Estimation of socioeconomic indicators through satellite imagery - Analysis of urban areas overlapping

GIUZIO, C. M. O.[1]; ZOELLER, I. V.[1]; COELHO, R. A.[2]

[1] Department of Electronic Systems Engineering - Polytechnic School of the University of São Paulo

[2] The Department of Computer Engineering and Digital Systems - Polytechnic School of the University of São Paulo

Email: carlosgiuzio@usp.br, igorzoeller@usp.br, rafaelcoelho@usp.br

INTRODUCTION

Brazil is the fifth largest country in the world and the NEXUS area (Figure 1), which includes the São Francisco and Parnaíba River basins, covers approximately 30% of the national territory, representing a rich diversity of natural resources. Therefore, its vast territorial extension encourages the application of preservation and sustainable development policies to protect its diversity, which are aided by monitoring the region through environmental and socioeconomic indicators.

However, this task is made difficult by the low temporal frequency, since surveys are carried out only every 10 years, and spatial coverage of censuses that collect this information. In addition, there is great availability of satellite imagery captured with high frequency by various sources.

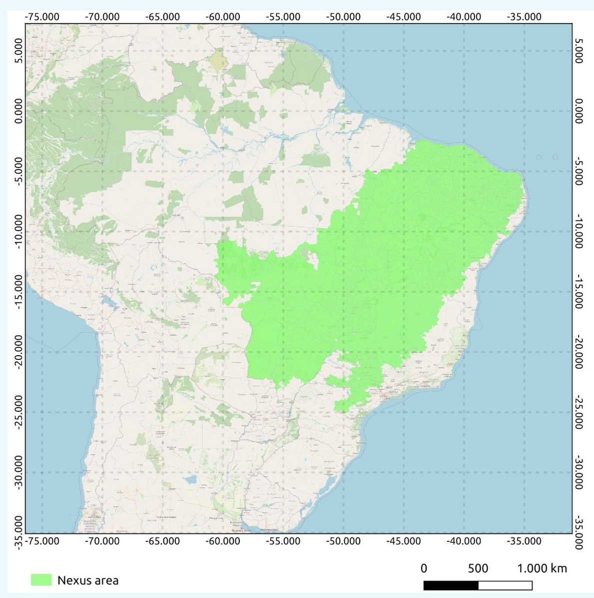

Figure 1 - NEXUS area map

OBJECTIVES

The objective of this study is to adapt the methodology presented in Yeh et. al [2] for the NEXUS area, in order to estimate the socioeconomic indicators and differentiate urban and rural areas. This will be useful to explore the different Brazilian regions, assisting future analyzes of the territory regarding the relationship between such indicators, showing how regions are influenced by being close or distant from protected areas.

The database of images to be collected, through the Google Earth Engine API [3], must meet three conditions to avoid model bias: (a) each image represents only a single type (urban or rural), (b) the amount of rural and urban images is in the same proportion and (c) avoid the presence of the same region in more than one image. Therefore, an analysis of the distribution of census tracts in the NEXUS area is necessary, exploring the extent and arrangement of urban and rural types, in order to determine the regions for obtaining images.

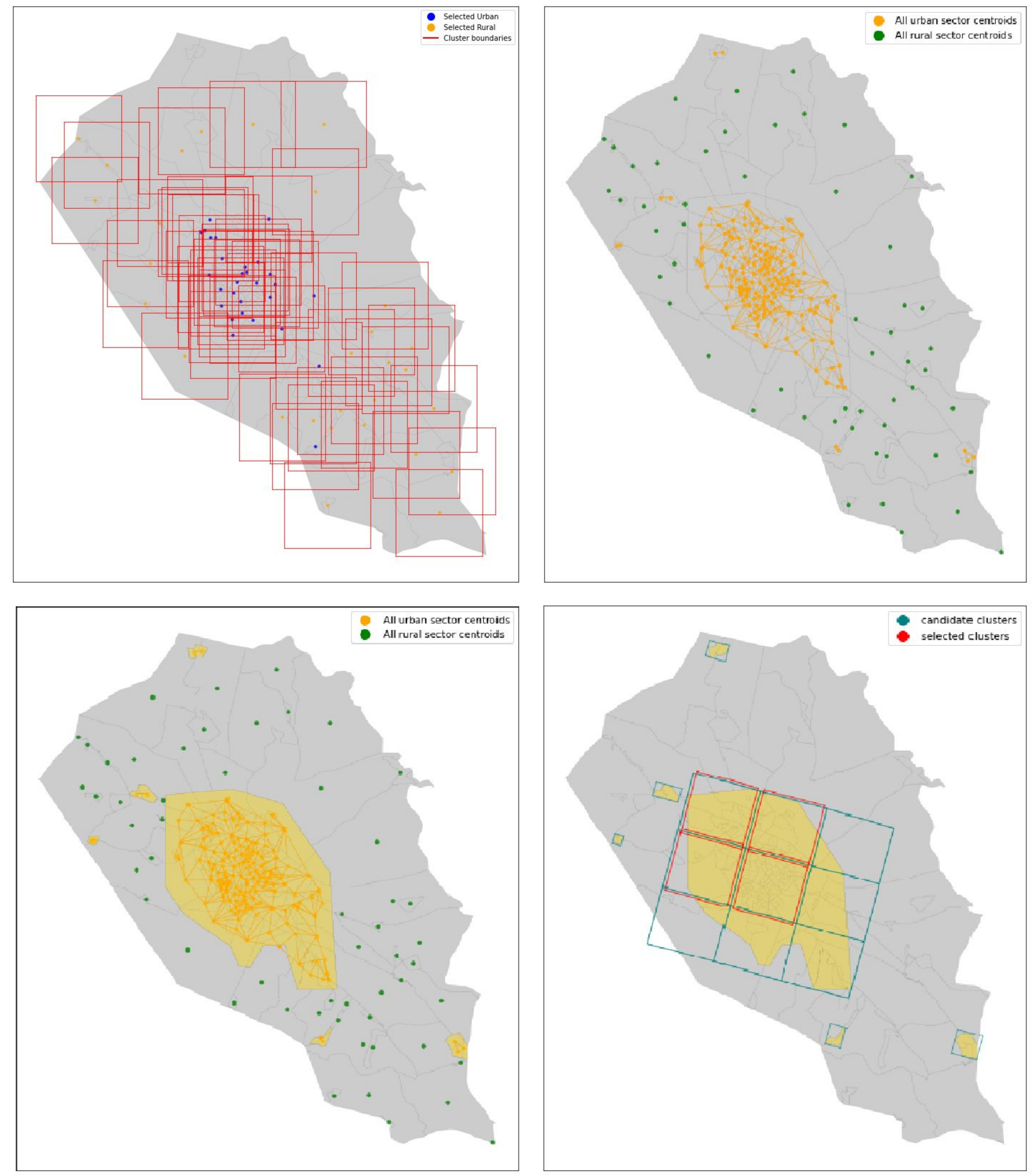

Figure 2 - Census tracts selection algorithm. Top left: Random Sampling. Other images: Neighbouring Graph workflow.

MATERIALS AND METHODS

The first step in this work is to analyze and compare the size of Brazil's census tracts with those in Africa to define the scale of the images in the dataset. IBGE provides a dataset with geometric information on the entirety of the Brazilian sector mesh, through which the distribution of tract areas is calculated. The main information to be defined are the image centroid (clusters). With a scale of 30 meters/pixel and a resolution of 224 x 224 pixels, each image covers an area 45.2 km^2. While this coverage presents almost no problem for rural tracts, the discrepancy in area of those to the much smaller urban tracts (Figure 3) imposes the problem that if careful selection of the position of urban clusters is not taken into account, the same sectors may be present in multiple images.

Random sampling of each municipality's tracts as the center of a cluster makes the problem above difficult to solve because, for many cases, the urban sectors will be so close to each other that it will be almost impossible to sample a good amount of random clusters. With that in mind, a second solution proposes to merge neighboring urban sectors into a single group and tries to fit clusters in these groups.

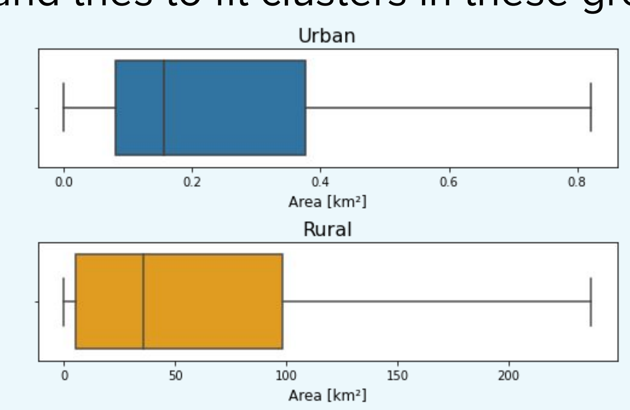

Figure 3 - Boxplot of Census Tracts areas separated by Type

RESULTS:

The urban groups are created through an algorithm that computes a disconnected graph which maps the adjacency of urban tracts in the entire municipality. Each subgraph is then merged to create a single polygon and all possible clusters are fitted into this new form. A cluster will fit the polygon if at least 70% of its area is filled with urban tracts, as shown by Figure 2.

This method is adequate since it solves the problem of having the same tract, i.e. the same information, in more than one cluster, while still allowing the selection of a good number of urban clusters, guaranteeing the proportion between clusters of both types.

References:

[1] Censo Demográfico - IBGE, Avaiable in: https://www.ibge.gov.br/

[2] Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nature Communications, 11(1), 1–11;

[3] H. Tamiminia, B. Salehi, et al. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review, ISPRS J. Photogrammetry Remote Sens., vol. 164, pp. 152-170.

Acknowledgements:

Dr. Pedro Andrade (Instituto Nacional de Pesquisas Espaciais - INPE)

Prof. Dr. Pedro Luiz Pizzigatti Corrêa (Parsec Project)

Dr. Marina Jeaneth Machicao Justo (Parsec Project)

Author affiliations: Carlos Massao Oishi Giuzio (University of São Paulo, BR); Igor Varela Zoeller (University of São Paulo, BR); Rafael Araujo Coelho (University of São Paulo, BR).