Producing Socio-Economic Information with Earth Observations and AI Innovations

Instructions: Click on the link to access each author's presentation.

Organiser and Chair: Elio Villaseñor

Participants:

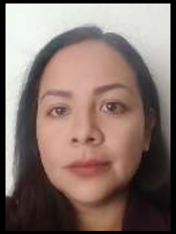
Joaquin Salas: Quick and reliable socio-economic assessments from remote sensing and machine learning

Abel Coronado: Identification of Agricultural Land Use through Remote Sensing and Machine Learning Techniques

Ranyart Rodrigo Suárez Ponce de León: Estimating Economic Activity Indicators at the Municipal Level in Mexico Using Nighttime Lights

Quick and reliable socio-economic assessments using remote sensing and machine learning

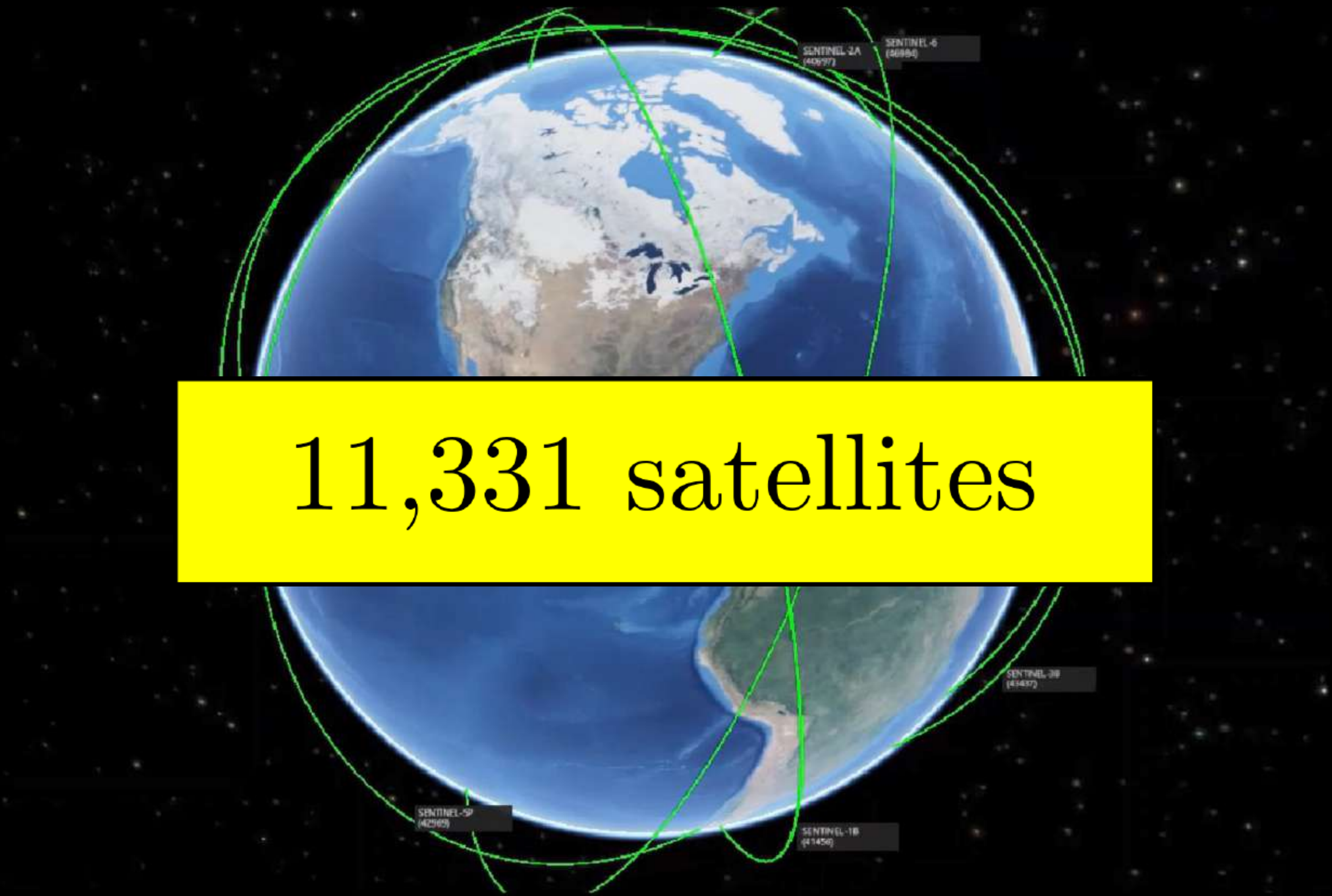
IAOS-ISI 2024
MEXICO CONFERENCE
INEGI
isi International Statistical Institute
IAOS IMPROVING OFFICIAL STATISTICS






Alejandra Figueroa • Roberto Manduchi • Ranyart Suarez • Pablo Vera • Elio-Atenógenes Villaseñor • Marivel Zea-Ortiz

Joaquin Salas
joaquín.salas@gmail.com
jsalasr@ipn.mx

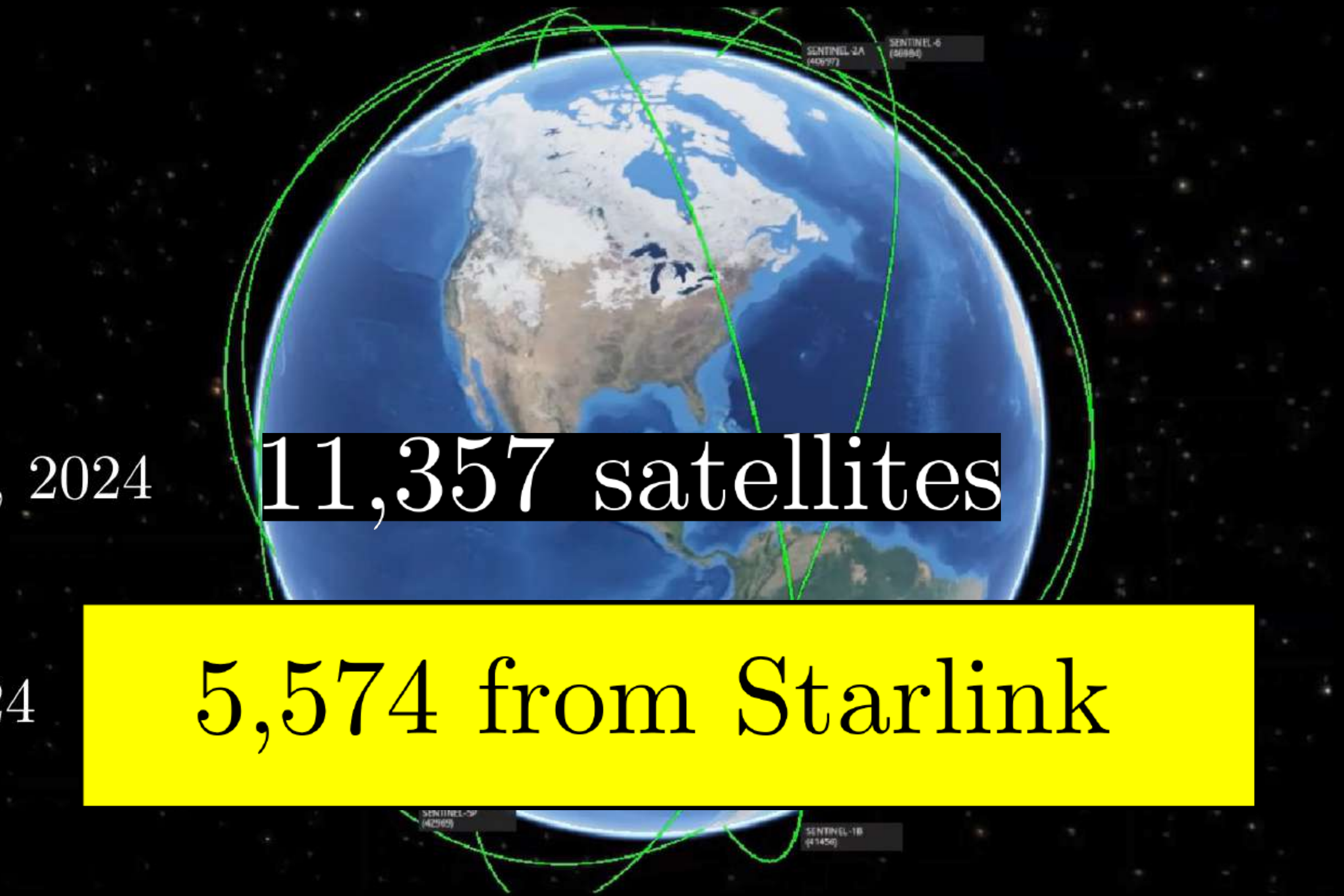


11,331 satellites

https://celestrak.org/

April 25, 2024

https://www.unoosa.org/oosa/osoindex/search-ng.jspx

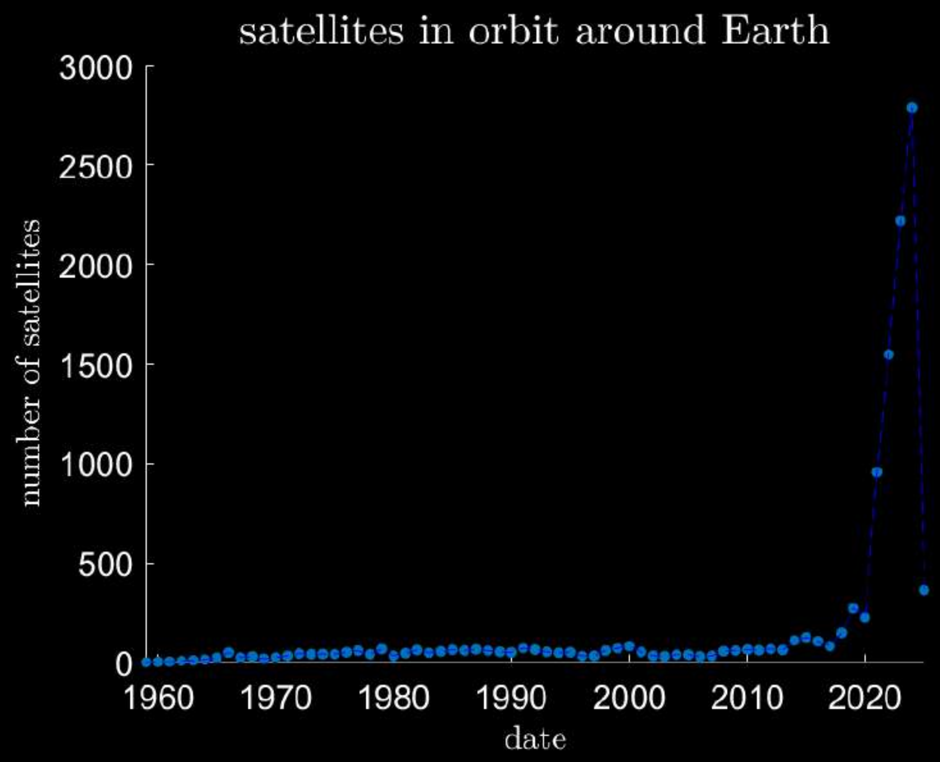


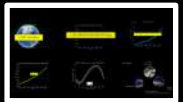
April 25, 2024 11,357 satellites

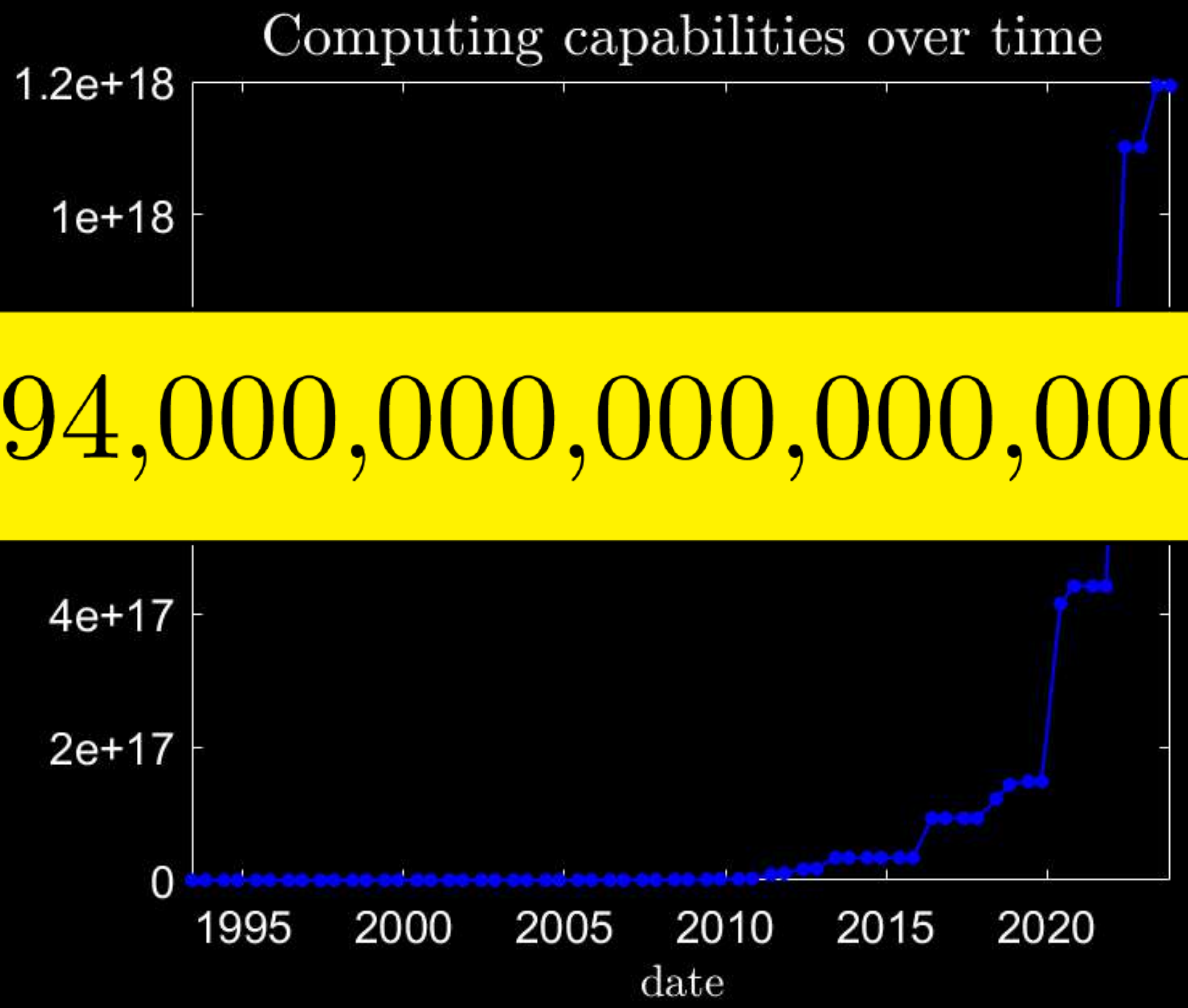
April 25, 2024 5,574 from Starlink

https://celestrak.org/

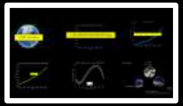
https://www.unoosa.org/oosa/osoindex/search-ng.jsp


satellites in orbit around Earth
number of satellites
3000
2500
2000
1500
1000
500
0
1960
1970
1980
1990
2000
2010
2020
date

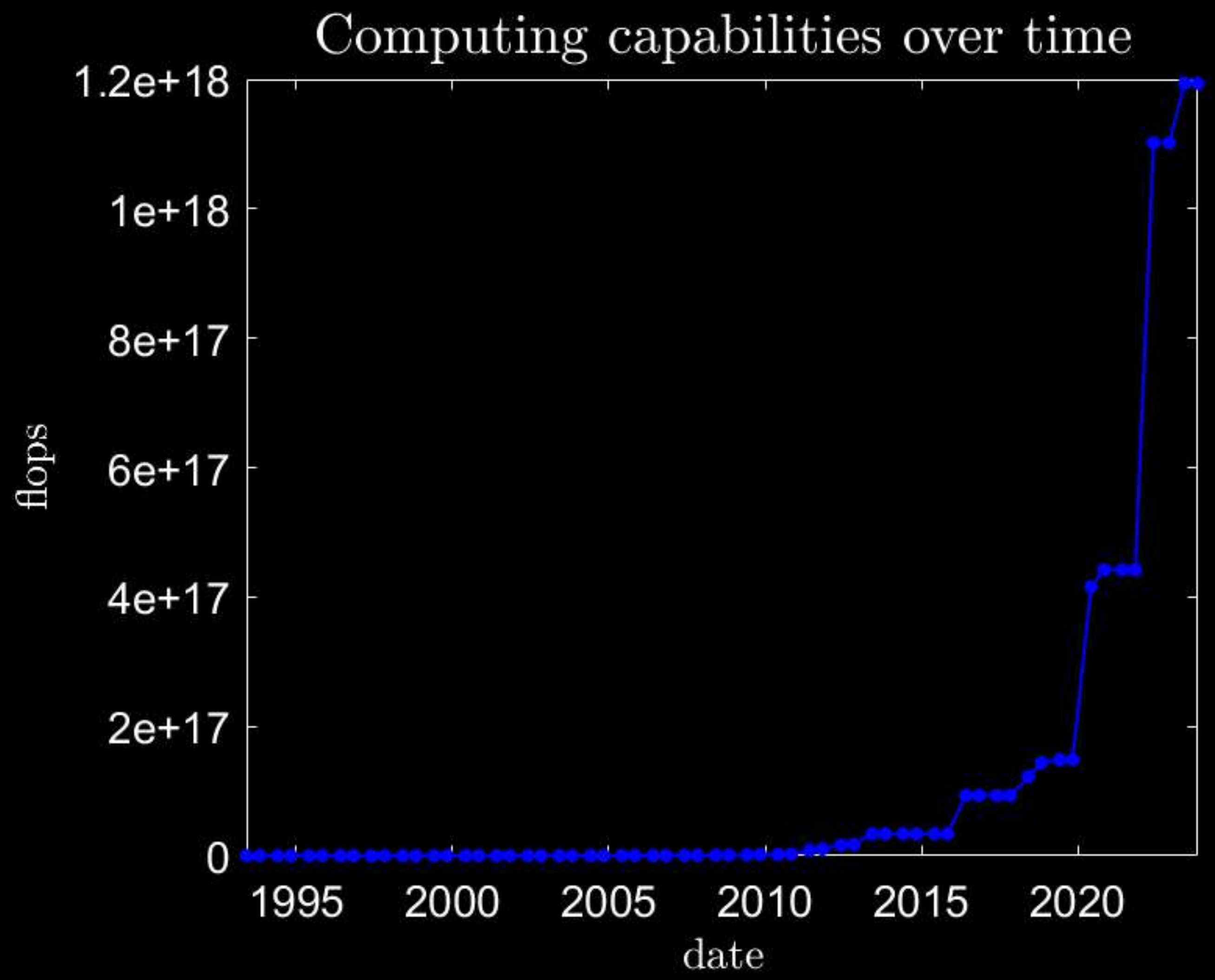




1,194,000,000,000,000,000 flops

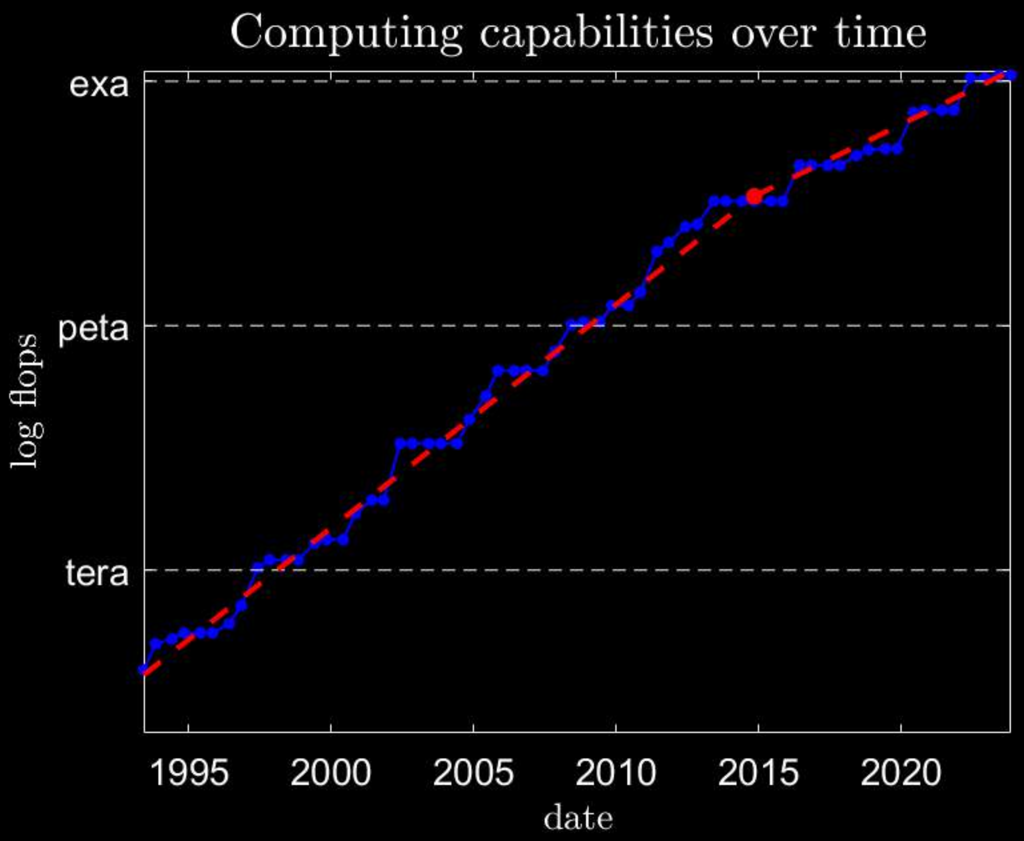


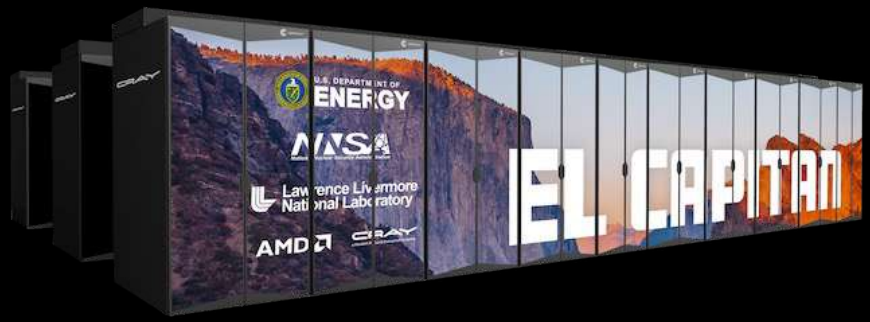
top500.org



top500.org

1,194,000,000,000,000,000 flops



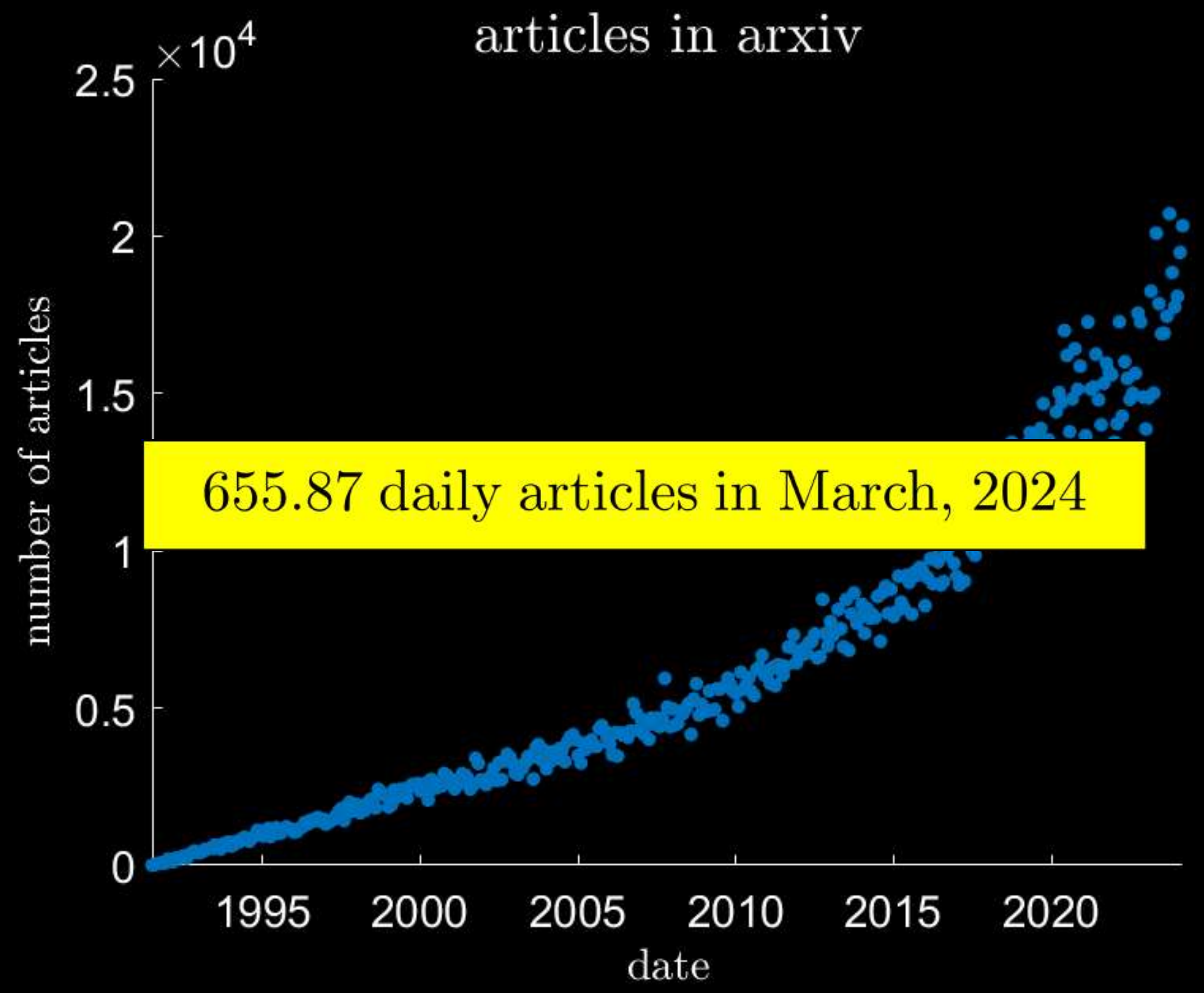


1,194,000,000,000,000,000 flops

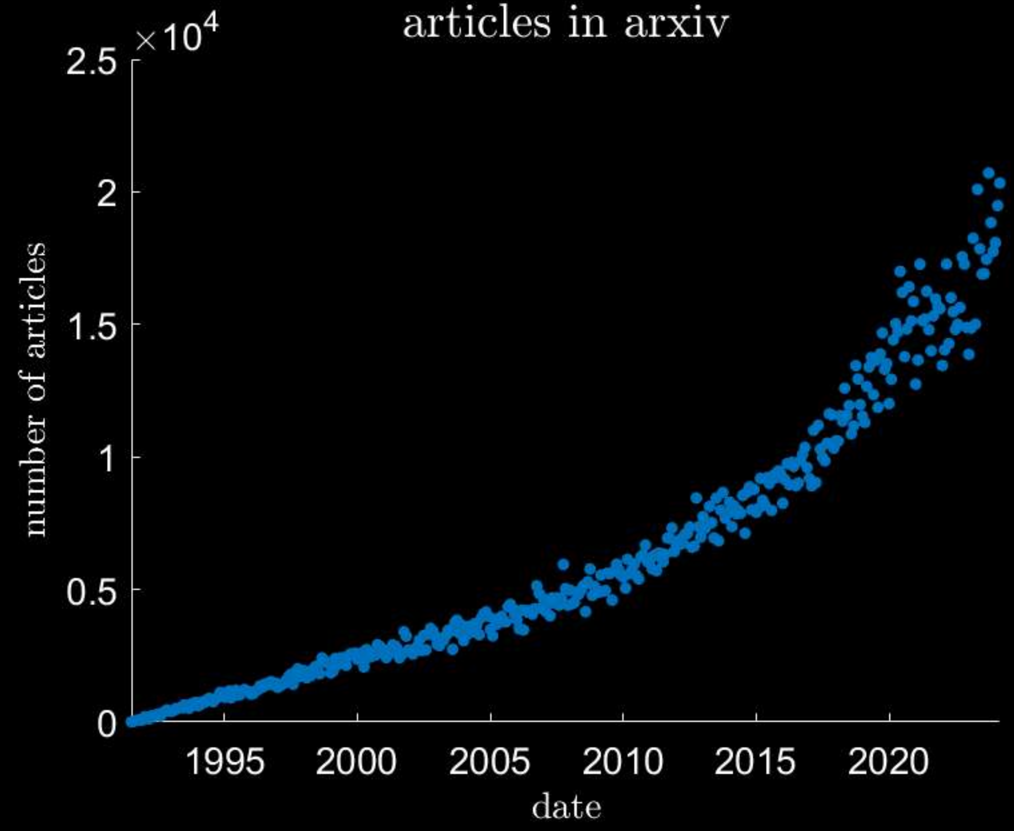
top500.org

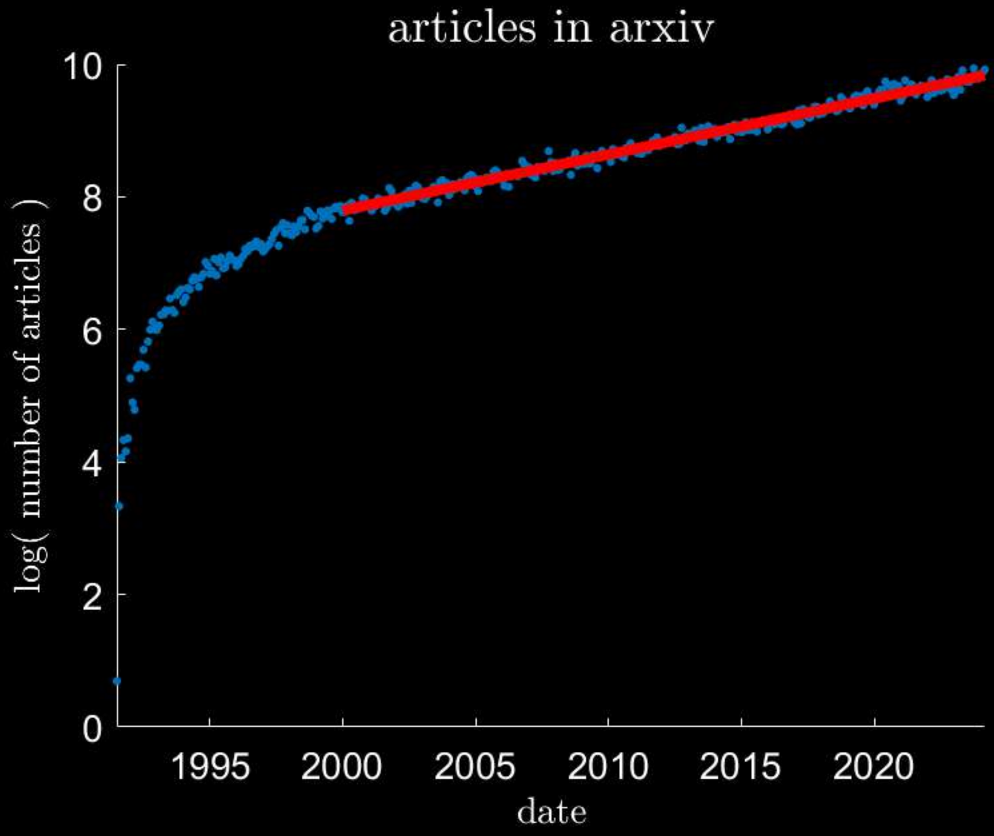
https://www.intel.com

https://www.llnl.gov/

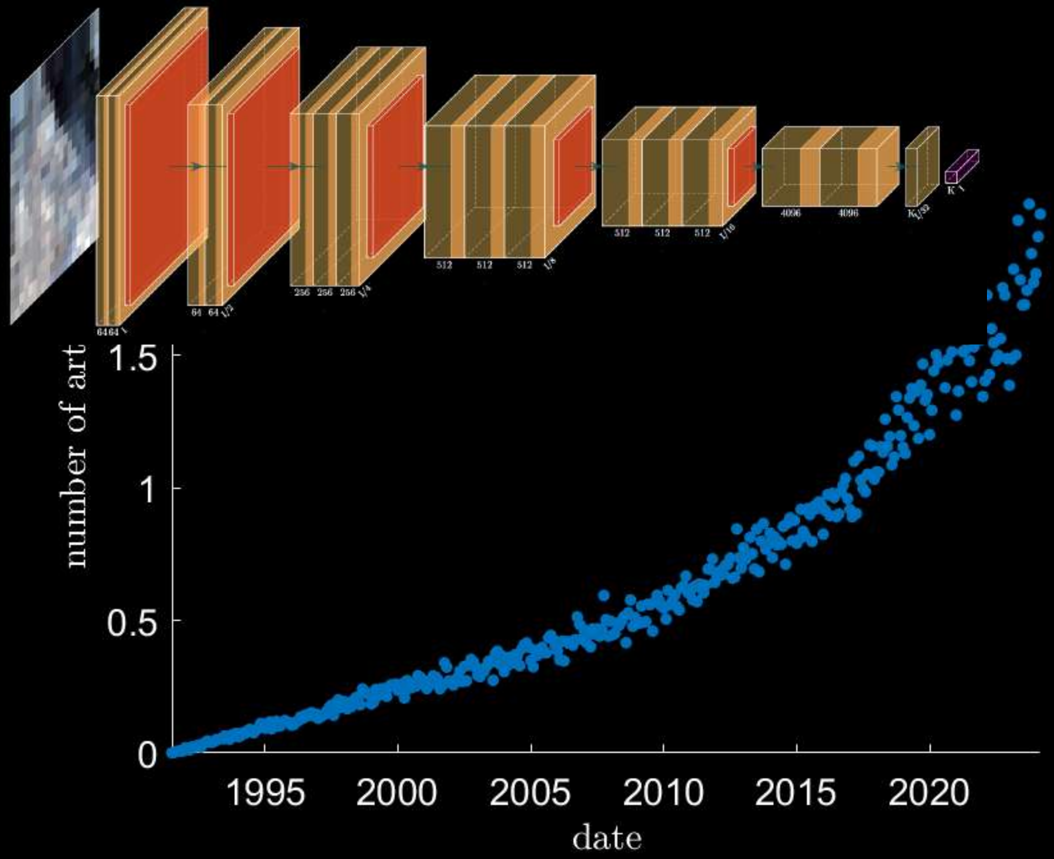


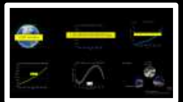
https://arxiv.org/stats/monthly_submissions

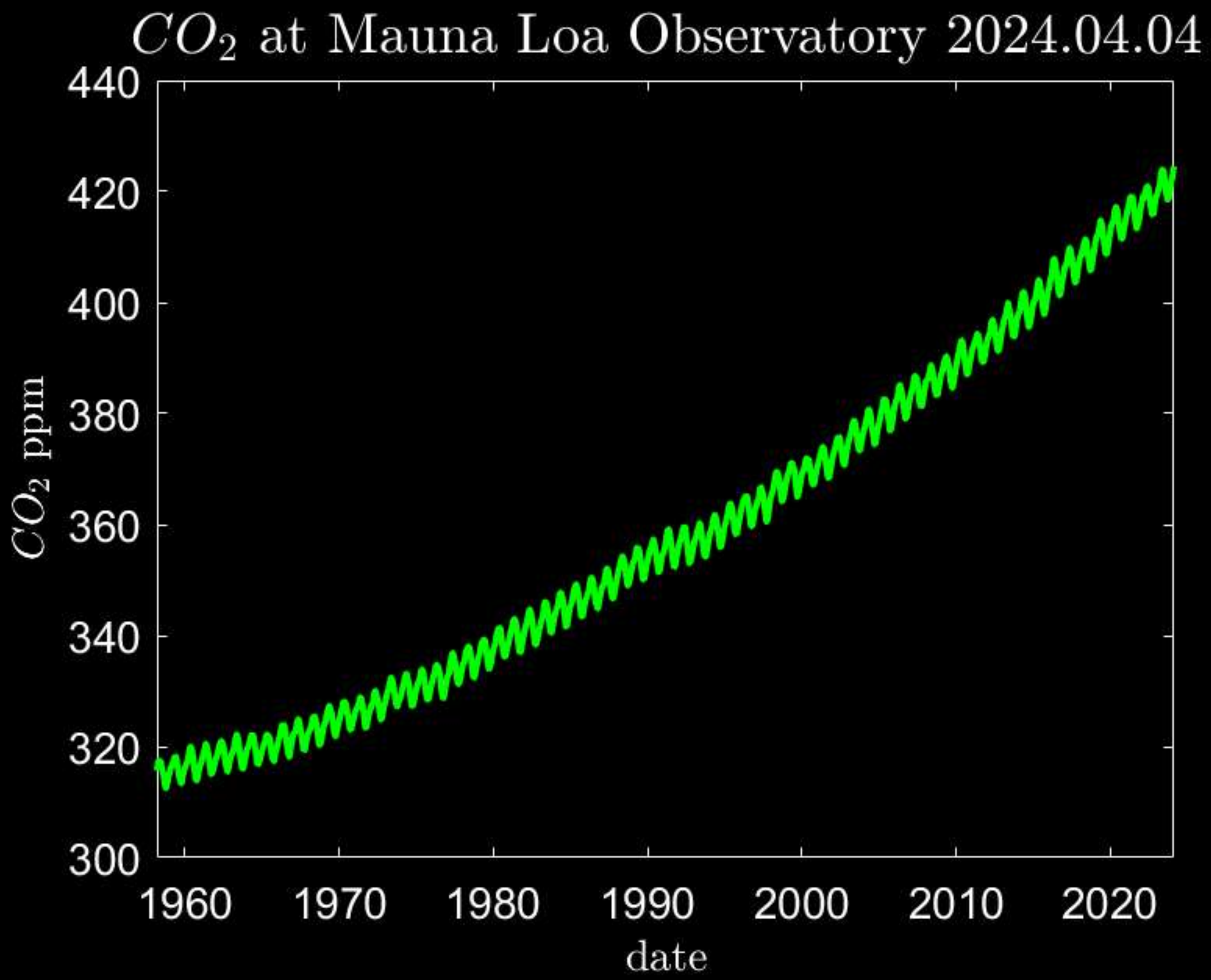


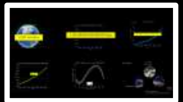


https://arxiv.org/stats/monthly_submissions

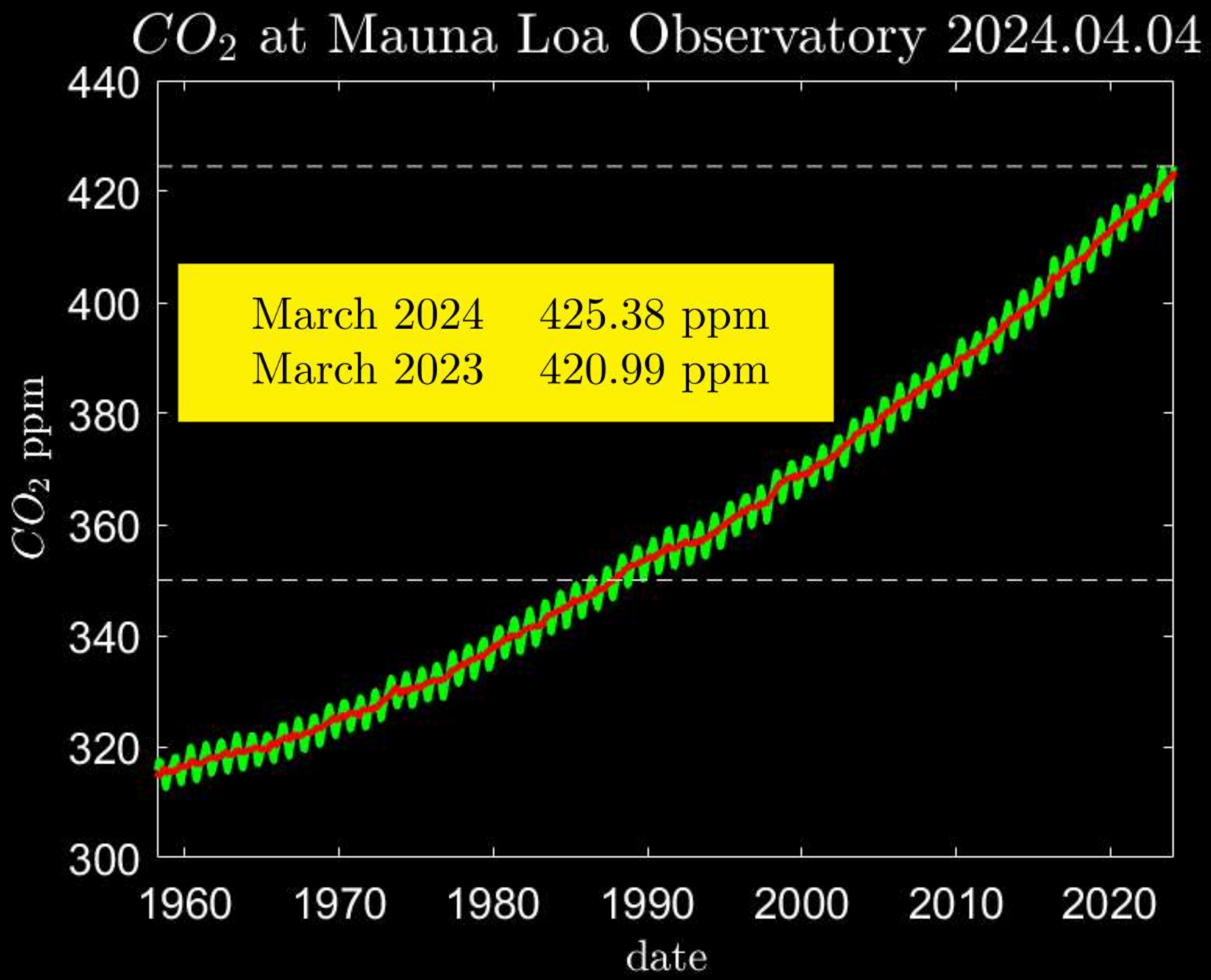

number of art
1.5
1
0.5
0
1995
2000
2005
2010
2015
2020
date







https://github.com/datasets/co2-ppm/blob/master/data/co2-mm-mlo.csv



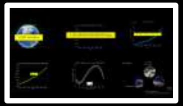
https://github.com/datasets/co2-ppm/blob/master/data/co2-mm-mlo.csv

Sustainability

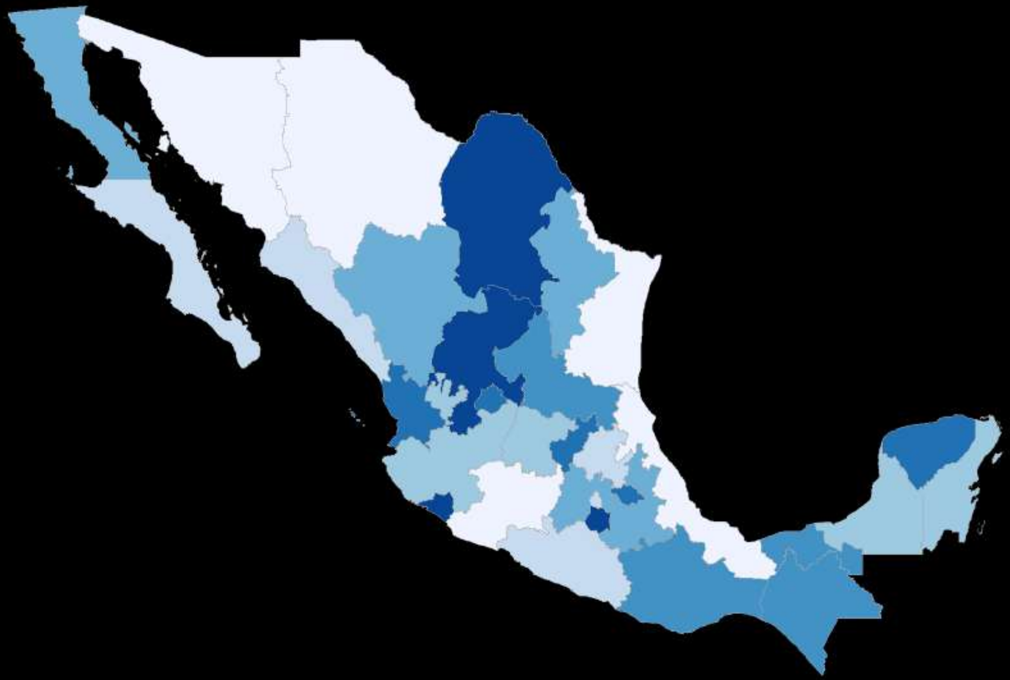
human
vulnerability

biodiversity

climate
change



CONEVAL
multidimensional
poverty model



State level
every two years
$\Longrightarrow$

CONEVAL
multidimensional
poverty model

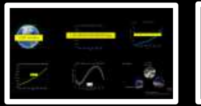
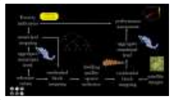
State level
every two years
$\Longrightarrow$

CONEVAL
multidimensional
poverty model

Municipal level
every five years
$\Longrightarrow$

2022 assessment

- 105,525 goal survey size
- $\approx$ 3,300 for each of 32 states
- 2,712 field workers
- 412 office workers
- August 11 to November 28, 2022
- 84.4% complete interviews
- 61.6% direct informant, 22.7% indirect informant
- results published nine months later

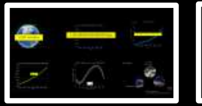
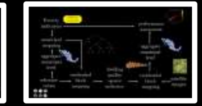



2022 assessment

- 105,525 goal survey size
- $\approx$ 3,300 for each of 32 states
- 2,712 field workers
- 412 office workers
- August 11 to November 28, 2022
- 84.4% complete interviews
- 61.6% direct informant, 22.7% indirect informant
- results published nine months later

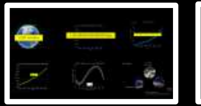
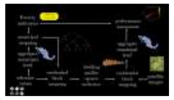
2022 assessment

- 105,525 goal survey size
- ≈ 3,300 for each of 32 states
- 2,712 field workers
- 412 office workers
- August 11 to November 28, 2022
- 84.4% complete interviews
- 61.6% direct informant, 22.7% indirect informant
- results published nine months later

2022 assessment

- 105,525 goal survey size
- $\approx$ 3,300 for each of 32 states
- 2,712 field workers
- 412 office workers
- August 11 to November 28, 2022
- 84.4% complete interviews
- 61.6% direct informant, 22.7% indirect informant
- results published nine months later

2022 assessment

- 105,525 goal survey size
- $\approx$ 3,300 for each of 32 states
- 2,712 field workers
- 412 office workers
- August 11 to November 28, 2022
- 84.4% complete interviews
- 61.6% direct informant, 22.7% indirect informant
- results published nine months later

2022 assessment

- 105,525 goal survey size
- $\approx$ 3,300 for each of 32 states
- 2,712 field workers
- 412 office workers
- August 11 to November 28, 2022
- 84.4% complete interviews
- 61.6% direct informant, 22.7% indirect informant
- results published nine months later

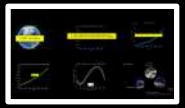
2022 assessment

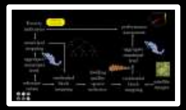
- 105,525 goal survey size
- $\approx$ 3,300 for each of 32 states
- 2,712 field workers
- 412 office workers
- August 11 to November 28, 2022
- 84.4% complete interviews
- 61.6% direct informant, 22.7% indirect informant
- results published nine months later

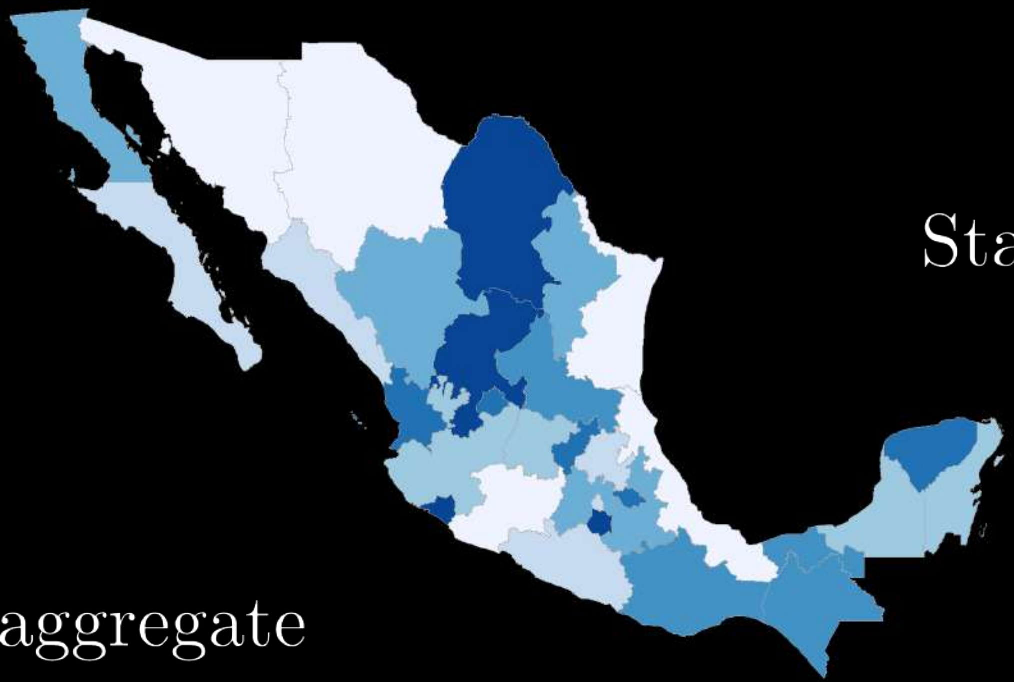
2022 assessment

- 105,525 goal survey size
- $\approx$ 3,300 for each of 32 states
- 2,712 field workers
- 412 office workers
- August 11 to November 28, 2022
- 84.4% complete interviews
- 61.6% direct informant, 22.7% indirect informant
- results published nine months later

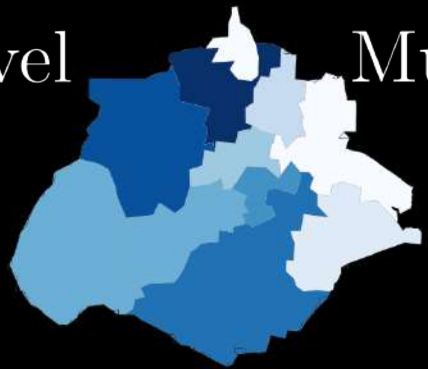
reference
values







State level $\Longrightarrow$



Municipal level $\Longrightarrow$

aggregate
municipal
level

$\uparrow$

reference
values

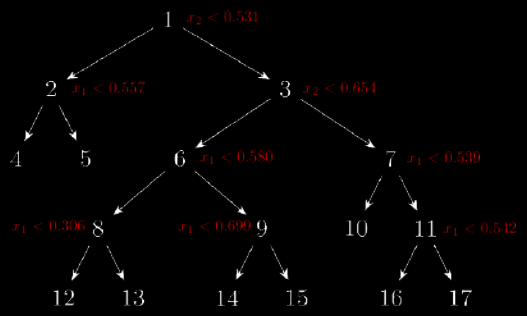
Poverty indicators

municipal mapping

aggregate municipal level

reference values

CONEVAL multidimensional poverty model



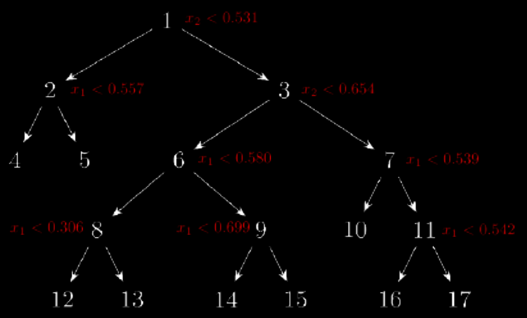
Poverty indicators

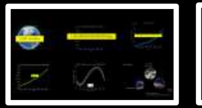
municipal mapping

aggregate municipal level

reference values

residential block mapping





Poverty indicators

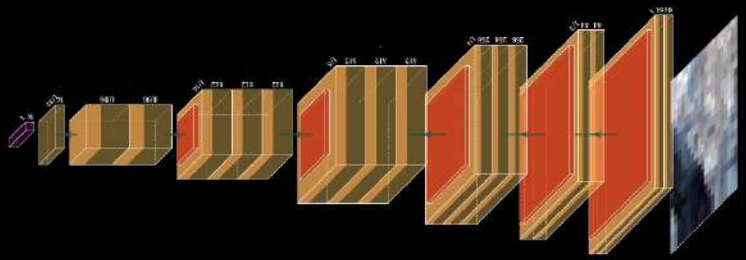
municipal mapping

aggregate municipal level

reference values

residential block mapping

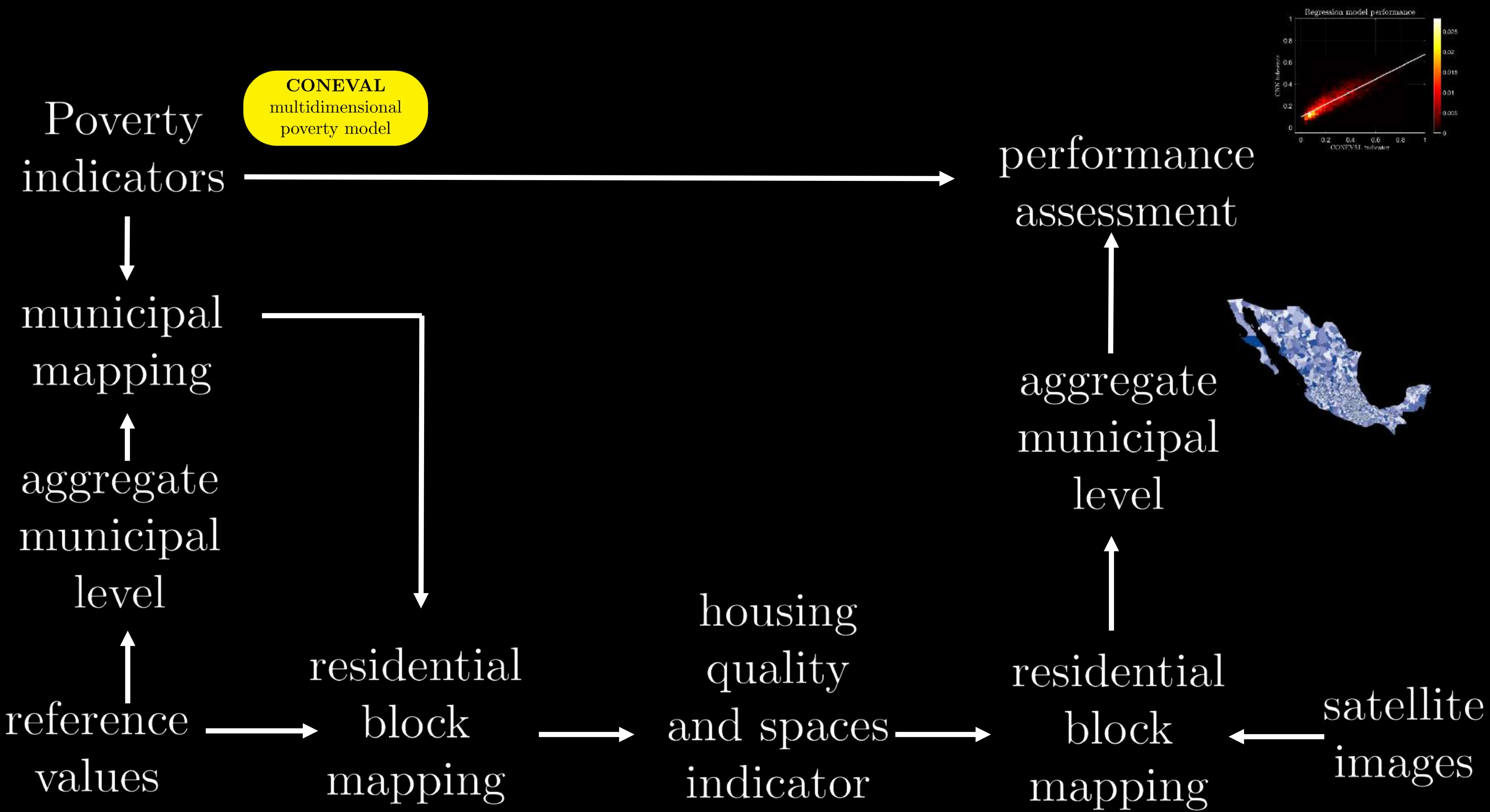
CONEVAL
multidimensional poverty model

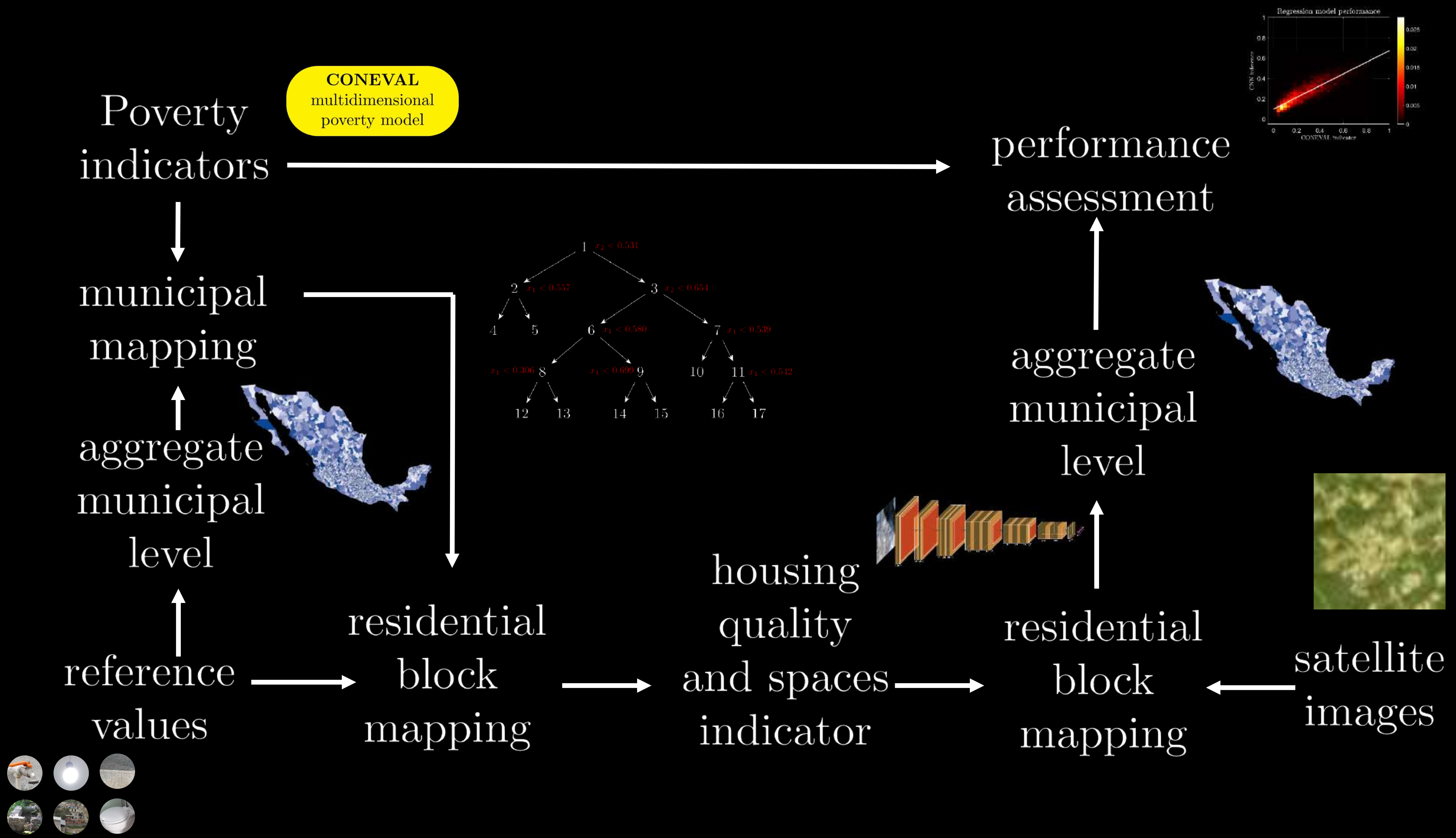


dwelling quality spaces indicator

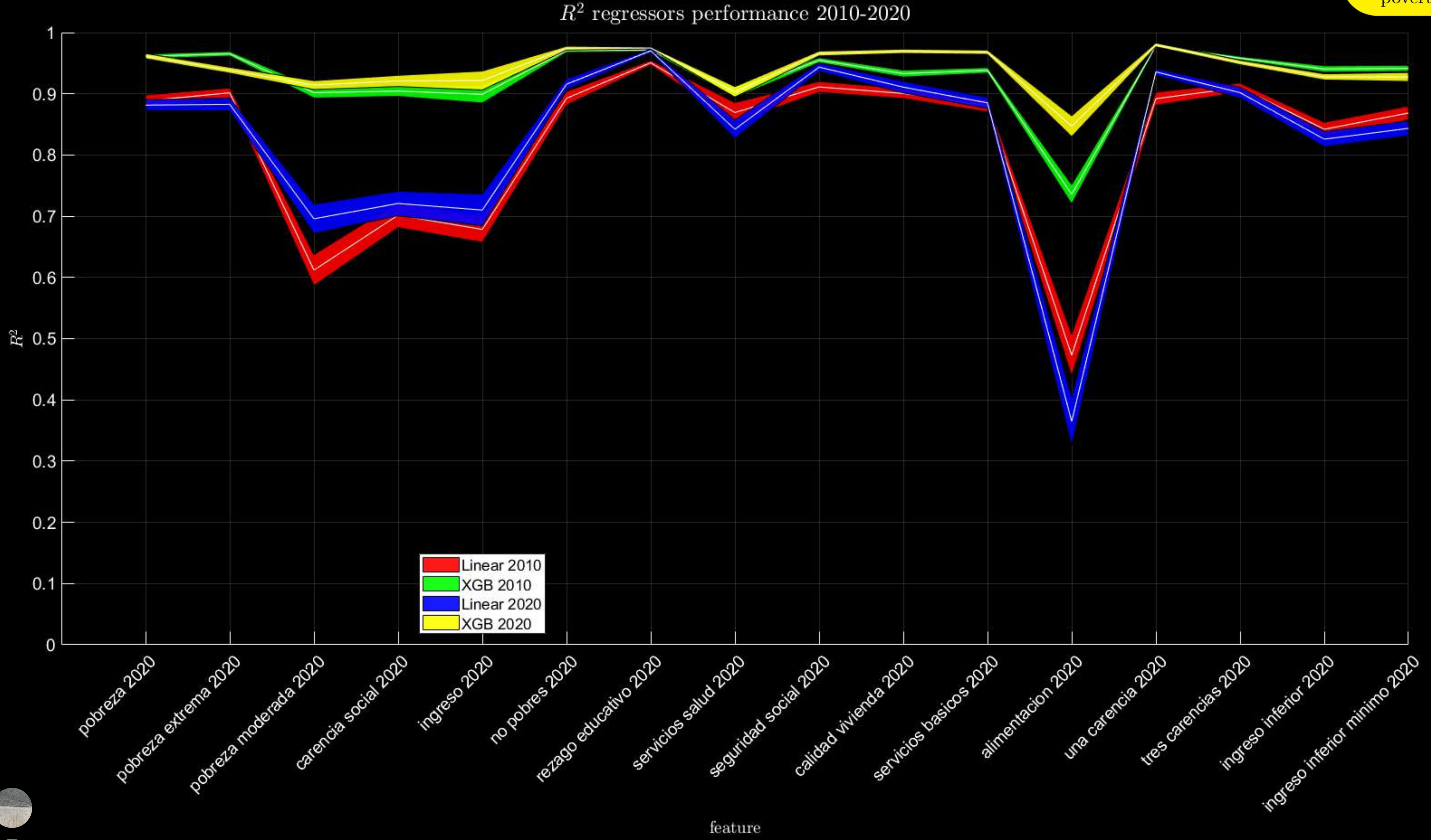
residential block mapping

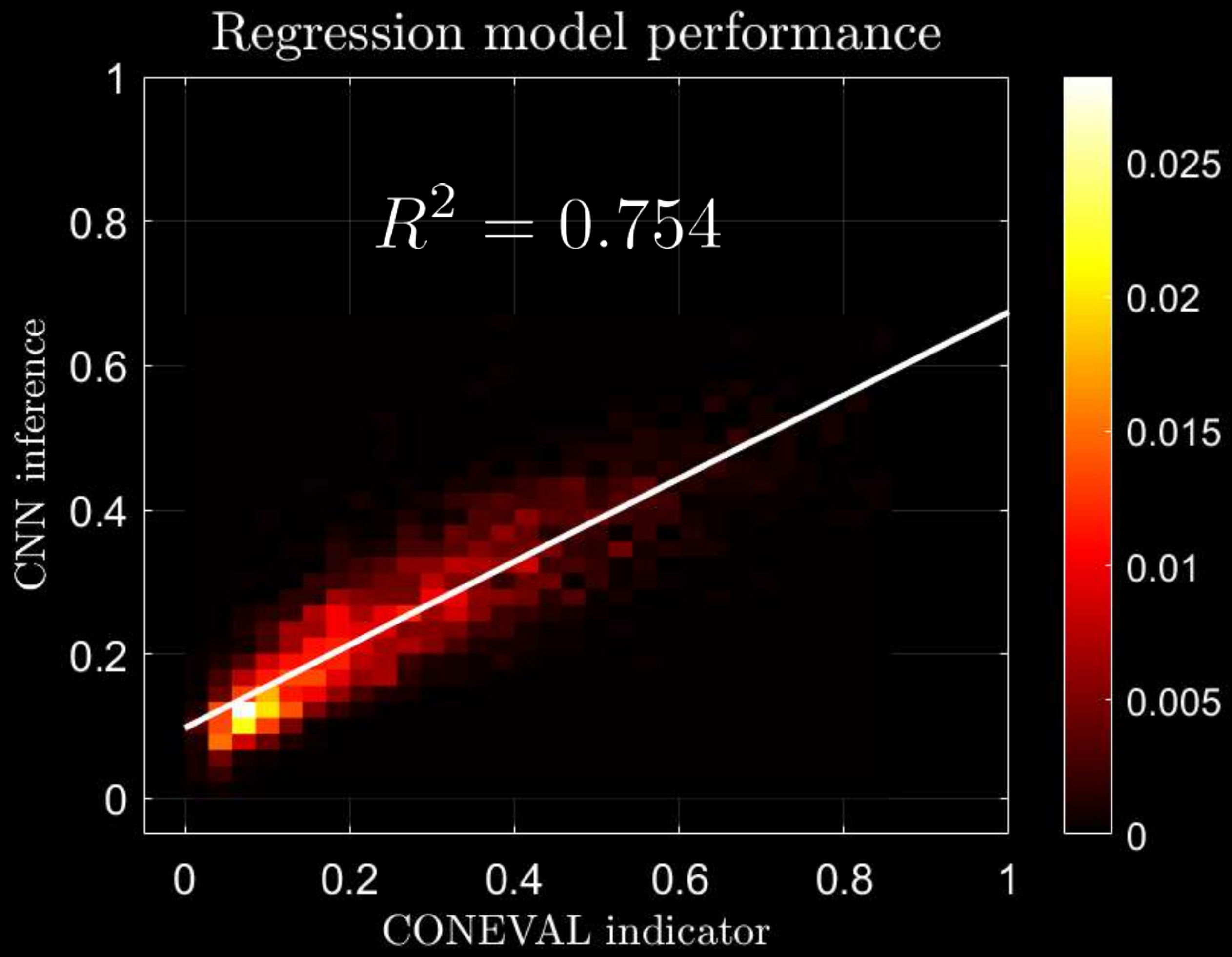
satellite images


Poverty indicators
CONEVAL multidimensional poverty model
municipal mapping
aggregate municipal level
reference values
residential block mapping
housing quality and spaces indicator
residential block mapping
satellite images
aggregate municipal level
performance assessment
Regression model performance
CNN inference
CONEVAL indicator
0
0.2
0.4
0.6
0.8
1
0.025
0.02
0.015
0.01
0.005

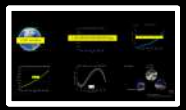

Poverty indicators
CONEVAL
multidimensional
poverty model
performance assessment
Regression model performance
CONEVAL indicator
municipal mapping
aggregate municipal level
aggregate municipal level
reference values
residential block mapping
housing quality and spaces indicator
residential block mapping
satellite images

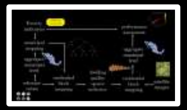
CONEVAL
multidimensional
poverty model

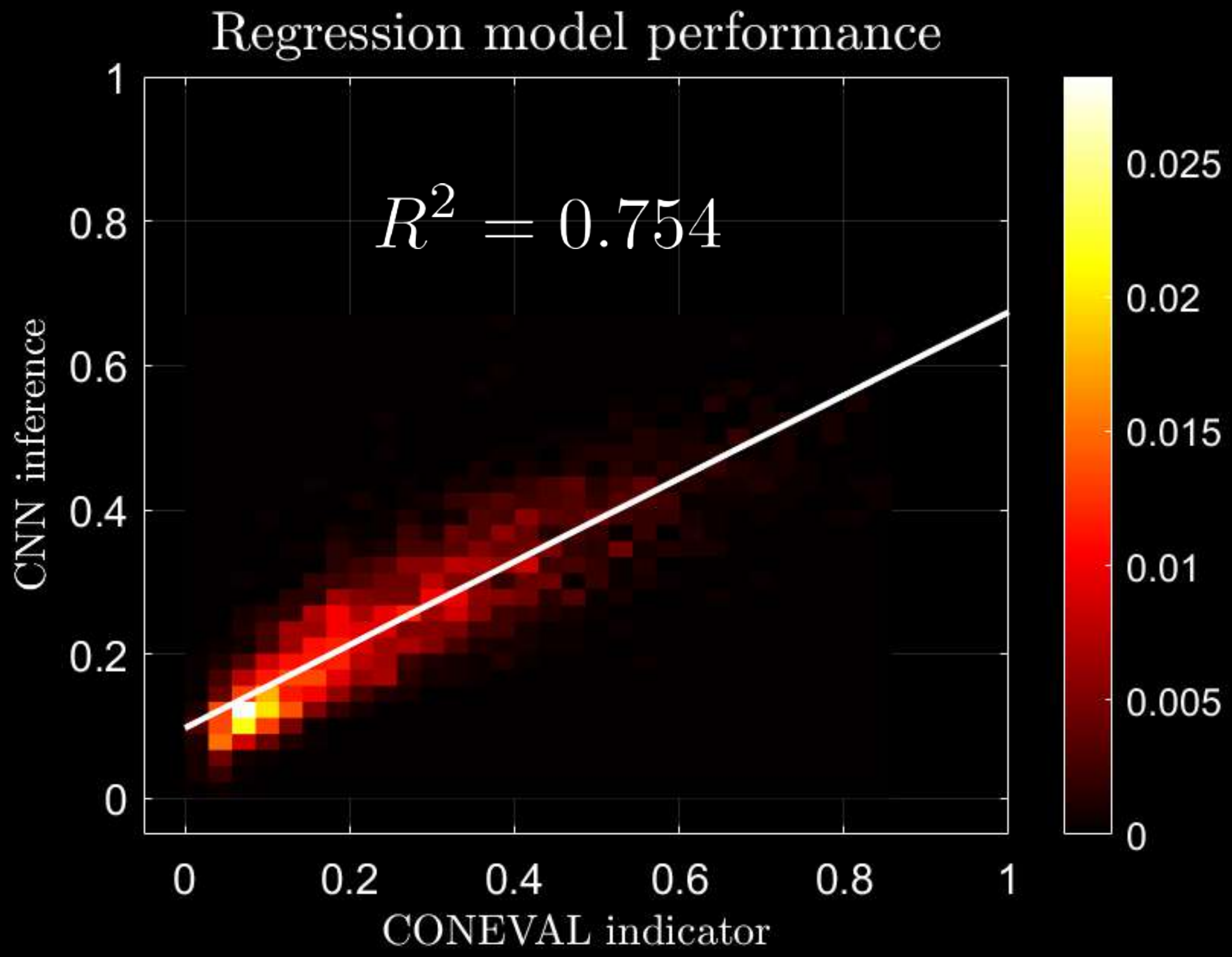




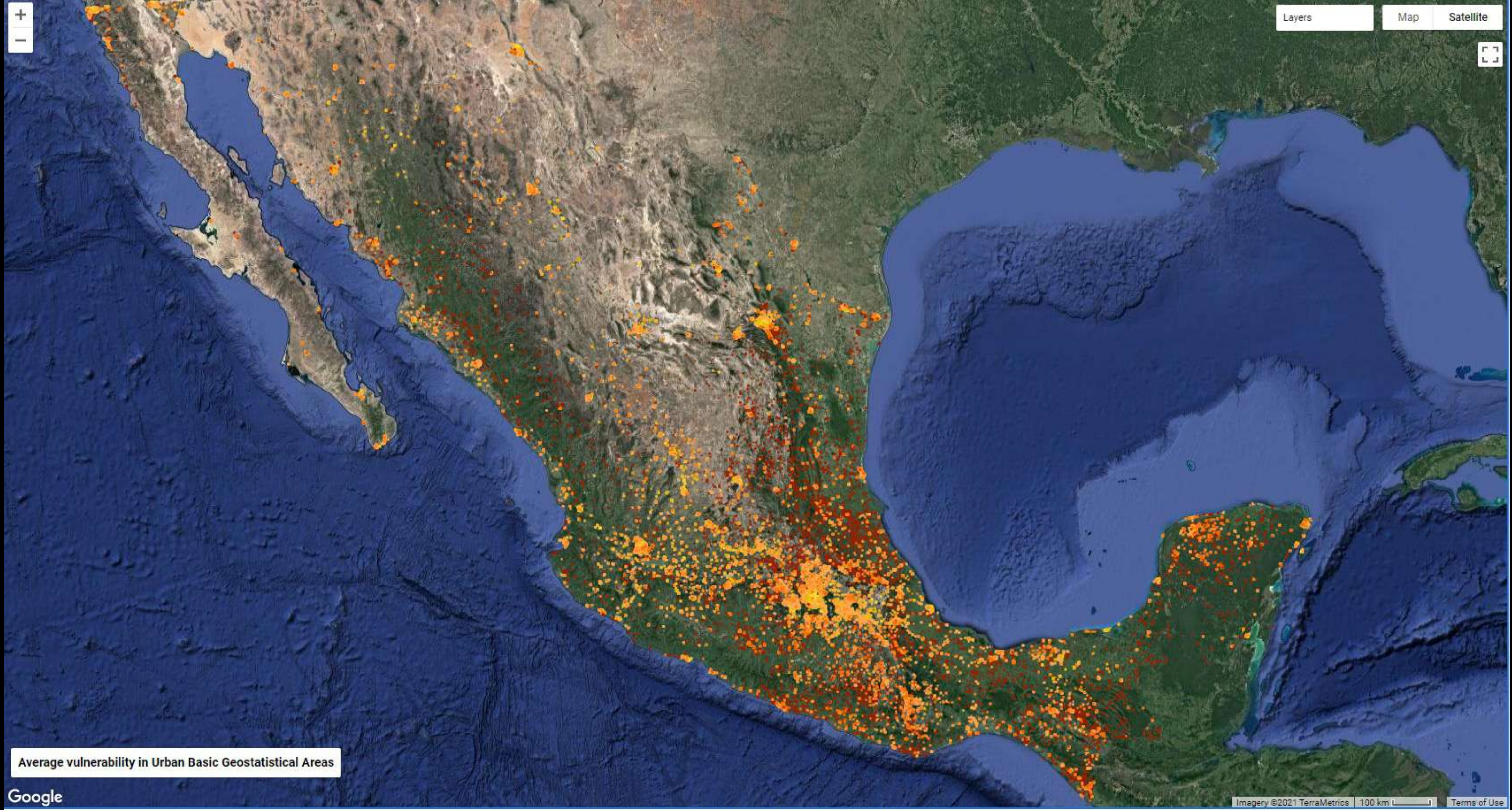
housing
quality
and spaces





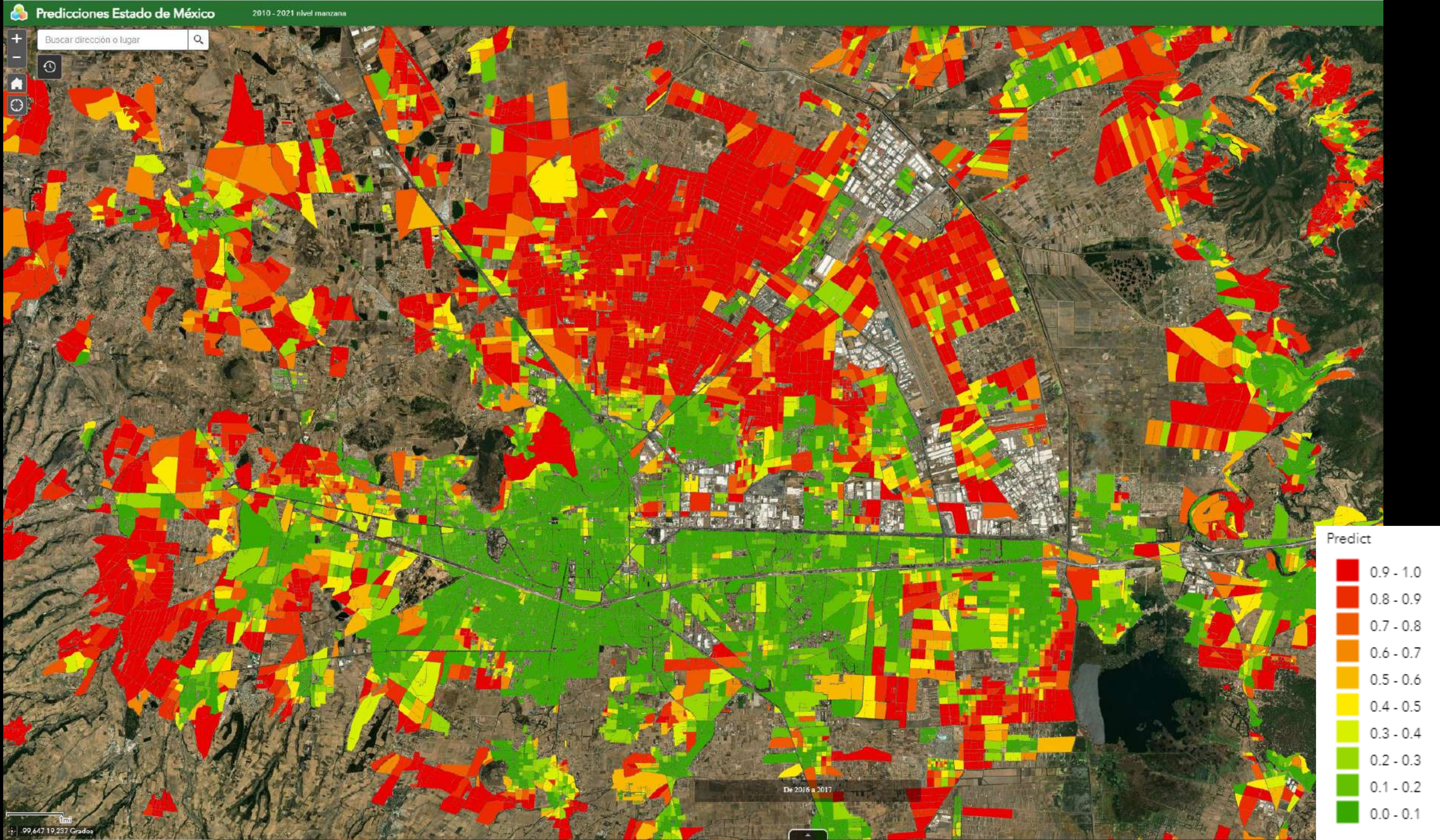


housing
quality
and spaces

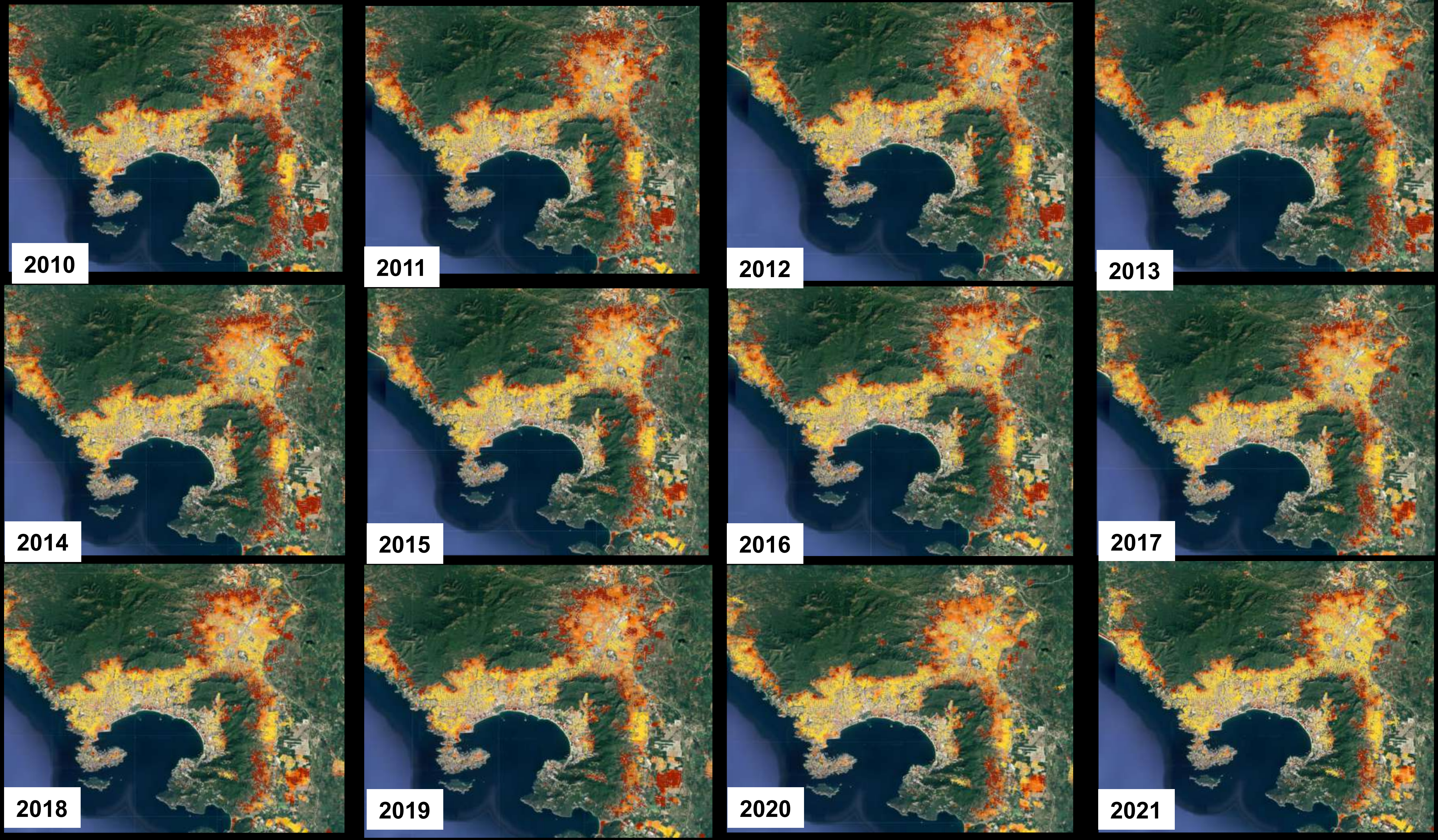


Large-Scale Vulnerable Communities

Vulnerability level

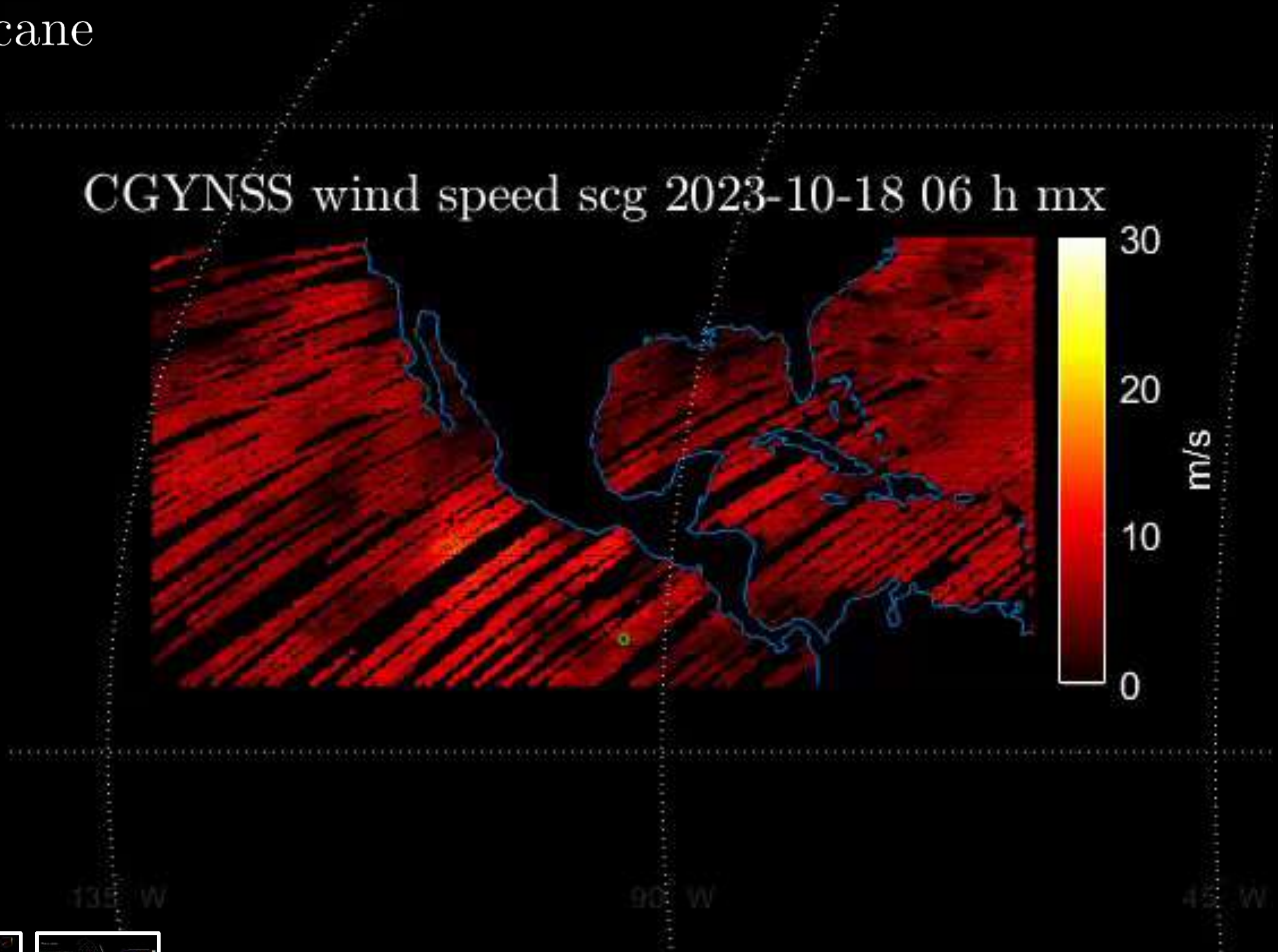

Predicciones Estado de México
2010 - 2021 nivel manzana
Buscar dirección o lugar
Predict
0.9 - 1.0
0.8 - 0.9
0.7 - 0.8
0.6 - 0.7
0.5 - 0.6
0.4 - 0.5
0.3 - 0.4
0.2 - 0.3
0.1 - 0.2
0.0 - 0.1
De 2016 a 2017
1mi
-99.647 19.237 Grados


Inicio
Datos
Servicios
Transparencia
Investigación
Vulnerability level 2010
0.0 to 0.1
0.1 to 0.2
0.2 to 0.3
0.3 to 0.4
0.4 to 0.5
0.5 to 0.6
0.6 to 0.7
0.7 to 0.8
0.8 to 0.9
0.9 to 1.0
Vulnerability level 2020
0.0 to 0.1
0.1 to 0.2
0.2 to 0.3
0.3 to 0.4
0.4 to 0.5
0.5 to 0.6
0.6 to 0.7
0.7 to 0.8
0.8 to 0.9
0.9 to 1.0



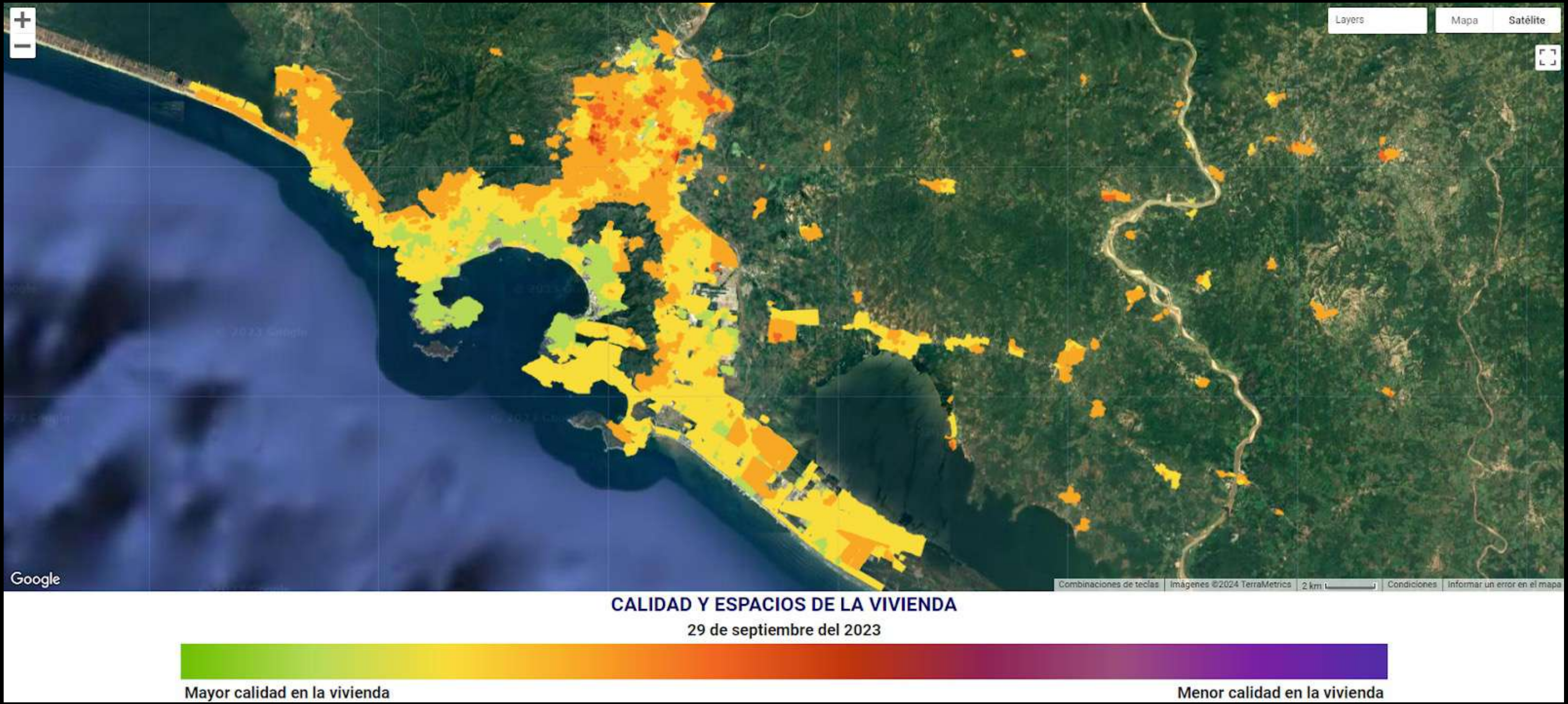
Acapulco, Guerrero

Otis hurricane

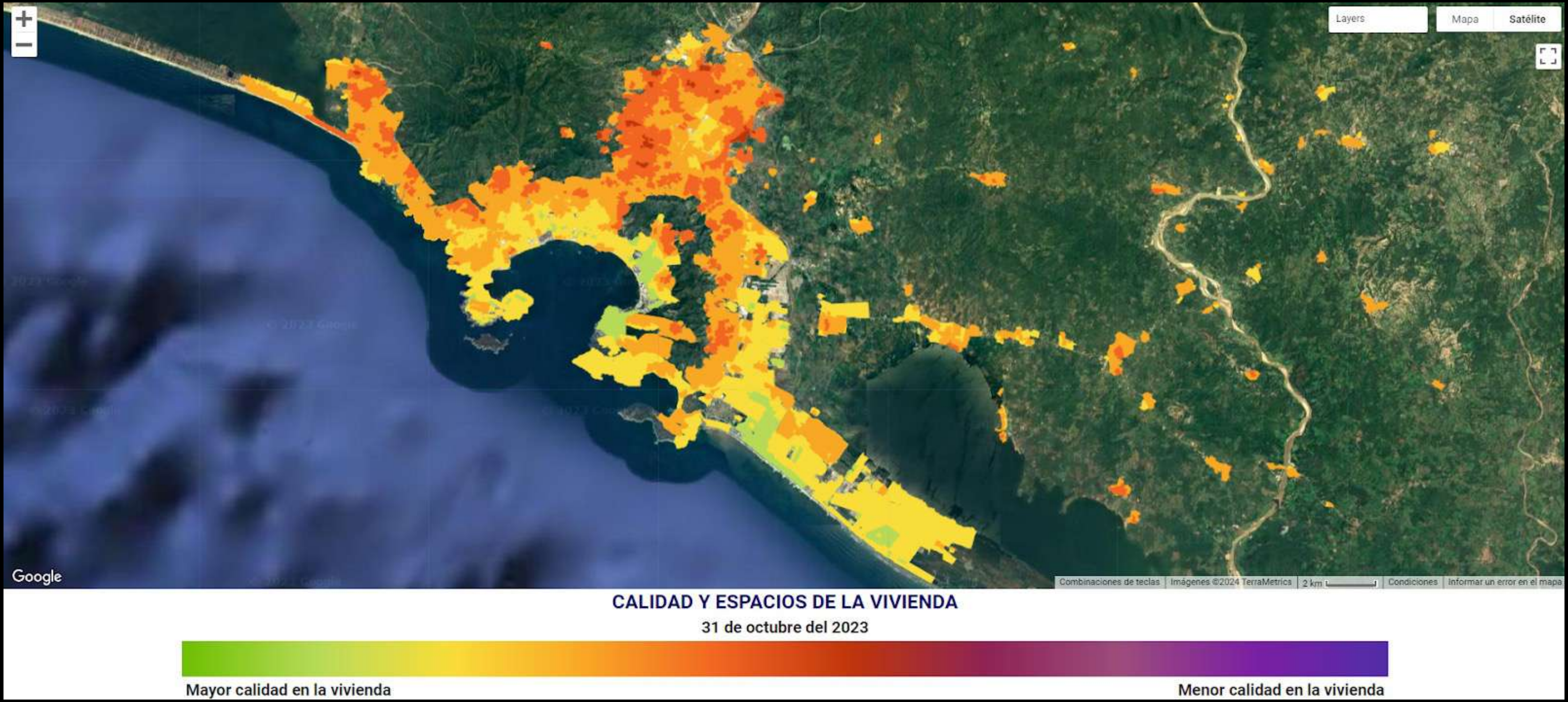


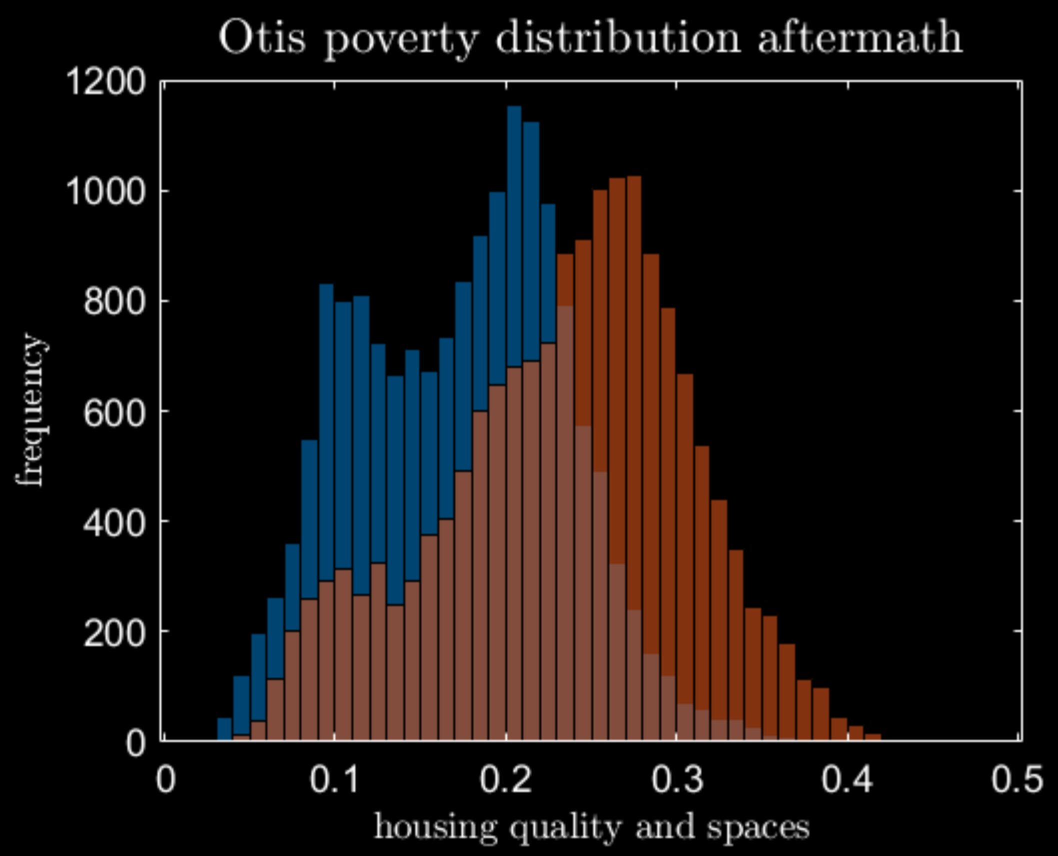

Otis Max Wind Speed
80
60
40
20
0
m/s
Category 5
Category 1
TS
TD
Oct 18
Oct 19
Oct 20
Oct 21
Oct 22
Oct 23
Oct 24
Oct 25
Oct 26
observation
2023

Septiembre 29, 2023

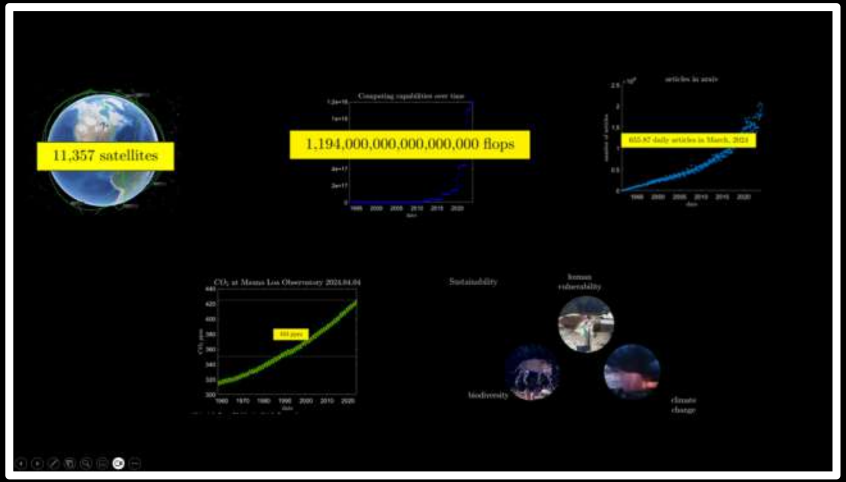


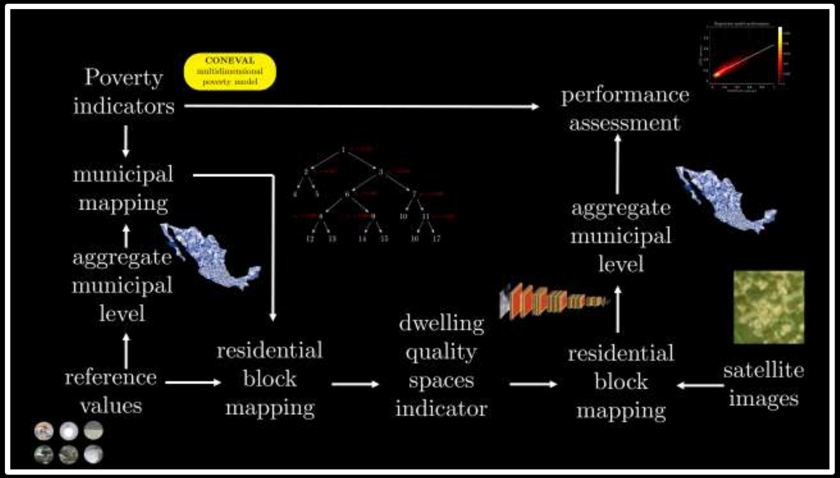
Octubre 31, 2023

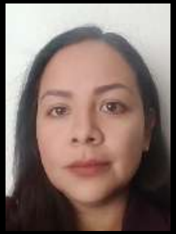




https://datos.nexos.com.mx/desentranando-el-impacto-en-la-pobreza-del-huracan-otis/


11,357 satellites
Computing capabilities over time
1,194,000,000,000,000,000 flops
articles in arxiv
655.87 daily articles in March, 2024
CO_2 at Mauna Loa Observatory 2024.04.04
Sustainability
human vulnerability
biodiversity
climate change


Poverty indicators
CONEVAL multidimensional poverty model
performance assessment
municipal mapping
aggregate municipal level
aggregate municipal level
reference values
residential block mapping
dwelling quality spaces indicator
residential block mapping
satellite images





Alejandra Figueroa • Roberto Manduchi • Ranyart Suarez • Pablo Vera •
Elio-Atenógenes Villaseñor • Marivel Zea-Ortiz

Joaquin Salas
joaquín.salas@gmail.com
jsalasr@ipn.mx

Thank you

IAOS-ISI 2024
MEXICO CONFERENCE

Identification of Agricultural Land Use through Remote Sensing and Machine Learning Techniques

Dr. Abel Coronado

INEGI

International Statistical Institute

IAOS IMPROVING OFFICIAL STATISTICS

Introduction

Mapping Agricultural Land Use

Our research explores the potential of utilizing readily available satellite imagery and advanced machine learning algorithms to map agricultural land use. We will delve into the methodological aspects of our approach, including the use of data sources like Landsat and Sentinel-2, the application of Geographic Object-Based Image Analysis (GEOBIA) for segmentation, and the extraction of features for machine learning classification.

Using Technology to Map Agriculture

Our study shows that our maps of agricultural land use, made using remote sensing, match well with the official agricultural statistics. We achieved more than 90% accuracy with certain machine learning techniques and data sets. Our methods are fast, cost-effective, and provide detailed images (30 meters with Landsat and 10 meters with Sentinel-2). However, there are some challenges like needing accurate data for training, possible errors in complex or fragmented agricultural areas, and the need for clear skies for satellite images. Looking ahead, we plan to combine our methods with INEGI's existing statistical processes. This integration will allow more frequent updates on how agricultural land is used, improve the accuracy of traditional census data, and provide better information on different crop types and their locations across the country.

Methodology

Methodological Overview

In this section, we introduce our methodology for estimating agricultural land use. We begin with the integration of data from the 2019 Agricultural Frontier, which we combine with extensive satellite imagery from Landsat (Phase 1) and Sentinel-2 (Phase 2).

To refine our analysis, we apply Geographic Object-Based Image Analysis (GEOBIA), which allows us to go beyond the limits of traditional pixel-based classification methods.

We then utilize advanced machine learning algorithms, including Random Forest and MLP (Multilayer Perceptron), to classify the land use with precision. The effectiveness of our approach is confirmed by the high accuracy rates achieved in our 2019 evaluations.

Data Sources

Data Sources - Phase 1: Landsat (30m x 30m)

2019 National Geomedian Mosaic from Landsat.
The dimensions of the mosaic are 110,000 pixels by 70,000 pixels.

•**Agricultural Frontier 2019 Data**: Utilized as ground truth, providing a base for comparison and analysis.

•**Satellite Imagery Management**: Managed with the Open Data Cube, handling data from Landsat, representing over three decades of agricultural imaging.

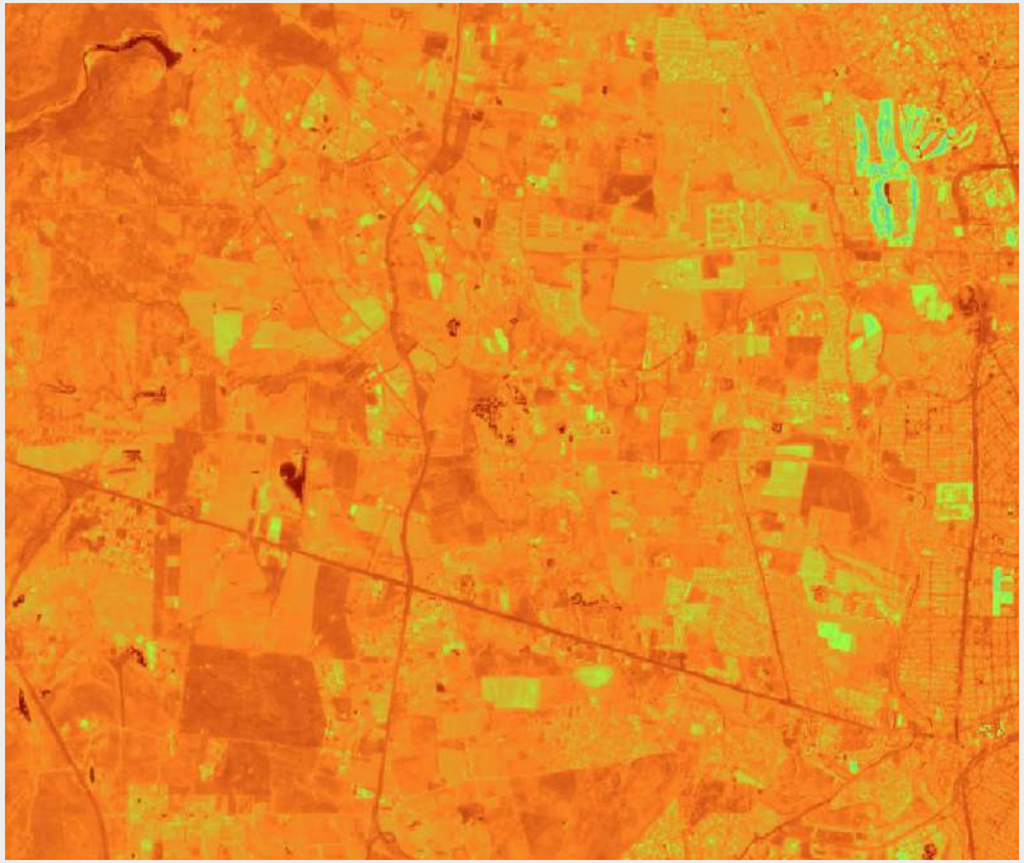
•**Data Processing**: Utilization of the Open Data Cube for the generation of Geomedian mosaics, Median Absolute Deviation (MAD) layers, and aggregated Normalized Difference Vegetation Index (NDVI) over time using Landsat imagery.

•**Additional Data**: Integration of DEM and precipitation data to enrich the models.

•**Spectral Indices**: Computation of 20 spectral indices from Landsat imagery to improve land cover analysis.

IGOS-ISI 2024
MEXICO CONFERENCE

Data Sources - Phase 2: Sentinel-2 (10m x 10m)



Infrared band from Sentinel-2 used as a base for texture filter calculations.

•**Agricultural Frontier 2019 Data**: Continues as the foundational data layer for accurate agricultural mapping.

•**Data Processing**

Geomedian Composites: Utilizes Google Earth Engine to create Geomedian composites, effectively summarizing satellite data by using the geometric median which reduces noise and enhances the true signal in Sentinel imagery.

Spectral Indices: 20 different indices are computed to provide diverse spectral features.

Texture Filters: Implements a suite of 48 Leung-Malik texture filters to characterize the surface texture of the landscape.

GEOBIA (Geographic Object-Based Image Analysis)

GEOBIA in Phase 1: Landsat

9,181,382 Clumps nationwide

- **Introduction to GEOBIA:** Introduced as a method to address the limitations of pixel-based classification, focusing on object segmentation to enhance landscape analysis.
- **Segmentation Process:** Utilized Shepherd segmentation method within the RSGISLib library, setting the number of clusters to 60.
- **Segmentation Results:** Achieved 9,181,382 distinct clumps, clearly delineating different land uses. This showcases the effectiveness of GEOBIA in capturing and analyzing the complexity of varied landscapes.

GEOBIA in Phase 2: Sentinel-2

261,547,763 Clumps nationwide

- **Introduction to GEOBIA:** Continued application of GEOBIA with the higher resolution of Sentinel-2 imagery, utilizing the same segmentation methodology to enhance detail and accuracy.

- **Segmentation Process:** Applied Shepherd segmentation with the RSGISLib library, with the cluster number parameter consistently set at 60, to handle the increased complexity and detail provided by Sentinel-2 data.

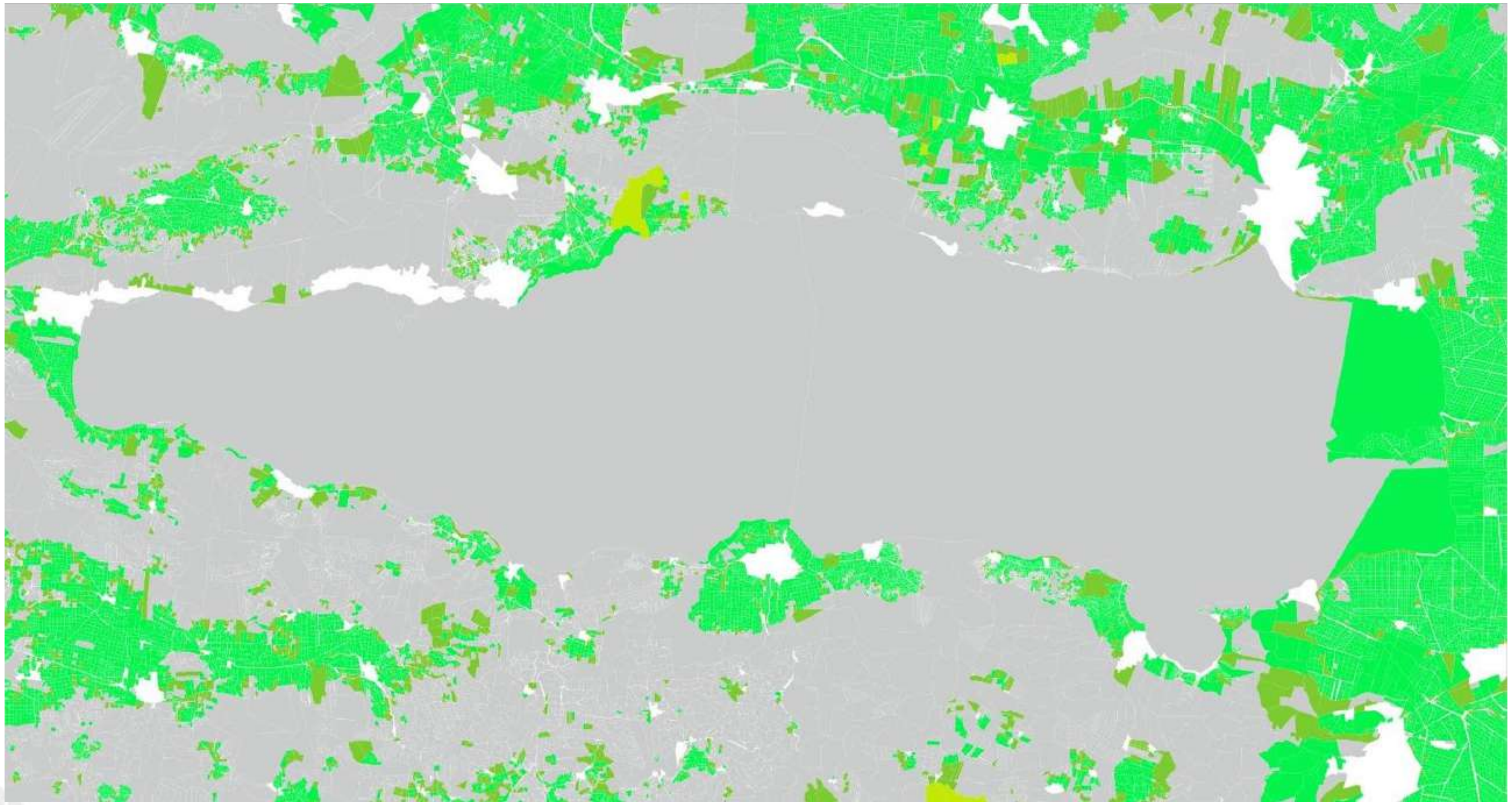
- **Segmentation Results:** Achieved 261 million segments.

Labeling and Feature Extraction

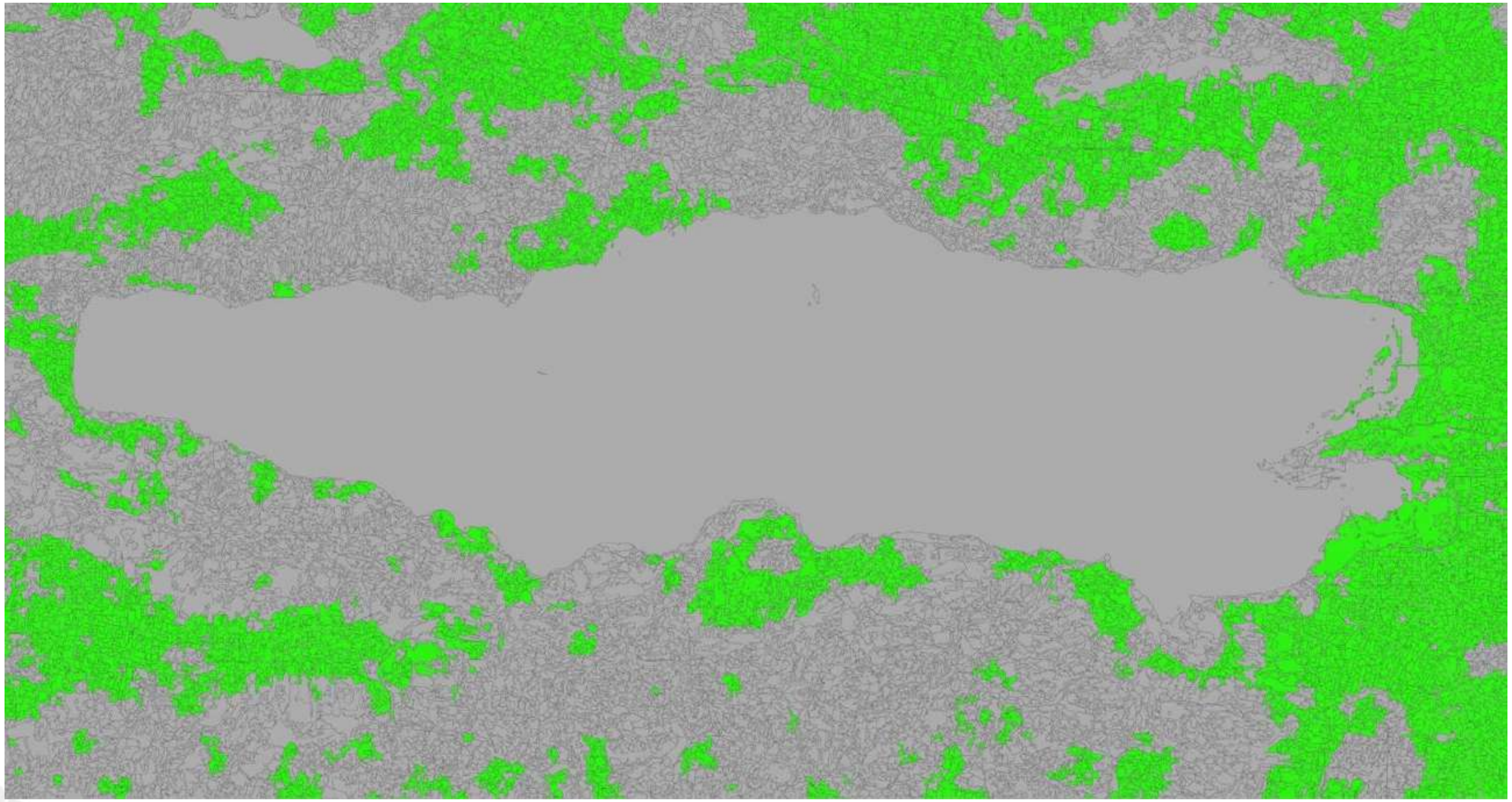
Labeling Process

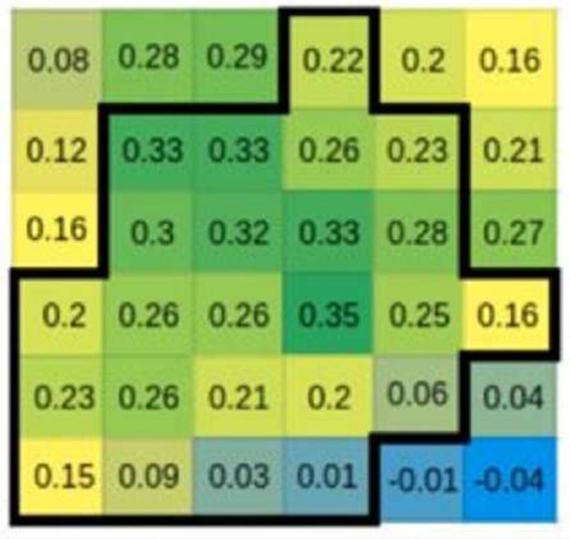
Labels assigned based on the proportion of agricultural vs. non-agricultural pixels within each segmented object.

Clump Labeling (Landsat & Sentinel-2)



Clump Labeling (Landsat & Sentinel-2)





Spectral Index Values within a Clump: This image depicts a single clump, showcasing the distribution of spectral index values across its extent. Each cell within the clump grid represents a different value of the spectral index.

Feature Extraction - Phase 1: Landsat (30m x 30m)

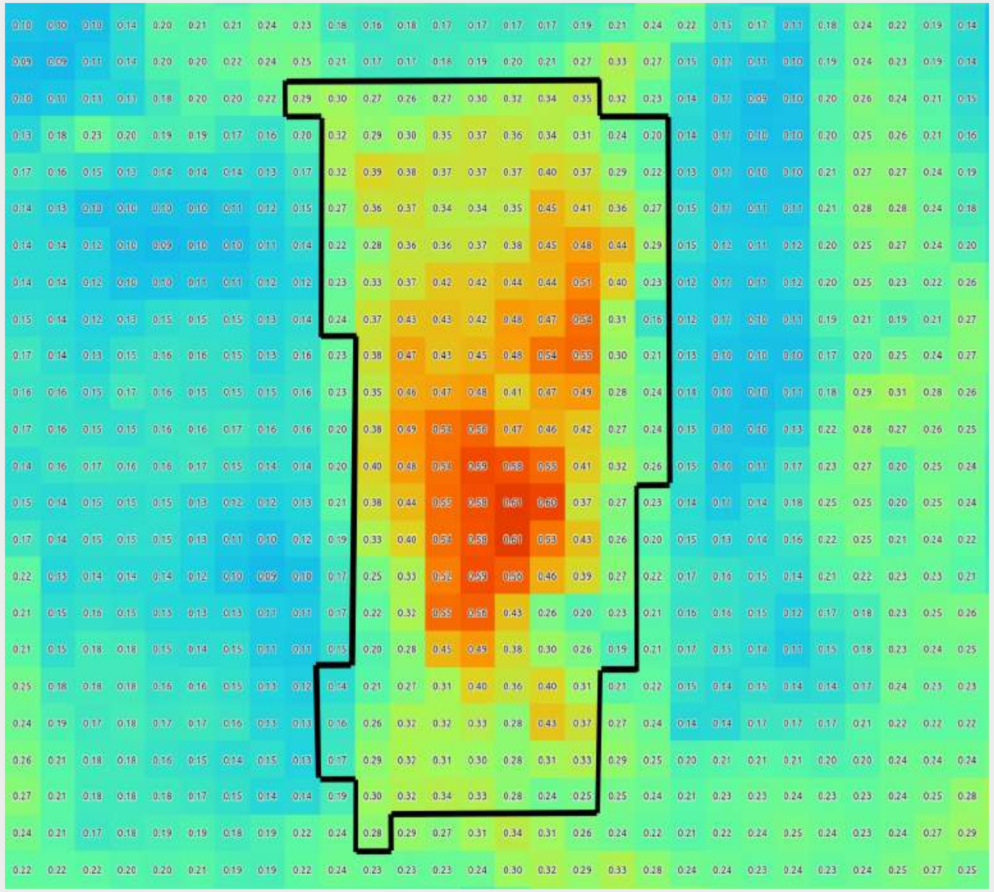
Data Layers Analyzed: Statistical summaries extracted from 37 data layers for each object, encompassing Geomedian components (Blue, Green, Red, Near-Infrared, Shortwave Infrared 1 and 2), and various indices like NDVI (Max, Medium, Median, Min, Standard Deviation), along with terrain features (Altitude, Slope) and precipitation totals.

Feature Extraction (Landsat)

Clump	Class label	Geomedian-Blue Min	Geomedian-Blue Max	Geomedian-Blue Median	...	VARI Sum	VARI Std Dev
1	2	256	3,235	1,570	...	0.26	0.07
...	...	...	...	...	...	...	...
9,181,382	1	129	2,500	1,120	...	0.39	0.19

This table presents statistical summaries for 9,181,382 clumps across 185 variables, featuring key metrics such as minimum, maximum, median, and sum calculated for each band.

Feature Extraction: Sentinel-2



Spectral Index Values within a Clump: This image depicts a single clump, showcasing the distribution of spectral index values across its extent. Each cell within the clump grid represents a different value of the spectral index.

- ALL LAYERS
  - minimum
  - maximum
  - media
  - sum
  - standard deviation
- TEXTURE FILTERS
  - 10th - 90th percentile.

Feature Extraction (Sentinel-2)

Clump	Class label	Geomedian-Blue Min	Geomedian-Blue Max	Geomedian-Blue Median	...	Texture Filter 80th Percentile	Texture Filter 90th Percentile
1	2	256	3,235	1,570	...	0.26	0.07
...	...	...	...	...	...	...	...
261,547,763	1	129	2,500	1,120	...	0.39	0.19

This table showcases comprehensive statistical summaries for a total of 261,547,763 segments. It details the distribution of values across 832 variables, including the minimum, maximum, median, sum of all the bands, and specific percentiles from texture bands, highlighting the extensive data analysis performed.

Results

Agricultural Land Cover Classification - Phase 1: Landsat

Machine Learning Algorithms:

- **Overview:** Applied AutoSkLearn 2.0 tailored specifically for Landsat imagery to classify agricultural land cover.
- **Performance:** Achieved an optimal accuracy of 83% in national validation, showcasing the effectiveness of the model in agricultural settings.

Agricultural Land Cover Classification - Phase 1: Landsat

AMCA 2016				AMCA 2016	1st Iteration Landsat	2nd Iteration Landsat	3rd Iteration Landsat	4th Iteration Landsat
COD_ACT	DESCRIPTION	Hectares		Hectares	Hectares	Hectares	Hectares	Hectares
A	Fully agricultural	20,025,775	AGRICULTURAL	31,376,931	55,797,681	43,915,831	40,326,623	38,806,173
C	At least 30% agriculture	10,978,330						
M	Mixed	372,826			A,C,M,	A	A	A
F	Was agricultural	500,709	NON AGRICULTURAL	160,636,548	139,432,164	151,314,013	154,903,222	156,423,671
N	Non Agricultural activity	11,766,282						
U	Urban	434,264						
V	Non Agricultural activity	147,123,210						
W	Water Bodies	809,802						
B	Road	1,649						
I	floodable areas							
	Accuracy				79%	80%	82%	83%

IOS-ISI 2024 MEXICO CONFERENCE

Agricultural Land Cover Classification - Phase 2: Sentinel-2

Machine Learning Algorithms:

- **Optimization:** Employed the Random Forest algorithm with a focus on regional training sets derived from Sentinel-2 imagery.

- **Enhanced Accuracy:** Increased overall accuracy to 91%, demonstrating substantial improvement in the classification of agricultural land cover.

Agricultural Land Cover Classification - Phase 2: Sentinel-2

AMCA 2016				AMCA 2016	1st Iteration Landsat	2nd Iteration Landsat	3rd Iteration Landsat	4th Iteration Landsat	1st Iteration Sentinel (2019)
COD_ACT	DESCRIPTION	Hectares		Hectares	Hectares	Hectares	Hectares	Hectares	Hectares
A	Fully agricultural	20,025,775	AGRICULTURAL	31,376,931	55,797,681	43,915,831	40,326,623	38,806,173	32,457,571
C	At least 30% agriculture	10,978,330							
M	Mixed	372,826			A,C,M,	A	A	A	A
F	Was agricultural	500,709	NON AGRICULTURAL	160,636,548	139,432,164	151,314,013	154,903,222	156,423,671	162,772,274
N	Non Agricultural activity	11,766,282							
U	Urban	434,264							
V	Non Agricultural activity	147,123,210							
W	Water Bodies	809,802							
B	Road	1,649							
I	floodable areas								
	Accuracy				79%	80%	82%	83%	91%

Applications in Production Processes

MEXICO CONFERENCE

Refinement and Comparison with Sentinel-2 Data

Internal Validation: Machine learning outputs have been used internally to validate and cross-reference data from the National Agricultural Survey (ENA) 2019, ensuring the accuracy of reported figures.

International Relevance: While primarily used internally, these results also support presentations in international forums, showcasing advancements in agricultural data estimation.

Future Prospects: Ongoing improvements promise to integrate these methods more deeply into official statistical processes, potentially enhancing the reliability of agricultural data nationally.

.

Conclusion

Achievements

- **Technological Integration:** Successfully implemented machine learning techniques to redefine agricultural boundary estimation using satellite imagery from Landsat and Sentinel platforms.

- **Accuracy and Validation:** Enhanced accuracy in land classification has been achieved, with preliminary assessments helping align satellite data with statistical findings for improved decision-making.

Collaborative Efforts

- **Refinement of Methods:** Ongoing improvements are aimed at optimizing machine learning algorithms to provide even more precise and up-to-date data for each agricultural cycle.

- **Continuous Engagement:** We remain committed to working closely with all stakeholders to refine and enhance the utility of machine learning in agricultural land classification, ensuring that these advancements lead to practical, actionable insights for agricultural planning.

Thank you

MEXICO CONFERENCE

Estimating Mexican municipal-level economic activity indicators using nighttime lights

Francisco Corona
José López
Ranyart Rodrigo Suárez
Elio Villaseñor

International Statistical Institute

IMPROVING OFFICIAL STATISTICS

Introduction

IOS-ISI 2024
MEXICO CONFERENCE

Mexican context of IAEM

- In Mexico, the availability of economic information with high levels of geographic and temporal disaggregation is scarce
- Its availability is of importance for policy design at the municipal, state and federal levels
- In the Economic Censuses (EC) published INEGI every five years, one can obtain relevant information like census gross added value (CGAV), labor statistics and company assets, which can be obtained at municipal level
  - This information is of relevance and meets the geographic needs to carry economic analysis at municipal level, its temporal frequency impedes to analyze economic municipalities evolution across time
- In previous works Corona and López-Pérez (2019), López-Pérez and Corona (2020) have presented methodologies to obtain Mexican municipal-level economic activity indicators (IAEM in Spanish) based on the information of EC

Main motivation

There is practically no alternative information that allows analyzing the economic evolution of the municipalities over time. However, there is empirical evidence that corroborates a significant statistical association between economic activity and the intensity of nighttime luminosity.

Related work

Economic activity and nighttime lights

- In the works of Donaldson and Storeygard (2016) and Li et al. (2016), it is documented various applications in which it is found significant relationships between variables extracted from satellite databases and time series of economic activity.
- Galimberti (2020) makes an application to the economy to improve forecasts of the Gross Domestic Product (GDP). In such work, they conclude that night light observations provide predictive information on GDP growth rates as long as economic activity shows serial autocorrelation.
- Hu and Yao (2022), analyze the assumption of the relationship between luminosity and GDP growth for different countries, concluding that this may even be nonlinear depending on the type of economy of the country.
- In Guerrero and Mendoza (2019), which refer methodologically to Henderson et al. (2012), the concept of "*true economic activity*" is mentioned, whose essential idea is that the economic activity, frequently measured through GDP by national statistics agencies, can be thought as an approximation of true GDP or economic activity.

Economic activity and nighttime lights

- Guerrero and Mendoza (2019) generate estimates of the true economic activity of China, Chile and Mexico, as a convex function of the official economic activity and the intensity of the lights. It is concluded that, in the case of Mexico, the GDP was underestimated by an annual average of 0.45% for the analyzed period (1992-2008).
- Other works with applications for the case of Mexico are
  - Rangel-González and Llamosas-Rosas (2019) who estimate the size of the informal economy for the years 2015 and 2016, specifically the authors take into consideration the fact that the correlation between night luminosity and economic activity differs by sector.
  - Llamosas-Rosas et al. (2021) measure the economic activity in the main beaches of Mexico from 1993 to 2017 and apply the elasticity obtained in the period 1993–2013, to extend the results until 2017.

  - Llamosas-Rosas et al. (2021) measure the economic activity in the main beaches of Mexico from 1993 to 2017 and apply the elasticity obtained in the period 1993–2013, to extend the results until 2017.

Data sources

Traditional sources

Economic Censuses

INEGI's EC information

1994 1999 2004 2009 2014 2019

Microdata of INEGI's EC

According to the North American Industrial Classification System (NAICS), the IAEM consider only the information of three groups of subsectors: S31–33 Manufacturing industries (99% coverage with respect to the number of NAICS classes), S43–46 & S72 Commerce and Temporary accommodation and food preparation services and beverages (99%) and S54-S81 non-financial services (70%).

Mexico's System of National Accounts

The GDP by federal entity is used to reconcile the municipal activity figures obtained based on the EC with the state figures.

Non-traditional sources

Satellite images

OLS

VIIRS

The satellite images used are obtained from Google Earth Engine, a repository that combines information from different satellite sources for the entire planet. Nighttime lights images used which consist of grids of observations of light intensities captured for the entire globe at night. To obtain an appropriate measure, we add the luminous intensity within the polygon of each municipality using the INEGI geostatistical framework.

- **Operational Linescan System (OLS)** available from 1992 to 2013 with annual frequency, the Stable Lights band is used, which contains lights from cities, towns and other places with persistent lighting. Ephemeral events such as fires are ruled out. The values are in units called digital numbers (DN) in the range 0–63.
- **Visible Infrared Imaging Radiometer Suite (VIIRS)** from April 2012 to date with annual frequency. Unlike the DN, the records of this source correspond to the enhanced night light radiation with high radiometric resolution. The median of the monthly information is used to generate an annual series.

Methodology

Methods

The methodologies used in this work can be found in the works of López-Pérez and Corona (2020) in relation to estimating IAEM based solely on official information, while Guerrero and Mendoza (2019) show how "official" estimates can be refined using nighttime light information. In this section we summarize the most important of each of these methodologies, but the reader is invited to review the original works for more details.

Municipal-level economic activity indicators based on the economic censuses: 1993–2018

The methodology to estimate the IAEM based on the information provided by the EC is explained in López-Pérez and Corona (2020), which in turn refer to Corona and López-Pérez (2019). To summarize what is most important, the most important steps are described below:

1. Select from the EC, those types of economic activity in which the observation units record information regarding the place where they perform (see Sect. 2.1). For those observation units with negative CGAV, this value is reestimated using the adjustment obtained through a logarithmic regression, which considers CGAVas the dependent variable only for units with positive values in the sample and, as regressors, man-hours and total fixed assets.
2. For each entity and census year, build CGAV ratios for each group of economic activity sectors considered, in such a way that the sum of the municipalities is equivalent to 1.

Municipal-level economic activity indicators based on the economic censuses: 1993–2018 (cont.)

3. For economic activity sector group, estimate the intercensal years using linear interpolators. The founding date of the municipalities is taking into consideration, especially for new created municipalities.For each entity and census year, build CGAV ratios for each group of economic activity sectors considered, in such a way that the sum of the municipalities is equivalent to 1.
4. Apply the Combination Rule (CR) of Guerrero and Peña (2003) where the weighting matrix is obtained following Corona and López-Pérez (2019) to reconcile figures, in such a way that the municipal-EAI of the previous step add to their respective data granted by the Mexico's NSA. Note that this reconciliation phase allows expressing the municipal-EAI to base year 2013
5. Obtain the suggested aggregations: manufacturing industries, commerce, temporary accommodation services and food and beverage preparation, and non-financial services.
6. Add to obtain the state total (which only considers these three groups of subsectors).

Estimation of municipal economic activity using night luminosity

Once described the procedure to estimate $\lambda*$ (see Corona (2022) for the formulae of $\lambda*$), the estimation algorithm of the IAEM using official information and satellite images of night luminosity is carried out using the following algorithm for each municipality:

1. The methodology of López-Pérez and Corona (2020) is used to generate IAEM based on the EC for 1993 to 2018.
2. The annual variation of the IAEM from 1994 to 2013 is estimated with nighttime luminosity information using the OLS information source.
3. The annual variation of the IAEM from 2012 to 2018 is estimated with nighttime luminosity information using the VIIRS information source.
4. Repeat steps 1–3 to obtain $\lambda*$ considering a partition of $R = 1000$ values.
5. Once selected $\lambda*$, the level of IAEM with luminosity-VIIRS is recovered assuming as initial value the 2011 value of the IEAM estimated in point 1 above. We use cumulative sums to recover the levels.
6. We splice the series obtained in points 3 and 5 by attributing the annual variation of IAEM with luminosity-OLS along 1994–2011 to the levels of IAEM with luminosity-VIIRS along 2012–2018. We assume as initial value the 1993 value of the IEAM estimated in point 1 above

Results

IOS-ISI 2024
MEXICO CONFERENCE

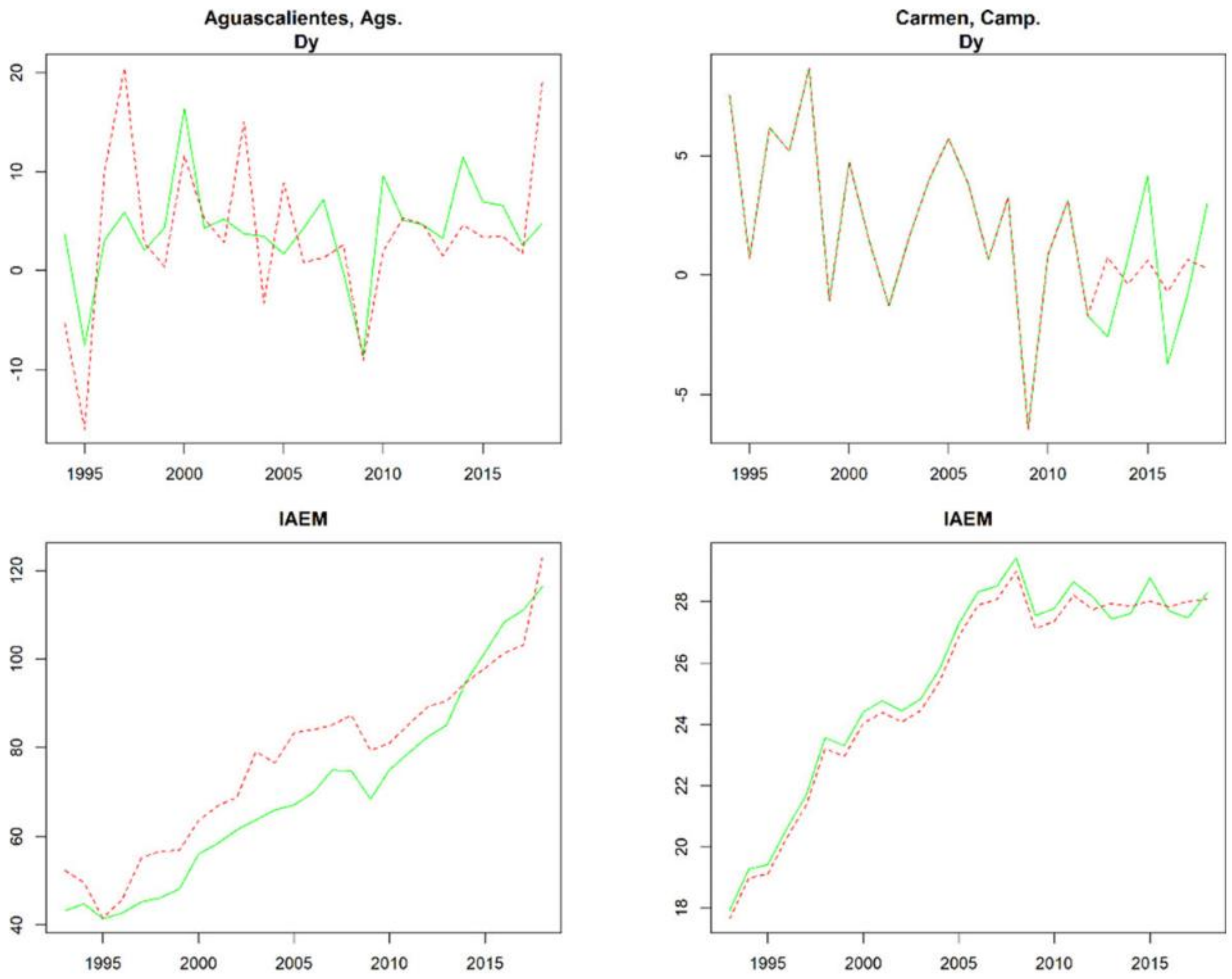
Considerations

- IAEM estimates were obtained for all the municipalities of Mexico, for the period 1993–2018. As an illustration, this section shows the estimates for a conventional sample of 12 municipalities.
- Such municipalities were chosen because they constitute a diverse selection of the most representative states of the national economy and heterogeneous; thus, we consider that these municipalities generate an interesting picture of Mexico. For instance,
    - Aguascalientes, capital of the Aguascalientes state, is a emergent city;
    - Tijuana is the biggest city in the north border of Mexico;
    - Carmen is an oil city in the southeast;
    - Benito Juárez and Miguel Hidalgo are municipalities of the capital of the country, which are very representative of the services sector;
    - León is a representative city of the “Bajío” zone;
    - Guadalajara-Zapopan and San Pedro Garza García-Monterrey are part of two representative metropolitans zone of Mexico, known as Guadalajara and Monterrey, respectively;
    - Finally, Hermosillo and Centro are two antipodal municipalities, while Hermosillo is the capital of Sonora, in the north of Mexico, Centro is the capital of Tabasco, in the south of the country.

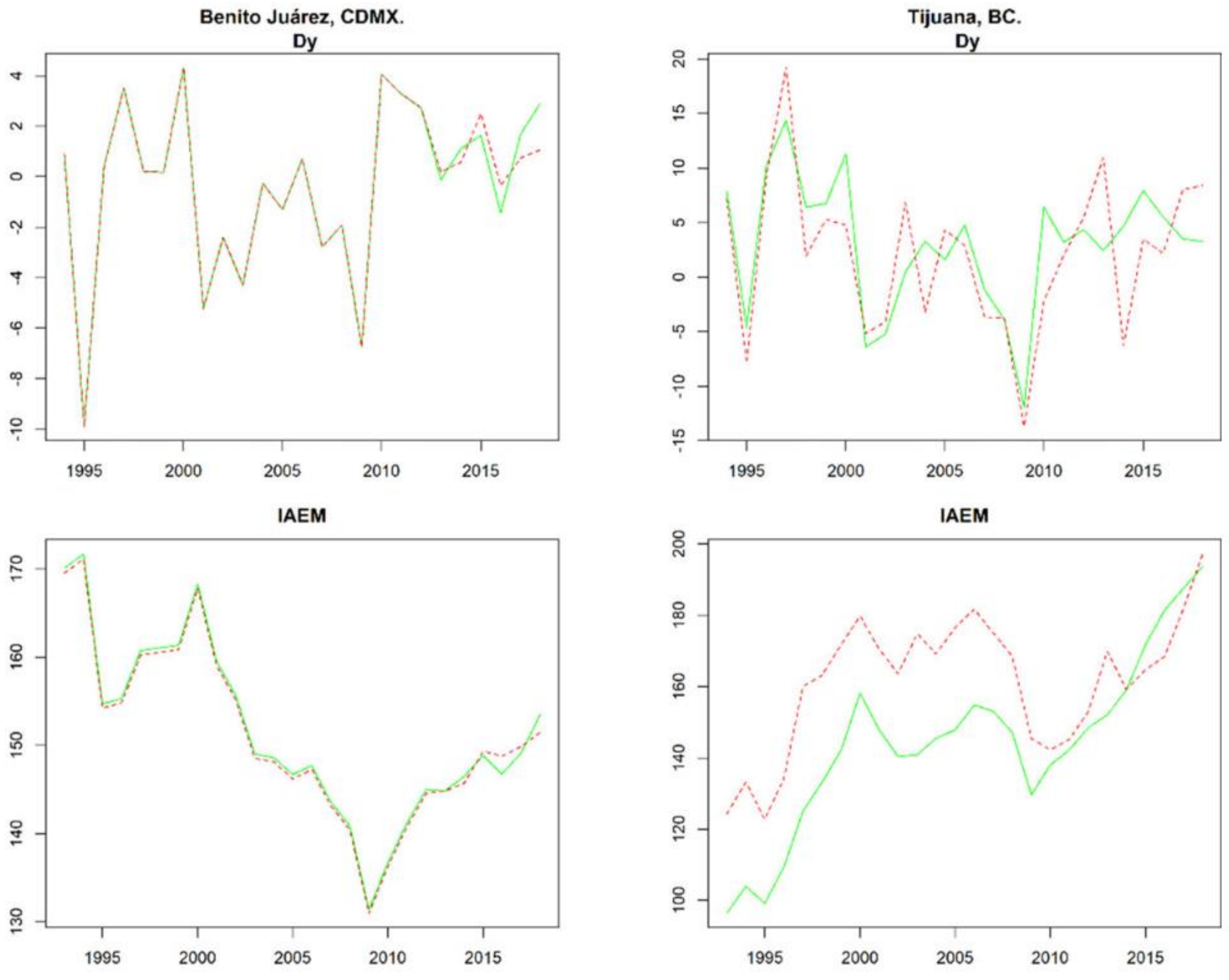
Considerations (cont.)

- Each of the graphs shows, in the upper panel, the annual percentage variations for both the IAEM adjusted for luminosity Dy , (red dotted line) and the preliminary series extracted using only the information from the EC, that is say Dz (continuous green line); while the lower panel shows the levels, that is, the IAEM for both cases.
- Considerations were made for the treatment of the information from the OLS series, which is truncated at both ends, so it is common to find municipalities with Var(Dx) = 0, so the model described by the expression (3). In these cases, only the preliminary IAEM series is considered, that is, no adjustment is made for luminosity.
- The evaluation of the results is carried out in two ways.
  - First through a graphic inspection for the municipalities determined as important mentioned above (reviewed in this presentation).
  - Afterward, the results are evaluated as a whole by applying a statistical criterion between the series with and without considering luminosity (depicted in F. Corona et al. (2023))

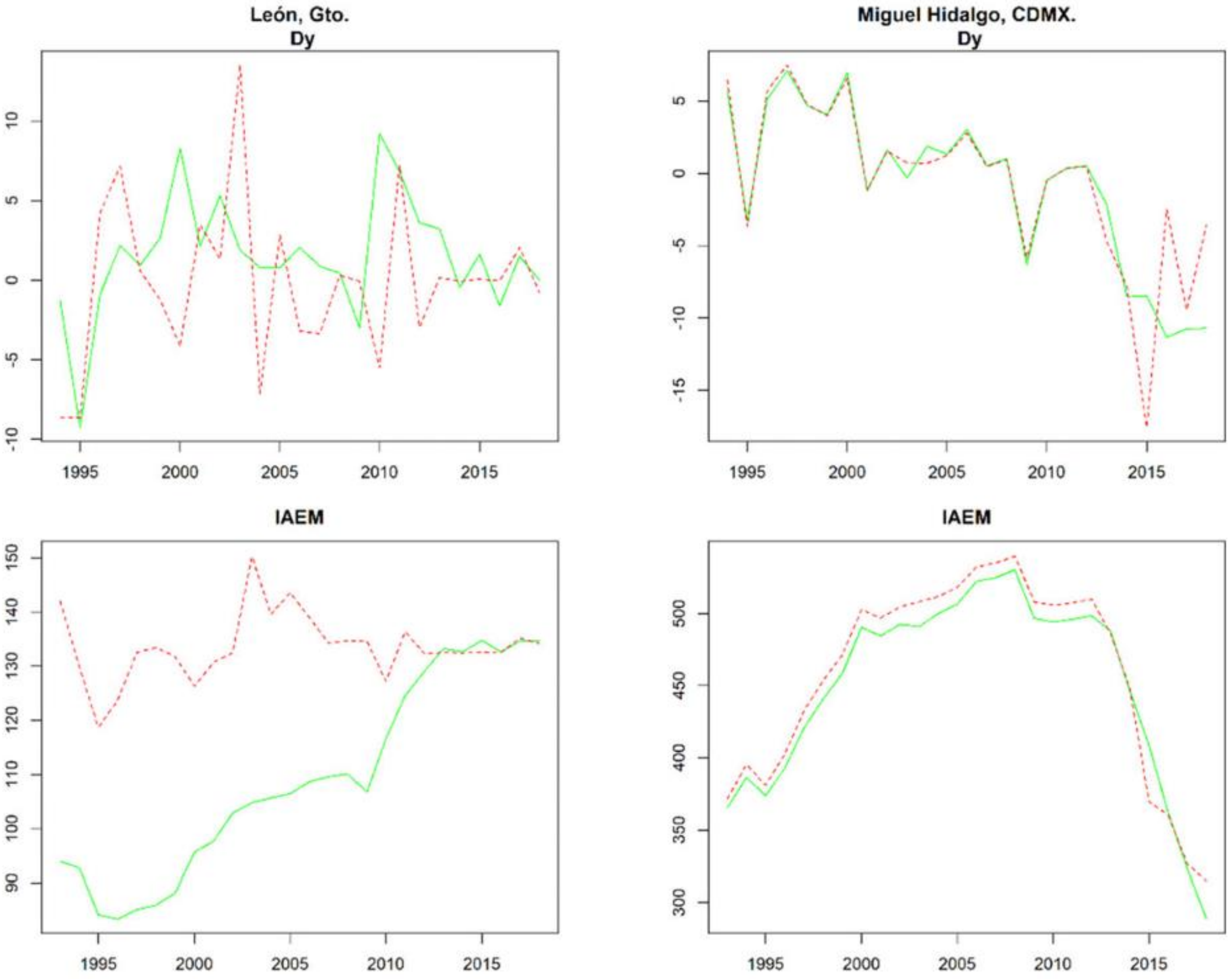
Graphic evaluation of the results



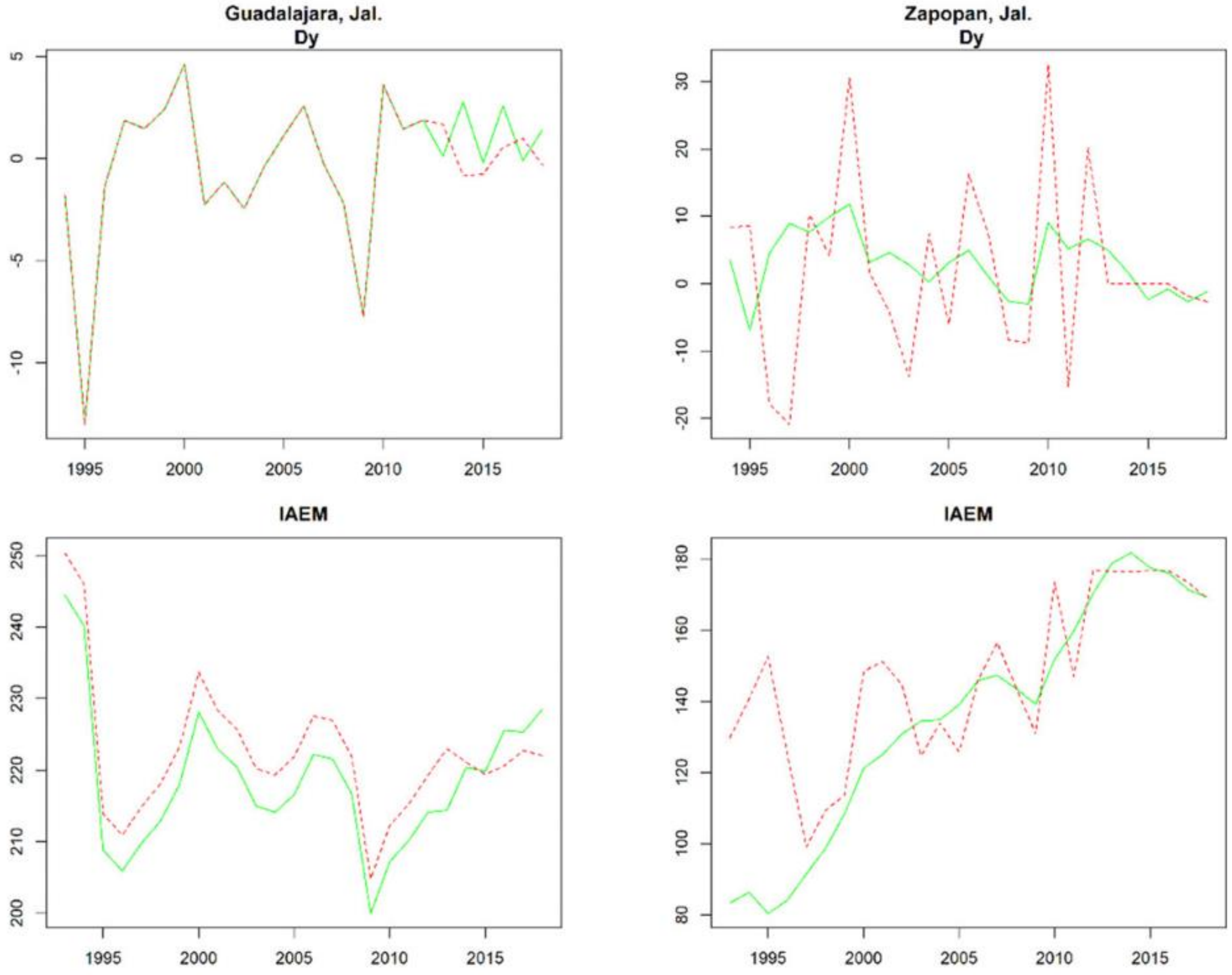
Graphic evaluation of the results (cont.)



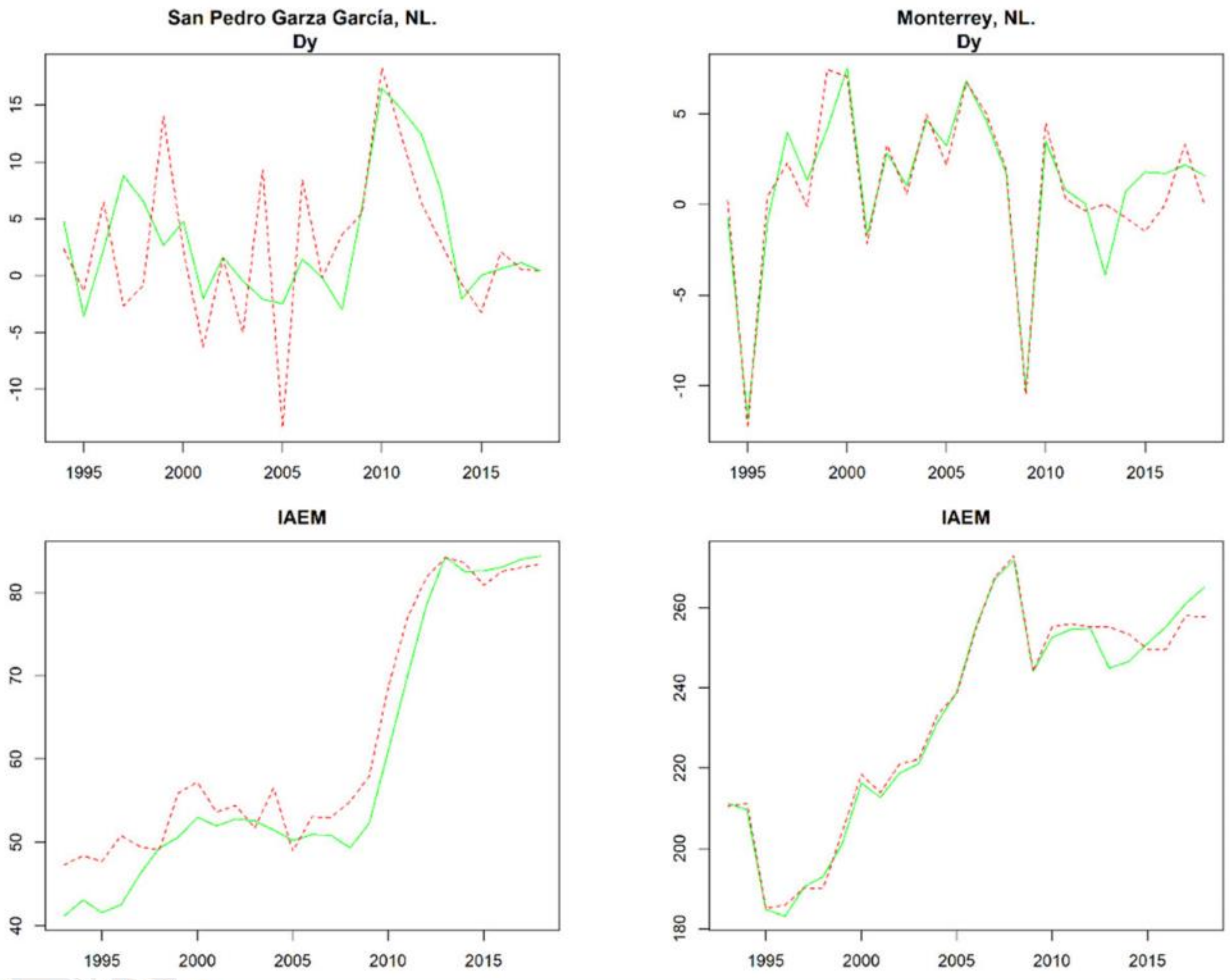
Graphic evaluation of the results (cont.)



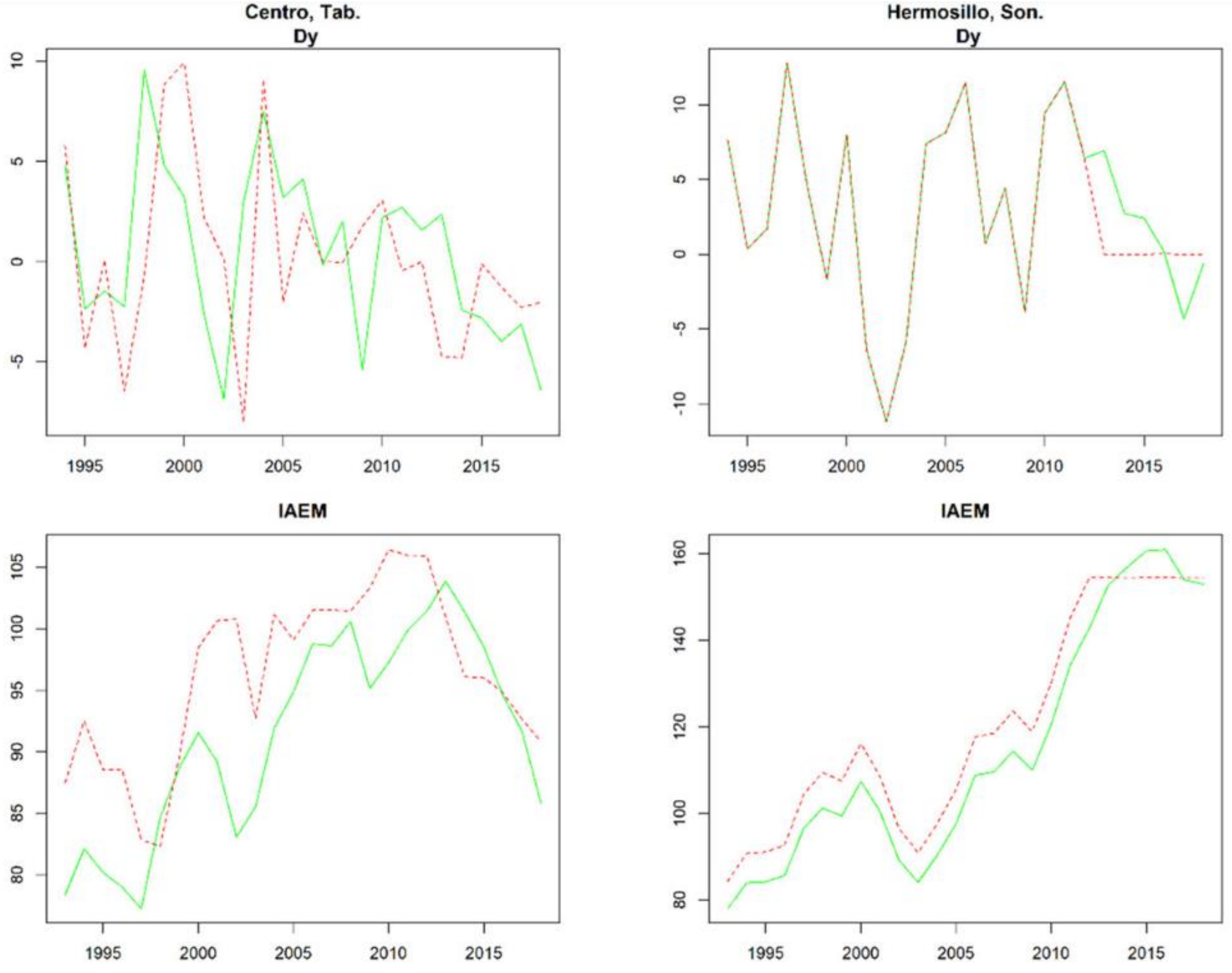
Graphic evaluation of the results (cont.)



Graphic evaluation of the results (cont.)



Graphic evaluation of the results (cont.)



Discussion

- We can observe that the series of annual variations of the series with luminosity register a dynamic similar to that of the preliminary series (continuous green line), for example, in municipalities such as Aguascalientes, Ags., Tijuana, BC. Monterey, NL. or, Center, Tabasco.
- In other cases, such as Carmen, Camp., Benito Juárez, CDMX, Guadalajara, Jal. and Hermosillo, Son., the variations of both series in the period from 1994 to 2012 are identical, due to the consideration that is taken when Var(Dx) = 0 described above.
- In all the previous cases, the series in levels also reflect similar movements between the series with and without luminosity. However, for example, the case of León, Gto. requires particular attention, since, although the percentage variations maintain similar
- The case of León, Gto. requires particular attention, since, although the percentage variations maintain similar dynamics and magnitude, the series in IAEM levels that include luminosity observes a considerable increase with respect to the IAEM series that only considers EC.

Conclusion

Conclusions and further lines

Using traditional information provided by EC and non-traditional information, such as that provided by satellite images of nighttime brightness, IAEM were generated which meet two essential requirements:

1. Geographical representativeness, basically by generating in a first stage, the IAEM as proposed by López-Pérez and Corona (2020)
2. Refine these estimates in a temporal sense by exploiting the relationship that exists between economic activity and nighttime luminosity, the above specifically for non-census years.

In this way, the IAEM generated in this work are a function of both the EC and the intensity of the night light in the municipalities. This is important for modelers, since it provides the most important elements to generate IAEM under this approach;

- For analysts, since they can unravel and understand the results and promote possible public policies and finally;
- In the academic field, since it represents an interesting application that combines traditional techniques of statistics and econometrics as in the field of Big Data, since large amounts of information are required to generate estimates.

Conclusions and further lines (cont.)

Main future lines:

- The main future lines are aimed at updating the IAEM without depending on the census year. This is possibly natural, when assigning an Autoregressive and Moving Average behavior for the part of Dzt and considering what was observed for Dxt since this information can be obtained in real time.
- Also, consider possible estimates for the groups of subsectors considered and not just the aggregate as has been done in this work. Additionally, it is interesting to analyze the reasons of the discrepancies between Dz and $\widehat{Dy}$, about all, for municipalities where the economic activity is small.
- Finally, it is intended to consider the original suggestion of Guerrero and Mendoza (2019) regarding the coverage percentage of the official estimates with respect to the interval generated by the estimates that also consider the nighttime luminosity.

Thank you