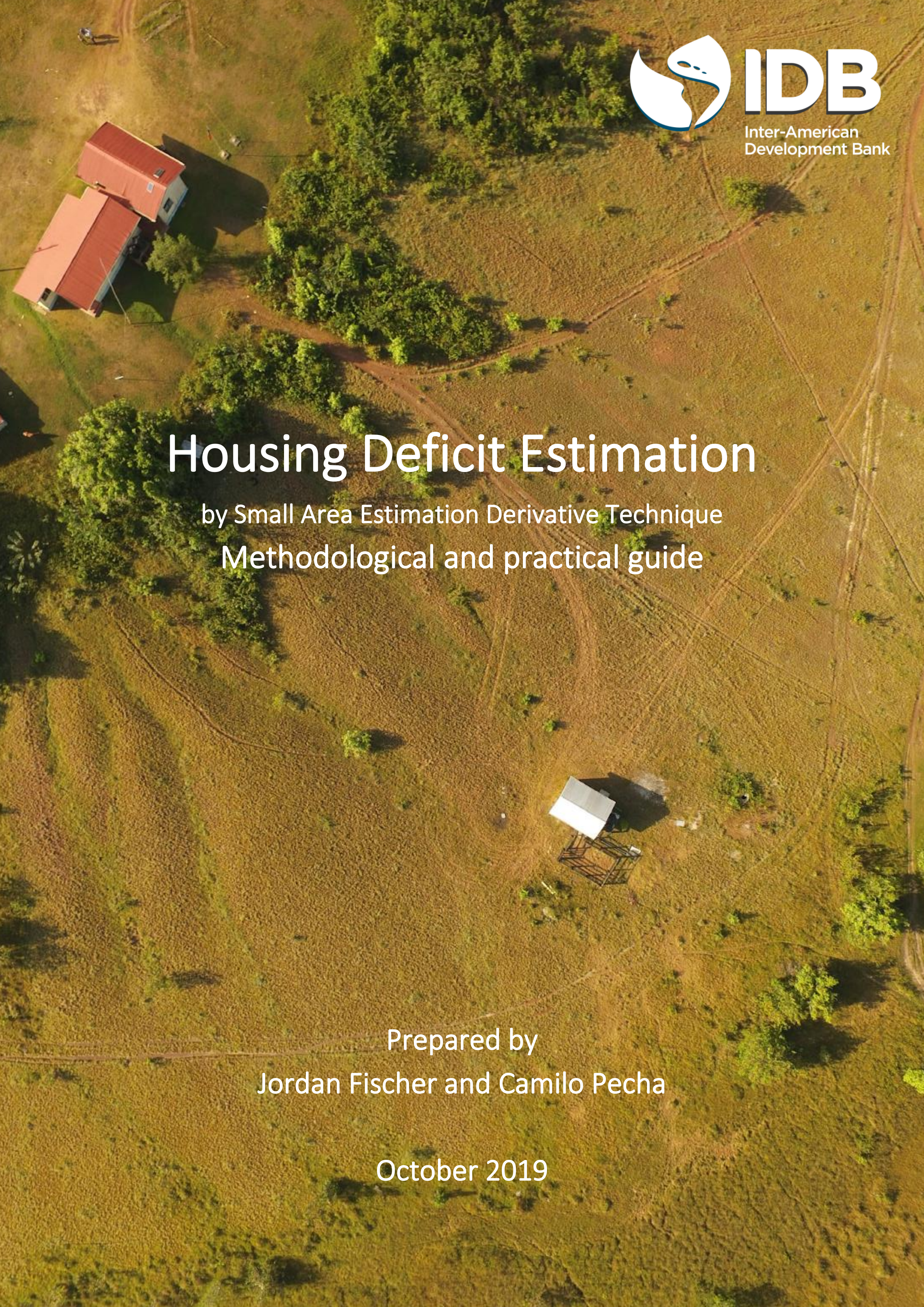


IDB
Inter-American Development Bank

Housing Deficit Estimation

by Small Area Estimation Derivative Technique

Methodological and practical guide

Prepared by
Jordan Fischer and Camilo Pecha

October 2019

Acknowledgment

This report details theoretical methodology and practical guidelines to replicate the estimation of updated housing deficit using census data and satellite imagery. We want to thank Patricio Zambrano-Barragán, Edgar Lemus Pablo, and Muchen Zhu for their support.

TABLE OF CONTENTS

I. INTRODUCTION

Understanding housing deficit is crucial in creating housing policy. Exactly how it is defined varies somewhat depending on the source[1], which may be necessary to take into account cultural and climatological factors, but the standard method for estimating housing deficit relies on census data. Census data is ideal for its thorough data collection, but it is only available once a decade. Housing and infrastructure projects are bound to take place in between censuses, and housing deficit will be an essential input to inform such projects.

Using open source software and the input of context experts, the georeferenced estimation of housing deficit can be automated to produce standardized maps of housing deficit within and across countries, and even, where conditions allow, to now-cast updated estimations using satellite imagery for contexts where recent census data is not available.

This document presents a methodology for the estimation of qualitative and quantitative housing deficits at a highly granular geospatial level and now-casting to estimate housing deficit at the time of the most recently available satellite images. The methodology was tested on data from three countries: Guyana, Trinidad and Tobago, and Peru with encouraging results.

The first section lays out data requirements and important definitions; the second explains the default methodology to be used in cases where a country does not have a nationally prescribed housing deficit estimation methodology; the third will describe in detail how to access and process satellite imagery using QGIS (the same process can be completed using a different software, but QGIS is open-source, free of charge, and more accessible than most alternatives); the next section will explain the now-casting element of the exercise in which a regression is used to 'predict' updated deficit; and finally, the last section will cover mapping and comparing the results of all these calculations.

[1] The CEPAL (Comisión Económica para América Latina y el Caribe) explores definitions and variations in the 1996 publication "Déficit habitacional y datos censales sociodemográficos: una metodología": https://repositorio.cepal.org/bitstream/handle/11362/9781/S9600043_es.pdf?sequence=1&isAllowed=y

II. DATA REQUIREMENTS, DEFINITIONS, AND DEFICIT MEASUREMENT

This section will describe the tabular and geospatial data required to calculate georeferenced quantitative and qualitative housing deficit indicators. The conditions necessary to now-cast updated estimates will be explained in section IV (Using public GIS and satellite information to estimate housing deficit). Assuming access to the necessary data, this section will also outline fundamental concepts of housing deficit estimation and provide reference to prevailing literature on the topic.

Data requirements

The feasibility of this exercise will depend on the availability of georeferenced microdata from a given country's most recent Census, the availability of granular administrative division shapefiles[2], the ability to associate these two datasets, and finally, satellite images of the same resolution, correction, and metrics available for the land mass in question from approximately the year of the census and the most recent year available. It is of crucial importance that the analyst be able to associate the administrative divisions used in the census to the administrative divisions present in the shapefiles. Furthermore, the census data must include georeferenced deficit indicator variables necessary to construct methodologically sound deficit indices. While this may vary from country to country (for example, inclusion of data on flooring but not roofing, or vice versa), the data should include data on construction materials, access to services, and density of habitation.

The R script "Data_prep.R" includes basic steps for data cleaning and standardizing that should be carried out on the census microdata. The exact code used to process the data for Guyana, Trinidad and Tobago, and Peru are contained within the scripts "GUY_data_prep.R", "TT_data_prep.R", and "PER_data_prep.R". These scripts are meant to serve as guidelines only, as data from a new area of study will require subjective scrutiny to determine exactly what steps must be taken to be cleaned and standardized.

The cleaning steps included in "Data_prep.R" address common issues found in census and census-style data. While these steps may be appropriate for some datasets, that does not mean they will be appropriate for all datasets. For example, one very tricky but very common problem found in large datasets is null values. In R, these values show up as "NA", and are excluded from some functions – for this reason, the code uses the function:

```
Table(Census$variable, useNA = 'always')
```

Instead of:

[2] To download shapefiles for administrative division of most countries, visit http://www.diva-gis.org/gdata where you can select the country and download various type of GIS data.

```
Table(Census$variable)
```

Null values represent unanswered questions. Especially in cases where the census includes a code for the respondent declining to answer, it is worth exploring why a census data engineer did not include any answer to this question. This can happen because the question was not asked – in which case, why was the question not asked? Do census-takers consider certain types of households unqualified to answer this question? Why? The data dictionary will not explain null values by nature, which means it is up to the analyst to delve deeper into the situation. Where possible, the best course of action is to follow up with the statistical institute responsible for producing the data in question.

Housing Deficit Definitions

The housing deficit standards and definitions laid out in this section are based on international literature on housing deficit. Sources of information like the UN HABITAT and DANE will give important insights in terms of definitions and how to measure those issues. Some countries' statistical institutions publish a specific set of definitions and methodology for calculating housing deficit within that country, as is the case with Peru[3], and international bodies such as CEPAL, MINURVI, and UN Habitat provide general guidelines in the absence of a country-specific methodology. In countries where no national methodology exists, this exercise will use a methodology based on UN Habitat and MINURVI guidelines, and the mthodology used by Colombia's Departamento Administrativo Nacional de Estadística (DANE, the National Statistics Office)[4], explained in the "Deficit indicators" section below.

Quantitative housing deficit

MINURVI[5] has defined quantitative housing deficit as "…the amount of new housing that is needed so that all households that need housing have a decent space that allows them to develop their reproductive, family and social activities."[6] With this, a basic indicator for quantitative deficit will refer to the difference between the total number of households and the number of occupied housing units.

Qualitative housing deficit

On the other hand, MINURVI has defined qualitative housing deficit as: "…dwellings that have deficiencies in the structure of the floor, space, availability of public services and, therefore, [require the provision of] public services, improvement or expansion of the housing unit"[7]

[3] Instituto nacional de estadística e informática de Perú (https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1442/cap13.pdf)
[4] DANE 2009
[5] General Assembly of Ministers and High Authorities of Housing and Urban Development of Latin America and the Caribbean
[6] MINURVI cited by UN HABITAT (2015)
[7] MINURVI cited by UN HABITAT (2015)

More specifically, ECLAC[8] has defined three aspects to take into consideration to measure qualitative deficit: 1. Materials of walls, roof, and floor, 2. Overcrowding (number of people per room at home), and 3. Access to utilities (potable water, sewerage, and electricity)[9].

Deficit indicators

In the absence of a nationally prescribed housing deficit estimation, a methodology can be derived using examples from other countries and guidance from the studies of relevant international organizations. Total housing deficit is constructed based on total dwelling needs in terms of quality and quantity. To calculate this, it is important to determine the number of households living in suboptimal conditions in terms of building quality and overcrowding. Based on questions commonly found in household Census data, a series of indicators were chosen to best represent housing deficit based on the information available, with certain categories considered 'inadequate', i.e., indicating a household in housing deficit.

The categories considered adequate or inadequate will vary in different climatological and cultural contexts. For example, indigenous communities may use traditional building materials for primarily cultural reasons; a colder climate will necessitate more insular building materials than a warmer one; and cultural considerations of 'crowding' vs. 'overcrowding' may change over decades and across regions. It is therefore imperative that this methodology be reviewed carefully and modified according to specific contexts and available information before any replication.

Housing deficit for both Guyana and Trinidad and Tobago was calculated using the following indicators, which provide valuable insight into households' shelter needs. A table detailing variables, categories, and values assigned to create the housing deficit indices can be found in the Annex.

Wall material quality: The Census questionnaire includes the variable 'Main wall material' wherein materials like "Makeshift", "Galvanized", "Troolie palm" are considered inadequate.

Roof material quality: The Census variable 'Main roofing material' indicates insufficient quality by materials like "Makeshift", "Sheet metal", "Troolie palm", and others.

Cohabitation: The Census also identifies cases of cohabitation, defined as two or more households sharing the same dwelling.

Overcrowding: This indicator is determined by the number of habitants per room. Where it is possible to differentiate between urban and rural households, it may be pertinent to apply

[8] Economic Commission for Latin America and the Caribbean
[9] ECLAC cited by UN HABITAT (2015)

different rules regarding overcrowding in urban and rural settings. In the case of Guyana, the census data does not include any indicator articulating this distinction, and thus more than 3 persons per room is considered overcrowding across the country.

Acute overcrowding: More than 5 people per bedroom is considered acute overcrowding for all parts of the country.

Availability of utilities: The Census contains information regarding households' main source of lighting (where not having access to electricity is considered deficit), water (deficit will be defined by absence of piped water and access to water from spring/river/pond), sewerage (deficit is defined by not having WC connected to a sewerage), and garbage disposal (dumping on land, burning, dumping/throwing into river/sea/pond is considered to be deficit).[10]

Estimation process.

Housing deficit is treated as a binary variable for this estimation – a household is either in deficit (indicators aggregate to a number greater than zero) or it is not (indicators aggregate to zero). Once the deficit situation for each household has been defined, the scores of all indicators can be aggregated as follows in order to determine the size of quantitative, qualitative, and total housing deficits.

To estimate the **Quantitative Deficit**, the following characteristics can be aggregated:

1. Cohabitation: more than one household per dwelling;
2. Acute overcrowding: more than 5 people per bedrooms;
3. Inadequate dwelling: The census questionnaire uses the variable 'Type of building' to identify the type of construction in which a dwelling is located. Households living in buildings type "Other" or "Community Service" are identified as not residential, and therefore qualitatively deficient;
4. Inadequate unit: The variable 'Type of dwelling unit', where the categories "barracks", "makeshift", and "other" are identify households in deficit, is used to provide additional insight into the adequacy or inadequacy of dwelling types.

To estimate **Qualitative Deficit**, the following characteristics can be aggregated:

1. Dwellings with low quality wall materials;
2. House with low quality roofing material;
3. Overcrowding: more than three people per room (may vary by urbanization level);
4. Access to utilities:
    a. piped water,
    b. sewerage,

[10] In particular, since there is no way to identify rural/urban areas from census data, it was important to incorporate potential indication of deficits for water access using an aggregation of urban indicator (lack of access to piped water) and rural indicator (main access to water from spring/river/pond).

c. electricity,
d. garbage disposal.

The estimation of Total Housing Deficit is calculated by summing up the quantitative and qualitative deficit. The share of population in housing deficit is equal to the number of households in deficit divided by the total number of households in the country. The indicators will then be estimated by a relatively granular geographical location (often second administrative division) and shown in a map.

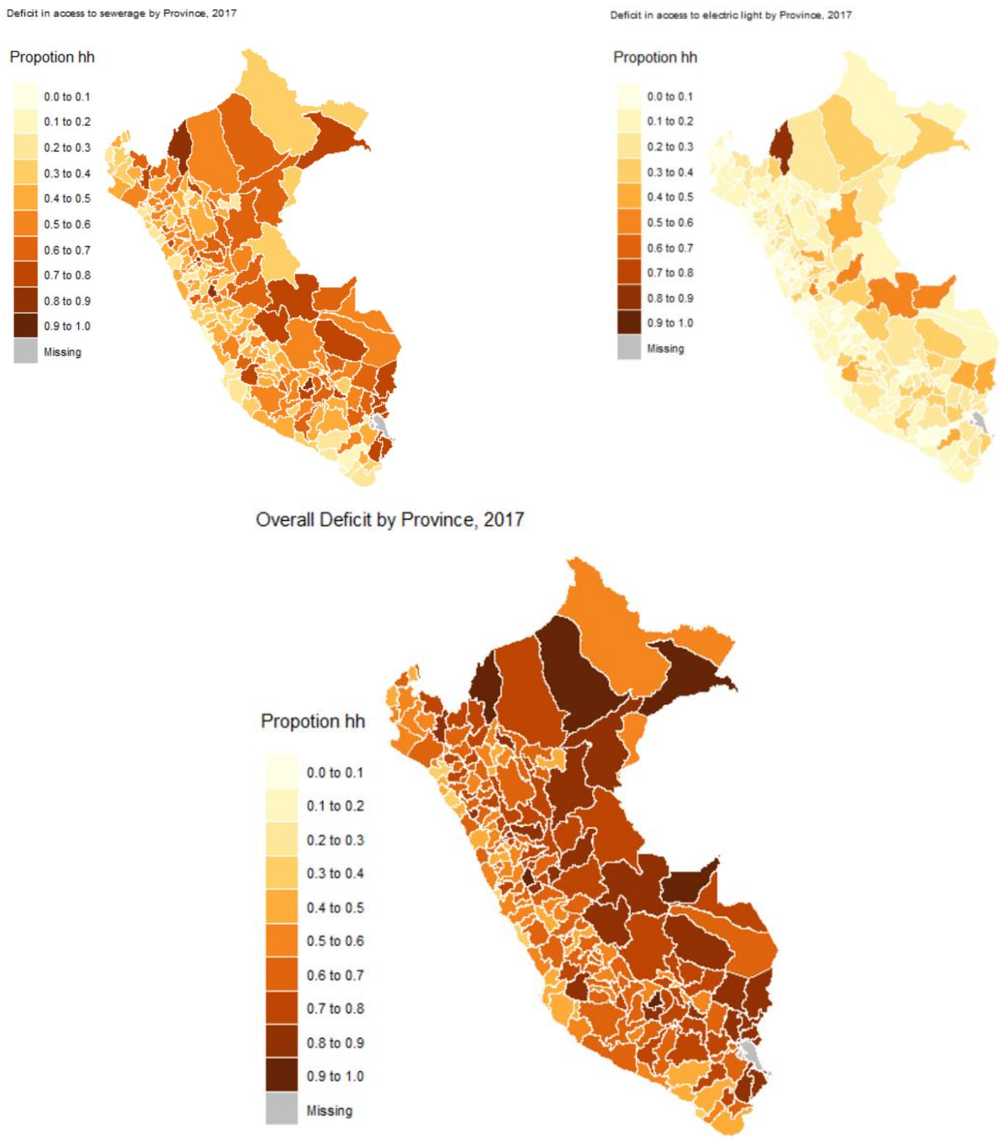


Figure 4: 2nd level administrative division maps of housing deficit drivers, overall deficit in Peru, 2017
Top left: sewerage access by province; Top right: electricity access by province; Bottom: total housing deficit

Throughout the region, qualitative deficit is widely considered to affect more households than quantitative deficit[11]. This was found to be the case in Guyana, Trinidad and Tobago, and Peru for the most recent census year available (2012, 2011, and 2017, respectively). In Trinidad and Tobago, overcrowding was also a major driver. This suggests that electric light is tied closely to overall housing deficit in these three diverse countries, an important assumption that precedes the nocturnal luminosity regression in section V.

[11] For details, see the IDB's and the World Bank's "The Quality of Life in Latin American Cities": http://documents.worldbank.org/curated/en/325571468045577440/pdf/544310PUB0EPI01OX0349415B01Public10.pdf

III. INDICATORS’ ESTIMATION BASED ON DECISION RULES

To estimate the indicators for housing deficits, a set of rules is defined. In particular, the following rules will help to identify the deficit condition of a household through some conditions related to the variables described above. The following are the rules designed to measure housing deficit.[12]

The rules for cohabitation and acute overcrowding are used to measure the size of quantitative deficit. In principle, a household will be in quantitative housing deficit if at least one of the following rules is satisfied:

$$Cohabitation = \begin{cases} 1 \; if \; household \; share \; dwelling \\ 0 \; otherwise \end{cases}$$

Meaning that, in the dataset, a household that shares its dwelling with other(s) household(s) will has a 1 and 0 otherwise.

$$Acute \; overcrowding = \begin{cases} 1 \; if \; more \; than \; 5 \; people \; per \; bedroom \\ 0 \; otherwise \end{cases}$$

Thus, taking the ratio between the household size and the number of bedrooms, a household which ration goes above 5 people per room will be assigned with 1 and 0 otherwise.

The size of the quantitative housing deficit in region *j* will be determined by the fraction of households that comply with any of the mentioned rules relative to the total number of households in region *j*:

$$QuantiHD_j = \frac{\sum_i (Cohabitation_{ij \; where=1} \; OR \; Acute \; overcrowding_{ij \; where=1})}{total \; households_j}$$

Where the numerator is defined by the sum of all of households *i* in region *j* that are in cohabitation **or** in acute overcrowding.[13] The denominator is the total number of households in region *j*.

[12] Table A1 in the annex presents all the variables used in the estimations, the corresponding categories, and values assigned in terms of deficit.
[13] In the case a household presents cohabitation AND acute overcrowding, the indicator will take a value of one for qualitative deficit.

As explained in the previous section, qualitative deficit is measure using quality of walls and roof materials, and access to public utilities. The rules to measure the size of qualitative housing deficit are the following.

$$Walls = \begin{cases} 1 \; if \; main \; material \; is \; \text{“}Makeshift\text{”}, \text{“}Galvanized\text{”}, \text{“}Troolie \; palm\text{”} \\ 0 \; otherwise \end{cases}$$

$$Roof = \begin{cases} 1 \; if \; main \; material \; is \; \text{"}Makeshift\text{”}, \text{“}Sheet \; metal\text{”}, \text{“}Troolie \; palm\text{”} \\ 0 \; otherwise \end{cases}$$

In terms of dwellings' structure, measures of the quality of construction materials will give an idea of dwellings' durability. In particular, Guyana's census incorporates questions regarding wall and roofing materials, however no information regarding floors is captured.[14] The rules are defined so that those households living in dwellings with non-durable materials are considered to be in housing deficit and therefore are valued as a 1 in the data, and 0 otherwise.

$$Overcrowding = \begin{cases} 1 \; if \; more \; than \; 3 \; people \; per \; room \\ 0 \; otherwise \end{cases}$$

With respect of the available space of a dwelling to develop social and biological activities, a less stringent measure of acute overcrowding is used and is defined by the ratio between the household size and the number of rooms.[15] If the ratio goes above 3 people per room, the household is consider to be in a situation of overcrowding and therefore it is assigned a 1 in the data, or 0 otherwise.

With respect to access to basic services, the availability of electricity, piped potable water, sewerage, and/or garbage collection help to determine if a dwelling has sufficiently habitable conditions.

$$Lighting = \begin{cases} 1 \; if \; no \; access \; to \; electricity \\ 0 \; otherwise \end{cases}$$

The most common way to identify the availability of electricity is through information related to main sources of lighting in the dwelling, collected by the census. A household is considered in deficit if it lacks access to electricity by private or public means, so it is assigned a 1 for lighting; those that have access to electricity are assigned a 0.[16]

---

[14] Internationally speaking, some methodologies incorporate the quality of outer walls materials as part of the indicators or rules for quantitative deficit calculations. If a weak dwellings' structure is considered to be important for quantifying the replacement need of units, the quality of walls' material should be considered as part of quantitative deficit. In the case of Guyana, due to the absence of information on floors' material, the quality of walls and floors were used as measures of housing material quality as recommended by CEDLAC (UN-HABITAT 2015)

[15] Note that this rule takes into consideration number of rooms *not including* bedrooms. In particular, Guyana's census gathers information on number of rooms other than bedrooms, and number of bedrooms separately. If it is not the case, the use of number of rooms for both acute and not acute overcrowding could be acceptable.

[16] Currently, Guyana's census incorporates information related to solar or inverter access to lighting which was considered as positive access to electricity.

$$Water = \begin{cases} 1 \; if \; no \; access \; to \; piped \; water \; and/or \; main \; acces \; is \; spring, \; river, \; or \; similar \\ 0 \; otherwise \end{cases}$$

Access to quality source of water is considered in deficit when dwelling has no access to piped water or when its primary water source is identified as spring, river, or similar; sources that would not warrant sufficient water quality for human consumption (DANE 2009).

$$Sewerage = \begin{cases} 1 \; if \; WC \; is \; not \; connected \; to \; sewerage \\ 0 \; otherwise \end{cases}$$

Adequate access to waste water management is considered in deficit when the type of toilet facility does not have a proper connection to sewerage (either sewerage system or septic tank. In those cases where toilet facilities are not connected to sewerage, the household is assigned a 1, and 0 otherwise.

$$Garbage = \begin{cases} 1 \; if \; no \; garbage \; collection \\ 0 \; otherwise \end{cases}$$

Finally, access to proper collection of waste or garbage could be associated with optimal habitational environmental conditions.[17] Deficit in garbage collection could be associated with a lack of proper housing environment, affecting suitable living conditions. Households reporting no garbage collection (including burning, burying, dumping in rivers, etc) will be assign 1 in the dataset and 0 otherwise.
Following these rules and households' compliance, the share of them under qualitative deficit, relative to the total number of households at the regional level is defined as:

$$QualiHD_j = \frac{\sum_i (W_{ij \; where=1} \; OR \; R_{ij \; where=1} \; OR \; O_{ij \; where=1} \; OR \; L_{ij \; where=1} \; OR \; U_{ij \; where=1})}{total \; households_j}$$

Where the numerator is defined by the sum of all households *i* in region *j* that indicate they have walls OR roofs made of non-durable materials OR are living in overcrowding, OR lack access to electricity, OR access to utilities (for the sake of equation space, $U_{ij \; where=1}$ incorporates adequate access to water, OR sanitation, OR garbage collection). A household assigned 1 for any of these indicators, will be considered in qualitative housing deficit. The denominator is the total number of households in region *j*.[18]

[17] This particular information is collected by the Guyana's census as one of the questions in the facilities available for use. Other potential variables related to conditions outside dwelling could be distances to work, nearest school, hospital or health center, and availability of public streetlighting.

[18] Note that, in this case, neither qualitative nor quantitative is divided by a regional characteristic like urban or rural, costal or interior, given that there is no specific variable available in Guyana's census to discriminate on these criteria. In the case there is a way to determine regional characteristics like cultural division, or rural or costal classifications, and the rules' set is available, what is called here "region *j*" can be adjusted to the selected identification.

These rules are put into code form in the “indicators.R” script. This same script includes code that can be used to plot the results directly within R by merging it with the administrative division shapefiles. If the user prefers the graphics of QGIS, the merged file, now called a *SpatialPolygonsDataFrame* (“GUY_map” for Guyana, “PER_map” for Peru, or “TT_map” for Trinidad and Tobago) can be saved as a shapefile using the *writeORG* function.

IV. USING PUBLIC GIS AND SATELLITE INFORMATION TO ESTIMATE HOUSING DEFICIT

Access to proper public amenities like public lighting, adequate roads, schools, and hospitals, provide additional insight into the housing deficit. Apart from a household's own access to electricity, piped water, and adequate sewerage systems, a household's neighborhood also comprises an important part of its conditions. Lack of public lighting could indicate risks related to crime, lack of schools would imply less access to education, and poorly paved roads would cause an increase in travel times from home to any destination, all of which reduce the quality of living for residents. These neighborhood indicators can provide a map across which to project data from contexts where conditions are known, to contexts where conditions are yet unknown. This section explores using public lighting based on satellite imagery to predict housing conditions for years when census data is unavailable.

Recently, satellite imagery has been used as a rich source of information to study social issues related to poverty and inequality. A strong relationship has been found between night lights and the economic dynamic measured by GDP (NOAA, 2010) and poverty (Engstrom, et.al, 2017); satellite information has been found useful in the detection of slums (Divyani, et.al, 2016) and prediction of housing prices (Bency, et.al, 2017).

To estimate an 'update' of the housing deficit in 2019 based on the most recent available census data, this note will demonstrate how the methodology designed by Henderson et.al, 2012, using night lights to estimate economic growth, can be successfully applied to estimate qualitative housing deficit in Guyana and Peru. The process incorporates the estimations of qualitative housing deficit based on the most recently available Census data (from section IV) and satellite data on nighttime lights.

The reasoning behind the use of night light data is that the level of luminosity detected by sensors could be associated with the level of urban development (Muzzini, et.al, 2016) in each region. Qualitative housing deficit, found to be the primary driver of total housing deficit, has been measured with rules derived from the availability of public services or utilities, variables that correlate with the presence or absence of public infrastructure like street lighting. Exploiting this relationship between dwellings' access to services and luminosity would help to understand the behavior of qualitative housing deficit in areas or periods of time where and when no statistical microdata is available. In this study, since the most recent census in Guyana was implemented in 2012, the idea is to use nightlight data to estimate the size and distribution of qualitative deficit in 2019 using data from 2012.[19]

[19] The nighttime lights data is not the only data that could be used for such a prediction. Depending on availability, it is possible to also include data on availability of schools and hospitals (from OpenStreetMap) to determine access to services that correlate with health and education levels. In this case, reliable OpenStreetMap data was not available from 2012 in Guyana.

Checking if the methodology will be appropriate

While some countries' housing deficit follows the patterns laid out above, other countries' deficit may be driven by things like cohabitation of multiple families in a single cramped – but well-lit – dwelling. Some areas may be too densely populated and well-lit for the model to differentiate between a dwelling providing electric lighting and the unlit dwelling next to it. Similarly, more densely populated areas also facilitate things like households illegally tapping electric grids that are meant to feed only their paying neighbors. This means that the model may not be appropriate for use in areas with higher population density and urbanization rates or that its accuracy may suffer under such circumstances (this phenomenon is well documented in nocturnal luminosity methodologies, including the World Bank's much-cited Poverty from Space).

To determine if a nocturnal luminosity regression will be appropriate for estimating updated housing deficit in a given area, we run two tests on the data.

Association Analysis

An association analysis is applied to indicators causing deficit to highlight patterns among sub-optimal conditions. This rules-based machine learning technique reveals relationships between variables, i.e., how likely it is that a household exhibiting one characteristic will also exhibit another. The main metric to look at in an association analysis is called Lift. Lift shows the relationship between the left-hand side (LHS) and the right-hand side (RHS), i.e., how likely it is that the right-hand side will occur in a case where the left-hand side is already occurring.

- Lift > 1 indicates that LHS and RHS are dependent
- Lift = 1 indicates that LHS and RHS are independent
- Lift < 1 indicates that lhs and rhs replace each other

In all three countries where we tested the methodology, access to electricity/access to electric light was strongly associated with other drivers of qualitative and overall deficit.

- Guyanese households lacking access to electricity are 64% more likely to also lack adequate sewerage;
- In Trinidad and Tobago, overcrowded households with inadequate water and sewerage access are twice as likely to lack access to electricity
- Peruvian households that experience overcrowding and inadequate walls and sewerage are 75% more likely to also lack access to electricity

Also, information from satellite imagery that contains data on the reflectance of roads could provide information about road characteristics like pavement status such that the density of unpaved/paved roads at the region level would provide insight on the level of public infrastructure and how this correlates with qualitative deficit. Again, in the case of Guyana adequate information was not identified for this study, but could be incorporated in future iterations, or in other contexts.

This consistent association supports the idea that nocturnal luminosity (which is overwhelmingly driven by access to electric light) will be a strong indicator of qualitative and overall housing deficit.

Coefficient of Variance

In order to determine the statistical accuracy of the indicators, a coefficient of variation (CV) is calculated. The CV will capture the indicators' dispersion so that the researcher can define how informative the indicator is in statistical terms. The CV is defined as the standard deviation of the indicator divided by indicator's mean multiplied by 100. The lower the CV, the more precise the estimation is. DANE, in Colombia, has defined a scale to determine how informative or precise a housing indicator is:

Table 2. Definition of accuracy of indicators based on CV value.

Qualification	Value of coefficient
Good	CV<7.5
Acceptable	7.5≤CV<15
Low precision	15≤CV<30
Bad	30≤CV

Source: Original construction based on DANE (2009)

The following table presents the value of coefficient of variation for housing deficit drivers in the three test countries. Noticeably, the CV for access to electric light is higher (less precise) in denser Trinidad and Tobago than it is for Guyana and Peru, where populations are more spread out.

Table 3. CV values for housing deficit drivers by test country.

	CV_walls	CV_overc	CV_water	CV_sewar	CV_light	CV_cohab	CV_roof	CV_garba	CV_acute
Guyana	1.9	11.8	26.4	36.8	11.1	3.1	3.0	57.4	3.3
T&T	1.7	15.8	7.6	24.3	15.5	1.2	0.5	NA	23.9
Peru	3.6	7.1	18.6	34.3	12.0	11.9			

Source: Original construction

Less variation in access to electric light will negatively affect the confidence of the regression results.

Nocturnal luminosity data from satellite imagery

In order to predict levels of qualitative housing deficit, this project utilizes nightlight data extracted from satellite imagery (luminosity regression). Where the outcomes of the association analysis indicate that access to electric light is strongly associated with other drivers of housing deficit in the area of study, and that the coefficient of variance indicates at least acceptable precision, we can proceed with the reasonable assumption that the data will be appropriate for a luminosity regression.

This section will explain the procedure to obtain, extract, and process satellite imagery information to produce the statistical data used in the estimation of housing deficit. At the end of this section the user should be able to obtain information in numerical format to be used in statistical packages in combination with other information, like census microdata, to do estimations.

Accessing the satellite imagery

The data used for this project was obtained from the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band. The VIIRS is a sensor designed by Raytheon Company and it has been installed in two satellites: Suomi NPP and National Oceanographic and Atmospheric Administration (NOAA)-20.[20] The VIIRS project cleans the imagery by filtering lights from lunar reflectance, clouds, etc. The data is then averaged monthly and annually, and the resolution is 15 arc-seconds (approximately 450 meters per pixel).[21] The following are the steps to download the data from Earth Observations Group (EOG) VIIRS data web page.

First, visit https://eogdata.mines.edu/download_dnb_composites.html and scroll down to find the download site as shown in figure 1. This data is stored in a panel structure with the world divided in 6 tiles, each tile with information from March 2012 to present.

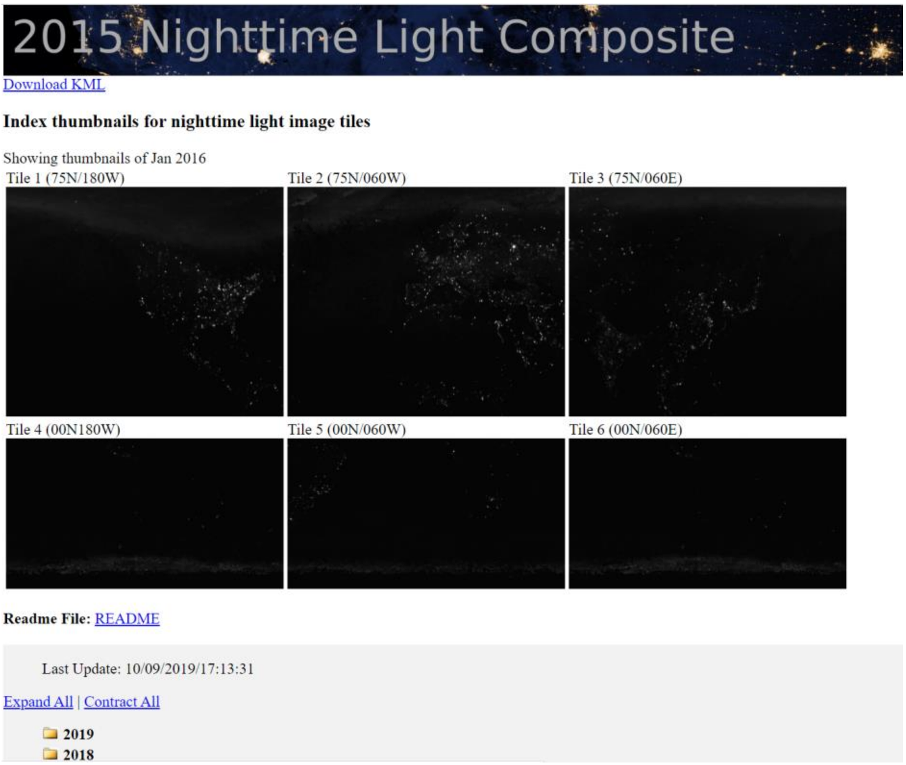


Figure 1: EOG download site for VIIRS data

[20] https://en.wikipedia.org/wiki/Visible_Infrared_Imaging_Radiometer_Suite
[21] For more information, please visit: https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html

The tiles selected should correspond to the part of the world where the area of study is found. Trinidad and Tobago is contained within Tile 1; Guyana is captured in both Tile 1 and Tile 2; and Peru is contained within Tile 4. From the menu located in the lower-left corner, select the folder from the base year (year of the last census)[22] by clicking on it. Select "Monthly", and then the earliest available month from the base year. For the tile(s) of interest, select the files named "VCMCFG" containing zipped folders with the tiff images, as shown in figure 2. The same process is followed for the most recent month available.

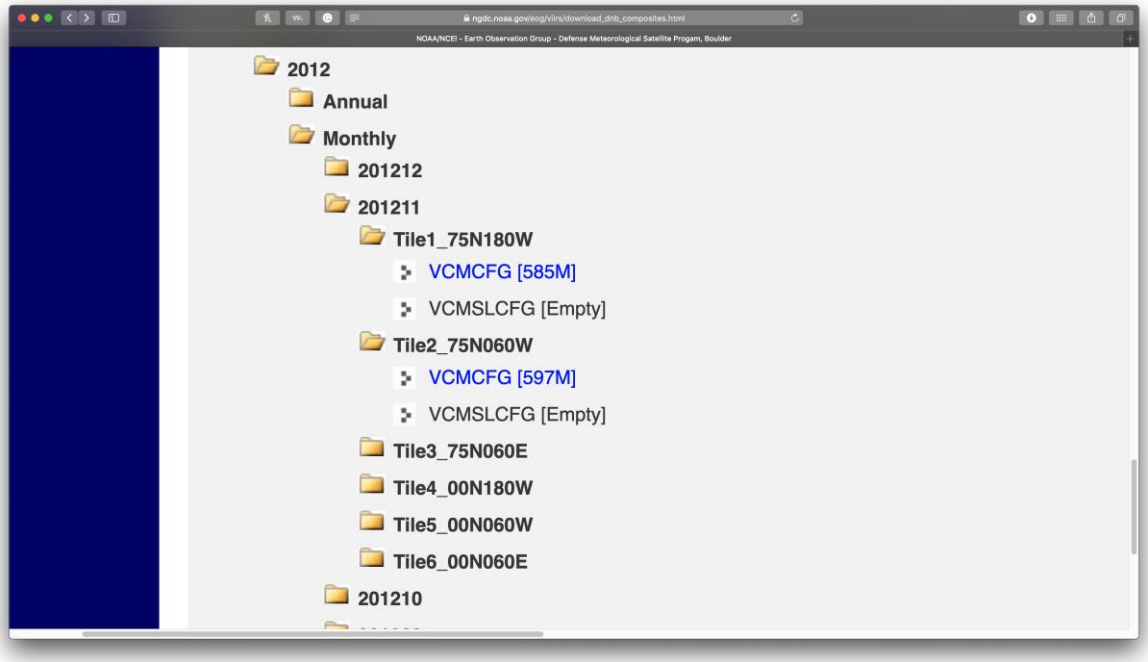


Figure 2: Files selection in EOG VIIRS site

Each zipped file is greater than 0.5GB; when unzipped, each folder's contents will occupy over 2GB of disc space, so it is important to have in mind the amount of free hard disc space needed for processing.

The following sections ('Opening data in QGIS', 'Merging geotiffs', 'Clipping the merged geotiff', and 'Converting the raster data to numerical data'), should be carried out first on the file(s) from the base year (year of the last census), and then on the file(s) from the most recent year. The following steps will show how to convert raster data into information compatible with other datasets like census microdata: the result for all of Section IV should be two .csv files – one for base year luminosity data and one for the most recent year's luminosity data.

[22] The most recent census taken in Trinidad and Tobago was in 2011; for this exercise, 2012 data was used.

Opening data in QGIS

The first step is to open the data in QGIS. Each downloaded folder should contain 2 .tif files – select the file ending in '…avg_rade9h.tif' for each tile since this is the file corrected by radiation. Select the files to be open in the main window as shown in figure 3 by selecting the "Layer -> add new -> Add raster layer" from the menu, or press the icon from the left side of the main window, or drag the file from the list of files on the left panel of the window into the primary panel to the right.[23] Second, we load the shapefile data to verify that the full the area of study is contained within the raster data we have loaded. To do this, open the shapefile for the level 0 of administrative division selecting "Layer -> add new -> Add vector layer" from the menu, or simply press icon.

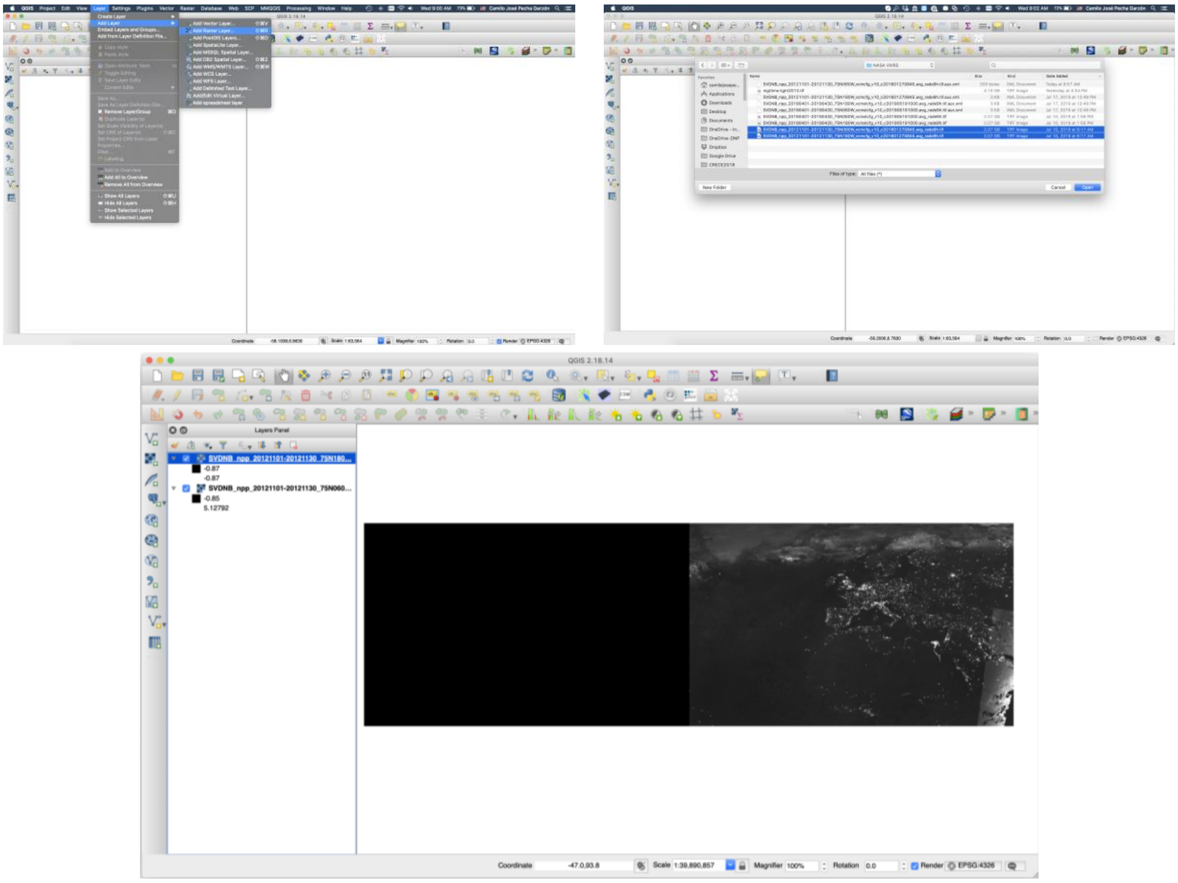


Figure 3: Opening raster files in QGIS

Figure 4 presents how to open the Guyana's level 0 of administrative division shapefile (adm0, country boundaries).

[23] This tutorial is based on the QGIS 2.18.14, the most stable version of the software for Mac-OS users.

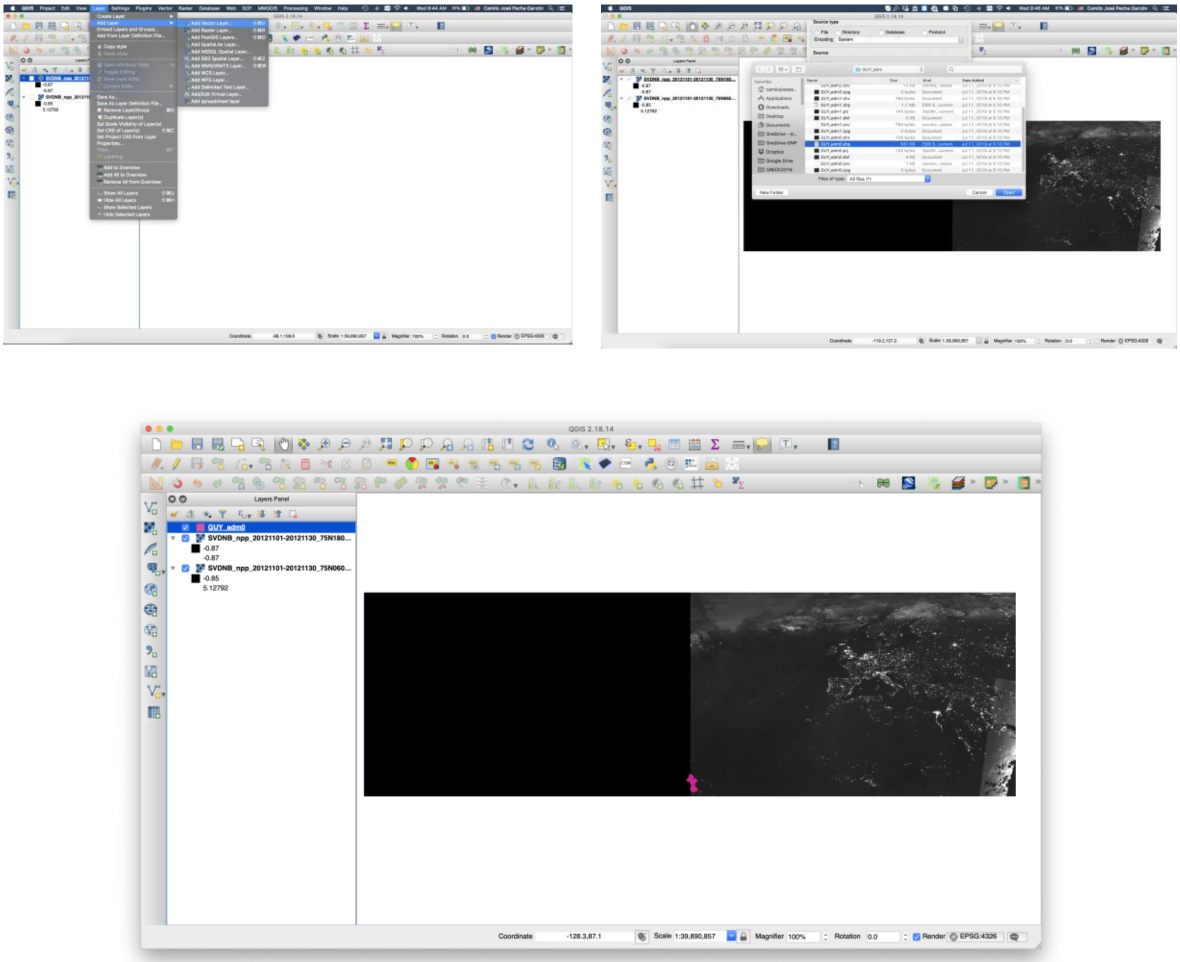


Figure 4: Opening vector files (shapefiles) in QGIS

Merging geotiffs (only necessary for some countries)

In cases like the country of Guyana, where the country's territory falls across multiple tiles, the two original tiles must be merged into a single Raster file. **This step is not necessary for countries that fall entirely within a single tile.** To merge two raster files, select "Raster -> Miscellaneous -> Merge…", then in the prompted window select the raster files to be merged as the "input files" field and set the output file's name and location, then click "OK". A new raster file is produced and loaded into the canvas as shown in figure 5. This step is computationally expensive and most computers will take time to complete it.

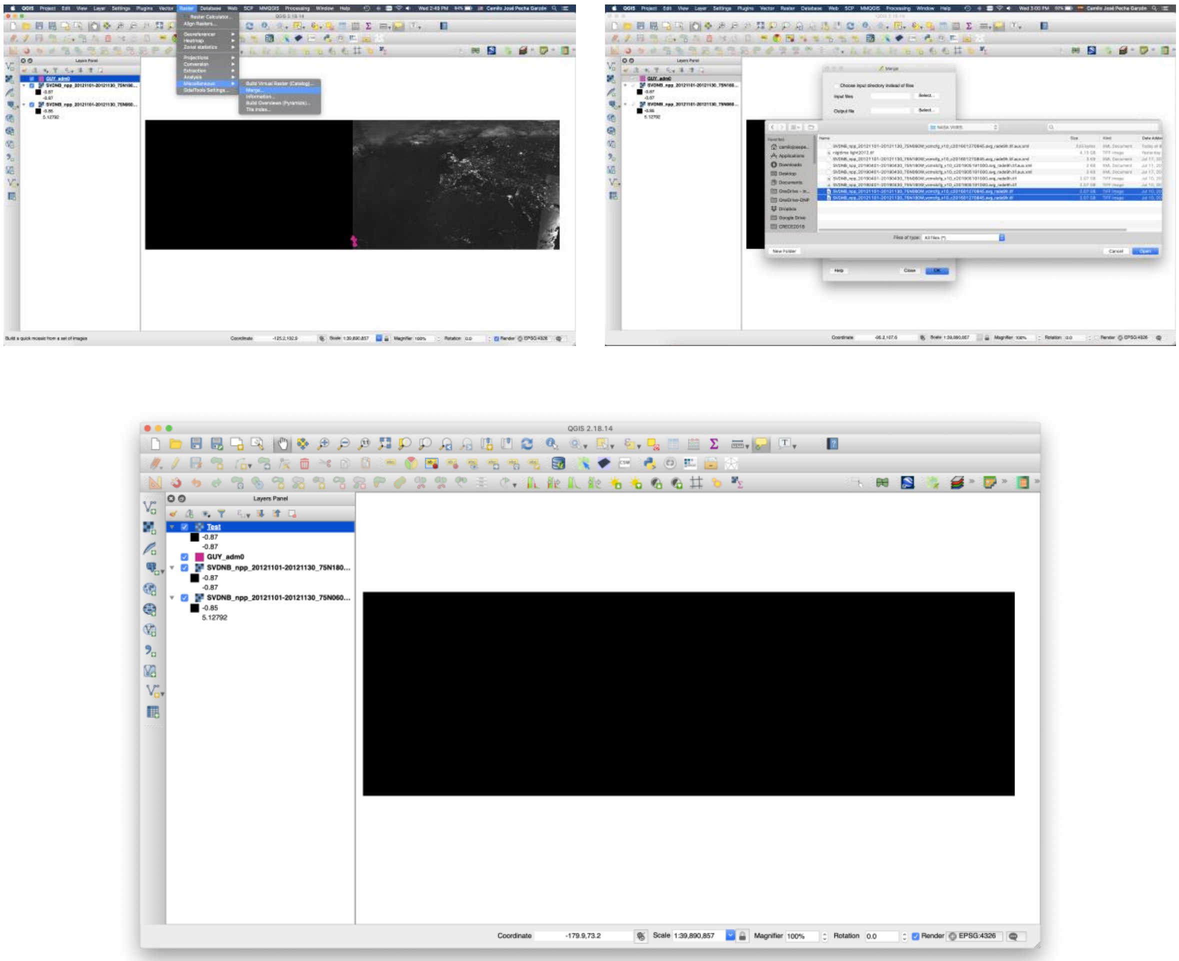


Figure 5: Merging Raster files in QGIS

If there are errors, check the box at the bottom of the prompt window. This box will show the python code that is used to run this function within QGIS. The code used in this case is shown in the box below:

```
gdal_merge.py -of GTiff -o /Users/…/Downloads/merged.tif /Users/…/Downloads/ SVDNB_npp_20121101-20121130_75N180W_vcmcfg_v10_c201601270845\SVDNB_npp_20121101-20121130_75N060W_vcmcfg_v10_c201601270845.avg_rade9h.tif /Users/…/Downloads/transaccion-7.pdf
```

Clipping the geotiff

Once a single raster containing the entire area of study is loaded into the QGIS canvas, we would next start to process it. However, the large size of raster files makes processing computationally expensive and we need only the data for area of study, so we use the shapefile for the country boundaries (adm0) to extract information only for that area.

To “cut” or select only the information inside the area of study, select “Raster -> Extractions -> Clipper…” from menu. In the prompted window, select the merged file in the “Input file (raster)” field, define an “Output file” that will be the file containing the information for only Guyana. In the “Clipping mode” section, select “Mask layer” and select the area of study’s full administrative shapefile as "Mask layer" – for an entire country, this will be the Adm0 shapefile, or the full territory of the country with no subdivisions. Finally, check mark "Crop the extent of the target dataset to the extent of the cutline" and click ok. The process is shown in figure 6.

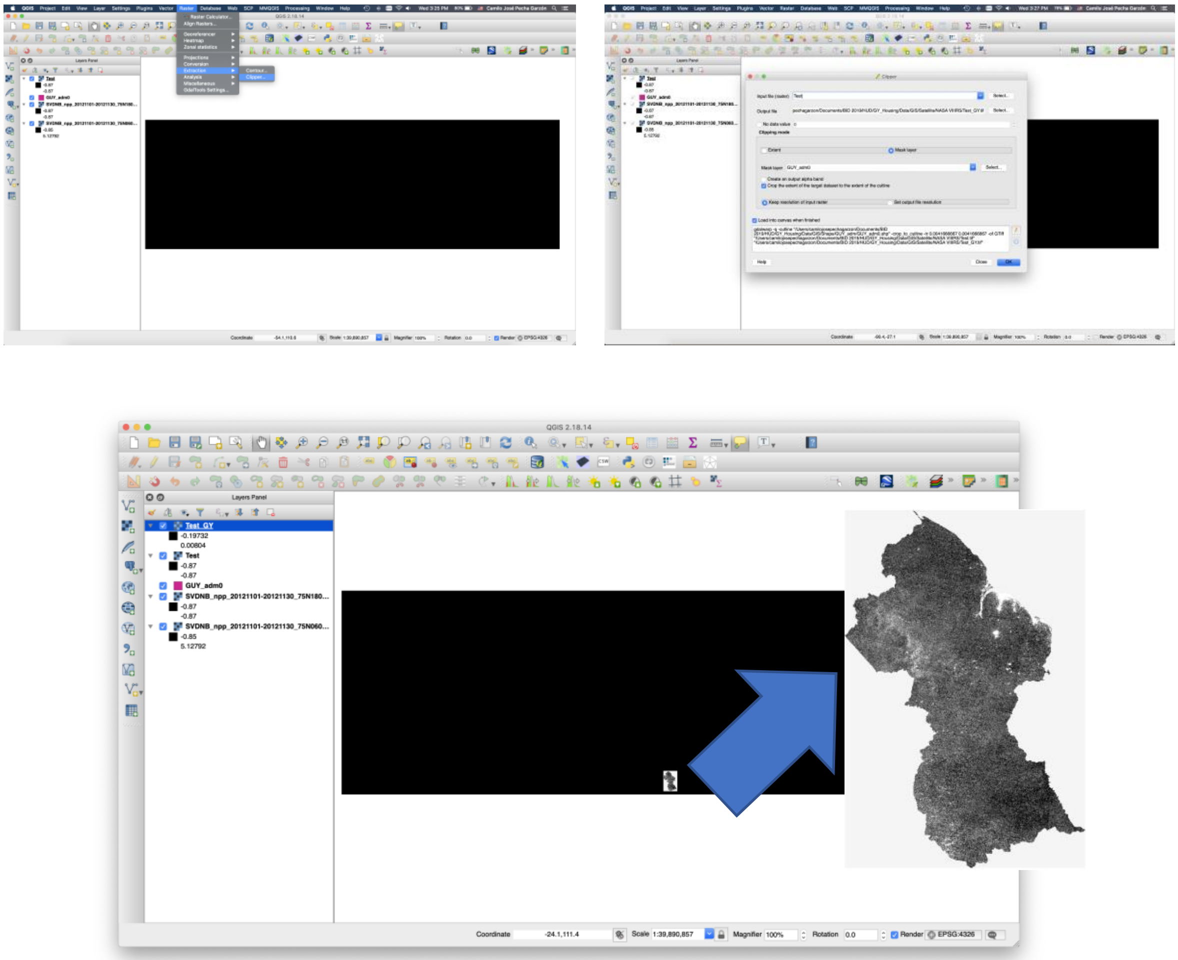


Figure 6: Extracting nighttime light information from raster file and shape file

If there are errors, check the box at the bottom of the prompt window. This box will show the python code that is used to run this function within QGIS. The code used in this case is shown in the box below:

```
gdalwarp -q -cutline "/Users/.../Data/GIS/Shape/GUY_adm/GUY_adm2.shp" -crop_to_cutline -tr 0.0041666667
0.0041666667 -of GTiff "/Users/.../Test.tif" "/Users/.../Test_GY.tif"
```

Converting the raster data to numerical data

The final step is to convert this raster data into numerical data that can be used in a statistical software. This step includes three sub-steps:

1. convert the raster data into vector data (polygonize);

2. combine the new vector data with the shapefile of the administrative divisions to be used in the analysis, recommended Adm2, i.e., level 2 of administrative division (spatial join);
3. convert the new shapefile resulted from previous step into a tabular data version (tabularization).

Polygonization

To convert raster to vector, select "Raster -> Conversion -> Poligonize" from menu. In the prompted window, select as "Input file (raster)" the raster file containing the information extracted in the previous step (figure 6), set the "Output file for polygons (shapefile)", define the name of the field containing the nighttime light data, and check "Load into canvas when finished". Finally, click "OK" as shown in figure 7.

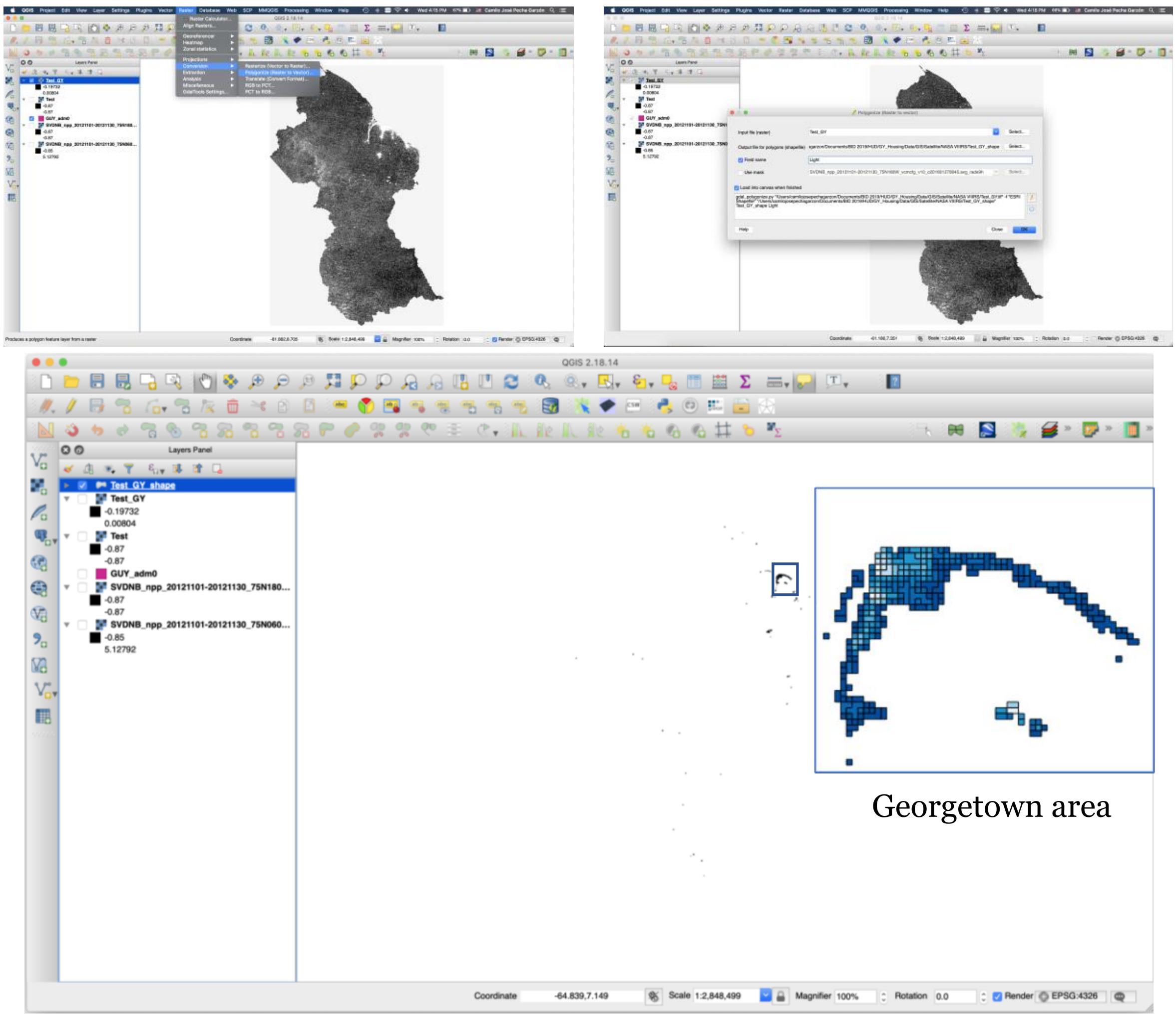


Figure 7: Converting nighttime light information from raster file to shape file[24]

[24] To obtain the shaded version of the map, select the shapefile with double click, "Properties" window will appear. In the left window, select "Style" and in the first dropdown menu select "Graduated". In the dropdown menu select the variable you want to show and in the "Color ramp" chose one of your interest. Select the number of classes you want to plot from the selector at the right-hand side under the "Classes" window. Click on "Classify" under the "Mode" dropdown menu. Finally, click "OK at the bottom left corner. You should obtain a graduated version of the map. This process will be explained in the following section.

Spatial join

The next step is to combine the new nighttime light shapefile with the administrative divisions to be used in the analysis (that correspond with the divisions in the census data), often second-level administrative divisions (Adm2). To do this, first, open the shapefile. To combine datasets, select "Vector -> Data Management Tools -> Join attributes by location" from the menu[25]. In the prompted window, select the administrative division as "Target vector layer", and select the nighttime light shapefile as "Join vector layer", checkmark all the "Geographic predicate" options, in the "Attribute summary" chose "Takes summary of intersecting features", define the output "Joined layer" and click "Run" as shown in figure 8.

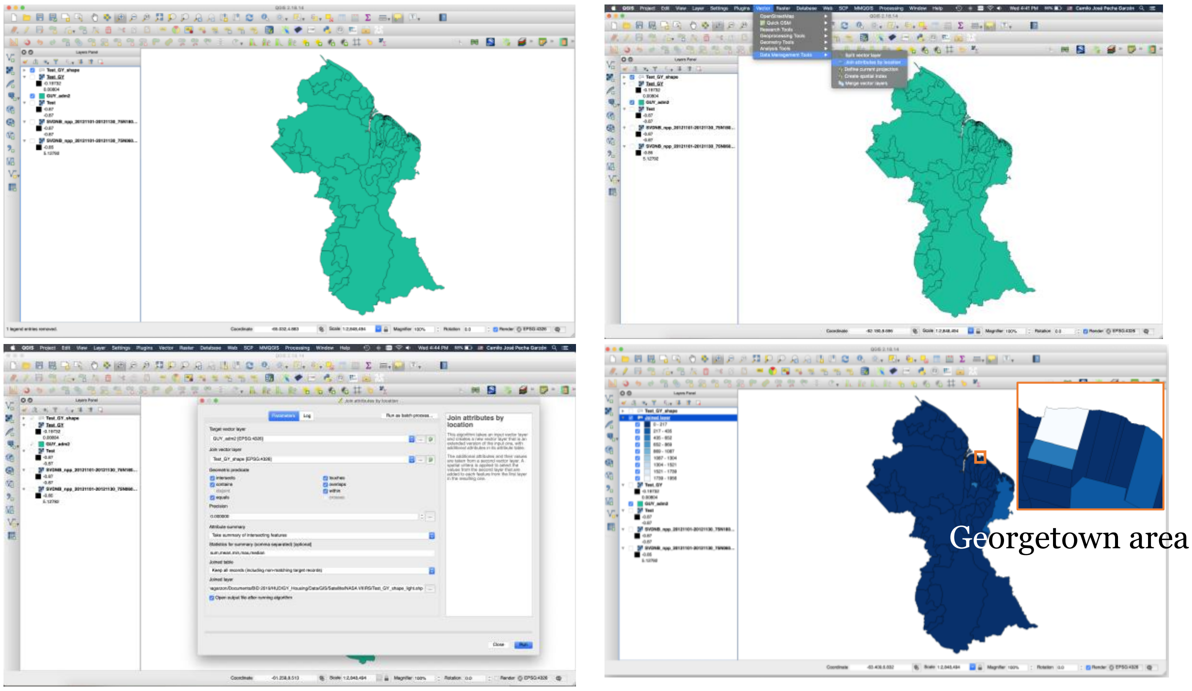


Figure 8: Combining nighttime light shapefile with administrative division shapefile

To check that this step was completed successfully, right click on the joined layer, select "open attribute table" and verify that new columns have been added as seen in the figure below.

[25] For QGIS versions >3.0, follow this tutorial: https://www.qgistutorials.com/en/docs/3/performing_spatial_joins.html

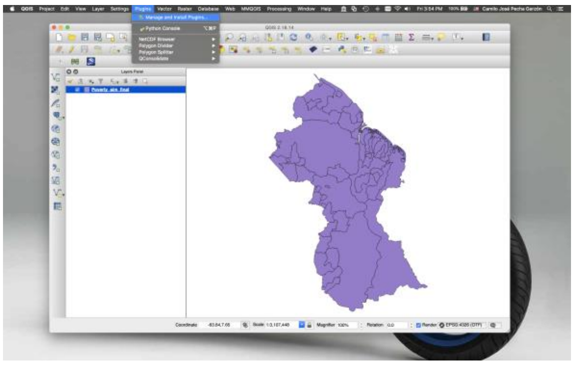
sumDN	meanDN	minDN	maxDN	medianDN	count
0.00000	0.00000	0.00000	0.00000	0.00000	1.00000
0.00000	0.00000	0.00000	0.00000	0.00000	1.00000
3.00000	1.00000	0.00000	2.00000	1.00000	3.00000
1.00000	0.00000	0.00000	1.00000	0.50000	2.00000
18.00000	3.00000	0.00000	6.00000	3.00000	6.00000
61.00000	2.00000	0.00000	7.00000	2.00000	25.00000
47.00000	3.00000	0.00000	17.00000	1.50000	14.00000
5.00000	1.00000	0.00000	3.00000	1.00000	4.00000
2.00000	0.00000	0.00000	1.00000	1.00000	3.00000
15.00000	1.00000	0.00000	3.00000	1.00000	11.00000
7.00000	1.00000	0.00000	3.00000	1.00000	5.00000
12.00000	1.00000	0.00000	4.00000	2.00000	7.00000

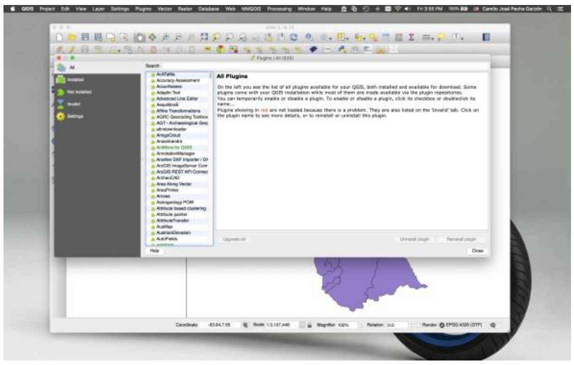
Figure 9: Example of shapefile's "attribute table"

Tabularization

The final step is to export the information stored into the shapefile to a tabular format. This can be done using a plugin called "MMQGIS". To install the MMQGIS plugin:

1. Go to menu "Plugins" > "Manage and Install Plugins…", and a window will appear
2. Using the **search** field in the window, type **mmqgis** and press install as shown in the following figures:





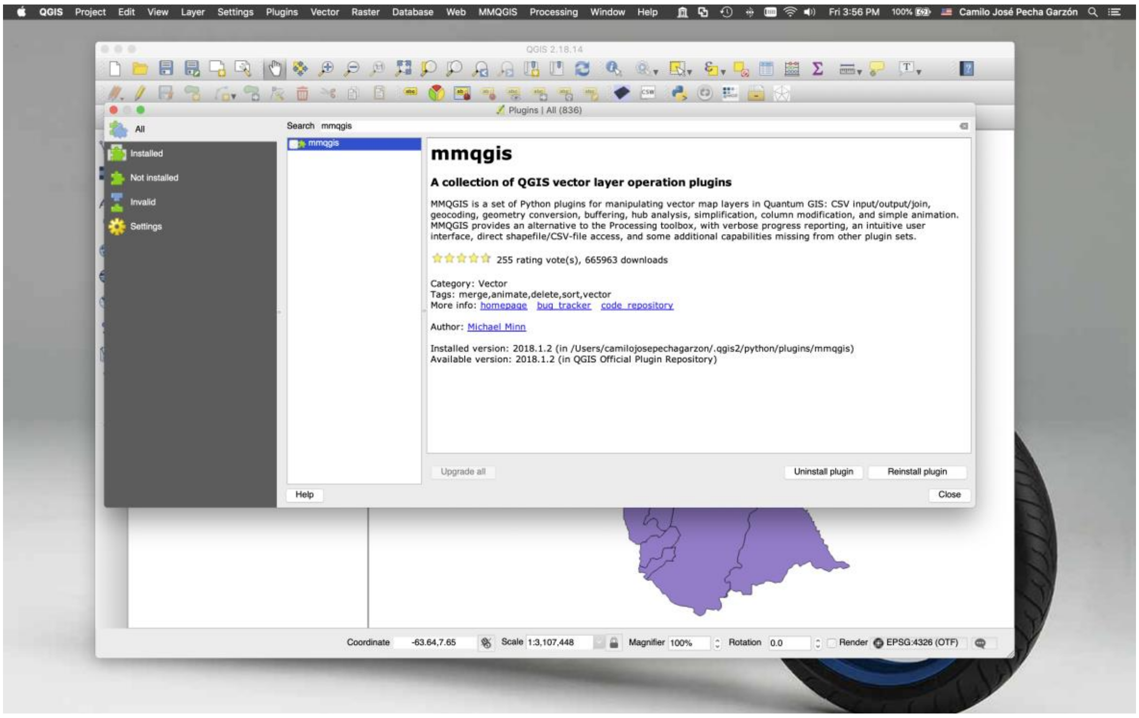
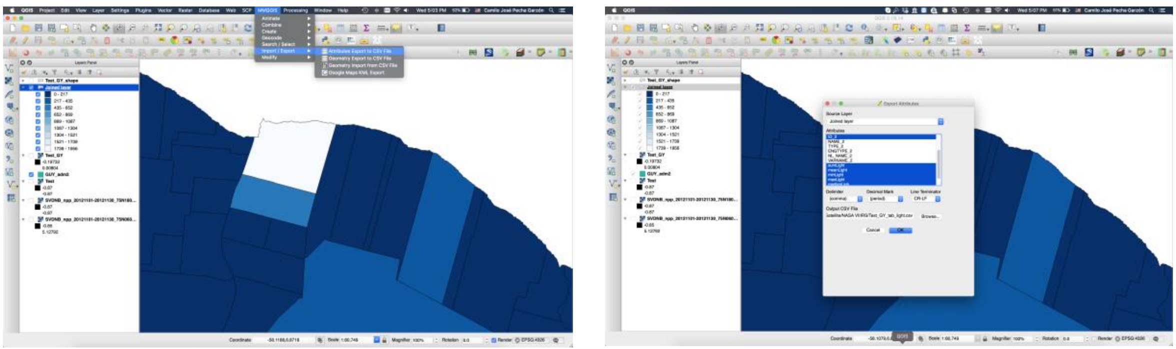
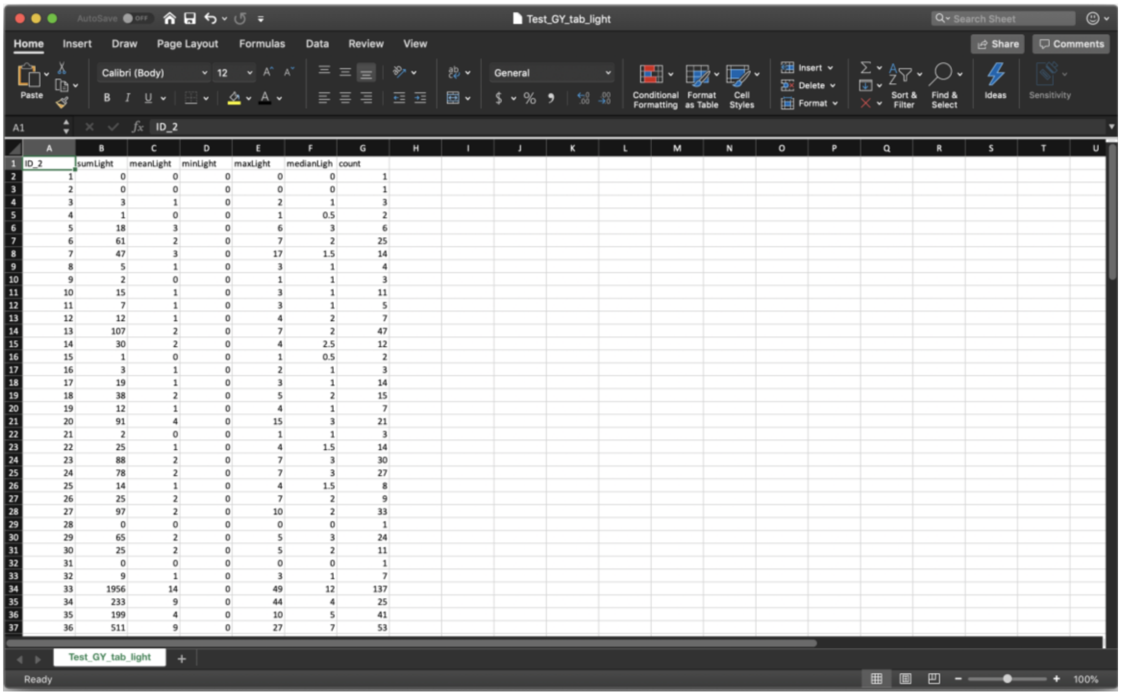


Figure 10. Installing MMQGIS plugin

Once MMQGIS is installed, it will appear as one of the options on the toolbar. Select "MMQGIS -> Import/Export -> Attributes Export to CSV file" in the menu. In the prompted window, select the desired attributes, define the name and path and click "OK". Figure 9 presents the process and an excerpt of the csv file opened in excel that contains the administrative division ID (in this case the second-level administrative division of Guyana, NDC) and the summary statistics of the nighttime light data. This CSV file can now be loaded directly into the "Prediction.R" script for processing[26].



[26] The "Prediction.R" script imports these files as "night_base_year.csv" and "night_base_year.csv" in lines 20-24 of the code.



ID_2	sumLight	meanLight	minLight	maxLight	medianLigh	count
1	0	0	0	0	0	1
2	0	0	0	0	0	1
3	3	1	0	2	1	3
4	1	0	0	1	0.5	2
5	18	3	0	6	3	6
6	61	2	0	7	2	25
7	47	3	0	17	1.5	14
8	5	1	0	3	1	4
9	2	0	0	1	1	3
10	15	1	0	3	1	11
11	7	1	0	3	1	5
12	12	1	0	4	2	7
13	107	2	0	7	2	47
14	30	2	0	4	2.5	12
15	1	0	0	1	0.5	2
16	3	1	0	2	1	3
17	19	1	0	3	1	14
18	38	2	0	5	2	15
19	12	1	0	4	1	7
20	91	4	0	15	3	21
21	2	0	0	1	1	3
22	25	1	0	4	1.5	14
23	88	2	0	7	3	30
24	78	2	0	7	3	27
25	14	1	0	4	1.5	8
26	25	2	0	7	2	9
27	97	2	0	10	2	33
28	0	0	0	0	0	1
29	65	2	0	5	3	24
30	25	2	0	5	2	11
31	0	0	0	0	0	1
32	9	1	0	3	1	7
33	1956	14	0	49	12	137
34	233	9	0	44	4	25
35	199	4	0	10	5	41
36	511	9	0	27	7	53

Figure 10: Example of final shapefile and exported attribute table

As explained above, once all luminosity data processing steps have been completed for the base year, the same steps should be repeated for the most recent year of available satellite data, producing one CSV file for each.

Functional code

Throughout this process, the prompt windows will include a box at the bottom with the python code used to run the various functions within QGIS. For the most part, users do not need to work with this box, however, if there are errors, the code contained therein can provide insight into what went wrong. The code used in this section is shown below:

Merging Geotiffs

```
gdal_merge.py -of GTiff -o /Users/…/ SVDNB_npp_20121101-
20121130_75N180W_vcmcfg_v10_c201601270845\SVDNB_npp_20121101-
20121130_75N060W_vcmcfg_v10_c201601270845.avg_rade9h.tif /Users/…/ SVDNB_npp_20121101-
20121130_75N180W_vcmcfg_v10_c201601270845\SVDNB_npp_20121101-
20121130_75N180W_vcmcfg_v10_c201601270845.avg_rade9h.tif
```

Clipping the merged geotiff

```
gdalwarp -q -cutline "/Users/…/Data/GIS/Shape/GUY_adm2.shp" -crop_to_cutline -tr 0.0041666667 0.0041666667 -of
GTiff "/Users/…/Test.tif" "/Users/…/Test_GY.tif"
```

Polygonizing

```
gdal_polygonize.py "/Users/…/Test_GY.tif" -f "ESRI Shapefile" "/Users/…/Test_GY_shape_light.shp"
Test_GY_shape_light Light
```

V. REGRESSION TO PREDICT HOUSING DEFICIT FROM NOCTURNAL LUMINOSITY

Three sets of data are needed to run this luminosity regression: qualitative deficit estimations for the base year, luminosity data from the base year, and luminosity data from the target year (the year of most recently available satellite imagery).

Using the baseline year figures for qualitative deficit estimated in Section III, the regression can be run in two steps.

First, the level and significance of the relationship between the base year qualitative deficit figures and the base year nighttime lights data must be estimated. This estimation will validate (through a test of statistical significance) that the reasoning explained at the beginning of Section IV is applicable to the area of study. The econometric model used is the following:

$$QualD_{b,i} = \beta_1 + \beta_2 L_{b,i} + \varepsilon \qquad (1)$$

Where $QualD_{b,i}$ is the variable that contains the qualitative deficit for the base year *b* in region *i*. β_1 is the intercept and the parameter of interest is β_2, which accompanies $L_{b,i}$, or the mean value of observed nighttime light for the base year *b* in region *i*. Finally, ε is an error term. The estimation process of this model produces the estimated values for the parameters and the error term, namely $\widehat{\beta_{1,b}}$, $\widehat{\beta_{2,b}}$, and $\hat{\varepsilon}_b$, respectively. In particular, and consistently with the assumption, it is expected that $\widehat{\beta_{2,b}}$ will be negative and statistically significant since an increase in luminosity should be correlated with a decrease in qualitative housing deficit.

Second, the levels of qualitative deficit are predicted for the target year. At this point, the only way to obtain figures for qualitative deficit in the target year is through predicting those values using the nighttime light information in the target year with the parameters obtain in equation 1. The process is summarized by the following equation:

$$PQD_{t,i} = \widehat{\beta_{1,b}} + \widehat{\beta_{2,b}}\, L_{t,i} + \hat{\varepsilon}_b \qquad (2)$$

Where $PQD_{t,i}$ is the predicted qualitative deficit value in the target year *t* for region *i*, obtained using the estimated parameters $\widehat{\beta_{1,b}}$, $\widehat{\beta_{2,b}}$, and $\hat{\varepsilon}_b$ in addition to the mean value of nighttime light in region *i* for target year *t*, $L_{t,i}$. The result of the exercise is summarized in the following tables.

Table 4. Results for estimation of equation 1 for qualitative deficit.

```
> fit_q <- lm(def_quali ~ light_mean, data=night2017_indicators)
> summary(fit_q)

Call:
lm(formula = def_quali ~ light_mean, data = night2017_indicators)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.28729 -0.08773 -0.00236  0.06909  0.35286 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.717699   0.015190  47.248   <2e-16 ***
light_mean  -0.022464   0.002469  -9.098   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1282 on 192 degrees of freedom
Multiple R-squared:  0.3012,	Adjusted R-squared:  0.2976 
F-statistic: 82.76 on 1 and 192 DF,  p-value: < 2.2e-16
```

Source: Calculations using "PER_predictions.R" script

Tables 4 and 5 present the results of the estimation of equation 1 for Peru. In both equations, the parameter of interest is "light_mean" as a predictor of deficit. As expected, the parameter is negative and statistically significant which supports the logic explained in Section IV, that the more nighttime lights will indicate lower levels of qualitative and total housing deficit.

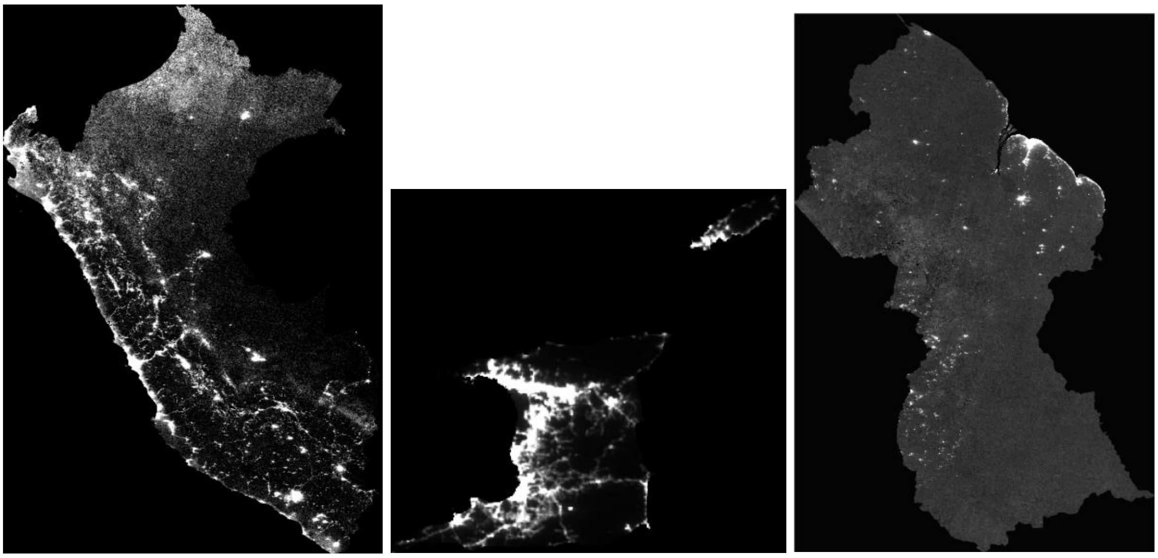
Table 5. Results for estimation of equation 1 for total housing deficit.

```
> fit <- lm(total_def ~ light_mean, data=night2017_indicators)
> summary(fit)

Call:
lm(formula = total_def ~ light_mean, data = night2017_indicators)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27650 -0.08058 -0.01036  0.06435  0.31078 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.739819   0.013417  55.139  < 2e-16 ***
light_mean  -0.016592   0.002181  -7.607 1.22e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1132 on 192 degrees of freedom
Multiple R-squared:  0.2316,	Adjusted R-squared:  0.2276 
F-statistic: 57.87 on 1 and 192 DF,  p-value: 1.221e-12
```

Source: Calculations using "PER_predictions.R" script

The results of equations 1 and 2 can be found in Annexes 2 and 3, respectively. However, it is important to note that the results of equation 1 for Trinidad and Tobago were only significant with 75% confidence for qualitative deficit and 90% confidence for total deficit. This was likely to happen as discussed in Section IV, where the CV for access to electric light indicated low precision for the island nation. Effectively, the country is too densely populated and too well-lit for this methodology to be applied with good precision. This can be seen in a side-by-side comparison of the night lights imagery for 2019 in each country.

Figure 6. 2019 night lights in the studied countries



From left to right: Peru, Trinidad and Tobago, Guyana.
Source: Earth Observatory Group VIIRS data, visualized in QGIS.

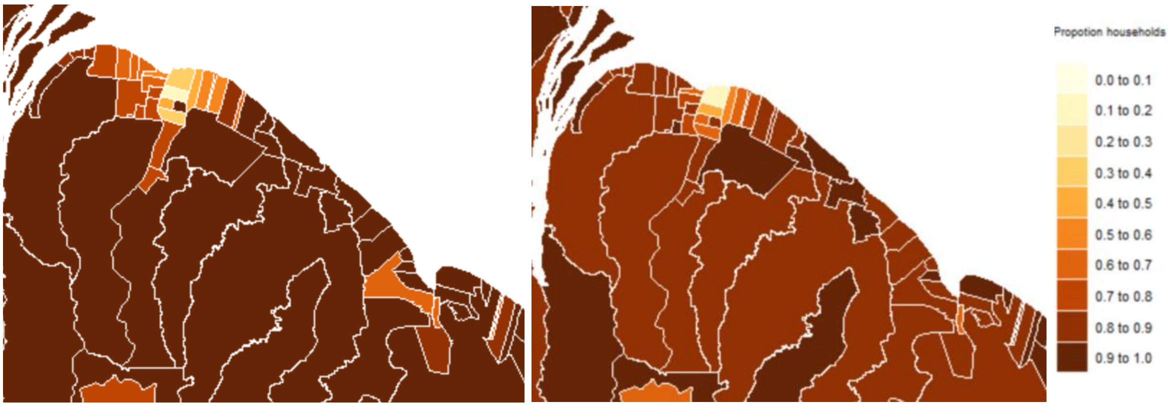
In all three countries, both qualitative and total housing deficits were estimated, on the basis of the fact that qualitative deficit was the overwhelming driver of overall deficit.

VI. RESULTS AND NEXT STEPS

The results of the predictions can be mapped using the tmap-based code in the "Predictions.R" script, by merging it with the administrative division shapefiles. If the user prefers the graphics of QGIS, the merged file, now called a *SpatialPolygonsDataFrame* ("GUY_map" for Guyana, "PER_map" for Peru, or "TT_map" for Trinidad and Tobago) can be saved as a shapefile using the *writeORG* function.

Table 5 shows the 2012 total housing deficit estimates obtained from Guyana's 2012 census data for the capital area, compared to the 2019 'now-cast' total housing deficit obtained in Section V for the same area.

Table 5. Total housing deficit in the Georgetown peri-urban area, Guyana



Left: 2012 estimates; Right: 2019 now-cast 'predictions'
Source: Calculations using "GUY_indicators.R" and "GUY_predictions.R" scripts

Metrics on qualitative housing deficit by administrative division are shown in the table below for all three studied countries:

	Base year				Target year			
Country	Low	High	Range	Median	Low	High	Range	Median
Guyana	0.12	1.0	0.88	0.98	0.10	0.99	0.89	0.89
Trinidad and Tobago	0.18	0.51	0.33	0.29	0.26	0.39	0.13	0.34
Peru	0.28	0.99	0.71	0.62	0.24	0.74	0.5	0.64

This methodology should be applied to as many countries as allowed by available data, and, ideally, tested on study areas for which the target year also has census data available for use in validation[27].

[27] This was not possible with the counties studied due to the fact that the source of satellite data, NOAA, changed the methodology the used to obtain and process satellite imagery in 2012, therefore data from prior years would not be comparable and this methodology would not be useable. Using this satellite imagery, both the base and test year must be approximately ≥ 2012.

VII. Bibliography

A. J. Bency, S. Rallapalli, R. K. Ganti, M. Srivatsa and B. S. Manjunath, "Beyond Spatial Auto-Regressive Models: Predicting Housing Prices with Satellite Imagery," 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, 2017, pp. 320-329. https://vision.ece.ucsb.edu/sites/default/files/publications/bency_wacv_17.pdf

Departamento Administrativo Nacional de Estadística. Metodología déficit de Vivienda. Bogotá, Colombia, 2009.

Divyani Kohli, Richard Sliuzas & Alfred Stein (2016) Urban slum detection using texture and spatial metrics derived from satellite imagery, Journal of Spatial Science, 61:2, 405-426, DOI: 10.1080/14498596.2016.1138247

Engstrom, Ryan; Hersh, Jonathan Samuel; Newhouse, David Locke. 2017. Poverty from space: using high-resolution satellite imagery for estimating economic well-being (English). Policy Research working paper; no. WPS 8284; Paper is funded by the Strategic Research Program (SRP). Washington, D.C. : World Bank Group.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." American Economic Review, 102 (2): 994-1028.

"Muzzini, Elisa; Eraso Puig, Beatriz; Anapolsky, Sebastian; Lonnberg, Tara; Mora, Viviana. 2016. Leveraging the Potential of Argentine Cities : A Framework for Policy Action. Directions in Development--Countries and Regions;. Washington, DC: World Bank. © World Bank.

NOAA, Ghosh, T., Powell, R., Elvidge, C. D., Baugh, K. E., Sutton, P. C., & Anderson, S. (2010)

ONU-Habitat. Déficit habitacional de América Latina y el Caribe: Una herramienta para el diagnóstico y el desarrollo de políticas efectivas en vivienda y hábitat. Programa de la Naciones Unidas para los Asentamientos Humanos (ONU-Habitat) | P.O. Box 30030, GPO Nairobi 00100, Kenia. 2015

VIII. Annex 1: Guyana Indicator Values

Indicators	Census's variable used	Options	Deficit (1=Yes, 0=No)
Inadequate	H1.1 What type of building is	1 Residential	0
		2 Residential/Commercial	0
		3 Residential/Office	0
		4 Community Service	1
		5 Other (specify)	1
	H2.2 What type of dwelling	1 Separate house/Detached	0
		2 Part of a private house/Attached	0
		3 Flat/Apartment/Condominium	0
		4 Townhouse	0
		5 Double house/Duplex	0
		6 Combined business & Dwelling	0
		7 Barracks	1
		8 Makeshift	1
		9 Other	1
Wall material	H1.2 What is the main 1 0	1 Wood	0
		2 Concrete	0
		3 Wood & Concrete	0
		4 Stone	0
		5 Adobe & Troolie Palm	1
		6 Makeshift	1
		7 Brick only (Clay Brick)	0
		8 Stone and brick	0
		9 Galvanize	1
		10 Wood & Brick	0
		11 Other (specify)	1
Roof material	H1.3 What is the main	1 Sheet metal (zinc, aluminium, galvanize)	0
		2 Shingle (asphalt)	0
		3 Shingle (wood)	0
		4 Shingle (other)	0
		5 Tile	0
		6 Concrete	0
		7 Thatched/Troolie Palm	1
		8 Makeshift	1
		9 Other (specify)	1
Cohabitation	Dewlling id and Household id	One household living in dwelling	0
		More than one household sharing	1
Overcrowding	Total number of people in		
	H4.7 How many rooms does	less than 3 people per room	0
		3 or more people per room	1
Acute	H4.8 How many bedrooms are	less than 5 people per room	0
		5 or more people per room	1
Acces to services	H4.2 What is the main source	1 Gas Lantern	1
		2 Kerosene	1
		3 Electricity (Public)	0
		4 Electricity (Private)	0
		5 Solar/Inverter	0
		6 Other (specify)	1
Acces to services	H4.3 What is the main source	1 Private, piped into dwelling	0

		2 Private catchments/rain water	0
		3 Private, piped into yard/plot	0
		4 Public, piped into dwelling	0
		5 Public, piped into yard/plot	0
		6 Public standpipe or hand pump	0
		7 Public well	0
		8 Spring/river/pond	1
		9 Truck borne	0
		10 Dug well/borehole	0
		11 Other (specify)	1
	H4.4 What is the main source	1 Piped into dwelling	0
		2 Piped into yard/plot	0
		3 Public standpipe	0
		4 Tube-well/borehole with pump	0
		5 Protected dug well/spring	0
		6 Bottled water	0
		7 Rain water collection	1
		8 Unprotected dug-well/spring	1
		9 Pond/river/stream	1
		10 Vendor/private supplier	0
		11 Other (specify)	1
Sewerage	H4.5 What type of toilet	1 W.C. (Flush toilet) linked to sewer	0
		2 W.C. (Flushtoilet)linkedtoseptictank/soak-away	0
		3 Ventilated Pit Latrine (VIP)	1
		4 Trad. Pit Latrine with slab	1
		5 Trad. Pit Latrine w/out slab	1
		6 None	1
		7 Other (specify)	1
Garbage	H4.9 How does this household	1 Dumping on land	0
		2 Compost	0
		3 Burning	1
		4 Dumping/throwing into river/sea/pond	1
		5 Burying	0
		6 Garbage truck/skip/bin - Public	0
		7 Garbage truck - Private	0
		8 Other (specify)	1

Table A1. Dummy variables for deficit defined based on Guyana’s 2012 census

IX. Annex 2: Regression Results

Housing deficit regression results: Guyana, 2012

```
Call:
lm(formula = def_quali ~ meanDN, data = night2012_indicators)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.84255 -0.04957  0.03549  0.07604  0.44562 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.964505   0.020902  46.145  < 2e-16 ***
meanDN      -0.045570   0.007563  -6.025 2.12e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1651 on 114 degrees of freedom
Multiple R-squared:  0.2415,    Adjusted R-squared:  0.2349 
F-statistic:  36.3 on 1 and 114 DF,  p-value: 2.119e-08
```

```
Call:
lm(formula = total_def ~ meanDN, data = night2012_indicators)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.84211 -0.04756  0.03594  0.07491  0.42862 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.964057   0.020460  47.119  < 2e-16 ***
meanDN      -0.043631   0.007403  -5.893 3.92e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1616 on 114 degrees of freedom
Multiple R-squared:  0.2335,    Adjusted R-squared:  0.2268 
F-statistic: 34.73 on 1 and 114 DF,  p-value: 3.92e-08
```

Housing deficit regression results: Trinidad and Tobago, 2011

```
Call:
lm(formula = def_quali ~ light_mean, data = night2012_indicators)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.106760 -0.052031 -0.002487  0.021538  0.181411 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.337593   0.032131  10.507 1.01e-07 ***
light_mean  -0.001089   0.000857  -1.271    0.226    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08254 on 13 degrees of freedom
Multiple R-squared:  0.1105,	Adjusted R-squared:  0.04204 
F-statistic: 1.614 on 1 and 13 DF,  p-value: 0.2261
```

```
Call:
lm(formula = total_def ~ light_mean, data = night2012_indicators)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.10322 -0.04088 -0.01657  0.01874  0.15954 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.4040122  0.0271756  14.867 1.54e-09 ***
light_mean  -0.0013137  0.0007248  -1.812   0.0931 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06981 on 13 degrees of freedom
Multiple R-squared:  0.2017,	Adjusted R-squared:  0.1403 
F-statistic: 3.285 on 1 and 13 DF,  p-value: 0.09308
```

Housing deficit regression results: Peru, 2017

```
Call:
lm(formula = def_quali ~ light_mean, data = night2017_indicators)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.28729 -0.08773 -0.00236  0.06909  0.35286 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.717699   0.015190  47.248   <2e-16 ***
light_mean  -0.022464   0.002469  -9.098   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1282 on 192 degrees of freedom
Multiple R-squared:  0.3012,	Adjusted R-squared:  0.2976 
F-statistic: 82.76 on 1 and 192 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = total_def ~ light_mean, data = night2017_indicators)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27650 -0.08058 -0.01036  0.06435  0.31078 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.739819   0.013417  55.139  < 2e-16 ***
light_mean  -0.016592   0.002181  -7.607 1.22e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1132 on 192 degrees of freedom
Multiple R-squared:  0.2316,	Adjusted R-squared:  0.2276 
F-statistic: 57.87 on 1 and 192 DF,  p-value: 1.221e-12
```

X. Annex 3: Prediction Results

Prediction Results: Peru, 2019					
Province	pred. total def.	pred. qual.def.	Province	pred. total def.	pred. qual.def.
Abancay	0.65543	0.60156	Jaen	0.67566	0.62891
Acobamba	0.71442	0.68105	Jauja	0.69488	0.65504
Acomayo	0.70729	0.67145	Jorge Basadre	0.58353	0.50408
Aija	0.68970	0.64764	Julcan	0.71512	0.68033
Alto Amazonas	0.66580	0.61568	Junin	0.66858	0.61901
Ambo	0.71317	0.67934	La Convencion	0.49344	0.38228
Andahuaylas	0.67366	0.62630	La Mar	0.65436	0.59693
Angaraes	0.70444	0.66765	La Union	0.71719	0.68518
Anta	0.68939	0.64741	Lamas	0.70820	0.67258
Antabamba	0.70901	0.67393	Lambayeque	0.66754	0.61770
Antonio Raymondi	0.72158	0.69071	Lampa	0.68160	0.63707
Arequipa	0.51575	0.40958	Lauricocha	0.69324	0.65281
Ascope	0.66464	0.61409	Leoncio Prado	0.65027	0.59430
Asuncion	0.72238	0.69176	Lima	0.47982	0.36382
Atalaya	0.70173	0.66430	Loreto	0.68241	0.63338
Ayabaca	0.71407	0.68051	Lucanas	0.70307	0.66595
Aymaraes	0.67611	0.62954	Luya	0.71475	0.68172
Azangaro	0.66244	0.61113	Manu	0.71203	0.67765
Bagua	0.67760	0.63165	Maranon	0.71322	0.67921
Barranca	0.64454	0.58685	Mariscal Caceres	0.67351	0.62558
Bellavista	0.70327	0.66613	Mariscal Luzuriaga	0.71917	0.68788
Bolivar	0.70273	0.66556	Mariscal Nieto	0.58183	0.50142
Bolognesi	0.69454	0.65443	Mariscal Ramon Castilla	0.71335	0.67876
Bongara	0.71390	0.68039	Maynas	0.53586	0.43944
Cajabamba	0.63280	0.57076	Melgar	0.64492	0.58539
Cajamarca	0.57419	0.49164	Moho	0.67986	0.63460
Cajatambo	0.73118	0.70240	Morropon	0.68460	0.64108
Calca	0.69973	0.66107	Moyobamba	0.65269	0.59770
Camana	0.68418	0.64054	Nazca	0.63897	0.57931
Canas	0.74167	0.70985	Ocros	0.72889	0.70086
Canchis	0.66547	0.61523	Otuzco	0.70654	0.67019
Candarave	0.71034	0.67557	Oxapampa	0.66041	0.60802
Canete	0.66345	0.61250	Oyon	0.69396	0.65376
Cangallo	0.70923	0.67412	Pacasmayo	0.65046	0.59481
Canta	0.71010	0.67507	Pachitea	0.67883	0.63316
Carabaya	0.69409	0.65364	Padre Abad	0.69340	0.65261
Caraveli	0.70408	0.66736	Paita	0.59133	0.51468
Carhuaz	0.70045	0.66233	Pallasca	0.70663	0.67060
Carlos Fermin Fitzcarrald	0.71869	0.68713	Palpa	0.69115	0.64958

Casma	0.65397	0.59909
Castilla	0.65099	0.59560
Castrovirreyna	0.72235	0.69206
Caylloma	0.68578	0.64214
Celendin	0.69537	0.65566
Chachapoyas	0.63019	0.56719
Chanchamayo	0.66241	0.61091
Chepen	0.66580	0.61567
Chiclayo	0.58286	0.50335
Chincha	0.67298	0.62318
Chincheros	0.70164	0.66398
Chota	0.69733	0.65794
Chucuito	0.67185	0.62366
Chumbivilcas	0.64574	0.58852
Chupaca	0.66197	0.60948
Churcampa	0.70228	0.66507
Concepcion	0.67740	0.63138
Condesuyos	0.69701	0.65778
Condorcanqui	0.72990	0.70227
Contralmirante Villar	0.68970	0.64741
Contumaza	0.69989	0.66165
Coronel Portillo	0.57137	0.48755
Corongo	0.72027	0.68740
Cotabambas	0.55479	0.46496
Cusco	0.39142	0.24191
Cutervo	0.72540	0.69009
Daniel Alcides Carrion	0.71825	0.68638
Dos de Mayo	0.71676	0.68408
El Collao	0.64409	0.58580
El Dorado	0.71407	0.68080
Espinar	0.57710	0.49514
Ferrenafe	0.65791	0.60398
General Sanchez Cerro	0.70813	0.67290
Gran Chimu	0.71082	0.67620
Grau	0.72061	0.68956
Huacaybamba	0.71552	0.68243
Hualgayoc	0.58966	0.51189
Huallaga	0.70975	0.67472
Huamalies	0.71782	0.68566
Huamanga	0.56308	0.47601
Huanca Sancos	0.72584	0.69516
Huancabamba	0.71256	0.67847

Parinacochas	0.68288	0.63873
Paruro	0.70882	0.67391
Pasco	0.63136	0.56851
Pataz	0.67862	0.63256
Paucar del Sara Sara	0.62555	0.56116
Paucartambo	0.66083	0.60809
Picota	0.69549	0.65553
Pisco	0.61792	0.55036
Piura	0.57147	0.48767
Pomabamba	0.71417	0.68001
Prov. Const. del Callao	0.47452	0.35631
Puerto Inca	0.68545	0.64203
Puno	0.57168	0.48671
Purus	0.69643	0.65701
Quispicanchi	0.70046	0.66233
Recuay	0.69387	0.65335
Requena	0.65992	0.60764
Rioja	0.67652	0.63012
Rodriguez de Mendoza	0.71605	0.68310
San Antonio de Putina	0.67181	0.62378
San Ignacio	0.69491	0.65509
San Marcos	0.71691	0.68452
San Martin	0.59725	0.52244
San Miguel	0.69402	0.65304
San Pablo	0.71517	0.68221
San Roman	0.48846	0.37516
Sanchez Carrion	0.67319	0.62154
Sandia	0.70679	0.67073
Santa	0.59713	0.52156
Santa Cruz	0.67869	0.63314
Santiago de Chuco	0.68395	0.64022
Satipo	0.68229	0.63791
Sechura	0.67671	0.63041
Sihuas	0.70537	0.66900
Sucre	0.72021	0.68878
Sullana	0.59548	0.52036
Tacna	0.58327	0.50350
Tahuamanu	0.69019	0.64856
Talara	0.54955	0.45657
Tambopata	0.60046	0.52665
Tarata	0.69878	0.66014
Tarma	0.66513	0.61476

Huancane	0.67366	0.62585
Huancavelica	0.66798	0.61862
Huancayo	0.54527	0.45248
Huanta	0.66115	0.60686
Huanuco	0.63113	0.56856
Huaral	0.70970	0.67510
Huaraz	0.60117	0.52720
Huari	0.62412	0.55876
Huarmey	0.69169	0.65063
Huarochiri	0.67685	0.63060
Huaura	0.65756	0.60453
Huaylas	0.69505	0.65515
Huaytara	0.72221	0.69144
Ica	0.59873	0.52468
Ilo	0.64473	0.58658
Islay	0.68665	0.64346

Tayacaja	0.69574	0.65600
Tocache	0.68198	0.63747
Trujillo	0.59887	0.52226
Tumbes	0.61328	0.54385
Ucayali	0.69636	0.65677
Urubamba	0.66495	0.61436
Utcubamba	0.66973	0.62087
Victor Fajardo	0.70278	0.66548
Vilcas Huaman	0.71962	0.68796
Viru	0.69498	0.65511
Yarowilca	0.72545	0.69573
Yauli	0.63350	0.57170
Yauyos	0.70279	0.66561
Yungay	0.69366	0.65325
Yunguyo	0.63465	0.57351
Zarumilla	0.62243	0.55681

Prediction Results: Guyana, 2019					
NDC number	pred. total def.	pred. qual.def.	NDC number	pred. total def.	pred. qual.def.
1	0.89726	0.89427	59	0.94214	0.94112
2	0.98452	0.98541	60	0.89726	0.89427
3	0.93678	0.93564	61	0.89315	0.89007
4	0.94089	0.93984	62	0.89315	0.89007
5	0.71942	0.70860	63	0.98041	0.98121
6	0.84819	0.84315	64	0.93545	0.93429
7	0.93757	0.93645	65	0.94089	0.93984
8	0.93678	0.93564	66	0.89315	0.89007
9	0.94089	0.93984	67	0.89394	0.89088
10	0.89315	0.89007	68	0.98452	0.98541
11	0.93678	0.93564	69	0.98452	0.98541
12	0.93678	0.93564	70	0.84952	0.84450
13	0.89182	0.88872	71	0.93678	0.93564
14	0.89182	0.88872	72	0.93545	0.93429
15	0.85363	0.84870	73	0.93678	0.93564
16	0.93678	0.93564	74	0.93678	0.93564
17	0.89315	0.89007	75	0.89182	0.88872
18	0.84819	0.84315	76	0.93678	0.93564
19	0.89315	0.89007	77	0.63215	0.61746
20	0.68036	0.66770	78	0.89182	0.88872
21	0.98452	0.98541	79	0.94089	0.93984
22	0.80589	0.79893	80	0.93678	0.93564
23	0.76093	0.75201	81	0.89182	0.88872
24	0.80456	0.79758	82	0.93678	0.93564
25	0.89315	0.89007	83	0.93678	0.93564
26	0.84819	0.84315	84	0.80456	0.79758
27	0.76093	0.75201	85	0.85363	0.84870
28	0.94089	0.93984	86	0.63215	0.61746
29	0.84819	0.84315	87	0.98452	0.98541
30	0.89182	0.88872	88	0.98452	0.98541
31	0.98452	0.98541	89	0.98452	0.98541
32	0.93678	0.93564	90	0.93757	0.93645
33	0.13879	0.10048	91	0.85031	0.84531
34	0.97524	0.97493	92	0.98452	0.98541
35	0.68036	0.66770	93	0.98452	0.98541
36	0.49530	0.47366	94	0.98452	0.98541
37	0.81125	0.80441	95	0.94089	0.93984
38	0.81125	0.80441	96	0.89394	0.89088
39	0.85031	0.84531	97	0.93678	0.93564
40	0.84819	0.84315	98	0.94089	0.93984
41	0.68619	0.67366	99	0.85363	0.84870
42	0.68036	0.66770	100	0.85031	0.84531
43	0.76762	0.75884	101	0.93545	0.93429

44	0.80589	0.79893		102	0.84819	0.84315
45	0.59893	0.58252		103	0.98452	0.98541
46	0.80668	0.79974		104	0.84819	0.84315
47	0.85031	0.84531		105	0.85031	0.84531
48	0.89315	0.89007		106	0.89394	0.89088
49	0.71942	0.70860		107	0.91079	0.90809
50	0.89315	0.89007		108	0.80668	0.79974
51	0.94089	0.93984		109	0.85488	0.84998
52	0.93678	0.93564		110	0.93545	0.93429
53	0.93545	0.93429		111	0.85363	0.84870
54	0.93678	0.93564		112	0.89394	0.89088
55	0.80456	0.79758		113	0.68619	0.67366
56	0.89315	0.89007		114	0.98452	0.98541
57	0.89394	0.89088		115	0.98452	0.98541
58	0.89182	0.88872		116	0.93678	0.93564

Prediction Results: Trinidad and Tobago, 2019[28]		
MUNICIPALITY	**pred. total def.**	**pred. qual.def.**
Borough of Arima	0.36134	0.31015
Borough of Chaguanas	0.37961	0.32374
Borough of Point Fortin	0.37082	0.31685
City of Port of Spain	0.30768	0.26447
City of San Fernando	0.35543	0.30573
Couva/ Tabaquite/ Talparo	0.40589	0.34679
Diego Martin	0.40013	0.34332
Mayaro/ Rio Claro	0.42228	0.36376
Penal/ Debe	0.40841	0.34888
Princes Town	0.44316	0.38858
San Juan / Laventille	0.38817	0.33099
Sangre Grande	0.41879	0.35828
Siparia	0.41490	0.35478
Tobago	0.41369	0.35368
Tunapuna/ Piarco	0.39197	0.33429

[28] Predictions for Trinidad and Tobago are only significant with 75% confidence for qualitative deficit and 90% confidence for total deficit.