

SABRINA LOURDES PEREIRA DE CRISTO

**ANÁLISE DA DETECÇÃO E DA INFLUÊNCIA DE OUTLIERS NA AVALIAÇÃO
DA ACURÁCIA POSICIONAL DE PRODUTOS CARTOGRÁFICOS**

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

Orientador: Nerilson Terra Santos

Coorientadores: Afonso de Paula dos Santos
Eduardo Campana Barbosa

**VIÇOSA – MINAS GERAIS
2021**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

C933a
2021

Cristo, Sabrina Lourdes Pereira de, 1996-
Análise da detecção e da influência de outliers na avaliação
da acurácia posicional de produtos cartográficos / Sabrina
Lourdes Pereira de Cristo. – Viçosa, MG, 2021.
1 dissertação eletrônica (82 f.): il. (algumas color.).

Inclui apêndice.

Orientador: Nerilson Terra Santos.

Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Estatística, 2021.

Referências bibliográficas: f.74-76.

DOI: <https://doi.org/10.47328/ufvbbt.2021.191>

Modo de acesso: World Wide Web.

1. Outliers (Estatística). 2. Cartografia - Controle de
Qualidade. 3. Coordenadas geográficas. 4. Dados geoespaciais.
I. Santos, Nerilson Terra, 1966-. II. Universidade Federal de
Viçosa. Departamento de Estatística. Programa de
Pós-Graduação em Estatística Aplicada e Biometria. III. Título.

CDD 22. ed. 519.5

Bibliotecário(a) responsável: Renata de Fátima Alves CRB6/2578

SABRINA LOURDES PEREIRA DE CRISTO

**ANÁLISE DA DETECÇÃO E DA INFLUÊNCIA DE OUTLIERS NA AVALIAÇÃO
DA ACURÁCIA POSICIONAL DE PRODUTOS CARTOGRÁFICOS**

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

APROVADA: 27 de agosto de 2021.

Assentimento:

Sabrina Lourdes Pereira de Cristo

Sabrina Lourdes Pereira de Cristo

Autora

Nerilson Terra Santos

Orientador

*Aos meus pais Júlio César e Maria de Lourdes.
As minhas irmãs Samila e Júlia.*

AGRADECIMENTOS

Tornar-me mestre é um sonho e só consegui chegar até aqui graças ao apoio de muitas pessoas. Primeiramente agradecer a Deus pelo dom da vida, pela sabedoria e por me guiar e conduzir até aqui, sendo o motivo de força para nunca desistir.

Aos meus pais Maria de Lourdes e Júlio César, meus pilares e exemplos de seres humanos, por me darem todas as condições para que eu pudesse estudar e me dedicar, por todo amor, carinho, alento e palavras de incentivos durante a trajetória.

As minhas irmãs Samila e Júlia, por todos os momentos bons de descontração e felicidade. A minha tia Gilma, por estar sempre presente, até mesmo na distância, sempre me questionando, mostrando interesse no andamento do meu trabalho e por fazer de tudo para me ajudar a não desistir do caminho a ser percorrido.

As amigas construídas no mestrado, em especial Marianna e Gleyner, pelos apertos e alegrias divididas, por serem meus fiéis amigos em todo o mestrado, por toda ajuda nos estudos e na vida.

Ao meu orientador Nerilson, por acreditar na minha pesquisa, por desde sempre me apoiar nas minhas escolhas e por todo ensinamento e orientação dada durante esses anos, que permitiu a construção deste trabalho. Agradeço também aos meus coorientadores professores Afonso e Eduardo, por toda a ajuda que me foi dada.

À todos os professores do Programa de Pós Graduação em Estatística Aplicada e Biometria, por todo o conhecimento e ensinamentos que foram passados durante esse período e aos demais funcionários do DET que sempre estiveram dispostos a ajudar.

À FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais e à CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pela bolsa concedida que permitiu a execução deste trabalho.

Por fim, a todas as pessoas que torceram para que eu pudesse concluir mais essa etapa em minha vida! Obrigada!

“Florescer onde a vontade do pai me colocou, saber esperar e entregar-me sem reservas, até que a tristeza e a alegria dos outros sejam minhas próprias tristezas e minhas próprias alegrias.”

(Trecho da Oração à São José)

RESUMO

CRISTO, Sabrina Lourdes Pereira de, M.Sc., Universidade Federal de Viçosa, agosto de 2021. **Análise da Detecção e da influência de Outliers na Avaliação da Acurácia Posicional de Produtos Cartográficos**. Orientador: Nerilson Terra Santos. Coorientadores: Afonso de Paula dos Santos e Eduardo Campana Barbosa.

Outliers são dados que se diferem do conjunto de dados ao qual pertence e podem ser identificados aplicando métodos próprios para esta finalidade. A detecção de candidatos a *outliers* é objeto de interesse de várias áreas, dentre elas o controle de qualidade cartográfica, que tem por objetivo avaliar a qualidade de um produto cartográfico e assegurar para qual finalidade o produto cartográfico pode ser utilizado. Um dos elementos do controle de qualidade cartográfica é a acurácia posicional, que mede a qualidade da posição geográfica de dados geoespaciais. A análise da qualidade da posicional deve ser avaliada em termos de precisão e tendência. A precisão é feita seguindo as recomendações do Decreto 89.817, ET-ADGV e ET-CQDG, enquanto para a tendência podem ser utilizados testes inferenciais, como o Teste t de Student, e a estatística espacial, usando a média direcional e a variância circular. A presença de *outliers* pode comprometer tanto a precisão quanto a tendência e apesar da importância da detecção de candidatos a *outliers*, essa etapa não é bem definida nas normas, que não a tratam como obrigatória. Dessa forma, objetivou-se analisar o impacto da detecção de candidatos a *outliers* na avaliação da acurácia posicional e avaliar o desempenho de diferentes métodos de detecção de candidatos a *outliers*. Para isso, foram avaliados cinco produtos cartográficos quanto a sua acurácia posicional de duas formas, a primeira sem realizar a etapa de detecção de candidatos a *outliers*, fazendo a análise de precisão e tendência, e a segunda maneira acrescentando a detecção de candidatos a *outliers*, aplicando sete métodos para cada base de dados, sendo eles: Método 3σ , Diagrama Box-Plot, Método Standard Deviation, Método Hampel, Teste de Dixon, Teste de Grubbs e Teste de Chauvenet. Com isso, notou-se que a presença de pontos *outliers* afeta a qualidade do produto cartográfico e que após sua remoção, a classificação dos dados quanto à acurácia posicional tem uma melhoria.

Palavras-chave: *Outliers*. Controle de Qualidade Cartográfica. Acurácia Posicional.

ABSTRACT

CRISTO, Sabrina Lourdes Pereira de, M.Sc., Universidade Federal de Viçosa, August 2021. **Analysis of the Detection and Influence of Outliers in the Evaluation of Positional Accuracy of Cartographic Products.** Adviser: Nerilson Terra Santos. Co-advisers: Afonso de Paula dos Santos and Eduardo Campana Barbosa.

Outliers are data that differ from the dataset to which they belong and can be identified by applying specific methods for this purpose. The detection of possible outliers is something of interest in several areas, including cartographic quality control, which aims to assess the quality of a cartographic product and ensure for what purpose the cartographic product can be used. One of the elements of cartographic quality control is positional accuracy, which measures the quality of the geographic position of geospatial. The analysis of the positional quality must be evaluated in terms of accuracy and trend. Precision is performed following the recommendations of Brazilian Law 89,817 and the ET-ADGV and ET-CQDG standards, as for the trend, inferential tests can be used, such as Student's t-test, and spatial statistics, using directional mean and circular variance. The presence of outliers can compromise both precision and trend. Despite the importance of detecting possible outliers, this step is not well defined in the standards, which do not consider it mandatory. Thus, the objective was to analyze the impact of detecting possible outliers in the assessment of positional accuracy and to evaluate the performance of different methods of detecting possible outliers. For this, five cartographic products were evaluated for their positional accuracy in two ways, the first without performing the possible outlier detection step, but performing the precision and trend analysis, and the second way adding the possible outliers detection, applying seven methods for each database, namely: 3σ , Box-Plot, Standard Deviation, Hampel's, Dixon's Q Test, Grubbs' Test and Chauvenet's Criterion. Thus, it was noted that the presence of outliers affects the quality of the cartographic product and that after its removal, the classification of data regarding positional accuracy has improved.

Keywords: *Outliers*. Cartographic Quality Control. Positional Accuracy.

LISTA DE FIGURAS

Figura 1 - Mundo real (a) com os pontos de referência e produto cartográfico (b) com os respectivos pontos de teste.	18
Figura 2 - Ilustração da comparação entre os dados de referência e de teste onde (a) não existe discrepância posicional, (b) com discrepância posicional.	20
Figura 3 - Diagrama BoxPlot com os limites inferior e superior, os candidatos a <i>outliers</i> detectados e os quartis, sendo a amplitude interquartil a diferença entre Q3 e Q1.	33
Figura 4 - Fluxograma das etapas realizadas para a avaliação da acurácia posicional, com vistas à comparações dos métodos de detecção de <i>outliers</i>	40
Figura 5 - Distribuição espacial dos pontos de controle das cinco bases de dados (AL26, IT30, VC30, VC24 e VC33) em que para VC33 foi feita uma ampliação para visualização dos pontos de teste e referência.	46
Figura 6 - Função K aplicada nas cinco bases de dados.	48
Figura 7 - Comparação entre os métodos descritivos e testes inferenciais.	67

LISTA DE TABELAS

Tabela 1: Critérios de Acurácia Posicional, em que "esc" é a escala da carta testada e "ev" é a equidistância vertical das curvas de nível.....	21
Tabela 2: Classes estabelecidas de acordo com a ET-ADGV e o Decreto 89.817, em que “esc” é a escala do produto cartográfico testado e “ev” é a equidistância das curvas de nível.	22
Tabela 3: Letra código para amostragem lote a lote de acordo com o nível de inspeção, em que tamanho do lote significa a quantidade de células válidas da amostragem.	24
Tabela 4: Tamanho da amostra e limite de qualidade aceitável de acordo com a letra código.	25
Tabela 5: Conversão do limite de qualidade aceitável para qualidade limite.	25
Tabela 6: Tamanho da amostra e número de aceitação da amostra de acordo para amostragem lote isolado, em que tamanho do lote é a quantidade de células válidas da amostragem.....	26
Tabela 7: Equações de Dixon.	35
Tabela 8: Bases de dados.	38
Tabela 9: Estatísticas descritivas para as bases de dados (AL26, IT30, VC30, VC24 e VC33), em que dx representa o valor de discrepância na componente X , dy a discrepância em Y , $d2D$ a discrepância planimétrica resultante e dZ a discrepância em Z	50
Tabela 10: Detecção de <i>outliers</i> pelos métodos descritivos aplicados a base AL26, em que “x” significa que o referido ponto é candidato a <i>outlier</i>	53
Tabela 11: Detecção de <i>outliers</i> pelos métodos descritivos aplicados a base IT30, em que “x” significa que o referido ponto é candidato a <i>outlier</i>	54
Tabela 12: Detecção de <i>outliers</i> pelos métodos descritivos aplicados a base VC30, em que “x” significa que o referido ponto é candidato a <i>outlier</i>	54
Tabela 13: Detecção de <i>outliers</i> pelos métodos descritivos aplicados a base VC24, em que “x” significa que o referido ponto é candidato a <i>outlier</i>	55

Tabela 14: Detecção de <i>outliers</i> pelos métodos descritivos aplicados a base VC33, em que “x” significa que o referido ponto é candidato a <i>outlier</i>	56
Tabela 15: Extrato de discrepâncias obtidas de 10 pontos de um total de 33 pontos da base VC33.....	57
Tabela 16: Resultado do teste de Shapiro-Wilk para as discrepâncias posicionais obtidas para os pontos de checagem de cada uma das cinco bases de dados.	58
Tabela 17: Resultados dos métodos inferenciais para a detecção de candidatos a <i>outliers</i> aplicados às discrepâncias das bases AL26, IT30, VC30, VC24 com dados planimétricos, em que os números identificam os pontos apontados como candidatos a <i>outlier</i>	59
Tabela 18: Resultados dos métodos inferenciais para a detecção de candidatos a <i>outliers</i> aplicados às discrepâncias da base de dados VC33, em que os números identificam os pontos apontados como candidatos a <i>outlier</i>	60
Tabela 19: Avaliação da acurácia posicional planimétrica aplicada para as cinco bases de dados originais e para as bases após a remoção de <i>outliers</i>	62
Tabela 20: Avaliação da acurácia posicional altimétrica aplicada para a base VC33 sem e com remoção de pontos <i>outliers</i>	64
Tabela 21: Coeficiente de variação das cinco bases de dados sem e com a remoção de pontos <i>outliers</i>	65
Tabela 22: Resultado do teste de Shapiro-Wilk para as discrepâncias posicionais obtidas para os pontos de checagem de cada uma das cinco bases de dados após a remoção dos pontos <i>outliers</i>	69
Tabela 23: Base de dados AL26 e suas respectivas discrepâncias posicionais.	77
Tabela 24: Base de dados IT30 e suas respectivas discrepâncias posicionais.....	78
Tabela 25: Base de dados VC30 e suas respectivas discrepâncias posicionais.....	79
Tabela 26: Base de dados VC24 e suas respectivas discrepâncias posicionais.....	80

Tabela 27: Base de dados VC33 e suas respectivas discrepâncias posicionais.....	81
---	----

SUMÁRIO

1. Introdução	14
2. Objetivo	16
2.1. Objetivos específicos.....	16
3. Revisão bibliográfica	17
3.1. Controle de qualidade cartográfica.....	17
3.1.1. Acurácia posicional	17
3.1.2. Decreto 89.817 e ET-CQDG	20
3.1.2.1. Amostragem	22
3.1.2.2. Avaliação do padrão de distribuição espacial dos pontos de checagem	26
3.1.2.3. Etapas da avaliação da acurácia posicional.....	27
3.1.2.4. Teste t de Student	28
3.1.2.5. Média direcional e variância circular	29
3.2. Métodos para detecção de <i>outlier</i> no controle de qualidade cartográfica	30
3.2.1. Métodos descritivos	31
3.2.1.1. Método 3σ	31
3.2.1.2. Diagrama Boxplot	32
3.2.1.3. Método Standard Deviation (SD).....	33
3.2.1.4. Método Hampel.....	34
3.2.2. Métodos inferenciais.....	34
3.2.2.1. Teste de Dixon	35
3.2.2.2. Teste de Grubbs.....	36
3.2.2.3. Teste de Chauvenet	37
4. Materiais e métodos	38
4.1. Materiais	38

4.2. Métodos	39
5. Resultados.....	45
5.1. Análise da distribuição espacial dos pontos de checagem	45
5.2. Cálculo das discrepâncias e análise preliminar a partir das estatísticas descritivas ...	49
5.3. Aplicação de métodos de detecção de <i>outliers</i>	52
5.4. Comparação entre os métodos e remoção de <i>outliers</i>	60
5.5. Avaliação da acurácia posicional	61
6. Discussão.....	65
7. Conclusão.....	73
Referências	74
APÊNDICE 1	77

1. Introdução

Segundo Silva (2004), “um *outlier* é um elemento que desvia de um padrão do conjunto de dados ao qual ele pertence”. Além disso, esses elementos discrepantes acontecem em pequena proporção dentro do conjunto de dados que estão inseridos, pois se ocorressem em grande quantidade não seriam discrepantes.

Quando se suspeita que um dado seja um *outlier*, ou valor discrepante, há duas alternativas: realizar um teste para decidir se os mesmos devem ou não ser mantidos no conjunto ao qual pertencem, ou ignorar a suspeita da discrepância e não removê-los do conjunto de dados. No primeiro caso, a suspeita pode ser avaliada por testes subjetivos ou análise gráfica (FIGUEIRA, 1998).

Quando se utiliza um teste subjetivo e a presença de um *outlier* é confirmada, tem-se outro problema, que é a decisão a respeito da eliminação ou não deste dado. Segundo Figueira (1998), em alguns casos, a presença de um ou mais *outliers* pode comprometer os resultados da pesquisa, mas em outros, pode ser útil para avaliar tanto a qualidade quanto a precisão do estudo. Por isso, deve ser escolhido o teste adequado de acordo com a situação, a origem e o tipo do conjunto de dados em estudo.

A detecção de *outlier* é objeto de interesse de muitas ciências, como por exemplo, a cartografia. Na geração de produtos cartográficos, os dados coletados que serão utilizados para representar o mundo real precisam ser confiáveis, caso contrário, terá uma representação cartográfica que não descreve com precisão a realidade. A avaliação de dados é realizada por meio do controle de qualidade cartográfica, que tem por objetivo analisar a qualidade do produto cartográfico e verificar para qual finalidade ele se adéqua.

No controle de qualidade cartográfica são utilizadas amostras de feições de checagem, que podem ser pontuais ou lineares, para verificar a qualidade de um produto cartográfico. Quando são utilizados pontos, estes são denominados pontos de checagem e são subdivididos em pontos de teste, que são retirados do próprio produto em avaliação, e pontos de referência, que são coletados em campo ou retirados de outra fonte com acurácia maior que o mapa avaliado, sendo os pontos de teste e referência homólogos, ou seja, com a mesma correspondência no campo e no produto cartográfico.

Porém, os pontos de checagem não estão livres de erros, assim como o processo de criação do produto cartográfico. Tais falhas podem alterar as coordenadas geográficas dos pontos, implicando em alteração da localização mais provável deste, prejudicando a qualidade do produto cartográfico final e as análises posteriores em que o produto cartográfico seja empregado. Para contornar esse problema, os pontos de checagem divergentes podem ser identificados utilizando um método de detecção de *outliers* adequado, por isso esta é uma etapa importante a ser feita no processo de avaliação da qualidade.

Assim, no controle de qualidade cartográfica, ao se detectar um *outlier*, primeiramente deve-se analisar se o processo de coleta das feições em campo foi bem executado, assim como se as feições de teste foram bem identificadas. Caso esta identificação não tenha sido bem feita, ou seja, tenha ocorrido um erro grosseiro, deve-se então coletar novamente o dado ou descartar esse ponto de checagem. Mas, se a identificação e a coleta da feição tiverem sido executadas corretamente e mesmo assim for detectado como um *outlier*, este ponto deve permanecer no conjunto de dados, pois é representativo da qualidade posicional dos pontos e importante para o resultado da avaliação (SANTOS et al., 2016).

Apesar da importância da detecção de *outliers*, muitas vezes esta etapa é negligenciada tanto pelos cartógrafos como pelas normas que regulamentam o controle de qualidade cartográfica, que não abordam o tema, já que há poucos estudos com ênfase na detecção de *outliers* aplicado ao controle de qualidade cartográfica. Das pesquisas existentes cita-se o trabalho de Nero (2005), que utiliza o Método 3σ e o trabalho de Carvalho (2017), que faz uma análise da influência dos valores *outliers* na avaliação de modelos digitais de elevação utilizando o método Boxplot ajustado. Essa escassez de trabalhos é ainda maior quando os métodos de detecção de *outliers* são aplicados ao controle de qualidade cartográfica com ênfase na acurácia posicional.

O processo de detecção de *outliers* é aplicado nas discrepâncias posicionais, que são obtidas através dos pontos de teste e de referência. Assim, devem ser escolhidos métodos adequados para identificar os *outliers* presentes nos produtos cartográficos, para garantir a assertividade durante a avaliação do produto. Dessa forma, surge a necessidade de avaliar se os métodos utilizados atualmente pelos cartógrafos, Boxplot e 3σ , são realmente eficientes na detecção de *outliers*, visto que existem métodos descritivos e inferenciais nunca aplicados no controle de qualidade, que podem auxiliar na detecção de candidatos a *outliers* e tornar o processo de avaliação da acurácia posicional mais completo. Além disso, é necessário avaliar

importância da aplicação dos métodos de detecção de *outliers* e a influência dessa detecção no resultado da avaliação de um produto cartográfico.

2. Objetivo

Comparar métodos descritivos e inferenciais utilizados para a detecção de candidatos a *outliers* e avaliar o efeito da remoção de *outliers* detectados por aqueles métodos na acurácia posicional de produtos cartográficos.

2.1. Objetivos específicos

- Comparar métodos descritivos e inferenciais utilizados para a detecção de candidatos a *outliers*;
- Estabelecer um critério para definir quando que um candidato a *outlier* é de fato um *outlier*;
- Aplicar métodos de detecção de *outliers* aos produtos cartográficos e comparar os candidatos a *outlier* obtidos pelos métodos em diferentes bases de dados;
- Analisar a influência da remoção de *outliers* na classificação de um produto cartográfico;
- Avaliar o efeito do agrupamento de pontos na detecção de *outliers*;
- Analisar o desempenho de métodos inferenciais de detecção de *outliers* frente a violação da pressuposição de normalidade.

3. Revisão bibliográfica

3.1. Controle de qualidade cartográfica

Primeiramente deve-se entender o que é um serviço cartográfico, pois é através dele que é gerado o produto cartográfico. Segundo o Decreto nº 89.817 de 1984, um serviço cartográfico envolve toda a operação de reprodução de uma porção ou de toda a superfície terrestre, por meio de imagens, cartas, plantas, dentre outras formas de expressão. A produção cartográfica deve obedecer às normas exigidas neste decreto, como também nas especificações estabelecidas pela Comissão de Cartografia – COCAR (BRASIL, 1984). Em 1994 este órgão sofreu alteração de nomenclatura, passando a se chamar Comissão Nacional de Cartografia – CONCAR (CONCAR, 2020), entretanto, a comissão foi extinta através do Decreto 9.759 de 11 de abril de 2019 (BRASIL, 2019).

Para avaliar a qualidade de um produto cartográfico é feito um controle de qualidade cartográfica, que tem por objetivo verificar o nível de qualidade e determinar para qual finalidade o produto cartográfico pode ser utilizado. Assim, quanto maior for o nível de qualidade de um produto cartográfico, maior a chance de ter acertos nas decisões tomadas ao utilizar esse produto cartográfico (SANTOS, 2010).

A regulamentação do controle de qualidade cartográfica está definida nas Especificações Técnicas de Controle de Qualidade de Dados Geoespaciais (ET-CQDG), que tem por objetivo padronizar a avaliação da qualidade dos produtos cartográficos, descrevendo as medidas, procedimentos e parâmetros usados na avaliação, fazendo com que todos os produtos cartográficos sejam avaliados usando uma estrutura comum. Empregando como parâmetro a norma internacional ISO 19157, a ET-CQDG dispõe os elementos de qualidade a serem avaliados em cinco categorias, sendo eles a completude, consistência lógica, acurácia posicional, acurácia temporal e acurácia temática (DSG, 2016a).

3.1.1. Acurácia posicional

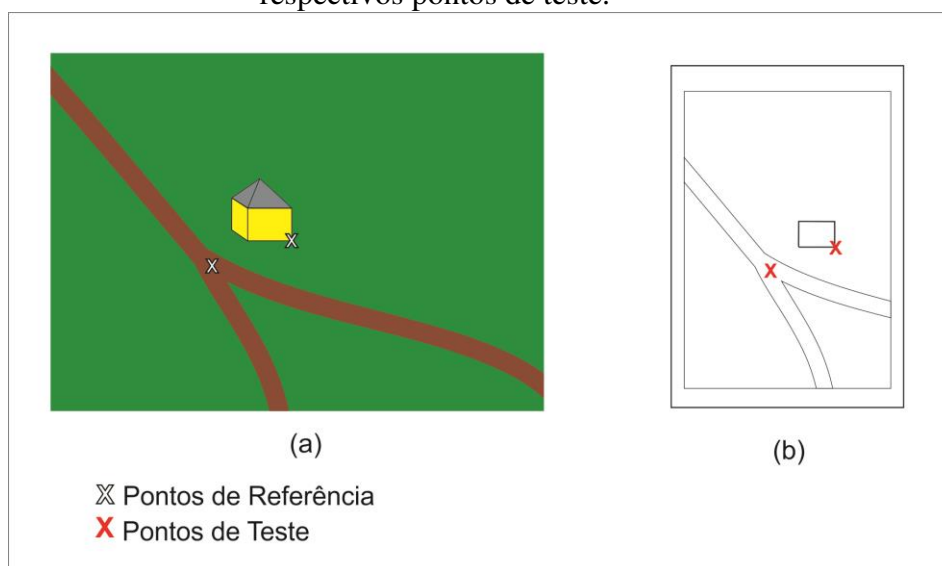
A acurácia posicional, objeto de estudo deste trabalho, pode ser definida como a medida da qualidade da posição geográfica de dados geoespaciais (DSG, 2016a).

Para realizar a avaliação de um produto cartográfico quanto a sua acurácia posicional é necessário fazer uma comparação entre as mensurações de posicionamento dos pontos do

produto a ser avaliado, obtidas no próprio produto cartográfico, e as mensurações dos mesmos pontos obtidas a partir de uma fonte mais confiável, definida como o padrão de comparação. Os pontos desta fonte padrão são denominados de pontos de referência e são obtidos por meio de um levantamento de campo ou de uma coleta sobre um mapa mais acurado. Já os pontos do produto em avaliação são chamados de pontos de teste. Ressalta-se que os pontos de teste e referência devem ser homólogos, ou seja, com mesma correspondência espacial (NERO, 2005).

A Figura 1 ilustra os pontos de checagem em que se tem (a) o mundo real com os pontos de referência coletados no campo e (b) o produto cartográfico em avaliação com seus respectivos pontos de teste, sendo os pontos de checagem em posições homólogas. Dá-se o nome de pontos de checagem o conjunto formado pelos pontos de referência e de teste.

Figura 1 - Mundo real (a) com os pontos de referência e produto cartográfico (b) com os respectivos pontos de teste.



Fonte: A autora.

Segundo Nero (2005), para a aquisição dos pontos de referência, deve ser usado um método que seja pelo menos três vezes mais acurado que o procedimento usado para determinação dos pontos de teste no produto avaliado. As coordenadas dos pontos de referência podem ser obtidas por meio de levantamentos de campo utilizando topografia, geodésia ou outro método de posicionamento, desde que se tenha cuidado na escolha da

localização dos pontos e na sua correta identificação no campo, para que sejam efetivamente pontos de referência confiáveis para serem usados na avaliação.

Os pontos de teste são obtidos no próprio produto cartográfico em avaliação, visto que todo mapa, carta ou ortofoto deve conter um grid de coordenadas geográficas que permita a determinação das coordenadas dos pontos desejados. Assim, de posse dos pontos de referência e dos pontos de teste, é possível obter a avaliação da acurácia posicional.

A acurácia posicional pode ser classificada como absoluta (ou externa), relativa (ou interna) ou do grid. De acordo com a ET-CQDG (DSG, 2016a), a acurácia:

Absoluta ou externa: refere-se ao caso em que há discrepância entre as coordenadas da representação da feição e as coordenadas das feições na fonte de maior precisão; **Relativa ou interna:** refere-se ao caso em que há discrepância entre as posições relativas dos objetos e as posições relativas reais ou aceitas como sendo certas; e **Do grid:** consiste da aplicação do conceito de acurácia posicional absoluta para os dados que se caracterizam por estarem em uma malha regular. Para todos os efeitos desta Especificação Técnica serão enquadrados nesta classe as avaliações que forem feitas sobre malhas de pontos de Modelos Digitais de Elevação (MDE).

Contudo, a ET-CQDG trata somente do primeiro e terceiro caso, visto que considera que os erros posicionais relativos decorrem dos erros posicionais absolutos e ainda que o erro relativo é também um erro absoluto, não fazendo distinção entre esses tipos de erros. (DSG, 2016a).

O cálculo da acurácia posicional absoluta se baseia no cálculo das diferenças entre as coordenadas dos pontos teste e de referência, sendo este resultado conhecido como discrepância posicional, que pode ser obtida através das equações 1 a 4, que descrevem o cálculo para as discrepâncias (d_i) altimétrica, X, Y e a resultante planimétrica, respectivamente.

$$dz_i = Z_{Ti} - Z_{Ri} \quad (1)$$

$$dx_i = X_{Ti} - X_{Ri} \quad (2)$$

$$dy_i = Y_{Ti} - Y_{Ri} \quad (3)$$

$$d_{2Di} = \sqrt{(dx_i)^2 + (dy_i)^2} \quad (4)$$

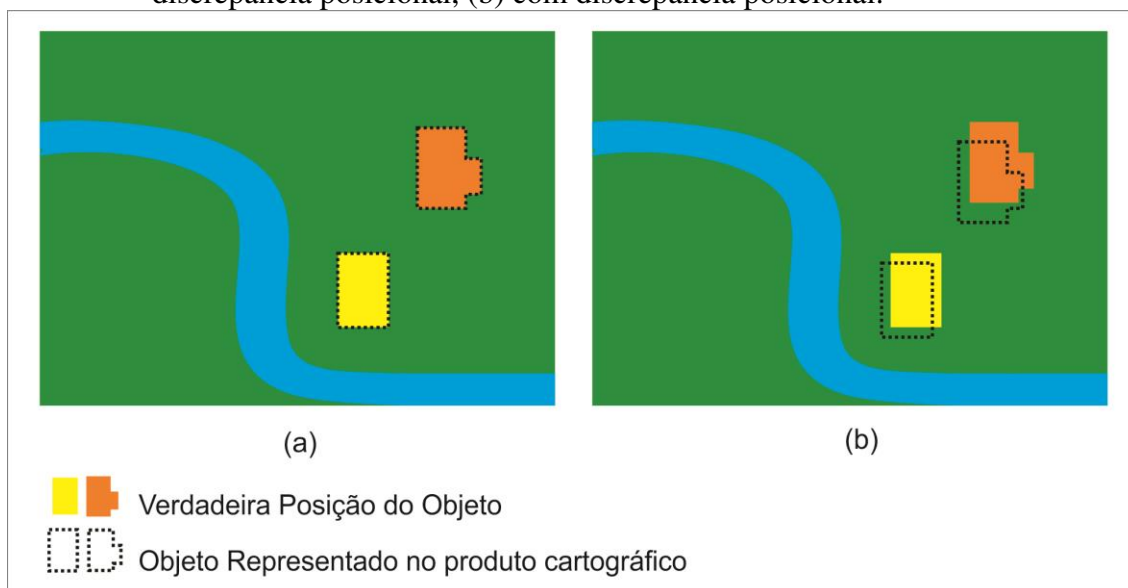
Em que:

- Z_{Ti} é a coordenada Z (ou altimétrica) do i-ésimo ponto de teste;
- Z_{Ri} é a coordenada Z (ou altimétrica) do i-ésimo ponto de referência;

- X_{Ti} é a coordenada X do i-ésimo ponto de teste;
- X_{Ri} é a coordenada X do i-ésimo ponto de referência;
- Y_{Ti} é a coordenada Y do i-ésimo ponto de teste;
- Y_{Ri} é a coordenada Y do i-ésimo ponto de referência.

A Figura 2 ilustra dois casos sendo que em um deles não há discrepância (a) entre as coordenadas de checagem e em outro caso há discrepância (b).

Figura 2 - Ilustração da comparação entre os dados de referência e de teste onde (a) não existe discrepância posicional, (b) com discrepância posicional.



Fonte: A autora.

Após o cálculo de suas discrepâncias, o produto cartográfico é classificado quanto a sua acurácia posicional através da comparação entre as coordenadas dos pontos de teste e referência, seguindo as condições descritas no Decreto 89.817, ET-CQDG e ET-ADGV (Especificação Técnica para Aquisição de Dados Geoespaciais). Com base nessa acurácia é possível definir para qual finalidade o produto cartográfico pode ser utilizado.

3.1.2. Decreto 89.817 e ET-CQDG

A classificação do produto cartográfico em relação a sua acurácia posicional é descrita pelo Decreto 89.817 de 1984 nos artigos 8º e 9º. A avaliação de um produto cartográfico é feita considerando a escala (para produtos planimétricos) e a equidistância vertical entre

curvas de nível (para produtos altimétricos), utilizando duas tolerâncias chamadas de Padrão de Exatidão Cartográfica (PEC) e Erro-Padrão (EP). O Decreto 89.817 classifica os produtos cartográficos gerados em três classes, quais sejam A, B e C.

Para classificar um produto cartográfico para uma dada classe e escala de interesse, é necessário que os dados planimétricos do produto avaliado atenda a duas condições: 1) 90% das discrepâncias calculadas tenham valores iguais ou inferiores ao PEC planimétrico; 2) que a raiz quadrada do erro quadrático médio seja menor ou igual ao valor do EP planimétrico. Já para dados altimétricos, é necessário que 90% das discrepâncias altimétricas tenham valor inferior ou igual ao PEC altimétrico e a raiz quadrada do erro quadrático médio menor ou igual que o EP altimétrico, de acordo com a equidistância vertical entre as curvas de nível do produto avaliado (BRASIL, 1984).

A Tabela 1 apresenta os critérios para o enquadramento do produto cartográfico em uma das classes, bem como os respectivos valores do PEC e EP definidos pelo Decreto 89.817.

Tabela 1: Critérios de Acurácia Posicional, em que "esc" é a escala da carta testada e "ev" é a equidistância vertical das curvas de nível.

Classe Decreto-Lei	Planimetria		Altimetria	
	PEC	EP	PEC	EP
A	$\leq 0,5 \text{ mm} * \text{esc}$	$\leq 0,3 \text{ mm} * \text{esc}$	$\leq 1/2 * \text{ev}$	$\leq 1/3 * \text{ev}$
	$> 0,5 \text{ mm} e \leq 0,8 \text{ mm} * \text{esc}$	$> 0,3 \text{ mm} e \leq 0,5 \text{ mm} * \text{esc}$	$> 1/2 e \leq 3/5 * \text{ev}$	$> 1/3 e \leq 2/5 * \text{ev}$
B	$> 0,8 \text{ mm} e \leq 1,0 \text{ mm} * \text{esc}$	$> 0,5 \text{ mm} e \leq 0,6 \text{ mm} * \text{esc}$	$> 3/5 e \leq 3/4 * \text{ev}$	$> 2/5 e \leq 1/2 * \text{ev}$
C				

Segundo Santos (2010), o Decreto 89.817 possui algumas inconsistências, como, por exemplo, a falta do estabelecimento de testes a serem usados para realizar o controle de qualidade cartográfica. Além disso, o Decreto 89.817 considera que as estatísticas descritivas desvio padrão, erro padrão e erro quadrático médio tem o mesmo significado, o que estatisticamente não é verdade. Tal ambiguidade torna a norma confusa o que pode resultar em diferentes interpretações e, conseqüentemente, aplicada de diferentes maneiras pelos cartógrafos.

O Decreto 89.817 de 1984 criou o PEC com o objetivo de quantificar a qualidade cartográfica do produto analógico. Porém, com a crescente evolução tecnológica viu-se a necessidade de se estabelecer novos critérios para avaliação dos produtos cartográficos

digitais. Assim, a ET-ADGV criou uma classe mais restritiva denominada de PEC-PCD, além da norma ET-CQDG que descreve o processo metodológico para a classificação da acurácia posicional estabelecida pelo Decreto 89.817.

As classes criadas pela ET-ADGV para avaliação do PEC estão apresentadas na Tabela 2 (DSG, 2016b).

Tabela 2: Classes estabelecidas de acordo com a ET-ADGV e o Decreto 89.817, em que “esc” é a escala do produto cartográfico testado e “ev” é a equidistância das curvas de nível.

Decreto 89.817	ET- ADGV	Planimetria		Altimetria	
		PEC	EP	PEC	EP
	A	$\leq 0,28mm * esc$	$\leq 0,17mm * esc$	$\leq 0,27 * ev$	$\leq 1/6 * ev$
		$> 0,28mm e \leq 0,5 mm$	$> 0,17 e$	$> 0,27 e$	$> 1/6 e$
A	B	$* esc$	$\leq 0,3 mm * esc$	$\leq 1/2 * ev$	$\leq 1/3 * ev$
		$> 0,5mm e \leq 0,8 mm$	$> 0,3 mm e$	$> 1/2 e$	$> 1/3 e$
B	C	$* esc$	$\leq 0,5 mm * esc$	$\leq 3/5 * ev$	$\leq 2/5 * ev$
		$> 0,8mm e \leq 1,0 mm$	$> 0,5mm e$	$> 3/5 e$	$> 2/5 e$
C	D	$* esc$	$\leq 0,6 mm * esc$	$\leq 3/4 * ev$	$\leq 1/2 * ev$

Assim, através das Tabelas 1 e 2 é possível classificar um produto cartográfico quanto a sua acurácia posicional. As etapas que devem ser realizadas para classificar o produto cartográfico de acordo com a ET-CQDG (Tabela 2) são descritas a seguir.

3.1.2.1. Amostragem

Para avaliar um produto cartográfico em relação a sua acurácia posicional, devem-se obter pontos de referência e seus respectivos pontos de teste. Porém, devido à impossibilidade de mapear todos os pontos de checagem da superfície terrestre que está representada no produto cartográfico para fazer a avaliação, faz-se uma amostragem, com o objetivo de representar toda a população (SANTOS, 2010).

Para realizar a amostragem, devem-se compreender quatro conceitos fundamentais, que foram baseados nas normas ISO 2859-1 e ISO 2859-2, os quais são: lote, item, limite de qualidade aceitável (LQA) e qualidade limite (QL). O lote é definido como a unidade mínima em que podem ser aplicados os procedimentos de qualidade, é o produto cartográfico a ser avaliado. Já o item é a “menor unidade do lote que pode ser inspecionada”, são as feições geográficas representadas no produto cartográfico. O LQA é o pior resultado aceitável em

uma avaliação lote a lote e, por fim, a QL é o nível de qualidade de um lote isolado (DSG, 2016a).

No processo de amostragem, o produto avaliado é particionado em células de 4x4cm na escala do produto cartográfico, usando sempre valores inteiros. Por exemplo, um mapa com escala 1:10000 terá células de 400x400m. Após a partição, é verificado cada célula a fim de determinar se ela fará parte da população que será avaliada no lote. Em uma amostragem de pontos de checagem, é considerada uma célula válida apenas aquelas que possuam feições em que é possível identificar um ponto no terreno, por exemplo edificações, cruzamento de vias, cruzamento, curvas com cercas, entre outros (DSG, 2016a).

Em relação ao número de pontos necessários, a ET-CQDG adota como referência a ISO 2859-1, para avaliar um conjunto de lotes com valor mínimo de 10 lotes, realizando a inspeção lote a lote e a ISO 2859-2, que avalia lotes isolados entre um e nove. Cada uma dessas normas apresenta tabelas para definir o tamanho da amostra (n) e o número de aceitação tolerável (Ac) dentro da amostra (DSG, 2016a). Assim, se o número de lotes for inferior a 10, o que acontece nos casos em que é avaliado uma ortoimagem ou imagem de satélite por exemplo, deve-se usar a metodologia de lotes isolados, mas se o número de lotes for igual ou maior que 10, quando é avaliada uma base cartográfica que tem 10 layers ou mais, ou seja, mais de 10 grupos de feições cartográficas como hidrografia, vias, edificações, etc, usa-se amostragem lote a lote.

Se for realizar uma inspeção lote a lote, deve-se consultar a Tabela 3 considerando o tamanho do lote e o nível de inspeção. A ET-CQDG especifica que o nível geral de inspeção II deve ser utilizado, assim, obtém a letra código para o tamanho da amostra.

Tabela 3: Letra código para amostragem lote a lote de acordo com o nível de inspeção, em que tamanho do lote significa a quantidade de células válidas da amostragem.

Tamanho do Lote	Níveis especiais de inspeções				Níveis gerais de inspeções		
	S-1	S-2	S-3	S-4	I	II	III
2 a 8	A	A	A	A	A	A	B
9 a 15	A	A	A	A	A	B	C
16 a 25	A	A	B	B	B	C	D
26 a 50	A	B	B	C	C	D	E
51 a 90	B	B	C	C	C	E	F
91 a 150	B	B	C	D	D	F	G
151 a 280	B	C	D	E	E	G	H
281 a 500	B	C	D	E	F	H	J
501 a 1200	C	C	E	F	G	J	K
1201 a 3200	C	D	E	G	H	K	L
3201 a 10000	C	D	F	G	J	L	M
10001 a 35000	C	D	F	H	K	M	N
35001 a 150000	D	E	G	J	L	N	P
150001 a 500000	D	E	G	J	M	P	Q
500001 ou mais	D	E	H	K	N	Q	R

Fonte: ISO (1999).

Em seguida, deve-se definir qual será o limite de qualidade aceitável (LQA) para a avaliação, de modo a encontrar o valor do tamanho amostral (n) e valor de A_c (Tabela 4).

Tabela 4: Tamanho da amostra e limite de qualidade aceitável de acordo com a letra código.

Letra Código	Tamanho da amostra	Limite de qualidade aceitável (LQA) em %									
		0,4	0,65	1,0	1,5	2,5	4,0	6,5	10	15	25
A	2	↓	↓	↓	↓	↓	↓	0	↓	↓	1
B	3	↓	↓	↓	↓	↓	0	↑	↓	1	2
C	5	↓	↓	↓	↓	0	↑	↓	1	2	3
D	8	↓	↓	↓	0	↑	↓	1	2	3	5
E	13	↓	↓	0	↑	↓	1	2	3	5	7
F	20	↓	0	↑	↓	1	2	3	5	7	10
G	32	0	↑	↓	1	2	3	5	7	10	14
H	50	↑	↓	1	2	3	5	7	10	14	21
J	80	↓	1	2	3	5	7	10	14	21	↑
K	125	1	2	3	5	7	10	14	21	↑	↑
L	200	2	3	5	7	10	14	21	↑	↑	↑
M	315	3	5	7	10	14	21	↑	↑	↑	↑
N	500	5	7	10	14	21	↑	↑	↑	↑	↑
P	800	7	10	14	21	↑	↑	↑	↑	↑	↑
Q	1250	10	14	21	↑	↑	↑	↑	↑	↑	↑
R	2000	14	21	↑	↑	↑	↑	↑	↑	↑	↑

Fonte: ISO (1999).

Se o tamanho da amostra não tiver um valor para o nível de aceitação, ou seja, cair nas setas, o valor apontado pela seta deve ser adotado pelo cartógrafo. Caso seja feita a inspeção lote isolado, o primeiro passo é converter o LQA determinado para QL, através da Tabela 5:

Tabela 5: Conversão do limite de qualidade aceitável para qualidade limite.

Tamanho do lote	LQA (%)		
	1,0	4,0	10
16 a 25	12,5	32	32
26 a 50	12,5	20	32
51 a 150	8,0	20	32
151 a 1200	5,0	20	32
1201 a 10000	3,15	12,5	20
10001 a 150000	3,15	8,0	20
150001 e maiores	2,0	8,0	20

Fonte: ISO, 1985.

Depois, de posse do tamanho do lote e do QL, basta consultar estes valores na Tabela 6 e será encontrado o tamanho da amostra (n) e o número de aceitação (Ac).

Tabela 6: Tamanho da amostra e número de aceitação da amostra de acordo para amostragem lote isolado, em que tamanho do lote é a quantidade de células válidas da amostragem.

Tamanho do lote		Qualidade Limite (QL) em %								
		0,8	1,25	2,0	3,15	5,0	8,0	12,5	20	32
16 a 25	n						17	13	9	6
	Ac	↓	↓	↓	↓	↓	0	0	0	0
26 a 50	n				50	28	22	15	10	6
	Ac	↓	↓	↓	0	0	0	0	0	0
51 a 90	n			50	44	34	24	16	10	8
	Ac	↓	↓	0	0	0	0	0	0	0
91 a 150	n		90	80	55	38	26	18	13	13
	Ac	↓	0	0	0	0	0	0	0	1
151 a 280	n	170	130	95	65	42	28	20	20	13
	Ac	0	0	0	0	0	0	0	1	1
281 a 500	n	220	155	105	80	50	32	32	20	20
	Ac	0	0	0	0	0	0	1	1	3
501 a 1200	n	255	170	125	125	80	50	32	32	32
	Ac	0	0	0	1	1	1	1	3	5
1201 a 3200	n	280	200	200	125	125	80	50	50	50
	Ac	0	0	1	1	3	3	3	5	10
3201 a 10000	n	315	315	200	200	200	125	80	80	80
	Ac	0	1	1	3	5	5	5	10	18
10001 a 35000	n	500	315	315	315	315	200	125	125	↑
	Ac	1	1	3	5	10	10	10	18	
35001 a 150000	n	500	500	500	500	500	315	200	↑	↑
	Ac	1	3	5	10	18	18	18		
150001 a 500000	n	800	800	800	800	↑	↑	↑	↑	↑
	Ac	3	5	10	18					
500001 e maiores	n	1250	1250	1250						
	Ac	5	10	18	↑	↑	↑	↑	↑	↑

Fonte: ISO (1985).

3.1.2.2. Avaliação do padrão de distribuição espacial dos pontos de checagem

Após definir o planejamento de quais pontos de checagem serão utilizados, é necessário avaliar o padrão de distribuição dos pontos de checagem antes de realizar a coleta de campo deles, pois Santos et al. (2016) recomendam que os pontos de checagem não apresentem padrão de distribuição espacial aleatório ou disperso. A avaliação do padrão é realizada usando técnicas de estatística espacial como o vizinho mais próximo ou função K de Ripley, conforme sugerido por Santos (2010). Esse planejamento é importante, pois reduz

custo e tempo e garante que os pontos de checagem coletados serão efetivamente úteis na avaliação e constituem uma amostra representativa de toda a área em estudo. Após a verificação do padrão de distribuição espacial dos pontos de checagem, e a subsequente coleta deles, pode-se prosseguir para a avaliação da acurácia posicional.

3.1.2.3. Etapas da avaliação da acurácia posicional

Finalizada a fase de definição do número de pontos de checagem e a escolha dos locais de coleta, pode-se dar início a avaliação da acurácia posicional planimétrica, seguindo os procedimentos descritos pela ET-CQDG, sendo eles (DSG, 2016a):

- 1) Obter os pontos de referência, que devem ter precisão três vezes superior dos pontos de teste e dos demais dados que foram utilizados para gerar o produto cartográfico;
- 2) Calcular a discrepância de cada componente planimétrica (equações 2, 3 e 4) sendo os pontos de teste denotados pelo sufixo “T” e os de referência pelo sufixo “R”, resultando na discrepância posicional dos erros d_i , descrita na Equação 5;
- 3) Calcular a Raiz quadrada do erro quadrático médio (RMS);

$$RMS = \sqrt{\frac{\sum_{i=1}^n (d_i)^2}{n}} \quad (5)$$

- 4) Comparar cada valor de d_i com o PEC estabelecido de cada classe da Tabela 2. O produto se enquadra onde tiver pelo menos 90% de d_i inferior ou igual ao PEC da classe testada.
- 5) Comparar o valor de RMS com o erro-padrão da classe que foi identificada no passo 4. Se o RMS for menor ou igual ao EP, tem-se como resultado a classe selecionada, caso contrário, deve-se testar com a classe abaixo, até encontrar uma classe que satisfaça as condições dos passos 4 e 5.

Para obter a acurácia posicional altimétrica, faz-se os mesmos procedimentos descritos acima, alterando apenas o cálculo do item 2 para a discrepância altimétrica como demonstrado na Equação 1. Salienta-se que há um erro conceitual na norma supracitada, pois no item 2 é utilizado pela ET-CQGD o termo “Erro médio quadrático (EMQ)” que deveria ser escrito como “Raiz quadrada do erro quadrático médio”.

A análise da acurácia posicional envolve a utilização de diversos métodos estatísticos, entre eles, citam-se análise de precisão, tendência e detecção de *outliers*. Monico et al. (2009), salientam que a acurácia envolve os efeitos sistemáticos e aleatórios, enquanto a precisão está associada apenas a efeitos aleatórios e a tendência relacionada com os efeitos sistemáticos, com isso, tem-se que a acurácia envolve a medida de precisão e tendência. Ao aplicar estas definições ao PEC, tem-se que a análise da acurácia é realizada em termos de precisão e tendência.

Dessa forma, para que um produto cartográfico tenha uma boa qualidade posicional das informações representadas, este deve ser preciso e livre de tendência. Para verificar a tendência, Nero (2005) sugere a utilização de um teste para avaliar se a média das discrepâncias é nula. Para essa finalidade, pode ser utilizado o teste t de Student ou o teste Z dependendo do conhecimento ou não da variância populacional. Como esses testes exigem normalidade, recomenda-se que *a priori* seja avaliado se as discrepâncias seguem uma distribuição normal.

No trabalho de Santos et al. (2016), para verificação da normalidade dos dados foi utilizado o teste de Shapiro-Wilk. Já para a análise de tendência foram utilizados os testes t de Student e as estatísticas espaciais média direcional e variância circular. No mesmo estudo, os autores aplicaram as metodologias do Decreto 89.817 para realizar a análise de precisão. Portanto, segundo a metodologia apresentada, um dado espacial só é considerado acurado se ele for preciso e livre de tendência.

3.1.2.4. Teste t de Student

O teste t de Student para tendência é aplicado sobre as discrepâncias de cada componente posicional (d_x, d_y, d_{2D} e d_z). Esse teste, considerando como exemplo a componente X, tem como hipótese de nulidade $H_0: \mu_{d_x} = 0$, ou seja, a componente X não apresenta tendência. Já a hipótese alternativa $H_A: \mu_{d_x} \neq 0$, ou seja, a componente X apresenta tendência.

A estatística é calculada de acordo com as Equações 6, 7 e 8:

$$t_{calc\ x} = \frac{\bar{d}_x}{\frac{s_{d_x}}{\sqrt{n}}} \quad (6)$$

Em que,

- $\bar{d}_x$ é a média das discrepâncias para a componente X , ou seja,

$$\bar{d}_x = \frac{\sum_{i=1}^n d_{xi}}{n} \quad (7)$$

- s_{d_x} é o desvio padrão das discrepâncias da componente X das discrepâncias, ou seja,

$$s_{d_x} = \sqrt{\frac{\sum_{i=1}^n (d_{xi} - \bar{d}_x)^2}{n - 1}} \quad (8)$$

- n é o número de amostras.

Se $|t_{calc\ x}| \geq t_{tab}$ rejeita-se a hipótese nula, caso contrário, não rejeita a hipótese nula. Para encontrar o valor tabelado, deve-se consultar uma tabela bilateral para o teste t de Student e verificar o valor de acordo com o número de amostras e o nível de significância:

$$t_{tab} = t_{n-1, \frac{\alpha}{2}}.$$

Para as demais componentes Y e Z seguem-se os mesmos passos supracitados para a análise de tendência.

3.1.2.5. Média direcional e variância circular

Uma vantagem da utilização da média direcional é que por ser uma estatística espacial, não tem a pressuposição de normalidade.

Para seu cálculo é necessário que sejam feitas retas ligando os pontos de referência aos respectivos pontos de teste e, em seguida, calcula-se o valor do azimuth do alinhamento formado (WONG; LEE, 2005). Por definição, o azimuth é o ângulo horário medido no plano horizontal com início na referência Norte até o alinhamento do vetor. Dessa forma, a média direcional θ_R é obtida pela Equação 9.

$$\tan(\theta_R) = \frac{\sum \sin(\theta_V)}{\sum \cos(\theta_V)} \quad (9)$$

Em que θ_V é o azimuth. Entretanto, a média circular apenas descreve a tendência dos dados em uma direção. Dessa forma, para avaliar a tendência para mais de uma direção é necessário combinar a média direcional com a variância circular, que é calculada a partir do comprimento do vetor resultante (C_R), conforme mostrado na Equação 10.

$$C_R = \sqrt{(\sum \cos(\theta_V))^2 + (\sum \sin(\theta_V))^2} \quad (10)$$

A partir do comprimento do vetor resultante, calcula-se o valor da variância circular que pode assumir valores entre 0 e 1.

$$S_c = 1 - \frac{C_R}{n} \quad (11)$$

Em que n é o número de vetores. Se o valor resultante da Equação 11 for próximo de 0, significa que a média direcional teve resultado significativo e que há tendência nos dados, mas se o valor for próximo de 1, isso indica que os vetores têm direções opostas e que a média direcional não é representativa.

Especificamente, este trabalho se concentra na avaliação dos diferentes métodos estatísticos para a detecção de *outliers*, pois a presença dos *outliers* pode comprometer tanto a avaliação da tendência bem como da precisão de um produto cartográfico.

3.2. Métodos para detecção de *outlier* no controle de qualidade cartográfica

Os métodos para a detecção de *outliers* se baseiam nas discrepâncias posicionais dos pontos de referência e dos pontos de teste. A obtenção destas discrepâncias (d_i) foi estabelecida nas equações 1, 2, 3 e 4. Os métodos se baseiam em diferentes critérios para avaliar se cada um dos valores d_i representam discrepâncias candidatas à *outliers* ou não.

Discrepâncias candidatas à *outliers* detectadas por um ou mais métodos podem levar a uma das três conclusões: 1) o levantamento de campo dos pontos de referência não foi bem executado e contém erros grosseiros por desatenção do cartógrafo; ou 2) o produto cartográfico avaliado tem os valores de coordenadas de teste muito distintos dos verdadeiros valores que deveriam assumir; ou 3) o cartógrafo identificou o ponto de teste ou de referência no local errado. Caso seja constatado o primeiro ou terceiro caso, os valores *outliers* devem ser excluídos e se necessário uma nova coleta de pontos de teste e de referência deve ser feita. Por outro lado, se constatado o segundo caso, ou seja, se foi feita uma análise e as coordenadas de teste e referência estão corretas, os *outliers* não podem ser excluídos, pois eles serão importantes para definição da qualidade do produto cartográfico.

Os métodos para a detecção de discrepâncias candidatas à *outliers* podem ser classificados como métodos descritivos e inferenciais. Cada método será aplicado sobre a discrepância de cada componente posicional, sendo elas X , Y , Z e $2D$.

3.2.1. Métodos descritivos

Os métodos descritivos se baseiam no cálculo de estatísticas descritivas das discrepâncias. Essas estatísticas são específicas para cada método.

3.2.1.1. Método 3σ

O Método 3σ é um dos critérios mais conhecidos para detectar candidatos a *outliers*. Porém, esse método pode não ter um desempenho satisfatório nos casos em que se dispõe de uma amostra pequena, como, por exemplo, com menos de dez elementos amostrais. Nestes casos, a média e o desvio-padrão das discrepâncias se tornam muito sensíveis a *outliers* (ZHAO et al., 2013).

Tem-se como premissa que discrepâncias candidatas à *outliers* ocorrem em pequenas proporções em um conjunto de dados. Segundo Ghilani (2010), quaisquer dados que diferem da média em mais de 3σ podem ser considerados como erros e devem ser removidos de um conjunto de dados. Este autor ainda afirma que rejeitar observações maiores que 3σ significa que cerca de 99,7% de todas as observações serão mantidas. Em outras palavras, apenas cerca de 0,3% das observações em um conjunto de erros aleatórios normalmente distribuídos estão fora do intervalo $\pm 3\sigma$.

Para esse método, uma discrepância é considerada como uma candidata a *outlier* se satisfizer o critério estabelecido na Equação 12 (ZHAO et al., 2013).

$$|d_i| > |\bar{d}| + 3 * s \quad (12)$$

Em que d_i ($i = 1, 2, \dots, n$) é a discrepância estabelecida nas equações 1, 2, 3 e 4, $\bar{d}$ é média das discrepâncias de cada componente posicional testada (d_x, d_y, d_{2D} e d_z) e s é o desvio-padrão amostral das discrepâncias de cada componente posicional testada.

Apesar de ser originalmente conhecido como 3σ , o método se baseia no desvio-padrão amostral, o qual é denotado por s e não σ , que normalmente representa o desvio-padrão populacional não conhecido. A notação s foi utilizada para manter a coerência em relação aos demais métodos, que também utilizam o desvio-padrão amostral.

O Método 3σ é também aplicado na Cartografia para a detecção de candidatos a *outliers*. Nesta aplicação é analisado se a discrepância posicional é superior à média das discrepâncias da componente testada, somado ao valor de três vezes o erro-padrão (EP)

(NERO, 2005), que tem seu valor presente no Decreto n° 89.817, conforme mostrado na Tabela 1. Neste caso, utiliza-se geralmente o valor contido na classe A e a escala do produto cartográfico a ser avaliado.

Entretanto, muitos autores que utilizam esse método no controle de qualidade cartográfica, aqui denominado de 3σ Cartográfico, fazem uma supressão da média, conforme descrito na Equação 13.

$$|d_i| > 3 * EP \quad (13)$$

O motivo da supressão da média da Equação 13 é explicado pelo fato de que, quando é realizado o controle da qualidade posicional, espera-se que as discrepâncias posicionais sejam mínimas ou nulas. Estes valores mínimos indicariam um produto com boa qualidade posicional, portanto, com a média das discrepâncias também assumindo um valor mínimo.

Partindo desse princípio, para aplicação no controle de qualidade cartográfica será utilizada nesse trabalho, para efeitos de comparação, a Equação 12 sem a média e utilizando o desvio-padrão, sendo denominado de método 3σ Alternativo:

$$|d_i| > 3 * s \quad (14)$$

Dessa forma, ter-se-á uma equação similar à utilizada na Cartografia (Equação 13), mas utilizando o desvio-padrão que é o proveniente do método tradicional.

3.2.1.2. Diagrama Boxplot

O diagrama de Boxplot, ou gráfico de caixa e bigodes, pode ser utilizado para analisar a distribuição das discrepâncias. Ele é obtido a partir da ordenação crescente dos valores das discrepâncias (d_i) e a identificação dos percentis 25, 50 e 75. Estes percentis são comumente conhecidos como primeiro quartil (Q_1), mediana e terceiro quartil (Q_3).

A diferença entre o terceiro e primeiro quartil é denominada amplitude interquartil (AIQ), que compreendem 50% das discrepâncias. Uma propriedade interessante da amplitude interquartil é que ela é resistente em relação ao efeito das distorções causadas pelas discrepâncias candidatas à *outliers* (WITTE; WITTE, 2005).

Para construir os bigodes da caixa, faz-se necessário o cálculo dos limites inferior (LI) e superior (LS) definidos nas Equações 16 e 17, respectivamente. Os bigodes se estendem dos quartis aos limites inferior e superior.

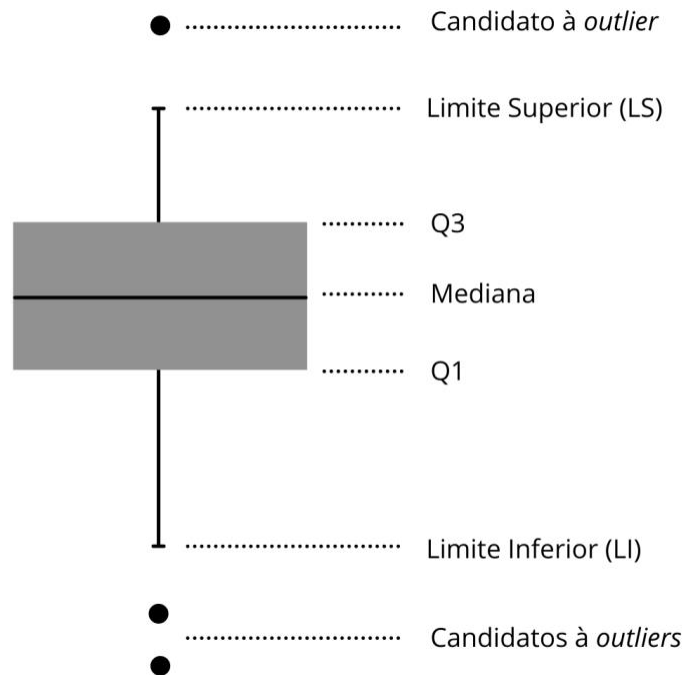
$$AIQ = Q_3 - Q_1 \quad (15)$$

$$LI = Q_1 - 1,5 * AIQ \quad (16)$$

$$LS = Q_3 + 1,5 * AIQ \quad (17)$$

Ao usar o gráfico de caixa e bigodes, uma discrepância é considerada ser candidata a *outlier* se seu valor for maior que o limite superior ou menor que o limite inferior. A Figura 3 ilustra um gráfico de caixa e bigodes.

Figura 3 - Diagrama BoxPlot com os limites inferior e superior, os candidatos a *outliers* detectados e os quartis, sendo a amplitude interquartil a diferença entre Q3 e Q1.



Fonte: A autora.

3.2.1.3. Método Standard Deviation (SD)

O método *Standard Deviation* é considerado clássico e simples de usar para detecção de candidatos a *outliers*. Basicamente este método utiliza de 2 limites, 2SD e 3SD, que podem ser obtidos com base nas Equações 18 e 19 respectivamente (SEO; MARSH, 2006).

$$2D \text{ METHOD} = \bar{d} \pm 2 * s \quad (18)$$

$$3D \text{ METHOD} = \bar{d} \pm 3 * s \quad (19)$$

Em que $\bar{d}$ é a média amostral das discrepâncias e s o desvio padrão amostral das discrepâncias.

Para cada uma das Equações 18 e 19, tem-se como resultado dois limites, o superior calculado utilizando a parte da soma e o inferior utilizando a subtração. Dessa forma, são consideradas candidatas a *outliers* as discrepâncias que ficarem acima do limite superior ou abaixo do limite inferior, ou seja, fora dos intervalos definidos nas equações.

Diferentemente do Método 3σ em que a discrepância é comparada utilizando o seu módulo, conforme mostra a parte esquerda da Equação 12, no método *Standard Deviation* as discrepâncias são comparadas utilizando o sinal que acompanha seu valor.

3.2.1.4. Método Hampel

Segundo Freitas (2019), o método Hampel aponta as discrepâncias candidatas à *outliers* utilizando somente a mediana das discrepâncias consideradas e o desvio mediano absoluto. Além disso, este método pode ser usado para vários tipos de conjunto de dados, sem restrição. A aplicação do método é descrita a seguir (OLIVEIRA; CAROLI; AMARAL, 2014):

- 1) Calcula-se a mediana (Me) das discrepâncias;
 - 2) Calcula-se o desvio médio (r_i) das discrepâncias;
- $$r_i = (d_i - Me) \quad (20)$$
- 3) Calcula-se a mediana dos desvios médios r_i ($Me_{|r_i|}$);
 - 4) Para cada ponto “ i ”, avaliar se a condição $|r_i| \geq 4,5 * Me_{|r_i|}$ é verdadeira. Caso afirmativo, a discrepância d_i é considerada candidata a *outlier*.

3.2.2. Métodos inferenciais

Os métodos inferenciais se baseiam no estabelecimento de duas hipóteses estatísticas, na obtenção de estatísticas (calculada e tabelada) e na comparação destas estatísticas para decidir por uma das duas hipóteses estabelecidas. Embora os métodos se assemelhem nos passos a serem seguidos, eles diferem nas hipóteses estabelecidas, na obtenção das estatísticas calculada e tabelada e, também, nas pressuposições para a aplicação deles, conforme apresentado a seguir.

3.2.2.1. Teste de Dixon

O teste de Dixon é um teste bilateral cuja hipótese nula defende a inexistência de discrepância candidata a *outlier* no conjunto das discrepâncias (FERGUSON, 1961).

Este teste tem como pressuposição a distribuição normal das discrepâncias e que o número de discrepâncias seja menor ou igual a 40, ou seja, $n \leq 40$. Para realizar o Teste de Dixon é necessário calcular o valor da estatística de Dixon, D , cuja equação depende das iterações que são feitas, de acordo com as discrepâncias detectadas (DIXON, 1950).

Tabela 7: Equações de Dixon.

Iteração	Equação
1 ^a	$D_1 = \frac{d_2 - d_1}{d_n - d_1} \text{ e } D_2 = \frac{d_n - d_{n-1}}{d_n - d_1}$

Fonte: Dixon (1950).

Uma vez em ordem crescente os valores das discrepâncias, d_1 representa o primeiro valor da discrepância, d_2 o segundo valor, até n -ésimo valor d_n . Primeiramente deve ser feito o cálculo de D_1 , demonstrado na Tabela 7, para verificar se d_1 , ou seja, a menor discrepância, é uma candidata à *outlier*. De acordo com o resultado obtido para o d_1 , segue-se os seguintes passos descritos a seguir:

- Se d_1 é candidata à *outlier*, esta discrepância deve ser removida da série e as demais discrepâncias reordenadas. Depois, realiza-se o cálculo de D_2 para verificar se d_n é uma candidata à *outlier*. Salienta-se que o novo valor de d_1 será o segundo valor da série, já que o primeiro foi eliminado.
- Se d_1 não é candidata à *outlier* mantêm-se todas as discrepâncias da série e realiza-se o cálculo de D_2 para verificar se d_n é uma candidata à *outlier*, usando o valor da primeira discrepância como d_1 .

Se d_n é uma candidata à *outlier*, esta discrepância deverá ser removida da série de dados e as discrepâncias reordenadas. Assim, para o próximo cálculo de D_1 , a discrepância a ser usada em d_n será o valor que está na posição $n - 1$. Logo, a discrepância detectada como candidata à *outlier* sempre é eliminada da série para fazer o próximo cálculo. Esse processo é

repetido até que o menor e maior valor da série de dados não sejam detectados como candidatos a *outliers* (DIXON, 1950).

Para decidir se a discrepância é ou não candidata a *outlier*, compara-se o valor de D_1 ou D_2 com o valor de D tabelado, o qual é obtido com base no nível de significância α e no número de discrepâncias n , ou seja, $D_{\text{tabelado}} = D_{\alpha}(n)$. Se $D_i (i=1 \text{ ou } 2) \geq D_{\text{tabelado}}$, tem-se a presença de uma discrepância candidata à *outlier*.

Uma desvantagem do teste de Dixon é que aumentando n , aumenta-se também a probabilidade de ocorrer grandes desvios no conjunto de medidas (OLIVEIRA, 2008).

3.2.2.2. Teste de Grubbs

Existem vários critérios para testar valores discrepantes e alguns deles incluem a observação suspeita de ser um candidato a *outlier* no cálculo, cujo valor resultante é comparado com um valor crítico e assim realiza-se uma análise para decidir se a observação deve ser mantida ou retirada. A maioria dos métodos se baseiam em uma distribuição normal, como é o caso do teste de Grubbs (1969).

Segundo Grubbs (1969), a hipótese nula testada é que todas as discrepâncias da amostra vêm da mesma população, ou seja, não são candidatas à *outliers*. Seja uma amostra com n discrepâncias posicionais, em ordem crescente denotada por $d_1 \leq d_2 \leq \dots d_j \dots \leq d_n$, sendo d_j o valor duvidoso, ou seja, que pode ser o menor ou maior valor. Para analisar um único candidato a *outlier*, a estatística T_j é calculada utilizando a Equação 21.

$$T_j = \frac{d_j - \bar{d}}{s} \quad (21)$$

Em que $\bar{d}$ é a média das n discrepâncias posicionais e s é o desvio padrão amostral, com $n - 1$ graus de liberdade. A Equação 21 também pode ser usada para testar se o valor mais baixo, d_1 , do conjunto de discrepâncias é um candidato a *outlier*. Porém, não pode ser usada para testar o valor mais alto e mais baixo, pois se acredita que é improvável que isso aconteça simultaneamente (GRUBBS, 1969).

O valor tabelado do teste de Grubbs é obtido por $T_{\text{tabelado}} = T_{\alpha}(n)$ em que α é o nível de significância e n é o número de discrepâncias. Se o valor de T_n for maior que o valor contido na Tabela de Grubbs, rejeita-se a hipótese nula e tem-se uma discrepância candidata à *outlier*.

3.2.2.3. Teste de Chauvenet

O teste de Chauvenet baseia-se em uma amostra com distribuição normal e na hipótese nula de que as discrepâncias não são *outliers* (CALLEGARO, 2014).

Para verificar se uma discrepância é candidata à *outlier*, deve-se calcular a estatística do teste R_i conforme Equação 22 (GLEN, 2016):

$$R_i = \frac{|d_i - \bar{d}|}{s} \quad (22)$$

Em que d_i é a discrepância candidata a *outlier*, $\bar{d}$ é a média amostral e s o desvio padrão das discrepâncias.

Para que seja realizado o teste de Chauvenet, as discrepâncias devem estar em ordem crescente. O teste deve ser aplicado calculando a estatística R_i (Equação 22) para a menor e a maior discrepância, e posteriormente comparar esses valores de R_i com o valor tabelado para o critério de Chauvenet o qual é obtido por $R_{tab} = R(n)$, em que n é o número de discrepâncias. Neste teste não é necessário definir um nível de significância, visto que a tabela é única e necessita apenas do número de discrepâncias.

Se $R_i > R_{tab}$, a discrepância testada é considerada um candidato a *outlier*. Se uma das discrepâncias for acusada, deve-se eliminá-la da amostra e refazer os cálculos para média e desvio padrão, assim como obter um novo valor tabelado, e seguir com o cálculo de R_i para os novos valores de menor e maior discrepância. O procedimento deve ser repetido até que não sejam mais detectados candidatos a *outliers* (FERREIRA, 2017).

Se a Equação 22 for comparada a Equação 21 proveniente do teste de Grubbs, nota-se uma semelhança entre elas. Entretanto, deve-se salientar que além das diferenças existentes como o módulo presente na Equação 22 e nos valores tabelados, já que cada teste tem sua respectiva tabela, ou seja, as metodologias de aplicação são distintas. Enquanto o método de Grubbs limita a sua aplicação apenas para a menor e maior discrepância, o teste de Chauvenet é um método iterativo, que permite a sua aplicação nas discrepâncias até que não sejam mais detectadas candidatas a *outliers*.

4. Materiais e métodos

4.1. Materiais

As bases de dados utilizadas neste estudo foram obtidas de trabalhos realizados por outros autores, cujas informações podem ser vistas na Tabela 8.

Tabela 8: Bases de dados.

Base	Fonte	Nº Pontos	Equidistância Vertical (m)	Resolução Espacial (m)	Escala do produto	Método de obtenção dos pontos de referência
AL26	(SANTOS, 2010)	26	-	2,5	1:25.000	Receptor GNSS
IT30	(MAROTTA e VIEIRA, 2005)	30	-	15	1:50.000	Cartas topográficas 1:50.000
VC30	(MAROTTA e VIEIRA, 2005)	30	-	15	1:50.000	Receptor GNSS
VC24	(SANTOS e VIEIRA, 2009)	24	-	2,5	1:25.000	Receptor GNSS
VC33	(SANTOS e VIEIRA, 2009)	33	5	1	1:10.000	Receptor GNSS

A primeira base de dados, aqui renomeada de AL26, foi gentilmente cedida pelo Professor Afonso de Paula dos Santos do Departamento de Engenharia Civil da Universidade Federal de Viçosa (UFV), e foi utilizada na elaboração de sua dissertação de mestrado (SANTOS, 2010). Esta base de dados é composta por uma ortoimagem ALOS do sensor PRISM pancromática, de julho de 2008, com resolução espacial de 2,5 metros que cobre uma porção do município de Alvinópolis - MG.

A segunda e a terceira base de dados, denominadas de IT30 e VC30, respectivamente, utilizadas por Marotta e Vieira (2005), compreendem duas imagens orbitais do sensor ASTER com 15 metros de resolução espacial, sendo uma obtida no dia 02/06/2002, que contempla o município de Itabirito – MG, e outra do dia 29/03/2001, que contém o município de Viçosa - MG. A quarta base de dados, VC24, utilizada por Santos e Vieira (2009), é composta por uma imagem orbital do satélite CBERS-2B do sensor HRC (Câmera Pancromática de Alta Resolução) com resolução espacial 2,5 metros, que também cobre a cidade de Viçosa – MG.

A quinta base de dados, também retirada de Santos e Vieira (2009), aqui denominada de VC33, consiste em uma ortoimagem IKONOS de 1 metro de resolução espacial, que contém além das informações planimétricas, valores de altimetria representados por meio de curvas de nível com equidistância de 5 metros.

Para cada uma das bases de dados, foi utilizado um método para obtenção dos pontos de referência. Na primeira base de dados AL26, foram coletados 26 pontos de referência por meio de um posicionamento estático com receptor GNSS (*Global Navigation Satellite System*) Leica, modelo GPS 900 CS. Para o processamento dos dados coletados em campo foi tido como base as estações VICO e MGBH da Rede Brasileira de Monitoramento Contínuo dos Sistemas GNSS (RBMC), localizadas em Viçosa-MG e Belo Horizonte-MG, respectivamente (SANTOS, 2010).

Para a base de dados IT30 foram utilizadas como referência cartas topográficas na escala 1:50.000, enquanto que para a base VC30, os dados de referência foram 30 pontos obtidos utilizando um receptor GNSS (MAROTTA; VIEIRA, 2005). A quarta e a quinta base de dados, VC24 e VC33, contaram com 24 e 33 pontos de referência, respectivamente. Sendo que ambas tiveram seus pontos coletados utilizando receptores GNSS Promark 2 da Ashtech, no modo de posicionamento estático (SANTOS; VIEIRA, 2009).

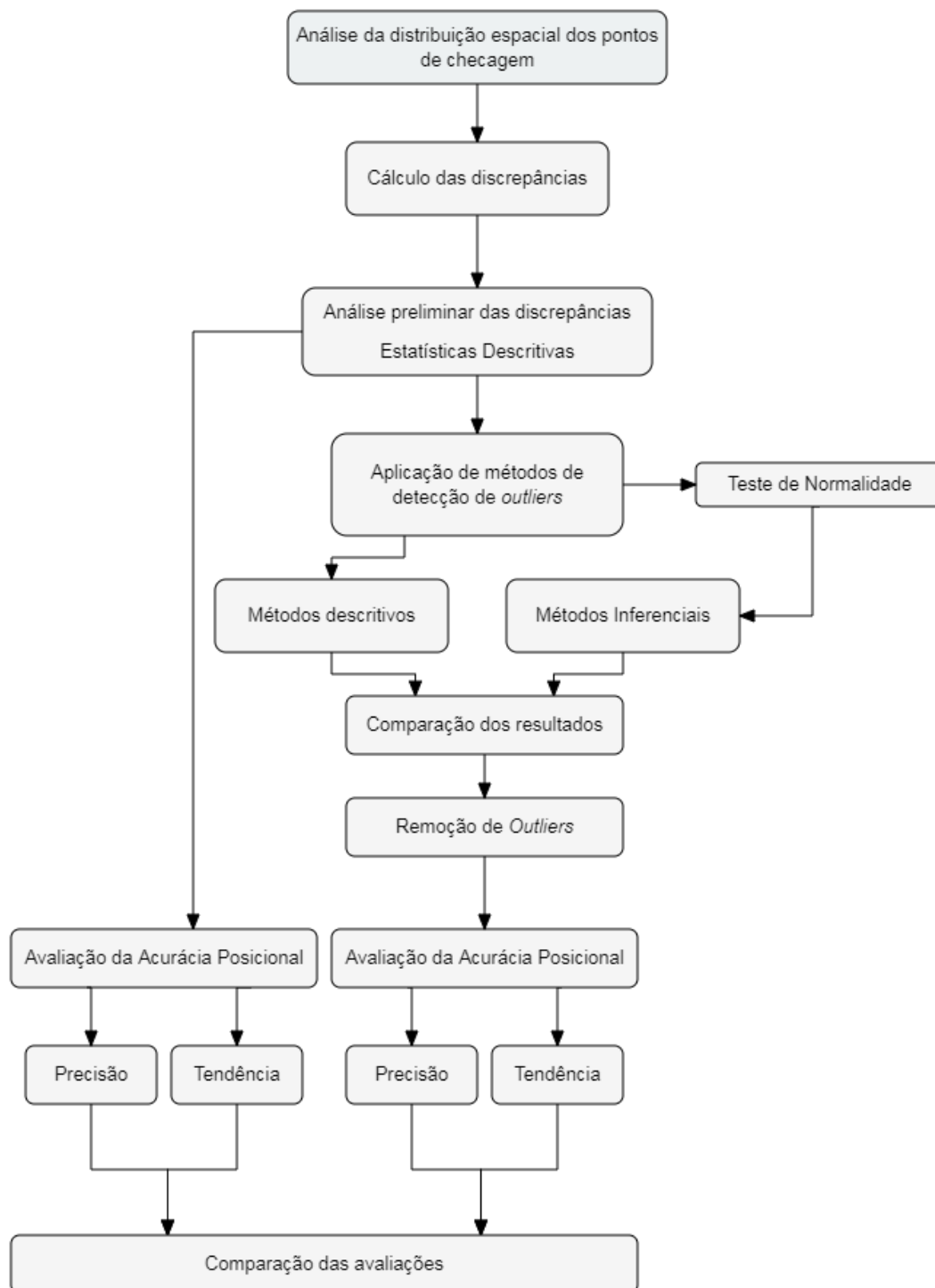
Alguns autores, como Nero (2005) e Santos (2010), recomendam que os pontos de referência sejam três vezes mais acurados que o produto cartográfico em avaliação. Dessa forma, as bases de dados AL26, VC30, VC24 e VC33, que tiveram seus pontos de referências coletados em campo por levantamento geodésico (posicionamento GNSS), atendem a esse requisito.

Entretanto, os pontos de referência da base IT30 foram adquiridos através da digitalização de outra carta. Este fato a princípio não seria um problema, porém a escala da carta de referência utilizada foi a mesma utilizada na classificação do produto cartográfico, sendo que a escala da carta de referência deveria ser três vezes maior. Logo, para a base IT30 a recomendação para os pontos de referência não foi seguida.

4.2. Métodos

As etapas utilizadas nesta pesquisa são apresentadas, resumidamente, na Figura 4. A descrição de cada uma destas etapas é apresentada a seguir.

Figura 4 - Fluxograma das etapas realizadas para a avaliação da acurácia posicional, com vistas à comparações dos métodos de detecção de *outliers*.



Fonte: A autora.

Salienta-se que antes de iniciar o controle de qualidade cartográfica, o ideal seria definir o número de amostras necessárias para cada uma das áreas de estudo e escolher os

locais onde seriam coletados os pontos de teste e de referência, conforme a metodologia apresentada na Revisão de Literatura (item 3.1.2.1). Contudo, esse passo não foi realizado, visto que as bases de dados utilizadas já contêm os pontos de checagem supracitados e não foram realizadas novas coletas de dados.

Assim sendo, a etapa inicial consistiu na espacialização das coordenadas X e Y dos pontos de teste e de referência de cada uma das cinco bases de dados, a fim de gerar uma análise visual da distribuição espacial dos pontos.

Posteriormente, para avaliar o padrão de distribuição espacial dos pontos de checagem foi utilizada a função K de Ripley, descrita em Wong e Lee (2005). Esta função utiliza a área mapeada em seu cálculo, mas como as bases disponibilizadas não possuíam os dados referentes à área real de estudo, foi utilizado como recurso um retângulo mínimo que envolve os pontos, ou seja, foi criado uma área mínima ao redor dos pontos.

Após a verificação do padrão da distribuição espacial dos pontos, com as coordenadas dos pontos de teste e das dos pontos de referência, foram calculadas as discrepâncias planimétricas e altimétricas, fazendo uso das Equações 1, 2, 3 e 4, para cada uma das cinco bases de dados. A partir dos valores dessas discrepâncias, foram calculadas algumas estatísticas descritivas, como média, mediana, desvio-padrão, coeficiente de variação, erro-padrão, primeiro e terceiro quartil, e maior e menor valor.

De posse de todas as discrepâncias planimétricas e altimétricas (dx_i , dy_i , d_{2D_i} e dz_i para $i = 1, 2, \dots, n$), aplicaram-se os métodos descritivos de detecção de *outliers* Método 3σ (subdividido em 3σ , 3σ Alternativo e 3σ Cartográfico), Diagrama Box-Plot, Método *Standard Deviation* (2D e 3D) e o Método de Hampel, descritos na Revisão de Literatura nos itens 3.2.1.1., 3.2.1.2, 3.2.1.3 e 3.2.1.4, respectivamente.

Em seguida, o conjunto das discrepâncias de cada uma das bases de dados foi submetido ao teste de Shapiro-Wilk para avaliar a suposição de distribuição de probabilidades normal. Posteriormente, cada conjunto de discrepâncias foi submetido aos métodos inferenciais para detecção de *outliers*, considerando o nível de significância de 5%, sendo eles o Teste de Dixon, o Teste de Grubbs e o Teste de Chauvenet, descritos na Revisão de Literatura nos itens 3.2.2.1, 3.2.2.2 e 3.2.2.3, respectivamente. Com isso, foi observado o desempenho dos métodos em detectar *outliers*, verificando qual método foi mais ou menos conservador à presença deles.

Após a aplicação dos métodos descritivos e inferenciais, calculou-se para cada um deles e em cada uma das bases de dados, a sua taxa de discrepâncias detectadas como candidatas à *outlier*. Essa taxa foi obtida pela Equação 23.

$$Tx_{kj} = \frac{d_{kj}}{n_j} \quad (23)$$

Em que

- Tx_{kj} : taxa de detecção de discrepâncias candidatas à *outlier* pelo método k (tal que representa cada um dos métodos descritivo e inferencial) na base j (tal que $j = \text{AL26, IT30, VC30, VC24 e VC33}$);
- d_{kj} : número de pontos que tiveram pelo menos uma de suas componentes de discrepância (dx_i, dy_i, dz_i) detectadas como candidatas a *outlier*;
- n_j : número total de pontos da base de dados.

No cálculo dessa taxa, mesmo que um ponto tivesse mais que uma componente de discrepância detectada como candidata a *outlier*, foi considerado apenas uma unidade no cálculo da taxa.

Em geral, para verificar se uma discrepância candidata a *outlier* é realmente um *outlier* deve ser realizada uma avaliação para averiguar o motivo do alto valor de discrepância detectado. Este valor pode estar relacionado com a qualidade do próprio produto cartográfico, ou com a ocorrência de algum erro grosseiro ou sistemático durante a obtenção do ponto de teste, no produto cartográfico avaliado, ou durante a coleta do ponto de referência. Para o primeiro caso, deve-se manter esta discrepância e seu respectivo ponto associado, pois este valor é importante para avaliar a qualidade do produto cartográfico. Para o segundo caso, ter-se-ia realmente um *outlier*, visto que o valor alto de discrepância provém de um erro ocorrido durante a coleta dos pontos de referência ou de teste utilizados na avaliação.

Como as bases de dados neste estudo foram obtidas de trabalhos anteriores, não foi possível a realização da análise quanto à forma de suas obtenções. Dessa forma, houve a necessidade de se estabelecer um critério para definir se um ponto de checagem é ou não é um *outlier*. Esse critério teve como base os resultados dos métodos descritivos e inferenciais aplicados a cada uma de suas componentes:

- Se nenhuma das componentes do ponto foi detectada como candidata a *outlier* por nenhum dos métodos, então o ponto não é *outlier* e, nesse caso, o ponto foi

mantido na base de dados para avaliação da acurácia do seu produto cartográfico;

- Se pelo menos uma das componentes do ponto foi detectada como candidata a *outlier* por um método descritivo e um método inferencial, então o ponto é um *outlier* e, nesse caso, o ponto foi removido da base de dados para avaliação da acurácia do seu produto cartográfico.

Este critério foi estabelecido considerando que os métodos descritivos e inferenciais são complementares. Uma vez que os inferenciais têm pressuposições para sua aplicação, caso alguma pressuposição não seja atendida e isso implique na falha do método inferencial, o método descritivo seria um indicativo para apontar a presença de um candidato a *outlier*.

As normas que regulamentam o controle de qualidade cartográfica utilizam apenas das discrepâncias 2D e Z para realizar a avaliação planimétrica e altimétrica do produto cartográfico, respectivamente. Dessa forma, na prática a detecção de *outliers* é feita apenas nas componentes 2D e Z. Contudo, a detecção de *outliers* deveria ser aplicada em todas as componentes (X, Y, Z e 2D), pois quando se aplica os testes para verificação da tendência, como o teste t de Student, ele é aplicado em todas as componentes. Assim, neste trabalho os métodos de detecção foram aplicados em todas as discrepâncias, sendo elas X, Y, Z e 2D.

Para auxiliar na visualização dos pontos que foram apontados como candidatos a *outliers* pelos métodos descritivos e inferenciais, foi construído um Diagrama de Venn contemplando todas as detecções.

Após a remoção ou não dos *outliers*, cada produto foi avaliado quanto a sua acurácia posicional planimétrica e/ou altimétrica e para isso, analisou-se a precisão e a tendência. A análise de precisão e de tendência foi realizada seguindo a metodologia proposta por Santos et al. (2016), descritas na Revisão de Literatura nos itens 3.1.2.3, 3.1.2.4 e 3.1.2.5.

A avaliação da acurácia posicional foi feita utilizando-se o *software* livre GeoPEC, disponível na página do curso de Engenharia de Agrimensura e Cartográfica da UFV <<http://www.eam.ufv.br/>>, cujos detalhes podem ser obtidos em Santos (2010).

Adicionalmente foi realizada a avaliação da acurácia posicional de cada produto cartográfico sem a remoção de *outliers*, realizando-se apenas as análises de precisão e de tendência (Figura 4).

Assim, cada um dos cinco produtos cartográficos foi avaliado quanto a sua acurácia posicional com e sem a remoção de pontos *outliers*, com o objetivo de avaliar se a remoção de

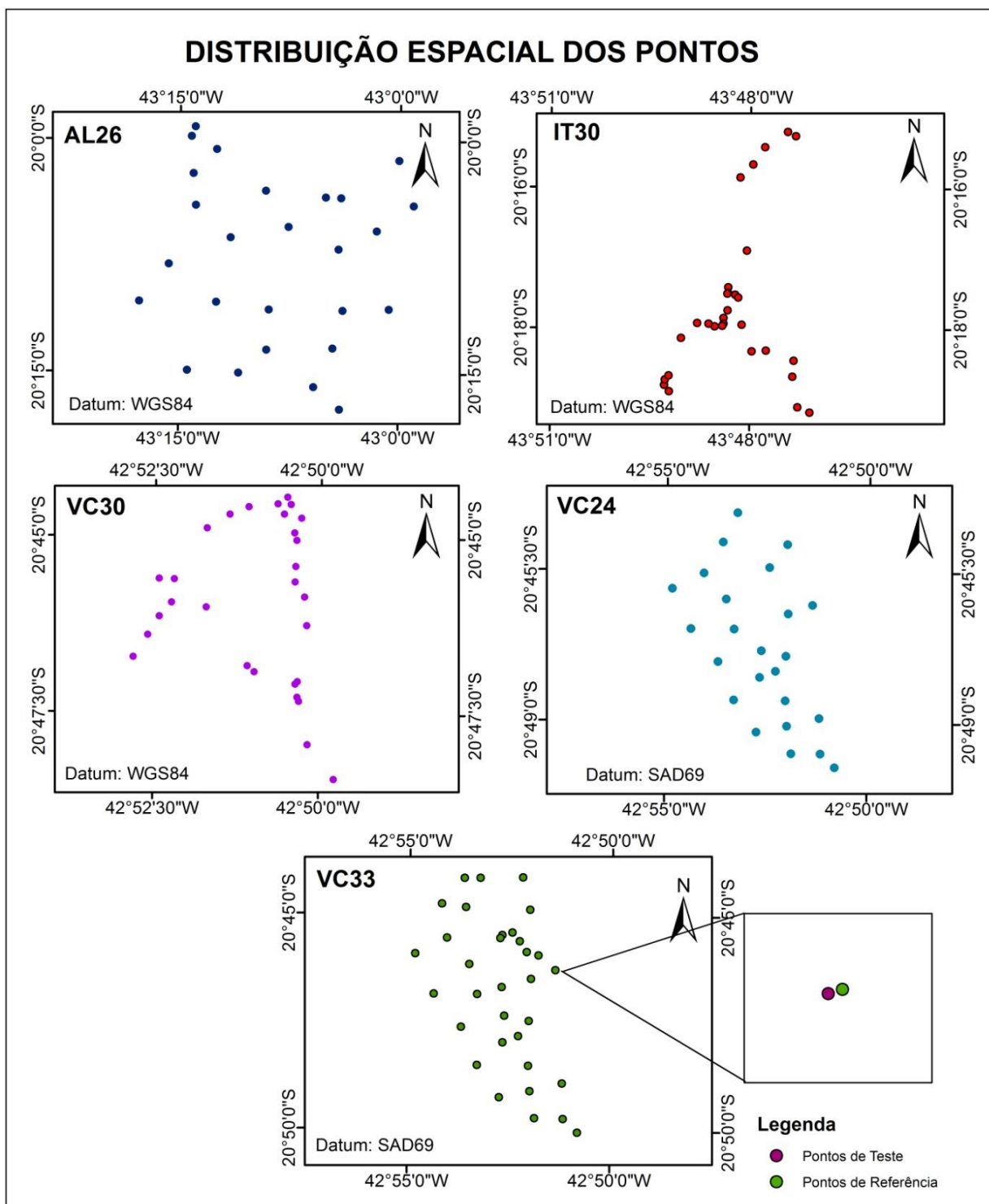
outliers influencia na qualidade do produto cartográfico. Além disso, foi definida a forma de combinação entre os métodos de detecção de *outliers* para ser aplicado no controle de qualidade cartográfica, no qual deve ser utilizado pelo menos um método descritivo e um inferencial. Se um ponto for detectado simultaneamente por pelo menos um método descritivo e um método inferencial, este ponto é um *outlier* e deverá ser removido da base de dados.

5. Resultados

5.1. Análise da distribuição espacial dos pontos de checagem

A primeira etapa deste estudo consistiu na espacialização das coordenadas dos pontos de checagem (teste e referência), para cada uma das bases de dados. Devido à pequena escala do desenho utilizada na Figura 5, os pontos de teste e referência ficaram sobrepostos. Contudo, se fosse utilizado uma escala maior para a representação, seria possível visualizar as discrepâncias entre as coordenadas de pontos de teste e de referência, conforme demonstrado na ampliação de um dos pontos de checagem da base VC33.

Figura 5 - Distribuição espacial dos pontos de controle das cinco bases de dados (AL26, IT30, VC30, VC24 e VC33) em que para VC33 foi feita uma ampliação para visualização dos pontos de teste e referência.

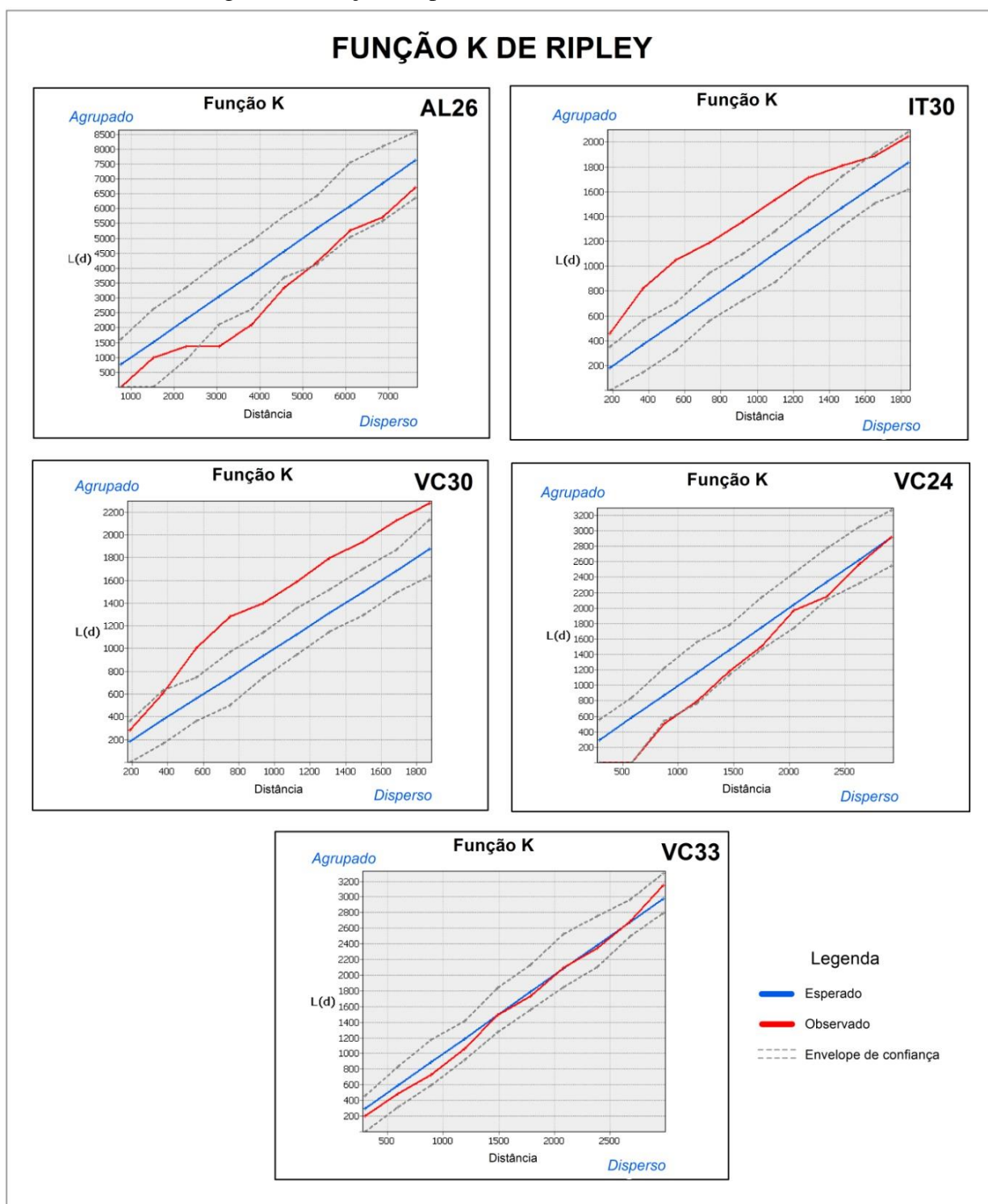


Fonte: A autora.

A partir de uma análise visual da Figura 5 é possível concluir que as bases de dados AL26, VC24 e VC33 possuem uma distribuição espacial dos pontos melhor do que as bases

IT30 e VC30. Porém, para que esta análise não seja apenas subjetiva, foi aplicada a Função K de Ripley sobre os pontos de referência de cada base de dados. Essa análise tem como objetivo avaliar se o padrão da distribuição espacial dos pontos é aleatório, agrupado ou disperso. Os gráficos obtidos com a aplicação da Função K de Ripley encontram-se na Figura 6.

Figura 6 - Função K aplicada nas cinco bases de dados.



Fonte: A autora.

A interpretação da Figura 6 tem como base a linha simbolizada na cor vermelho, cujo comportamento representa o padrão da distribuição espacial observado nos pontos de

referência, pois a linha na cor azul representa o padrão esperado de pontos para uma distribuição aleatória. Quando, para uma dada distância, a linha vermelha está acima do limite superior do envelope (linha preta tracejada superior na cor preta) conclui-se que, para aquela distância, o padrão da distribuição espacial dos pontos é agrupado. Por outro lado, se para uma dada distância, a linha vermelha está abaixo do limite inferior do envelope (linha preta tracejada inferior na cor preta) conclui-se que, para aquela distância, o padrão da distribuição espacial dos pontos é disperso. Se a linha na cor vermelha se encontra dentro do envelope o padrão de distribuição dos pontos é aleatório.

Para a maioria das distâncias, as bases AL26, VC24 e VC33 apresentaram um padrão de distribuição espacial dos pontos de referência aleatório. Por outro lado, as bases IT30 e VC30, para a maioria das distâncias, apresentaram padrão de distribuição espacial dos pontos de referência agrupado. De acordo com Santos (2010), para realizar uma avaliação da qualidade de um produto cartográfico é necessário que os pontos de checagem apresentem padrão de distribuição espacial disperso ou aleatório, pois esses padrões indicam que os pontos coletados estão bem distribuídos espacialmente, ou seja, tem informações que conseguem descrever toda a área de estudo. O padrão agrupado indica que os pontos estão concentrados em apenas uma parte da área de estudo. Dessa forma, não há uma representatividade de toda área mapeada e, assim, tais pontos não seriam ideais para serem utilizados para mensurar a qualidade de um produto cartográfico.

Partindo desse princípio, as bases AL26, VC24 e VC33 atendem ao padrão exigido e as bases IT30 e VC30 não atendem. Contudo, para efeito de estudo, as bases IT30 e VC30 foram mantidas neste trabalho para avaliar o efeito do agrupamento de pontos na detecção e remoção de *outliers*.

5.2. Cálculo das discrepâncias e análise preliminar a partir das estatísticas descritivas

Após a verificação do padrão de distribuição espacial dos pontos de referência, foram calculadas as discrepâncias posicionais e suas estatísticas descritivas para cada uma das bases (Tabela 9).

Tabela 9: Estatísticas descritivas para as bases de dados (AL26, IT30, VC30, VC24 e VC33), em que d_x representa o valor de discrepância na componente X, d_y a discrepância em Y, d_{2D} a discrepância planimétrica resultante e d_z a discrepância em Z.

Discrepância	nº de pontos	Média (m)	Mediana (m)	Desvio Padrão (m)	Coefficiente de variação (%)	Erro padrão (m)	Q1	Q3	Mínimo (m)	Máximo (m)
Estatísticas Descritivas - AL26										
d_x	26	-0,074	0,258	2,831	-3825,67%	0,555	-2,187	1,021	0,008	6,295
d_y		0,714	0,924	2,592	363,02%	0,508	-1,113	1,982	0,023	5,788
d_{2D}		3,356	3,221	1,884	56,14%	0,370	1,830	4,474	0,176	7,230
Estatísticas Descritivas - IT30										
d_x	30	-3,309	-3,676	12,865	-388,79%	2,523	-8,565	4,068	0,221	34,342
d_y		-8,489	-6,861	12,842	-151,28%	2,518	-16,752	0,503	0,277	40,512
d_{2D}		17,698	14,662	9,606	54,28%	1,884	10,169	22,266	6,266	40,548
Estatísticas Descritivas - VC30										
d_x	30	-0,243	1,032	14,107	-5805,35%	2,767	-12,220	11,724	0,752	30,395
d_y		1,256	0,611	11,061	880,65%	2,169	-5,787	5,894	0,019	44,618
d_{2D}		15,440	14,375	8,743	56,62%	1,715	10,463	16,725	4,173	46,318
Estatísticas Descritivas - VC24										
d_x	24	0,214	2,095	7,858	3671,96%	1,541	-3,097	4,536	0,082	19,523
d_y		-2,044	-1,077	6,481	-317,07%	1,271	-3,470	2,614	0,239	25,824
d_{2D}		7,923	5,466	6,533	82,46%	1,281	3,563	9,719	1,171	29,486
Estatísticas Descritivas - VC33										
d_x	33	-4,843	-4,951	1,345	-27,77%	0,264	-5,362	4,948	0,083	8,363
d_y		-0,600	-0,753	0,996	-166,00%	0,195	-1,193	-0,375	0,005	2,367
d_z		0,211	0,227	1,400	663,50%	0,275	-0,441	1,046	0,056	4,699
d_{2D}		5,014	5,028	1,199	23,91%	0,235	4,422	5,582	0,938	8,363

Pode ser observado na Tabela 9 que, para a base AL26, os máximos valores de discrepâncias tendem a ocorrer para a componente X e $2D$, com valores superiores a sete metros. Ainda nesta base, a componente $2D$ é a que possui maior média de discrepâncias. Entretanto, ao analisar seu coeficiente de variação notou-se que ele é o menor valor entre todas as componentes, mostrando que a componente $2D$ é a que possui maior homogeneidade em relação aos valores de discrepância das componentes X e Y .

Ao analisar a base IT30, notou-se que as discrepâncias tenderam a apresentar valores superiores aos observados para as bases de dados AL26, VC24 e VC33, atingindo os máximos valores em módulos de 34,342 metros, 40,512 metros e 40,548 metros para as discrepâncias X , Y e $2D$, respectivamente. Um dos fatores que podem explicar os altos valores de IT30 é a forma de coleta dos pontos de referência. Nesta base de dados os pontos de referência foram extraídos da carta SF-23-X-AIII-3 na escala 1:50.000, através de uma mesa digitalizadora (MAROTTA; VIEIRA, 2005). Segundo Francisco (1993), a digitalização é um processo moroso, sujeito a erros humanos. Adicionalmente, tem-se outro agravante que é o nível de desatualização das cartas. Outra possibilidade para explicar os altos valores de discrepâncias encontradas nessa base é a própria qualidade do produto.

Analisando-se os resultados para a base VC30, disponíveis na Tabela 9, observou-se que a discrepância $2D$ tende a possuir maior média de discrepâncias, mas é a mais homogênea em comparação com d_x e d_y , pois possui menor coeficiente de variação. O erro-padrão para a discrepância resultante $2D$ também é o que possui menor valor. O comportamento observado para a VC30 é semelhante ao que ocorreu com VC24, pois a discrepância em X é a que tende a possuir, em módulo, a menor média de discrepâncias e o maior desvio-padrão, o que consequentemente faz com que essa base de dados tivesse um alto coeficiente de variação. Com base nessa informação, concluiu-se que a componente X é a mais heterogênea e seus valores de discrepância variaram de 0,082 a 29,486 metros na base de dados VC24.

Conforme visto, as bases anteriores não têm a componente Z de discrepância, visto que foram coletados apenas pontos planimétricos X e Y . Já para a base VC33 foram coletados os valores de altitude geométrica. Sendo assim, foi possível fazer também o cálculo da discrepância altimétrica Z (d_z). Dentre todas as bases de dados, VC33 foi a que apresentou um dos menores valores de discrepâncias, sendo ele igual a 0,005 metros.

Na Tabela 9, notou-se ainda que algumas bases de dados são mais homogêneas, ou seja, possuem discrepâncias que giram em torno do mesmo valor e outras mais heterogêneas,

com grandes variações nas discrepâncias. Esse fato é confirmado através do coeficiente de variação, que mostrou que para todas as bases, a discrepância $2D$ foi a mais homogênea.

Em relação a esta estatística descritiva, notou-se que o coeficiente de variação possui altos valores e alguns deles com sinal negativo. Os valores negativos encontrados no coeficiente de variação são justificados pelo sinal da média. Muitas bases de dados têm mais discrepâncias negativas do que positivas, acarretando média negativa e, por consequência, coeficiente de variação negativo. Os altos valores de coeficiente de variação, que são superiores a 100, ocorreram devido ao baixo valor de média, visto que as discrepâncias na maioria dos casos são inferiores a unidade, e o alto valor de desvio-padrão.

Ainda em relação à Tabela 9, verificou-se que a média de discrepâncias para a componente $2D$ foi a maior, se comparado às outras componentes, para todas as bases de dados. Segundo Santos (2010), quando se avalia o comportamento das componentes da discrepância planimétrica X e Y isoladamente, é considerando apenas a translação do dado espacial, porém, ao se avaliar a resultante ($2D$), é analisada a escala do produto cartográfico.

5.3. Aplicação de métodos de detecção de *outliers*

Após realizar a análise preliminar descritiva e do padrão da distribuição espacial dos pontos, foram aplicados quatro métodos descritivos de detecção de *outliers* em cada uma das cinco bases de dados. Foram utilizados o Método 3σ (sendo aplicados o Método 3σ , 3σ alternativo e 3σ Cartográfico com EP da classe A para a respectiva escala do produto em avaliação), BoxPlot, Método *Standard Deviation* 2D e 3D e Método Hampel.

Em termos didáticos, pode-se pensar que cada um desses métodos realiza um teste de hipóteses empírico para cada uma das discrepâncias de uma base de dados. A hipótese de nulidade é que a discrepância não é uma candidata a *outlier* e a hipótese alternativa é que a discrepância é uma candidata a *outlier*. Diz-se que é um teste de hipótese empírico, pois a conclusão é apenas subjetiva, não se baseando em uma distribuição de probabilidades.

Assim, após aplicar os procedimentos foi verificado se o método foi mais conservador, ou seja, conservou um número maior de hipóteses de nulidade como verdadeira, ou se o método foi menos conservador, ou seja, conservou um número menor de hipóteses de nulidade como verdadeira.

Na Tabela 10 são apresentados os resultados dos métodos aplicados a cada discrepância posicional da base de dados AL26. Nessa tabela são apresentados apenas os

pontos que foram identificados como candidatos a *outlier*, ou seja, aqueles pontos que apresentaram pelo menos uma de suas componentes de discrepâncias como candidata a *outlier*, em pelo menos um método. Além disso, é apresentada a taxa de detecção de cada método, que é a porcentagem de pontos detectados como candidatos a *outliers* em relação ao total de pontos da base de dados.

Tabela 10: Detecção de *outliers* pelos métodos descritivos aplicados a base AL26, em que “x” significa que o referido ponto é candidato a *outlier*.

Método							
Ponto	3 σ			3 σ cartográfico Classe A 1:25.000			Hampel
	X	Y	2D	X	Y	2D	
17	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-
24	-	-	-	-	-	-	-
Taxa de detecção	0%			0%			0%

Considerando os resultados de todos os métodos, na base AL26 foram identificados quatro pontos (n° 17, 21, 22 e 24) como candidatos a *outliers*. O ponto 17 foi identificado como candidato a *outlier* por três métodos, já o ponto 24 foi identificado como candidato a *outlier* por dois métodos. Um aspecto a ser destacado é o fato do Método 3 σ , 3 σ Cartográfico, *Standard Deviation* 3D e o Hampel não terem identificado nenhum ponto, indicando que esses métodos são mais conservadores. Por outro lado, o Método 3 σ Alternativo foi menos conservador a presença de *outliers*, identificando os quatro pontos como candidatos a *outlier*. Se for comparada a taxa de detecção, este método apresentou uma detecção de 15% dos pontos da base AL26 como candidatos a *outliers*.

A base IT30 (Tabela 11) apresentou resultados similares aos observados para a base AL26, nos quais os métodos 3 σ , 3 σ Cartográfico e *Standard Deviation* 3D não acusaram nenhum ponto como um candidato a *outlier*. Os métodos 3 σ alternativo, BoxPlot e Método *Standard Deviation* 2D apresentaram a mesma taxa de detecção, de 10%, enquanto o método Hampel apresentou taxa de detecção de 7%.

Tabela 11: Detecção de *outliers* pelos métodos descritivos aplicados a base IT30, em que “x” significa que o referido ponto é candidato a *outlier*.

Método																					
Ponto	3σ			3σ alternativo			3σ cartográfico Classe A 1:50.000			BoxPlot			Standard Deviation 2D			Standard Deviation 3D			Hampel		
	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D
11	-	-	-	-	-	X	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-
13	-	-	-	-	X	X	-	-	-	-	-	X	-	X	X	-	-	-	-	-	X
14	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-
15	-	-	-	-	-	X	-	-	-	X	-	-	X	-	X	-	-	-	X	-	X
Taxa de deteccção	0%			10%			0%			10%			10%			0%			7%		

Para a base VC30 (Tabela 12), todos os métodos descritivos acusaram pelo menos um candidato a *outlier*, sendo o BoxPlot e o Método 3 σ alternativo os métodos menos conservadores, com a maior taxa de detecção tendo o valor de 10%.

Tabela 12: Detecção de *outliers* pelos métodos descritivos aplicados a base VC30, em que “x” significa que o referido ponto é candidato a *outlier*.

Método																					
Ponto	3σ			3σ alternativo			3σ cartográfico Classe A 1:50.000			BoxPlot			Standard Deviation 2D			Standard Deviation 3D			Hampel		
	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D
19	-	x	x	-	x	x	-	-	x	-	x	x	-	x	x	-	x	x	-	x	x
26	-	-	-	-	-	x	-	-	-	-	-	x	x	-	x	-	-	-	-	-	x
29	-	-	-	-	-	x	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-
Taxa de detecção	7%			10%			3%			10%			7%			3%			7%		

Cada um dos métodos descritivos aplicados a base VC24 (Tabela 13) detectaram pelo menos um candidato a *outlier*, sendo o Método Hampel o menos conservador.

Tabela 13: Detecção de *outliers* pelos métodos descritivos aplicados a base VC24, em que “x” significa que o referido ponto é candidato a *outlier*.

Método																						
Ponto	3σ			3σ alternativo			3σ cartográfico Classe A 1:25.000			BoxPlot			Standard Deviation 2D			Standard Deviation 3D			Hampel			
	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	X	Y	2D	
3	-	x	x	-	x	x	-	x	x	-	x	x	-	x	x	-	x	x	x	x	x	
6	-	-	-	-	-	x	-	-	-	x	-	x	x	-	-	-	-	-	-	x	-	x
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	x
Taxa de detecção	4%			8%			4%			8%			8%			4%			13%			

As bases de dados AL26, IT30, VC30 e VC 24 possuem apenas informações planimétricas para as quais foram obtidas as respectivas discrepâncias. Por outro lado, a base VC33 possui também a componente altimétrica. Sendo assim, os métodos de detecção de *outliers* foram aplicados também para a componente Z da discrepância posicional (Tabela 14).

Tabela 14: Detecção de *outliers* pelos métodos descritivos aplicados a base VC33, em que “x” significa que o referido ponto é candidato a *outlier*.

VC33							
Método							
Ponto	3 σ	3 σ alternativo	3 σ cartográfico Classe A 1:10.000	BoxPlot	Standard Deviation 2D	Standard Deviation 3D	Hampel
	X Y Z 2D	X Y Z 2D	X Y Z 2D	X Y Z 2D	X Y Z 2D	X Y Z 2D	X Y Z 2D
1	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
2	- - - -	- - - x	- - - -	- - - -	- - - -	- - - -	- - - -
3	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
4	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
5	- - - -	- - - x	- - - -	- - - -	- - - -	- - - -	- - - -
6	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
7	- - - -	- - - x	- - - -	- x - -	- x - -	- - - -	- x - -
8	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
9	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
10	- - - -	- - - -	- - - -	x - - x	x - - x	x - - x	x - - x
11	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
12	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
13	- - - -	x - - x	- - x -	- - - -	- - - -	- - - -	- - - -
14	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
15	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
16	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
17	- - - -	x - - x	- - x -	- - - -	- - - -	- - - -	- - - -
18	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
19	- - - -	- - - x	- - x -	- - x -	- - x -	- - - -	- - - -
20	- - - -	x - - x	- - - -	x - - x	x - - x	- - - -	x - - x
21	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
22	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
23	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
24	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
25	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
26	- - - -	x - - x	- - x -	- - - -	- - - -	- - - -	- - - -
27	- - - -	x - - x	- - x -	- - - -	- - - -	- - - -	- - - -
28	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
29	- - - -	x - - x	- - - -	- x - -	- x - -	- - - -	- x - -
30	- - - -	x - - x	- - x -	- x - -	- - - -	- - - -	- x - -
31	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
32	- - - -	x - - x	- - - -	- - - -	- - - -	- - - -	- - - -
33	- - - -	x - x x	- - x -	- - x -	- - x -	- - - -	- - x -
Taxa de detecção	0%	94%	21%	21%	18%	3%	18%

Para a base VC33 nota-se um comportamento diferente do observado para as bases AL26, IT30, VC30 e VC 24, visto que o método 3σ Alternativo detectou 94% dos pontos como candidatos a *outliers*, sendo que essas detecções foram observadas apenas para as componentes X e $2D$. Vale ressaltar que para o método 3σ Alternativo uma discrepância é candidata à *outlier* se o seu módulo for maior que três vezes o desvio padrão. Ao analisar um extrato de discrepâncias obtidas de 10 pontos de um total de 33 pontos da base VC33 (Tabela 15) observa-se que grande parte das discrepâncias é negativa. O sinal negativo foi observado porque as coordenadas dos pontos de referência (coletadas em campo) foram sistematicamente maiores que as coordenadas dos pontos de teste (coletadas na imagem).

Logo ao aplicar o módulo, a discrepância que era negativa ficou positiva e com valor superior ao limiar definido para este método. Ao analisar os sinais das discrepâncias das demais bases de dados, notou-se que em alguns pontos as coordenadas dos pontos de teste são superiores aos de referência e vice-versa, ou seja, há aleatoriedade nos sinais das discrepâncias. O que aconteceu com a base VC33 dá indícios de um erro sistemático durante a mensuração das coordenadas dos pontos de referência em campo, por isso houveram muitas detecções. Esse resultado indica que o Método 3σ alternativo, apesar da sensibilidade ao sinal das discrepâncias, poderia ser utilizado para detectar erros sistemáticos ocorridos em uma base de dados. Como desvantagem esse método acusa vários pontos como candidatos a *outliers* que podem não ser *outliers*, tendo em vista que não foram detectados pelos outros métodos descritivos.

Tabela 15: Extrato de discrepâncias obtidas de 10 pontos de um total de 33 pontos da base VC33.

Discrepâncias posicionais (m)				
Ponto	d_x	d_y	d_z	d_{2D}
1	-5,091	-1,598	0,386	5,336
2	-3,793	-1,558	0,608	4,101
3	-5,362	-1,553	-1,417	5,582
4	-5,329	-1,634	-0,375	5,574
5	-4,025	-0,962	0,926	4,138
6	-6,516	-0,723	0,227	6,556
7	-3,817	2,367	1,489	4,491
8	-6,423	-1,064	-0,772	6,511
9	-5,545	-0,198	-1,344	5,549
10	0,083	-0,934	0,269	0,938

Após a aplicação dos métodos descritivos, foi aplicado o teste Shapiro-Wilk a 5% de probabilidade para verificar a normalidade de cada componente da discrepância, cujos resultados, para cada uma das bases de dados, são apresentados na Tabela 16.

Tabela 16: Resultado do teste de Shapiro-Wilk para as discrepâncias posicionais obtidas para os pontos de checagem de cada uma das cinco bases de dados.

Teste de Shapiro-Wilk									
Base de dados	Nº	Valor-p d_x	Resultado	Valor-p d_y	Resultado	Valor-p d_z	Resultado	Valor-p d_{2D}	Resultado
AL26	26	0,3935	Normal	0,9462	Normal	-	-	0,5825	Normal
IT30	30	0,5179	Normal	0,6581	Normal	-	-	0,0298	Não Normal
VC30	30	0,2496	Normal	0,0004	Não Normal	-	-	0,0003	Não Normal
VC24	24	0,1159	Normal	0,0002	Não Normal	-	-	0,0003	Não Normal
VC33	33	0,0021	Não Normal	0,0031	Não Normal	0,4082	Normal	0,0080	Não Normal

Notou-se que para todas as bases de dados, a componente $2D$ não apresentou distribuição normal, exceto para AL26, que apresentou normalidade para todas as componentes de discrepância. Para a componente X , a falta de normalidade também foi observada na base de dados VC33, e para a componente Y nas bases VC30, VC24 e VC33.

Conforme observado na Tabela 16, algumas bases de dados não apresentaram normalidade e os métodos inferenciais descritos nos tópicos 3.2.2.1, 3.2.2.2 e 3.2.2.3 têm como pressuposição a normalidade dos dados. Apesar disso, esses métodos foram aplicados para todas as bases de dados, em todas as suas componentes de discrepância posicional. Tal procedimento foi realizado com o objetivo de analisar o desempenho do método frente à violação da pressuposição de normalidade da distribuição das discrepâncias, pois alguns métodos podem ser robustos aos desvios da normalidade, ou seja, continuam eficientes mesmo que a premissa não seja satisfeita. De acordo com Montgomery (2013), violações pequenas a moderadas da normalidade não afetam consideravelmente o desempenho dos testes.

Os resultados dos métodos inferenciais aplicados às bases AL26, IT30, VC30 e VC24, que possuem apenas informações planimétricas, são apresentados na Tabela 17. Nesta tabela são apresentados os pontos que tiveram pelo menos uma componente de discrepância apontada como candidata à *outlier*, bem como a taxa de detecção de candidatos a *outlier* por

cada um dos métodos. Os pontos que foram detectados pelos métodos inferenciais e por pelo menos um método descritivo foram identificados pelo símbolo “#”.

Tabela 17: Resultados dos métodos inferenciais para a detecção de candidatos a *outliers* aplicados às discrepâncias das bases AL26, IT30, VC30, VC24 com dados planimétricos, em que os números identificam os pontos apontados como candidatos a *outlier*.

AL26				
Método	Candidatos a <i>outliers</i>			Taxa de detecção
	X	Y	2D	
Teste de Dixon	-	-	-	0%
Teste de Grubbs	-	-	-	0%
Teste de Chauvenet	-	-	-	0%
IT30				
Método	Candidatos a <i>outliers</i>			Taxa de detecção
	X	Y	2D	
Teste de Dixon	-	-	-	0%
Teste de Grubbs	-	-	-	0%
Teste de Chauvenet	15 [#]	11 [#] 13 [#]	-	10%
VC30				
Método	Candidatos a <i>outliers</i>			Taxa de detecção
	X	Y	2D	
Teste de Dixon	-	19 [#]	-	3%
Teste de Grubbs	-	19 [#]	19 [#]	3%
Teste de Chauvenet	-	19 [#]	19 [#]	3%
VC24				
Método	Candidatos a <i>outliers</i>			Taxa de detecção
	X	Y	2D	
Teste de Dixon	-	3 [#]	-	4%
Teste de Grubbs	-	3 [#]	3 [#]	4%
Teste de Chauvenet	6 [#]	3 [#]	3 [#]	8%

#: Pontos que também foram detectados como candidatos a *outliers* por pelo menos um dos métodos descritivos.

Para a base de dados AL26, nenhum dos testes detectou pontos candidatos a *outliers*. Para a base IT30, apenas o teste de Chauvenet detectou que as discrepâncias das componentes X e Y dos pontos, 15, 11 e 13, fossem candidatas a *outliers*.

Observando-se os resultados da base VC30, nota-se que para o ponto 19 as discrepâncias das componentes Y e 2D foram detectadas como candidatas a *outliers* pelos testes de Dixon e de Grubbs, respectivamente.

Por fim, na base VC24, todos os três testes detectaram a discrepância ocorrida na componente *Y* do ponto 3 como candidata a *outlier*. Para esse mesmo ponto a discrepância da componente *2D* foi detectada pelos testes de Grubbs e Chauvenet. Esse teste também detectou a componente *X* do ponto 6 como candidata a *outlier*. Vale ressaltar que o teste de Grubbs detectou a maior e menor discrepância como candidatas a *outliers* para a componente *2D*. De acordo com Grubbs (1969), é pouco provável que o menor e o maior valor de um conjunto de dados sejam detectados como candidatos a *outliers* conjuntamente.

Finalmente, os métodos inferenciais foram aplicados à base VC33, a única que contém informações altimétricas. O resultado da aplicação de cada um dos testes pode ser visto na Tabela 18.

Tabela 18: Resultados dos métodos inferenciais para a detecção de candidatos a *outliers* aplicados às discrepâncias da base de dados VC33, em que os números identificam os pontos apontados como candidatos a *outlier*.

VC33					
Teste	Candidatos a <i>outliers</i>				Taxa de detecção
	X	Y	Z	2D	
Teste de Dixon	-	-	-	-	0%
Teste de Grubbs	10 [#]	7 [#]	33 [#]	10 [#] e 20 [#]	12%
Teste de Chauvenet	10 [#] e 20 [#]	7 [#] , 29 [#] e 30 [#]	33 [#]	10 [#] e 20 [#]	18%

[#]: Pontos que também foram detectados como candidatos a *outliers* por pelo menos um dos métodos descritivos.

O Teste de Dixon não apontou nenhuma discrepância candidata a *outlier*. O Teste de Grubbs detectou como candidata a *outlier* a componente *X* do ponto 10, a componente *Y* do ponto 7, a componente *Z* do ponto 33 e a componente *2D* dos pontos 10 e 20. Já no teste de Chauvenet, as componentes *X* e *2D* dos pontos 10 e 20 foram detectadas como candidatas a *outliers*. Esse mesmo teste detectou como candidata a *outlier* a componente *Y* dos pontos 7, 29 e 30; e a componente *Z* do ponto 33.

Comparando os resultados das Tabelas 17 e 18, a base de dados VC33 foi a que mais teve componentes de pontos identificadas como candidatas a *outliers* pelos métodos inferenciais.

5.4. Comparação entre os métodos e remoção de *outliers*

Após a detecção das discrepâncias candidatas a *outliers*, aplicou-se o critério para a remoção dos pontos de checagem considerados *outliers* para permitir a avaliação da qualidade do produto cartográfico. A remoção de um ponto ocorreu se ele teve pelo menos uma

componente detectada como candidata a *outlier* simultaneamente por pelo menos um método descritivo e um método inferencial. Esses pontos estão identificados pelo símbolo “#” nas Tabela 17 e Tabela 18. Vale salientar que nesse trabalho foi adotado o critério que um ponto é considerado candidato a *outlier* se ao menos uma de suas discrepâncias posicionais (X, Y, Z ou 2D) é candidata a *outlier*.

Partindo-se desse princípio, para a base AL26 não foi removido nenhum ponto. Para a base IT30 foram removidos os pontos 11, 13 e 15. Já para o conjunto VC30, foi retirado o ponto 19. Para VC24 foram removidos os pontos 3 e 6. E, por fim, para VC33 houve o maior número de pontos removidos, sendo eles o 7, 10, 20, 29, 30 e 33.

5.5. Avaliação da acurácia posicional

Tanto as bases de dados originais como as bases de dados com os pontos *outliers* removidos foram avaliadas quanto a sua acurácia posicional, seguindo a metodologia de Santos et al. (2016), para a escala de cada produto. Os resultados das avaliações da acurácia posicional planimétrica para as cinco bases de dados estão disponíveis na Tabela 19.

Tabela 19: Avaliação da acurácia posicional planimétrica aplicada para as cinco bases de dados originais e para as bases após a remoção de *outliers*.

Planimetria									
Base de dados	Escala testada	Sem a remoção de outliers				Com a remoção de outliers			
		Análise de Precisão Decreto 89.817 e ET-CQDG	Análise de Tendência		Análise da Acurácia	Análise de Precisão Decreto 89.817 e ET-CQDG	Análise de Tendência		Análise da Acurácia
			Teste t de Student	Média Direcional			Teste t de Student	Média Direcional	
AL26	1:25.000	Classe A	Não tendencioso	Não tendencioso	Acurado	Classe A	Não tendencioso	Não tendencioso	Acurado
IT30	1:50.000	Classe C	Tendencioso	Não Tendencioso	Não acurado	Classe C	Tendencioso	Não Tendencioso	Não acurado
VC30	1:50.000	Classe C	Inconclusivo	Não tendencioso	Acurado	Classe C	Não tendencioso	Não tendencioso	Acurado
VC24	1:25.000	Classe C	Inconclusivo	Não tendencioso	Acurado	Classe B	Não tendencioso	Não tendencioso	Acurado
VC33	1:10.000	Classe D	Tendencioso	Tendencioso	Não Acurado	Classe D	Tendencioso	Tendencioso	Não Acurado

Pode ser observado (Tabela 19) que não houve nenhuma mudança para a base AL26, já que nenhum ponto foi removido e que essa base de dados atingiu a classificação máxima, ou seja, classe A com acurácia.

Verificou-se que a base IT30 original com todos os pontos e após a remoção de *outliers*, apesar de ser qualificada na Classe C segundo o Decreto e a ET-ADGV, não pode ser considerada acurada. Isso porque para ambos os casos a base foi apontada como tendenciosa pelo teste t de Student e, de acordo com Santos et al. (2016), uma base só é acurada se for livre de tendência e precisa.

Similar à base IT30, a base VC30 também teve o padrão de distribuição de pontos agrupado, mas o agrupamento não impediu, neste caso, que a base fosse classificada. Apesar do resultado inconclusivo da aplicação do teste t de Student, com o emprego da média direcional em conjunto com a variância circular, foi possível verificar a ausência de tendência e, assim, realizar a classificação para VC30, que foi enquadrada na Classe C antes e após a remoção dos *outliers*.

Ao comparar as classificações obtidas para o produto cartográfico da base VC24, notou-se, através da Tabela 19, que a remoção dos *outliers* influenciou no resultado da avaliação da acurácia. Isso porque, para a escala 1:25.000, o resultado utilizando todos os pontos da base de dados foi Classe C, enquanto que, ao remover os pontos 3 e 6, o resultado para a mesma escala foi a Classe B. Além disso, ao remover os pontos *outliers* o teste t de Student passou a ser conclusivo (não tendencioso) ao contrário do resultado obtido para a base de dados original. Esse caso é esperado, já que ao retirar os valores *outliers* há a possibilidade dos dados apresentarem normalidade e assim ser possível a aplicação do teste t de Student.

Já a base de dados VC33 não foi considerada acurada tanto na avaliação planimétrica da base de dados original, quanto na avaliação planimétrica após o procedimento de remoção de *outliers*. O aspecto que impediu que essa base fosse acurada foi a tendência, na qual o teste t de Student e a média direcional/variância circular apontaram um comportamento tendencioso.

Além da avaliação da acurácia planimétrica, para essa base de dados foi realizada também sua avaliação quanto à acurácia posicional altimétrica, visto que a base possui a componente altimétrica. O resultado da avaliação encontra-se na Tabela 20.

Tabela 20: Avaliação da acurácia posicional altimétrica aplicada para a base VC33 sem e com remoção de pontos *outliers*.

VC 33 – Altimetria						
Equidistância da curva de nível (m)	Sem a remoção de outliers			Com a remoção de outliers		
	Análise de Precisão Decreto 89.817 e ET-CQDG	Análise de Tendência Teste t de Student	Análise da Acurácia	Análise de Precisão Decreto 89.817	Análise de Tendência Teste t de Student	Análise da Acurácia
5	Classe B	Não tendencioso	Acurado	Classe B	Não tendencioso	Acurado

Analisando a Tabela 20, notou-se que para a altimetria a base de dados foi classificada na classe B e que se mostrou livre de tendência, logo sendo acurada posicionalmente.

6. Discussão

As bases de dados utilizadas neste estudo provêm de diferentes localidades e se diferenciam quanto à sua origem, método de obtenção dos pontos de checagem, escala e tamanho da área, etc. Ao comparar as estatísticas descritivas, nota-se que as componentes X e $2D$ da base de dados VC33 foram as que apresentaram as maiores magnitudes de médias de discrepâncias e menor coeficiente de variação (Tabela 9). Além disso, essas componentes apresentaram o maior número de detecções de candidatos a *outlier* pelo método 3σ Alternativo (Tabela 14).

A fim de avaliar o efeito da presença de pontos *outliers* na homogeneidade de cada base de dados, foram calculados os valores do coeficiente de variação após a remoção de *outliers* e observou-se que para algumas componentes de discrepâncias o coeficiente de variação tende a abaixar, mas para outros aumenta, mostrando uma aleatoriedade. Este fato mostra que, a princípio, não há uma correlação entre a remoção de pontos *outliers* e o coeficiente de variação, conforme pode ser observado na Tabela 21.

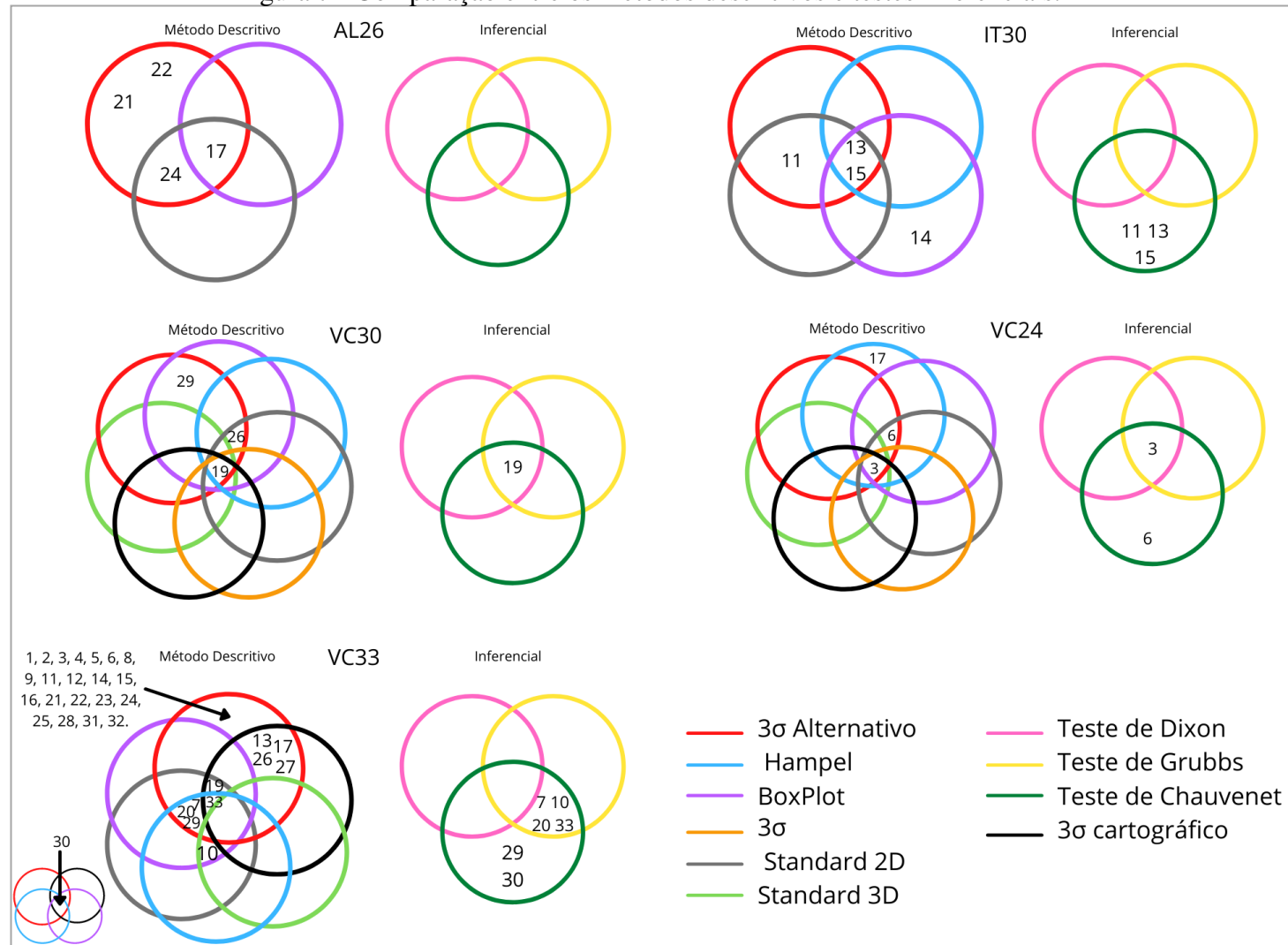
Tabela 21: Coeficiente de variação das cinco bases de dados sem e com a remoção de pontos *outliers*.

Base de dados	Discrepância	Coeficiente de variação (%)	
		Sem remoção de <i>outliers</i>	Com remoção de <i>outliers</i>
AL26	d_x	-3825,67%	-3815,10%
	d_y	363,02%	363,17%
	d_{2D}	56,14%	56,14%
IT30	d_x	-388,79%	591,77%
	d_y	-151,28%	173,30%
	d_{2D}	54,28%	44,82%
VC30	d_x	-5805,35%	2079,13%
	d_y	880,65%	3168,29%
	d_{2D}	56,62%	46,12%
VC24	d_x	3671,96%	342,30%
	d_y	-317,07%	-453,91%
	d_{2D}	82,46%	61,57%
VC33	d_x	-27,77%	-16,94%
	d_y	-166,00%	-72,91%
	d_z	663,50%	2291,99%
	d_{2D}	23,91%	15,70%

Além da estatística descritiva, analisou-se também o efeito do agrupamento nos resultados da detecção de *outliers*. Conforme visto na Figura 6, as bases IT30 e VC30 tiveram padrão de distribuição dos pontos agrupado, que a princípio não são indicados para realização do controle de qualidade cartográfica. Contudo, notou-se que mesmo a distribuição dos pontos não sendo aleatória ou dispersa, esse fato não impediu a detecção de candidatos a *outliers*, visto que foi possível aplicar tanto os métodos descritivos, quanto os inferenciais. Mas, salienta-se que em uma base de dados com pontos agrupados deve ser feito um estudo mais criterioso dos resultados, já que os pontos não são representativos para toda a área de estudo.

Para avaliar a interseção da detecção de pontos candidatos a *outliers* pelos métodos descritivos e de inferências foi construído um diagrama de Venn (Figura 7), no qual são apresentados os pontos em que pelo menos uma de suas componentes de discrepância, X , Y , Z ou $2D$, foram apontadas como candidatos a *outliers* nas bases de dados AL26, IT30, VC30, VC24 e VC33. O fato de um ponto ser detectado como candidato a *outlier* por um ou mais métodos (descritivo e inferencial) representa um forte indício de que ele seja de fato um *outlier*. No diagrama de Venn (Figura 7) foram apresentados apenas os resultados dos métodos que detectaram pelo menos um ponto.

Figura 7 - Comparação entre os métodos descritivos e testes inferenciais.



Fonte: A autora.

O ponto 17 da base de dados AL26 foi detectado como candidato a *outlier* pelos métodos descritivos 3σ Alternativo, BoxPlot e Método Standard Deviation 2D, mas nenhum ponto foi identificado pelos testes inferenciais (Figura 7). Na base IT30 os pontos 11, 13 e 15 foram detectados como *outliers* pelo Teste de Chauvenet e pelos métodos descritivos 3σ alternativo, Hampel, Boxplot e Standard Deviation 2D. Os pontos 19 e 3 das bases VC30 e VC24, respectivamente, foram apontados como candidatos a *outliers* em todos os métodos inferenciais e descritivos. A base de dados VC33 teve muitos valores de discrepâncias negativas, o que ocasionou em inúmeras detecções de candidatos a *outliers* pelos métodos descritivos (Tabela 14). Entretanto, a aplicação dos testes inferenciais não detectou esses pontos, dessa forma, a falta de normalidade dessa base pode ter influenciado no desempenho dos testes inferenciais, mas não pode ser considerada como a única responsável pelo desempenho dos testes.

A princípio, os métodos descritivos tendem a identificar um número maior de pontos como candidatos a *outliers*. Isso ocorre pois, em sua metodologia, as discrepâncias, de todos os pontos são testadas, o que aumenta a chance de detecção de mais pontos. Já para os métodos inferenciais isso não se aplica, pois cada método tem um procedimento diferente de aplicação. Os testes Dixon e Chauvenet têm metodologias semelhantes, testando inicialmente se o menor e o maior valor de cada componente da discrepância presente na base de dados são candidatas a *outliers*, sendo os próximos valores somente testados se o menor ou maior forem detectados, caso contrário, o cálculo é interrompido. Já o Teste de Grubbs somente testa o menor e maior valor de cada componente, ou seja, independente do resultado, as outras discrepâncias subsequentes não são testadas. Dessa forma, o teste de Grubbs limita a detecção de candidatos a *outliers*, visto que a base de dados pode ter mais do que duas discrepâncias como *outliers* e consequentemente mais do que dois pontos como *outliers* e esses pontos não serão identificados. Já os Testes de Dixon e Chauvenet permitem testar em mais de dois pontos, dessa forma, poderia ser aplicado a todos os pontos da base de dados, o que é uma vantagem, pois dessa forma todos os pontos podem ser verificados.

Outra diferença dos testes inferenciais em relação aos métodos descritivos é que os inferenciais pressupõem a normalidade dos dados. Quanto ao desempenho dos testes inferenciais aplicados a dados não normais, os resultados obtidos indicam que os métodos avaliados são robustos à violação a pressuposição de normalidade, visto que todos os pontos que foram acusados pelos métodos inferenciais foram mencionados por pelo menos um

método descritivo. Dessa forma, mesmo que a premissa de normalidade não tenha sido atendida, os testes se mostraram eficientes.

Em um estudo realizado por Fonseca, Marques e Gripp (2016), o método 3σ cartográfico foi utilizado na etapa de detecção de candidatos a *outliers* na avaliação da acurácia posicional de uma ortoimagem. Neste caso em questão, este método detectou dois pontos como candidatos a *outliers*. Entretanto, após uma investigação, notou-se que não se tratava de *outliers*, mas de um problema de distorção na imagem, ou seja, um erro sistemático. Por isso antes de remover um ponto considerado como candidato a *outlier* deve ser feita uma análise criteriosa para verificar se realmente se trata de um *outlier*.

Em relação à normalidade, que é uma premissa para aplicação dos testes inferenciais, notou-se pela Tabela 22 que após a remoção dos pontos *outliers*, para as bases IT30 e VC33 a componente 2D que anteriormente eram não normais (Tabela 16), passaram a apresentar normalidade. As bases VC30 e VC24 para a componente Y também apresentaram normalidade após a remoção dos *outliers*, assim como a componente X da base VC33.

Tabela 22: Resultado do teste de Shapiro-Wilk para as discrepâncias posicionais obtidas para os pontos de checagem de cada uma das cinco bases de dados após a remoção dos pontos *outliers*.

Teste Shapiro-Wilk após remoção de <i>outliers</i>								
Base de dados	Valor-p d_x	Resultado	Valor-p d_y	Resultado	Valor-p d_z	Resultado	Valor-p d_{2D}	Resultado
AL26	0,3935	Normal	0,9462	Normal	-	-	0,5825	Normal
IT30	0,7381	Normal	0,8842	Normal	-	-	0,2693	Normal
VC30	0,3641	Normal	0,4560	Normal	-	-	0,0234	Não Normal
VC24	0,6545	Normal	0,1890	Normal	-	-	0,0469	Não Normal
VC33	0,7216	Normal	0,8164	Normal	0,9369	Normal	0,5863	Normal

No trabalho de Silva (2020), no qual foi realizada a avaliação da acurácia posicional altimétrica de Modelos Digitais de Elevação (MDE), a detecção de *outliers* foi feita pelo método BoxPlot e, após a remoção de quatro pontos, notou-se que a distribuição da amostra que inicialmente era não normal, se tornou normal.

Essa mudança no comportamento da normalidade dos dados também foi observada no trabalho de Carvalho (2017), no qual foram avaliados Modelos Digitais de Elevação (MDE), Modelos Digitais de Superfície (MDS) e Modelos Digitais de Terreno (MDT) quanto à

acurácia altimétrica, sendo a etapa de detecção de candidatos a *outliers* feita através do método BoxPlot ajustado. Na comparação feita com os dados originais e aqueles livres de *outliers* quanto à normalidade, após a aplicação do teste Shapiro-Wilk, percebeu-se que em alguns casos os MDS's não apresentaram normalidade após remoção de *outliers*, comportamento contrário ao ocorrido na base de dados original e que em alguns casos os MDT's se tornaram normais, contrário ao resultado da base original. Assim, o autor concluiu que os pontos *outliers* têm a capacidade de interferir nos resultados dos testes de normalidade, mas não podem ser considerados como os únicos responsáveis por esta alteração.

Para definir se um ponto candidato a *outlier* pelos métodos descritivos e inferenciais é de fato um *outlier*, deve ser realizado um estudo acerca da forma de coleta dos pontos de teste e de referência obtidos. Porém, não foi possível a realização desta análise visto que as bases de dados foram retiradas de outros trabalhos, logo se utilizou o critério de coincidência de detecção de pontos por pelo menos um método descritivo e inferencial.

Após remover os pontos considerados *outliers*, fez-se a avaliação da acurácia posicional nas bases de dados originais, ou seja, com todos os pontos presentes e na base livre de *outliers*.

Notou-se a partir dos resultados da avaliação da acurácia posicional planimétrica apresentados na Tabela 19 que a base de dados AL26 foi aquela que teve a melhor classificação em comparação com as demais, atingindo classe A. Para essa base não foi removido nenhum ponto, já que nenhum método inferencial detectou pontos candidatos a *outliers* (Tabela 17). O resultado da avaliação neste trabalho foi semelhante ao encontrado no trabalho de Santos (2010), no qual essa base também obteve classe A para a escala 1:25.000.

Verificou-se que a base IT30 com e sem a remoção de *outliers* não foi considerada acurada. Ao avaliar o resultado presente na Figura 6 concluiu-se a base IT30 tem padrão de distribuição dos pontos de referência agrupado, portanto o agrupamento pode ter afetado o resultado da avaliação, visto que os pontos utilizados não foram representativos de toda a extensão da área. Dessa forma, para finalidades que exijam um rigor quanto à qualidade, essa base de dados não é indicada para ser utilizada.

Ao verificar o resultado da avaliação da acurácia posicional planimétrica para essa base de dados feita por Marotta e Vieira (2005), os autores classificaram a base IT30 como classe B para a escala 1:50.000 utilizando como critério os valores presentes do Decreto nº 89.817 (Tabela 1), que corresponde a Classe C da ET-ADGV (Tabela 2), que foi o critério utilizado neste trabalho. Ainda no estudo citado, o procedimento para análise de tendência

apontou dados livres de tendência, porém no presente estudo o teste t de Student acusou tendência para a componente Y, dessa forma, como há tendência, a base IT30 não é acurada.

Já para a base VC30, no artigo no qual foram extraídos os dados, Marotta e Vieira (2005) obtiveram como resultado a Classe B, utilizando o Decreto como referência, que corresponde a Classe C da ET-ADGV, que foi o resultado obtido nesse trabalho, já que foi considerado como referência a Tabela 2. Os autores citados também aplicaram o teste t de Student para avaliar a tendência e tiveram um resultado semelhante ao obtido no presente estudo, tendo como resultado dados livres de tendência e considerados acurados.

Ao comparar as classificações obtidas para o produto cartográfico da base VC24 sem e com remoção de pontos *outliers*, notou-se que houve uma melhora no resultado da classificação, pois a classificação obtida antes da remoção dos pontos *outliers* que era Classe C passou a ser Classe B. No estudo feito por Santos e Vieira (2009), não foram utilizados os 24 pontos, mas apenas 21, obtendo-se como resultado que a base de dados é acurada para Classe A.

Por fim, conforme visto na Tabela 19, a base VC33 não foi considerada acurada para a acurácia planimétrica e as discrepâncias das componentes X, Y e 2D dos pontos da base VC33 não apresentaram distribuição de probabilidades normal (Tabela 16). Essa falta de normalidade pode ter influenciado na não classificação desta base de dados para a acurácia planimétrica.

Porém, para essa mesma base sua avaliação altimétrica (Tabela 20) resultou em uma base acurada para a classe B, resultado semelhante ao obtido no estudo de Santos e Vieira (2009).

A partir dos resultados das avaliações das cinco bases de dados, notou-se que a ausência de *outliers* proporciona uma melhora no desempenho da aplicação do teste t de Student. Isso porque em alguns casos, como mostrado na Tabela 19, este teste havia se mostrado inconclusivo ao ser aplicado nas bases de dados originais, mas após a remoção de *outliers* este teste foi possível de ser aplicado, apresentando resultados conclusivos. Ainda, em algumas bases como a VC24, notou-se uma melhora nos resultados da classificação com a ausência de *outliers*.

A variação na classificação dos produtos cartográficos também ocorreu no trabalho de Carvalho (2017), visto que houve situações em que um produto original que anteriormente possuía classificação PEC inferior a D passou a ter classificação B e outro caso em que um Modelo Digital de Superfície com a presença de *outliers* foi classificado como inferior a D e passou para classe A, na escala 1:25.000, após a remoção dos *outliers*.

No trabalho de Silva (2020), após a remoção de quatro *outliers* da área 01, o produto que anteriormente havia atingido classe D na análise de precisão e estava sem classificação final devido à tendência, alcançou a classe A e se tornou livre de tendência. Em contrapartida, ao avaliar o segundo produto, da área 02, no qual foi removido apenas um *outlier*, a tendência presente nos dados e a classificação atingida não foram alteradas com a retirada do ponto *outlier*.

Dessa forma, o autor concluiu que os pontos *outliers* existentes em uma amostra de discrepâncias influenciam a avaliação posicional altimétrica de um Modelo Digital de Elevação e que a detecção e remoção desses podem melhorar os resultados dessa avaliação, sugerindo que seja inserida a etapa de detecção de *outliers* nos padrões de qualidade cartográfica estabelecidos no Brasil (SILVA, 2020).

Assim, como visto nos trabalhos citados e nos resultados aqui expostos, a presença de *outliers* afeta o resultado da avaliação da acurácia posicional em produtos cartográficos, alterando a classificação, como também na tendência e normalidade dos dados.

7. Conclusão

A detecção de candidatos a *outliers* é uma etapa importante que deve ser levada em consideração durante o controle de qualidade cartográfica, visto que *outliers* presentes nos pontos de checagem podem levar a avaliações errôneas, cujo resultado não estará condizente com a realidade. Em especial para a avaliação da acurácia posicional, notou-se que a presença de *outliers* pode afetar tanto a precisão, quanto a tendência. Dessa forma, o produto poderá ter sua classificação comprometida, levando a conclusões equivocadas e decisões não assertivas em que o produto for utilizado.

Quanto aos procedimentos utilizados para a detecção de candidatos a *outliers*, os métodos descritivos apresentaram como vantagem a ausência de pressuposições para sua aplicação e a possibilidade de aplicação dos métodos em todos os pontos e para todas as discrepâncias posicionais, permitindo que toda a base de dados seja testada. Além disso, conforme visto, o método 3σ Alternativo se mostrou eficiente na detecção de erros sistemáticos presentes em uma das bases de dados.

Por outro lado, os testes inferenciais têm a pressuposição de normalidade para sua aplicação e alguns deles não permitem que todos os pontos sejam testados, mas apesar disso, se mostraram robustos em alguns casos em que a premissa da normalidade não foi cumprida. Diante disso, nota-se que os métodos descritivos e testes inferências se comportam como complementares, pois atuam de formas distintas, mas como o mesmo objetivo final. Dessa forma, a junção dos métodos descritivos e testes inferenciais na detecção de *outliers* se torna uma boa ferramenta, pois combinam metodologias diferentes aplicadas em uma mesma base de dados, permitindo uma conclusão mais assertiva.

Os métodos utilizados neste trabalho foram eficientes para as bases de dados em que foram aplicadas, mas sugere-se que outros métodos sejam testados, a fim de encontrar aqueles que melhor se adequem aos dados utilizados. Assim, sugere-se que durante a avaliação da acurácia posicional sejam aplicados pelo menos um método descritivo e um teste inferencial para detecção de candidatos a *outliers* e que as novas revisões das normas que regulamentam o controle de qualidade cartográfica incluam como obrigatória a etapa de detecção de *outliers*.

Sugerem-se, ainda, para trabalhos futuros, que sejam analisadas novas metodologias para detecção de candidatos a *outliers* e que seja avaliado seu efeito nas outras componentes do controle de qualidade cartográfica, como a completude, consistência lógica, acurácia temporal e acurácia temática.

Referências

- BRASIL. **Decreto nº 89.817, de 20 de Junho de 1984, Estabelece as Instruções Reguladoras das Normas Técnicas da Cartografia Nacional.**, 1984. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/1980-1989/D89817.htm>
- BRASIL. **Decreto nº 9.759, de 11 de abril de 2019, Extingue e estabelece diretrizes, regras e limitações para colegiados da administração pública federal.**, 2019.
- CALLEGARO, R. F. **Uma Arquitetura Para Fusão De Dados De Sensores De Baixo Custo Em Redes De Sensores Sem Fio.** [s.l.] (Dissertação de Mestrado) - Universidade Federal de Santa Catarina, 2014.
- CARVALHO, J. A. B. DE. **Avaliação da Acurácia Altimétrica de Modelos Digitais de Elevação no Brasil: Um Estudo Sobre a Distribuição dos Erros Altimétricos e a Influência de Valores Outliers Para o Processo de Análise.** [s.l.] Universidade Federal de Pernambuco, 2017.
- CONCAR. **Panorama Histórico.** Disponível em: <<https://www.concar.gov.br/panoramaHist.aspx>>. Acesso em: 5 jun. 2020.
- DIXON, W. J. Analysis of Extreme Values. **The Annals of Mathematical Statistics**, v. 21, n. 4, p. 488–506, 1950.
- DSG, (DIRETORIA DE SERVIÇO GEOGRÁFICO). **Norma Da Especificação Técnica Para Controle De Qualidade De Dados Geospaciais (ET-CQDG)**Brasil, 2016a.
- DSG, (DIRETORIA DE SERVIÇO GEOGRÁFICO). **Especificação Técnica para a Aquisição de Dados Geospaciais Vetoriais**, 2016b. Disponível em: <<http://www.geoportal.eb.mil.br/index.php/inde2?id=140/>>
- FERGUSON, T. S. Rules for Rejection of Outliers. **International Statistical Institute**, v. 29, n. 3, p. 29–43, 1961.
- FERREIRA, A. L. DOS S. **Comparação de diferentes técnicas para detecção e tratamento de outliers na determinação de fatores de medidores.** [s.l.] Pontífica Universidade Católica do Rio de Janeiro, 2017.
- FIGUEIRA, M. Identificação de Outliers. **Millenium**, n. 1983, 1998.
- FONSECA, I. G. R.; MARQUES, É. T.; GRIPP, J. J. **Avaliação da Acurácia Posicional de Uma Ortoimagem de Barão de Cocais para aplicação no Cadastro Territorial Multifinalitário Urbano.** 12º Congresso de Cadastro Técnico Multifinalitário e Gestão Territorial. **Anais...**Florianópolis: 2016
- FRANCISCO, C. N. **O USO DE SCANNERS NA DIGITALIZAÇÃO DE MAPAS DESTINADOS A SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS.** Congresso de Brasileiro de Cartografia. **Anais...**Rio de Janeiro: 1993, 1993

FREITAS, I. W. S. DE. **Um estudo comparativo de técnicas de detecção de outliers no contexto de classificação de dados**. [s.l.] Universidade do Estado do Rio Grande do Norte, 2019.

GHILANI, C. **Adjustament Computations**. 5. ed. New Jersey: [s.n.].

GLEN, S. **Chauvenet's Criterion**. Disponível em: <<https://www.statisticshowto.com/chauvenets-criterion/>>. Acesso em: 11 maio. 2021.

GRUBBS, F. E. Procedures for Detecting Outlying Observations in Samples. **Technometrics**, v. 11, n. 1, p. 1–21, 1969.

ISO. **Sampling procedures for inspection by attributes - Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection**. International Organization for Standardization (ISO), 1985.

ISO. **ISO 2859-1:1999. Sampling procedures for inspection by attributes - Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection**. International Organization for Standardization (ISO), 1999.

MAROTTA, G. S.; VIEIRA, C. A. O. **Aplicação do Padrão de Exatidão Cartográfica em Imagens Orbitais ASTER Para Fins de Atualizações de Mapeamentos**. In: Congresso Brasileiro de Cartografia, 22. **Anais...**2005

MONICO, J. F. G. et al. Acurácia e Precisão: Revendo Os Conceitos De Forma Acurada. **Boletim de Ciências Geodésicas**, v. 15, n. 3, p. 469–483, 2009.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 8. ed. [s.l: s.n.].

NERO, M. A. **Proposta para o controle de Qualidade de Bases Cartográficas com Ênfase na Componente Posicional**. [s.l.] (Tese de Doutorado, Escola Politécnica da Universidade de São Paulo) - Universidade de São Paulo, 2005.

OLIVEIRA, CLEDSON D; CAROLI, ADHEMAR A.; AMARAL, A. S. . V. O. L. Detecção de Fraudes, Anomalias e Erros em Análise de Dados Contábeis: Um Estudo com Base em Outliers. **Revista Eletrônica do Departamento de Ciências Contábeis & Departamento de Atuária e Métodos Quantitativos (REDECA)**, v. 1, n. 1, p. 102–127, 2014.

OLIVEIRA, E. C. DE. **Comparação das Diferentes Técnicas para a Exclusão de “OUTLIERS”**. In: Congresso da Qualidade em Metrologia. **Anais...**São Paulo: 2008

SANTOS, A. P. **Avaliação da Acurácia Posicional em Dados Espaciais com o uso de Estatística Espacial**. Viçosa: (Dissertação de Mestrado em Engenharia Civil) - Universidade Federal de Viçosa), 2010.

SANTOS, A. DE P. DOS et al. Avaliação da Acurácia Posicional em Dados Espaciais Utilizando Técnicas de Estatística Espacial: Proposta de Método e Exemplo Utilizando a Norma Brasileira. **Boletim de Ciências Geodésicas**, p. 630–650, 2016.

SANTOS, A. DE P. DOS; VIEIRA, C. A. O. **Avaliação do Padrão de Exatidão**

Cartográfico em imagens orbitais IKONOS e CBERS-2B, na bacia do Ribeirão São Bartolomeu em Viçosa-MG. In: Simpósio Brasileiro de Sensoriamento Remoto, 15. Anais...Curitiba: 2009

SEO, S.; MARSH, G. M. A review and comparison of methods for detecting outliers in univariate data sets. **Department of Biostatistics, Graduate School of Public Health**, p. 1–53, 2006.

SILVA, D. DA C. E. **Estudo Da Qualidade Posicional E Análise De Padrões Espaciais Na Distribuição De Erros Altimétricos Em Modelos Digitais De Elevação.** [s.l.] Universidade Federal de Pernambuco, 2020.

SILVA, F. R. **Uma abordagem para Detecção de Outliers em Dados Categóricos.** [s.l.] (Dissertação de Mestrado no Instituto de Computação) - Universidade Estadual de Campinas, 2004.

WITTE, R. S.; WITTE, J. S. **Estatística.** 7. ed. Rio de Janeiro: LTC, 2005.

WONG, D. W. S.; LEE, J. **Statistical Analysis of Geographic Information System With ArcView GIS and ArcGIS (SAGI).** New Jersey: Inc. Hoboken, 2005.

ZHAO, Y. et al. Outlier detection rules for fault detection in solar photovoltaic arrays. **Conference Proceedings - IEEE Applied Power Electronics Conference and Exposition - APEC**, p. 2913–2920, 2013.

APÊNDICE 1

Tabela 23: Base de dados AL26 e suas respectivas discrepâncias posicionais.

AL26					Discrepâncias Posicionais		
N°	E teste (m)	N teste (m)	E referência (m)	N referência (m)	d_x (m)	d_y (m)	d_{2D} (m)
1	684405,087	7787965,778	684407,944	7787963,872	-2,857	1,906	3,434
2	684862,797	7789104,451	684864,903	7789108,431	-2,106	-3,980	4,503
3	684610,271	7783543,003	684609,840	7783544,746	0,431	-1,743	1,795
4	687404,600	7786399,575	687403,647	7786402,428	0,953	-2,853	3,008
5	693198,021	7781420,890	693197,048	7781420,913	0,973	-0,023	0,973
6	702155,480	7780512,252	702153,962	7780511,053	1,518	1,199	1,934
7	700340,034	7780606,478	700342,248	7780605,376	-2,214	1,102	2,473
8	709071,590	7784948,764	709075,683	7784945,252	-4,093	3,512	5,393
9	684901,527	7779743,317	684900,886	7779744,122	0,641	-0,805	1,029
10	689009,141	7775877,464	689009,133	7775872,760	0,008	4,704	4,704
11	695890,825	7777099,256	695888,721	7777098,732	2,104	0,524	2,168
12	701844,226	7774412,988	701844,077	7774411,327	0,149	1,661	1,668
13	706379,407	7776556,264	706379,459	7776554,489	-0,052	1,775	1,776
14	702302,800	7767121,600	702307,166	7767121,322	-4,366	0,278	4,375
15	710776,554	7779534,763	710776,308	7779532,460	0,246	2,303	2,316
16	707804,097	7767252,014	707807,507	7767250,007	-3,410	2,007	3,957
17	678133,125	7768365,479	678126,830	7768367,309	6,295	-1,830	6,556
18	681668,259	7772775,759	681665,994	7772777,359	2,265	-1,600	2,773
19	687292,880	7768232,073	687291,843	7768231,326	1,037	0,747	1,278
20	693532,969	7767297,965	693532,220	7767293,642	0,749	4,323	4,387
21	683811,677	7760139,597	683806,822	7760135,605	4,855	3,992	6,285
22	689889,879	7759791,862	689889,608	7759786,074	0,271	5,788	5,794
23	693225,581	7762511,979	693225,443	7762511,869	0,138	0,110	0,176
24	701098,588	7762645,391	701104,282	7762649,846	-5,694	-4,455	7,230
25	698807,095	7758048,848	698803,536	7758050,063	3,559	-1,215	3,761
26	701872,028	7755359,852	701875,357	7755358,722	-3,329	1,130	3,516

Tabela 24: Base de dados IT30 e suas respectivas discrepâncias posicionais.

Nº	IT30		Discrepâncias Posicionais				
	E referência (m)	N referência (m)	E teste (m)	N teste (m)	d_x (m)	d_y (m)	d_{2D} (m)
1	626284,568	7760015,534	626296,711	7760019,503	12,143	3,969	12,775
2	626486,908	7759908,171	626506,117	7759908,448	19,209	0,277	19,211
3	625703,330	7759606,336	625703,109	7759620,879	-0,221	14,543	14,545
4	625393,664	7759157,526	625389,484	7759168,697	-4,180	11,171	11,927
5	625053,925	7758844,627	625056,867	7758826,251	2,942	-18,376	18,610
6	625203,425	7756909,624	625221,694	7756908,186	18,269	-1,438	18,326
7	624749,316	7755948,728	624735,228	7755952,157	-14,088	3,429	14,499
8	624919,001	7755763,864	624914,398	7755757,092	-4,603	-6,772	8,188
9	625006,336	7755674,513	624993,974	7755680,828	-12,362	6,315	13,882
10	625088,236	7754966,099	625082,193	7754967,754	-6,043	1,655	6,266
11	625354,189	7754305,847	625345,852	7754270,701	-8,337	-35,146	36,121
12	625739,833	7754304,348	625715,173	7754289,847	-24,660	-14,501	28,608
13	626446,063	7754063,010	626444,362	7754022,498	-1,701	-40,512	40,548
14	626439,003	7753601,853	626410,930	7753601,286	-28,073	-0,567	28,079
15	626573,960	7752820,911	626539,618	7752803,500	-34,342	-17,411	38,503
16	626861,852	7752689,195	626856,058	7752664,131	-5,794	-25,064	25,725
17	624726,102	7755809,953	624727,080	7755799,761	0,978	-10,192	10,239
18	624714,517	7755793,590	624711,345	7755784,965	-3,172	-8,625	9,190
19	624702,304	7755356,884	624716,025	7755344,886	13,721	-11,998	18,227
20	624615,957	7755153,001	624609,488	7755146,568	-6,469	-6,433	9,123
21	624605,543	7755000,622	624609,986	7754995,683	4,443	-4,939	6,643
22	624565,950	7754948,506	624573,126	7754941,556	7,176	-6,950	9,990
23	624374,067	7754940,608	624374,396	7754925,833	0,329	-14,775	14,779
24	624208,623	7755004,465	624215,551	7754993,851	6,928	-10,614	12,675
25	623903,301	7755035,556	623918,088	7755013,499	14,787	-22,057	26,555
26	623511,964	7754618,666	623492,762	7754622,807	-19,202	4,141	19,643
27	623062,680	7753415,574	623054,039	7753394,591	-8,641	-20,983	22,693
28	623183,124	7753245,033	623177,848	7753224,722	-5,276	-20,311	20,985
29	623083,459	7753535,185	623073,330	7753535,763	-10,129	0,578	10,145
30	623170,193	7753641,425	623167,277	7753638,343	-2,916	-3,082	4,243

Tabela 25: Base de dados VC30 e suas respectivas discrepâncias posicionais.

Nº	VC30				Discrepâncias Posicionais		
	E referência (m)	N referência (m)	E teste (m)	N teste (m)	d_x (m)	d_y (m)	d_{2D} (m)
1	722604,783	7704280,330	722603,125	7704270,500	-1,658	-9,830	9,969
2	723200,124	7704641,765	723209,001	7704630,445	8,877	-11,320	14,386
3	723697,630	7704835,032	723694,313	7704832,500	-3,317	-2,532	4,173
4	724460,043	7704912,558	724445,354	7704904,401	-14,689	-8,157	16,802
5	724624,632	7704638,519	724612,338	7704631,324	-12,294	-7,195	14,245
6	724896,900	7704146,279	724887,462	7704142,464	-9,438	-3,815	10,180
7	724949,776	7703951,493	724939,353	7703951,526	-10,423	0,033	10,423
8	724924,335	7703268,896	724903,771	7703270,571	-20,564	1,675	20,632
9	724906,537	7702860,513	724880,925	7702861,702	-25,612	1,189	25,640
10	725150,446	7702463,944	725136,688	7702457,500	-13,758	-6,444	15,192
11	725077,813	7704532,500	725092,147	7704538,153	14,334	5,653	15,408
12	725206,340	7701713,445	725193,747	7701710,646	-12,593	-2,799	12,900
13	724961,015	7700250,481	724970,875	7700250,500	9,860	0,019	9,860
14	724994,704	7699733,234	725005,859	7699732,814	11,155	-0,420	11,163
15	725217,764	7698593,141	725232,329	7698601,841	14,565	8,700	16,966
16	725902,836	7697680,336	725914,750	7697688,500	11,914	8,164	14,443
17	724953,786	7699836,477	724961,688	7699845,500	7,902	9,023	11,994
18	724902,562	7700183,526	724915,625	7700189,500	13,063	5,974	14,364
19	723829,516	7700509,143	723841,951	7700553,761	12,435	44,618	46,318
20	723645,766	7700667,276	723660,063	7700675,500	14,297	8,224	16,494
21	722575,562	7702209,498	722576,875	7702220,000	1,313	10,502	10,584
22	721744,167	7702950,321	721732,167	7702939,829	-12,000	-10,492	15,940
23	721347,240	7702962,383	721351,458	7702966,992	4,218	4,609	6,248
24	721669,813	7702337,799	721675,063	7702339,162	5,250	1,363	5,424
25	721345,436	7701976,218	721346,188	7701988,500	0,752	12,282	12,305
26	721041,793	7701490,683	721072,188	7701475,500	30,395	-15,183	33,976
27	720661,708	7700916,020	720658,313	7700919,500	-3,395	3,480	4,862
28	725077,813	7704532,500	725092,147	7704538,153	14,334	5,653	15,408
29	724712,717	7705085,042	724684,813	7705082,000	-27,904	-3,042	28,069
30	724801,819	7704893,740	724787,500	7704881,500	-14,319	-12,240	18,837

Tabela 26: Base de dados VC24 e suas respectivas discrepâncias posicionais.

VC24					Discrepâncias Posicionais		
N°	E teste (m)	N teste (m)	E referência (m)	N referência (m)	d_x (m)	d_y (m)	d_{2D} (m)
1	724071,705	7694725,257	724060,886	7694729,831	10,819	-4,574	11,746
2	723467,237	7695313,278	723455,329	7695316,02	11,908	-2,742	12,22
3	722216,591	7695327,259	722230,822	7695353,083	-14,231	-25,824	29,486
4	722028,577	7696512,913	722025,693	7696512,03	2,884	0,883	3,016
5	723420,568	7696837,723	723416,141	7696840,825	4,427	-3,102	5,406
6	719756,429	7697637,015	719775,952	7697639,838	-19,523	-2,823	19,726
7	720874,881	7698604,639	720872,535	7698612,241	2,346	-7,602	7,956
8	722001,28	7699516,737	722001,057	7699521,941	0,223	-5,204	5,209
9	719086,209	7699286,704	719094,454	7699283,073	-8,245	3,631	9,009
10	721550,222	7698869,221	721539,537	7698871,799	10,685	-2,578	10,992
11	720718,371	7696257,972	720721,239	7696255,682	-2,868	2,29	3,67
12	720946,259	7699747,774	720943,516	7699749,379	2,743	-1,605	3,178
13	717928,92	7700702,5	717925,152	7700705,535	3,768	-3,035	4,838
14	722099,144	7701330,958	722094,28	7701330,719	4,864	0,239	4,87
15	723146,66	7701696,783	723142,67	7701702,809	3,99	-6,026	7,227
16	717137,242	7702435,711	717137,426	7702434,555	-0,184	1,156	1,171
17	718490,168	7703093,353	718503,096	7703104,544	-12,928	-11,191	17,099
18	719314,203	7704418,455	719314,285	7704415,607	-0,082	2,848	2,849
19	719943,7	7705672,359	719938,822	7705669,764	4,878	2,595	5,525
20	722076,049	7704302,541	722068,179	7704297,595	7,87	4,946	9,295
21	719445,283	7701969,367	719449,067	7701966,134	-3,784	3,233	4,977
22	721969,626	7697604,24	721967,783	7697601,571	1,843	2,669	3,243
23	719780,223	7700684,428	719784,691	7700681,127	-4,468	3,301	5,555
24	721306,341	7703321,658	721308,136	7703322,207	-1,795	-0,549	1,877

Tabela 27: Base de dados VC33 e suas respectivas discrepâncias posicionais.

VC33							Discrepâncias posicionais			
Nº	E teste (m)	N teste (m)	h teste (m)	E referência (m)	N referência (m)	h ref (m)	d_x (m)	d_y (m)	d_z (m)	d_{2D} (m)
1	724055,795	7694728,233	782,411	724060,886	7694729,831	782,025	-5,091	-1,598	0,386	5,336
2	723451,536	7695314,462	746,000	723455,329	7695316,020	745,392	-3,793	-1,558	0,608	4,101
3	722225,460	7695351,530	773,920	722230,822	7695353,083	775,337	-5,362	-1,553	-1,417	5,582
4	722020,364	7696510,396	735,539	722025,693	7696512,030	735,914	-5,329	-1,634	-0,375	5,574
5	723412,116	7696839,863	726,365	723416,141	7696840,825	725,439	-4,025	-0,962	0,926	4,138
6	719769,436	7697639,115	722,832	719775,952	7697639,838	722,605	-6,516	-0,723	0,227	6,556
7	720868,718	7698614,608	690,967	720872,535	7698612,241	689,478	-3,817	2,367	1,489	4,491
8	721994,634	7699520,877	734,771	722001,057	7699521,941	735,543	-6,423	-1,064	-0,772	6,511
9	719088,909	7699282,875	705,000	719094,454	7699283,073	706,344	-5,545	-0,198	-1,344	5,549
10	721539,620	7698870,865	698,000	721539,537	7698871,799	697,731	0,083	-0,934	0,269	0,938
11	720716,415	7696254,906	739,648	720721,239	7696255,682	738,447	-4,824	-0,776	1,201	4,886
12	720939,474	7699748,686	680,000	720943,516	7699749,379	679,785	-4,042	-0,693	0,215	4,101
13	717920,689	7700705,540	762,263	717925,152	7700705,535	764,295	-4,463	0,005	-2,032	4,463
14	720833,830	7700977,625	672,306	720838,319	7700977,471	672,151	-4,489	0,154	0,155	4,492
15	721912,881	7702479,481	659,000	721917,197	7702480,261	657,787	-4,316	-0,780	1,213	4,386
16	721615,420	7702945,203	655,000	721619,822	7702945,627	655,296	-4,402	-0,424	-0,296	4,422
17	722416,493	7702335,119	665,000	722421,272	7702336,095	663,006	-4,779	-0,976	1,994	4,878
18	722091,007	7701329,966	764,832	722094,280	7701330,719	765,896	-3,273	-0,753	-1,064	3,359
19	723139,474	7701700,544	677,304	723142,670	7701702,809	680,467	-3,196	-2,265	-3,163	3,917
20	717129,063	7702434,541	778,885	717137,426	7702434,555	778,648	-8,363	-0,014	0,237	8,363
21	718497,074	7703104,111	672,893	718503,096	7703104,544	672,474	-6,022	-0,433	0,419	6,038
22	719308,991	7704414,975	659,083	719314,285	7704415,607	659,387	-5,294	-0,632	-0,304	5,332
23	718283,129	7704567,534	659,371	718288,705	7704567,145	659,315	-5,576	0,389	0,056	5,590
24	719933,486	7705668,523	716,975	719938,822	7705669,764	717,416	-5,336	-1,241	-0,441	5,478
25	719257,674	7705665,561	625,000	719262,612	7705666,171	626,538	-4,938	-0,610	-1,538	4,976

26	720865,528	7703210,505	658,208	720870,633	7703212,029	655,815	-5,105	-1,524	2,393	5,328
27	720784,582	7703077,976	660,000	720789,533	7703078,854	656,955	-4,951	-0,878	3,045	5,028
28	722063,893	7704297,171	670,281	722068,179	7704297,595	669,948	-4,286	-0,424	0,333	4,307
29	721759,974	7705684,038	663,617	721764,180	7705682,083	662,557	-4,206	1,955	1,060	4,638
30	719443,537	7701967,403	745,909	719449,067	7701966,134	748,272	-5,530	1,269	-2,363	5,674
31	721962,474	7697599,839	751,465	721967,783	7697601,571	750,419	-5,309	-1,732	1,046	5,584
32	719779,419	7700680,752	680,000	719784,691	7700681,127	679,883	-5,272	-0,375	0,117	5,285
33	721302,106	7703321,014	655,000	721308,136	7703322,207	650,301	-6,030	-1,193	4,699	6,147