Estimación de población urbana a partir de imágenes satelitales

Tomás Rey Deutsch
Brian Koblinc

Capas de asentamiento humano global (GHSL)

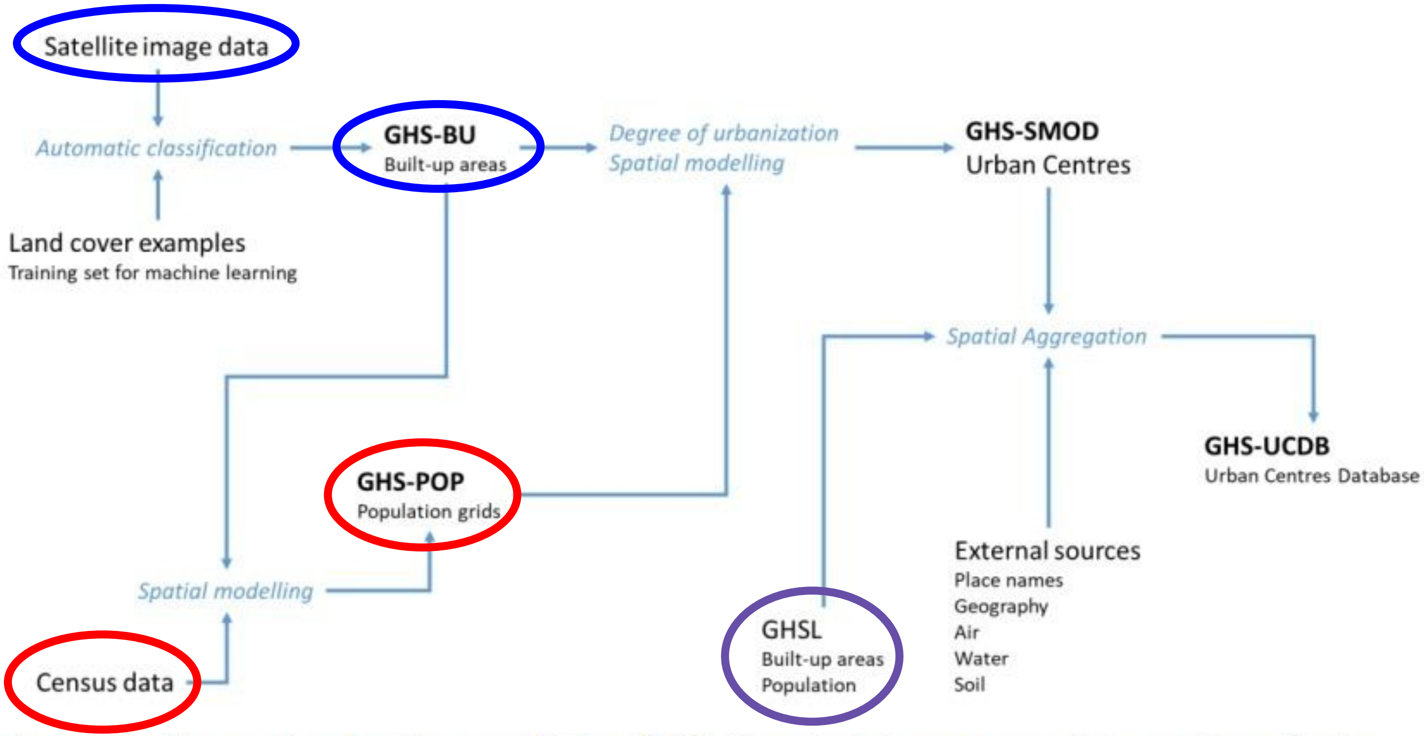
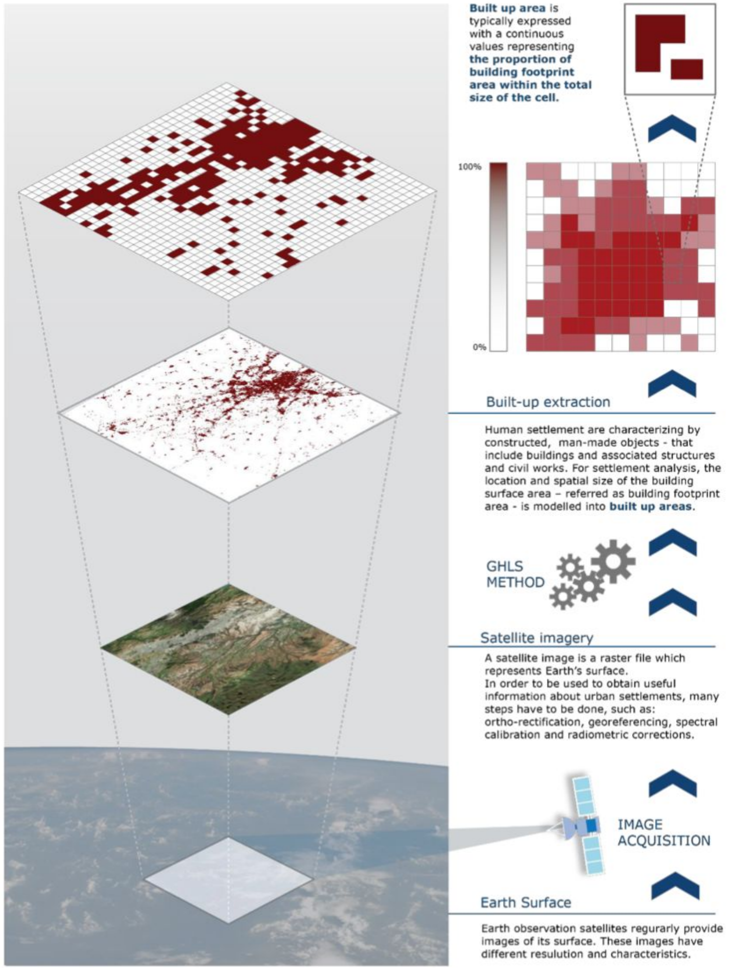
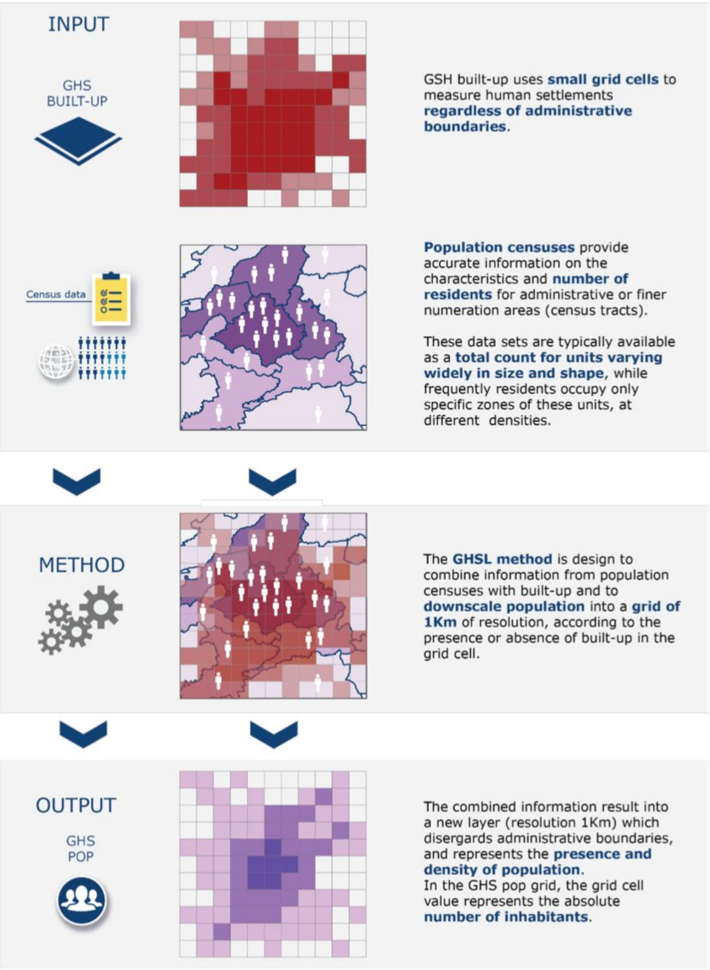


Figure 1. Conceptual schema of the GHSL input data, processing and products.


Built up area is typically expressed with a continuous values representing the proportion of building footprint area within the total size of the cell.
100%
0%
Built-up extraction
Human settlement are characterizing by constructed, man-made objects - that include buildings and associated structures and civil works. For settlement analysis, the location and spatial size of the building surface area – referred as building footprint area - is modelled into built up areas.
GHLS
METHOD
Satellite imagery
A satellite image is a raster file which represents Earth's surface.
In order to be used to obtain useful information about urban settlements, many steps have to be done, such as: ortho-rectification, georeferencing, spectral calibration and radiometric corrections.
IMAGE
ACQUISITION
Earth Surface
Earth observation satellites regurarly provide images of its surface. These images have different resulution and characteristics.


INPUT
GHS
BUILT-UP
GSH built-up uses small grid cells to measure human settlements regardless of administrative boundaries.
Census data
Population censuses provide accurate information on the characteristics and number of residents for administrative or finer numeration areas (census tracts).
These data sets are typically available as a total count for units varying widely in size and shape, while frequently residents occupy only specific zones of these units, at different densities.
METHOD
The GHSL method is design to combine information from population censuses with built-up and to downscale population into a grid of 1Km of resolution, according to the presence or absence of built-up in the grid cell.
OUTPUT
GHS
POP
The combined information result into a new layer (resolution 1Km) which disergards administrative boundaries, and represents the presence and density of population.
In the GHS pop grid, the grid cell value represents the absolute number of inhabitants.

Algunas de las variables más relevantes

Built-up area

Superficie que ocupa la zona edificada de una urbe

GDP (PBI)

Magnitud macroeconómica que expresa el valor monetario de la producción de bienes y servicios de demanda final de una región

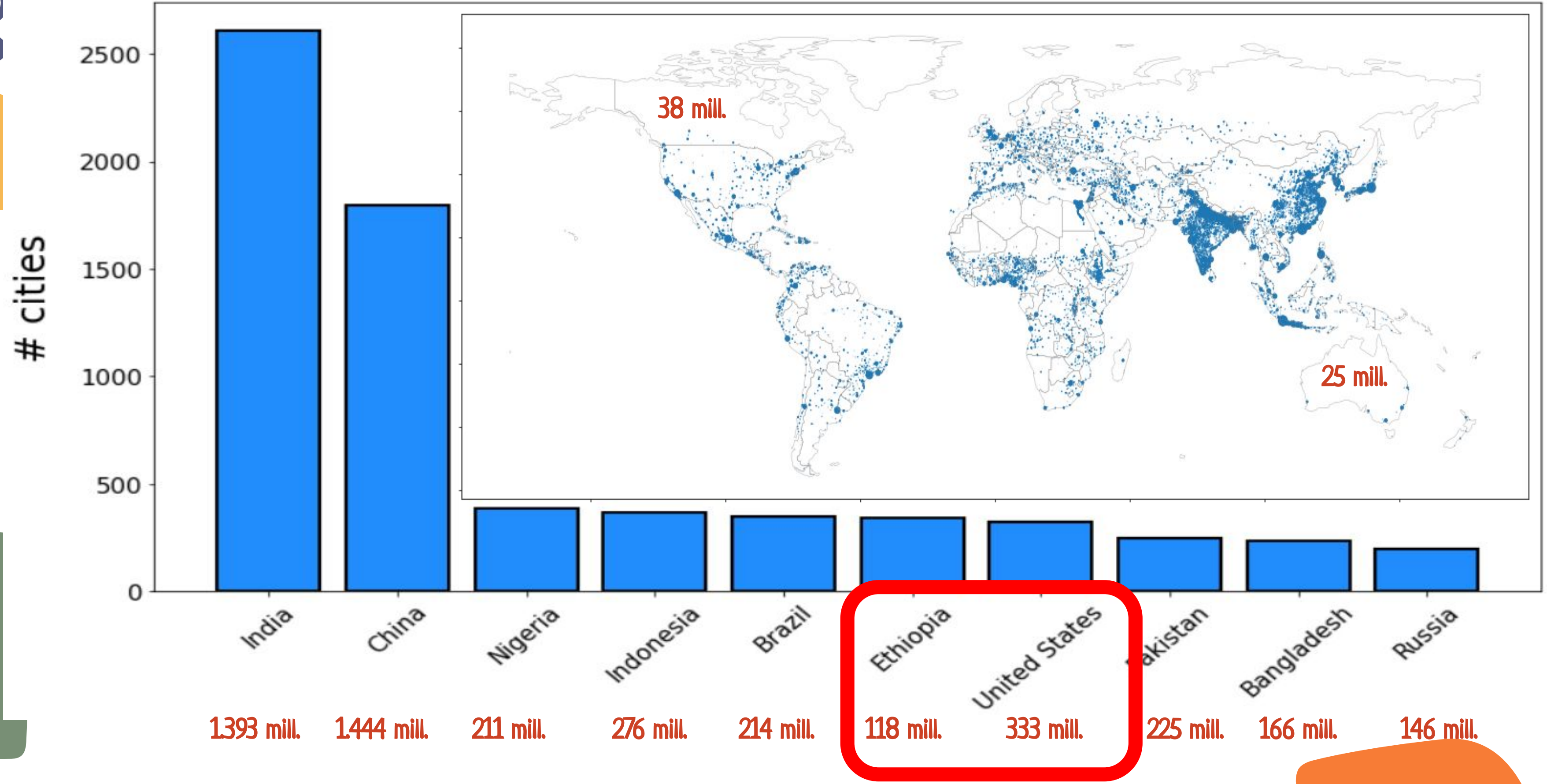
PM2.5

Concentración de partículas suspendidas en el aire con un un diámetro de menor a 2,5 μm

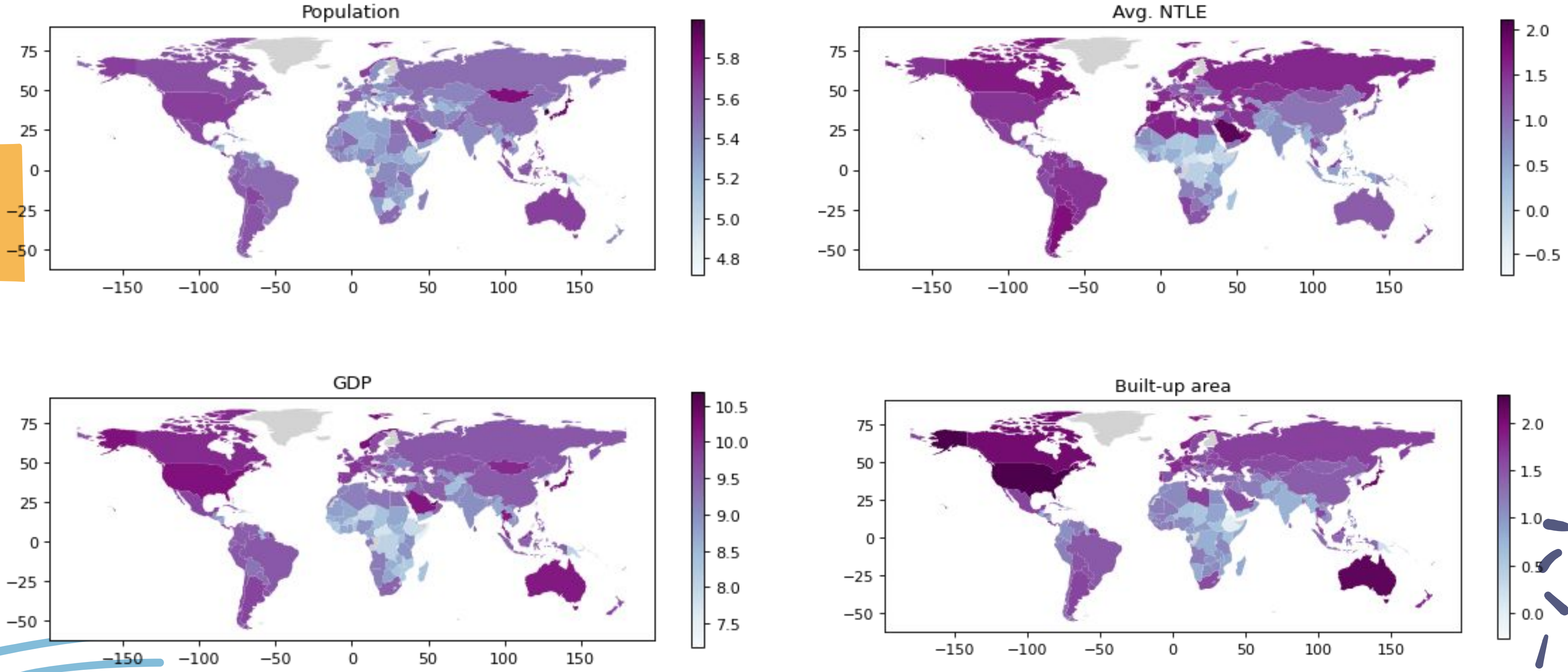
Avg. NTLE

Cantidad promedio de luz colectada por noche mediante un conjunto de radiómetros de imágenes infrarrojas-visibles (VIIRS)

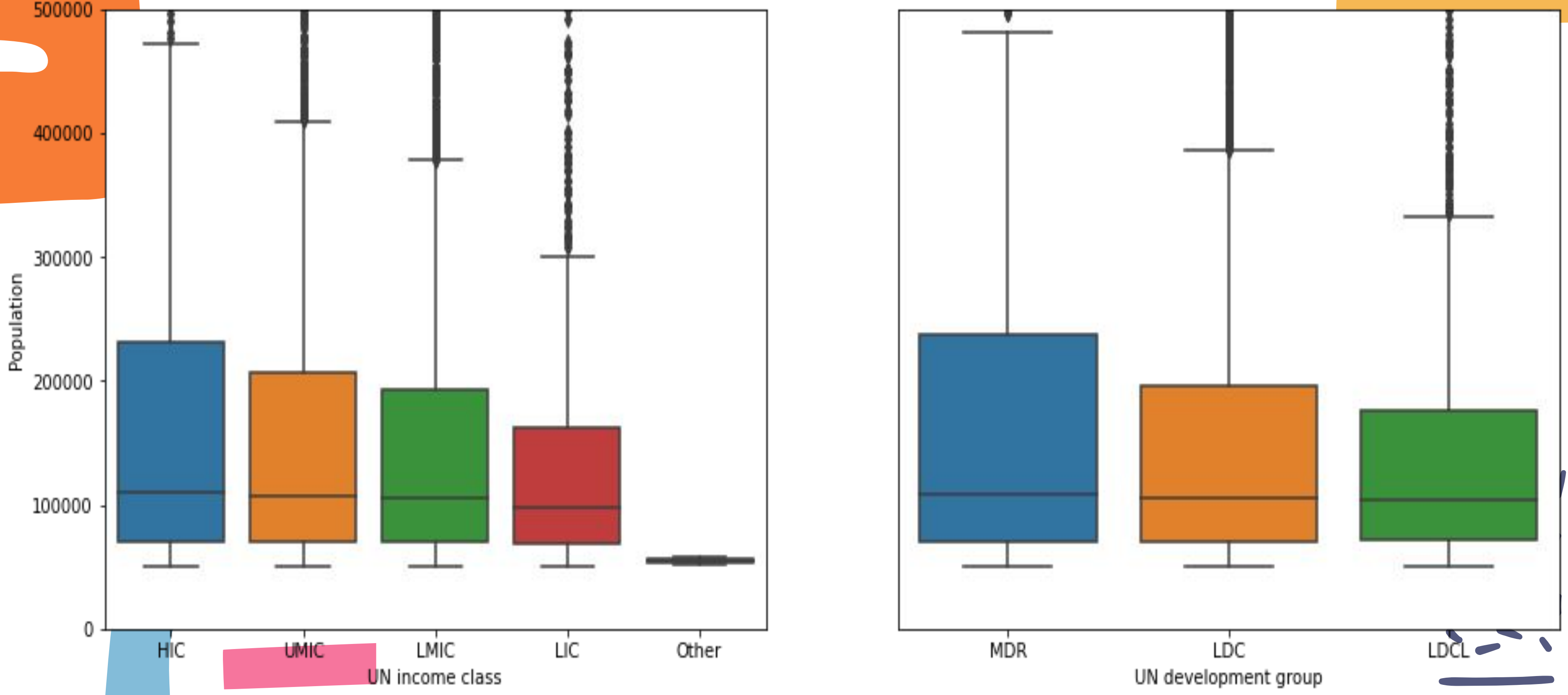
Distribución espacial de las urbes



Valores promedio por país en escala logarítmica

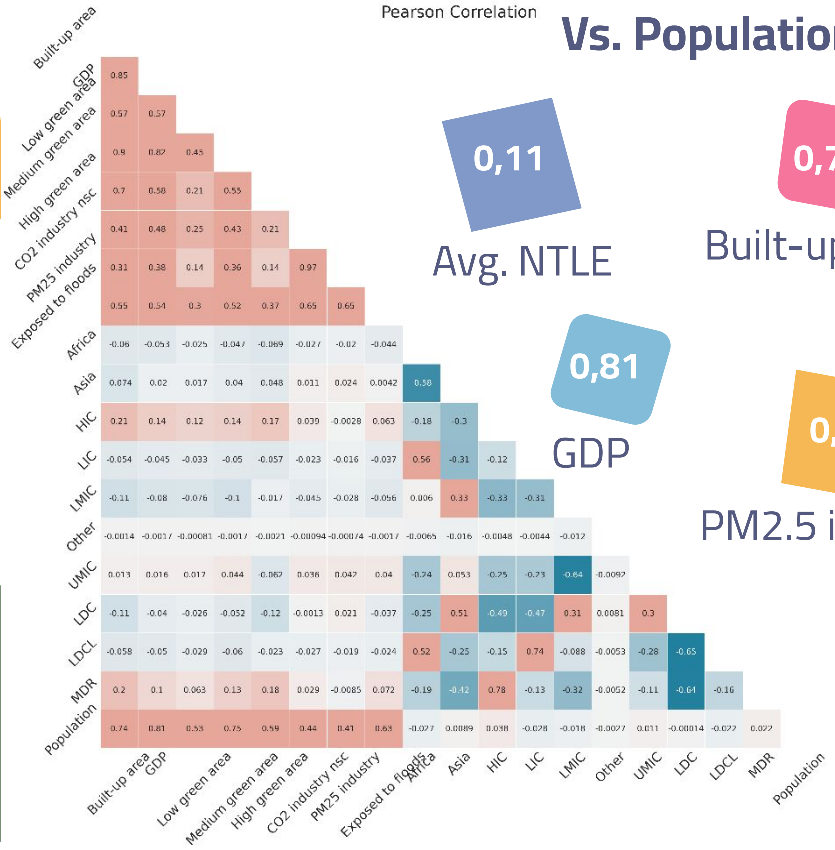


Clasificación ONU vs Población



Correlación entre variables

Vs. Population

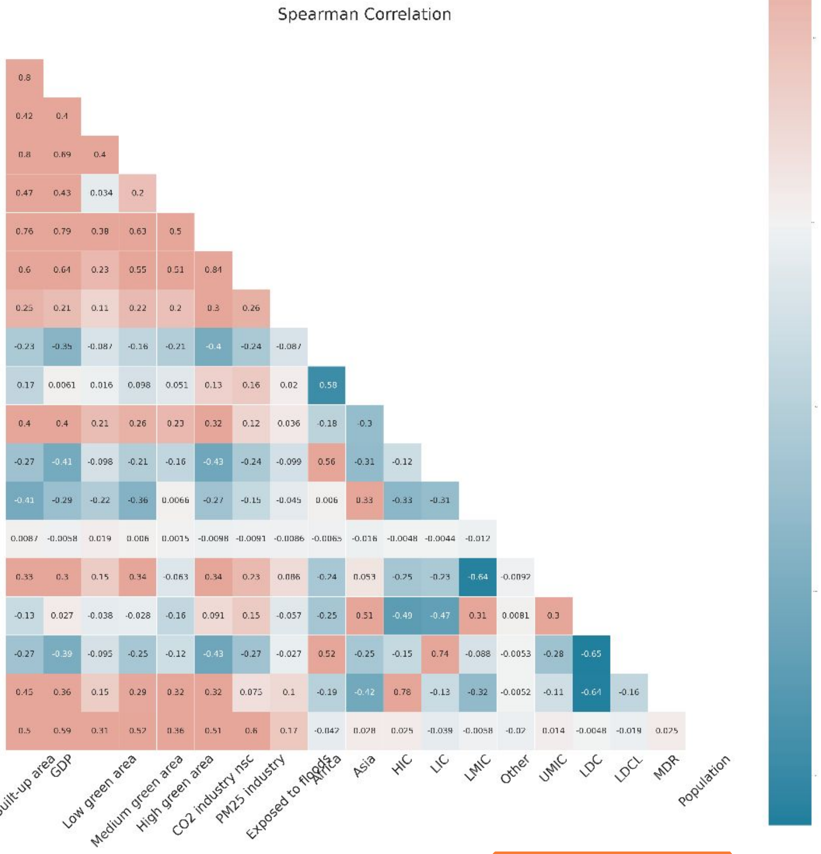


0,11
Avg. NTLE

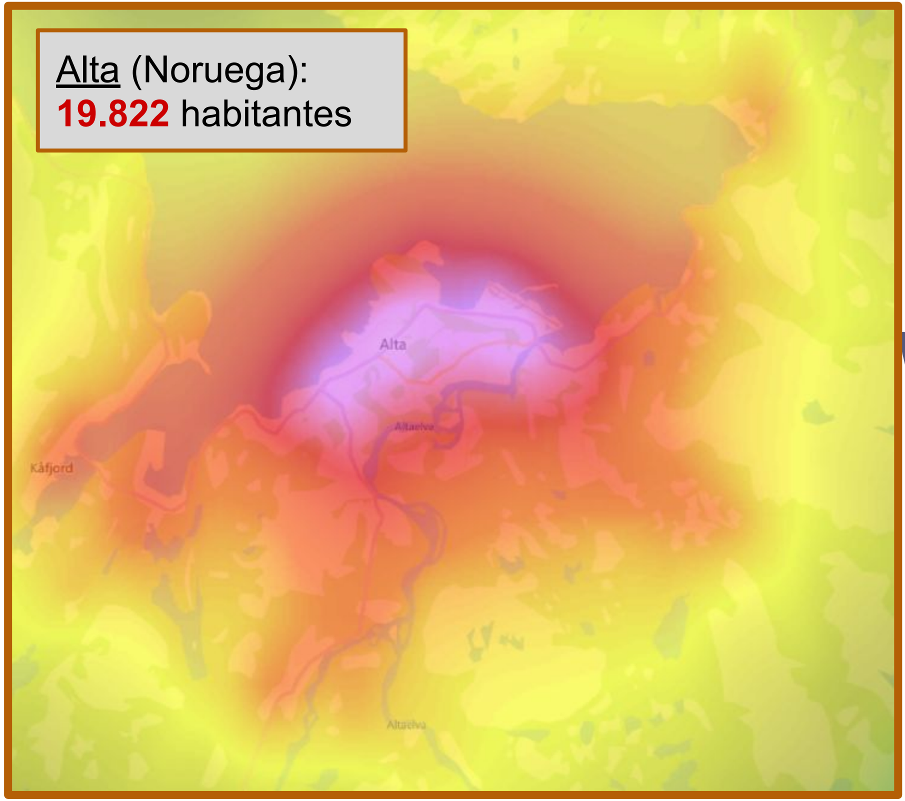
0,74
Built-up area

0,81
GDP

0,41
PM2.5 industry



Emisión de luz artificial



https://www.lightpollutionmap.info/

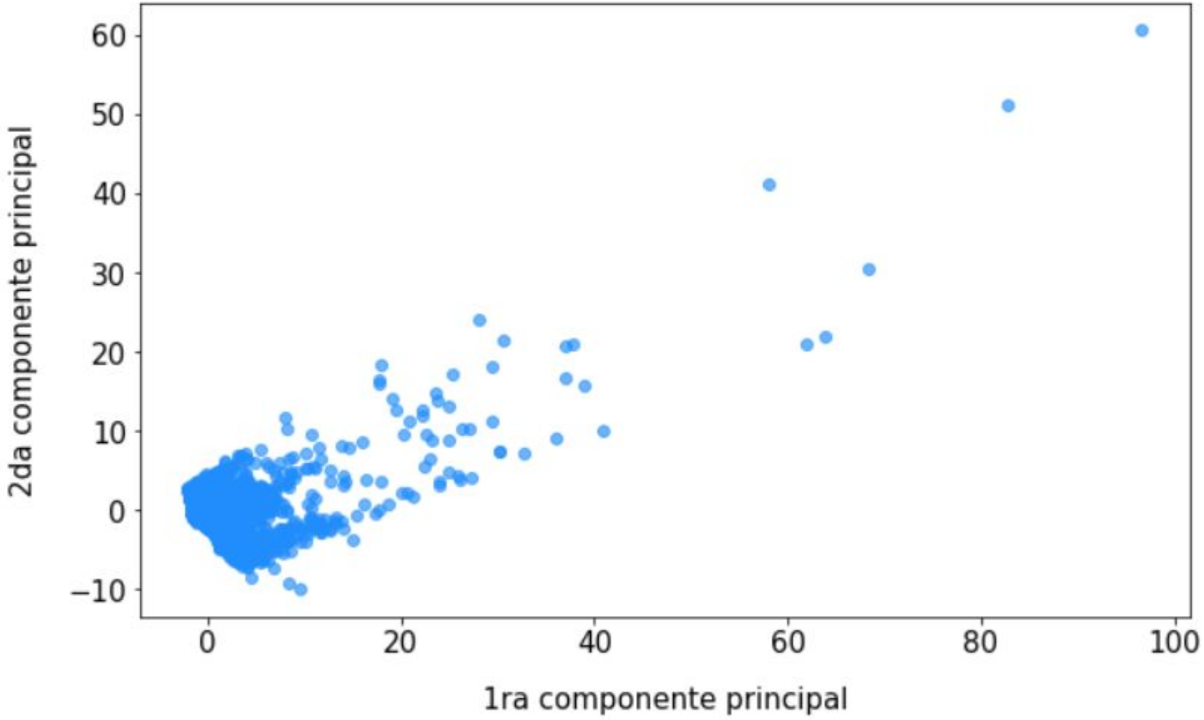
Emisión de luz artificial

Rubaya (Rep. Dem. Congo):
342.564 habitantes

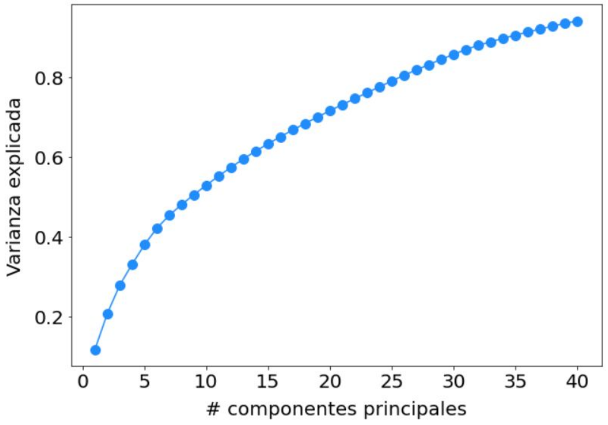
Alta (Noruega):
19.822 habitantes

https://www.lightpollutionmap.info/

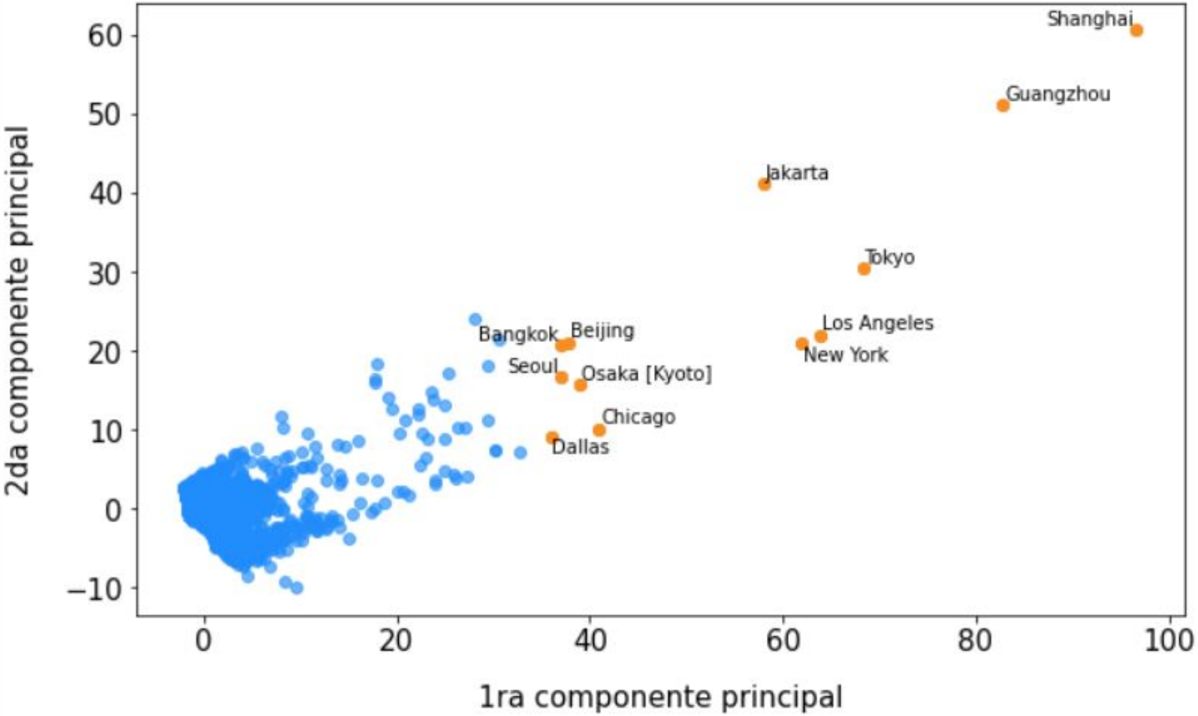
PCA



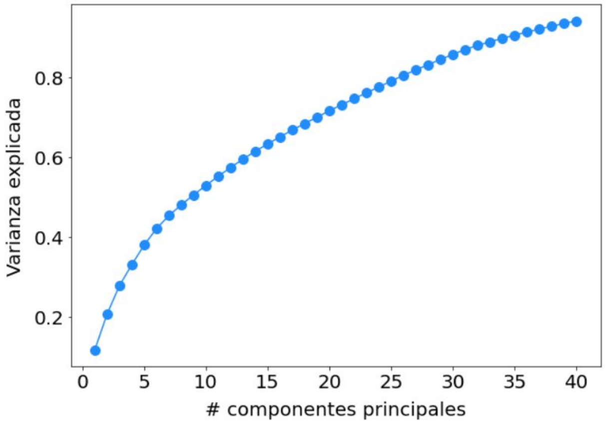
1ra comp. ppal		2da comp. ppal	
Area	**0.26**	Avg. temp.	**0.23**
Built-up area	**0.26**	LMIC	**0.22**
CO2 transp.	**0.25**	Asia	**0.22**
GDP	**0.24**	Longitude	**0.19**
MDR	**0.23**	PM25 total	**0.18**



PCA



1ra comp. ppal		2da comp. ppal	
Area	0.26	Avg. temp.	0.23
Built-up area	0.26	LMIC	0.22
CO2 transp.	0.25	Asia	0.22
GDP	0.24	Longitude	0.19
MDR	0.23	PM25 total	0.18



Modelos predictivos de regresión

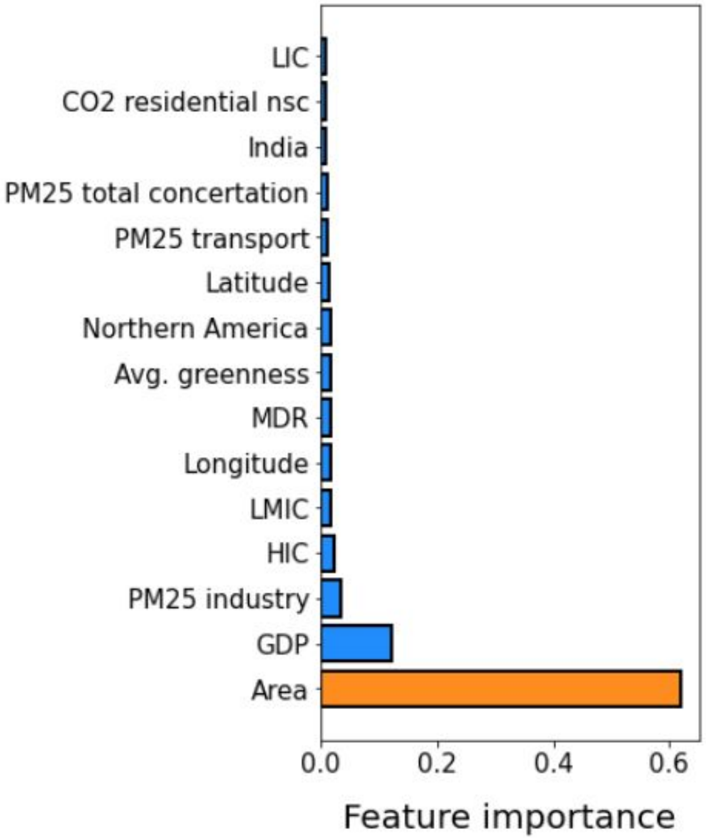
Modelos	Train score	Test score
KNN	-	0.48
Lin. Reg.	0.76	0.67
SVR	0.91	0.78
Rand. forest	0.98	0.90

Modelos predictivos de regresión

Modelos	Train score	Test score
KNN	-	0.48
Lin. Reg.	0.76	0.67
SVR	0.91	0.78
Rand. forest	0.98	0.90
AdaBoost	0.73	0.72
Gradient Boost	0.99	0.92
XGBoost	0.99	0.93

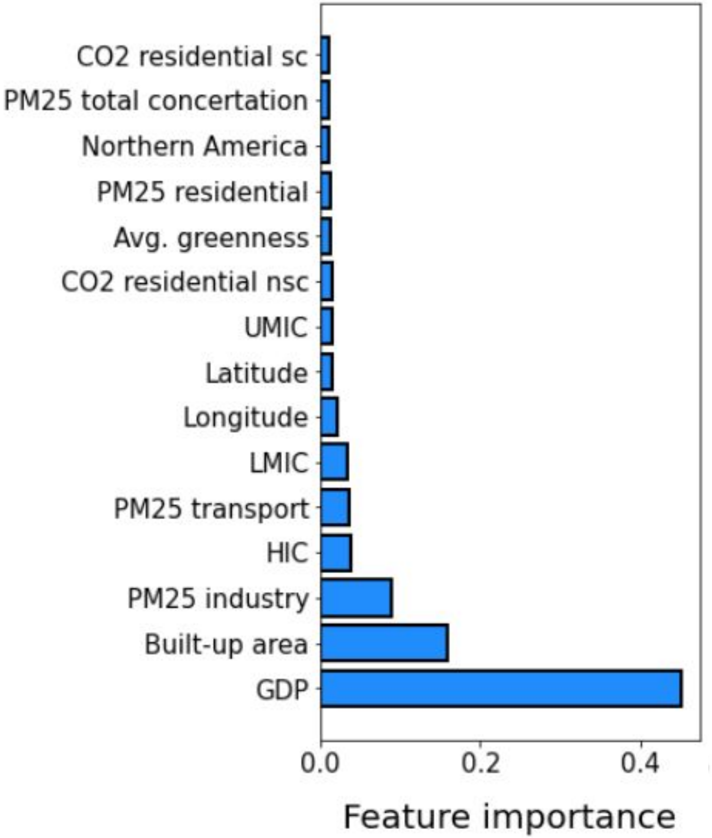
Modelos predictivos de regresión

Modelos	Train score	Test score
KNN	-	0.48
Lin. Reg.	0.76	0.67
SVR	0.91	0.78
Rand. forest	0.98	0.90
AdaBoost	0.73	0.72
Gradient Boost	0.99	0.92
XGBoost	0.99	0.93



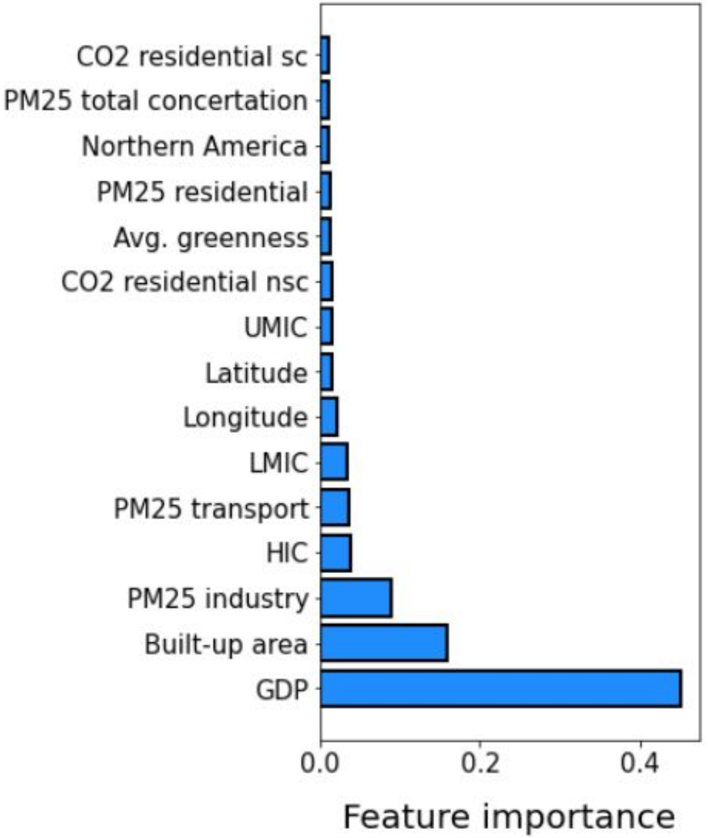
Modelos predictivos de regresión

Modelos	Train score	Test score
KNN	-	0.48
Lin. Reg.	0.76	0.67
SVR	0.91	0.78
Rand. forest	0.98	0.90
AdaBoost	0.73	0.72
Gradient Boost	0.99	0.89
XGBoost	0.99	0.89



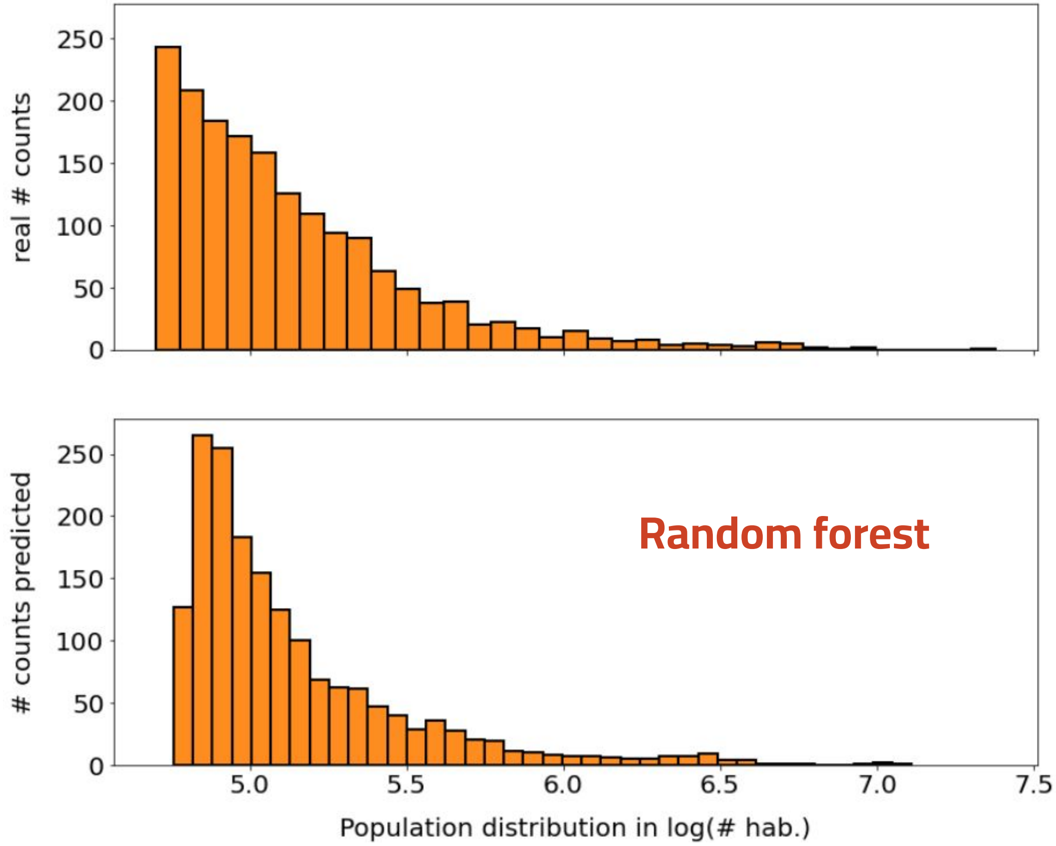
Modelos predictivos de regresión

Modelos	Train score	Test score
KNN	-	0.48
Lin. Reg.	0.76	0.67
SVR	0.91	0.78
Rand. forest	0.98	0.90
AdaBoost	0.73	0.72
Gradient Boost	0.99	0.89
XGBoost	0.99	0.89

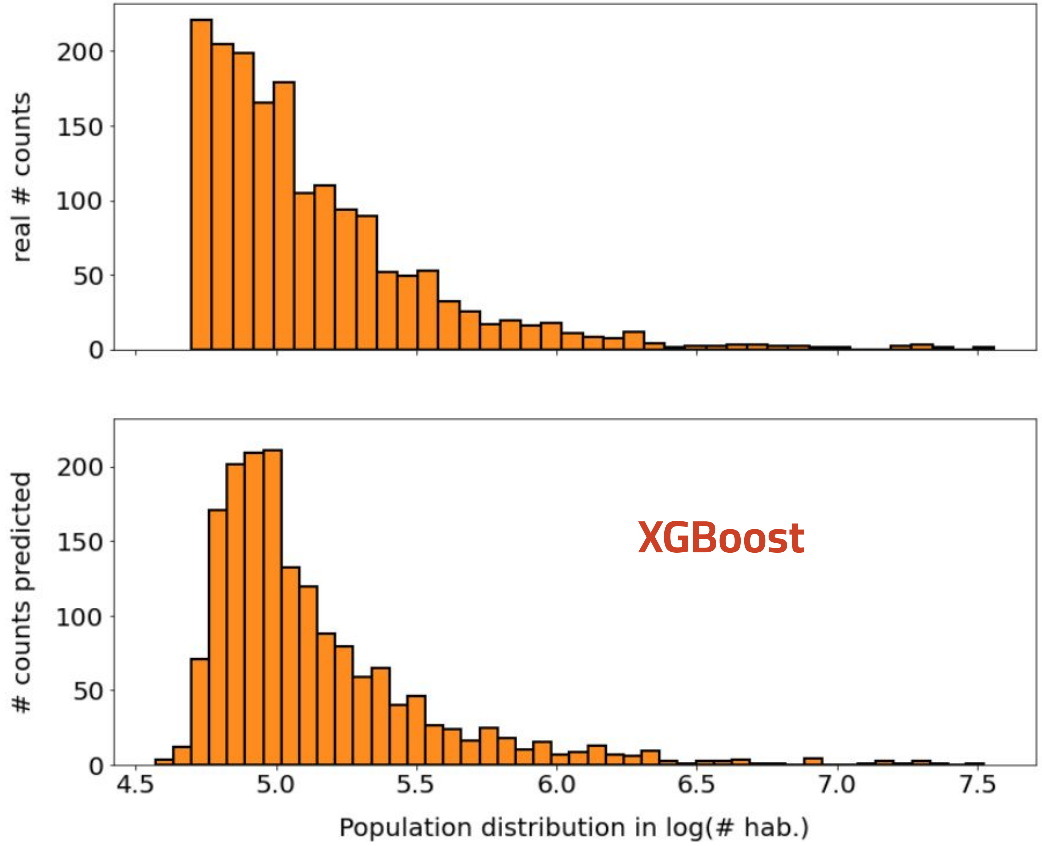


Urbe	Real (mill. hab.)	Pred (mill. hab.)	Error rel.
Bs. As.	13.91	14.43	3%
Rosario	1.10	0.93	15%
Córdoba	1.37	1.41	3%
Tokyo	33.03	27.47	16%
Beijing	17.98	15.99	11%
L.A.	14.28	14.11	1%
Paris	9.71	9.65	<1%
Berlin	3.27	2.92	10%

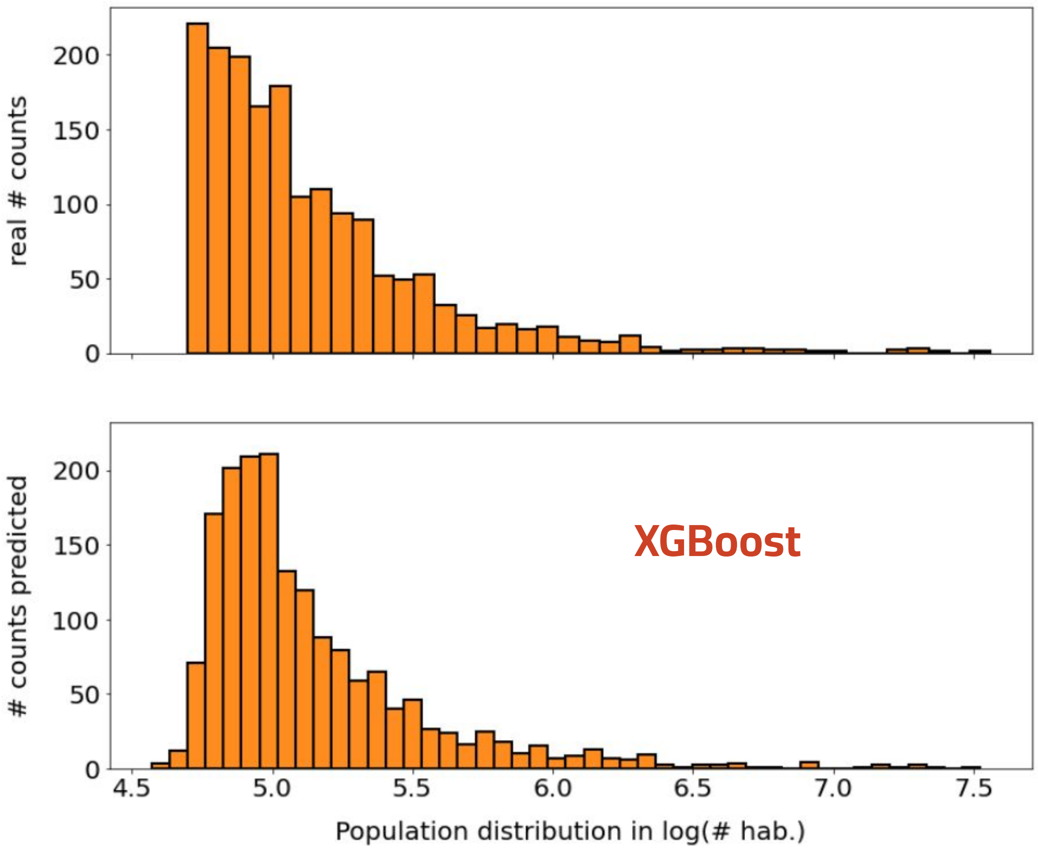
Distribución de errores

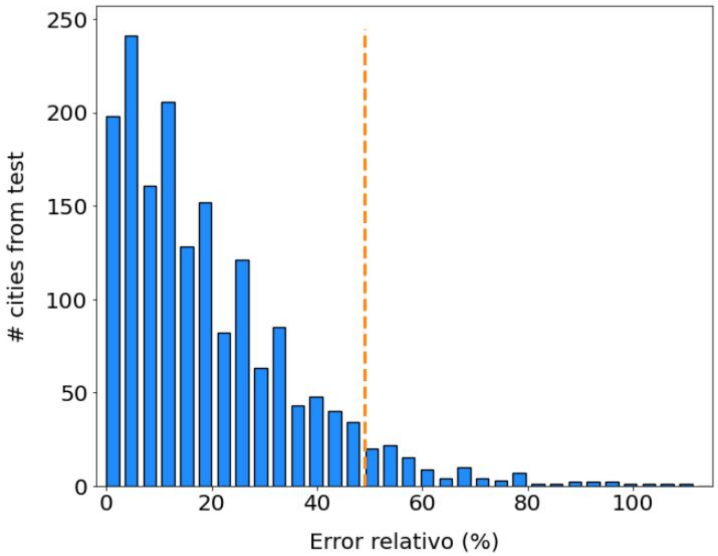


Distribución de errores

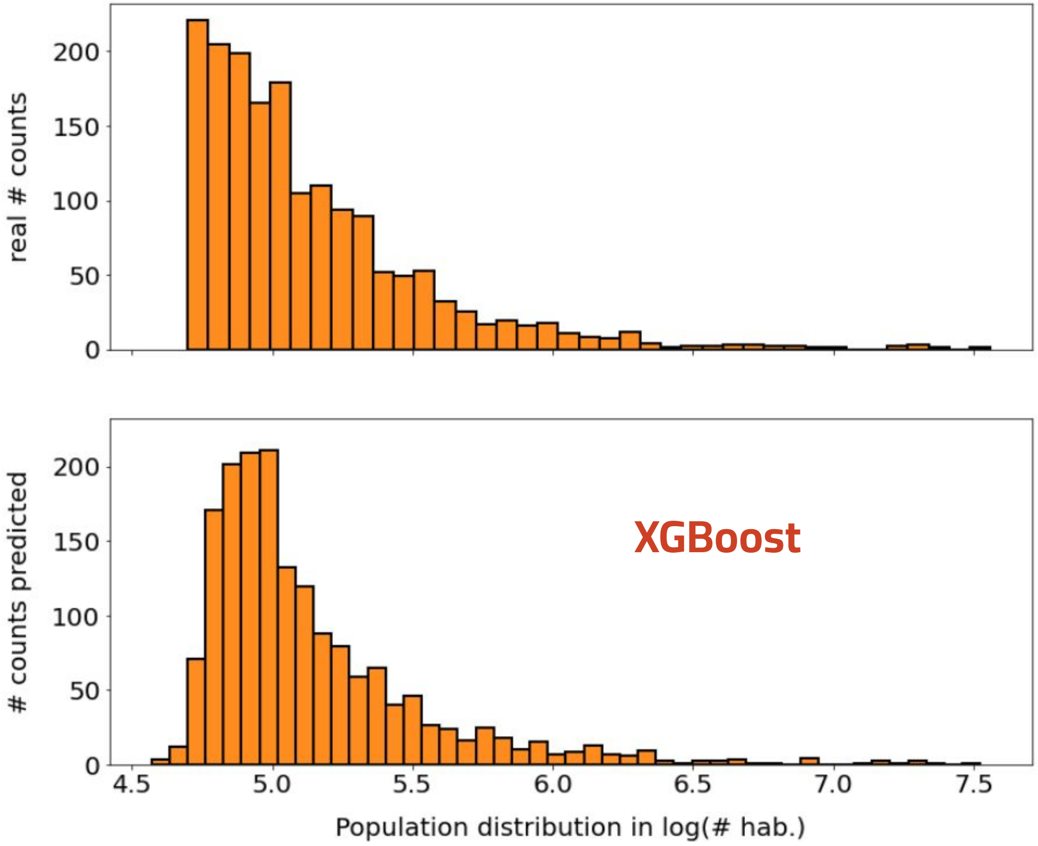


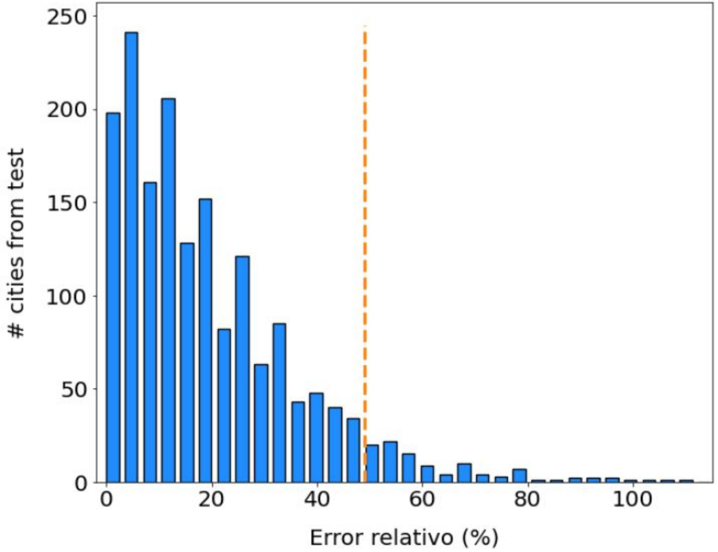
Distribución de errores





Distribución de errores





Conclusiones

1. Más barato y rápido que un censo pero menos preciso.

2. La emisión de luz artificial no es un buen predictor de poblaciones.

3. Se podría tratar de hacer modelos que predigan índices de urbanización o hagan forecasting para tratar de predecir crecimientos poblacionales.