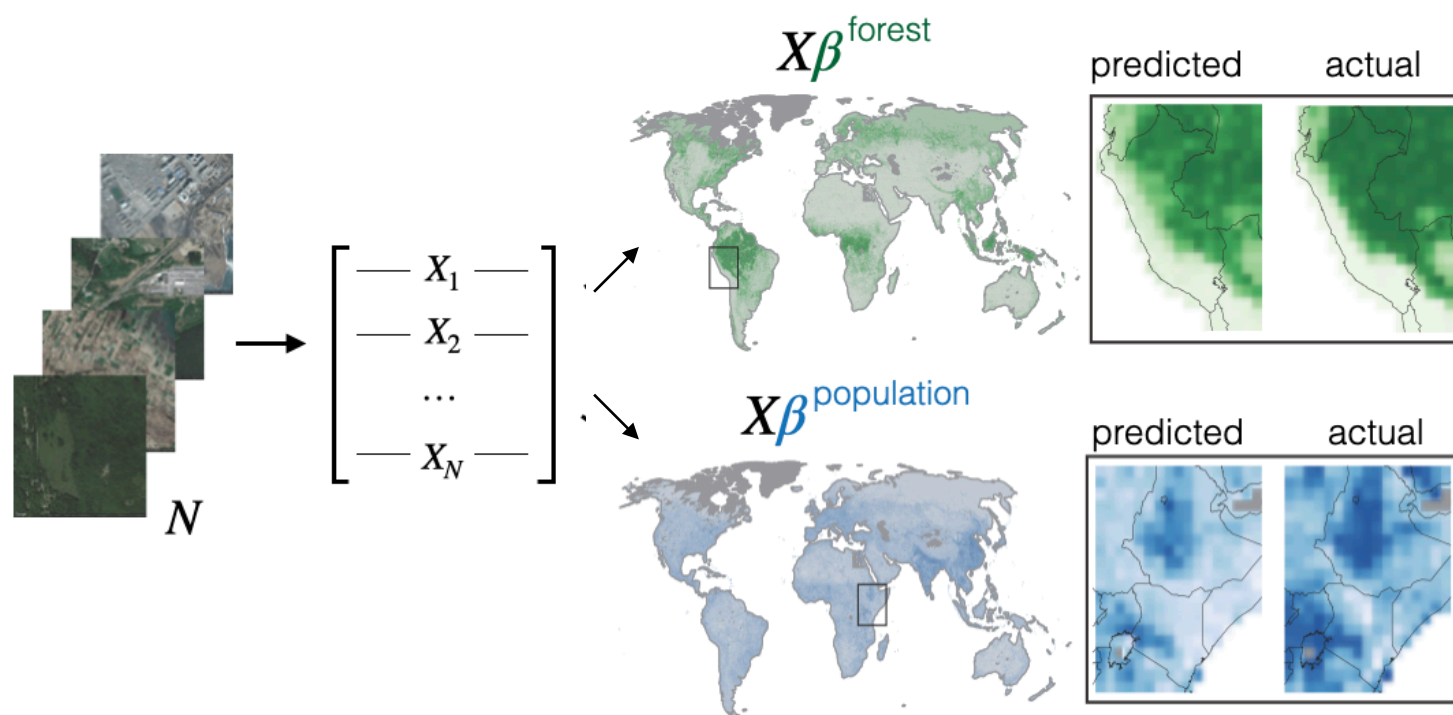


MOSAIKS:

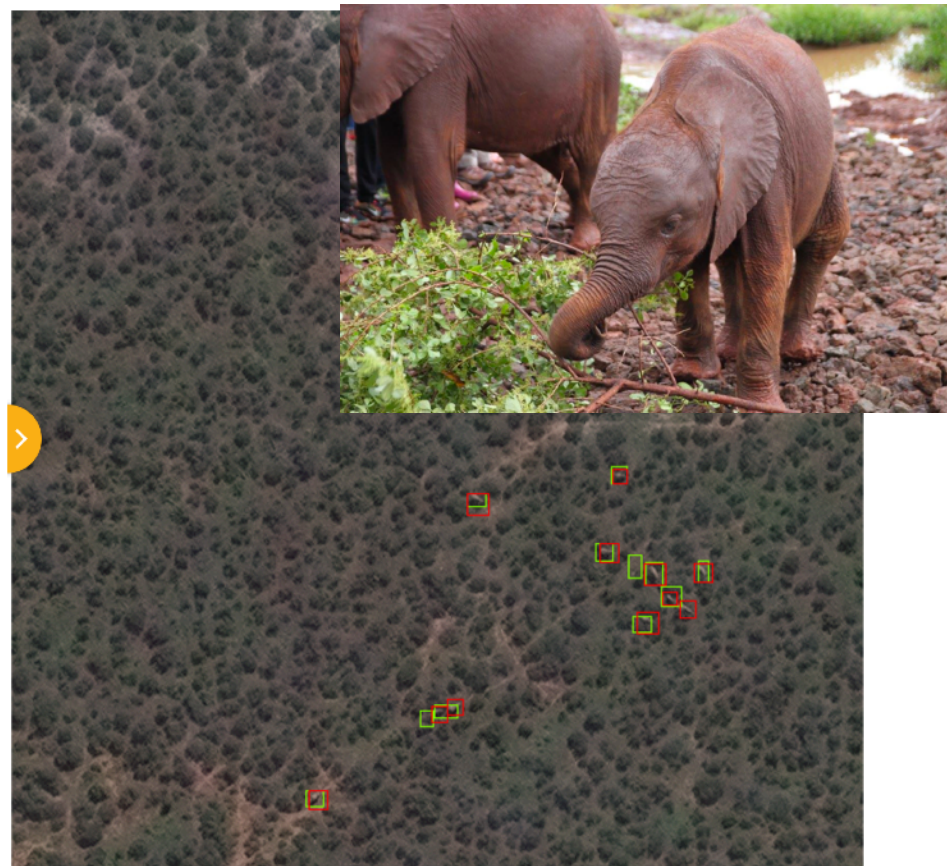
A generalizable and accessible approach to machine learning with global satellite imagery



...and how to use it!

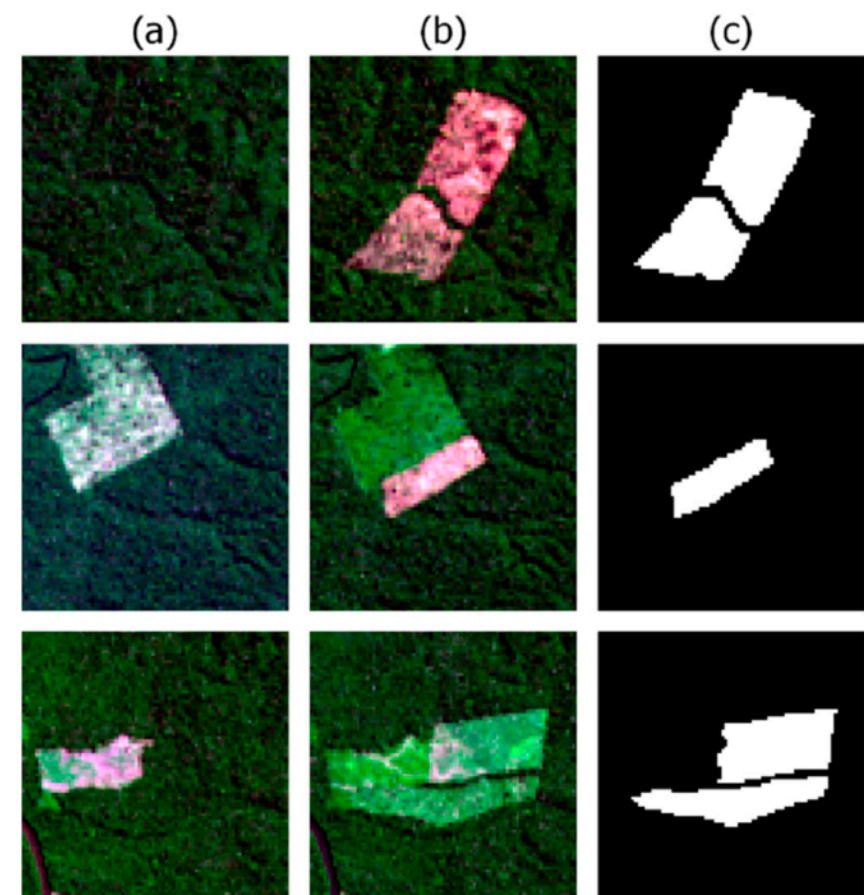
Combining satellite imagery and machine learning (SIML) has **the potential** help monitor our world across multiple domains:

Wildlife



Duporge et al 2020

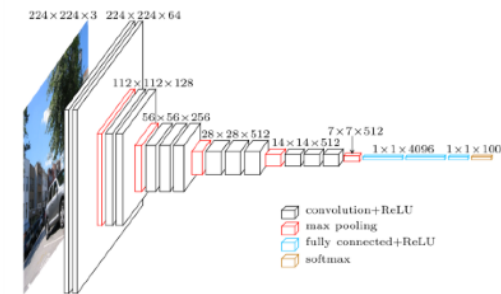
Deforestation



de Bem et al 2020

Currently, transforming satellite imagery into relevant statistics is:

- costly (computation and expertise)
- domain-specific



e.g. ResNet-18

Currently, transforming satellite imagery into relevant statistics is:

- costly (computation and expertise)
- domain-specific

However, in SIML problems, the same images summarize multiple outcomes:



tree cover : 10%

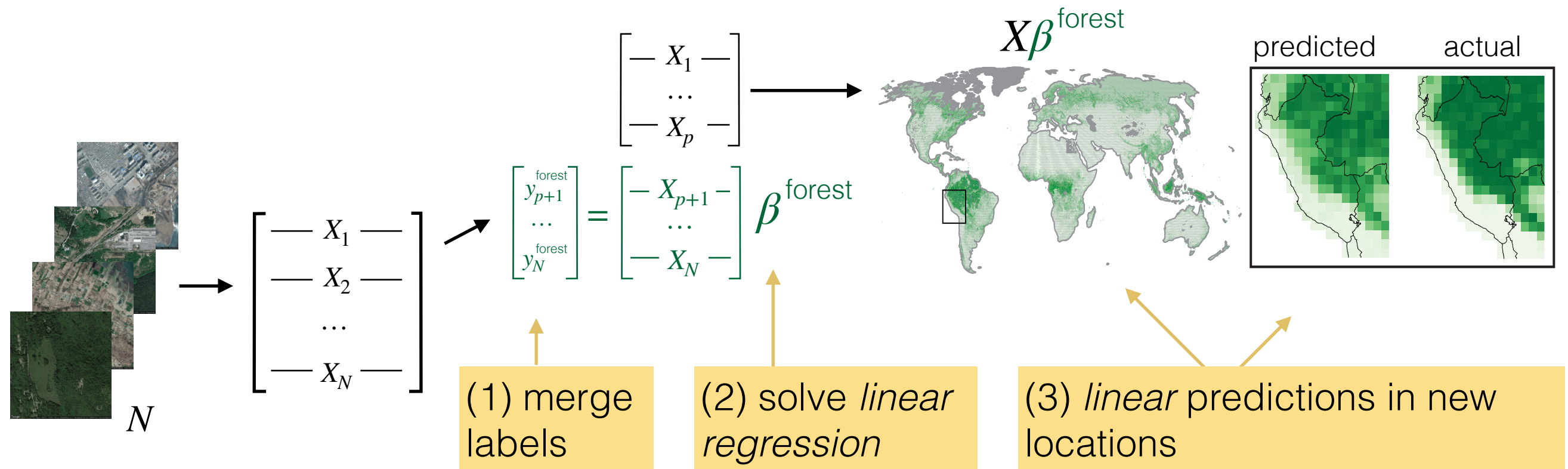
population : 264 people

avg. income : \$67,000

road length : 2.7km

MOSAIKS:

Multi-task **O**bservation with **S**atellite **I**magery and random **K**itchen **S**inks



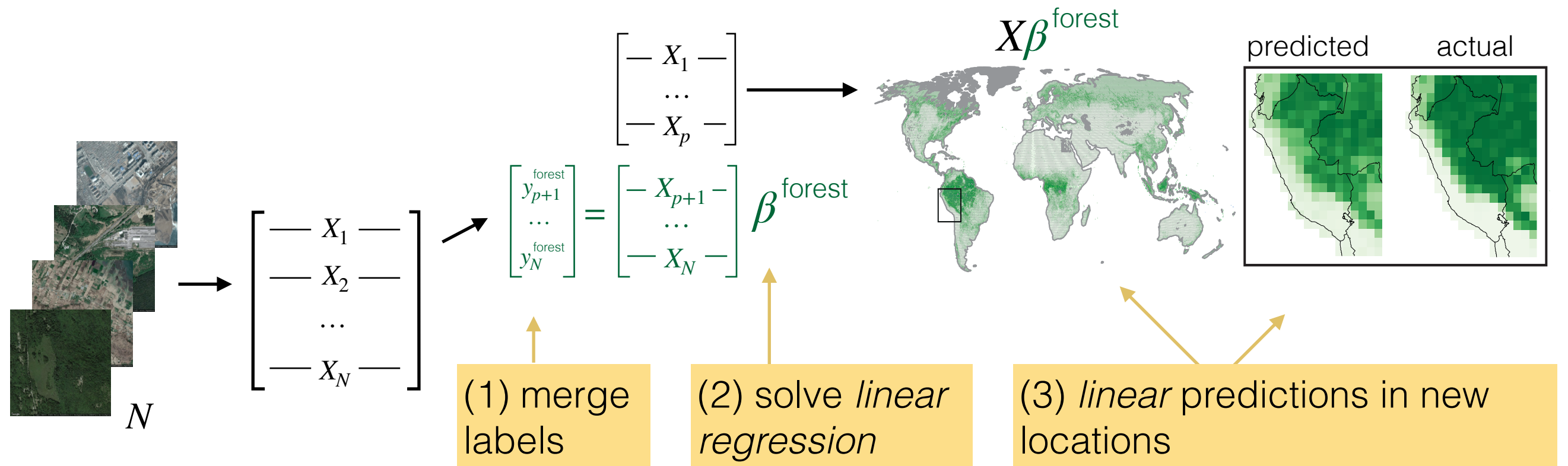
Accessibility

Generalizability

Simplicity

MOSAICS:

Multi-task **O**bservation with **S**atellite **I**magery and random **K**itchen **S**inks



Accessibility

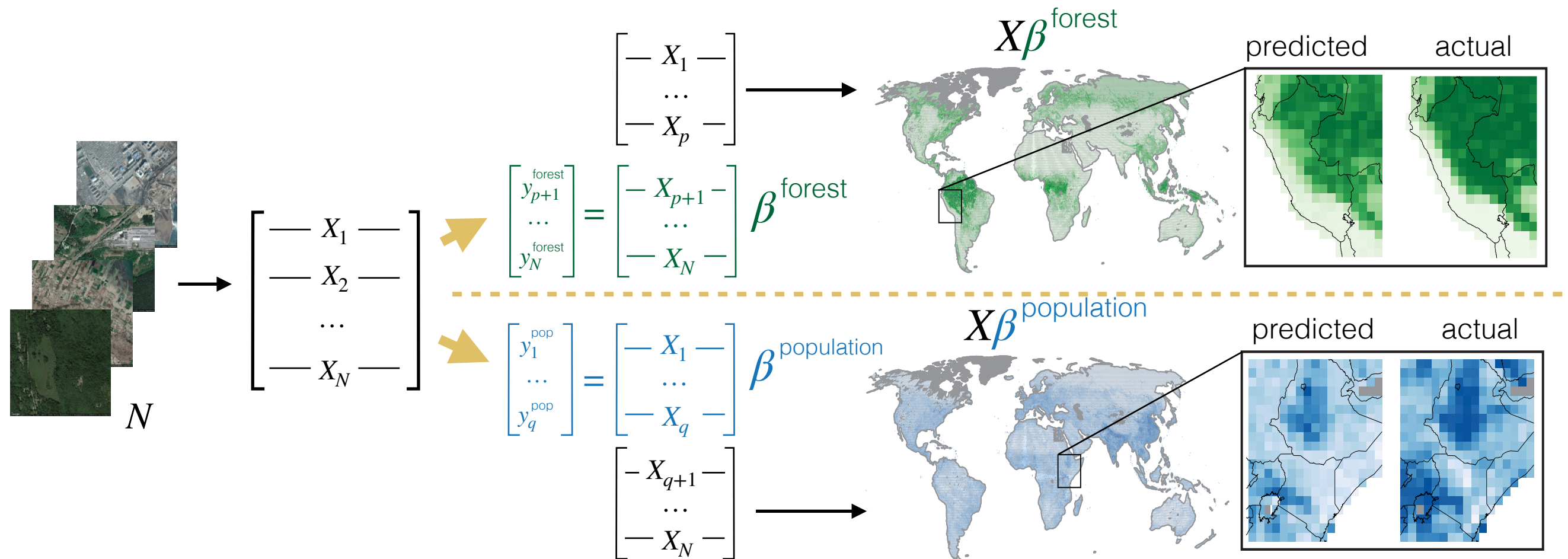
Generalizability

Simplicity

The prediction step should be **simple to use** and **computationally efficient** for users.

MOSAICS:

Multi-task **O**bservation with **SA**atellite **I**magery and random **K**itchen **S**inks



Accessibility

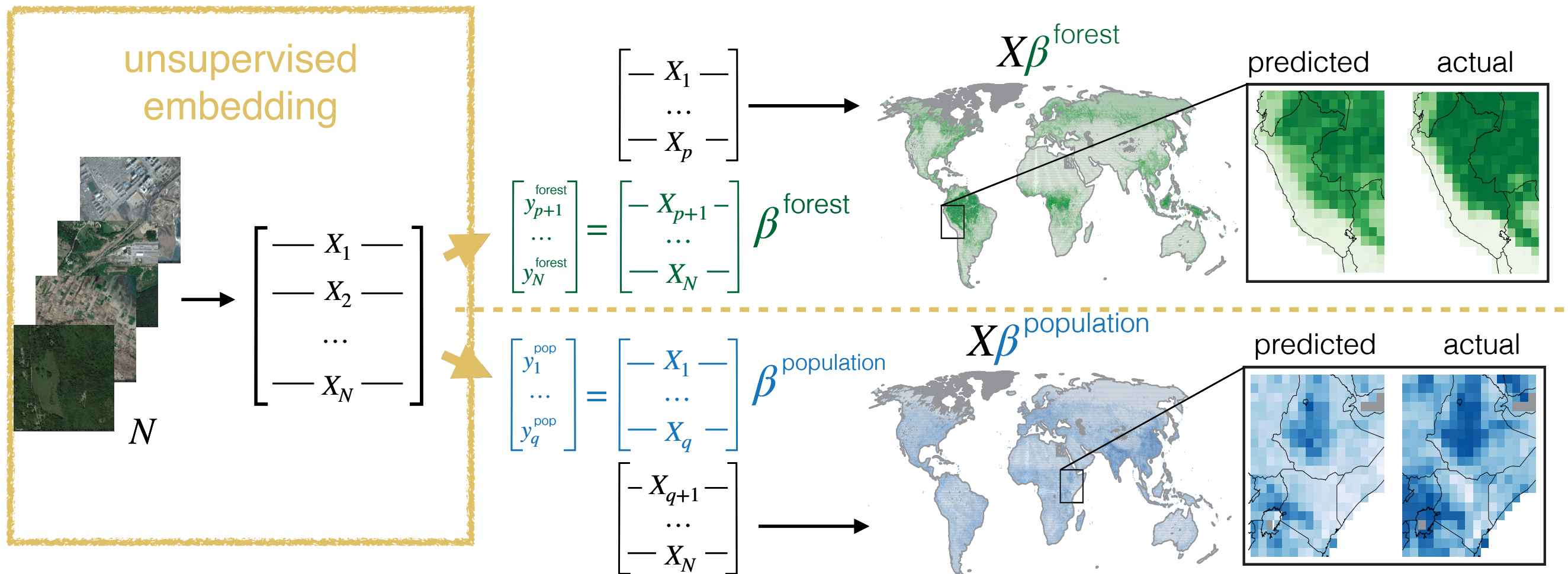
Generalizability

Simplicity

The same system should be useful for many prediction **tasks** using a common source of satellite imagery

MOSAICS:

Multi-task Observation with **S**atellite Imagery and random **K**itchen **S**inks



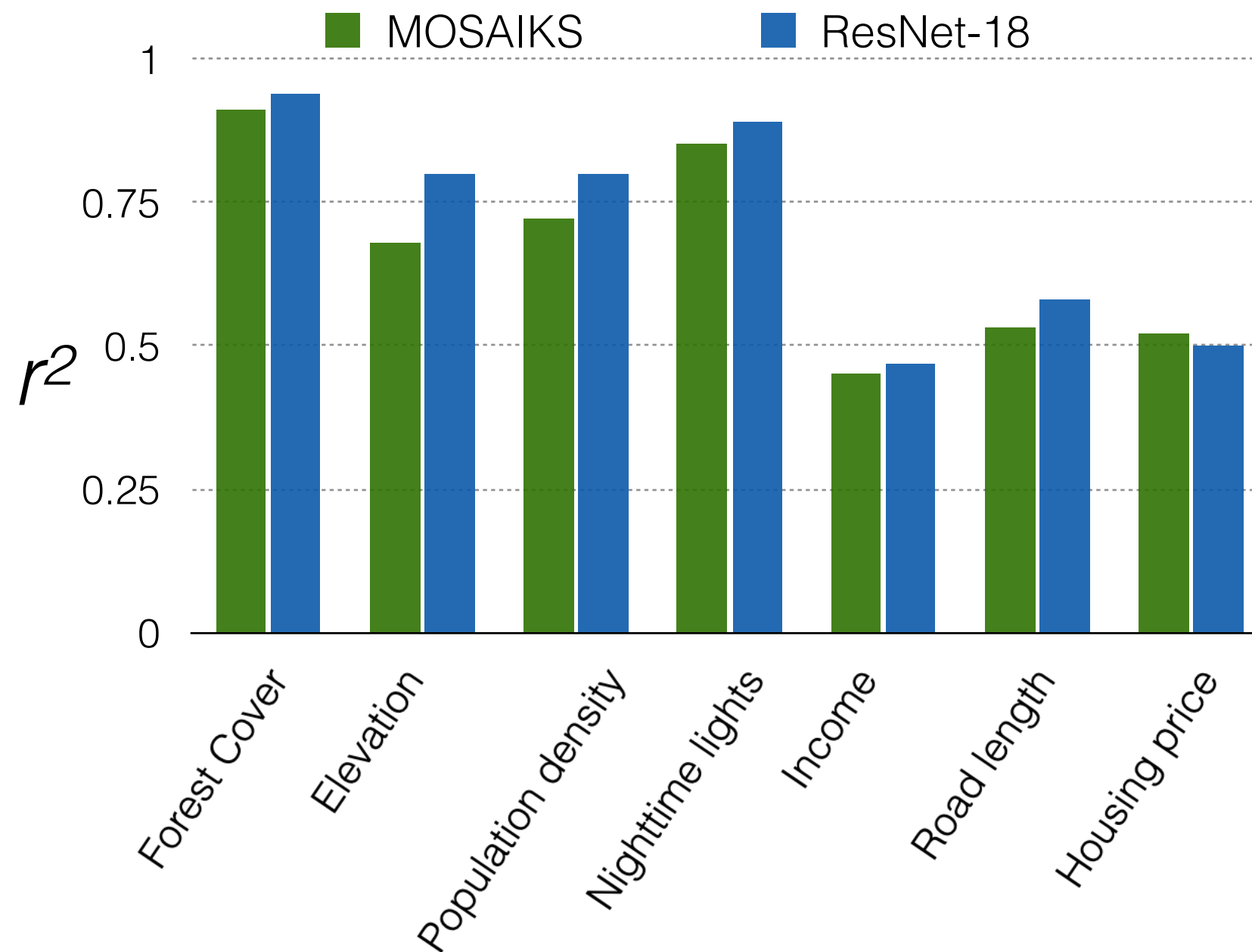
Accessibility

Generalizability

Simplicity

The **algorithms** behind the prediction system should be as simple as possible.

MOSAICS compared to fine-tuned ResNet-18 (within US)



Training times:

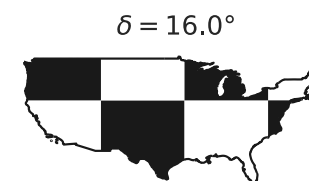
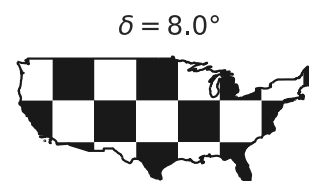
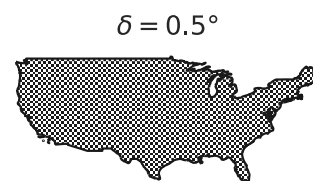
MOSAICS: 2 minutes
(Laptop)

Fully trained ResNet-18: 7.9 hours
(AWS EC2 p3.xlarge, Tesla V100 GPU)

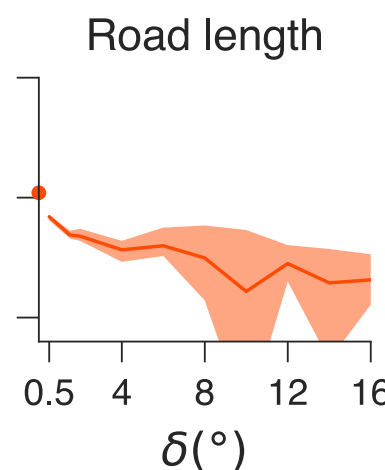
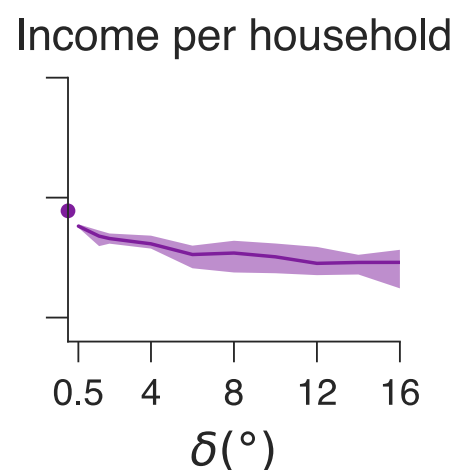
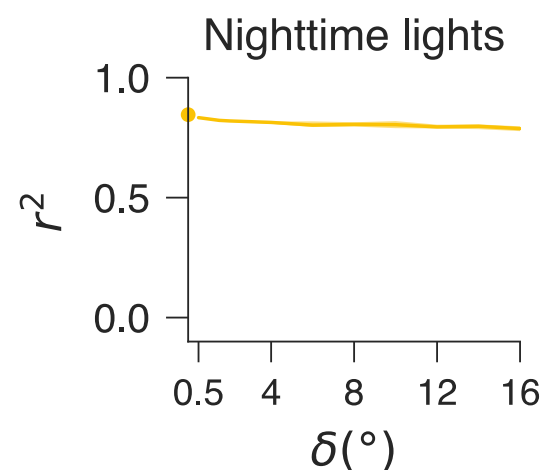
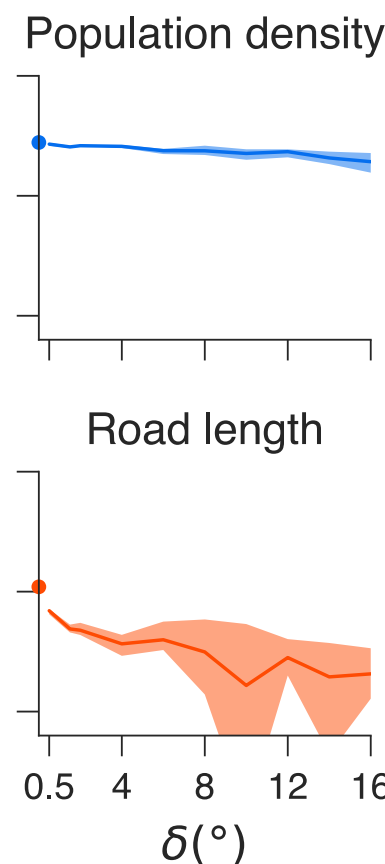
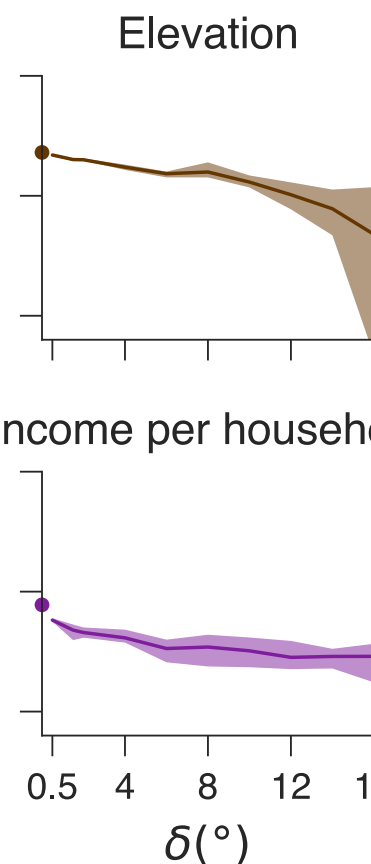
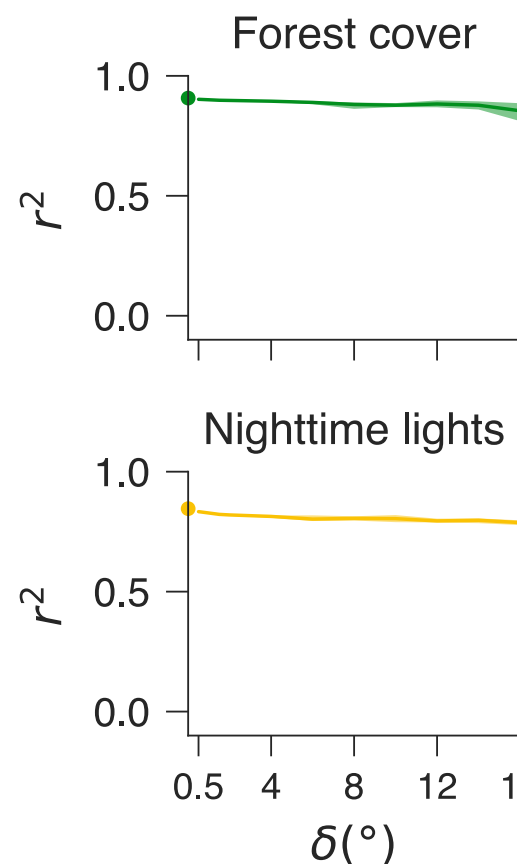
Fast retraining of MOSAIKS → simplifies context-aware sensitivity analyses.

Example: as degree of spatial extrapolation (δ) increases, how does performance degrade differently across task domains?

Method: change conditions, retrain



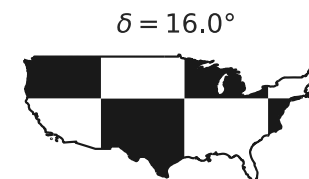
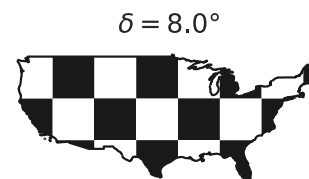
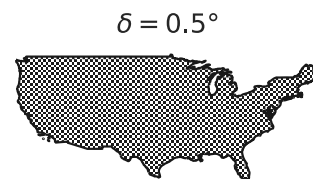
□ Training data
■ Validation data



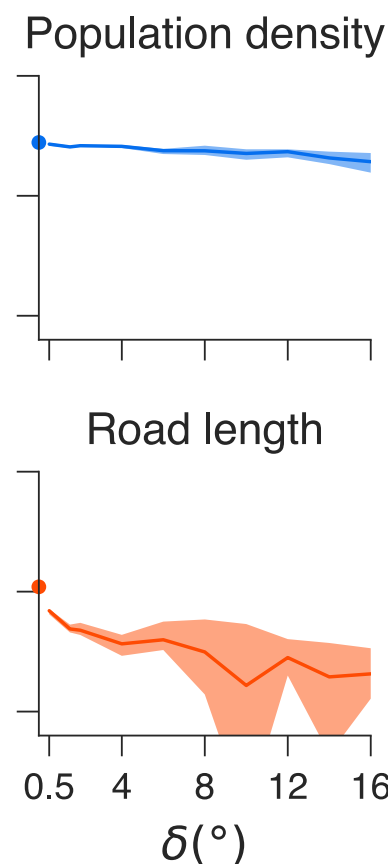
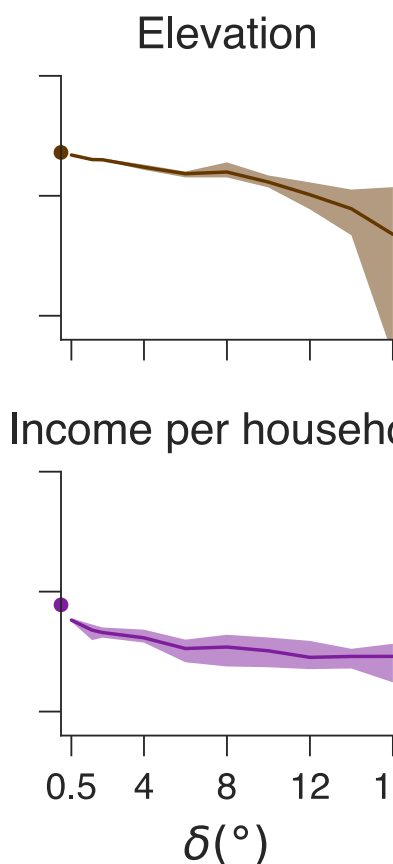
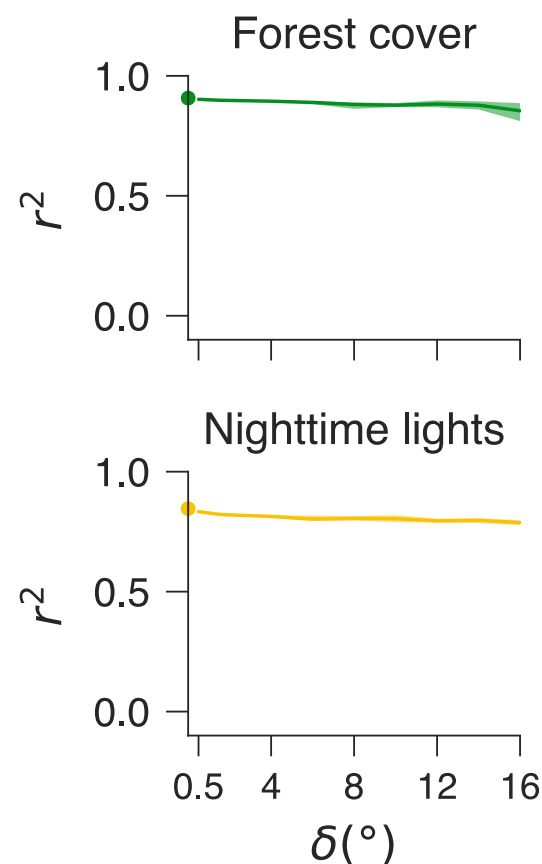
Fast retraining of MOSAIKS → simplifies context-aware sensitivity analyses.

Example: as degree of spatial extrapolation (δ) increases, how does performance degrade differently across task domains?

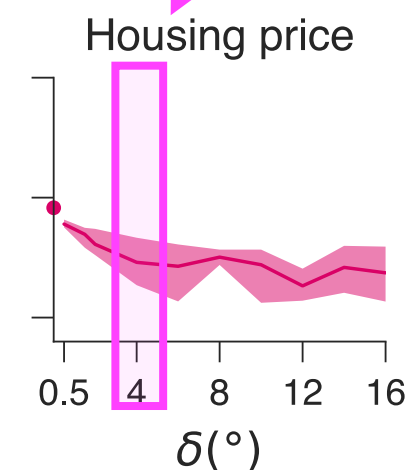
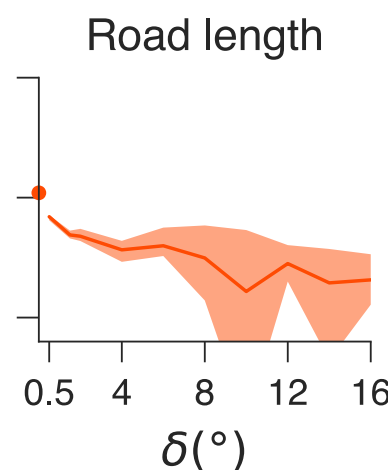
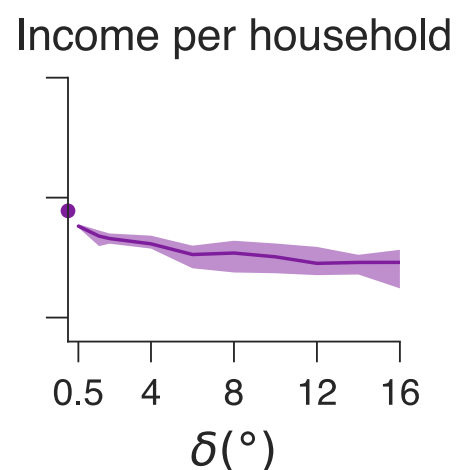
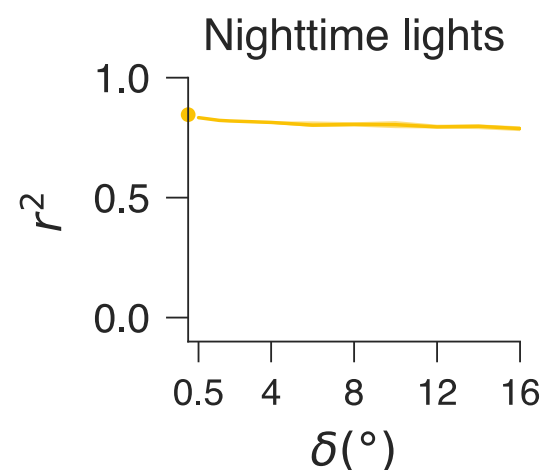
Method: change conditions, retrain



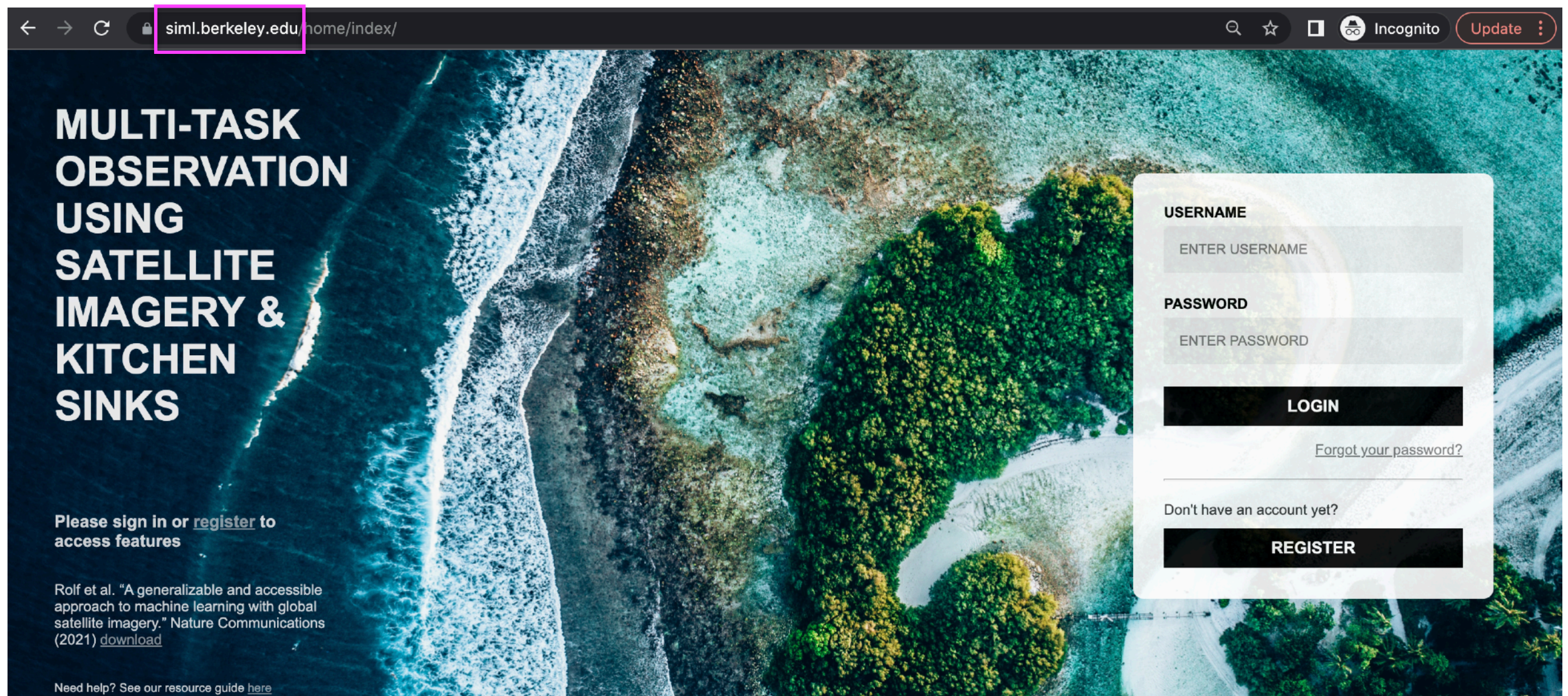
□ Training data
■ Validation data



For each training condition,
~2 min vs. hours



Now available! A public API where users can query for features and run their own scientific analyses:



siml.berkeley.edu

Our team: Tamma Carleton, Trinetta Chong, Hannah Druckenmiller, Eugenio Noda, Jonathan Proctor, Esther Rolf and Solomon Hsiang

Now available! A public API where users can query for features and run their own scientific analyses:

The screenshot shows the SIML Berkeley File Query interface. The browser address bar displays `siml.berkeley.edu/portal/file_query/`. The navigation bar includes the MOSAIKS logo, a BETA badge, and links for Map Query, File Query (highlighted with a pink box), My Files, Resources, and Contact Us. A SIGN OUT button is located on the right. The main content area is titled "3 Identify coordinates".

On the left, a CSV preview for `california_population.csv` is shown. The table has columns for Type, Lon, Lat, and Population. The first 10 rows are displayed, showing various longitude and latitude values for different locations in California.

On the right, the "Select Coordinates Columns" section prompts the user to specify columns for Latitude and Longitude. The Latitude dropdown is set to "lat" with a green "Selection is valid" message. The Longitude dropdown is set to "lon" with a blue "Selection is valid" message. Below this, the "Submit" section contains a message: "When you are satisfied with the selection, you can submit your query." and a blue "Create Feature File" button.

	Lon	Lat	Population
Type	NUMERIC	NUMERIC	NUMERIC
1	-117.095	32.595	9.25743103
2	-116.995	32.595	86.1885910
3	-116.695	32.595	0.17026083
4	-116.595	32.595	0.39001014
5	-116.495	32.595	3.39810228
6	-114.795	32.595	0.41147568
7	-114.695	32.595	0.65670961
8	-114.595	32.595	79.9315490
9	-117.195	32.695	63.8588600
10	-117.095	32.695	4852.96533

N. Records: 3116

siml.berkeley.edu

Now available! A public API where users can query for features and run their own scientific analyses:

MOSAIKS

BETA

Precomputed Files

Map Query

File Query

My Files

Resources

Contact Us

SIGN OUT

			15:18:58		15:52:49			
7	california_treecover.zip	File	2022-07-19 15:49:19	Completed	2022-07-19 15:51:12	242.3MiB	Download	Delete
8	california_population_2.zip	File	2022-07-18 18:45:16	Completed	2022-07-18 18:46:06	104.3MiB	Download	Delete
9	20220718_184341.zip	Map	2022-07-18 18:43:41	Completed	2022-07-18 18:46:47	336.6MiB	Download	Delete
10	california_population.zip	File	2022-07-18 01:50:06	Completed	2022-07-18 01:50:58	104.3MiB	Download	Delete

siml.berkeley.edu