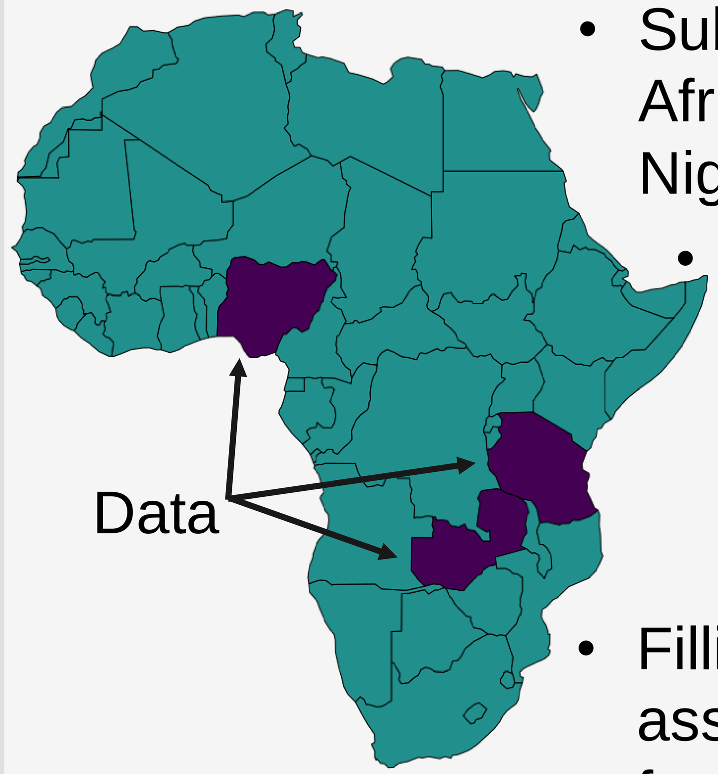
An Open-Source Pipeline for Remote Sensing of Crop Yields Under Environmental Change in Sub-Saharan Africa

Steven Cognac, Juliet Cohen, Grace Lewin, Cullen Molitor
Advisor: Dr. Tamma Carleton

UC SANTA BARBARA
Bren School of Environmental Science & Management

Expanding accessibility to satellite imagery and machine learning methods to fill environmental data gaps

Data Limitations in Africa



- Subnational crop yield data in Africa is limited to three countries: Nigeria, Zambia, and Tanzania
  - This data gap presents a challenge for researchers and decision-makers who need accurate data to study this region
- Filling this data gap will aid in assessing current and future food security risks
- Having these predictions will allow scientists to study how crop yields respond to changing regional environmental variables such as temperature, precipitation, and in turn, global climate change

MOSAIKS Approach

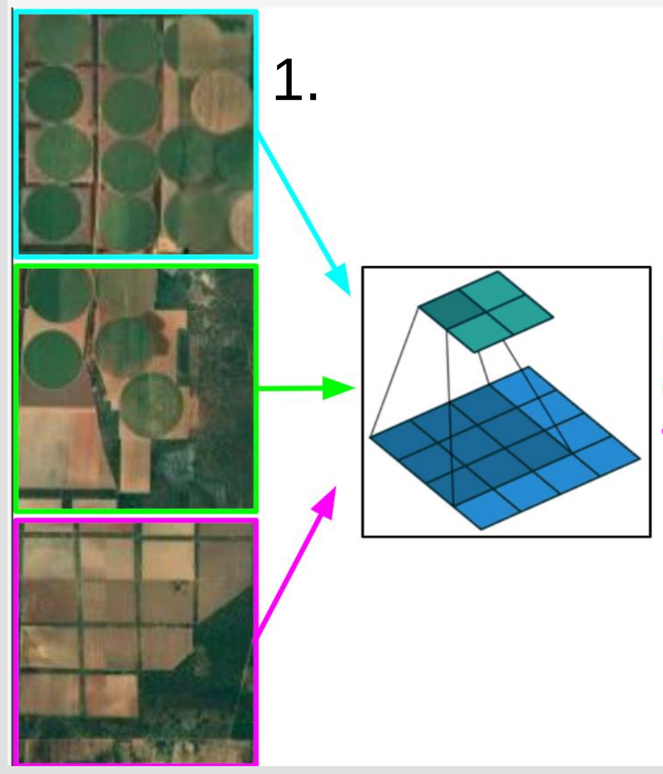
Satellite imagery with machine learning (SIML) methods have been shown to make accurate predictions for a given task. We are using the new MOSAIKS approach, which stands for **M**ulti-task **O**bservation using **Sa**tellite **I**magery & **K**itchen **S**inks. This uses SIML methods to create tabular features encoded with satellite information that are task-agnostic. This means they can be used to predict any task of interest such as population, forest cover, income, or crop yield.

Filling the Data Gap With MOSAIKS

We use the MOSAIKS approach to create features for 1984–2021 and use those features to model crop yield. This model uses data from three countries: Zambia, Tanzania, and Nigeria. The model will then be applied to make predictions for the rest of sub-Saharan Africa.

1. Take satellite images at 1 km^2 resolution
2. Compute random convolutional features
3. Collapse features to district boundaries and merge with crop yield data
4. Use features to model crop yield and make predictions

1.



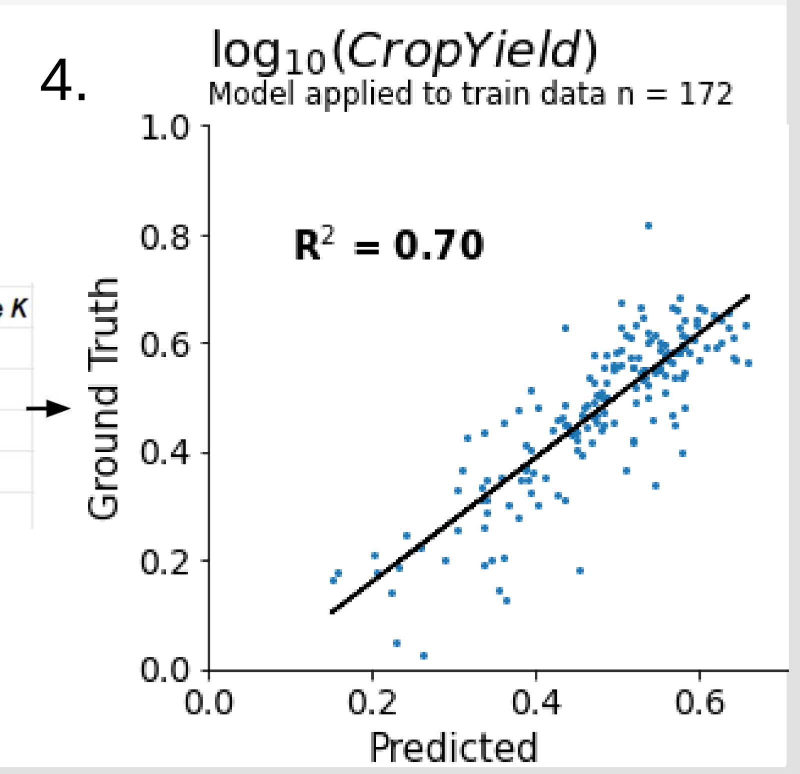
2.

Image	Geometry	Feature 1	...	Feature *K*
1	lon, lat	4	...	12
2	lon, lat	2	...	11
3	lon, lat	3	...	6
...	...	...	...	...
N	lon, lat	10	...	2

3.

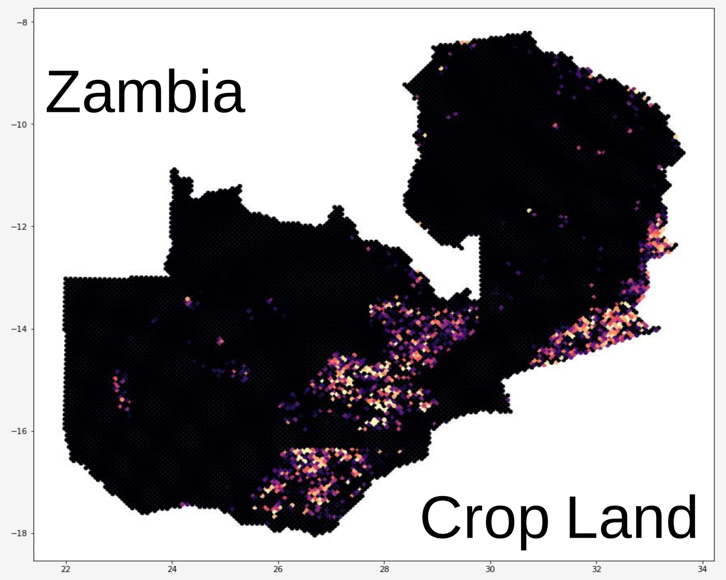
District	Crop Yield	Feature 1	...	Feature *K*
Chibombo	1520	4	...	12
Kabwe	1878	2	...	11
Mkushi	2078	3	...	6
...	...	...	...	...
Kitwe	2383	10	...	2

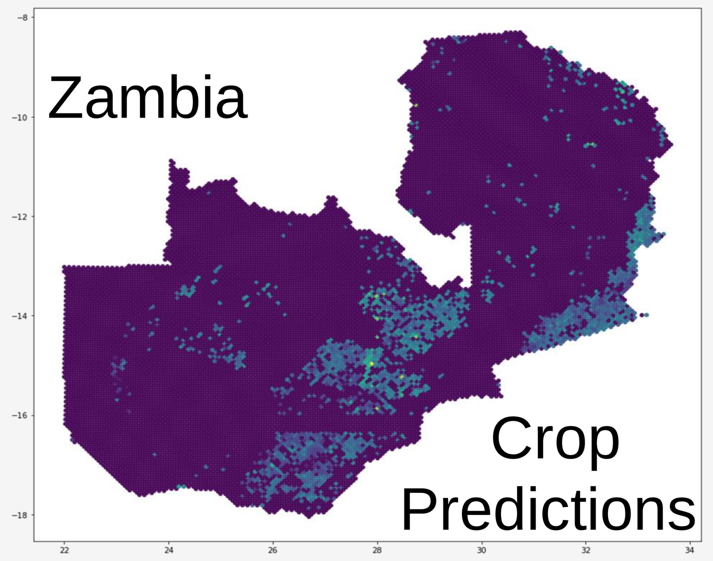
4.



Preliminary Results

- For the country of Zambia, our model predicts crop yields with a very promising R^2 value of 0.70
  - The top right a plot shows the crop land area of Zambia. The bottom right plot shows the predicted crop yields at 1 km^2 resolution
  - This model uses cross validated ridge regression with 172 training points
    - The model is tested on the remaining 44 data points
  - The model is applied to the full set of features to make predictions at finer resolution than original crop data





An Open-access Tool for Future Applications

- Our team is contributing task-agnostic features created from publicly available satellite imagery to the MOSAIKS API
  - These features can be downloaded by global users to predict domain-specific tasks (forest cover, population, income, crop yield, etc.)
  - Executing linear regression on this joined dataset yields high-resolution predictions at low computational costs
- Open-source code examples of the linear modeling process will be available
- The MOSAIKS API database will be continually updated as imagery from more satellites and for more years are processed
- Predictions made using these features can be used to fill important gaps in many environmental datasets going back to the beginning of the satellite record

Why is Increasing Access Important?

The MOSAIKS system and its resulting products can provide policy-makers, researchers, and users in sub-Saharan Africa with a tool to address environmental questions in less time with fewer resources. Innovative remote sensing tools and machine learning can fill in the current gaps in data for many regions of the world. By filling in these data gaps, decision-makers can study how a region might be affected by global climate change and develop mitigation efforts accordingly.

Acknowledgments

Thank you to our advisors Dr. Tamma Carleton, Dr. Allison Horst, and Dr. Jonathan Proctor, the MOSAIKS team, and the Microsoft Planetary Computer team for their support and contributions to this project.

Literature Cited

Rolf, Esther, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht, and Solomon Hsiang. "A Generalizable and Accessible Approach to Machine Learning with Global Satellite Imagery." *Nature Communications* 12, no. 1 (December 2021): 4392. https://doi.org/10.1038/s41467-021-24638-z.