Employment topography from geocoded jobs: identification of subcenters for selected Brazilian urban areas

Marcello L. Souza Junior[1]
Alex N Almeida[2]
Carlos R. Azzoni[3]

Abstract

This paper uses nonparametric smoothing methods to identify subcenters (SBD) of four labor market areas (LMA) inland Brazil in 2006 and 2016. We use a two-stage method based on locally weighted regressions, with an endogenously determined cut-off critical value. We use matched employee/employer geocoded data from the Ministry of Labor. Given the labor market dynamics in the period, we found new subcenters in some cases, a stable situation in others, and the suppression of a subcenter in one case. Even in cases with just one center, we found growth in the area and sectoral diversification. The results reveal important structural changes in the sectoral composition of jobs in the dynamic centers. These changes constitute relevant inputs for public policies, particularly in planning the supply of public transportation and creating new interactive economic centers in these regions.

Keywords: Urban structure, Subcenters, Employment Surface, LOESS regressions
Classificação JEL: C55; O18; R12; R23.

1. Introduction

During the 20th century, large metropolitan areas experienced a restructuring of the urban space, shaped by changes in land use, population dispersion, and job distribution. This transformation was a result of strategic decisions made by firms and individuals regarding their spatial locations, leading to a shift from the traditional monocentric model to polycentric agglomerations (Schnore, 1957; Erickson & Gentry, 1985; Erickson, 1986; Roca Cladera et al., 2009).

In the context of decentralized urban development, significant technological innovations were observed in communication tools and transportation of goods and people. These innovations were complemented by centrifugal forces resulting from the effects of diseconomies of agglomeration (Anas et al., 1998). Moreover, there were active efforts to promote policies aimed at fostering the development of specific locations and understanding their subsequent impact on the concentration of economic activities in the urban area, as well as on the distribution of the population (Bollinger & Ihlanfeldt, 2003).

The interactions between market forces have led to a process of decentralization in the spatial distribution of population and jobs within metropolitan areas. As a result, new interactive economic centers (Erickson, 1986), commonly referred to as subcenters (SBDs), have emerged.

[1] ESALQ, University of São Paulo, Brazil.
[2] ESALQ, University of São Paulo, Brazil.
[3] FEA, University of São Paulo, Brazil.

These locations are typically defined as regions with a high relative density of employment, significantly influencing the occupational structure of neighboring areas (McMillen, 2001).

The understanding of polycentric urban systems and the identification of secondary centers within urban agglomerations play a vital role in guiding the development of public policies that prioritize the common interest. Some examples of such initiatives include actions to optimize infrastructure and the provision of public transportation (Kneib, 2008). Moreover, it is crucial to encourage the economic development of each location, promoting the growth of subcenters and capitalizing on productivity gains from agglomeration economies (Bollinger & Ihlanfeldt, 2003; Li & Liu, 2018).

Policy makers should also consider initiatives that anticipate the demands of productive sectors concerning infrastructure and access to specialized labor (Kneib, 2008; Campos & Chagas, 2017).

Given the benefits stemming from the identification of subcenters, studies have focused on this topic in metropolitan regions in U.S. (Giuliano & Small, 1991; Hughes, 1993; McMillen & McDonald, 1997; McMillen, 2001; McMillen & Lester; 2003), Europe (Roca Cladera et al., 2009; Veneri, 2013), China (Liu & Wang, 2016; Huang et al., 2017) and Brazil (Kneib, 2008; Kneib, 2016; Siqueira, 2014; Campos & Chagas, 2017; Nadalin et al., 2018; Campos, 2018).

This work proposes adopting a set of procedures for identifying secondary centers in urban agglomerations inland Brazil in 2006 and 2016. The strategy relies on an econometric evaluation that utilizes nonparametric smoothing methods to estimate employment surfaces in urban settlements. It specifically focuses on the populations of Americana and Santa Bárbara D'Oeste (SBO), Campinas, Piracicaba, and Jundiaí, located in the inland region of the Brazilian state of São Paulo.

It is worth noting that a significant portion of the aforementioned territorial area includes municipalities located in the Administrative Region (A.R.) of Campinas, which has gained considerable importance in the domestic economic scenario. This region has benefited from the industrial deconcentration observed in the neighboring Metropolitan Region of São Paulo. Between 2002 and 2018, the economy of the municipalities in the A.R. of Campinas experienced substantial growth of 58.5% in their Gross Domestic Product (GDP), representing 18.3% of the state GDP in 2020. The region is only behind the Metropolitan Region of São Paulo, which accounted for 53.4% of the state GDP.[4]

In recent decades, there has been a centrifugal movement towards the axis of the renowned “Anhanguera Bandeirantes” highways found in the Administrative Region (A.R.) of Campinas. This movement has played a significant role in alleviating aspects of diseconomies of scale faced by industrial activity in the capital, São Paulo, and its surrounding areas (Dias et al., 2013). Additionally, the metropolitan region of Campinas serves as an example of dispersed urban expansion, with its development extending to other population centers such as Americana and Santa Bárbara D’Oeste, Piracicaba, and Jundiaí (Bernardini, 2018). In the annex, Figure 1A and Table 1A present a list of these integrated municipalities being the objects of the present study.

[4] https://www.seade.gov.br/pib-da-regiao-de-campinas-cresce-585-entre-2002-e-2018-entenda-os-motivos/ and https://www.seade.gov.br/crescimento-do-pib-paulista-atinge-maioria-das-regioes (Access in: Jun 25th, 2022).

To date, there have been few works discussing the transformations that occurred in the urban structure of Campinas and its surroundings, particularly the expansion of the urban area guided by vectors linked to the local road systems (Caiado & Pires, 2006; Pereira, 2008; Cunha, 2016).

In this study, we geocode the data from 2006 and 2016, which comes from the mandatory Annual Social Information Report (RAIS—*Relação Anual de Informações Sociais*) produced by the Brazilian Ministry of Labor. This data cover all formally established (incorporated) organizations, both public and private, and includes workers with a labor card, making it a comprehensive census of formal workers. However, it excludes informal organizations, non-wage labor relations (self-employed, temporary work, etc.), and public sector organizations. The exclusion of public sector organizations is necessary because the locational decision process for governmental activities differs from that of the private sector and could potentially bias the results (Campos & Azzoni, 2021).

The main hypothesis of this research is that the urbanization process in the region has influenced economic agents' perception of the positive effects of agglomeration economies. It is further hypothesized that the advantages of central locations diminish in comparison to the cost reductions achieved by companies and families moving away from traditional centers, leading to the emergence of new job sub-centers. This study contributes to urban literature by providing relevant information for public policy planning and data management, especially concerning population agglomerations in the inland areas of São Paulo state.

The rest of this paper is organized as follows: The next section briefly reviews the literature related to the scope of the paper, while section 3 describes the data e methodology adopted. Section 4 presents the results. Section 6 concludes with a discussion of our major findings and some brief evidence-based policy recommendations.

2. Literature Review

The formation process of dynamic nuclei may be a reflection of incorporating old cities resulting from the urban expansion of the metropolis. Alternatively, it could be a consequence of centrifugal forces that expel firms and individuals from the traditional centers (Anas et al., 1998).

Numerous factors contribute to the structuring of urban space (Fujita, 1989). Spatial heterogeneity, for instance, refers to to the initial distribution of natural resources in a particular region, allowing for productive specialization and, consequently, the planning of the city's territory based on these natural advantages (Cronon, 1991).

On the other hand, the existence of agglomeration economies concerns the economic benefits observed by firms when they coexist in the same location. Such gains can come from locational economies guided by the existence of intra-industry trade, as well as from intersectoral interactions, promoting the dispersion of technological innovations and access to a group of qualified workers (Anas et al., 1998).

The occurrence of these factors, especially locational economies, shaped the economic and spatial growth of large cities until the beginning of the 20th century. The results are monocentric structures represented by traditional centers (concentrators of economic activities) and suburbs (Jacobs, 1969). Consequently, the formation of a subcenter takes place in a context where the agglomeration economies observed by agents are outweighed by the cost reductions in rents and transportation through the relocation to the periphery (Odland, 1978).

The microeconomic aspects related to the monocentric model were developed based on the contributions made by von Thünen to the spatial organization of agricultural production and the relationship between commodity prices, transport costs, and the value of rent of the land (Clark, 1967).

Alonso (1964) adapted von Thünen's ideas to the urban context, modeling the optimal behavior of firms and individuals in relation to willingness to pay for land use (Anas et al., 1998). According to Alonso, in short, the model predicts the existence of N citizens with identical preference structures and whose utility is derived from the acquisition of land (L) and other assets (z). However, each individual has an exogenous income (y), used to pay for the consumption of z. L is priced in terms of rent $(r(x))$ which in turn is indirectly related to the distance to the center (x) and transport costs $T(x)$. Thus, by maximizing the net income to be spent on location, conditional on a utility $\bar{u}$, one obtains the maximum rent that an agent is willing to pay per unit of space, given $\bar{u}$. This result reveals that, to maintain consumer indifferent between two or more locations, regions distant from the traditional center must have lower rent values.

Alternatively, when considering that the demand for factors of production depends on the distance from the traditional center, White (1976) discusses the rent function of firms. He reveals that the firm's willingness to pay a certain land use value is a reflection of the interaction between potential savings achieved by paying lower wages and higher expenses with freight for the flow of production as the productive activity moves towards the suburbs. Furthermore, the relationship between the slopes of the rent functions of firms and individuals defines the way in which the land will be used at each point of the urban territory (White, 1976).

Empirical studies conducted in recent decades demonstrate that the behavior of rent gradients, population density, and jobs throughout the urban territory of large metropolitan areas diverges from the results predicted by the monocentric model. This finding has contributed to the development of a new theoretical basis regarding urban spatial structure. Based on a polycentric system with the presence of multiple dynamic economic centers in a particular region, has led to the existence of cross-commuting between different subcenters (Richardson, 1988).

The de-characterization of the monocentric model is also attributed to technological advances observed in urban transport, the advent of new communication tools, and the creation of logistical access routes in the peripheral areas of large centers (Schnore, 1957; Odland, 1978; Erickson, 1986; Vicino, 2008). Studies, such as Odland (1978) and Fujita and Ogawa (1982), have sought to propose models capable of representing the conditions which an urban system becomes polycentric. Therefore, it is necessary to understand the procedures that have usually been adopted in the literature to identify these urban centers.

According to Campos & Azzoni (2020), the strategy of (i) critical values, (ii) parametric and nonparametric estimations, (iii) spatial statistics and (iv) network analysis (individual flow) are procedures commonly adopted for the identification of subcenters.

The literature attributes to McDonald (1987) the pioneering work of proposing empirical strategies to verify the existence of subcenters in large cities. Specifically, the author suggests constructing an employment density indicator as the ratio between the total number of employed people and the total area. A location is considered a subcenter if it has a higher employment density than its neighboring areas.

The classification of urban areas based on fixed reference values or thresholds from contiguous areas to analyzed subcenters can be seen as an ad hoc procedure, as the number of identified subcenters may vary significantly depending on the thresholds used (Giuliano & Small, 1991; McDonald & McMillen, 1989). Nevertheless, such a method has the potential to compromise the comparative analysis conducted between two or more metropolitan areas or cities (Redfearn, 2007).

Roca Cladera et al. (2009) emphasize the use of strategies based on parametric regression analysis. In these cases, the identification of subcenters is based on the analysis of the residuals of an econometric model, typically represented by the negative exponential function (McMillen & Lester, 2003; Wang et al., 2014).

Recognizing the limitations of parametric methods due to weak or wrong functional form specifications, several studies have explored nonparametric approaches, particularly based on Locally Weighted Regression (LWR) Models (Roca Cladera et al., 2009). An advantage of this empirical approach is the ability to estimate surfaces of employment and population density, which facilitates the identification of employment and housing peaks throughout the urban space (Roca Cladera et al., 2009). Examples of this procedure can be found in the works of Lopez (2006) and Redfern (2007) for the metropolitan regions of Barcelona and Los Angeles, respectively.

Nadalin et al. (2018) investigated whether the traditional centers of the main Brazilian metropolitan regions lost preference as job locations from 2002 to 2013. They used georeferenced employment data from the Annual Social Information Report (RAIS) and created heat maps for their analysis. Similarly, Campos (2018) adopted RAIS geocoded data for the Metropolitan Region of São Paulo in 2002, 2008, and 2016 to develop a novel empirical approach for defining subcentralities in the urban space. This approach involved using geographically weighted regressions and the critical value rules.

Spatial exploratory analysis is also adopted for identifying subcenters. It aims to establish a relationship between densities observed in a specific area and their correlation with values observed in neighboring areas (Kneib, 2008). For instance, Baumont et al. (2004) have employed this approach to analyze the city of Dijon. Siqueira (2014) identified and classified Brazilian subcenters by combining exploratory analysis of spatial data (AEDE) with cut criteria (critical values) for job density indicators.

The study conducted by Campos & Chagas (2017) developed a methodology for identifying subcenters using a two-stage approach that combined elements of exploratory analysis and spatial econometric models. The authors began with the theory of hedonic prices applied to the real estate market in the city of São Paulo, Brazil.

The few studies mentioned earlier on subcenters define them based on demographic and structural aspects, including population and employment density, total jobs, and worker-to-resident ratio. However, the urban hierarchy involves functional relationships between cities, such as the flow of information, consumers, workers, and interdependence in goods and services (Roca Cladera et al., 2009; Burger & Meijers, 2012). Bourne (1989) and Veneri (2013) used graph and social network analyses to understand urban structure, emphasizing the genuine components—relationships between cities—shaping the urban hierarchy, as opposed to relying solely on structural or demographic data.

3. Data and Methodology

The data used in this study to identify subcenters are from the formal employment locations provided by the Annual Social Information Report (RAIS - *Relação Annual de Informações Sociais*) for 2006 e 2016 (MTE, 2006; 2016). Additionally, shapefiles containing addresses of firms across the national territory are employed to create a match of census territories[5] with the population agglomerations of interest. These shapefiles are available by the Brazilian Institute of Geography and Statistics (IBGE), the Official Bureau of Statistics of the Federal Government.

Once the data has been compiled, the geocoding process of firms' locations is carried out. Addresses are converting into geographic coordinates to represent them on a cartographic surface (Siqueira, 2014; Campos & Chagas, 2017). To enable georeferencing, a preliminary step of standardizing the data describing the addresses of legal firms identified in the RAIS is necessary. Implementing such treatments helps minimize the occurrence of unsuccessful conversions. However, there were cases where latitudes and longitudes of addresses were not found. They represent a small percentage of the total firms analyzed. For instance, the georeferencing success rates were 99.49% of addresses.

The geocoding of firms within the RAIS is made possible through the construction of an R language script. This script performs a mass query of addresses in the Google Maps database using the geocode[6] function from the ggmap package. The function returns pairs of longitude and latitude coordinates.

The identification of latitude and longitude coordinates for firms in the selected periods enables the aggregation of the total number of employment relationships under each CNPJ (*Cadastro Nacional da Pessoa Jurídica* or National Registry of Legal Firms) within the respective census sectors. This, in turn, allows the calculation of job density (workers/1,000m^2) in these territorial areas. Job density is defined as the ratio between the total number of workers in the formal labor market registered in the i-th census sector and the total area of the census sector, as defined by IBGE, measured in thousand square meters.

Following Readfern (2007) and Campos (2018) our empirical strategy is based on a two-step procedure to identify subcenters by computing employment densities. Specifically, the adopted strategies are:

1) Estimation of a Locally-weighted Regression model (LOESS) from geocoded employment data;
2) Based on smoothed job density estimates, a cut-off rule based on endogenous critical values is used.

Nonparametric smoothing techniques, introduced by Cleveland (1979), are widely used in geoeconomic data analysis. These techniques are preferred for their capability to accurately depict the diverse distribution observed in variables such as: employment; population; or rent values across urban areas (Meese & Wallace, 1991; Redfern, 2007).

The local nonparametric regression method predicts the estimation of an unknown function $g(x)$ in the vicinity of each point $x = x_o$ to be analyzed. Thus, definitions are required regarding

[5] Census territories are the smallest official territorial units in the country. In this study, the definition used is the one provided by IBGE for 2010 (IBGE, 2011).

[6] For more information about the ggmap package and the geocode function, we refer to Kahle (2013) and the documentation available at http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

the coverage area (bandwidth h) of each point and the influence of neighboring observations on the centroid in question (kernel function) (Marquetti & Viali, 2004). The econometric model proposed in this study consists of a multivariate version of weighted local regression, described in expression (1) as:

$$y_i = g(x_{1i}, x_{2i}) + \varepsilon_i \qquad (1)$$

where y_i represents the employment density calculated in the i-th census sector, $g(\cdot)$ refers to an unknown functional form to be estimated from the relationship between the dependent variable y_i, while the regressors, x_{1i} and x_{2i} are the length of i-th census sector analyzed. The ε_i refers to the error term.

The selection of the bandwidth (h) is based on an Akaike Information Criterion (AIC) technique using various estimated models with different polynomial degrees for local data fitting (Campos, 2018). The local estimation process involves Ordinary Least Squares (OLS) or Tukey's Descending M-Estimator (Hurvich et al., 1998)[7]. The optimal bandwidth is determined by the smallest AIC value. Additionally, a tri-cubic function is used as a weighting criterion (McMillen & Smith, 2003; Redfern, 2007).

As a result of Eq. 1, we expect an employment density surface of dimension $k + 1$, where k represents the number of independent variables. Once the employment density surfaces have been estimated for the delimited regions, a second step is running to determine the subcenters and the census areas that make up them.

In order to eliminate the arbitrariness inherent in the approach to cutting rules, as well as the need for prior knowledge of the object of study, a decision rule based on an endogenous critical value from the distribution of $\hat{y}_i$, is adopted following Campos (2018). This step consists of an inequality, where:

$$SBD = \hat{y}_i \geq c_j \qquad (2)$$

where SBD refers to the classification of a census sector as a subcentrality, $\hat{y}_i$ the estimated employment density and c_j the critical value extracted from the model itself, where $j \in \{90, 95, 99\}$ for different quantiles of the distribution of $\hat{y}_i$. In this research, SBDs will be defined as sets of neighboring census areas that satisfy Eq. 2.

As mentioned earlier, any discussion regarding to job density, it is essential to consider firms whose operational activities are largely dispersed across the urban territory while potential biases that such occurrences can introduce in the analysis (Nadalin et al., 2018; Campos, 2018). These are the cases associated with sectors such as: Public Administration; Defense, and Social Security; Electricity and Gas; Water; Sewage; Waste Management; Decontamination Activities; Land and Air Transport; among others. They were excluded in the analysis.

[7] If the chosen function is "Gaussian," the adjustment of polynomial equations to the local data structure is done using the Ordinary Least Squares Method. However, if the option is "Symmetric," the estimations are performed using a Redescending M estimator. This nomenclature follows the input recommendations of the loess.as function from the R package fANCOVA, designed for the analysis of nonparametric models. See https://cran.r-project.org/web/packages/fANCOVA/index.html for details.

4. Results

Beginning the results with the dispersion of jobs across the urban territory of the analyzed regions, Table 1 presents descriptive statistics to the population agglomeration (PA) of interest for the years 2006 and 2016 related to: (i) the number of census areas with formal employment records; (ii) the mean employment density; (iii) the minimum value; (iv) the maximum value; and (v) the standard deviation (SD).

The data indicates a trend of jobs spreading across urban areas, reflecting the process in which centrifugal forces drive families and firms away from traditional urban centers. This finding is supported by the evolution of the total number of census areas that reported at least one formal job between 2006 and 2016. In fact, all four population agglomerations analyzed showed an increase in the number of census areas with formal jobs. Additionally, between 2006 and 2016 there was a positive evolution in the average employment density in the regions.

Understanding the distribution of job density y_i among the census sectors across population agglomerations is crucial for characterizing the urbanization process rather than relying on average values. Thus, an increase in the standard deviation of y_i suggests a trend towards employment concentration, observed in the PA of Americana-Santa Bárbara D´Oeste (SBO) and Campinas. Conversely, the reduction in the standard deviation of employment density between 2006 and 2016 in the agglomerations of PA of Jundiaí and PA of Piracicaba indicates a greater homogenization of employment locations throughout their urban areas.

Table 1. Number of census areas with formal employment records, average and maximum employment density and the respective standard deviation, by municipality and population agglomeration (PA) in 2006 and 2016.

	2006				**2016**			
Population Agglomerations	**Census areas with formal employment records**	**Average Density**	**Maximum Density**	**SD**	**Census areas with formal employment records**	**Average Density**	**Maximum Density**	**SD**
PA of Americana - Santa Bárbara D´Oeste (SBO)	531	0,22	23,91	1,88	571	0,29	48,63	2,92
PA of Piracicaba	645	0,06	66,07	3,28	690	0,07	23,07	2,57
PA of Campinas	2.209	0,25	174,34	6,39	2.441	0,37	199,78	6,83
PA of Jundiaí	934	0,13	83,81	3,32	1.028	0,20	29,29	2,56

Source: Own elaboration based on RAIS.

The job density data are crucial, but they exhibit a positively skewed distribution due to null values in the first quartile of the sample (Campos, 2018). As discussed by Campos, this outcome directly results from using the total area in the process of microlocation and employment aggregation across territorial segments, considering the entirety of space rather than just the functional areas.

As stated, subcenters are identified by estimating local regressions that smooth the distribution of employment density across the urban area. This process involves fitting polynomial functions to a selected subset of the analyzed data, with appropriate weighting.

Considering that the optimal model fit is determined by the lowest AIC value, the Population Agglomeration of Americana - Santa Bárbara D'Oeste shows estimated coefficients measured with a window of 0.2634 in 2006 and 0.3935 in 2016. This implies that 26.34% and 39.35% of the observations are considered in each local adjustment for the initial and final periods, respectively, when estimating the employment surface for this Population Agglomeration.

In the PA of Piracicaba, the topography of jobs in 2006 is obtained using a span of 0.1487. In 2016, the lowest AIC is achieved with a span of 0.3008. For the PA of Campinas, bands of 0.0625 and 0.0821 were selected for the first year and 2016, respectively. Finally, for the PA of Jundiaí, the optimal spans are 0.2594 in 2006 and 0.1814 in 2016.

Based on the models selected using information criteria (AIC), smoothed employment densities are estimated for each existing census sector. The results of these regressions are presented in Table 2 and Figure 1 for the selected PAs. The variable is the smoothed estimate of employment density $\hat{y}_i$ per 1,000m².

Table 2. Results of the locally weighted regression model estimates for the Population Agglomerations selected in 2006 and 2016.

Variables	**2006**				**2016**			
	Population Agglomerations of				Population Agglomerations of			
	Americana -SBO	Piracicaba	Campinas	Jundiaí	Americana-SBO	Piracicaba	Campinas	Jundiaí
Minimum	0.00	0.00	0.00	0.000	0.000	0.000	0.00	0.00
1st Quartile	0.38	0.18	0.24	0.17	0.43	0.18	0.41	0.23
Median	0.64	0.52	0.49	0.39	0.84	0.54	0.64	0.54
3rd Quartile	1.01	1.35	0.91	0.85	1.33	1.76	1.13	1.10
Maximum	2.54	6.88	15.66	6.08	2.55	6.34	14.52	7.07
Standard Error	1.76	2.84	5.73	3.03	2.81	2.02	6.16	2.11
Observations	623	808	2.849	1.269	623	808	2.849	1.26

Source: Own elaboration based on RAIS.
Note: Values related to the quartiles of the estimated coefficients for each census area centroid.

The illustrations in the upper part of Figure 1 include the representation of ordered triplets (y, x_1, x_2), where x_1 and x_2 refer to the longitude and latitude of the centroid of each of the 623 existing census areas in the Population Agglomeration of Americana - SBO (both in decimals). Meanwhile, y corresponds to the estimated employment density per 1,000m² obtained as a result of locally weighted regression (identified as loess in the legends of each figure).

When analyzing the job topography in the Population Agglomeration of Americana – SBO, a significant concentration of jobs is evident in the centers of Americana, Santa Bárbara, and Nova Odessa municipalities. Moreover, there has been an expansion process in terms of the area covered by peaks in formal job density until 2016, contributing to the potential aggregation of the centralities of Americana and Santa Bárbara D'Oeste.

Regarding the Population Agglomeration in Piracicaba, the measured employment surfaces only reveal concentrations of formal job links in the central regions of Piracicaba and Rio das Pedras (both in 2006 and 2016) when considering the traditional centers of the cities covered in the analysis.

As for the Population Agglomeration of Campinas, the city plays a vital role in the region's job dynamics. However, comparing the temporal evolution of employment areas between 2006 and 2016 shows distinct patterns. The ramifications extend from the global maximum point (center of Campinas) towards centralities with lower job density, including the cities of Valinhos, Indaiatuba, Hortolândia, and Sumaré.

Lastly, the Population Agglomeration of Jundiaí is significantly influenced by logistical junctions on the job topography in the region. Particularly, diagonal "corridors" arise from the highest job density observed in the central area of Jundiaí, stretching from Louveira and Itupeva on the far left of the image to Várzea Paulista and Campo Limpo Paulista during the analyzed period.

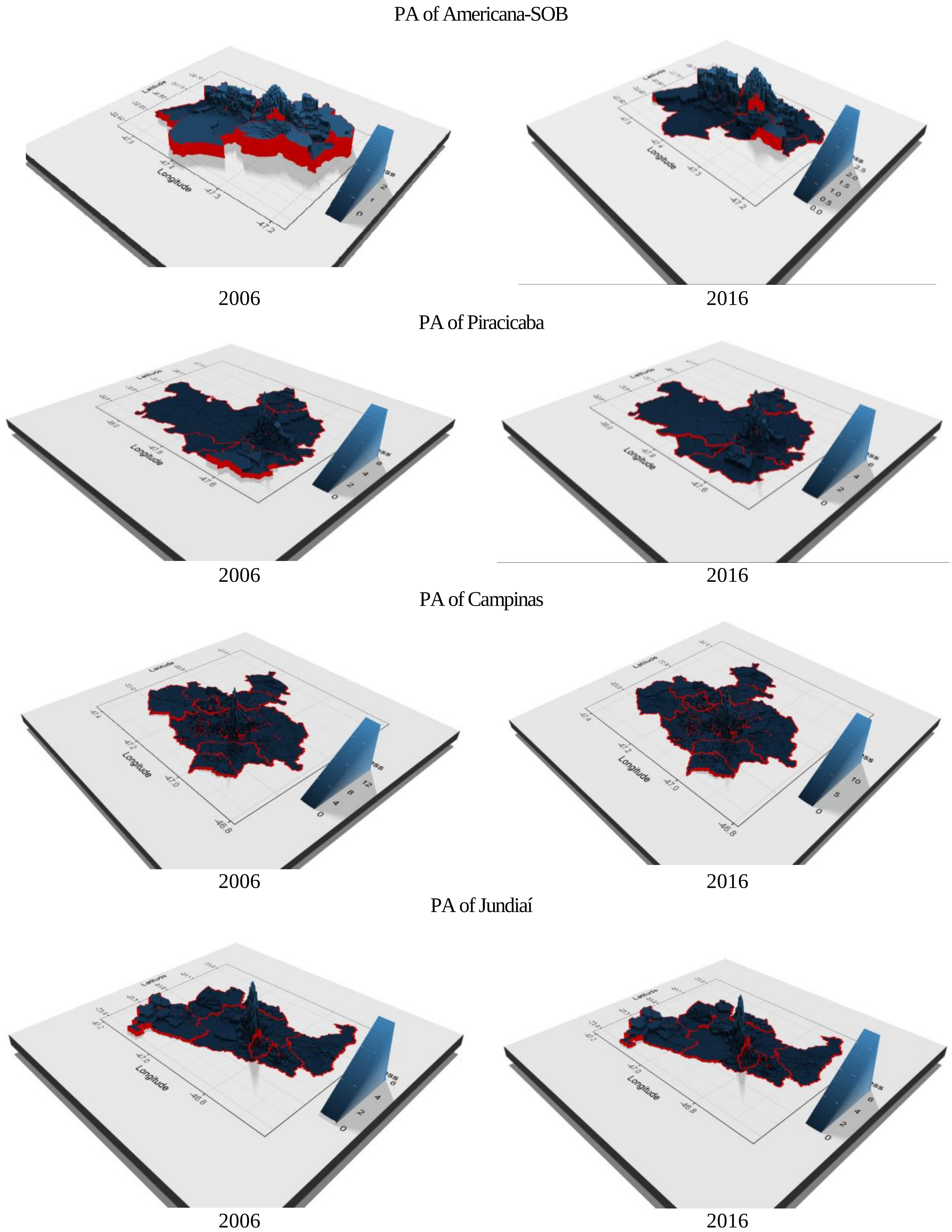


Fig. 1. Topography of jobs for selected Population Agglomerations (PA).
Source: Own elaboration based on RAIS.

After presenting the employment topographies derived from locally weighted regressions, the focus of the results shifts to identifying subcentralities using the endogenous critical value strategy, as explained in Section 3. Subcenters (SBDs) are shaped by grouping neighboring census areas that exhibit higher employment density than the cutoff statistics by the 99th, 95th, and 90th quantiles of the distribution of $\widehat{y_i}$. Table 3 shows the distribution quantiles of estimated employment densities, cut-off values, selected census sectors, size of areas (km^2) and SBDs identified in 2006 and 2016.

In the case of the Population Agglomeration of Americana - Santa Bárbara D'Oeste, the cutoff parameters for the year 2006 for the 99th, 95th, and 90th quantiles were 2.43, 2.01, and 1.49, respectively. However, in 2016, the distribution shows a more even range, leading to critical values of 2.38, 2.08, and 1.82.

Based on a more rigorous metric, the selection criteria identified 7 census areas of interest (99th quantile) in both years of the analysis. By expanding the reference value scope (quantile 90th), the total number of census sectors increased to 63. It is important to emphasize that the number of sectors remained unchanged between 2006 and 2016. However, the composition of these regions varied over time such as the number of census areas and the total area of centralities, concerning expansion or concentration. In the case of the PA of Americana – Santa Bárbara D'Oeste, a polycentric urban structure exists only in the 90th quantile. In this scenario, two dynamic centers are observed, one in Americana and the other in Santa Bárbara D'Oeste, with the latter emerging from 2016. However, for the more restricted case (99th quantile), only one center is identified in both periods. Particularly, in this restricted case, there is an increase of 5.43% in the area identified as the central region of Santa Bárbara D'Oeste.

The PA of Piracicaba had just one SBD across all analyzed scenarios, situated within the municipality of Piracicaba. From 2006 to 2016, the SBD experienced an expansion, incorporating sectors of the city beyond its conventional urban center.

On the other hand, the PA of Campinas are revealing as it was the only one to observe the exclusion of an SBD between 2006 and 2016. In 2006, the selection criteria (90th) indicated the existence of 2 SBDs, both located in Campinas. However, due to job densification during the reference decade, one of the centers expanded its geographic coverage. Consequently, there was a positive growth in the total area (km^2) of selected census sectors, despite a decrease in the number of SBDs.

Regarding the PA of Jundiaí, the selected cutoff values in both periods showed the presence of only 1 SBD in the urban area, situated in the central region of the municipality. An analysis over time reveals that the identified SBD expanded from 11.87 km^2 in 2006 (90th quantile) to 12.57 km^2 in 2016.

In two out of the four investigated locations, there is only one SBD maintained between 2006 and 2016, and it expands its area—a concept known as concentrated deconcentration (Azzoni, 1986). Our findings are aligned with Campos' (2018) analysis of the job distributions in the São Paulo metropolitan area.

Table 3. Distribution quantiles of estimated employment densities, cut-off values, selected census sectors, area of centralities and SBDs identified in 2006 and 2016.

Year	Population Agglomeration	Quantiles	Cut-off values [1]	Sector	Sector sizes (km^2)	SBDs
2006	PA of Americana – SBO	99	2.43	7	0.96	1
	PA of Americana – SBO	95	2.01	32	4.57	1
	PA of Americana – SBO	90	1.49	63	10.87	1
	PA of Piracicaba	99	6.21	9	0.38	1
	PA of Piracicaba	95	4.41	41	2.06	1
	PA of Piracicaba	90	3.12	81	4.34	1
	PA of Campinas	99	13.30	29	0.79	1
	PA of Campinas	95	5.69	143	6.29	1
	PA of Campinas	90	2.10	285	19.88	2
	PA of Jundiaí	99	5.30	13	0.92	1
	PA of Jundiaí	95	3.53	64	5.13	1
	PA of Jundiaí	90	1.97	127	11.87	1
2016	PA of Americana – SBO	99	2.38	7	1.01	1
	PA of Americana – SBO	95	2.08	32	4.40	1
	PA of Americana – SBO	90	1.82	63	10.14	2
	PA of Piracicaba	99	6.20	9	0.40	1
	PA of Piracicaba	95	5.40	41	1.64	1
	PA of Piracicaba	90	4.01	81	4.56	1
	PA of Campinas	99	13.06	29	0.78	1
	PA of Campinas	95	7.569	143	6.57	1
	PA of Campinas	90	2.89	285	20.85	1
	PA of Jundiaí	99	6.06	13	1.11	1
	PA of Jundiaí	95	4.14	64	4.87	1
	PA of Jundiaí	90	2.24	127	12.57	1

Source: Own elaboration based on RAIS.
[1] Jobs per 1.000m^2

To provide a clearer view of the identified SBDs, Figure 2 shows maps that highlight the census sectors constituting a subcentrality. The maps consider critical values from the 90th, 95th, and 99th quantiles of the distribution of $\hat{y_i}$. Readers can access these dynamic visualization maps on the Rpubs platform, as indicated in the notes below each figure.

In Figure 2, the only identified SBD in the PA of Americana – Santa Bárbara D'Oeste (SOB) for 2006 is displayed, based on various criteria. In the perspective of 90th quantile, this polygon covers 2.27% of the entire territory of Americana, Santa Bárbara D´Oeste, and Nova Odessa, yet it accommodates 26.63% of the jobs in the region. In terms of sectors, the service sector was the primary economic activity employing formal labor in the SBD in 2006, with a smaller contribution from the Textile Industry.

In 2016, two distinct SBDs (90th quantile) were identified. The first center, situated in Americana, resembled the structure observed in 2006. However, there was a notable change in the sectoral composition, with the SBD including only activities related to services.

The second SBD of the PA of Americana – Santa Bárbara D´Oeste, located in Santa Bárbara D´Oeste, focused on jobs in the sector of services. When considering both centralities, the dynamic centers were responsible for 26.94% of all formal jobs in the region's municipalities in 2016, even though they represented only 2.12% of the entire urban network.

In the Population Agglomeration of Piracicaba, regardless of the critical values used, only 1 SBD was identified in both analyzed periods. Using the most comprehensive criterion (90th quantile), it became apparent that over the ten-year reference period, the Piracicaba subcenter shifted towards the northern part of the city. Regarding the main economic activities in the dynamic center, in 2006, both the industrial sector and services coexisted. However, over a decade, the activities shifted to exclusively focus on the service sectors. Additionally, in 2016, the SBD accounted for 19.62% of the formal labor market in the municipalities of Piracicaba, Charqueada, Rio das Pedras, and Saltinho, while occupying only 0.24% of the total area.

Continuing the analysis, the Population Agglomeration of Campinas revealed two distinct SBDs in 2006 (90th quantile), with the first one situated in the expanded center of the municipality of Campinas. This area can be regarded as a dynamic center for various service-related activities.

Regarding the second identified SBD, which covers a smaller area, located along the "Dom Pedro I" highway at the junction with "Guilherme Campos" Avenue, a predominance of service activities is also observed, especially those provided by companies. Together, for the period, the two SBDs comprised 25.60% of all formal jobs registered in the municipalities of Cosmópolis, Hortolândia, Monte Mor, Paulínia, Sumaré, Valinhos, Vinhedo, and Campinas, despite representing only 1.12% of the entire territorial expanse of the region.

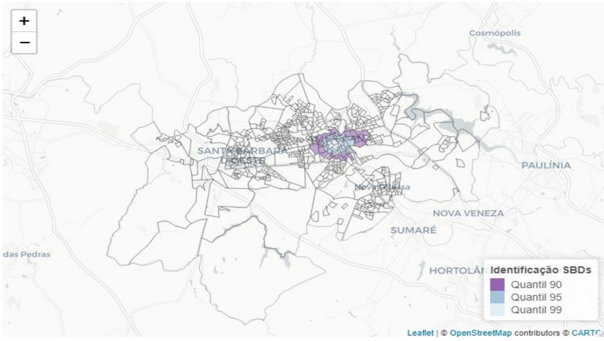
In 2016, the shifts observed in the local labor market led to the preservation of a single SBD. The resilient dynamic hub was situated in the central region of Campinas, and it continued to be significant for services and public administration.

Finally, the critical values observed for the PA of Jundiaí indicates the presence of a single SBD in the region in both 2006 and 2016, regardless of the quantile considered. The density of jobs observed in the PA contributed to a 5.89% expansion of the area of the identified subcenter within the highlighted interval based on the SBD obtained from the 90th quantile.

In terms of participation, the selected SBD accounted for 27.56% of formal jobs in the municipalities of Itupeva, Louveira, Jarinu, Campo Limpo Paulista, Várzea Paulista, Cabreúva, and Jundiaí, covering an area of 11.87 km2. By 2016, this representation decreased to 23.63% in a larger urban polygon, spanning 12.57 km2.

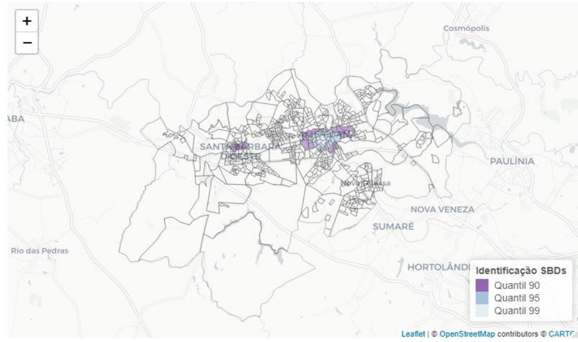
Upon investigating the composition of the primary economic activities in the region, it is evident that it remained a dynamic center for services in both periods. Particularly, in 2016, there was a more pronounced emphasis on activities related to Information Technology and services provided to companies.

PA of Americana-SOB



2006

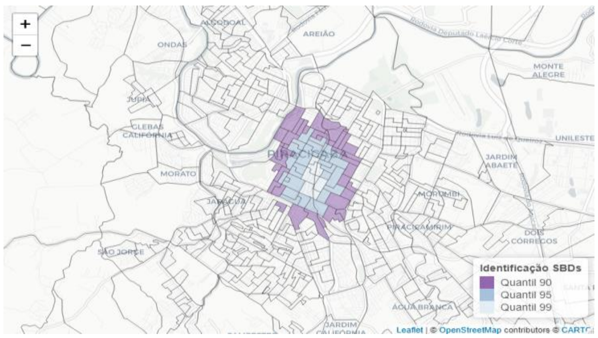
See https://rpubs.com/MarcelloSouza/sbd-ame-2006



2016

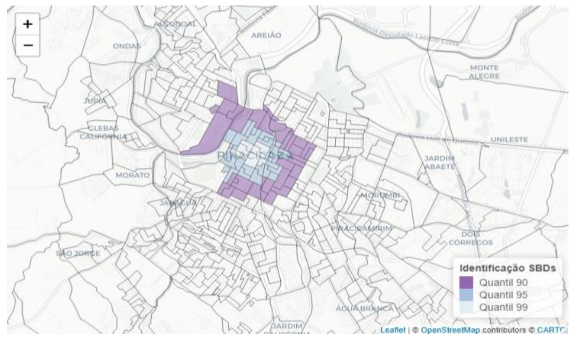
See https://rpubs.com/MarcelloSouza/sbd-ame-2016

PA of Piracicaba



2006

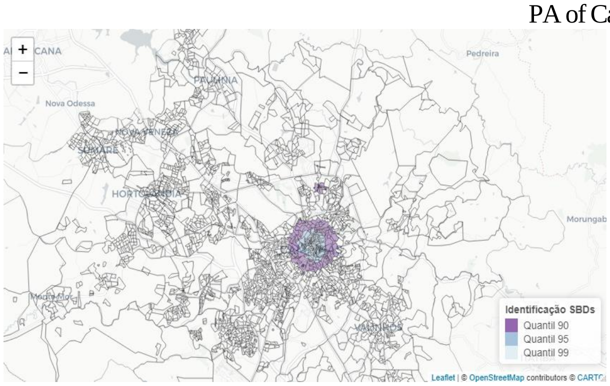
See https://rpubs.com/MarcelloSouza/sbd-pira-2006



2016

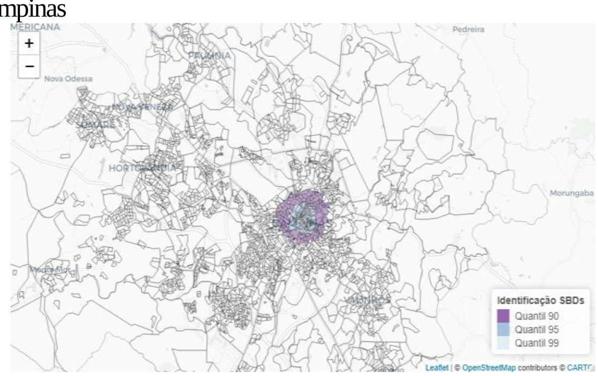
See https://rpubs.com/MarcelloSouza/sbd-pira-2016

PA of Campinas



2006

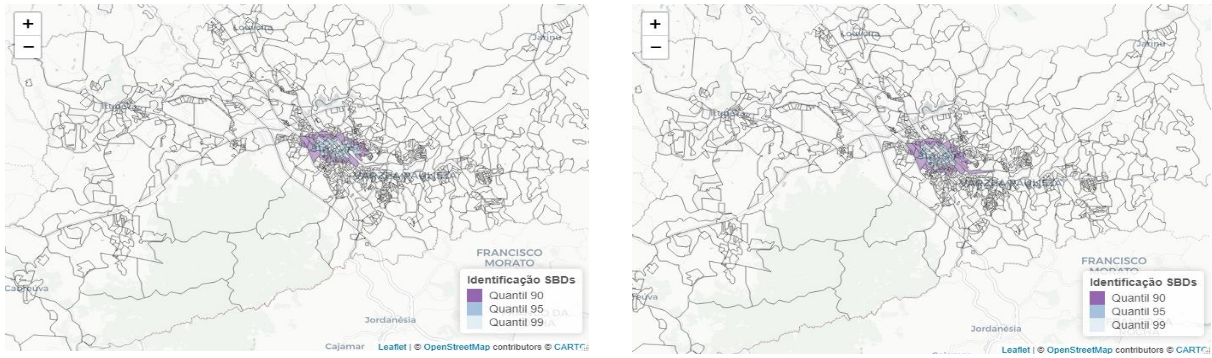
See https://rpubs.com/MarcelloSouza/sbd-camp-2006



2016

See https://rpubs.com/MarcelloSouza/sbd-camp-2016

PA of Jundiaí



2006 — See https://rpubs.com/MarcelloSouza/sbd-jund-2006

2016 — See https://rpubs.com/MarcelloSouza/sbd-jund-2016

Fig 2. Identification of SBDs for selected areas.
Source: Own elaboration based on RAIS.

5. Concluding Remarks

The intense urbanization process and technological innovations in transportation and communication during the 20th century reshaped urban systems significantly. Studies like ours, which aim to map the transformation of urban agglomerations over time, contribute to urban economics and public policy formulation. Employment densities are measured using census sectors, the smallest territorial units available, to mitigate spatial influences on the data structure. However, it was not possible to entirely eliminate biases resulting from the choice of area units, a limitation of the study.

The empirical strategy based on locally weighted regressions enables estimating employment surfaces for regions of interest, avoiding potential biases from functional model specification. The use of information criteria (AIC) and endogenous critical values proposed by Campos (2018) helped eliminate subjective interventions from the analysis.

The study focused on the Population Agglomerations of Americana – Santa Bárbara D´Oeste, Piracicaba, Campinas, and Jundiaí, important regions in the Sao Paulo state but underexplored in this context of the urban economics literature.

The results indicated the emergence of a new SBD in the PA of Americana – Santa Bárbara D´Oeste in 2016, based on the most comprehensive definition (90th quantile). For PA of Piracicaba, there was only one dynamic center in both periods, but its area expanded, and its functional structure shifted towards the services sector.

The results for PA of Campinas revealed the suppression of an SBD between 2006 and 2016. However, the surviving SBD expanded in area. As for PA of Jundiaí, employment data indicated the existence of only one SBD, which expanded its area and incorporated new activities, including information technology.

These findings provide valuable insights for public policy planning, particularly in transportation supply for improved access to the new SBD and areas recently incorporated by pre-existing subcentralities. Future studies should focus on characterizing subcenters, including employee profiles and the behavior of rent and salary vectors throughout the urban territory, particularly around the identified SBDs.

References

Alonso, W. Location and Land Use. Cambridge, MA: *Harvard U. Press*, 1964.

Anas, A.; ARNOTT, R.; SMALL, K. A. Urban spatial structure. *Journal of Economic Literature*, 36(3), 1426-1464, 1998.

Azzoni, C. R. A lógica da dispersão da indústria no Estado de São Paulo. *Estudos Econômicos* n.16, 1986.

Baumont, C.; Ertur, C.; Le Gallo, J. Spatial analysis of employment and population density: the case of the agglomeration of Dijon 1999. *Geographical analysis*, 36(2), 146-176, 2004.

Bollinger, C. R; Ihlanfeldt, K. R. The intraurban spatial distribution of employment: which government interventions make a difference? *Journal of Urban Economics*, v. 53, p. 396-412, 2003.

Bourne, L. S. Are new urban forms emerging? Empirical tests for Canadian urban areas. *Canadian Geographer/Le Géographe canadien*, 33(4), 312-328, 1989.

Burger, M.; Meijers, E. Form follows function? Linking morphological and functional polycentricity. *Urban studies*, 49(5), 1127-1149, 2012.

Campos, R. B. A.; Chagas, A. S. Identificação e caracterização das subcentralidades de emprego no Município de São Paulo. In: *45º Encontro Nacional De Economia*, Natal, 2017.

Campos, R. B. A. Subcentralidades e Prêmio Salarial Intra-Urbano na Região Metropolitana de São Paulo. *Tese (Doutorado em Economia)*, Departamento de Economia, Universidade de São Paulo, São Paulo, 2018.

Campos, R. B. A.; AzzonI, C. R. Dispersão concentrada do emprego intraurbano: modelos teóricos e abordagens empíricas. *Revista Brasileira de Estudos Regionais e Urbanos*, v. 13, p. 606-627, 2020.

Clark, C. Von Thunen's isolated state. *Oxford Economic Papers*, 19(3), 370-377, 1967.

Cleveland, W; Loader, C. R. Smoothing by local regression: principles and methods. Em: Hardle, W. Schimeck, M. G. *Statistical theory and computational aspects of smoothing*. Heidelberg (Alemanha): Physica-Verlag, p.10-49, 1996.

Cronon, W. *Nature's Metropolis*: Chicago and the Great West, NY: Norton, 1991.

Erickson, R. A. Multinucleation in metropolitan economies. *Annals of the Association of American Geographers*, 76(3), 331-346, 1986.

Fujita, M. *Urban Economic Theory: Land Use and City Size*, Cambridge, UK: Cambridge U. Press, 1989.

Fujita, M.; Ogawa, H. Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations. *Regional Science & Urban Economics*, 12:2, pp.161-96, 1982.

Giuliano, G.; Small, K. A. Subcenters in the Los Angeles region. *Regional science and urban economics*, 21(2), 163-182, 1991.

Huang, D; Liu, Z; Zhao, X; Zhao, P. Emerging polycentric megacity in China: An examination of employment subcenters and their influence on population distribution in Beijing. *Cities*, v. 69, p. 36-45, 2017.

Hughes, H. L. Metropolitan structure and the suburban hierarchy. *American Sociological Review*, 417-433, 1993.

Hurvich, C.M.; Simonoff, J.S.; TSAI, C.-L. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of the Royal Statistical Society. *Series B (Statistical Methodology)*, 60 (2), 271–293, 1998.

IBGE. *Base de informações do Censo Demográfico 2010: Resultados do Universo por setor censitário*. IBGE, Rio de Janeiro, 2011.

IBGE. *Arranjos populacionais e concentrações urbanas no Brasil*. 2ª Edição. IBGE, Rio de Janeiro, 2016.
Jacobs, J. *The economy of cities*. Random House, New York, 1969.
Kahle, D; Wickham, H. ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161, 2013.
Kneib, E. C. Subcentros urbanos: contribuição conceitual e metodológica à sua definição e identificação para planejamento de transportes. *Tese (Doutorado)*. Universidade de Brasília, Brasília, 2008.
Kneib, E. C. Centralidades urbanas e sistemas de transporte público em Goiânia, Goiás. URBE. *Revista Brasileira de Gestão Urbana*, 8(3), 2016.
Li, Y; Liu, X. How did urban polycentricity and dispersion affect economic productivity? A case study of 306 Chinese cities. *Landscape and Urban Planning*, v. 173, p. 51-59, 2018.
Liu, X; Wang, M. How polycentric is urban China and why? A case study of 318 cities. *Landscape and Urban Planning*, v. 151, p. 10-20, 2016.
Lopez, M. A. G. Estructura espacial del empleo y economias de aglomeración: el caso de la industria en la Region Metropolitana de Barcelona. *Tese (Doutorado)*. Universitat Autónoma de Barcelona, Barcelona, 2006.
Marquetti, A; Viali, L. Princípios e aplicações de regressão local. *Análise Econômica*, ano 22, n. 42, p. 253-277, 2004.
Mcdonald, J. F. The identification of urban employment subcenters. *Journal of Urban Economics*, 21(2), 242-258, 1987.
Mcmillen, D. P.; Mcdonald, J. F. Selectivity bias in urban land value functions. *Land Economics*, 65(4), 341, 1989.
Mcmillen, D.P; Mcdonald, J.F. A nonparametric analysis of employment density in a polycentric city. *Journal of Regional Science*, v. 37, n. 4, p. 591-612, 1997.
Mcmillen, D. P. Nonparametric employment subcenter identification. *Journal of Urban economics*, 50(3), 448-473, 2001.
Mcmillen, D. P; Lester, T. W. Evolving subcenters: employment and population densities in Chicago, 1970–2020. *Journal of Housing Economics*, 12(1), 60-81, 2003.
Mcmillen, D. P; Smith, S. C. The number of subcenters in large urban areas. *Journal of Urban Economics*, v. 53, nº3, p. 321-338, 2003.
Meese, R; Wallace, N.E. Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices. *American Real State and Urban Economic Association Journal*, 19, p. 308-332, 1991.
MTE. MINISTÉRIO DO TRABALHO E EMPREGO. *Relações Anuais de Informações Sociais* (RAIS). Brasília. MTE, 2006.
MTE. MINISTÉRIO DO TRABALHO E EMPREGO. *Relações Anuais de Informações Sociais* (RAIS). Brasília. MTE, 2016.
Nadalin, V. G.; Furtado, B. A.; Rabetti, M. Concentração intraurbana de população e empregos: os centros antigos das cidades brasileiras perderam primazia? *Revista Brasileira de Estudos de População*, 35(3), 2018.
Odland, J. The conditions for multi-center cities. Economic Geography, 54(3), 234-244, 1978.
Redfearn, C. L. The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area. *Journal of Urban Economics*, 61(3), 519-541, 2007.
Richardson, H. W. Monocentric vs. policentric models: The future of urban economics in regional science. *The Annals of Regional Science*, 22(2), 1-12, 1988.

Roca Cladera, J.; Marmolejo Duarte, C. R.; MOIX, M. Urban structure and polycentrism: Towards a redefinition of the sub-centre concept. *Urban Studies*, 46(13), 2841-2868, 2009
Schnore, L. F. Metropolitan growth and decentralization. *American Journal of Sociology*, 63(2), 171-180, 1957.
Siqueira, L. P. R. Proposta metodológica para a identificação de subcentros urbanos: estudo de caso na região metropolitana de São Paulo. *Economia Aplicada*, 18(1), 139-165, 2014.
Veneri, P. The identification of sub-centres in two Italian metropolitan areas: A functional approach. *Cities*, 31, 177-185, 2013.
Wang, F.; Tang, Q.; Wang, L. Post-Katrina Population Loss and Uneven Recovery in New Orleans, 2000–2010. *Geographical Review*, 104(3), 310-327, 2014.
White, M. J. Firm Suburbanization and Urban Subcenters. *Journal of Urban Economics*, 3:4 pp. 323-43, 1976.

Annex

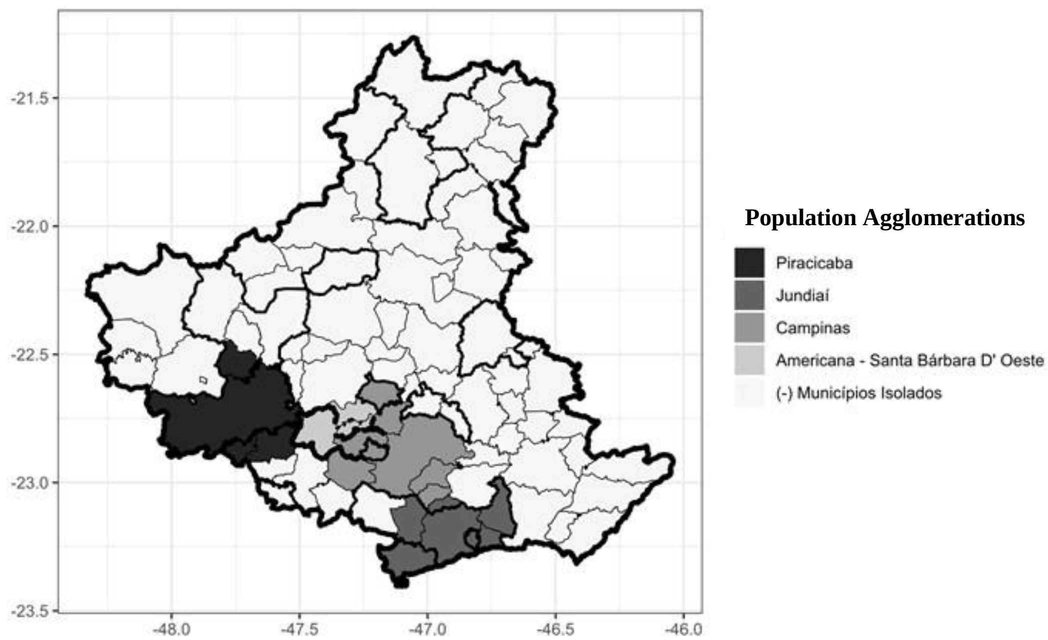


Fig. 1A. Geographical representation of the Population Agglomerations of Americana and Santa Bárbara D' Oeste, Piracicaba, Jundiaí, and Campinas, located in the Administrative Region of Campinas.
Source: Own elaboration based on IBGE (2016) definitions.

Table 1A. The municipalities that compound the Population Agglomerations (PA) of Americana and Santa Bárbara D'Oeste (SBO), Piracicaba, Jundiaí, and Campinas, located in the Administrative Region of Campinas.

Population Agglomerations	Number of Municipalities	Names of municipalities
Americana-Santa Bárbara D'Oeste	3	Americana, Nova Odessa e Santa Bárbara D´Oeste.
Campinas	8	Campinas, Cosmópolis, Hortolândia, Monte Mor, Paulínia, Sumaré, Valinhos e Vinhedo.
Jundiaí	7	Itupeva, Louveira, Jarinu, Campo Limpo Paulista, Várzea Paulista, Cabreúva e Jundiaí.
Piracicaba	4	Charqueada, Rio das Pedras, Piracicaba e Saltinho.

Source: Own elaboration based on IBGE (2016) definitions.

Acknowledgements

The first author thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES/BR) - Code 001 for the financial support.

The authors thank the Brazilian Ministry of Labor (MTB) for providing access to the RAIS database. The microdata used in this research are subjected to a Confidentiality Agreement signed between the Center for Advanced Studies in Applied Economics (CEPEA/ESALQ/USP) and the Ministry of Labor.