The Use of Geospatial Technology for Epidemiological Chagas Analysis in Bolivia

Natalia I. Vargas-Cuentas[1], Alicia Alva Mantari[2], Avid Roman-Gonzalez[3]
Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, Lima, Peru

***Abstract*—Chagas disease is caused by the parasite Trypanosoma Cruzi and transmitted by the Vinchuca. Bolivia is the country with the highest prevalence in the South American region; for example, in 2015, there was a prevalence of 33.4%. This disease causes severe intestinal and cardiac problems in the long term, 30% of the cases register cardiac symptoms, and 10% have alterations in the esophagus or colon. This research aims to analyze the relationship between environmental factors and Chagas outbreaks in an area of Bolivia to identify the environmental conditions in which the disease is transmitted, using epidemiological, meteorological data and also environmental indexes extracted from Landsat 8 satellite images. Through a Principal Components Analysis (PCA) of the environmental indexes extracted from the satellite images and the meteorological information, has been found that the environmental conditions that have a correlation with the occurrence of cases are: temperature, relative humidity, visibility, Normalized Difference Soil Index (NDSI) and Modified Normalized Difference Water Index (MNDWI).**

***Keywords*—Trypanosoma Cruzi; Vinchuca; Landsat 8; PCA; Normalized Difference Soil Index (NDSI); Modified Normalized Difference Water Index (MNDWI)**

I. Introduction

In the world, the environmental characteristics and climatic changes of the different ecosystems have a decisive influence on some diseases that affect man, especially those associated with vectors such as Chagas.

Chagas disease is one of the most worrisome vector diseases in Latin America; the World Health Organization (WHO) declares it as one of the most critical public health problems in America.

According to [1] since 1990, significant success has been achieved in the control of the parasite and the vector in Latin America. [2] In the last 20 years, different Initiatives achieved a reduction in the transmission by domestic vectors as well as blood transfusions. The burden of Chagas disease has been reduced significantly (from around 30 million to 7 -8 million). [3] Chagas disease is caused by the flagellated protozoan, Trypanosoma Cruzi, which is transmitted to humans and other mammals mainly through the droppings of blood-sucking hemipteran insects on their host immediately after feeding.

The transmission of the disease to humans is due to the ability of the insect to explode and adapt to environments in different areas. The distribution of the disease is not uniform. Previous studies like [4] show environmental differences in the situation of the problem of this disease. For the planning of protection measures, it is essential to know how this disease is distributed at the provincial level. Also, it is necessary to identify the risk that each sector has, in order to assess the needs for control, care and planning according to local needs.

The conception of the study of space as a new perspective for epidemiological analysis in order to identify possible epidemiological outbreaks and the development of the disease as a global process in a population space is called panoramic epidemiology. In [5] for example, the reality of Argentina is described regarding the occurrence of dengue outbreaks, in order to carry out epidemiological surveillance for the control of the disease-causing vector. In work an exploration of the problem is carried out from the discipline of panoramic epidemiology to analyze the distribution, ecology, and behavior of the vector. Besides, an exploration of techniques in classification and image processing is developed, in order to generate a temporal space dispersion model of the vectors. Predictive maps were developed, of the focal density of Aedes aegypti, based on environmental information derived from SPOT 5 HRG1 high-resolution spatial images and images of average spatial resolution of surface temperature derived from Landsat 5 satellite information. A predictive model of biweekly aedic indices was generated, which was based on macro-environmental information from Landsat TM and ETM sensors, and vector monitoring and control information. Then, space-time epidemiological patterns and population parameters of the vector were estimated; the effectiveness of vector control measures during the outbreak was also estimated.

Also in [6] as part of the investigation of CONAE (National Commission for Space Activities) and the Ministry of Health of Argentina, a project was developed to implement a Dengue Early Warning computer system, which calculates Dengue's environmental risk in Argentina. The risk is assessed based on a static component related to historical environmental conditions and a dynamic component related to current environmental conditions. The stratification of the localities of Argentina is shown as a result according to their environmental risk of Dengue.

The study published by Neteler [7] analyzes the environmental conditions of the spread of the Asian tiger mosquito Aedes Albopictus in Europe, as part of a constant concern for public health due to outbreaks transmitted by vectors. Data from the reconstructed daily satellite time series (MODIS Land Surface Temperature maps, LST) of northeastern Italy were used, as well as the annual average temperature by areas, in order to reconstruct areas and compare them with the current known distribution of Ae. Albopictus in northeastern Italy. LST maps show peculiar microclimatic characteristics. From this data, surface maps are reconstructed

to predict the infection areas of the vector with an accuracy of 200m pixels. This is an important study that could be applied to other arthropod species in which the temperature is a relevant factor.

According to [8], the dispersal capacity of a disease that is transmitted by vectors depends on multiple environmental, climatological, biophysical and social variables [1]. The information extracted from the satellite images will allow us to observe the climatic changes related to epidemiological outbreaks indirectly [8]. With satellite images of different spectral bands, different environmental parameters can be identified and measured [2].

Images from the Landsat 8 satellite and the OLI-TRIS sensors will be used. With this analysis, we try to describe if there are tools to generate models that describe the propagation habitat of vectors that are the transmitters of endemic diseases.

It is from these images that for certain period of time we analyze the indices that are extracted from them to generate a model that gives us a simplified representation of reality. This model allows finding variables to understand the values of the incidence and prevalence, as well as the process of transmission of the disease in a certain period.

The importance of this work is based on the fact that space is a privileged place. For this reason, satellite images can give us an overview of the area that one wishes to analyze. The idea is to have global information about areas with a possibility of epidemiological outbreaks. All this information can contribute to better decision making when planning prevention tasks and epidemiological management.

The limitations of the present study are that the temporal resolution of the Landast 8 satellite is 16 days, which implies having only one image to analyze per month, in addition in the study area there are not many weather stations to complement the data obtained from the satellite images.

The content of this work is broadly divided into three main parts: in the first part the Chagas disease is described. Afterwards a spatial temporal epidemiology analysis is developed. Finally an analysis of the relationship between all the data extracted is implemented.

II. Description of Chagas Disease

This section seeks to expose the basic aspects of Chagas disease, besides shows the growing presence of Chagas in South America and mainly in Bolivia in recent years.

A. Definition and basic Aspects of the Disease

The disease of American trypanosomiasis or Chagas, it is a type of zoonosis, a disease typical of the encounter with animals of contact with people, typical of this continent, because has existed in Latin America since before the conquest.

In 2018 WHO reported 6 or 7 million people infected in 21 countries by the parasite Trypanosoma Cruzi [3], which causes this disease and 90 million people are at risk of contracting the infection.

Chagas can be transmitted through the consumption of contaminated food, stinging of the infected vector, blood transfusion, transmission of the pregnant mother to the child, transplantation of infected organs.

The process of the disease has an incubation period of 4 to 10 days, mostly asymptomatic. Then the three phases are distinguished, acute, indeterminate and chronic.

The acute phase lasts between 2 to 4 months, is asymptomatic, so the diagnosis is difficult at this stage, it can also present very mild symptoms. It is characterized by the high concentration of parasites in blood. Some of the symptoms presented on occasion are: fever, headache, swollen lymph nodes, slight pallor, muscle pain, breathing with difficulty, abdominal pain.

The indeterminate phase begins after the acute phase and lasts for 8-10 weeks, regardless of the symptomatology. During this phase, the disease is usually asymptomatic. Although it still has quantities of the parasite in the blood, it is only possible to diagnose it in 20% to 60% of the cases. It is during this phase that the contagion becomes a great problem of public health, due to the ignorance of the presence of the disease.

During the chronic phase the parasite usually hides in the cardiac and digestive organs, where they multiply and begin to generate irreversible tissue damage. Only 30% of cases have any cardiac symptoms, and about 10% have alterations in the esophagus or colon. Moreover during this stage the amount of parasitaemia is low. The symptoms will depend on the damage of the parasite in the host organ. Chronic myocarditis is the most common heart affection due to Chagas disease.

Among the symptoms that derive from the presence of the parasite in the heart, are myocardial damage, arrhythmia and heart failure.

Ventricular fibrillation is probably the most frequent mechanism of sudden death in chronic Chagasic patients.

The disease registers a percentage of mortality that oscillates between 55% and 65%, mainly due to cardiac complications.

The treatment is especially useful in the acute phase of the disease, where it can cure up to 100% of cases. The effectiveness of the treatment decreases in an inverse manner over time, because the parasite can cause irreversible damage to some organs.

This treatment is based mainly on benznidazole and nifurtimox capable of killing the parasite depending on the phase of infection.

In some chronic cases, the treatment although it may not kill the parasite completely, can prevent or delay the progression of the disease, although it is important to consider the adverse consequences of dosing them over time.

Unfortunately there is no vaccine against Chagas disease, but the prevention method is the control of the vector in the areas of greatest incidence.

The T. Cruzi can infect several species, depending on the geography of the area, for this reason it is recommended for its prevention the fumigation of areas of greater risk, cleaning and improvement in the conditioning of homes, hygiene, adequate conservation of the food and develop regular serological tests in risk areas.

B. Presence of Chagas in Latin America

Chagas is a parasitic disease that has become a public health problem in Latin America [9], because it has a presence in at least 21 countries in the region.

In America the infection is located from the south of the United States to Argentina and Chile. According to [10] due to this disease in this continent about 50,000 people die each year and more than 100 million people are at risk of becoming infected.

In 2013 according to the World Health Organization [11], it has been estimated that of the 8 million people carrying T. Cruzi, the largest number of cases are concentrated in Latin America, although important number of cases are also calculated in the United States, Canada, Spain, Japan and Australia.

Chagas disease in Latin America can be associated to multiple factors, including dwelling houses built with materials such as adobe, mud and straw, this type of structures can be seen in rural and suburban areas of the region.

There are around of 140 species of vinchucas in the world, of which the majority are distributed throughout the American continent, only a few species are present in Asia, Africa and Australia. [11].

In Latin America, together with the Pan American Health Organization (PAHO) and the World Health Organization (WHO), a horizontal technical cooperation strategy was developed among countries to prevent and control Chagas disease in the region. [3] For example, these initiatives have been developed in the Southern Cone (1992), the Andean countries (1998) and the Amazonian countries (2003), among other cooperation strategies in the region.

This cooperation has contributed to the elimination of allochthonous species of vectors, the detection of congenital cases, the reduction of prevalence in children and the improvement of the quality of treatment of infected and sick people, among others.

In the specific case of South America, the vector Triatoma Infestans predominates in Argentina, Bolivia, Brazil, Chile, Paraguay, Peru and Uruguay. [12] Besides in Colombia and Venezuela the predominant vector is R. Prolixus.

According to figures calculated by WHO in 2010, Bolivia is the country with the highest incidence of Chagas in South America.

C. Presence of Chagas in Bolivia

According to the Institute of Development Health Research (IINSAD) of the Universidad Mayor de San Andrés (UMSA) in [13] there are 140 species of vinchucas in the world, in Bolivia there are 21 types of vinchucas identified, the most common vinchuca is the Triatoma Infestans, which is responsible for the largest number of Chagas (CH) cases recorded in the country.

The country with the largest dispersion area of the CH vector (Triatoma Infestans) is Bolivia. [14] In the endemic vector map of the Ministry of Health Chagas disease is dispersed in approximately 60% of the Bolivian territory.

In the country, three endemic zones were identified: the valley area comprised by the departments of Cochabamba, Chuquisaca, Tarija and Potosí, the Chaco area comprised by the departments of Santa Cruz, Chuquisaca and Tarija, and finally the Amazon area departments of Beni, Pando, part of Santa Cruz, north of La Paz and north of Cochabamba.

According Médecins Sans Frontières (MSF) in [15] the South American region, Bolivia register more than 600,000 people infected with the disease. Besides, an average of 8,000 new cases of Chagas is registered each year.

The Chagas Prevention Program of the Ministry of Health of Bolivia indicates in [13] that the country register the highest prevalence in the South American region, for example in 2015 there was a prevalence of 33.4%.

In 2016 the Ministry of Health of Bolivia recorded 17,892 new cases, in the department of Santa Cruz, 57.72% of the total cases of Chagas were concentrated, followed only by the department of Cochabamba, which concentrated a total of 23.13% of infected persons [14].

III. Spatial Temporal Epidemiology Analysis

For this research it is necessary to select a study area of interest in Bolivia, acquire the different types of data and performs the treatment of the satellite images, this section exposes these main procedures.

A. Selection of the Study Area

Chagas has become an emerging disease in Bolivia, since it is dispersed in approximately 60% of the Bolivian territory. The CH vector (Triatoma Infestans) it has found in the warm departments of the country, such as Santa Cruz, Tarija, Chuquisaca and Cochabamba, the adequate conditions to prevail and transmit the disease.

We can also observe that Chagas disease has presence in the nine departments of Bolivia. For this reason we must select among the most affected departments of the country by this viral disease and where there is a high number of cases.

In order to select a suitable research study area the National Program of Preventive and Control of Chagas in Bolivia was revised, where the epidemiological situation of the country in 2017 can be observed:

As we can observe in Table I, there is a high number of Chagas cases in three departments: Santa Cruz, Cochabamba and Tarija. Also we can observe a moderate number of cases in three departments: Potosi, Chuquisaca and Beni. Finally there is a low number of cases in three departments: La Paz, Pando and Oruro.

TABLE. I. TOTAL CASES OF CHAGAS DISEASE IN BOLIVIA IN 2016

Departments	MALE	FEMALE	TOTAL
BENI	82	85	167
CHUQUISACA	350	483	833
COCHABAMBA	1,705	2,433	4,138
LA PAZ	79	131	210
ORURO	4	17	21
PANDO	6	5	11
POTOSI	408	448	856
SANTA CRUZ	4,334	5,994	10,328
TARIJA	545	783	1,328
TOTAL:	7,513	10,379	17,892

In the nine departments of Bolivia there is a total of 15911 cases of Chagas that have been recorded throughout the year of 2017, as can be seen in Fig. 1.

According to the collected data, the department of Santa Cruz, is the department that has registered the most high number of cases in 2017, with a total of 7777 cases, in second place is the department of Cochabamba with a total of 4794 cases and in third place is the department of Tarija with a total of 1457 cases of Chagas.

It can be observed that in eight of the Bolivian departments the number of female cases is greater. The departments of Santa Cruz and Cochabamba are those that report the highest number of female cases.

In order to correctly identify the study area, the incidence rate and point prevalence of Chagas disease in the nine departments of Bolivia were calculated in Table II.

As can be seen in Table III, the departments with the highest incidence rate and prevalence percentage are: Tarija, Santa Cruz and Cochabamba respectively. However, the department of Santa Cruz has the highest number of cases, because it has a larger population (more than three millions of inhabitants), for this reason Santa Cruz is selected as the area of analysis.

Since the department of Santa Cruz is the largest department in Bolivia, is necessary to select only one specific area to carry out the research.

The department of Santa Cruz has 15 provinces, as can be seen in Fig. 2, of which according to the Ministry of Health in Bolivia 14 provinces have the presence of Chagas in 2017, the information collected can be observed in Table III.

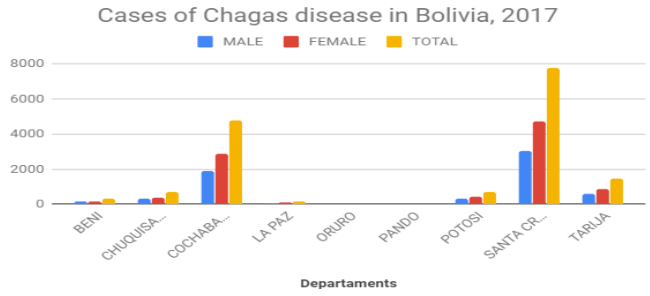

Fig. 1. Chagas Disease in the Bolivian Departments in 2017.

TABLE. II. TOTAL CASES OF CHAGAS DISEASE IN BOLIVIA IN 2017

	TOTAL		
Departments	**MALE**	**FEMALE**	**TOTAL**
BENI	153	147	300
CHUQUISACA	293	379	672
COCHABAMBA	1,915	2,879	4,794
LA PAZ	60	99	159
ORURO	5	6	11
PANDO	4	15	19
POTOSI	317	405	722
SANTA CRUZ	3,050	4,727	7,777
TARIJA	603	854	1,457

TABLE. III. INCIDENCE RATE AND POINT PREVALENCE OF CHAGAS

Departments	TOTAL	Population	Incidence rate (100,000 inhabitants)	Point prevalence (%)
BENI	300	462,081	64.924	0.065
CHUQUISACA	672	621,148	108.187	0.108
COCHABAMBA	4,794	1,943,429	246.677	0.247
LA PAZ	159	2,862,504	5.555	0.006
ORURO	11	531,890	2.068	0.002
PANDO	19	139,018	13.667	0.014
POTOSI	722	880,651	81.985	0.082
SANTA CRUZ	7,777	3,151,676	246.758	0.247
TARIJA	1,457	553,373	263.294	0.263

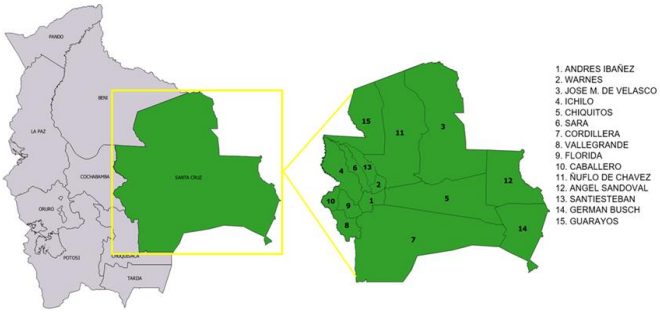

Fig. 2. Provinces of the Department of Santa Cruz.

As can be seen in Table IV, the provinces with the highest number of cases in the department of Santa Cruz are: Andres Ibañez (with 6,315 cases), Obispo Santistevan (with 811 cases) and Warnes (with 330 cases).

As can be observed in Fig. 3, in 9 of the 14 provinces affected by Chagas in Santa Cruz it can be seen a higher number of cases in the female population, except in the province of German Busch where it can be seen that the number of cases is the same in the female and male population.

Andrés Ibañez province represents 81.9% of the total number of cases registered in the department of Santa Cruz throughout 2017, followed in the second place by the province of Obispo Santistevan, which represents 11.2% of the cases and in the third place is the province Warnes which represents 2.2% of the total cases, as can be seen in Fig. 4.

TABLE. IV. CASES OF CHAGAS IN THE PROVINCES OF SANTA CRUZ IN 2017

	TOTAL		
Province	**MALE**	**FEMALE**	**TOTAL**
ANDRES IBAÑEZ	2,498	3,817	6,315
CHIQUITOS	11	46	57
CORDILLERA	14	18	32
FLORIDA	1	4	5
GERMAN BUSCH	4	4	8
GUARAYOS	13	16	29
ICHILO	14	15	29
MANUEL MARIA CABALLERO	5	2	7
ÑUFLO DE CHAVEZ	43	40	83
OBISPO SANTISTEVAN	341	470	811
SARA	31	19	50
VALLEGRANDE	5	12	17
VELASCO	3	1	4
WARNES	67	263	330

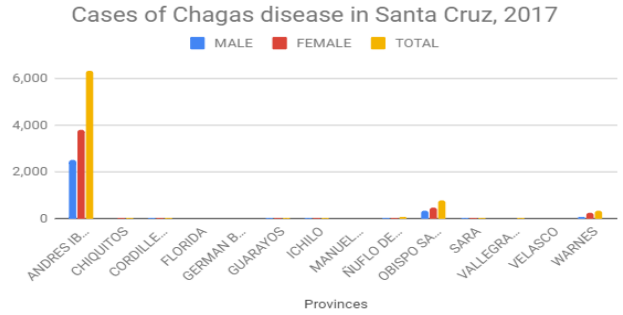

Fig. 3. Cases of Chagas Disease in the Provinces of Santa Cruz.

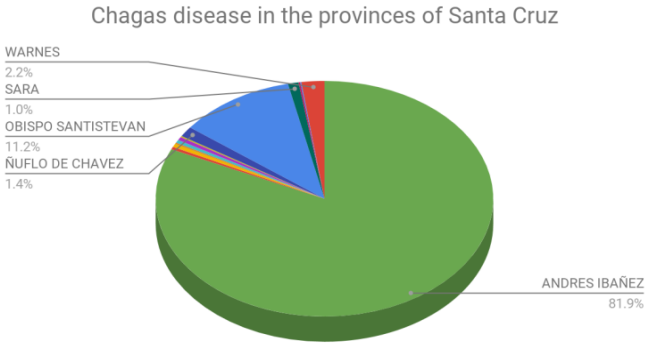

Fig. 4. Chagas Disease in the Provinces of Santa Cruz.

Thus these three provinces of Santa Cruz, representing the 95.3% of Chagas cases, are selected as the study area analyzed in the present project, as can be observed in Fig. 5.

The province Andrés Ibañez it is the most important province of the department of Santa Cruz, has five municipalities, which are: Cotoca, El Torno, La Guardia, Porongo and Santa Cruz de la Sierra. It is located at the coordinates: 17°50'00"S 63°18'00"W. Besides, the province Obispo Santistevan has five municipalities, which are: Montero, Saavedra, Mineros, Fernández Alonso and San Pedro. It is located at the coordinates: 16°30'00"S 63°30'00"W. Finally the province Warnes has two municipalities, which are: Warnes and Okinawa. It is located at the coordinates: 17°20'00"S 63°00'00"W.

B. Data Acquisition

For the present work, one needs three types of data on which the analysis, correlations, and conclusions will be obtained. These data are: satellite images, epidemiological data, and meteorological data. The description of each of these data is detailed below.

a) Satellite images: There are different satellites that provide us multispectral images, from which it is possible to collect environmental information, which in this case will be used to identify the environmental parameters that influence in Chagas outbreaks in three provinces of Santa Cruz in Bolivia.

For this research we collected information from the Landsat 8 satellite, launched on February 11, 2013, these satellite images are downloaded for free from the web: https://earthexplorer.usgs.gov/.

The Landsat 8 satellite has two sensors, OLI and TIRS. Among the most important features observed in [16], can be mentioned that it is heliosynchronous, at an orbital height of 705 km, WRS-2 (Worldwide Reference System), with an inclination of 98.2°, a temporal resolution of 16 days and having a radiometric resolution of 12 bits, spatial resolution of 30m, and spectral resolution of nine bands 5 of them in the visible field and the others in the non-visible field.

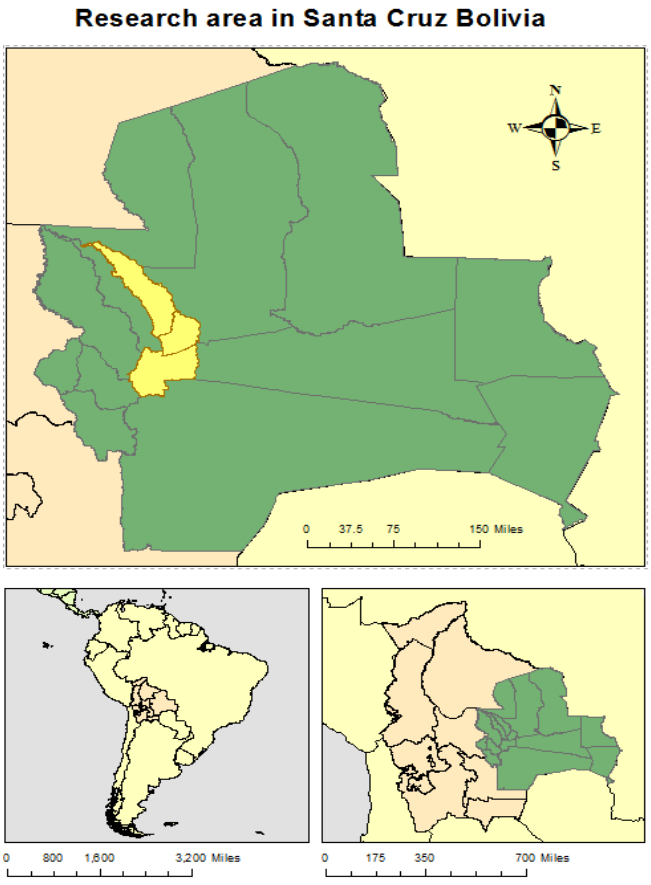

Fig. 5. Research Area in Santa Cruz Bolivia.

The images selected for this research were compiled from the USGS website. It was identified that the scene that covers 83% of the study area, is in the location Path 231 and Row 72, according to the WRS system.

The scenes of all the months of the year 2017 were analyzed, choosing only those that have a minimum cloud percentage, the details can be observed in the following Table V.

Those images registering a minimum cloudiness of less than 40% were selected. The selected scenes have an average cloudiness of 21.80%, the selected months are: February, July, August, October and November.

Of the five datasets selected, we have been downloaded: 11 raster bands (GeoTIFF), the BQA file (16 bits quality control file), and the metadata file (MTL), in total 14 files.

The spatial subset of the 11 bands of the image is: 7622x7732 pixels, with a vertical and horizontal resolution of 96pp, and a depth of 32 bits. The image corresponding to band 8, is the only one that has a resolution of 15242x15242 pixels, with a vertical, horizontal and depth values similar to those already mentioned.

b) Epidemiological data: The Ministry of Health of Bolivia, through its Departmental Health Service (SEDES), by its acronym in Spanish "Servicio Departamental de Salud", in a joint effort of the nine departments of Bolivia, has developed and implemented the National Health Information System.

This information system contains fourteen-year data from 2005 to 2018, also it contains different groups of variables, such as immuno-preventable diseases, sexually transmitted diseases and vector-borne diseases, among many others.

The national information system has been developed to maintain an epidemiological surveillance program in all the health establishments of the nine departments of Bolivia. The tool is available on the web portal of the Ministry of Health Bolivia in the informatics tools section available at: https://snis.minsalud.gob.bo/.

This informatic tool allows to perform an epidemiological surveillance of vector-borne diseases, including: Dengue, Zika, Chikungunya and Chagas disease among others.

The tool allows a search of data by departments, provinces, municipalities, networks and health establishments throughout Bolivia and in the different months of the year of interest.

For this study, the epidemiological database corresponding to the 2017 period is used, as can be seen in Table VI.

In the database collected, as can be seen in Table IV, the number of cases of acute Chagas reported in the study area is obtained, but in addition these cases are classified by gender and by age, from less than 6 months to over sixty years of age.

It can be observed in Fig. 6 that the age group with the highest number of cases of Chagas is the group of 60 years and over, with 2,560 cases. Moreover, the age group with the least presence of Chagas is the group of six months to less than one year, with 40 cases.

The database compiled also contains data on the types of health facilities where cases of Chagas were reported in 2017. It is important to mention that in Bolivia there are first, second and third level hospitals, health centers and medical posts among other smaller health facilities.

TABLE. V. SATELLITE IMAGE DATASETS OF THE STUDY AREA

Data set	Path-Row	Date acquired	% Cloud coverage
LC08_L1TP_231072_20170119_20170311_01_T1	231 - 72	19/01/2017	44.76
LC08_L1TP_231072_20170220_20170301_01_T1	231 - 72	20/02/2017	33.55
LC08_L1TP_231072_20170308_20170317_01_T1	231 - 72	08/03/2017	64.03
LC08_L1TP_231072_20170425_20170502_01_T1	231 - 72	25/04/2017	69.88
LC08_L1TP_231072_20170527_20170615_01_T1	231 - 72	27/05/2017	71.7
LC08_L1TP_231072_20170612_20170628_01_T1	231 - 72	12/06/2017	82.26
LC08_L1TP_231072_20170714_20170726_01_T1	231 - 72	14/07/2017	10.9
LC08_L1TP_231072_20170831_20170915_01_T1	231 - 72	31/08/2017	5.61
LC08_L1GT_231072_20170916_20170929_01_T2	231 - 72	16/09/2017	100
LC08_L1TP_231072_20171018_20171025_01_T1	231 - 72	18/10/2017	20.27
LC08_L1TP_231072_20171103_20171109_01_T1	231 - 72	03/11/2017	38.69
LC08_L1GT_231072_20171205_20171222_01_T2	231 - 72	05/12/2017	100

TABLE. VI. DESCRIPTION OF THE COLLECTED EPIDEMIOLOGICAL DATA

Database	Form	Disease	Year	Months	Weeks
Notification for epidemiological surveillance	302 a	Acute Chagas	2017	January to December	52
Monthly notification for epidemiological surveillance	302 b	Acute Chagas	2017	January to December	52

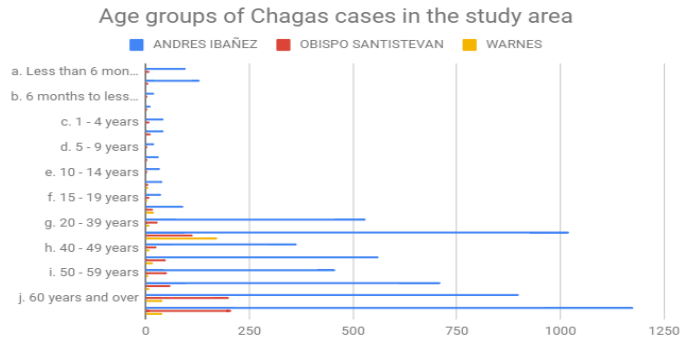

Fig. 6. Age Groups of Chagas Cases in the Study Area.

As it can be observed in Fig. 7 the Chagas disease has been found in 11 types of health facilities, both public and private, belonging to the health network of the Bolivian state. It can be seen that there is a greater presence of Chagas in the Ambulatory Health Centers, on the contrary it can be seen that there is a lower presence of the disease in the poly-medical centers.

Table VII is a summary of the epidemiological information extracted and shows the total cases of Chagas in three provinces selected as the study area. We can only observe the cases of Chagas of certain months, this is due to the cloud coverage, as was previously explained we only select the months of the year with a low percentage of clouds.

c) Meteorological data: The meteorological data are essential since they are the data that we seek to equate with the environmental parameters that will be obtained from the satellite images to correlate them with the epidemiological data.

The meteorological data come from the meteorological stations of the National Service of Meteorology and Hydrology (SENAMHI) of Bolivia (http://senamhi.gob.bo/index.php /inicio).

The extracted data are daily precipitation, daily relative humidity, and the daily visibility.

In Table VIII, one can see a table with the collected meteorological data for February, July, August, October, and November of 2017.

The data collected corresponds to the year 2017 (which corresponds to the year that is being analyzed in the present work with the satellite images) in 852450 SLET meteorological station, since this station is in the area under analysis. One chose February, July, August, October, and November because these months correspond to the satellite images.

C. Treatment of Satellite Images

It is necessary to extract the environmental parameters of the satellite images, for this the images must be cropped with the shape of the area of interest, corrections must be made as part of the preprocessing stage and finally an algorithm must be programmed in MATLAB to extract the necessary information from the satellite images.

a) Preprocessing: To begin with the preprocessing of the images, the maps of departmental, provincial and municipal limits were used in shape file format, selecting those that correspond to the area of interest identified in the epidemiological analysis.

The shapefiles were downloaded from the website of GeoBolivia, which were based on the maps created by the Bolivian Geographical Institute, which has the information provided by the Ministry of Autonomies with data updated to April 2015, this divides Bolivia into 339 municipalities.

These geographic information files were created on April 15, and its publication as an available web resource occurred on May 10, 2015, under the OGC protocol: WMS-1.1.1-http-get-map.

After obtaining the shapefile of the area of interest, which as indicated above, comprises three provinces: Andrés Ibañez, Obispo Santistevan and Warnes. The next process was to crop the 11 bands and the BQA raster of each multispectral image using the obtained shapefile, in each subset the spatial resolution of the input satellite image is preserved.

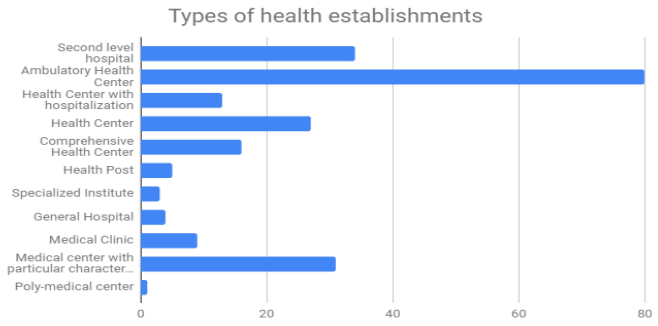

Fig. 7. Type of Health Establishments in the Study Area.

TABLE. VII. CHAGAS CASES IN THE STUDY PROVINCES

Province	Municipality	February	July	August	October	November
Andres Ibañez	COTOCA	3	13	23	23	31
	EL TORNO	5	1	2	9	7
	LA GUARDIA	7	22	15	6	11
	PORONGO	0	0	0	0	0
	SANTA CRUZ DE LA SIERRA	472	531	640	469	466
Obispo Santistevan	FERNANDEZ ALONSO	17	1	1	2	4
	GENERAL SAAVEDRA	0	0	2	2	0
	MINEROS	11	0	0	0	0
	MONTERO	44	50	18	55	26
	SAN PEDRO	3	4	2	18	13
Warnes	OKINAWA	2	0	0	1	1
	WARNES	41	49	6	0	0

TABLE. VIII. METEOROLOGICAL DATA FOR 2017 (852450 SLET METEOROLOGICAL STATION)

	February	July	August	October	November
Temperature	27.5	25.4	22.4	31.4	28.4
Relative humidity	73.0	60.0	34.0	52.0	67.0
Visibility	11.3	11.4	12.9	6.9	12.9

For achieve this objective QGIS 3.4 Madeira was used, this procedure is important in order to obtain the data only from the area of interest, in this case the area contained only in the path 231 and row 72.

After cutting the raster according to the area of interest, we proceed to perform the correction procedures using the ENVI 5.3 software.

Due to interferences in the satellite instruments, different effects can occur that can affect the images acquired, which is why, in order to acquire valid information from the satellite images, two corrections are made, the first is a radiometric correction and the second an atmospheric correction.

Radiometric correction is a technique used to reduce anomalies caused by the sensor system or the conditions of the image capture, this correction resets the digital values of the image.

In order to change the original multispectral images obtained by the sensors at a radiometric scale, the spectral radiance is calculated [17]. Radiometric correction is an indispensable step in the creation of high-level images for subsequent processes.

The following parameters, observed in Table IX, were obtained after performing the radiometric correction.

It must be remembered that the electromagnetic radiation of the earth is captured through the satellite sensors. This energy is known as radiance [18].

The atmospheric correction is the technique of evaluate and eliminate the atmospheric and terrain distortions that are introduced in the values of radiance that arrive at the sensor from the Earth's surface [19]. The main objective of this correction is recovering the physical parameters of the terrestrial surface including the reflectance of the surface, the visibility of the ground and the temperature.

The model used for atmospheric correction is Fast Line of Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH). For the use of this module in ENVI 5.3, some input parameters that detail the characteristics of the image are required. [20] Besides, the input data type must be a radiance file of floating-point and in the format of Band Interleaved by Line (BIL).

Moreover, for the appropriate atmospheric correction of the image, meteorological data are needed, also information such as height, date and time, the type of sensor that models the relative spectral response. For which it was identified that the average altitude of the study area is 384 m.a.s.l and the meteorological data were extracted from the Trompillo weather station, located at the GPS point (Latitude -17.8, Longitude -63.16).

TABLE. IX. RESULTS OF THE RADIOMETRIC CORRECTION

Variables	20/02	14/07	31/08	18/10	3/11
Sun Azimuth	82.68	39.1	50.84	76.58	88.72
Sun elevation	57.45	40.64	51.09	63.95	65.63
Cloud cover	33.55	10.9	5.61	20.27	38.69
Earth sun distance	0.99	1.01	1.01	1	0.99

Within the specific configuration for this correction, the multispectral settings must be configured, an important step is to select specific bands for aerosol retrieval, for which the Kaufman-Tanre 1997 method is used [21], these method identify dark pixels used for the visibility estimate, the recommended wavelength ranges for this model is (640 - 680) nm for the lower channel and (2100 - 2250) nm for the upper channel.

In this case, the SWIR2 band (2.21010) is used as the upper channel and the red band (0.6546) as the lower channel.

In summary, the parameters used for the proper configuration of the FLAASH atmospheric correction tool are:

1) Scene center: Given by the geographical coordinates of the center of the scene, in this case corresponding to Path 231 and row 72. (Latitude: -17'20"47.42 and Longitude: -63'41"28.03).
2) Sensor type: Multispectral - Landsat8 OLI.
3) Sensor altitude: 705 km.
4) Ground elevation: 0.384 km.
5) Pixel size: 30 m.
6) Date of image acquisition: See Table X.
7) Flight time: See Table X.
8) Atmospheric model: Tropical.
9) Aerosol model: Urban.
10) Visibility: See Table X.

b) Processing: The first step is to identify the environmental parameters that are extracted from the different spectral bands of the satellite images.

In Table XI, we can observe a list of the indexes calculated, their respective definitions [22], [23], [24] and their equations.

We proceeded to program an algorithm to extract the environmental parameters, for this the MATLAB software was used. The algorithm develops three major processes, as can be seen in Fig. 8.

TABLE. X. DATA FOR THE ATMOSPHERIC CORRECTION

Variable	20/02	14/07	31/08	18/10	3/11
Temperature	27.5	25.4	22.4	31.4	28.4
Time (HH:MM:SS)	14:16:50	14:16:44	14:17:01	14:17:13	14:17:13
VV (average visibility)	11.3	11.4	12.9	6.9	12.9
Average relative humidity	73	60	34	52	67
Atmospheric Model	Tropical	Tropical	Tropical	Tropical	Tropical

TABLE. XI. ENVIRONMENTAL INDEXES

Name	Definition	Equation
NDSI	Normalized Difference Soil Index	$NDSI = \frac{SWIR - NIR}{SWIR + NIR}$
NDMI	Normalized Difference Moisture Index	$NDMI = \frac{NIR - SWIR_1}{NIR + SWIR_1}$
NDVI	Normalized Difference Vegetation Index	$NDVI = \frac{NIR - Red}{NIR + Red}$
NDWI	Normalized Difference Water Index (Content in leaves)	$NDWI = \frac{NIR - SWIR_1}{NIR + SWIR_1}$
MNDWI	Modified Normalized Difference Water Index	$MNDWI = \frac{Green - SWIR}{Green + SWIR}$

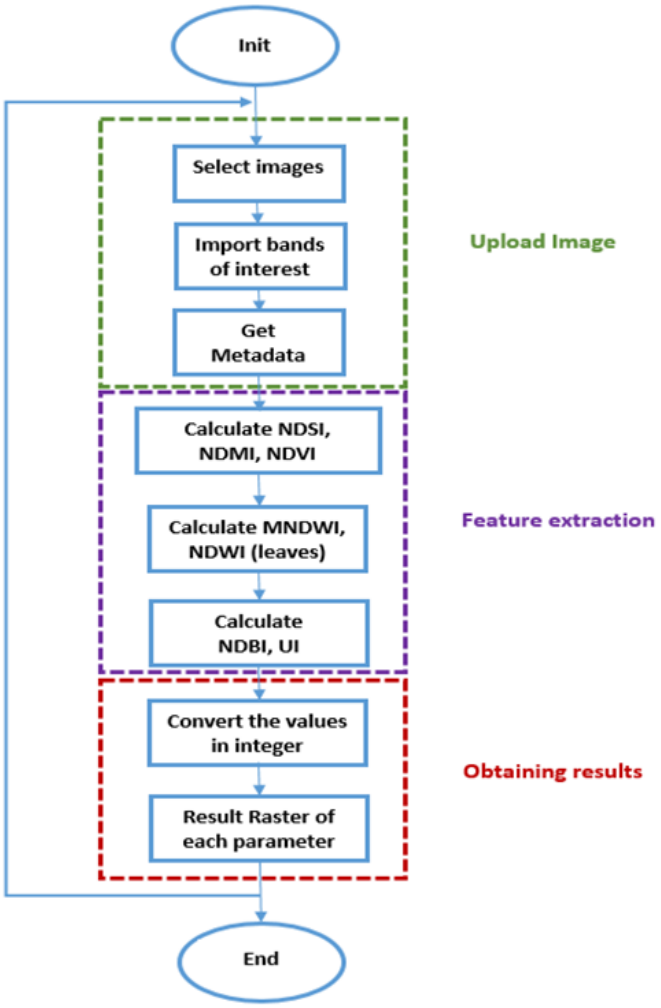

Fig. 8. Flow Diagram of the Processing Algorithm.

Upload the image: In this stage the satellite image must be selected, the bands of interest used for the calculations must be imported and the metadata must be read.

Feature extraction: In this step, different parameters must be calculated, for which different equations must be programmed using different spectral bands of the image, in order to obtain different indexes such as the NDVI, NDSI, among others.

Obtaining results: In this stage the raster resulting from the calculation of each of the parameters must be obtained, an evaluation of the results obtained must be carried out to corroborate that the obtained indexes are within the maximum ranges established in the theory, to confirm that the calculations made are correct.

IV. ANALYSIS OF THE RELATIONSHIP OF ALL THE DATA EXTRACTED

Among the obtained data, one has seven parameters or indices extracted from the satellite images, three parameters from the meteorological stations, and the data related to the cases of Chagas in the area under analysis. The parameters obtained from the satellite images and the data from the meteorological station, both dataset form a total of ten features.

The first analysis to be performed is the calculation of the correlation between the ten selected features.

In Table XII, one can see the Pearson correlation between the eight selected features where a good correlation is shown between the NDMI and the NDWI. The p-value of the correlations previously mentioned (see Table XIII) is 0.0001 in both cases; this value being less than 0.05 confirms the significance of both correlations.

Having ten features, one applies the principal component analysis (PCA) to reduce the space of variables to be analyzed.

After the PCA, it is observed that the workspace is reduced only to four principal components (see Table XIV).

Among the four main components shown in Table XV, they explain the 100% of the data collected. As shown in Table XII and Fig. 9, Component 1 explains more than 94% of the data.

Since Component 1 is the one that practically explains the data thoroughly, we look for the correlation of Component 1 with the epidemiological data that represent the number of cases of Chagas in the area under analysis.

In Table XVI, it can be seen that there is a high correlation (Pearson correlation) between Component 1 and epidemiological data. This absolute value of correlation is equal to 0.8469.

In Fig. 10, we can observe a correlation map between Chagas cases and the results found in the analysis of the principal components extracted from Table XIV.

The characteristics of component 1 obtained by PCA were combined to create heat maps where the blue areas represent high values and the orange areas correspond to low values.

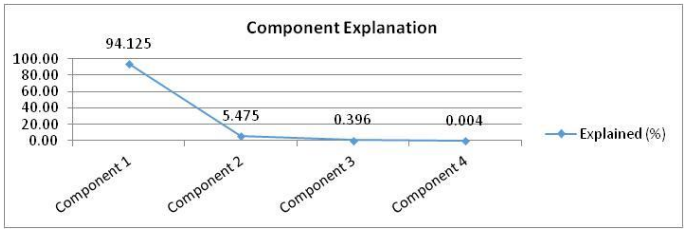

Fig. 9. Distribution of the Percentage of Explanation for the Four Principal Components after the PCA.

TABLE. XII. CORRELATION BETWEEN ALL TEN FEATURES FROM SATELLITE IMAGES AND METEOROLOGICAL STATION

	NDSI	NDMI	NDVI	NDWI (Leaves)	MNDWI	Temperature	Relative humidity	Visibility
NDSI	1,0000	-0,9976	-0,5035	-0,9976	-0,6463	-0,1595	0,4019	-0,0570
NDMI	-0,9976	1,0000	0,4551	1,0000	0,6818	0,1573	-0,4033	0,1024
NDVI	-0,5035	0,4551	1,0000	0,4551	-0,3268	-0,3267	-0,0563	-0,0513
NDWI (Leaves)	-0,9976	1,0000	0,4551	1,0000	0,6818	0,1573	-0,4033	0,1024
MNDWI	-0,6463	0,6818	-0,3268	0,6818	1,0000	0,4816	-0,3471	0,0405
Temperature	-0,1595	0,1573	-0,3267	0,1573	0,4816	1,0000	-0,7676	-0,7429
Relative humidity	0,4019	-0,4033	-0,0563	-0,4033	-0,3471	-0,7676	1,0000	0,4133
Visibility	-0,0570	0,1024	-0,0513	0,1024	0,0405	-0,7429	0,4133	1,0000

TABLE. XIII. P-VALUE FROM CORRELATION SHOWED IN TABLE XII

	NDSI	NDMI	NDVI	NDWI (Leaves)	MNDWI	Temperature	Relative humidity	Visibility
NDSI	1,0000	0,0001	0,3872	0,0001	0,2387	0,7978	0,5024	0,9274
NDMI	0,0001	1,0000	0,4412	0,0000	0,2049	0,8006	0,5008	0,8698
NDVI	0,3872	0,4412	1,0000	0,4412	0,5915	0,5915	0,9284	0,9347
NDWI (Leaves)	0,0001	0,0000	0,4412	1,0000	0,2049	0,8006	0,5008	0,8698
MNDWI	0,2387	0,2049	0,5915	0,2049	1,0000	0,4115	0,5671	0,9484
NDBI	0,0001	0,0000	0,4412	0,0000	0,2049	0,8006	0,5008	0,8698
UI	0,0003	0,0018	0,3208	0,0018	0,2921	0,7973	0,5077	0,9971
Temperature	0,7978	0,8006	0,5915	0,8006	0,4115	1,0000	0,1297	0,1503
Relative humidity	0,5024	0,5008	0,9284	0,5008	0,5671	0,1297	1,0000	0,4892
Visibility	0,9274	0,8698	0,9347	0,8698	0,9484	0,1503	0,4892	1,0000

TABLE. XIV. RESULT OF THE PCA. THE WORKSPACE IS REDUCED TO FOUR PRINCIPAL COMPONENTS

	Comp. 1	Comp. 2	Comp. 3	Comp. 4
Coefficients	-0,00346636	0,00350483	0,02501616	-0,47097695
	0,00344560	-0,00393762	-0,01981733	0,45440768
	0,00044936	-0,00374080	-0,05152317	0,04637818
	0,00344560	-0,00393762	-0,01981733	0,45440768
	0,00347139	0,00234426	0,03949607	0,60243097
	0,11338987	0,76616650	0,63035961	0,00760066
	0,99352569	-0,08817349	-0,07090754	-0,00778362
	-0,00119390	-0,63651515	0,76941084	0,00395777

TABLE. XV. PERCENTAGE OF EXPLANATION OF THE FOUR PRINCIPAL COMPONENT AFTER THE PCA

	Explained (%)
Component 1	94.125
Component 2	5.475
Component 3	0.396
Component 4	0.004

TABLE. XVI. CORRELATION BETWEEN THE COMPONENT 1 AND THE EPIDEMIOLOGICAL DATA

	Pearson
Correlation	-0.8469

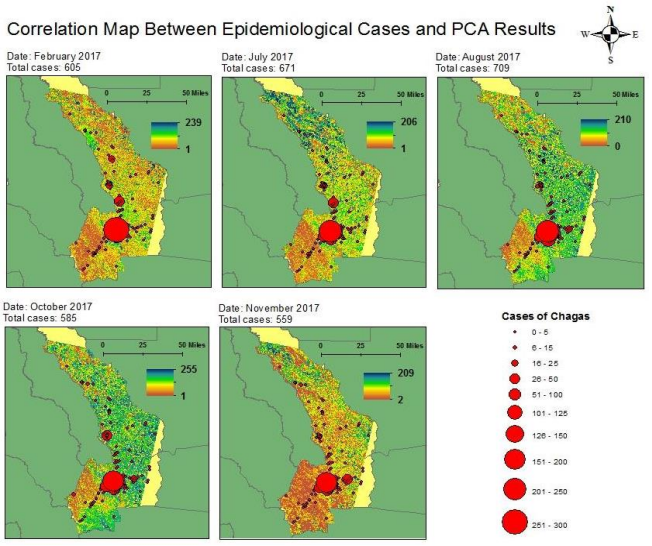

Fig. 10. Correlation Map between Epidemiological Cases and PCA.

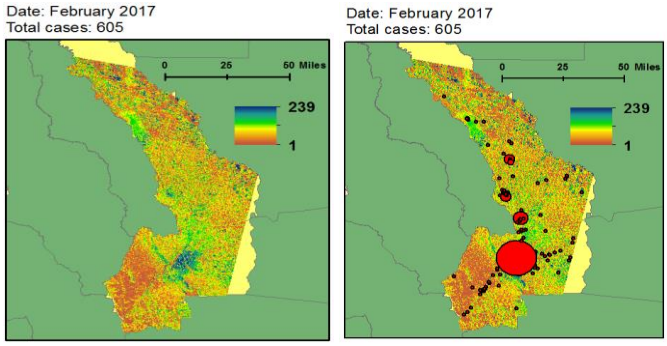

Fig. 11. Comparison between the Correlation Map and the Heat Map.

It can be observed in the five months of the study that the areas with the highest number of Chagas cases also present high and concentrated values of the characteristics of component 1, as can be seen in Fig. 11 that shows the results of February.

V. Conclusions

Five environmental parameters were extracted from the satellite images, this are the NDSI, NDMI, NDVI, NDWI, and MNDWI. These environmental parameters were complemented with weather data as temperature, relative humidity, and visibility extracted from meteorological station. Through the principal components analysis (PCA) of all the environmental variables, it was established that only one component is necessary to explain almost 94% of the data. The parameter that has a strong positive influence on the appearance of Chagas outbreaks according to the first component is the Relative Humidity. The parameter that has a strong positive influence according to the second component is the Temperature. The parameter that has a strong positive influence according to the third component is the Visibility. Finally, the parameter that has a strong positive influence on the appearance of Chagas outbreaks according to the fourth component is the MNDWI.

As a general conclusion of the present study, it has been possible to obtain a comprehensive understanding of the disease based on the use of geospatial technology with 84.69% of correlation. One studied the five months with larger number of cases, between them one establish that the months that present an increase of cases of Chagas in the study area are the months of July and August, this months present environmental conditions that have a strong influence on the occurrence of cases.

As one can see in the study of state of the art, different satellites were used for the environmental analysis related to the propagation of vectors. The contribution of this work is represented mainly in the demonstration of the use of Lansat 8 images for this type of application of epidemiological study specifically Chagas in Bolivia.

As a future work, we want to include data from multiple satellites to reduce the temporal resolution of the satellite used in the present study, besides it is desired to build a geographic information system that contains all the data extracted from the analysis of the satellite images, in addition it is planned to analyze multiple years and finally it is planned to develop an alert platform that can indicate the potential areas that may present Chagas cases according to their environmental conditions.

References

[1] Orellana-Halkyer N, Arriaza-Torres B. Enfermedad de Chagas en poblaciones prehistóricas del norte de Chile. Rev Chil Hist Nat 2010; 83: 531–541.

[2] Chagas CI de la IA de C de la TV y T de la E de. Definición de variables y criterios de riesgos para la caracterización epidemiológica e identificación de áreas prioritarias en el control y vigilancia de la transmisión vectorial de la Enfermedad de Chagas. Uniandes. Fac. de Ciencias, 2004.

[3] O. P. de la Salud, «Enfermedad de Chagas: guía para vigilancia, prevención, control y manejo clínico de la enfermedad de Chagas aguda transmitida por alimentos», Ser. Man. Téc. 12, 2009.

[4] Barbosa M das GV, Ferreira JMBB, Arcanjo ARL, et al. Chagas disease in the State of Amazonas: history, epidemiological evolution, risks of endemicity and future perspectives. Rev Soc Bras Med Trop 2015; 48: 27–33.

[5] C. H. Rotela, «Desarrollo de Modelos e Indicadores Remotos de Riesgo Epidemiológico de Dengue en Argentina.», abr. 2019.

[6] M. Neteler, D. Roiz, D. Rocchini, C. Castellani, y A. Rizzoli, «Terra and Aqua satellites track tiger mosquito invasion: modelling the potential distribution of Aedes albopictus in north-eastern Italy», *Int. J. Health Geogr.*, vol. 10, n.º 1, p. 49, ago. 2011.

[7] Lanfri S, Frutos N, Porcasi X, Rotela C, Peralta G, De Elia E, Lanfri M, Scavuzzo M. -Algoritmos para el Alerta Temprana de Dengue en un Ambiente Geomático. - Instituto de Altos Estudios Espaciales Mario Gulich, Comisión Nacional de Actividades Espaciales. Centro Espacial Teófilo Tabanera, Córdoba, Argentina - (89 - 104).

[8] Díaz ML, González CI. Enfermedad de Chagas agudo: transmisión oral de Trypanosoma Cruzi como una vía de transmisión re-emergente. Rev Univ Ind Santander Salud 2014; 46: 177–188.

[9] J. A. Pérez-Molina y I. Molina, «Chagas disease», Lancet Lond. Engl. vol. 391, n.o 10115, pp. 82-94, 06 2018.

[10] Uribe AG, Bernal GB. Ministro de salud y Protección social. 34.

[11] R. A. Kolliker-Frers, I. Insua, G. Razzitte, y F. Capani, «Chagas disease prevalence in pregnant women: migration and risk of congenital transmission», J. Infect. Dev. Ctries., vol. 10, n.o 09, pp. 895-901, sep. 2016.

[12] F. M. Sano, M. Sano, y O. P. de la Salud, «La enfermedad de Chagas. A la puerta de los 100 años del conocimiento de una endemia ancestral», A doença de Chagas. Perto dos 100 anos do conhecimento de uma endemia americana ancestral, 2007.

[13] C. Forsyth, «From Lemongrass to Ivermectin: Ethnomedical Management of Chagas Disease in Tropical Bolivia», Med. Anthropol., vol. 37, n.o 3, pp. 236-252, 2018.

[14] Ministry of Health Bolivia, "Boletín de vigilancia epidemiológica, Chagas Agudo", 2016.

[15] A. Rassi, A. Rassi, y J. Marcondes de Rezende, «American trypanosomiasis (Chagas disease)», Infect. Dis. Clin. North Am., vol. 26, n.o 2, pp. 275-291, jun. 2012.

[16] Roy DP, Kovalskyy V, Zhang HK, et al. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote Sens Environ 2016; 185: 57–70.

[17] Chander G, Markham BL, Helder DL. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sens Environ 2009; 113: 893–903.

[18] Aguilar-Arias H, Mora-Zamora R, Vargas-Bolaños C. Metodología para la Corrección Atmosférica de Imágenes Aster, Rapideye, Spot 2 y Landsat 8 con el Módulo Flaash del Software ENVI. Rev Geográfica América Cent 2014; 2: 39–59.

[19] Tardy B, Rivalland V, Huc M, et al. A software tool for atmospheric correction and surface temperature estimation of landsat infrared thermal data. Remote Sens; 8. Epub ahead of print 2016. DOI: 10.3390/rs8090696.

[20] Kruse F. Comparison of (ATREM), (ACORN), and (FLAASH) Atmospheric Corrections using Low-Altitude (AVIRIS) Data of Boulder, Colorado. 2004.

[21] Remer LA, Tanré D, Kaufman YJ. Algorithm for remote sensing of tropospheric aerosol from MODIS: Collection 005.

[22] Deng Y, Wu C, Li M, et al. RNDSI: A ratio normalized difference soil index for remote sensing of urban/suburban environments. Int J Appl Earth Obs Geoinformation 2015; 39: 40–48.

[23] Xu H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int J Remote Sens 2006; 27: 3025–3033.

[24] Zha Y, Gao J, Ni S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. Int J Remote Sens 2003; 24: 583–594.