



Mapping Deforestation Trends in the Brazilian Amazon Using Machine Learning

Courtney Whalen, Data Scientist

Jason Brown, Senior Data Scientist

Goal

- Identify areas of the planet where deforestation is occurring
 - Monitoring deforestation is important for understanding impacts on climate change

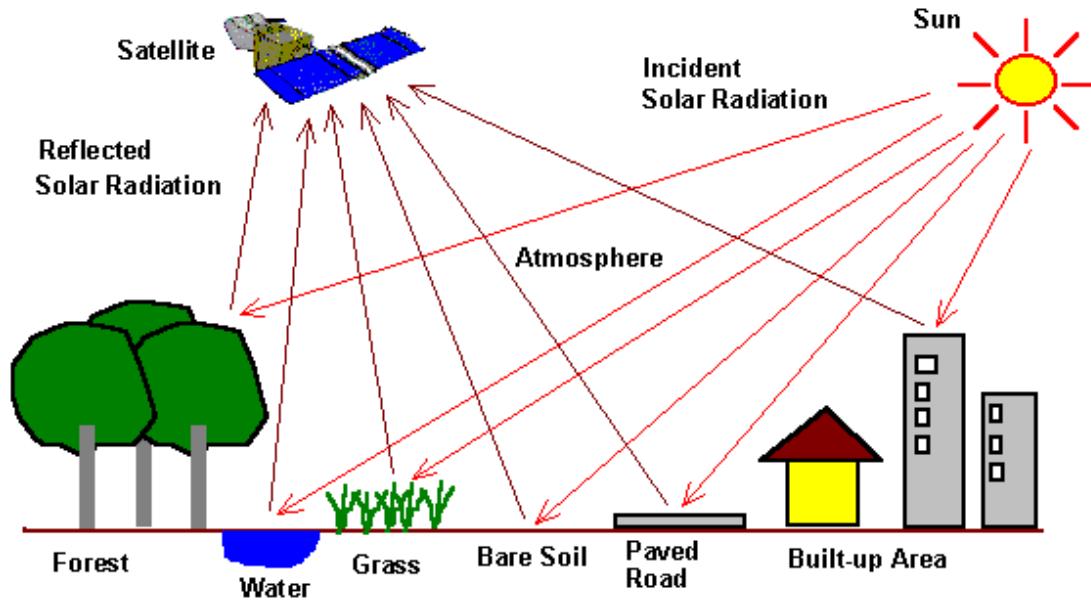


Methodology

- Supervised classification model using remote sensing data
 - Binary target
- Post-classification change detection



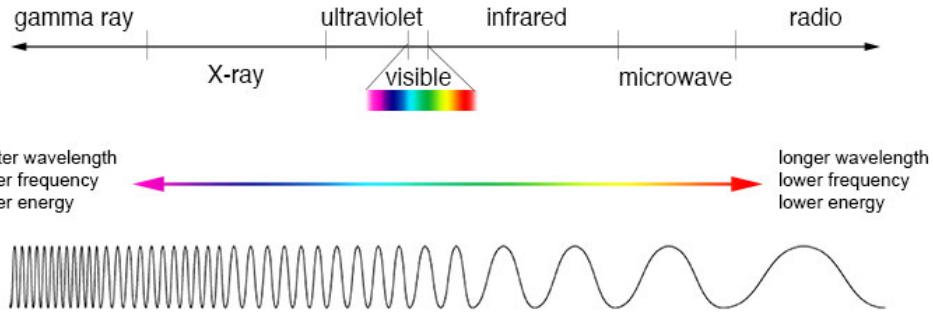
Remote Sensing



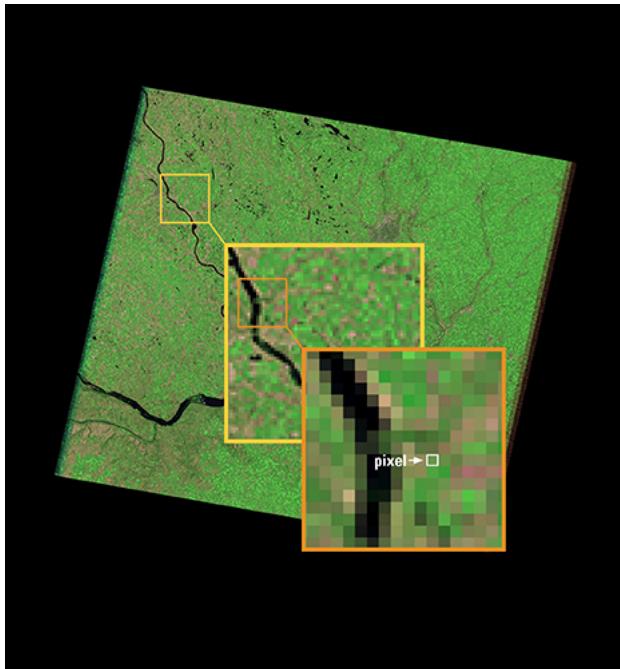
Satellite Data

- Moderate Resolution Imaging Spectroradiometer (MODIS)
 - Revisit rate: 1-2 days
 - Resolution: 500 m
 - 7 spectral bands
 - Atmospheric corrections

Band	Bandwidth (nm)	Resolution (m)
Band 1 (Red)	620 - 670	250
Band 2 (NIR)	841 - 876	250
Band 3 (Blue)	459 - 479	500
Band 4 (Green)	545 - 565	500
Band 5 (SWIR)	1230 - 1250	500
Band 6 (SWIR)	1628 - 1652	500
Band 7 (SWIR)	2105 - 2155	500



Raster Data



pixel	red	blue	green	nir	...	forest
1	0.9	0.1	0.2	0.9	...	TRUE
2	0.5	0.6	0.8	0.6	...	FALSE
3	0.3	0.4	0.7	0.2	...	FALSE
⋮	⋮	⋮	⋮	⋮	⋮	⋮
52956	0.8	0.4	0.3	0.9	...	TRUE

Training Data

- System for Terrestrial Ecosystem Parameterization (STEP) reference data
 - ~2,000 hand-labeled sites scattered across every continent
 - Database of training polygons
 - 17 land cover types
 - 5 forest types



RasterFrames

- Software Library that unlocks Spark for global-scale geospatial machine learning
- Available in Scala and Python
- Free and Open Source Software under Eclipse Foundation - LocationTech

Learn more at: rasterframes.io

The screenshot shows the RasterFrames documentation site. The left sidebar includes links for Getting Started, Creating RasterFrames, Spatial Queries, Applications, GeoTrellis Operations, Computing NDVI, Machine Learning, Raster Statistics, Clustering, Classification, Exporting, RasterFrames Reference, and Release Notes. The main content area features a title 'RasterFrames' and a brief description: 'RasterFrames™ brings the power of Spark DataFrames to geospatial raster data, empowered by the map algebra and tile layer operations of GeoTrellis.' It also mentions the source code on GitHub. Below this is a detailed description of the library's purpose and how it integrates with Spark Catalyst. A 'Getting Started' link is provided. On the right, there's a navigation bar with 'On this page:' and links for 'RasterFrames', 'Detailed Contents', and 'Related Links'. At the bottom, a flow diagram illustrates the data pipeline: GeoTIFF and GeoTrellis Layers feed into a 'RasterFrame' component, which then performs operations like 'join', 'Layer Operations', 'Map Algebra', 'Statistical Analysis', 'Machine Learning', and 'Visualization' before being consumed by 'Your Application'.

Feature Engineering

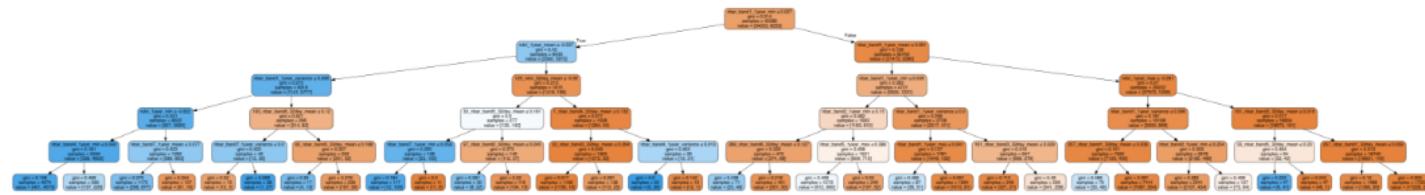
- 128 time windowed features using seven different spectral bands
 - Monthly aggregates
 - Mean for each band over 32 day span
 - Yearly aggregates
 - Min, max, mean, variance for each band over year span
 - Normalized difference vegetation index (NDVI)
- Land cover types transformed into a binary target: forest/not forest

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}$$



Model Training

- Trained on MODIS 2012 data at pixel level
 - 80% train, 20% test
 - Split data so all pixels from each site kept within the same set
 - Balanced forest/not forest pixels within each set



Model Selection

Model	Parameters	Accuracy
Decision Tree	Depth = 10	0.939
	Depth = 15	0.926
	Depth = 20	0.914
Random Forest	Depth = 10	0.943
	Depth = 15	0.951
	Depth = 20	0.952
Gradient Boosted Trees	Depth = 10	0.941
	Depth = 15	0.928
	Depth = 20	0.912
Neural Net	5 layers	0.935
Logistic Regression	Elastic net = 0.5	0.918

Final model: Random Forest

- 1000 trees
- Depth of 15
- Minimum of 10 instances per node

		Predicted	
		Forest?	
Actual	True	True	False
	True	1642	452
	False	70	8496

Precision = 0.96

Recall = 0.78

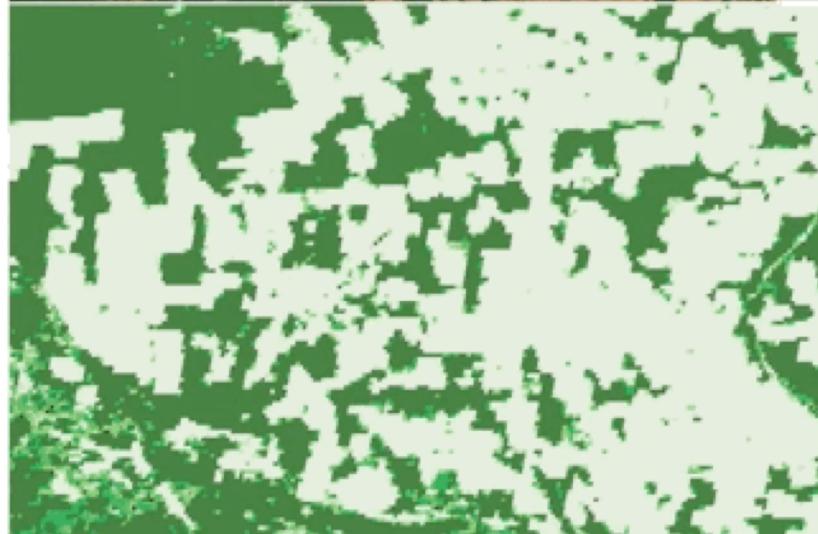
Feature Importance

1. Red band - year aggregate minimum
2. Red band - year aggregate mean
3. NDVI - year aggregate mean
4. Green band - year aggregate mean
5. NDVI - year aggregate maximum
6. Blue band - year aggregate minimum
7. Blue band - year aggregate mean
8. Green band - year aggregate minimum
9. Green band - year aggregate maximum
10. NDVI - June aggregate mean
11. Red band - August aggregate mean
12. NDVI - year aggregate minimum
13. Blue band - year aggregate minimum
14. Red band - year aggregate minimum
15. Red band - June aggregate mean

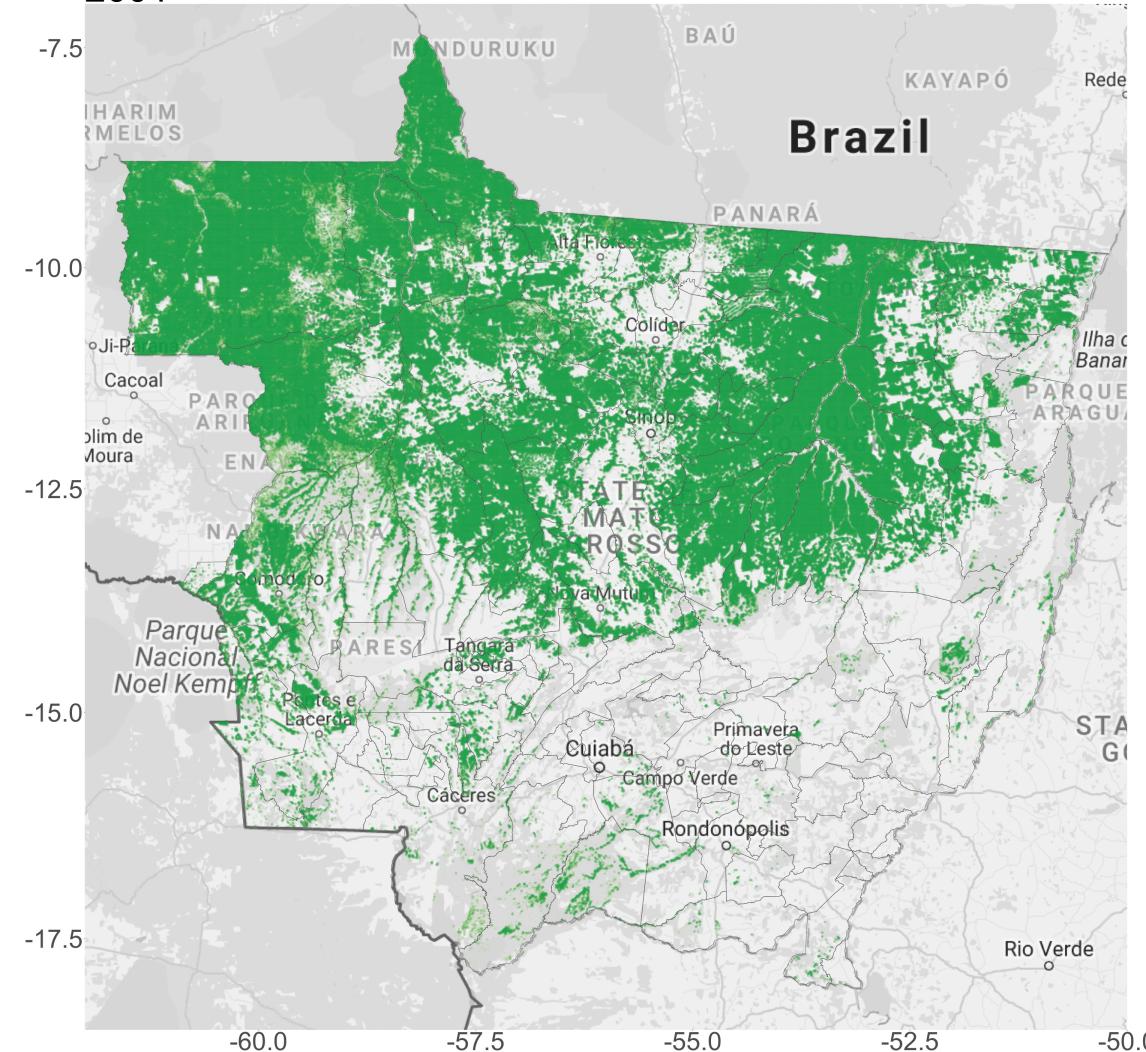
2001-Aug Landsat 7



2002-Sept Landsat 7



2001



Percentage of Forest Cover in Mato Grosso, Brazil



Next Steps

Improving model:

- Add more features
- Higher spatial resolution
- Better handling of seasonal differences across world

Applying model:

- Shorten time window for change detection
- Score the whole world



Questions?

astraea.earth

rasterframes.io



Appendix

IGBP Land Cover Classification System

Class	Class name	Description
1	Evergreen needleleaf forests	Lands dominated by needleleaf woody vegetation with a percent cover >60% and height exceeding 2 m. Almost all trees remain green all year. Canopy is never without green foliage.
2	Evergreen broadleaf forests	Lands dominated by broadleaf woody vegetation with a percent cover >60% and height exceeding 2 m. Almost all trees and shrubs remain green year round. Canopy is never without green foliage.
3	Deciduous needleleaf forests	Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 m. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
4	Deciduous broadleaf forests	Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 m. Consists of broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
5	Mixed forests	Lands dominated by trees with a percent cover >60% and height exceeding 2 m. Consists of tree communities with interspersed mixtures or mosaics of the other four forest types. None of the forest types exceeds 60% of landscape.
6	Closed shrublands	Lands with woody vegetation less than 2 m tall and with shrub canopy cover >60%. The shrub foliage can be either evergreen or deciduous.
7	Open shrublands	Lands with woody vegetation less than 2 m tall and with shrub canopy cover between 10% and 60%. The shrub foliage can be either evergreen or deciduous.
8	Woody savannas	Lands with herbaceous and other understory systems, and with forest canopy cover between 30% and 60%. The forest cover height exceeds 2 m.
9	Savannas	Lands with herbaceous and other understory systems, and with forest canopy cover between 10% and 30%. The forest cover height exceeds 2 m.
10	Grasslands	Lands with herbaceous types of cover. Tree and shrub cover is less than 10%.
11	Permanent wetlands	Lands with a permanent mixture of water and herbaceous or woody vegetation. The vegetation can be present either in salt, brackish, or fresh water
12	Croplands	Lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems). Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type
13	Urban and built-up lands	Land covered by buildings and other man-made structures.
14	Cropland/natural vegetation mosaics	Lands with a mosaic of croplands, forests, shrubland, and grasslands in which no one component comprises more than 60% of the landscape.
15	Snow and ice	Lands under snow/ice cover throughout the year.
16	Barren	Lands with exposed soil, sand, rocks, or snow and never have more than 10% vegetated cover during any time of the year.
17	Water bodies	Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or saltwater bodies.

Forest Threshold

Threshold	Accuracy
0	0.196
0.1	0.807
0.2	0.943
0.3	0.957
0.4	0.959
0.5	0.951
0.6	0.938
0.7	0.920
0.8	0.897
0.9	0.866
1	0.804

Accuracy vs. Threshold

