Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Comparative Analysis Between Methods for Determining Demographic and Socioeconomic Data in Ottobasins: A Case Study of the Paranapanema Basin

Análise Comparativa entre Métodos para Determinação de Dados Demográficos e Socioeconômicos em Ottobasins: Estudo de Caso da Bacia Hidrográfica do Rio Paranapanema

Alexei Nowatzki [1], Luciene Stamato Delazari [2] and Eduardo Vedor de Paula [3]

1 Universidade Federal do Paraná, Curitiba, Brasil. nowatzki.a@gmail.com.
ORCID: https://orcid.org/0000-0003-2634-5122
2 Universidade Federal do Paraná, Curitiba, Brasil. luciene@ufpr.br.
ORCID: https://orcid.org/0000-0003-0018-085X
3 Universidade Federal do Paraná, Curitiba, Brasil. edugeo@ufpr.br.
ORCID: https://orcid.org/0000-0002-1847-0161

Received: 07.2022 | Accepted: 01.2023

Abstract: Demographic data, such as absolute population, can be queried in spatial units of census tracts, yet socioeconomic data, such as the Human Development Index (HDI) and the Gross Domestic Product (GDP), can be provided on a municipal scale. However, watersheds frequently used in environmental studies have different limits about these units, makings it difficult to determine these data. This work aimed to determine demographic and socioeconomic data for a set of ottobasins in the Paranapanema River watershed, comprising the states of São Paulo and Paraná. For this, simple intersection and weighted area methods were used, and for population data, the dasimetric mapping technique was used, performed from land use. Finally, comparisons were made between the methods by ottobasin size grouping. The results showed that the simple intersection compromises the determination of these data, as all spatial units intersected with the ottobasins have the same weight in the calculation, even if it represents a small percentage. Therefore, the intersection of the sums and averages of weighted areas made the determination closer to reality within the context of the ottobasins. For the population data, it was observed that there is a greater variation between the methods in ottobasins with larger area values in square kilometers.
Keywords: Spatial Analysis. Water Resources Management. Dasimetric Mapping. Census Tracts. Geographic Information System.

Resumo: Dados demográficos, como os de população absoluta, podem ser consultados em unidades espaciais de setores censitários, já dados socioeconômicos, como o Índice de Desenvolvimento Humano (IDH) e o Produto Interno Bruto (PIB), podem ser fornecidos em escala municipal. Contudo, bacias hidrográficas que são usadas com frequência em estudos ambientais possuem limites diferentes em relação a essas unidades, o que dificulta a determinação desses dados. O objetivo deste trabalho foi determinar dados demográficos e socioeconômicos para um conjunto de ottobacias, na bacia hidrográfica do rio Paranapanema, compreendendo os estados de São Paulo e Paraná. Para isso foram utilizados métodos de interseção simples e de áreas ponderadas e para dados de população, foi utilizada a técnica de mapeamento dasimétrico, realizado a partir do uso da terra. Por fim, foram realizadas comparações entre os métodos pelo agrupamento de tamanho de ottobacias. Os resultados apontaram que a interseção simples compromete a determinação desses dados, pois todas as unidades espaciais interseccionadas com as ottobacias têm o mesmo peso no cálculo, mesmo que ela represente uma pequena porcentagem. Portanto a interseção a partir das somas e médias de áreas ponderadas fizeram a determinação mais próxima da realidade dentro do contexto das ottobacias. Para os dados de população foi observado que há uma maior variação entre os métodos em ottobacias com maiores valores de área em quilômetros quadrados.
Palavras-chave: Análise Espacial. Gestão de Recursos Hídricos. Mapeamento Dasimétrico. Setores Censitários. Sistema de Informações Geográficas.

1 INTRODUCTION

Generating detailed information on human populations' structure and spatial distribution is essential for decision-making in the most diverse spheres. The main source of this information is population censuses, which are complex statistical surveys made at regular intervals and carried out by almost all countries. The census is motivated by the need for demographic information to plan public policies in the medium and long term, using them in various analyses. However, census variables have limitations related to their spatial and temporal scales, compromising their use in some applications, such as the environment, civil defense, and public health (FRANCE; STRAUCH; AJARA, 2014).

Throughout its history, the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística - IBGE) has produced quite significant sets of research for the national territory, regarding the information on the varied demographic and socioeconomic characteristics of the Brazilian population, in this case, the demographic census (OLIVEIRA; SIMÕES, 2005). Duarte; Silva; Brito (2016) emphasize that its relevance to the public statistics system is indisputable from the point of view of its thematic and territorial scope. However, its complexity and size lead to challenges in ensuring the quality of its results, given the Brazilian case, whose nation lives in a territory with continental dimensions.

As pointed out by Oliveira and Simões (2005) and Duarte, Silva, and Brito (2016), the censuses carried out in Brazil have offered wide possibilities for the studies of quantification, composition, structure, and political-administrative distribution of the population, which is understood as a set of all the inhabitants of a given place, which can be a residence, census tract, municipality, state, country among others. However, Abreu et al. (2011) argue that the larger the population of a place, the smaller the real picture of human development due to the presence of different territorialities in the proposed area. Thus, the scale of work must be compatible with the human dynamics of a region, such as the areas called census tracts. According to IBGE (2021), these sectors are "constituted by continuous polygons, fully contained in an urban or rural area, whose size, number of households and establishments allow the census taker to fulfill their activities within a certain period".

In addition to counting the absolute population, a census can report data referring to the Human Development Index (HDI), proposed by the Pakistani economist Mahbub ul Haq, an indicator widely used by the United Nations (UN) organizations to measure life quality. This index is understood as a concept elaborated to obtain more data on social development besides purely economic information, considering the population's life quality as a whole. Currently, according to IBGE (2020), the calculation takes into account the relationship between three main factors: gross income *per capita* of the population, that is, the production of wealth added to the remittance of foreign exchange received by a country or territory; life expectancy or life expectancy of the population at birth and; access to education, by literacy rates, education and the number of enrollments made.

For Silva (2016) and IBGE (2020), the HDI is an index that serves, for example, as a comparison between countries to measure the degree of economic and life quality development offered to the population, and the closer to 1, the more developed the country is. This index is also used to determine the development of cities, states, and regions, given that the HDI is considered low, being below 0.5. In contrast, values between 0.5 and 0.8 are considered medium human development, and those with HDI greater than 0.8 are considered high. The MHDI can be used for municipal indicators, the Municipal Human Development Index, whose spatial cutoff is the municipal limits (UNDP, 2022).

Another socioeconomic data is the Gross Domestic Product (GDP), which measures the level of economic activity of a country, state, or municipality. Silva (2016) states that it is common to say that GDP is a good indicator of growth rather than development, which involves a qualitative transformation of the country's economic, social, and cultural structure. The GDP is calculated in several ways: one of them is by the sum of the riches produced within the country, including national and foreign companies, considering that this calculation includes the results of the final sale of agricultural, industrial, and service products. Another way to measure the GDP is to assess demand, in this case, considering household consumption, government consumption, government investment, private companies, and the sum of imports and exports (SILVA, 2016; IBGE, 2020b). Silva (2016) points out that GDP *per capita* is calculated from the division of total GDP by the number of inhabitants, which indicates how much each inhabitant produced in a given period. However, GDP *per capita* is an indicator that needs to be carefully evaluated, as the HDI is a kind of counterpoint to the Gross Domestic Product (GDP) *per capita*, in the sense that while it is concerned only with the economic dimension, the HDI encompasses, in addition to this issue, the longevity and education components, with the three levels having the same weight in the calculation of the indicator (ABREU et al., 2011).

GDP is, therefore, only a synthesis indicator of an economy, as it helps to understand a country but does not express important factors, such as income distribution, quality of life, education, and health. For the calculation of the Brazilian GDP (IBGE, 2020b), the following data are used as a basis: Balance of Payments (Central Bank); Declaration of Economic-Fiscal Information of Legal Person - DIPJ (Internal Revenue Service); Broad Producer Price Index - IPA (FGV); National Extended Consumer Price Index - IPCA (IBGE); Municipal Agricultural Production - PAM - (IBGE); Annual Trade Survey - PAC (IBGE); Annual Services Survey - PAS (IBGE); Household Budget Survey - POF (IBGE); Annual Industrial Survey - Company - PIA-Company (IBGE); Monthly Industrial Survey - Physical Production - PIM-PF (IBGE); Monthly Trade Survey - PMC (IBGE); and Monthly Services Survey - PMS (IBGE).

From the censuses, the IBGE makes socioeconomic data available for consultations at national, state, and municipal levels or even in census tracts. Thus, any other spatial unit lacks integral information, thus requiring determinations of values that it can adapt to the reality of those desired units. In this sense, France, Strauch, and Ajara (2014) highlight that census zoning often does not coincide with other regions of analysis, such as landscape units, administrative regions of education and health, or even municipal neighborhoods. Another example of a spatial unit not contemplated by the census is the watershed, widely used in water resources management studies and understood as a territory delimited by water dividers, whose watercourses, in general, converge to a single mouth located at the lowest point of the region (ANA, 2021).

In this context, the ottocoding of watersheds proposed by Pfafstatter (1989) can improve basins' management and allow greater control of actions in these areas and the consequences they can cause. As Gomes and Barros (2011) pointed out, it is a hierarchical method based on the land's topography, allowing detailing of the water system with a system of numerical digits. Gomes and Barros (2011) and Tozzi et al. (2013) highlight that among the advantages of Ottobasin (Ottobacias), as this methodology became known, one can mention the economy of characters in the code, the characteristic of having the topological information of the basin embedded in the digits and its global applicability. From this method, it is possible to optimally control the activities implemented in the region, as it is possible to quickly detect which basins are upstream or downstream of the one studied. It is still possible to identify the river basin, its main river, and its relationship with the other basins of the region. Therefore, it is a tool with great potential to be used as an instrument of water resources management.

Due to its broad applicability, the ottobasins method has been adopted by institutions and government agencies such as the National Water Agency (ANA), together with the River Basin Committees, which are entities of the National Water Resources Management System, whose representatives of the community of a basin discuss and deliberate on water resources management, sharing management responsibilities with the government, such as the Paranapanema Water Resources Management Unit (UGRH Paranapanema), which is divided into six Water Management Units (UGHs) (ANA, 2021).

In the Paranapanema river basin, the socioeconomic data available for consultation are found in spatial units of municipalities and census tracts according to the 2010 IBGE census (2012) for absolute population data. Thus, for studies of water resources management, the most used spatial units are the watersheds, sub-basins, or even ottobasins, which differs from the spatial units where these socioeconomic data are available, according to the notes mentioned above of France, Strauch, and Ajara (2014).

Thus, the main objective of this study was to determine for a set of ottobasins in the Paranapanema basin, demographics data of the absolute population and socioeconomic of HDI and GDP *per capita*, considering that for this purpose, simple intersection methods and weighted areas were used. A dasimetric mapping was also used for the population demographics data, carried out by mapping land uses present in Pirh (2016). Finally, the last step consisted of comparisons between the methods from the grouping by the size of the ottobasins.

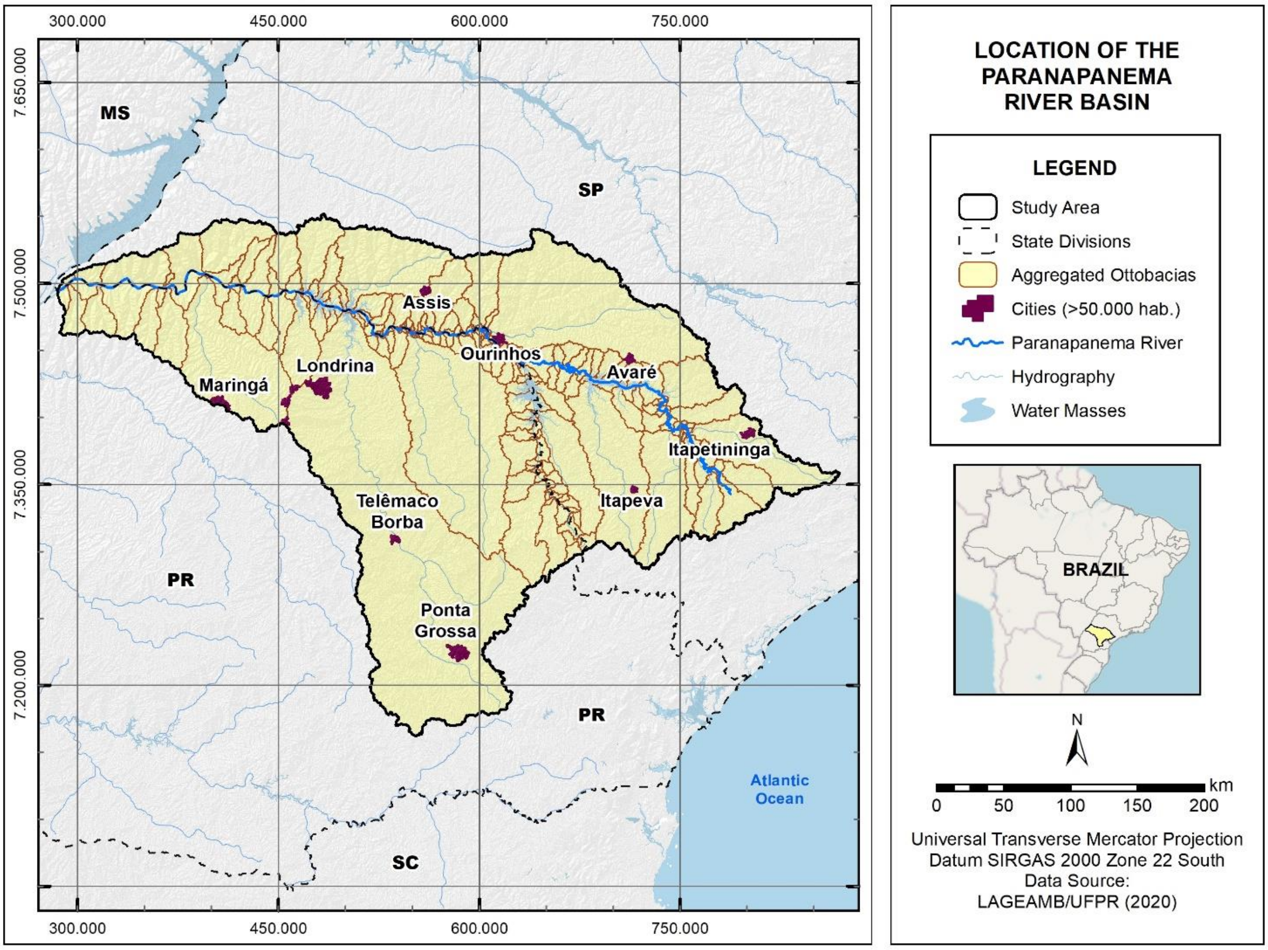
 DOI: http://dx.doi.org/10.14393/rbcv75n0a-66498

2 METHODOLOGY

2.1. Study area location

According to the Integrated Water Resources Plan of UGRH Paranapanema (PIHR, 2016), the Paranapanema river basin is located between the states of Paraná and São Paulo, in a territory of 105,921 km², covering a total of 247 Municipalities (212 with urban headquarters in the Basin) – 51% of the territory is in Paraná, with 132 municipalities (123 urban headquarters) and 49% in São Paulo, with 115 municipalities (99 urban headquarters). The total population of the Basin was estimated (from the weighting of municipal areas within the UGRH, according to data from the 2010 census) in more than 4,680,000 inhabitants, 87.5% urban and 12.4% rural, and also 62.9% in Paraná and 37.1% in São Paulo. Figure 1 shows the location of the Paranapanema River basin, especially for cities with a population above 50,000 inhabitants, according to the 2010 census.

Figure 1 - Location of the Paranapanema river basin



Elaboration: The authors (2022)

The Paranapanema River is born in the Serra de Agudos Grandes, municipality of São Miguel Arcanjo, in the southeast of the State of São Paulo, about 100 km from the Atlantic Coast and at approximately 900m of altitude, and has its mouth on the Paraná River, after covering about 900 km. In the Paranapanema River Hydrographic Basin, there are seven committees installed: three state committees in the Paranaense strand (CBHs Tibagi, Norte Pioneiro, and Piraponema); three state committees in the São Paulo strand (CBHs Alto Paranapanema, Médio Paranapanema, and Pontal do Paranapanema); and the Interstate Committee of the Paranapanema River Hydrographic Basin, which is the focus of this work, highlighting that it is an integrating committee of the basin, this being a space where representatives of the community of a river basin discuss and deliberate about the management of water resources, sharing management responsibilities with the public power (PIRH, 2016).

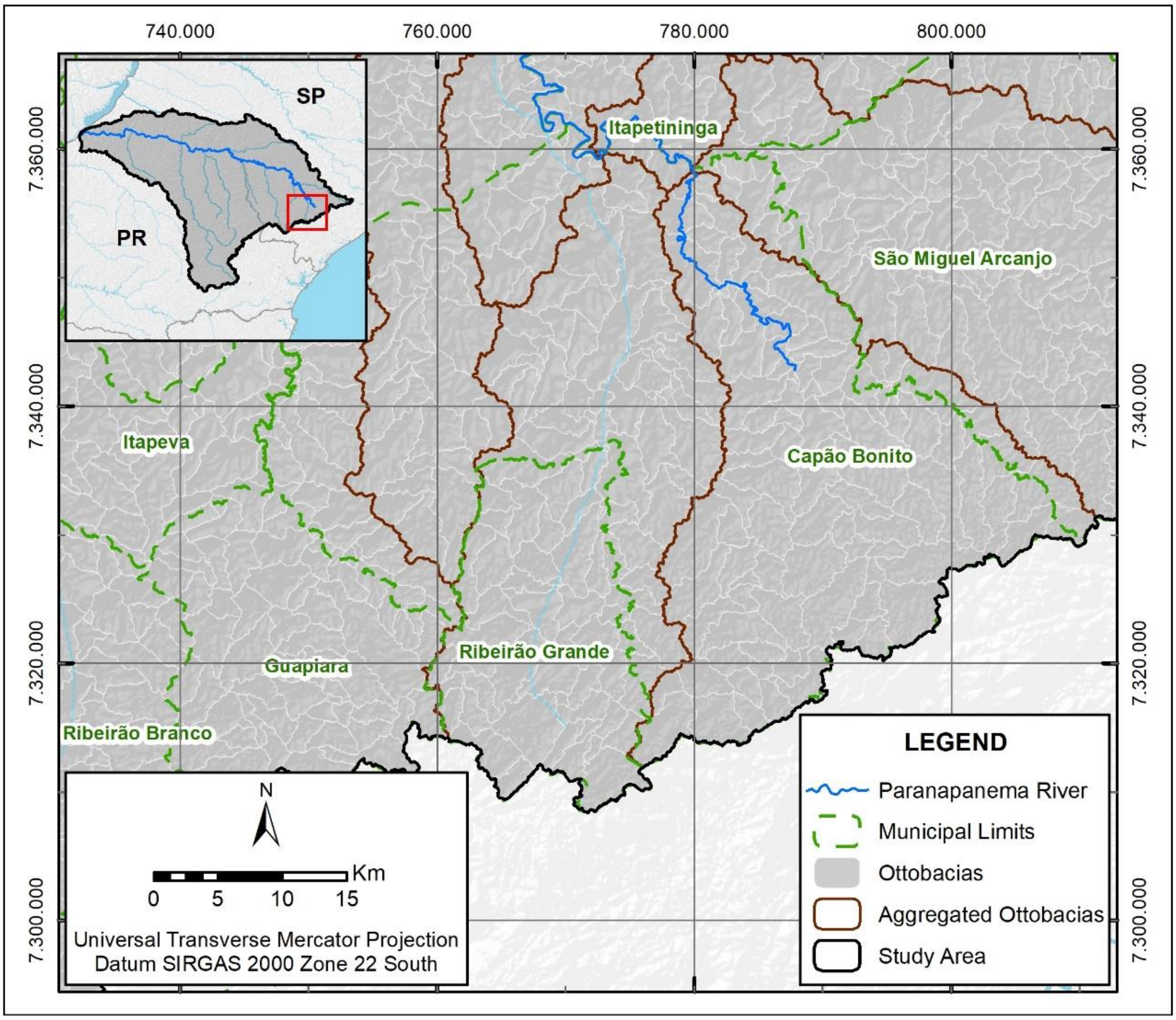
2.2. Database organization and definition of spatial analysis units

The construction of the geospatial database was based on PIRH (2016), considering that this database has data produced in the 2010 Brazilian census (IBGE, 2012). The organization and standardization of these data followed the parameters present in Paz, Dal Pai, and Paula (2020), detailed in Lageamb (2020), emphasizing the adoption of the cartographic projection of the Mercator Universal Transverse System (22 - South - Central Meridian: 51°), *Datum* SIRGAS 2000, in addition to the application of topology corrections (such as *gap* and overlap correction), through ArcGIS 10.5 software. For this work, the data of absolute population, Gross Domestic Product (GDP), and Municipal Human Development Index (MHDI) were considered (SILVA, 2016; IBGE, 2020).

In studies involving the management of water resources, the ottobasins analysis unit is used. The ottobasins are the current official division of the Brazilian watersheds and are divided into several levels (GOMES; BARROS, 2011). For this work, level 6 was considered, totaling 648 ottobasins in the study area. For simulations and modeling of watersheds used in other studies in the Paranapanema river, the sixth-level ottobasins contribute to the main river of Paranapanema and the Itararé river (tributary) were aggregated, reaching a number of 241 units, called aggregated ottobasins.

Due to the discrepancy in the representation of demographic and socioeconomic data (municipality and census tract) and limits of ottobasins (Figure 2), three possible solutions for data transposition were listed, namely: sum and arithmetic mean (simple), sum and weighted mean and also dasimetric mapping.

Figure 2 - Differences in spatial limits between the raw data and the aggregated ottobasins



Elaboration: The authors (2022)

The following items highlight the logic used to determine demographic and socioeconomic data in aggregated ottobasins. All procedures were performed using ArcGIS™ 10.5 *software*, developed by the *Environmental Systems Research Institute* (ESRI), with the *Intersect* tool available in the *Analysis Tools/Overlay* module, whose logic is to intersect two or more layers of information, since the output layer only presents what overlaps in all input layers, combining the attribute tables. The tools used in the *software* were the *Field Calculator and Calculated Geometry*, and the calculations were performed directly in the attribute table. To compare the determined data, descriptive statistical analyses were used, present in Bussab and Morettin (2017), namely: Average, Median, Mode, Standard Deviation, Interval (amplitude), Minimum Value, Maximum Value, and also the Sum (only for the results of the population data).

2.3. Sum and arithmetic mean operations

The first method used to determine demographic and socioeconomic data in spatial clippings of aggregated basins was performed through mathematical sum operations for absolute population data from census tracts and mean arithmetic operations (simple) (ANDRIOTTI, 2003; BUSSAB; MORETTIN, 2017; MOORE, 2017) for MHDI and GDP *per capita* data, both in spatial units at the municipal level. To determine the population values in each aggregated ottobasin, the intersection between the data of the sectors was performed and then summed according to Equation 1 below.

$$Ss = \sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n \tag{1}$$

Where:

Ss = simple sum; x = data value; n = amount of data

For the determinations of municipal MHDI and GDP *per capita* in aggregate ottobasins, measures of central tendency of the data expressed employing the simple arithmetic mean (Ma) (Equation 2) was used, which is the result of dividing the sum of the values of the data of the observations by the total amount of the data.

$$Ma = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{2}$$

Where:

Ma = arithmetic mean (simple); x = data value; n = amount of data

2.4. Weighted sum and weighted average function-weight operations

The intersection between these layers of information was also performed for the calculations of weighted areas of population values, MHDI, and GDP per capita resampled in aggregated ottobasins. However, by weighted logic, it was first necessary to obtain the area value of each municipality in the intersected aggregate ottobasins (UMBELINO; BARBIERI, 2008; MACEDO; MAGALHÃES JR., 2010). As described in Bussab and Morettin (2017), this percentage (P) is the ratio between the area value of the municipality after the intersection with the ottobasins and the total area value of the municipality. The calculation of the percentage for the weighting by area is determined by Equation 3, in which the maximum value that can be reached is 1, that is if the municipality is in its entirety within the aggregate ottobasin.

$$P = \frac{Aoa}{Atm} \tag{3}$$

Where:

 DOI: http://dx.doi.org/10.14393/rbcv75n0a-66498

P = percentage of the area of the municipality in the aggregate ottobasin; Aoa = area of the municipality intersected with the aggregate ottobasin; Atm = total area of the municipality.

For the population data, the logic of the weighted sum (Sp) of all absolute population values of the census tracts contained within the aggregate ottobasin in question was used, according to Equation 4.

$$Sp = \sum_{i=1}^{n} x_i \times P_i = (x_1 \times P_1) + (x_2 \times P_2) + \cdots + (x_n \times P_n) \quad (4)$$

Where:

Sp = weighted sum of the intersected census tracts with the aggregate ottobasin; x = value of the data; P = percentage of the area of the municipality in the aggregate ottobasin

Regarding the data of HDI and GDP *per capita*, sampled at the municipal level, it was necessary to stipulate the weighted average (Mp) from the raw values of each of the municipalities intersected with the aggregated ottobasins, given by Equation 5:

$$Mp = \frac{\sum_{i=1}^{n} x_i \times P_i}{\sum_{i=1}^{n} P_i} = \frac{(x_1 \times P_1) + (x_2 \times P_2) + \ldots + (x_n \times P_n)}{P_1 + P_2 + \cdots + P_n} \quad (5)$$

Where:

Mp = weighted average of municipalities intersected with the aggregate ottobasin; x = value of the data; P = percentage of the area of the municipality in the aggregate ottobasin

2.5. Population data by dasimetric mapping

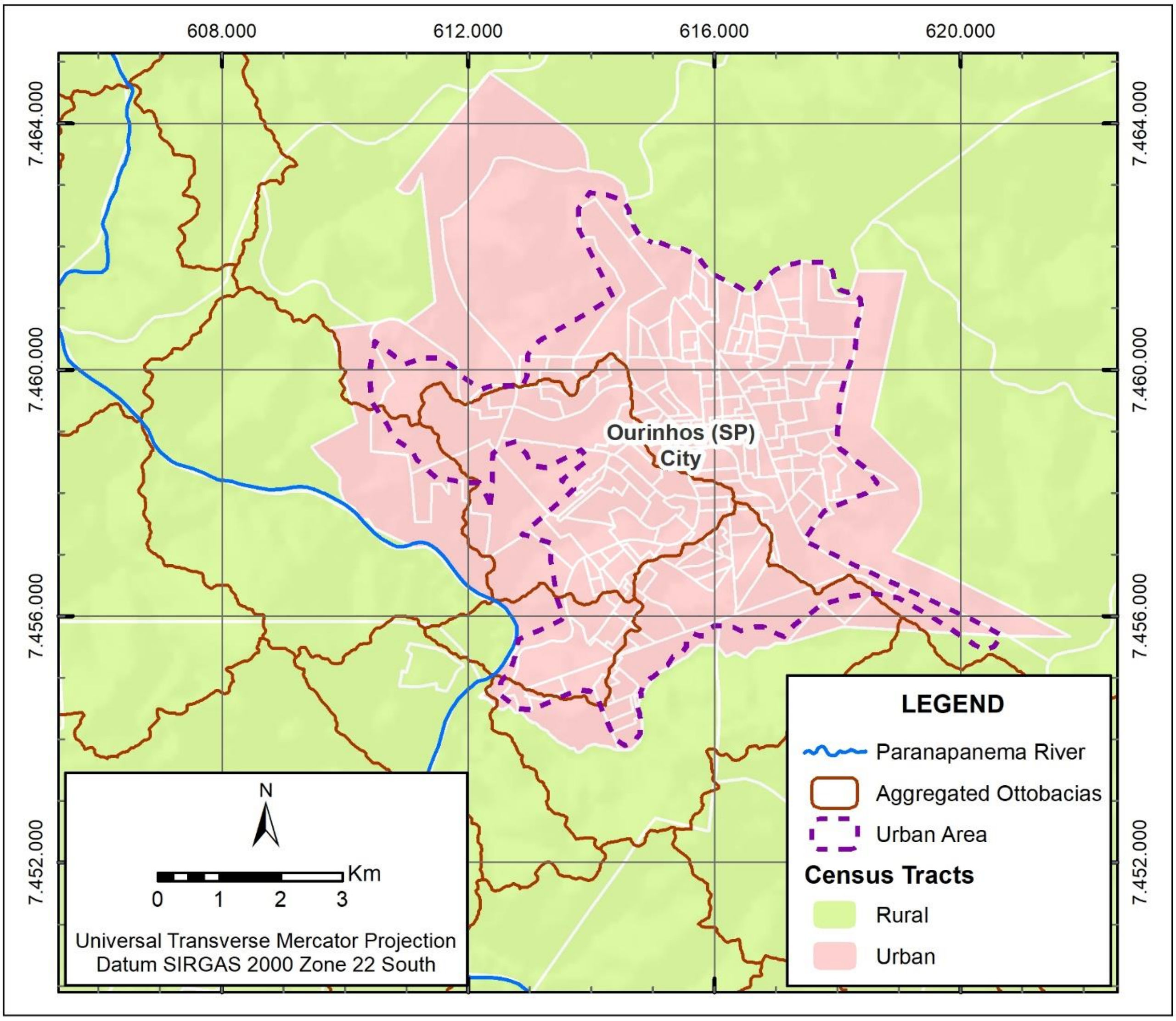
According to France, Strauch, and Ajara (2014) and Morais Junior and Silva (2019), a dasimetric mapping can be understood as a specific type of zonal interpolation based on a cartographic technique that uses auxiliary information to disaggregate demographic variables such as total population and population density. Recently, several approaches to the dasimetric method have been developed, refined, and compared at different scales, with land use classes derived from satellite images as traditionally used information (MOON; FARMER, 2001; MENNIS; HULTGREN, 2005; WEBER, 2010),

It is noteworthy that although it is used almost exclusively to represent and analyze the resident population or population density, dasimetric interpolation can be applied to disaggregate any quantitative variable, provided that a set of auxiliary data strongly correlated with this variable is available (FRANCE; STRAUCH; AJARA, 2014; MORAIS JUNIOR; SILVA, 2019).

Neves, Strauch, and Ajara (2017) and Santos, Holmes, and Ramos (2018) highlight that this type of mapping employs two classes of land use to differentiate areas that are inhabited by human populations, considering that traditional choropleth maps equally fill the entire length of a given observation unit, that is, it considers where the population resides within the census polygon or the municipality, the rest is considered as an area of exclusion.

Initially, urban-class census tracts were considered to perform the dasimetric mapping, as shown in Figure 3. However, even with this IBGE denomination, it was observed that urban areas do not fully coincide with these sectors. The Vegetable Coverage and Land Use map, made available by the Paranapanema River Basin Committee (PIRH, 2016), was used to select urbanized areas. This mapping was carried out by acquiring images from the LANDSAT 8 satellite of 2014, using bands from 2 to 7 and a spatial resolution of 30 meters. The delimited classes are those of (in order from largest to smallest area): Pasture, Temporary Culture, Forestry, Silviculture, Rural, Lakes, dams and water bodies, Urbanized Area, and Permanent Culture.

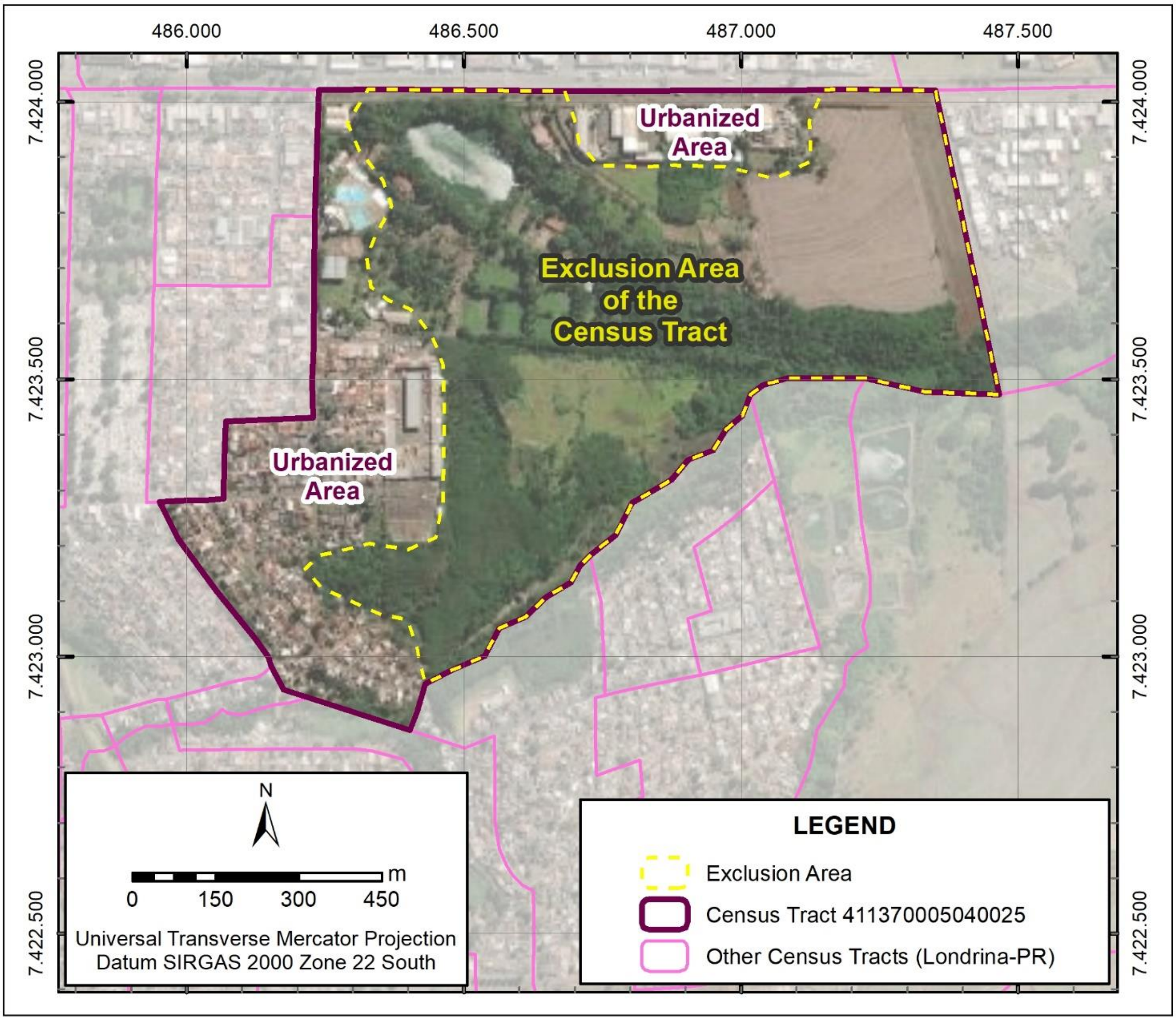
Figure 3 - Differences in spatial limits between census tracts, urban areas, and aggregate ottobasins



Elaboration: The authors (2022)

Figure 4 represents an urban census tract in the municipality of Londrina, in Paraná, whose area is mostly occupied by non-urban land uses, such as agriculture and areas with natural vegetation. These data may influence the determination of population values of aggregated ottobasins. If a watershed separates this sector, the urbanized areas may be in only one of the ottobasins. Therefore, this mapping type is chosen to compare the data generated for the entire census tract. The source of the geospatial data was IBGE (2012), PIRH (2016), and ArcGIS 10.5 (Basemap - Imagery) – GeoEye01 – Resolution: 0.46 m – Date: 14/08/2016.

Figure 4 - Urbanized and exclusion area for dasimetric mapping in the census tract of code 411370005040025 in the municipality of Londrina (PR)



Elaboration: The authors (2022)

2.6. Selection of comparative analyses between methods of determining demographic and socioeconomic data

From the generation of the data considered in this study, it was observed that there were differences in the values generated between the simple and weighted methods. They varied according to the area size of the aggregated ottobasins, especially concerning population data. First, the difference (amplitude) was made between the results of the determinations for each aggregate ottobasin (subtracting the highest value from the lowest). Then, the ottobasins were grouped according to the size of their areas. The division of classes (break values) was performed using different classification methods: the standard deviation, natural break (*Jenks*), quantile, and geometric interval. The method adopted was the geometric interval, as it allows better visualization of data that do not have normal distribution or have the distribution too distorted (SLOCUM et al., 2009; SILVEIRA, 2019).

As Silveira (2019) points out, the geometric interval method is similar to a progression classification (binary, geometric, or logarithmic) but with the addition of a coefficient. Class breaks, in turn, are based on class intervals with a geometric series. Said classification denotes greater variation in the data due to class breaks occurring in a constant geometric increase of the interval preceding the breaks – it can be double (2, 4, 8…) or triple (1, 3, 9…). The values of class breaks are highlighted in Table 1, along with their areas. It should be noted that the area values in square kilometers were rounded from the histogram analysis.

 DOI: http://dx.doi.org/10.14393/rbcv75n0a-66498

Table 1 – Division of classes by aggregate ottobasin size by geometric intervals

CLASS	AREA (KM^2)	AGGREGATE OTTOBASIN SIZE
1	>10	Very small
2	10-70	Small
3	71-500	Avg.
4	501-3500	Big
5	3500<	Too large

Elaboration: The authors (2022)

After performing this division into classes, arithmetic means (according to Equation 2) of the amplitude values of the size classes of the ottobasins were performed to observe which ones had the greatest differences between the methods used in the present work.

3 RESULTS AND DISCUSSION

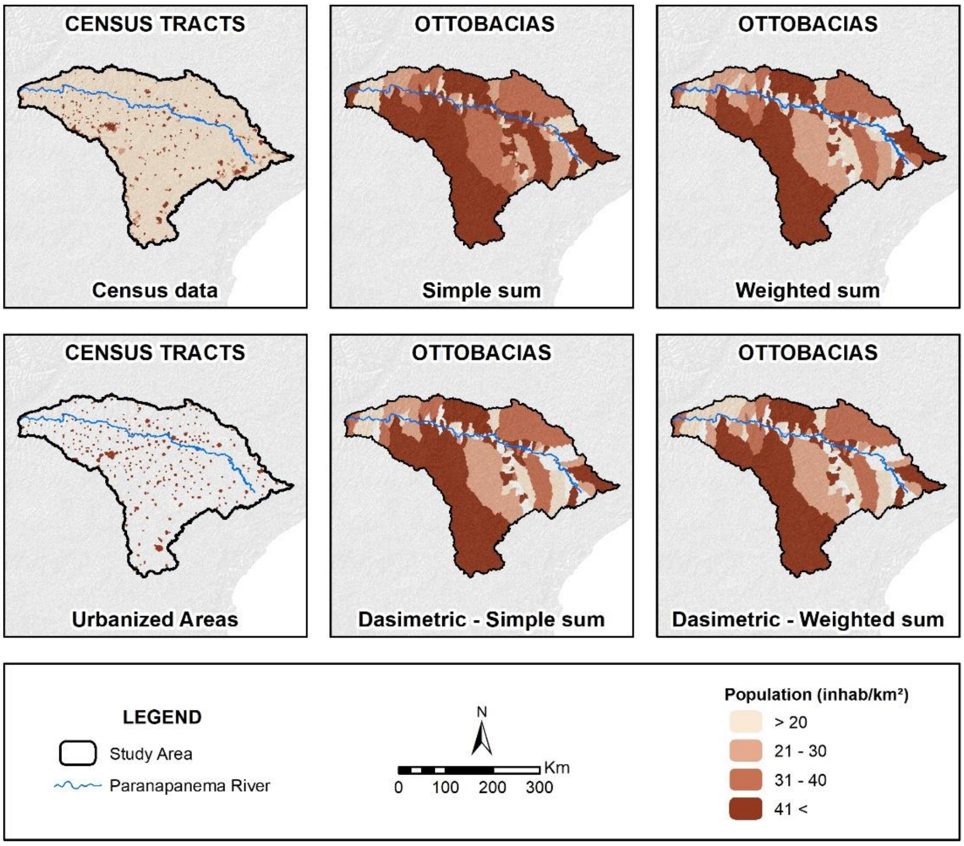
3.1. Spatialization of demographic and socioeconomic data in aggregated ottobasins (ottobacias) and discussions on their applicability in water resources management

According to the surveys carried out by the 2010 census by IBGE (2012), Figure 5 was prepared from the absolute population data to illustrate the population density (ratio between the absolute population and the square kilometer of the census tract) of the census tracts present in the Paranapanema river hydrographic basin. It was decided to illustrate the population density in the figures of this work, and not in the absolute population data itself, due to the question of its cartographic representation as a function of the area of the ottobasin because in choropleth maps, the data must be represented in the form of rates or ratios. In total, 8490 census tracts were considered, and their population values were resampled for the aggregated ottobasins under study through simple and weighted sums. Figure 5 further illustrates the dasimetric mapping performed.

It is observed that the determination of the data by the census values, considering all census tracts (rural and urban), generally presents values higher than those determined by dasimetric mapping. This information corresponds to what is expected, considering that for this type of mapping, only those sectors inserted within the polygons of the urbanized area of the land use map are considered. It is also possible to highlight that in both situations, the weighted sum has lower values than the simple one since each sector intersected with the ottobasin is considered in the equation of spatial analysis. The details of these measures can be seen in Table 2.

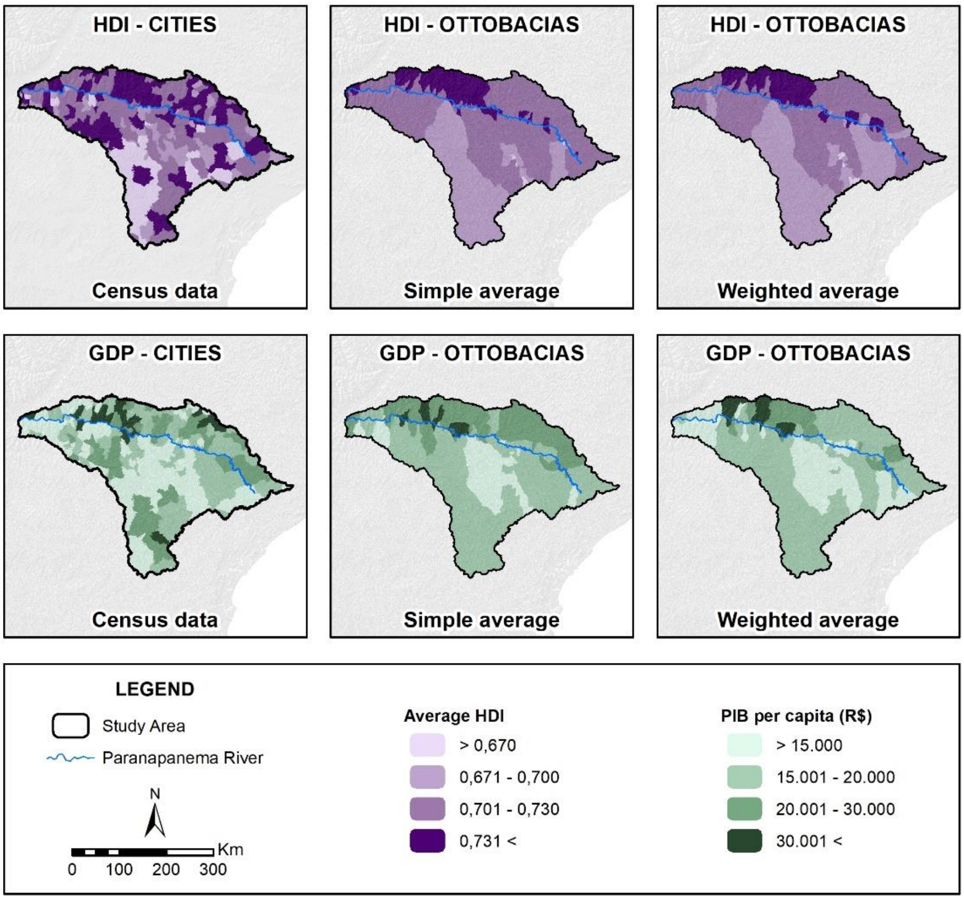
Figure 6 represents the resampling of the 247 municipalities covered in the Paranapanema basin for the data of average MHDI and GDP *per capita* in reais (R$). It is noted that in Table 2, the simple arithmetic means, for both cases, resampled values higher than the weighted means, in the same way as the weighted sums presented in the determination of population data.

Figure 5 - Determination of demographic data in aggregated ottobasins (ottobacias)



Elaboration: The authors (2022)

Figure 6 - Determination of MHDI and GDP *per capita* in R$ in aggregate ottobasins (ottobacias)



Elaboration: The authors (2022)

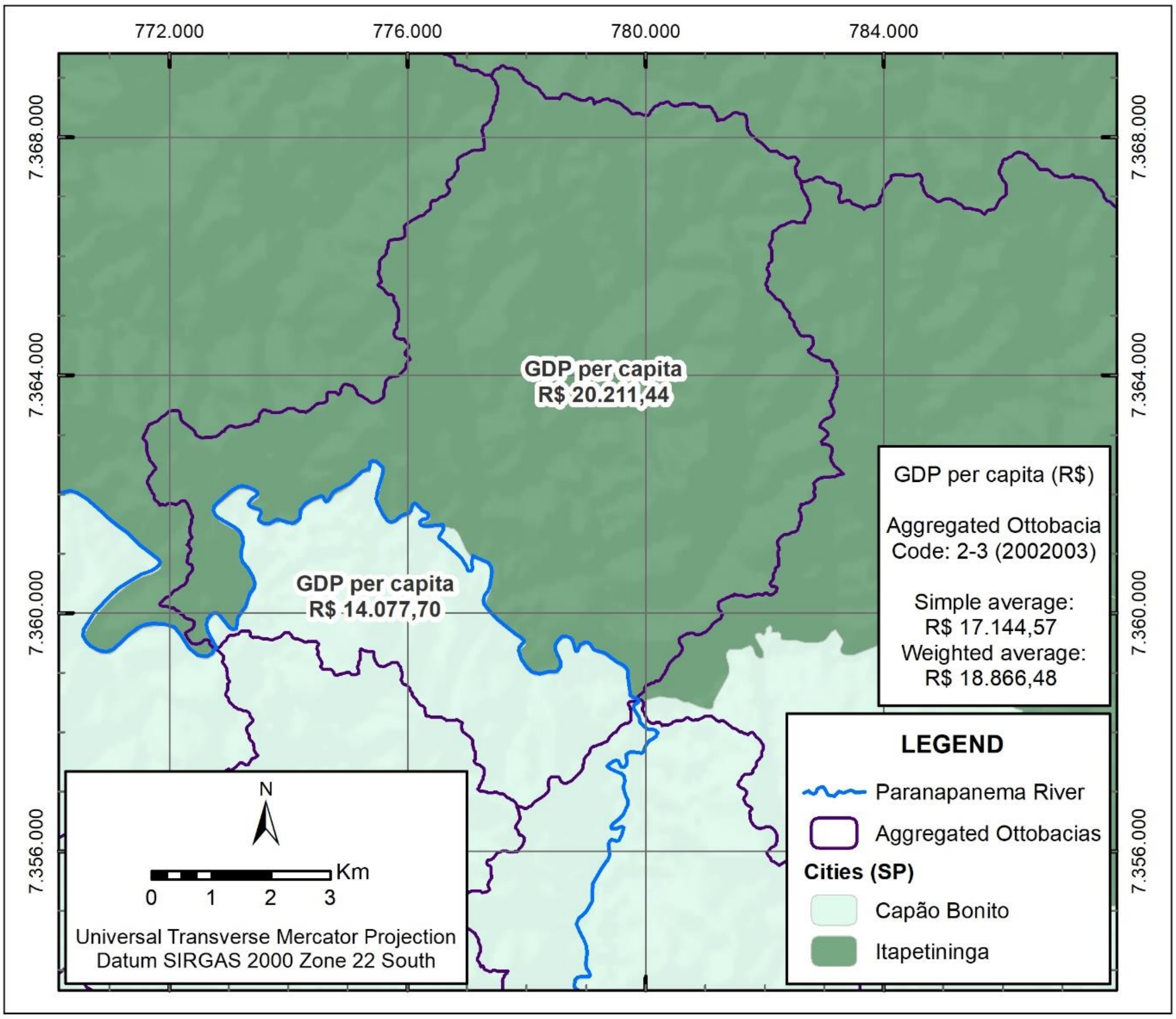
Table 2 - Synthesis of the results of the determination of demographic and socioeconomic data in aggregated ottobasins (ottobacias) by descriptive statistical analysis

SOCIOEC. DATA	ABSOLUTE POPULATION				MHDI		GDP *per capita* (R$)	
METHOD	Ss (Das.)	Sp (Das.)	Ss	Sp	Ma	Mp	Ma	Mp
Arithmetic Mean	18.276	16.856	22.365	18.981	0,718	0,716	19.545,54	19.440,82
Median	0	0	2.079	370	0.719	0,716	17.284,91	16.733,51
Mode	0	0	319	1	0,723	0,729	14.285,39	19.377,60
Standard deviation	108.806	105.361	121.898	117.314	0,022	0,024	6.741,57	8.448,59
Break	1.500.531	1.463.174	1.690.049	1.640.810	0,117	0,117	31.664,30	46.427,23
Minimum	0	0	125	0	0,662	0,661	10.318,92	9.541,07
Maximum	1.500.531	1.463.174	1.690.174	1.640.810	0,778	0,778	41.983,22	55.968,30
Sum	4.404.407	4.062.330	5.389.845	4.574.387	-	-	-	-

Legend: Ss = Simple Sum; Sp = Weighted Sum; Ma = Arithmetic Mean; Mp = Weighted Average; Das. = Dasimetric Mapping. Elaboration: The authors (2022)

Figure 7 illustrates these differences through the two methods. For the selected ottobasin, two municipalities are contemplated, and one contains a larger occupied area than the other. In a simple average determination, the value is equal for the two municipalities. In contrast, in the weighted average, the municipality with the highest prevalence has a greater weight in the equation, influencing the result.

Figure 7 – Example of determining GDP *per capita* data using different methods



Elaboration: The authors (2022)

Also, in Table 2, it is possible to highlight that in addition to the average values, the median and mode of the three socioeconomic data were also lower for the sums and weighted averages, except for the population determined by dasimetric mapping, whose median and mode result was equal to zero since the ottobasins that did not have an urban area were considered to have a non-existent population. The total population density of the Paranapanema basin was also calculated, resulting in the values for inhabitants per square kilometer: 41,6 (Ss Das.); 38,4 (Sp Das.); 50,9 (Ss) and; 43,2 (Sp).

For Silva, Morato, and Kawakubo (2013), demographic density is mapped using the population values present in the census tracts, divided by the area covered by the sector (a method known in thematic cartography as choroplethetic). The problem with this approach is that this population density calculation procedure often masks the actual population density, as non-residential areas are included in the calculation, causing the smoothing effect. An alternative method for calculating population density is the dasimetric method. The dasimetric methodology proposes a more realistic description in which the areas/rates are modified according to the homogeneity criterion obtained through complementary information. However, it was observed that this logic might be more accurate in spatial analyses with scales of greater detail, such as the behavior of the population within the census tracts. In determinations for areas of large watersheds, as in the case of the Paranapanema River, where there are rural census tracts with large numbers of residents, it can negatively impact the results, determining absolute population values lower than reality.

In the 1980s, Prochnow (1988) mentioned that data gaps negatively impact watershed management. With the release of the 2000 Demographic Census, when several socioeconomic data were made available, was already highlighted by Umbelino and Barbieri (2008) and Macedo and Magalhães Jr. (2010), the challenge of researchers to reconcile distinct spatial units and aggregate data to physical clippings that grew in importance, such as the watershed. In general, in studies that adopt the river basin as a unit of analysis, socioeconomic data are used inappropriately to encompass, respectively, areas larger or smaller than the actual area of the river basin. Frequent analyses use data and statistical information from the municipalities that make up the basin without these being entirely contained in these watersheds (OLIVEIRA; MAGALHÃES JR., 2013). Other studies detail the information a little more and use data from the census tracts, but to make the cut out of the census tracts compatible with the boundaries of the basins; these studies divide the census data in proportion to the percentage (weighted logic) of the sector within the hydrographic basin (UMBELINO; BARBIERI, 2008; MACEDO; MAGALHÃES JR., 2010). It is still possible to consider the preponderant municipality or census tract within the basin (or ottobasin) so that the total value of that unit of analysis is considered.

Oliveira and Magalhaes Jr. (2013) point out that the determinations by weighted areas, although more precise, also do not allow reaching a degree of certainty, given that this division can oversize or undersize the selected data since the spatial distribution of the occupation of the population in the sectors is not homogeneous. For these authors, if the current methods of collecting and disseminating socioeconomic data do not generate information compatible with the cutout of watersheds, the degree of accuracy of these data cannot even be determined. It is impossible to know the population of watersheds by a reliable and accurate method, only with indirect statistical determinations, as in the case of sums and weighted averages. Even with the inaccuracy of municipal data transposed to the basins, it is noteworthy that these are widely used to guide policies and decisions on water resources in the country.

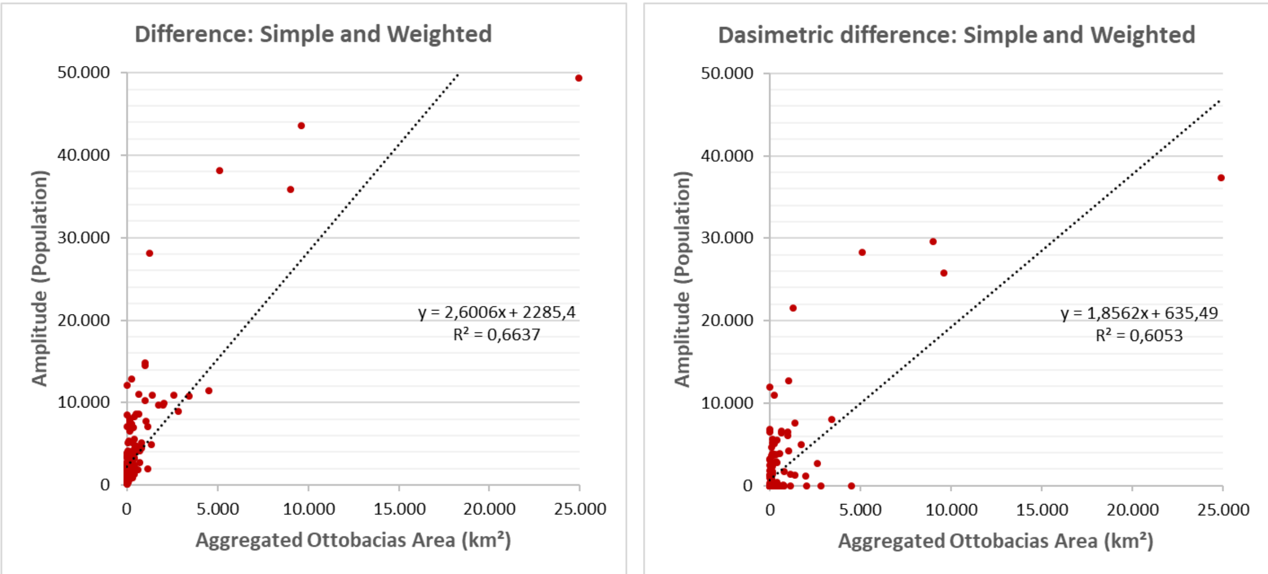
From the knowledge of each basin's reality, it is possible to build more realistic diagnoses and scenarios through more elaborated data and indexes, basing decision-making processes and management policies appropriate to each specificity. With more accurate information, managers can make more appropriate decisions considering that the efficiency and optimization of monitoring processes, inspection, and financial investments, in general, also benefit from the more transparent exposure to the physical and socio-environmental realities of watersheds. Oliveira and Magalhaes Jr. (2013) propose that there is a consideration and possible inclusion of watersheds, or even ottobasins, in the planning of the cartographic network of IBGE census tracts.

However, it is pertinent to point out that several studies have been done to reach reliable estimates for determining data in different spatial cutouts. The case of IBGE (2016) can be cited, which launched the question of the Statistical Grid, which is a system of regular grids used for the dissemination of census data, which was precisely thought to allow the integration of data from different sources, as well as being adaptable to various types of spatial cutouts and which can be used in works involving watersheds, or as Pedro and Queiroz Filho (2017) demonstrated in studies evaluating the statistical grid in subnormal clusters.

3.2. Differences in values between methods of determining demographic and socioeconomic data

The first analysis in Figure 8 allows us to perform the size of the aggregated ottobasins (abscissas) about the amplitude of the population data values between the two methods of simple and weighted sum (ordered). It was found that there is a tendency that the larger the area of the ottobasin, the greater the amplitude between the methods. This finding can be applied both to the determination of population data by census tracts and to the dasimetric mapping data, observed both by the linear trend line and by the value of the coefficient of determination (R^2 or R-square), whose results showed values close to 0.6, considering that the higher the R^2, the more explanatory the linear model is, that is, the better it fits the sample.

Figure 8 – Dispersion diagrams of the population data from the amplitudes between the methods

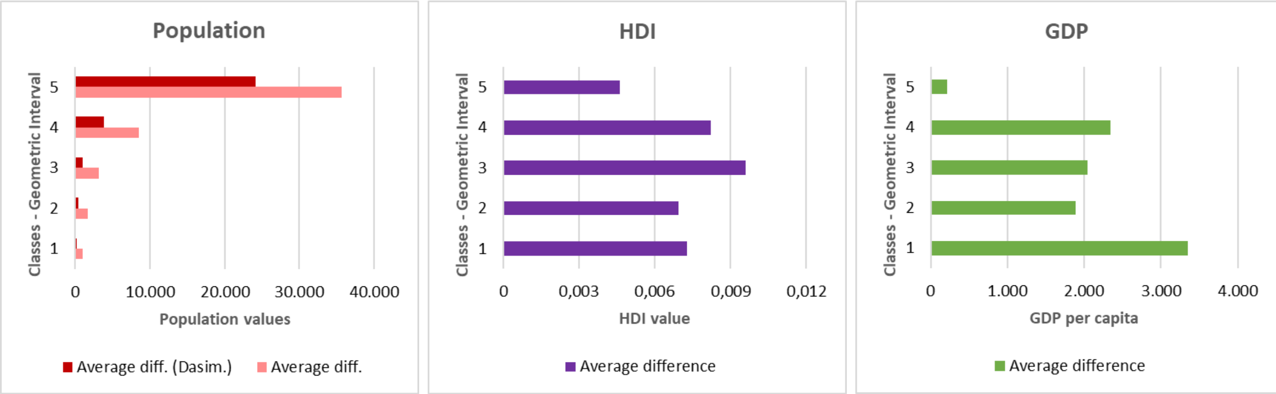


Elaboration: The authors (2022)

However, for the MHDI and GDP *per capita* data, this trend did not show the same characteristics as the population data. Therefore, it is hypothesized that this lack of relationship is because these socioeconomic data are an index and a monetary value, and not absolute count data, as is the population. It should still be taken into account the fact that the origin of the data is municipal limits and not census tracts. It was observed that the relationship between the amplitude of the methods and the size of the areas of the aggregated ottobasins was practically nil, with R-squared values close to zero.

As highlighted in the methodology, Table 1 presented the division into five classes of ottobasins that were classified by geometric intervals to search for patterns of differences between the determination methods proposed in the present work. From this classification, it was possible to make the dispersion analysis between the methods used in this study. Figure 9 represents the arithmetic mean of the amplitudes between the methods used in this study, according to the aggregated ottobasins size classes proposed by the geometric interval.

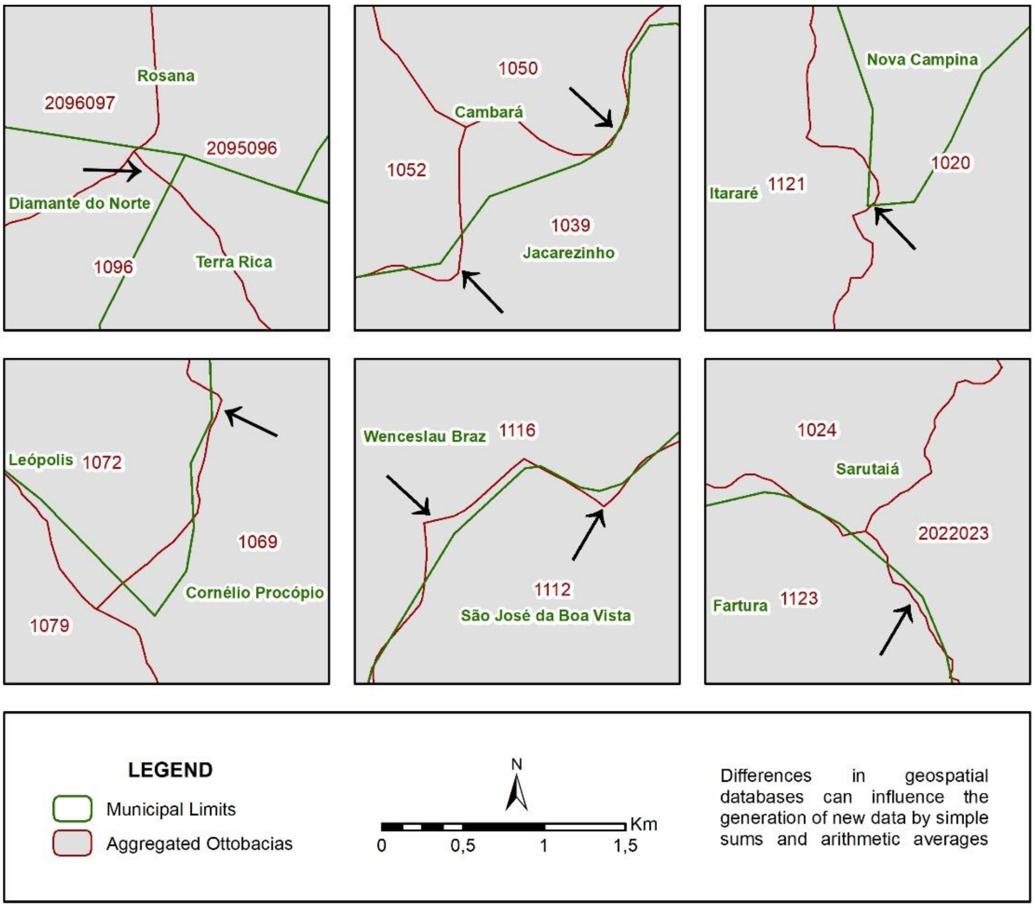
Figure 9 – Differences between methods by ottobasins (ottobacias) size



Elaboration: The authors (2022)

Corroborating what was presented in Figure 8, the population data tend to have larger amplitudes between the methods as the size of the aggregated ottobasins increases, both for the estimation by census tracts and by dasimetric mapping, that is, for very small ottobasins (class 1) the weighted sum method has values very close to the simple sum. However, for the MHDI and GDP *per capita* data, the very large ottobasins (class 5) were the ones that presented the lowest amplitudes between the results of the two methods. However, the ones with the greatest differences for MHDI were class 3 ottobasins, while for GDP were class 1. However, it is noteworthy that for MHDI and GDP, the lowest mean differences occurred in class 5. One hypothesis may be related to the size of the ottobasin itself because the larger the ottobasin, the greater the occurrence of different municipalities. If the MHDI and GDP are similar for the municipalities that make up a given ottobasin, the result will be a smaller amplitude in any size of ottobasin. Finally, Figure 10 illustrates six distinct situations that can influence simple sums and averages to the detriment of weighted ones.

Figure 10 – Differences between the database of municipal limits and aggregated ottobasins



Elaboration: The authors (2022)

It is noted that the differences between the municipal limits and the ottobasins cause small areas that stand out, and with this, they enter the intersection in the GIS environment. Therefore, they influence the determination of socioeconomic data and harm the quality of the final values, as they are influenced by areas that are practically not even part of the ottobasin in question. Thus, it is noteworthy that this factor is a limitation of the data, not of the study's method. Therefore, new stages of work can be suggested in which the researcher is committed to directing the research towards methods capable of reducing the error in data determinations in different spatial cutouts.

However, it should be noted that the data used in this research are available in political-administrative sections (municipalities) and information collection sections (census tracts). They were not necessarily designed to carry out politics and management, nor do they necessarily need to have this qualification. It is only necessary to highlight the limitations inherent to the data used.

4 FINAL CONSIDERATIONS

The determination of data of demographic or socioeconomic origin, whatever they may be, in units of analysis such as watersheds is always a challenge, especially when the database does not match its limits. Determining these data originating in census tracts or municipalities tends to cause some inconsistency, and the researcher must always highlight the limitations of the chosen method. The present work tried to demonstrate that, despite these difficulties, it is possible to make this determination of data; however, depending on the method, the results may vary, thus compromising the analysis and discussions about the reality of the watershed or ottobasin in question.

The viability of using GIS as an aid tool in generating demographic and socioeconomic data that can serve as subsidies for work involving the management of water resources was demonstrated. However, considering all census tracts or municipalities in the equation, the simple intersection can generate values far above reality.

The dasimetric mapping, carried out for the population data, proved to be limiting for this type of study as it disregards any area outside urbanized areas, given that there is a considerable number of residents in rural areas in the study area. In this way, it is emphasized that it is always necessary to consider the objective of the research before using a methodology, in the sense that the same data could generate satisfactory results if only one analysis were carried out with a scale compatible with the census tracts, demonstrating the areas of exclusion of sectors, depending on land use (urban and non-urbanized areas). It should also be noted that the data source can change the result, as in the case of the Landsat 30 m images used in this research. If another source (such as another satellite) had been used, there could have been a change in the quality of the final data presentation.

This study also identified issues that still require further investigation, such as the fact that the differences between the methods vary according to the size of the aggregated ottobasins, as in the case of population data that showed a trend that the larger the ottobasin, the greater the amplitude between the methods of analysis.

Finally, a discussion is presented in the sense that there is a gap in the generation of socioeconomic data in spatial units other than census tracts or municipalities, as in the case of watersheds. Indirect methods make all data of this nature generated in basins and only manage to make estimates of the reality of the data. It should be noted that despite these gaps, data collected by censuses are still the most accurate source available in the Brazilian reality and that the impairment or non-performance of a census can harm any studies of a social, geographical, and cartographic nature, among others, that take into account the characteristics of the population.

Acknowledgments

To National Water Agency (ANA) for the funding of the research project (TED - Decentralized Execution Term No. 11/2018/ANA), to the Graduate Program in Geodetic Sciences of Federal University of Paraná (UFPR), to CNPq for the research productivity grant – process 310312/2017-5, to Laboratory of

 DOI: http://dx.doi.org/10.14393/rbcv75n0a-66498

Geoprocessing and Environmental Studies (LAGEAMB) of the Department of Geography of UFPR, to Professor Cristovão Vicente Scapulatempo Fernandes of the Graduate Program in Water and Environmental Resources Engineering (Department of Hydraulics and Sanitation of UFPR) and finally, to the reviewers of the Brazilian Journal of Cartography for the suggestions and criticisms that added value to the manuscript.

Authors' contributions

A. N.: Conceptualization, Data Curation, Formal Analysis, Research, Methodology, Project Administration, Validation, Visualization, Writing - initial draft and Writing - revision and editing; L. S. D.: Conceptualization, Research, Methodology, Resources, Supervision, Validation, and Writing - revision and editing; EVP: Conceptualization, Acquisition of funding, Research, Resources, Supervision, Validation, and Writing - revision and editing.

Conflicts of Interest

The authors have no conflicts of interest to declare. The funders did not interfere in the development of the study, in the analysis or interpretation of the data, in the writing of the manuscript, or in the decision to publish the results. This article is part of the study of the postdoctoral project of the first author, Alexei Nowatzki, developed under Luciene Stamato Delazari and presented in 2021 at the Postgraduate Program in Geodetic Sciences at Federal University of Paraná (UFPR).

References

ABREU, M. V. S.; OLIVEIRA, J. C. DE; ANDRADE, V. D. A.; MEIRA, A. D. Proposta metodológica para o cálculo e análise espacial do IDH intraurbano de Viçosa – MG. **Revista Brasileira de Estudos de População**, v. 28, n. 1, p. 169-186, 2011. DOI: 10.1590/S0102-30982011000100009.

NATIONAL WATER AGENCY (ANA) **Watershed Committees**. Brasília, 2021. Available at: < https://www.ana.gov.br/aguas-no-brasil/sistema-de-gerenciamento-de-recursos-hidricos/comites-de-bacia-hidrografica> Accessed on: Apr 09, 2021.

ANDRIOTTI, J. L. S. **Fundamentos de estatística e geoestatística**. São Leopoldo: Ed. Unisinos, 165p. 2003.

BUSSAB, W.O.; MORETTIN P.A. **Estatística Básica**, Ed. Saraiva, São Paulo, 9 ed. 576p. 2017.

DUARTE, L. T.; SILVA, D.B.N.; BRITO, J.A.M. Análise de para dados do Censo Demográfico 2010: uma investigação de fatores associados a erros não amostrais do levantamento de dados. **Revista Brasileira de Estudos de População**, Rio de Janeiro, v.33, n.3, p.679-701, 2016. DOI: 10.20947/S0102-30982016c0011.

FRANÇA, V. O.; STRAUCH, J. C. M.; AJARA, C. Método dasimétrico inteligente: uma aplicação na mesorregião metropolitana de Belém. **Revista Brasileira de Cartografia**, Uberlândia, v. 66, n. 6, 2014. DOI: 10.14393/rbcv66n6-44721.

GOMES, J. V. P.; BARROS, R. S. A importância das Ottobacias para gestão de recursos hídricos. In: XV **Simpósio Brasileiro de Sensoriamento Remoto**, 2011, Curitiba - PR. XV SBSR, 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **General Census of Brazil in 2000** (for units of the federation and municipalities). Rio de Janeiro, 2000.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Brazilian Census of 2010**. Rio de Janeiro: IBGE, 2012.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Statistical Grid**. Rio de Janeiro: IBGE, 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **IBGE Portal**, Brasília, 2020. Available at: < https://www.ibge.gov.br/>. Accessed on: Aug 26, 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **What is GDP**. Brasília, 2020b.

Available at: < https://www.ibge.gov.br/explica/pib.php/> Accessed on: Aug 27, 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Mesh of Census Sectors**. Brasília, 2021. Available at: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/26565-malhas-de-setores-censitarios-divisoes-intramunicipais.html?=&t=saiba-mais-edicao).> Accessed on: Oct 11, 2021

LABORATORY OF GEOPROCESSING AND ENVIRONMENTAL STUDIES (LAGEAMB). Geospatial Database and Project Document Management. Org..: PAULA, E. V.; PAZ, O. L. S.; SILVA, V. C.; NOWATZKI, A.; WROBLEWSKI, C. A.; PILATTI, D. M.; RIBEIRO, E.; SILVA, E. N.; SENGER, G.; COUTO, J. P. B.; CORDEIRO, M. F. O.; PAI, M. O. D.; SANTOS, R. V.; VIKOU, S. V. P. 54p. **Technical Note**. The Federal University of Paraná Curitiba. 2020..

MACEDO, D. R.; MAGALHÃES JR. A. P. Evaluation of urban stream restoration project through water quality analysis and survey of the neighborhood residents. In: **International Conference Sustenaible Techniques and Strategies in Urban Water Management**, n. 7, p. 1-9, Lyon, France: Graie, 2010.

MENNIS, J.; HULTGREN, T. Dasimetric mapping for disaggregating coarse resolution population data. **Proceedings of the 22nd Annual International Cartographic Conference**, p. 9-16. 2005.

MOON, Z. K; FARMER, F. L. Population density surface: a new approach to an old problem. **Society & Natural Resources**, v. 14, n. 1, p. 39-51, 2001. DOI: 10.1080/08941920120013.

MOORE, D. S. **A estatística básica e sua prática**. Rio de Janeiro: LTC, 7th ed., 2017.

MORAIS JUNIOR, H. DE S.; SILVA, S. L. B. DA. Dasimetric mapping and interpolation of census data and orbital imaging for analysis of the dilation of the urban site of the city of Belém do Pará. **Geonorte Magazine**, v. 10, n. 36, p. 95-109, 2019. DOI: https://doi.org/10.21170/geonorte.2019.V.10.N.36A95.109

NEVES, O. M. S.; STRAUCH, J. C. M.; AJARA, C. Dasymetric methods applied to Jacarepaguá watershed. **Boletim de Ciências Geodésicas,** Curitiba, v. 23, n. 4, p. 606-622, 2017. DOI: 10.1590/S1982-21702017000400040.

OLIVEIRA, L. A. P.; SIMÕES, C.C.S. The IBGE and population surveys **Revista Brasileira de Estudos de População**, São Paulo, v. 22, n. 2, p. 291-302, 2005. DOI: 10.1590/S0102-30982005000200007.

OLIVEIRA, F. F.; MAGALHÃES JR., A. P. Disponibilidade de dados socioeconômicos para a gestão de bacias hidrográficas no Brasil. Em: **XX Simpósio Brasileiro de Recursos Hídricos**, v. 1, p. 102-113, 2013.

PAZ, O. L. S.; DAL PAI, M. O.; PAULA, E. V. Methodological proposal for elaborating a geospatial database as a subsidy for environmental studies: application in conservation units on the northern coast of Paraná. **Revista Brasileira de Geografia Física**, Recife, v. 13, n. 2, p. 613-629, 2020. DOI: 10.26848/rbgf.v13.2.p613-629.

PEDRO, A. A.; QUEIROZ FILHO, A. P. de. Evaluation of the statistical grid in subnormal clusters: A case study of the São Mateus - SP subprefecture. **Revista Brasileira de Cartografia**, [S. l.], v. 69, n. 2, 2017. DOI: 10.14393/rbcv69n2-44018.

INTEGRATED WATER RESOURCES PLAN (PIRH). **Unidade de Gestão de Recursos Hídricos Paranapanema** (PIRH Paranapanema), Brasília, 335p, 2016. Available at: < https://www2.paranapanema.org/plano-de-bacia/> Accessed on May 09. 2021.

UNITED NATIONS DEVELOPMENT PROGRAM (UNDP). **What is human development**. Brasília, 2022. Available at: <https://www.br.undp.org/content/brazil/pt/home/idh0.html> Accessed on: Oct 11, 2022.

PFAFSTETTER, O. **Classificação de Bacias Hidrográficas** – Metodologia de Codificação. Rio de Janeiro, RJ: DNOS, 1989.

PROCHNOW, M. C. R. Planning of watersheds – Environment Week, In **Anais Livroceres**. Campinas, Brazil. Campinas: Fund. Cargill; p. 31-39, 1988.

SANTOS, A. M.; HOMES, D. C. S. C.; RAMOS, H. F. Demographic density: a comparative study of two methodologies from orbital and suborbital images in Aparecida de Goiânia/Goiás. **Ateliê Geográfico**, [S. l.], v. 12, n. 1, p. 175–200, 2018. DOI: 10.5216/ag.v12i1.45968.

SILVA, J. G. Secrets of Statistics for Geography. **Cadernos Geográficos**. Florianópolis, no. 35, 128 p. 2016.

SILVA, A. P; MORATO, R. G.; KAWAKUBO, F. S. Mapping the spatial distribution of the population using the dasimetric method: case example in southeastern Brazil. **Revista Brasileira de Geografia Física**, Recife, v. 6, p. 37-48, 2013. DOI: 10.26848/rbgf.v06.1.p037-048.

SILVEIRA, R. M. P. **Thematic Cartography**. 1. ed. Curitiba: Intersaberes, 2019.

SLOCUM, T. A.; MCMASTER, R. B.; KESSLER, F. C.; HOWARD, H. H. **Thematic Cartography and Geovisualization**. (3rd ed.). EUA: Prentice Hall Series in Geographic Information Science, 2009.

TOZZI, B. K. M.; KRÜGER, C. M.; GALLEGO, C. E. C.; RAMINA, R. H.; TOZZI, R. F.; TASCHELMAYER, C. Proposed methodology for determining surface water availability in the planning of water resources, using spatial interpolation and construction of flow surfaces. In: XX **Simpósio Brasileiro De Recursos Hídricos**, Bento Gonçalves – RS, 2013.

UMBELINO, G. J. M.; BARBIERI, A. F. Methodology for the compatibility of census tracts and urban perimeters between the 1991, 2000, and 2010 censuses. In: XVI National Meeting of Population Studies, 2008, Caxambu - MG. **Anais do XVI Encontro Nacional de Estudos Populacionais**, 2008.

WEBER, E. M. **DasyTools: A user-friendly dasymetric mapping toolkit with automated accuracy assessments**. 2010, 105 f. Thesis (Ph.D. in Geography), University of Kansas, Lawrance, 2010.

Biography of the main author

Alexei Nowatzki was born in São Paulo (SP) in 1986. Geographer graduated from UFPR (Federal University of Paraná). Ph.D. and Master's by the Postgraduate Program in Geography (UFPR) with the research line "Landscape and Environmental Analysis", with research development in digital soil mappings. Post-doctorate in Geodetic Sciences (UFPR) by the research line "Cartography and GIS" and with the partnership of ANA (National Water Agency), in the theme of water resources and territory management. He has experience in Pedology, Geomorphology, Cartography, Geoprocessing, and Geography Education. He is a substitute professor at IFPR (Federal Institute of Paraná) at the Curitiba Campus.