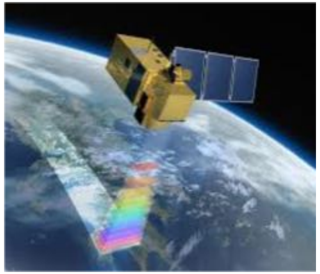
Earth Observation Big Data & AI for Land Cover Mapping & Change Detection

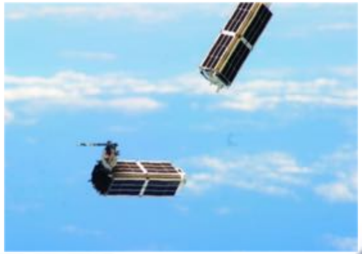
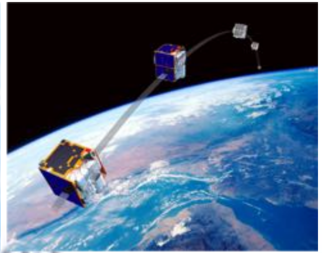
UN: Only 11 Years Left to Prevent Irreversble Damage from Climate Change

Earth Observation Big Data

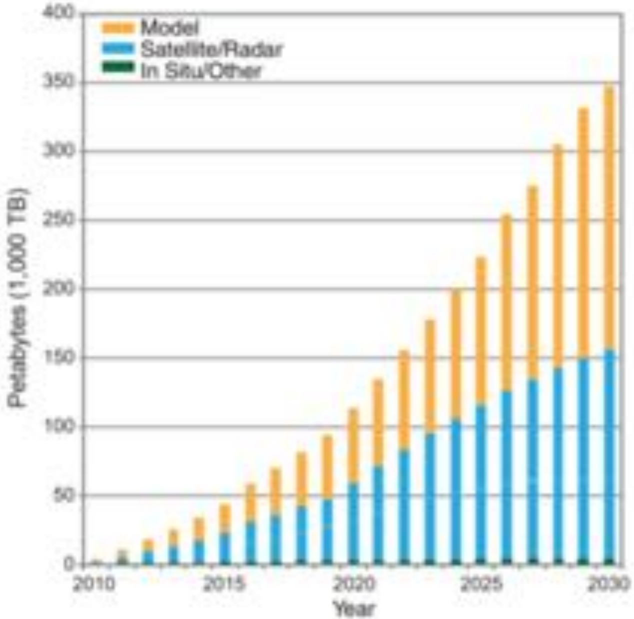
Where do we stand on Earth Observation?

- Thanks to the fast growth of satellite technology we are moving forward into a new era of Earth Observation (EO).
- Both National/International space agencies and innovative companies are supporting various EO programs acquiring huge amounts of data every day

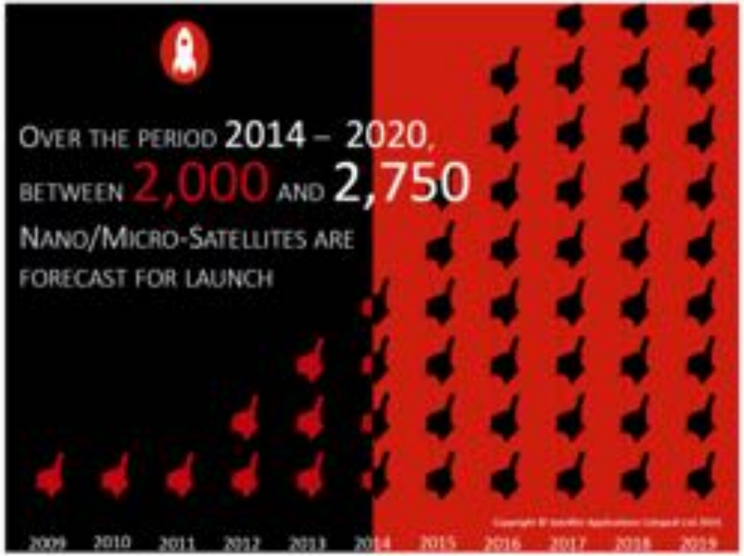




Earth Observation Big Data



Jonathan T. Overpeck et al., Science 2011

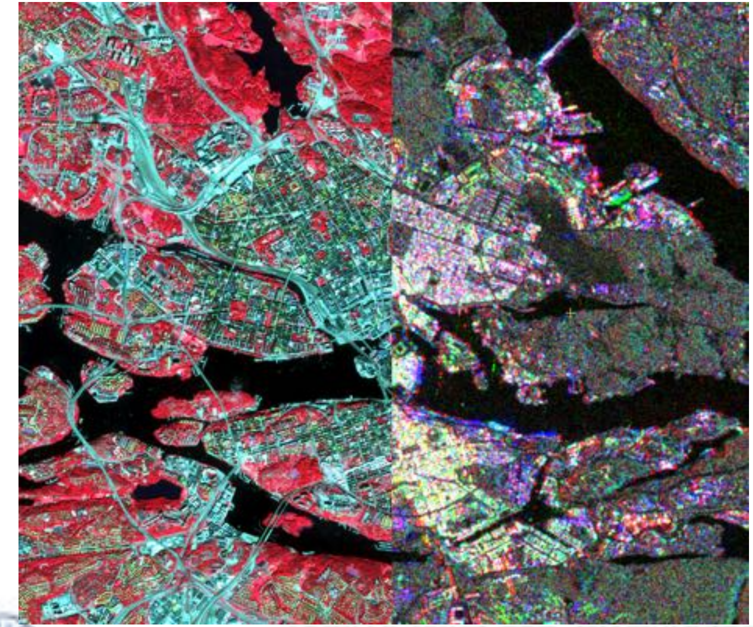


KTH MIST - MIniature STudent satellite

https://mistsatellite.space/

ESA Copernicus Program

- ESA: Sentinel-1 & -2 satellites with global coverage of multispectral images every 5 days & SAR images every 6 days
- Freely available imagery for both scientific and commercial use

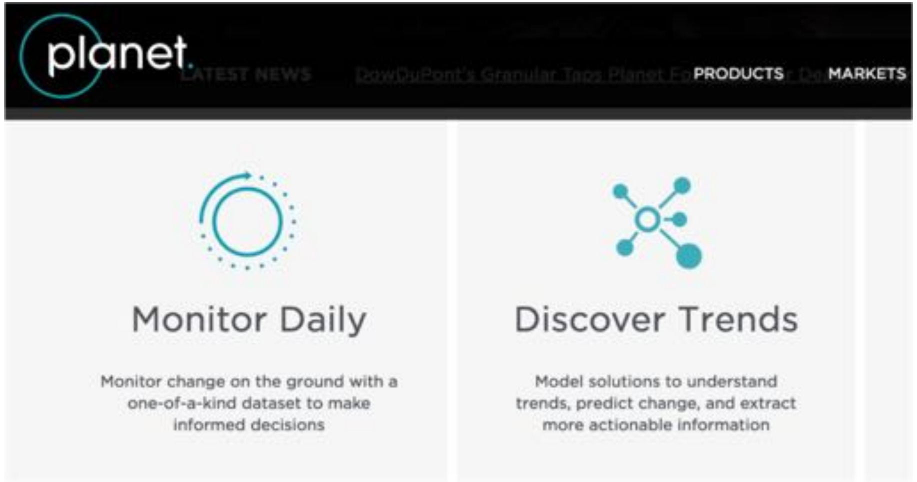


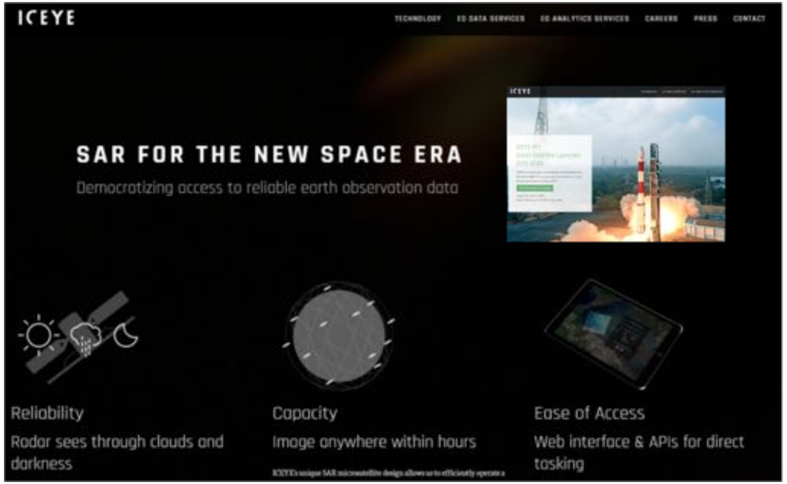
Earth Observation Big Data: New Trends

Planet: more than 175 Doves, 13 SkySats and 5 RapidEye satellites that acquire multispectral imagery of the entire Earth's landmass daily

Capella Space is deploying a SAR CubeSat satellite constellation that will provide hourly imagery with a global coverage

ICEYE is launching a constellation of 18 SAR satellites by the end of 2020





Earth Observation Big Data: Opportunities & Challenges

- **Opportunities**
  - Near-real time monitoring of phenomena affecting built and natural environment
  - Dense time series for analysis of global environmental changes
  - New possibility to deploy operational and reliable services

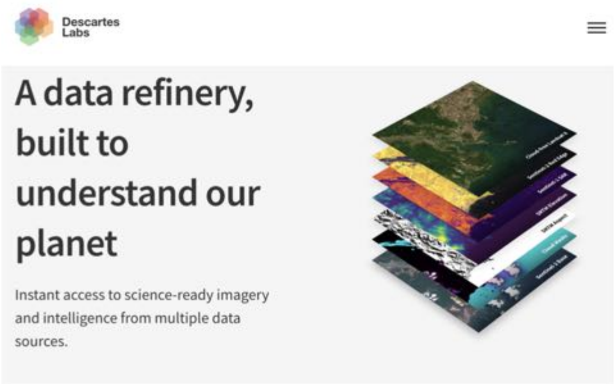
- **Challenges**
  - Deploy innovative computing infrastructure to handle, store and process the data
  - Develop new methods and algorithms to extract valuable information
  - Integrate the analysis of the EO imagery with other geospatial big data (i.e. social media, ground sensors, crowdsourced data)

EO Cloud Processing Platform

Several EO CPP are under development with contributions from open source communities (Open Data Cube), space agencies (ESA Thematic Exploitation Platform, DIAS) and private companies (Google Earth Engine, AWS, Sentinel-Hub, Descartes lab)

"Often it turns out to be more efficient to move the questions than to move the data." The Fourth Paradigm – Tony Hey et al.

OPEN DATA CUBE

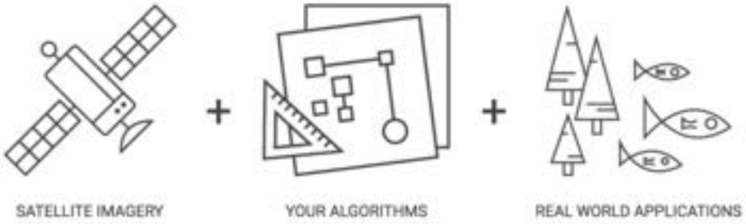


Google Earth Engine platform

Google Earth Engine (GEE) is a computing platform recently released by Google "for petabyte-scale scientific analysis and visualization of geospatial datasets":

- GEE enables researchers to **access geospatial information** and satellite imagery, for global and large scale remote sensing applications (**over than 20 petabytes** of geospatial data)

- GEE can be used to **perform geospatial analysis**, exploiting a dedicated HPC infrastructure, also running **user-developed software** through the **GEE API**



What Is Earth Engine?

"Big Data" analysis and visualization platform

- Inherently parallel system
- Designed for scientists, not software engineers
- Goals: make it easy, **enable non-traditional users**

Focused on society's biggest challenges

- Deforestation
- Drought
- Disaster
- Disease
- Climate Change
- Conflict
- Global Food Security
- Sustainability

20PB Public Data Catalog

Satellite Imagery:

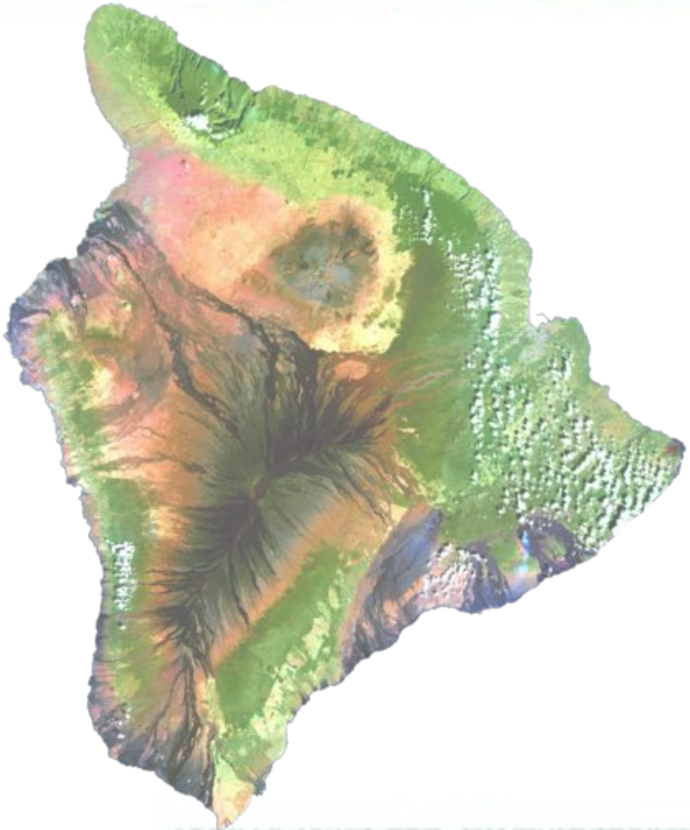
- Landsat 4-8 7 bands, 30m
- MODIS 250m Daily Global
- Sentinel-1 10m SAR
- Sentinel-2 12 bands, 10/20/60m
- (both TOA and BOA imagery)
- Alos data collection

Geophysical datasets:

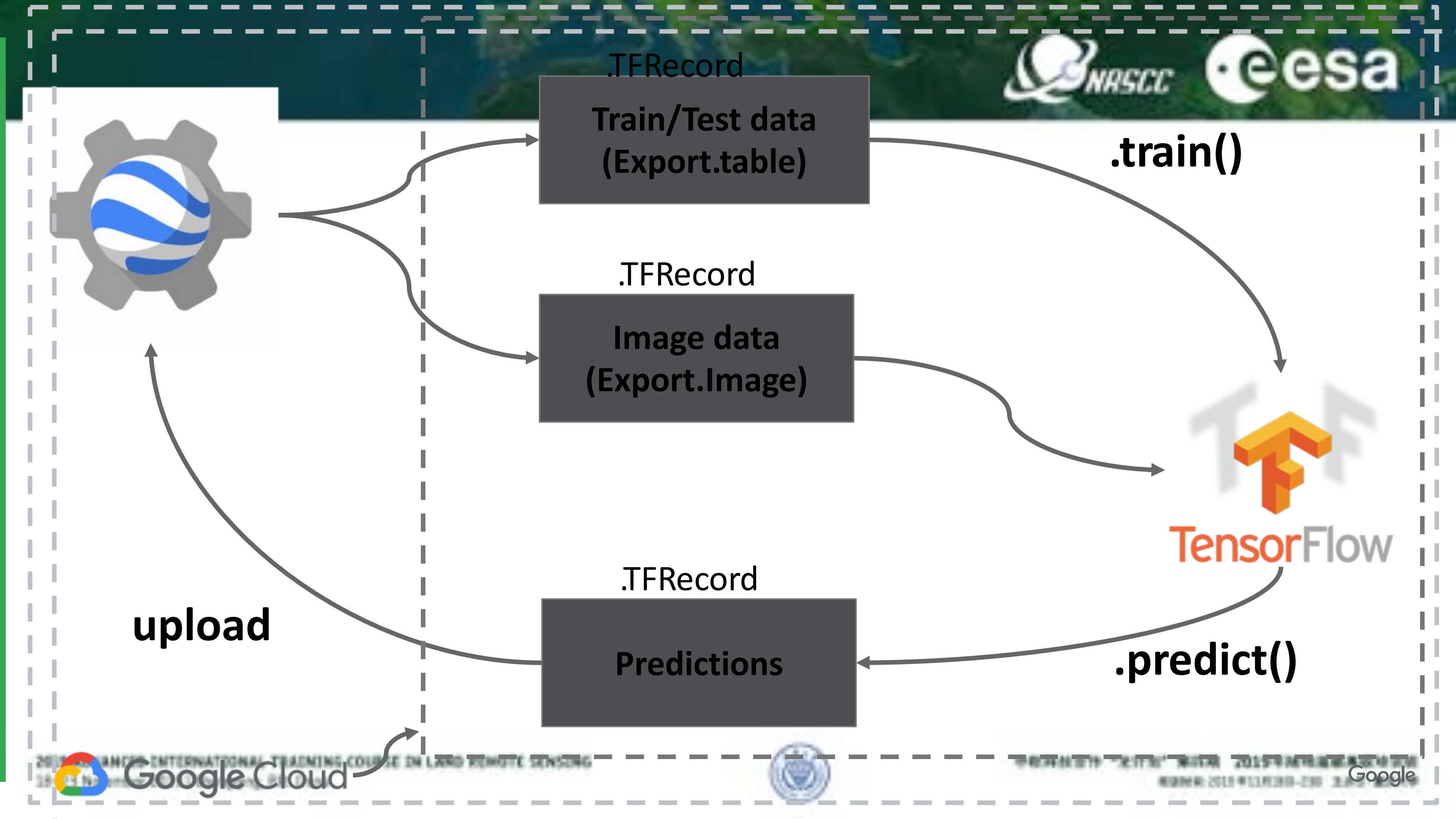
- Digital Elevation (SRTM, ALOS)
- Land Cover (USGS)
- Surface Temperature, etc.

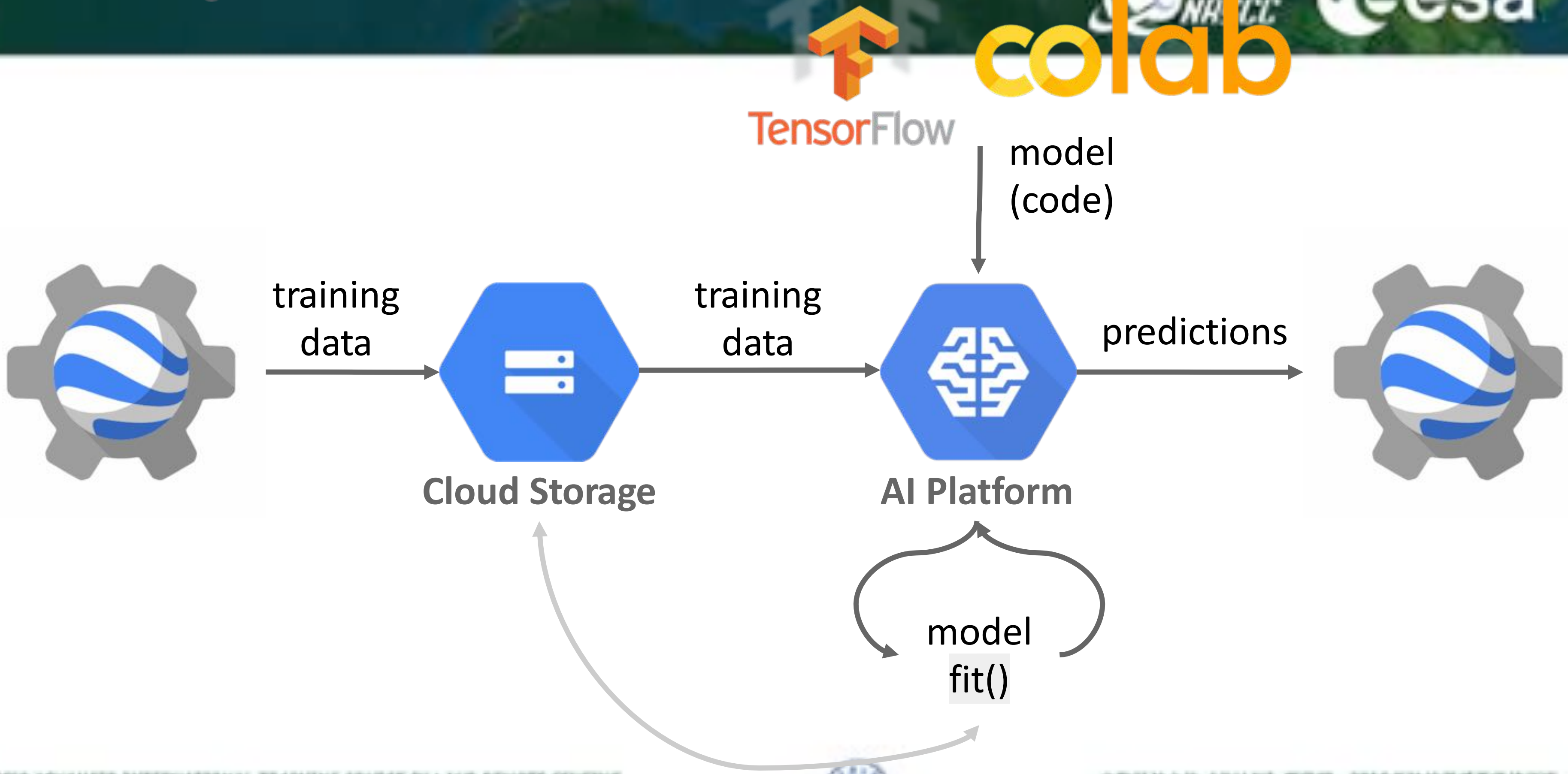
Weather Forecasts, Climate Models

- +300 more analysis ready datasets



's TensorFlow


.TFRecord
Train/Test data
(Export.table)
.train()
.TFRecord
Image data
(Export.Image)
TensorFlow
.TFRecord
Predictions
.predict()
upload
Google Cloud


TensorFlow
colab
model
(code)
training
data
Cloud Storage
training
data
AI Platform
predictions
model
fit()

Machine learning: Recipe Learned from Data

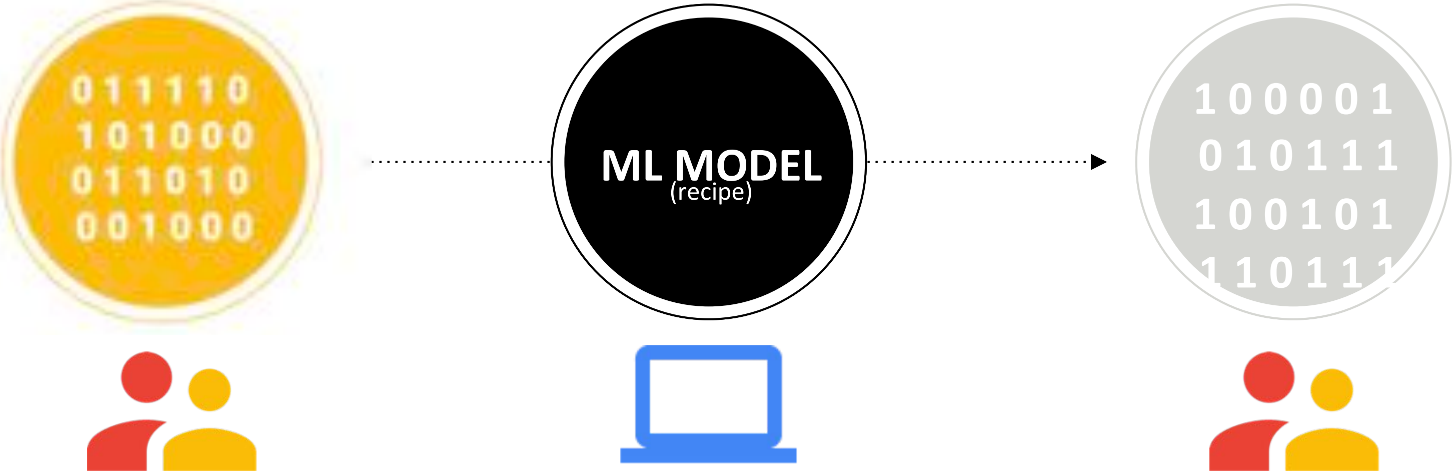
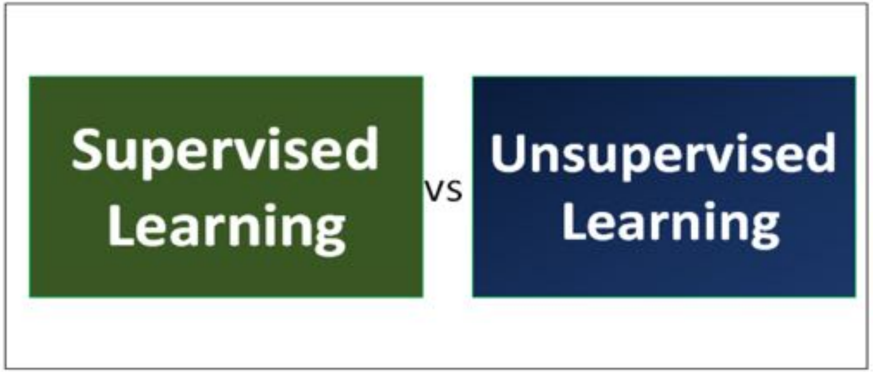
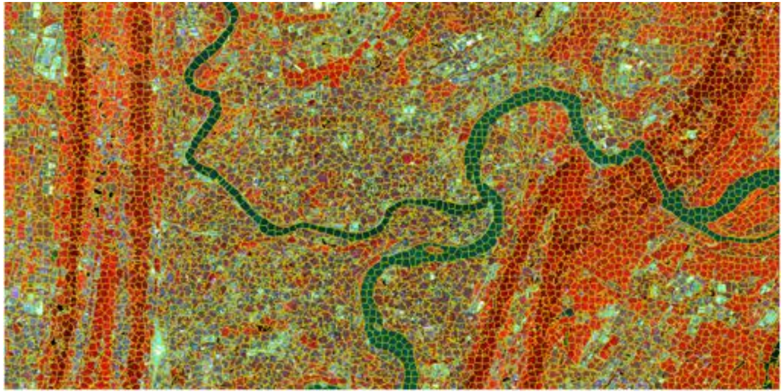


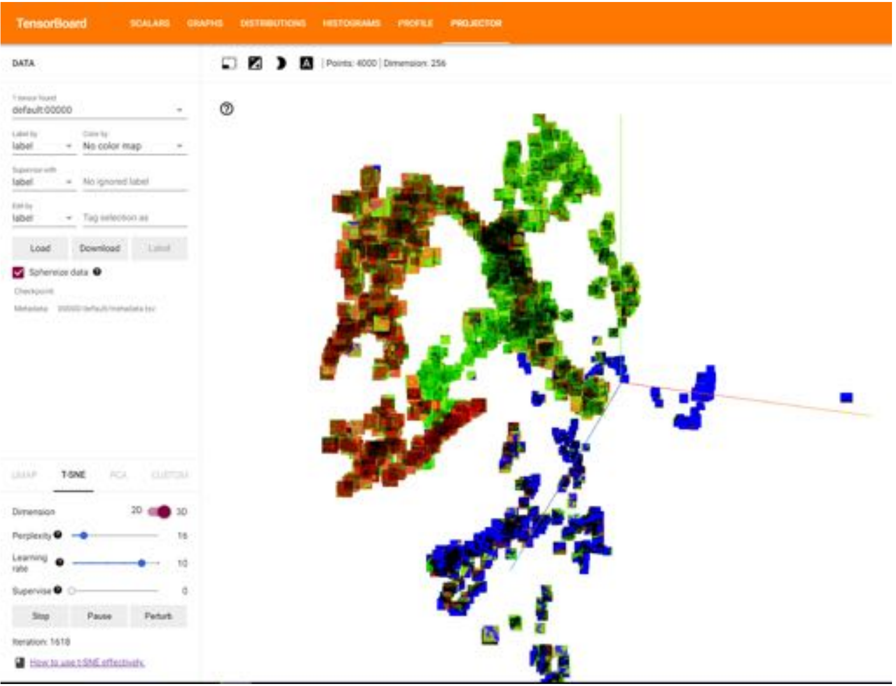
Image Classification Methods



Pixel-based and Object-based Classification



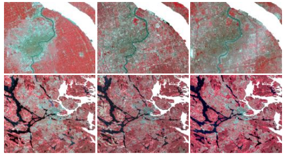
Land Cover Classification Algorithms

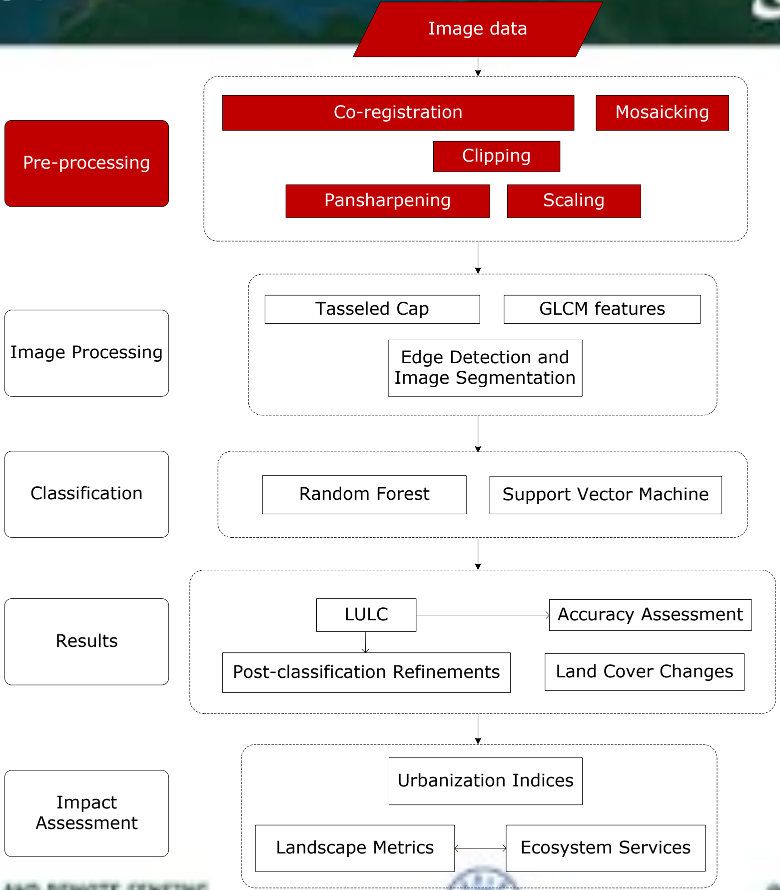


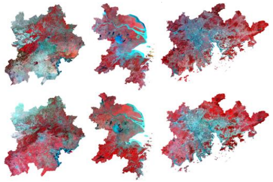
Algorithms

- Support Vector Machine
- Decision Tree
- Neural Network

Methodology





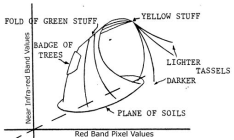


Methodology

Image data

Pre-processing
- Co-registration
- Mosaicking
- Clipping
- Pansharpening
- Scaling

Image Processing
- Tasseled Cap
- GLCM features
- Edge Detection and Image Segmentation



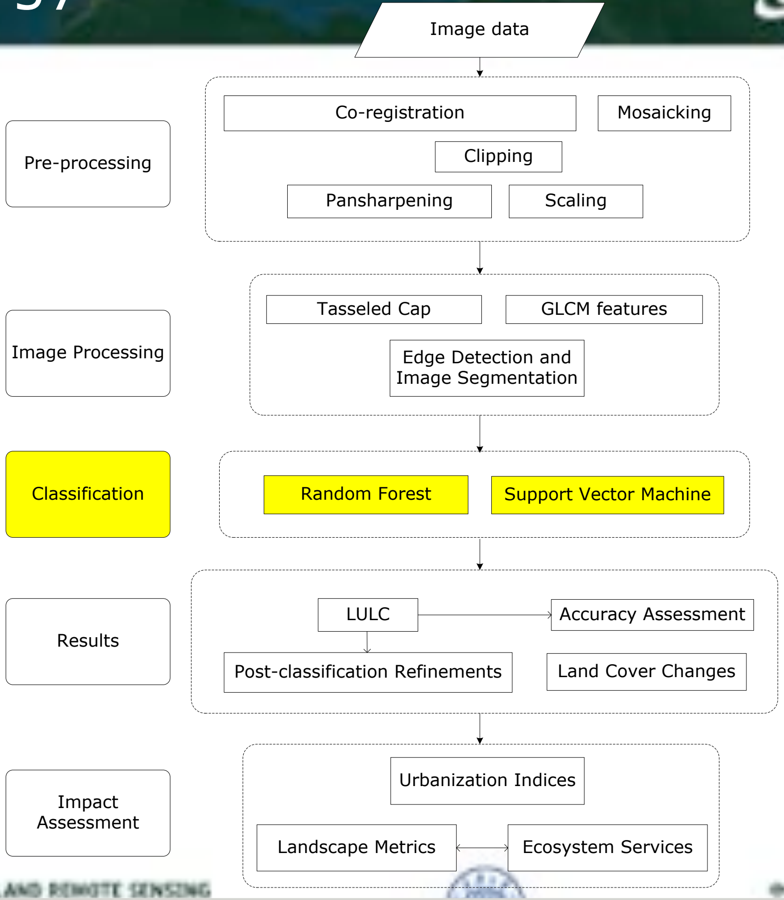
0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

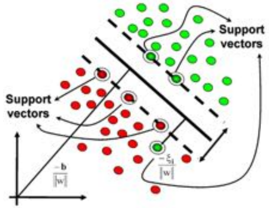
Classification
- Random Forest
- Support Vector Machine

Results
- LULC
- Accuracy Assessment
- Post-classification Refinements
- Land Cover Changes

Impact Assessment
- Urbanization Indices
- Landscape Metrics
- Ecosystem Services

Methodology





Methodology

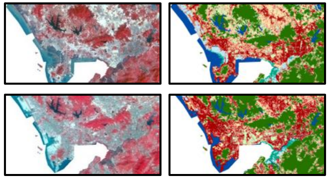
Image data

Pre-processing
- Co-registration
- Mosaicking
- Clipping
- Pansharpening
- Scaling

Image Processing
- Tasseled Cap
- GLCM features
- Edge Detection and Image Segmentation

Classification
- Random Forest
- Support Vector Machine

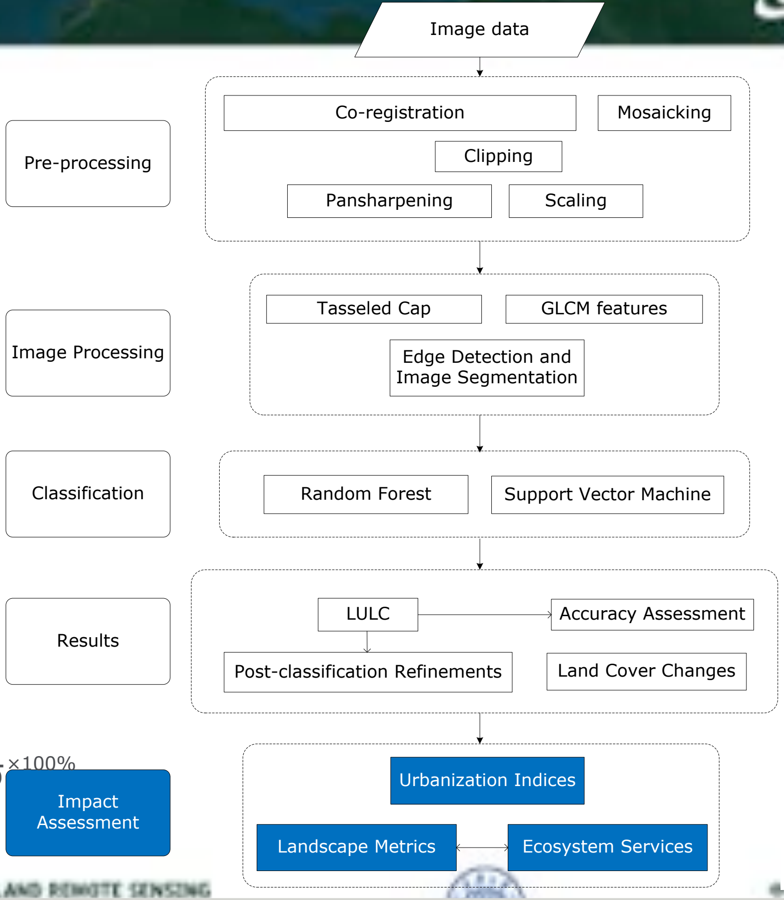
Results
- LULC
- Accuracy Assessment
- Post-classification Refinements
- Land Cover Changes



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	UA
C1	100	0	0	0	0	0	0	0	0	0	0	0	**100**
C2	0.6	91.8	3.6	0	1.5	0	0.6	0.1	1.8	0	0.1	0	**91.7**
C3	0	0	95.4	0	0	0	0	0	4.6	0	0	0	**95.4**
C4	0	0	0	79.9	0.3	0	1.2	14.1	0	0	0.4	4.1	**79.9**
C5	0	0	1.2	0.7	81	0	1.9	4	0.5	4.1	1	5.5	**81.1**
C6	0	0	0	0	0	72.4	0.7	12.8	13.5	0	0.6	0.1	**72.4**
C7	0	0	0	0	6.4	0.1	86.9	1.7	2.1	0	0.3	2.5	**87**
C8	0	0	0	39.3	0.8	2.1	10.7	31.8	4	5.2	5.8	0.4	**31.8**
C9	0	0	0	0	1.1	2.4	0	0	95.1	0	0	1.4	**95.1**
C10	0	0	1.8	0	12.1	0	5.1	1	0	80	0	0	**80**
C11	0	0	0	3.7	13.8	0.5	0.9	1.6	14.8	0.5	51.1	13.2	**51.1**
C12	0	0	0	0	13	0	0	0	6.6	0	0	80.4	**80.4**
PA	**99.4**	**100**	**93.8**	**65.3**	**66.1**	**93.8**	**81.2**	**47.4**	**68.7**	**89.4**	**79.7**	**77.9**	

Impact Assessment
- Urbanization Indices
- Landscape Metrics
- Ecosystem Services

Methodology



$$UGI = \frac{UGS_{t2} - UGS_{t1}}{(HDB_{t2} + LDB_{t2}) - (HDB_{t1} + LDB_{t1})} \times 100\%$$

$$UI = \frac{UL}{TL} \times 100\%$$

$$UX_r = \frac{UL_{t2} - UL_{t1}}{UL_{t1}} \times 100\%$$

What is deep learning?

"**Deep learning** is a particular kind of machine learning that achieves great power and flexibility by representing the world as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones."

Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning"

What is TensorFlow?

TensorFlow™ is an open source software library for high performance numerical computation. Its flexible architecture allows easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices. Originally developed by researchers and engineers from the Google Brain team within Google's AI organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains.

https://www.tensorflow.org/

What is Colab?

Colaboratory is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud.

With Colaboratory you can write and execute code, save and share your analyses, and access powerful computing resources, all for free from your browser. Colab notebooks are stored in Google Drive, and can be shared just as you would with Google Docs or Sheets.

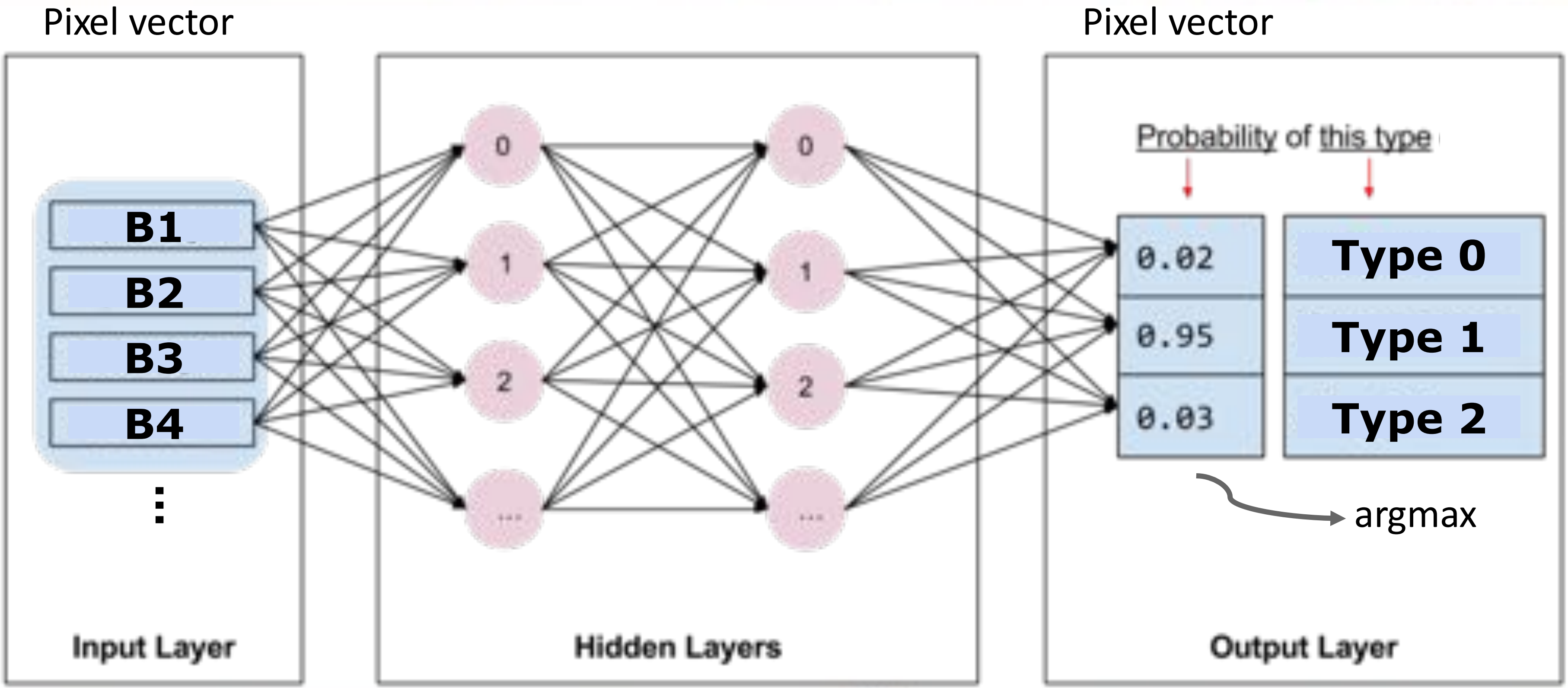
https://colab.sandbox.google.com/notebooks/welcome.ipynb

Artificial Neural Networks (ANNs)

Deep feedforward networks, also called **feedforward neural networks**, or **multilayer perceptrons** (MLPs), are the quintessential deep learning models.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning"

What is an ANN?

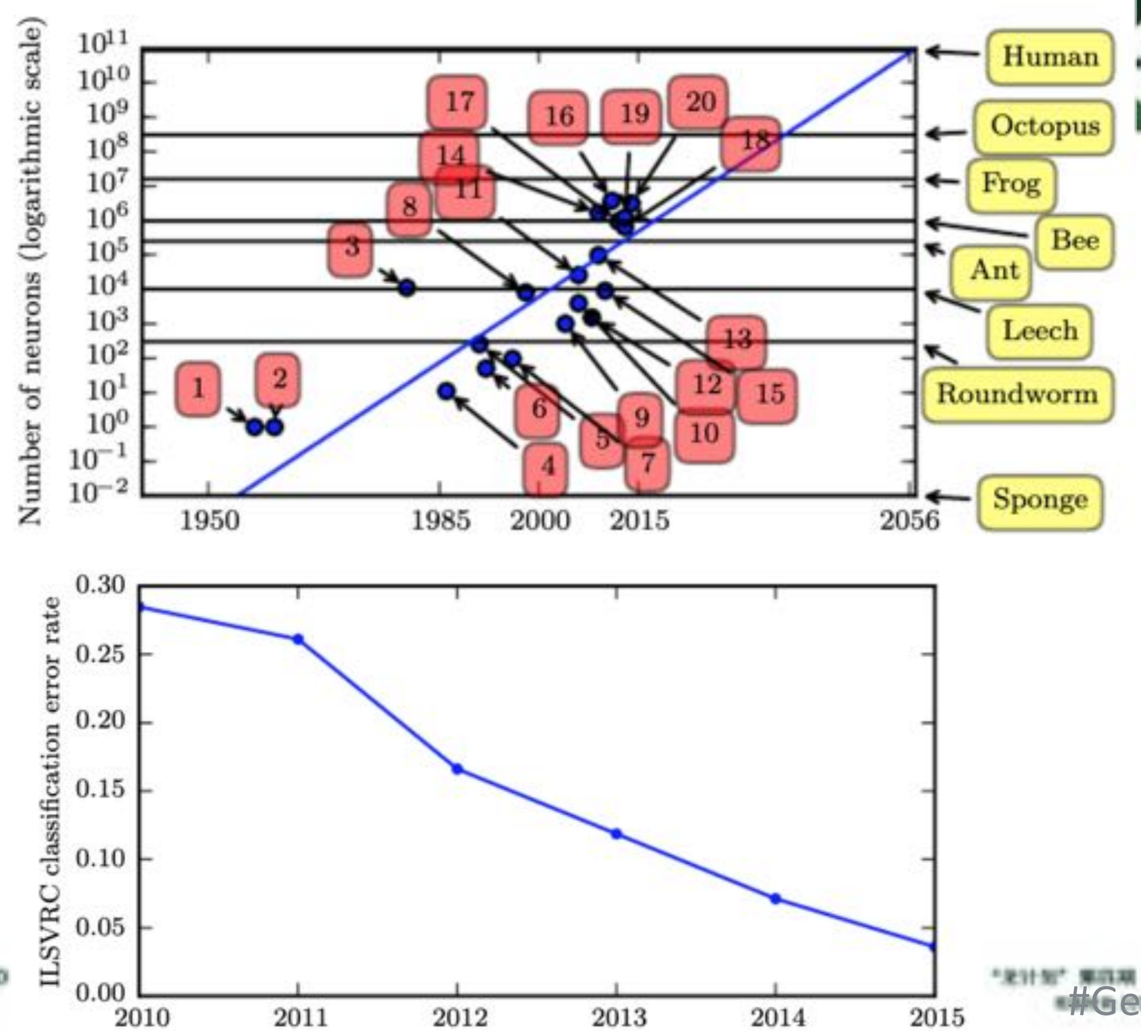


What is an ANN?

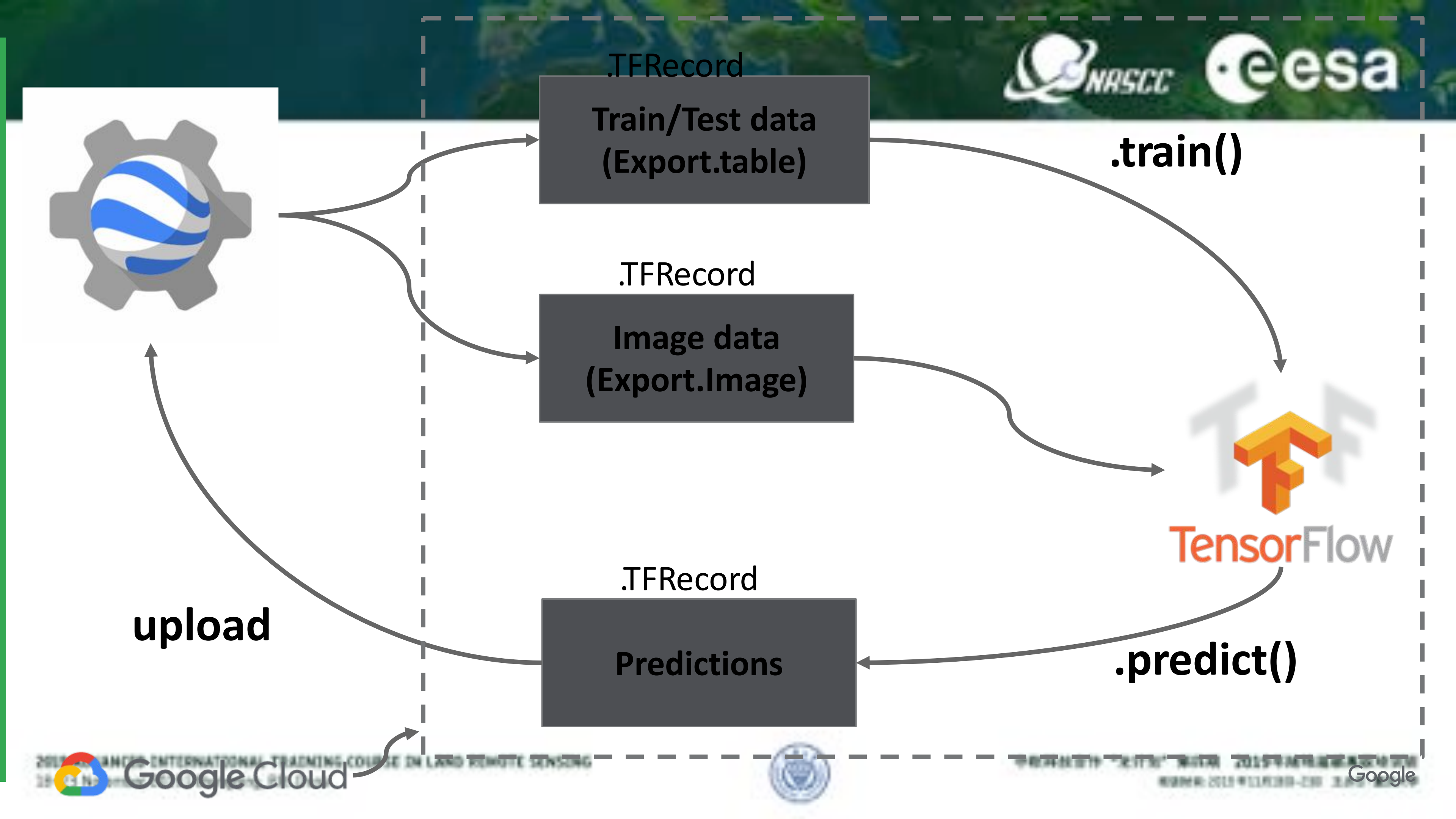
for a classifier, $y = f^*(\boldsymbol{x})$ maps an input $\boldsymbol{x}$ to a category y. A feedforward network defines a mapping $\boldsymbol{y} = f(\boldsymbol{x}; \boldsymbol{\theta})$ and learns the value of the parameters $\boldsymbol{\theta}$ that result in the best function approximation.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning"

Why ANNs?



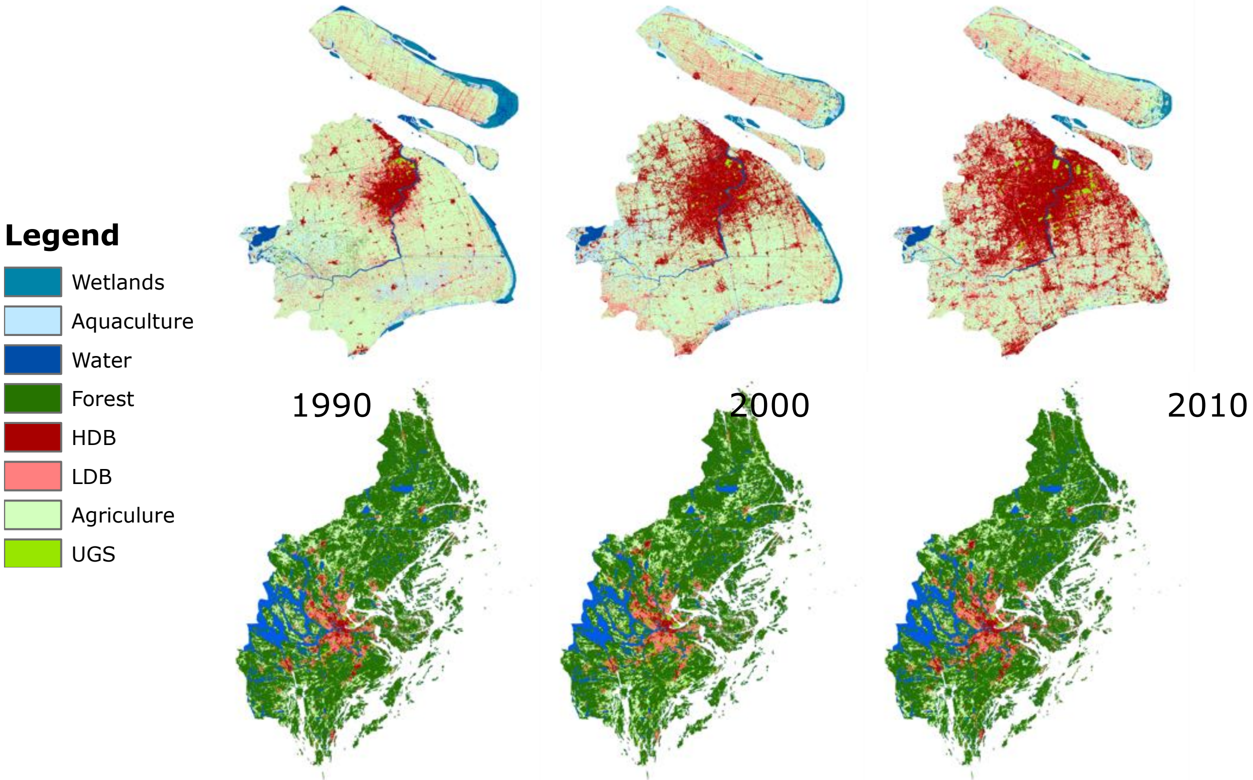
Goodfellow et al. 2016


.TFRecord
Train/Test data
(Export.table)
.train()
.TFRecord
Image data
(Export.Image)
TensorFlow
.TFRecord
Predictions
.predict()
upload
Google Cloud

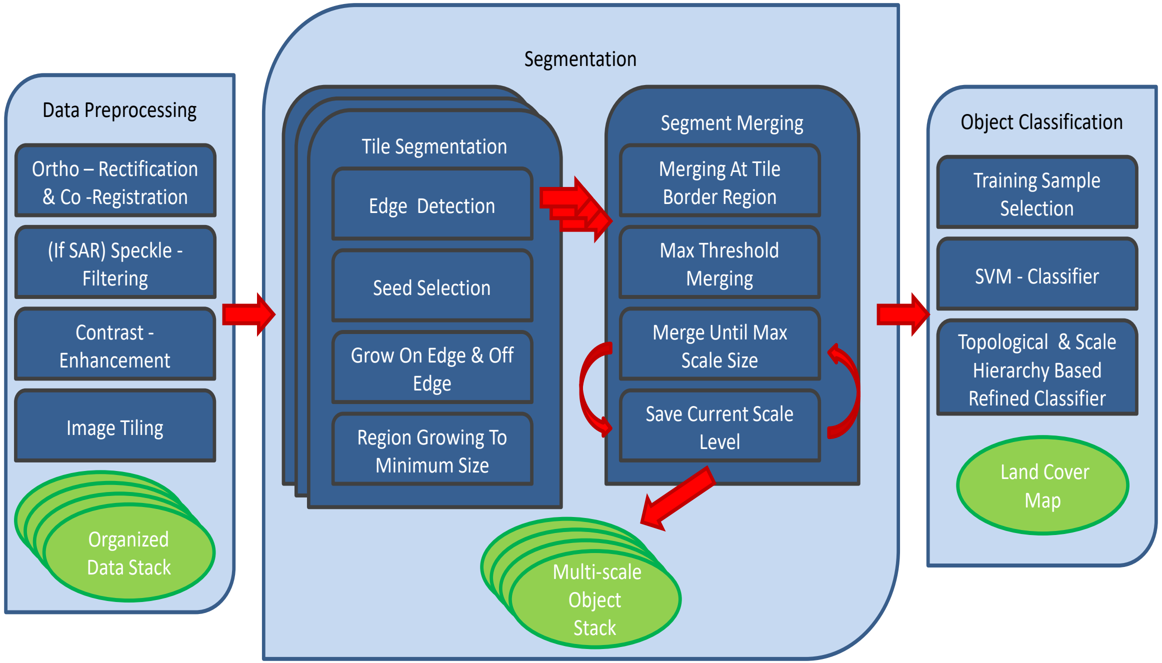
's

TensorFlow

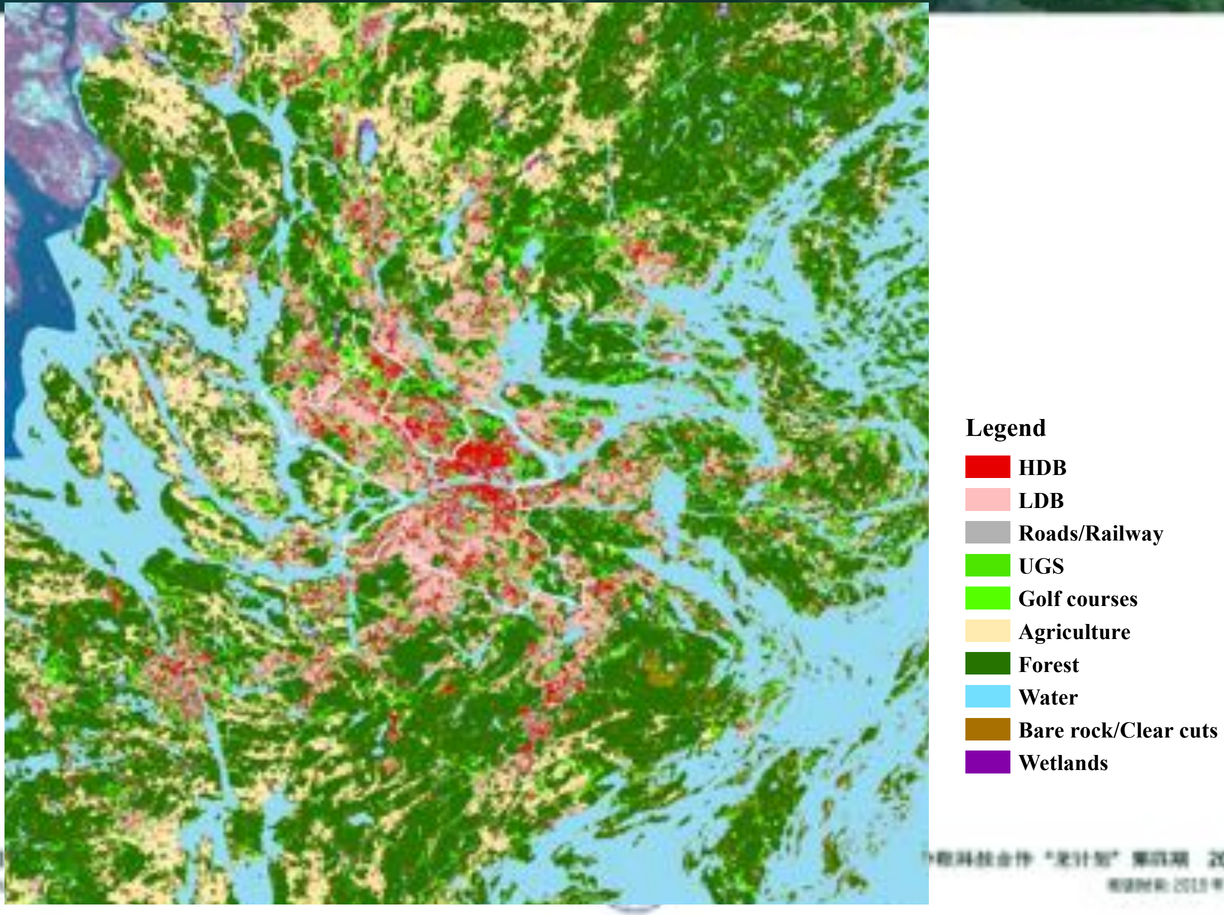
Stockholm/Shanghai



KTH-SEG: An Edge-Aware Region Growing and Merging Algorithm



Classification (subsection) of 2015 Sentinel-2 Data



Land Cover Change Detection

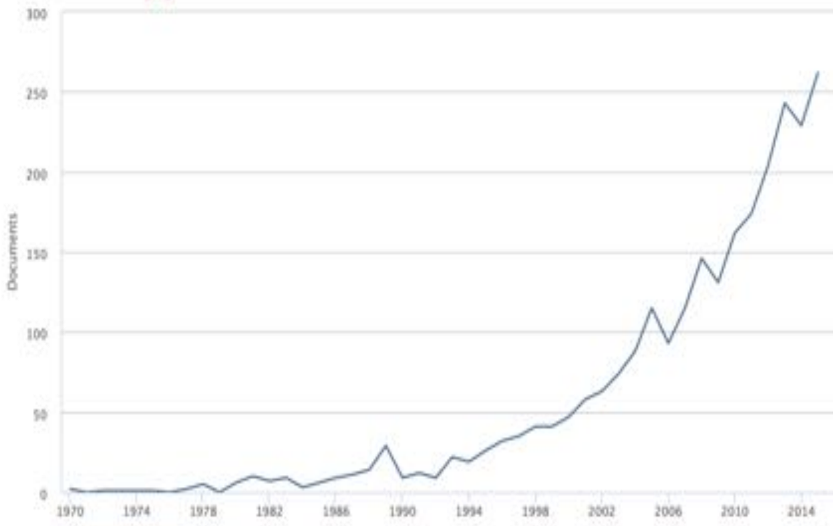
1 Multitemporal data for change detection: current trend

2 Change detection in VHR multispectral images

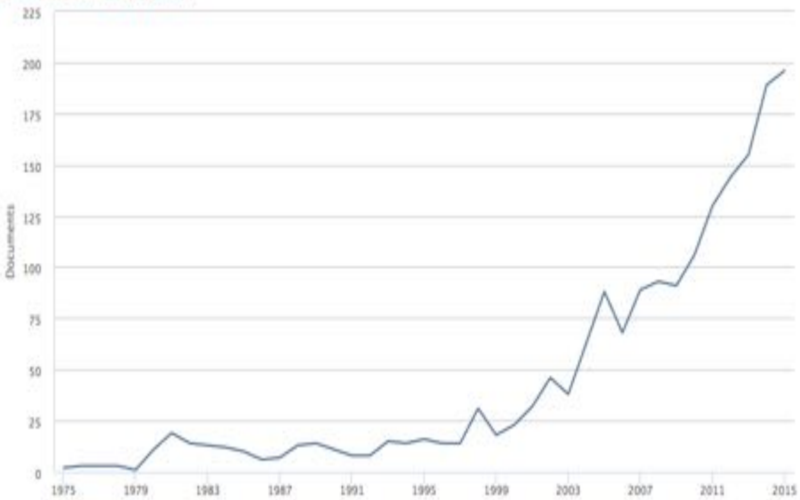
3 Change detection in VHR SAR images (Not included)

4 Change detection in multisensor images

Multitemporal data for change detection: current trend

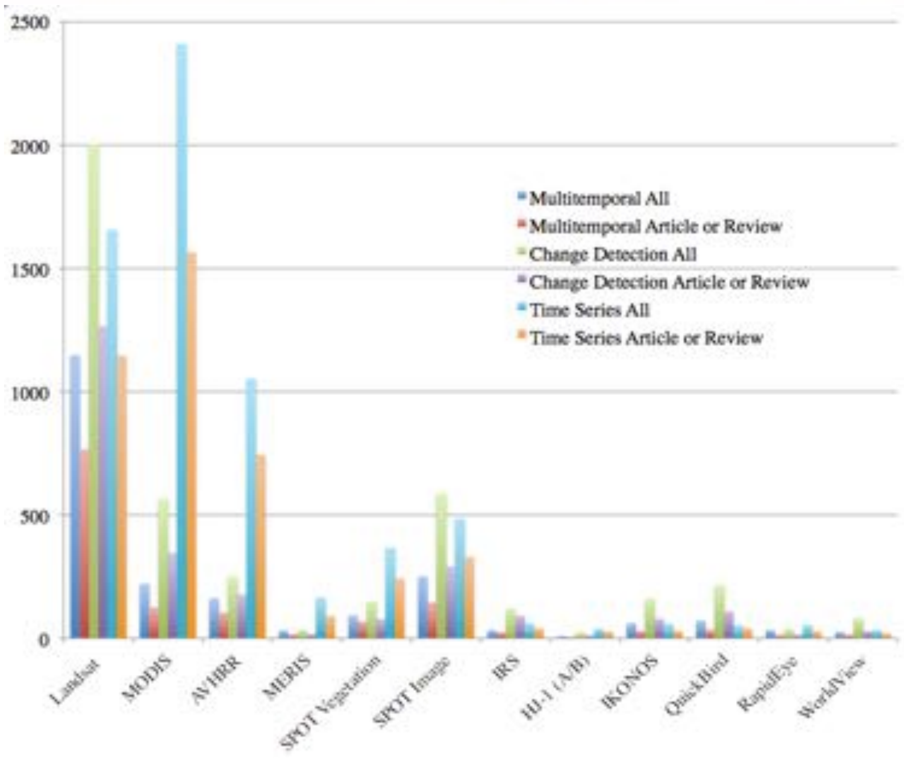


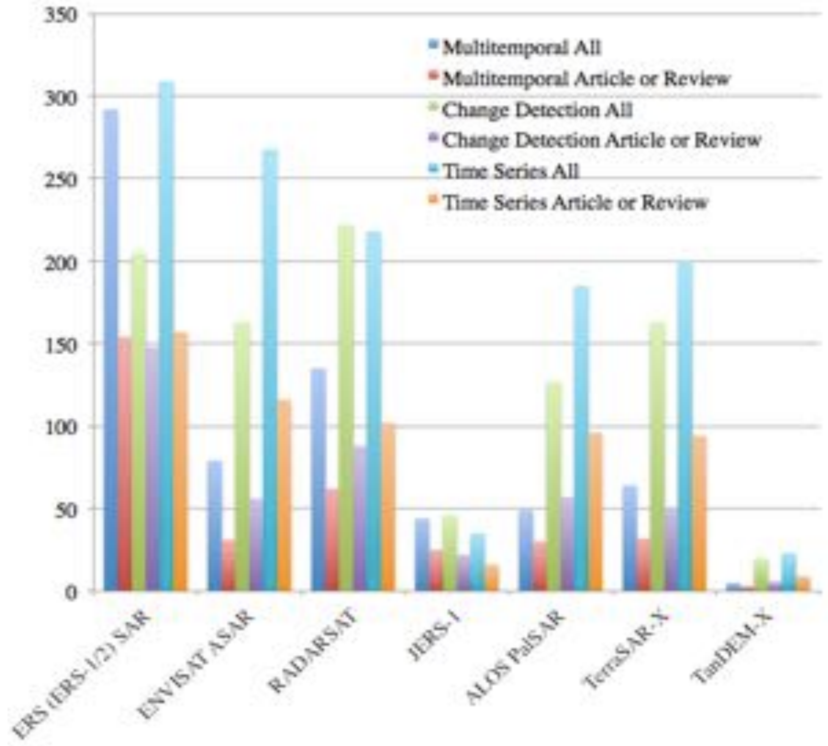
The number of articles (including review articles) per year on change detection derived from a Scopus search on July 15, 2016



Number of Publications per Year from Scopus Search for ‘Change Detection’ and ‘Landsat’

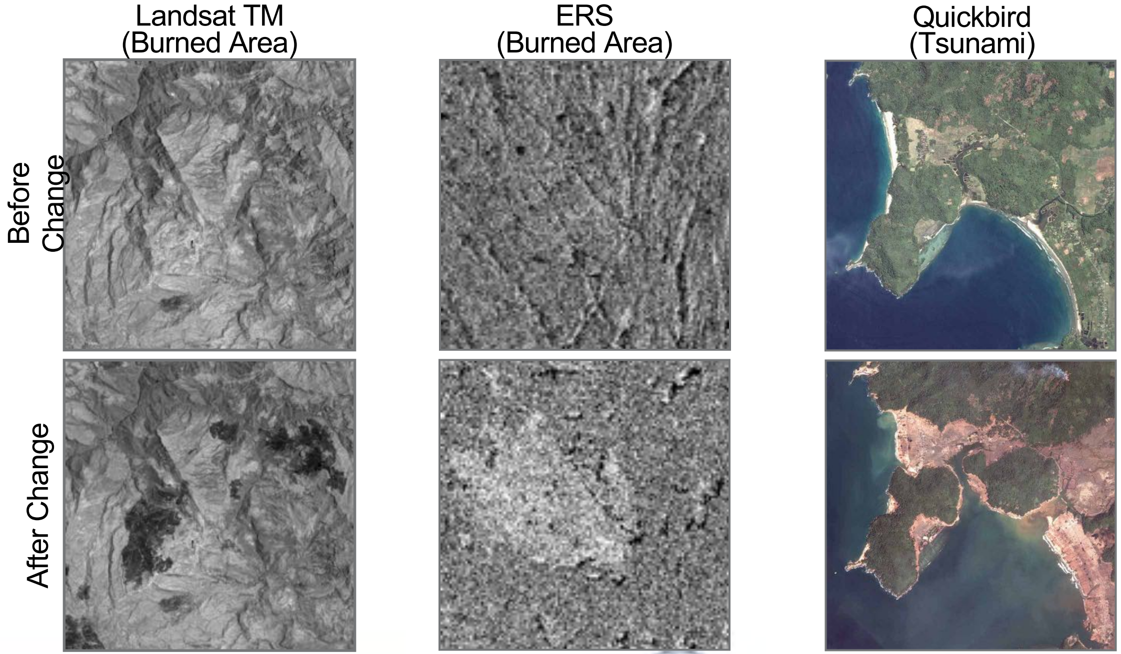
Multitemporal data for change detection: current trend



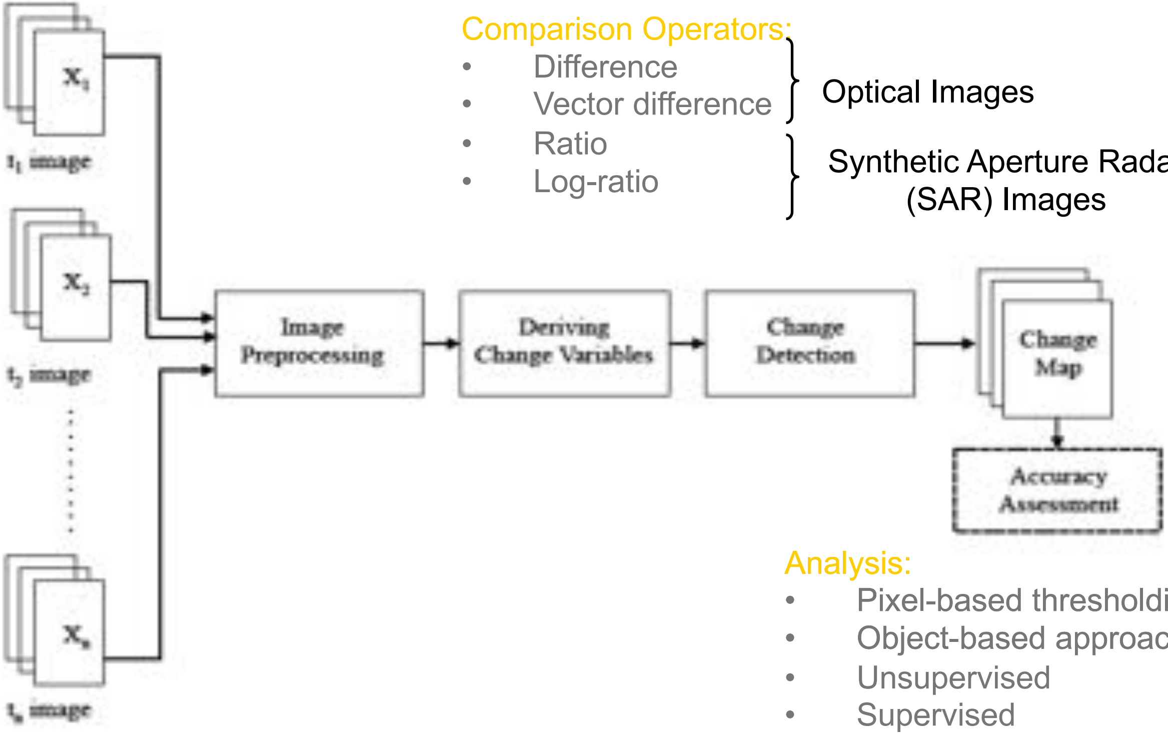


Change Detection: Traditional Definition

Change detection (CD): process that analyzes ***bi-temporal*** remote sensing images acquired on the same geographical area for identifying changes occurred between the considered acquisition dates.



Change Detection



Comparison Operators:

- Difference
- Vector difference
- Ratio
- Log-ratio

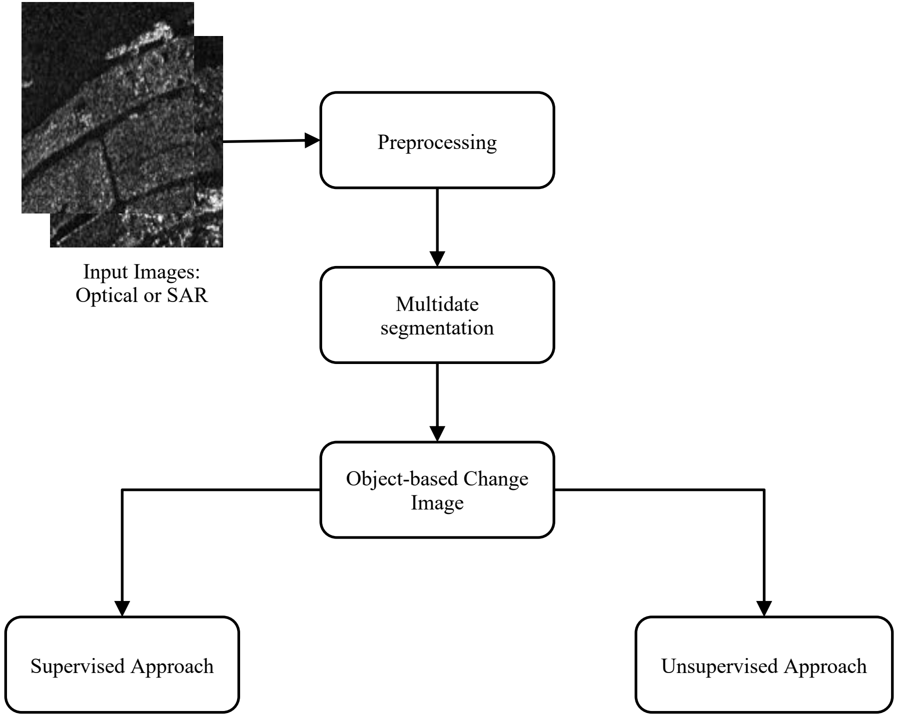
Difference, Vector difference: Optical Images

Ratio, Log-ratio: Synthetic Aperture Radar (SAR) Images

Analysis:

- Pixel-based thresholding
- Object-based approaches
- Unsupervised
- Supervised

Object-based Change Detection

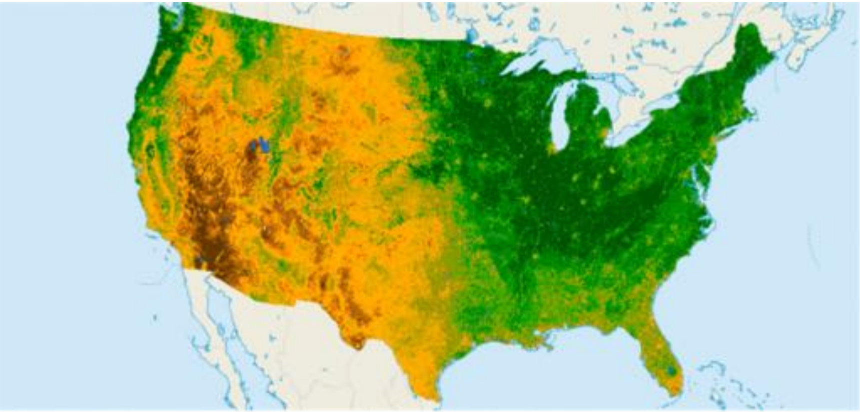


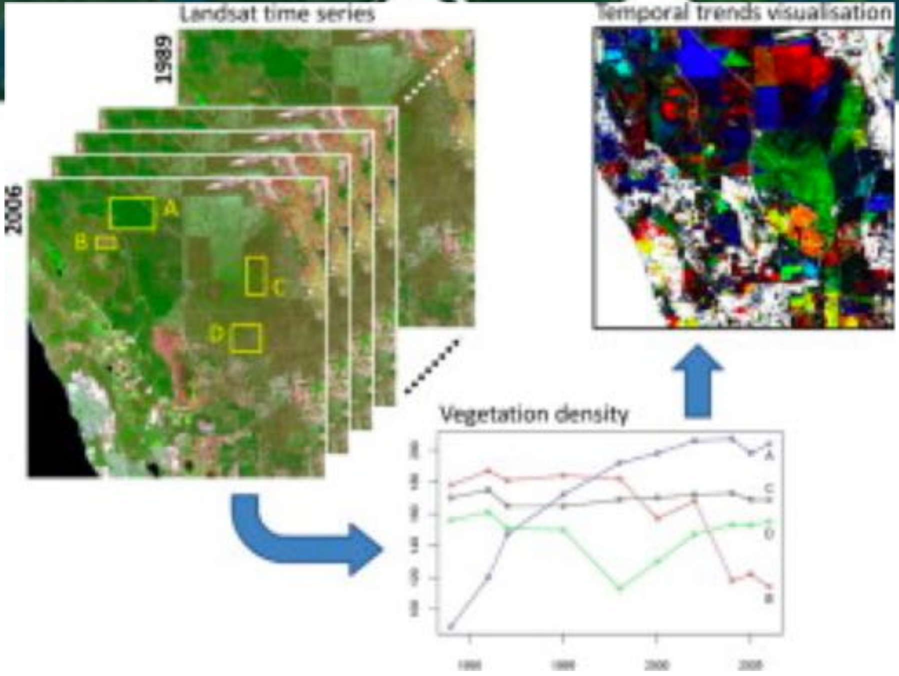
Time Series Analysis in GEE

We can analyzed different trends:

- Sentinel-1 VV, VH backscatter time series
- Sentinel-2 or Landsat-8 NDVI time series

The goal is to detect changes -> anomalies in image trends

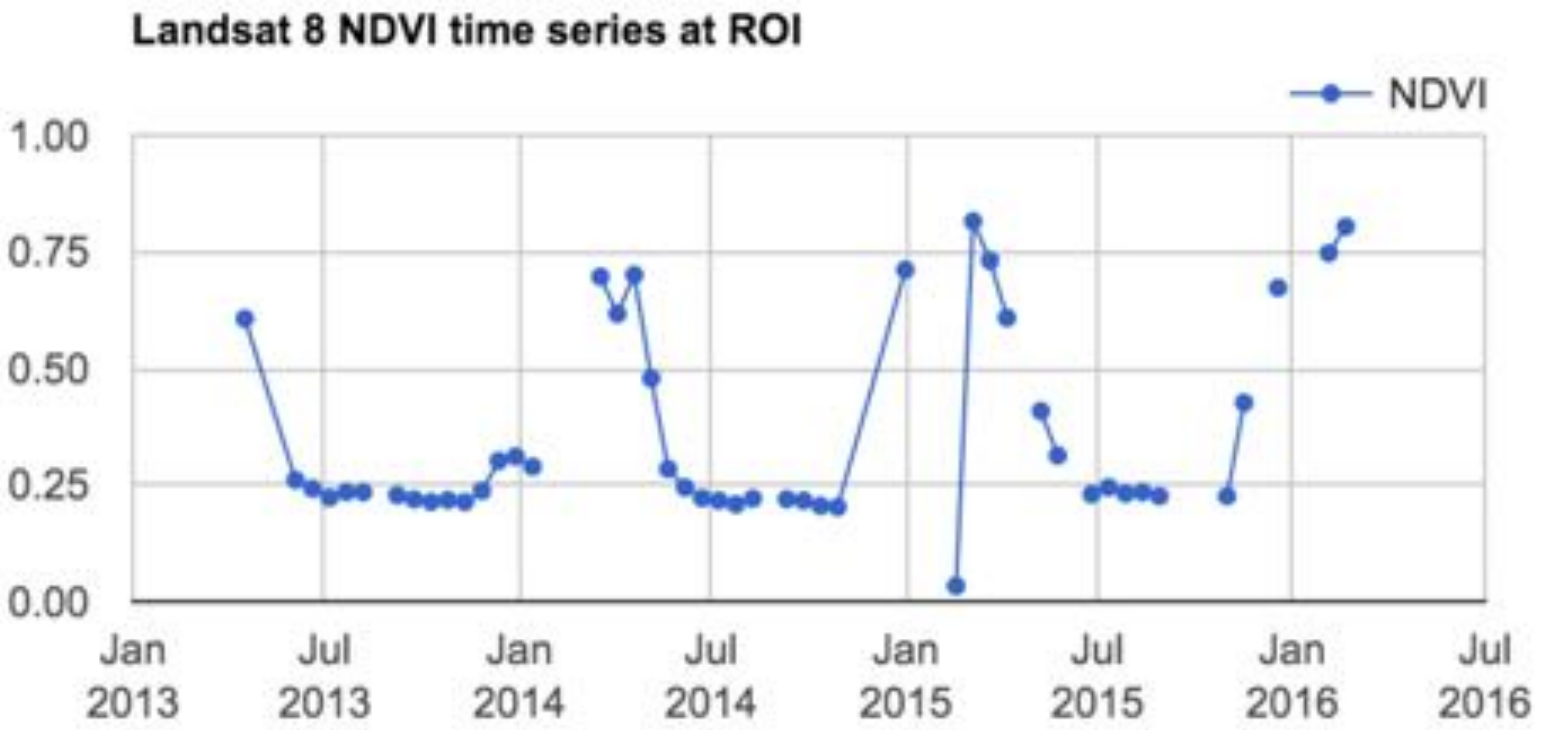




Earth Engine Time Series modelling examples:

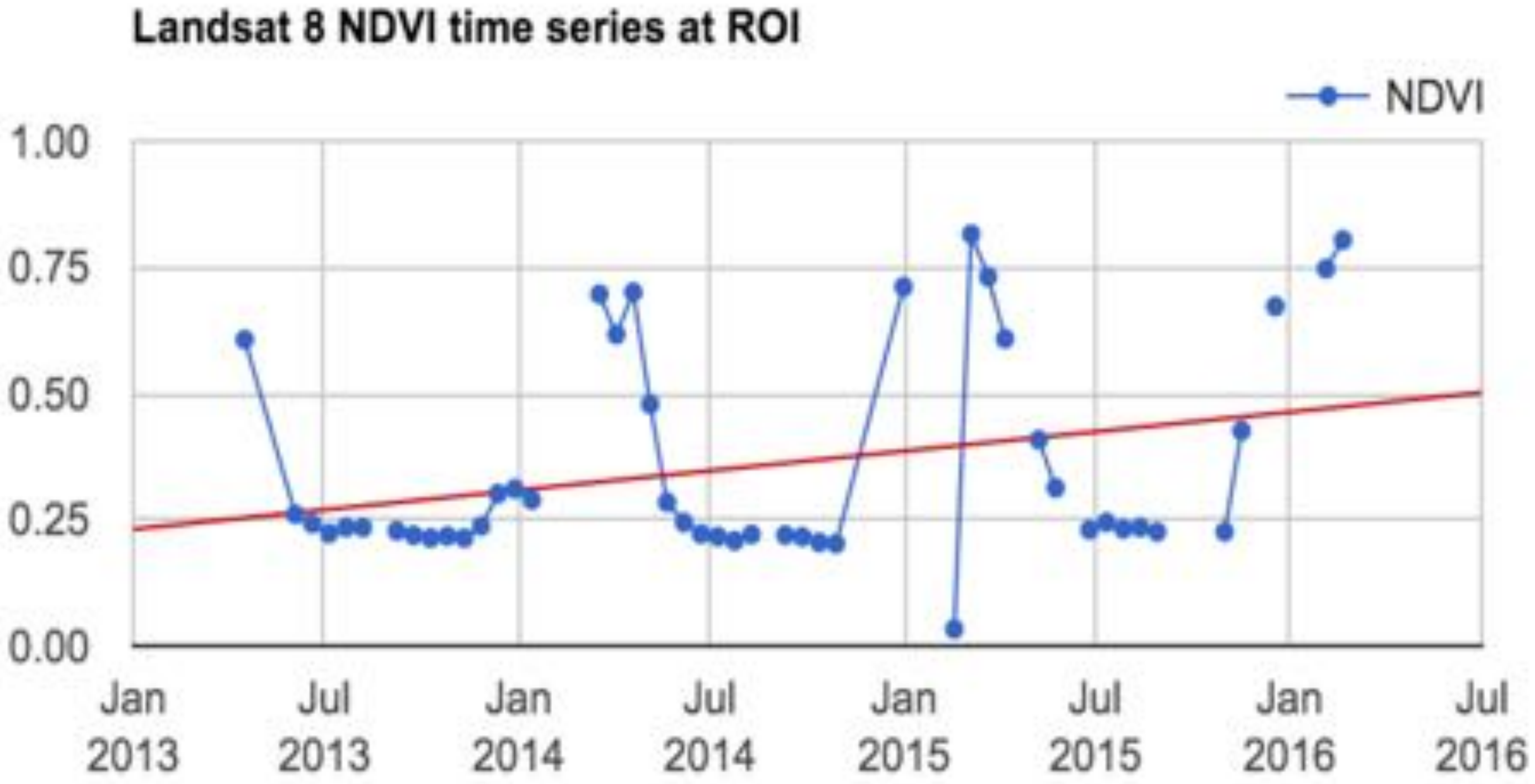
- Linear modeling
- Harmonic modeling

Model fitting: irregularly spaced points, missing data, noisy data

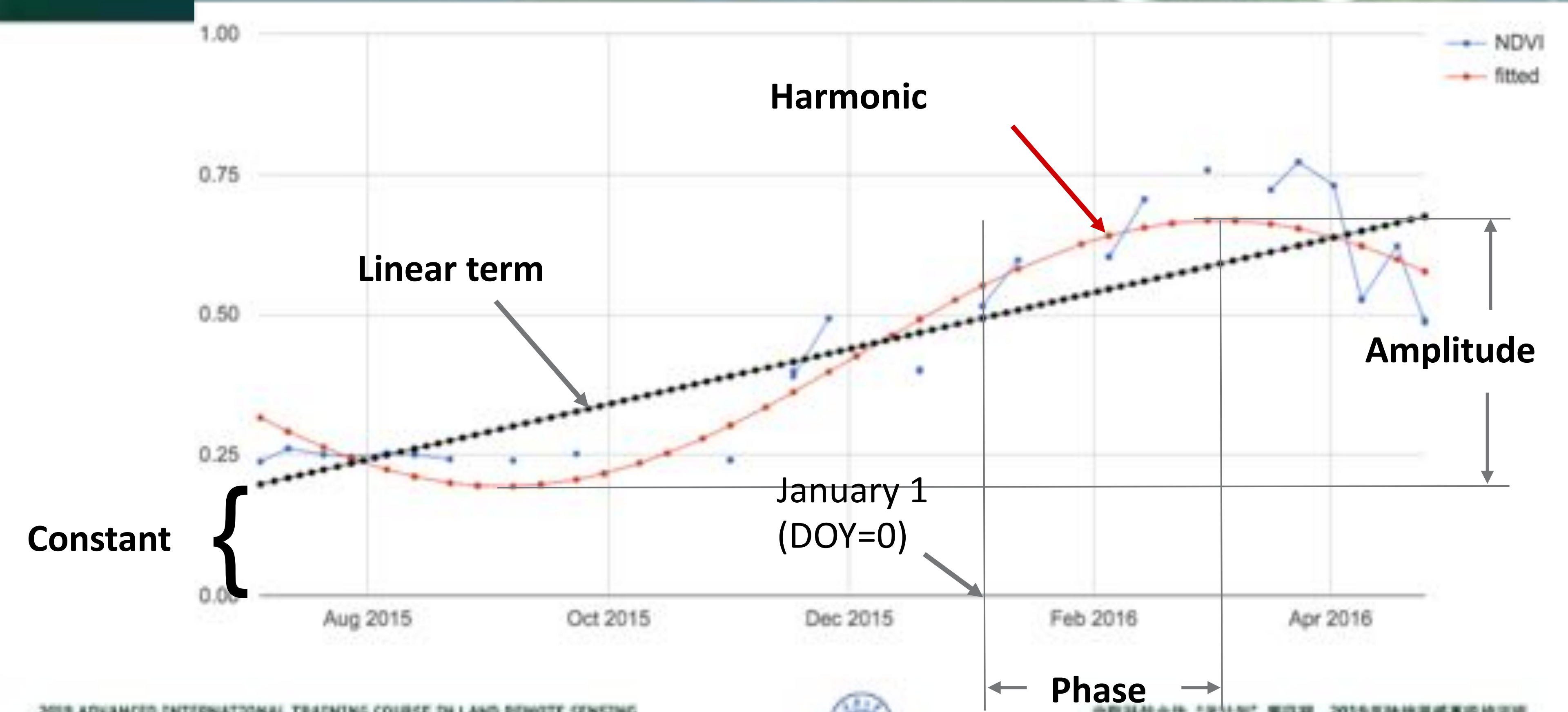


https://code.earthengine.google.com/0827ce3bbe401c769090e1d192d47b74

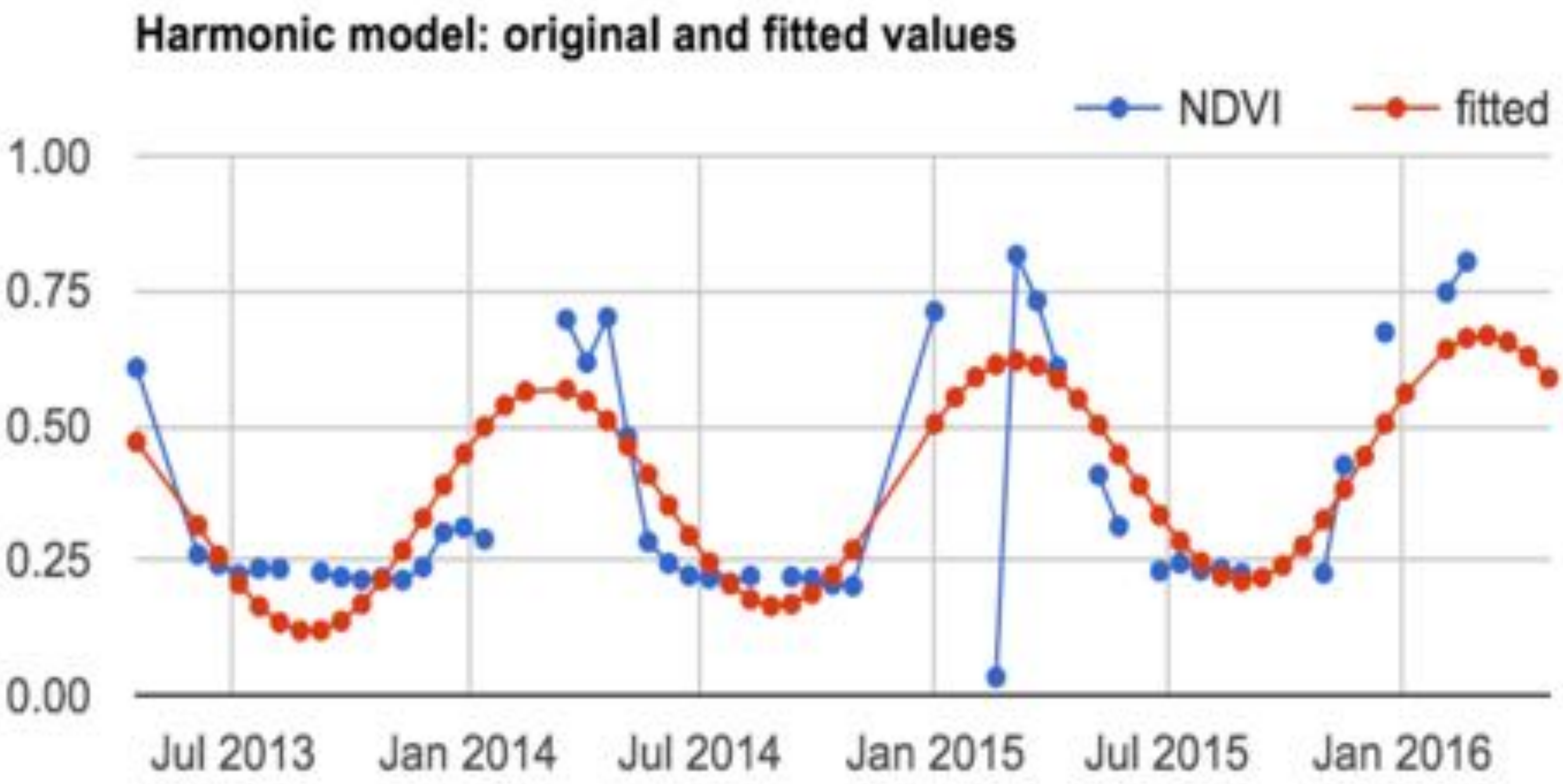
Model fitting: linear trend



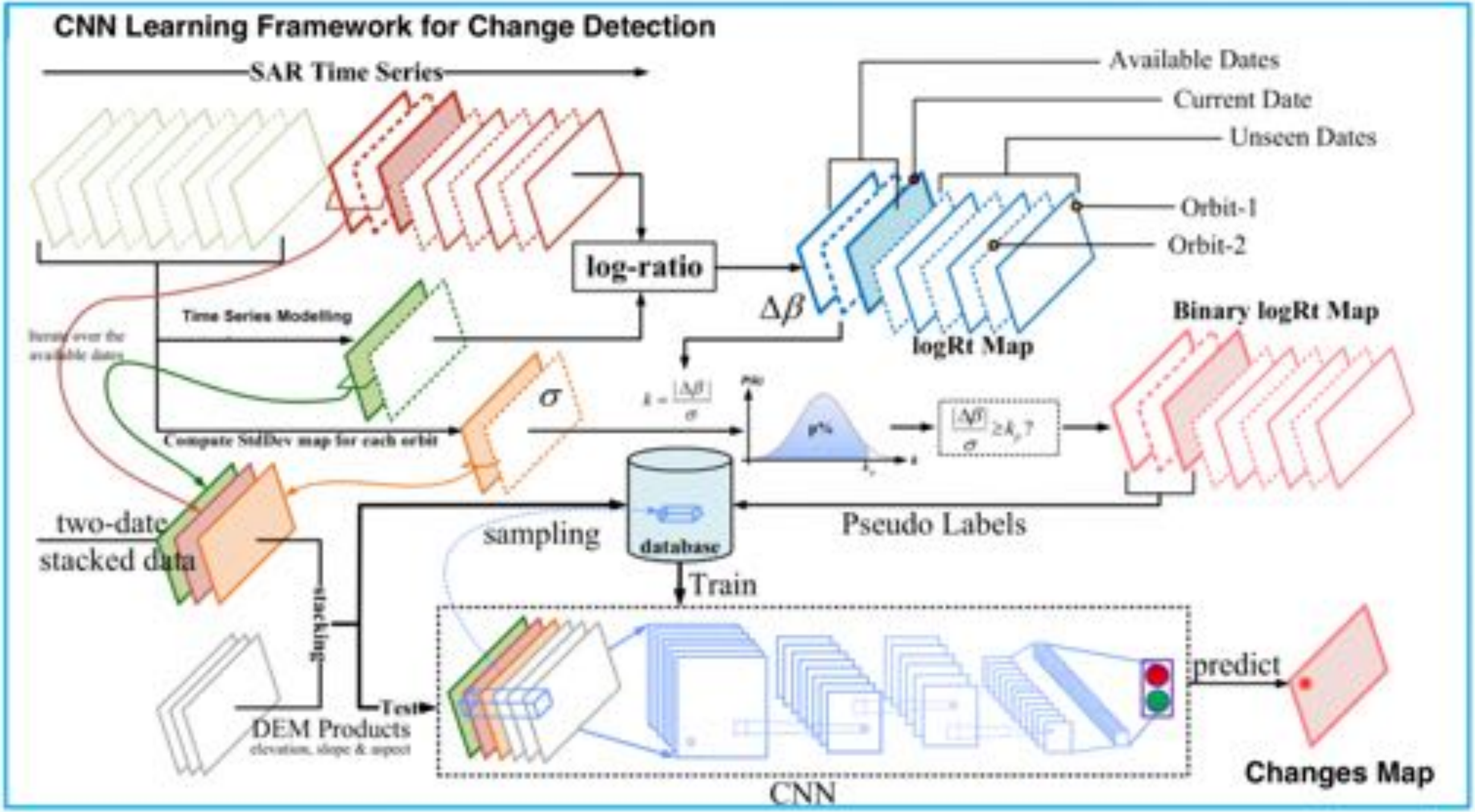
Model Fitting: Harmonic model



Harmonic model



CNN-Based Change Detection

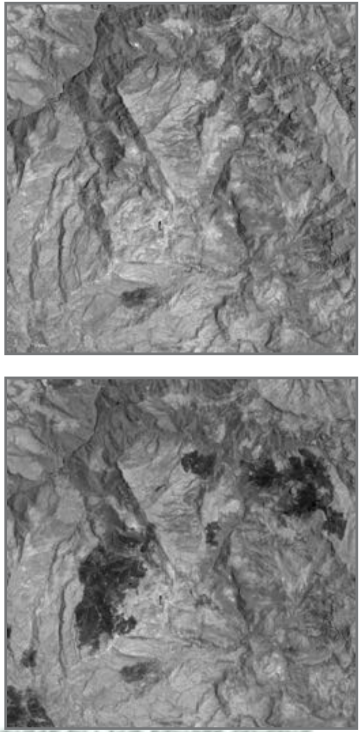


Multispectral Data: CD few years ago…..

Landsat TM (burned area detection)

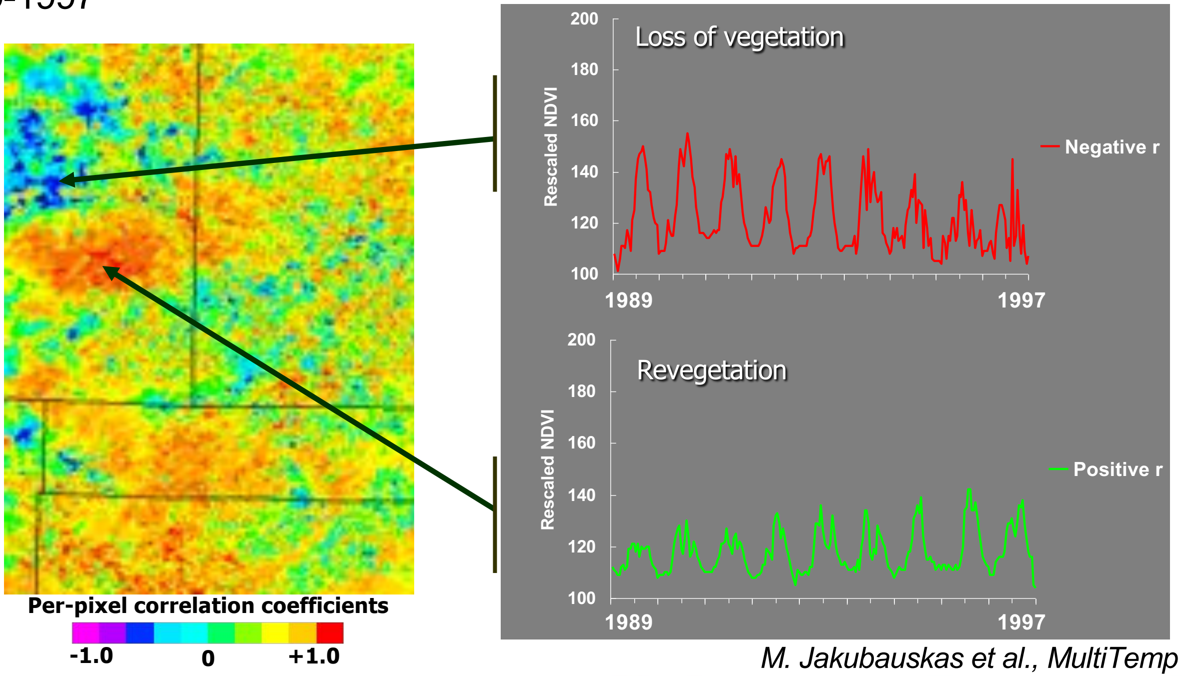
Before Change

After Change



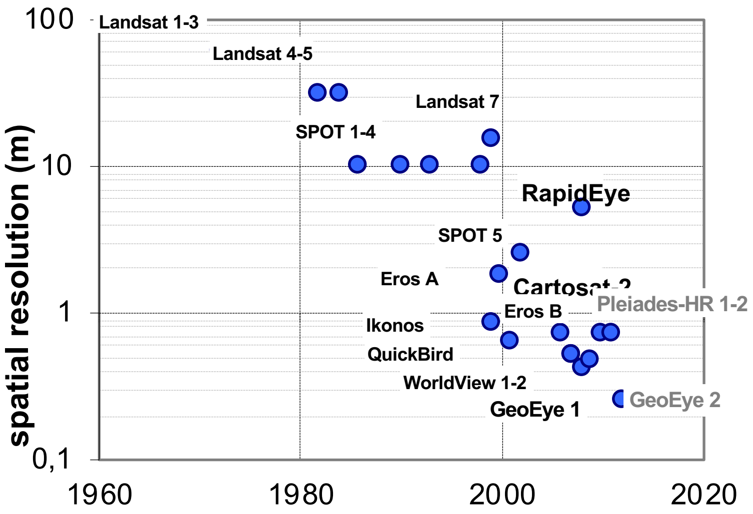
Multispectral Data: CD few years ago…..

First term amplitude change,
1989-1997



M. Jakubauskas et al., MultiTemp 2001

New Satellites with VHR Multispectral (MS) Sensors



Multispectral Data: CD Now…..

July 2006

October 2005

Quickbird images of the city of Trento (Italy)

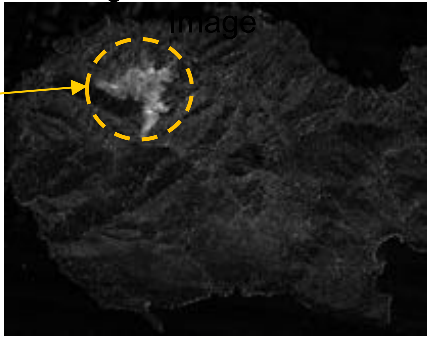
CD in Multitemporal HR Images: Example

Landsat TM, Pre-event

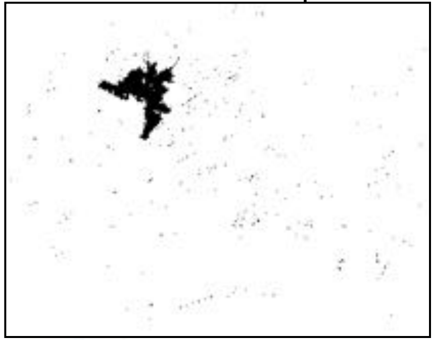
Landsat TM, Post-event

Magnitude Difference image

Burned area

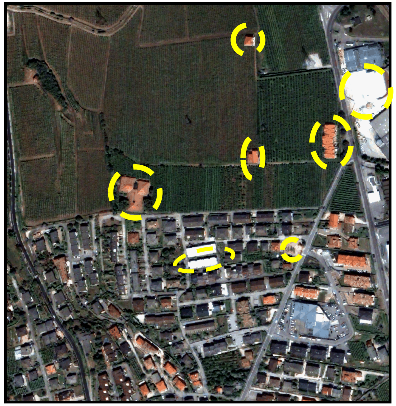


Pixel-Based Change Detection Map



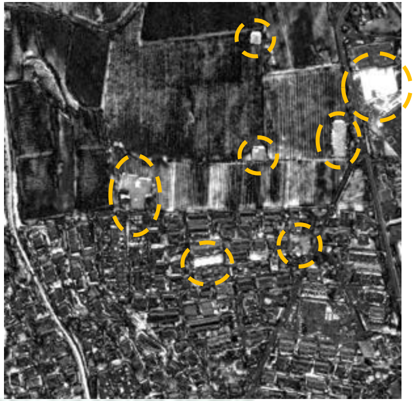
CD in Multitemporal VHR Images: Example

Quickbird, October 2004 (true color composition)

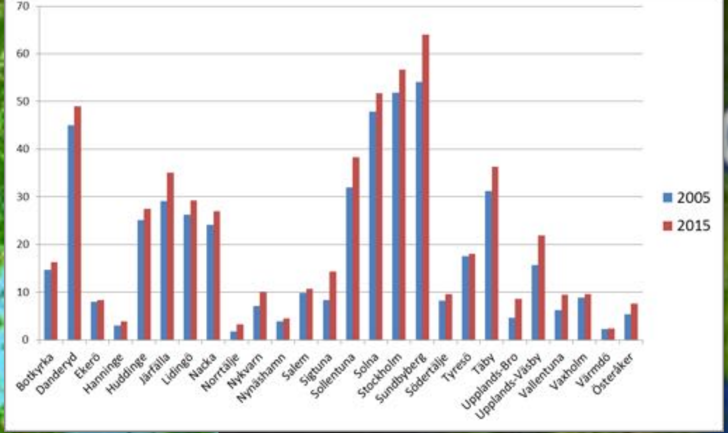


Quickbird, July 2006 true color composition

Magnitude Difference Image



Pixel-Based Change Detection Map



Stockholm växer, grönområdena krymper

TEMA En mänskligare stad

Change Detection in VHR Images

Main assumption of "traditional" techniques: unsupervised change-detection techniques generally assume that multitemporal images are similar to each other except for the presence of changes occurred on the ground.

Problems: This assumption is seldom satisfied in VHR images due to:

- the complexity of the objects present in the scene (which may show different spectral behaviors at two different dates even if their semantic meaning does not change);
- the differences in the acquisition conditions (e.g., sensor acquisition geometry, atmospheric and sunlight conditions, etc.).

New challenges: Passive Sensors

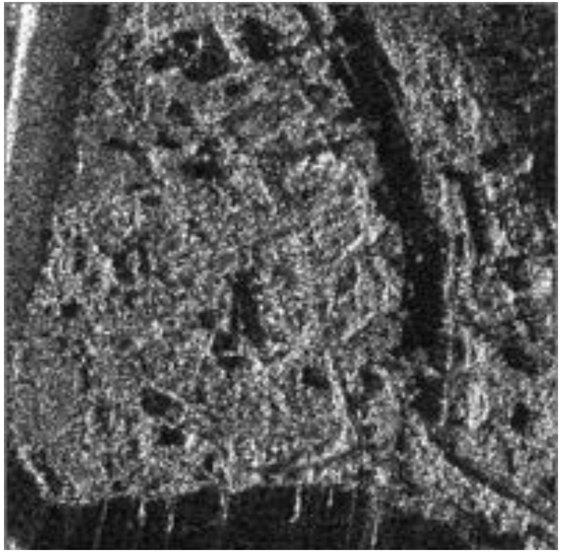
- ✓ Increased spatial resolution:
  - Increased complexity in the scene and in the image analysis;
  - Strong impact of the acquisition parameters in the comparison of images;
  - Radiometric changes often do not represent real changes occurred on the ground.

- ✓ Long time series of VHR images:
  - How optimizing the pre-processing chain?
  - How extracting relevant information?
  - New applications?

- ✓ Increased spectral resolution:
  - Time series of hyperspectral images;
  - New satellite missions with hyperspectral sensor on board

New Challenges: Data Fusion

Quickbird image before earthquake



COSMO-Skymed image after earthquake

Earthquake of Sichuan province, China, May, 2008

D. Brunner, G. Lemoine, L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48, No. 5, pp.2403-2420, 2010.

New Challenges with EO Big Data

✓ Increased amount of data and longer time series. Crucial for studies on:

- Climate change.
- Urbanization
- Vegetation dynamics.
- Forestry.
- Desertification.
- …..

✓ Data mining in very long time series (detection of spatio-temporal events)

Urbanization Monitoring