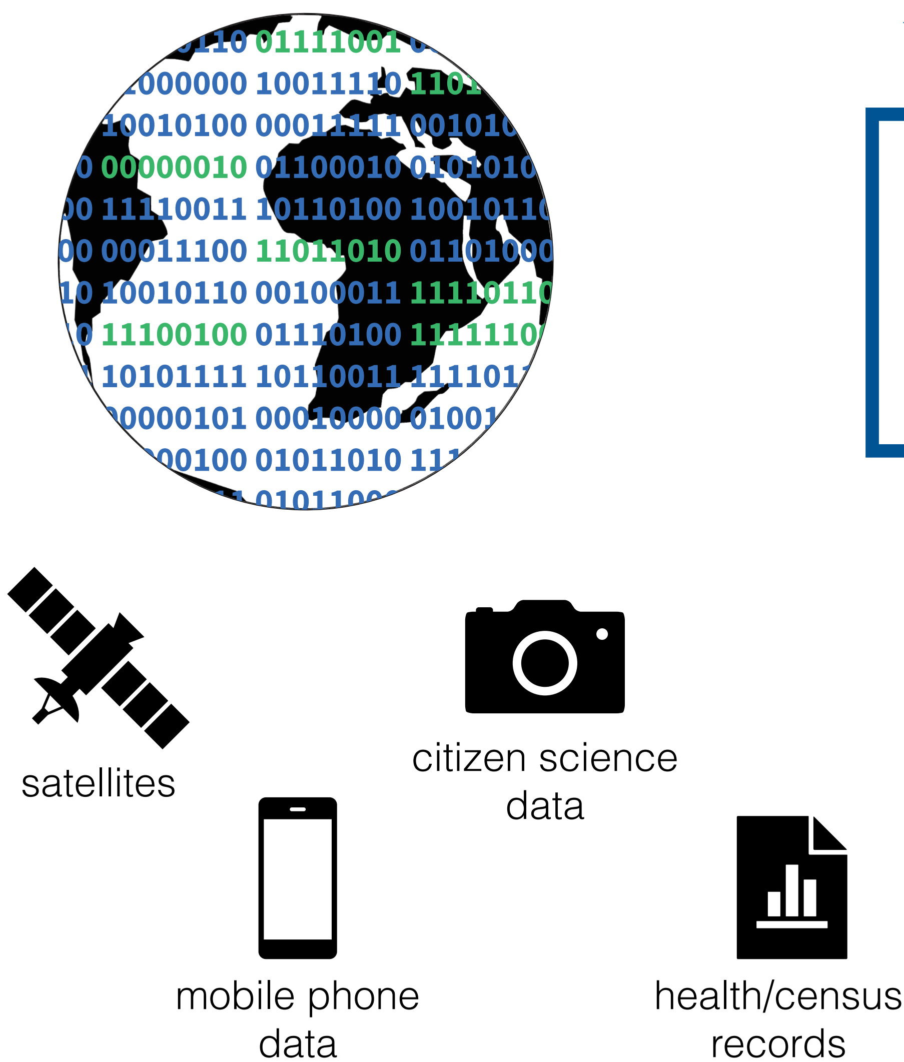
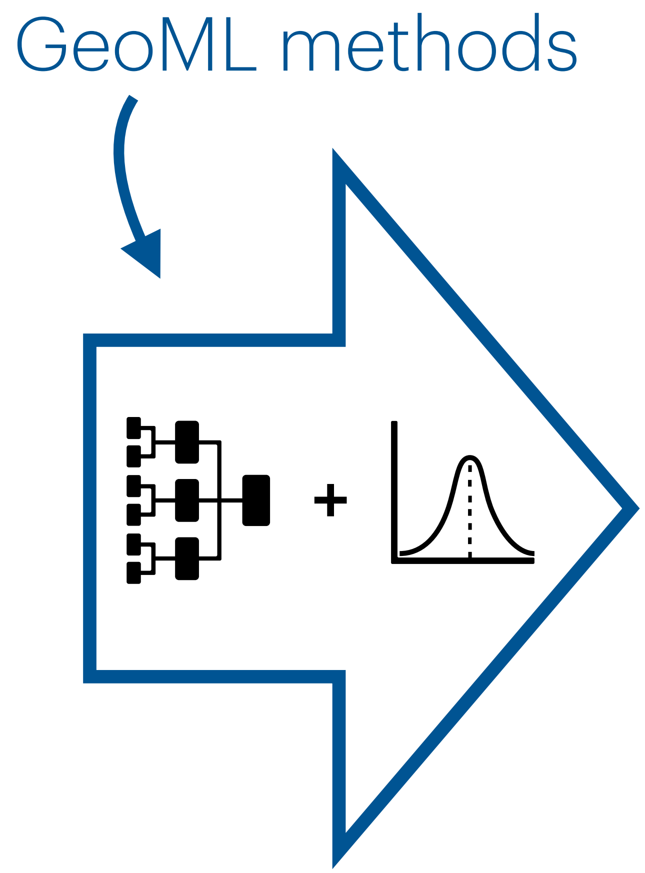
We need to talk (more) about uncertainty in geospatial machine learning

Esther Rolf
Assistant Professor,
CU Boulder Computer Science

Workshop on Bayesian Decision-making and Uncertainty at NeurIPS, Dec 14, 2024

ML can transform environmental monitoring by extracting crucial information from geospatial data





Mapping tree canopy height and its connection to biodiversity.

Detecting artisanal mining and its connection to health outcomes.

1000+ earth observation satellites collect over **90TB** data / day.

Combining satellite imagery and machine learning (**SatML**) can help researchers and policymakers monitor our world and act in it.

Combining satellite imagery and machine learning (**SatML**) can help researchers and policymakers monitor our world and act in it.

Environmental monitoring

- tracking species populations
- mapping built infrastructure
- biodiversity mapping

Article

Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks

Pablo Pozzobon de Bem, Osmar Abílio de Carvalho Junior *, Renato Fontes Guimarães and Roberto Arnaldo Trancoso Gomes

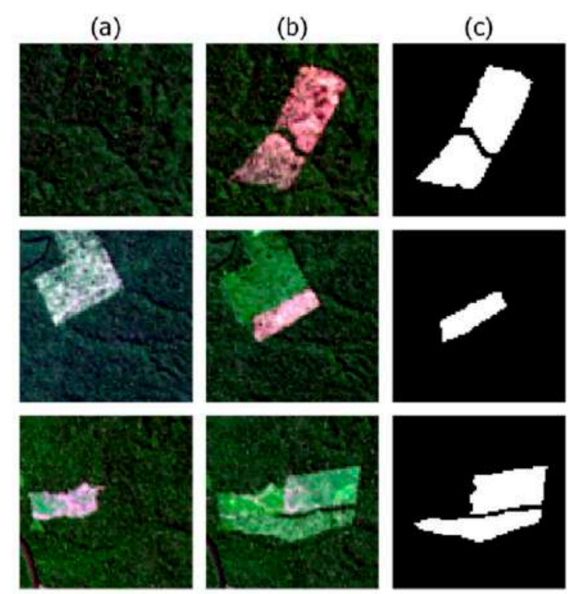


Figure 4. Example of the change mapping in three locations between **(a)** 2017 and **(b)** 2018 and the respective **(c)** rasterized deforestation mask.

Figure from de Bem et al., Remote Sensing 2020.

Combining satellite imagery and machine learning (**SatML**) can help researchers and policymakers monitor our world and act in it.

Environmental monitoring

- tracking species populations
- mapping built infrastructure
- biodiversity mapping

Disaster response

- estimating damages from natural disasters with building detection
- finding and prioritizing the most vulnerable or affected areas

Article: https://www.technologyreview.com/2023/02/20/1068824/ai-actually-helpful-disaster-response-turkey-syria-earthquake/

Combining satellite imagery and machine learning (**SatML**) can help researchers and policymakers monitor our world and act in it.

Environmental monitoring

- tracking species populations
- mapping built infrastructure
- biodiversity mapping

Disaster response

- estimating damages from natural disasters with building detection
- finding and prioritizing the most vulnerable or affected areas

Global policy

- food security / yield prediction
- fine grained poverty estimates

Microestimates of wealth for all low- and middle-income countries

Guanghua Chi[a,1,2], Han Fang[b], Sourav Chatterjee[b], and Joshua E. Blumenstock[a,1,3]

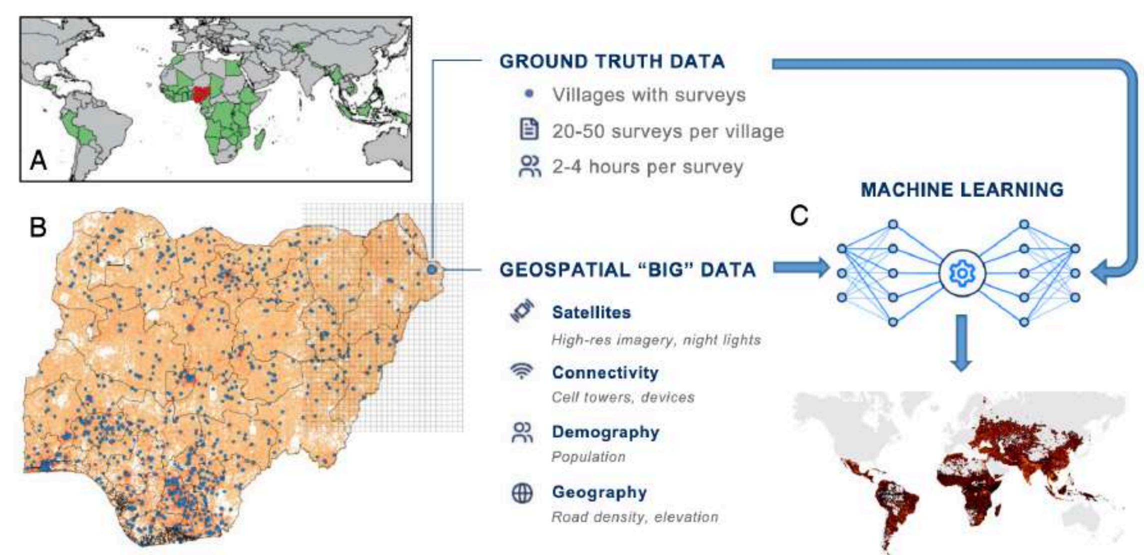
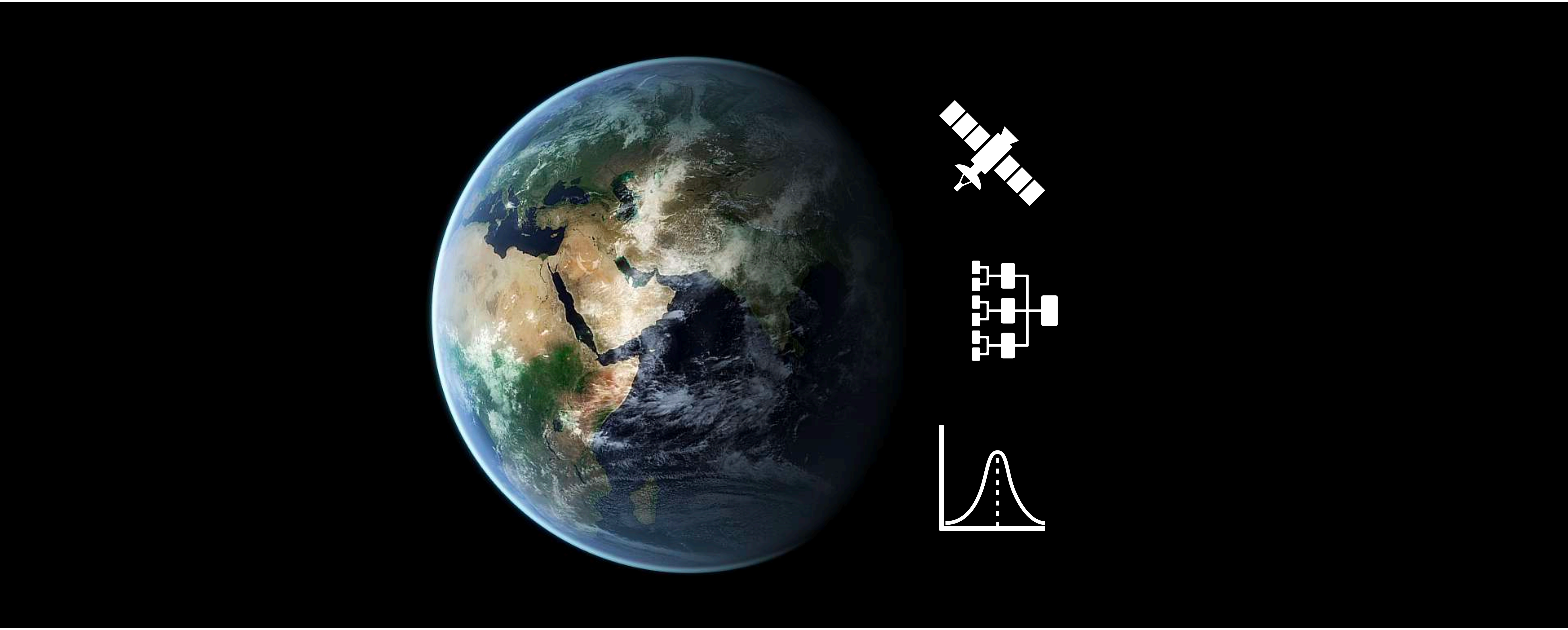


Fig. 2. Overview of approach. (*A*) Nationally representative household survey data are obtained from 56 different countries around the world. (*B*) In Nigeria, for example, there are 40,680 households surveyed in 899 unique survey locations ("villages"). Geospatial "big" data from satellites and other existing sensors are also sourced from each location. (*C*) These data are used to train a machine-learning algorithm that predicts microregional poverty from nontraditional data, even in regions where no ground-truth data exists.

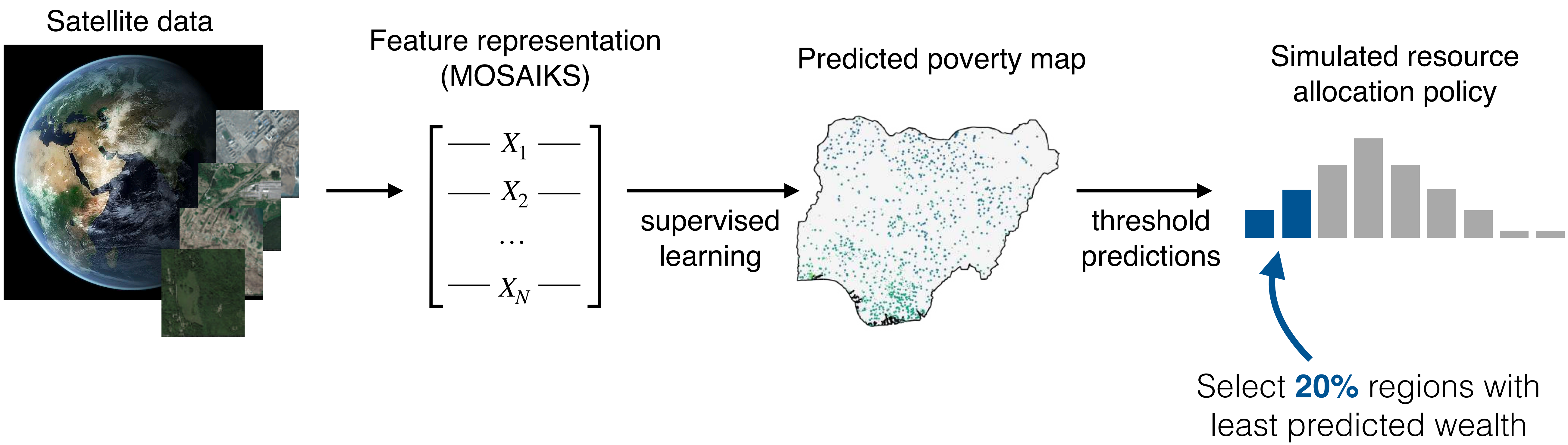
Figure from Chi et al., PNAS 2022

It is easier than ever to make maps with satellite imagery and machine learning…



… it is crucial that we understand and convey the limitations and uncertainty of mapped predictions.

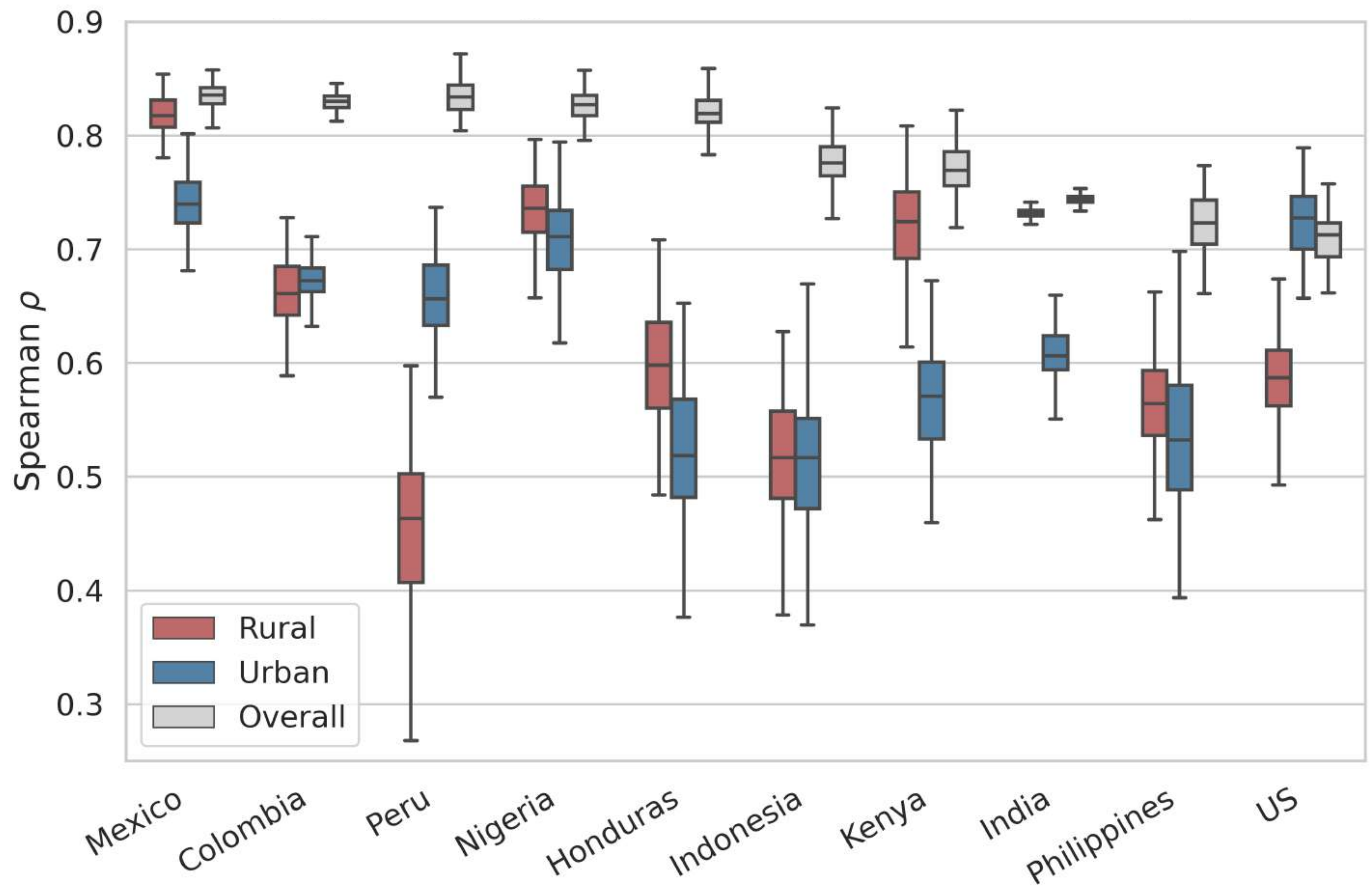
Example: (simplified) poverty prediction with SatML



Aiken*, Rolf*, Blumenstock. *Fairness and representation in satellite-based poverty maps: Evidence of urban-rural disparities and their impacts on downstream policy.* IJCAI 2023.



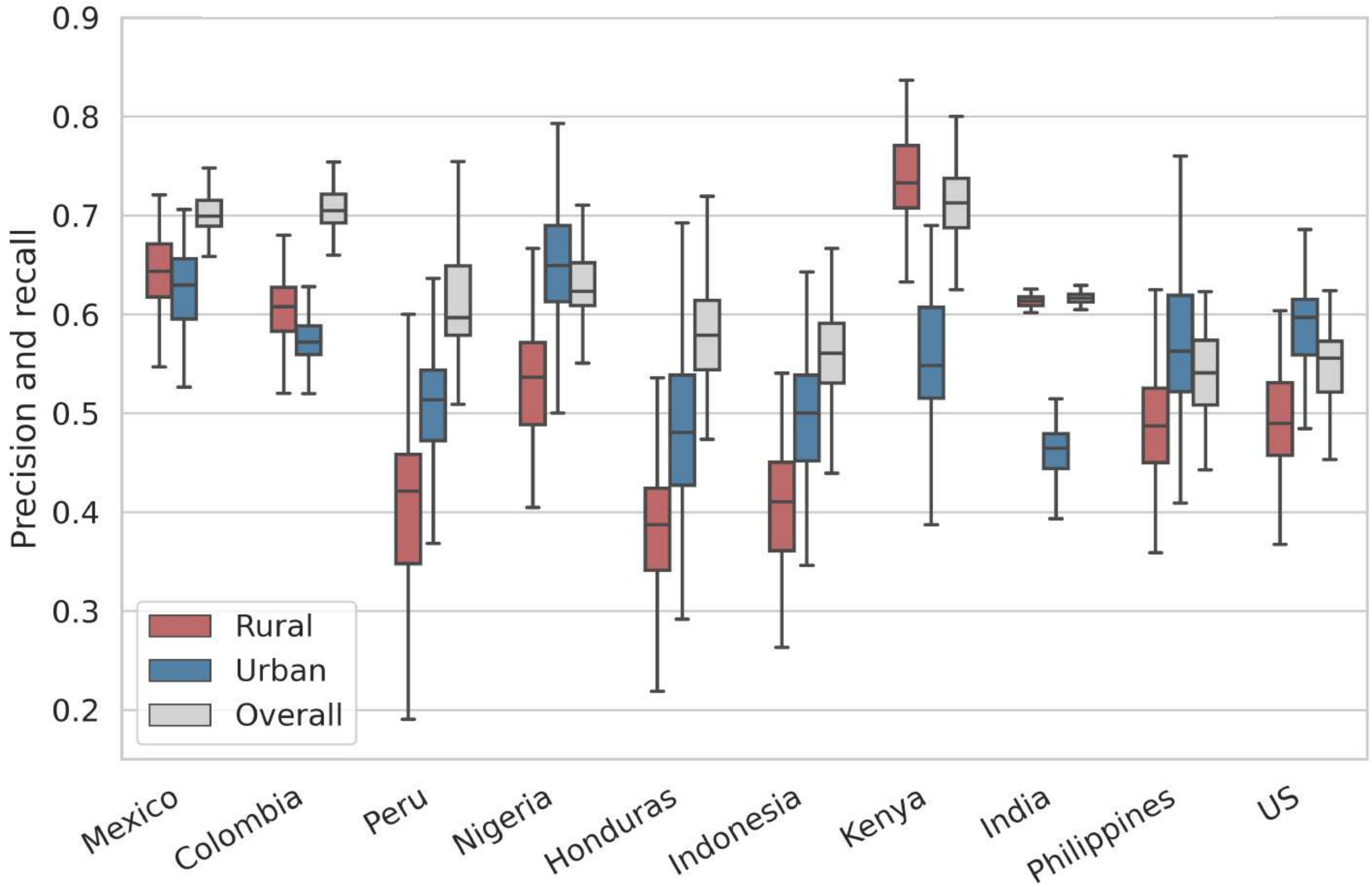
Prediction performance is lower if evaluate only within rural or urban regions vs. in each country as a whole.



This systematically replicates analysis in [1] for ten countries across the globe.

[1] Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nature communications, 2020.

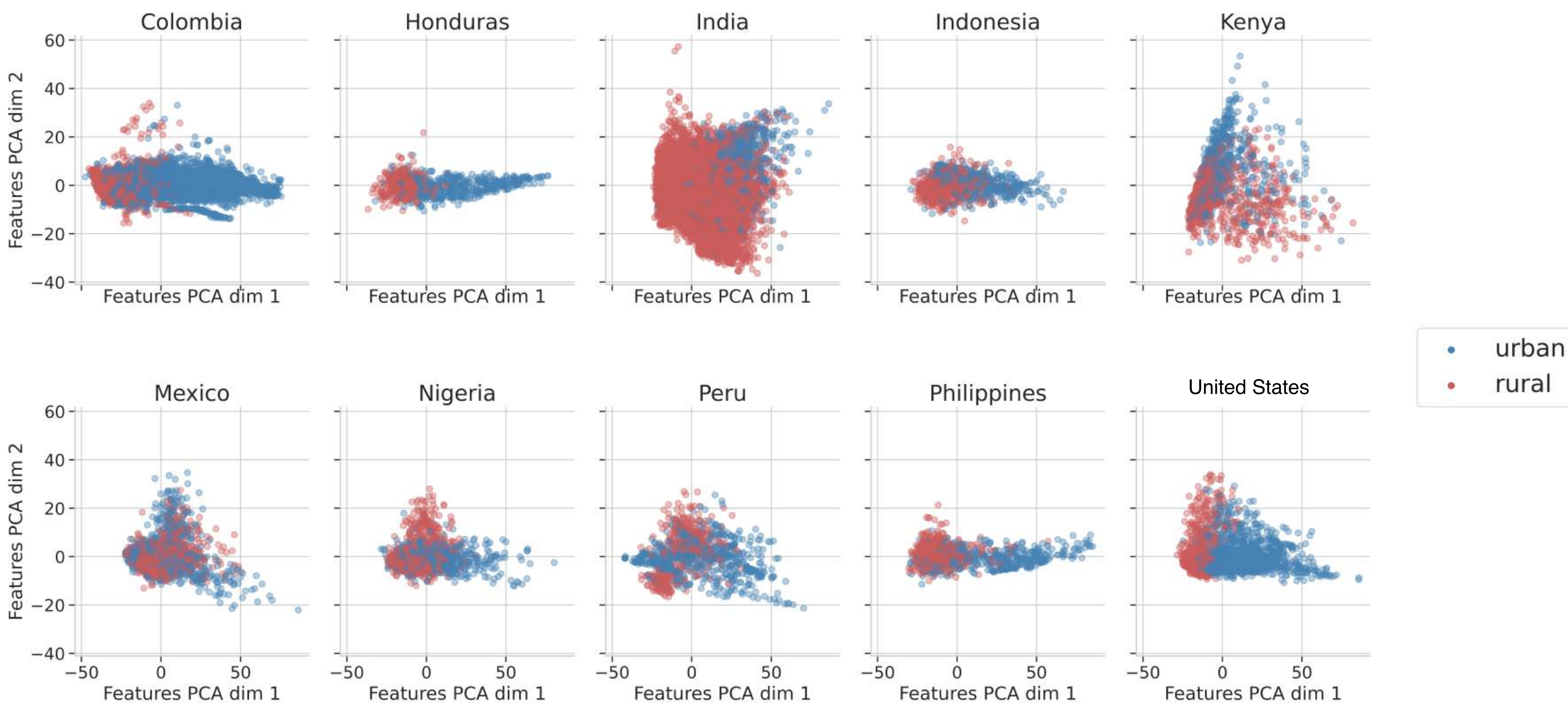
Targeting effectiveness is lower if aid program allocates resources just within rural or urban areas.

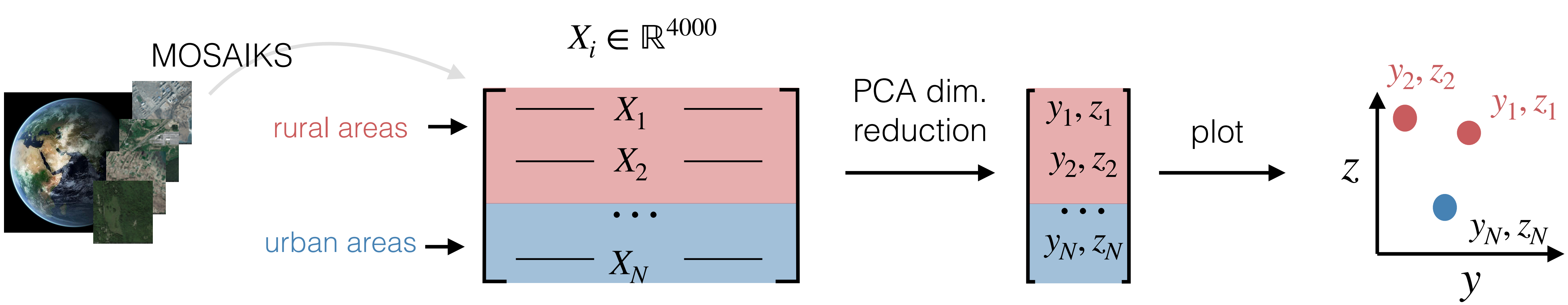


What information does satellite imagery convey about relative wealth across regions?

Sentinel 2 satellite imagery accessed via Microsoft Planetary Computer

Image embedding distributions: urban and rural

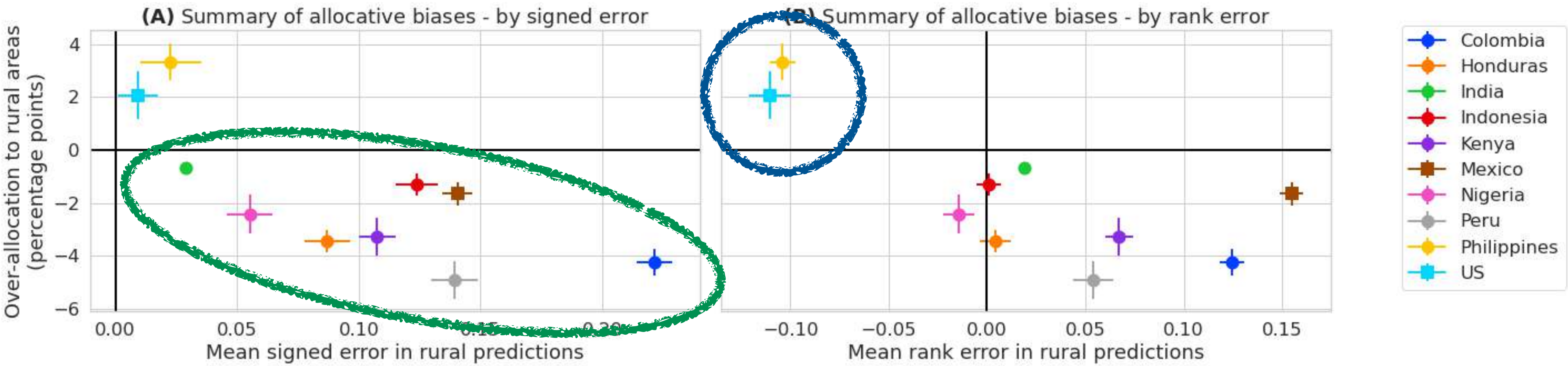




Two (competing) phenomena contribute to allocational disparities:

1) (Over)reliance on correlation between wealth and urbanization

2) Reversion of predictions to the population means



Satellite data only gives a partial representation of many ground phenomena we wish to map

➡ Systematic errors in predictions

Satellite data only gives a partial representation of many ground phenomena we wish to map

➡ Systematic errors in predictions

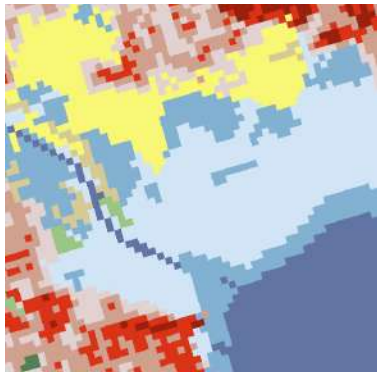
What if we had a more structured, uncertainty based approach for learning with geospatial data?

➡ Ideally, multiple modalities of geospatial data

Often in GeoML, data do not provide appropriate "ground truth," but indirect guidance on label values.

Coarse/incomplete landcover observations

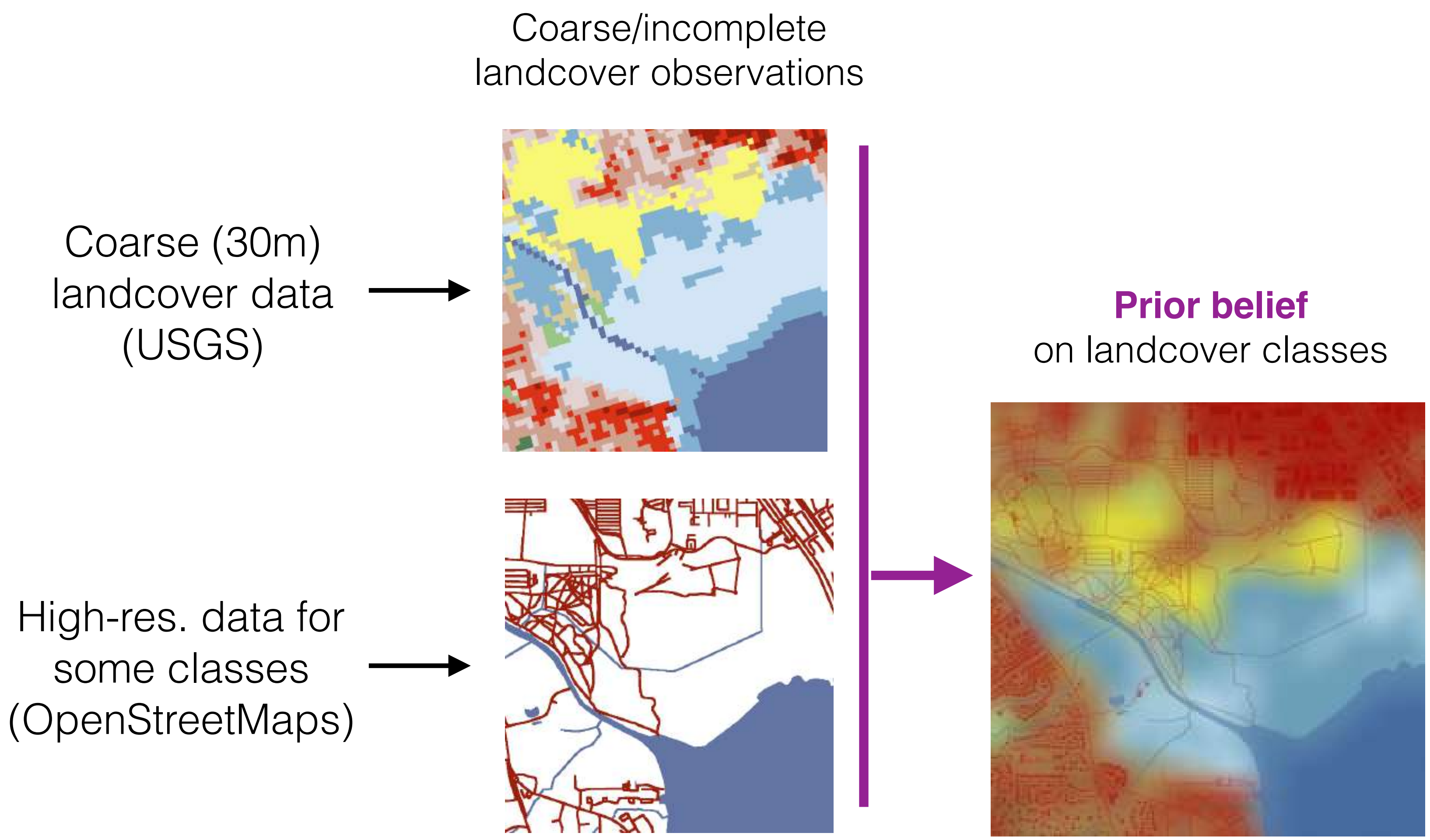
Coarse (30m) landcover data (USGS) →



High-res. data for some classes (OpenStreetMaps) →

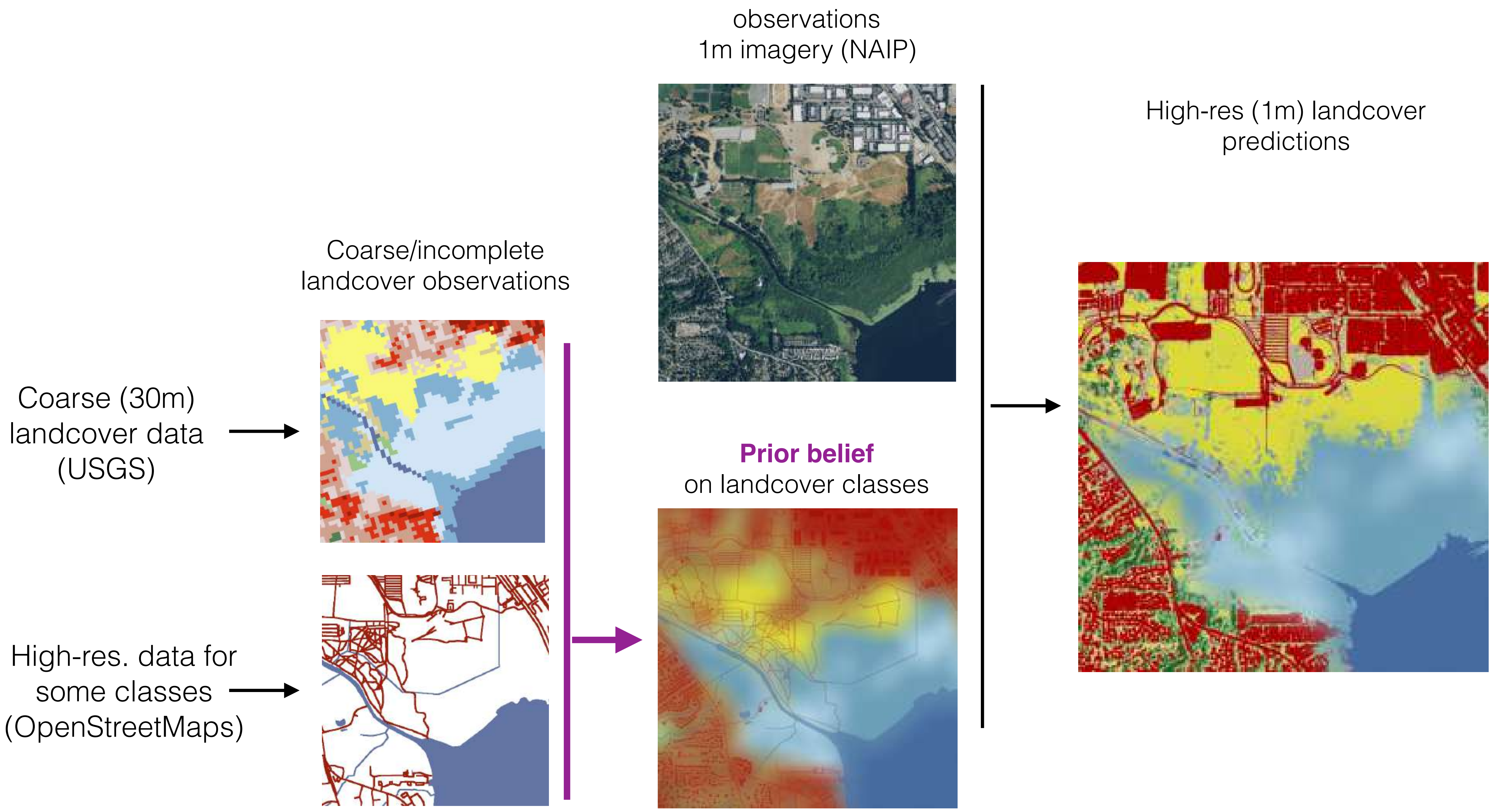
Rolf*, Malkin*, Graikos, Jojic, Robinson, Jojic. *Resolving label uncertainty with implicit posterior models,* UAI 2022.

Often in GeoML, data do not provide appropriate "ground truth," but indirect guidance on label values (→ a prior belief).



Rolf*, Malkin*, Graikos, Jojic, Robinson, Jojic. *Resolving label uncertainty with implicit posterior models,* UAI 2022.

Often in GeoML, data do not provide appropriate "ground truth," but indirect guidance on label values (→ **a prior belief**).

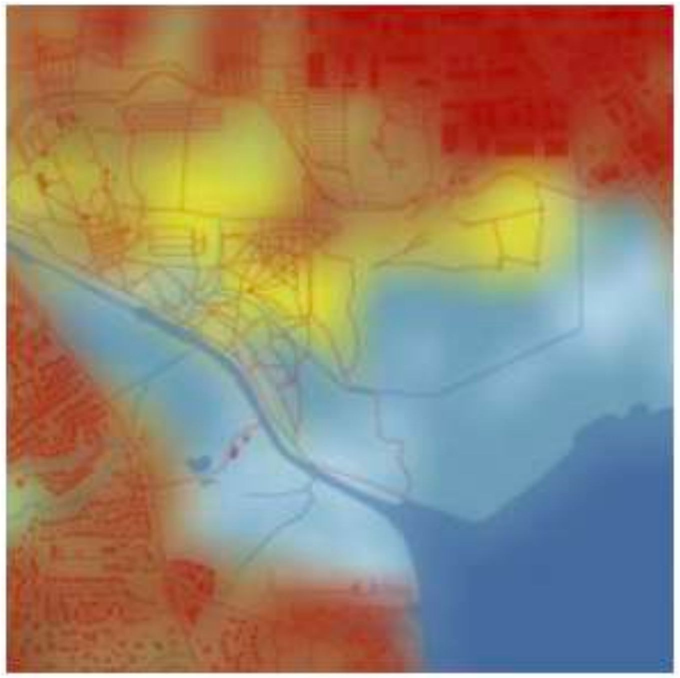


Learning from uncertain labels is hard:

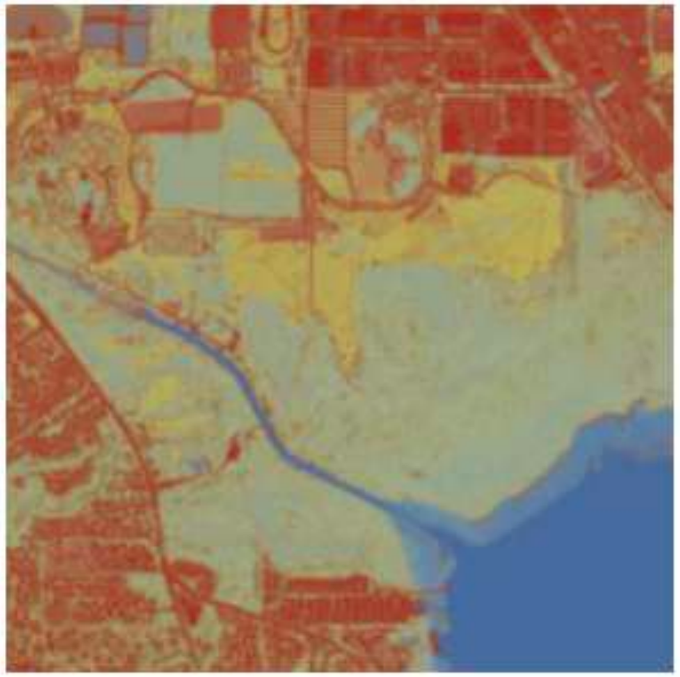
Supervised learning:

- uncertain labels → uncertain predictions
- ✓ amenable to rich model classes
- ✓ simple training w/ standard loss fxns

Prior belief
on landcover classes



Predictions from model
trained with CE loss on prior



Learning from uncertain labels is hard:

Supervised learning:

- uncertain labels → uncertain predictions
- ✓ amenable to rich model classes
- ✓ simple training w/ standard loss fxns

Generative modeling:

- ✓ high certainty in posterior w/ soft priors
- ✓ opportunities to model rich structure in the prior beliefs
- typically more expensive to train (requires sampling, 2x parameters)

Learning from uncertain labels is hard:

Supervised learning:

✓ amenable to rich model classes

✓ simple training w/ standard loss fxns

Generative modeling:

✓ high certainty in posterior w/ soft priors

✓ opportunities to model rich structure in the prior beliefs

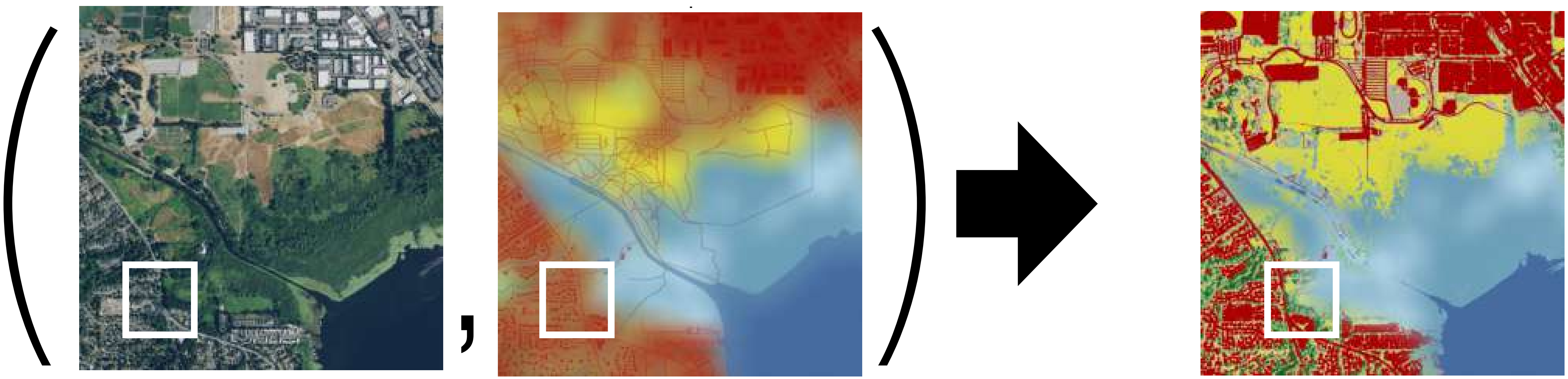
Our approach: match output of supervised learning model with a generative model involving provided prior belief

➡ merges flexibility of generative modeling with ease of supervised learning

Setting: learning from a prior belief

Goal: use observations of (x_i, p_i) pairs to disambiguate uncertainty in prior!

$\{x_i\}_{i \in \text{pixels}}$ $\{p_i\}_{i \in \text{pixels}}$ $p(\ell \mid x_i) \propto p(x_i \mid \ell) p_i(\ell)$



classes ℓ :

- Open Water
- Developed Open Space
- Developed Low Intensity
- Developed Med. Intensity
- Developed High Intensity
- Barren Land
- Pasture/Hay
- Evergreen Forest
- Mixed Forest
- Woody Wetlands
- Emergent Herbaceous Wetlands

Optimizing implicit posterior models

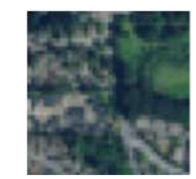
Assume generative model $p(x \mid \ell)$ exists, but **unknown,** then posterior is

$$p(\ell \mid x_i) = c_i \cdot p(x_i \mid \ell) p_i(\ell)$$

Notation

ℓ : classes

x_i : observations



p_i : priors over ℓ

c_i : normalizing constant

Optimizing implicit posterior models

Assume generative model $p(x \mid \ell)$ exists, but **unknown,** then posterior is

$$p(\ell \mid x_i) = c_i \cdot p(x_i \mid \ell) p_i(\ell)$$

Estimate posterior distribution with a parametrized model $q(\ell \mid x_i; \theta)$, then we can minimize:

$$\boxed{\sum_i \mathrm{KL}\left(q(\ell \mid x_i; \theta) \| c_i \cdot \underbrace{p(x_i \mid \ell)}_{??} p_i(\ell)\right) \quad (\star)}$$

Notation

ℓ : classes

x_i : observations

p_i : priors over ℓ

c_i : normalizing constant

$q_i : q(\ell \mid x_i; \theta)$ (e.g., neural net)

Optimizing implicit posterior models

Assume generative model $p(x \mid \ell)$ exists, but **unknown,** then posterior is

$$p(\ell \mid x_i) = c_i \cdot p(x_i \mid \ell) p_i(\ell)$$

Estimate posterior distribution with a parametrized model $q(\ell \mid x_i; \theta)$, then we can minimize:

$$\boxed{\sum_i \mathrm{KL}\left(q(\ell \mid x_i; \theta) \| c_i \cdot p(x_i \mid \ell) p_i(\ell)\right) \quad (\star)}$$

Fix $q(\ell \mid x_i; \theta)$ $\longrightarrow$ $\underset{p(x_i \mid \ell)}{\arg\min} \; (\star)$ s.t. $\sum_i p(x_i \mid \ell) \leq 1$ $= \dfrac{q(\ell \mid x_i; \theta)}{\sum_j q(\ell \mid x_j; \theta)}$

Let direct model **imply** $p(x_i \mid \ell)$

approximates all existing data with $i \in$ batch

Notation

ℓ : classes

x_i : observations

p_i : priors over ℓ

c_i : normalizing constant

$q_i : q(\ell \mid x_i; \theta)$ (e.g., neural net)

Optimizing implicit posterior models

$$\arg\min_{\theta} \sum_{i} \mathrm{KL}\left(q(\ell \mid x_i; \theta) \,\middle\|\, c_i \frac{q(\ell \mid x_i; \theta)}{\sum_j q(\ell \mid x_j; \theta)} p_i(\ell) \right)$$

Notation

ℓ : classes

x_i : observations

p_i : priors over ℓ

c_i : normalizing constant

$q_i : q(\ell \mid x_i; \theta)$

(e.g., neural net)

Optimizing implicit posterior models

$$\arg\min_{\theta} \sum_{i} \mathrm{KL}\left(q(\ell \mid x_i; \theta) \,\middle\|\, c_i \frac{q(\ell \mid x_i; \theta)}{\sum_j q(\ell \mid x_j; \theta)} p_i(\ell) \right)$$

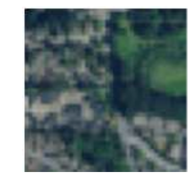
Loss function in 2 lines (PyTorch):

```
def qr_loss(log_q, prior):
    log_r = (log_q.log_softmax(0) + prior.log()).log_softmax(1)
    return (log_q * log_q.exp()).sum(1) - (log_r * log_q.exp()).sum(1)
```

Notation

ℓ : classes

x_i : observations

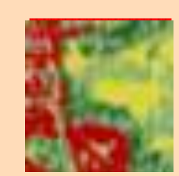


p_i : priors over ℓ

c_i : normalizing constant

$q_i : q(\ell \mid x_i; \theta)$

(e.g., neural net)

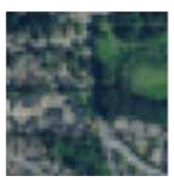


Optimizing implicit posterior models

$$\arg\min_{\theta} \sum_{i} \mathrm{KL}\left(q(\ell \mid x_i;\theta) \,\middle\|\, c_i \frac{q(\ell \mid x_i;\theta)}{\sum_j q(\ell \mid x_j;\theta)} p_i(\ell) \right)$$

Two outputs:

q_i : direct model output of the variational posterior

$q($  $;\theta) =$ 

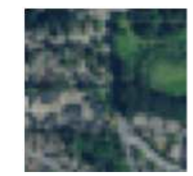
r_i : implied posterior of q_i and p_i

$z(q_i) \times$  $=$ 

Notation

ℓ : classes

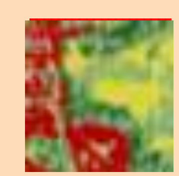
x_i : observations



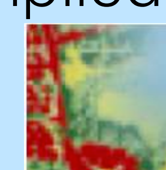
p_i : priors over ℓ

c_i : normalizing constant

$q_i : q(\ell \mid x_i;\theta)$ (e.g., neural net)



r_i : implied posterior

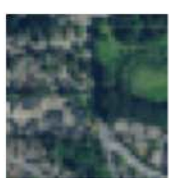


Optimizing implicit posterior models

$$\arg\min_{\theta} \sum_i \mathrm{KL}\left(q(\ell \mid x_i; \theta) \,\middle\|\, c_i \frac{q(\ell \mid x_i; \theta)}{\sum_j q(\ell \mid x_j; \theta)} p_i(\ell) \right)$$

Two outputs:

q_i : direct model output of the variational posterior

$q($  $; \theta) =$ 

r_i : implied posterior of q_i and p_i

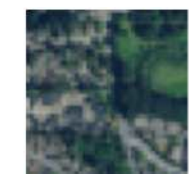
$z(q_i) \times$  $=$ 

At model convergence, r_i and q_i should agree to the amount possible given data and model class

Notation

ℓ : classes

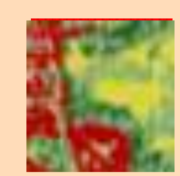
x_i : observations



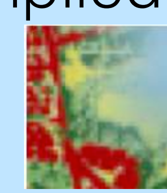
p_i : priors over ℓ

c_i : normalizing constant

$q_i : q(\ell \mid x_i; \theta)$ (e.g., neural net)



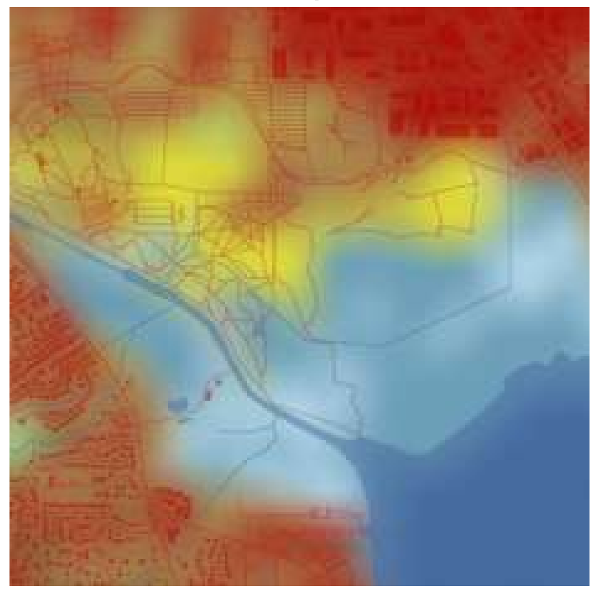
r_i : implied posterior



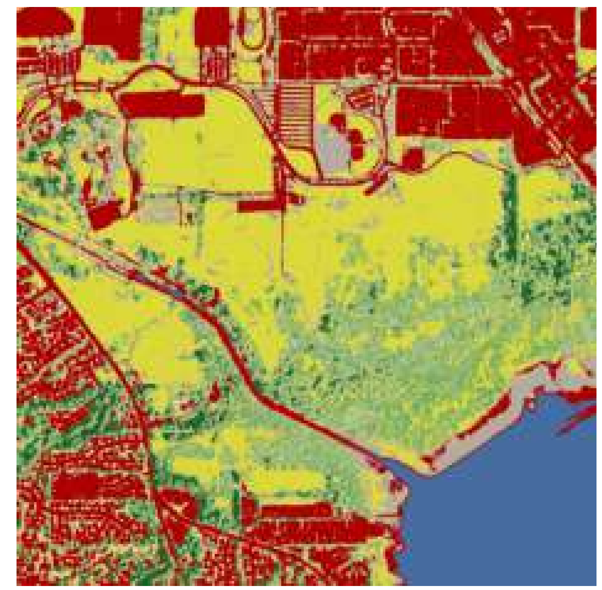
Example: land-cover super-resolution

Observations (x_i) : 1m-resolution

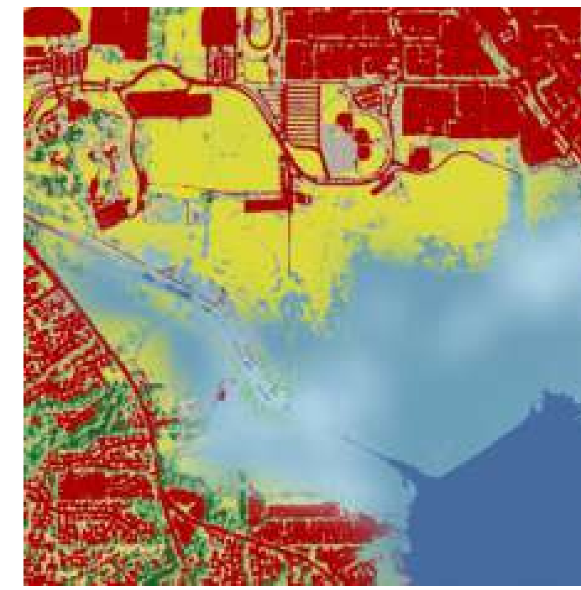
Prior beliefs (p_i) :



Direct model output (q_i) :



Implied posterior (r_i) :



NLCD Legend

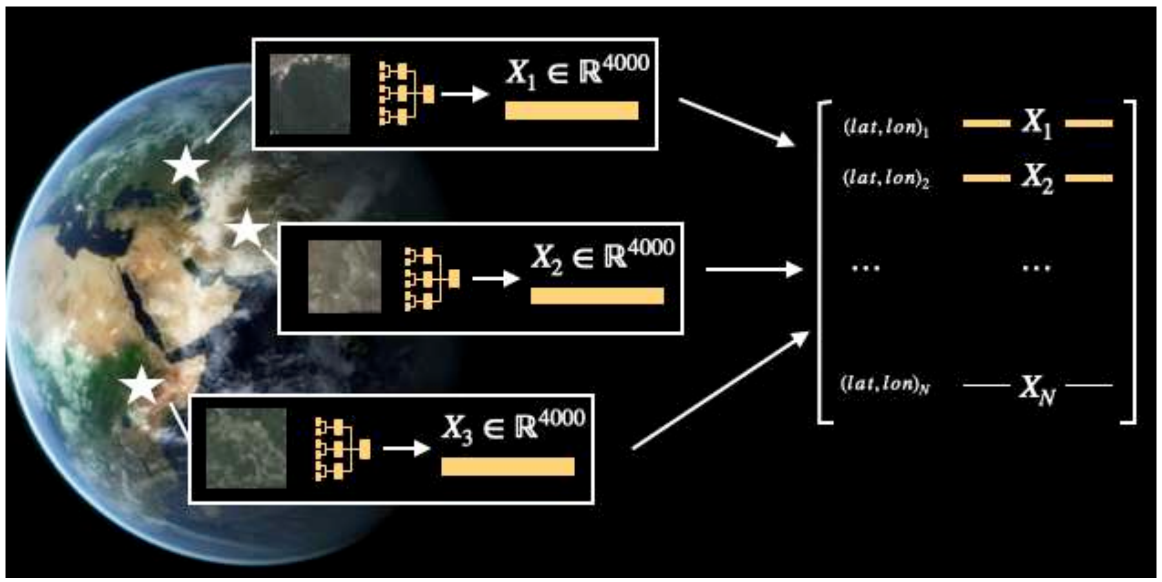
- Open Water
- Developed Open Space
- Developed Low Intensity
- Developed Med. Intensity
- Developed High Intensity
- Barren Land
- Pasture/Hay
- Evergreen Forest
- Mixed Forest
- Woody Wetlands
- Emergent Herbaceous Wetlands

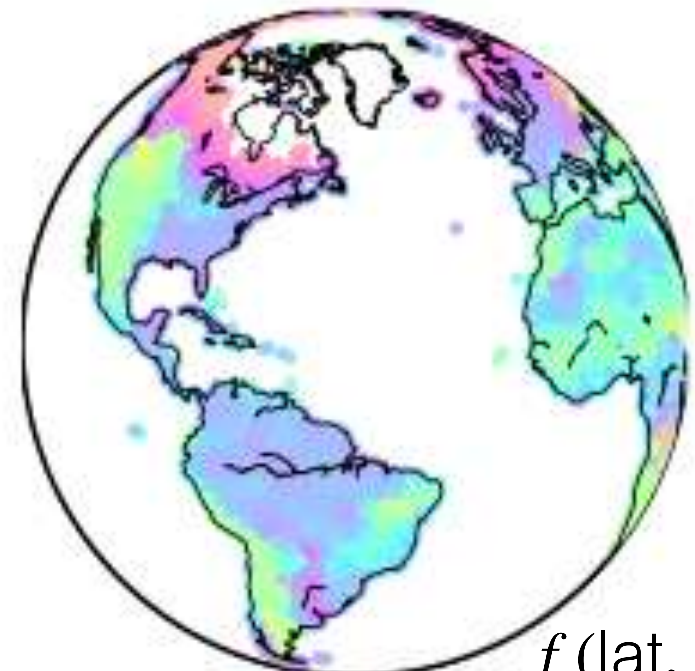
Highlighted results:

- On benchmark land cover super-resolution dataset, QR achieves **72.1% IoU** vs. 59.7% by treating low-resolution labels as ground truth (also higher than previous best at 69.7%)
- on a cross-geography domain adaptation problem, QR/RQ allows for **in-sample predictions from weak sources** (NLCD/ OpenStreetmaps), better than applying models trained on **high-resolution** labels in another region

How do you get a good (global) prior?

Global location embeddings: a "latent space" of geospatial data





f_c(lat, lon)

MOSAIKS
precomputed image embeddings on a 0.01° grid

mosaiks.org > access

SatCLIP
pretrained location encoder

https://github.com/microsoft/satclip

A Generalizable and Accessible Approach to Machine Learning with Global Satellite Imagery.
Rolf, Proctor, Carleton, Bolliger, Shankar, Ishihara, Recht, Hsiang. Nature Communications 2021.

SatCLIP: Global, General-Purpose Location Embeddings with Satellite Imagery.
Klemmer, Rolf, Rußwurm, Robinson, Mackey. AAAI 2025

🤔

Why aren’t we talking (enough) about uncertainty in GeoML?

Mission Critical – Satellite Data is a Distinct Modality in Machine Learning

Esther Rolf [*1 2] Konstantin Klemmer [3] Caleb Robinson [4] Hannah Kerner [*5]

Abstract

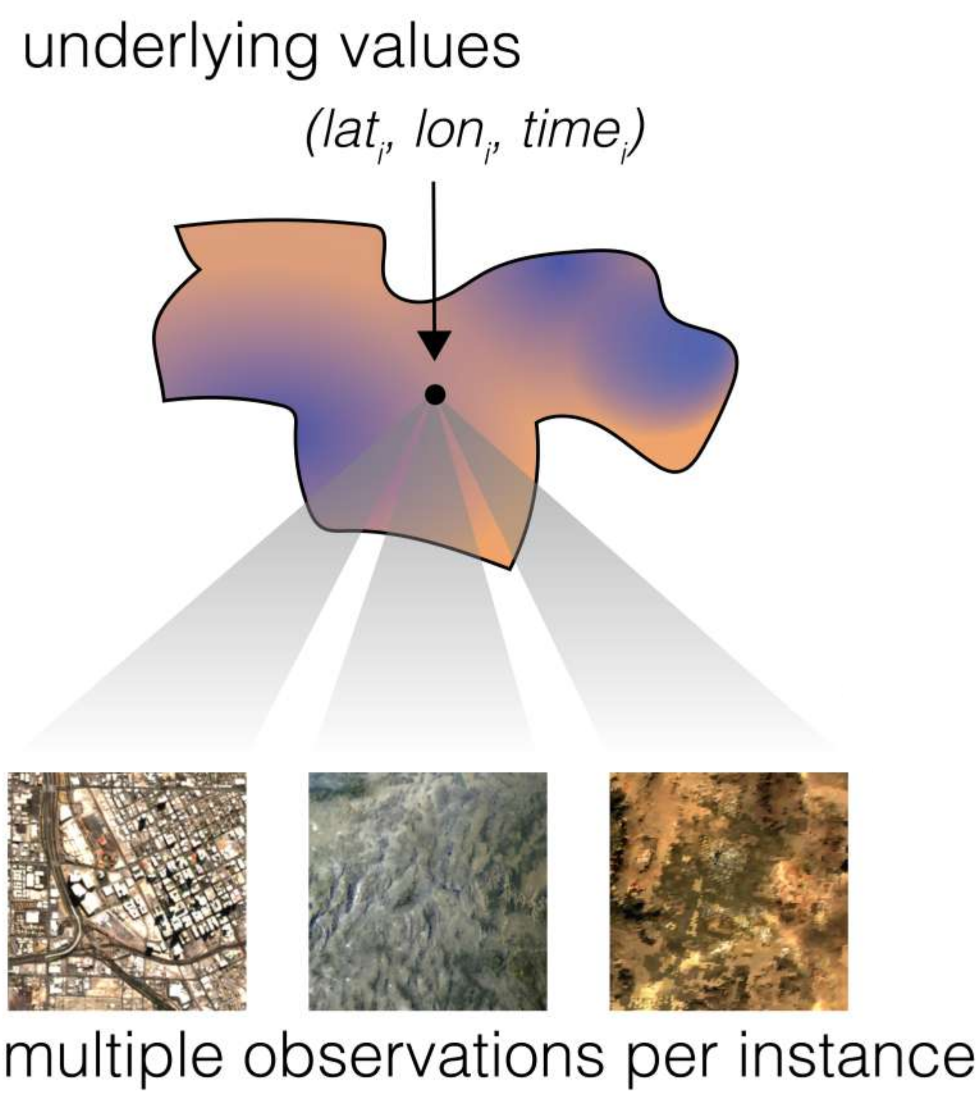
Satellite data has the potential to inspire a seismic shift for machine learning—one in which we rethink existing practices designed for traditional data modalities. As machine learning for satellite data (SatML) gains traction for its real-world impact, our field is at a crossroads. We can either continue applying ill-suited approaches, or we can initiate a new research agenda that centers around the unique characteristics and challenges of satellite data. This position paper argues that satellite data constitutes a distinct modality for machine learning research and that we must recognize it as such to advance the quality and impact of SatML research across theory, methods, and deployment.

Satellite data presents challenges and opportunities distinct from other data modalities (Figure 1). Unlike natural images, the size of targets in satellite images span a logarithmic scale from < 1m (e.g., trees) to > 1km (e.g., forests). Temporal patterns in satellite time series also span logarithmic scales, from hours or days (e.g., floods) to years or decades (e.g., sea level rise). Data are acquired using a variety of sensors that capture diverse spectral channels (beyond 3-channel RGB) and precise measurements (beyond 8 bits). Satellites collect data over the entire surface of the Earth at fixed time intervals and spatial resolutions. Observations are acquired from an overhead perspective from fixed altitudes and lack a "natural" orientation, unlike natural images.

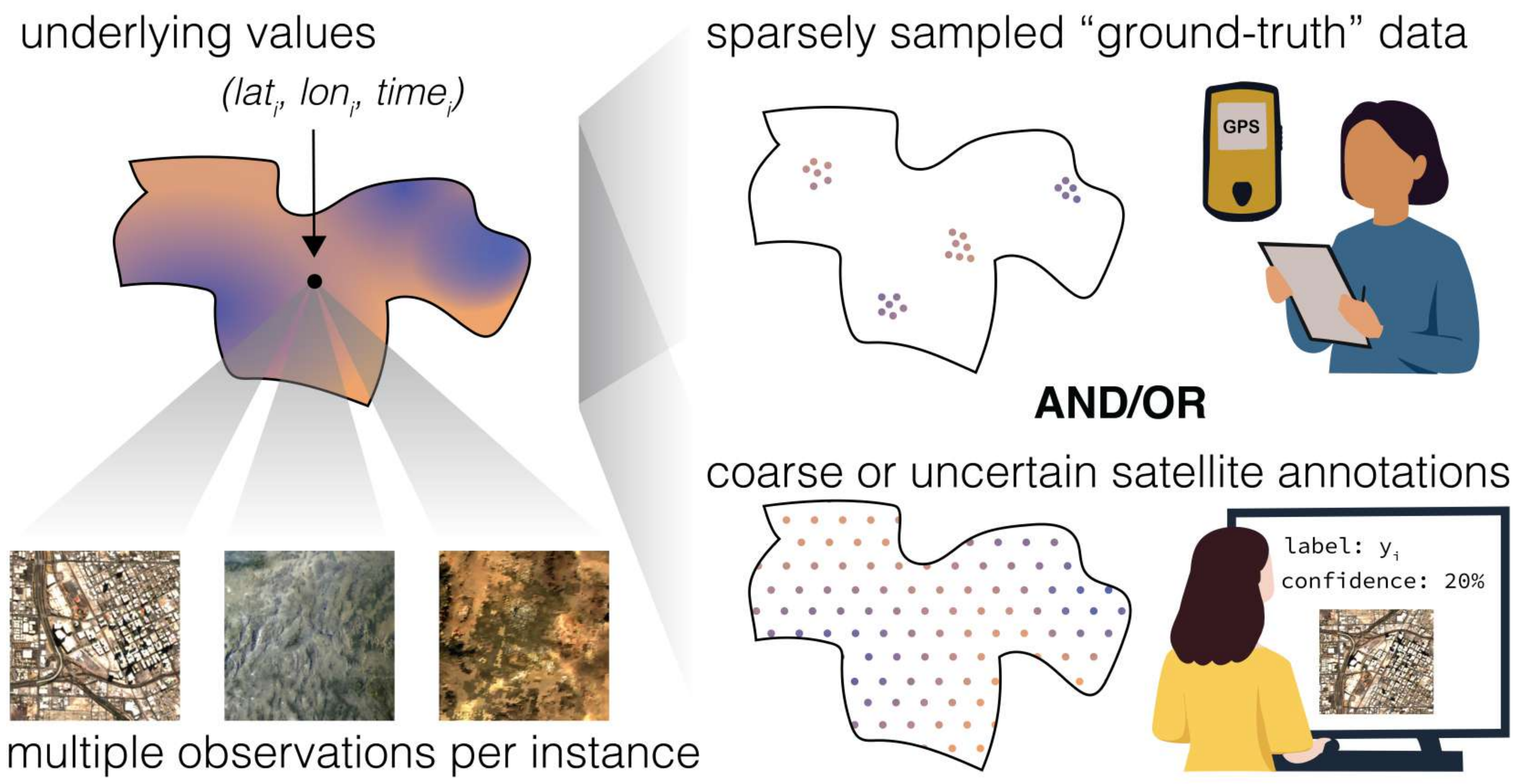
While there has been increasing interest in ML for satellite data (SatML) (Zhu et al., 2017; Table A1), SatML research

Rolf*, Klemmer, Robinson, Kerner*. *Position: Mission critical — Satellite data is a distinct modality in ML,* ICML 2024.

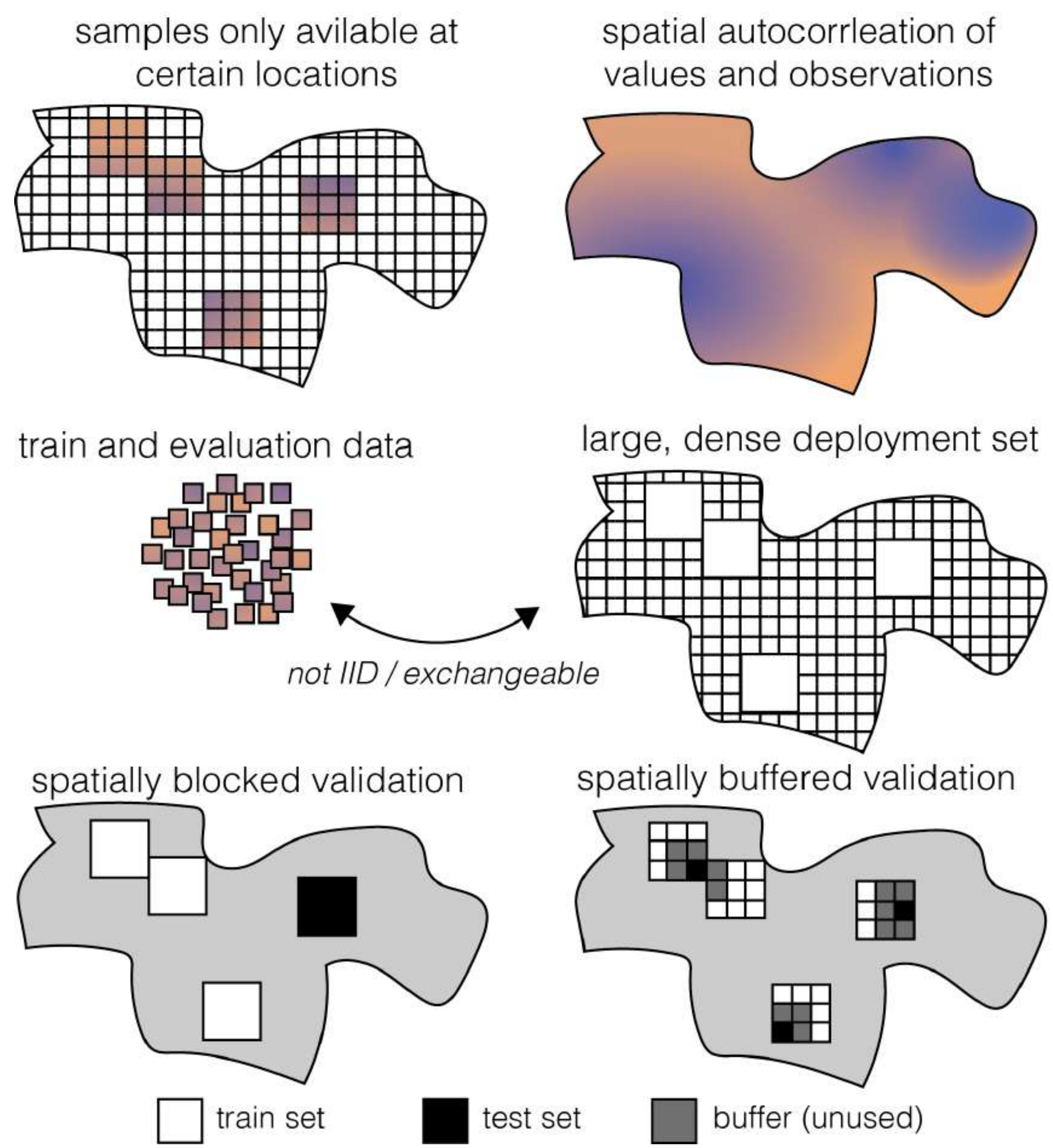
GeoML warrants specialized **methods.**



GeoML warrants specialized **methods.**



GeoML warrants specialized **evaluation.**



Evaluation challenges for geospatial ML
Rolf, Workshop on ML for Remote Sensing at ICLR 2023.

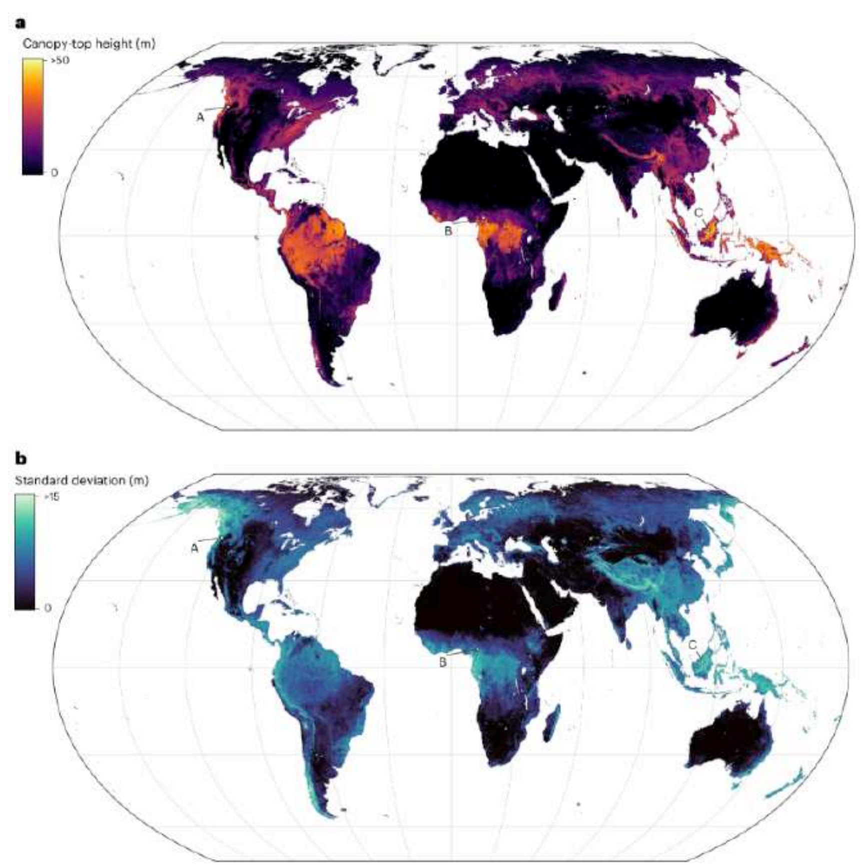
Position: Mission critical — Satellite data is a distinct modality in Machine Learning.
Rolf, Klemmer, Robinson, Kerner. ICML 2024.

We need to talk (more) about uncertainty in geospatial machine learning

We need to talk (more) about uncertainty in geospatial machine learning

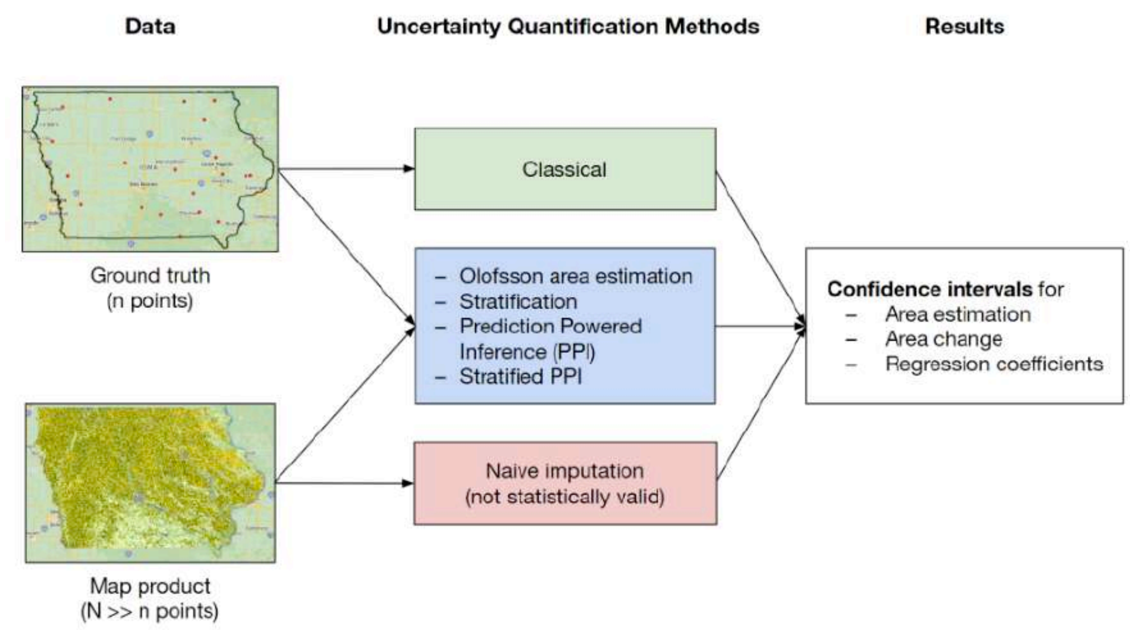
There are some excellent integrations of uncertainty in GeoML predictions!

ensemble-based uncertainty



A high-resolution canopy height model of the Earth
Lang et al. 2023

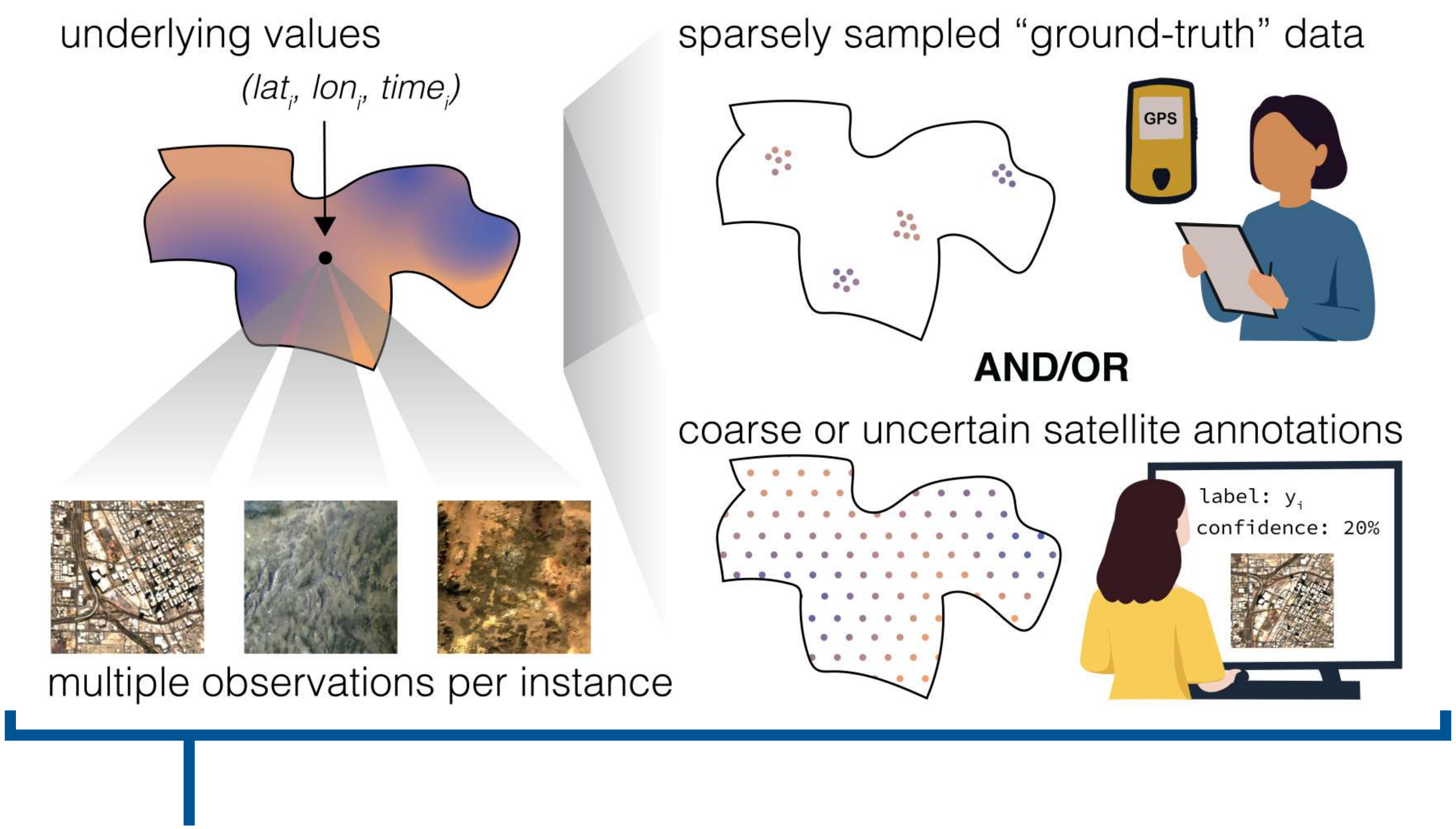
calibration-based intervals



Quantifying uncertainty in area and regression coefficient estimation from remote sensing maps
Lu et al. 2024

We need to talk (more) about uncertainty in geospatial machine learning

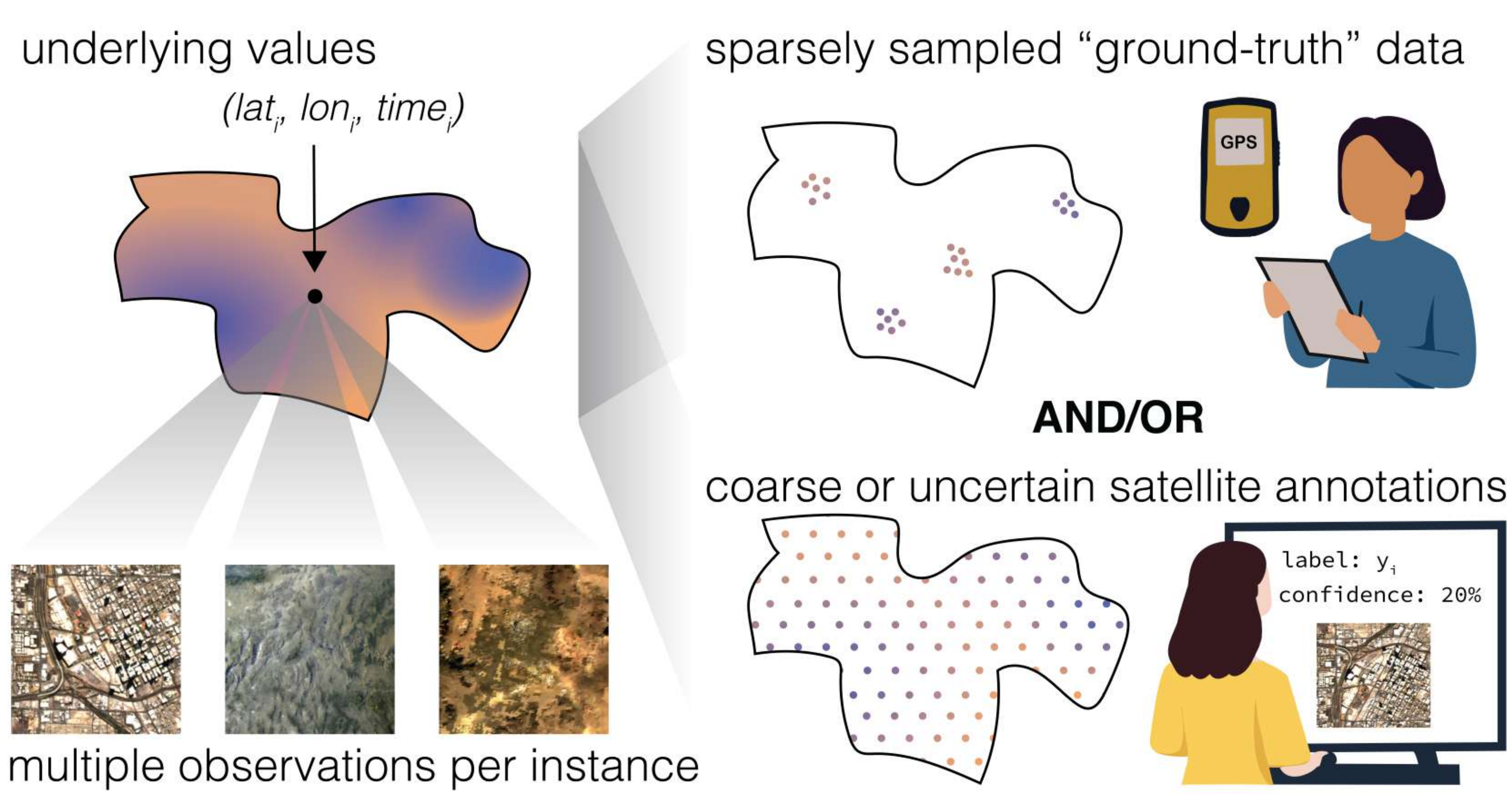
We need methods tailored for the uncertainties in geospatial data:



1. Uncertainty due to **partial representativity** of geospatial data

We need to talk (more) about uncertainty in geospatial machine learning

We need methods tailored for the uncertainties in geospatial data:



2. Uncertainty due to **spatially biased** distributions of data

1. Uncertainty due to **partial representativity** of geospatial data

We need to talk (more) about uncertainty in geospatial machine learning

Esther Rolf
Assistant Professor, CU Boulder Computer Science

estherrolf.com
estherrolf.bksy

I'm hiring! Apply by Dec 15:

- New! **Postdoctoral Fellowship in ML + Environment**
- PhD in CS at CU Boulder

HDSI | Harvard Data Science Initiative

CRCS Center for Research on Computation and Society
Harvard John A. Paulson School of Engineering and Applied Sciences

Microsoft Research

NSF GRFP

Google PhD Fellowship Program