

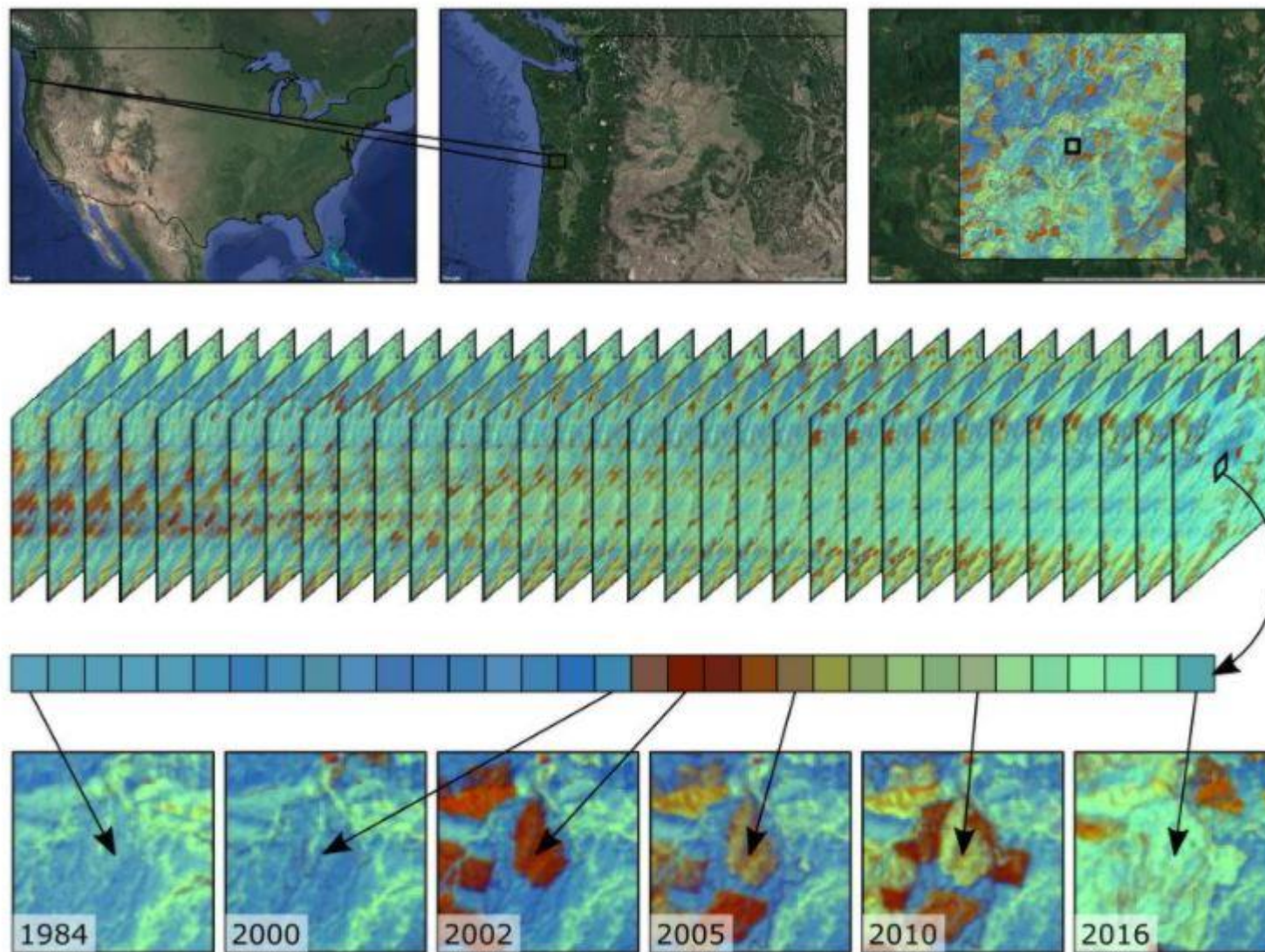
Detecção de Áreas de Florestas Invariantes em Séries Temporais Utilizando Random Forest

Eduardo Ribeiro Lacerda

Fluminense Federal University (UFF)
International Institute for Sustainability (IIS)



Algoritmos de detecção de mudança



Algoritmos de detecção de mudança

- VCT – *Vegetation Change Tracker* (2010)
- Landtrendr (2012)
- ITRA - *Image Trends from Regression Analysis* (2012)
- MIICA - *Multi-index Integrated Change Analysis* (2013)
- EWMACD - *Exponentially Weighted Moving Average Change Detection* (2014)
- CCDC - *Continuous Change Detection and Classification* (2014)
- Shapes-NBR (2016)
- VerDET - *Vegetation Regeneration and Disturbance Estimates through Time* (2017)
- COLD - *Continuous Monitoring of Land Disturbance* (2020)



O problema

- A limpeza dos dados na etapa de pós processamento desses algoritmos é essencial para a obtenção de resultados de boa qualidade
- Áreas de floresta que possuem alta declividade, por exemplo, tendem a sofrer mais com certos ruídos
- A limpeza desses dados normalmente acontece utilizando dados de projetos como o Mapbiomas como base para ignorar áreas de não interesse



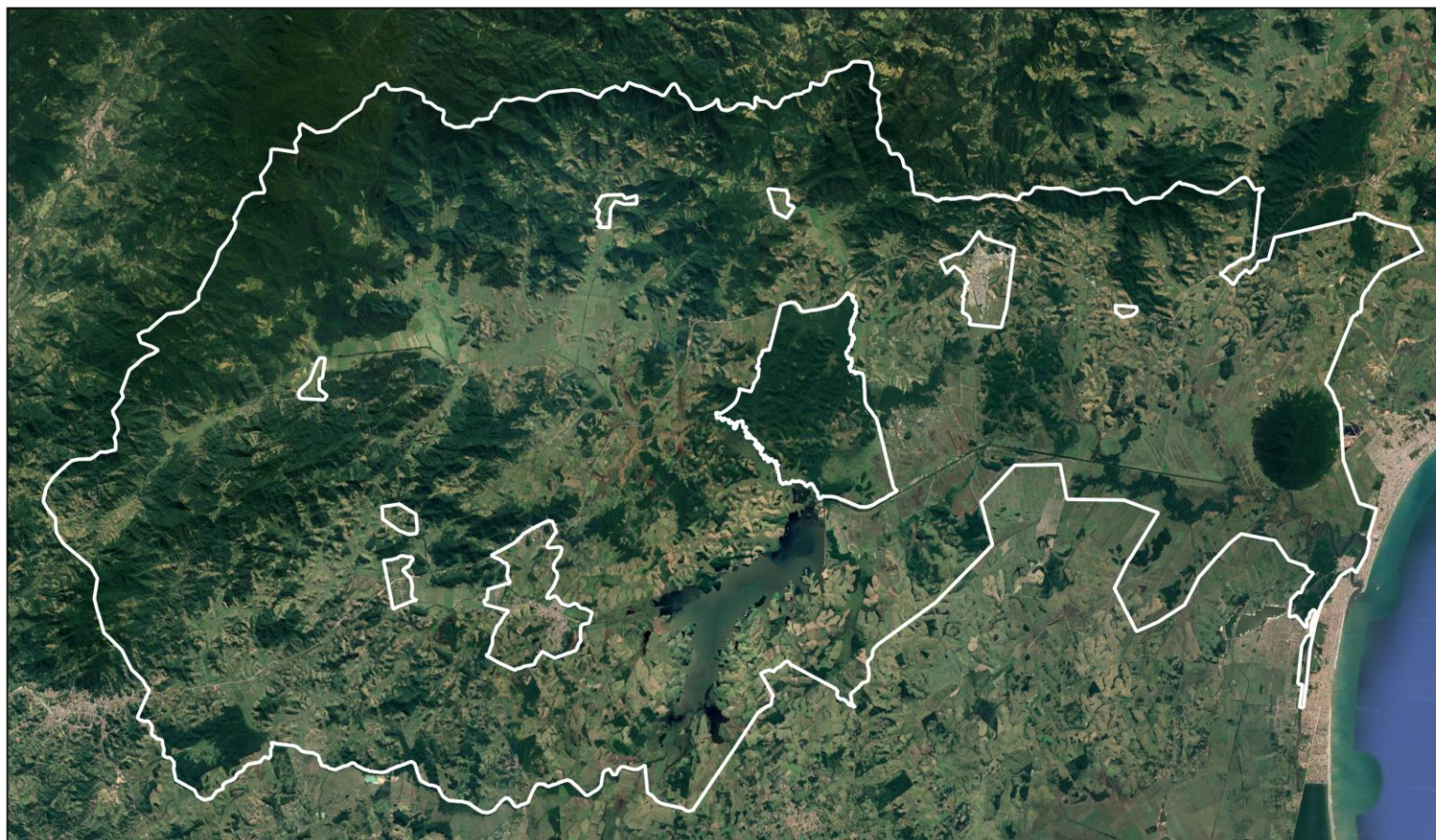
Como resolver?

- O presente trabalho busca elaborar uma alternativa a essa técnica
- Criar uma forma mais precisa de gerar camadas de áreas de estabilidade/invariância
- Utilizamos o algoritmo Random Forest na plataforma Google Earth Engine (GEE) além da linguagem R e do pacote MLR (Machine Learning in R)



Área de Estudo

Limite da Bacia Hidrográfica do Rio São João



Sistema de Projeção UTM, Datum WGS84.
Autor: Eduardo Ribeiro Lacerda, 2020.

0 10 20 km

Elementos do Mapa

 Bacia Hidrográfica do Rio São João

Máximo NDVI

- Landsat 5, 7 e 8 (Surface Reflectance Tier 1)
- Retirou todos os pixels com nuvem/sombra
- Mediana de cada ano (1985 – 2018)
- Extração do valor máximo do NDVI em uma única camada final (output)



Definição do limiar para o NDVI Max

- Encontrar um limiar para classificar o que é floresta e o que não é
- Mapbiomas (floresta binário e pasto/agricultura binário) -> raster pra ponto (vetor)
- Seleção aleatória de 2000 pontos para cada classe
- Extração utilizando os pontos dos valores do raster de NDVI max
- Teste T de Student -> valor $p = 2.2e-16$ (demonstrando diferença significativa entre as classes)
- Extração do valor mínimo encontrado nas 2000 amostras de NDVI max pela classe floresta = 0.83 ndvi
- Criação de uma nova camada raster binária com todos os valores iguais ou maiores que 0.83



Criação de camada de florestas

- Multiplicação da camada do Mapbiomas de florestas invariantes com a nova camada com o max ndvi



Camada Landtrendr (florestas que sofreram mudança)

- Greatest Loss
- 1985 – 2018
- Período: 1 de janeiro até 31 de dezembro
- Camada: NDVI
- Filtros: Magnitude maior que 200 (> 0.2 de NDVI)



Camada Outros

- Tudo que não foi classificado nem como floresta na camada de florestas e nem como mudança pela camada do Landtrendr
- Água, solo exposto, pasto, etc...
- Amostras selecionadas de forma aleatória (Raster to Point + Random Selection)



Criação da série temporal para classificação

- Série Landsat 5, 7 e 8
- Limpeza de nuvens de sombras (no data)
- Mediana para a composição das camadas anuais
- Bandas: blue, green, red, nir, swir1, swir2, ndvi, ndmi, ndwi, savi, greenness, wetness, brightness



Validação Cruzada (cross validation) usando o MLR



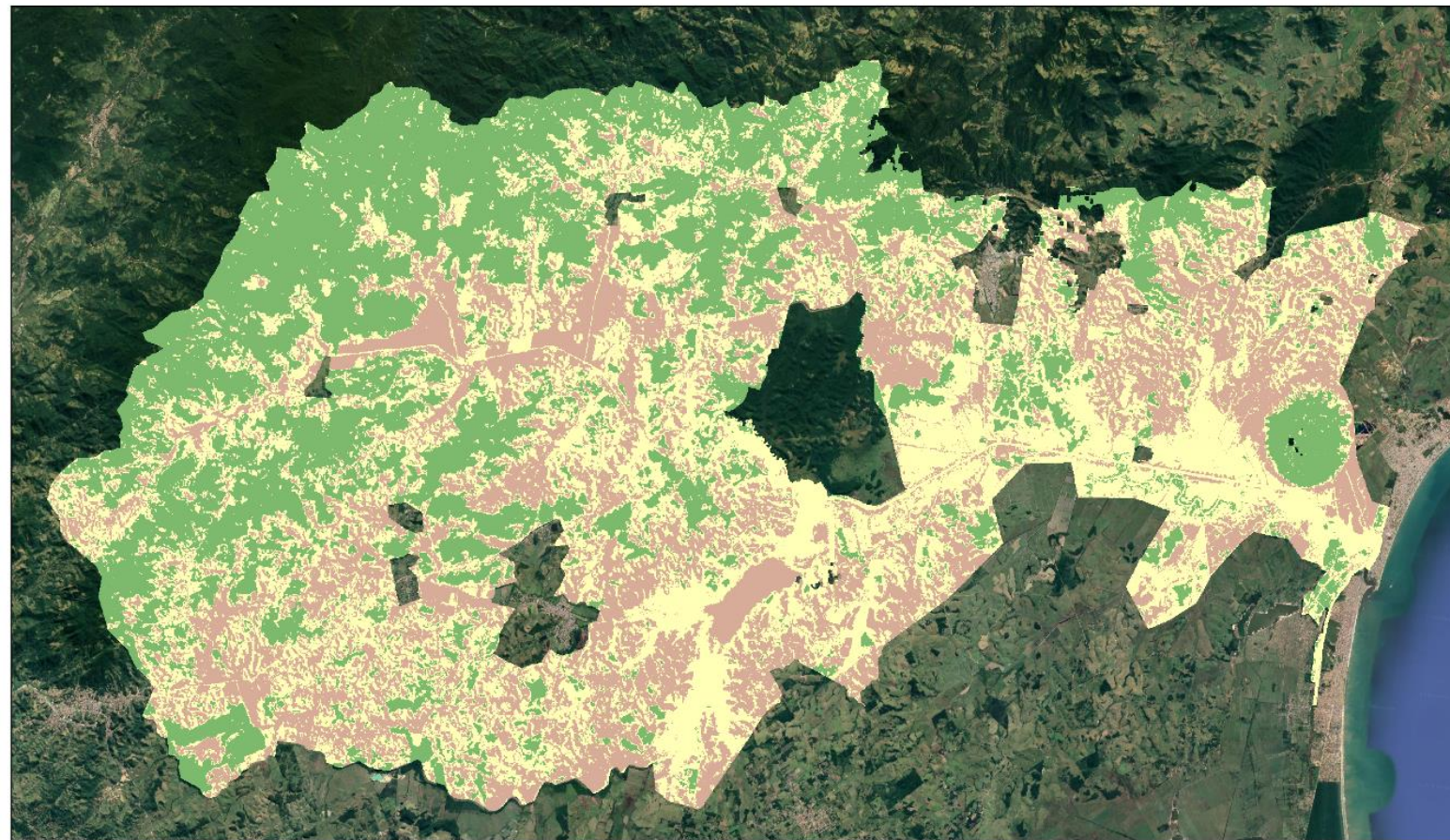
- O processo de validação cruzada foi realizado no R utilizando o pacote MLR
- K-fold:10 (90% treino e 10% teste)
- Iterações: 100 (resampling)
- Algoritmo de classificação: Random Forest (100 árvores)
- 24 testes com combinações de bandas diferentes totalizando 24.000 processos de classificação diferentes

Resultados de acordo com a combinação de bandas utilizada

BANDA	Kappa
Blue + Green + Red + NIR + SWIR1 + SWIR2 + NDVI + NDWI + NDMI + SAVI	0,844754
Blue + Green + Red + NIR + SWIR1 + SWIR2 + NDVI + NDWI + NDMI + SAVI + Greenness + Wetness + Brightness	0,8398718
Blue + Green + Red + NIR + SWIR1 + SWIR2	0,8388044
Blue + Green + Red + NIR	0,8333813
Blue + Green + Red	0,8327737
Greenness + Wetness + Brightness + NDVI + NDWI + NDMI + SAVI	0,8295032
NDVI + NIR + NDMI	0,8198751
NDVI + NIR	0,8198327
Greenness + Wetness + Brightness	0,8126214
Red	0,8124956
SWIR2	0,8089151
Green	0,8055279
Blue	0,8000793
NDVI + NDWI + NDMI + SAVI	0,7966787
NDVI + NDMI	0,7833422
SWIR1	0,7788457
Wetness	0,7675543
Greenness	0,761319
NDVI	0,7554752
SAVI	0,7552823
Brightness	0,7159552
NDMI	0,7060223
NDWI	0,6343948
NIR	0,5760713

Classificação final no Google Earth Engine (Random Forest)

Resultado da Classificação Random Forest



Classes
Floresta Invariante
Perda Landtrendr
Outros

0 10 20 km

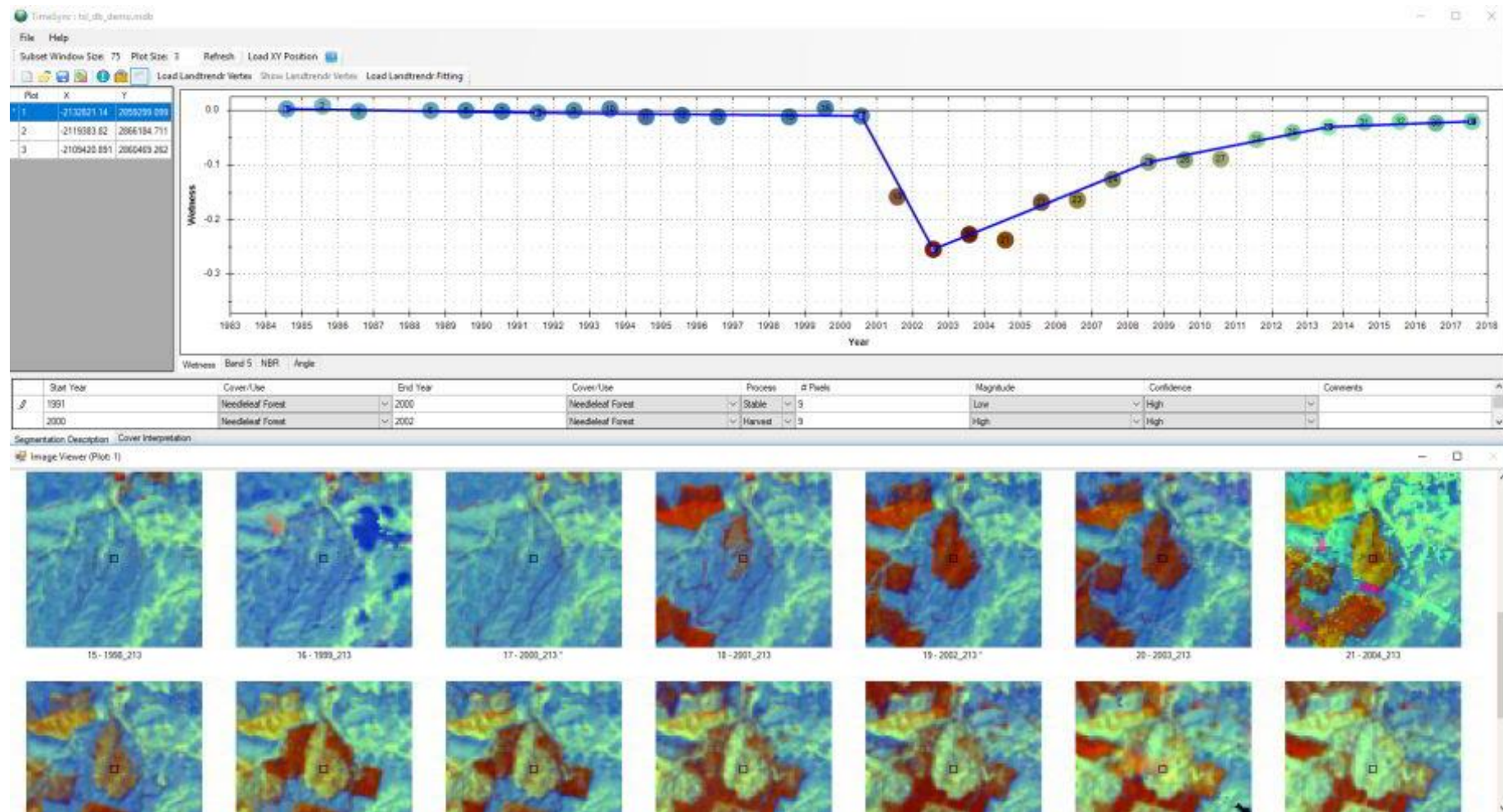


Diferenças entre o resultado obtido e a técnica antiga (Mapbiomas)

- 2000 mil amostras aleatórias para cada classe
- (FOR) – Áreas de floresta classificadas pelos dois mapeamentos
- (RF) - Áreas de floresta classificadas apenas pela técnica proposta (random forest)
- (Mapbiomas) - Áreas de floresta classificadas apenas pelo Mapbiomas
- ANOVA – 0,0116 (mostrando uma diferença estatisticamente significativa)
- Teste de Tukey

Combinação (Teste de Tukey)	Valor-p
Mapbiomas - FOR	0,0008489
RF - FOR	0,1841643
RF - Mapbiomas	0,0000002

Cálculo do erro associado a cada classe (TimeSync)



Cálculo do erro associado a cada classe (TimeSync)

- 300 amostras por classe
- Mapbiomas (somente mapbiomas) – 55% acerto
- RF (somente random forest) – 76% acerto
- FOR (ambos) – 97% acerto
- Mapeamento final – 91,7% acerto



Link para os dados

Todos os resultados, amostras, imagens, arquivos vetoriais, códigos e materiais para validação utilizados neste trabalho estão disponíveis para visualização e possível reprodução através deste link:

<https://github.com/sacridini/Artigo-JGEOTEC-2020>

