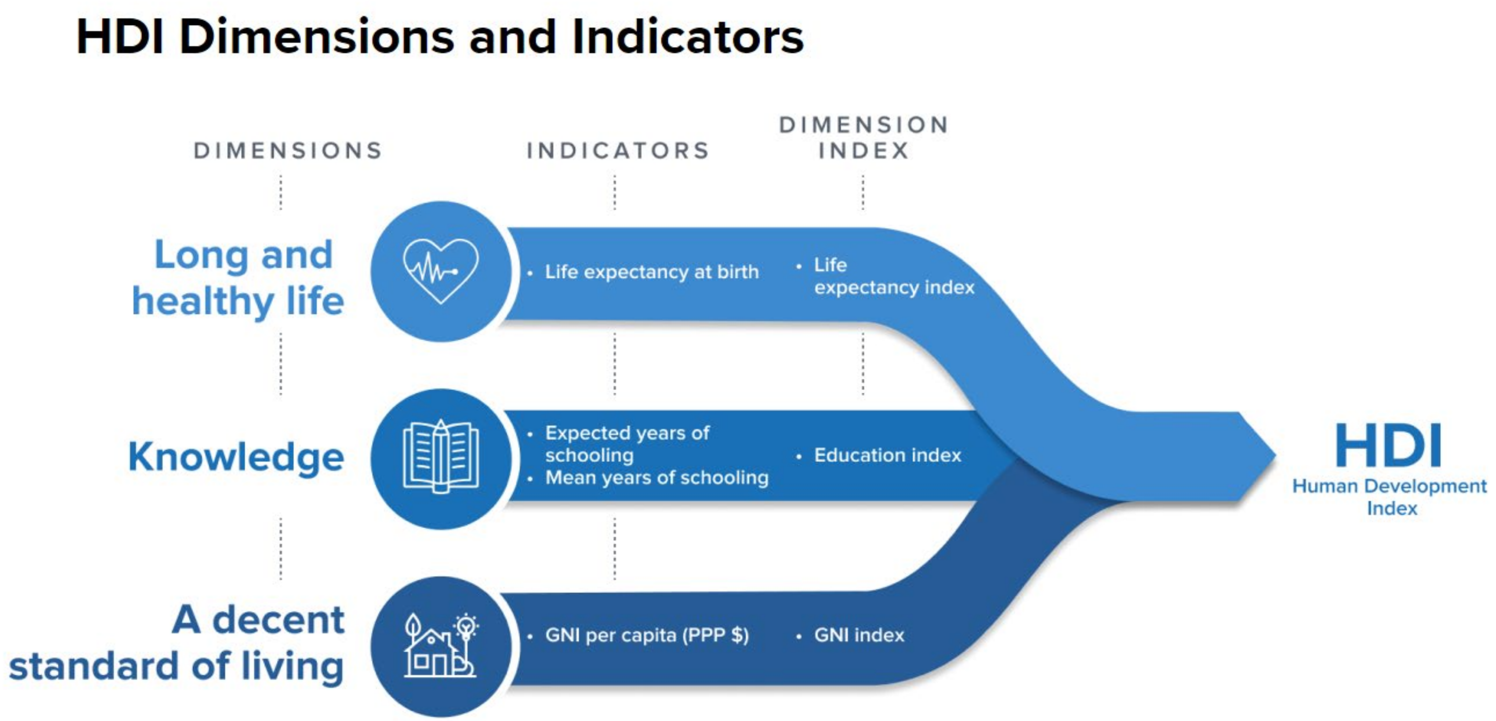
Estimating HDI at high resolution using Satellite Imagery

Luke Sherman (UC Berkeley)
Jonathan Proctor (Harvard University)
Hannah Druckernmiller (Resources for the Future)
Heriberto Tapia (Human Development Report Office, UNDP)
Solomon Hsiang (UC Berkeley)

Human Development Index (HDI)



- Indicator of **human capabilities**
- Probably the main indicator to assess the multidimensional development of countries

Human Development Index: beyond averages?

- **Official HDI** computed at the level of countries.
- Several projects compute HDI at a disaggregated level within certain countries.
- But, no systematic disaggregated data level, globally comparable.
- Exception: Permanyer and Smits (2019), at the provincial level.

Question: How to use satellite images to predict development variables globally at different levels of disaggregation?

There are over 700 Earth observation satellites in orbit.
Collectively, the retrieve > 100TB of data per day.

Potential to bring new standards (for comparability)

Human Development beyond averages

This presentation: emphasis on the **HDI (Experimental, not official)** and on **downscaling** (tackling the problem of limited information).

Administrative divisions (used here)
Countries (**ADM0**): **189** with data out of UN 193 member states
"Provinces" or "States" (**ADM1**): **1,707** units with 'unofficial' data
"Municipalities" or "Counties" (**ADM2**): around **61,172** units.

Methodology

Flexible use of Satellite Images

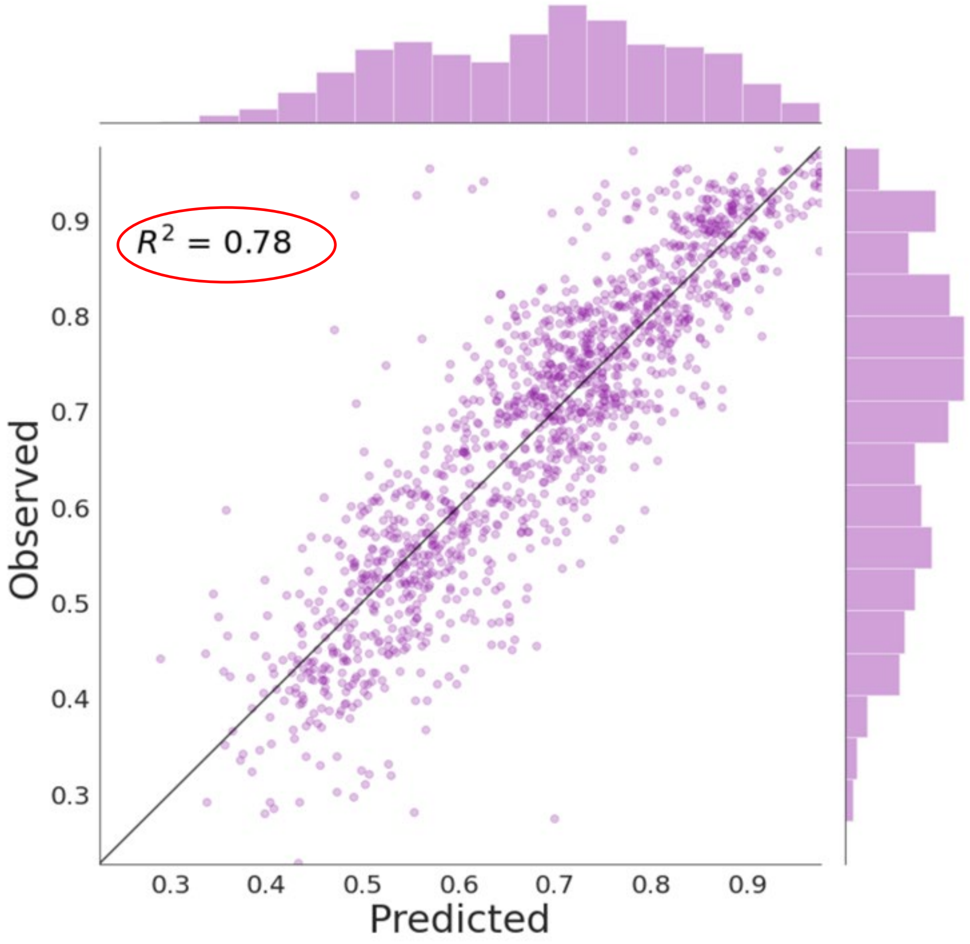
- MOSAIKS (Satellite data using daytime images), complemented with Nighttime Lights

First Stage: MOSAIKS transforms satellite images into large data vectors (through an unsupervised algorithm—Random Convolutional Features). Rolf et al (2021).

Second Stage: Ridge regressions (to accommodate big data). Using training data sets on labeled data.

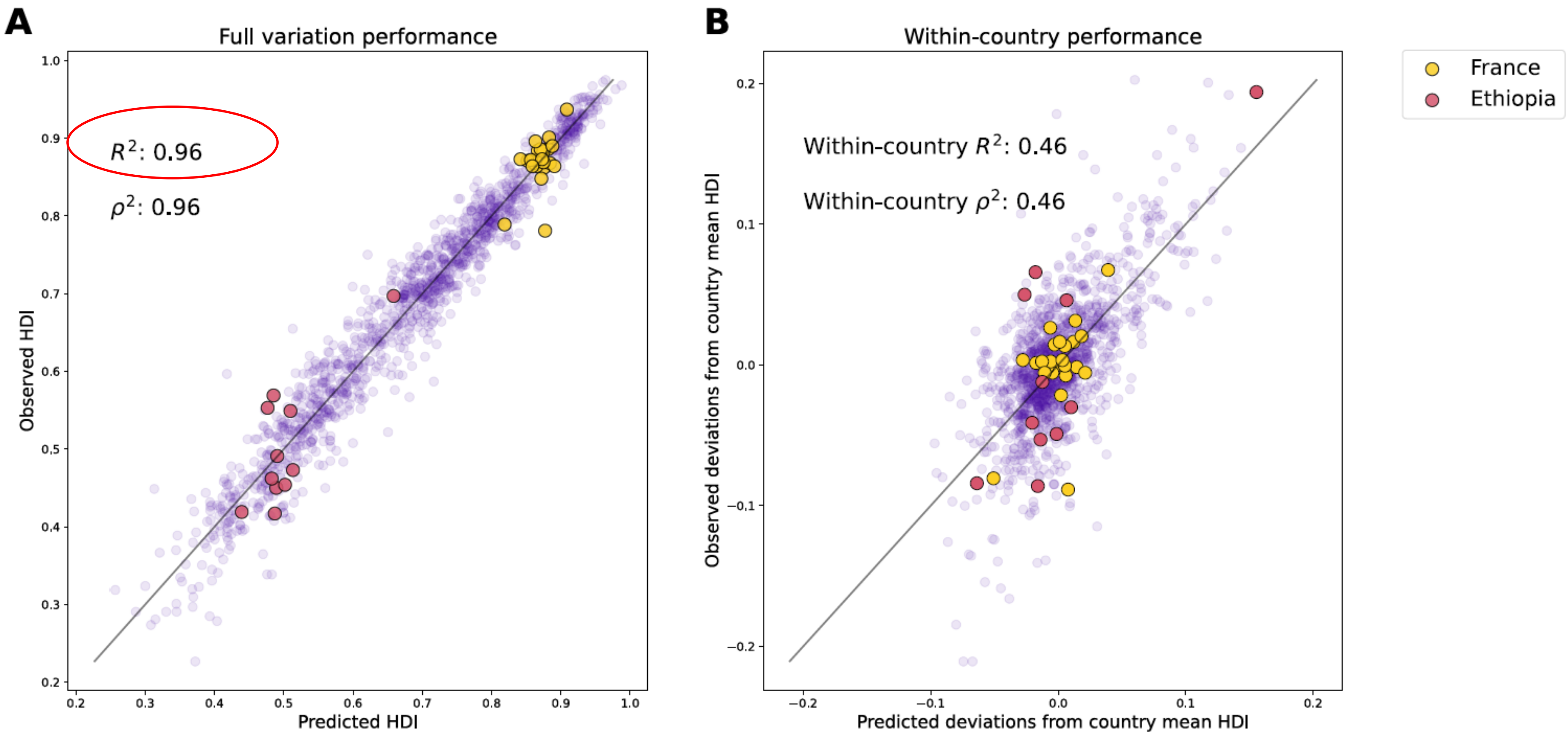
Predicting HDI: Province level (ADM1 to ADM1)

Training and predicting on provincial (ADM1) level directly



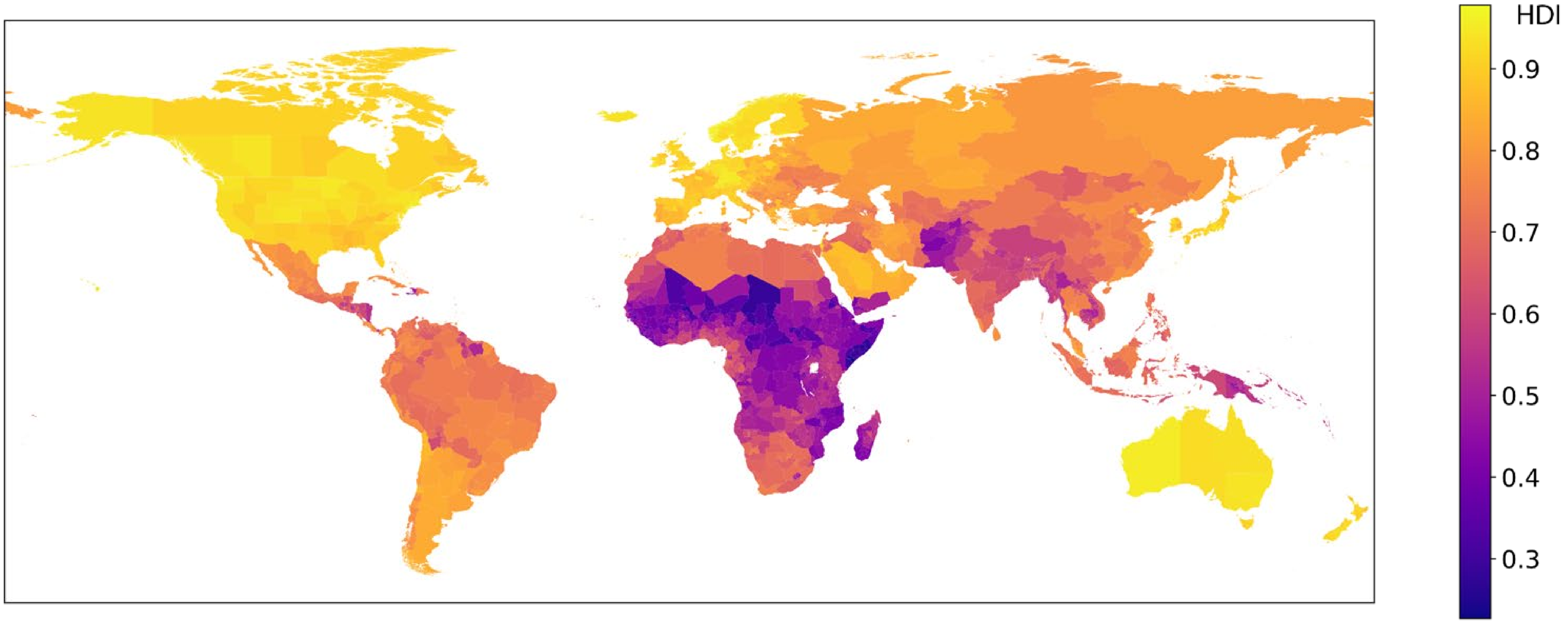
Downscaling HDI: ADM0 to ADM1

HDI predicted at the province level (n=1,344)



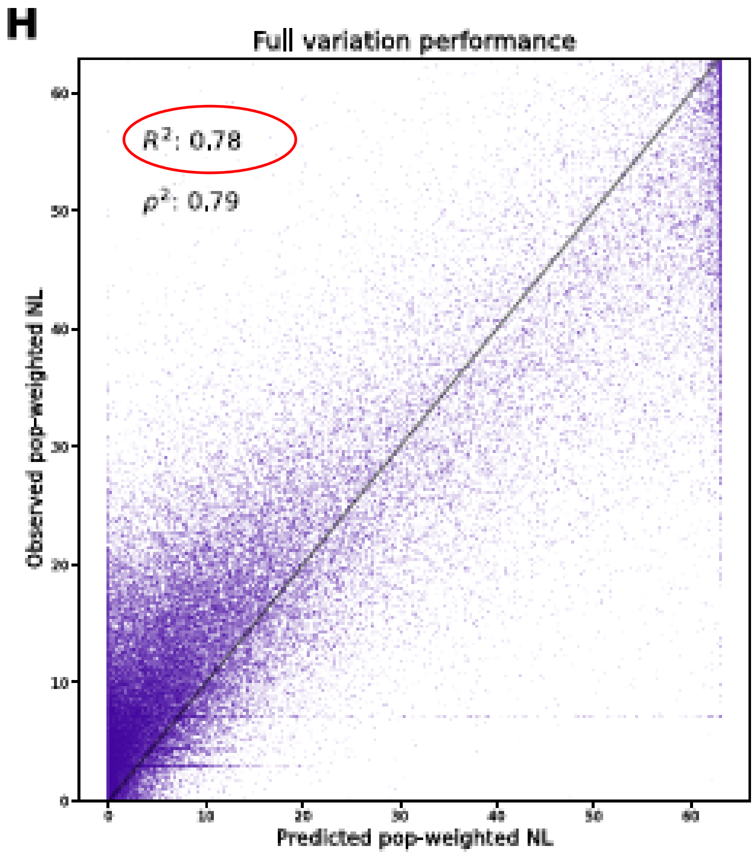
Training on country-level data (ADM0) and predicting on provincial (ADM1) level.
Using demeaned prediction, then adding back the country mean.

Downscaling HDI: Country to Province (ADM0 to ADM1)



Downscaling from ADM1 to ADM2, can we?

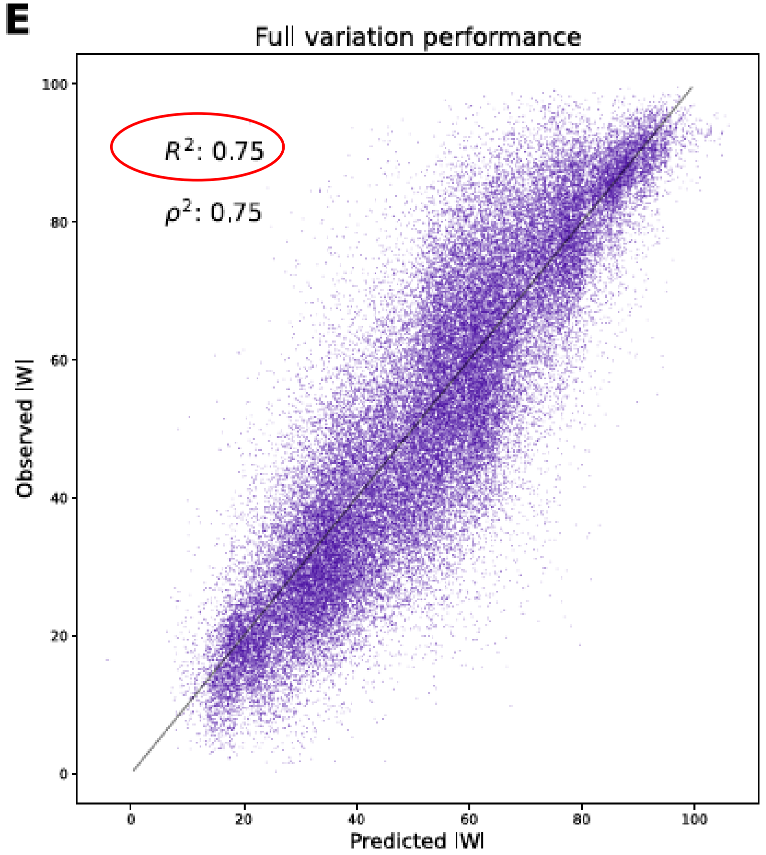
Nightlights



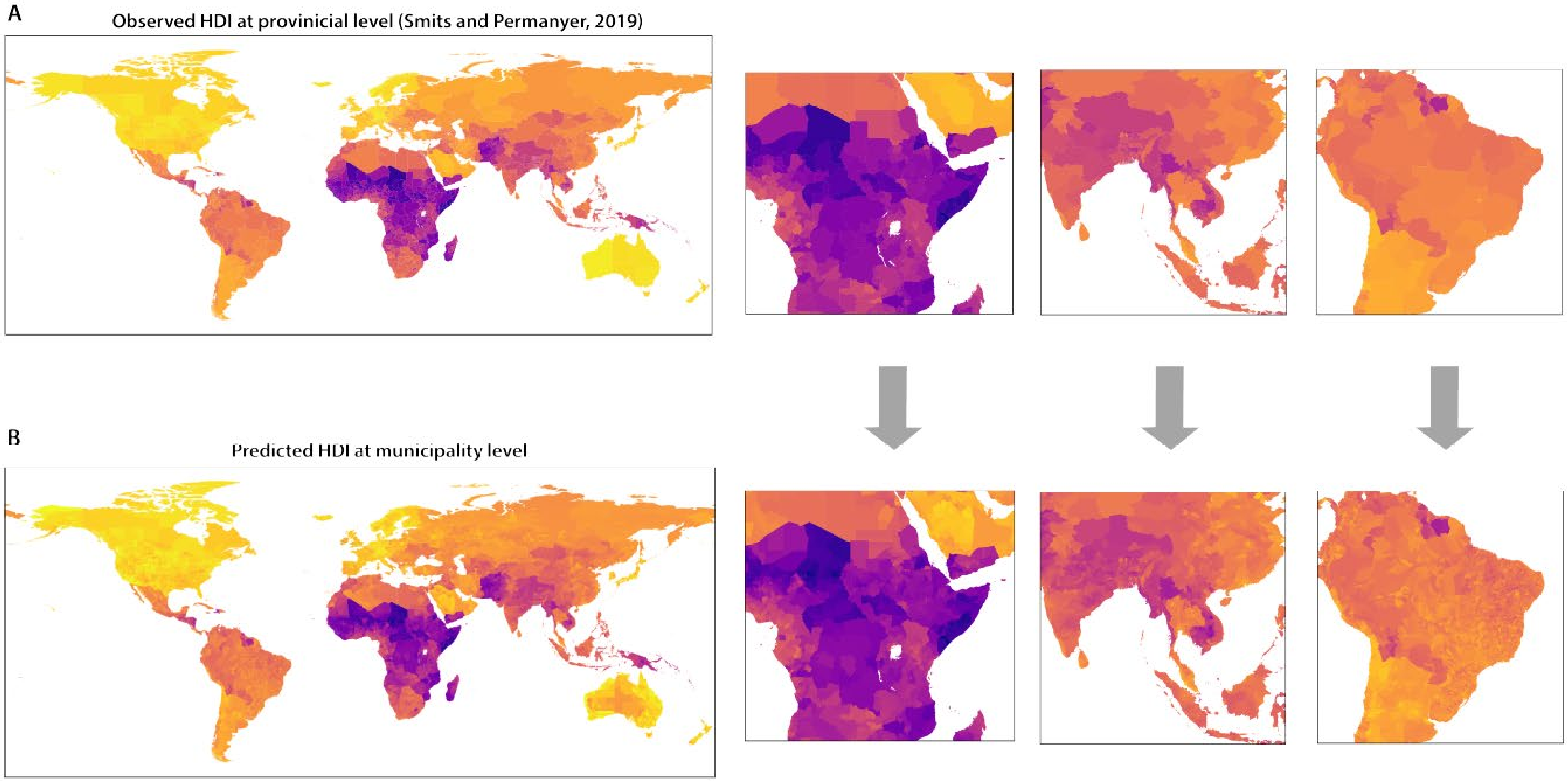
No consistent validation dataset for HDI

…But we can use other variables known at the municipal level

International Wealth Index



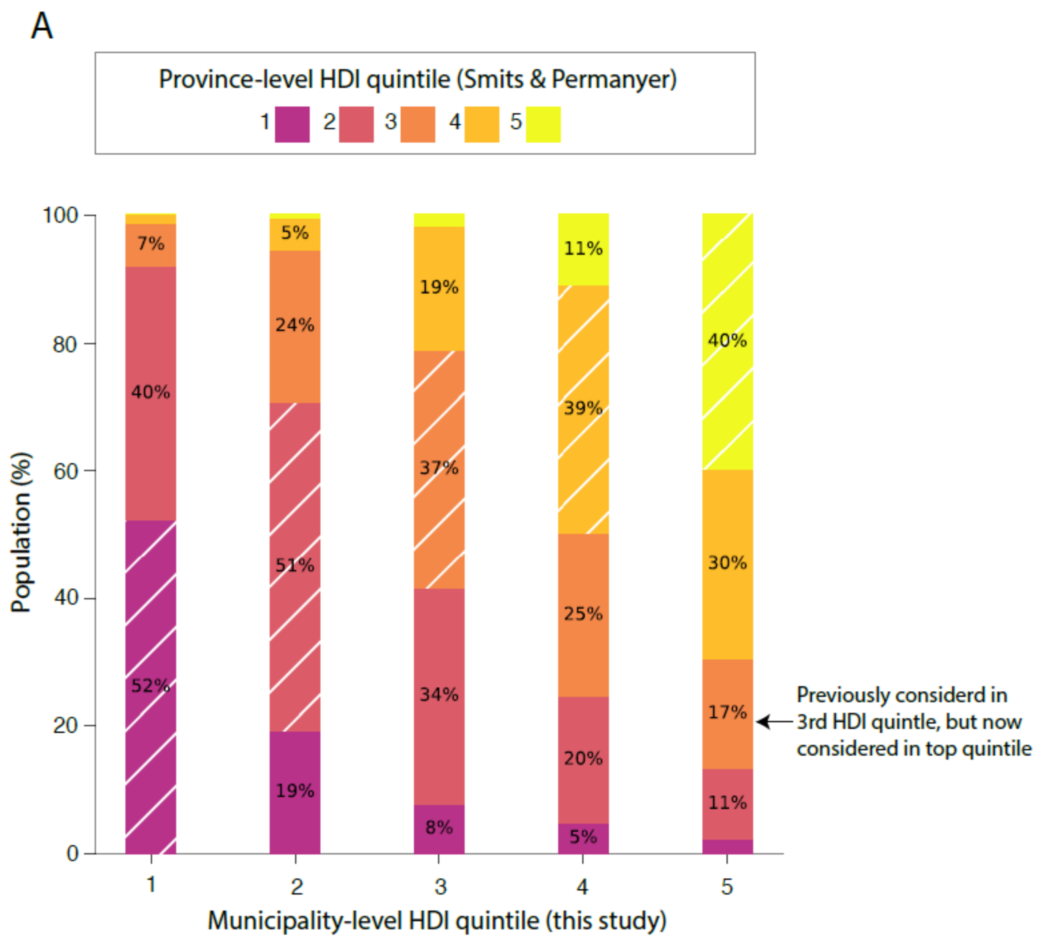
Finally: Downscaling HDI **From ADM1 to ADM2**



Why this is important? Better local picture for policies (Leaving No One Behind)

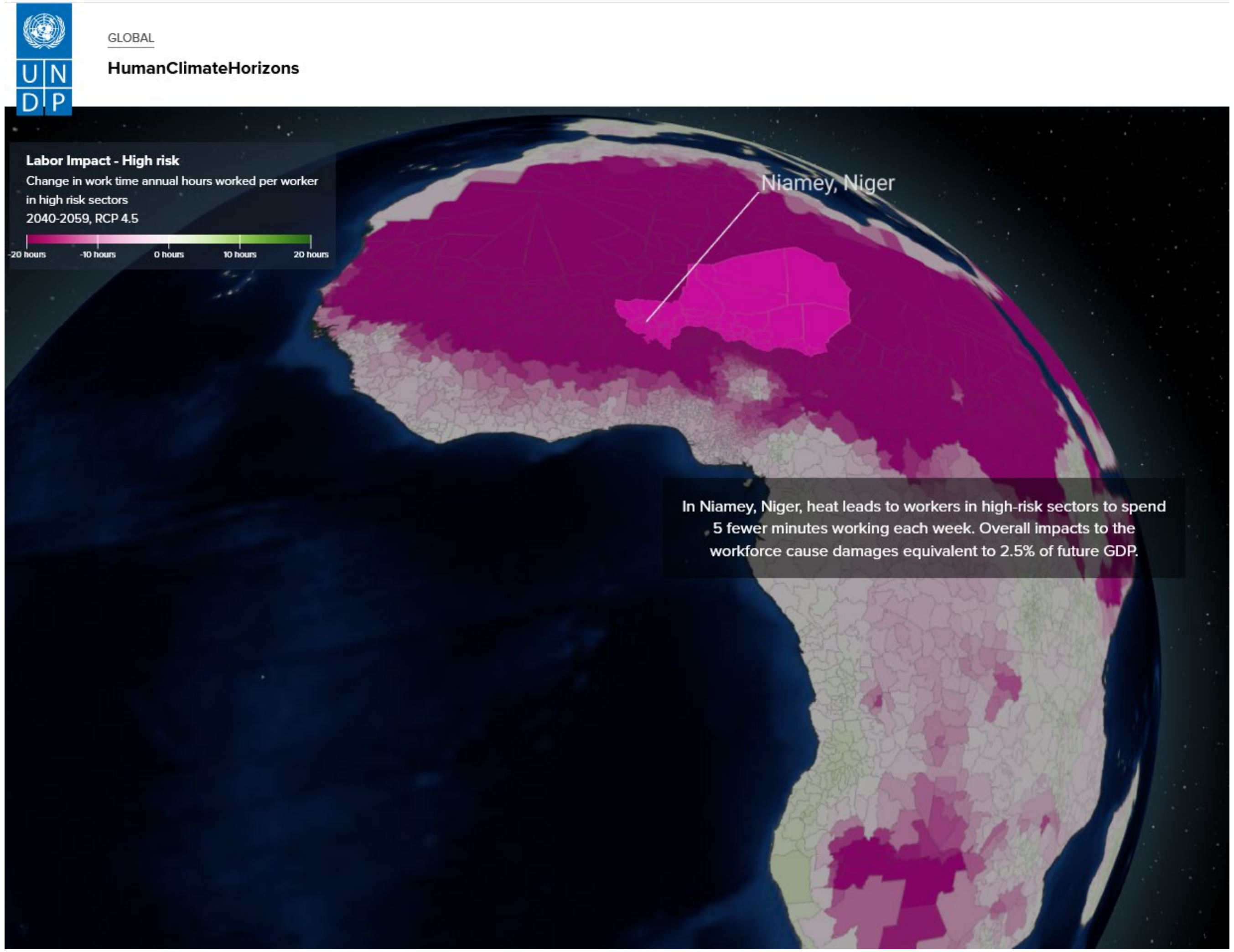
Differences in estimated HDI quintile, within countries

province-level vs municipality-level data.



More spatially granular HDI measures can improve decision-making.

Why this is important?
Ability to interact with new generation of data



#Humanclimatehorizons **platform:**

Hyper localized effects of climate change on human development:
2020-2100

Launched expected October 2022

Conclusions

Promising results in the disaggregation of HDI using SIML

- Predicting ADM1 from ADM1 (R-squared 0.78)
- Downscaling from ADM0 to ADM1 (R-squared 0.96)
- We present first global HDI disaggregated at the municipality level (ADM2).

We + our partners are working on extensions (inequalities, other development variables), further validation, and accessibility.

This agenda will play a critical role in designing policies both for LNOB and to respond to the global challenges of the Anthropocene.

Thank you

Heriberto.Tapia@undp.org

MOSAIKS

nature COMMUNICATIONS

ARTICLE

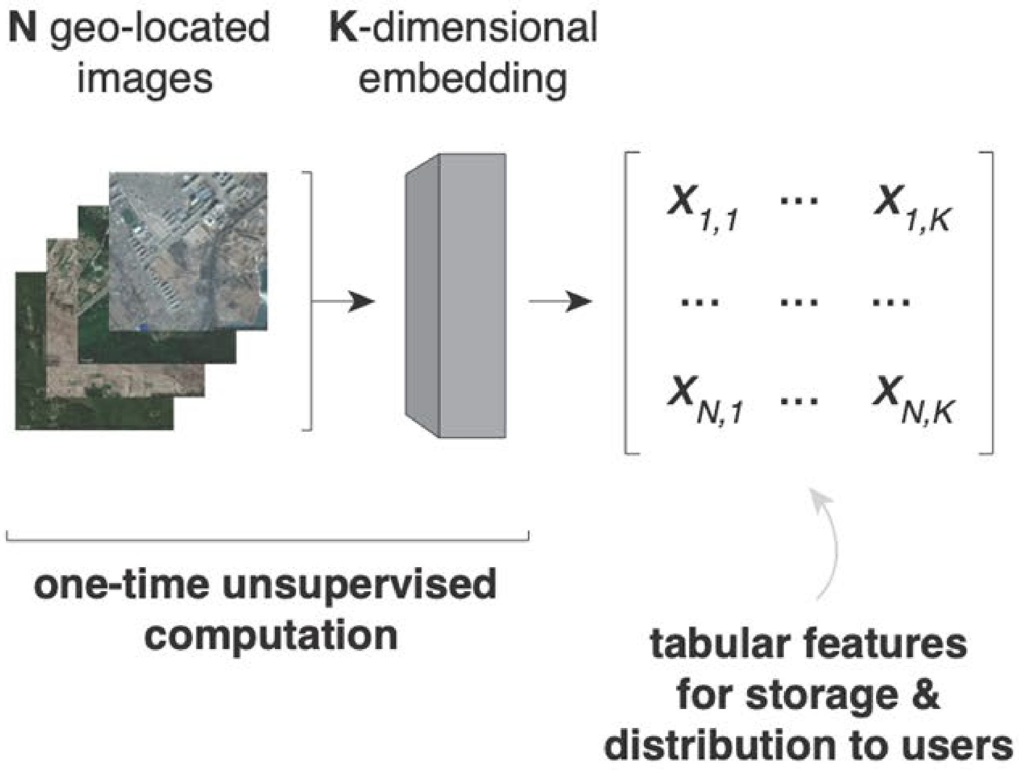
https://doi.org/10.1038/s41467-021-24638-z OPEN

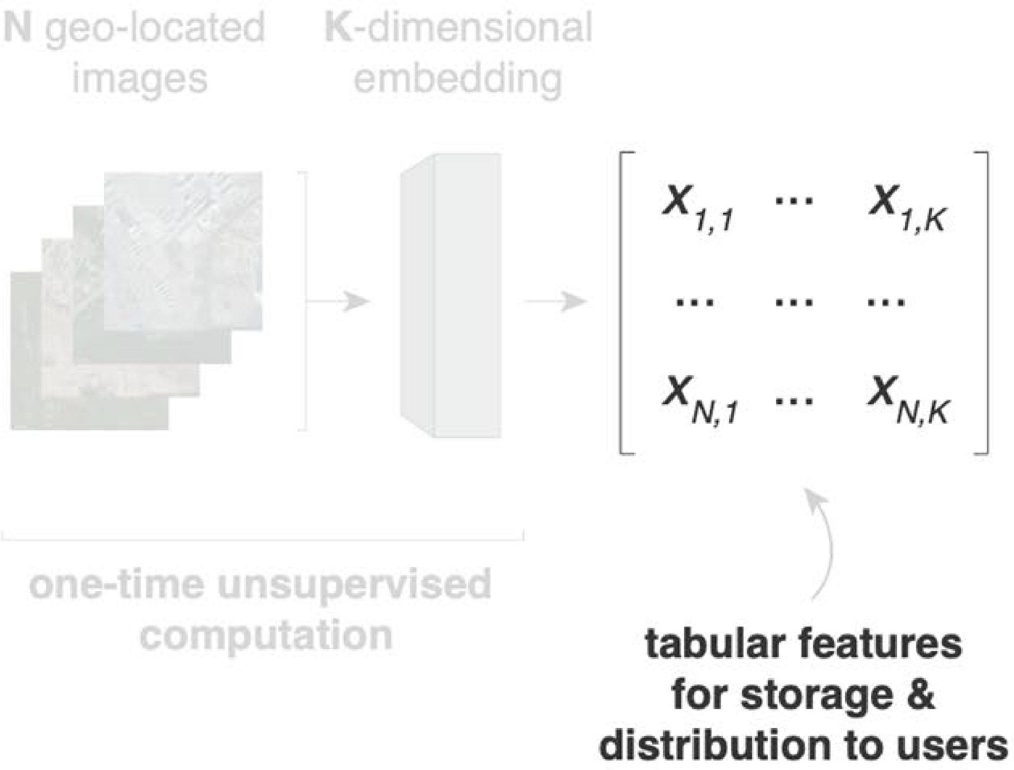
A generalizable and accessible approach to machine learning with global satellite imagery

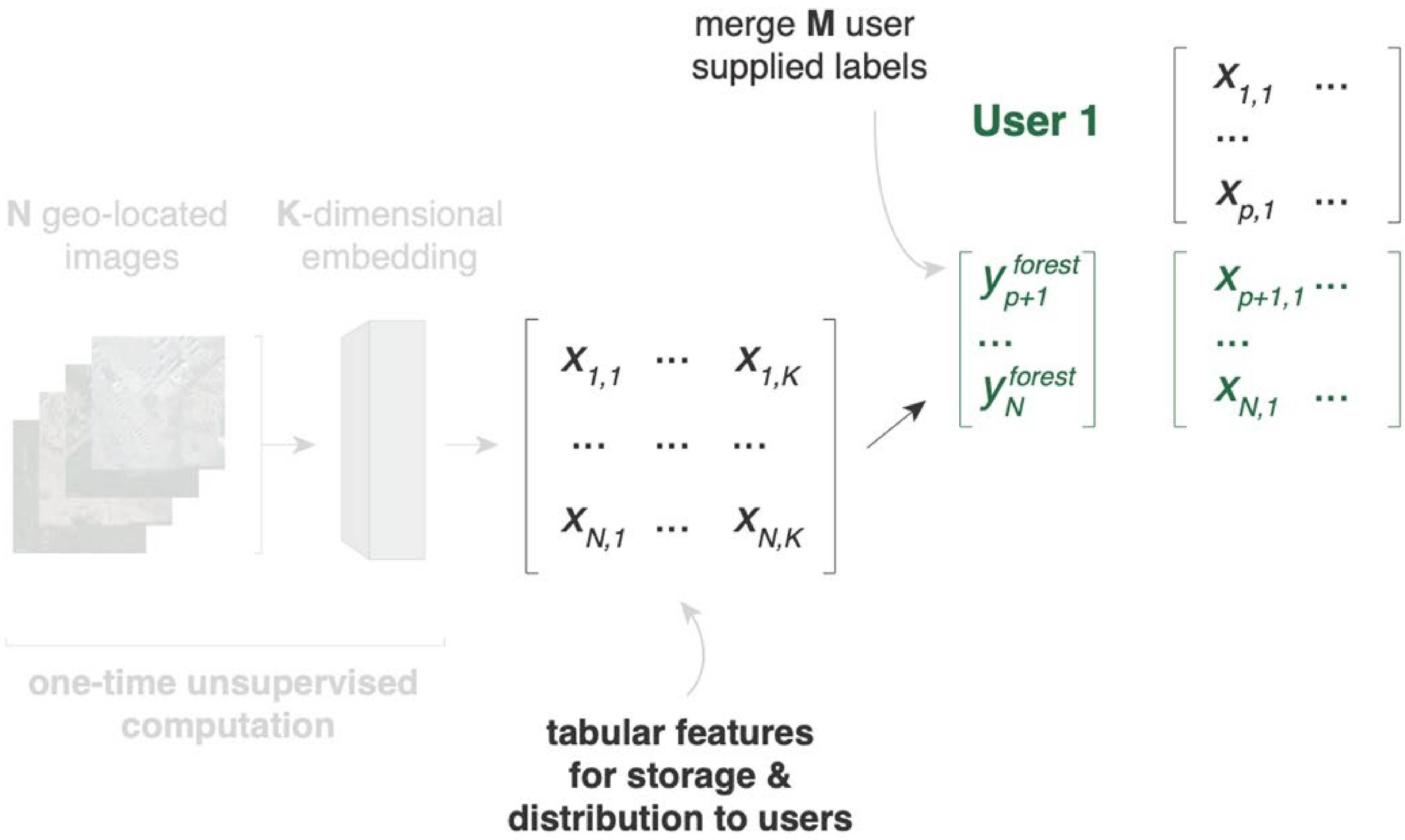
Esther Rolf[1,2,9], Jonathan Proctor[3,9], Tamma Carleton[4,5,9], Ian Bolliger[2,6,9], Vaishaal Shankar[1,9], Miyabi Ishihara[2,7], Benjamin Recht[1] & Solomon Hsiang[2,5,8]✉

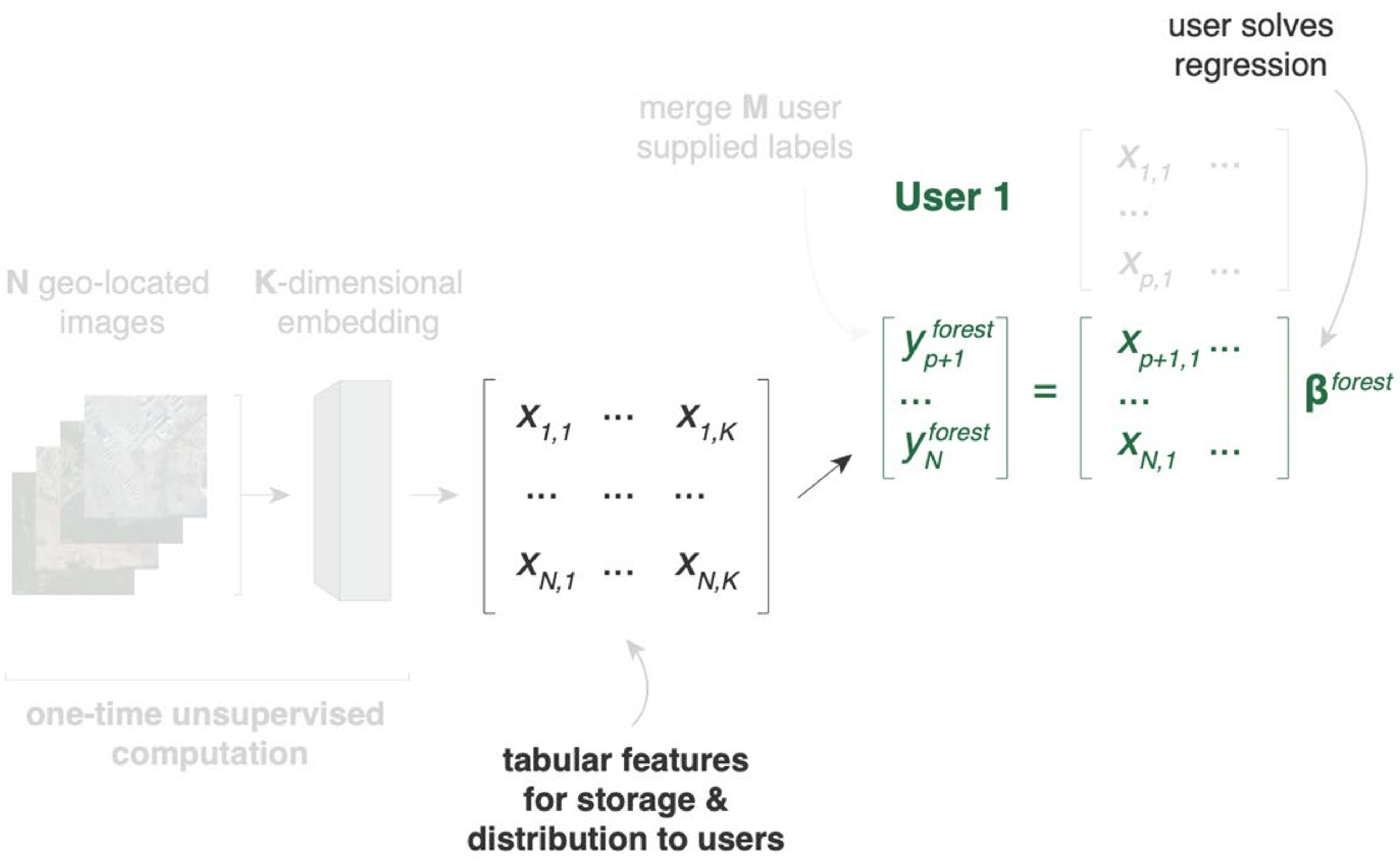
https://siml.berkeley.edu/

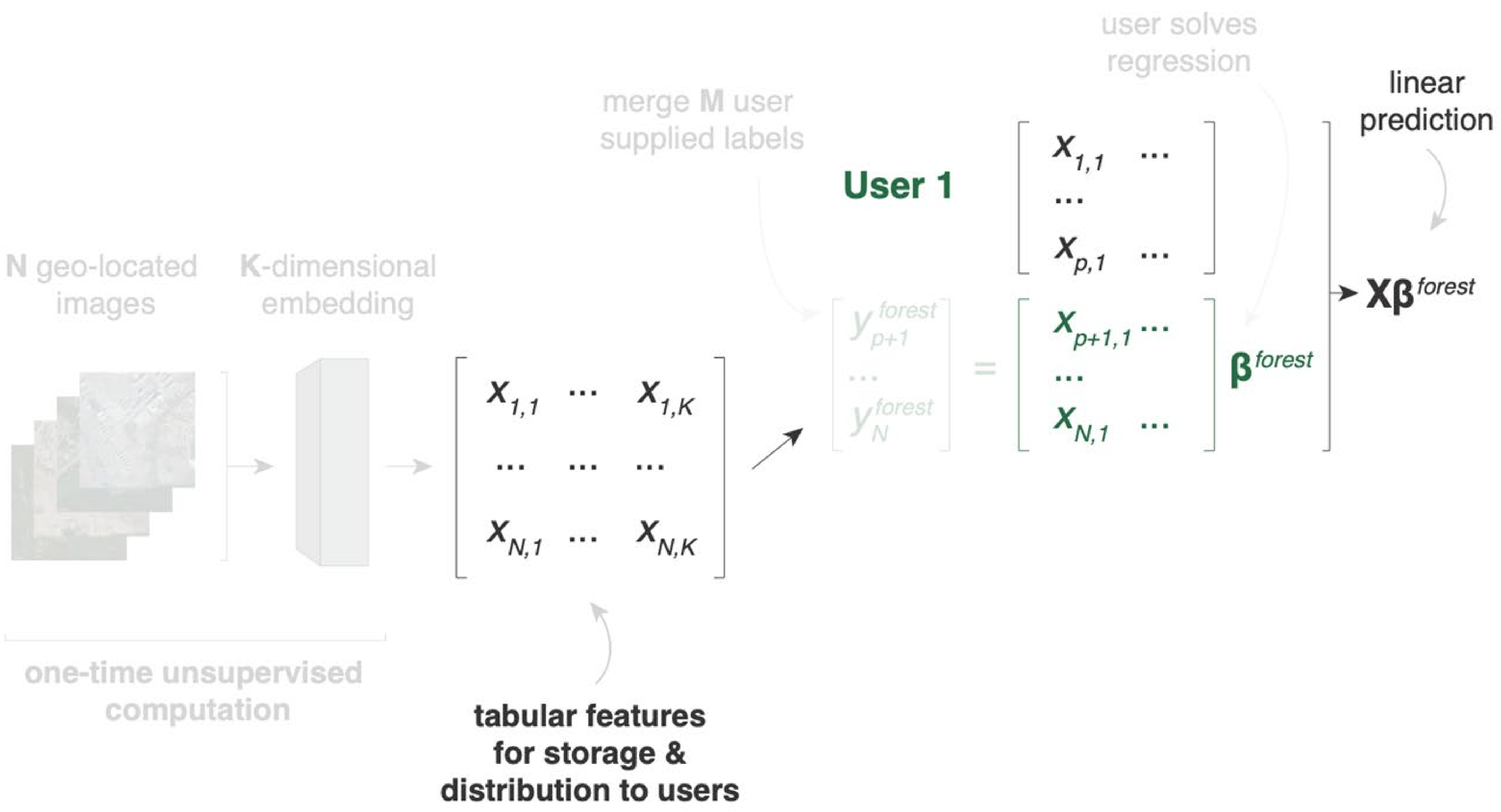
N geo-located images


N geo-located images
K-dimensional embedding
$x_{1,1}$ ... $x_{1,K}$
... ...
$x_{N,1}$ ... $x_{N,K}$
one-time unsupervised computation
tabular features for storage & distribution to users

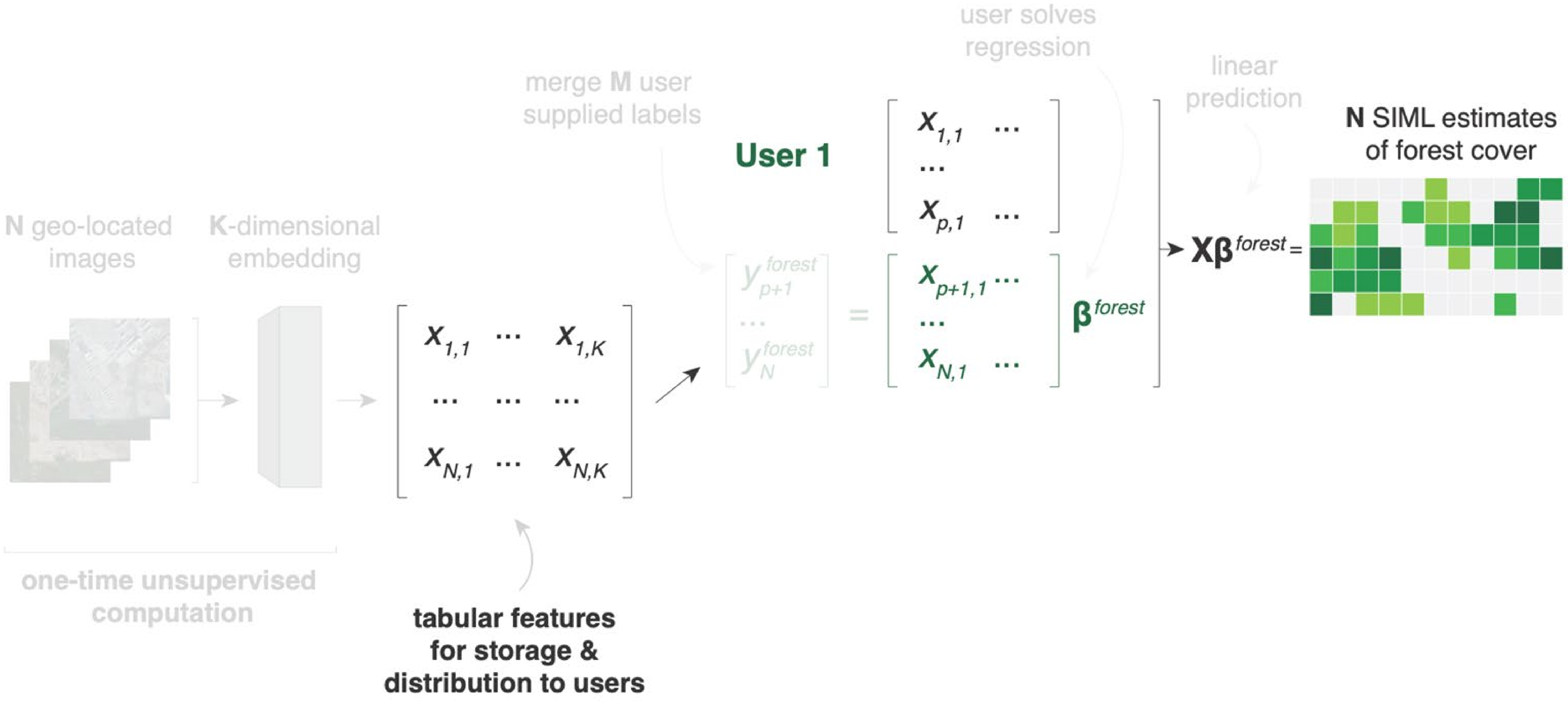

N geo-located images
K-dimensional embedding
$x_{1,1}$ ... $x_{1,K}$
... ...
$x_{N,1}$ ... $x_{N,K}$
one-time unsupervised computation
tabular features for storage & distribution to users


merge **M** user supplied labels
User 1
$x_{1,1}$ ... $x_{p,1}$ ...
y^{forest}_{p+1} ... y^{forest}_{N}
$x_{p+1,1}$ ... $x_{N,1}$ ...
N geo-located images
K-dimensional embedding
$x_{1,1}$ ... $x_{1,K}$
... ...
$x_{N,1}$ ... $x_{N,K}$
one-time unsupervised computation
tabular features for storage & distribution to users

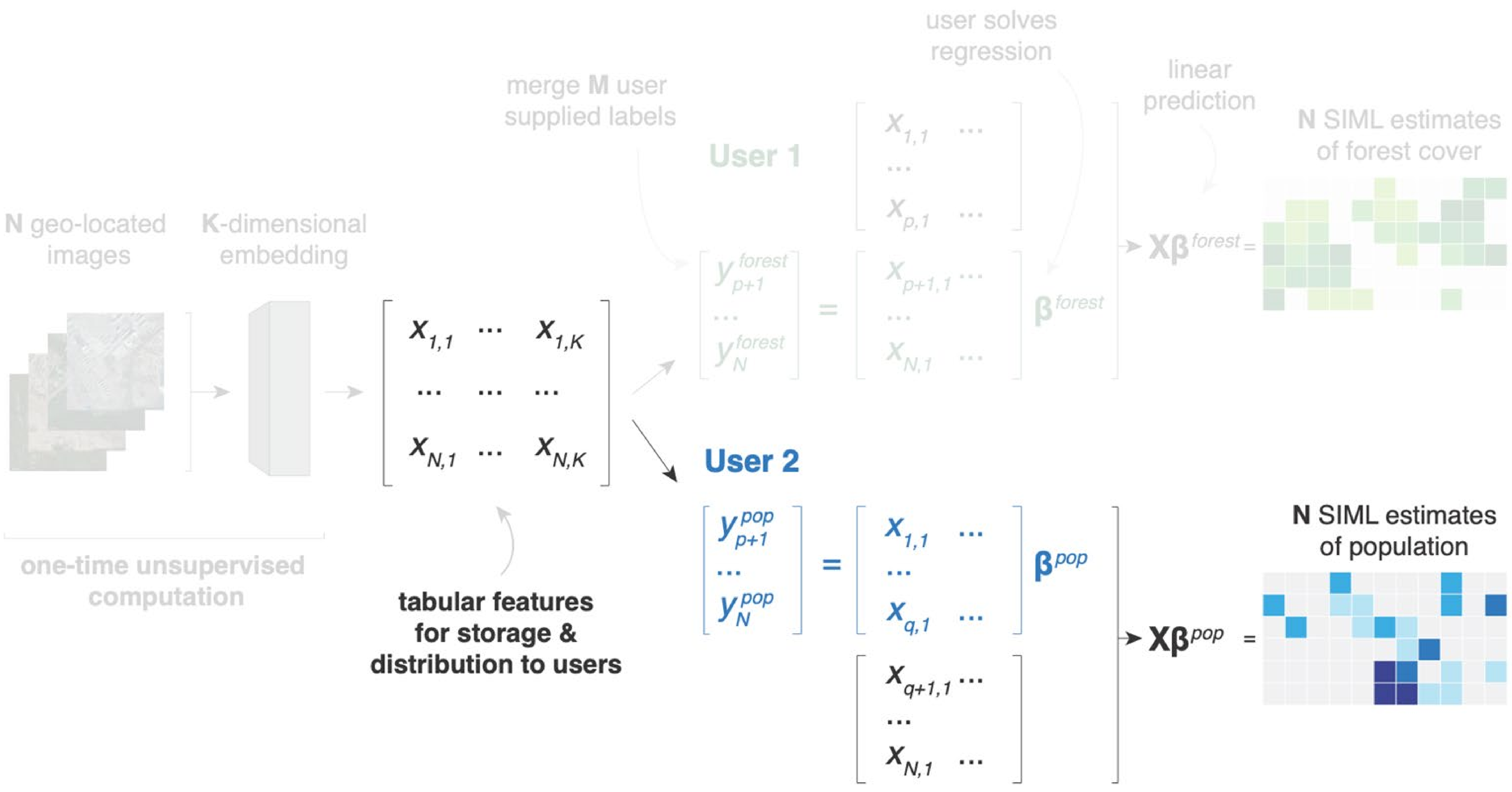

user solves regression
merge M user supplied labels
User 1
N geo-located images
K-dimensional embedding
β^{forest}
one-time unsupervised computation
tabular features for storage & distribution to users



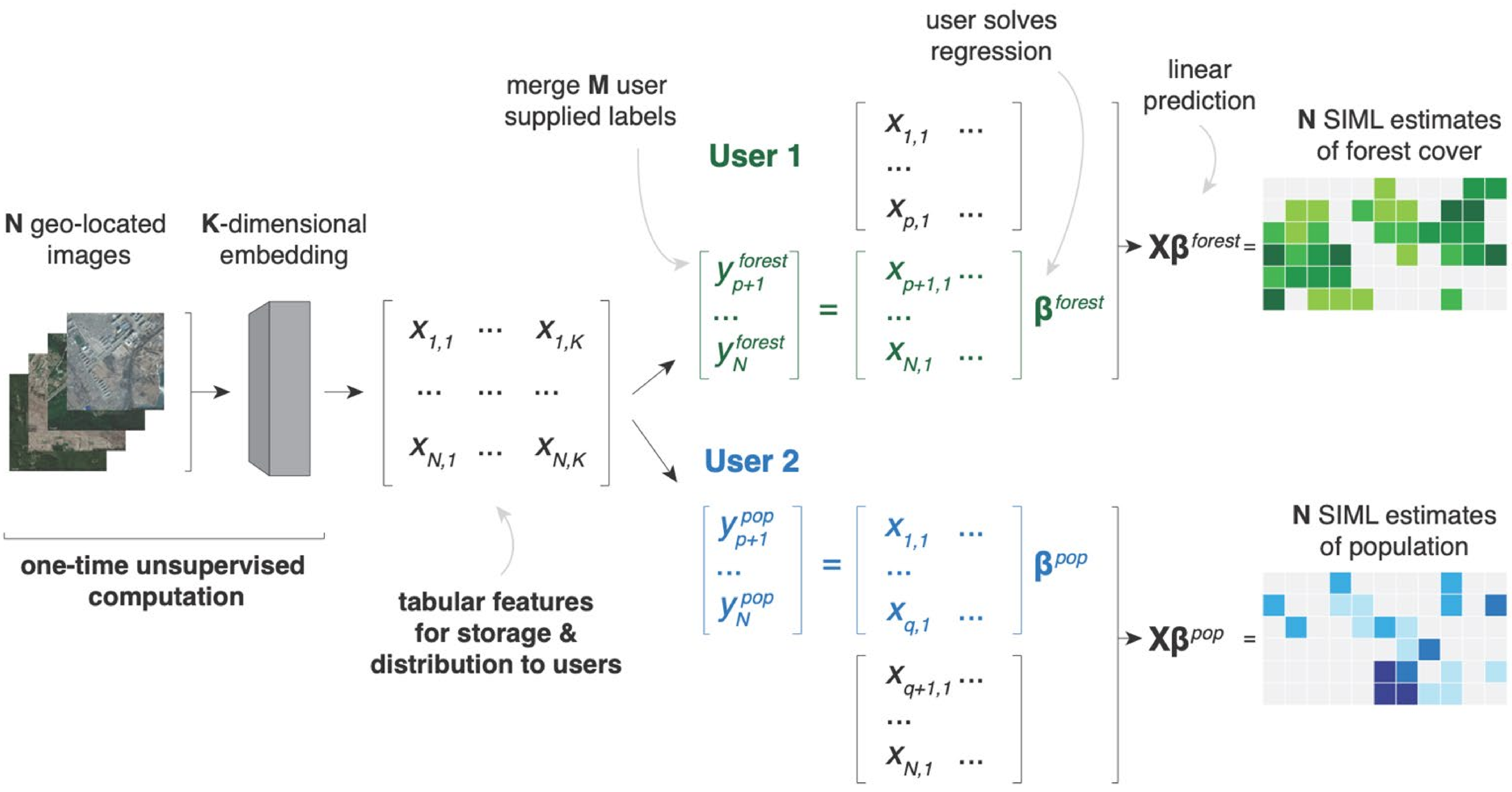
N geo-located images
K-dimensional embedding
one-time unsupervised computation
tabular features for storage & distribution to users
merge M user supplied labels
user solves regression
User 1
linear prediction
$X\beta^{forest}$
β^{forest}



N geo-located images
K-dimensional embedding
one-time unsupervised computation
tabular features for storage & distribution to users
merge M user supplied labels
User 1
user solves regression
linear prediction
N SIML estimates of forest cover



N geo-located images
K-dimensional embedding
one-time unsupervised computation
merge M user supplied labels
tabular features for storage & distribution to users
User 1
user solves regression
linear prediction
N SIML estimates of forest cover
User 2
N SIML estimates of population


N geo-located images
K-dimensional embedding
one-time unsupervised computation
tabular features for storage & distribution to users
merge M user supplied labels
user solves regression
linear prediction
User 1
User 2
N SIML estimates of forest cover
N SIML estimates of population
$X\beta^{forest} =$
$X\beta^{pop} =$
β^{forest}
β^{pop}