Geotemas, Pau dos Ferros, RN, Brasil
ISSN: 2236-255X, v. 16, 2026.

REMOTE SENSING FOR POPULATION ESTIMATES IN SMALL TOWNS

Sensoriamento remoto para estimativas populacionais em cidades de pequeno porte

Teledetección para estimaciones poblacionales en ciudades pequeñas

Analissa Haga
Centro Universitário de Adamantina (FAI)
E-mail: analissahaga@hotmail.com

Paulo Roberto da Silva Ruiz
Faculdade de Tecnologia do Estado de São Paulo (FATEC)
E-mail: paulo.ruiz@cps.sp.gov.br

Article History
Received: 03 april, 2026
Accepted: 31 may, 2026
Published: 15 july, 2026

ABSTRACT

Studies on the dynamics of occupation and urban expansion impact the planning of public policies, as well as the allocation of resources for their implementation, and they gain relevance especially because one of the UN's Sustainable Development Goals (SDGs) is to develop sustainable cities and communities. Updating urban population data is important for securing financial resources to manage and plan cities; however, due to the complexity of such surveys, obtaining updated data is not trivial to support decision-making by urban managers. In this context, this study investigates Remote Sensing as an alternative for estimating urban population during inter-census periods. The objective was to verify the linear correlation between urban population and urbanized area in municipalities of the Nova Alta Paulista region - Adamantina, Dracena, Junqueirópolis, Lucélia, and Osvaldo Cruz - for the years 2000, 2010, and 2022. The methodology involved collecting population data from IBGE and surveying the urbanized area using images from the Landsat satellite, with linear regression models being developed, whose accuracy was assessed using mean absolute error metrics. The results indicated a strong correlation between the growth of urbanized area and population increase; however, the mean error indicated that the model has limitations for population estimation, and testing additional variables is recommended. Finally, it was possible to project the population of these cities for the year 2030. The development of this research concludes that Remote Sensing is a supportive tool for population estimates, contributing to the planning and the formulation of urban public policies.

Keywords: Demographic census; Spatial data science; Linear regression; Urbanization.

https://doi.org/10.33237/2236-255X.2026.7936

RESUMO

Estudos sobre a dinâmica da ocupação e expansão urbana impactam no planejamento de políticas públicas, bem como na alocação de recursos para sua implementação e ganham relevância especialmente porque um dos Objetivos de Desenvolvimento Sustentável (ODS) da ONU é desenvolver cidades e comunidades sustentáveis. A atualização da população urbana é importante para a obtenção de recursos financeiros para gerir e planejar as cidades, porém, devido à complexidade desses levantamentos, a obtenção de dados atualizados não é trivial para auxiliar na tomada de decisão dos gestores urbanos. Nesse sentido este trabalho estuda o Sensoriamento Remoto como uma alternativa para estimar a população urbana em períodos intercensitários. O objetivo foi verificar a correlação linear entre população urbana e área urbanizada em municípios da Nova Alta Paulista - Adamantina, Dracena, Junqueirópolis, Lucélia e Osvaldo Cruz - nos anos de 2000, 2010 e 2022. A metodologia envolveu coleta de dados populacionais do IBGE e levantamento da área urbanizada por meio de imagens do satélite Landsat, sendo desenvolvidos modelos de regressão linear cuja acurácia foi avaliada por métricas de erro médio absoluto. Os resultados indicaram forte correlação entre crescimento da área urbanizada e aumento da população; contudo o erro médio indicou que o modelo apresenta limitações para estimativa populacional, recomendando-se testar variáveis adicionais. Por fim, foi possível projetar a população dessas cidades para o ano de 2030. O desenvolvimento desta pesquisa conclui que o Sensoriamento Remoto é uma ferramenta auxiliar para estimativas populacionais, colaborando com o planejamento e a formulação de políticas públicas urbanas.

Palavras-chave: Censo demográfico; Ciência de dados espaciais; Regressão linear; Urbanização.

RESUMEN

Estudios sobre la dinámica de la ocupación y expansión urbana impactan en la planificación de políticas públicas y en la asignación de recursos para su implementación. Dado que uno de los Objetivos de Desarrollo Sostenible (ODS) de la ONU es desarrollar ciudades y comunidades sostenibles, estos estudios cobran relevancia. La actualización de la población urbana es importante para obtener recursos financieros para gestionar las ciudades, pero debido a la complejidad de estos levantamientos, la obtención de datos actualizados no es trivial para apoyar la toma de decisiones de los gestores urbanos. Este trabajo estudia la Teledetección como alternativa para estimar la población urbana en períodos intercensales, con el objetivo de verificar la correlación lineal entre población urbana y área urbanizada en municipios de la Nova Alta Paulista (Adamantina, Dracena, Junqueirópolis, Lucélia y Osvaldo Cruz) en los años 2000, 2010 y 2022. La metodología implicó la recolección de datos poblacionales del IBGE y el levantamiento del área urbanizada mediante imágenes del satélite Landsat, desarrollando modelos de regresión lineal cuya precisión fue evaluada con error absoluto medio. Los resultados indicaron fuerte correlación entre crecimiento del área urbanizada y aumento poblacional; sin embargo, el error medio señaló limitaciones para la estimación, recomendándose probar variables adicionales. Finalmente, se proyectó la población para 2030, concluyendo que la Teledetección es una herramienta auxiliar para estimaciones poblacionales que colabora con la planificación y formulación de políticas públicas urbanas.

Palabras clave: Censo demográfico; Ciencia de datos espaciales; Regresión lineal; Urbanización.

1 INTRODUCTION

The United Nations Sustainable Development Goals (SDGs) (UN, 2025) seek to meet the expectations of sustainable cities and communities. In this context, the occupation and expansion of urban areas are important factors for territorial planning and the formulation of effective public policies, since understanding the dynamics of these phenomena is fundamental for developing strategies, allocating government resources, developing infrastructure, and promoting social well-being (Rosa, 2018). The lack of updated data on urban populations can lead to less efficient decisions, which may hinder the socioeconomic development of municipalities and the quality of life of their inhabitants.

In Brazil, according to the National Confederation of Municipalities (2025), the criteria for the transfer of funds to municipalities vary depending on the type of resource, but generally include population, per capita income, demographic and epidemiological profile, and the municipality's share in the local budget. For federal transfers, such as the Municipal Participation Fund (FPM), population is a key factor: the larger the population, the greater the participation coefficient and, consequently, the larger the amount received, reflecting the need for more public services (Mesquita, 2025).

The Brazilian Institute of Geography and Statistics (IBGE) is the body responsible for conducting Demographic Censuses, which provide detailed information on the population and its characteristics. However, the decennial frequency of these surveys, combined with their logistical and financial complexity, generates significant information gaps between censuses (IBGE, 2022). This lag can lead to outdated indicators and impair the proper distribution of investments and public services, affecting the capacity of municipalities to meet the needs of their population (Guiarra; Yallouz; Silva, 2019).

This update is particularly relevant for small municipalities, whose revenues depend on limited sources of taxation. Santos (1982) identifies three categories of cities: large cities, characterized by the selectivity of modern productive activity; medium or intermediate cities; and small cities. According to REGIC – Regions of Influence of Cities (IBGE, 2018), cities with up to 20,000 inhabitants are classified as small – a category that encompasses more than 80% of municipalities – and are also referred to as Local Centers. Despite their weak centrality in business and public management activities, these municipalities hold significant relevance in the national context due to their number and territorial reach. Although they constitute the majority, many of them lack a master plan, as the City Statute mandates its preparation only for municipalities with a population exceeding 20,000 inhabitants, legally exempting small cities with up to that limit. These local centers rely on other higher-hierarchy urban centers as references (IBGE, 2018; Ramos, 2011).

Given this scenario, Remote Sensing (RS) emerges as an alternative for monitoring and updating territorial data, complementing the efforts of traditional censuses. This technology allows the acquisition of information about the Earth's surface from electromagnetic radiation reflected by objects, captured by sensors aboard satellites (Jensen, 2009). The ability of orbital sensors to convert this energy into images enables the mapping of territorial dynamics and urban expansion, providing valuable inputs for land use and land cover planning (Ruiz, 2017). Guiarra, Yallouz, and Silva (2019) highlight that this technology results in time and resource savings for decision-makers in regional public policy planning, such as mayors, municipal secretaries of planning and environment, regional development councilors, and technical staff of state territorial management agencies.

Based on this premise, this study seeks to answer the following question: is it possible to use satellite images of urbanized areas to estimate population during inter-censual periods in small municipalities? The investigation contributes to the development of complementary tools to traditional censuses, allowing for more frequent and less costly population estimates. Such estimates are important for more efficient urban and regional planning in the Brazilian context, assisting public managers and researchers in decision-making and resource optimization, thereby promoting more sustainable city development.

Thus, the present study aimed to verify the linear correlation between urban population and urbanized area in selected municipalities of the Nova Alta Paulista region – Adamantina, Dracena, Junqueirópolis, Lucélia, and Osvaldo Cruz – for the years 2000, 2010, and 2022, using satellite images and linear regression models. The goal is to generate insights for urban and regional planning through inter-censual population projections.

2 REMOTE SENSING IN URBAN ANALYSIS

Remote Sensing consists of the set of activities that allows obtaining information about objects composing the Earth's surface without the need for direct contact with them (Novo, 2010), operating through sensors installed on artificial satellites that capture solar radiation reflected by objects on the Earth's surface. Different materials absorb or reflect incident radiation, where part of it remains trapped in the atmosphere and another part reaches space. The sensors attached to satellites capture this radiation, which is called reflectance, and convert it into the physical property termed radiance. Each object reflects solar radiation differently along the electromagnetic spectrum, thus making it possible to differentiate them through their radiance transformed into an image (Jensen, 2009).

According to Jensen (2009), optical sensors are classified according to their resolutions: spatial, spectral, radiometric, and temporal. Spatial resolution refers to the size of the smallest detectable object, related to the produced pixel size. Spectral resolution corresponds to the number of bands the sensor has, whereby better spectral distinction of objects is achieved. Radiometric resolution is the ability to detect subtle variations in radiation intensity through gray-level quantization in each spectral band. Temporal resolution refers to the sensor's revisit time, i.e., the frequency with which new images of the same location are generated.

One of the main advantages of RS is the ability of multispectral sensors to capture information across various ranges of the electromagnetic spectrum, going beyond visible light. While a conventional color image (RGB) is composed of the combination of red, green, and blue bands – reproducing color as perceived by the human eye – multispectral sensors record data in additional bands, such as the infrared, enabling much deeper analyses of the Earth's surface (Richardson, Wiegand, 1977; Ponzoni, Shimabukuru, 2010). Commercial satellites such as WorldView-3 obtain images in 16 bands, others such as Geoeye in 8 bands, all considered multispectral. Hyperspectral sensors, in turn, use narrow bands, which allow identifying material details and differentiating surfaces; one example is NASA's AVIRIS with 224 spectral bands (Ruiz, 2017).

The precursor to current satellites is ERTS-1 or Landsat-1, launched in 1972, which was developed to work directly in natural resource research with Return Beam Vidicon (RBV) cameras and the Multispectral Scanner System (MSS), present on Landsat 2 (1975) and Landsat 3 (1978). Landsat 4 (1982) included the Thematic Mapper (TM) sensor, which remained on Landsat-5 (1984). Landsat 6 was designed with the Enhanced Thematic Mapper (ETM) sensor, with configurations similar to its predecessor, adding the panchromatic band 8 with 15 meters of spatial resolution. The ETM sensor evolved into the ETM+ (Enhanced Thematic Mapper Plus) sensor on Landsat 7 (1999), improving system accuracy within the same spectral intervals, extending the spatial resolution of band 6 (thermal infrared) to 60 meters, and the panchromatic band allowed the generation of color composites with 15 meters of resolution. Landsat 8 (2013) operates with the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) instruments and includes two new spectral bands, one designed for coastal area studies and another for detecting cirrus clouds (EMBRAPA, 2018).

In 2021, Landsat 9 was launched, representing the most recent advancement in the Earth observation program of the North American Space Agency (NASA), continuing a series of over 50 years of orbital monitoring. Inheriting the polar orbit and equatorial crossing of its predecessor, the satellite operates with two main instruments: the Operational Land Imager 2 (OLI-2), which captures data in nine spectral bands with 30-meter spatial resolution (15 meters for the panchromatic band), and the Thermal Infrared Sensor 2 (TIRS-2), which monitors surface temperature in two thermal bands with 100-meter resolution. This combination of spatial, spectral, and radiometric resolutions allows the detection of subtle changes on the Earth's surface, with vast applications, including monitoring deforestation in agriculture, water resource management, tracking urban growth, mapping natural disasters, and studying the impacts of climate change (Xu; Ren; Lin, 2024).

2.1 Applications in Urban Expansion

Mazroa et al. (2024) examine urban expansion in the Vila Velha agglomeration using Landsat images. Urban growth patterns were analyzed using Shannon Entropy, indicating a high expansion rate. The Modules for Land Use Change Simulations (Molusce) and Multilayer Perceptron Neural Networks were used to predict the expansion of the urban fabric for 2034, projecting an increase of 51.64 km² in 2024 and 62.31 km² in 2034. Campos (2020) estimates the population density of the Belo Horizonte metropolitan region using medium spatial resolution Landsat 7 ETM images and census tract data from the IBGE Census between 2000 and 2010, employing regression models. The mean reflectances of the spectral bands from the images, observed at the census tract and statistical grid levels, were used as explanatory variables. Population density of census tracts was used as the dependent variable. In a second stage, reflectances and population distributed at the pixel level of the images were observed through an iterated regression procedure. The pixel-level method yielded consistent results for municipalities with high population density (0.7%), while the census tract-level method showed satisfactory results for municipalities with low population density.

Aguiar and Baptista (2023) evaluated the growth of the urban footprint in neighborhoods of the Federal District (DF) through pixel-by-pixel classification of Landsat 5 images, which were subjected to surface reflectance conversion and geometric correction to generate maps with reduced spectral mixing and to identify classes. Land cover classification was performed using the K-Nearest Neighbor (KNN) algorithm, constructing a confusion matrix to verify accuracy and the Kappa index for validation and analysis. The urban growth trend curve was estimated, in which Ceilândia shows accelerated and fragmented expansion, with occupation of peripheral areas and reduction of vegetation. In contrast, Sol Nascente/Pôr do Sol exhibits more controlled growth, with less encroachment on natural areas, due to zoning policies. The authors highlight that the use of hybrid methods combined with the Random Forest algorithm can improve results in complex urban areas.

In studies integrating diverse geospatial data, Huang et al. (2021) applied deep neural networks, such as DenseNet and VGG, to correlate Sentinel-2 images with LandScan population data, demonstrating the ability to capture demographic variations at high resolution, although with a tendency to overestimate sparsely populated areas and underestimate densely occupied ones. Complementarily, Luo et al. (2019) combined high-resolution images from the GaoFen-2 satellite with Open Street Map data, using morphological attributes and a road hierarchy strategy within the Random Forest classifier to distinguish land cover classes with similar spectral characteristics, achieving more accurate classification in complex urban environments.

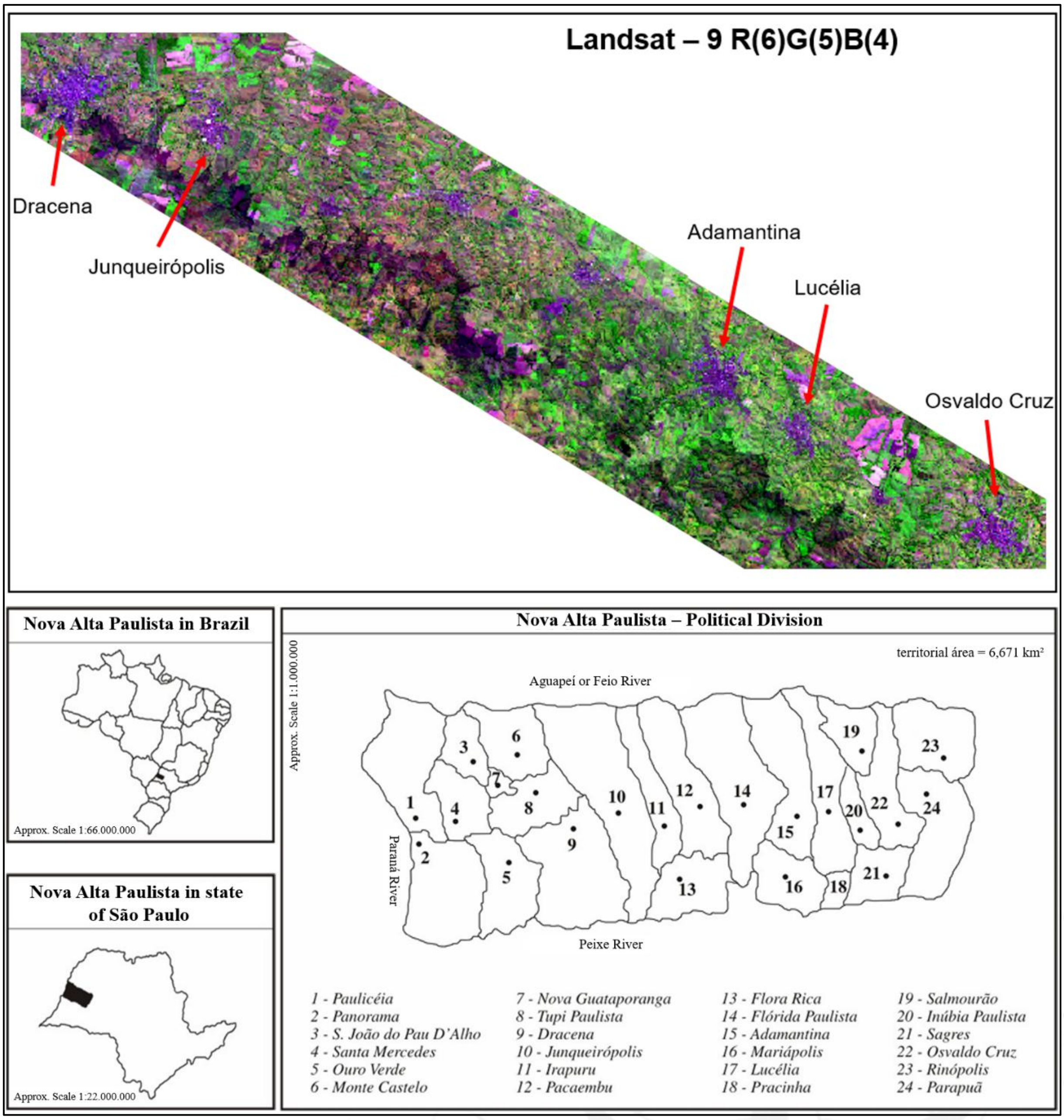
3 METHODOLOGY

3.1 Study Area

For this study, the cities of Adamantina, Dracena, Junqueirópolis, Lucélia, and Osvaldo Cruz were considered. These municipalities are located in the interior of the State of São Paulo, in the Nova Alta Paulista region, and have a population exceeding 20,000 inhabitants according to the 2022 census. Nova Alta Paulista, which is part of the far western portion of the state, is a region of relatively recent geographic formation and human occupation, whose settlement intensified mainly from the 20th century onward with the expansion of the coffee frontier and subsequently of crops such as cotton and sugarcane, which have shaped its economy. Characterized by an urban matrix in which medium and small towns play a central role in territorial organization, the region exhibits population and land use dynamics marked by strong interaction between urban and rural areas (Gil, 2004).

Figure 01 presents the location of this region and a false-color Landsat 9 image, indicating urban areas in purple. These cities were selected because they exemplify the urban expansion characteristic of regions where urbanization interacts directly with agribusiness. According to the latest Population Census (IBGE, 2022), Dracena is the municipality with the largest population, totaling 45,474 inhabitants. Adamantina ranks as the second municipality in terms of population, with 34,687 inhabitants. Osvaldo Cruz holds the third position, with 31,272 inhabitants. Finally, Junqueirópolis and Lucélia have 20,448 and 20,372 inhabitants, respectively.

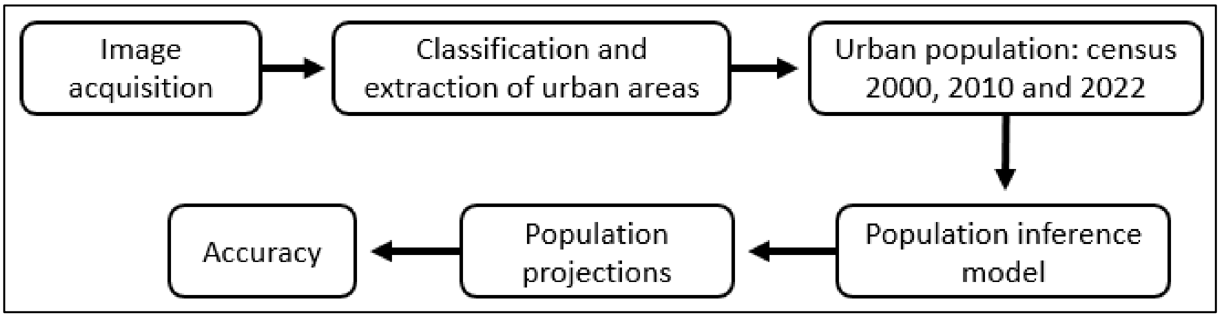
Figure 01 – Nova Alta Paulista region and Landsat-9 image with the selected cities.



Source: Created by the authors based on Landsat 9 and Gil (2004).

This study is divided into the following stages: data acquisition, followed by image classification for extracting the area of the selected cities. Concurrently, urban population data from the censuses were collected to create the population inference model, concluding with the verification of the model's accuracy. These stages are presented in Figure 02.

Figure 02 – Methodological flowchart



Source: Authors (2025).

The following sections address each of the stages presented in the methodological flowchart.

3.2 Data and extraction of urban areas

For this study, images from the Landsat satellite series were used, prioritizing images from the month of June of the IBGE census years, aiming to minimize cloud interference and ensure similar illumination conditions and solar angles across years. The image for the year 2000 was obtained from Landsat 5, dated June 5. For the year 2010, an image from June 11 was obtained from Landsat 7. For 2022, an image from June 29 from Landsat 8 was used.

After download, the images undergo preprocessing, which includes atmospheric correction using the Dark Object Subtraction (DOS) algorithm to convert digital numbers into surface reflectance, a fundamental step for temporal comparability of data (Chander; Markham; Helder, 2009). Following the selection of cities, land cover classes were defined as binary: urban and non-urban. The urban class comprises built-up areas, asphalt, and other impervious surfaces, which generally exhibit high reflectance in the mid-infrared and low reflectance in the near-infrared. The non-urban class aggregates all other land covers, such as vegetation, bare soil, and water bodies.

Training sample collection consists of selecting pixels that adequately represent each land cover class. These samples are collected considering the spectral attributes of the image bands, ensuring they are representative of that class. With a substantial set of training samples, the Support Vector Machine (SVM) algorithm was trained to recognize the spectral pattern associated with each class and then applied to the entire scene to classify each pixel individually.

SVM seeks to find an optimal hyperplane to maximize the separation between different classes in the feature space (Vapnik, 1999). It works by identifying the decision boundary that best separates points from different classes, using a concept called maximum margin. When the data are not linearly separable, it can use kernel functions, such as the Radial Basis Function (RBF) and the polynomial kernel, to project the data into a higher-dimensional space where separation becomes feasible. This method has demonstrated high accuracy in complex classifications and has been applied in land use and land cover studies due to its generalization capability and robustness under different environmental conditions (Mountrakis; Im; Ogole, 2011). Finally, validation and assessment of classification quality are performed. This step was carried out through visual inspection, comparing the generated map and the satellite images by overlaying the areas.

From the thematic map, saved in shapefile format, the area of the "urban" class was calculated. This procedure was performed in QGIS software using the field calculator. For this, it was necessary to transform the coordinate system to metric using the Universal Transverse Mercator (UTM) projection. This system allows the extraction of measurements from a vector file because its coordinates are given in meters, not degrees.

3.3 Population Inference model

For each census year, the Pearson linear correlation coefficient was calculated between the urbanized area (independent variable) and the urban population (dependent variable), according to the formula in Eq. 1, established in the statistical literature (Callegari-Jacques, 2009).

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \qquad (1),$$

where x_i represents the urbanized area of municipality i; y_i its urban population; $\bar{x}$ and $\bar{y}$ the respective means. Statistical significance was assessed using Student's t-test, considering $p < 0{,}05$ as the significance threshold. Subsequently, simple linear regression models were developed for each period, according to Eq. 2.

$$P = a * A + b \tag{2}$$

where P represents the estimated population; A the urban area; a s the slope coefficient, indicating the expected change in urban population for each unit increase in urban area; and b is the intercept coefficient, representing the estimated urban population when urban area is zero, according to the methodology described by Fávero and Belfiore (2017).

Model validation was performed using an inter-censual prediction test, applying the coefficients from one period to estimate the population for the subsequent period. Specifically, the 2000 model was applied to predict 2010 and 2022, and the 2010 model to predict 2022. Accuracy was measured using the following metrics: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The MAPE was classified as <10% (excellent), 10–20% (good), 20–50% (reasonable), and >50% (unacceptable).

MAE measures the average magnitude of errors in a set of predictions, without considering the direction of the errors. It is the average of the absolute differences between predicted values and actual values (Eq. 3).

$$AE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

where n is the number of observations, y_i is the actual value of observation i and $\hat{y}_i$ is the predicted value for observation i.

MAPE measures the magnitude of the error relative to the size of the actual value, expressing accuracy as a percentage, according to Eq. 4.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{4}$$

where n is the number of observations, y_i is the actual value of observation i and $\hat{y}_i$ is the predicted value for observation i.

Based on the validated 2022 model, population projections for 2030 were developed considering two urban expansion scenarios: Scenario 1 – Maintenance of the historical area growth rate, and Scenario 2 – Moderate growth (half of the historical rate).

In Scenario 1, the average annual growth rate (AAGR) of the urban area was calculated using the concept of exponential growth, which is analogous to compound interest calculation in financial mathematics (Eq. 5). The formula determines what constant rate, when applied repeatedly year after year over 22 periods, produces the total growth observed between 2000 and 2022.

$$TMA = \left[\left(\frac{área_{2022}}{área_{2000}} \right)^{\left(\frac{1}{22}\right)} - 1 \right] * 100 \quad (5),$$

where $\left(\frac{área_{2022}}{área_{2000}}\right)$ represents the total growth factor over the period; and $\frac{1}{22}$ distributes this growth uniformly across each year.

This rate was then projected for the following 8 years using Eq. 6, which expands the urban area maintaining the historical growth rate.

$$Área_{2030} = área_{2022} * (1 + TMA/\,100)^{8} \quad (6),$$

Finally, the corresponding population in Scenario 1 was estimated by applying the coefficients of the 2022 model in Eq. 7.

$$Pop_{2030} = \, a_{2022} * \, Área_{2030} + \, b_{2022} \quad (7),$$

where a_{2022} represents the slope coefficient (inhabitants per unit area) and b_{2022}, the intercept of the regression line.

In Scenario 2, only the territorial expansion assumption was modified, using half of the historical rate calculated with the same area projection formula (Eq. 8). Furthermore, the application of the regression model for population estimation remained unchanged.

$$Área_{2030} = \, a_{2022} * (1 + \frac{\left(\frac{TMA}{2}\right)}{100})^{8} \quad (8),$$

The multi-scenario approach follows the population projection methodology for Brazil developed by the Ministry of Social Development and Fight against Hunger for municipal population projections (Brasil, 2013).

4 RESULTS

The population data and the estimated areas of the municipalities obtained via satellite are presented in Table 01.

Table 01 – Population Data and estimated área of minicipalities.

	2000			2010			2022		
Municipality	**Total Pop.**	**Urban Pop.**	**Est. Area (Km²)**	**Total Pop.**	**Urban Pop.**	**Est. Area (Km²)**	**Total Pop.**	**Urban Pop.**	**Est. Area (Km²)**
Adamantina	33,497	30,368	7.415	33,797	31,948	10.658	34,687	33,536	13.241
Dracena	40,500	37,153	7.351	43,258	39,946	11.893	45,474	42,064	17.343
Junqueirópolis	17,005	13,420	3.047	18,726	15,399	4.711	20,448	18,958	6.807
Lucélia	18,316	15,698	3.561	19,882	17,219	4.182	20,061	16,654	5.856
Osvaldo Cruz	29,648	26,141	5.406	30,917	27,782	7.561	31,272	28,953	9.664

Source: Adapted from IBGE (2025).

Table 02 presents a summary of the correlations. In all years evaluated, the correlations were very strong, highly significant, and well represented by the linear model.

Table 02 – Summary of correlation results.

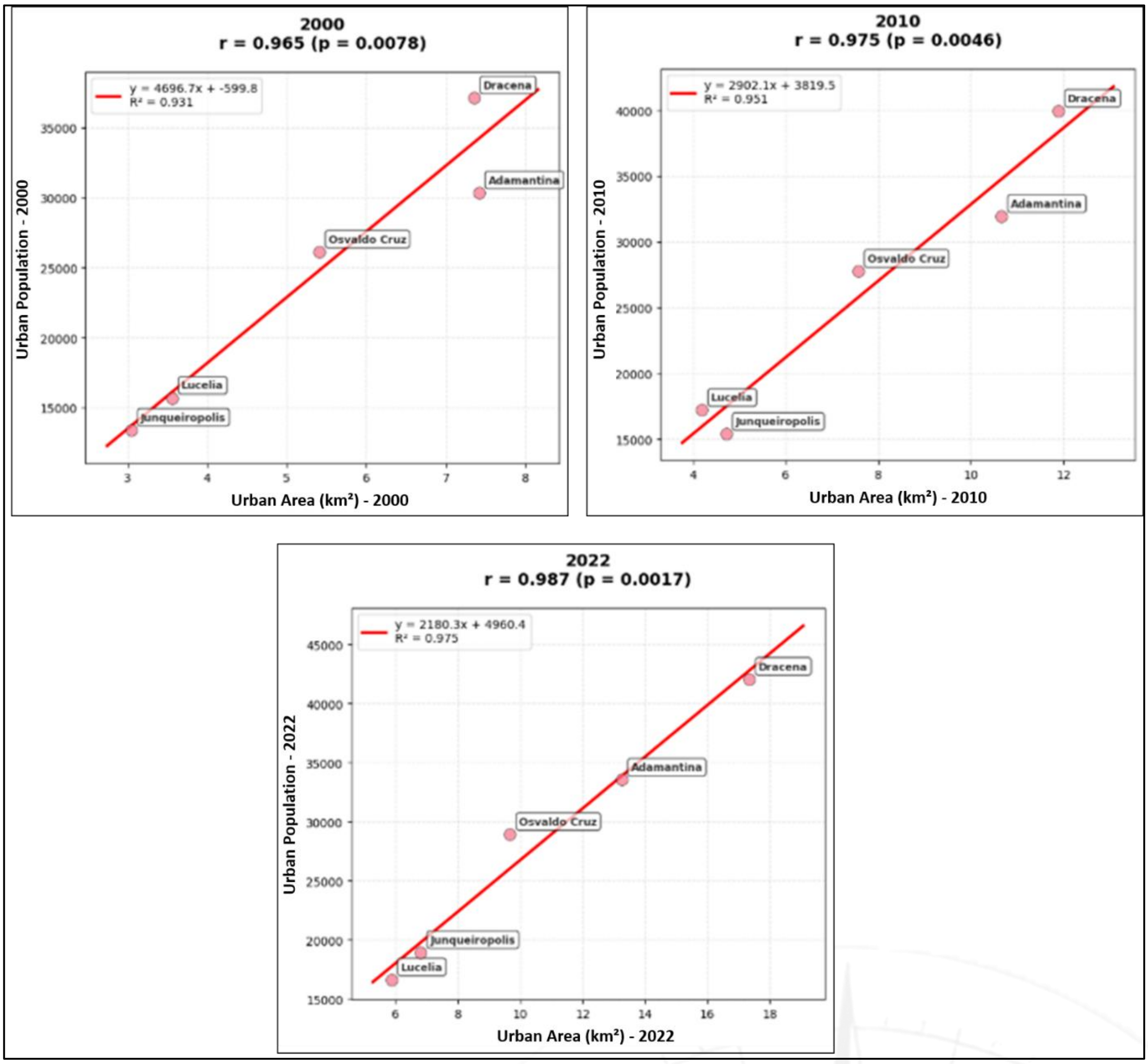
Year	Correlation	p-value	R^2
2000	0.9649	0.0078	0.9311
2010	0.9754	0.0046	0.9514
2022	0.9874	0.0017	0.9750

Source: Authors (2025)

Figure 03 presents the linear regressions. In 2000, although the linear model represents 96% of the data, Dracena and Adamantina show a greater distance from the linear intercept, indicating a growth pattern different from the other cities. In 2010, Adamantina and Dracena become closer to the intercept, increasing the correlation and making the model more balanced, although the other cities deviate slightly from the line.

Finally, in 2022, Dracena, Adamantina, Junqueirópolis, and Lucélia are very close to the intercept, exhibiting the highest correlation.

Figure 03 – Linear regression for the years 2000, 2010, and 2022

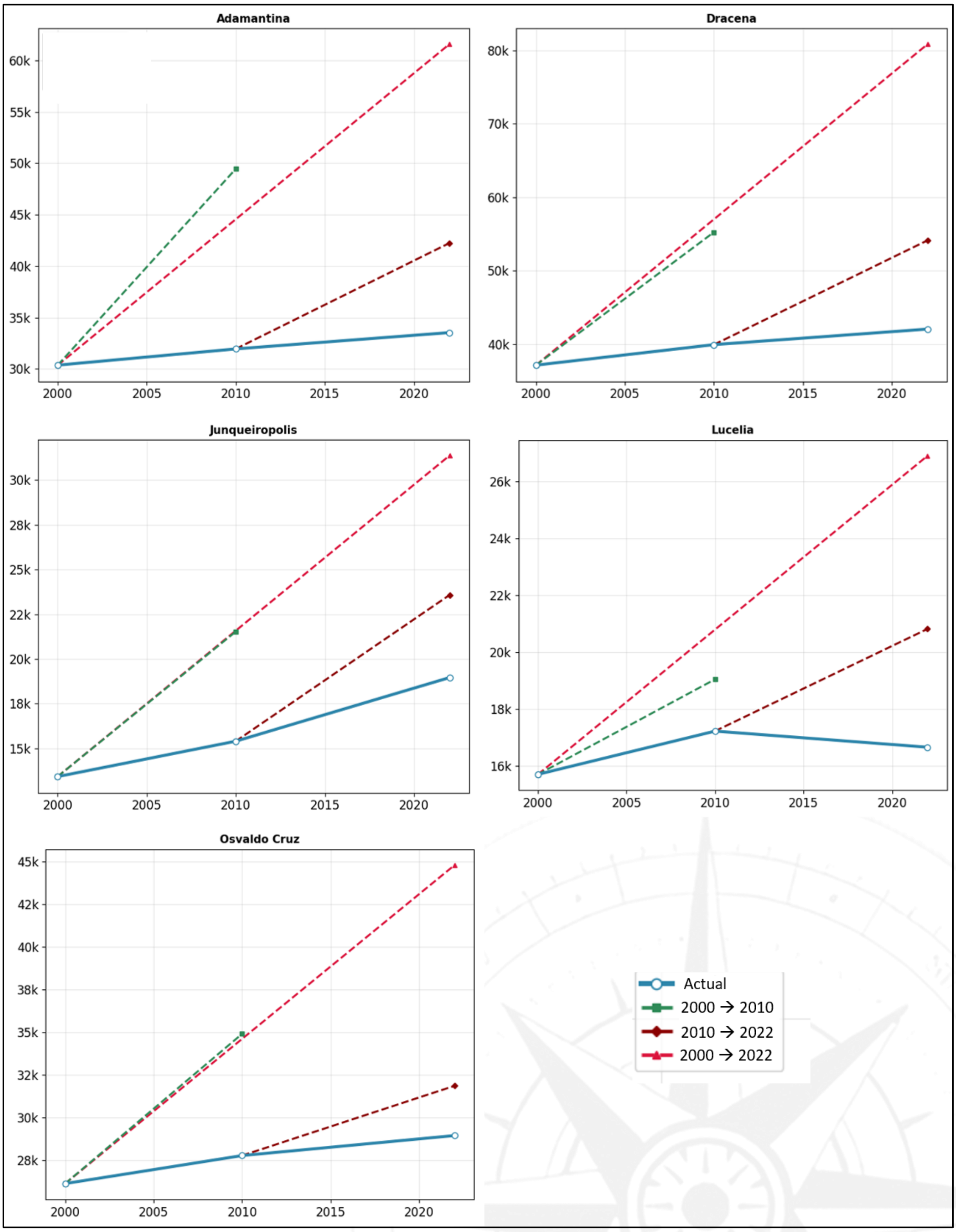


Source: Authors (2025).

The strong positive correlation (r), significance (p-value), and model representation (R^2) observed corroborate the literature used in this study, indicating that RS can be an effective complementary alternative for monitoring urban dynamics (Saberifar, 2013; Campos, 2020; Huang et al., 2021). This finding indicates that the urban footprint is a relevant indicator for population growth.

Figure 04 presents population projections based on the models generated by the correlation equation for each city. The dashed lines correspond to the population projection for the next census year, with the 2000 model projecting populations for 2010 and 2022, and the 2010 model projecting the population for 2022.

Figure 04 – Population projections by municipality – modeling scenarios



Source: Authors (2025)

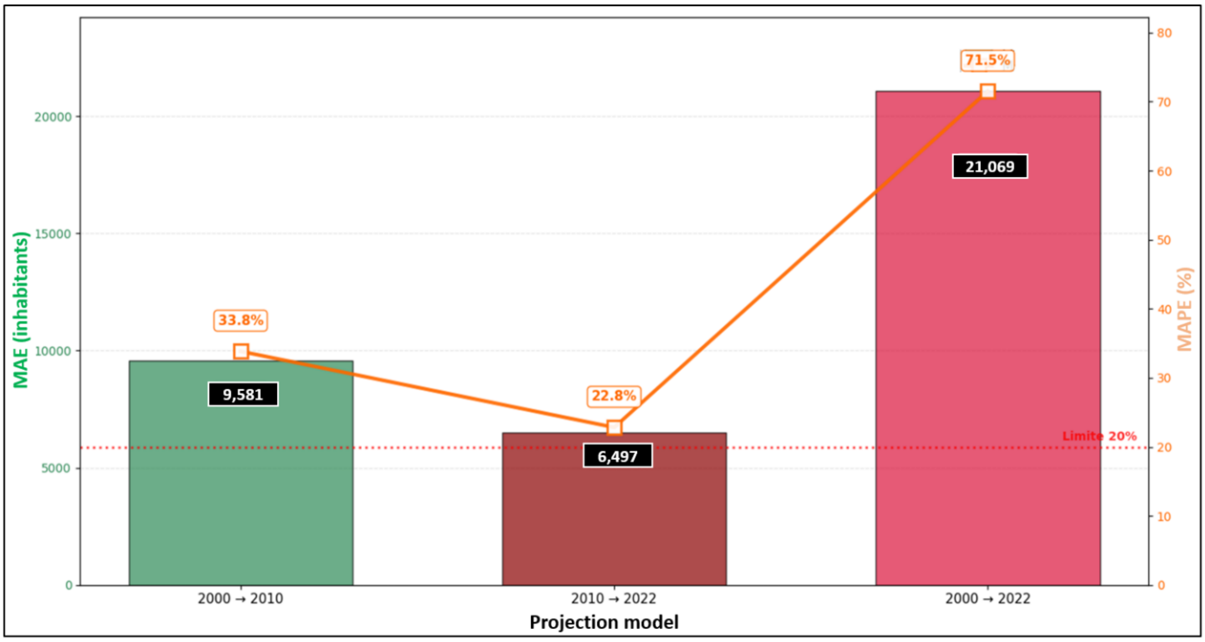
Figure 05 presents the accuracy of model predictions for each city, revealing, through MAE (bars – inhabitants) and MAPE (line – percentage), temporal patterns in population projections. The 2000 → 2010 model already showed significant errors, ranging from 10.6% in *Lucélia* to 54.8% in *Adamantina*, indicating changes in urban density patterns during the first decade. The updated 2010 → 2022 model demonstrated substantial improvement, with errors between 10.1% and 28.7%, confirming that recent models produce more reliable predictions. The 2000 → 2022 model, in turn, evidenced the limited lifespan of these area-population relationships, with enormous errors ranging from 54.7% to 92.2%, highlighting that long-term projections require periodic parameter recalibration.

Figure 05 – Model accuracy by city

Source: Authors (2025)

Figure 06 presents the overall accuracy of the models based on the average of the five cities. The superiority of the 2010 → 2022 model is evident, showing the best balance between error metrics: an absolute error of 6,497 inhabitants, corresponding to a 22.8% error relative to reality. In contrast, the 2000 → 2022 model demonstrates the temporal degradation of parameters, with a large error of 21,069 inhabitants, exceeding the error of the best model by more than threefold.

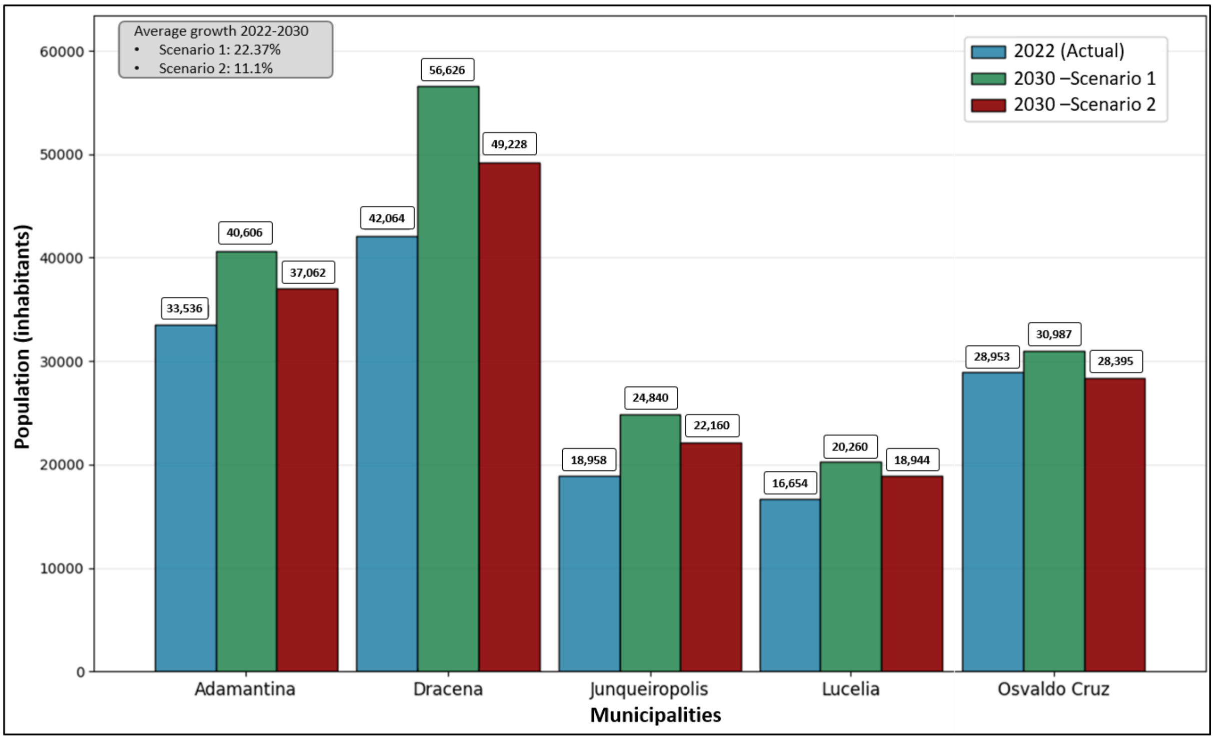
Figure 06 – Model accuracy – average of cities



Source: Authors (2025)

The population projection graph for 2030 presents three sets of comparative bars for each municipality, where the blue bars represent the actual population in 2022, the green bars indicate projections for Scenario 1 (maintenance of historical territorial expansion rates), and the wine-colored bars show estimates for Scenario 2 (moderate urban area growth) (Figure 07). Analyzing the projections, Dracena shows the largest absolute growth, increasing from 42,064 inhabitants in 2022 to over 54,000 in Scenario 1 by 2030, reflecting its high historical territorial expansion rate of 3.82% per year. Junqueirópolis stands out with the highest percentage growth, projecting an increase of 25–35% due to its intense urban expansion (3.49% per year), while Adamantina and Osvaldo Cruz maintain moderate growth between 10–20%, consistent with their expansion rates of 2.58% and 2.50%, respectively. Lucélia, despite a territorial growth rate of 2.25% per year, shows the smallest population increase, indicating possible changes in population density patterns.

Figure 07 – Population projections for 2030 based on the 2022 model



Source: Authors (2025)

Although the linear model presents good parameters, due to its simplicity, it cannot fully explain urban population growth through urban footprint expansion, as there are limitations regarding heterogeneous urbanization dynamics. The metrics suggest that local specific factors may influence the relationship between area and population in distinct ways. Factors such as urban verticalization, the presence of large non-residential areas, or the socioeconomic distribution of the population directly influence density and were not addressed in this study.

It is emphasized that this study aims to complement and support the use of tools that facilitate demographic studies; however, it does not replace the need for census surveys.

5 FINAL REMARKS

This study aimed primarily to verify the linear correlation between urban population and urbanized area in five municipalities of the Nova Alta Paulista region (Adamantina, Dracena, Junqueirópolis, Lucélia, and Osvaldo Cruz) for the years 2000, 2010, and 2022, in order to assess the feasibility of using satellite images to estimate population during inter-censual periods. To this end, data collection and organization, development and validation of linear regression models, analysis of the temporal evolution of urban occupation patterns, and the elaboration of population projections were carried out.

The results indicate a strong and significant positive correlation between the expansion of urbanized area, detected by RS, and population growth in the studied municipalities. These findings positively answer the research question: "Is it possible to use satellite images of urbanized areas to estimate population between censuses in medium and small municipalities?" pointing to the applicability of RS as an indicator of demographic and territorial dynamics and for monitoring urban development, which over time show convergence in occupation patterns.

However, despite the high correlation (r) and explanatory power of the models (R^2), the accuracy metrics indicated that linear regression models present limitations for generating population estimates. The mean absolute percentage error suggests that the complexity of population dynamics within urban areas is not fully captured when using a single predictor variable, necessitating additional variables such as urban verticalization, as suggested for future work with other samples and in larger numbers.

This research intends to contribute to the monitoring of urban expansion and the inference of population trends between demographic censuses for territorial planning, resource allocation, and the formulation of more efficient and sustainable public policies, and to reinforce the viability of RS as a complementary tool to traditional demographic data collection methods.

The limitations of this study include the simplicity of the linear regression model, which, although identifying trends, does not guarantee accuracy for population estimates in all situations; the relationship between urbanized area and population may not capture the heterogeneity of population expansion. Furthermore, the study was restricted to five municipalities in Nova Alta Paulista, which, although providing urban support for analyses, generalization to other regions or types of municipalities should be made with caution, as urban growth dynamics may vary.

It is recommended to study other variables and models, such as nighttime light data, energy consumption, land use type (residential, commercial, industrial), or verticalization indices, density, socioeconomic factors in non-linear models, and in different regions to test the robustness and adaptability of models built using orbital data.

REFERENCES

AGUIAR, B. S.; BAPTISTA, G. M. M. Aplicação de aprendizado de máquina para análise do crescimento urbano: o caso das regiões administrativas Ceilândia e Pôr do Sol no Distrito Federal. **Revista Brasileira de Geomática**, v. 11, n. 2, p. 465-487, 2023.

BRASIL. Ministério do Desenvolvimento Social e Combate à Fome. **Projeções populacionais para o Brasil**. Brasília, DF: MDS, 2013. 45 p. Disponível em: https://acervodigital.mds.gov.br/xmlui/bitstream/handle/123456789/320/ETEC_23_2013%20Proje%c3%a7%c3%b5es%20populacionais%20para%20o%20Brasil.pdf?sequence=1. Acesso em: 18 out. 2025.

CALLEGARI-JACQUES, Sidia M. **Bioestatística: princípios e aplicações**. Artmed Editora, 2009.

CAMPOS, J.; RIGOTTI, J. I. R.; BAPTISTA, E. A.; MONTEIRO, A. M. V.; REIS, I. A. Population Estimates from Orbital Data of Medium Spatial Resolution: Applications for a Brazilian Municipality. **Sustainability**, v. 12, n. 9, p. 3565, 2020. DOI: 10.3390/su12093565.

CHANDER, G.; MARKHAM, B. L.; HELDER, D. L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. **Remote Sensing of Environment**, v. 113, n. 5, p. 893-903, 2009.

CONFEDERAÇÃO NACIONAL DE MUNICÍPIOS. **Transferências constitucionais.** Brasília, [S.d]. Disponível em: https://cnm.org.br/. Acesso em: 16 out. 2025.

EMBRAPA. **Satélites de Monitoramento**. Campinas, 2018. Disponível em: < https://www.embrapa.br/satelites-de-monitoramento >. Acesso em: 20 mai. 2025.

FÁVERO, L. P.; BELFIORE, P. **Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®**. Elsevier Brasil, 2017.

FLORENZANO, T. G. **Imagens de satélite para estudos ambientais**. São Paulo: Oficina de Textos, 2002.

GIL, I. C. Territorialidade e desenvolvimento contemporâneo. **Revista Nera**, n. 4, p. 5- 19, 2004.

GUIARRA, A. P. M.; YALLOUZ, G. A. N.; SILVA, I. T. Estudo bibliométrico sobre o uso de geotecnologias aplicadas ao planejamento de cidades médias. **Terr@ Plural**, v. 14, p. 1–14, 2019.

HUANG, X.; ZHU, D.; ZHANG, F.; LIU, T.; LI, X.; ZOU, L. **Sensing population distribution from satellite imagery via deep learning**: Model selection, neighboring effects, and systematic biases. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, v. 14, p. 5137–5151, 2021. DOI: 10.1109/JSTARS.2021.3076630.

IBGE. **A rede urbana brasileira**. Rio de Janeiro: IBGE, 2018. Disponível em: https://www.ibge.gov.br/apps/regic/pdf/03_rede_urbana.htm. Acesso em: 30 jan. 2026.

IBGE. **Censo 2022**. 2022. Disponível em: https://censo2022.ibge.gov.br/panorama/. Acesso em: 13 ago. 2025.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: uma perspectiva em Recursos Terrestres**. Tradução de JCN Epiphanio. São José dos Campos, SP: Parênteses, 2009. 598 p.

LUO, N.; WAN, T.; HAO, H.; LU, Q. Fusing High-Spatial-Resolution Remotely Sensed Imagery and OpenStreetMap Data for Land Cover Classification Over Urban Areas. **Remote Sensing**, v. 11, n. 1, p. 88, 7 jan. 2019. DOI: 10.3390/rs11010088.

MAZROA, A.; MAASHI, M.; KOUKI, F.; OTHMAN, K. M.; SALIH, N.; ELFAKI, M. A.; BEGUM S, S. An analysis of urban sprawl growth and prediction using remote sensing and machine learning techniques. **Journal of South American Earth Sciences**, v. 142, p. 104988, 2024. ISSN 0895-9811. DOI: 10.1016/j.jsames.2024.104988.

MESQUITA, J. M. **Análise crítica do regime de distribuição do fundo de participação dos municípios: uma avaliação ex-post e ex-ante das propostas de inovação legislativa**. 2024. 78 f. Dissertação (Mestrado em Administração Pública) – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, Brasília, 2024. Disponível em: https://repositorio.idp.edu.br//handle/123456789/5487. Acesso em: 30 jan. 2026.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: A review. **ISPRS journal of photogrammetry and remote sensing**, v. 66, n. 3, p. 247- 259, 2011.

NOVO, E. M. L. **Sensoriamento Remoto: princípios e aplicações**. Editora Blucher, 2010.

ONU - Organização das Nações Unidas (Brasil). **Objetivos de Desenvolvimento Sustentável no Brasil**. [s.l.], [s.d.]. Disponível em: https://brasil.un.org/pt-br/sdgs. Acesso em: 17 out. 2025.

PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento Remoto no Estudo da Vegetação**. São José dos Campos: Parêntese, 2010, 127p.

RAMOS, E. F. **As cidades médias em perspectiva: uma análise dos papéis e tendências de um grupo de cidades médias brasileiras**. 2011. Dissertação (Mestrado em Geografia) – Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

RICHARDSON, A. J.; WIEGAND, C. L. Distinguishing vegetation from soil background information. **Photogrammetric engineering and remote sensing**, v. 43, n. 12, p. 1541-1552, 1977.

ROSA, M. R. Classificação do padrão de ocupação urbana de São Paulo utilizando aprendizagem de máquina e Sentinel 2. **Revista do Departamento de Geografia**, p. 15-21, 2018.

RUIZ, P. R. S. **Classificação da cobertura do solo urbano usando árvores de decisão a partir de cenas WorldView-2 e WorldView-3 para diferentes níveis de legenda**. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2017, 181 p.

SABERIFAR, R. Estimation of urban population by remote sensing data. **J Emerg Trends Econ Manag Sci**, v. 4, n. 6, p. 565, 2013.

SANTOS, M. **Espaço e sociedade: ensaios**. 2. ed. Petrópolis: Vozes, 1982.

XU, H.; REN, M.; LIN, M. Cross-comparison of Landsat-8 and Landsat-9 data: A threelevel approach based on underfly images. **GIScience & Remote Sensing**, v. 61, n. 1, p. 2318071, 2024.

VAPNIK, V. **The nature of statistical learning theory**. Springer Science & Business Media, 1999.

***