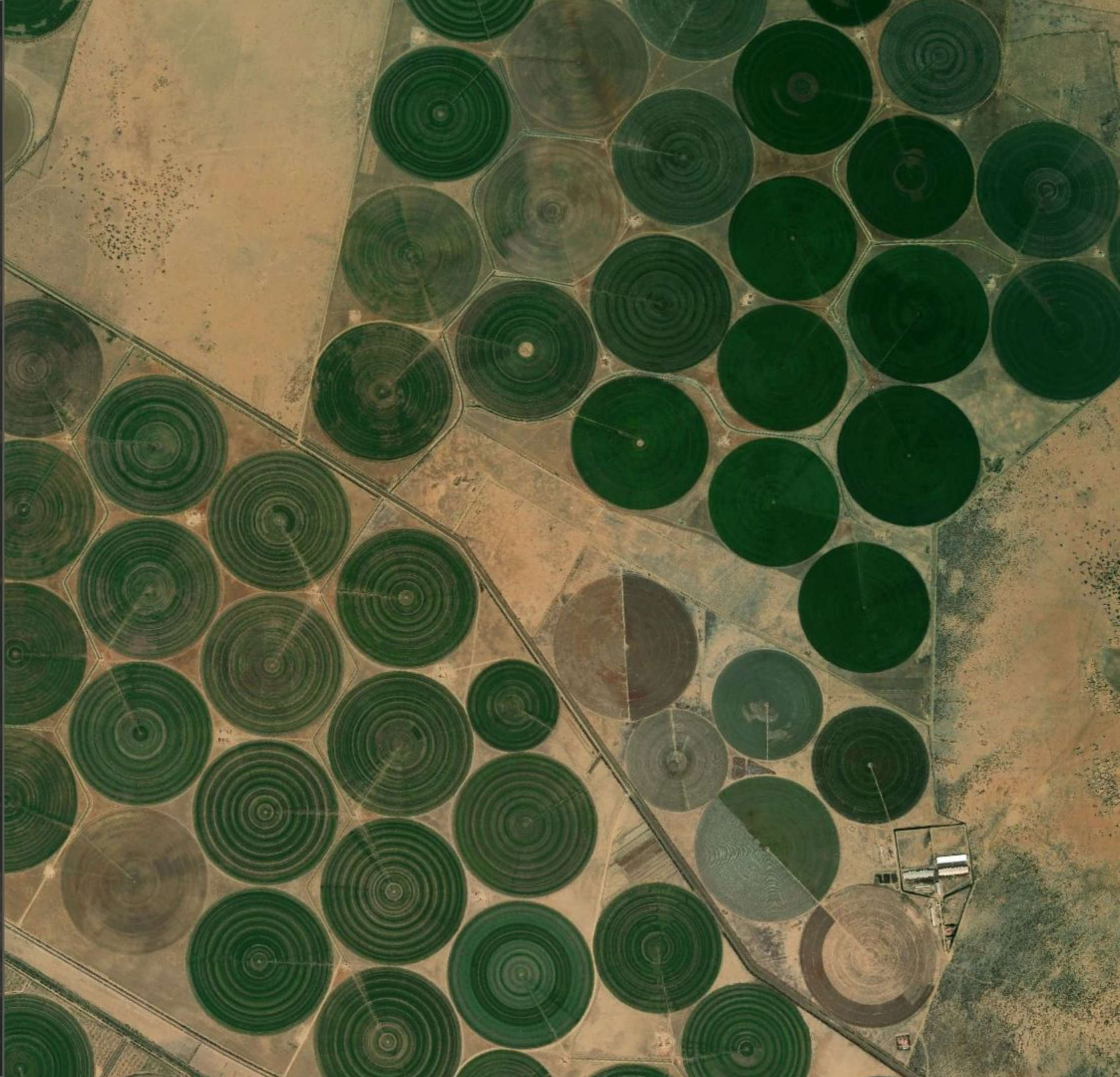
Updating Irrigation Maps

Using Landsat Images

Laura Pintos
Carlos Sancini
Sam Tosaria, CFA

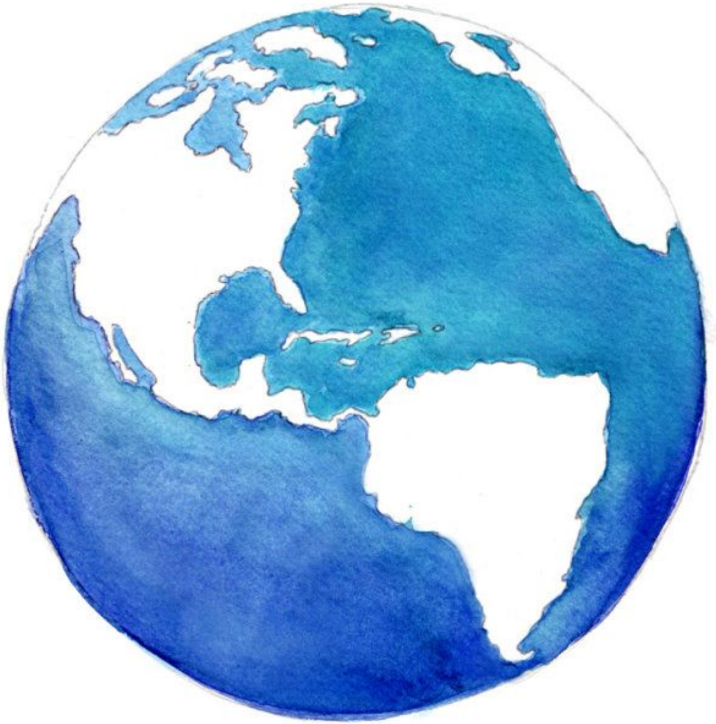
Special thanks to
Prof. Paolo D'Odorico
Fred Nugen, PhD
Lorenzo Rosa
Alberto Todeschini, PhD



Agenda

- Project Objective
- Background + Dataset
- Feature Engineering
- Pipeline & Model
- Inference

Water - The Essence of Life



96% Oceans

2.5% Fresh Water

1% Easily Accessible

Image : www. npr.org, Benjamin Arthur

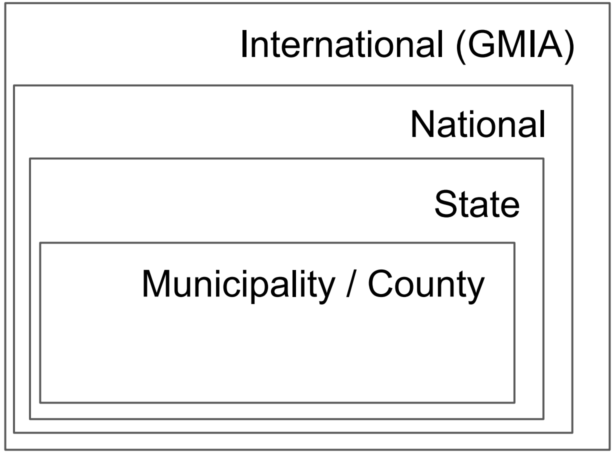
Water - The Essence of Food

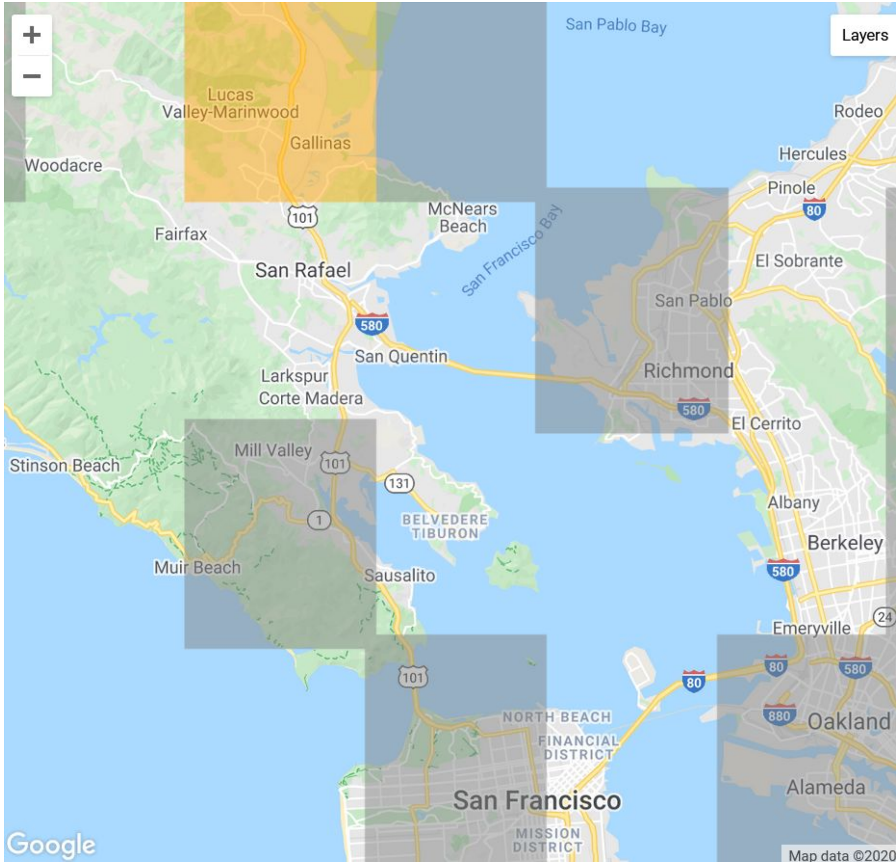
70%

Agriculture irrigation

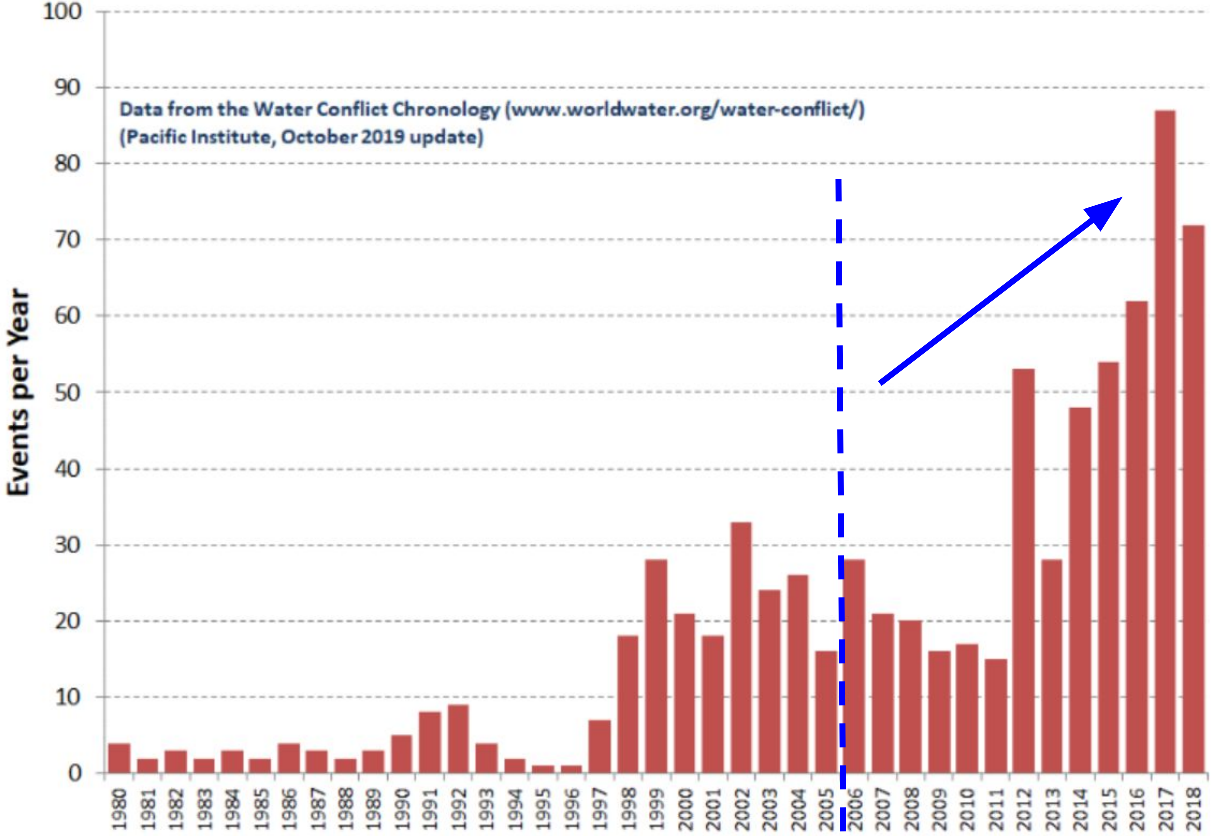
Global Map of Irrigation Areas (GMIA - 2005)

Irrigation Data Aggregation





Political Conflict Involving Water (1980 - 2018)

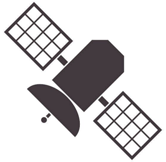


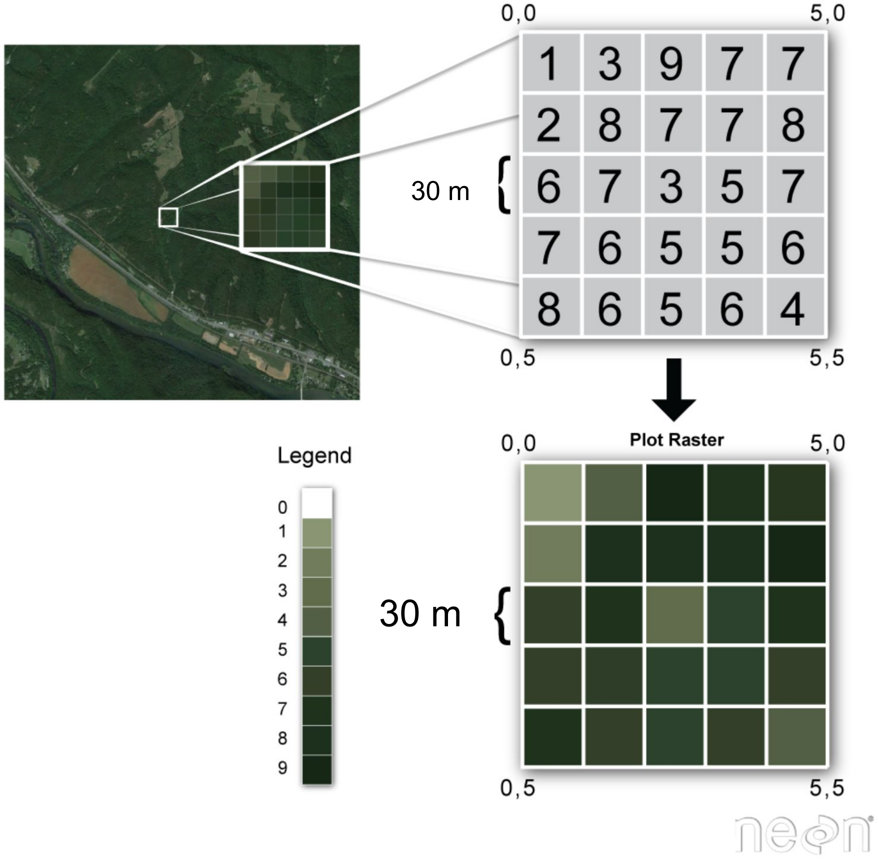
Source: Pacific Institute

Background

Fresno, California, USA

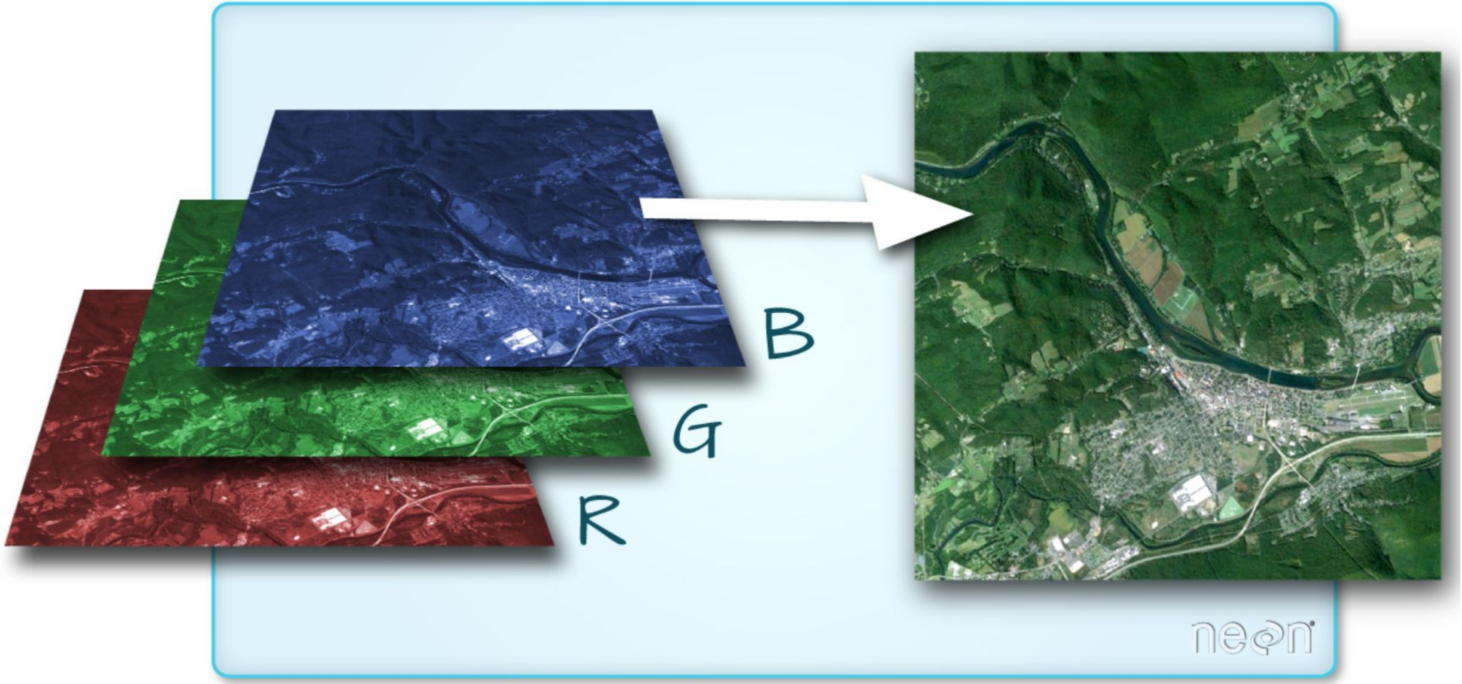
What is Raster Data?





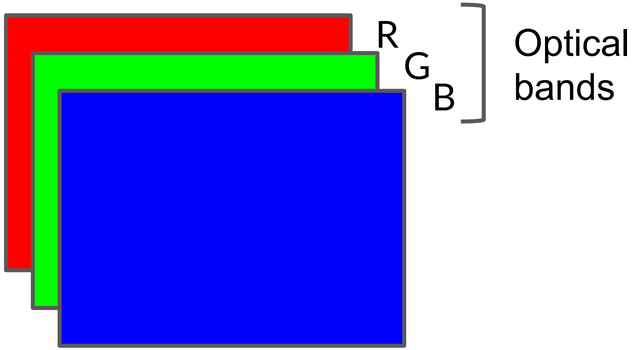
Source: National Ecological Observatory Network (NEON)

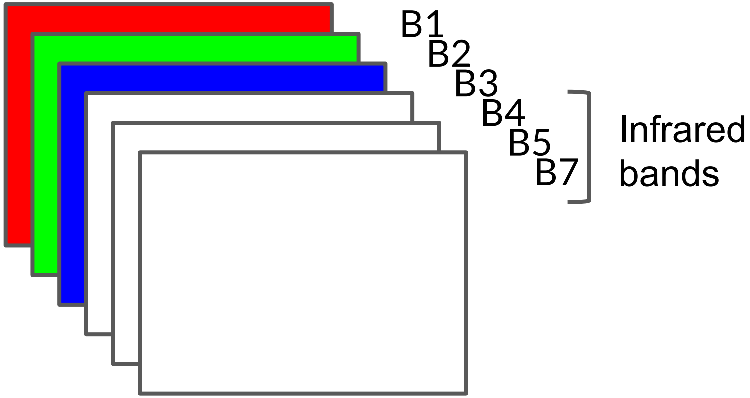
How Do We Create An Image From a Raster?



(Source: National Ecological Observatory Network (NEON).)

What is unique about a satellite image?





What is the resolution of the Landsat images?

Each Landsat 8 pixel is 30m x 30m or $900m^2$

How does it compare with what I see on Google Maps?

Satellite Data Spatial Resolution

What does ‘Top of Atmosphere’ and ‘Surface Reflectance’ mean?

winter		summer	
TOA	SR	TOA	SR

But what about clouds and image composites?

winter		summer	
Composite based on Cloud score	No aggregation (first image on the stack)	Composite based on Cloud score	No aggregation (first image on the stack)

How do the images differ with seasons?

winter	spring	summer	fall

Datasets: Remote Sensing Data (Landsat 2005)

PCT_AEI: 12%

ID: 181771
(California)

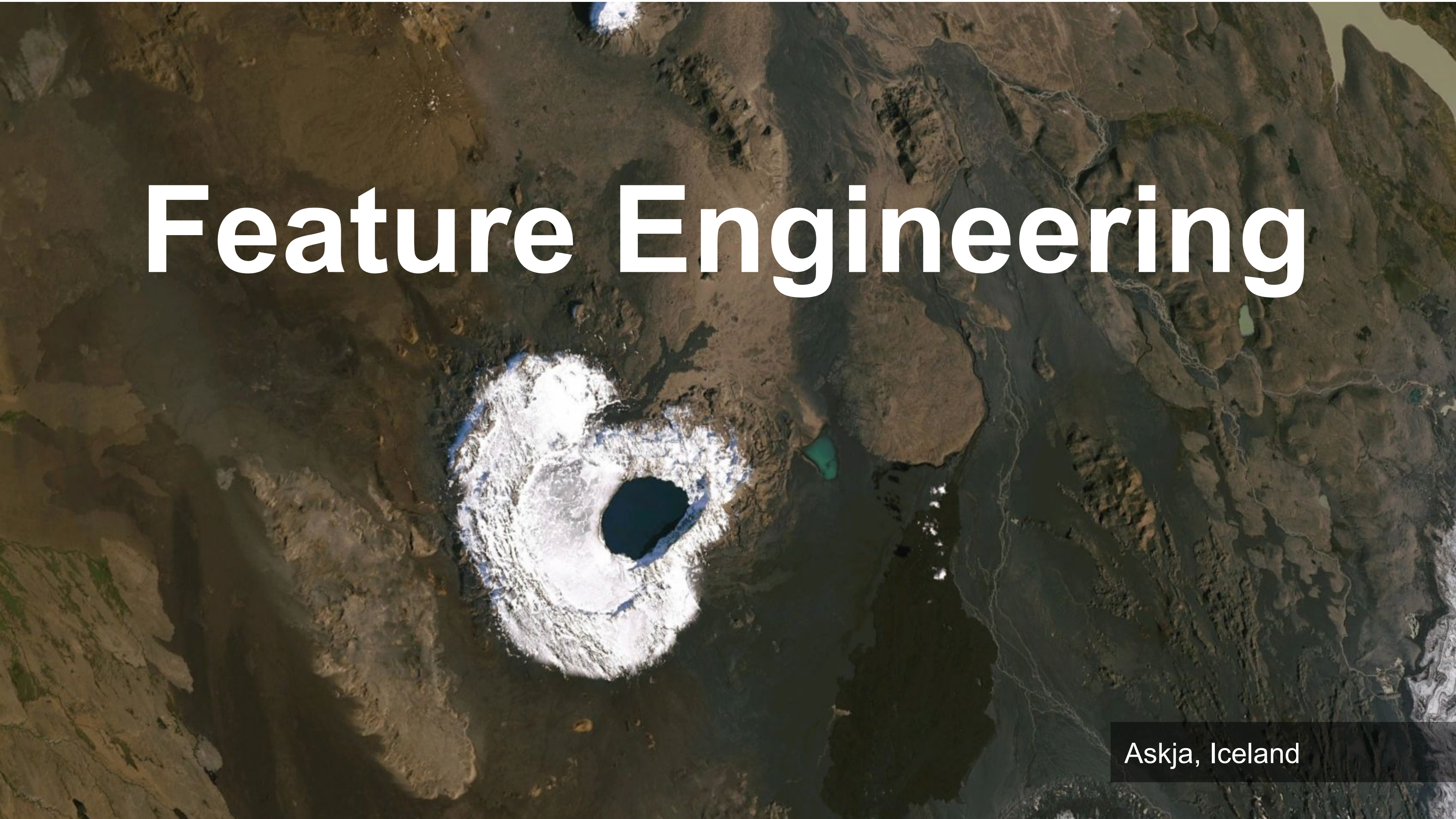
PCT_AEI: 35%

ID: 182783
(California)

PCT_AEI: 79%

ID: 183785
(California)

Feature Engineering



Askja, Iceland

Array Payload

Feature	Bandwidth	Spectrum
Blue	Band B1 (0.45-0.52 μm)	Visual
Green	Band B2 (0.52-0.60 μm)	Visual
Red	Band B3 (0.63-0.69 μm)	Visual
Near Infrared (NIR)	Band B4 (0.77-0.90 μm)	Infrared
Medium Infrared (MIR)	Band B5 (1.55-1.75 μm)	Infrared
Long Infrared (LIR)	Band B7 (1.55-1.75 μm)	Infrared

Feature Engineering Using Channel Depth

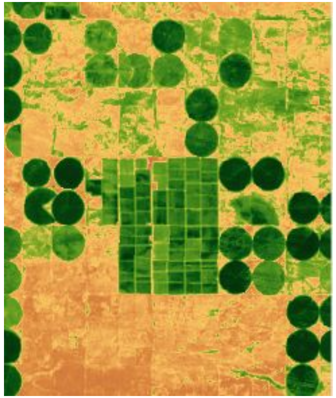
Feature	Calculation	Intuition
NDVI	(NIR - Red) / (NIR + Red)	Vegetation robustness based chlorophyll content
EVI	2.5*((NIR - Red)/(NIR + 6*Red - 7.5*Blue+1))	Similar to NDVI but more sensitive to canopy structural variations
NDMI	(NIR - MIR) / (NIR + MIR)	Vegetation robustness based on water content/ moisture
NDWI	(Blue - NIR) / (Blue + NIR)	Identification of flooded land used in semi aquatic crops
MNDWI	(Blue - MIR) / (Blue + MIR)	Sensitive to vegetation/soil presence in the flooded crop
WRI	(Green + Red)/(NIR + MIR)	A ratio of “non water” to water wavelength content

Source: Water Feature Extraction and Change Detection Using Multitemporal Landsat Imagery - Rokni, Ahmad, Selamat, Hazini, Remote Sensing (2014)

Visual Display of Added Features

RGB

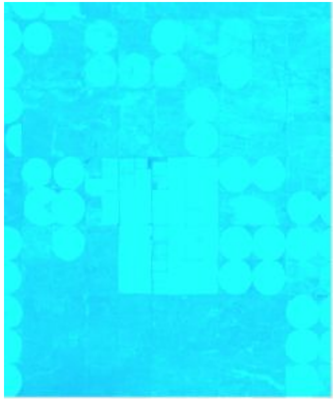
NDVI



EVI

NDMI

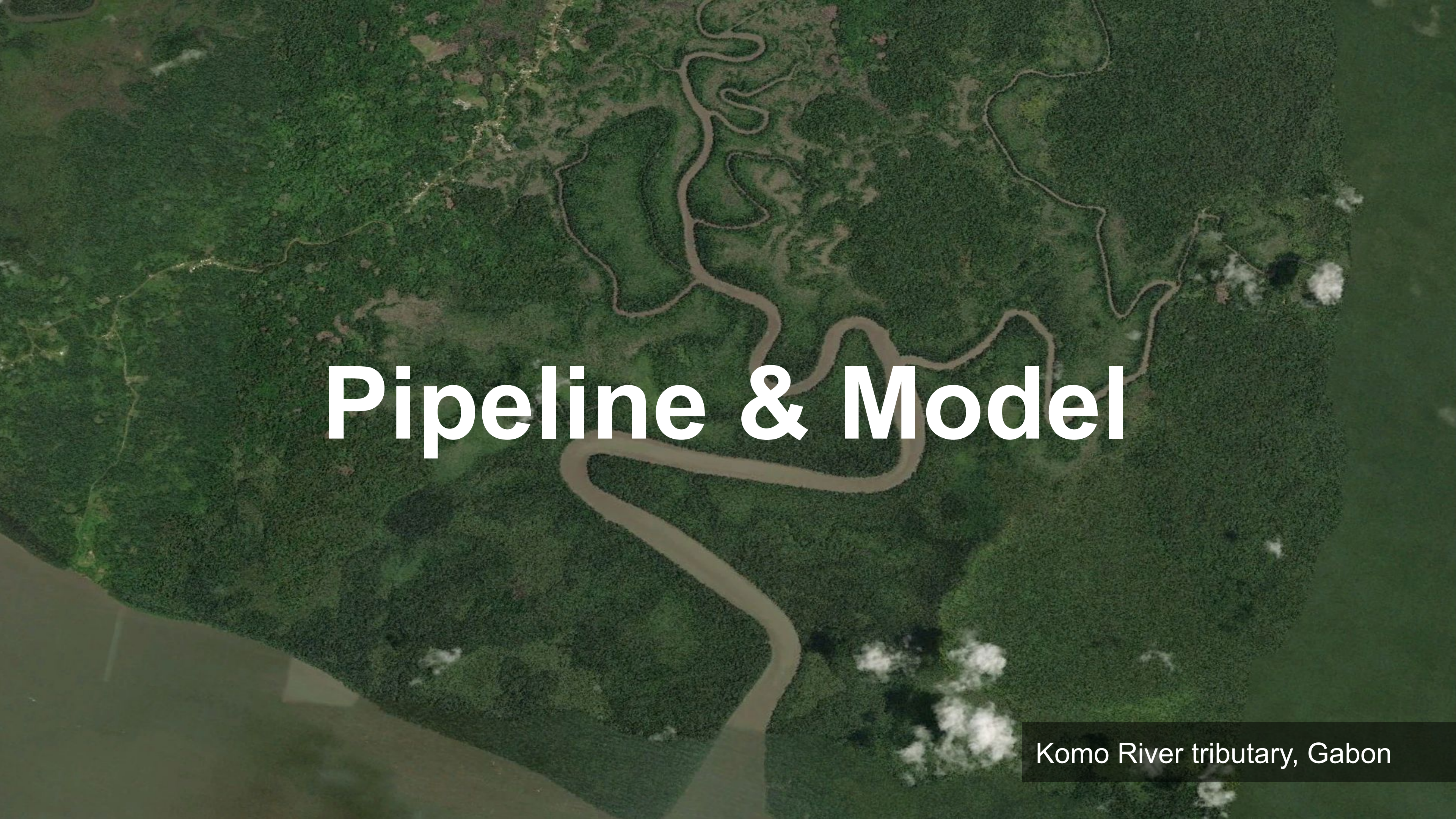
NDWI



MNDWI

Tile: 240405
(Arizona)

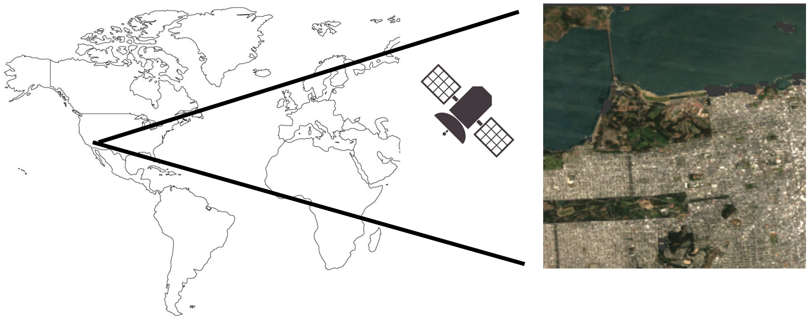
Pipeline & Model



Komo River tributary, Gabon

Pipeline Overview

Google Earth Engine + GMIA geo-coordinates



Landsat 7
GeoTIFF
~ 5.6 TB

Preprocessing: Cloud, scan lines, seasons

6 channels
256 x 256 pixel
~ 4 MB

Feature Engineering: Geographic indices

TFRecord
x
23,688 examples

12 channels
256 x 256 pixel
~ 100 GB

Model

TensorFlow 2.0

Output

Google Earth Engine
ArcGIS

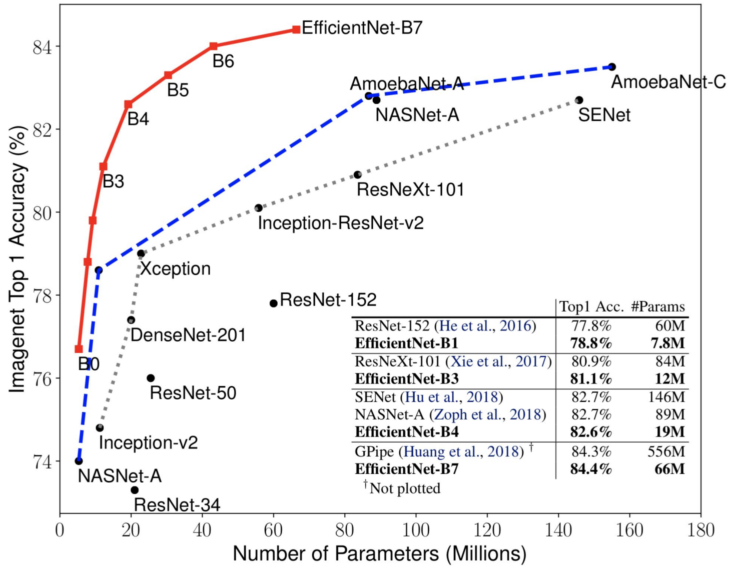
Update + Expand
1,500 examples

Pipeline Overview

Data Extraction and Preparation	*Training*	*Inference*
GMIA georeferenced polygons matched with 2005 Landsat 7 multispectral images in Google Earth Engine for the US area.	Multiple deep learning model architectures	Expand sections of Sudan, Rio Grande basin and Argentina (2019).
Images were re-scaled to tiles of 256x256 pixels, which encompassed an area of $10km^2$.	Tournament of models with hyperparameter tuning	Inference of selected using best performing model.
Usage of cloud correction algorithm: averaging high score cloud of the summer season.	Multiple labels thresholds (binning) and continuous response tests	Map irrigation changes from 2005 to 2019
The tiles with were extracted in tensorflow TFRecord format: 6 raw bands, GMIA IDs, and area equipped for irrigation response variable	Error analysis and visual inspections.	
Impose ‘schema on read’ for NDVI, EVI, NDMI, NDWI, MNDWI and WRI - total 6 + 6 bands.		

The Best Performing Model - Efficient Net

Efficient Net Deep Learning Architecture



Our best performing model

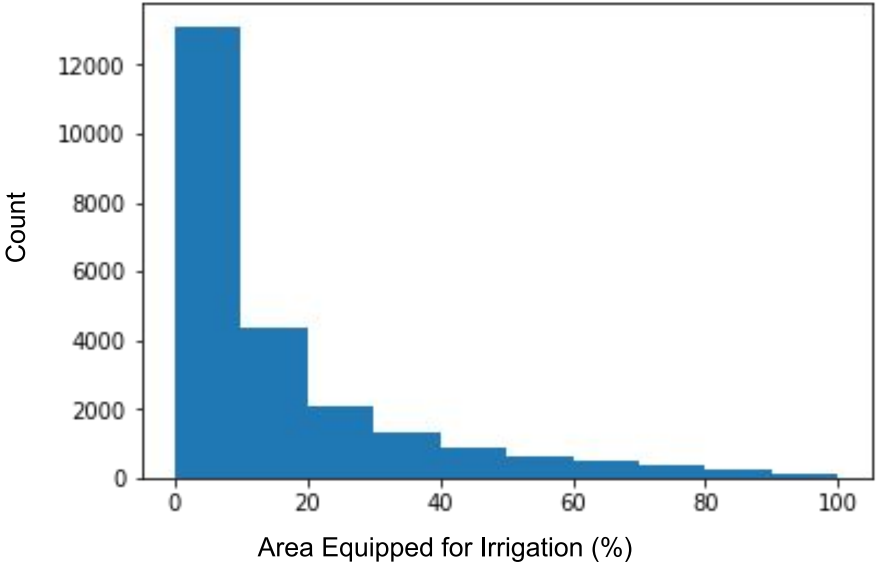
1. Use of B5 Architecture
2. Adaptation of the number of bands of the input layer, from the traditional 3-band RGB to the 12-band multispectral features
3. Continuous response variable binned in three buckets: small, medium and large
4. Inference applied with a binary approach to maximize accuracy (small vs medium/large)

Source: Mingxing Tan, Quoc V. Le 2019, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, https://arxiv.org/abs/1905.11946

Labels

Distribution of US Irrigated Area

256x256 multispectral images (10km^2)

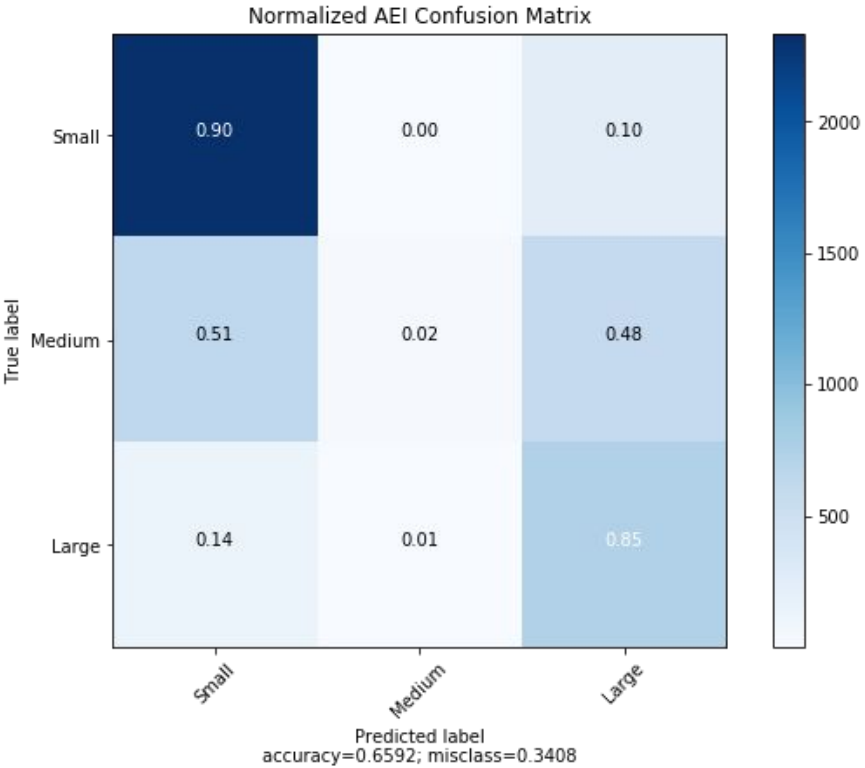


Bins

Label	% Irrigated Area	% distribution of observations
Small	0 - 3.6	33
Medium	3.6 - 15	33
Large	15 - 100	33

23,663 observations
80% Training, 20% Validation

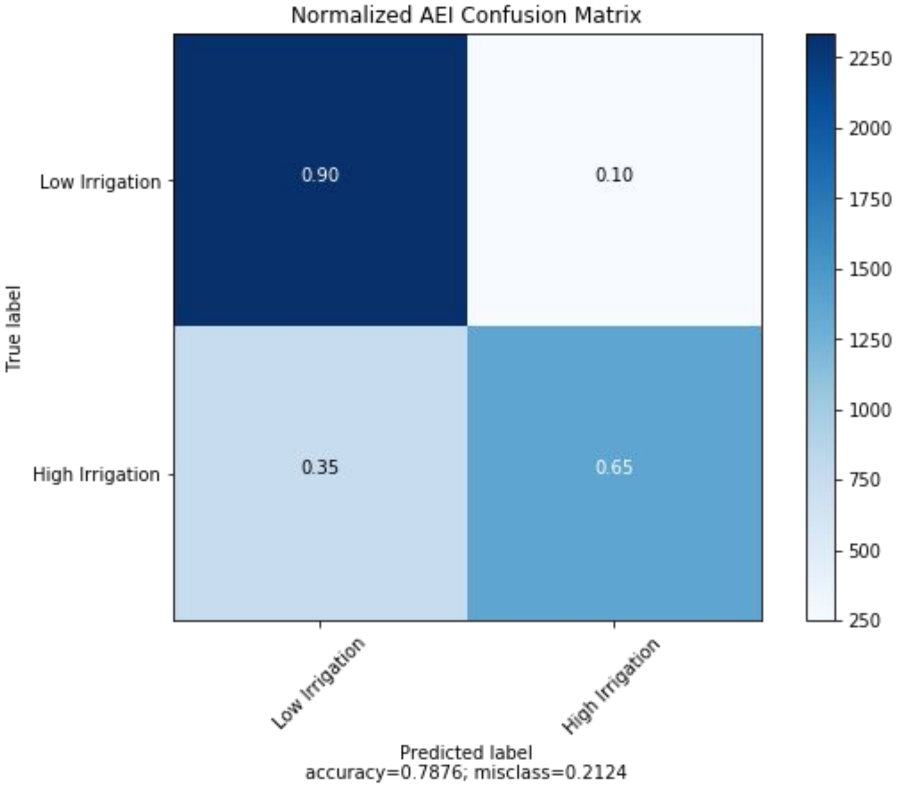
3-Bin Model Results



Analysis

- Good separation between small vs large irrigated areas
- This level of separation was not achieved with other models (e.g. decile or binary bins)
- Medium label with poor performance
- Derivation of a binary: small (low irrigation) vs medium/large (high irrigation)

Binary Model Results

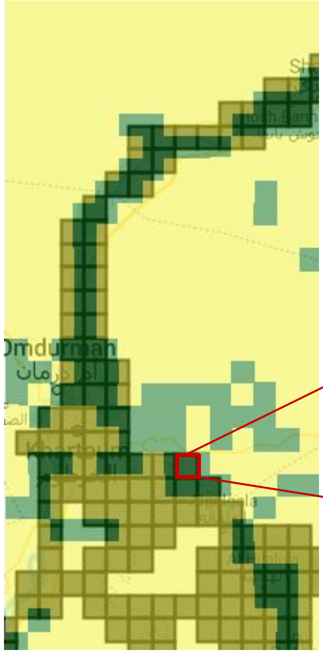


Binary Accuracy	**0.79**
Binary Accuracy for RGB model	**0.70**

Inference

Dahlak Archipelago, Red Sea

Inference - Sudan (2019)

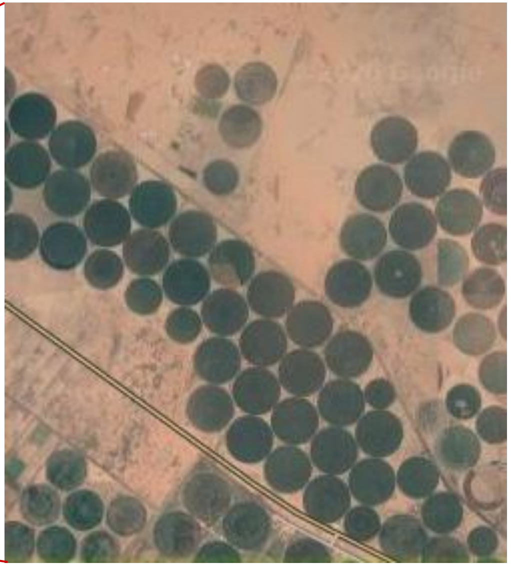


Low irrigation

High irrigation

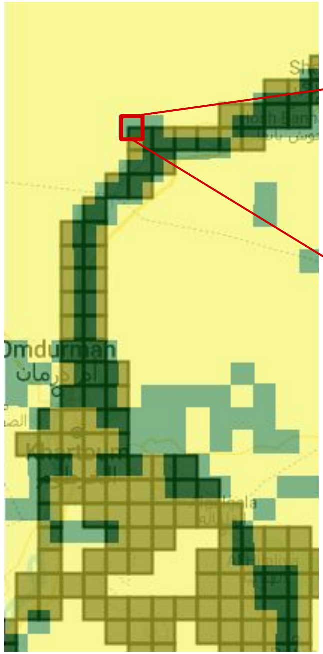
GMIA original tiles

406 tiles



PCT_AEI: 4%
ID: 344851

Inference - Sudan (2019)

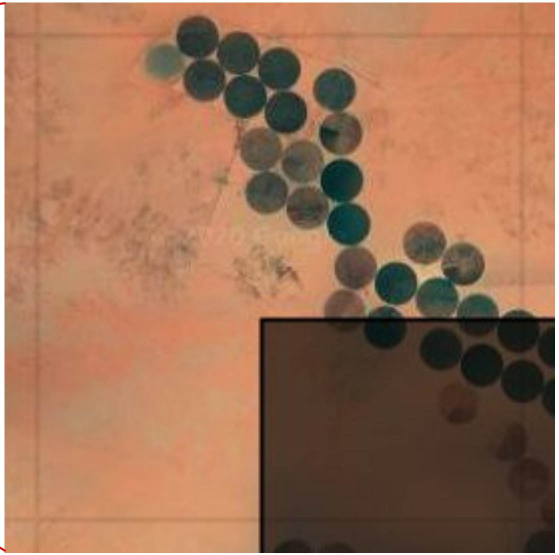


406 tiles

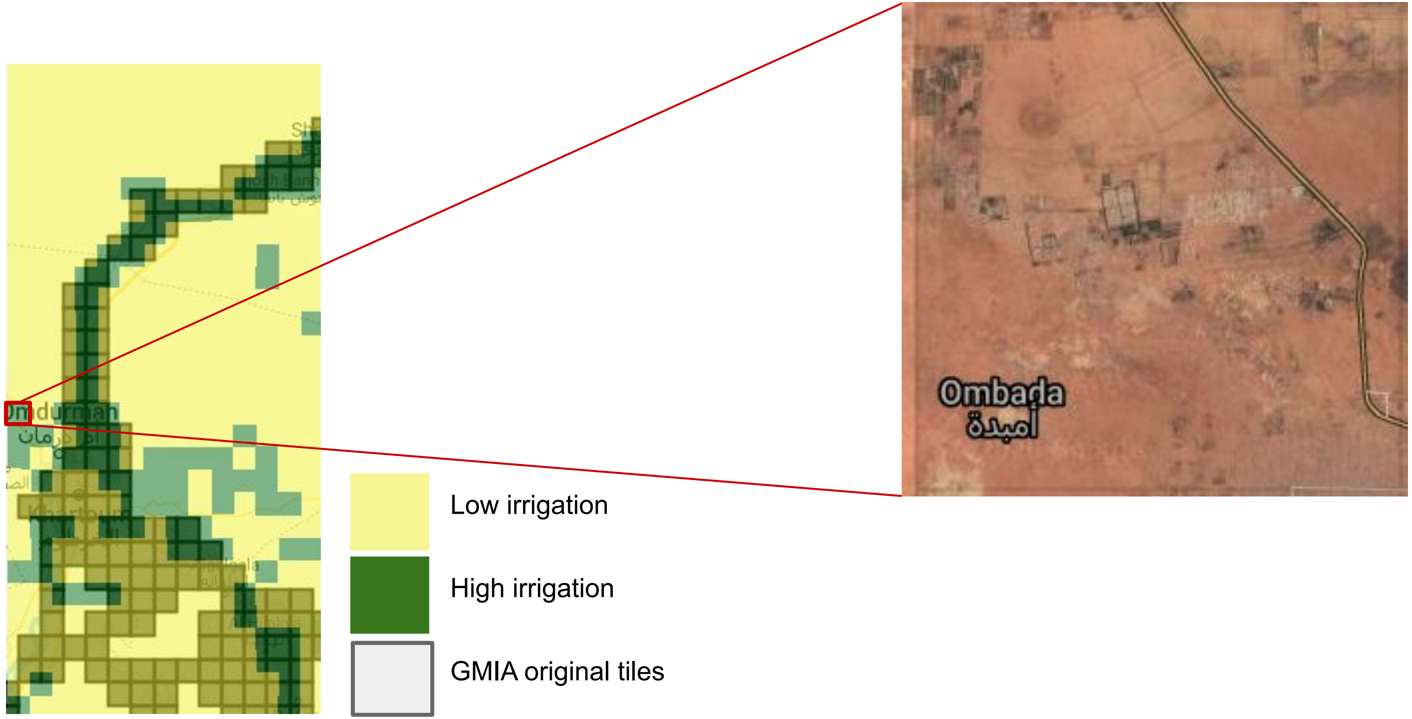
Low irrigation

High irrigation

GMIA original tiles

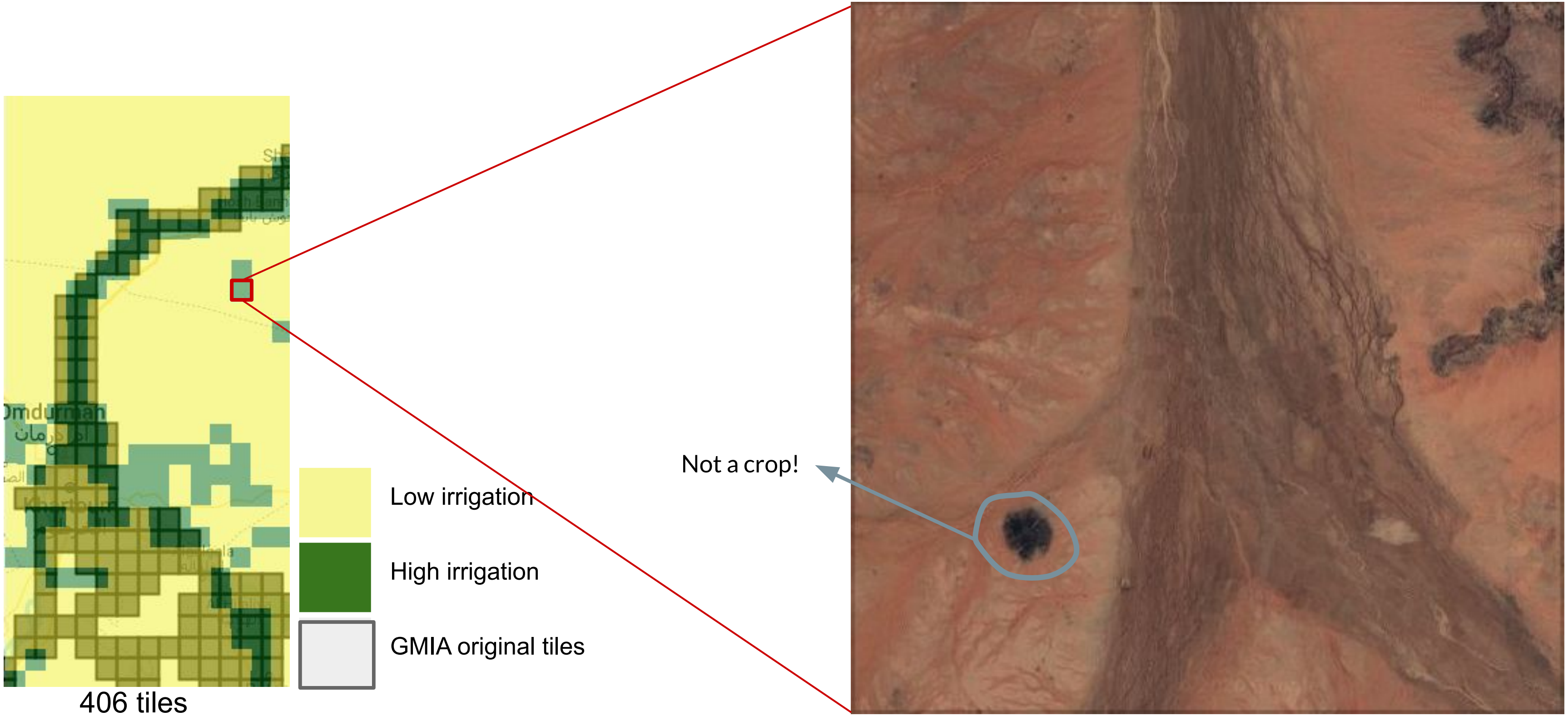


Inference - Sudan (2019)



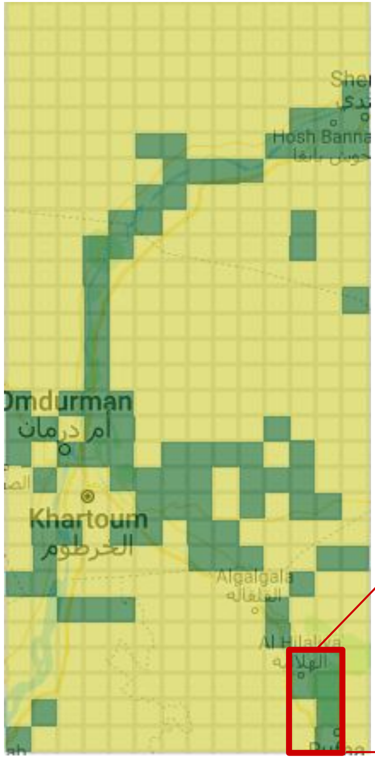
406 tiles

Inference - Sudan (2019)



406 tiles

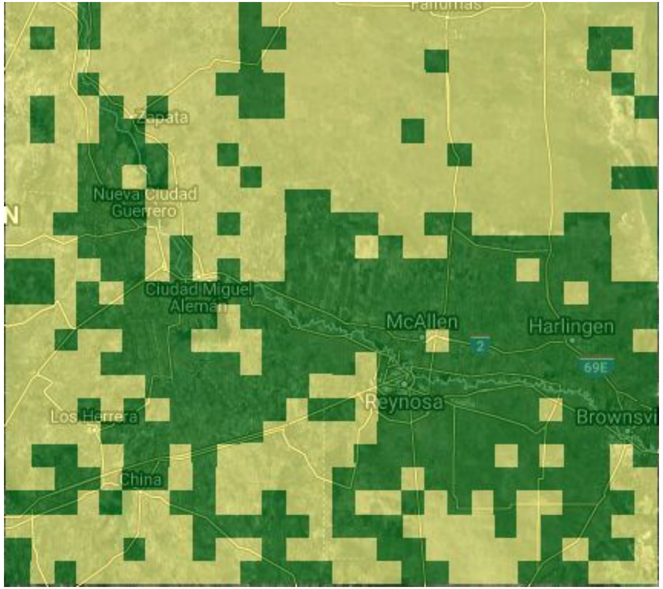
Inference - Sudan (2019)



Low irrigation

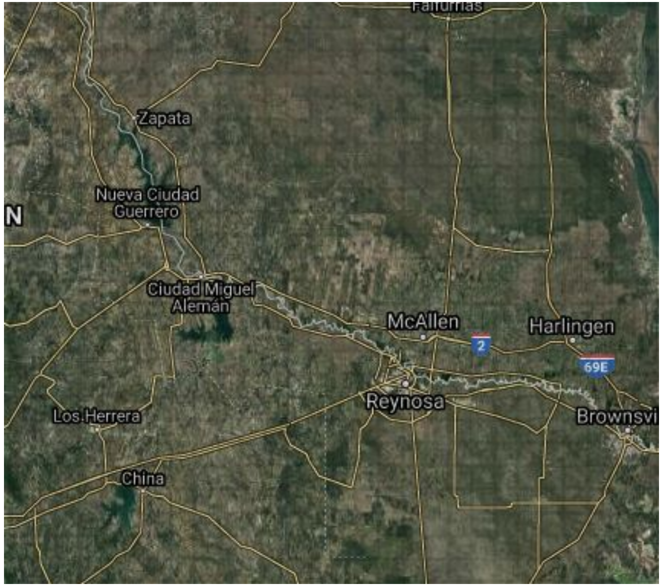
High irrigation

Inference - Rio Grande Basin (2019)



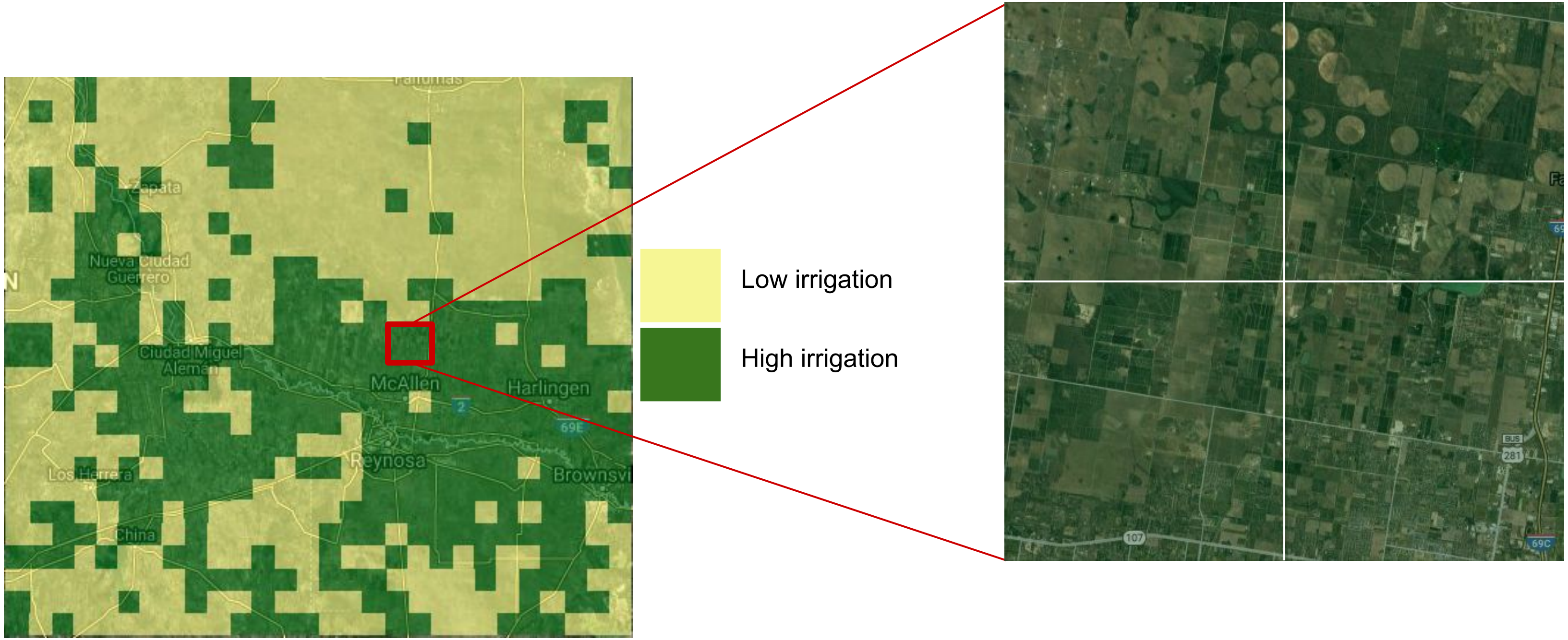
Low irrigation

High irrigation



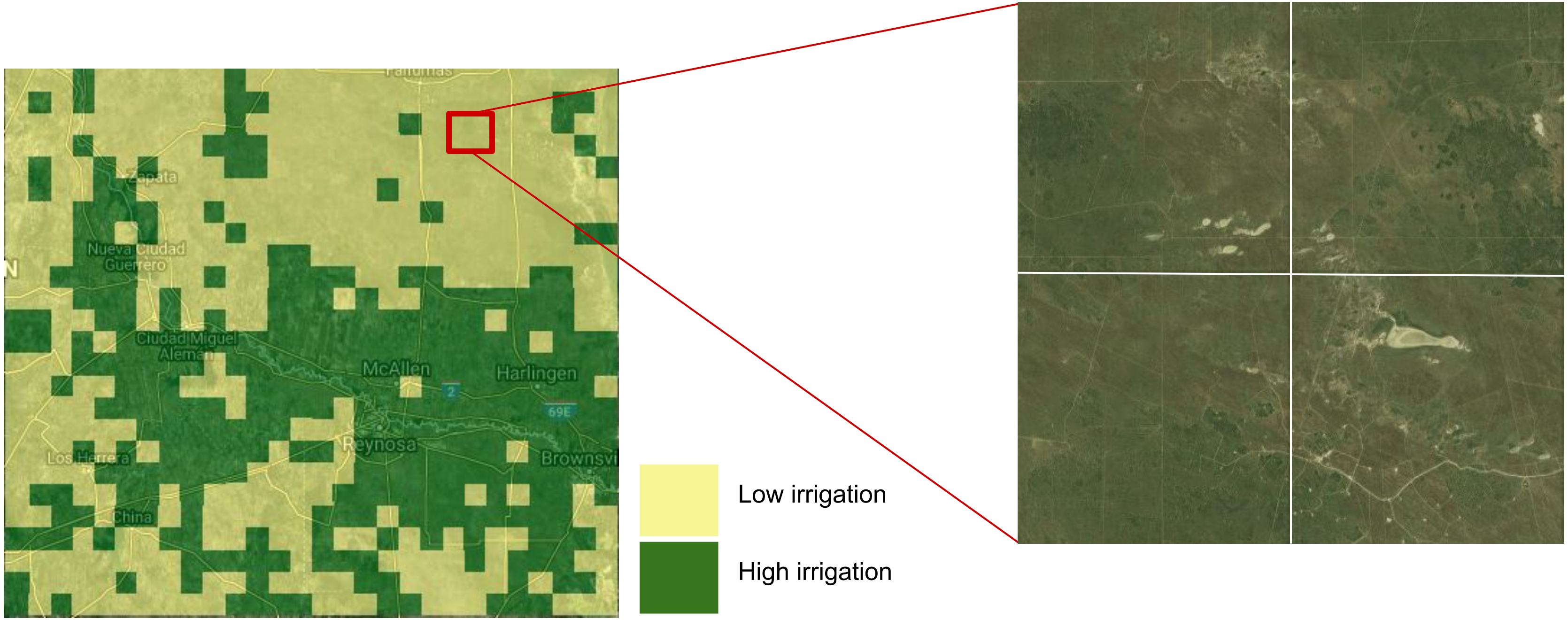
700 tiles

Inference - Rio Grande Basin (2019)



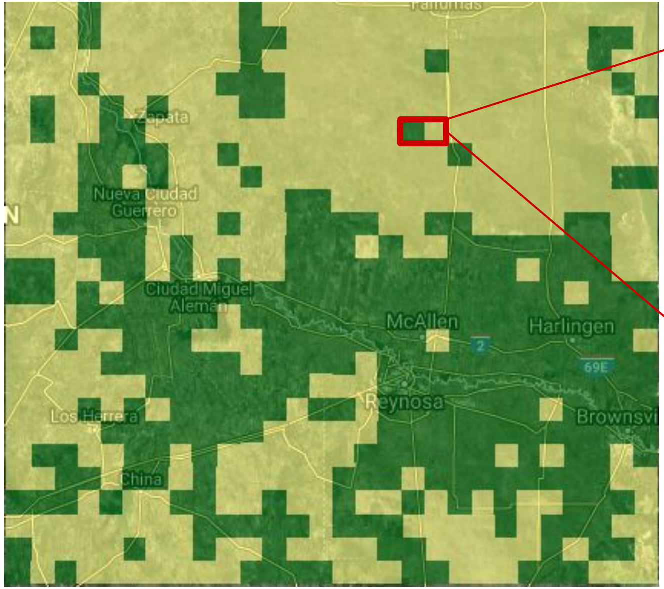
700 tiles

Inference - Rio Grande Basin (2019)



700 tiles

Inference - Rio Grande Basin (2019)



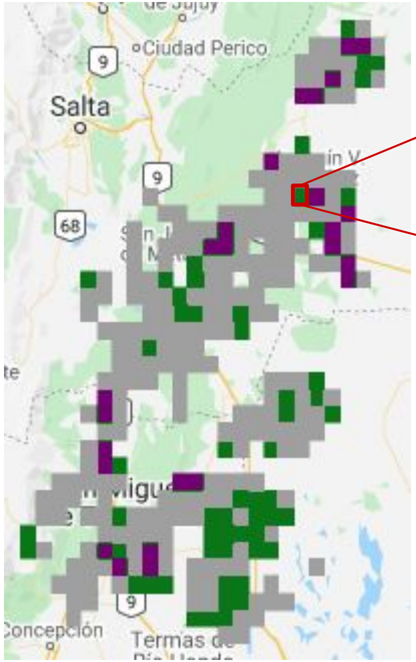
700 tiles

Low irrigation

High irrigation

Inference - Chaco (2019)

Inference - Gain / Loss - Chaco (2019)



2005
PCT_AEI: 2%

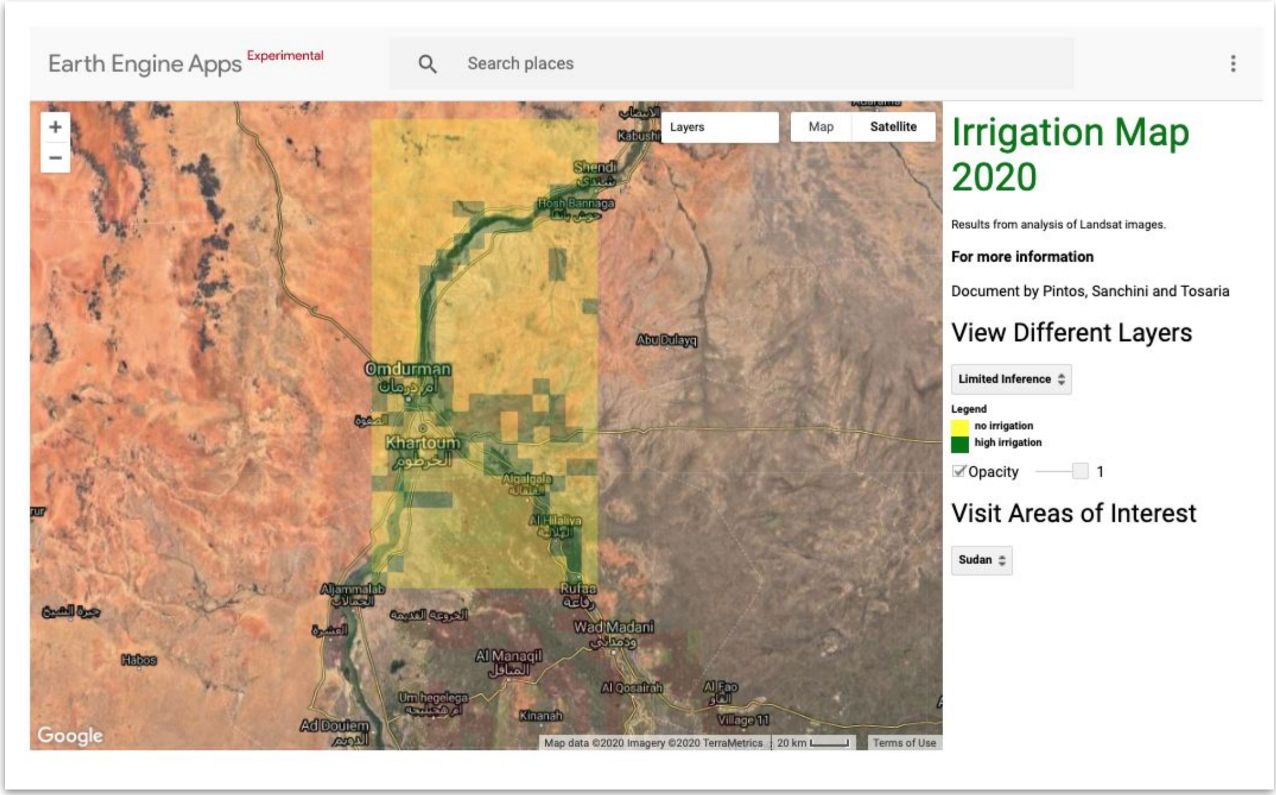
2019
High irrigation

Tile: 440747
Chaco

Web Deliverable

Published in
Google Earth Engine

Irrigation Map 2020



Future Plans

- Include temporal analysis of satellite images to capture the vegetation change across seasons
- Conduct feature importance between raw bands and derived indicators
- Train with more data beyond US
- Geography specific models
- Increase the number of bins (e.g. deciles)
- Apply inference to the global map, derive a new global estimate and trends of irrigated areas

Thanks!

Please visit

www.irrigation2019.org

Sandwich Island

Appendix

Region Selection for Inference

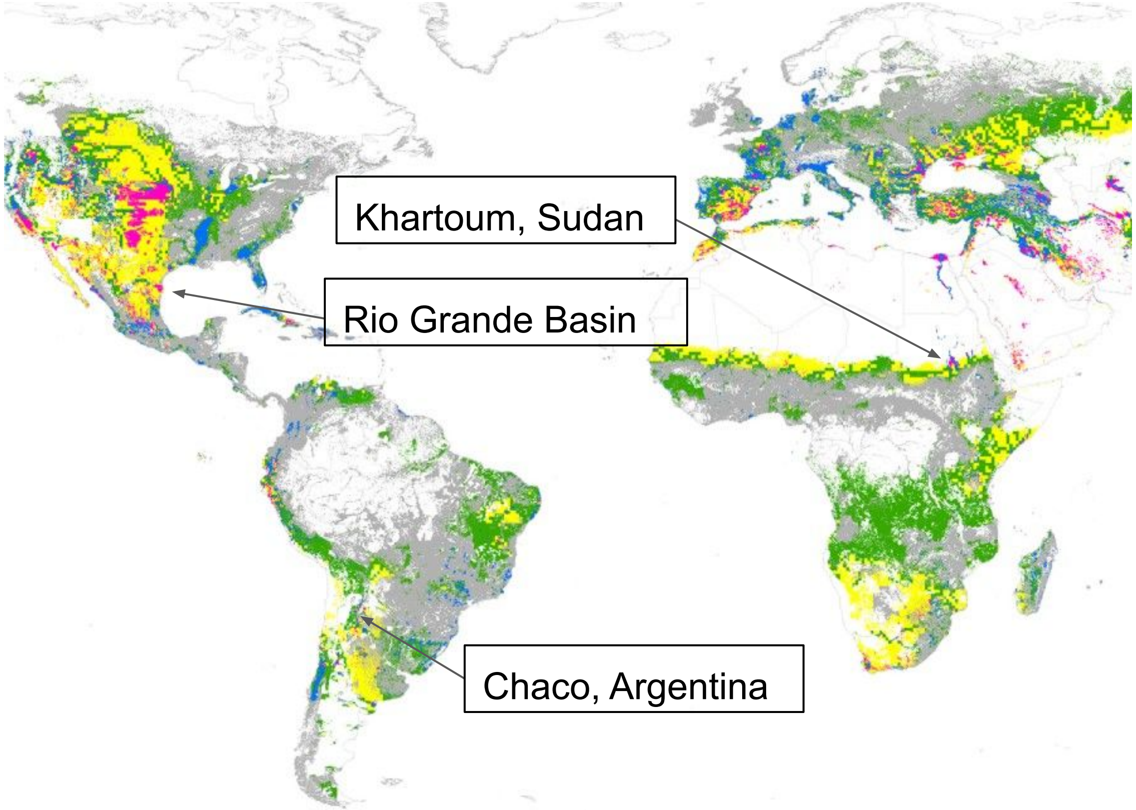
Current Sustainable

Current Unsustainable

Yield Gap Closure Sustainable

Yield Gap Closure Unsustainable

Primarily Rainfed



Souce: Closing the yield gap while ensuring water sustainability - Rosa,Rulli,Davis,Chiarelli,Passera,D'Odorico - Environmental Research Letters - 2018

Performance Metrics of the Binary Model

Use Case: Looking for low irrigated areas		Use Case: Looking for high irrigated areas	
TP	0.90	TP	0.65
FP	0.35	FP	0.10
TN	0.65	TN	0.90
FN	0.10	FN	0.35
Precision	**0.75**	**Precision**	**0.87**
Recall	**0.90**	**Recall**	**0.65**

The Project

Context

The global distribution of irrigated areas is unknown and a global updated map of irrigated areas is missing.

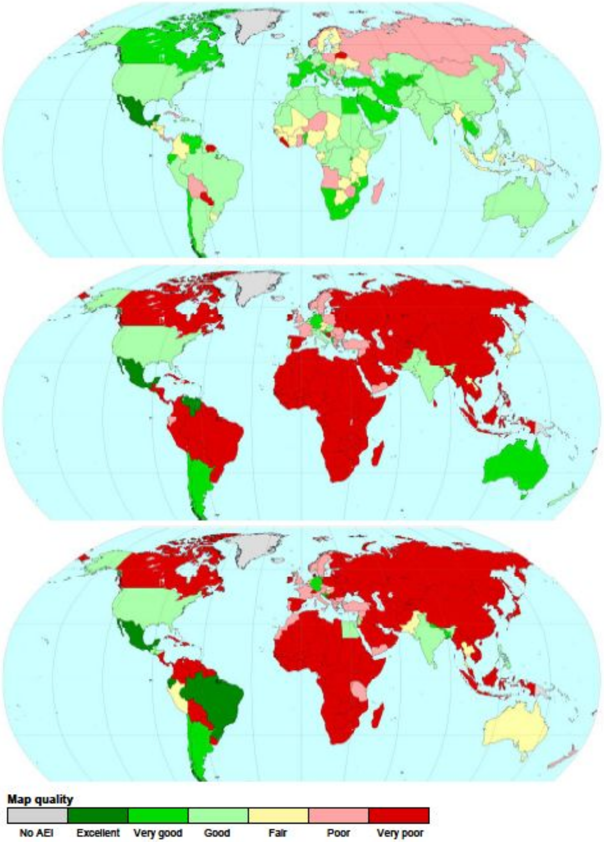
Current maps rely on census and statistical from a wide range of sources all over the world.

FAO from UN maintains and disseminates global info to support agricultural and rural development through sustainable use of water and land.

Goal

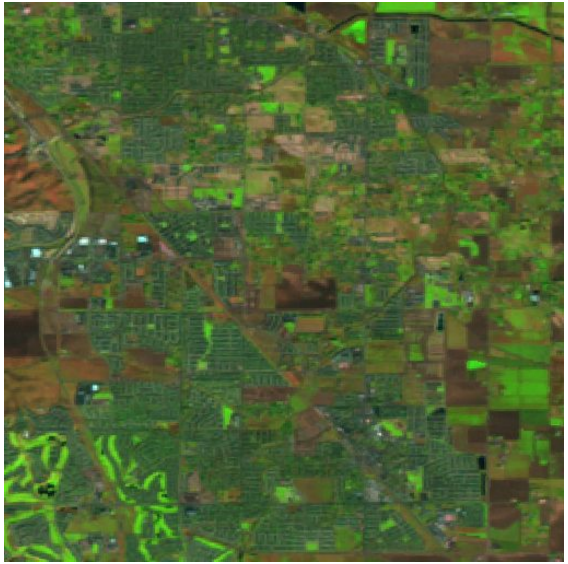
Rely on current curated database census, statistical and maps together with satellite data to train a machine learning model that could update the world map of irrigated crops.

Update the irrigation data with a novel automated method.



Map quality of the layers on area equipped for irrigation (top), area actually irrigated (center) and the water source for irrigation (buttom).

Tile extraction and prediction



Classification

- Area equipped for irrigation?
- Cropland classification?

Datasets: Remote Sensing Data (Landsat)

2005

PCT_AEI: 12%
ID: 181771
(California)

PCT_AEI: 35%
ID: 182783
(California)

PCT_AEI: 79%
ID: 183785
(California)

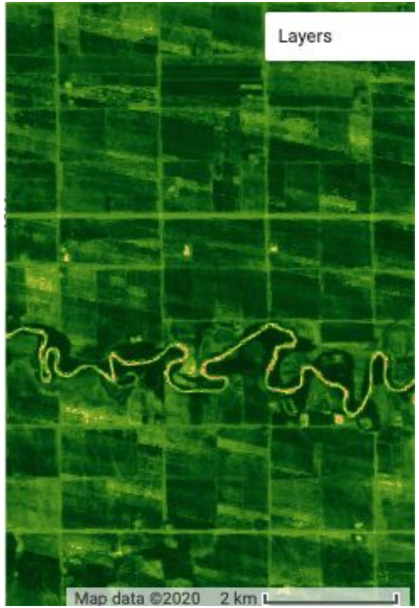
2019

PCT_AEI: ?
ID: 181771
(California)

PCT_AEI: ?
ID: 182783
(California)

PCT_AEI: ?
ID: 183785
(California)

Classification Examples

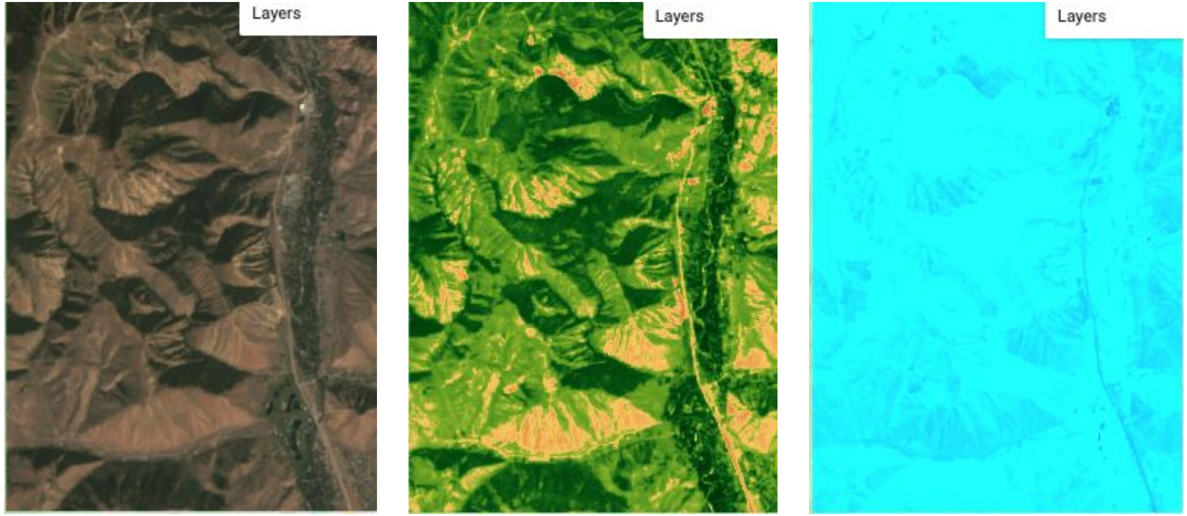


RGB	NDVI	NDWI

True: small
Pred: large

AEI: 1%
ID: 69252

Classification Examples

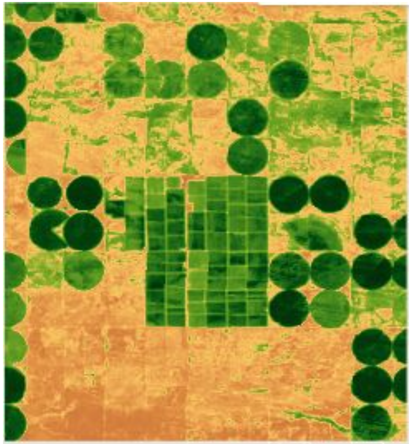


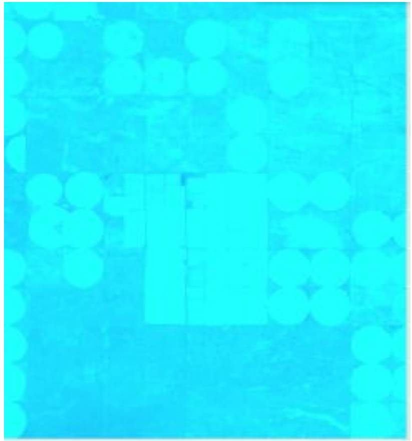
RGB	NDVI	NDWI

True: small
Pred: large

AEI: 1%
ID: 111193

Classification Examples





RGB	NDVI	NDWI

True: large
Pred: large

AEI: 20%
ID: 240405

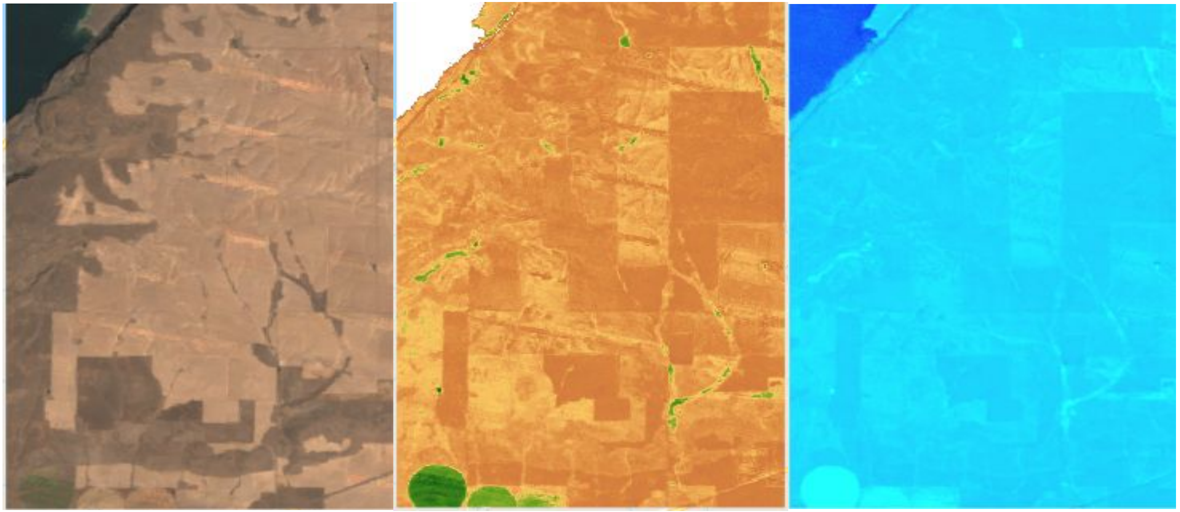
Classification Examples

RGB	NDVI	NDWI

True: large
Pred: large

AEI: 58%
ID: 284701

Classification Examples

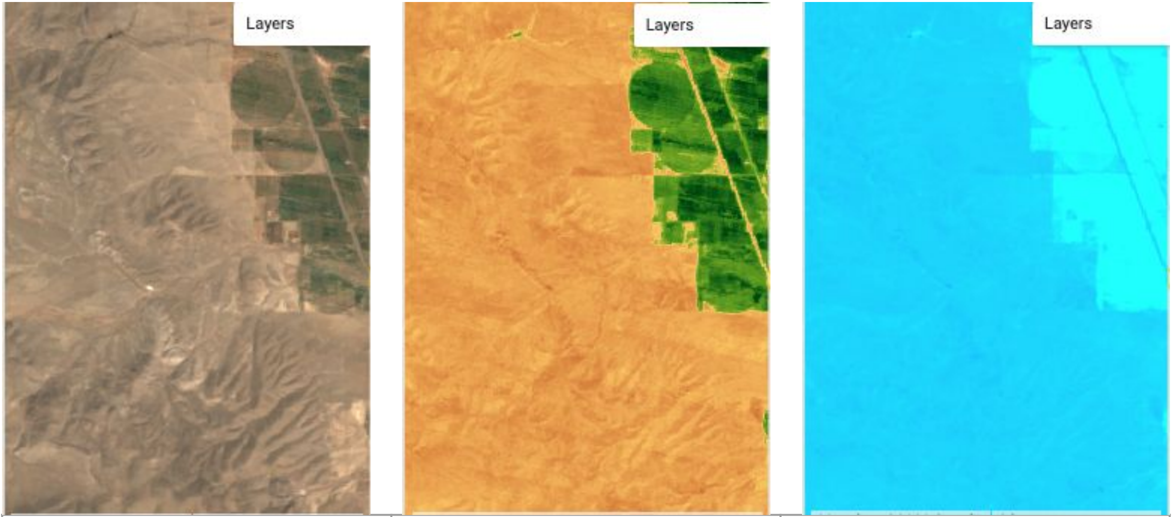


RGB	NDVI	NDWI

True: large
Pred: small

AEI: 70%
ID: 69801

Classification Examples

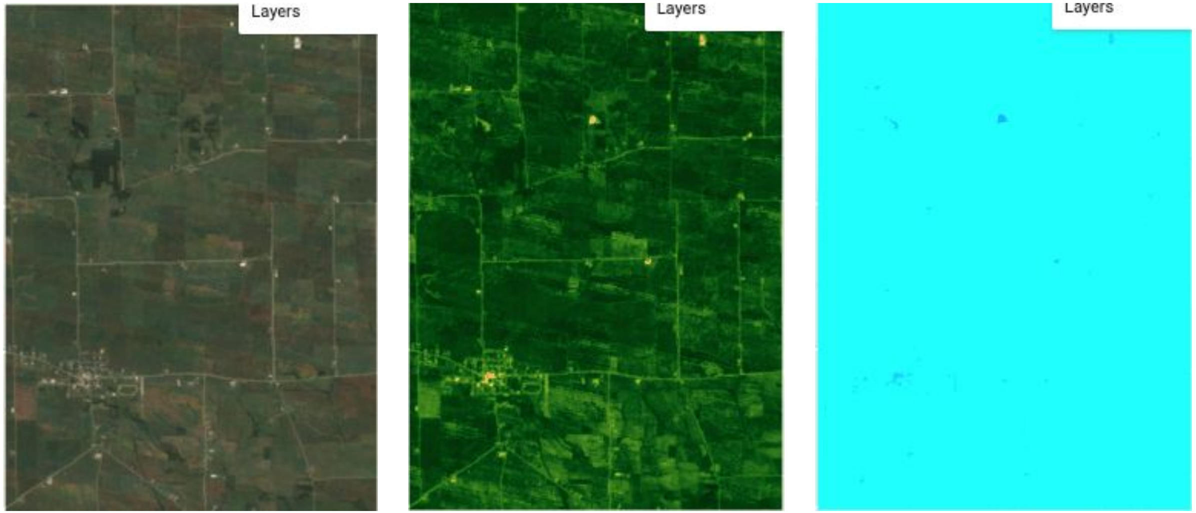


RGB	NDVI	NDWI

True: large
Pred: small

AEI: 36%
ID: 93445

Classification Examples

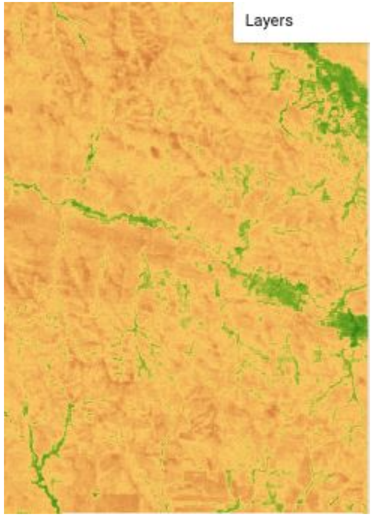


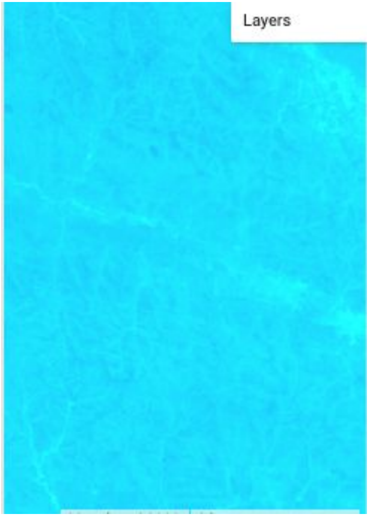
RGB	NDVI	NDWI

True: small
Pred: small

AEI: 3%
ID: 133110

Classification Examples





RGB	NDVI	NDWI

True: small
Pred: small

AEI: 1%
ID: 92640

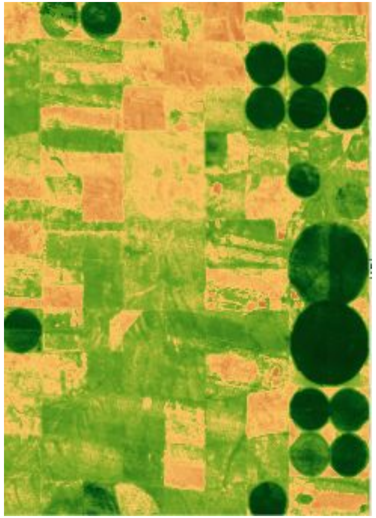
Classification Examples

RGB	NDVI	NDWI

True: medium
Pred: medium

AEI: 12%
ID: 215762

Classification Examples

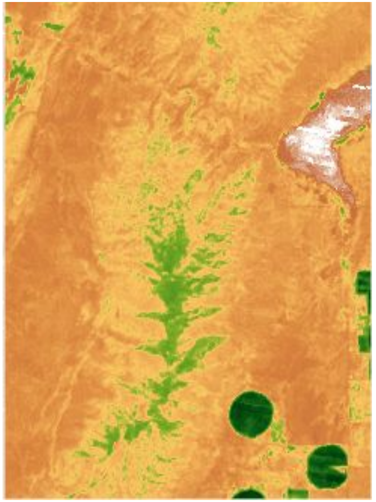
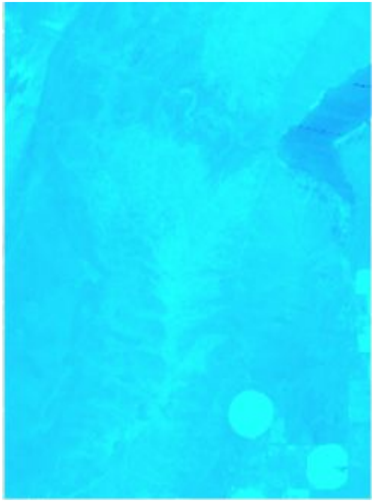




RGB	NDVI	NDWI

True: medium
Pred: medium

AEI: 15%
ID: 164541

Classification Examples

RGB	NDVI	NDWI

True: medium
Pred: large

AEI: 4%
ID: 179809

Our modeling attempts

Baseline Models

Simple multiple blocks composed of convolutions and max pooling, stacked together with a flattened fully connected classification layer.

Semantic Segmentation

Pixel level predictions using U-Net neural network architectures.

Leverage of other irrigation maps such as GFSAD.

Data Augmentation

Multiple combinations of rotation and image flipping.

Feature Selection

Tournament of models with different combinations of features (raw and engineered)

The GFSAD Model: U-NET Architecture

Our baseline to predict the cropland pixel classification

2

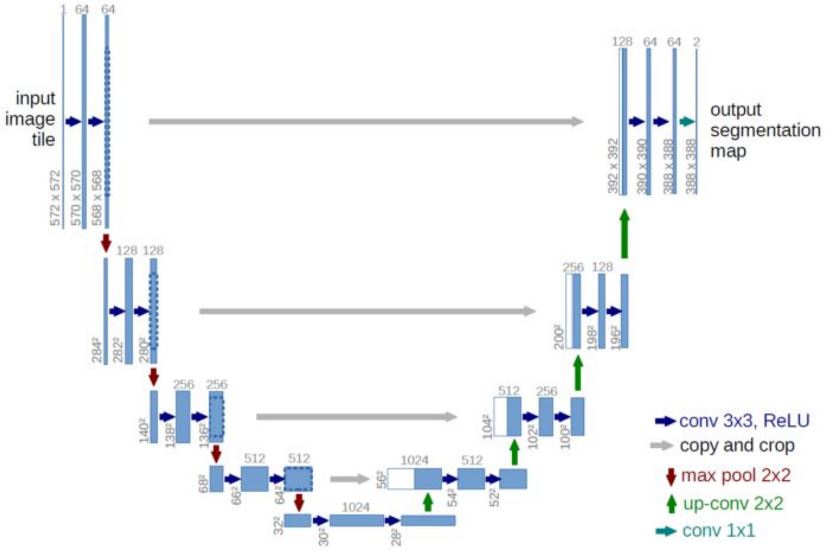
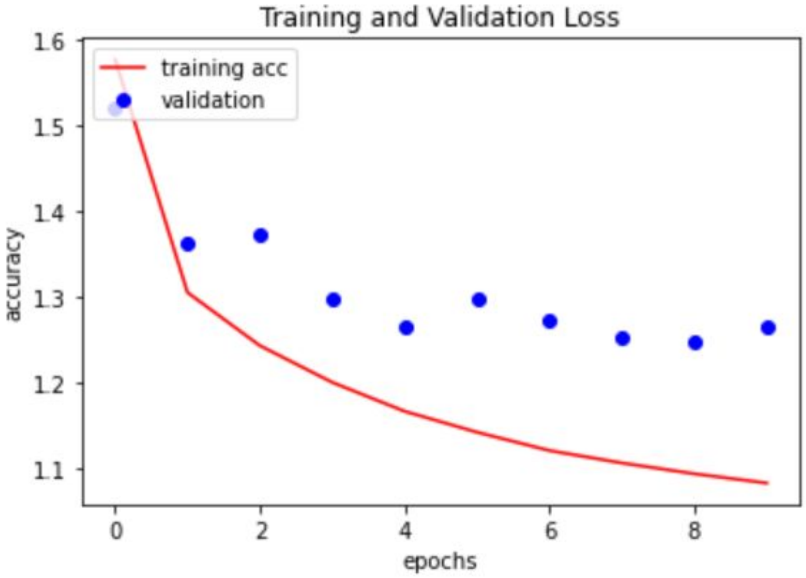


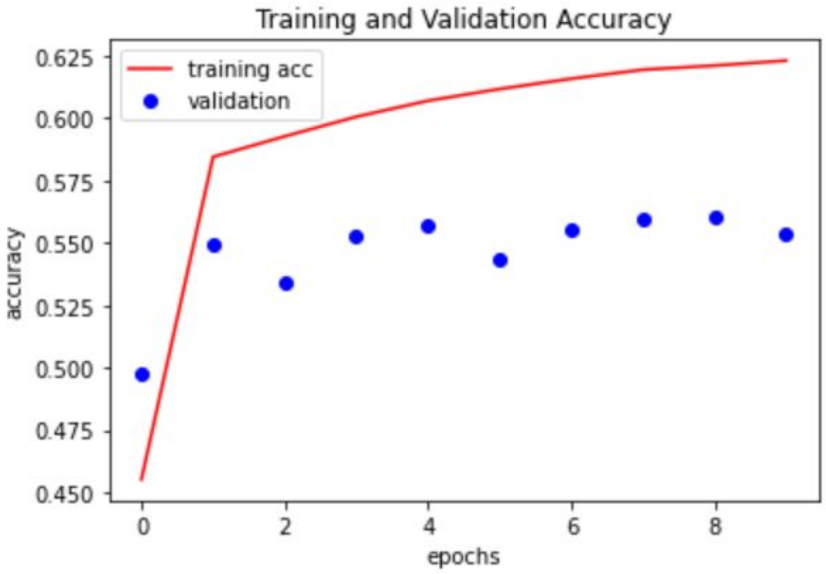
Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Color	Description
black	Non-croplands
orange	Croplands: irrigation major
brown	Croplands: irrigation minor
02a50f	Croplands: rainfed
green	Croplands: rainfed, minor fragments
yellow	Croplands: rainfed, very minor fragments

GFSAD Pixel Segmentation Model: Results

Training data ~23,300
EPOCHS - 10

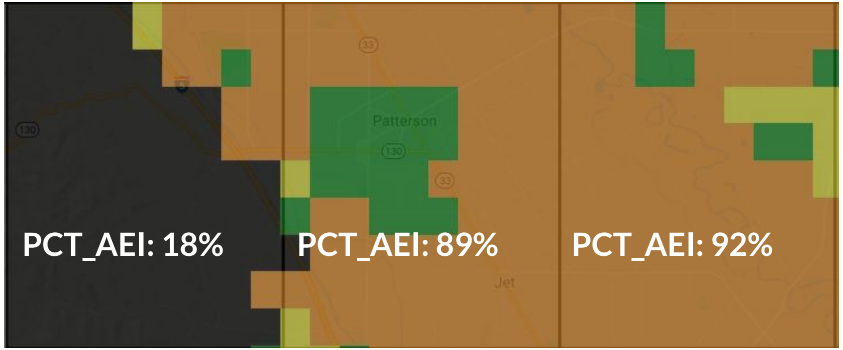




Relation between AEI and GFSAD

Input - [256, 256] - Pixel class [0..5]

Output - Area under Irrigation class [0...5]



CLASS 4 CLASS 6 CLASS 6

The Model: Next steps to increase performance

Some challenges

Size in disk: basic statistics, such as counting the number of observations, takes time must be made with a batch approach

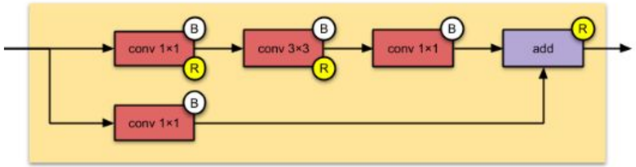
Shuffling and splitting: for the baseline we shuffled and split (train/test) based on filename approach (200MB each)

Model size: baseline model already has 42 million parameters

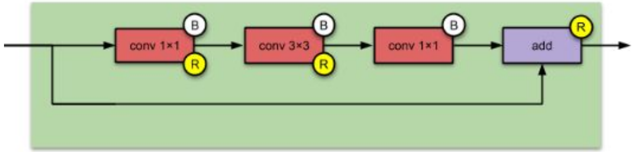
Labels: Highly concentrated in small areas (less than 5%)

Challengers Models

Ideas from ResNet-50

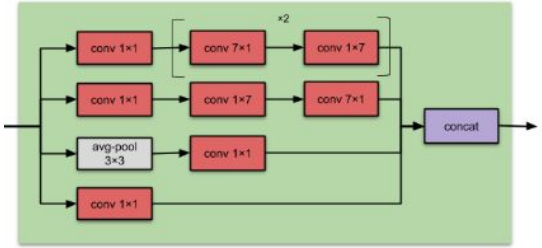


Conv block



Identity block

Ideas from Inception-V4



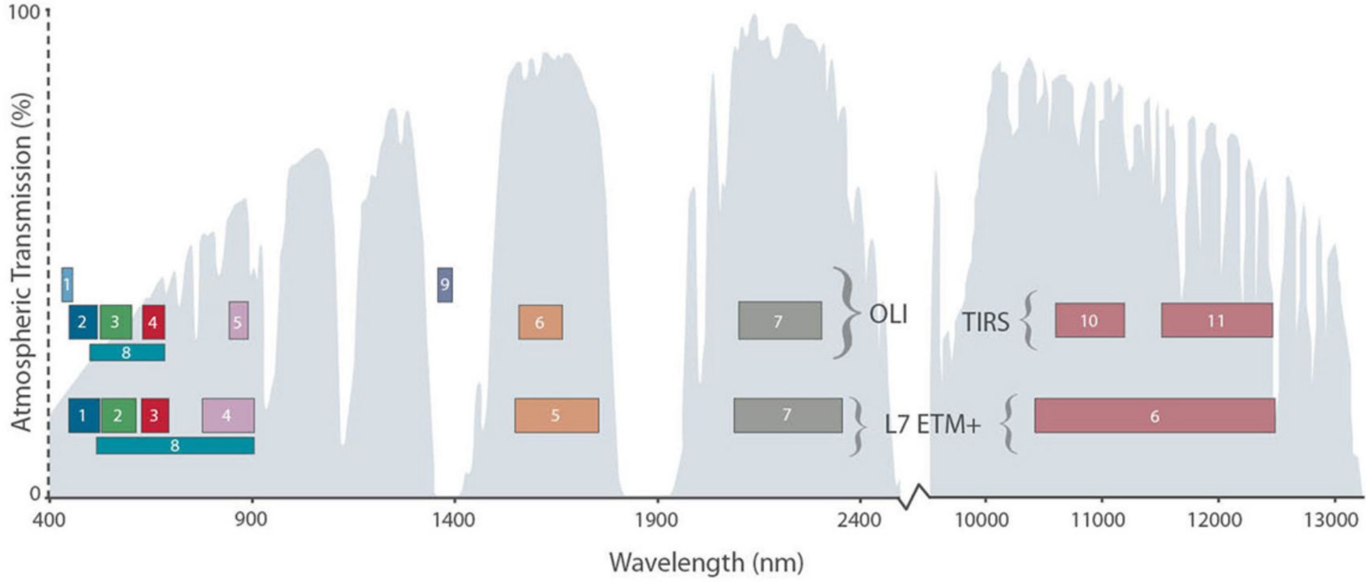
Inception-B

Next Steps

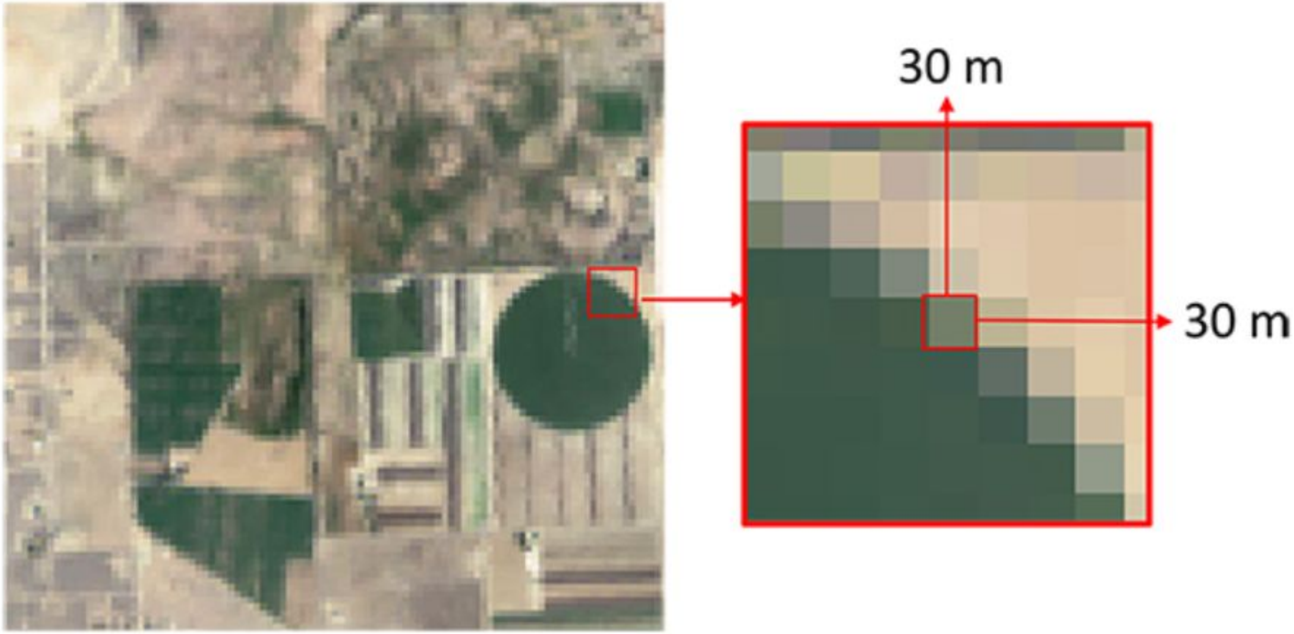
	March 19	March 26	April 2	April 9	April 16
Training	- Error analysis - Challenger models design	- Best model results			
Inference	- Landsat 2020 image extractions	- Preliminary inference results	- Inference on selected US states		
Data Product	- Check with Prof. D'Odorico design of web deliverable	- Check with Prof. D'Odorico design of web deliverable	- Check with Prof. D'Odorico design of web deliverable	Deploy Irrigation 2020 map to GEE	- Deploy Irrigation 2020 map to GEE - Final Presentation

Individual Contributions
Research, EDA: all
Feature engineering/Image treatment: all
Tensors Extraction: Sam
Model Training, Inference: Carlos & Sam
Web Deliverable: Laura
Presentation materials: all

What is Surface Reflectance ?



What is a resolution?



Each Landsat 8 pixel is 30m x 30m or $900m^2$

Overview

Detecting irrigated areas using AI

	Our Focus	Where we are
1 Label Data	GMIA data from FAO - area equipped for irrigation as percentage of the area (10km resolution)	Evaluating other potential label datasets such as GFSAD1000 from NASA: fine grained labels and greater resolution (1km)
2 Input Data	LANDSAT 7 satellite - wide range of wavelengths used in combination as indicators the characterize vegetation: Visual RGB, NDVI, EVI, NDMI, Temperature	Minor definitions on how to approach temperature and bad image quality (clouds)
3 Data Viz & Prep	Google Earth Engine Geopandas Rasterio	Extraction of standardized images for each label entry
4 Modeling	Convolutional Neural Networks with Tensorflow	Creation of tensors from images

Workflow

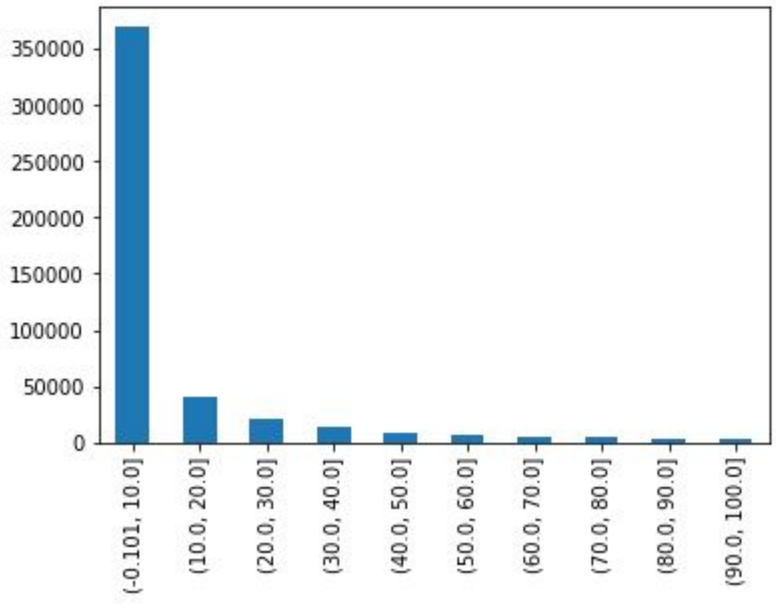
Analysis	Geospatial Mapping	Model	Deployment
Python Notebooks	Google Earth Engine ArcGIS	Tensorflow	Google Earth Engine Platform

Python Shape Objects → Javascript Variables → Python Image Tensors → Javascript Variables

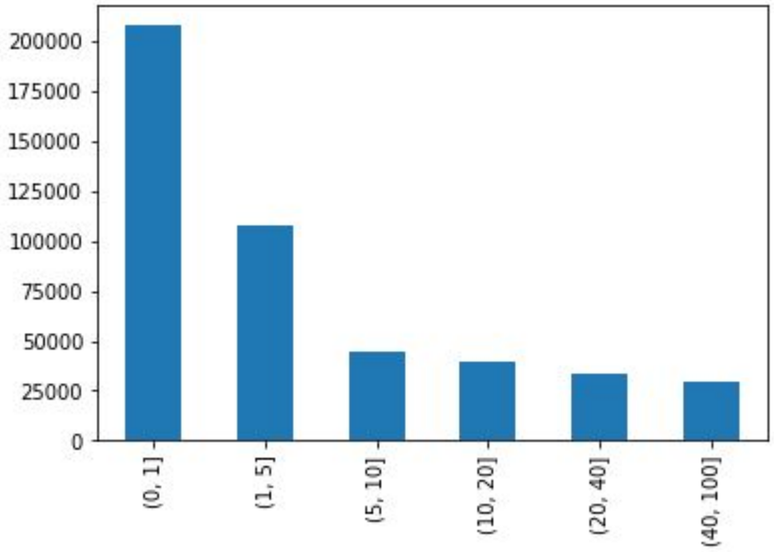
GMIA EDA

GMIA - Area Equipped for Irrigation (AEI)

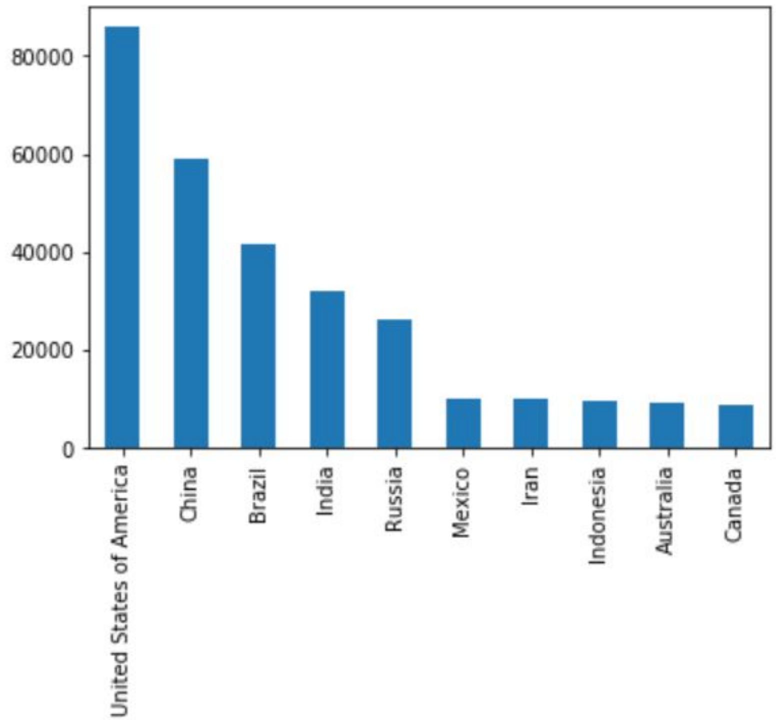
Deciles



Custom bins



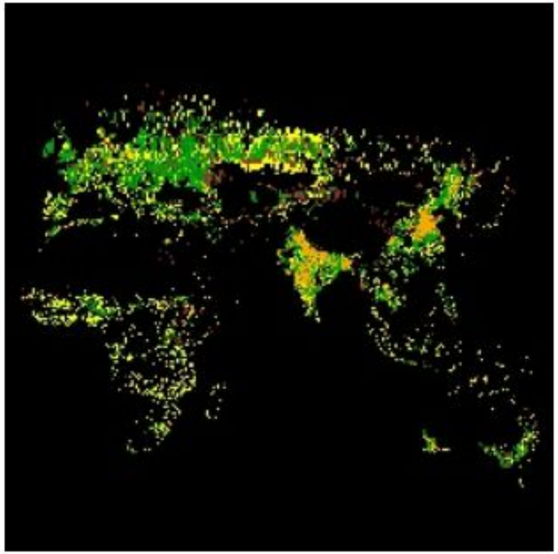
GMIA - Area Equipped for Irrigation (AEI) - by country



Datasets: GFSAD1000: Cropland Extent 1km Multi-Study Crop Mask, Global Food-Support Analysis Data

NASA funded project to provide high resolution global cropland data and their water use that contributes towards global food security in the twenty-first century. The GFSAD30 products are derived through multi-sensor remote sensing data (e.g., Landsat, MODIS, AVHRR), secondary data, and field-plot data and aims at documenting cropland dynamics from 1990 to 2017.

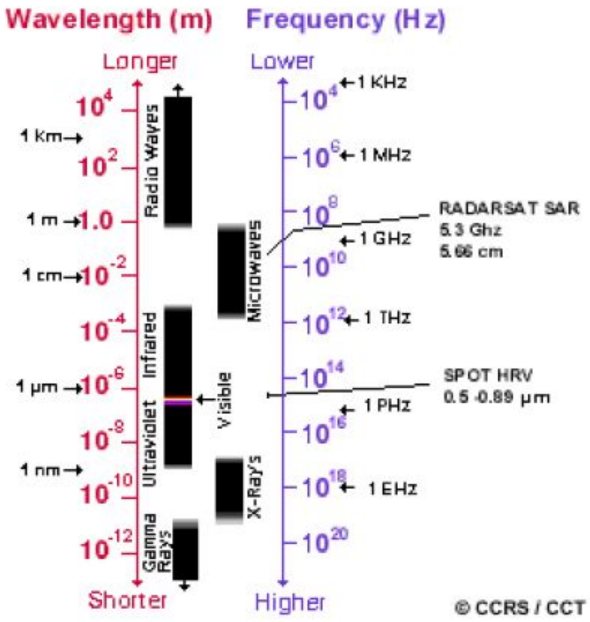
Color	Description
black	Non-croplands
orange	Croplands: irrigation major
brown	Croplands: irrigation minor
02a50f	Croplands: rainfed
green	Croplands: rainfed, minor fragments
yellow	Croplands: rainfed, very minor fragments



http://geography.wr.usgs.gov/science/croplands/

Landasat 7 Satellite

Spectral range



Landsat Sensors

Channel	Wavelength Range (μm)	Application
TM 1	0.45 - 0.52 (blue)	soil/vegetation discrimination; bathymetry/coastal mapping; cultural/urban feature identification
TM 2	0.52 - 0.60 (green)	green vegetation mapping (measures reflectance peak); cultural/urban feature identification
TM 3	0.63 - 0.69 (red)	vegetated vs. non-vegetated and plant species discrimination (plant chlorophyll absorption); cultural/urban feature identification
TM 4	0.76 - 0.90 (near IR)	identification of plant/vegetation types, health, and biomass content; water body delineation; soil moisture
TM 5	1.55 - 1.75 (short wave IR)	sensitive to moisture in soil and vegetation; discriminating snow and cloud-covered areas
TM 6	10.4 - 12.5 (thermal IR)	vegetation stress and soil moisture discrimination related to thermal radiation; thermal mapping (urban, water)
TM 7	2.08 - 2.35 (short wave IR)	discrimination of mineral and rock types; sensitive to vegetation moisture content

Types of Images

RAW

At sensor data which do not account for solar or atmospheric interferences in the radiation.

TOA Reflectance

Top of Atmosphere (TOA) Reflectance is a unitless measurement which provides the ratio of radiation reflected to the incident solar radiation on a given surface.

Corrects for effects of different solar zenith angles due to the time difference between data acquisitions

Compensates for different values of the solar irradiance arising from spectral band differences.

Surface Reflectance (our choice)

Additional to TOA corrections, Surface reflectance (SR) improves comparison between multiple images over the same region by accounting for atmospheric effects such as aerosol scattering and thin clouds, which can help in the detection and characterization of Earth surface change. Less bands available.

Remote Sensing Indicators for Crop Monitoring

NDVI

Formula
(B4 – B3)/(B4+B3)

Concept

Normalized Difference Vegetation Index (NDVI) quantifies vegetation by measuring the difference between near-infrared (NIR) (which vegetation strongly reflects) and red light (which vegetation absorbs).

Healthy vegetation (chlorophyll) reflects more near-infrared (NIR) and green light compared to other wavelengths. But it absorbs more red and blue light.

EVI

Formula
2.5 * ((B4 - B3)/ (B4 + 6*B3 - 7.5*B1 + 1)

Concept

The Enhanced Vegetation Index simultaneously corrects NDVI results for atmospheric influences and soil background signals, especially in areas of dense canopy.

Compared to NDVI, it has improved sensitivity in high biomass regions.

NDMI

Formula
(B4 – B5)/(B4+B5)

Concept

The NDMI (Normalized Difference Moisture Index) describes the crop’s water stress level and is calculated as the ratio between the difference and the sum of the refracted radiations in the near infrared (NIR) and short-wave infrared SWIR).

The interpretation of the absolute value of the NDMI makes possible to immediately recognize the areas of the farm or field with water stress problems.

Global Map of Irrigated Areas (GMIA)

Irrigation Data Aggregation

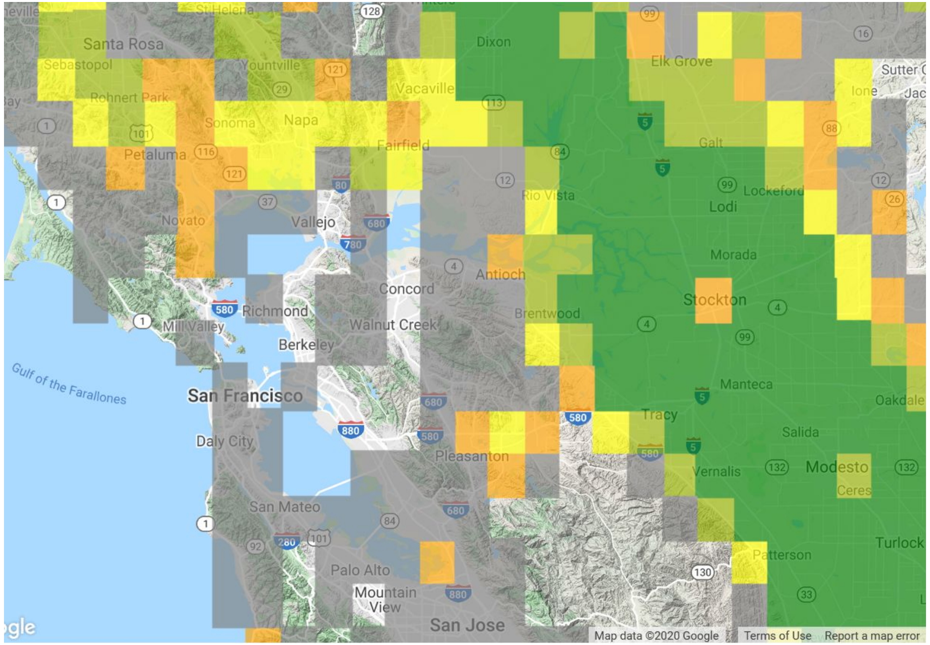
International (GMIA)

National

State

County

Municipality



Opodepe, Sonora, Mexico

30.041666, -110.625000

VISUAL RGB

Left
Sep/Oct 2005
End of the rainy season

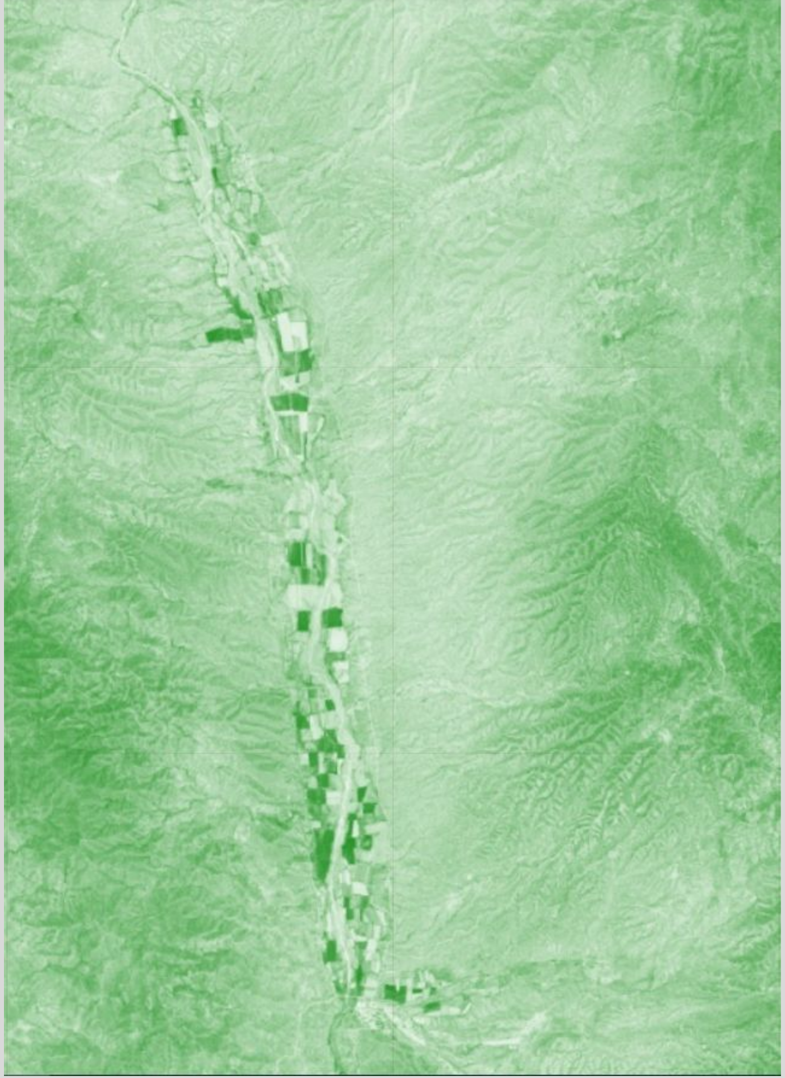
Right
Mar/Apr 2005
End of the dry season

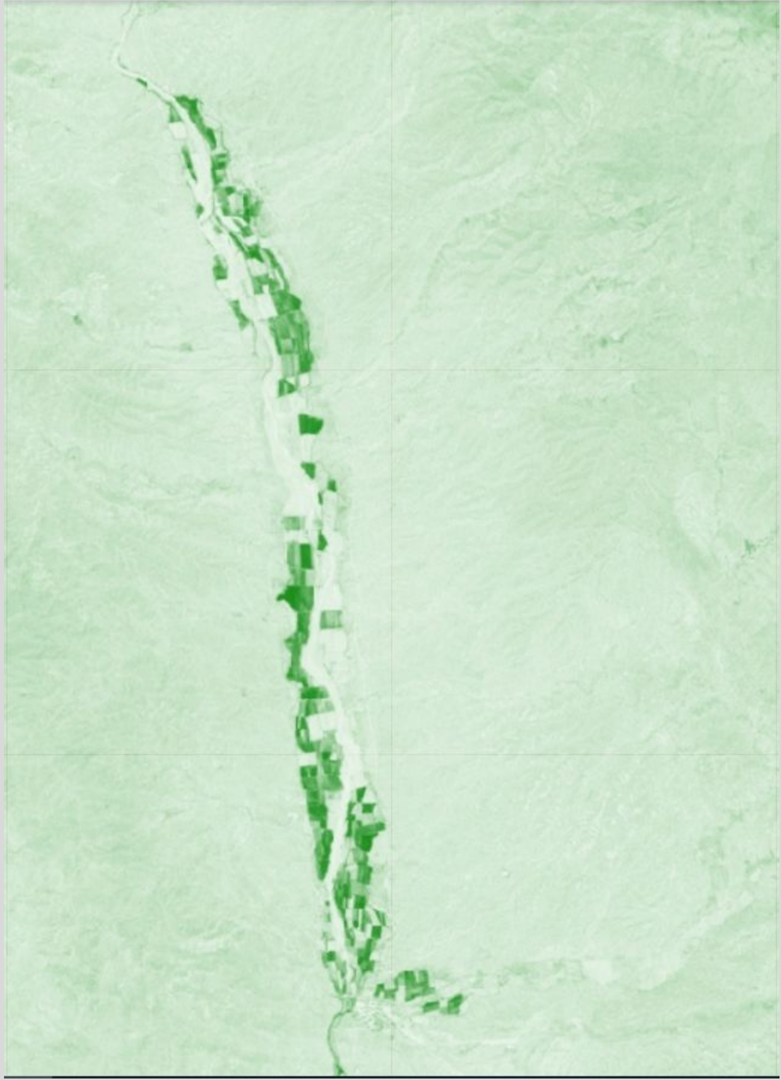
NDVI

Left
Sep/Oct 2005
End of the rainy season

Right
Mar/Apr 2005
End of the dry season

Legend
Darker green represents a more robust vegetation.



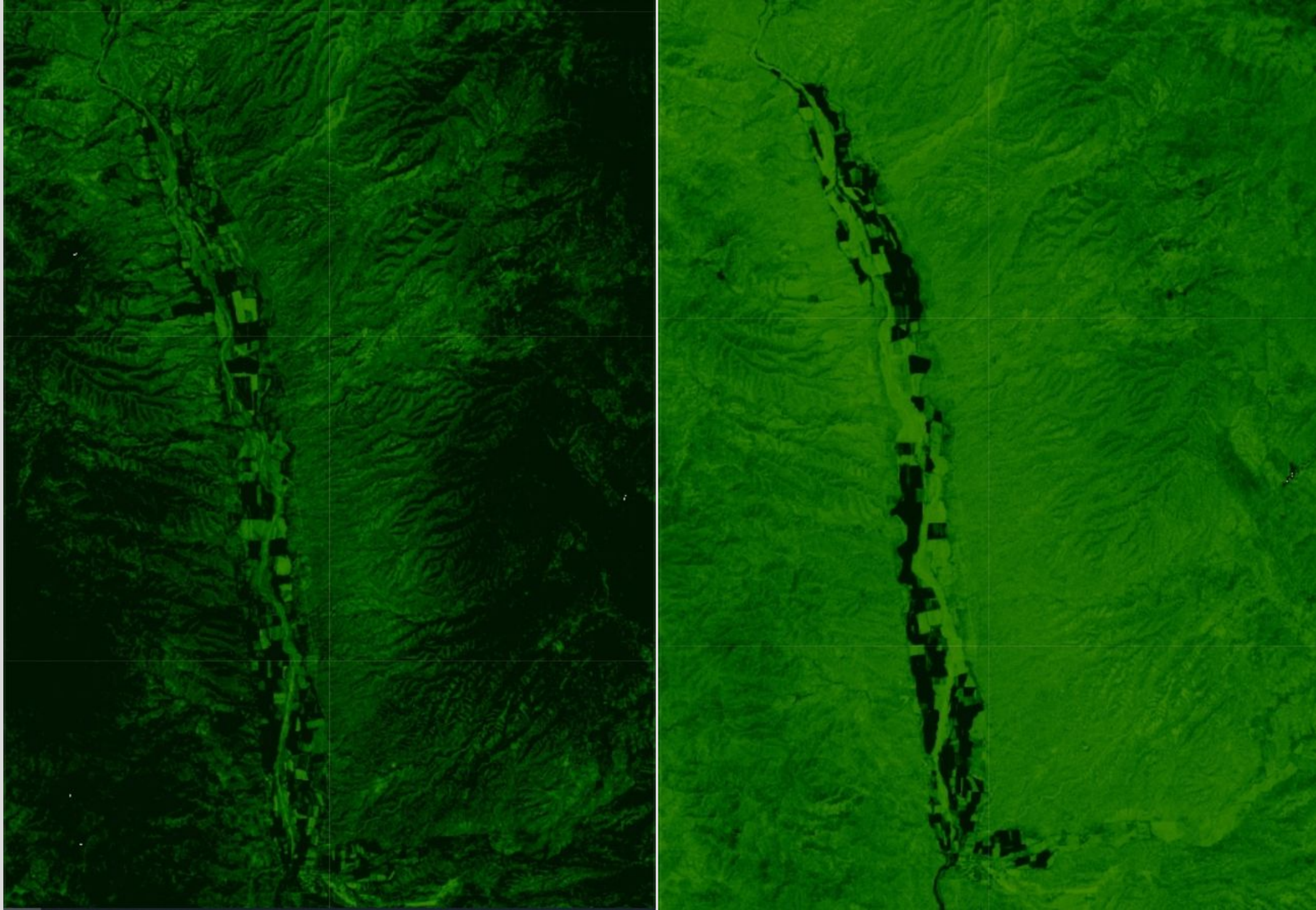


EVI

Left
Sep/Oct 2005
End of the rainy season

Right
Mar/Apr 2005
End of the dry season

Legend
Darker green represents a more robust vegetation.

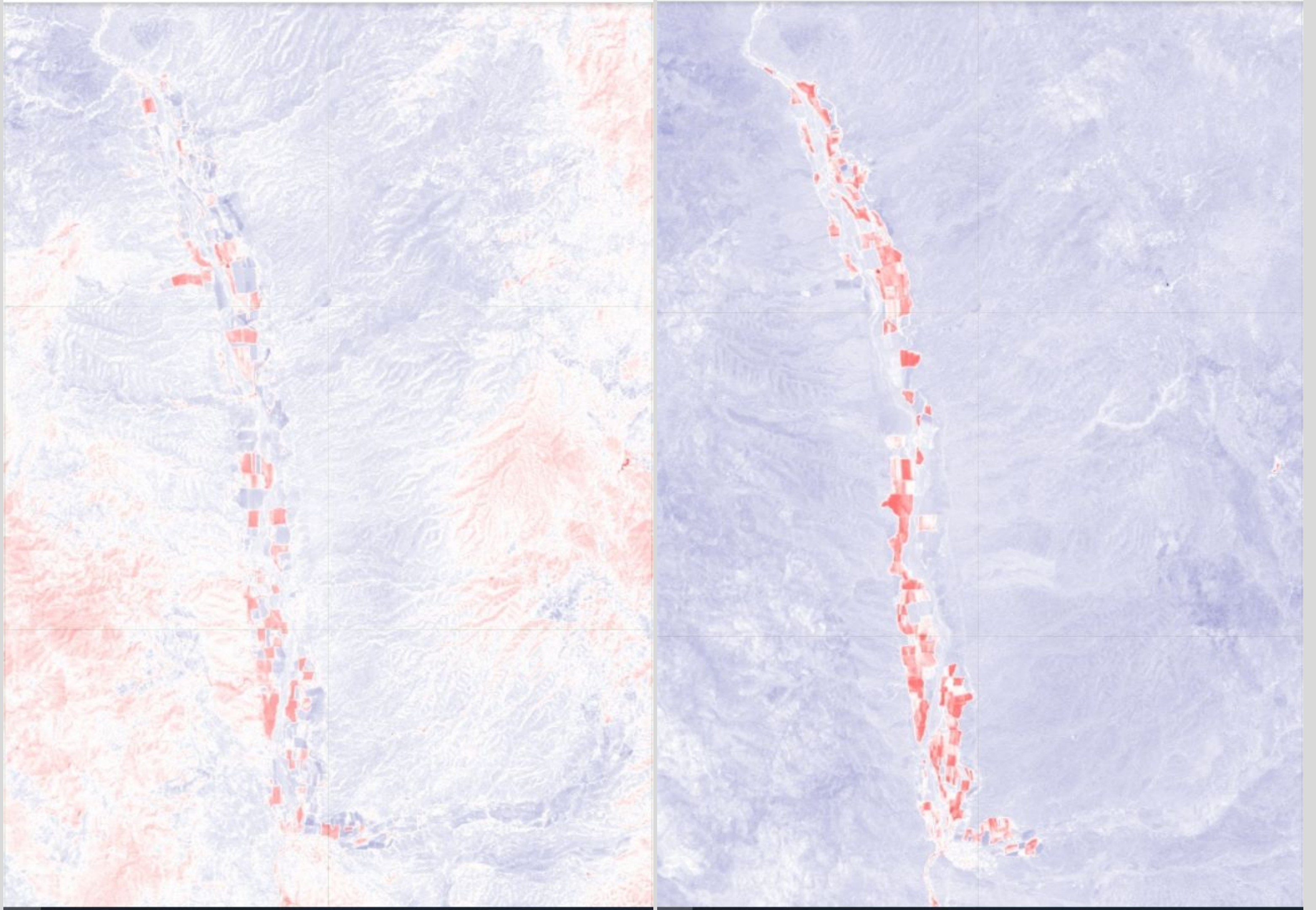


NDMI

Left
Sep/Oct 2005
End of the rainy season

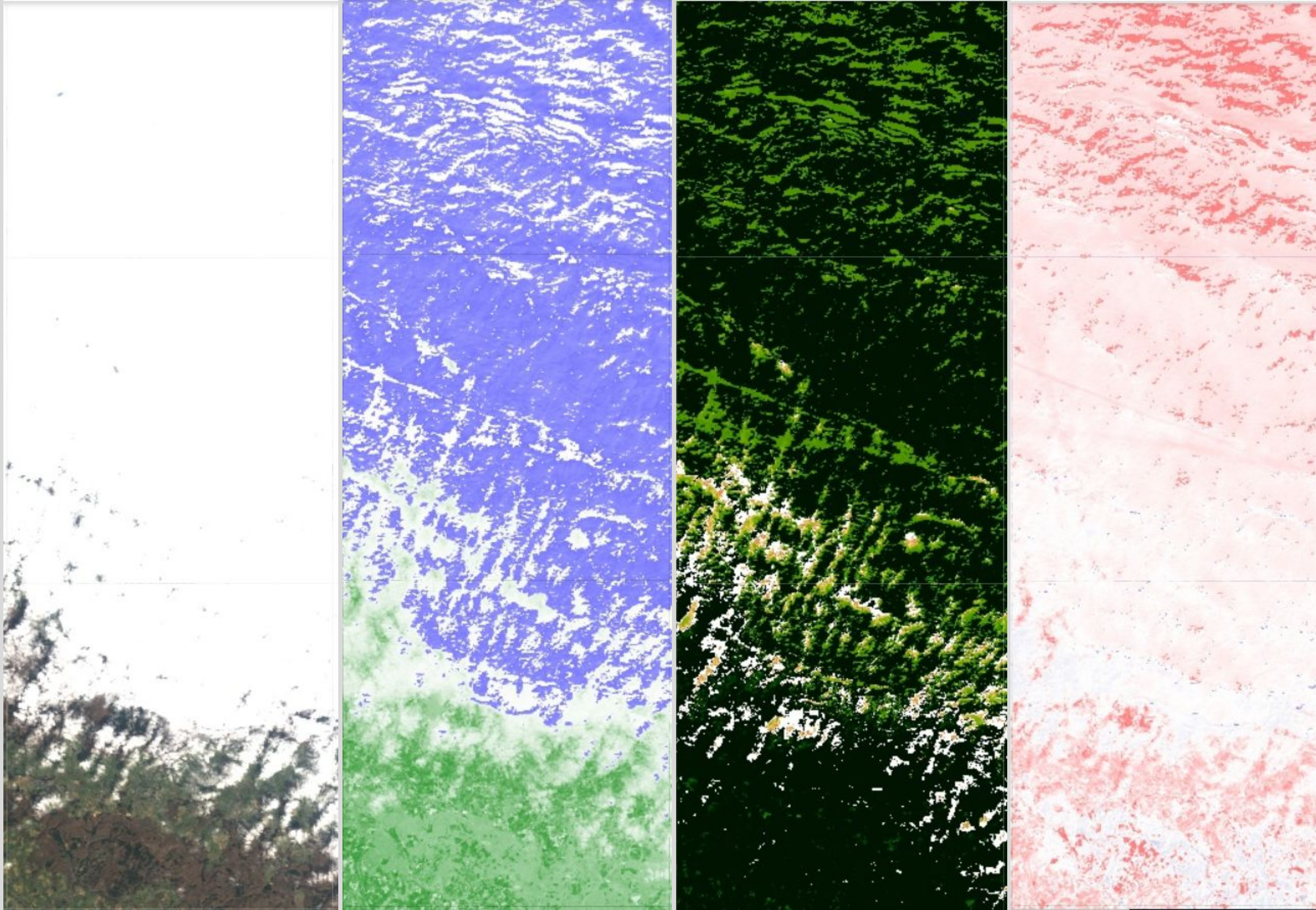
Right
Mar/Apr 2005
End of the dry season

Legend
Blue more water stress
Red less water stress

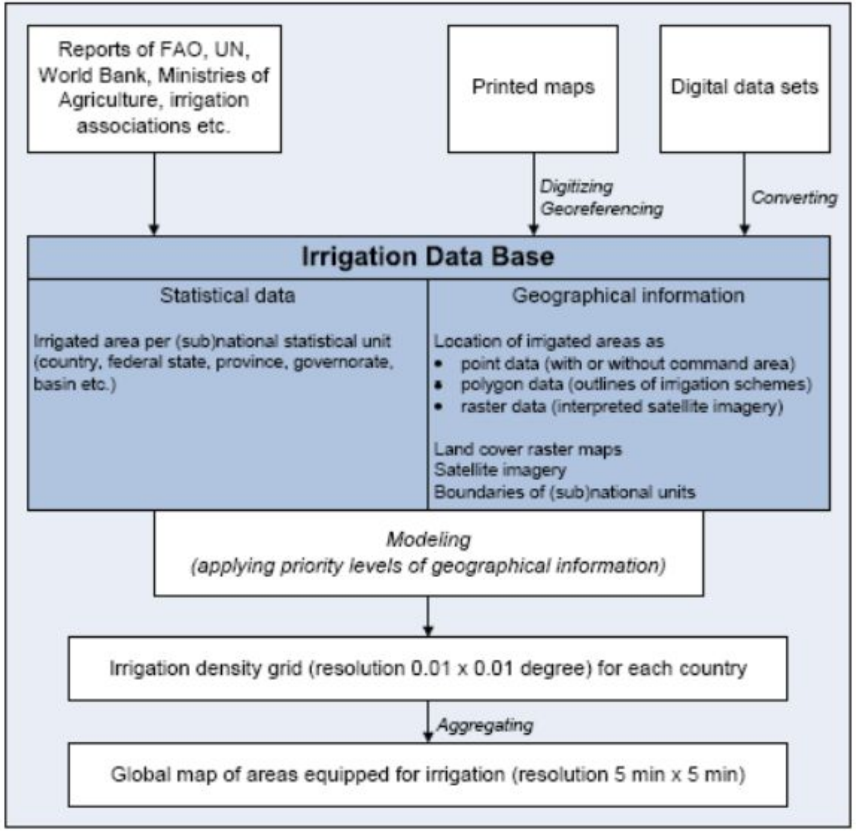


VISUAL RGB
NDVI
EVI
NDMI

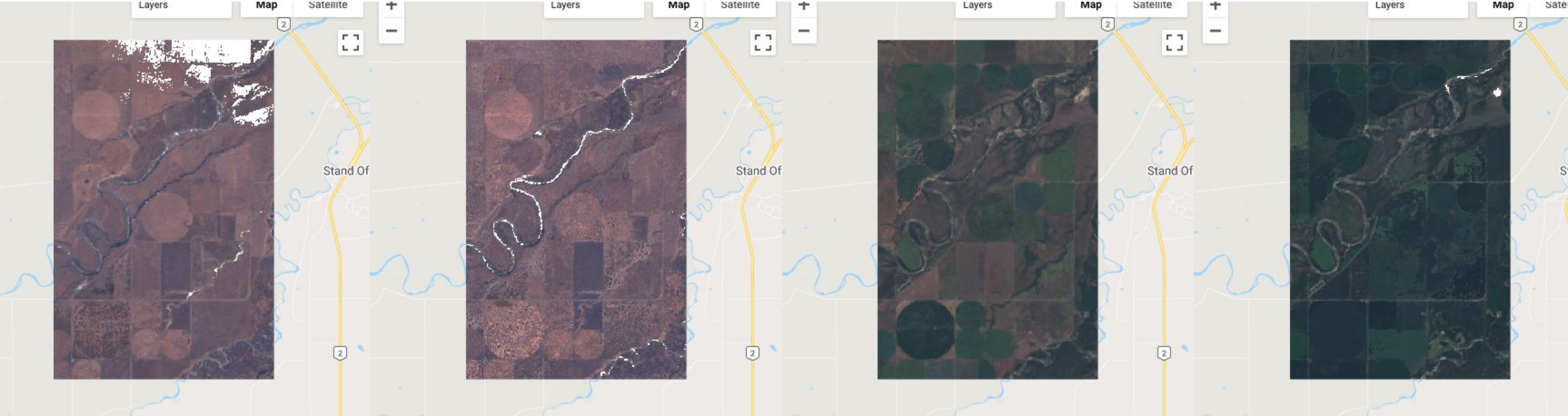
All from Mar 2005
The problem of sensor data from a cloudy day



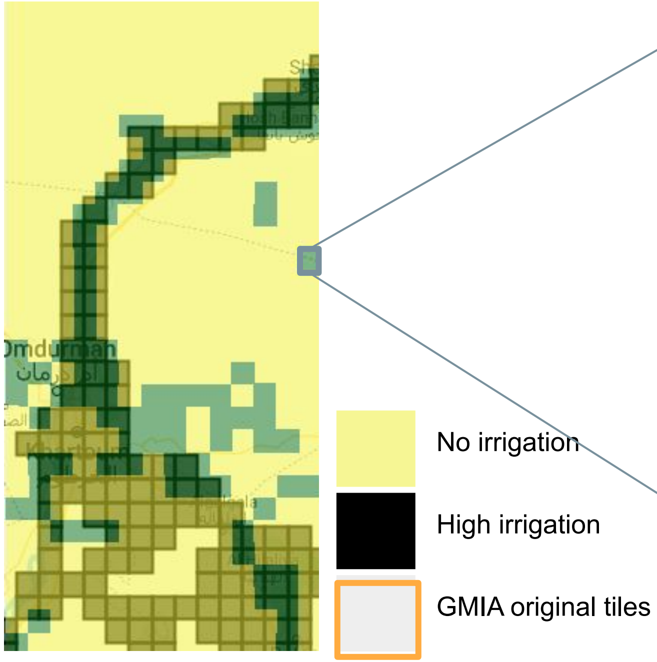
Datasets: The Global Map of Irrigated Areas (GMIA)



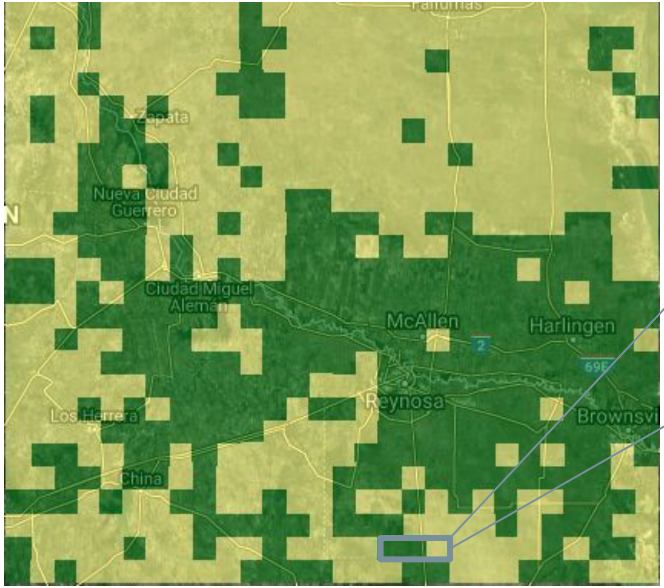
Tile Examples: Image Composite

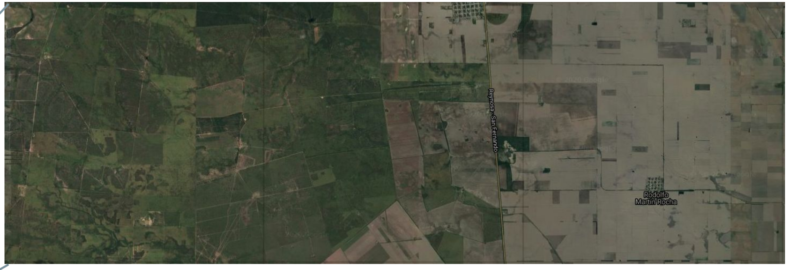


Inference - Sudan



Inference - Rio Grande Basin

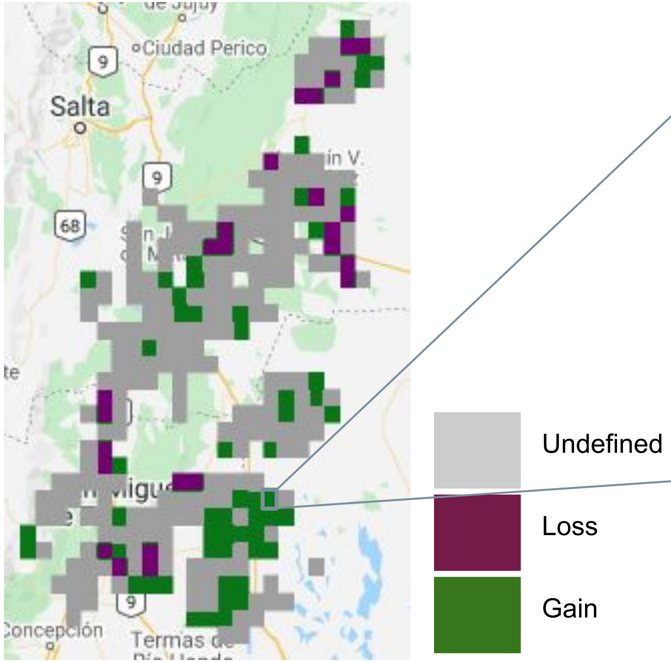




No irrigation

High irrigation

Irrigation Gain / Loss - Chaco



2005
Irrigation < 1%

2019
High irrigation

Tile: 443150
Chaco