Fuzzy and spatial analysis of cutaneous leishmaniasis in Pará State, Brazilian Amazon: an ecological and exploratory study

Simone BN Costa1, Claudia do SC Miranda 1, Bruna C de Souza1, Heloisa Maria M e S Guimarães1, Camylle MC Faria1, Pedro S da S Campos2, Taiana MA Koury1, José Gabriel M da Paixão1, Alessandra L Leal1, Maria de Fátima P Carrera1, Silvana R de Brito2, Nelson V Gonçalves1,2

1 Amazon Epidemiology and Geoprocessing Laboratory, Faculty of Medicine, Pará State University, Belém, Brazil
2 Cyberspace Institute, Amazon Rural Federal University, Belém, Brazil

Abstract
Introduction: This study sought to analyze the relationships between cutaneous leishmaniasis and its epidemiological, environmental and socioeconomic conditions, in the 22 microregions of Pará state, Brazil, for the period from 2017 to 2022.

Methodology: In this ecological and exploratory study, the microregions were used as spatial units because they are formed by contiguous municipalities with similar characteristics. The epidemiological, environmental, socioeconomic, and public health policy data employed were obtained from the official information systems at the Ministry of Health, National Institute for Space Research, and Brazilian Institute of Geography and Statistics. A fuzzy system was developed to identify risk factors for the disease, using Python programming language. The results were analyzed with the bivariate Global Moran spatial analysis technique.

Results: It was observed that the Altamira microregion had the highest risk percentage for the disease, while Breves had the lowest, with significant differences in the relevance of its conditioning factors, mainly related to land use and cover patterns, in addition to demography and living conditions index, education and public health policies.

Conclusions: The fuzzy system associated with the geostatistical technique was satisfactory for identifying areas with health vulnerability gradients related to deforestation, pasture, poverty, illiteracy, and health services coverage, as its conditioning variables. Thus, it was demonstrated that deforestation was the main risk factor for the disease. The system can also be used in environmental and epidemiological surveillance.

Key words: Cutaneous leishmaniasis; epidemiology; spatial analysis; fuzzy logic.

Introduction
Cutaneous leishmaniasis (CL) is an infectious disease caused by protozoans of the genus Leishmania. The main vectors for this disease are phlebotomine flies of the genus Lutzomyia, usually known in English as sandflies. CL is considered a major public health problem due to the work disabilities that it causes and its relationship with environmental and socioeconomic conditions [1,2].

In recent decades over 1,000,000 cases of CL have been recorded in Latin America, with an average of 53,387 per year and a decreasing trend, although in 2020 Brazil saw a slight increase in cases, which may be related to the COVID-19 pandemic [3]. During the same period, the state of Pará, which has 7,581,051 inhabitants and is located in the northern region of Brazil has undergone major environmental and socioeconomic changes due to the implementation of developmental policies that entail adverse epidemiological scenarios in its territory. These changes may be observed at distinct levels in its 144 municipalities, which make up 22 microregions [4].

In that context, a significant increase in risk factors for CL was found, such as deforestation, mining, and pastures associated with the establishment of settlements that have precarious infrastructure and public services, including health coverage and health establishments (CHE). These conditions are found at different levels in the municipalities that form the microregions of Pará. It is thus possible that their association is related to the unequal occurrence of the disease in these territories [5].

Digital cartography and spatial statistics are currently widely employed in Epidemiological and Environmental Surveillance for CL. However, the deterministic dimension of their techniques limits the production of contextualized information because of possible inaccuracies generated in the relation between...
the disease and environmental and socioeconomic variables [6], as well as inducing elements (migratory flow and increase in disorderly urbanization) whose consequences are observed at different geographical scales.

In this context, the use of systems based on fuzzy logic can contribute to the production of more significant information about the occurrence of the disease in the state of Pará. This is especially true when compared to deterministic systems that operate with binary logic and present limitations to the processing of imprecise qualitative data, such as the classification of deforestation as very low, low, moderate, high and very high [7]. This technique, that was originally developed in Artificial Intelligence science, enables the processing of imprecise data and has been used for diagnostic and prognostic systems in the health field, and more recently in epidemiology [8-10].

In light of the facts presented above and in order to contribute to the production of an epidemiological memory of the environmental production of CL in the Amazon, the purpose of this study was to analyze the relation between the risk of CL establishment (RCLE) and the epidemiological, environmental and socioeconomic conditionings in microregions of the state of Pará, for the 2017 to 2022 period, using a fuzzy system.

Methodology

In this ecological and exploratory study, a fuzzy system entitled EPISIS – FUZZY V. Alpha was developed for estimating the risk factors for CL in 22 microregions of the state of Pará. For that purpose, data on environmental variables (deforestation, forest cover, pasture/mining, and secondary vegetation) were obtained from the TerraClass project of the Brazilian Space Research Institute [11]. Socioeconomic data on education and income were obtained from the 2010 Census of the Brazilian Institute of Geography and Statistics. Epidemiological data on the disease endemicity based on prevalence and CHE in the microregions were obtained from the Notifiable Disease Information System and the National Registry of Health Establishments of the Ministry of Health.

To operationalize the system, it was divided into three modules (input, processing, and output), which executed different routines. Thus, the input module was responsible for fuzzifying the data from the previously specified variables. For the environmental variables, the percentage of their occurrence in the microregions was used. For epidemiological variables, it was necessary to calculate the number of CL cases and of health establishments in relation to their population. For socioeconomic variables, the illiteracy rate and average per capita income per microregion were considered.

In terms of processing, the system was developed to operate with an inference engine based on a Mamdani algorithm [12], whose rules, were modeled by epidemiology specialists and drawn up according to the following syntax: If “antecedent” then “consequence”, with two or more fuzzified data being associated through logical operators and the consequent resulting in output values for the RCLE (Figure 1).

Thus, 10 rules were developed by combining the input variables, in order to correlate them with the output data. This process sought to combine the weights of the consequences of rules that were activated (with a degree of pertinence greater than zero) into a single outlet, through a normal distribution of degrees of activation. Next, defuzzification was performed based on the centroid method to transfer the fuzzy value into a real value (crisp) [12].

In the output module the response variable was the risk of CL establishment (RCLE) for each microregion, classified into five ranges: very low (0.0 – < 0.2), low (0.2 - < 0.4), moderate (0.4 - < 0.6), high (0.6 - < 0.8) and very high (0.8 - < 1.0), according to the percentage of occurrence of the variables, considering the gaussian output function. The system was developed in the Google Colab virtual environment, using the scikit-fuzzy 0.4.2 library of the Python version 3.7 programming language. The source code is available at the library repository of the Federal Rural University of the Amazon.

In order to identify the spatial context related to the results obtained by the fuzzy system, the Global Moran’s bivariate index (I) technique was used, which correlated the input system variables considering their occurrence in the municipalities that form the

Figure 1. Example of a rule developed for assessing high RCLE.

IF endemicity is high OR moderate AND (forest is low OR moderate AND (deforestation is high OR (mining is high OR moderate)) OR (pasture is high OR moderate) OR secondary_veg is high) AND ((income is low OR moderate) OR (literacy is low OR moderate) OR (CSES is low OR moderate)) THEN FR_leish is high

Description of a fuzzy code from the rule base.
microregions associated with the highest and lowest RCLE. In this way, the relationships between the areas with the occurrence of CL cases and with the environmental, socioeconomic, and epidemiological variables were analyzed intraregionally, considering a direct relationship for $I > 0$, an inverse relationship for $I < 0$, and a strong relationship for close values of the variation limits ($-1$ and $1$). The indexes were expressed in a thematic map produced with ArcGIS 10.5.1 software.

Seeking to contextualize the results obtained, two observational fieldwork activities were carried out to confirm the relationships identified. This study was approved (no. 3.245.271/2019) by the Research Ethics Committee of the Universidade do Estado do Pará (Pará State University), in accordance with the norms of Resolution no. 466/12 of the National Health Council.

**Results**

After the fuzzy processing of environmental, socioeconomic, and epidemiological data for the 22 microregions in the state of Pará it was noted that the Altamira microregion, presented the highest risk for CL establishment (0.9195) and Breves presented the lowest (0.3275), as shown in Figure 2.

The Altamira microregion identified as having the highest RCLE in the state presented the following output variables values as the result of defuzzification: moderate-income (0.6130), high illiteracy (0.8540), high pasture/mining (0.7810), high secondary vegetation (0.8050), low forest (0.4023), high deforestation (0.7838), very high endemicity (0.8015) and high CHE (0.7243) (Figure 3).

As for the Breves microregion where the lowest RCLE in the state was found, the results of defuzzifying the output variables were: very low income (0.1254), very high illiteracy (0.9533), very low pasture/mining (0.1745), very low secondary vegetation (0.1898), very high forest (0.9787), very low deforestation (0.1567), very low endemicity (0.1098) and moderate CHE (0.4276) (Figure 4).

The spatial analysis, using the Moran technique, considering the occurrence of the disease in relation to the epidemiological, environmental, and socioeconomic variables from the municipalities that make up the Altamira microregion showed a weak direct autocorrelation for all variables. In contrast, in Breves the CHE, income, pasture/mining, and deforestation variables showed an inverse and weak relation while education was inverse and strong. As for secondary vegetation and forest, a direct and weak spatial dependence was observed, as shown in the Moran’s index scale in Figure 5.

**Discussion**

The occurrences of a higher risk for CL in Altamira and a lower percentage in Breves generally follow the indicators for the occurrence of the disease in the state of Pará over the last two decades [13]. That scenario may be related to the evidence for similar land use and cover patterns, the implementation of public health policies and the living conditions of the populations in those two microregions, as observed in the analysis of spatial autocorrelation between them using the Moran technique and the fieldwork.

In the Altamira Microregion, which has one of the largest populations in the state, the fuzzy system showed a low percentage of forest cover and high percentages of both deforestation and secondary vegetation.
Figure 3. Defuzzification of the most expressive RCLE socioeconomic, environmental and epidemiological variables for the Altamira microregion, state of Pará.

Membership functions: very low, low, moderate, high, very high, crisp values.

Figure 4. Defuzzification of the most expressive RCLE socioeconomic, environmental and epidemiological variables for the Breves microregion, state of Pará.

Membership functions: very low, low, moderate, high, very high, crisp values.
Those human impacts in the municipalities that make up this microregion may be associated with its historical occupation process that occurred more significantly after the construction of the BR-230 highway known as the Transamazônica, which sparked an intense and continuous migratory inflow in the region, encouraged by a macro-developmentalist colonization policy [14,15]. That activity brought impacts on public health in that region, especially an increase in infectious diseases [5].

This process was accompanied by aggressive deforestation that led to the continuous removal of vegetation cover and the disorganized establishment of urban centers. As a result of those activities, large areas may be found in the municipalities of this region that have ecological succession processes characterized by early and advanced stages of secondary vegetation. This scenario may be associated with an expansion of sites with active transmission of CL as their vectors adapt to the human-impacted environment, as has been observed in several other territories that have the same characteristics [16,17].

In the Altamira Microregion, one may also observe a high percentage of areas in pasture/mining activities, which illustrates the anthropic pressure that these municipalities have faced because of their economic potential, as illustrated by the expansion of agroindustry and mining [18]. In that context, the existence of large areas of placer gold mining and pastures in municipalities such as Altamira and Pacajá have produced risk factors for CL and have exposed their populations to that disease. This fact highlights the inefficiency of public policies for sustainable forest use, which encourage the environmental production of CL, as well as illustrating the need for environmental surveillance in such areas.

**Figure 5.** Land use and cover: A) Altamira Microregion; B) Breves Microregion.
In Altamira Microregion high illiteracy and moderate income reflect the socioeconomic vulnerability of the population and are risk factors for CL. The low levels of education observed may be associated with a lack of knowledge regarding the occurrence of the disease and the appropriate prophylactic methods [5]. In light of this reality, health must be expanded in the basic education curriculum, particularly in areas with low education levels and environmental conditions that are favorable for transmitting the disease. This situation is also observed with several other infectious diseases in the Amazon [19].

The high percentage of health coverage services observed in Altamira compared to other microregions may be related to the geographical and demographic conditionings observed in its municipalities, since most of their inhabitants live in urban and peri-urban areas. That circumstance favors the organization of public health services, which tend to concentrate on municipal seats, with access facilitated by the BR-230 highway [20,21]. The relation between demography, geography, and availability of health services is most striking in Altamira, with a population that is 85% urban and an elevated level of health service coverage.

The very high percentage of forest cover and very low rates of deforestation and secondary vegetation in Breves may be related to the low population level. Furthermore, its population is made up largely of traditional riverbank inhabitants, descended mostly from Indigenous and Quilombola groups, whose economic activities have a low level of impact [22]. In this microregion, the respect its inhabitants have for forest conservation leads to an environmental balance, which may limit transmission cycles for CL and other vector-borne diseases that occur due to work practices in natural environments (forest product harvesting, fishing and subsistence agriculture), with exposure to with risk factors for the disease [23].

The very low percentage of pasture/mining in this microregion as large-scale activities may be related to environmental characteristics such as type of soil and geomorphology, which do not favor such activities in the region’s municipalities. The low rates of occurrence for those activities thus reduce the production of risk factors for CL compared to other microregions where ranching and mining are more widespread [24,25].

Illiteracy is very high in the Breves microregion, and income levels are very low. This highlights the social inequalities faced by populations in this territory due to precarious implementation of public policies for education and employment to encourage its inhabitants to remain in the region. This lack has direct implications for health conditions among these populations, who as natives of the region have a profound respect for the forest and the rivers. Access to information on sustainable environmental management in formal education programs may reduce the production of risk factors for the disease [5,19,25].

The moderate percentage of health service coverage services in this microregion may be related to the large distances from population settlements and the difficulties in accessing such centers, since transportation most often occurs via rivers, channels, and creeks. This fact suggests that an epidemiological silence may be occurring in those areas due to difficulties in implementing public health policies such as epidemiological and environmental surveillance for diseases [26,27].

This study differs from others due to its exploratory characteristics in order to estimate the risk of CL, considering the occurrence of the disease risk factors at different magnitudes in the microregions of Pará, and also to be capable of processing linguistic and semantic information. In this way, the system was made suitable for processing information that occurs at different scales. This fact is not observed in the techniques based on logistic regression [28].

Thus, the fuzzy system proved to be satisfactory for processing information related to the occurrence of environmental, socioeconomic, and epidemiological conditions for CL at different scales and to identify the microregions with the highest and lowest risk for the establishment of the disease. This was especially the case when the system was used in conjunction with Moran’s spatial statistics technique, which showed the different spatial dependence relationships between the variables studied in the municipalities that form the microregions of Altamira and Breves.

The spatial distribution of the disease conditionings and its geographic extension constituted limitations to more precise analyses, especially at local scales. This problem was solved by carrying out two fieldworks to observe the relationships between the variables that were processed by the fuzzy system in the laboratory. To improve the system, different membership functions e.g., trapezoidal or triangular can be used, thus including a greater number of functions in the processing of input variables, which would imply the need for the development of more rules.

Conclusions
The fuzzy system proved to be satisfactory for analyzing the different relationships between CL and its
epidemiological, environmental and socioeconomic conditionings in the microregions of the state of Pará, identifying the ones with the highest and lowest risk of disease establishment (RCLE). The joint use of the system with Moran's spatial statistic was effective in explaining vulnerabilities related to land use and cover, such as the gradients of deforestation in the studied areas. However, the computational effort of the developed fuzzy system can be improved with the inclusion of a greater number of input variables and rules seeking to interrelate them more completely. The system can be used in decision-making processes related to environmental and epidemiological surveillance, in order to intensify them, in addition to contributing to the establishment of an epidemiological memory of the environmental production of the parasitic diseases in the Amazon.

Acknowledgements

To the Laboratory of Epidemiology and Geoprocessing of Amazon (EPIGEO), from Pará State University (UEPA).

References


Corresponding author
Nelson Veiga Gonçalves, PhD
Laboratory of Epidemiology and Geoprocessing, Department of Community Health,
Center for Biological and Health Sciences, Pará State University,
Belém, Pará, Brazil / Cyberspace Institute,
Federal Rural University of Amazon, Belém,
Pará, Brazil. Tv. Perebebuí, no. 2623 - Marco, Belém -PA,
CEP: 66087-670
Tel: +55 (91) 3131-1759
Email: nelsoncg2009@gmail.com

Conflict of interests: No conflict of interests is declared.