

Predicting and monitoring local development from outer space

Evidence from the Bolivian municipalities

Carlos Mendez

<https://carlos-mendez.org>

Graduate School of International Development

Nagoya University

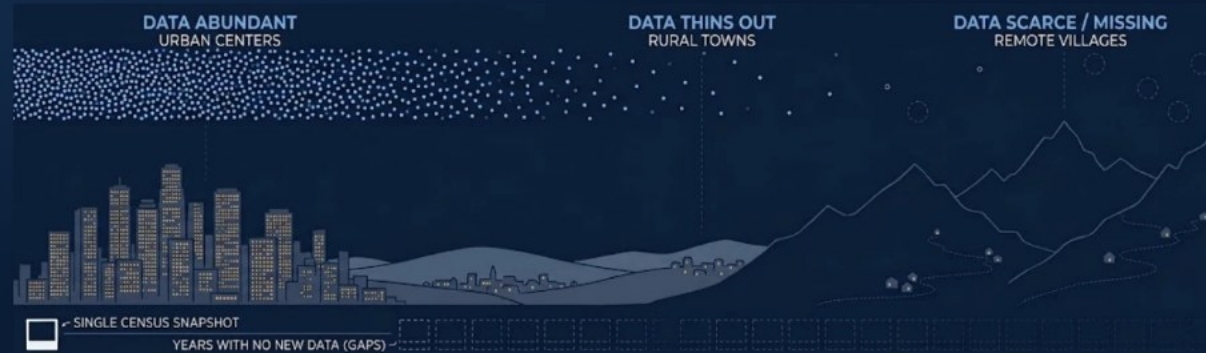
Japan



Motivation and questions

Where development data is scarcest, it is needed most

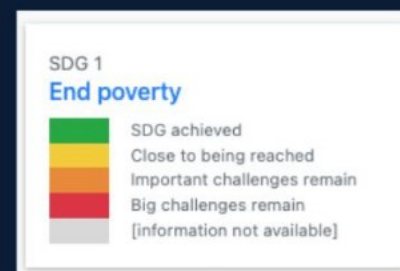
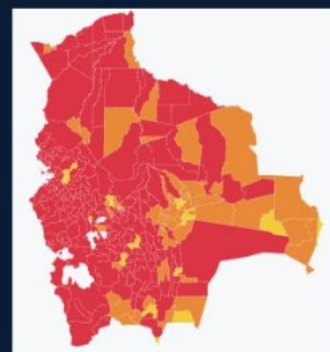
THE PROBLEM



- Local development statistics are scarce, expensive, and outdated.
- Some subnational regions are too small or too remote to survey often.
- A census is a snapshot, not a series.

The places that most need planning and monitoring are the ones surveys reach last.

A new dataset to study subnational development in Bolivia

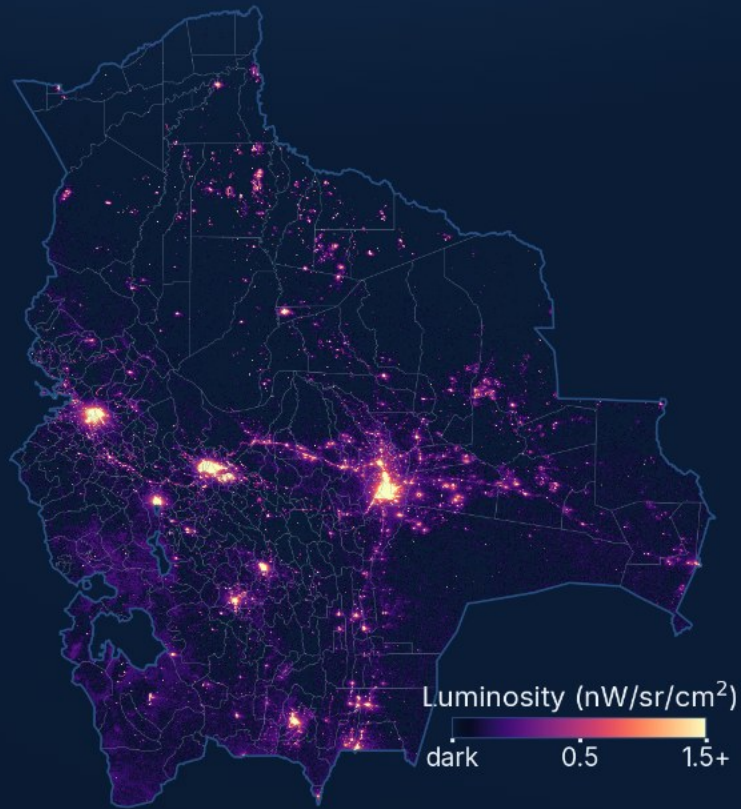


15 development (SDG) indicators for 339 municipalities around the year 2017

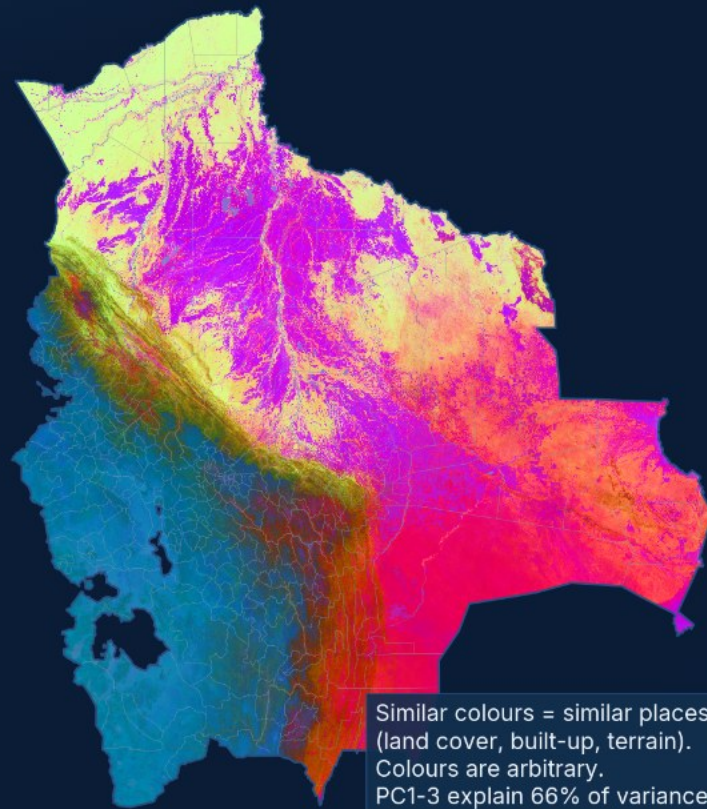
Satellite imagery: A modern instrument for measuring local development

THREE WAYS TO SEE BOLIVIA

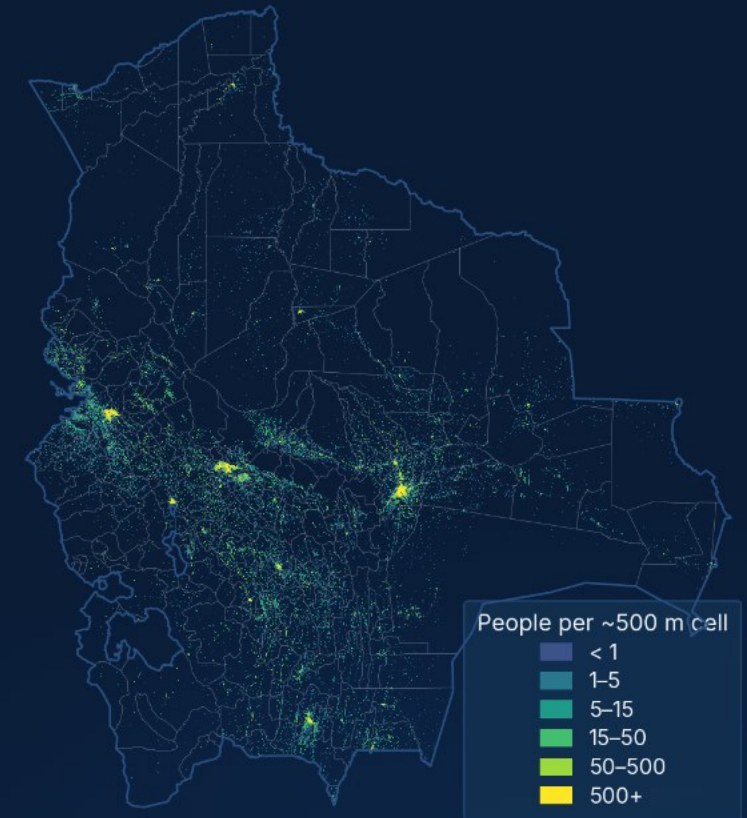
Nighttime lights



Daytime embeddings



Population



Research questions

TWO COMPLEMENTARY YET DIFFERENT QUESTIONS

- Can satellites **read (predict)** development — what is their predictive accuracy?
- Can satellites **map (monitor)** it — where development clusters and outliers sit?

Global predictive accuracy may not be useful for local monitoring.



Data and methods

The prediction targets: 15 development goals

THE OUTCOME VARIABLE



62 indicators from official statistics

15 goal indices, each scored 0–100
— higher is better

a 2017 value for each of 339 municipalities

SDG 1
Poverty

SDG 2
Hunger

SDG 3
Health

SDG 4
Education

SDG 5
Gender

SDG 6
Water

SDG 7
Energy

SDG 8
Jobs

SDG 9
Infrastructure

SDG 10
Inequality

SDG 11
Cities

SDG 13
Climate

SDG 15
Land

SDG 16
Institutions

SDG 17
Partnerships

Municipal Atlas of the Sustainable Development Goals in Bolivia (Andersen et al., 2020) · only two of the 62 indicators come from satellites

Nighttime lights: the predictor economists already know



MEAN · MEDIAN

Typical brightness of each pixel across every cloud-free night of 2017.

MASKED MEAN · MEDIAN

The same, with background glow, fires and aurora removed: stable human light only.

MINIMUM · MAXIMUM

The dimmest and brightest nights: how steady or episodic the light is.

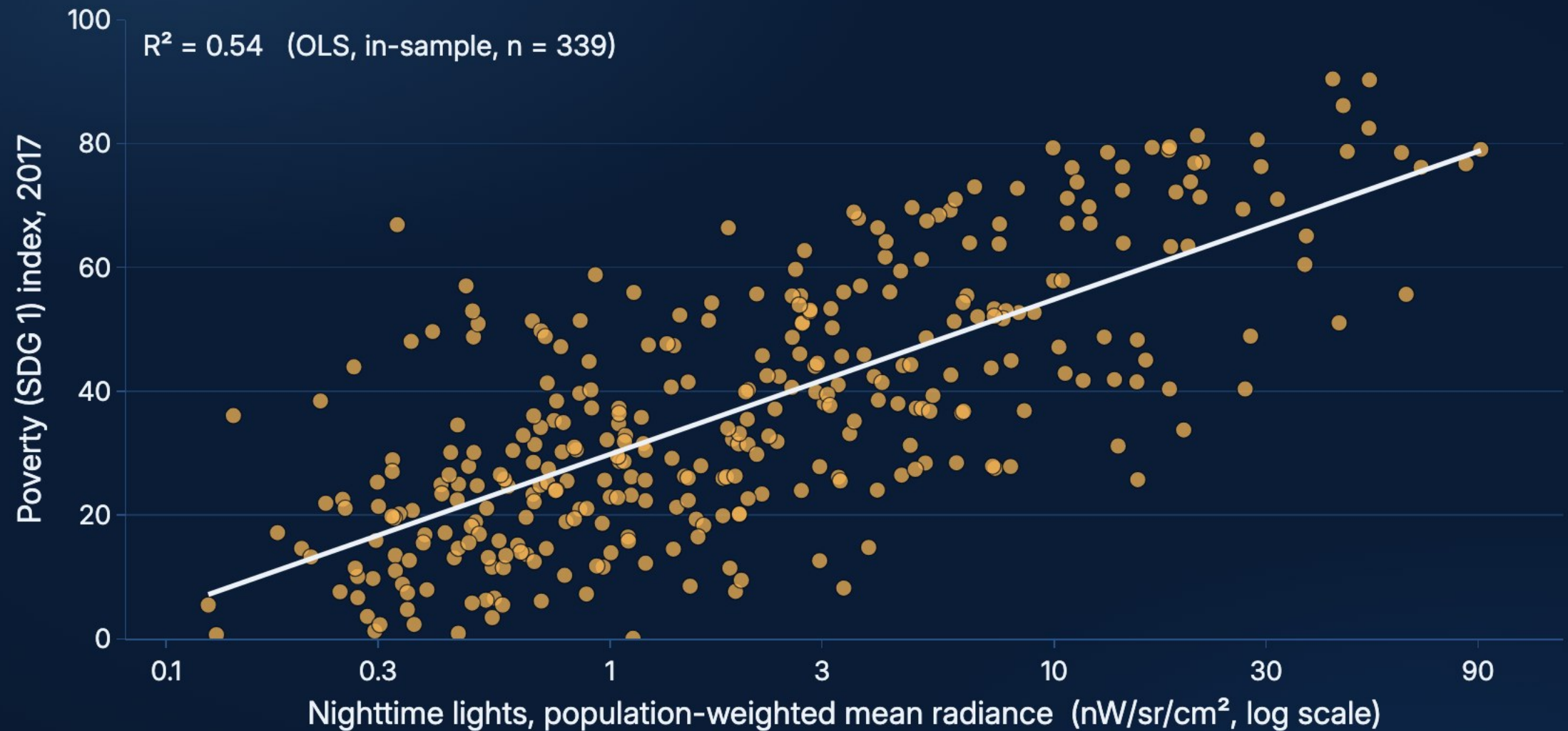
TWO COVERAGE COUNTS

How many nights were observed, and how many were cloud-free: data quality, not light.

Six radiance statistics plus two counts, aggregated to each municipality: all 8 enter the model together.

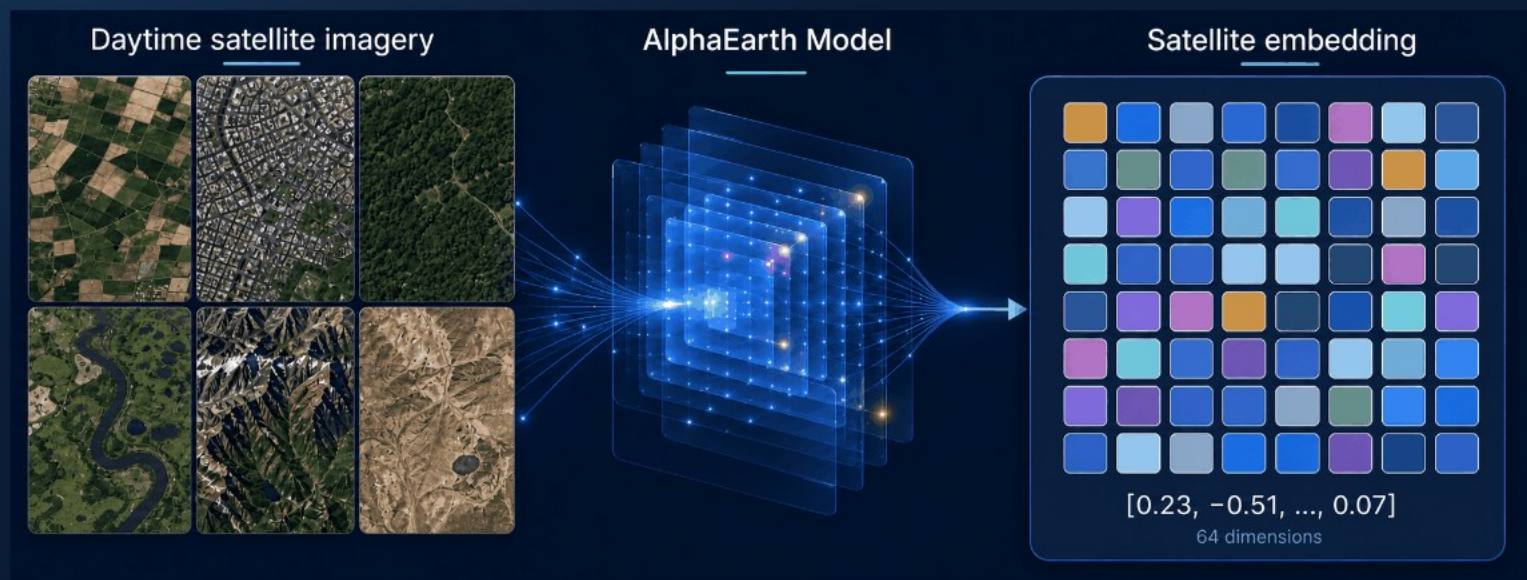
NASA Black Marble · Suomi NPP VIIRS day/night band, annual composite · radiance in $\text{nW/cm}^2/\text{sr}$

Example: Nighttime lights and SDG 1 (no poverty)



VIIRS annual composite, population-weighted municipal means · 2017 · ordinary least squares, in-sample

Daytime embeddings: what the new predictors are



1 · A YEAR OF IMAGES

Each patch of land, observed again and again through 2017.

2 · COMPRESSED BY A MODEL

A neural network trained worldwide, with no social, economic, or nighttime-light data.

3 · INTO 64 NUMBERS

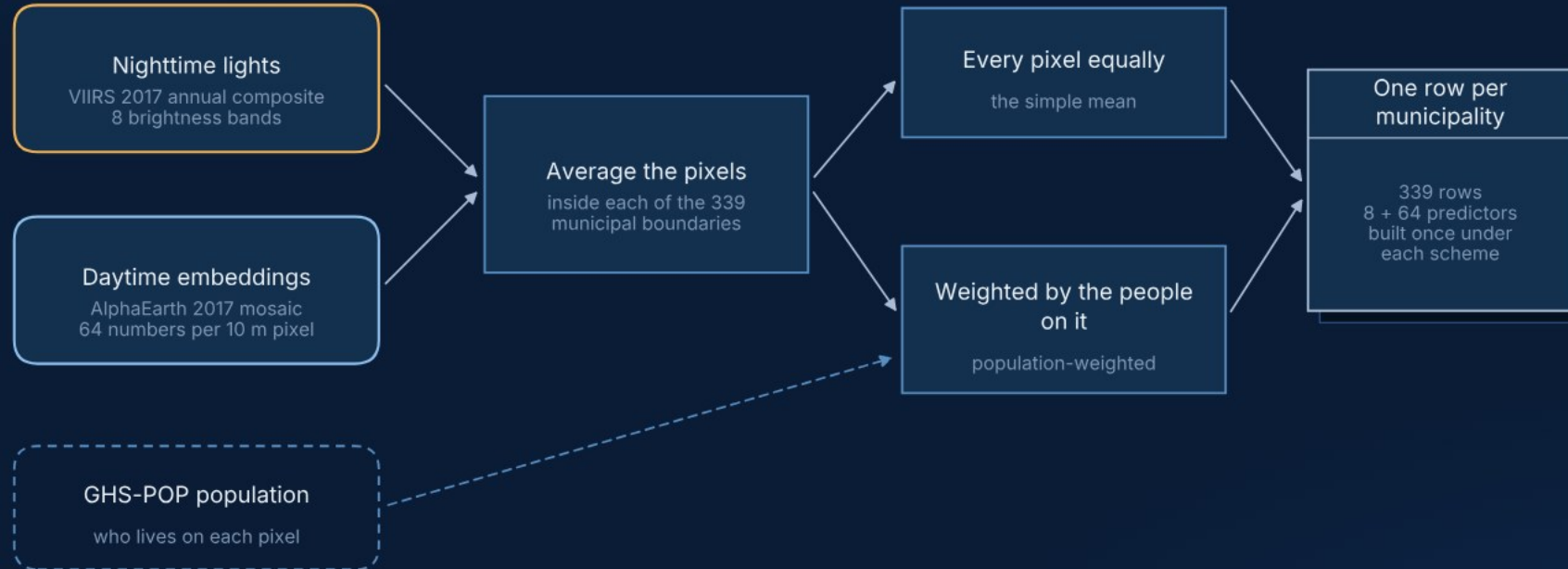
Places that look alike get similar numbers. No single number has a name.

Intuition: a principal-components index of the landscape, with **64** components instead of one.

AlphaEarth Foundations satellite embeddings · [explore how they work, in pictures ↗](#)

From pixels to municipalities

Average every pixel equally or weight each pixel by the people on it.



Prediction and monitoring methods

HOW EACH QUESTION IS ANSWERED

PREDICT



MANY TREES
MAKE A FOREST



BAYESIAN
TUNING



TESTED ON
UNSEEN DATA

MONITOR



NEIGHBOURS



GLOBAL
CLUSTERING



LOCAL
CLUSTERS

PREDICTION · MACHINE LEARNING

- Random forest, one model for each goal index.
- Tuned by Bayesian search, inside a nested cross-validation.
- Scored out of sample: random folds, then whole departments held out.

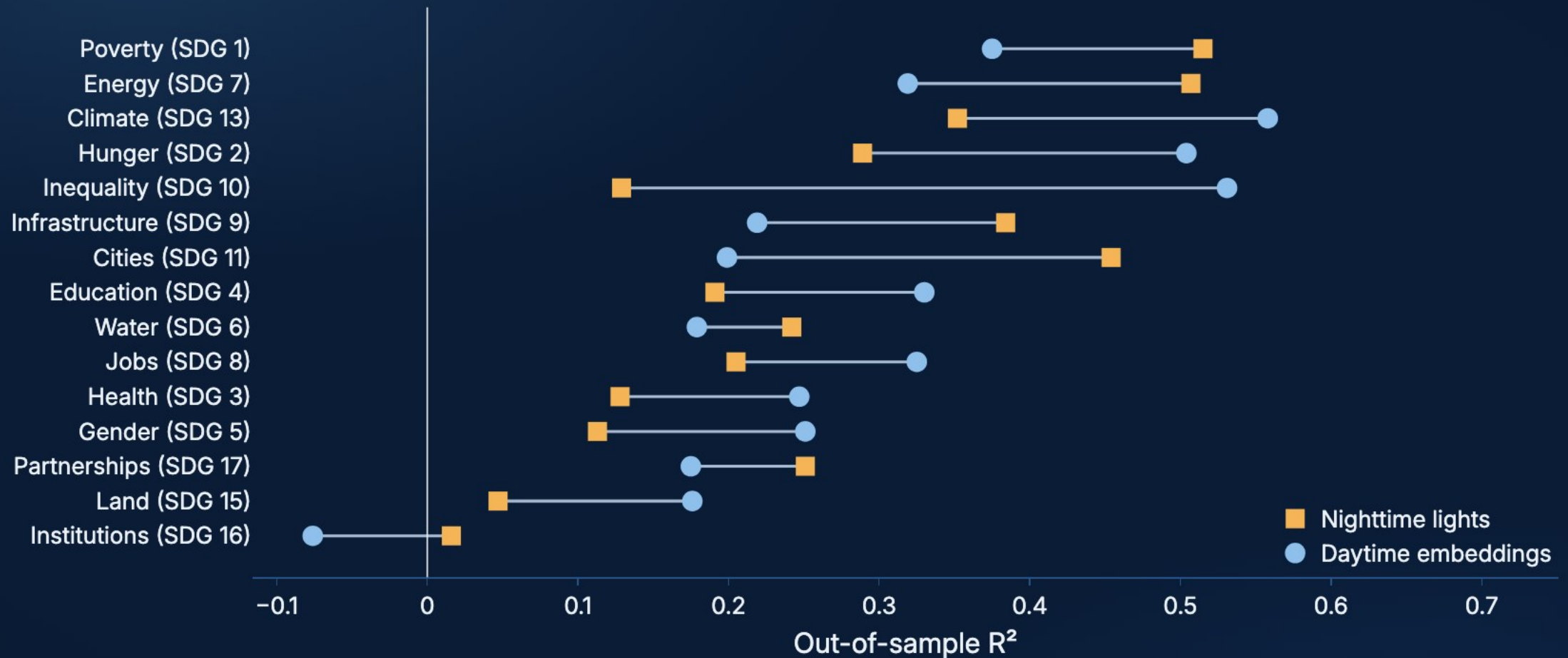
MONITORING · SPATIAL DEPENDENCE

- Neighbours are municipalities whose borders touch — queen contiguity.
- Global Moran's I — how strong is the clustering process.
- Local indicators of spatial association — hotspot, coldspot, or outlier.

Can satellites read development?

PART ONE · PREDICTION

On the raw average, nightlights win headline goals

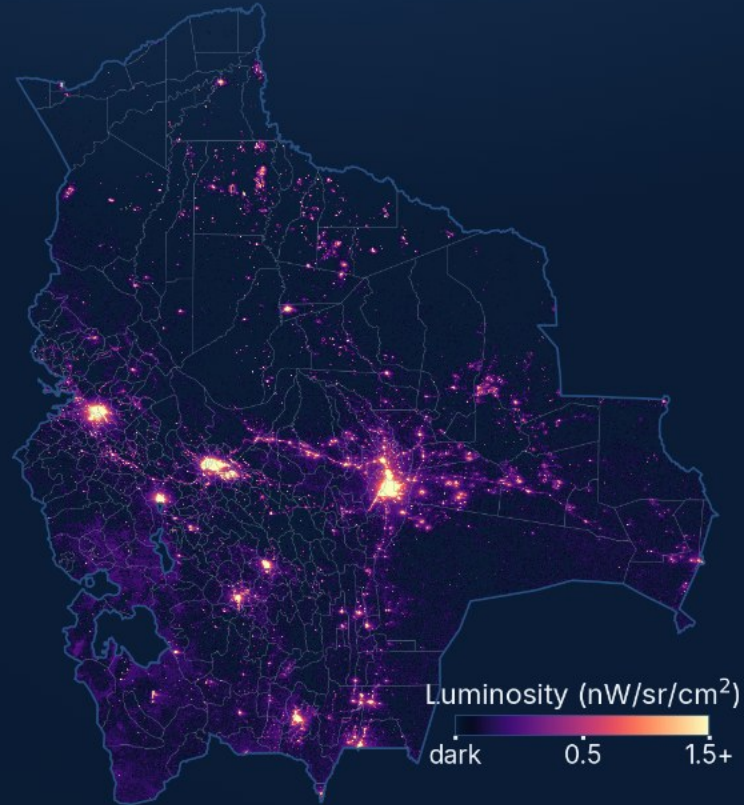


Simple average · out of sample · 339 municipalities · daytime embeddings lead on only 8 of 15 — and trail on poverty and energy

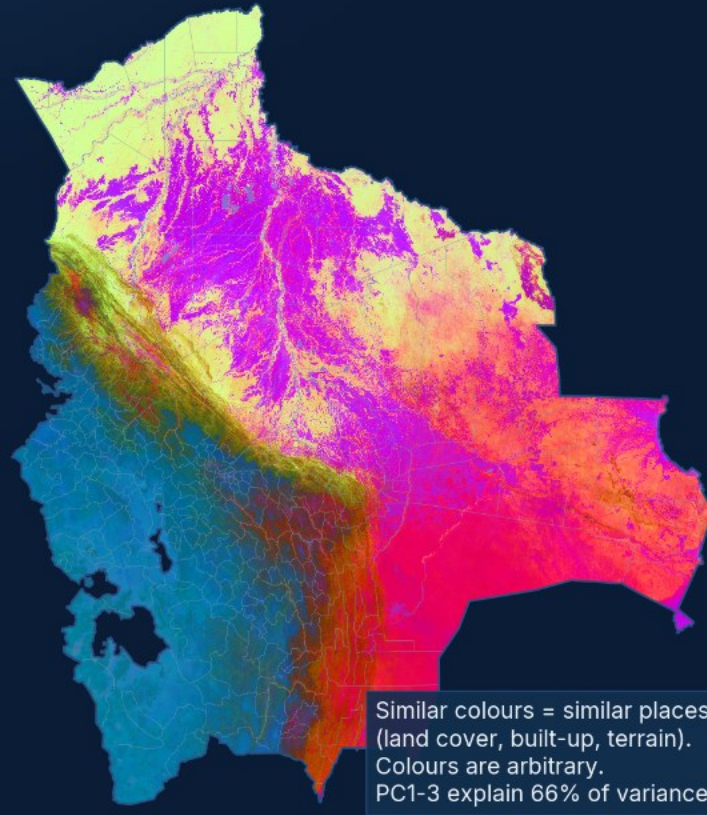
Most land is empty, let's weight each pixel by the people on it

KEY METHODOLOGICAL CHOICE

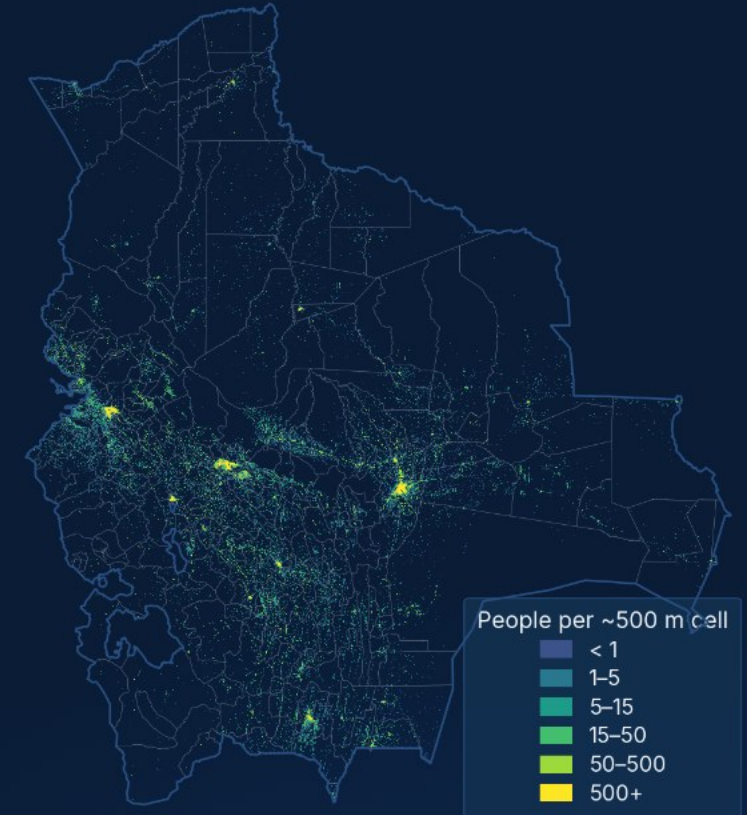
Nighttime lights



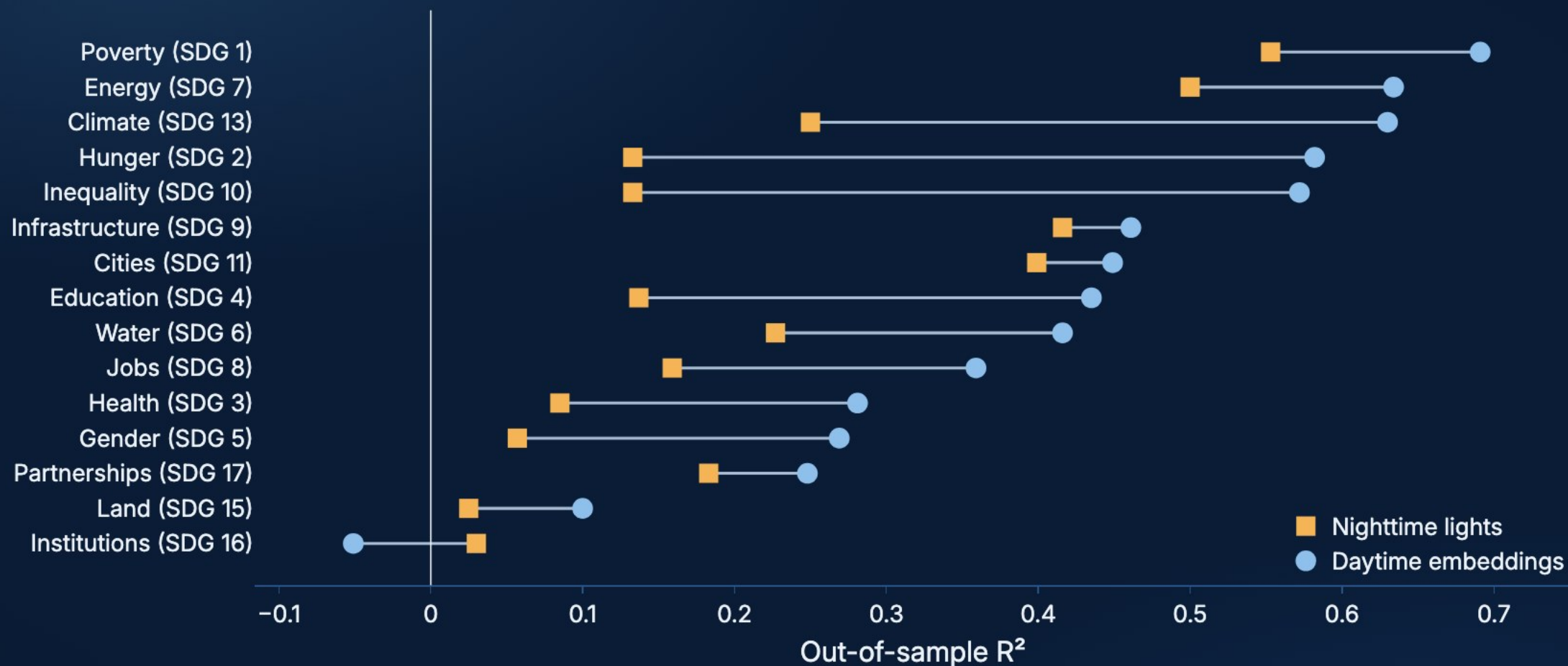
Daytime embeddings



Population



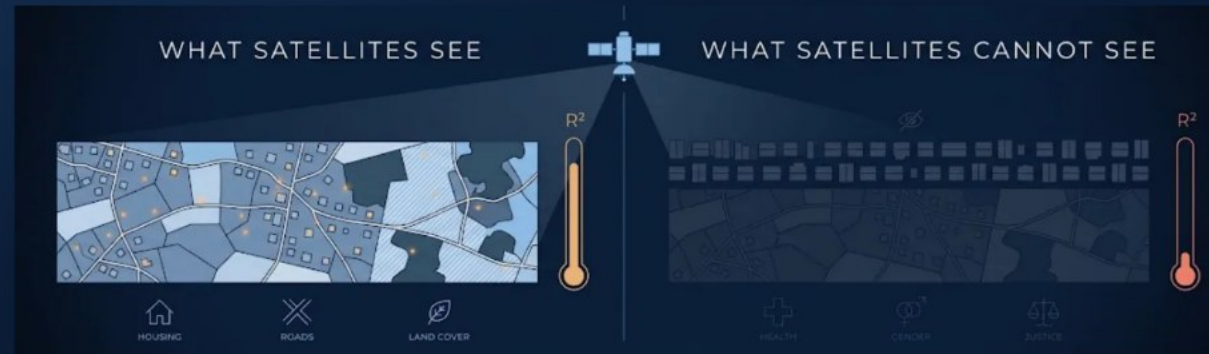
Weight pixels by population and daytime embeddings win



Weighted by population · out of sample · daytime embeddings now lead on 14 of 15 goals (8 → 14) — though the lead clears fold noise on only 5 of the 15

What a goal is made of decides whether satellites see it

DISCUSSION



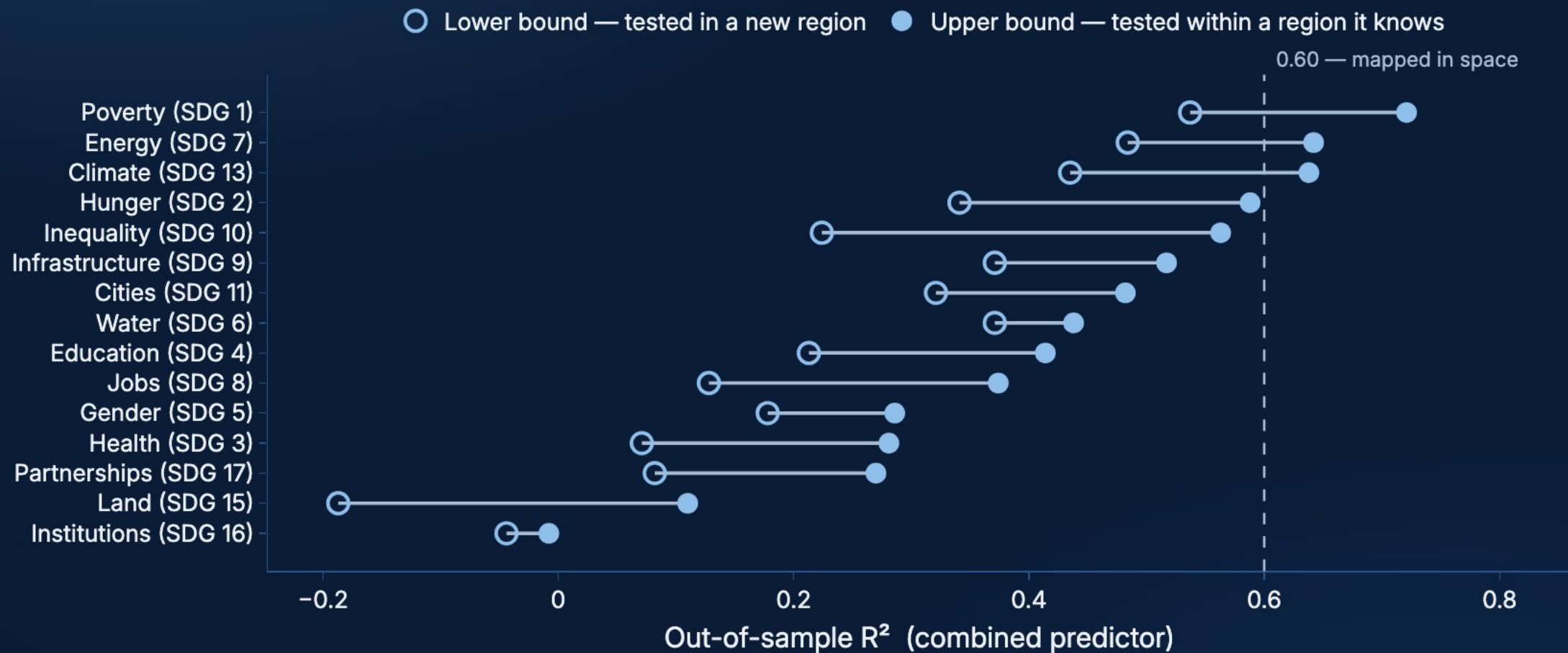
- Some goals are built from things satellites can **see**: housing, roads, land cover.
- Others from things they **cannot**: health, gender, justice.

Poverty leaves a footprint visible from outer space ($R^2 = 0.69$); institutions leave none ($R^2 = -0.05$).

Which goals can we map and how well?

PART TWO · MONITORING

Integrating nighttime lights and daytime embeddings provides a better prediction



Each goal spans the two ways of testing · The three goals above the 0.60 line are the ones we map next

The combined satellite view recovers the geography of no poverty (SDG1)

(a) Actual
reference · $I = 0.50$

(b) Nighttime lights
77% of all clusters · 53% of hot/cold
Moran's I 0.33 (actual 0.50)

(c) Daytime embeddings
78% of all clusters · 80% of hot/cold
Moran's I 0.60 (actual 0.50)

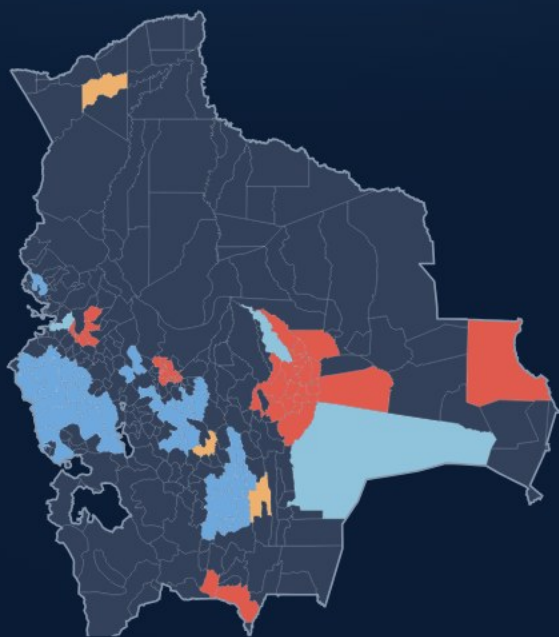
(d) Combined
80% of all clusters · 75% of hot/cold
Moran's I 0.50 (actual 0.50)

■ Low-poverty cluster ■ Poverty trap ■ Outlier: high among low ■ Outlier: low among high ■ Not significant

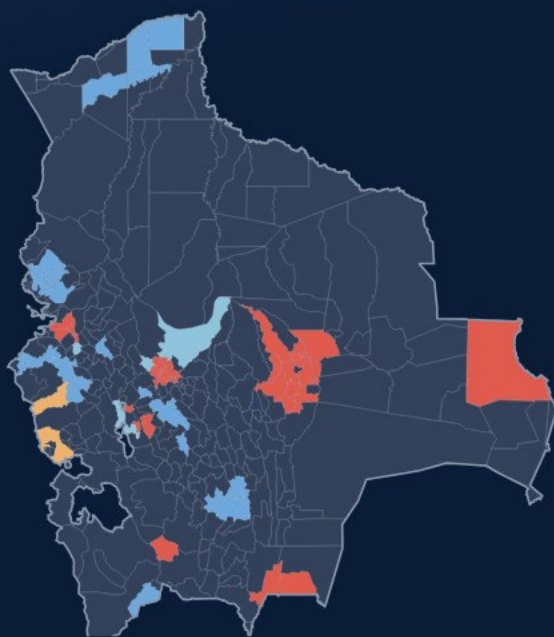
Red: clusters of low poverty · Blue: poverty traps · Each panel reports its agreement with the actual map and its clustering strength

The combined satellite view recovers the geography of affordable and clean energy (SDG7)

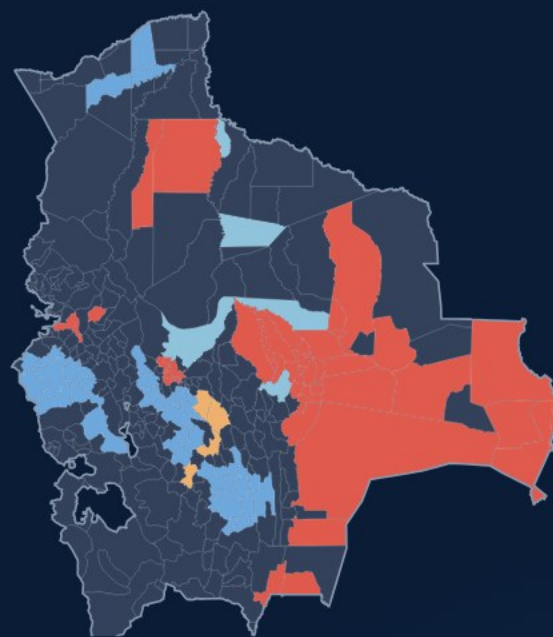
(a) Actual
reference · $I = 0.43$



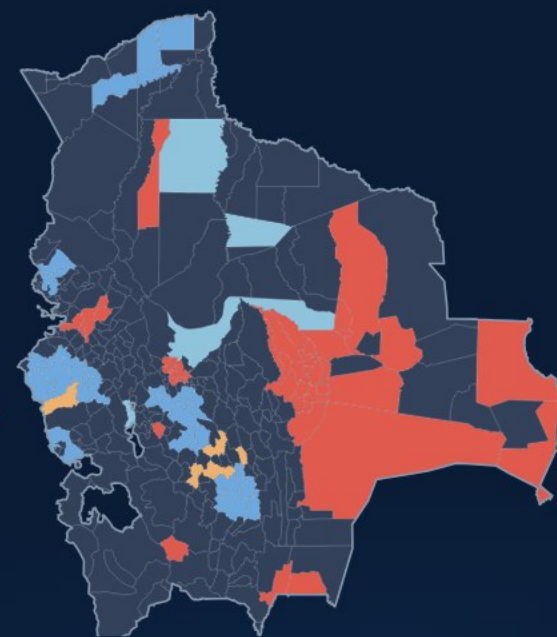
(b) Nighttime lights
77% of all clusters · 49% of hot/cold
Moran's I 0.28 (actual 0.43)



(c) Daytime embeddings
83% of all clusters · 75% of hot/cold
Moran's I 0.49 (actual 0.43)



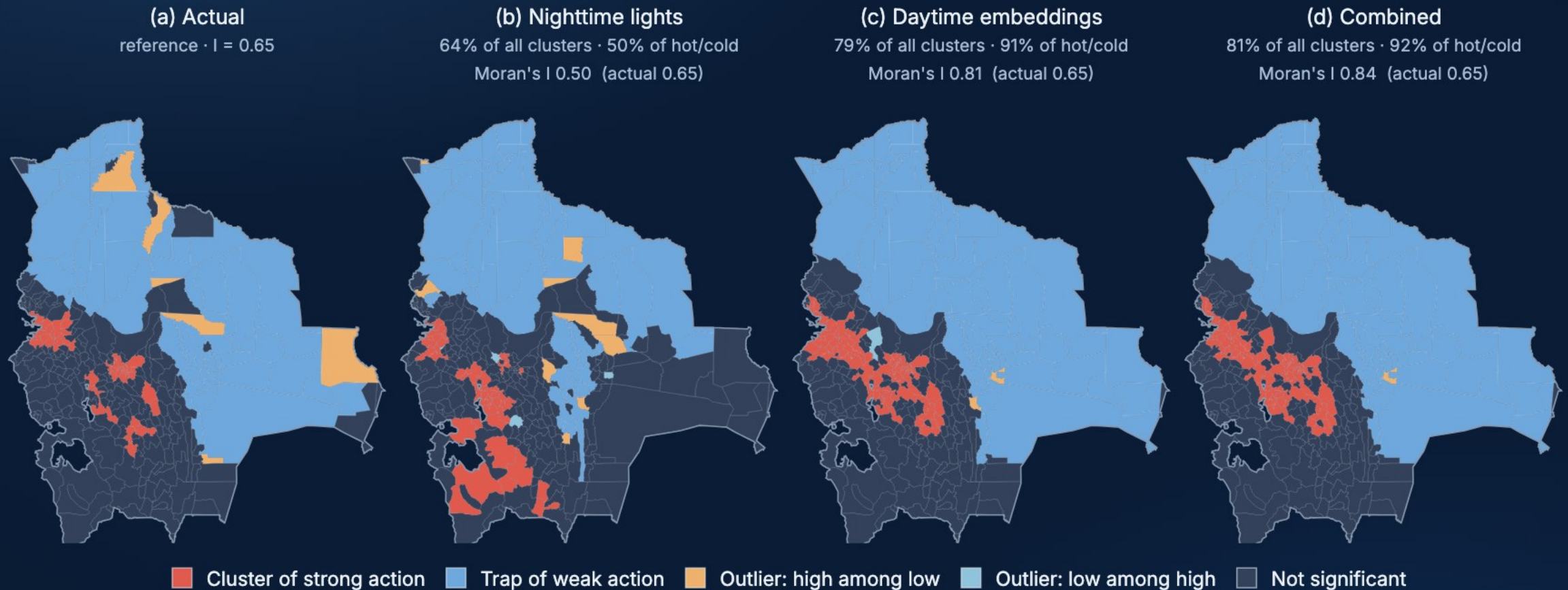
(d) Combined
82% of all clusters · 70% of hot/cold
Moran's I 0.40 (actual 0.43)



■ Cluster of good access ■ Trap of poor access ■ Outlier: high among low ■ Outlier: low among high ■ Not significant

Red: clusters of good access · Blue: traps of poor access · The lights reach $r = 0.73$ on the level, and half that on the pattern

The combined satellite view recovers the geography of climate action (SDG13)



Red: clusters of strong action · Blue: traps of weak action · Best cluster recovery of the three goals — but both embedding views make the map look more clustered than it is

Satellites can monitor the material geography of development

CONCLUDING REMARKS

1 · WEIGHT BY WHERE PEOPLE LIVE

Daytime embeddings' average R^2 rises from **0.29** to **0.41**; they lead on **14** of 15 goals.

2 · WHAT A GOAL IS MADE OF MATTERS

Poverty reads at **0.69**; health, gender and institutions barely register.

3 · COMBINE BOTH VIEWS

Lights plus embeddings read poverty at **0.54 to 0.72** out of sample, across the two tests.

4 · MAP THE CLUSTERS

Poverty's clusters reappear: **80%** agreement with the actual map.

Use multiple population-weighted satellite data to better monitor local development **between** surveys and census.

Predicting and monitoring local development from outer space

Evidence from the Bolivian municipalities

Carlos Mendez

<https://carlos-mendez.org>

Graduate School of International Development

Nagoya University

Japan

Thank you

