

GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS APPROACH TO EVALUATE ELECTRICITY CONSUMPTION BEHAVIOUR

NATHALIA TEJEDOR-FLORES^{1,2,3}, PURIFICACIÓN VICENTE-GALINDO⁴
& PURIFICACIÓN GALINDO-VILLARDÓN⁴

¹Centro de Investigaciones Hidráulicas e Hidrotécnicas, Universidad Tecnológica de Panamá, Panama

²Sistema Nacional de Investigación, Panama

³Centro de Estudios Multidisciplinarios en Ciencias, Ingeniería y Tecnología-AIP, Panama

⁴Departamento de Estadística, Universidad de Salamanca, Spain

ABSTRACT

Electricity consumer behaviour is primarily based on individual decisions, which are often driven by external factors such as economic incentives, existing demographics, environmental variables, social norms, and infrastructure. This study aims to understand the amount of electricity used per hour dedicated in the household (HH) sectors and the paid sectors – including agriculture (AG), other productive sectors (PS) and the service and government sector (SG) – using Geographically Weighted Principal Components Analysis (GWPCA). In the literature, we found that a standard Principal Components Analysis (PCA) can be replaced with a GWPCA when we want to account for a certain spatial heterogeneity. To use the GWPCA to compare the results with a standard PCA, we took the data used in a previous investigation and applied both analyses in order to find a better way to understand the electricity consumer behaviour, using a multivariate analysis. The standard PCA reveals that the first three components collectively account for 73.66% of the variation in the data. Using GWPCA, we found a clear geographical variation in the percentage of total variance data, with higher percentages (90%–95%) located in the south-west and a small part of the north-east of the case of study used. Also, the electrical Energy Throughput in Paid Work (ET_{PW}), and the amount of energy used per hour dedicated to the Paid Work sector (EMR_{PW}), appears to play an important part in defining the local structure in the south-west (coastal region) and in the northern part of the case of study used, respectively. The comparison results suggest that GWPCA provides better fitness than the standard PCA model by considering spatial heterogeneity.

Keywords: Geographically Weighted Principal Components Analysis, multivariate analysis, sustainable development, electricity consumption.

1 INTRODUCTION

According to Bybee in 1991, environmental issues have expanded from local to global; they have evolved from minor to major, and they have extended from personal concerns to public policies. In this period, the concept of Sustainability appears as “the most necessary unifying central idea at this moment in human history” [1]. “A threatened future” is the title of the first chapter of *Our Common Future, Report of the World Commission on Environment and Development* and the first attempts to introduce the concept of sustainable development, defined as “development, which meets the needs of current generations without compromising the ability of future generations to meet their own needs” [2].

Daly in 1992, defines sustainable development in a biophysical perspective that distinguishes three fundamental aspects: (1) Renewable resources, indicating that the rate of use must not exceeds the rate of regeneration; (2) Pollution, the rate of waste generation must not exceed the assimilation capacity of the environment; and (3) Non-renewable resources, indicating the development of renewable substitutes for those resources [3]. Meanwhile, in 1995 Goodland, affirms that “sustainable development should integrate social, environmental, and economic sustainability and use these three to start to make development



sustainable” [4]. There are different definitions of sustainable development, such as Berkes and Folke in 1998, in the book “Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience” present an integrated vision of the “human being in nature”[5], in this vision the ecosystems are integrated with human society.

At the beginning of this 21st century, a new scientific domain, Sustainability Science, started to develop with the explicit aim of facing the present situation of planetary emergency to make possible the transition to sustainable societies [6]. This new science, called “Sustainability Science” was first announced by 23 researchers from different areas of knowledge, in an article published in 2011, in the journal *Science*: “A new field of sustainability science is emerging that seeks to understand the fundamental character of interactions between nature and society” [7]. Sustainability science integrates different fields, for example, economy, the study of biodiversity, energy efficiency, and many others, which have in common human actions that affect nature.

The Water-Energy-Food Nexus has emerged as a useful concept to addresses interactions and feedback between human and natural systems. It focuses on the resource base, including both biophysical and socio-economic resources, on which we depend to achieve social, environmental and economic goals pertaining to water, energy and food [8]. To understand the electricity consumption of a country, using the Water-Energy-Food Nexus, the Food and Agriculture Organization of the United Nations (FAO) and other scientific research, propose to use the Multi-Scale Integrated Analysis of Societal and Ecosystem Metabolism (MuSIASEM) approach [9]–[14]. The MuSIASEM is an innovative approach to accounting that integrates quantitative information generated by distinct types of conventional models based on different dimensions and scales of analysis [15]. Using the MuSIASEM variables, the aims of this study is to understand the amount of electricity used per hour dedicated in the household (HH) sectors and the paid sectors, included agriculture (AG), other productive sectors (PS) and the service and government sector (SG), in this case, in a country, applying Geographically Weighted Principal Components Analysis (GWPCA), in order to account for a certain spatial heterogeneity. The text is divided into five main sections: Section 1 contains the introduction; Section 2 presents the theoretical foundations of the GWPCA; Section 3 the description of variables, database and methodology used; Section 4 contains the summary of key findings and Section 5 presents the conclusions of this study.

2 THEORETICAL FOUNDATIONS

Principal Components Analysis (PCA) is widely used in many contexts for reducing the dimensionality of multivariate data [16] but, according to Harris et al. in 2011 [17], standard PCA can be replaced with a geographically weighted PCA (GWPCA), when we want to account for a certain spatial heterogeneity.

2.1 Geographically Weighted Principal Components Analysis

GWPCA is an extension of the classical PCA, while PCA analysis can provide information regarding the global internal structure; it fails to consider that the covariance structure of the data can change spatially. In essence, GWPCA performs a local PCA analysis by considering a neighbourhood around each spatial feature [18].

2.1.1 From PCA to GWPCA

Given a data matrix X composed by n rows (individuals) and m columns (variables) with a matrix of covariances Σ . The trace of Σ is the total variance in the data. If the columns in X are standardised with zero mean and unit variance, the values in Σ will be those in the



correlation matrix for X with its trace equivalent to the number of columns in X. A standard result in linear algebra states that:

$$LVL^T = R, \tag{1}$$

where V is a diagonal matrix of eigenvalues, L is a matrix of eigenvectors and the matrix R is symmetric and positive definite [17].

The principal components are found by post-multiplying the original data values X by L; the correlation matrix for XL is an identity matrix, Fig. 1.

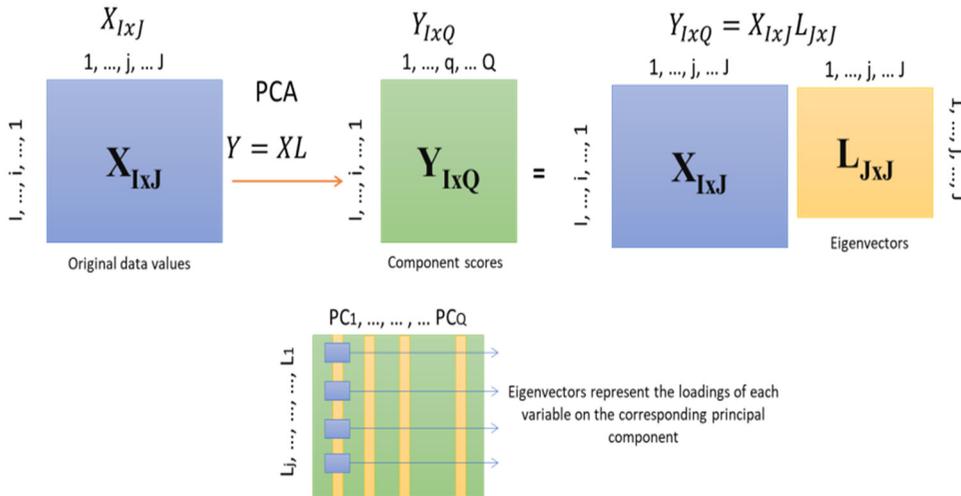


Figure 1: Principal Components Analysis.

In GWPCA, a vector of variables x_i has spatial location i is assumed to have a multivariate normal distribution with mean vector μ and variance–covariance matrix Σ that is [17]:

$$x_i \sim N(\mu, \Sigma). \tag{2}$$

Furthermore, if spatial location i has coordinates (u, v) , then PCA with local geographical effects involves regarding x_i as conditional on u and v , and making μ and Σ functions of u and v ; thus [17]:

$$x_i | (u, v) \sim N(\mu(u, v), \Sigma(u, v)). \tag{3}$$

As μ and Σ are functions of u and v , this implies that each element of $\mu(u, v)$ and $\Sigma(u, v)$ is also a function of u and v . To obtain the GW principal components, the decomposition of the GW variance–covariance matrix provides the GW eigenvalues and GW eigenvectors. The product of the i th row of the data matrix with the GW eigenvectors for the i th location provides the i th row of GW component scores. The GW variance–covariance matrix is [17]:

$$\Sigma(u, v) = X^T W(u, v) X, \tag{4}$$

where $W(u, v)$ is a diagonal matrix of geographic weights that can be generated using kernel function presented in Table 1 [19].

Table 1: Six kernel functions [19].

Global Model	$w_{ij} = 1$	(5)
Gaussian	$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right)$	(6)
Exponential	$w_{ij} = \exp\left(-\frac{ d_{ij} }{h}\right)$	(7)
Box-car	$w_{ij} = \begin{cases} 1 & \text{if } d_{ij} < h, \\ 0 & \text{otherwise} \end{cases}$	(8)
Bi-square	$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2 & \text{if } d_{ij} < h, \\ 0 & \text{otherwise} \end{cases}$	(9)
Tri-cube	$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^3\right)^3 & \text{if } d_{ij} < h, \\ 0 & \text{otherwise} \end{cases}$	(10)

w_{ij} is the j th element of the diagonal of the matrix of geographical weights $W(u_i, v_i)$, and d_{ij} is the distance between observations i and j , and h is the bandwidth.

The GW principal components for the location (u_i, v_i) is given by (11):

$$LVL^T|(u_i, v_i) = \sum(u_i, v_i), \quad (11)$$

where $\sum(u_i, v_i)$ is the GW variance–covariance matrix for location (u_i, v_i) .

3 METHODOLOGY

3.1 Case study

Ecuador, like most Latin American countries, does not have an extensive statistical culture. The dismantling of public planning during the 1980s and 1990s, plus the financial crises of the late 1990s, made the production of public statistics a low priority for the governments [20]. Others aspects which have caused serious social and environmental impacts and also, characterize the historical, economic and environmental development, not only from Ecuador but also from other Latin American countries, are deforestation, changes in land use, high rates of population growth, and the deterioration of tropical forests, causing important losses of biodiversity in this areas [9]. We used the unit of administrative and territorial division of Ecuador called a “canton”. Cantons are the second-level subdivisions of Ecuador, below the provinces and they are formed by an urban centre, and its rural parishes. A total of 221 cantons were used in this study because with studies at this geographical level we can reach more detailed conclusions.



3.2 Description of variables

The database used includes 26 variables that measure characteristics of human societies in the 221 cantons of Ecuador. This database has been used in a previous study, related also with multivariate analysis, using BIPLLOT Approach [21]. See Table 2 [22].

We used three main economic sectors (activity i): Agriculture, livestock and fishing sector (AG); Industrial sector (PS); and Service and Government sector (SG). The population and activity of economic sectors data were obtained from the statistics published by the National Institute of Statistics and Censuses of Ecuador Government, gross domestic product from the Central Bank of Ecuador and energy data from the Statistics of the Electric Sector of the Ecuador Government. See Table 3.

Table 2: Description of the MuSIASEM variables used [22].

Variable	Description	Unit
Total Energy Throughput (TET)	Total energy sources that are used for generating electricity in an economy in one year	Mega joules (MJ)
Total Human Activity (THA)	Total human time a society has available for conducting different activities	Hours (h)
Gross Domestic Product (GDP)	Added value generated by an economy in one year	US dollars (\$)
Energy Throughput in activity i (ET_i)	Energy sources that are used for generating electricity in activity i , in one year	Mega joules (MJ)
Human Activity in activity i (HA_i)	Human time a society has allocated to activity i	Hours (h)
Added value generated by activity i (GDP_i)	Sum of the value added from the various economic sectors	US dollars (\$)
Exosomatic Metabolic Rate, average of the society (EMR_{SA})	The amount of electricity per hour of human time for the whole society	MJ/h
Exosomatic Metabolic Rate for activity i (EMR_i)	The amount of electricity used per hour dedicated to each sector	MJ/h
Economic labour productivity for activity i (ELP_i)	Added value per hour of working time in activity i	\$/h
Economic Energy Intensity (EEI)	Energy consumption in electricity per unit of added value	(MJ/\$)

Table 3: Summary of the online sources used in the database.

Database	Source
Population	https://bit.ly/3hgCueB
Employment	https://bit.ly/31k0IUZ
GDP	https://bit.ly/34qEZYm
Energy	https://bit.ly/3hknexq



3.3 Data analysis

The database from the previous step was organized in an $I \times J$ data matrix, where rows I correspond to the 221 Cantons of Ecuador and columns J correspond to 26 variables. We used the Gwmodel R package, developed by Gollini et al. in 2015 [19], to apply the GWPCA in our database. The outputs of this analysis are mapped to provide a useful exploratory tool into the nature of the data spatial heterogeneity. The Gwmodel R package includes functions for: GW summary statistics, GW principal components analysis, GW regression, and GW discriminant analysis; some of which are provided in basic and robust forms [19]. We added to the data matrix the geographic location (u_i, v_i) in WGS 84/UTM zone 18N coordinate system, of the 221 cantons and used a bi-square function to generated the diagonal matrix of geographic weights $W(u, v)$, Fig. 2.

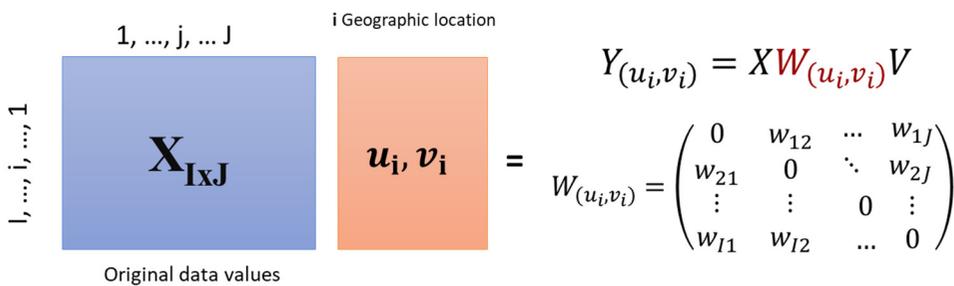


Figure 2: Geographically Weighted Principal Components Analysis.

4 RESULTS

4.1 Principal Components Analysis

We standardise the data and specify the global PCA with the covariance matrix, the same standardised data are also used in the GWPCA calibrations, which are similarly specified with covariance matrices. The effect of this standardization is to make each variable have equal importance in the subsequent analysis [19]. The standard PCA reveals that the first three components collectively account for 73.66% of the variation in the data. See Table 4.

Table 4: Variability % explained of global PCA.

Components	% of Variance	Cumulative %
1	54.93	54.93
2	10.41	65.34
3	8.32	73.66

In Table 5, component 1 appears to represent the Total Human Activity (THA) and Human Activity in Household (HAHH); component 2, the Economic Labour Productivity in Paid Work (ELPPW) and component 3, the Exosomatic Metabolic Rate in Household (EMRHH). These are whole-country statistics and interpretations [19], and they may not represent the local social structure of each Ecuador canton.

Table 5: Loadings of global PCA.

Variable	Comp.1	Comp.2	Comp.3
THA	-0.252	-0.059	0.020
HA _{AG}	-0.192	-0.061	0.023
HA _{PS}	-0.249	-0.062	0.025
HA _{SG}	-0.250	-0.062	0.023
HA _{PW}	-0.250	-0.063	0.024
HA _{HH}	-0.252	-0.058	0.020
TET	-0.244	0.030	-0.139
ET _{HH}	-0.199	0.043	-0.360
ET _{PW}	-0.245	0.018	0.012
ET _{SG}	-0.235	0.055	0.016
ET _{AG}	-0.238	-0.032	0.012
ET _{PS}	-0.238	-0.005	0.008
EMR _{SA}	-0.027	0.275	-0.559
EMR _{HH}	-0.014	0.139	-0.598
EMR _{SG}	-0.012	0.376	0.093
EMR _{AG}	-0.196	0.066	0.012
EMR _{PS}	-0.024	0.294	-0.029
EMR _{PW}	-0.038	0.474	0.042
GDP	-0.247	0.001	0.068
GDP _{SG}	-0.245	-0.065	0.024
GDP _{PS}	-0.224	0.100	0.136
GDP _{AG}	-0.168	0.067	0.016
ELP _{PW}	-0.021	0.412	0.262
ELP _{SG}	-0.049	0.046	0.033
ELP _{PS}	-0.012	0.370	0.264
ELP _{AG}	-0.036	0.278	0.007

4.2 Geographically Weighted Principal Components Analysis

A major challenge in GWPCA is in the estimation of the bandwidth [19]. For this study, we calibrated the GWPCA with a bi-square kernel using adaptive bandwidths whose sizes are chosen automatically and objectively via cross-validation [23]. The bi-square function is useful as it can provide an intermediate weighting between the box-car and the Gaussian functions [24].

For visualizing the output from GWPCA, we focused on: how data dimensionality varies spatially and how the original variables influence the components [19]. To visualize and interpret the results, we mapped the spatial distribution of the local percentage of variance explained and the first three principal components.

Fig. 3(a) shows the spatial distribution of the percentage of variance explained for the first three principal components. The fact that the variance changes throughout the study area also suggests the advantage of using GWPCA against standard PCA [26]. The percentage of variance explained for the first three local components in the GWPCA exceeds the percentage of variance explained for the first three components in the standard PCA (73.66%), in some cantons of Ecuador. There is a clear geographical variation in the results of the percentage of variance explained, with higher percentages (90%–95%) located in the south-west (coastal

region) and a small part of the northeast (Amazon region), Fig. 3(b). The GWPCA offers the possibility of making not only a conjunct interpretation for all data but also as many analyses as there are data, according to their location [26].

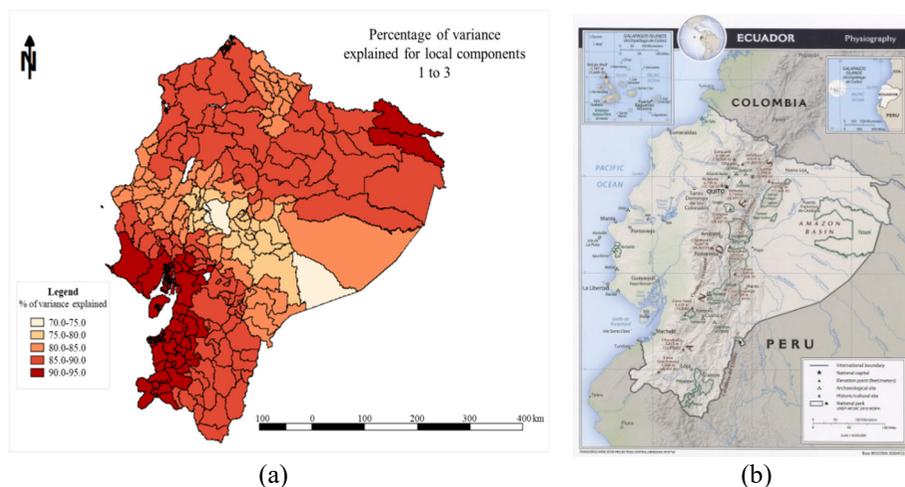


Figure 3: (a) Percentage of variance explained for the first three local components; and (b) Physiography map of Ecuador [25].

In Figs 4, 5, and 6 we can see the spatial distribution of the winning (highest absolute local loading in the corresponding component [17]) variables for the first three geographically weighted components. In Fig. 4, we can observe that the electrical Energy Throughput in Paid Work (ET_{PW}) and the amount of energy used per hour dedicated to the Paid Work sector (EMR_{PW}), appears to represent an important part in defining the local structure in the south-west (coastal region) and in the northern part of Ecuador, respectively. In the literature, we can find that the Exosomatic Metabolic Rate of the productive sector (EMR_{PW}) can be used as a proxy for the level of capitalization of the economy [27].

In the principal component 2, the energy sources that are used for generating electricity in the industrial sector (ET_{PS}) plays an important role in the north-east part of Ecuador (coastal region); and the variable Economic Labour Productivity per hour (ELP_{PW}) appears to be important in the north-western part of Ecuador (Amazon region). In the cantons located in the central part of Ecuador, the variable related to Economic Labour Productivity in the Services and Government sector (ELP_{SG}) is the one to appear important in this part of the country (see Fig. 5).

In component 3, the amount of electricity used per hour dedicated in the Agriculture sector (EMR_{AG}) is the winning variable in the north-east and the variable related with the amount of electricity used per hour in society (EMR_{SA}) is the winning variable in the north-west and part of the cantons located in the centre of Ecuador. In the cantons located in the southern part of Ecuador, the winning variable is the one related to the Gross Domestic Product generated by the Agriculture sector (GDP_{AG}). See Fig. 6.

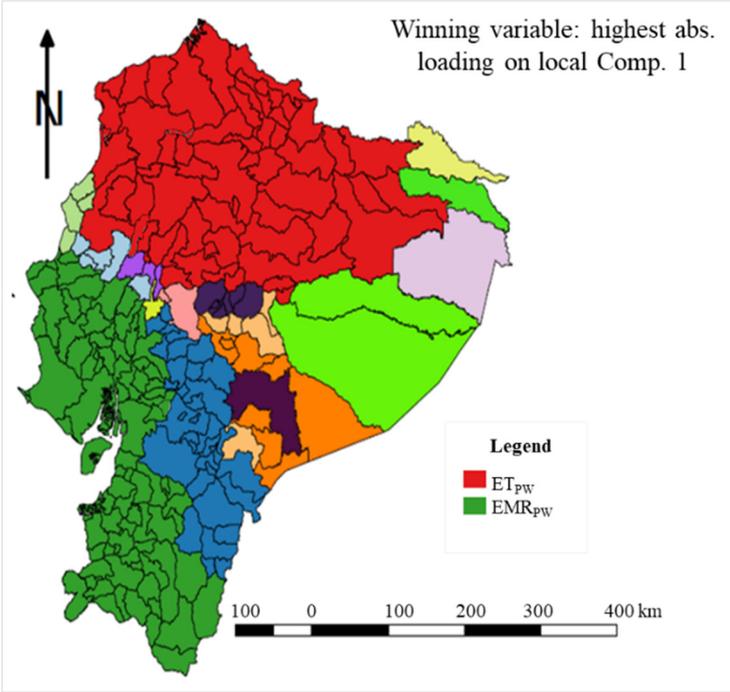


Figure 4: Winning variable on local component 1.

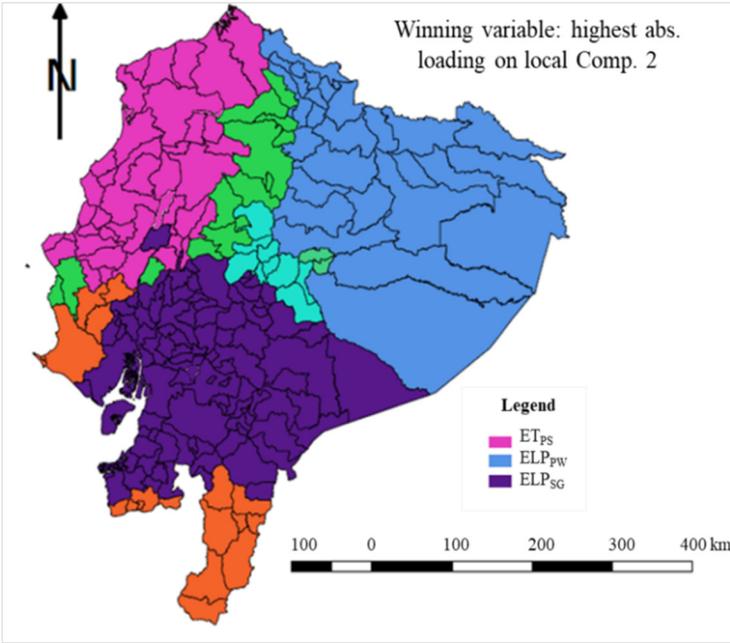


Figure 5: Winning variable on local component 2.

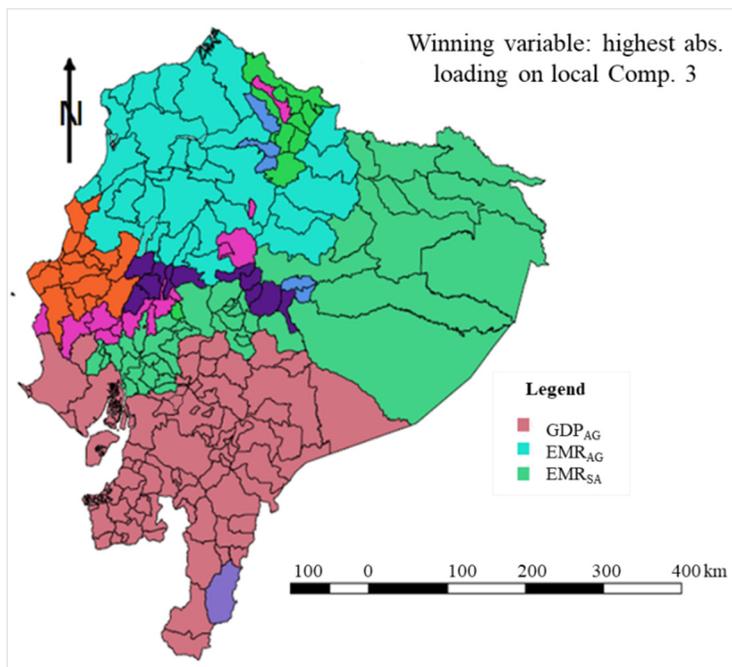


Figure 6: Winning variable on local component 3.

5 CONCLUSIONS

Concluding remarks:

1. In the bibliographic review, we found that similar investigations usually study large geographic units (for example countries, regions or provinces). This investigation at a smaller geographical level (cantons) provides more detailed results that facilitate the implementation of local improvement and development strategies in the country.
2. In this study we used the GWPCA, to account for a certain spatial heterogeneity of the database, and we found an optimal bandwidth using the Kernel Bi-square function.
3. The percentage of variance explained, using the GWPCA, for the first three local components exceeds the percentage of variance explained for the first three components in the standard PCA (73.66%), in some cantons of Ecuador. The higher percentages of variance explained (90%–95%) are located in the south-west (coastal region) and a small part of the northeast (Amazon region).
4. One of the advantages of using the GWPCA is to be able to visualize and interpret the factors that influence the analysis of sustainable development, focusing on how data dimensionality varies spatially, thus obtaining, a greater degree of detail and precision that can be used to support local decision making.

ACKNOWLEDGEMENT

The authors express their appreciation to the National System of Investigation of Panama (SNI) at SENACYT for supports research activities of Nathalia Tejedor-Flores.



REFERENCES

- [1] Bybee, R.W., Planet Earth in crisis: How should science educators respond? *The American Biology Teacher*, **53**(3), pp. 146–153, 1991.
- [2] Brundtland, G.H., Report of the World Commission on Environment and Development: Our Common Future, The Brundtland Report, 1987.
- [3] Daly, H.E., Allocation, distribution, and scale: Towards an economics that is efficient, just, and sustainable. *Ecological Economics*, **6**(3), pp. 185–193, 1992.
- [4] Goodland, R., The concept of environmental sustainability. *Annual Review of Ecology, Evolution and Systematics*, **26**, pp. 1–24, 1995.
- [5] Berkes, F. & Folke, C., *Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience*, Resilience Alliance: Cambridge, 1998.
- [6] Vilches, A., Ciencia de la Sostenibilidad: ¿Una nueva disciplina o un nuevo enfoque para todas las disciplinas? *Revista Iberoamericana de Educación*, **69**(1), pp. 39–60, 2015.
- [7] Kates, R.W. et al., Sustainability science. *Science*, **292**(5517), pp. 641–642, 2001.
- [8] Flammini, A., Puri, M., Pluschke, L. & Dubois, O., Walking the nexus talk: Assessing the water-energy-food nexus in the context of the sustainable energy for all initiative. <https://bit.ly/2XwIvvO>. Accessed on: 30 Jul. 2020.
- [9] Falconí-Benítez, F., Integrated assessment of the recent economic history of Ecuador. *Population and Environment*, **22**(3), pp. 257–280, 2001.
- [10] Ramos-Martin, J., Giampietro, M. & Mayumi, K., On China's exosomatic energy metabolism: An application of multi-scale integrated analysis of societal metabolism (MSIASM). *Ecological Economics*, **63**(1), pp. 174–191, 2007.
- [11] Ramos-Martín, J., Cañellas-Boltà, S., Giampietro, M. & Gamboa, G., Catalonia's energy metabolism: Using the MuSIASEM approach at different scales. *Energy Policy*, **37**(11), pp. 4658–4671, 2009.
- [12] Sorman, A.H. & Giampietro, M., Generating better energy indicators: Addressing the existence of multiple scales and multiple dimensions. *Ecological Modelling*, **223**(1), pp. 41–53, 2011.
- [13] Recalde, M. & Ramos-Martin, J., Going beyond energy intensity to understand the energy metabolism of nations: The case of Argentina. *Energy*, **37**(1), pp. 122–132, 2012.
- [14] Biggs, E.M. et al., Sustainable development and the water-energy-food nexus: A perspective on livelihoods. *Environmental Science and Policy*, **54**, pp. 389–397, 2015.
- [15] Giampietro, M., Aspinall, R.J., Ramos-Martin, J. & Bukkens, S.G.F., *Resource Accounting for Sustainability Assessment: The Nexus Between Energy, Food, Water and Land Use*, 1st ed., Routledge: Abingdon, 2014.
- [16] Lloyd, C.D., Analysing population characteristics using geographically weighted principal components analysis: A case study of Northern Ireland in 2001. *Computers, Environment and Urban Systems*, **34**(5), pp. 389–399, 2010.
- [17] Harris, P., Brunson, C. & Charlton, M., Geographically weighted principal components analysis. *International Journal Geographical Information Science*, **25**(10), pp. 1717–1736, 2011.
- [18] Fernández, S., Cotos-Yáñez, T., Roca-Pardiñas, J. & Ordóñez C., Geographically Weighted Principal Components Analysis to assess diffuse pollution sources of soil heavy metal: Application to rough mountain areas in Northwest Spain. *Geoderma*, **311**, pp. 120–129, 2018.



- [19] Gollini, I., Lu, B., Charlton, M., Brunson, C. & Harris, P., GWmodel: An R package for exploring spatial heterogeneity. *Journal of Statistical Software*, **63**(17), pp. 1–50, 2015.
- [20] Villacís, B. & Carrillo, D., Estadística demográfica en el Ecuador: Diagnóstico y propuestas. <https://bit.ly/3ieRI3H>. Accessed on: 30 Jul. 2020.
- [21] Tejedor-Flores, N., Vicente-Galindo, P. & Galindo-Villardón, P., Sustainability multivariate analysis of the energy consumption of Ecuador using MuSIASEM and BIPLLOT approach. *Sustainability*, **9**(6), pp. 1–15, 2017.
- [22] Ginard-Bosch, F.J. & Ramos-Martín, J., Energy metabolism of the Balearic Islands (1986–2012). *Ecological Economics*, **124**, pp. 25–35, 2016.
- [23] Wei, C., Cabrera-Barona, P. & Blaschke, T., Local geographic variation of public services inequality: Does the neighborhood scale matter? *International Journal of Environmental Research and Public Health*, **13**(10), p. 981, 2016.
- [24] Lu, B., Harris, P., Charlton, M. & Brunson, C., The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geospatial Information Science*, **17**(2), pp. 85–101, 2014.
- [25] Central Intelligence Agency, Physiography map of Ecuador. <https://bit.ly/3hIzuXK>. Accessed on: 30 Jul. 2020.
- [26] Roca-Pardiñas, J., Ordóñez, C., Cotos-Yáñez, T.R. & Pérez-Álvarez, R., Testing spatial heterogeneity in geographically weighted principal components analysis. *International Journal Geographical Information Science*, **31**(4), pp. 676–693, 2017.
- [27] Velasco-Fernández, R., Ramos-Martín, J. & Giampietro, M., The energy metabolism of China and India between 1971 and 2010: Studying the bifurcation. *Renewable and Sustainable Energy Review*, **41**(1), pp. 1052–1066, 2015.

