Meiqing Li
CPLN 680 Advanced Topics in GIS
Professor Dana Tomlin
May 3, 2018

Estimation of Urban Intensity
Through Day and Night Remote Sensing Imagery

Abstract

Remote sensing imagery with its objectivity and consistency across space and time, serves as a good source to explore land uses. Both daytime and nighttime remote sensing imagery can shade light on spatial patterns of human activities and settlements, especially in cities where social and economic data are scarce or with limited access. While nighttime light (NTL) imagery is a proxy for urban intensity, daytime land cover data well indicate the built-up areas. This project develops a generalizable tool using Google Earth Engine to estimate urban intensity by normalized difference urban index (NDUI), and validates this model by socio-economic data in different contexts. The exploratory analysis tested the validity of county/MSA[1] level NDUI of California. It then extended the scope of analysis applying the model to the 100 largest MSAs, as cases in developed and developing countries respectively. The results show NDUI is more significantly associated with road density in cities of US, while it better indicates job density in Mexican cities.

Introduction

As a planner, my particular interest in land use lies in patterns of urban development, urban form as well as its relationship with human activities. This project was inspired by an attempt to study the unrevealed facts underlying progressive housing development in some Chinese cities where actual occupancy data are mysterious. It occurred to me that night light which naturally distinguish most of human activities might be an ideal source to identify areas with high concentration of human settlements and activities. By comparing day and night imageries, I will be able to identify the areas with housing oversupply. This idea raised a series of questions that needs to be tackled before achieving the final goals. *How do we measure human activities? What is the spatial resolution of NTL that associates with human activities? Does it indicate more of street lightings or actual human settlements?* These are the questions I want to address using the available data, tools and skills.

[1] Metropolitan Statistical Area

Goals & Objectives

The primary goal of this project is to develop an urban intensity index based on day and night remote sensing imagery, and assess its validity through correlation with social and economic indicators. The secondary goal is to develop user-cases, and conduct case studies to test the generalizability of this tool.

A further step yet to reach is to develop a predictive model based on exploratory studies that could be generally applied to various contexts.

Literature Review

Urban intensity or the intensity of urban environments is a common urban form measure that captures the density or concentration of people or other urban elements. Guan and Rowe (2016) defines urban intensity as related to compactness, diversity, density, and connectivity, regarding the distribution of resource consumption, economic opportunity, social integration and environmental performance. Some of these measures such as compactness and connectivity consider more of morphology whereas density and diversity rely more on aggregate values of socio-economic indicators. Most of the studies on NTL focus on the relationship between light intensity and non-morphological measures of urban intensity, probably due to the fact that morphological measures such as compactness or concentration can be derived from socio-economic indicators of disaggregate units of analysis. Therefore, the significance of these measures (compactness, concentration) depends on the quality of disaggregate data.

In terms of methods, several researches have validated the relationship between NTL and urban activities applying variations of DMSP/OLS, Landsat and MODIS images as well as in different contexts. For example, Mellander et al. (2015) explore the correlation between nighttime light (NTL) and socio-economic dynamics using fine-grained statistical data from Sweden. Akiyama (2012) conducted statistical tests and regression on light intensity and urban spatial factors including road and building density, as well as dynamic population distribution. Some researchers attempt to improve the robustness of estimation through calibration by combinations of satellite imagery. Zhang et al. (2013, 2015) introduce two different measures of urban intensity integrating DMSP/OLS NTL, Landsat and MODIS NDVI composites. Since most of existing researches rely on specific datasets pertinent to study area, the models tend to lack generalizability.

Data & Tools

The primary platform for raster imagery processing and computation in this project is Google Earth Engine, a planetary-scale platform that allows users to run algorithms on georeferenced imagery and vectors stored on Google's infrastructure. I utilized the Earth Engine (EE) Code Editor (a web-based IDE for the Earth Engine JavaScript API)'s capacity to obtain time-series urban intensity index (NDUI) from day and night satellite imageries.

The NDUI computation incorporated DMSP/OLS NTL and MODIS 16-day NDVI composite as two major raster dataset in this study. Depending on data availability, it took the annual average of years from 2000 to 2013 for longitudinal analysis. As for the cross-sectional comparison among different scopes of analysis, I used images from the most recent available year 2013.

DMSP/OLS NTL

Prepared by NOAA, the DMSP-OLS Nighttime Lights Time Series Version 4 provides cloud free composite making use of all available data from each calendar year from 1992 through 2013. It has a 'stable lights' band which is a very good source to indicate persisting lighting in urban areas. However, pixels in urban areas are often saturated and suffer from overglow effects (Zhang, et al., 2015). With 30 arc second (about 1km) grids, its resolution is coarse. Therefore, it would be ideal to have other data complementing NTL.[2]

MODIS 16-day NDVI Composite

The MODIS imagery has a median resolution at the range of 250m to 1km. Its 16-day NDVI composite can thus be a substitute when good-quality Landsat-7 imagery is not available or well-calibrated. This imagery source has a time-series collection since 2000.

Vector data collection acquired the geographic boundaries of different scopes of analysis (i.e. counties/MSAs in California, 100 largest MSAs in the US, and 100 largest cities in Mexico). The Google Earth Engine API requires data preprocessing to convert shapefiles to compatible Google Fusion Table format.[3]

Major sources of data for validation are census, jobs and road shapefiles. For the US, land area, population and median household income come from 2011-2015 ACS 5-year estimates; the number of jobs for each city is matched and aggregated from census tract level Longitudinal Employer-Household Dynamics (LEHD) data to MSA level. For Mexico, most of the socio-economic data come from Insitituto Nacional de Estadistica y Geografia (INEGI). To ensure consistency across countries, I extracted road information of cities in both countries from Open Street Map (OSM) in January, 2018.[4]

The above-mentioned data collection and assembly, as well as statistical analysis are mainly conducted in ArcGIS and RStudio, an open-source IDE for R.

Methodology

The overall NDUI algorithm is based on Zhang (et al., 2015). It is a normalized difference index analogous to NDVI (Normalized Difference Vegetation Index), a commonly used index to

[2] Landsat-7 imagery from USGS is a fine-grained dataset with 30m resolution. The annual NDVI composite could thus be an ideal source to accompany NTL. However, there has been ETM+ scan line corrector (SLC) failure since May 2003, resulting in 22% missing pixels. Without calibration, the Landsat-7 NDVI Composite is not considered a good source to indicate land cover conditions. Therefore this study opt for MODIS NDVI composite as an alternative source, though with lower resolution.

[3] I consulted *Shape Escape* (http://shpescape.com/) for help.

[4] This is not so consistent in terms of time alignment of data but generally applicable since temporal changes are not significant compared to cross-sectional differences.

determine the density of greenness on a patch of land. The NDUI composite combines the coarse-resolution NTL and finer-grained MODIS NDVI composite to maximize the separation between urban areas and easily confused barren lands.

The NTL value ranges from 0 to 63, and NDVI has a value between -1 and 1. The calculation of NDUI follows the rationale that there is an inverse relationship between vegetation density and luminosity. Therefore, when both NTL and NDVI are normalized, their normalized difference between -1 and 1 enhances the robustness of the estimated urban intensity. As a result, in urban core areas with sparse vegetation cover, NDUI will have a high value close to 1, while unlit rural areas with more vegetation will have lower values close to 1.[5]

$$NDUI = (NTL - NDVI)/(NTL + NDVI)$$

The project starts from an exploratory analysis that validate the appropriate geographic unit to interpret NDUI. At this stage, I used simple Pearson Correlation to test the relationship between county/MSA level NDUI and two density indicators: population density and road density.

$$\rho_{x,y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - [E(X)]^2}\sqrt{E(Y^2) - [E(Y)]^2}}$$

After validating the spatial specificity at county/MSA level, I applied the same estimation method to the top 100 cities in the US and Mexico, specifically the 100 largest MSAs in the US and 100 largest SUNs[6], the Mexican counterpart of US MSAs. With a larger sample size and more variation across countries and regions, I ran a simple linear regression to examine the pattern and relationship between NDUI and multiple socio-economic variables.

The key hypothesis is that urban intensity estimate is significantly associated with other urban development indicators (Table 1). However cities in developed and developing countries might exhibit different patterns due to special characteristics of urban form.

Table 1. Variables and expected relationships

Variables	Expected Relationship
NDUI	Dependent variable, a proxy for urban development intensity
Population Density (ppl/ha)	Independent variable, positively correlated with urban intensity
Job Density (#jobs/pa)	Independent variable, positively correlated with urban intensity
Road Density (feet/ha)	Independent variable, positively correlated with urban intensity
Median Household Income (1k)	Independent variable, positively correlated with urban intensity

[5] A tool for Google Earth Engine NDUI calculation was developed as a final project for CPLN 670 Geospatial Software Design in Fall, 2017. See appendix for more details of this tool.
[6] Sistema Urbano Nacional

Exploratory Analysis

Though the NDUI calculation tool can be easily applied to different contexts and scales for estimation purpose, data quality and the tool's processing capacity pose limitations on sample selection. Bearing in mind these limitations, I first selected California as the study area to help me determine the appropriate geographic unit for NDUI aggregation.

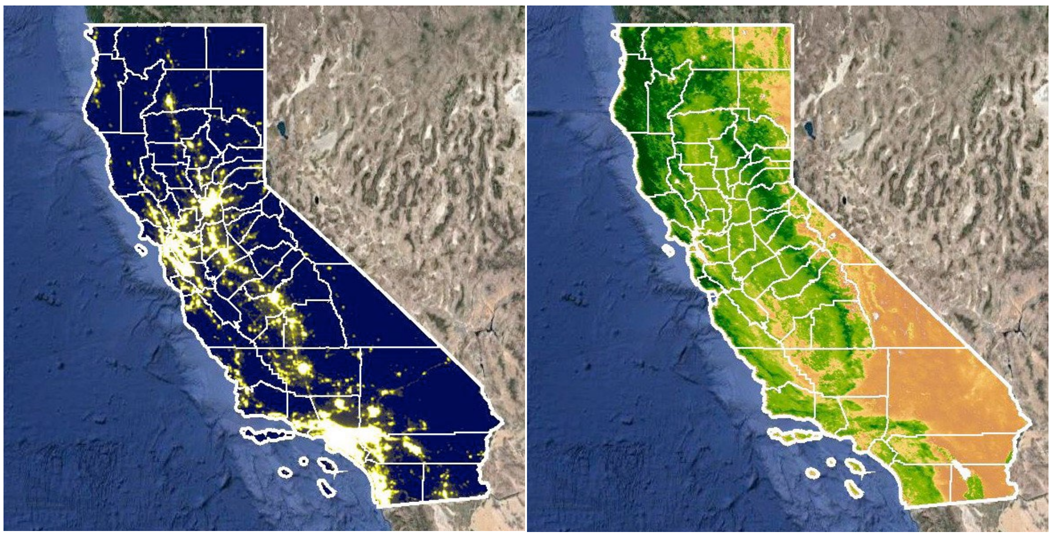


Figure 1. DMSP/OLS NTL and MODIS NDVI 16-day composite of California counties in 2013

There are in total 58 counties in California. The tool output annual estimates of these counties between 2000 and 2013 (Figure 2). The time-series output shows a reasonable trend of urban intensification of cities in California in the most recent years. The changes by magnitude, however, are not so significant.

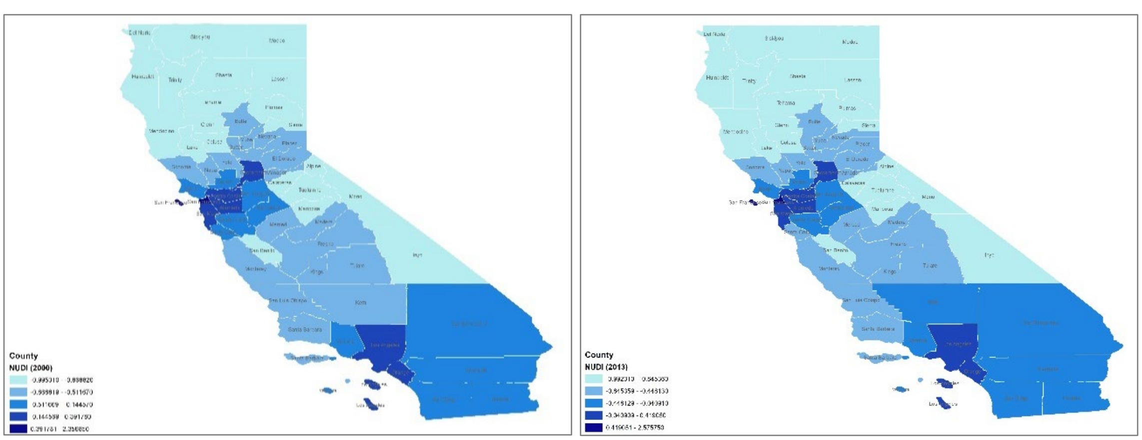


Figure 2. NDUIs for counties of California in year 2000 and 2013

Statistically, Pearson correlation tests the expected relationships between NDUI and population and road densities. The one between NDUI and population density is **0.875**; the other one between NDUI and road density is **0.659**. The strong positive correlation supports the assumption that county would an appropriate unit that captures the spatial pattern of urban intensity.[7]

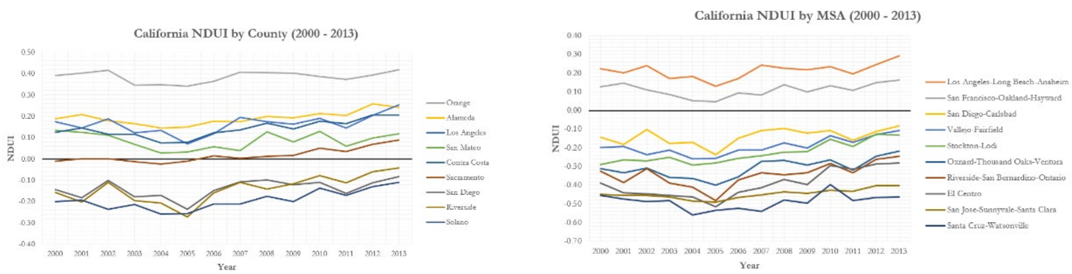


Figure 3. NDUIs for counties and MSAs of California from year 2000 to 2013

The exploratory analysis also reveals a fact that despite high urban intensity some of the exceptionally small counties such as San Francisco could not be captured due to the limitation of spatial resolution. Although county and MSA usually share boundaries and are of comparable scales, the MSA definition might be better at eliminating outliers and creating consistent results (Figure 3). Therefore, instead of counties I sampled from MSAs (and their Mexican counterparts) in the regression models.

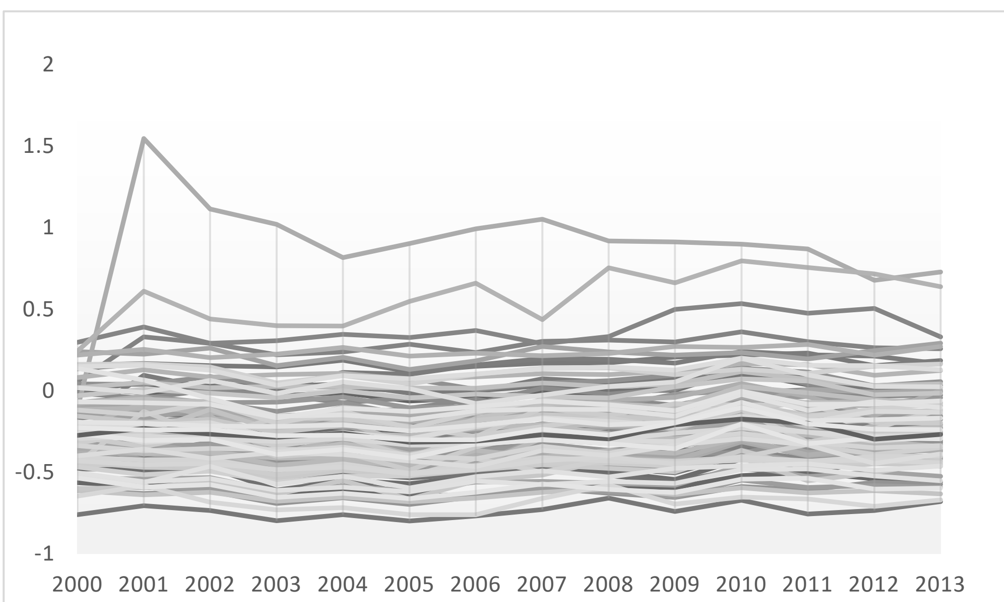


Figure 4. NDUIs for 100 largest MSAs in the US from year 2000 to 2013

[7] The CPLN 670 final report "Nighttime Light and Urban Intensity Estimation and Validation with Google Earth Engine and ArcPy" elaborates more about the tool and method.

Considering the limitation of the number of samples for EE to process at the same time, I sampled 100 largest cities in each country (US and Mexico) to examine the statistical relationship between NDUI and socio-economic indicators. Since time-series data are available, my first idea was to run panel data analysis with city-fixed effect. Same as in the exploratory analysis, I plotted the trend of NDUIs of 100 largest MSAs from year 2000 to 2013. Countrywide there is no significant temporal pattern as the levels of urban intensity differ, whose variation is larger than the temporal changes in urban intensity (Figure 4). A few cities that demonstrate novel patterns with exceptionally high NDUIs are Milwaukee-Waukesha-West Allies, WI, New Orleans-Metairie, LA, and Cleveland-Elyria, OH MSAs.

Based on the findings from exploratory analysis, I figured it makes more sense to develop a statistical model based on single year data across cities for each country. The 100 largest cities in each country were selected based on total population. Figure 6 and 7 illustrate the general urban structure of cities in both countries by simple classification of DMSP/OLS NTL imagery of 2013.[8] They showcase the difference in urban development structure of the two countries. Spatially cities in the US are more evenly distributed with two major corridors on the northeast and southwest of the country; while Mexico City being the single largest city in Mexico, urban development in the rest of country is marginalized.

Density plots demonstrate the distinguishable patterns too. There is an almost normal distribution of NDUI in the US. In Mexico however, there is a cluster of cities at 0 NDUI level. The overall US cities exhibit higher NDUIs than Mexican cities (Figure 5).

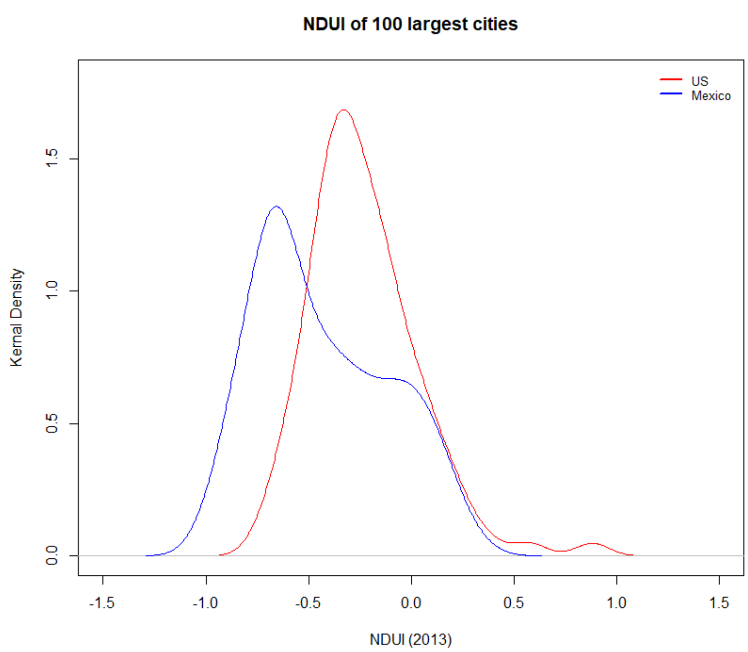


Figure 5. DMSP/OLS NTL classification of urban areas in 2013 (100 largest MSAs in the US)

[8] Areas in cities are roughly classified to urban (62,63], suburban (15,62], and rural [0,15] based on NTL value [0,63]. By this means, urban areas are identified as those with NTL saturation/overglow.

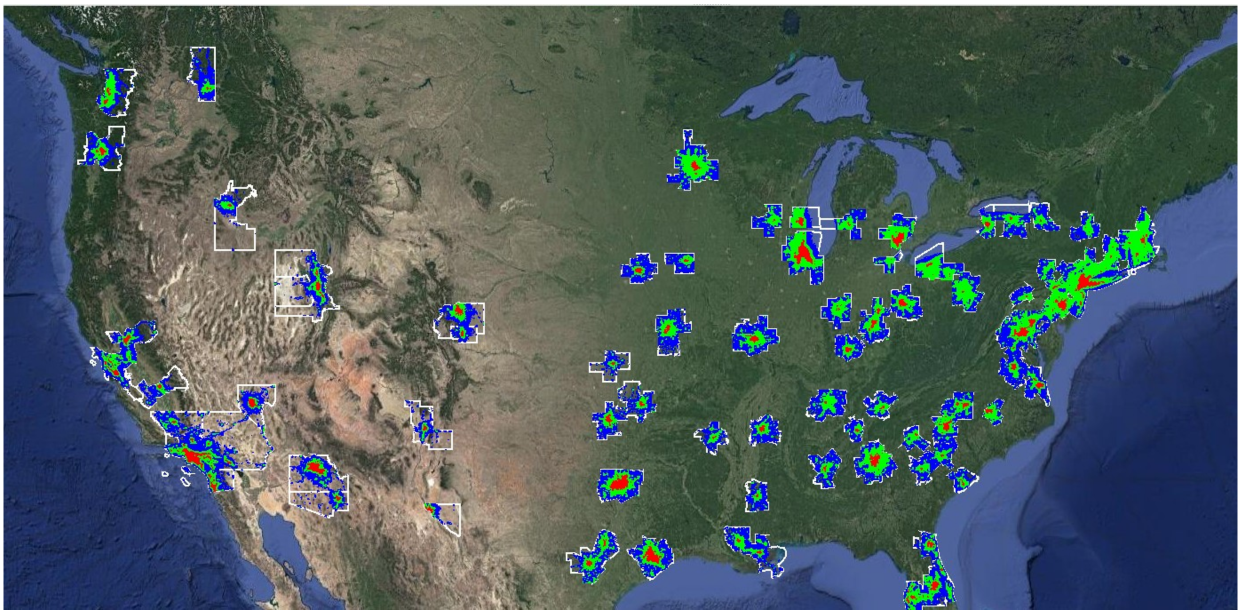


Figure 6. DMSP/OLS NTL classification of urban areas in 2013 (100 largest MSAs in the US)

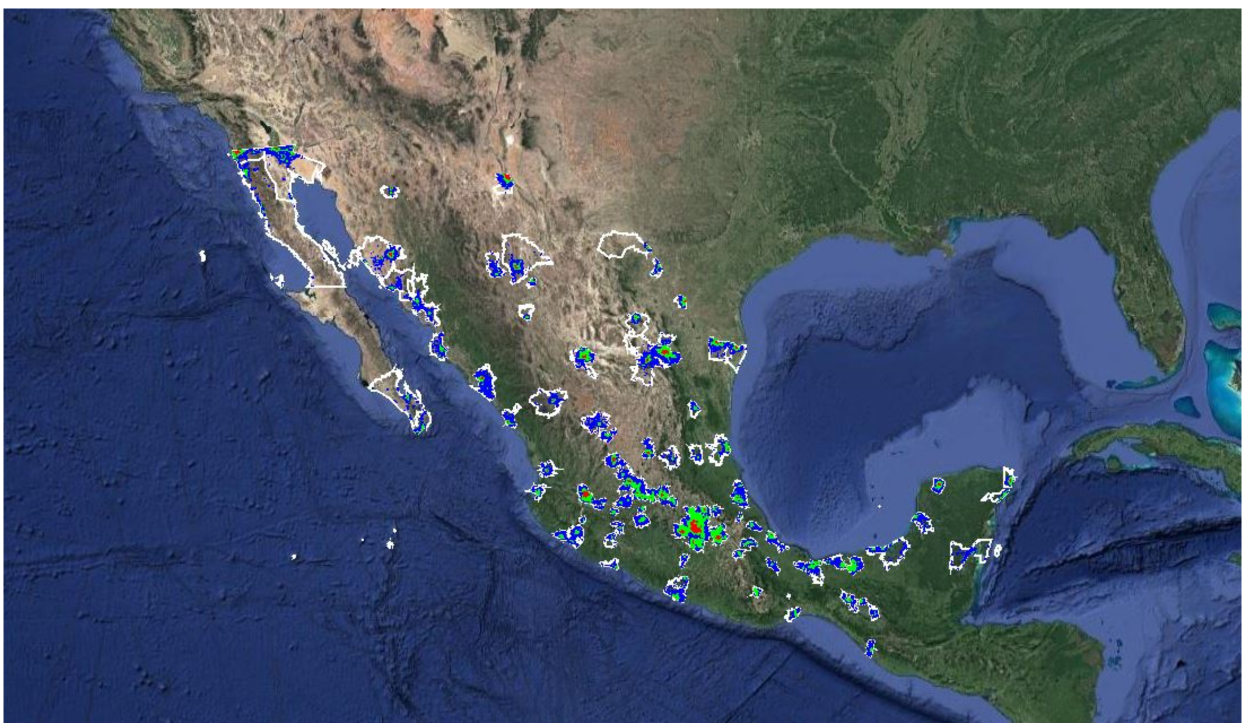


Figure 7. DMSP/OLS NTL classification of urban areas in 2013 (100 largest SUNs in Mexico)

OLS Regression

The dependent variable is citywide annual average NDUIs in 2013 (the most recent year available), computed by the previously developed tool. Figure 8 and 9 exhibit the aggregate urban intensity index in each country. The distributions of NDUI correspond with previous observation.

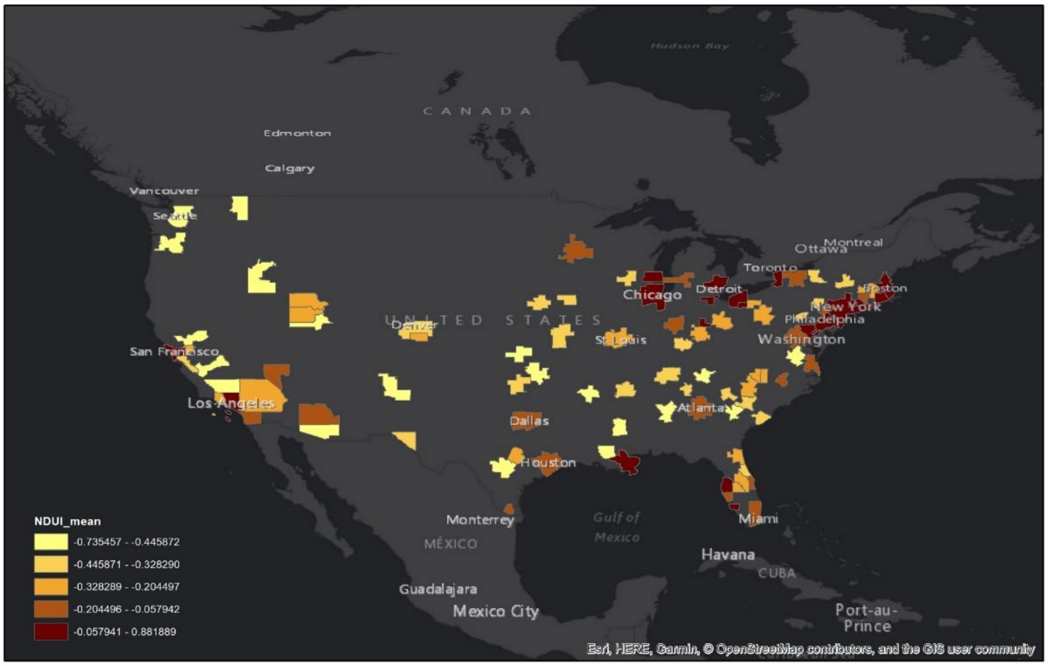


Figure 8. NDUIs of 100 largest MSAs in the US (2013)

Figure 9. NDUIs of 100 largest SUNs in Mexico (2013)

Because the dependent variable is almost normally distributed, I adopted simple OLS regression for both countries. Table 2 and 3 summarize all input variables of the model for each country.

Table 2. Summary statistics of dependent and independent variables (MSAs)

Variables	Obs	Mean	Std.dev.	Min	Max
NDUI	100	-0.27	0.26	-0.74	0.88
Population density (ppl/ha)	100	2.12	1.84	0.21	10.47
Job density (job/ha)	100	0.97	0.86	0.097	4.79
Road density (feet/ha)	100	110.23	49.19	0.31	255.96
Median household income (1k USD)	100	55.46	10.43	35.04	95.97

Table 3. Summary statistics of dependent and independent variables (SUNs)

Variables	Obs	Mean	Std.dev.	Min	Max
NDUI	100	-0.43	0.32	-0.95	0.29
Population density (ppl/ha)	100	51.25	16.04	17.85	103.08
Job density (job/ha)	100	8.50	3.15	2.70	20.96
Road density (feet/ha)	100	465.2	137.43	156.5	762.10
Median household income (1k USD)[9]	100	2.96	0.50	1.90	4.33

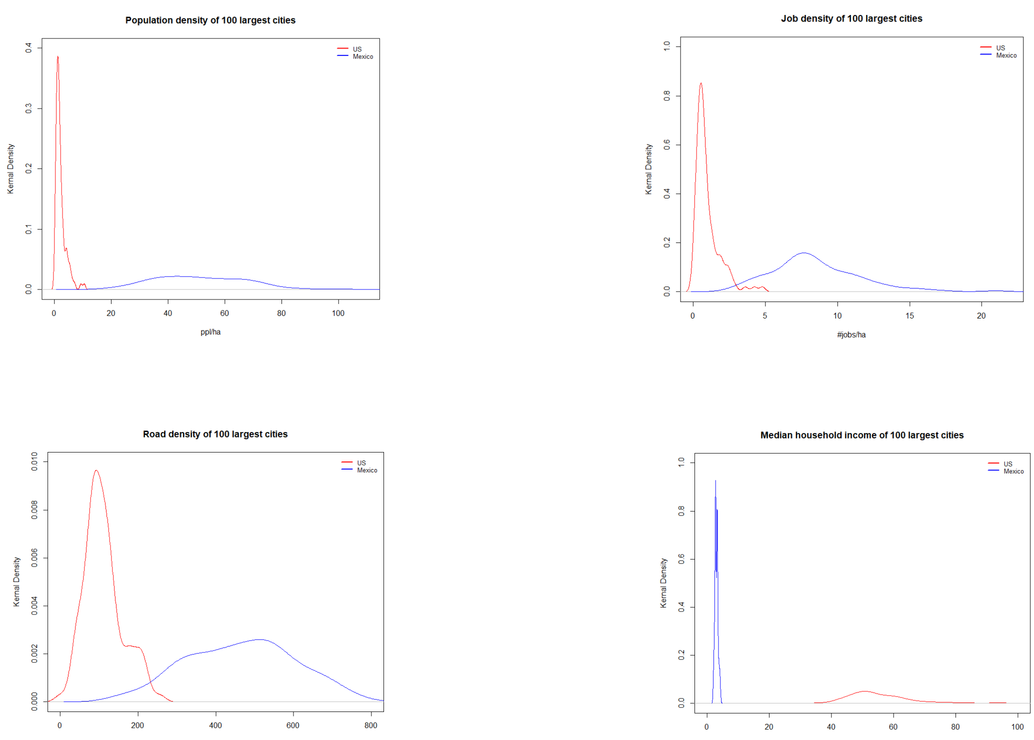


Figure 10. Density plots of each independent variable of both countries

[9] Converted from Mexican pesos (1USD = 19 Mexican pesos)

Figure 10 plots out the distributions of the socio-economic indicators in cities of both countries, which serve as independent variables for the OLS regression model. The three density indicators (i.e. population density, job density, and road density) demonstrate a contrast between the two countries: though US cities have a much higher level NDUI, they are much more spread out than cities in Mexico, revealed by the lower magnitude in terms of density (red lines in Figure 10). To ensure consistency, median household income of Mexico is converted to USD (1 USD = 19 Mexican pesos. In this regard, the average household income is much lower in Mexico compared with that in the US, which could potentially be positively correlated with levels of overall light intensity or NDUI.

Based on above analysis, I developed the following OLS regression model of city-level NDUI with different combinations of independent variables:

$$NDUI = \beta_0 + \beta_1 Pop\_den + \beta_2 Job\_den + \beta_3 Road\_den + \beta_4 Inc\_median$$

where:

- NDUI is the aggregate normalized difference urban index by city;
- Pop_den is the city-level population density (people/hectare);
- Job_den is the city-level job density (number of jobs/hectare);
- Road_den is the city-level road density (feet/hectare);
- Inc_median is the median household income of the city (1000USD).

I first ran separate models of each country with and without median household income added to the three density indicators. In order to capture the income effect across US and Mexico, with significantly difference levels of NDUI and income, I ran two additional models that combine US and Mexico datasets. Table 4 shows the results from these models.

Overall, the US model performs better with higher adjusted R^2. Household income is a significant predictor associated with NDUI. Incorporating the income variable to model in additional to densities enhances the model's predictive power. Interestingly, within each country household income is negatively correlated with the measure of urban intensity, while there is a positive correlation in the combined model. It is possible that income effect outweighs other factors when there exists a huge income gap between the two countries which determines the intensity of urban development. Nevertheless within each country, it is likely density plays a role: wealthier cities tend to spread more. Also, NDUI turns out to be a better proxy for different urban form indicators across the two countries. Comparing first and second models, I found NDUI serves as a significant indicator for road density in the US while in the Mexican case a better indicator for job density. This implies in the US case, instead of lightings from human settlements, NTL imagery more likely captures highway lightings. Whereas in Mexico, cities are less spread out as US cities, making it easier for highly concentrated cities to stand out. Moreover, a significant portion of jobs in Mexico are in the informal sector. Job counts excluding informal jobs on the other hand, likely distinguish intensively developed areas where more formal jobs congregate. Lastly none of the country-specific model indicates significant relationship between NDUI and population density.

Table 4. OLS regression models of NDUI and socio-economic variables in the US and Mexico

OLS Regression Results

	Dependent variable:					
	MSA	SUN	MSA	SUN	Combine	
	(1)	(2)	(3)	(4)	(5)	(6)
Population density	-0.017	-0.003	0.040	-0.003	-0.007***	-0.004*
	(0.071)	(0.003)	(0.070)	(0.003)	(0.002)	(0.002)
Job density	0.224	0.044***	0.070	0.037**	0.050***	0.047***
	(0.156)	(0.014)	(0.149)	(0.014)	(0.013)	(0.013)
Road density	0.001**	-0.0004	0.001**	-0.001**	-0.0005**	-0.0003
	(0.001)	(0.0003)	(0.001)	(0.0003)	(0.0002)	(0.0002)
Median household income	-0.006**	-0.145**				0.004***
	(0.002)	(0.069)				(0.002)
Constant	-0.244**	-0.007	-0.529***	-0.341**	-0.230***	-0.477***
	(0.120)	(0.204)	(0.051)	(0.130)	(0.036)	(0.095)
Observations	100	100	100	100	200	200
R^2	0.536	0.125	0.503	0.084	0.141	0.174
Adjusted R^2	0.516	0.089	0.487	0.056	0.128	0.157
Residual Std. Error	0.185 (df = 95)	0.306 (df = 95)	0.191 (df = 96)	0.312 (df = 96)	0.291 (df = 196)	0.286 (df = 195)
F Statistic	27.397*** (df = 4; 95)	3.405** (df = 4; 95)	32.359*** (df = 3; 96)	2.946** (df = 3; 96)	10.714*** (df = 3; 196)	10.292*** (df = 4; 195)
Note:						*p<0.1; **p<0.05; ***p<0.01

Discussion & Conclusion

Despite the acceptable validity of this model at MSA level, it is far from a generalizable tool to estimate or predict city-level urban intensity. However the models direct us to the most-likely predictors that bridge universal measures derived from raster imagery with vector-based objective urban indicators, such as job and road density. Building such connection not only helps us interpret urban development patterns from remote sensing imagery, but also estimate the actual socio-economic indicators wherever such data are not directly available.

This study reveals NDUI in both US and Mexico cases, due to its coarse resolution, serves better as a proxy for urban infrastructure development intensity, rather than human settlement or activity intensity. It would be worthwhile to study the urban form characteristics of cities of interest before applying NDUI estimation. The contrasting nature of cities in the US and Mexico well exemplify the complementary relationship between qualitative and quantitative analysis.

Enhancing generalizability of the model requires more statistical rigor and empirical intuition. A further step might be to specify a more suitable statistical model instead of simple linear regression, especially for cases like Mexico where levels urban development intensity are not so normally distributed. Another stretch goal could be to categorize patterns and stages of urban development in terms of relationships with measures derived from remote sensing imagery. By developing typologies, it could potentially relieve the conflict between model specificity and generalizability.

References

Akiyama, Y. (2012). "Analysis of Light Intensity Data by the DMSP/OLS Satellite Image Using Existing Spatial Data for Monitoring Human Activity in Japan", *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume I-2, XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia.

Guan, Chenghe, Peter G. Rowe (2016). "The Concept of Urban Intensity and China's Townization Poligy: Cases from Zhejiang Province", *Cities*, 55: 22-41.

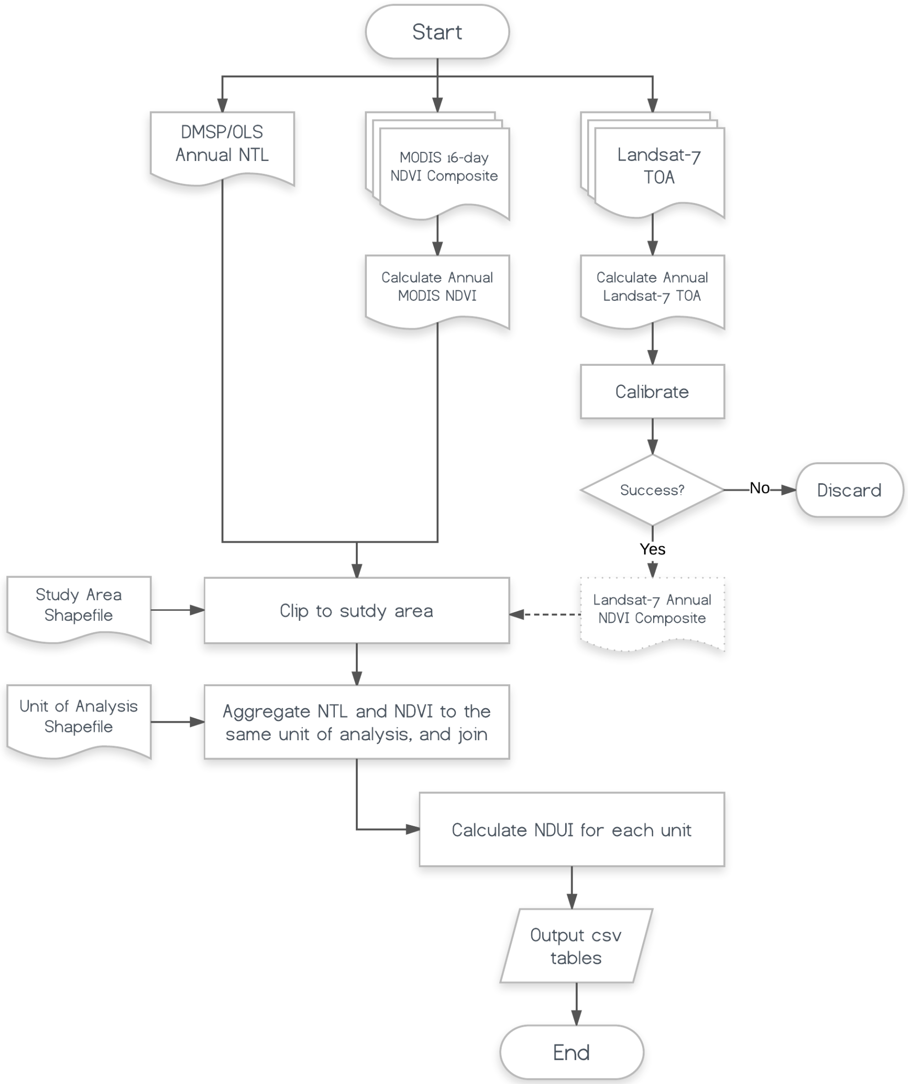
Mellander C, Lobo J, Stolarick K, Matheson Z (2015). "Night-Time Light Data: A Good Proxy Measure for Economic Activity?", *PLoS ONE* 10(10).

Wang, Weixing, Nan Yang, Yi Zhang, Fengping Wang, Ting Cao, Patrik Eklund (2016). "A Review of Road Extraction from Remote Sensing Images", *Journal of Traffic and Transportation Engineering* (English Edition) 2016: 3(3): 271-282.

Zhang, Qingling, Bin Li, David Thau, and Rebecca Moore (2015). "Building a Better Urban Picture: Combining Day and Night Remote Sensing Imagery", *Remote Sensing* 2015(7).

Zhang, Qingling, Christal Schaaf, Karen C. Seto (2013). "The Vegetation Adjusted NTL Urban Index: A new approach to reduce saturation and increase variation in nighttime luminosity", *Remote Sensing of Environment* 129(2013).

Appendix: flowchart of NDUI calculation



Links to Google Earth Engine code:
MSA: https://code.earthengine.google.com/369b98603439e800e5469daac043c7a2
SUN: https://code.earthengine.google.com/0544a0ccf56932b01ace1f78a1abaf1e