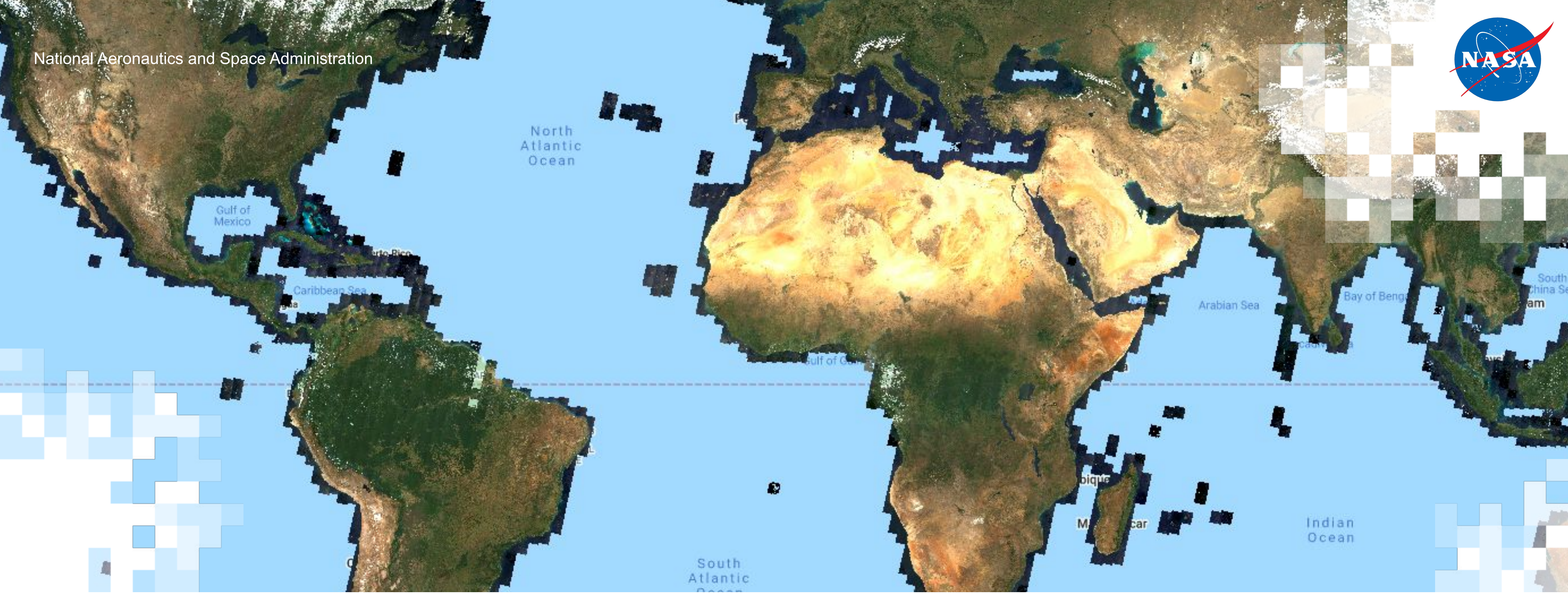
National Aeronautics and Space Administration



Part 2: Using Google Earth Engine for Land Monitoring Applications

Zach Bengtsson, Britnay Beaudry, Juan Torres-Pérez, and Amber McCullum

June 23, 2021

Course Structure and Materials

- Three 2-hour sessions on June 16, 23, & 30
- Sessions will be presented once in English 12:00-14:00 EDT
- One Google Form homework due on July 14
- Webinar recordings, PowerPoint presentations, and the homework assignment can be found after each session at:
  - https://appliedsciences.nasa.gov/join-mission/training/english/arset-using-google-earth-engine-land-monitoring-applications
  - Q&A following each lecture and/or by email at:
    - bengtsson@baeri.org
    - juan.l.torresperez@nasa.gov
    - amberjean.mccullum@nasa.gov

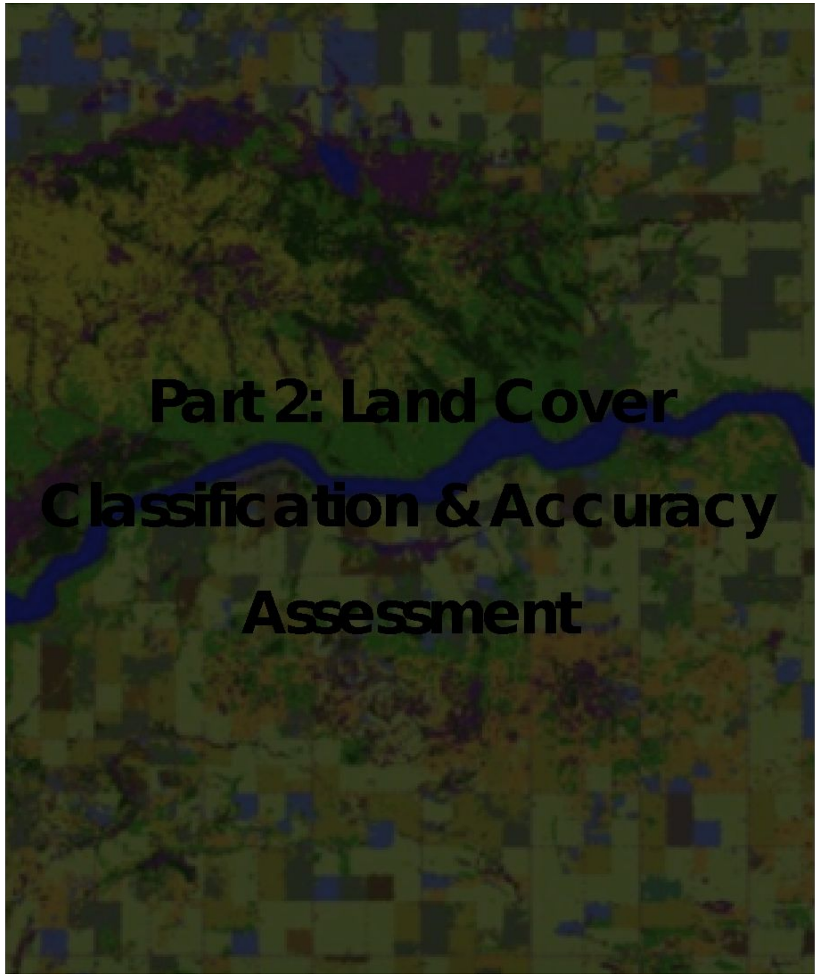
Webinar Agenda

Part 1: Google Earth Engine Basics & General Applications

Part 2: Land Cover Classification & Accuracy Assessment

Part 3: Time Series Analysis & Change Detection

Webinar Agenda




Zach Bengtsson

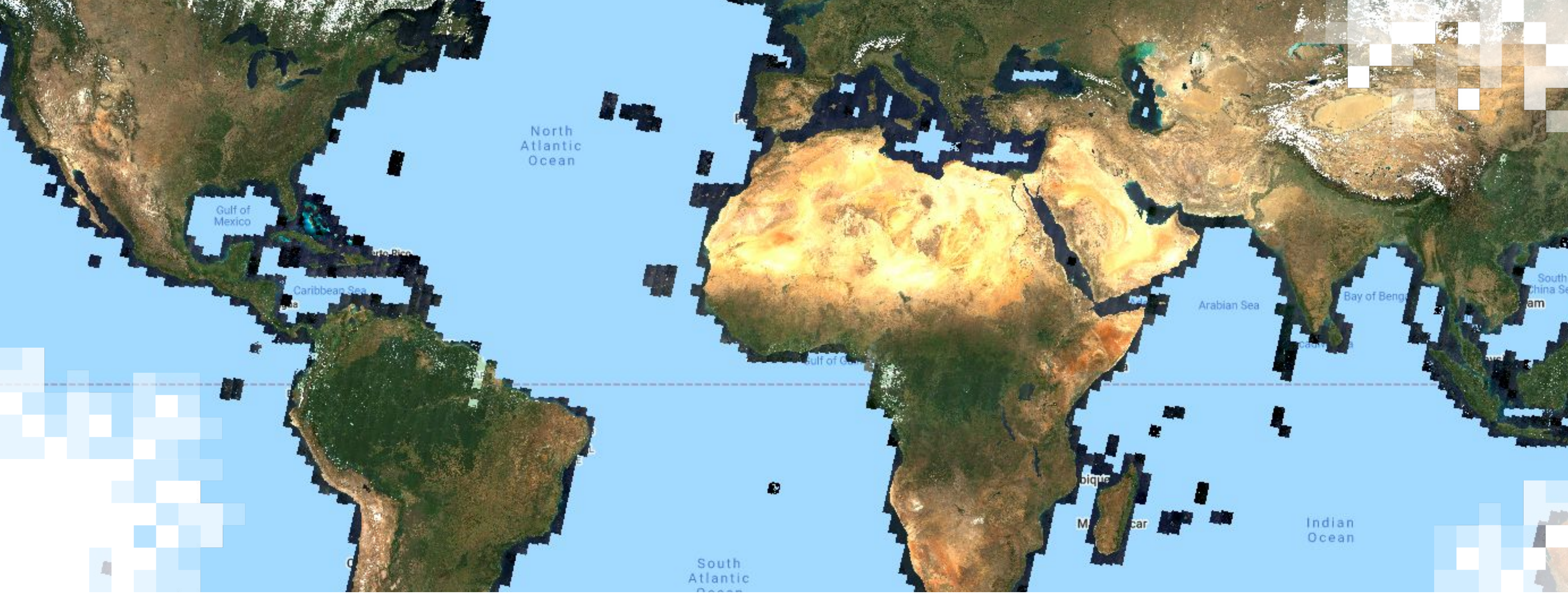
Britnay Beaudry
NASA DEVELOP
Guest Speaker


Juan Torres-Pérez


Amber McCullum

Session Outline

- **Land Cover Classification and Accuracy Assessment Overview**
  - Land cover classification and accuracy assessment basics
  - Classification methods available in GEE
  - Techniques for accuracy assessment and validation in GEE
- **GEE Code Editor Activity**
  - Landsat data retrieval and imagery preparation
  - Supervised land cover classification
  - Accuracy assessment of land cover classifications
- **Question & Answer Session**



Land Classification and Accuracy Assessment Overview

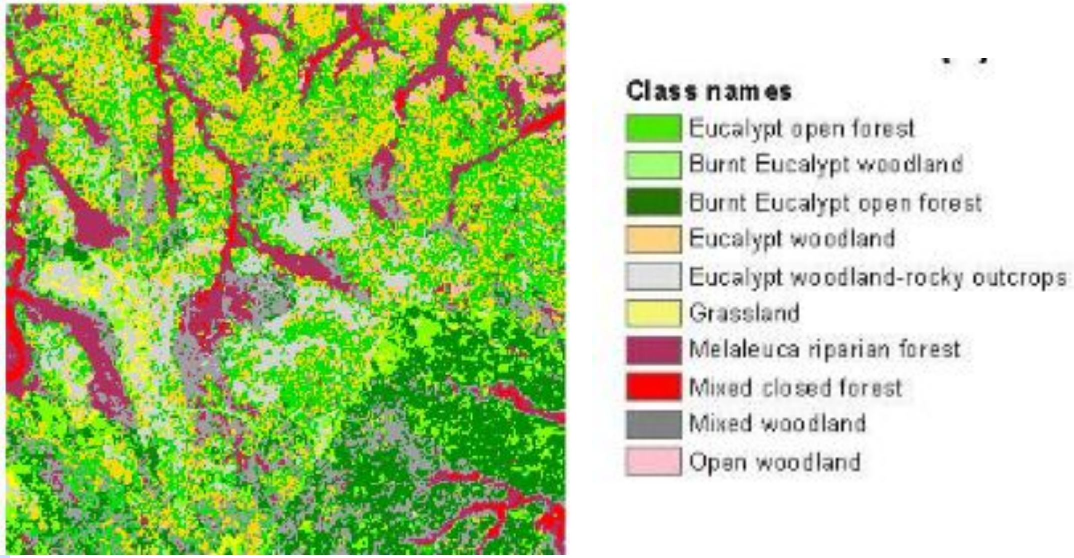
Land Cover Classification Overview

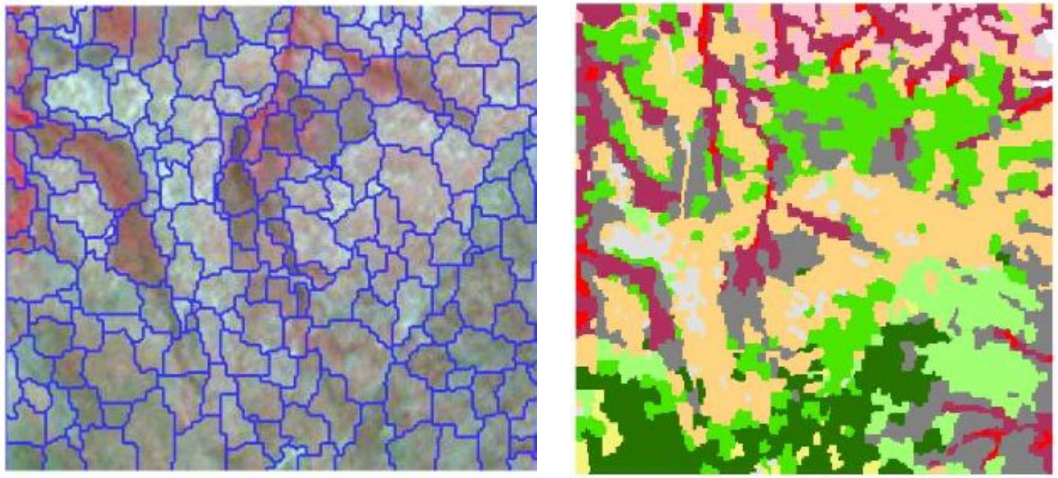
- Land cover classification is the process of grouping spectral classes and assigning them informational class names.
- Spectral Classes:
  - Groups of pixels that are uniform with respect to their pixel values in several spectral bands
- Informational classes
  - Categories of interest to users of the data (like water, forest, urban, agriculture, etc.)

Classic example of a land cover assessment from the USGS National Land Cover Database. Image Credit: USGS

Image Classification

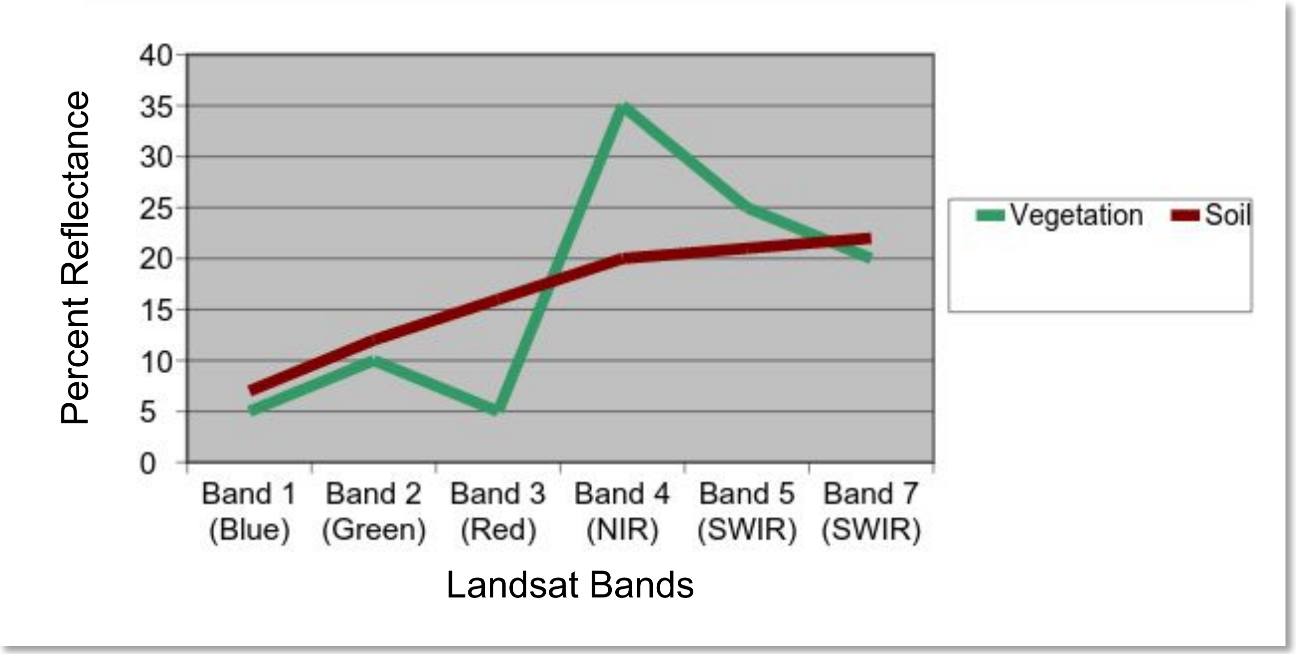
- **Pixel-Based**
  - Each pixel is grouped in a class
  - Useful for multiple changes in land cover within a short period of time
  - Best for complete data coverage and ensuring time series consistency at the pixel level
- **Object-Based**
  - Pixels with common spectral characteristics are first grouped together (segmentation)
  - Useful for reducing speckle noise in radar images and high-resolution imagery





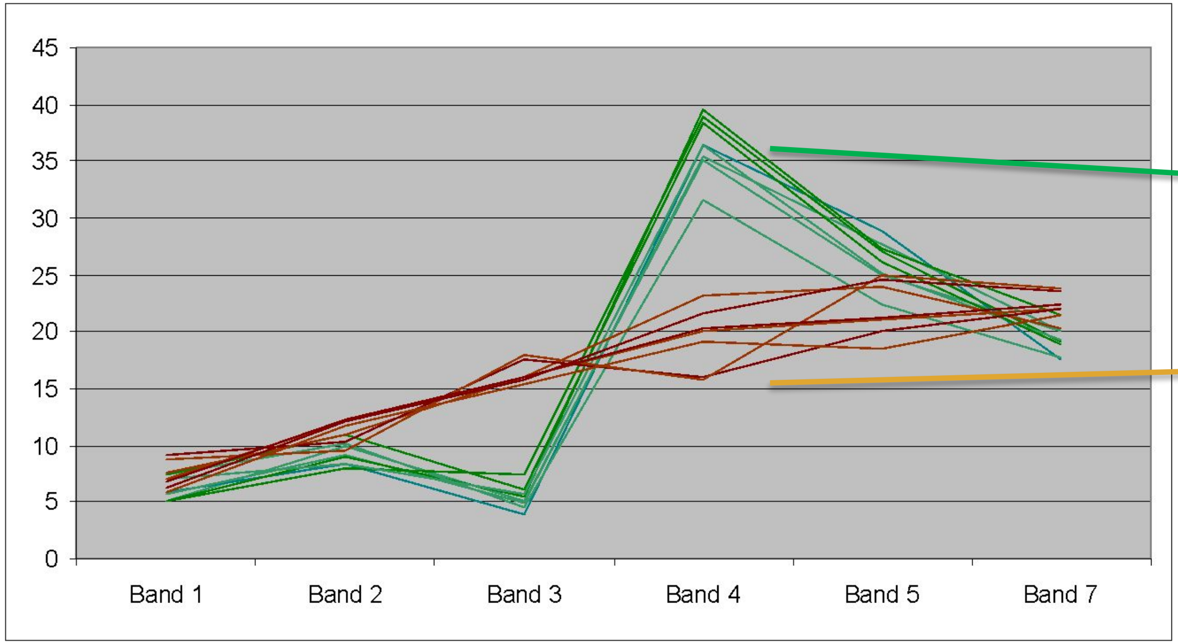
Pixel-Based Classification

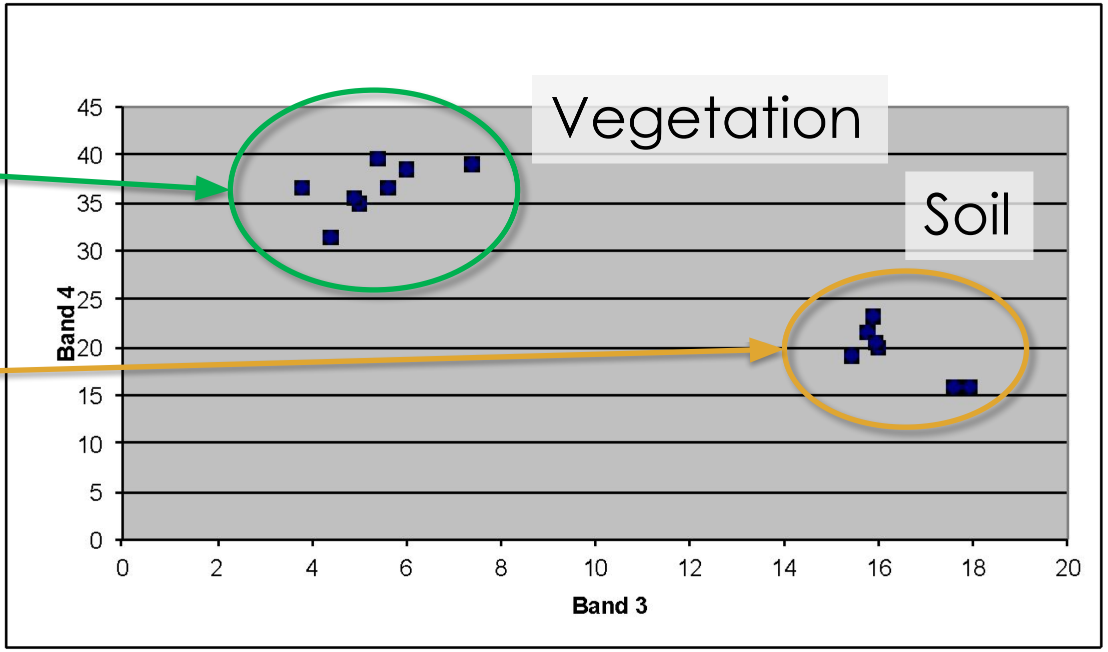
- Pixel-based classification uses the spectral information represented by the digital numbers in sensor data spectral bands and attempts to classify each individual pixel based on this spectral information.
- Spectral Signature:
  - Objects on the ground reflect electromagnetic radiation differently in different wavelengths.
- Example: Green vegetation absorbs red wavelengths but reflects near-infrared (NIR) wavelengths.



Spectral Variation

- It's easier to distinguish **between** broad classes.
  - E.g., forest, agriculture, bare soil
- It's more difficult to distinguish **within** broad classes.
  - E.g., vegetation species

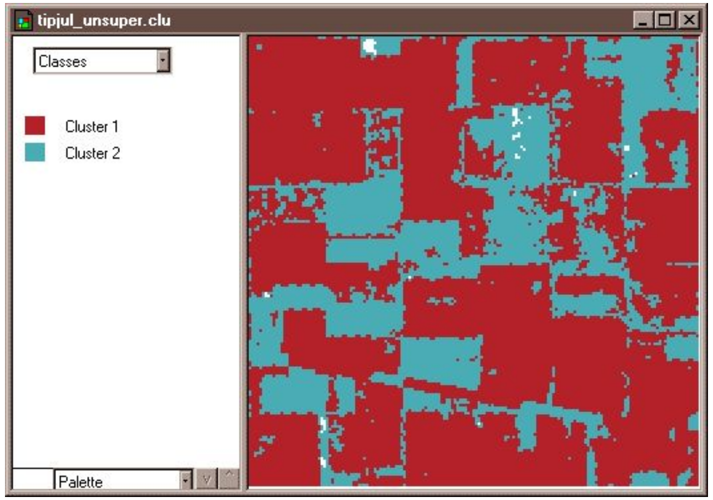




Types of Land Cover Classifications

- **Unsupervised Classification**
  - Uses classification algorithms to assign pixels into one of a number of user-specified class groupings.
  - Interpreters assign each of the groupings of pixels a value corresponding to a land cover class.

- **Supervised Classification**
  - Uses expert-defined areas of known vegetation types (training areas) to tune parameters of classification algorithms.
  - The algorithm then automatically identifies and labels areas similar to the training data.



Simple example of supervised and unsupervised classifications. Supervised classifications assign a class through the use of training data. Unsupervised classifications cluster similar pixels for later classification. Image Credits: PSU Department of Geography

Supervised vs. Unsupervised Classification Workflow

Unsupervised

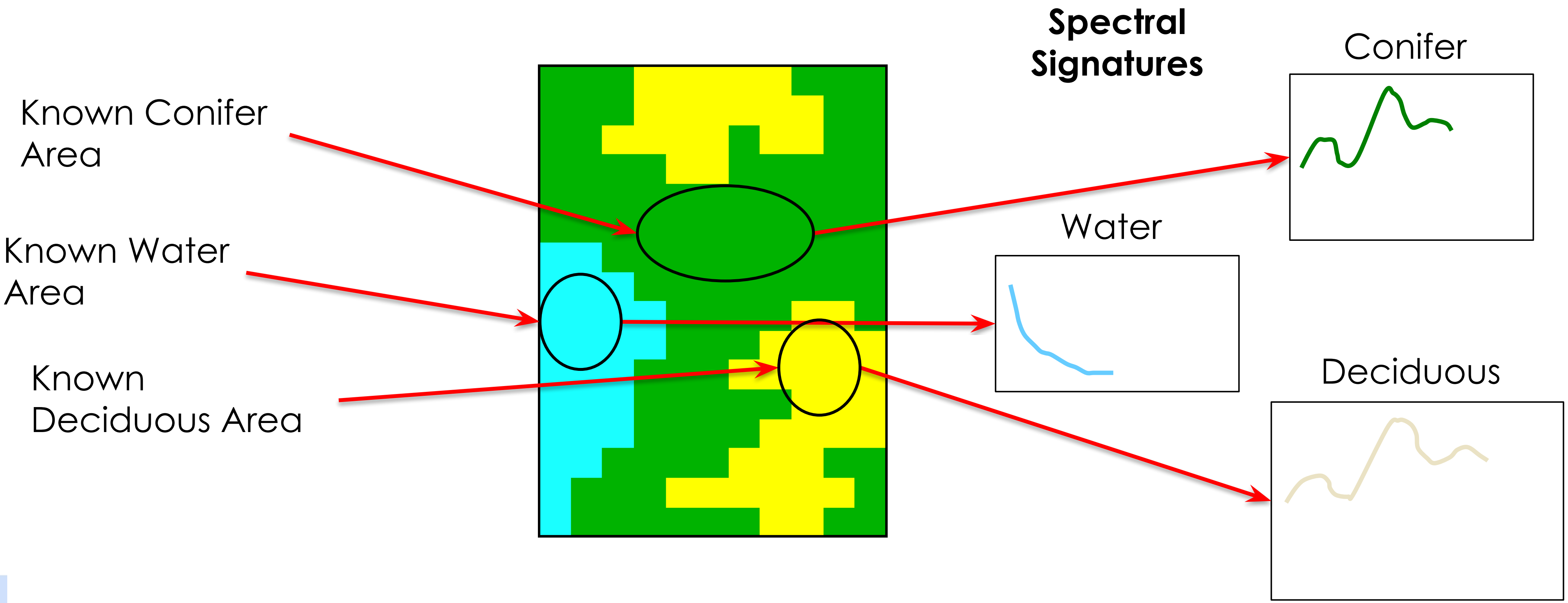
1. Run Clustering Algorithm
2. Identify Classes
3. Edit/Evaluate Signatures
4. Evaluate Classification

Supervised

1. Select Training Fields
2. Edit/Evaluate Signatures
3. Classify Image
4. Evaluate Classification

Training a Supervised Land Classification

- Supervised classification requires the analyst to select training areas or points where the land cover class on the ground is known.
- These reference points are used by an algorithm to classify remaining pixels.



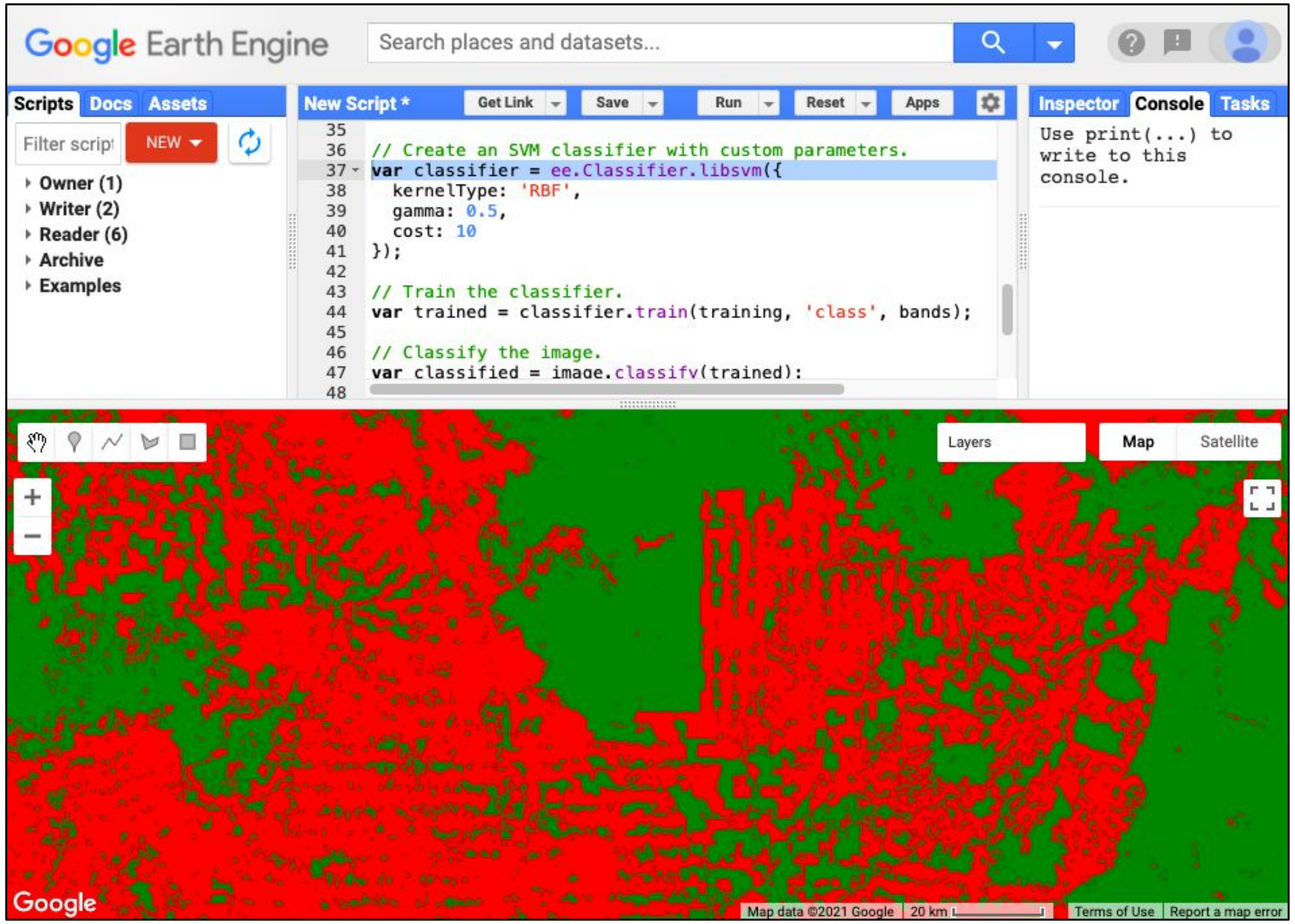
Machine Learning Algorithms for Land Cover Classification

- Machine learning algorithms use inputs (imagery and training data) to group and classify pixels. The algorithm is a “black box”, meaning we are only able to examine inputs and outputs.

Input and Training Data → Machine Learning Algorithm “Black Box” → Output Classifications

Supervised Classifications Algorithms Available in GEE

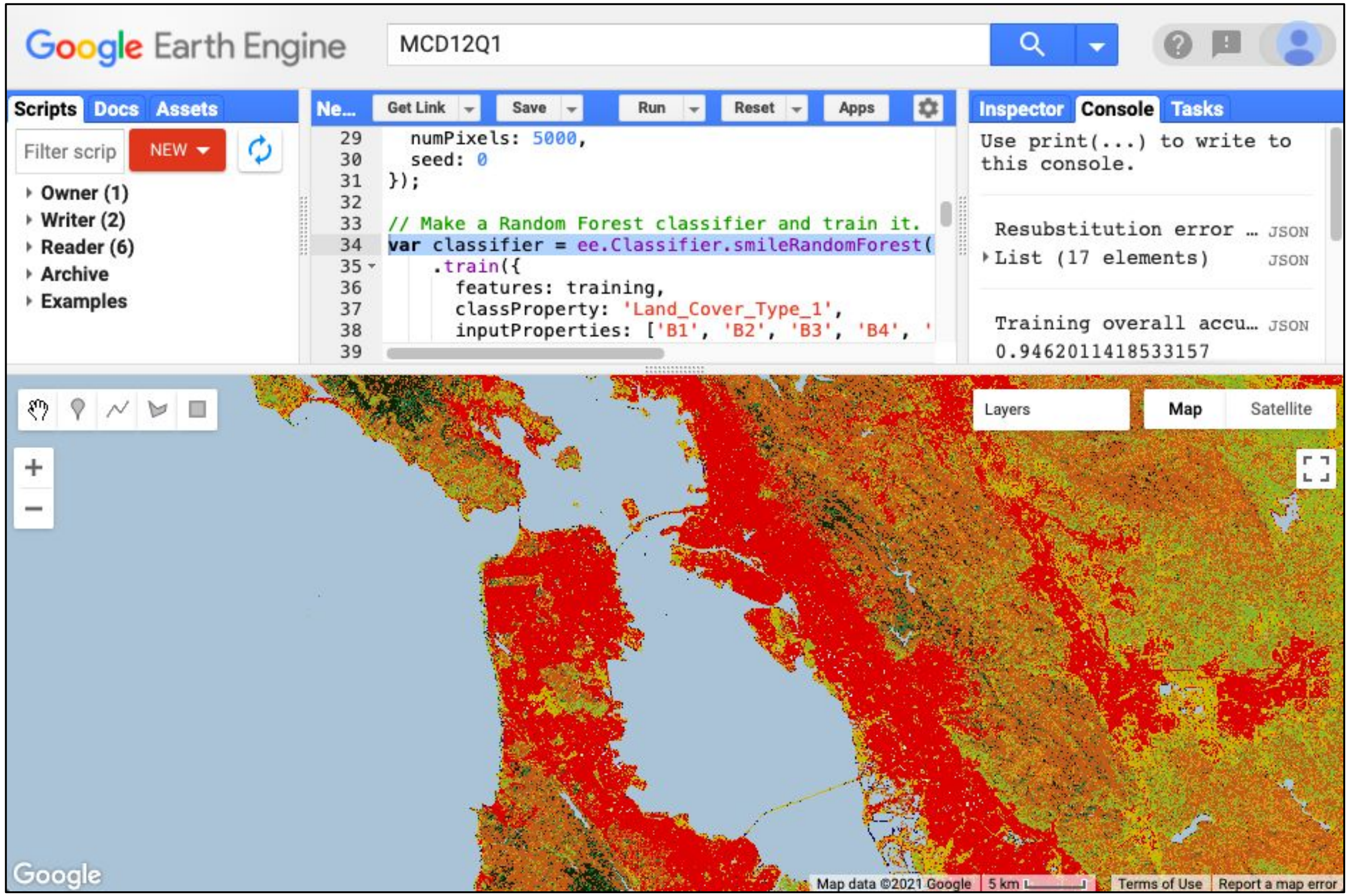
- Classification and Regression Trees (CART)
  - ee.Classifier.smileCart
- Naive Bayes
  - ee.Classifier.smileNaiveBayes
- Support Vector Machine (SVM)
  - ee.Classifier.libsvm
- RandomForest
  - ee.Classifier.smileRandomForest
- In this session we will be using the RandomForest algorithm.



Simple Support Vector Machine (SVM) classifier example in the GEE code editor mapping deforestation in red. Credit: GEE Developers

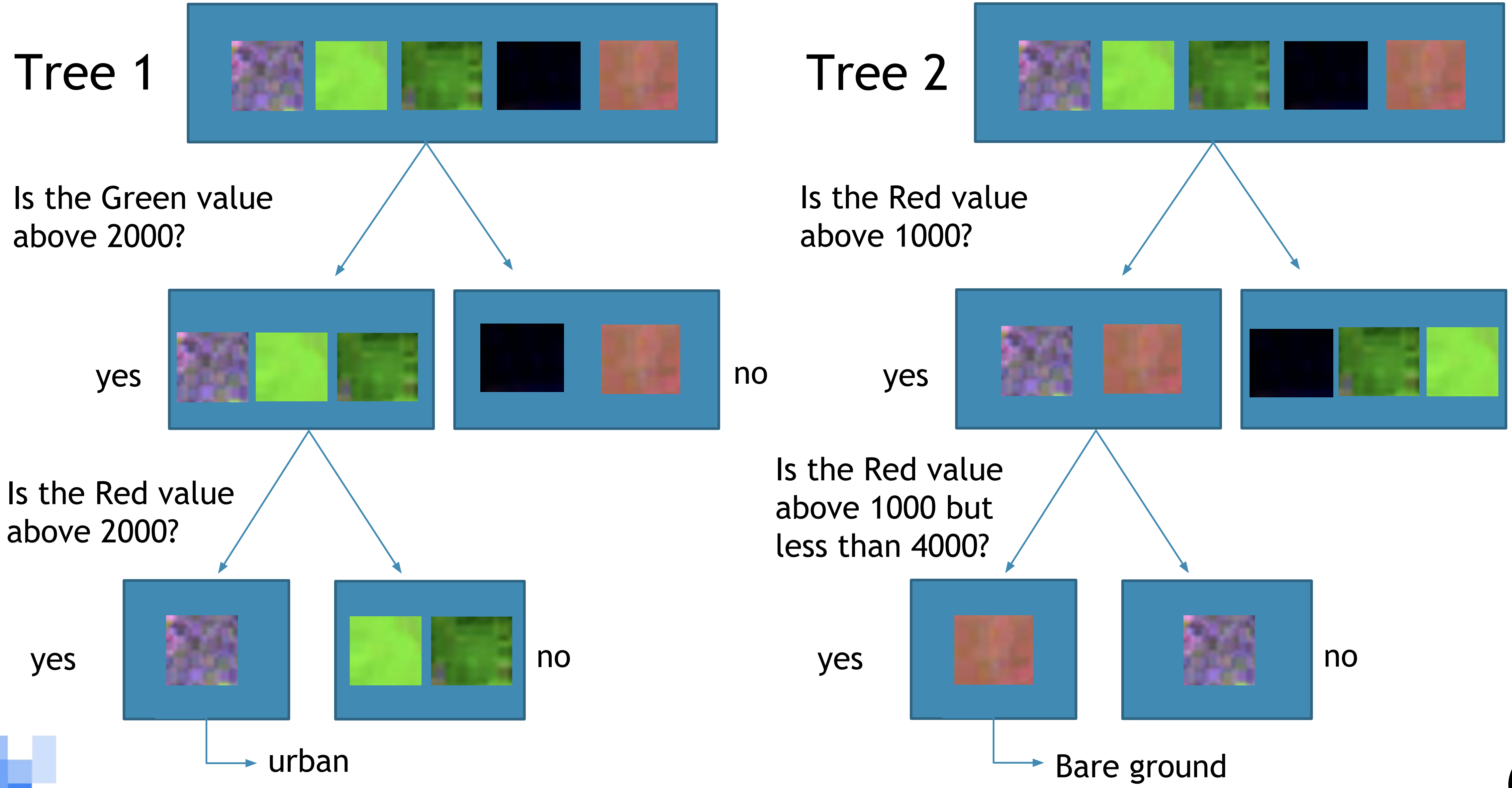
Random Forest Classification

- Learns from training data and identifies statistical patterns in large datasets
- Tree-based machine learning algorithm
  - Uses a series of decision trees to select the best classification for all pixels within imagery
  - Iterative use of decision trees allows algorithm to “vote” for the best solution

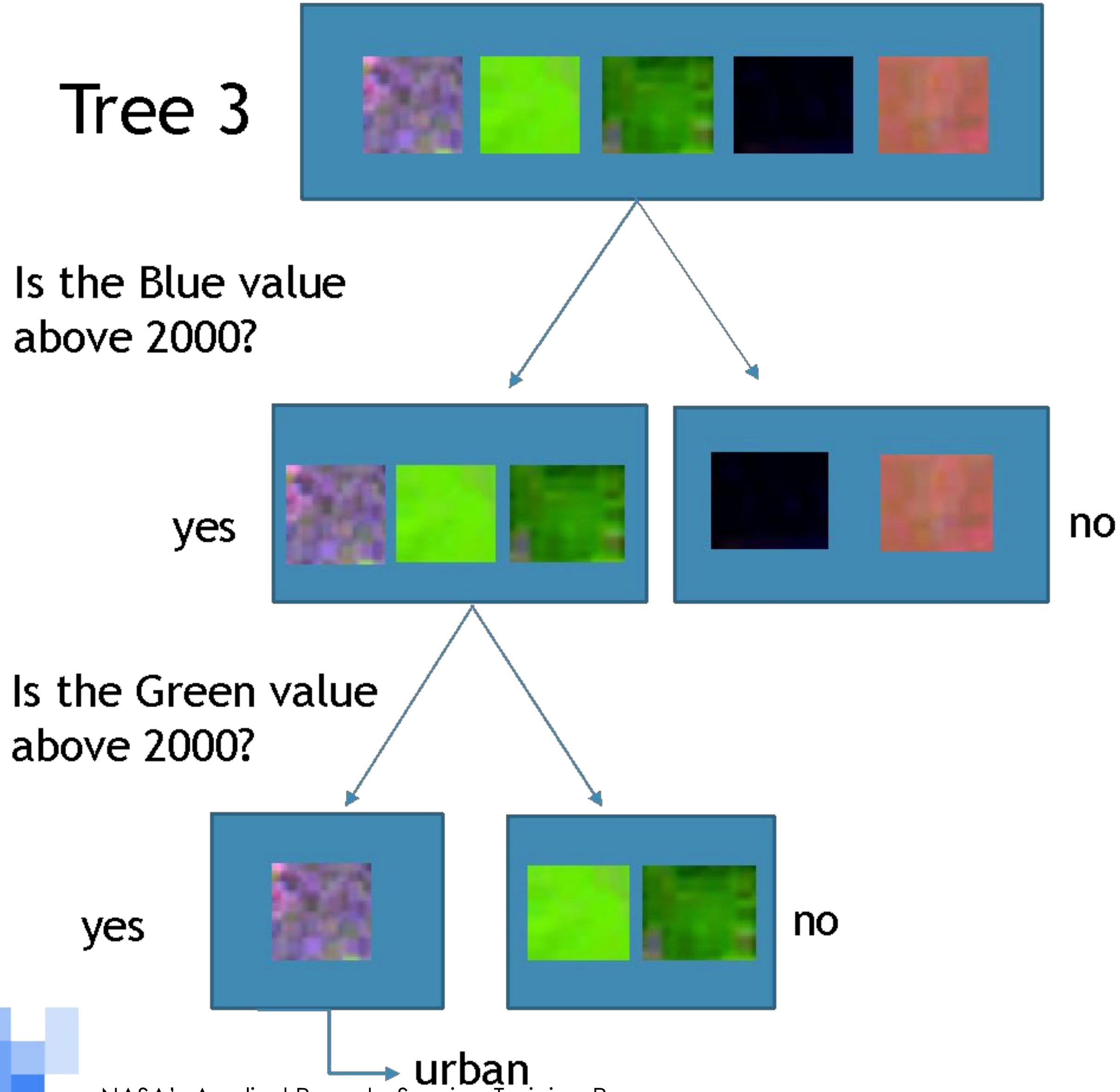


Simple random forest classification completed for the San Francisco Bay Area of California in the GEE interface. Credit: GEE Developers

Random Forest Classification



Random Forest Classification



Tree 1 = Urban

Tree 2 = Bare Ground

Tree 3 = Urban

Classification

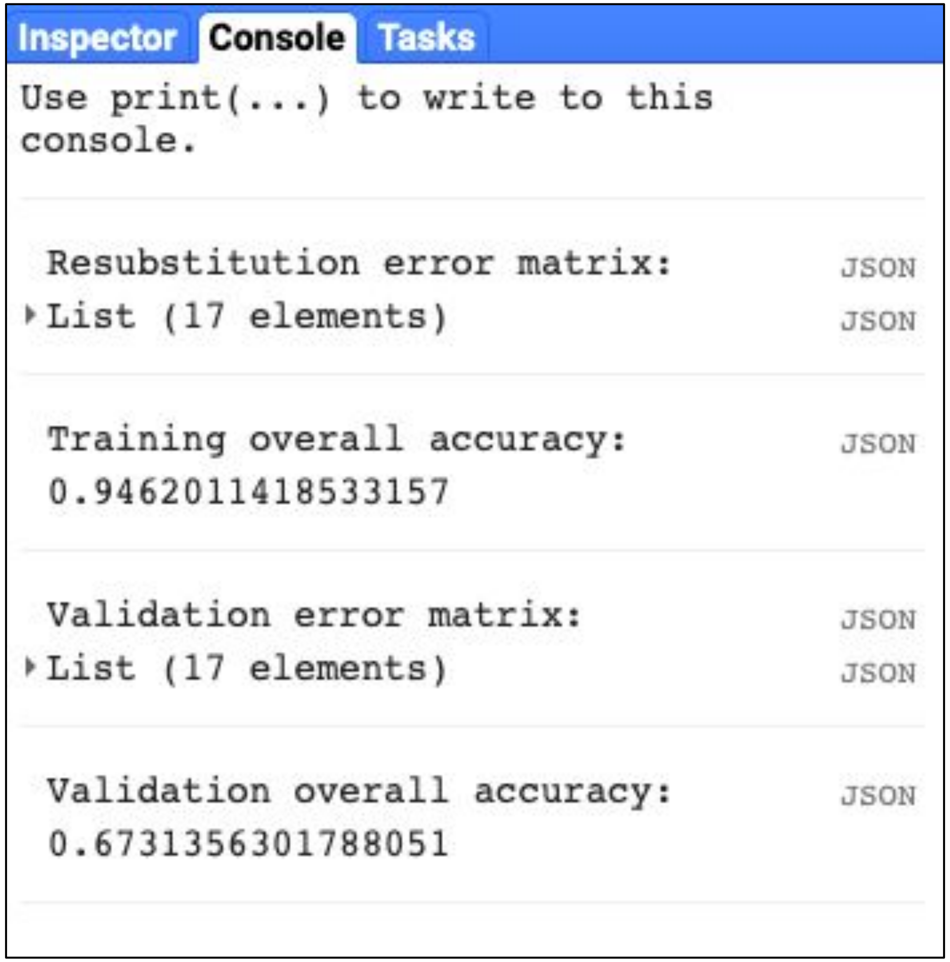
Urban

Advantages and Limitations of the Random Forest Algorithm

- **Advantages:**
  - Use of multiple trees reduces the risk of overfitting
  - Training time is shorter and not sensitive to outliers in training data
  - Runs efficiently and produces high accuracy for large datasets
  - Easy to parameterize
- **Limitations:**
  - Algorithm cannot predict spectral range beyond training data
  - Training data must capture the entire spectral range

Accuracy Assessment

- Accuracy refers to the degree of correspondence between classification and reality.
- Accuracy assessment is the process by which the accuracy or correctness of an image classification is evaluated.
- This involves the comparison of the image classification to reference data that are assumed to be true.
  - References can include ground data or a subset of training points withheld for accuracy assessment purposes.



Example of error matrix display in the GEE console for accuracy assessment. Credit: GEE Developers

Determining Classification Accuracy

- Agreement between reference data and algorithm classifications indicates the classifier performed well.
- If there is not agreement, the classifier has incorrectly measured the land class resulting in error.
- Comparison of reference data and classifications is typically done using a confusion (or error) matrix to compile these comparisons.

Plot ID	Class in reference source	Class in classification map	Agreement
1	Urban	Urban	Yes
2	Bare Ground	Urban	No
3	Forest	Forest	Yes
4	Forest	Agriculture	No
5	….	….	….

Error Matrix

- Table of reference classes to predicted classes:
  - Reference classes are assumed to be correct (columns).
  - Mapped classes are the output of the classification (rows).

	Reference Classes				
Classification	**Urban**	**Agriculture**	**Forest**	**Bare Ground**	**Row Total**
Urban	45	4	12	24	85
Agriculture	6	91	5	8	110
Forest	0	8	55	9	72
Bare Ground	4	7	3	55	69
Column Total	55	110	75	96	**336**

Mapped Classes

The number of correctly classified pixels is shown along the diagonal.

Error Matrix

- Table of reference classes to predicted classes:
  - Reference classes are assumed to be correct (columns).
  - Mapped classes are the output of the classification (rows).

Mapped Classes		Reference Classes				
	Classification	**Urban**	**Agriculture**	**Forest**	**Bare Ground**	**Row Total**
	Urban	45	4	12	24	85
	Agriculture	6	91	5	8	110
	Forest	0	8	55	9	72
	Bare Ground	4	7	3	55	69
	Column Total	55	110	75	96	**336**

Column totals are the total number of reference pixels in each class.

Error Matrix

- Table of reference classes to predicted classes:
  - Reference classes are assumed to be correct (columns).
  - Mapped classes are the output of the classification (rows).

	Reference Classes				
Classification	**Urban**	**Agriculture**	**Forest**	**Bare Ground**	**Row Total**
Urban	45	4	12	24	85
Agriculture	6	91	5	8	110
Forest	0	8	55	9	72
Bare Ground	4	7	3	55	69
Column Total	55	110	75	96	**336**

Mapped Classes

Row totals are the total number of classified pixels in each class.

Error Matrix

- Table of reference classes to predicted classes:
  - Reference classes are assumed to be correct (columns).
  - Mapped classes are the output of the classification (rows).

	Reference Classes				
Classification	**Urban**	**Agriculture**	**Forest**	**Bare Ground**	**Row Total**
Urban	45	4	12	24	85
Agriculture	6	91	5	8	110
Forest	0	8	55	9	72
Bare Ground	4	7	3	55	69
Column Total	55	110	75	96	**336**

Mapped Classes

Off-diagonal numbers represent errors of commission and omission.

Map Accuracy

$$\text{Overall Accuracy} = \frac{\text{Number of correctly classified pixels (sum of diagonal)}}{\text{Number of total sampled pixels}}$$

$$\text{Overall Accuracy} = \frac{45+91+55+55}{336} * 100 = \mathbf{73\%}$$

	Reference Classes				
Classification	Urban	Agriculture	Forest	Bare Ground	Row Total
Urban	45	4	12	24	85
Agriculture	6	91	5	8	110
Forest	0	8	55	9	72
Bare Ground	4	7	3	55	69
Column Total	55	110	75	96	**336**

Mapped Classes

Map accuracy is defined by the percent of correctly classified pixels.

Kappa Statistic

- Takes into account the possibility of the agreement occurring by chance
- Proportion of agreement after chance agreement has been removed
- Calculated from an error matrix as an additional accuracy check
- A higher kappa value means higher accuracy
- Does not usually provide more information about accuracy than the error matrix overall accuracy calculation

Kappa	Interpretation
< 0	No agreement
0.0 - 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

Image Credit: Okwuashi et al. 2012

Kappa Statistic

$$k = \frac{p_0 - p_c}{1 - p_c}$$

$\mathbf{p_0}$= Observed accuracy. $\Sigma\, p_{ii}$ is the sum of relative frequency in the diagonal of the error matrix

$\mathbf{p_c}$= Chance agreement. $\Sigma\, p_{i+} * p_{+i}$ is the relative frequency of a random allocation of observations to the cells of the error matrix

Calculation of kappa

	Reference Classes				
Classification	Urban	Agriculture	Forest	Bare ground	Row total
Urban	45	4	12	24	85
Agriculture	6	91	5	8	110
Forest	0	8	55	9	72
Bare ground	4	7	3	55	69
Column total	55	110	75	96	**336**

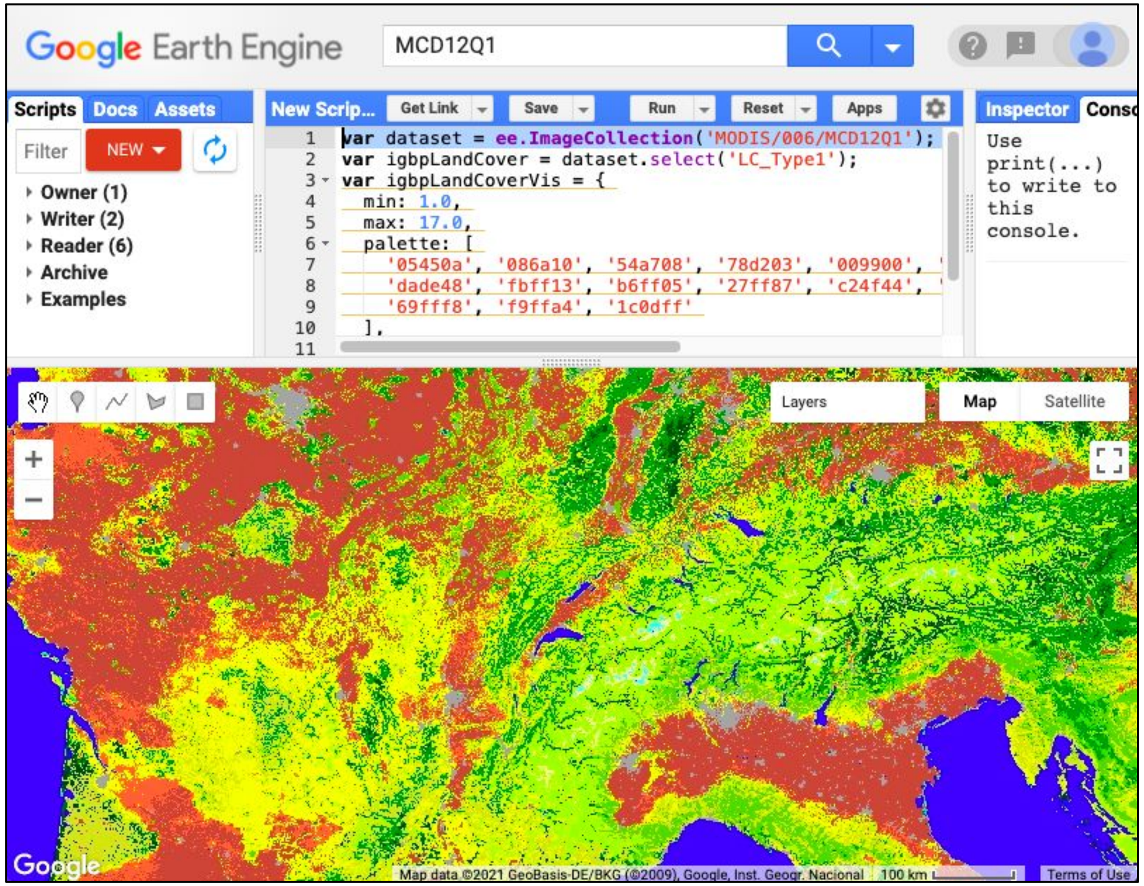
$p_0 = (45+91+55+55)/336 = 0.7321$

$$p_c = \frac{(85*55)/336 + (110*110)/336 + (72*75)/336 + (69*96)/336 = 0.2551}{336}$$

$$k = \frac{0.7321 - 0.2551}{1 - 0.2551} = \mathbf{0.64}$$

Accuracy Assessment in GEE: Confusion/Error Matrix

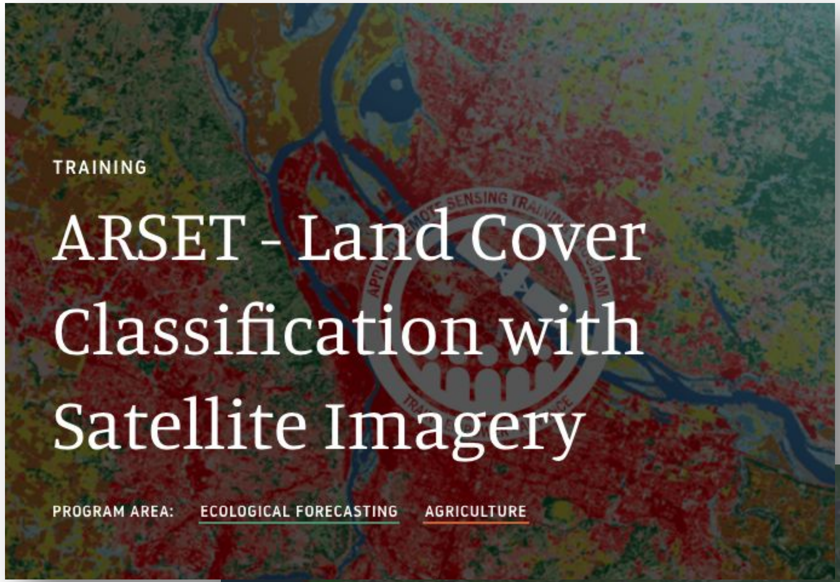
- **Confusion/Error Matrix:**
  - Describes how well the classifier was able to correctly label training data the classifier has already seen.
  - Can also compare predicted values to actual values the classifier has not seen.
  - A subset of training data can be withheld as validation data and used to test the ability of the trained classifier to accurately predict land cover class.
  - Other land cover products like MODIS land cover product (MCD12Q1) can be used as validation data.

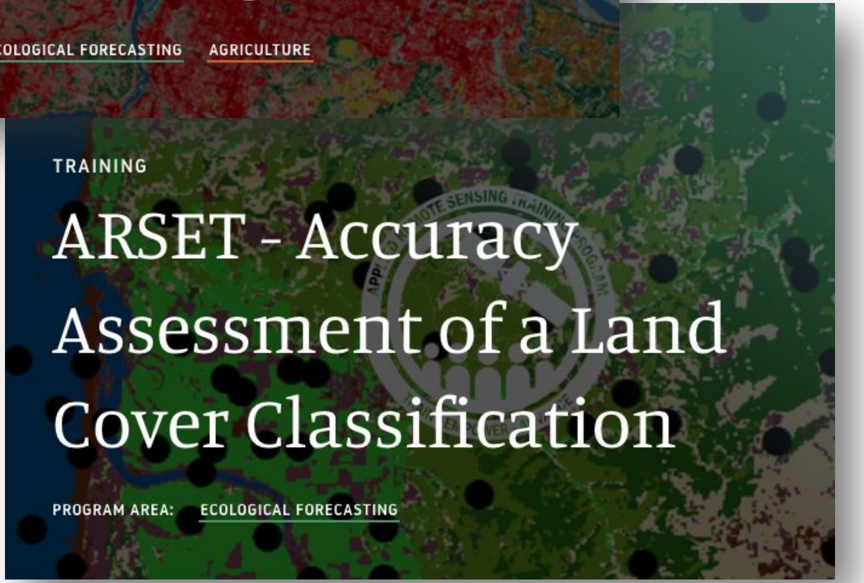


MODIS annual, global land cover data product that is mentioned in the GEE developer repository as a source of points for accuracy assessment of your own land classification. Credit: GEE Developers

Previous ARSET Trainings

- Past ARSET trainings relevant to land classification and accuracy assessment:
  - [Land Cover Classification with Satellite Imagery]
  - [Accuracy Assessment of a Land Cover Classification]
  - [Remote Sensing for Mangroves in Support of the UN Sustainable Development Goals]

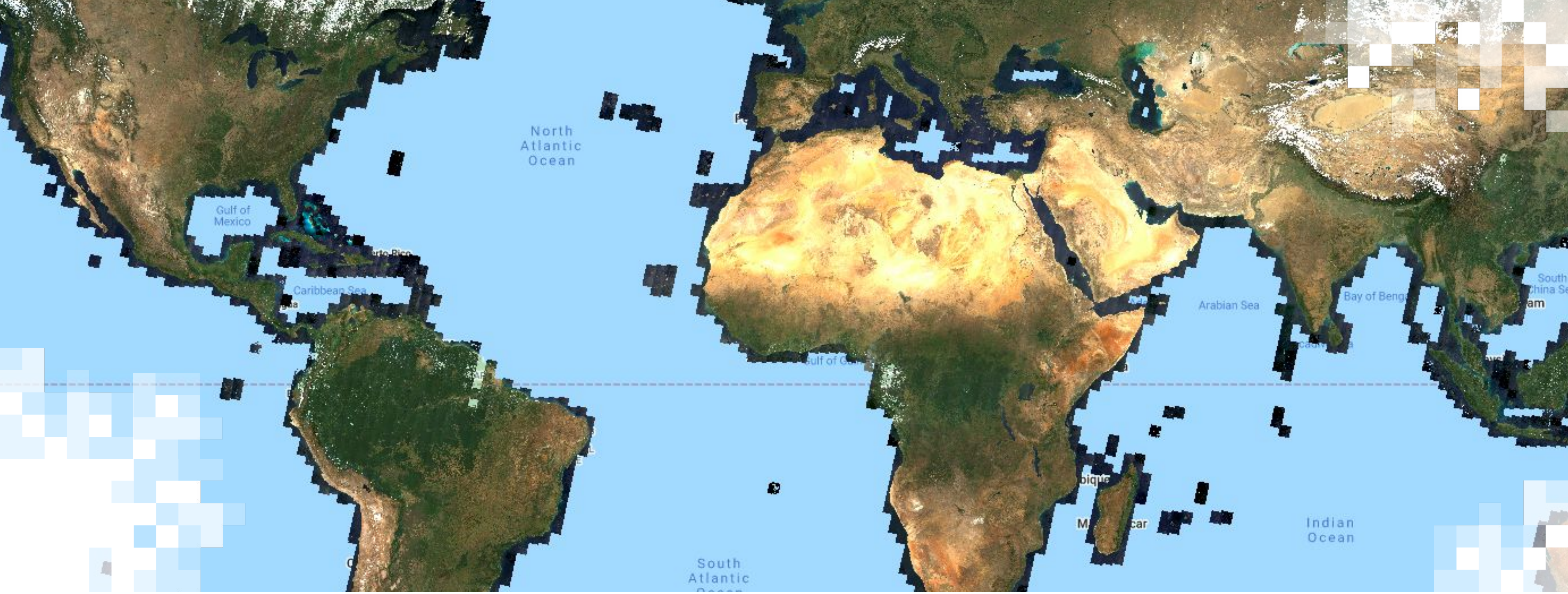




GEE Developer Guides

- Relevant guides to GEE features and JavaScript Code:
  - FeatureCollection Overview
  - Compositing, Masking, and Mosaicking
  - Machine Learning in Earth Engine
  - Supervised Classification (including accuracy assessment)
- The full list of guides and tutorials made available by the developers:
  - JavaScript and Python Guides

Image Credit: Google Earth Engine



Supervised Land Classification and Accuracy Assessment in Google Earth Engine

https://code.earthengine.google.com/a2bde57085bc1d47c49a83f44ea15918

Summary

- Land classification in GEE is a valuable tool for mapping and monitoring land cover.
- To complete a supervised land classification, the user must first establish training points of known land cover classes to train the classifier.
- GEE provides many machine learning classification algorithms built into the API:
  - Such as Classification and Regression Trees, Naïve Bayes, Support Vector Machine, and RandomForest.
- Random Forest is a tree-based machine learning algorithm that uses a series of decision trees to select the best classification for all pixels within imagery.
- Simple accuracy assessment in GEE can be completed through the use of confusion/error matrices to compare predicted classifications to withheld training data and validation data.
- We demonstrated these functionalities in our activity classifying land cover in Cumberland County, Maine.
- Session 3: Time Series Analysis and Change Detection

Homework and Certificate

- One homework assignment:
  - Answers must be submitted via Google Form, accessed from the ARSET website.
  - Due date for homework: **July 14, 2021**
- A certificate of completion will be awarded to those who:
  - Attend all live webinars
  - Complete the homework assignment by the deadline
  - You will receive a certificate approximately two months after the completion of the course from: marines.martins@ssaihq.com

Contacts

- Trainers:
  - Zach Bengtsson: bengtsson@baeri.org
  - Britnay Beaudry: britnay.beaudry@ssaihq.com
  - Juan Torres-Pérez: juan.l.torresperez@nasa.gov
  - Amber McCullum: amberjean.mccullum@nasa.gov

Follow us on Twitter
@NASAARSET

- Training Webpage:
  - https://appliedsciences.nasa.gov/join-mission/training/english/arset-using-google-earth-engine-land-monitoring-applications
- ARSET Website:
  - https://appliedsciences.nasa.gov/what-we-do/capacity-building/arset

Thank You!