Multi-Task Observation Using Satellites and Kitchen Sinks (MOSAIKS)

A generalizable and accessible approach to machine learning that democratizes access to global satellite imagery

Esther Rolf, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht & Solomon Hsiang

Motivation

There are over **700 earth observation** (EO) satellites in orbit around the earth generating massive amounts of imagery that—if used properly—can help inform and appraise policy-making around the world. Academic research has already shown the value of using satellite images with machine learning (SIML) to better understand **forest cover**, **land use**, **poverty rates**, and **population density**, among other observed outcomes. SIML has vast potential to improve the quality and frequency of information used by decisionmakers, particularly in low- and middle-income countries (LMICs) where accurate and timely data is often difficult, if not impossible, to come by.

Lack of technical capacity has posed a significant obstacle to the uptake of these innovations in LMIC contexts. Machine learning models used to derive useful insights from satellite imagery require domain-specific knowledge, technical expertise, and large computational resources. Accordingly, SIML-based tools are largely inaccessible to researchers and practitioners in the LMIC environments where imagery could have the largest potential for positive impact.

The Tool

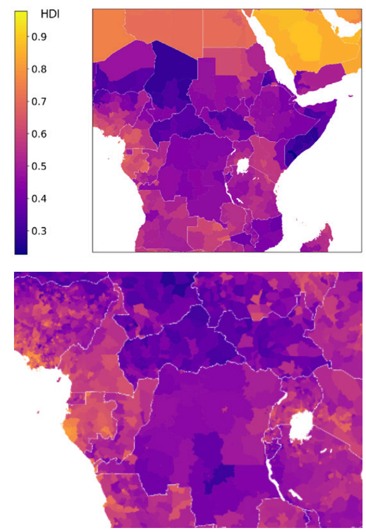
To meet this need, a team of researchers at UC Santa Barbara and UC Berkeley have developed a simple tool called **MOSAIKS** (short for **"Multi-Task Observation Using Satellites and Kitchen Sinks"**) that transforms satellite imagery from almost anywhere on earth into summary information that can be used to predict a wide range of different outcomes ("tasks") in a matter of minutes using a standard laptop computer. This means users can leverage imagery information to solve a variety of customized measurement challenges using basic statistics and a laptop, without ever needing to store or directly manipulate raw imagery, which requires significant computational resources.

MOSAIKS is an important departure from earlier SIML work that created highly specialized tools for transforming satellite imagery into single task estimates. In contrast, MOSAIKS makes SIML much more computationally efficient by collecting and housing satellite imagery on servers in the United States and employing an automated, centralized process to extract "features"— summary pieces of information indicating color, texture, and spatial structure. These features are then put into a regression model to predict the outcome of interest: the task.

This approach performs similarly to more computationally intense, decentralized, and technically demanding approaches that use convolutional neural networks (CNNs). It is also flexible enough to incorporate other sources of georeferenced information, like other imagery data or household survey data.

MOSAIKS + HDI

Application of MOSAIKS to improve the resolution of Human Development Index data in Africa. Above: "state" level HDI estimates. Below: MOSAIKS estimates of HDI at the "county" level.



Source: mosaiks.org/hdi

Potential Applications of MOSAIKS

Food security: crop mapping and agricultural production prediction

Housing: housing stock estimation and affordable housing development

Illicit crops: detecting illicit crop cultivation

Environmental protection: monitoring to mitigate deforestation, informal mining, and other resource extraction activities

MOSAIKS can be used to increase the spatial coverage of existing survey data or to increase the spatial resolution of existing aggregate administrative data.

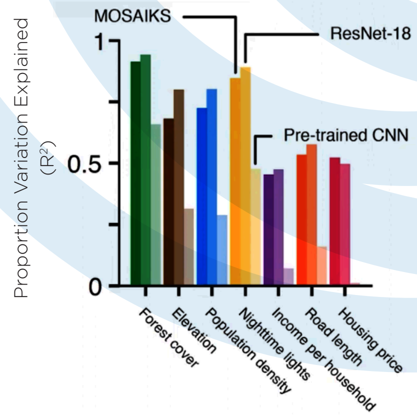
Impact

Deploying MOSAIKS for decisionmakers In 2022, the MOSAIKS team demonstrated how to apply the tool to key research questions for the Bureau for Development, Democracy, and Innovation (DDI) at USAID. Over the following months, Tamma Carleton's team provided direct support to DDI on specific research questions of interest related to resilience programs in the Sahel. Today, USAID's Development Impact Lab is supporting the development of a public API that will further dissemination of MOSAIKS by enabling government partners to test and train with the tool.

Improving future policy evaluations In departing from earlier developments in satellite image machine learning, MOSAIKS advanced the field as a user-friendly, computationally-efficient application that performs as accurately as its predecessors. In March 2023, Solomon Hsiang and his team at the Global Policy Lab put the approach to use, developing the first estimates of the UN Human Development Index for municipalities and counties. When coupled with quasi-experimental approaches, this data will improve future impact evaluations of policies and programs.

MOSAIKS Performance

With just 7.5 minutes of training, MOSAIKS predicts outcomes comparably to neural networks requiring hours of training and use of cloud computing.



Source: mosaiks.org/tutorials

Using MOSAIKS

Data Formatting

Users **download pre-computed imagery "features" from MOSAIKS** in areas where ground truth labels are available. Labels are user-collected and geocoded data that contain information about the desired outcomes to predict.

Users **combine features and ground truth labels** using a spatial merge dependent on longitude and latitude.

Model Training

Using a single set of features, users **apply a simple linear ridge regression technique**, which automatically identifies the right set of weights to assign to each of these features in order to best predict the labeled data.

Evaluate & Apply

Users **evaluate how well the model predicts** the task they assigned it using different metrics.

Users **apply the trained model** to high-resolution, globally-comprehensive satellite imagery data to predict the outcome in locations where ground-truth label data is unavailable.

CEGA
Center for Effective Global Action

cega.berkeley.edu

go.cega.org/X
go.cega.org/linkedin

University of California, Berkeley
207 Giannini Hall
Berkeley, CA 94720-3310
United States

About CEGA

The Center for Effective Global Action (CEGA) is a hub for research, training, and innovation headquartered at the University of California, Berkeley. We generate insights that leaders can use to improve policies, programs, and people's lives. Our academic network includes more than 160 faculty, 65 scholars from low- and middle-income countries, and hundreds of graduate students — from diverse academic disciplines and around the world — that produce rigorous evidence about what works to expand education, health, and economic opportunities for people living in poverty.

Learn more by subscribing to the CEGA newsletter at go.cega.org/newsletter.