



UNIVERSIDADE FEDERAL DA BAHIA
ESCOLA POLITÉCNICA
Programa de Pós-Graduação em Engenharia Civil



WENDELL SANTANA FAGUNDES

**AVALIAÇÃO DA QUALIDADE DE DADOS ALTIMÉTRICOS “ALOS WORLD 3D –
30 m”**

SALVADOR

2023

WENDELL SANTANA FAGUNDES

**AVALIAÇÃO DA QUALIDADE DE DADOS ALTIMÉTRICOS “*ALOS WORLD 3D –
30 m*”**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Civil da Escola Politécnica da Universidade Federal da Bahia, como requisito para obtenção do título de Mestre em Engenharia Civil.

ORIENTADOR: Prof. Dr. Mauro J. Alixandrini Jr

SALVADOR

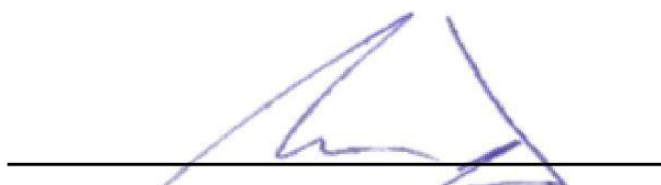
2023

MEMBROS DA BANCA EXAMINADORA DA DISSERTAÇÃO DE MESTRADO

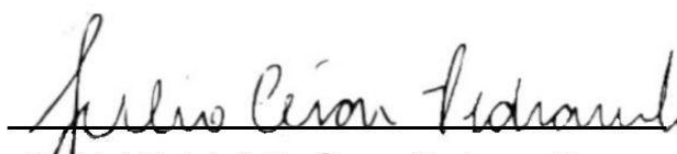
_____ **Wendell Santanta Fagundes** _____

APRESENTADA AO MESTRADO EM ENGENHARIA CIVIL, DA UNIVERSIDADE FEDERAL DA BAHIA, EM 10 DE JULHO DE 2023.


BANCA EXAMINADORA



Prof.(a) Dr.(a) **Mauro José Alixandrini Junior**
Orientador
PPEC-UFBA



Prof.(a) Dr.(a) **Julio Cesar Pedrassoli**
PPEC- UFBA



Prof.(a) Dr.(a) **Marcio Augusto Reolon Schmidt**
PPGCG-UFU

“Toma cuidado com o homem de um só livro”

Santo Tomás de Aquino

DEDICATÓRIA: Aos meus pais, Paulo e Adélia, meus maiores apoiadores e exemplos de vida.

AGRADECIMENTOS

Ao Deus Uno e Trino, Onipotente e Eterno, O Senhor da minha vida.

À Maria Santíssima, Mãe do meu Salvador e minha Mãe, que nunca me desampara.

À todos meus familiares: tios, tias, afilhadas, primos, primas, avós, avôs, pai, mãe e irmã, pelo apoio de sempre.

À minha noiva Sabrina Oliveira, pelo companheirismo, apoio e compreensão.

Ao meu orientador Mauro José Alixandrini Junior pelo apoio, paciência e conhecimentos passados.

Ao amigo e colega de trabalho Thiago Carneiro Neto pelos incentivos, *feedbacks*, conselhos e conhecimentos em estatística passados.

Ao amigo e colega de trabalho Raomax Moreira pelo apoio e por intermediar a obtenção de alguns dados.

Aos amigos e colegas de trabalho Carlos Henrique Cavalcante e Luciano Guimarães pela compreensão, apoio e *feedbacks*.

A todos amigos e amigas que sempre apoiaram, torceram e oraram pelas minhas conquistas.

RESUMO

Modelos Digitais de Elevação (MDE) a nível global são fontes de dados para diversas disciplinas do conhecimento, implicando numa crescente busca pelos melhores modelos globais em termos de precisão. Assim, o presente estudo teve como objetivo central avaliar a qualidade altimétrica e propor correções a nível nacional para o MDE global Alos World 3D – 30m (AW3D30), avaliando também correlações dos erros verticais do AW3D30 com classes de uso e cobertura do solo, oriundas do projeto MapBiomias, e aplicando as correções com base nestas correlações. Por fim, validou o método de correção com enquadramento do PEC-PCD nos MDE original e corrigido. Os materiais deste estudo consistiram em dados altimétricos que representam a verdade de campo, o MDE AW3D30 e Raster de uso e cobertura do solo proveniente da Plataforma MapBiomias. A metodologia foi estruturada iniciando na aquisição, tratamento e normalização dos dados supramencionados, em seguida foram estimados erros no MDE AW3D30 frente aos dados da verdade de campo, em seguida estudos estatísticos avaliaram a consistência dos dados e a qualidade altimétrica do AW3D30 por classe do MapBiomias. Na sequência, o trabalho estabeleceu parâmetros de correção altimétrica utilizando uma modelagem empírica, esse modelo foi aplicado e validado em diferentes regiões. Diante da metodologia exposta, este estudo conseguiu construir modelos AW3D30 corrigidos para três áreas de validação, nos estados de Pernambuco, São Paulo e Bahia. Os resultados apontaram significativa melhora da acurácia altimétrica, medida pelo RMSE, cujos valores saíram de 8,22m, 4,06m e 1,87m antes da correção para 4,55m, 2,48m e 1,26m após a correção, respectivamente. Os resultados encorajam a aplicação da metodologia corretiva proposta neste estudo em todo território nacional, visto que reduziu o RMSE do AW3D30 em 38,70%, na média, para as áreas de validação.

Palavras-chave: Modelos digitais de elevação globais; MDE; MDT; AW3D30; Alos; acurácia de MDE; correções de MDE; PE3D; GeoSampa; MapBiomias; SRTM; ASTER GDEM; TOPODATA.

ABSTRACT

Digital Elevation Models (DEM) at a global level are sources of data for several disciplines of knowledge, implying a growing search for the best global models in terms of accuracy. Thus, the present study had as its main objective to evaluate the altimetric quality and propose corrections at a national level for the global MDE Alos World 3D – 30m (AW3D30), also evaluating correlations of the vertical errors of the AW3D30 with land use and land cover classes, derived from of the MapBiomias project, and applying corrections based on these correlations. Finally, it validated the correction method by framing the PEC-PCD in the original and corrected MDE. The materials of this study consisted of altimetric data that represent the field truth, the MDE AW3D30 and Raster of land use and land cover from the MapBiomias Platform. The methodology was structured starting with the acquisition, treatment and normalization of the aforementioned data, then errors were estimated in the MDE AW3D30 against field truth data, then statistical studies evaluated the consistency of the data and the altimetric quality of the AW3D30 by MapBiomias class . In the sequence, the work established parameters of altimetric correction using an empirical modeling, this model was applied and validated in different regions. Given the exposed methodology, this study was able to build corrected AW3D30 models for three validation areas, in the states of Pernambuco, São Paulo and Bahia. The results showed a significant improvement in altimetric accuracy, measured by the RMSE, whose values increased from 8.22m, 4.06m and 1.87m before correction to 4.55m, 2.48m and 1.26m after correction, respectively. The results encourage the application of the corrective methodology proposed in this study throughout the national territory, since it reduced the RMSE of the AW3D30 by 38.70%, on average, for the validation areas.

Keywords: Global digital elevation models; DEM; DTM; AW3D30; Alos; DEM accuracy; corrections in DEM; PE3D; GeoSampa; MapBiomias; SRTM; ASTER GDEM; TOPODATA.

LISTA DE FIGURAS

		Pág.
Figura 01	Curva da distribuição normal (gaussiana) de probabilidade	23
Figura 02	Demonstrativo de gráfico <i>box plot</i>	26
Figura 03	Geometria de observação do sensor PRISM em modo triplo (OB1) e em modo estéreo (OB2)	33
Figura 04	Mapa de localização das áreas de estudo e validação	36
Figura 05	Fluxograma da metodologia empregada neste trabalho	38
Figura 06	Demonstrativo da lógica aplicada na modelagem de correção empírica.	48
Figura 07	Gráfico <i>Box Plot</i> dos erros - Brasil	53
Figura 08	Gráfico <i>Box Plot</i> dos erros - PE-01	54
Figura 09	Gráfico <i>Box Plot</i> dos erros - PE-02	54
Figura 10	Gráfico <i>Box Plot</i> dos erros - SP-01	55
Figura 11	Gráfico <i>Box Plot</i> dos erros - SP-02	55
Figura 12	Gráfico de densidade de probabilidade dos erros para a região de estudo “Brasil”	61
Figura 13	Gráfico de densidade de probabilidade dos erros para a região de estudo “PE-01”	62
Figura 14	Gráfico de densidade de probabilidade dos erros para a região de estudo “PE-02”	63
Figura 15	Gráfico de densidade de probabilidade dos erros para a região de estudo “Brasil”	64
Figura 16	Gráfico de densidade de probabilidade dos erros para a região de estudo “Brasil”	65
Figura 17	Gráficos para estabelecer amostragem mínima na estimativa de erro médio (a) e RMSE (b). Dados da classe “Mosaico de agricultura e pastagem” (21).	73
Figura 18	Gráficos para estabelecer amostragem mínima na estimativa de erro médio (a) e RMSE (b). Dados da classe “Pastagem” (15)	73
Figura 19	Gráficos para estabelecer amostragem mínima na estimativa de erro médio (a) e RMSE (b). Dados da classe “Formação Florestal” (3)	74
Figura 20	Fluxograma da metodologia de correção empírica proposta neste trabalho	75
Figura 21	Fórmula condicional usada na correção do AW3D30.	77
Figura 22	Código usado na calculadora raster do QGIS para calcular a superfície de correção do AW3D30.	78
Figura 23	Mapa hipsométrico elaborado com o AW3D30 corrigido para área de validação no Estado de Pernambuco.	79

Figura 24	Perfil topográfico para o alinhamento A-B do mapa hipsométrico da Figura 23	79
Figura 25	Mapa hipsométrico elaborado com o AW3D30 corrigido para área de validação no Estado de São Paulo.	80
Figura 26	Perfil topográfico para o alinhamento A-B do mapa hipsométrico da Figura 25	80
Figura 27	Mapa hipsométrico elaborado com o AW3D30 corrigido para área de validação no Estado da Bahia	81
Figura 28	Perfil topográfico para o alinhamento A-B do mapa hipsométrico da Figura 27	82

LISTA DE TABELAS

		Pág.
Tabela 01	Resumo de alguns valores de RMSE dos MDE SRTM (30m), ASTER GDEM, AW3D30 e TOPODATA, encontrados na literatura nos últimos 10 anos	29
Tabela 02	Erro médio e RMSE por região amostral, calculados com <i>outliers</i> .	52
Tabela 03	Erro médio e RMSE por região amostral, calculados sem <i>outliers</i> .	52
Tabela 04	Estatísticas por classe acerca dos <i>outliers</i> da região “Brasil”	57
Tabela 05	Estatísticas por classe acerca dos <i>outliers</i> da região “PE-01”	58
Tabela 06	Estatísticas por classe acerca dos <i>outliers</i> da região “PE-02”	58
Tabela 07	Estatísticas por classe acerca dos <i>outliers</i> da região “SP-01”	59
Tabela 08	Estatísticas por classe acerca dos <i>outliers</i> da região “SP-02”	59
Tabela 09	<i>Outliers</i> médios por classe de uso e cobertura do solo	60
Tabela 10	Resultados do teste de Tukey para média dos erros por região amostral.	66
Tabela 11	Resultados do teste de Tukey para média dos erros quadráticos (RMSE) por região amostral.	67
Tabela 12	Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral	68
Tabela 13	Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral “PE-01”	68
Tabela 14	Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral “PE-02”	69
Tabela 15	Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral “SP-01”	70
Tabela 16	Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral “SP-02”	70
Tabela 17	Erro médio por classe do MapBiomas nas áreas de estudo.	72
Tabela 18	Quantidade de unidades amostrais por área de estudo.	72
Tabela 19	Coefficientes de correção altimétrica para o MDE AW3D30 em função das classes do MapBiomas.	76
Tabela 20	Avaliação da qualidade altimétrica do MDE AW3D30 antes e depois da correção empírica proposta neste estudo.	82

SIGLAS E ABREVIATURAS

MDE - Modelo Digital de Elevação

MDT - Modelo Digital de Terreno

MDS - Modelo Digital de Superfície

GNSS - *Global Navigation Satellite Systems*

SR - Sensoriamento Remoto

SRTM - *Shuttle Radar Topography Mission*

ASTER GDEM - Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model

JAXA - *Japan Aerospace Exploration Agency*

ALOS - *Advanced Land Observing Satellite*

AW3D30 - *Alos World 3D - 30 meters*

PRISM - *Panchromatic Remote-sensing Instrument Stereo Mapping*

RMSE - *Root Mean Square Error*

PEC-PCD - Padrão de Exatidão Cartográfica para Produtos Cartográficos Digitais

PEC - Padrão de Exatidão Cartográfica

DSG - Departamento de Serviços Geográficos do Exército Brasileiro

ANOVA - Análise de Variância

LIDAR - *Light Detecting And Ranging*

ET-ADGV - Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais

IBGE - Instituto Brasileiro de Geografia e Estatística

EGM96 - *Earth Gravitational Models from 1996*

SGB - Sistema Geodésico Brasileiro

RAAP - Rede Altimétrica de Alta Precisão do SGB

PE3D - Plataforma Pernambuco Tridimensional

UTM - *Universal Transverse Mercator*

SIG - Sistemas de Informação Geográfica

SUMÁRIO

1 INTRODUÇÃO	15
1.1 OBJETIVOS	18
1.1.1 Geral	18
1.1.2 Específicos	18
1.2 JUSTIFICATIVA	19
2 REVISÃO DE LITERATURA	20
2.1 MDE	20
2.2 ACURÁCIA DE MDE	21
2.2.1 Estatísticas na averiguação da qualidade altimétrica de MDE	22
2.2.1.1 Parâmetros básicos envolvidos na validação de MDE	22
2.2.1.2 Teste de normalidade dos dados	23
2.2.1.3 Detecção de outliers	25
2.2.1.4 Análise de Variância e Teste de Tukey	26
2.2.1.5. Padrão de Exatidão Cartográfico para Produtos da Cartografia Digital PEC-PCD	27
2.2.2 Acurácia de MDE globais encontrada na Literatura	28
2.3 RETIFICAÇÃO DE MDE	29
2.4 MDE ALOS WORLD 3D - 30M	32
2.5 PROJETO MAPBIOMAS	34
3 METODOLOGIA	35
3.1 ÁREAS DE ESTUDO E VALIDAÇÃO	35
3.2 DADOS E RECURSOS	35
3.3 MÉTODOS	37
3.3.1 Aquisição dos dados	38
3.3.2 Tratamento dos dados	40
3.3.2.1 Tratamento das nuvens de pontos (Grupo 01)	41
3.3.2.2 Tratamento dos pontos da Rede Altimétrica (Grupo 02)	42
3.3.2.3 Processamento dos dados levantados com receptores GNSS	42
3.3.3 Cálculo dos erros	42
3.3.3.1 Cálculo de erros do Grupo 01	43
3.3.3.2 Cálculo de erros do Grupo 02	43
3.3.4 Aplicação do estudo estatístico	43
3.3.4.1 Eliminação de Outliers	43
3.3.4.2 Teste de normalidade do conjunto amostral “erros”	44
3.3.4.3 Comparação de média das regiões amostrais	44
3.3.4.4 Comparação de média das classes do MapBiomias	45
3.3.5 Elaboração de uma metodologia para modelagem da superfície de correção	45

3.3.5.1 Estudo do erro médio por classe do MapBiomias	45
3.3.5.2 Modelo de correção empírico baseado em coeficientes de correção	46
3.3.5.3 Determinação dos coeficientes de correção	46
3.3.5.4 Número mínimo de unidades amostrais como entrada no modelo	47
3.3.5.5 Modelagem de correção	48
3.3.6 Aplicação da metodologia de correção e validação dos resultados	49
4 RESULTADOS E DISCUSSÕES	50
4.1 INDICADORES ESTATÍSTICOS	50
4.1.1 Erro altimétrico médio e Raiz do erro quadrático médio (RMSE)	50
4.1.2 Análise dos outliers por Box plot	53
4.1.2.1 Outliers por classe do Map Biomias	56
4.1.3 Normalidade do conjunto erros	60
4.1.4 Teste de média para averiguar homogeneidade das 5 regiões amostrais	66
4.1.5. Teste de média para averiguar correlação entre o porte de vegetação e os valores de erro	67
4.1.6 Erro médio por classe	71
4.1.7 Número mínimo de unidades amostrais para estimar o erro	72
4.2 MODELAGEM DE CORREÇÃO DAS ALTITUDES	74
4.2.1 Fluxograma da correção	74
4.2.2 Coeficientes de correção	76
4.2.3 Fórmula condicional para conversão raster	77
4.2.4 AW3D30 corrigido	78
4.2.5 Validação dos resultados através do PEC-PCD	82
5 CONSIDERAÇÕES FINAIS	84
REFERÊNCIAS BIBLIOGRÁFICAS	86

1 INTRODUÇÃO

Com o início da popularização do uso do computador, a partir da década de 80 (CORREIA, 2021), as representações analógicas de superfícies geográficas começaram entrar em desuso, ficando com o Modelo Digital de Elevação (MDE) o papel de oferecer a projetistas e pesquisadores a visão topográfica de um determinado espaço, para as diversas finalidades deste modelo.

Dentre as várias aplicações de MDE se incluem hidrologia e recursos hídricos, geologia, geomorfologia, projetos de engenharia civil, levantamento de vegetação, glaciologia, vulcanologia, modelagem de riscos naturais (MOORE, GRAYSON e LADSON, 1991), ortorretificação de imagens orbitais, aplicações em projetos de irrigação, entre outros. Independentemente da aplicação, quanto maior a exatidão de um MDE, mais acurados serão os resultados nas aplicações, otimizando a tomada de decisão.

A acurácia de um MDE está intimamente ligada à sua construção, a qual, por sua vez, pode advir de distintas técnicas, das quais podem ser citadas a interpolação de dados da topografia clássica ou do levantamento por GNSS, interferometria por radar, estereoscopia orbital ou suborbital, perfilamento a *laser*, entre outros. Independente da técnica, a inerência de erros na geração de um MDE é algo factual e constantemente discutido, principalmente no meio científico, buscando-se sempre mensurar a magnitude dos erros bem como soluções para dirimi-los, uma das pautas deste trabalho.

O cometimento de erros na geração de MDE é inevitável, porque os modelos são visões simplificadas da realidade, submetidas a um processo de generalização, onde os parâmetros modelados são imprecisos (FELICÍSSIMO, 2001). No entanto, este fato não pode implicar em resignação total à existência de imprecisões, fazendo-se necessário buscar a diminuição destes erros por meio de inovações nas técnicas de levantamento e processamento de dados, resultando na geração de modelos cada vez mais acurados. Não é a toa que muitos estudos acadêmicos buscam a retificação de MDE globais obtidos por métodos de sensoriamento remoto, destacando-se aqui alguns trabalhos recentes como o de O'loughlin (2016), Gallant e Read (2016), Girão et al (2019), Magruder et al (2021) e Hawker et al (2022).

A modelagem de terreno com base em dados do sensoriamento remoto (SR) é uma importante técnica capaz de construir modelos de elevação a nível global, fornecendo à comunidade internacional informações topográficas em larga escala. Nas análises voltadas à superfície terrestre, este tipo de dado possibilita estudos prévios que não demandem altas

precisões, ao passo em que ajudam a cortar custos, dado a gratuidade da grande maioria dos MDE globais. Quando o estudo exige maior nível de detalhes e acurácia do MDE, o refinamento desses modelos globais podem ser suficientes para algumas análises espaciais, como, por exemplo, delimitação de sub-bacias hidrográficas, projeto básico de irrigação, modelagens geomorfológicas, de riscos naturais, entre outros.

Existem vários modelos digitais de elevação gratuitos com cobertura quase global advindos do SR, tais como o SRTM (*Shuttle Radar Topography Mission*) oriundo de interferometria por radar cuja versão mais recente possui resolução espacial de 30 metros; o ASTER GDEM também com tamanho de pixel de 30 metros, porém sua geração foi baseada na técnica de estereoscopia orbital; e mais recentemente foi disponibilizado pela JAXA (*Japan Aerospace Exploration Agency*) o MDE Alos World 3D (AW3D30), com uma resolução espacial também de 30 metros, tem extração de altimetria baseada em estereoscopia orbital. Independentemente da fonte, os MDE globais têm sido ferramentas indispensáveis em diversas áreas do conhecimento, sendo aplicados em inúmeros estudos acadêmicos e técnicos. Landau e Guimarães (2011) comentam que os modelos digitais de elevação construídos por métodos como a estereoscopia orbital e interferometria por radar se apresentam como uma eficiente alternativa para constituir bases topográficas padronizadas, cobrindo áreas relativamente extensas em menor tempo.

O MDE AW3D30 foi gerado a partir de conjuntos estereoscópicos coletados pelo radiômetro pancromático PRISM (*Panchromatic Remote-sensing Instrument Stereo Mapping*), a bordo do satélite japonês ALOS (*Advanced Land Observing Satellite*), o qual permaneceu em órbita entre janeiro de 2006 e maio de 2011. Originalmente, o modelo foi gerado com resolução espacial de 5 metros, por meio de análise estereoscópica de imagens ortoretificadas com 2,5 metros de tamanho de pixel. Porém, a JAXA disponibiliza de modo gratuito uma versão reamostrada do modelo, com resolução espacial de 30 metros (JAXA, 2021). Uma observação acerca do modelo AW3D30 foi realizada nos estudos de Santillan e Makinano-Santillan (2016), os quais apontaram superioridade dos dados AW3D30 sobre o SRTM de 30 metros, em termos de precisão vertical, em função dos menores erros médios e RMSE para sua área de estudo.

Gallant & Read (2016) chamam a atenção para o fato de que o uso rotineiro de MDE globais para fins analíticos como extração de rede hidrológica, delimitação de bacia de inundação, mapeamento do ambiente e do solo, podem ser seriamente afetados pela presença de artefatos nos dados, principalmente os deslocamentos devido à cobertura de árvores e ao

ruído aleatório. Embora esses erros possam ser comumente atribuídos à vegetação, sua correção requer conhecimento sobre as alturas do dossel e profundidades de penetração do radar (O'LOUGHLIN et al., 2016). Nessa mesma linha de raciocínio, Paiva et al (2013) conclui em seu trabalho de validação de modelagem hidrológica e hidrodinâmica para a bacia Amazônica que alternativas para diminuir erros nos modelos podem ser o uso de novas técnicas de processamento dos dados de sensoriamento remoto, onde se corrija o MDE removendo altura da vegetação em áreas florestadas e assim estimar a altimetria do solo das várzeas.

Morais (2017) observaram que nos trabalhos de avaliação altimétrica de MDE globais, as áreas de estudo escolhidas para avaliação apresentam características homogêneas, sendo comumente evitadas regiões com vegetação diversificada ou com presença de edificações, já que essas feições podem interferir na precisão altimétrica dos modelos. Este fato reforça ainda mais a necessidade de um estudo que busque averiguar a magnitude dos erros altimétricos em regiões com forte presença vegetativa, pois deste modo, o usuário final poderá ter informações da qualidade dos dados altimétricos em função da classe de cobertura de solo proeminente.

Conforme os estudos de Silva (2020), constatou-se que a metodologia mais comum para a avaliação de acurácia de MDE é comparar os dados de uma amostra de pontos do MDE a ser analisado, ditos pontos homólogos, com as coordenadas tidas como verdade de campo, calculando-se, entre outros indicadores, a raiz quadrada do erro médio quadrático (RMSE) da elevação. Entretanto, de acordo observações de Wise (2000) e Chagas et al (2010), dois problemas se vinculam a esta metodologia tradicional: o primeiro se relaciona ao fato de que costumeiramente poucos pontos amostrais são aplicados (20 a 30 amostras), e o segundo, e talvez mais importante, é que a referida abordagem metodológica não investiga as fontes de erro. Frente a este fato, neste trabalho se propôs analisar o erro pixel a pixel do AW3D30 nas regiões de estudo, eliminando-se apenas aqueles que forem considerados *Outliers*. Além disso, um dos principais objetivos deste trabalho é investigar uma possível influência de diferentes classes de uso e cobertura do solo na qualidade final do modelo.

Além das variáveis estatísticas utilizadas para estudar acurácia de modelos de terreno, como a raiz quadrada do erro médio quadrático (RMSE) da elevação, erro médio (EM) e desvio padrão (DP), no Brasil o normativo cartográfico prevê o Padrão de Exatidão Cartográfica para Produtos Cartográficos Digitais (PEC-PCD) como forma de validar a confiabilidade vertical e horizontal da produção cartográfica no país. Estabelecido pela norma

“Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV)”, o PEC-PCD é comumente aplicada em estudos acadêmicos que avaliam a acurácia de MDE, citando-se, a seguir, alguns destes estudos:

- o trabalho de Machado et al (2020) avaliou o PEC-PCD do MDE SRTM para a escala de 1:50.000;
- a pesquisa de Barbosa et al (2019) comparou a acurácia dos MDE ALOS PALSAR e ALOS PRISM, por meio do PEC-PCD, nas escalas 1:25.000, 1:50.000 e 1:100.000;
- Moraes et al (2017) abordaram a avaliação altimétrica dos Modelos Digitais de Elevação ASTER GDEM v2 e SRTM v3, tendo como referência dados *Light Detection And Ranging* (LIDAR) e metodologia baseada em métodos estatísticos e PEC-PCD.
- Iorio et al (2012) compararam a acurácia altimétrica de um modelo digital de elevação baseado em estereoscopia de imagens ALOS/PRISM com os MDE gratuitos SRTM, ASTER GDEM e TOPODATA, tendo como metodologia de referência o enquadramento no PEC.

Isto posto, pode-se pressupor a frequente utilização do PEC-PCD como forma de validar a acurácia de modelos digitais de elevação, sendo esse um indicador de dispersão relativo a 90% de probabilidade, que define a acurácia (exatidão) do trabalho cartográfico (DSG, 2016b).

1.1 OBJETIVOS

Perante o que foi explicitado anteriormente, este trabalho tem os seguintes objetivos:

1.1.1 Geral

Avaliar a acurácia altimétrica e propor um método de correção para o modelo digital de elevação AW3D30, aplicável no Brasil.

1.1.2 Específicos

A) Verificar se as amostras tomadas em diferentes regiões do país possuem acurácias que podem ser consideradas estatisticamente similares;

B) Avaliar a magnitude de erros verticais do MDE AW3D30 em função do uso e cobertura do solo usando o projeto MapBiomass, e verificar se há alguma correlação entre estas duas variáveis;

C) Validar o processo e verificar se a correção foi efetivamente impactante na acurácia do MDE AW3D30 corrigido.

1.2 JUSTIFICATIVA

O presente trabalho, dentro do contexto e objetivos apresentados, mostra sua relevância não só em investigar a relação entre classes de uso e cobertura do solo e os erros inerentes ao MDE AW3D30, mas também em averiguar a possibilidade de uma modelagem corretiva aplicada à MDE, o que, conseqüentemente, poderá melhorar as análises espaciais que utilizam estes modelos como base de dados. Deste modo, um modelo corrigido e com padrão de precisão conhecida possibilitará análises espaciais mais confiáveis, para as diversas aplicações mencionadas anteriormente.

Este estudo, por meio de investigações estatísticas e a proposição de uma metodologia simplificada para corrigir erros em MDE, após ter sua validação, poderá ser aplicado em território brasileiro para o AW3D30. Isso representa um avanço na aplicação de MDE globais no Brasil, visto que, costumeiramente, estes modelos são aplicados diretamente da forma que vêm da fonte original. Isso não tira o respaldo que têm estes modelos em sua versão original, mas aprimorar tais modelos de acordo ao contexto regional traz ganhos de qualidade, tanto nas representações, como nas aplicações.

Além disso, todo arcabouço de resultados e revisão bibliográfica que embasam este estudo serão fonte de informação técnico-científica sobre a qualidade altimétrica de MDE e sua correlação com o uso e cobertura do solo, apoiando futuros trabalhos dentro desta temática.

2 REVISÃO DE LITERATURA

O presente capítulo se reserva à exposição do arcabouço de conhecimentos encontrados na bibliografia pesquisada, acerca da temática proposta neste estudo, o qual foi imprescindível para o alcance dos objetivos.

2.1 MDE

Segundo Felicísimo (1994), o modelo digital de elevação (MDE) é definido como uma estrutura de dados numéricos para representar a distribuição espacial da altitude na superfície terrestre. De um modo geral, a unidade básica de informação em um MDE é o valor de altitude z , acompanhado pelos valores correspondentes de x e y , expressos dentro de um sistema de projeção geográfica para um referenciamento espacial preciso (FELICÍSIMO, 1994). No conceito de Santos et al (2014) o MDE é explicado como “uma representação digital composta por uma matriz de *pixels* com coordenadas planimétricas (x,y) e com um valor de intensidade do *pixel* que corresponde à elevação”.

Atualmente o conceito de MDE se divide em dois termos, o modelo digital de terreno (MDT) e o modelo digital de superfície (MDS). Assim, partindo do entendimento do MDE como uma matriz georreferenciada cuja intensidade de pixel representa as altitudes, se desdobram dois outros conceitos, conforme Silva (2020), como segue: quando se representa os acidentes naturais e artificiais sobre o solo, estamos falando de um Modelo Digital de Superfície (MDS). Se, por outro lado, na representação são desconsiderados esses acidentes, ou seja, considerando apenas o solo exposto, trata-se de um Modelo Digital do Terreno (MDT). Esta noção conceitual também está expressa pelo DSG (Diretoria de Serviço Geográfico do Exército Brasileiro) em sua norma EB80-N-72.001 de 2016, onde o MDS é entendido como a representação do solo exposto e dos acidentes encontrados acima do solo, como edificações, vegetação, dentre outros, de forma contínua e suavizada, a partir de dados adequadamente estruturados e amostrados do mundo real; Já, o MDT trata-se da representação do solo exposto de forma contínua e suavizada, a partir de dados adequadamente estruturados e amostrados da superfície física da Terra (DSG, 2016a).

Diante de tudo que foi exposto, foi determinado que neste trabalho se utilizará o termo MDE fazendo referência às matrizes digitais que têm a altitude como variável “ z ”, vinculada a uma localização geográfica (x,y), sendo essa altitude relativa ao solo ou a qualquer outra feição que se incorpore ao mesmo (como vegetação, edificações, etc).

2.2 ACURÁCIA DE MDE

O modelo digital de elevação é um modelo de informação para a geração de outros modelos, total ou parcialmente dependentes da topografia. Por este motivo, a utilidade e validade dos resultados derivados estão estritamente relacionados com a acurácia do modelo base (FELICÍSIMO, 2001). Como já abordado anteriormente, a inerência de erros na formulação de modelos digitais de elevação é um fato inevitável, entretanto, avaliar tendências de viés, buscar melhora na similaridade com o espaço representado e determinar um padrão de acurácia são passos imprescindíveis para obtenção de melhores resultados, diante de projetos que usam modelos de terreno. A literatura relativa ao tema tem abordado essas questões, de forma a investigar e buscar novas soluções, como será mostrado nos parágrafos seguintes.

Adaptando-se a categorização dos erros na modelagem de altitude feita por Veregin (1989), é possível afirmar que os erros em modelos de elevação se dividem em dois tipos: erros planimétricos e erros altimétricos. O primeiro tipo de erro, o planimétrico, implica em imprecisões na localização geográfica da altitude, seja na trajetória das curvas de nível, na planimetria de pontos ou de pixels de um MDE, portanto, afetam a situação no plano horizontal (plano XY); erros altimétricos envolvem inexatidão da própria altitude associada à curva de nível, ponto ou número digital do pixel do MDE, vinculando-se, assim, à dimensão vertical (eixo Z).

Em termos de MDE produzidos com dados de SR, os erros planimétricos podem estar relacionados ao georreferenciamento das imagens espectrais ou de radar, usadas na construção do MDE. Quanto aos erros altimétricos, o foco deste estudo, estes podem estar relacionados a diversos fatores. Segundo Magruder et al (2021), eles são provenientes da complexidade da região ou da dinâmica da topografia, em outros casos são inerentes às capacidades do sensor ou às técnicas de processamento. Na perspectiva de autores como Paiva et al (2013), O'loughlin et al (2016) e Gallant & Read (2016), os erros altimétricos em MDE podem ser atribuídos principalmente à presença de vegetação, ou ainda ao ruído aleatório.

O autor Silva (2020) constatou que a metodologia mais comum, encontrada na literatura recente, para avaliação de acurácia de MDE, tem como foco a comparação dos dados altimétricos extraídos do MDE a ser analisado com as altimetrias tidas como verdadeiras, originárias de uma fonte mais acurada. Na prática, isto é feito com um conjunto amostral de pontos, ditos pontos homólogos, em cuja localização geográfica se extrai a elevação do MDE avaliando bem como a elevação de referência. Tal comparação se

concentra, portanto, no cálculo de média e desvio padrão do conjunto de valores das diferenças entre as altitudes do MDE e as de referência, para os pontos homólogos.

O grau de acurácia dos MDE pode ser então definido tanto pelo erro altimétrico médio, ou seja, a média das diferenças entre as elevações estimadas pelo modelo e as obtidas em campo, bem como pela raiz quadrada do erro médio quadrático (RMSE) da elevação. Entretanto, dois problemas estão comumente vinculados a esta metodologia de avaliação de erros em MDE's: o primeiro se relaciona ao fato de que costumeiramente poucos pontos amostrais são aplicados (20 a 30 amostras), e o segundo, e talvez mais importante problema, é que a referida abordagem metodológica não investiga as fontes de erro; apesar disso, esta metodologia é a mais utilizada para avaliar a qualidade de MDE gerados a partir de distintas fontes e técnicas (WISE, 2000; CHAGAS et al, 2010).

2.2.1 Estatísticas na averiguação da qualidade altimétrica de MDE

Os fundamentos de estatística são, sem dúvidas, unânimes em aplicações nos mais diversos estudos científicos, e para o presente caso não seria diferente. Neste tópico serão abordados alguns dos principais conceitos e técnicas estatísticas aplicadas nos estudos que envolvem a quantificação de erros em um modelo digital de elevação. As estatísticas são imprescindíveis neste estudo pois, a partir delas, é possível determinar um indicador da qualidade geral do modelo, a distribuição dos erros no modelo, a identificação de possíveis tendências, bem como de possíveis correlações com outras variáveis.

2.2.1.1 Parâmetros básicos envolvidos na validação de MDE

O primeiro parâmetro na avaliação da qualidade de um modelo digital de elevação é o erro altimétrico, cujo cálculo está na simples diferença entre a altitude do modelo e a altitude tomada como verdadeira, para cada ponto homólogo. Esta variável possui valores positivos e negativos, aspecto o qual permite a inferência de diversas características sobre acurácia do modelo avaliando, como por exemplo: muitos erros positivos concentrados numa determinada região do modelo podem indicar a presença de artefatos, como uma cobertura vegetal densa; erros negativos podem indicar inconsistências do tipo *sinks* (buracos) no modelos.

Conforme Zanetti et al (2018), a análise de acurácia de um produto cartográfico é realizada por meio de estimadores estatísticos para um certo nível de confiança, consoante com o normativo recomendado para cada país. Levando-se em conta a definição de acurácia revista por Monico et al.(2009), a qual envolve tanto efeitos aleatórios (precisão) quanto

efeitos sistemáticos (tendência), o principal indicador de acurácia seria o RMSE (*root mean square error*) das discrepâncias das coordenadas (Equação 01), traduzido para nossa língua vernácula como raiz do erro médio quadrático.

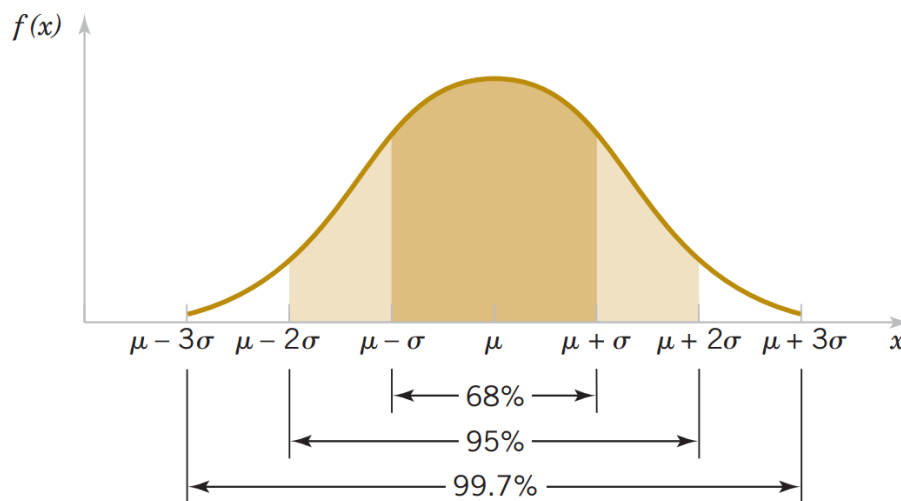
$$RMSE = \sqrt{\sum_{i=1}^n \frac{e_i^2}{n}} \quad \text{Eq. 01}$$

Onde e^2 é o erro quadrático e n é o número de unidades amostrais.

2.2.1.2 Teste de normalidade dos dados

A avaliação sobre a normalidade da distribuição de um conjunto de dados é essencial para uma descrição adequada da amostra e de sua análise inferencial. A distribuição das probabilidades em “forma de sino”, conhecida como curva normal ou de Gauss (Figura 01), possui propriedades que a torna especial para a estatística, principalmente sua simetria, coincidência entre moda, média e mediana, a da possibilidade de ser representada e quantificada com base nos valores de média e desvio padrão (MIOT, 2017).

Figura 01: Curva da distribuição normal (gaussiana) de probabilidade.



Fonte: UFSC (2022)

Neste trabalho de avaliação de erros altimétricos do AW3D30, testar a normalidade do conjunto de erros é passo imprescindível visto que a normalidade de tais dados é pressuposto para duas etapas subsequentes: o estudo do PEC-PCD e a análise de variância (ANOVA) com Teste de Tukey para comparação de médias. Sobre o PEC-PCD, Carvalho e Silva (2018) destacam que ambos são calculados como sendo 1,6449 vezes o valor do erro padrão, relativo a probabilidade de 90% dentro da curva da distribuição normal. Portanto, é legítimo aceitar que o PEC-PCD é somente válido se a variável erro tiver distribuição

gaussiana, conforme os autores. Quanto à ANOVA, é um método estatístico já consolidado no meio acadêmico, usado para informar se existe um tratamento discrepante dentre os demais. Porém, tal método exige o atendimento de algumas premissas, tais como: distribuição normal dos resíduos e homogeneidade de variância (CEGATTA, 2016).

Para se iniciar a avaliação da normalidade de um conjunto de dados deve ser plotado seu histograma e gráfico de densidade de probabilidade, com intuito de identificar assimetrias excessivas, picos multimodais e descontinuidades de dados. Para uma testagem mais rigorosa há uma dezena de testes estatísticos que apuram o ajuste dos dados à distribuição gaussiana, a partir de diferentes pressupostos e algoritmos. Estes testes presumem a hipótese de normalidade dos dados (H_0), retornando um *p-valor* maior que 0,05 se resultarem na adesão aos parâmetros de normalidade. Diversas simulações apontam desempenho superior para os testes de Shapiro-Wilk e Shapiro-Francia (MIOT, 2017).

De acordo com os estudos bibliográficos de Silva (2020), o teste de Shapiro-Wilk é um dos mais eficientes para testar a aderência de um conjunto amostral à normalidade, sobretudo em amostras com poucas observações (até cinquenta). O teste de Shapiro-Wilk consiste num teste de hipóteses, onde a hipótese nula é aceita quando o valor “W”, obtido através da Equação 02, é maior que o valor crítico “ $W\alpha$ ” tabelado, e ainda que o *p-value* seja maior do que o nível de significância (0,05), levando-se em consideração o número de observações e o nível de confiança do teste. A aceitação da hipótese nula significa que os dados são provenientes de uma distribuição Gaussiana, em caso de rejeição da hipótese nula a normalidade é descartada (SILVA, 2020).

$$W = \frac{\sum_{i=1}^n a_i y_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{Eq. 02}$$

onde “ a_i ” são constantes calculadas por meio da Equação 03, usando-se médias, variâncias e covariâncias da estatística de ordem de uma amostra de tamanho “n” de uma distribuição normal.

$$(a_1, a_2, \dots, a_n) = \frac{m^t V^{-1}}{(m^t V^{-1} V^{-1} m)^{1/2}} \quad \text{Eq. 03}$$

sendo $V = [\text{cov}(y_i, y_j)]$ a matriz de covariância daquela estatística de ordem; e $m = (m_1, m_2, \dots, m_n)$ o vetor dos valores esperados da estatística de ordem de amostras aleatórias independentes e identicamente distribuídas (SILVA, 2020).

2.2.1.3 Detecção de outliers

Segundo Zanetti et al (2018), apesar da maioria das normas de controle de qualidade posicional em produtos cartográficos (inclusive a brasileira) não preverem testes para detecção de *outliers*, esta análise é importante, pois a utilização de um dado tendencioso impossibilita a integração entre dados, podendo gerar análises inconsistentes e influenciar negativamente os processos decisórios, findando na geração de produtos que não coincidem com a verdade de campo. Zandbergen (2008) aponta que as discrepâncias (erros) posicionais em pontos levantados por *Lidar* terão distribuição que se aproxima da normal apenas quando os valores *outliers* forem excluídos dos dados. Diante do exposto, percebe-se a importância de aplicar um tratamento aos *outliers*, como forma de acurar a avaliação de qualidade do produto em teste, bem como normalizar a distribuição de erros, esta última sendo premissa para outras análises, como já mencionado anteriormente.

Com base num estudo conceitual em diversas fontes, Figueira (1998) chegou a uma definição de *outliers* sumarizada na seguinte ideia: um *outlier* é uma observação cuja característica precípua se fundamenta na relação de distanciamento para com as demais observações que fazem parte da amostra. Tais observações podem também ser designadas como anormais, contaminantes, estranhas, extremas ou aberrantes.

Os valores *outliers* não são coerentes com o restante da série de dados, podendo ser falhas ou ruídos ocasionados por erros nos equipamentos, e cuja eliminação é desejável numa etapa de pré-processamento. Há diversos métodos de detecção de *outliers* os quais podem ser estatísticos, baseados na distância ou na densidade dos dados. Um método simples é aquele que considera *outlier* o dado que sai do limite estabelecido por um valor mínimo e outro máximo calculados (VENTURA et al. 2013; CHEN et al. 2010).

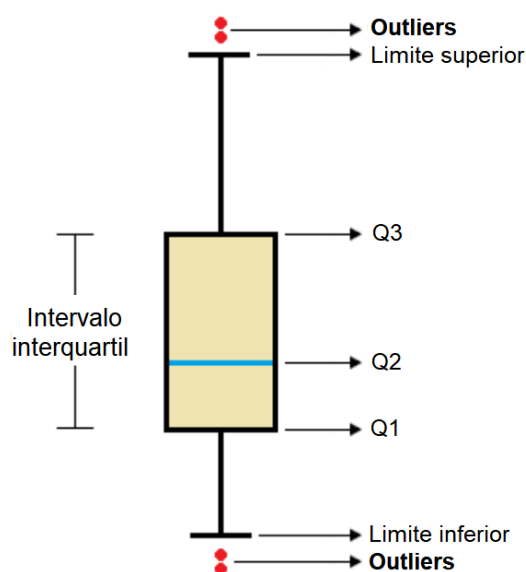
Neste estudo, optou-se por utilizar o teste *box plot* para determinação dos *outliers*, por ser um método simples, de fácil entendimento e amplamente utilizado na literatura. O *box plot* é uma representação gráfica da dispersão dos dados (Figura 02). Este gráfico representa o quartil inferior (Q1) e o quartil superior (Q3) juntamente com a mediana (Q2), sendo que o quartil inferior representa 25% da amostra, a mediana é o percentil 50 dos dados e o quartil superior é o percentil 75. Como pode ser observado na Figura 02, as partes superior e inferior da caixa são definidas por uma distância fixa do intervalo entre os quartis Q1 e Q3 ($Q3 - Q1$), denominado intervalo interquartil. O limite inferior para determinação de *outliers* é calculado a partir da diferença entre o Q1 e o intervalo interquartil multiplicado por 1,5 (Equação 04). Já o limite superior é a soma de Q3 com intervalo interquartil multiplicado por 1,5 (Equação 05).

Qualquer observação fora desses limites é considerada um potencial *outlier*. Mesmo quando os dados não são normalmente distribuídos, um *box plot* pode ser usado por depender da mediana e não da média dos dados (WALFISH, 2006).

$$\text{Limite inferior} = Q_1 - 1,5(Q_3 - Q_1) \quad \text{Eq. 04}$$

$$\text{Limite superior} = Q_3 + 1,5(Q_3 - Q_1) \quad \text{Eq. 05}$$

Figura 02: Demonstrativo de gráfico *box plot*.



Fonte: Adaptado de Silva (2020).

2.2.1.4 Análise de Variância e Teste de Tukey

Segundo Silva (2020) numa avaliação da qualidade posicional de MDE, inicialmente, parâmetros de qualidade são obtidos para cada uma das áreas de estudo separadamente, ou seja, as estatísticas básicas calculadas, como a média, desvio padrão e RMSE são analisadas dentro de cada amostra de pontos de controle, sem levar em consideração uma comparação dos resultados entre as amostras. Com o objetivo de verificar se as mesmas fazem parte de uma mesma população, mesmo sendo amostras coletadas em áreas diferentes, são necessários cálculos estatísticos para comprovar essa premissa. Um método que pode ser utilizado para tirar este tipo de conclusão é conhecido como análise de variância ou ANOVA.

Conforme Riboldi (1995) a análise de variância é uma técnica para verificar diferenças entre dois ou mais tratamentos, utilizada para decompor a variabilidade total nas múltiplas causas de variação. Segundo Gomes (1954) a comparação entre médias pela análise de variância era um velho problema sem solução satisfatória até então. FISHER havia proposto

uma solução baseada em “teste t ”, mas tal solução não era exata, já que trata apenas de um método empírico grosseiro. BRIEGER introduziu um método bastante razoável, porém igualmente baseado no teste t e empírico. Até que em 1952 SCHEFFÉ e TUKEY lançaram as bases de dois métodos, ambos exatos, que são de grande utilidade na análise estatística. Neste trabalho optou-se pela utilização do método de Tukey, o qual será explicado adiante.

Conforme Gomes (1954), o método de Tukey está baseado na amplitude total "estudentizada" ("*studentized range*"), se aplicando preferencialmente em casos de comparação de duas médias, escolhidas entre as k médias de tratamentos, todos com a mesma quantidade de repetições. Suponha-se, por exemplo, que se tem k tratamentos e r repetições, e uma estimativa s para o erro residual, com n graus de liberdade. Assim, o intervalo de confiança para a diferença $\bar{y}_i - \bar{y}_j$ entre duas médias de tratamentos é dada pela Equação 06 a seguir:

$$\bar{y}_i - \bar{y}_j \pm \frac{s}{\sqrt{r}}q \quad \text{Eq. 06}$$

onde q é valor da amplitude total "estudentizada" para o nível α de probabilidade (geralmente 5% ou 1%), para o caso de k tratamentos e n graus de liberdade na estimativa do erro. O valor q é tabelado.

2.2.1.5. Padrão de Exatidão Cartográfico para Produtos da Cartografia Digital PEC-PCD

No Brasil, a avaliação de um produto geoespacial tem sua qualidade estimada através do Padrão de Exatidão Cartográfica (PEC), posteriormente atualizado para Padrão de Exatidão Cartográfica dos Produtos Cartográficos Digitais (PEC-PCD), o qual categoriza os dados geoespaciais conforme o seu nível de qualidade (CARVALHO e SILVA, 2018).

O Decreto nº 89.817, de 20 de Junho de 1984, surgiu para estabelecer critérios de classificação das cartas topográficas quanto à sua exatidão e à distribuição de erros ao longo das mesmas, através de um indicador estatístico de qualidade denominado “Padrão de Exatidão Cartográfica (PEC)”. Com a evolução tecnológica, a disseminação do conhecimento, a popularização de equipamentos de geotecnologias e demandas de usuários indicaram a necessidade do estabelecimento de novos padrões de qualidade para os produtos cartográficos (DSG, 2016b). Para atender a estas necessidades, o DSG (Departamento de Serviço Geográfico) do Exército Brasileiro, através de uma série de Especificações Técnicas (ET), definiu normas, padrões e especificações para controle de qualidade da produção cartográfica nacional. Nesse contexto, o PEC-PCD surgiu como uma evolução do antigo PEC, e os valores

referentes ao mesmo estão apresentados na Norma EB80-N-72.003 do ano 2016, também conhecida como Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV).

Segundo o DSG (2016b), para um produto cartográfico digital ser aceito como produto de referência do Sistema Cartográfico Nacional, a exemplo do previsto para o PEC (produtos impressos em papel), noventa por cento [90% ou $1,6449 \cdot \text{Erro Padrão (EP)}$] dos erros dos pontos coletados no produto cartográfico, comparando-os com as suas coordenadas levantadas em campo por método de alta precisão, devem apresentar valores iguais ou inferiores aos previstos ao PEC-PCD, devendo ainda apresentar os valores de EP também iguais ou inferiores aos previstos na ET-ADGV.

Quanto aos modelos digitais de elevação, a ET-ADGV estabelece valores específicos para esse tipo de produto, os quais podem ser observados no Quadro 01. As classes “B”, “C” e “D” do PEC-PCD correspondem, em ordem, às classes “A”, “B”, “C” do PEC Altimétrico previstas no Decreto 89.817, de 20 de junho de 1984, enquanto que a classe “A” (PEC-PCD) teve seus valores definidos a partir de adaptações dos estudos realizados por MERCHANT (1982) e ASPRS (1989), nos quais o PEC-PCD = $0,27 \cdot (\text{Equidistância do produto cartográfico})$ e o EP = $1/6 \cdot (\text{Equidistância do produto cartográfico})$ (DSG, 2016b).

Quadro 01 – Padrão de Exatidão Cartográfica Altimétrica dos Pontos Cotados e do MDT, MDE e MDS para a produção de Produtos Cartográficos Digitais.

PEC - PCD ⁽³⁾	1:1.000 (Eqd = 1 m)		1:2.000 (Eqd = 1 m)		1:5.000 (Eqd = 2 m)		1:10.000 (Eqd = 5 m)		1:25.000 (Eqd = 10 m)		1:50.000 (Eqd = 20 m)		1:100.000 (Eqd = 50 m)		1:250.000 (Eqd = 100 m)	
	PEC (m)	EP (m)	PEC (m)	EP (m)	PEC (m)	EP (m)	PEC (m)	EP (m)	PEC (m)	EP (m)	PEC (m)	EP (m)	PEC (m)	EP (m)	PEC (m)	EP (m)
A	0,27	0,17	0,27	0,17	0,54	0,34	1,35	0,84	2,70	1,67	5,50	3,33	13,70	8,33	27,00	16,67
B	0,50	0,33	0,50	0,33	1,00	0,66	2,50	1,67	5,00	3,33	10,00	6,66	25,00	16,66	50,00	33,33
C	0,60	0,40	0,60	0,40	1,20	0,80	3,00	2,00	6,00	4,00	12,00	8,00	30,00	20,00	60,00	40,00
D	0,75	0,50	0,75	0,50	1,50	1,00	3,75	2,50	7,50	5,00	15,00	10,00	37,50	25,00	75,00	50,00

Fonte: DSG, 2016b.

2.2.2 Acurácia de MDE globais encontrada na Literatura

Na literatura são encontrados muitos trabalhos técnicos e científicos voltados à medição do grau de acurácia dos MDE globais, a grande maioria com base em RMSE e desvio padrão. Diante desta vasta literatura acerca do tema, foram elencados alguns autores desde 2013, trazendo-se os valores de RMSE encontrados por cada um em diversas partes do Brasil e do mundo. Estas informações estão resumidas na Tabela 01, a qual contém valores de RMSE encontrados na literatura para alguns dos modelos digitais de elevação mais utilizados no Brasil (SRTM -30m, ASTER GDEM, AW3D30, e TOPODATA).

Tabela 01: Resumo de alguns valores de RMSE dos MDE SRTM (30m), ASTER GDEM, AW3D30 e TOPODATA, encontrados na literatura nos últimos 10 anos.

Autor	RMSE do MDE				Área de estudo	Origem dos dados de validação	Ano
	SRTM 30m	ASTER GDEM	AW3D30	TOPODATA			
Mukherjee et al (2013)	9,2	6,08			Utaracanda - IND	Receptor GNSS e Cart. Topográficas	2013
Araújo et al (2014)		9,69		10,83	Mato Grosso - BRA	Nivelamento trigonométrico	2014
Moura, Blas e Brites (2014)		7,03		7,75	Distrito Federal - BRA	Vértices de Rede Geodésica	2014
Souza (2015)		4,31		5,39	Paraíba - BRA	Receptor GPS Topográfico	2015
Santillan e Makinano-Santillan (2016)	8,28	11,98	5,68		Midanao - PHL	Nivelamento Geométrico	2016
O'Loughlin et al (2016)	14,12				Global	Lidar (ICESat)	2016
Shiramizu, Doi e Aoyama (2017)		25,83	8,5		Antártica	Receptor GNSS-RTK	2017
Moraís (2017)	4,95	6,44		8,9	Minas Gerais - BRA	Lidar	2017
Silva, Santiago e Silva (2017)	11,69		12,45		Goiás - BRA	Altímetro eletrônico	2017
Elkhrachy (2018)	7,92	7,45			Najran - SAU	Base Cartográfica e Rec. GNSS-RTK	2018
Girão, Vicens e Fernandes (2019)		7,66			Rio de Janeiro - BRA	Receptor GNSS	2019
Barbosa et al (2019)			5,34		Minas Gerais - BRA	Receptor GNSS Geodésicos	2019
Machado et el (2019)	7,02				São Paulo - BRA	Base de dados ITESP e SIGEF	2019
Pereira e Caten (2019)	8,65			12,92	Santa Catarina - BRA	Receptor GNSS Geodésico	2019
Viel, Rosa e Júnior (2020)	8,47	8,35	9,18		Rio Grande do Sul - BRA	Aerofotogrametria	2020
Uemaa et al (2020)	6,59	10,36	7,92		Estônia	Lidar	2020
	10	13,52	7,18		China	Estereoscopia do Pleiades-1A	
	13,07	11,77	11,42		Nova Zelândia	Lidar	
		9,22	4,98		Noruega	Lidar	
Nikolakopoulos (2020)			8,58		Grécia	Vértices de Rede Geodésica	2020
Magruder et al (2021)	5,5				Carolina do Norte - EUA	Lidar	2021
Tahire Din (2022)	2,38	5,79	1,24		Basrah - IRQ	Receptor GNSS-RTK	2022
Freire-Silva et al (2022)	4,58	20,67	6,96	9,64	Pernambuco - BRA	Lidar	2022
RMSE Médio Geral	8,16	10,38	7,45	9,24			
DP Geral	3,21	5,68	3,00	2,58			
RMSE Médio Brasil	7,56	9,16	8,48	9,24			
DP Brasil	2,65	5,34	3,08	2,58			
	SRTM 30m	ASTER GDEM	AW3D30	TOPODATA			

Fonte: Próprio autor (2023)

2.3 RETIFICAÇÃO DE MDE

A fase de detecção de erros deve ser seguida pela de correção, onde os dados imprecisos são retificados. No caso de modelos vetoriais, a correção pode ocasionalmente limitar-se à remoção de dados escusos sem perda apreciável de informação. Em MDE, modelos de espécie matricial, é necessário substituir o valor errado pelo correto uma vez que não é aceitável a existência de lacunas numa matriz (FELICÍSIMO, 2001). Estas imprecisões têm fontes distintas, podendo estar vinculadas ao tipo de cobertura do solo, a topografia ou até mesmo a falhas na aquisição e processamento dos dados. Deste modo, existem duas formas de se corrigir MDE para aumento de sua acurácia: a primeira maneira se baseia na identificação da magnitude e natureza das fontes de erro, para que então as mesmas possam ser minimizadas; a segunda forma é baseada na identificação de relações matemáticas entre os dados altimétricos provenientes do imageamento por sensoriamento remoto e os dados de referência obtidos por métodos de alta precisão, para a mesma área (AUDETTE, FERRIE e PETERS, 2000; GRUEN e AKCA, 2005; GIRÃO et al, 2019).

De todo modo, a busca por melhoria na acurácia de MDE é assunto por vezes pautado na literatura, buscando-se sempre formulações e metodologias corretivas que melhor se enquadrem a cada tipo de dado, região ou cobertura do solo. Deste modo, serão apresentados a seguir alguns estudos que buscaram minimizar os erros relativos a modelos digitais de elevação, tanto pela abordagem de identificação das fontes de erro, como pela busca de relações matemáticas entre dados avaliados e dados de validação.

O'loughlin et al (2016) trabalharam na correção global do viés de vegetação no modelo SRTM. Em seu relatório descreveram uma metodologia baseada na estimativa de altura de dossel, onde aplicaram dados de dossel advindos de outros autores e instituições. Deste modo, testaram cinco equações empíricas diferentes, todas em função de dados de altura e densidade de dossel, como o mapa de altura da vegetação HVEG produzido por Simard et al. (2011); o produto MODIS VCE que estima densidade e profundidade de penetração do SRTM no dossel; e dados altimétricos do ICESat, um altímetro a laser. Assim, aplicaram álgebra de bandas para cada uma das equações, testando seus resultados finais através do RMSE e média de erros (ME), usando dados *Lidar* como referência para validação. Assim, os autores concluíram seus estudos com uma redução significativa no RMSE do SRTM de mais de 14 m para 6 m, em escala global.

Gallant e Read (2016) num projeto semelhante ao anterior, também buscaram retificar o MDE SRTM em áreas de floresta. Os dados aplicados para correções foram o *glc30* (Classificação de cobertura da terra – GlobeLand30), imagens de radar do Alos-Palsar, usando os dados de polarização HH e o Tree cover 2000 (Cobertura da copa das árvores no ano 2000). Inicialmente, delimitaram áreas com a presença e ausência de árvores, através da classificação de imagens do Landsat e de dados Tree cover 2000. Em seguida, estimaram a altura de dossel, tendo como base os dados supracitados, produzindo um modelo de deslocamento por presença de árvores. Por fim, fizeram a subtração das alturas da vegetação no SRTM. Suas conclusões são parciais, não tendo ainda comparativo de acurácia antes e depois da correção.

Girão et al (2019) abordaram a correção de erros pressupondo a aleatoriedade das fontes. Desta forma, modelaram superfícies de compensação de erro por meio de interpolações determinísticas e geoestatísticas para o aumento da acurácia vertical do MDE ASTER GDEM, tendo como região de estudo o Estado do Rio de Janeiro. Para atingir seu objetivo os autores, num primeiro momento, selecionaram os pontos de controle e validação oriundos do projeto RJ25 (IBGE) com valores de altitudes ortométricas; em seguida

calcularam as diferenças de erro para cada ponto amostral, entre a altitude levantada em campo e a capturada no ASTER GDEM; após, eles interpolaram os valores da fase anterior por meio de dois interpoladores determinísticos, o TIL (Triangulação com Interpolação Linear) e o IDW (Inverso da Distância Ponderada), e um interpolador geoestatístico, a Krigagem Ordinária; estes produtos, denominados superfícies de compensação de erro, foram somados ao MDE ASTER GDEM. Os resultados levaram os autores a concluir que o interpolador geoestatístico ofereceu os melhores resultados, tendo um MDE corrigido de classe B para o PEC-PCD na escala 1:50.000, e classe A nas escalas 1:100.000 e 1:250.000.

Magruder et al. (2021) propuseram correções para o MDE SRTM na região leste do estado norte-americano da Carolina do Norte. O objetivo do estudo foi usar dados ICESat-2 e Landsat VCE como a base para corrigir elevações de terreno do SRTM em uma aplicação regional. Sua metodologia consistiu em modelar uma equação usando as diferenças de elevação entre o SRTM e ICESat-2, associada a uma técnica de regressão múltipla que utiliza essas diferenças verticais, a declividade do terreno determinada pelo SRTM e a porcentagem de cobertura vegetal derivada do Landsat VCE, como dados de entrada do modelo. Os coeficientes do modelo de correção são determinados por meio de uma técnica de decomposição ortogonal-triangular, onde a cobertura do dossel é tratada como ortogonal e a declividade como matrizes triangulares superiores. Os resultados mostraram que o RMSE geral do modelo SRTM original saiu de 5,5 metros para 3,2 metros no modelo corrigido, uma melhora significativa de mais de 2,0 metros. Além disso, foi mostrado que a modelagem proposta pelos autores tem mais eficácia em regiões de alta declividade e cobertura vegetal, onde em regiões planas e de pouca cobertura vegetativa o ganho em RMSE foi de apenas 0,44 metro. Os autores concluem que as correções do modelo podem atingir uma redução dos vieses em mais de 50% em algumas áreas, significando um sucesso para correções no SRTM no local de estudo. A modelagem proposta fornece uma base para pesquisas futuras para explorar a viabilidade de aplicar a técnica em ecossistemas e estruturas morfológicas diferentes. Além disso, a abordagem do modelo de correção mostra potencial para ser usado na correção de outros MDE's provenientes de radar e fotogrametria estéreo.

Hawker et al (2022) utilizaram aprendizado de máquina com o algoritmo *Random Forest* para remover edifícios e florestas do MDE Copernicus, produzindo um mapa global de elevação sem edifícios e florestas, em resolução espacial de 30 metros. Para tal, os autores treinaram o *Random Forest* para criar um modelo de correção utilizando um conjunto de dados de elevação de referência de 12 países, cobrindo uma ampla gama de zonas climáticas e

extensões urbanas. Para corrigir os erros devido à floresta utilizaram 15 preditores como entrada na regressão, enquanto que para a remoção de edifícios utilizaram 14 preditores. Conforme os autores, essa abordagem tem aplicabilidade mais ampla se comparado às modelagens anteriores as quais utilizaram dados de treinamento de um único país. Os resultados da pesquisa mostraram uma redução do erro vertical médio absoluto em áreas construídas de 1,61 para 1,12 m, e em florestas de 5,15 para 2,88 m.

2.4 MDE ALOS WORLD 3D - 30M

O MDE Alos World 3D é um dos produtos da missão ALOS (*Advanced Land Observing Satellite*), cujo satélite lançado em 24 de janeiro de 2006 pela JAXA, tem por objetivo contribuir na cartografia, observação regional, monitoramento de desastres e levantamento de recursos, por meio do avanço das tecnologias de observação terrestre (JAXA, 2008). O ALOS está equipado com três instrumentos de observação: o radiômetro pancromático PRISM (*Panchromatic Remote-sensing Instrument Stereo Mapping*), o radiômetro multiespectral AVNIR-2 (*Advanced Visible and Near Infrared Radiometer 2*) e o radar de abertura sintética PALSAR (*Phased Array type L-band Synthetic Aperture Radar*). A operação do ALOS foi concluída em 12 de maio de 2011.

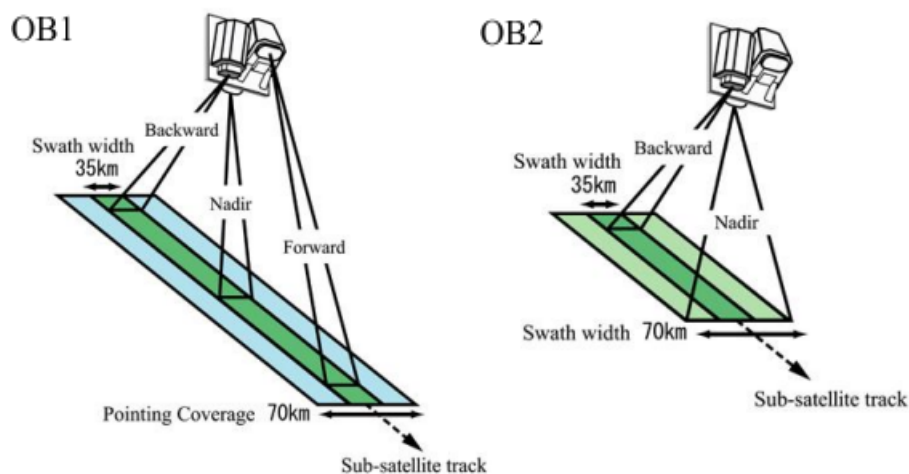
O radiômetro pancromático PRISM, sensor que forneceu os dados para a geração do DEM AW3D30, é um instrumento óptico que consiste em três radiômetros pancromáticos independentes apontados para nadir (NDR), para frente (FWD), e para trás (BWD), formando uma imagem estereoscópica tripla com resolução espacial de 2,5 m e largura de faixa de imageamento de 35 km, no modo de observação tripla (OB1). Também tem a capacidade de obter imagens a nadir (NDR) e para trás (BWD), no modo de observação estéreo (OB2) (TADONO et al., 2009). O Quadro 02 mostra algumas informações técnicas do PRISM, enquanto que a Figura 01 esboça o esquema de imageamento dos dois modos de aquisição do PRISM.

Quadro 02: Características técnicas do sensor PRISM.

Características do PRISM	
ITEM	DESCRIÇÃO
Número de Bandas	1 (Pancromática)
Comprimento de onda	0,52-0,77 micrômetros
Número de ópticos	3 (NDR, FWD e BWD)
Resolução espacial	2,5 m (NDR)
Largura da faixa	35 km (OB1) / 70 km (OB2)
Resolução radiométrica	8 bits

Fonte: Adaptado de Tadono *et al* (2014)

Figura 03: Geometria de observação do sensor PRISM em modo triplo (OB1) e em modo estéreo (OB2).



Fonte: Tadono *et al* (2014)

Conforme Santos et al (1999), por meio da estereoscopia orbital é possível obter dados tridimensionais com a observação de um par de imagens planas (pares estereoscópicos) adquiridas de uma mesma cena, com ângulos de incidência diferentes. A construção do modelo tridimensional Alos World 3D se deu pelo processamento estereoscópico do conjunto de imagens gerado pelo PRISM, formando um modelo digital de elevações da superfície com 5 metros de resolução espacial, batizado de AW3D DSM. A precisão do produto foi definido em 5 m (RMSE) na vertical e também 5 m (RMSE) na horizontal (TAKAKU et al, 2016). O produto em resolução original de 5 metros pode ser adquirido comercialmente. Já para disponibilização gratuita, a JAXA lançou o “ALOS World 3D-30m” (AW3D30), oriundo de uma reamostragem do produto AW3D DSM.

Atualmente, para o Brasil, está disponível a versão 3.1 dos dados AW3D30. Conforme a JAXA (2021), esta é uma versão criada a partir da reconsideração do formato nas áreas de altas latitudes, aplicação de dados auxiliares e novos métodos de processamento. Diferentes espaçamentos de pixels para cada zona de latitude foram adotados, dados da costa, que são um dos dados auxiliares, foram alterados e novos dados suplementares foram aplicados, além disso, à versão 3 do AW3D também foi aplicada como fonte de dados para o Japão, e houve melhora no método de detecção de valores anômalos (JAXA, 2021).

Os dados do modelo AW3D30 foram disponibilizados com altitudes referenciadas ao geoide EGM96, e, dado que discrepância entre o modelo EGM96 e o Datum Vertical Brasileiro é algo em torno de 10 centímetros (LUZ et al, 2002), neste trabalho serão desconsideradas conversões entre os dois sistemas pois a discrepância entre eles é irrisória

diante do RMSE de MDE globais comumente encontrado na literatura, que fica numa média de 7,45m para o AW3D30, como demonstrado na Tabela 01.

2.5 PROJETO MAPBIOMAS

O projeto MapBiomias trata-se de um mapeamento anual da cobertura e uso do solo, bem como monitoramento da superfície de água e cicatrizes de fogo, com dados a partir do ano 1985. Além disso, desde 2019 produz relatórios para cada evento de desmatamento detectado no Brasil. Tal projeto é empreendido por uma rede colaborativa, formada por ONGs, universidades e *startups* de tecnologia, com o propósito de revelar transformações do território brasileiro por meio da ciência, tornando acessível o conhecimento sobre a cobertura e o uso da terra, para buscar a conservação e o manejo sustentável dos recursos naturais, como forma de combate às mudanças climática (PROJETO MAPBIOMAS, 2022).

O conjunto de dados de entrada para o mapeamento consiste em imagens orbitais dos satélites Landsat-5 (Sensor TM), Landsat-7 (sensor ETM) e Landsat-8 (sensores OLI e TIRS), em resolução espacial de 30 metros. Tais dados são processados com tecnologia Google, a qual inclui processamento de imagem em infraestrutura de computação em nuvem, dentro da plataforma *Google Earth Engine* e armazenamento de dados usando *Google Cloud Storage*, com emprego de programação Javascript e Python (SOUZA e AZEVEDO, 2022).

Os resultados, disponibilizados de modo gratuito na plataforma MapBiomias (<https://mapbiomas.org/>), consistem em mapas e gráficos os quais esboçam uma série histórica de 33 anos, até então, acerca da evolução do uso e cobertura do solo no Brasil. Para usuários em geral estão disponíveis mapas de cobertura e uso da terra, mapas de cicatrizes de fogo, mapas de superfície da água, mosaicos de imagens de satélite, todos em resolução espacial de 30 metros, os quais podem ser acessados na Plataforma *Web*, na qual estão disponíveis imagens, mapas e estatísticas das coleções do MapBiomias. Para usuários especializados estão disponíveis ainda notas metodológicas, coleção de *scripts* no GitHub, plugIns para QGis, ferramentas de acesso e download dos dados no *Google Earth Engine* e notas técnicas (PROJETO MAPBIOMAS, 2022).

3 METODOLOGIA

Neste capítulo serão mostrados todos os procedimentos metodológicos para atingir os objetivos propostos neste trabalho, como segue.

3.1 ÁREAS DE ESTUDO E VALIDAÇÃO

Neste trabalho houveram cinco regiões amostrais, como forma de melhorar a consistência estatística das análises. Estas amostras estão divididas em dois grandes grupos: grupo de amostras contínuas (Grupo 01) e grupo de amostras pontuais (Grupo 02). Ressalta-se que estas regiões amostrais serviram de base para extração de todos os parâmetros estatísticos que embasam a metodologia de correção.

No grupo de amostras contínuas (dados matriciais), Grupo 01, existem quatro recortes de estudos, dois no estado de Pernambuco e dois no estado de São Paulo (áreas em vermelho no mapa de localização). A seleção destas regiões, em específico, se justifica pelas fontes dos dados de controle e validação, as quais se limitam aos estados supracitados. No grupo de amostras pontuais (dados vetoriais), o Grupo 02, há diversos pontos espalhados pelo Brasil (pontos azuis no mapa de localização), advindos da Rede Altimétrica do IBGE (Instituto Brasileiro de Geografia e Estatística). As áreas possuem grande variabilidade de classes de uso e cobertura do solo, bem como de classes de declividade, de modo a ter amostragem para diversas características de terreno. O mapa de localização das chamadas “áreas de estudo” está expresso na Figura 04.

Além das áreas de estudo descritas anteriormente, foi necessário ainda mais 3 áreas para validar a metodologia de correção proposta. Tais áreas foram chamadas de “áreas de validação”, e se localizam nos estados de São Paulo, Pernambuco e Bahia. O mapa expresso na Figura 04 mostra a disposição espacial das áreas de validação, as quais estão destacadas em verde.

3.2 DADOS E RECURSOS

Durante os estudos foram utilizados dados vetoriais e matriciais de fontes externas e levantadas pelo próprio autor, além de programas computacionais para modelagem dos dados, sendo listados a seguir:

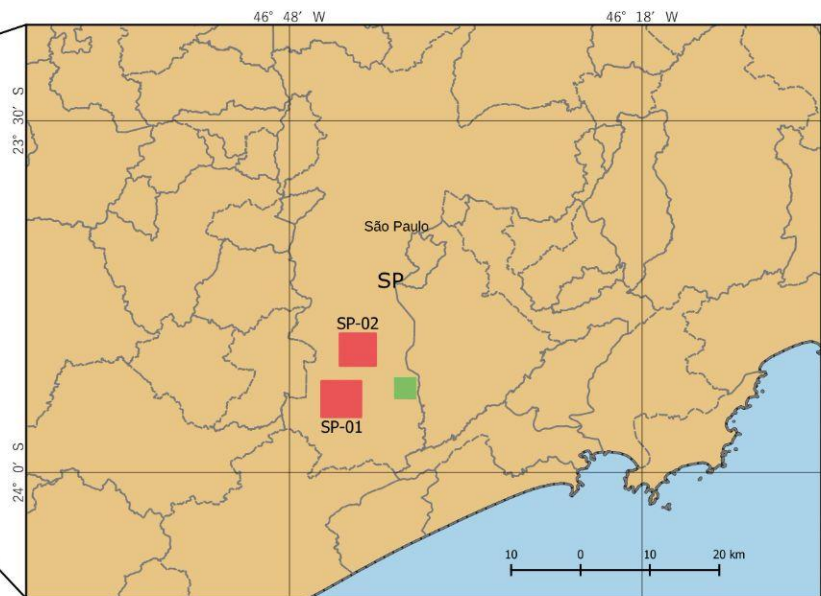
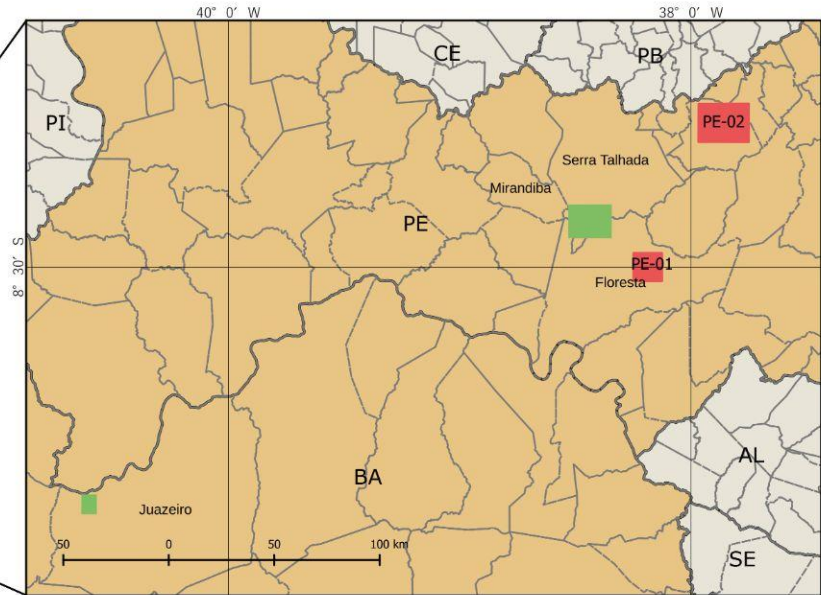
- a) Recortes do MDE AW3D30 para as áreas de estudo e validação;
- b) Raster de uso e cobertura do solo do ano 2006 advindo da plataforma MapBiomias;

FIGURA 04: MAPAS DE LOCALIZAÇÃO DAS ÁREAS DE ESTUDO E VALIDAÇÃO



LEGENDA

- | | |
|--------------------------------------|----------------------|
| Fronteiras dos países sul-americanos | AMOSTRAS |
| Oceanos | Amostras do Grupo 01 |
| LIMITES DE ESTADO | Amostras do Grupo 02 |
| Estados em estudo | Áreas de validação |
| Demais estados | |



Fonte da figura: Próprio autor (2023)

c) Dados vetoriais de pontos da Rede Altimétrica de Alta Precisão (RAAP) do Sistema Geodésico Brasileiro (SGB), estabelecida através da técnica de nivelamento geométrico de alta precisão, para determinação das altitudes referenciadas ao Nível Médio do Mar (NMM) em Imbituba –SC.

d) Dados *Lidar* (*Light Detection and Ranging*) advindos de duas fontes distintas:

1. Dados *Lidar* do município de São Paulo disponíveis na plataforma GeoSampa;
2. Dados *Lidar* do Estado de Pernambuco disponíveis na plataforma Pernambuco Tridimensional;

e) Pontos tridimensionais levantados com receptores GNSS pelo método estático rápido;

f) *Software* proprietário de planilha eletrônica *Excel*;

g) *Software* SIG (sistema de informação geográfica) de código aberto QGIS 3.28;

h) Plataforma de processamento em nuvem *Google Earth Engine*;

i) *Software* *Sisvar* para cálculos estatísticos;

j) *Software* *iNZight* para plotagem de gráficos;

k) *Software* *CloudCompare*, aplicativo de código aberto para processamento de nuvem de ponto 3D e TIN (*Triangle Irregular Network*);

3.3 MÉTODOS

Basicamente, a metodologia para atingir os objetivos deste trabalho está seccionada em 6 etapas, listadas abaixo e organizadas no fluxograma da Figura 05:

1. Aquisição dos dados necessários a esta pesquisa, os quais foram mencionados no item 3.2;

2. Tratamento dos dados altimétricos de controle, advindos do GeoSampa, PE3D, RAAP e do levantamento GNSS;

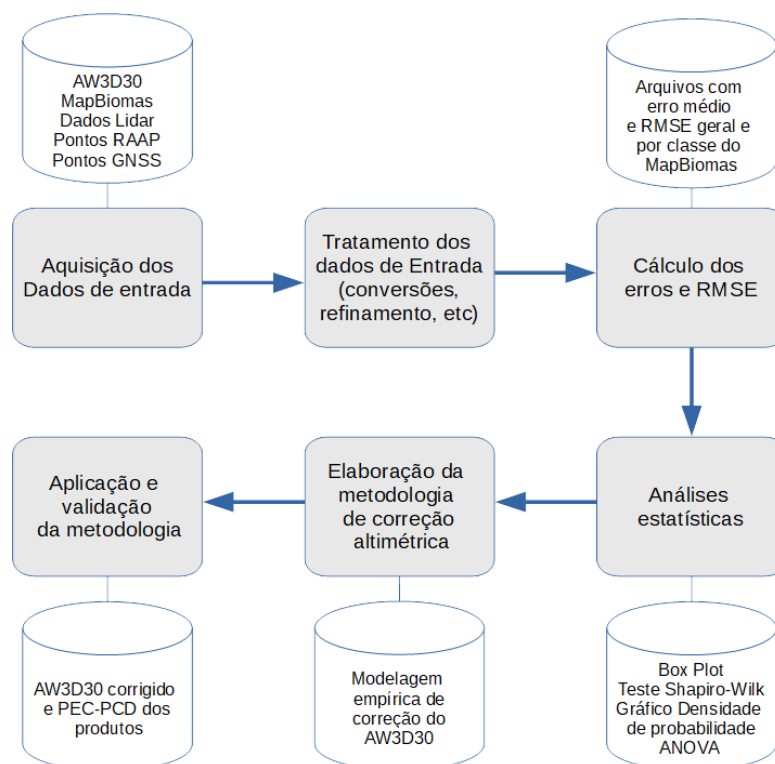
3. Cálculo das diferenças entre as elevações do MDE AW3D30 e as de controle (erros), bem como do erro em módulo para cada unidade amostral;

4. Estudos estatístico para avaliar a consistências dos dados e averiguar a qualidade altimétrica do AW3D30 por classe do MapBiomass;

5. Elaboração de uma metodologia para modelagem da superfície de correção em função da magnitude de erros das classes do MapBiomass;

6. Aplicação da metodologia para corrigir o MDE AW3D30 e validação dos resultados;

Figura 05: Fluxograma da metodologia empregada neste trabalho



Fonte: Próprio autor (2023).

3.3.1 Aquisição dos dados

O início deste estudo, como de praxe, consistiu na captura de dados necessários à execução da pesquisa proposta. Os dados aqui utilizados vieram de distintas fontes, e foram obtidos por meio de sítios na internet ou levantamento em campo. Convém mencionar que, embora estes dados derivem de fontes distintas, as entidades de fornecimento dos mesmos possuem grande respaldo técnico e acadêmico, consistindo, na sua totalidade, em órgãos governamentais de todas as esferas: municipais, estaduais ou federais. A seguir, expor-se-á os detalhes de obtenção dos conjuntos de dados utilizados.

a) Dados Alos World 3D

Estes dados foram obtidos no portal da própria JAXA, onde o modelo é disponibilizado gratuitamente, sendo necessária a realização de um cadastro e consequente geração de login e senha para baixar os dados. Este conjunto de dados foi obtido no formato GeoTiff, com cada pixel representando 30x30 metros em solo, resolução radiométrica de 16 bits, Datum horizontal WGS-84 em coordenadas geográficas e Datum vertical EGM96.

b) Raster de uso e cobertura do solo MapBiomas

Este conjunto de dados, também obtido gratuitamente, é originário do projeto MapBiomas, o qual tem por objetivo central produzir um mapeamento anual da cobertura e

uso da terra, desde 1985, para todo território brasileiro. O *download* do raster de uso e cobertura do Brasil no ano de 2006 foi realizado pela Plataforma Google Earth Engine, através de um código (*Toolkit*) disponibilizado pelo próprio MapBiomias. Ressalta-se que o ano 2006 foi o escolhido pelo fato de que este é o mesmo ano da tomada dos dados que originaram o DEM Alos World 3D, visando, assim, evitar eventuais discrepâncias devido às alterações na cobertura terrestre ao longo do tempo. Este conjunto de dados foi obtido no formato GeoTiff, com cada pixel representando 30x30 metros em solo, resolução radiométrica de 8 bits, Datum horizontal SIRGAS-2000 em coordenadas geográficas.

c) Dados da Rede Altimétrica de Alta Precisão

Conforme o IBGE (2022), a Rede Altimétrica de Alta Precisão (RAAP) fornece as referências verticais para todo território nacional, e compõem o Sistema Geodésico Brasileiro (SGB). Também denominadas Referências de Nível (RRNN), esta rede de vértices foi estabelecida desde 1945, utilizando-se o método de nivelamento geométrico de alta precisão, e ao longo dos anos novas linhas de nivelamento foram adicionadas através de novas técnicas de medição e processamento. Esta rede foi baixada em formato vetorial *Shapefile* do tipo ponto, no Portal Brasileiro de Dados Abertos do governo federal (BRASIL, 2022). Suas georreferências são: Datum horizontal SIRGAS-2000 em coordenadas geográficas, com as altitudes ortométricas referenciadas ao marégrafo de Imbituba - SC.

d) Dados Lidar do município de São Paulo

Os dados Lidar de São Paulo foram adquiridos na plataforma GeoSampa, a qual reúne informações detalhadas e georreferenciadas sobre a cidade de São Paulo. Através do sítio eletrônico, o usuário pode consultar e baixar informações de várias fontes, em um único mapa, sem a necessidade de cadastro de usuário e senha (SÃO PAULO, 2022). Os dados *Lidar* obtidos foram capturados através do sensor Sensor *Optech Gemini*, cuja acurácia vertical varia entre 5 e 30 cm. Este conjunto de dados trata-se de uma nuvem de pontos em formato LAZ, com Datum horizontal SIRGAS-2000 em coordenadas planas na projeção UTM, e altitudes ortométricas referenciadas ao marégrafo de Imbituba - SC. Todas as informações elencadas acima foram passadas pela equipe do Geosampa através de e-mail.

e) Dados Lidar do estado de Pernambuco

Os dados *Lidar* de Pernambuco foram adquiridos na plataforma Pernambuco Tridimensional (PE3D), a qual foi criada pelo governo de Pernambuco para disponibilizar gratuitamente dados geoespaciais do estado à comunidade (PERNAMBUCO, 2022). Em especial, destaca-se os dados *Lidar*, cuja densidade é 1 ponto por 1,3 m², em média, com

acurácia igual ou melhor que 25 cm. Este conjunto de dados trata-se de uma nuvem de pontos obtida em formato *XYZ*, Datum horizontal SIRGAS-2000 em coordenadas planas da projeção UTM. Não se pode apontar com plena certeza o referencial altimétrico destes dados, visto que tais informações não foram encontradas durante buscas no próprio site da plataforma, bem como na bibliografia acerca do tema. Assim, foi enviado um e-mail para a equipe do PE3D, entretanto, até o momento não foi obtida resposta. Deste modo, foi arbitrado a este conjunto de dados altitudes ortométricas referenciadas ao Datum vertical de Imbituba-SC, visto que se pressupõe que foram seguidos os normativos nacionais quanto à produção de dados geoespaciais.

f) Pontos levantados por Posicionamento GNSS

Para auxiliar a validação da metodologia de correção proposta neste estudo, foi necessário aplicá-la numa área com condições diferentes daquelas que embasaram sua construção, ou seja, numa área, preferencialmente, distinta de São Paulo e Pernambuco, e que os dados para validação não sejam oriundos da RAAP, PE3D ou GeoSampa. Diante disso, foram coletados em campo, na cidade de Juazeiro-BA, pontos amostrais utilizando-se um par de receptores GNSS modelo Hiper II, do fabricante Topcon. O método de posicionamento aplicado foi o estático rápido, onde um dos receptores permanece num ponto de coordenadas conhecidas (chamado de Base) enquanto os pontos de interesse são coletados a partir do segundo receptor (chamado de Rover). O tempo de permanência do Rover nos pontos de interesse é de até 20 minutos, no método estático rápido.

3.3.2 Tratamento dos dados

Todo trabalho científico com dados pressupõem um tratamento inicial para minerar os dados brutos de modo a evitar que o estudo seja enviesado pela presença de amostras não representativas, dentro do escopo proposto. Além disso, como o presente caso utiliza dados de fontes distintas, se faz necessário trabalhar a compatibilização dos mesmos, pois eles vêm em formatos primários, projeções, Data verticais e horizontais, etc, todos diferentes, evocando a necessidade de normalizar tais formatações, e só assim iniciar o estudo propriamente dito. Deste modo, a seguir serão mostradas todas as compatibilizações e tratamentos iniciais aplicados aos dados utilizados neste estudo.

3.3.2.1 Tratamento das nuvens de pontos (Grupo 01)

As nuvens de pontos utilizadas neste estudo são muito densas, e a aplicação da totalidade destes pontos no estudo inviabilizaria o processamento, dado a limitação operacional da máquina de processamento utilizada (notebook com intel core i7, 8 GB de memória RAM, e placa de vídeo com 2 GB de memória dedicada). As nuvens de pontos de Pernambuco e São Paulo utilizadas têm, respectivamente, 692 pontos e 210 pontos por pixel do AW3D30, em média. Diante disso, foi realizada uma amostragem dentro do espaço ocupado por cada pixel para reduzir o número de pontos *Lidar* por pixel do AW3D30. Na prática, esta amostragem foi realizada no *software CloudCompare*, mais especificamente com a ferramenta “*subsample a point cloud*”. Através desta ferramenta, arbitrou-se a amostragem de um ponto a cada 15 metros em solo, aleatoriamente. Isto significa que para cada pixel do AW3D30 foram selecionados de 2 a 5 pontos da verdade de campo, possibilitando extração de média.

Após a obtenção dos pontos amostrais, foi realizada a conversão do raster AW3D30 em uma camada do tipo vetor, onde cada pixel do raster foi transformado em um polígono, e os valores de intensidade dos pixels (valores da altitude) foram automaticamente alocados em uma coluna da tabela de atributos da camada vetor. Uma segunda coluna desta mesma tabela recebeu identificadores únicos (ID) para cada polígono-pixel. Esta etapa foi realizada com a função “Pixel de raster para polígono” do QGIS 3.28.

Por conseguinte, por meio da ferramenta “Associar pela localização”, foi atribuído a cada ponto da verdade de campo o “ID” do polígono-pixel cujo predicado geométrico fosse “Dentro”, em outras palavras, todo ponto amostral recebe, na tabela de atributos, o valor “ID” do polígono que o envolve. Isto é requisito para execução da próxima ferramenta: “Estatística por categoria”, através da qual foram calculadas estatísticas básicas (média, mediana, max, min, etc) das elevações dentro da categoria “ID”, ou seja, uma altitude de referência para cada pixel. A saída desta função é uma tabela, a qual foi atrelada, por meio do campo “ID”, à camada vetorial de pixels vetorizados. O resultado final destas sub-etapas foi a camada de pixels vetorizados com uma tabela de atributos contendo os campos “ID” (do pixel), “Altitude do AW3D30”, estatísticas gerais das altitudes dos pontos *Lidar*, assumindo-se a média como a altitude verdadeira de campo, para o posterior cálculo dos erros.

O processo executado no QGIS, como descrito acima, foi transformado em um Modelo de Ferramenta de processamento, como forma de otimizar a execução do mesmo, dado que ele seria repetido pelo menos quatro vezes. A ideia, a posteriori, é transformar este

Modelo num Plugin para o QGIS, de modo a contribuir com a comunidade de usuários do programa que necessitem realizar a tarefa descrita.

3.3.2.2 *Tratamento dos pontos da Rede Altimétrica (Grupo 02)*

Nos pontos da rede altimétrica o tratamento foi mais simples, dado que as elevações já estavam prontas na planilha de atributos. Deste modo, realizou-se uma filtragem nestes dados com o objetivo de se utilizar apenas os pontos cujas altitudes e posições horizontais foram determinadas por métodos de posicionamento de maior confiabilidade quanto a acurácia. Assim, somente os pontos que tiveram coordenadas planimétricas determinadas por GPS Geodésico, Poligonação ou triangulação e altitudes determinadas por Nivelamento Geométrico ou Trigonométrico foram selecionados. Esta filtragem foi executada no QGIS e resultou num total de 1662 pontos espalhados por todo território nacional.

3.3.2.3 *Processamento dos dados levantados com receptores GNSS*

Os pontos do levantamento de campo na cidade de Juazeiro-BA foram processados, inicialmente, utilizando o *software* Topcon Tools v7.1. Para tal, os arquivos originais de levantamento em formato *.TPS foram transferidos do par de receptores para o computador, e no ambiente digital foram processados aplicando-se ainda correções através de efemérides precisas. As coordenadas tridimensionais destes pontos foram exportados em formato *.TXT, e tiveram suas altitudes geométricas convertidas em altitudes ortométricas através do *software* livre MAPGEO2015, do IBGE. Este programa converte altitudes referenciadas ao elipsóide, advindas de levantamentos por GNSS, em altitudes ortométricas, relacionadas à superfícies física da terra (geóide), através do modelo de conversão hgeoHNOR2020.

3.3.3 **Cálculo dos erros**

Neste item se descreve como foi realizado o cálculo de erro altimétrico e erro em módulo, tanto para as amostras do Grupo 01 como do Grupo 02. Há de se destacar que o RMSE, na prática, foi calculado como a média dos erros em módulo para cada unidade amostral. Portanto, neste trabalho, quando se mencionar o termo “erro em módulo”, refere-se ao valor vinculado a cada unidade amostral (*pixel* ou ponto) usado como base para estimativa do RMSE.

3.3.3.1 Cálculo de erros do Grupo 01

Para execução desta etapa utilizou-se a tabela de atributos dos pixels vetorizados da etapa descrita em 3.3.2.1, e, dentro do programa de edição de planilhas eletrônicas, subtraiu-se a coluna referente às altitudes do AW3D30 da coluna contendo os valores médio de altitude dos pontos *Lidar*, para obtenção do erro altimétrico. Além disso, foi calculado também o erro em módulo, o qual consiste na raiz da diferença elevada ao quadrado da altitude do AW3D30 pela altitude de referência (Ver Equação 07). Isto foi repetido para cada região de estudo do Grupo 01. Por fim, para cada região do Grupo 01 foi gerada uma planilha simples contendo apenas duas colunas: erro e erro em módulo, como forma de facilitar as análises estatísticas.

$$Em = \sqrt{(Amde - Aref)^2} \quad \text{Eq. 07}$$

Onde “Em” é o erro em módulo; “Amde” é a altitude do MDE AW3D30; “Aref” é a altitude de referência.

3.3.3.2 Cálculo de erros do Grupo 02

Para cálculo do erro dos pontos do Grupo 02 foi necessário, antes de tudo, atribuir o valor de altitude do AW3D30 aos pontos da RAAP em sua tabela de atributos. Para tal, foi implementado um código (APÊNDICE A) na plataforma Google Earth Engine, a qual possui a base de dados do AW3D30 em nuvem, evitando assim fazer *download* do raster de todo o Brasil, já que a camada de pontos do RAAP se estende por todo território nacional. Obteve-se como resultado uma planilha com valores de altitude dos pontos da RAAP e do AW3D30, numa mesma linha, a partir da qual foi realizado o cálculo de erro e erro em módulo, de modo similar ao descrito no item 3.3.3.1. Por fim, a região amostral do Grupo 02 também recebeu uma planilha simples contendo apenas duas colunas: erro e erro em módulo.

3.3.4 Aplicação do estudo estatístico

Nesta seção serão apresentados os procedimentos metodológicos aplicados neste estudo, os quais objetivaram a obtenção dos dados e informações estatísticas que embasaram a determinação da acurácia do AW3D30, bem como a modelagem de correção das altitudes.

3.3.4.1 Eliminação de Outliers

Para eliminação dos *outliers* dos conjuntos amostrais foi necessário a plotagem de gráfico *Box Plot*. Esta sub-etapa foi executada usando-se o *software iNZight*, no qual são

apontados também os limites de corte de *outliers*, tanto superior quanto inferior. Para simples conferência, estes pontos de corte também foram calculados manualmente com auxílio do Excel, usando-se as Equações 04 e 05. Por fim, através também do Excel, realizou-se filtragem de valores *outliers* e posterior remoção dos mesmos nas planilhas de erro.

3.3.4.2 Teste de normalidade do conjunto amostral “erros”

O teste de normalidade da variável erro é premissa para as próximas etapas de análise estatística. Assim, para se analisar se os erros dos cinco conjuntos amostrais possuem ou não uma distribuição normal de probabilidade, foram construídos gráficos de densidade de probabilidade bem como realizou-se o teste Shapiro-Wilk.

Para elaboração dos gráficos utilizou-se o *software iNZight*, no qual foram carregadas planilhas eletrônicas geradas em 3.3.3, e através do programa mencionado foram plotados gráficos do tipo Densidade de probabilidade, o qual possui no eixo X a variável erro e no eixo Y as probabilidades distribuídas conforme seus respectivos valores de erro, sob a forma de uma curva.

Em seguida, usando-se o programa *Sisvar* procedeu-se a uma análise de normalidade conhecida como Teste de Shapiro-Wilk. Para tal, foi necessário realizar uma nova amostragem aleatória dentro de cada um dos cinco conjuntos de dados, reduzindo-os para 50 amostras cada. Isto foi feito perante a limitação do Teste de Shapiro-Wilk, o qual admite apenas conjuntos de até 50 amostras. Os cinco novos conjuntos de erros (Brasil, PE-01, PE-02, SP-01 e SP-02) foram postos em uma única planilha no formato *dBase (*.dbf)*, a qual foi posteriormente carregada no programa para execução do teste Shapiro-Wilk para cada coluna. Este processo só foi realizado após a remoção dos *outliers* dos conjuntos.

3.3.4.3 Comparação de média das regiões amostrais

Para testar se há homogeneidade dos erros encontrados nas 5 regiões amostrais, distribuídas nos dois grupos já descritos, aplicou-se análise de variância com teste de Tukey, os quais ajudam a presumir se os conjuntos amostrais pertencem a uma mesma população. Na prática, esta análise foi realizada com o *software Sisvar*, onde foi necessário organizar uma planilha contendo os erros e os erros em módulo categorizados por região amostral. Devido às limitações operacionais do *software* e do exorbitante número de pontos homólogos do Grupo 1, foi feita uma amostragem aleatória simples de 500 pontos em cada uma das cinco regiões

de estudo. Portanto, na ANOVA e teste de Tukey foram usadas 500 unidades amostrais para cada região de estudo, totalizando 2.500 unidades.

3.3.4.4 Comparação de média das classes do MapBiomias

Para verificar se há correlação entre as variáveis erros e classes do MapBiomias, também foi executada análise de variância com teste de Tukey. Dado que tal análise aponta diferença entre as médias de erro categorizadas pelas classes de uso e cobertura do solo do Map Biomias, espera-se que as categorias com presença de vegetação mais densa possuam maior média de erro altimétrico. Para realizar o teste de Tukey foi construída uma tabela contendo três colunas: uma com o erro absoluto, o erro em módulo e a respectiva classe do MapBiomias, e então foi repetido o processo descrito em 3.4.4.3, obtendo-se relatório de análise de variância e teste de Tukey.

3.3.5 Elaboração de uma metodologia para modelagem da superfície de correção

A modelagem de uma superfície de correção para o AW3D30 está baseada em álgebra de mapas, onde uma fórmula empírica condicional é aplicado em rasters do MapBiomias para se obter uma superfície que, no fim, é subtraída do raster AW3D30. Nos próximos itens desta seção serão mostrados alguns detalhes da elaboração da metodologia corretiva proposta neste trabalho.

3.3.5.1 Estudo do erro médio por classe do MapBiomias

Para embasamento da modelagem de correção, de início foi efetuada uma análise sobre os indicadores estatísticos dos erros por classes de uso e cobertura do MapBiomias. Assim, foram levantadas várias informações tais como erros por classe do MapBiomias e por região de estudo, incidência das classes por região, correlação dos erros com outras variáveis como declividade, por exemplo, entre outros. Estas análises permitiram decidir qual metodologia se iria utilizar para modelar a correção das altitudes. Assim, perante as características que foram apreendidas decidiu-se que a modelagem iria ser empírica ao invés de uma regressão, dado que não houveram correlações matemáticas significativas entre os erros e as outras variáveis, como a declividade, e dado ainda que a variável classes do MapBiomias, por ser qualitativa nominal, não poderia ter sua correlação com os erros medida por indicadores tradicionais, como Pearson ou Spearman, por exemplo.

3.3.5.2 *Modelo de correção empírico baseado em coeficientes de correção*

A modelagem empírica de correção proposta neste trabalho está arraigada em algumas premissas lógicas que foram apreendidas ao longo do estudo, e estão listadas a seguir:

a) O erro altimétrico médio não é uniforme para todas as áreas de estudo, com valores que variam entre 1,95 metros e 11,10 metros. Logo, uma possível correção tem de ser baseada no erro local ao invés de um mesmo valor para todas as áreas. Tal erro local deverá ser medido por um número mínimo de unidades amostrais. Este número mínimo terá sua estimativa demonstrada no item 3.3.5.4.

b) Só serão aplicadas correções para classes do MapBiomias as quais tiveram amostragem significativa para extração dos parâmetros estatísticos. Assim, foi arbitrado que apenas nas classes ocorridas em pelo menos 3 das 5 áreas de estudo seriam aplicadas algum tipo de correção.

c) A modelagem corretiva se baseará em coeficientes de correção, os quais estarão diretamente vinculados às classes do MapBiomias. Assim, cada classe do MapBiomias terá seu próprio coeficiente de correção, e este, por sua vez, será multiplicado pelo erro médio local para se determinar o valor que irá corrigir a altitude no AW3D30 no pixel em questão.

d) Os coeficientes de correção foram determinados por meio de análises estatísticas associadas a testes empíricos. Este procedimento está detalhado no item 3.3.5.3.

e) Após alguns testes, determinou-se a exclusão dos dados da área de SP-02 para cálculo dos coeficientes de correção, dado que esta área de estudo foi a única a não apresentar uma distribuição normal de probabilidade. Isto indica algum tipo de tendência nos dados desta área, que poderia, de algum modo, afetar o modelo de correção.

3.3.5.3 *Determinação dos coeficientes de correção*

Como foi identificado que o erro médio para o modelo AW3D30 têm grande variação conforme a localização geográfica da área de estudo, não é conveniente trabalhar uma correção que leve em consideração um mesmo erro médio para todos os locais onde se aplica a metodologia. Deste modo, optou-se pelo estabelecimento de coeficientes de correção, extraídos das análises estatísticas dos dados amostrais, os quais vão variar de acordo ao tipo de uso e cobertura do solo do local. Por fim, o valor da correção do modelo resultará da multiplicação deste coeficiente de correção pelo valor do erro médio local, determinado por uma pequena amostragem na área onde se pretende corrigir o MDE.

Diante disso, a estimação dos coeficientes foi baseada numa análise estatística nos valores de erro por classe do MapBiomias, onde se calculou um coeficiente para cada classe que ocorreu em pelo menos 3 das 4 áreas de estudo utilizadas para se extrair os parâmetros. Os coeficientes foram estimados através da fórmula empírica expressa na Equação 08:

$$Cc = \frac{Ec - Em}{Em} \quad \text{Eq. 08}$$

Onde Cc é o coeficiente de correção da classe; Ec é o erro médio da classe; Em é a média dos erros por classe (média das médias);

Na prática, a determinação dos coeficientes de correção se deu através de uma planilha de cálculos onde foram elencadas todas as classes de uso e cobertura do solo, que ocorreram nas áreas de estudo, correlacionando-as com o erro altimétrico médio das mesmas, para cada uma das cinco áreas de estudo. Após alguns testes empíricos, e levando-se em consideração o fato de que a área de estudo SP-02 não apresentou distribuição normal nos dados, a mesma foi descartada para cálculo dos coeficientes. Assim, com as outras 4 áreas de estudo que restaram (Brasil, PE-01, PE-02 e SP-01) procedeu-se o cálculo dos coeficientes de correção para as classes de uso e cobertura do solo ocorridas em pelo menos 3 das 4 áreas amostrais. Para as demais classes foi determinado a não aplicação da correção.

3.3.5.4 Número mínimo de unidades amostrais como entrada no modelo

Para se determinar uma quantidade mínima de unidades amostrais necessárias à aplicação da metodologia de correção numa determinada área, foi realizado um estudo estatístico como descrito a seguir.

Foram selecionados 11 conjuntos amostrais aleatórios dentro de cada uma das três classes a seguir: “Pastagem” (15), “Mosaico de agricultura e pastagem” (21) e “Formação Florestal” (3). A seleção destas classes seguiu dois critérios: o primeiro é a ocorrência em todas áreas de estudo, o segundo consiste nas maiores quantidades de unidades amostrais somando todas as áreas de estudo. A análise destes critérios foi apoiada por uma planilha de cálculos, a qual está expressa nos resultados deste trabalho, especificamente no item 4.1.7.

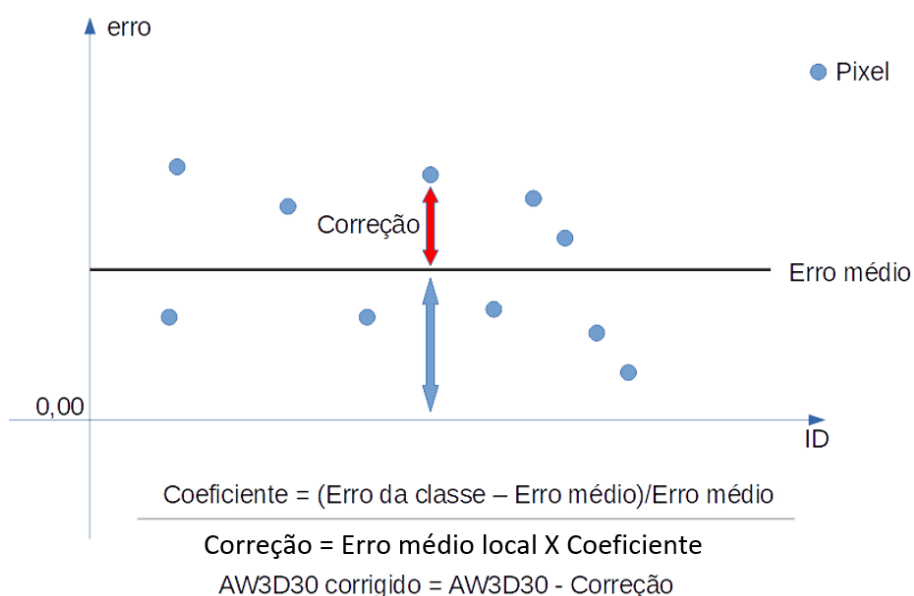
Os 11 conjuntos amostrais contêm de 10 a 110 unidades, escalonados de 10 em 10 unidades. A partir do cálculo de erro altimétrico médio e RMSE destas 11 amostras, foram construídos gráficos para se analisar a evolução dos valores médios, observando-se a partir de qual amostra estes valores começam a estabilizar. Este ponto indica qual amostragem mínima suficiente para ser considerada representativa na estimação do erro altimétrico médio do AW3D30. Assim, foram feitos dois gráficos para cada uma das três classes supracitadas, um

gráfico com a variável erro altimétrico médio e o outro com o RMSE. Nestes gráficos foram inseridos ainda valores de referência, que consistem em médias (do erro médio e RMSE) utilizando todas as unidades amostrais das 5 regiões de estudo para a classe em questão.

3.3.5.5 Modelagem de correção

A modelagem de correção se baseou na diminuição dos erros altimétricos médios locais através da aplicação de coeficientes de correção vinculados às classes de uso e cobertura do solo do MapBiomias. Para tal, foram realizados estudos estatísticos para se determinar o erro médio por classe, e através do mesmo juntamente com a média dos erros médios das classes, levando-se em conta todas as classes estudadas, foi efetuada uma espécie de normalização, utilizando a fórmula descrita na Equação 08, para determinação dos coeficientes de correção.

Figura 06: Demonstrativo da lógica aplicada na modelagem de correção empírica.



Fonte: Próprio autor (2023).

Assim, a intenção nesta modelagem empírica foi diminuir os efeitos dos erros altimétricos por meio de uma redução dos erros que excedem o erro médio local, como demonstrado na Figura 06, acima. A referência para essa redução são as classes de uso e cobertura do solo do MapBiomias, sendo que apenas para classes ocorridas em no mínimo 3 das 4 áreas de estudo escolhidas como referência (Brasil, PE-01, PE-02 e SP-01) foram calculados coeficientes. Para as demais classes determinou-se a não aplicação de correção, frente a baixa quantidade de dados estatísticos. Reitera-se ainda que das 5 áreas de estudo iniciais, uma delas foi descartada na determinação dos coeficientes, a SP-02, dado que a

mesma não apresentou distribuição normal na variável erro altimétrico. Na Tabela 19, na parte dos resultados, estão expressos os valores dos coeficientes e erro médio por classe do MapBiomias.

3.3.6 Aplicação da metodologia de correção e validação dos resultados

Na sexta etapa deste estudo foi realizada aplicação da modelagem corretiva através de álgebra de mapa, nos dados originais AW3D30. Assim, através da metodologia proposta em 3.3.5 foi construída uma superfície de correção a qual foi subtraída do AW3D30 original para obtenção do modelo corrigido. Para se validar a aplicabilidade do modelo empírico foram selecionadas três áreas de validação, diferentes das 5 áreas de estudo iniciais, sendo elas:

- Uma na cidade de São Paulo (utilizando dados Lidar do GeoSampa);
- Uma no Estado de Pernambuco (Utilizando dados Lidar do PE3D);
- Uma área no Estado da Bahia (Utilizando dados advindos de levantamento GNSS)

É importante destacar que as duas primeiras áreas de validação (em São Paulo e Pernambuco), por possuírem dados com a mesma origem dos que foram usados para construir o modelo, servem para apontar se a modelagem empírica tem validade dentro do próprio contexto delimitador deste estudo (região geográfica, fonte de dados, tipo de dados de validação, etc). Enquanto que a terceira área, na Bahia, serve para demonstrar se o modelo tem alguma aplicabilidade em regiões fora do contexto limitador deste estudo.

Diante do exposto, a metodologia de correção empírica foi aplicada nestas três áreas seguindo o fluxograma de trabalho descrito a seguir:

I - Aquisição de 80 pontos de pontos com elevações de controle, particionados em duas amostras: uma com 40 pontos para estimar o erro médio local, valor referencial para a correção do MDE (nomeada de “amostra base”); e a segunda amostra com os demais 40 pontos, para quantificar a acurácia do MDE antes e após a correção (nomeada de “amostra de validação”);

II - Cálculo do “erro médio local” para o AW3D30 não corrigido, usando a primeira amostra;

III - Conversão do raster do MapBiomias em “raster dos coeficientes de correção”, utilizando a fórmula condicional expressa na Figura 21;

IV - Multiplicação do “raster dos coeficientes de correção” pela constante “erro médio local” para obtenção da “superfície de correção”;

V - Alinhamento dos pixels da “superfície de correção” com os pixels do MDE AW3D30;

VI - Aplicação de álgebra de mapa para subtrair o raster “superfície de correção” do raster AW3D30, obtendo, assim, o raster “AW3D30 corrigido”;

VII - Utilização do segundo conjunto amostral de pontos para estimar a acurácia do MDE antes e depois da correção, para fins de validação do método, incluindo a aplicação do PEC-PCD.

4 RESULTADOS E DISCUSSÕES

No presente capítulo apresentar-se-á e discutir-se-á os resultados obtidos, no intento de demonstrar consecução dos objetivos que nortearam esta dissertação. Assim, serão apresentados os indicadores estatísticos da qualidade altimétrica do AW3D30, tanto de um modo geral, quanto correlacionados às classes de uso e cobertura do solo, para as áreas de estudo. Além disso, está apresentada a metodologia empírica de correção do AW3D30 elaborada ao longo deste estudo, possibilitando aos leitores a reprodução da mesma. Sempre que possível, foram elencados resultados de estudos de outros autores, similares a este, de modo a estabelecer uma autocriticidade.

4.1 INDICADORES ESTATÍSTICOS

Nesta seção, os resultados do estudo estatístico serão mostrados. Ressalta-se que a apreensão de todos estes indicadores foi fundamental no estabelecimento da metodologia de correção das altitudes, aplicada no MDE Alos Word 3D - 30m.

4.1.1 Erro altimétrico médio e Raiz do erro quadrático médio (RMSE)

O erro altimétrico e o erro quadrático são indicadores que podem parecer, até certo ponto, similares, porém os dois têm importâncias distintas numa análise de qualidade de produtos cartográficos. Ao longo deste estudo, foi possível observar que enquanto o erro quadrático é significativo para apontar a acurácia geral do produto, o erro médio é mais assertivo para avaliar a distribuição espacial da acurácia, quando a área é classificada em categorias de cobertura do solo, por exemplo. Isso se dá pelo fato de que o erro altimétrico, por mostrar valores negativos e positivos, apresenta melhor as tendências dentro da área analisada, indicando, por exemplo, a presença de possíveis artefatos no modelo quando uma grande quantidade de erros positivos se concentram numa região, ou possíveis “*sinks*” quando há presença de erros negativos. E, no caso desta pesquisa, isso é muito importante, visto que

um dos pressupostos é o viés da vegetação no AW3D30, conseqüentemente, para provar essa hipótese, é necessário identificar a presença de erros positivos concentrados numa região de cobertura florestal. Por outro lado, a raiz do erro médio quadrático mostra melhor a acurácia geral do modelo, pois os erros são tratados em módulo, anulando a possibilidade dos erros negativos dirimirem os positivos no cálculo de média. Isso é explicado pela própria teoria dos erros aleatórios, os quais, possuindo distribuição normal, tendem a anular-se.

Os erros encontrados neste estudo estão esboçados nas Tabelas 02 e 03, as quais expõem os erro médio e RMSE de cada região amostral analisada, sendo a Tabela 02 voltada aos erros “brutos”, ou seja, sem o descarte de *outliers*, e a Tabela 03 mostra os erros “limpos”, ou seja, com os *outliers* eliminados.

Observando a Tabela 02 (com *outliers*) é possível verificar, já de início, a grande discrepância entre as médias de cada região amostral, sendo um indício de que os tratamentos não possuem homogeneidade, o que será tratado com mais rigor no item 4.1.5. Além disso, nota-se que os valores de erro médio e RMSE são bem próximos dentro de cada região amostral do Grupo 01. Isso se deve, obviamente, a uma maior quantidade de erros positivos em relação aos negativos, para as regiões de estudo. Isso já pode ser considerado um indicativo de que artefatos sobre o solo, como cobertura vegetal e infraestruturas humanas, estão causando a superestimação das elevações pelo AW3D30.

O mesmo não ocorre com os valores do Grupo 02, onde o erro médio possui um valor muito baixo em relação ao RMSE. Com base apenas nestes indicadores, cogita-se que essa característica se vincule ao fato dos pontos da RAAP estarem localizados, com maior frequência, em áreas com cobertura vegetal rasteira ou sem nenhuma cobertura vegetal, o que, em teoria, reduziria a quantidade dos erros positivos, deixando a distribuição dos erros “mais normal”. Conseqüentemente, esse maior equilíbrio entre valores de erro positivo e negativo reduz o valor do erro altimétrico médio. Porém, esta tese só poderá ser comprovada, de fato, no item 4.1.6, onde serão analisados a correlação entre os erros e a cobertura do solo.

Tabela 02: Erro médio e RMSE por região amostral, calculados com *outliers*.

Com Outliers					
Grupo amostral	Região amostral	Erro médio	RMSE	QNT de unidades amostrais	Tipo de Unidade
Grupo 01	PE-01	3,29	3,34	201472	PIXELS
	PE-02	4,66	4,81	476110	PIXELS
	SP-01	11,38	11,47	37050	PIXELS
	SP-02	5,80	6,72	32399	PIXELS
Grupo 02	Brasil	0,78	2,29	1662	PONTOS

Fonte: Próprio autor (2022).

No tocante à Tabela 03 (sem *outliers*), percebe-se queda nos valores de RMSE para todas regiões amostrais, enquanto que os valores de erro médio tiveram queda nas amostras do Grupo 1, e um aumento em se tratando da região Brasil (Grupo 02). Este aumento se deve aos limites inferiores para corte de *outliers*, os quais tendem a diminuir o número de unidades amostrais com valores negativos. A consequência disso é que esta quantidade pouca ou nula de erros negativos deixam de anular os erros positivos, terminando por elevar a média após o corte de *outliers*. Isso é aceitável, visto que a intenção aqui é mostrar qual o erro padrão para a região em estudo, e se valores negativos não são representativos para o erro nominal, estes devem ser retirados da amostra. Este fato reforça ainda mais a hipótese de que os artefatos advindos principalmente da cobertura vegetal estão elevando os valores de altitude medidos no AW3D30.

Tabela 03: Erro médio e RMSE por região amostral, calculados sem *outliers*.

Sem Outliers					
		Erro médio	RMSE	QTD de unidades amostrais	Tipo de Unidade
Grupo 01	PE-01	3,22	3,22	193047	PIXELS
	PE-02	4,28	4,28	450642	PIXELS
	SP-01	11,04	11,10	36432	PIXELS
	SP-02	5,66	6,56	32131	PIXELS
Grupo 02	Brasil	0,96	1,95	1590	PONTOS

Fonte: Próprio autor (2022).

Os valores de erro e RMSE encontrados neste trabalho possuem certa variabilidade entre as regiões de estudo. Partindo-se do valor RMSE de 5 metros como a exatidão nominal dos dados AW3D30 (SILVA, SANTIAGO e SILVA 2018), pode-se dizer que quatro das cinco regiões de estudo estão dentro deste limiar, sendo SP-01 o contraponto, com um valor de RMSE bem acima do valor considerado como nominal.

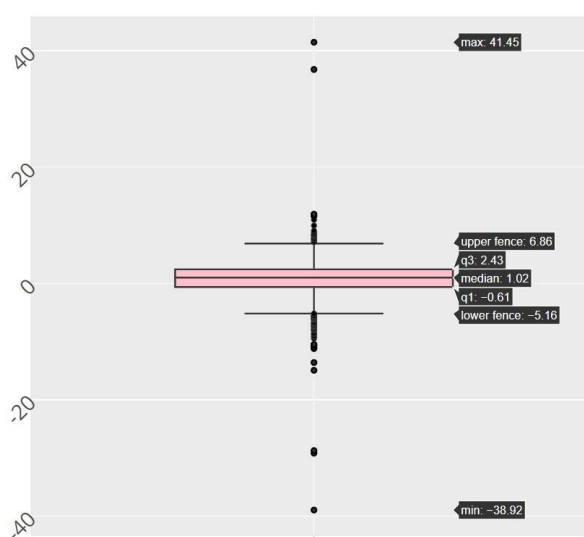
No trabalho de Santillan e Makinano-Santillan (2016) foi encontrado um RMSE de 5,68 metros para a região de Mindanao, nas Filipinas, com 274 pontos de controle. Este limiar condiz bastante com os valores de SP-02, PE-01 e PE-02. Barbosa et al (2019) encontraram

para o AW3D30 um valor de RMSE de 5,34 para região de estudo da Serra de São Domingos, Formoso–MG, valor também dentro de uma faixa de valores próximos aos encontrados neste estudo. Já Silva, Santiago e Silva (2018), em suas análises calcularam um RMSE de 12,45 metros, trabalhando na bacia hidrográfica Ribeirão São Domingos, Goiás, este valor mais se aproxima de SP-01. Diante disso, percebe-se que até mesmo na literatura se identifica certa variabilidade nos valores de RMSE encontrados, mas a maior parte se aproxima da exatidão nominal de 5 metros. Das cinco regiões de estudo deste trabalho, quatro delas (PE-01, PE-02, SP-01 e SP-02) estão com RMSE próximos a valores encontrados na literatura. A região amostral “Brasil” possui valor de RMSE bem abaixo, destacando-se por uma acurácia acima da média.

4.1.2 Análise dos outliers por Box plot

No teste *Box plot* foi possível verificar não apenas os limites para eliminação de *outliers*, mas também como se comporta o conjunto amostral em torno da mediana. O primeiro gráfico apresentado (Figura 07) é o que mostra a situação da região amostral “Brasil”.

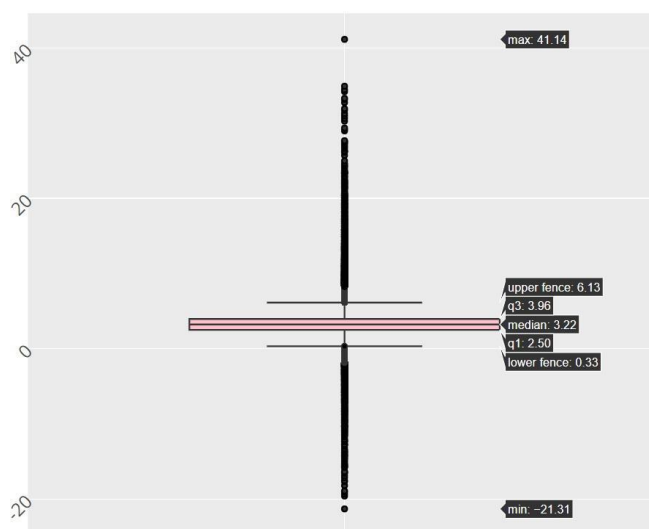
Figura 07: Gráfico Box Plot dos erros - Brasil



Fonte: Próprio autor (2022).

Por meio da análise do *box plot* de “Brasil” nota-se que o limite inferior foi de -5,16, enquanto que o limite superior foi de 6,86 para eliminação de valores anômalos. Por meio deste teste, foram removidos um total de 72 pontos amostrais, sendo 52 (3,13%) abaixo do limite inferior e 20 (1,20%) acima do limite superior. Os *outliers* eliminados representam 4,33% da amostra inicial. O próximo gráfico apresentado (Figura 08) refere-se ao *Box plot* da região amostral PE-01.

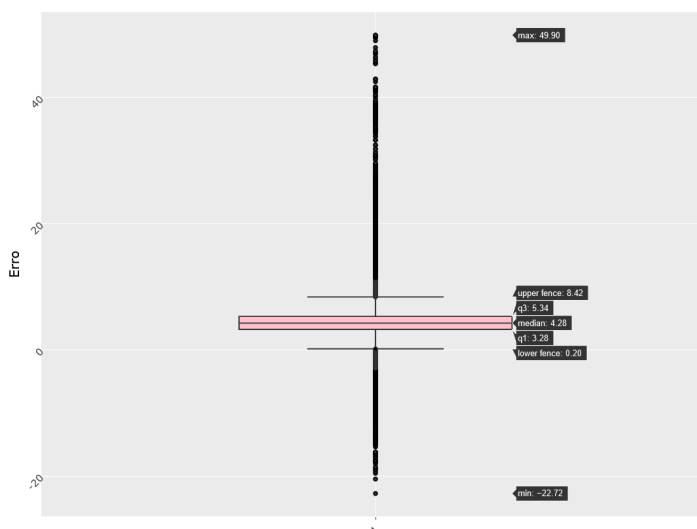
Figura 08: Gráfico Box Plot dos erros - PE-01



Fonte: Próprio autor (2022).

Para PE-01 o teste *box plot* indica um valor de corte inferior de 0,33, enquanto que o corte superior foi de 6,13 para eliminação de valores anômalos. Diante destes valores, foram removidos um total de 8425 pixels da amostra, sendo 2133 (1,06%) abaixo do corte inferior e 6290 (3,12%) acima do corte superior. Os *outliers* eliminados representam 4,18% da amostra inicial. O próximo gráfico apresentado (Figura 09) refere-se à situação da região amostral PE-02.

Figura 09: Gráfico Box Plot dos erros - PE-02

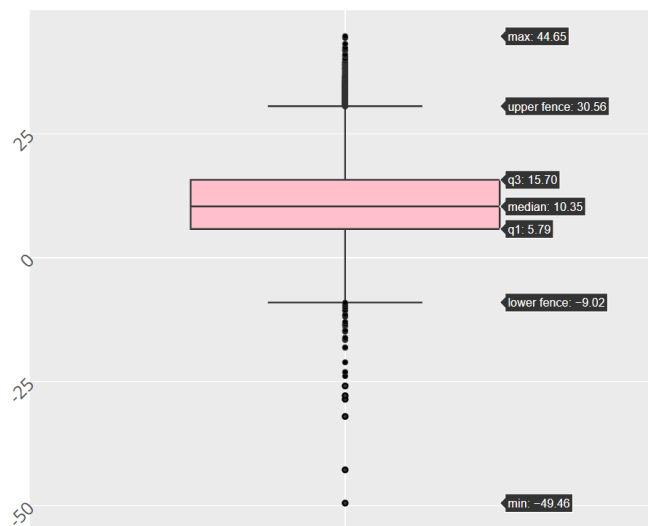


Fonte: Próprio autor (2022).

Por meio da análise do *box plot* de PE-02 nota-se que o limite inferior foi de 0,20, enquanto que o limite superior foi de 8,42 para eliminação de valores anômalos. Por meio deste teste, foram removidos um total de 25.468 pixels amostrais, sendo 6.343 (1,33%) abaixo

do limite inferior e 19.125 (4,02%) acima do limite superior. Os *outliers* eliminados representam 5,35% da amostra inicial. O próximo gráfico apresentado (Figura 10) refere-se ao *box plot* da região amostral SP-01.

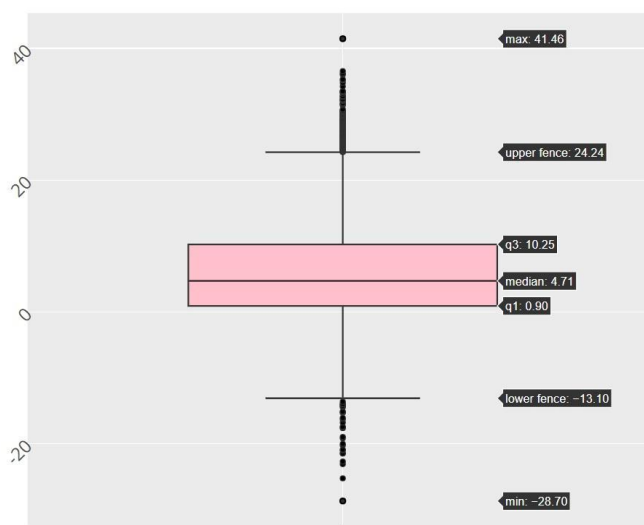
Figura 10: Gráfico Box Plot dos erros - SP-01



Fonte: Próprio autor (2022).

Para SP-01 o teste *box plot* indica um valor de corte inferior de -9,02, enquanto que o corte superior foi de 30,56 para eliminação de valores anômalos. Diante destes valores, foram removidos um total de 618 pixels da amostra, sendo 29 (0,08%) abaixo do corte inferior e 589 (1,59%) acima do corte superior. Os *outliers* eliminados representam 1,67% da amostra inicial. O próximo gráfico apresentado (Figura 11) refere-se à situação da região amostral SP-02.

Figura 11: Gráfico Box Plot dos erros - SP-02



Fonte: Próprio autor (2022).

Por meio da análise do *box plot* de SP-02 nota-se que o limite inferior foi de $-13,10$, enquanto que o limite superior foi de $24,24$ para eliminação de valores anômalos. Por meio deste teste foram removidos um total de 268 pixels amostrais, sendo 27 (0,08%) abaixo do limite inferior e 241 (0,75%) acima do limite superior. Os *outliers* eliminados representam 0,83% da amostra inicial, sendo este o menor percentual de *outliers* entre todas as cinco regiões de estudo.

O teste *box plot* se mostrou importante na eliminação de *outliers*. As porcentagens de *outliers* eliminadas das amostras ficaram abaixo ou muito próxima da especificada na literatura, de até 5% conforme Silva (2020 *apud* DOMINGUES et al, 2018). A retirada destas unidades contaminantes faz com que o conjunto amostral fique mais representativo, mais aderente à normalidade, e que o RMSE seja reduzido, como já mostrado no item 4.1.1. Resultados do estudo de Silva (2020), por exemplo, mostraram que a remoção dos *outliers* por *Box Plot* permitiu que uma das áreas de estudo, que antes não tinha distribuição normal, passasse a ter distribuição gaussiana, e além disso, os RMSE caíram significativamente em todas as áreas de estudo, após serem eliminadas as unidades espúrias. No trabalho de Machado et al (2019), a eliminação dos *outliers* fez com que o RMSE caísse de 14,72 m para 4,32 m, testando o SRTM e usando dados do SIGEF (Sistema de Gestão Fundiária - INCRA) para validação.

4.1.2.1 *Outliers por classe do Map Biomas*

Dada a hipótese da concentração de erros positivos em áreas com cobertura vegetal, neste tópico estão apresentadas tabelas que esboçam as quantidades absolutas e percentuais de *outliers* por classe de uso e cobertura do solo (advindas do Map Biomas) para cada região amostral. Estas tabelas possuem nove colunas, sendo que na primeira há o código da classe de uso e cobertura do Map Biomas, na segunda a quantidade de *outliers* positivos ou acima do limite superior da classe, na terceira a quantidade de *outliers* negativos ou abaixo do limite inferior da classe, na quarta a quantidade total de *outliers* da classe, na quinta a quantidade total de unidades amostrais da classe, na sexta o percentual de *outliers* positivos em relação ao total de unidades amostrais da classe, na sétima o percentual de *outliers* negativos em relação ao total de unidades amostrais da classe, na oitava o percentual total de *outliers* em relação ao total de unidades amostrais da classe, e na nona coluna o percentual de *outliers* da classe em relação a toda amostra. O parâmetro principal para as análises destes dados será a identificação das classes com o limiar acima dos 5% de *outliers*, considerado como aceitável

na literatura. Além disso, sempre elencar-se-á pelo menos três classes com a maior incidência de *outliers*, mesmo que uma ou mais não excedam os 5%.

A primeira tabela deste tópico (Tabela 04) é relativa aos dados da região amostral “Brasil”, e nela se observa 17 diferentes classes de uso e cobertura do solo, a região com ocorrência de mais classes, o que já era esperado dado que essa região amostral possui a maior extensão espacial dentre as cinco amostras.

Tabela 04: Estatísticas por classe acerca dos outliers da região “Brasil”

Classe MapBiomias	Nº de Outliers Positivos	Nº de Outliers Negativos	Nº Total de Outliers	Nº de Unidades amostrais da classe	% de Outliers Positivos em relação á classe	% de Outliers Negativos relação á classe	% total de Outliers em relação a classe	% total de Outliers em relação a toda amostra
3	2	2	4	35	5,71	5,71	11,43	0,24
4	0	4	4	156	0,00	2,56	2,56	0,24
9	1	0	1	1	100,00	0,00	100,00	0,06
11	0	0	0	6	0,00	0,00	0,00	0,00
12	0	6	6	142	0,00	4,23	4,23	0,36
13	0	0	0	1	0,00	0,00	0,00	0,00
15	5	11	16	529	0,95	2,08	3,02	0,96
20	1	0	1	9	11,11	0,00	11,11	0,06
21	2	10	12	239	0,84	4,18	5,02	0,72
23	0	0	0	8	0,00	0,00	0,00	0,00
24	9	14	23	401	2,24	3,49	5,74	1,38
25	0	2	2	68	0,00	2,94	2,94	0,12
29	0	0	0	1	0,00	0,00	0,00	0,00
33	0	3	3	24	0,00	12,50	12,50	0,18
39	0	0	0	24	0,00	0,00	0,00	0,00
41	0	0	0	16	0,00	0,00	0,00	0,00
48	0	0	0	2	0,00	0,00	0,00	0,00
TOTAL	20	52	72	1662				4,33

Fonte: Próprio autor (2022).

Na análise dos dados da região “Brasil” se nota que as três classes com maior ocorrência de *outliers* positivos (ou acima do limite superior de corte) são as classes 3 (Formação Florestal) com 5,71% de incidência, a 9 (Silvicultura) com 100% de ocorrência e a classe 20 (Cana) com 11,11% de incidência. Todas as três classes elencadas estão com porcentagem de *outliers* acima do comum, isso já indica uma certa tendência para erros acima do normal para regiões destas classes. Outro ponto muito importante a ser destacado é que estas três classes são de cobertura vegetal, reforçando a hipótese de que a vegetação enviesada o AW3D30. A próxima tabela (Tabela 05) é relativa aos dados da região amostral “PE-01”, na qual há ocorrência de 8 diferentes classes de uso e cobertura do solo.

Tabela 05: Estatísticas por classe acerca dos outliers da região “PE-01”

Classe MapBiomias	Nº de Outliers limite superior	Nº de Outliers limite inferior	Nº Total de Outliers	Nº de Unidades amostrais da classe	% de Outliers Positivos em relação à classe	% de Outliers Negativos em relação à classe	% total de Outliers em relação a classe	% total de Outliers em relação a toda amostra
3	252	9	261	1238	20,36	0,73	21,08	0,13
4	5426	1984	7410	146767	3,71	1,44	5,05	3,68
12	4	0	4	106	3,77	0,00	3,77	0,00
15	56	7	63	8416	0,68	0,08	0,75	0,03
21	395	108	503	43625	0,91	0,25	1,15	0,25
25	24	8	32	739	3,25	1,08	4,33	0,02
30	6	0	6	55	10,91	0,00	10,91	0,00
33	127	19	146	526	24,10	3,80	27,76	0,07
TOTAL	6290	2135	8425	201472				4,18

Fonte: Próprio autor (2022).

Diante dos dados da região “PE-01” verifica-se que as três classes com maior ocorrência de *outliers* acima do limite superior de corte são as classes 3 (Formação Florestal) com 20,36% de incidência, a 30 (Mineração) com 10,91% de ocorrência e a classe 33 (Rio, Lago e Oceano) com 24,10% de incidência. Todas as três classes elencadas estão com porcentagem de *outliers* acima do comum. Além disso, salienta-se que a classe 3, a de maior ocorrência em *outliers*, caracteriza vegetação arbórea, reforçando a hipótese de que a vegetação enviesada o AW3D30. A próxima tabela (Tabela 06) é relativa aos dados da região amostral “PE-02”, na qual há ocorrência de 8 diferentes classes de uso e cobertura do solo.

Tabela 06: Estatísticas por classe acerca dos outliers da região “PE-02”

Classes Mapiomas	Nº de Outliers Positivos	Nº de Outliers Negativos	Nº Total de Outliers	Nº de Unidades amostrais da Classe	% de Outliers Positivos em relação à classe	% de Outliers Negativos em relação à classe	% Total de Outliers em relação à classe	% Total de Outliers em relação a toda amostra
3	3705	296	4001	8248	44,92	3,59	48,51	0,84
4	15119	4039	19158	270912	5,58	1,49	7,07	4,02
12	7	0	7	72	9,72	0,00	9,72	0,00
15	262	309	571	85418	0,31	0,36	0,67	0,12
21	632	782	1414	104697	0,60	0,75	1,35	0,30
24	38	11	49	3239	1,17	0,34	1,51	0,01
25	118	119	237	3276	3,60	3,63	7,23	0,05
29	0	0	0	7	0,00	0,00	0,00	0,00
33	23	8	31	232	9,91	3,45	13,36	0,01
41	0	0	0	9	0,00	0,00	0,00	0,00
TOTAL	19904	5564	25468	476110				5,35

Fonte: Próprio autor (2022).

Ao se verificar os dados da região “PE-02” foram identificadas quatro classes com porcentagem de *outliers* positivos (ou acima do limite superior de corte) acima dos 5%. Estas classes são a 3 (Formação Florestal) com 44,92% de incidência, a 4 (Formação savânica) com 5,58% de ocorrência, a 12 (Formação Campestre) com 9,72% de ocorrência e a classe 33

(Rio, Lago e Oceano) com 9,91% de incidência. Além disso, salienta-se que as classes 3, 4 e 12 são de cobertura vegetal, ressaltando mais uma vez a tendência de viés no AW3D30 em áreas com vegetação de médio e grande porte. Já a classe 33 aparece mais uma vez com alta incidência de *outliers*, e o que pode explicar esse fato é o complexo comportamento espectral da superfície da água. A próxima tabela (Tabela 07) é relativa aos dados da região amostral “SP-01”, na qual há ocorrência de 7 diferentes classes de uso e cobertura do solo.

Tabela 07: Estatísticas por classe acerca dos outliers da região “SP-01”

Classe MapBiomas	Nº de Outliers Positivos	Nº de Outliers Negativos	Nº Total de Outliers	Nº de Unidades amostrais da classe	% de Outliers Positivos em relação á classe	% de Outliers Negativos relação á classe	% total de Outliers em relação a classe	% total de Outliers em relação a toda amostra
3	589	22	611	28564	2,06	0,08	2,14	1,65
11	0	0	0	66	0,00	0,00	0,00	0,00
15	0	5	5	1225	0,00	0,41	0,41	0,01
21	0	2	2	6556	0,00	0,03	0,03	0,01
24	0	0	0	562	0,00	0,00	0,00	0,00
25	0	0	0	27	0,00	0,00	0,00	0,00
33	0	0	0	50	0,00	0,00	0,00	0,00
TOTAL	589	29	618	37050				1,67

Fonte: Próprio autor (2022).

Perante os dados da região “SP-01” pode ser notado que apenas uma classe das sete apresenta ocorrência de *outliers* positivos (ou acima do limite superior de corte) que é a classe 3 (Formação Florestal), com 2,06% de incidência. Apesar disso, está com um limiar abaixo dos 5%. Destaca-se que, mais uma vez, a classe 3, caracterizada como vegetação arbórea, aparece neste rol de classes com alta incidência de *outliers* positivos. A próxima tabela (Tabela 08) é relativa aos dados da região amostral “SP-02”, na qual há ocorrência de 6 diferentes classes de uso e cobertura do solo.

Tabela 08: Estatísticas por classe acerca dos outliers da região “SP-02”

Classe MapBiomas	Nº de Outliers Positivos	Nº de Outliers Negativos	Nº Total de Outliers	Nº de Unidades amostrais da classe	% de Outliers Positivos em relação á classe	% de Outliers Negativos relação á classe	% total de Outliers em relação a classe	% total de Outliers em relação a toda amostra
3	237	18	255	19661	1,21	0,09	1,30	0,79
15	0	0	0	1379	0,00	0,00	0,00	0,00
21	4	9	13	7737	0,05	0,12	0,17	0,04
24	0	0	0	3503	0,00	0,00	0,00	0,00
25	0	0	0	2	0,00	0,00	0,00	0,00
33	0	0	0	117	0,00	0,00	0,00	0,00
TOTAL	241	27	268	32399				0,83

Fonte: Próprio autor (2022).

Ao se verificar os dados da região “SP-02” percebe-se que só duas classes apresentam ocorrência de *outliers* positivos (ou acima do limite superior de corte): as classes 3 (Formação Florestal) com 1,21% de incidência e 21 (Mosaico de agricultura e pastagem) com 0,05% de ocorrência. Nenhuma das duas apresenta porcentagem de *outliers* acima dos 5%. Convém

mencionar que a classe 3, caracterizada por vegetação do tipo arbórea, apareceu em todas as áreas amostrais deste estudo e sempre teve alta incidência de *outliers* acima do limite superior. Quanto a classe 21, que aparece em SP-02, é uma tipologia de cobertura vegetal do MapBiomias, e este resultado também corrobora com a pressuposto de que artefatos vegetais afetam a acurácia altimétrica do AW3D30. A Tabela 09 traz um resumo de tudo que foi abordado neste tópico, mostrando as médias de *outliers*, por classe do Map Biomias, encontradas nas cinco regiões de estudo deste trabalho.

Tabela 09: *Outliers* médios por classe de uso e cobertura do solo

Classe MapBiomias	% Média de Outliers > Limite Superior	% Média de Outliers < Limite inferior
3	14,85	2,04
4	1,86	1,10
9	20,00	0,00
11	0,00	0,00
12	2,70	0,85
13	0,00	0,00
15	0,39	0,59
20	2,22	0,00
21	0,48	1,07
23	0,00	0,00
24	0,68	0,77
25	1,37	1,53
29	0,00	0,00
30	2,18	0,00
33	6,80	3,95
39	0,00	0,00
41	0,00	0,00
48	0,00	0,00

Fonte: Próprio autor (2022).

Perante esta tabela nota-se claramente que as classes com maior ocorrência de outliers positivos (ou acima do limite superior de corte) trata-se das classes 9 (Silvicultura) com 20%, a classe 3 (Formação florestal) com 14,85% e a classe 33 (Rio, Lago e oceano) com 6,80%. Duas destas três classes são tipologias de vegetação, um forte indicativo da interferência da vegetação densa na acurácia altimétrica do AW3D30.

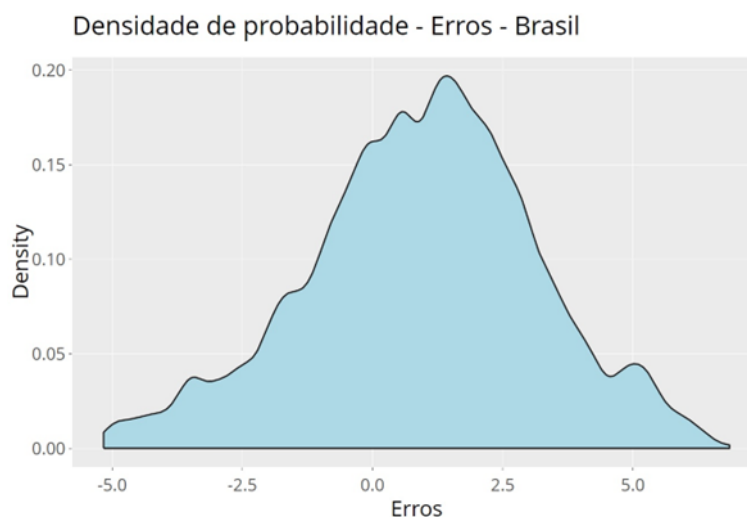
4.1.3 Normalidade do conjunto erros

Os testes de normalidade aqui realizados, baseados na plotagem de gráfico densidade de probabilidade e no teste Shapiro-Wilk foram relevantes nas análises de qualidade do AW3D30. A seguir serão mostrados tais gráficos, bem como os resultados do teste Shapiro-Wilk, para se comprovar ou não a normalidade dos conjuntos de erro por região amostral. Ressalta-se que os testes foram realizados após a remoção dos *outliers*.

a) Região amostral “Brasil” - Grupo 02

É possível verificar pela análise visual do gráfico de densidade de probabilidades gerado (Figura 12), que a curva tem algumas características de uma curva gaussiana, como, por exemplo, certa simetria em relação ao centro da figura (forma de sino). Entretanto, sendo um tanto quanto subjetiva este tipo de análise visual, foi aplicado teste de Shapiro-Wilk na amostra, a um nível de confiança de 95%, resultando nos valores apresentados no quadro 03.

Figura 12: Gráfico de densidade de probabilidade dos erros para a região de estudo “Brasil”



Fonte: Próprio autor (2022)

Quadro 03: Estatísticas do Teste Shapiro-Wilk para a região de estudo “Brasil”

Teste Shapiro-Wilk – Brasil			
$W\alpha$	W	<i>p-value</i>	Hipótese H0
0,9470	0,9663	0,1632	aceita
Nível de confiança = 95%			
Nível de significância = 0,05			
Nº de unidades amostrais = 50			

Fonte: Próprio autor (2022)

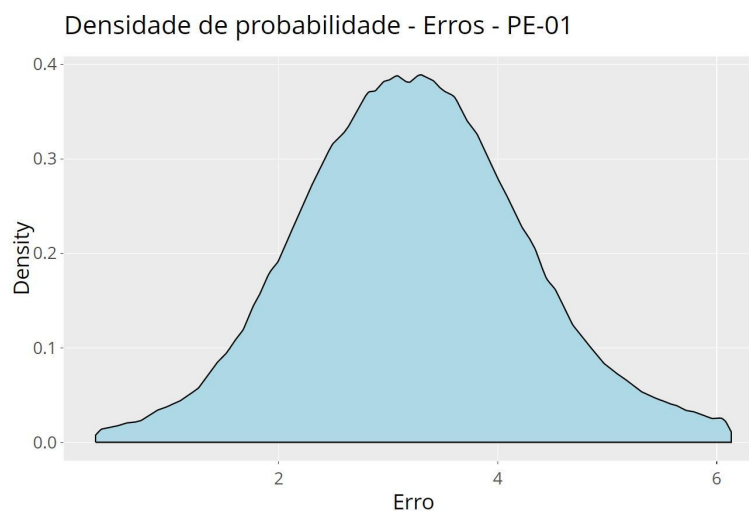
Considerando a amostra com cinquenta unidades e um nível de significância de 5%, o valor $W\alpha$ tabelado é de 0,947. Sendo o valor calculado de W igual a 0,9663, maior que $W\alpha$, e, ainda, o *p-value* encontrado maior que o nível de significância de 0,05, aceita-se a hipótese nula a um nível de confiança de 95%, concluindo-se, portanto, que a amostra é normalmente distribuída.

b) Região amostral PE-01 - Grupo 01

Se verifica pela análise visual do gráfico de densidade de probabilidades (Figura 13) gerado, que a curva tem características de uma gaussiana, como a simetria em relação ao centro da figura (forma de sino), e proximidade entre valores de média, moda e mediana, constituindo-se o gráfico mais próximo (visualmente) de uma amostra perfeitamente normal,

dentre as cinco regiões amostrais deste trabalho. Ainda assim, esta análise visual é subjetiva, por isso foi aplicado teste de Shapiro-Wilk na amostra, a um nível de confiança de 95%, resultando nos valores apresentados no quadro 04.

Figura 13: Gráfico de densidade de probabilidade dos erros para a região de estudo “PE-01”



Fonte: Próprio autor (2022)

Quadro 04: Estatísticas do Teste Shapiro-Wilk para a região de estudo PE-01

Teste Shapiro-Wilk – PE-01			
$W\alpha$	W	<i>p-value</i>	Hipótese H0
0,9470	0,9841	0,7331	aceita
Nível de confiança = 95%			
Nível de significância = 0,05			
Nº de unidades amostrais = 50			

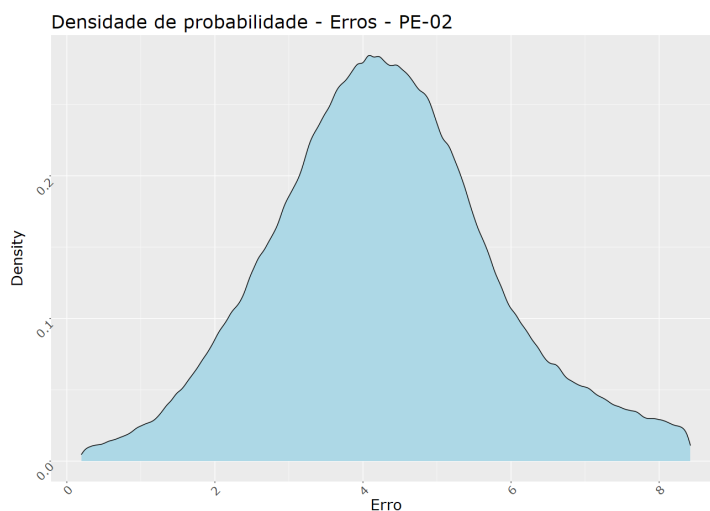
Fonte: Próprio autor (2022)

Considerando a amostra com cinquenta unidades e um nível de significância de 5%, o valor $W\alpha$ tabelado é de 0,947. Sendo o valor calculado de W igual a 0,9841, maior que $W\alpha$. E, ainda, o *p-value* encontrado maior que o nível de significância de 0,05, aceita-se a hipótese nula a um nível de confiança de 95%, concluindo-se, portanto, que a variável erros altimétricos da amostra PE-01 é normalmente distribuída.

c) Região amostral PE-02 - Grupo 01

Pode-se verificar pela análise visual do gráfico de densidade de probabilidades (Figura 14) gerado, que a curva tem características de uma gaussiana, como certa simetria em relação ao centro da figura (forma de sino), e proximidade entre valores de média, moda e mediana. Entretanto, sendo um tanto quanto subjetiva este tipo de análise visual, foi aplicado teste de Shapiro-Wilk na amostra, a um nível de confiança de 95%, resultando nos valores apresentados no quadro 05.

Figura 14: Gráfico de densidade de probabilidade dos erros para a região de estudo “PE-02”



Fonte: Próprio autor (2022)

Quadro 05: Estatísticas do Teste Shapiro-Wilk para a região de estudo PE-02

Teste Shapiro-Wilk – PE-02			
W_α	W	<i>p-value</i>	Hipótese H0
0,9470	0,9600	0,0890	aceita
Nível de confiança = 95%			
Nível de significância = 0,05			
Nº de unidades amostrais = 50			

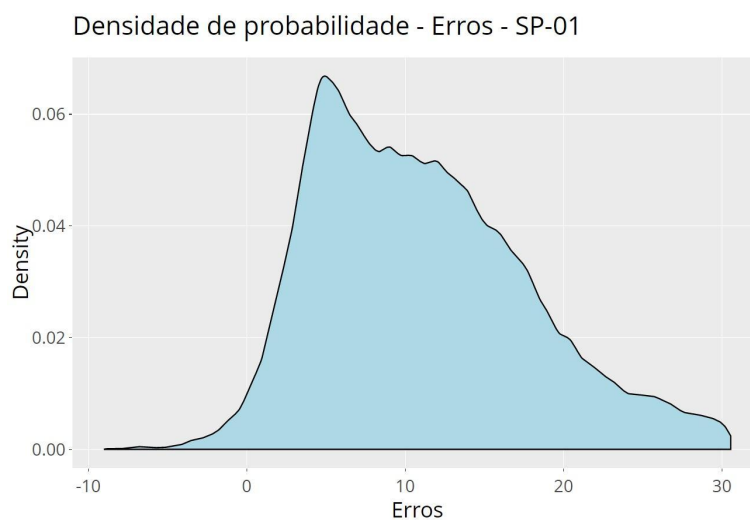
Fonte: Próprio autor (2022)

Considerando um W_α tabelado de 0,947, referente a uma amostra com cinquenta unidades e um nível de significância de 5%, tem-se um W calculado maior que W_α , sendo seu valor igual a 0,9600. E, ainda, o *p-value* encontrado maior que o nível de significância de 0,05, aceita-se, portanto, a hipótese nula a um nível de confiança de 95%, concluindo-se que a variável erros altimétricos da amostra PE-02 é normalmente distribuída.

d) Região amostral SP-01 - Grupo 01

Fazendo-se análise visual do gráfico de densidade de probabilidades (Figura 15) gerado, não é possível afirmar sobre a normalidade, visto que falta simetria em relação ao centro do gráfico, apesar do lado direito da figura ter forma de sino. Porém, dada a subjetividade deste tipo de análise visual, foi aplicado teste de Shapiro-Wilk na amostra, a um nível de confiança de 95%, resultando nos valores apresentados no quadro 06.

Figura 15: Gráfico de densidade de probabilidade dos erros para a região de estudo “SP-01”



Fonte: Próprio autor (2022)

Quadro 06: Estatísticas do Teste Shapiro-Wilk para a região de estudo SP-01

Teste Shapiro-Wilk – SP-01			
$W\alpha$	W	<i>p-value</i>	Hipótese H0
0,9470	0,9635	0,1251	aceita
Nível de confiança = 95%			
Nível de significância = 0,05			
Nº de unidades amostrais = 50			

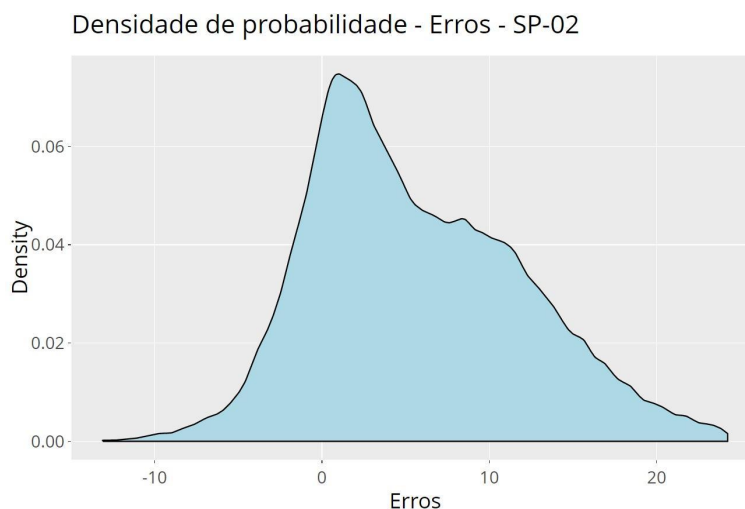
Fonte: Próprio autor (2022)

Considerando a amostra com cinquenta unidades e um nível de significância de 5%, o valor $W\alpha$ tabelado é de 0,947. Sendo o valor calculado de W igual a 0,9635, maior que $W\alpha$. E, ainda, o *p-value* encontrado maior que o nível de significância de 0,05, aceita-se a hipótese nula a um nível de confiança de 95%, concluindo-se, portanto, que a variável erros altimétricos da amostra SP-01 é normalmente distribuída.

e) Região amostral SP-02 - Grupo 01

É possível verificar pela análise visual do gráfico de densidade de probabilidades (Figura 16), que a curva não possui características de uma curva gaussiana, dada a assimetria em relação ao centro da figura, com um pico muito acentuado fora do centro do gráfico. Entretanto, sendo um tanto quanto subjetiva este tipo de análise visual, foi aplicado teste de Shapiro-Wilk na amostra, a um nível de confiança de 95%, resultando nos valores apresentados no quadro 07.

Figura 16: Gráfico de densidade de probabilidade dos erros para a região de estudo “SP-02”



Fonte: Próprio autor (2022)

Quadro 07: Estatísticas do Teste Shapiro-Wilk para a região de estudo PE-01

Teste Shapiro-Wilk – SP-02			
$W\alpha$	W	<i>p-value</i>	Hipótese H0
0,9470	0,9470	0,0258	rejeitada
Nível de confiança = 95%			
Nível de significância = 0,05			
Nº de unidades amostrais = 50			

Fonte: Próprio autor (2022)

Considerando a amostra com cinquenta unidades e um nível de significância de 5%, calculou-se um valor W de 0,947. Sendo então este valor igual ao $W\alpha$ tabelado de 0,947, e o p-value encontrado menor que o nível de significância de 0,05, o resultado do teste mostra que a hipótese nula foi rejeitada ao nível de confiança de 95%, concluindo-se que a amostra SP-02 não é normalmente distribuída.

As distribuições de PE-01, PE-02 e Brasil tiveram gráficos de distribuição das probabilidades mais similares ao formato de sino, sendo todos estes conjuntos considerados normais com base no teste de Shapiro-Wilk. Quanto às amostras do estado de São Paulo, pode ser notada certa similaridade entre os gráficos das duas, sendo marcante um pico modal à esquerda do centro da curva. Porém, ainda assim, o teste Shapiro-Wilk mostrou que SP-01 é considerada normal, o mesmo não ocorrendo para SP-02, visto que este último teve um gráfico muito destoante do que se espera para uma distribuição Gaussiana.

O estudo sobre a normalidade da distribuição dos erros das cinco regiões amostrais foi bem proveitoso, dado que foi possível visualizar como estão distribuídos os erros e onde se centra o valor mais representativo quanto a acurácia do AW3D30. A observação das

centralidades dos cinco gráficos indica uma tendência positiva dos valores de erros, dado que a maior parte da curva fica na parte positiva do plano cartesiano. Esta constatação é mais um indicativo que há uma tendência para superestimação das elevações pelo AW3D30, devido a artefatos sobre o solo, muito provavelmente a cobertura vegetal e infraestruturas humanas.

4.1.4 Teste de média para averiguar homogeneidade das 5 regiões amostrais

Neste tópico serão apresentados o resultado da ANOVA e teste de Tukey para verificar se as regiões amostrais deste estudo podem ser consideradas de uma mesma população. Ressalta-se que o teste de média aplicado ajuda a averiguar se os resultados de erro médio e RMSE de cada região amostral podem ser considerados estatisticamente de um mesmo grupo populacional. Na tabela 10 a seguir são mostrados os resultados do teste de Tukey para erros por região amostral. Quando na coluna “Resultados do Teste” a letra “a” estiver acompanhada de um mesmo número significa que as médias não diferem entre si pelo teste de Tukey a 1% de probabilidade.

Tabela 10: Resultados do teste de Tukey para média dos erros por região amostral.

Tratamentos (Regiões amostrais)	Erro médio	Resultado do Teste
Brasil	0,96	a1
PE-01	3,21	a2
PE-02	4,30	a3
SP-02	6,02	a4
SP-01	10,87	a5
Média harmônica do número de repetições (r): 500		
Erro padrão: 0,134		
Conclusão: Significativo a 1% de probabilidade conforme ANOVA		

Fonte: Próprio autor (2022)

Os dados da tabela acima demonstram que as médias de erro das regiões amostrais deste estudo não pertencem a uma mesma população, conforme análise de variância e teste de Tukey. Este fato demonstra que o AW3D30 não possui uma homogeneidade quanto à acurácia em toda sua cobertura espacial, indicando que o tratamento dado a possíveis correções deve ser realizado com ressalvas, e avaliando as especificidades locais. A seguir mostra-se uma segunda tabela (Tabela 11) com resultados do teste de Tukey, só que agora a variável analisada foi o RMSE. Quando na coluna “Resultados do Teste” a letra “a” estiver acompanhada de um mesmo número significa que as médias não diferem entre si pelo teste de Tukey a 1% de probabilidade.

Tabela 11: Resultados do teste de Tukey para média dos erros quadráticos (RMSE) por região amostral.

Tratamentos (Regiões amostrais)	RMSE	Resultado do Teste
Brasil	1,97	a1
PE-02	2,06	a1
PE-01	3,80	a2
SP-02	6,45	a3
SP-01	11,31	a4
Média harmônica do número de repetições (r): 500		
Erro padrão: 0,118		
Conclusão: Significativo a 1% de probabilidade conforme ANOVA		

Fonte: Próprio autor (2022)

Para a variável RMSE, a ANOVA com teste de Tukey demonstrou diferenças significativas entre as regiões amostrais deste estudo, seguindo o mesmo padrão da variável erro médio.

4.1.5. Teste de média para averiguar correlação entre o porte de vegetação e os valores de erro

Para verificar se há correlação entre as variáveis erro e porte vegetativo, serão apresentados os resultados de uma análise de variância com teste de Tukey aplicados a um conjunto de dados que relaciona os erros com os tipos de uso e cobertura do solo identificados em cada unidade amostral. O teste de Tukey poderá apontar se há diferença significativa entre as médias de erro por categorias do MapBiomas, a qual inclui tipologias de cobertura vegetal, presumindo-se que cada uma destas categorias terão uma média diferente, ou que pelo menos categorias cuja vegetação é arbórea tenham um valor de erro médio estatisticamente maior que as demais categorias. A análise dos dados se centra em averiguar as três categorias do MapBiomas de maior erro médio (positivo) e ainda se há diferença estatística destas categorias para as demais. A seguir, a Tabela 12 apresenta os dados do teste de Tukey para a região amostral “Brasil”. Quando na coluna “Resultados do Teste” a letra “a” estiver acompanhada de um mesmo número significa que as médias não diferem entre si pelo teste de Tukey a 5% de probabilidade.

Analisando-se os dados da Tabela 12 pode ser observado que a maior parte das categorias de uso e cobertura do solo têm médias de erro consideradas estatisticamente iguais pelo teste de Tukey. Porém destaca-se a classe 3 (Formação florestal) a qual possui o maior erro altimétrico médio (2,39 m) ao tempo que também tem média destacada das demais, conforme teste de Tukey. As classes subsequentes com maiores valores de erro altimétrico são a 41 (Outras lavouras temporárias) com 1,72 metros de erro médio, e a classe 24 (área urbanizada) com 1,52 metros. Apesar de estarem elencadas entre as três de maior erro médio,

estas duas últimas categorias mencionadas são consideradas estatisticamente iguais à maioria das classes ocorridas na área de estudo, conforme resultado do teste. Na Tabela 13 estão apresentados os dados do teste de Tukey para a região amostral “PE-01”. Quando na coluna “Resultados do Teste” a letra “a” estiver acompanhada de um mesmo número significa que as médias não diferem entre si pelo teste de Tukey a 5% de probabilidade.

Tabela 12: Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral “Brasil”

Tratamentos (Classes Map Biomas)	Erro médio	Resultado do Teste
29	-2,67	a1
48	-0,68	a1 a2
33	-0,63	a1 a2
11	-0,51	a1 a2
13	-0,39	a1 a2
12	0,15	a1 a2
15	0,55	a1 a2
39	0,76	a1 a2
23	0,90	a1 a2
20	0,96	a1 a2
25	1,04	a1 a2
21	1,12	a1 a2
4	1,49	a1 a2
24	1,52	a1 a2
41	1,72	a1 a2
3	2,39	a2

Média harmônica do número de repetições (r): 4,896
 Erro padrão: 0,972
 Conclusão: Significativo a 5% de probabilidade conforme ANOVA

Fonte: Próprio autor (2022)

Tabela 13: Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral “PE-01”

Tratamentos (Classes Map Biomas)	Erro médio	Resultado do Teste
25	3,01	a1
12	3,07	a1
21	3,21	a1
15	3,23	a1 a2
4	3,29	a1 a2
30	3,52	a2 a3
3	3,74	a3 a4
33	3,89	a4

Média harmônica do número de repetições (r): 225,786
 Erro padrão: 0,069
 Conclusão: Significativo a 5% de probabilidade conforme ANOVA

Fonte: Próprio autor (2022)

Observando os dados da Tabela 13 acima, destaca-se a classe 33 (Rio, Lago e Oceano) com erro altimétrico médio de 3,89 metros, ao tempo que também tem média destacada das demais, conforme teste de Tukey. As classes subsequentes com maiores valores de erro altimétrico são: a classe 3 (Formação florestal) com 3,74 metros de erro médio, e a classe 30 (Mineração) com 3,52 metros. Estas duas últimas categorias mencionadas também são

consideradas estatisticamente destacadas das demais classes ocorridas na área de estudo, conforme resultado do teste. Cabe ressaltar que a classe 3, aparece novamente entre as classes de maiores erros médio, reforçando a ideia de que o dossel arbóreo impacta a qualidade altimétrica do AW3D30. A seguir a Tabela 14 apresenta os dados do teste de Tukey para a região amostral “PE-02”. O significado da codificação da coluna “Resultado do Teste” é o mesmo da tabela anterior

Tabela 14: Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral “PE-02”

Tratamentos (Classes Map Biomas)	Erro médio	Resultado do Teste
41	3,13	a1
25	3,78	a1 a2
21	3,83	a1 a2 a3
12	3,83	a1 a2 a3
25	3,83	a1 a2 a3
33	4,33	a2 a3
24	4,55	a2 a3
4	4,63	a2 a3
29	4,72	a3
3	5,86	a4

Média harmônica do número de repetições (r): 44,654
 Erro padrão: 0,210
 Conclusão: Significativo a 5% de probabilidade conforme ANOVA

Fonte: Próprio autor (2022)

Diante dos dados da tabela acima, observa-se que a classe 3 (Formação florestal) possui o maior erro altimétrico médio (5,86 m) ao tempo que também tem média destacada das demais. Em sequência se tem a classe 29 (afloramento rochoso) com erro altimétrico médio de 4,72 e depois a classe 4 (Formação savânica) com 4,63 metros de erro médio, sendo que estas duas possuem média estatisticamente igual a maioria das classes ocorridas na área de estudo conforme teste de Tukey. Nota-se, mais uma vez, que a classe 3 desponta com o maior erro médio dentre as classes da área de estudo, e ainda com diferença significativa das demais. Não se pode deixar de destacar também que a classe 4 (Formação savânica), de característica arbórea, está entre as classes de maior erro. A seguir a Tabela 15 apresenta os dados do teste de Tukey para a região amostral “SP-01”. O significado da codificação da coluna “Resultado do Teste” é o mesmo das tabelas anteriores.

Tabela 15: Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral “SP-01”

Tratamentos (Classes Map Biomas)	Erro médio	Resultado do Teste
25	1,93	a1
11	2,91	a1 a2
15	3,45	a1 a2
24	4,83	a2
21	5,06	a2
33	8,61	a3
3	12,93	a4
Média harmônica do número de repetições (r): 91,783		
Erro padrão: 0,604		
Conclusão: Significativo a 5% de probabilidade conforme ANOVA		

Fonte: Próprio autor (2022)

Analisando-se os dados da tabela acima pode ser observado que a classe 3 (Formação florestal) possui o maior erro altimétrico médio (12,93 m) ao tempo que também tem média estatisticamente diferente das demais, conforme teste de Tukey. As classes subsequentes com maiores valores de erro altimétrico são a 33 (Rio, Lago e Oceano) com 8,61 metros de erro médio, e a classe 21 (Mosaico de agricultura e pastagem) com 5,06 metros. Estas duas últimas categorias mencionadas tem média de erros estatisticamente diferentes entre si, conforme resultado do teste. Destaca-se, mais uma vez, a classe 3 possuindo a maior média de erros.

A seguir a Tabela 16 apresenta os dados do teste de Tukey para a região amostral “SP-02”. Para esta região amostral ressalva-se que a ANOVA com teste de Tukey não são recomendados, devido ao fato da variável erro médio não possuir distribuição normal. Porém, para fins analíticos deste estudo, foi realizado mesmo assim a ANOVA, para se extrair alguns parâmetros que se julga indispensáveis nesta pesquisa, como a classe de maior erro, por exemplo. O significado da codificação da coluna “Resultado do Teste” é o mesmo das tabelas anteriores.

Tabela 16: Resultados do teste de Tukey para média dos erros por classe do Map Biomas para a região amostral “SP-02”

Tratamentos (Classes Map Biomas)	Erro médio	Resultado do Teste
15	-0,38	a1
21	1,01	a1
33	1,62	a1
24	1,72	a1
25	2,07	a1
3	8,66	a2
Média harmônica do número de repetições (r): 11,771		
Erro padrão: 1,496		
Conclusão: Significativo a 5% de probabilidade conforme ANOVA		

Fonte: Próprio autor (2022)

Analisando-se os dados da tabela acima pode ser observado que quase todas categorias de uso e cobertura do solo têm médias de erro consideradas estatisticamente iguais pelo teste de Tukey, menos a classe 3 (Formação florestal). A mesma possui o maior erro altimétrico médio (8,66 m) desta região amostral, além disso tem média diferente das demais, conforme teste de Tukey. As classes subsequentes com maiores valores de erro altimétrico são: a 25 (Outras áreas não vegetadas) com 2,07 metros de erro médio, e a classe 24 (área urbanizada) com 1,72 metros. Salienta-se, mais uma vez, o fato da classe 3 despontar como a de maior erro altimétrico.

No contexto geral dos dados expostos neste tópico é perceptível que a classe Formação Florestal teve um erro médio elencado sempre entre os maiores, em todas as cinco regiões amostrais deste estudo. Diante disso, pode ser afirmado que a ocorrência desta categoria do MapBiomias tem forte correlação com os erros contidos no AW3D30, e, obviamente, estes erros acima da média se vinculam aos deslocamentos causados pela cobertura vegetal durante o imageamento pelo satélite. Assim, esta constatação é um pressuposto importante na execução de futuras correções do modelo. Não se pode deixar de mencionar também o fato da classe 33 (Rio, Lago e Oceano) ter aparecido bastante entre as classes de maiores erros neste estudo, e isto se deve ao fato da superfície da água possuir uma assinatura espectral inconstante, gerando confusões durante os processos matemáticos de conversão de valores digitais das imagens em elevações.

4.1.6 Erro médio por classe

Com base nos dados de erro médio por classe, em cada região amostral, se construiu ainda uma tabela mostrando o erro médio geral por classe, ou seja, tirando a média das regiões amostrais, como pode ser observado na Tabela 17.

Tendo como foco as classes que ocorreram em pelo menos 3 regiões amostrais, as 5 classes do MapBiomias com maiores erros identificados foram "Formação Florestal", "Rio, Lago e Oceano", "Formação Savânica", "Área urbanizada" e "Mosaico de agricultura e pastagem". Diante destes dados, temos mais uma evidência de que o MDE tem sua qualidade impactada em regiões de vegetação densa e em áreas construídas, visto que das cinco classes de uso e cobertura citadas acima, três delas são relativas a cobertura vegetal (Classes 3, 4 e 21) e uma à área com alta densidade de construções civis (Classe 24). A única classe que destoa dessas características é a 33, que se trata de superfície de água.

Tabela 17: Erro médio por classe do MapBiomias nas áreas de estudo.

Classes do MapBiomias		Erro por Área de estudo					Erro Médio	Ocorrência em pelo menos 3 áreas de estudo
LEGENDA	CÓDIGO	BRASIL	PE-1	PE-2	SP-1	SP-2		
Formação Florestal	3	2,39	3,74	5,81	12,93	8,66	6,71	X
Mineração	30		3,52				3,52	
Rio, Lago e Oceano	33	-0,63	3,89	4,10	8,61	1,62	3,52	X
Formação Savânica	4	1,49	3,29	4,56			3,11	X
Área Urbanizada	24	1,52		4,23	4,83	1,72	3,08	X
Mosaico de Agricultura e Pastagem	21	1,12	3,21	3,89	5,06	1,01	2,86	X
Outras Lavouras Temporárias	41	1,72		3,13			2,42	
Formação Campestre	12	0,15	3,07	4,02			2,42	X
Outras Áreas não Vegetadas	25	1,04	3,01	3,99	1,93	2,07	2,41	X
Pastagem	15	0,55	3,23	3,85	3,45	-0,38	2,14	X
Campo Alagado e Área Pantanosa	11	-0,51			2,91		1,20	
Afloramento Rochoso	29	-2,67		4,72			1,03	
Cana	20	0,96					0,96	
Praia, Duna e Areal	23	0,90					0,90	
Soja	39	0,76					0,76	
Outras Formações não Florestais	13	-0,39					-0,39	
Outras Lavouras Perenes	48	-0,68					-0,68	

Fonte: Próprio autor (2023).

4.1.7 Número mínimo de unidades amostrais para estimar o erro

Em função da necessidade de se ter uma amostragem mínima para estimar o erro médio do local onde se aplica a metodologia de correção do AW3D30, proposta neste estudo, analisou-se a evolução do erro médio de 11 conjuntos amostrais contendo entre 10 e 110 unidades. Esta análise foi feita para três classes do MapBiomias: “Mosaico de agricultura e pastagem” (21), “Pastagem” (15), e “Formação Florestal” (3). A seleção destas classes seguiu dois critérios: o primeiro é a ocorrência em todas áreas de estudo, o segundo consiste nas maiores quantidades de unidades amostrais (pixels ou pontos) somando todas as áreas de estudo. A análise destes critérios está representada na Tabela 18, a seguir.

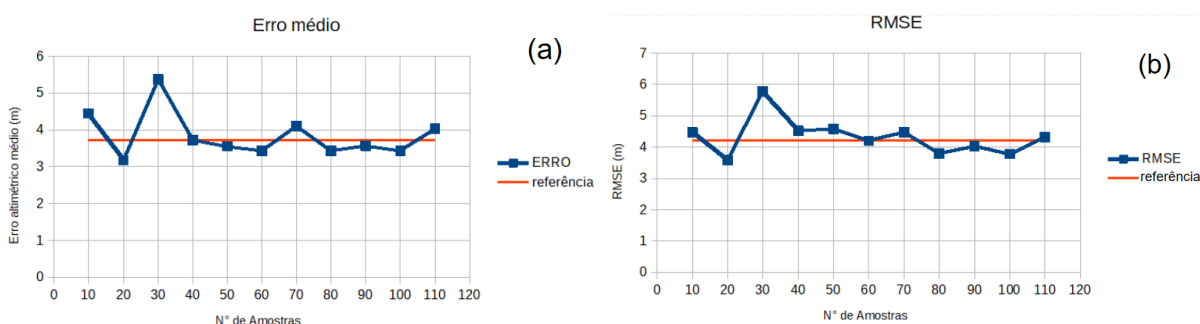
Tabela 18: Quantidade de unidades amostrais por área de estudo.

ID da Classe MapBiomias	Unidades amostrais por Áreas de Estudo					Total de unidades amostrais	% em relação ao total	Ocorre em todas as áreas?
	BRASIL	PE-01	PE-02	SP-01	SP-02			
4	155	146.894	255.973	0	0	403.022	55,47	
21	238	43.626	103.545	6.556	7.212	161.177	22,18	SIM
15	529	8.416	84.953	1.225	1.302	96.425	13,27	SIM
3	35	1.238	4.755	28.564	18.561	53.153	7,32	SIM
24	400	0	3.208	562	3.280	7.450	1,03	
25	68	739	3.068	27	2	3.904	0,54	SIM
33	24	527	202	50	124	927	0,13	SIM
12	142	106	66	0	0	314	0,04	
11	6	0		66	0	72	0,01	
30	0	55		0	0	55	0,01	
41	16	0	9	0	0	25	0,00	
39	24	0		0	0	24	0,00	
29	1	0	7	0	0	8	0,00	
20	9	0		0	0	9	0,00	
23	8	0		0	0	8	0,00	
48	2	0		0	0	2	0,00	
9	1	0		0	0	1	0,00	
13	1	0		0	0	1	0,00	
TOTAL	1.658	201.601	455.786	37.050	30.481	726.576	100,00	

Fonte: Próprio autor (2023).

Assim, após a escolha das classes que iriam ser testadas, foi realizada uma seleção aleatória das amostras dentro dos conjuntos de dados por classe. E, com tais dados, foram construídos os gráficos mostrados a seguir, sendo a Figura 17 relativa aos gráficos da classe 21, a Figura 18 relativa aos gráficos da classe 15, e a Figura 19 mostra os gráficos da classe 3.

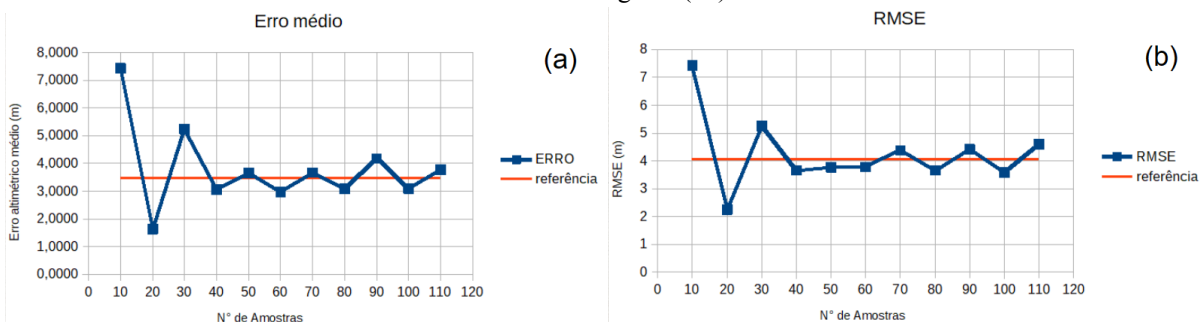
Figura 17: Gráficos para estabelecer amostragem mínima na estimativa de erro médio (a) e RMSE (b). Dados da classe “Mosaico de agricultura e pastagem” (21).



Fonte: Próprio autor (2023)

Com base nos gráficos relativos à classe 21, nota-se que conjuntos amostrais contendo a partir de 40 unidades começam a oscilar em torno do valor de referência para classe, com baixa variabilidade, o que denota certa estabilização. Convém destacar que o valor de referência foi calculado com todas as unidades amostrais das 5 regiões de estudo para a classe em questão.

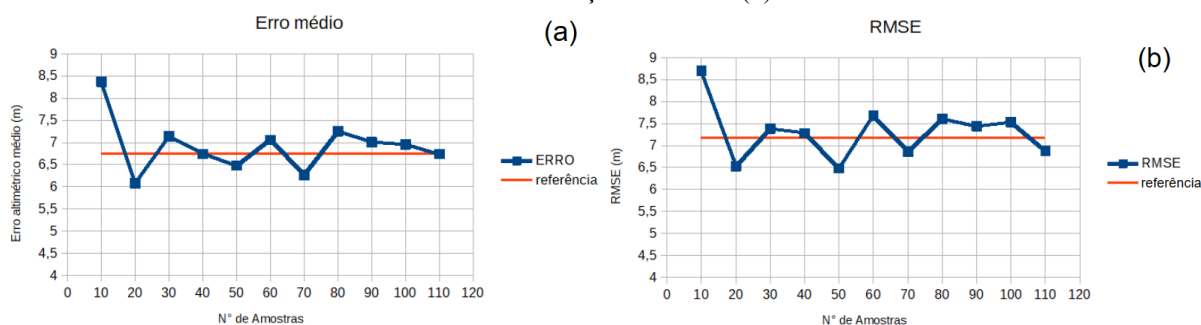
Figura 18: Gráficos para estabelecer amostragem mínima na estimativa de erro médio (a) e RMSE (b). Dados da classe “Pastagem” (15).



Fonte: Próprio autor (2023)

Com base nos gráficos relativos à classe 15, nota-se que conjuntos amostrais contendo a partir de 40 unidades começam a oscilar em torno do valor de referência para classe, com variação relativamente baixa e constante, denotando certa estabilização.

Figura 19: Gráficos para estabelecer amostragem mínima na estimativa de erro médio (a) e RMSE (b). Dados da classe "Formação Florestal" (3).



Fonte: Próprio autor (2023)

Com base nos gráficos relativos à classe 3, nota-se que conjuntos amostrais contendo a partir de 30 unidades começam a oscilar em torno do valor de referência para classe, com pouca variação. Nesta classe, a quantidade de unidades amostrais mínima para início da estabilização foi um pouco menor que as duas primeiras.

Diante das análises feitas com os dados contidos nos gráficos acima, concluiu-se que, para se obter um valor de erro altimétrico médio do MDE AW3D30 a nível local, são necessárias pelo menos 40 unidades numa amostra. Esse valor é importante pois, a partir dele, será feita a amostragem base para metodologia de correção proposta neste estudo, e cujos resultados estão expressos na seção 4.2.

4.2 MODELAGEM DE CORREÇÃO DAS ALTITUDES

A modelagem de correção empírica desenvolvida neste estudo teve dois grandes princípios: a efetividade em melhorar significativamente a acurácia altimétrica do MDE AW3D30 e a facilidade na sua aplicação. O primeiro princípio terá sua demonstração mais adiante, nos itens 4.2.4 e 4.2.5. Quanto à demonstração de como é simples aplicar a metodologia de correção está disposta no próximo item 4.2.1 a seguir, com coeficientes e fórmula apresentados em 4.2.2 e 4.2.3, respectivamente.

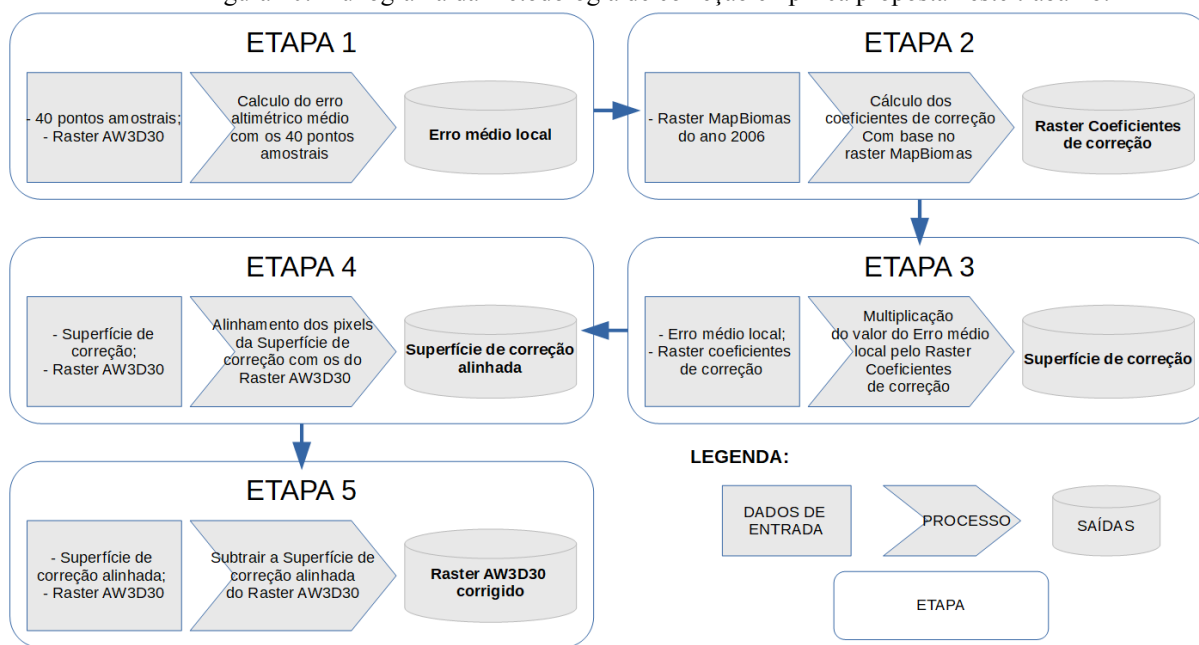
4.2.1 Fluxograma da correção

A aplicação da metodologia segue um passo-a-passo em cinco etapas, como demonstrado no fluxograma da Figura 20. Neste processo é necessário um programa SIG, de livre escolha, para processar os dados de entrada, como descritos a seguir.

Para a primeira etapa são necessários, no mínimo, 40 pontos amostrais na área a ser corrigida, os quais devem ter suas respectivas altitudes ortométricas tomadas por algum método de alta precisão. Além desse dado, na primeira etapa é necessário também o raster

AW3D30 original da área selecionada, o qual pode ser adquirido junto à JAXA de modo gratuito. Com estes dados, faz-se a comparação das altitudes da verdade de campo com as do MDE AW3D30, para cada ponto amostral, obtendo-se o erro altimétrico e , em sequência, calculando-se a média destes erros.

Figura 20: Fluxograma da metodologia de correção empírica proposta neste trabalho.



Fonte: Próprio autor (2023).

Na segunda etapa da correção é necessário o dado MapBiomias, especificamente para o ano 2006 (ano em que os dados ALOS foram tomados para construção do AW3D30). Com esse raster, se aplica a fórmula condicional com auxílio de uma calculadora raster, presente na maioria dos *softwares* SIG. Tal fórmula, apresentada na Figura 21, converte os valores de pixel do MapBiomias (referentes às classes de uso e cobertura do solo) em valores dos coeficientes de correção. O subproduto desta etapa é um raster cujos pixels carregam os coeficientes de correção altimétrica, o qual foi denominado de raster dos coeficientes de correção.

Na terceira etapa da correção, as entradas consistem nos subprodutos das duas etapas iniciais: O erro médio dos 40 pontos amostrais (chamado de erro médio local) e o raster dos coeficientes de correção. Assim, nesta etapa é feita a simples multiplicação destas entradas, resultando num raster que foi denominado de superfície de correção, e cujos pixels já carregam os valores que irão corrigir o AW3D30 original.

Na quarta etapa é necessário realizar uma operação de alinhamento raster, para que os pixels da superfície de correção (originados do MapBiomias) sejam forçados a ocupar o

mesmo espaço geográfico dos pixels do AW3D30. Diante desta necessidade, pode haver algum efeito de borda, mas seus impactos são mínimos dado o mesmo tamanho de pixel e a coincidente proximidade espacial da cobertura MapBiomias e AW3D30. O resultado desta etapa foi chamada de superfície de correção alinhada.

Na quinta etapa é, enfim, efetuada operação raster de subtração, onde os valores da superfície de correção são subtraídos das elevações do AW3D30, resultando no MDE AW3D30 corrigido.

Diante do exposto é notável a simplicidade de aplicação desta metodologia, frente a algumas metodologias propostas na literatura que têm aplicabilidade mais complexa, as quais, por exemplo, usam regressões com múltiplas variáveis para estimar a compensação de erros, como no caso dos trabalhos de Gallant e Read (2016), O’loughlin et al (2016) e Magruder et al (2021). Outra metodologia um tanto quanto complexa trata-se da proposta de Hawker et al (2022), aplicando aprendizado de máquina com até 15 variáveis de entrada. Em contrapartida, a proposta de correção mais simplista que foi elencada, mais simples até que a deste trabalho, foi a de Girão et al (2019), centrada na interpolação de uma superfície de correção utilizando pontos amostrais e interpoladores determinísticos comuns.

4.2.2 Coeficientes de correção

Para aplicação da metodologia apresentada acima são necessários os coeficientes de correção por classe de uso e cobertura do solo do MapBiomias, cujo estabelecimento se deu ao longo deste estudo. Na Tabela 19, abaixo, são apresentados estes coeficientes correlacionados com as respectivas classes do MapBiomias.

Tabela 19: Coeficientes de correção altimétrica para o MDE AW3D30 em função das classes do MapBiomias.

Classe	Código	Erro Médio*	Coeficiente de correção
Formação Florestal	3	6,22	1,8133
Rio, Lago e Oceano	33	3,99	0,8072
Área Urbanizada	24	3,53	0,5968
Mosaico de Agricultura e Pastagem	21	3,32	0,5026
Formação Savânica	4	3,11	0,4074
Pastagem	15	2,77	0,2539
Outras Áreas não Vegetadas	25	2,49	0,1281
Formação Campestre	12	2,42	0,0928
MÉDIA DO ERRO DAS CLASSES = 2,21 **			
* O Erro médio levando-se em conta todas áreas de estudo, menos SP-02			
** Média do erro de todas as classes encontradas neste estudo, excluindo-se SP-02			

Fonte: Próprio autor (2023).

Ressalta-se, novamente, que as demais classes do MapBiomias não listadas na Tabela 18, não passam por correção na metodologia aqui proposta. Isso se deve ao fato que tais classes tiveram pouco ou nenhum dado nas áreas de estudo, impossibilitando o estabelecimento dos parâmetros para cálculo dos coeficientes de correção.

Na Tabela 18, a coluna “coeficientes de correção” é resultante da aplicação da Equação 08, a qual utiliza os dados da coluna “Erro médio” de cada classe e a constante “média dos erros das classes”. Esta última, diga-se de passagem, é a média dos erros de todas as classes que tiveram alguma unidade amostral encontrada nas áreas de estudo, excluindo-se SP-02.

4.2.3 Fórmula condicional para conversão raster

A fórmula condicional proposta neste estudo tem por objetivo converter os valores dos pixels do MapBiomias, correspondentes às classes, em valores dos coeficientes de correção. Como a linguagem computacional para aplicação de fórmulas condicionais variam de acordo o programa SIG utilizado, abaixo (Figura 21) está demonstrada a lógica desta conversão de um modo didático, para apreensão do conceito. Assim, na aplicação prática da fórmula o usuário deve adaptá-la à codificação da calculadora raster utilizada.

Figura 21: Fórmula condicional usada na correção do AW3D30..

SE	Raster MapBiomias =	3	ENTÃO	Superfície de correção =	1,8133
SE	Raster MapBiomias =	33	ENTÃO	Superfície de correção =	0,8072
SE	Raster MapBiomias =	24	ENTÃO	Superfície de correção =	0,5968
SE	Raster MapBiomias =	21	ENTÃO	Superfície de correção =	0,5026
SE	Raster MapBiomias =	4	ENTÃO	Superfície de correção =	0,4074
SE	Raster MapBiomias =	15	ENTÃO	Superfície de correção =	0,2539
SE	Raster MapBiomias =	25	ENTÃO	Superfície de correção =	0,1281
SE	Raster MapBiomias =	12	ENTÃO	Superfície de correção =	0,0928
SE	Raster MapBiomias =	outros valores	ENTÃO	Superfície de correção =	0,0000

Fonte: Próprio autor (2023)

Neste estudo foi utilizada a calculadora raster do QGIS, onde a fórmula condicional teve de ser adaptada conforme o código expresso na Figura 22.

Figura 22: Código usado na calculadora raster do QGIS para calcular a superfície de correção do AW3D30.

```

if( "MAPBIOMAS@1"= 3 , 1.8133 ,
if( "MAPBIOMAS@1"= 33 , 0.8072 ,
if( "MAPBIOMAS@1"= 24 , 0.5968 ,
if( "MAPBIOMAS@1"= 21 , 0.5026 ,
if( "MAPBIOMAS@1"= 4 , 0.4074 ,
if( "MAPBIOMAS@1"= 15 , 0.2539 ,|
if( "MAPBIOMAS@1"= 25 , 0.1281 ,
if( "MAPBIOMAS@1"= 12 , 0.0928 ,0)))))))))

```

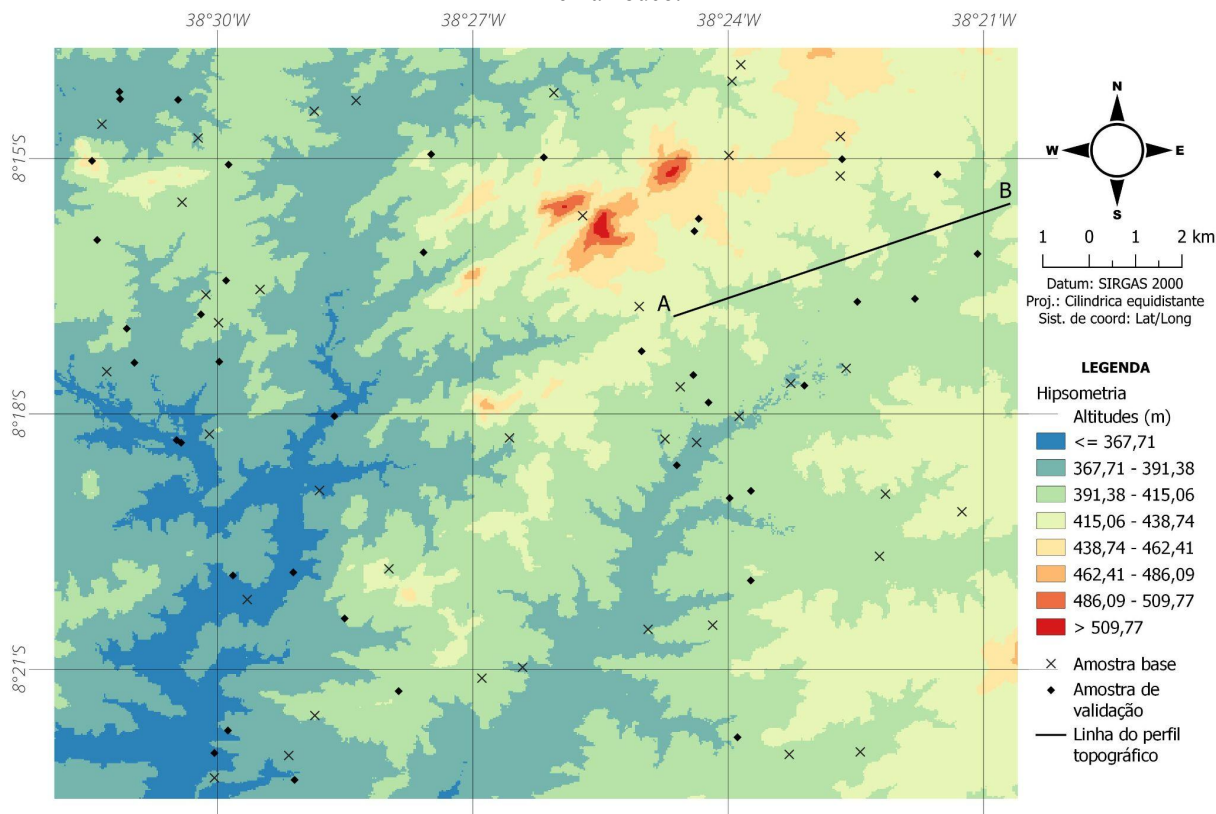
Fonte: Próprio autor (2023)

4.2.4 AW3D30 corrigido

O um dos maiores objetivos deste trabalho foi obter um modelo digital de elevação corrigido, baseado no AW3D30, nas áreas de validação. Deste modo, nesta seção serão apresentados os MDE AW3D30 corrigidos para as três áreas de validação já mencionadas, onde na Figura 23 se apresenta o raster relativo á área em Pernambuco; na Figura 25 é retratado o raster corrigido para a área em São Paulo, e, finalmente, na Figura 27 é mostrado o resultado para área na Bahia. Nesta última área de validação não foi possível traçar perfil topográfico representando a verdade de campo, visto que foram coletados apenas os 80 pontos para construir a “amostra base” e “amostra de validação”, quantidade insuficiente para interpolar uma superfície raster necessária na geração de um perfil topográfico.

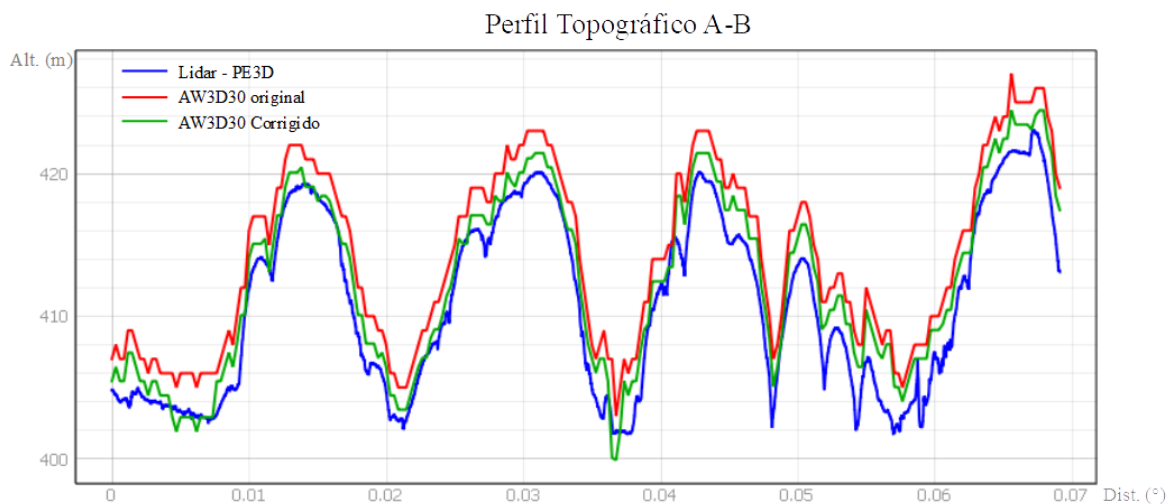
Na área de validação em Pernambuco, observando o MDE AW3D30 corrigido com a metodologia de correção empírica (Figura 23), nota-se que as altitudes variam entre 367 m e 510 m, aproximadamente. A maior parte das altitudes da região são de até 439 m. Esta área de validação ocupa 147,15 km². No tocante ao perfil topográfico (Figura 24) mostrando as três superfícies: Lidar PE3D (altitudes de referência), AW3D30 original (sem correção) e o AW3D30 corrigido, verifica-se que a versão corrigida do AW3D30 se aproxima mais da verdade de campo que o modelo sem correção. Isto é um indicativo que a modelagem proposta neste estudo tem efetividade em tornar o modelo AW3D30 mais verossímil.

Figura 23: Mapa hipsométrico elaborado com o AW3D30 corrigido para área de validação no Estado de Pernambuco.



Fonte: Próprio autor (2023).

Figura 24: Perfil topográfico para o alinhamento A-B do mapa hipsométrico da Figura 23..

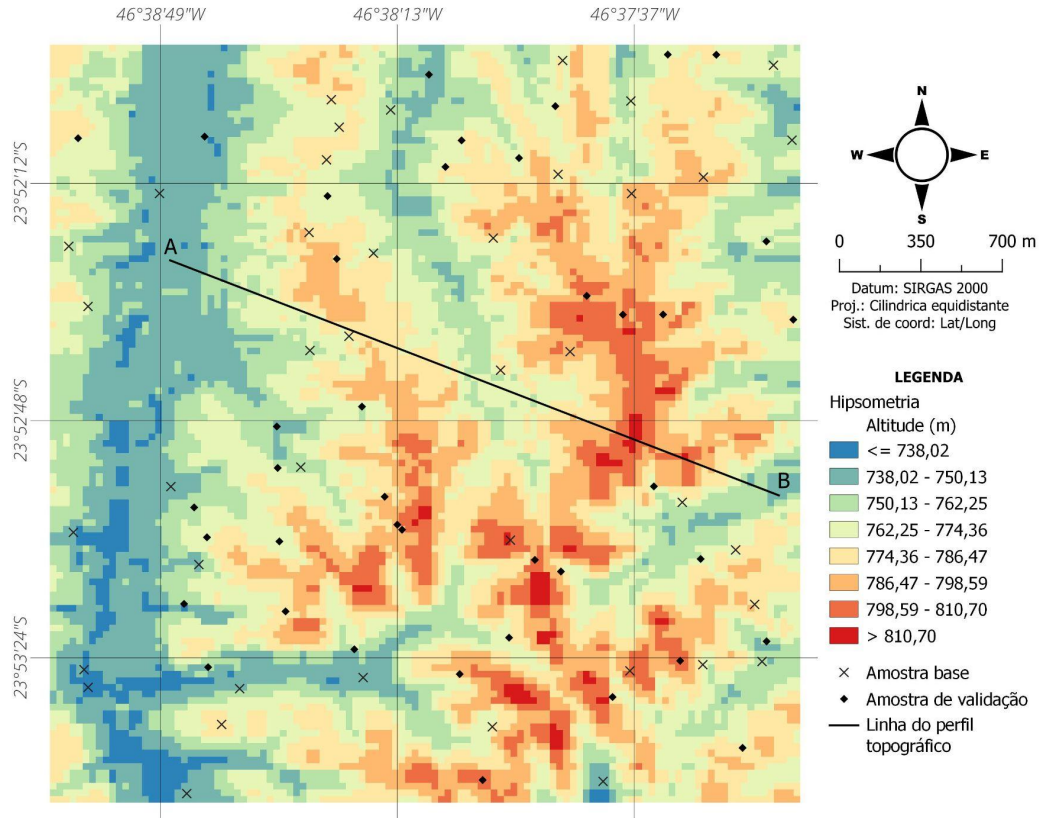


Fonte: Próprio autor (2023)

Na área de validação em São Paulo, observando o MDE AW3D30 corrigido com a metodologia de correção empírica (Figura 25), nota-se que as altitudes variam entre 738 m e 811 m, aproximadamente. A maior parte das altitudes da região são de até 787 m. Esta área de validação ocupa 10,46 km². No tocante ao perfil topográfico (Figura 26) mostrando as três

superfícies: Lidar GeoSampa (altitudes de referência), AW3D30 original (sem correção) e o AW3D30 corrigido, verifica-se que a versão corrigida do AW3D30 se aproxima mais da verdade de campo que o modelo sem correção. Isto indica que a modelagem proposta neste estudo tem efetividade em tornar o modelo AW3D30 mais verossímil.

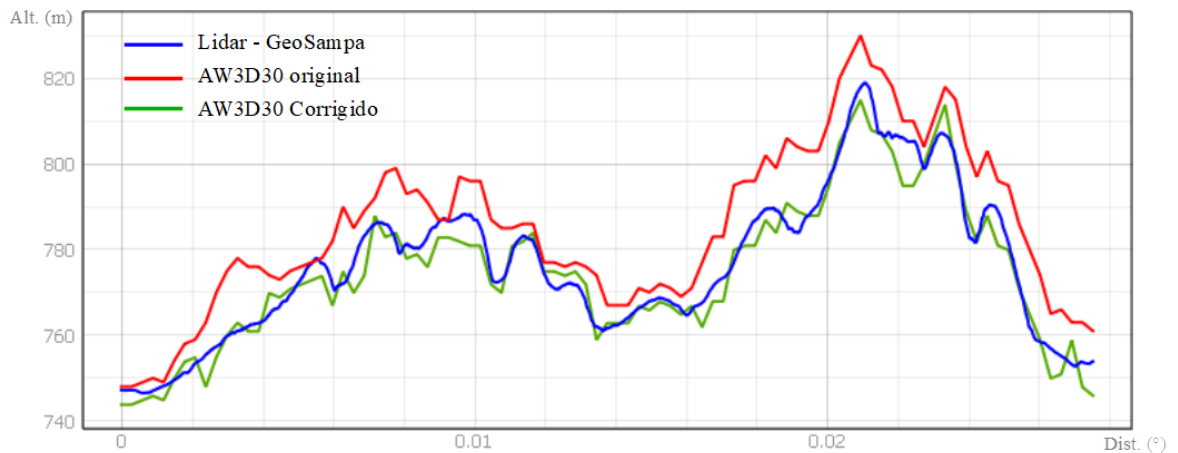
Figura 25: Mapa hipsométrico elaborado com o AW3D30 corrigido para área de validação no Estado de São Paulo.



Fonte: Próprio autor (2023)

Figura 26: Perfil topográfico para o alinhamento A-B do mapa hipsométrico da Figura 25.

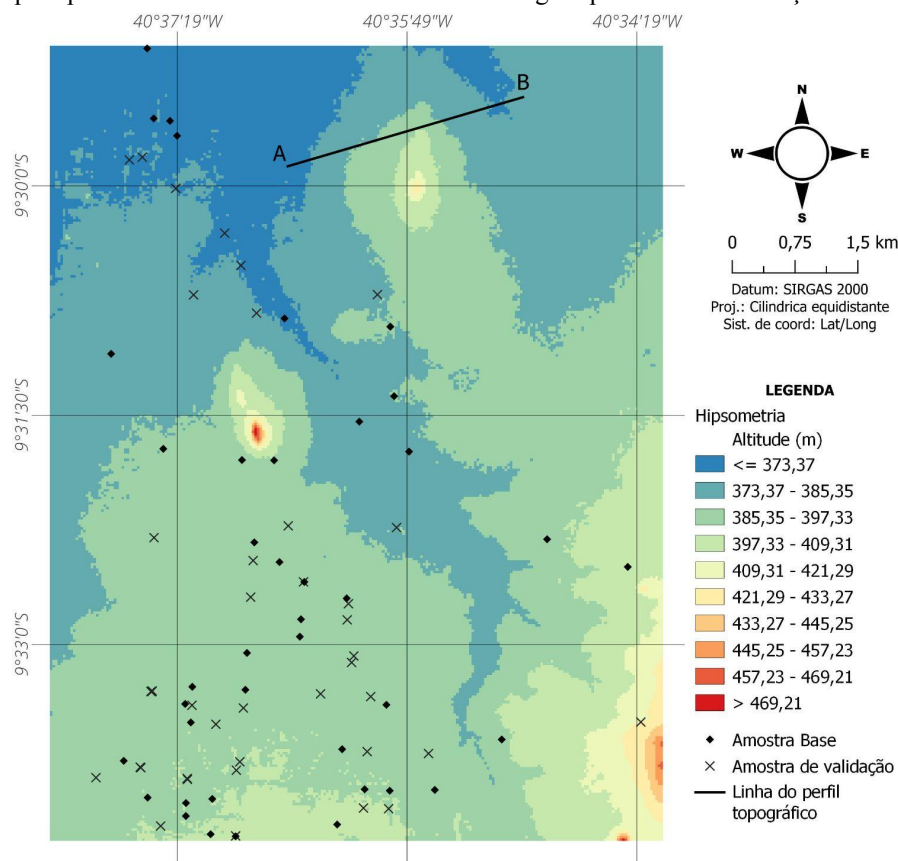
Perfil Topográfico A-B



Fonte: Próprio autor (2023)

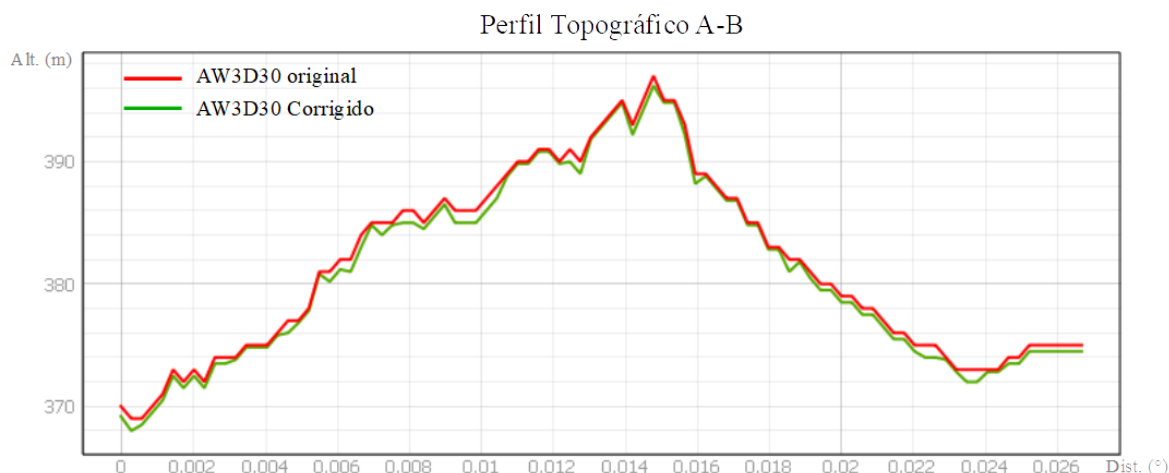
Na área de validação na Bahia, observando o MDE AW3D30 corrigido com a metodologia de correção empírica (Figura 27), nota-se que as altitudes variam entre 373 m e 470 m, aproximadamente. A maior parte das altitudes da região são de até 422 m. Esta área de validação ocupa 67,39 km². Quanto ao perfil topográfico (Figura 28), nota-se uma redução das altitudes de um modo geral do modelo original para o corrigido, o que, baseado nos casos das outras duas áreas de validação, significa uma aproximação para a superfície real do terreno. Convém reiterar o fato desta área de validação não possuir perfil topográfico da verdade de campo, devido a quantidade insuficiente de dados altimétricos para se interpolar um MDE.

Figura 27: Mapa hipsométrico elaborado com o AW3D30 corrigido para área de validação no Estado da Bahia.



Fonte: Próprio autor (2023)

Figura 28: Perfil topográfico para o alinhamento A-B do mapa hipsométrico da Figura 27..



Fonte: Próprio autor (2023)

4.2.5 Validação dos resultados através do PEC-PCD

Com os estudos do PEC-PCD para os modelos AW3D30 original e corrigido, nas áreas de validação, foi possível averiguar significativa melhora na acurácia altimétrica após as correções no AW3D30. Na Tabela 20 estão apresentados os resultados deste estudo, onde 40 pontos amostrais em cada uma das três áreas foram usados para se extrair o RMSE, e conseqüentemente, enquadrar as versões original e corrigida do AW3D30 no PEC-PCD, conforme a Norma EB80-N-72.003 do DSG.

Tabela 20: Avaliação da qualidade altimétrica do MDE AW3D30 antes e depois da correção empírica proposta neste estudo.

ÁREA DE VERIFICAÇÃO	DADO	RMSE (m)	Escala 1:50.000		Escala 1:100.000		GANHO EM RMSE (m)	% DO GANHO
			PEC-PCD	% < PEC (%)	PEC-PCD	% < PEC (%)		
Pernambuco	Sem correção	4,06	B	100,00	A	100,00	1,59	39,06
	Após correção	2,48	A	95,00	A	100,00		
São Paulo	Sem correção	8,22	Não se enquadra	-	B	100,00	3,67	44,62
	Após correção	4,55	C	95,00	A	100,00		
Bahia	Sem correção	1,87	A	100,00	A	100,00	0,61	32,43
	Após correção	1,26	A	100,00	A	100,00		

Fonte: Próprio autor (2023)

Observando os resultados da área de validação em Pernambuco, se nota uma redução do RMSE de 4,06 m para 2,48 m, representando um ganho de 1,58 m ou 39,06% na qualidade altimétrica do modelo nesta região. Essa melhora foi suficiente para alterar o enquadramento do PEC-PCD na escala de 1:50.000 da classe “B” para a classe “A”, demonstrando o impacto positivo da metodologia de correção.

Quanto aos resultados da área de validação em São Paulo, observa-se uma redução do RMSE de 8,22 m para 4,55 m, representando um ganho de 3,67 m ou 44,62% na qualidade altimétrica do modelo nesta região. Esse foi o maior ganho em qualidade dentre as áreas de validação, após a correção. Na análise do PEC-PCD, nota-se que antes da correção o MDE

não se enquadra em nenhuma classe para escala de 1:50.000, e logo após ser corrigido teve enquadramento na classe “C”. Em se tratando da escala de 1:100.000, passou da classe “B” antes da correção, para classe “A” após ser corrigido.

Observando os resultados da área de validação na Bahia, verifica-se redução do RMSE de 1,87 m para 1,26 m, representando um ganho de 0,61 m ou 32,43% na qualidade altimétrica do modelo nesta região. No enquadramento do PEC-PCD, não houve alteração da classe, dado que antes da correção o AW3D30 já estava na classe “A”, tanto na escala de 1:50.000 quanto na 1:100.000. Ademais, a melhora de 32,43% na acurácia altimétrica comprova efetividade na modelagem de correção altimétrica para o MDE AW3D30 proposta neste estudo.

Os resultados positivos demonstrando melhora significativa na qualidade altimétrica do AW3D30, em todas as áreas de validação, respaldam o uso da metodologia de correção proposta neste estudo em áreas no Brasil. Os ganhos variam entre 32% e 44%, tendo um ganho médio de 38,70% em RMSE, aproximadamente. Valores similares foram atingidos por Hawker et al (2022) na correção global do MDE Copernicus, que conseguiu melhoras de 30% e 44%, em áreas construídas e de floresta, respectivamente. Magruder et al. (2021) melhoraram o SRTM em 41% em sua correção para o estado da Carolina do Norte/EUA. Já O’loughlin et al (2016) informaram um ganho de 58% em RMSE corrigindo o SRTM globalmente. Girão et al (2019), aplicando uma correção simples no MDE ASTER, apontaram uma redução de 5% no RMSE do modelo no estado do Rio de Janeiro.

5 CONSIDERAÇÕES FINAIS

Os trabalhos para atingir os objetivos do presente estudo enfrentaram desafios, os quais foram contornados através da associação de técnicas e conceitos consagrados na literatura com testes empíricos, exultando numa modelagem de correções das altitudes para o MDE global AW3D30. Além disso, todo arcabouço de resultados e revisão bibliográfica aqui apresentados fornecem informações acerca da qualidade altimétrica do AW3D30 e sua correlação com o uso e cobertura do solo, apoiando futuros trabalhos dentro desta temática.

O presente estudo, por meio de investigações estatísticas e a proposição de uma metodologia simplificada para corrigir erros em MDE, conseguiu a validação da modelagem, encorajando sua aplicação em território brasileiro para o AW3D30. Há de se dizer também que, apesar dos resultados positivos, esta modelagem possui suas limitações, e pode ser aperfeiçoada ou ainda cooperar com novos estudos que visam melhorar a qualidade altimétrica de MDE globais. Diante disso, destaca-se as principais conclusões deste estudo a seguir:

- A acurácia altimétrica do AW3D30 original apresentou resultados dentro do esperado para o modelo na maioria das áreas testadas neste estudo, para as quais o indicador RMSE aponta boa acurácia;
- As áreas de estudo e validação em Pernambuco, Bahia, e as amostras pontuais da RAAP espalhadas por todo o Brasil, tiveram RMSE para o AW3D30 original em limiares entre 1,90m e 4,30m, aproximadamente. Tais valores estão bem abaixo do RMSE médio encontrado na literatura mundial, que é 7,45 metros;
- Já para o caso das áreas estudadas no estado de São Paulo, em duas delas foram encontrados valores de RMSE iguais a 6,59m e 8,22m, bem próximos ao limiar médio (7,45m) da literatura elencada neste estudo. O caso que extrapolou tal limiar foi na área SP-01, onde o RMSE foi de 11,10m, porém na literatura se encontra valores similares a este.
- As análises dos valores de RMSE das áreas de estudo e validação indicam que não há uma homogeneidade na acurácia do AW3D30 em todo território do Brasil. Fato comprovado também pela ANOVA com teste de média, além dos dados de outros autores. Isso permite concluir que a qualidade do modelo AW3D30 possui grande variabilidade espacial, merecendo uma investigação minuciosa para identificação dos prováveis agentes;

- Esta variabilidade espacial dos erros no MDE AW3D30 não encoraja correções globais com fatores uniformes, os quais não levem em consideração as realidades locais. Esta conclusão direcionou este estudo a optar por uma modelagem de correção das altitudes que levasse em conta o erro médio local;
- Os maiores erros no AW3D30 são encontrados em regiões com cobertura florestal, fato amplamente comprovado pelos diversos resultados deste trabalho. Obviamente, isso se explica pelos deslocamentos na altitude causados pelas copas das árvores durante o imageamento pelo sensor em órbita.
- As análises acerca do erro por classes de uso e cobertura do solo permitem concluir ainda que o MDE AW3D30 tem sua qualidade sempre afetada em regiões com presença de vegetação ou construções. Outra cobertura do solo com erros acima da média é a superfície de água;
- A metodologia empírica para correções no AW3D30, baseada em erros proporcionais (coeficientes) pela classe de uso e cobertura do solo, mostrou-se efetiva pois reduziu o RMSE do AW3D30 em 38,70%, na média, para as áreas de validação;
- Os resultados encorajam a aplicação da metodologia corretiva proposta neste estudo em todo território nacional, visto que foram testadas áreas dentro e fora do contexto delimitador (região geográfica, fonte de dados, tipo de dados de validação, etc), e todos acusaram melhora significativa na acurácia do MDE;
- A cobertura MapBiomas foi essencial na elaboração da metodologia de correção altimétrica aqui proposta. Os dados oriundos do projeto MapBiomas estão cumprindo seu papel multifacetado de apoio a estudos e análises acerca do uso e cobertura do solo brasileiro;
- A metodologia proposta tem sua fragilidade no tocante a propor correção apenas para oito classes de uso e cobertura do solo do MapBiomas, visto que as demais não tiveram dados amostrais suficientes para se estabelecer parâmetros de modo estatisticamente seguro. Assim, trabalhos futuros poderiam analisar estatisticamente mais áreas pelo país, buscando aperfeiçoar os coeficientes já propostos e estabelecer os coeficientes para as demais classes de uso e cobertura do MapBiomas;
- Os principais impactos deste trabalho residem na investigação da relação entre os erros inerentes ao MDE AW3D30 e as classes de uso e cobertura do solo, e na proposição de uma modelagem corretiva aplicada à MDE globais, o que, conseqüentemente, poderá melhorar as análises espaciais que utilizam estes modelos como base de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

AUDETTE, M. A.; FERRIE, F. P.; PETERS, T. M. *An algorithmic overview of surface registration techniques for medical imaging*, Medical Image Analysis, vol 2, 2000. pp. 201-217.

ARAÚJO, O. .; BRUM, E. V.; SILVA, E. .; CAIONI, C. .; CLAUDINO, W. V. . ACURÁCIA POSICIONAL DO MODELO DIGITAL DE TERRENO COM OS MODELOS DIGITAIS DE ELEVAÇÃO: ASTER GDEM, SRTM E TOPODATA . ENCICLOPEDIA BIOSFERA, [S. l.], v. 10, n. 19, 2014. Disponível em: <https://www.conhecer.org.br/ojs/index.php/biosfera/article/view/2226>. Acesso em: 12 abr. 2023.

BARBOSA, R.F.; CICERELLI, R. E.; DE ALMEIDA, T. Análise comparativa entre os Modelos Digitais de Elevação (MDE) do satélite Alos. In: XIX SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 19., 2019, Santos. Anais. São José dos Campos: INPE, 2019.

BRASIL. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). . Rede Altimétrica. 2022. Disponível em: <https://www.ibge.gov.br/geociencias/informacoes-sobre-posicionamento-geodesico/rede-geodesica/16283-rede-altimetrica.html?=&t=sobre>. Acesso em: 10 set. 2022.

BRASIL, GOVERNO FEDERAL. Informações digitais de dados altimétricos - Referência de Nível. Disponível em: https://dados.gov.br/dados/conjuntos-dados/cged_bdg_rn. Acesso em: 10 set. 2022.

CARVALHO, J. A. B.; SILVA, D. C. Métodos para avaliação da acurácia posicional altimétrica no Brasil. Revista Brasileira de Cartografia, v. 70, n. 2, 2018, p. 725- 744. Doi: 10.14393/rbcv70n2-45404.

CEGATTA, Ítalo. ANOVA e teste de Tukey. 2016. Disponível em: <https://italocegatta.github.io/anova-e-teste-de-tukey/>. Acesso em: 18 set. 2022.

CHAGAS, César S. et al . Avaliação de modelos digitais de elevação para aplicação em um mapeamento digital de solos. Rev. bras. eng. agríc. ambient., Campina Grande, v. 14, n. 2, p. 218-226, Feb. 2010. Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-43662010000200014&lng=en&nrm=iso>. access on 20 Nov. 2020. <https://doi.org/10.1590/S1415-43662010000200014>.

CHEN, S., WANG, W., ZUYLEN H. (2010) “A comparison of outlier detection algorithms for ITS data”. Expert Systems with Applications 37, 1169-1178.

CORREIA, L. F. S. O Futuro Chegou! A popularização dos computadores e o presentismo nos anúncios publicitários brasileiros na década de 1980. Anais da ReACT-Reunião de Antropologia da Ciência e Tecnologia, v. 5, n. 5, 2022.

DIRETORIA DO SERVIÇO GEOGRÁFICO DO EXÉRCITO BRASILEIRO (DGS). EB80-N-72.001: Especificação Técnica para Produtos de Conjuntos de Dados Geoespaciais (ET PCDG Defesa F Ter). 2ª Edição. Brasília: DSG, 2016a.

DIRETORIA DO SERVIÇO GEOGRÁFICO DO EXÉRCITO BRASILEIRO (DGS). EB80-N-72.003: Norma da Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais de Defesa da Força Terrestre (ET ADGV Defesa F Ter). 2ª Edição. Brasília, 2016b.

DOBSON, M. C. (2000). *Forest information from Synthetic Aperture Radar. Journal of Forestry*. n.98, p.41-43.

DOYLE, F.J. *Digital terrain models: an overview. Photogrammetric Engineering and Remote Sensing*, v.44, n.12, p.1481-1485, 1978.

ELKHRACHY, I., 2018. Vertical accuracy assessment for SRTM and ASTER Digital Elevation Models: A case study of Najran city, Saudi Arabia. *Ain Shams Eng. J.* 9 (4), 1807–1817. <https://doi.org/10.1016/J.ASEJ.2017.01.007>.

FELICÍSIMO, A. M. *Modelos digitais de terreno. Introducción y aplicaciones en las ciencias ambientales*. Oviedo, Pentalfa, 118 S. 2001.

FIGUEIRA, M. M. C. Identificação de outliers. *Revista Millenium Online*, Viseu, n. 12. Out. 1998. Disponível em: . Acesso em: 18 set. 2022.

FREIRE-SILVA, J.; MIRANDA, R. Q.; CANDEIAS, A. L. B. Análise dos Modelos Digitais de Elevação (PE3D, SRTM-30, SRTM-90, ASTER GDEM, TOPODATA, TANDEM-X, ALOS PALSAR e ALOS AW3D30) e a necessidade da produção de dados altimétricos em excelência no Brasil. *Revista Brasileira de Geografia Física*, [S.l.], v. 15, n. 3, p. 1543-1555, jun. 2022. ISSN 1984-2295. Disponível em: <<https://periodicos.ufpe.br/revistas/rbgfe/article/view/251253>>. Acesso em: 11 abr. 2023. doi:<https://doi.org/10.26848/rbgf.v15.3.p1543-1555>.

GALLANT, J. C.; READ, A. M. **A near-global Bare-Earth MDE from SRTM**. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B4, XXIII ISPRS Congress, Prague, Czech Republic, 2016.

GIRÃO, S. R.; VICENS, R. S., & FERNANDES, P. J. F. (2019). Correção Altimétrica do ASTER GDEM Através do Método de Superfícies de Compensação de Erros. *Revista Brasileira de Cartografia*, 71(2), 392-433.

GOMES, F.P. A comparação entre médias de tratamentos na análise de variância. São Paulo: Escola superior de agricultura “Luiz de Queiroz”, 1954.

GRUEN, A.; AKCA, D. *Least squares 3D surface matching*, Anais...ASPRS 2005 Annual Conference, Baltimore, Maryland, 2005. pp. 7-11.

HAWKER, L., UHE, P., PAULO, L., SOSA, J., SAVAGE, J., SAMPSON, C., & NEAL, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, v. 17, n. 2, p. 024016, 2022.

IORIO, M. M.; LASTORIA, G.; MIOTO, C. L.; ALBREZ, E. A.; PARANHOS FILHO, A.C. (2012). Avaliação de Modelos Digitais de Elevação Extraídos de Imagem Alos/Prism e Comparação com os Modelos Disponibilizados Gratuitamente na Web. São Paulo, UNESP: Geociências (ISSN 1980-900X), v. 31, n. 4, p. 650-664, 2012.

JAXA. ALOS HANDBOOK: *Alos Data Users Handbook*. Disponível em http://www.eorc.jaxa.jp/ALOS/en/doc/fdata/ALOS_HB_RevC_EN.pdf. 2008 Acesso em novembro de 2020.

JAXA. *ALOS Global Digital Surface Model (DSM) “ALOS World 3D-30m” (AW3D30) Ver.3.2/3.1 Product Description*. Earth Observation Research Center - Japan Aerospace Exploration Agency (JAXA EORC), 2021.

LANDAU, E. C.; GUIMARÃES, D. P. Análise comparativa entre os modelos digitais de elevação ASTER, SRTM e TOPODATA. In: Embrapa Milho e Sorgo-Artigo em anais de congresso (ALICE). In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., 2011, Curitiba. Anais... São José dos Campos: INPE, 2011., 2011.

LUZ, R. R.; FREITAS, S. R. C.; DALAZOANA, R. Acompanhamento do Datum Altimétrico Imbituba através das redes altimétrica e maregráfica do Sistema Geodésico Brasileiro. Anais...In: CONGRESSO INTERNACIONAL DE CIENCIAS DE LA TIERRA, 7., 2002, Santiago, CL. Santiago: IGM, 2002. Disponível em: <https://ww2.ibge.gov.br/home/geociencias/geodesia/artigos/2002-Acompanhamento_Imbituba_atraves_Redes_SGB.pdf>. Acesso em: 20 set. 2022.

MACHADO, V. S. et al. Validação da Acurácia Vertical de Modelos Digitais de Superfície Utilizando o Banco de Dados do Sistema de Gestão Fundiária: Um Estudo de Caso no Oeste do Estado de São Paulo. Anuário do Instituto de Geociências, v. 42, n. 4, p. 139-147, 2020.

MAGRUDER, L., NEUENSCHWANDER, A. & KLOTZ, B. *Digital terrain model elevation corrections using space-based imagery and ICESat-2 laser altimetry*. In: Remote Sensing of Environment, Volume 264, 2021, ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2021.112621>.

MIOT, H. A.. Avaliação da normalidade dos dados em estudos clínicos e experimentais. *Jornal Vascular Brasileiro* [online]. 2017, v. 16, n. 2 [Acessado 15 Setembro 2022], pp. 88-91. Disponível em: <<https://doi.org/10.1590/1677-5449.041117>>. ISSN 1677-7301. <https://doi.org/10.1590/1677-5449.041117>.

MOORE, I. D., GRAYSON, R. B., & LADSON, A. R. (1991). Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5(1), 3–30. <http://dx.doi.org/10.1002/hyp.3360050103>.

MORAIS, J. D., FARIA, T. S., ELMIRO, M. A. T., NERO, M. A., SILVA, A. D. A., & NOBREGA, R. A. D. A. (2017). Avaliação Altimétrica dos Modelos Digitais de Elevação ASTER GDEM v2 e SRTM v3: Um Estudo de Caso na Área Urbana de Belo Horizonte, MG, Brasil. *Boletim de Ciências Geodésicas*, 23(4), 654-668, 2017.

MORAIS, J. D. Avaliação de modelos digitais de elevação provenientes de dados de sensoriamento remoto de distribuição gratuita [manuscrito] / Josyceyla Duarte Morais. – 2017. 94 f., enc.: il. (principalmente color.) Orientador: Marcos Antônio Timbó Elmiro. Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Geociências, 2017.

MOURA, L. Z.; BIAS, E. de S.; BRITES, R. AVALIAÇÃO DA ACURÁCIA VERTICAL DE MODELOS DIGITAIS DE ELEVAÇÃO (MDES) NAS BACIAS DO PARANOÁ E SÃO BARTOLOMEU. *Revista Brasileira de Cartografia*, [S. l.], v. 66, n. 1, 2014. DOI: 10.14393/rbcv66n1-43892. Disponível em: <https://seer.ufu.br/index.php/revistabrasileiracartografia/article/view/43892>. Acesso em: 11 abr. 2023.

MUKHERJEE, S., JOSHI, P.K., MUKHERJEE, S., GHOSH, A., GARG, R.D., MUKHOPADHYAY, A., 2013. Evaluation of vertical accuracy of open source Digital Elevation Model (DEM). *Int. J. Appl. Earth Observ. Geoinform.* 21, 205–217.

NIKOLAKOPOULOS, K. G. “Accuracy Assessment of ALOS AW3D30 DSM and Comparison to ALOS PRISM DSM Created with Classical Photogrammetric Techniques.”

European Journal of Remote Sensing, 2020, 53 (sup2): 39–52.
Doi:10.1080/22797254.2020.1774424.

O'LOUGHLIN, F. E.; PAIVA, R. C. D.; DURAND, M.; ALSDORF, D.E.; BATES, P. D. **A multi-sensor approach towards a global vegetation corrected SRTM MDE product.** Remote Sensing of Environment, v. 182, p. 49-59, 2016.

PAIVA, R. C. D., BUARQUE, D. C., COLLISCHONN, W., BONNET, M. -P., FRAPPART, F., CALMANT, S., & BULHÕES MENDES, C. A. (2013). Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. Water Resources Research, 49(3), 1226–1243 (2013).

PEREIRA, G. E.; TEN CATEN, A. Exatidão cartográfica de três MDE disponíveis para o Estado de Santa Catarina. Revista de Ciências Agroveterinárias, Lages, v. 18, n. 2, p. 258-262, 2019. DOI: 10.5965/223811711812019258. Disponível em: <https://www.revistas.udesc.br/index.php/agroveterinaria/article/view/13159>. Acesso em: 11 abr. 2023.

PERNAMBUCO, GOVERNO DO ESTADO DE. Agência Pernambucana de Águas e Climas: PE3D. Disponível em: <http://www.pe3d.pe.gov.br/>. Acesso: 03 de fevereiro de 2022.

PROJETO MAPBIOMAS – Coleção 6 da Série Anual de Mapas de Uso e Cobertura da Terra do Brasil do ano 2006, acessado em 14 de fev. 2021 através do link: <http://mapbiomas.org>.

REBELO, R. Aplicação do INSAR na Geração de Modelos Digitais do Terreno. 2007. 230 f. Dissertação (Mestrado em Engenharia Geográfica e Geoinformática) – Universidade de Lisboa, Lisboa, 2007.

REUTER, ISAAK, NELSON, ANDY, JARVIS, ANDREW, 2007. An evaluation of void-filling interpolation methods for SRTM data. International Journal of Geographical Information Science 21 (9), 983–1008. <https://doi.org/10.1080/13658810601169899>.

RIBOLDI, J.. Análise de variância. **Cadernos de matemática e estatística. Série B, Trabalho de apoio didático. Porto Alegre. No. 27 (mar. 1995), f. 1-105**, 1995.

SÃO PAULO, PREFEITURA MUNICIPAL DE. GeoSampa v. 10, 2019.. Disponível em: http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx. Acesso em 30 de fevereiro de 2022.

SANTILLAN, J.; MAKINANO-SANTILLAN, M. Vertical accuracy assessment of 30-M resolution alos, aster, and SRTM global dems over northeastern Mindanao, Philippines. In: Isprs Congress: The International Archives Of The Photogrammetry, Remote Sensing And Spatial Information Sciences, 23., 2016, Prague, Czech Republic. Analls... Prague: ISPRS, 2016. v. XLI-B4, p. 149-155.

SANTOS, A. R.; PARADELLA, W. R.; VENEZIANI, P; MORAIS, M. C. A estereoscopia com imagens RADARSAT-1: uma avaliação geológica na Província Mineral de Carajás. Revista Brasileira de Geociências. V. 29, n. 4, p. 627-632, 1999.

SANTOS, F. M.; LOLLO, J. A.; MAUAS, F. F. Caracterização da rede de drenagem da bacia hidrográfica do córrego Caçula – Ilha Solteira/SP. In: X encontro nacional de águas, 10., 2014. São Paulo. Anais. São Paulo: ABRH, 2014.

SHIRAMIZU, K.; DOI, K.; AOYAMA, Y. Generation of a high-accuracy regional DEM based on ALOS/PRISM imagery of East Antarctica. Polar Sci. 2017, 14, 30–38.

SILVA, Daniel da Costa e. Estudo da qualidade posicional e análise de padrões espaciais na distribuição de erros altimétricos em modelos digitais de elevação. 2020. Dissertação (Mestrado em Ciências Geodésicas e Tecnologias da Geoinformação)- Universidade Federal de Pernambuco, Recife, 2020.

SILVA, A. S. da, SANTIAGO, O. R. P. L., SILVA, C. R. da, 2018. Análise de Exatidão entre MDEs: AW3D, SRTM-30m e Projeto SPMGGO50. *Geografia (Londrina)* v. 27. n° 2, 87 – 98. DOI: 10.5433/2447-1747.2018v27n2p87

SOUZA, J. O. P. de. ANÁLISE DA PRECISÃO ALTIMÉTRICA DOS MODELOS DIGITAIS DE ELEVAÇÃO PARA ÁREA SEMIÁRIDA DO NORDESTE BRASILEIRO. *Revista do Departamento de Geografia*, [S. l.], v. 30, p. 56-64, 2015. Disponível em: <https://www.revistas.usp.br/rdg/article/view/98555>. Acesso em: 11 abr. 2023.

SOUZA, C.; AZEVEDO, Tasso. *MapBiomass general handbook*. MapBiomass: São Paulo, Brazil, p. 1-51, 2022.

TAKAKU, J., TADONO, T., TSUTSUI, K., & ICHIKAWA, M. (2016). *Validation Of a 3d Global Dsm Generated From Alos Prism*. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, v. 3, n. 4, 2016.

TADONO, T., SHIMADA, M., MURAKAMI, H., e TAKAKU, J., 2009. *Calibration of PRISM and AVNIR-2 onboard ALOS "Daichi"*. *Transactions on Geoscience and Remote Sensing, IEEE*, 47(12), pp. 4042-4050.

TADONO, T., ISHIDA, H., ODA, F., NAITO, S., MINAKAWA, K., e IWAMOTO, H., 2014. *"Precise Global DEM Generation by ALOS PRISM"*, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol.II-4, pp.71-76, 2014.

TAHIR, H. AND DIN, A. H. M.: Vertical Accuracy Assessment For Open-Source Digital Elevation Model: A Case Study Of Basrah City, Iraq, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-4/W6-2022, 355–361, <https://doi.org/10.5194/isprs-archives-XLVIII-4-W6-2022-355-2023>, 2023.

T. M. VENTURA; H. O. MARQUES; A. G. OLIVEIRA; C. A. MARTINS; M. C. J. A. NOGUEIRA; W. R. S. TEIXEIRA; J. M. FIGUEIREDO; A. G. BONFANTE. Detecção de Outliers em Dados Micrometeorológicos. IX Congresso Brasileiro de Agroinformática, Cuiabá, 2013.

UFSC (Brasil). Distribuição Normal: gaussiana. *Gaussiana*. 2022. Disponível em: <https://www.inf.ufsc.br/~andre.zibetti/probabilidade/normal.html>. Acesso em: 20 set. 2022.

VEREGIN, H. (1989): "A review of error models for vector to raster conversion". *The Operational Geographer*, 7(1): 11-15. 1989.

VIEL, Jorge Antônio; ROSA, Kátia Kellem da; MENDES JUNIOR, Cláudio Wilson. Avaliação da Acurácia Vertical dos Modelos Digitais de Elevação SRTM, ALOS World 3D e ASTER GDEM: Um Estudo de Caso no Vale dos Vinhedos, RS – Brasil. *Revista Brasileira de Geografia Física*, [S.l.], v. 13, n. 5, p. 2255-2268, jul. 2020. ISSN 1984-2295. Disponível em: <<https://periodicos.ufpe.br/revistas/rbgfe/article/view/243773/36912>>. Acesso em: 10 abr. 2023. doi:<https://doi.org/10.26848/rbgf.v13.5.p2255-2268>.

WALFISH, S.. "A review of statistical outlier methods," *Pharmaceutical Technol.*, vol. 30, no. 11, pp. 1–5, 2006.

WISE, S. *Assessing the quality for hydrological applications of digital elevation models derived from contours*. Hydrological Processes, v.14, p.1909-1929, 2000.

UUEMAA, E.; AHI, S.; MONTIBELLER, B.; MURU, M.; KMOCH, A. Vertical Accuracy of Freely Available Global Digital Elevation Models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). *Remote Sensing*, 2020, v. 12, No 21: 3482. <https://doi.org/10.3390/rs12213482>

ZANETTI, J.; BRAGA, F. L. S.; SANTOS, A. P. Comparativo das normas de controle de qualidade posicional de produtos cartográficos do Brasil, da ASPRS e da OTAN. *Revista Brasileira de Cartografia*, v. 70, n. 1, 2018, p. 359-390. Doi: 10.14393/rbcv 70n1-45259. Disponível em: <http://www.seer.ufu.br/index.php/revistabrasileira-cartografia/article/view/45259>. Acesso em: 12 set. 2022.

ZANDBERGEN, Paul A.. Positional accuracy of spatial data: Non-normal distributions and a critique of the national standard for spatial data accuracy. *Transactions In GIS*. Nova Jersey, 2008. pp. 103-130.

APÊNDICE A

Código do *GEE* para extrair elevação do AW3D30 para uma camada vetorial de pontos

```
var table = ee.FeatureCollection('projects/ee-wendellfagundes20/assets/REDEALT_IBGE');
var img = ee.ImageCollection("JAXA/ALOS/AW3D30/V3_2").select('DSM')
  var mean_img = img.reduce(ee.Reducer.max())
  var samples = mean_img.reduceRegions({
    'collection': table,
    'scale': 30,
    'reducer': 'max'}).getInfo()
print(samples)
var result=ee.FeatureCollection(samples)

Export.table.toDrive({
  collection: result,
  folder: '2021',
  description:'dadosAlos32',
  fileFormat: 'KML'

});
```