

Digital image processing techniques for the identification of the appropriate flood areas for the malaria vector procreation

Mariane Carvalho Assis



Geógrafa, Mestre em Sensoriamento Remoto (INPE). Analista Operacional do Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN). Cachoeira Paulista [SP], Brasil. <mariane@dpi.inpe.br>.

Claudia Durand Alves



Arquiteta Urbanista, Mestre em Sensoriamento Remoto (INPE). Bolsista PCI-CNPq no Instituto Nacional de Pesquisas Espaciais (INPE). São José dos Campos [SP], Brasil. <durand@dsr.inpe.br>.

Vassiliki Terezinha Galvão Boulomytis



Engenheira Civil, Mestre em Engenharia Civil, Doutoranda em Sensoriamento Remoto (INPE). Docente do Instituto Federal de Ciência e Tecnologia de São Paulo (IFSP). São José dos Campos [SP], Brasil. <likitgb@dsr.inpe.br>.

Abstract

The present work intends to identify the areas which have appropriate environment for the procreation of the disease vector in the northern part of Labrea municipality, in the state of Amazonas. The environmental conditions found in the study region are highly favorable to the elevate density of mosquitoes and to their procreation in the larvae phase, as it is known that the ideal environment considered to its procreation is the aquatic one, where occurs around $\frac{3}{4}$ of the mosquito life cycle. The digital image processing techniques have been considered as an important tool to identify these areas and have aided the decision making in the public health segment. In order to achieve the purpose of this work, two approaches were made: the object-oriented analysis classification of an optical image and the proposal of a new methodological approach for the terrain description, resulting from the application of the algorithm called HAND. In this approach, the purpose is to characterize both the water bodies identified during the image classification and the possible flood regions, thus, favorable for the procreation of the Anopheles vector. In the spatial analysis stage, a local spatial regression model was used. As a result, it was concluded that de descriptor terrain model in the flood areas of the study region is an explanatory factor of the malaria incidence in the northern part of Labrea.

Keywords

Digital image processing techniques; malaria; HAND algorithm.

Técnicas de processamento digital de imagens para identificação de áreas alagáveis apropriadas à procriação do vetor da malária

Resumo

O presente trabalho pretende identificar áreas que possuem ambiente propício a procriação do vetor da doença no norte do município de Lábrea, estado do Amazonas. As condições ambientais encontradas na região de estudo são altamente favoráveis à elevada densidade de mosquitos e à procriação dos mesmos em sua fase larvária, pois se sabe que o ambiente considerado ideal para sua procriação é o aquático, onde ocorre cerca de $\frac{3}{4}$ do ciclo de vida do mosquito. As técnicas de processamento digital de imagens vêm sendo consideradas uma importante ferramenta para identificação dessas áreas e vêm auxiliando na tomada de decisão na área de saúde pública. Para atingir o objetivo aqui proposto foram comparadas duas abordagens: a classificação por meio da análise orientada a objeto de imagem óptica e a proposta de uma nova abordagem metodológica para descrição do terreno, resultado da aplicação de um algoritmo denominado HAND. Nesta, se busca caracterizar, além dos corpos d'água identificados pela classificação da imagem, regiões passíveis de inundação e, portanto favoráveis a procriação do vetor anofelino. Na etapa da análise espacial utilizou-se um modelo de regressão espacial local. Como resultado obteve-se que o modelo

descritor do terreno nas áreas de várzea da região de estudo é um fator explicativo para a incidência de malária no norte de Lábrea.

Palavras-chave

Técnicas de processamento digital de imagens; malária; algoritmo HAND.

1. Introduction

Among the three large epidemics of the 21st century are malaria, as well as AIDS and tuberculosis. According to Marques (1994), malaria assumes from the 80's a focal characteristic of transmission in the Amazonic region, establishing a serious public health problem in the region. Therefore, innumerable studies present the region as an interesting area of study. Consequently, the present work shall investigate the dynamics of malaria in the northern part of Labrea municipality, state of Amazonas.

Malaria is characterized for being a debilitating disease due to the symptoms of anemia that the individual presents, as the parasite protozoarium of the Plasmodium species uses and destroys massively the red globules in order to reproduce. Besides the parasite protozoarium, the transmission cycle of malaria is also composed by the Anopheles vector, or else, the mosquito of the Anopheles species, which is the human host of the disease. As the relationship between the vector and man is highly complex and strongly influenced by the social-environmental conditions, it is important to know the environment where the disease occurs, in order to understand better this dynamics and contribute to the action towards the health condition of the population who lives in the area by the public policies.

Regarding the environmental issue, the Amazonic region is in general highlighted by the influence of the humid rainforest and by the considerable presence of rivers and lakes, which is the ideal and favorable environment for the disease. It is known that the Anopheles vector breeding is represented by collections of clean water, with a certain depth, in the shade, compounded by floating or emergent aquatic plants, and poor in salt minerals and organic material (FORATTINI, 2002). The Anopheles species, common in the Amazonic region, is called Anopheles darlingi. It is considered typically fluvial, with about ¾ of the mosquito life cycle in the aquatic environment. Hence, the presence of water bodies is an important environmental indicator to comprehend the dynamics of the disease in the region, defining potential environments for the procreation of the malaria vector.

In order to comprehend and characterize these environments, digital image processing methodologies using images from orbital remote sensors have been widely used, mainly through the image classification of optical sensors. Although these methodologies offer relevant contributions for the determination of the appropriate environment for the procreation of the disease vector, they present limitations.

The means of acquiring images is determinant, as it is not possible to obtain them during the rain period due to the presence of clouds. The characteristics of the operated sensor are also another limitation, given that information cannot be acquired in the terrain under the crown of trees. Thus, it is very common to obtain information only about the main rivers, as they are permanent and visible in the dry season, and, wide enough so that they are not completely covered by the surrounding vegetation, making it possible to identify these targets in the optical images.

In this context, it is proposed in this work to present a new approach in order to add an important contribution towards the determination of appropriate environments for the procreation of the *Anopheles* vector.

The algorithm called HAND (Height Above the Nearest Drainage) (RENNÓ *et al.*, 2008) was exploited, which was capable of generating as output data the terrain areas with conditions for the existence of water, or else, possible flood areas.

The use of the proposed methodology in addition to the optical image classification may be considered a methodological innovation for characterizing the environment of *Anopheles darlingi* at larval stage.

2. Material and Methods

2.1. Study area

Labrea is a municipality, located in the north of the state of Amazonas. The studied area corresponds to the northern portion of Labrea municipality (Figure 1), where its environment is well preserved.

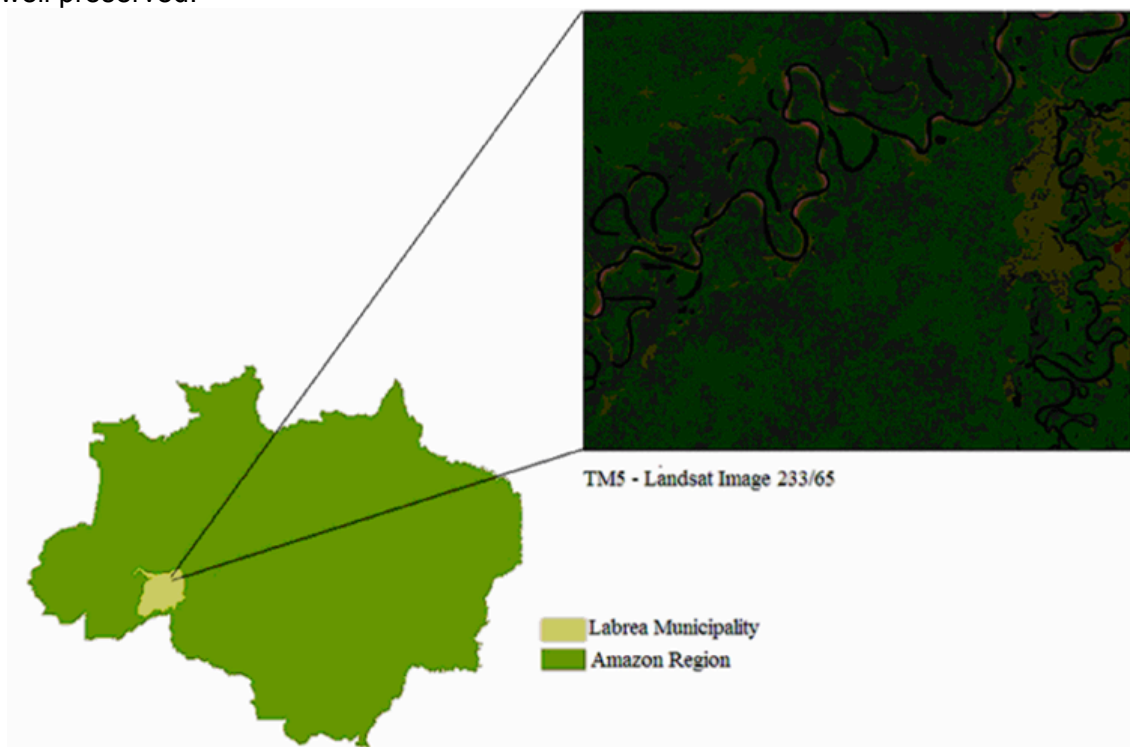


Figure 1. Location of the study area.

The interest in the area is to understand better the relationship between the *Anopheles darlingi* habitat and the incidence of malaria, prior to the occurrence of severe changes in the environment and interference in the transmission dynamics of the disease.

2.2. Environmental data

In order to obtain the identification of the potential environment for the procreation of malaria vector in the area of study, characterized primarily by the presence of water bodies or areas

liable to flood, environmental data were generated, which are consisted of the results obtained from the optical image classification and the HAND algorithm.

2.2.1. Optical image classification

The TM5-Landsat image used in the work was selected from an image data bank provided by the Global Land Cover Facility – GLCF at the National Aeronautics and Space Administration – NASA Landsat Program (2009). The bounding area of study corresponds to 200,000 hectares (50km X 40km), which is part of the image achieved on July 11, 2005, orbit/point 233/65.

During the image pre-processing stage, a process of restoration was held in the bands 2, 3, 4, 5 and 7, in order to improve the spatial resolution from 30m to 15m, and reduce the degradation inherent to the system while the image was been achieved. This procedure was done in the free software SPRING, version 4.3.3 (CÂMARA *et al.*, 1996), released by the National Institute of Space Research (INPE).

Continuing the image digital processing, the classification phase was initialized by the use of the software DEFINIENS, version 7.0. This kind of classification is based in the objects or segments of the image. In the image, an object represents an identity that may be individualized by its attributes and properties from the class it was originated by (DEFINIENS, 2006). In this process, the segmentation was done and, afterwards, the image classification. The image segmentation is divided by homogeneous and contiguous objects and different layers are created having the objects in different scales, according to size, shape and color criteria, related to each other.

Different from the traditional procedures, other attributes may be considered in this one, besides the spectral characteristics from the objects, such as the topological characteristics, texture, shape and size.

Initially a legend of the land cover was defined from the visual analysis and knowledge of the area of study, which includes the following classes: water bodies, forest, bare soil and sand.

The TM images were segmented in a first level (MRS5), using the parameters of scale (5), shape (0.1) and color (1.0), defined based on the spatial characteristics of the data and target to be identified and in the work developed by Alves *et al.* (2009b). The MRS5 segmentation was considered appropriate essentially because of making possible the identification of low-dimension water bodies, disconnected from the main rivers, even though it had generated a large number of segments. In order to solve this problem, it was used another kind of segmentor, available at the DEFINIENS software, called Spectral Difference Segmentation.

In this sense, the spectral difference segmentation was applied on the MRS5, already carried out for the assortment of existing segments from a scale parameter defined empirically. As a measure of scale, the value 10 was selected, that permitted the gathering of contiguous segments belonging to water bodies and vegetation classes, as shown in Figure 2.

The classification was held in the SDS10 segmentation level, in order to obtain the previously defined classes for the legend. Training samples, for the determination of the used attributes during the classification, were collected. The selection of attributes, that compound the class descriptors, is done based on the knowledge of the interpreter according to the analyzed data and the object of study in its various aspects: spectral, spatial, dimensional, textural,

contextual, and relational, among others. The most relevant attributes were spatialized, based on the previous knowledge of the interpreters, with the purpose of verifying the influence of each attribute in the sample discrimination. The spatialization of each attribute value like an image (Figure 3) allows the visualization and evaluation of its behavior in the group of samples, making the process of attribute selection easier.

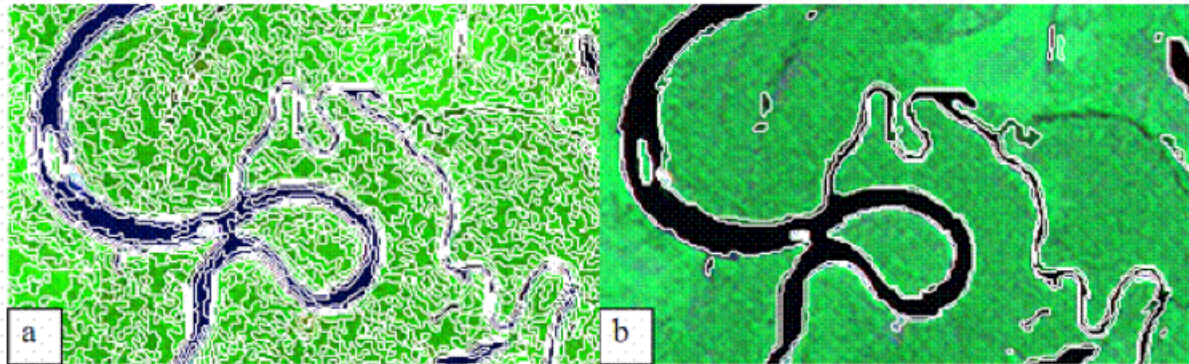


Figure 2. (a) MRS5 and (b) SDS10 segmentation in TM5-Landsat images.

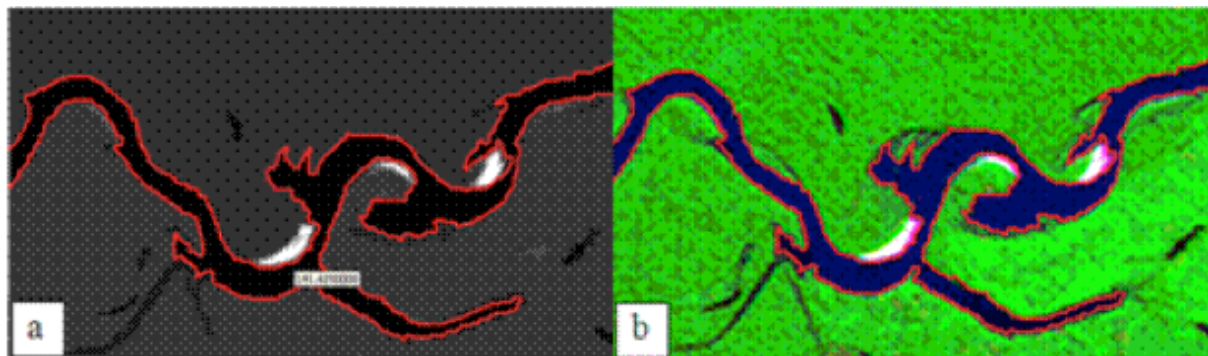


Figure 3. (a) Spatialization of the *brightness* attribute and its value for the sand (141.425) and (b) the TM5-Landsat image.

In Figure 3 (a), the level of gray may be observed in a nearly-white shade that was applied to the sand for the brightness attribute contrasting from the other targets of the image, vegetation and water, which may be observed in the corresponding TM image (in Figure 3 b). The attributes presented in Table 1 were selected for a better separability among the sample objects that represent the classes of interest.

Table 1. Attributes selected for the classes of interest.

| Classes | Attributes | Type of Attribute |
|--------------|------------|-------------------|
| water bodies | NDWI | spectral |
| sand | brightness | spectral |
| bare soil | not-forest | relational |
| forest | NDVI | spectral |

The selected attribute for water bodies was based on the methodology described at the INPE/ANEEL report discussed by Feeters (1996) in which the TM5-LANDSAT images and the index called Normalized Difference Water Index – NDWI) were used. The NDWI is defined by the Equation 1:

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad [1]$$

where: NIR - Band sensible to the infrared radiation (band 4)
 Green - Band sensible to the green radiation (band 2)

After the attribute selection, intervals of each attribute value were established. Thus, graphs were generated in order to determine the corresponding function for the inclusion of objects in each class. This procedure is done based on the values found in the sample polygons selected for each class of interest. From the generated histograms for the selected attributes, pertinent functions based on the fuzzy logic were established which define the inclusion of an object in a determined class based on a pertinence level set by imprecise limits.

As a product of the classification process, a thematic map was obtained with the previously defined classes. This product was exported from the DEFINIENS and inserted in a databank created in the SPRING (CÂMARA *et al.*, 1996).

In the final stage of the TM5-landsat image classification process, it was done the matrix edition of the map obtained from the automatic classification at DEFINIENS in order to reduce the product errors, representing the hybrid aspect of the classification, according to what was proposed by Durand *et al.* (2007), Alves *et al.* (2009a and 2009b). A qualitative analyses of the classification was carried out from the comparison of the map with the political map of Amazonas (IBGE, 2009), composed by the main hydrographic basin of the studied area.

2.2.2. Terrain descriptor

The HAND algorithm (RENNÓ *et al.*, 2008) might be considered a new quantitative terrain descriptor based on the data of the Shuttle Radar Topographic Mission (SRTM) through the digital model of terrain elevation (MDE), constructed from the achievement of information of SAR radar images of band C. More information about the grid generation and data manipulation may be found at Rennó (2004). At the end of the process, a monochromatic image is obtained as a result, varying the gray levels according to the vertical distance of the terrain regarding the drainage network.

Concerning this image it is possible to define thresholds according to the terrain description and to apply the slicing technique to facilitate the comprehension of the result. Five classes were obtained corresponding to the slicing procedure of the gray levels.

2.3. Epidemiological data

Malaria, according to the official demands settled by the State Health Department (Secretaria de Vigilância em Saúde do Ministério da Saúde), is an endemic disease in the list of compulsory notification, which must be notified to the responsible public health authority. When a screening exam is carried out, an investigation file is filled; all the information is recorded in a databank and sent to the authorities. Therefore, it is possible to obtain an epidemiological summary and information about the place where the disease was possibly transmitted. The database is released by SIVEP (Sistema de Informação de Vigilância Epidemiológica do Ministério da Saúde) which is responsible for the notification of all the malaria cases.

The essential data extracted from the summaries for the production of this work is the Annual Parasite Incidence (API) of malaria. API is calculated according to the number of annual positive cases divided by the number of local inhabitants and multiplied by one thousand. It corresponds to a risk measure of exposition to malaria defined by the responsible authorities

in order to improve the public policies towards the definition of the disease control and awareness.

2.4. Spatial data analysis

Due to the organization and different resolution of the environmental datum, compound by the terrain description variable (HAND) and epidemiological one, through the malaria incidence information (API), it was chosen to work in a cellular space compound by [2x2]km cells. This resolution was defined according to the understanding of the studied phenomenon, as it is related to the flight scope of the mosquito in regular conditions. Thus, after the creation of cellular plan, it is possible to fill in the cells with these variables: HAND e API. The procedures were done by the use of the free software TERRAVIEW, version 3.3.1, released by INPE, through the plug-in filling of the cells.

From the creation and promotion of the databank of the studied area, it was possible to initialize the analysis, searching for answers about the existence of a relationship between the descriptor terrain variable and the incidence of malaria in the studied area. In other words, it was possible to investigate the environments defined as favorable for the procreation of malaria vector, if they actually present a positive association with the incidence of the disease. It was preferable to use a regression model that incorporates the local spatial effects, such as the Geographically Weighted Regression – GWR (FOOTHERINGHAM *et al.*, 2002). For the implementation of the GWR, the free software R, version 2.9.2 (R DEVELOPMENT CORE TEAM, 2010).

First the bandwidth, which is essential to the local regression analysis, was calculated, simultaneously applying the cross-validation – CV technique in this stage. This technique is a preview and estimative evaluation method, through the exclusion of an observation each time, predicting the values based on the remaining data, and comparing the results with the observed values.

After the definition of this bandwidth, it is possible to apply the GWR equation, with the incidence of malaria as a response variable and the environmental data of the terrain descriptor as an explicative variable of the phenomenon. The results might be visualized and mapped by the use of TERRAVIEW.

3. Results and Discussion

From the used criteria for the segmentation and classification of the optical image a thematic map was obtained as a result according to Figure 4.

Four classes were obtained: water bodies in blue color, which characterize the identification of the main rivers and the streams; sand, that corresponds to the sediment deposition areas along the rivers; bare soil, which represents the areas with human settlement or, as in the eastern side of the Figure, periodic flood areas by the rivers; and forest, illustrating the largest area of study, characterized by the preservation of its vegetation.

According to the purpose of the work to identify the potential areas for the procreation of malaria vector, it may be said that, based on the image classification, the environment considered more appropriate is the class identified with the presence of water bodies. However, as it was said before, these images are limited for the determination of these environments, mainly regarding the small water bodies and flood areas in the dry periods.

For this reason, the information obtained to identify the ideal environment for the *Anopheles darlingi* was complemented by a new approach, applying the HAND algorithm. The map achieved in this procedure (Figure 5) has more information added to the description of the terrain.

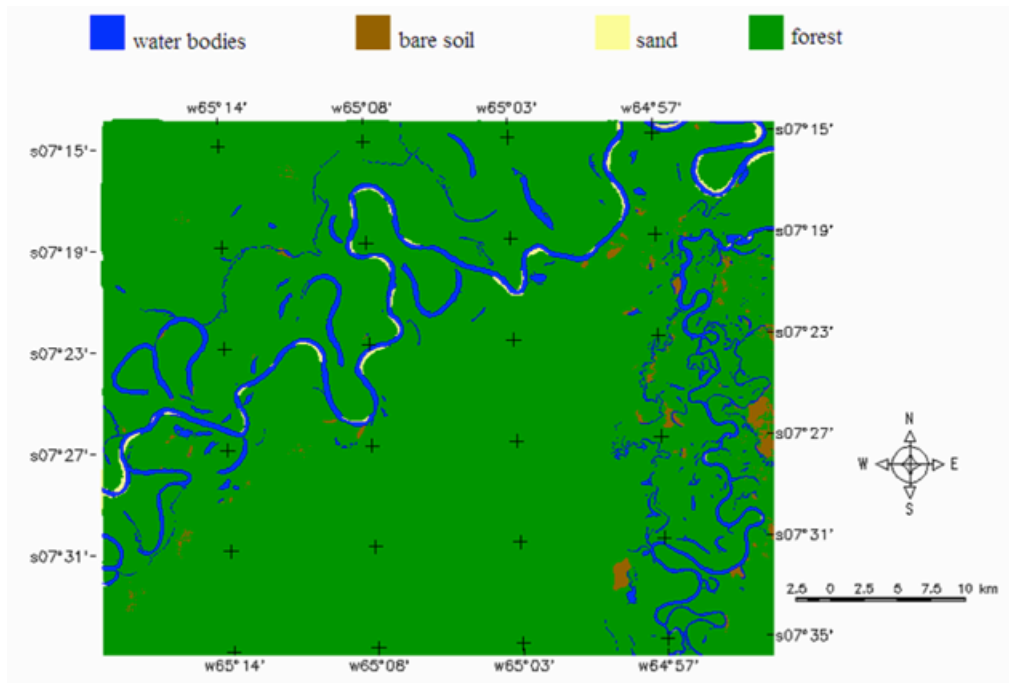


Figure 4. TM5-Landsat image classification.

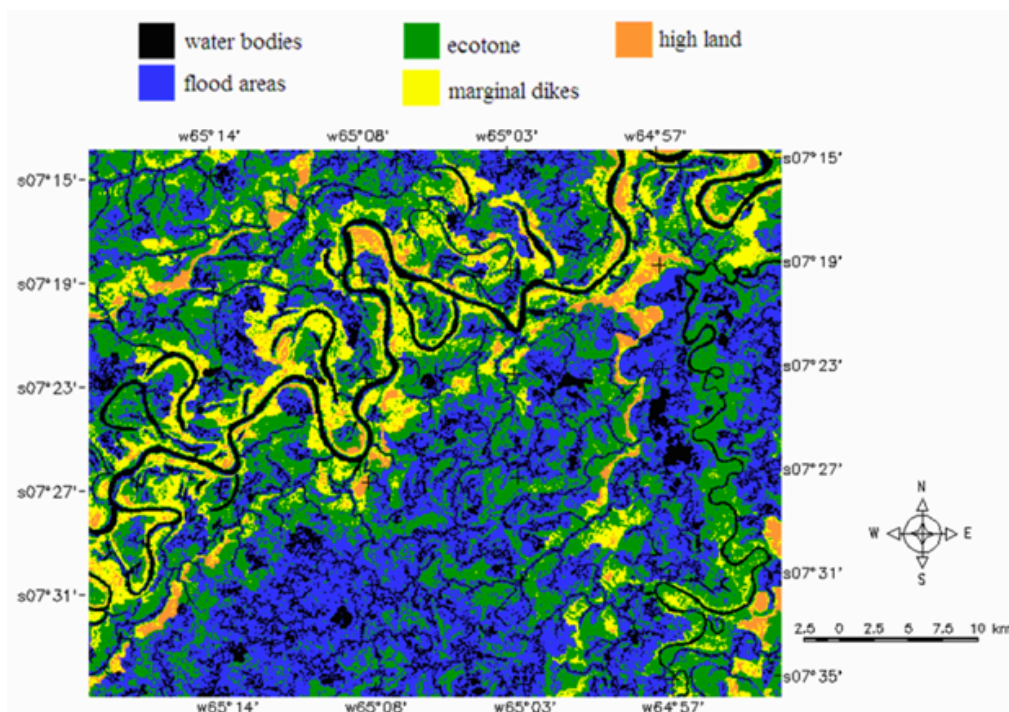


Figure 5. Descriptive terrain map regarding the water bodies.

In Figure 5, it is possible to observe the delimitation of the appropriate terrain areas for the procreation of the malaria vector, such as: water bodies (black color) and flood areas (blue

color). These flood areas correspond to a more favorable environment for the procreation of Anopheles, as they may have collections of clean and still water, ideal for its habitat in the larvae phase. The water bodies have the same characteristic for the mosquito environment, although it tends to be more turbulent due to its own nature and also to the influence of the population who uses it for transportation, mainly with large boats, that can make it more difficult for the procreation of the mosquito in it.

The green areas are called ecotone or transition zone, defined as a humid region with underground water near the land surface, but not necessarily flood one. The marginal dikes, in yellow, are on higher places than the transition zone and they are the places where thinner sediments concentrate, and even large ones after raining. Finally, the high land, in orange, is the area where it will be very difficult for the occurrence of flood.

To effect the investigation about the relationship between the terrain description, result from the HAND algorithm, and the incidence of malaria in the northern part of Labrea, it is necessary to apply an investigation method among the variables regarding the incidence of malaria that analyzes its spatial relationship, which would not happen by the use of a linear regression model. Thus, it is more effective to use a spatial regression method. In the work it was chosen to use the GWR (Geographically Weighted Regression), as it has the advantage of generating results with two kinds of coefficients: a global coefficient from the association between the explaining variable and the response variable; and another local coefficient in each cell of the studied area (Table 2), allowing the visualization of the coefficient variation.

Table 2. Global coefficient as a result from the GWR.

| Covariable | Global Coefficient |
|------------|--------------------|
| HAND | 63.5266 |

From the positive value of the global coefficient, 63.5266, product from the application of the GWR model, it is possible to affirm that the HAND covariable has a positive relationship with the response variable of the studied phenomenon. Hence, it is considered an explanatory factor of the incidence of malaria in the studied area.

However the most interesting result that the GWR regression model may present is how the local coefficient values vary spatially, in this case, for each cell of [2x2]km as it was said before (Figure 6).

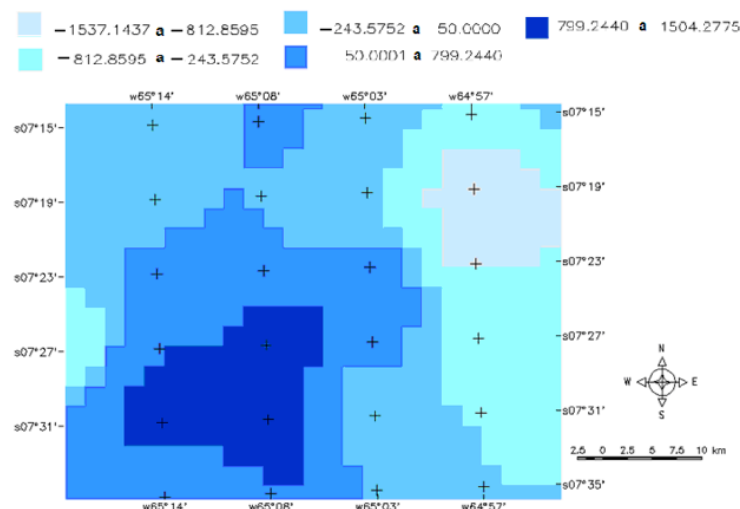


Figure 6. Spatialized result from the coefficients for the HAND covariable.

The area that has a higher positive coefficient represented by the darker blue color matches the limits of the flood area. In other words, in this area the HAND covariable is positively explanatory for the response variable of the incidence of malaria. Thus, these flood areas are determining for the occurrence of the disease. At last, the proximity between men and the vector in the larvae phase, illustrated before in flood areas, is a risk factor related to the incidence of malaria in the northern part of Labrea.

4. Conclusions

The environmental data produced from the application of digital processing techniques constitute an important tool for the identification of appropriate areas for the procreation of the malaria vector.

The descriptor terrain model, resulted from the application of the HAND algorithm, is one important methodological improvement for the identification of these environments, complementing the generated data exclusively from the optical images.

The spatial data analysis from the use of a local spatial regression model indicated that this approach may facilitate the prevention and control of disease, since it is possible to identify the ideal environment for breeding the *Anopheles darlingi* and act directly to control the proliferation and spread of the vector in these areas.

5. References

- ALVES, C. D.; PEREIRA, M. N.; FLORENZANO, T. G. Mapeamento das novas formas de ocupação urbana por meio da análise orientada a objeto. Estudo de caso: aglomeração urbana de Piracicaba. In: Simpósio Brasileiro de Sensoriamento Remoto, 14., Natal. **Anais...** São José dos Campos: INPE, 2009a. p. 569-576. CD-ROM.
- ALVES, C.D.; PEREIRA, M.N.; FLORENZANO, T. G.; SOUZA, I.M. Análise orientada a objeto no mapeamento de áreas urbanas com imagens LANDSAT. **Boletim de Ciências Geodésicas**, Curitiba, v.15, n. 1, p. 120-141, 2009b.
- CÂMARA, G.; SOUZA, R.C.M.; FREITAS, U.M.; GARRIDO, J.; LI, F. M.; SPRING: Integrating Remote Sensing and GIS by Object-Oriented Data Modelling. **Computer and Graphics**, v.20, n. 3, p. 395-403, 1996.
- DEFINIENS. **eCognition: User Guide 5**. 2005. Disponível em: <http://www.definiens.com/services/faq/Definiens_Professional5_InstallationGuide.pdf>. Acesso em dez. 2009.
- DURAND, C.; SOUZA, I. M., PEREIRA, M.N. Análise Orientada a Objeto para a Extração Automática de Áreas Urbanizadas. In: Simpósio Brasileiro de Sensoriamento Remoto, 13., Florianópolis. **Anais...** São José dos Campos: INPE, 2007. p. 5187-5189. CD-ROM.
- FEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International Journal of Remote Sensing**, v. 17, n.7, p.1425-1432, 1996.
- FORATTINI, O. P. **Culicidologia Médica: identificação, biologia e epidemiologia**. São Paulo: Edusp, 2002, v.2, 860 p.
- FOTHERINGHAM, A.S.; BRUNSDON, C.; CHARLTON, M.E. **Geographically Weighted Regression: The Analysis of Spatially Varying Relationships**, Chichester: Wiley, 2002, 269p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Mapeamento Geográfico – Produtos: Mapas Estaduais e Regionais**. Disponível em: <http://geoftp.ibge.gov.br/mapas/tematicos/politico/AM_Politico.pdf>. Acesso em dez. 2009.

MARQUES, A. C.; GUTIERRES H. C. Combate à malária no Brasil: evolução, situação atual e perspectiva. **Revista da Sociedade Brasileira de Medicina Tropical**, v.27, p. 91-108, 1994.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA) LANDSAT PROGRAM. **Landsat TM5**, scene orbit/point 233/65, GeoCover, USGS, Sioux Falls, jul. 2005. Disponível em: <<http://glcf.umiacs.umd.edu/data/landsat/>>. Acesso em: dez. 2009.

R DEVELOPMENT CORE TEAM. **SPGWR package**. Disponível em: <www.r-project.org>. Acesso em: jan. 2010.

RENNÓ, C. D. **Construção de um sistema de análise e simulação hidrológica: aplicação a bacias hidrográficas**, 2003, 158 p. Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2004.

RENNÓ, C. D.; NOBRE, A. D.; CUARTAS, L. A.; SOARES, J. V.; HODNETT, M. G.; TOMASELLA, J.; WATERLOO, M. J. HAND, A new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. **Remote Sensing of Environment**, v. 112, p. 3469–3481, 2008.

SECRETARIA DE VIGILÂNCIA EM SAÚDE, MINISTÉRIO DA SAÚDE (SVS). **Situação epidemiológica da malária no Brasil de 2007**. Disponível em: <http://portal.saude.gov.br/portal/arquivos/pdf/folder_malaria_2008_final.pdf>. Acesso em jan. 2010.