The use of principal component analysis for the construction of the Water Poverty Index

Uso de análise de componentes principais na construção do Índice de Pobreza Hídrica (WPI)

Larynne Dantas de Senna[1], Adelena Gonçalves Maia[1] (ID) **and Joana Darc Freire de Medeiros[1]** (ID)

[1]Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil

E-mails: larynnedantas@yahoo.com.br (LDS), adelenam@gmail.com (AGM), joanadarc.medeiros@gmail.com (JDFM)

ABSTRACT

In relation to water resources, indexes can be created to express the multiple dimensions involved with it to aid the planning and management of basins. In this regard, the Water Poverty Index is globally used, but one of its criticisms includes the subjectivity associated with how the sub-indexes are weighted. Therefore, in this study, we applied principal component analysis (PCA) to determine the sub-indexes' weight: resource, access, capacity, use, and environment of the Seridó river basin. This new index with PCA presents an average range with broader values compared to methodologies without, allowing clear identification of the disparities among the cities and the possibility to better prioritize investments concerning water poverty reduction. Our results show that this approach makes it possible to qualitatively identify geographical locations that have greater water poverty compared to others. Additionally, with this approach, it can be determined whether water poverty is caused due to natural characteristics or deficits in water infrastructure investment, providing insight into social fragilities as well. Overall, the presented hierarchical tool in this study has a high value to improve the planning of water resource uses.

Keywords: Water scarcity; Multivariate analysis; Semiarid region.

RESUMO

Em relação aos recursos hídricos, índices vem sendo criados para expressar as múltiplas dimensões envolvidas no planejamento e gestão das bacias hidrográficas. Nesse sentido, o Índice de Pobreza Hídrica tem sido globalmente usado. Uma de suas críticas inclui a subjetividade associada a como os sub-índices são ponderados. Neste estudo, aplicamos a análise de componentes principais (PCA) para determinar o peso dos sub-índices: recursos, acesso, capacidade, uso e ambiente da bacia do rio Seridó. O novo índice obtido com o uso de PCA apresenta uma faixa média com valores mais amplos em comparação com as metodologias sem o uso de PCA, permitindo uma identificação clara das disparidades entre as cidades e a possibilidade de priorizar melhor os investimentos relativos à redução da pobreza hídrica. Nossos resultados mostram que essa abordagem possibilita identificar qualitativamente as localizações geográficas que têm maior pobreza hídrica em comparação com outras. Além disso, com essa abordagem, pode-se determinar se a pobreza hídrica é causada devido a características naturais ou déficits de investimento em infraestrutura hídrica, fornecendo também insights sobre as fragilidades sociais. No geral, a ferramenta hierárquica apresentada neste estudo tem um alto valor para melhorar o planejamento dos usos dos recursos hídricos.

Keywords: Escassez hídrica; Análise multivariada; Região semiárida.

INTRODUCTION

Indexes can be created with the intention to express multiple dimensions (e.g., socioeconomic, physical, environmental, and institutional) in a simple way for interpretation. The index approach can also be used in relation to water resources (MLOTE; SULLIVAN; MEIGH, 2002). In general, indexes can serve as an instrument to make large quantities of data more manageable by reducing the data's raw size while retaining the essential information (OTT, 1978). Therefore, for water resources, indexes can contribute to the formulation of policies for management of a river basin.

The Water Poverty Index (WPI) is an index that can capture the complete range of issues related to water resource availability and its relationship to human and ecological needs (LAWRENCE; MEIGH; SULLIVAN, 2002; MLOTE; SULLIVAN; MEIGH, 2002). The WPI aims to correlate water with poverty through various indicators, such as water for sanitation, hygiene, and health, as well as the generation of jobs, equality among social classes, and the rights of lower social classes' access to water (RIJSBERMAN, 2003). In multiple studies, the WPI has been used to evaluate water scarcity, in which most of them used the index to identify a specific set of indicators for different locations (GINE; FOGUET, 2009; LAWRENCE; MEIGH; SULLIVAN, 2002; LISA, 2014; MANANDHAR; PANDEY; KAZAMA, 2012; ZHANG et al., 2015). Thanks to these studies, it has led to the awareness that questions regarding water scarcity and indicators to represent them are location specific. As such, the indicators must be carefully chosen (MANANDHAR; PANDEY; KAZAMA, 2012).

Since the development of the WPI, various criticisms had emerged as well (FEITELSON; CHENOWETH, 2002; JIMÉNEZ; MOLINERO; PÉREZ-FOGUET, 2009; GINE; FOGUET, 2009; CHO; OGWANG; OPIO, 2010). One of these criticisms describes the redundancy in the WPI between variables and the decision of weight attribution to the sub-indexes. Previously, Martinez-Alier, Munda, and O'Neill (1998) reported that the process of weight attribution to the sub-indexes is sometimes arbitrary and that there is no justification of the weight distribution attributed to a specific indicator rationally. Magalhães Júnior (2007) reinforced this criticism, noting that the weight attribution can occur based on different criteria but that at times, it is not possible to do it without incorporating some level of discretion or subjectivity.

The use of principal component analysis (PCA) of the sub-indexes can solve the difficulty regarding the random choice of weights used in the WPI (CHO; OGWANG; OPIO, 2010). For this reason, we evaluated the use of the PCA for assigning weights to sub-indexes that comprise the WPI of the Seridó river basin, located in the semiarid region of the north-east of Brazil.

MATERIAL AND METHODS

Study area

In Brazil, one of the regions experiencing considerable water scarcity is the semiarid region, which covers more than 57% of the north-eastern territorial area. In this region, natural factors—such as long periods without precipitation, an elevated evaporation rate, and low storage capacity of water in the subsoil, which is associated with economic activities (e.g., agriculture and livestock, which has water as a basic input)—result in low water availability. In this region, there is the Seridó river basin (a sub-basin of the Piranhas-Açu river basin), which has a total area of 10,092 km²—consisting of 6,645 km² in the Rio Grande do Norte state and 3,447 km² in the Paraíba state. Furthermore, the Seridó river basin contains

the municipal offices of 28 cities (18 in the RN and 10 in PB state) (ABRANTES, 2011).

In this study, the cities within the Seridó river basin, with the exception of eight cities due to inconsistency or deficiency in the acquisition of the selected variable for the index formation, were included (Figure 1).

The Seridó river basin has an annual average temperature between 26 and 28°C and a relative humidity of 64%. Furthermore, the majority of the Seridó basin's climate is predominantly classified as a BSw 'h' type (based on the Köppen climate classification), which describes hot and semiarid weather. The basin is also entirely located within the Caatinga biome and, in geological terms, the basin presents is a predominance of precambrian crystalline rocks (ABRANTES, 2011).

Water Poverty Index

The construction of the WPI can be realized with various methods, such as the composite index approach, the Gap method, the matrix approach, and the time analysis approach (SULLIVAN, 2002). In this study, the composite index approach was used due to its main advantage regarding its ability to address the multidimensional nature of water poverty. Using this approach, the WPI was constructed based on five sub-indexes: resource, access, capacity, use, and environment. Each of them is composed using a linear combination of their representative variables (e.g., the population with access to potable water and water use per capita for agriculture among others). The sub-indexes can be viewed as separate indicators in dimensions that, when aggregated, obtain the index (SICHE et al., 2007).

The resource sub-index reflects the physical availability of superficial and underground water, while the access sub-index reflects the population's access to water. Then, the capacity sub-index reflects the ability of citizens to obtain or manage water (or both), while the use sub-index reflects how the water is used (e.g., for domestics, agricultural, and non-agricultural). Finally, the environment sub-index indicates ecologic integrity, which can reveal whether there is a capacity to handle water stress and to ensure the sustainable use of the water resources.

So, the WPI can be calculated according to Equation 1 (LAWRENCE; MEIGH; SULLIVAN, 2002):

$$WPI = w_r R + w_a A + w_c C + w_u U + w_e E \qquad (1)$$

In Equation 1, w_r, w_a, w_c, w_u, and w_e are the applied weights for each sub-index, R is the resource sub-index value, A is the access sub-index value, C is the capacity sub-index value, U is the use sub-index value, and E is the environment sub-index value. The weights vary between 0 and 1 and are used to incorporate the distinct importance of the sub-indexes, which are used to highlight the main problems that need to be addressed by policy goals (MLOTE; SULLIVAN; MEIGH, 2002).

The WPI values range from 0 to 100. A higher WPI value reflects a lower degree of water poverty. Then, with the use of the WPI value, it is possible to determine the water poverty relative position of the cities.
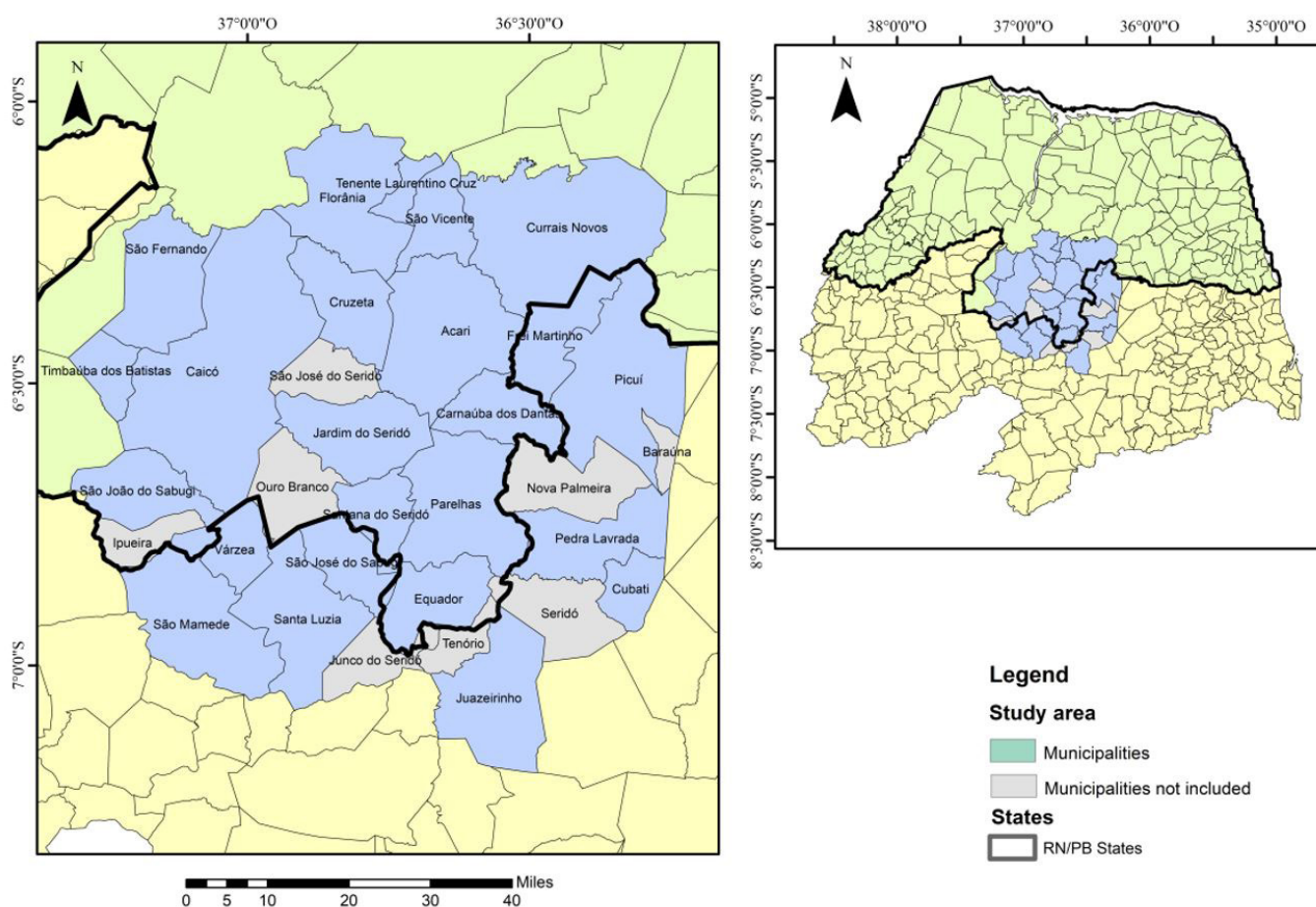
Figure 1. Study area.

Definition of the WPI sub-index variables

The variables utilized in this study were selected based on the following considerations: They had to be indicators already internationally consolidated and widely used, their data had to be available for analysis, and they could be applied to the semiarid region's characteristics (Table 1). All variables were standardized using the minimum-maximum method, see Equation 2:

$$Var_{standardized} = \frac{var_i - var_{min}}{var_{max} - var_{min}} \qquad (2)$$

In Equation 2, $Var_{standardized}$ is the original standardized variable of city i, in which var_{min} is the lowest value of the variable among the cities and var_{max} the highest one. The determination of each sub-index value was made with the weighted average of the standardized variable's value.

For the resource sub-index, two variables were used: the annual average precipitation and the regulated flow per capita (with a 90% guarantee). Considering that the municipalities in the Seridó river basin are powered by reservoirs, precipitation occurring on the dam that supply one city was attributed to this respective city. In a situation where the municipality water supply comes from more than one reservoir, the average of these variables were used.

The access sub-index accounts for situations where the water poverty is not associated with an insufficient water availability, but with an inadequate infrastructure to make this water resource available

Table 1. Resource, access, capacity, use, and environment sub-index variables.

Sub-indexes	Variables
Resource	Annual average precipitation (mm.year[-1])[1, 2]
	Regulated flow per capita (hm[3].s[-1].inhab[-1])[3]
Access	Population access to treated water (%)[4]
Capacity	IDH-M[4]
	Literacy rate (%)[4]
	Economically Active Population (%)[4]
Use	Domestic use of water per capita (m[3].day[-1].inhab[-1])[5]
	Water use for irrigation (m[3].year[-1])[3]
	Water use for livestock (m[3].year[-1])[3]
Environment	Area with natural vegetation (%)[6]
	Total Phosphorus Total (mg.L[-1])[2,7]

Sources:[1]Secretariat of Water Resources and Environment from Rio Grande do Norte State (SEMARH, 2014). [2]Executive Agency of Water Management for Paraíba State (AESA, 2014). [3]National Water Agency (ANA, 2014). [4]Demographic Census, 2010, from the Instituto Brasileiro de Geografia e Estatística (IBGE, 2015). [5]National Research of Basic Sanitation (IBGE, 2008). [6]Agricultural Census (IBGE, 2006). [7]Water Management Institute of RN (IGARN, 2015).

to the population. Therefore, in the present study, the percentage of the population with access to potable water was included in the index. In terms of the capacity sub-index, socioeconomic indicators already internationally consolidated and widespread

were used, including Municipal Human Development Indicators (IDH-M), the literacy rate, and the economically active population. Together, these reflect the economic and social development of the region. Then, for use sub-index we used the water consumption for public supply water, irrigation and use of livestock.

Finally, for the environment sub-index, we used the percentage of area with natural vegetation, representing how healthy the environment is, and the total phosphorus percentage present in water reservoirs. This percentage is one of the parameters indicating the water's eutrophication, which is an inversely proportional variable to the WPI index (i.e., higher values are associated with worse hydric situations). For this reason, it was necessary to use the minimum-maximum method for standardization in the reverse direction, see Equation 3:

$$\text{Var}_{\text{standardized}} = \frac{\text{var}_{\text{max}} - \text{var}_{\text{i}}}{\text{var}_{\text{max}} - \text{var}_{\text{min}}} \tag{3}$$

Finally, the sub-index values were generated through an arithmetic average of the standardized variables.

For a better interpretation of the WPI and its sub-indexes, the classification proposed by El-Gafy (2018) was used, where 0 to 20 = very poor, >20 to 40 = poor, >40 to 60 = good, >60 to 80 = very good, and >80 to 100 = excellent.

Principal component analysis

For the weight definition of each sub-index (Equation 1), a PCA was used. In general, a PCA is used to transform a large set of correlated variables into a smaller set of uncorrelated variables, termed principal components, that account for most of the variation in the original set of variables (DUNTEMAN; 1989; MORRISON, 1967). So, a PCA transforms the original variables into a new set of variables that are (1) linear combinations of the original ones, (2) uncorrelated with each other, and (3) ordered according to the amount of variation in the original variables, which can be accounted for by the new variables (EVERITT; HOTHORN, 2011). In mathematical terms, a PCA involves the following steps:

(1) standardization of variables X1, X2, etc. for the mean zero and unit variance

(2) calculation of the correlation matrix R

(3) determination of the eigenvalues λ1, λ2, ..., and λp and the corresponding eigenvectors a1, a2, ..., and ap through the solution of Equation 4, where "I" is the identity matrix:

$$\left| R - I\lambda \right| = 0 \tag{4}$$

(4) elimination of components that have little contribution to the variance of the original data set

(5) application of matrices of eigenvectors as the factors in a linear combination of standardized variables for the composition of the principal components (NOORI et al., 2010).

The first generated principal component explains the higher proportion of the total variance from the original database, while the second captures the higher proportion of the total variance not represented by the first, etc. For a database of k variables, for example, the maximum number of extracted components would be k, regardless of whether there is a high correlation among its variables, in which a much smaller number of components would be enough to represent the highest portion of the total variance from the original variables (CHO; OGWANG; OPIO, 2010). In our work, only the main components that obtained an eigenvalue greater than 0.7 were used, similar to the criteria used by Cho, Ogwang, and Opio (2010), and Jemmali and Matoussi (2013).

Each principal component is associated with an eigenvector that provides weight to the sub-indexes. However, because there is more than one component, more than one weight may be available for each sub-index as well. For this reason, an aggregation method of the principal components (PÉREZ-FOGUET; GINÉ GARRIGA, 2011) was used to calculate the weight of each sub-index, see w_i, Equation 5. In this approach, the higher the variance proportion is expressed by the component of a determined self-vector, the higher the weight will be to compose a final weighting:

$$w_i = \Sigma_{k=1}^{n} \left(a_{k,i} \frac{\sqrt{\lambda_k}}{\Sigma_{j=1}^{n} \sqrt{\lambda_j}} \right) \tag{5}$$

In Equation 5, w_i is the final weight used for sub-index i, k is the number of principal components, $a_{k,i}$ are the self-vectors that vary from 1 to k (the principal component numbers) and from 1 to i (the sub-indexes), λ_k are the self-values of the principal components k, and $\Sigma_{j=1}^{n} \sqrt{\lambda_j}$ is the sum of the j-adopted self-values after the selection criteria application.

In the present study, statistical analysis was performed with R software (R CORE TEAM, 2012).

RESULTS AND DISCUSSION

Determination of the WPI sub-indexes

The results of the sub-indexes—including resource, access, capacity, use, and environment—are presented in Figures 2 (a) to 6 (b). The results are divided into two states, Rio Grande do Norte and Paraíba, because many of the proposed investments are supplied by those respective state governments.

The resource sub-index presents low values in most of the cities, indicating that the water availability in the Seridó river basin does not adequately meet the demand for water (Figure 2). Interestingly, even cities containing a large reservoir in their territories, such as Parelhas and Caicó, showed low resource sub-index values. This finding can be associated with its population density as the regulated water flow was divided by the population to compose the regulated flow per capita variable. The cities that presented the best values for the resource sub-index (i.e., São Mamede, Santa Luzia, and Tenente Laurentino Cruz) obtain their water supply from other basins and were classified as "very good" in the resource sub-index.

It is important to emphasize that the resource sub-index was capable of accurately portraying the region's water availability
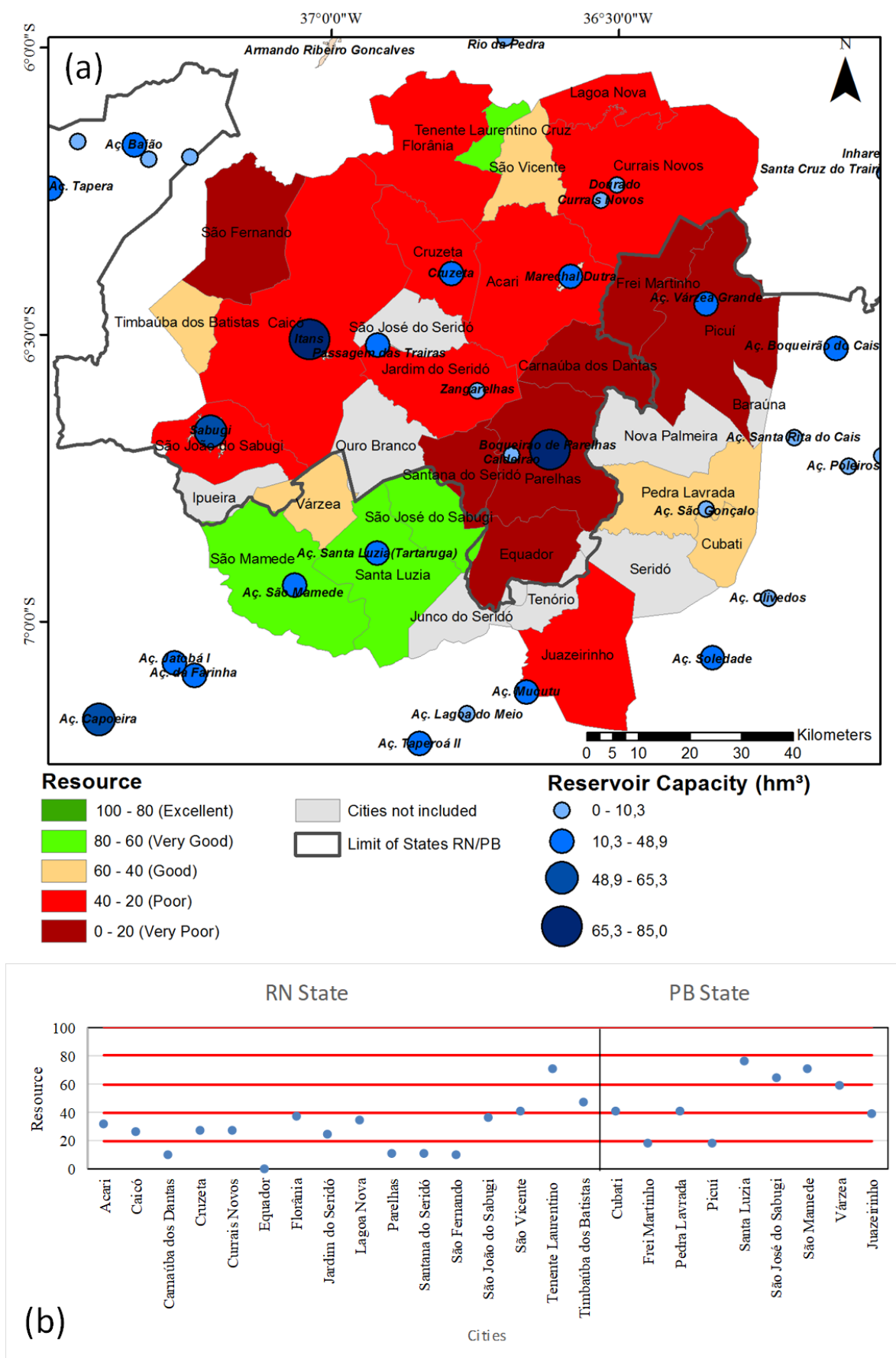
Figure 2. Resource sub-index spatial distribution (a) and its classification for the different locations (b).

situation. The Seridó river and the whole semiarid region of north-eastern Brazil had experienced a long period of below-average precipitation since 2012 resulting in very low reservoir water levels. The cities with the worst resource sub-index values had their water supply systems collapsed, indicating the accuracy of the sub-index. For instance, in 2015, the cities of Carnaúba dos Dantas, Acari, Currais Novos, Santana do Seridó, Caicó, Cruzeta, Equador, and Jardim do Seridó experienced at least one month with a collapsed water distribution system and were classified (Figure 2b) as very poor or as poor in the resource sub-index, indicating the lack of available water again to support the region's water demand.

The seven municipalities listed in the very poor class should have the highest priority regarding the allocation of monetary investments to increase water availability. New water sources need to be investigated, and investments should make water depositions in underground reservoirs available (depending on each location characteristic) or make it feasible to transport water from other neighboring basins.

Then, the access sub-index (Figure 3), which constructed only by one variable based on standardized data. Here, it is important to mention that cities with a lower access sub-index do not necessarily have a low population percentage that has access to potable water, but it is the city with the worst result. In the Seridó basin, Pedra Lavrada was the only city classified as very poor for the access sub-index, where only about 44% of its population has access to treated water. For this class (very poor), investments should be made to increase the percentage of the population with access to water, and in this case, Pedra Lavrada should have the highest priority. Additionally, the other evaluated cities should receive investments as well, depending on their classification score of the access sub-index presented in Figure 3b, in order to increase their water supply system. In addition, the access sub-index indicated that cities in the state of Rio Grande do Norte have better average values than cities in Paraíba. This occurred probably due to a more robust sanitation policy in the state of Rio Grande do Norte.

The capacity sub-index (Figure 4) is not associated with a region's socioeconomic development; therefore, the high values observed for cities such as Caicó and Currais Novos (excellent class) were expected as they are major cities in which most of the region's economic activities are concentrated. In contrast, the two cities classified as very poor (Lagoa Nova and Cubati), should receive primary attention to develop their areas in health, education, and economy. After these two cities, attention should be given to the five cities that are classified as poor.

For the use sub-index (Figure 5), only one city was classified as very good and two as good. These three cities have a higher level of irrigation usage, which has aided their higher values as irrigation use tends to have a significant impact on the sub-index composition. The cities classified as very good and good should still have the efficiency of their water usage analyzed. Once the water resources is being used efficiently, it can bring economic development to the region. However, if the water usage is determined inefficient, then policies encouraging wise water usage should be implemented in these cities, particularly for irrigation uses.

Complementary studies should be conducted on cities classified as poor and very poor to verify the possibility to increase

the use of water. As an effect, this would improve the region's development. However, this increase is limited by the amount of water available in the region. Another recommended action is to invest in more efficient irrigation systems or develop economic activities that require less water.

For the environment sub-index (Figure 6), cities in Paraíba demonstrated the best results. Here, developed cities such as Caicó and Currais Novos registered low values on the environment sub-index. Generally, when a population develops, more natural resources are used, and there is greater access to them. Consequently, there is degradation of the environment that surrounds the population. The lack of sanitary sewage system management is the main reason for high percentages of phosphorus present in reservoirs, causing eutrophication in most of these systems in the region. However, even the cities classified in the environment sub-index as good, very good, and excellent can have environmental issues. For this reason, the studied cities should implement public policies to increase the areas with natural vegetation, reduce deforestation, and improve the sanitary sewer system.

Principal component analysis

When applying the PCA to the weight attribution of the sub-indexes, the first principal component generated is CP1. This component captured the largest information and portrayed the phenomenon most optimally by presenting the largest variability of the original data, capturing 47.81% of the total variance. The second principal component (CP2) captured 23.34%, and the third (CP3) 13.89%. Together, these three components were capable of explaining 85.04% of the data variability. Then, the remaining two, CP4 and CP5, captured 10.12% and 4.84%, respectively.

Based on the criteria of Jolliffe (1973), eigenvalues with a value higher than 0.7 were selected for the principal components. The sequence of obtained values was $\Lambda_1 = 2.4$, $\Lambda_2 = 1.2$, $\Lambda_3 = 0.7$, $\Lambda_4 = 0.5$, and e $\Lambda_5 = 0.2$. Therefore, the first three principal components (CP1, CP2, and CP3) were selected to compose the sub-indexes' weight (Table 2). With this approach, using the aggregation method—suggested by Pérez-Foguet and Giné Garriga (2011) (Equation 5)—it is possible to find values of the final weighting for each sub-index. By rescheduling the weights (so the sum results in 1), the WPI1 index has the following composition (Equation 6):

$$WPI1 = 0.074R + 0.185A + 0.333C + 0.389U + 0.019E \qquad (6)$$

In Equation 6, R is the resource sub-index value, A is the access sub-index value, C is the capacity sub-index value, U is the use sub-index value, and E is the environment sub-index value.

Our findings show that the WPI1 is especially influenced by the use, capacity, and access sub-indexes, which represented more

Table 2. Weightings of CP1, CP2, and CP3.

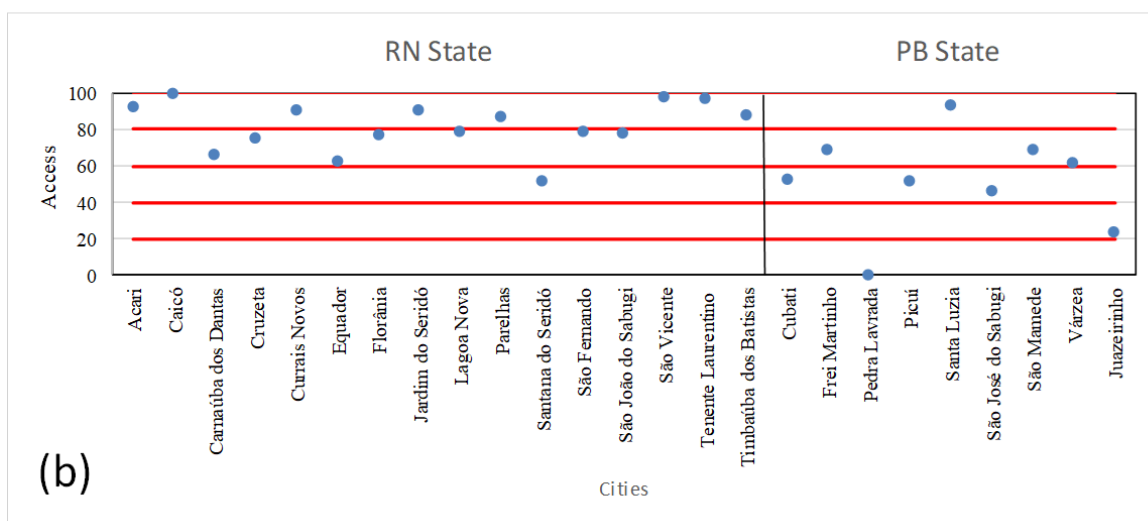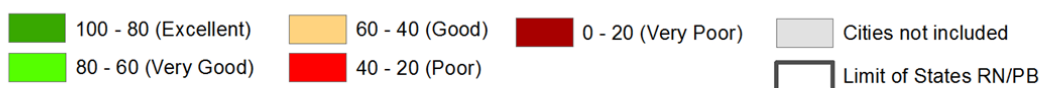| Sub-indexes	CP1	CP2	CP3
Access	0.505	0.313	−0.509
Capacity	0.486	0.127	0.448
Use	0.462	0.323	0.463
Environment	−0.491	0.374	0.492
Resource	−0.232	0.801	−0.289
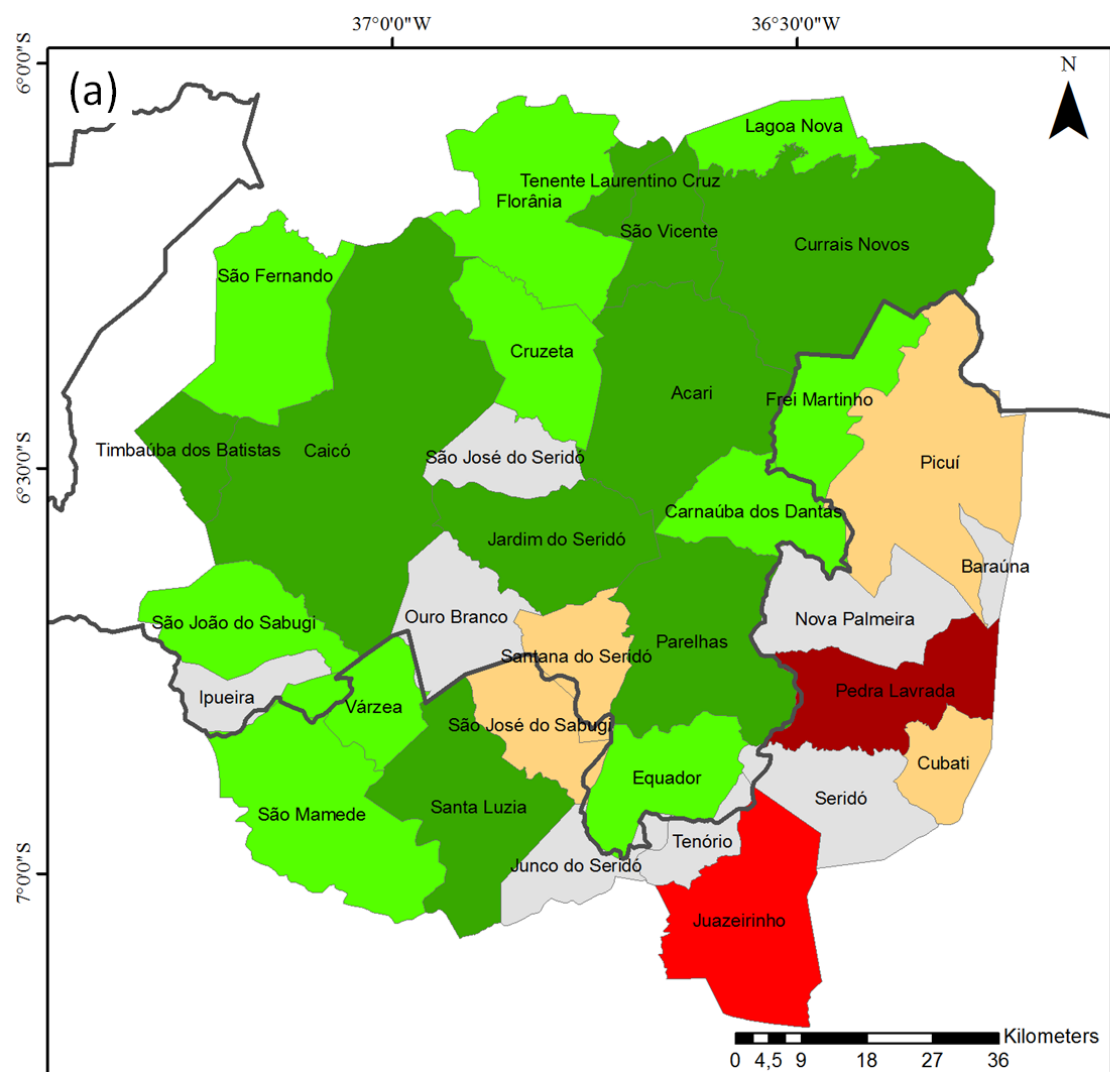
Figure 3. Access sub-index spatial distribution (a) and its classification for the different locations (b).
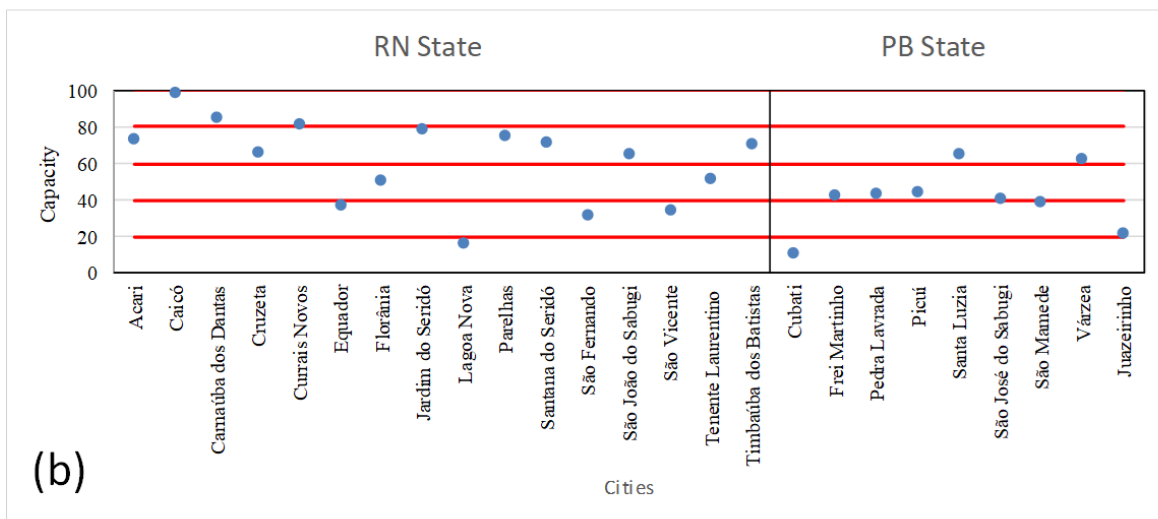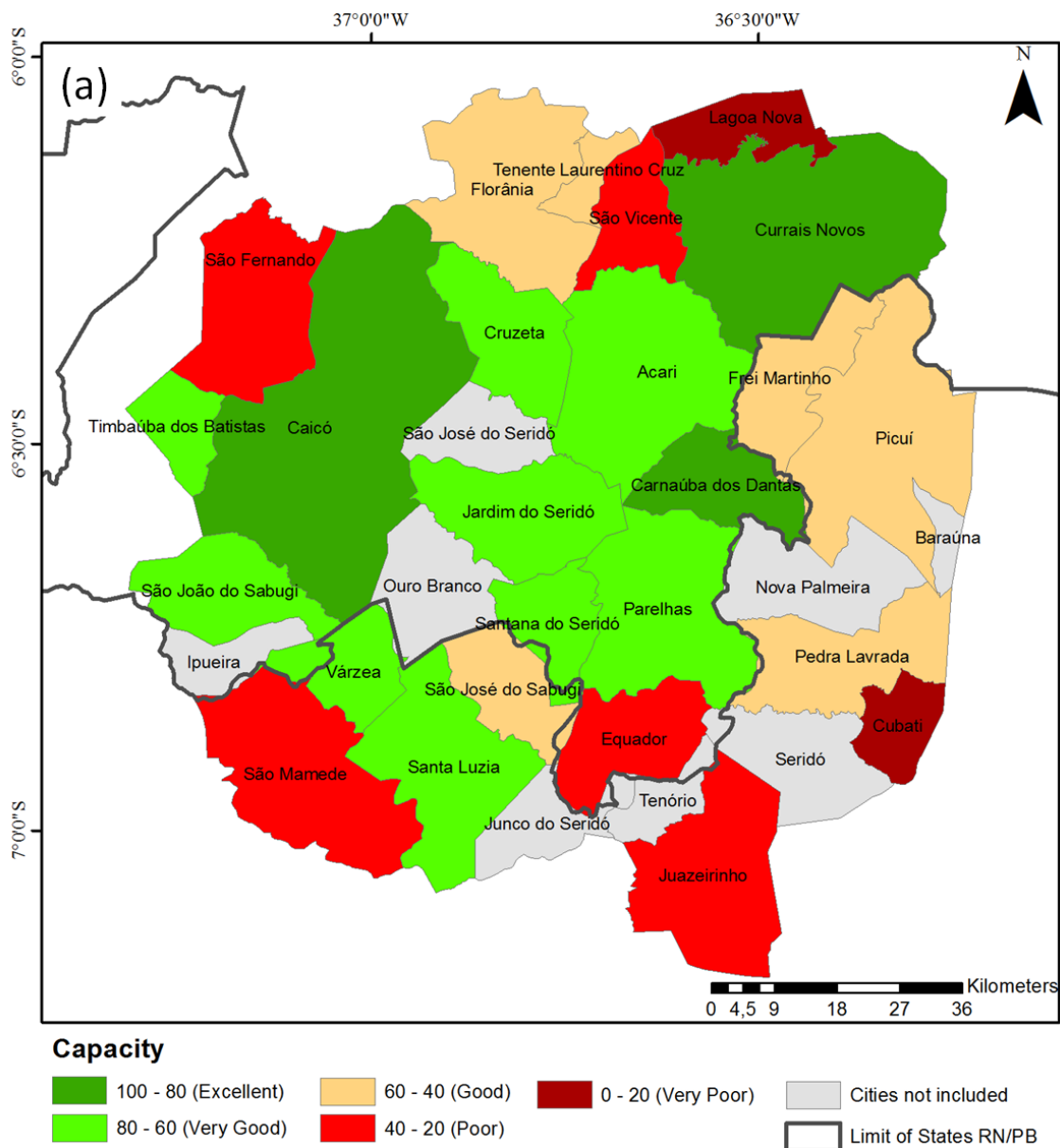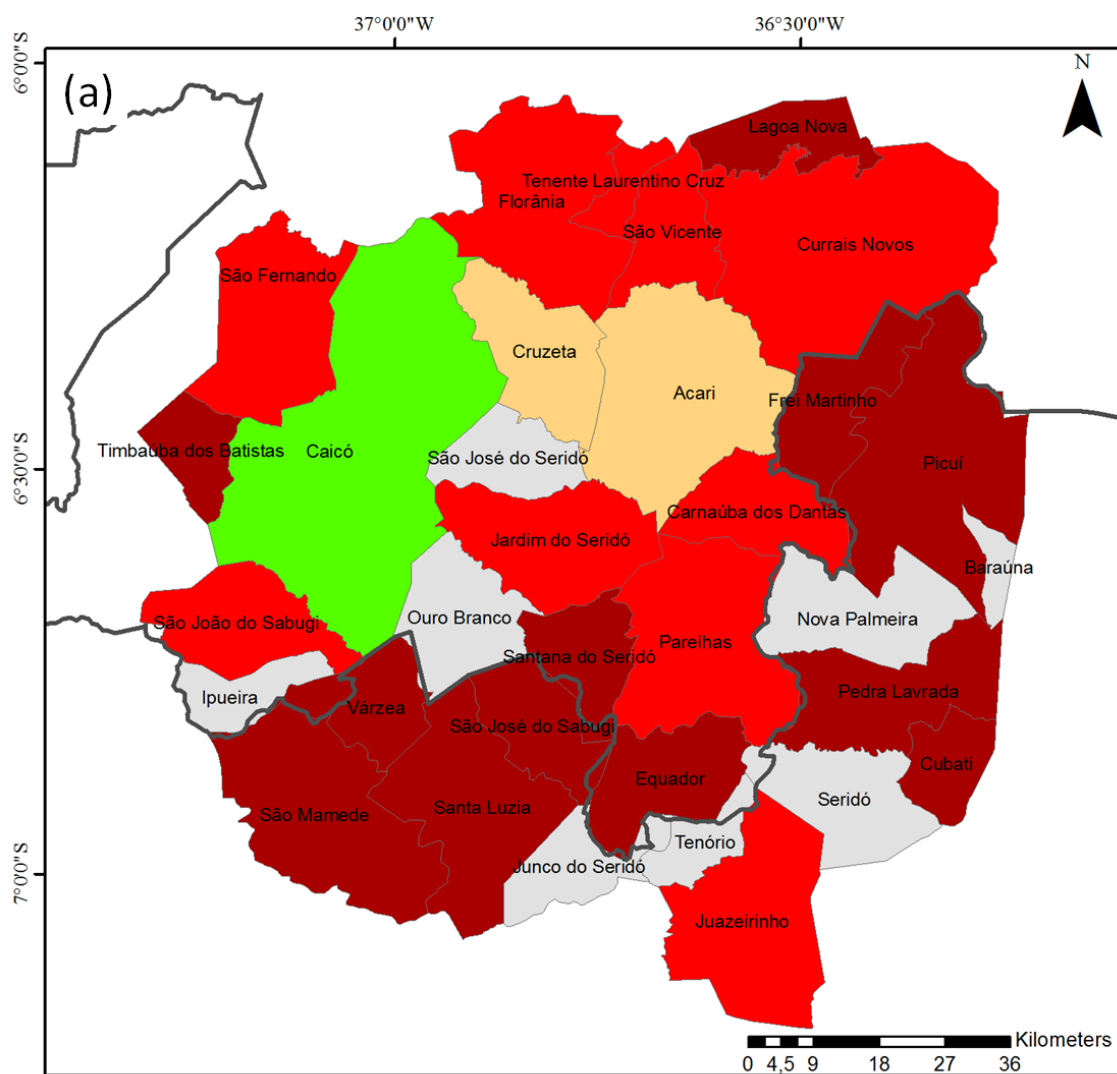
Figure 4. Capacity sub-index spatial distribution (a) and its classification for the different locations (b).
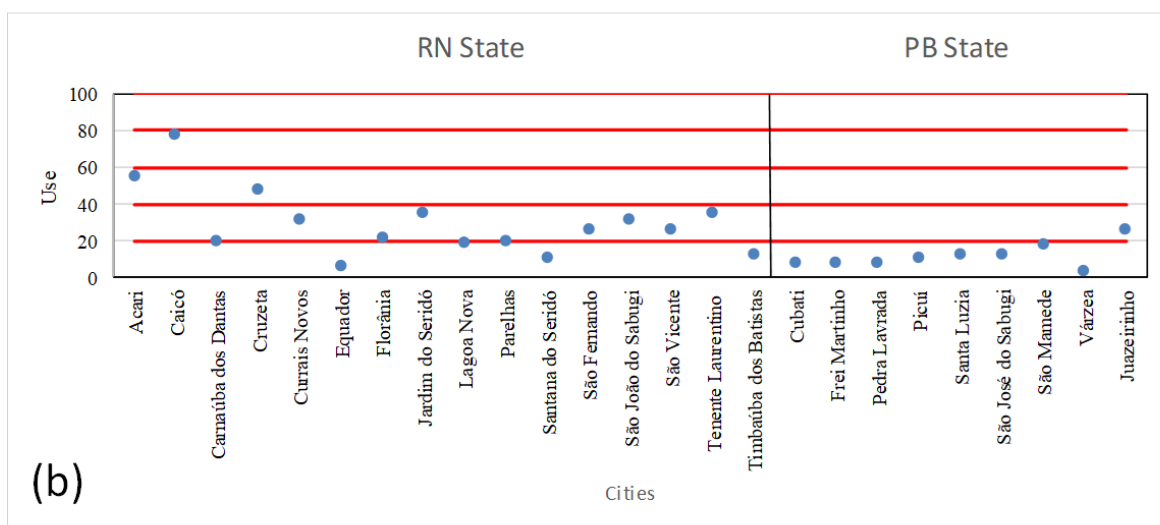
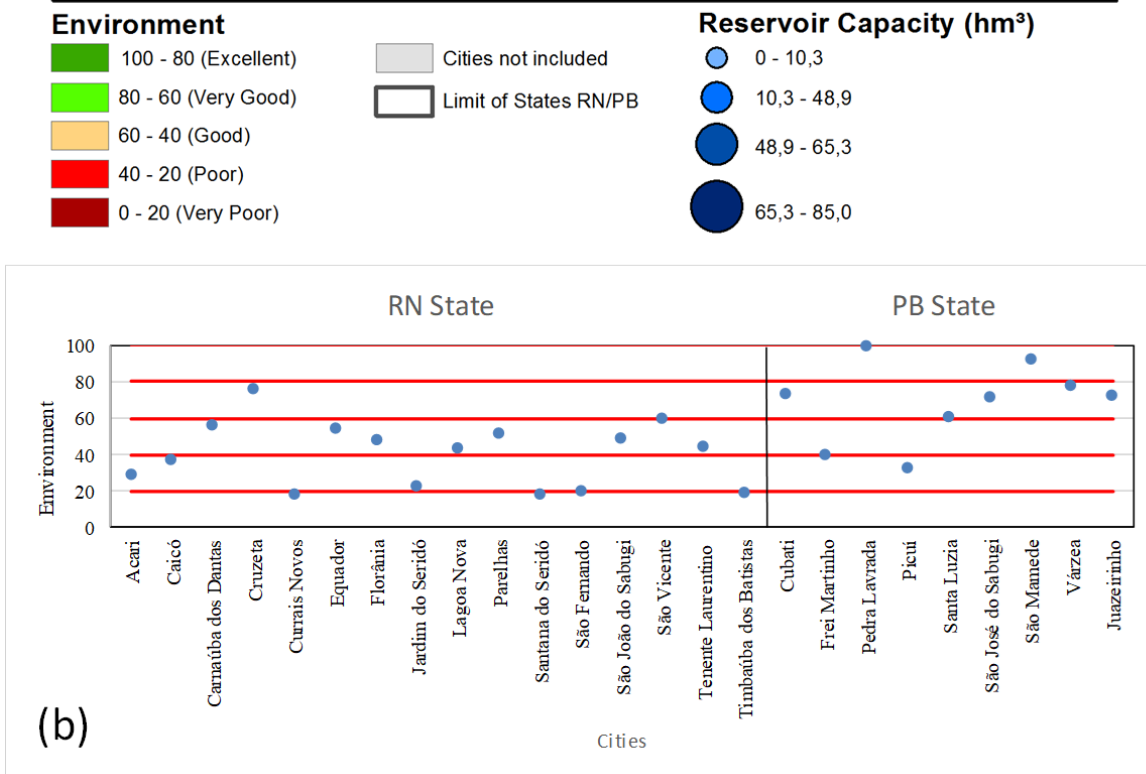**Figure 5.** Use sub-index spatial distribution (a) and its classification for the different locations (b).
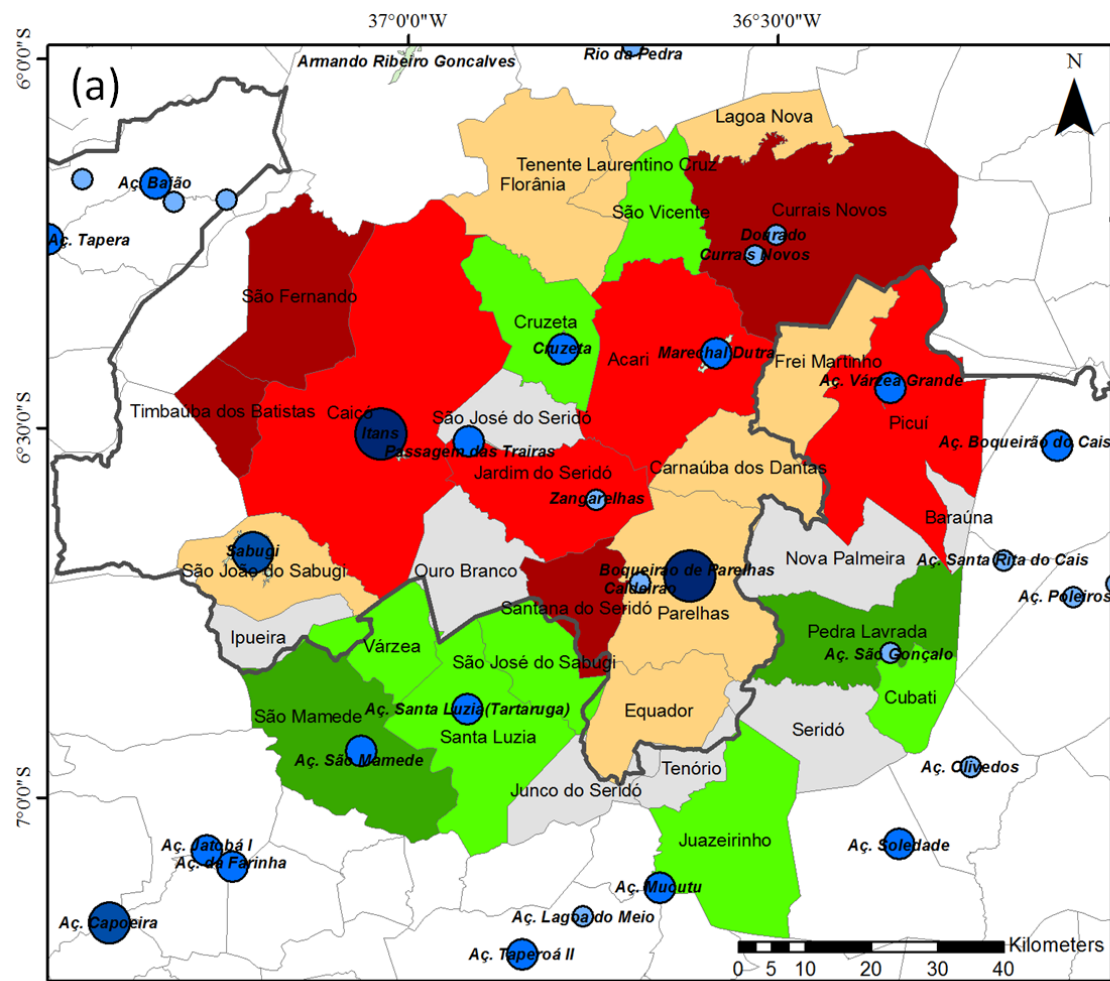
Figure 6. Environment sub-index spatial distribution (a) and its classification for the different locations (b).

than 90% of the total index value; the resource and environment sub-indexes had much lower weights. These low weights might be caused due to the basin has the same overall characteristics of water availability and environmental quality. Manandhar, Pandey, and Kazama (2012) had investigated different spatial scales and observed that there is no clear trend in the attribution of the weights of the sub-index, suggesting the need for location-specific analyses for each case.

After considering the low weights of the resource and environment sub-indexes, they were excluded from the WPI2 calculation, and the sub-indexes' weights that remained were adjusted to sum the unit, see Equation 7:

$$WPI2 = 0.2A + 0.4C + 0.4U \qquad (7)$$

Additionally, the original WPI was obtained by assigning the same weights for all sub-indexes in order to compare with each other, see Equation 8:

$$WPI = 0.2(R + A + C + U + E) \qquad (8)$$

Comparison of indexes

The results of WPI1, WPI2, and WPI from the involved cities in this study are presented in Figure 7a, b, and c. Our results show that the use of principal components to define the sub-indexes' weights influences the cities' ranking. Certain cities, such as Santa Luzia presenting the highest values for the resource and environment
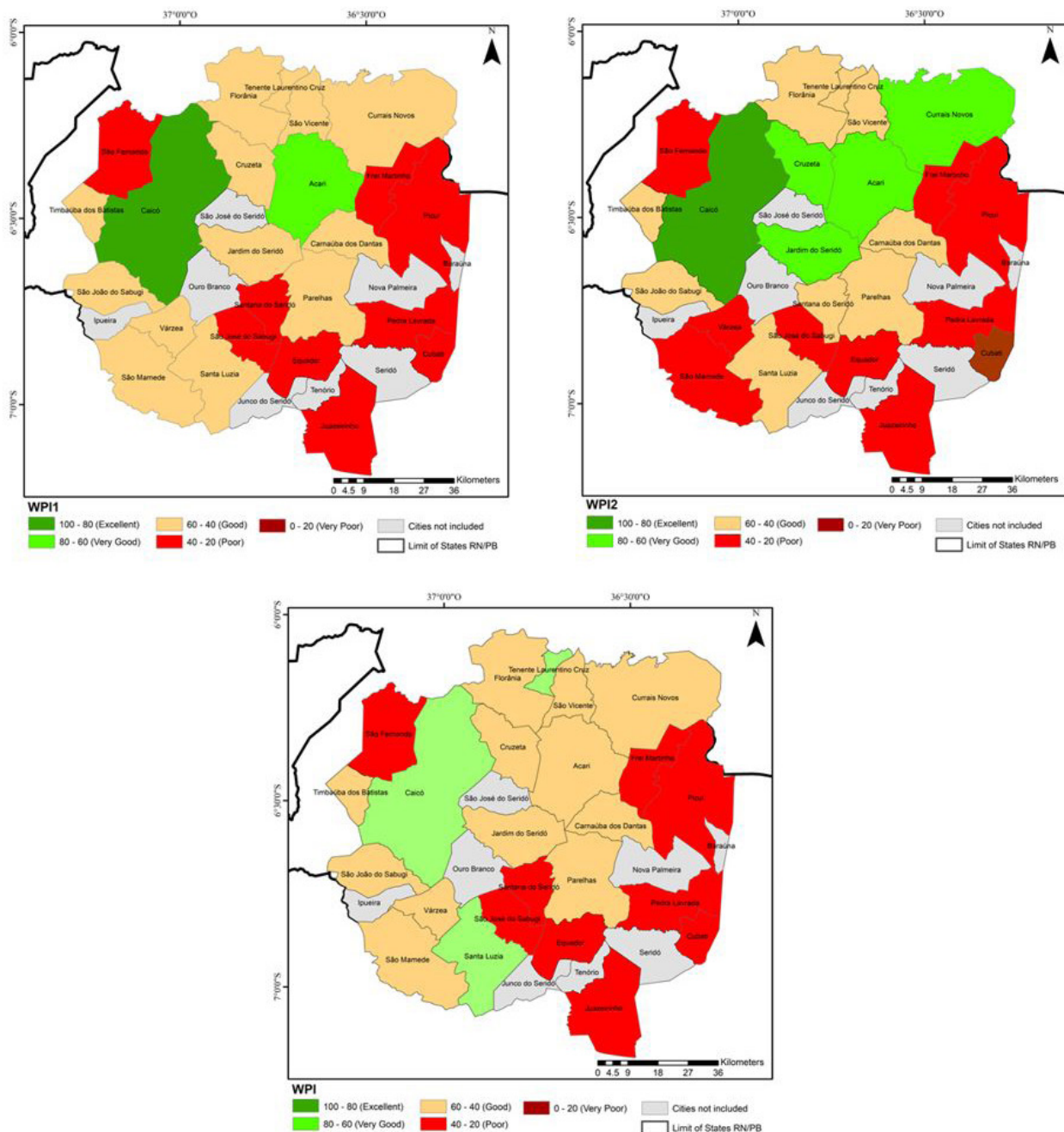

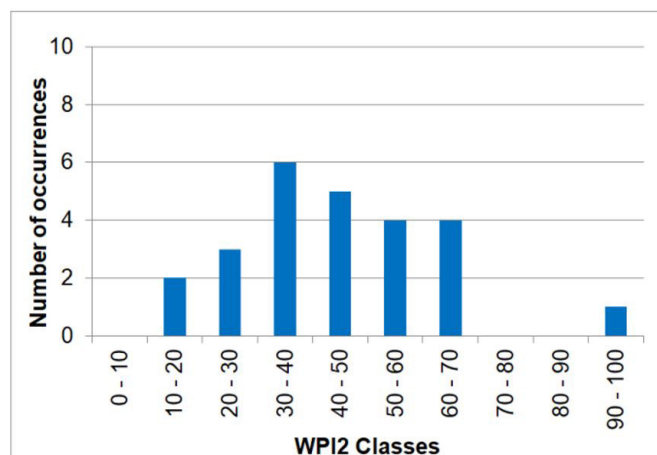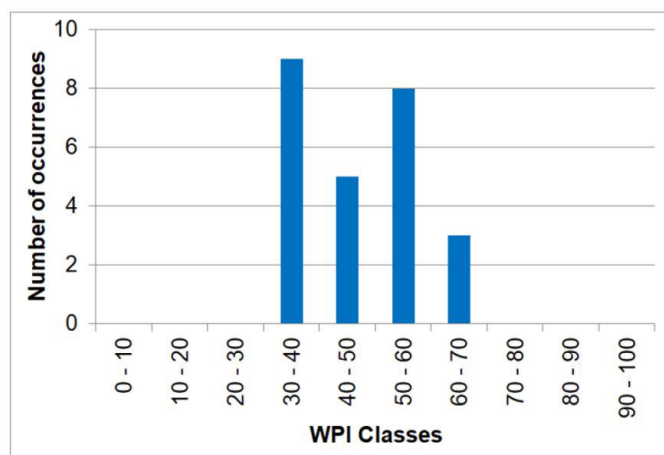**Figure 7.** Spatial distribution of WPI1, WPI2, and WPI.

Figure 8. Histogram of value frequencies from WPI and WPI2.

sub-indexes, demonstrate good WPI results and decrease in the ranking for WPI1. However, the differences are small and only four municipalities have their classes modified when principal components show weighting.

A comparison between the WPI and the WPI2 indicates that an expansion of the indexes' average values occur, in which the WPI2 presents broader results (Figure 8). In this regard, the greatest amplitude of the index allows an easier identification of the differences between the cities. However, despite the differences of the WPI values compared to the WPI2 (in the function of the different adopted weights), there is an elevated correlation among these indexes with a value of 0.73. This indicates that using the WPI2 with the lowest variable levels is a viable option for the Seridó river basin. The use of fewer variables allows the acquirement of WPI2 at a lower cost, especially in poor regions (e.g., north-eastern Brazil) where it is difficult to obtain reliable information.

Together, our results depict the general framework of our study area, and the use of WPI2 can identify critical municipals, indicating the priority of effort that should be employed to improve their water availability, as well as social, economic, and environmental quality development. Nevertheless, to determine the investment priorities, it is important to consult the sub-indexes' results as well to make a proper judgment depending on the specific area requiring further development. In fact, the association of WPI2 with the sub-indexes' results allow to point out which geographical spaces—in this case, the municipalities—have greater water poverty. Additionally, it also allows the evaluation of whether water poverty is associated with uncontrollable natural characteristics (e.g., annual precipitation) or investment deficits in the water infrastructure (the construction of reservoirs and water supply systems), which allows an assessment of whether water poverty is associated with social fragilities.

It must be highlighted that our results become more valuable in a scenario where the region has a low investment capacity. In this case, the lack of resources to invest in the fulfillment of demands require a hierarchical tool (such as WPI2) without any vices or tendencies. This tool can then direct investments to the most critical region areas to promote economic and social development and, consequently, improve the population's quality of life.

CONCLUSIONS

Using PCA to find the sub-indexes' weight has shown to be a robust methodology to assess water poverty. By applying the WPI on the Seridó basin, we were able to identify the sub-indexes use, capacity, and access as the ones with the greatest importance, which allowed the construction of a simpler index that requiring less variables (WPI2). Furthermore, the indexes obtained from the PCA methodology had the advantage of acquiring values with a broader range, making the identification of disparities among cities easier.

The evaluation of sub-indexes in the studied area made it also possible to highlight information that is usually overlooked by the global value of the index. For this reason, the sub-indexes has the potential to aid decision makers to make educated decisions regarding water resource management issues.

In conclusion, WPI2's association with the sub-indexes' allows investigators to point out which geographical space—in this case, the municipalities—presents greater water poverty. Furthermore, it makes it possible to distinguish whether water poverty is associated with natural characteristics or due to investment deficits in the water infrastructure, which in turn allows the assessment of whether the determined water poverty is associated with social fragilities.

ACKNOWLEDGMENTS

REFERENCES

ABRANTES, P. C. M. *Avaliação do desenvolvimento sustentável na bacia hidrográfica do Rio Seridó em face das políticas públicas e da gestão dos recursos hídricos*. 2011. 436 f. Dissertação (Mestrado em Saneamento Ambiental, Meio Ambiente, Recursos Hídricos e Hidráulica). Universidade Federal do Rio Grande do Norte, Natal, 2011.

AESA – AGÊNCIA EXECUTIVA DE GESTÃO DE ÁGUAS. João Pessoa: AESA; 2014. Available from: <www.aesa.pb.gov.br>. Access on: Dezembro, 2014.

ANA – AGÊNCIA NACIONAL DE ÁGUAS. Brasília: ANA; 2014. Available from: <www.ana.gov.br>. Access on: Junho, 2014.

CHO, D. I.; OGWANG, T.; OPIO, C. Simplifying the water poverty index. *Social Indicators Research*, v. 97, n. 2, p. 257-267, 2010. http://dx.doi.org/10.1007/s11205-009-9501-2.

DUNTEMAN, G. H. *Principal components analysis*. Newbury Park: Sage, 1989. http://dx.doi.org/10.4135/9781412985475.

EL-GAFY, I. K. E. The water poverty index as an assistant tool for drawing strategies of the Egyptian water sector. *Ain Shams Engineering Journal*, v. 9, n. 2, p. 173-186, 2018. http://dx.doi.org/10.1016/j.asej.2015.09.008.

EVERITT, B.; HOTHORN, T. *An Introdution to applied multicariate analysis with R (Use R!)*. New York: Springer, 2011. http://dx.doi.org/10.1007/978-1-4419-9650-3.

FEITELSON, E.; CHENOWETH, J. Water poverty: towards a meaningful indicator. *Water Policy*, v. 4, n. 3, p. 263-281, 2002. http://dx.doi.org/10.1016/S1366-7017(02)00029-6.

GINE, R.; FOGUET, P. Enhancing sector data management to target the water poor, water sanitation and hygiene: sustainable development and multisectoral approaches. In: WEDC INTERNATIONAL CONFERENCE, 34., 2009, Addis Ababa, Ethiopia. *Proceedings...* Canada: IISD, 2009. p. 286-295. Available from: <https://upcommons.upc.edu/handle/2117/7938>. Access on: 29 may 2018.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo agropecuário*. Rio de Janeiro: IBGE; 2006. Available from: <http://biblioteca.ibge.gov.br/visualizacao/periodicos/51/agro_2006.pdf>. Access on: Agosto, 2014.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo Demográfico – 2010*. Rio de Janeiro: IBGE; 2015. Access on, 2014.

IBGE – Instituto Brasileiro de Geografia e Estatística. *Pesquisa Nacional de Saneamento Básico*. Rio de Janeiro: IBGE; 2008. Available from: <https://sidra.ibge.gov.br/pesquisa/pnsb/tabelas>. Access on: Agosto, 2014.

IGARN – INSTITUTO DE GESTÃO DAS ÁGUAS DO ESTADO DO RIO GRANDE DO NORTE. Natal: IGARN; 2015. Available from: <www.igarn.rn.gov.br>. Access on: Janeiro, 2015.

JEMMALI, H.; MATOUSSI, M. S. A multidimensional analysis of water poverty at local scale: application of improved Water Poverty Index for Tunisia. *Water Policy*, v. 15, p. 98-115, 2013.

JIMÉNEZ, A.; MOLINERO, J.; PÉREZ-FOGUET, A. Monitoring water poverty: a vision from development practitioners. In: LLAMAS, M.R.; MARTINEZ-CORTINA, L.; MUKHERJI, A. (Eds.). *Water ethic:* marcelino botin water forum. London: Taylor & Francis Group, 2009. p. 151-170. http://dx.doi.org/10.1201/9780203875438-c10.

LAWRENCE, P.; MEIGH, J.; SULLIVAN, C. *The water poverty index:* an international comparison. Staffordshire: Keele University, 2002. 24 p. (Keele Economics Research Papers, n. 19). Available from: <https://econwpa.ub.uni-muenchen.de/econ-wp/dev/papers/0211/0211003.pdf>. Access on: 29 may 2018.

LISA, G. The Water Poverty Index in rural Cambodia and Viet Nam: a holistic snapshot to improve water management planning. *Natural Resources Forum*, v. 38, n. 3, p. 203-219, 2014. http://dx.doi.org/10.1111/1477-8947.12051.

MAGALHÃES JÚNIOR, A. P. *Indicadores ambientais e recursos hídricos:* realidade e perspectivas para o Brasil a partir da experiência francesa. Rio de Janeiro: Bertrand Brasil Ltda, 2007.

MANANDHAR, S.; PANDEY, V. P.; KAZAMA, F. Application of Water Poverty Index (WPI) in nepalese context: a case study of Kali Gandaki River Basin (KGRB). *Water Resources Management*, v. 26, n. 1, p. 89-107, 2012. http://dx.doi.org/10.1007/s11269-011-9907-x.

MARTINEZ-ALIER, J.; MUNDA, G.; O'NEILL, J. Weak comparability of values as a foundation for ecological economics. *Ecological Economics*, v. 26, n. 3, p. 277-286, 1998. http://dx.doi.org/10.1016/S0921-8009(97)00120-1.

MLOTE, S. D. M.; SULLIVAN, C.; MEIGH, J. Water Poverty Index: a tool for integrated water management. In: WATERNET/WARFSA SYMPOSIUM – WATER DEMAND MANAGEMENT FOR SUSTAINABLE DEVELOPMENT, 3., 2002, Dar es Salaam. *Proceedings...* Pennsylvania: The Pennsylvania State University, 2002.

MORRISON, D. F. *Multivariate statistical methods*. New York: McGraw-Hill, 1967.

NOORI, R.; SABAHI, M. S.; KARBASSI, A. R.; BAGHVAND, A.; TAATI ZADEH, H. Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. *Desalination*, v. 260, n. 1-3, p. 129-136, 2010. http://dx.doi.org/10.1016/j.desal.2010.04.053.

OTT, W. R. *Environmental indices:* theory and practice. Michigan: Ann Arbor Science, 1978.

PÉREZ-FOGUET, A.; GINÉ GARRIGA, R. G. Analyzing water poverty in basins. *Water Resources Management*, v. 25, n. 14, p. 3595-3612, 2011. http://dx.doi.org/10.1007/s11269-011-9872-4.

R CORE TEAM. *R: A language and environment for statistical computing*. Viena: R Foundation for Statistical Computing, 2012. Available from: <http://www.R-project.org/>. Access on: 29 may 2018.

RIJSBERMAN, F. Can development of water resources reduce poverty? *Water Policy*, v. 5, n. 5-6, p. 399-412, 2003. http://dx.doi.org/10.2166/wp.2003.0025.

SEMARH – SECRETARIA DE ESTADO DOS RECURSOS HÍDRICOS. Natal: SEMARH; 2014. Available from: <www.semarh.rn.gov.br>. Access on: Junho, 2014.

SICHE, R.; AGOSTINHO, F.; ORTEGA, H.; ROMEIRO, A. Índices versus indicadores: precisões conceituais na discussão da sustentabilidade de países. *Ambiente & Sociedade*, v. 10, n. 2, p. 137-148, 2007. http://dx.doi.org/10.1590/S1414-753X2007000200009.

SULLIVAN, C. Calculating a water poverty index. *World Development*, v. 30, n. 7, p. 1195-1210, 2002. http://dx.doi.org/10.1016/S0305-750X(02)00035-9.

ZHANG, Q.; LIU, B.; ZHANG, W.; JIN, G.; LI, Z. Assessing the regional spatio-temporal pattern of water stress: a case study in Zhangye City of China. *Physics and Chemistry of the Earth*, v. 79-82, p. 1474-7065, 2015.

Authors contributions

Larynne Dantas de Senna: Participated in data processing and analysis, bibliographical revision, discussion discussion of the results and paper writing.

Adelena Gonçalves Maia: General design, discussion of the results, writing and review.

Joana Darc Freire de Medeiros: General design, discussion of the results, writing and review.