NASA DEVELOP National Program
North Carolina – NCEI
Summer 2022

Mato Grosso Agriculture

Enhancing Crop Classification Mapping Using Optical and Radar Satellite Sensors to Enhance Agricultural Management and Policy Making in Mato Grosso, Brazil

DEVELOP Technical Report

Max Rock (Project Lead)
Kate Reynolds
Elijah Dalton
Aidan Harvey

Advisors:
Dr. Garrett Graham (NOAA National Centers for Environmental Information, North Carolina Institute for Climate Studies)
Molly Woloszyn (NOAA National Centers for Environmental Information, National Integrated Drought Information System)

Fellow:
Katie Lange (NCEI Fellow)

1. Abstract

Ranked as the fourth largest food producer in the world, Brazil is an agricultural powerhouse. Agricultural production at this scale warrants accurate crop monitoring and classification, however, this tropical area is frequently concealed by dense cloud cover in standard optical imagery. To improve the accuracy and spatial coverage of current crop monitoring operations, the team incorporated radar data capable of penetrating cloud coverage to classify second season corn and cotton fields. Utilizing optical imagery from Landsat 8 Operational Land Imager (OLI), as well as radar imagery from Sentinel-1 C-band Synthetic-Aperture Radar (C-SAR), the NASA DEVELOP team worked with the United States Department of Agriculture (USDA) Foreign Agricultural Service (FAS) and World Agricultural Outlook Board to generate a crop classification procedure using a random forest model for accurate mapping and crop area estimates. Additionally, accuracy assessments were performed to ensure confidence in classification accuracy and to allow for comparison with previous classification maps of the area. Classification maps and area estimates produced will be used by the USDA FAS to generate accurate estimates of available commodities as well as assist in policy decision making. The DEVELOP team's classification procedures were shared with partners, to increase crop classification capacity.

Key Terms

remote sensing, radar imagery, accuracy assessment, sensitivity analysis, SAR, random forest, cloud cover, Mato Grosso

2. Introduction

2.1 Background Information Brazil is one of the world's largest agricultural producers, consistently ranking among the top global producers of soybean, corn, and cotton (Valdes, 2021; Ustinova, 2020). Brazil ranked 4th in the USDA's 2019–2020 Estimate and 2020–2021 Forecast for Top Cotton Global Producers and remained the third-largest global corn producer for almost a decade (Ustinova, 2020; Yadav-Pauletti, 2021). In terms of soybean, Brazil is the top exporter in the world, producing 28.6 billion dollars of the crop in 2020. The state of Mato Grosso, located in western Brazil, accounts for the majority of corn, soybean, and cotton production in the nation. In tropical regions such as Mato Grosso, Brazil, dense cloud overage often limits the use of optical imagery that is used to produce crop classification maps and commodity estimates.

Figure 1: Map of study area

2.2 Study Area

The climate in Mato Grosso, Brazil is heavily influenced by its proximity to the Amazon River basin located just to the north of the state. Northern Mato Grosso has a tropical monsoon climate while the south is characterized by a tropical savanna climate. The regions within Mato Grosso where cropland is concentrated straddle the interface of the rainforests and savannas, benefiting from a consistent heavy rainy season to the north in austral summer and a dry austral winter (Arvor et al, 2013). Because of the extreme rainfall, Mato Grosso hosts two agricultural growing seasons per year. The majority of soybean production occurs in the first season, with planting starting in October and harvesting in March. Once the soybeans are harvested, fields undergo crop rotation to second season corn and cotton, which undergo planting in January and are harvested up to September. Second season corn and cotton occupy approximately 60–65% of the previously inhabited soybean acreage, with corn accounting for 55% and cotton taking up 10% of the fields. From 2017–2019, 43% of Brazil's second corn yield was in Mato Grosso (USDA Corn Production, 2019). In the same time frame, 65% of cotton in Brazil was grown in Mato Grosso, with 87% grown in the second season (USDA Cotton Production, 2019; Ustinova, 2020).

2.3 Project Partners & Objectives

In this project, the team partnered with the United States Department of Agriculture (USDA) Foreign Agriculture Service (FAS), International Production Assessment Division for Brazil and the USDA Office of the Chief Economist and World Agricultural Outlook Board. The USDA FAS supports the US in decision making regarding agricultural exports and foreign policy by collecting intelligence on current agricultural conditions in Brazil. The USDA FAS for Brazil relies on spatially comprehensive crop classification maps to produce commodity and yield estimates using optical satellite imagery, however, the dense cloud coverage brought on by the tropical climate of Brazil presents an obstacle to collecting clear

and consistent data, particularly about corn and cotton distribution in Mato Grosso.

To assist the partners with crop classification in Mato Grosso, the team came up with three main objectives. These objectives were 1) to produce 2015 and 2020 crop classification maps for corn and cotton that enhance map inventories limited by cloud coverage, 2) to perform an accuracy assessment to increase partner confidence in the optical and radar data fusion approach to crop classification, and 3) to generate F1 scores used for a more specified analyses of corn and cotton, the classes of interest for crop classification.

2.4 Scientific Basis

Previous literature undertaking crop classification in Brazil has favored a data fusion approach, utilizing radar, topographic, and optical imagery. In a previous study done by Pott et al. (2021) radar data from Sentinel-1, and elevation data from the Shuttle Radar Topographic Mission (SRTM) were utilized to supplement optical data from Sentinel-2, in creating a crop map layer of soybean, corn, and rice in the state of Rio Grande do Sul, Brazil with an overall accuracy of 0.95 (Pott et al., 2021). The practice of data fusion has been used in crop classifications globally in regions where the sole use of optical imagery proves unreliable. Forkuor et al. 2014) combined optical and dual polarized synthetic aperture radar (SAR) data to map cotton crops in northwestern Benin in West Africa, a region characterized by excessive cloud cover (Forkuor et al., 2014). McNairn et al., (2008) suggests the practice of data fusion should be mainstreamed due to the simple risk of relying on a single data source; instead, they integrated SAR and optical imagery to produce multitemporal crop classification maps of Canada given the nation’s diverse landscapes and climate. They found that the sole usage of Landsat 8 imagery across Canada yielded crop classification accuracies consistently below 75% while when fused with SAR data, these percentages increased 3%–18% (McNairn et. al., 2008). Such literature supplemented the team’s confidence in pursuing a fusion of optical, radar, and topographic data for crop classification in Mato Grosso, Brazil.

3. Methodology

3.1 Data Acquisition

To create crop classification maps the team first acquired optical, radar, and topographic input imagery. Using the Python API for Google Earth Engine (GEE) the team compiled optical imagery from Landsat 8 OLI, radar imagery from Sentinel-1 C-SAR, and topographic imagery from the Shuttle Radar Topography Mission (SRTM) (Table 3). From Landsat 8 OLI all bands were utilized and from Sentinel-1 C-SAR both VV and VH bands were used. The team sourced ground truth data from the University of Kansas and S3 Dataset Land Cover Map (Kastens, et al., 2017). From Landsat 8 the team used the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI). NDVI is calculated from the visible and near infrared light reflected by vegetation. Healthy vegetation reflects a plethora of near infrared light, while absorbing most of the visible light. On the other hand, sparse or unhealthy vegetation does the opposite, reflecting more visible light and less near infrared light. Many satellites Vegetation Indices calculate NDVI to quantify vegetation density on Earth. EVI was utilized with the

same justification as NDVI however with the added benefits of mitigation of atmospheric effects due the inclusion of the blue band.

Table 1
Predictors derived from satellite observations and temporal ranges of data used

Predictor Variable	**Source**	**Native Resolution**	**Image Dates**
Normalized Difference Vegetation Index (NDVI)	Landsat 8 OLI	30m	4/1–5/15 (2014)
Enhanced Vegetation Index (EVI)	Landsat 8 OLI	30m	4/1–5/15 (2014)
Surface Reflectance	Landsat 8 OLI	30m	4/1–5/15 (2014)
Elevation	SRTM	30m	N/A
Slope	SRTM	30m	N/A
Vertical Polarization (VV)	Sentinel-1 C-SAR	N/A	4/1–5/15 (2015)

The team received two sets of ground truth data from the partners for the years 2014 and 2020. However, the 2020 ground truth data did not contain enough points to act as a reliable layer to train the machine learner on. Because of this, the team decided to focus on generating 2014 crop classification maps with an emphasis on conveying and building the methodology for future application upon the acquisition of sufficient ground truth data. However only being able to utilize 2014 ground truth data presented an issue for temporal consistency across the layers. Sentinel-1 C-SAR does not have imagery for Brazil in the year 2014. The team remedied this by acquiring 2015 Sentinel-1 C-SAR data. The team then decided upon the monthly temporal range to set all of the imagery and data to.

Due to time restrictions, the team prioritized second-season corn and cotton over first-season soybean for their analysis (Table 2). The team determined the second season range to be from early February to late July to incorporate both second season corn and cotton. However even within this timeframe, the team had to be cognizant of the two crops spectral signatures given their different stages of growth. This was especially a concern with cotton. Due to the distinct stages of growth that cotton transitions through during the growing season, confidently identifying cotton crops presented a challenge without narrowing the temporal

range. During its growth cycle, cotton produces white flowers that soon turn a dark shade of red. After this stage, the flowers are replaced by green bolls, which are rounded seed capsules found on the plant (Xun et al. 2021). To account for different spectral profiles present at each stage of growth, information on cotton phenology was gathered, and additional visual inspections were performed to ensure classification accuracy. Utilizing the Global Inventory Monitoring and Modeling Studies (GIMMS) Global Agricultural Modeling Tool as well as USDA FAS Mato Grosso Production infographics from 2017–2020, the team was able to view the vegetation peaks for both corn and cotton in previous years and make an estimate for the temporal range that would encapsulate both vegetation peaks for corn and cotton. The team determined this range to be from April 1st to May 15th. During data preprocessing and cleaning, the team set this as the temporal range for all of the imagery layers in 2014 and for 2015Senintel-1 C-SAR data.

Table 2
Crop type and Season used to Determine the Temporal Range for Training Data (USDA GIMMS; USDA Brazil Crop Production Maps 2017–2019)

CROP TYPE	PLANT	MID-SEASON	HARVEST
CORN (FIRST SEASON)	Oct–Dec	Jan–Feb	Feb–Jul
CORN (SECOND SEASON)	Jan–Mar	Mar–Jun	Jun–Sep
SOYBEAN	Oct–Jan	Jan–Mar	Mar–Jun
COTTON (SECOND SEASON)	Jan–Feb	Feb–Jul	Jul–Sept

3.2 Data Processing

After the acquisition of training and validation data, cleaning and preprocessing in Google Earth Engine and Google Colab, and the combination of relevant features, the team undertook direct steps to produce crop classification maps, accuracy matrices, and F1 scores. To pursue all of the objectives, the team utilized a random forest (RF) model (Figure 3). RF Modeling is a classification algorithm consisting of many decision trees. It uses bagging and feature randomness to create individual trees which creates a forest of uncorrelated trees whose prediction together is more accurate than any individual tree (Dimitriadis and Liparas, 2018). Utilizing this model in Google Colab produces maps, confusion matrices, and F1 values. Within the RF model, 696 training points were generated using provided data, with each of the 8 classes receiving an equal distribution of points. This training data was then split up between training and validation data with 70%

going to training, and 30% to validation. RF has been utilized in a number of crop classification studies such as Tatsumi et al (2015), Ok et al (2017), and Reynolds et al (2016) in addition to previously mentioned studies that have specifically undertaken a data fusion methodology in crop classification like Forkour et al (2014) and Ajadi et al (2020), the latter of which produced crop classification maps for Mato Grosso, Brazil as well.

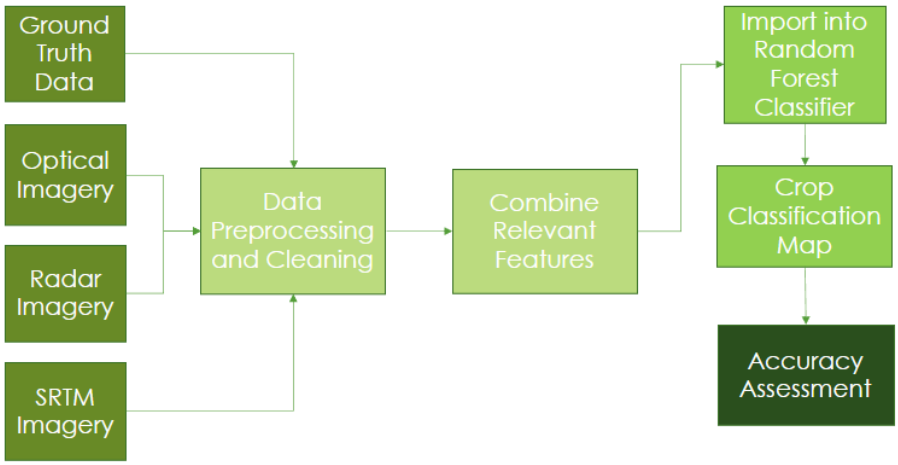


Figure 3: Workflow diagram depicting preprocessing of remotely sensed imagery, combination of relevant features, and output of classification maps from the machine learning algorithm.

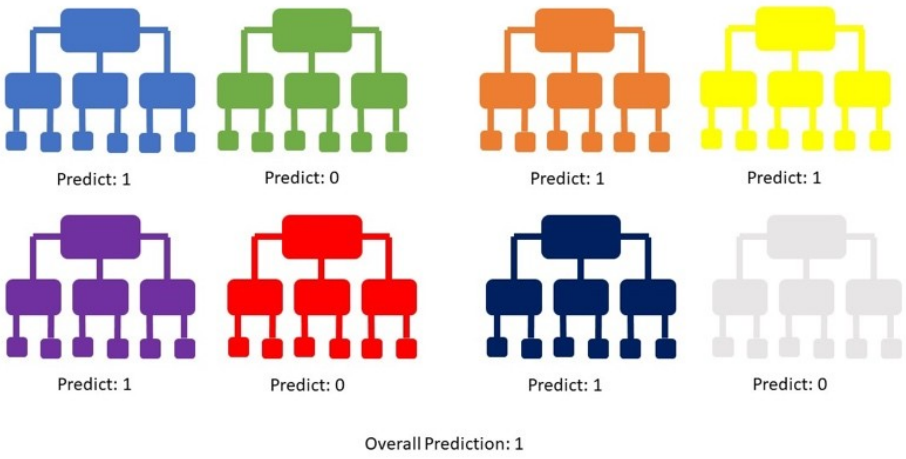


Figure 4: Depiction of how a classifier works using a Random Forest algorithm.

3.2.1 Crop Classification Maps

To identify vegetation indices for the study area, the team processed data from Landsat 8 OLI, Sentinel 1 C-SAR, SRTM, and EVI and NDVI. Using Google Colab, NDVI and the Enhanced Vegetation Index (EVI) were calculated and deployed in the team's stack of data used as input within the Random Forest Model (RF). Once the data was acquired from Google Earth Engine using the Google Colab's Python API, the team applied a bitmask to mask cloud cover within Mato Grosso. Then, the team processed Landsat 8 OLI data to display the NDVI and EVI using the equations below. To prepare for analysis, the team also generated a cropland/no cropland map (Figure B2). This was generated in ArcGIS Pro using a shapefile of different crop types within Mato Grosso from the University of Maryland Global Land Analysis and Discovery Laboratory (GLAD) (2019).

Table 3
Equations for predictor variable indices within the RF

Predictor Variable	**Equation**
Normalized Differenced Vegetation Index	$\frac{NIR - RED}{NIR + RED}$
Enhanced Vegetation Index	$2.5 \times \frac{NIR - RED}{NIR + 6 \times RED - 7.5 \times BLUE + 1}$

3.2.2 Accuracy Assessments and F1 scores
The team initially planned to perform a sensitivity analysis in order to determine which variables in the classifier contributed most to its accuracy; however, due to time constraints, the team opted to calculate F1 values in its place. The F1 values were used for a more specified analysis of the classes of interest, and they provide information on how each class is performing within the model by examining the harmonic mean of precision and recall.

3.3 Data Analysis
In order to determine the significance of the addition of radar and topographic data in crop classification using the RF, the team generated three maps with corresponding training and validation confusion matrices, accuracy scores, and F1 values within Google Colab. The first map only utilized optical data to gauge what accuracies the partners were dealing with. The second version fused optical and radar data to see the impact adding radar had on accuracy. The third iteration added topographic data to optical and radar to see the impact or necessity for topographic data in Mato Grosso Crop Classification.

The confusion matrices provided the team with a baseline visualization of how well the machine learner performed with the classes that it was trained in. Confusion matrices show how many times the machine learner classified classes correctly, as well as depicting false positives and false negatives. Once normalized, these matrices provided the team with an understanding of how well corn and cotton classified in every combination of data layers and how well they compared to the classification of other classes. Obstacles regarding spectral similarities were also

revealed in this process, identifying classes in which corn and cotton had a tendency to be falsely classified. Training and validation matrices were the foundation for the generating of accuracy and F1 scores, allowing the team to analyze the RF performance trends with more detail and confidence.

The training and validation accuracy percentages provided overall insight to how well the machine learner performed in every iteration. Accuracy is calculated using the equation below where TP stands for True Positives, TN stands for True Negatives, FP stands for False Positives, and FN stands for False Negatives (Google, 2022).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

There are two types of accuracies that the team analyzed, training and validation accuracy. Training accuracy shows the accuracy when the RF is run on the set of data used to train it. The validation accuracy is calculated upon running the RF on data it is not trained on, allowing the team to assess how well the RF would react to new information. Inherently, the training accuracy is higher than the validation accuracy, but having the two values is important in identifying levels of overfitting. The greater the difference between the validation and training accuracies, the greater the overfitting tendency of the model is. This can be addressed through hyperparameter optimization. Given the time constraint and the team's main goal of seeing how the introduction of radar and topographic data would impact accuracy from using a sole optical approach, hyperparameter optimization was not a focus. The overall accuracy gave the team insight into the general performance of the machine learner, but to understand its performance in enhancing corn and cotton classification, the team looked at the F1 values.

In order to better understand the predictive accuracy of the machine learner, the team calculated F1 values for corn and cotton. The F1 value can provide valuable insight into how well the model classified data by calculating the harmonic mean of two important variables: precision, and recall (Lipton, 2015). Precision can tell us the extent of error in the model that was caused by false positives, while recall tells us how much error can be attributed to false negatives. The harmonic mean is particularly useful in machine learning because it ensures that an appropriately low score is given to a class if either precision or recall is very low. This is done on a scale of 0-1. This contrasts with the arithmetic mean, where a low recall or precision score could still produce a relatively high F1 value.

The mathematical definition of an F1 value can be seen below where r stands for recall, and p stands for precision (Lipton, 2015):

$$F1 = 2/(\frac{1}{r} + \frac{1}{p})$$

4. Results & Discussion

4.1 Analysis of Results

For each classification model used by the team, a map was produced that depicts the land coverage of Mato Grosso classified into 8 separate classes. These visualizations provide insight into the initial performance of the algorithm by showing which classes are most abundant. The first trials were indicative of slight

overfitting which called the team to perform hyperparameter optimization. These optimizing techniques improved the output of confusion matrices and accuracy scores for both the training and validation data. These parameters included having a max number of 1000 decision trees, a max number of 61 nodes, and at least 5 variables per split. Below are the final maps that the team produced with the different combinations of data.

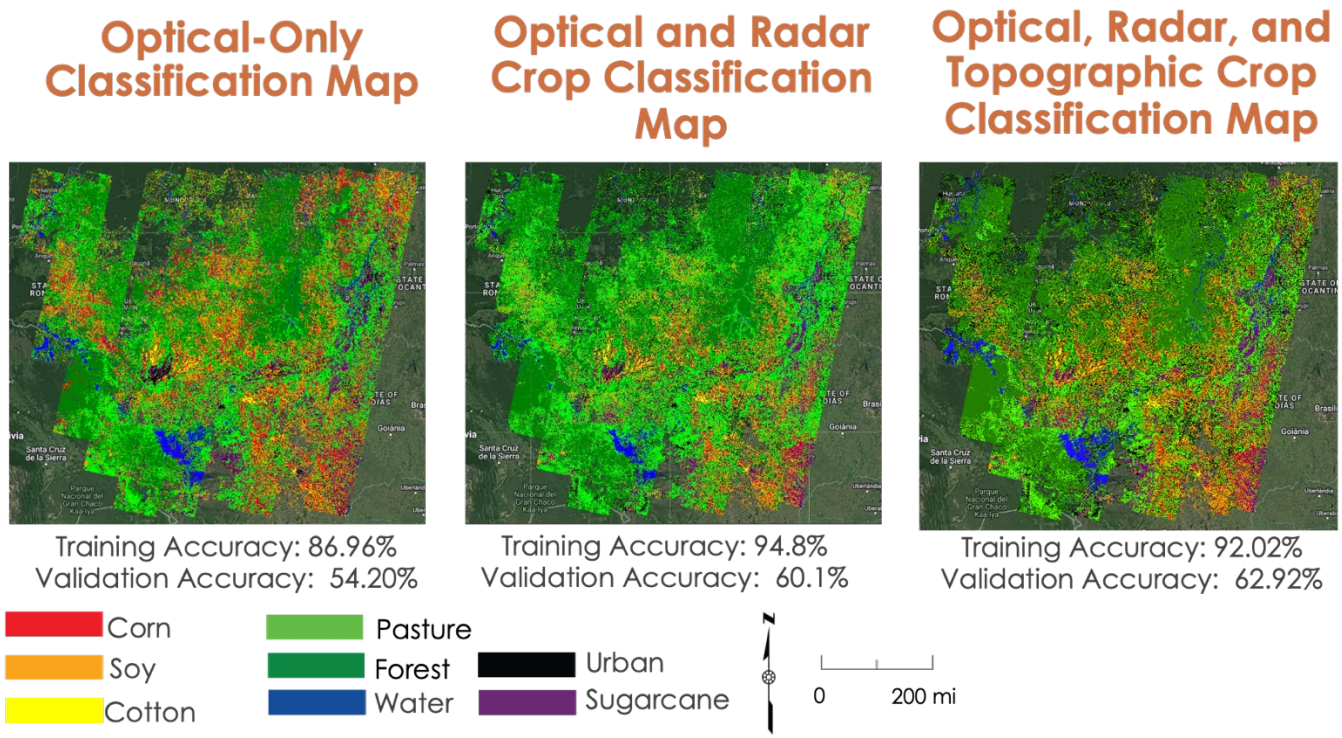
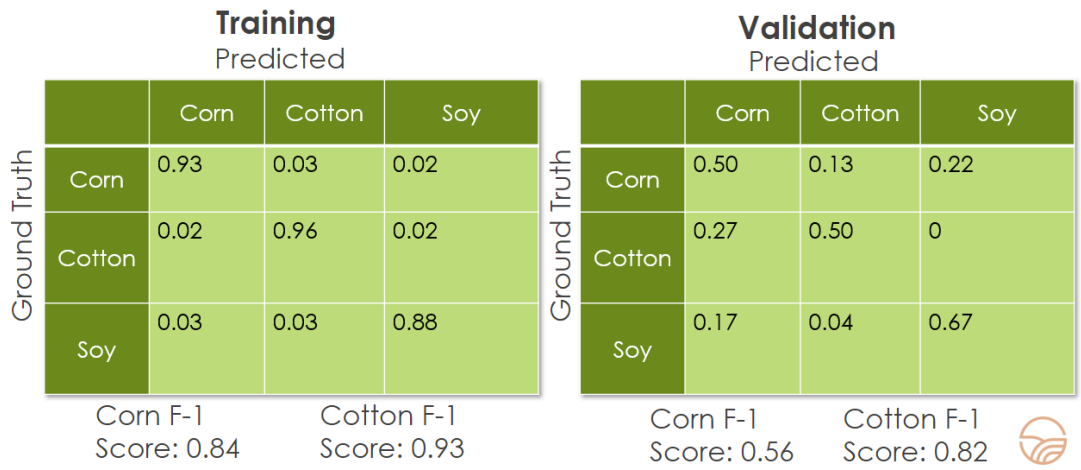


Figure 4: Maps conveying optical-only, Optical & Radar, and Optical, Radar and Topographic classifications

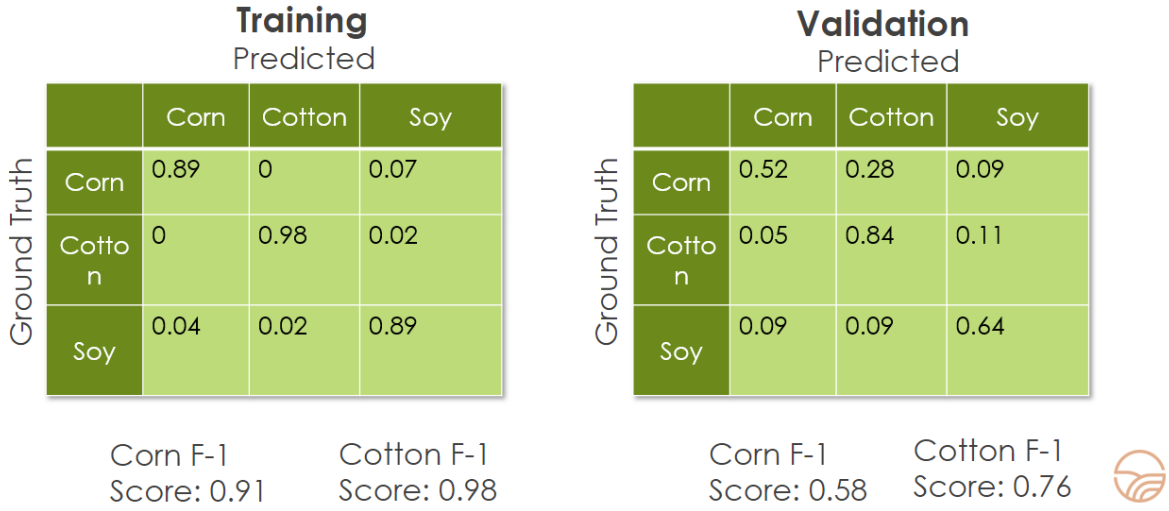
With the classified map produced using optical imagery alone, an exceptionally large amount of soy corn, and soy were represented. Additionally, the algorithm classified less land as pasture or forest than the other models (Figure 4). The classification map produced using optical and radar imagery, shows a decrease in corn and soy. An additional increase in the amount of pasture and forest land is also present in this map (Figure 4). This map, much like the optical only classification, has much more corn, soy, and cotton present. This could be caused by the overfitting of the model. However, the team's confusion matrices and F1 scores give a detailed look at classification accuracy across the various data fusions (Tables 4, 5, and 6). To fine tune the RF algorithm, the team created confusion matrices for each classification. Confusion matrices provide insight into the performance of a given classification model by showing what each pixel was classified as by the algorithm (Appendix A).

Table 4
Optical-Only Normalized Training and Validation Matrices where values on the diagonal represent recall.



When looking at the matrices generated by the optical only model, there is an overall training accuracy of 86.96%, while the validation matrix had an accuracy of 54.20%. Looking at the validation matrix, cotton and corn were observed to have particularly poor levels of accuracy. For the optical only classification model, the training F1 values were 0.84 for corn, and 0.93 for cotton. The validation F1 values were 0.56 for corn, and 0.82 for cotton.

Table 5
Optical and Radar Normalized Training and Validation Matrices where values on the diagonal represent recall.



When radar data is added to optical imagery, the overall training accuracy increased 7.84% to 94.8% and the validation accuracy increased 5.9% to 60.1%. Focusing on the results from the validation matrix, corn classified 2% better than the optical-only classification and cotton had a significant increase in accuracy of 24%. For the optical and radar model, training F1 values were 0.91 for corn and

0.98 for cotton. Validation F1 scores for this model were 0.58 for corn and 0.76 for cotton.

Table 6
Optical, Radar, and Topographic Normalized Training and Validation Matrices

Training

Predicted

Ground Truth	Corn	Cotton	Soy
Corn	0.78	0.03	0.15
Cotton	0.01	0.97	0.01
Soy	0.03	0.09	0.95

Corn F-1 Score: 0.85 Cotton F-1 Score: 0.97

Validation

Predicted

Ground Truth	Corn	Cotton	Soy
Corn	0.52	0.22	0.09
Cotton	0.05	0.84	0.10
Soy	0.09	0.09	0.64

Corn F-1 Score: 0.58 Cotton F-1 Score: 0.76

* Values on the diagonal represent recall.

In the final training confusion matrix, there is a training accuracy of 92.02% which is a 2.78% decrease from the matrix of just optical and radar data. With the validation confusion matrix, there is an accuracy of 62.92%, which is a 2.82% increase from the accuracy from optical and radar without topographic data. Overall, Validation accuracy did not increase for corn, cotton or soy from the optical-radar fused approach. In the final model, which utilized optical, radar, and topographic imagery, training F1 values for corn were 0.86, while cotton was 0.91. For this model's validation F1 scores, corn was 0.58, and cotton was 0.76.

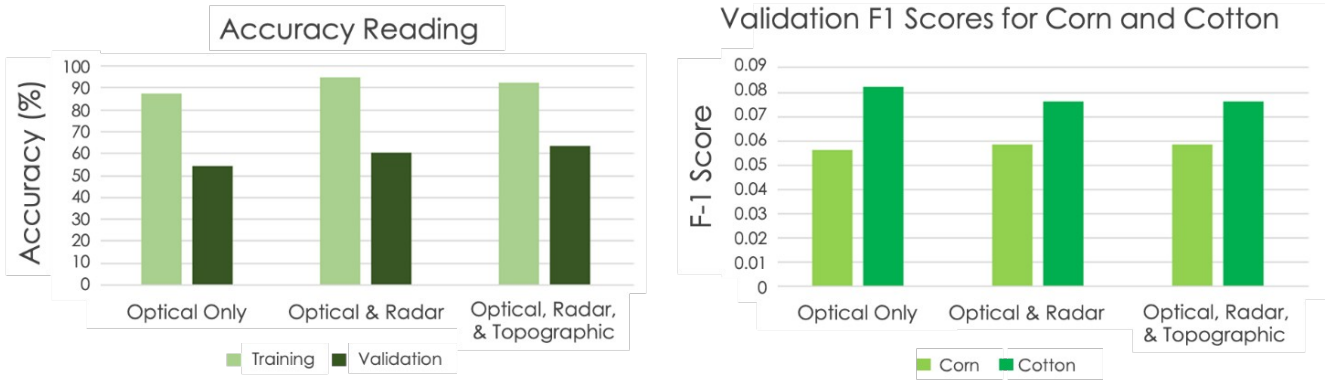


Figure 5: Synthesized Overall Training and Validation Accuracy Readings and Validation F1 scores for Corn and Cotton

When looking at the trends of accuracy and F1 scores, certain patterns are revealed as radar and topographic data layers are added. When looking at the accuracy readings, training accuracy increases upon the addition of radar data and then decreases slightly when topographic data is further added. However, validation accuracy, which provides insight into the ability for the RF to classify new data, was the team's primary interest given that it provides the most

predictive power for future applications of the RF. The validation accuracy increased upon adding radar data and slightly increased again once adding topographic data.

The team pulled from the validation set of data to inform analysis of corn and cotton specifically. The corn F1 score increased upon the addition of radar data and stayed the same when topographic data was further added. Even though the cotton F1 score was higher, it actually decreased upon adding the radar data and also stayed the same upon adding the topographic data. The lack of F1 impact from adding the topographic data was anticipated by the team due to Mato Grosso's plateaued farmlands. This is indicative of a normal machine learning phenomenon where certain class accuracies are sacrificed to enhance accuracies of other classes.

4.2 Future Work

This project laid the groundwork for machine learning in crop classification in Mato Grosso, Brazil, and identified many different avenues to enhance its accuracy moving forward. Due to the time restrictions of the project, the RF model was selected for use without comparing the advantages of alternate algorithms. A comparison of different machine learning algorithms' results would be useful and could provide feedback on best practices for future classification efforts.

Future work on this project should incorporate rainfall data from Mato Grosso to better understand how rainfall may have affected the growth, harvest, or health of crops produced in the area. Rainfall data may also help identify areas in the classification maps that may be experiencing cloud contamination. This context would allow us to examine how great an effect cloud contamination may have had on the team's overall classification accuracy. Furthermore, analyses of cropland at various stages of the growing season, as well as across multiple different years, would provide partners with valuable information on the changes in variety and coverage of crops in the area.

The limited availability of ground truth data in Mato Grosso posed an issue for the team in training a more recent RF model. Ground truth data in Brazil is limited due to consistent agricultural expansion and the associated time and expenditure. Future efforts aiming to classify crops in more recent years using machine learning could draw from non-local ground truth data to supplement local ground truth data. This has been done successfully for soybean and corn classification in Brazil by Ajadi et al. (2021), pulling from the USDA National Agricultural Statistics Service (NASS) Cropland Data Layer in GEE to aid in training the RF model.

Incorporating Sentinel-2 MSI and additional vegetation indices could also increase accuracy. Three additional indices of interest are the Green Normalized Difference Vegetation Index (GNDVI), Green Chlorophyll Vegetation Index (GCVI), and the Normalized Difference Red Edge Index (NDRE) given their abilities to function in dense vegetation environments and enhanced sensitivity. Moving forward, the inclusion of Sentinel-2 MSI imagery is possible as well. The team was unable to use Sentinel-2 MSI data in this project given the temporal limitations of the ground truth data. However, these data would be a helpful addition to future iterations of the machine learner as ground truth data becomes more readily available in the

region. Performing a traditional sensitivity analysis with the addition of these potential data inclusions would help to identify the relative importance of their presences, which the team was not able to do during this project timeframe.

5. Conclusions

This project's findings provided a more effective method for generating accurate crop classification maps. By utilizing Landsat 8 OLI, Sentinel-1 C-SAR, and SRTM imagery, the team produced crop classification maps displaying the extent of cotton and corn growth in Mato Grosso. The projects classification methods can be leveraged to improve the scale and accuracy of crop classification maps in tropical regions that have frequent cloud cover. Results benefited the USDA Foreign Agriculture Service, International Production Assessment Division, as well as the USDA Office of the Chief Economist and World Outlook Board by providing comprehensive and accurate coverage of important agricultural areas that can be used to enhance commodity estimates impacting US and foreign policy.

The accuracy assessment provided confidence in the results as the team saw an overall increase in both training and validation accuracy when radar imagery was added to the classification model. Additionally, the team investigated whether adding SRTM topographical data to the model would have an impact on overall accuracy. Through the results, the team determined that due to Mato Grosso's consistently plateaued geography, adding topographical data did not have a significant overall effect on the accuracy, however the team believes that in areas where there are large elevation differences throughout, SRTM imagery may provide an improvement. But the team also had to consider the cotton F1 scores, which actually decreased when applying additional data layers. For corn, even though accuracy and F1 scores never dropped, it did not increase significantly when adding the additional data layers. This suggests that the model has provided a basis for crop classification, however it may need continued alterations to reliably classify both corn and cotton.

The project evaluated the benefits provided by a fusion radar and optical approach to crop classification in order to improve classification and commodity estimates. Due to the changes in agricultural areas in Brazil, it is crucial to have up to date and accurate information providing data on crop type and quantity in order to make informed policy decisions. This classification method will provide enhanced classification capabilities in areas that have struggled to produce comprehensive and accurate classification maps due to atmospheric interferences present.

6. Acknowledgments

The Mato Grosso Agriculture team would like to thank the following for their constant support and guidance, each of whom provided invaluable advice throughout the term

- Kathleen Lange (NASA DEVELOP National Program)
- Dr. Garrett Graham (NOAA National Centers for Environmental Information, North Carolina Institute for Climate Studies)
- Molly Woloszyn (NOAA National Centers for Environmental Information, National Integrated Drought Information System)
- Dr. Sunita Yadav-Pauletti (USDA Foreign Agricultural Service, International Production Assessment Division for Brazil)

- Mark Brusberg (USDA Office of the Chief Economist and World Agricultural Outlook Board)

Finally, we thank DEVELOP Project Coordination Fellows Tamara Barbakova and Sophia Skoglund for their guidance and edits.

This material contains modified Copernicus Sentinel data (2015), processed by ESA.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C.

7. Glossary

Earth observations - Information about the Earth's physical, chemical, and biological systems collected remotely through satellites and sensors
Surface reflectance – A measure of the amount of light reflected off of the Earth's surface
SAR – Synthetic aperture radar
NDVI – Normalized difference vegetation index
NDRE – Normalized difference red edge index
GCVI – Green chlorophyll vegetation index
EVI2 – Enhanced vegetation index 2

8. References

Ajadi, O.A., Barr, J., Liang S., Ferreira, R., Kumpatla, S. P., Patel, R., Swatantran, A. (2021). Large-scale crop type and crop area mapping across Brazil using synthetic aperture radar and optical imagery. *International Journal of Applied Earth Observation and Geoinformation, 97*, https://doi.org/10.1016/j.jag.2020.102294

Ok, A.O., Akar, O., & Gungor, O. (2017). Evaluation of random forest method for agricultural crop classification. *European Journal of Remote Sensing, 45*(1), https://doi.org/10.5721/EuJRS20124535

Anulacion, M. (2019, August 12). *Commodity Intelligence Report- Brazil's 2018/19 Record Cotton Harvest Underway*. USDA Foreign Agricultural Service. https://ipad.fas.usda.gov/highlights/2019/08/brazil/index.pdf

Arvor, D., Ronchail, J., Dubreuil V., Simoes M. (2014). Spatial patterns of rainfall regimes related to levels of double cropping agriculture systems in Mato Grosso (Brazil). *International Journal of Climatology, 34*(8), 2622-2633. https://doi.org/10.1002/joc.3863

Dimitriadis, S. I. & Liparas, D. (2018). How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database. *Neural Regeneration Research*, *13*(6), 962–970. https://doi.org/10.4103/1673-5374.233433

ESA (2014, October 3). *Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected, log scaling.* Earth Engine Data Catalog. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD

Forkuor, G., Conrad, C., Thiel, M., Ullmann, T., Zoungrana, E. (2014). Integration of Optical and Synthetic Aperture Radar Imagery for Improving Crop Mapping in Northwestern Benin, West Africa. *Remote Sensing Journal, 6*(7) https://doi.org/10.3390/rs6076472

University of Maryland Global Land Analysis and Discovery Laboratory (2019). *Global cropland expansion in the 21st century.* Global Land Analysis and Discovery. https://glad.umd.edu/dataset/croplands

Google. (2022, July 18). *Classification: Accuracy*. Machine Learning. Retrieved August 10, 2022, from https://developers.google.com/machine-learning/crash-course/classification/accuracy

Kastens J.H., Brown J.C., Coutinho A.C., Bishop C.R., Esquerdo J.C.D.M. (2017) Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. *PLOS ONE*, 12(4), e0176168. https://doi.org/10.1371/journal.pone.0176168

Lipton, Z.C., Elkan, C., Naryanaswamy, B. (2015). Optimal Thresholding of Classifiers to Maximize F1 Measure, Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2014. Lecture Notes in Computer Science, vol 8725. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44851-9_15

McNairn, H., Champagne, C., Shang, J., Holmstrom, D., Reichert, G. (2009). Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(5), 434-449. https://doi.org/10.1016/j.isprsjprs.2008.07.006

NASA, USGS (2000, February 11). *NASA SRTM Digital Elevation 30m.* Earth Engine Data Catalog.

https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003

Pott, L. P., Carneiro Amando, T.J., Schwalbert, R.A., Corassa, G.M., Ciampitti, I.A. (2021). Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing*, *17*, 196-210. https://doi.org/10.1016/j.isprsjprs.2021.04.015

Reynolds, J. (2016). Using Remote Sensing and Random Forest to Assess the Conservation Status of Critical Cerrado Habitats in Mato Grosso do Sul, Brazil. *Land*, *5*(2). https://doi.org/10.3390/land5020012

Tatsumi, K., Yamashiki, Y., Canales Torres, M.A., Ramos Taipe, C.L. (2014). Crop Classification of upland fields using Random forest of time-series Landsat 7 ETM+ data, *Computers and Electronics in Agriculture, (115)* 171-179. http://dx.doi.org/10.1016/j.compag.2015.05.001

United States Department of Agriculture, & The Brazilian Institute of Geography and Statistics - Produção Agrícola Municipal. (2019). *Brazil: Cotton Production.* https://ipad.fas.usda.gov/rssiws/al/crop_production_maps/Brazil/Municipality/Brazil_Cotton.png

United States Department of Agriculture, & The Brazilian Institute of Geography and Statistics- Produção Agrícola Municipal. (2019). Brazil: Second Season Corn Production. https://ipad.fas.usda.gov/rssiws/al/crop_production_maps/Brazil/Municipality/Brazil_SecondSeason_Corn.png

USDA (2022), GIMMS Global Agricultural Monitoring, USDA IPAD. https://glam1.gsfc.nasa.gov/

USGS (2013, March 18). *USGS Landsat 8 Level 2, Collection 2, Tier 1.* Earth Engine Data Catalog. https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2

Ustinova, E. (2020, November 25). *Cotton and Products Update - USDA Brazil*. Global Agricultural Information Network. https://usdabrazil.org.br/wp-content/uploads/2021/05/Cotton-and-Products-Update_Brasilia_Brazil_11-30-2020.pdf

Xun Lan, Jiahua Zhang, Dan Cao, Jingwen Wang, Sha Zhang, Fengmei Yao, (2021, February). Mapping cotton cultivated area combining remote sensing with a fused representation-based classification algorithm, *Computers and Electronics in Agriculture*, *181.* https://doi.org/10.1016/j.compag.2020.105940

Valdes, C. (2021, October 20). *International Markets & U.S. Trade: Brazil*. USDA Economic Research Service. https://www.ers.usda.gov/topics/international-markets-u-s-trade/

Yadav-Pauletti, S. (2021, September 13). *Commodity Intelligence Report: Brazil Corn 2020/21: A Delayed Start in Planting and Severe Drought Reduce Yields*. USDA Foreign Agriculture Service. https://ipad.fas.usda.gov/highlights/2021/09/Brazil/index.pdf

9. Appendices

Appendix A

Table A1

Crop type and Season used to Determine the Temporal Range for Training Data

CROP TYPE	PLANT	MID-SEASON	HARVEST
CORN (FIRST CROP) < 6%	Oct–Dec	Jan–Feb	Feb–Jul
CORN (SECOND CROP) 43%	Jan–Mar	Mar–Jun	Jun–Sep
SOYBEAN 25%	Oct–Jan	Jan–Mar	Mar–Jun
COTTON (FULL SEASON) 65%	Nov–Dec	Dec–May	May
COTTON (SECOND) (PLACEHOLDER FOR % YIELD FOR SECOND SEASON THEORETICALLY, 87% OF 65%)	Jan–Feb	Feb–Jul	Jul–Sept

Table A2

Optical-Only Confusion Matrix (Training)

	Corn	Soy	Cotton	Pasture	Forest	Water	Urban	Sugarcane
Corn	**0.93**	0.02	0.03	0.02	0	0	0	0
Soy	0.03	**0.88**	0.03	0	0	0	0.02	0.03
Cotton	0.02	0.02	**0.96**	0	0	0	0	0
Pasture	0.05	0.06	0	**0.77**	0.03	0.01	0.02	0
Forest	0.02	0.02	0	0	**0.95**	0.02	0	0
Water	0.02	0	0	0	0.1	**0.86**	0	0.02
Urban	0.02	0.05	0	0.03	0.02	0	**0.84**	0.03
Sugarcane	0.06	0.06	0	0.02	0.02	0	0.02	**0.85**

Table A3

Optical-Only Confusion Matrix (Validation)

	Corn	Soy	Cotton	Pasture	Forest	Water	Urban	Sugarcane
Corn	**0.50**	0.22	0.13	0.05	0	0	0.05	0.05
Soy	0.17	**0.67**	0.04	0	0	0	0.04	0.08
Cotton	0.27	0	**0.50**	0.05	0.14	0	0	0.05
Pasture	0.23	0.09	0	**0.41**	0.05	0.05	0.05	0.09
Forest	0.08	0	0	0.12	**0.56**	0.16	0	0.08
Water	0	0	0	0.11	0	**0.84**	0	0.05
Urban	0.04	0.22	0	0.04	0	0	**0.61**	0.09
Sugarcane	0.12	0.12	0	0.24	0.08	0.04	0.08	**0.32**

Table A4

Optical and Radar Confusion Matrix (Training)

	Corn	Soy	Cotton	Pasture	Forest	Water	Urban	Sugarcane
Corn	**0.89**	0.07	0	0	0	0	0.01	0.01
Soy	0.04	**0.89**	0.02	0	0	0.04	0	0.18
Cotton	0	0.02	**0.98**	0	0	0	0	0
Pasture	0.04	0.05	0	**0.86**	0.04	0.02	0	0
Forest	0	0	0	0	**1**	0	0	0
Water	0	0	0	0	0.02	**0.98**	0	0
Urban	0	0.04	0	0	0.04	0	**0.96**	0
Sugarcane	0	0.05	0	0	0.02	0	0	**0.93**

Table A5

Optical and Radar Confusion Matrix (Validation)

	Corn	Soy	Cotton	Pasture	Forest	Water	Urban	Sugarcane
Corn	**0.52**	0.09	0.22	0	0	0	0	0.17
Soy	0.09	**0.64**	0.09	0	0	0.05	0	0.14
Cotton	0.05	0.11	**0.84**	0	0	0	0	0
Pasture	0.05	0.19	0	**0.48**	0.10	0	0	0.19
Forest	0	0	0	0.05	**0.74**	0.16	0	0.05
Water	0	0	0	0.11	0.11	**0.72**	0.06	0
Urban	0.04	0.14	0	0	0.18	0.04	**0.61**	0.07
Sugarcane	0.04	0.18	0	0	0.19	0	0.04	**0.57**

Table A6

Optical, Radar, and Topographic Confusion Matrix (Training)

	Corn	Soy	Cotton	Pasture	Forest	Water	Urban	Sugarcane
Corn	**0.78**	0.15	0.03	0	0	0	0.02	0.02
Soy	0.03	**0.95**	0	0	0	0.02	0	0
Cotton	0.01	0.01	**0.97**	0	0	0	0	0
Pasture	0	0.09	0	**0.84**	0.03	0.02	0	0.02
Forest	0	0.02	0	0	**0.98**	0	0	0
Water	0	0	0	0	0.04	**0.96**	0	0
Urban	0	0.06	0	0	0	0	**0.94**	0
Sugarcane	0	0.06	0	0	0	0	0	**0.94**

Table A7

Optical, Radar, and Topographic Confusion Matrix (Validation)

	Corn	Soy	Cotton	Pasture	Forest	Water	Urban	Sugarcane
Corn	**0.52**	0.09	0.22	0	0	0	0	0.17
Soy	0.09	**0.64**	0.09	0	0	0.04	0	0.13
Cotton	0.05	0.10	**0.84**	0	0	0	0	0
Pasture	0.05	0.19	0	**0.48**	.09	0	0	.19
Forest	0	0	0	0.05	**0.74**	0.16	0	0.05
Water	0	0	0	0.11	0.11	**0.72**	0.06	0
Urban	0.04	0.14	0	0.11	0.04	0	**0.60**	0.07
Sugarcane	0.04	0.18	0	0	.18	0	0.04	**0.57**