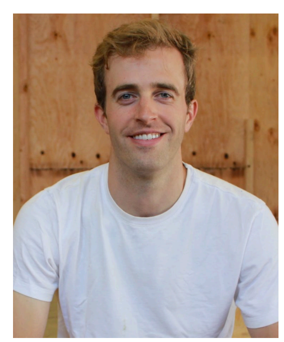
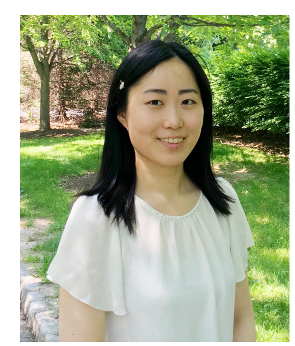
MOSAIKS:

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

<u>Esther Rolf</u>*, Jonathon Proctor*, Tamma Carleton*, Ian Bolliger*, Vaishaal Shankar*, Miyabi Ishihara, Benjamin Recht, Solomon Hsiang

Presented at:
AI + Climate Seminar
November 30, 2021

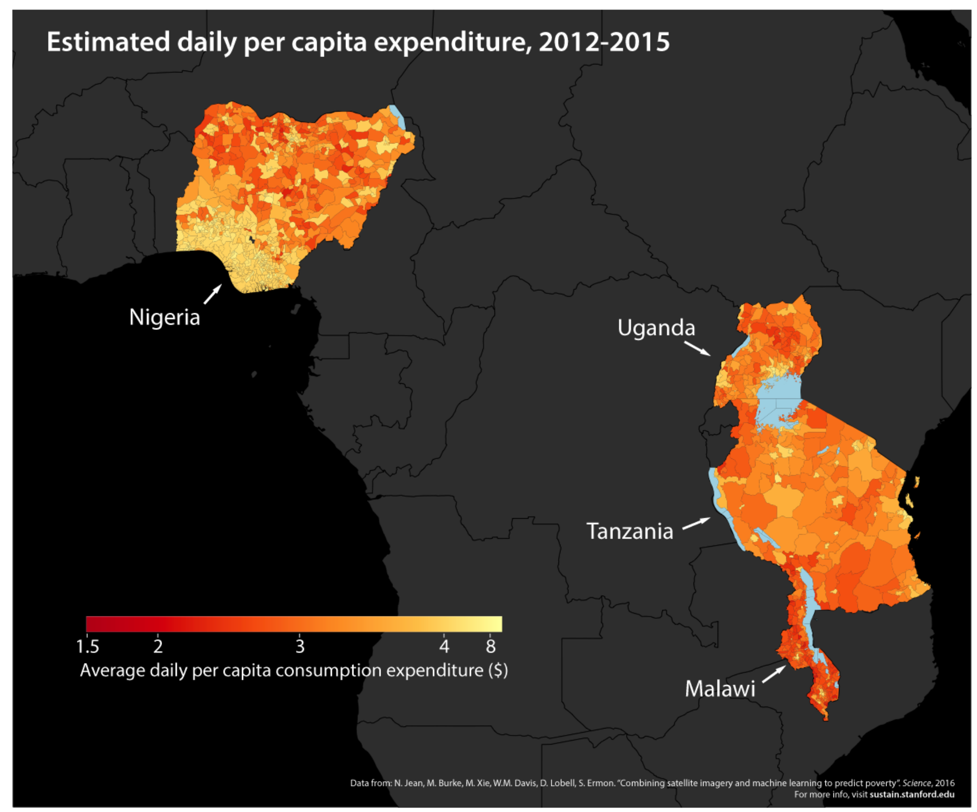
700+ satellites are orbiting the earth, collecting over **90TB** of data daily, enabling researchers to monitor our world across multiple domains:

Esther Rolf (UC Berkeley)

Image Credit: https://landsat.gsfc.nasa.gov/

700+ satellites are orbiting the earth, collecting over **90TB** of data daily, enabling researchers to monitor our world across multiple domains:

Poverty mapping



Jean et al. 2016, Science

SUSTAINABLE DEVELOPMENT GOALS

Esther Rolf (UC Berkeley)

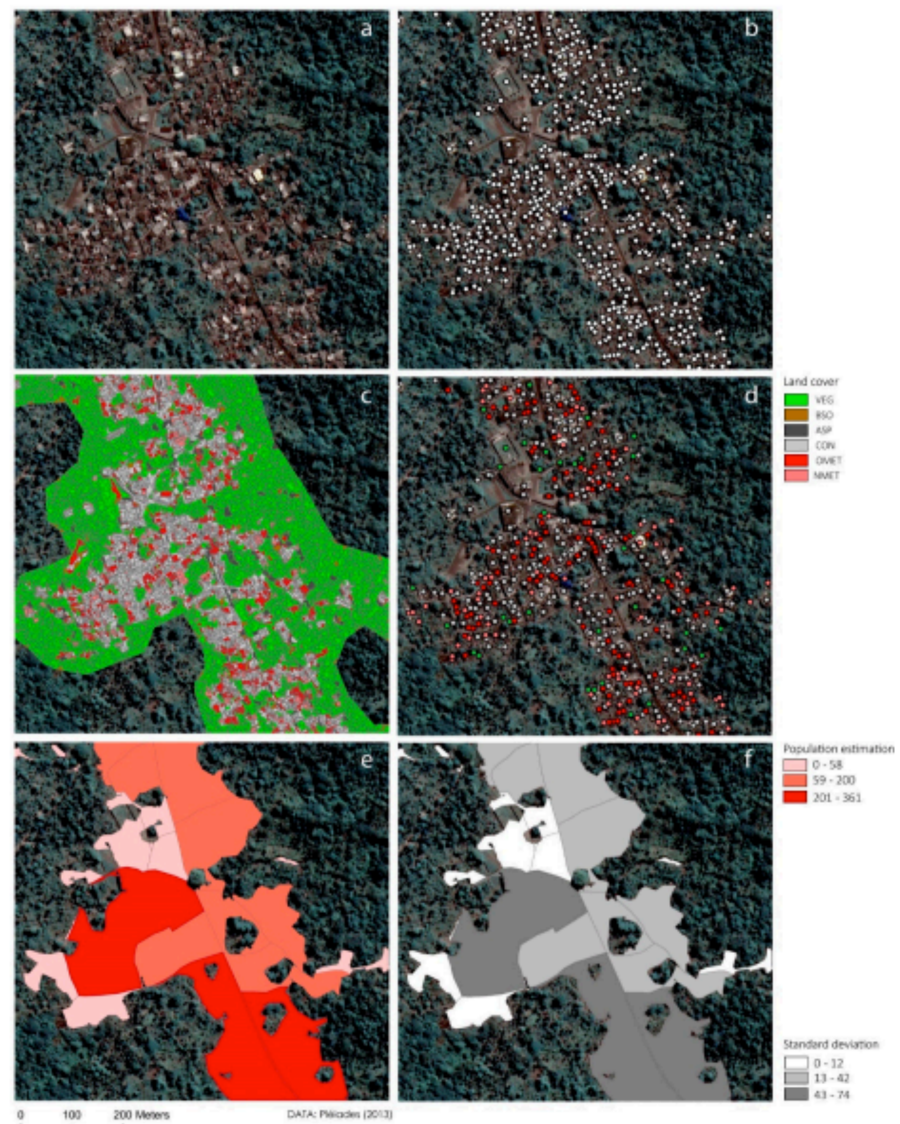
Figure: http://sustain.stanford.edu/predicting-poverty

700+ satellites are orbiting the earth, collecting over **90TB** of data daily, enabling researchers to monitor our world across multiple domains:

1 NO POVERTY

11 SUSTAINABLE CITIES AND COMMUNITIES

Mapping Population Distribution



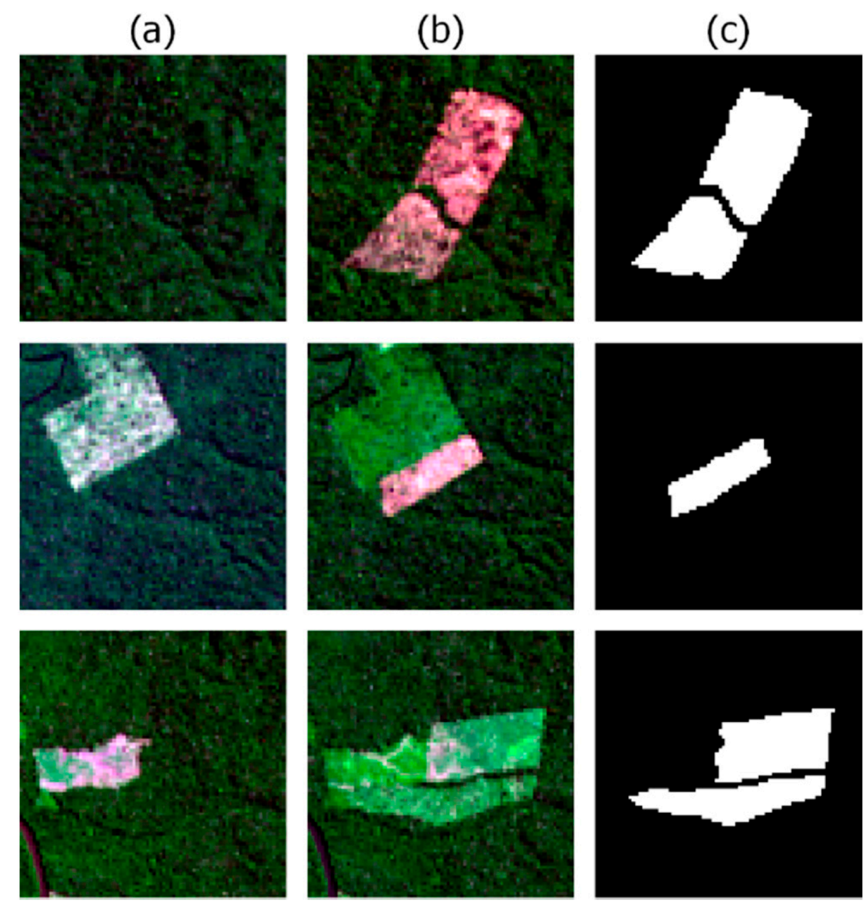
Mossoux et al. 2018

Esther Rolf (UC Berkeley)

Figure from: https://www.mdpi.com/2072-4292/10/9/1409/pdf-vor (Mossoux et al. Paper)

700+ satellites are orbiting the earth, collecting over **90TB** of data daily, enabling researchers to monitor our world across multiple domains:

Monitoring Deforestation



de Bem et al. 2020

SUSTAINABLE DEVELOPMENT GOALS

Esther Rolf (UC Berkeley)

Photos: https://www.mdpi.com/2072-4292/12/6/901 (de Bem et al. Paper)

700+ satellites are orbiting the earth, collecting over **90TB** of data daily, enabling researchers to monitor our world across multiple domains:

Wildlife Detection and Monitoring

Duporge et al. 2020

SUSTAINABLE DEVELOPMENT GOALS

Esther Rolf (UC Berkeley)

Photos: https://blog.maxar.com/earth-intelligence/2021/deep-learning-detects-elephants-in-maxar-satellite-imagery-on-par-with-human-accuracy

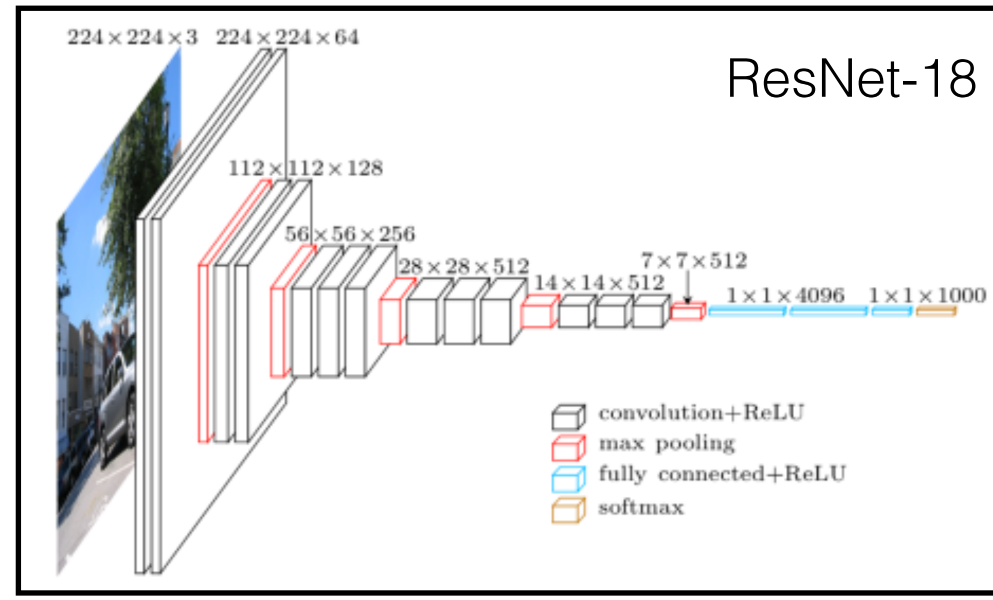
700+ satellites are orbiting the earth, collecting over **90TB** of data daily, <u>enabling researchers to monitor our world</u> across multiple domains:

SUSTAINABLE DEVELOPMENT GOALS

See, e.g. Burke et al. 2021

700+ satellites are orbiting the earth, collecting over **90TB** of data daily, <u>enabling researchers to monitor our world</u> across multiple domains:

However, transforming satellite imagery into relevant statistics is **costly** (computation and expertise) and most solutions are domain-specific.



Our approach: a *<u>general method</u>* that allows researchers to easily predict *<u>any variable</u>* from space.

In this talk:

goal: **learn** functions that take in *satellite images*, and **predict** *variables of interest*

In this talk:

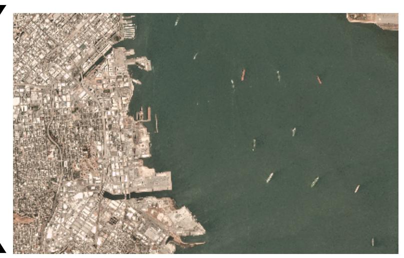
goal: **learn** functions that take in *satellite images*, and **predict** *variables of interest*

f() ≈ 45% forest

In this talk:

goal: **learn** functions that take in *satellite images*, and **predict** *variables of interest*

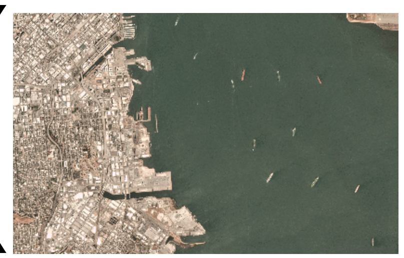
f() ≈ 45% forest

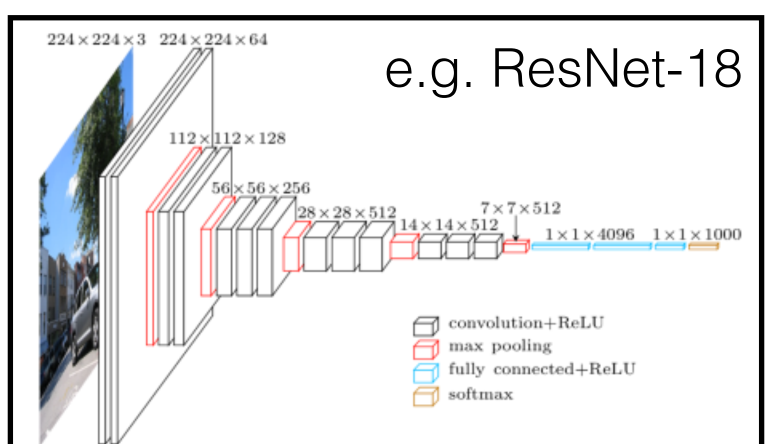
g() ≈ 14 ships

In this talk:

goal: **learn** functions that take in *satellite images*, and **predict** *variables of interest*

f() ≈ 45% forest

g() ≈ 14 ships



MOSAIKS design goals

Multi-task Observation using Satellite Imagery & Kitchen Sinks

Accessibility

Simplicity

Generalizability

MOSAIKS design goals

Accessibility

The machine learning system should be **simple to use** and **computationally efficient** for users.

Simplicity

Generalizability

MOSAIKS design goals

Accessibility

Simplicity

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

Generalizability

MOSAIKS design goals

Accessibility

Simplicity

Generalizability

One machine learning system could be useful for many prediction ***tasks,*** using a common source of satellite imagery

MOSAIKS design goals

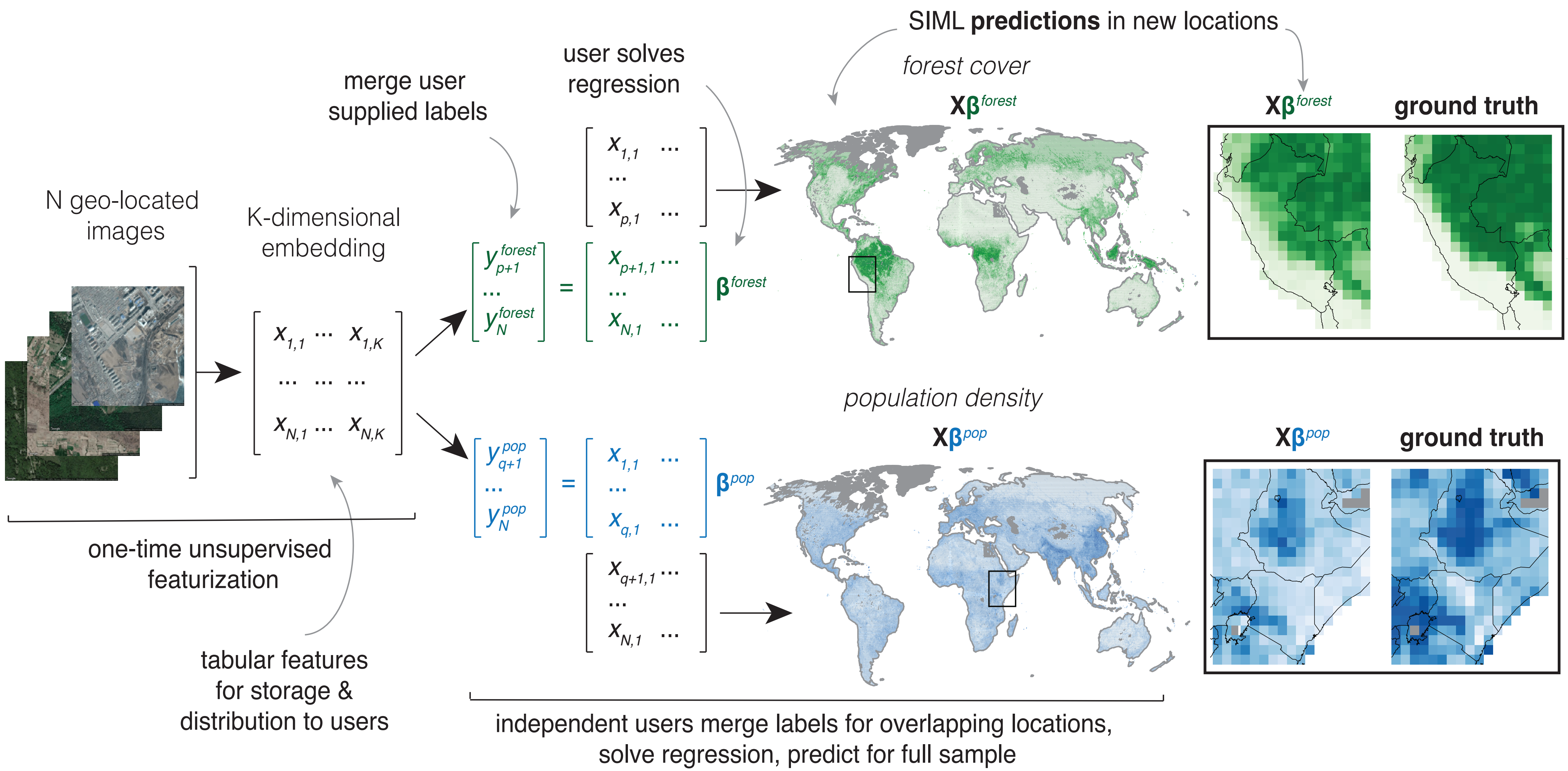
Accessibility

Simplicity

Generalizability

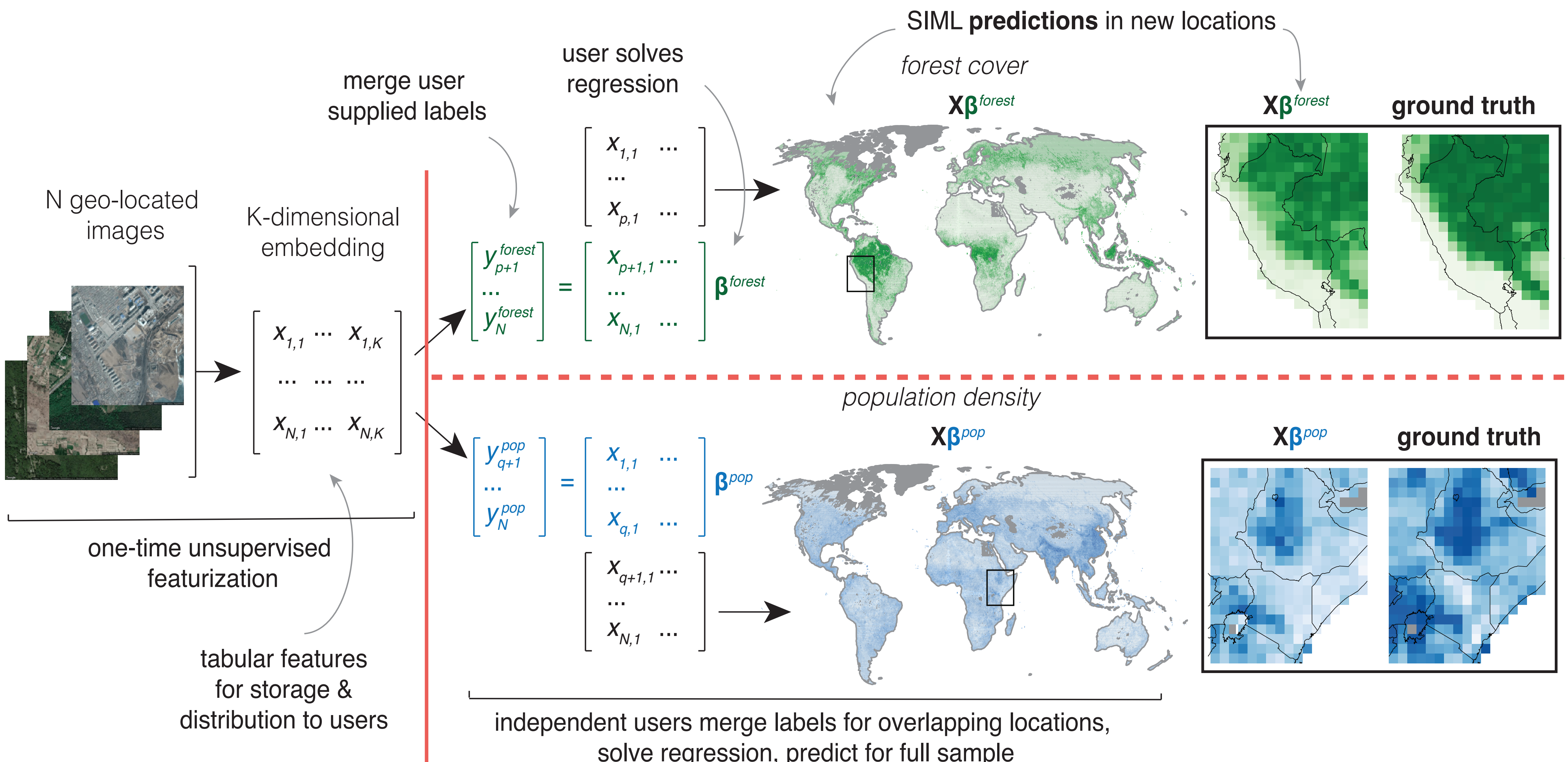
Without sacrificing accuracy or applicability.

MOSAIKS design

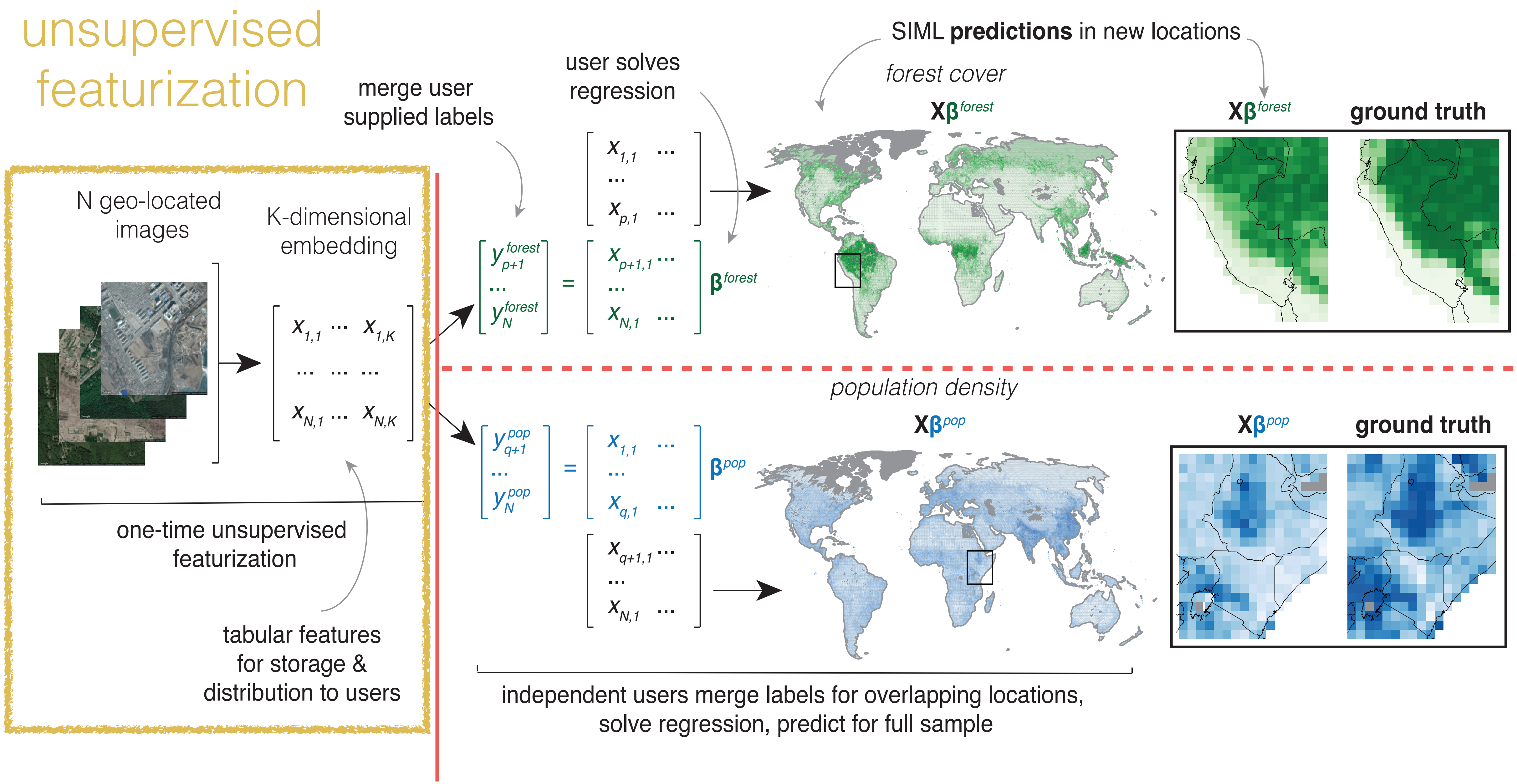


Esther Rolf (UC Berkeley)

MOSAIKS design

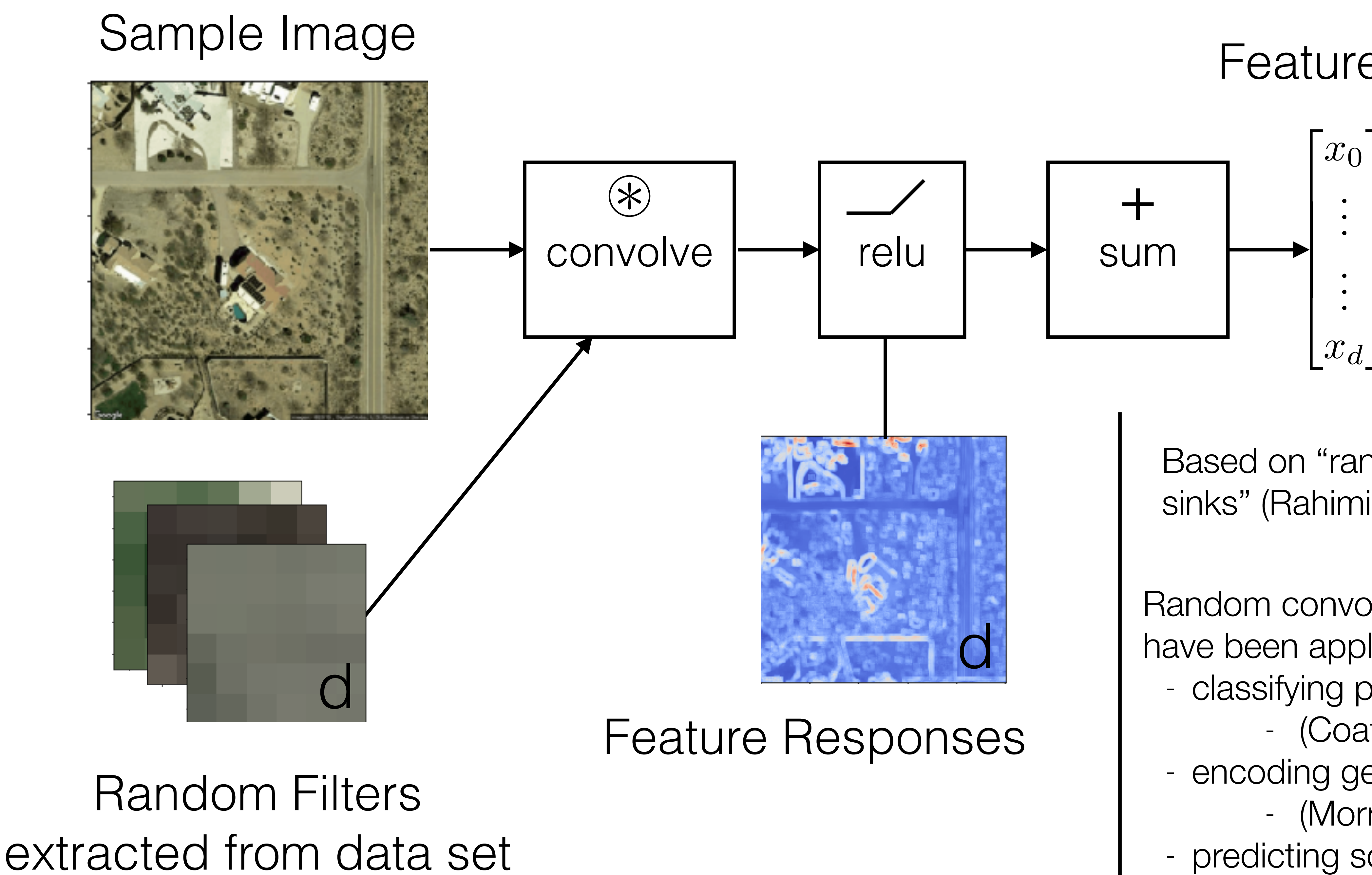


MOSAIKS design



Esther Rolf (UC Berkeley)

Random Convolutional Features



Based on “random kitchen sinks” (Rahimi & Recht 2008)

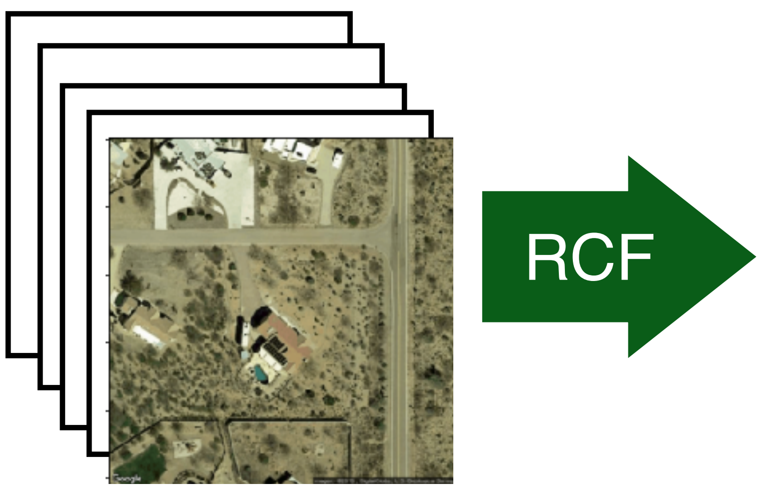
Random convolutional features have been applied to:

- classifying photographs
    - (Coates & Ng, 2011)
- encoding genomic sequences
    - (Morrow et al., 2017)
- predicting solare flares
    - (Jonas et al., 2018)

Esther Rolf (UC Berkeley)

Prediction with Random convolutional features (RCF)

n Images



Feature Matrix in R$^{n \times d}$

$$\begin{bmatrix} \text{---} \; x^{(0)} \; \text{---} \\ \vdots \\ \text{---} \; x^{(n)} \; \text{---} \end{bmatrix} \longrightarrow \min_{w} \{ \|Xw - y\|_2^2 + \lambda \|w\|_2^2 \}$$

n labels $\begin{bmatrix} y^{(0)} \\ \vdots \\ y^{(n)} \end{bmatrix}$

Image features are created without knowledge of labels!

Prediction with unsupervised features

n Images

CNN

Feature Matrix in $R^{n \times d}$

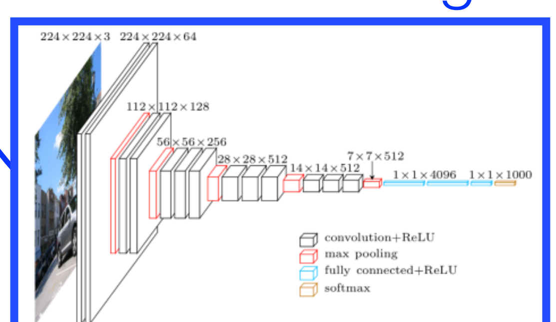
$$\begin{bmatrix} — & x^{(0)} & — \\ & \vdots & \\ — & x^{(n)} & — \end{bmatrix}$$

n labels

$$\begin{bmatrix} y^{(0)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

$$\min_{w} \{ \|Xw - y\|_2^2 + \lambda \|w\|_2^2 \}$$

$\boldsymbol{\omega}$: model weights

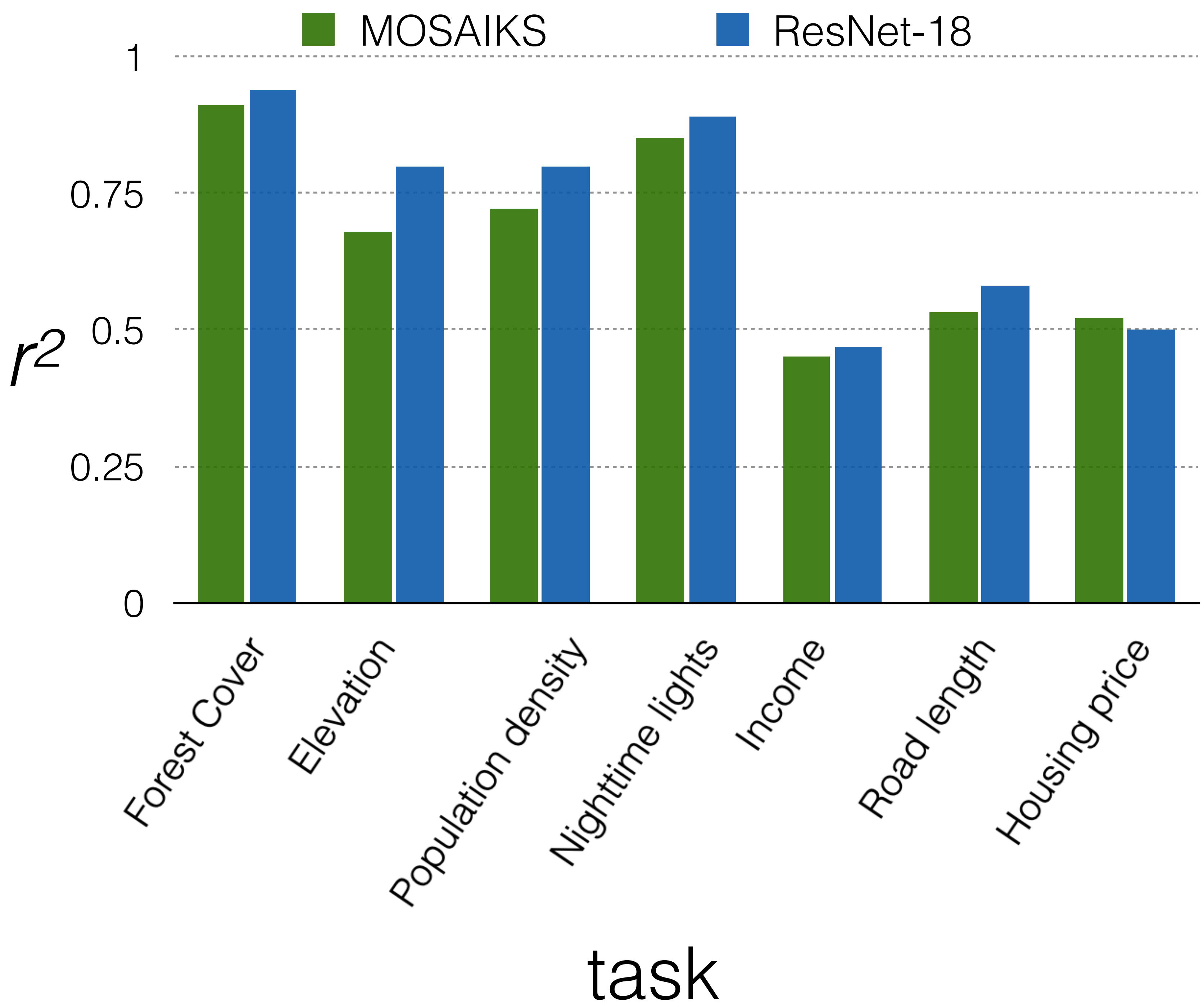


In contrast, Convolutional Neural Networks (CNNs) learn domain-specific image features

Prediction Domains

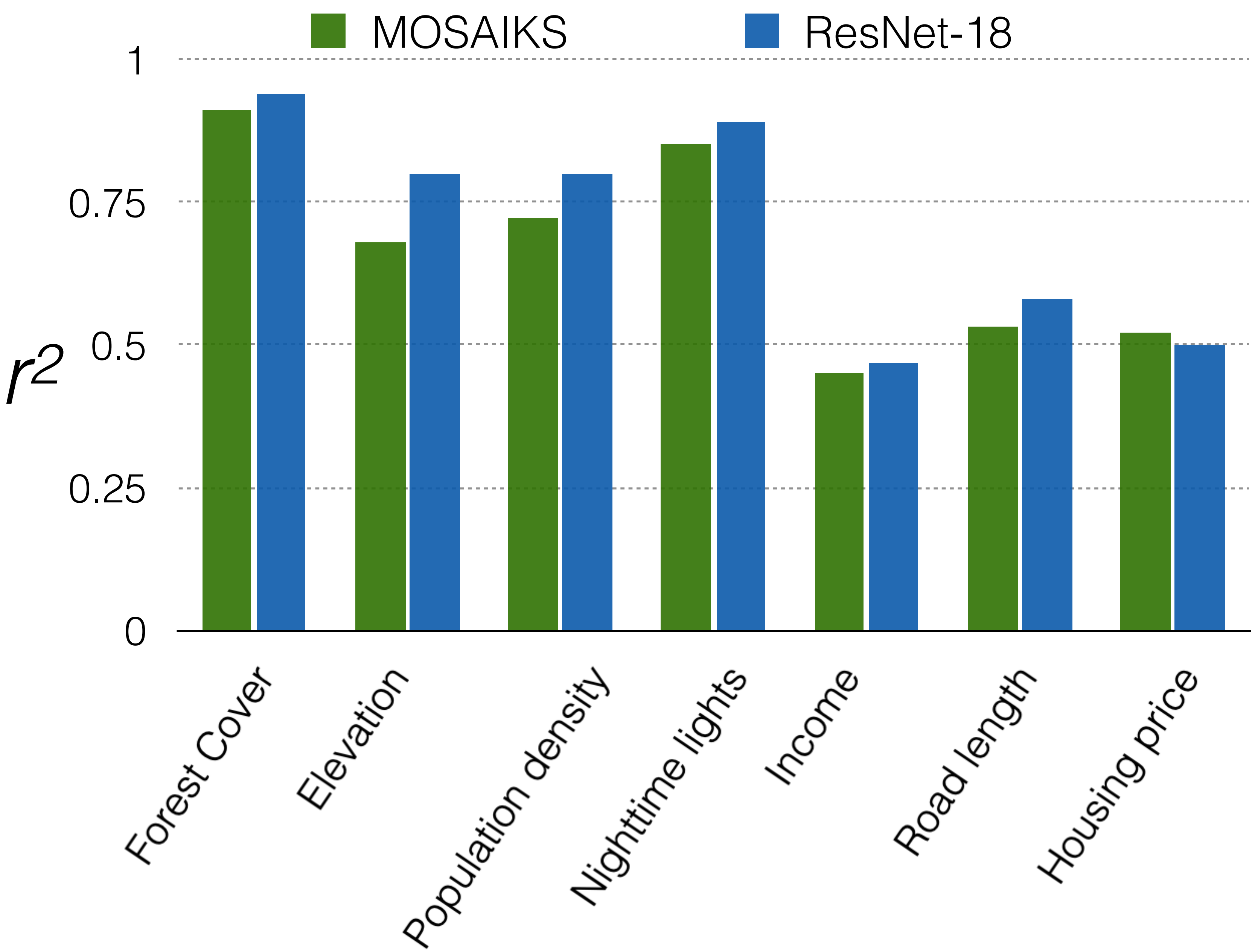
- Forest Cover
- Elevation
- Population
- Nighttime Luminosity
- Income
- Road Length
- Housing Price

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



Esther Rolf (UC Berkeley)

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



Training times:

MOSAIKS: **1 minute**
(CPU, 10 cores)

Esther Rolf (UC Berkeley)

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

The most commonly used CNNs are designed for classification of "natural imagery"

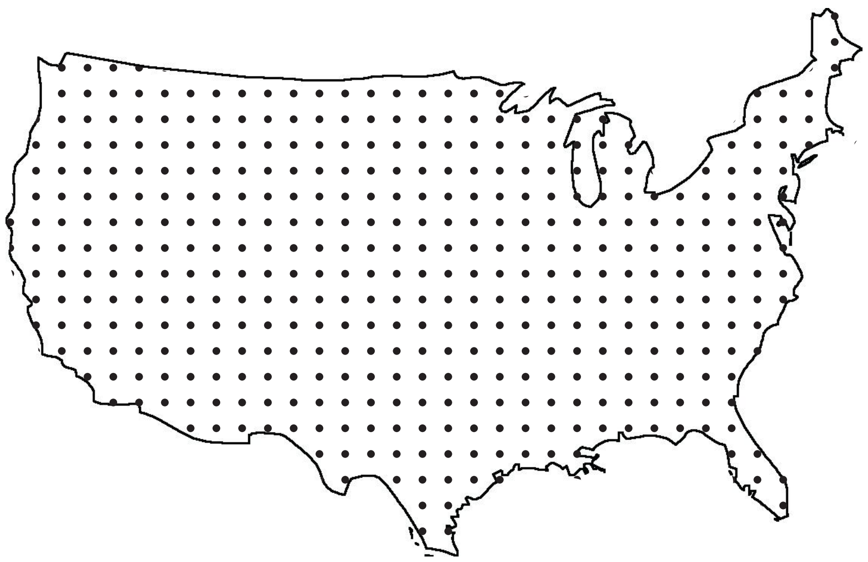

Satellite images are largely **scale** and **orientation** invariant.


The structures in satellite imagery can explain why performance of simpler methods would match that of more complex methods.

Esther Rolf (UC Berkeley)

Methods

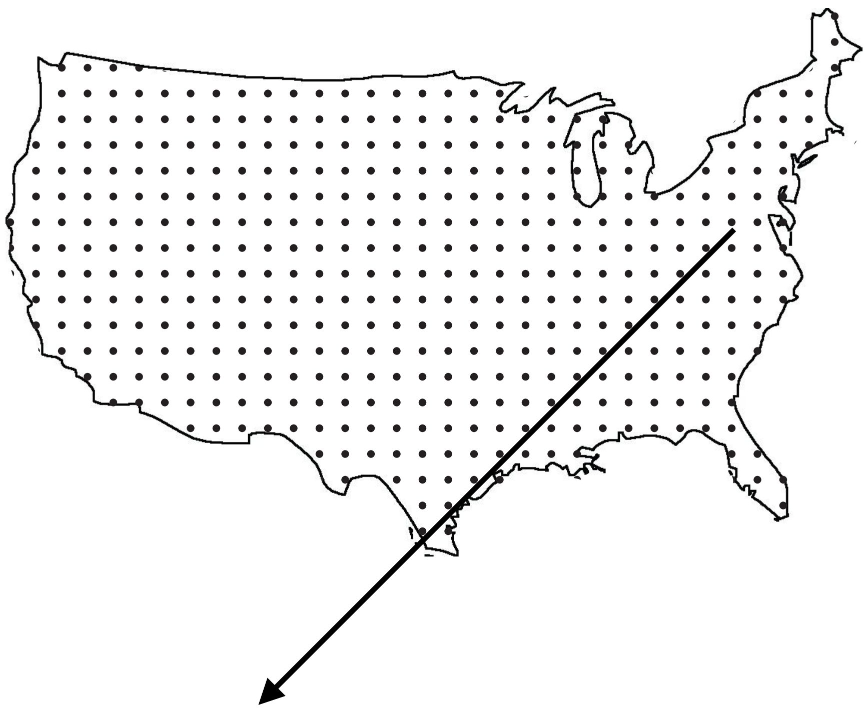
1. Grid the U.S.

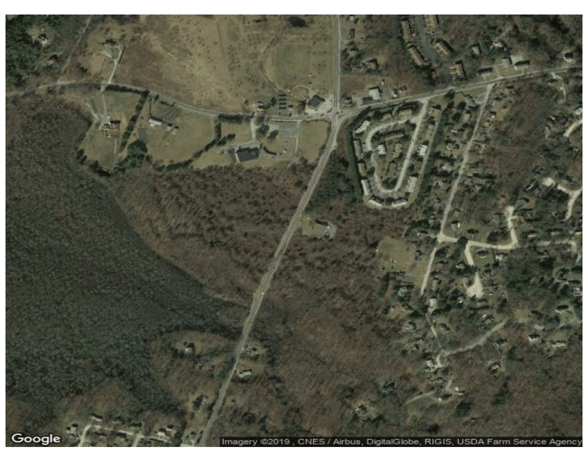


Esther Rolf (UC Berkeley)

Methods

1. Grid the U.S.

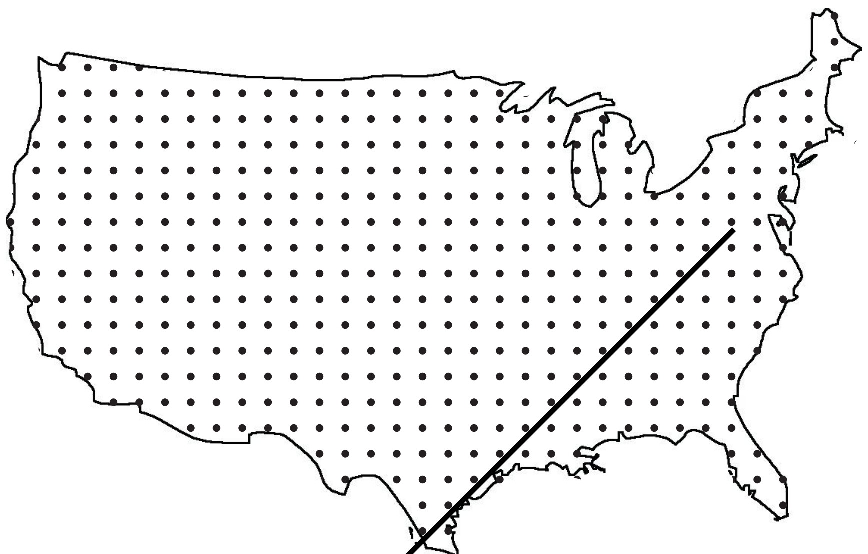




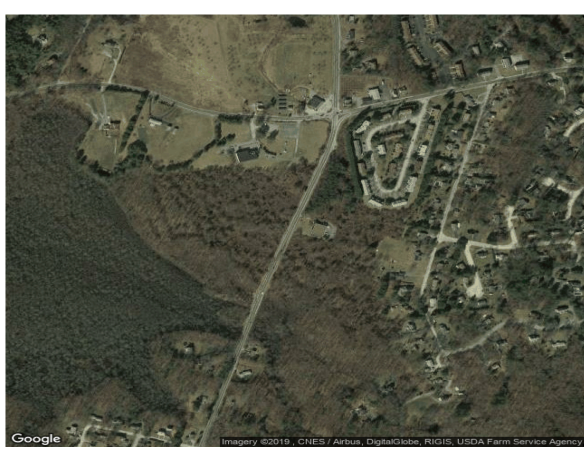
~1km x 1km

Methods

1. Grid the U.S.

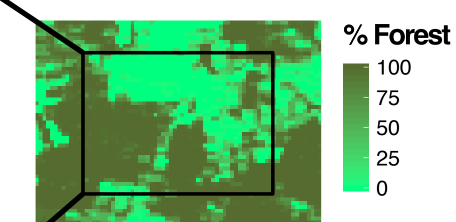


format and spatial resolution
of the labeled data varies across tasks



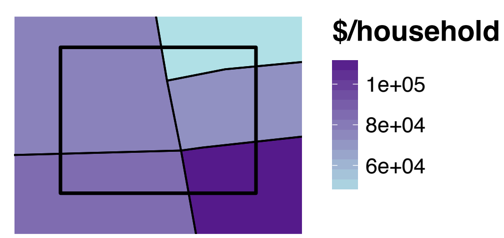
~1km x 1km

Forest cover



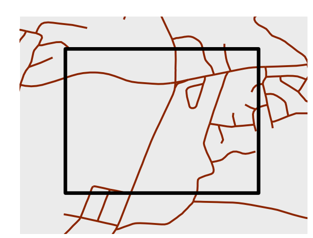
Value = average of values within grid cell

Income



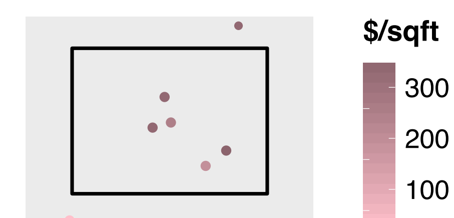
Value = area-weighted average of block group income

Road length



Value = sum of road length

Housing price



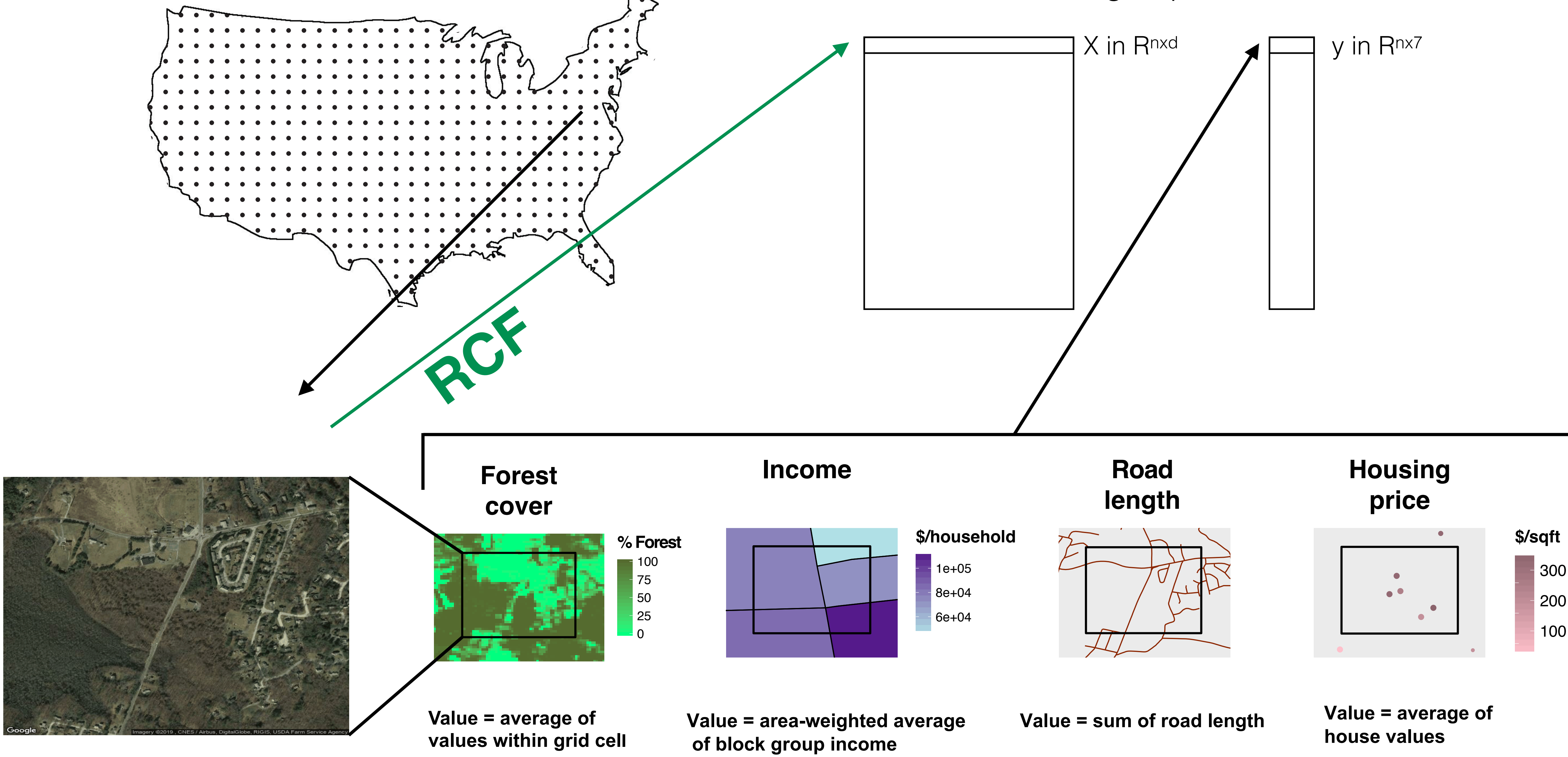
Value = average of house values

Esther Rolf (UC Berkeley)

Methods

1. Grid the U.S.

2. Download and featurize Images (RCF)
3. Associate with each grid point each of 7 labels

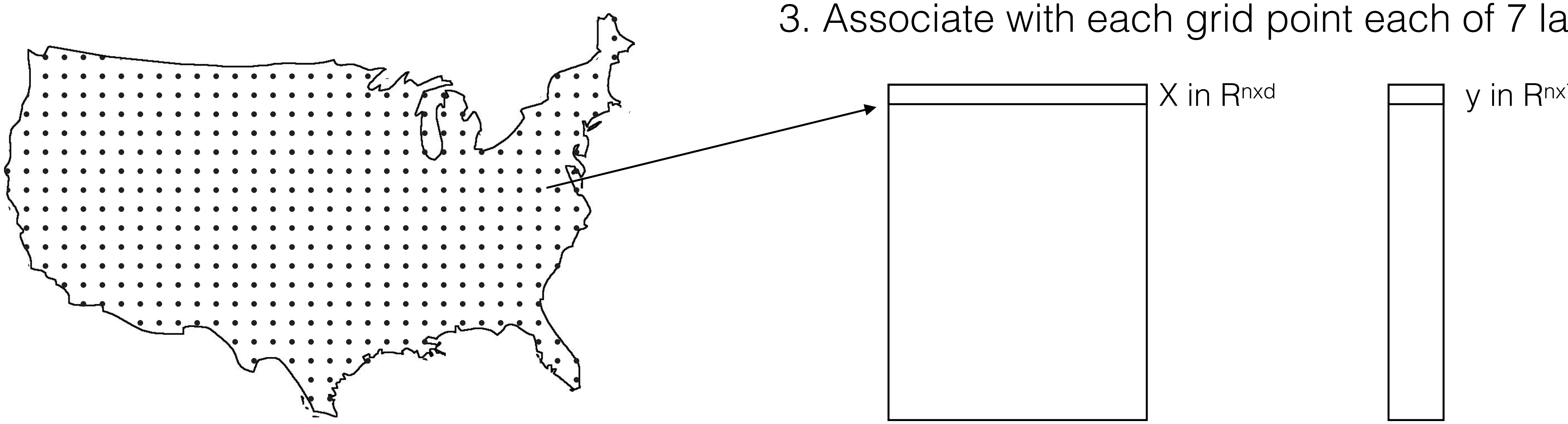


~1km x 1km

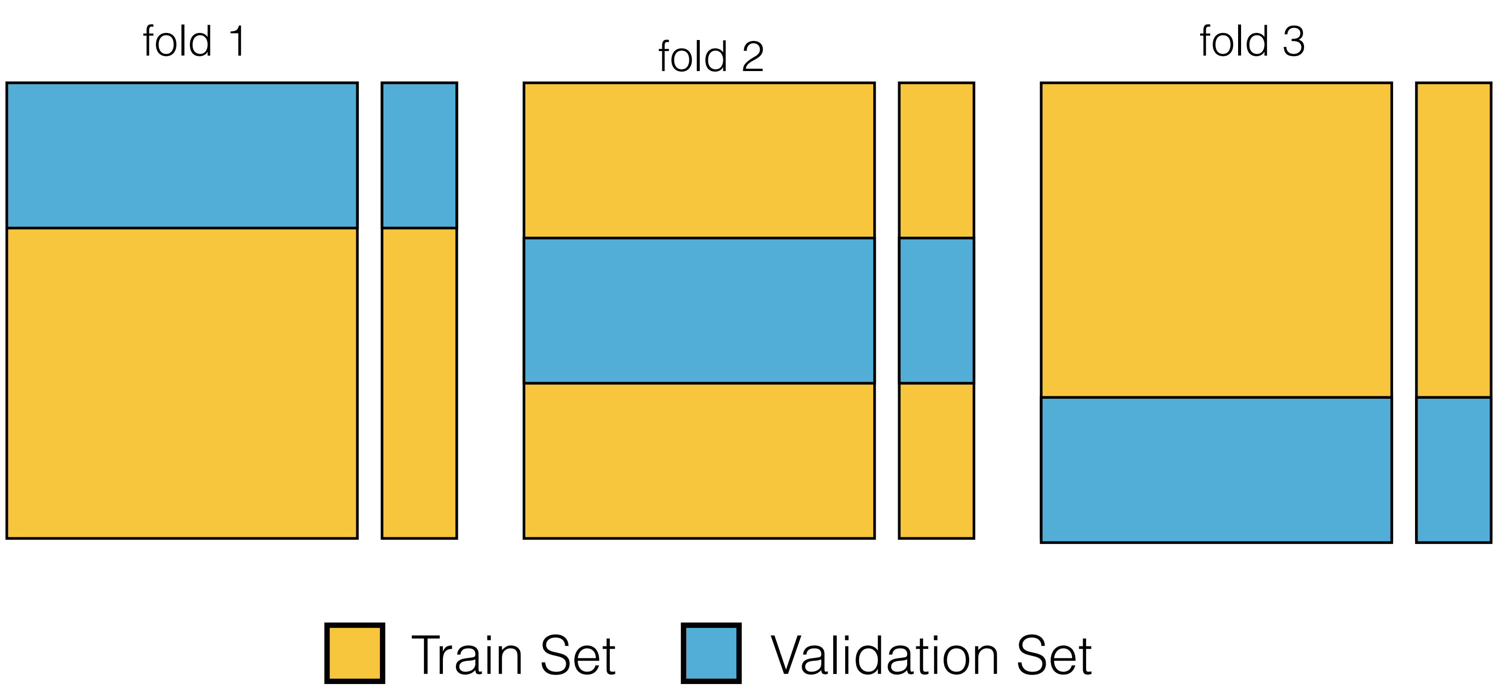
Methods

1. Grid the U.S.

2. Download and featurize Images (RCF)
3. Associate with each grid point each of 7 labels



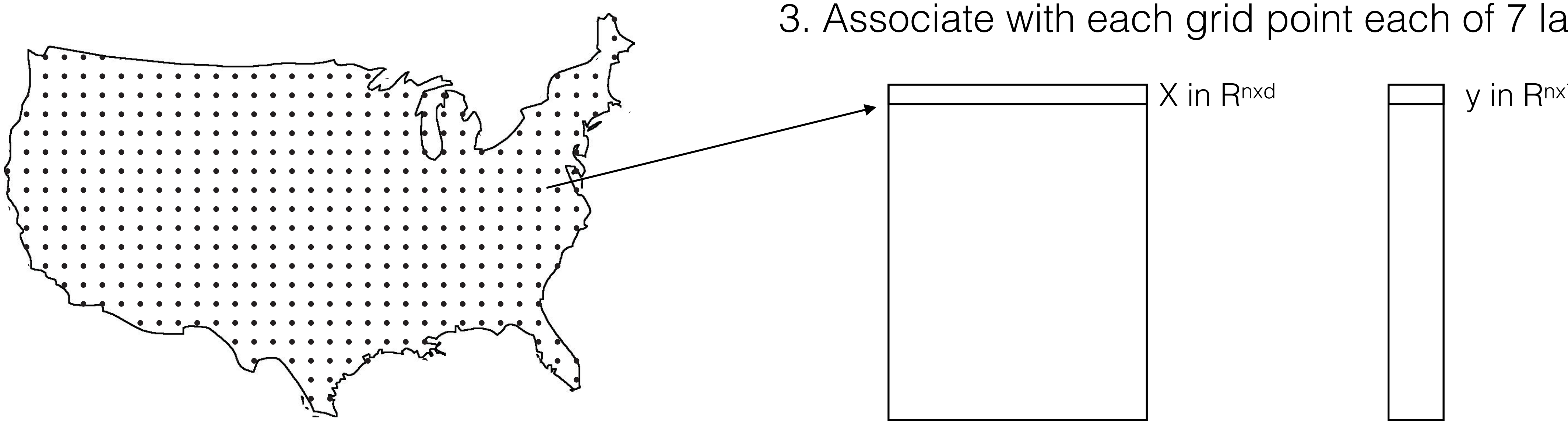
4. k-fold cross validation; pick model parameters



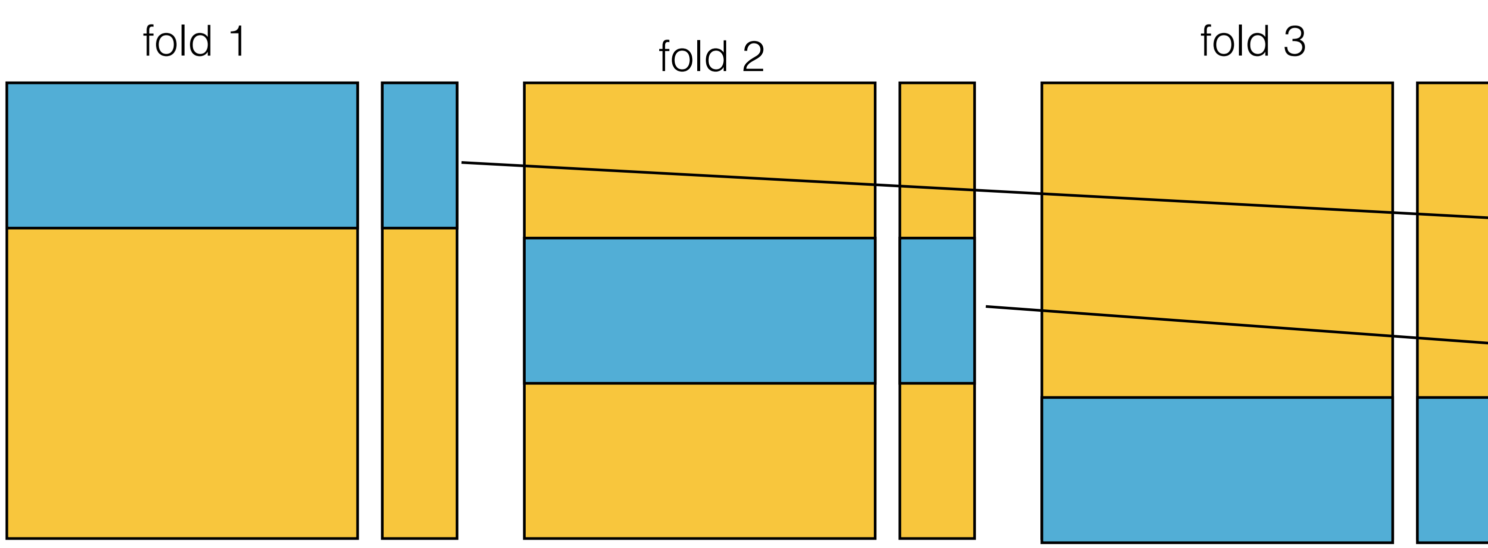
Methods

1. Grid the U.S.

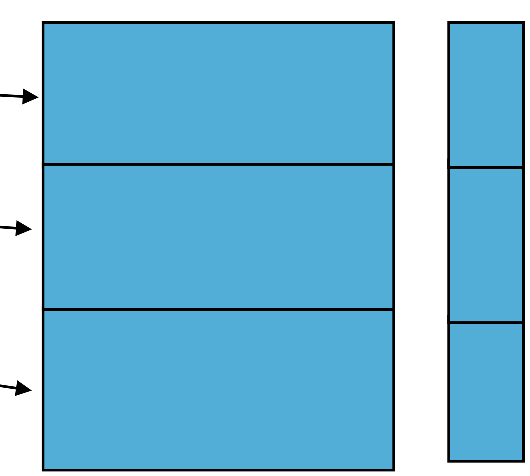
2. Download and featurize Images (RCF)
3. Associate with each grid point each of 7 labels



4. k-fold cross validation; pick model parameters



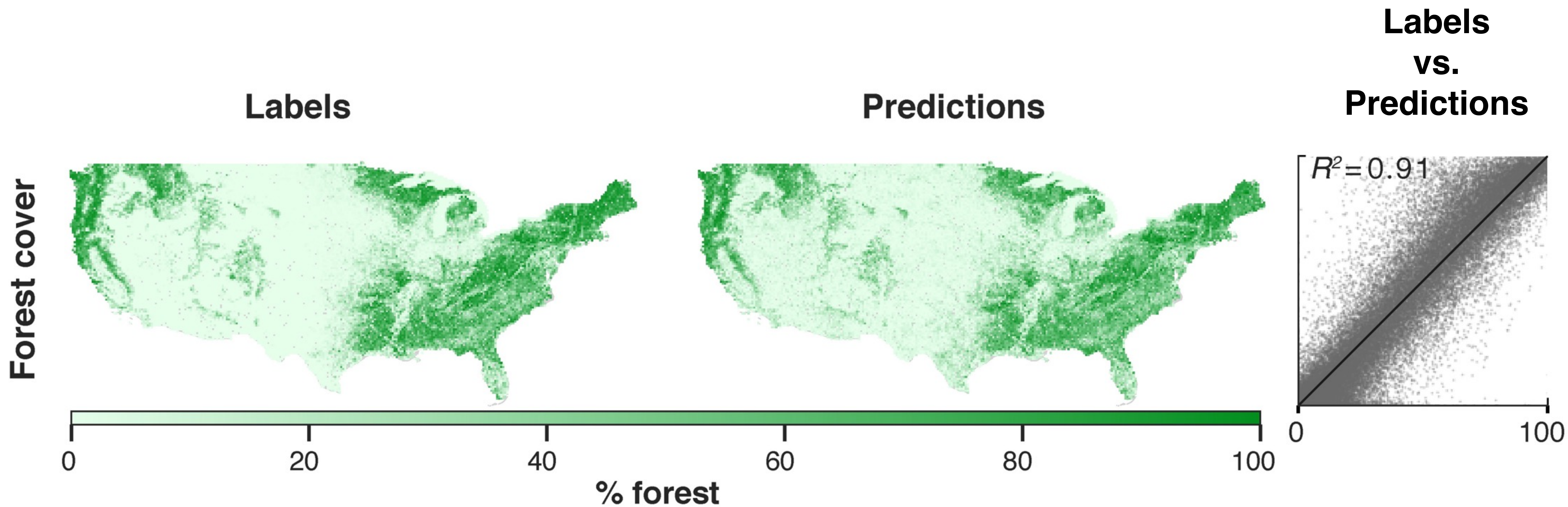
5. Report accuracy (R^2) for concatenated validation sets.



Train Set   Validation Set

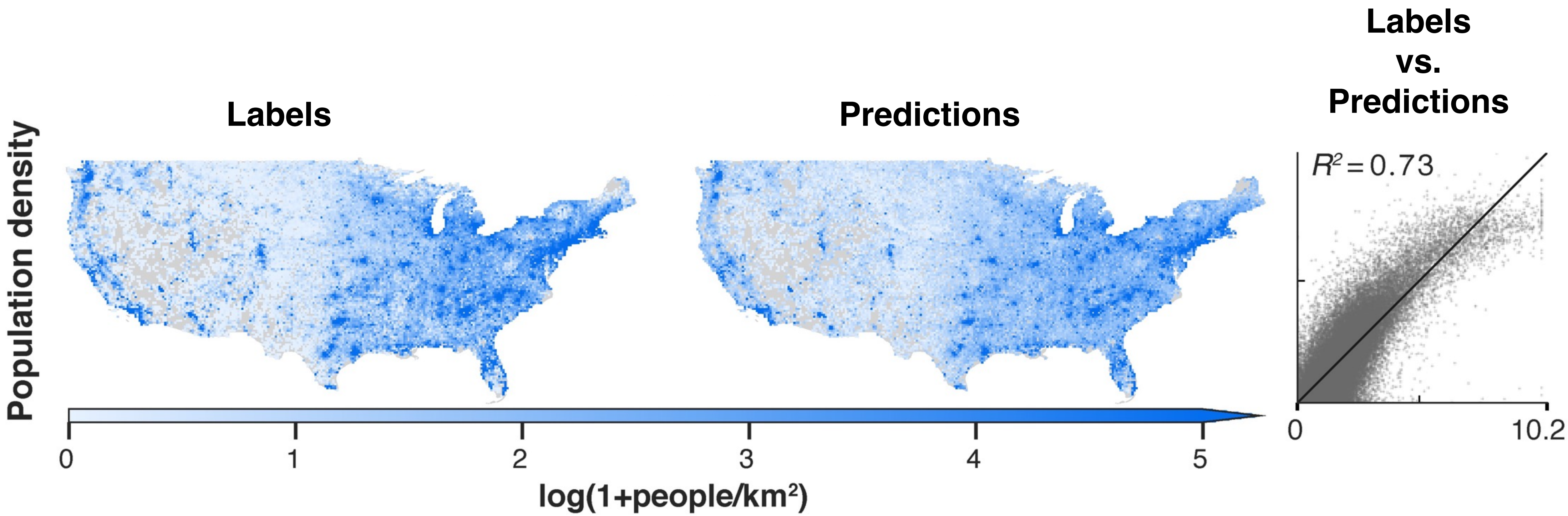
Esther Rolf (UC Berkeley)

US Outcome #1: **Forest Cover**

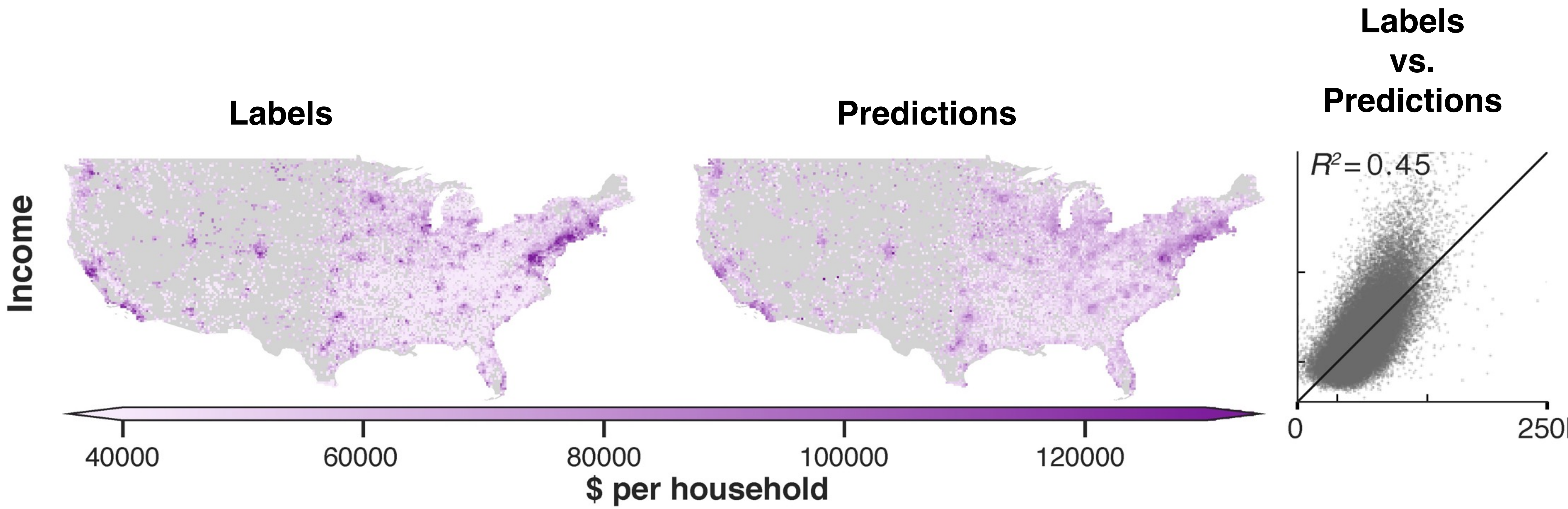


Esther Rolf (UC Berkeley)

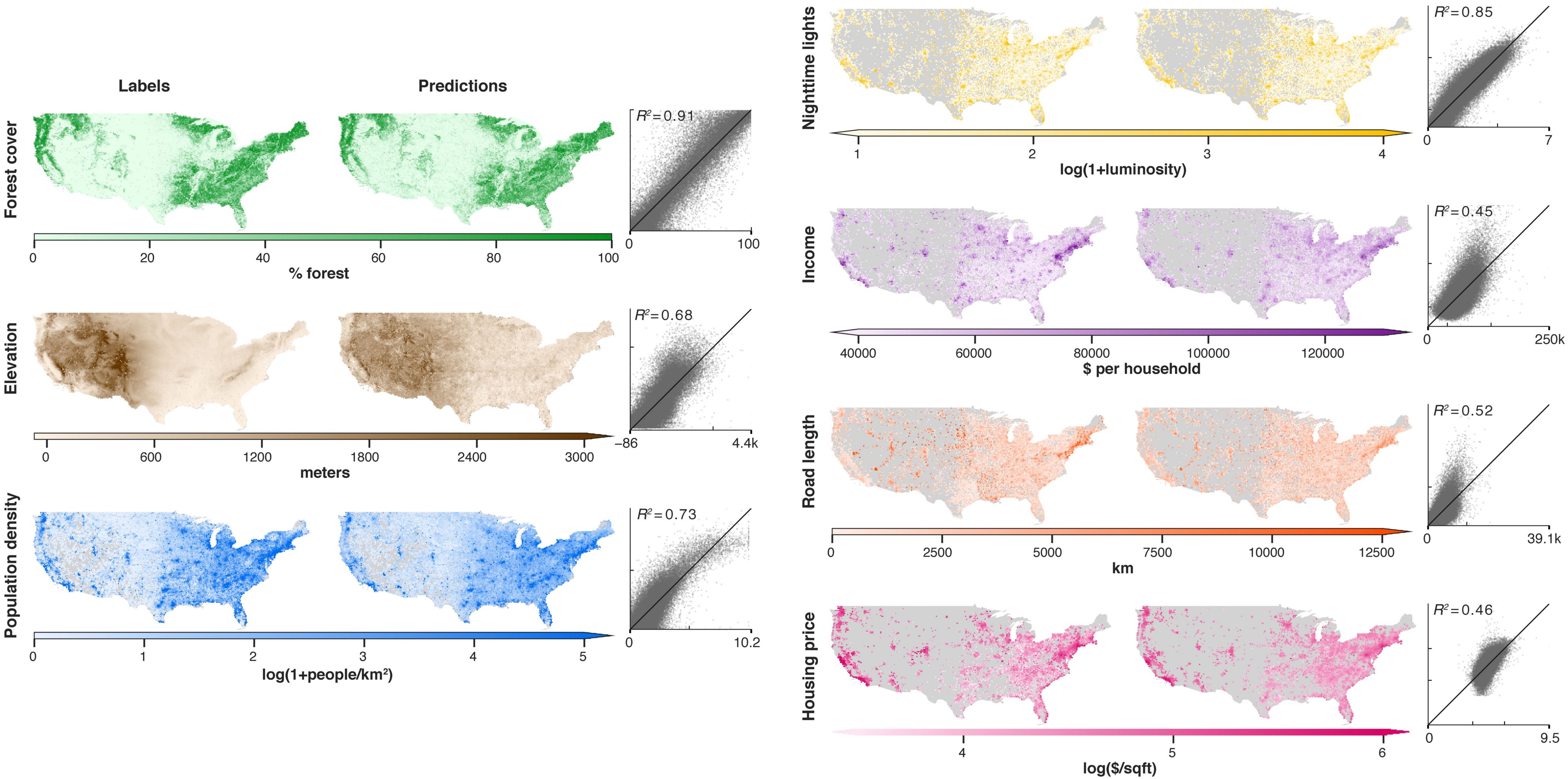
US Outcome #3: **Population Density**



Outcome #5: **Income (per household)**



A common featurization allows use to directly compare performance across outcomes.



Esther Rolf (UC Berkeley)

Global Results

- Forest cover, nighttime luminosity, elevation
- Train on ~700k image label pairs, test on ~100k
  - Using the exact same featurization as in the U.S.
  - Report accuracy (R^2) on test set.

Global Results

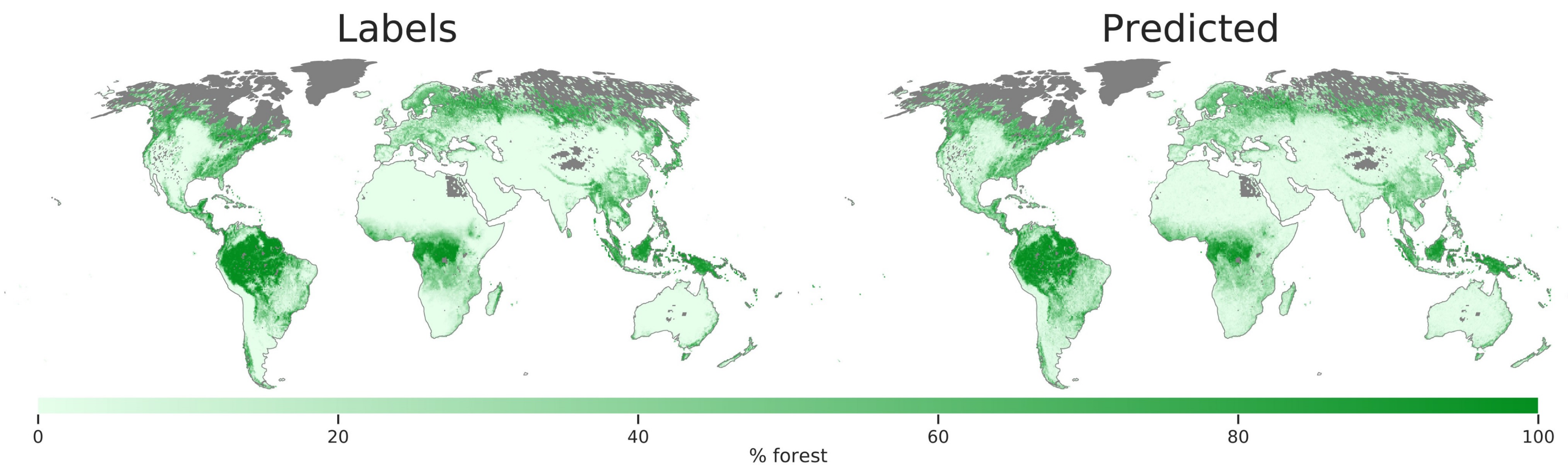
- Forest cover, nighttime luminosity, elevation
- Train on ~700k image label pairs, test on ~100k
  - Using the exact same featurization as in the U.S.
  - Report accuracy (R^2) on test set.

challenges to scaling globally:

- more imagery
- label distribution mismatch U.S. to global
- images differing quality, many are “missing”

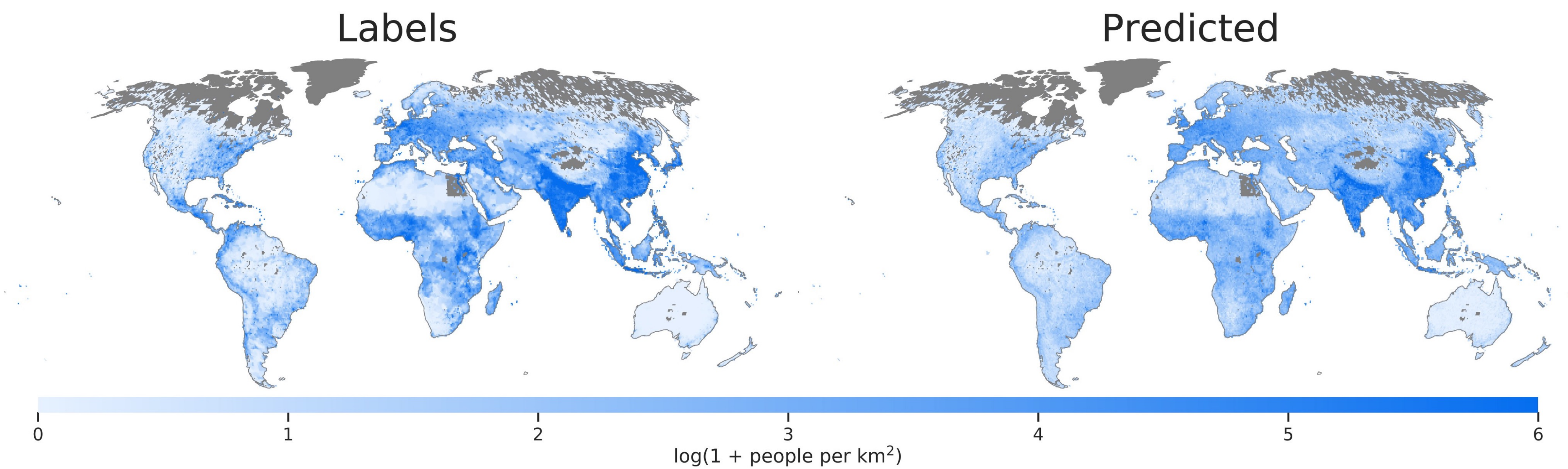
Global Outcome #1: **Forest Cover**

$r^2 = 0.85$



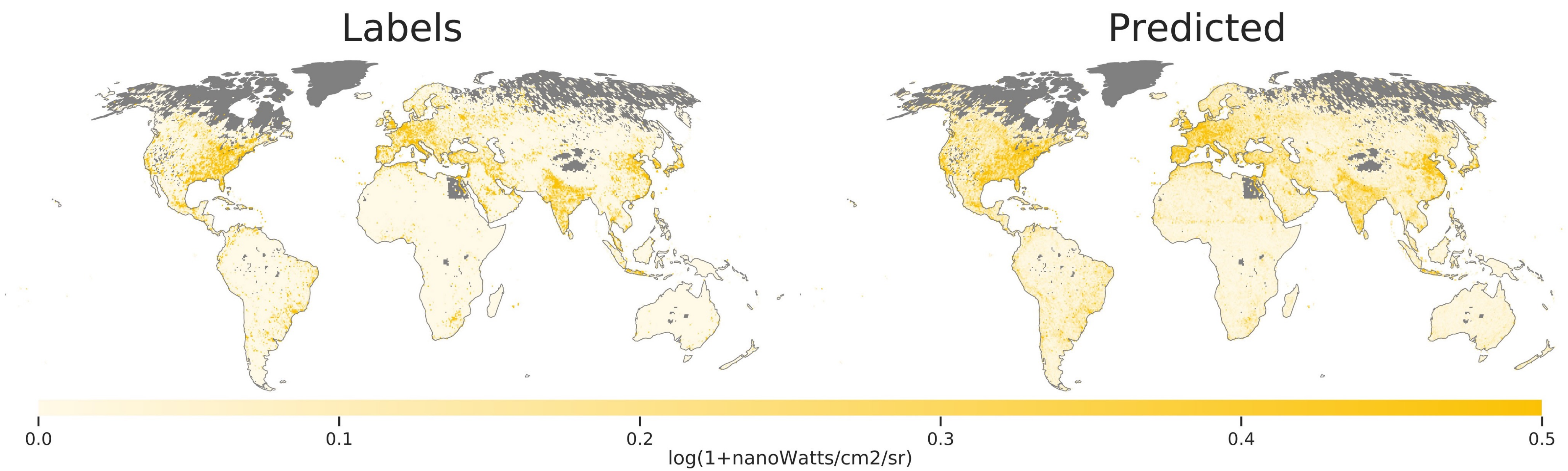
Global Outcome #2: **Population Density**

$r^2 = 0.62$



Global Outcome #3: **Nighttime Lights**

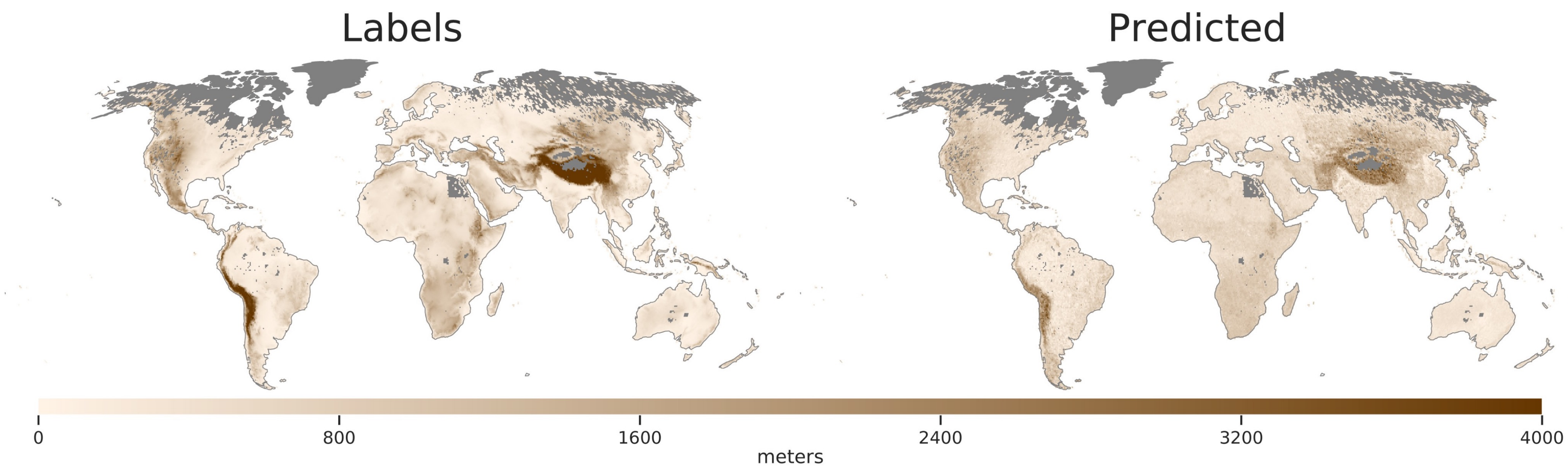
$r^2 = 0.49$



Esther Rolf (UC Berkeley)

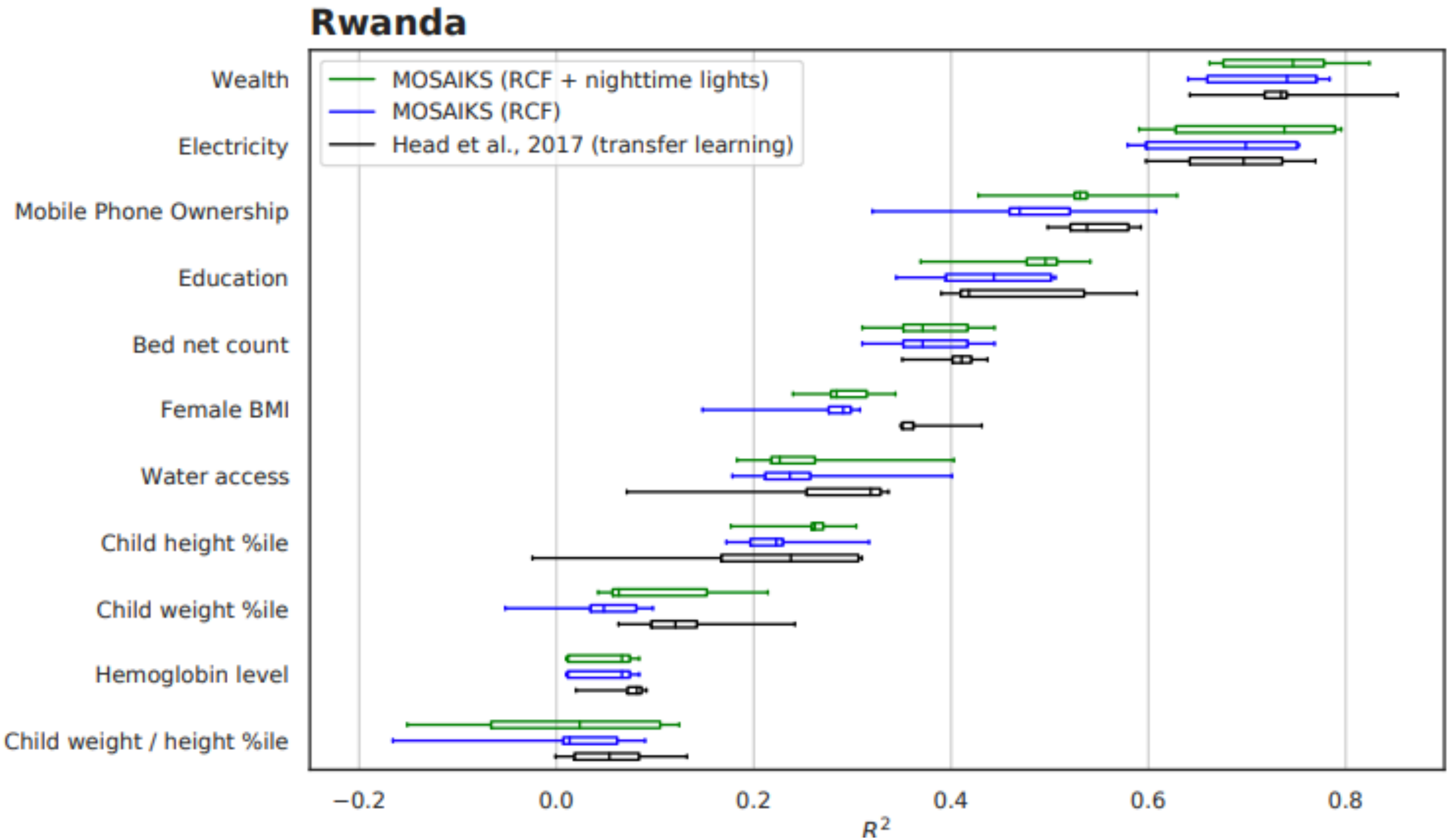
Global Outcome #4: **Elevation**

$r^2 = 0.45$



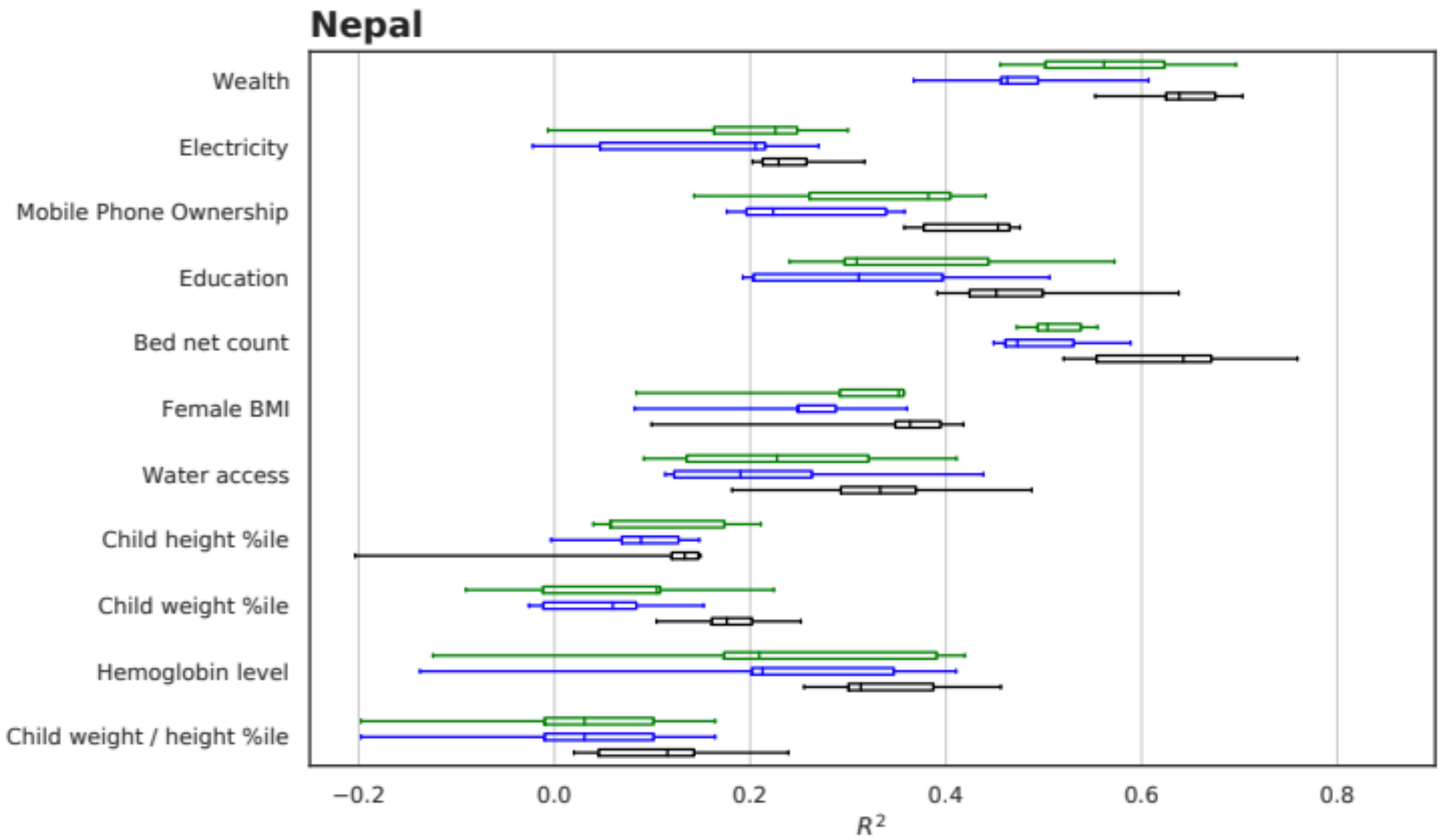
Esther Rolf (UC Berkeley)

MOSAIKS for measuring human development



Esther Rolf (UC Berkeley)

MOSAIKS for measuring human development



Esther Rolf (UC Berkeley)

Many analyses end here.

Our fixed featurization allows further exploration of our predictions, and their use in practice

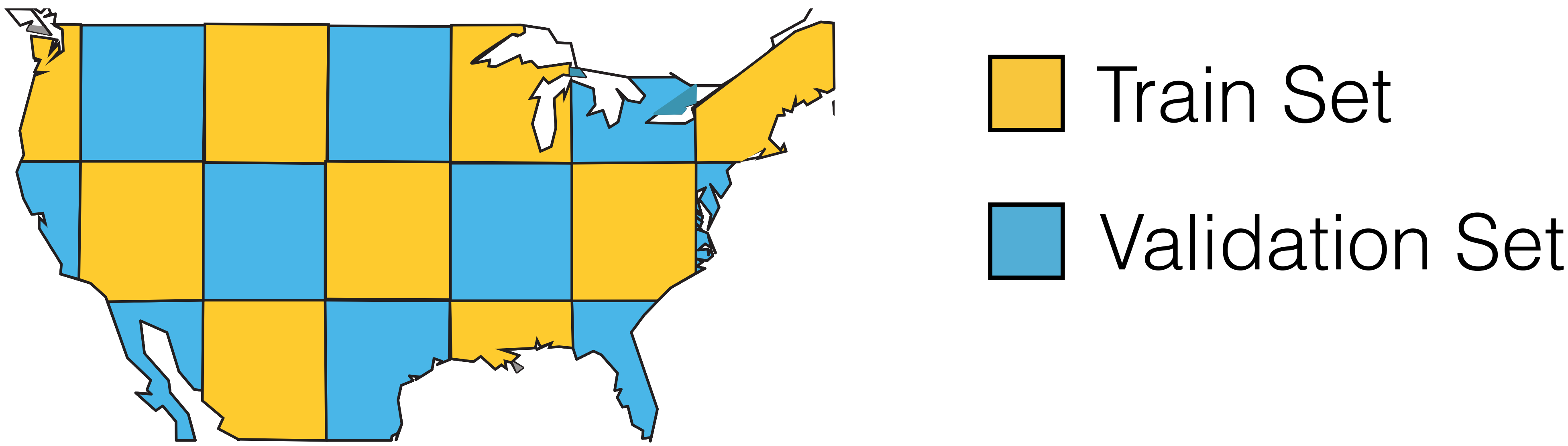
Generalizing across space

Goal: assess whether our method is generalizing over space, or just learning locality.

Generalizing across space

Goal: assess whether our method is generalizing over space, or just learning locality.

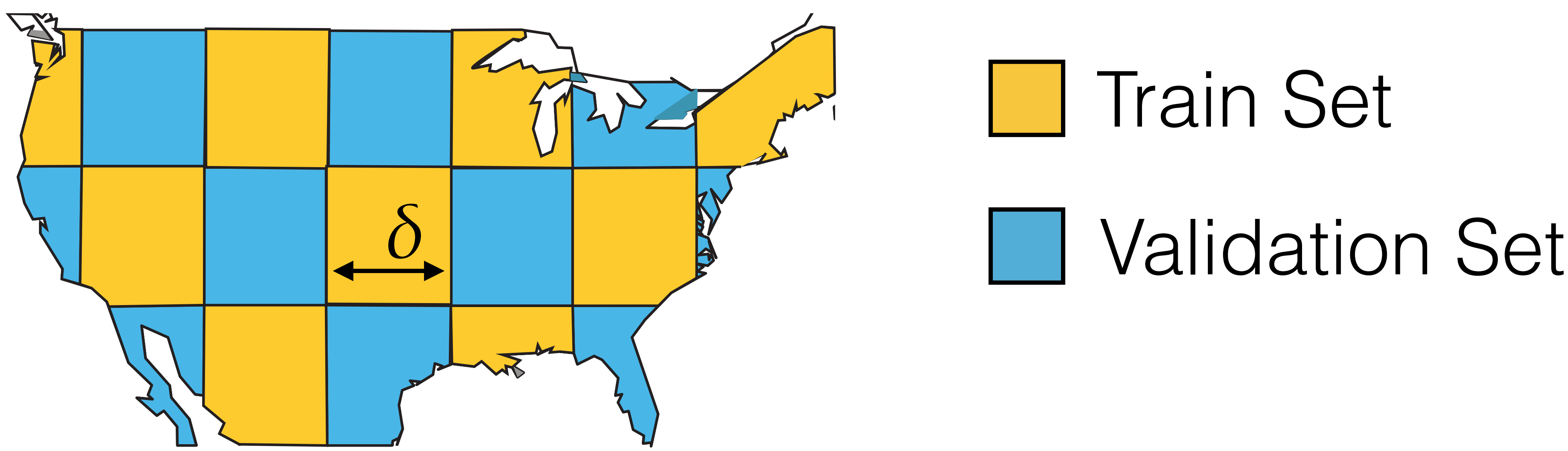
- Procedure: Split up the U.S. into geographically disjoint train and validation sets:



Generalizing across space

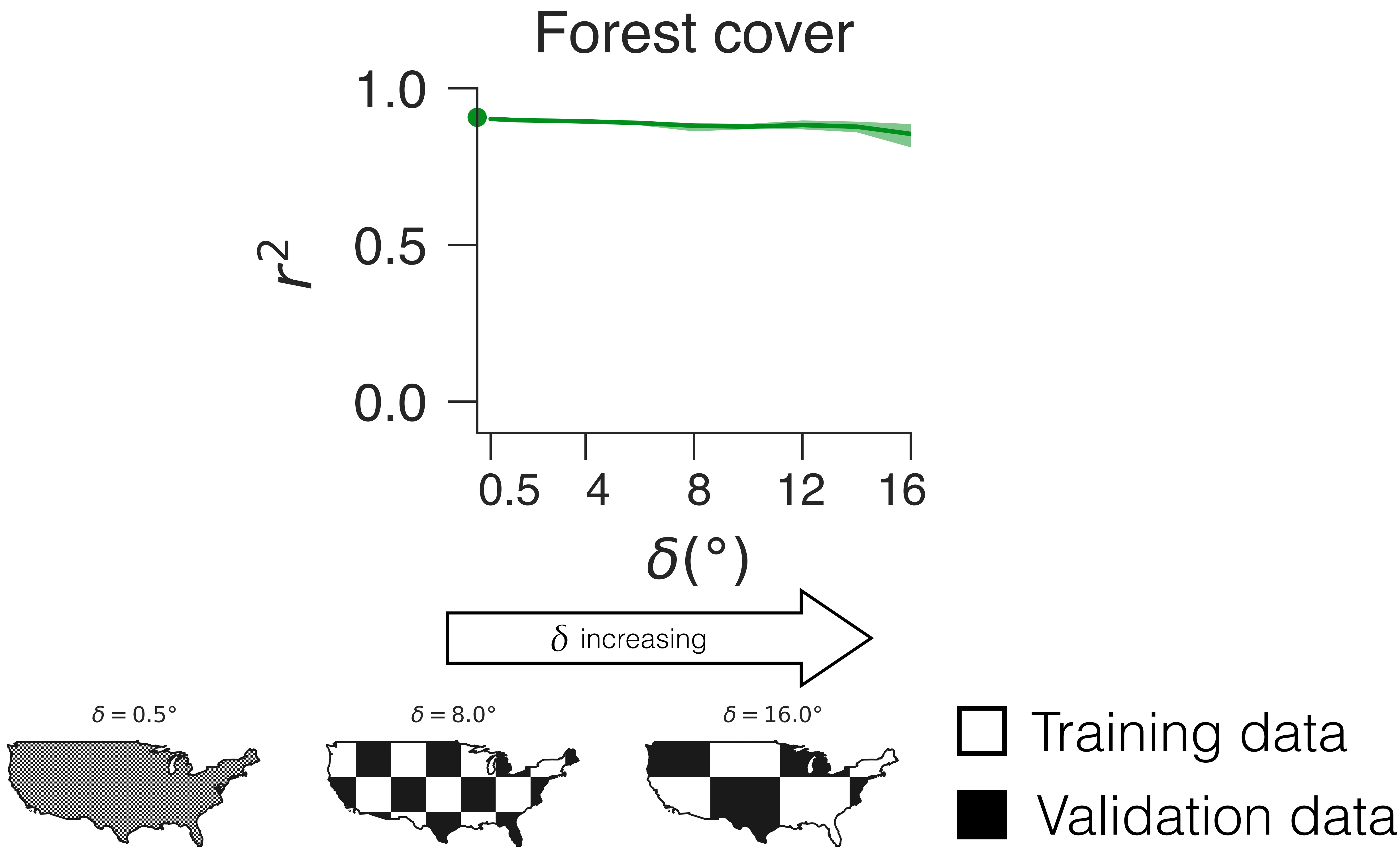
Goal: assess whether our method is generalizing over space, or just learning locality.

- Procedure: Split up the U.S. into geographically disjoint train and validation sets:



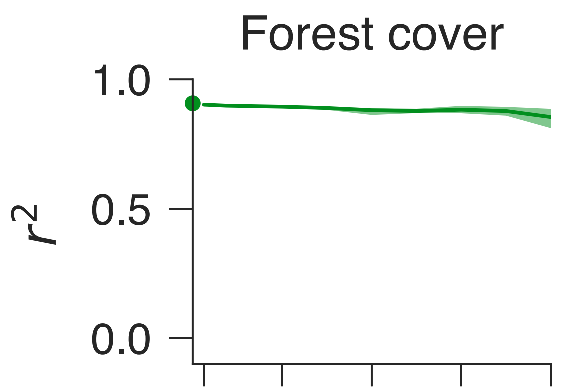
- Vary δ; at what distance from the training set can we predict points in the validation set?

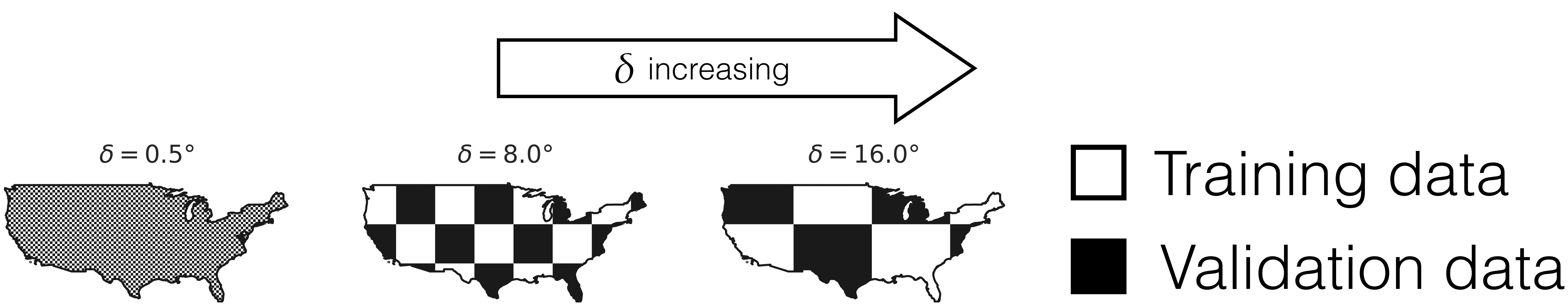
As degree of spatial extrapolation (δ) increases, performance degrades differently across domains.



Esther Rolf (UC Berkeley)

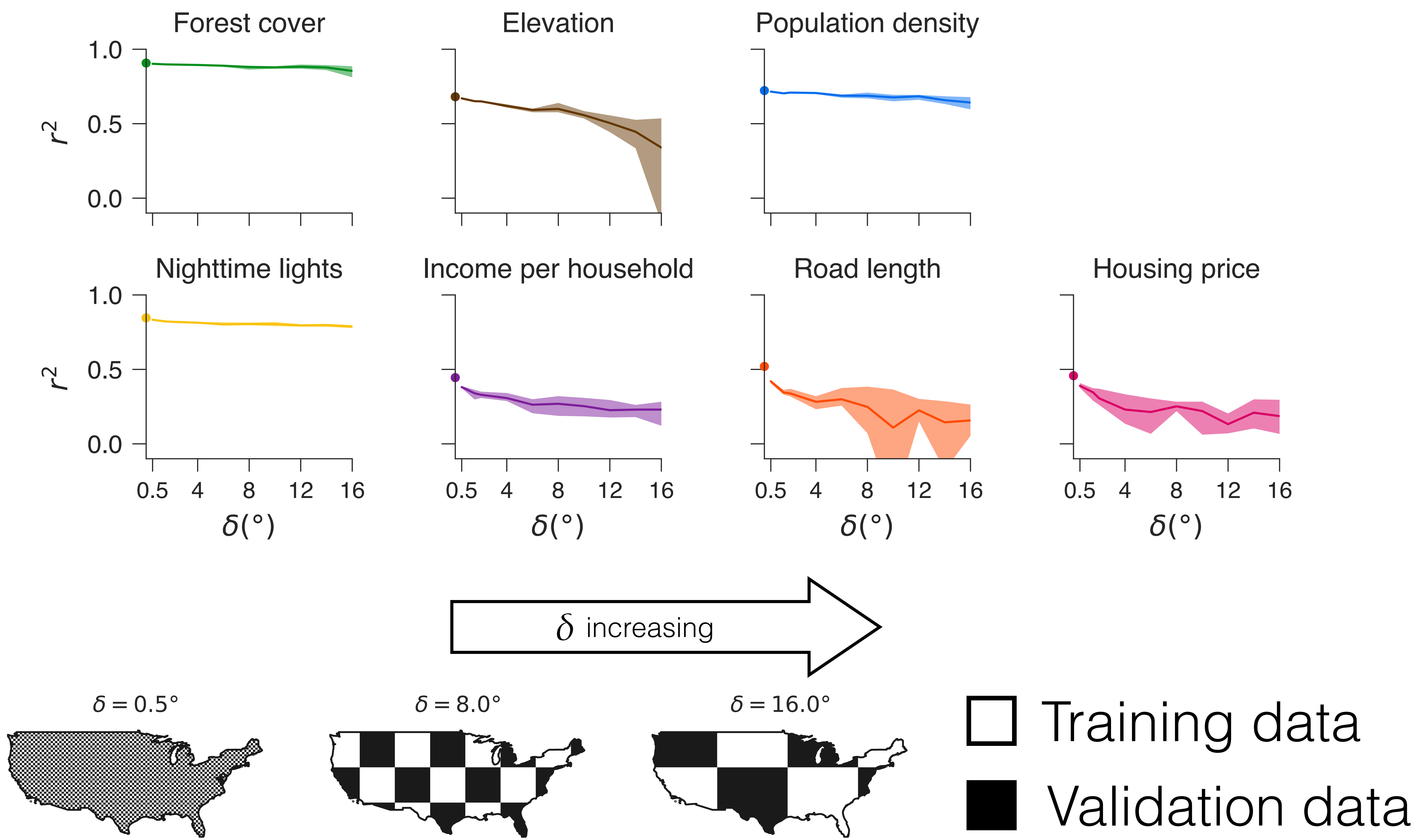
As degree of spatial extrapolation (δ) increases, performance degrades differently across domains.





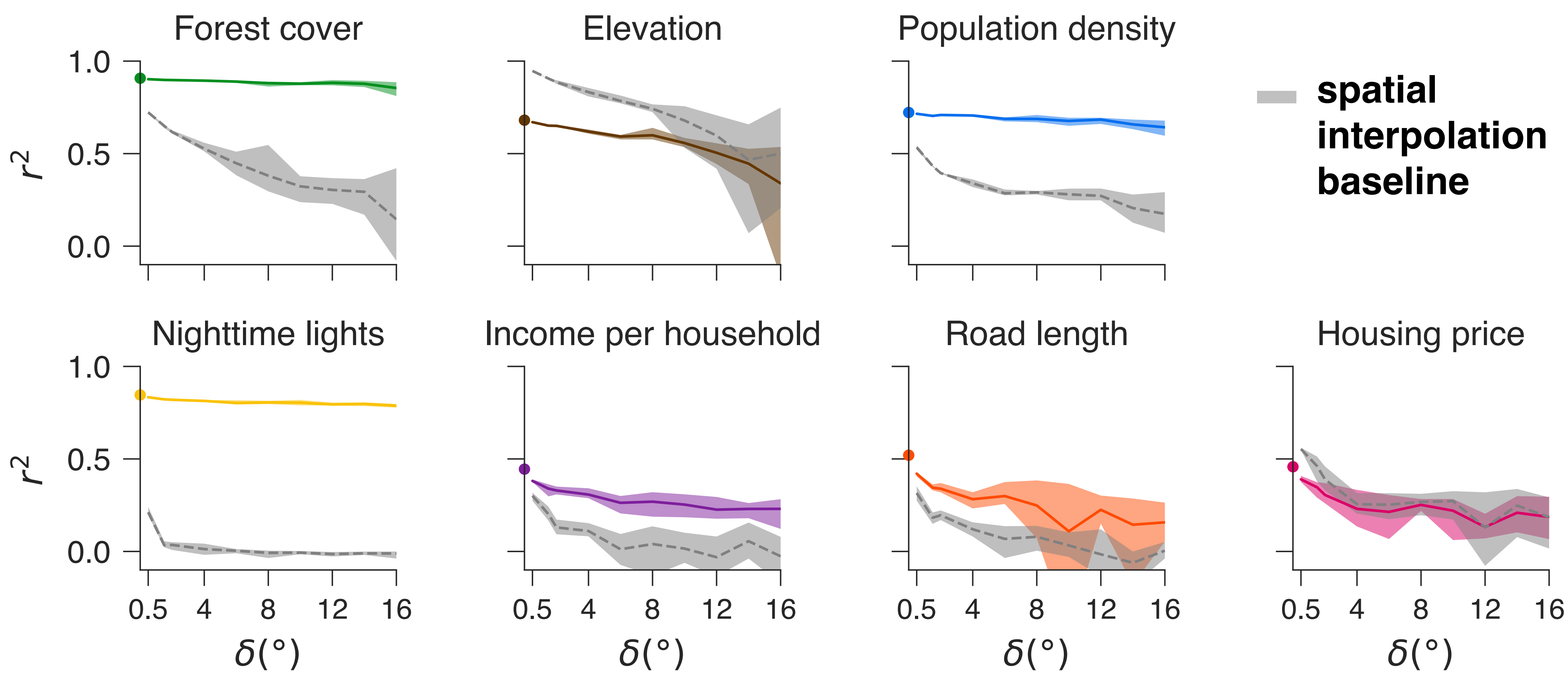
Esther Rolf (UC Berkeley)

As degree of spatial extrapolation (δ) increases, performance degrades differently across domains.

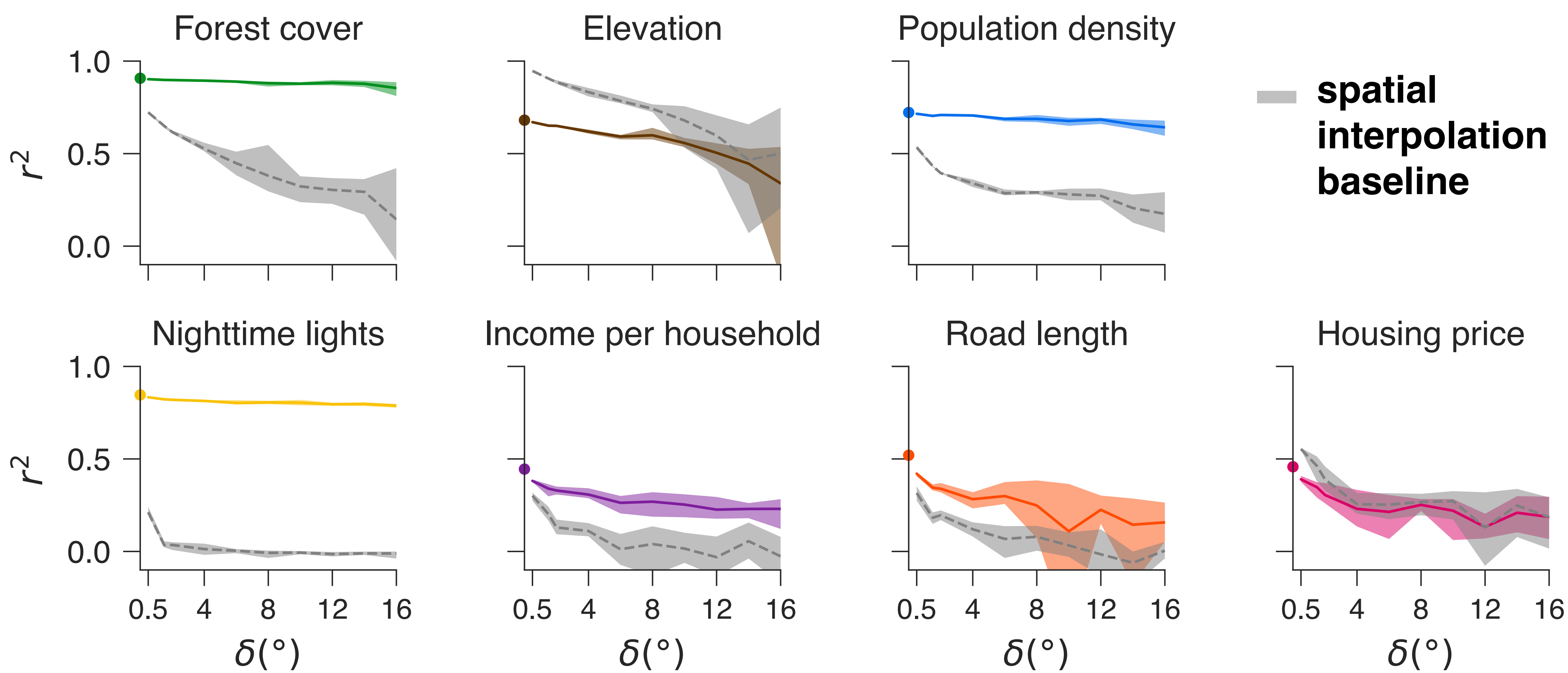


Esther Rolf (UC Berkeley)

Compared to a **spatial interpolation** baseline, our predictions have higher performance on 5/7 tasks.



Compared to a **spatial interpolation** baseline, our predictions have higher performance on 5/7 tasks.

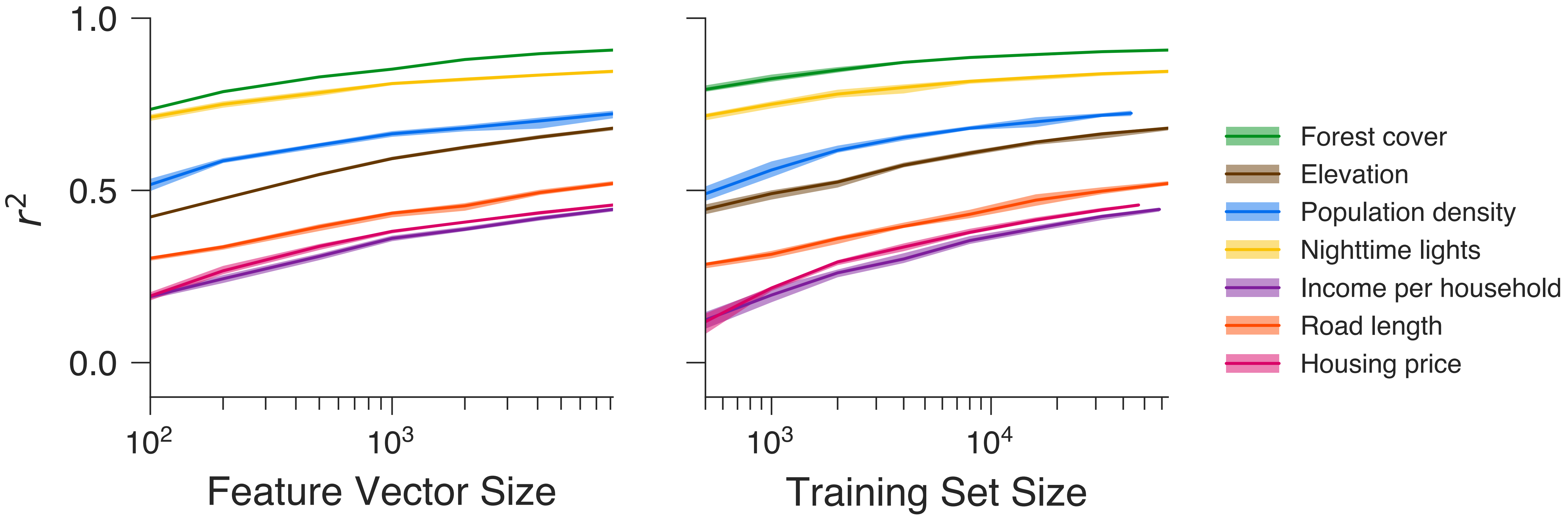


Domains where performance is worse than the baseline are known to exhibit high spatial correlation.

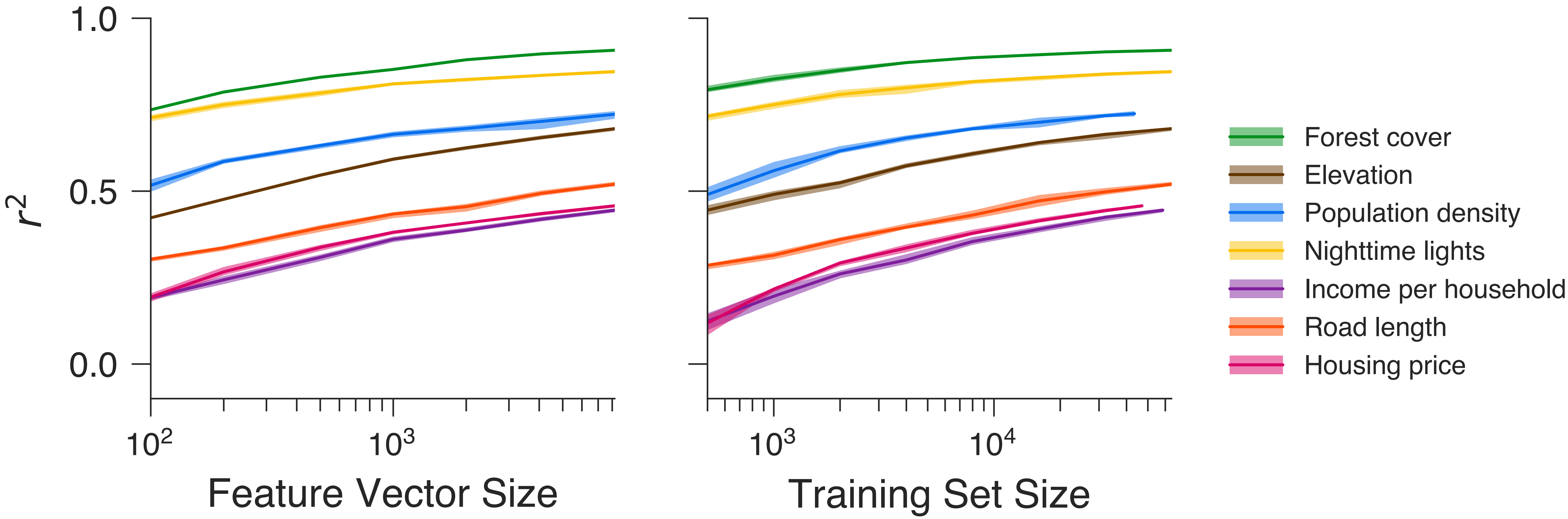
Takeaway: for certain domains, augmenting with location could be beneficial.

Esther Rolf (UC Berkeley)

Model Diagnostics



Model Diagnostics



reiterate: with standard neural nets,
we would need to **retrain for each horizontal datapoint.**

Context switch:
we’ve expanded to global scale,
but can we get finer resolution?

super-resolution

Can we predict variables at *sub-image* resolution?

predicted road length (km):

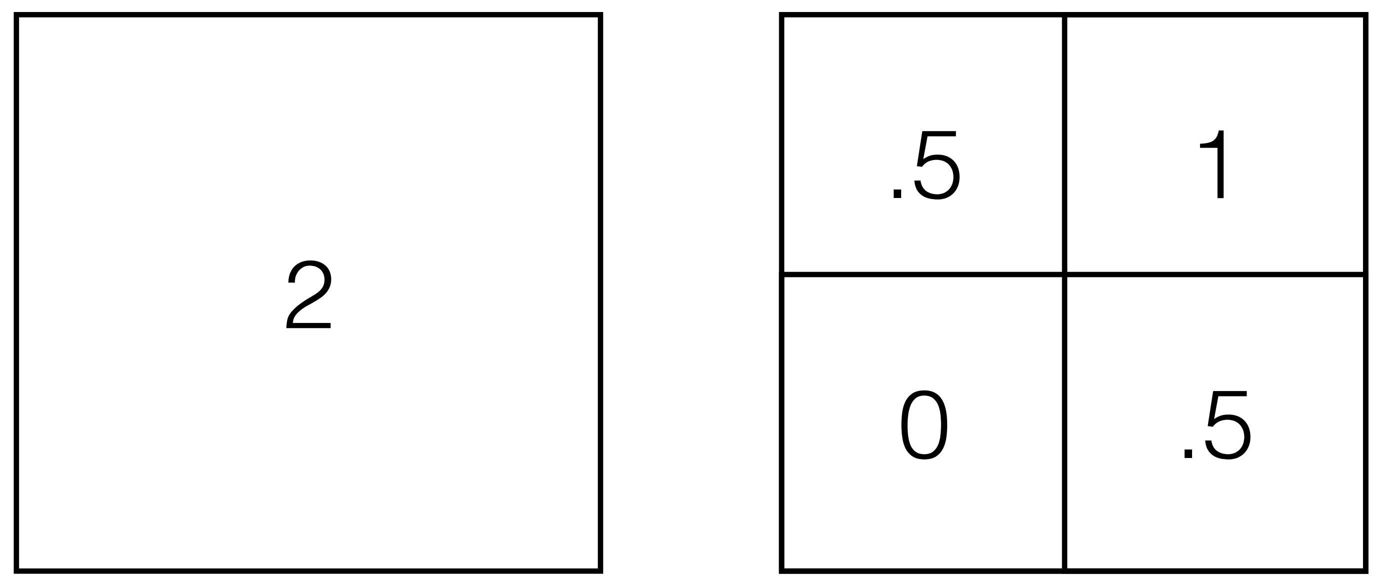
2

what we've done so far

super-resolution

Can we predict variables at *sub-image* resolution?

predicted road length (km): predicted road length (km):

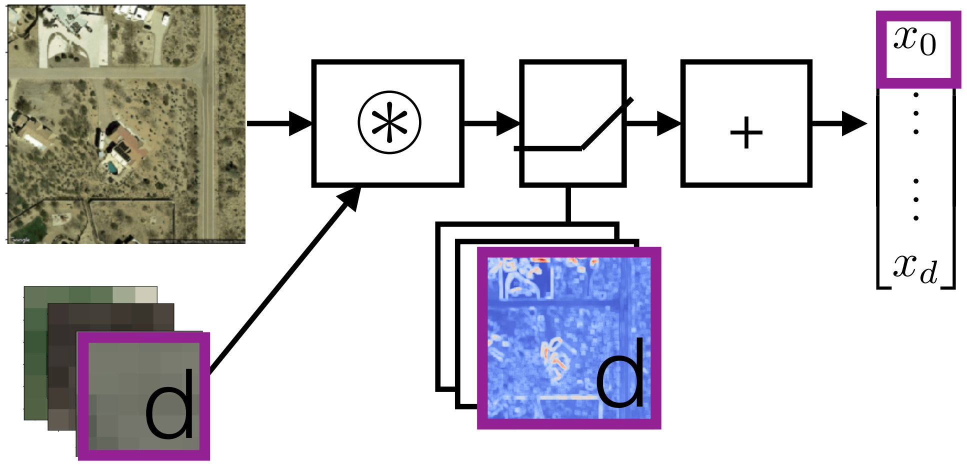


what we've done so far

superresolution

super-resolution

$$\boxed{x_j} = \sum_{p \in \text{pixels}} \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)$$

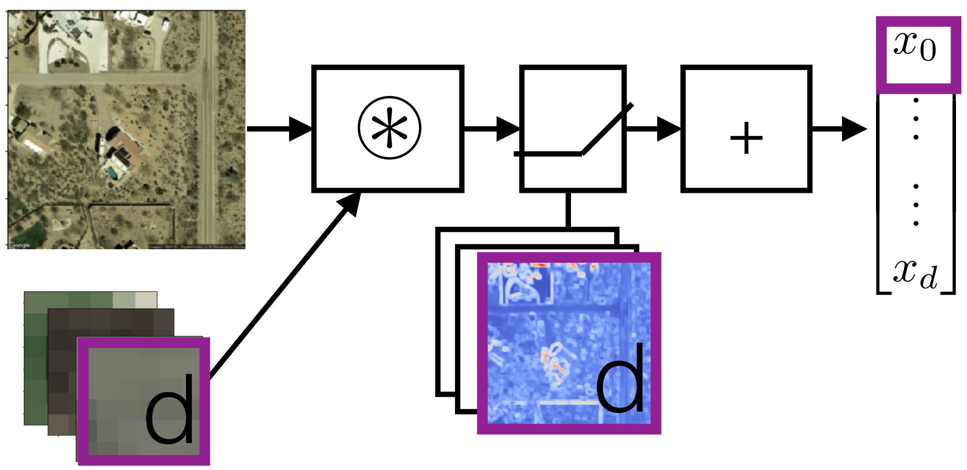


super-resolution

$$\boxed{x_j} = \sum_{p \in \text{pixels}} \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)$$

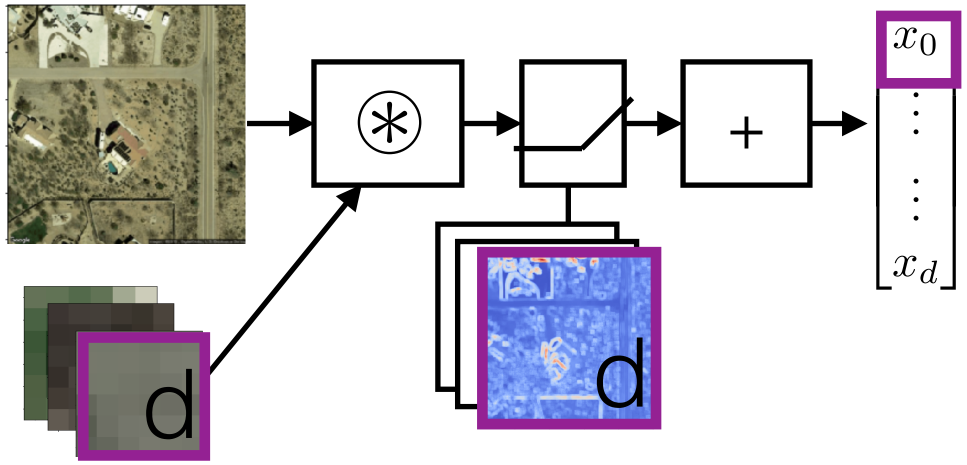
optimal regression weights

$$\widehat{y} = x^\top w$$



super-resolution

$$\boxed{x_j} = \sum_{p \in \text{pixels}} \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)$$



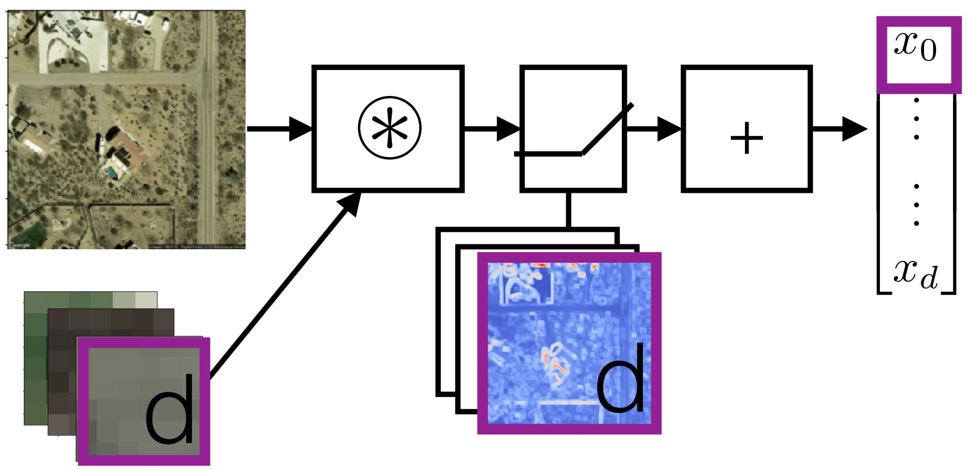
optimal regression weights

$$\hat{y} = x^\top w$$

$$= \sum_{j=1}^{d} w_j x_j$$

super-resolution

$$\boxed{x_j} = \sum_{p \in \text{pixels}} \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)$$



optimal regression weights

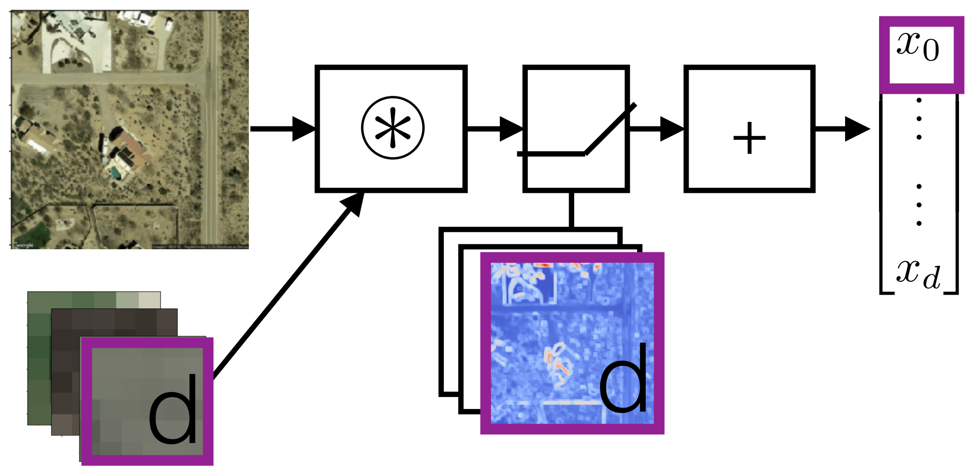
$$\hat{y} = x^\top w$$

$$= \sum_{j=1}^{d} w_j x_j$$

$$= \sum_{j=1}^{d} w_j \sum_{p \in \text{pixels}} \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)$$

super-resolution

$$\boxed{x_j} = \sum_{p \in \text{pixels}} \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)$$

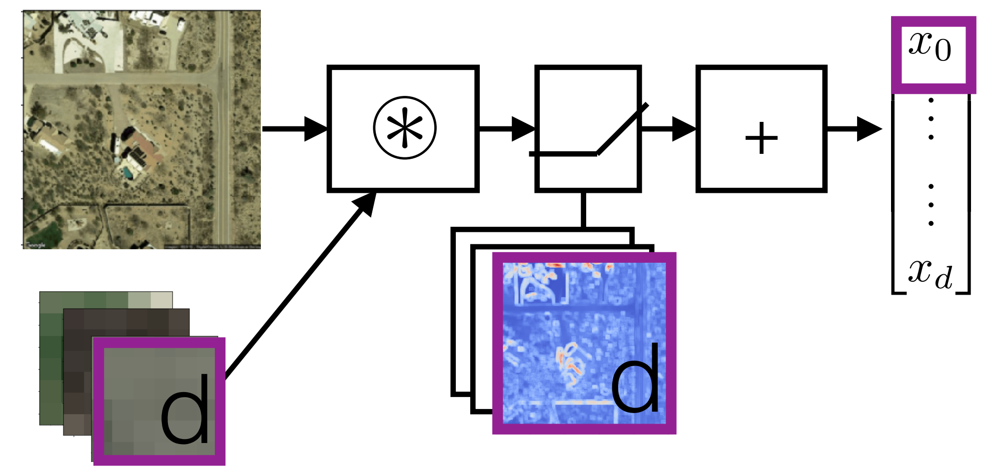


optimal regression weights

$$\begin{aligned}
\widehat{y} &= x^\top w \\
&= \sum_{j=1}^{d} w_j x_j \\
&= \sum_{j=1}^{d} w_j \sum_{p \in \text{pixels}} \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right) \\
&= \sum_{p \in \text{pixels}} \sum_{j=1}^{d} \left(w_j \cdot \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)\right)
\end{aligned}$$

super-resolution

$$\boxed{x_j} = \sum_{p \in \text{pixels}} \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)$$



optimal regression weights

$$\widehat{y} = x^\top w$$

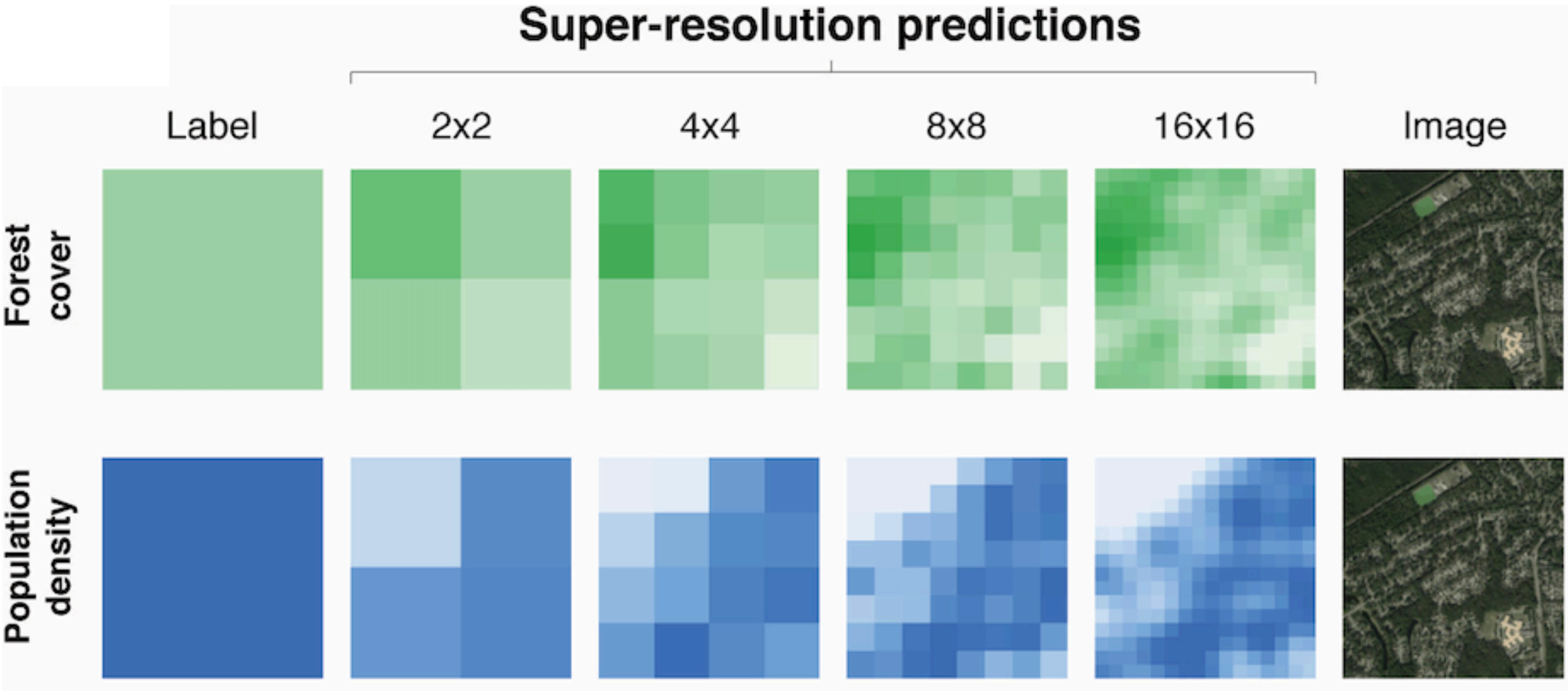
$$= \sum_{j=1}^{d} w_j x_j$$

$$= \sum_{j=1}^{d} w_j \sum_{p \in \text{pixels}} \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)$$

$$= \sum_{p \in \text{pixels}} \boxed{\sum_{j=1}^{d} \left(w_j \cdot \text{ReLu}\left(\text{Image}[p] \circledast \text{Filter}[j]\right)\right)}$$

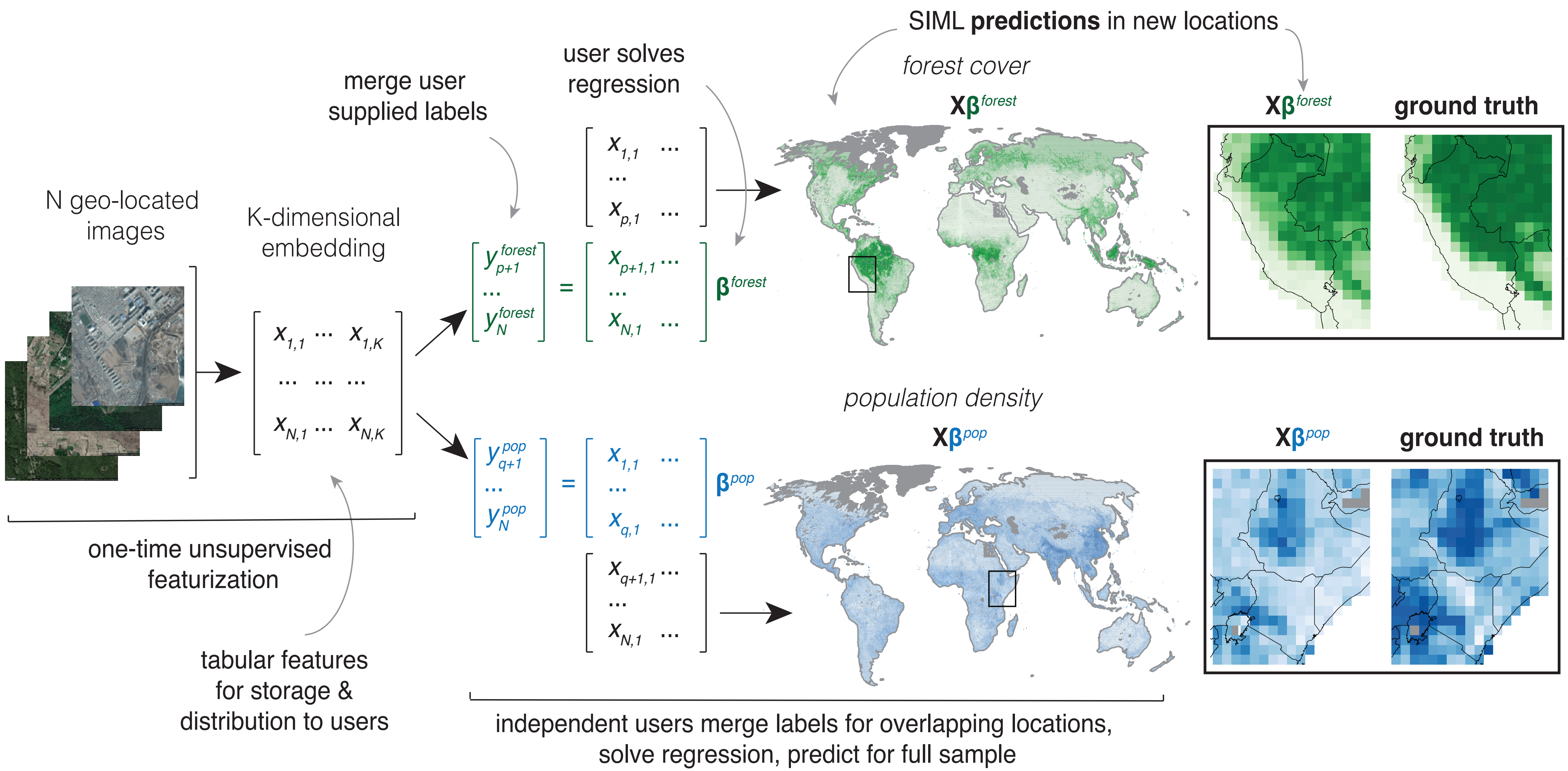
per-pixel predictions!
a sum of filter responses
scaled by regression
weights

Bonus: “super-label-resolution predictions”



Esther Rolf (UC Berkeley)

Recap



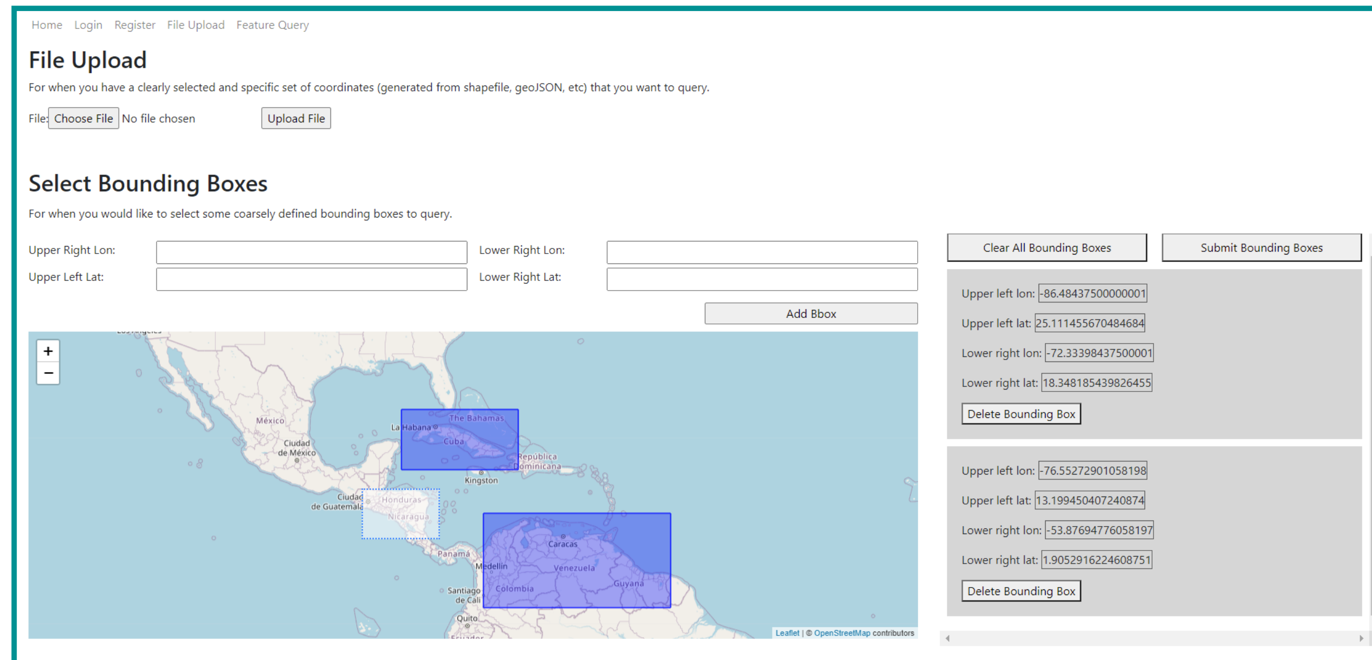
N geo-located images
K-dimensional embedding
merge user supplied labels
user solves regression
SIML **predictions** in new locations
forest cover
Xβ *forest*
ground truth
population density
Xβ *pop*
ground truth
one-time unsupervised featurization
tabular features for storage & distribution to users
independent users merge labels for overlapping locations, solve regression, predict for full sample

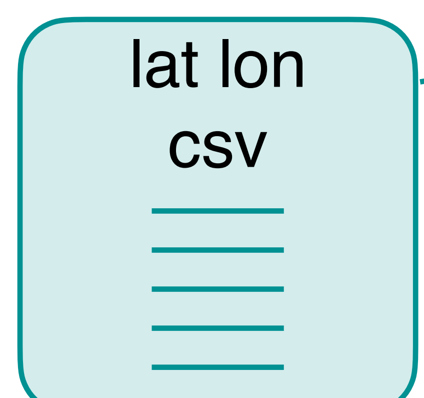
In progress: a public API where users can query for features to run their own scientific analyses.

web-based UI

RCF

$$\begin{bmatrix} \text{---}\ x^{(0)}\ \text{---} \\ \vdots \\ \text{---}\ x^{(n)}\ \text{---} \end{bmatrix}$$





Esther Rolf (UC Berkeley)

screenshot by
Alex Gao, UC Berkeley

Summary: unsupervised RCF features enable a generalizable and accessible system for SIML.

What's next:

- Better unsupervised/self-supervised representations for satellite imagery.
- Adapting MOSAIKS design principles to other SIML prediction settings (segmentation, etc).

General SIML resources:

- MOSAIKS code, data, and tutorials: http://www.globalpolicy.science/mosaiks
- Torchgeo (https://torchgeo.readthedocs.io/en/latest/)
- Thorough list of resources at https://github.com/robmarkcole/satellite-image-deep-learning

References

Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., Recht, B. & Hsiang, S.. A Generalizable and Accessible Approach to Machine Learning with Global Satellite Imagery. *Nature Communications (2021).*

Jean, Neal, et al. "Combining satellite imagery and machine learning to predict poverty." *Science* 353.6301 (2016): 790-794.

Mossoux, Sophie, et al. "Mapping population distribution from high resolution remotely sensed imagery in a data poor setting." *Remote Sensing* 10.9 (2018): 1409

Duporge, Isla, et al. "Using very‐high‐resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes." *Remote Sensing in Ecology and Conservation* (2020).

de Bem, Pablo Pozzobon, et al. "Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks." *Remote Sensing* 12.6 (2020): 901.

Rahimi, A., & Recht, B. (2008, December). Weighted sums of random kitchen sinks: replacing minimization with randomization in learning. In *Nips* (pp. 1313-1320).

Coates, A., Ng, A., & Lee, H. (2011, June). An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 215-223). JMLR Workshop and Conference Proceedings.

Jonas, E., Bobra, M., Shankar, V., Hoeksema, J. T., & Recht, B. (2018). Flare prediction using photospheric and coronal image data. *Solar Physics*, *293*(3), 1-22.

Morrow, A., Shankar, V., Petersohn, D., Joseph, A., Recht, B., & Yosef, N. (2017). Convolutional kitchen sinks for transcription factor binding site prediction. *arXiv preprint arXiv:1706.00125*.

Head, A, Manguin, M, Tran, N, and Blumenstock, JE (2017). Can Human Development be Measured with Satellite Imagery?, *Proceedings of the 9th IEEE/ACM International Conference on Information and Communication Technologies and Development (ICTD 2017)*

Burke, M., Driscoll, A., Lobell, DB., and Ermon, S. Using satellite imagery to understand and promote sustainable development. *Science (2021)*