Augusto, Nadja Luna, Patrícia Vida, Paulo Gimenez, Paulo Trezena, Rafael Balbi, Renata Curi, Sulamita Aquino, Vânia Nagem, Vânia Rasga, Vitor França, Walter Subiza, Wesley Fernandes e a Dona Marina.

Ao Luis Felipe Louzada, da Coordenação de Treinamento do IBGE pelo incentivo e orientações.

Um agradecimento especial ao Professor Maurício Galo, da Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP, pelas relevantes contribuições incorporadas à esta pesquisa.

À Universidade Federal de Viçosa - UFV, ao programa de pós graduação em Engenharia Civil, e ao IBGE pela oportunidade deste desenvolvimento profissional.

Aos amigos e professores da UFV, em especial ao Afonso de Paula, Daniel Duarte, Bruno Ventrone, Edilson José, Flávio Oliveira, Gabriel Diniz, Jarbas Nunes, Marília Sanglard, Paulo Caixeta, Rafaela Nogueira, Salomão Martins, Wagner Souza e Wellington Guimarães, espero revê-los em breve.

E aos que passaram por minha vida desejando o bem ao próximo, sintam-se agradecidos.

## SUMÁRIO

<b>LISTA DE FIGURAS .....</b>	<b>viii</b>
<b>LISTA DE TABELAS .....</b>	<b>xi</b>
<b>RESUMO .....</b>	<b>xiii</b>
<b>ABSTRACT .....</b>	<b>xiv</b>
<b>1. INTRODUÇÃO .....</b>	<b>1</b>
1.1. Objetivo.....	3
1.2. Motivação.....	4
<b>2. FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>5</b>
2.1. Qualidade de dados: uma abordagem geoespacial .....	5
2.1.1. Qualidade do produto e qualidade do processo .....	8
2.1.2. Visão produtor e visão do usuário .....	8
2.1.3. Qualidade de dados em Cartografia.....	9
2.1.4. Normas de qualidade de dados espaciais.....	10
2.1.5. Elementos de qualidade de dados espaciais.....	15
2.2. Generalização cartográfica: princípios de cartografia aplicados na produção de mapas.....	21
2.2.1. Conceitos de generalização cartográfica .....	23
2.2.2. Operadores de generalização cartográfica .....	24
2.3. Técnicas de amostragem aplicadas a dados geoespaciais.....	31
2.3.1. Conceitos de amostragem.....	31
2.3.2. Métodos de amostragem.....	33
2.3.3. Tamanho da amostra .....	37
2.3.4. Estratégias de amostragem para dados espaciais .....	39
2.4. Geoestatística .....	40
2.4.1. Conceitos de Geoestatística .....	41
2.4.2. A teoria das variáveis regionalizadas .....	42
2.4.3. Análise estrutural.....	43
2.4.4. Estimativas Geoestatísticas .....	52
2.5. Metadados: informando a qualidade de dados ao usuário .....	60
2.5.1. Padrões de metadados geoespaciais .....	62
2.5.2. A norma ISO 19115:2003 .....	63
2.5.3. Perfil de Metadados Geoespaciais do Brasil (MGB).....	64
2.5.4. Informações de qualidade de dados espaciais .....	65
<b>3. MATERIAIS E MÉTODOS .....</b>	<b>68</b>
3.1. Descrição geral do conjunto de dados em estudo.....	68
3.1.1. Dados e insumos utilizados .....	69
3.2. Equipamentos e softwares .....	70
3.3. Metodologia .....	71
3.3.1. Avaliação da qualidade .....	72

3.3.2. Método de inspeção.....	76
3.3.3. Análise espacial da qualidade.....	96
3.3.4. Metadados da avaliação de qualidade .....	102
<b>4. RESULTADOS E DISCUSSÃO .....</b>	<b>103</b>
4.1. Avaliação da qualidade dos dados espaciais em relação a completude.....	103
4.1.1. Identificar a medida de qualidade.....	104
4.1.2. Método de inspeção.....	104
4.2. Análise espacial da qualidade: Geoestatística aplicada na avaliação da completude .....	105
4.2.1. Preparo da análise geoestatística em termos de comissão e omissão .....	105
4.2.2. Análise exploratória dos dados de comissão .....	106
4.2.3. Análise variográfica dos dados de comissão .....	108
4.2.4. Krigagem dos dados de comissão.....	110
4.2.5. Mapa de qualidade de comissão .....	110
4.2.6. Análise exploratória dos dados de omissão .....	112
4.2.7. Análise variográfica dos dados de omissão .....	113
4.2.8. Krigagem dos dados de omissão .....	115
4.2.9. Mapa de qualidade de omissão .....	115
4.3. Metadados da avaliação de qualidade da completude .....	117
4.3.1. Metadados de qualidade: taxa de comissão de elementos de hidrografia.....	117
4.3.2. Metadados de qualidade: taxa de omissão de elementos de hidrografia.....	118
4.4. Avaliação da qualidade dos dados espaciais em relação a acurácia posicional.....	119
4.4.1. Identificar a aplicação de elementos de qualidade e o âmbito da avaliação .....	119
4.4.2. Identificar a medida de qualidade.....	119
4.4.3. Método de inspeção para acurácia posicional absoluta .....	120
4.5. Análise espacial da qualidade: Geoestatística aplicada na avaliação da acurácia posicional .....	121
4.5.1. Preparo da análise geoestatística em termos de acurácia posicional absoluta .....	121
4.5.2. Análise exploratória dos dados da acurácia posicional absoluta .....	122
4.5.3. Análise variográfica dos dados da acurácia posicional absoluta .....	124
4.5.4. Krigagem dos dados da acurácia posicional absoluta.....	126
4.5.5. Mapa de qualidade da acurácia posicional absoluta .....	126
4.6. Metadados da avaliação da qualidade da acurácia posicional .....	127
4.6.1. Metadados de qualidade: acurácia posicional absoluta da BC250 .....	127
4.7. Considerações gerais.....	128
<b>5. CONCLUSÕES E RECOMENDAÇÕES .....</b>	<b>131</b>
5.1. Conclusões .....	131
5.2. Recomendações.....	134
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>137</b>
<b>APÊNDICES.....</b>	<b>145</b>
APÊNDICE A - Medidas de qualidade .....	146
Apêndice A.1 - Medida de qualidade para itens em excesso.....	146
Apêndice A.2 - Medida de qualidade para itens ausentes .....	147
Apêndice A.3 - Medida de qualidade para acurácia posicional absoluta .....	148

## LISTA DE FIGURAS

Figura 1 - Aspectos gerais de qualidade de dados espaciais.....	9
Figura 2 - Visão geral de informações da qualidade de dados.....	11
Figura 3 - Pirâmide de implementação de normas ISO. ....	11
Figura 4 - Acurácia versus precisão. ....	14
Figura 5 - Exemplos de omissão (B) e comissão (C).....	15
Figura 6 - Relação entre o processo de abstração da realidade e a completude.....	16
Figura 7 - Esquema de significados de completude.....	17
Figura 8 - Exemplo de avaliação da consistência lógica.....	18
Figura 9 - Exemplo de avaliação da acurácia posicional .....	19
Figura 10 - Operadores de generalização cartográfica.....	25
Figura 11 - Operador de generalização cartográfica: amálgama .....	26
Figura 12 - Operador de generalização cartográfica: exagero .....	27
Figura 13 - Operador de generalização cartográfica: fusão .....	28
Figura 14 - Operador de generalização cartográfica: indução .....	28
Figura 15 - Operador de generalização cartográfica: refinamento.....	29
Figura 16 - Operador de generalização cartográfica: simplificação .....	30
Figura 17 - Exemplo de amostragem não aleatória orientada por área.....	34
Figura 18 - Exemplo de amostragem aleatória orientada por áreas .....	35
Figura 19 - Relacionamento entre as estratégias de amostragem .....	39
Figura 20 - Os três variogramas da Geoestatística.....	44
Figura 21 - Exemplo de construção de um semivariograma.....	45
Figura 22 - Principais parâmetros do semivariograma .....	45
Figura 23 - Distribuição regular de amostras e a construção de variogramas experimentais .....	46
Figura 24 - Distribuição irregular de amostras e a construção de variogramas experimentais .....	47
Figura 25 - Relação entre os alcances dos principais variogramas teórico com patamar	49
Figura 26 - Tipos de comportamento na origem do variograma.....	50
Figura 27 - Variograma indicando efeito buraco .....	50
Figura 28 - Variograma indicando estrutura imbricada .....	51
Figura 29 - Fluxograma do ajuste do variograma teórico ao experimental. ....	52
Figura 30 - Krigagem ou interpolação .....	53

Figura 31 - Processo de estimativa geoestatística.....	53
Figura 32 - Exemplo de estimação de um valor através da krigagem ordinária.....	57
Figura 33 - Síntese do método geoestatístico.....	60
Figura 34 - Perfil de metadados de uma comunidade.....	64
Figura 35 - Diagrama UML das seções de metadados do Perfil MGB.....	64
Figura 36 - A BC250 e seus blocos de trabalho antes da disponibilização na INDE.....	69
Figura 37 - Fluxograma para avaliação da qualidade de dados geoespaciais.....	71
Figura 38 - Legenda dos símbolos usados nos fluxogramas da dissertação.....	72
Figura 39 - Ilustração das medidas de qualidade elaboradas para avaliação da BC250.....	73
Figura 40 - Procedimentos para elaboração de classe(s) de ocorrência.....	74
Figura 41 - Exemplo de classes de ocorrências criadas para completude.....	75
Figura 42 - Exemplo de classes de ocorrências criadas para acurácia posicional.....	76
Figura 43 - Procedimentos para inspeção por amostragem.....	77
Figura 44 - Procedimentos para inspeção completa.....	78
Figura 45 - Malha regular baseada nas articulações do mapeamento sistemático.....	80
Figura 46 - Procedimentos pra elaboração da grade regular contínua.....	81
Figura 47 - Delimitação político-administrativa do Estado de Minas Gerais.....	82
Figura 48 - Grade regular na estrutura vetorial sem índice linha e coluna.....	83
Figura 49 - Distribuição espacial das 571 unidades de amostragem para completude.....	86
Figura 50 - Procedimentos para seleção de feições geográficas dentro das unidades de amostragem.....	87
Figura 51 - Trecho de drenagem significativo ausente na BC250 (Rio Coruripe).....	88
Figura 52 - Exemplos de comissão e omissão de um trecho de drenagem.....	89
Figura 53 - Exemplo de contagem de ocorrências por unidade de amostragem.....	90
Figura 54 - Tabela de atributos contendo os dados observados na inspeção da completude por unidade de amostragem.....	91
Figura 55 - Distribuição espacial das áreas para coleta de pontos amostrais referentes à avaliação da acurácia posicional absoluta.....	93
Figura 56 - Detalhe da amostragem sistemática para acurácia posicional absoluta.....	94
Figura 57 - Detalhe de uma situação de aferição da acurácia posicional absoluta.....	95
Figura 58 - Procedimentos da análise geoestatística.....	97
Figura 59 - Fluxograma do ajuste do variograma teórico ao experimental.....	98
Figura 60 - Procedimentos para o georrefenciamento da krigagem.....	99
Figura 61 - Ferramenta de georreferenciamento no gvSIG.....	100
Figura 62 - Tela principal para o georreferenciamento no gvSIG.....	101

Figura 63 - Dados observados da taxa de comissão e omissão no formato texto .....	106
Figura 64 - Histograma referente à taxa de comissão de feições geográficas de hidrografia da BC250.....	107
Figura 65 - Distribuição espacial dos dados observados referentes à taxa de comissão de feições geográficas de hidrografia da BC250 .....	107
Figura 66 - Variograma experimental dos dados de comissão .....	108
Figura 67 - Ajuste do variograma teórico ao experimental dos dados de comissão.....	109
Figura 68 - Mapa de qualidade referente à taxa de comissão de feições geográficas de hidrografia da BC250.....	111
Figura 69 - Histograma referente à taxa de omissão de feições geográficas de hidrografia da BC250.....	112
Figura 70 - Distribuição espacial dos dados observados referentes à taxa de omissão de feições geográficas de hidrografia da BC250 .....	113
Figura 71 - Variograma experimental dos dados de omissão .....	113
Figura 72 - Ajuste do variograma teórico ao experimental dos dados de omissão.....	114
Figura 73 - Mapa de qualidade referente à taxa de omissão de feições geográficas de hidrografia da BC250.....	116
Figura 74 - Dados observados de acurácia posicional absoluta em formato texto .....	121
Figura 75 - Histograma referente ao erro planimétrico da BC250.....	122
Figura 76 - Distribuição espacial dos dados observados referentes à acurácia posicional absoluta .....	123
Figura 77 - Variograma experimental dos dados referentes ao erro planimétrico.....	124
Figura 78 - Ajuste do variograma teórico ao experimental para o erro planimétrico...	125
Figura 79 - Mapa de qualidade da acurácia posicional absoluta da BC250.....	127
Figura 80 - Generalização cartográfica considerada em massas d'água em excesso....	146
Figura 81 - Generalização cartográfica considerada em massas d'água ausentes.....	147

## LISTA DE TABELAS

TABELA 1 - Normas, padrões e especificações para mapeamento da INDE.....	12
TABELA 2 - Padrão de Exatidão Cartográfica - PEC.....	13
TABELA 3 - Níveis de abstração do mundo real ao terreno nominal.....	16
TABELA 4 - Elementos e subelementos de qualidade.....	21
TABELA 5 - Classificação dos graus de aleatoriedade.....	47
TABELA 6 - Equações de modelos de variogramas teóricos.....	48
TABELA 7 - Entidades e elementos do Perfil MGB sumarizado.....	65
TABELA 8 - Componentes de metadados de qualidade de dados espaciais.....	66
TABELA 9 - Montagem da expressão SQL para seleção aleatória sistemática em SIG 85	
TABELA 10 - Descrição dos campos da unidade de amostragem de completude.....	91
TABELA 11 - Estatística descritiva referente à taxa de comissão de feições geográficas de hidrografia da BC250.....	106
TABELA 12 - Pontos extremos das unidades de amostragem de completude.....	108
TABELA 13 - Parâmetros do ajuste do variograma teórico ao experimental da taxa de comissão de feições geográficas de hidrografia da BC250.....	109
TABELA 14 - Resultados da autovalidação do ajuste do variograma teórico ao experimental da taxa de comissão de feições geográficas de hidrografia da BC250....	110
TABELA 15 - Estatística descritiva referente à taxa de omissão de feições geográficas de hidrografia da BC250.....	112
TABELA 16 - Parâmetros do ajuste do variograma teórico ao experimental da taxa de omissão de feições geográficas de hidrografia da BC250.....	114
TABELA 17 - Resultados da autovalidação do ajuste do variograma teórico ao experimental da taxa de omissão de feições geográficas de hidrografia da BC250.....	115
TABELA 18 - Metadados de qualidade: taxa de comissão de feições geográficas de hidrografia da BC250.....	117
TABELA 19 - Metadados de qualidade: taxa de omissão de feições geográficas de hidrografia da BC250.....	118
TABELA 20 - Níveis de conformidade para a acurácia posicional absoluta.....	120
TABELA 21 - Síntese da classificação da BC250 segundo o PEC.....	120
TABELA 22 - Estatística descritiva referente à acurácia posicional absoluta (erro planimétrico) da BC250.....	122
TABELA 23 - Pontos extremos das unidades de amostras da acurácia posicional.....	123

TABELA 24 - Parâmetros do ajuste do variograma teórico ao experimental para acurácia posicional absoluta.....	125
TABELA 25 - Resultados da autovalidação do ajuste do variograma teórico ao experimental referente ao erro planimétrico da BC250.....	126
TABELA 26 - Metadados de qualidade de dados referentes à acurácia posicional absoluta da BC250.....	128

## RESUMO

SANTOS, Alex da Silva, M. Sc., Universidade Federal de Viçosa, outubro de 2013. **Geoestatística aplicada na avaliação da qualidade de dados geoespaciais.** Orientadora: Nilcilene das Graças Medeiros. Coorientadores: Gérson Rodrigues dos Santos e Jugurta Lisboa Filho.

No processo de produção cartográfica é natural a busca pela qualidade do conjunto de dados espaciais. Ao longo do tempo, a evolução tecnológica propiciou aos profissionais da área um crescimento constante desta produção e na demanda destas informações geoespaciais por parte dos usuários. Entretanto, para acompanhar esta evolução, sem perder a qualificação dos dados espaciais, é necessária a adoção de padrões. Além disto, a complexidade das informações geográficas, em diversos aspectos, demanda técnicas de amostragem para mensurar e aferir a qualidade do conjunto de dados espaciais, a fim de garantir com embasamento estatístico, uma quantidade significativa e uma adequada distribuição espacial de suas amostras. Na área de Geociências é intuitivo pensar na correlação espacial de um fenômeno em estudo e a Geoestatística dispõe de ferramentas para identificar e representar o comportamento desta dependência. A análise espacial dos resultados de uma inspeção de qualidade de um produto cartográfico raramente é abordada na literatura, restrita apenas à conclusões descritivas e tabulares, baseadas no pressuposto da Estatística Clássica de independência dos dados observados. No presente momento está em implantação a Infraestrutura Nacional de Dados Espaciais do Brasil (INDE), onde diversos produtos cartográficos são disponibilizados à sociedade, junto com seus metadados geográficos. Os metadados facilitam o acesso ao dado espacial, mas atendem em parte aspectos relativos à qualidade do conjunto de dados. Desta forma, a presente dissertação propõe uma metodologia de inspeção da qualidade de conjuntos de dados espaciais da INDE, baseada em normas internacionais, conceitos de generalização e elementos de qualidade definidos na literatura, bem como a análise espacial dos resultados desta inspeção, usando métodos geoestatísticos e a representação desta aferição por meio de mapas de qualidade. O método de avaliação da qualidade de dados foi desenvolvido e aplicado sobre a Base Cartográfica Contínua do Brasil na escala 1:250.000 - BC250, com foco em determinados aspectos de qualidade. Os mapas de qualidade gerados mostraram regionalizações dos indicadores de qualidade inspecionados, confirmadas pela equipe produtora da referida base cartográfica, subsidiando usuários e os próprios produtores no entendimento do comportamento espacial da qualidade do produto cartográfico em estudo.

## ABSTRACT

SANTOS, Alex da Silva, M. Sc., Universidade Federal de Viçosa, October, 2013. **Geostatistics applied in assessment of geospatial data quality.** Adviser: Nilcilene das Graças Medeiros. Co-advisers: Gérson Rodrigues dos Santos and Jugurta Lisboa Filho.

In the process of production cartographic is natural the pursuit of quality spatial dataset. Throughout time technological evolution provided at the professionals in the area a constant growth in this production and in demand of these geospatial information by the users. However, to support this evolution without losing the qualification of space data is necessary to adopt standards. Furthermore, the complexity of geographic information in many aspects, require sampling techniques to measure and assess the quality of the spatial dataset in order to ensure with statistical basis, a significant amount and an appropriate spatial distribution of its samples. In Geosciences is intuitive to think in spatial correlation of a phenomenon under study and Geostatistics provides tools to identify and represent the behavior of this dependence. The spatial analysis of results from an inspection of a quality cartographic product is seldom addressed in literature, restricted only to conclusions descriptive and tabular, based on the assumption of classical statistical of independence from the observed data. At the present time is in deploying the National Spatial Data Infrastructure of Brazil (INDE), where several cartographic products are made available to the society, along with their geographic metadata. The metadata facilitates access to spatial data, but serve partly aspects concerning the quality of the dataset. Therefore, this thesis proposes a methodology for the quality inspection of spatial datasets of the INDE, based on international standards, concepts of generalization and elements of quality defined in literature as well as the spatial analysis of results of this inspection by using geostatistical methods and the representation of this assessment by means of quality maps. The method of quality evaluation data developed and applied on the Base Cartographic Continuous of Brazil at the 1:250.000 scale - BC250 with a focus on certain aspects of quality. The quality maps generated showed regionalizations in the indicators of quality inspected, confirmed by the staff that producer of cartographic base, subsidizing the producers themselves and users in understanding the spatial behavior quality product cartographic under study.

## 1. INTRODUÇÃO

Desde quando os primeiros cartógrafos, como Mercator, demonstraram sua preocupação com a representação adequada dos continentes e mares para permitir a rota de navios, usuários e produtores de dados geoespaciais<sup>1</sup>, esforçam-se para quantificar, modelar e melhorar a qualidade dos dados (STEIN *et al.*, 2009).

O processo cartográfico incorporado aos meios computacionais possibilitou o número crescente de produtores e principalmente de usuários de dados espaciais. Para atender as demandas de mapeamento, com produtos de qualidade satisfatória e acompanhar a evolução tecnológica, normas e parâmetros mínimos devem ser adotados (DALMOLIN e LEAL, 2001).

A transição do processo de produção cartográfica do meio analógico para o digital trouxe uma série de vantagens, dentre as quais destacam-se o aumento da capacidade produtiva e da demanda por informações espaciais. Entretanto, embora seja reconhecido o valor dos dados geoespaciais, tanto pelo governo quanto pela sociedade, a utilização eficiente dos mesmos é inibida por seu desconhecimento, pela falta de documentação e padronização do conjunto de dados, além da qualidade inadequada, conforme constatado na Conferência das Nações Unidas sobre o Meio Ambiente e Desenvolvimento - CNUMAD, realizada em 1992, constituindo um desafio na implantação de uma Infraestrutura de Dados Espaciais - IDE (CONCAR, 2010).

O termo NSDI (*National Spatial Data Infrastructure*), de abrangência nacional, apareceu pela primeira vez em 13 de abril de 1994, e seu conceito foi estendido à definição da GSDI (*Global Spatial Data Infrastructure*), de amplitude mundial. O conceito de Infraestruturas de Dados Espaciais - IDE engloba tecnologia, políticas, padrões e recursos humanos necessários para adquirir, processar, armazenar, distribuir e melhorar à utilização de dados geoespaciais (BILL, 1994).

Num esforço conjunto de instituições públicas brasileiras, produtoras e usuárias de informações geoespaciais, foi instituída no dia 27 de novembro de 2008 através da publicação do Decreto Presidencial n.º 6.666 a Infraestrutura Nacional de Dados Espaciais - INDE, definida como:

---

<sup>1</sup> Segundo o Art. 2º, do Decreto Lei n.º 6.666 de 27 de novembro de 2008, entende-se por "*dado ou informação geoespacial: aquele que se distingue essencialmente pela componente espacial, que associa a cada entidade ou fenômeno uma localização na Terra, traduzida por sistema geodésico de referência, em dado instante ou período de tempo, podendo ser derivado, entre outras fontes, das tecnologias de levantamento, inclusive as associadas a sistemas globais de posicionamento apoiados por satélites, bem como de mapeamento ou de sensoriamento remoto*" (BRASIL, 2008).

*"um conjunto integrado de tecnologias, políticas, mecanismos e procedimentos de coordenação e monitoramento, padrões e acordos, necessário para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal" (CONCAR, 2010).*

Depois da publicação do Decreto Presidencial n.º 6.666, a Comissão Nacional de Cartografia - CONCAR, com representantes a nível Federal, Estadual e Municipal, elaborou o Plano de Ação da INDE, sendo um instrumento norteador, estruturado em 3 (três) ciclos, compostos de prioridades e metas: Ciclo I - até dezembro de 2010; Ciclo II - de 2011 até 2014 e Ciclo III - de 2015 até 2020 (CONCAR, 2010).

No primeiro ciclo, a meta da INDE foi a implantação dos requisitos mínimos de hardware, software, telecomunicações e instalação do Diretório Brasileiro de Dados Geográficos - DBDG, além do Portal SIG Brasil com ferramentas de busca, exploração, acesso aos dados e metadados geoespaciais. No segundo e atual ciclo a meta é tornar a INDE a principal ferramenta de busca, exploração e acesso de dados e metadados geoespaciais do Brasil, a consolidação do DBDG em nível de governo federal e extensiva aos níveis estadual e municipal, além da sedimentação de normas e padrões. No terceiro e último ciclo espera-se que a INDE esteja ramificada em todos os setores produtivos da sociedade, além do governo, e consolidada como uma referência para busca, exploração e acesso de dados e metadados geoespaciais no Brasil, assim como integrada com outras IDE (CONCAR, 2010).

Em Robinson *et al.* (1995) há uma previsão do cenário atual das IDE, em relação a preocupação com padrões, o aumento de unidades locais alimentando dados centrais a nível estadual, nacional e global, a forma diretora de organizações responsáveis pelo conjunto de normas, facilidade no acesso e distribuição dos dados, a fim de atender a qualidade, o intercâmbio e a interoperabilidade de hardwares e softwares, conforme preconiza a INDE do Brasil.

Desta forma, a INDE tem o papel de catalogar, integrar e harmonizar os dados existentes nas instituições públicas do Brasil. A proposta do Plano de Ação, aliada a adoção do Perfil de Metadados Geoespaciais do Brasil, facilita o acesso às informações espaciais produzidas, mas atende em parte a evolução do processo de qualidade do dado espacial que é a adoção de padrões (ARIZA, 2002).

Os metadados geoespaciais facilitam a localização, identificação e até mesmo prevê a descrição qualitativa e quantitativa da qualidade de um conjunto de dados geoespaciais. Entretanto, devido a complexidade e a própria natureza do produto cartográfico é necessário a adoção de padrões para aferir e garantir sua qualidade, assim como possibilitar o acompanhamento das constantes evoluções tecnológicas. Segundo

Galo e Camargo (1994); e Nogueira Junior (2003) o controle de qualidade de um produto cartográfico é um procedimento extremamente importante e não se conhece efetivamente como tem sido realizado no Brasil.

A série de normas ISO 19100 são destinadas a informações geográficas, especificamente as normas ISO 19113, 19114 e 19115 abordam a qualidade do conjunto de dados espaciais. Todas são de grande valia para área de Geociências, mas os resultados da inspeção de qualidade, completa ou por amostragem, propostas por estes padrões, são meramente descritivos e tabulares. A representação espacial dos resultados da inspeção de qualidade, apesar de trivial e intuitiva, raramente é abordada na literatura, assim como acontece com o Padrão de Exatidão Cartográfico – PEC. Esta representatividade da qualidade é definida pela análise espacial da variável em estudo, identificando áreas correlacionadas em relação ao fenômeno analisado. Em função da constatação desta demanda de representação espacial da inspeção, além da aferição da qualidade baseada nos padrões da ISO, adotou-se neste trabalho a Geoestatística e suas ferramentas, a fim de possibilitar a identificação e visualização da correlação espacial, e o uso de metadados de qualidade para relatar os procedimentos de avaliação de forma sintetizada e facilitar o acesso a esta informação por parte dos usuários do produto cartográfico.

Desenvolver uma metodologia para avaliar e relatar a qualidade de um conjunto de dados geoespaciais, contribuir com os propósitos da INDE, e aproximar usuários, produtores e a realidade representada no produto cartográfico, motivaram a presente pesquisa composta por conceitos de qualidade aplicados a dados espaciais, generalização cartográfica, técnicas de amostragem, métodos geoestatísticos e metadados.

### **1.1. Objetivo**

O objetivo principal desta pesquisa é estudar o uso da Geoestatística na identificação de prováveis regionalizações da qualidade posicional e de completude, aliada a adoção de normas internacionais, para avaliação da qualidade de produtos cartográficos e a representação espacial do resultado desta aferição. Os objetivos específicos incluem: a elaboração de uma metodologia para inspeção da qualidade de um produto cartográfico, com base em normas internacionais; estudo de técnicas de amostragem e de métodos geoestatísticos; estabelecer o preenchimento de metadados de qualidade, focando em Projetos de Bases Cartográficas Contínuas da Coordenação de Cartografia (CCAR) do IBGE, componentes da INDE.

## 1.2. Motivação

Dentro de uma própria instituição pode haver produtores e usuários de dados geoespaciais, com seus pontos de vistas distintos sobre a qualidade do dado geoespacial. Geralmente, os produtores focam na qualidade posicional, semântica, consistência lógica, temporal e completude dos dados; e os usuários, por outro lado, preocupam-se o quanto tais dados são adequados ao seu uso, considerando aspectos não previstos pelo produtor.

Independente do ponto de vista adotado, o controle sobre a qualidade de dados, é fundamental, tanto no processo de produção como na aplicação dos produtos cartográficos gerados. As demandas da INDE, aliada a padronização e normalização do modelo de dados da Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV), assim como os metadados, são indicativos da necessidade de controle de qualidade na produção do dado cartográfico.

O uso de padrões internacionais na produção cartográfica, tanto durante o processo de produção do conjunto de dados espaciais quanto na avaliação do produto cartográfico final, pode agregar uma certificação à qualidade dos dados representados na INDE e por consequência uma melhoria na informação geoespacial disponível.

No processo de atualização ou concepção de bases cartográficas geoespaciais é inerente almejar sua qualidade. Entretanto, nesta busca de melhoria da qualidade do conjunto de dados espaciais, deve-se estabelecer e considerar parâmetros essenciais, relativos ao "o que" e "como avaliar". A premissa é a acurácia posicional e semântica, oriunda desde o mapeamento analógico, mas outros fatores, pertinentes ao próprio desenvolvimento tecnológico, devem ser levados em consideração, como a consistência lógica do conjunto de dados espaciais.

O conceito e a evolução da concepção de IDE abordam a interoperabilidade de dados, onde a informação geoespacial compartilha características que são comuns a usuários diferentes. A ideia do uso comum vem em paralelo à adoção de padrões e normas, tanto na descrição e compartilhamento de dados, como na evolução tecnológica da área de Geociências, como as de Geoprocessamento. Logo, o estudo da adoção de padrões internacionais relativos a qualidade de dados espaciais, a identificação de parâmetros e fatores para avaliação, a análise espacial da qualidade, a descrição do processo de atualização e avaliação do conjunto de dados espaciais, a interação com os diversos usuários são de fundamental importância para o cenário atual.

## **2. FUNDAMENTAÇÃO TEÓRICA**

Este capítulo é dedicado a abordagem dos aspectos referentes ao tema qualidade de dados espaciais. Os tópicos perpassam conceitos e normas através do foco em aspectos e características dos dados espaciais, compondo um ciclo conceitual para auxiliar na elaboração da metodologia de avaliação da qualidade de dados geoespaciais. O capítulo é constituído de 5 (cinco) tópicos:

Qualidade de dados: uma abordagem geoespacial, onde são abordados temas desde a evolução do conceito de qualidade, normas referentes a qualidade de dados espaciais e seus elementos de qualidade.

Generalização cartográfica: princípios de cartografia aplicados a produção de mapas, onde são abordados conceitos e os operadores de generalização cartográfica;

Técnicas de amostragem, onde são abordados conceitos de Estatística Clássica, tamanho de amostra e planos de amostragem.

Geoestatística, onde são apresentados conceitos, suas ferramentas de análise: o variograma e a krigagem para identificar a dependência e correlação espacial do fenômeno espacial em estudo.

Metadados: informando a qualidade de dados ao usuário, a fim de mostrar conceitos e sua relevância para documentar e informar a avaliação de qualidade quantitativa dos dados espaciais.

### **2.1. Qualidade de dados: uma abordagem geoespacial**

Os conceitos de qualidade, sua concepção ao longo do tempo e aplicação de seus princípios são fundamentais na avaliação do conjunto de dados espaciais. Desta forma, para abordar a qualidade de dados, deve-se considerar a descrição dos elementos de qualidade inerentes a conjunto de dados espaciais, passando pelas perspectivas de qualidade do produtor e do usuário de informações espaciais, até a situação atual de desenvolvimento de normas internacionais de qualidade e sua adoção em nível nacional.

Nota-se que quando o termo qualidade é abordado, independente de sua finalidade, a palavra padrão pode ser associada, seja na comparação e adequação de um modelo existente ou como referência num processo de aferição da qualidade de um conjunto de dados.

Um produto cartográfico é uma representação da realidade e a qualidade é a ligação entre o conjunto de dados espaciais e este mundo real. Desta forma, os conceitos de qualidade de dados espaciais e das próximas seções devem ser associados ao processo de produção cartográfica, como fonte de melhoria do conjunto de dados

gerado e identificação de inconsistências nesta relação entre a realidade e sua representação, compondo uma das partes essenciais na avaliação qualitativa do conjunto de dados espaciais.

O termo "qualidade" tem sua origem no latim "*Qualitas*", inspirado do grego "*ποιον*" e é atribuído ao filósofo Cícero, baseado em "*qualis*" que significa "o que", isto é, a natureza de algo. Entretanto, é atribuído ao engenheiro Frederick Winslow Taylor, as primeiras obras relacionadas ao domínio da qualidade, onde seus princípios construíram a base do "taylorismo". A evolução dos conceitos relacionados a qualidade e sua gestão seguiram com Henry Ford, William Edwards Deming e o conceito de Gestão da Qualidade Total (TQM - *Total Quality Management*), Joseph M. Juran, Armand V. Feigenbaum e mais posteriormente Shigeo Shing, Kaoru Ishikawa, Genichi Taguchi e Philip Crosby (DEVILLERS e JEANSOULIN, 2006; PEARSON, 2011).

O interessante ao observar o parágrafo anterior e os autores que perpassaram a evolução do conceito de qualidade e sua aplicação é notar o quanto todos ressaltaram sua importância, mas o quão divergentes foram suas definições sobre o tema. Para alguns a qualidade é relacionada a ausência de erro, ou a conformidade com as especificações utilizadas, para outros, é a necessidade de atender as demandas do cliente. Posteriormente, as obras destes autores foram padronizadas pela ISO (Comitê Técnico 176-1979) resultando na norma ISO 9000 em 1987, relativa a gestão da qualidade.

Nos dias atuais existe um consenso na definição de qualidade, como expressado em Ariza (2002), na norma ISO 19113:2002 e no Plano de Ação da INDE (CONCAR, 2010), onde é descrita como "*a totalidade de características de um produto ou serviço que lhe conferem aptidões para satisfazer necessidades explícitas e implícitas*".

Em paralelo a evolução dos conceitos de qualidade e sua gestão, a aplicação da estatística no controle de qualidade surge praticamente na mesma época. Em razão do aumento da capacidade produtiva, a inspeção completa dos produtos tornou-se cada vez mais inviável, houve demanda pela inspeção por amostragem e o uso de ferramentas estatísticas (PEARSON, 2011).

### ***A organização ISO***

A ISO (*International Organization for Standardization*) é uma federação mundial dos organismos nacionais de normalização (organismos membro da ISO), com origem no campo da eletrotécnica através de sua precursora a IEC (*International*

*Electrotechnical Commission*) criada em 1906<sup>2</sup>. O termo ISO foi escolhido intencionalmente com o objetivo de padronizar a sigla em diferentes idiomas, sua origem vem do grego "isos" e significa "igual" (PEARSON, 2011).

No ano de 1979 foi criado o Comitê Técnico - TC-176 da ISO para tratar de qualidade. Este Comitê Técnico padronizou as obras relativas a qualidade, conforme citado nesta dissertação, e publicou a ISO 9000 em 1987, relativa a gestão de qualidade (MARIANI, 2006).

Em 1994 foi criado o Comitê Técnico - TC-211 da ISO para tratar de informações geográficas, com o objetivo de estabelecer um conjunto estruturado de normas para informações sobre objetos ou fenômenos que estão direta ou indiretamente associados a uma localização relativa à Terra. A família de normas ISO 19100 abordam especificamente a informação geográfica (BURITY e SILVEIRA, 2012; ARIZA e PASCUAL, 2008).

Atualmente, existem diversos organismos envolvidos ativamente nos trabalhos da ISO/TC 211. Entre estes incluem organismos nacionais de normalização, o Consórcio OpenGIS (OGC), organismos internacionais profissionais (como FIG - *International Federation of Surveyors* e ICA - *International Cartographic Association*), agências da ONU (Organização Mundial das Nações Unidas) e organismos setoriais (como DGIWG- *Defence Geospatial Information Working Group* e ICAO - *International Civil Aviation Organization*) (ISO, 2012).

A ISO é formada por representantes dos países-membros. No Brasil, a ABNT (Associação Brasileira de Normas Técnicas) é a entidade responsável por representar o país perante a ISO.

O INMETRO (Instituto Nacional de Metrologia, Normalização e Qualidade Industrial) é o organismo de acreditação para o Brasil, ligado à ABNT. O INMETRO é a entidade responsável em determinar as diretrizes dos Organismos Credenciados de Certificação (OCC) na emissão de certificados ISO para entidades corporativas.

Existe um Fórum Internacional de Acreditação (*International Accreditation Forum* - IAF) que é uma associação mundial de Organismos de Acreditação, Associações de Organismos de Certificação e outras organizações envolvidas em atividades de avaliação de conformidade em uma variedade de campos, incluindo sistemas de gestão, produtos, serviços e pessoal.

Segundo o INMETRO, a acreditação representa o reconhecimento formal da competência técnica de organizações que realizam a avaliação da conformidade, sendo

---

<sup>2</sup> Mais informações em <<http://www.iso.org/>>

uma maneira segura de identificar aqueles que oferecem a máxima confiança em seus serviços. A certificação de produtos ou serviços, sistemas de gestão e pessoas é o procedimento realizado por uma organização independente acreditada para avaliação da conformidade destes com os requisitos especificados.

Os OCCs são encarregados de interpretar a norma ISO 9001, avaliar sua aplicação às situações de negócio da companhia e determinar se o sistema de qualidade da empresa está em conformidade com a norma, para posterior certificação. Atualmente, existe certificação apenas para as normas ISO 9001 e ISO 14000, através dos OCCs.

### **2.1.1. Qualidade do produto e qualidade do processo**

A ideia de controlar um processo para conseguir melhorias é totalmente diferente da ideia de inspecionar produtos para identificar os não conformes, embora os dois procedimentos utilizem em parte as mesmas ferramentas estatísticas (JURAN, 1997).

Segundo Pascual (2012) existem dois aspectos de qualidade que surgem de pontos de vista completamente diferentes na hora de abordar o problema: a qualidade do produto versus a qualidade do processo. O primeiro, que considera a qualidade do produto final, visa determinar de forma estatística e objetiva a qualidade do produto, sem preocupar-se com o processo de produção, que se configura numa autentica caixa preta na qual interessa somente a saída, ou seja, o resultado final. O segundo, que considera a qualidade do processo, visa na maior ou menor adequação do processo de produção e fornecimento de um bem ou de um serviço e, num sentido mais geral, da qualidade da organização produtora, e por consequência, se processo de produção for correto e adequado, logicamente o resultado também o será. Além disto, inclui uma série de aspectos de grande relevância que envolve o bem produzido: prazo de entrega, garantia, serviço pós-venda, atualização do produto, preço, instruções de uso, e outros de interesse do produtor, dos usuários e de suas aplicações.

### **2.1.2. Visão produtor e visão do usuário**

A qualidade de dados espaciais depende internamente da percepção do produtor, e externamente da perspectiva do usuário. Para os produtores os fatores que determinam a qualidade são: a precisão geométrica e semântica, a genealogia dos dados, a consistência lógica, bem como a integridade dos dados. A preocupação do usuário, por outro lado, é a adequação ao uso, ou o nível de aptidão entre os dados e suas necessidades, definidos em termos de acessibilidade, interpretação, relevância,

completude, facilidade de compreensão e custo (MOSTAFAVI *et al.*, 2004). A Figura 1 ilustra a percepção destes dois pontos de vista.

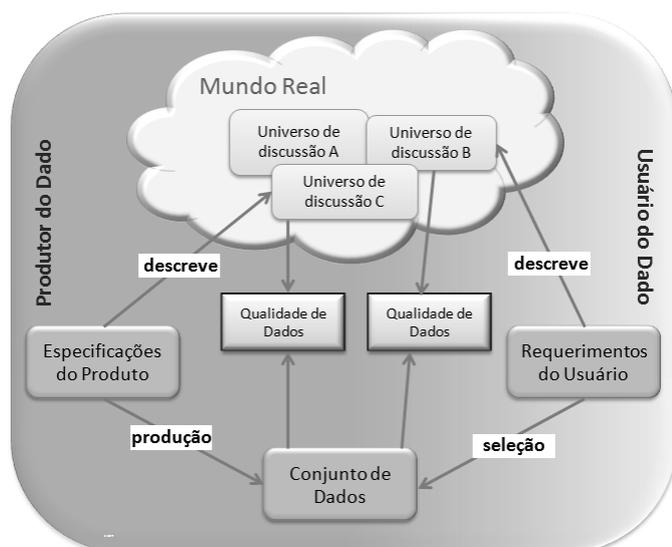


Figura 1 - Aspectos gerais de qualidade de dados espaciais.

Fonte: adaptado da norma ISO 19113:2002

Segundo Nogueras (2005), usuários de informação geográfica digital *online*, tendem a confiar nos dados provenientes de instituições respeitáveis, sem preocupar-se com garantias de qualidade documentadas e a procedência dos insumos de origem, assim como eles fazem quando adquirem os dados *offline*. O autor ressalta o uso dos metadados como um peça essencial para o desenvolvimento de uma Infraestrutura de Dados Espaciais.

### 2.1.3. Qualidade de dados em Cartografia

Atualmente, na era da Tecnologia da Informação (TI), há uma diversidade de ferramentas ligadas a produção de dados espaciais, propiciando um aumento da capacidade produtiva e em sua qualidade. Em relação a identificação e tratamento de inconsistências em dados na estrutura matricial (*rasters*) em um ambiente SIG (Sistema de Informação Geográfica) existe, na literatura, uma maturidade tecnológica e metodológica. No entanto, ocorre o contrário em relação ao tratamento de inconsistências de dados na estrutura vetorial, incluindo a modelagem de erros em feições com geometrias do tipo pontos, linhas e polígonos, bem como a modelagem de propagação de incertezas. Para complementar, alguns softwares para análise, avaliação e visualização da qualidade de dados de SIG tem sido desenvolvidos (STEIN *et al.*, 2009).

Segundo Pascual (2012), no manuseio de um conjunto de dados geográficos a qualidade é tida como a diferença entre a medida do dado e a realidade que este

representa, ou seja, a qualidade ancora os dados a realidade, descreve e dá ideia de um vácuo existente entre o modelo do mundo real e o próprio mundo real.

#### **2.1.4. Normas de qualidade de dados espaciais**

O Perfil de Metadados Geoespaciais do Brasil é baseado na norma ISO 19115:2003 (*Geographic Information – Metadata*), desenvolvido pelo Comitê Técnico 211 (TC/211) da ISO (the *International Organization for Standardization*), sendo parte de um grupo de especificações internacionais para garantir a qualidade da Informação Geográfica (IG). Em relação à qualidade de dados geoespaciais, existem outras normas internacionais vinculadas a ISO 19115:2003.

A norma ISO 19113:2002, relativa aos Princípios de Qualidade, especifica componentes para retratar a qualidade de dados espaciais e conceitos para manipulação da informação gerada. A norma ISO 19114:2003 é relativa aos Procedimentos para Avaliação da Qualidade, que consiste em orientações sobre métodos para especificar ou avaliar a qualidade de dados espaciais (BURITY e SILVEIRA, 2012).

A norma ISO 19138:2006 define os componentes e a estrutura do conteúdo para registrar um conjunto de medidas de qualidade dos dados referentes aos elementos de qualidade identificados na norma ISO 19113:2002. Os metadados de qualidade seguem orientações das normas ISO 19114:2003 e ISO 19115:2003 e foram preenchidos nesta pesquisa.

Outras normas internacionais relacionadas à qualidade de dados espaciais são<sup>3</sup>: a ISO 19122 referente a Qualificação e certificação de pessoal, a ISO 19127 relativa a Códigos e parâmetros geodésicos, a ISO 19131 sobre a Especificação de Produto dos dados geográficos e a ISO 19139 a respeito da implementação do esquema de Metadados em XML (*eXtensible Markup Language*).

As normas ISO de qualidade de dados espaciais encontram-se em constante desenvolvimento. As mesmas apresentam desde informações conceituais até exemplos práticos relatando a aplicação de medidas de qualidade, baseadas nos elementos de qualidade e nas especificações técnicas do conjunto de dados.

As normas são relacionadas entre si, e a aplicação de uma geralmente recai sobre requisitos presentes em outra. A própria norma 19115:2003 relativa aos metadados está vinculada as demais normas de qualidade do dado espacial. A Figura 2 mostra uma

---

<sup>3</sup> Em pesquisa bibliográfica foi encontrada a norma ISO 19157 (*Geographic Information - Data Quality*), ainda não homologada. Segundo informações da ISO foi enviada uma prova desta norma, em 20 de agosto de 2013, ao secretariado para votação de sua aprovação em 2 (meses). Esta norma é uma revisão das normas de qualidade de dados espaciais ISO 19113:2002, ISO 19114:2003 e ISO 19138:2006. Mais informações em <<http://www.iso.org>>

visão geral das informações da qualidade de dados, sendo uma parte não quantitativa e outra quantitativa, presentes nas normas ISO 19113, 19114 e 19115.

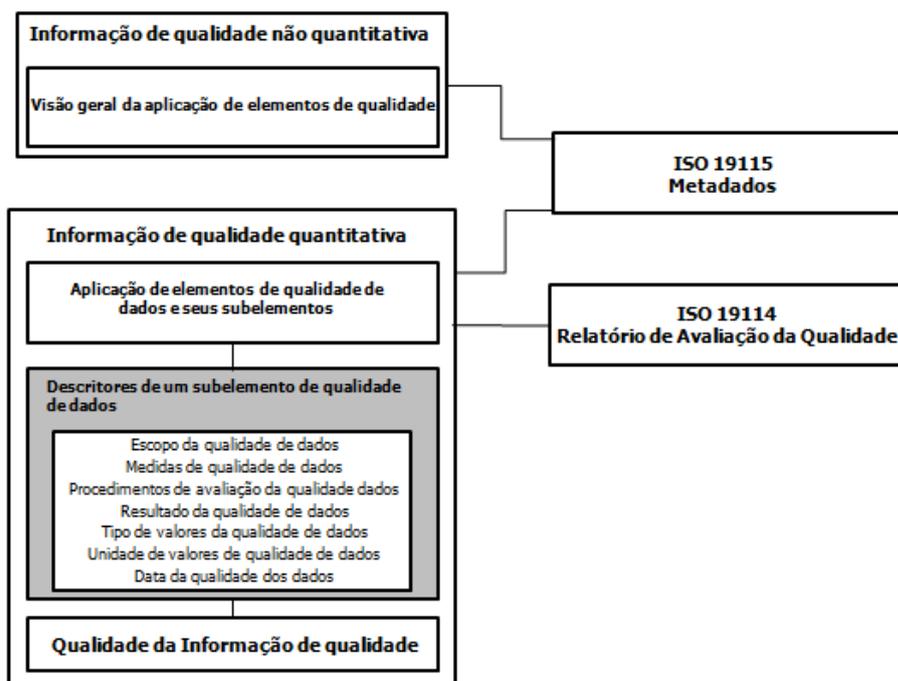


Figura 2 - Visão geral de informações da qualidade de dados.

Fonte: adaptado da Norma ISO 19113:2002

Em Ceballos e Gatica (2011), há um estudo onde as normas ISO 19113, 19114 e 19138, específicas sobre qualidade de dados espaciais, junto com a norma ISO 19115 referente aos metadados geoespaciais, formam uma pirâmide de implementação das normas, tendo como base a norma ISO 9001 que aborda o gerenciamento de sistemas de qualidade, atendendo assim os usuários, conforme ilustra a Figura 3, numa visão de controle do processo de produção cartográfica.



Figura 3 - Pirâmide de implementação de normas ISO.

Fonte: adaptado de Ceballos e Gatica (2011)

#### 2.1.4.1. Normas, padrões e especificações para mapeamento da INDE

Para a evolução da qualidade num processo de produção é essencial a adoção de padrões. Neste sentido, uma Infraestrutura de Dados Espaciais pressupõe qualidade, a elaboração e adoção de padrões fornece condições para controlar o processo de produção de dados espaciais e estabelece níveis de conformidade de qualidade dos produtos cartográficos gerados. A INDE do Brasil está em fase de consolidação, além das normas existentes, outras estão sendo elaboradas, conforme Tabela 1 (ARIZA, 2002; CONCAR, 2010, LUNARDI, 2012).

Ressalta-se que de acordo com o Decreto Lei 243 de 28 de fevereiro de 1967 é de competência da DSG (Diretoria de Serviço Geográfico do Exército) o estabelecimento de Normas Técnicas para a cartografia brasileira na série de cartas gerais, das escalas 1:250.000 e maiores.

TABELA 1 - Normas, padrões e especificações para mapeamento da INDE

ESTRUTURA DO DADO	SIGLA	FINALIDADE	SITUAÇÃO/ INSTITUIÇÃO EXECUTORA
Matricial	ET-PCDG	Padrões dos produtos de conjuntos de dados geoespaciais do tipo carta matricial	Em elaboração / DSG
	ET-EDGV	Modelo conceitual p/ dados vetoriais garantindo a consistência lógica (supre consistência conceitual e dos domínios alfanuméricos)	Elaborada/ CONCAR/ CEMND
Vetorial	ET-ADGV	Regras de aquisição da geometria dos dados garantindo a consistência lógica do atributo geometria e consistência topológica.	Elaborada/ DSG
	ET-PCDG	Padrões dos produtos de conjuntos de dados geoespaciais vetoriais	Em elaboração / DSG
	ET-RDG	Garante a consistência na representação das mesmas classes de objetos.	Em elaboração / DSG
	ET-CQPCDG	Procedimentos para o controle de qualidade dos produtos de conjuntos de dados geoespaciais vetoriais	Em elaboração / DSG

Fonte: CONCAR (2010)

#### 2.1.4.2. Padrão de Exatidão Cartográfica (PEC)

No Brasil, a avaliação da qualidade posicional é baseada no Padrão de Exatidão Cartográfica - PEC, instituído no Decreto Lei 89.817 de 20 de junho de 1984. Este foi concebido para aferição de produtos cartográficos analógicos (BRASIL, 1984).

No Decreto Lei 89.817 além do PEC, constam outros fatores de qualidade gerais a respeito de cartas analógicas, como a presença de grade de coordenadas, escala, projeção, classificação planimétrica e altimétrica, entre outros itens.

Os referenciais planimétrico e altimétrico para a Cartografia Brasileira são aqueles que definem o Sistema Geodésico Brasileiro - SGB, conforme estabelecido pela Fundação Instituto Brasileiro de Geografia e Estatística - IBGE, em suas especificações e normas (BRASIL, 2005). A Tabela 2 sintetiza o PEC planimétrico e altimétrico.

TABELA 2 - Padrão de Exatidão Cartográfica - PEC

<b>Padrão de Exatidão Cartográfico Planimétrico</b>		
<b>Classe</b>	<b>PEC (mm x DE)</b>	<b>EP (mm x DE)</b>
A	0,5	0,3
B	0,8	0,5
C	1,0	0,6
<b>Padrão de Exatidão Cartográfico Altimétrico</b>		
<b>Classe</b>	<b>PEC (mm x DE)</b>	<b>EP (mm x DE)</b>
A	1/2 eq*	1/3 eq*
B	3/5 eq*	2/5 eq*
C	3/4 eq*	1/2 eq*

onde:  
 DE - Denominador da escala da carta;  
 EP - Erro padrão;  
 \*Equidistância entre as curvas de níveis

Fonte: Brasil (1984).

### ***Conceitos de erro padrão, desvio padrão e erro médio quadrático abordados no PEC***

O Decreto Lei nº 89.817 (BRASIL, 1984) considera equivalentes os termos "erro padrão" (EP), "desvio padrão" ( $\sigma$ ) e "erro médio quadrático" (EMQ), sem apresentar as formulações matemáticas para o cálculo, ocasionando interpretações dúbias. O desvio padrão e o erro padrão retratam a variabilidade dos dados em torno da média aritmética do espaço amostral. O EMQ está associado à diferença entre um valor observado, referente ao conjunto de dados espaciais, e o tomado como referência para mensurar o erro planimétrico ( $e_p$ ) e/ou altimétrico ( $e_a$ ). As Equações 1, 2 e 3 são formulações para o cálculo do desvio padrão, erro padrão e EMQ respectivamente (OLIVEIRA, 2011):

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (e_{pi} - \bar{e}_p)^2}{n-1}}, \quad (1)$$

$$EP = \frac{\sigma}{\sqrt{n}}, \quad (2)$$

$$EMQ = \sqrt{\frac{\sum_{i=1}^n (e_{pi})^2}{n-1}}, \quad (3)$$

onde:

- $e_{pi}$  erro planimétrico de um ponto identificado no conjunto de dados espaciais.
- $\bar{e}_p$  média amostral do erro planimétrico.
- $n$  número de pontos observados.

Ainda segundo Oliveira (2011) a classificação da acurácia posicional de produtos cartográficos deve atender aos dois critérios do PEC simultaneamente, sendo o EMQ associado à precisão e o PEC à tendência dos dados observados. Conforme recomendado pelo autor nesta dissertação será adotado o EMQ para considerar o erro padrão do produto cartográfico avaliado.

### ***Acurácia versus precisão***

Segundo Gemael (1994), nos textos de língua inglesa existem os dois termos *accuracy* e *precision*, e que os mesmos não são sinônimos, sendo traduzidos respectivamente por "acurácia" e "precisão". O termo precisão é vinculado a efeitos aleatórios, ou seja, a dispersão das observações, enquanto que acurácia vincula-se a efeitos aleatórios e sistemáticos.

Segundo Mônico *et al.* (2009), a acurácia reflete a proximidade de uma grandeza estatística à estimativa de um parâmetro (ou valor verdadeiro) e a precisão expressa o grau de aderência desta grandeza observada com sua média. Ressalta-se que a precisão considera apenas efeitos aleatórios, enquanto a acurácia avalia os efeitos aleatórios e sistemáticos. Na ausência de um valor de referência para avaliar a tendência, a alternativa é utilizar o valor da precisão como um indicativo da acurácia, neste caso a tendência é desconhecida, mas pode existir. A Figura 4 sintetiza os conceitos de precisão e acurácia.

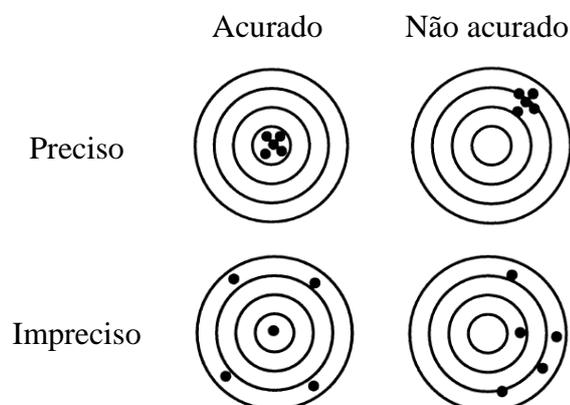


Figura 4 - Acurácia versus precisão.

Fonte: adaptado de Ghilani (2010).

### 2.1.5. Elementos de qualidade de dados espaciais

Dentre os componentes para avaliação da qualidade de um conjunto de dados espaciais, a norma ISO 19113:2002 agrupa estes em elementos e subelementos de qualidade. Os elementos de qualidade são: completude, consistência lógica, acurácia posicional, acurácia temporal e acurácia temática, e para cada um destes elementos é vinculado um conjunto de subelementos, conforme apresentado a seguir.

#### 2.1.5.1. Completude

A completude indica o grau em que os objetos representados em um conjunto de dados relacionam-se com as instâncias de classes de um universo abstrato. Este parâmetro de qualidade descreve a presença e/ou ausência de instâncias de feições, relacionamentos e seus atributos, numa relação entre os objetos representados no conjunto de dados espaciais e os que conformam com o modelo cartográfico da realidade. O seu conceito é de difícil medição, existindo pouco desenvolvimento e discussão sobre sua definição formal (ARIZA, 2002).

O conceito de completude subdivide-se em: omissão e comissão, sendo afetado por regras de generalização e escala. Os casos de omissão consideram que as entidades são removidas frente as entidades da realidade e os casos de comissão são entidades falsas frente as entidades à representar no conjunto de dados. Em resumo a completude monitora a falta (omissão), assim como o excesso (comissão) de informação presentes no conjunto de dados, conforme mostra a Figura 5 (DEVILLERS e JEANSOULIN, 2006).

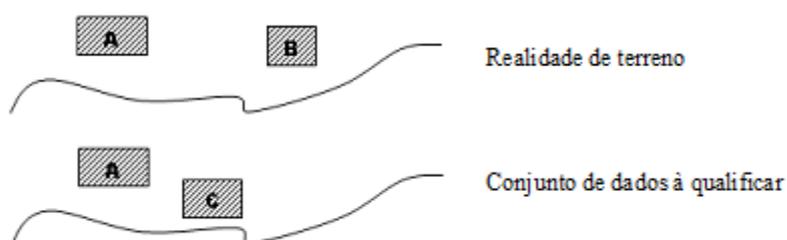


Figura 5 - Exemplos de omissão (B) e comissão (C).

Fonte: adaptado de Servigne *et al.* (2006).

Segundo Ariza (2004), a completude deve ser entendida em função das demandas ou necessidades de um produto cartográfico, na qual sua avaliação é a conformidade com um padrão presente nas especificações técnicas do conjunto de dados e não uma comparação direta com o mundo real. A Tabela 3 descreve os níveis de abstração de dados geoespaciais, a completude é avaliada através da comparação com o terreno nominal, ou seja, suas especificações técnicas.

TABELA 3 - Níveis de abstração do mundo real ao terreno nominal

Mundo Real	Conjunto de fenômenos geoespaciais que podem ou não ser cognitivos pelo ser humano, é a realidade em sua completude.
Mundo conceitual ou universo abstrato	Conjunto de entidades percebidas da realidade, a linguagem natural de nomear as coisas abstrai a realidade e define uma referência para a obtenção do conjunto de dados.
Mundo Geoespacial	Contempla ferramentas para estabelecer uma nova simplificação da realidade e facilita o tratamento da informação, através de restrições e peculiaridades, adequando a implementação do mundo conceitual a um banco de dados geográficos.
Universo de discussão	Visão do mundo real, ou hipotético, contemplando o que é relevante para uma aplicação geoespacial concreta, adaptada pelo processo de abstração do mundo conceitual, das restrições do modelo geoespacial e das especificações do produto.
Terreno nominal	O terreno nominal contempla o conjunto ideal de dados geográficos, com regras especificadas ao qual se compara com o conjunto real para aferir sua qualidade, é a própria especificação técnica. Desta forma, a cada produto cartográfico, com suas distintas regras de implementação é recomendável a contestação de sua qualidade.

Fonte: adaptado de Ariza (2004)

A cartografia é dedicada a representar modelos do mundo real e a acuracidade de um produto cartográfico é determinada por sua conformidade com um modelo abstrato da realidade, a qual é descrita nas especificações técnicas do conjunto de dados espaciais ou terreno nominal. A Figura 6 mostra a relação entre o processo de abstração e a completude (ARIZA, 2002; LONGLEY *et al.*, 2013).

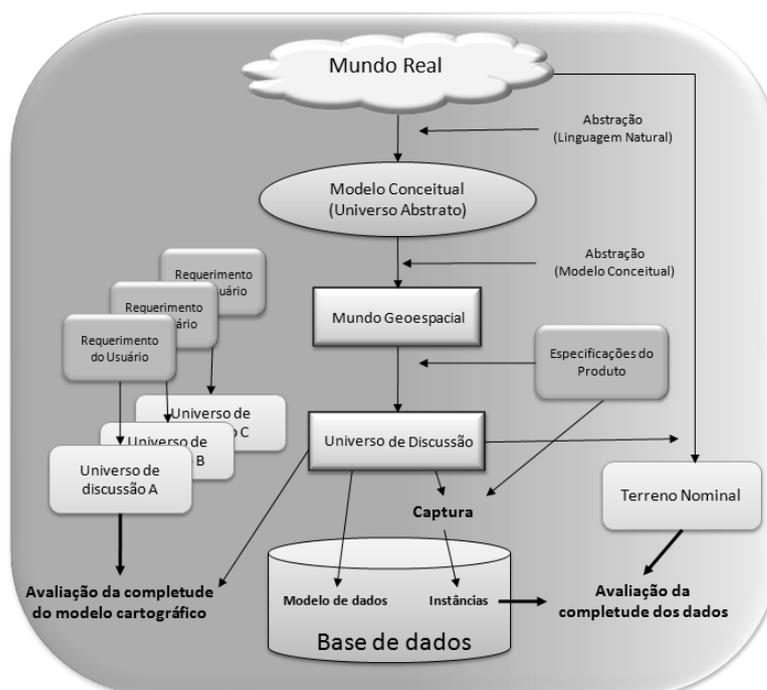


Figura 6 - Relação entre o processo de abstração da realidade e a completude.

Fonte: adaptado de Ariza (2004)

Na literatura, em relação ao termo "completude", pode-se encontrar distintos significados (Figura 7), como a completude dos dados (qualidade técnica), relativa aos objetos representados (omissão e comissão) na base de dados conforme prescrições estabelecidas e a completude do modelo cartográfico (qualidade teórica) referente a presença de classes, atributos e relacionamentos necessários a implementação adequada da realidade do modelo cartográfico da realidade (universo de discussão). Neste trabalho, durante o processo de avaliação do conjunto de dados será considerada apenas a completude dos dados, ou seja, a qualidade técnica (ARIZA, 2002).

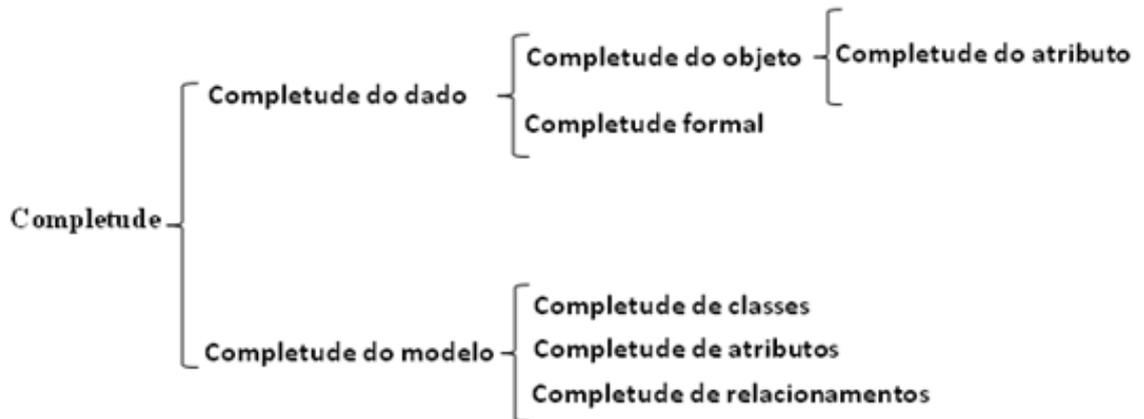


Figura 7 - Esquema de significados de completude

Fonte: Ariza (2004)

Uma característica entre os elementos de qualidade é a possibilidade da correlação dos mesmos, onde ocorrências de um podem influenciar o outro. Segundo Ariza (2002), no caso específico da completude, por exemplo, a ausência de feições ou atributos pode ocasionar inconsistências na avaliação da consistência lógica do conjunto de dado. Nestas situações o autor recomendava optar pela inclusão da ocorrência em um dos elementos, porém, em Ariza (2004), o mesmo autor recomenda que a sinalização seja feita em ambos os elementos, conforme adotado nesta dissertação.

Segundo Salgé (1995), a completude está incluída na acurácia temática, onde a taxa de comissão ( $\tau^+$ ) e taxa de omissão ( $\tau^-$ ) dos dados são fornecidas pelas Equações 4 e 5:

$$\tau^- = \frac{N^-}{\max(N, N^0)} \quad (\text{taxa de omissão}) \quad (4)$$

$$\tau^+ = \frac{N^+}{\max(N, N^0)} \quad (\text{taxa de comissão}) \quad (5)$$

onde:

- $N^-$  é o número de ocorrência que existem na realidade percebida e que não foram reconhecidas, ou seja, erro de omissão.
- $N^+$  é o número de ocorrências na amostra que não ocorrem na realidade percebida, ou seja, erro de comissão.
- $N^0$  é o número de ocorrências na realidade percebida.
- $N$  é o número de elementos da amostra, dado por:

$$N = N^0 + N^+ - N^- . \quad (6)$$

Os índices acima se aplicam tanto em questões relativas a completude quanto a acurácia temática.

### 2.1.5.2. Consistência lógica

A representação de fenômenos do mundo real é fundamental para qualquer processamento e entendimento do mesmo. O processo de modelagem de dados espaciais visa representar esta realidade num banco de dados geográficos, onde os objetos retratam as entidades do mundo real através do armazenamento de seus atributos (espaciais e descritivos) e relacionamentos espaciais relevantes (KAINZ, 1995; CASANOVA *et al.*, 2005, LISBOA FILHO, 2000).

A consistência lógica é o elemento de qualidade de dados espaciais cujo objetivo é avaliar a integridade estrutural de um conjunto de dados, através do grau de aderência deste conjunto de dados às regras lógicas do modelo de dados espacial utilizado. Inicialmente, foi utilizada em verificações de integridade de dados não espaciais, e sua aplicação para dados espaciais ocorreu na época das primeiras análises topológicas. A Figura 8, mostra uma situação relativa a consistência lógica (KAINZ, 1995; ARIZA, 2002; SERVIGNE *et al.*, 2006).

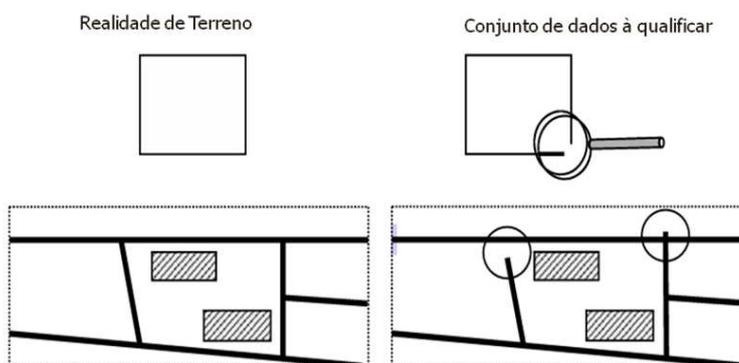


Figura 8 - Exemplo de avaliação da consistência lógica

Fonte: Devillers e Jeansoulin (2006)

Segundo a norma ISO 19113:2002 o elemento de qualidade consistência lógica subdivide-se em 4 (quatro) subelementos: consistência conceitual, consistência de domínio, consistência de formato e a consistência topológica. Todos são vinculados a regras de integridade estrutural do conjunto de dados.

A consistência conceitual analisa a aderência do conjunto de dados espaciais às regras do esquema conceitual, ou seja, ao universo de discussão. A consistência de domínio avalia a conformidade em relação a valores de domínios pré-estabelecidos. A consistência de formato considera o grau em que os dados são armazenados de acordo com a estrutura física do conjunto de dados, isto é, examina o modelo de armazenamento para uma estrutura de dados de um hardware específico ou ambiente de software. A consistência topológica refere-se aos aspectos geométricos e topológicos da informação espacial, como situações de adjacência e pertinência (BORGES, 1997).

### 2.1.5.3. Acurácia posicional

De acordo com Ariza (2002), a acurácia posicional é uma propriedade tradicional e emblemática das produções cartográficas, existindo um grande número de referências em relação ao tema. Esta refere-se a acuracidade planimétrica e altimétrica do conjunto de dados espaciais e segundo a norma ISO 19113:2002 existem 3 (três) tipos de acurácia posicional:

Acurácia absoluta ou externa: proximidade dos valores das coordenadas observadas com os valores aceitos como verdadeiros.

Acurácia relativa ou interna: proximidade dos valores de coordenadas observadas com a posição relativa de feições do conjunto de dados e suas respectivas posições aceitáveis ou consideradas como verdadeiras (Figura 9).

Acurácia de um *gride* de dados: proximidade da posição de um *gride* de dados com o valor aceitável ou verdadeiro.

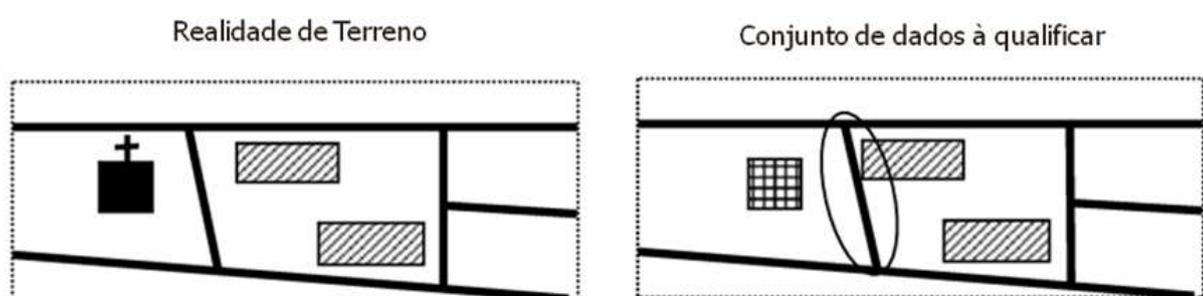


Figura 9 - Exemplo de avaliação da acurácia posicional

Fonte: Devillers e Jeansoulin (2006)

#### **2.1.5.4. Acurácia temporal**

Segundo Ariza (2002), o tempo é uma característica fundamental para julgar a qualidade dos dados, mas para gerenciar esta informação deve-se considerar a que tempo se refere: ao tempo lógico do evento, ou seja, quando ocorreram as mudanças no mundo real; ao tempo da observação da evidência; ou ao tempo em que é realizada a carga das mudanças no banco de dados. Desta forma, para norma ISO 19113:2002 a acurácia temporal é composta de 3 (três) subelementos:

Acurácia de uma medida temporal: correta referência temporal do item.

Consistência temporal: correta ordem de eventos ou sequência, se relatado.

Validade temporal: é a validade dos dados em respeito a um determinado tempo.

A forma mais simples de incluir o tempo é como um atributo, entretanto o desejável seria poder gerenciá-lo como uma dimensão a mais (ARIZA, 2002).

A data da entrada de dados, ou a data da sua revisão, é um fator importante para o usuário quando julgar a qualidade dos dados, no sentido de aptidão para o uso. A acurácia temporal diz respeito às datas de aquisição de dados, tipos de atualizações e períodos de validade (DEVILLERS e JEANSOULIN, 2006).

#### **2.1.5.5. Acurácia temática**

Segundo Ariza (2002), do ponto de vista de qualidade, a componente temática refere-se a acurácia dos atributos de um conjunto de dados espaciais em relação a verdade do terreno. Esta componente de qualidade não é exclusiva de mapas temáticos, pois qualquer elemento representado em um mapa pertence a um tema, como construções, nascentes, vértices geodésicos, que também podem submeter-se a um estudo sobre sua classificação correta ou não.

Segundo a norma ISO 19113:2002, a acurácia temática é a classificação correta entre as classes e seus atributos em relação a um universo de discussão, definindo dois tipos de atributos: quantitativos e não quantitativos. Esta define 3 (três) subelementos: correção da classificação, grau de correção de atributos não quantitativos e acurácia dos atributos quantitativos.

Em relação aos elementos de qualidade citados nesta dissertação, a Tabela 4 mostra uma síntese dos mesmos. Os elementos e subelementos listados (Tabela 4) são baseados na norma ISO 19113:2002, e a mesma cita outros elementos não quantitativos que podem ser descritos no padrão internacional, como o propósito, uso e a linhagem do conjunto de dados espaciais.

TABELA 4 - Elementos e subelementos de qualidade

<b>Elemento</b>	<b>Subelemento</b>
<b>Compleitude</b>	Comissão
	Omissão
<b>Consistência lógica</b>	Consistência conceitual
	Consistência de domínio
	Consistência de formato
	Consistência topológica
<b>Acurácia posicional</b>	Acurácia absoluta
	Acurácia relativa
	Acurácia posicional dos dados em grade
<b>Acurácia temporal</b>	Acurácia de uma medida temporal
	Consistência temporal
	Validade temporal
<b>Acurácia temática</b>	Correção da classificação
	Grau de correção dos atributos não quantitativos
	Acurácia dos atributos quantitativos

Fonte: baseado na norma ISO 19113:2002

Os elementos de qualidade citados nesta seção são implícitos ao conjunto de dados geoespaciais e fundamentais no preparo da aferição da qualidade. Estes, aliados às especificações técnicas do produto cartográfico em estudo, mostram como as características presentes nos objetos do mundo real devem ser aferidas com a sua representação no conjunto de dados e desta forma realizar a avaliação da qualidade.

Esta seção apresentou os conceitos gerais de qualidade de dados, os elementos de qualidade de dados espaciais, a visão do produtor e do usuário dos dados e o desenvolvimento atual de normas internacionais e nacionais relativas à avaliação de um conjunto de dados espaciais, indicando a resposta sobre "o que avaliar". A próxima seção complementa os conceitos apresentado até aqui, com o objetivo de auxiliar a composição de procedimentos de aferição da qualidade de um produto cartográfico.

## **2.2. Generalização cartográfica: princípios de cartografia aplicados na produção de mapas**

Nesta seção será apresentada uma visão geral dos conceitos de generalização e os chamados operadores de generalização. O objetivo é complementar os conceitos presentes nos elementos de qualidade citados anteriormente com os existentes nos processos de generalização cartográfica, de forma a ressaltar a relevância desta ferramenta conceitual, que em diversas vezes confunde-se com a própria produção cartográfica.

No processo de produção cartográfica, distintos conceitos de generalização são postos em prática, em diversos momentos mostram-se intuitivos, apesar de descritos e algumas vezes modelados na literatura. Todavia, justamente por se tratar de

procedimentos intuitivos, a generalização cartográfica influencia o processo de aferição da qualidade do conjunto de dados espaciais. Em diferentes situações é contestado se determinado operador de generalização foi aplicado adequadamente ou se o mesmo (ou outro) melhoraria determinada característica do conjunto de dados. Dentre os diversos fatores que influenciam tanto o processo de produção como o de generalização cartográfica são citados o da realidade, o propósito do mapa, a qualidade e quantidade (causados tanto pela falta como pelo excesso) de dados avaliados para a escala do mapa, as condições de uso e os limites técnicos (ROBINSON *et al.*, 1995).

A aplicação do processo de generalização cartográfica surge da necessidade da comunicação dos dados requeridos para o propósito do mapa. O processo deve ser executado de forma a manter um equilíbrio entre a quantidade de dados representados e o objetivo do mapeamento, onde a escala do mapa tem impacto significativo na quantidade de generalização a ser utilizada, quanto menor a escala, normalmente um maior grau de generalização será necessário (ROBINSON *et al.*, 1995; VIANNA, 1997).

A quantificação da generalização mostra-se uma necessidade notória, a fim de permitir compreender a qualidade de cada um de seus aspectos, mas sua natureza intuitiva torna difícil uma avaliação completamente objetiva. Ressalta-se que, apesar da generalização ser um processo que afeta os componentes de qualidade, existe um desenvolvimento limitado, e um escasso registro histórico ou de linhagem dos operadores de generalização aplicados sobre o conjunto de dados (ARIZA, 2002).

Outros fatores que afetam a generalização são os chamados limites gráficos, podendo ser divididos em dois fatores: o limite técnico estabelecido pelas ferramentas e o limite da percepção do olho humano. Todos os tipos de limitações seja física (equipamentos, materiais e habilidade operacional), fisiológica e psicológica (percepção e reações às variáveis visuais do mapa) propiciam uma demanda por controle sobre a quantidade de generalização a ser usada com sucesso (ROBINSON *et al.*, 1995).

A generalização cartográfica é um processo fundamental na produção de mapas, contendo um conjunto de regras e procedimentos para sua aplicação. Entretanto, por se tratar de um processo subjetivo, depende do conhecimento geográfico e cartográfico do técnico responsável e, apesar de afetar significativamente os elementos de qualidade, existe um desenvolvimento limitado a fim de torná-lo sistemático e capaz de ser repetido através da definição de parâmetros (ARIZA, 2002; VIANNA, 2003; LONGLEY *et al.*, 2013).

### **2.2.1. Conceitos de generalização cartográfica**

A generalização cartográfica é um processo chave para o mapeamento, apesar de ser de um procedimento subjetivo de ajuste das feições selecionadas à escala do mapa, dependente do conhecimento geográfico e cartográfico do técnico responsável pela sua execução para concretizar uma comunicação eficaz (VIANNA, 1997).

Um documento cartográfico é um modelo, em escala do mundo real, e para sua criação é preciso extrair da realidade todas as informações necessárias, reduzir a quantidade de elementos, simplificar relações e criar representações para atender a objetivos específicos do usuário (ISSMAEL, 2003).

Esta redução do espaço permite observar a disposição geográfica dos fenômenos, mas é acompanhada por mudanças inevitáveis. As distâncias entre feições, larguras e comprimentos são reduzidas, assim como itens adjacentes antes discretos mostram-se cada vez mais aglomerados. Há um aumento da complexidade e aglomeração das feições promovendo confusão visual (ROBINSON *et al.*, 1995). A restrição do espaço e a representação dos objetos de forma simbólica ocupam um lugar proporcionalmente igual ou maior do que possuem na realidade. Isto gera uma concorrência ou conflito, existindo uma clara necessidade de regras de seleção para assegurar a homogeneidade da escolha de feições e a redução da influência da interpretação pessoal de cada operador. As especificações de um produto cartográfico definem como o mundo real deve ser incluído no conjunto de dados, e a generalização cartográfica contempla normas e procedimentos aos quais permitem escolher e representar, de forma simplificada, os detalhes necessários em cada mudança de escala, sempre inferior, e de propósito, sobre o conjunto de dados geoespacial (ARIZA, 2002).

As informações geradas são obtidas através de transformações aplicadas nos dados originais, sejam eles obtidos diretamente ou do mundo real (restituição fotogramétrica, interpretação de imagens orbitais e/ou dados de campo) ou de cartas de escalas maiores (compilação). Entretanto, independente do método há a diminuição da capacidade de apresentar detalhes e informações, pressupondo perdas, caracterizando um processo de generalização (VIANNA, 1997).

A definição de regras e parâmetros é necessária para cada tema e circunstância, contribuindo para prever as perdas e conhecer as possibilidades de transformações da informação manuseada. Para neutralizar esses efeitos indesejáveis são realizadas duas

operações fundamentais nas características das feições geográficas: seleção e generalização (ROBINSON *et al.*, 1995; ARIZA, 2002)<sup>4</sup>.

Na operação de seleção é decidido o que é necessário em termos de classe de feições para atender ao propósito do mapa, mas nenhuma modificação é realizada, e a escolha consiste em representar ou não determinada feição. Em relação ao processo de seleção, a cartografia digital permite a experimentação, sem aumentar significativamente o tempo e o custo do mapeamento. Ao despendar tempo suficiente no processo de seleção pode haver uma redução nos custos globais do mapeamento (ROBINSON *et al.*, 1995).

Uma vez selecionadas as feições e atributos de mapeamento a próxima etapa é a generalização que é realizada através dos chamados operadores de generalização (ROBINSON *et al.*, 1995).

### **2.2.2. Operadores de generalização cartográfica**

O nível de detalhe de uma base cartográfica determina o grau de aproximação com o mundo real e sua complexidade, onde técnicas de generalização são aplicadas geralmente para remover detalhes e transformar espacialmente as feições (LONGLEY *et al.*, 2013). Na generalização cartográfica é importante abordar cinco termos: classificação, simplificação, exagero, simbolização e indução (ROBINSON *et al.*, 1995).

A classificação consiste na ordenação de escala e agrupamento de feições por suas características e valores de atributos. A simplificação determina as características relevantes de feições de atributos e elimina detalhes indesejados. O exagero aumenta ou enfatiza características importantes das feições. A simbolização é utilizada para codificar a informação para visualização e a indução ocorre quando há inferências a partir das inter-relações entre feições no mapa. O objetivo dos operadores de generalização é simplificar um processo complexo a proporções controláveis, entretanto exigem uma série de decisões. Na realidade, os operadores de generalização são correlacionados, pois é difícil controlar um elemento sem afetar os demais (ROBINSON *et al.*, 1995). Além disto, o tratamento independente dos elementos do mapa pode levar a sobreposições e coberturas de símbolos fazendo ilegível a combinação (ARIZA, 2002).

---

<sup>4</sup> O autor distingue seleção e generalização, onde no primeiro caso não há alteração das feições geográficas ao contrário do segundo.

Os principais operadores de generalização, encontrados na literatura sobre o assunto, são mostrados na Figura 10 conforme citados em Issmael (2003). Em função da dimensão das ilustrações de alguns exemplos, os demais são apresentados nos tópicos com as respectivas descrições.

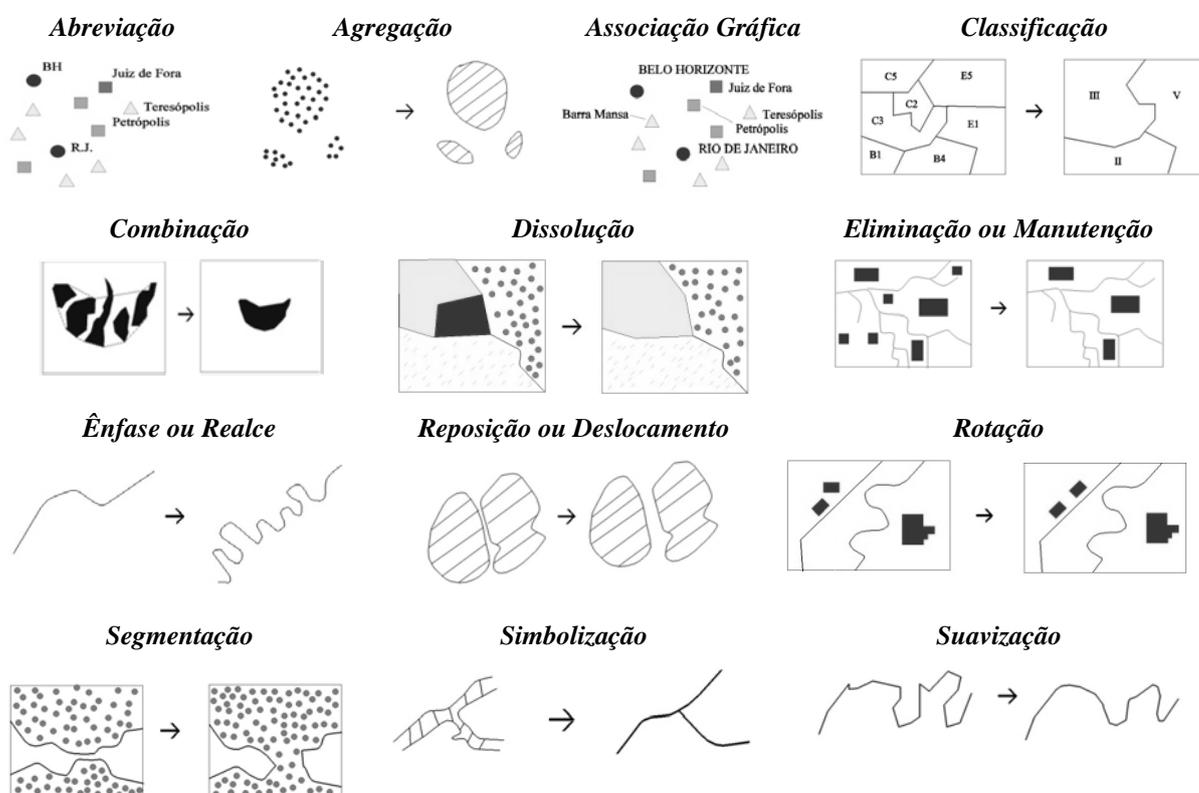


Figura 10 - Operadores de generalização cartográfica

Fonte: baseado em Issmael (2003)

### ***Abreviação***

A abreviação é o operador de generalização que permite a escrita de uma informação textual de forma reduzida (MONMONIER<sup>5</sup>, 1991 apud ISSMAEL, 2003).

### ***Agregação***

A agregação é o operador de generalização que visa a junção de elementos que estejam muito próximos uns dos outros, permitindo mudanças dimensionais crescentes: ponto em linha, ponto em área, ponto em volume, linha em área, linha em volume e área em volume (MONMONIER, 1991 apud ISSMAEL, 2003 e AZEVEDO, 2003).

### ***Associação Gráfica***

A associação gráfica é o operador que permite a associação de uma informação textual a um elemento pontual através de linha ou código (MONMONIER, 1991 apud ISSMAEL, 2003).

<sup>5</sup> MONMONIER, M. **How to Lie With Maps**. The University of Chicago Press, Chicago, 1991.

### **Amálgama**

A amálgama é o operador que permite a união de áreas contíguas com atributos similares em uma área maior que representa a soma das iniciais, eliminando as fronteiras entre elas, ou seja, substitui vários objetos do tipo área por um único objeto, conforme mostra a Figura 11 (ISSMAEL, 2003 e LONGLEY *et al.*, 2013).

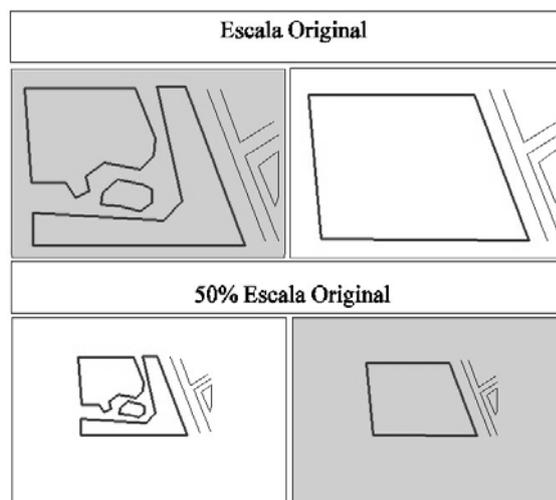


Figura 11 - Operador de generalização cartográfica: amálgama

Fonte: McMaster e Shea<sup>6</sup> (1992) apud Issmael (2003)

### **Classificação**

A classificação é a operação através da qual são agrupados os elementos geográficos que compartilham, no mínimo, uma característica em comum, fazendo com que a individualidade e detalhes pertencentes a cada elemento sejam desprezados mantendo, apenas uma feição mais geral que representa o grupo de objetos originais (VIANNA, 1997 e AZEVEDO, 2003).

Segundo Robinson *et al.* (1995), os métodos de classificação são colapso, tipificação, estatística do vizinho mais próximo (para feições pontuais), tipificação (para feições lineares), agregação de áreas, agregação de volumes (para feições do tipo área).

### **Combinação**

A combinação junta feições de uma classe num único elemento de mesma geometria, quando são posicionadas muito próximas na redução de escala (KEATES<sup>7</sup>, 1989 apud VIANNA, 1997).

<sup>6</sup> MCMMASTER, R. B.; SHEA, K. S. **Generalization in Digital Cartography**. 1.ed. Washington: Association of American Geographers, 1992.

<sup>7</sup> KEATES, J. S. **Cartographic Design and Production**. Longman Scientific & Technical, New York, 1989.

### ***Dissolução***

A dissolução é o operador de generalização que permite remover os limites de um determinado objeto com o objetivo de classificá-lo de acordo com o elemento envolvente, ou seja, trata-se de um tipo de eliminação (MONMONIER, 1991 apud ISSMAEL, 2003).

### ***Eliminação ou Manutenção***

O operador de generalização de eliminação permite a remoção de elementos geográficos menos significativos devido a proximidade de elementos geográficos mais relevantes ou a manutenção (permanência) de elementos pertinentes durante a mudança de escala ou propósito do mapa (AZEVEDO, 2003 e ISSMAEL, 2003).

### ***Ênfase ou Realce***

O operador de generalização de ênfase ou realce permite a alteração das características de um símbolo, a fim de torná-lo mais adequado à visualização em escalas menores, detalhes são acrescentados a simbologia de modo a deixá-la mais real, atendendo requisitos específicos do mapa (VIANNA, 1997 e AZEVEDO, 2003).

### ***Exagero***

O exagero é o operador de generalização que permite o aumento das dimensões de objetos geográficos considerados relevantes, para que sua representação seja adequada a escala do mapa final, caso contrário as dimensões verdadeiras seriam pequenas demais para visualização, conforme mostra a Figura 12 (AZEVEDO, 2003 e ISSMAEL, 2003).

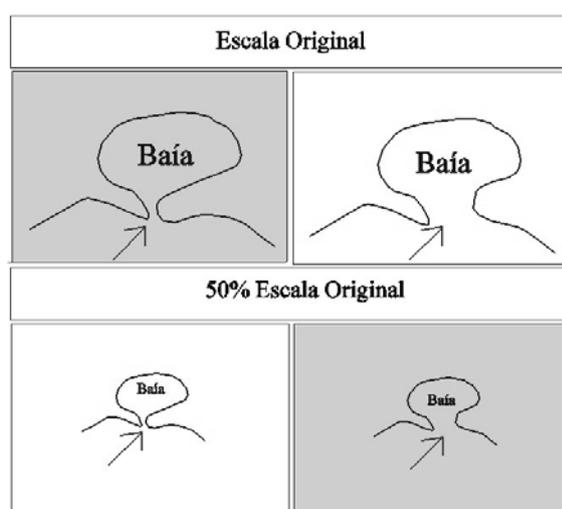


Figura 12 - Operador de generalização cartográfica: exagero

Fonte: McMaster e Shea (1992) apud Issmael (2003)

### ***Fusão ou unificação***

A fusão é o operador de generalização que permite substituir objetos com um determinado tipo de geometria, por outro de mesma geometria como uma densidade menor (pontos em ponto, linhas em linha, áreas em área e volumes em volume), conforme mostra a Figura 13 (ISSMAEL, 2003).

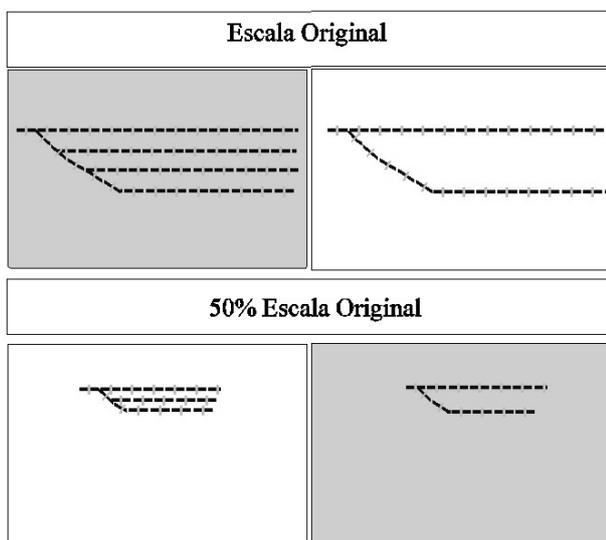


Figura 13 - Operador de generalização cartográfica: fusão

Fonte: McMaster e Shea (1992) apud Issmael (2003)

### ***Indução***

A generalização por indução pode ser incentivada por um bom e claro projeto cartográfico, após a classificação, simplificação, exagero e simbolização. A indução nunca pode ser planejada por completo, uma vez que diferentes usuários do conjunto de dados espaciais, com seus conhecimentos pessoais, descobrem hipóteses que não podem ser previstas ou estabelecidas, a não ser pelo uso. As informações presentes e especificadas no produto cartográfico são ampliadas, evidenciando uma das principais vantagens de um mapa, conforme mostra a Figura 14 (ROBINSON *et al.*, 1995).

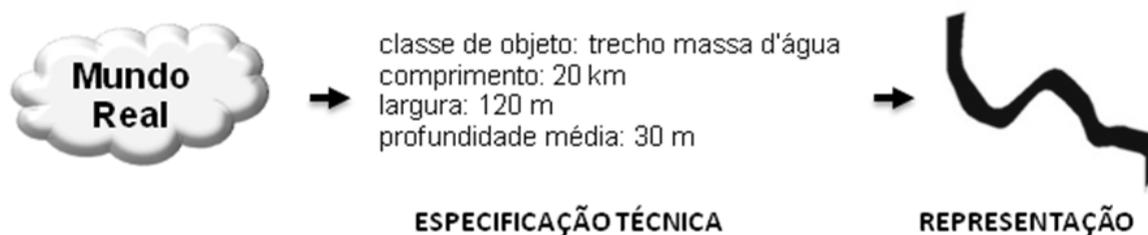


Figura 14 - Operador de generalização cartográfica: indução

Fonte: adaptado de Issmael (2003)

### ***Reposição ou Deslocamento***

O operador de generalização deslocamento visa a mudança intencional da posição de um objeto, para destacá-lo de outro muito próximo, mantendo a posição relativa entre as ocorrências. Para este operador, deve-se criar uma sequencia de prioridades de deslocamento para os objetos (AZEVEDO, 2003 e ISSMAEL, 2003).

### ***Refinamento***

O operador de generalização de refinamento permite a seleção de um determinado número de objetos de mesma origem, com a finalidade de manter ou eliminar os que não acrescentam detalhes significativos na impressão, conforme mostra a Figura 15 (ISSMAEL, 2003).

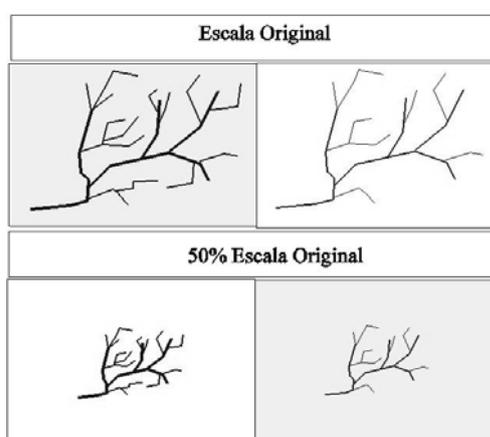


Figura 15 - Operador de generalização cartográfica: refinamento

Fonte: McMaster e Shea (1992) apud Issmael (2003)

### ***Rotação***

A rotação é o operador de generalização que permite girar objetos pontuais em torno de uma determinada posição (ISSMAEL, 2003).

### ***Segmentação***

A segmentação é o operador de generalização onde um trecho do objeto é eliminado ou dissolvido, e a representação do objeto fica segmentada e preenchida com a simbologia do objeto envolvente (ISSMAEL, 2003).

### ***Simbolização***

A simbolização é o operador de generalização responsável por transformar as feições do mundo real em objetos gráficos (símbolos) num mapa, e tornar as representações significativas, segundo o propósito e escala do mapa, ou seja, é realizada uma codificação gráfica (ISSMAEL, 2003 e VIANNA, 1997).

Na realidade, todos os objetos gráficos em um mapa são símbolos, os mesmos podem ser utilizados para simbolizar um conceito, uma série de fatos ou uma característica de uma distribuição geográfica (ROBINSON *et al.*, 1995).

### ***Simplificação***

A simplificação determina as características relevantes das feições e elimina detalhes indesejados (vértices), contemplando dois objetivos principais: reduzir a quantidade de informações no mapa de forma legível na escala escolhida e manter a aparência semelhante dos fenômenos geográficos mapeados, sendo aplicada a elementos lineares e contornos de áreas, conforme mostra a Figura 16 (VIANNA, 1997, ISSMAEL, 2003 e ROBINSON *et al.*, 1995).

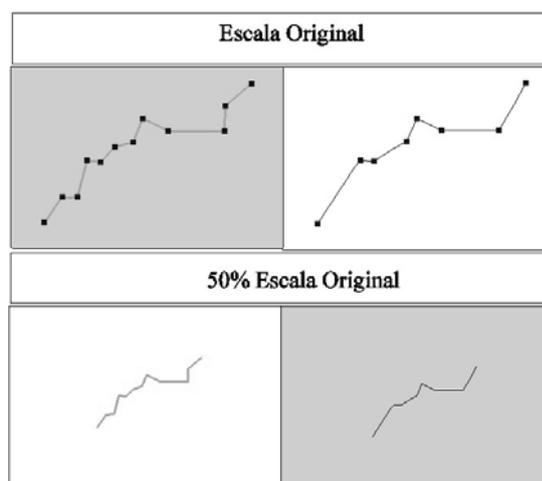


Figura 16 - Operador de generalização cartográfica: simplificação

Fonte: McMaster e Shea (1992) apud Issmael (2003)

Apenas a seleção de feições necessárias ao mapeamento não é suficiente, na maioria dos casos, a informação a ser mapeada precisa ser simplificada. Esta simplificação consiste na eliminação e suavização, redução de detalhes das feições restantes, sem implicar em perda da qualidade ou precisão (AZEVEDO, 2003; ROBINSON *et al.*, 1995).

### ***Suavização***

A suavização é o operador de generalização que permite criar e reposicionar os pontos originais de linhas e contornos de área, com o objetivo de eliminar e minimizar pequenas perturbações entre traços sucessivos, delineando as tendências mais significativas da linha, abstraindo as pequenas variações locais (AZEVEDO, 2003 e ISSMAEL, 2003).

Conforme demonstrado, a generalização cartográfica é um processo fundamental na produção de mapas, contendo um conjunto de regras e procedimentos para sua aplicação. Ressalta-se que, nesta dissertação, a seleção foi considerada como operador de pré-generalização, pois não transforma espacialmente as feições ao contrário dos demais operadores de generalização, conforme Issmael (2003) e Robinson *et al.* (1995).

### 2.3. Técnicas de amostragem aplicadas a dados geoespaciais

Uma etapa fundamental ao processo de avaliação da qualidade de um produto cartográfico é a amostragem dos dados. Na maioria das vezes, ao lidar com um conjunto de dados espaciais nota-se sua complexidade devido a dimensões territoriais e a quantidade de informações representadas. Desta forma, em diversas situações é inviável realizar uma inspeção completa do conjunto de dados, exceto nos processos que podem ser automatizados, como algumas avaliações de consistência lógica, as demais demandam amostras que representem o produto cartográfico como um todo.

Neste tópico será abordado conceitos gerais de amostragem e dos métodos mais usuais para coleta de amostras, além de descrever os parâmetros característicos de uma análise exploratória dos dados de um fenômeno em estudo. Entende-se que na avaliação de qualidade é necessária uma noção das diferentes técnicas de amostragem a fim de identificar o método mais adequado a sua demanda. Nesta pesquisa, a amostragem de regiões e feições geográficas para avaliação de sua qualidade é baseada nos elementos de qualidade definidos nas normas ISO e nas especificações técnicas de um produto cartográfico em estudo.

#### 2.3.1. Conceitos de amostragem

Para a tomada de decisão é sempre conveniente ter um conhecimento completo da situação, mas muitas vezes isto não é possível, neste caso preocupa-se em ter um conhecimento que, embora aproximado, tenha base estatística (ARIZA, 2004).

*"(...) a amostragem é naturalmente usada em nossa vida diária. Por exemplo, para verificar o tempero de um alimento em preparação, podemos provar (observar) uma pequena porção. Estamos fazendo uma amostragem, ou seja, extraindo do todo (população) uma parte (amostra), com o propósito de termos uma ideia (inferirmos) sobre a qualidade do tempero de todo o alimento" (BARBETTA, 2012).*

A amostragem é o processo de seleção de uma amostra, que possibilita o estudo das características (parâmetros) da população. Para estimar o comportamento da população a partir de amostras é necessário que o subconjunto seja coletado de tal forma que cada observação tenha a mesma chance de ser selecionada (LANDIM, 1998).

A população ( $N$ ) é o conjunto formado por indivíduos ou objetos que têm pelo menos uma variável comum e observável. Na realidade, esta variável comum é uma medida referente a uma propriedade específica de um grupo de objetos, e não o grupo de objetos. Além disto, a definição desse conjunto de objetos deve considerar sua localização no espaço e no tempo (LANDIM, 1998; FERREIRA, 2009; MORETTIN, 2010).

A amostra ( $n$ ) é uma porção dos elementos de uma população, ou seja, qualquer subconjunto de seus elementos, que deve reproduzir a distribuição e variabilidade da variável em estudo, tanto em termos quantitativos como em relação a disposição dos pontos amostrais no domínio em análise. O parâmetro ( $\theta$ ) é uma medida que reproduz determinada particularidade dos elementos da população, como a média ( $\mu$ ), a variância ( $\sigma^2$ ) e o coeficiente de correlação ( $\rho$ ). O espaço amostral ( $\Omega$ ) é o conjunto dos resultados de um experimento aleatório, onde cada elemento do mesmo é denominado de ponto amostral. O erro amostral ( $\varepsilon$ ) é a incerteza decorrente do uso de amostras. A estimativa ( $\hat{\theta}_0$ ) é o valor calculado com base na amostra cujo objetivo é avaliar aproximadamente um parâmetro, estipulado pelo estimador (MORETTIN, 2010; BARBETTA, 2012).

O estimador ( $\hat{\theta}$ ) é uma propriedade numérica estipulada na amostra, denominado estatística de um parâmetro populacional, como a média amostral ( $\bar{x}$ ), a variância amostral ( $s^2$ ) e o coeficiente de correlação amostral ( $r$ ) (MORETTIN, 2010).

O coeficiente de correlação linear ( $r$ ) é uma medida referente ao grau da relação linear entre duas variáveis X e Y em uma amostra, o seu valor varia entre  $-1$  (correlação inversa) e  $+1$  (correlação direta). Por exemplo, considerando a Equação 7, num caso hipotético, onde  $r^2 = 0,83$ , o grau de dependência da variável Y em relação a X é de 83%, significando que 17% da variação (dispersão) total permanece sem explicação (ANDRIOTTI, 2003),

$$r^2 = \frac{\text{variação explicada}}{\text{variação total}} \quad \text{ou} \quad (7)$$

$$r^2 = (\text{variação total} - \text{variação não explicada}) / (\text{variação total})$$

A regressão linear simples é dada pela Equação 8 de uma reta (ANDRIOTTI, 2003):

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (8)$$

onde:

- $\beta_0$  representa o coeficiente angular da reta;
- $\beta_1$  representa o ponto em que a reta intercepta o eixo das ordenadas;
- $\varepsilon$  representa o resíduo entre o valor amostrado e o estimado.

O resultado da regressão idealiza que  $\beta_1$  seja próximo de 1 e  $\beta_0$  seja próximo de zero, indicando uma boa correlação entre as variáveis X e Y.

A unidade de amostragem é a parte a ser selecionada para obter os elementos da população. Esta pode ser o próprio elemento da população, ou outro exemplar de fácil seleção vinculado aos elementos da população (BARBETTA, 2012).

Quando os dados de uma amostra são utilizados para se conhecer os parâmetros de uma população, é realizada uma inferência estatística. Os valores calculados com base na amostra são chamados de estimativas destes parâmetros (BARBETTA, 2012). A inferência estatística diz respeito às relações entre amostras e a população amostrada (LANDIM, 1998).

Segundo Guerra (1988), uma das limitações básicas da amostragem é consequência justamente do conceito de "representatividade", já que dela depende em grande parte das "interpretações" e "estimativas" posteriores.

A representatividade está ligada ao conceito de "heterogeneidade" ou "homogeneidade", onde num corpo totalmente homogêneo, qualquer fragmento é representativo do conjunto, ao contrário de um corpo heterogêneo, cujo fragmento não representa nem mesmo seu vizinho (GUERRA, 1988).

Para Ariza (2004), estimar é trabalhar com uma aproximação estatística, assim existem riscos quanto a essa aproximação. Os mesmos são denominados:

- a) risco do produtor (probabilidade de reprovar um bom trabalho);
- b) risco do cliente ou usuário (probabilidade de aceitar como bom um trabalho ruim).

Em relação à amostragem, a norma ISO 19114:2003 orienta o uso das séries da normas ISO 2859 e ISO 3951-1, que apesar de serem originalmente destinadas ao uso não espacial, as mesmas podem ser aplicadas aos dados geográficos.

A norma ISO 19114:2003 destaca dois conceitos no método de inspeção por amostragem: lote e item. O lote é a unidade mínima a qual se aplica a avaliação da qualidade, por exemplo, uma determinada classe de feição ou uma área, e item é a unidade mínima a ser inspecionada, como uma instância de feição pertencente a um lote, ambos devem ser definidos de acordo com as especificações técnicas do conjunto de dados.

### **2.3.2. Métodos de amostragem**

O processo de amostragem é a metodologia adequada para a seleção de elementos que serão efetivamente observados e deve ser realizada com base num planejamento, que define a coleta das unidades de amostragem de forma aleatória simples, aleatória estratificada ou sistemática. As amostragens são ditas probabilísticas quando cada elemento da população pode pertencer a amostra sorteada, caso contrário,

se por alguma razão isto não ocorrer, a amostragem é dita não probabilística (FERREIRA, 2009, BARBETTA, 2012; YAMAMOTO e LANDIM, 2013).

A seguir são descritos os tipos de amostragens e suas características.

### 2.3.2.1. Amostragens não aleatórias

Na prática, uma amostragem não aleatória pode ser mais adequada do que uma amostragem aleatória, em função da dificuldade ou impossibilidade da seleção da amostra, procurando, de forma geral, criar amostras que representem significativamente a população de onde foram extraídas (BARBETTA, 2012).

No caso de amostragens orientadas por área, onde a cobertura total de uma determinada região geográfica é necessária, os locais de amostras devem ser determinados de acordo com um padrão regular ou semi regular. A Figura 17 mostra uma situação de amostragem não aleatória guiada por área regular sobre o município de Leopoldina/MG (ISO 19114:2003).

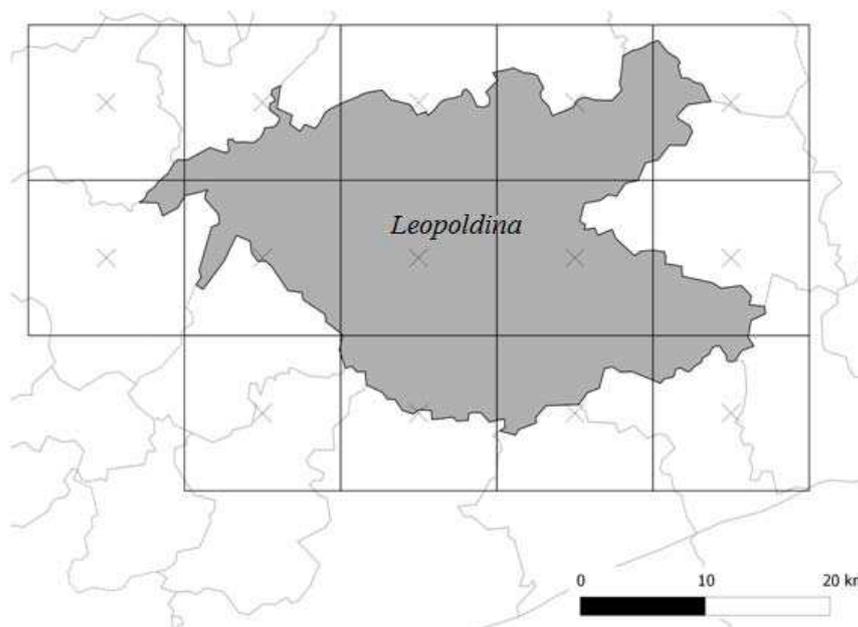


Figura 17 - Exemplo de amostragem não aleatória orientada por área

Fonte: adaptada da norma ISO 19114:2003

### ***Amostragem por cotas***

Na amostragem por cotas a população é vista de forma segregada, discriminada em subgrupos, similar a amostragem estratificada proporcional (abordada no tópico das amostragens aleatórias). A cota de seleção dentro de cada subgrupo é proporcional ao tamanho do mesmo. No entanto, ao contrário da amostragem aleatória estratificada proporcional, a seleção não precisa ser aleatória e usualmente costuma-se dividir a população num grande número de subgrupos para compensar a falta de aleatoriedade (BARBETTA, 2012).

### 2.3.2.2. Amostragens aleatórias

A amostragem aleatória consiste na seleção casual de  $n$  unidades sorteadas sem reposição de uma população constituída de  $N$  unidades sequencialmente numeradas, na qual qualquer elemento da população tem a mesma probabilidade de ser selecionado. Nos estudos de fenômenos espaciais a seleção aleatória pode levar em consideração as coordenadas geográficas para serem escolhidas aleatoriamente (YAMAMOTO, 2013).

Segundo Landim (1998) e Barbetta (2012) para obter a amostragem aleatória simples é necessário ter uma relação completa de todos os elementos da população ( $N$ ). Este tipo de amostragem consiste em selecionar a amostra através de um sorteio, sem restrição. Além disto, esta é dita com reposição caso os elementos da população possam entrar mais de uma vez na amostra, sendo considerada estatisticamente independente, caso contrário, é denominada amostragem sem reposição e considerada estatisticamente dependente.

Segundo a norma ISO 19114:2003 a amostragem aleatória simples é útil quando a população de interesse é relativamente homogênea em relação a característica em análise, ou seja, sem grandes padrões e agrupamentos. Uma vez que, no caso de dados geoespaciais existe o risco da ocorrência de uma amostra concentrada numa pequena região, não representando uma cobertura significativa da área de interesse, o que pode ser indesejável. A Figura 18 ilustra uma seleção aleatória, orientada por áreas, no município de Carlos Chagas/MG, na prática as quadrículas possuem a dimensão das articulações do mapeamento 1:25.000 correspondendo a uma área de aproximadamente 13 km<sup>2</sup>.

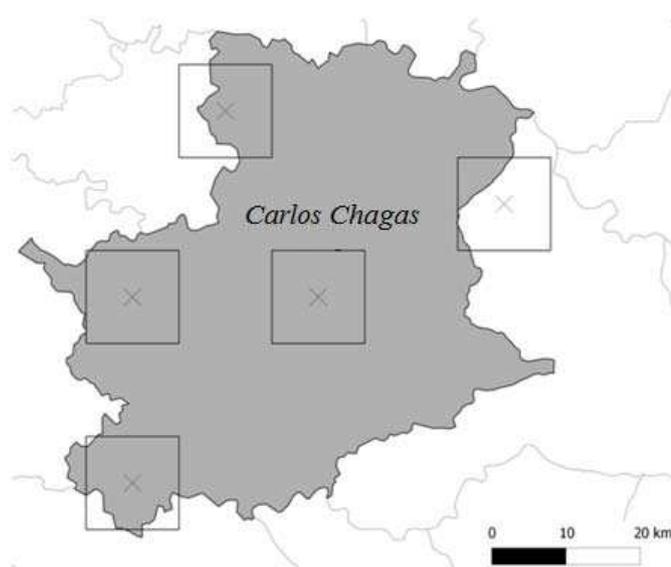


Figura 18 - Exemplo de amostragem aleatória orientada por áreas

Fonte: adaptada da norma ISO 19114:2003

### ***Amostragem sistemática***

A amostragem aleatória sistemática é utilizada quando o sistema de referência geral para toda a população é dispensado e amostras são sistematicamente coletadas por sorteio segundo um padrão pré-determinado (LANDIM, 1998).

Para Barbeta (2012) a amostragem aleatória sistemática é um processo de características similares à seleção aleatória simples e uma amostra sistemática poderá ser tratada como uma amostra aleatória simples se os elementos da população estiverem ordenados aleatoriamente.

Para realizar a amostragem aleatória sistemática é necessário que a população ( $N$ ) esteja ordenada e o tamanho da amostra ( $n$ ) definido. A relação  $s = N/n$ , denominada como fator de sistematização ou intervalo de seleção, é calculada. O primeiro elemento da amostra é obtido através do sorteio de um número " $m$ " no intervalo de 1 e  $N$ , o restante " $(n-1)$ " são coletados de forma sistemática, somando o intervalo de seleção ( $s$ ) ao número sorteado até o  $n$ -ésimo elemento da amostra, ou seja, o número " $(n-1)s + m$ " (MORETTIN, 2010).

Na norma ISO 19114:2003, a amostragem aleatória sistemática é denominada semi aleatória, descrita como uma maneira prática e fácil de garantir a cobertura de uma área e ressaltando a necessidade de definir uma regra para incluir ou excluir os objetos que não estão contidos dentro da área de interesse amostrada. Na realidade, é notória a boa relação entre a amostragem aleatória sistemática e métodos de análise espacial dos resultados de inspeção por amostragem, em função da abrangência na distribuição espacial das amostras (YAMAMOTO e LANDIM, 2013).

### ***Amostragem estratificada***

A amostragem estratificada é utilizada quando se pressupõem uma grande variabilidade nas observações. Neste caso, a população é dividida em subgrupos, denominados estratos, através de uma variável "critério", onde cada um é submetido a uma amostragem aleatória simples. Desta forma, apesar de existir uma grande heterogeneidade entre os subgrupos, esta dispersão é menor dentro de cada subgrupo, ou seja, os estratos são homogêneos (LANDIM, 1998; MORETTIN, 2010).

Segundo a norma ISO 19114:2003, para a aplicação da amostragem estratificada a população precisa ser separada em camadas não sobrepostas ou subpopulações. O potencial deste método de amostragem está na maior precisão das estimativas da média e variância amostral dos estratos, em relação a uma estratégia de amostragem não estratificada sobre a população.

No método de amostragem estratificada são realizadas seleções aleatórias sobre os diversos estratos, de forma independente, o mesmo é classificado em proporcional e uniforme. Na amostragem estratificada proporcional a quantidade de elementos é proporcional ao tamanho de cada estrato. Na amostragem estratificada uniforme a mesma quantidade de elementos é selecionada em cada estrato, com o objetivo de comparar os diferentes estratos através de estimativas separadas (BARBETTA, 2012).

### ***Amostragem por conglomerados***

Na amostragem por conglomerados a mudança principal é o sorteio por agrupamento de elementos de população, com as seguintes características: dentro de cada aglomeração há uma grande dispersão dos elementos (heterogeneidade), mas entre os conglomerados a dispersão é menor, ou seja, são mais homogêneos. Na prática a seleção aleatória das amostras por conglomerados acontecem em vários estágios, onde novas seleções aleatórias são realizadas sucessivamente. Entretanto, as estimativas de uma amostra de conglomerados, ao contrário da amostragem estratificada, tendem a gerar resultados mais distantes dos parâmetros populacionais, quando comparada com uma amostra aleatória simples de mesmo tamanho (MORETTIN, 2010; BARBETTA, 2012).

A relação  $\frac{n}{N}$  é denominada fração de amostragem, ou seja, a proporção da população que será efetivamente observada, caso esta seja constante para todos os conglomerados selecionados, então todos os elementos da população tem a mesma probabilidade de pertencer à amostra (BARBETTA, 2012).

### **2.3.3. Tamanho da amostra**

Segundo a norma ISO 19114:2003, o tamanho de uma população ( $N$ ) e conseqüentemente o tamanho da amostra ( $n$ ) podem ser definidas para diferentes bases de itens, mas exige a indicação explícita dos mesmos, como feições, área de cobertura, curvas ou vértices. A heterogeneidade da população e os tipos de parâmetros a estimar são pontos relevantes na determinação do tamanho da amostra. Além disto, é necessário especificar o erro amostral tolerável, ou seja, o quanto se admite errar na avaliação do(s) parâmetro(s) de interesse (BARBETTA, 2012).

A especificação do erro amostral tolerável deve ser feita sob um enfoque probabilístico, visto que, por maior que seja a amostra, existe o risco do sorteio gerar amostras com propriedades bem diferentes das características da população de origem. Além disto, restrições financeiras podem ser consideradas na determinação do tamanho da amostra ( $n$ ) e conseqüentemente no cálculo do erro amostral tolerável pretendido (BARBETTA, 2012).

Desta forma, segundo Barbetta (2012), um primeiro cálculo do tamanho da amostra ( $n_0$ ), baseado apenas no erro amostral tolerável ( $E_0$ ), independente do tamanho da população ( $N$ ) pode ser determinado, através da Equação 9<sup>8</sup>,

$$n_0 = \frac{1}{E_0^2}. \quad (9)$$

Ainda segundo o autor, caso a população seja considerada grande, cita-se mais que vinte vezes o valor calculado em  $n_0$ , então  $n_0$  pode ser adotado como tamanho da amostra ( $n = n_0$ ), caso contrário é sugerida a seguinte correção:

$$n = \frac{N \cdot n_0}{N + n_0}. \quad (10)$$

O cálculo do tamanho da amostra para subgrupos, como na amostragem estratificada ou por conglomerados, é realizado para cada agrupamento. O fato repercute no tamanho total da amostra, ao qual corresponderá à soma das amostras dos subgrupos. Desta forma, um erro amostral tolerável maior é aceitável, em função desta elevada quantidade de amostras no somatório de todos subgrupos (BARBETTA, 2012).

Segundo Nogueira Junior (2003), o tamanho mínimo de uma amostra ( $n$ ) para uma população finita ( $N$ ), ao fazer uma estimativa da média ( $\mu$ ), definir um erro amostral tolerável ( $E_0$ ) e estabelecer um nível de confiança de  $(1-\alpha)$  desejado é dada por:

$$n = \frac{z^2 \sigma^2 N}{(N-1)E_0 + z^2 \sigma^2}, \quad (11)$$

onde:

- $z$  é o valor padronizado obtido através da tabela de distribuição normal em função do intervalo de confiança desejado;
- $\sigma$  é o desvio padrão;
- $n$  é o tamanho da amostra.

Segundo Barbetta (2012) o valor  $z$ , conhecido como valor padronizado, é uma medida relativa que indica o quanto o valor de uma variável  $x$  se afasta da média ( $\mu$ ), em unidade de desvio padrão ( $\sigma$ ). A transformação do valor da variável  $x$  neste valor padronizado da distribuição normal padrão<sup>9</sup> é calculado segundo a Equação 12:

$$z = \frac{x - \mu}{\sigma}. \quad (12)$$

<sup>8</sup> Segundo Barbetta (2012) esta abordagem é bastante genérica e pode fornecer um tamanho de amostra superior ao tamanho de amostra que seria necessário para uma dada situação específica.

<sup>9</sup> Segundo Barbetta (2012) a distribuição normal padrão é uma distribuição normal com média 0 (zero) e desvio padrão 1(um).

### 2.3.4. Estratégias de amostragem para dados espaciais

Na elaboração de um plano de amostragem, os objetivos do estudo, a população a ser amostrada e os parâmetros a estimar devem estar bem definidos, além de constar a unidade de amostragem, forma de seleção dos elementos da população e o tamanho da amostra (BARBETTA, 2012).

Nos métodos de amostragem de dados geográficos os seguintes aspectos precisam ser levados em consideração: as áreas cobertas por um conjunto de dados geográficos podem formar um espaço contínuo e, caso exista uma divisão por lotes, uma atenção deve ser dada às omissões e comissões aos itens cruzando o limite destes lotes; a qualidade dos insumos de origem e a habilidade dos operadores podem afetar a qualidade dos dados geográficos, ocasionando erros sistemáticos ou discrepantes; e o produtor de dados deve ter cuidado na especificação técnica do conjunto de dados espaciais para alcançar a homogeneidade em termos de qualidade (ISO 19114:2003; YAMAMOTO e LANDIM, 2013).

A norma ISO 19114:2003 apresenta uma diretriz para a definição de amostras e métodos de amostragem para dados geográficos, conforme Figura 19, e considera dois aspectos principais para a estratégia de coleta de amostras: os itens a serem amostrados (por área ou feição), e a maneira pela qual os mesmos são selecionados: probabilidade ou julgamento.

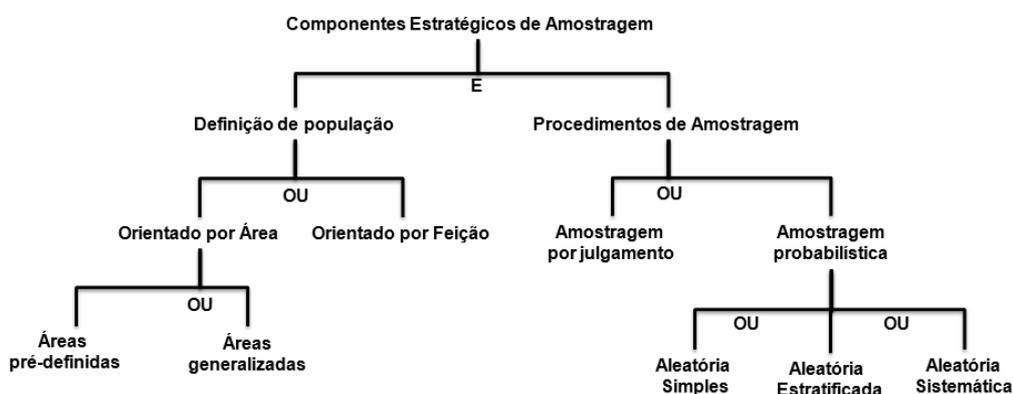


Figura 19 - Relacionamento entre as estratégias de amostragem<sup>10</sup>

Fonte: adaptado da norma ISO 19114:2003

A fim de inspecionar o conjunto de dados espaciais fundamentado estatisticamente, foram apresentados os conceitos e técnicas de amostragem, apontando a resposta sobre "o quanto" e "como" avaliar, ou seja, aborda o erro amostral tolerável e os métodos por amostragem quando a inspeção completa não se aplica ao fenômeno em estudo.

<sup>10</sup> A norma ISO 19114:2003 denomina a amostragem aleatória sistemática como semi aleatória.

## 2.4. Geoestatística

Nesta seção, são apresentados conceitos, características e ferramentas da Geoestatística para analisar, identificar e representar a correlação espacial presente no fenômeno em estudo, desde seu histórico, passando pela teoria das variáveis regionalizadas, métodos geoestatísticos, cálculo e modelagem do variograma, modelos teóricos, análise variográfica, até o seu método de estimação, a krigagem.

A geoestatística nasceu num contexto geológico, onde na década de 1950, os sul-africanos Daniel Gerhardus Krige<sup>11</sup> (engenheiro de minas) e Herbert Simon Sichel (estatístico) iniciaram a publicação de resultados de seus estudos sobre dados relativos a atividades desenvolvidas em minas de ouro do Rand (entre 1947 e 1960), na África do Sul (GUERRA, 1988; ANDRIOTTI, 2003).

O matemático francês George Matheron, no *Centre de Morphologie Mathématique*, em Fontainebleau, França, influenciado pelos estudos de Krige e outros desenvolvidos por Wijs sobre jazidas uraníferas (1952 e 1953) também na África do Sul, foi o primeiro a utilizar o termo Geoestatística e a noção de variograma, sendo estes conceitos introduzidos através do "*Traité de Géostatistique Appliquée*". (GUERRA, 1988; ANDRIOTTI, 2003).

A escola francesa, liderada por Matheron, integrou os conceitos das escolas sul-africana e norte americana, onde a primeira admitia a dependência espacial da variável em estudo e a influência do tamanho de suas amostras, e a segunda baseada na Estatística Clássica que considera a independência das variáveis (ANDRIOTTI, 2003).

Segundo Matheron<sup>12</sup> (1965) apud Andriotti (2003), "*a Geoestatística é a aplicação do formalismo das funções aleatórias ao reconhecimento e à estimação dos fenômenos naturais*". Na realidade, conforme destaca Guerra (1988), o formalismo matemático da Geoestatística não é utilizado na prática, mas foi mistificado e evoluiu em função dos trabalhos práticos da escola sul-africana, desenvolvida por Krige, e a necessidade do uso de computadores de maior eficiência, inexistentes na época.

No Brasil, as primeiras publicações referentes a geoestatística datam dos anos 80. Inicialmente, restritas a situações geológicas e posteriormente abordando outros campos, principalmente em relação a problemas associados ao meio ambiente (LANDIM, 1998).

---

<sup>11</sup> Aproveito para prestar esta breve homenagem à memória de Daniel Gerhardus Krige, 26 de agosto de 1919 à 3 de março de 2013, que faleceu durante o desenvolvimento desta pesquisa aos 93 anos.

<sup>12</sup> MATHERON, G. *Les variables régionalisées et leur estimation, une application de la théorie des fonctions aléatoires aux sciences de la nature*. Ed. Masson et Cie, Paris, 1965.

Segundo Landim (1998), no estudo espacial, as variáveis que assumem valores definidos para cada localização no espaço e as dependentes do tempo, apresentam um comportamento demasiadamente complexo para análise por métodos estatísticos usuais. A chamada Estatística Clássica pressupõe a aleatoriedade dos valores amostrais, as posições relativas das amostras são ignoradas e consideradas independentes.

Segundo Camargo (2011):

*"(...) os métodos tradicionais de interpolação espacial, como triangulação, média local das amostras e método da distância inversa, estão amplamente disponíveis nos SIG do mercado. No caso de dados esparsos, no entanto, tais métodos possuem limitações na representação da variabilidade espacial, porque desconsideram a anisotropia e a continuidade do fenômeno que se quer observar".*

Em função da simplicidade e bom desempenho dos métodos geoestatísticos na análise de dados espaciais sua aplicação foi expandida à outras áreas, além do âmbito de sua gênese, com aplicações na área florestal, criminal, epidemiológica, meteorológica, entre outras (ROSA, 2003).

#### **2.4.1. Conceitos de Geoestatística**

Segundo Landim (1998) o termo geoestatística encontra-se consagrado como um tópico especial da estatística aplicada que trata de problemas referentes às variáveis regionalizadas, as quais têm um comportamento espacial, que mostram características intermediárias entre as variáveis verdadeiramente casuais e as totalmente determinísticas.

Entretanto, conforme citado e recomendado em Andriotti (2003) faz-se necessário um estudo estatístico preliminar dos dados para posterior aplicação da Geoestatística. Segundo Guerra (1988), a geoestatística não cria informação, apenas trata a informação disponível, e se esta informação é de má qualidade, os resultados também o serão.

Segundo Andriotti (2003):

*"As técnicas geoestatísticas podem ser usadas para descrever e modelizar padrões espaciais (variografia), para prever valores em locais não amostrados (krigagem), para obter a incerteza associada a um valor estimado em locais não amostrados (variância de krigagem) e para otimizar malhas de amostragem".*

As vantagens reconhecidas da Geoestatística sobre outras técnicas convencionais de predição são: a análise variográfica, a suavização, o desagrupamento das observações, a determinação de anisotropia<sup>13</sup>, a precisão na estimativa e a incerteza da estimativa (ANDRIOTTI, 2003).

---

<sup>13</sup> Segundo Guerra (1988) a anisotropia é uma característica das variáveis regionalizadas, conhecida como zonalidade, e refere-se a possibilidade de existir certa direção privilegiada ao longo da qual os valores não variam de forma significativa, ao contrário de outras direções.

## 2.4.2. A teoria das variáveis regionalizadas

Segundo Andriotti (2003), o termo variável regionalizada foi escolhido por Matheron para destacar os aspectos aleatórios e estruturais dos fenômenos espaciais estudados. O autor ainda cita que quase todas as variáveis quantitativas referentes as ciências da Terra podem ser consideradas variáveis regionalizadas.

A Geoestatística é baseada na teoria das variáveis regionalizadas (VR), e entendida como uma função que varia de um lugar para outro no espaço com certa aparência de continuidade (GUERRA, 1988).

A continuidade geográfica atribuída se manifesta pela tendência que a variável tem de apresentar valores muito próximos em dois pontos vizinhos e mais distintos à medida que os pontos vão ficando mais distantes (LANDIM, 1998).

Segundo Guerra (1988), as variáveis regionalizadas têm em comum duas características essenciais: o aspecto aleatório onde os valores numéricos observados podem variar consideravelmente de um ponto a outro do espaço; e o aspecto espacial onde apesar da dispersão dentro do espaço, os valores numéricos observados não são inteiramente independentes. Dentre outras características qualitativas ligadas à estrutura do fenômeno natural o autor cita: a localização, a continuidade e a anisotropia.

Em Landim (1998) há uma descrição matemática em relação aos conceitos de variáveis regionalizadas:

*"(...) sejam  $\mathbf{x}(\mathbf{i})$  e  $\mathbf{x}(\mathbf{i}+\mathbf{h})$  dois valores de uma variável regionalizada obtidos nos pontos  $(\mathbf{i})$  e  $(\mathbf{i}+\mathbf{h})$ , separados entre si por uma distância  $(\mathbf{h})$ , o vetor com direção e orientação específica em um espaço de uma, duas ou três dimensões. A diferença entre esses dois valores é outra variável casual  $[\mathbf{x}(\mathbf{i})-\mathbf{x}(\mathbf{i}+\mathbf{h})]$ . A hipótese intrínseca estabelece que  $[\mathbf{x}(\mathbf{i})-\mathbf{x}(\mathbf{i}+\mathbf{h})]$  é estacionária de segunda ordem, ou seja, para qualquer deslocamento  $(\mathbf{h})$ , os dois primeiros momentos da diferença  $[\mathbf{x}(\mathbf{i})-\mathbf{x}(\mathbf{i}+\mathbf{h})]$  são independentes da localização de  $(\mathbf{x})$  e função apenas de  $(\mathbf{h})$ ".*

Segundo Santos (2010b):

*"uma função  $Z(x)$  é estacionária de 1ª ordem se,  $E[Z(x)]=\mu$ , ou seja, o valor esperado existe e não depende da posição  $x$ . Analogamente, uma função aleatória  $Z(x)$  é estacionária de 2ª ordem se for estacionária de 1ª ordem e, para cada par de variáveis  $Z(x)$  e  $Z(x+h)$ , a covariância é dada por  $cov(h) = E[Z(x)Z(x+h)] - \mu^2$ , ou seja, a covariância existe e está em função apenas de  $h$ ".*

O autor demonstra que, partindo dos princípios citados acima, "uma função aleatória é intrínseca quando satisfaz a estacionariedade de 1ª ordem e o incremento  $Z(x)$  e  $Z(x+h)$  tem variância finita e não depende de  $x$ , para qualquer vetor  $h$ ", ou seja,

$$\gamma(h) = \frac{1}{2} E\{[Z(x) - Z(x+h)]^2\}. \quad (13)$$

Desta forma, Santos (2010b) descreve as principais hipóteses da Geoestatística, estacionariedade de 1ª e 2ª ordem e intrínseca.

Segundo Landim (1998), no estudo do comportamento das variáveis regionalizadas, duas ferramentas são fundamentais para os métodos geoestatísticos: o semivariograma e a krigagem.

### ***Variâncias da estimativa e de dispersão***

Uma das vantagens da Geoestatística é diferenciar bem e atribuir valores a duas variâncias: a de estimação e a de dispersão (ANDRIOTTI, 2003).

A variância da estimativa mede a qualidade da avaliação, apresentando a diferença quadrática entre o valor real e o valor estimado, ou seja, é possível calcular a precisão da estimativa  $Z$ , por meio de um estimador  $\hat{Z}_k$  (GUERRA, 1988; ANDRIOTTI, 2003),

$$\sigma_k^2 = E[Z - \hat{Z}_k]^2. \quad (14)$$

A variância de dispersão mede a variabilidade dos valores das amostras, em torno de um valor médio do fenômeno espacial em estudo (GUERRA, 1988).

### **2.4.3. Análise estrutural**

A análise estrutural é um dos objetivos principais da Geoestatística<sup>14</sup> e visa extrair da aparente desordem dos dados disponíveis, uma visualização de sua variabilidade e uma medida da correlação de valores tomados entre 2 (dois) pontos com localização distintas no espaço, realizado através do variograma (GUERRA, 1988).

#### **2.4.3.1. Variograma**

A determinação do variograma é descrita como o primeiro e mais importante passo no procedimento de uma estimativa geoestatística, pois o mesmo influenciará em todo processo de krigagem, nos resultados e conclusões. Na prática é a ferramenta matemática que permite, analisar a dispersão natural das variáveis regionalizadas, representando o grau de continuidade do fenômeno estudado (Equação 15) (GUERRA, 1988),

$$2\hat{\gamma}(h) = \frac{1}{n(h)} \sum_{i=1}^{n(h)} [Z(x_i) - Z(x_i + h)]^2, \quad (15)$$

---

<sup>14</sup> Segundo Guerra (1988) são dois objetivos principais da Geoestatística: a análise estrutural da variável em estudo através do variograma e sua estimativa por meio da krigagem. Os dois são ligados pela operação de amostragem do fenômeno em estudo.

onde:

- $Z(x_i)$  é o valor da variável no ponto  $x_i$ .
- $Z(x_i + h)$  é o valor da variável no ponto  $x_i + h$ .
- $n(h)$  é o número de pares, a uma distância  $h$ .

Segundo Andriotti (2003) é comum o uso na literatura da palavra variograma quando, na realidade está se tratando de semivariograma, e na prática o valor de um é o dobro do outro. O variograma ( $2\hat{\gamma}(h)$ ) é um parâmetro natural a estimar a partir das diferenças quadráticas médias e a krigagem utiliza a quantidade do semivariograma

De forma similar ao método estatístico, existem 3(três) tipos de variograma: o observado ou experimental, o verdadeiro e o teórico. Na prática, apenas o variograma experimental é conhecido e o variograma teórico serve de referência para estimar o variograma verdadeiro, a Figura 20 ilustra os três tipos de variograma (GUERRA, 1988).

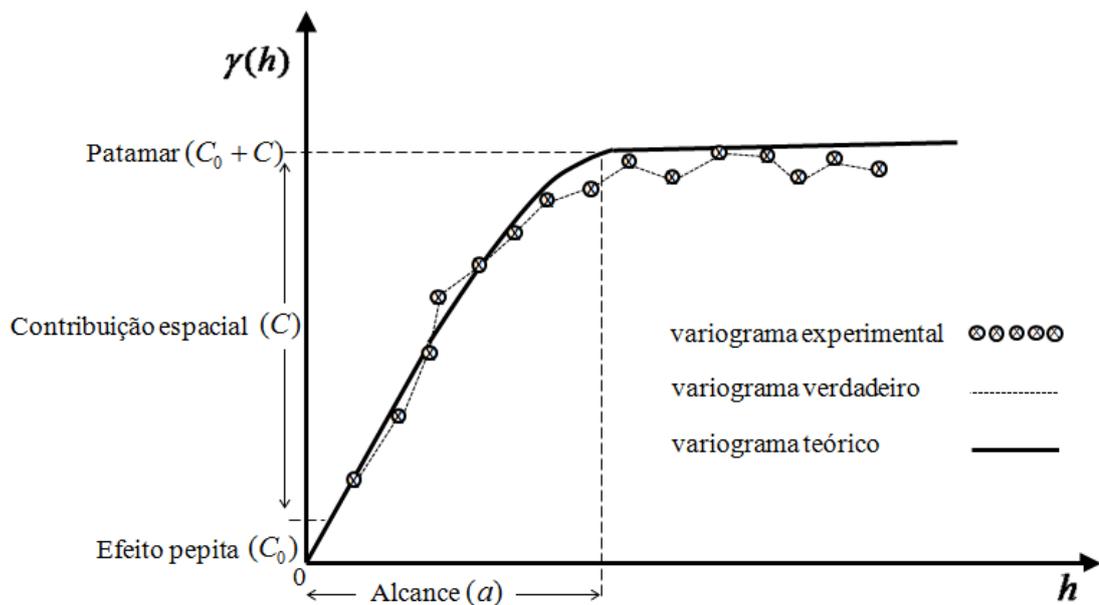


Figura 20 - Os três variogramas da Geoestatística

As diferenças quadráticas entre valores observados, considerando a distância  $h$  que os separa, permitem a construção do semivariograma que é a ferramenta base da Geoestatística, através da Equação 15 e ilustrado na Figura 21 (GUERRA, 1988).

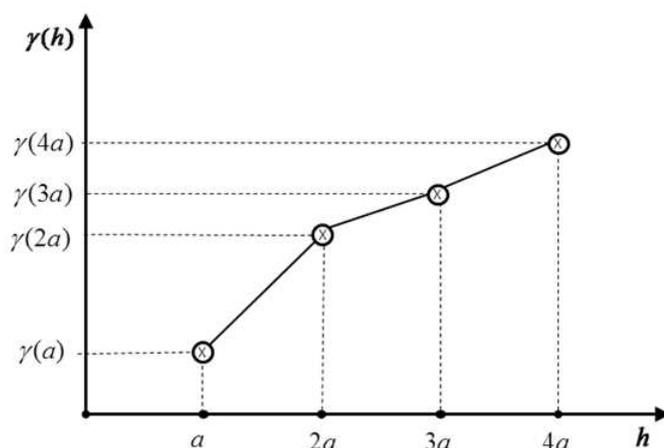


Figura 21 - Exemplo de construção de um semivariograma

Fonte: adaptado de Guerra (1988)

Os principais parâmetros de um semivariograma são ilustrados na Figura 22.

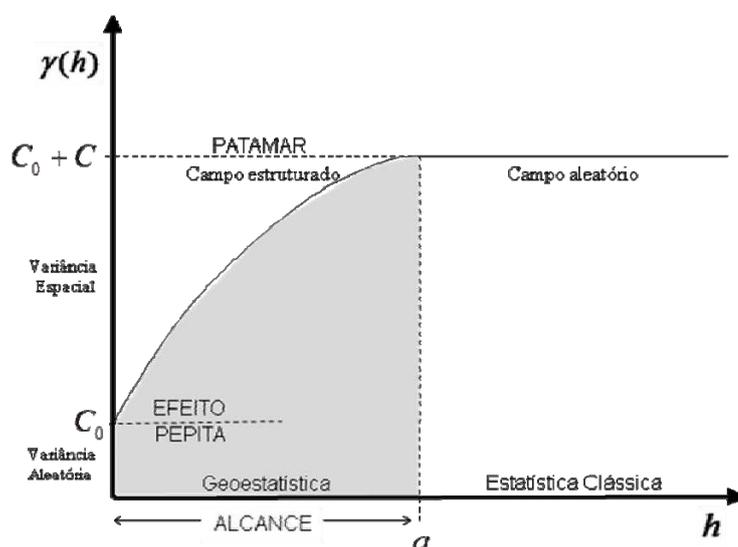


Figura 22 - Principais parâmetros do semivariograma

Fonte: adaptado de Andriotti (2003).

O alcance ( $a$ ) de um variograma é a distância a partir da qual as amostras passam a ser independentes, ou seja, a dispersão entre dois valores observados são casuais e sem correlação espacial, objeto da Estatística Clássica (ANDRIOTTI, 2003).

O efeito pepita<sup>15</sup> ( $C_0$ ) é a descontinuidade na origem do variograma sendo uma particularidade na correlação entre amostras, onde  $\gamma(0) = 0$  e  $\gamma(e) = Cov(0)$  (efeito pepita), sendo  $e$  a menor distância entre amostras. Este pode ser atribuído a erros de medição ou ao fato de que os dados não foram coletados a intervalos suficientemente pequenos para mostrar o comportamento espacial subjacente do fenômeno em estudo (GUERRA, 1988; LANDIM, 1998; VIEIRA, 2000).

<sup>15</sup> O efeito pepita recebeu este nome por ter sido utilizado a primeira vez nos depósitos auríferos da África do Sul (ANDRIOTTI, 2003).

O patamar ( $C_0+C$ ) de um variograma representa o valor da variância total da variável em estudo e equivalente ao ponto de sua estabilização, a variância espacial ou de dispersão é representada por ( $C$ ). A teoria mostra que a variância dos dados observados, da população estudada, equivale a este ponto de estabilização e limita o crescimento do semivariograma com patamar em função de  $h$  (GUERRA, 1988 e ANDRIOTTI, 2003).

O semivariograma expressa o comportamento espacial da variável regionalizada e seus resíduos, como: o tamanho da zona de influência em torno de uma amostra, a anisotropia e a continuidade através da forma do variograma<sup>16</sup> (LANDIM, 1998).

### ***Variograma experimental***

A construção do variograma experimental é sensível à localização dos pontos amostrais e ao tipo de distribuição estatística apresentada pelos dados observados. Em relação a distribuição espacial dos dados experimentais, esta pode ser regular ou irregular (YAMAMOTO e LANDIM, 2013).

No caso de uma estrutura regular dos pontos (Figura 23), por meio da Equação 15, o variograma é calculado diretamente através da formação de pares de pontos amostrais a uma determinada distância  $h$  (YAMAMOTO e LANDIM, 2013).

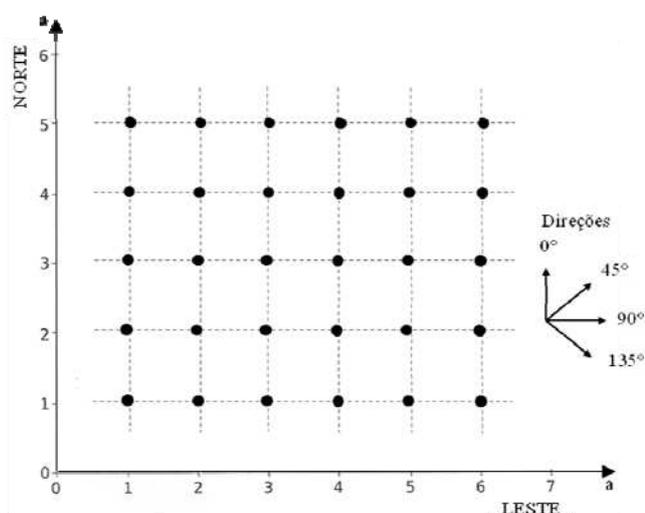


Figura 23 - Distribuição regular de amostras e a construção de variogramas experimentais

Fonte: adaptado de Camargo (2011)

A distância máxima recomendada para calcular o variograma experimental, denominada de campo geométrico, é igual à metade do comprimento da linha na direção considerada (JOURNAL e HUIJBREGTS, 2003).

<sup>16</sup> Segundo Guerra (1988) o comportamento na origem reflete o grau de continuidade do fenômeno em estudo, pois quanto mais rápido cresce o variograma, maior é a descontinuidade.

No caso da distribuição irregular, existe a necessidade de definir parâmetros adicionais além da direção e distância, agrupando os pontos amostrados em classes de ângulos e distâncias (Figura 24), para formar os pares de pontos de amostras e utilizar a Equação 15 para construção do variograma (YAMAMOTO e LANDIM, 2013).

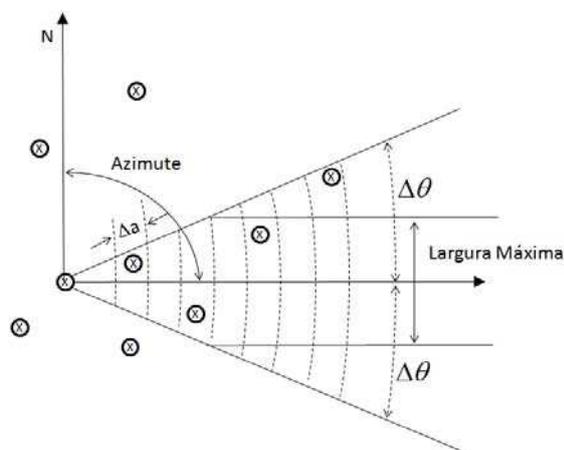


Figura 24 - Distribuição irregular de amostras e a construção de variogramas experimentais

Fonte: adaptado de Guerra (1988) e Yamamoto e Landim (2013).

Os parâmetros, ilustrados na Figura 24, para a construção de variogramas experimentais de dados observados de forma irregular são: o azimute ( $Az$ ), a tolerância angular ( $\Delta\theta$ ), passo ( $lag^{17}$ ), a tolerância do passo ( $\Delta a$ ) e a largura máxima, sendo que  $\Delta a$  e  $\Delta\theta$  devem ser o menor possível para se evitar distorções de análise de resultados (GUERRA, 1988; YAMAMOTO e LANDIM, 2013).

Segundo Guerra (1988) através da dispersão espacial ( $C$ ) e do efeito pepita ( $C_0$ ) é possível determinar um grau de aleatoriedade ( $\varepsilon$ ) presente nos dados em relação a sua regionalização,

$$\varepsilon = C_0 / C. \quad (16)$$

Entretanto, o autor ressalta que, é praticamente impossível diferenciar as causas do efeito pepita, sejam estas inconsistências na análise ou a baixa representatividade das amostras. A Tabela 5 sintetiza as respectivas classificações.

TABELA 5 - Classificação dos graus de aleatoriedade

Grau de aleatoriedade	Componente
$\varepsilon < 0,15$	Pequena
$0,15 \leq \varepsilon \leq 0,30$	Importante
$\varepsilon > 0,30$	Bastante importante

Fonte: baseado em Royle<sup>18</sup> (1979) apud GUERRA (1988); Yamamoto e Landim (2013)

<sup>17</sup> Segundo Camargo (2011) o termo "Lag" refere-se a uma distância pré-definida para o cálculo do semivariograma.

<sup>18</sup> ROYLE, A. G. **Why Geostatistics?** Engineering and Mining Journal. New York, 1979.

Segundo Andriotti (2003), o efeito pepita puro indica a ausência de correlação espacial entre amostras, onde os resultados da Geoestatística são similares aos da Estatística Clássica. Além disso, um valor de efeito pepita superior a cerca de 30% do patamar ( $C_0+C$ ) é considerado elevado.

De posse do variograma experimental, um modelo teórico é ajustado a este variograma observado.

### **Variogramas teóricos**

Segundo Guerra (1988) não é possível obter uma estimativa e a precisão desta estimativa utilizando os mesmos dados experimentais sem fazer uma hipótese teórica, assim é necessária uma forma matemática teórica do variograma. O estudo variográfico consiste em analisar qual variograma teórico melhor se ajusta ao variograma experimental, onde a partir do modelo teórico são feitas as inferências em relação ao variograma verdadeiro. Estes são classificados em: modelos teóricos com patamar e modelos teóricos sem patamar, conforme Tabela 6.

TABELA 6 - Equações de modelos de variogramas teóricos

<b>MODELOS TEÓRICOS COM PATAMAR</b>	
Esférico (Esquema de Matheron)	$\gamma(h) = C_0 + C \left[ (3/2)(h/a) - (1/2)(h^3/a^3) \right]$ quando $0 < h \leq a$ $\gamma(h) = C_0 + C$ quando $h > a$ $\gamma(0) = 0$
Exponencial (Esquema de Formery)	$\gamma(h) = C_0 + C(1 - e^{-h/a})$ quando $0 < h \leq a$ $\gamma(h) = C_0 + C$ quando $h > a$ $\gamma(0) = 0$
Gaussiano (Parabólico)	$\gamma(h) = C_0 + C(1 - e^{-h^2/a^2})$ quando $0 < h \leq a$ $\gamma(h) = C_0 + C$ quando $h > a$ $\gamma(0) = 0$
<b>MODELOS TEÓRICOS SEM PATAMAR</b>	
Linear	$\gamma(h) = ph$ , sendo $p$ a inclinação da reta;
Logarítmico (Esquema de Wijs)	$\gamma(h) = C_0 + 3\alpha \log_e(h)$ , sendo $\alpha$ é uma constante conhecida como "dispersão absoluta".

Fonte: baseado em Guerra (1988) e Yamamoto e Landim (2013).

Segundo Yamamoto e Landim (2013), dos modelos teóricos com patamar, o esférico, o exponencial e o gaussiano explicam a maioria dos fenômenos espaciais. A Figura 25 ilustra a relação entre os alcances destes variogramas teóricos com patamar.

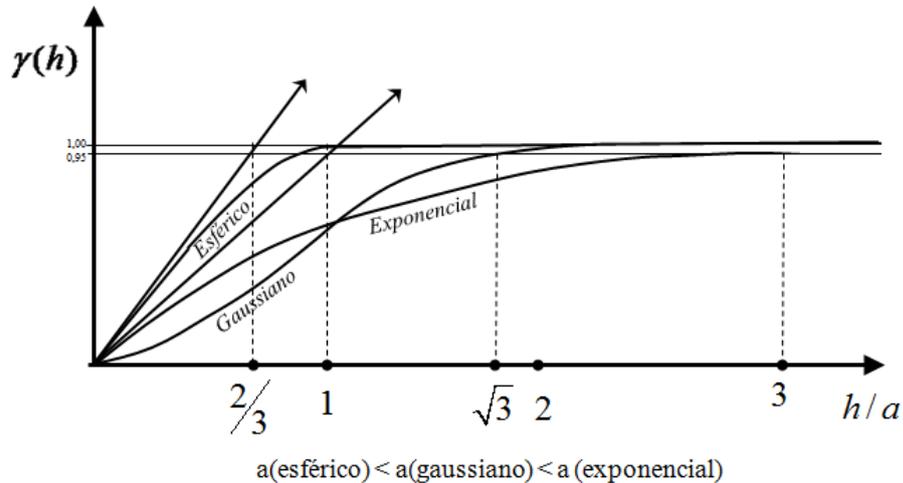


Figura 25 - Relação entre os alcances dos principais variogramas teórico com patamar

Fonte: adaptado de Guerra (1988)

O conhecimento dos parâmetros que caracterizam um variograma (alcance, patamar, efeito pepita, o comportamento na origem e o passo) permitem analisar qual modelo teórico melhor se ajusta ao variograma experimental. Entre outros modelos teóricos de variogramas cita-se: o efeito seno e o efeito cosseno, comuns quando os valores de  $h$ , representados no eixo das abscissas, são referentes ao tempo e não a distâncias (ANDRIOTTI, 2003).

#### 2.4.3.2. Características estruturais do variograma

O variograma pode representar características estruturais específicas de uma variável regionalizada, auxiliando no entendimento do fenômeno espacial em estudo. Estas características são: comportamento na origem, anisotropias, fenômenos de transição, tendências, efeito proporcional, efeito buraco, estruturas imbricadas, corregionalizações e a noção de suporte (GUERRA, 1988).

O comportamento na origem do variograma diz respeito a continuidade do fenômeno em estudo e pode ser diferenciado em quatro tipos: (a) parabólico, indicando uma continuidade estrita ou grande extensão; (b) contínuo na origem, mas com uma tangente oblíqua indicando uma continuidade média; (c) tangente vertical na origem indicando um efeito pepita ( $\gamma(h)=C_0$ ) quando a distância ( $h$ ) tende a zero; (d) efeito pepita puro indicando a aleatoriedade do fenômeno em estudo. As situações descritas são ilustradas na Figura 26 (GUERRA, 1988):

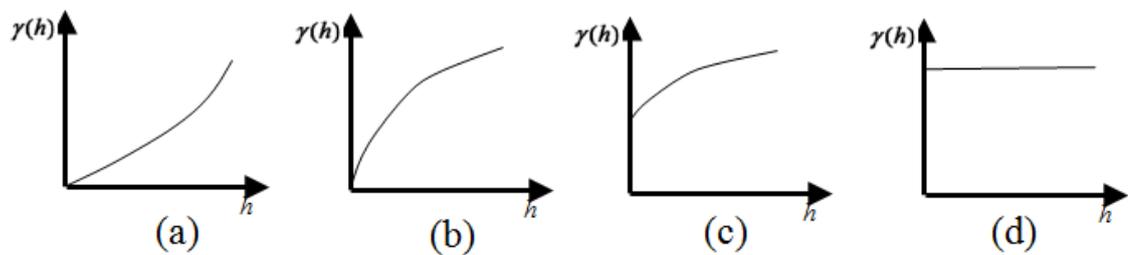


Figura 26 - Tipos de comportamento na origem do variograma

Fonte: Guerra (1988)

A anisotropia ocorre quando o fenômeno espacial em estudo apresenta variações em diferentes direções, caso contrário ocorre a isotropia, e ambas refletem a gênese da variável analisada. A partir da análise do variograma, a anisotropia é classificada em dois tipos: a anisotropia geométrica (mesmos patamares e diferentes alcances) e a anisotropia zonal (patamares diferentes) (ANDRIOTTI, 2003).

O fenômeno de transição é caracterizado pela tendência da variância dos dados observados atingir um patamar, determinado pela distância (alcance) a partir da qual um ponto da variável em estudo não tem mais influência sobre o ponto vizinho, definindo o início da zona de pura aleatoriedade (GUERRA, 1988).

As tendências ou presença de deriva ocorrem quando a variável em estudo tende a assumir um valor, por exemplo, crescente ou decrescente, conforme a distância  $h$  é aumentada, produzindo discrepâncias cada vez maiores. Nestas situações é necessário remover a tendência e construir o variograma dos resíduos (GUERRA, 1988).

O efeito proporcional caracteriza-se quando diversos variogramas absolutos se correspondem por uma afinidade que depende das médias experimentais. A princípio, pode ser confundido com um fenômeno de anisotropia, ou em alguns casos esconder a existência da mesma (GUERRA, 1988).

O efeito buraco reflete uma variável cujo fenômeno em estudo apresenta duas regionalizações distintas numa mesma direção, apresentando a estrutura conforme o variograma representado na Figura 27 (GUERRA, 1988).

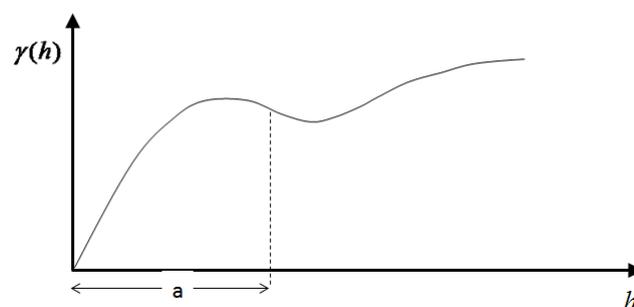


Figura 27 - Variograma indicando efeito buraco

Fonte: Guerra (1988)

As estruturas imbricadas do variograma experimental apresentam uma série de alcances e patamares, onde cada uma das estruturas na área de interesse aparece com um fenômeno de transição particular, conforme Figura 28 (GUERRA, 1988).

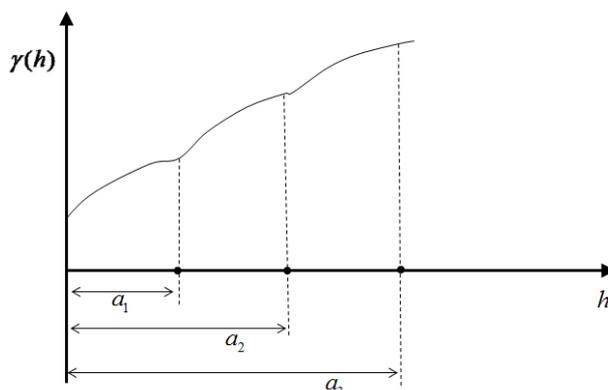


Figura 28 - Variograma indicando estrutura imbricada

Fonte: Guerra (1988)

A correionalização ocorre quando uma variável existente numa área de interesse indica a presença de outro fenômeno espacial relevante, para sua averiguação são realizadas as chamadas coestimativas<sup>19</sup>, através de variogramas cruzados<sup>20</sup> (GUERRA, 1998).

#### 2.4.3.3. Ajuste de variogramas

Em Geoestatística, a análise estrutural é a denominação referente às operações iniciais de compreensão do fenômeno em estudo, elaboração dos semivariogramas experimentais e o ajuste de modelos teóricos conhecidos. Neste estudo estrutural (estudo variográfico), a parte mais importante é o ajuste dos variogramas e requer experiência e paciência (GUERRA, 1988; LANDIM, 1998; ANDRIOTTI, 2003).

O ajuste dos variogramas teóricos aos experimentais é feito de modo iterativo, testando parâmetros de variograma (modelo de variograma, efeito pepita, alcance, patamar e número de estruturas presentes) até que o ajuste seja considerado satisfatório, conforme mostra o fluxograma da Figura 29. Uma vez escolhido o variograma teórico são aproveitadas suas propriedades para análise variográfica e fundamentalmente para avaliação do fenômeno em estudo, possibilitando inferências em relação ao variograma verdadeiro. Ressalta-se que o modelo ajustado não passa por todos os pontos, somente ajusta-se a eles, por isto é importante um bom ajuste aos primeiros passos do

<sup>19</sup> "A cokrigagem é similar à krigagem e permite estimar uma variável a partir das informações que se tem sobre ela própria e também a partir das informações disponíveis sobre outras variáveis que tenham correlação espacial com ela" (ANDRIOTTI, 2003).

<sup>20</sup> A aplicação da validação cruzada possibilita uma outra forma de aferir a qualidade das estimativas realizadas pela krigagem, onde o valor de uma amostra é eliminado e novamente estimado com base nos dados adjacentes, ou seja, este ponto de amostra é considerado como não coletado (ANDRIOTTI, 2003).

variograma, onde a confiabilidade é superior devido a maior quantidade de pares de pontos de amostras formados (GUERRA, 1988; ANDRIOTTI, 2003).

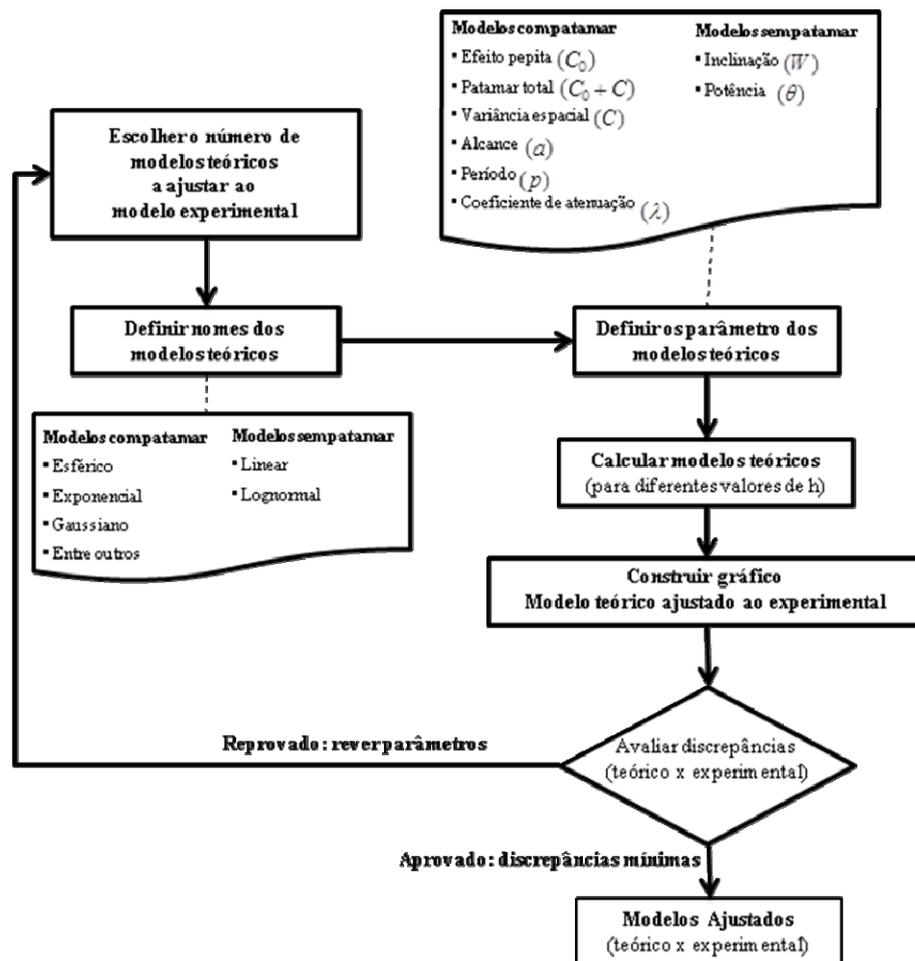


Figura 29 - Fluxograma do ajuste do variograma teórico ao experimental.

Fonte: baseado em Guerra (1988)

#### 2.4.4. Estimativas Geoestatísticas

A estimativa geoestatística mostra-se de um modo geral superior aos demais métodos de interpolação, uma vez que visa determinar a distribuição espacial e a acurácia do fenômeno em estudo, através da Krigagem conhecida como BLUP (*Best Linear Unbiased Predictor*). A qualidade da estimativa é vinculada a ideia da acuracidade (variabilidade) de cada ponto estimado, dimensionando o erro associado (GUERRA, 1988; ANDRIOTTI, 2003; YAMAMOTO e LANDIM, 2013).

Segundo Yamamoto e Landim (2013), quando não há correlação espacial, outros métodos de interpolação preconizados pela Estatística Clássica podem ser utilizados, conforme a Figura 30.

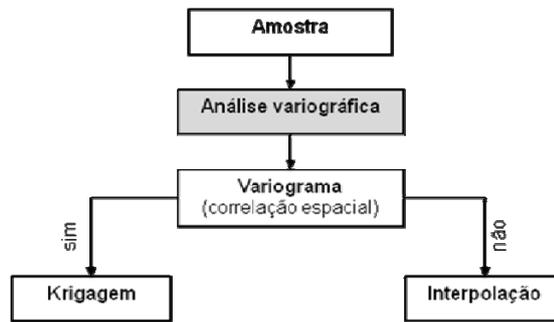


Figura 30 - Krigagem ou interpolação

Fonte: adaptado de Yamamoto e Landim (2013)

As variáveis regionalizadas podem ser contínuas ou discretas. As variáveis contínuas podem apresentar distintas características identificadas pela forma de seu respectivo histograma. As transformações de dados, em diversas circunstâncias são necessárias antes da estimativa geoestatística, ou seja, a metodologia geoestatística pode ser aplicada diretamente sobre os dados originais ou sobre os dados transformados conforme mostra a Figura 31 (YAMAMOTO e LANDIM, 2013).

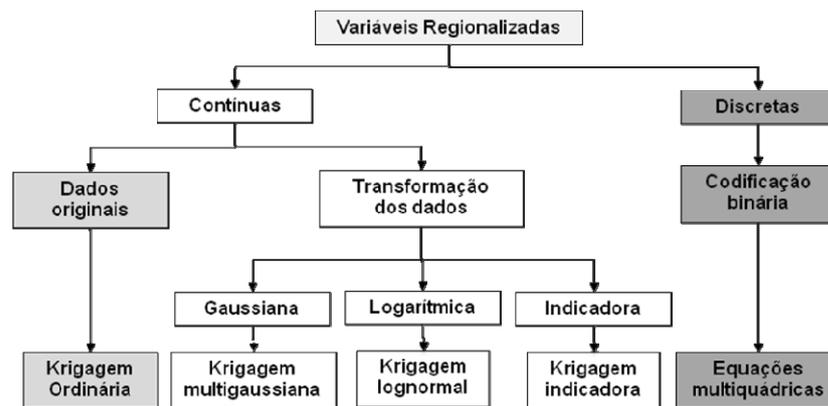


Figura 31 - Processo de estimativa geoestatística

Fonte: adaptado de Yamamoto e Landim (2013)

#### 2.4.4.1. Krigagem

A krigagem, do francês *krigeage* e do inglês *kriging*, é um termo cunhado pela escola francesa de geoestatística, desenvolvida por George Matheron, em homenagem ao engenheiro de minas sul-africano e pioneiro na aplicação de técnicas estatísticas na avaliação mineira, Daniel Gerhardus Krige (LANDIM, 1998).

A Krigagem é um processo geoestatístico capaz de estimar valores de variáveis distribuídas num espaço mensurável contínuo, como o geoespacial e o tempo, a partir de medidas adjacentes consideradas correlacionadas na análise variográfica e fornecer a acurácia desta estimativa. Para isto, a Krigagem considera informações oriundas do variograma para definição dos pesos ótimos associados a cada dado experimental, entre as quais cita-se: a quantidade de amostras, a localização dessas sobre a área de interesse

e seu intervalo de coleta, a região a ser estimada e a continuidade espacial do fenômeno em estudo. Ressalta-se que, estes ponderadores são alterados conforme ocorrem mudanças na distribuição espacial das amostras (LANDIM, 1998; ANDRIOTTI, 2003; YAMAMOTO e LANDIM, 2013).

Em Geoestatística, a krigagem é um sistema de equações que possibilita calcular o valor de cada um dos ponderadores, onde a variância de estimativa é mínima e a soma dos ponderadores corresponde a 1. Desta forma, a krigagem é um estimador não-enviesado ("*unbiased*"), ou seja, sem viés, conforme expressa a Equação 17 (GUERRA, 1988 e ANDRIOTTI, 2003),

$$\hat{Z}_k(x_0) = \sum_{i=1}^n \lambda_i \cdot Z(x_i), \quad (17)$$

onde:

$$\sum_{i=1}^n \lambda_i = 1 \text{ (condição de universalidade, sem tendência).}$$

$$\sigma_k^2 = E[Z - \hat{Z}_k]^2 \text{ é mínimo (condição de otimalidade).}$$

Na Equação 17,  $\lambda_i$  são os ponderadores associados aos dados experimentais  $Z(x_i)$ ;  $n$  é o número total desses dados;  $\hat{Z}_k$  é o estimador de krigagem e fornece a variância da estimação, o "chapéu" é uma designação de estimador; e  $Z$  é o valor observado (GUERRA, 1988; ANDRIOTTI, 2003).

Segundo Guerra (1988) e Andriotti (2003) o método de Lagrange<sup>21</sup> permite a resolução do sistema de equações da krigagem, onde conforme mencionado, é necessário minimizar a variância da estimação e respeitar a condição de não-enviesamento (tendência), estas são expressas matricialmente,

$$\begin{pmatrix} \bar{\gamma}(x_1, x_1) & \bar{\gamma}(x_1, x_2) \dots & \bar{\gamma}(x_1, x_n) & 1 \\ \bar{\gamma}(x_2, x_1) & \bar{\gamma}(x_2, x_2) \dots & \bar{\gamma}(x_2, x_n) & 1 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \bar{\gamma}(x_n, x_1) & \bar{\gamma}(x_n, x_2) \dots & \bar{\gamma}(x_n, x_n) & 1 \\ 1 & 1 \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \bar{\gamma}(x_1, V) \\ \bar{\gamma}(x_2, V) \\ \cdot \\ \cdot \\ \bar{\gamma}(x_n, V) \\ 1 \end{pmatrix}. \quad (18)$$

$$[A] \cdot [\lambda] = [B]$$

Desta forma, a variância de krigagem é dada por (ANDRIOTTI, 2003):

$$\sigma_k^2 = \sum_{i=1}^n \lambda_i \bar{\gamma}(x_i, V) + \mu - \bar{\gamma}(V, V). \quad (19)$$

<sup>21</sup> Segundo Andriotti (2003) Lagrange desenvolveu a técnica dos multiplicadores em 1797.

A Equação 18 representa um sistema de krigagem com  $n+1$  equações com  $n+1$  incógnitas,  $i= 1$  a  $n$  é a quantidade de observações,  $x_i$  as observações,  $\lambda_i$  os ponderadores de krigagem e o parâmetro  $\mu$  o Multiplicador de Lagrange,  $V$  é o bloco a estimar,  $\bar{\gamma}(x_1, x_2)$  corresponde ao variograma médio entre  $x_1$  e  $x_2$ ;  $\bar{\gamma}(x_1, V)$  é o variograma médio entre  $x_1$  e o bloco  $V$ ;  $\bar{\gamma}(V, V)$  é o variograma médio do bloco  $V$  consigo próprio. A matriz  $A$  é denominada Matriz de Matheron e é simétrica, ou seja, os valores de covariâncias  $\bar{\gamma}(x_i, x_j) = \bar{\gamma}(x_j, x_i)$  são iguais para quaisquer  $i$  e  $j$ . Todos os termos da diagonal principal da matriz  $A$  valem zero, pois a covariância  $\bar{\gamma}(x_i, x_i) = 0$  para qualquer  $i$ . Ressalta-se que o sistema de krigagem permanece inalterado se o sistema de coordenadas for modificado a partir da origem dos dados. Nas margens de uma região estudada, cresce as variâncias de estimação de krigagem em função de um menor número de vizinhos, o que é válido para a amostragem aleatória e em grade (GUERRA, 1988 e ANDRIOTTI, 2003).

Os métodos de estimativa podem ser classificados segundo as características da distribuição do histograma em linear e não linear. Dentre os métodos de estimativas cita-se a krigagem simples, krigagem ordinária, krigagem universal e krigagem intrínseca, para o tratamento de dados lineares; a krigagem lognormal, krigagem multigaussiana, krigagem indicativa, krigagem probabilística, e krigagem disjuntiva, para o tratamento de dados não lineares. Mas na realidade, os métodos não lineares fazem uso do estimador de krigagem ordinária<sup>22</sup> (LANDIM, 1998; YAMAMOTO e LANDIM, 2013).

Segundo Landim (2000), os métodos de estimativa linear mais usuais são a krigagem simples e a ordinária, podendo ser aplicadas diretamente sobre os dados originais ou transformados, e entre os métodos não lineares destaca-se a krigagem indicativa. **Krigagem linear**

#### ***Krigagem simples ou estacionária (KS)***

A krigagem simples (KS) é utilizada quando a média é assumida como estatisticamente idêntica para toda área. O conhecimento da média em uma determinada área de estudo exige muitos dados anteriores ao atual, permitindo uma estimativa da média da variável na região de interesse. Na prática, a média da variável regionalizada é considerada conhecida e o peso da média fornece um critério que permite testar a qualidade da krigagem (ANDRIOTTI, 2003).

---

<sup>22</sup> Segundo Yamamoto e Landim (2013) a tradução do termo em francês "*krigeage ordinarie*" deveria ser "krigagem normal", mas o termo "krigagem ordinária" está consagrado no Brasil.

Seja um local não amostrado  $x_0$  e  $n$  valores obtidos em pontos adjacentes. Uma estimativa linear ponderada desse local é dada por (JOURNEL<sup>23</sup>, 1989 apud YAMAMOTO e LANDIM, 2013),

$$Z_{KS}^*(x_0) = m_0 + \sum_{i=1}^n \lambda_i [Z(x_i) - m_i], \quad (20)$$

onde:

$m_i = E[Z(x_i)]$  são as médias, as quais são assumidas como conhecidas;

$m_0$  é a média no ponto  $x_0$  e  $\lambda_i, \{i=1, n\}$  são os pesos associados aos  $n$  dados.

Na condição de estacionaridade de segunda ordem, a média e a variância de todos os locais são constantes, dependendo apenas das distâncias euclidianas que os separam, sendo a média assumida para toda área (YAMAMOTO e LANDIM, 2013):

$$E[Z(x)] = m, \text{ ou seja, } m_0 = m_i. \quad (21)$$

Desta forma, o estimador da krigagem simples equivale a (JOURNEL, 1989 apud YAMAMOTO e LANDIM, 2013):

$$Z_{KS}^*(x_0) = m + \sum_{i=1}^n \lambda_i [Z(x_i) - m]. \quad (22)$$

### ***Krigagem ordinária (KO)***

A krigagem ordinária (KO) é utilizada quando a média é considerada flutuante ou móvel para toda área, ou seja, não exige o conhecimento da média e da estacionariedade da área em estudo, sendo por isto mais utilizada do que a krigagem simples (KS). A mesma, estima os valores do fenômeno em estudo em qualquer lugar, com exceção dos locais dos pontos de observação, onde o valor medido é mantido, e o erro médio quadrático desaparece (ANDRIOTTI, 2003).

A krigagem ordinária (KO) é a krigagem simples (KS) com a média local calculada pela krigagem da média (KM), onde seu estimador é (YAMAMOTO e LANDIM, 2013):

$$Z_{KO}^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i). \quad (23)$$

A fim de facilitar o entendimento, a Figura 32 mostra a distribuição de pontos amostrados para estimar um valor de um ponto  $x_0$ , na localização ( $x = 28,75; y = 21,25$ ). Os cálculos são detalhados em Yamamoto e Landim (2013), onde o autor a partir de um conjunto de dados disponibilizados para a aplicação da

---

<sup>23</sup> JOURNEL, A. G. **Fundamentals of geostatistics in five lessons**. Washington: American Geophysical Union, 1989.

krigagem ordinária, define um modelo teórico a utilizar, no caso o esférico, monta o sistema de equações de krigagem, vistos na Equação 18, determina os ponderadores da krigagem ordinária e desta forma estima o valor do ponto ( $x_0$ ), ou seja,  $Z_{KO}^*(x_0) = 11,131$  e a respectiva variância da krigagem  $\sigma_{KO}^2 = 9,084$ .

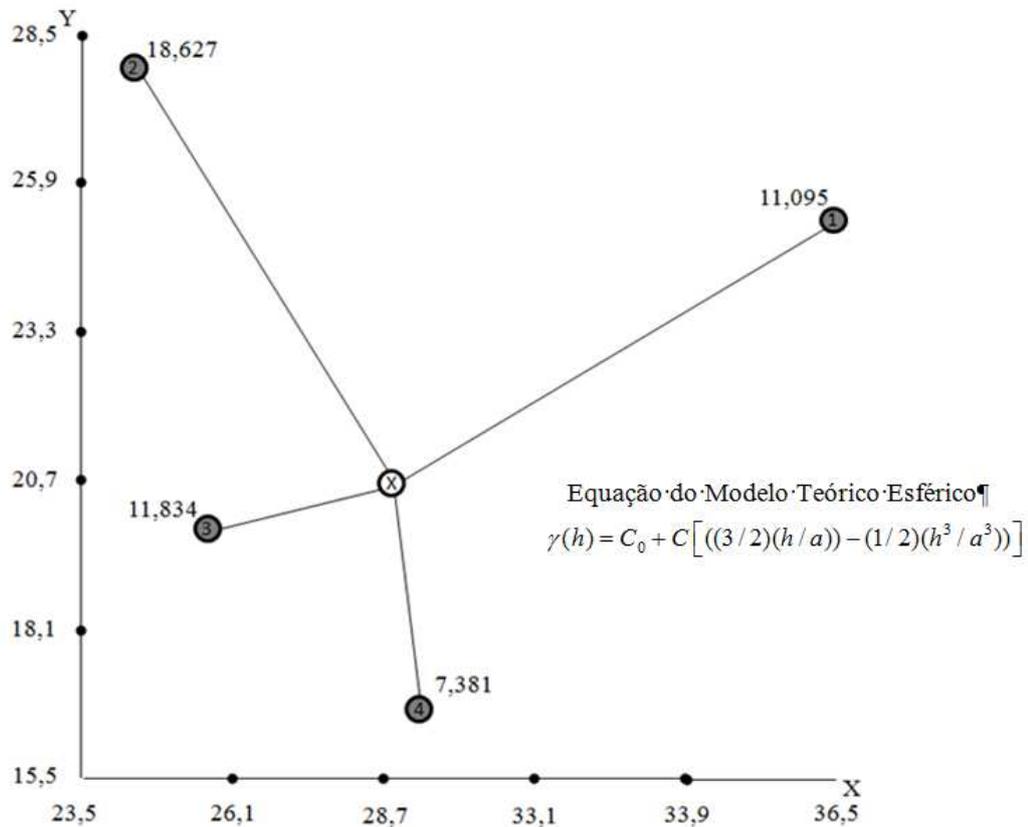


Figura 32 - Exemplo de estimação de um valor através da krigagem ordinária

Fonte: adaptado de Yamamoto e Landim (2013)

### ***Krigagem da média (KM)***

A krigagem simples (KS) presume hipoteticamente que a média seja conhecida e considerada constante em toda região dos pontos de observação. Entretanto, isto nem sempre é possível e a média deve ser considerada constante. Para isto, a média é estimada em torno de uma área de interesse por uma vizinhança com  $n$  pontos mais próximos. Desta forma, a média estimada é calculada por (WACKERNAGEL<sup>24</sup>, 1995 apud YAMAMOTO e LANDIM, 2013):

$$m^* = \sum_{i=1}^n \lambda_i^{KM} Z(x_i). \quad (24)$$

Sendo a média assumida para toda área:

$$E[Z(x)] = m \quad (25)$$

<sup>24</sup> WACKERNAGEL, H. **Multivariate geostatistics**. Springer-Verlag, 1995.

Seguindo os preceitos da krigagem, o erro de estimativa  $E[(\hat{m} - m)] = 0$ . O estimador é também não-enviesado ("*unbiased*"), ou seja, sem viés (tendência) (ANDRIOTTI, 2003 e YAMAMOTO, 2013).

$$\sum_{i=1}^n \lambda_i^{KM} = 1 \text{ (condição de universalidade)} \quad (26)$$

ou

$$E[\hat{m} - m] = 0$$

#### 2.4.4.3. Krigagem não linear

##### *Krigagem indicadora (KI)*

A krigagem indicadora (KI) (indicatriz ou indicativa) consiste na aplicação da krigagem ordinária (KO) sobre a variável transformada, ou seja, a variável resultante da aplicação da função não linear  $f(z) = 0$  ou  $1$ . O uso de variáveis indicadoras evita o problema de contaminação em função da presença de alguns valores extremos na interpolação de regiões com valores baixos (LANDIM, 2000; YAMAMOTO e LANDIM, 2013).

A krigagem indicativa exige que os valores da variável em estudo sejam transformados em indicadores binários (0 ou 1) e um conjunto de dados por indicação  $I(x)$  seja gerado a partir de um valor de corte (*cutoff*) ( $Z_c$ ), pré-estabelecido, aplicado a um conjunto de dados observados  $Z(x)$  (ANDRIOTTI, 2003):

$$I_c(x_i, Z_i) = \begin{cases} 1, & \text{se } Z(x_i) \geq Z_c \\ 0, & \text{se } Z(x_i) < Z_c \end{cases}, \text{ ou vice-versa para os sinais da desigualdade.} \quad (27)$$

O variograma das variáveis binárias pode ser útil por mostrar padrões de continuidade espaciais, não identificados no variograma dos dados originais devido a presença de valores discrepantes (*outliers*) (ANDRIOTTI, 2003).

O estimador da krigagem indicativa é expresso conforme a Equação 28 (JOURNEL<sup>25</sup>, 1980 apud YAMAMOTO e LANDIM, 2013),

$$I_{KO}^*(x_o, Z_c) = \sum_{i=1}^n \lambda_i I(x_i, Z_c). \quad (28)$$

A acurácia da estimativa indicativa pode ser medida através da variância de interpolação, expressa segundo a Equação 29 (YAMAMOTO, 2012),

$$S_O^2(x_o; Z_c) = I_{KO}^*(x_o; Z_c)(1 - I_{KO}^*(x_o; Z_c)). \quad (29)$$

<sup>25</sup> JOURNEL, A. G. **The lognormal approach to predicting local distributions of selective mining unit grades.** Mathematical Geology, v. 12, n. 4, p.285-303. 1980.

A krigagem indicativa mostra uma superfície de probabilidades vinculada a ocorrência de um determinado valor de corte ( $Z_c$ ), indicando a presença de anomalias. Esta superfície probabilística é associada a um mapa de variâncias o qual mostra as incertezas em função da interpolação superficial (ANDRIOTTI, 2003).

A presente seção mostrou os conceitos de Geoestatística e suas ferramentas necessárias à análise do comportamento espacial da variável em estudo, fornecendo subsídios de "como" avaliar e representar os resultados das inspeções de qualidade por meio de mapas de qualidade.

#### **2.4.4.4. Validação cruzada**

Segundo Andriotti (2003) uma outra forma de aferir a qualidade das estimativas realizadas pela krigagem é a aplicação da validação cruzada, onde a mesma pressupõe que uma amostra não tenha sido coletada, para isto seu valor é eliminado e é novamente estimado com base nos dados vizinhos. Na prática, a validação cruzada é o procedimento no qual cada amostra é removida do conjunto de dados sendo realizada uma krigagem para aferir seu valor, a diferença entre valores amostrados e estimados será o mais próximo de zero quanto melhor for a estimativa. O intuito é utilizar os resíduos para verificar a performance do modelo, entretanto a validação cruzada não prova que o modelo de variograma esteja correto, mas garante que o mesmo não seja de má qualidade. Desta forma, se a validação cruzada apresentar os resultados sobre uma reta de regressão próxima da bissetriz e com pequena variabilidade dos pontos em torno dessa reta, conclui-se que foi realizado uma boa estimativa dos pontos. A regressão linear simples é dada pela Equação 30 de uma reta:

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (30)$$

onde:

- $\beta_0$  representa o coeficiente angular da reta
- $\beta_1$  representa o ponto em que a reta intercepta o eixo das ordenadas
- $\varepsilon$  representa o resíduo entre o valor amostrado e o estimado.

Desta forma, o resultado da regressão da validação cruzada idealiza que  $\beta_1$  seja próximo de 1 e  $\beta_0$  seja próximo de zero, indicando um bom ajuste do variograma teórico ao variograma experimental.

O fluxograma ilustrado na Figura 33, sintetiza os procedimentos do método geoestatístico: análise exploratório dos dados observados; análise variográfica do fenômeno espacial em estudo e ajuste do variograma teórico ao experimental; a validação cruzada; e a krigagem.

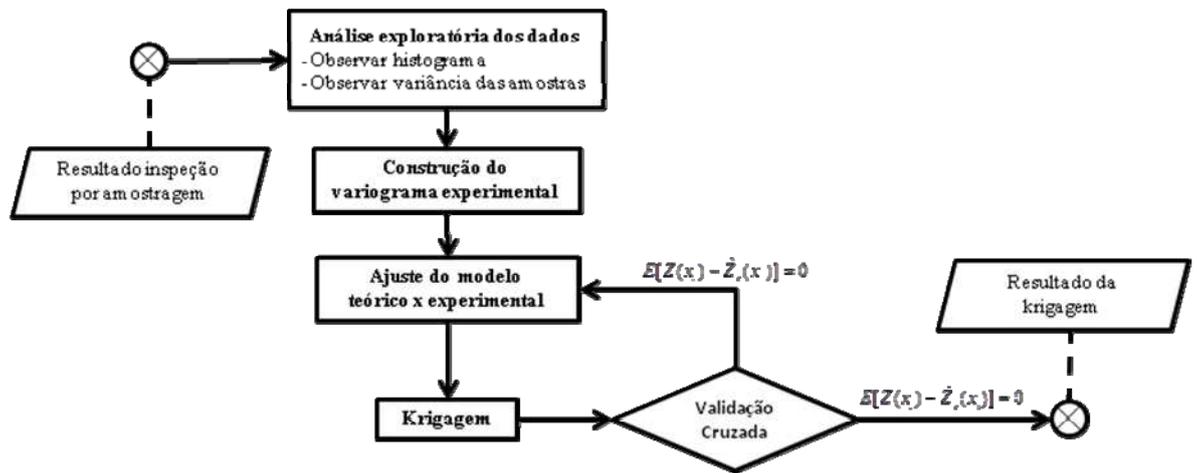


Figura 33 - Síntese do método geoestatístico

## 2.5. Metadados: informando a qualidade de dados ao usuário

Neste tópico são apresentados os conceitos referentes a metadados, a definição de perfis de metadados e sua importância na construção de IDE's e informação ao usuário, focando nos aspectos relativos a descrição da qualidade de dados espaciais.

O Perfil de Metadados Geoespaciais do Brasil (MGB) é baseado na norma ISO 19115:2003, que faz parte de um conjunto de normas referentes à informação espacial. A norma referente aos metadados, junto com os elementos de qualidades descritos nas normas 19113:2002 e 19114:2003, auxiliam na composição de medidas de qualidade, que se apresentam como um catálogo de rotinas de avaliação do conjunto de dados espaciais (CONCAR, 2009).

Os metadados, a princípio, aparentam apresentar um caráter meramente descritivo dos dados, mas sua aplicação e uso vão, além disto, sendo também capazes de qualificar um conjunto de dados, informando ao usuário os relatos das avaliações de qualidade realizadas sobre o produto em questão.

A popularização do uso de Sistemas de Informação Geográfica (SIG), por especialistas em geociências e diversos outros profissionais, ampliou a produção de dados geoespaciais, mas em muitos casos as especificações técnicas e de qualidade do produto cartográfico são desconhecidas, inviabilizando a reutilização e integração com outros conjuntos de dados. A cartografia digital retrata esta situação, onde o aumento da disponibilidade de banco de dados geoespaciais e seu mau uso, explodiu na demanda por metadados com informações relativas à acurácia dos dados. A manipulação de grandes volumes de dados espaciais tornou-se uma realidade em várias instituições brasileiras. A fim de criar condições mais favoráveis de gerenciar, localizar, recuperar e distribuir dados geoespaciais com informações de sua qualidade através das IDE é

recomendado a utilização de metadados (ROBINSON *et al.*, 1995; FREITAS, 2005; NOGUERAS, 2005).

O conhecimento humano sempre se apoiou no registro de dados e informações, ou seja, a documentação de informações tem acompanhado a evolução da humanidade (IBGE, 2011). A documentação sobre os dados gerados, denominada de metadados, deve ser padronizada e estruturada, promovendo aos usuários o conhecimento de sua descrição técnica e a adequabilidade dos dados a sua aplicação (FREITAS, 2005).

Segundo a CONCAR (2009), conforme definido no Art. 2º do Decreto Lei nº 6.666 de 27 de novembro de 2008, referente à INDE, os metadados são "*um conjunto de informações descritivas sobre os dados, incluindo as características de seu levantamento, produção, qualidade e estrutura de armazenamento, essenciais para promover a sua documentação, integração e disponibilização, bem como possibilitar sua busca e exploração*".

*"metadados é um vocábulo composto pela junção do prefixo 'meta' que significa 'além de', e da palavra 'dados' que é a medida da presença de um determinado conjunto de valores de uma variável ou fenômeno ou tema, ou ainda o valor para uma media observada. Este vocábulo metadados adquire o significado popularizado de 'os dados sobre o dado'"(IBGE, 2011).*

Ao usar fontes de mapa existentes para a compilação, na cartografia analógica era cumprida a regra "*sempre compilar a partir da escala maior para a escala menor*". Na cartografia digital, não é possível identificar se um banco de dados geoespacial seguiu esta regra. Assim, é imperativo que todos os dados indiquem a qualidade dos mesmos para evitar o uso indevido (ROBINSON *et al.*, 1995).

O país passa por um processo de reordenação de sua produção cartográfica com várias normas em uso e em desenvolvimento para o mapeamento sistemático, assim como a obrigatoriedade da elaboração de metadados pela instituição pública produtora de um mapeamento.

Segundo Robinson *et al.* (1995), um catálogo de metadados é a chave para identificar os dados desejáveis. Idealmente todos os elementos de metadados deveriam ser registrados automaticamente pelo software SIG, mas apenas algumas informações são coletadas. Alguns elementos, como os relacionados à qualidade dos dados, exigem mais pesquisas para desenvolver medidas descritivas quantitativas a serem gravadas nos metadados.

Conforme identificado por Freitas (2005), a construção de um catálogo de metadados traz vantagens significativas para as organizações produtoras de dados e para os usuários finais, fornecendo informações para identificação, exploração e utilização dos dados e bases geoespaciais. Para atingir estes objetivos, esforços internos e externos

nas instituições partícipes, de preferência em nível nacional e internacional, são necessários. Além disto, a autora ainda ressalta que:

*"o desenvolvimento de catálogos de metadados espaciais nas instituições provedoras de dados e informações identifica o seu compromisso com o compartilhamento de informações sobre a realidade espacial e corrobora com a ampliação do conhecimento na sociedade sobre o país" (FREITAS, 2005).*

### **2.5.1. Padrões de metadados geoespaciais**

Os metadados geoespaciais visam a descrição textual, de forma normatizada, do conjunto de dados geográficos proporcionando a sua identificação e compartilhamento (LEME, 2006). Entretanto, ao abordar o tema "padrões" existe uma contradição, constatada por Lisboa Filho *et al.*, (2012), na qual na busca por um "padrão" único houve uma "despadronização".

Os padrões de metadados mais utilizados e difundidos em diversas áreas do conhecimento são o *Dublin Core*, o padrão americano FGDC do *United States Federal Geographic Data Committee*, e o padrão da ISO (*International Organization of Standards*).

A norma ISO 19115:2003 é um dos padrões mais populares para metadados geoespaciais, pois combina aspectos de diversos padrões de metadados compondo um padrão universal de armazenamento e distribuição de metadados geoespaciais. Um dos aspectos negativos da norma elaborada pela ISO é sua complexidade e o fato de não ser gratuita (CONCAR, 2009; PRADO *et al.*, 2010).

Segundo CONCAR (2009), os padrões de metadados estão conceituados e estruturados em seções com funções específicas de: identificar o produtor e a responsabilidade técnica de produção, padronizar a terminologia utilizada, garantir a transferência de dados, viabilizar a integração de informações, possibilitar o controle de qualidade e garantir os requisitos mínimos de disponibilização.

As seções definidas nos diversos padrões de metadados constituem 3 (três) níveis: metadados de descoberta, metadados de exploração e metadados de utilização (GSDI, 2004). Segundo Freitas (2005), os metadados de descoberta informam ao usuário o conteúdo do conjunto de dados, o formato e a área de abrangência, contemplando questões referentes ao "*o que, quem, onde, como e quando*", permitindo ao usuário decidir se o conjunto de dados é potencialmente útil. Os metadados de exploração são correlacionados com a qualidade de dados, pois informam ao usuário se o conjunto de dados contém informações relevantes e aptas a sua aplicação, garantindo um uso correto e coerente dos mesmos. Os metadados de utilização informam como o conjunto de dados pode ser obtido, os requisitos computacionais, os direitos autorais e

as restrições e responsabilidades de uso. Neste nível são informados, opcionalmente, contatos adicionais para quaisquer dúvidas na utilização dos dados.

### **2.5.2. A norma ISO 19115:2003**

A norma ISO 19115:2003 é ampla, possui mais de 300 elementos, e de difícil aplicação, pois foi elaborada com o objetivo de caracterizar qualquer tipo de informação geográfica. O relacionamento entre classes é confuso e, algumas vezes, determinadas classes acabam sendo compostas por outras. Além disso, a ISO 19115 não é disponibilizada gratuitamente (CONCAR, 2009; PRADO *et al.*, 2010).

A norma ISO 19115:2003 combina aspectos de muitos outros padrões de metadados, visando um padrão universal para o armazenamento e distribuição de metadados geoespaciais. Embora, em termos gerais, a estrutura seja similar à do padrão FGDC (*The Federal Geographic Data Committee*), ou seja, um padrão com elementos obrigatórios e opcionais, a partir do qual podem ser derivados perfis. O padrão ISO 19115 apresenta diferenças marcantes em sua implementação, através do recurso de modelagem orientada a objetos, baseada no esquema UML (FREITAS *et al.*, 2010).

No Brasil, a Infraestrutura Nacional de Dados Espaciais - INDE consolidou o Perfil de Metadados Geoespaciais do Brasil (Perfil MGB), que foi criado pelo Comitê de Estruturação de Metadados Geoespaciais (CEMG) da Comissão Nacional de Cartografia (CONCAR) coordenado pelo IBGE. A criação desse perfil foi realizada com base na norma ISO 19115:2003, principalmente por ser uma norma reconhecida e adotada internacionalmente, por suportar, além do inglês, outros idiomas; além de fazer uso de listas controladas de códigos (*codelists*) no lugar de textos livres, tornando a interoperabilidade de metadados mais efetiva nesse padrão.

Segundo CONCAR (2009), a norma ISO 19115:2003 é composta por um conjunto de 326 elementos de metadados, organizados em 92 classes que caracterizam a informação, as aplicações e os serviços relacionados à informação geográfica. Destes elementos, 8 (oito) são obrigatórios para a criação de um determinado perfil de metadados baseado nesse padrão. Esta base mínima de elementos seria o denominador comum a todos os perfis de metadados, o que garante a interoperabilidade entre as diversas implementações. Esse padrão apresenta 23 elementos obrigatórios contemplados nas seções de: identificação, identificação do conjunto de dados geográficos - CDG, informação de restrição, qualidade, informação de manutenção, informação da representação espacial, sistema de referência, informação do conteúdo, distribuição e metadados.

### 2.5.3. Perfil de Metadados Geoespaciais do Brasil (MGB)

Um perfil de metadados geoespaciais, conforme ilustra a Figura 34, deve conter um conjunto básico de elementos que retrate as características dos produtos cartográficos derivados daqueles dados, e garanta sua identificação, exploração e utilização consistente. Esse conjunto básico é proposto como o núcleo comum a todos os tipos de produtos cartográficos. Os produtos de mapeamento especial, cadastral e temático requerem maior detalhamento dos itens de algumas seções dos metadados para retratar suas especificidades (CONCAR, 2010).

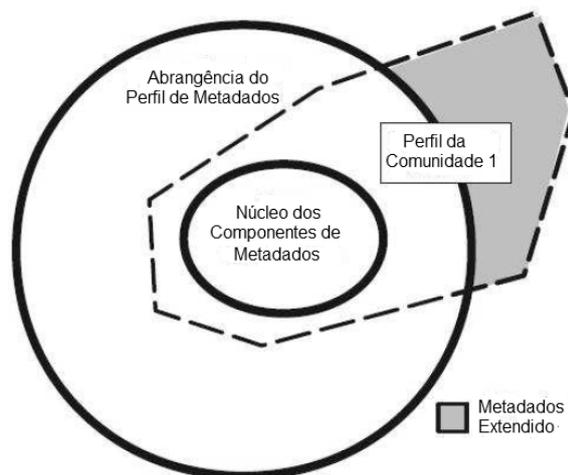


Figura 34 - Perfil de metadados de uma comunidade  
Fonte: adaptado da ISO 19115:2003

As seções de metadados presentes no Perfil MGB são ilustradas na Figura 35.

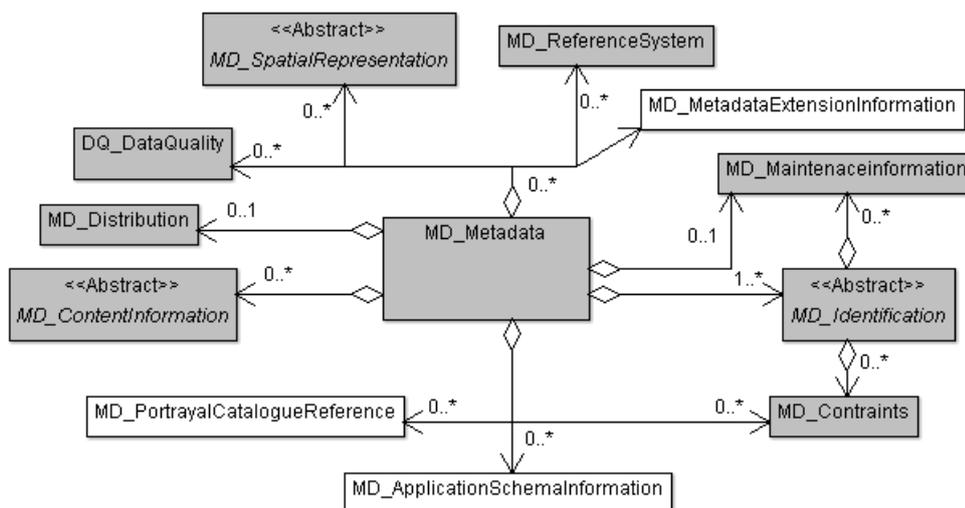


Figura 35 - Diagrama UML das seções de metadados do Perfil MGB  
Fonte: adaptado de CONCAR (2009)

Uma versão resumida do Perfil MGB foi criada para atender as organizações e usuários que porventura não disponham de informações que contemplem o perfil completo, conforme Tabela 7.

TABELA 7 - Entidades e elementos do Perfil MGB sumarizado

ENTIDADE/ELEMENTO	CONDIÇÃO	ENTIDADE/ELEMENTO	CONDIÇÃO
1 Título	obrigatório	13 Sistema de referencia	obrigatório
2 Data	obrigatório	14 Linhagem	opcional
3 Responsável	obrigatório	15 Acesso online	opcional
4 Extensão geográfica	condicional	16 Identificador metadados	opcional
5 Idioma	obrigatório	17 Nome padrão de metadados	opcional
6 Código de caracteres	condicional	18 Versão da norma de metadados	opcional
7 Categoria temática	obrigatório	19 Idioma de metadados	condicional
8 Resolução espacial	opcional	20 Código de caracteres do metadado	condicional
9 Resumo	obrigatório	21 Responsável pelos metadados	obrigatório
10 Formato de distribuição	obrigatório	22 Data dos metadados	obrigatório
11 Extensão temporal e altimétrica	opcional	23 Status	obrigatório
12 Tipo de representação espacial	opcional		

Fonte: CONCAR (2009)

#### 2.5.4. Informações de qualidade de dados espaciais

Uma das tarefas mais difíceis da cartografia é indicar aos usuários dos mapas a sua qualidade. *"Ao escrever ou falar, são utilizadas palavras como "quase", "cerca de", e "aproximadamente", nestas palavras as pessoas sabem o quanto preciso quero ser, mas não é fácil fazer a mesma coisa em um mapa"* (ROBINSON *et al.*, 1995).

A honestidade intelectual é vital em cartografia, pois o mapa possui aspecto oficial de verdade e precisão. Assim, é necessário assegurar que a informação representada está correta e ter a certeza de que os mapas não transmitem uma sensação maior de integridade e confiabilidade a qual se justificam (ROBINSON *et al.*, 1995).

A seção de qualidade de dados presente no Perfil MGB descreve, de maneira generalizada, a qualidade de uma série de conjunto de dados, feições, atributos, etc, e contempla informações sobre as fontes dos insumos e uma explicação geral do processo de produção (linhagem) usados para a criação dos dados (IBGE, 2011).

Segundo a norma ISO 19115:2003, o pacote referente a Qualidade de Dados (*DQ\_DataQuality*), representado na Figura 35, apresenta uma avaliação geral da qualidade do conjunto de dados, entretanto a mesma é opcional e contempla o âmbito da avaliação da qualidade. Esta entidade possui duas classes agregadas a Linhagem (*LI\_Lineage*) e o Elemento de Qualidade de Dados (*DQ\_Element*), este último por sua vez pode ser especificado como Completude (*DQ\_Completeness*), Consistência Lógica (*DQ\_LogicalConsistency*), Acurácia Posicional (*DQ\_PositionalAccuracy*), Acurácia Temática (*DQ\_ThematicAccuracy*) e Acurácia Temporal (*DQ\_TemporalAccuracy*), referentes aos elementos de qualidade de dados, podendo conter mais de uma subclasse para os respectivos subelementos de qualidade de dados, os atributos desta classe são relativos as medidas de qualidade, citadas na ISO 19114:2003.

No Perfil MGB observa-se que a proposta da seção de qualidade é uma descrição geral de como o conjunto de dados foi produzido, ou seja, sua linhagem. As informações relativas a avaliação dos elementos de qualidade não é ressaltada.

Segundo Freitas (2005) a existência de medidas de qualidade de dados é fundamental para avaliar a confiabilidade dos resultados obtidos a partir de aplicações efetuadas com o conjunto de dados. Em relação ao relato dos resultados da aplicação destas medidas de qualidade, a Tabela 8 apresenta os componentes dos metadados de qualidade de um conjunto de dados geoespacial, conforme as normas ISO 19114:2003 e ISO 19115:2003.

TABELA 8 - Componentes de metadados de qualidade de dados espaciais

<b>Componente DQ (<i>Data Quality</i>)</b>	<b>Descrição / Exemplos</b>		
<i>Data Quality Scope</i> (DQ_Scope)	Âmbito ao qual se aplica a informação da qualidade de dados. Exemplo: Todos os itens classificados como rios		
<i>Data Quality Element</i> (DQ_Element)	Elementos de Qualidade de Dados (Domínio enumerado) 1 - Completude 2 - Consistência Lógica 3 - Acurácia Posicional 4 - Acurácia Temporal 5 - Acurácia Temática		
<i>Data Quality Subelement</i> (DQ_Subelement)	(Completude) 1 - Comissão 2 - Omissão	(Consistência Lógica) 1 - Consistência Conceitual 2 - Consistência de Domínio 3 - Consistência de Formato 4 - Consistência Topológica	(Acurácia Posicional) 1 - Acurácia Absoluta/ Externa 2 - Acurácia Relativa/ Interna 3 - Acurácia dos dados em grade
		(Acurácia Temporal) 1 - Acurácia do tempo medido 2 - Consistência Temporal 3 - Validade Temporal	(Acurácia Temática) 1 - Classificação Correta 2 - Atributos não quantitativos corretos 3 - Acurácia dos atributos quantitativos
<i>Data Quality Measure</i> (DQ_Measure)			
<i>Data Quality Measure Description</i> (DQ_MeasureDesc)	Descrição do método de avaliação (texto livre) Exemplo: Aprovar / Reprovar		
<i>Data Quality Measure Identification</i> (DQ_MeasureID)	Código identificando um procedimento padrão registrado (domínio enumerado) Exemplos: 10101, 10201		
<i>Data Quality Evaluation Method</i> (DQ_EvalMethod)	Tipo de método usado para avaliar a qualidade do conjunto de dados (domínio enumerado).		
<i>Data Quality Evaluation Method Type</i> (DQ_EvalMethodType)	1 - Interno 2 - Externo 3 - Indireto		
<i>Data Quality Evaluation Method Description</i> (DQ_EvalMethodDesc)	Descrição do método de avaliação Exemplo: Comparar a contagem de itens no conjunto de dados contra a contagem de itens no universo de discussão.		
<i>Data Quality Result</i> (DQ_QualityResult)			
<i>Data Quality Value Type</i> (DQ_ValueType)	Tipo de valor (domínio enumerado) 1 - Variável Booleana 2 - Número 3 - Proporção (ratio) 4 - Porcentagem 5 - Amostra		6 - Tabela 7 - Imagem binária 8 - Matriz 9 - citação 10 - Texto Livre 11 - Outro
<i>Data Quality Value</i> (DQ_Value)	Depende do tipo de valor Exemplo: Falso (Tipo de Valor = Variável Booleana)		
<i>Data Quality Value Unit</i> (DQ_ValueUnit)	Unidade do valor (Depende do tipo de valor) Exemplo: NA		
<i>Data Quality Date</i> (DQ_Date)	Data ou período de data em que a medida de qualidade foi aplicada. Exemplo: 2000-03-05		
<i>Data Quality Conformance Level</i> (DQ_ConformanceLevel)	Valor (ou conjunto de valores) obtidos a partir da aplicação de uma medida de qualidade de dados ou do resultado da avaliação do valor obtido (ou conjunto de valores), contra um nível de qualidade aceitável conforme especificada. Exemplo: Zero comissões no conjunto de dados		

Fonte: adaptado da norma ISO 19114:2003

Iniciativas como a IDE Geominas<sup>26</sup>, podem facilitar o acesso às pesquisas acadêmicas, metodologias e seus dados, corroborar os princípios da INDE e compor um de seus nós, possibilitando a integração proativa entre instituições acadêmicas, órgãos públicos e/ou empresas privadas de mapeamento (LISBOA FILHO *et al*, 2011).

Nesta seção foi apresentada uma visão geral do tema metadados, incluindo conceitos, padrões de metadados e perfis, associados à informações relativas a qualidade do conjunto de dados espaciais, a fim de promover o acesso as informações de qualidade aos usuários do produto cartográfico. Esta seção encerra o ciclo de fundamentos teóricos relativos a aferição, representação e relato da qualidade de dados geoespaciais presentes nesta pesquisa. O ciclo se inicia na produção cartográfica, focando nos elementos de qualidade e operadores de generalização, em sequência pela aferição através de normas internacionais, análise geoestatística dos resultados para identificar a correlação espacial e se finaliza com a informação de qualidade ao usuário final por meio dos metadados. Dentre os termos citados anteriormente: "o que", "quanto" e "como" avaliar, pode-se incluir o último e mais importante "para quem" avaliar.

---

<sup>26</sup> O portal da IDE Geominas é < <http://www.ide.ufv.br/geominas>>

### 3. MATERIAIS E MÉTODOS

Neste capítulo, seguindo o objetivo principal desta pesquisa, de estudar o uso da Geoestatística aliada a adoção de normas internacionais para avaliação da qualidade de produtos cartográficos disponibilizados na INDE e representação espacial do resultado desta aferição, foram utilizados dados e insumos disponíveis à sociedade, no sistema de coordenadas geográficas e referencial geodésico SIRGAS 2000, além de softwares livres da área de geoprocessamento e geoestatística. Os métodos são descritos e apresentados de forma a compor uma metodologia de inspeção de produtos cartográficos, baseados em normas internacionais de qualidade de dados espaciais, técnicas de amostragem, análise espacial dos resultados da inspeção de qualidade e preenchimento de metadados de qualidade.

#### 3.1. Descrição geral do conjunto de dados em estudo

Dentre os conjuntos de dados espaciais utilizados, a Base Cartográfica Contínua do Brasil na escala 1:250.000 - BC250 foi escolhida para aferição de sua completude, em termos de hidrografia, e da acurácia posicional, em termos de planimétricos. Esta escolha atende a um dos propósitos desta pesquisa de avaliar a qualidade de dados geoespaciais disponibilizados na INDE. Este produto cartográfico, até a presente data, encontra-se em processo de integração dos 11 (onze) blocos de trabalho, divididos aproximadamente, segundo Reis *et al.* (2011), em função dos biomas do Brasil, conforme mostra a Figura 36, sendo acessíveis na página do *site* do IBGE<sup>27</sup>. Nesta dissertação, a referida base cartográfica foi tratada como contínua, sem abordar a análise por blocos, visto que, caso ocorra discrepâncias entre os blocos nas características de qualidade avaliadas, esta será representada no resultado final através dos mapas de qualidade.

Em função da complexidade e quantidade de informações representadas na BC250 e dos elementos de qualidade, citados na fundamentação teórica, para avaliação da completude (omissão e comissão) foi escolhida a categoria de informação hidrografia e as classes: trecho drenagem, trecho massa d'água e massa d'água conforme descritas nas especificações técnicas e na ET-EDGV (Especificação Técnica para a Estruturação de Dados Geoespaciais Vetoriais). Para aferição da acurácia posicional foram consideradas todas as classes de feições presentes no conjunto de dados, principalmente as de fácil identificação no terreno como: vias, entroncamento de vias, passagens de nível e pontes.

---

<sup>27</sup> Disponível em <ftp://geoftp.ibge.gov.br/mapeamento\_sistematico/base\_vetorial\_continua\_escalas\_250mil>.

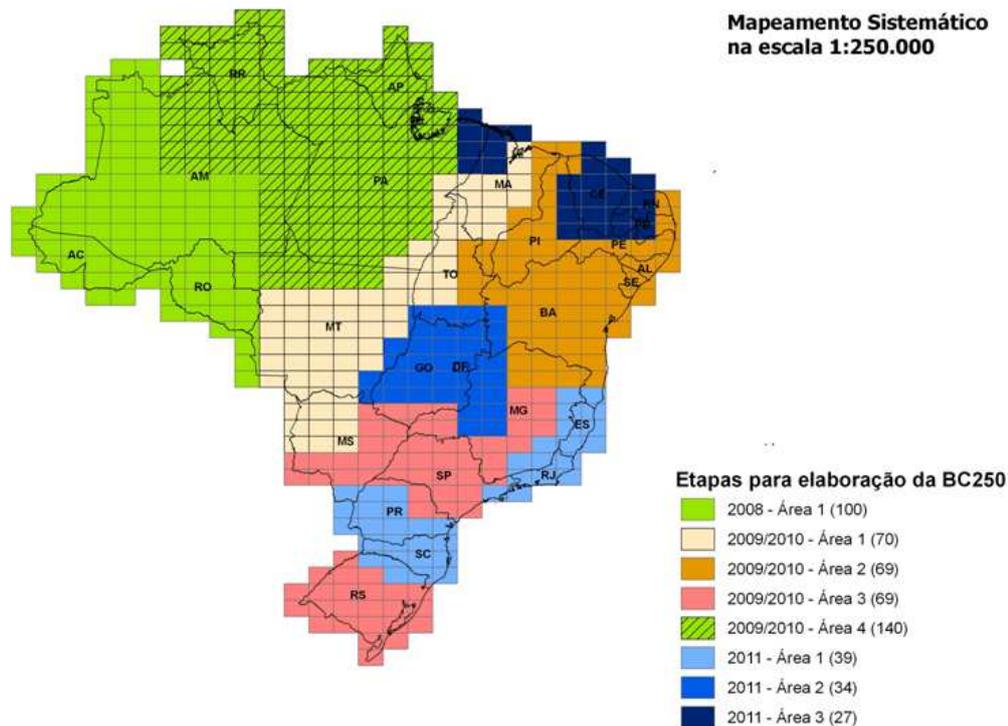


Figura 36 - A BC250 e seus blocos de trabalho antes da disponibilização na INDE

Fonte: REIS *et al.* (2011)

### 3.1.1. Dados e insumos utilizados

Além do produto cartográfico citado no tópico anterior, foram utilizados insumos de atualização da referida base cartográfica e outros materiais, listados a seguir:

- Imagens Landsat /TM (sensor *Thematic Mapper*), ortorretificadas, com resolução espacial de 30 metros, com ano de aquisição posterior a 2006.

- Pontos de controle referentes a pontos coletados em campo, em distintos projetos da Coordenação de Cartografia do IBGE (CCAR), presentes no banco de dados de pontos de apoio (Projeto PT-CON)<sup>28</sup>.

- Documentos cartográficos do mapeamento sistemático, acessados através do geoserviço do CIGEX (Centro de Imagens e Informações Geográficas do Exército)<sup>29</sup>.

- Rotinas (*scripts*) em ambiente R, elaboradas pelo Prof. Dr. Gerson Rodrigues dos Santos, no curso de extensão "Introdução a Geoestatística no R", ministrado na UFV em 2011, adaptadas ao presente estudo.

<sup>28</sup> Na presente data, segundo a área produtora do Projeto BC250, estes pontos de controle do Projeto PT-CON serão disponibilizados à sociedade.

<sup>29</sup> Este tipo de Geoserviço, denominado WMS (*Web Map Service*), é acessado por meio de ambiente SIG, permitindo a visualização de dados presentes em um servidor remoto. O Geoserviço do CIGEX está disponível no endereço <<http://mapas.mma.gov.br/cgi-bin/mapserv?map=/opt/www/html/webservices/baseraster.map>>

### 3.2. Equipamentos e softwares

Conforme mencionado, esta pesquisa foi realizada através da utilização de softwares livres da área de geoprocessamento e geoestatística a fim de facilitar sua reutilização e entendimento. As ferramentas utilizadas são listadas e descritas a seguir:

- Quatum GIS<sup>30</sup> (QGIS) é um software livre da área de geoprocessamento, vinculado oficialmente como um projeto do *Open Source Geospatial Foundation* (OSGeo), que suporta diversos formatos vetoriais, matriciais, banco de dados, além de suas funcionalidades. O ambiente de trabalho é amigável e permite realizar consultas espaciais e semânticas, bem como a edição de dados em formato *shapefile*. Está em constante desenvolvimento e, atualmente encontra-se em sua versão 1.8;
- gvSIG<sup>31</sup> é similar ao Quantum GIS, e faz parte do mesmo consórcio OSGeo. Atualmente encontra-se em sua versão 1.11.0 e foi utilizado para complementar a funcionalidade de georreferenciamento do QGIS;
- O ambiente R, ou simplesmente R<sup>32</sup> é um software livre com linguagem orientada a objetos com o qual é possível manipular e analisar dados através de rotinas (*scripts*) presentes em pacotes específicos, relativos principalmente a análises estatísticas. Além disto, com o R é possível realizar análises geoestatísticas e manipular estruturas de dados vetoriais e matriciais, como por exemplo, arquivos *shapefiles* e imagens Geotiff (PETERNELLI e MELLO, 2012).
- TinnR<sup>33</sup> é um software livre editor e/ou processador de texto ASCII (*American Standard Code for Information Interchange*), cujo ambiente de trabalho possui interface gráfica e ambiente de desenvolvimento integrado, com o objetivo principal de facilitar o aprendizado e uso do R.
- *Google Earth* é um software proprietário da *Google* disponibilizado gratuitamente que permite a visualização de imagens orbitais de distintas fontes, inclusive num modelo tridimensional da Terra. Esta ferramenta foi utilizada para complementar o ambiente QGIS, durante a inspeção de qualidade do conjunto de dados, com o objetivo de auxiliar na interpretação de acidentes naturais e antrópicos do mundo real, além da visualização e identificação da localização

---

<sup>30</sup> Mais informações em <<http://www.qgis.org>>

<sup>31</sup> Mais informações em <<http://www.gvsig.org>>

<sup>32</sup> Mais informações em <<http://www.rproject.org>>

<sup>33</sup> Mais informações em <<http://ngbcgib.uesc.br/lec/software/desenvolvidos/editores/tinn-r/pt>>

adequada dos pontos de apoio de campo, representados nas imagens Landsat/TM.

- O software livre *LibreOffice Calc*<sup>34</sup> foi utilizado para auxiliar nos cálculos e preparo de tabelas, tanto nas etapas de avaliação como na análise espacial da qualidade do conjunto de dados.

### 3.3. Metodologia

Este tópico sintetiza os conceitos e normas apresentados na fundamentação teórica, transformando-os em ferramentas práticas de inspeção, análise, representação e descrição da qualidade de um conjunto de dados espaciais em estudo. O fluxograma ilustrado na Figura 37, apresenta a visão geral da metodologia de avaliação da qualidade de produtos cartográficos, baseada nas normas ISO 19113:2002, ISO 19114:2003 e ISO 19115:2003, com ênfase na análise espacial dos resultados da inspeção por amostragem, através da Geoestatística e posterior relato desta aferição nos metadados. Ressalta-se que, os procedimentos e rotinas complementares são detalhadas nesta dissertação.

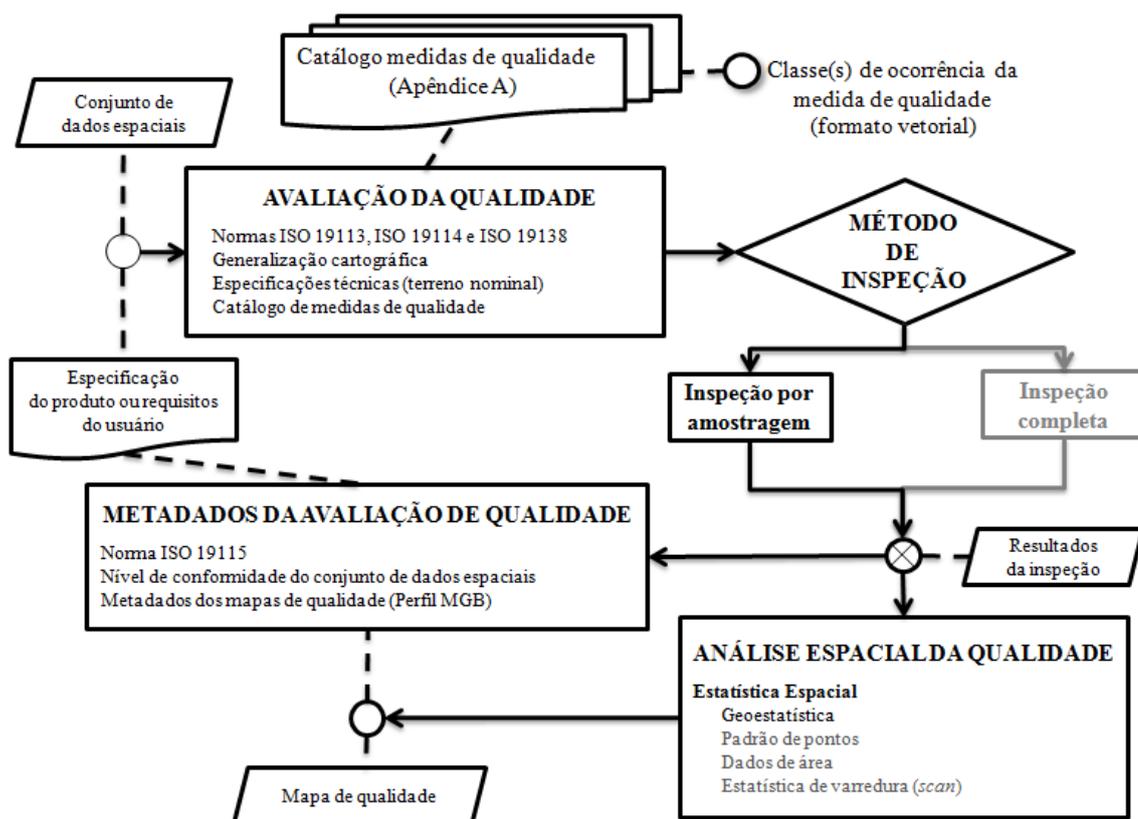


Figura 37 - Fluxograma para avaliação da qualidade de dados geoespaciais<sup>35</sup>

<sup>34</sup> Mais informações em <<http://pt-br.libreoffice.org>>

<sup>35</sup> Nesta pesquisa não foram abordados o método de inspeção completa e os seguintes tópicos de Estatística Espacial: padrão de pontos; dados de área; e estatística de varredura (*scan*).

Para facilitar a leitura dos fluxograma apresentados nesta dissertação, a Figura 38 apresenta uma legenda com a notação adotada<sup>36</sup>.

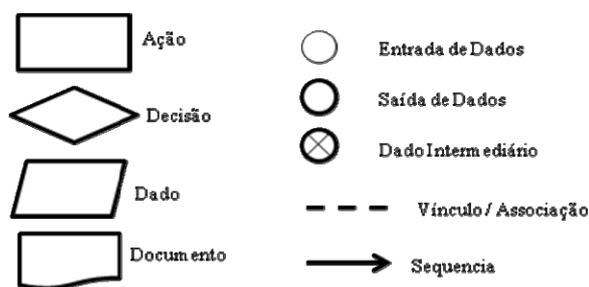


Figura 38 - Legenda dos símbolos usados nos fluxogramas da dissertação

O processo metodológico foi baseado na aplicação da norma ISO 19114:2003 relativa a qualidade de dados espaciais e a ISO 19115:2003 referente aos metadados geoespaciais. Além das orientações e recomendações das normas ISO citadas, foi incluído a análise espacial do resultado obtido no processo de inspeção de qualidade, seja esta completa ou por amostragem, a fim de fornecer uma representatividade visual a esta avaliação, a princípio descritiva e pontual, para auxiliar na identificação de possíveis regionalizações da qualidade causadas por prováveis erros sistemáticos.

No final do ciclo de avaliação, análise e representação espacial da qualidade do produto cartográfico existe uma seção dedicada a informar ao usuário do conjunto de dados espaciais sobre os procedimentos e resultados desta qualificação, os metadados.

### 3.3.1. Avaliação da qualidade

Num primeiro momento, recomenda-se uma análise preliminar do conjunto de dados espaciais a avaliar e um estudo das respectivas especificações técnicas, antes de iniciar a avaliação da qualidade em si. O objetivo deste estudo inicial é identificar o elemento de qualidade a ser avaliado, conforme Tabela 4, e o âmbito desta inspeção de qualidade, conforme orienta a norma ISO 19114:2003. Esta etapa é essencial para identificar/elaborar a medida de qualidade a ser executada. Os procedimentos, documentos, dados de entrada e saída, são descritos a seguir.

#### 3.3.1.1. Conjunto de dados espaciais

O conjunto de dados espaciais representa, no caso desta dissertação, o produto cartográfico ao qual se deseja avaliar a qualidade. Entretanto, pode ser representado por um pequeno grupo de dados, limitado por uma extensão espacial ou tipos de feições, alocados em um conjunto de dados maior. Teoricamente, o mesmo pode ser tão pequeno quanto uma simples feição ou um atributo de feição, contido num conjunto de

<sup>36</sup> Baseado e adaptado da Notação para Modelagem de Processos de Negócio (BPMN - *Business Process Modeling Notation*), mais informações em <<http://www.bpmn.org>>



### ***Preparo de classe de ocorrências e de amostras das medidas de qualidade***

As classes de ocorrências são vinculadas aos objetivo das medidas de qualidade, com a função de discriminar cada ocorrência identificada no conjunto de dados espaciais avaliado, facilitar sua localização e posterior contagem das mesmas, auxiliando nas conclusões dos resultados da inspeção e na sua análise espacial.

A classe de ocorrências pode ser uma réplica, inicialmente sem registros, da classe de feições em aferição ou uma classe auxiliar onde é representada a ocorrência segundo uma determinada variável demandada pela medida de qualidade correspondente. A Figura 40 mostra o fluxograma deste procedimento e as etapas são descritas a seguir:

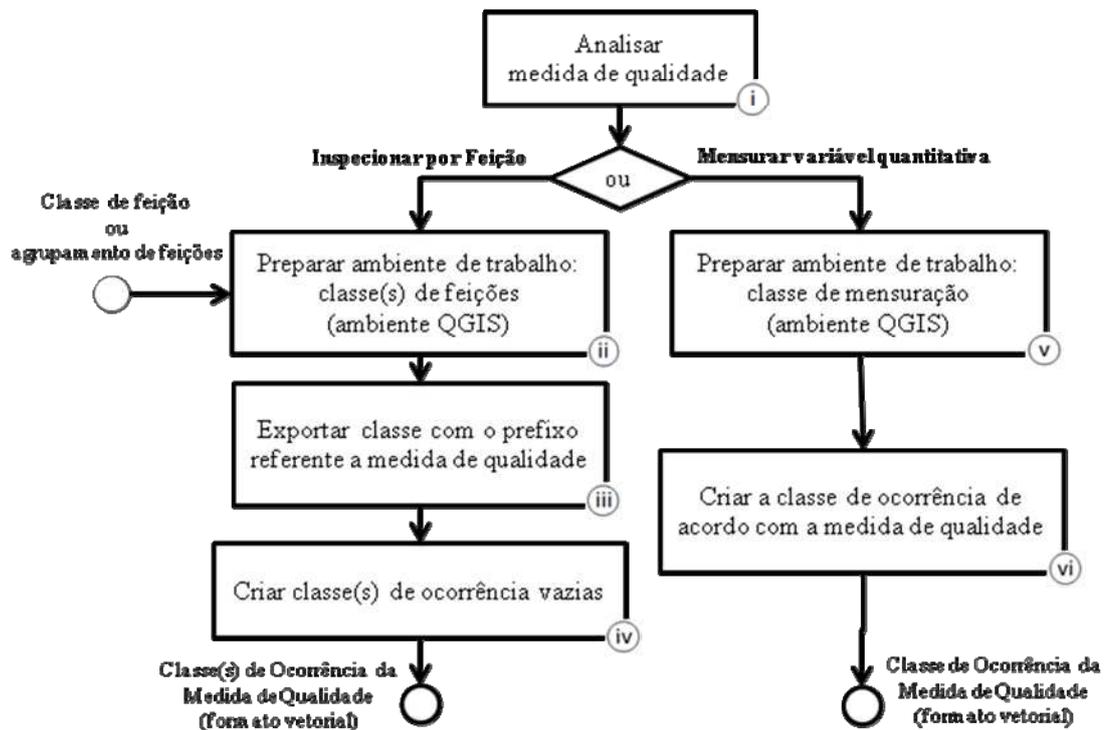


Figura 40 - Procedimentos para elaboração de classe(s) de ocorrência

#### i) Analisar medida de qualidade

Inicialmente deve ser feito um estudo da finalidade da medida de qualidade. A classe de ocorrência visa auxiliar na identificação, localização, contagem e/ou cálculo da variável em estudo. Logo, de acordo com a proposta de avaliação, cada ocorrência da medida de qualidade pode ser observadas através da própria feição do conjunto de dados espaciais ou por meio de uma feição de auxílio. Na presente dissertação foram criadas classes de ocorrência tanto baseadas em classes existentes, no caso da completez, como auxiliares para medir o erro planimétrico, no caso da acurácia posicional absoluta.

Caso a avaliação seja referente a uma determinada classe ou a um conjunto de classes específicos, os passos "ii", "iii" e "iv" devem ser executados. Caso a avaliação seja referente a uma determinada variável a determinar ou mensurar, os passos "v" e "vi" devem ser executados. As ocorrências são propriedades que devem estar adequadas a sua representação no conjunto de dados espaciais em estudo.

ii) Preparar ambiente de trabalho: classe(s) de feições (ambiente QGIS)

No ambiente Quantum GIS, iniciar um novo projeto e verificar se o sistema de coordenadas cartográficas e geodésicas são compatíveis com a proposta do conjunto de dados espaciais a avaliar, no caso da BC250 sistema de coordenadas geográficas e referencial geodésico SIRGAS2000. Adicionar as camadas de feições referentes a medida de qualidade em avaliação. No caso desta dissertação, as classes de feições carregadas são relativas a análise da completude, tanto para omissão quanto comissão, dos elementos da categoria de informação hidrografia: trecho de drenagem, trecho massa d'água e massa d'água.

iii) Exportar a classe(s) de feições com o prefixo referente a medida de qualidade

Cada classe carregada foi exportada para uma pasta específica onde foi incluído o código de identificação da medida de qualidade na denominação do shapefile; Por exemplo: "*M10101\_HID\_MassaDagua.shp*" e "*M10201\_HID\_Trecho\_Drenagem.shp*", referentes aos casos de comissão e omissão das feições trecho massa d'água e trecho de drenagem.

iv) Criar classe(s) de ocorrência vazias

As classes de ocorrência exportadas foram então editadas com o objetivo de apagar os registros originais e deixar o *shapefile* vazio e pronto para receber as ocorrências relativas a medida de qualidade.

Na presente dissertação, no caso da completude, foram criadas 3 (três) classes de ocorrência, relativas ao agrupamento de feições avaliadas: trecho de drenagem, trecho massa d'água e massa d'água, tanto para os casos de omissão quanto comissão, conforme Figura 41.

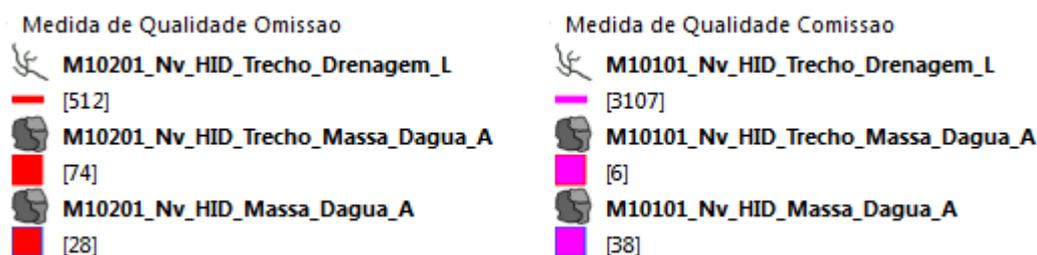


Figura 41 - Exemplo de classes de ocorrências criadas para completude.

iv) Preparar ambiente de trabalho: classe de mensuração

No ambiente Quantum GIS, iniciar um novo projeto e verificar se o sistema de coordenadas cartográficas e geodésicas são compatíveis com a proposta do conjunto de dados espaciais a avaliar, no caso da BC250 sistema de coordenadas geográficas e referencial geodésico SIRGAS2000.

vi) Criar a classe de ocorrência de acordo com a medida de qualidade

Uma classe de ocorrência com atributo(s) adequado(s) deve ser criada, conforme a proposta de avaliação da medida de qualidade.

Na presente dissertação, no caso da acurácia posicional absoluta foi criada uma feição com geometria tipo linha para medir o erro planimétrico entre o ponto representado na base cartográfica e seu homólogo coletado por GPS, considerado como verdadeiro no terreno. A criação de classes de ocorrências para auxiliar na mensuração de uma variável requisitada na medida de qualidade será subjetiva, mas deverá atender ao propósito da avaliação. Deste forma, foi criada a classe de ocorrência "*M30101\_ErroPlanimetrico.shp*" com o atributo "*ErroLinear*", conforme Figura 42.

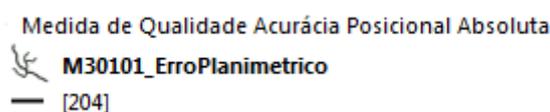


Figura 42 - Exemplo de classes de ocorrências criadas para acurácia posicional.

### 3.3.2. Método de inspeção

Nesta dissertação, conforme mencionado, foram avaliados dois elementos de qualidade: a completude (omissão e comissão) e a acurácia posicional. Ressalta-se que, para ambos os elementos de qualidade, o método de inspeção selecionado foi o por amostragem. O tópico "métodos de inspeção por amostragem" apresenta a visão geral dos procedimentos utilizados nesta pesquisa para realizar a inspeção por amostras. O tópico "método de inspeção completa" foi descrito, entretanto não foi realizado nesta pesquisa. Nas seções seguintes, os procedimentos são detalhados conforme o elemento de qualidade avaliado: completude e acurácia posicional.

#### *Métodos de inspeção por amostragem*

Neste método de inspeção por amostragem ocorre a avaliação parcial do conjunto de dados, baseado numa seleção amostral. No capítulo referente a amostragem, além dos conceitos em si, foram apresentados distintos métodos e seleção de amostras onde um número suficiente de itens da população a testar é necessário, a fim de atingir um resultado de qualidade. O fluxograma apresentado na Figura 43 sintetiza os

procedimentos de inspeção por amostragem de um conjunto de dados, baseado na norma ISO 19114:2003.

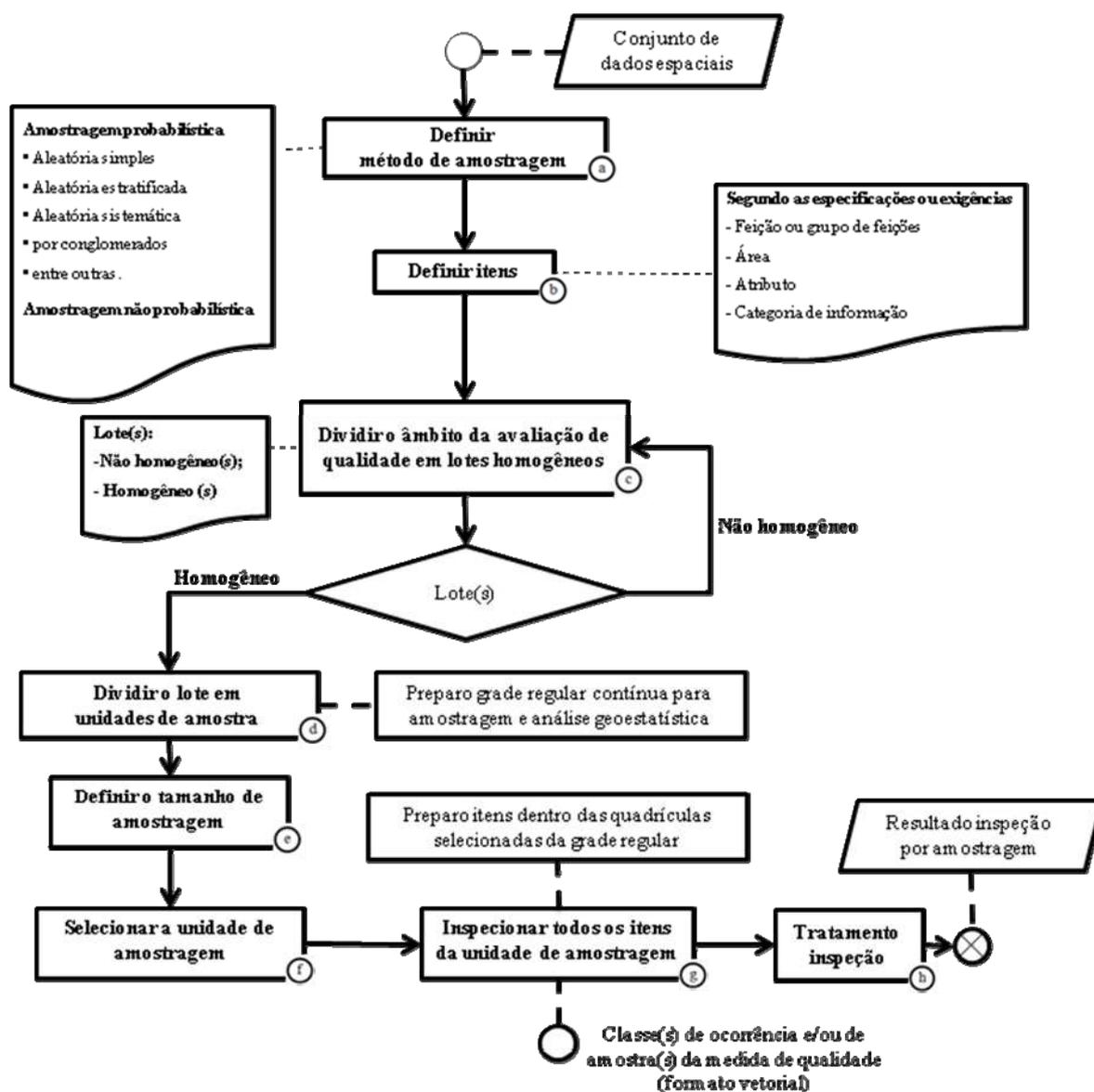


Figura 43 - Procedimentos para inspeção por amostragem

Fonte: baseado na norma ISO 19114:2003

Seguindo o fluxograma de inspeção por amostragem proposto nesta dissertação (Figura 43), foram realizadas as ações descritas e agrupadas nos tópicos referentes a inspeção do elemento de qualidade avaliado, no caso desta pesquisa: completude e acurácia posicional.

### *Método de inspeção completa*

O método de inspeção completa não foi aplicado nesta dissertação, no entanto, é apresentado um fluxograma de inspeção completa, a fim de ilustrar tal método, para casos que necessitem deste tipo de inspeção.

O método de inspeção completa requer testar todos os itens da população especificada no âmbito de qualidade de dados.

A Figura 44 mostra os passos para realizar a inspeção completa no conjunto de dados espaciais.

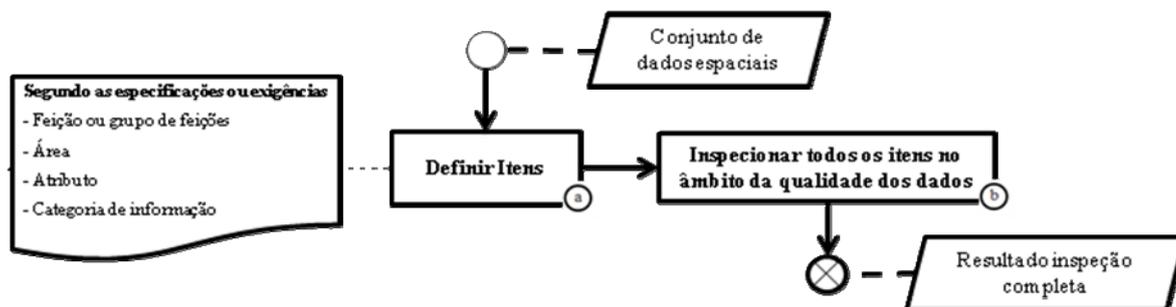


Figura 44 - Procedimentos para inspeção completa

Fonte: baseado na norma ISO 19114:2003

Os tópicos da inspeção completa são descritos a seguir, conforme a norma ISO 19114:2003:

#### **a) Definir itens**

Um item é a unidade mínima a ser inspecionada, o mesmo pode ser uma feição ou um grupo de feições, uma área de interesse, um atributo de feição, uma categoria de informação ou um relacionamento.

#### **b) Inspeccionar itens no âmbito da qualidade de dados**

Todos os itens presentes no âmbito da qualidade de dados devem ser inspecionados.

A inspeção completa é apropriada a pequenas populações ou em testes que podem ser realizados por métodos automatizados, como por exemplo, a validação topológica, que pertence ao elemento de qualidade de consistência lógica.

O tópico referente ao "preparo de itens dentro da grade regular contínua" pode ser utilizado na inspeção completa, a fim criar um indicador de qualidade, auxiliar na contagem e no tratamento posterior dos itens conformes (ou não conformes) e identificar erros sistemáticos na produção cartográfica.

O resultado da inspeção completa pode ser refletido nas classe(s) de ocorrência, correspondente(s) à medida de qualidade executada, criada conforme o tópico "preparo de classes de ocorrência e de amostras das medidas de qualidade".

### **3.3.2.1. Método de inspeção para completude**

Conforme mencionado foi realizado um estudo preliminar das especificações técnicas da BC250, a partir da qual foi identificado o elemento de qualidade a avaliar, no caso a completude, e o agrupamento de feições de hidrografia, no caso as classes: trecho de drenagem, trecho massa d'água e massa d'água, de acordo com as especificações técnicas do produto cartográfico e a ET-EDGV (CONCAR, 2008). Em relação ao âmbito da inspeção, considerando o propósito desta dissertação de auxiliar na avaliação da qualidade de dados da INDE, foi escolhida toda a extensão territorial do Brasil, afim de produzir um mapa de qualidade global e não apenas regional. A aplicação da metodologia para completude é similar em termos de comissão e omissão, com alteração apenas nos resultados apresentados e na análise espacial em si.

#### ***Métodos de inspeção por amostragem: comissão e omissão***

Seguindo o fluxo proposto nesta dissertação, mostrado na Figura 43, as ações foram definidas e realizadas conforme os passos a seguir.

##### ***a) Definir o método de amostragem***

Seleção aleatória sistemática, conforme recomendado nesta dissertação (YAMAMOTO e LANDIM, 2013).

##### ***b) Definir itens***

Os itens são um grupo de feições da categoria de informação hidrografia, composto pelas classes: trecho de drenagem, trecho massa d'água e massa d'água.

##### ***c) Lotes homogêneos***

Todo conjunto de dados a ser avaliado é homogêneo, apesar de estar em lotes de produção distintos, os mesmo foram produzidos segundo as mesmas especificações técnicas, praticamente num mesmo intervalo de tempo e ambiente de hardware e software de trabalho das empresas contratadas. Segundo a norma ISO 19114:2003, a homogeneidade está vinculada às seguintes condições: os dados de mesma origem de produção têm quase a mesma qualidade; sistemas de produção (hardware, software, habilidade do operador) são essencialmente os mesmos; e outros fatores que podem afetar a probabilidade de ocorrência de não conformes, como a complexidade e a densidade de feições, são essencialmente os mesmos. Um menor tamanho de lote é desejável quando há uma correlação espacial significativa na ocorrência de não conformidade.

#### ***d) Dividir o lote em unidade de amostragem***

A unidade de amostragem pode ser uma área geográfica existente ou alguma outra partição do universo de discussão de que a inspeção é conduzida. Quando a unidade de amostragem é uma área geográfica, as regras devem ser fornecidas para a inclusão de itens parcialmente localizados sobre as unidades de amostragem. Nesta dissertação as unidades de amostragem são quadrículas de aproximadamente 13km<sup>2</sup>, baseada na articulação do mapeamento sistemático 1:25.000. Em relação as articulações do mapeamento sistemático, numa folha 1:250.000 existem 6 (seis) folhas 1:100.000, compostas por 4 (quatro) folhas 1:50.000, que são compostas por 4 (quatro) folhas 1:25.000, ou seja, cada folha na escala 1:250.000 contempla 96 folhas na escala 1:25.000, conforme mostra a Figura 45.

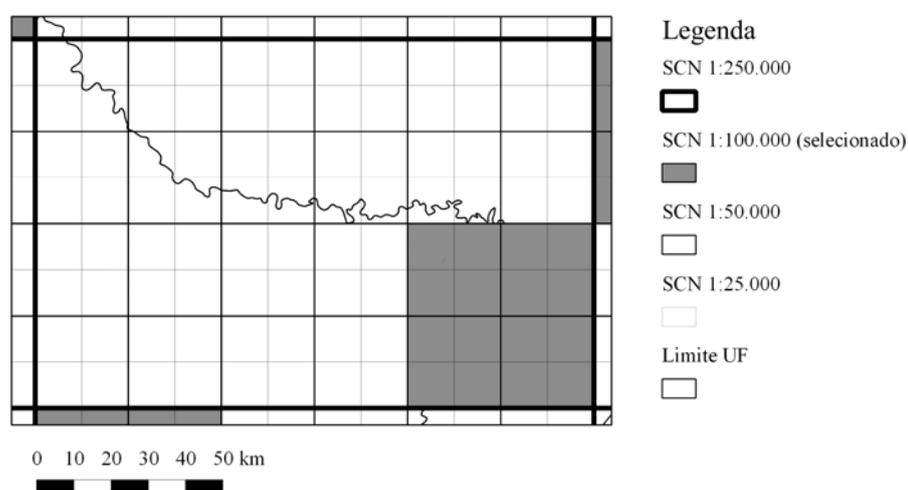


Figura 45 - Malha regular baseada nas articulações do mapeamento sistemático

As dimensões refletidas nas quadrículas 1:25.000 atendem ao propósito da dissertação, mas foram escolhidas de forma empírica. As dimensões poderiam ser menores, e podem ser objeto de estudo quanto a determinação de tamanhos mais adequados. . Na prática, deve-se considerar as dimensões em centímetros na escala a qual a medida de qualidade será executada. Por exemplo, uma quadrícula de 4x4cm corresponde a aproximadamente 10x10km na escala do 1:250.000, o que representa aproximadamente 1% da área total de uma folha desta escala. Desta maneira, na escala 1:1.000.000 esta dimensão de 4x4cm pode atender ao propósito da mesma medida de qualidade aplicada na escala 1:250.000, só que representando na realidade uma região maior de 40x40km, com correspondência de 1% aproximadamente da folha 1:1.000.000.

Ressalta-se que para a aplicação da análise geoestatística pressupõem-se que as amostras sejam contínuas, desta forma, as unidades de amostragem consideradas foram

áreas de avaliação, baseadas nos tamanhos de quadrícula do mapeamento sistemático 1:25.000, de aproximadamente 13km<sup>2</sup>, sendo considerada a inclusão de itens parcialmente presentes na unidade de amostragem. O tópico a seguir demonstra a criação de uma grade regular contínua para auxiliar na amostragem aleatória sistemática e análise geoestatística, tornando o espaço amostral contínuo, para posterior estimação das áreas não amostradas através de métodos geoestatísticos.

### ***Preparo da grade regular contínua para amostragem e análise geoestatística***

Na presente dissertação foi utilizada a articulação do mapeamento sistemático 1:25.000 para associar as unidades de amostragem a esta escala de trabalho. Esta foi uma escolha empírica, pelo fato das dimensões da quadrícula atenderem ao propósito da pesquisa. Entretanto, outras grades de amostragem podem ser criadas para regiões específicas, com áreas de abrangência maiores ou menores, de acordo com o fenômeno espacial em estudo. Desta forma, o presente tópico apresenta, através dos ambientes Quantum GIS e R, a criação de uma grade regular contínua, sobre uma região de interesse e a inserção do índice linha e coluna referente a posição relativa de cada quadrícula, conforme mostra o fluxograma da Figura 46. Esta grade regular contínua possibilita estratégias de amostragem sobre fenômenos espaciais descontínuos e análises geoestatísticas.

Neste tópico, por questões didáticas foi considerado um exemplo hipotético de criação de uma grade regular contínua para o estado de Minas Gerais, onde cada unidade de amostragem deve possuir uma dimensão aproximada de 111km<sup>2</sup>. Ressalta-se que esta grade, para fins geoestatísticos, deve ser criada no sistema de coordenadas geográficas ou em um adequado sistema de coordenadas projetado, a fim de evitar distorções e variações nas dimensões das quadrículas.

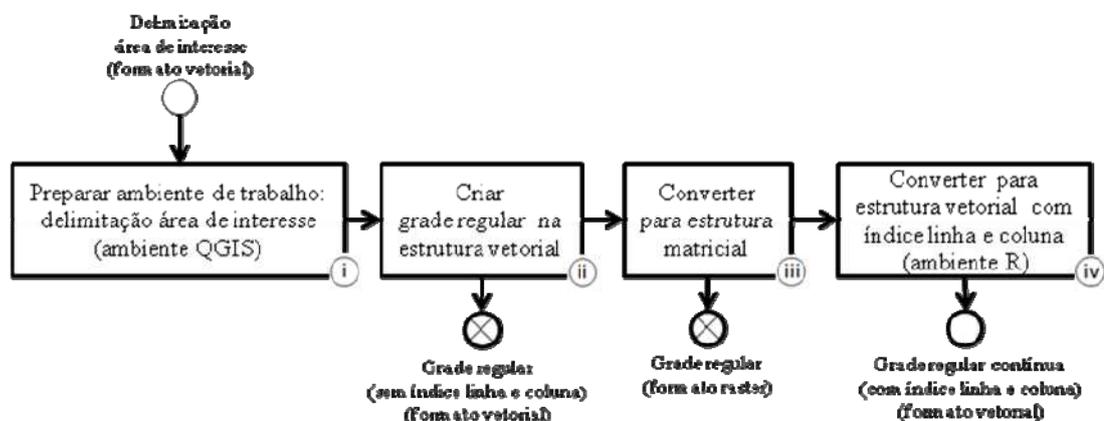


Figura 46 - Procedimentos pra elaboração da grade regular contínua

i) Preparar ambiente de trabalho: delimitação área de interesse (ambiente QGIS)

No ambiente de trabalho Quantum GIS deve ser carregada a área delimitando a região de interesse, neste exemplo foi utilizado o limite político administrativo do Estado de Minas Gerais, representado na BCIM (Base Cartográfica Contínua do Brasil, ao milionésimo)<sup>37</sup>, conforme Figura 47.



Figura 47 - Delimitação político-administrativa do Estado de Minas Gerais

ii) Criar grade regular na estrutura vetorial

Carregada a região de interesse no ambiente de trabalho e utilizar a ferramenta "*Grade vetorial*" do QGIS, presente no menu "*Vetor > Investigar*". A delimitação da grade deve ser ajustada a partir da extensão da camada de interesse, no caso Minas Gerais. A dimensão das quadrículas é definida no item "*distância entre linhas*"<sup>38</sup>, no caso 111km<sup>2</sup> por quadrícula, ou seja, foi definido o valor de 1(um) grau. Por último, selecionar a opção de "*lançar a grade como polígono*" e denominar o shapefile de saída como "*GradeRegular.shp*".

Na grade regular gerada, ressalta-se a presença de um atributo contendo uma numeração sequencial para cada célula, adequada a amostragem aleatória sistemática, com origem no canto superior esquerdo, conforme ilustra a Figura 48. Entretanto, as células da grade estão sem as informações de um referencial linha e coluna necessário para as análises geoestatísticas propostas nesta dissertação.

<sup>37</sup> Primeira base cartográfica contínua a compor a INDE do Brasil, disponível na seção de Geoserviços do Portal da INDE <<http://www.inde.gov.br>>

<sup>38</sup> Nesta dissertação, o sistema de coordenadas cartográfica do projeto QGIS é o geográfica, conforme o conjunto de dados espaciais da BC250, desta forma, a unidade de medida é graus, 1º grau equivale a aproximadamente 111km ao longo da linha do equador

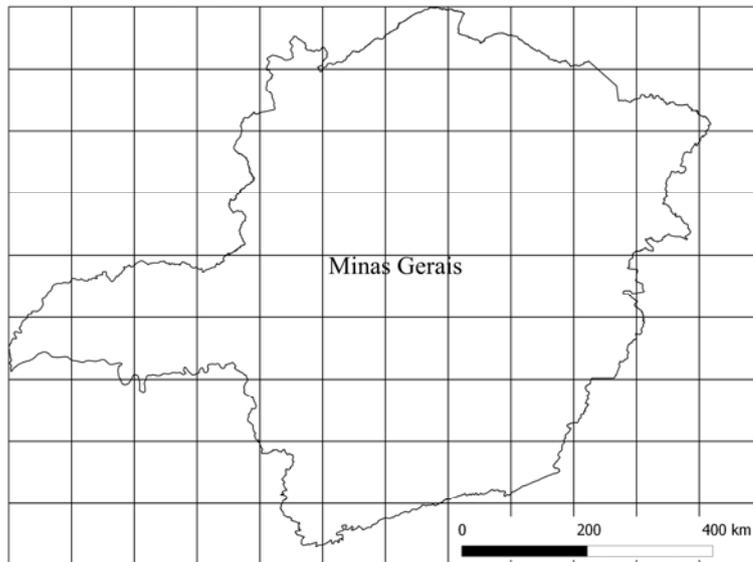


Figura 48 - Grade regular na estrutura vetorial sem índice linha e coluna

iii) Converter para estrutura matricial

Desta forma, de posse da grade regular gerada no passo anterior é iniciado o processo para inserir o índice linha e coluna. No ambiente QGIS exportar a grade regular para a estrutura matricial. Neste exemplo, foi atribuído a quantidade de 12 colunas e 9 linhas, conforme observado na Figura 48.

iv) Converter para estrutura vetorial com índice linha e coluna (ambiente R).

No ambiente R, através de funções presentes no pacote *rgdal*, foram inseridas as posições relativas de linha e coluna de cada quadrícula.

De posse da grade regular, gerada no passo anterior, a mesma é carregada no R, através da função *raster ()*. Em seguida é montada uma matriz a partir da leitura do arquivo raster por meio da função *"cbind ()"*, conforme expressão a seguir<sup>39</sup>. No resultado esta presente a posição coluna (*col*), linha (*row*) e o valor da célula, no caso a identificação da quadrícula.

```
"matriz<-cbind((colFromCell(r,1:ncell(r))), (ymax) - rowFromCell(r,1:ncell(r)),getValues(r))"
```

Na prática foi criada uma matriz com 3(três) dimensões. Antes de criar a grade regular com estas informações é necessário usar a função *"stack ()"* para separar estes atributos, feito isto esta estrutura matricial é convertida para polígono, através da função *"rasterToPolygon()"*, e posteriormente salva no formato shapefile, por meio da função *writeOGR()*.

<sup>39</sup> Para auxiliar na interpretação e georreferenciamento da krigagem, recomenda-se que a origem do sistema de coordenadas cartesiano, criado em linha e coluna, seja o canto inferior esquerdo da grade de amostragem. Desta forma, os valores das posições de coluna, eixo das abscissas (y), devem ser subtraídos do total de colunas mais 1, por isto a presença da variável "ymax" no script em R.

Desta forma, a "grade regular contínua para amostragem e análise geoestatística" foi criada. Esta contém informações da posição linha, coluna e a identificação das quadrículas ou células, representadas no shapefile "*GradeRegularContinua.shp*", destacando que a posição no eixo das abscissas ( $x$ ) refere-se a coluna e a posição no eixo das ordenadas ( $y$ ) refere-se a linha. Uma variável  $Z$ , correspondente ao resultado de uma medida de qualidade ou a um fenômeno espacial em estudo pode ser associada a cada quadrícula amostrada<sup>40</sup>, propiciando a análise do seu comportamento espacial através da Geoestatística.

Ressalta-se que, cada célula da grade regular contínua pode representar uma unidade de amostragem. Os itens vinculados as unidades de amostragem auxiliam no cálculo dos indicadores de qualidade estabelecidos nas medidas de qualidade.

#### ***e) Definir o tamanho da amostragem***

Após identificar o método e a unidade de amostragem, calcular o tamanho da amostragem com base no erro amostral tolerável desejado, conforme Equação 9. Nesta pesquisa, além do erro amostral tolerável, o tamanho da amostra levou em consideração o fato da BC250 ser composta pela integração das folhas 1:250.000. O objetivo foi tentar obter pelo menos uma amostra por folha 1:250.000, ou seja, em cada folha aproximadamente 1% de sua área seria analisada através da quadrícula de amostragem de 13 km<sup>2</sup>, conforme pode ser observado na Figura 45.

Entretanto, apenas o mapeamento 1:100.000 possui numeração sequencial inteira, representado no mapa índice (mi), passível de ser selecionada de forma aleatória e sistemática<sup>41</sup>. Desta forma, antes da seleção da quadrícula na escala 1:25.000, foram selecionadas as quadrículas das folhas 1:100.000 sistematicamente, cujo intervalo de mapa índice varia entre 1 e 3036, ou seja, uma população ( $N$ ) de 3036 quadrículas. Desta forma, para realizar a amostragem aleatória sistemática e o cálculo do erro amostral tolerável foram realizados os seguintes passos:

Inicialmente foi sorteado um número aleatório entre (1 e 3036) através do software livre *LibreOffice Calc*, através da função "*ALEATÓRIOENTRE()*" e o resultado foi o número 1364.

Partindo do número sorteado ( $m=1364$ ), os demais números de mapa índice (mi) a serem selecionados foram calculados, acima e abaixo do mi sorteado, através do software livre *LibreOffice Calc*, considerando um fator de sistematização ( $s$ ) igual a 5

---

<sup>40</sup> Conforme citado nesta dissertação, recomenda-se o uso da amostragem aleatória sistemática por apresentar uma melhor distribuição espacial (YAMAMOTO e LANDIM, 2013).

<sup>41</sup> Nas demais escalas do mapeamento sistemático o mapa índice é uma codificação sequencial alfanumérica, dificultando sua seleção aleatória sistemática.

(cinco), com o objetivo de selecionar um ponto de amostra por folha 1:250.000, em função da composição do mapeamento sistemático, como ilustra a Figura 45. Desta forma, foram selecionadas  $n = \left\lceil \frac{N}{s} \right\rceil$ , ou seja, 607 quadrículas na escala 1:100.000, visando selecionar posteriormente uma unidade de amostragem por folha 1:250.000.

Após o sorteio do número aleatório e o cálculo das demais amostras a serem selecionadas foi necessário criar uma expressão SQL (*Structured Query Language*), com o auxílio do software livre *LibreOffice Calc*, devido a quantidade de quadrículas selecionadas na escala 1:100.000, conforme ilustra parcialmente a Tabela 9, para uso em ambiente SIG. A expressão "(mi = '1364' OR mi = '1369' OR mi = '1374' .... OR mi = '14' OR mi = '9' OR mi = '4')", mostra parcialmente a consulta por atributo criada e utilizada no software Quantum GIS para seleção das folhas 1:100.000.

TABELA 9 - Montagem da expressão SQL para seleção aleatória sistemática em SIG

<b>Número aleatório sorteado "1364"</b>			
<b>MI acima</b>	<b>SQL (superiores)</b>	<b>MI abaixo</b>	<b>SQL(inferiores)</b>
1364	mi = '1364'	1364	mi = '1364'
1369	OR mi = '1369'	1359	OR mi = '1359'
1374	OR mi = '1374'	1354	OR mi = '1354'
....	.....	....	.....
Até 3036	OR mi = '3034'	Até 0	OR mi = '4'

Posteriormente à seleção aleatória sistemática, citada através da expressão SQL, foi realizada a seleção das unidades de amostragem nas dimensões da quadrícula desejada. Esta seleção foi realizada da seguinte forma:

#### ***Unidades de amostragem para completude***

Dentro de cada folha 1:100.000 selecionada anteriormente foi gerado um ponto aleatório, por meio do software livre Quantum GIS, através da função "*Pontos aleatórios*", presente no menu "*Vetor > Investigar*".

Em seguida foram selecionadas as unidades de amostragem de 13 km<sup>2</sup>, geradas no tópico relativo a grade regular contínua, que tocam estes pontos aleatórios. Alguns pontos aleatórios gerados sobre a articulação 1:100.000 caíram em locais extremos do país onde não há articulações 1:25.000, sendo que em 2 (duas) unidades de amostragem não ocorreram amostras, por isto a ausência de 36 unidades de amostragem. A Figura 49 ilustra a distribuição espacial das 571 unidades de amostragem selecionadas.



Figura 49 - Distribuição espacial das 571 unidades de amostragem para completeude

Assim, considerando o tamanho da amostra de 571 unidades e utilizando a Equação 9, o erro amostral tolerável calculado para completeude foi de 4,18%, conforme expressão:

$$E_0 = \sqrt{\frac{1}{571}} = 0,0418. \quad (31)$$

***f) Selecionar as unidades de amostragem***

Conforme descrito anteriormente, as unidades de amostragem foram selecionadas através de pontos aleatórios dentro de cada uma das folhas 1:100.000 sistematicamente selecionadas, totalizando 571 amostras a serem avaliadas, representando áreas de aproximadamente 13 km<sup>2</sup>, e correspondendo a quadrículas do mapeamento sistemático 1:25.000, que delimitaram o agrupamento de feições de hidrografia: trecho de drenagem, trecho massa d'água e massa d'água.

***g) Inspeccionar as unidades de amostragem***

A inspeção pode ser de itens conformes, não conformes ou uma determinada medida, mas em todos os casos é recomendado separar a característica avaliada numa classe de ocorrência específica relativa a medida de qualidade considerada, conforme descrito anteriormente. Este procedimento visa facilitar a identificação, localização, contagem, medição ou cálculo do parâmetro em análise. No caso de itens não conformes, o posterior tratamento das inconsistências fica facilitado. Caso a unidade amostra seja uma área de avaliação, é recomendada a seleção dos itens que interceptam as áreas selecionadas, conforme detalha o tópico "*preparo de itens dentro da grade regular contínua*".

### ***Preparo de itens dentro da grade regular contínua***

Nesta etapa é realizada a seleção dos itens definidos na medida de qualidade, no caso desta dissertação relativos a completude de feições de hidrografia. Esta metodologia é aplicada nos casos de inspeção de classes de feições dentro do conjunto de dados com a finalidade de facilitar os procedimentos de avaliação e contagem dos itens selecionados por unidade de amostra.

Desta forma, de posse das quadrículas da grade regular contínua para amostragem, criada conforme descrito anteriormente, é realizada uma consulta espacial para identificar as feições que tocam as unidades de amostragem selecionadas e posteriormente salvas num *shapefile* correspondente, criando classe(s) de amostras. O fluxograma da Figura 50 mostra os procedimentos descritos a seguir.

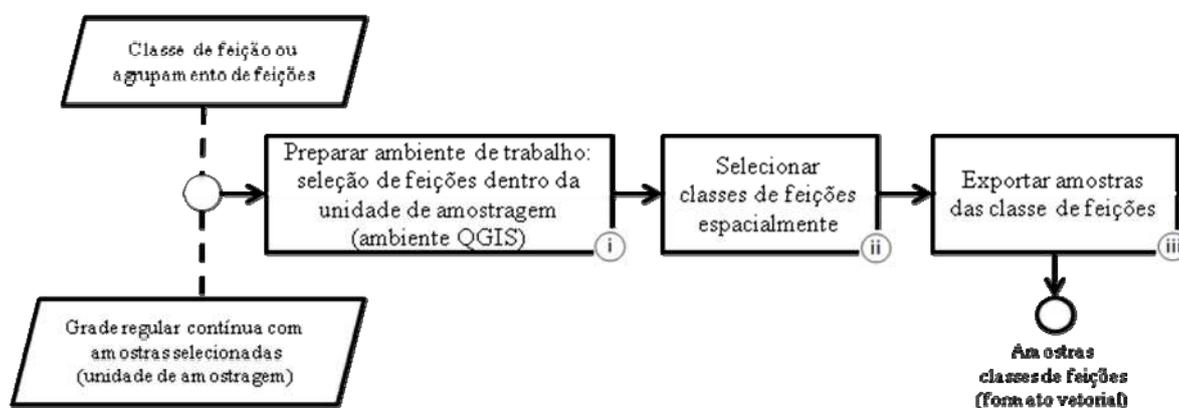


Figura 50 - Procedimentos para seleção de feições geográficas dentro das unidades de amostragem

#### **i) Preparar ambiente de trabalho: seleção de feições dentro da unidade de amostragem**

No ambiente Quantum GIS, adicionar as camadas de feições referentes à medida de qualidade em avaliação e a grade regular com as unidades de amostragem selecionada pelo método aleatório sistemático. No caso desta dissertação, as classes de feições carregadas são relativas a análise da completude, tanto para omissão quanto comissão, dos elementos da categoria de informação hidrografia: trecho de drenagem, trecho massa d'água e massa d'água.

#### **ii) Selecionar classes de feições espacialmente**

No ambiente SIG foi realizado uma consulta espacial, presente no menu "*Vetor > Consulta espacial*", onde classe a classe foram selecionadas as feições que interceptavam as unidades de amostragem. Nesta pesquisa, este procedimento foi realizado em cada um dos 11 (onze) blocos da BC250.

iii) Exportar as amostras da classes de feição

A cada nova seleção foi gerado um shapefile correspondente a classe consultada com o prefixo "Amostra", agrupadas em pastas segundo os blocos. Por exemplo: "*Amostra\_HID\_Massa\_Dagua.shp*". Relembrando que o objetivo é auxiliar na contagem de amostras e execução da inspeção da medida de qualidade.

Desta forma, as feições de hidrografia (trecho de drenagem, trecho massa d'água e massa d'água) que interceptam as unidades de amostragem foram selecionadas e separadas como amostras destas feições. Em relação as classes relativas, a ocorrência das medidas de qualidade "itens em excesso em hidrografia" e "itens ausentes em hidrografia" foram criadas classes de ocorrências, *shapefiles* vazios, para cada feição de hidrografia, segundo a medida de qualidade em estudo. Além deste preparo para inspeção, a unidade de amostragem foi representada e delimitada por quadrículas, articulações do mapeamento sistemático na escala 1:25.000. A Figura 51 ilustra uma unidade de amostragem avaliada, com as classes de ocorrências relativas as medidas de qualidade de comissão (*M10101*) e omissão (*M10201*); as amostras de feições de trechos de drenagem, trechos massa d'água e massa d'água; o mosaico Landsat no plano de fundo, onde foi constatado a ausência de alguns trechos de drenagem, os quais foram vetorizados na classe de ocorrência *M10201\_Nv\_HID\_Trecho\_Drenagem\_L*.

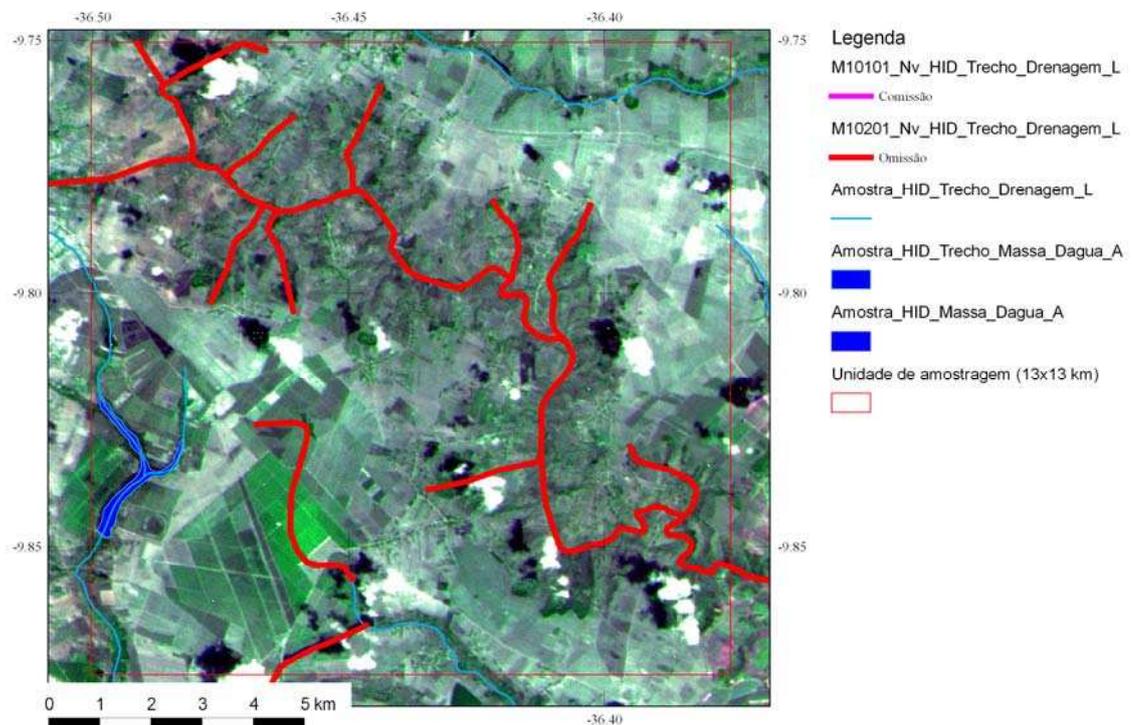


Figura 51 - Trecho de drenagem significativo ausente na BC250 (Rio Coruripe)

A Figura 52 mostra um trecho de drenagem onde o curso d'água verdadeiro não foi vetorizado, gerando um caso de comissão e outros dois de omissão.

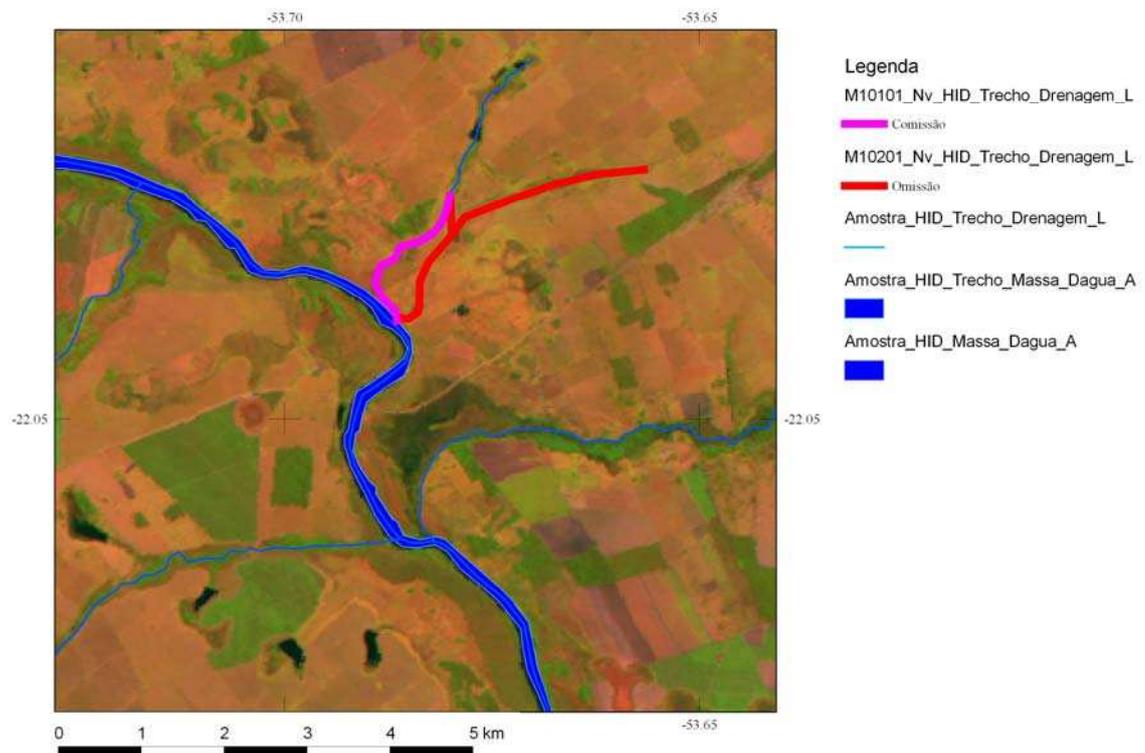


Figura 52 - Exemplos de comissão e omissão de um trecho de drenagem

A inspeção da completude (omissão e comissão) foi realizada levando em consideração as especificações técnicas, documentos cartográficos existentes da mesma escala ou maiores, acessados através do geoserviço do CIGEX<sup>42</sup>, mosaico de imagens Landsat, onde cada ocorrência de omissão ou comissão foi inserida na classe de ocorrência da medida de qualidade correspondente. Além destes insumos, foi utilizado, o *Google Earth* como ferramenta auxiliar para facilitar a visualização de feições, ausentes ou em excesso, no conjunto de dados. Para isto, as amostras selecionadas junto com a delimitação da unidade de amostragem foram convertidas para o formato kml (*Keyhole Markup Language*).

#### ***h) Tratamento da inspeção***

O resultado da inspeção por amostragem deve estar refletido na unidade de amostragem, entretanto para isto é necessário um tratamento a partir da(s) classe(s) de ocorrências da medida de qualidade executada e da contagem de amostras presentes em cada unidade de amostragem. Caso a unidade de amostragem seja uma área geográfica é recomendado o uso da ferramenta de "Junção espacial" (*Join spatial*) para realizar a contagem de ocorrências e o número de feições geográficas avaliadas por unidade de amostragem. Cálculos complementares na tabela de atributos referentes a unidade de amostragem podem ser necessários para atender a demandas da medida de qualidade.

<sup>42</sup> Disponível através do geoserviço:  
<http://mapas.mma.gov.br/cgi-bin/mapserv?map=/opt/www/html/webservices/baseraster.map>

Desta forma, para inserir este resultado na unidade de amostragem foram realizados procedimentos de junção de atributos espaciais, ou seja, utilizando o software Quantum GIS, através da ferramenta "*Unir atributos por localização*", presente em "*Vetor > Gerenciar dados*", conforme ilustra a Figura 53. Assim foi realizada a contagem de ocorrências e do número de amostras nas unidade de amostragem selecionadas.

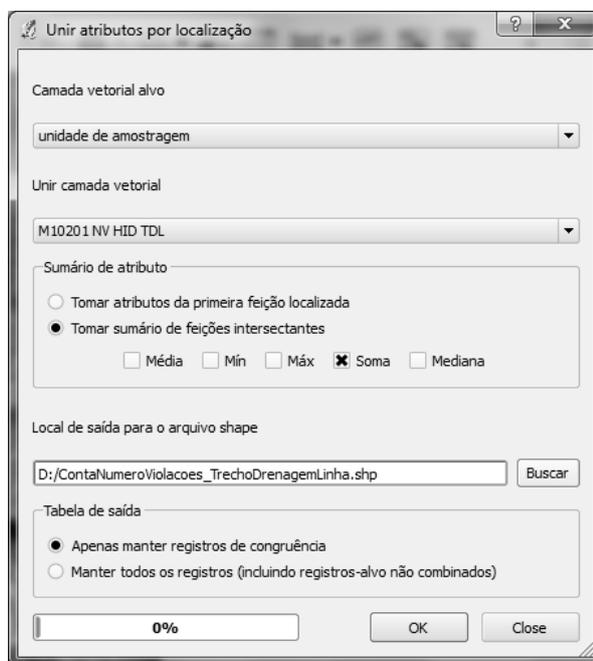


Figura 53 - Exemplo de contagem de ocorrências por unidade de amostragem

A cada *shapefile* criado foi adicionado um novo campo na tabela de atributos com uma denominação adequada, no qual foram inseridos os valores gerados pela ferramenta "*Unir atributo por localização*", através da função "*Calculadora de campo*" presente na própria tabela de atributos. Após renomear os campos de todos os *shapefiles* gerados, o procedimento de "*Unir atributos por localização*" foi realizado novamente para adicionar estes campos e seus valores a unidade de amostragem correspondente.

Assim, a avaliação da qualidade expressava um resultado quantitativo tabular, presente em cada unidade de amostragem. Em seguida, estes valores foram utilizados para calcular a taxa de comissão e taxa de omissão conforme Equações 4 e 5. Entretanto, para realizar o cálculo, por limitações do ambiente SIG, foi utilizado o software livre *LibreOffice Calc*, onde a tabela de atributos da unidade de amostragem de completude, no formato dbf (Arquivo dBase), foi editada com o objetivo de calcular as taxas de omissão e comissão. Após os cálculos, além de salvar a tabela no formato nativo do software livre *LibreOffice Calc*, a mesma foi salva no formato dbf com o mesmo nome do *shapefile*. Desta forma, o arquivo *shapefile* passou a conter as taxas de omissão e comissão, além dos atributos descritos anteriormente. A Figura 54 mostra os

atributos das unidades de amostragem de completude<sup>43</sup>, acrescidos das taxas de omissão e comissão calculadas, conforme descrição dos campos são descritos na Tabela 10.

LINHA	COLUNA	ID	TDL	TMD	MD	NvMD10101	NvTMD10101	NvMD10201	NvTMD10201	NvTDL10101	NvTDL10201	NvM10101	NvM10201	No	Nr	TxComissao	TxOmissao	ValorMax
0	137	128	13902	14	1	2	1	0	0	1	13	14	1	17	4	0.82353	0.05882	17
1	117	129	14113	4	0	2	2	0	1	0	2	0	4	6	3	0.66667	0.16667	6
2	280	138	15568	126	0	0	0	0	0	0	65	0	65	126	61	0.51587	0.00000	126
3	197	101	9683	55	0	0	0	0	0	0	28	0	28	55	27	0.50909	0.00000	55
4	268	34	2175	6	0	0	0	0	0	0	3	3	3	6	6	0.50000	0.50000	6
5	249	58	4571	77	19	0	0	0	0	0	47	6	47	96	55	0.48958	0.06250	96

Figura 54 - Tabela de atributos contendo os dados observados na inspeção da completude por unidade de amostragem

TABELA 10 - Descrição dos campos da unidade de amostragem de completude<sup>44</sup>

CAMPO	DESCRIÇÃO
LINHA	Posição relativa da unidade da amostragem verticalmente
COLUNA	Posição relativa da unidade de amostragem horizontalmente
ID	Codificação da unidade de amostragem de acordo com a grade regular criada
TDL	Número de elementos de amostras da classe "trecho drenagem linha"
TMD	Número de elementos de amostras da classe "trecho massa d'água"
MD	Número de elementos de amostras da classe "massa d'água"
NvTDL10101	Número de ocorrências de comissão de "trecho drenagem linha"
NvMD10101	Número de ocorrências de comissão de "massa d'água"
NvTDL10101	Número de ocorrências de comissão de "trecho massa d'água"
NvTDL10201	Número de ocorrências de omissão de "trecho drenagem linha"
NvMD10201	Número de ocorrências de omissão de "massa d'água"
NvTDL10201	Número de ocorrências de omissão de "trecho massa d'água"
Nv10101	Número de ocorrências de comissão na unidade de amostragem ( $N^+$ )
Nv10201	Número de ocorrências de omissão na unidade de amostragem ( $N^-$ )
No	Número de elementos da amostra calculada segundo a equação: $N^o = N^r + N^+ - N^-$
Nr	Número de ocorrências na realidade percebida, calculados através da equação: $N^r = N^o - N^+ + N^-$
TxComissao	Taxa de comissão calculada segundo a equação: $\tau^+ = \frac{N^+}{\max(N^o, N^r)}$
TxOmissao	Taxa de omissão calculada segundo a equação: $\tau^- = \frac{N^-}{\max(N^o, N^r)}$
ValorMax	Seleção do maior valor, este é o número de ocorrências na realidade percebida ( $N^r$ ) ou o número de elementos da amostra ( $N^o$ ) para utilizar no cálculo das taxas de omissão e comissão.

<sup>43</sup> As informações relativas a avaliação de qualidade de completude (omissão e comissão) foram sintetizadas e representadas num único *shapefile*, em função das unidades de amostragem serem idênticas para omissão e comissão.

<sup>44</sup> A nomenclatura da equação referente ao número de elementos na amostra foi alterada para facilitar o entendimento e associação da mesma com a tabela de atributos presente na unidade de amostragem referente a completude e não confundir com a denominação de população ( $N$ ).

Desta forma, os cálculos necessários para atender as medidas de qualidade avaliadas foram realizados e esta etapa foi concluída, atendendo as recomendações da norma ISO 19114:2003. Os resultados das medidas de qualidade em relação a itens em excesso e itens ausentes na BC250 de forma geral, sem considerar o comportamento espacial, são representados nas taxas de comissão e omissão, conforme Salgé (1995):

$$\tau^+ = \frac{N^+}{\max(N^o, N^r)} = \frac{3146}{24706} = 0.1273 = 12.73\% \text{ (taxa de comissão geral),}$$

$$\tau^- = \frac{N^-}{\max(N^o, N^r)} = \frac{614}{24706} = 0.0248 = 2.48\% \text{ (taxa de omissão geral).}$$

### **3.3.2.2. Método de inspeção para acurácia posicional**

Para avaliação da acurácia posicional absoluta da BC250 foi definida a inspeção por amostragem aleatória sistemática, devido ao propósito da medida de qualidade e a quantidade de pontos de apoio disponíveis. Entretanto, apesar do número significativo de pontos de apoio de campo em algumas áreas de inspeção, não foi possível realizar a correlação entre os pontos de campo e o seu correspondente no conjunto de dados espaciais, pelo fato do ponto de apoio de campo estar localizado num local não representável na escala 1:250.000, impossibilitando identificar seu homólogo.

#### ***Método de inspeção por amostragem: acurácia posicional absoluta***

Seguindo o fluxo proposto nesta dissertação, mostrado na Figura 43, as ações foram definidas e realizadas conforme os passos a seguir.

#### ***a) Definir o método de amostragem***

Seleção aleatória sistemática, conforme recomendado nesta dissertação (YAMAMOTO e LANDIM, 2013).

#### ***b) Definir itens***

O item a ser avaliado é o erro planimétrico que representa a discrepância posicional, medida linearmente, entre o ponto de apoio coletado em campo, considerado como “verdadeiro”, e o seu homólogo no conjunto de dados espaciais em estudo.

#### ***c) Lotes homogêneos***

Todo conjunto de dados foi considerado homogêneo, pois foram produzidos com base nas especificações técnicas do produto obedecendo aos ambientes de trabalho descritos em IBGE (2012).

**d) Dividir o lote em unidade de amostragem**

A distância entre o ponto de apoio de campo e seu homólogo identificado na BC250 foi associada a uma única unidade de amostragem, quadrícula de aproximadamente 13km<sup>2</sup>.

**e) Definir o tamanho da amostra**

Da mesma forma que na completude, o objetivo foi tentar medir ao menos um ponto por folha (na escala 1:250.000), dentro da área de busca delimitada pela articulação 1:100.000 selecionada. Entretanto, apesar de existirem aproximadamente 9500 pontos de apoio de campo, de diversos projetos realizados pelo IBGE, em diversos locais não foi possível identificar o homólogo na BC250 nas áreas selecionadas para inspeção. Além disto, existe uma ausência de pontos de controle na região da Amazônia Legal. Desta forma foram inspecionados ao todo 204 pontos e a Figura 55 ilustra a distribuição espacial das articulações 1:100.000 selecionadas.



Figura 55 - Distribuição espacial das áreas para coleta de pontos amostrais referentes à avaliação da acurácia posicional absoluta

Assim, considerando o tamanho da amostra de 204 unidades de amostragem representadas nas áreas de 13 km<sup>2</sup>, o erro amostral tolerável ( $E_0$ ) calculado foi de 7,01%, conforme expressão:

$$E_0 = \sqrt{\frac{1}{204}} = 0,0701.$$

#### f) Selecionar as unidades de amostragem

A articulação 1:100.000 delimitou as áreas onde poderiam ser correlacionados os pontos de apoio coletados em campo e seus homólogos na BC250. A Figura 56 apresenta em detalhe uma determinada região dentro da área em estudo. Conforme mencionado, não foi possível identificar pontos em todas as quadrículas 1:100.000 e posteriormente associar o erro planimétrico à articulação 1:25.000.

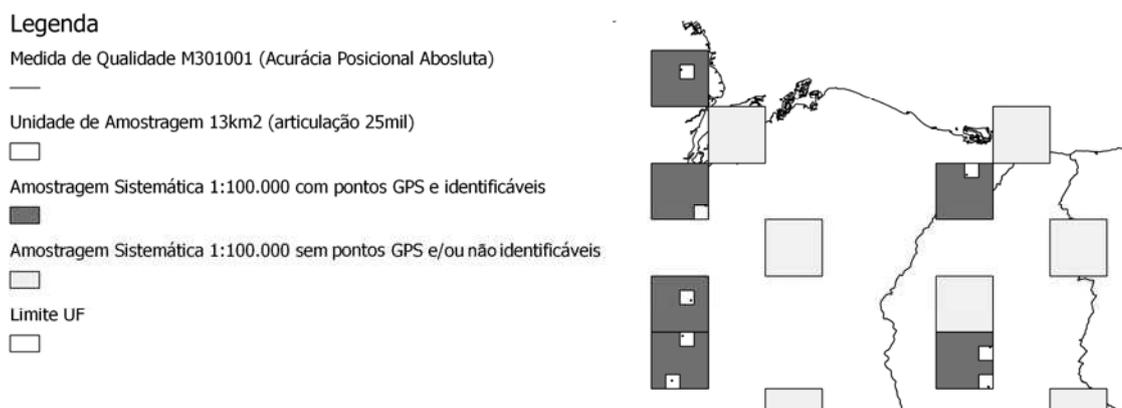


Figura 56 - Detalhe da amostragem sistemática para acurácia posicional absoluta

#### g) Inspeccionar as unidades de amostragem

Antes de iniciar a inspeção das unidades de amostragem, foram realizados os procedimentos do tópico "preparo de classe de ocorrências e de amostras das medidas de qualidade". Desta forma, foi criada a classe de ocorrência, um arquivo *shapefile*, denominado "*M30101\_Acuracia\_Absoluta*", relativa a medida de qualidade "incertezas posicionais", com o tipo de geometria linha e um campo para o cálculo do erro planimétrico ao final da inspeção.

O objetivo foi encontrar pontos de apoio de campo, considerados como referência, dentro da área delimitada por cada articulação 1:100.000 selecionada de forma aleatória sistemática aleatória. Esta região de varredura corresponde a aproximadamente 2700 km<sup>2</sup> e, entre cada posição identificada na BC250 com o respectivo ponto de apoio foi traçada uma linha, correspondente ao erro planimétrico, com o cuidado de coletar uma amostra dentro de cada unidade de amostragem, no caso áreas de 13 km<sup>2</sup> correspondentes a articulação 1:25.000 do mapeamento sistemático. Para cada ponto de apoio de campo localizado dentro da região de varredura, buscou-se identificar o seu homólogo na BC250. A Figura 57 ilustra uma unidade de amostragem avaliada.

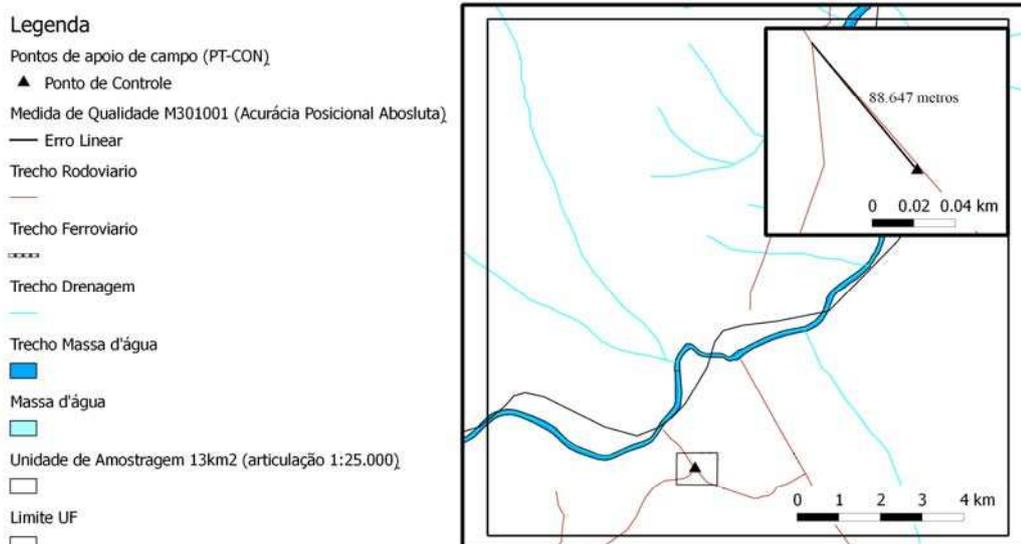


Figura 57 - Detalhe de uma situação de aferição da acurácia posicional absoluta

A inspeção da acurácia posicional absoluta foi realizada levando em consideração as feições presentes no conjunto de dados, pontos de controle coletados em campo e/ou documentos cartográficos de maior escala. Nesta pesquisa, além dos pontos de apoio de campo, foram utilizadas as imagens Landsat, documentos cartográficos existentes, acessados através do geoserviço do CIGEX, e o software *Google Earth* como ferramenta auxiliar para facilitar a identificação dos pontos de controle no terreno.

#### ***h) Tratamento da inspeção***

Ao final do processo de inspeção por amostragem, o resultado deve estar refletido na unidade de amostragem. No caso da acurácia posicional absoluta o resultado é a medida da extensão de cada erro planimétrico avaliado, presente na classe de ocorrência desta medida de qualidade. Para o cálculo desta extensão foi utilizada a Projeção Policônica, com os parâmetros descritos na respectiva medida de qualidade, baseado em IBGE (2010). Em seguida, cada erro planimétrico calculado foi associado a unidade de amostragem de 13km<sup>2</sup> sobre a qual o mesmo estava contido. Desta forma, para inserir este resultado na unidade de amostragem foi realizado o procedimento de junção de atributos espacialmente, utilizando o software Quantum GIS, através da ferramenta "*Unir atributos por localização*", presente em "*Vetor > Gerenciar dados*", e realizada a união dos atributos dos arquivos *shapefile* "M30101" à unidade de amostragem de 13km<sup>2</sup> que interceptava as discrepâncias lineares medidas, similar ao representado na Figura 53 referente a completude.

### **3.3.3. Análise espacial da qualidade**

Após a inspeção do conjunto de dados, por amostragem ou de forma completa, é relevante realizar uma análise espacial do resultado da medida de qualidade executada, a fim de gerar uma representatividade espacial da referida avaliação da qualidade.

O método de inspeção, completa ou por amostragem, determina a forma de análise espacial da avaliação da qualidade. Nesta dissertação, o foco foi o uso de métodos geoestatísticos para identificar a correlação espacial ou não do resultado da medida de qualidade. Na inspeção completa, outros métodos da Estatística Espacial podem ser aplicados, mas não foram abordados neste documento.

#### **3.3.3.1. Análise geoestatística**

A análise geoestatística pressupõe continuidade espacial do campo amostral. Na presente pesquisa os indicadores de qualidade avaliados são representados nas células da "grade regular contínua para amostragem e análise geoestatística". Cada unidade de amostragem selecionada é utilizada para prever valores em células não amostradas, tanto na aplicação de medidas de qualidade de completude como no caso da acurácia posicional. As células da grade regular contínua contemplam indicadores de qualidade, de acordo com a proposta da medida de qualidade.

Os métodos geoestatísticos devem ser utilizados para identificar o comportamento espacial da inspeção por amostragem da variável de qualidade em estudo. O objetivo é indicar o uso da Estatística Espacial ou da Estatística Clássica nesta avaliação, em razão da possibilidade de existir dependência ou independência espacial dos dados observados. Durante a própria etapa de inspeção da qualidade do conjunto de dados pode ser perceptível a correlação espacial do fenômeno em estudo. O fluxograma para análise geoestatística é mostrado na Figura 58 e as ações nele representadas são descritas nos passos a seguir.

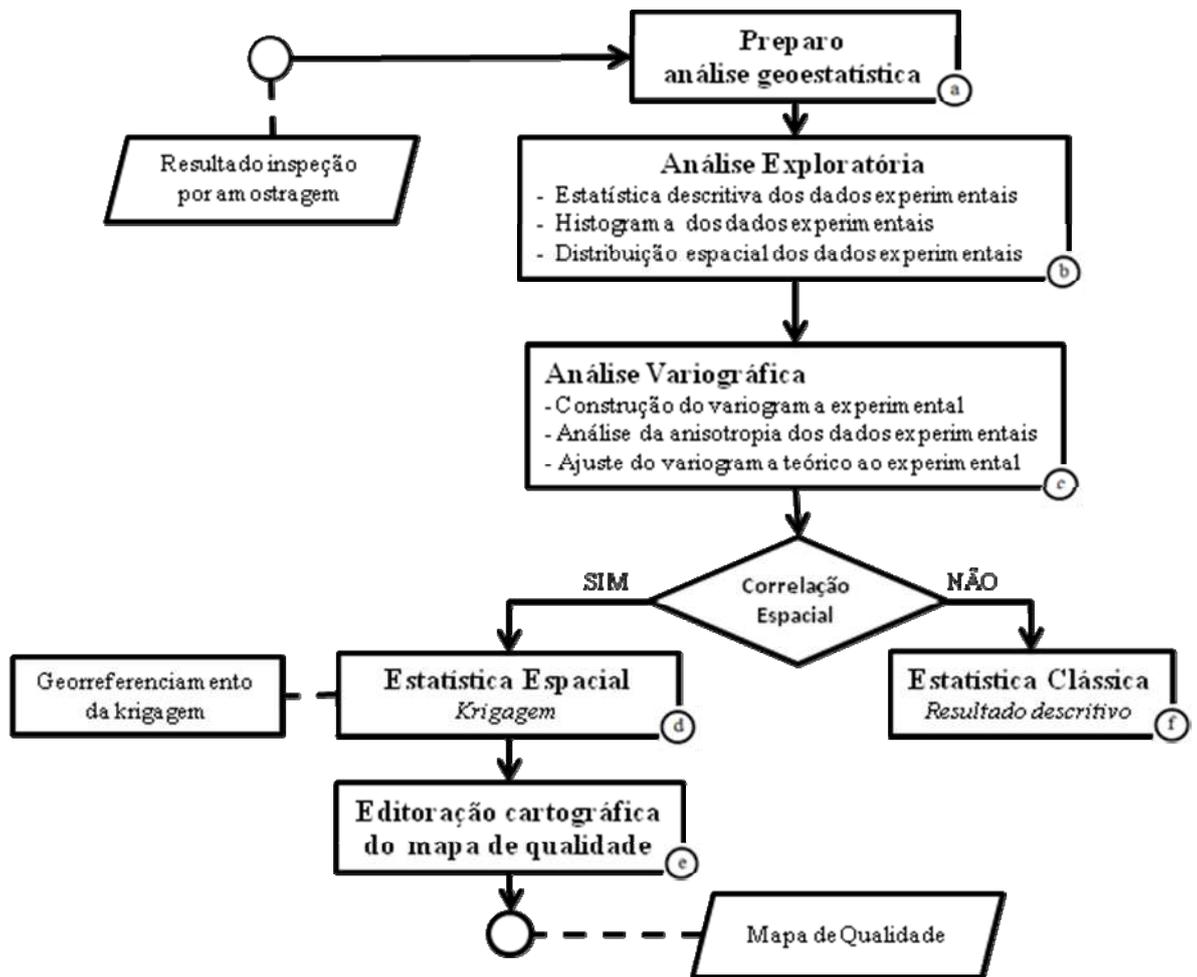


Figura 58 - Procedimentos da análise geostatística<sup>45</sup>

#### a) Preparo análise geostatística

Para análise geostatística os dados de entrada devem ser sintetizados e estruturados de forma a contemplar as coordenadas (X, Y) de cada unidade de amostragem (células da grade regular contínua selecionada) e o respectivo resultado da medida de qualidade avaliada (Z), após o tratamento da inspeção citado no tópico anterior. Assim, com base nestas três variáveis são iniciados os procedimentos para aplicação dos métodos geostatísticos.

#### b) Análise exploratória dos dados

Uma análise estatística preliminar é necessária para verificar o histograma do indicador de qualidade avaliado, determinado no resultado da medida de qualidade, e algumas informações estatísticas descritivas, como: o número de observações; a média, a variância dos dados observados; os valor máximo e mínimo.

<sup>45</sup> O georreferenciamento da krigagem é necessário devido ao uso de indicadores de qualidade e a "Grade Regular Contínua para Amostragem e Análise Geostatística", visto que esta possui um sistema de coordenadas auxiliar próprio para possibilitar o uso do método geostatístico.

Ressalta-se que, além das estimativas dos parâmetros estatísticos, é relevante observar situações gerais em relação a distribuição e localização das unidades de amostragem como: as coordenadas mínima e máxima, e as distâncias mínima e máxima entre as amostras.

**c) Análise variográfica.**

Esta é a etapa fundamental da análise geoestatística, onde é construído o variograma experimental, analisada a presença de anisotropia nos dados observados e realizado o ajuste do variograma teórico ao experimental. Inicialmente é construído e observado o variograma experimental com base nos dados observados e na distância máxima a qual devem ser formados os pares de amostras para calcular a variância. Em seguida, construído o variograma e observado a forma como as semivariâncias se apresentam no variograma é avaliada a presença de anisotropia nos dados, construindo outros variogramas com pares formados em direções diferentes da inicial. Finalizada esta etapa é realizado o ajuste de um variograma teórico ao variograma experimental. Para realizar o ajuste do variograma teórico ao experimental devem ser identificados parâmetros do modelo teórico selecionado, conforme ilustra o fluxograma da Figura 59.

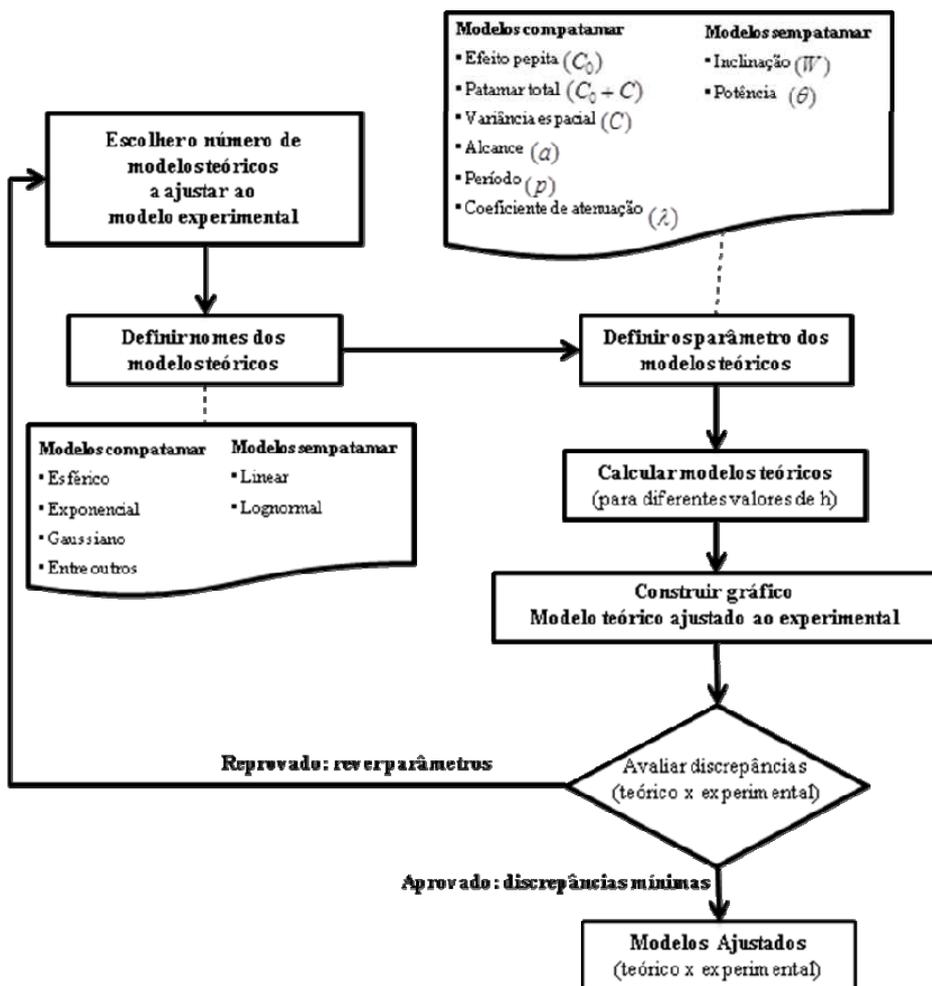


Figura 59 - Fluxograma do ajuste do variograma teórico ao experimental

No ajuste do variograma teórico ao experimental devem ser observados: o modelo teórico selecionado, o efeito pepita ( $C_0$ ), o patamar ( $C_0+C$ ), a variância espacial ( $C$ ) e o alcance ( $a$ ).

#### ***d) Estatística Espacial (krigagem)***

Caso na análise variográfica for identificada a correlação espacial no resultado da avaliação de qualidade, ou seja, a dependência espacial dos dados observados, a krigagem é executada para estimar os pontos não inspecionados na amostragem.

#### ***Georreferenciamento da krigagem***

Na presente dissertação, toda análise geoestatística foi realizada sobre áreas de amostragem contínuas, entretanto para isto foi necessário criar um sistema de referência auxiliar para atender aos preceitos da Geoestatística de continuidade do fenômeno em estudo. As posições linha e coluna foram associadas as células da grade regular contínua, criada no tópico "preparo da grade regular contínua para amostragem e análise geoestatística", e desta forma, a krigagem precisa ser georreferenciada para posterior criação dos mapas de qualidade.

A krigagem estima os valores das unidades de amostragem que não foram avaliadas, no caso desta pesquisa cada unidade de amostragem representa um índice de qualidade vinculado a respectiva célula da grade regular selecionada na amostragem.

Para realizar o georreferenciamento foi utilizado o software livre gvSIG, pois o Quantum GIS, apresentou inconsistências ao realizar esta tarefa em sua versão 1.8. Os passos mostrados na Figura 60 descrevem os procedimentos para georreferenciar a krigagem com base nas quadrículas extremas da medida de qualidade em estudo.



Figura 60 - Procedimentos para o georreferenciamento da krigagem

Para facilitar o entendimento, como exemplo, foi utilizado o georreferenciamento da krigagem dos dados observados de comissão de feições de hidrografia

#### **i) Preparar ambiente de trabalho: georreferenciamento da krigagem (ambiente gvSIG)**

Na caixa de diálogo inicial do gvSIG é necessário criar uma "Vista", em seguida no menu "*Vista > Propriedades*" foram determinadas as propriedades gerais do projeto

em relação ao sistema coordenadas cartográficas e o referencial geodésico, no caso desta pesquisa: coordenadas geográficas e referencial geodésico SIRGAS 2000 ("SIRGAS 2000 - EPGS 4674")<sup>46</sup>.

Os dados necessários para realizar o georreferenciamento são: a grade regular contínua, em formato shapefile, criada no tópico "preparo da grade regular contínua para amostragem e análise geoestatística"; e a krigagem na estrutura matricial, conforme descrito na metodologia desta dissertação, em formato "\*.asc"<sup>47</sup>.

ii) Selecionar os pontos extremos.

Os pontos extremos, da medida de qualidade avaliada, são as coordenadas extremas (linha e coluna). Para isto foi realizada uma consulta SQL na grade regular contínua, a fim de extrair as quadrículas extremas e auxiliar no georreferenciamento. Desta forma, a consulta SQL seleciona a linha mínima e máxima em conjunto com a coluna mínima e máxima.

Realizada a consulta, a seleção é exportada para um novo shapefile, denominado "*PontosExtremosCompleitude.shp*", para auxiliar no georreferenciamento da krigagem referente a comissão e omissão de feições geográficas de hidrografia. Desta forma, o ambiente para georreferenciar a krigagem está construído.

iii) Georreferenciar a krigagem

Os pontos extremos da krigagem presentes no arquivo raster "*Comissao.asc*" são associados aos extremos das quadrículas de amostragens, neste caso, presentes no shapefile "*PontosExtremosCompleitude.shp*". No gvSIG a ferramenta para esta etapa está presente no botão de "*Transformações geográficas > Georreferenciamento*", conforme Figura 61.

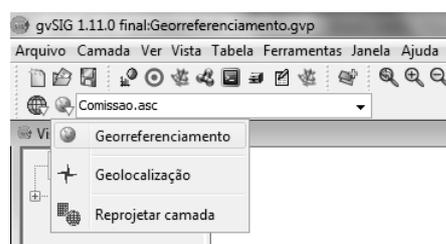


Figura 61 - Ferramenta de georreferenciamento no gvSIG.

A Figura 62 mostra a tela de georreferenciamento no gvSIG.

<sup>46</sup> *European Petroleum Survey Group* (EPGS) os parâmetros de diversos sistemas de coordenadas cartográficas e geodésicas são disponibilizados em <<http://www.spatialreference.org>>

<sup>47</sup> No ambiente R a krigagem foi salva na estrutura raster através da função "*writeGDAL()*" presente no pacote *rgdal*, no formato ASCII - *American Standard Code for Information Interchange* (Código Padrão Americano para o Intercâmbio de Informação)

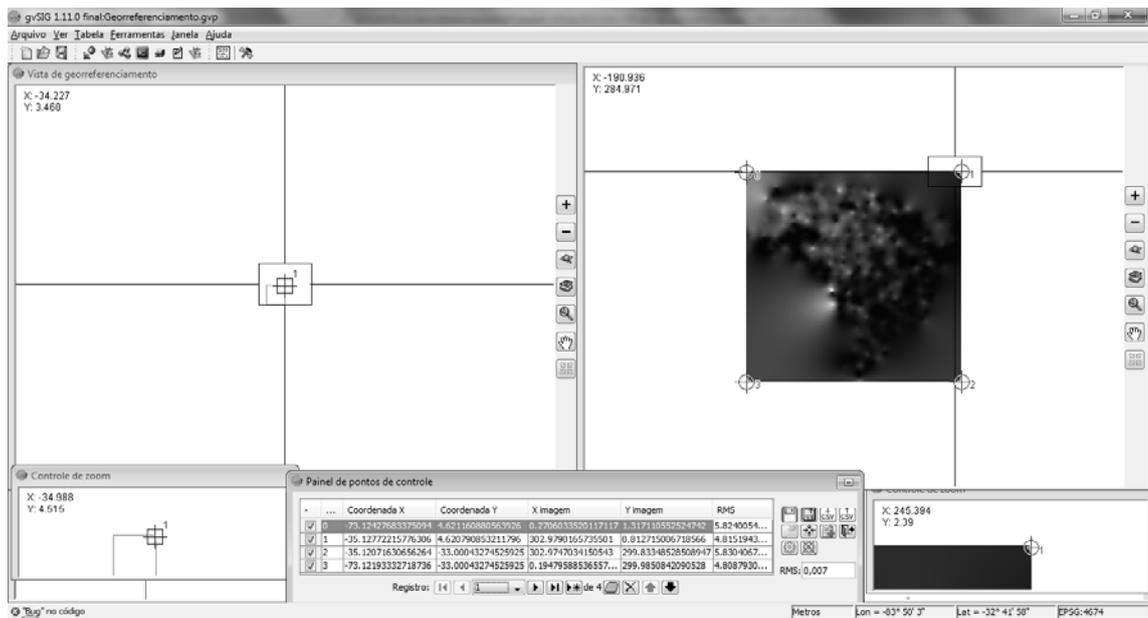


Figura 62 - Tela principal para o georreferenciamento no gvSIG

Desta forma, a parte de georreferenciamento é concluída e a krigagem pode ser associada espacialmente aos demais elementos da base cartográfica em estudo. O procedimento é válido para todos os resultados de inspeção por amostragem e aplicação da análise geoestatística.

#### ***e) Edição cartográfica<sup>48</sup> do mapa de qualidade***

Para contextualizar e localizar geograficamente o resultado da análise espacial é necessário realizar a edição cartográfica do mapa de qualidade, inserindo elementos básicos de um mapa e delimitando a área de interesse avaliada. Esta etapa conclui o processo de avaliação da qualidade e representatividade da mesma, sendo gerado um mapa de qualidade no final.

#### ***f) Estatística Clássica (resultado descritivo).***

Caso na análise variográfica for constatado que não há correlação espacial no resultado da avaliação de qualidade, ou seja, o variograma apresenta efeito pepita puro, a Estatística Clássica deve ser utilizada, assim como outro método de interpolação, visto que ocorre aleatoriedade nos valores analisados. Ressalta-se que nesta dissertação, outro método de interpolação não será abordado, mas o efeito pepita puro, indica o uso da Estatística Clássica.

<sup>48</sup> Segundo Fernandes e Ribeiro (2011) a edição cartográfica é o último passo na linha de produção de um mapa e visa atribuir à representação de objetos geográficos uma simbologia adequada e previamente estabelecida, segundo convenções cartográficas existentes.

### **3.3.4. Metadados da avaliação de qualidade**

Concluída as etapas anteriores de avaliação e análise espacial da qualidade do conjunto de dados em estudo, a próxima etapa é relatar os resultados e os métodos utilizados neste processo de aferição. Além disto, é definido o nível de conformidade da qualidade dos dados em relação a um parâmetro ou tolerância existente para comparação. Entretanto, nota-se a existência de parâmetros oficiais apenas para acurácia posicional, fornecidos pelo PEC, tornado inviável a definição e classificação do conjunto de dados para outros elementos de qualidade. Em relação ao preenchimento dos metadados de qualidade, conforme mencionado, a estrutura da medida de qualidade facilita este processo, a Tabela 8 mostra as seções presentes nos metadados de qualidade.

Além do preenchimento dos metadados de qualidade, é recomendada a geração de uma dos mapas de qualidade, em analogia às considerações presentes em IBGE (2011), cujo objetivo é fornecer uma percepção geral do conjunto de dados, neste caso de um mapa de qualidade, ao usuário final.

A presente seção descreveu a metodologia elaborada nesta dissertação, onde o foco é a avaliação da qualidade de dados na estrutura vetorial e a análise espacial através da Geoestatística. O próximo capítulo apresenta os resultados e análises da aplicação desta metodologia sobre um conjunto de dados espaciais disponibilizados no Portal da INDE, em relação a avaliação de dois elementos de qualidade: a completude e a acurácia posicional.

#### **4. RESULTADOS E DISCUSSÃO**

Um dos propósitos desta dissertação é fornecer uma metodologia de avaliação da qualidade de dados presentes no Portal da INDE. Assim, a Base Cartográfica Vetorial Contínua 1:250.000 - BC250 foi escolhida como o conjunto de dados a avaliar. Em relação aos elementos de qualidade, e considerando a extensão da base cartográfica a ser avaliada, foram selecionados os elementos de completude, tanto em termos de omissão (ausência de feições) quanto comissão (excesso de feições), e a acurácia posicional para análise do subelemento acurácia posicional absoluta. Ressalta-se a importância de um estudo preliminar do conjunto de dados através das especificações técnicas do produto cartográfico, no caso da BC250 descritas em IBGE (2011a) e IBGE (2011b), atendendo os requisitos iniciais da avaliação da qualidade (Figura 37).

Na completude foram avaliadas as classes de feições presentes na categoria hidrografia, mas especificamente trechos de drenagem, trechos massa d'água e massa d'água, conforme especificações técnicas do produto, onde foi delimitada uma área de avaliação para posterior cálculo das taxas de comissão e omissão. Para o caso da avaliação da acurácia posicional absoluta foram considerados todos os elementos do produto cartográfico, onde pontos notáveis na referida base cartográfica foram identificados, medidos e comparados aos elementos de referência coletados através de informações de campo. O PEC foi utilizado como parâmetro para avaliar a conformidade do conjunto de dados em relação a avaliação da qualidade posicional, ao contrário da completude, pois não existe um parâmetro de referência para comparação e determinação do nível da qualidade deste produto cartográfico.

Posteriormente, com base nos resultados de Completude, incluindo taxas de omissão e comissão, e na Acurácia Posicional e erro planimétrico em metros, foi realizada a análise espacial, através de métodos geoestatísticos a fim de identificar e representar uma possível correlação espacial dessas variáveis em estudo.

Finalizando a avaliação de qualidade foram preenchidos os metadados de qualidade em relação aos elementos de qualidade avaliados.

Os resultados obtidos são descritos em detalhe a seguir, sendo agrupados segundo os elementos de qualidade em avaliação.

##### **4.1. Avaliação da qualidade dos dados espaciais em relação a completude**

Conforme mencionado, foi realizado um estudo preliminar das especificações técnicas da BC250, a partir da qual foi identificado o elemento de qualidade a avaliar, no caso a completude, e o agrupamento de feições de hidrografia, no caso as classes:

trecho de drenagem, trecho massa d'água e massa d'água, conforme definido nas especificações técnicas do produto cartográfico, conforme a ET-EDGV (CONCAR, 2008). Em relação ao âmbito da inspeção, considerando o propósito desta dissertação de auxiliar na avaliação da qualidade de dados da INDE, foi escolhida toda a extensão territorial do Brasil, afim de produzir um mapa de qualidade integral e não apenas regional. A aplicação da metodologia para completude é similar em termos de comissão e omissão, com alteração apenas nos resultados apresentados e na análise espacial em si. Desta forma, seguindo o fluxograma mostrado na Figura 37, a aplicação desta metodologia é descrita abaixo. Ressalta-se que, como as avaliações de qualidade em termos de comissão e omissão são similares, na primeira parte demonstra-se de forma única os dois subelementos da completude. Apenas na análise espacial e preenchimento dos metadados de qualidade estes são mostrados de maneira distinta.

#### **4.1.1. Identificar a medida de qualidade**

Para este estudo foram criadas medidas de qualidade de completude (omissão e comissão) as quais encontram-se descritas no Apêndice A. Os Apêndices A.1 e A.2 descrevem as medidas de qualidade referentes aos casos de omissão e comissão e como deve ser realizada a inspeção de qualidade.

#### **4.1.2. Método de inspeção**

Para avaliação da comissão e omissão de feições presentes na BC250 foi definida a inspeção por amostragem das classes: trecho de drenagem, trecho massa d'água e massa d'água da categoria de informação hidrografia.

##### **4.1.2.1. Resultados da inspeção por amostragem para completude**

Os resultados das medidas de qualidade em relação a itens em excesso e itens ausentes na BC250 de forma geral, sem considerar o comportamento espacial, são representados nas taxas de comissão e omissão, conforme Salgé (1995):

$$\tau^+ = \frac{N^+}{\max(N^o, N^r)} = \frac{3146}{24706} = 0.1273 = 12.73\% \text{ (taxa de comissão geral)}$$

$$\tau^- = \frac{N^-}{\max(N^o, N^r)} = \frac{614}{24706} = 0.0248 = 2.48\% \text{ (taxa de omissão geral)}$$

Logo, na BC250 foi considerado, de acordo com os testes realizados, que existem 12,73% de itens em excesso e 2,48% de itens ausentes em relação às classes de feições de hidrografia avaliadas. Entretanto, esta porcentagem não é homogênea em toda a base, conforme foi percebido durante o processo de inspeção por amostragem e como foi comprovado nos resultados de análise espacial através de métodos geoestatísticos descritos a seguir.

## **4.2. Análise espacial da qualidade: Geoestatística aplicada na avaliação da completude**

Finalizada a etapa de avaliação da medida de qualidade foi iniciada a etapa de análise espacial dos resultados obtidos, representados nas unidades de amostragem de completude, tanto de omissão quanto comissão. Em seguida o objetivo foi analisar o comportamento espacial dos dados, no sentido de comprovar e determinar a correlação espacial perceptível durante o processo de inspeção por amostragem e representar espacialmente esta percepção. Apesar dos resultados de omissão e comissão estarem representados dentro de uma única tabela de atributos foram realizadas análises geoestatísticas distintas.

Esta etapa foi realizada utilizando os softwares R e Tinn-R, com base no fluxo proposto na Figura 58, onde os resultados foram obtidos através de rotinas (*scripts*) elaboradas<sup>49</sup>. O software R trabalha por meio de pacotes que são carregados conforme a necessidade da aplicação, assim para execução das rotinas os pacotes: *geoR* (geoestatística), *rgdal* (tratamento do resultado em estrutura matricial) *scatterplot3d* (gráficos) e *tcltk2* (auxiliar no ajuste do variograma) são recomendados. No final da análise geoestatística foi utilizado o ambiente SIG, visto que, para atender a preceitos da geoestatística de continuidade da variável em estudo, a análise com o sistema de referência utilizado foi a posição relativa linha e coluna da unidade de amostragem, contudo, ao final dos procedimentos foi realizado os passos descritos no tópico "*georreferenciamento da krigagem*".

### **4.2.1. Preparo da análise geoestatística em termos de comissão e omissão**

Os resultados representados na unidade de amostragem de completude foram preparados e separados em dois arquivos no formato "txt", compostos por uma estrutura de colunas X, Y e Z. Para isto foi utilizado o software livre *LibreOffice Calc*, onde a tabela de atributos da unidade de amostragem de completude, presente no arquivo de formato "dbf", foi editada de forma a conter apenas os campos "COLUNA, LINHA, TXComissão"<sup>50</sup>. Em seguida estes campos foram renomeados para X, Y, Z, todos os registros foram selecionados, copiados e salvos em formato texto, arquivo ASCII. O mesmo procedimento foi realizado para o caso da omissão. A Figura 63 mostra

---

<sup>49</sup> As rotinas (*scripts*) em R elaboradas pelo Prof. Dr. Gerson Rodrigues dos Santos no curso de extensão "Introdução a Geoestatística no R", ministrado na UFV em 2011, foram adaptadas a presente dissertação.

<sup>50</sup> A ordem de campos deve ser obedecida (X,Y,Z), onde o eixo X é representado pela COLUNA, o eixo Y é representado pela LINHA, e o valor Z em estudo é representado pelo resultado da medida de qualidade.

parcialmente os dados observados contidos nos arquivos "ComissaoXYZ.txt" e "OmissaoXYZ.txt" gerados.

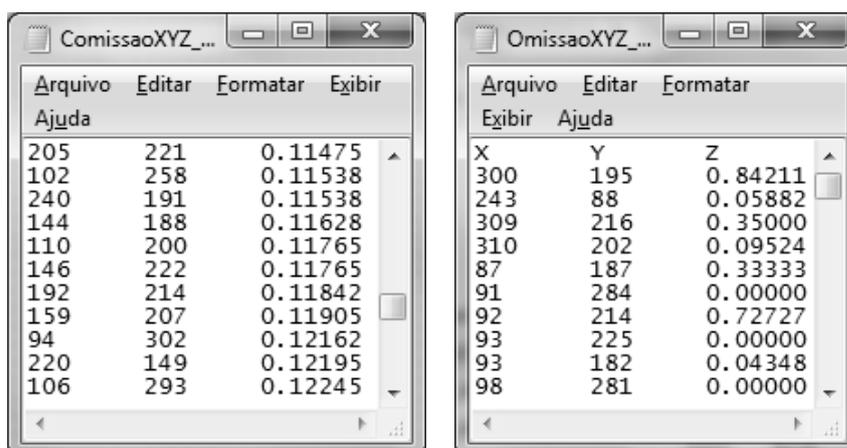


Figura 63 - Dados observados da taxa de comissão e omissão no formato texto

O arquivos texto gerados foram dados de entrada na rotina (*script*) em R, através da função "*read.geodata( )*". Desta forma, os parâmetros necessários para análise geoestatística foram observados, através do resultado de funções, e inseridos por meio de variáveis definidas na rotina. Em sequência são apresentados os resultados da análise geoestatística dos dados de comissão, ressaltando que os procedimentos são similares para qualquer estudo geoestatístico. Posteriormente são apresentados os resultados para os dados de omissão.

#### 4.2.2. Análise exploratória dos dados de comissão

Na análise exploratória dos dados foram obtidos os parâmetros da estatística descritiva presentes na Tabela 11 e o histograma dos dados observados conforme ilustrado na Figura 64.

TABELA 11 - Estatística descritiva referente à taxa de comissão de feições geográficas de hidrografia da BC250

ESTATÍSTICA DESCRITIVA	MEDIDA DE QUALIDADE
	M10101 (Comissão)
N	571
Média	0.0920 %
Desvio padrão	0.1122 %
Variância dos dados observados	0.0126
Máximo	0.8235 %
Quartil superior	0.1547 %
Mediana	0.0526 %
Quartil inferior	0%
Mínimo	0%

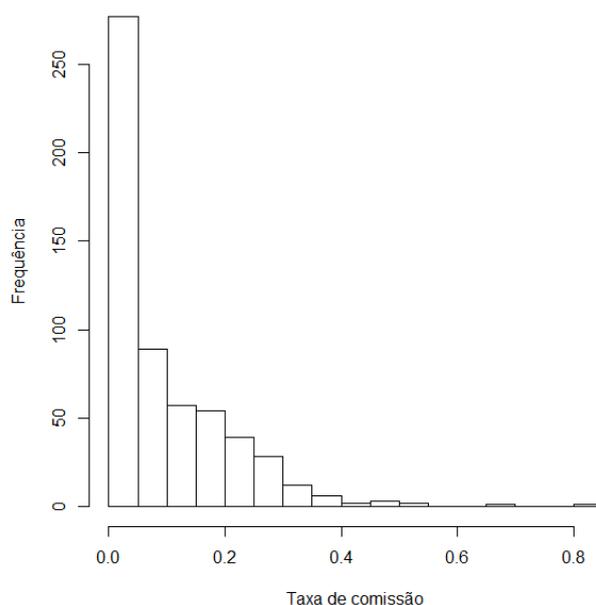


Figura 64 - Histograma referente à taxa de comissão de feições geográficas de hidrografia da BC250.

A distribuição espacial da variável em estudo é mostrada na Figura 65, o tamanho dos círculos são proporcionais aos valores observados. Além dos valores da Tabela 11 é interessante observar as coordenadas extremas dos dados, assim como as distâncias mínimas e máximas entre as amostras, conforme apresentados na Tabela 12.

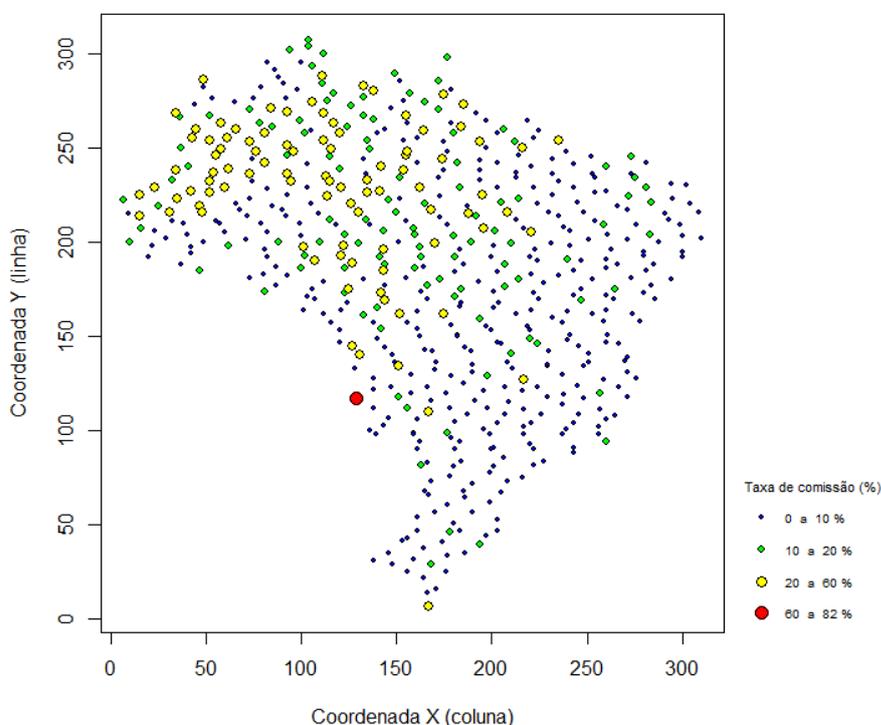


Figura 65 - Distribuição espacial dos dados observados referentes à taxa de comissão de feições geográficas de hidrografia da BC250 <sup>51</sup>

<sup>51</sup> Cada unidade de amostragem (quadrícula) representa uma área de aproximadamente 13km<sup>2</sup>, corresponde a uma célula da grade regular contínua, ou seja, cada intervalo de linha e coluna possui aproximadamente uma distância de 13km.

TABELA 12 - Pontos extremos das unidades de amostragem de completude

	X	Y	Distância (quadrículas)
Mínima	7	7	1.0000
Máxima	310	307	306.5436

Nota-se visualmente na Figura 65 um agrupamento mais intenso na região norte do país e outros pontos de maior magnitude isolados nas demais regiões. A Geoestatística dispõe de ferramentas, no caso o variograma, para identificar esta correlação espacial, conforme é apresentado na sequência.

#### 4.2.3. Análise variográfica dos dados de comissão

Realizada a análise exploratória dos dados, o procedimento de análise variográfica foi feita. Primeiro foi construído o variograma empírico, ou seja, o variograma experimental, com base na determinação da distância máxima (também denominada de "*lag*"). A literatura sobre o tema recomenda que seja considerado a metade do campo geométrico dos dados, ou seja, a metade da maior distância entre as unidades de amostragem, desta forma o valor do *lag* inicial foi de 153. Entretanto, como este valor foi insuficiente para observar o patamar no variograma experimental, então o mesmo foi alterado para 180 e o resultado é mostrado na Figura 66.

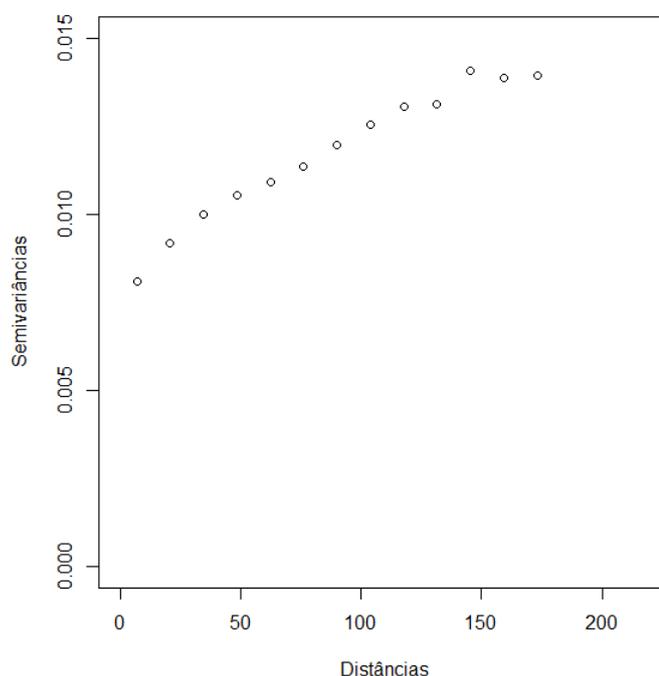


Figura 66 - Variograma experimental dos dados de comissão

Na análise anisotrópica foi observada variações zonais mínimas, em torno de 0,05% e o mesmo foi considerado isotrópico. A Tabela 13 apresenta os parâmetros determinados durante a análise variográfica e a Figura 67 mostra o variograma teórico ajustado ao variograma experimental.

TABELA 13 - Parâmetros do ajuste do variograma teórico ao experimental da taxa de comissão de feições geográficas de hidrografia da BC250

<b>MEDIDA DE QUALIDADE PARA ITENS EM EXCESSO (M10101)</b>	
Modelo teórico	Esférico
Efeito pepita ( $C_0$ )	0.0077
Patamar ( $C_0+C$ )	0.0145
Variância espacial ( $C$ )	0.0068
Variância dos dados observados	0.0126
Alcance ( $a$ )	150 (quadrículas)
Grau de aleatoriedade ( $E = C_0 / C$ )	1.132353

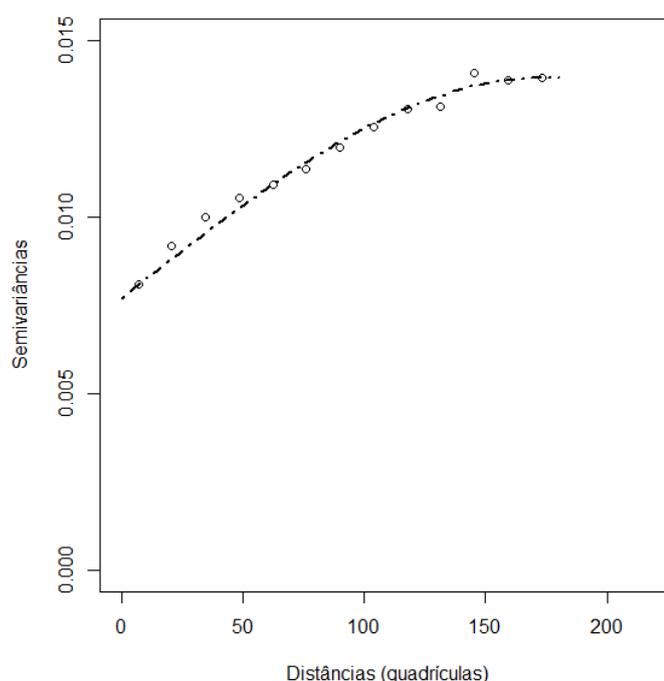


Figura 67 - Ajuste do variograma teórico ao experimental dos dados de comissão.

Segundo os dados presentes na Tabela 13, após o ajuste do variograma, o grau de aleatoriedade "bastante importante", conforme classificação dos graus de aleatoriedade mostrada na Tabela 5, em função do valor representativo do efeito pepita, indica que há uma aleatoriedade nos dados até atingir o valor do efeito pepita, em função da escala da grade regular contínua e da descontinuidade do fenômeno em estudo. Deste ponto em diante constata-se a dependência espacial nos resultados observados até atingir o patamar e voltar a apresentar aleatoriedade em função da distância entre as amostras. O comportamento do variograma indica uma contínua correlação espacial em relação a comissão de feições de hidrografia.

A Tabela 14 mostra os resultados da validação cruzada em relação ao ajuste do variograma teórico ao experimental em relação a comissão das feições geográficas de hidrografia.

TABELA 14 - Resultados da autovalidação do ajuste do variograma teórico ao experimental da taxa de comissão de feições geográficas de hidrografia da BC250

Parâmetros da autovalidação	Comissão
$\beta_1$	1.0172
$\beta_0$	- 0.0015

Através da regressão linear é possível observar a correlação entre os valores observados e o seus valores estimados. O valor de  $\beta_1$  foi de 1.0172 e  $\beta_0$  foi de - 0.0015, indicando um bom ajuste do variograma teórico ao experimental.

#### 4.2.4. Krigagem dos dados de comissão

Realizado o ajuste do variograma teórico ao experimental foi realizada a krigagem, a fim de estimar os valores das quadrículas não inspecionadas, ressaltando que antes deste passo foi necessário definir uma grade para inserir os valores estimados. A dimensão desta grade deve ser baseada na diferença entre as extremidades das quadrículas inspecionadas, conforme mostra a Tabela 12, para possibilitar o georreferenciamento da krigagem, conforme descrito no tópico sobre este assunto na metodologia.

#### 4.2.5. Mapa de qualidade de comissão

Para finalizar a etapa de análise espacial, em relação a qualidade dos dados considerando a comissão, fez-se o georreferenciamento da krigagem, e após foi realizado a editoração cartográfica no sentido de facilitar sua leitura e localização geográfica, inserindo elementos básicos de um mapa e delimitando a área de interesse avaliada pelos métodos geoestatísticos. Esta etapa concluiu o processo de avaliação da qualidade e representatividade da comissão, sendo gerado um mapa de qualidade no final, conforme Figura 68.

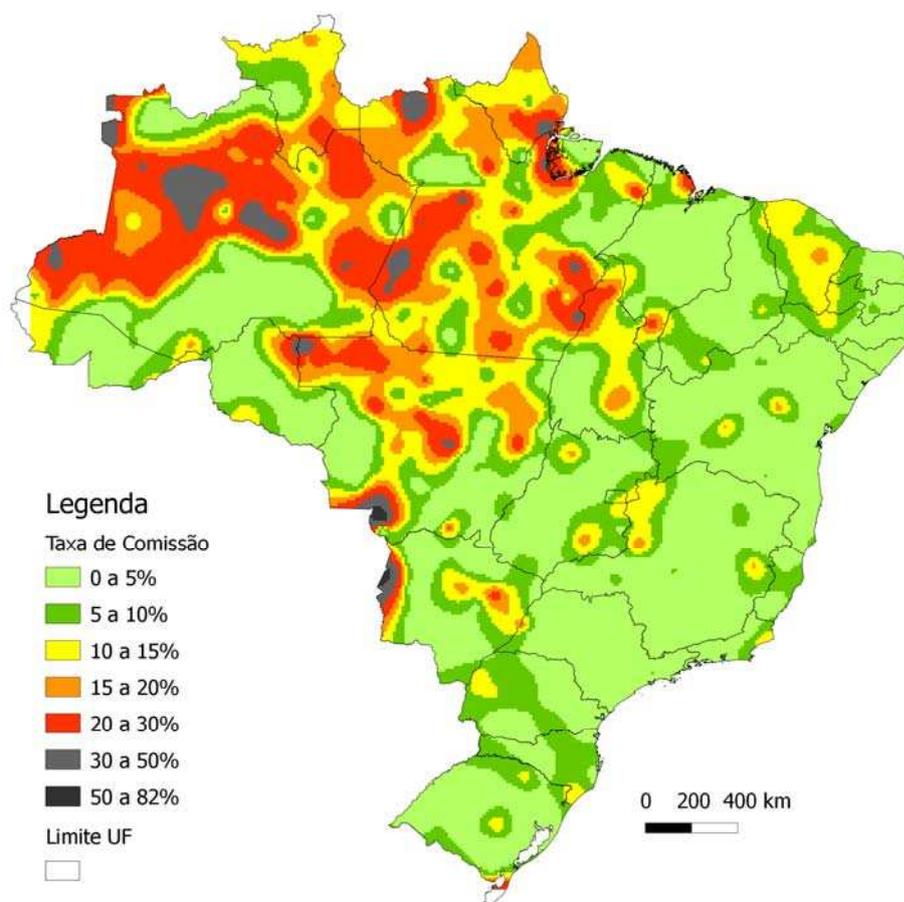


Figura 68 - Mapa de qualidade referente à taxa de comissão de feições geográficas de hidrografia da BC250

### ***Análise dos resultados do mapa de qualidade de comissão***

Segundo a equipe produtora da BC250 esta regionalização e densidade de itens em excesso ocorreu em função da linhagem dos insumos da Base Cartográfica Contínua na escala 1:250.000, mas especificamente das folhas topográficas e planimétrica na escala 1:250.000. Esta densidade de elementos de hidrografia iniciou-se no mapeamento topográfico desta escala na região amazônica no início da década de 1970, através do processo de restituição analógica realizado a partir de fotografias aéreas e da extração de feições a partir de imagens de radar, segundo as especificações da época. No final da década de 1990, com a criação do projeto SIVAM (Sistema de Vigilância da Amazônia), estas folhas topográficas foram convertidas do meio analógico para o digital (formato CAD - *Computer Aided Design* ou Desenho Assistido por Computador) e atualizadas por imagens do projeto Landsat, mantendo a densidade de feições de hidrografia existente. No presente momento, o projeto BC250 integrou todo mapeamento topográfico da escala 1:250.000 e converteu a estrutura CAD para a de Banco de Dados Geográficos, segundo as especificações da ET-EDGV, com o objetivo de compor a base cartográfica contínua desta escala na INDE. Estes erros sistemáticos

indicam futuras melhorias neste produto cartográfico com base em suas especificações técnicas

#### 4.2.6. Análise exploratória dos dados de omissão

A metodologia para omissão é similar ao realizado para comissão, o que realmente difere são os resultados, conforme é mostrado na Tabela 15 e no histograma dos dados observados ilustrado na Figura 69.

TABELA 15 - Estatística descritiva referente à taxa de omissão de feições geográficas de hidrografia da BC250

ESTATÍSTICA DESCRITIVA	MEDIDA DE QUALIDADE
	M10201 (Omissão)
N	571
Média	0.0444 %
Desvio padrão	0.1131 %
Variância dos dados observados	0.01281197
Máximo	0.8571 %
Quartil superior	0.0304 %
Mediana	0
Quartil inferior	0
Mínimo	0

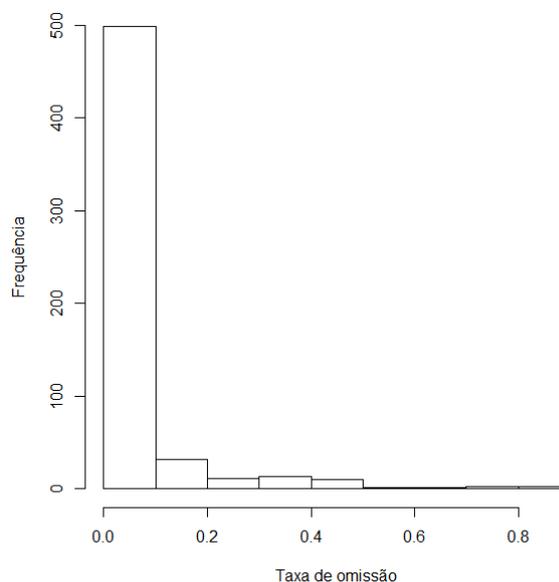


Figura 69 - Histograma referente à taxa de omissão de feições geográficas de hidrografia da BC250.

As coordenadas extremas para unidades de amostragem de omissão são as mesmas descritas para os casos de comissão, conforme mostra a Tabela 12. O mesmo é válido para a distribuição espacial das unidades de amostragem, variando apenas a intensidade dos valores, conforme mostra a Figura 70.

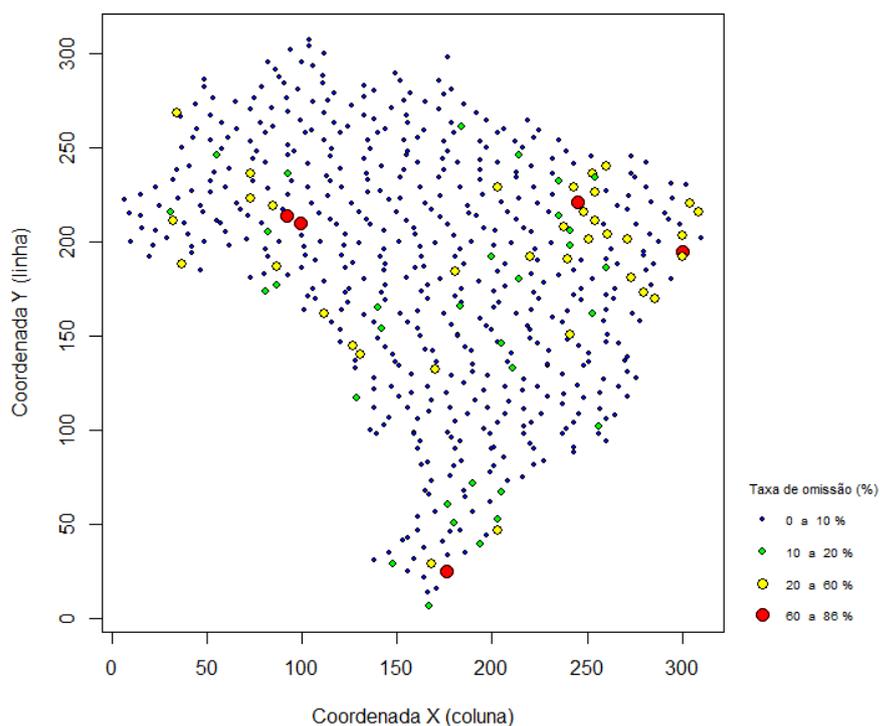


Figura 70 - Distribuição espacial dos dados observados referentes à taxa de omissão de feições geográficas de hidrografia da BC250

Nota-se visualmente na Figura 70 distintos agrupamentos mais intensos na região norte, nordeste e sul do país. A Geoestatística dispõe de ferramentas para identificar esta correlação espacial, conforme é apresentado na sequência.

#### 4.2.7. Análise variográfica dos dados de omissão

Desta forma, foi realizada a análise variográfica para os casos de omissão. Assim como em comissão foi utilizado a distância máxima (*lag*) de 250 quadrículas para a construção do variograma experimental, conforme ilustrado na Figura 71.

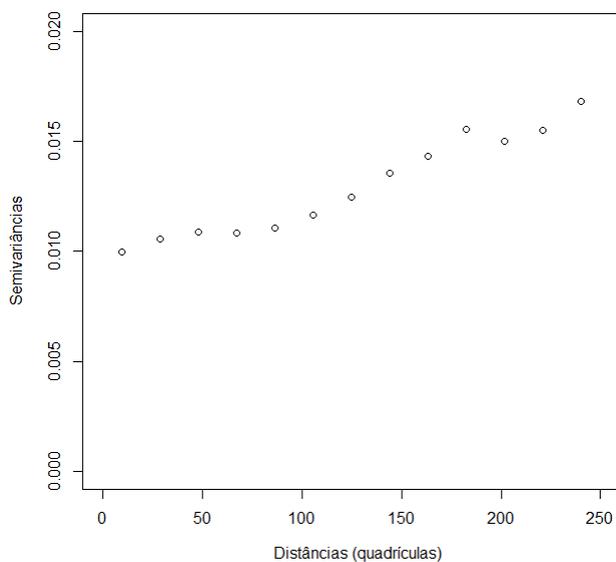


Figura 71 - Variograma experimental dos dados de omissão

Na análise anisotrópica foram observadas variações mínimas, em torno de 0,01%, sendo considerado isotrópico. A Tabela 16 apresenta os parâmetros determinados durante a análise variográfica e a Figura 72 mostra o variograma teórico ajustado ao variograma experimental.

TABELA 16 - Parâmetros do ajuste do variograma teórico ao experimental da taxa de omissão de feições geográficas de hidrografia da BC250

MEDIDA DE QUALIDADE PARA ITENS AUSENTES (M10201)	
Modelo teórico	Gaussiano
Efeito pepita ( $C_0$ )	0.0094
Patamar ( $C_0+C$ )	0.0137
Variância espacial ( $C$ )	0.0043
Variância dos dados observados	0.01281197
Alcance ( $a$ )	146 (quadrículas)
Grau de aleatoriedade ( $E = C_0 / C$ )	2.186047

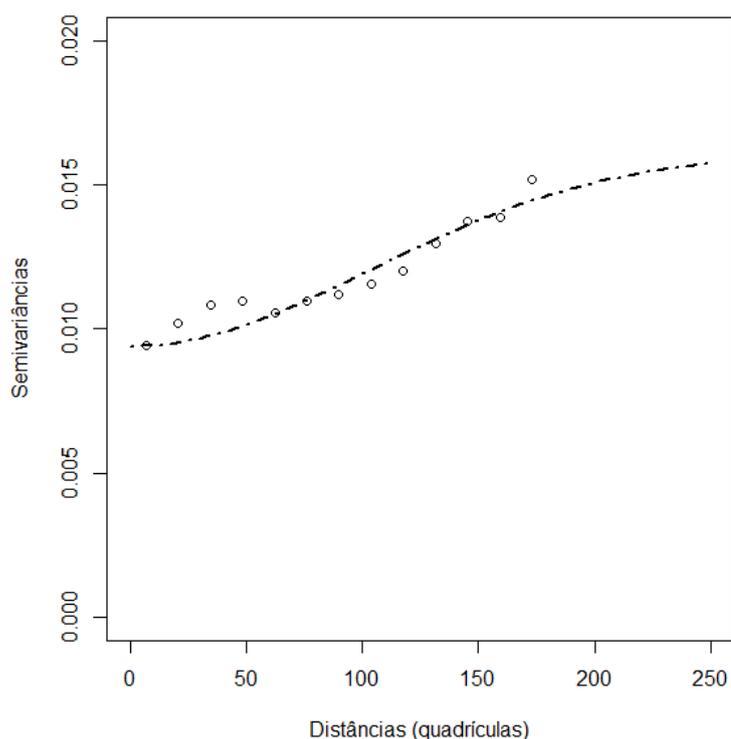


Figura 72 - Ajuste do variograma teórico ao experimental dos dados de omissão.

Segundo os dados presentes na Tabela 16, após o ajuste do variograma, o grau de aleatoriedade é "bastante importante", conforme classificação dos graus de aleatoriedade mostrada na Tabela 5, em função do valor representativo do efeito pepita, que indica que há uma aleatoriedade nos dados até atingir o valor do efeito pepita, em

função da escala da grade regular contínua e das discontinuidades do fenômeno em estudo, deste ponto em diante constata-se a dependência espacial nos resultados observados até atingir o patamar e voltar a apresentar aleatoriedade em função da distância entre as amostras. O comportamento do variograma indica regionalizações, em função de sua aparente estrutura imbricada, em relação a omissão de feições de hidrografia.

A Tabela 17 mostra os resultados da validação cruzada em relação ao ajuste do variograma teórico ao experimental em relação a omissão das feições geográficas de hidrografia.

TABELA 17 - Resultados da autovalidação do ajuste do variograma teórico ao experimental da taxa de omissão de feições geográficas de hidrografia da BC250

Parâmetros da Autovalidação	Omissão
$\beta_1$	0.9360
$\beta_0$	0.0028

Através da regressão linear é possível observar a correlação entre os valores observados e o seus valores estimados. O valor de  $\beta_1$  foi de 0.9360 e  $\beta_0$  foi de 0.0028, indicando um bom ajuste do variograma teórico ao experimental.

#### 4.2.8. Krigagem dos dados de omissão

Realizado o ajuste do variograma foi realizada a krigagem, da mesma forma como foi realizada para a comissão, e a imagem gerada foi georreferenciada.

#### 4.2.9. Mapa de qualidade de omissão

Para finalizar esta etapa, após o georreferenciamento da krigagem, foi feita a editoração cartográfica do mapa de qualidade de dados em relação a omissão e o resultado é apresentado na Figura 73.

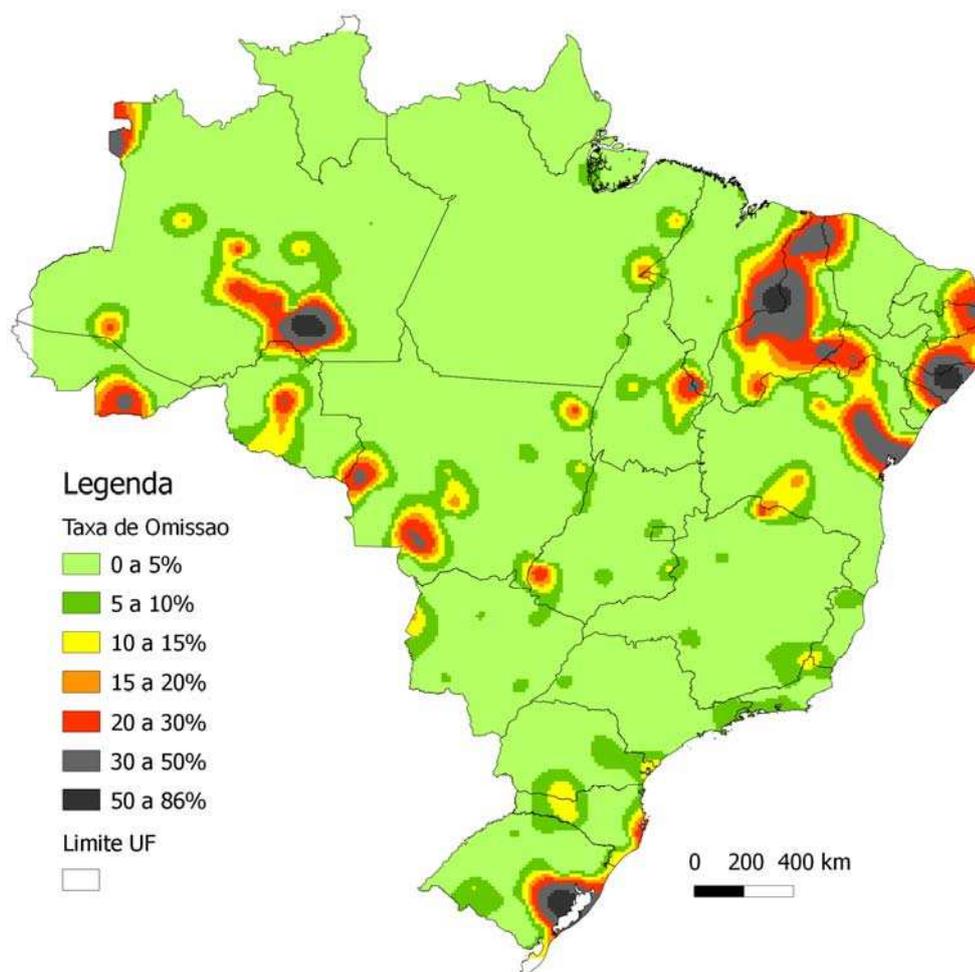


Figura 73 - Mapa de qualidade referente à taxa de omissão de feições geográficas de hidrografia da BC250

#### ***Análise dos resultados do mapa de qualidade de comissão***

A análise do resultado é similar a realizada para os casos de completude, onde segundo a equipe produtora da BC250 esta regionalização e densidade de itens em excesso ocorreu em função da linhagem dos insumos da Base Cartográfica Contínua na escala 1:250.000, mas especificamente das folhas topográficas e planimétrica na escala 1:250.000. Esta ausência de elementos de hidrografia iniciou-se no mapeamento topográfico desta escala na região amazônica no início da década de 1970, através do processo de restituição analógica realizado a partir de fotografias aéreas e da extração de feições a partir de imagens de radar, segundo as especificações da época, sobre a qual as feições não foram identificadas. Na região do Piauí, a origem dos dados são cartas planimétricas, onde a representação dos rios foi escassa. Um indicativo para atualização destas regiões e auxiliar na identificação das feições ausentes é utilizar insumos de sensoriamento remoto de épocas distintas do ano. Estes erros sistemáticos indicam futuras melhorias neste produto cartográfico com base em suas especificações técnicas.

### 4.3. Metadados da avaliação de qualidade da completude

Neste tópico é apresentado o preenchimento dos metadados, segundo a norma ISO 19115:2003, baseado nos resultados das inspeções realizadas referentes as medidas de qualidade de comissão e omissão de feições de hidrografia do conjunto de dados espaciais em análise, no caso a BC250. Em relação a definição do nível de conformidade da completude (comissão e omissão) não existem parâmetros para comparação, conforme mencionado em relação à medida de qualidade, assim este nível de conformidade é indefinido.

#### 4.3.1. Metadados de qualidade: taxa de comissão de elementos de hidrografia.

Em relação a taxa de comissão de elementos de hidrografia da BC250, concluídas as etapas anteriores de avaliação e análise espacial, e de posse das unidades de amostragem sintetizando os resultados da avaliação da qualidade e do mapa de qualidade, os metadados de qualidade são apresentados na Tabela 18.

TABELA 18 - Metadados de qualidade: taxa de comissão de feições geográficas de hidrografia da BC250

<b>Componente DQ (Data Quality)</b>	<b>Descrição / Exemplos</b>
<i>Data Quality Scope</i> (DQ_Scope)	Todos os elementos de trecho de drenagem, trecho massa d'água e massa d'água, presentes na categoria de informação hidrografia.
<i>Data Quality Element</i> (DQ_Element)	1 - Completude
<i>Data Quality Measure</i> (DQ_Measure)	1 - Comissão
<i>Data Quality Measure Description</i> (DQ_MeasureDesc)	A medida de qualidade consiste em determinar a taxa de comissão (itens em excesso) em relação as feições de hidrografia, mas especificamente: trecho de drenagem, trecho massa d'água e massa d'água.
<i>Data Quality Measure Identification</i> (DQ_MeasureID)	10101
<i>Data Quality Evaluation Method</i> (DQ_EvalMethod)	
<i>Data Quality Evaluation Method Type</i> (DQ_EvalMethodType)	2 - Externo
<i>Data Quality Evaluation Method Description</i> (DQ_EvalMethodDesc)	Comparar o item presente no conjunto de dados com o presente no universo de discussão, para posterior contagem da proporção de itens incoerentes.
<i>Data Quality Result</i> (DQ_QualityResult)	
<i>Data Quality Value Type</i> (DQ_ValueType)	4 - Porcentagem
<i>Data Quality Value</i> (DQ_Value)	-
<i>Data Quality Value Unit</i> (DQ_ValueUnit)	%
<i>Data Quality Date</i> (DQ_Date)	2013-02-17 a 2013-04-26
<i>Data Quality Conformance Level</i> (DQ_ConformanceLevel)	Num universo de 571 unidades de amostragem, a taxa de comissão para a feições de hidrografia avaliada (trecho drenagem, trecho massa d'água e massa d'água) foi de 12,73% na Base Cartográfica Contínua 1:250.000.

Além do preenchimento dos metadados de qualidade, é recomendada a geração da imagem referente ao mapa de qualidade dos casos de comissão de elementos de hidrografia, em analogia as considerações presentes em IBGE (2011), com o objetivo de

fornecer uma percepção visual da qualidade do conjunto de dados espaciais ao usuário final. Ressalta-se que, os mapas de qualidade gerados são produtos cartográficos, logo pode haver o preenchimento dos metadados dos mesmos segundo o Perfil MGB.

#### 4.3.2. Metadados de qualidade: taxa de omissão de elementos de hidrografia

Considerando a taxa de omissão de elementos de hidrografia da BC250, e após concluídas as etapas anteriores de avaliação e análise espacial, a próxima etapa foi relatar os resultados e métodos utilizados durante o processo de aferição. Como no caso da comissão, não existem parâmetros para comparação, assim o nível de conformidade é indeterminado. Em relação as demais seções dos metadados de qualidade, a Tabela 19 apresenta os metadados de qualidade relatados em relação a omissão, conforme orienta as normas ISO 19114:2003 e ISO 19115:2003.

TABELA 19 - Metadados de qualidade: taxa de omissão de feições geográficas de hidrografia da BC250

Componente DQ ( <i>Data Quality</i> )	Descrição / Exemplos
<i>Data Quality Scope</i> (DQ_Scope)	Todos os elementos de trecho de drenagem, trecho massa d'água e massa d'água.
<i>Data Quality Element</i> (DQ_Element)	1 - Completude
<i>Data Quality Measure</i> (DQ_Measure)	1 - Omissão
<i>Data Quality Measure Description</i> (DQ_MeasureDesc)	A medida de qualidade consiste em determinar a taxa de omissão (itens ausentes) em relação as feições de hidrografia, mas especificamente: trecho de drenagem, trecho massa d'água e massa d'água.
<i>Data Quality Measure Identification</i> (DQ_MeasureID)	10201
<i>Data Quality Evaluation Method</i> (DQ_EvalMethod)	
<i>Data Quality Evaluation Method Type</i> (DQ_EvalMethodType)	2 - Externo
<i>Data Quality Evaluation Method Description</i> (DQ_EvalMethodDesc)	Comparar o item presente no conjunto de dados com o presente no universo de discussão, para posterior contagem da proporção de itens incoerentes.
<i>Data Quality Result</i> (DQ_QualityResult)	
<i>Data Quality Value Type</i> (DQ_ValueType)	4 - Porcentagem
<i>Data Quality Value</i> (DQ_Value)	-
<i>Data Quality Value Unit</i> (DQ_ValueUnit)	%
<i>Data Quality Date</i> (DQ_Date)	2013-02-17 a 2013-04-26
<i>Data Quality Conformance Level</i> (DQ_ConformanceLevel)	Num universo de 571 unidades de amostragem, a taxa de omissão para as feições de hidrografia avaliadas foi de 2,48% de itens ausentes, agrupados em determinadas regiões conforme pode ser visualizado no mapa de qualidade de omissão da BC250.

Da mesma forma que para o caso da comissão, recomenda-se a geração da imagem referente ao mapa de qualidade para os casos de omissão de elementos de hidrografia, em analogia as considerações presentes em IBGE (2011). Ressalta-se que,

os mapas de qualidade gerados são produtos cartográficos, logo pode haver o preenchimento dos metadados dos mesmos segundo o Perfil MGB.

#### **4.4. Avaliação da qualidade dos dados espaciais em relação a acurácia posicional**

A avaliação da acurácia posicional é emblemática, existindo várias pesquisas sobre o assunto. Na presente dissertação o objetivo foi complementar a avaliação da qualidade requerida no PEC, segundo preconiza a ISO e por meio de análise espacial dos resultados através de métodos geoestatísticos. Para isto, uma medida de qualidade foi elaborada antes de executar a avaliação da qualidade posicional, e a mesma está presente no Apêndice A.3. A metodologia aplicada para aferição da acurácia posicional absoluta da BC250 é similar a descrita em relação a completude. Na prática, como a geoestatística pressupõe a continuidade da variável em estudo, cada erro planimétrico inspecionado e calculado foi associado a uma unidade de amostragem de 13km<sup>2</sup>. A amostragem ocorreu de forma aleatória sistemática, onde as quadrículas 1:100.000, utilizados na seleção das unidades de amostragem da completude, delimitaram regiões onde os pontos de campo (GPS) poderiam ser comparados com seu o homólogo na referida base cartográfica.

##### **4.4.1. Identificar a aplicação de elementos de qualidade e o âmbito da avaliação**

Em relação a acurácia posicional absoluta foram considerados todos os elementos do conjunto de dados, principalmente para os de fácil identificação no terreno (entroncamento de vias, passagens de nível e pontes), presentes nas categorias de informação hidrografia e sistema de transporte, conforme recomendado nas especificações técnicas do produto cartográfico e refletida na medida de qualidade correspondente. Entretanto, especialmente a análise foi limitada pela ausência de pontos de controle dentro da região da Amazônia Legal.

##### **4.4.2. Identificar a medida de qualidade**

A medida de qualidade de acurácia posicional absoluta utilizada nesta dissertação foi elaborada com base no PEC, e utiliza os níveis de conformidade apresentados na Tabela 20 e nas especificações técnicas da referida base cartográfica (IBGE, 2011a). O apêndice A.3 apresenta esta medida de qualidade em detalhes, com a descrição de como realizar a avaliação da acurácia posicional absoluta, baseada nas especificações técnicas do produto e o PEC.

TABELA 20 - Níveis de conformidade para a acurácia posicional absoluta

Classe	PEC (metros)	EP (metros)
A	125 m	75 m
B	200 m	125 m
C	250 m	150 m

EP - Erro padrão

#### 4.4.3. Método de inspeção para acurácia posicional absoluta

Para avaliação da acurácia posicional absoluta da BC250 foi definida a inspeção por amostragem aleatória sistemática, devido ao propósito da medida de qualidade e a quantidade de pontos de controle disponíveis. Entretanto, apesar do número significativo de pontos de apoio de campo em algumas áreas de inspeção não foi possível realizar a identificação entre todos os pontos de campo e o seu correspondente no conjunto de dados espaciais. Em distintas situações, o ponto de apoio de campo estava localizado num local ou feição geográfica não representável na escala 1:250.000, impossibilitando identificar seu homólogo.

##### 4.4.3.1. Resultado da inspeção por amostragem da acurácia posicional absoluta

O resultado da avaliação da qualidade da acurácia posicional absoluta, em relação ao erro planimétrico do conjunto de dados, expressa um resultado quantitativo tabular, presente em cada unidade de amostragem. Estes valores foram utilizados para calcular o percentual de erros planimétricos que estavam acima do PEC Classe A para escala de 1:250.000. Para realizar o cálculo, foi utilizado o software livre *LibreOffice Calc*, onde a tabela de atributos da unidade de amostragem da acurácia posicional, no formato dbf (Arquivo dBase), foi editada para este fim. A Tabela 21 sintetiza os quantitativos e percentagens observados.

TABELA 21 - Síntese da classificação da BC250 segundo o PEC

Classe	Intervalo	Percentual de Pontos no Intervalo	Quantidade de Pontos no intervalo
A	0 a 125 metros	84,31%	172
B	0 a 200 metros	94,12%	192
C	0 a 250 metros	95,59%	195

O EMQ calculado foi de 123,63 metros para os 204 pontos observados do conjunto de dados espaciais em estudo. Desta forma, conclui-se que a acurácia posicional absoluta da BC 250 pode ser classificada como Classe B, em conformidade com o PEC para a escala 1:250.000 e o Erro Padrão considerado. Entretanto, uma análise espacial desta inspeção de qualidade identificou regionalizações destas incertezas posicionais, comprometendo a conformidade do conjunto de dados com a Classe A do PEC, além da dependência espacial dos dados observados.

#### 4.5. Análise espacial da qualidade: Geoestatística aplicada na avaliação da acurácia posicional

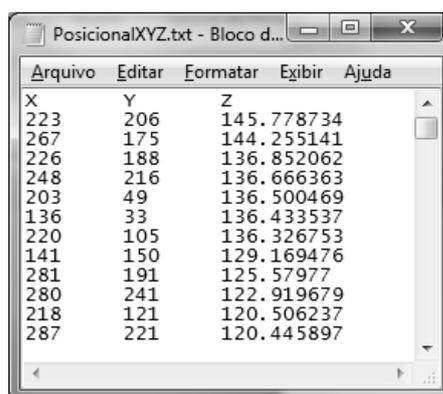
O objetivo nesta etapa foi analisar o comportamento espacial e identificar regiões com boa ou má acurácia posicional, segundo as classes de qualidade do PEC, através de métodos geoestatísticos.

Esta etapa foi realizada utilizando os software R e Tinn-R, com base no fluxo proposto na Figura 58, onde os resultados foram obtidos através de rotinas (*scripts*) elaboradas com este propósito. O software R trabalha por meio das rotinas: *geoR* (geoestatística), *rgdal* (tratamento do resultado em estrutura matricial) *scatterplot3d* (gráficos) e *tcltk2* (auxiliar no ajuste do variograma). No final da análise geoestatística foi utilizado o ambiente SIG para concluir esta etapa.

Durante o processo de análise geoestatística o sistema de referência foi alterado para o sistema de linha e coluna da unidade de amostragem e ao final do processo foi realizado os procedimentos de georreferenciamento da krigagem.

##### 4.5.1. Preparo da análise geoestatística em termos de acurácia posicional absoluta

Os resultados representados na unidade de amostragem considerando a acurácia posicional absoluta foram preparados e separados em um arquivo no formato "txt", composto por uma estrutura de colunas X, Y e Z. Para isto foi utilizado o software livre *LibreOffice Calc*, onde a tabela de atributos da unidade de amostragem, presente no arquivo de formato "dbf", foi editada de forma a conter apenas os campos "COLUNA, LINHA, "ErroLinear"<sup>52</sup>. Todos os registros foram selecionados, copiados e colados no Bloco de Notas para depois serem salvos no formato "txt". A Figura 74 mostra parcialmente os dados contidos no arquivo "PosicionalXYZ.txt" gerado.



X	Y	Z
223	206	145.778734
267	175	144.255141
226	188	136.852062
248	216	136.666363
203	49	136.500469
136	33	136.433537
220	105	136.326753
141	150	129.169476
281	191	125.57977
280	241	122.919679
218	121	120.506237
287	221	120.445897

Figura 74 - Dados observados de acurácia posicional absoluta em formato texto

<sup>52</sup> A ordem de campos deve ser obedecida (X,Y,Z), onde o eixo X é representado pela COLUNA, o eixo Y é representado pela LINHA, e o valor Z em estudo é representado pelo resultado da medida de qualidade.

O arquivo texto gerado foi o dado de entrada na rotina (*script*) em R. Desta forma, os parâmetros necessários para análise geoestatística foram observados, através do resultado de funções, e inseridos por meio de variáveis definidas na rotina. As próximas seções apresentam os resultados da análise geoestatística dos dados.

#### 4.5.2. Análise exploratória dos dados da acurácia posicional absoluta

Conforme mencionado, a metodologia para acurácia posicional absoluta é similar a realizada para comissão e omissão, sendo distinta apenas nos resultados, conforme é demonstrado na Tabela 22 e no histograma dos dados observados ilustrado na Figura 75.

TABELA 22 - Estatística descritiva referente à acurácia posicional absoluta (erro planimétrico) da BC250

ESTATÍSTICA DESCRITIVA	MEDIDA DE QUALIDADE
	M30101
N	204
Média	82.84969 m
Desvio padrão	91.57781 m
Variância dos dados observados	8386.495 m <sup>2</sup>
Máximo	789.600 m
Quartil superior	98.900 m
Mediana	57.840 m
Quartil inferior	33.320 m
Mínimo	3.916 m

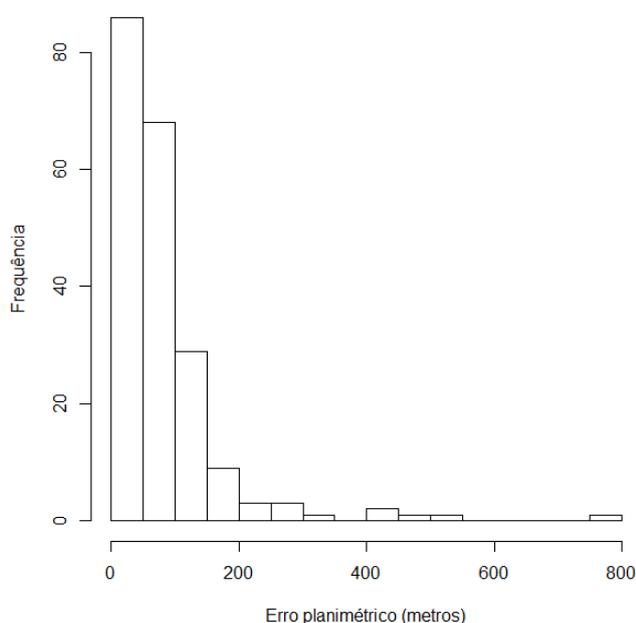


Figura 75 - Histograma referente ao erro planimétrico da BC250.

A Figura 76 mostra a distribuição espacial dos pontos e a densidade dos valores em estudo, gerada através da função *points()* no R. Além dos valores da Tabela 22 é interessante observar as coordenadas extremas dos dados, assim como as distâncias mínimas e máximas entre as amostras, estes valores são apresentados na Tabela 23.

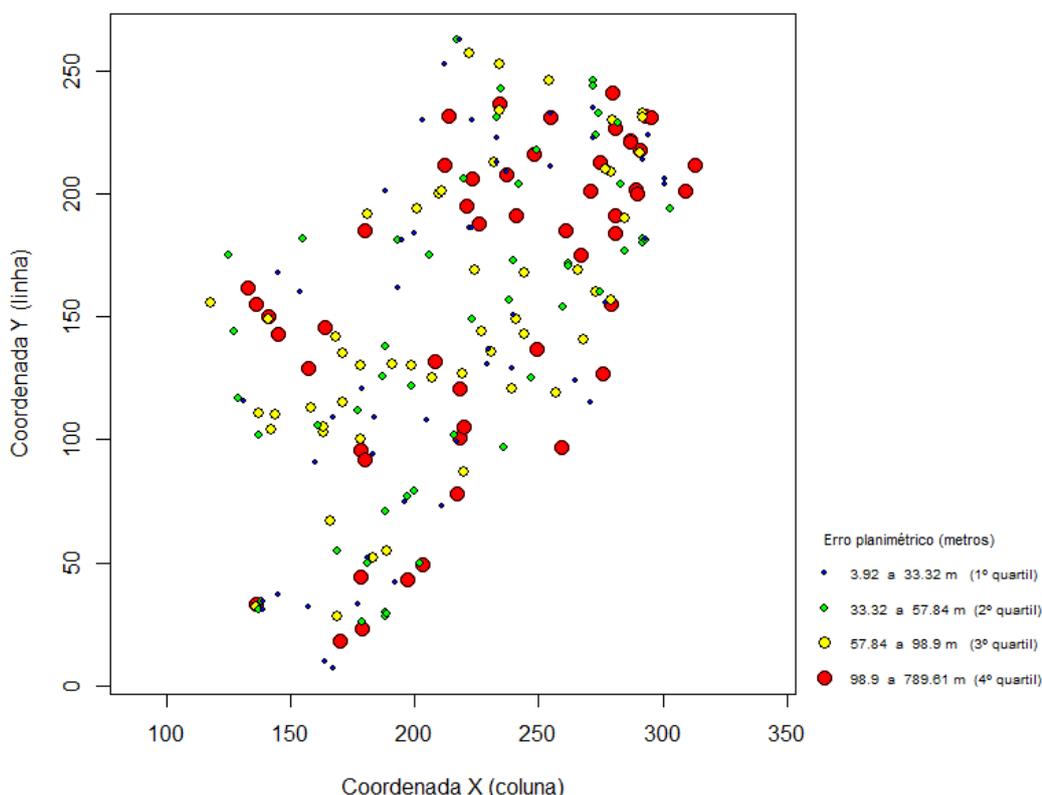


Figura 76 - Distribuição espacial dos dados observados referentes à acurácia posicional absoluta<sup>53</sup>

TABELA 23 - Pontos extremos das unidades de amostras da acurácia posicional

	X	Y	Distância (quadrícula)
<b>Mínima</b>	118	7	1.0000
<b>Máxima</b>	313	263	261.0479

Conforme mencionado, a falta de dados observado na região norte do Brasil (Figura 76) ocorre em função da ausência de pontos de apoio de campo no Projeto PT-CON, o que inviabilizou a análise de resultados nesta região.

Para atender ao PEC, foram realizados o teste de tendência (t de Student) e de normalidade (Shapiro-Wilk), conforme recomendado por Santos (2010a), onde foi detectado a tendência nos dados observados e que os mesmos não seguem a distribuição normal ( $p\text{-valor} = 2.2e-16 < 0,05$ , para o nível de confiança de 90%). No próximo tópico esta conclusão é comprovada pela Geoestatística, pois foi identificada a dependência espacial dos dados observados.

<sup>53</sup> Cada intervalo de linha e coluna possui aproximadamente uma distância de 13km.

### 4.5.3. Análise variográfica dos dados da acurácia posicional absoluta

Foi realizada a análise variográfica para os casos de incerteza posicional. A distância máxima (*lag*) foi de 70 quadrículas, dando origem ao variograma experimental ilustrado na Figura 77.

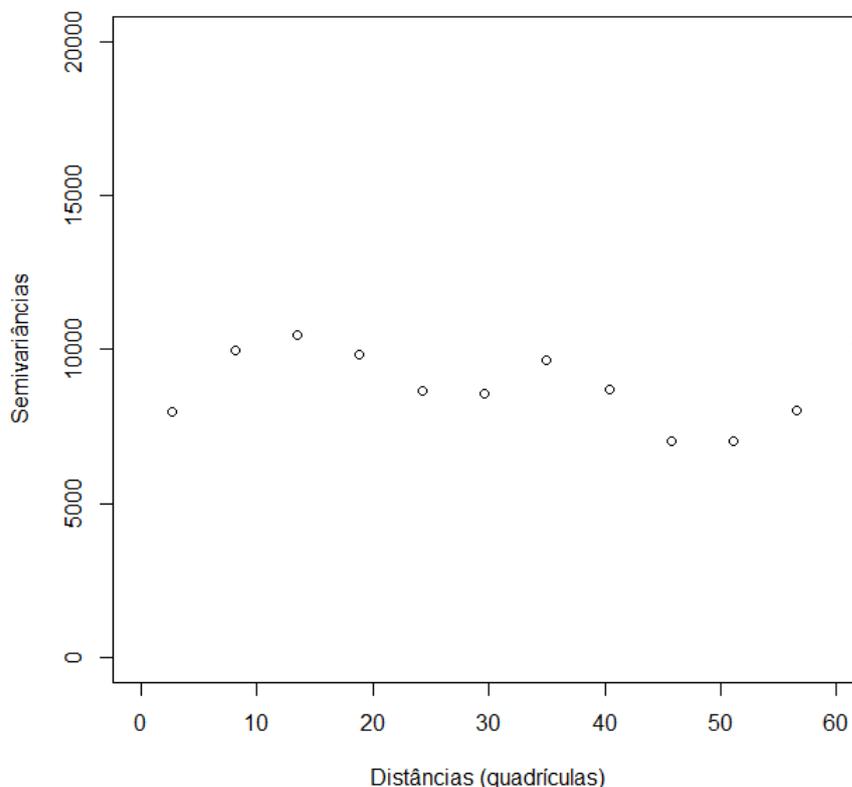


Figura 77 - Variograma experimental dos dados referentes ao erro planimétrico

Foi observado uma pequena dependência espacial indicando o uso da Estatística Espacial, visto que um preceito da Estatística Clássica foi quebrado, ou seja, o de independência dos dados observados. A estrutura imbricada indica que ocorre regionalizações com variabilidade próximas em áreas distintas do mapa, ou seja, existe correlação em distâncias próximas e esta dependência espacial volta a ser observada quando forma-se pares mais distantes.

Na análise anisotrópica foram observadas variações mínimas, e os dados foram considerados isotrópicos. Para a escolha do melhor ajuste do variograma teórico ao variograma experimental, foram considerados três modelos teóricos: esférico, exponencial e gaussiano. Na sequência os parâmetros de cada um foram alterados até atingir o melhor ajuste ao variograma experimental, sendo obtido pelo modelo esférico, conforme mostra a Tabela 24 com os respectivos parâmetros do variograma ajustado, ilustrado na Figura 78.

TABELA 24 - Parâmetros do ajuste do variograma teórico ao experimental para acurácia posicional absoluta

<b>MEDIDA DE QUALIDADE PARA ACURÁCIA POSICIONAL ABSOLUTA (M30101)</b>	
Modelo teórico	Esférico
Efeito pepita ( $C_0$ )	6100
Patamar ( $C_0+C$ )	9000
Variância espacial ( $C$ )	2900
Variância dos dados observados	8386.495
Alcance ( $a$ )	10
Grau de aleatoriedade ( $E = C_0 / C$ )	0.6777778

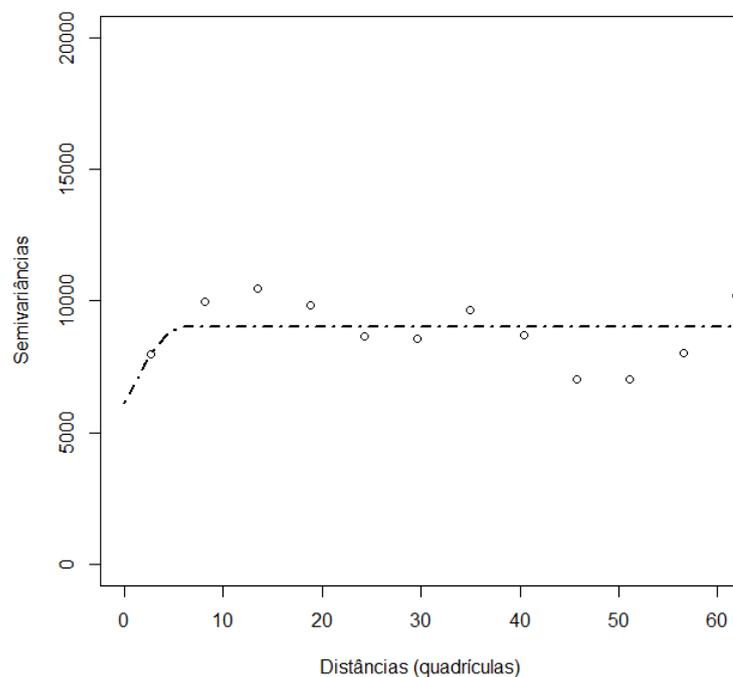


Figura 78 - Ajuste do variograma teórico ao experimental para o erro planimétrico.

Segundo os dados presentes na Tabela 24, após o ajuste do variograma, o grau de aleatoriedade é "bastante importante", em função do valor considerável do efeito pepita, ocasionado pela escala da grade regular contínua, a quantidade de pontos observados e das descontinuidade do fenômeno em estudo, com contribuição espacial relativamente baixa, mas mostrando dependência espacial para pontos amostrados em agrupamentos menores. Em relação ao comportamento do variograma, este indica a existência de regionalizações em áreas distintas, mas com dependência espacial semelhantes. Nota-se correlações espaciais distintas e com pouco alcance, porém um bom ajuste aos pontos iniciais garante uma boa interpolação através da krigagem.

A Tabela 25 mostra os resultados da validação cruzada em relação ao ajuste do variograma teórico ao experimental, referente a acurácia posicional absoluta do conjunto de dados em estudo. Através da regressão linear é possível observar a correlação entre os valores observados e o seus valores estimados. O valor de  $\beta_1$  foi de -0.4397 e  $\beta_0$  foi de 119.2699, indicando não houve um bom ajuste do variograma teórico ao experimental, conforme é perceptível na Figura 78. Entretanto, este fato não invalida o ajuste realizado, visto que a correlação espacial foi identificada.

TABELA 25 - Resultados da autovalidação do ajuste do variograma teórico ao experimental referente ao erro planimétrico da BC250

Parâmetros da Autovalidação	Acurácia Posicional Absoluta
$\beta_1$	- 0.4397
$\beta_0$	119.2699

#### 4.5.4. Krigagem dos dados da acurácia posicional absoluta

Foi realizada a krigagem, a fim de estimar os valores das quadrículas não inspecionadas, ressaltando que antes foi necessário definir uma grade para inserir os valores estimados. O detalhe esta nas dimensões desta grade que deve ser a mesma levando em consideração as extremidades das quadrículas inspecionadas, conforme descrição da Tabela 23.

#### 4.5.5. Mapa de qualidade da acurácia posicional absoluta

Após o georreferenciamento da krigagem, foi realizado um editoração cartográfica do mapa de qualidade referente a acurácia posicional. No caso, além dos elementos de referência dos mapas de qualidade foi acrescentada a delimitação da Amazônia Legal para destacar a área onde não existem pontos de controle de campo para aferir a qualidade posicional do conjunto de dados em estudo. Esta etapa concluiu o processo de avaliação da qualidade e gerou sua representação espacial, visualizada através do mapa de qualidade, conforme Figura 79.

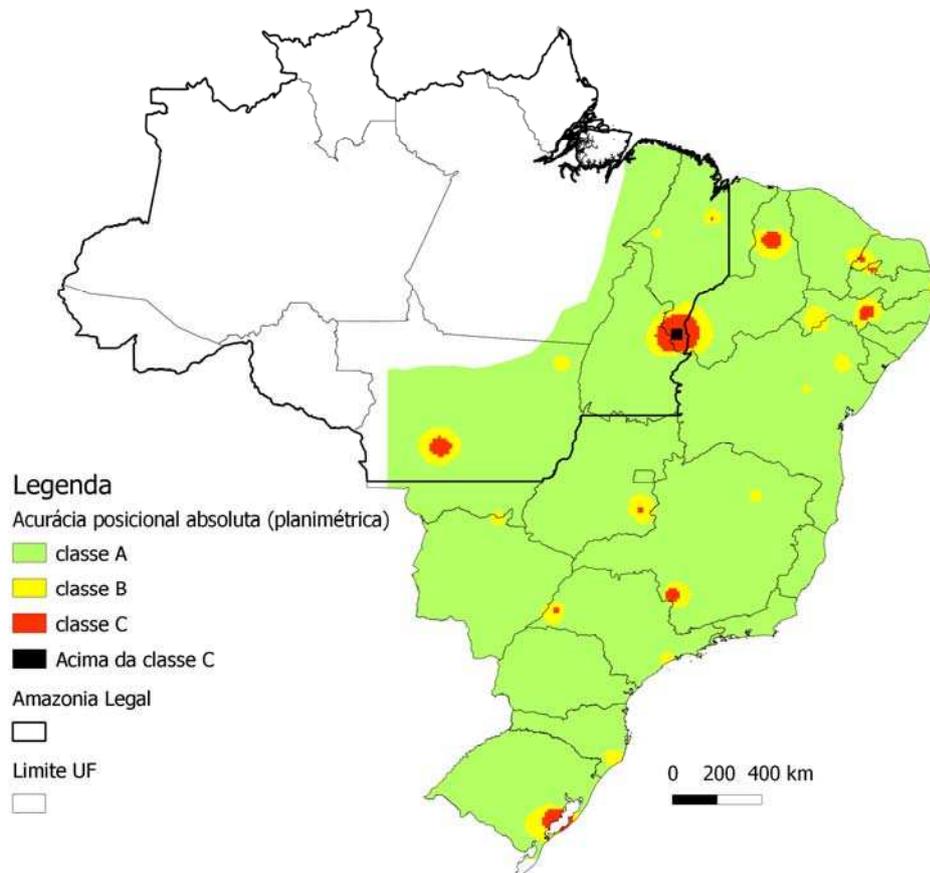


Figura 79 - Mapa de qualidade da acurácia posicional absoluta da BC250

#### ***Análise dos resultados do mapa de qualidade de acurácia posicional absoluta***

Na análise espacial do mapa de qualidade, nota-se a concentração de regiões com o erro planimétrico elevado próximos a delimitação da Amazônia Legal, indicando possíveis erros sistemáticos, segundo relato da equipe produtora da BC250, são locais de difícil acesso como o Pantanal Mato-grossense, ou áreas com excesso de nuvens, o que ocasionou dificuldades na coleta e identificação de pontos para ortorretificar as imagens orbitais do projeto.

#### **4.6. Metadados da avaliação da qualidade da acurácia posicional**

Neste tópico é apresentado o preenchimento dos metadados, segundo a norma ISO 19115:200, baseado nos resultados das inspeções realizadas referentes a medidas de qualidade de acurácia posicional absoluta do conjunto de dados espaciais em análise, no caso a BC250.

##### **4.6.1. Metadados de qualidade: acurácia posicional absoluta da BC250**

A próxima etapa foi relatar os resultados e métodos utilizados durante o processo de aferição. Em relação a definição do nível de conformidade da acurácia posicional o parâmetro adotado foi o PEC, através da análise percentual simples e o

EMQ. Em relação as demais seções dos metadados de qualidade, a Tabela 26 apresenta os metadados de qualidade relatados, conforme orienta as normas ISO 19114:2003 e ISO 19115:2003.

TABELA 26 - Metadados de qualidade de dados referentes à acurácia posicional absoluta da BC250

<b>Componente DQ (Data Quality)</b>	<b>Descrição / Exemplos</b>
<i>Data Quality Scope</i> (DQ_Scope)	Todos os elementos da base cartográfica
<i>Data Quality Element</i> (DQ_Element)	3 - Acurácia Posicional
<i>Data Quality Measure</i> (DQ_Measure)	01 - Acurácia Posicional Absoluta
<i>Data Quality Measure Description</i> (DQ_MeasureDesc)	Classificar a Base Cartográfica segundo o PEC
<i>Data Quality Measure Identification</i> (DQ_MeasureID)	30101
<i>Data Quality Evaluation Method</i> (DQ_EvalMethod)	
<i>Data Quality Evaluation Method Type</i> (DQ_EvalMethodType)	2 - Externo
<i>Data Quality Evaluation Method Description</i> (DQ_EvalMethodDesc)	Identificar um a localização representado no conjunto de dados e correlacionar com os pontos de controle coletados em campo, considerados como verdadeiros, e medir o erro planimétrico entre as duas posição observadas para posterior classificação segundo o PEC.
<i>Data Quality Result</i> (DQ_QualityResult)	
<i>Data Quality Value Type</i> (DQ_ValueType)	4 - Porcentagem
<i>Data Quality Value</i> (DQ_Value)	-
<i>Data Quality Value Unit</i> (DQ_ValueUnit)	%
<i>Data Quality Date</i> (DQ_Date)	2013-02-17 a 2013-04-26
<i>Data Quality Conformance Level</i> (DQ_ConformanceLevel)	Classe B. Num universo de 204 unidades de amostragem, 32 estavam acima do PEC Classe A, ou seja, 84,31% dos pontos estão abaixo de 125 metros no caso da escala 1:250.000. Em relação a Classe B, 12 pontos estão acima do PEC Classe B, ou seja, 94,11% dos pontos medidos estão abaixo de 200 metros de erro. Em relação a Classe C, 9 pontos estão acima do PEC Classe C, ou seja, 95,58% dos pontos medidos estão abaixo de 250 metros de erro.

Também para o caso do preenchimento dos metadados de qualidade quanto a posicional absoluta da BC250, é recomendada a geração da imagem referente ao mapa de qualidade.

#### 4.7. Considerações gerais

A dependência espacial identificada pela Geoestatística é identificada pela Estatística Clássica através do teste de normalidade. Entretanto, há um quebra de pressuposto da Estatística Clássica, visto que existe dependência dos dados, indicando o uso da Estatística Espacial.

Toda variabilidade de um fenômeno, condicionada ao longo de um espaço mensurável e contínuo, seja este temporal ou dimensional, pode ser analisada pela

Geoestatística. Logo, a aplicação dos métodos geoestatísticos pode ser estendida à diversas áreas de Cartografia e Geoprocessamento, como Maregrafia, Geodésia e Sensoriamento Remoto.

Em relação à metodologia proposta, sua aplicação pode ser utilizada tanto na avaliação de um produto cartográfico final, em meio digital ou analógico, como no estudo preliminar de identificação de áreas de interesse de atualização e/ou melhoria da qualidade de um conjunto de dados espaciais específico. O mapa de qualidade pode auxiliar na orientação de um processo de produção de produtos cartográficos, bem como na identificação de regiões com maior demanda de atualização e/ou de possíveis erros sistemáticos. A qualidade de dados geoespaciais não é um fenômeno natural, mas a Geoestatística pode ser utilizada para analisar o comportamento espacial dos elementos de qualidade, onde a inspeção por amostras pode ser associada a uma grade regular contínua para fins geoestatísticos.

A Geoestatística apresenta uma representatividade espacial da variabilidade da qualidade de um produto cartográfico (mapa de qualidade) e a acurácia desta estimativa. A definição adequada do método de amostragem possibilita a avaliação dos elementos de qualidade de dados espaciais e a análise espacial desta qualidade, através dos métodos geoestatísticos.

Mesmo para os casos onde a Geoestatística apontar a ausência de correlação espacial da variável de qualidade analisada, ou seja, onde ocorra um efeito pepita puro na análise variográfica, o método valida o uso da Estatística Clássica, que considera a aleatoriedade do fenômeno em estudo. Nestes casos, uma conclusão meramente descritiva e/ou tabular representará a inspeção da qualidade do conjunto de dados. Na prática, a regionalização da qualidade indica prováveis erros sistemáticos que podem ser corrigidos, e a qualidade de dados geoespaciais deseja que estes erros sejam aleatórios. Desta forma, reafirma-se a importância de um estudo geoestatístico para verificar a correlação espacial dos resultados da inspeção de qualidade.

Ressalta-se que outros métodos devem ser estudados e aplicados na representação e identificação da correlação espacial nos resultados da inspeção completa. A visualização da qualidade do conjunto de dados espaciais através dos mapas de qualidade facilitam o entendimento e adequação ao uso por parte dos usuários.

O órgão responsável pelo produto cartográfico deve ser responsável por sua qualidade. Numa hierarquia de responsabilidades, existem órgãos competentes responsáveis por determinadas camadas de informações representadas em distintos universos de discussão. Esta hierarquia deve ser respeitada, antes de realizar uma

atualização da base cartográfica, para garantir a qualidade fornecida pelo órgão responsável e evitar a duplicação de esforços. Num plano de trabalho ideal, diferentes órgãos seriam responsáveis por suas respectivas camadas de informação numa Infraestrutura de Dados Espaciais. A equipe produtora deve identificar as medidas de qualidade necessárias para aferição e melhoria da qualidade do conjunto de dados de sua responsabilidade.

As normas referentes à qualidade de dados espaciais orientam a aplicação, mas atualmente não existe uma certificação de qualidade quando uma determinada instituição se utiliza das mesmas, com exceção, apenas no caso da ISO 9001 e ISO 14000, conforme informa o Inmetro - Instituto Nacional de Metrologia, Qualidade e Tecnologia<sup>54</sup>.

A documentação produzida pode ser utilizada como referência no controle de qualidade de produtos cartográficos e atende ao objetivo de aplicação direta na produção cartográfica da Coordenação de Cartografia do IBGE e até mesmo outras áreas da Diretoria de Geociências, agregando valor aos dados disponibilizados na INDE.

---

<sup>54</sup> Mais informações em < <http://www.inmetro.gov.br>>

## 5. CONCLUSÕES E RECOMENDAÇÕES

### 5.1. Conclusões

No caso da inspeção por amostragem da base cartográfica BC250, notou-se empiricamente que as quadrículas de 13km<sup>2</sup> atenderam ao propósito da pesquisa na avaliação da completude da categoria de informação hidrografia. Entretanto, estas poderiam ser menores, tornando mais eficiente o processo de inspeção por unidade de amostragem, segundo o elemento de qualidade em estudo. Ressalta-se que, estas dimensões devem adequar-se ao propósito da medida de qualidade e ao fenômeno espacial em estudo. Em relação à distribuição espacial das amostras, é recomendada a amostragem aleatória sistemática, uma vez que esta reproduz melhor a distribuição e variabilidade espacial da variável de interesse.

Os testes da metodologia aqui proposta, por meio da avaliação da completude e acurácia posicional do conjunto de dados da BC250, mostrou-se eficiente, apresentando especificidades representadas nas “manchas” dos respectivos mapas de qualidade, confirmadas pela equipe produtora da referida base cartográfica, fornecendo subsídios para uma próxima atualização.

O entendimento dos elementos de qualidade e dos operadores de generalização cartográfica, confundem-se com a própria produção cartográfica, e auxiliam na identificação, elaboração e aplicação de medidas de qualidade. Desta forma, um catálogo de medidas de qualidade pode ser criado e a presente metodologia pode ser aplicada a cada nova aferição de qualidade identificada. Este catálogo mínimo de medidas de qualidade pode ser elaborado pelo órgão produtor do conjunto de dados espaciais. Nesta dissertação foram elaboradas 3 (três) medidas de qualidade, baseadas na norma ISO 19114:2003, mas outras podem ser criadas e desdobradas para as demais escalas do mapeamento sistemático e produtos da cartografia de uma maneira geral. Logo, o estudo referente a amostragem, citados no parágrafo anterior, e a elaboração do catálogo de medidas de qualidade, podem proporcionar um padrão de itens mínimos a serem aferidos no mapeamento sistemático independentes de suas escalas de representação. A acurácia de um mapeamento está vinculada ao terreno nominal (especificações técnicas) ao qual o conjunto de dados deve ser comparado e no mapeamento sistemático o modelo de dados é o mesmo, só variando a escala de representação.

Em diversas situações onde ocorre má qualidade do conjunto de dados espaciais é perceptível, não só a incompatibilidade com a respectiva especificação técnica do

produto, mas também o mau uso dos operadores de generalização. Desta forma, ao se avaliar a qualidade de um conjunto de dados espaciais e até mesmo apontar possíveis soluções de melhoria, é necessário ter um conhecimento geral dos conceitos de generalização.

Conforme mencionado, a representação espacial do resultado da inspeção da qualidade do produto cartográfico é essencial e os métodos geoestatísticos atendem a esta proposta e propiciam a identificação da correlação espacial.

As classes de ocorrências e de amostras correspondentes às respectivas medidas de qualidade, no caso específico das inspeções de qualidade da completude e do agrupamento de feições de hidrografia, permitem avaliações individualizadas para cada classe de feição geográfica ou até mesmo a análise da densidade de feições por unidade de amostragem. Ressalta-se que as classes de ocorrência e de amostras visam facilitar a contagem e/ou mensuração da variável em estudo, e facilitar sua localização e sua correção ou adequação posteriormente.

Assim como na evolução da cartografia analógica para a digital, as informações referentes à qualidade de um conjunto de dados espaciais, não estão mais restritas as dimensões de um papel. O usuário carece desta informação e a tecnologia, apesar de facilitar e agilizar o processo de produção cartográfica, fez com que o controle de qualidade do conjunto de dados geoespaciais fosse de certa forma omitido. Assim, a identificação e o detalhamento das fases de produção e controle de qualidade do produto cartográfico, certificada segundo normas de uso internacional, atendem à demanda da INDE e permitem o acesso à informação pela sociedade.

Os metadados de qualidade, referentes à completude e à acurácia posicional, podem ser disponibilizados junto com o conjunto de dados da BC250, de preferência com os respectivos mapas de qualidade. Na prática, os mapas de qualidade gerados nesta dissertação são produtos cartográficos, logo podem ser descritos segundo o Perfil MGB.

O padrão de metadados faz com que o produtor documente e reflita sobre as características do seu produto. As normas de qualidade fazem com que o produtor analise o seu processo de produção e atente para as demandas do usuário. A adoção de normas internacionais em Infraestruturas de Dados Espaciais propicia credibilidade aos dados, viabiliza a interoperabilidade e o uso das informações nela representadas. No Brasil, assim como em outros países, existem comitês técnicos analisando e estudando

padrões e normas a serem aplicadas em suas infraestruturas de dados espaciais<sup>55</sup>. Especialistas em produção de dados espaciais, e áreas afins, devem estabelecer padrões e normas para harmonizar e compatibilizar o uso e acesso aos dados. A CONCAR, mais especificamente, a Diretoria de Serviço Geográfico do Exército - DSG, está desenvolvendo normas para avaliação da qualidade de produtos cartográficos vetoriais na escala de 1:250.000 e maiores e o IBGE para as escalas menores que 1:250.000, bem como de outros produtos cartográficos vinculados a sua produção<sup>56</sup>.

O importante ao se avaliar um conjunto de dados espaciais é ter em mente a ideia do universo de discussão e não do mundo real em si, pois o excesso de informações dificulta o processo, podendo gerar avaliações inconsistentes sobre o conjunto de dados. Por isto, as especificações técnicas do produto, junto com seu modelo de dados espaciais, são fundamentais. A acurácia de um produto cartográfico é vinculada a sua especificação, e desta forma uma documentação técnica detalhada, até mesmo em termos de execução da produção cartográfica, deve ser desenvolvida, a fim de evitar desperdício de esforços, tanto na produção quanto na avaliação do produto final. Isto induz ao desenvolvimento de ferramentas e metodologias e aplicação de conceitos de controle de qualidade no decorrer da produção cartográfica. Além disso, as estruturas de hardware, software e insumos devem ser adequadas ao propósito do projeto de produção, sem divergências ou falta de homogeneidade, para auxiliar e facilitar o processo de atualização e garantir a qualidade do conjunto final de dados espaciais.

Para a acurácia posicional de produtos cartográficos, o PEC parte do pressuposto que os dados observados para avaliação do conjunto de dados espaciais, no caso desta dissertação o erro planimétrico, sejam independentes. Entretanto, como observado nesta pesquisa, pode existir a dependência espacial e mesmo que seja mínima, ela faz quebrar um pressuposto da Estatística Clássica. A Estatística Espacial deve ser utilizada para analisar os dados.

---

<sup>55</sup> Comitê de Estruturação da Mapoteca Nacional Digital (CEMND), Comitê de Estruturação de Metadados Geoespaciais (CEMG), Comitê de Implantação da Infraestrutura Nacional e Dados Espaciais (CINDE), Comitê de Nomes Geográficos (CNGEO), Comitê de Normatização do Mapeamento Cadastral (CNMC). Mais informações na página da CONCAR <<http://www.concar.ibge.gov.br>>

<sup>56</sup> Segundo Decreto Lei 243 de 28 de fevereiro de 1967 é de competência da DSG o estabelecimento de Normas Técnicas para a cartografia brasileira na série de cartas gerais, das escalas 1:250.000 e maiores; e do IBGE a competência de estabelecimento de Normas Técnicas para a série de cartas gerais, das escalas de 1:1.000.000 e menores. Nessas escalas constam produtos cartográficos como: Atlas Nacional, Atlas Escolares e Mapas do Brasil, Regionais, Estaduais e Municipais para levantamentos estatísticos.

Por fim, conclui-se que a avaliação da qualidade de dados espaciais, com base nas orientações da norma ISO 19114:2003, foi eficiente e representada nas unidades de amostragem de cada medida de qualidade realizada. A Geoestatística mostrou-se adequada na análise espacial quanto a aferição de qualidade, identificando a correlação espacial ou a aleatoriedade dos resultados da inspeção por amostragem. A técnica de amostragem aleatória sistemática, baseada nas dimensões das articulações do mapeamento sistemático, tanto na escala 1:100.000 quanto na escala 1:25.000, se mostrou eficiente, apresentando uma distribuição espacial apropriada para a base cartográfica em estudo. Entretanto, o controle do processo de produção e a avaliação dos elementos de qualidade durante a elaboração da base cartográfica pode ser estudado e até mesmo modelado, e medidas de qualidade podem ser elaboradas para aferir a qualidade durante este processo de produção cartográfica. Porém, a inspeção do produto cartográfico no final deste processo de produção garante a confiabilidade do mesmo à sociedade.

## **5.2. Recomendações**

No processo de aferição da qualidade de um conjunto de dados é óbvio, ao final, sentenciar sua conformidade ou não com uma determinada especificação ou parâmetro. Todavia, para distintos elementos de qualidade é necessário definir estes parâmetros para comparação e definir um nível de conformidade, assim como existe o PEC para a acurácia posicional. Nota-se uma ausência de parâmetros para a cartografia digital e as conclusões de diferentes avaliações de qualidade tornam-se vagas por falta desta variável, impossibilitando ancorar os dados a sua realidade, pois até que ponto determinado fenômeno de qualidade em estudo é significativo ou não. Além de uma consistente e detalhada especificação técnica do produto cartográfico, um ambiente de hardware e softwares homogêneos auxiliam na produção de dados de qualidade. Alterações no decorrer do processo de produção podem acarretar na variação da qualidade do produto final em distintas regiões. Da mesma forma, um treinamento adequado com os técnicos responsáveis pela execução da atualização pode evitar regionalizações de qualidade, criados por estes operadores.

Como já dito, as ocorrências em distintos elementos de qualidade podem afetar outros elementos, de forma similar aos operadores de generalização, e este vínculo é perceptível inclusive na comparação de produtos de distintas escalas do mapeamento sistemático. Desta forma, recomenda-se um estudo geoestatístico sobre a correlação espacial destes fenômenos, com o objetivo de garantir a qualidade entre escalas, operadores de generalização e elementos de qualidade, por meio da cokrigagem.

Ressalta-se o uso eficiente dos softwares livres de geoprocessamento e análise geoestatística aplicados nesta dissertação, os quais podem proporcionar e facilitar a reprodução desta pesquisa. As respectivas listas de discussão, mostram-se solistas na resolução dos problemas apresentados. Nos ambientes SIG (Quantum GIS e gvSIG), destaca-se a presença de módulos referentes a análise geoestatística e a integração com o ambiente R, possibilitando a execução de rotinas (*scripts*) desenvolvidas fora da aplicação. Desta forma, recomenda-se o estudo de sua aplicação e avaliação de seus resultados.

No caso específico da acurácia posicional para escala 1:250.000, em relação ao uso de técnicas de amostragem e coletas de pontos de controle em campo, é indicada a utilização de uma grade regular de busca, análoga à articulação 1:100.000, com dimensões superiores às quadrículas necessárias para análise geoestatística, similar à articulação 1:25.000. Na prática, fatores como difícil acessibilidade de algumas regiões, ou a presença de nuvens nas imagens analisadas, entre outros, dificultam o levantamento e/ou a identificação de pontos de controle dos projetos. Assim, uma área de varredura com dimensões maiores facilita a coleta, escolha e identificação dos pontos de apoio para mensurar o erro planimétrico e realizar a análise geoestatística. Recomenda-se um estudo sobre o assunto e assim como no caso da definição da área das unidades de amostragem, devem ser consideradas inicialmente suas dimensões na escala do produto para posterior associação com o tamanho real da área de varredura.

Recomenda-se para trabalhos futuros, um estudo em relação ao tamanho da amostra, sua distribuição espacial e a dimensão das unidades de amostragem, considerando a escala nominal do produto cartográfico (especificações técnicas), o elemento de qualidade ou fenômeno espacial a aferir, o respectivo erro amostral tolerável desejável e os fatores financeiros. Um outro fator a considerar neste estudo é analisar a influência da escala, em relação as dimensões das células da grade regular contínua para fins geoestatísticos, sobre o efeito pepita, ou seja, o quanto esta variação de escala das células pode afetar na descontinuidade na origem do variograma.

Particularmente, em relação a acurácia posicional do conjunto de dados espaciais, demanda-se encontrar um denominador comum entre a quantidade de pontos de controle a observar no campo, sua distribuição espacial e até onde é necessário densificar a coleta na área de estudo, em função da identificação da dependência espacial pela Geoestatística. Ressalta-se que, caso os dados sejam coletados num intervalo superior ao alcance da dependência espacial dos mesmos, a Geoestatística não identificará a dependência espacial e a Estatística Clássica poderá ser utilizada. Desta

forma, recomenda-se um estudo sobre a forma de coletar os pontos de controle no campo, tornando eficiente a relação entre sua distribuição espacial, a quantidade de pontos a observar e o uso de métodos geoestatísticos para diminuir custos e possibilitar uma acuracidade maior na avaliação do produto cartográfico, aliando a Estatística Clássica à Geoestatística, de forma a atender as especificações do PEC. A revisão da forma de utilizar o PEC, considerando este aspecto de dependência espacial, pode ser relevante para tornar a avaliação da qualidade do conjunto de dados espaciais mais acurada.

Além disso, recomenda-se a elaboração de outras medidas de qualidade, como as observadas durante a inspeção dos elementos de qualidade, dentre as quais cita-se: a avaliação dos Nomes Geográficos, tanto em relação à completude (omissão e comissão de toponímias) como em termos da correta grafia do mesmo; a correta classificação das feições, o posicionamento relativo adequado de elementos pontuais e a densidade de vértices relativos a elementos com geometria tipo linha e área, como trechos de drenagem e rodoviários.

Ressalta-se que, nesta dissertação, a inspeção foi realizada por amostragem para aplicação da Geoestatística. Todavia a inspeção completa e sua representação espacial podem constituir trabalhos futuros. Por mais que os processos de inspeção completa possam ser automatizados, é relevante uma análise espacial desta variação da qualidade para evitar erros sistemáticos em produtos futuros. Classes de ocorrências para inspeção completa podem ser criadas e a análise através da Estatística Espacial pode ser aplicada.

Recomenda-se que, numa próxima revisão do Perfil MGB, as seção referentes a avaliação dos elementos de qualidade sejam incluídas, para permitir o relato das aferições nos metadados do produto disponibilizados ao usuário, no caso da INDE à sociedade.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ANDRIOTTI, J. L. S. **Fundamentos de Estatística e Geoestatística**. Ed. UNISINOS, São Leopoldo, 165p. 2003.
- ARIZA, F. J. **Calidad en La Producción Cartográfica**. Editora RA-MA. Madrid, 2002.
- ARIZA, F. J. **Casos prácticos de calidad em La producción cartográfica**. Universidade de Jaén. Editora RA-MA. Madrid, 2004.
- ARIZA, F. J.; PASCUAL, A. F. R. **Introducción a la normalización en información geográfica: la familia ISO 19100**. Grupo de Investigación en Ingeniería Cartográfica. Universidade de Jaén, Espanha, Jaén, 2008.
- AZEVEDO, J. B. **Algoritmos para a Simplificação da Representação Cartográfica**. Dissertação de Mestrado. UERJ, 2003.
- BARBETTA, P. A. **Estatística aplicada às Ciências Sociais**. 8. ed. - Florianópolis: E. da UFSC, 2012. 318p
- BILL, C. Executive Order 12906: Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure, signed by President Bill Clinton on April 11, 1994. **The U.S. National Archives and Records Administration**. Disponível em <<http://www.archives.gov/federal-register/executive-orders/pdf/12906.pdf>>, Acesso em 30 de maio de 2012.
- BORGES, K. A. V. **Modelagem de Dados Geográficos: uma extensão do modelo OMT para aplicações a geográficas**. Belo Horizonte: Fundação João Pinheiro. Dissertação de Mestrado, Escola do Governo, 1997.
- BRASIL, Decreto n.º 89.817, de 20 de junho de 1984. Dispõe sobre as instruções reguladoras das normas técnicas da cartografia nacional. **Diário Oficial** [da República Federativa do Brasil], Poder Executivo, Brasília, 22 de junho de 1984.
- BRASIL, Decreto n.º 5.334, de 06 de janeiro de 2005. Dá nova redação ao art. 21 e revoga o art. 22 do Decreto n.º 89.817, de 20 de junho de 1984, que estabelece as Instruções Reguladoras das Normas Técnicas da Cartografia Nacional. **Diário Oficial** [da República Federativa do Brasil], Poder Executivo, Brasília, DF, 6 de janeiro de 2005.
- BRASIL, Decreto n.º 6.666, de 27 de novembro de 2008. Institui, no âmbito do Poder Executivo Federal, a Infraestrutura Nacional de Dados Espaciais – INDE. **Diário Oficial** [da República Federativa do Brasil], Poder Executivo, Brasília, DF, 28 de novembro de 2008, Seção 1, p. 57.

- BRASSEL, K.; BUCHER, F.; STEPHAN, E.-M.; VCKOVSKI, A. **Completeness**. Elements of spatial data quality[Ed] Stephen C. Gruptill, International Cartographic. Cap. 5 p. 81-108.
- BURITY, E. F.; SILVEIRA, J. C. **Requisitos de qualidade no contexto da série de normas ISO 9000**. Disponível em <[http://www.cartografia.org.br/xxi\\_cbc/211-C47.pdf](http://www.cartografia.org.br/xxi_cbc/211-C47.pdf)>. Acesso em 10 de março de 2012.
- CAMARGO, E. C. G. **Geoestatística: Fundamentos e Aplicações**. In: Geoprocessamento para Projetos Ambientais. INPE, 1998. Disponível em: <[http://www.dpi.inpe.br/gilberto/tutoriais/gis\\_ambiente/](http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/)>. Acesso em: 20 de out. de 2011.
- CASANOVA, M. A.; CÂMARA, G.; DAVIS JR, C. A.; VINHAS, L.; QUEIROZ, G. R. **Banco de Dados Geográficos**. MundoGEO, Curitiba, 2005.
- CEBALLOS, J. A. R.; GATICA, G. L. R. **Implementation of a Management and Quality Control System Under ISO STANDARDS 9001:2000, 19113, 19114, 19138 and 19115 in Cartographic Production**. Disponível em <[http://icaci.org/documents/ICC\\_proceedings/ICC2009/html/refer/3\\_2.pdf](http://icaci.org/documents/ICC_proceedings/ICC2009/html/refer/3_2.pdf)>. Acesso em 2 de maio de 2011.
- CLARK, D. M. **Lineage** In: GUPTILL, S. C.; MORRISON, J. L., Elements of Spatial Data Quality, Elsevier Science, 1995. Cap. 2, p. 13-30.
- CONCAR, Comissão Nacional de Cartografia. **Especificação Técnica para a Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV)**. CONCAR, 2008.
- CONCAR, Comissão Nacional de Cartografia. **Perfil de Metadados Geoespaciais do Brasil – Perfil MGB**. CONCAR, 2009.
- CONCAR, Comissão Nacional de Cartografia. **Plano de Ação para Implantação da INDE – Infraestrutura Nacional de Dados Espaciais**. CONCAR, 2010.
- CRAGLIA, M; ANNONI, A.; MASSER, I. **Geographic Information Policies in Europe: National and Regional Perspectives**. EUROGI-EC Data Policy Workshop, Amersfoort 15th November 1999. European Commission - Space Applications Institute, European Communities. Disponível em <<http://www.ecgis.org/reports/policies.pdf>>, Acesso em 15 de maio de 2012.
- DALMOLIN, Q.; LEAL, E. M. **Análise da qualidade posicional em bases cartográficas geradas em CAD**. Boletim de Ciências Geodésicas, v. 7, nº 1. UFPR. Curitiba, 2001.

- DEVILLERS, R.; JEANSOULIN, R. **Fundamentals of spatial data quality**. Editora ISTE. Londres, 2006.
- DRUMMOND, J., **Positional Accuracy** In: GUPTILL, S. C.; MORRISON, J. L., Elements of Spatial Data Quality, Elsevier Science, 1995. Cap. 3, p. 31-58.
- EGENHOFER, M. J.; FRANK, A. U.; JACKSON, J. P. **A topological data model for spatial databases**. In: SYMPOSIUM DESIGN AND IMPLEMENTATION OF LARGE SPATIAL DATA BASES. 1989. pp.271-286.
- FERNANDES, W. S.; RIBEIRO, J. A. **Realização do processo de editoração cartográfica utilizando aplicativos livres de geoprocessamento**. Anais do XV Simpósio Brasileiro de Sensoriamento Remoto - SBSR, de 30 de abril a 5 de maio de 2011. Curitiba, PR. 2011.
- FERREIRA, D. F. **Estatística Básica**. 2ª Edição Revisada. Editora UFLA. Lavras, 2009.
- FGDC, Federal Geographic Data Committee. **Content standard for digital geospatial metadata**. Washington: D.C.: Federal Geographic Data Committee, 1998.
- FREITAS, A. L. B. de. **Catálogo de Metadados de Dados Cartográficos como suporte a implementação de Clearinghouse Nacional**. Dissertação de Mestrado. Rio de Janeiro: Instituto Militar de Engenharia, 2005.
- FREITAS, A. L. B.; SANTOS, A. S.; AZEVEDO, H. G.; NAGATOMI, R. C. M. E; GOUVEIA, A. L. Acesso e Uso de Informação Geográfica: Introdução ao ambiente SIG: QGIS. Programa de Atualização Permanente da BCIM. IBGE, Rio de Janeiro, 2010.
- GALO, M.; CAMARGO, P. O. **Utilização do GPS no controle de qualidade de cartas**. In: COBRAC - 1994, 1º Congresso Brasileiro de Cadastro Técnico Multifinalitário. Tomo II, p. 41-48, Florianópolis, 1994.
- GEMAEL, C. **Introdução ao ajustamento de observações**. Editora UFPR. Curitiba, PR. 1994.
- GOODCHILD, M. F. **Attribute Accuracy** In: GUPTILL, S. C.; MORRISON, J. L., Elements of Spatial Data Quality, Elsevier Science, 1995. Cap. 4, p. 59-80.
- GSDI, *Global Spatial Data Infrastructure*. **Developing Spatial Data Infrastructures: The SDI Cookbook**. GSDI, 2004.
- GUERRA, P. A. G. **Geoestatística Operacional**. Departamento Nacional da Produção Mineral, Brasília, 1988.
- IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. **Noções Básica de Cartografia**. IBGE, 1998.

- IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. **Base Cartográfica Contínua do Brasil, ao milionésimo - BCIM**. IBGE, 2010.
- IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. **Especificações Técnicas para Execução dos Serviços de Mapeamento da Escala 1:250 000**. IBGE, 2011a.
- IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. **Manual de procedimentos Técnicos para Fiscalização, Controle de Qualidade e Validação da Base Cartográfica Contínua na Escala 1:250 000**. IBGE, 2011b.
- ISSMAEL, L. S. **Generalização Cartográfica: Determinação de Operadores e de Escalas Catastróficas**. Dissertação de Mestrado. Rio de Janeiro: Instituto Militar de Engenharia, 2003.
- ISO 19113. **Geographic Information – Quality principles**. International Standardization for Organization. ISO TC/211 Secretariat, Genebra, Suíça. 2002.
- ISO 19114. **Geographic Information – Quality evaluation procedures**. International Standardization for Organization. ISO TC/211 Secretariat, Genebra, Suíça. 2003.
- ISO 19115. **Geographic Information – Metadata**. International Standardization for Organization, Genebra, Suíça. 2003
- JOURNEL, A. G. **The lognormal approach to predicting local distributions of selective mining unit grades**. *Mathematical Geology*, v. 12, n. 4, p.285-303. 1980.
- JOURNEL, A. G. **Fundamentals of geostatistics in five lessons**. Washington: American Geophysical Union, 1989.
- JOURNEL, A. G.; HUIJBREGTS, J. C. H. **Mining geostatistics**. The Blackburn Press: New Jersey, 2003.
- JURAN, J. M. **A qualidade desde o projeto: novos passos para o planejamento da qualidade em produtos e serviços** / J. M. Juran; tradução de Nivaldo Montingelli Jr. Pioneira. São Paulo, 1997. 3.ed. 551p. : il.
- KAINZ, W. **Logical Consistency** In: GUPTILL, S. C.; MORRISON, J. L., *Elements of Spatial Data Quality*, Elsevier Science, 1995. Cap. 8, p. 109-138.
- KEATES, J. S. **Cartographic Design and Production**. Longman Scientific & Technical, New York, 1989.

- KEMP, Z.; THEARLE, R. *Modelling relationships in spatial databases*. In: INTERNATIONAL SYMPOSIUM ON SPATIAL DATA HANDLING, 5th, Arlington, 1992.
- LANDIM, P. M. B. **Análise Estatística e dados geológicos**. São Paulo: Fundação Editora da UNESP, 1998.
- LANDIM, P. M. B. **Análise estatística de dados geológicos multivariados**. Texto Didático 03. Laboratório de Geomática, DGA, IGC, UNESP, Rio Claro, 2000. Disponível em <<http://www.rc.unesp.br/igce/aplicada/DIDATICOS/LANDIM/multivariados.pdf>> Acessado em: 25 de março de 2013.
- LEME, L. A. P. P. **Uma arquitetura de software para catalogação automática de dados geográficos**. Dissertação de Mestrado em Informática. PUC-Rio, Rio de Janeiro. 2006.
- LISBOA FILHO, J.; IOCHPE, C.; BEARD, K. *Applying Analysis Patterns in the GIS Domain*. in Proceedings of SIRC`98, Dunedin, NZ, 1998.
- LISBOA FILHO, J. **Projeto Conceitual de Banco de Dados Geográficos através da Reutilização de Esquemas, utilizando Padrões de Análise e um Framework Conceitual**. Tese de Doutorado em Ciência da Computação. Porto Alegre: PPGC da UFRGS, 2000. 212p. il.
- LISBOA FILHO, J.; VEGI, L. F. M.; SOUZA, W. D.; LAMAS, J. P. C.; COSTA, G. L. S.; OLIVEIRA, W. M.; CARRASCO, R. S.; FERREIRA, T. G.; BAIA, J. W. **Uma infraestrutura de dados espaciais para o projeto Geominas com metadados definidos no Perfil MGB da INDE**. Revista Brasileira de Cartografia, n.65, v.1, p.123-138, 2011.
- LONGLEY, P. A.; GOODCHILD, D. J M; RHIND, D. W. **Sistema de Informações Geográficas**. 3ª Edição. Editora BOOKMAN COMPANHIA EDITORA LTDA. 2013
- LUNARDI, O. A. **Controle de Qualidade de Dados Espaciais**. Mundo GeoConnect2012. São Paulo, SP. 30 de maio de 2012.
- MARIANI, E. J. **As Normas ISO**. Revista Científica Eletrônica de Administração. FAEF - Faculdade de Agronomia e Engenharia Florestal de Garça . São Paulo. 2006.
- MASSER, I. **Building European Spatial Data Infrastructures**. 2ª Edição. ESRI Press. EUA, 2010.

- MATHERON, G. **Les variables régionalisées et leur estimation, une application de la théorie des fonctions aléatoires aux sciences de la nature.** Ed. Masson et Cie, Paris, 1965.
- MCMMASTER, R. B.; SHEA, K. S. **Generalization in Digital Cartography.** 1.ed. Washington: Association of American Geographers, 1992.
- MÔNICO, J. F. G.; DAL PÓZ, A. P.; GALO, M. ; SANTOS, M. C.; OLIVEIRA, L. C. **Acurácia e Precisão: revendo os conceitos de forma acurada.** Boletim de Ciências Geodésicas, sec. Comunicações, Curitiba, v. 15, nº 3, p.469-483, jul-set, 2009.
- MONMONIER, M. **How to Lie With Maps.** The University of Chicago Press, Chicago, 1991.
- MORETTIN, L. G. **Estatística Básica: Probabilidade e Inferência,** volume único. São Paulo: Pearson Prentice Hall, 2010.
- MORRISON, J. L. **Spatial Data Quality** In: GUPTILL, S. C.; MORRISON, J. L., Elements of Spatial Data Quality, Elsevier Science, 1995. Cap. 1, p. 1-12.
- MOSTAFAVI, M. A.; G. EDWARDS, R. JEANSOULIN. **An ontology-based method for quality assessment of spatial databases.** In: A. U. Frank and E. Grum (compilers), Proceedings of the ISSDQ'04, Vol. 1. Geo-Info 28a, pp 49-66. Dept. for Geoinformation and Cartography, Vienna University of Technology. 2004.
- NAGATOMI, R. C. M. E.; FERNANDES, W. S.; SILVA, R. L.; BORBA, R. L. R.; SANTOS, A. S. **Migration of Cartographic Database of Brazil, in the millionth, according standards of the INDE.** Latin America Geospatial Forum - LAGF, de 16 a 19 de agosto de 2011. Disponível em <<http://www.lagf.org/2011/proceeding/Renata1.pdf>>. Acesso em 2 de maio de 2012.
- NOGUERAS, J. **Geographic Information Metadata for Spatial Data Infrastructures.** Universidade de Zaragoza, Espanha, 2005.
- NOGUEIRA JÚNIOR, J. B. **Controle de qualidade de produtos cartográficos: uma proposta metodológica.** Dissertação de Mestrado em Ciências Cartográficas. UNESP. Presidente Prudente, 2003.
- OLIVEIRA, L. F. **Análise de modelos digitais de superfície gerados a partir de imagens do sensor PRISM/ALOS.** Dissertação de Mestrado em Informações Espaciais. UFV. Viçosa, 2011.

- PASCUAL, A. F. R. **Normas de Calidad: ISO 19113 e ISO 19114**. Instituto Geográfico Nacional da Espanha (IGN). Curso de IDE. Disponível em <[http://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CFoQFjAA&url=http%3A%2F%2Fwww.cp-idea.org%2Fdocumentos%2FnormasEspecificaciones%2F3.B.3a\\_Calidad.doc&ei=GNfYT55KkbLxBITvmO4D&usg=AFQjCNFlAz\\_VygMqNv910Orju92jchgjRg&sig2=SSqhzW45IEEruEmYQcoGWg](http://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CFoQFjAA&url=http%3A%2F%2Fwww.cp-idea.org%2Fdocumentos%2FnormasEspecificaciones%2F3.B.3a_Calidad.doc&ei=GNfYT55KkbLxBITvmO4D&usg=AFQjCNFlAz_VygMqNv910Orju92jchgjRg&sig2=SSqhzW45IEEruEmYQcoGWg)>. Acesso em 8 de junho de 2012.
- PEARSON, A. **Gestão da Qualidade**. 1ª ed. Série Didática. Pearson Education, 2011.
- PEUQUET, D. J. *A conceptual framework and comparasion of spatial data modeles*. Cartographica, v.21, n.º 4, pp. 66-113, 1984.
- PETERNELLI, L. A.; MELLO, M. P. **Conhecendo o R: uma visão estatística**. 1ªed. Série Didática. Viçosa-MG. Editora UFV, 2012.
- PRADO, B. R.; HAYAKAWA, E. H.; BERTANI, T. C.; SILVA, G. B. S.; PEREIRA, G.; SHIMABUKURO, Y. E. **Padrões para metadados geográficos digitais: modelo ISO 19115:2003 e modelo FGDC**. Revista Brasileira de Cartografia, n.62, v.1, p.33-41, 2010.
- REIS, R. B; XAVIER , L. A.; KRUMBIEGEL, M.; BARBOSA, F. O.; PEREIRA, P. B. M.; AZEVEDO, J. B. **Mapeamento da base contínua do Brasil na escala 1:250.000 com auxílio de sensores orbitais - Projeto BC250**. Anais do XV Simpósio Brasileiro e Sensoriamento Remoto - SBSR, Curitiba, 2011.
- ROBINSON, A. H.; MORRISON, J. L.; MUEHRCKE, P. C.; KIMERLING, A. J.; GUPTIL, S. C. **Elements of Cartography**. New York: John Wiley & Sons, 1995. 6. ed.
- RODRIGUES, M., ALMEIDA, O. W. R. *Modelagem de Dados Espaciais para Sistemas de Informações Geográficas*. In: CONGRESSO E FEIRA PARA USUÁRIOS DE GEOPROCESSAMENTO, 1., GIS Brasil 94. Curitiba: Sagres, 1994.
- ROSA, G. **Avaliando a Qualidade dos Estimadores de Variograma (Variograma Experimental) e do Método de Mínimos Quadrados Ponderados para Estimação dos Parâmetros do Modelo de Variograma Teórico do Processo**. Dissertação de Mestrado em Estatística. UFMG, Belo Horizonte, 2003.
- ROYLE, A. G. **Why Geostatistics?** Engineering and Mining Journal. New York, 1979.

- SALGÉ, F., **Semantic Accuracy** In: GUPTILL, S. C.; MORRISON, J. L., Elements of Spatial Data Quality. Elsevier Science, 1995. Cap. 7, p. 139-152.
- SANTOS, A.P. **Avaliação da acurácia posicional em dados espaciais com o uso da estatística espacial**. Dissertação de Mestrado em Informações Espaciais. UFV. Viçosa, MG, 2010a.
- SANTOS, G. R. **Hierarquização geométrica dos preditores geoestatísticos**. Tese de Doutorado em Estatística e Experimentação Agropecuária. UFLA, Lavras, 2010b.
- SERVIGNE, S.; LESAGE, N.; LIBOUREL, T. **Quality Components, Standards and metadata**. In: Fundamentals of Spatial Data Quality. Editora ISTE. Londres, 2006. Cap. 10, p. 179-210.
- SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. **Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core**. Ci. Inf., Brasília, v. 29, n. 1, p. 93-102, jan./abr. 2000.
- STEIN, A.; SHI, W.; BIJKER, W. **Quality aspects in spatial data mining**. CRC Press, 2009.
- VIEIRA, S. R. **Geoestatística em estudos de variabilidade espacial do solo**. In: Tópicos em ciência do solo, Viçosa, Sociedade Brasileira de Ciência do Solo, 2000. v.1, p.1-54.
- VIANNA, C. R. F. **Generalização Cartográfica em Ambiente Digital – Escala 1:250.000 a partir de Dados Cartográficos Digitais na Escala 1:50.000**. Dissertação de Mestrado, Instituto Militar de Engenharia/ IME, Rio de Janeiro, 1997.
- WACKERNAGEL, H. **Multivariate geostatistics**. Springer-Verlag, 1995.
- YAMAMOTO, J. K.; MAO, X. M.; KOIKE, K.; CROSTA, A. P.; LANDIM, P. M. B.; HU, H. Z.; WANG, C. Y.; YAO, L. Q. **Mapping an uncertainty zone between interpolated types of a categorical variable**. Computers and Geosciences, v. 40, p. 146-152, 2012.
- YAMAMOTO, J. K.; LANDIM, P. M. B.. **Geoestatística: conceitos e aplicações**. São Paulo: Oficina de Textos, 2013.

## **APÊNDICES**

## APÊNCICE A - Medidas de qualidade

Os apêndices A.1, A.2 e A.3 apresentam as medidas de qualidade elaboradas durante a dissertação, para avaliar a omissão, comissão e acurácia posicional absoluta da Base Cartográfica Contínua, na escala 1:250.000, baseadas na norma ISO 19114:2003 e nas Especificações Técnicas da BC250 descritas em IBGE (2011a).

### Apêndice A.1 - Medida de qualidade para itens em excesso

Linha	Componente	Descrição
1	Identificador	10101
2	Nome	Taxa de itens em excesso em hidrografia
3	Alias	Excesso Hidrografia
4	Nome do Elemento	Comissão
5	Medida Básica	Taxa de Comissão
6	Definição	O resultado indica o percentual de feições de hidrografia em excesso no conjunto de dados em estudo, em relação a quantidade de elementos representados, em relação as classes: trecho de drenagem, trecho massa d'água e massa d'água
7	Descrição	<p>Medida de qualidade aplicada ao grupo de feições da categoria de informação hidrografia: trecho de drenagem, trecho massa d'água e massa d'água, devido ao vínculo espacial e de representação entre as mesmas. Segundo as especificações técnicas da BC250 (IBGE, 2011a): O <b>curso de água (trecho de drenagem)</b> é adquirido quando sua <b>extensão for superior a 2.500m</b> (esta restrição não é considerada para cursos d'água que compõe limites políticos administrativos). A nascente do curso d'água principal será a que estiver em maior altitude. Quando a <b>largura do curso de água for superior a 125m</b> será adquirido com <b>geometria do tipo polígono (trecho massa d'água)</b>, sendo considerado como margem dupla na especificação, <b>mesmo em trechos com extensão inferior a 1.250 metros</b>. A <b>massa d'água</b> é adquirida quando a <b>área for superior ou igual a 62.500m<sup>2</sup></b>. Quando duas ou mais massas de água estiverem separadas por uma distância inferior a 125m essas deverão ser combinadas numa única, exceto quando pelo menos uma delas possuir nome, conforme Figura 80. Os critérios para combinação são descrito abaixo. Quando a área do polígono, após a combinação, for inferior a 62.500m<sup>2</sup> elas não serão adquiridas.</p>  <p>Figura 80 - Generalização cartográfica considerada em massas d'água em excesso</p> <p>Fonte IBGE (2011a)</p> <p>Para o cálculo da taxa de comissão utilizar a equação definida por Salgé (1995),</p> $\tau^+ = \frac{N^+}{\max(N, N^0)}$ <p>onde:</p> <p><math>N^-</math> é o número de ocorrência que existem na realidade percebida e que não foram reconhecidas, ou seja, omissão.</p> <p><math>N^+</math> é o número de ocorrências na amostra que não ocorrem na realidade percebida, ou seja, comissão.</p> <p><math>N^0</math> é o número de ocorrências na realidade percebida.</p> <p><math>N</math> é o número de elementos da amostra, dado por:</p> $N = N^0 + N^+ - N^-$
8	Parâmetro	Não existe. Em relação a completude não há um percentual definido como aceitável, tornando inviável a comparação
9	Tipo de Valor	4 - Porcentagem
10	Estrutura do Valor	-
11	Referência	SALGÉ, F., <b>Semantic Accuracy</b> In: GUPTILL, S. C.; MORRISON, J. L., Elements of Spatial Data Quality, Elsevier Science, 1995. Cap. 7, p. 139-152. IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. <b>Especificações Técnicas para Execução dos Serviços de Mapeamento da Escala 1:250 000</b> . IBGE, 2011a. IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. <b>Manual de procedimentos Técnicos para Fiscalização, Controle de Qualidade e Validação da Base Cartográfica Contínua na Escala 1:250 000</b> . IBGE, 2011b.
12	Exemplo	0,82 (17 trechos de drenagem estão no conjunto de dados, mas 14 não existem no universo de discussão)
13	Classe(s) de Ocorrência	"M10101_HID_Trecho_Drenagem.shp" com geometria tipo linha. "M10101_HID_Trecho_MassaDagua.shp" com geometria tipo área. "M10101_HID_MassaDagua.shp" com geometria tipo área.

## Apêndice A.2 - Medida de qualidade para itens ausentes

Linha	Componente	Descrição
1	Identificador	10201
2	Nome	Taxa de itens ausentes em hidrografia
3	Alias	Ausência Hidrografia
4	Nome do Elemento	Omissão
5	Medida Básica	Taxa de Omissão
6	Definição	O resultado indica o percentual de feições de hidrografia, mas especificamente: trecho de drenagem, trecho massa d'água e massa d'água ausentes no conjunto de dados em estudo, em relação a um universo de discussão.
7	Descrição	<p>Medida de qualidade aplicada ao grupo de feições da categoria de informação hidrografia: trecho de drenagem, trecho massa d'água e massa d'água, devido ao vínculo espacial e de representação entre as mesmas. Segundo as especificações técnicas da BC250:</p> <p><b>O curso de água (trecho de drenagem)</b> é adquirido ou atualizado quando sua <b>extensão for superior a 2.500m</b> (esta restrição não é considerada para cursos d'água que compõe limites políticos administrativos). A nascente do curso d'água principal será a que estiver em maior altitude.</p> <p>Quando a <b>largura do curso de água for superior a 125m</b> será adquirido com <b>geometria do tipo polígono (trecho massa d'água)</b>, sendo considerado como margem dupla na especificação, <b>mesmo em trechos com extensão inferior a 1.250 metros</b>.</p> <p>A <b>massa d'água</b> é adquirida ou atualizada quando a <b>área for superior ou igual a 62.500m<sup>2</sup></b>. Quando duas ou mais massas de água estiverem separadas por uma distância inferior a 125m essas deverão ser combinadas numa única, exceto quando pelo menos uma delas possuir nome, conforme Figura 81. Os critérios para combinação são descrito abaixo. Quando a área do polígono, após a combinação, for inferior a 62.500m<sup>2</sup> elas não serão adquiridas.</p>  <p>Figura 81 - Generalização cartográfica considerada em massas d'água ausentes</p> <p>Fonte IBGE (2011a)</p> <p>Para o cálculo da taxa de omissão utilizar a equação definida por Salgé (1995),</p> $\tau^- = \frac{N^-}{\max(N, N^0)}$ <p>onde:</p> <p><math>N^-</math> é o número de ocorrência que existem na realidade percebida e que não foram reconhecidas, ou seja, omissão.</p> <p><math>N^+</math> é o número de ocorrências na amostra que não ocorrem na realidade percebida, ou seja, comissão.</p> <p><math>N^0</math> é o número de ocorrências na realidade percebida.</p> <p><math>N</math> é o número de elementos da amostra, dado por:</p> $N = N^0 + N^+ - N^-$
n8	Parâmetro	Não existe. Em relação a completude não há um percentual definido como aceitável, tornando inviável a comparação
9	Tipo de Valor	4 - Porcentagem
10	Estrutura do Valor	-
11	Referência	SALGÉ, F., <b>Semantic Accuracy</b> In: GUPTILL, S. C.; MORRISON, J. L., Elements of Spatial Data Quality, Elsevier Science, 1995. Cap. 7, p. 139-152. IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. <b>Especificações Técnicas para Execução dos Serviços de Mapeamento da Escala 1:250 000</b> . IBGE, 2011a. IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. <b>Manual de procedimentos Técnicos para Fiscalização, Controle de Qualidade e Validação da Base Cartográfica Contínua na Escala 1:250 000</b> . IBGE, 2011b.
12	Exemplo	0,058 (17 trechos de drenagem estão no conjunto de dados, mas 1 trecho estava ausente).
13	Classe(s) de Ocorrência	"M10201_HID_Trecho_Drenagem.shp" com geometria tipo linha. "M10201_HID_Trecho_MassaDagua.shp" com geometria tipo área. "M10201_HID_MassaDagua.shp" com geometria tipo área.

### Apêndice A.3 - Medida de qualidade para acurácia posicional absoluta<sup>57</sup>

Linha	Componente	Descrição												
1	Identificador	30101												
2	Nome	Incertezas posicionais												
3	Alias	Incertezas Posicionais												
4	Nome do Elemento	Acurácia Absoluta ou Externa												
5	Medida Básica	Medida Linear												
6	Definição	Medida linear referente a distância entre a posição de medição e a que é considerada como a respectiva posição verdadeira (em metros).												
7	Descrição	<p>Medida de qualidade aplicada em todo conjunto de dados, segundo as especificações técnicas. O mapeamento da BC250 deve ser compatível com a escala 1:250.000, ou seja, <b>90% dos pontos bem definidos na base cartográfica</b>, quanto medidos, <b>não podem apresentar erros planimétricos superiores a 125m</b>. A aferição desta acurácia é realizada em pontos bem identificados na BC250, como: entroncamento de vias, pontes ou passagens de nível, estes são comparados com pontos tomados como verdadeiro, medindo a distância entre os dois. Para comprovação desta acurácia, podem ser tomados como verdadeiro as seguintes referências:</p> <p><b>Pontos medidos em campos</b> utilizando tecnologia GNSS, que apresentem erro médio quadrático planimétrico igual ou inferior a 15m e erro médio quadrático altimétrico igual ou inferior a 20m (a altura elipsoidal, fornecida pelo GNSS, deve ser convertida em altitude ortométrica).</p> <p><b>Pontos obtidos em folhas topográficas</b> do Sistema Cartográfico Nacional na escala de 1:50.000 ou maiores.</p> <p><b>Pontos obtidos através de ortomosaico</b> de imagens com resolução espacial igual ou melhor que 20 m, precisão compatível com a escala de 1:50.000 ou maiores e que não tenham sido usados na produção do produto testado.</p> <p>A medida de qualidade deve ser calculada num sistema de coordenadas projetado e adequado, como o descrito abaixo (IBGE, 2010):</p> <p style="text-align: center;"><b>Parâmetros Projeção Policônica</b></p> <p style="text-align: center;">Longitude origem: -54° Latitude origem: 0° Unidade de trabalho: metros</p>												
8	Parâmetro	<p>O parâmetro é o PEC para a escala 1:250.000</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Classe</th> <th>PEC (mm x DE)</th> <th>EP (mm x DE)</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>125 m</td> <td>75 m</td> </tr> <tr> <td>B</td> <td>200 m</td> <td>125 m</td> </tr> <tr> <td>C</td> <td>250 m</td> <td>150 m</td> </tr> </tbody> </table> <p>DE - Denominador da escala da carta; EP - Erro padrão;</p>	Classe	PEC (mm x DE)	EP (mm x DE)	A	125 m	75 m	B	200 m	125 m	C	250 m	150 m
Classe	PEC (mm x DE)	EP (mm x DE)												
A	125 m	75 m												
B	200 m	125 m												
C	250 m	150 m												
9	Tipo de Valor	4 - Porcentagem												
10	Estrutura do Valor	-												
11	Referência	BRASIL, Decreto n.º 89.817, de 20 de junho de 1984. Dispõe sobre as instruções reguladoras das normas técnicas da cartografia nacional. Diário Oficial [da República Federativa do Brasil], Poder Executivo, Brasília, 22 de junho de 1984. IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. <b>Base Cartográfica Contínua do Brasil, ao Milionésimo - BCIM</b> . IBGE, 2010.												
12	Exemplo	Classe A. Num universo de 100 pontos de controle de campo identificados no conjunto de dados, 9 estavam acima do PEC classe A, ou seja, 91% dos pontos estão abaixo de 125 metros no caso da escala 1:250.000.												
13	Classe(s) de Ocorrência	"M30101_AcuraciaPosicionalAbsoluta.shp" com geometria tipo linha e o atributo "ErroLinear" tipo real.												

<sup>57</sup> O cálculo do erro planimétrico foi baseado em IBGE (2010), onde a Projeção Policônica foi aplicada no cálculo de extensões representadas na Base Cartográfica Contínua, ao milionésimo - BCIM, na escala 1:1000.000, similar a BC250, com dimensões continentais, esta foi a primeira base cartográfica a ser contemplada na INDE.