

Metodologia de Análise no Controle de Qualidade de Vetores de Hidrografia e Modelos Digitais de Terreno no Mapeamento do Estado de São Paulo

Rafael Duarte¹
Mateus Pedrucci Romanholi²

¹ Instituto Geográfico e Cartográfico – IGC
Gerência de Cartografia
Rua Boa Vista, 150 - Centro, São Paulo - SP, Brasil
rduarte@sp.br

² Escola Politécnica da Universidade de São Paulo – Poli/USP
Departamento de Engenharia de Transportes - PTR
Av. Prof. Almeida Prado, Travessa 2, nº 83 - Butantã – São Paulo - SP, Brasil
mateus.romanholi@usp.br

Abstract. This paper describes how cartographic literature has been applied to validate and control the quality, in terms of positional accuracy and topologic consistence, of the vector hidrography and digital terrain models produced in the scope of the mapping project being carried out by the Geographic and Cartographic Institute of the State of São Paulo (Brazil) at 1:25.000 scale.

Keywords: Methodology of Validation, quality control, cartography Metodologia de Validação, Controle de Qualidade, Cartografia

1. Introdução

Segundo Leal e Dalmolin (2002) devido a demanda por produtos cartográficos de qualidade satisfatória é necessário estabelecer parâmetros mínimos para o desenvolvimento destes produtos. Neste sentido observa-se a necessidade de trabalhos que descrevam processos de controle de qualidade, conforme proposto neste trabalho.

Com a necessidade de normatizar produtos geoespaciais, a International Organization for Standardization - ISO criou a série ISO 19000, que propõem a padronização de dados geoespaciais. A ISO:19157 (2011), em especial, define formas para classificar a qualidade da informação geoespacial de duas maneiras: posicional e semântica.

Para a análise semântica dos produtos digitais, foi proposta como norma para aquisição de informação geoespacial as Especificações Técnicas para Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV), da Infraestrutura Nacional de Dados Espaciais (INDE), que propõe a padronização da aquisição de dados geoespaciais (CONCAR, 2011). A ET-ADGV propõe novos parâmetros em relação ao Decreto nº 89.817/84, possibilitando o controle de qualidade de produtos digitais.

2. Revisão Bibliográfica

Segundo Gupstill e Morrison (2013) e Fegeas et. al (1992) podemos diferenciar as inconformidades das seguintes formas: qualidade posicional, a linhagem, fidelidade de atributos, completeza, consistência lógica e fidelidade à semântica.

Segundo Miller e Laflamme (1958), o Modelo Digital de Terreno (MDT) ou Digital Terrain Model (DTM), do inglês, é uma representação estatística da superfície contínua do terreno, por

meio de pontos selecionados, com coordenadas X, Y, Z determinadas, em um dado sistema de coordenadas.

Segundo Blak (2007) as análises de qualidade do MDT são divididos em análises quantitativas e qualitativas. As análises quantitativas são análises que levam em consideração observações das diferenças encontradas entre o modelo e pontos de referência tidos como pontos verdadeiros. As análises qualitativas, por sua vez, levam em consideração comparações que utilizam como referência a estatística e a distribuição de amostras pontuais.

Para análises quantitativas de MDT segundo CONCAR (2011) recomenda-se a utilização da ET-ADGV, na qual são definidas classes qualitativas de precisão. Este indicador estatístico é o Padrão de Exatidão Cartográfica dos Produtos Cartográficos Digitais (PEC-PCD) na figura 1. Quanto a avaliação posicional estatística absoluta segundo Merchant (1982) quando comparados pontos de melhor acurácia com pontos homólogos no produto torna-se possível a aplicação de testes medindo a tendência e precisão do produto. A utilização desta proposta para, visando porem atender ao decreto nº 89.817 de 20 de junho de 1984 (BRASIL,1984), foi documentada por Galo e Camargo (1994).

PEC-PCD (%)	1:1.000		1:2.000		1:5.000		1:10.000		1:25.000		1:50.000		1:100.000		1:250.000	
	PEC (m)	EP (m)														
A	0,27	0,17	0,27	0,17	0,54	0,34	1,35	0,84	2,70	1,67	5,50	3,33	13,70	8,33	27,00	16,67
B	0,50	0,33	0,50	0,33	1,00	0,66	2,50	1,67	5,00	3,33	10,00	6,66	25,00	16,66	50,00	33,33
C	0,60	0,40	0,60	0,40	1,20	0,80	3,00	2,00	6,00	4,00	12,00	8,00	30,00	20,00	60,00	40,00
D	0,75	0,50	0,75	0,50	1,50	1,00	3,75	2,50	7,50	5,00	15,00	10,00	37,50	25,00	75,00	50,00

Figura 1: Valores PEC-PCD para modelos digitais de superfície segundo (CONCAR, 2011)

Em análises qualitativas de MDTs Blak (2007, p. 443) sugere utilizar o relevo sombreado (*shaded relief* ou *hillshade*) do MDT, que auxilia na visualização de anomalias no terreno. Quando disponíveis modelos ou linhas de quebra (*BreakLine*) podem ser utilizados como boas referências para a verificação.

3. Desenvolvimento

Na execução da análise, inicialmente é identificado o produto a ser realizada a análise. Na execução deste processo, é necessário efetuar controle e separação conforme demonstrado pelo diagrama da figura 2. Após o processo de controle dos arquivos, são executadas as análises dos vetores de hidrografia e do MDT. Após o processo de controle dos arquivos serão descritas as análises do vetor de hidrografia editada e do MDT.

3.1. Descrição das análises efetuadas nos Vetores de Hidrografia

A seguir serão detalhadas as análises efetuadas na rede hidrográfica vetorial que foram agrupadas em: Consistência Lógica (fidelidade de atributos, fidelidade à semântica), Completitude e Acurácia Posicional relativa.

3.1.1. Consistência Lógica (fidelidade de atributos, fidelidade à semântica)

Utilizando as definições propostas por Guptill e Morrison (2013), a análise de consistência lógica é executada conforme as etapas descritas a seguir:

1. Inconsistência de coordenadas X, Y e Z entre os arquivos ESRI shapefile® (SHP) e Microstation® DGN: uma vez que os produtos vetoriais foram solicitados às empresas

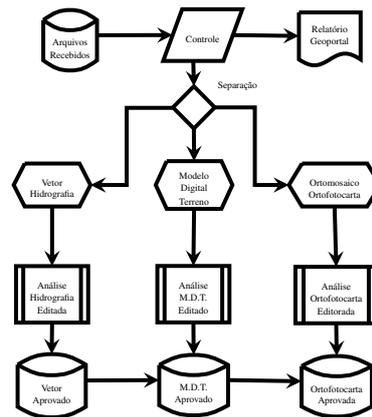


Figura 2: Diagrama de separação e organização para análise

executoras nos formatos DGN e SHP, a análise inicial compara as coordenadas dos vértices das feições homólogas entre os dois formatos de arquivos, identificando também erros de omissão ou comissão;

2. Inconsistência de Z de barragens: são verificadas inconsistências entre a coordenadas Z das barragens e da rede de drenagem próxima;
3. Sentido de restituição (eixos e rios) não confere: sentido de restituição de montante para jusante da hidrografia linear, respeitando o sentido natural do fluxo d'água na superfície;
4. Desconectividade entre elementos (gaps): Se caracteriza por ocorrências de desconectividades entre os elementos da rede de drenagem;
5. Inconsistência no fechamento de polígonos: ocorre quando a linha que define a borda de um polígono não possui os vértices inicial e final coincidentes e, portanto, não define uma região fechada;
6. Inconsistência de atributos do shapefile de linhas ou áreas: Inconsistência no preenchimento de atributos conforme a padronização definida pelas Especificações Técnicas para Estruturação de Dados Geospaciais Vetoriais (ET-EDGV);
7. Atributos entre ShapeFile e CAD-DGN incompatíveis -Diferenças de toponímias de feições homólogas entre os arquivos SHP e DGN;
8. Polígonos sobrepostos/não-vazados e Feições duplicadas: verificação de erros em locais onde polígonos não deveriam se sobrepor, e sim compartilharem as bordas, como nos casos de ilhas no interior de massas d'água;
9. Inconsistências na divisão de trechos: verificação de ausência de quebras entre trechos de drenagem nas confluências e também da existência de pseudo-nós, isto é, quebras desnecessárias entre trechos de drenagem contíguos;

3.1.2. Completitude

Salgé (1995) define a proposta da acurácia semântica como sendo a distância semântica entre objetos geográficos e a realidade pertencente a um modelo selecionado. Para representar estatisticamente, utilizaremos a taxa de comissão r^+ e taxa de omissão r^- , sendo estas representadas pelas equações 1 e 2.

$$r^+ = \frac{N^+}{\max(N, N^0)} \quad (1)$$

$$r^- = \frac{N^-}{\max(N, N^0)} \quad (2)$$

Onde N^- é o número de ocorrências que existem na realidade percebida e que não foram reconhecidas, constituindo erros de omissão, enquanto que N^+ representa o oposto. O elemento N representa o número de elementos da amostra e N^0 o número de ocorrências da realidade percebida. Para calcularmos o valor de N utilizamos a equação 3.

$$N = N^0 + N^+ - N^- \quad (3)$$

3.1.3. Acurácia Posicional relativa

Segundo Joksić e Bajat (2004), na avaliação relativa são calculadas as diferenças entre as coordenadas de feições coletadas no processo e locais onde elas deveriam estar. Segundo Gesch et al. (2014), diferentemente da análise absoluta, que testa erros aleatórios e sistemáticos, a análise posicional relativa calcula somente erros aleatórios.

Seguindo a metodologia proposta por Shi (2009, p. 364) para avaliar a acurácia relativa, foram coletados pontos em estação fotogramétrica que foram comparados com pontos sobre as feições de hidrografia restituídas. Para comparações foi utilizada a metodologia proposta por Duarte e Romanholi (2016) onde podemos encontrar um ponto sobre a linha e em seguida calcular o Valor Relativo Vertical $\Delta\bar{Z}$ (equação 4) e os valores relativos horizontais na ordenada $\Delta\bar{X}$ (equação 5) e $\Delta\bar{Y}$ (equação 6).

$$\Delta\bar{Z} = Z_i - Z_k \quad (4)$$

$$\Delta\bar{X} = X_i - X_k \quad (5)$$

$$\Delta\bar{Y} = Y_i - Y_k \quad (6)$$

Com os valores $\Delta\bar{X}$ e $\Delta\bar{Y}$, calculamos a resultante de ambas as ordenadas a partir da equação 7.

$$\Delta\bar{XY} = \sqrt{(\Delta\bar{X}_i^2 + \Delta\bar{Y}_i^2)} \quad (7)$$

A partir do valor da resultante $\Delta\bar{XY}$, utilizamos a equação 8 para calcular o Erro Médio Quadrático ou Root Mean Square Error (RMSE).

$$RMSE_{\Delta\bar{XY}} = \sqrt{\frac{\sum_{i=1}^n \Delta\bar{XY}_i^2}{n}} \quad (8)$$

Para a componente vertical $\Delta\bar{Z}$, podemos calcular o RMSE das observações utilizando a equação 9, considerando o número de observações n da amostra.

$$RMSE_{\Delta\bar{Z}} = \sqrt{\frac{\sum_{i=1}^n \Delta\bar{Z}_i^2}{n}} \quad (9)$$

Para a hidrografia vetorial, o edital de contratação da primeira área do mapeamento IGC (Lote 1) exigiu um RMSE de 1,67m, aceitando 5,01m de erro máximo e tolerando um erro de 2,20m em até 90% dos pontos.

3.2. Descrição das análises efetuadas no Modelo Digital de Terreno

Neste item serão detalhadas as análises efetuadas nos Modelos Digitais de Terreno (MDTs), que foram agrupadas em: Análise Qualitativa e Análise Quantitativa.

3.2.1. Análise Qualitativa

Foram analisadas as coordenadas das *breakLine* e erros de edição nos MDTs, conforme as etapas descritas a seguir:

1. Discrepâncias entre as *breakLines* e a hidrografia vetorial restituída: foram verificados erros de omissão e comissão e diferenças nas coordenadas de vértices entre as breaklines utilizadas no MDT e as feições homólogas da hidrografia vetorial;
2. Diferenças entre arquivos no formato texto (ASCII) e no formato imagem (TIF): foram comparados os dois formatos de arquivo (GEOTIFF e ASCII XYZ) de um mesmo MDT, que foram solicitados às empresas executoras;
3. Análise de bordas: identificação de altimetria discrepante na região de sobreposição entre dois MDTs vizinhos;
4. Edição do terreno: discrepâncias notáveis e bruscas, que indicassem a presença de edifícios, por exemplo, foram procuradas visualmente no arquivo de relevo sombreado (*hillshade*) gerado a partir do MDT;
5. Represas e lagos não-planos: foi verificado se os valores de cota do MDT estavam constantes no interior dos polígonos de feições de lagos e represas (reservatórios);
6. Erros de altitude entre o MDT e barragens: foram verificadas diferenças maiores que 1,67m entre as cotas das barragens, que foram utilizadas como breaklines, e as cotas do MDT;
7. Ilhas achatadas: feições de ilhas que tiveram as cotas altimétricas de seus vértices alteradas para valores constantes, assim como lagos.
8. Erro de altitude entre o MDT e os vértices das linhas restituídas: foram identificadas diferenças maiores que 4m entre os valores de cota dos rios restituídos e seus respectivos valores no MDT.

3.2.2. Análise Quantitativa

Neste tópico serão demonstradas as análises efetuadas através de comparações entre amostras pontuais e o MDT, sendo elas as de Acurácia posicional relativa e absoluta (JOKSIC; BAJAT, 2004).

3.2.2.1 Acurácia Posicional Relativa

Segundo Salgé (1995), a avaliação relativa, também conhecida como ponto-a-ponto ou interna, é um processo no qual são calculadas as diferenças entre as coordenadas de feições coletadas no processo de mapeamento e as coordenadas dos pontos onde as mesmas feições deveriam estar.

Segundo Gesch et al. (2014), em alguns casos a acurácia relativa é mais importante que a absoluta, pois, diferentemente da análise absoluta, que testa erros aleatórios e sistemáticos,

a análise posicional relativa calcula somente erros aleatórios. Para avaliar a acurácia relativa foram utilizados pontos que foram comparados com o MDT. Podemos calcular o Valor Relativo Vertical $\Delta\bar{Z}$ com a equação 10, onde Z_i representa o ponto observado e Z_k o ponto coletado.

$$\Delta\bar{Z} = Z_i - Z_k \quad (10)$$

Com os valores de $\Delta\bar{Z}$ podemos calcular o RMSE das observações utilizando a equação 11, considerando o número de observações n da amostra.

$$RMSE_{\Delta\bar{Z}} = \sqrt{\frac{\sum_{i=1}^n \Delta\bar{Z}_i^2}{n}} \quad (11)$$

Para o MDT, o edital de contratação da primeira área do mapeamento IGC (Lote 1) exigiu um RMSE de 1,35m, aceitando 4,05m de erro máximo e tolerando um erro de 2,50m em até 90% dos pontos.

3.2.2.2 Acurácia Posicional Absoluta

Nesta etapa pretende-se verificar a qualidade geométrica do MDT através de testes estatísticos de tendência e acurácia. O objetivo é classificar o produto conforme as classes estabelecidas pela ET-ADGV ou conforme edital de contratação. Em uma amostra de 10% do total de cartas 1:25.000 do mapeamento, foram coletadas coordenadas de feições em campo e comparadas com as coordenadas das feições homólogas. Para a validação estatística da análise foram aplicados os testes descritos a seguir.

3.2.2.3 Análise da tendência

Nesta etapa são analisadas as discrepâncias entre a carta e os valores coletados em campo utilizando a equação 12.

$$\Delta Z_i = Z_i - Z_i^r \quad (12)$$

Logo após calculamos a média das amostras pela a equação 13.

$$\Delta\bar{Z} = \frac{1}{n} \sum_{i=1}^n \Delta Z_i \quad (13)$$

Depois calcula-se o desvio-padrão das discrepâncias utilizando a equação 14.

$$S_{\Delta z} = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (Z_i - \Delta\bar{Z})^2} \quad (14)$$

Para o teste de tendência, são testadas as hipóteses 15.

$$H_0 : \Delta\bar{Z} = 0 \quad \text{ou} \quad H_1 : \Delta\bar{Z} \neq 0 \quad (15)$$

O teste t amostral para as discrepâncias entre o produto e os valores coletados em campo é calculado pela equação 16.

$$t_z = \frac{\Delta\bar{Z}}{S_z} \sqrt{n} \quad (16)$$

Deve ser considerado ainda o intervalo de confiança calculado pela equação 17.

$$|t_i| = t_{(n-1, \alpha/2)} \quad (17)$$

Caso o produto apresente um valor de t_z fora do intervalo de confiança, consideramos que a carta não está livre de tendências.

3.2.2.4 Análise da precisão

A análise de precisão foi executada através da comparação das discrepâncias entre as coordenadas e o erro-padrão (EP) esperado para a classe do PEC exigida no mapeamento (Classe A). Nesta etapa, são testadas as hipóteses 18.

$$H_0 : S_z^2 = z^2 \quad \text{contra} \quad H_1 : S_z^2 > z^2 \quad (18)$$

Sendo S_z o desvio-padrão (ou erro-padrão) esperado para a coordenada Z , podemos calcular com a equação 19 o EP esperado para cada classe do PEC utilizando os valores altimétricos constantes nas normas.

$$\delta x = \frac{EP}{\sqrt{2}} \quad (19)$$

Com o valor da variância esperada obtido na equação 19, calcularemos os valores do teste qui-quadrado para a componente Z utilizando a equação 20.

$$\tilde{\chi}_x^2 = (n - 1) \frac{S_{\Delta z}^2}{\delta_x^2} \quad (20)$$

Com a equação 21, podemos verificar se o valor calculado na equação 20 está no intervalo de aceitação.

$$\tilde{\chi}_x^2 \leq \chi_{(n-1, \alpha)}^2 \quad (21)$$

Caso a equação 21 seja verdadeira concluímos que a hipótese será aceita com um nível de confiança 90%.

4. Conclusão

Para o Mapeamento do Estado de São Paulo na escala 1:25.000 contratado pelo IGC, a execução da metodologia descrita neste artigo se mostrou efetiva para a análise de qualidade dos produtos cartográficos gerados, atendendo os parâmetros estabelecidos na ET-ADGV. Para trabalhos futuros recomenda-se que sejam utilizados parâmetros que atendam as análises topológicas e lógicas propostas por Kainz (1995) e abordadas por Júnior (2003). Sugerimos, ainda, que para a análise dos erros de classificação seja testada a metodologia Zero-Defect proposta por Ho et. al (2010).

Referências

- BLAK, A. T. Dem quality assessment. In: MAUNE, D. F. (Ed.). *Digital elevation model technologies and applications: the DEM users manual*. [S.l.]: Asprs Publications, 2007. cap. 14, p. 425–473.
- BRASIL, L. D. **Decreto no 89.817, de 20 de junho. estabelece as instruções reguladoras das normas técnicas da cartografia nacional (irntcn)**. Brasília: Diário Oficial da União (DOU), v. 22, p. 06–84, 1984.
- CONCAR, D. S. G. **Especificação técnica para a aquisição de dados geoespaciais vetoriais (ET-ADGV)**. Ministério da Defesa, Exército Brasileiro, Departamento de Ciência e Tecnologia. Brasília-DF, 2ª edição, 2011.

DUARTE, R.; ROMANHOLI, M. P. **Algoritmo para cálculo de distâncias tridimensionais entre pontos e linhas em banco de dados geográficos para avaliação de acurácia posicional de dados cartográficos vetoriais.** In: . Florianópolis-SC: Universidade Federal de Santa Catarina: CONGRESSO BRASILEIRO DE CADASTRO TÉCNICO MULTIFINALITÁRIO, 2016.

FEGEAS, R. G.; CASCIO, J. L.; LAZAR, R. A. **An overview of fips 173, the spatial data transfer standard.** *Cartography and Geographic Information Systems*, Taylor & Francis, v. 19, n. 5, p. 278–293, 1992.

GALO, M.; CAMARGO, P. **Utilização do gps no controle de qualidade de cartas.** In: . Florianópolis-SC: Universidade Federal de Santa Catarina: Anais CONGRESSO BRASILEIRO DE CADASTRO TÉCNICO MULTIFINALITÁRIO, 1994. p. 41–48.

GESCH, B. et al. **Accuracy assessment of the u.s.geological survey national elevation dataset, and comparison with other large-area elevation datasets – srtm and aster.** In: . U.S. Geological Survey, Reston, Virginia: USGS O. F. R. 2014-1008, 2014. p. 5–6.

GUPTILL, S. C.; MORRISON, J. L. *Elements of spatial data quality.* [S.l.]: Elsevier, 2013.

HO, L. L.; QUINTANILHA, J. A.; QUININO, R. C. **The design of a zero-defect sampling with rectification procedure in the presence of classification errors: an application to database with spatial data.** *Brazilian Journal of Operations & Production Management*, v. 3, n. 1, p. 87–89, 2010.

JOKSIĆ, D.; BAJAT, B. **Elements of spatial data quality as information technology support for sustainable development planning.** *Spatium*, n. 11, p. 77–83, 2004.

JÚNIOR, J. B. N. **Controle de qualidade de produtos cartográficos: uma proposta metodológica.** Universidade Estadual Paulista (UNESP), 2003.

KAINZ, W. Semantic accuracy. In: GUPTILL, S. C.; MORRISON, J. L. (Ed.). *Elements of spatial data quality.* [S.l.]: Oxford Elsevier Science, 1995. v. 202, cap. 8, p. 109–138.

LEAL, E. d. M.; DALMOLIN, Q. *Considerações Sobre a Classificação de Documentos Cartográficos.* [S.l.]: Curitiba, 2002.

MERCHANT, D. Spatial accuracy standards for large scale line maps. In: *Technical papers of the American Congress on Surveying and Mapping.* [S.l.: s.n.], 1982. v. 1, p. 222–231.

MILLER, C. L.; LAFLAMME, R. A. *The Digital Terrain Model-: Theory & Application.* [S.l.]: MIT Photogrammetry Laboratory, 1958.

SALGÉ, F. Semantic accuracy. In: GUPTILL, S. C.; MORRISON, J. L. (Ed.). *Elements of spatial data quality.* [S.l.]: Oxford Elsevier Science, 1995. v. 202, cap. 7, p. 139–151.

SHI, W. *Principles of modeling uncertainties in spatial data and spatial analyses.* [S.l.]: CRC Press, 2009.

STANDARIZATION, I. O. for. *ISO 19157: Geographic Information - Data Quality. International Organization for Standarization.* [S.l.]: International Organization for Standardization, 2011.