Geospatial Model for Residential Property Tax Collection using the Census Block Groups as Homogenous Zones

Lisandra Benítez
Master in Geospatial Science and Technologies
Raúl Matos Flores, PhD.
Civil & Environmental Engineering and Land Surveying
Polytechnic University of Puerto Rico

Abstract *—This investigative project proposes a tax collection model based on a property value homogeneous area using the U.S. Census block group as the base region for residential tax type. Geospatial technology was used to compare market value sales with owner-occupied dwellings Census data applying geostatistics, spatial join visualization, and spatial distribution. The developed model uses the Census block group as geographic region with the market property sales as baseline for current fiscal economy. A sample of 288 records of 5,614, for years 2013 – 2018, was set aside to calculate the Root Mean Square Error after comparing each set of data. This model confers an improvement in out-of-sample prediction accuracy of up to 5% proving that market value sales resembles current fiscal economy reality.*

Key Terms *— geospatial technologies, homogeneous zone, market value, real estate appraisal process*

INTRODUCTION

The Municipal Revenue Collection Center (CRIM by its Spanish acronym) is a Municipal Entity independent of the other government agencies. It was created by the approval of Act No. 80 of August 30, 1991, as amended, as part of the Municipal Reform process. Likewise, with the creation of the CRIM, the Administration of the Municipal Property Tax Law was delegated to the Municipalities. The CRIM provides fiscal services in favor of the Municipalities and has the responsibility of notifying, assessing, collecting, receiving and distributing public funds from the property tax, the state subsidy, funds from the Electronic Lottery and any other funds that are provided by law for the benefit of the Municipalities of Puerto Rico.

CRIM uses a formula to calculate property taxes that is outdated. The formula used to calculate the property taxes does not comes from what is commonly known as real-estate appraisals based on current market values, but rather on property values dating back to 1957. This scenario presents an outdated methodology [1] not to be discussed furthermore in this investigation. However, its relevance creates the path for the chosen methodology to develop a much-needed tax collection model representative of the current economy.

The U.S. Census Bureau, 5-Year American Community Survey (ACS) is one source of data on residential real estate sales across the country. Since the Census form asks homeowners how much they think the house and the land, apartment, or mobile home and the lot will be sold if it were for sale, this made it possible to use the 5-Year ACS Census to study residential property value. Although the American Community Survey (ACS) produces population, demographic and housing unit estimates, it is the Census Bureau's Population Estimates Program that produces and disseminates the official estimates of the population for the nation, states, counties, cities and towns and estimates of housing units for states and counties. Estimates of urban and rural population, housing units, and characteristics reflect boundaries of urban areas defined based on Census 2010 data. As a result, data for urban and rural areas from the ACS do not necessarily reflect ongoing urbanization patterns. Data are based on a sample and are subject to sampling variability. The degree of uncertainty for an estimate arising from sampling variability is represented using a margin of error. The value shown here is the 90 percent margin of error. The margin of error can be interpreted roughly as providing a 90 percent probability that the

interval defined by the estimate minus the margin of error and the estimate plus the margin of error (the lower and upper confidence bounds) contains the true value.

The investigation utilizes the Census Block Group as the established homogeneous area [2]. Block Groups (BGs) are clusters of blocks within the same census tract. Each census tract contains at least one BG, and BGs are uniquely numbered within census tracts. BGs have a valid code range of 0 through 9. BGs have the same first digit of their 4-digit census block number from the same decennial census. BG's generally contain between 600 and 5,000 people. A BG usually covers a contiguous area but never crosses county or census tract boundaries. They may, however, cross the boundaries of other geographic entities like county subdivisions, places, urban areas, voting districts, etc. [3].

Another source of data is the Puerto Rico Comparable Sales Data System Web-Edition by Luis Abreu & Associates. Luis Abreu & Associates is the first, and for many years, the only company in Puerto Rico, dedicated to providing comparable sales information throughout the Island. It was established in 1989 as a tool for appraisers, brokers and professionals in the industry of real estate in Puerto Rico. Information is updated daily and/or as soon as received from the financial institution.

This investigation will utilize CRIM parcel map as a mean to georeferenced market sales into the geographic location for the Municipality of San Juan in order to compare geostatistically with the census value (owner-occupied) within the homogeneous area of the BG.

Land Value Appraisal

An appraisal is an opinion of value. For estate planning, financial planning, or sale price decisions, individuals or a trusted advisor usually orders an appraisal. An appraiser will review recent sales and listings of comparable properties [3].

Comparable are recently sold or listed properties that have similar utility, quality, age and amenities as the subject property and are in the subject property's market area. The comparable method of appraisal presents a realistic and applicable method to the development of the model applied in this investigation [7].

Geospatial Science and Techniques

Applying geospatial science and technologies through Geographic Information System (GIS) software is undoubtedly useful on mapping, displaying patterns, and detecting deficiencies any property tax system may have [4]. Spatial analysis is the core competency for every geospatial scientist. Analyzing certain amount of land characteristics in an objective way along with construction development, buying patterns, and economic trends [5] can provide a solution to bring up-to-date Puerto Rico's property tax using homogeneous zones as the base unit for tax type.

Location Geography

San Juan is the capital of Puerto Rico among 78 municipalities (Figure 1), The Municipality of San Juan (Figure 2) is 48 square miles (125 square km), 18 administrative divisions (ward), 368 Census BG's, and a population of 365,326 inhabitants (Table 1). As the core of the San Juan-Metropolitan Area and characteristically to all metropolitan areas, San Juan displays several aspects such as the amount of population, economic activities, mobility, activities of the population, and the structure of the region. The population density and location closer to higher paid jobs becomes a major consideration in determining housing purchases and tax valuation.

Figure 1
78 Municipalities of Puerto Rico

Figure 2
Municipality of San Juan

Economic and Demographic Pattern

In the last decade, high-rise housing developments proliferation targeting high purchasing power sectors have gained most of the market scenario in San Juan. Newer housing developments are capitalized for high income individuals leaving the lower- and middle-class group looking for property locations out of the San Juan Metropolitan Area where housing sales price are more suitable to their income [6].

Table 1
Municipality of San Juan Demographics

Estimates	Value
Population, Census April 1, 2010	395, 326
Population percentage (April 1, 2010 to July 1, 2019)	-19.50%
Owner-occupied housing unit rate (2014 - 2018)	53.50%
Median value of owner-occupied housing unit (2014 - 2018)	$155,000

There is a correlation between inflation and house price. In San Juan, where land availability is limited, you can see a more pronounced effect of inflation. There are a lot more factors that affect house price.

- **Consumer Price Index (CPI):** The Consumer Price Index (Table 2) measures the average change in prices over time that consumers pay for a basket of goods and services
- **CPI**: is the most widely used measure of inflation and, by proxy, of the effectiveness of the government's economic policy.

Median Value of Owner-Occupied Housing Units Census Table

This tabulation includes only specified owner-occupied housing units--one-family houses on less than 10 acres without a business or medical office on the property. These data exclude mobile homes, houses with a business or medical office, houses on 10 or more acres, and housing units in multi-unit structures.

- **Owner-Occupied:** A housing unit is owner-occupied if the owner or co-owner lives in the unit, even if it is mortgaged or not fully paid for. The owner or co-owner must live in the unit and usually is Person 1 on the questionnaire. The unit is "Owned by you or someone in this household with a mortgage or loan" if it is being purchased with a mortgage or some other debt arrangement such as a deed of trust, trust deed, contract to purchase, land contract, or purchase agreement. The unit also is considered owned with a mortgage if it is built on leased land and there is a mortgage on the unit. Mobile homes occupied by owners with installment loan balances also are included in this category.
- **The median:** divides the value distribution into two equal parts: one-half of the cases falling below the median value of the property (house and lot) and one-half above the median. Median value calculations are rounded to the nearest hundred dollars.
- **Value**: is the respondent's estimate of how much the property (house and lot) would sell for if it were for sale.

METHODOLOGY

The study area selection criteria depended on bulk data availability for a significant timeframe of more than two consecutive years, high municipal revenue, real market sales values availability, and recurrent community surveys. The criteria are met

100% with the Municipality of San Juan for the years 2013 - 2018. Data sourcing came from three different institutions:

- **CRIM** – Geodata: the cadastral map was exclusively used to geocodify market sales from the Puerto Rico Comparable Sales Data System. Only the cartographic representation was needed along with the parcel number (XXX-XXX-XXX-XX-XXX (cadaster number nomenclature) (Figure 3) to spatially joined the data. Value attributes, nor owner information was necessary in the analysis due to the outdated information. A total of 133,632 polygons were originally added to the map, once spatially joined, the universe was reduced to 5,614 records.

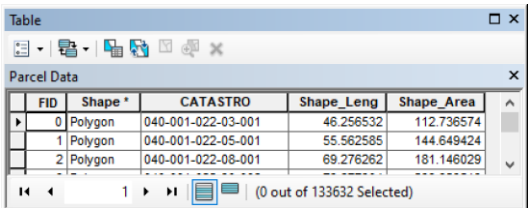

Figure 3
Parcel Map Attribute Table

- **U.S. Census** – Geodata: TIGER/Line Shapefile, 2019, state, Puerto Rico, Current Block Group State-based (Figure 4). This geodata set serves as basemap for the comparison between values. Contrary to any assumption, CRIM parcel data map cannot be used to create homogeneous zones for residential tax type [8].

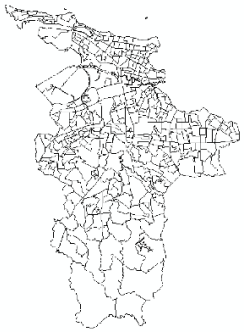

Figure 4
U.S. Census Block Group

Attributes: 2009-2013, and 2014-2018 ACS 5-Year Estimates Value Table B25075 were analyzed for content. Median sales values per BG (Figure 5) were analyzed and display to understand special behavior and patterns.

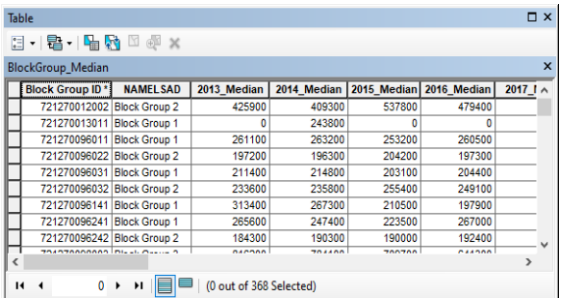

Figure 5
U.S. Census Block Group Attribute Table

- **Puerto Rico Comparable Sales Data** – Attribute: 9,261 records of comparable sales values (Figure 6) in San Juan were obtained with a $48.00 one-month subscription.

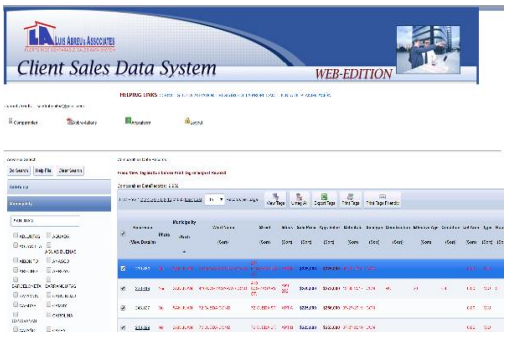

Figure 6
Puerto Rico Comparable Sales Data System Web-Edition

This data represented the greatest challenge due to the lack of consistency with the cadastral number. The data needed to be scrubbed and geocoded to CRIM parcel data (Figure 7).

Figure 7
Puerto Rico Comparable Sales Data System Web-Edition

After data validation, only 5,614 records were worth using. While market sales information is available from 2011 – 2020, the lack of cadastral

numbers on some records, and/or incorrect nomenclature, limited this investigation to years 2013 – 2018. Market values were adjusted to reflect January 2020 CPI.

Institute of Statistics of Puerto Rico

The Institute prepares and development policy of the statistical public function, provides complete, reliable, and quick access. Also, supports objectively government management, promotes the use of universal norms and standards to stimulate investigative capacity among professionals.

House Price Indexes presents two price indices of repeated sales of single-family homes in Puerto Rico. It only includes property purchases and refinances whose values are estimated using the sales price (not an appraisal). Fannie Mae, Freddie Mac, Federal Housing Administration, (FHA), and the Federal Home Loan Bank of New York (FHLBNY) as the sources of information used to prepare the Puerto Rico indices.

CPI for housing goods from 2013 – 2018 was adjusted to reflect price changes in the economy to be used as guide in order to make a valid comparison between market value property sales and census data. (Table 2).

Table 2
Puerto Rico's Consumer Price Index: Housing

Year	CPI
2020	117.98
2018	116.74
2017	116.03
2016	115.19
2015	114.67
2014	114.13
2013	112.76

For the purpose of this investigation, the CPI from 2013 – 2018 was adjusted, the deflate formula (Figure 8) was used for this purpose to reflect price changes in the economy to be used as guide in order to make a valid comparison between market value property sales and census data.

$$\text{Sales Price Adjusted} = \text{Provided Sales Price X} \frac{\text{Final CPI}}{\text{Initial CPI}}$$

Figure 8
Deflation Formula

After deflating Market Value sales, a sample was exported for comparison. A margin of error of 5% and a confidence level of 95% provided a sample size of 288 records from a universe of 5614. This sample will serve as reference for each dataset; the value closes to the sample is the closest to reality.

A series of steps was followed prior, during the analysis process to ensure data compatibility and usability (Table 3).

Table 3
Exploratory Techniques for Geospatial Data

Data Preparation	• Census Tiger BG shapefile → Clip tool → input: San Juan Municipality • Census BG → Project → output: NAD 83 • Parcel map → Join → Market Value Sales • Deflate Market Value Sales (every year) → Calculate Field → output: expression • Get a sample from Market Value Sales → Python_9.3 Code Block: import random def rand(): return random.random(), Exprt → output: new layer • Deflate Census Median Average (every year) → Calculate Field → output: expression • Calculate average for Median Sale Category per BG → Calculate Field → output: expression
Spatial Distribution Visualization	• Aggregate features based on BG • Market Value → Dissolve > optional statistics field > MEAN → output: new layer

The project design assumed the premise that market values were the closest to reality, contrary to the Census median value data. Using ArcMap 10.7.1 with Spatial Analyst, and Geostatistical Analyst extensions, all datasets were processed and geovisualized.

Root Mean Square Error Calculation

RMSE measures how much error there is between two datasets. RMSE was used to compare a predicted value and an observed value (Figure 9).

- Predicted value: Market Value/Census Median
- Observed value: Sample from Market Value

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}$$

Figure 9
RMSE Formula

Microsoft Excel was employed to calculate the RMSE by simply subtracting one set of values from the other (column fashion) and applying the square mean root to all records operations (Table 4).

Table 4
RMSE Calculation for
Sample vs Market Value & Sample vs Census Median

Sample vs Market Value	Sample vs Census Median
$112,150.21	$263,753.05

GEOSPATIAL ANALYSIS

The geovisualization process discovered and punctuated on the initial assumption: market value sales are closer to reality and tax type can be applied using the Census BG as homogeneous zone.

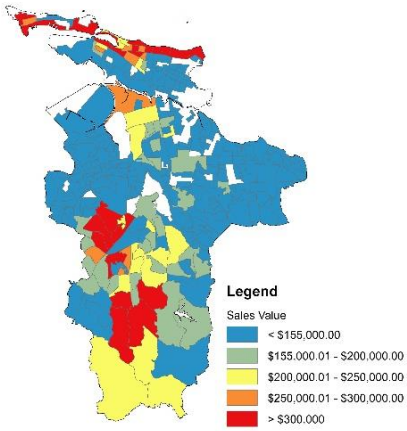

Figure 10
Comparable Market Sales

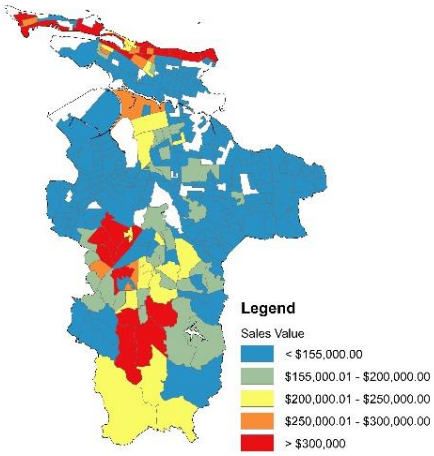

Figure 11
Sample from Comparable Market Sales

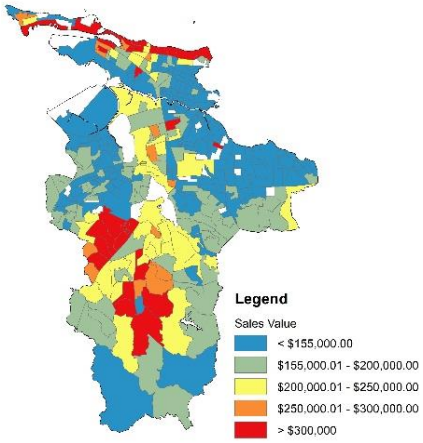

Figure 12
Census Median Values

The RMSE value closest to the sample corresponds to the calculation between Sample and Market Value: $112,150.20. RMSE quantified how much different census median is to residential properties real estate value.

All datasets were divided into five classes for property value [3] were used: less than $155,00, between $155,000 and $200,000, between $200,00 and $250,000, between $250,000 and $300,000, and greater than $300,000.

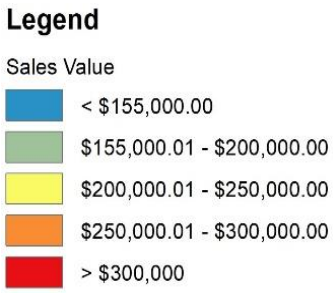

Figure 13
Sales Classes in ArcMap

The census states the following: "BG's generally contain between 600 and 5,000 people. A BG usually covers a contiguous area but never crosses county or census tract boundaries. They may, however, cross the boundaries of other geographic entities like county subdivisions, places, urban areas, voting districts, etc." If this statement remains true, tax types can be applied using the block group as a homogeneous zone according to the previously established housing development or new ones.

CONCLUSION

This investigation concludes that a homogeneous zone for tax type and real estate is suitable to bring up-to-date revenue from residential property sales. An institutional region exists within the U.S. Census and it is used for gathering owner-occupied information regarding the housing sales price. Although median value proved to be far from residential property value reality, the geographic region serves its purpose for classification of tax type. Market sales data provided clarity as of what patterns housing development are following, the block group gathers inhabitants per unit. Geospatially speaking, "The key to developing an accurate land-value assessment roll is the process of land-value mapping." (Oliver Oldman, Harvard Law School) [7]. Mapping homogeneous zones and passing Public Policy over this investigations model, is a step forward in bringing up-to-date municipal finances. Planned development is critical in an economy of uncertainty. The alternative presented in this investigation should be evaluated at depth including other factors not approached, like CRIM tax formula and home improvements.

REFERENCES

[1] German, J., et. al. (2000). Traditional Methods and New Approaches to Land Valuation. Land Policy. Acceded on Mar 6, 2020. [Online]. Available: https://www.lincolninst.edu/publications/articles/traditional-methods-new-approaches-land-valuation

[2] Pierluigi & Manganelli, Benedetto. (2014). Estimating the Market Value of the Building Sites for Homogeneous Areas. Advanced Materials Research. Acceded on Apr 1, 2020. [Online]. Available: www.scientific.net/AMR.869-870.14.

[3] Demetriou, D. (2018) Automating the land valuation process carried out in land consolidation schemes. Land Use Policy. Acceded on Feb 7, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0264837717304465

[4] *A guide to understand a residential appraisal*, 6144., Abbrev. The Appraisal Foundation., Washington, DC, 2012, pp. 4-6.

[5] Li, Lin & Pussella, Pgrni & Gunathilake, M.D.E.K. & Munasinghe, Deepika & Karadana, Chinthana. (2015). Land Valuation Systems using GIS Technology Case of Matara Urban Council Area, Sri Lanka. Bhumi, The Planning Research Journal. 4. 7. 10.4038/bhumi.v4i2.6.

[6] Matos, Raúl. (2018). El Componente Espacial En La Valoración De Inmuebles Residenciales: Uso De La Regresión Geográficamente Ponderada. Geografía y Sistemas de Información Geográfica. Available: https://www.researchgate.net/publication/324017412_el_componente_espacial_en_la_valoracion_de_inmuebles_residenciales_uso_de_la_regresion_geograficamente_ponderada

[7] Morano, P., & Manganelli, B. (2013). Estimating the Market Value of the Building Sites for Homogeneous Areas. Advanced Materials Research, 869–870, 14–19. https://doi.org/10.4028/www.scientific.net/amr.869-870.14

[8] "TIGER/Line Shapefile, 2019, state, Puerto Rico, Current Block Group State-based" U.S. Census Bureau. Washington, DC. 2019