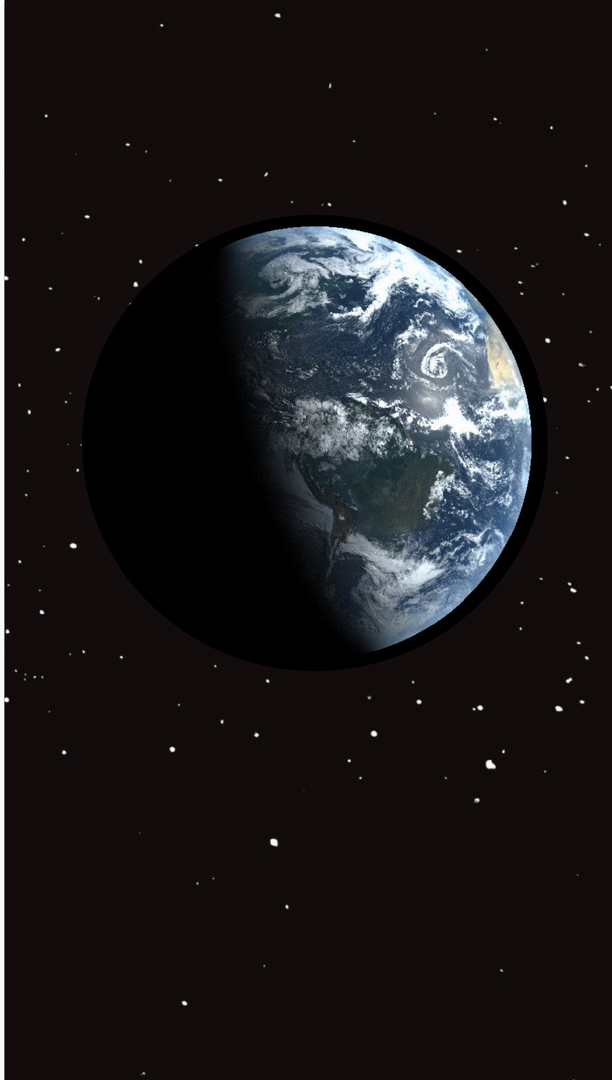
Google

Super-charging EO analysis with Geospatial AI

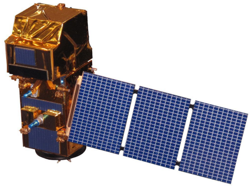
Valerie Pasquarella
Research Scientist, Google Earth Engine

Google Geo for Gov Summit
July 31, 2024



10
million
Landsat
TM / ETM+ / OLI

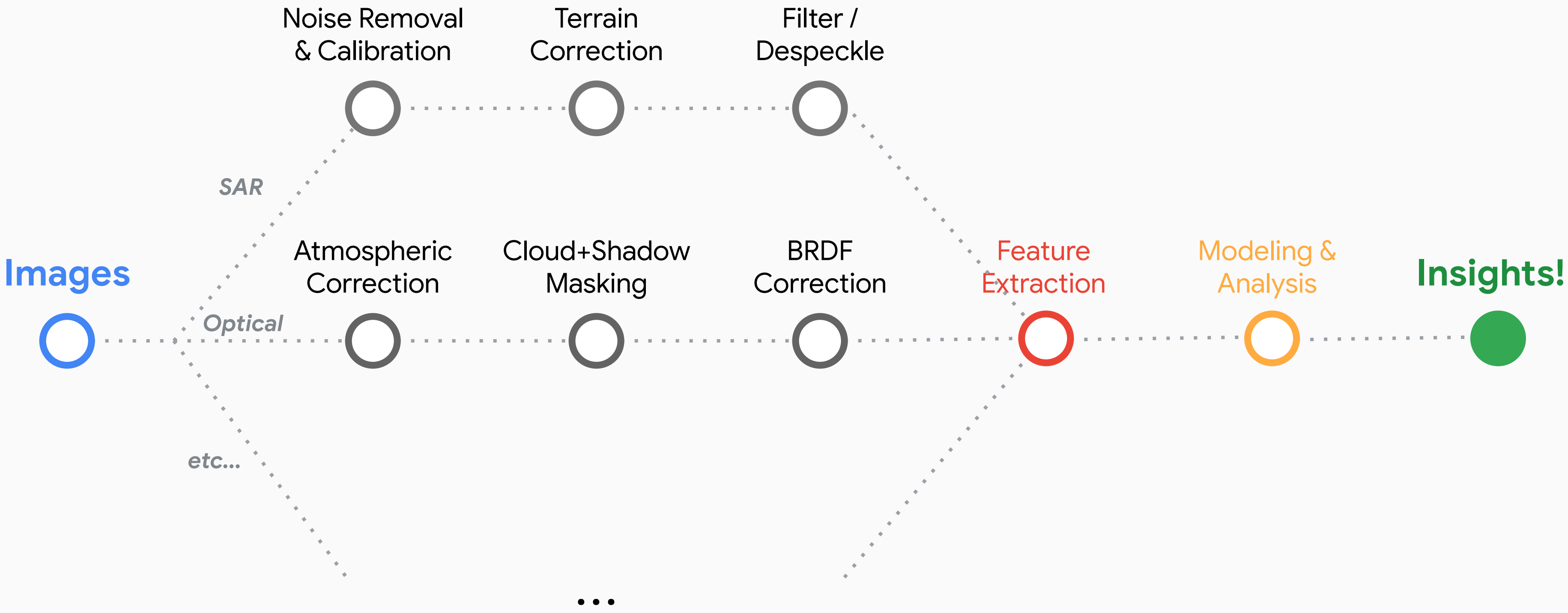
31
million
Sentinel-2
L1C

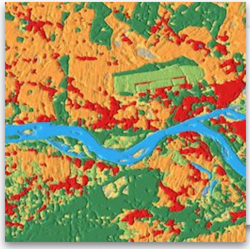


2
million
Sentinel-1
GRD

Google

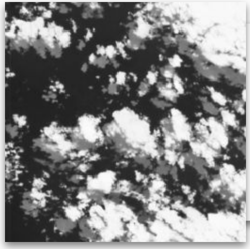
images != insights


Noise Removal & Calibration
Terrain Correction
Filter / Despeckle
SAR
Images
Atmospheric Correction
Cloud+Shadow Masking
BRDF Correction
Feature Extraction
Modeling & Analysis
Insights!
Optical
etc...
...
Google



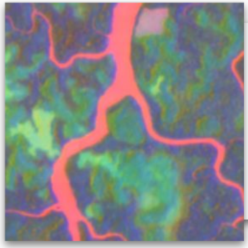
Dynamic World

Near-real-time land cover classification



Cloud Score+

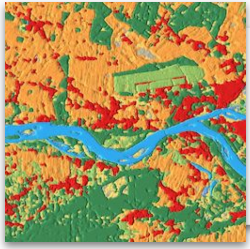
First-of-its-kind comprehensive per-pixel QA score



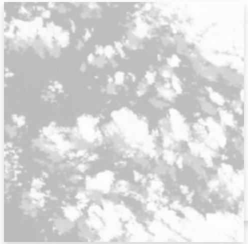
Embedding Fields Model

Generating foundational features for accelerating EO workflows

Google

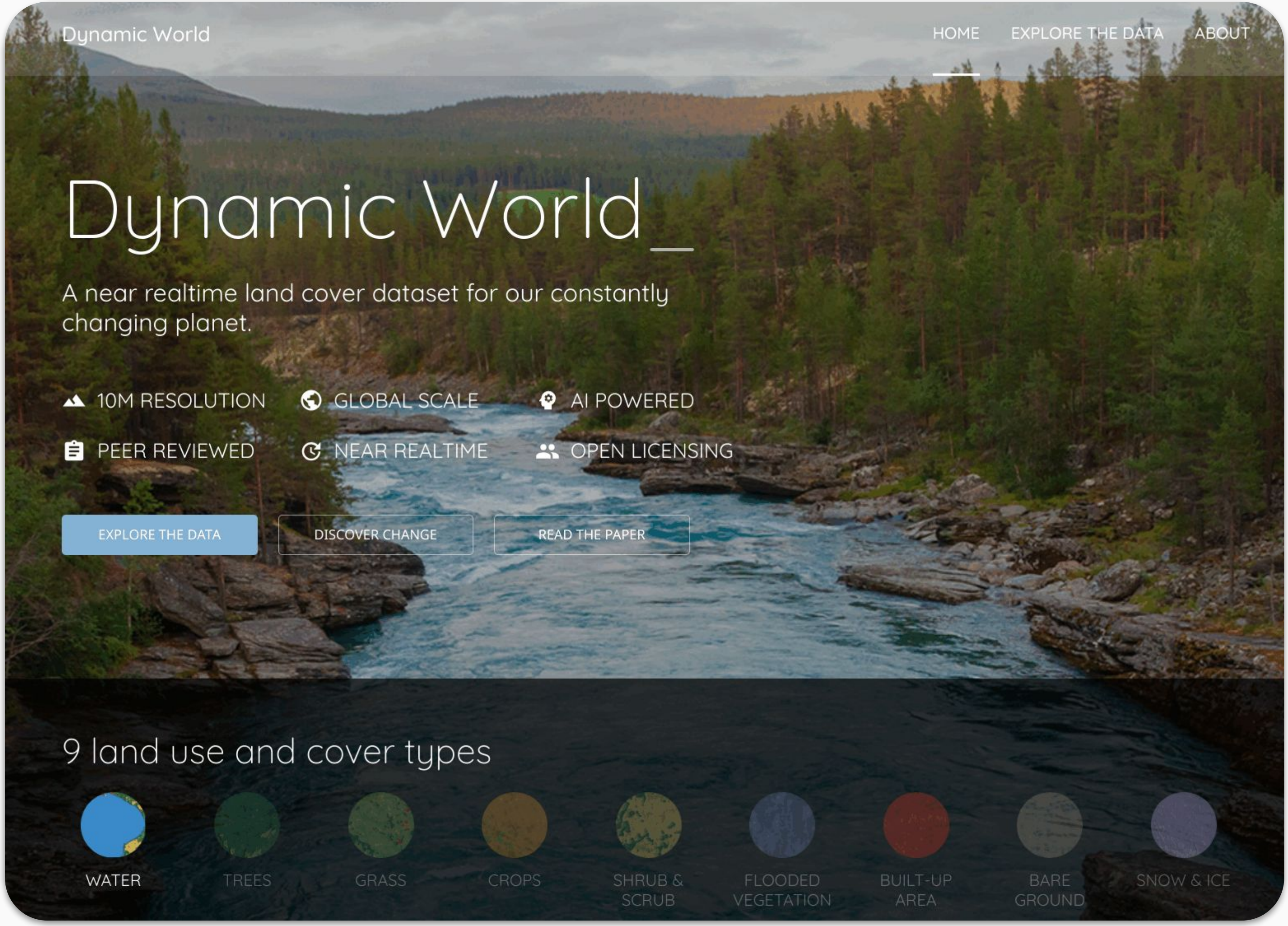


Dynamic World
Near-real-time land cover classification

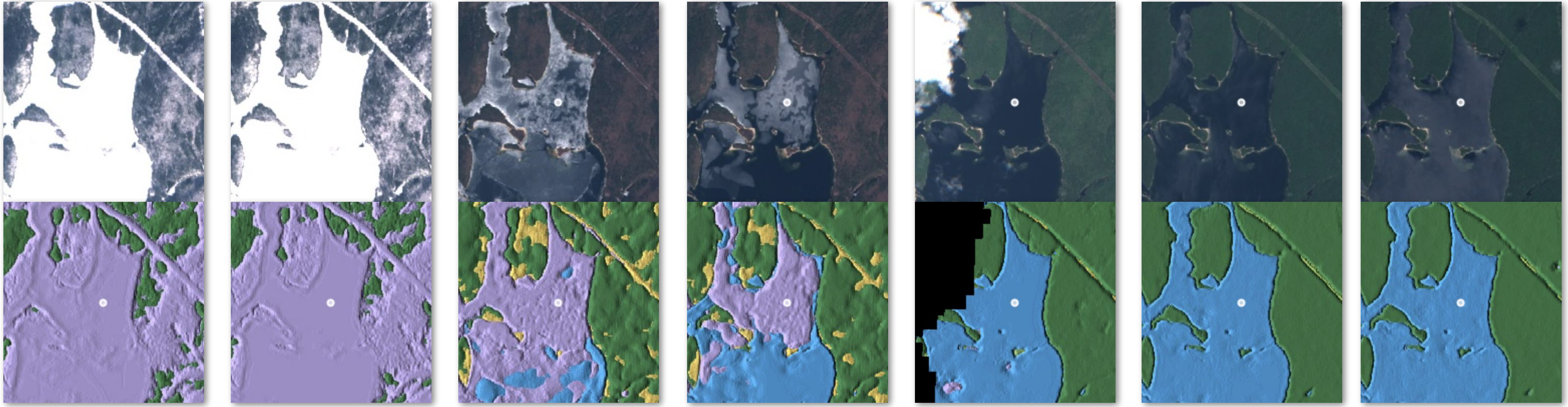


Cloud Score+
First-of-its-kind comprehensive per-pixel QA score

Embedding Fields Model
Generating foundational features for accelerating EO workflows


Dynamic World
HOME
EXPLORE THE DATA
ABOUT
Dynamic World_
A near realtime land cover dataset for our constantly changing planet.
10M RESOLUTION
GLOBAL SCALE
AI POWERED
PEER REVIEWED
NEAR REALTIME
OPEN LICENSING
EXPLORE THE DATA
DISCOVER CHANGE
READ THE PAPER
9 land use and cover types
WATER
TREES
GRASS
CROPS
SHRUB & SCRUB
FLOODED VEGETATION
BUILT-UP AREA
BARE GROUND
SNOW & ICE

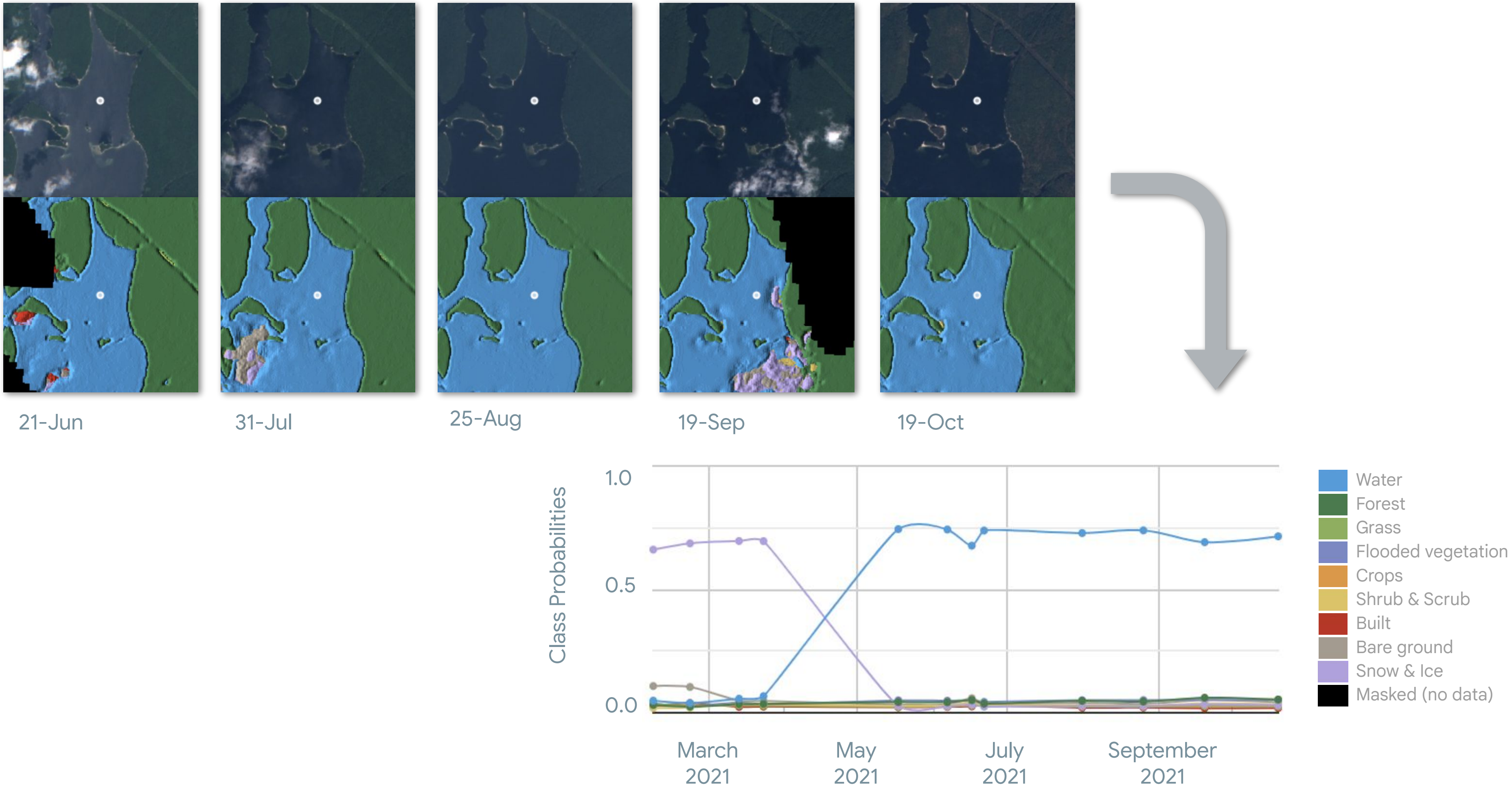
2021



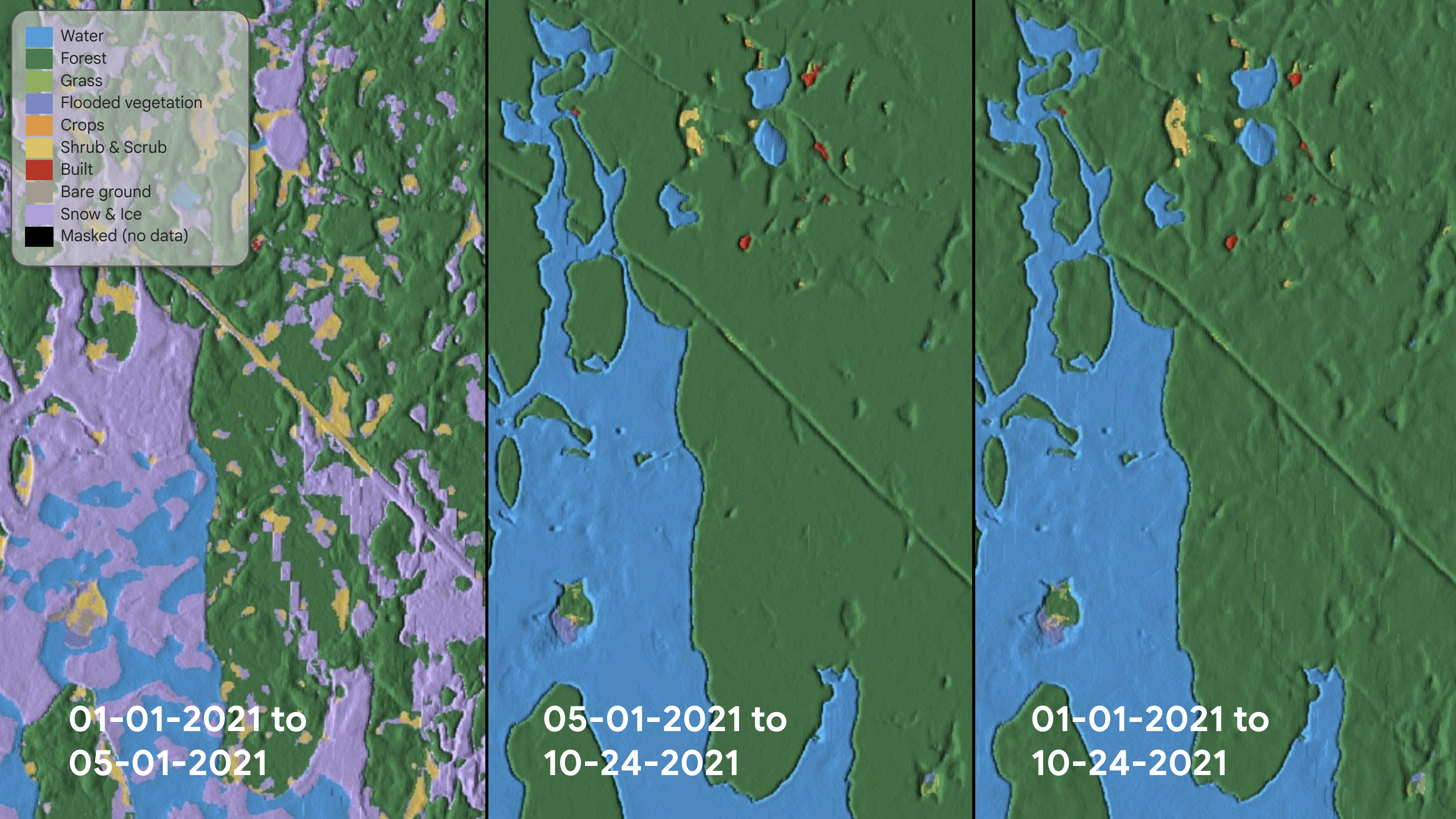
06-Feb | 21-Feb | 13-Mar | 23-Mar | 17-May | 06-Jun | 16-Jun

- Water
- Forest
- Grass
- Flooded vegetation
- Crops
- Shrub & Scrub
- Built
- Bare ground
- Snow & Ice
- Masked (no data)

Single-image classification ⟶ **NRT** mapping



21-Jun
31-Jul
25-Aug
19-Sep
19-Oct
Class Probabilities
1.0
0.5
0.0
March 2021
May 2021
July 2021
September 2021
Water
Forest
Grass
Flooded vegetation
Crops
Shrub & Scrub
Built
Bare ground
Snow & Ice
Masked (no data)


Water
Forest
Grass
Flooded vegetation
Crops
Shrub & Scrub
Built
Bare ground
Snow & Ice
Masked (no data)
01-01-2021 to
05-01-2021
05-01-2021 to
10-24-2021
01-01-2021 to
10-24-2021

This is how researchers contributed to an NRK documentary about the degradation of nature in Norway

Researchers used both satellite images and AI in collaboration with journalists at NRK. They are now documenting how we have lost bit by bit Norwegian nature.

Anne Olga Syverhuset
COMMUNICATIONS ADVISOR

Norwegian Institute for Natural Research (NINA)

Saturday 13 January 2024 - 04:31

The NRK case *Norway in red, white and gray* made frightening figures visible: Norway has on average lost at least 79 square meters of nature per minute over the last five years.

It is 207 square kilometers in total. The story behind the numbers is also fascinating.

There was no overview of built-up nature

In 2022, Zander Venter, who is a researcher at NINA, was contacted by Mads Nyborg Støstad with a simple question: Do we have data that shows us how much nature our society consumes in Norway? Støstad had asked Statistics Norway, the Norwegian Institute for Bioeconomy (NIBIO) and the Norwegian Environment Agency for a complete dataset of all nature interventions over a year, and no one could give him a simple answer. Venter also pointed out that there are no readily available maps of decommissioning.

- I had recently published a scientific article on satellite-based maps of land use and was coincidentally in the process of including the maps in NINA's projects on ecosystem accounting when Støstad contacted me, says Venter.

NINA uses a bird's eye view and AI

Venter used Google's dataset called Dynamic World . It uses openly available images from the two satellites Sentinel-2A and Sentinel-2B. They are managed by the European Space Agency (ESA).

The satellites fly continuously over the earth and take pictures of the whole world, including Norway, with a resolution of ten metres.

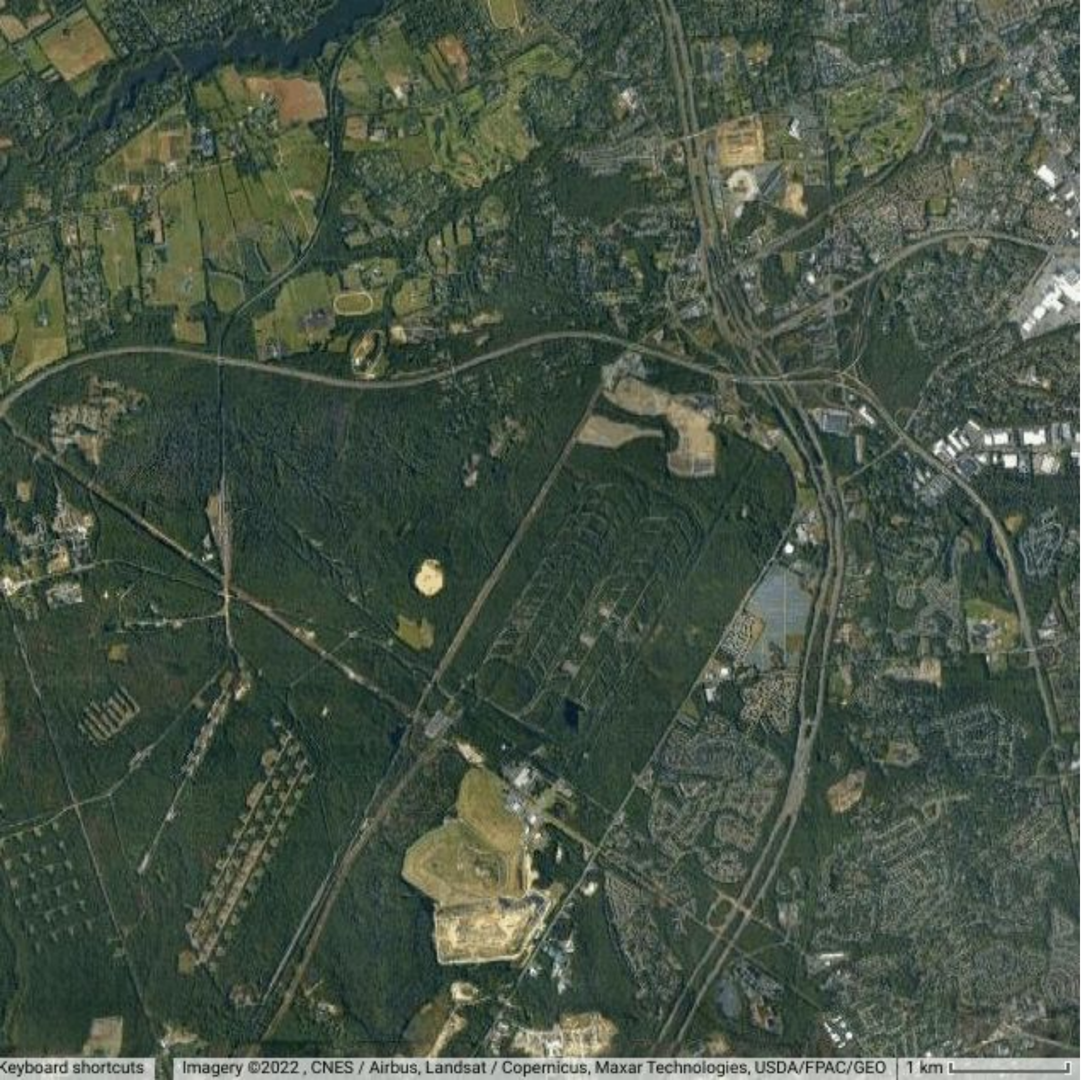
The images are rather unclear, but Google uses artificial intelligence (AI/AI) to analyze such satellite images and recognize different types of land use.

The artificial intelligence can recognize nine different categories: water, trees, grass, flooded vegetation, snow and ice, bushes and undergrowth, bare land, cultivated land and built-up areas.

Zander used the raw data from the AI model to reveal whether natural areas have been built on.

More specifically, the artificial intelligence gives a probability that each and every data point of 10x10 meters on the map of Norway is built on or not. Using time series analysis, Venter was able to identify data points, and ultimately areas of data points, that have gone from natural to built-up cover.

Source: forskning.no



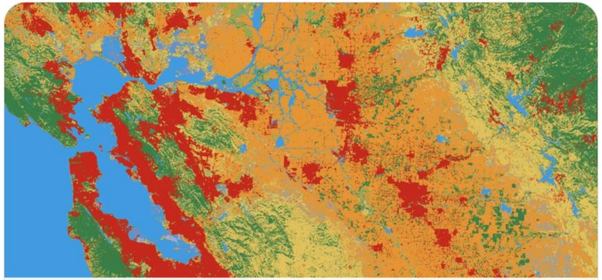
Class **probabilities >**
Discrete **labels**

Google

EARTH BLOX

Earth Blox dataset review: Dynamic World Landcover (10m)

Added to Satellite Data June 7, 2024



Google WORLD RESOURCES INSTITUTE

Per-pixel probabilities

As well as providing a classification map, Dynamic World also offers probability bands (as a %) for each class. The class assigned to the pixel is the class that has the highest probability. It is useful to be able to see the percentage because some pixels may have one dominant class, but for others, the classes may all score very similar values. You can therefore use the % values as an indicator of confidence in the classification. For example, imagine a pixel has been assigned to the TREES class. Sometimes that might be because TREES is 100% and the other eight classes are 0%. But the pixel would have the same class if TREES was only 12% and the other eight classes scored 11%. Clearly, the former pixel is a more reliable classification than the latter. In Earth Blox, it is straightforward to add a block that could, for instance, mask out any pixels that don't have a high % classification so that users can be confident they are seeing only the most reliable results.

The other way to use the band of % values is to use this as an input into a supervised classification. This is particularly powerful if you are operating in a landscape where

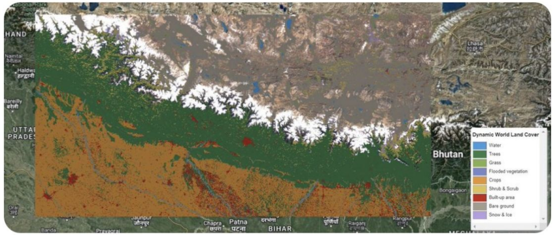
The other way to use the band of % values is to use this as an input into a supervised classification. This is particularly powerful if you are operating in a landscape where the 9 default land classes don't especially align with the land cover types you want to identify.

Source: earthblox.io

AMRIT THAPA
@amrit2044

Follow

#DynamicWorld A great dataset, however, the cloud mask prevents us from seeing snow over the Himalayan region.

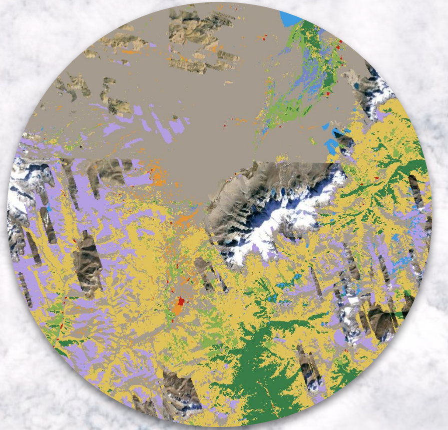


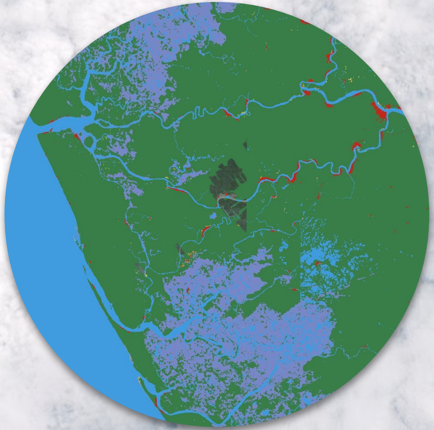
2:22 PM · 6/10/22 From Earth

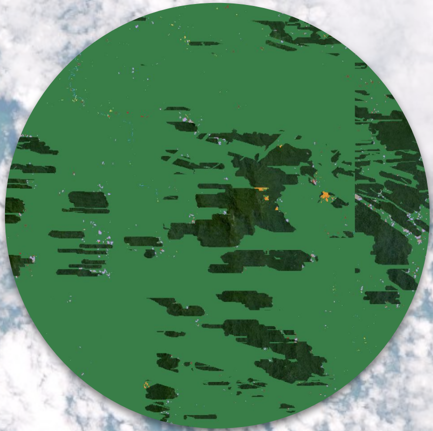
1 Repost **30** Likes

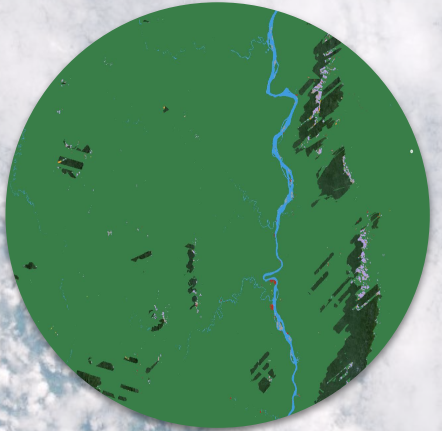
AMRIT THAPA @amrit2044 · 6/10/22

However, it looks better for Ali Abad, Hunza. #DynamicWorld

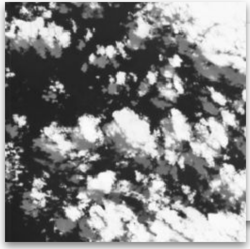








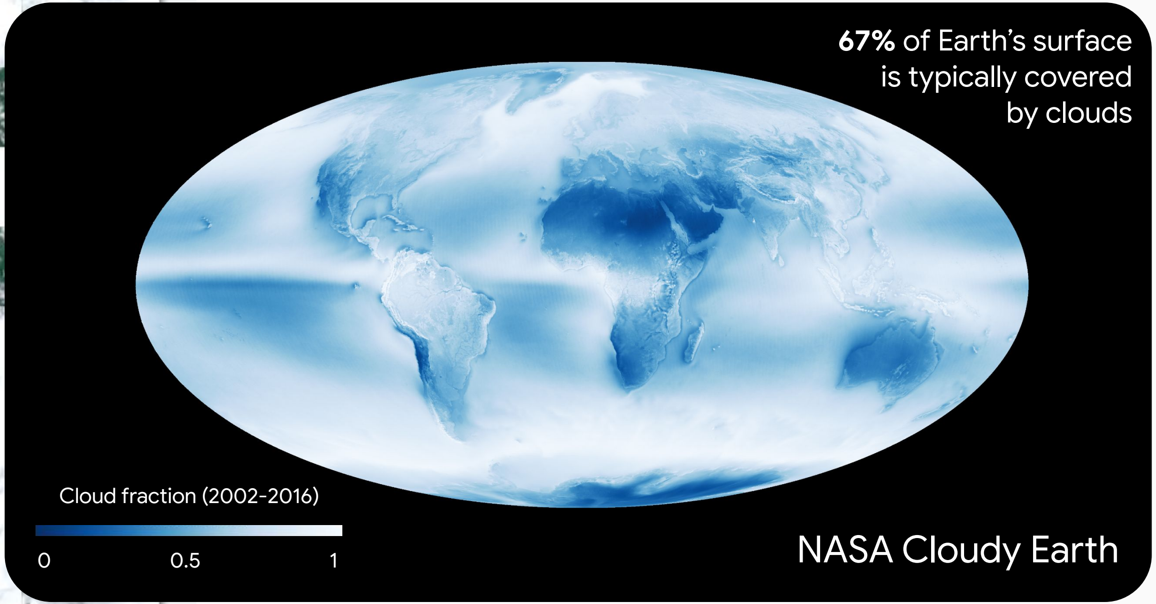
Dynamic World
Near-real-time land cover classification

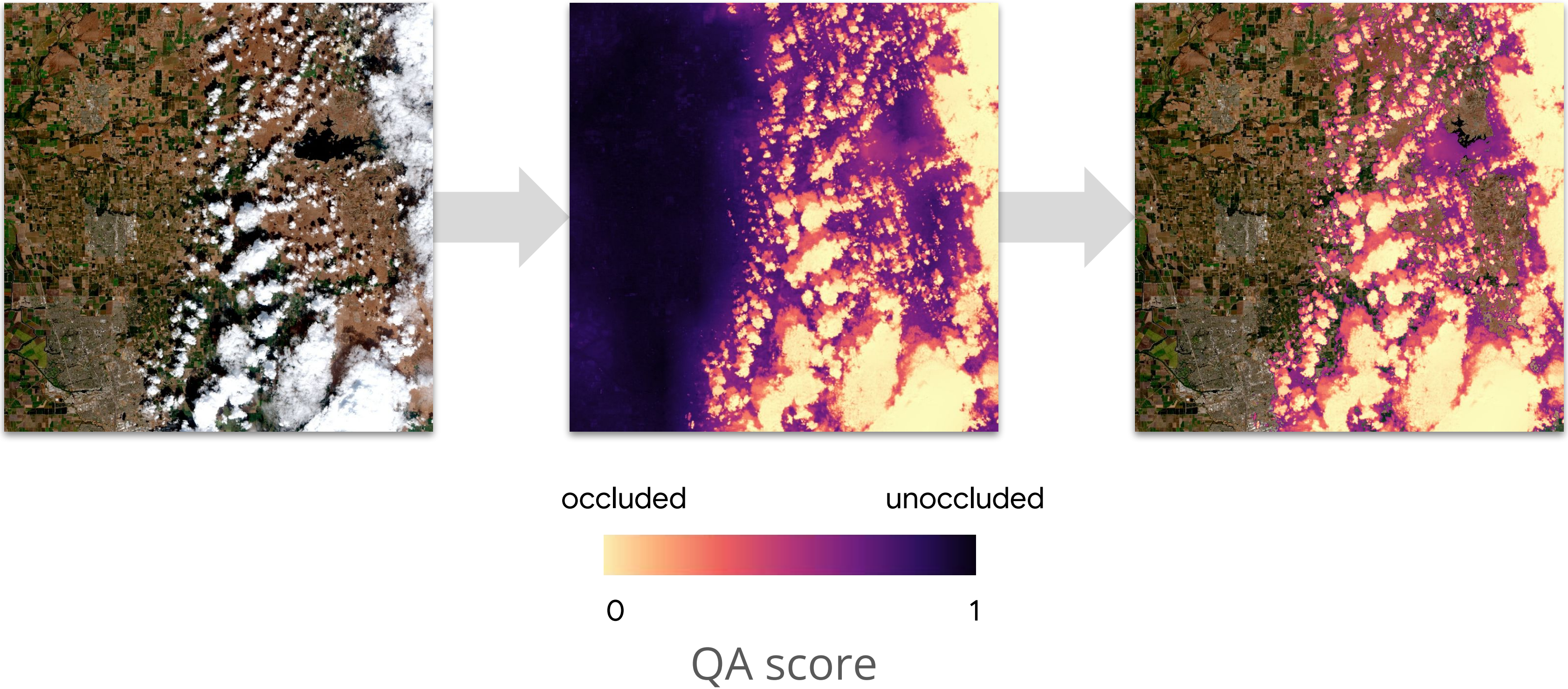


Cloud Score+
First-of-its-kind comprehensive per-pixel QA score

Embedding Fields Model
Generating foundational features for accelerating EO workflows

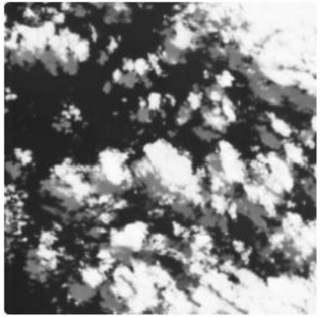
Imbabura, Ecuador





Google

Cloud Score+ S2_HARMONIZED V1



Dataset Availability

2015-06-27T00:00:00 -

Dataset Provider

Google Earth Engine

Collection Snippet

```
ee.ImageCollection("GOOGLE/CLOUD_SCORE_PLUS/V1/S2_HARMONIZED")
```

See example

Tags

cloud google sentinel2-derived

DESCRIPTION | BANDS | IMAGE PROPERTIES | TERMS OF USE | CITATIONS

Cloud Score+ is a quality assessment (QA) processor for medium-to-high resolution optical satellite imagery. The Cloud Score+ S2_HARMONIZED dataset is being operationally produced from the harmonized Sentinel-2 L1C collection, and Cloud Score+ outputs can be used to identify relatively clear pixels and effectively remove clouds and cloud shadows from L1C (Top-of-Atmosphere) or L2A (Surface Reflectance) imagery.

The Cloud Score+ S2_HARMONIZED dataset includes two QA bands, `cs` and `cs_cdf`, that both grade the usability of individual pixels with respect to surface visibility on a continuous scale between 0 and 1, where 0 represents "not clear" (occluded), while 1 represents "clear" (unoccluded) observations. The `cs` band scores QA based on a spectral distance between the observed pixel and a (theoretical) clear reference observation, while the `cs_cdf` band represents the likelihood an observed pixel is clear based on an estimated cumulative distribution of scores for a given location through time. In other words, `cs` can be thought of as a more instantaneous atmospheric similarity score (i.e., how similar is this pixel to what we'd expect to see in a perfectly a clear reference), while `cs_cdf` captures an expectation of the estimated score through time (i.e., if we had all the scores for this pixel through time, how would this score rank?).

Images in the Cloud Score+ S2_HARMONIZED collection have the same id and `system:index` properties as the individual Sentinel-2 L1C assets from which they were produced such that Cloud Score+ bands can be linked to source images based on their shared `system:index`.

Cloud Score+ backfill for the entire Sentinel-2 archive is currently in progress and Dataset

CLOSE IMPORT

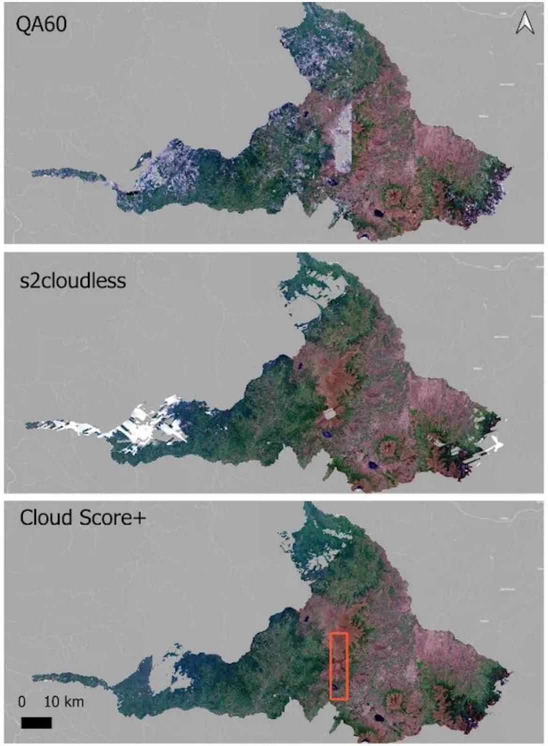
Cloud Score+ in Action: Land Cover Mapping in Ecuador

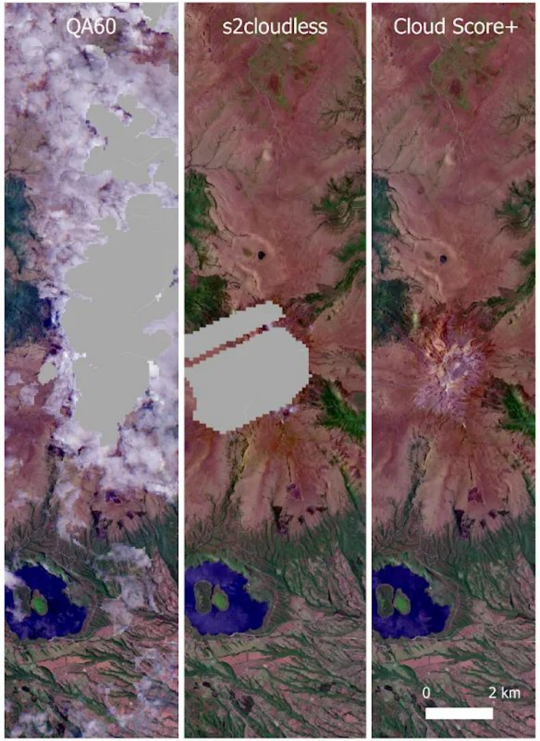
Google Earth · Follow
Published in Google Earth and Earth Engine · 9 min read · Mar 27, 2024

94

By Andréa P. Nicolau, Geospatial Data Scientist, Spatial Informatics Group, Earth Engine Google Developer Expert

Monitoring the Earth's surface has always depended on the availability of high-quality cloud-free imagery. Now, imagine trying to map and monitor land cover and land use in the world's cloudiest spots. Nightmare, right? But fear not, with the recent launch of Cloud Score+ for Sentinel-2, the clouds are parting ways for clearer composites. In this post, we'll dive into a case study that shines a light on how Cloud Score+ is revolutionizing land cover





Source: Google Earth Medium

Google

Cloud Score+ in Action: Land Cover Mapping in Ecuador

Google Earth · Follow
Published in Google Earth and Earth Engine · 9 min read · Mar 27, 2024

94

By Andréa P. Nicolau, Geospatial Data Scientist, Spatial Informatics Group, Earth Engine Google Developer Expert

Monitoring the Earth's surface has always depended on the availability of high-quality cloud-free imagery. Now, imagine trying to map and monitor land cover and land use in the world's cloudiest spots. Nightmare, right? But fear not, with the recent launch of Cloud Score+ for Sentinel-2, the clouds are parting ways for clearer composites. In this post, we'll dive into a case study that shines a light on how Cloud Score+ is revolutionizing land cover

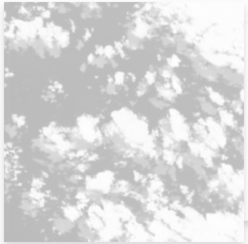
	User's Accuracy (%)			Producer's Accuracy (%)		
LULC classes	QA60 band	s2cloud less	CS+ (cs 0.4)	QA60 band	S2cloud less	CS+ (cs 0.4)
Forest	67	73	75	81	78	81
Natural	95	95	95	87	87	88
Shrublands	49	50	50	41	45	46
Croplands and Pasturelands	50	54	53	50	56	55
Moorlands	83	83	85	90	93	92
Urban areas	75	76	78	78	79	80
Bareland	77	76	78	69	68	75
Forest plantations	74	73	76	79	76	78
Infrastructure	65	65	64	64	63	62

Source: Google Earth Medium

Watch our [Geo for Good 2023 session] for a deep-dive into the CS+ model and applications

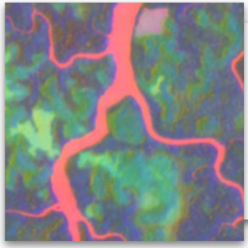
Dynamic World

Near-real-time land cover classification



Cloud Score+

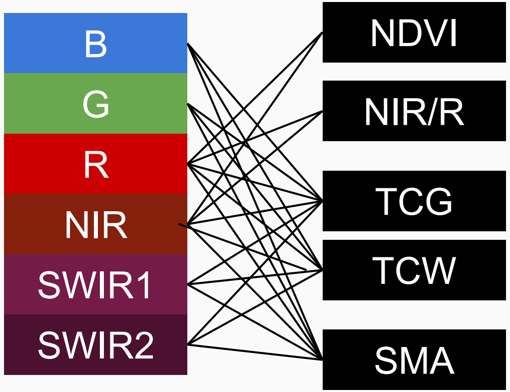
First-of-its-kind comprehensive per-pixel QA score



Embedding Fields Model

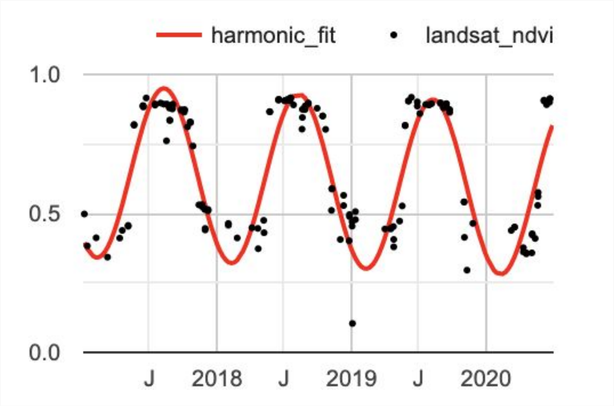
Generating learned features for accelerating EO workflows

Spectral



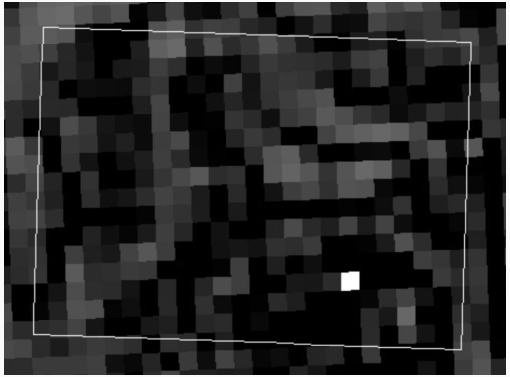
Multi-band combinations
Common spectral indices and transforms to emphasize key (surface) properties

Temporal



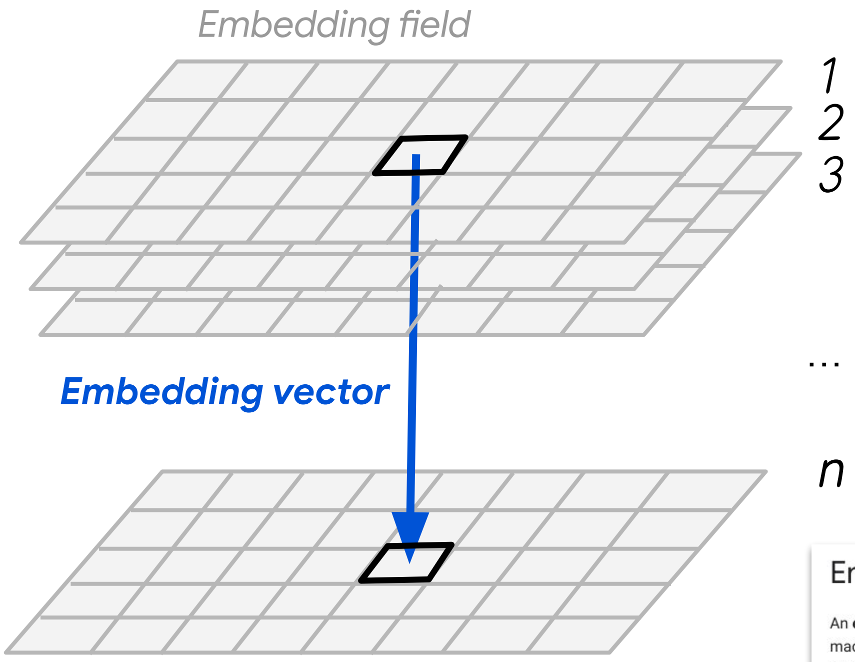
Spectral + time
Simple harmonic regression models fit to time series to understand seasonalities & phenologies of ecosystems, croplands, etc.

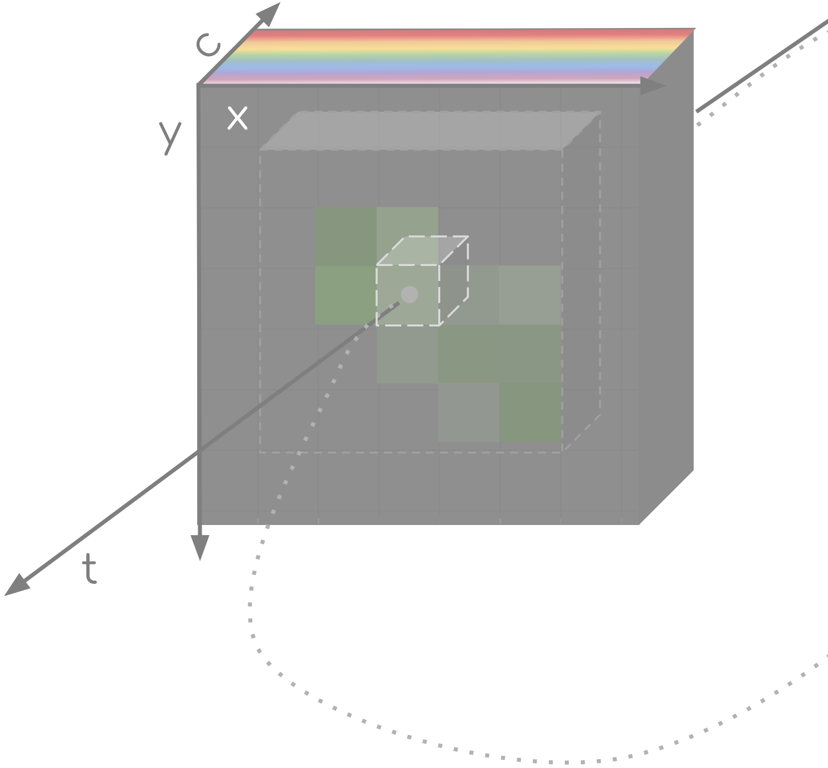
Spatial



Spectral + kernel
GLCM correlation, texture/kernel-based metrics, neighborhood analyses to understand pixel interrelatedness

What’s an embedding field?



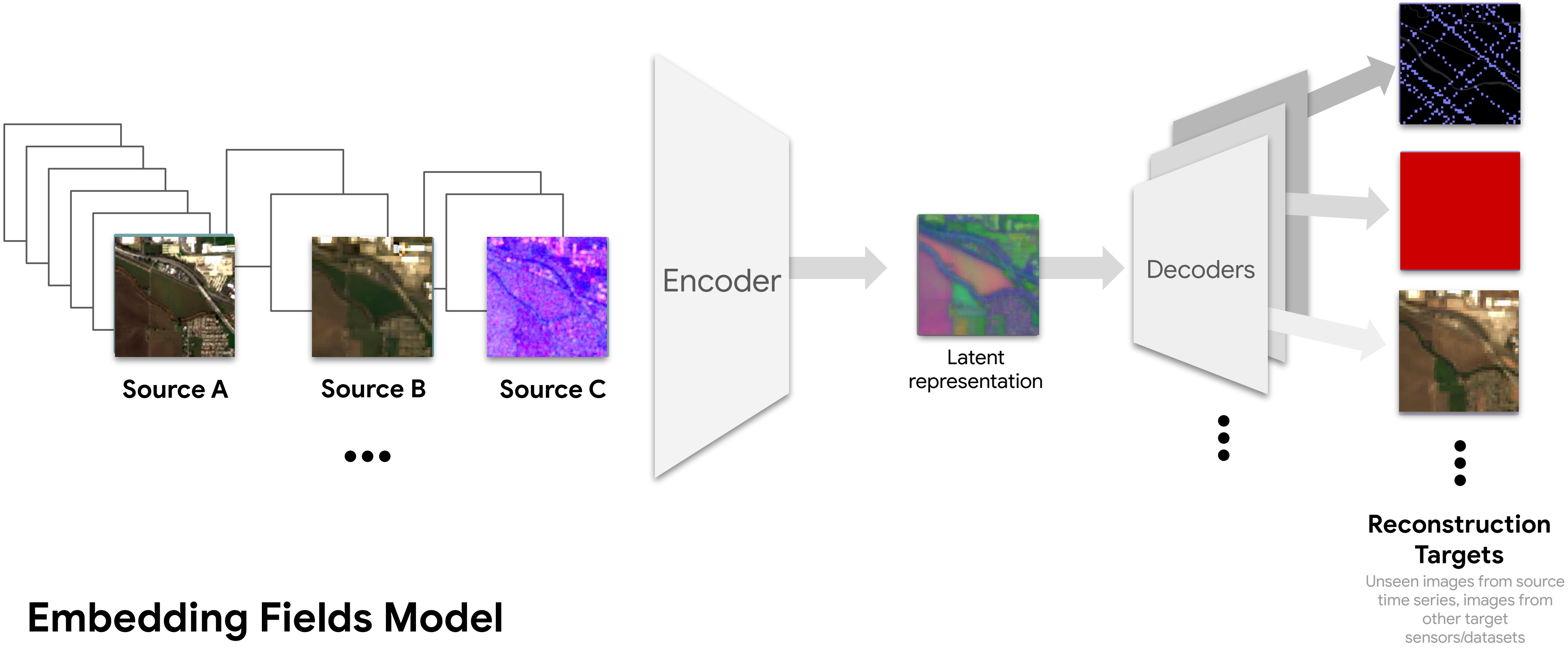


Embeddings

Send feedback

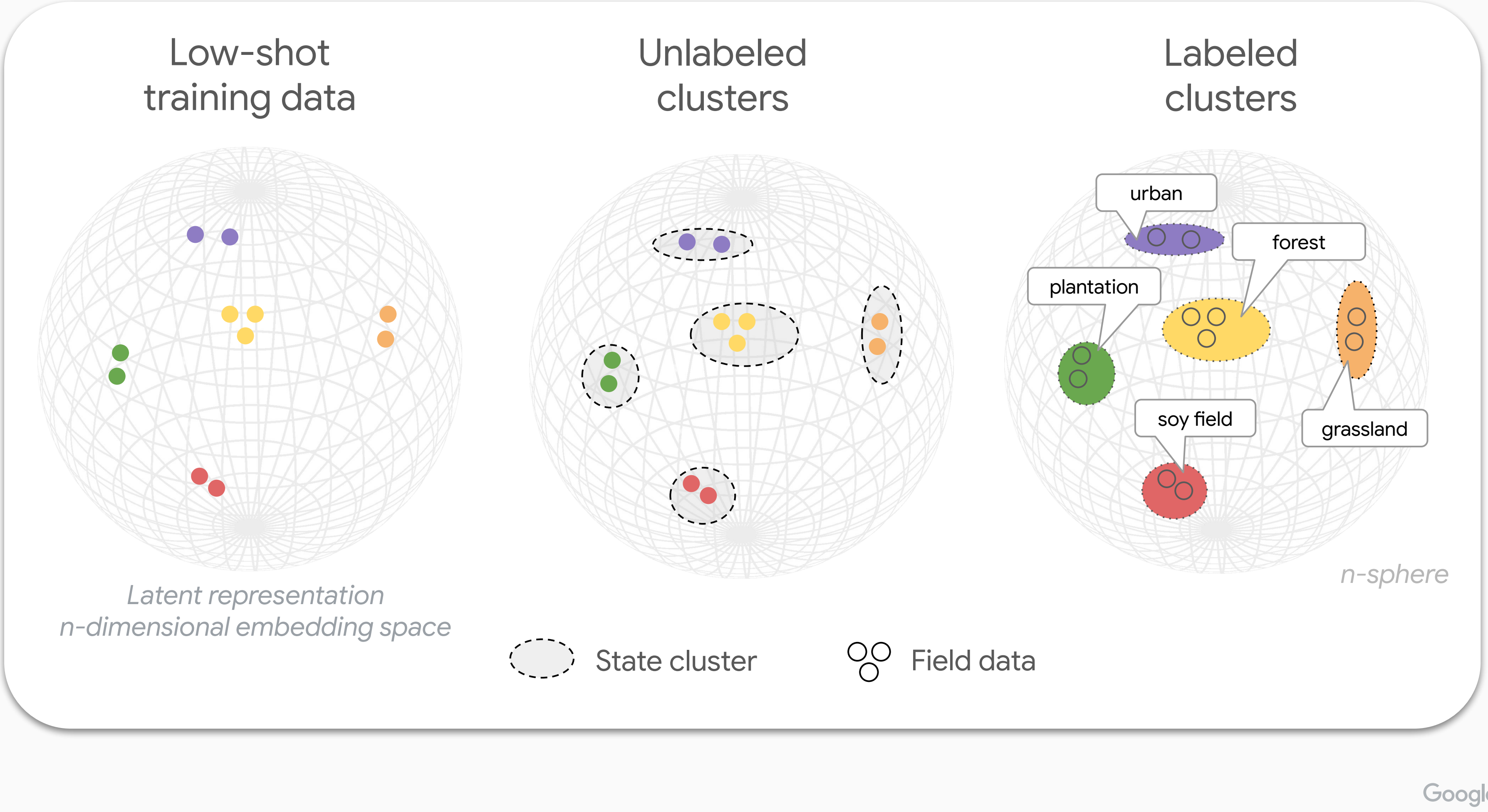
An **embedding** is a relatively low-dimensional space into which you can translate high-dimensional vectors. Embeddings make it easier to do machine learning on large inputs like sparse vectors representing words. Ideally, an embedding captures some of the semantics of the input by placing semantically similar inputs close together in the embedding space. An embedding can be learned and reused across models.

https://developers.google.com/machine-learning/crash-course/embeddings/video-lecture



Embedding Fields Model

Google


Low-shot training data
Unlabeled clusters
Labeled clusters
urban
forest
plantation
soy field
grassland
n-sphere
Latent representation
n-dimensional embedding space
State cluster
Field data
Google

Agriculture

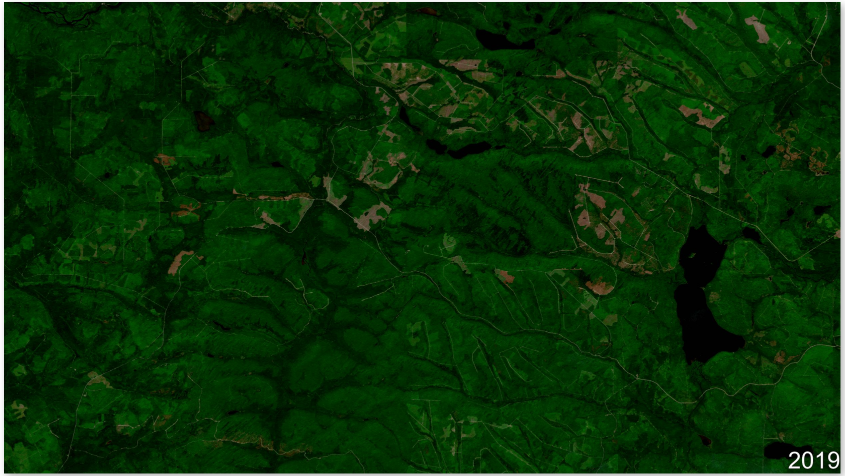
Wetlands

Golf Courses

Built

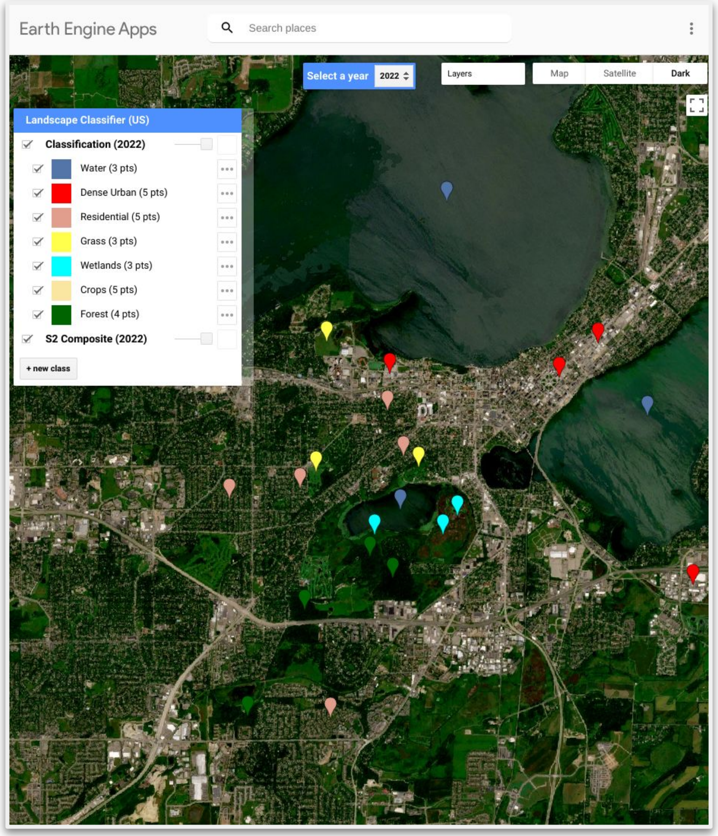
EFM used to proxy
USDA NASS
Cropland Data Layer

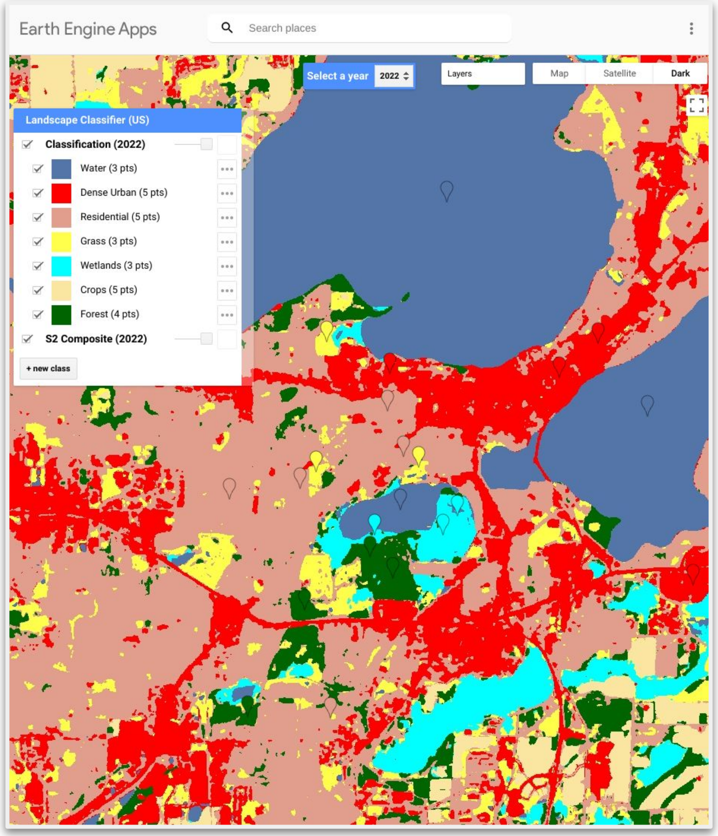
- Corn
- Soy
- Sweet Corn
- Dry Beans



Active forestry
Maine, USA

Change detection
Angle between
embedding vectors



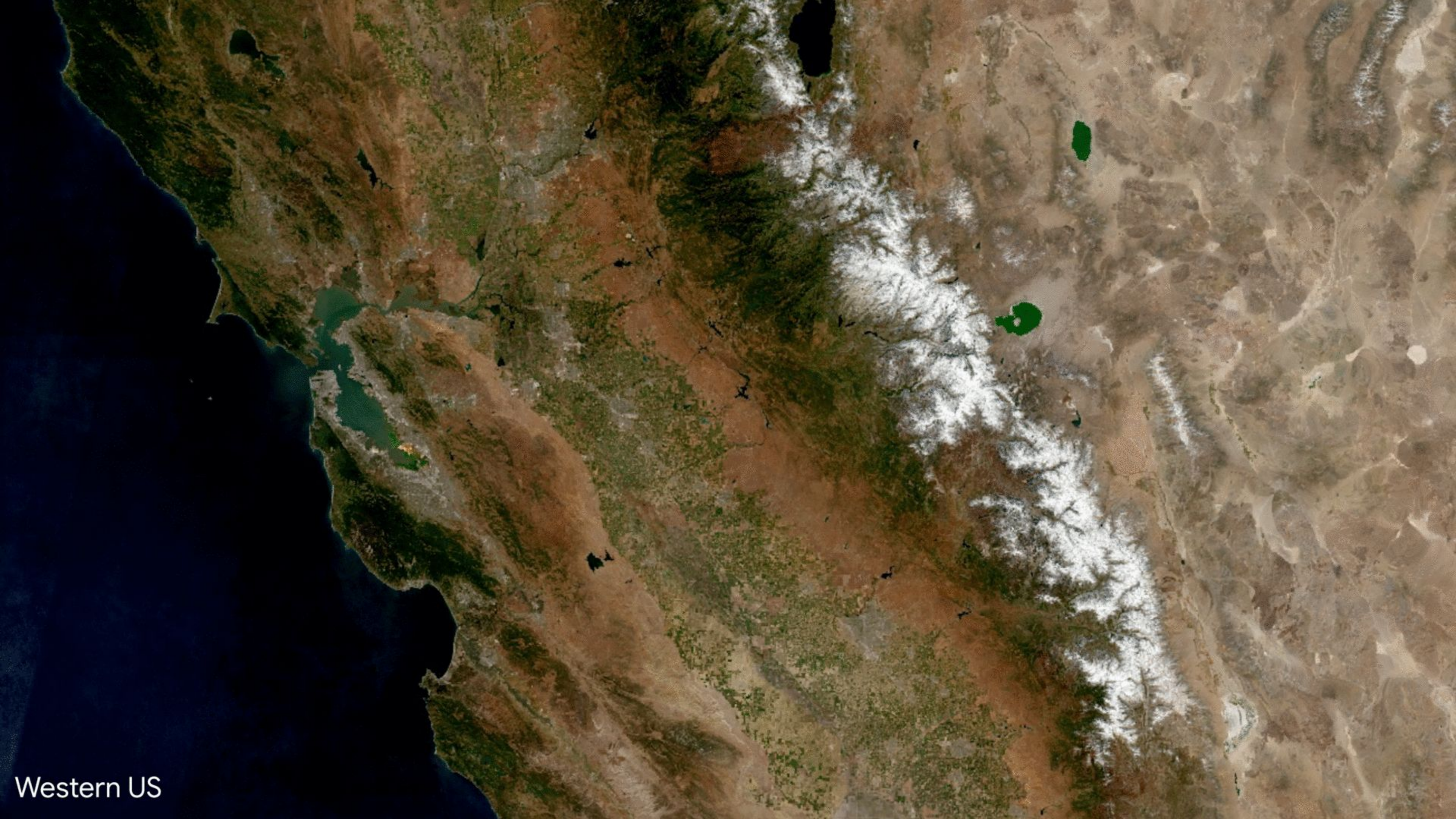


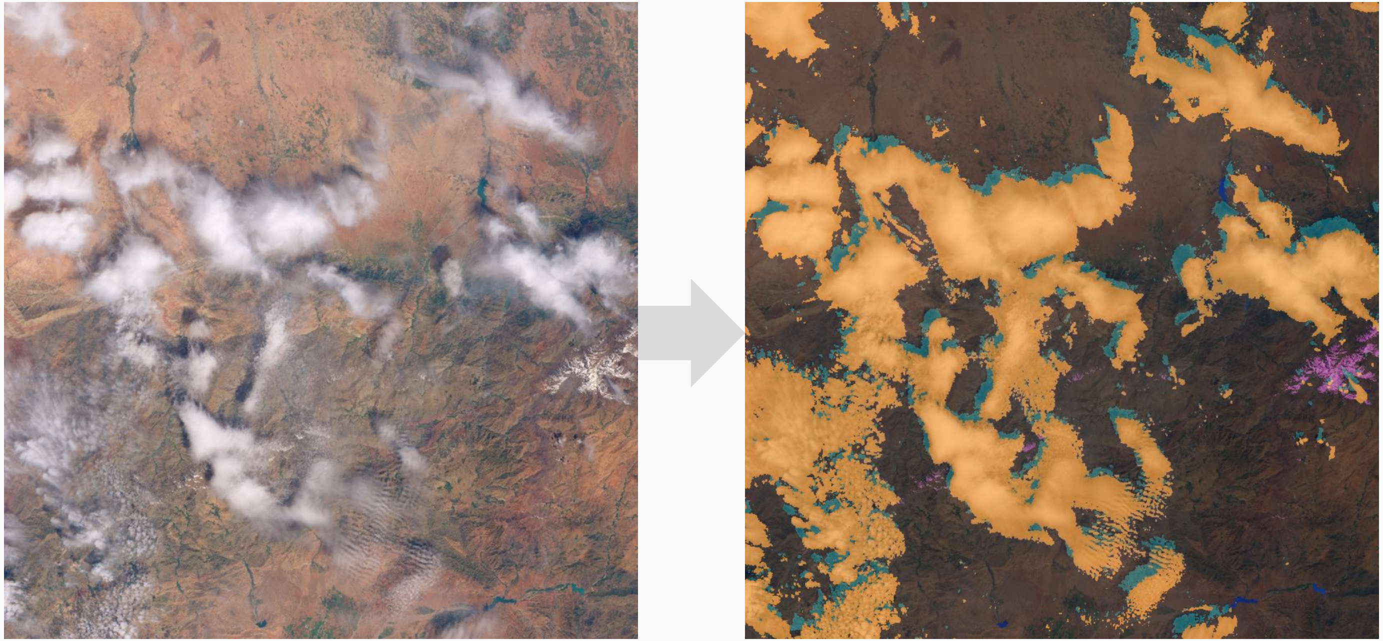
Come check out our **interactive** low-shot **classification demo!**

Google

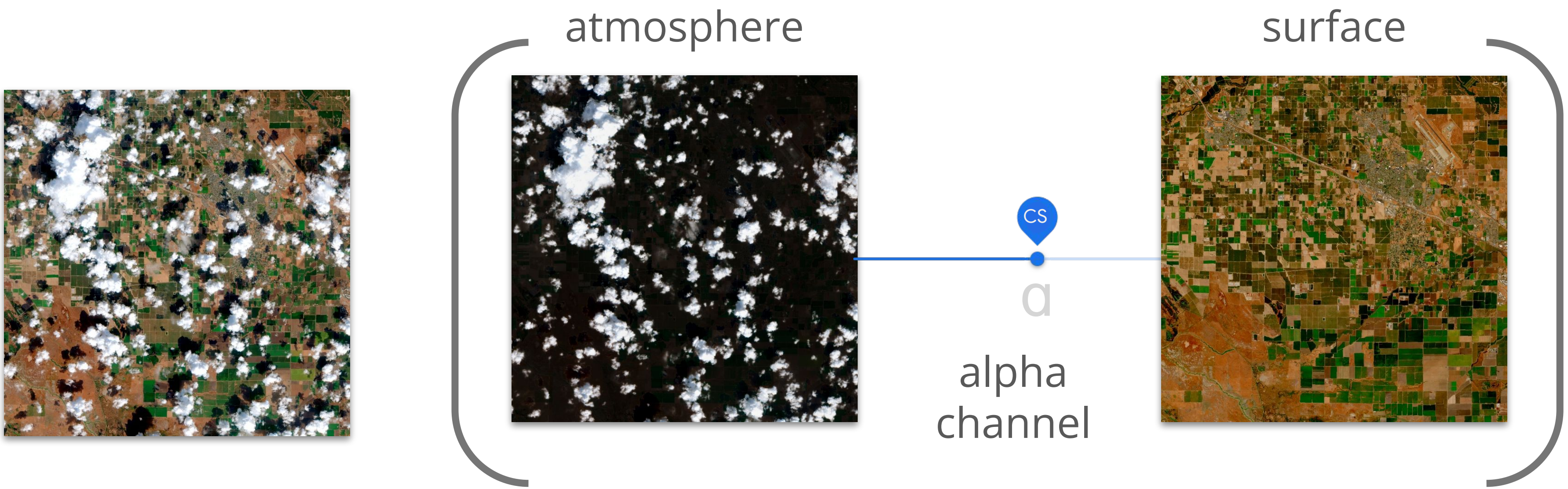
Thank you!

Google


Western US



Cloud and cloud shadow detection typically treated as a
classification problem
(semantic segmentation)



Factor image into components

`linkCollection()`

One-line joins for `ee.Image` or `ee.ImageCollection`

linkCollection(imageCollection, *linkedBands*, *linkedProperties*, *matchPropertyName*)

Links images in this collection to matching images from `imageCollection`.

For each source image in this collection, any specified bands or metadata will be added to the source image from the matching image found in `imageCollection` If the bands or metadata are already present they will be overwritten. If a matching image is not found, any new or updated bands will be fully masked and any new or updated metadata will be null. The output footprint will be the same as the source image footprint.

A match is determined if the source image and an image in `imageCollection` have a specific equivalent metadata property. If more than one collection image would match, the collection image selected is arbitrary. By default, images are matched on their 'system:index' metadata property.

This linking function is a convenience method for adding bands to a target image based on a specified shared metadata property and is intended to support linking collections that apply different processing/product generation to the same source imagery. For more expressive linking known as 'joining', see https://developers.google.com/earth-engine/guides/joins_intro.

Returns the linked image collection.

Arguments:

- **this:imagecollection (ImageCollection):**
  The ImageCollection instance.
- **imageCollection (ImageCollection):**
  The image collection searched to find matches from this collection.
- ***linkedBands (List<String>, optional):***
  Optional list of band names to add or update from the matching image.
- ***linkedProperties (List<String>, optional):***
  Optional list of metadata properties to add or update from the matching image.
- ***matchPropertyName (String, optional):***
  The metadata property name to use as a match criteria. Defaults to "system:index".

Returns: ImageCollection

CLOSE

```
var s2 = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED');

var csPlus = ee.ImageCollection('GOOGLE/CLOUD_SCORE_PLUS/V1/S2_HARMONIZED');
var csPlusBands = csPlus.first().bandNames();

// Link S2 and CS+ results.
var linkedCollection = s2.linkCollection(csPlus, csPlusBands);

// Function to mask pixels with low CS+ QA scores.
function maskLowQA(image) {
  var qaBand = 'cs';
  var clearThreshold = 0.60;
  var mask = image.select(qaBand).gte(clearThreshold);

  return image.updateMask(mask);
}

// Build a median composite.
var dateStart = '2023-01-01';
var dateEnd = '2024-01-01';

var composite = linkedCollection
    .filterDate(dateStart, dateEnd)
    .map(maskLowQA)
    .median();

var s2Viz = {bands: ['B4', 'B3', 'B2'], min: 0, max: 3000};
Map.addLayer(composite, s2Viz, 'median composite');
```