

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Escola de Engenharia
Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos

Alexandre Flávio Assunção

**USO DO SENSORIAMENTO REMOTO PARA AVALIAÇÃO DA DINÂMICA ESPAÇO-
TEMPORAL DE PARÂMETROS DE QUALIDADE DA ÁGUA DE UM RESERVATÓRIO
URBANO: ESTUDO DE CASO NA LAGOA DA PAMPULHA**

Belo Horizonte
2023

Alexandre Flávio Assunção

**USO DO SENSORIAMENTO REMOTO PARA AVALIAÇÃO DA DINÂMICA ESPAÇO-
TEMPORAL DE PARÂMETROS DE QUALIDADE DA ÁGUA DE UM RESERVATÓRIO
URBANO: ESTUDO DE CASO NA LAGOA DA PAMPULHA**

Dissertação apresentada ao Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Saneamento, Meio Ambiente e Recursos Hídricos.

Área de concentração: Recursos Hídricos

Linha de pesquisa: Sistemas de Recursos Hídricos

Orientadora: Dra. Talita Fernanda das Graças Silva

Coorientador: Dr. Lino Augusto Sander de Carvalho

Belo Horizonte
2023

A851u	<p>Assunção, Alexandre Flávio. Uso do sensoriamento remoto para avaliação da dinâmica espaço-temporal de parâmetros de qualidade da água de um reservatório urbano [recurso eletrônico] : estudo de caso na Lagoa da Pampulha / Alexandre Flávio Assunção. – 2023. 1 recurso online (183f.: il., color.) : pdf.</p> <p>Orientadora: Talita Fernanda das Graças Silva. Coorientador: Lino Augusto Sander de Carvalho.</p> <p>Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Engenharia.</p> <p>Apêndices: f. 145-183.</p> <p>Bibliografia: f. 127-144. Exigências do sistema: Adobe Acrobat Reader.</p> <p>1. Engenharia sanitária - Teses. 2. Recursos hídricos - Desenvolvimento - Teses. 3. Detectores óticos - Teses. 4. Monitoramento ambiental - Teses. 5. Água - Qualidade - Teses. 6. Reservatórios - Teses. 7. Pampulha, Lagoa da (MG) - Teses. I. Silva, Talita Fernanda das Graças. II. Carvalho, Lino Augusto Sander de. III. Universidade Federal de Minas Gerais. Escola de Engenharia. IV. Título.</p>
-------	--

CDU: 628(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
[ESCOLA DE ENGENHARIA]
CURSO DE PÓS-GRADUAÇÃO EM [SANEAMENTO, MEIO AMBIENTE E RECURSOS HÍDRICOS]

ATA DA DEFESA DE DISSERTAÇÃO DE **ALEXANDRE FLÁVIO ASSUNÇÃO**
NÚMERO DE REGISTRO: **743**

Às 08:00 horas do dia 02 de junho de 2023, reuniu-se na Escola de Engenharia da UFMG a Comissão Examinadora de DISSERTAÇÃO DE MESTRADO, indicada pelo Colegiado do PROGRAMA DE PÓS-GRADUAÇÃO EM SANEAMENTO, MEIO AMBIENTE E RECURSOS HÍDRICOS, para julgar, em exame final, o trabalho final intitulado:

"USO DO SENSORIAMENTO REMOTO PARA AVALIAÇÃO DA DINÂMICA ESPAÇO-TEMPORAL DE PARÂMETROS DE QUALIDADE DA ÁGUA DE UM RESERVATÓRIO URBANO: ESTUDO DE CASO NA LAGOA DA PAMPULHA"

requisito parcial para a obtenção do Grau de MESTRE em SANEAMENTO, MEIO AMBIENTE E RECURSOS HÍDRICOS, Área de Concentração: HIDRÁULICA E RECURSOS HÍDRICOS

Abrindo a sessão, a Orientadora, Prof^a. Talita Fernanda das Graças Silva, após dar conhecimento aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa pelo candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

BANCA EXAMINADORA	INSTITUIÇÃO	INDICAÇÃO
Prof ^a . Dra. Talita Fernanda das Graças Silva - Orientadora	UFMG	APROVADO
Prof. Dr. Lino Augusto Sander de Carvalho - Coorientador	UFRJ	APROVADO
Prof ^a . Dra. Maria Clara Vieira Martins Starling	UFMG	APROVADO
Prof. Dr. Felipe de Lucia Lobo	UFPEL	APROVADO

Pelas indicações, o candidato foi considerado APROVADO, devendo proceder a pequenas alterações no trabalho, para a sua edição definitiva, a ser entregue no prazo de 2 meses.

O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 02 de junho de 2023.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Talita Fernanda das Gracas Silva, Professora do Magistério Superior**, em 06/06/2023, às 10:17, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maria Clara Vieira Martins Starling, Professora do Magistério Superior**, em 06/06/2023, às 11:00, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Lino Augusto Sander de Carvalho, Usuário Externo**, em 06/06/2023, às 13:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Felipe de Lucia Lobo, Usuário Externo**, em 07/06/2023, às 16:39, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2352913** e o código CRC **A9584578**.

AGRADECIMENTOS

Agradeço aos(às) orientadores(as) do trabalho, Talita e Lino, por todos os ensinamentos, direcionamentos e por acreditarem no meu trabalho. Agradeço à banca participante na defesa da dissertação pela avaliação e comentários.

Agradeço à minha família por todo apoio necessário, principalmente por acreditar em mim repetidas vezes, respeitando os erros cometidos e celebrando as vitórias conquistadas.

Agradeço à Katelyn, que esteve ao meu lado acompanhando a evolução do trabalho e sempre me motivando a persistir.

Agradeço aos amigos(as) que tenho no ambiente acadêmico (UFMG), ambiente social (BD/BH) e ambiente profissional, que mesmo em momentos de desabafos e assuntos não interessantes a eles(as) seguiram como audiência, interessados(as) pela jornada e pelo meu bem-estar.

Agradeço aos projetos de pesquisa que participei, MOSAP e Projeto P&D Aneel/CEMIG2016-GT-607-2019-2023 (SIMOA), em que tive a oportunidade de expandir meus conhecimentos e trabalhar em paralelo com temas relacionados à minha pesquisa.

Agradeço à UFMG, CAPES, FAPEMIG e CNPq pelo amparo concedido para o desenvolvimento da pesquisa. Agradeço ao IGAM, à PBH, à USGS e à ESA pela disponibilidade de dados. Agradeço às disponibilidades de recursos computacionais providos pela Google. Agradeço aos desenvolvedores(as) que fazem parte da comunidade Python pelo desenvolvimento de bibliotecas livres utilizadas no presente trabalho.

*Ah, se o mundo inteiro me pudesse ouvir
Tenho muito pra contar, dizer que aprendi*

(Tim Maia)

RESUMO

A avaliação de padrões e tendências da qualidade da água de reservatórios e lagos é rotineiramente prejudicada pela carência de dados históricos regulares de monitoramento. Técnicas convencionais de monitoramento são limitadas pela baixa representatividade espacial e temporal, dificuldade de acesso e limitação de recursos financeiros. Nos últimos anos, o sensoriamento remoto tem ampliado a representatividade espacial e temporal do monitoramento da qualidade da água por intermédio da reconstrução de séries históricas e do monitoramento da extensão superficial dos corpos hídricos, tornando-se um complemento ao monitoramento *in situ*. Este estudo, por meio de imagens de satélites e respectivos sensores do Landsat-8 (OLI) e Sentinel-2 (MSI), corrigidas por algoritmo de correção atmosférica (SIAC), buscou estimar as concentrações de sólidos suspensos totais (TSM), clorofila-a (Chla), turbidez, e as medidas da profundidade do disco de Secchi (SDD) para um reservatório urbano, a Lagoa da Pampulha (Belo Horizonte, MG, Brasil). Os dados dos sensores foram concatenados com os dados dos programas de monitoramento limnológico históricos realizados no reservatório por órgãos governamentais entre 2013 e 2022. A base de dados foi então modelada por meio do ajuste de modelos existentes na literatura e modelos propostos para o presente trabalho. Para maior robustez na modelagem, investigou-se métodos de regressão linear e não linear, univariados e multivariados, com a implementação de validação cruzada combinada com simulação de Monte Carlo (10,000 repetições). O desempenho dos modelos foi avaliado, na calibração, por meio do coeficiente de determinação (R^2) e, na validação, pela raiz do erro médio quadrático (RMSE) e pelo RMSE normalizado pela média (nRMSE). Os modelos escolhidos do Landsat-8 foram: TSM (n=45) usando *Random Forest* com R^2 de 0,89, RMSE de 7,8 (mg/L) e nRMSE de 28%; Chla (n=49) usando regressão polinomial com R^2 de 0,77, RMSE de 91,8 ($\mu\text{g/L}$) e nRMSE de 95%; SDD (n=23) usando *Random Forest* com R^2 de 0,62, RMSE de 0,1 (m), nRMSE de 30%, e Turbidez (n=16) usando regressão linear com R^2 de 0,8, RMSE de 26,2 (NTU) e nRSMSE de 52%. Os modelos escolhidos do Sentinel-2 foram: TSM (n=53) usando *Random Forest* com R^2 de 0,72, RMSE de 8,5 (mg/L) e nRMSE de 42%, Chla (n=54) usando regressão linear com R^2 de 0,38, RMSE de 51,6 ($\mu\text{g/L}$) e nRMSE de 92%, SDD (n=33) usando *Random Forest* com R^2 de 0,89, RMSE de 0,1 (m) e nRMSE de 15%, e Turbidez (n=9) usando MLR com R^2 de 0,97, RMSE de 3,4 (NTU) e nRMSE de 15%. Os modelos selecionados foram aplicados em imagens sem interferência de nuvens do Landsat-8 e Sentinel-2 derivando uma matriz multidimensional única de dados dos parâmetros de qualidade de água modelados com dimensões (tempo= 155, y=71 e x=95). Por meio de técnica de agrupamento, cinco regiões homogêneas foram identificadas no reservatório e obteve-se diferença significativa (p-valor<5%) na qualidade da água para comparações entre período chuvoso e período seco, períodos em relação ao tratamento químico da Lagoa e anos mais ou menos úmidos, conforme indicação do índice padronizado de precipitação. Observou-se tendência de melhora da qualidade da água para o período modelado (2013-2022) com a ocorrência de redução dos parâmetros TSM, Chla e Turbidez e aumento do SDD, inferida por meio de coeficiente da reta derivado de teste estatístico de *Theil-Sen* combinado com teste de Mann-Kendall Seasonal. A ampliação da quantidade de dados na escala temporal e espacial, mais as análises realizadas com sustentação em informações quantitativas contribuem com diretrizes para estudos futuros e entidades responsáveis pelo gerenciamento do reservatório.

Palavras-chave: sensores ópticos, constituintes opticamente ativos, monitoramento.

ABSTRACT

The assessment of patterns and trends in the water quality of reservoirs and lakes is routinely hindered by the lack of regular historical monitoring data. Conventional monitoring techniques are limited by spatial and temporal representativeness, accessibility challenges, and financial constraints. In recent years, remote sensing has expanded the spatial and temporal representativeness of water quality monitoring through the reconstruction of historical series and monitoring of the surface extent of water bodies, becoming an add-on to in situ monitoring. This study, using satellite images obtained from their sensors from Landsat-8 (OLI) and Sentinel-2 (MSI), corrected by an atmospheric correction algorithm (SIAC), aimed to estimate the concentrations of total suspended matter, chlorophyll-a, and turbidity, as well as Secchi disk depth measurements for an urban reservoir, known as Lagoa da Pampulha, located in Belo Horizonte-MG, Brazil. Satellite-sensor data was combined with historical water quality monitoring data collected by government agencies between 2013 and 2022. The data was then modeled through the adjustment of existing literature models and models proposed by this study. To increase modeling robustness, univariate and multivariate linear and nonlinear regression methods were investigated, with the implementation of cross-validation method combined with Monte Carlo simulation (10,000 repetitions). Model performance was evaluated, in calibration, using the coefficient of determination (R^2), and, in validation, by the root mean square error (RMSE) and nRMSE, which is represent RMSE normalized by the observed sample mean. The chosen models from Landsat-8 were: TSM (n=45) using Random Forest model with R^2 of 0.89, RMSE of 7.8 (mg/L) and nRMSE of 28%; Chla (n=49) using polynomial regression model with R^2 of 0.77, RMSE of 91.8 ($\mu\text{g/L}$), and nRMSE of 95%; SDD (n=23) using Random Forest model with R^2 of 0.62, RMSE of 0.1 (m) and NRMSE of 30%; and Turbidity (n=16) using linear regression model with R^2 of 0.8, RMSE of 26.2 (NTU) and NRMSE of 52%. The chosen models from Sentinel-2 were: TSM (n=53) using Random Forest model with R^2 of 0.72 RMSE of 8.5 (mg/L), and nRMSE of 42%; Chla (n=54) using linear regression model with R^2 of 0.38, RMSE of 51.6 ($\mu\text{g/L}$), and nRMSE of 92%; SDD (n=33) using Random Forest with R^2 of 0.89, RMSE of 0.1 (m) and nRMSE of 15%; and Turbidity (n=9) using multiple linear regression with R^2 of 0.97, RMSE of 3.4 (NTU) and nRMSE of 13%. The selected models were applied to cloud-free images from Landsat-8 and Sentinel-2, resulting in a unique multidimensional array of water quality model parameters with dimensions (time=155, y=71, and x=95). Through cluster analysis, five homogeneous regions were identified in the reservoir, and there was a significant difference ($p\text{-value}<5\%$) in water quality when comparing the rainy and dry seasons, periods related to the chemical treatment of the Lagoon, and wetter or drier years, as indicated by the standardized precipitation index. An improving trend in water quality was observed for the modeled period (2013-2022) with a decrease in TSM, Chla, and turbidity parameters, and an increase in SDD, inferred through the slope coefficient derived from *Theil-Sen* statistical test combined with seasonal Mann-Kendall test. The expansion of data quantity in the temporal and spatial scale, along with the analyses supported by quantitative information, provide guidelines for future studies and entities responsible for reservoir management.

Keywords: optical sensors, optical active constituents, monitoring.

LISTA DE FIGURAS

Figura 1 – Exemplificações das resoluções de sensores ópticos.....	24
Figura 2 – Interação da radiação com a água e seus principais OACs.....	28
Figura 3 – Absorção registrada para os OACs.....	30
Figura 4 – Apresentação da área de estudo	45
Figura 5 – Registros da situação da Lagoa da Pampulha.....	46
Figura 6 – Captura de tela do banco de dados de qualidade de água concatenado das bases do IGAM e PBH	47
Figura 7 – Largura das bandas a serem utilizadas na obtenção de grandezas radiométricas para os sensores embarcados no Landsat-8 (L8) e Sentinel-2 (S2)...	48
Figura 8 – Diagrama esquemático das etapas envolvidas para a modelagem da qualidade da água por dados de sensoriamento remoto	50
Figura 9 – Exemplo de transformação das variáveis categóricas para variáveis numéricas (criação de <i>dummy variables</i>).....	53
Figura 10 – <i>GridSearch</i> combinado com validação cruzada para a determinação dos melhores hiperparâmetros	55
Figura 11 – Exemplo da separação do conjunto amostral em treinamento e teste para uma repetição.....	56
Figura 12 – Extensão da máscara de água respeitando os 50 metros da margem e estruturas presentes no reservatório.....	59
Figura 13 – Exemplo da estrutura da matriz multidimensional com a representação das coordenadas (eixos x e y) e tempo (eixo T).....	59
Figura 14 – Disponibilidade de dados de monitoramento <i>in situ</i> realizado pelo IGAM e PBH na Lagoa da Pampulha.....	63
Figura 15 – <i>Boxplots</i> com a distribuição dos valores dos principais parâmetros de qualidade de água do presente estudo monitorados pelo IGAM e PBH para a Lagoa da Pampulha	64
Figura 16 – Série temporal dos parâmetros de qualidade de água por fonte de monitoramento na Lagoa da Pampulha	65
Figura 17 – Distribuições e correlações entre os parâmetros de qualidade de água para as diferentes fontes de monitoramento da Lagoa da Pampulha	67
Figura 18 – Precipitação total anual para os anos de 1961 a 2023 para a estação INMET 83.587 localizada em Belo Horizonte.....	68

Figura 19 – Índice SPI ₁₂ para os dados de precipitação disponibilizados para a estação INMET 83.587 localizada em Belo Horizonte	69
Figura 20 – Distribuição dos valores de reflectância nos pontos de monitoramento para as bandas dos satélites utilizadas para a modelagem da qualidade da água	71
Figura 21 – Correlação de <i>Pearson</i> para as principais variáveis independentes de entrada para a modelagem de TSM para ambos os satélites.	73
Figura 22 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R ² e RMSE para o parâmetro TSM modelado a partir dos dados do Landsat-8	74
Figura 23 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R ² e RMSE para o parâmetro TSM modelado a partir dos dados do Sentinel-2	75
Figura 24 – Comparação entre valores estimados pelos modelos escolhidos e valores observados de TSM para ambos os satélites.....	76
Figura 25 – Correlação de <i>Pearson</i> para as principais variáveis independentes de entrada para a modelagem de Chla para ambos os satélites.	77
Figura 26 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R ² e RMSE para o parâmetro Chla modelado a partir dos dados do Landsat-8	78
Figura 27 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R ² e RMSE para o parâmetro Chla modelado a partir dos dados do Sentinel-2.	80
Figura 28 – Comparação entre valores estimados pelos modelos escolhidos e valores observados de Chla para ambos os satélites.....	81
Figura 29 – Correlação de <i>Pearson</i> para as principais variáveis independentes de entrada para a modelagem de SDD para ambos os satélites.	83
Figura 30 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R ² e RMSE para o parâmetro SDD modelado a partir dos dados do Landsat-8.	84
Figura 31 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R ² e RMSE para o parâmetro SDD modelado a partir dos dados do Sentinel-2.	85
Figura 32 – Comparação entre valores estimados pelos modelos escolhidos e valores observados de SDD para ambos os satélites.....	86

Figura 33 – Correlação de <i>Pearson</i> para as principais variáveis independentes de entrada para a modelagem do Turbidez para ambos os satélites.....	88
Figura 34 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R ² e RMSE para o parâmetro Turbidez modelado a partir dos dados do Landsat-8	89
Figura 35 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R ² e RMSE para o parâmetro Turbidez modelado a partir dos dados do Sentinel-2	90
Figura 36 – Comparação entre valores estimados pelos modelos escolhidos e valores observados de Turbidez para ambos os satélites	91
Figura 37 – Imagens de satélites utilizadas para a avaliação da dinâmica espacial e temporal da qualidade da água na Lagoa da Pampulha	95
Figura 38 – Série temporal para os parâmetros de qualidade de água (observados e modelados) nos pontos de monitoramento limnológico	98
Figura 39 – Mediana dos parâmetros de qualidade de água calculada por <i>pixel</i> para o período (2013 – 2022).....	100
Figura 40 – Agrupamento hierárquico para os parâmetros de qualidade de água modelados para a Lagoa da Pampulha.....	101
Figura 41 – Teste de <i>Mann-Whitney</i> para os valores medianos de qualidade de água comparando as distribuições de período chuvoso e período seco para cada grupo identificado na análise multivariada.....	102
Figura 42 – Teste de <i>Kruskal-Wallis</i> para a comparação das distribuições das classes de aplicação de produto químico para cada grupo identificado na análise multivariada	103
Figura 43 – Teste de <i>Kruskal-Wallis</i> para a comparação das distribuições das classes do índice SPI ₁₂ para cada grupo identificado na análise multivariada.....	104
Figura 44 – Valores medianos de TSM (mg/L) calculado por <i>pixel</i> considerando período chuvoso e período seco para cada ano durante o período estudado (2013-2022)	106
Figura 45 – Valores medianos de Chla (µg/L) por <i>pixel</i> considerando período chuvoso (6 meses) e período seco (6 meses) para cada ano durante o período estudado (2013-2022)	107

Figura 46 – Valores medianos de SDD (m) por <i>pixel</i> considerando período chuvoso (6 meses) e período seco (6 meses) para cada ano durante o período estudado (2013-2022)	108
Figura 47 – Valores medianos de Turbidez (NTU) por <i>pixel</i> considerando período chuvoso (6 meses) e período seco (6 meses) para cada ano durante o período estudado (2013-2022)	109
Figura 48 – Coeficiente da reta calculado por método de <i>Theil-Sen</i> para o teste de Mann Kendall com sazonalidade para os parâmetros de qualidade de água modelados para a Lagoa da Pampulha	110
Figura 49 – <i>Boxplots</i> com a distribuição dos valores dos parâmetros de qualidade de água modelados apresentados por grupos	121

LISTA DE QUADROS

Quadro 1 – Modelos empíricos e semi-empíricos da literatura	35
---	----

LISTA DE TABELAS

Tabela 1 – Características morfológicas da Lagoa da Pampulha	45
Tabela 2 – Algoritmos da literatura para a estimativa de TSM e Chla adaptados para os satélites utilizados	52
Tabela 3 – Hiperparâmetros testados no <i>GridSearch</i> para os modelos de aprendizado de máquina.....	55
Tabela 4 – Métricas para verificação da acurácia dos modelos.....	57
Tabela 5 – Principais estatísticas descritivas dos parâmetros de qualidade da água monitorados na Lagoa da Pampulha segregados por fonte de monitoramento	65
Tabela 6 – Dados dos satélites combinados com os parâmetros de qualidade da água monitorados (2013 – 2022)	69
Tabela 7 – Principais estatísticas dos parâmetros de qualidade da água monitorados na Lagoa da Pampulha de acordo com a correspondência +-2 dias de passagem do satélite.....	70
Tabela 8 – Modelos selecionados com suas principais características descritas para cada parâmetro e para cada satélite	93
Tabela 9 – Valores máximos e mínimos dos coeficientes da reta calculados por método de <i>Theil-Sen</i> para cada grupo obtido na análise multivariada.....	111

LISTA DE ABREVIATURAS E SIGLAS

λ – Comprimento de onda	NIR – Infravermelho Próximo
ρ_s – Coeficiente de correlação de <i>Spearman</i>	NO_3^- - Nitrato
6SV - <i>Second Simulation of the Satellite signal in the Solar Spectrum</i>	nRMSE – Raiz do Erro Médio Quadrático Normalizado
AOPs - <i>Apparent Optically Properties</i>	NTU – Unidades Nefelométricas de Turbidez
BDMEP – Banco de Dados Meteorológicos	OACs – <i>Optically Active Constituents</i>
CAMS - <i>Copernicus Atmospheric Monitoring Service</i>	OLI – <i>Operational Land Imager</i>
CDOM – <i>Colored Dissolved Organic Matter</i>	OLCI – <i>Ocean and Land Colour Imager</i>
Chla – Clorofila-a	PBH – Prefeitura de Belo Horizonte
ETAF – Estação de Tratamento de Águas Fluviais	PCA – <i>Principal Component Analysis</i>
GEE – <i>Google Earth Engine</i>	PHY – Phytoplankton
IGAM – Instituto Mineiro de Gestão das Águas	PNRH – Política Nacional de Recursos Hídricos
INMET – Instituto Nacional de Meteorologia	PROPAM – Programa de Recuperação e Desenvolvimento Ambiental
IOPs - <i>Inherent Optically Properties</i>	r – Coeficiente de correlação linear de <i>Pearson</i>
Kd – Coeficiente de atenuação difuso	R - Reflectância
MERIS - <i>Medium Resolution Imaging Spectrometer</i>	R ² ajustado – Coeficiente de Determinação ajustado
MLR – Regressão Linear Múltipla	R ² - Coeficiente de Determinação
MODIS - <i>Moderate Resolution Imaging Spectroradiometer</i>	RMSE – Raiz do Erro Médio Quadrático
MODTRAN - <i>Moderate Resolution Atmospheric Transmission</i>	Rrs – Reflectância de sensoriamento remoto
MSI – <i>MultiSpectral Imager</i>	SDD – Profundidade do disco de Secchi
NAPs – <i>Non-Algae Particles</i>	SeaWIFS – <i>Sea-viewing Wide Field-of-view Sensor</i>
NDCI – <i>Normalized Difference Chlorophyll Index</i>	SIAC - <i>Satellite Invariant Atmospheric Correction</i>
NetCDF – <i>Network Common Data Form</i>	SNR – Sinal-ruído
	SPI - <i>Standardized Precipitation Index</i>
	SR - Sensoriamento Remoto

SVR – *Support Vector Machine*

SWIR – *Shortwave Infrared*

TP – Fósforo Total

TSM – Sólidos suspensos totais

UNESCO – Organização das Nações

Unidas para a Educação, a Ciência e a Cultura

VANTs – Veículos Aéreos Não Tripulado

VIIRS – *Visible Infrared Imaging
Radiometer Suite*

E

Ed

Es

Eu

L

Lw

M

MERIS

MODIS

O

OLCI

P

PNRH

R

R

S

SNR

V

VANTs

SUMÁRIO

1	INTRODUÇÃO	19
2	OBJETIVOS	21
2.1	Objetivo Geral	21
2.2	Objetivos Específicos	21
3	REVISÃO DA LITERATURA	22
3.1	Reservatórios urbanos: relevância e monitoramento	22
3.2	O sensoriamento remoto aplicado à cor da água	23
3.3	A interação da radiação com os constituintes presentes na água	27
3.4	Estimativa de parâmetros de qualidade da água por sensoriamento remoto 31	
3.5	Avaliação da qualidade da água de reservatórios a partir de dados de sensoriamento remoto	39
3.6	Uso do sensoriamento remoto para avaliação da dinâmica da qualidade da água para reservatórios urbanos tropicais	42
4	MATERIAIS E MÉTODOS	44
4.1	Área de estudo	44
4.2	Dados limnológicos e climatológicos	46
4.3	Dados de sensoriamento remoto	48
4.4	Processamento dos dados de qualidade de água e sensoriamento remoto	49
4.5	Inferência de parâmetros de qualidade de água a partir de dados do sensoriamento remoto	53
4.6	Aplicação dos modelos para todo o reservatório e para imagens históricas	58
4.7	Avaliação da dinâmica espacial e temporal da qualidade da água na Lagoa da Pampulha	60
4.7.1	Estatística descritiva da dinâmica espaço-temporal da qualidade da água 60	
4.7.2	Análise multivariada para determinação de regiões homogêneas e heterogêneas no reservatório	60
4.7.3	Testes de hipótese para a avaliação da influência de variáveis categóricas externas na qualidade da água	61
4.7.4	Avaliação da tendência da qualidade da água do reservatório	62
5	RESULTADOS	63
5.1	Dados limnológicos, climatológicos e de sensoriamento remoto	63
5.1.1	Dados limnológicos	63
5.1.2	Dados climatológicos	67
5.1.3	Dados de sensoriamento remoto	69
5.2	Inferência de parâmetros de qualidade de água a partir de dados do sensoriamento remoto	71

5.2.1	Sólidos Suspensos Totais (mg/L)	72
5.2.2	Clorofila-a ($\mu\text{g/L}$)	76
5.2.3	Profundidade do Disco de Secchi (m).....	82
5.2.4	Turbidez (NTU)	87
5.2.5	Modelos selecionados e suas principais características	92
5.3	Aplicação dos modelos para todo o reservatório e para imagens históricas	95
5.3.1	Estatística descritiva da dinâmica espaço-temporal da qualidade da água	96
5.3.2	Análise multivariada para determinação de regiões homogêneas e heterogêneas no reservatório	100
5.3.3	Testes de hipótese para a avaliação da influência de variáveis categóricas externas na qualidade da água	101
5.3.4	Avaliação da tendência da qualidade da água do reservatório.....	105
6	DISCUSSÃO	111
6.1	Desempenho dos modelos na inferência de parâmetros de qualidade da água	111
6.2	Fatores, padrões, e tendências da qualidade da água para reservatórios urbanos.....	120
7	CONCLUSÕES	124
8	RECOMENDAÇÕES E PERSPECTIVAS FUTURAS	126
	REFERÊNCIAS	127
	APÊNDICE A – Colunas do banco de dados prévio à modelagem para o Landsat-8	145
	APÊNDICE B – Colunas do banco de dados prévio à modelagem para o Sentinel-2	148
	APÊNDICE C – Todas as imagens entre 2013 e 2022-05-01 para a região da Pampulha para o Landsat-8	153
	APÊNDICE D – Todas as imagens entre 2016 e 2022-05-01 para a região da Pampulha para o Sentinel-2.....	162
	APÊNDICE E – <i>Matchups</i> de imagens do Landsat-8 e Sentinel-2 com cálculo de correlação dos valores modelados para os parâmetros de estudo	178
	APÊNDICE F – Comparação e avaliação da série temporal dos parâmetros de qualidade de água modelados frente os parâmetros monitorados no monitoramento limnológico	179

1 INTRODUÇÃO

Ambientes intermediários entre sistemas lóticos e lênticos, os reservatórios apresentam relevância significativa para a população, haja vista seus usos múltiplos e serviços ecossistêmicos providos (SILVA et al., 2019). A rápida urbanização e a intensificação de atividades humanas em regiões de contribuição a reservatórios aumentam o aporte de nutrientes e sedimentos a esses corpos d'água e contribuem para sua eutrofização. Reservatórios urbanos recebem elevadas concentrações de poluentes com o carreamento de nutrientes e matéria inorgânica por estarem inseridos em contexto de alta densidade populacional associado com a falta de controle da expansão urbana e insuficiências de instrumentos de controle ambiental. Notoriamente, épocas chuvosas com a ocorrência de eventos extremos em contínua expansão, intensificam a poluição mencionada reforçando a necessidade de atenção para o monitoramento de reservatórios urbanos (COELHO, 1998).

A presença de florações de algas e assoreamento são problemas ambientais recorrentes em reservatórios que trazem impactos como a emissão de odores, a degradação da paisagem e a perda de volume útil (FALCONER, 1999). Para a avaliação da qualidade da água e dos impactos causados, estudos, diagnósticos e modelos utilizam-se de dados registrados em campanhas convencionais de monitoramento *in situ* (FURTADO; MONTE-MOR; COUTO, 2021; SILVA et al., 2019). Diversos estudos destacaram a importância da adequação da representatividade espacial e temporal do monitoramento *in situ* para a melhor predição de modelos (SILVA et al., 2016), para o cálculo do estado trófico da água (LOPES et al., 2020) e para a modelagem de impactos da poluição difusa na qualidade da água (SILVA et al., 2019). No entanto, muitas vezes, a grande extensão dos corpos d'água, a dificuldade de acesso e a limitação de recursos desafiam o aprimoramento da representatividade espacial e temporal do monitoramento *in situ*.

Como complemento, o sensoriamento remoto vem sendo aplicado para o monitoramento quantitativo e qualitativo dos recursos hídricos. As interações físicas da luz com os constituintes presentes na água (KIRK, 2011), permite a aplicação de métodos de inferência de parâmetros de qualidade de água, conforme observado inicialmente para a área de oceanografia (GORDON; BROWN; JACOBS, 1975) e com avanços e adaptações realizados para a limnologia (OGASHAWARA; MISHRA;

GITELSON, 2017). Isso em virtude dos avanços de sistemas satélite-sensor que aprimoraram resoluções espectrais, espaciais, radiométricas e temporais, favorecendo inferências de constituintes presentes na água continentais e suas variações de concentração (BARBOSA; NOVO; MARTINS, 2019). Além disso, o sensoriamento remoto por regularmente monitorar os sistemas aquáticos e terrestres vem complementando séries históricas de monitoramento de qualidade da água e o monitoramento para regiões de corpos hídricos sem a presença de monitoramento limnológico *in situ*. As complementações mencionadas favorecem em análises da dinâmica da qualidade da água, incluindo identificação de fatores, tendências e comportamento anômalos (HANSEN et al., 2020).

A Lagoa da Pampulha, reservatório urbano localizado no município de Belo Horizonte, MG, sofre pressões antrópicas há anos e possui classificação hipereutrófica devido ao aporte de nutrientes associados à falta de controle ambiental na área de contribuição e carreamento de sedimentos durante períodos chuvosos (SEIDL et al., 2020). Conhecida por ser patrimônio cultural da UNESCO, a Lagoa da Pampulha vem sendo monitorada por órgãos públicos visando ao atendimento das exigências legais. Apesar de existir monitoramento limnológico no reservatório, o presente estudo busca adaptar os avanços obtidos na área de sensoriamento remoto de sistemas aquáticos para permitir o monitoramento de um reservatório urbano visando a maximização dos dados de qualidade de água disponíveis e a avaliação de sua dinâmica espaço temporal.

2 OBJETIVOS

2.1 Objetivo Geral

Avaliar o uso do sensoriamento remoto para o monitoramento de parâmetros relacionados à qualidade da água em um reservatório urbano.

2.2 Objetivos Específicos

- Avaliar a utilização de regressões univariadas e multivariadas e dados do Sentinel-2 e Landsat-8 na modelagem de dados de qualidade da água;
- Investigar o potencial para maximização, a partir do sensoriamento remoto, da quantidade de dados sobre a qualidade da água de um reservatório urbano;
- Avaliar a dinâmica espaço-temporal de parâmetros ópticos relacionados à qualidade da água em um reservatório urbano a partir de dados de sensoriamento remoto para diferentes cenários: (i) período chuvoso/seco, (ii) tratamento biológico e químico realizado no reservatório e, (iii) precipitação anual.

3 REVISÃO DA LITERATURA

3.1 Reservatórios urbanos: relevância e monitoramento

Nota-se na literatura distinção entre lagos, reservatórios e outros ambientes lênticos baseados nas características geomorfológicas – por exemplo, área superficial e profundidade; e biogeoquímicas – por exemplo, ciclo de nutrientes e estado trófico; que estimulam autores a adotarem o corpo hídrico alvo do estudo como único. Entretanto, o impacto na qualidade da água causada pelo rápido desenvolvimento de áreas de contribuição em reservatórios urbanos merece destaque e distinção (SCHUELER; SIMPSON, 2001).

Diferente de outros ambientes lênticos, os reservatórios urbanos normalmente são rasos com baixa circulação de água apresentando fragilidade ecológica devido à limitada tolerância à poluentes (CHEN et al., 2020; PINEDA-MENDOZA et al., 2020). Esses desempenham papel fundamental no cotidiano urbano, no bem-estar social, na riqueza da biodiversidade e no desempenho da gestão dos recursos hídricos. Dentre as principais vantagens, cita-se a paisagem, lazer (PINEDA-MENDOZA et al., 2020); controle de cheias (HENNY; MEUTIA, 2014); habitat de inúmeras espécies (ZHANG et al., 2020); e conforto térmico (CHEN et al., 2020).

Em consequência da localização geográfica dos reservatórios urbanos, há forte influência da densidade populacional (HENNY; MEUTIA, 2014), grau de impermeabilização (RADOSAVLJEVIC et al., 2022) e atividades potencialmente poluidoras (SCHUELER; SIMPSON, 2001) na qualidade da água, agravada pela falta de planejamento urbano e carência de medidas de controle ambiental tão comuns nos países em desenvolvimento (SEIDL et al., 2020; SUN et al., 2022). Além disso, eventos climáticos extremos, como mudanças nos padrões de chuva e aumento da temperatura, têm intensificado o aporte de nutrientes e sedimentos para corpos hídricos (MAZHAR et al., 2021), e acelerado o processo de eutrofização (VISSER et al., 2016), respectivamente, o que coloca em risco os usos múltiplos dos reservatórios por meio da degradação da qualidade da água e da redução da vida útil (CAI et al., 2021).

Nas últimas décadas, a maioria dos pesquisadores focaram no entendimento da qualidade da água para oceanos e lagos continentais, enquanto os reservatórios urbanos têm recebido menor atenção (CHEN et al., 2020). A deficiência de entendimento de como reservatórios urbanos funcionam leva à ineficiência no gerenciamento desses (POPA et al., 2023). A fim de suprir essa lacuna, o monitoramento contínuo e consistente provê dados quali-quantitativos que promovem o esclarecimento da dinâmica da qualidade da água e apoiam o gerenciamento desses ambientes aquáticos.

Apesar da Política Nacional de Recursos Hídricos (PNRH), instituída pela Lei Federal nº 9.433/1997, prever a gestão sistemática dos recursos hídricos e promover instrumentos de enquadramento e outorga que reforçam a necessidade de monitoramento qualitativo de corpos hídricos (BRASIL, 1997), os programas de monitoramento normalmente estão associados à objetivos específicos, como a avaliação de impactos de descarga de efluentes, o que enfraquece a utilização desses para a avaliação da qualidade da água na escala de bacia. Por isso, abordagens de modelagem da qualidade da água, por exemplo, na utilização de traçadores, sondas e sensoriamento remoto (BEHMEL et al., 2016), complementam o gerenciamento de recursos hídricos (COSTA et al., 2021).

3.2 O sensoriamento remoto aplicado à cor da água

O sensoriamento remoto (SR) é a ciência de inferência de informações sobre um objeto, área ou fenômeno por meio de dados coletados por sensores que não entram em contato direto com os alvos (LORENZZETTI, 2015). Para o caso do SR aplicado à ambientes aquáticos são utilizados sensores embarcados em satélites, aeronaves e veículos aéreos não tripulados (VANTs) (TOTH; JÓZKÓW, 2016). Contudo, existem estudos que utilizam sensores adaptados em suportes (CAO et al., 2022), embarcações (SEIDEL et al., 2020), ou presentes em celulares e radiômetros (ELSAYED et al., 2021).

Sensores ópticos, que são de natureza passiva pelas características de utilização da energia solar, permitem a inferência da interação dos constituintes presentes na água com a radiação solar (OGASHAWARA; MISHRA; GITELSON, 2017). A quantidade e qualidade das informações fornecidas pelo SR dependem das resoluções: espaciais,

espectrais, radiométricas e temporais (Figura 1a-d). A resolução espacial é definida pela projeção do detector no terreno. A resolução espectral representa a quantidade de regiões e a largura de cada região, denominadas bandas, no espectro eletromagnético. A resolução radiométrica é definida como o número de níveis digitais que subdivide a radiância coletada pelo sensor. A resolução temporal representa a frequência de revisita do sensor (BARBOSA; NOVO; MARTINS, 2019).

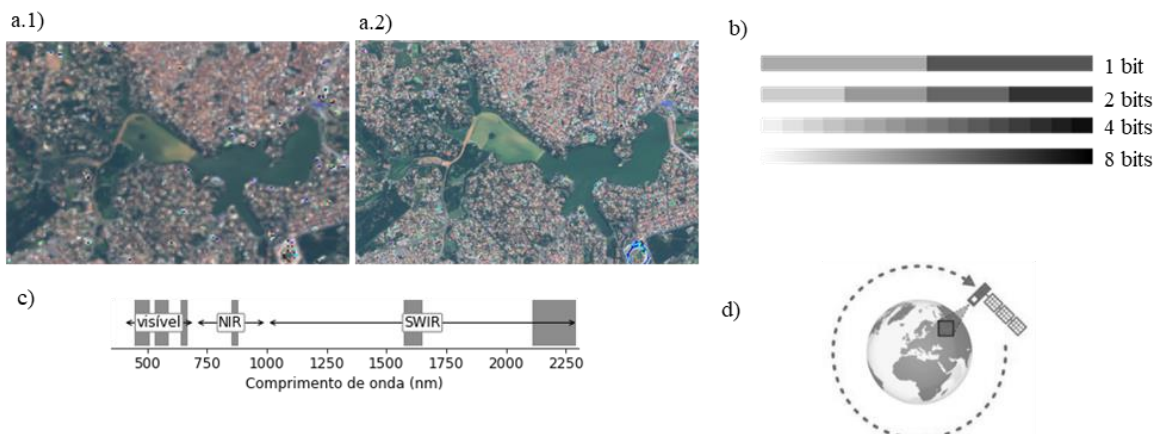
Por meio das informações registradas nos sensores orbitais é possível obter a reflectância de sensoriamento remoto (R_{rs}) (Equação 1) e a reflectância (R) (Equação 2).

$$R_{rs}(\theta, \phi, \lambda) = \frac{L_w(\theta, \phi, \lambda)}{E_s(\lambda)} \quad (1)$$

$$R(z, \lambda) = \frac{E_u(z, \lambda)}{E_d(z, \lambda)} \quad (2)$$

A R_{rs} representa a razão entre a radiância emergente da coluna d'água (L_w) e a irradiância incidente (E_s), caracterizando o campo de luz subaquático e o cálculo da radiação que incide e deixa a superfície do corpo d'água (MOBLEY, 1999). A R , também conhecida como razão de irradiâncias, representa a proporção entre a irradiância espectral ascendente (E_u) e a irradiância espectral descendente (E_d), apresentada na Equação 2 (BARBOSA; NOVO; MARTINS, 2019).

Figura 1 – Exemplificações das resoluções de sensores ópticos



Legenda: a) Diferenças entre resoluções espaciais a.1) 30m e a.2) 10m; b) Níveis de cinza são uma forma de representação da resolução radiométrica; c) Espectro eletromagnético particionado em bandas que representam a resolução espectral; d) A linha pontilhada representa o tempo que o satélite leva para monitorar o mesmo alvo.

Fonte: Próprio Autor (2022); Radiant Earth Foundation (2017).

Entre os principais satélites de que se derivam dados de reflectância, observam-se aqueles que possuem abrangência global, como o Aqua e Terra com o sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS), o Envisat com o sensor *Medium Resolution Imaging Spectrometer* (MERIS) e o Sentinel-3 com o sensor *Ocean and Land Colour Imager* (OLCI). Estes apresentam resolução espacial, nas bandas do visível e infravermelho próximo (VNIR), de 250, 300, 300 m, respectivamente, resolução temporal de 1 a 3 dias (GHOLIZADEH; MELESSE; REDDI, 2016) e se mostram promissores na modelagem de parâmetros de qualidade da água de oceanos e de grandes sistemas aquáticos continentais (GROOM et al., 2019).

Sistemas orbitais com abrangência regional favorecem o monitoramento de corpos d'água de menor porte. Dentre os satélites mais utilizados observa-se a relevância da série Landsat (5 ao 9), por exemplo do Landsat-8 com o sensor *Operational Land Imager* (OLI), e do satélite Sentinel-2 com o sensor *MultiSpectral Instrument* (MSI), cujas imagens disponíveis gratuitamente possibilitam a obtenção de dados com uma resolução espacial, nas bandas VNIR, respectivamente de 30 e 10m, e resolução temporal de 16 dias (Landsat-8) e 5 dias (Sentinel-2 A e B) (GHOLIZADEH; MELESSE; REDDI, 2016).

Sistemas orbitais com abrangência local, por exemplo WorldView-3, apresentam maior resolução espacial (1,2 m para o VNIR) e revisita de menos de 1 dia. Contudo, as imagens fornecidas por estes sensores, normalmente, são comerciais e pouco utilizadas no cunho acadêmico (MORTULA et al., 2020). O uso de sensores aeroportados, como VANTs (DIAS et al., 2021) e sensores *in situ*, como fixados em suportes (pontes, estrutura metálica, barcos) (SUN et al., 2022), vem avançando nos últimos anos. Entretanto tem-se a perda de escala por atuarem em lugares específicos, além das dificuldades de operação e padronização das informações (SIBANDA et al., 2021).

Existem fatores que influenciam nos valores registrados de reflectância pelos sensores abordo dos satélites. Os fatores estão associados com influências físicas associadas a variáveis externas como a presença de ondas na superfície d'água, ocorrência de nuvens e sombra de nuvens, a presença de aerossóis na atmosfera e

a influência do fundo e alvos adjacentes aos sistemas aquáticos (BARBOSA; NOVO; MARTINS, 2019; MOBLEY, 2001). Para a correção dessa influência, observa-se na literatura a utilização de métodos de correção atmosférica – para remoção de efeitos de aerossóis e espalhamento da radiação; correção de reflexão especular (*glint*) – para a remoção de reflexo e espalhamento da radiação presente no espelho d’água e no céu; e correção de efeitos de adjacência – para a remoção da influência da interação com a radiação de alvos terrestres na borda dos sistemas aquáticos na reflectância de alvos aquáticos. Além disso, o sensoriamento remoto de sistemas aquáticos apresenta desafio na inferência de parâmetros de qualidade da água por utilizar regiões de maior comprimento de onda (λ), que coincide com regiões que a água pura apresenta alta absorção configurando baixa relação sinal-ruído (SNR).

Existem inúmeros métodos para a correção atmosférica sobre sistemas aquáticos. O primeiro algoritmo desenvolvido, denominado *Coastal Zone Color Scanner*, objetivava a remoção da interferência da atmosfera no registro de radiância realizado pelos sensores MODIS, *Sea-viewing Wide Field-of-view Sensor* (SeaWiFS) e *Visible Infrared Imaging Radiometer Suite* (VIIRS) para a oceanografia (GORDON, 1978, 2021; GORDON; WANG, 1994). Por utilizar técnica de estimativa de aerossóis coincidente com região do espectro eletromagnético em que a água apresenta alta absorção – “*black pixel assumption* ou *dark spectrum fitting*”, esse método apresenta limitações de correção atmosférica para águas continentais, nas quais partículas em suspensão contribuem para uma maior turbidez e reduzem a capacidade de absorção da água para a região do espectro eletromagnético adotada (NIR) (MOSES et al., 2017).

Vanhellemont e Ruddick (2015) propuseram a adaptação do método *dark spectrum fitting* para a correção atmosférica de dados radiométricos registrados pelo Landsat-8 e Sentinel-2 para sistemas aquáticos continentais que apresentam turbidez. A proposição de utilizar a banda do SWIR corrigida pelo espalhamento *Rayleigh* para águas turvas permitiu aplicação do método para águas continentais, como no caso do estudo realizado para a região de várzea em rios turvos amazônicos (MARTINS et al., 2017). Contudo, limitações são observadas para o método, como o desempenho obtido comparado à modelos físicos, a baixa relação SNR para a banda utilizada e

extremidade mais curta da banda do SWIR para alguns sensores (MOSES et al., 2017).

Os modelos físicos, como o *Second Simulation of the Satellite signal in the Solar Spectrum* (6SV) (VERMOTE et al., 1997) e *Moderate Resolution Atmospheric Transmission* (MODTRAN) (BERK et al., 1999), baseiam-se na equação de transferência radiativa por meio da estimativa de espalhamento e absorção da atmosfera. A dificuldade de utilização desses modelos é a necessidade de conhecimento prévio sobre os parâmetros atmosféricos correspondentes as datas de passagem dos satélites. Mesmo que estes dados estejam disponíveis a partir de modelos climatológicos, existem adaptações para a utilização de produtos atmosféricos derivados de satélites, como no caso do *Satellite Invariant Atmospheric Correction* (SIAC) que utiliza dados climatológicos derivados do produto MCD43A3 do MODIS e do produto *Copernicus Atmospheric Monitoring Service* (CAMS) (YIN; LEWIS; GÓMEZ-DANS, 2022). O SIAC sobressai em relação a modelos empíricos devido sua capacidade de invariância em relação aos satélites e por estar disponível para ser aplicado em imagens do Sentinel-2 e Landsat-8 direto na plataforma de computação em nuvem *Google Earth Engine* (GEE) (YIN; LEWIS; GÓMEZ-DANS, 2022).

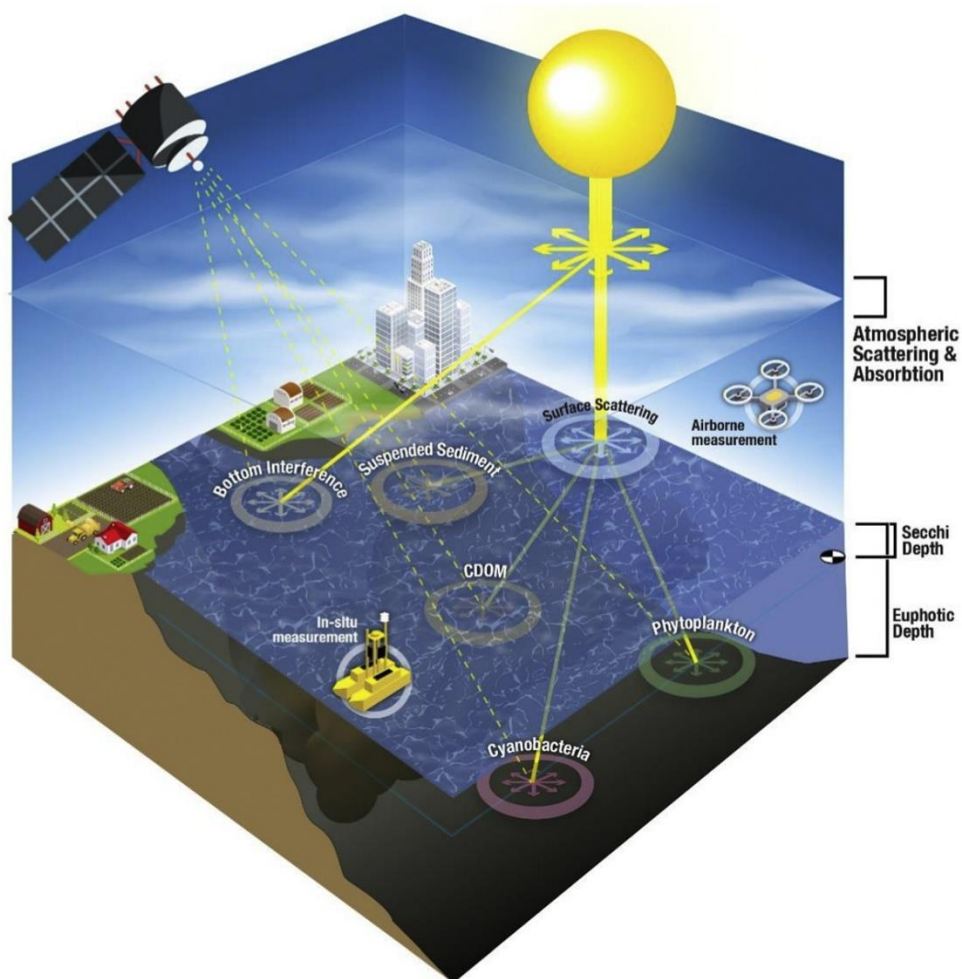
A plataforma do GEE permite análises espaciais em escala global aplicadas aos dados de sensoriamento remoto armazenados e processados por arquitetura de computação em nuvem disponibilizada pela *Google®* (GORELICK et al., 2017). Com lançamento em 2010, o GEE vem sendo utilizado amplamente no monitoramento de desmatamento, secas, incêndios, mudanças climáticas e gestão de recursos hídricos (YANG et al., 2022). Para o último item observa-se crescente utilização da plataforma para o monitoramento quali-quantitativo da água por meio de dados disponibilizados do Landsat-8 e Sentinel-2 na plataforma (PEKEL et al., 2016; WANG et al., 2020).

3.3 A interação da radiação com os constituintes presentes na água

No meio aquoso (Figura 2), encontram-se frações de constituintes suspensos e dissolvidos que possuem a capacidade de interação com a radiação, denominados constituintes opticamente ativos (OACs- *optically active constituents*) (MOBLEY, 1994). Os modelos bio-ópticos buscam inferir a concentração de OACs presentes na

água que interagem com a radiação solar por meio das (IOPs - *inherent optically properties*) e propriedades ópticas aparentes (AOPs - *apparent optically properties*). As IOPs, representadas pela absorção e retroespalhamento da radiação, não sofrem interferências externas e dependem apenas da composição e concentrações dos OACs. As AOPs representam relações entre grandezas radiométricas, como a reflectância de sensoriamento remoto, e o coeficiente de atenuação difuso (K_d), e sofrem interferência do meio e do comportamento da luz no ambiente (KIRK, 2011; MOBLEY, 1994).

Figura 2 – Interação da radiação com a água e seus principais OACs



Fonte: (SAGAN et al., 2020)

Os principais OACs presentes nos corpos hídricos são a água pura, a matéria orgânica dissolvida colorida (CDOM – *colored dissolved organic matter*), as partículas não algais (NAPs – *non algae particles*) e as partículas algais (PHY – *phytoplankton*), conhecidas como fitoplâncton (BRICAUD; ROESLER; ZANEVELD, 1995).

A água pura apresenta maior absorção da radiação a partir de 550 nm, que aumenta à medida que o λ aumenta. O retroespalhamento da água pura é alto para menores λ , correspondente à faixa do azul (KIRK, 2011).

O CDOM representa a parcela colorida do carbono orgânico dissolvido no meio aquoso formado por constituintes terrestres que são carregados para os corpos hídricos, podendo ser de fontes alóctones ou autóctones (SIEGEL et al., 2002). O CDOM apresenta forte absorção em pequenos λ e decai exponencialmente com o crescimento do λ (Figura 3d). Não se contabiliza o retroespalhamento do CDOM em função dos baixos valores registrados (RIDDICK et al., 2015).

Sedimentos minerais, orgânicos e inorgânicos presentes no meio aquoso são denominados NAPs e representam a parcela das partículas sólidas totais em suspensão (TSM – *total suspended matter*). O espectro de absorção do NAP apresenta função exponencial decrescente ao longo do espectro eletromagnético (Figura 3b), com dependência da composição e concentração orgânica ou inorgânica do NAP (BARBOSA; NOVO; MARTINS, 2019). O retroespalhamento dos NAPs ocorre principalmente em função da presença de frações inorgânicas e apresenta redução à medida que avança no espectro (KIRK, 2011).

Os organismos fitoplanctônicos (PHY) são rotineiramente estimadas pela presença da clorofila-a (Chla) (GHOLIZADEH; MELESSE; REDDI, 2016). Esta possui alta absorção nas bandas do azul (440 nm) e do vermelho (670 nm) (Figura 3c) (KIRK, 2011). Em águas com altas concentrações de CDOM e NAP, o pico de absorção na região do azul, normalmente utilizado para a estimativa de PHY em oceanos, é mascarado pela absorção de CDOM e NAPs (Figura 3a), levando autores a buscarem regiões do vermelho e da borda do vermelho (660-750 nm), onde os valores de absorção de CDOM e NAPs são menores, evidenciando melhor a absorção por PHY (DALL'OLMO; GITELSON, 2005; MISHRA; MISHRA, 2012). O espectro de retroespalhamento é fortemente dependente da forma, tamanho e índice de refração das células que compõem o fitoplâncton (BARBOSA; NOVO; MARTINS, 2019).

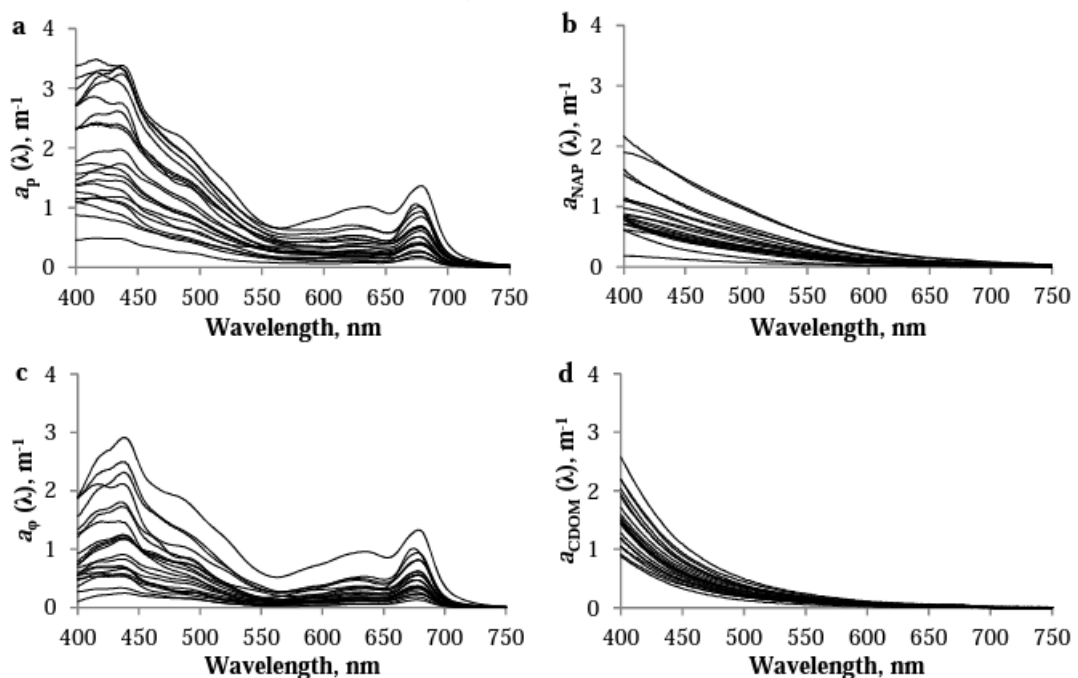
O TSM está associado com NAPs e organismos clorofilados microscópicos (fitoplâncton) para águas continentais (BARBOSA; NOVO; MARTINS, 2019). A Chla está presente em todas as espécies de fitoplâncton. Além desses, o parâmetro

turbidez e a medida da profundidade do disco de Secchi (SDD – *Secchi disk depth*) também são estimados por meio do sensoriamento remoto devido às contribuições que os OACs apresentam nesses e ambos interpretarem a capacidade de propagação da luz em ambientes aquáticos (DÖRNHÖFER; OPPELT, 2016; MATTHEWS, 2011). Além disso, a transparência da água, relacionada com o SDD, apresenta correlação inversa ao K_d (LEE et al., 2015).

O SDD é a medida da transparência que corresponde à profundidade em que o disco imerso na água não consegue mais ser visualizado da superfície medida em unidade de distância. Esta medida indica a quantidade de luz que é capaz de penetrar a coluna d'água, normalmente influenciada por partículas em suspensão, matéria orgânica dissolvida e fitoplâncton (GHOLIZADEH; MELESSE; REDDI, 2016).

A Turbidez é a medida da transparência da água derivada da quantidade de luz espalhada por constituintes presentes na água. Dentre os constituintes, cita-se a contribuição de matéria inorgânica, como sedimentos inorgânicos em suspensão, e matéria orgânica, como PHY e resíduos orgânicos. A medida da Turbidez é realizada por turbidímetro e as unidades obtidas são expressas em unidades nefelométricas de turbidez (NTU) (KITCHENER; WAINWRIGHT; PARSONS, 2017).

Figura 3 – Absorção registrada para os OACs



Legenda: Espectro representativo de absorção dos constituintes para 21 amostras que tiveram Chla variando entre 2,3 e 132,4 ($\mu\text{g/L}$): (a) todos os OACs, (b) NAPs, (c) PHY, (d) CDOM.

Fonte: (GURLIN, 2012)

Referidas IOPs se relacionam com as AOPs por meio da equação da teoria de transferência radiativa. A solução da equação permite a inferência das grandezas radiométricas como irradiâncias incidente e emergente que posteriormente podem ser derivadas nas AOPs (MOBLEY, 1994). Entretanto, mencionada equação apresenta alta complexidade (MOBLEY, 1994). Gordon, Brown e Jacobs (1975) propuseram a simplificação da equação para a aplicação na oceanografia e posteriormente foi complementada para águas continentais por Whitlock *et al.* (1981), Gordon *et al.* (1988), Morel e Gentili (1991) e Kirk (2011). A equação proposta relaciona a reflectância de sensoriamento remoto com as principais IOPs, absorção e espalhamento (Equação 3).

$$R_{rs}(\lambda) = f/Q \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \quad (3)$$

Onde f é um coeficiente dependente do ângulo zenital solar e da função de retroespalhamento do volume, Q representa a razão entre a irradiância emergente de subsuperfície e a radiância emergente de subsuperfície no campo de visão, $b_b(\lambda)$ e $a(\lambda)$ representam o retroespalhamento e a absorção, respectivamente.

3.4 Estimativa de parâmetros de qualidade da água por sensoriamento remoto

Os modelos bio-ópticos utilizados para se estimar parâmetros de qualidade da água podem ser considerados de natureza empírica (MOREL, 2001) ou analítica (LEE; CARDER; ARNONE, 2002; NECHAD; RUDDICK; PARK, 2010). Ogashawara (2015) determinou as seguintes classificações para os modelos: (i) modelos empíricos são orientados a dados, (ii) modelos semiempíricos são orientados a dados com a tentativa de replicar comportamentos físicos, (iii) modelos semianalíticos e quasianalíticos utilizam métodos físicos com a combinação de dados de monitoramento e (iv) modelos analíticos retratam apenas comportamentos físicos. Além das diferentes classificações dos modelos bio-ópticos, a construção da relação entre OACs e AOPs pode se dar através de métodos diretos ou inversos. O primeiro se constrói por meio da estimativa das concentrações dos OACs presentes na água, derivação dessas para as IOPs seguido da modelagem das AOPs. O segundo faz o

processo inverso, onde as AOPs são as variáveis monitoradas e as IOPs e concentrações dos OACs modeladas (DEKKER; VOS; PETERS, 2002).

Apesar de modelos analíticos apresentarem maior capacidade de generalização e aplicabilidade para outras regiões, esses necessitam do levantamento de inúmeras variáveis em campo e apresentam difícil solução da equação devido a interação de diversos constituintes (MATTHEWS, 2011; TOPP et al., 2020). Por isso, nota-se na literatura a utilização de dados *in situ* de OACs para o ajuste de modelos semianalíticos, como o estudo desenvolvido por Nechad, Ruddick e Park (2010) que utilizaram dados de radiômetro combinados com medidas de TSM *in situ* para a construção do algoritmo semianalítico para a estimativa do parâmetro.

Modelos semiempíricos e empíricos se destacam graças às possibilidades de se trabalhar apenas com as AOPs registradas pelos satélites e dados de qualidade monitorados *in situ* (MATTHEWS, 2011). Dados simultâneos – que coincidem a passagem do satélite com a realização do monitoramento *in situ*; ou quase simultâneos – diferença de alguns dias entre passagem e coleta; são modelados por meio de regressão para a estimativa de parâmetros de qualidade de água (ODERMATT et al., 2012). Apesar da simplicidade descrita, modelos empíricos frequentemente apresentam boas métricas apenas para os ambientes aquáticos em que os dados utilizados na modelagem foram obtidos, em razão de fatores físicos específicos de cada local e que alteram as características espectrais (HANSEN et al., 2020; MATTHEWS, 2011; TOPP et al., 2020).

Ainda que o interessante seja a ocorrência do monitoramento *in situ* na mesma data de passagem do satélite, alguns estudos mostram resultados satisfatórios quando utilizados dados quase simultâneos, especialmente se a quantidade de amostras aumenta (HANSEN et al., 2015). Zhang *et al.* (2021a) realizaram a comparação entre diferença de datas de +- 3 dias e +- 7 dias e observaram o aumento da quantidade de pares de dados, com pouca alteração nas métricas de desempenho obtidas para a estimativa de SDD por meio de imagens da série Landsat. Lobo *et al.* (2021) compararam o desempenho do modelo para estimativa de Chla por meio de imagens do Sentinel-2 corrigidas pelo método SIAC e perceberam que o intervalo de dois dias

entre a passagem do satélite e o monitoramento *in situ* foi o conjunto que apresentou a melhor relação entre quantidade de amostras ($n=136$) e performance ($R^2 = 0,86$).

Modelos empíricos e semiempíricos são encontrados vastamente na literatura e partem de modelos simples, como regressões lineares (ONDERKA; PEKÁROVÁ, 2008), até modelos mais complexos como regressões não lineares (ZHAO et al., 2020), modelos de aprendizado de máquina (ARIAS-RODRIGUEZ et al., 2021) e modelos de aprendizado profundo (CUI et al., 2022). No Quadro 1, apresenta-se exemplos de modelos que utilizaram dados de sensoriamento remoto (variáveis independentes) para estimar parâmetros de qualidade de água (variáveis dependentes).

Observa-se dentre os estudos a utilização de modelos simples como regressão linear univariada e multivariada, partindo da premissa que as variáveis apresentam comportamento linear (ONDERKA; PEKÁROVÁ, 2008). Em caso de variáveis não lineares, recorre-se para métodos não lineares que permitem a linearização dos dados, como o caso da regressão exponencial, ou para soluções não lineares como regressão polinomial (MISHRA; MISHRA, 2012), logarítmica (ZHAO et al., 2020) e potencial (LOBO et al., 2021).

Além dos modelos listados anteriormente, nota-se que avanços recentes na área de computação fortaleceram a utilização de diversas variáveis de forma conjunta para a construção de modelos de aprendizado de máquina, como o *Support Vector Regressor* (SVR) e *Random Forest Regressor* (YANG et al., 2022). Diferente dos métodos que utilizam mínimos quadrados para minimizar o erro entre valores preditos e observados, o método SVR busca ajustar um hiperplano para identificar a dependência linear entre os vetores n -dimensionais (variáveis independentes) e o vetor 1-dimensional de saída (variável dependente) (CERVANTES et al., 2020). O *Random Forest* representa um conjunto de modelos de árvores de decisão, que são formadas por nós e folhas com base nas variáveis independentes de entrada e cálculos de redução da variância para o avanço da profundidade da árvore (BREIMAN, 2001). As vantagens de modelos como SVR e *Random Forest* são a capacidade de suportar base de dados desbalanceadas, com presença de outliers e valores nulos (WAGLE; ACHARYA; LEE, 2020). Entretanto, além da complexidade de

implementação e execução desses modelos, observa-se as dificuldades desses em extrapolar informações desconhecidas da etapa de calibração.

O comportamento e desempenho dos modelos de aprendizado de máquina são influenciados pelos hiperparâmetros. Esses devem ser configurados pelos especialistas de domínio visando o melhor ajuste e desempenho (YANG et al., 2022). Entretanto, visto a extensa quantidade de hiperparâmetros e as inúmeras opções existentes, por exemplo, parâmetro de regularização, número máximo e profundidade máxima de árvores de decisão, *kernel*, entre outros, nota-se na literatura a utilização de métodos combinados de *Grid Search* e validação cruzada para a determinação dos hiperparâmetros (NIU et al., 2021).

O *Grid Search* de forma exaustiva busca dentre os hiperparâmetros a serem testados informados pelo usuário quais apresentam o melhor desempenho (PEDREGOSA et al., 2011). A validação cruzada permite a divisão do conjunto amostral em grupos de calibração e validação. Diferente de uma separação aleatória entre dois grupos de calibração e validação, a validação cruzada visa a definição de “k” subgrupos que posteriormente são alternados entre calibração e validação para garantir a generalização dos modelos (ARIAS-RODRIGUEZ et al., 2020).

Outra forma de se avaliar a capacidade de generalização dos modelos é por meio de inúmeras repetições de divisão dos dados em calibração e validação, nomeada simulação de Monte Carlo e utilizada por (MACIEL et al., 2019). Augusto-Silva et al. (2014) combinaram essa simulação com técnica de amostragem do conjunto de validação por equivalência de quartis, que permite o equilíbrio e garante a presença de toda a amplitude dos valores observados.

Quadro 1 – Modelos empíricos e semi-empíricos da literatura

Ref.	Variáveis independentes			Variáveis dependentes		Modelo			
	Satélite (Sensor)	Corr. Atm	Variáveis independentes (x)	Variável dependente (y)	Amplitude	n	Janela +- de dias	Método de regressão	Equação
(NECHAD; RUDDICK; PARK, 2010)	Envisat (MERIS)	Sim	$\frac{\rho(708)}{1 - \rho(708)/18,86}$	TSM	1,24 - 110,27 (mg/L)	72	0	Regressão não linear	$y = 537x + 1,15$
(ONDERKA; PEKÁROVÁ, 2008)	Landsat-7 (ETM)	Não	$\rho(830)$	TSM	19,5 - 57,5 (mg/L)	10	-	Regressão linear	$y = 4,17x - 43,22$
(CHEN et al., 2021)	Sentinel-2 (MSI)	Sim	$\frac{\rho(833) - \rho(665)}{\rho(833) + \rho(665)}$	TSM	-	65	0	Função de potência	$y = 0,99x^{-0,28}$
(ZHAO et al., 2020)	Landsat-8 (OLI)	Sim	$[\rho(480)^{-1} - \rho(560)^{-1}] \times \rho(655)$	TSM	5 - 50 (mg/L)	49	-	Regressão logarítmica	$y = 16,5\ln(x) - 8,4$
(SABERIOON et al., 2020)	Sentinel-2 (MSI)	Sim	8 bandas do MSI + 19 índices da literatura	TSM	2 - 195 (mg/L)	131	0	Aprendizado de máquina	-
(ARIAS-RODRIGUEZ et al., 2021)	Landsat-8 (OLI) Sentinel-2 (MSI)	Sim	7 bandas do OLI e 8 bandas do MSI	TSM	10 - 475 (mg/L)	134	3	Aprendizado de máquina	-
(GITELSON et al., 2008)	Envisat (MERIS)	Sim	2 bandas $\rho(665)^{-1} \times \rho(725)$	Chla	1 - 240 (µg/L)	145	-	Regressão linear	$y = 136,3x - 16,2$
(GITELSON et al., 2008)	Envisat (MERIS)	Sim	3bandas $\rho(671)^{-1} \times \rho(710)^{-1}] \times \rho(740)$	Chla	1 - 240 (µg/L)	145	-	Regressão linear	$y = 117,4x + 23,4$
(MISHRA; MISHRA, 2012)	Envisat (MERIS)	Sim	NDCI $\frac{\rho(708) - \rho(665)}{\rho(708) + \rho(665)}$	Chla	0,9 - 28 (µg/L)	29	0	Regressão polinomial	$y = 194,3x^2 + 86,1x + 14,04$

Ref.	Variáveis independentes			Variáveis dependentes		Modelo			
	Satélite (Sensor)	Corr. Atm	Variáveis independentes (x)	Variável dependente (y)	Amplitude	n	Janela +- de dias	Método de regressão	Equação
(LOBO et al., 2021)	Sentinel-2	Sim	$\frac{\rho(704) - \rho(665)}{\rho(704) + \rho(665)}$	Chla	0,6 - 486,2 (µg/L)	136	2	Função de potência	$y = 23,4(x + 1)^{7,95}$
(ARIAS-RODRIGUEZ et al., 2021)	Landsat-8 (OLI) Sentinel-2 (MSI)	Sim	7 bandas do OLI e 8 bandas do MSI	Chla	0 - 218 (µg/L)	134	3	Aprendizado de máquina	-
(ZHANG et al., 2021a)	Landsat-5, Landsat-7, Landsat-8	Sim	$\rho(\sim 650)$	SDD	0,04 - 14 (m)	864	1	Função de potência	$y = 0,0046x^{-1,26}$
(ARIAS-RODRIGUEZ et al., 2021)	Landsat-8 (OLI) Sentinel-2 (MSI)	Sim	7 bandas do OLI e 8 bandas do MSI	SDD	0,1 - 2,3 (m)	134	3	Aprendizado de máquina	-
(HOSSAIN; MATHIAS; BLANTON, 2021)	Landsat-8	Sim	$\rho(655)$	Turbidez	3,1 - 88,8 (NTU)	108	0	Função de potência	$y = 2677x^{1,86}$
(ARIAS-RODRIGUEZ et al., 2021)	Landsat-8 (OLI) Sentinel-2 (MSI)	Sim	7 bandas do OLI e 8 bandas do MSI	Turbidez	0,3 - 750 (NTU)	134	3	Aprendizado de máquina	-

Nota: ρ : reflectância, "-": não existe equação pré-definida por utilizar modelo de aprendizado de máquina ou não foi possível identificar a janela temporal com o intervalo de dias

Fonte: Próprio Autor (2023)

Como os estudos apresentados buscaram a estimativas de parâmetros de qualidade de água em águas turvas, observa-se a preferência de variáveis independentes com maiores comprimentos de onda, como a região do vermelho (660nm) até a região do NIR (840nm). Além disso, é comum observar a ocorrência de operações matemáticas entre bandas, como multiplicação, razão (DALL'OLMO; GITELSON, 2005) e construção de índices (MISHRA; MISHRA, 2012), para tentar realçar as relações entre variáveis dependentes e independentes.

A descoberta de regiões que melhor representam parâmetros de qualidade de água segue embasamento físico, conforme observado nos modelos desenvolvidos por Nechad, Ruddick e Park (2010) – algoritmo de Nechad; por Dall'omo e Gitelson (2005) – modelo de 2bandas e 3bandas; e por Mishra e Mishra (2012) – *Normalized Difference Chlorophyll Index* (NDCI). Todos esses utilizaram informações coletadas com radiômetros em campo para a identificação da região do espectro eletromagnético que apresenta correlação com o parâmetro de interesse. Após a identificação da região ótima para a modelagem, os estudos realizaram a adaptação do algoritmo para a correspondência com informações de satélites, por exemplo do sensor MERIS a bordo do Envisat, e a verificação da aplicabilidade dos modelos (GITELSON et al., 2008; MISHRA; MISHRA, 2012; NECHAD; RUDDICK; PARK, 2010).

A aplicabilidade, o desempenho e a fundamentação física dos modelos, fizeram com que esses fossem amplamente utilizados na literatura para a estimativa de TSM e Chla. Houve adaptações para outros satélites, como na adaptação do algoritmo NDCI realizado por Lobo *et al.* (2021) para utilizarem bandas do Sentinel-2 na estimativa de Chla, adaptação feita por Watanabe *et al.* (2019) para a estimativa de Chla pelos modelos 2bandas, 3bandas e NDCI por meio de dados de reflectância do Sentinel-2 simulados através de dados de radiômetros, e na adaptação dos algoritmos de Nechad e NDCI realizado por Coimbra, Alcântara e Souza Filho (2021) para utilizarem bandas do Landsat-8 na estimativa de TSM e Chla. Deve-se tomar cuidados com a adaptação dos modelos 2bandas, 3bandas e NDCI para o satélite Landsat-8 devido a lacuna de bandas para a região entre o vermelho (655nm) e NIR (865nm) (CAO et al., 2020).

Outra forma de identificar regiões do espectro eletromagnético e relações entre bandas ideais para o monitoramento é por meio de abordagem empírica, por exemplo, no cálculo da correlação entre variáveis dependentes e independentes combinada com inúmeras iterações promovidas por linguagem de programação (MATTHEWS, 2011; YANG et al., 2022). Saberioon *et al.* (2020) identificaram que a utilização de índices que usavam a banda do SWIR do Sentinel-2 ajuda no desempenho de modelos empíricos para a estimativa de TSM e Chla, mesmo que essa banda apresente alta absorção da radiação (KIRK, 2011).

Problemas de regressão podem ser otimizados por meio da utilização de variáveis categóricas, como a descrição de período chuvoso ou seco. A transformação das variáveis categóricas para variáveis discretas pode ser realizada por método denominado “*dummy variables*” que se traduz na atribuição de valores numéricos as classes observadas (ANDERSON, 1984). Por exemplo, no trabalho de Souza (2019) constatou-se que a separação da modelagem de acordo com a cor da água apresentava melhor desempenho para um reservatório que tinha uma estrutura física de contenção de sedimentos. Outro trabalho notou que a separação por período chuvoso ou seco pode assimilar efeitos de sazonalidade das propriedades ópticas aparentes e inerentes causados pela afluência de diferentes materiais orgânicos e inorgânicos ao longo do ano (BRANDÃO; STAEHR; BEZERRA-NETO, 2016).

Existem outros parâmetros de qualidade da água que não possuem características ópticas, como o nitrogênio, fósforo e oxigênio dissolvido, mas que podem estar correlacionados aos parâmetros bio-ópticos. Por isso, nota-se na literatura avanços na tentativa de monitoramento desses parâmetros que não interagem com a luz por meio de técnicas empíricas (HAJIGHOLIZADEH et al., 2021).

3.5 Avaliação da qualidade da água de reservatórios a partir de dados de sensoriamento remoto

Observa-se na literatura estudos que priorizam a construção de modelos e avaliação de desempenho desses na estimativa de parâmetros de qualidade da água por sensoriamento remoto (TOPP et al., 2020). Outros, preferem avaliar a aplicação do modelo para imagens de outras datas em que não houve monitoramento para a demonstração da aplicabilidade do modelo em preencher lacunas nas dimensões espaciais e temporais. De forma minoritária, constata-se estudos que fazem a aplicação dos modelos de qualidade de água para séries temporais de imagens e avaliam a dinâmica da qualidade da água (TOPP et al., 2020).

Zhao *et al.* (2020) identificaram a ocorrência de maiores concentrações de TSM na parte de montante do reservatório *Hedi* na China e que as concentrações de TSM apresentaram crescimento entre os anos de 2014 e 2016 concomitante ao avanço para regiões de jusante. Nos outros dois anos (2016-2018) observaram retrocesso na concentração de TSM e o menor valor de concentração para o período avaliado. Os dados de TSM modelados foram possíveis com a utilização de imagens do Landsat-8 entre 2014 e 2018.

Watanabe *et al.* (2015) estimaram concentrações de Chla para o reservatório de Barra Bonita por meio de razão de bandas ($\rho_{\text{NIR}}/\rho_{\text{Verde}}$) do Landsat-8 e verificaram que as concentrações de Chla do mês de janeiro de 2014 superou os outros meses analisados (maio/2013, junho/2013, agosto/2013, setembro/2013, setembro/2014 e outubro/2014). Dörnhöfer *et al.* (2018) estudaram o *Lake Kummerow* na Alemanha e verificaram que as concentrações de Chla estimadas por aplicação dos modelos obtidos para a abordagem multisensor (ETM+, OLI, MSI, MODIS) indicavam a evolução de eventos de floração de algas.

Olmanson, Bauer e Brezonik (2008) mapearam SDD para diversas regiões do estado de Minnesota nos Estados Unidos. Com a série do Landsat, imagens entre 1985 e 2005 foram modeladas derivando os valores de SDD que posteriormente indicaram que a transparência tende a ser baixa para a região sul e sudoeste do estado e alta para a região norte e nordeste do estado.

Quang *et al.* (2017) modelaram o parâmetro Turbidez para um estuário no Vietnã. Para a modelagem foram utilizados pares de dados de monitoramento *in situ* e imagens do Landsat-8 para o período de 2013 a 2016. Além de comparar as concentrações de Turbidez com distinção de período seco e período chuvoso, os autores também avaliaram os impactos do vento e ondas. Os resultados indicaram que as maiores concentrações de turbidez ocorrem em eventos de precipitação esporádicos no período seco, ou durante o período chuvoso.

Eventos atípicos e de grande escala, como a ocorrência de tufões, pandemia e rompimento de barragens podem impactar diretamente a qualidade de água. Caballero e Navarro (2021) mostraram a influência de tufões que ocorreram nas Filipinas na qualidade da água do *Laguna Lake* modelada com dados do satélite Sentinel-2. Os tufões influenciaram no aporte de TSM e intensificaram florações de alga na região de *Central Bay*, localizada na parte central do corpo hídrico. A pandemia do covid-19 permitiu aferir a melhora da qualidade da água de reservatórios localizados no Brasil, Itália e China entre períodos anteriores e durante o *lockdown* por meio de imagens de satélite (ALCÂNTARA *et al.*, 2021; TETA *et al.*, 2021; WANG *et al.*, 2022; YUNUS; MASAGO; HIJIOKA, 2020). Os rompimentos das barragens de Fundão e B1 – Córrego do Feijão, em Minas Gerais, e a avaliação dos impactos na qualidade da água foram complementados por estudos que utilizaram dados do Landsat-8 e Sentinel-2 (COIMBRA; ALCÂNTARA; DE SOUZA FILHO, 2021; SOUZA *et al.*, 2021).

Existem análises estatísticas que permitem auxiliar na inferência e entendimento da dinâmica da qualidade da água, como testes estatísticos não paramétricos, análise multivariada e análise de tendência (SPERLING; VERBYLA; OLIVEIRA, 2020). Testes estatísticos, como os testes de *Mann-Whitney* (MANN, 1945) e *Kruskal-Wallis* (KRUSKAL; WALLIS, 1952) referem-se a testes de hipóteses que visam ranqueamento e comparação entre grupos amostrais para indicar se esses fazem parte da mesma população. Análises multivariadas, como análise de agrupamentos (*cluster*) e análise de componentes principais (PCA - *principal component analysis*), funcionam como métodos exploratórios que auxiliam na interpretação e extração de informações de um conjunto amostral grande (TRINDADE *et al.*, 2017). Análises de tendência permitem a avaliação do comportamento de variáveis de qualidade da água

ao longo do tempo, por exemplo tem-se o método de Mann-Kendall (KENDALL, 1948; MANN, 1945), que posteriormente foi ajustado para diminuir efeitos de sazonalidade (HIRSCH; SLACK; SMITH, 1982) por meio do método denominado Mann-Kendall Seasonal. A combinação do método de Mann-Kendall com a associação do método de *Theil-Sen's slope* (THEIL, 1992) é amplamente utilizado em pesquisas atmosféricas e hidrológicas (COLLAUD COEN et al., 2020).

Para o sensoriamento remoto, Hansen *et al.* (2020) avaliaram o comportamento do parâmetro Chla modelado para 32 anos de imagens da série de satélites Landsat para três lagos na região oeste dos Estados Unidos e compararam padrões de médio prazo (estações do ano) e padrões de longo prazo (toda série temporal). Por meio de testes não paramétricos estatísticos de *Theil-Sen*, *Kruskal-Wallis* e *Kendall's Tau*, os autores observaram que a ocorrência de florações de alga adiantou entre 1,2 e 2,5 dias e que eventos climáticos de curto prazo, como chuvas intensas, temperaturas altas e ventos fortes, provocaram distúrbio na produção primária de fitoplâncton.

Xu *et al.* (2018) utilizou teste de Mann-Kendall para a avaliação da tendência de SDD no *Lake Liangzi*. Os resultados indicaram uma tendência significativa de diminuição da SDD durante o período estudado. O p-valor do teste foi de 0,002, o que reforça a significância dos resultados. Para a análise de cluster, Coffey *et al.* (2020) modelou floração de algas para todo os Estados Unidos continental por meio de dados do Sentinel-3 (OLCI) e utilizou método *Getis-Ord G** (ORD; GETIS, 1995) para agrupar regiões que apresentam características homogêneas em uma escala anual. Os autores observaram que existem grupos (*clusters*) de sistemas aquáticos continentais que atravessam estados, bacias hidrográficas e regiões ecológicas. O mesmo método de agrupamento foi aplicado para uma baía no Japão com dados de Chla modelados pelo Landsat-8 e foi verificado que regiões críticas na Baía identificada pela análise de *cluster* combinavam com as regiões de floração de algas (YUNUS; DOU; SRAVANTHI, 2015).

Apesar dos seguintes estudos não terem utilizado dados de sensoriamento remoto, esses utilizaram de análises estatísticas multivariadas para parâmetros de qualidade de água e merecem destaque. Nguyen *et al.* (2020) por meio de PCA mostraram que 14 de 21 parâmetros de qualidade da água explicam 75,1% da variabilidade da

qualidade da água para uma rede amostral localizada no *Mulde River* na Alemanha, o que retrata a aplicabilidade do método para a redução de dimensionalidade em casos de análise da qualidade da água. Na *Hilo Bay* localizada no Hawaii – EUA, Badlowski *et al.* (2021) utilizaram análise de agrupamentos por meio do método estatístico *Getis-Ord Gi** definido por Ord and Getis (1995) para a identificação de regiões com persistência de valores máximos e valores mínimos dos parâmetros de salinidade, turbidez e Chla. Na Turquia, Varol *et al.* (2012) buscaram avaliar a formação de grupos na dimensão temporal, e grupos na dimensão espacial para a avaliação dos meses que se agrupavam em relação ao período chuvoso e seco e estações amostrais que se agrupavam em relação às concentrações de parâmetros de qualidade de água monitorados.

3.6 Uso do sensoriamento remoto para avaliação da dinâmica da qualidade da água para reservatórios urbanos tropicais

A avaliação da qualidade da água em reservatórios urbanos é um tema que vem avançando nos últimos anos. Oliver, Corburn e Ribeiro (2019) por meio de revisão sistêmica identificaram que estudos entre 1992 e 2018 que utilizaram os temas reservatório urbano e cianobactérias totalizaram 357 artigos, e o Brasil lidera com 72 publicações.

Reservatórios com índice de estado trófico elevado apresentam presença de diferentes tipos de OACs que favorecem o monitoramento por sensoriamento remoto. Entretanto, na literatura nota-se uma baixa quantidade de estudos que utilizam o sensoriamento remoto para a avaliação da qualidade da água de reservatórios urbanos quando comparado com o avanço da utilização de sensoriamento remoto para águas continentais.

Na China, constata-se estudos que utilizaram do sensoriamento remoto aplicado a sistemas aquáticos inseridos em contexto urbano. Em um estudo, os autores utilizaram dados do Sentinel-2 para modelar a turbidez antes e durante o período de *lockdown* na pandemia do covid-19 (WANG *et al.*, 2022). Em outro estudo, Cai *et al.* (2021) classificaram diferentes tipos de água em relação a assinatura espectral e posteriormente combinaram com 12 parâmetros de qualidade de água para a classificação dos grupos de qualidade de água. Zhu *et al.* (2020) utilizaram dados do

Landsat-8 para avaliar CDOM e TSM no *West Lake*, inserido em área antrópica da cidade de Hangzhou na China. Os autores constataram que as principais fontes de poluição estão associadas às atividades de turismo e comerciais que intensificam as concentrações dos parâmetros na região norte do corpo hídrico.

No Brasil, existem estudos realizados para o Lago Paranoá, inserido no contexto urbano de Brasília em que os autores buscaram estimar Chla por meio de dados do Sentinel-2 (BORGES et al., 2020) e TSM por meio de dados de VANTs (OLIVETTI et al., 2020). No semiárido brasileiro, Souza Junior et al. (2021) relacionaram o índice de estado trófico com dados de reflectância do Landsat-8 para o Açude Grande, localizado na área urbana do município de Cajazeiras na Paraíba. Os reservatórios *Billings* e Guarapiranga, inseridos no contexto urbano de São Paulo, sofrem com pressões antrópicas e servem como estudo de caso para avaliação da qualidade da água por sensoriamento remoto, conforme estudado por Alcântara et al. (2021), Lobo et al. (2021), e Maciel et al. (2021). A Lagoa da Pampulha, em Belo Horizonte, apresenta estado hipereutrófico devido a sua inserção no contexto urbano da Região Metropolitana de Belo Horizonte (RMBH) e teve sua qualidade da água modelada por meio de sensoriamento remoto realizada por Souza (2019), que avaliou o parâmetro Turbidez inferido por modelos construídos com dados simulados do Landsat-8 e o parâmetro Chla modelado com dados de radiômetro.

Dentre os estudos citados, observa-se as dificuldades de utilização do sensoriamento remoto para a modelagem da qualidade da água de reservatórios urbanos, como: (i) a dificuldade de adaptação de modelos de águas continentais haja vista a inferior amplitude dos valores de concentração dos parâmetros; (ii) a alta frequência da dinâmica da qualidade da água devido às pressões antrópicas e processos hidrodinâmicos (mistura e estratificação); e (iii) fatores intervenientes que influenciam no registro da radiância, como efeito de adjacência, presença de resíduos no espelho d'água, poluição atmosférica, entre outros.

4 MATERIAIS E MÉTODOS

4.1 Área de estudo

O reservatório urbano a ser estudado é a Lagoa da Pampulha construída a partir do barramento do ribeirão Pampulha, em 1938, no município de Belo Horizonte, Minas Gerais (Figura 4a). Inicialmente, a finalidade principal do reservatório era o abastecimento de água, entretanto, a partir da década de 1970 a qualidade da água passou a ser severamente degradada como consequência da rápida urbanização com poucos instrumentos de controle ambiental na área de contribuição (98km²) do reservatório (CBH RIO DAS VELHAS, 2020). Dentre os principais afluentes, observa-se a contribuição majoritária (70% da vazão afluente) dos Córregos Ressaca e Sarandi, cujas áreas de contribuição localizam-se nos municípios de Contagem e Belo Horizonte e coincidem com regiões que a companhia de saneamento enfrenta dificuldades de universalização (TORRES; RESCK; COELHO, 2007) (Figura 4b).

Na tentativa de minimizar a poluição causada por ambos os córregos, em 2003 foi implementada uma estação de tratamento de águas fluviais que permite o tratamento de 750 L/s, correspondente à vazão média no período seco (COUTINHO; VON SPERLING, 2007). Outras medidas de proteção foram implementadas no âmbito da bacia hidrográfica, como o Programa de Recuperação e Desenvolvimento Ambiental da Bacia da Pampulha (PROPAM), iniciado em 1997. Em 2016, o complexo da Lagoa da Pampulha, que engloba o reservatório e os monumentos arquitetônicos no entorno, recebeu o título de Patrimônio Cultural da Humanidade, concedido pela UNESCO. Esse título intensificou a pressão sobre as organizações públicas em recuperar a qualidade da água do reservatório, levando à contratação de empresa especializada para a aplicação de remediador trófico para a captura do fósforo (*Phoslock*®) e biorremediador para redução da carga orgânica (*Enzilimp*®) (FURTADO; MONTE-MOR; DO COUTO, 2019).

Com 183 ha de espelho d'água, aproximadamente 1 milhão de m³ e profundidade máxima de aproximadamente 14 m (FURTADO; MONTE-MOR; COUTO, 2021), a Lagoa da Pampulha está localizada em região de clima tropical de altitude com distinção entre um período chuvoso com temperaturas quentes (primavera e verão) e outro período seco com temperaturas médias (outono e inverno) (NUNES; PINTO;

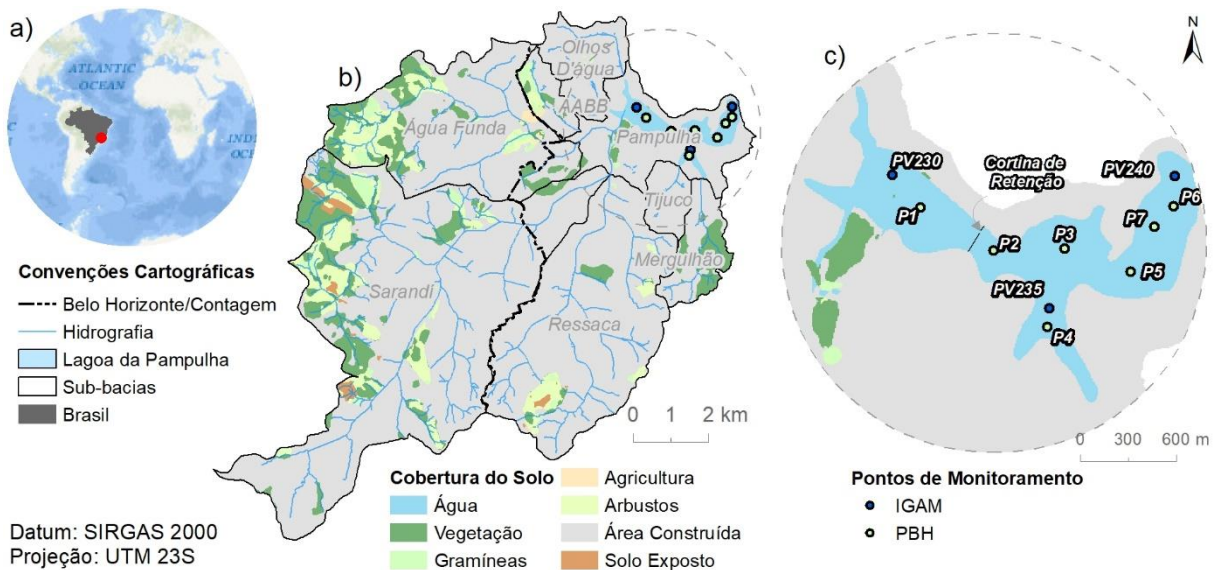
BAPTISTA, 2018). A temperatura do ar média varia entre 23°C na época do verão e 18°C na época do inverno (SILVA et al., 2019). Registros pluviométricos mostram que a precipitação anual média é aproximadamente 1500 mm (BEATO; DUTRA; MEDEIROS, 2001; NUNES; PINTO; BAPTISTA, 2018). As características morfométricas da Lagoa da Pampulha estão resumidas na Tabela 1 a seguir.

Tabela 1 – Características morfológicas da Lagoa da Pampulha

Características morfológicas	
Altitude	801 m
Profundidade média	5,1 m
Profundidade máxima	14 m
Profundidade mínima	1 m
Área Superficial	183 ha
Volume (nível normal)	1.000.000 m ³

Fonte: Adaptado de (FURTADO; MONTE-MOR; COUTO, 2021)

Figura 4 – Apresentação da área de estudo



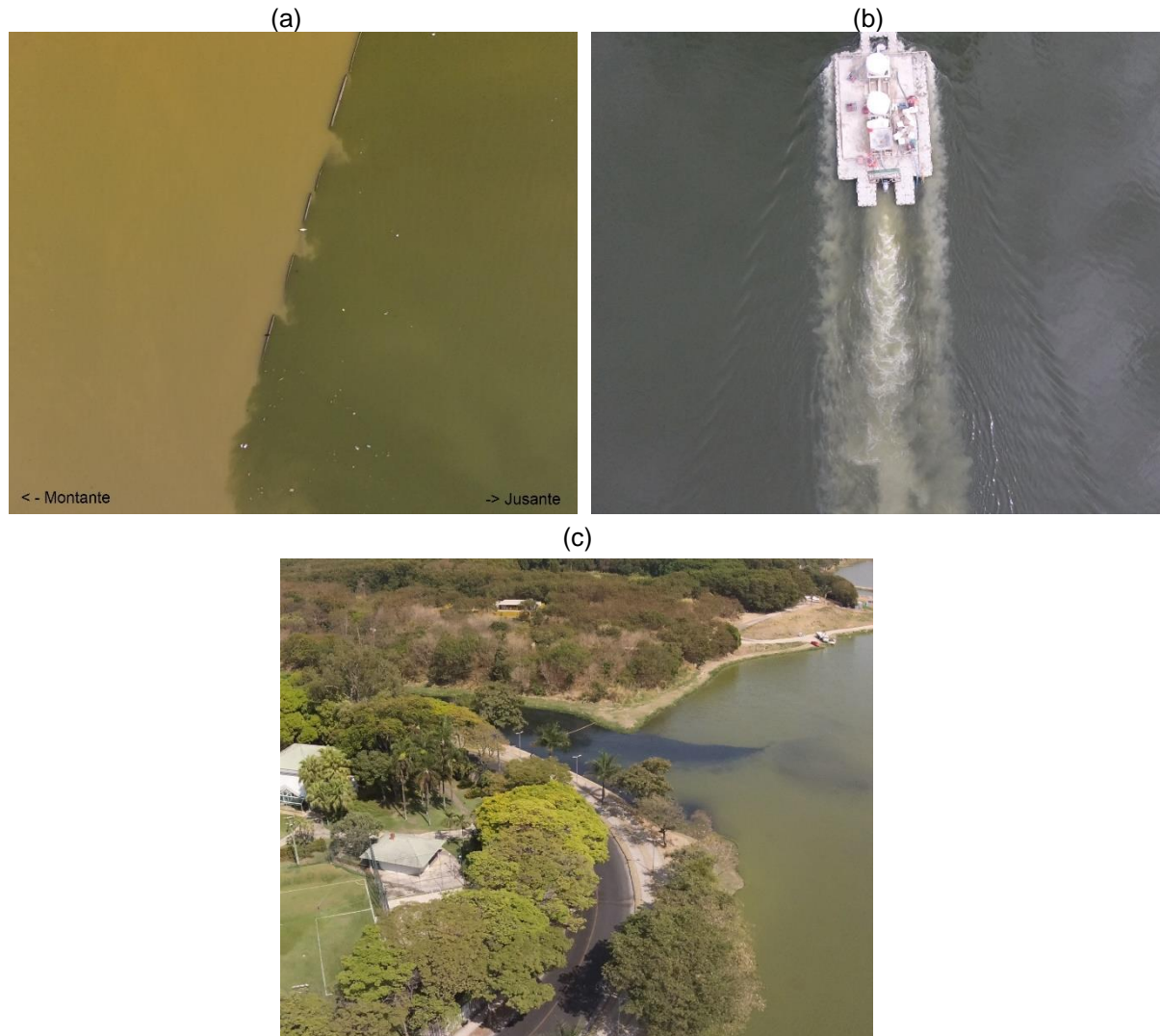
Legenda: a) Localização da área de estudo no globo terrestre; b) Bacia Hidrográfica da Lagoa da Pampulha e sua cobertura do solo; c) Lagoa da Pampulha e seus pontos de monitoramento de qualidade da água

Fonte: (IGAM, 2013; MICROSOFT; ESRI, 2022; Próprio Autor, 2022).

Nos registros fotográficos apresentados na Figura 5 nota-se a divisão da coloração da água para a região da cortina de retenção (Figura 5a), localizada conforme apresentada na Figura 4c, que vai da superfície até o fundo do reservatório (aproximadamente 2 m de profundidade para o local), a aplicação de produtos químicos por meio de uma balsa ao longo do reservatório (Figura 5b), e o encontro do

efluente da ETAF com a Lagoa da Pampulha com altas cargas de matéria orgânica (coloração escura) (Figura 5c).

Figura 5 – Registros da situação da Lagoa da Pampulha



Legenda: a) Cortina de retenção; (b) Aplicação de produtos químicos; (c) Efluente da ETAF chegando na Lagoa da Pampulha. Nota: Fotos cedidas pelo SIMOA-<https://simoa.eng.ufmg.br> Fonte: SIMOA (2023)

4.2 Dados limnológicos e climatológicos

O Instituto Mineiro de Gestão das Águas (IGAM) e a Prefeitura de Belo Horizonte (PBH) realizam monitoramento periódico da qualidade da água do reservatório por meio de pontos de monitoramento localizados no reservatório (Figura 4c). De acordo com os planos de monitoramento e regulamentações, as bases apresentam frequência de monitoramento de 3 meses e 1 mês, respectivamente.

O IGAM possui monitoramento em três pontos com coleta de água na subsuperfície: PV230 – Ilha dos Amores, PV235 – Igreja São Francisco e PV240 – Vertedouro (Figura 4c) (IGAM, 2013). Os dados monitorados pelo IGAM foram obtidos por meio do Portal InfoHidro (<http://portalinfohidro.igam.mg.gov.br/>), entretanto, desde o início do monitoramento em 2000, apenas dados posteriores a 2013 estão disponíveis no domínio. A PBH possui monitoramento em 7 pontos (P1 – P7) com coleta de água subsuperficial e registros desde 2016 (Figura 4c). Os dados foram obtidos por solicitação direta à PBH.

O IGAM monitora 192 parâmetros de qualidade de água enquanto a PBH monitora 20 parâmetros. Entretanto, para o presente estudo foram utilizadas informações gerais registradas nos bancos de dados de monitoramento e os dados de qualidade de água de TSM, Chla, SDD e Turbidez – parâmetros que apresentam interação com a radiação; coletados na sub-superfície da Lagoa da Pampulha. A Figura 6 demonstra exemplo do banco de dados de qualidade de água construído para a associação com imagens de satélite.

Figura 6 – Captura de tela do banco de dados de qualidade de água concatenado das bases do IGAM e PBH

	Ponto	Data Amostragem	Hora Amostragem	Fonte	Latitude	Longitude	TSM (mg/L)	Chla (µg/L)	SDD (m)	Turbidez (NTU)
0	PV230	2012-12-14	09:30:00	IGAM	-19.845856	-43.991425	40.0	23.059091	NaN	23.0
1	PV230	2013-03-27	09:30:00	IGAM	-19.845856	-43.991425	12.0	40.840000	NaN	21.9
2	PV230	2013-06-27	10:10:00	IGAM	-19.845856	-43.991425	30.0	26.619699	NaN	29.2
3	PV230	2013-07-31	10:55:00	IGAM	-19.845856	-43.991425	74.0	18.986667	NaN	37.3
4	PV230	2013-08-28	09:40:00	IGAM	-19.845856	-43.991425	38.0	3.204000	NaN	36.5
...
560	P3s	2021-10-01	NaN	PBH	-19.851366	-43.977531	32.0	92.500000	20.0	NaN
561	P4s	2021-10-01	NaN	PBH	-19.857328	-43.978897	NaN	119.000000	NaN	NaN
562	P5s	2021-10-01	NaN	PBH	-19.853088	-43.972201	NaN	53.400000	NaN	NaN
563	P6s	2021-10-01	NaN	PBH	-19.848126	-43.968757	NaN	82.400000	NaN	NaN
564	P7s	2021-10-01	NaN	PBH	-19.849696	-43.970317	NaN	80.800000	NaN	NaN

Fonte: Próprio autor (2022)

Os dados de precipitação são provenientes da estação 83.587 do Instituto Nacional de Meteorologia (INMET) localizada nas coordenadas 19,9344 S e 43,9522 O. Os dados foram baixados com agregação mensal e possuem registros de 1961 a 2022. Para baixar as informações, utilizou-se o Banco de Dados Meteorológicos (BDMEP - <https://bdmep.inmet.gov.br/#>). Meses que apresentaram valores nulos foram

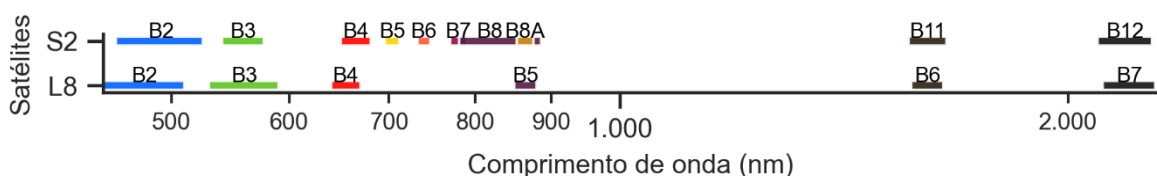
preenchidos utilizando técnica do cálculo de medianas para aquele mês em toda série temporal.

4.3 Dados de sensoriamento remoto

Todas as operações com as imagens de satélite foram realizadas no *Python* 3.10 por meio de interface com a biblioteca do GEE *Python API* (GORELICK et al., 2017). Para a inferência dos parâmetros de qualidade de água de interesse (TSM, Chla, SDD, Turbidez), utilizou-se de AOPs derivadas dos sensores a bordo do Landsat-8 e Sentinel-2 para os períodos de 2013-2022 e 2015-2022, respectivamente. O período escolhido vai de encontro com a disponibilidade de informações de qualidade de água e dos satélites com o ano de realização deste estudo.

O Landsat-8, lançado em 2013, apresenta resolução espacial de 30 m para 7 bandas espectrais do sensor OLI. A resolução radiométrica é de 12 bits com reamostragem para 16 bits e a taxa de revisita de 16 dias (USGS, 2019). O Sentinel-2, por meio da constelação Sentinel-2A lançado em 2015 e Sentinel-2B lançado em 2017, possui o sensor MSI com resolução de 10 m para bandas do VNIR, 20 m para bandas da borda do vermelho e SWIR. No total, têm-se 12 bandas. A resolução radiométrica e temporal é de 12 bits e 5 dias, respectivamente (ESA, 2016). Para o monitoramento da Lagoa da Pampulha foram utilizadas apenas bandas entre 400 e 2300 nm no espectro eletromagnético. Os comprimentos de onda e faixas adotadas para as bandas utilizadas no presente estudo estão apresentadas na Figura 7.

Figura 7 – Largura das bandas a serem utilizadas na obtenção de grandezas radiométricas para os sensores embarcados no Landsat-8 (L8) e Sentinel-2 (S2).

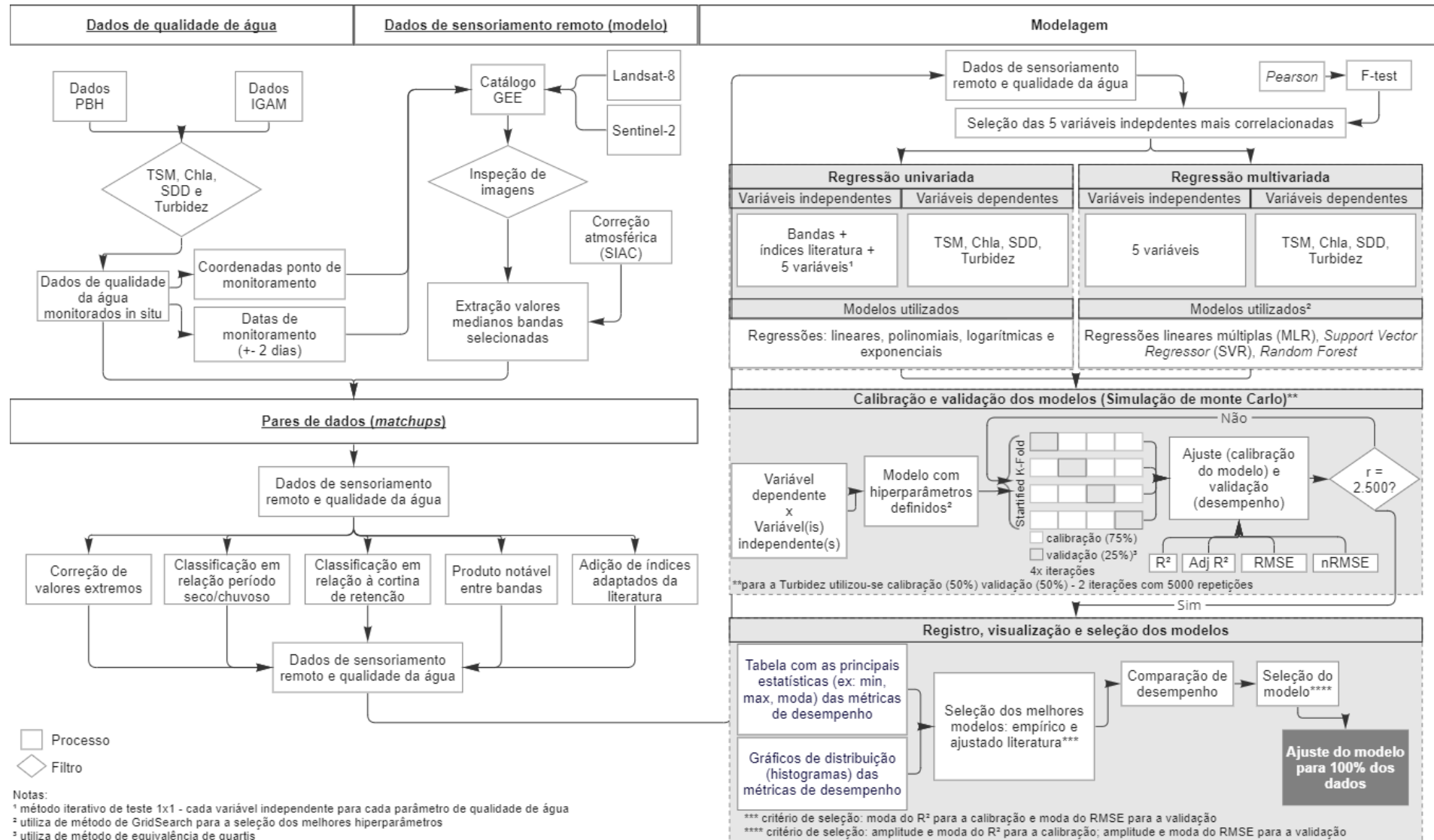


Nota: B2 – azul, B3 – verde, B4 – vermelho, B5 – borda do vermelho 1 para S2 e NIR para o L8, B6 – borda do vermelho 2 para S2 e SWIR 1 para L8, B7 – borda do vermelho 3 para o S2 e SWIR 2 para o L8, B8 – NIR para o S2, B8A – borda do vermelho 4 para o S2, B11 e B12 – SWIR 1 e SWIR 2 para o S2, respectivamente. Fonte: Adaptado de (ESA, 2016; USGS, 2019)

4.4 Processamento dos dados de qualidade de água e sensoriamento remoto

As etapas de processamento dos dados de qualidade de água, de processamento de dados de sensoriamento remoto, de união dos pares de dados (qualidade da água *in situ* com dados de reflectância de sensoriamento remoto), de união com dados de sensoriamento remoto, e de modelagem estão apresentados na Figura 8.

Figura 8 – Diagrama esquemático das etapas envolvidas para a modelagem da qualidade da água por dados de sensoriamento remoto



Fonte: Próprio Autor (2023). Nota: A construção de pares de dados e modelagem é por satélite-sensor, portanto se repete.

Os dados de qualidade de água provenientes do monitoramento do IGAM e PBH foram concatenados e filtrados para os parâmetros modelados no presente estudo conforme ilustrado na Figura 6 e Figura 8. Posteriormente, realizou-se análises descritiva e exploratória desses visando a compreensão da frequência de monitoramento realizado, a verificação da presença de *outliers*, e a correlação entre os parâmetros de qualidade de água. Para a realização das etapas descritas foram utilizadas bibliotecas do *Python*, como *Pandas* (REBACK et al., 2022), *Scipy* (VIRTANEN et al., 2020), *Numpy* (HARRIS et al., 2020), *Seaborn* (WASKOM, 2021), e *Matplotlib* (HUNTER, 2007).

Ao combinar as datas concomitantes de monitoramento realizados pelo IGAM e PBH com a passagem dos satélites, observou-se baixo número de pares ($n_{L8}=9$ e $n_{S2}=13$) e por isso buscou-se utilizar imagens com diferença de até 2 dias da data do monitoramento. Todas as imagens dos satélites obtidas foram visualizadas de forma manual para a seleção daquelas sem a interferência de nuvens ou presença de reflexão especular. A correção atmosférica foi realizada por meio do método *SIAC* utilizando o código disponibilizado em https://github.com/MarcYin/SIAC_GEE (YIN; LEWIS; GÓMEZ-DANS, 2022).

Os valores de reflectância foram extraídos considerando-se uma janela de 3×3 *pixels* no entorno do pixel correspondente ao ponto de monitoramento. Um valor mediano de reflectância foi calculado considerando-se os *pixels* incluídos na janela, buscando-se minimizar de erros de localização dos pontos de monitoramento e possíveis derivações das embarcações durante as campanhas.

Concatenou-se ao banco de dados os produtos notáveis de multiplicação (Equação 4), divisão (Equação 5), e índice (Equação 6), entre duas bandas.

$$Rrs(\lambda_i) \times Rrs(\lambda_j) \quad (4)$$

$$Rrs(\lambda_i) \div Rrs(\lambda_j) \quad (5)$$

$$\frac{Rrs(\lambda_i) - Rrs(\lambda_j)}{Rrs(\lambda_i) + Rrs(\lambda_j)} \quad (6)$$

Onde i e j se referem as bandas dos satélites e $i \neq j$.

Os principais índices ou razões entre bandas construídos por algoritmos consolidados na literatura foram adaptados para a utilização das bandas do Landsat-8 e Sentinel-2 (Tabela 2) e enriquecimento do banco de dados.

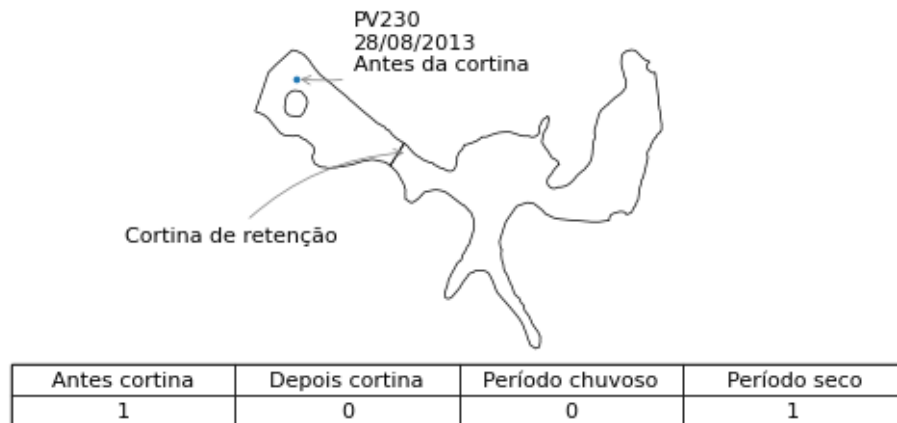
Tabela 2 – Algoritmos da literatura para a estimativa de TSM e Chla adaptados para os satélites utilizados

Nome	Satélite	Equação	Referência (Adaptado de)
Nechad		$\frac{R_{rs}(865)}{1 - R_{rs}(865)/21,2}$	(NECHAD; RUDDICK; PARK, 2010)
2bandas	Landsat-8	$R_{rs}^{-1}(655) \times R_{rs}(865)$	(DALL'OLMO; GITELSON, 2005)
3bandas		$[R_{rs}^{-1}(560) - R_{rs}^{-1}(655)] \times R_{rs}(865)$	
NDCI		$\frac{R_{rs}(865) - R_{rs}(655)}{R_{rs}(865) + R_{rs}(655)}$	(MISHRA; MISHRA, 2012)
Nechad		$\frac{R_{rs}(704)}{1 - R_{rs}(704)/18,9}$	(NECHAD; RUDDICK; PARK, 2010)
2bandas	Sentinel-2	$R_{rs}^{-1}(665) \times R_{rs}(704)$	(DALL'OLMO; GITELSON, 2005)
3bandas		$[R_{rs}^{-1}(665) - R_{rs}^{-1}(704)] \times R_{rs}(740)$	
NDCI		$\frac{R_{rs}(704) - R_{rs}(665)}{R_{rs}(704) + R_{rs}(665)}$	(MISHRA; MISHRA, 2012)

Fonte: Próprio Autor (2022)

Além disso, inseriu-se no banco de dados informações referentes a localização do ponto de monitoramento em relação à cortina de retenção de sedimentos e a correspondência entre a data do monitoramento e a estação do ano, se primavera e verão (período chuvoso) ou outono e inverno (período seco). Estas variáveis categóricas foram transformadas em variáveis numéricas com a representação de ocorrência da classe o valor “um” e a não ocorrência o valor “zero”, conforme exemplificado na Figura 9.

Figura 9 – Exemplo de transformação das variáveis categóricas para variáveis numéricas (criação de *dummy variables*)



Fonte: Próprio Autor (2023)

Dessa forma, o banco de dados ficou enriquecido com as variáveis das bandas dos satélites, variáveis de relações entre bandas conforme equações listadas nas Equações 4, 5, e 6, variáveis dos algoritmos adaptados da literatura (Tabela 2) e variáveis categóricas transformadas para variáveis discretas que representam processos físicos no reservatório. As listas de todas as colunas do banco de dados final por satélite podem ser verificadas no APÊNDICES A e B. Esse banco de dados (Figura 8 – Pares de dados (*matchups*)) foi fundamental para a etapa da modelagem e inferência dos parâmetros de qualidade de água.

4.5 Inferência de parâmetros de qualidade de água a partir de dados do sensoriamento remoto

A inferência dos parâmetros de qualidade de água foi tratada como um problema de regressão. Os parâmetros de qualidade da água obtidos *in situ* TSM, Chla, SDD e Turbidez compõem o conjunto de variáveis dependentes a serem preditas a partir das variáveis independentes as quais incluem os valores de reflectância correspondentes às bandas dos satélites selecionadas, as relações entre bandas e as variáveis de localização e período. Foram utilizados modelos de regressão univariadas, incluindo regressão linear (Equação 7), polinomial (Equação 8), exponencial (Equação 9) e logarítmica (Equação 10), e modelos de regressão multivariadas, incluindo MLR (Equação 11), SVR (Equação 12) e *Random Forest* (Equação 13).

$$y = \beta_0 + \beta_1 x \quad (7)$$

$$y = \beta_0 + \beta_2x + \beta_3x^2 \quad (8)$$

$$y = \beta_0 + e^{\beta_1x} \quad (9)$$

$$y = \beta_0 + \beta_1 \ln(x) \quad (10)$$

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_nx_n \quad (11)$$

$$\min(\omega, b, \xi, \xi^*) \frac{1}{2} \omega^T \omega + C \sum_{i=1}^1 \xi_i + C \sum_{i=1}^1 \xi_i^* \quad (12)$$

$$\hat{Y}(x_i) = \frac{1}{B} \sum_{b=1}^B T_b(x_i) \quad (13)$$

Onde y é a variável dependente, β_n equivale aos coeficientes e interseção da reta, x representa a variável independente para os modelos de regressão linear, polinomial, exponencial, logarítmica e linear múltipla. Para o SVR, ω representa um vetor de pesos que determina o hiperplano, b representa o viés, ξ e ξ^* são variáveis de brecha associadas as amostras de calibração nos dois lados do hiperplano, ω^T é o termo de regularização e C é o parâmetro de custo que possibilita a obtenção de um erro de calibração menor. Para o *Random Forest*, B equivale ao número de árvores de decisão, T_b é o nó final da árvore e \hat{Y} representa o resultado médio da inferência para as árvores presentes no método *Random Forest*.

Para encontrar as variáveis independentes que melhor explicavam as variáveis dependentes, realizou-se o cálculo do coeficiente de correlação linear de *Pearson* e, posteriormente, aplicou-se teste de hipótese *F-test* para a verificação da significância do resultado. As cinco variáveis independentes com maiores correlações e que demonstraram ser significantes (p-valor menor que 5%) foram selecionados para serem utilizados nas regressões univariadas e multivariadas. Para facilitar a interpretação categorizou-se a correlação em forte ($r \geq 0,7$ e $r \leq -0,7$), intermediária ($0,2 < r < 0,7$ e $-0,7 < r < -0,2$) e fraca ($-0,2 \leq r \leq 0,2$).

Para as regressões univariadas, haja vista a agilidade de processamento e obtenção dos resultados, além das cinco variáveis independentes que mais se correlacionavam com as variáveis dependentes, optou-se por testar todas as bandas dos satélites

consideradas no presente estudo, e as relações entre bandas adaptadas da literatura (Tabela 2). Para as regressões multivariadas, por dependerem maior tempo e recursos computacionais, foram utilizados apenas as cinco variáveis. A lógica de separação e utilização das variáveis explicadas pode ser visualizada nas seções de regressão univariada e multivariada apresentadas na Figura 8.

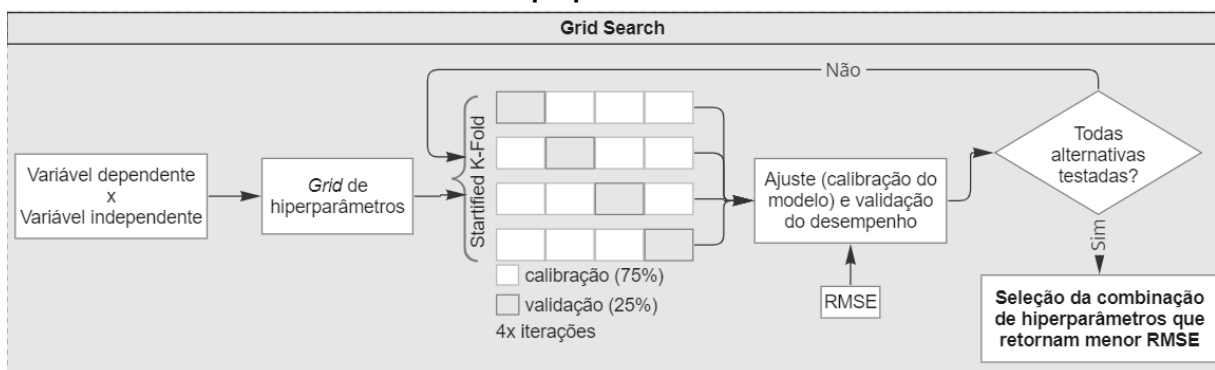
Para a determinação dos melhores hiperparâmetros dos modelos de aprendizado de máquina, utilizou-se de método de *GridSearch* combinado com validação cruzada (PEDREGOSA et al., 2011). A determinação das combinações dos melhores hiperparâmetros se deu por verificação do RMSE durante o processo de iteração dos modelos. Os hiperparâmetros testados por modelo estão apresentados na Tabela 3 e a lógica desenvolvida é explicada por meio da Figura 10.

Tabela 3 – Hiperparâmetros testados no *GridSearch* para os modelos de aprendizado de máquina

Modelo	Hiperparâmetro	Valores testados
SVR	Kernel	Linear, Polynomial, RBF, Sigmoid
	C	500, 1000, 10000
	ϵ	1, 2, 5, 10
	Γ	Scale, Auto
Random Forest	Nº de estimadores	10, 50, 100, 500, 1000
	Critério	Mínimos quadrados
	Profundidade Máxima	2, 10, 25, 50

Fonte: Próprio Autor (2022)

Figura 10 – *GridSearch* combinado com validação cruzada para a determinação dos melhores hiperparâmetros

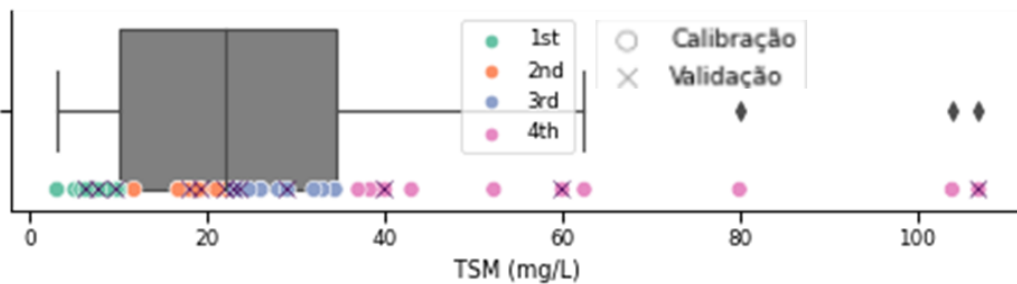


Fonte: Próprio Autor (2023)

Utilizou-se de método de validação cruzada para a separação do conjunto amostral de dados para as etapas de treinamento (calibração) e teste (validação). O método de validação cruzada *Stratified Repeated K-Fold*, disponível na biblioteca *Sklearn* (PEDREGOSA et al., 2011), foi implementado com uma quantidade de 4 grupos,

sendo 3 grupos (75% dos dados) para calibração e 1 grupo (25% dos dados) para a validação com 2.500 repetições. Para o parâmetro turbidez, devido à baixa quantidade de dados disponíveis, utilizou-se validação cruzada *Stratified Repeated K-Fold* com 50% dos dados para calibração e 50% para validação com 5.000 repetições (Figura 8 – calibração e validação dos modelos). Os métodos de seleção das amostras a serem utilizados na validação respeitou equivalência por quartis que garante que na validação exista a mesma quantidade de dados de cada quartil (Figura 11). Isso permite maior equilíbrio na obtenção das métricas de desempenho do modelo.

Figura 11 – Exemplo da separação do conjunto amostral em treinamento e teste para uma repetição



Fonte: Próprio Autor (2022)

Tendo em vista o reduzido tamanho amostral (n variou entre 9 e 54) e sua considerável extensão temporal (9 anos), a validação cruzada (2.500 repetições x 4 iterações – TSM, Chla, SDD, e 5.000 repetições x 2 iterações - Turbidez) permitiu para o presente trabalho a adoção de Simulação de Monte Carlo (10.000 repetições) para cada modelo para aumentar a robustez dos modelos e permitir a avaliação da distribuição das métricas de desempenho (Figura 8). As métricas utilizadas, apresentadas na Tabela 4 foram o coeficiente de determinação (R^2), o coeficiente de determinação ajustado (R^2 ajustado), a raiz do erro médio quadrático (RMSE) e o RMSE normalizado (nRMSE) pela média dos dados observados.

A seleção dos melhores modelos (Figura 8 – Registro, visualização e seleção dos modelos) foi dividida em modelo empírico, que se refere modelo que não necessariamente utiliza bandas ou relações de bandas que apresentam sentido físico na estimativa de OACs, e modelo ajustado da literatura, que se refere a modelo que utiliza de conceitos físicos para sua construção. A decisão para a escolha dos dois modelos baseou-se em maiores valores de moda ($n=10.000$) para o R^2 de calibração e menores valores de moda ($n=10.000$) para o RMSE de validação.

Tabela 4 – Métricas para verificação da acurácia dos modelos

Nome	Equação	Intervalo	Unidade
R²	$R^2 = \frac{\sum_{i=1}^n (Y_{pred} - \bar{y})^2}{\sum_{i=1}^n (Y_{obs} - \bar{y})^2}$	(0, 1)	Adimensional
R² ajustado	$AdjR^2 = \frac{(1 - R^2)(1 - n)}{(n - k - 1)}$	(0, 1)	Adimensional
RMSE	$RMSE = \left(\frac{\sum_{i=1}^n (Y_{pred} - Y_{obs})^2}{n} \right)^{\frac{1}{2}}$	(0, +∞)	Unidade dos parâmetros
nRMSE	$nRMSE = \frac{RMSE}{\bar{Y}}$	(0, +∞)	Percentual

Nota: Y_{pred} são os valores estimados, Y_{obs} são os valores observados e \bar{Y} é a média dos valores observados, k e n – número de variáveis independentes e número de dados utilizados na regressão, respectivamente.

Fonte: Próprio Autor (2022)

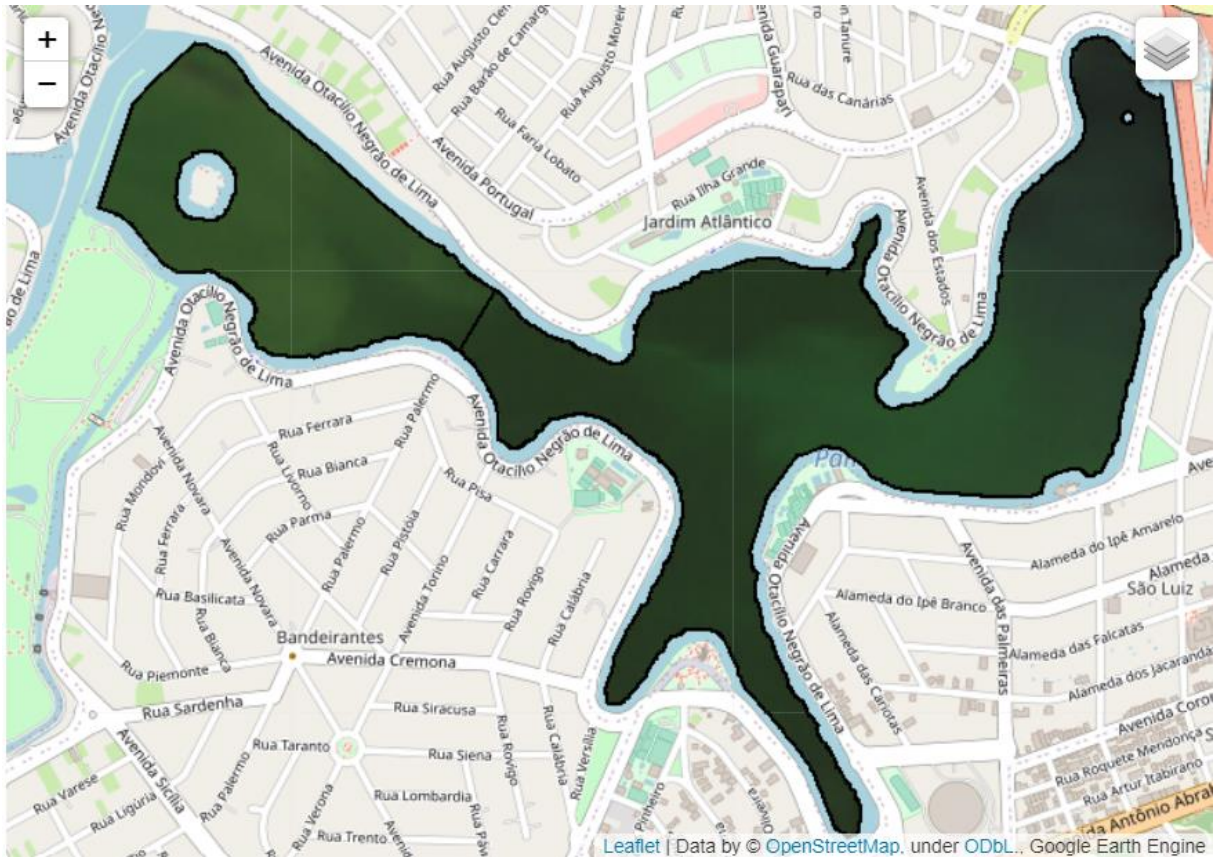
Com a escolha do modelo empírico e modelo ajustado da literatura para cada parâmetro de qualidade de água, comparou-se a distribuição de métricas obtidas para ambos os modelos e avaliou-se a capacidade de generalização de cada um. Selecionou-se apenas um modelo para cada parâmetro de qualidade de água por satélite. Os modelos selecionados, com suas configurações provenientes do processo de calibração e validação, foram ajustados novamente considerando todos os dados *in situ* disponíveis para o parâmetro modelado e satélite correspondente. Com isso, dois modelos (um por satélite) para cada parâmetro foram empregados para a inferência da qualidade da água em regiões do reservatório sem monitoramento *in situ* e para todas as datas com disponibilidade de imagens durante o período de estudo.

4.6 Aplicação dos modelos para todo o reservatório e para imagens históricas

A extrapolação dos resultados da modelagem foi executada por meio da aplicação do melhor modelo obtido por parâmetro para os demais pixels não monitorados e para datas complementares àquelas em que ocorreram as campanhas *in situ*. Restringiu-se a extrapolação para pixels que distam no mínimo 50 m da margem do reservatório, evitando assim interferências na reflectância provenientes de efeitos do fundo do reservatório e de efeitos de pixels nas margens, na interface entre água e terra (DÖRNHÖFER; OPPELT, 2016) Foram selecionadas de forma manual (visual) apenas imagens que não apresentaram nuvens, sombra de nuvens, e reflexão especular evitando a superestimação ou subestimação dos parâmetros devido a essas interferências.

A delimitação da máscara d'água foi construída por meio da adaptação de arquivo vetorial disponibilizado pela PBH com ajustes que incluíram: remoção da região de cabeceira que circunda o Parque Ecológico, remoção dos braços do reservatório na entrada dos córregos Água Funda e Braúnas, a oeste, remoção de estruturas que se sobrepõe ao espelho d'água do reservatório (Ilha dos Amores, a Casa do Baile, o vertedouro tulipa e a fonte de água próxima à barragem). Com o arquivo vetorial da máscara d'água atualizada moveu-se o limite 50 metros para o interior do reservatório para evitar os problemas descritos anteriormente (Figura 12).

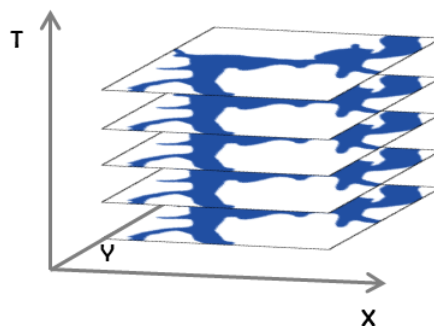
Figura 12 – Extensão da máscara de água respeitando os 50 metros da margem e estruturas presentes no reservatório.



Fonte: Próprio Autor (2022)

Todas as imagens disponíveis para ambos satélites utilizadas neste trabalho estão apresentadas nos APÊNDICES C e D. Criou-se uma lista com as imagens aptas para a aplicação dos modelos e por meio da biblioteca xarray (HOYER; HAMMAN, 2017) as imagens da área de estudo foram baixadas do servidor da Google® e foram criados arquivos locais *Network Common Data Form* (NetCDF) para a aplicação dos modelos e construção das matrizes multidimensionais (Figura 13).

Figura 13 – Exemplo da estrutura da matriz multidimensional com a representação das coordenadas (eixos x e y) e tempo (eixo T).



Fonte: Próprio Autor (2022)

Os dados modelados de qualidade de água do Sentinel-2 foram combinados com os dados modelados de qualidade de água do Landsat-8 para a maximização da frequência amostral dos parâmetros de qualidade de água. Para isso, os dados das bandas com resolução espacial de 10 m e 20 m do Sentinel-2 foram reamostrados para 30 m utilizando-se método de interpolação linear.

4.7 Avaliação da dinâmica espacial e temporal da qualidade da água na Lagoa da Pampulha

4.7.1 Estatística descritiva da dinâmica espaço-temporal da qualidade da água

Para os dados modelados combinados (Landsat-8 e Sentinel-2) neste trabalho, recorreu-se à métodos estatísticos descritivos para a visualização do comportamento espacial e temporal dos parâmetros de qualidade da água. Comparou-se os dados modelados com os dados observados por meio de gráficos de séries temporais para o período estudado (2013-2022). Calculou-se a mediana para cada parâmetro na dimensão do tempo para a visualização do comportamento espacial da qualidade da água no reservatório.

4.7.2 Análise multivariada para determinação de regiões homogêneas e heterogêneas no reservatório

No âmbito espacial, agruparam-se os parâmetros de qualidade de água modelados para todos os períodos em um banco de dados único, calculando-se a medida de similaridade e dissimilaridade por meio da distância euclidiana e agrupando-se subconjuntos amostrais por método de *Ward* (WARD JR., 1963). A seleção do número de grupos foi definida por meio da visualização do dendrograma. Os grupos construídos foram plotados em mapa para a identificação das regiões homogêneas (*pixels* agrupados) e heterogêneas (*pixels* de grupos diferentes). Para reduzir problemas de diferentes amplitudes dos parâmetros de qualidade de água, os dados foram normalizados antes da realização da análise de *cluster* (VAROL et al., 2012).

Escolheu-se distância para a divisão do número de grupos com base na separação da Lagoa da Pampulha em mais de dois grupos, para que se pudesse avaliar outras regiões além das regiões de montante (anterior à cortina de retenção) e de jusante (posterior à cortina de retenção) (SOUZA, 2019).

4.7.3 Testes de hipótese para a avaliação da influência de variáveis categóricas externas na qualidade da água

O teste univariado de *Mann-Whitney* (MANN, 1945) e multivariado de *Kruskal-Wallis* (KRUSKAL; WALLIS, 1952) foram aplicados para se avaliar a influência (i) do período chuvoso e seco; (ii) do tratamento da Lagoa da Pampulha com os produtos *Phoslock®* e *Enzilimp®*, considerando-se o período anterior ao tratamento, durante o tratamento e os períodos de interrupções temporárias e; (iii) de anos muito secos, secos, normais, úmidos ou muito úmidos assim definidos por meio do *Standard Precipitation Index* (SPI).

Ambos testes comparam amostras de dados independentes e seguem abordagem não paramétrica prevista para dados que não seguem a normalidade, como é comum no caso de variáveis ambientais (SPERLING; VERBYLA; OLIVEIRA, 2020). O teste de *Mann-Whitney* é utilizado para a comparação de até dois grupos amostrais, ao passo que o teste de *Kruskal-Wallis* se aplica à comparação de três ou mais grupos. Os testes foram aplicados considerando-se um nível de significância de 5%. Para a aplicação do teste, calculou-se o valor da mediana para cada região identificada na análise de agrupamento, para cada data e parâmetro de qualidade da água avaliado.

A distinção entre dados de período chuvoso e período seco foi realizada por meio da classificação em relação a data de monitoramento. Dados monitorados com data entre 01/outubro e 31/março foram classificados como período chuvoso e dados monitorados entre 01/abril e 30/setembro foram classificados como período seco.

A aplicação de produtos químicos para o tratamento do reservatório teve início em abril de 2016, conforme informado por Barçante *et al.* (2020). Os eventos de paralisação da aplicação foram informados pela PBH e correspondem com os meses de março/2018, abril/2018, junho/2018, novembro/2019, janeiro/2020, abril/2020, maio/2021, outubro/2021 e novembro/2021. Dados anteriores a abril de 2016 foram considerados anteriores à aplicação de produtos químicos.

O cálculo do índice SPI₁₂ foi realizado por meio dos dados de precipitação acumulado mensal da estação do INMET 83.587 para um período de 12 meses considerando o ano hidrológico tendo início no mês de outubro. A metodologia de cálculo e

classificação do índice seguiu a proposição de McKee, Doesken e Kleist (1993) e foi implementada por meio da biblioteca *standard-precip* (https://github.com/e-baumer/standard_precip) no *Python*.

4.7.4 Avaliação da tendência da qualidade da água do reservatório

Para a avaliação de tendência os dados tiveram que ser reamostrados para uma frequência menor de monitoramento, favorecendo a redução de efeitos de dependência temporal (HIRSCH; SLACK; SMITH, 1982; KENDALL, 1948) e reduzindo o número de lacunas existentes em meses que não foi possível obter imagens de satélite. Para isso, utilizou-se de métodos da biblioteca *xarray* (HOYER; HAMMAN, 2017) para a formação de períodos – optou-se por adotar distinção de períodos chuvosos e períodos secos; e cálculo da mediana na dimensão do tempo. Os valores das medianas obtidos para cada período e parâmetro foram plotados para fins de avaliação da evolução da qualidade da água na Lagoa da Pampulha para o período estudado (2013-2022).

Para ratificar a tendência dos parâmetros de qualidade de água entre os anos 2013 e 2022, utilizou-se de teste de hipótese não paramétrico denominado *Mann-Kendall Seasonal* disponível por meio da biblioteca *pymannkendall* (HUSSAIN; MAHMUD, 2019), que considera sazonalidade nos dados e permite a ocorrência de alteração da qualidade da água ao longo dos anos (HIRSCH; SLACK; SMITH, 1982; KENDALL, 1948; MANN, 1945). O teste consistiu em verificar, para cada pixel do reservatório, a condição de padrão de qualidade da água crescente, decrescente ou nulo. Por meio de ajuste linear e cálculo de medianas dos coeficientes da reta, o método de *Theil-Sen* foi combinado com o anterior para indicar o valor do coeficiente da reta, permitindo inferir se a tendência é crescente ou decrescente (SEN, 1968; THEIL, 1992).

Por meio dos grupos definidos na análise multivariada, extraiu-se a quantidade de *pixels*, os valores mínimos e, os valores máximos obtidos do coeficiente de reta para auxiliar na interpretação da tendência da qualidade da água no reservatório.

5 RESULTADOS

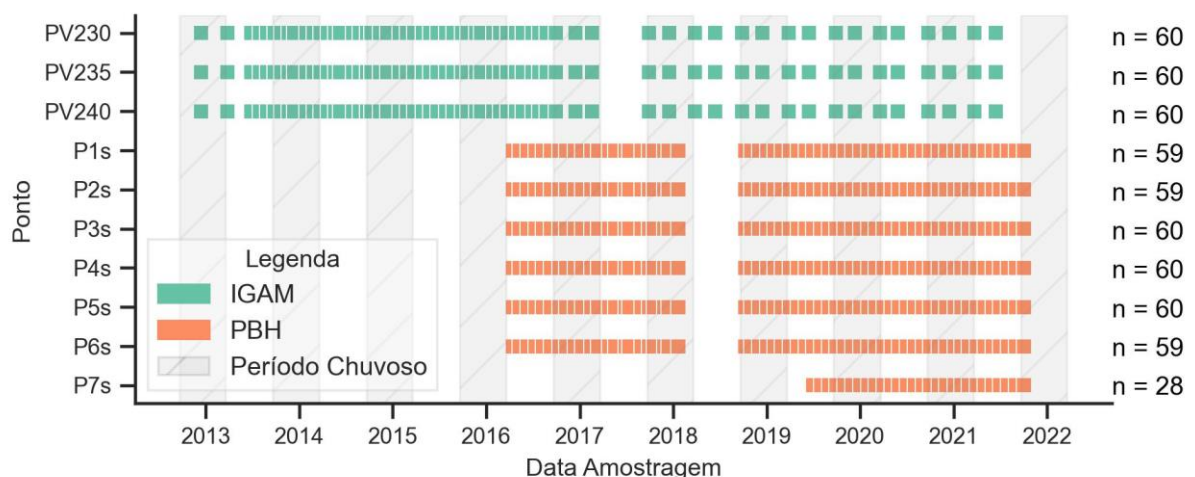
5.1 Dados limnológicos, climatológicos e de sensoriamento remoto

5.1.1 Dados limnológicos

A disponibilidade de dados de qualidade de água do IGAM para a Lagoa da Pampulha inicia-se no ano de 2013. Observa-se variação na frequência de monitoramento, com frequência trimestral até o ano de 2014 e, nos anos de 2014, 2015 e 2016, verifica-se um aumento da frequência para o monitoramento mensal (Figura 14). A partir de 2017 o monitoramento voltou a ter frequência trimestral. Para o presente estudo foram utilizados dados até maio de 2021. Para os três pontos monitorados (PV 230, PV 235 e PV 240) observa-se um total de 60 registros por ponto.

O monitoramento da PBH teve início em 2016, a partir da instituição do Consórcio Pampulha Viva. Observa-se a frequência de monitoramento mensal (Figura 14), entretanto, nota-se a interrupção do monitoramento pela maior parte do ano de 2018. Para os pontos monitorados (Figura 4c), observa-se semelhança na quantidade de dados de monitoramento ($n=60$), exceto para o P6, que possui 59 registros e o P7, que foi incorporado ao monitoramento a partir de 2019, apresentando 28 registros. Para o presente estudo, foram utilizados dados até outubro de 2021.

Figura 14 – Disponibilidade de dados de monitoramento *in situ* realizado pelo IGAM e PBH na Lagoa da Pampulha



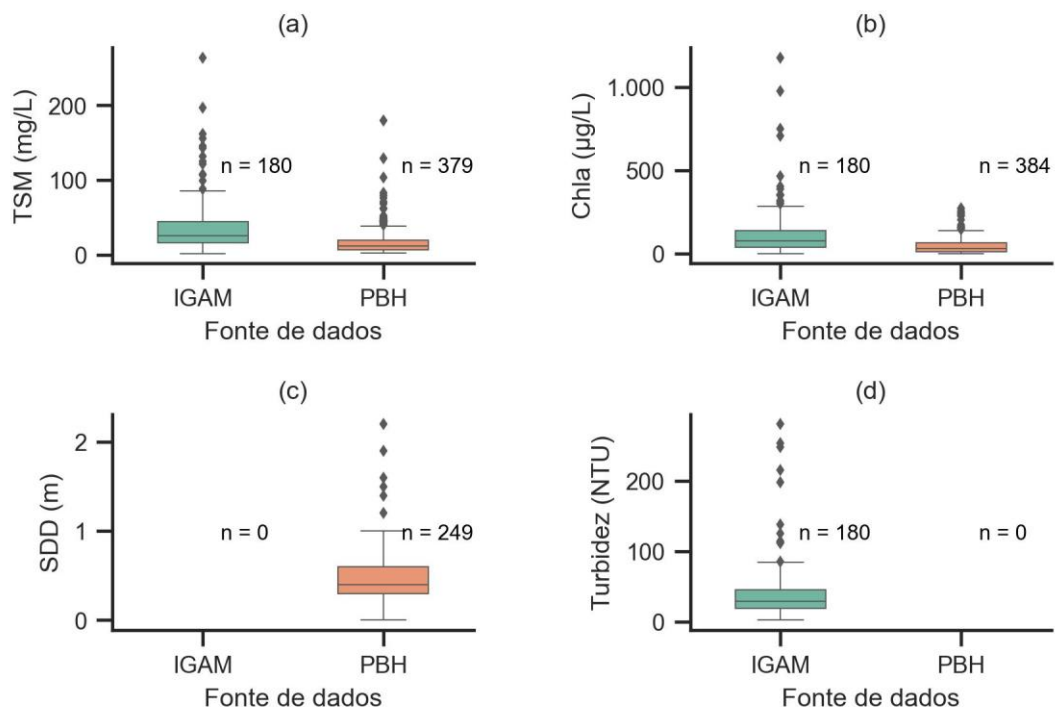
Fonte: Próprio Autor (2022)

Para as medições de SDD, constatam-se valores próximos ou maiores que a profundidade do reservatório (Tabela 1). Ao avaliar as características desses valores

extremos, observa-se que ocorreram na mesma data de monitoramento e para pontos localizados na parte de montante do reservatório (P1s e P2s), em região pouco profunda e caracterizada por baixos valores de transparência. Por isso, interpreta-se estes valores preenchidos com a unidade de medida errada, decidindo-se dividi-los por 100 para adequar à escala de medida. Na continuação do estudo, os valores de SDD serão apresentados com a mencionada correção.

A análise descritiva dos parâmetros de qualidade de água alvo do presente estudo mostrou maior disponibilidade de dados da PBH em relação ao IGAM. Para os parâmetros TSM e Chla, observa-se quase o dobro do total de registros da PBH em relação ao IGAM (Figura 15a e Figura 15b). Para os parâmetros SDD e Turbidez, observa-se a não realização do monitoramento do primeiro pelo IGAM (Figura 15c) e do segundo, pela PBH (Figura 15d). Os valores dos parâmetros de qualidade de água foram plotados em *boxplots* (Figura 15a-d) e suas principais estatísticas para a Lagoa da Pampulha foram calculadas e estão apresentadas na Tabela 5.

Figura 15 – *Boxplots* com a distribuição dos valores dos principais parâmetros de qualidade de água do presente estudo monitorados pelo IGAM e PBH para a Lagoa da Pampulha



Legenda: Distribuição dos dados de a) TSM; b) Chla; c) SDD; d) Turbidez.

Fonte: Próprio Autor (2022)

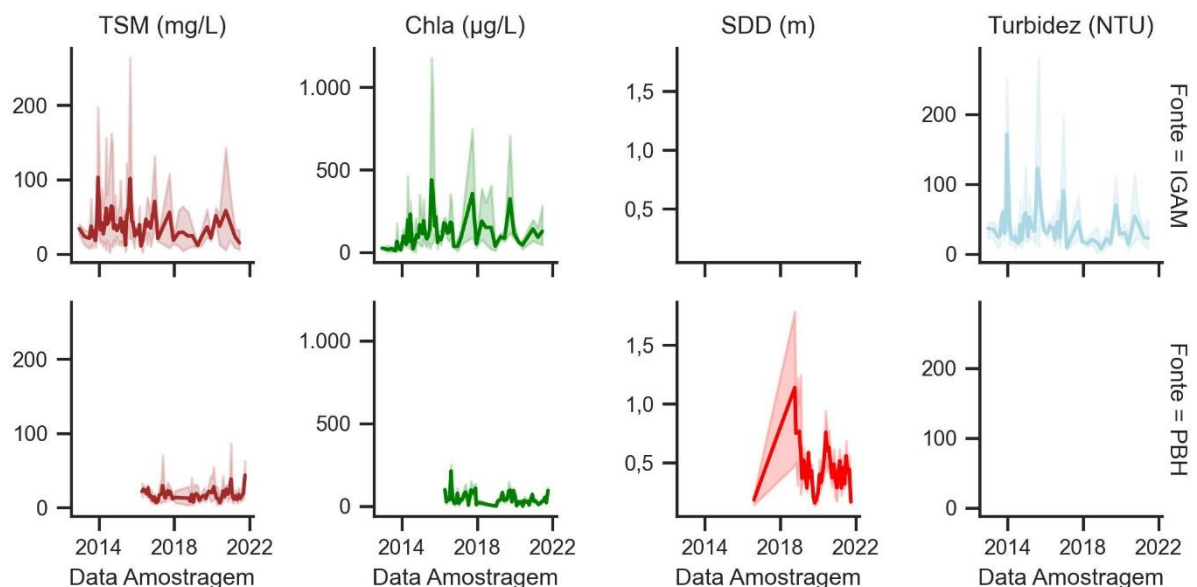
Tabela 5 – Principais estatísticas descritivas dos parâmetros de qualidade da água monitorados na Lagoa da Pampulha segregados por fonte de monitoramento

Fonte	Parâmetro	N	Min	Máx	Média	Desvio Padrão
IGAM	TSM (mg/L)	180	2	264	37,7	37,0
	Chla ($\mu\text{g/L}$)	180	0,006	1.177,5	115,53	149,9
	SDD (m)	0	-	-	-	-
	Turbidez (NTU)	180	2,37	281	39,8	41,3
PBH	TSM (mg/L)	379	3	179,5	17,1	17,2
	Chla ($\mu\text{g/L}$)	384	0,1	272	43,9	44,6
	SDD (m)	249	0	2,3	0,45	0,31
	Turbidez (NTU)	0	-	-	-	-

Fonte: Próprio Autor (2023)

Ao avaliar a série histórica por parâmetro e por fonte de monitoramento, observa-se uma variação sazonal e anual da concentração dos parâmetros avaliados, inclusive com padrões semelhantes entre esses (Figura 16). Para os dados de TSM, Chla e Turbidez monitorados pelo IGAM, evidencia-se picos ocorrendo no final de 2013, 2015, e 2020. Para dados monitorados pela PBH, percebe-se picos para os parâmetros TSM e Chla próximos aos anos de 2017, 2020 e 2021.

Figura 16 – Série temporal dos parâmetros de qualidade de água por fonte de monitoramento na Lagoa da Pampulha



Legenda: Devido à existência de mais de um ponto com medição por data, a linha sólida representa a média dos registros ao longo do tempo e a hachura representa a variação do parâmetro para todos os pontos.

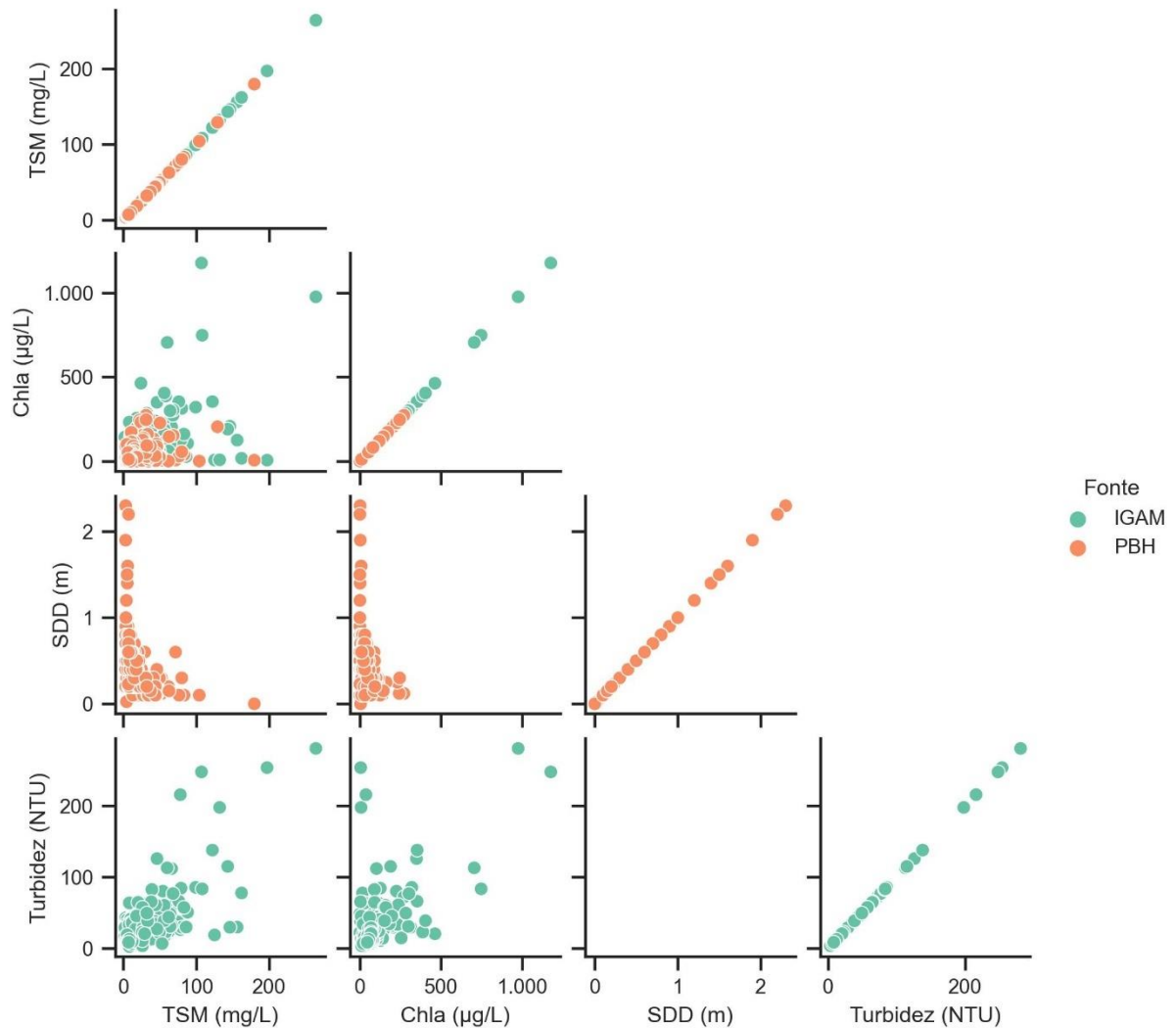
Fonte: Próprio Autor (2022)

Os dados de ambas as bases de monitoramento demonstram assimetria à direita com valores médios maiores que a mediana e a moda, característica que é rotineiramente observada em dados ambientais (SPERLING; VERBYLA; OLIVEIRA, 2020). Ao compararem-se valores monitorados *in situ* de Chla e TSM, para o período que se teve monitoramento de ambas as fontes, observam-se valores de maior magnitude monitorados pelo IGAM (Figura 17). Infelizmente, por se tratar de dados secundários e o monitoramento não ser realizado de forma concomitante por ambas as fontes, não há elementos suficientes para explicar tal diferença. Cabe avaliar em estudos futuros, as datas de monitoramento e de aplicação dos produtos químicos para o tratamento da Lagoa da Pampulha.

Destaca-se que o monitoramento do IGAM iniciou em 2013, incluindo período anterior ao início do tratamento da Lagoa da Pampulha e os dados da PBH começaram a partir de 2016, que corresponde ao início do tratamento.

Existem correlações entre as variáveis TSM e Turbidez ($\rho_s = 0,54$) e Chla e Turbidez ($\rho_s = 0,36$) para os dados do IGAM, com o coeficiente ρ_s representando a correlação de *Spearman*. Para o SDD, com dados monitorados apenas pela PBH, notam-se correlações negativas com a Chla ($\rho_s = -0,37$) e com o TSM ($\rho_s = -0,69$), o que é esperado, devido à maior transparência da água quando há a redução da concentração desses constituintes. Para os dados de TSM e Chla de ambas as bases, a correlação entre os parâmetros resultou em um valor $\rho_s=0,35$, indicando que a variação de TSM no reservatório pode ser explicada parcialmente pela variação da Chla (Figura 17).

Figura 17 – Distribuições e correlações entre os parâmetros de qualidade de água para as diferentes fontes de monitoramento da Lagoa da Pampulha

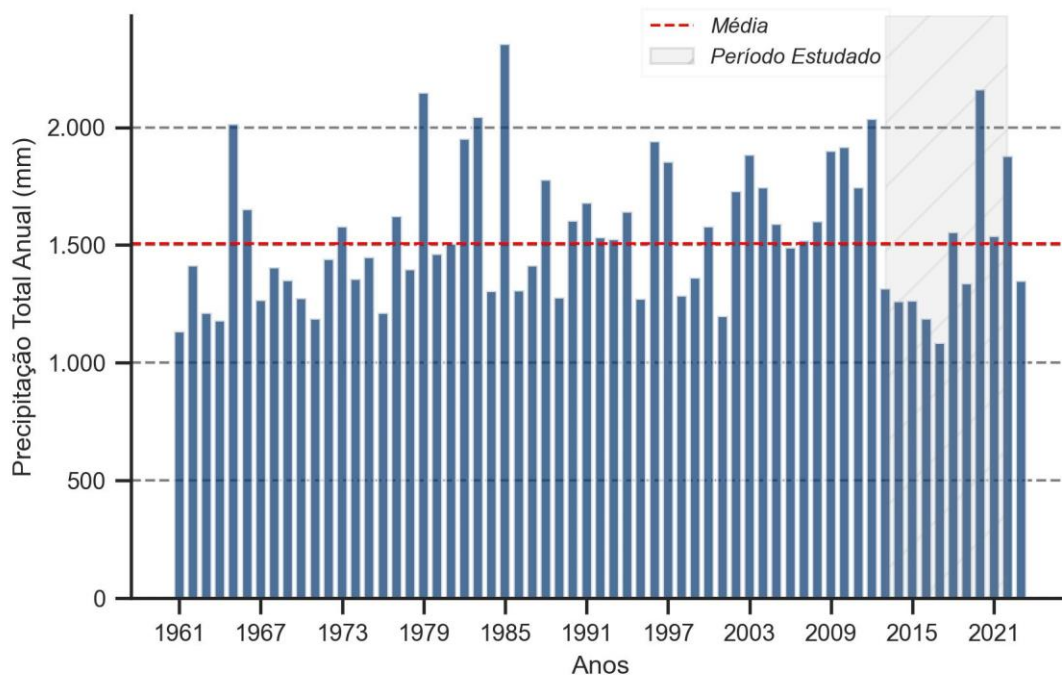


Fonte: Próprio Autor (2022)

5.1.2 Dados climatológicos

Os dados de precipitação total anual para os anos entre 1961 e 2023 obtidos na estação 83.587 do INMET, segregados por ano hidrológico são apresentados na Figura 18. Observa-se valor médio anual para a série histórica de aproximadamente 1500 mm. Para o período estudado, os anos mais chuvosos que a média são 2018, 2020, 2021 e 2022; e os anos menos chuvosos que a média são os anos de 2013, 2014, 2015, 2016, 2017 e 2019.

Figura 18 – Precipitação total anual para os anos de 1961 a 2023 para a estação INMET 83.587 localizada em Belo Horizonte.

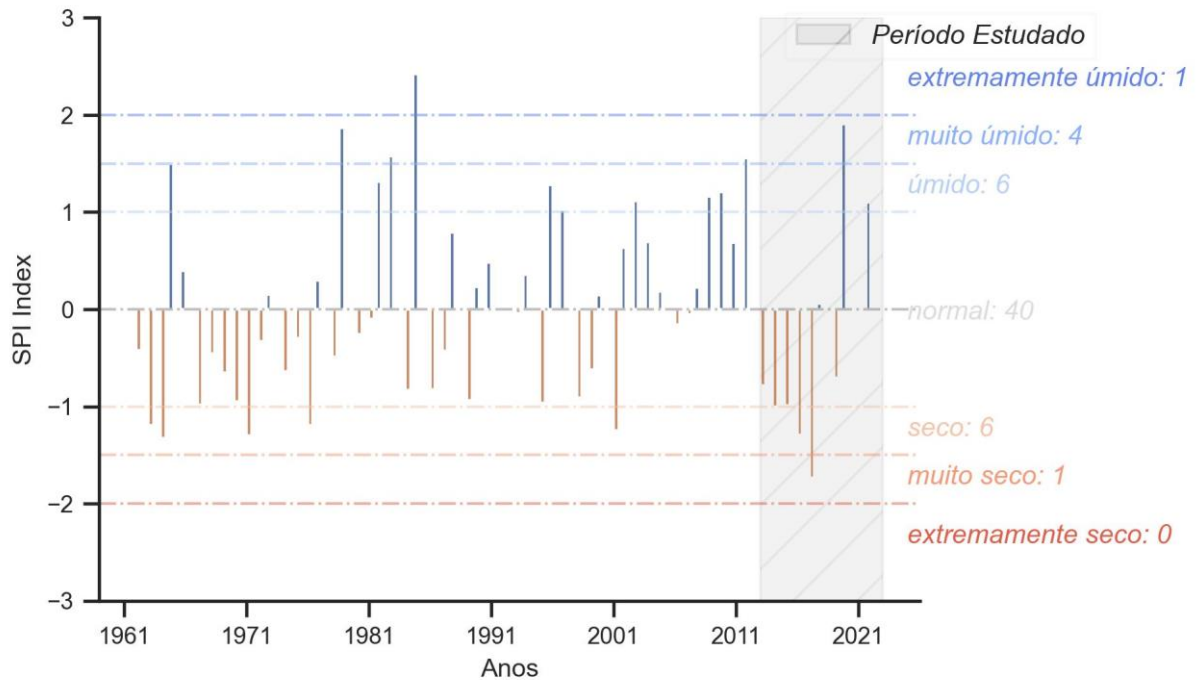


Nota: Precipitação total anual acumulada por ano hidrológico, com início em outubro e fim em setembro para cada ano.

Fonte: Próprio Autor (2022)

O cálculo do índice SPI_{12} para a estação pluviométrica INMET 83.587 indica maioria dos anos (69%) categorizados como normal (40). Entre os anos com valor negativo do índice SPI_{12} , registram-se seis (1963, 1971, 1976, 2001, 2014, 2016) anos na categoria seco (10%), um ano (2017) na categoria muito seco (2%) e nenhum ano na categoria extremamente seco. Entre os anos com valor positivo do índice SPI_{12} , registram-se seis anos (1965, 1982, 1996, 2003, 2009, 2022) na categoria úmido (10%), quatro anos (1979, 1983, 2012, 2020) na categoria muito úmido (7%) e um ano (1985) na categoria extremamente úmido (2%). Para o período estudado (2013 a 2022), o SPI_{12} calculado demonstra variação entre as cinco classes citadas, sendo o ano de 2017 categorizado como muito seco, os anos de 2014 e 2016 como secos, os anos de 2013, 2015, 2018, 2019 e 2021 como normais, o ano de 2022 como úmido e o ano de 2020 como muito úmido (Figura 19).

Figura 19 – Índice SPI₁₂ para os dados de precipitação disponibilizados para a estação INMET 83.587 localizada em Belo Horizonte



Fonte: Próprio Autor (2022)

5.1.3 Dados de sensoriamento remoto

Para o período estudado, a Lagoa da Pampulha conta com 196 e 389 imagens disponíveis, respectivamente, para o Landsat-8 e Sentinel-2. Destas, aquelas que apresentam imagens sem nuvens e sem reflexão especular na área do reservatório e, dentro do intervalo de +- 2 dias em relação à data do monitoramento limnológico, correspondem a apenas 6,1% (11) das imagens do Landsat-8 e 2,6% (10) das imagens do Sentinel-2 (Tabela 6). A quantidade de registros por parâmetro é variável devido às lacunas de monitoramento para algumas datas e a não realização do monitoramento de SDD pelo IGAM e de turbidez pela PBH.

Tabela 6 – Dados dos satélites combinados com os parâmetros de qualidade da água monitorados (2013 – 2022)

Satélite	Quantidade de Imagens	Quantidade de dados por parâmetro			
		TSM	Chla	SDD	Turbidez
Landsat-8	11	45	49	23	16
Sentinel-2	10	53	54	16	9

Fonte: Próprio Autor (2022)

Para o Landsat-8, sete imagens representam o período chuvoso e quatro imagens o período seco. Dentre as dez imagens do Sentinel-2, uma imagem representa o

período chuvoso e as outras nove representam o período seco. Conjuntos de dados desbalanceados, conforme notado para o Sentinel-2, refletem em construção de modelos favoráveis a certa época do ano e dificulta o avanço na análise de padrões da qualidade para épocas distintas (CRESSON et al., 2019; YANG et al., 2022).

Para os dados *in situ* de qualidade da água correspondentes à passagem do Landsat-8 com a tolerância de +- 2 dias, percebem-se valores de TSM (n=45) de 3 a 107 mg/L, Chla (n=49) de 0,1 a 1.177,5 µg/L, SDD (n=23) de 0,02 a 0,6 m e Turbidez (n=16) de 3,4 a 248 NTU (Tabela 7). Para os dados de qualidade da água correspondentes à passagem do Sentinel-2 com a tolerância de +- 2 dias, notam-se valores *in situ* de TSM (n=53) de 3,0 a 80 mg/L, Chla (n=54) de 1,9 a 353,2 µg/L, SDD (n=16) de 0,1 a 0,8 m e Turbidez (n=9) de 6,6 a 66,1 NTU (Tabela 7).

Ressalta-se que os valores máximos observados de TSM, Chla e Turbidez referente a datas de passagem do Landsat-8 ocorreram no mesmo dia (30/07/2015) no ponto PV 230 em monitoramento realizado pelo IGAM.

Tabela 7 – Principais estatísticas dos parâmetros de qualidade da água monitorados na Lagoa da Pampulha de acordo com a correspondência +-2 dias de passagem do satélite

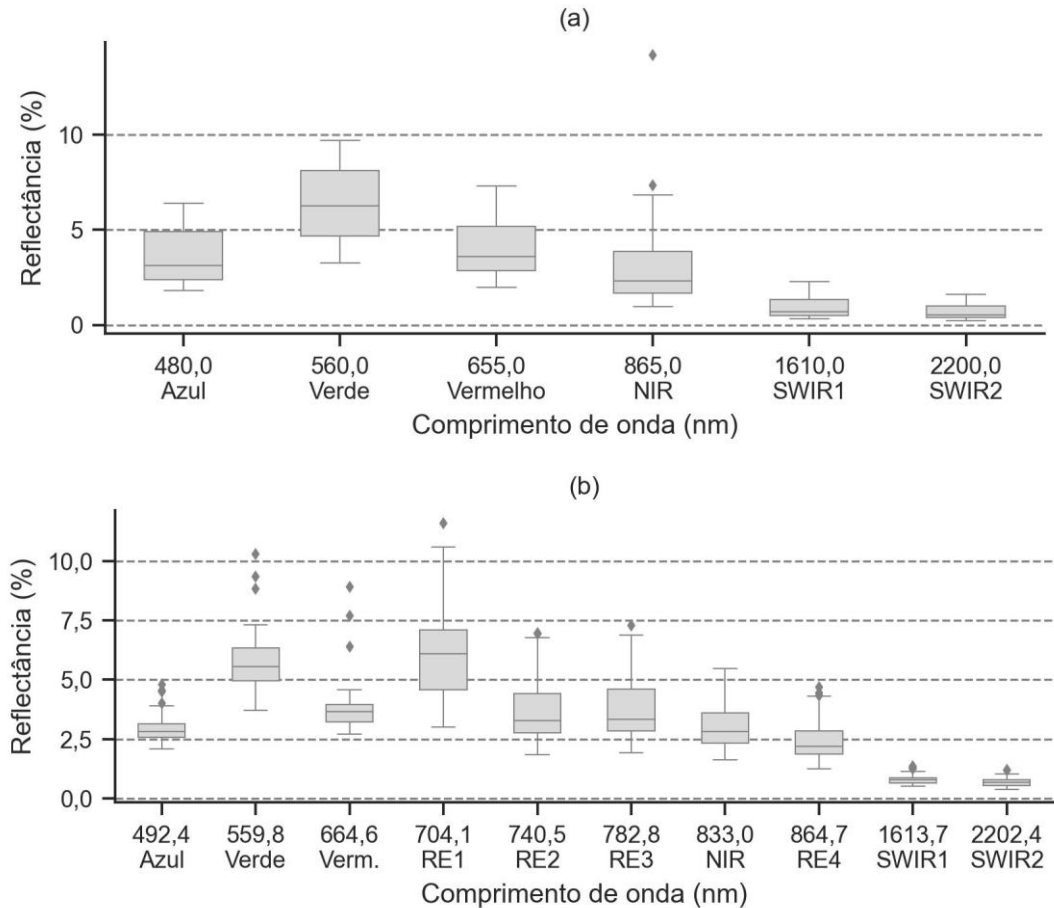
Satélite	Parâmetro	N	Min	Máx	Média	Desvio Padrão
Landsat-8	TSM (mg/L)	45	3,0	107,0	28,0	23,8
	Chla (µg/L)	49	0,1	1.177,5	96,4	194,0
	SDD (m)	23	0,02	0,6	0,32	0,16
	Turbidez (NTU)	16	3,4	248,0	50,0	60,0
Sentinel-2	TSM (mg/L)	53	3,0	80,0	20,1	16,3
	Chla (µg/L)	54	1,9	353,2	55,9	66,2
	SDD (m)	33	0,1	0,8	0,41	0,19
	Turbidez (NTU)	9	6,6	66,1	25,2	20,6

Fonte: Próprio Autor (2023)

Haja vista a inexistência de informações de espectrorradiômetro para validação do método de correção atmosférica SIAC, tomou-se o cuidado de inspecionar cada imagem, buscando não selecionar imagens que apresentassem influência de nuvens sobre os pontos de monitoramento. Nota-se que o comportamento espectral das imagens selecionadas (Figura 20) encontra-se conforme esperado para águas continentais ricas em Chla, com pico de reflectância no verde (550 nm) e na região da

borda do vermelho/NIR (710 nm) devido ao espalhamento causado pela presença de sólidos suspensos inorgânicos e do fitoplâncton (ODERMATT et al., 2012).

Figura 20 – Distribuição dos valores de reflectância nos pontos de monitoramento para as bandas dos satélites utilizadas para a modelagem da qualidade da água



Legenda: (a) Landsat-8 (n=49), (b) Sentinel-2 (n=54). Nota: “n” representa a quantidade de dados de reflectância por banda.

Fonte: Próprio Autor (2022)

5.2 Inferência de parâmetros de qualidade de água a partir de dados do sensoriamento remoto

Foram realizadas regressões univariadas e multivariadas para os parâmetros de qualidade da água selecionados e as variáveis independentes obtidas por meio dos dados de sensoriamento remoto, considerando-se cada satélite separadamente. Para a inferência dos parâmetros por meio de regressão univariada, utilizaram-se todas as bandas selecionadas para cada satélite, os índices listados na Tabela 2 e as melhores correlações entre bandas – calculada pelo coeficiente de correlação linear de *Pearson*; podendo ser a multiplicação, a divisão ou índices entre bandas. Na Figura

21, Figura 25, Figura 29 e Figura 33, são apresentadas todas as variáveis independentes testadas nas regressões univariadas. Para a regressão multivariada foram utilizadas as variáveis independentes destacadas em negrito nestas Figuras.

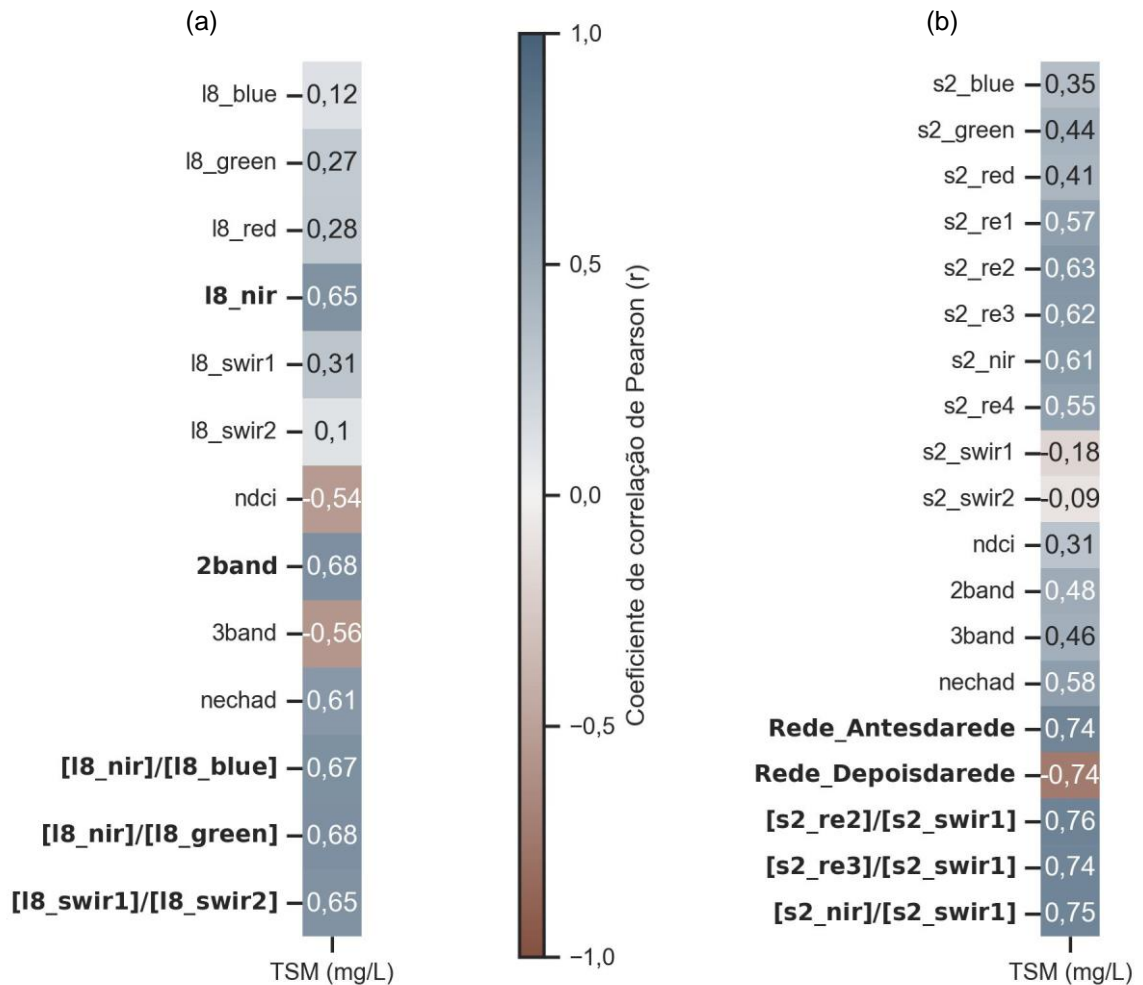
Levando em conta a distinção de satélites, tipos de regressão, parâmetros de qualidade de água e, possibilidades de combinação entre as informações do sensoriamento remoto (variáveis independentes), muitos modelos foram obtidos. Todos estes modelos foram analisados, mas somente são aqui mostrados aqueles que apresentaram os melhores resultados, seja de base empírica, seja através de adaptação de modelos da literatura. A escolha dos melhores modelos para cada satélite fundamentou-se nos maiores valores da moda do R^2 na etapa de calibração e nos menores valores da moda do RMSE na etapa de validação.

5.2.1 Sólidos Suspensos Totais (mg/L)

A correlação do parâmetro TSM com dados do Landsat-8, apresentada na Figura 21a, mostra-se promissora para a utilização da banda do infravermelho ($I8_nir$), com coeficiente de correlação linear de *Pearson* (r) equivalente a $r = 0,65$. A razão de duas bandas ($2bandas$) e as razões da banda do infravermelho ($I8_nir$) com as bandas azul ($I8_blue$) e verde ($I8_green$) mostram uma correlação positiva intermediária ($r > 0,67$). A correlação positiva intermediária ($r = 0,65$) entre as bandas com maiores comprimentos de onda ($I8_swir1/I8_swir2$) indica que a concentração e as características do TSM presente no reservatório impactam no aumento dos valores de reflectância registrados.

Correlações do parâmetro TSM com dados do Sentinel-2, apresentadas na Figura 21b, mostram que a banda individual que apresentou maior correlação foi a RedEdge-2 ($s2_re2$), com $r = 0,63$. A divisão dessa pela banda SWIR1 ($s2_swir1$) apresenta melhora da correlação para $r = 0,76$. Também apresentam forte correlação positiva com o parâmetro, a banda RedEdge-3 ($s2_re3$) e banda NIR ($s2_nir$) divididas pela SWIR1 ($s2_swir1$) com $r = 0,74$ e $0,75$, respectivamente. A posição do ponto de monitoramento limnológico em relação à cortina de retenção também apresenta forte correlação com o TSM, o que reflete a capacidade da cortina em barrar sólidos e formar distintos ambientes em termos de concentração de TSM na Lagoa da Pampulha.

Figura 21 – Correlação de Pearson para as principais variáveis independentes de entrada para a modelagem de TSM para ambos os satélites.

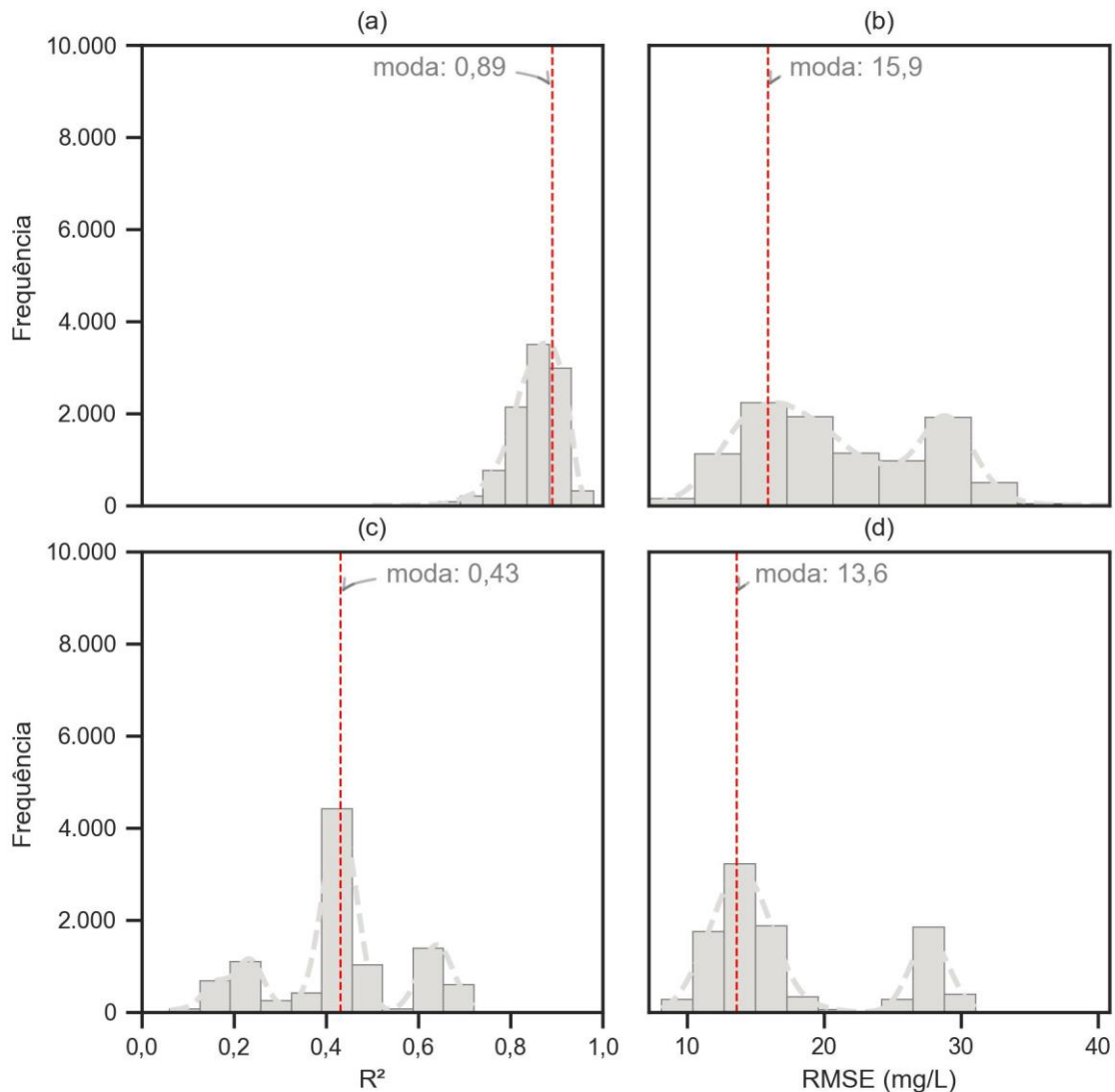


Legenda: (a) Landsat-8, (b) Sentinel-2. Nota: os termos em negrito representam as cinco variáveis independentes mais correlacionadas com o parâmetro modelado.

Fonte: Próprio Autor (2022)

Quando utilizado os dados do Landsat-8 para modelar TSM, o modelo empírico é aquele que utiliza o método de regressão *Random Forest*, cujas variáveis independentes de entrada estão destacadas em negrito na Figura 21a. A moda do R^2 para a calibração é igual a 0,89 e a moda do RMSE para validação é igual a 15,9 mg/L (Figura 22a). O modelo da literatura ajustado é aquele que utiliza a banda do infravermelho (i8_nir) para a estimativa de TSM por meio de regressão linear. As métricas calculadas são $R^2_{moda} = 0,43$ na calibração e $RMSE_{moda} = 13,6$ mg/L na validação (Figura 22b).

Figura 22 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R² e RMSE para o parâmetro TSM modelado a partir dos dados do Landsat-8

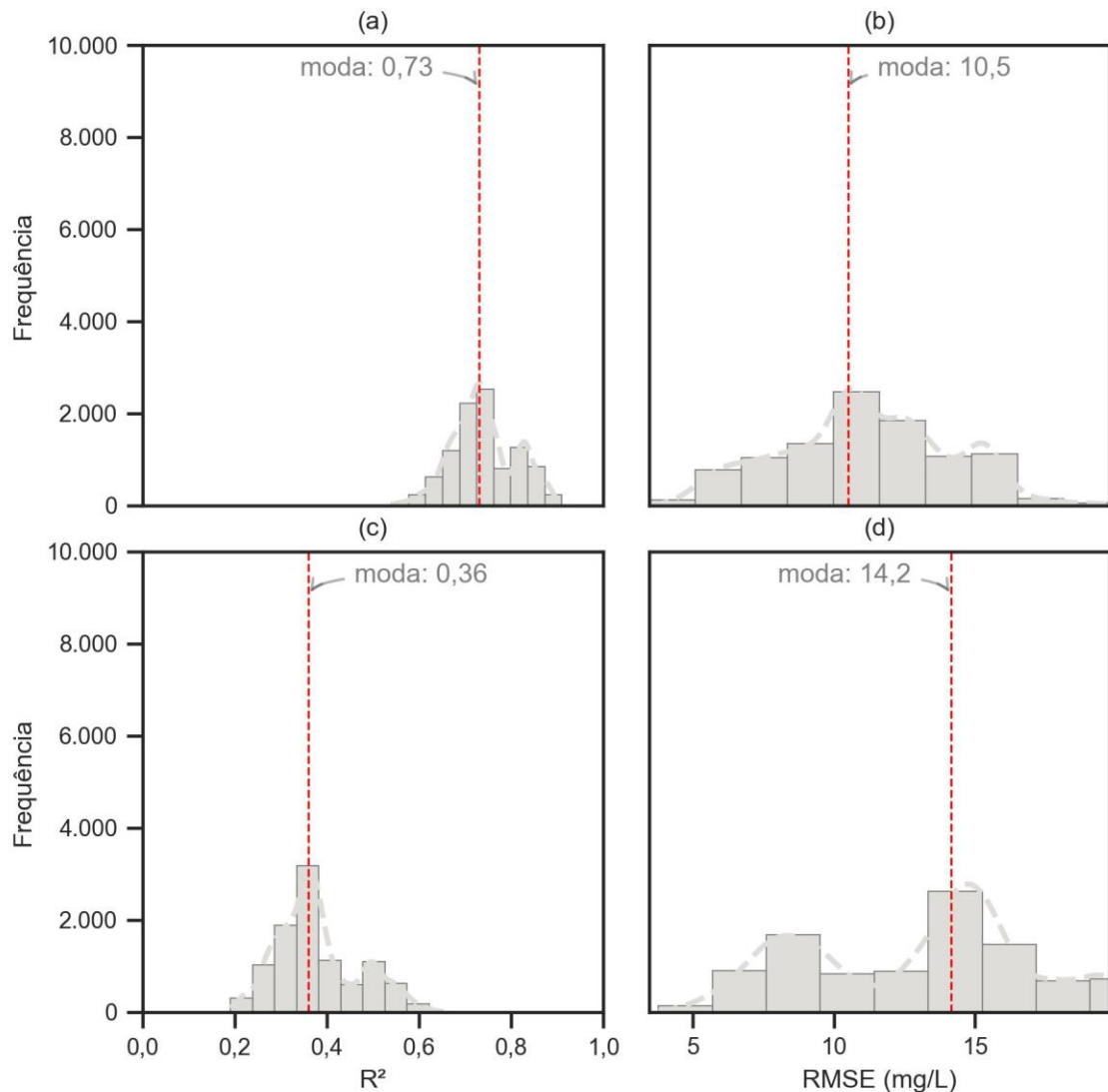


Legenda: (a) e (b) modelo empírico, (c) e (d) modelo da literatura ajustado com os dados limnológicos da Lagoa da Pampulha. Nota: R² se refere à calibração e RMSE à validação

Fonte: Próprio Autor (2022)

Quando utilizado os dados do Sentinel-2 modelando TSM, o modelo empírico é aquele que utiliza o método de regressão *Random Forest* com variáveis independentes de entrada destacadas em negrito na Figura 21b. A moda do R² para a calibração é igual a 0,73 e a moda do RMSE para validação é igual a 10,5 mg/L (Figura 23a). Para o modelo da literatura, observaram-se melhores métricas quando utilizada a regressão linear entre o parâmetro e a banda do infravermelho (s2_nir), com moda do R² de 0,36 e a moda do RMSE é igual a 14,2 mg/L (Figura 23b).

Figura 23 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R² e RMSE para o parâmetro TSM modelado a partir dos dados do Sentinel-2



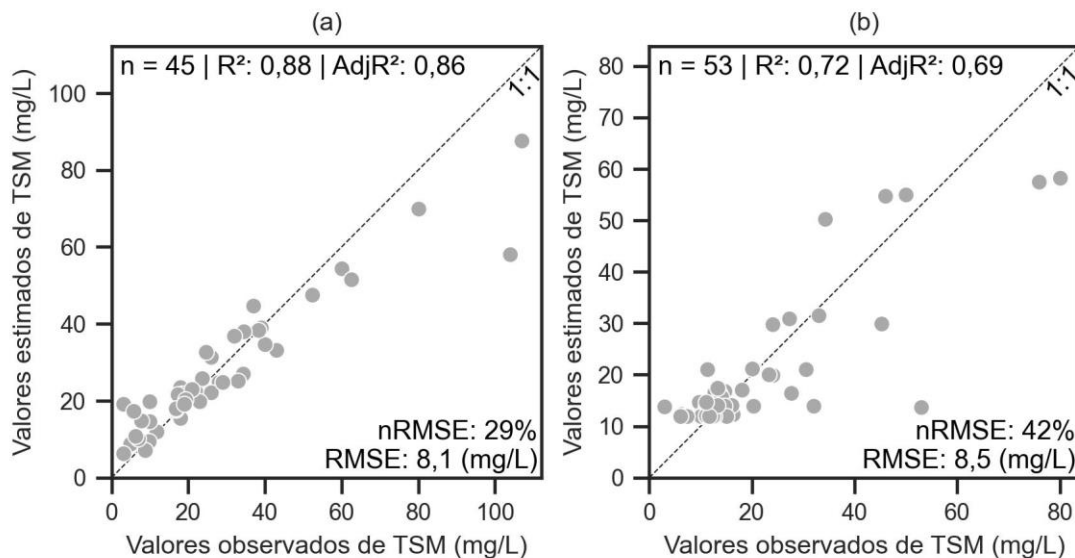
Legenda: (a) e (b) modelo empírico, (c) e (d) modelo da literatura ajustado com os dados limnológicos da Lagoa da Pampulha. Nota: R² se refere à calibração e RMSE à validação

Fonte: Próprio Autor (2022)

Dentre os modelos apresentados foram escolhidos aqueles obtidos empiricamente visto que apresentaram as melhores métricas de desempenho na modelagem de TSM. Para ambos os satélites, percebe-se que quando comparados aos modelos ajustados da literatura, os modelos empíricos superam em duas vezes o valor da moda do R² na etapa de calibração. Para o Sentinel-2, a moda do RMSE do modelo empírico apresenta valor inferior, assim como a distribuição favorece a utilização do modelo empírico.

Ao aplicar os modelos obtidos considerando-se toda a amostra de dados de TSM disponíveis para o Landsat-8 (n=45), verifica-se um valor de R^2 igual a 0,88, R^2 ajustado igual a 0,86, RMSE igual a 8,1 mg/L, e nRMSE igual a 29% (Figura 24a). Para o Sentinel-2, com os dados de TSM disponíveis (n=53), verifica-se R^2 igual a 0,72, R^2 ajustado de 0,69, RMSE igual a 8,5 mg/L e nRMSE igual a 42% (Figura 24b).

Figura 24 – Comparação entre valores estimados pelos modelos escolhidos e valores observados de TSM para ambos os satélites



Legenda: (a) Landsat-8, (b) Sentinel-2

Fonte: Próprio Autor (2022)

No caso do Landsat-8, o modelo apresenta melhor ajuste aos dados e se aproxima mais da diagonal 1:1, principalmente para baixas concentrações de TSM (<40 mg/L). Para o Sentinel-2 nota-se agrupamento de valores próximo à região de concentração TSM entre 10 e 20 mg/L e para valores elevados nota-se a superestimação e subestimação do modelo com o distanciamento da diagonal 1:1.

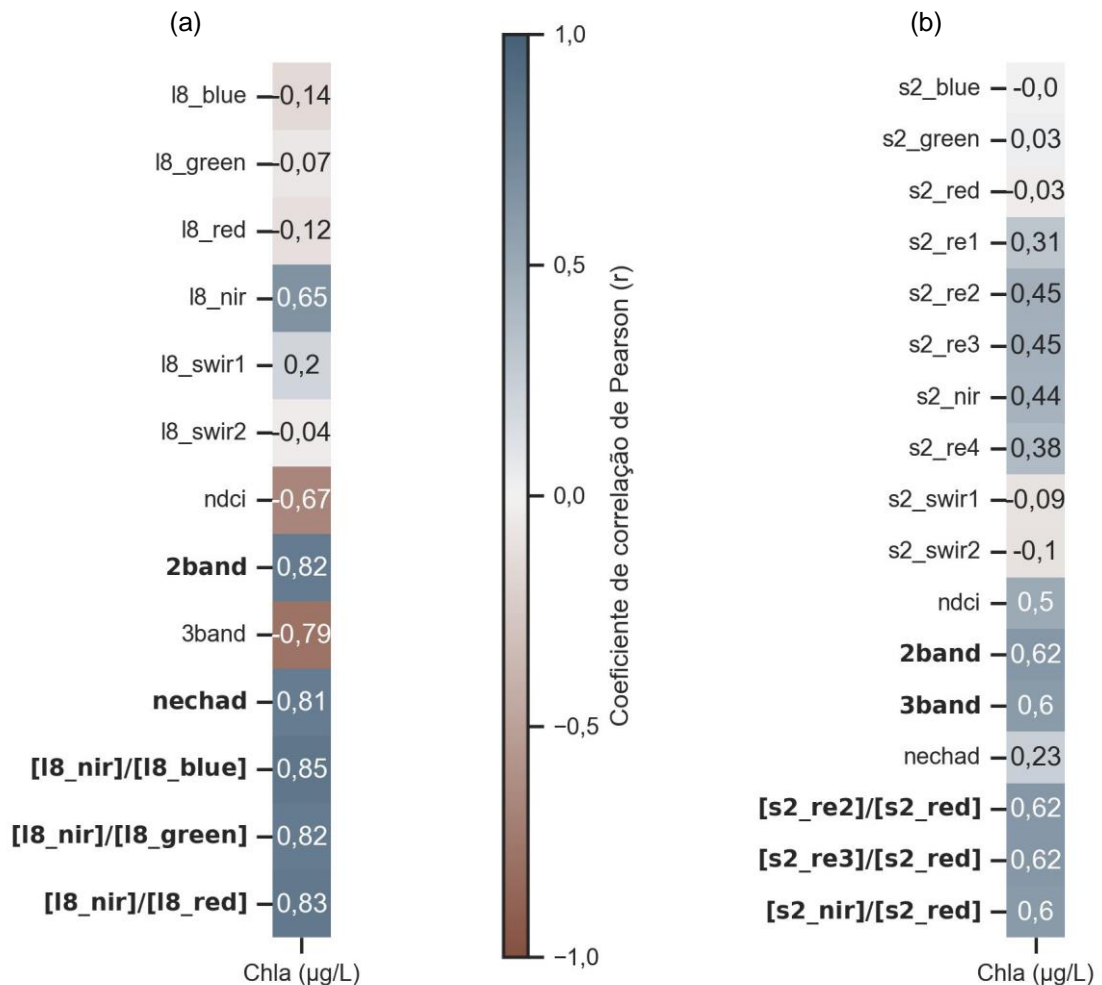
5.2.2 Clorofila-a ($\mu\text{g/L}$)

A correlação do parâmetro Chla com dados do Landsat-8, apresentada na Figura 25a, demonstra, dentre as bandas individuais selecionadas do satélite, $r=0,65$ com a banda do infravermelho (I8_nir). As variáveis 2bandas e Nechad apresentam fortes correlações positivas com a Chla, sendo $r=0,82$ e $r=0,81$, respectivamente. Razões entre bandas com a banda do NIR (I8_nir) no numerador e as bandas do azul

(I8_blue), verde (I8_green) e vermelho (I8_red) no denominador também apresentam fortes correlações positivas, sendo $r=0,85$, $r=0,82$ e $r=0,83$, respectivamente.

Para as correlações do parâmetro com dados do Sentinel-2, apresentadas na Figura 25b, as bandas individuais que apresentaram maior correlação são as bandas RedEdge-2 (s2_re2) e RedEdge-3 (s2_re3), ambas com $r=0,45$. Correlações fortes e positivas são observadas para as variáveis 2bandas e 3bandas ($r>0,6$), e razões entre bandas RedEdge-2 (s2_re2) com vermelho (s2_red), RedEdge-3 (s2_re3) com vermelho (s2_red) e NIR (s2_nir) com vermelho (s2_red).

Figura 25 – Correlação de Pearson para as principais variáveis independentes de entrada para a modelagem de Chla para ambos os satélites.

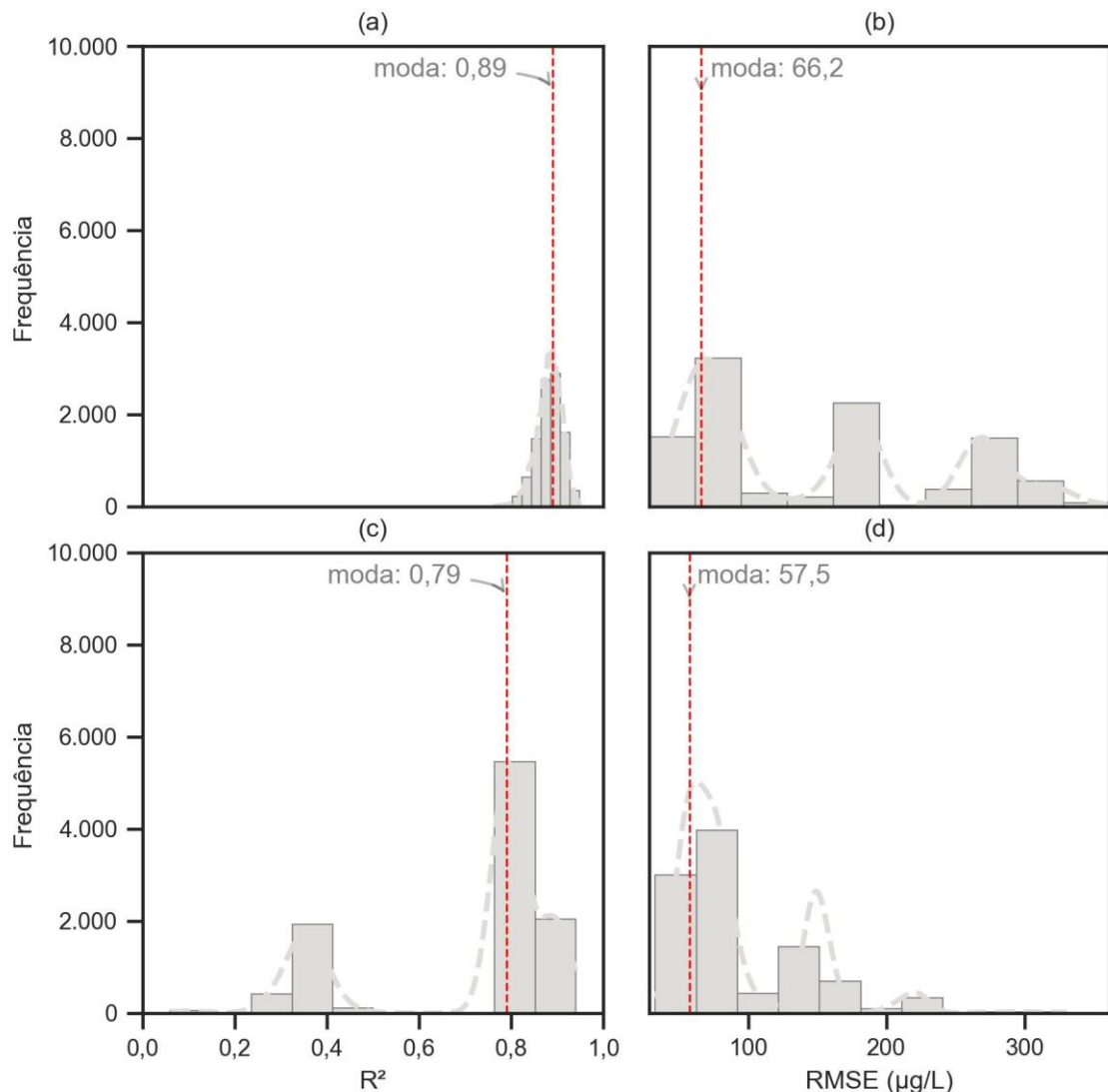


Legenda: (a) Landsat-8, (b) Sentinel-2. Nota: os termos em negrito representam as cinco variáveis independentes mais correlacionadas com o parâmetro modelado.

Fonte: Próprio Autor (2022)

Para o Landsat-8, o modelo empírico obtido para a Chla é aquele que utiliza o método de regressão *Random Forest* com variáveis independentes de entrada destacadas em negrito na Figura 25a. A moda do R^2 para a calibração é igual a 0,89 (Figura 26a) e a moda do RMSE para validação é igual a 66,2 $\mu\text{g/L}$ (Figura 26b). O modelo da literatura ajustado é aquele que utiliza o índice 2bandas para a estimativa de Chla por meio de regressão linear. As métricas calculadas são $R^2_{\text{moda}} = 0,79$ (Figura 26c) na calibração e $\text{RMSE}_{\text{moda}} = 57,5 \mu\text{g/L}$ na validação (Figura 26d). O modelo ajustado da literatura apresentou um comportamento bimodal para a distribuição dos valores de R^2 (Figura 26c), o que indica que o mesmo não é capaz de estimar com o mesmo desempenho valores muito distintos de Chla.

Figura 26 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R^2 e RMSE para o parâmetro Chla modelado a partir dos dados do Landsat-8

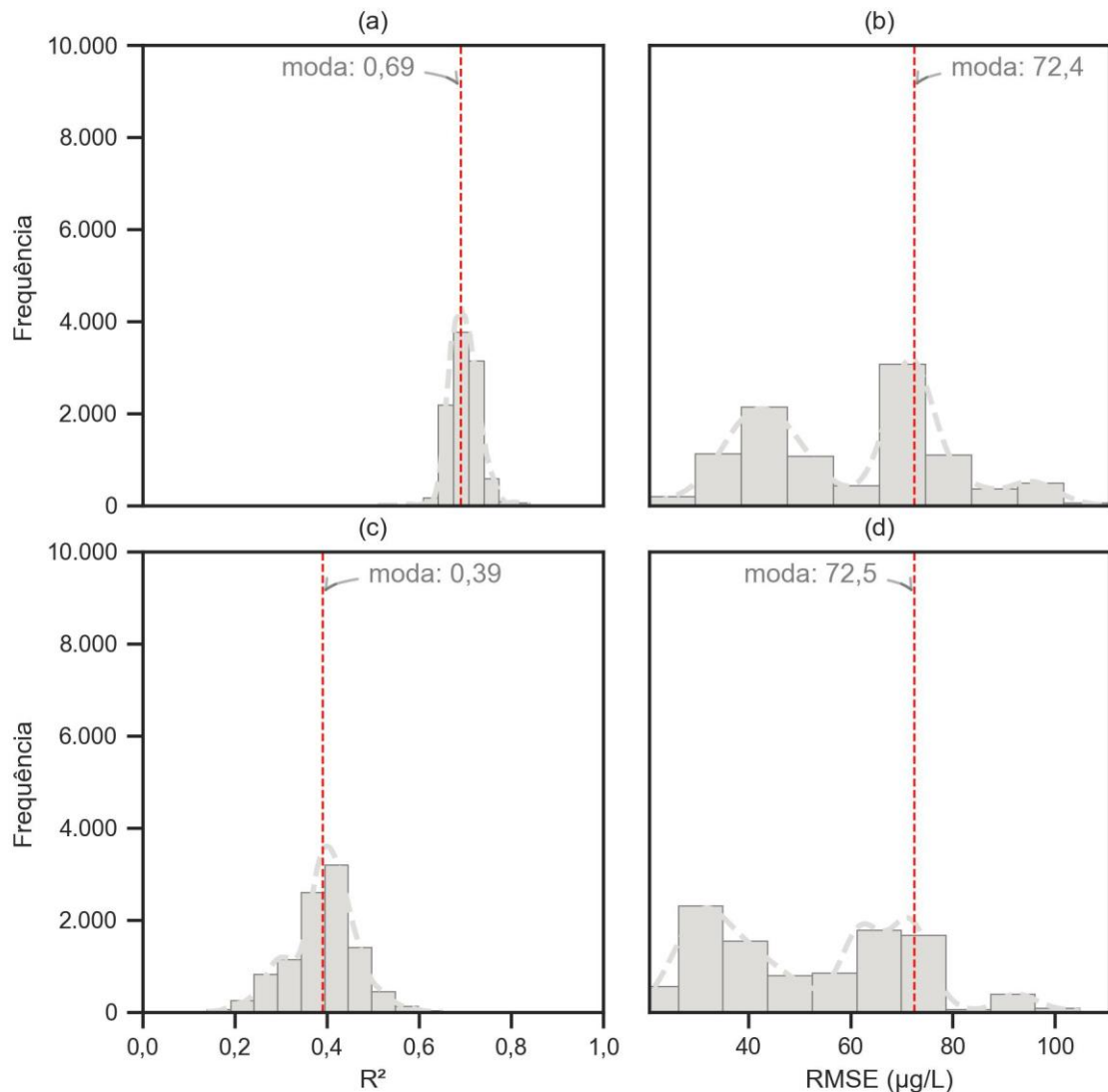


Legenda: (a) e (b) modelo empírico, (c) e (d) modelo da literatura ajustado com os dados limnológicos da Lagoa da Pampulha. Nota: R^2 se refere à calibração e RMSE à validação

Fonte: Próprio Autor (2022)

Para o Sentinel-2, o modelo empírico obtido para a Chla é aquele que utiliza o método de regressão *Random Forest* com variáveis independentes de entrada destacadas em negrito na Figura 25b. A moda do R^2 para a calibração é igual a 0,69 e a moda do RMSE para validação é igual a 72,4 $\mu\text{g/L}$ (Figura 27a). O modelo da literatura ajustado é aquele que utiliza a variável independente 2bandas para a estimativa de Chla por meio de regressão linear. As métricas calculadas são $R^2_{\text{moda}} = 0,39$ na calibração e $\text{RMSE}_{\text{moda}} = 72,5 \mu\text{g/L}$ na validação (Figura 27b). Apesar de observar alta frequência para a região de 30 $\mu\text{g/L}$, o valor que representa a moda está próximo da região de 70 $\mu\text{g/L}$.

Figura 27 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R² e RMSE para o parâmetro Chla modelado a partir dos dados do Sentinel-2.



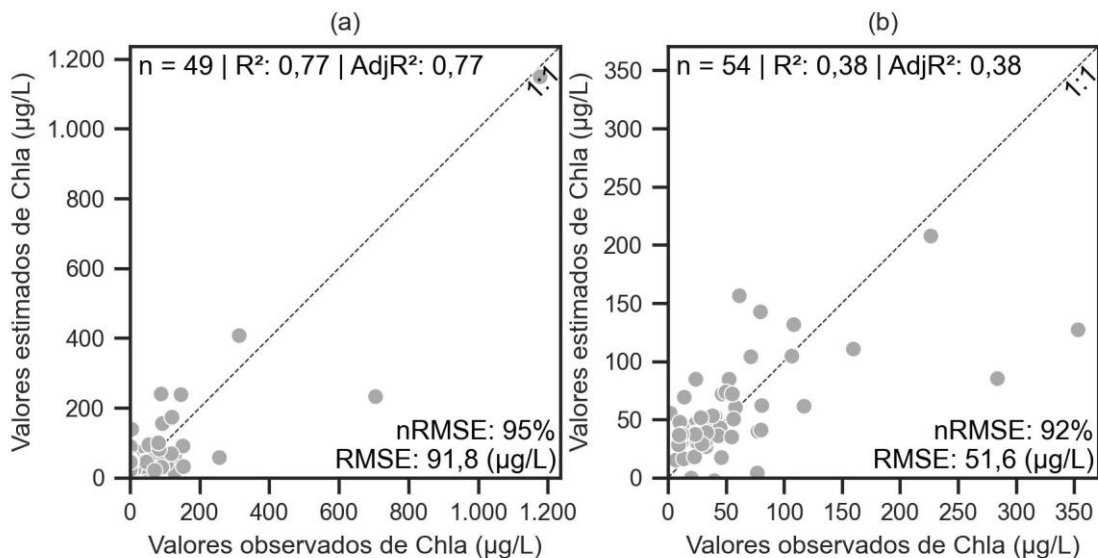
Legenda: (a) e (b) modelo empírico, (c) e (d) modelo da literatura ajustado com os dados limnológicos da Lagoa da Pampulha. Nota: R² se refere à calibração e RMSE à validação

Fonte: Próprio Autor (2022)

No caso do Landsat-8, observa-se proximidade entre as modas do R² e valor menor da moda do RMSE para o modelo ajustado da literatura em relação ao modelo empírico. Para o Sentinel-2, apesar do R² ser menor para o modelo da literatura, nota-se que para a validação o RMSE ficou próximo do modelo empírico. O RMSE do modelo ajustado da literatura para ambos os satélites apresenta amplitude de valores inferior em relação a amplitude de valores do modelo empírico. Estes fatores favoreceram a escolha dos modelos ajustados da literatura, para ambos os satélites, em relação aos modelos empíricos para a estimativa de Chla.

Ao aplicar os modelos obtidos considerando-se toda a amostra de dados de Chla disponíveis para o Landsat-8 (n=49), verifica-se valores de R^2 e R^2 ajustado iguais a 0,77, RMSE igual a 91,8 $\mu\text{g/L}$, e nRMSE igual a 95% (Figura 28a). Para o Sentinel-2, com os dados de Chla disponíveis (n=54), obtêm-se R^2 e R^2 ajustado iguais a 0,38, RMSE igual a 51,6 $\mu\text{g/L}$ e nRMSE igual a 92% (Figura 28b).

Figura 28 – Comparação entre valores estimados pelos modelos escolhidos e valores observados de Chla para ambos os satélites



Legenda: (a) Landsat-8, (b) Sentinel-2

Fonte: Próprio Autor (2022)

Para os modelos nota-se proximidade dos pontos da diagonal 1:1 para concentrações menores (<100 $\mu\text{g/L}$) e maior distanciamento para maiores concentrações. Devido à amplitude dos valores de Chla observados nos pares de dados do Landsat-8 (>1000 $\mu\text{g/L}$), o R^2 apresenta valores altos devido à correlação espúria e o RMSE de validação apresenta um valor maior que o valor de RMSE obtido para o Sentinel-2. Contudo, ao comparar o nRMSE para ambos os satélites, os valores são próximos, mostrando a proximidade de desempenho dos modelos. Para as maiores concentrações observadas de Chla, nota-se uma subestimação do modelo, sugerindo a ocorrência de uma saturação nos valores de reflectância. Isso pode estar associado à ocorrência de maiores concentrações de Chl-a e não necessariamente a absorção na região do vermelho aumentar e ao fato que maiores concentrações de Chl-a não levarem efeito proporcional ao espalhamento para regiões da borda do vermelho.

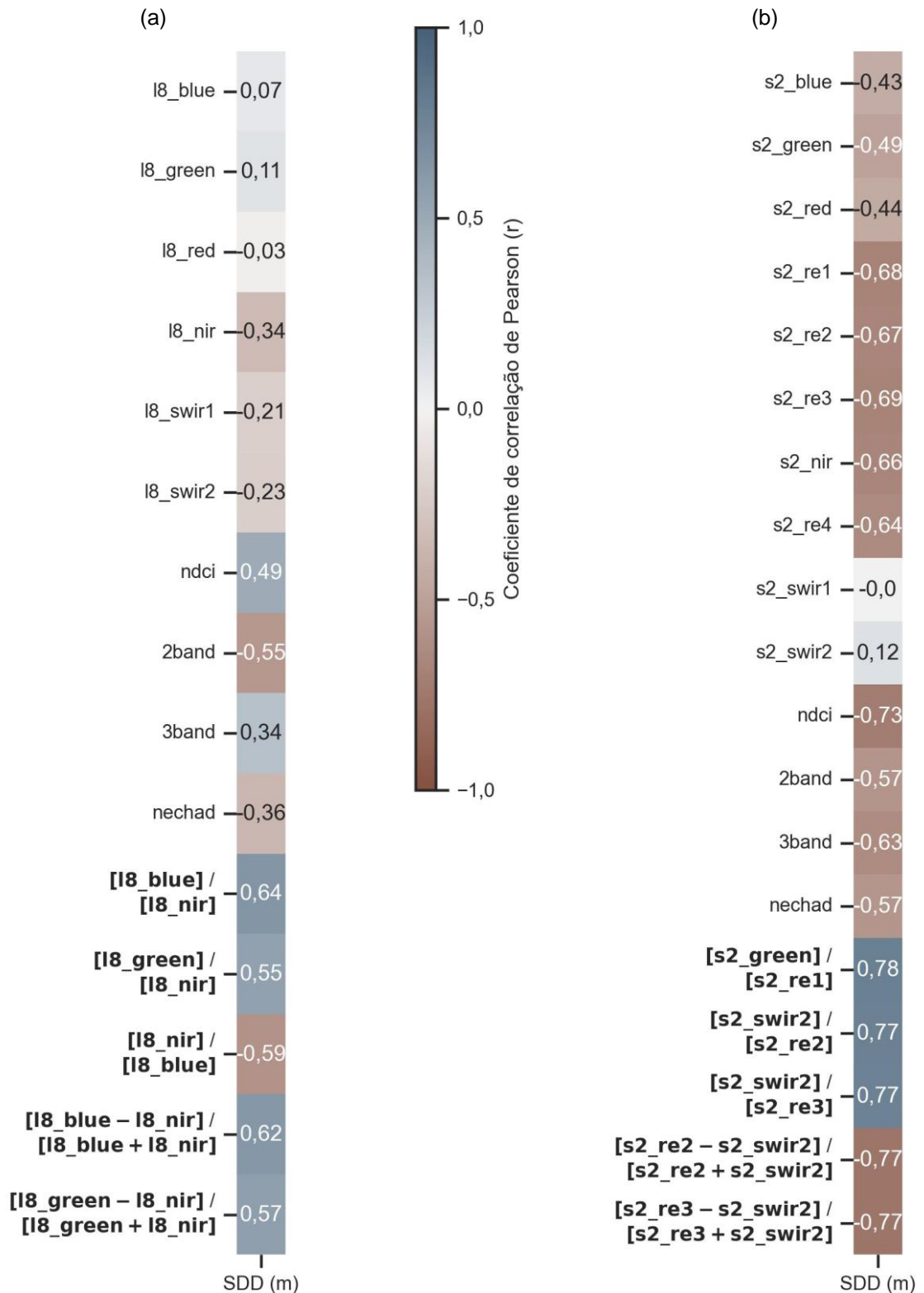
5.2.3 Profundidade do Disco de Secchi (m)

A correlação do parâmetro SDD com dados Landsat-8, apresentada na Figura 29a, demonstra, dentre as bandas individuais selecionadas do satélite, $r=-0,34$ com a banda do NIR (l8_nir). Correlações intermediárias ($r>0,54$ e $r<-0,54$) são observadas para razões entre bandas e índices que utilizaram as bandas do azul (l8_blue), verde (l8_green), e NIR (l8_nir).

Para as correlações do SDD com dados do Sentinel-2, apresentadas na Figura 29b, a banda individual que apresentou maior correlação foi a RedEdge-3 (s2_re3) com $r=-0,69$. Correlações fortes ($r>0,75$ e $r<-0,75$) são observadas para razões entre bandas e índices que utilizaram as bandas do verde (s2_verde), RedEdge-1 (s2_re1), RedEdge-2 (s2_re2), RedEdge-3 (s2_re3), e SWIR-2 (s2_swir2).

Conforme apresentado, os valores do coeficiente de correlação linear de *Pearson* (r) entre bandas e SDD normalmente são negativos, indicando que quanto maior a transparência, maior a absorção da radiação e menor a reflectância.

Figura 29 – Correlação de *Pearson* para as principais variáveis independentes de entrada para a modelagem de SDD para ambos os satélites.

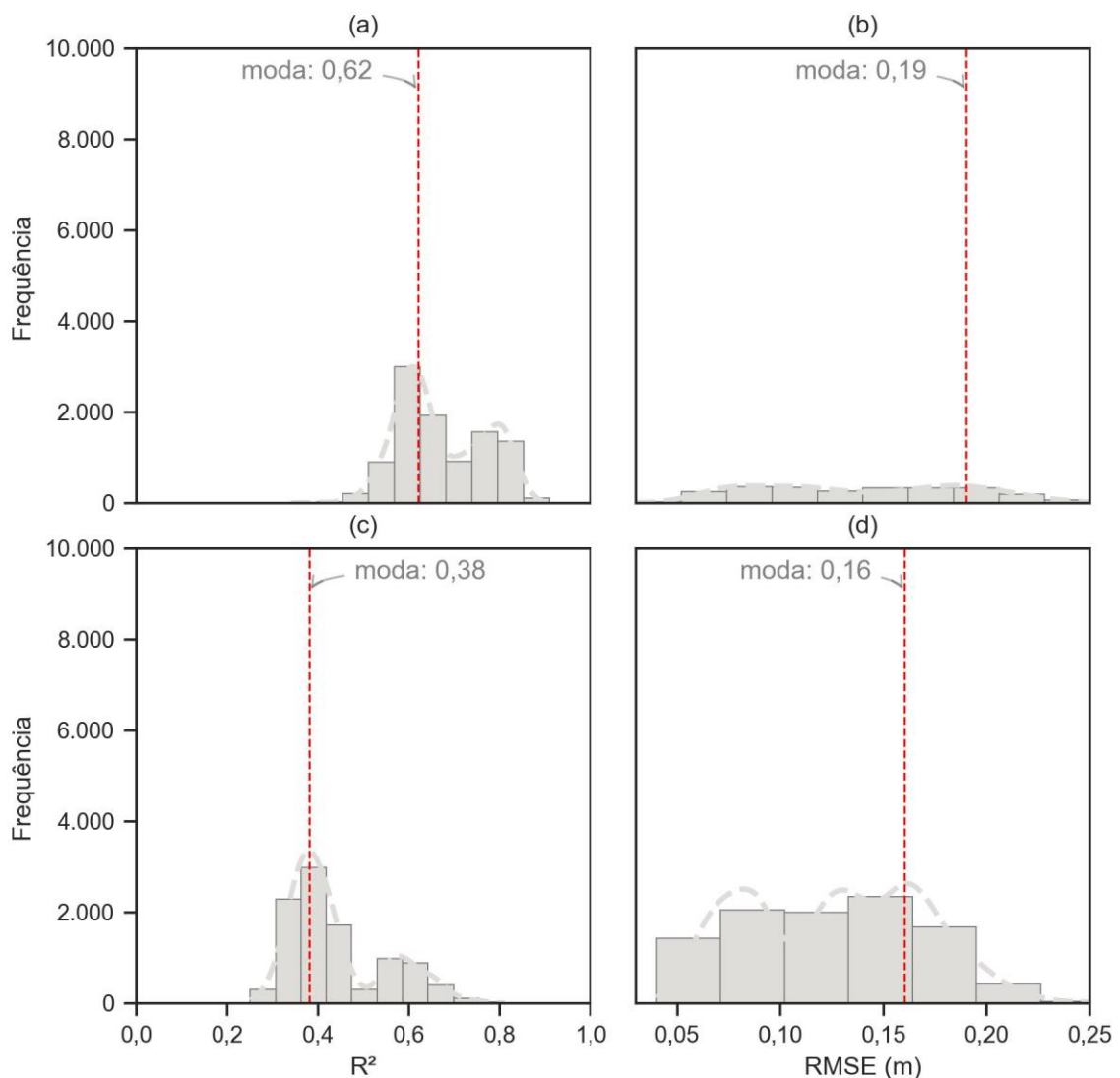


Legenda: (a) Landsat-8, (b) Sentinel-2. Nota: os termos em negrito representam as cinco variáveis independentes mais correlacionadas com o parâmetro modelado.

Fonte: Próprio Autor (2022)

Para o Landsat-8, o modelo empírico obtido para o SDD é aquele que utiliza o método de regressão *Random Forest*, com variáveis independentes de entrada destacadas em negrito na Figura 29a. A moda do R^2 para a calibração é igual a 0,62 e a moda do RMSE para validação é igual a 0,19 m (Figura 30a). O modelo da literatura ajustado é aquele que utiliza a razão de bandas $I8_nir/I8_blue$ para a estimativa de SDD por meio de regressão polinomial. As métricas calculadas são $R^2_{moda} = 0,38$ na calibração e $RMSE_{moda} = 0,16$ m na validação (Figura 30b).

Figura 30 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R^2 e RMSE para o parâmetro SDD modelado a partir dos dados do Landsat-8.

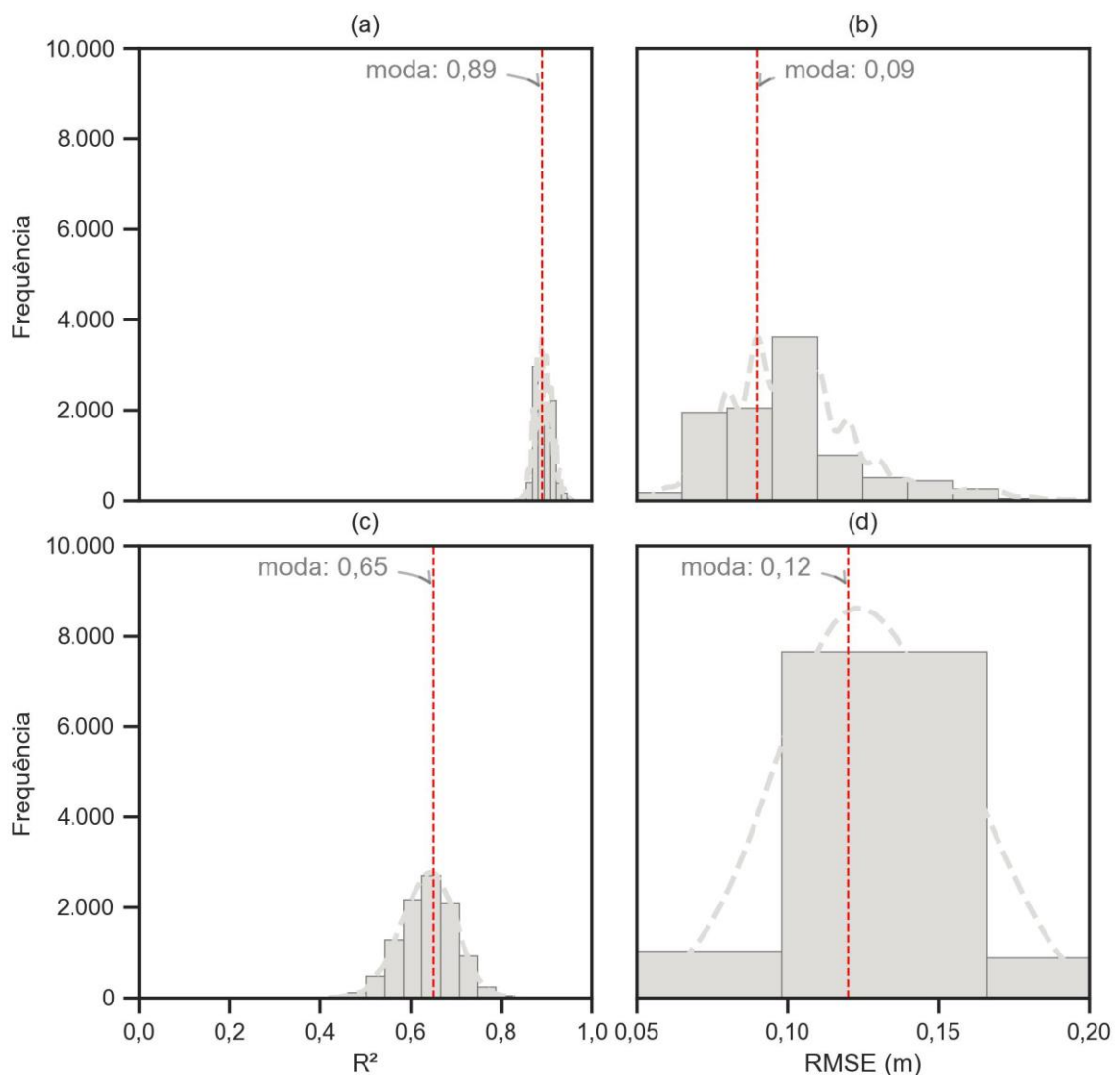


Legenda: (a) e (b) modelo empírico, (c) e (d) modelo da literatura ajustado com os dados limnológicos da Lagoa da Pampulha. Nota: R^2 se refere à calibração e RMSE à validação

Fonte: Próprio Autor (2022)

Para o Sentinel-2, o modelo empírico obtido para o SDD é aquele que utiliza o *Random Forest* com variáveis independentes de entrada destacadas em negrito na Figura 29b. A moda do R^2 para a calibração é igual a 0,89 e a moda do RMSE para validação é igual a 0,09 m (Figura 31a). O modelo da literatura ajustado é aquele que utiliza a banda RedEdge1 (s2_re1), para a estimativa de SDD por meio de regressão linear. As métricas calculadas são $R^2_{\text{moda}} = 0,65$ na calibração e $\text{RMSE}_{\text{moda}} = 0,12$ m na validação (Figura 31b).

Figura 31 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R^2 e RMSE para o parâmetro SDD modelado a partir dos dados do Sentinel-2.



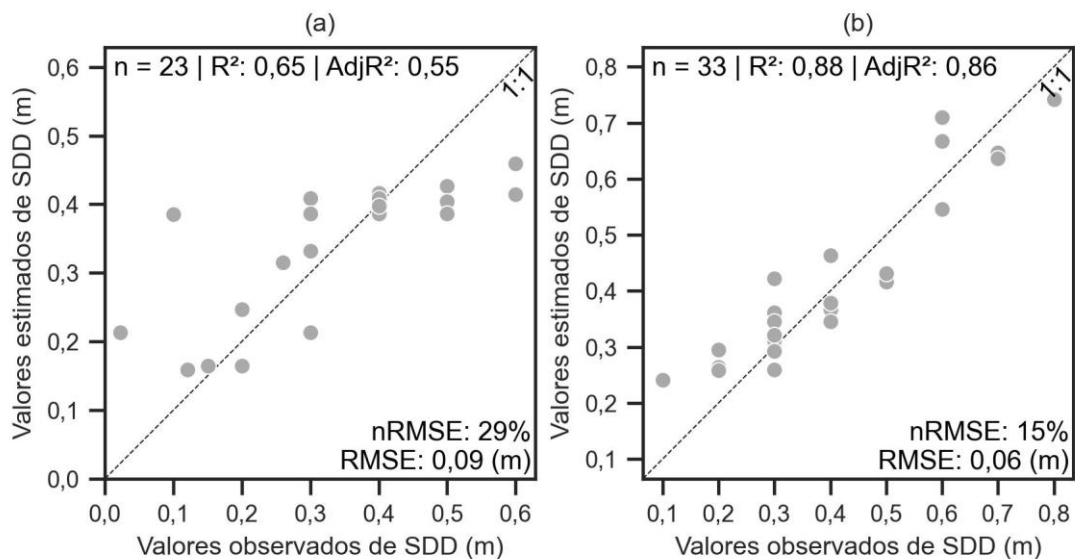
Legenda: (a) e (b) modelo empírico, (c) e (d) modelo da literatura ajustado com os dados limnológicos da Lagoa da Pampulha. Nota: R^2 se refere à calibração e RMSE à validação

Fonte: Próprio Autor (2022)

No caso do Landsat-8, observa-se maior valor da moda do R^2 para o modelo empírico, entretanto nota-se comportamento contrário quando comparado às modas do RMSE. Para o Sentinel-2, tanto a distribuição quanto a moda demonstram superação de desempenho do modelo empírico em relação ao modelo da literatura ajustado. Por isso, adotou-se os modelos empíricos de ambos os satélites para a modelagem de SDD.

Ao aplicar os modelos obtidos considerando-se toda a amostra de dados de SDD disponíveis para o Landsat-8 ($n=23$), verifica-se um valor de R^2 igual a 0,65, R^2 ajustado igual a 0,55, RMSE igual a 0,1 m e nRMSE igual a 29% (Figura 32a). Para o Sentinel-2, com os dados de SDD disponíveis ($n=33$), obtêm-se R^2 igual a 0,88, R^2 ajustado igual a 0,86, RMSE igual a 0,1 m e nRMSE igual a 15% (Figura 32b).

Figura 32 – Comparação entre valores estimados pelos modelos escolhidos e valores observados de SDD para ambos os satélites



Legenda: (a) Landsat-8, (b) Sentinel-2

Fonte: Próprio Autor (2022)

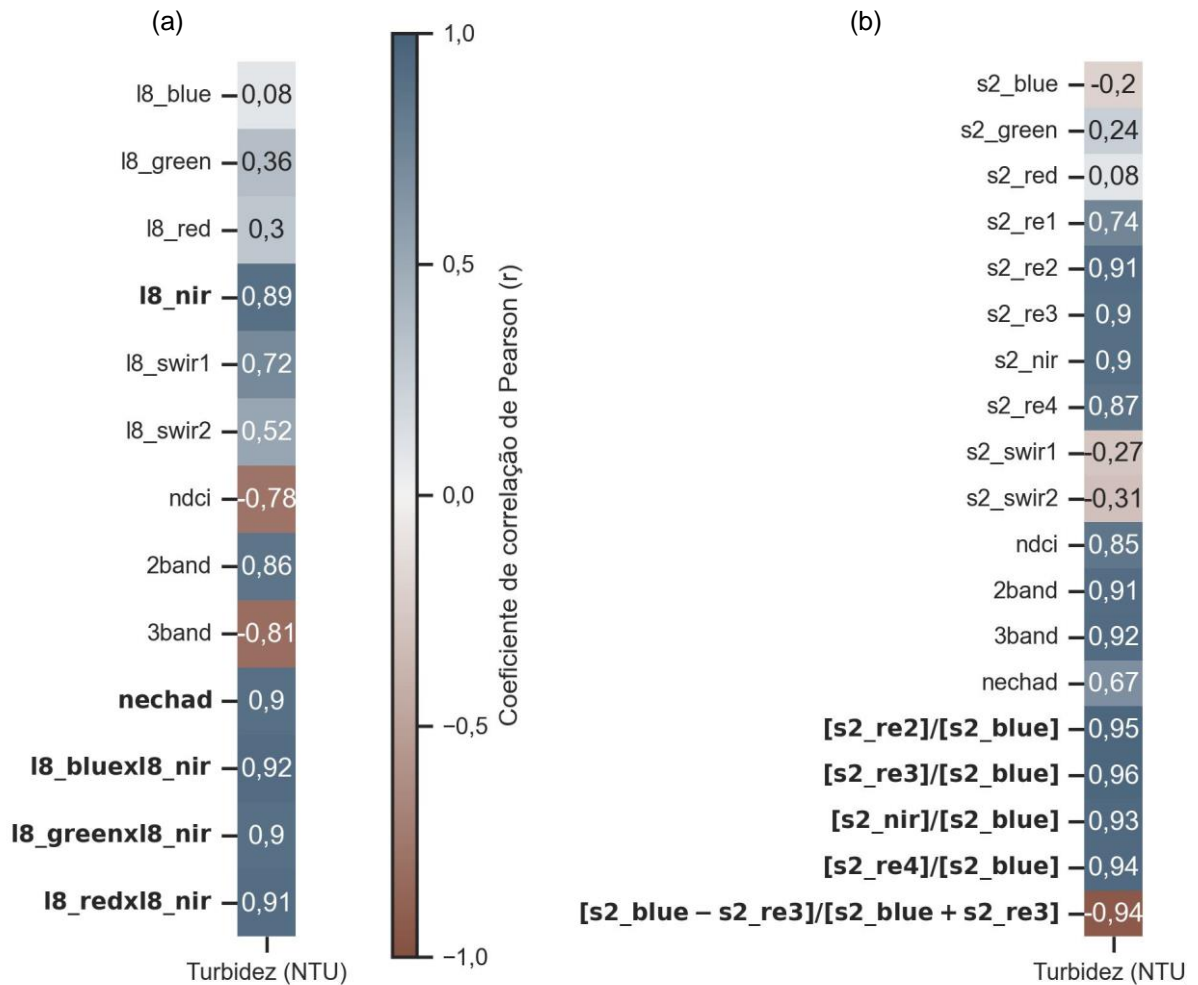
Para os modelos apresentados, nota-se superestimativa para valores baixos de SDD e a inversão para a subestimativa para valores maiores de SDD. No caso do Sentinel-2, como os dados apresentam maior amplitude, o RMSE de 0,1 m, equivalente ao RMSE obtido para o Landsat-8 apresenta menor erro normalizado, indicando melhor desempenho.

5.2.4 Turbidez (NTU)

A correlação do parâmetro Turbidez com dados do Landsat-8, apresentada na Figura 33a, mostra-se promissora para a utilização da banda do infravermelho (I8_nir), equivalente a $r = 0,65$. Além disso, essa banda, utilizada na variável independente de Nechad, ou na multiplicação entre bandas, por exemplo, do azul (I8_blue) com NIR (I8_nir), do verde (I8_green) com NIR (I8_nir) e do vermelho (I8_red) com NIR (I8_nir), demonstra forte correlação positiva com a Turbidez ($r > 0,9$).

Correlações do parâmetro Turbidez com dados do Sentinel-2, apresentadas na Figura 33b, mostram que a banda RedEdge-2 (s2_re2), com $r = 0,91$, foi a que apresentou correlação mais forte dentre as bandas do Sentinel-2. Correlações fortes ($r > 0,91$ e $r < -0,91$) são observadas para razões entre bandas e índices que utilizaram as bandas do azul (s2_blue), do infravermelho (s2_nir) do RedEdge-2 (s2_re2), RedEdge-3 (s2_re3) e RedEdge-4 (s2_re4).

Figura 33 – Correlação de *Pearson* para as principais variáveis independentes de entrada para a modelagem do Turbidez para ambos os satélites.

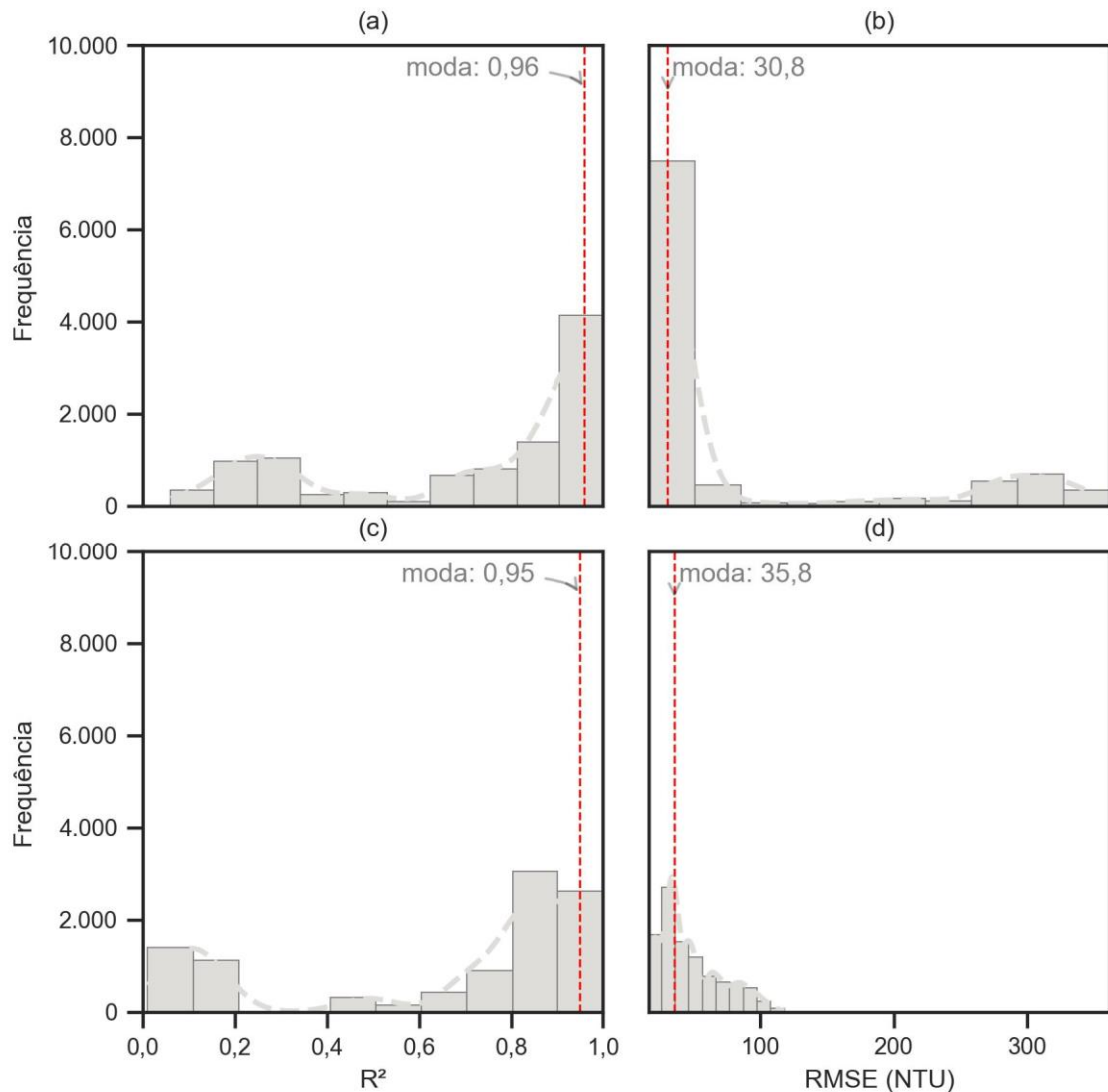


Legenda: (a) Landsat-8, (b) Sentinel-2. Nota: os termos em negrito representam as cinco variáveis independentes mais relacionadas com o parâmetro modelado.

Fonte: Próprio Autor (2022)

Quando utilizado os dados do Landsat-8 para modelar Turbidez, o modelo empírico é aquele que utiliza o método de regressão polinomial utilizando a variável independente de multiplicação entre bandas do azul e do NIR ($l8_blue \times l8_nir$). A moda do R^2 para a calibração é igual a 0,96 e a moda do RMSE para validação é igual a 30,8 NTU (Figura 34a). O modelo da literatura ajustado é aquele que utiliza a banda do infravermelho ($l8_nir$) para a estimativa da Turbidez por meio de regressão polinomial. As métricas obtidas são $R^2_{moda} = 0,95$ na calibração e $RMSE_{moda} = 35,8$ NTU na validação (Figura 34b).

Figura 34 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R² e RMSE para o parâmetro Turbidez modelado a partir dos dados do Landsat-8

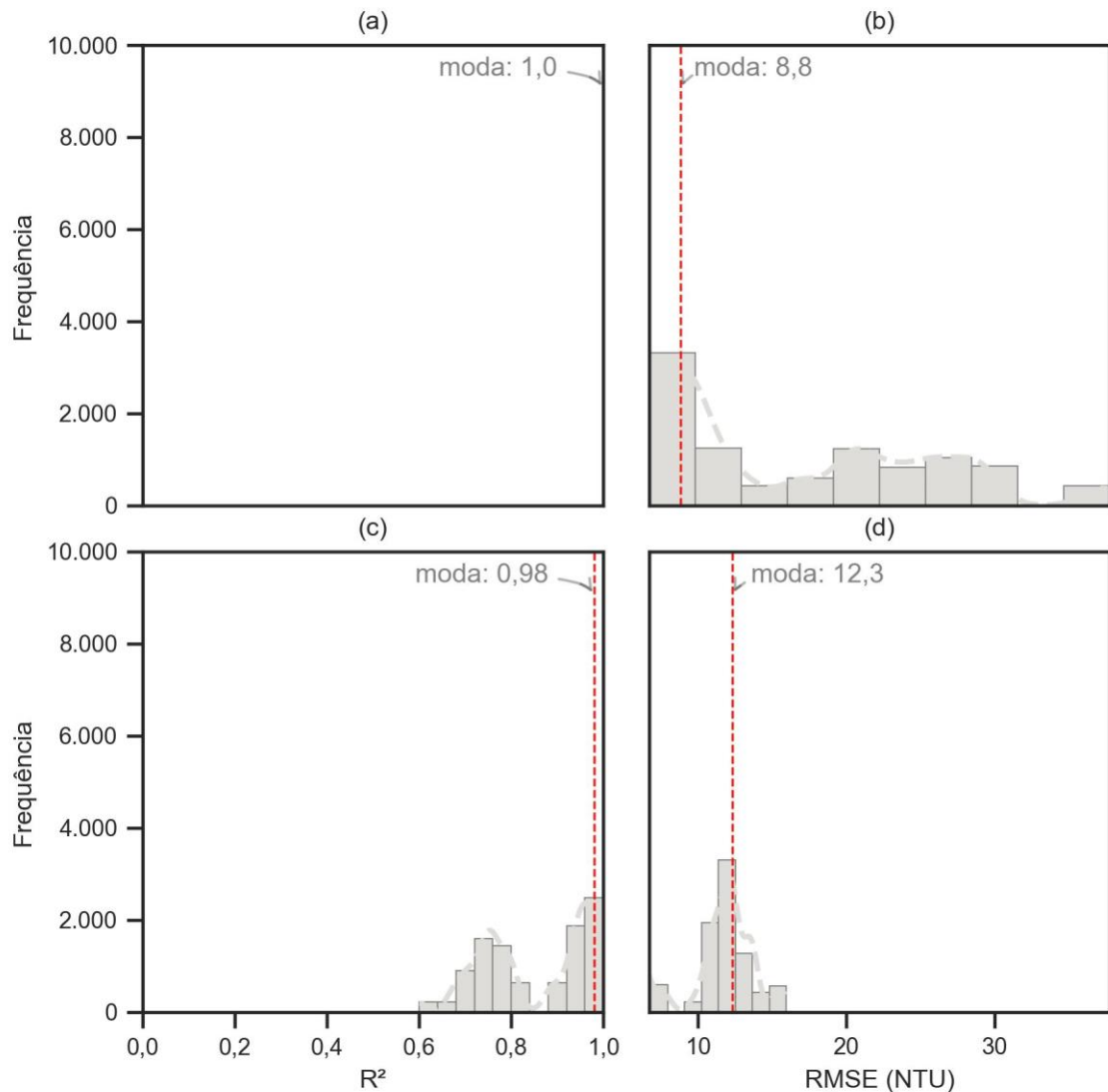


Legenda: (a) e (b) modelo empírico, (c) e (d) modelo da literatura ajustado com os dados limnológicos da Lagoa da Pampulha. Nota: R² se refere à calibração e RMSE à validação

Fonte: Próprio Autor (2022)

Para o Sentinel-2, o modelo empírico para Turbidez é aquele que utiliza o método de regressão linear multivariada com as variáveis independentes de entrada destacadas em negrito na Figura 33b. A moda do R² para a calibração é igual a 1,0 e a moda do RMSE para validação é igual a 8,8 NTU (Figura 34a). O modelo da literatura ajustado é aquele que utiliza banda do infravermelho (s2_nir) para a estimativa da Turbidez por meio de regressão linear. As métricas calculadas são R²_{moda} = 0,98 na calibração e RMSE_{moda} = 12,3 NTU na validação (Figura 35b).

Figura 35 – Histogramas de calibração (n=10000) e validação (n=10000) com as métricas R² e RMSE para o parâmetro Turbidez modelado a partir dos dados do Sentinel-2



Legenda: (a) e (b) modelo empírico, (c) e (d) modelo da literatura ajustado com os dados limnológicos da Lagoa da Pampulha. Nota: R² se refere à calibração e RMSE à validação

Fonte: Próprio Autor (2022)

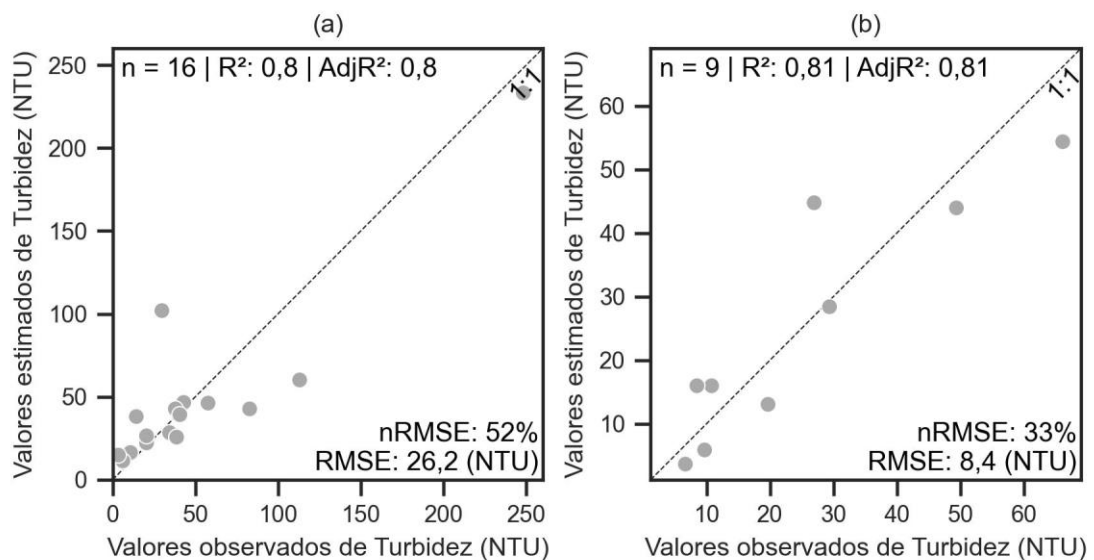
No caso do Landsat-8, observa-se proximidade nas modas do R² e modas do RMSE, entretanto, a distribuição do RMSE de validação para o modelo empírico apresenta erros ultrapassando 200 NTU. Por isso, decidiu-se adotar o modelo da literatura ajustado para a estimativa de Turbidez para o Landsat-8.

Para o Sentinel-2, o modelo empírico, por utilizar regressão linear multivariada com 5 parâmetros se ajusta perfeitamente aos dados sorteados para a calibração em todas as simulações, indicando superparametrização, ao passo que o modelo ajustado da literatura apresenta R² de calibração variando entre 0,6 e 1,0 com moda de 0,98. Para

a validação, verifica-se o melhor desempenho do modelo empírico, que apresenta moda do RMSE igual a 8,8 NTU quando comparado à moda do RMSE de validação do segundo igual a 12,3 NTU. Entretanto, a amplitude dos valores de RMSE do modelo empírico chegam a 300 NTU. Portanto, para o Sentinel-2, também preferiu-se escolher o modelo ajustado da literatura para a continuidade dos trabalhos.

Ao aplicar os modelos obtidos considerando-se toda a amostra de dados de Turbidez disponíveis para o Landsat-8 (n=16), verifica-se valores de R^2 e R^2 ajustado iguais a 0,8, RMSE igual a 26,2 NTU, e nRMSE igual a 52% (Figura 36a). Para o Sentinel-2, com os dados de Turbidez disponíveis (n=9), obtêm-se R^2 e R^2 ajustado iguais a 0,81, RMSE igual a 8,4 NTU e nRMSE de 33% (Figura 36b).

Figura 36 – Comparação entre valores estimados pelos modelos escolhidos e valores observados de Turbidez para ambos os satélites



Legenda: (a) Landsat-8, (b) Sentinel-2

Fonte: Próprio Autor (2022)

Para os modelos adotados, nota-se a proximidade dos pontos da diagonal 1:1 para toda amplitude de valores. Para o modelo aplicado aos dados do Landsat-8, nota-se maior amplitude de valores e maiores erros (RMSE e nRMSE) associados na estimativa de turbidez, resultando em um desempenho inferior comparado ao modelo aplicado aos dados do Sentinel-2, que apresentou valor de nRMSE menor.

5.2.5 Modelos selecionados e suas principais características

Os modelos selecionados apresentados nos itens anteriores foram organizados com suas principais informações e estão apresentados na Tabela 8. Para a coluna de hiperparâmetros ou equações, apresenta-se os hiperparâmetros para os modelos que utilizaram aprendizado de máquina e equações para modelos que utilizaram regressões lineares, exponenciais, logarítmicas ou polinomiais.

Para o prosseguimento do estudo, os modelos selecionados serão utilizados para o mapeamento da qualidade da água de regiões não monitoradas na Lagoa da Pampulha e para datas que ocorreram as passagens dos satélites Landsat-8 e Sentinel-2 e não ocorreram campanhas de monitoramento *in situ*.

Nota-se na Lagoa da Pampulha, para o parâmetro TSM e SDD, baixos valores de RMSE (8 mg/L e 0,10 m, respectivamente) em relação aos valores mínimo e máximo observados no reservatório, 2,37 mg/L a 281 mg/L e 0 a 2,3 m (Tabela 5), respectivamente.

Para o parâmetro Chla, ambos os satélites apresentam desempenho razoável, com valores do nRMSE chegando a aproximadamente 100%. Entretanto, destaca-se que este parâmetro apresenta desafios em sua modelagem, visto que é mais sensível a variações e à influência das espécies fitoplanctônicas presentes no meio. A Turbidez apresenta desempenho satisfatório para o Sentinel-2, entretanto para o Landsat-8 o desempenho pode ser considerado razoável, com o RMSE chegando a 26,2 NTU que pode representar a mudança de enquadramento do corpo d'água em situações de avaliação das legislações vigentes de gestão dos recursos hídricos.

Tabela 8 – Modelos selecionados com suas principais características descritas para cada parâmetro e para cada satélite

Variável dependente (y)	Modelo	Tipo de regressão	Hiperparâmetros ou Equação	Satélite	Variável independente (x)	Calibração e Validação		Modelo selecionado aplicado em todo conjunto de dados da variável dependente			
						R ²	RMSE	R ²	R ² Aj	RMSE	nRMSE
						MODA	MODA				
TSM (mg/L)	Modelo empírico	Random Forest	critério: mínimos quadrados, prof_max: 10, n_est: 10	L8	l8_nir, [l8_nir/l8_red], [l8_nir/l8_blue], [l8_nir/l8_green], [l8_swir1/l8_swir2]	0,89	15,9 mg/L	0,88	0,86	8,1 mg/L	29%
TSM (mg/L)	Modelo empírico	Random Forest	critério: mínimos quadrados, prof_max: 2, n_est: 100	S2	Antes da rede, Depois da rede, [s2_re2 / s2_swir1], [s2_re3 / s2_swir1], [s2_nir / s2_swir1]	0,73	10,5 mg/L	0,72	0,69	8,5 mg/L	42%
Chla (µg/L)	Modelo ajustado literatura	Regressão polinomial	$y = 289,2x^2 + 3,1$	L8	[l8_nir/l8_red]	0,79	57,5 µg/L	0,77	0,77	91,8 µg/L	95%
Chla (µg/L)	Modelo ajustado literatura	Regressão linear	$y = 134,6x - 74,7$	S2	[s2_re1/s2_red]	0,39	72,5 µg/L	0,38	0,38	51,6 µg/L	92%
SDD (m)	Modelo empírico	Random Forest	critério: mínimos quadrados, prof_max: 2, n_est: 10	L8	[l8_blue / l8_nir], [l8_green / l8_nir], [l8_nir / l8_blue], [(l8_blue - l8_nir) / (l8_blue + l8_nir)], [(l8_green - l8_nir) / (l8_green + l8_nir)]	0,62	0,19 m	0,65	0,55	0,09 m	29%
SDD (m)	Modelo empírico	Random Forest	critério: mínimos quadrados, prof_max: 2, n_est: 100	S2	[s2_green / s2_re1], [s2_swir2 / s2_re2], [s2_swir2 / s2_re3],	0,89	0,09 m	0,88	0,86	0,06 m	15%

Variável dependente (y)	Modelo	Tipo de regressão	Hiperparâmetros ou Equação	Satélite	Variável independente (x)	Calibração e Validação		Modelo selecionado aplicado em todo conjunto de dados da variável dependente			
						R ² MODA	RMSE MODA	R ²	R ² Aj	RMSE	nRMSE
					[[s2_re2 - s2_swir2] / [s2_re2 + s2_swir2]], [[s2_re3 - s2_swir2] / [s2_re3 + s2_swir2]]						
Turbidez (NTU)	Modelo ajustado literatura	Regressão linear	$y = 1683,2x - 4,7$	L8	l8_nir	0,95	35,8 NTU	0,8	0,8	26,2 NTU	52%
Turbidez (NTU)	Modelo ajustado literatura	Regressão linear	$y = 2043,6x - 29,7$	S2	s2_nir	0,98	12,3 NTU	0,81	0,81	8,4 NTU	33%

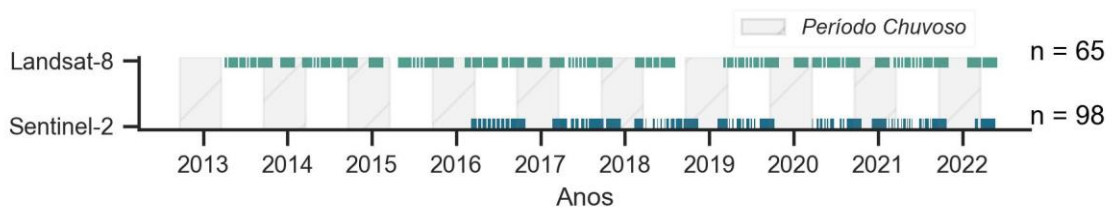
Nota: L8 – Landsat-8, S2 – Sentinel-2

Fonte: Próprio Autor (2023)

5.3 Aplicação dos modelos para todo o reservatório e para imagens históricas

No período de 01/01/2013 a 01/05/2022, para a Lagoa da Pampulha, foram obtidas 195 imagens do Landsat-8 e 389 imagens do Sentinel-2 (A e B), sendo que apenas 65 imagens (33%) do Landsat-8 e 98 (25%) do Sentinel-2 não apresentaram interferência de nuvens, sombra de nuvens, ou reflexão especular sobre o espelho d'água. Na Figura 37, observa-se que a maior parte das imagens se concentra no período seco ($n_{L8}=50$, $n_{S2}=90$) e poucas imagens correspondem ao período chuvoso ($n_{L8}=15$, $n_{S2}=8$).

Figura 37 – Imagens de satélites utilizadas para a avaliação da dinâmica espacial e temporal da qualidade da água na Lagoa da Pampulha



Fonte: Próprio Autor (2022)

Apesar das incertezas inerentes aos modelos propostos para estimativa dos parâmetros de qualidade da água pelo sensoriamento remoto, o presente estudo buscou a avaliação de fatores, padrões e tendências na qualidade da água para a Lagoa da Pampulha. Para isso, aplicaram-se os modelos escolhidos na Subseção 5.2 para extrair os parâmetros de qualidade de água a partir das reflectâncias registradas nas imagens selecionadas. Como resultado, obteve-se matrizes multidimensionais com dimensões tempo=65, y=71 e x=95, para o Landsat-8 e, dimensões tempo=98, y=210 e x=284, para o Sentinel-2. O tempo representa a data e hora da imagem, o y representa a coordenada vertical e x a coordenada horizontal.

Com a junção dos dados de qualidade de água modelados de ambos os satélites por meio da reamostragem dos dados do Sentinel-2 para a resolução de 30m, duas matrizes multidimensionais transformaram-se em uma com dimensões: tempo=155, y=71 e x=95. A quantidade de imagens (tempo=155) para o período (2013-2022) não corresponde ao somatório das imagens selecionadas para o Landsat-8 e Sentinel-2 devido à ocorrência da passagem simultânea (mesmo dia) para ambos os satélites em 8 ocasiões. A correlação entre valores dos parâmetros de qualidade de água modelados para imagens do mesmo dia para ambos os satélites foi calculada (n=8) e

os resultados mostram que o parâmetro Chla foi o que teve maior correlação média ($\rho=0,81$), ao passo que a Turbidez apresentou menor correlação ($\rho = 0,48$). O APÊNDICE E apresenta os valores de correlação obtidos para cada parâmetro e entre imagens.

Para essas datas extraiu-se o valor médio do parâmetro de qualidade de água modelado a partir de dados de ambos os satélites. A partir deste momento, os resultados por parâmetro de qualidade de água serão apresentados sem distinção do modelo ou do satélite utilizado.

5.3.1 Estatística descritiva da dinâmica espaço-temporal da qualidade da água

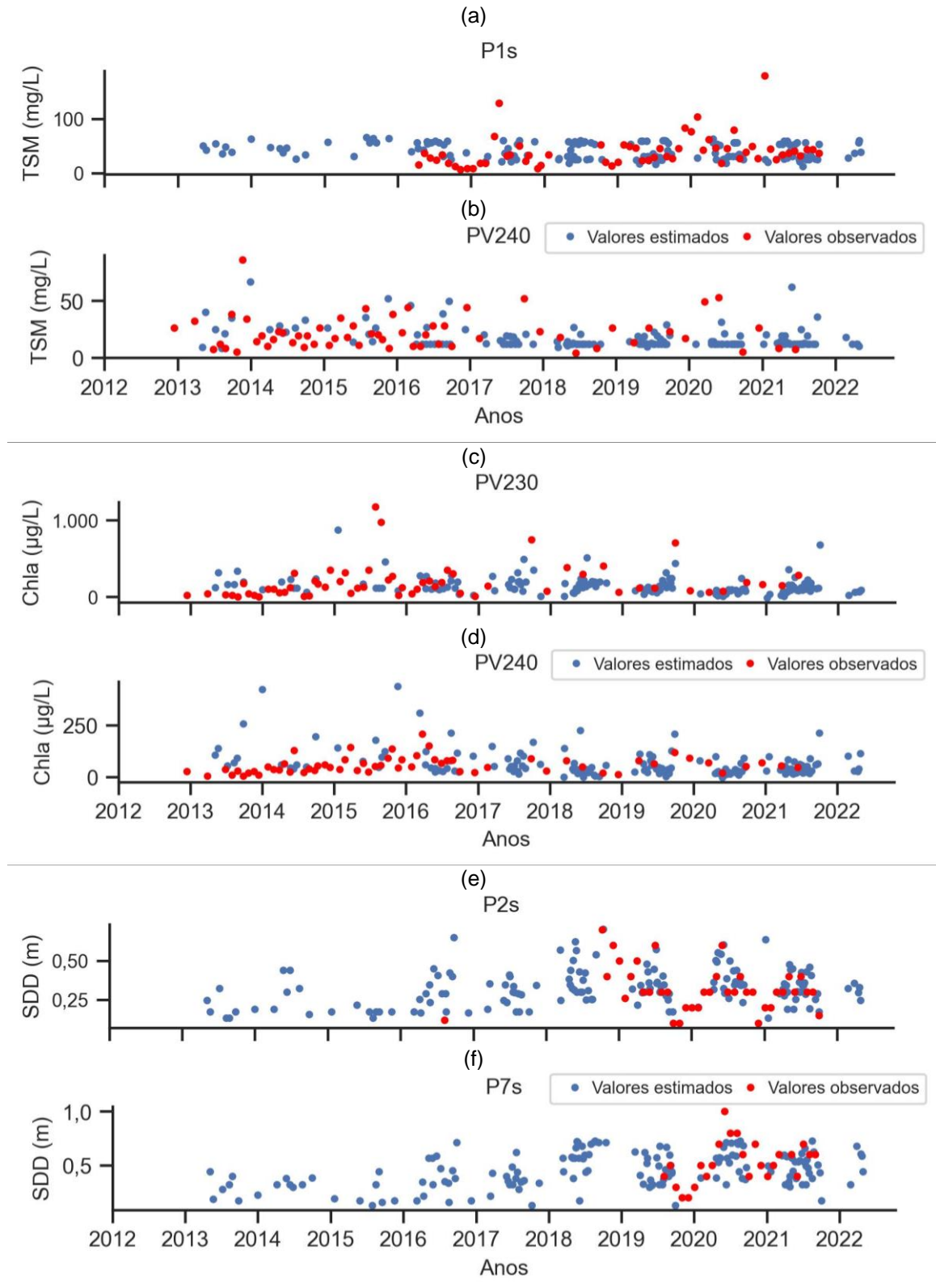
Para os pontos de monitoramento limnológico utilizados pela PBH e IGAM foram extraídas as séries temporais dos parâmetros modelados para a avaliação de seu comportamento ao longo do tempo e para comparação com os valores observados. Devido à existência de dez pontos de monitoramento e quatro parâmetros de qualidade da água, preferiu-se apresentar exemplos das séries temporais de pontos localizados a montante e a jusante para cada parâmetro (Figura 38). As séries temporais dos demais pontos de monitoramento podem ser visualizadas no APÊNDICE F.

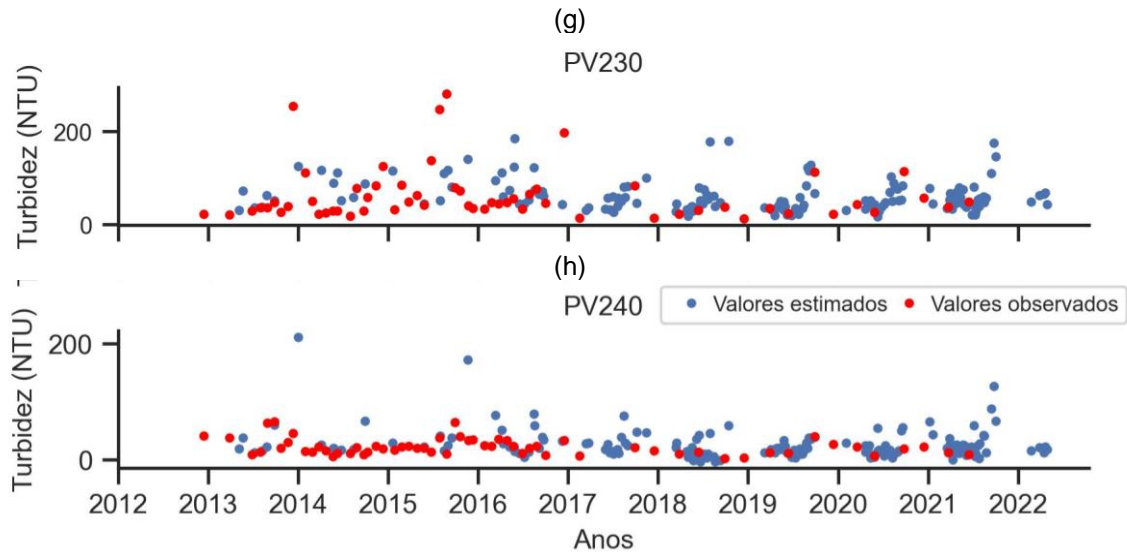
Os resultados apresentados na Figura 38 permitem notar uma maior amplitude de variação dos parâmetros de qualidade da água para os pontos localizados à montante (Figura 38 a, c, e, g) quando comparada com os pontos localizados mais à jusante do reservatório (Figura 38 b, d, f, h). Em todos os pontos verifica-se comportamento sazonal, com a ocorrência de valores máximos em épocas entre outubro e março, que correspondem ao período chuvoso para a região e redução dos valores durante a transição para o período seco (abril-setembro) para os parâmetros TSM, Chla e Turbidez. Para o SDD também se nota o comportamento sazonal, entretanto contrário aos parâmetros supramencionados, haja vista a natureza de valores maiores representarem maior transparência da água.

Ao comparar os valores estimados com os valores observados, nota-se a correspondência dos pontos ao longo do período analisado, com a mesma ordem de grandeza e comportamento sazonal bem representados. Desvios, como valores

estimados de Chla superiores aos valores observados no ponto de montante (PV230) em 2014 (Figura 38c) e valores estimados de TSM abaixo dos valores observados no ponto de montante P1s em 2021 (Figura 38a), podem estar associados: (i) à alta variabilidade da qualidade de água para a região devido à influência imediata dos afluentes ao reservatório, (ii) à dificuldade dos modelos em se ajustar aos dados *in situ* devido a fatores intervenientes existentes no processo de registro e transformação do valor de reflectância da água no valor correspondente ao parâmetro de qualidade de água.

Figura 38 – Série temporal para os parâmetros de qualidade de água (observados e modelados) nos pontos de monitoramento limnológico



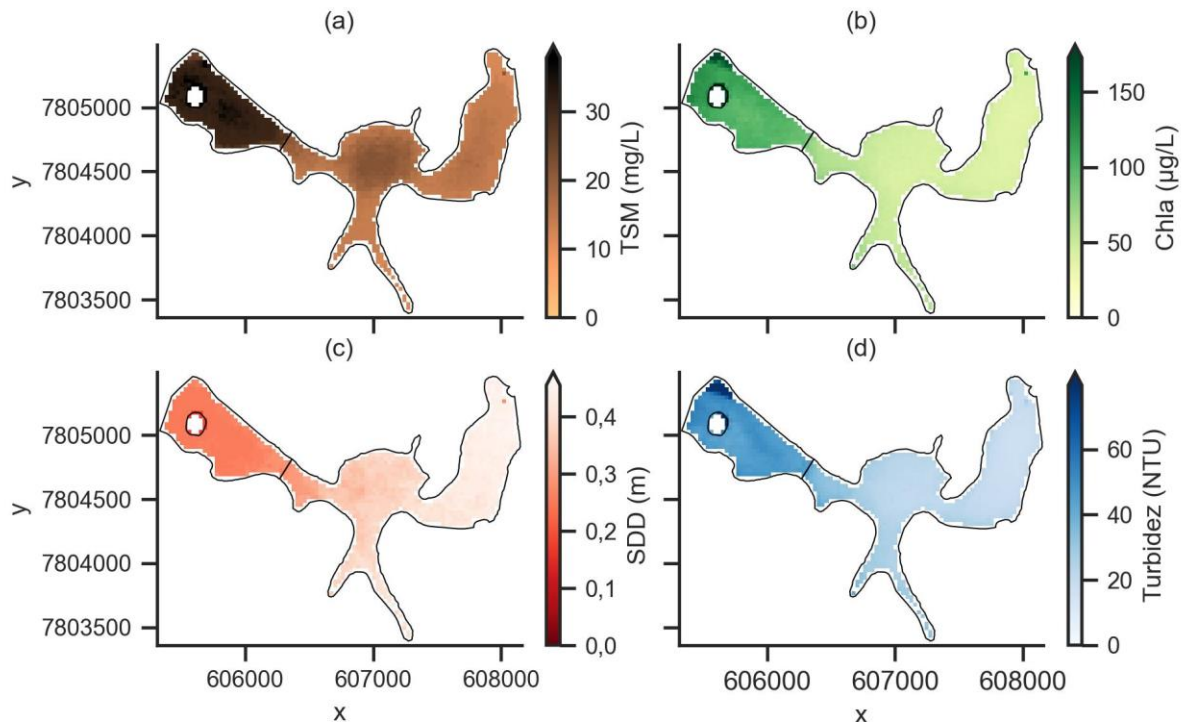


Nota: a, c, e, g – pontos de montante, b, d, f, h – pontos de jusante (em relação a cortina de retenção)

Fonte: Próprio Autor (2022)

O cálculo da mediana para cada *pixel* para o período entre 2013 e 2022 demonstra divisão clara da concentração dos parâmetros de qualidade de água para regiões de montante e jusante à cortina de retenção. Os parâmetros TSM (Figura 39a), Chla (Figura 39b) e Turbidez (Figura 39d) apresentam maiores valores de concentrações medianas nas proximidades da Ilha dos Amores, que corresponde à região de chegada dos principais afluentes, Ressaca e Sarandi. Para o SDD, observa-se a diminuição da transparência da água à medida que se desloca de jusante (SDD ~ 0,6m) para montante (SDD ~ 0,2 m) com valores intermediários na região central do reservatório (SDD ~ 0,3 m) (Figura 39c).

Figura 39 – Mediana dos parâmetros de qualidade de água calculada por *pixel* para o período (2013 – 2022)



Legenda: (a) TSM (mg/L), (b) Chla (µg/L), (c) SDD (m), (d) Turbidez (NTU)

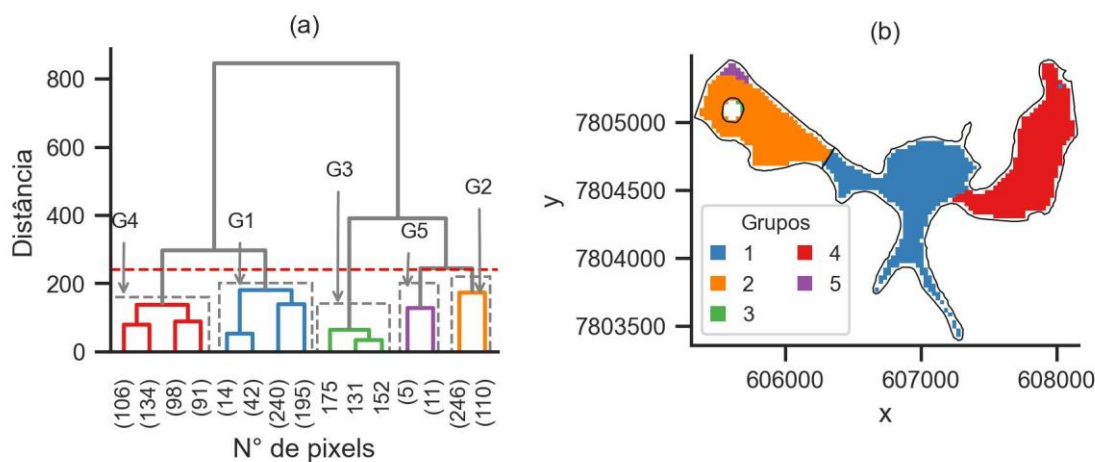
Fonte: Próprio Autor (2022)

5.3.2 Análise multivariada para determinação de regiões homogêneas e heterogêneas no reservatório

Para todos os valores modelados dos parâmetros de qualidade de água entre (2013-2022), que compõem a matriz multidimensional ($t=155$, $y=71$, $x=95$), a análise de agrupamento hierárquico com cálculo de distância Euclidiana e método de aglomeração de *Ward* apresentou distâncias acumuladas variando entre 0 e 800. Adotou-se valor de distância igual a 240 o que refletiu na separação da Lagoa da Pampulha em cinco grupos de *pixels* (valores modelados de qualidade da água), conforme apresentado na Figura 40b.

A divisão de grupos demonstra que há bem definido um grupo homogêneo a montante da cortina de retenção com maioria dos *pixels* da região ($n_{\text{pixels}} = 356$), denominado “G2”, e outros dois grupos, um no entorno da Ilha dos Amores (“G3”, $n_{\text{pixels}}=3$) e um na afluência do Ressaca/Sarandi (“G5”, $n_{\text{pixels}} = 11$). Os outros dois grupos (“G1”; $n_{\text{pixels}}=491$ e “G4”; $n_{\text{pixels}}=429$) estão a jusante da cortina de retenção e são divididos por região de estreitamento localizado no centro-leste do reservatório (Figura 40b).

Figura 40 – Agrupamento hierárquico para os parâmetros de qualidade de água modelados para a Lagoa da Pampulha



Legenda: (a) Dendrograma com os grupos definidos e linha de corte, (b) Grupos representados na dimensão espacial. Nota: Valores com “()” indicam grupos de pixels e as cores dos gráficos são correspondentes.

Fonte: Próprio Autor (2022)

Visando à associação dos grupos com a localização desses, nomeou-se os grupos da seguinte forma: G1: região central, G2: montante, G3: entorno Ilha dos Amores, G4: jusante e G5: afluência Ressaca/Sarandi.

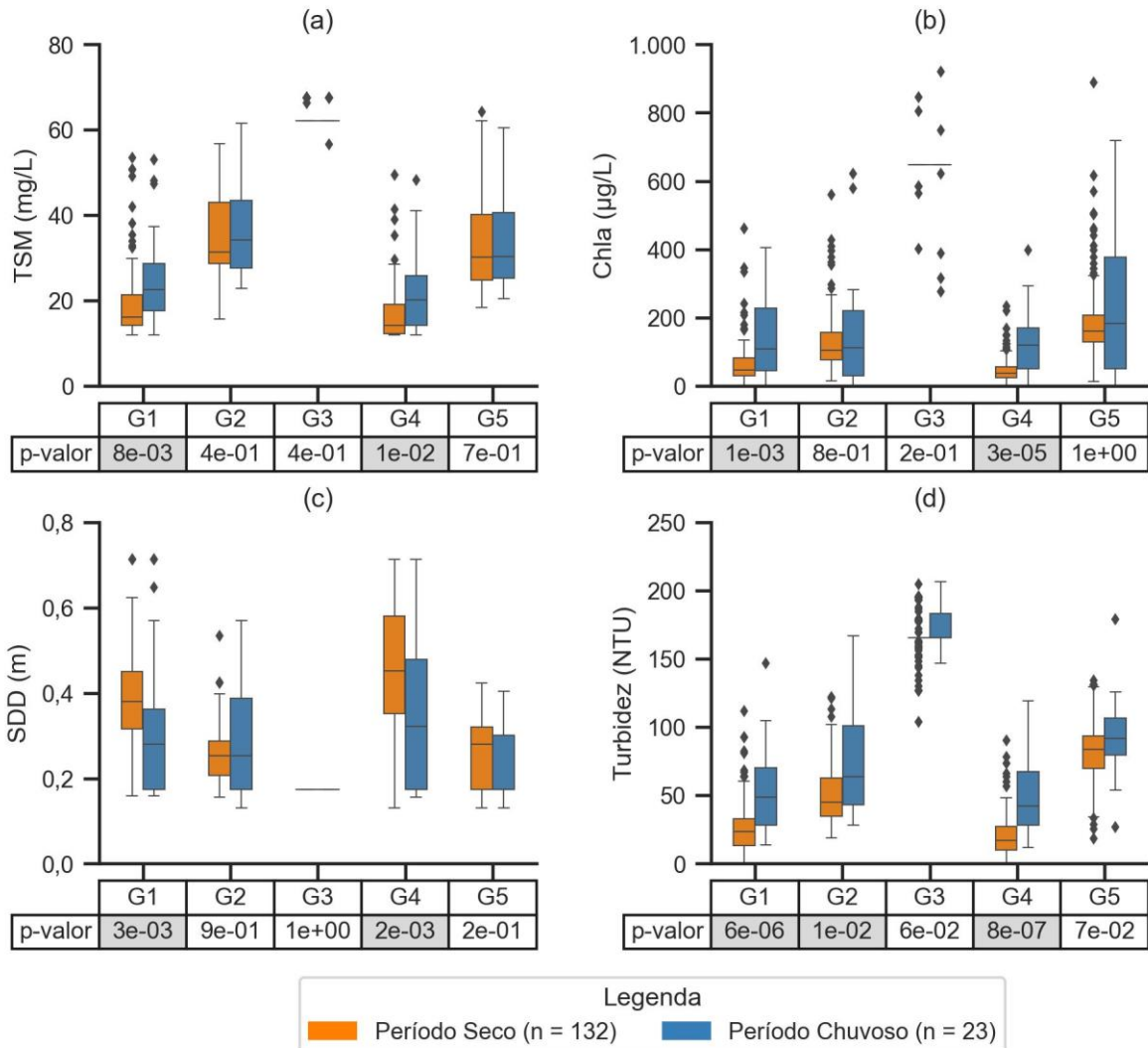
5.3.3 Testes de hipótese para a avaliação da influência de variáveis categóricas externas na qualidade da água

A redução de dimensionalidade por meio do cálculo da mediana dos parâmetros de qualidade da água para cada grupo, definido na Seção 5.3.2 e apresentados na Figura 40, transformou a matriz multidimensional de dimensões $t=155$, $y=71$ e $x=95$ para dimensões $t=155$, $gh=5$, onde gh representa o número de grupos definidos.

Em relação ao período chuvoso e período seco, das 155 imagens, 132 imagens representam o período seco e 23 imagens representam o período chuvoso. Por meio dos p-valores obtidos por teste de *Mann-Whitney* e apresentados nas tabelas da Figura 41, observa-se que o G1 (região central) e G4 (jusante) apresentam rejeição da hipótese nula (p -valor $< 1\%$) para todos os parâmetros de qualidade de água, indicando a diferença entre as amostras referentes ao período seco e período chuvoso, o que coincide com os *boxplots* apresentados (Figura 41a-d). No caso da Turbidez, o teste realizado para G2 (montante) também rejeita a hipótese nula. Os

outros grupos não apresentam distribuição distinta entre período chuvoso e período seco.

Figura 41 – Teste de *Mann-Whitney* para os valores medianos de qualidade de água comparando as distribuições de período chuvoso e período seco para cada grupo identificado na análise multivariada



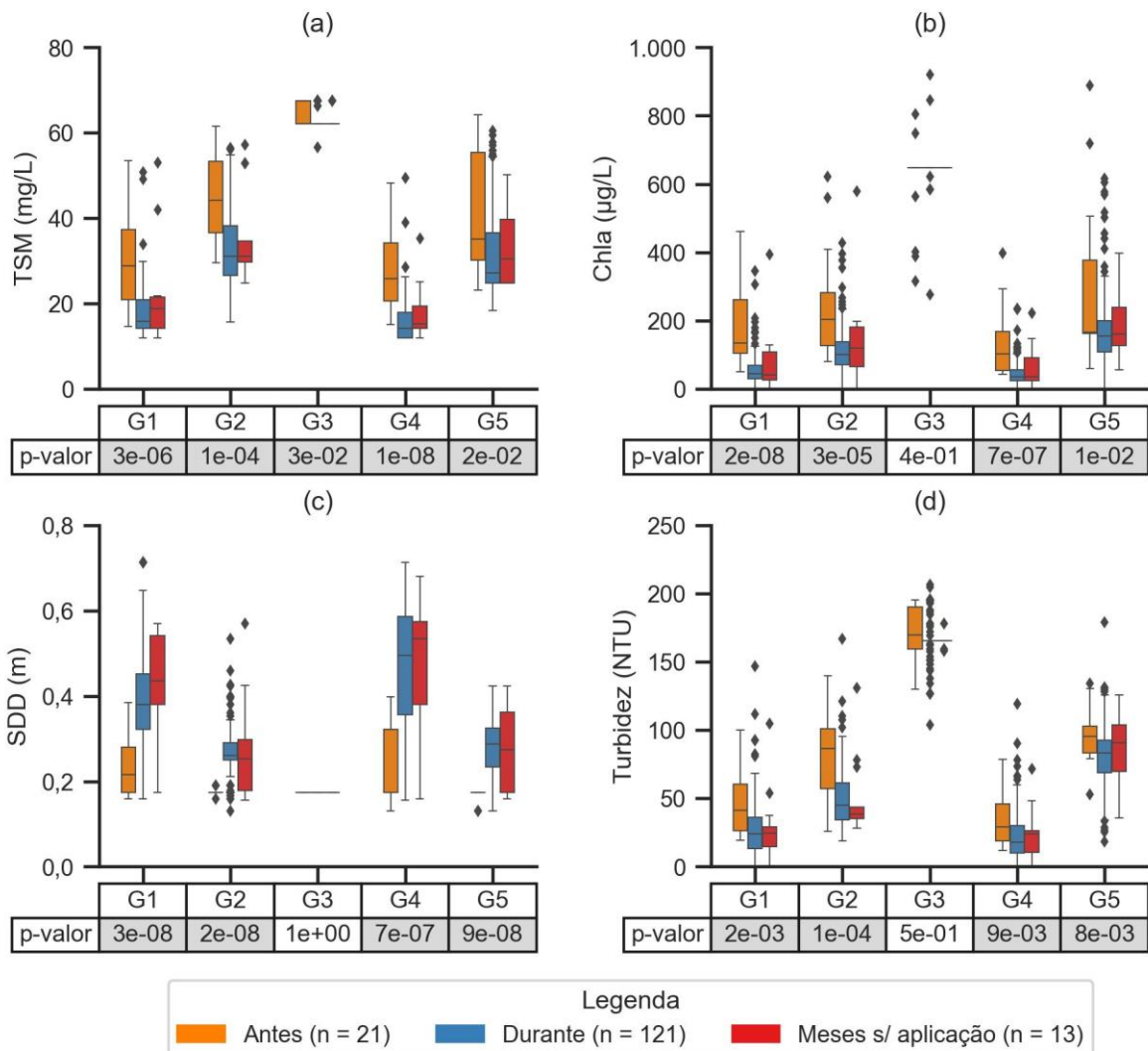
Legenda: Eixo y: parâmetros; Eixo x: grupos definidos para o reservatório com a indicação da significância do teste *Mann-Whitney* que comparou período chuvoso e período seco. Nota: p-valor destacados em cinza apresentam valores inferiores à 5%.

Fonte: Próprio Autor (2023)

Em relação à aplicação de produto químico para o tratamento da Lagoa da Pampulha, têm-se as classes: anterior à aplicação, durante a aplicação e meses sem aplicação. Por meio da Figura 42a-d nota-se que os parâmetros ópticos modelados apresentavam condições piores nos períodos anteriores à aplicação do tratamento ($n_{imagens} = 21$), com a redução dos valores modelados de TSM, Chla e Turbidez durante

a aplicação ($n_{\text{imagens}} = 121$). Para meses com paralisações da aplicação ($n_{\text{imagens}} = 13$) a distribuição volta a apresentar maior variância. Isso se traduz nos testes de *Kruskal-Wallis* realizados para todos os grupos, que apresentam (p -valor $< 2\%$) para as distribuições dos quatro parâmetros de qualidade de água, exceto para o G3 (entorno Ilha dos Amores) que apresenta p -valor $< 5\%$ apenas para o TSM.

Figura 42 – Teste de *Kruskal-Wallis* para a comparação das distribuições das classes de aplicação de produto químico para cada grupo identificado na análise multivariada



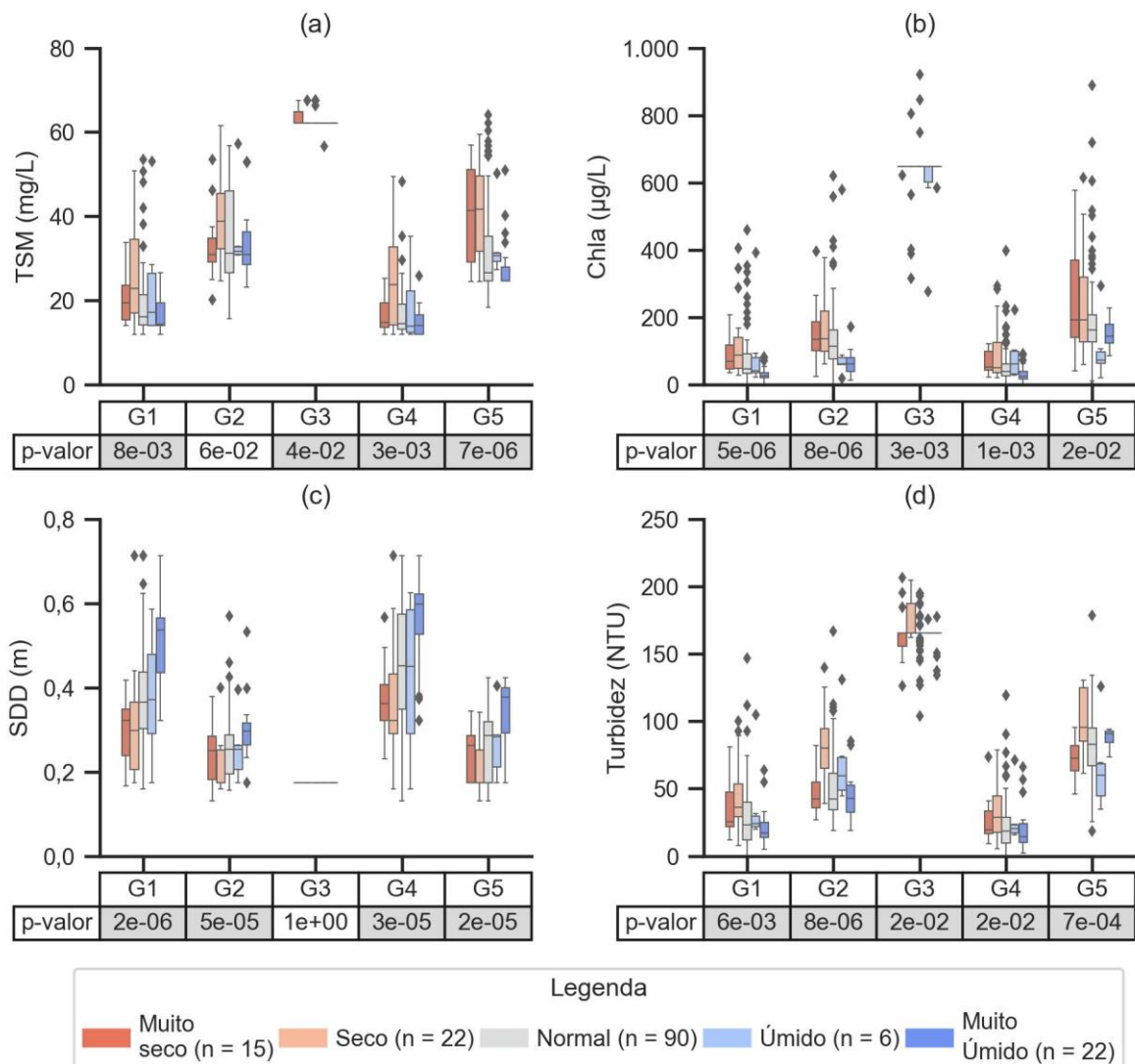
Legenda: Eixo y: parâmetros; Eixo x: grupos definidos para o reservatório com a indicação da significância do teste *Kruskal-Wallis* que comparou a aplicação de produto químico para o tratamento da Lagoa da Pampulha. Nota: p-valor destacados em cinza apresentam valores inferiores à 5%.

Fonte: Próprio Autor (2023)

Para as classes do índice SPI_{12} , observam-se 15 imagens para a classe “Muito seco”, 22 imagens para a classe “Seco”, 90 imagens para a classe “Normal”, 6 imagens para a classe “Úmido”, e 22 imagens para a classe “Muito Úmido”, totalizando as 155

imagens. Com exceção do G2 (montante) para TSM e G3 (entorno Ilha dos Amores) para SDD, todos os outros grupos apresentam rejeição da hipótese nula, pelo teste de *Kruskal-Wallis*, indicando diferença nas distribuições das classes do SPI₁₂ para os parâmetros de qualidade de água. No entanto, ao observar os *boxplots* (Figura 43a-d), nota-se que maiores valores dos parâmetros modelados coincidentes com períodos chuvosos (conforme apresentado na análise de período chuvoso e seco na Figura 41a-d), não se materializa para o índice SPI₁₂ calculado para o período de 12 meses.

Figura 43 – Teste de *Kruskal-Wallis* para a comparação das distribuições das classes do índice SPI₁₂ para cada grupo identificado na análise multivariada



Legenda: Eixo y: parâmetros; Eixo x: grupos definidos para o reservatório com a indicação da significância do teste *Kruskal-Wallis* que comparou as classes do índice SPI₁₂. Nota: p-valor destacados em cinza apresentam valores inferiores à 5%.

Fonte: Próprio Autor (2023)

5.3.4 Avaliação da tendência da qualidade da água do reservatório

O cálculo de medianas para o período chuvoso e período seco para o período estudado (2013-2022) permitiu a redução de dimensionalidade da matriz multidimensional que passou a apresentar dimensões $t=19$, $y=71$, $x=95$.

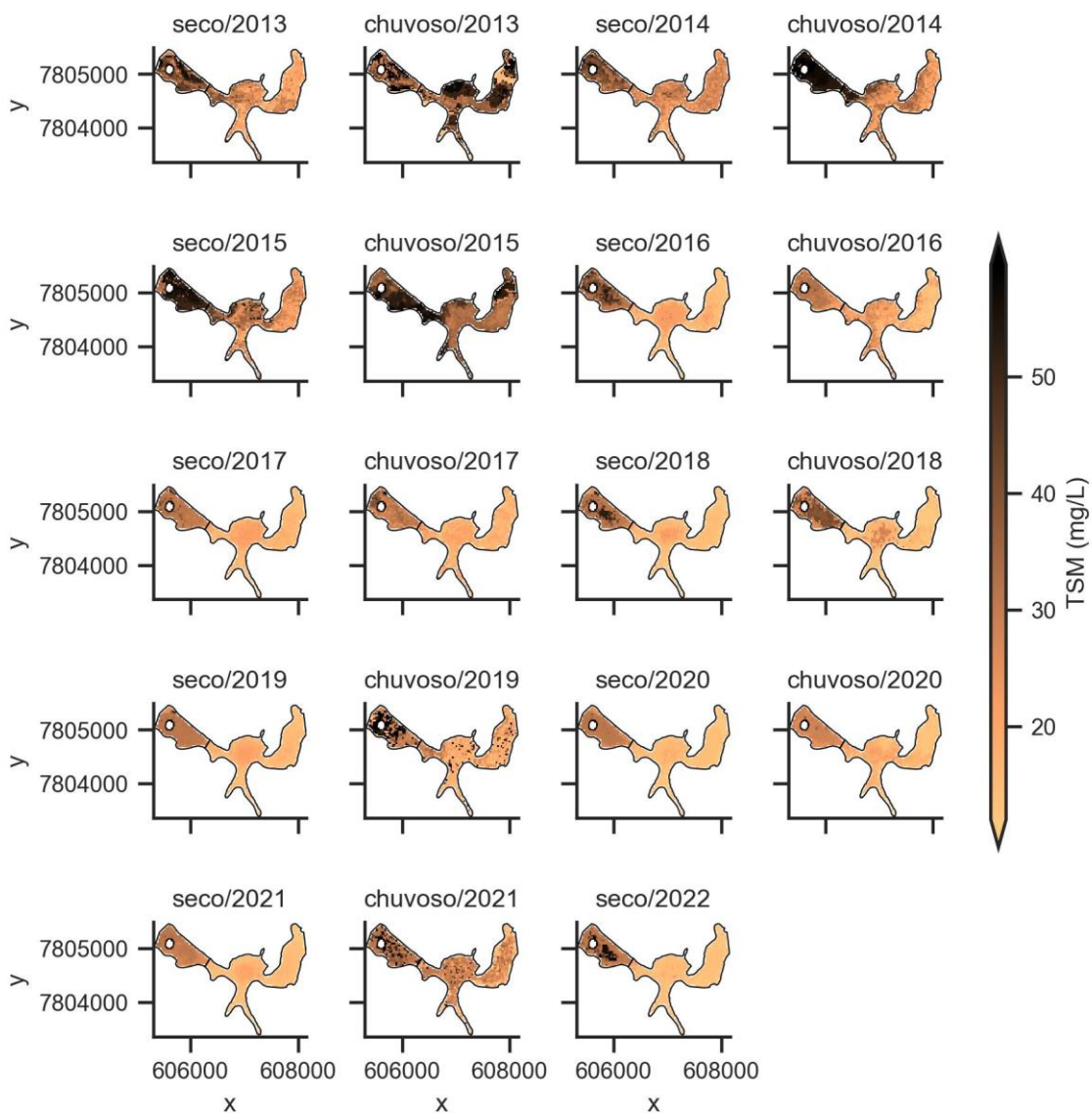
Para o período estudado (2013-2022) nota-se maiores concentrações de TSM durante os primeiros anos de concentrações modeladas. Períodos chuvosos do ano de 2013, 2014 e 2015 se destacam com predominância de maiores concentrações de TSM em todo reservatório (Figura 44). Para o período seco de 2015 verifica-se concentrações na mesma ordem de grandeza do período chuvoso de 2014 e período chuvoso de 2015, indicando possível correlação com o período anterior e subsequente. Para todos os períodos constata-se distinção das regiões de montante e jusante em relação à cortina de retenção, com concentrações mais elevadas para a região de montante (Figura 44).

A Chla apresenta diminuição das concentrações ao longo do período estudado (2013-2022) (Figura 45). O período seco de 2013, chuvoso de 2014 e seco de 2015 demonstram elevadas concentrações de Chla provenientes do principal afluente do reservatório, o Ressaca/Sarandi. O período seguinte, chuvoso/2015 demonstra o avanço de maiores concentrações de Chla para as regiões centrais e de jusante do reservatório (Figura 45).

Haja vista a natureza inversa do SDD, observam-se menores transparências do reservatório para os períodos iniciais, incluindo valores baixos para os períodos secos de 2014 e 2015 e período chuvoso de 2014. Para o restante dos períodos observa-se maior amplitude de valores medidos para as regiões de jusante quando comparadas com regiões de montante (Figura 46).

Entre 2013 e 2015 observam-se maiores concentrações de Turbidez, principalmente para regiões a montante da cortina de retenção. Este comportamento está diretamente relacionado ao comportamento de TSM e é o contrário do observado para SDD. De forma geral, a Turbidez diminui ao longo do período estudado (2013-2022) apresentando maior sensibilidade à sazonalidade que os demais parâmetros (Figura 47).

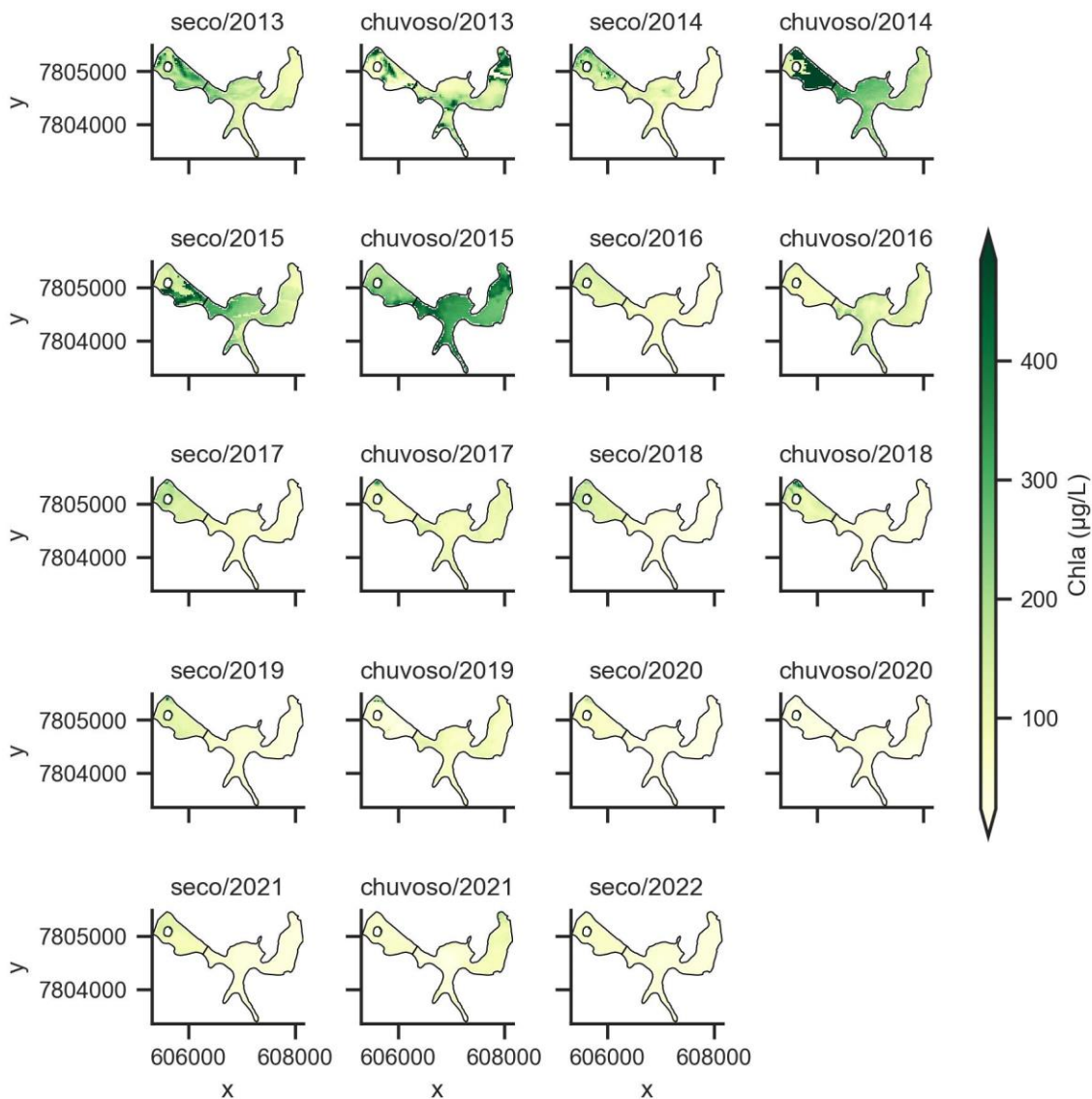
Figura 44 – Valores medianos de TSM (mg/L) calculado por *pixel* considerando período chuvoso e período seco para cada ano durante o período estudado (2013-2022)



Nota: O período seco refere-se aos meses de abril a setembro para o ano apresentado no título de cada gráfico. O período chuvoso refere-se aos meses de outubro a dezembro para o ano apresentado no título e janeiro a março do ano seguinte.

Fonte: Próprio Autor (2023)

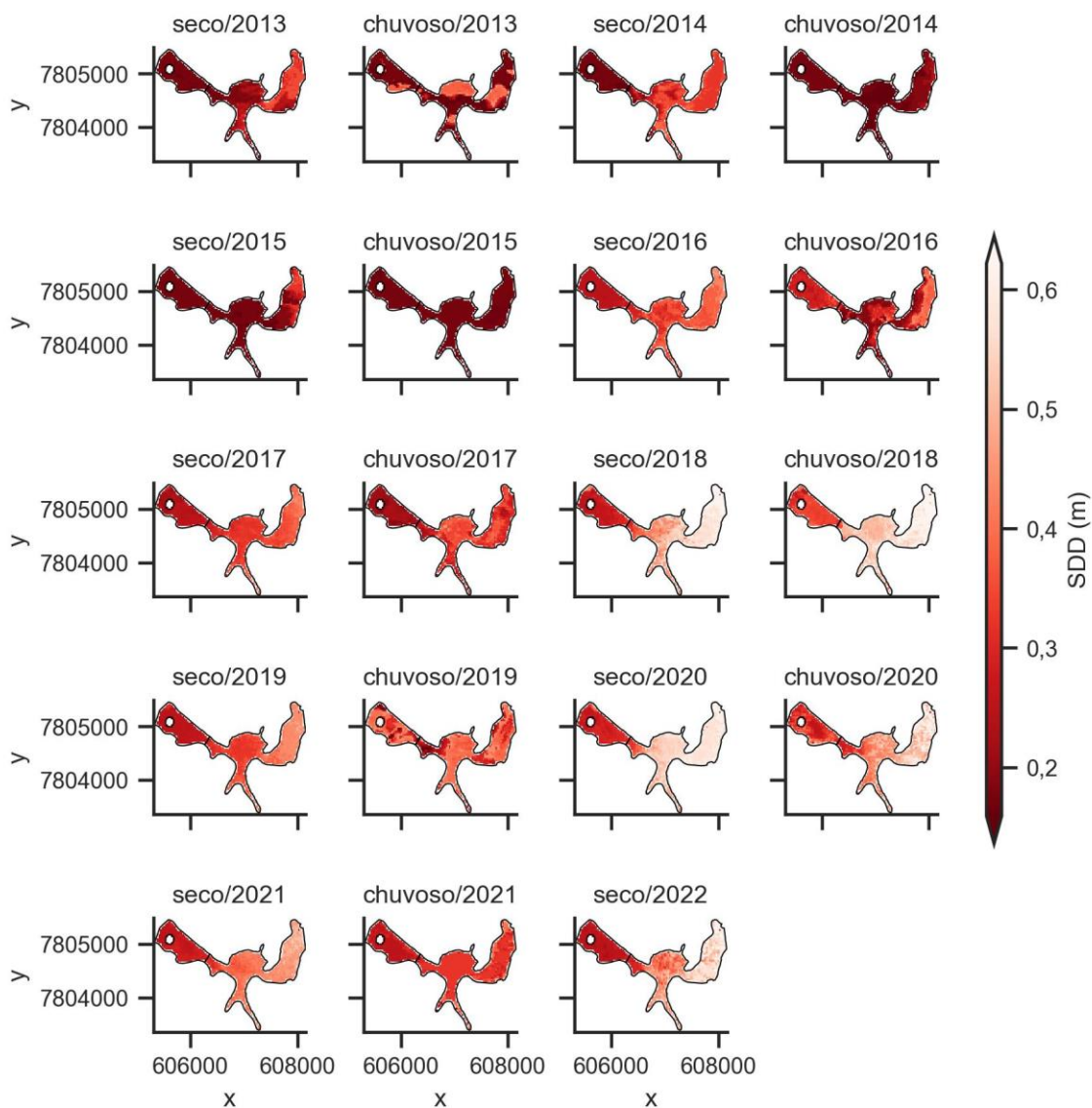
Figura 45 – Valores medianos de Chla ($\mu\text{g/L}$) por *pixel* considerando período chuvoso (6 meses) e período seco (6 meses) para cada ano durante o período estudado (2013-2022)



Nota: O período seco refere-se aos meses de abril a setembro para o ano apresentado no título de cada gráfico. O período chuvoso refere-se aos meses de outubro a dezembro para o ano apresentado no título e janeiro a março do ano seguinte.

Fonte: Próprio Autor (2023)

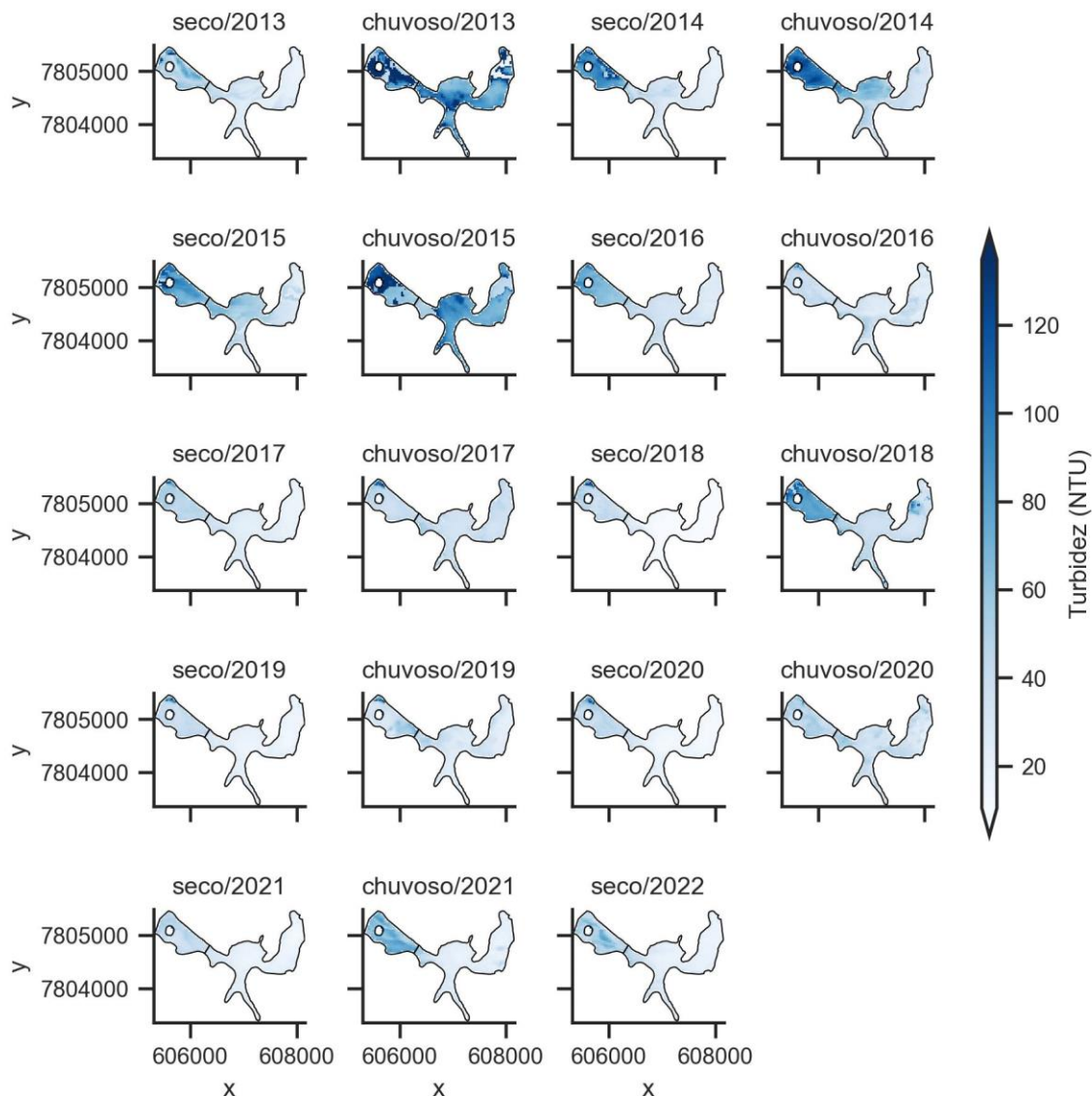
Figura 46 – Valores medianos de SDD (m) por *pixel* considerando período chuvoso (6 meses) e período seco (6 meses) para cada ano durante o período estudado (2013-2022)



Nota: O período seco refere-se aos meses de abril a setembro para o ano apresentado no título de cada gráfico. O período chuvoso refere-se aos meses de outubro a dezembro para o ano apresentado no título e janeiro a março do ano seguinte.

Fonte: Próprio Autor (2023)

Figura 47 – Valores medianos de Turbidez (NTU) por *pixel* considerando período chuvoso (6 meses) e período seco (6 meses) para cada ano durante o período estudado (2013-2022)



Nota: O período seco refere-se aos meses de abril a setembro para o ano apresentado no título de cada gráfico. O período chuvoso refere-se aos meses de outubro a dezembro para o ano apresentado no título e janeiro a março do ano seguinte.

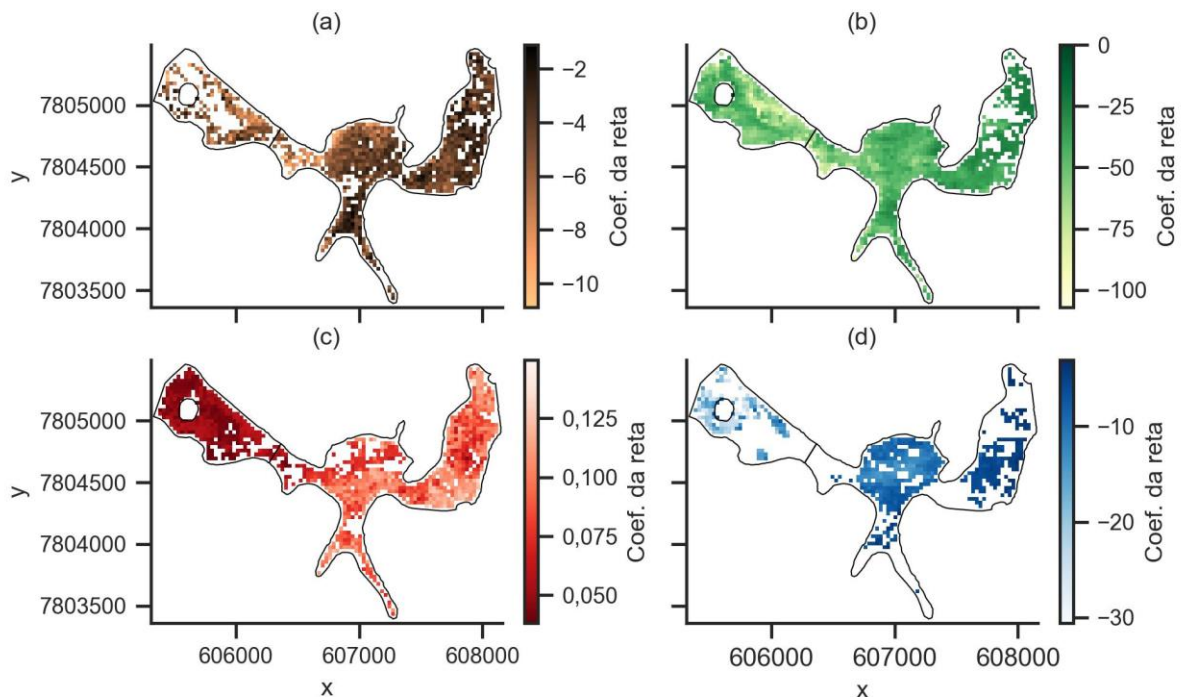
Fonte: Próprio Autor (2023)

O teste de *Mann-Kendall Seasonal* indica a melhora da qualidade da água durante o período estudado (2013-2022) para todos os parâmetros (Figura 48). Combinado com o método de *Theil-Sen*, os coeficientes da reta para os parâmetros TSM, Chla e Turbidez são negativos. Para o SDD, devido à natureza inversa aos outros parâmetros, os coeficientes são positivos. Para fins de simplicidade no entendimento, os resultados do coeficiente da reta serão descritos a seguir em valores absolutos (módulo).

Observam-se coeficientes da reta para o parâmetro TSM variando entre 2 e 10 ($n_{\text{pixels}}=935$) (Figura 48a), com valores mais baixos do coeficiente ocorrendo após a cortina de retenção, coincidente a região delimitada pelo G1 (região central) e G4 (jusante). Para os valores máximos do coeficiente de TSM, observa-se ocorrência na região G1 (região central) e G2 (montante) (Tabela 9).

Para a Chla, os coeficientes da reta variam entre 0 e 100 ($n_{\text{pixels}}=1.142$) (Figura 48b) em todo o reservatório. A região G4 (jusante) é a que apresenta valor mínimo do coeficiente (Tabela 9). Regiões que apresentam maior redução da Chla ao longo do período estudado (2013-2022) são o G1 (região central) e o G2 (montante) (Tabela 9).

Figura 48 – Coeficiente da reta calculado por método de *Theil-Sen* para o teste de Mann Kendall com sazonalidade para os parâmetros de qualidade de água modelados para a Lagoa da Pampulha



Legenda: (a) TSM, (b) Chla, (c) SDD, (d) Turbidez; Nota: Regiões sem valores de coeficiente da reta apresentaram p-valor > 5%.

Fonte: Próprio Autor (2023)

Os valores mínimos do coeficiente da reta para o SDD em todo o reservatório (Figura 48c) e para os grupos definidos pela análise multivariada são próximos (0,03~0,04) (Tabela 9). Para os valores máximos do coeficiente da reta (>0,13), que indicam

aumento da transparência da água do reservatório para o período (2013-2022), têm-se os grupos G1 (região central) e G4 (jusante) (Tabela 9).

Dentre os parâmetros de qualidade de água, a Turbidez é o que mais apresenta lacunas do coeficiente de reta devido a não obtenção de p-valores menores que 5% ($n_{\text{pixels}}=610$) (Figura 48d). De qualquer forma, pode-se observar valores significativos de redução da Turbidez para o período (2013-2022) com a ocorrência de valor máximo (30,6) para a região de montante do reservatório (G1).

Tabela 9 – Valores máximos e mínimos dos coeficientes da reta calculados por método de Theil-Sen para cada grupo obtido na análise multivariada

Grupo	TSM			Chla			SDD			Turbidez		
	N	Mín	Máx	N	Mín	Máx	N	Mín	Máx	N	Mín	Máx
1	409	1,13	10,9	490	23,1	107,1	351	0,03	0,13	301	3,1	15,3
2	179	1,71	10,9	348	24,1	97,5	304	0,03	0,07	141	5,3	30,6
3	0	-	-	1	0	0	0	-	-	0	-	-
4	345	1,1	7,9	296	15,5	59,3	402	0,04	0,15	162	3,1	9,1
5	2	5,7	9,4	7	32,5	69,0	13	0,03	0,05	6	11,3	25,6

Nota: Os resultados estão apresentados em valores absolutos (módulo). N amostral é diferente para os grupos

Fonte: Próprio Autor (2023)

6 DISCUSSÃO

Nesta seção, compara-se os resultados obtidos a partir dos modelos para inferência parâmetros de qualidade de água abordados com resultados da literatura. Além disso, discute-se os resultados da dinâmica da qualidade da água para um reservatório urbano obtidos a partir dos testes estatísticos.

6.1 Desempenho dos modelos na inferência de parâmetros de qualidade da água

A utilização do coeficiente de correlação linear de *Pearson* (r) permitiu o destaque de variáveis independentes que mais se correlacionaram com os parâmetros de qualidade de água, para o Landsat-8 foram 5 de 91 variáveis e para o Sentinel-2, 5 de 215 variáveis. Árias-Rodríguez *et al.* (2021) para a modelagem de TSM, Chla, SDD e Turbidez em lagos no México também utilizaram o coeficiente de correlação de *Pearson* para a seleção de variáveis e Árias-Rodríguez *et al.* (2020), por meio de análise de sensibilidade, observaram que a seleção de três a seis variáveis

independentes apresentou desempenho constante e satisfatório para modelos MLR, *Random Forest* e SVR, concordando com o número de cinco variáveis selecionados para o presente estudo.

A correlação intermediária obtida entre a banda do NIR e o parâmetro TSM ($r=0,65$ Landsat-8 e $r=0,61$ Sentinel-2) demonstra a capacidade desta região do espectro eletromagnético em capturar o aumento das concentrações de TSM possivelmente causado pela afluência de sedimentos inorgânicos ao corpo d'água (KIRK, 2011; MOBLEY, 1994). Situação ratificada para a região de montante da Lagoa da Pampulha, visto a correlação forte ($r=0,74$) observada entre o TSM e a localização do ponto de monitoramento em relação à cortina de retenção (Rede_Antesdarede). Saberioon *et al.* (2020) obtiveram correlação similar para TSM e a banda do NIR do Sentinel-2 ($r=0,61$).

Para a Chla, a correlação obtida para banda do NIR do Landsat-8 ($r=0,65$) e correlações próximas obtidas entre duas bandas da região do RedEdge (s2_re2 e s2_re3) e NIR (s2_nir) do Sentinel-2 ($r=0,45$ e $r=0,44$, respectivamente) indicam que regiões com maiores λ estão capturando o aumento da concentração de Chla no reservatório, o que pode estar associado com a correlação da concentração de Chla com a concentração de TSM. Saberioon *et al.* (2020) obtiveram correlação inferior para Chla e a banda do RedEdge-2 (s2_re2) do Sentinel-2 ($r=0,61$).

Para o SDD, o presente estudo apresentou correlações mais fortes do parâmetro com a banda do NIR (l8_nir) do Landsat-8 ($r=-0,34$) e do Sentinel-2 (s2_nir) ($r=-0,66$) quando comparadas com os resultados obtidos por Árias-Rodríguez *et al.* (2021) que apresentaram $r=-0,19$ e $r=0,48$, respectivamente. A instabilidade na correlação do SDD pode estar associada a incertezas nas medições devido à variabilidade na percepção do observador sobre o desaparecimento do disco (LEE et al., 2017).

Dentre as bandas do Landsat-8 e Sentinel-2, a banda do NIR (l8_nir) do Landsat-8 e a banda do RedEdge (s2_re2) foram as que apresentaram maior correlação com o parâmetro Turbidez, $r=0,89$ e $r=0,91$, respectivamente. Esses valores superaram os resultados de correlação encontrados para essas bandas no estudo de Árias-Rodríguez *et al.* (2021) para o Landsat-8 ($r=0,2$) e Sentinel-2 ($r=0,34$).

Conforme apresentado, para todos os parâmetros de qualidade de água observou-se dentre as bandas disponíveis dos satélites a obtenção de correlações mais fortes das bandas do NIR para o Landsat-8 ($I8_nir$) e região do RedEdge/NIR para o Sentinel-2 ($s2_re1$ a $s2_nir$). Isso indica a covariação que os parâmetros de qualidade de água apresentam e a dificuldade de distinguir os comportamentos espectrais apresentados por eles (DEKKER; VOS; PETERS, 2002), uma vez que a concentração de Chla contribui na concentração de TSM, e ambas contribuem na concentração de Turbidez e na medida do SDD.

Para as variáveis independentes que representam relações entre bandas observou-se que para a estimativa de TSM, o algoritmo de Nechad apresentou correlação inferior à banda do NIR ($I8_nir$), resultado também observado por Maciel *et al.* (2019), ressaltando a dificuldade de generalização dos coeficientes obtidos em águas costeiras por Nechad, Ruddick e Park (2010) para águas continentais. Razões de bandas que utilizaram a banda do SWIR podem estar amplificando o efeito de ruídos (reflexão especular e fatores atmosféricos), cuja relação SNR é baixa para esta região do espectro, e merecem atenção (OBAID *et al.*, 2021).

Razões entre bandas construídas para a potencialização da correlação das IOPs observadas para regiões entre 650 e 750nm com as concentrações de Chla (DALL'OLMO; GITELSON, 2005; MISHRA; MISHRA, 2012) permitiram a seleção da variável independente 2bandas para o Landsat-8 ($r=0,82$) e Sentinel-2 ($r=0,62$). Para o Landsat-8, menciona-se que a forte correlação obtida está diretamente associada com a amplitude de dados de Chla ($>1000 \mu\text{g/L}$), se tratando de uma relação estatística, e não física conforme preconizado pela relação de duas bandas (2bandas).

Para o SDD, correlações positivas de bandas de menores comprimentos de onda ($r_{I8_blue} = 0,07$ e $r_{I8_green} = 0,11$) combinadas com a banda do NIR ($I8_nir$) promoveram correlações mais fortes ($r > 0,55$ e $r < -0,55$), o que pode estar associado com baixas concentrações de carbono orgânico para regiões de jusante do reservatório (BRANDÃO; STAEHR; BEZERRA-NETO, 2016). No caso do Sentinel-2, a banda do SWIR apresentou correlação positiva com o SDD ($r=0,12$) e isso pode estar associado a fatores intervenientes como efeitos de adjacência, influência atmosférica e reflexão especular no espelho d'água (OGASHAWARA; MISHRA; GITELSON, 2017).

Para a Turbidez obteve-se para as cinco variáveis selecionadas correlação superior a 0,89 para ambos os satélites. No caso do Landsat-8, a influência da banda do infravermelho foi complementada com a multiplicação de bandas de baixo λ que apresentaram correlações fracas ($r < 0,36$) para a seleção das variáveis independentes ($r > 0,89$). No caso do Sentinel-2, a correlação das bandas do RedEdge (s2_re2 e s2_re3) ($r = 0,9$) cresceram para $r > 0,93$, quando combinadas com a banda do azul ($r = 0,2$).

As variáveis independentes descritas acima foram utilizadas na modelagem por meio de regressões para o ajuste de equações que permitissem a inferência dos parâmetros de qualidade da água (variáveis dependentes). Observa-se na literatura a utilização de métricas como R^2 , RMSE e nRMSE para a avaliação do desempenho dos modelos (GITELSON et al., 2008; MISHRA; MISHRA, 2012). Topp *et al.* (2020) por meio de estudo de revisão bibliográfica observaram valor médio de R^2 igual a 0,76, com amplitude entre 0,25 e 1,0, para a calibração de modelos de inferência de parâmetros de qualidade de água para sistemas aquáticos continentais. A alta variabilidade no desempenho de modelos encontrada na literatura está associada aos dados e métodos utilizados, ora pelas variáveis independentes – satélite escolhido, fatores intervenientes associados e derivações de relações entre bandas; ora pelas variáveis dependentes – devido à amplitude dos dados de qualidade da água, ao período considerado na modelagem, e aos modelos de regressão utilizados.

Para modelos semiempíricos e semianalíticos consolidados na literatura, observou-se a estimativa de TSM com a obtenção de R^2 de 0,82, RMSE de 13,4 mg/L e nRMSE de 52% por meio da modelagem de dados do sensor MERIS obtidos para a região *Southern North Sea* (NECHAD; RUDDICK; PARK, 2010). Maciel *et al.* (2019) para a modelagem de TSM em rios da Amazônia simularam dados de espectrorradiômetros para a representação das bandas do Landsat-8 e Sentinel-2 e obtiveram R^2 de 0,89, 0,91 e RMSE de 10,84, 9,81 mg/L, respectivamente. Para os resultados de ambos os satélites foram consideradas a regressão logarítmica e a banda do infravermelho (l8_nir e s2_nir). Para o presente estudo, considerando a modelagem de TSM por meio da banda do infravermelho para o Landsat-8 e Sentinel-2, obteve-se R^2_{moda} de 0,43, 0,36 e $\text{RMSE}_{\text{moda}}$ de 13,6 mg/L, 14,2 mg/L, respectivamente. Apesar do desempenho inferior, ao selecionar o modelo *Random Forest* para a estimativa de

TSM, conforme apresentado na Tabela 8, obteve-se R^2 igual a 0,88, RMSE igual a 8,1 mg/L e nRMSE de 29% para o Landsat-8, e R^2 igual a 0,72, RMSE igual a 8,5 mg/L e nRMSE igual a 42% para o Sentinel-2, o que supera o desempenho dos modelos supramencionados. Wen *et al.* (2022) ao compararem modelo de *Random Forest* com modelo de regressão linear simples (banda do vermelho), preferiram a utilização do primeiro para a estimativa de TSM por imagens do Landsat-8 com desempenho do R^2 igual a 0,63, e RMSE de 16,1 mg/L. Saberioon *et al.* (2020) utilizaram relações de bandas do Sentinel-2 para inferência de TSM utilizando modelo de aprendizado de máquina denominado *Cubist*. Os autores obtiveram R^2 de 0,96 e RMSE de 19,55 mg/L.

Para a estimativa de Chla, Gitelson *et al.* (2008) aplicaram o modelo 2bandas para dados do sensor MERIS e obtiveram desempenho do R^2 igual a 0,94, RMSE igual a 14,6 $\mu\text{g/L}$ e nRMSE de 45%. Augusto-Silva *et al.* (2014) simularam dados do OLCI por meio de medidas hiperespectrais e por meio do modelo de 2bandas obtiveram R^2 de 0,9 com RMSE igual a 9,65 $\mu\text{g/L}$ e nRMSE 4,75% para a estimativa de Chla no Reservatório de Funil. Os autores também propuseram a utilização de Simulação de Monte Carlo, que retornou R^2 entre 0,5 e 1,0, com valor R^2_{moda} entre 0,8 e 0,9. Watanabe *et al.* (2018) estimaram Chla para dados simulados do sensor OLI do Landsat-8 por meio do modelo 2bandas construído pela razão de bandas do vermelho ($\lambda=655\text{nm}$) e verde ($\lambda=560\text{nm}$) e obtiveram R^2 de 0,08. Pouco tempo depois, para dados do sensor MSI do Sentinel-2, Watanabe *et al.* (2019) testaram o modelo 2bandas para a estimativa de Chla em uma cascata de reservatórios em São Paulo, e obtiveram R^2 igual a 0,83, RMSE igual a 42,6 $\mu\text{g/L}$ e nRMSE de 21% ($\text{RMSE} / (y_{\text{max}} - y_{\text{min}})$). Souza (2019) testou o modelo 2bandas com dados de espectrorradiômetros para a estimativa de Chla na Lagoa da Pampulha e decidiu por descartar o modelo visto o baixo desempenho obtido ($R^2=0,05$) para a amplitude de concentrações no reservatório (40 - 664 $\mu\text{g/L}$). Para o presente estudo o desempenho de R^2 de 0,77, RMSE de 91,8 $\mu\text{g/L}$, nRMSE de 95% para o Landsat-8 e R^2 de 0,38, RMSE de 51,6 $\mu\text{g/L}$ e nRMSE de 92% para o Sentinel-2, obtido para o modelo 2bandas, demonstrase superior ao obtido para o estudo anterior realizado na Lagoa da Pampulha e inferior aos estudos apresentados em outros locais. Cabe mencionar que além da utilização de Simulação de Monte Carlo para que a aleatoriedade dos modelos pudesse abranger toda a variabilidade dos OACs, o cálculo do nRMSE utilizado por Watanabe

et al. (2019) diferencia-se do cálculo do nRMSE utilizado neste estudo (média dos dados observados), o que sugere cautela na interpretação e comparação dos resultados apresentados.

Zhang *et al.* (2021a) obtiveram desempenho de R^2 igual a 0,73 e RMSE de 0,63 m e nRMSE de 55,4% para a estimativa de SDD em diversos lagos na China com dados da banda do vermelho modelados do Landsat 5, 7 e 8. Os modelos selecionados no presente estudo para a modelagem de SDD (Tabela 8) apresentaram R^2 de 0,65, RMSE de 0,09 e nRMSE de 29% para o Landsat-8, e R^2 de 0,86, RMSE de 0,06m e nRMSE de 15% para o Sentinel-2. Resultados próximos foram observados em Maciel *et al.* (2021), que também utilizaram *Random Forest*, com dados do Sentinel-2, para a estimativa de SDD no reservatório do *Billings* em São Paulo. O desempenho obtido apresentou coeficiente de correlação linear de *Pearson* (r) igual a 0,66 e RMSE de 0,18 m. Cui *et al.* (2022) estimaram SDD no *Chaohu Lake* na China e obtiveram R^2 igual a 0,88 e RMSE 14,8 cm com dados do Sentinel-2 e Landsat-8.

Pela banda do vermelho do Landsat-8, Hossain, Mathias e Blanton (2021), ajustaram modelo para a estimativa de Turbidez no *Tennessee River* que apresentou desempenho de R^2 igual a 0,94 e RMSE igual a 18,1 NTU. Arias-Rodriguez *et al.* (2020) recorreram a um modelo de aprendizado de máquina para o ajuste de dados do MERIS para a estimativa de Turbidez e obtiveram R^2 igual a 0,83 e RMSE igual a 1,10 NTU. Com a banda do infravermelho dos satélites Landsat-8 e Sentinel-2, o presente estudo apresentou desempenho melhor por meio do ajuste linear com os dados de Turbidez, uma vez que o quantitativo de amostras é baixo e desfavorece a utilização de modelos de aprendizado de máquina (MACIEL *et al.*, 2021). Os resultados obtidos R^2 igual a 0,8, 0,81, RMSE igual 26,2, 8,4 NTU e nRMSE 52%, 33%, para os satélites Landsat-8 e Sentinel-2, respectivamente, mostram-se satisfatórios e com melhor desempenho do que o modelo proposto por Souza (2019) que utilizou a banda NIR do Landsat-8 (I8_nir) e obteve nRMSE igual a 62%.

Apesar dos modelos semianalíticos e semiempíricos, graças ao embasamento físico, serem passíveis de generalização para outras áreas de estudo (MATTHEWS, 2011), apenas os parâmetros Chla e Turbidez apresentaram desempenhos melhores quando utilizadas em sua modelagem variáveis independentes com as quais possuem relação

física. Os outros parâmetros TSM e SDD foram ajustados por meio do método *Random Forest*. Dentre os possíveis fatores de distinção dos modelos selecionados por parâmetro, menciona-se a alta variabilidade de concentrações do parâmetro Chla (0 a 1.177,5 $\mu\text{g/L}$) e o baixo número de registros de turbidez ($n=16$ para o Landsat-8 e $n=9$ para o Sentinel-2), que podem forçar a ocorrência de superparametrização dos modelos, como observado para a modelagem da Turbidez (R^2_{moda} de 0,96 para o Landsat-8 e R^2_{moda} de 1,0 para o Sentinel-2).

Para comparar o desempenho entre os satélites utilizados, deve-se ter em mente que o conjunto de dados utilizados para a modelagem de cada satélite é diferente (amplitude de valores dos parâmetros e n_{amostral} diferentes). Por meio dos resultados apresentados, o Landsat-8 superou o Sentinel-2 na estimativa apenas de TSM ($n\text{RMSE}_{L8}=29\%$ e $n\text{RMSE}_{S2}=42\%$). De forma contrária, para este parâmetro, Maciel *et al.* (2019) obtiveram melhores desempenhos dos modelos do Sentinel-2 em relação ao Landsat-8, o que reforça a dificuldade de adaptação de domínio e o desafio da generalização do desempenho dos modelos de cada satélite (OBAID *et al.*, 2021).

No caso da Chla, na Lagoa da Pampulha, o desempenho similar obtido entre satélites ($n\text{RMSE}_{L8}=95\%$ e $n\text{RMSE}_{S2}=93\%$) está associado com a presença de um valor extremos de Chla no conjunto de dados do Landsat-8 (Chla: 1.177,5 $\mu\text{g/L}$). Na literatura observou-se a melhor desempenho do Sentinel-2 na estimativa de Chla (ARIAS-RODRIGUEZ *et al.*, 2021) devido a presença de bandas entre a região do vermelho e NIR que potencializam a identificação das IOPs do parâmetro (LOBO *et al.*, 2021). Por exemplo, para a estimativa de Chla, Watanabe *et al.* (2018) obtiveram desempenho melhor quando utilizada a banda RedEdge ($\lambda=705\text{nm}$) do Sentinel-2 quando comparado com modelos que utilizaram as bandas do vermelho e NIR do Landsat-8. No caso do SDD o Sentinel-2 apresentou melhor desempenho ($n\text{RMSE}_{S2}=15\%$) que o Landsat-8 ($n\text{RMSE}_{L8}=29\%$), concordando com a comparação realizada por Pizani *et al.* (2020) ($n\text{RMSE}_{S2}=37\%$, $n\text{RMSE}_{L8}=41\%$). A modelagem da Turbidez para a Lagoa da Pampulha apresentou melhor desempenho com os dados do Sentinel-2 ($n\text{RMSE}_{S2}=33\%$) comparado ao desempenho do Landsat-8 ($n\text{RMSE}_{L8}=52\%$). Pizani *et al.* (2020) para a modelagem de Turbidez no reservatório de Três Marias, reservatório de usina hidroelétrica com, elevado tempo de detenção

hidráulica e estado oligotrófico, obtiveram valores próximos entre os satélites comparados ($nRMSE_{S2}=42,9\%$, $nRMSE_{L8}=42,7\%$).

Mesmo que o Landsat-8 tenha apresentado desempenho inferior para a maioria dos parâmetros, comenta-se que a data de lançamento desse (2013) em comparação a de lançamento do Sentinel-2 (A: 2015, B: 2017) permitiu uma avaliação mais extensa da qualidade da água. Por isso, e em razão da existência de outros satélites da série Landsat anteriores (Landsat-5, Landsat-7) e posteriores ao Landsat-8 (Landsat-9), recomenda-se a utilização desse no monitoramento da qualidade da água na Lagoa da Pampulha.

Na literatura, notou-se estudos que utilizaram datas específicas (SOUZA, 2019), períodos do ano específicos (BORGES et al., 2020), e outros que consideraram mais de um ano (ZABALETA; ACHKAR; AUBRIOT, 2021) para a modelagem da qualidade da água e generalização nas dimensões espaciais e temporais. O presente estudo apresentou dificuldades em maximizar o desempenho dos modelos ao utilizar aproximadamente 9 anos de dados para a modelagem da qualidade da água na Lagoa da Pampulha. Isso ocorreu devido à complexidade de fatores associados à qualidade da água, incluindo variáveis climáticas, antrópicas e, ambientais que favorecem problemas de superparametrização, ruídos e multicolineariedade na modelagem realizada (HANSEN et al., 2020). Entretanto, desde que construídos e avaliados de forma robusta – com validação em amostras independentes, e utilização de métodos estatísticos consolidados; modelos com maior representatividade amostral, como os modelos propostos no presente trabalho, permitem maior capacidade de generalização e previsibilidade dos fatores supramencionados (TOPP et al., 2020).

Outro ponto de destaque é a dependência que dados de sensoriamento remoto de satélites apresentam em relação a métodos de correção atmosférica. Apesar de existirem modelos de qualidade da água que utilizam informações sem correção atmosférica (ARANHA et al., 2022; PEPPA; VASILAKOS; KAVROUDAKIS, 2020), notou-se a preferência de utilização de modelos com correção para a remoção de interferências da atmosfera nas IOPs dos *pixels* correspondentes às águas continentais (MOSES et al., 2017). No GEE encontra-se dados de reflectância do Landsat-8 e Sentinel-2 com correção atmosférica, LaRSC e Sen2Cor,

respectivamente. Contudo, faltam imagens com correção atmosférica para o período inicial de imageamento do Sentinel-2 e os métodos mencionados são voltados para a correção de dados de sensoriamento remoto de observação da terra, não sendo adequadas para sistemas aquáticos. Como alternativa, o método SIAC proposto por Yin, Lewis e Gómez-Dans (2022) apresenta abordagem multisensor (Landsat-8 e Sentinel-2) e vem sendo empregado em estudos de sensoriamento remoto aplicado à cor da água, como na aplicação do SIAC em dados do Sentinel-2 para a estimativa de Chla, realizado por Lobo *et al.* (2021) e para a estimativa de SDD, realizada por Maciel *et al.* (2021). O primeiro estudo obteve coeficiente de correlação de *Pearson* (r), igual a 0,96 e erro médio percentual (MAPE) igual a 70,6% quando comparadas todas as bandas utilizadas com dados de espectrorradiômetro coletados em campo. O segundo obteve $r=0,44$, MAPE = 20,14% para a banda do azul e $r=0,91$ e MAPE=41% para a banda do NIR. No presente estudo não foi possível avaliar a precisão do método de correção atmosférica devido à inexistência de dados de espectrorradiômetro, entretanto observou-se valores maiores de reflectância para a região do SWIR em ambos os satélites, que indicam possível insuficiência do método de correção atmosférica em remover interferências para essa região do espectro ou indicam necessidade de correções adicionais como remoção de reflexão especular e efeitos de adjacência.

Por fim, problemas de efeito de adjacência podem ter resultado na anomalia observada com a formação de um grupo com três pixels próximo à Ilha dos Amores no reservatório. Paulino *et al.* (2022) estudaram efeito de adjacência para reservatórios no Brasil e verificaram um total de efeitos de adjacência de 3% em relação a área do reservatório *Billings*, com maiores valores ocorrendo ao longo dos braços estreitos dos afluentes. Liu *et al.* (2021) mencionaram que quando desconsiderados cinco *pixels* da margem do *Lake Xingyun*, um lago eutrófico com área de 31,7 km², a influência dos efeitos de adjacência pôde ser desconsiderada para dados registrados pelo Landsat-8 e Sentinel-2. Infelizmente, para a Lagoa da Pampulha, por apresentar apenas 2km², existem regiões em que os braços apresentam largura inferior à 120 metros, o que impossibilita a desconsideração dessa quantidade de *pixels*. Para estudos futuros, sugere-se a avaliação dos efeitos de adjacência, principalmente para os grupos G1 (região central), G3 (Ilha dos Amores) e G5 (afluência Ressaca/Sarandi)

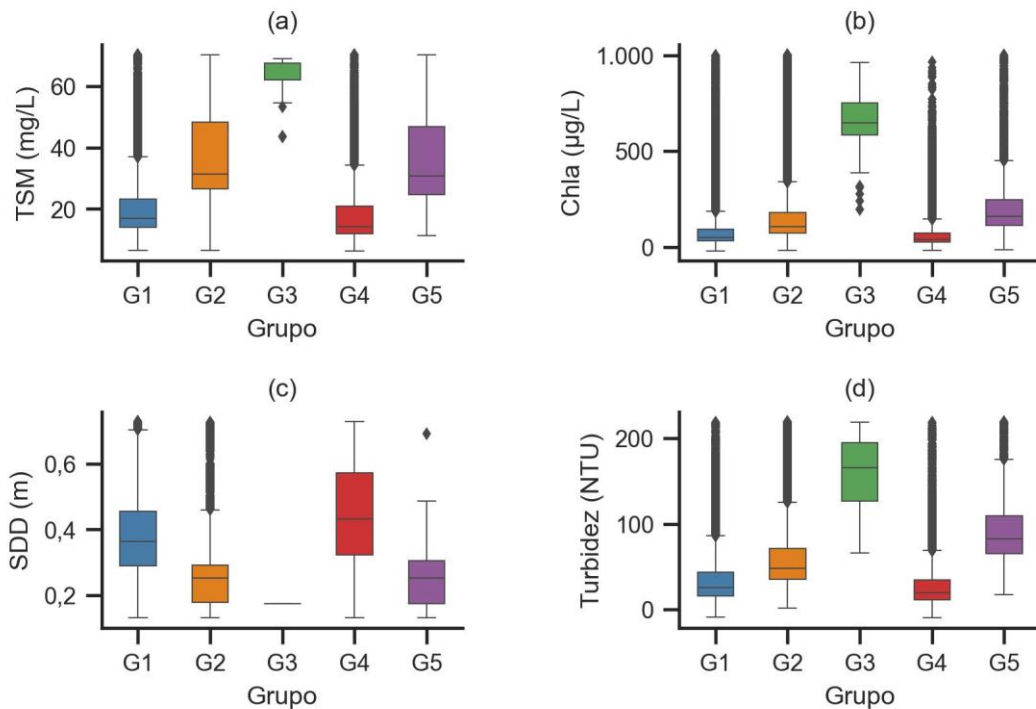
6.2 Fatores, padrões, e tendências da qualidade da água para reservatórios urbanos

Apesar das incertezas inerentes aos modelos propostos para estimativas da qualidade da água pelo sensoriamento remoto, os resultados alcançados permitiram a avaliação de fatores, padrões e tendências na qualidade da água para a Lagoa da Pampulha.

De forma exploratória, Souza (2019) identificou o melhor desempenho de modelos em situações que a Lagoa da Pampulha era separada em duas regiões. Apesar da separação ocorrer por estrutura física (cortina de retenção), o presente estudo preferiu abranger uma quantidade maior de grupos para a avaliação do comportamento de diferentes regiões do reservatório. A análise multivariada indicou a presença de cinco grupos heterogêneos na Lagoa da Pampulha que se diferenciaram em relação aos parâmetros modelados (TSM, Chla, SDD e Turbidez) para o período de (2013-2022). Os grupos formados foram associados à localização geográfica e os afluentes que contribuem para cada uma delas, assim como observado por Varol *et al.* (2012).

A Figura 49 demonstra a pior qualidade da água, com valores mais altos de TSM, Chla e Turbidez, e valores mais baixos de Secchi, ocorrendo para o G2 (montante), que representa região de chegada dos afluentes Ressaca/Sarandi que aportam mais de 1500 ton/ano de TSM (FURTADO; MONTE-MOR; COUTO, 2021) e representa região com altas concentrações de nutrientes e carbono orgânico dissolvido devido a atividades antrópicas (BEZERRA *et al.*, 2022). Para os G1 e G4, regiões centrais e de jusante do reservatório, respectivamente, observa-se menores valores de TSM, Chla e Turbidez e maiores valores de SDD visto que os processos físicos, químicos, biológicos no reservatório favorecem a sedimentação e depuração de poluentes (TRINDADE *et al.*, 2017)

Figura 49 – Boxplots com a distribuição dos valores dos parâmetros de qualidade de água modelados apresentados por grupos



Legenda: a) TSM; b) Chla; c) SDD; d) Turbidez.

Fonte: Próprio Autor (2023)

A confirmação da diferença entre amostras para o período chuvoso e período seco observada para os parâmetros TSM, Chla, SDD e Turbidez nos grupos G1 (região central) e G4 (jusante) demonstram a influência que o período chuvoso apresenta no reservatório, levando a piora da qualidade da água. Isso também foi observado por Silva *et al.* (2019) ao modelar a contribuição de fósforo total (TP), nitrato (NO_3^-) e TSM da área de contribuição na Lagoa da Pampulha, apresentando relação 60:1 para TSM, 10:1 para NO_3^- e 4:1 para TP, para o período chuvoso e período seco, respectivamente. Esperava-se que a região de montante, por receber a carga orgânica e inorgânica dos principais afluentes do reservatório apresentasse diferença entre período chuvoso e seco. A insuficiência do teste estatístico em capturar este comportamento pode estar associada à inexistência de imagens livre de nuvens concomitantes aos eventos chuvosos. Em bacias hidrográficas no Nordeste do Brasil, Freire, Costa e Lima Neto (2021) também confirmaram por teste de hipótese que a Turbidez monitorada apresenta maiores concentrações no período chuvoso. De forma oposta aos resultados obtidos no presente estudo, Freire, Costa e Lima Neto (2021) observaram para a maioria das bacias hidrográficas estudadas maiores

concentrações de Chla e TSM durante o período seco. Assume-se que resultados distintos apresentados estão associados com a natureza do estudo realizado na Lagoa da Pampulha, um reservatório urbano com regime lântico, e estudo realizado no nível de bacias hidrográficas com pontos de monitoramento localizados em rios, que apresentam regime lótico e inseridos em contexto rural. Além disso, o primeiro utiliza dados de sensoriamento remoto, enquanto o segundo dados monitorados *in situ*, que podem resultar em diferentes indicativos da dinâmica da qualidade da água, conforme observado por Hansen *et al.* (2020).

O teste de hipótese comparando a aplicação de produto químico para o tratamento da Lagoa da Pampulha indicou melhoria da qualidade da água por meio da redução dos parâmetros TSM, Chla e Turbidez e aumento da transparência (SDD). A redução da concentração de fósforo foi observada em estudo realizado por Barçante *et al.* (2020) que compararam momentos anteriores e posteriores à aplicação de *Phoslock®* na Lagoa da Pampulha. Bezerra *et al.* (2022) também constataram a redução de fósforo para períodos posteriores a aplicação de produtos para tratamento do reservatório, entretanto, observaram aumento da concentração de carbono na coluna d'água do reservatório e dos afluentes. Apesar do fósforo não ter sido modelado no presente estudo, esse atua como fator limitante e principal variável associada com ocorrência de floração de algas, incluindo espécies tóxicas de cianobactérias (SILVA *et al.*, 2019), o que permite a associação da redução de fósforo observada pelos autores supramencionados com a redução das concentrações de TSM, Chla e Turbidez observada para o presente estudo. Mesmo com as reduções observadas das concentrações dos parâmetros posteriores ao tratamento, a Lagoa da Pampulha manteve-se em estado eutrófico/hipereutrófico (FURTADO; MONTE-MOR; COUTO, 2021) com dificuldades de atingir o limite legal estabelecido no enquadramento do reservatório (Chla: 60 µg/L, Turbidez: 100 NTU) (BRASIL, 2005). Outro problema é a dependência observada da aplicação do produto, uma vez que a paralisação demonstrou maior variabilidade das concentrações dos parâmetros, indicando a permanência do aporte de nutrientes externos (poluição difusa e pontual) (SILVA *et al.*, 2019) e aportes internos (ressuspensão de nutrientes e sedimentos presentes no fundo do reservatório) (BARÇANTE *et al.*, 2020).

Os índices SPI₁₂ obtidos estão de acordo com observações sobre a ocorrência de períodos secos comentada por Jardim *et al.* (2019) para a região de Belo Horizonte. Freire, Costa e Lima Neto (2021) observaram o aumento da concentração de nutrientes para anos categorizados como secos (índice SPI₁₂) em bacias hidrográficas no Nordeste, confirmando que eventos extremos, como chuvas intensas e secas severas alteram a dinâmica da qualidade da água (XUE *et al.*, 2016). Para o presente estudo, as classes do SPI₁₂ apresentaram diferença significativa entre os cinco grupos classificados para a Lagoa da Pampulha, entretanto a relação de piora da qualidade da água para períodos chuvosos observada anteriormente não foi observada para o presente caso. Dentre os motivos, supõe-se que:

- A utilização do período de 12 meses para a construção do SPI₁₂ refletiu na obtenção de valores dos parâmetros modelados de um número maior de imagens do período seco do que do período chuvoso (ex: índice SPI₁₂ normal $n_{img}=90$, $n_{img_seco}=77$ $n_{img_chuovoso}=13$);
- Anos categorizados como muito secos e secos (2014, 2016, e 2017) são anos que estão próximos do início do período estudado. Anos categorizados como úmido ou muito úmidos (2020, 2021 e 2022) são anos que estão próximos do final do período estudado. Conforme análise de tendência realizada, obteve-se melhora da qualidade da água para os parâmetros modelados durante o período (2013-2022).

Portanto, a relação do SPI₁₂ com a qualidade da água para a Lagoa da Pampulha é fortemente influenciada pelo conjunto amostral modelado e pelas condições de melhora da qualidade da água observada no reservatório, dificultando a distinção da contribuição do índice SPI₁₂ na variação da qualidade da água.

Em relação à tendência da qualidade da água, notou-se que os testes de Mann-Kendall indicaram redução das concentrações de TSM, Chla, Turbidez, e aumento nas medidas da transparência (SDD) para o período (2013-2022). Ao avaliar a evolução da transparência no *Lake Liangzi* na China, Xu *et al.* (2018) observaram redução das medidas de SDD que estão associadas com a intensificação de atividades antrópicas na área de contribuição. Em contraste, Zhang *et al.* (2021b)

estudaram diversos sistemas lênticos na Mongólia e observaram o aumento da transparência significativa (p -valor $<5\%$) para 205 de 594 corpos hídricos.

Hansen *et al.* (2020) observaram redução da Chla modelada por imagens do Landsat-8 para o *Utah Lake*, com uma taxa de $-0,25$ ($\mu\text{g/L/ano}$) calculada por método *Theil-Sen*. Na Pampulha, a avaliação e constatação da melhora da qualidade da água ao longo dos anos foi realizada por Furtado, Monte-Mor e Do Couto (2019). Além da aplicação de produtos químicos (*Phoslock®* e *Enzilimp®*), os autores mencionaram diversas atividades realizadas no reservatório e bacia hidrográfica, como obras de dragagem para o desassoreamento e remoção de resíduos do espelho d'água que auxiliaram para o melhor desempenho ecológico do reservatório.

7 CONCLUSÕES

A partir deste estudo foi possível avaliar o uso do sensoriamento remoto para o monitoramento da qualidade da água em um reservatório urbano com os principais resultados:

- Os satélites utilizados (Sentinel-2 e Landsat-8) permitiram a inferência dos parâmetros TSM, Chla, SDD e Turbidez. O Sentinel-2 superou o Landsat-8 para os parâmetros SDD e Turbidez, teve desempenho similar para o parâmetro Chla e apresentou desempenho pior para a estimativa de TSM.
- Modelos de regressão univariados permitiram estimar os parâmetros Chla e Turbidez. Modelos mais robustos que utilizaram método *Random Forest* apresentaram desempenho superior na estimativa de TSM e SDD.
- Para a dimensão de tempo, observou-se incremento modesto ($n=36$) da quantidade de registros disponíveis quando utilizado o sensoriamento remoto ($n=155$ datas) para a complementação do monitoramento *in situ* ($n=119$ datas). Isso ocorreu devido à dificuldade de utilização de imagens durante o período chuvoso (presença de nuvens e sombra de nuvens).
- Para a dimensão espacial (x e y), notou-se aumento significativo da quantidade de dados de qualidade da água quando utilizado o sensoriamento remoto ($n=1285$) para a complementação do monitoramento *in situ* ($n=10$). Deve-se atentar para fatores intervenientes na modelagem de dados de reflectância para algumas regiões, por exemplo, no entorno do reservatório.

Os dados de sensoriamento remoto modelados permitiram a avaliação da dinâmica da qualidade da água de um reservatório urbano por meio da construção de uma matriz multidimensional (tempo=155, y=71, x=95). Os tópicos apresentados a seguir retratam a utilização de métodos estatísticos para a avaliação da qualidade da água pelos dados modelados.

- A delimitação de cinco grupos no reservatório indicou a definição de regiões heterogêneas com relação aos parâmetros de qualidade da água. Dos cinco grupos, três apresentaram maior relevância pela quantidade de *pixels* e amplitude dos valores de concentração dos parâmetros, estando esses definidos a seguir:
 - G1: região central do reservatório, apresentou concentrações dos parâmetros de qualidade de água inferiores ao G2 e superiores ao G4.
 - G2: região de montante, por ser a região de encontro dos afluentes Ressaca/Sarandi apresentou concentrações de TSM, Chla e Turbidez mais elevadas.
 - G4: região de jusante, apresentou as menores concentrações de TSM, Chla e Turbidez, e maior transparência.
- Para os grupos definidos, notou-se a distinção entre os períodos chuvosos e secos, com maiores valores dos parâmetros TSM, Chla, e Turbidez, e menores valores de SDD ocorrendo para o período chuvoso.
- O tratamento com produtos químicos na Lagoa da Pampulha apresentou distinção entre as classes anterior ao tratamento, durante o tratamento e paralisações no tratamento, com menores concentrações em todos os grupos de TSM, Chla e Turbidez para os meses com a aplicação de *Phoslock®* e *Enzilimp®*. Verificou-se também a dependência do reservatório na continuidade da aplicação haja vista o aumento das concentrações dos parâmetros durante paralisações da aplicação que refletem a disponibilidade de nutrientes advindos de fontes internas e externas ao reservatório.
- A associação do índice SPI₁₂ com parâmetros de qualidade de água para os grupos espaciais determinados (G1, G2 e G4) foi inconclusiva devido ao período utilizado para a construção do SPI₁₂ (12-meses), ao desbalanceamento entre os dados do período chuvoso e período seco, e à tendência observada na qualidade da água no reservatório ao longo dos anos.

- O teste de Mann-Kendall combinado com *Theil-Sen* indicou melhora para o período (2013-2022) da qualidade da água da Lagoa da Pampulha por meio da obtenção de coeficientes negativos para TSM, Chla e Turbidez, e coeficiente positivo para SDD.

8 RECOMENDAÇÕES E PERSPECTIVAS FUTURAS

Haja vista a quantidade de informações disponíveis e os avanços computacionais recentes, vislumbra-se a utilização de dados de sensoriamento remoto dos satélites Landsat-8 e Sentinel-2 no nível L1C integrados com métodos de correções atmosféricas específicos para a corpos hídricos continentais e processamento em nuvem. Inclusive, há espaço para o desenvolvimento de todo o processo de modelagem, obtenção das matrizes multidimensionais e análise da dinâmica da qualidade da água no GEE. Recomenda-se também o desenvolvimento de uma aplicação na plataforma para a visualização dos resultados de forma dinâmica.

Para maior assertividade na construção de modelos que inferem parâmetros de concentração de qualidade de água, sugere-se o investimento em monitoramento *in situ* orientado à construção de modelos por sensoriamento remoto, incluindo monitoramento concomitante à passagem dos satélites, utilização de espectro radiômetro para a avaliação de métodos de correção atmosférica e efeitos de adjacência com a estipulação de um plano de trabalho para a padronização das coletas e realização das análises.

As análises e avaliações a respeito da dinâmica da qualidade da água apresentada no presente estudo focaram mais em técnicas exploratórias. Para estudos futuros sugere-se uma abordagem voltada à avaliação da gestão dos recursos hídricos, tanto para a inferência de classes de parâmetros de qualidade da água (ex: Chla < 60 µg/L) com foco na legislação vigente, quanto para a combinação com variáveis governantes da qualidade da água no reservatório (temperatura, vento, precipitação, uso e cobertura do solo, e atividades antrópicas) associada com métodos avançados estatísticos (métodos estocásticos, estatística Bayesiana, entre outros).

REFERÊNCIAS

- ALCÂNTARA, E. et al. A satellite-based investigation into the algae bloom variability in large water supply urban reservoirs during COVID-19 lockdown. **Remote Sensing Applications: Society and Environment**, v. 23, p. 100555, 2021.
- AMORIM, C. **Projeto P&D Aneel/CEMIG2016-GT-607-2019-2023** . , 2023.
- ANDERSON, J. A. Regression and ordered categorical variables. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 46, n. 1, p. 1–22, 1984.
- ARANHA, T. R. B. T. et al. Remote Analysis of the Chlorophyll-a Concentration Using Sentinel-2 MSI Images in a Semiarid Environment in Northeastern Brazil. **Water (Switzerland)**, v. 14, n. 3, 1 fev. 2022.
- ARIAS-RODRIGUEZ, L. F. et al. Monitoring Water Quality of Valle de Bravo Reservoir, Mexico, Using Entire Fifespan of Meris Data and Machine Learning Approaches. **Remote Sensing**, v. 12, n. 10, 1 maio 2020.
- ARIAS-RODRIGUEZ, L. F. et al. Integration of Remote Sensing and Mexican Water Quality Monitoring System Using an Extreme Learning Machine. **Sensors**, v. 21, n. 12, 2 jun. 2021.
- AUGUSTO-SILVA, P. B. et al. Analysis of MERIS Reflectance Algorithms for Estimating Chlorophyll-a Concentration in a Brazilian Reservoir. **Remote Sensing**, v. 6, n. 12, p. 11689–11707, 2014.
- BADLOWSKI, G. A.; ADOLF, J. E.; FOUAD, G. Spatial analysis of water quality parameters in Hilo Bay, Hawai'i, using a combination of interpolated surfaces and hot spot analysis. **Environmental Monitoring and Assessment**, v. 193, n. 3, 1 mar. 2021.
- BARBOSA, C. C. F.; NOVO, E. M. L. M.; MARTINS, V. S. **Introdução ao Sensoriamento Remoto de Sistemas Aquáticos**. 1st. ed. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2019.

BARÇANTE, B. et al. Cyanobacteria dynamics and phytoplankton species richness as a measure of waterbody recovery: Response to phosphorus removal treatment in a tropical eutrophic reservoir. **Ecological Indicators**, v. 117, 1 out. 2020.

BEATO, D. A. C.; DUTRA, G. M.; MEDEIROS, M. J. **Projeto Pampulha: Estudo Hidrogeológico da Bacia da Lagoa da Pampulha**. Belo Horizonte: [s.n.].

BEHMEL, S. et al. Water quality monitoring strategies — A review and future perspectives. **Science of the Total Environment**, v. 571, p. 1312–1329, 15 nov. 2016.

BERK, A. et al. **MODTRAN4 radiative transfer modeling for atmospheric correction**. Proc.SPIE. **Anais...20** out. 1999. Disponível em: <<https://doi.org/10.1117/12.366388>>

BEZERRA, M. P. et al. Water quality evaluation and dissolved organic matter characterization of a tropical hypereutrophic reservoir and its streams treated with Phoslock® and microbial bioremediation Enzilimp®. **Environmental Science and Pollution Research**, v. 29, p. 1375–1390, 2022.

BORGES, H. D. et al. Monitoring cyanobacteria occurrence in freshwater reservoirs using semi-analytical algorithms and orbital remote sensing. **Marine and Freshwater Research**, v. 71, n. 5, p. 569–578, 2020.

BRANDÃO, L. P. M.; STAEHR, P. A.; BEZERRA-NETO, J. F. Seasonal changes in optical properties of two contrasting tropical freshwater systems. **Journal of Limnology**, v. 75, n. 3, p. 508–519, 2016.

BRASIL. **Lei nº 9.433, de 8 de janeiro de 1997. Institui a Política Nacional de Recursos Hídricos, cria o Sistema Nacional de Gerenciamento de Recursos Hídricos, regulamenta o inciso XIX do art. 21 da Constituição Federal, e altera o art. 1º da Lei nº 8.001, de 13 de março de 1990, que modificou a Lei nº 7.990, de 28 de dezembro de 1989**. Diário Oficial da União, de 09 de jun. de 1997, p. 470, , 1997.

BRASIL. **Resolução CONAMA nº 357, de 17 de março de 2015. Dispõe sobre a classificação dos corpos de água e diretrizes ambientais para o seu**

enquadramento, bem como estabelece as condições e padrões de lançamento de efluentes, e dá outras providências. Diário Oficial da União, 18 de mar. de 2005, , 2005.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5–32, 2001.

BRICAUD, A.; ROESLER, C.; ZANEVELD, J. R. V. In situ methods for measuring the inherent optical properties of ocean waters. **Limnology and Oceanography**, v. 40, n. 2, p. 393–410, 1995.

CABALLERO, I.; NAVARRO, G. Monitoring cyanoHABs and water quality in Laguna Lake (Philippines) with Sentinel-2 satellites during the 2020 Pacific typhoon season. **Science of the Total Environment**, v. 788, 20 set. 2021.

CAI, X. et al. Urban Water Quality Assessment based on Remote Sensing Reflectance Optical Classification. **Remote Sensing**, v. 13, n. 20, 1 out. 2021.

CAO, Q. et al. Monitoring Water Quality of the Haihe River based on Ground-Based Hyperspectral Remote Sensing. **Water (Switzerland)**, v. 14, n. 1, 1 jan. 2022.

CAO, Z. et al. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. **Remote Sensing of Environment**, v. 248, p. 111974, 2020.

CBH RIO DAS VELHAS. **Poluição ainda desafia Lagoa da Pampulha**. Disponível em: <<https://cbhvelhas.org.br/noticias/poluicao-ainda-desafia-lagoa-da-pampulha/>>. Acesso em: 22 maio. 2022.

CERVANTES, J. et al. A comprehensive survey on support vector machine classification: Applications, challenges and trends. **Neurocomputing**, v. 408, p. 189–215, 2020.

CHEN, S. et al. Geographical Patterns of Algal Communities Associated with Different Urban Lakes in China. **International Journal of Environmental Research and Public Health**, v. 17, n. 3, 2020.

CHEN, W. et al. Possibility of using multiscale normalized difference vegetation index data for the assessment of total suspended solids (TSS) concentrations in surface water: A specific case of scale issues in remote sensing. **Environmental Research**, v. 194, 1 mar. 2021.

COELHO, R. M. P. Effects of eutrophication on seasonal patterns of mesozooplankton in a tropical reservoir: a 4-year study in Pampulha Lake, Brazil. **Freshwater Biology**, v. 40, n. 1, p. 159–173, 1 ago. 1998.

COFFER, M. M. et al. Quantifying national and regional cyanobacterial occurrence in US lakes using satellite remote sensing. **Ecological Indicators**, v. 111, 1 abr. 2020.

COIMBRA, K. T. O.; ALCÂNTARA, E.; DE SOUZA FILHO, C. R. Satellite evidence for pervasive water eutrophication in the Doce River reservoirs following the collapse of the Fundao dam in Brazil☆. **Environmental Pollution**, v. 272, 1 mar. 2021.

COLLAUD COEN, M. et al. Effects of the prewhitening method, the time granularity, and the time segmentation on the Mann–Kendall trend detection and the associated Sen's slope. **Atmospheric Measurement Techniques**, v. 13, n. 12, p. 6945–6964, 2020.

COSTA, D. DE A. et al. From Monitoring and Modeling to Management: How to Improve Water Quality in Brazilian Rivers? A Case Study: Piabanha River Watershed. **Water**, v. 13, n. 2, 2021.

COUTINHO, W.; VON SPERLING, M. **EMPREGO DA FLOTAÇÃO A AR DISSOLVIDO NO TRATAMENTO DE CURSOS D'ÁGUA: Avaliação de Desempenho da Estação de Tratamento dos Córregos Ressaca e Sarandi Afluentes à Represa da Pampulha**. Belo Horizonte: Universidade Federal de Minas Gerais, 2007.

CRESSON, R. et al. **Optical image gap filling using deep convolutional autoencoder from optical and radar images**. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. **Anais...2019**.

CUI, Y. et al. Deep learning-based remote sensing estimation of water transparency in shallow lakes by combining Landsat 8 and Sentinel 2 images. **Environmental Science and Pollution Research**, v. 29, p. 4401–4413, 2022.

DALL'OLMO, G.; GITELSON, A. A. Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. **Appl. Opt.**, v. 44, n. 3, p. 412–422, jan. 2005.

DE SOUZA JUNIOR, T. G. et al. Inferência sobre a perda histórica da qualidade da água de um açude urbano utilizando o sensoriamento remoto Revista Brasileira de Geografia Física. p. 3138–3155, 2021.

DEKKER, A. G.; VOS, R. J.; PETERS, S. W. M. Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data. **International Journal of Remote Sensing**, v. 23, n. 1, p. 15–35, 10 jan. 2002.

DIAS, R. L. S. et al. Machine learning models applied to TSS estimation in a reservoir using multispectral sensor onboard to RPA. **Ecological Informatics**, v. 65, 1 nov. 2021.

DÖRHNHÖFER, K. et al. Multi-sensor satellite and in situ monitoring of phytoplankton development in a eutrophic-mesotrophic lake. **Science of The Total Environment**, v. 612, p. 1200–1214, 2018.

DÖRHNHÖFER, K.; OPPELT, N. Remote sensing for lake research and monitoring - Recent advances. **Ecological Indicators**, v. 64, p. 105–122, 1 maio 2016.

ELSAYED, S. et al. Assessment of Water Quality in Lake Qaroun using Ground-Based Remote Sensing Data and Artificial Neural Networks. **Water (Switzerland)**, v. 13, n. 21, 1 nov. 2021.

ESA. **Sentinel-2 MSI Technical Guide**. Disponível em: <<https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi>>. Acesso em: 17 maio. 2022.

FALCONER, I. R. An Overview of problems caused by toxic blue–green algae (cyanobacteria) in drinking and recreational water. **Environmental Toxicology**, v. 14, n. 1, p. 5–12, 1 fev. 1999.

FREIRE, L. L.; COSTA, A. C.; LIMA NETO, I. E. Spatio-temporal Patterns of River Water Quality in the Semiarid Northeastern Brazil. **Water, Air, and Soil Pollution**, v. 232, n. 11, 1 nov. 2021.

FURTADO, A. P. F. V.; MONTE-MOR, R. C. DE A.; COUTO, E. DE A. DO. Evaluation of reduction of external load of total phosphorus and total suspended solids for rehabilitation of urban lakes. **Journal of Environmental Management**, v. 296, 15 out. 2021.

FURTADO, A. P. F. V.; MONTE-MOR, R. C. DE A.; DO COUTO, E. DE A. **AVALIAÇÃO DAS AÇÕES DE REABILITAÇÃO DA LAGOA DA PAMPULHA EM BELO HORIZONTE-MG**. XXIII SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS. **Anais...**2019.

GHOLIZADEH, M. H.; MELESSE, A. M.; REDDI, L. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. **Sensors (Switzerland)**, v. 16, n. 8, 16 ago. 2016.

GITELSON, A. A. et al. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. **Remote Sensing of Environment**, v. 112, n. 9, p. 3582–3593, 15 set. 2008.

GORDON, H. R. Removal of atmospheric effects from satellite imagery of the oceans. **Appl. Opt.**, v. 17, n. 10, p. 1631–1636, maio 1978.

GORDON, H. R. et al. A semianalytic radiance model of ocean color. **Journal of Geophysical Research: Atmospheres**, v. 93, n. D9, p. 10909–10924, 20 set. 1988.

GORDON, H. R. **Evolution of ocean color atmospheric correction: 1970–2005**. **Remote Sensing**MDPI, , 1 dez. 2021.

GORDON, H. R.; BROWN, O. B.; JACOBS, M. M. Computed Relationships Between the Inherent and Apparent Optical Properties of a Flat Homogeneous Ocean. **Appl. Opt.**, v. 14, n. 2, p. 417–427, fev. 1975.

GORDON, H. R.; WANG, M. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. **Applied optics**, v. 33, n. 3, p. 443–452, 1994.

GORELICK, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, p. 18–27, 1 dez. 2017.

GROOM, S. B. et al. Satellite Ocean Colour: Current Status and Future Perspective. **Frontiers in Marine Science**, v. 6, 2019.

GURLIN, D. **Near infrared-red models for the remote estimation of chlorophyll-a concentration in optically complex turbid productive waters: From in situ measurements to aerial imagery.** [s.l.] University of Nebraska, Lincoln, jan. 2012.

HAJIGHOLIZADEH, M. et al. Land–Lake Linkage and Remote Sensing Application in Water Quality Monitoring in Lake Okeechobee, Florida, USA. **Land**, v. 10, n. 2, p. 1–17, 1 fev. 2021.

HANSEN, C. H. et al. Reservoir water quality monitoring using remote sensing with seasonal models: case study of five central-Utah reservoirs. **Lake and Reservoir Management**, v. 31, n. 3, p. 225–240, 2015.

HANSEN, C. H. et al. Evaluating historical trends and influences of meteorological and seasonal climate conditions on lake chlorophyll a using remote sensing. **Lake and Reservoir Management**, v. 36, n. 1, p. 45–63, 2 jan. 2020.

HARRIS, C. R. et al. Array programming with NumPy. **Nature**, v. 585, n. 7825, p. 357–362, 2020.

HENNY, C.; MEUTIA, A. A. Urban Lakes in Megacity Jakarta: Risk and Management Plan for Future Sustainability. **Procedia Environmental Sciences**, v. 20, p. 737–746, 2014.

HIRSCH, R. M.; SLACK, J. R.; SMITH, R. A. Techniques of trend analysis for monthly water quality data. **Water Resources Research**, v. 18, n. 1, p. 107–121, 1982.

HOSSAIN, A. K. M. A.; MATHIAS, C.; BLANTON, R. Remote Sensing of Turbidity in the Tennessee River using Landsat 8 Satellite. **Remote Sensing**, v. 13, n. 3785, p. 1–24, 2 set. 2021.

HOYER, S.; HAMMAN, J. xarray: N-D labeled Arrays and Datasets in Python. **Journal of Open Research Software**, v. 5, n. 1, p. 10, 5 abr. 2017.

HUNTER, J. D. Matplotlib: A 2D graphics environment. **Computing in Science & Engineering**, v. 9, n. 3, p. 90–95, 2007.

HUSSAIN, MD.; MAHMUD, I. pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. **The Journal of Open Source Software**, v. 4, n. 39, p. 1556, jul. 2019.

IGAM. **Avaliação da Qualidade das Águas da Bacia da Lagoa da Pampulha**. Belo Horizonte: [s.n.].

JARDIM, C. H. et al. O CLIMA EM ÁREAS LIMÍTROFES AO PLANALTO MERIDIONAL DO ESPINHAÇO: BELO HORIZONTE, SETE LAGOAS E CONCEIÇÃO DO MATO DENTRO, MINAS GERAIS-BRASIL. **Revista Brasileira de Climatologia**, v. 25, p. 549–570, 2019.

KENDALL, M. G. **Rank correlation methods**. 3d ed. ed. London,: C. Griffin, 1948.

KIRK, JOHN. T. O. **Light and Photosynthesis in Aquatic Ecosystems**. 3rd. ed. Cambridge: Cambridge University Press, 2011.

KITCHENER, B. G. B.; WAINWRIGHT, J.; PARSONS, A. J. A review of the principles of turbidity measurement. **Progress in Physical Geography**, v. 41, n. 5, p. 620–642, 2017.

KRUSKAL, W. H.; WALLIS, W. A. Use of Ranks in One-Criterion Variance Analysis. **Journal of the American Statistical Association**, v. 47, n. 260, p. 583–621, 1952.

LEE, Z. et al. Secchi disk observation with spectral-selective glasses in blue and green waters. **Opt. Express**, v. 25, n. 17, p. 19878–19885, ago. 2017.

LEE, Z. P. et al. Secchi disk depth: A new theory and mechanistic model for underwater visibility. **Remote Sensing of Environment**, v. 169, p. 139–149, 1 nov. 2015.

LEE, Z. P.; CARDER, K. L.; ARNONE, R. A. Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters. **Applied Optics**, v. 41, n. 27, p. 5755–5772, 2002.

LIU, M. et al. Sentinel-2 and Landsat-8 Observations for Harmful Algae Blooms in a Small Eutrophic Lake. **Remote Sensing**, v. 13, n. 4479, p. 1–18, 1 nov. 2021.

LOBO, F. DE L. et al. Algaemap: Algae Bloom Monitoring Application for Inland Waters in Latin America. **Remote Sensing**, v. 13, n. 15, 1 ago. 2021.

LOPES, F. A. et al. Challenges for contact recreation in a tropical urban lake: assessment by a water quality index. **Environment, Development and Sustainability**, v. 22, n. 6, p. 5409–5423, 1 ago. 2020.

LORENZZETTI, J. A. **Princípios Físicos de Sensoriamento Remoto**. [s.l.] Editora Blucher, 2015.

MACIEL, D. et al. Retrieving Total and Inorganic Suspended Sediments in Amazon Floodplain Lakes: A Multisensor Approach. **Remote Sensing**, v. 11, n. 15, 1 ago. 2019.

MACIEL, D. A. et al. Water clarity in Brazilian water assessed using Sentinel-2 and machine learning methods. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 182, p. 134–152, 2021.

MANN, H. B. Nonparametric Tests Against Trend. **Econometrica**, v. 13, n. 3, p. 245–259, 1945.

MARTINS, V. S. et al. Assessment of Atmospheric Correction Methods for Sentinel-2 MSI Images Applied to Amazon Floodplain Lakes. **Remote Sensing**, v. 9, n. 4, 2017.

MATTHEWS, M. W. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. **International Journal of Remote Sensing**, v. 32, n. 21, p. 6855–6899, 2011.

MAZHAR, N. et al. Effects of climatic factors on the sedimentation trends of Tarbela Reservoir, Pakistan. **SN Applied Sciences**, v. 3, n. 1, p. 122, 2021.

MCKEE, T. B.; DOESKEN, N. J.; KLEIST, J. **THE RELATIONSHIP OF DROUGHT FREQUENCY AND DURATION TO TIME SCALE** Eighth Conference on Applied Climatology. [s.l.: s.n.].

MICROSOFT; ESRI. **Sentinel-2 10m Land Use/Land Cover Timeseries**. Disponível em:

<<https://www.arcgis.com/home/item.html?id=d3da5dd386d140cf93fc9ecbf8da5e31>>.

Acesso em: 27 maio. 2022.

MISHRA, S.; MISHRA, D. R. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. **Remote Sensing of Environment**, v. 117, p. 394–406, 15 fev. 2012.

MOBLEY, C. **Light and Water: Radiative Transfer in Natural Waters Inverse methods for estimating inherent optical properties from radiometric measurements View project**. 1st. ed. [s.l.] Ocean Optics, 1994.

MOBLEY, C. D. Radiative Transfer in the Ocean. Em: STEELE, J. H. (Ed.). **Encyclopedia of Ocean Sciences**. Oxford: Academic Press, 2001. p. 2321–2330.

MOREL, A. Bio-optical Models. Em: STEELE, J. H. (Ed.). **Encyclopedia of Ocean Sciences**. Oxford: Academic Press, 2001. p. 317–326.

MOREL, A.; GENTILI, B. Diffuse reflectance of oceanic waters: its dependence on Sun angle as influenced by the molecular scattering contribution. **Applied Optics**, v. 30, n. 30, p. 4427–4438, 1991.

MORTULA, M. et al. Towards monitoring of nutrient pollution in coastal lake using remote sensing and regression analysis. **Water (Switzerland)**, v. 12, n. 7, 1 jul. 2020.

MOSES, W. J. et al. Atmospheric Correction for Inland Waters. Em: **Bio-optical Modeling and Remote Sensing of Inland Waters**. [s.l.] Elsevier Inc., 2017. p. 69–100.

NECHAD, B.; RUDDICK, K. G.; PARK, Y. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. **Remote Sensing of Environment**, v. 114, n. 4, p. 854–866, 15 abr. 2010.

NGUYEN, T. H. et al. Quantifying the Information Content of a Water Quality Monitoring Network Using Principal Component Analysis: A case study of the Freiburger Mulde River Basin, Germany. **Water (Switzerland)**, v. 12, n. 2, 1 fev. 2020.

NIU, C. et al. Deep learning based regression for optically inactive inland water quality parameter estimation using airborne hyperspectral imagery. **Environmental Pollution**, v. 286, p. 117534, 2021.

NUNES, A. DE A.; PINTO, E. J. DE A.; BAPTISTA, M. B. Detection of trends for extreme events of precipitation in the Metropolitan Region of Belo Horizonte through statistical methods. **Revista Brasileira de Recursos Hídricos**, v. 23, 2018.

OBAID, A. A. et al. Assessing the utility of using current generation high-resolution satellites (Sentinel 2 and Landsat 8) to monitor large water supply dam in South Africa. **Remote Sensing Applications: Society and Environment**, v. 22, 1 abr. 2021.

ODERMATT, D. et al. Review of constituent retrieval in optically deep and complex waters from satellite imagery. **Remote Sensing of Environment**, v. 118, p. 116–126, 15 mar. 2012.

OGASHAWARA, I.; MISHRA, D. R.; GITELSON, A. A. Chapter 1 - Remote Sensing of Inland Waters: Background and Current State-of-the-Art. Em: MISHRA, D. R.; OGASHAWARA, I.; GITELSON, A. A. (Eds.). **Bio-optical Modeling and Remote Sensing of Inland Waters**. [s.l.] Elsevier, 2017. p. 1–24.

OLIVER, S.; CORBURN, J.; RIBEIRO, H. Challenges Regarding Water Quality of Eutrophic Reservoirs in Urban Landscapes: A Mapping Literature Review.

International Journal of Environmental Research and Public Health, v. 16, n. 1, 2019.

OLIVETTI, D. et al. Low-cost unmanned aerial multispectral imagery for siltation monitoring in reservoirs. **Remote Sensing**, v. 12, n. 11, 1 jun. 2020.

OLMANSON, L. G.; BAUER, M. E.; BREZONIK, P. L. A 20-year Landsat water clarity census of Minnesota's 10,000 lakes. **Remote Sensing of Environment**, v. 112, n. 11, p. 4086–4097, 15 nov. 2008.

ONDERKA, M.; PEKÁROVÁ, P. Retrieval of suspended particulate matter concentrations in the Danube River from Landsat ETM data. **Science of The Total Environment**, v. 397, n. 1, p. 238–243, 2008.

ORD, J. K.; GETIS, A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. **Geographical Analysis**, v. 27, n. 4, p. 286–306, 1995.

PAULINO, R. S. et al. Assessment of Adjacency Correction over Inland Waters Using Sentinel-2 MSI Images. **Remote Sensing**, v. 14, n. 8, p. 1829, 11 abr. 2022.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, n. 85, p. 2825–2830, 2011.

PEKEL, J.-F. et al. High-resolution mapping of global surface water and its long-term changes. **Nature**, v. 540, n. 7633, p. 418–422, 2016.

PEPPA, M.; VASILAKOS, C.; KAVROUDAKIS, D. Eutrophication Monitoring for Lake Pamvotis, Greece, Using Sentinel-2 Data. **ISPRS International Journal of Geo-Information**, v. 9, n. 3, 2020.

PINEDA-MENDOZA, R. M. et al. Seasonal changes in the bacterial community structure of three eutrophicated urban lakes in Mexico city, with emphasis on *Microcystis* spp. **Toxicon**, v. 179, p. 8–20, 2020.

PIZANI, F. M. C. et al. **Estimation of Water Quality in a Reservoir from Sentinel-2 MSI and Landsat-8 OLI Sensors**. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. **Anais...**Copernicus GmbH, 3 ago. 2020.

POPA, C. L. et al. Land use impact on the levels of fluorescent dissolved organic matter, phytoplankton and zooplankton in urban lakes. **Limnologica**, v. 99, p. 126062, 2023.

QUANG, N. H. et al. Spatiotemporal Variation of Turbidity Based on Landsat 8 OLI in Cam Ranh Bay and Thuy Trieu Lagoon, Vietnam. **Water**, v. 9, n. 8, 2017.

RADIANT EARTH FOUNDATION. **Remote Sensing Infographic**. [s.l: s.n.].

RADOSAVLJEVIC, J. et al. Salinization as a driver of eutrophication symptoms in an urban lake (Lake Wilcox, Ontario, Canada). **Science of The Total Environment**, v. 846, p. 157336, 2022.

REBACK, J. et al. pandas-dev/pandas: Pandas 1.4.2. 2 abr. 2022.

RIDDICK, C. A. L. et al. Spatial variability of absorption coefficients over a biogeochemical gradient in a large and optically complex shallow lake. **Journal of Geophysical Research: Oceans**, v. 120, n. 10, p. 7040–7066, 1 out. 2015.

SABERIOON, M. et al. Chlorophyll-a and total suspended solids retrieval and mapping using Sentinel-2A and machine learning for inland waters. **Ecological Indicators**, v. 113, 1 jun. 2020.

SAGAN, V. et al. **Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing**. **Earth-Science Reviews** Elsevier B.V., , 1 jun. 2020.

SCHUELER, T.; SIMPSON, J. Introduction: Why urban lakes are different. **Watershed Protection Techniques**, v. 3, n. 4, 2001.

SEIDEL, M. et al. Underwater Use of a Hyperspectral Camera to Estimate Optically Active Substances in the Water Column of Freshwater Lakes. **Remote Sensing**, v. 12, n. 11, 1 jun. 2020.

SEIDL, M. et al. Impact of urbanisation (trends) on runoff behaviour of Pampulha watersheds (Brazil). **Environmental Science and Pollution Research**, v. 27, n. 13, p. 14259–14270, 1 maio 2020.

SEN, P. K. Estimates of the Regression Coefficient Based on Kendall's Tau. **Journal of the American Statistical Association**, v. 63, n. 324, p. 1379–1389, 1 dez. 1968.

SIBANDA, M. et al. **Application of drone technologies in surface water resources monitoring and assessment: A systematic review of progress, challenges, and opportunities in the global south. DronesMDPI**, , 1 set. 2021.

SIEGEL, D. A. et al. Global distribution and dynamics of colored dissolved and detrital organic materials. **Journal of Geophysical Research: Oceans**, v. 107, n. 12, 15 dez. 2002.

SILVA, T. F. DAS G. et al. Modelagem da Lagoa da Pampulha: Uma ferramenta para avaliar o impacto da bacia hidrográfica na dinâmica do fitoplâncton. **Engenharia Sanitaria e Ambiental**, v. 21, n. 1, p. 95–108, 1 jan. 2016.

SILVA, T. F. G. et al. Impact of Urban Stormwater Runoff on Cyanobacteria Dynamics in a Tropical Urban Lake. **Water (Switzerland)**, v. 11, n. 5, 1 maio 2019.

SOUZA, A. P. D. et al. Application of remote sensing in environmental impact assessment: a case study of dam rupture in Brumadinho, Minas Gerais, Brazil. **Environmental Monitoring and Assessment**, v. 193, n. 9, 1 set. 2021.

SOUZA, M. E. V. **USO DO SENSORIAMENTO REMOTO PARA ESTIMAR A CONCENTRAÇÃO DE COMPONENTES OPTICAMENTE ATIVOS NA LAGOA DA PAMPULHA-MG**. Belo Horizonte: Universidade Federal de Minas Gerais, 2019.

SPERLING, M. VON; VERBYLA, M. E.; OLIVEIRA, S. M. A. C. **Assessment of Treatment Plant Performance and Water Quality Data A Guide for Students, Researchers and Practitioners.** [s.l.] IWA Publishing, 2020.

SUN, X. et al. Monitoring water quality using proximal remote sensing technology. **Science of the Total Environment**, v. 803, 10 jan. 2022.

TETA, R. et al. Monitoring Cyanobacterial Blooms during the COVID-19 Pandemic in Campania, Italy: The Case of Lake Avernus. **Toxins**, v. 13, n. 7, 1 jul. 2021.

THEIL, H. A Rank-Invariant Method of Linear and Polynomial Regression Analysis. Em: RAJ, B.; KOERTS, J. (Eds.). **Henri Theil's Contributions to Economics and Econometrics: Econometric Theory and Methodology.** Dordrecht: Springer Netherlands, 1992. p. 345–381.

TOPP, S. N. et al. Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications. **Water (Switzerland)**, v. 12, n. 1, 1 jan. 2020.

TORRES, I. C.; RESCK, R. P.; COELHO, R. M. P. Mass balance estimation of nitrogen , carbon , phosphorus and total suspended solids in the urban eutrophic , Pampulha reservoir , Brazil. **Acta Limnologica Brasiliensia**, v. 19, n. 1, p. 79–91, 2007.

TOTH, C.; JÓZKÓW, G. Remote sensing platforms and sensors: A survey. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 115, p. 22–36, 1 maio 2016.

TRINDADE, A. L. C. et al. Tendências temporais e espaciais da qualidade das águas superficiais da sub-bacia do Rio das Velhas, estado de Minas Gerais. **Engenharia Sanitaria e Ambiental**, v. 22, n. 1, p. 13–24, jan. 2017.

USGS. **Landsat 8 (L8) Data Users Handbook.** [s.l: s.n.].

VANHELLEMONT, Q.; RUDDICK, K. Advantages of high quality SWIR bands for ocean colour processing: Examples from Landsat-8. **Remote Sensing of Environment**, v. 161, p. 89–106, 2015.

VAROL, M. et al. Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. **Catena**, v. 92, p. 11–21, maio 2012.

VERMOTE, E. F. et al. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: an overview. **IEEE Transactions on Geoscience and Remote Sensing**, v. 35, n. 3, p. 675–686, 1997.

VIRTANEN, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, v. 17, n. 3, p. 261–272, 2020.

VISSER, P. M. et al. How rising CO₂ and global warming may stimulate harmful cyanobacterial blooms. **Harmful Algae**, v. 54, p. 145–159, 2016.

WAGLE, N.; ACHARYA, T. D.; LEE, D. H. Comprehensive review on application of machine learning algorithms for water quality parameter estimation using remote sensing data. **Sens. Mater**, v. 32, n. 11, p. 3879–3892, 2020.

WANG, L. et al. Mapping Freshwater Chlorophyll-a Concentrations at a Regional Scale Integrating Multi-Sensor Satellite Observations with Google Earth Engine. **Remote Sensing**, v. 12, n. 20, p. 1–18, 2 out. 2020.

WANG, L. et al. Small Water Body Detection and Water Quality Variations with Changing Human Activity Intensity in Wuhan. **Remote Sensing**, v. 14, n. 1, 1 jan. 2022.

WASKOM, M. seaborn: statistical data visualization. **Journal of Open Source Software**, v. 6, n. 60, p. 3021, 6 abr. 2021.

WATANABE, F. et al. Remote sensing of the chlorophyll-a based on OLI/Landsat-8 and MSI/Sentinel-2a (Barra Bonita Reservoir, Brazil). **Anais da Academia Brasileira de Ciências**, v. 90, n. 2, p. 1987–2000, 2018.

WATANABE, F. et al. Mapping the chlorophyll-a horizontal gradient in a cascading reservoirs system using MSI Sentinel-2A images. **Advances in Space Research**, v. 64, n. 3, p. 581–590, 2019.

WATANABE, F. S. Y. et al. Estimation of Chlorophyll-a Concentration and the Trophic State of the Barra Bonita Hydroelectric Reservoir Using OLI/Landsat-8 Images. **International Journal of Environmental Research and Public Health**, v. 12, n. 9, p. 10391–10417, 2015.

WEN, Z. et al. Remote sensing of total suspended matter concentration in lakes across China using Landsat images and Google Earth Engine. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 187, p. 61–78, 2022.

WHITLOCK, C. H. et al. Comparison of reflectance with backscatter and absorption parameters for turbid waters. **Applied Optics**, v. 20, n. 3, p. 517–522, 1981.

XU, X. et al. Long-Term Changes in Water Clarity in Lake Liangzi Determined by Remote Sensing. **Remote Sensing**, v. 10, n. 9, 2018.

XUE, Y. et al. Nitrate Pollution and Preliminary Source Identification of Surface Water in a Semi-Arid River Basin, Using Isotopic and Hydrochemical Approaches. **Water**, v. 8, n. 8, 2016.

YANG, L. et al. Towards Synoptic Water Monitoring Systems: A Review of AI Methods for Automating Water Body Detection and Water Quality Monitoring Using Remote Sensing. **Sensors**, v. 22, n. 6, 1 mar. 2022.

YIN, F.; LEWIS, P. E.; GÓMEZ-DANS, J. L. Bayesian atmospheric correction over land: Sentinel-2/MSI and Landsat 8/OLI. **EGUsphere**, v. 2022, p. 1–62, 2022.

YUNUS, A. P.; DOU, J.; SRAVANTHI, N. Remote sensing of chlorophyll-a as a measure of red tide in Tokyo Bay using hotspot analysis. **Remote Sensing Applications: Society and Environment**, v. 2, p. 11–25, 1 dez. 2015.

YUNUS, A. P.; MASAGO, Y.; HIJIOKA, Y. COVID-19 and surface water quality: Improved lake water quality during the lockdown. **Science of the Total Environment**, v. 731, 20 ago. 2020.

ZABALETA, B.; ACHKAR, M.; AUBRIOT, L. Hotspot analysis of spatial distribution of algae blooms in small and medium water bodies. **Environmental Monitoring and Assessment**, v. 193, n. 4, 1 abr. 2021.

ZHANG, Y. et al. Remote sensing estimation of water clarity for various lakes in China. **Water Research**, v. 192, 15 mar. 2021a.

ZHANG, Y. et al. Water clarity response to climate warming and wetting of the Inner Mongolia-Xinjiang Plateau: A remote sensing approach. **Science of the Total Environment**, v. 796, 20 nov. 2021b.

ZHANG, Z. et al. Urban lake spatial openness and relationship with neighboring land prices: Exploratory geovisual analytics for essential policy insights. **Land Use Policy**, v. 92, p. 104479, 2020.

ZHAO, J. et al. Remote sensing evaluation of total suspended solids dynamic with markov model: A case study of inland reservoir across administrative boundary in south china. **Sensors (Switzerland)**, v. 20, n. 23, p. 1–29, 1 dez. 2020.

ZHU, W. et al. Landsat 8-observed water quality and its coupled environmental factors for urban scenery lakes: A case study of West Lake. **Water Environment Research**, v. 92, n. 2, p. 255–265, 1 fev. 2020.

APÊNDICE A – Colunas do banco de dados prévio à modelagem para o Landsat-8

Índice	Nome da Coluna	Valores não-nulos	Formato do dado
1	Ponto	49	object
2	img_id	49	object
3	satellite	49	object
4	date	49	datetime64[ns]
5	timestamp	49	datetime64[ns]
6	l8_blue	49	float64
7	l8_green	49	float64
8	l8_red	49	float64
9	l8_nir	49	float64
10	l8_swir1	49	float64
11	l8_swir2	49	float64
12	Data Amostragem	49	datetime64[ns]
13	Hora Amostragem	16	object
14	Fonte	49	object
15	Latitude	49	float64
16	Longitude	49	float64
17	TSM (mg/L)	45	float64
18	Chla (µg/L)	49	float64
19	SDD (m)	23	float64
20	Turbidez (NTU)	16	float64
21	Data_inferior	49	datetime64[ns]
22	Data_superior	49	datetime64[ns]
23	Estacao do ano	49	object
24	Periodo_Chuvoso	49	uint8
25	Periodo_Seco	49	uint8
26	Rede_Antes da rede	49	uint8
27	Rede_Depois da rede	49	uint8
28	l8_blue x l8_green	49	float64
29	l8_blue x l8_red	49	float64
30	l8_blue x l8_nir	49	float64
31	l8_blue x l8_swir1	49	float64
32	l8_blue x l8_swir2	49	float64
33	l8_green x l8_red	49	float64
34	l8_green x l8_nir	49	float64
35	l8_green x l8_swir1	49	float64
36	l8_green x l8_swir2	49	float64
37	l8_red x l8_nir	49	float64
38	l8_red x l8_swir1	49	float64
39	l8_red x l8_swir2	49	float64
40	l8_nir x l8_swir1	49	float64
41	l8_nir x l8_swir2	49	float64

42	$l8_swir1 \times l8_swir2$	49	float64
43	$l8_blue / l8_green$	49	float64
44	$l8_blue / l8_red$	49	float64
45	$l8_blue / l8_nir$	49	float64
46	$l8_blue / l8_swir1$	49	float64
47	$l8_blue / l8_swir2$	49	float64
48	$l8_green / l8_blue$	49	float64
49	$l8_green / l8_red$	49	float64
50	$l8_green / l8_nir$	49	float64
51	$l8_green / l8_swir1$	49	float64
52	$l8_green / l8_swir2$	49	float64
53	$l8_red / l8_blue$	49	float64
54	$l8_red / l8_green$	49	float64
55	$l8_red / l8_nir$	49	float64
56	$l8_red / l8_swir1$	49	float64
57	$l8_red / l8_swir2$	49	float64
58	$l8_nir / l8_blue$	49	float64
59	$l8_nir / l8_green$	49	float64
60	$l8_nir / l8_red$	49	float64
61	$l8_nir / l8_swir1$	49	float64
62	$l8_nir / l8_swir2$	49	float64
63	$l8_swir1 / l8_blue$	49	float64
64	$l8_swir1 / l8_green$	49	float64
65	$l8_swir1 / l8_red$	49	float64
66	$l8_swir1 / l8_nir$	49	float64
67	$l8_swir1 / l8_swir2$	49	float64
68	$l8_swir2 / l8_blue$	49	float64
69	$l8_swir2 / l8_green$	49	float64
70	$l8_swir2 / l8_red$	49	float64
71	$l8_swir2 / l8_nir$	49	float64
72	$l8_swir2 / l8_swir1$	49	float64
73	$l8_blue - l8_green / l8_blue + l8_green$	49	float64
74	$l8_blue - l8_red / l8_blue + l8_red$	49	float64
75	$l8_blue - l8_nir / l8_blue + l8_nir$	49	float64
76	$l8_blue - l8_swir1 / l8_blue + l8_swir1$	49	float64
77	$l8_blue - l8_swir2 / l8_blue + l8_swir2$	49	float64
78	$l8_green - l8_red / l8_green + l8_red$	49	float64
79	$l8_green - l8_nir / l8_green + l8_nir$	49	float64
80	$l8_green - l8_swir1 / l8_green + l8_swir1$	49	float64
81	$l8_green - l8_swir2 / l8_green + l8_swir2$	49	float64
82	$l8_red - l8_nir / l8_red + l8_nir$	49	float64
83	$l8_red - l8_swir1 / l8_red + l8_swir1$	49	float64
84	$l8_red - l8_swir2 / l8_red + l8_swir2$	49	float64
85	$l8_nir - l8_swir1 / l8_nir + l8_swir1$	49	float64
86	$l8_nir - l8_swir2 / l8_nir + l8_swir2$	49	float64
87	$l8_swir1 - l8_swir2 / l8_swir1 + l8_swir2$	49	float64

88	ndci	49	float64
89	2band	49	float64
90	3band	49	float64
91	nechad	49	float64

APÊNDICE B – Colunas do banco de dados prévio à modelagem para o Sentinel-2

Índice	Nome da Coluna	Valores não-nulos	Formato do dado
1	Ponto	54	object
2	img_id	54	object
3	satellite	54	object
4	date	54	datetime64[ns]
5	timestamp	54	datetime64[ns]
6	s2_blue	54	float64
7	s2_green	54	float64
8	s2_red	54	float64
9	s2_re1	54	float64
10	s2_re2	54	float64
11	s2_re3	54	float64
12	s2_nir	54	float64
13	s2_re4	54	float64
14	s2_swir1	54	float64
15	s2_swir2	54	float64
16	Data Amostragem	54	datetime64[ns]
17	Hora Amostragem	9	object
18	Fonte	54	object
19	Latitude	54	float64
20	Longitude	54	float64
21	TSM (mg/L)	53	float64
22	Chla (µg/L)	54	float64
23	SDD (m)	33	float64
24	Turbidez (NTU)	9	float64
25	Data_inferior	54	datetime64[ns]
26	Data_superior	54	datetime64[ns]
27	Estacao do ano	54	object
28	Periodo_Chuvoso	54	uint8
29	Periodo_Seco	54	uint8
30	Rede_Antes da rede	54	uint8
31	Rede_Deapos da rede	54	uint8
32	s2_blue x s2_green	54	float64
33	s2_blue x s2_red	54	float64
34	s2_blue x s2_re1	54	float64
35	s2_blue x s2_re2	54	float64
36	s2_blue x s2_re3	54	float64
37	s2_blue x s2_nir	54	float64
38	s2_blue x s2_re4	54	float64
39	s2_blue x s2_swir1	54	float64
40	s2_blue x s2_swir2	54	float64
41	s2_green x s2_red	54	float64

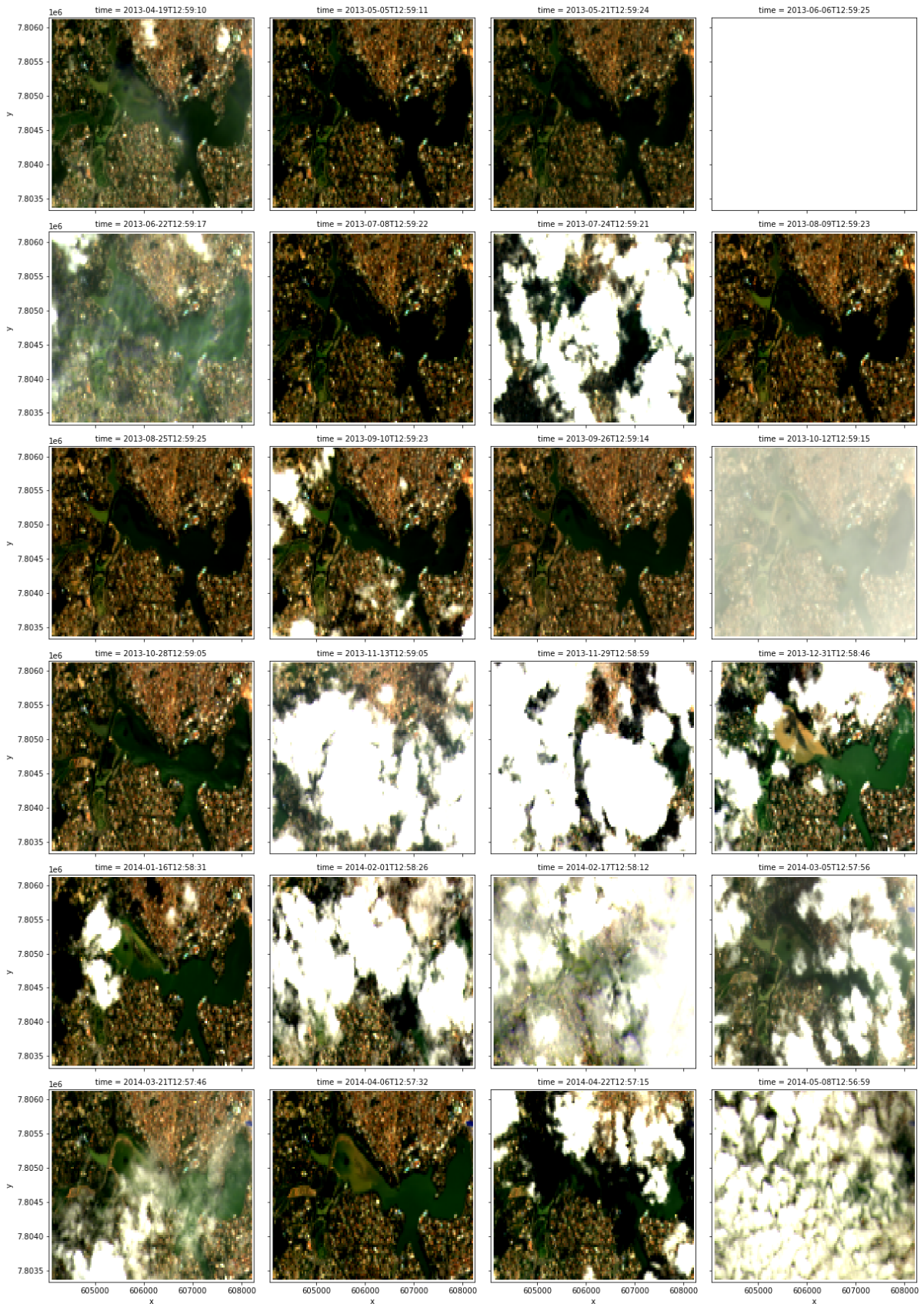
42	s2_green x s2_re1	54	float64
43	s2_green x s2_re2	54	float64
44	s2_green x s2_re3	54	float64
45	s2_green x s2_nir	54	float64
46	s2_green x s2_re4	54	float64
47	s2_green x s2_swir1	54	float64
48	s2_green x s2_swir2	54	float64
49	s2_red x s2_re1	54	float64
50	s2_red x s2_re2	54	float64
51	s2_red x s2_re3	54	float64
52	s2_red x s2_nir	54	float64
53	s2_red x s2_re4	54	float64
54	s2_red x s2_swir1	54	float64
55	s2_red x s2_swir2	54	float64
56	s2_re1 x s2_re2	54	float64
57	s2_re1 x s2_re3	54	float64
58	s2_re1 x s2_nir	54	float64
59	s2_re1 x s2_re4	54	float64
60	s2_re1 x s2_swir1	54	float64
61	s2_re1 x s2_swir2	54	float64
62	s2_re2 x s2_re3	54	float64
63	s2_re2 x s2_nir	54	float64
64	s2_re2 x s2_re4	54	float64
65	s2_re2 x s2_swir1	54	float64
66	s2_re2 x s2_swir2	54	float64
67	s2_re3 x s2_nir	54	float64
68	s2_re3 x s2_re4	54	float64
69	s2_re3 x s2_swir1	54	float64
70	s2_re3 x s2_swir2	54	float64
71	s2_nir x s2_re4	54	float64
72	s2_nir x s2_swir1	54	float64
73	s2_nir x s2_swir2	54	float64
74	s2_re4 x s2_swir1	54	float64
75	s2_re4 x s2_swir2	54	float64
76	s2_swir1 x s2_swir2	54	float64
77	s2_blue / s2_green	54	float64
78	s2_blue / s2_red	54	float64
79	s2_blue / s2_re1	54	float64
80	s2_blue / s2_re2	54	float64
81	s2_blue / s2_re3	54	float64
82	s2_blue / s2_nir	54	float64
83	s2_blue / s2_re4	54	float64
84	s2_blue / s2_swir1	54	float64
85	s2_blue / s2_swir2	54	float64
86	s2_green / s2_blue	54	float64
87	s2_green / s2_red	54	float64

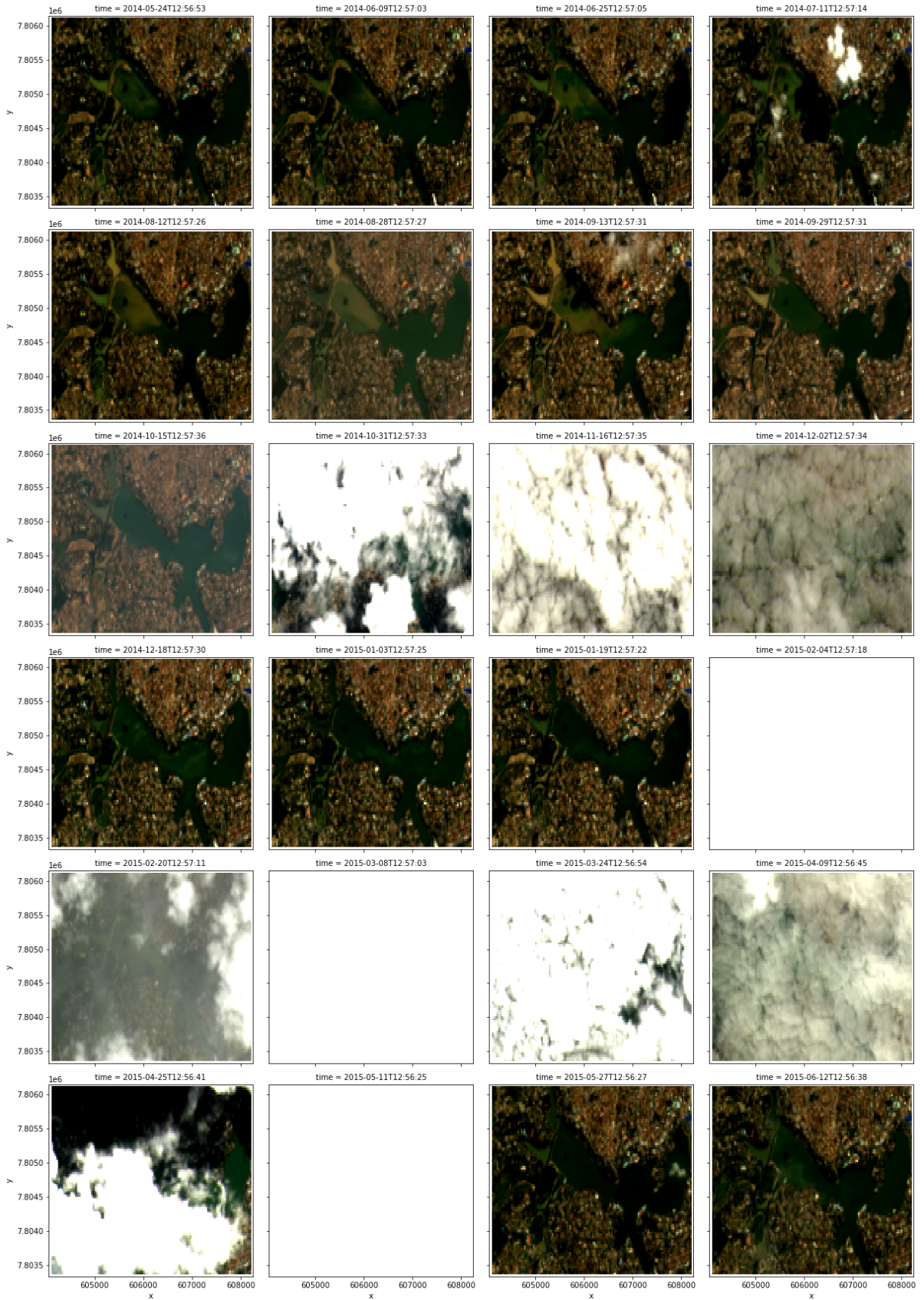
88	s2_green / s2_re1	54	float64
89	s2_green / s2_re2	54	float64
90	s2_green / s2_re3	54	float64
91	s2_green / s2_nir	54	float64
92	s2_green / s2_re4	54	float64
93	s2_green / s2_swir1	54	float64
94	s2_green / s2_swir2	54	float64
95	s2_red / s2_blue	54	float64
96	s2_red / s2_green	54	float64
97	s2_red / s2_re1	54	float64
98	s2_red / s2_re2	54	float64
99	s2_red / s2_re3	54	float64
100	s2_red / s2_nir	54	float64
101	s2_red / s2_re4	54	float64
102	s2_red / s2_swir1	54	float64
103	s2_red / s2_swir2	54	float64
104	s2_re1 / s2_blue	54	float64
105	s2_re1 / s2_green	54	float64
106	s2_re1 / s2_red	54	float64
107	s2_re1 / s2_re2	54	float64
108	s2_re1 / s2_re3	54	float64
109	s2_re1 / s2_nir	54	float64
110	s2_re1 / s2_re4	54	float64
111	s2_re1 / s2_swir1	54	float64
112	s2_re1 / s2_swir2	54	float64
113	s2_re2 / s2_blue	54	float64
114	s2_re2 / s2_green	54	float64
115	s2_re2 / s2_red	54	float64
116	s2_re2 / s2_re1	54	float64
117	s2_re2 / s2_re3	54	float64
118	s2_re2 / s2_nir	54	float64
119	s2_re2 / s2_re4	54	float64
120	s2_re2 / s2_swir1	54	float64
121	s2_re2 / s2_swir2	54	float64
122	s2_re3 / s2_blue	54	float64
123	s2_re3 / s2_green	54	float64
124	s2_re3 / s2_red	54	float64
125	s2_re3 / s2_re1	54	float64
126	s2_re3 / s2_re2	54	float64
127	s2_re3 / s2_nir	54	float64
128	s2_re3 / s2_re4	54	float64
129	s2_re3 / s2_swir1	54	float64
130	s2_re3 / s2_swir2	54	float64
131	s2_nir / s2_blue	54	float64
132	s2_nir / s2_green	54	float64
133	s2_nir / s2_red	54	float64

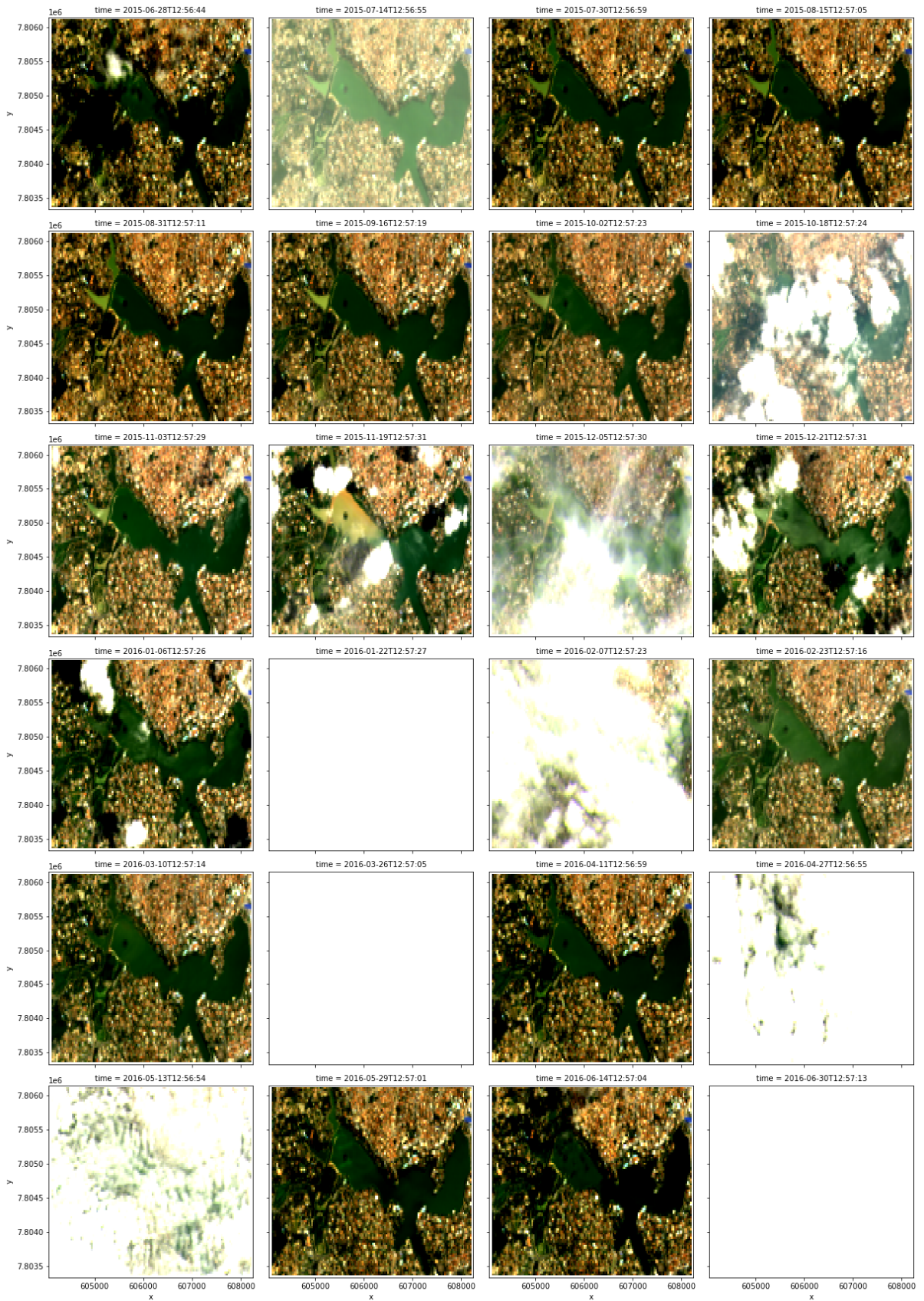
134	s2_nir / s2_re1	54	float64
135	s2_nir / s2_re2	54	float64
136	s2_nir / s2_re3	54	float64
137	s2_nir / s2_re4	54	float64
138	s2_nir / s2_swir1	54	float64
139	s2_nir / s2_swir2	54	float64
140	s2_re4 / s2_blue	54	float64
141	s2_re4 / s2_green	54	float64
142	s2_re4 / s2_red	54	float64
143	s2_re4 / s2_re1	54	float64
144	s2_re4 / s2_re2	54	float64
145	s2_re4 / s2_re3	54	float64
146	s2_re4 / s2_nir	54	float64
147	s2_re4 / s2_swir1	54	float64
148	s2_re4 / s2_swir2	54	float64
149	s2_swir1 / s2_blue	54	float64
150	s2_swir1 / s2_green	54	float64
151	s2_swir1 / s2_red	54	float64
152	s2_swir1 / s2_re1	54	float64
153	s2_swir1 / s2_re2	54	float64
154	s2_swir1 / s2_re3	54	float64
155	s2_swir1 / s2_nir	54	float64
156	s2_swir1 / s2_re4	54	float64
157	s2_swir1 / s2_swir2	54	float64
158	s2_swir2 / s2_blue	54	float64
159	s2_swir2 / s2_green	54	float64
160	s2_swir2 / s2_red	54	float64
161	s2_swir2 / s2_re1	54	float64
162	s2_swir2 / s2_re2	54	float64
163	s2_swir2 / s2_re3	54	float64
164	s2_swir2 / s2_nir	54	float64
165	s2_swir2 / s2_re4	54	float64
166	s2_swir2 / s2_swir1	54	float64
167	s2_blue - s2_green / s2_blue + s2_green	54	float64
168	s2_blue - s2_red / s2_blue + s2_red	54	float64
169	s2_blue - s2_re1 / s2_blue + s2_re1	54	float64
170	s2_blue - s2_re2 / s2_blue + s2_re2	54	float64
171	s2_blue - s2_re3 / s2_blue + s2_re3	54	float64
172	s2_blue - s2_nir / s2_blue + s2_nir	54	float64
173	s2_blue - s2_re4 / s2_blue + s2_re4	54	float64
174	s2_blue - s2_swir1 / s2_blue + s2_swir1	54	float64
175	s2_blue - s2_swir2 / s2_blue + s2_swir2	54	float64
176	s2_green - s2_red / s2_green + s2_red	54	float64
177	s2_green - s2_re1 / s2_green + s2_re1	54	float64
178	s2_green - s2_re2 / s2_green + s2_re2	54	float64
179	s2_green - s2_re3 / s2_green + s2_re3	54	float64

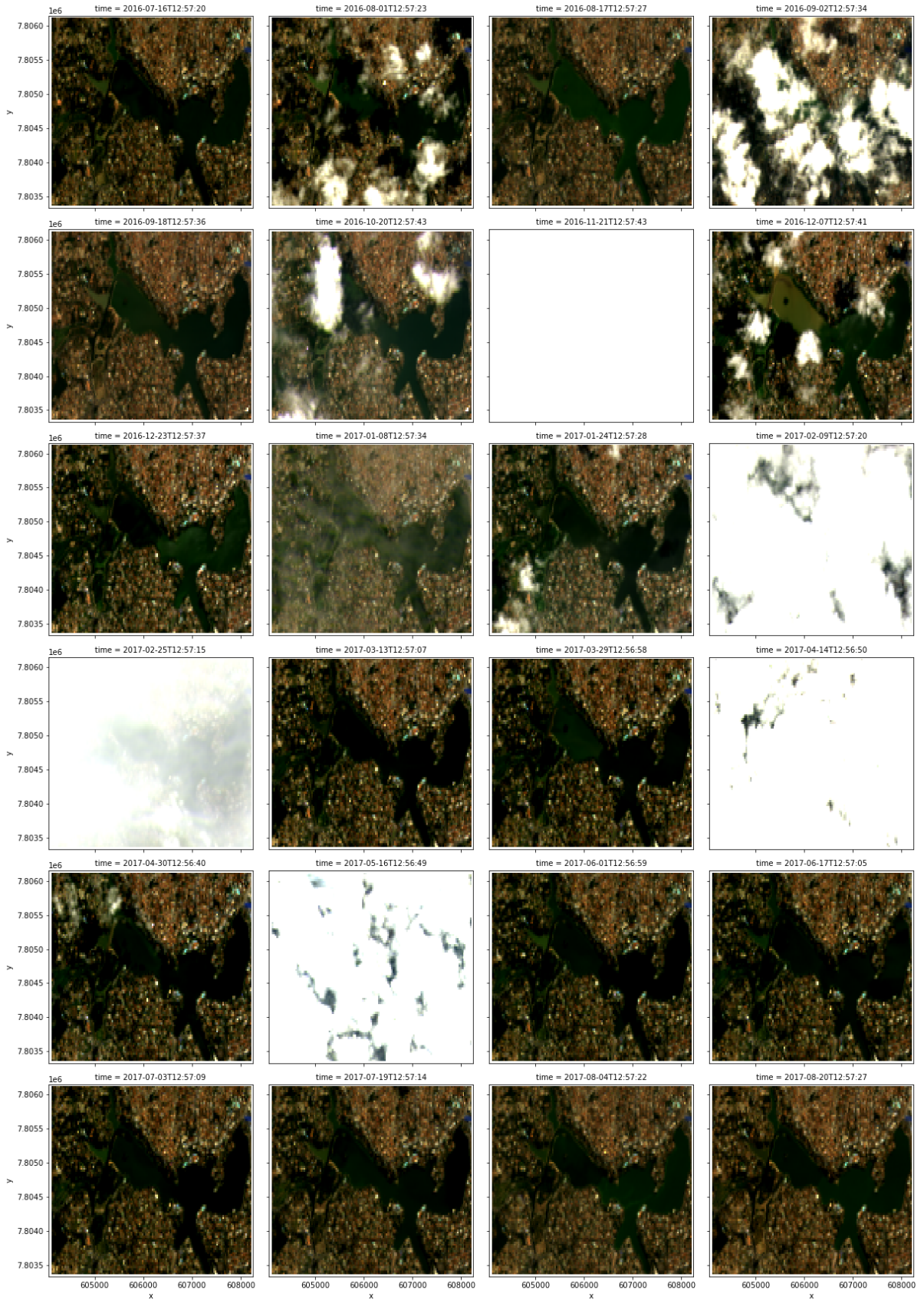
180	$s2_green - s2_nir / s2_green + s2_nir$	54	float64
181	$s2_green - s2_re4 / s2_green + s2_re4$	54	float64
182	$s2_green - s2_swir1 / s2_green + s2_swir1$	54	float64
183	$s2_green - s2_swir2 / s2_green + s2_swir2$	54	float64
184	$s2_red - s2_re1 / s2_red + s2_re1$	54	float64
185	$s2_red - s2_re2 / s2_red + s2_re2$	54	float64
186	$s2_red - s2_re3 / s2_red + s2_re3$	54	float64
187	$s2_red - s2_nir / s2_red + s2_nir$	54	float64
188	$s2_red - s2_re4 / s2_red + s2_re4$	54	float64
189	$s2_red - s2_swir1 / s2_red + s2_swir1$	54	float64
190	$s2_red - s2_swir2 / s2_red + s2_swir2$	54	float64
191	$s2_re1 - s2_re2 / s2_re1 + s2_re2$	54	float64
192	$s2_re1 - s2_re3 / s2_re1 + s2_re3$	54	float64
193	$s2_re1 - s2_nir / s2_re1 + s2_nir$	54	float64
194	$s2_re1 - s2_re4 / s2_re1 + s2_re4$	54	float64
195	$s2_re1 - s2_swir1 / s2_re1 + s2_swir1$	54	float64
196	$s2_re1 - s2_swir2 / s2_re1 + s2_swir2$	54	float64
197	$s2_re2 - s2_re3 / s2_re2 + s2_re3$	54	float64
198	$s2_re2 - s2_nir / s2_re2 + s2_nir$	54	float64
199	$s2_re2 - s2_re4 / s2_re2 + s2_re4$	54	float64
200	$s2_re2 - s2_swir1 / s2_re2 + s2_swir1$	54	float64
201	$s2_re2 - s2_swir2 / s2_re2 + s2_swir2$	54	float64
202	$s2_re3 - s2_nir / s2_re3 + s2_nir$	54	float64
203	$s2_re3 - s2_re4 / s2_re3 + s2_re4$	54	float64
204	$s2_re3 - s2_swir1 / s2_re3 + s2_swir1$	54	float64
205	$s2_re3 - s2_swir2 / s2_re3 + s2_swir2$	54	float64
206	$s2_nir - s2_re4 / s2_nir + s2_re4$	54	float64
207	$s2_nir - s2_swir1 / s2_nir + s2_swir1$	54	float64
208	$s2_nir - s2_swir2 / s2_nir + s2_swir2$	54	float64
209	$s2_re4 - s2_swir1 / s2_re4 + s2_swir1$	54	float64
210	$s2_re4 - s2_swir2 / s2_re4 + s2_swir2$	54	float64
211	$s2_swir1 - s2_swir2 / s2_swir1 + s2_swir2$	54	float64
212	ndci	54	float64
213	2band	54	float64
214	3band	54	float64
215	nechad	54	float64

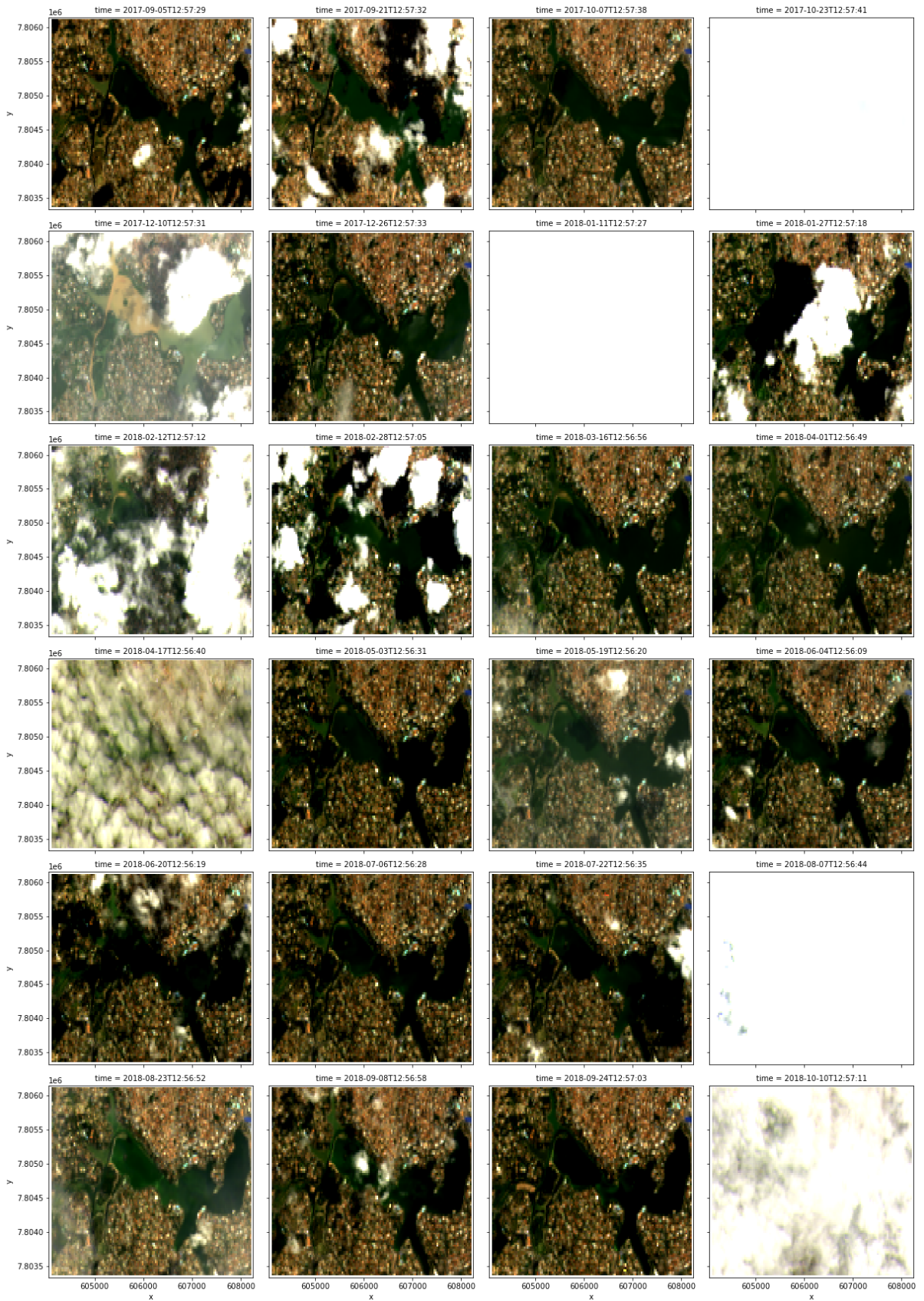
APÊNDICE C – Todas as imagens entre 2013 e 2022-05-01 para a região da Pampulha para o Landsat-8

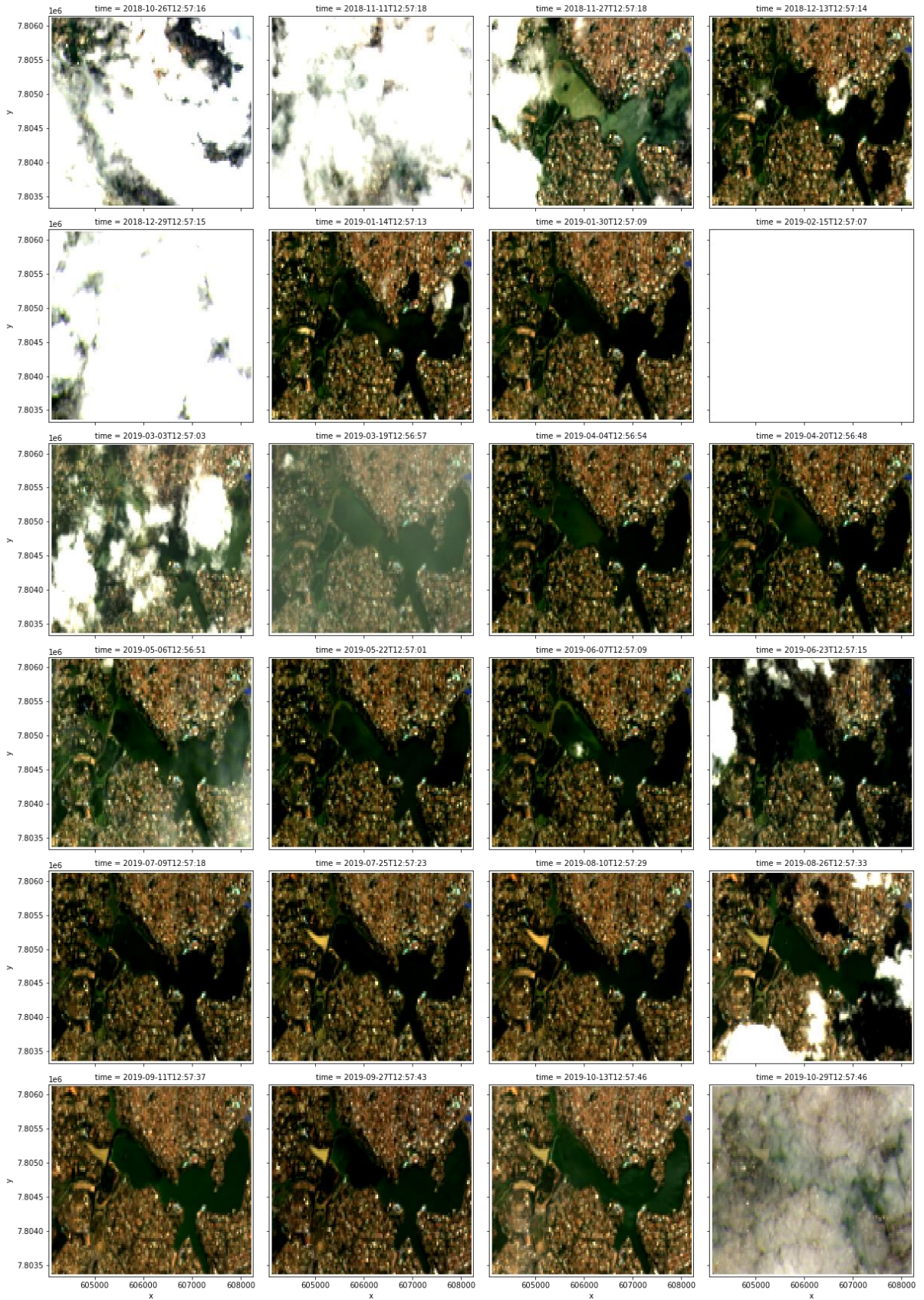


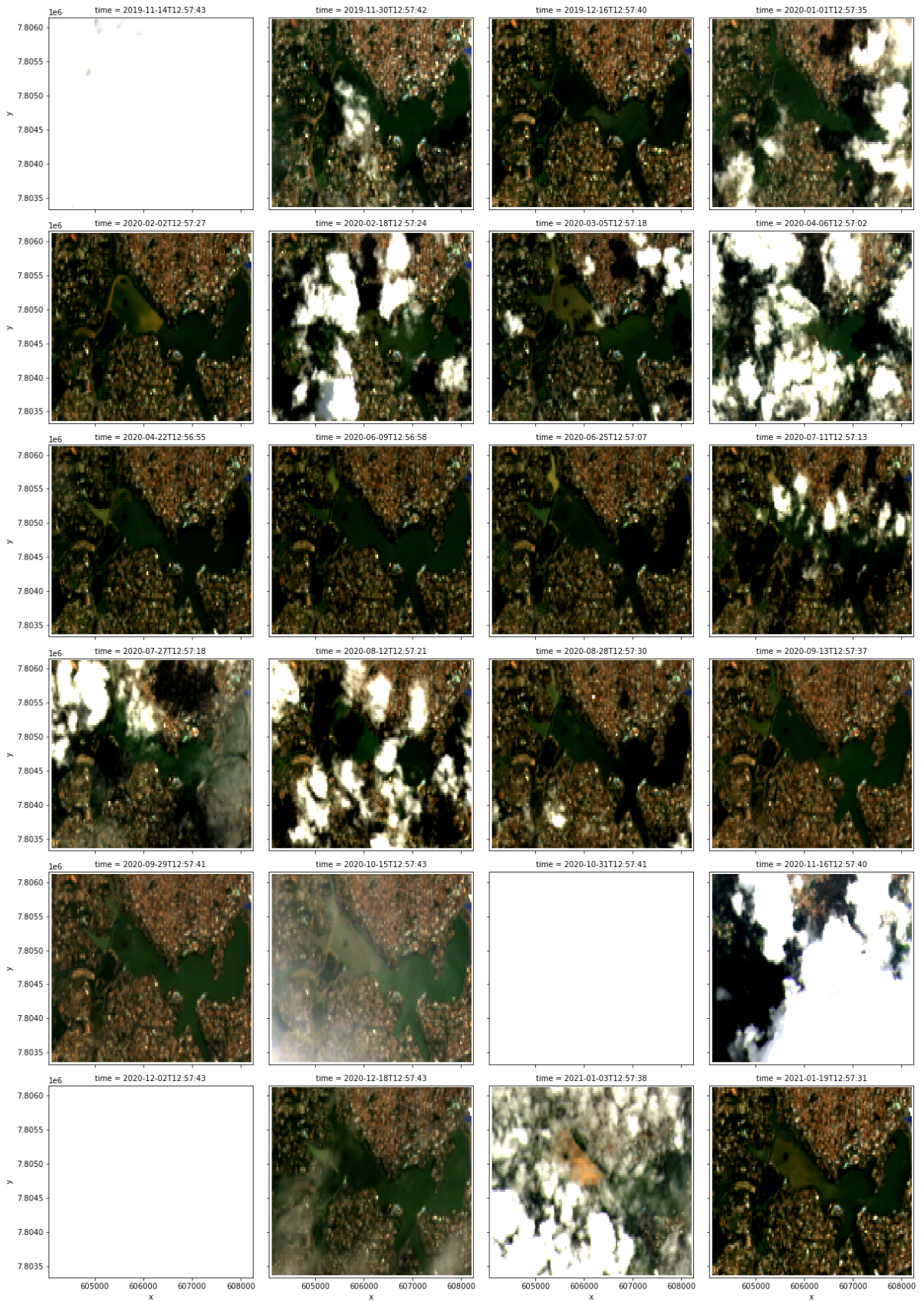


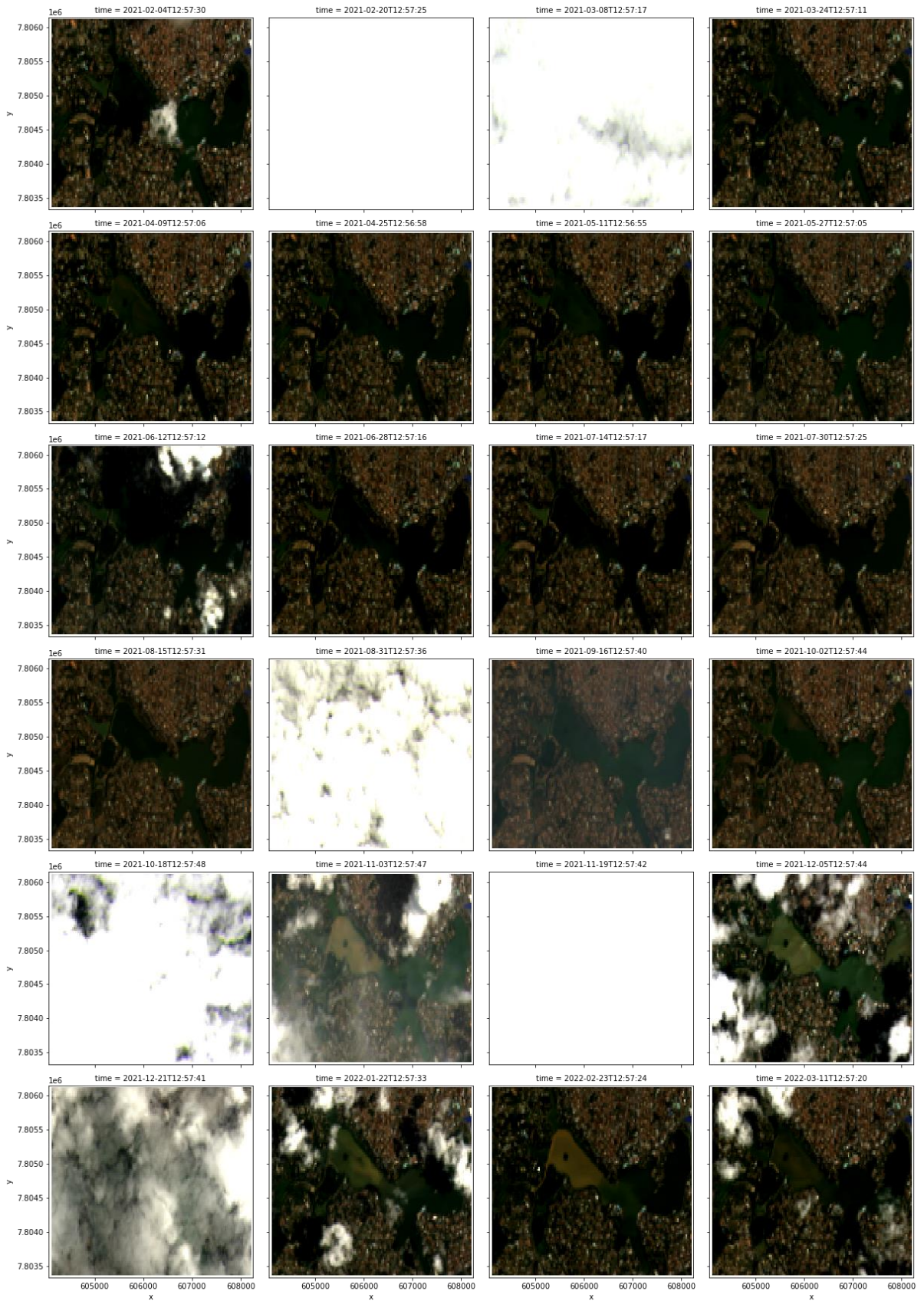


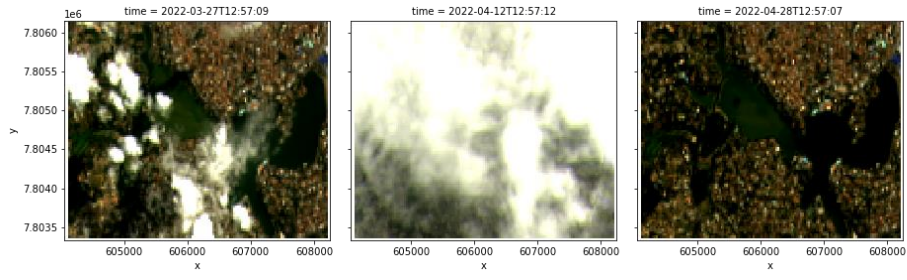




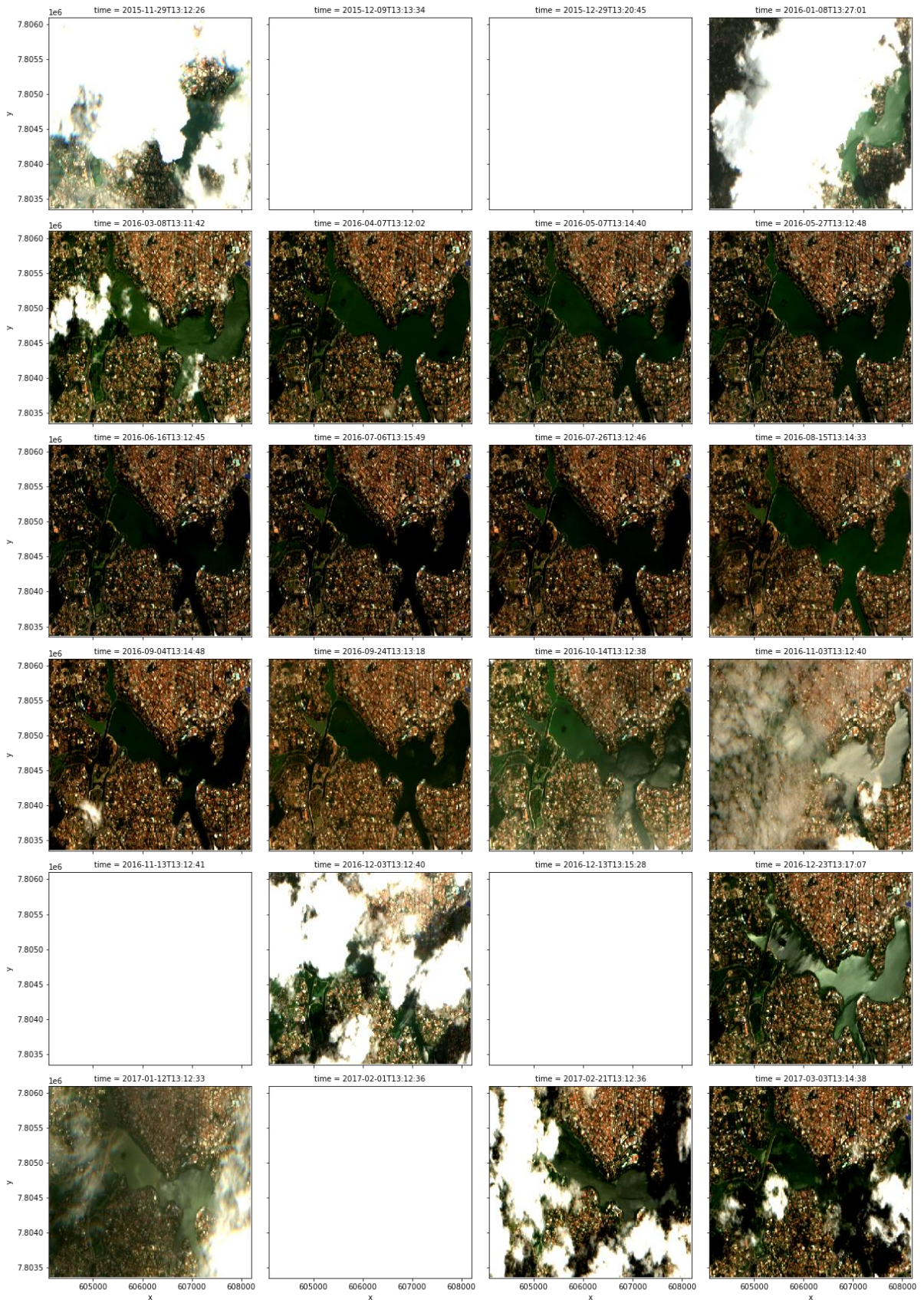


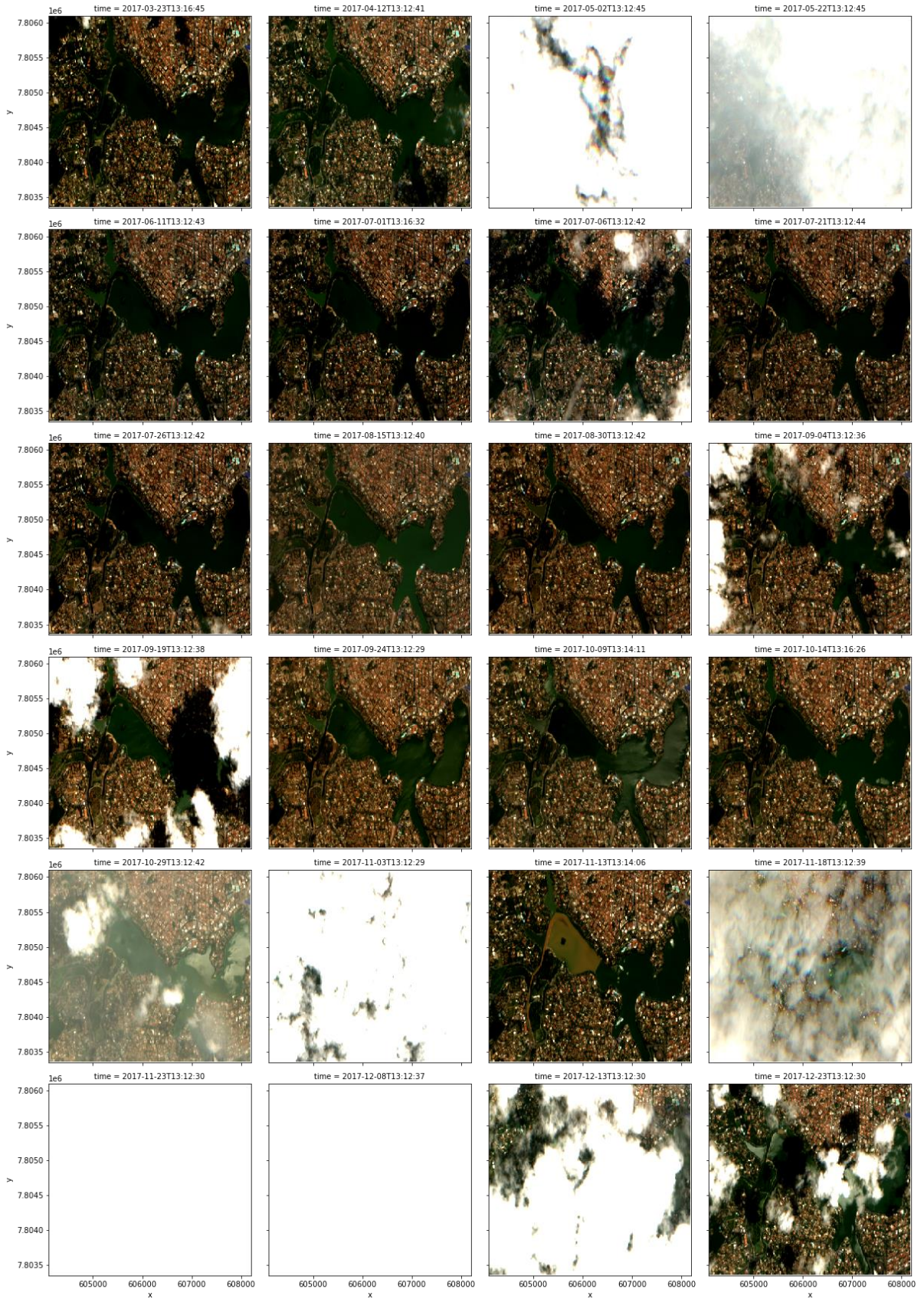


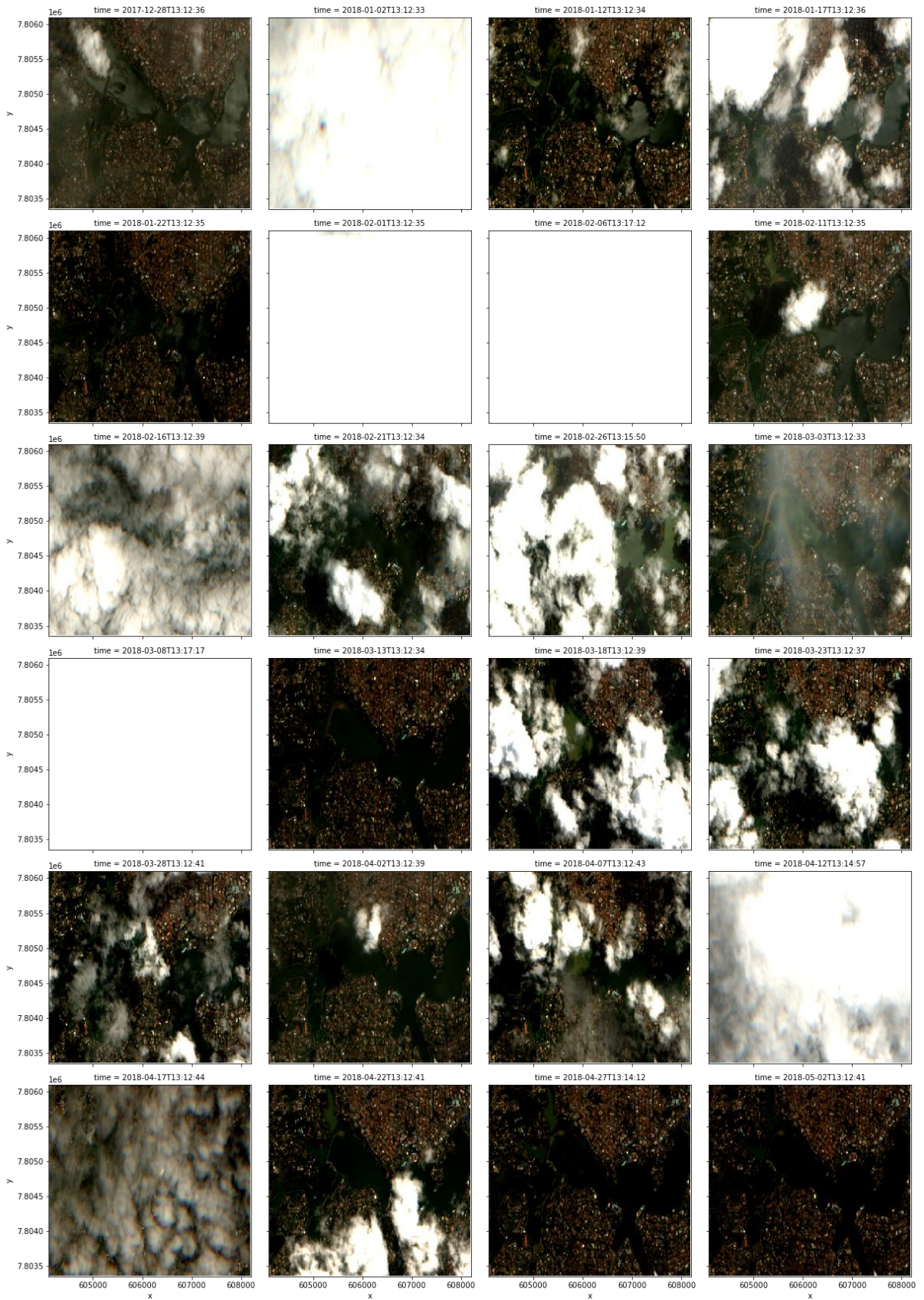


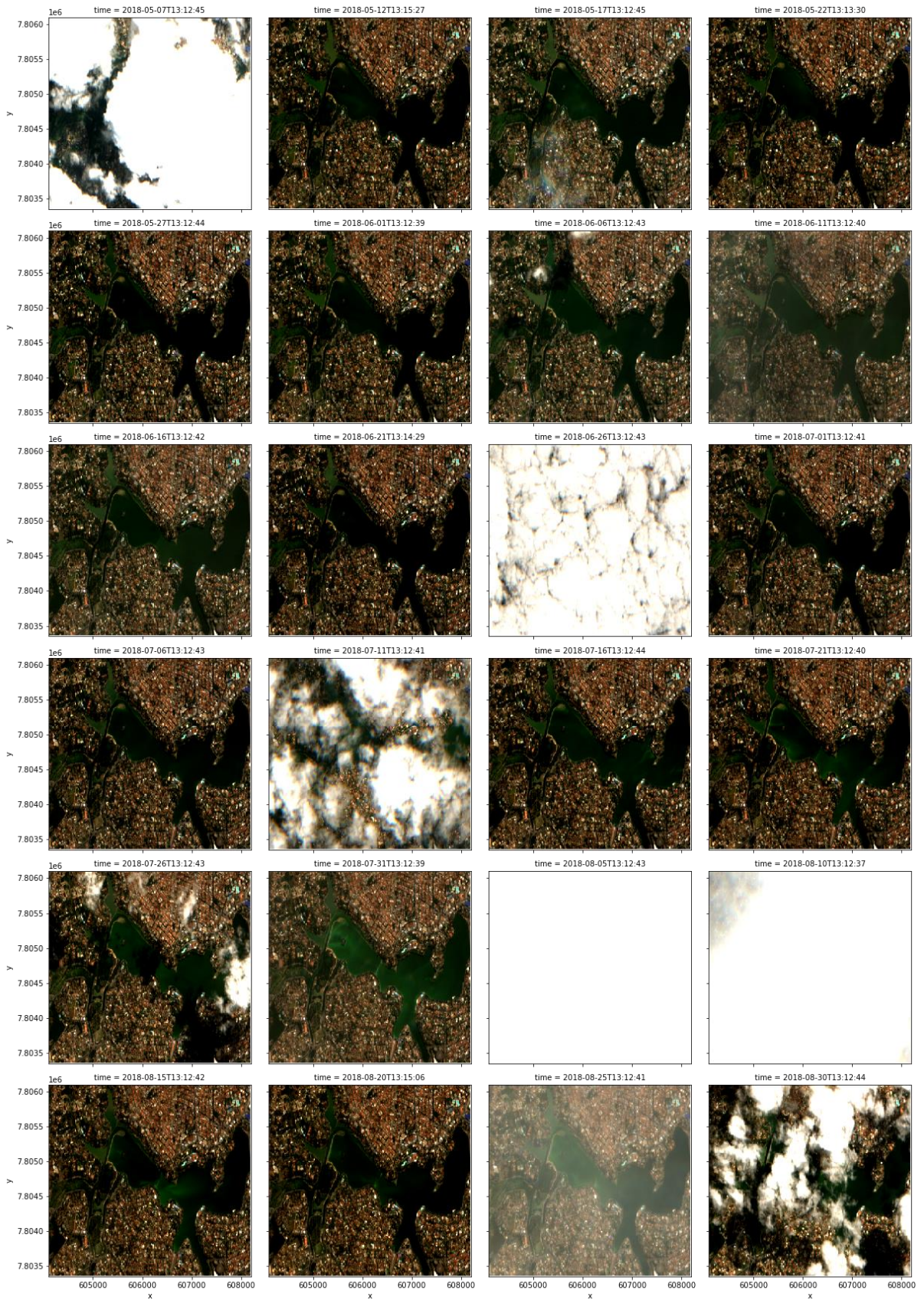


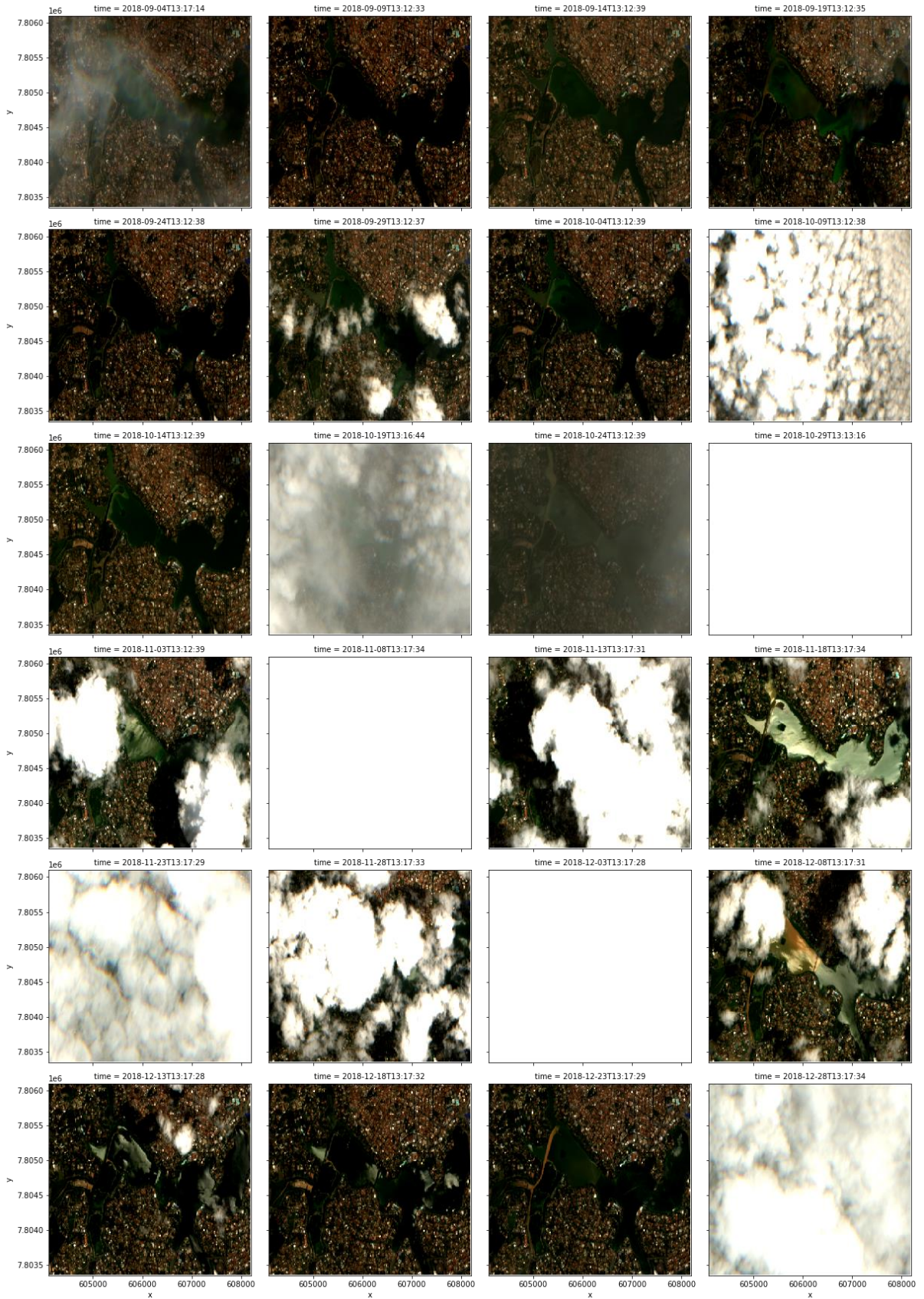
APÊNDICE D – Todas as imagens entre 2016 e 2022-05-01 para a região da Pampulha para o Sentinel-2

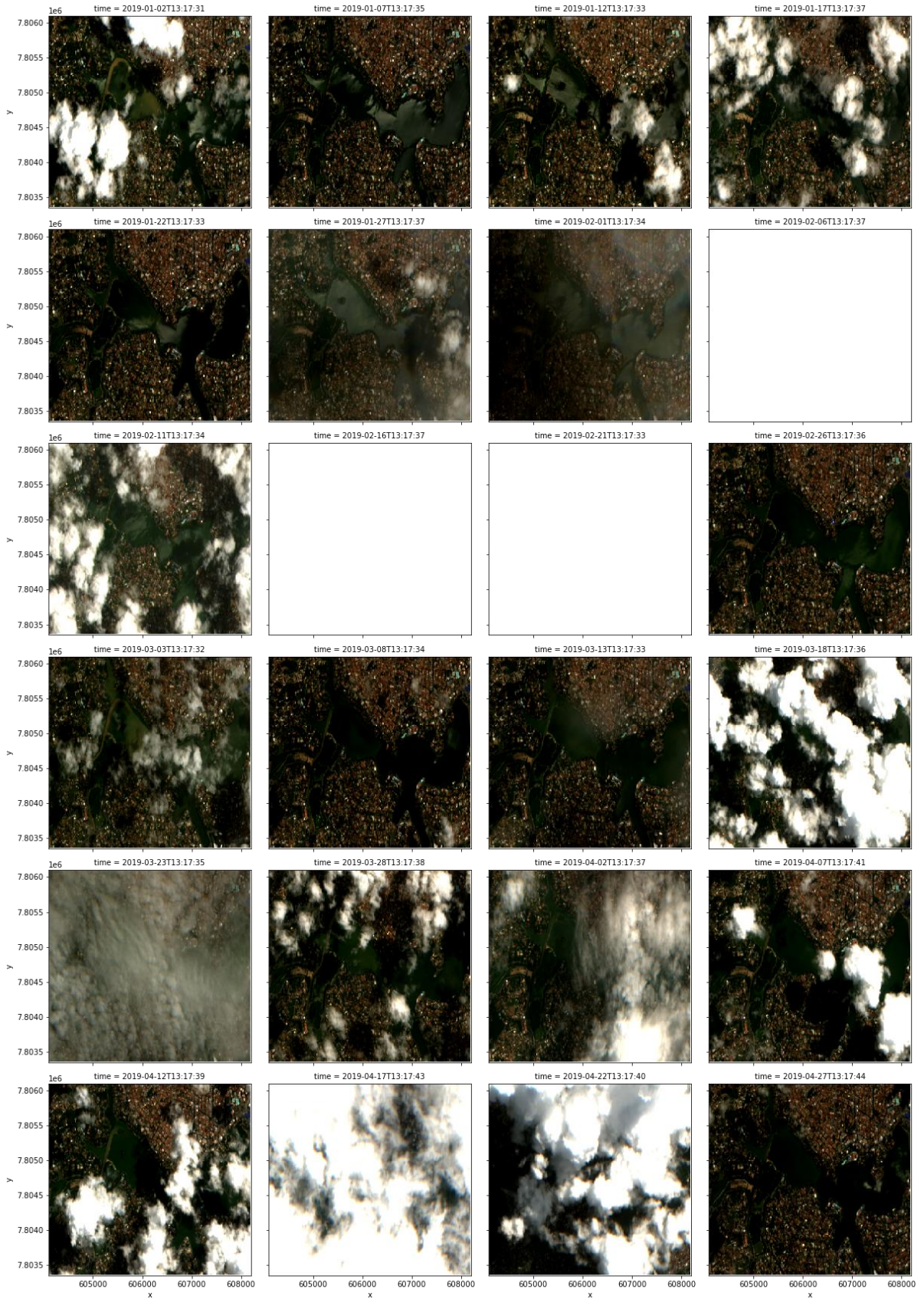


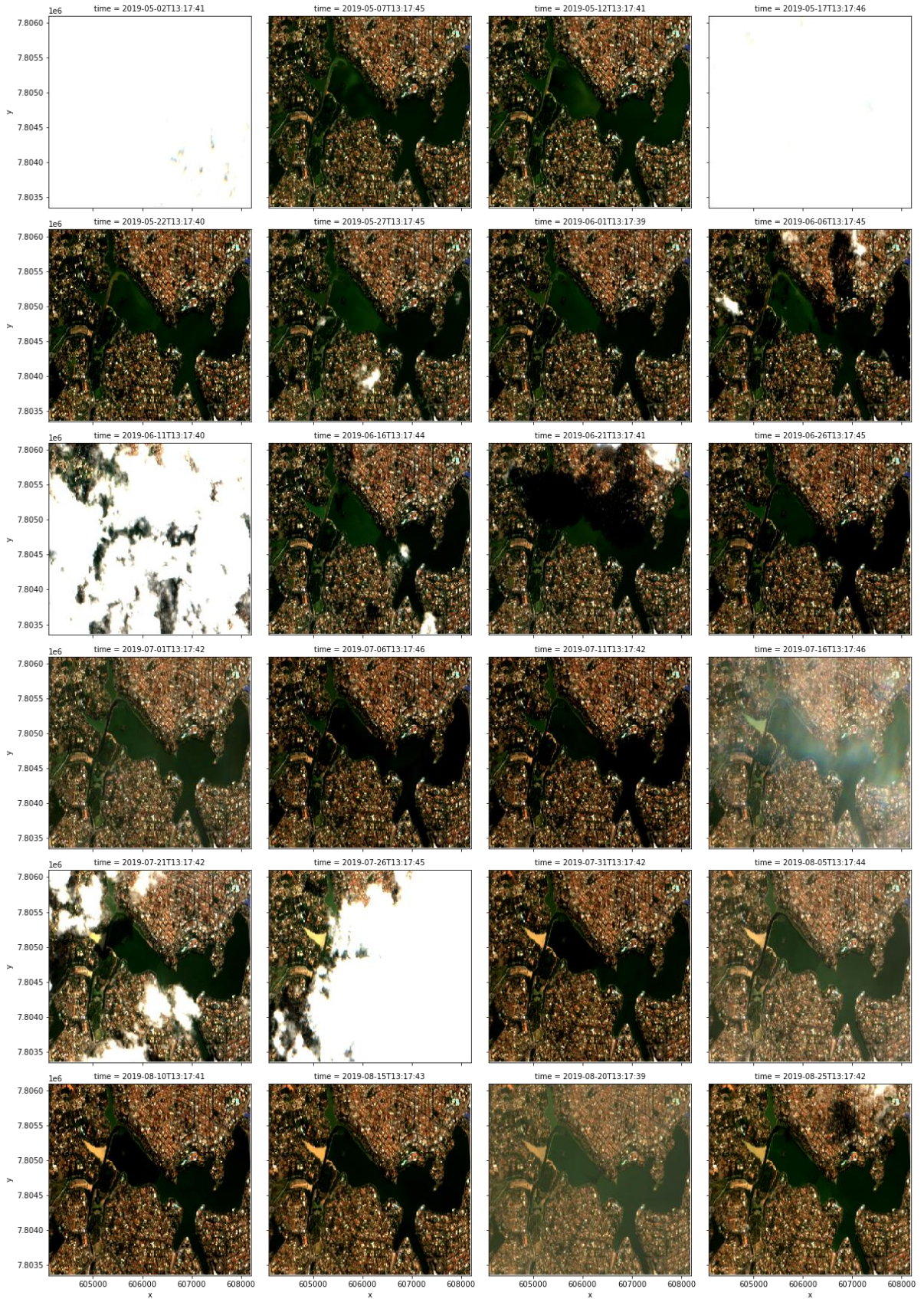


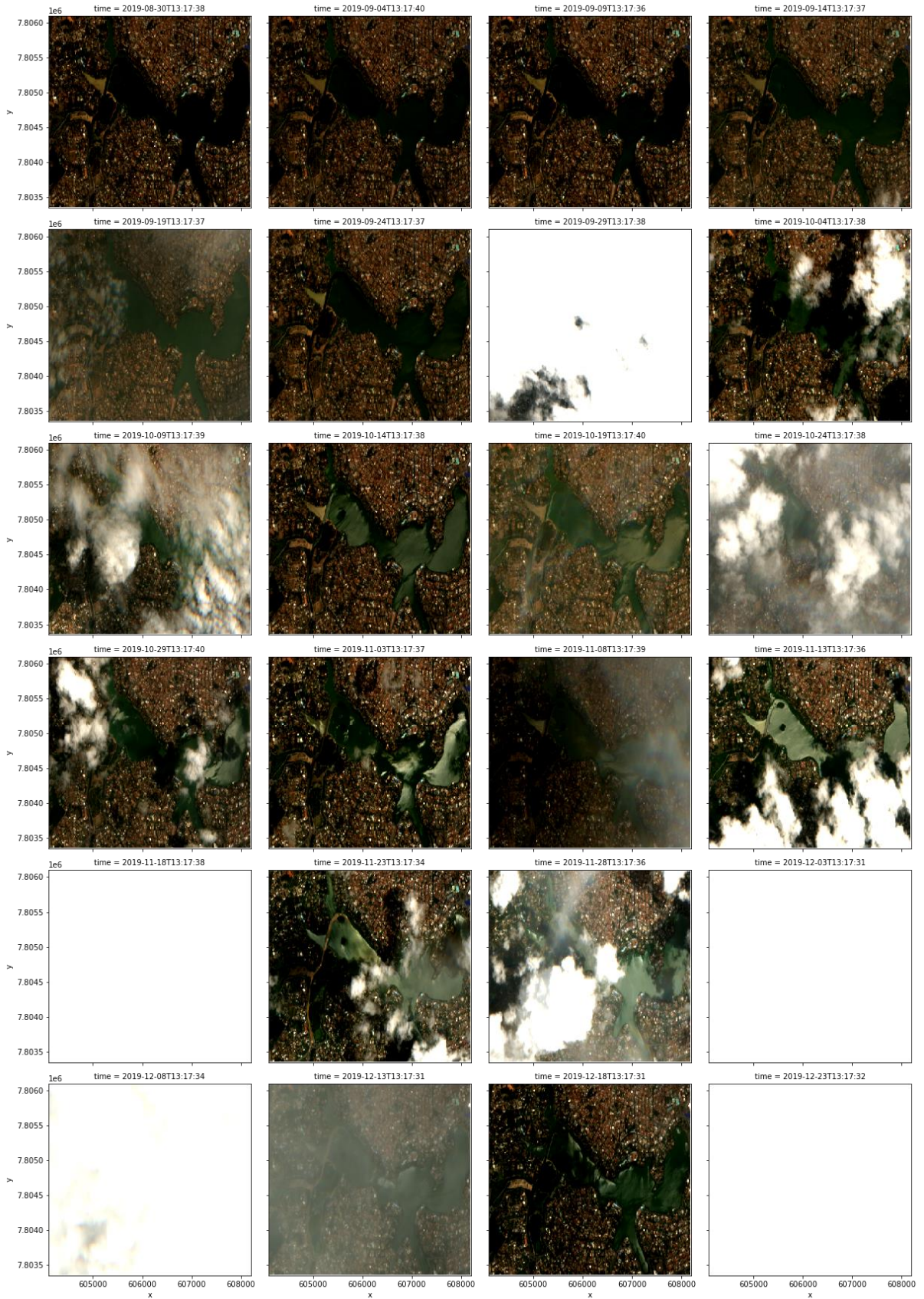


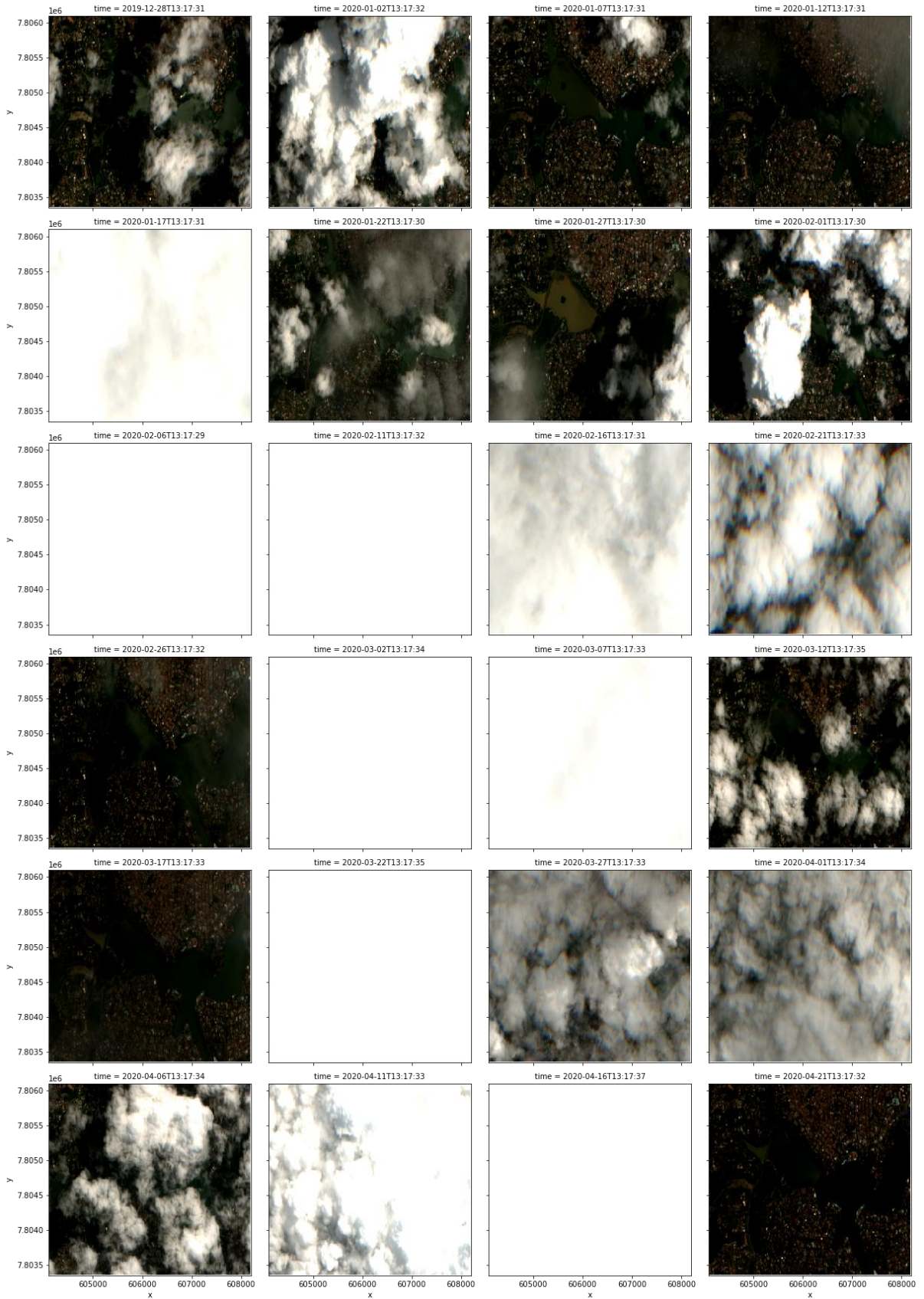


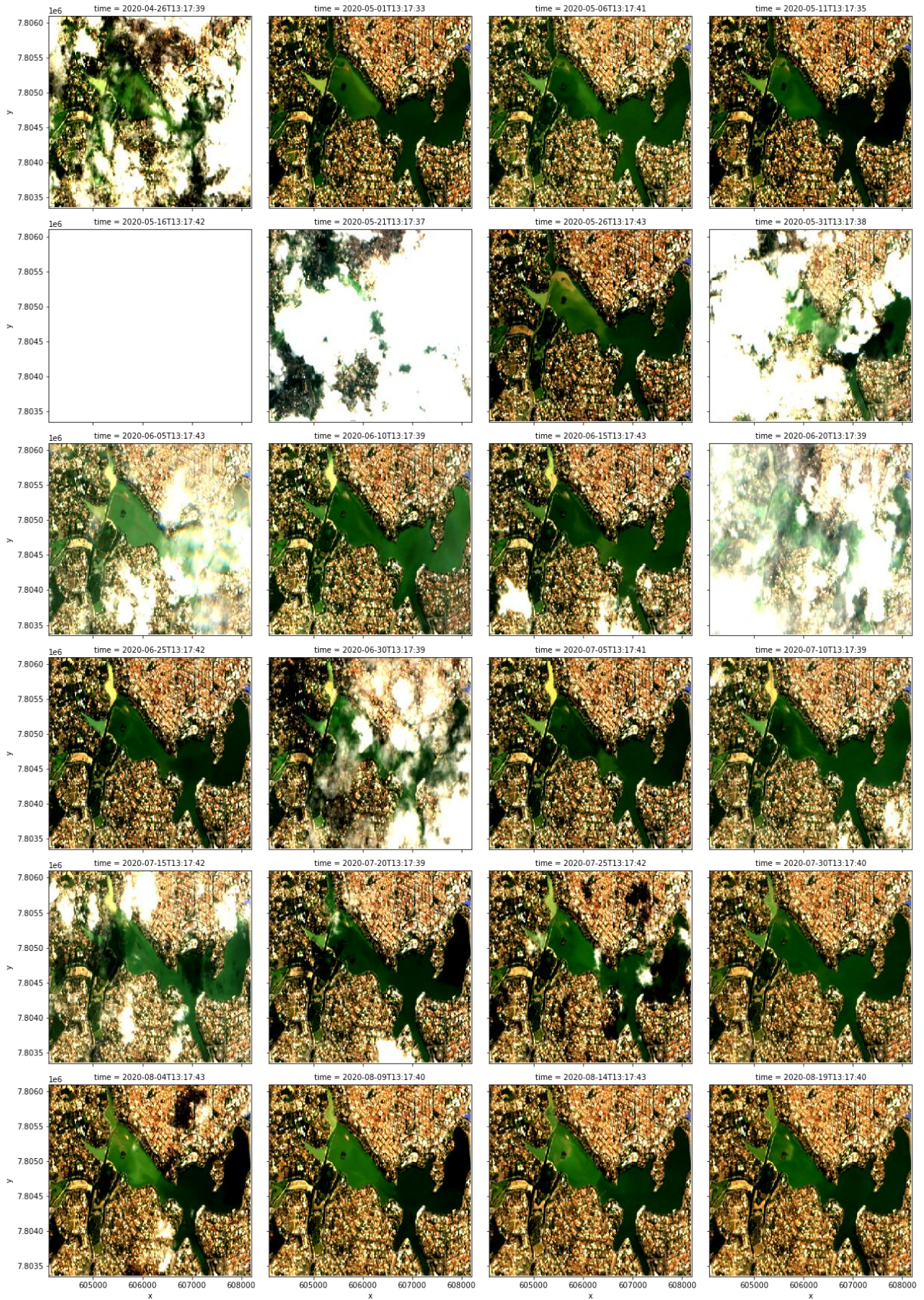


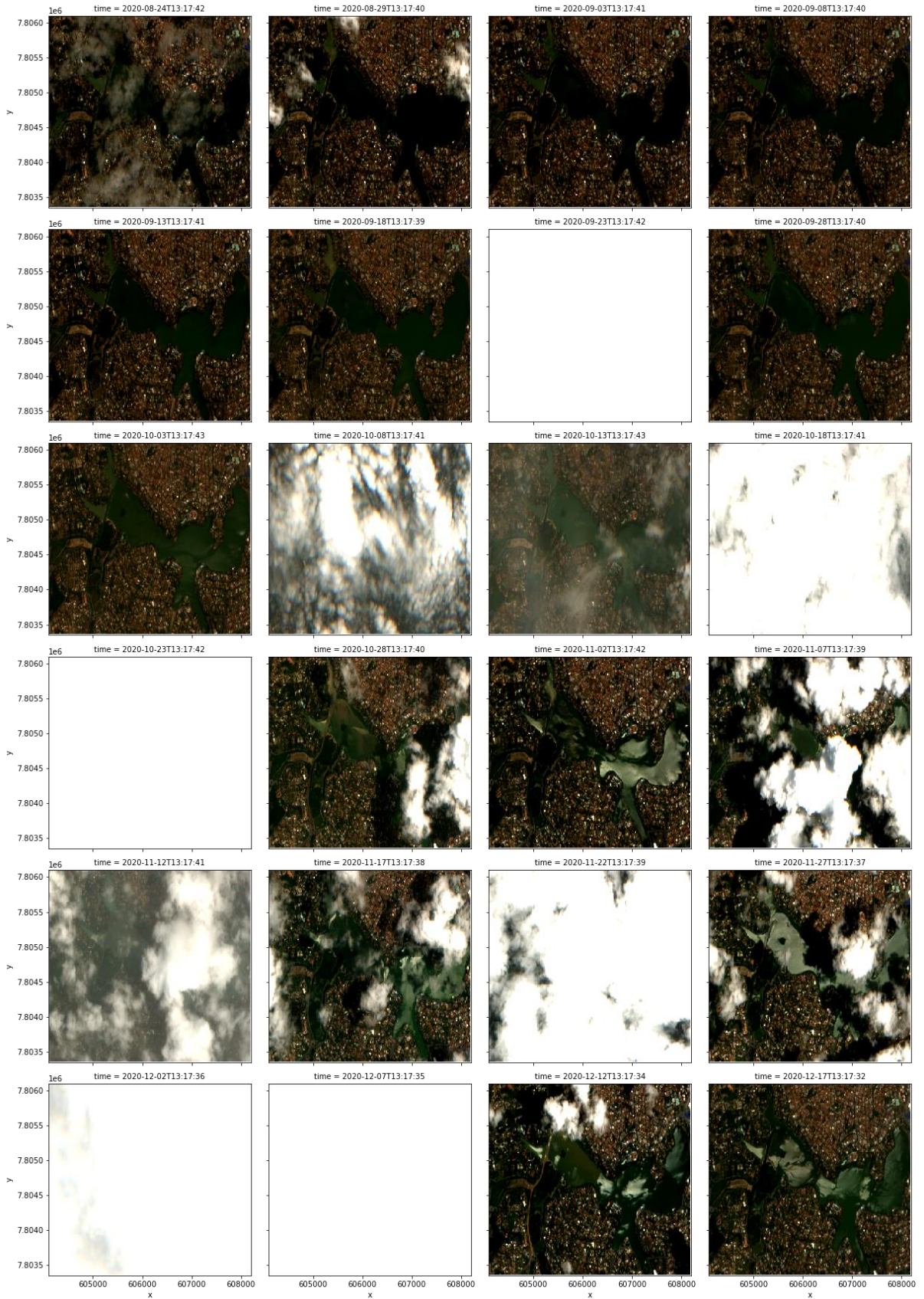


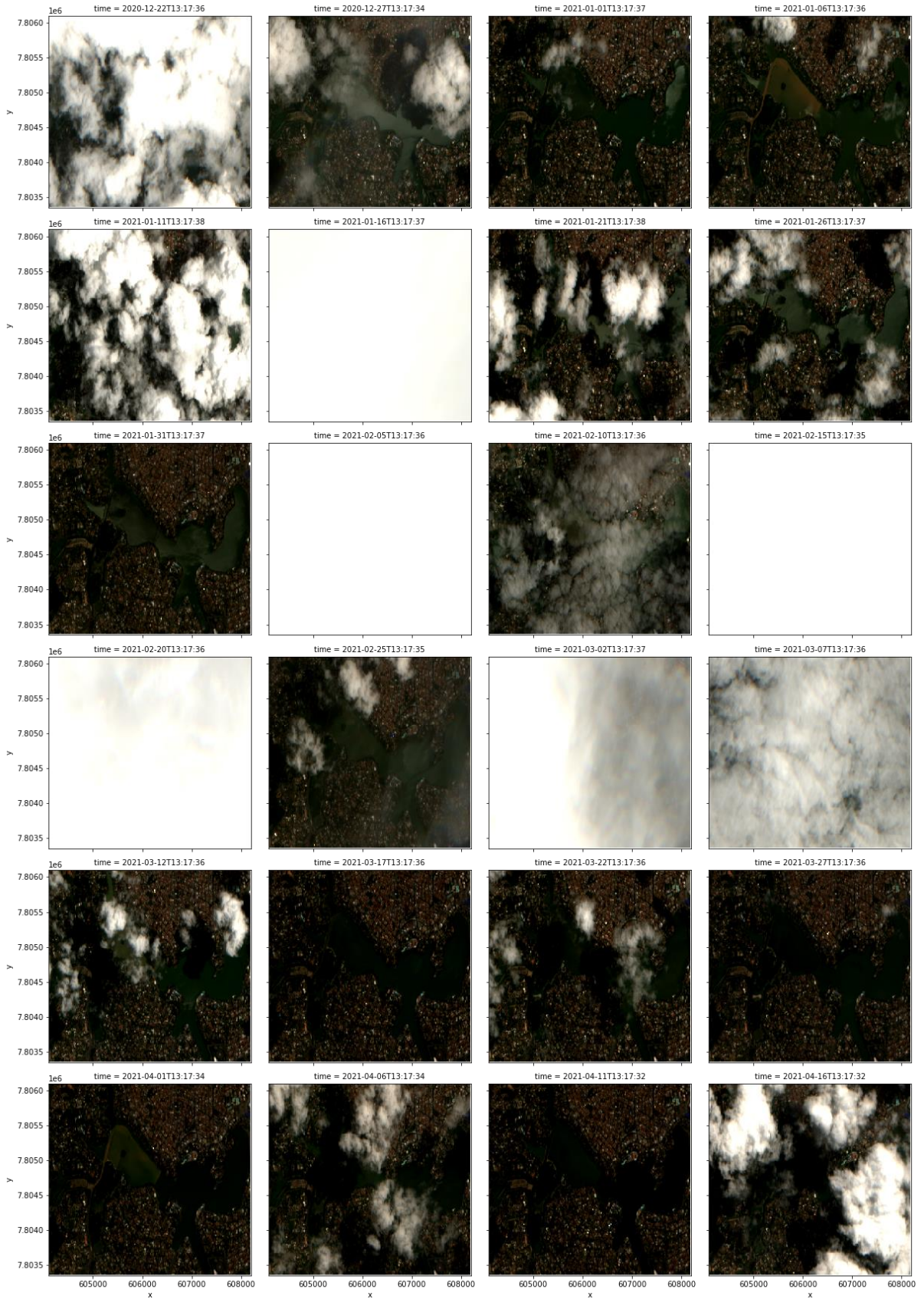


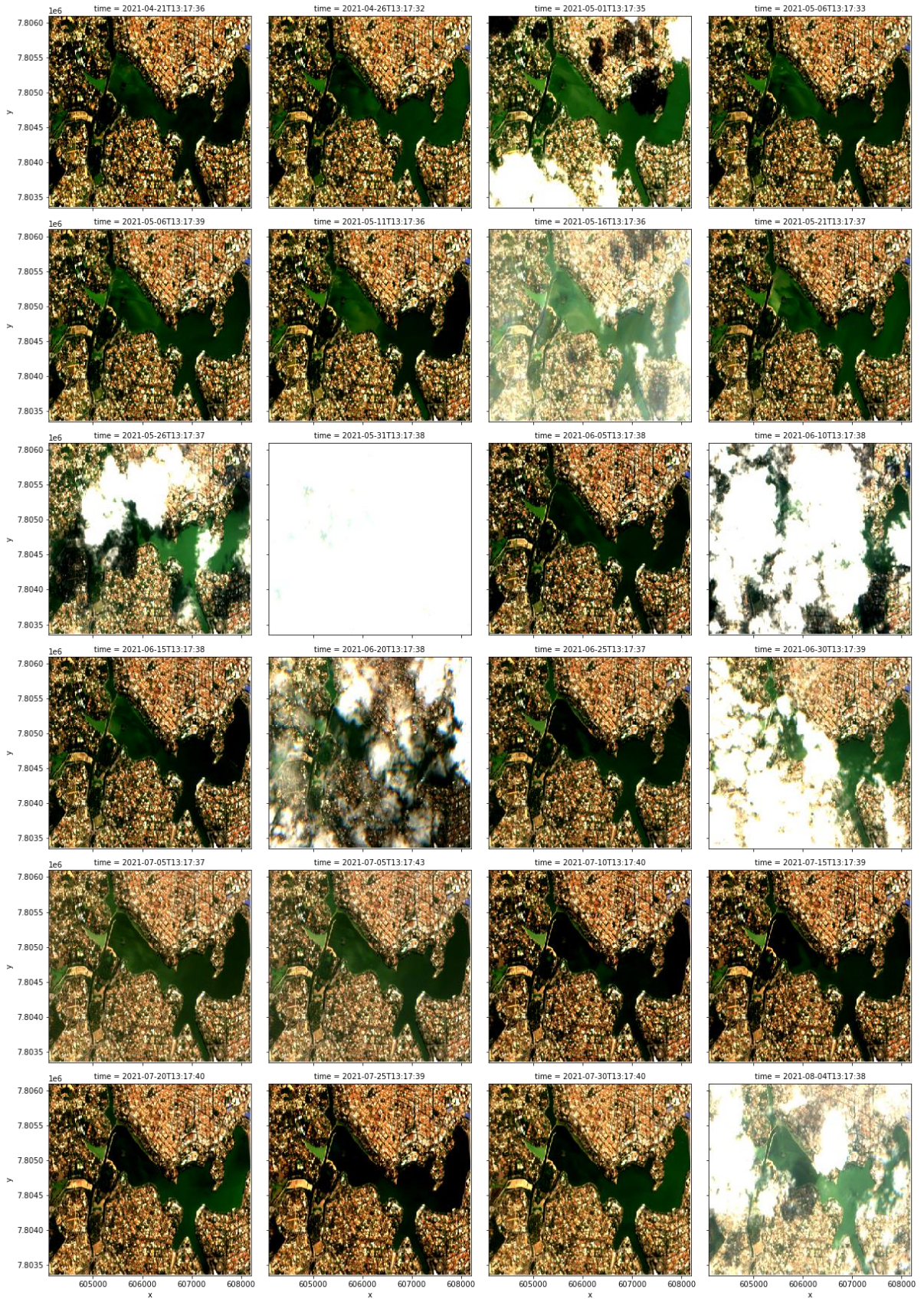


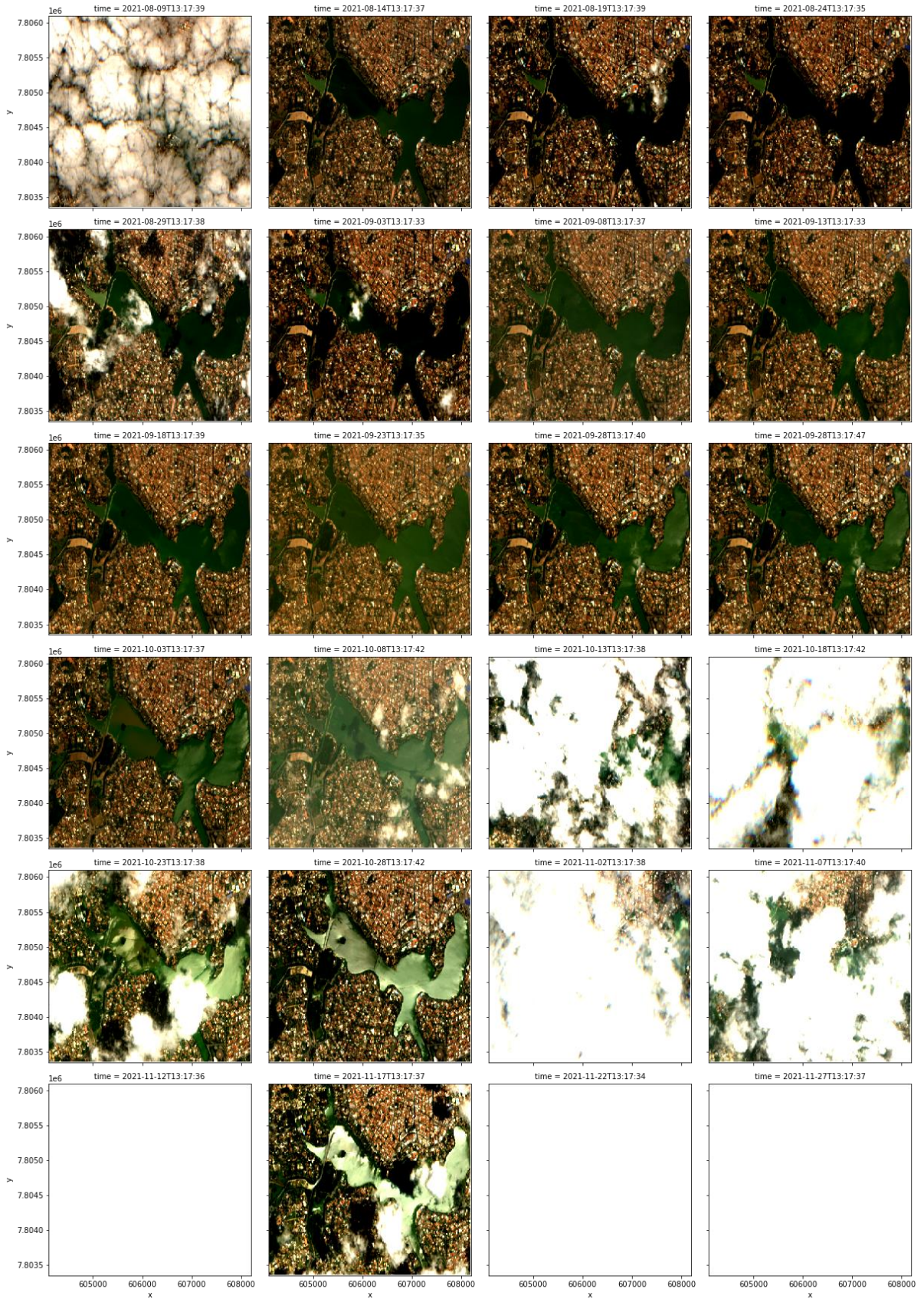


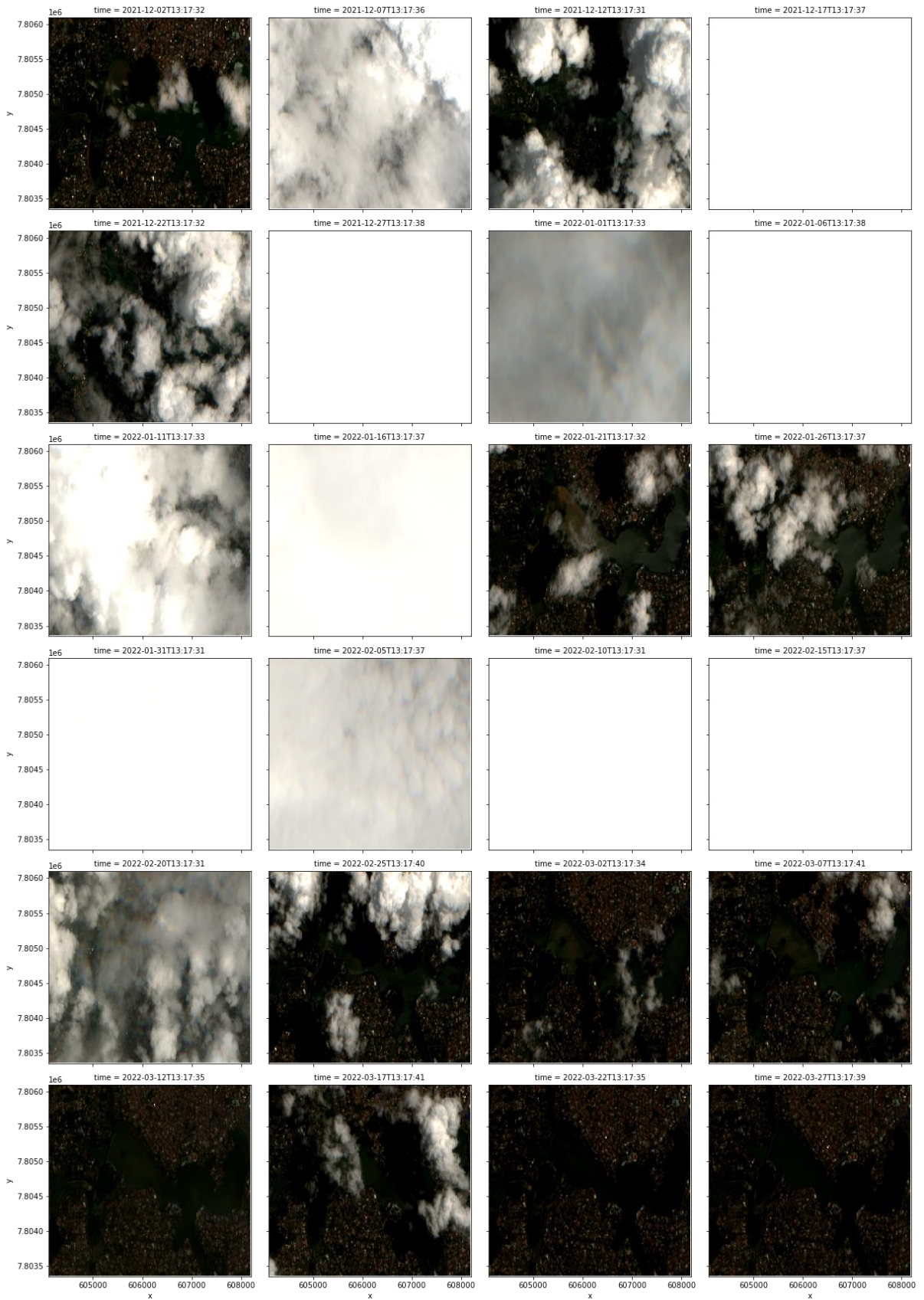


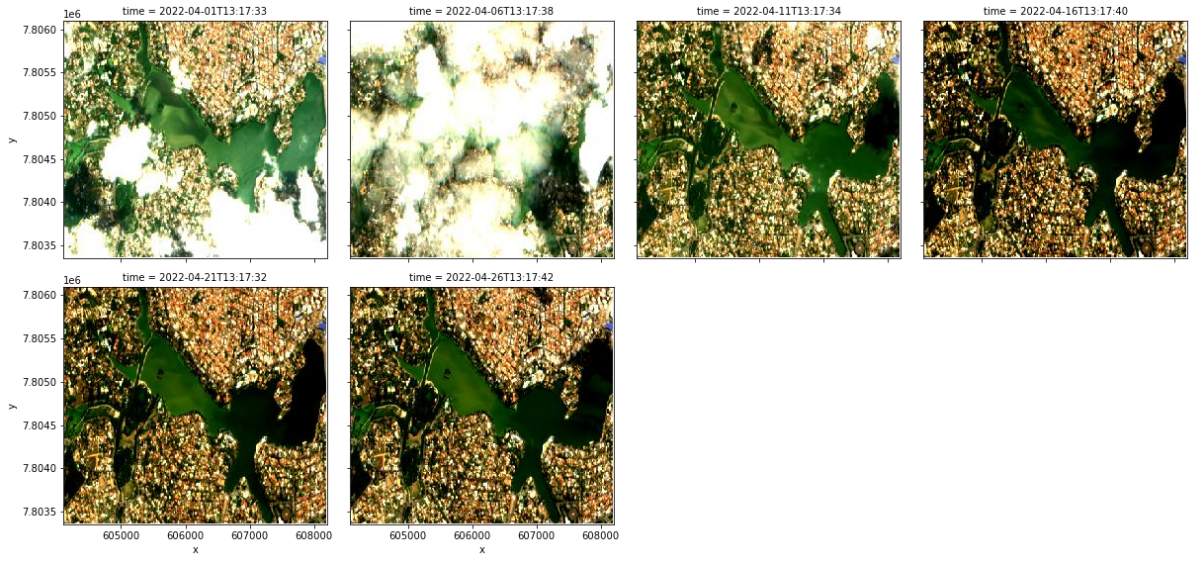












APÊNDICE E – Matchups de imagens do Landsat-8 e Sentinel-2 com cálculo de correlação dos valores modelados para os parâmetros de estudo

ID	Data Landsat-8 (UTC)	Data Sentinel-2 (UTC)	Coeficiente de correlação linear de Pearson (ρ)			
			TSM (mg/L)	Chla ($\mu\text{g/L}$)	SDD (m)	Turbidez (NTU)
0	2018-07-06 12:56:28	2018-07-06 13:12:43	0,67	0,75	0,55	0,39
1	2019-05-22 12:57:01	2019-05-22 13:17:40	0,24	0,62	0,46	0,12
2	2019-08-10 12:57:29	2019-08-10 13:17:41	0,53	0,83	0,80	0,19
3	2020-06-25 12:57:07	2020-06-25 13:17:42	-0,18	0,75	0,36	0,19
4	2020-09-13 12:57:37	2020-09-13 13:17:41	0,53	0,74	0,68	0,22
5	2021-05-06 13:17:33	2021-05-06 13:17:39	1,00	1,00	1,00	1,00
6	2021-05-11 12:56:55	2021-05-11 13:17:36	0,46	0,90	0,65	0,87
7	2021-07-30 12:57:25	2021-07-30 13:17:40	0,73	0,90	0,82	0,83
	Média		0,50	0,81	0,67	0,48

Fonte: Próprio Autor (2023)

APÊNDICE F – Comparação e avaliação da série temporal dos parâmetros de qualidade de água modelados frente os parâmetros monitorados no monitoramento limnológico

