

UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ
FACULTADES DE CIENCIAS QUÍMICAS, INGENIERÍA, MEDICINA
Y CIENCIAS SOCIALES Y HUMANIDADES
PROGRAMA MULTIDISCIPLINARIO DE POSGRADO EN CIENCIAS AMBIENTALES
AND
TH KÖLN - UNIVERSITY OF APPLIED SCIENCES
INSTITUTE FOR TECHNOLOGY AND RESOURCES MANAGEMENT IN THE TROPICS AND
SUBTROPICS

MAPPING CHAGAS DISEASE TRANSMISSION RISK USING VECTOR AND
RESERVOIR DISTRIBUTION MODELS IN SAN LUIS POTOSÍ, MÉXICO

THESIS TO OBTAIN THE DEGREE OF
MAESTRÍA EN CIENCIAS AMBIENTALES
DEGREE AWARDED BY UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ
AND
MASTER OF SCIENCE
NATURAL RESOURCES MANAGEMENT AND DEVELOPMENT
DEGREE AWARDED BY TH KÖLN – UNIVERSITY OF APPLIED SCIENCES

PRESENTS:

SAÚL ALFREDO GÁMEZ HIDALGO

CO-DIRECTOR OF THESIS (*PMPCA*)
DR. MARÍA GUADALUPE GALINDO MENDOZA

CO-DIRECTOR OF THESIS (*ITT*)

DR. CLAUDIA RAEDIG

ASSESSOR:

DR. GILBERTO BALLESTEROS RODEA

PROYECTO REALIZADO EN:

SAN LUIS POTOSÍ, MÉXICO Y COLONIA, ALEMANIA

INSTITUTO O FACULTAD

PMPCA E ITT

UNIVERSIDAD

UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ

Y

TH KÖLN - UNIVERSITY OF APPLIED SCIENCES

CON EL APOYO DE:

CONSEJO NACIONAL DE CIENCIA Y TECNOLOGÍA (CONACYT)

**LA MAESTRÍA EN CIENCIAS AMBIENTALES RECIBE APOYO A TRAVÉS DEL PROGRAMA
NACIONAL DE POSGRADOS (PNPC - CONACYT)**



Mapping Chagas disease transmission risk using vector and reservoir distribution models in San Luis Potosi, Mexico by Sául Alfredo Gámez Hidalgo is licensed under a [Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Erklärung / Declaración

Name / Nombre: Saúl Alfredo Gámez Hidalgo

Matrikel-Nr. / N° de matrícula: 11131786 (TH Köln), 0302300 (UASLP)

Ich versichere wahrheitsgemäß, dass ich die vorliegende Masterarbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen sind, sind als solche kenntlich gemacht.

Aseguro que yo redacté la presente tesis de maestría independientemente y no use referencias ni medios auxiliares a parte de los indicados. Todas las partes, que están referidas a escritos o a textos publicados o no publicados son reconocidas como tales.

Die Arbeit ist in gleicher oder ähnlicher Form noch nicht als Prüfungsarbeit eingereicht worden.

Hasta la fecha, un trabajo como éste o similar no ha sido entregado como trabajo de tesis.

Köln, den /el 11 septiembre del 2020

Unterschrift / Firma: Alfredo Hidalgo

Ich erkläre mich mit einer späteren Veröffentlichung meiner Masterarbeit sowohl auszugsweise, als auch Gesamtwerk in der Institutsreihe oder zu Darstellungszwecken im Rahmen der Öffentlichkeitsarbeit des Institutes einverstanden.

Estoy de acuerdo con una publicación posterior de mi tesis de maestría en forma completa o parcial por las instituciones con la intención de exponerlos en el contexto del trabajo investigación de las mismas.

Unterschrift / Firma: Alfredo Hidalgo

Agradecimientos

Quiero agradecer primeramente a mi comité tutelar; la Dra. Guadalupe Galindo, la Dra. Claudia Raedig y el Dr. Gilberto Ballesteros, por sus consejos y apoyo durante el proceso de investigación, por su ayuda en la conceptualización de la idea de investigación, y por la retroalimentación recibida.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) de México por otorgarme las becas que me permitieron estudiar la maestría en México y Alemania, así como al PMPCA por nominarme a estas. También agradezco al CONACYT de El Salvador por brindarme apoyo económico para instalarme en México.

Agradezco a Sandra por toda la ayuda brindada con el papeleo de la maestría por parte del ITT y también por su ayuda con la inscripción a los cursos de alemán y el papeleo del seguro médico en Alemania. También agradezco a Maricela y a Laura por su ayuda con el papeleo por parte del PMPCA y CONACYT.

Agradezco a mis padres, Miguel Gámez y Patricia Hidalgo por apoyarme siempre, por alentarme a cumplir mi sueño de estudiar en el extranjero y por estar pendientes de mi durante estos dos años de maestría.

Agradezco a mis compañeros/amigos/familia de la maestría, por su apoyo y sus consejos durante el curso de estos dos años, así como por los buenos momentos y experiencias que compartimos en estos dos años.

Agradezco al equipo técnico-científico del LaNGIF por su compañía y sus consejos en el laboratorio, especialmente a Deniss por introducirme al modelado de distribución de especies usando Maxent.

Agradezco a Ricardo, por el apoyo emocional en momentos de estrés y por hacerme ver que los problemas de aquellos momentos sí tenían solución.

Agradezco a mi tío Armando por recibirme en México y estar pendiente de mi durante los dos años de maestría.

Agradezco a Guillermo Recinos por darme sus consejos y contarme sus experiencias sobre el modelado de distribución de especies y a Ana Miriam González por dejarme usar su foto de triatomos.

Agradezco a mis amigos de El Salvador, los del colegio y la U, por compartir sus ocurrencias en redes sociales conmigo y a su vez, por dejarme contarles experiencias buenas, malas y chistosas que viví durante la maestría.

Finalmente, quiero agradecer a todos los demás amigos y familiares que estuvieron apoyándome a distancia.

Abstract

Keywords: Chagas disease, species distribution modelling, vectors, reservoirs, risk mapping.

Chagas disease is a parasitic infection endemic to America, caused by the protozoan *Trypanosoma cruzi* and mainly transmitted to humans by contact with insect species of the Triatominae subfamily (Hemiptera). The disease is known to affect disproportionately rural impoverished human communities where it is known to cause premature death and is considered a social and economic burden. The Mexican government has made important progress into the detection, surveillance, treatment, and prevention of the disease in the last decades, however, Chagas disease has also been reported in areas where it had not been previously reported, and there are still barriers for access to treatment. In the state of San Luis Potosi, the disease is more studied in the east, nevertheless, it has been estimated that the reported cases of the entire state have been underestimated. New approaches to detect Chagas risk areas could help prioritize locations for Chagas disease education and prevention programs, detect cases of the disease in a timely manner, and provide access to the necessary treatments. The objective of this study was to identify risk areas for the transmission of Chagas disease in San Luis Potosí using species distribution modelling to estimate vectors and reservoirs' distributions. To do this, firstly, important vectors and one reservoir species of *T. cruzi* were identified by reviewing their reported infection rates in literature and the number of times reported in Mexico. Next, species distribution models were calculated for the chosen vector and reservoir species present in the state. The models were done using the Maxent algorithm. Lastly, the resulting distribution models were combined into a risk map by thresholding the model outputs to produce binary predictions and then performing an overlap spatial analysis. Vector species were found to have suitable areas in 36.08% of the state's territory while areas suitable for both vectors and reservoir were 7.4% of the state's total area. While this figure may look small at first glance, the analysis suggests that 30% of the rural population and 52% of the urban population of the state are living in an area suitable for vectors and reservoir and therefore at risk. Species distribution modelling can be a powerful tool for identifying human populations at risk of contracting Chagas disease. In the future, including different species of reservoirs into the analysis could help to discover new risk areas in the state.

Resumen

Palabras clave: Enfermedad de Chagas, modelado de distribución de especies, vectores, reservorios, mapas de riesgo.

La enfermedad de Chagas es una infección parasitaria endémica de América, causada por el protozoo *Trypanosoma cruzi* y transmitida principalmente a los humanos por contacto con insectos de la subfamilia Triatominae (Hemiptera). Se sabe que la enfermedad afecta de manera desproporcionada a las comunidades humanas empobrecidas de las zonas rurales, donde causa muerte prematura y donde se considera una carga social y económica. El Gobierno de México ha mejorado en la detección, vigilancia, tratamiento y prevención de la enfermedad en las últimas décadas; sin embargo, también se ha reportado la enfermedad de Chagas en zonas en las que previamente no se había reportado anteriormente, a su vez, sigue habiendo obstáculos para el acceso al tratamiento. En el estado de San Luis Potosí, la enfermedad se ha estudiado más en el oriente, sin embargo, se ha calculado que el número de casos que se maneja de manera oficial subestima el número real de casos. La utilización de nuevos enfoques para detectar áreas de riesgo para la enfermedad de Chagas podría ayudar en varios aspectos, como la selección de lugares prioritarios donde realizar intervenciones específicas, detectar los casos de la enfermedad de manera oportuna y proporcionar acceso a los tratamientos necesarios. El objetivo de este estudio fue identificar áreas de riesgo para la transmisión de la enfermedad de Chagas en San Luis Potosí, utilizando modelos de distribución de vectores y reservorios. Para ello se identificaron vectores importantes y una especie de reservorio de *T. cruzi*, por medio de la revisión de sus tasas de infección reportadas en la literatura y el número de veces que se reportaron en estudios en México. A continuación, se calcularon los modelos de distribución de especies vectores y del reservorio escogidos presentes en el estado. Los modelos se elaboraron utilizando el algoritmo de Maxent. Por último, los modelos de distribución resultantes se combinaron en un mapa de riesgo, primero produciendo mapas de distribución binarios y luego realizando un análisis espacial de superposición. Se encontró que las especies de vectores tienen áreas adecuadas en el 36,08% del territorio del estado, mientras que las áreas adecuadas tanto para vectores como para reservorios son el 7,4% del área total del estado. Si bien esta cifra puede parecer pequeña a primera vista, el análisis sugiere que el 30% de la población rural y el 52% de la población urbana del estado viven en una zona adecuada para vectores y reservorios y, por lo tanto, están en riesgo. El modelado de distribución de especies puede ser una herramienta poderosa para identificar las poblaciones humanas en riesgo de contraer la enfermedad de Chagas. En el futuro, la inclusión de diferentes especies de reservorios en el análisis podría ayudar a descubrir nuevas zonas de riesgo en el estado.

Index of contents

List of tables	1
List of figures	2
Abbreviations	3
1 Introduction	4
1.1 Chagas disease.....	4
1.1.1 Phases of Chagas disease in humans	4
1.1.2 Transmission of Chagas disease	4
1.1.3 Wild reservoirs of <i>T. cruzi</i>	5
1.1.4 Transmission cycles of <i>T. cruzi</i>	6
1.1.5 Chagas disease in Mexico.....	7
1.1.6 Chagas disease in San Luis Potosí.....	7
1.2 Species distribution and ecological niche modeling	7
1.2.1 Ecological niche modelling concepts	8
1.2.2 Modelling methods	9
1.2.3 Model calibration and evaluation	10
1.2.4 Niche modelling of diseases and risk mapping	10
1.3 Problem statement.....	11
3 Objective and research questions	12
3.1 General objective.....	12
3.2 Research questions	12
4 Methods	13
4.1 Study area.....	13
4.2 Methods overview	15
4.3 Selection of important vector and reservoir species.....	15
4.3.1 Vector species	15
4.3.2 Reservoir species	16
4.4 Distribution modelling of vector and reservoir species	16
4.4.1 Species niche description.....	16
4.4.2 Species occurrence data collection and cleaning	16

4.4.3	Environmental predictor variables collection and cleaning.....	17
4.4.4	Modelling.....	17
4.5	Mapping risk areas	18
5	Results	19
5.1	Selection of important vector and reservoir species.....	19
5.1.1	Important reservoir species identification.....	19
5.2	Distribution modelling of vector and reservoir species	25
5.2.1	Niche description of species to model	25
5.2.2	Triatomine occurrence data collection and cleaning	29
5.2.3	Reservoir occurrence data collection and cleaning.....	32
5.2.4	Model training and evaluation	34
5.3	Mapping risk areas	43
6	Discussion and conclusions	48
6.1	Selection of important reservoir species	48
6.2	Distribution modelling of vectors and reservoir species	49
6.3	Mapping risk areas	52
	Reference list	56
	Annexes	66

List of tables

Table 1: list of parameters used in model training.	18
Table 2: List of mammal individuals found infected with <i>T.cruzi</i> in studies about reservoirs of Chagas disease in Mexico.	21
Table 3: List of species reported to be infected by <i>T. cruzi</i> in the literature in Mexico, with reporting status in San Luis Potosí.	24
Table 4: Niche description of <i>Triatoma dimidiata</i>	25
Table 5: Niche description of <i>Triatoma gerstaeckeri</i>	26
Table 6: Niche description of <i>Triatoma mexicana</i>	28
Table 7: Institution general data and classification by trusted source.	31
Table 8: Number of occurrence points used for modelling	32
Table 9: Model parameters ran and corresponding AIC values..	34
Table 10: Settings and average evaluation statistics of the species distribution models for <i>T. dimidiata</i>	35
Table 11: Settings and average evaluation statistics of the species distribution models for <i>T. gerstaeckeri</i>	37
Table 12: Settings and average evaluation statistics of the species distribution models for <i>T. mexicana</i>	39
Table 13: Settings and average evaluation statistics of the species distribution models for <i>D. virginiana</i>	41
Table 14: Total of rural and urban population living in areas suitable for Chagas disease vectors and reservoirs	45
Table 15: Total predicted presence areas per species in San Luis Potosi.....	47

List of figures

Figure 1: Image of various life stages <i>Triatoma dimidiata</i> , member of the Triatominae subfamily..	5
Figure 2: Wild and domestic transmission cycles of <i>T. cruzi</i>	7
Figure 3: Map of the state of San Luis Potosi divided into its four regions.	13
Figure 4: Climates of San Luis Potosi.	14
Figure 5: Graphic overview of methods used.	15
Figure 6: Mammal Individuals reported infected with <i>T. cruzi</i> by order.	19
Figure 7: Mammal Individuals reported infected with <i>T. cruzi</i> by 7 states of the 32 states of Mexico	20
Figure 8: Rate of reported infections with <i>T. cruzi</i> by species	20
Figure 9: Distribution of occurrences by species and year between 2000 and 2015 with outliers removed	29
Figure 10: Distribution of all triatomine occurrences by year between 1990 and 2015 with outliers removed.	30
Figure 11: Distribution of <i>Didelphis virginiana</i> occurrence data from 2000 to 2020	32
Figure 12: Map of cleaned occurrence points of vectors and reservoir chosen to be used in modelling	33
Figure 13: <i>Triatoma dimidiata</i> Maxent distribution model, logistic output.	36
Figure 14: <i>Triatoma gerstaeckeri</i> Maxent distribution model, logistic output.	38
Figure 15: <i>Triatoma mexicana</i> Maxent distribution model, logistic output.	40
Figure 16: <i>Didelphis virginiana</i> Maxent distribution model, logistic output.	42
Figure 17: Binary prediction of species suitability for vector species <i>T. dimidiata</i> , <i>T. gerstaeckeri</i> , <i>T. mexicana</i> and reservoir species <i>D. virginiana</i> .	44
Figure 18: Vector and reservoir overlapping areas	45
Figure 19: Identified overlaps suitable areas between vectors and reservoir of <i>T. cruzi</i> .	46

Abbreviations

AIC: Akaike information criterion.

AUC: Area under the ROC curve.

MTP: Minimum training presence

MTPTOR: Minimum Training Presence Test Omission Rate.

10PTPTOR: 10 Percentile Training Presence Test Omission Rate.

MTSPS: Maximum test Sensitivity Plus Specificity.

LQH: Linear, quadratic and hinge features.

LQHP: Linear, quadratic, hinge, and product features.

1 Introduction

1.1 Chagas disease

Chagas disease is a parasitic infection caused by the protozoan, *Trypanosoma cruzi* (Chagas 1909). It affects humans and other mammals and is endemic to the American continent, with confirmed cases from southern USA, Mexico, Central America, up to northern Argentina and Chile. It is estimated that eight million people live with the disease worldwide (WHO, 2018). Furthermore, Chagas disease is part of the list of Neglected Tropical Diseases (NTDs), a group of chronic and debilitating illnesses that disproportionately affect rural impoverished communities, especially those of indigenous descent (Hotez *et al.*, 2008).

1.1.1 *Phases of Chagas disease in humans*

The disease includes three phases, an acute phase, an intermediate phase, and a chronic phase. The acute phase lasts around two months and is characterized by a high quantity of parasites in the patient's bloodstream, this is the only phase where the disease is curable. Usually, there are no symptoms in this phase, but a small percentage of people can develop mild fevers, ocular swelling (*romaña* sign) and skin lesions (chagoma). If untreated, the disease can enter an intermediate phase that lasts between 10 and 30 years. In this intermediate phase the parasites enter organ tissues where they cause no clinical manifestations. In about 30 % of patients, the disease advances to the chronic phase where parasites can cause heart and gut disease. In this final phase the disease may cause disability and is potentially fatal (Echeverria and Morillo, 2019; Pérez-Molina and Molina, 2018).

1.1.2 *Transmission of Chagas disease*

Transmission of a disease or infection can be defined as the process by which an infectious agent is spread from a source to an organism, the source can be the environment or another organism (Porta, 2014).

Transmission of Chagas disease to humans (human transmission) can occur primarily through an insect vector, other ways of transmission are direct and include transmission through blood donations contaminated with the parasite, through direct ingestion of the parasite (oral transmission), and transmission from mother to fetus during pregnancy (congenital transmission). There are other less frequent ways of human transmission which include include transmission by laboratory accidents, handling of infected animals, ingestion of uncooked meat contaminated with *T. cruzi* or sexual transmission (Coura, 2015).

The most important route of human transmission is vector-borne transmission, which occurs through contact with the infected feces of blood-sucking bugs of the Triatominae subfamily (Heteroptera) (Figure 1). This subfamily is currently comprised of 140 species, from which only some have been found in human dwellings or infected with parasites, thus having higher epidemiological importance (Galvão and Justi, 2015; Schofield and Galvão, 2009). The best-

known vector species in public health are *Triatoma infestans*, *Triatoma brasiliensis*, and *Panstrongylus megistus* in South America, and *Rhodnius prolixus* and *Triatoma dimidiata* in the Andean countries and Central America. This importance is due to their ability to colonize dwellings and thus, increase the risk of contact with humans (Catalá, Noireau and Dujardin, 2017). These triatomines often enter vulnerable human dwellings (made of adobe or bahareque) and hide in cracks, clutter, and ceilings until nighttime. At night, they come out and feed on sleeping people and domestic animals. These bugs often defecate while feeding, this behavior allows the *T. cruzi* parasites in its feces to enter the bloodstream through contact with the feeding wound (Lent and Wygodzinsky, 1979).



Figure 1: Image of various life stages *Triatoma dimidiata*, member of the *Triatominae* subfamily. Image taken by González (2020) and used with permission.

1.1.3 Wild reservoirs of *T. cruzi*

Triatomine vectors are the most important mechanism of transmission to humans, however, not all triatomines are infected with *T. cruzi* because there is no transovarial transmission of the parasite between triatomines, thus, next generations of bugs are free of *T. cruzi* and only acquire it when feeding from infected mammals (Jurberg and Galvao, 2006). Consequently, *T. cruzi* populations are not maintained by triatomines in the wild, but by wild mammals called reservoirs.

A reservoir is a host species in which the parasite normally lives and reproduces, in this relationship the parasite depends on the reservoir for its survival and future transmission to susceptible hosts and vectors (Porta, 2014). There are roughly 180 species of reservoir species that have been reported to host *T. cruzi* (Herrera, 2010). In México there have been reports of wild animals such as opossums, bats, skunks, racoons, white nosed coatis, mice, rats, and howling monkeys (López-Cancino *et al.*, 2015; Martínez-Hernández *et al.*, 2014; Panti-May *et al.*, 2017; Rovirosa-Hernández *et al.*, 2013; Ruiz-Piña *et al.*, 2018a). There have also been reports of domestic animals infected with *T. cruzi*, such as dogs, cats, pigs, sheep, and horses (Galaviz-Silva *et al.*, 2017; Jiménez-Coello *et al.*, 2012; Longoni *et al.*, 2012; López-Cancino *et al.*, 2015; Ruiz-Piña *et al.*, 2018a).

1.1.4 Transmission cycles of *T. cruzi*

T. cruzi circulate between suitable mammal hosts in wild and in domestic environments through sylvatic and domestic transmission cycles (Figure 2). The parasite's transmission cycle in the wild takes place between mammals and triatomine bugs, and among mammal individuals. Transmission between mammals and triatomines occur when uninfected triatomine bugs feed on infected mammals and become infected themselves. On the contrary, infected triatomines can feed on uninfected mammals and transmit the parasite. Other ways the mammals can acquire the parasite from the bugs are by ingesting them. Transmission between mammals takes place when carnivores feed from infected prey (Herrera, 2014; Roque *et al.*, 2008).

In the domestic transmission cycle, infected triatomines enter human dwellings and seek suitable shelters. They can transmit the parasite to domestic animals which turn into domestic reservoirs themselves and help maintain *T. cruzi* populations in the human dwellings and increase risk of transmission to humans. Triatomines can also transmit the parasite while feeding from humans, they can contaminate food with infected feces or they may be accidentally ingested by people (Filigheddu, Górgolas and Ramos, 2017; Herrera, 2014; Roque *et al.*, 2008).

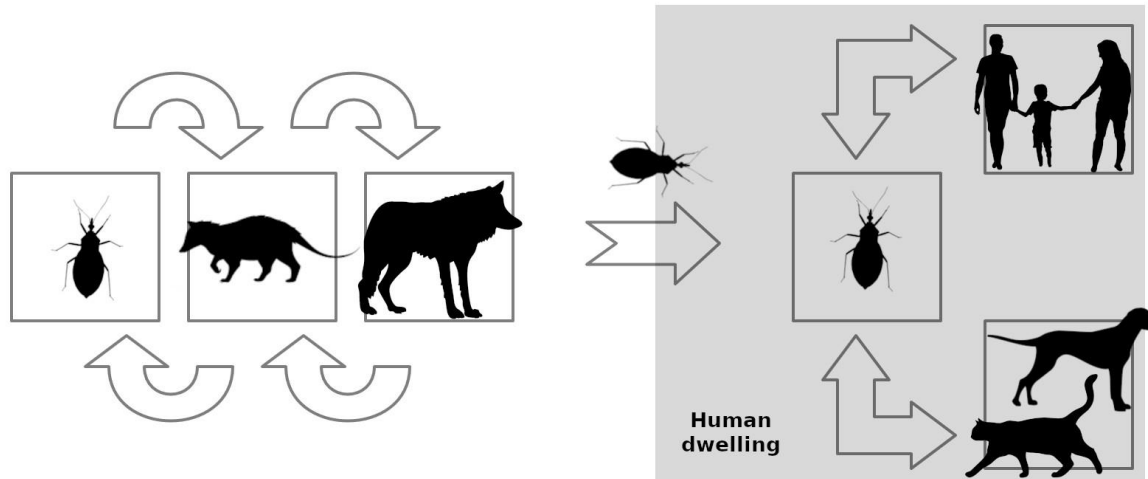


Figure 2: Wild and domestic transmission cycles of *T. cruzi*. Modified from Herrera 2014. Arrows symbolize *T. cruzi* transmission between hosts.

1.1.5 Chagas disease in Mexico

The first human case of Chagas disease in México was described in 1928, but it was considered an exotic disease until the 1970s decade, when it was recognized as a native disease as more nationwide published epidemiological studies were available. In the same decade, research centers were created in the southern states of the country (Chiapas and Yucatán) which continued generating scientific research about the disease (Cruz-Reyes and Pickering-López, 2006). Historically, southern states have registered more cases of the disease and this trend continues to be true. Between 2007 and 2016 there was an increase of 253% detected cases of the disease in Mexico, where southern and central states (Veracruz, Chiapas, Quintana Roo, Oaxaca, Morelos and Yucatán) were the most affected. This indicates that the disease is still a problem in the country (Cruz-Reyes and Pickering-López, 2006; Ibáñez-Cervantes *et al.*, 2018). Currently the main strategies to prevent the disease in the country are home improvement, to eliminate sites where the vector can stay and reproduce, and chemical control, using insecticides (Secretaría de Salud, 2015).

1.1.6 Chagas disease in San Luis Potosí

In the state of San Luis Potosí, Chagas disease has mostly been reported in the Huasteca Region, where, between 2003 and 2012, a total of 257 cases were reported. Most cases were distributed in the south, where indigenous populations have more presence (Medina-Garza *et al.*, 2016). Another study found 6.5 % of Teenek indigenous populations infected, which is higher compared to the national mean of 1.6 % of the total Mexican population (Juarez-Tobias *et al.*, 2009).

1.2 Species distribution and ecological niche modeling

Species distribution and ecological niche modelling are part of a recent growing field of research which uses different theories to extrapolate species distribution data in space based

on environmental data and statistical models (Franklin, 2012; Peterson *et al.*, 2011). These models have been used for different tasks, like assessing invasive species proliferation, for predicting the impact of climate change in species distribution patterns, for conservation planning and reserve selection, and more (Franklin, 2012; Guisan and Thuiller, 2005).

Franklin (2012), describes the elements required modelling: First a theoretical conceptual model of abiotic and biotic factors controlling the species distribution. Second, data on species occurrence. Third, a model linking species occurrence to predictor variables. Fourth, a tool for applying the model to produce maps (GIS), as well as, data and criteria to validate the predictions.

1.2.1 Ecological niche modelling concepts

Species distribution and ecological niche modeling are based on the “ecological niche” concept from Hutchinson (1957), whom described niche as a hypervolume of environmental variables within which a species can survive. There are two types of niches (Franklin, 2012; Peterson *et al.*, 2011; Phillips, Anderson and Schapire, 2006): The fundamental niche which is defined as all the conditions of the environment which would permit the species to survive in long term. The realized niche is defined as the part of the fundamental niche where the species actually occupies, regardless of competition and other biotic interactions. After Hutchinson, many authors have changed some aspects of his definition, mainly the ones referring to the persistence of the species in the environment, the type of variables that are included in the hypervolume, and the feedback loops including the species’ populations (Peterson *et al.*, 2011).

The concepts of ecological niche model and species distribution models have been used interchangeably and it has been argued that this can lead to conceptual inconsistencies and interpretation errors (Peterson and Soberón, 2012). The first difference is that ecological niche modelling is focused on estimating the fundamental niche and the appropriate abiotic space of a species, while species distribution models focus on estimating the area that a species occupies; one takes place exclusively in the hypervolume of environmental variables while the other focuses on the geography. The second difference is that species distribution modelling should consider the dispersal capabilities of the species in order to restrict the species distribution in the geographical space (Peterson and Soberón, 2012; Soberón, Osorio-Olvera and Peterson, 2017).

1.2.2 Modelling methods

There are three types of methods to model the niches and distribution of species (Peterson, Papeş and Soberón, 2015): Mechanistic models which are based on physiological characteristics measured from a few individuals, correlative approaches, which are the most popular approaches for niche and species distribution modelling and associate occurrence points with environmental factors to make predictions, and process based approaches, which can estimate occupied areas taking into account dispersal capabilities of species.

There are many modelling algorithms being used for correlative methods. A modelling algorithm is the “procedure, rule, or mathematical function used to estimate the species’ ecological niche as a function of a suite of environmental variables” (Peterson *et al.*, 2011). It is convenient to classify these according to their required occurrence input, there are presence–absence approaches, and presence only approaches (Franklin, 2012; Peterson *et al.*, 2011). Presence-absence approaches require data about the occurrence of a species in an area and data about its true absence which may require high sampling effort to achieve, while presence only approaches deal with the lack of this data in their algorithms in different ways (Franklin, 2012).

The advantage of using presence-absence approaches is that different kind of niches and distribution can be estimated more accurately (Franklin, 2012), nevertheless, acquiring absence data requires a significant effort to systematically survey a region, consequently, such data is usually not available in museum and herbaria for most species (Elith *et al.*, 2011; Raedig, 2010). Modelling using presence-only approaches helps avoid some of the disadvantages of presence-absence approaches, but it also has its own disadvantages like the fact that presence-only data tends to lack information about time and scale and it also tends to be subjected to sampling bias (Elith *et al.*, 2011). One of the most used software for presence-only modelling is Maxent, which applies the Maximum Entropy Modelling technique, a statistical and machine learning method. This algorithm takes presence-only disadvantages into account (Elith *et al.*, 2011; Franklin, 2012; Merow, Smith and Silander, 2013; Phillips, Anderson and Schapire, 2006). An additional advantage of the Maxent algorithm is that it has shown to produce accurate models using low sample sizes compared to other algorithms (Wisz *et al.*, 2008).

1.2.3 *Model calibration and evaluation*

Regardless of the modelling method type chosen, models must be calibrated and evaluated, often using different sets of data (Peterson *et al.*, 2011). Calibration refers to the use of a dataset for the estimation of model internal parameters (like one would estimate the coefficients of a regression model). To evaluate the predictions of the calibrated model, another dataset is often used. Moreover, validation refers to the selection of models that are useful for the intended purposes through evaluation measures (Franklin, 2012; Peterson *et al.*, 2011).

1.2.4 *Niche modelling of diseases and risk mapping*

Traditional ecological niche modelling frameworks have been shown appropriate for predicting distributions of free-living organisms. Nevertheless, certain conceptual considerations must be taken when modelling distributions of disease causing organisms. Peterson (2008), noted that the focus of traditional coarse-scale niche modelling is on abiotic conditions while undermining the effect of biotic interactions. However, biotic interactions are much relevant for disease causing parasites, and these must be considered to accurately predict disease niches and distributions.

There are two approaches for modelling disease distributions, black box approaches and component-based approaches (Johnson, Escobar and Zambrana-Torrel, 2019; Peterson, 2014): Black box approaches consider only the final result of the interactions of host and parasites into model calibration, that is, occurrences of the disease in humans or animals. On the other hand, component-based approaches consider the species in the transmission system behind the registered disease cases.

Johnson, Escobar and Zambrana-Torrel (2019), published a framework that takes into account previous conceptual considerations and approaches. Here, biotic interactions have an integral part in the modelling. Other interesting emerging approaches include the use of community ecology to manage parasite populations in nature taking into account intraspecific interactions (Johnson, Roode and Fenton, 2015), and approaches that differentiate between the hazard and risk concepts, and evaluate how biodiversity and human populations affect both (Hagenlocher *et al.*, 2014; Hammer *et al.*, 2019; Hosseini *et al.*, 2017).

1.3 Problem statement

Since it was discovered and recognized as native, Chagas disease is now better understood in Mexico, with more cases being reported every year due to better diagnostic methods and epidemiological surveillance (Cruz-Reyes and Pickering-López, 2006; Ibáñez-Cervantes *et al.*, 2018). However, Chagas disease is a disease difficult to diagnose since it doesn't always show symptoms, and when it does, they can be confused with other diseases (Pérez-Molina and Molina, 2018). Thus, Chagas disease is still an important disease in Mexico that still causes disability and loss of human lives. Additionally, the number of reported Chagas disease cases in Mexico have been underestimated, as well as in some states like San Luis Potosi (Arnal *et al.*, 2019). This is important because the Mexican health authorities rely on predictions of outbreaks of Chagas disease to acquire medicine to treat the disease in all its phases (Manne *et al.*, 2013). Underestimating the transmission of Chagas disease in the state of San Luis Potosí could lead to unavailability of treatment and loss of lives.

Additionally, in San Luis Potosí, Chagas disease has been studied the most in Huasteca region (Juarez-Tobias *et al.*, 2009; Medina-Garza *et al.*, 2016), but few studies exist in other regions, even though previous coarse-scale distribution models of triatomines show presences and predict suitable habitat for three species of triatomines all over the state of San Luis Potosi (Ramsey *et al.*, 2015).

Therefore, Chagas disease may be underreported in San Luis Potosi and there is a need to find currently undiagnosed cases. A risk map based on distribution models of vectors and reservoir species of the disease could better predict potential distribution of the disease in the state and to discover populations at risk in previously unsampled areas.

3 Objective and research questions

3.1 General objective

The objective of this study is to identify risk areas for the transmission of Chagas disease in San Luis Potosí using species distribution modelling of vectors and reservoirs, and to use anthropogenic environmental data as predictor variables into the risk maps.

3.2 Research questions

- Which are the relevant vector and reservoir species for Chagas disease transmission in San Luis Potosí?
- What is the potential distribution area of relevant vectors and reservoirs of *T. cruzi* in Media and Huasteca regions in San Luis Potosí?
- Is there available anthropogenic environmental data which could be included into species distribution modelling at municipal scale as predictor variables?
- Which areas have the highest risk for Chagas disease infection of human populations in San Luis Potosí?

4 Methods

4.1 Study area

The study area was the Mexican state of San Luis Potosi. The state is often culturally divided in four parts, these are, Altiplano, Centro, Media and Huasteca (Figure 3).

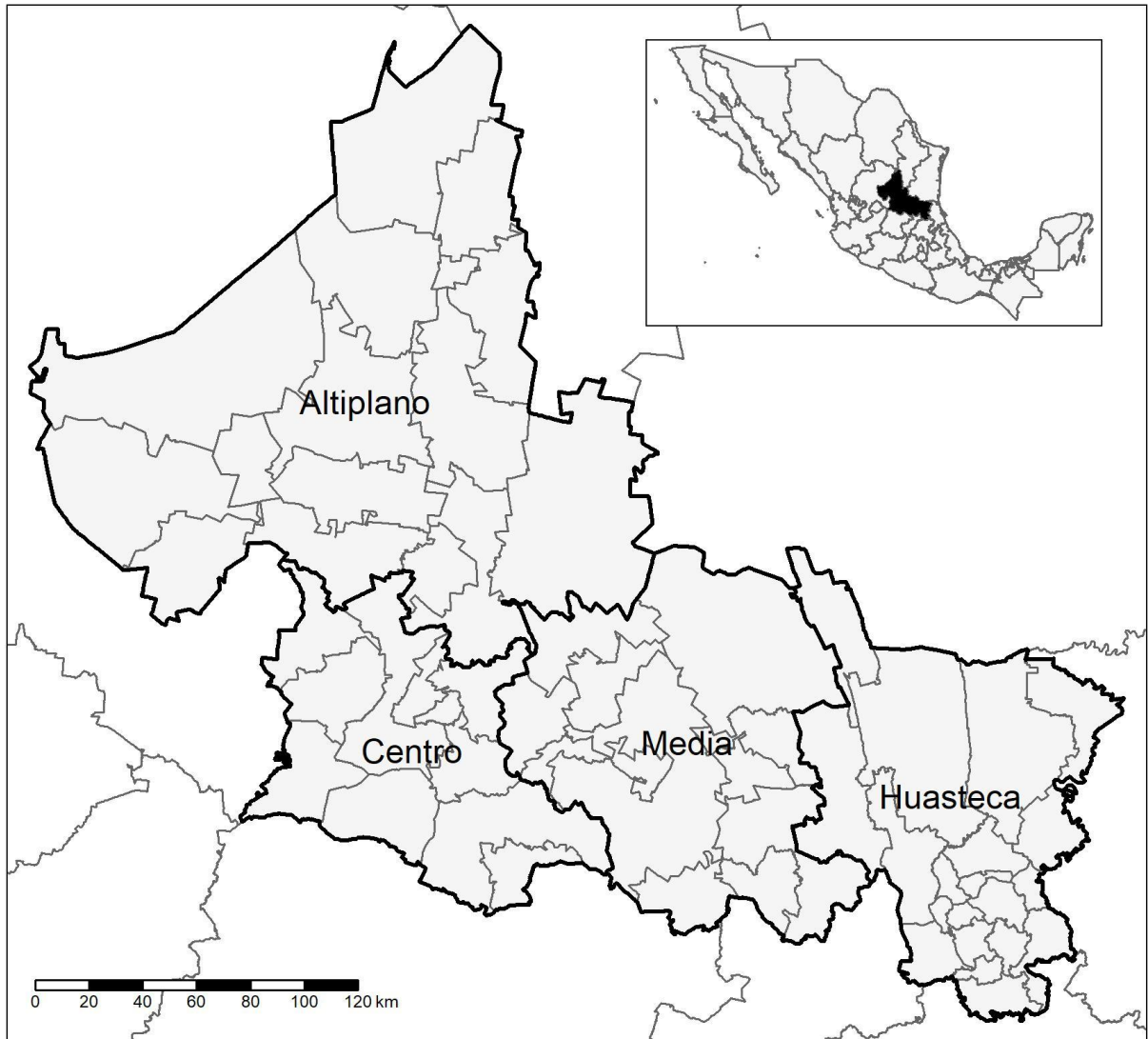


Figure 3: Map of the state of San Luis Potosi divided into its four regions.

San Luis Potosi has an annual precipitation of 250 and 500 mm in its driest sections and more than 2000 mm in its most humid parts. The state has a combination of tropical, dry, and temperate climates all over its territory. Approximately 70% of the state is in some degree arid, with the most tropical and wet climates on the south-east side of the state. The climate is modified by the different topographic features present, being the most important the Sierra Madre Oriental mountain range (Rezedowski, 1965). Climate maps based on the Köppen classification (Peel, Finlayson and McMahon, 2007) system and the modified version of García (1998) for Mexico are shown in Figure 4.

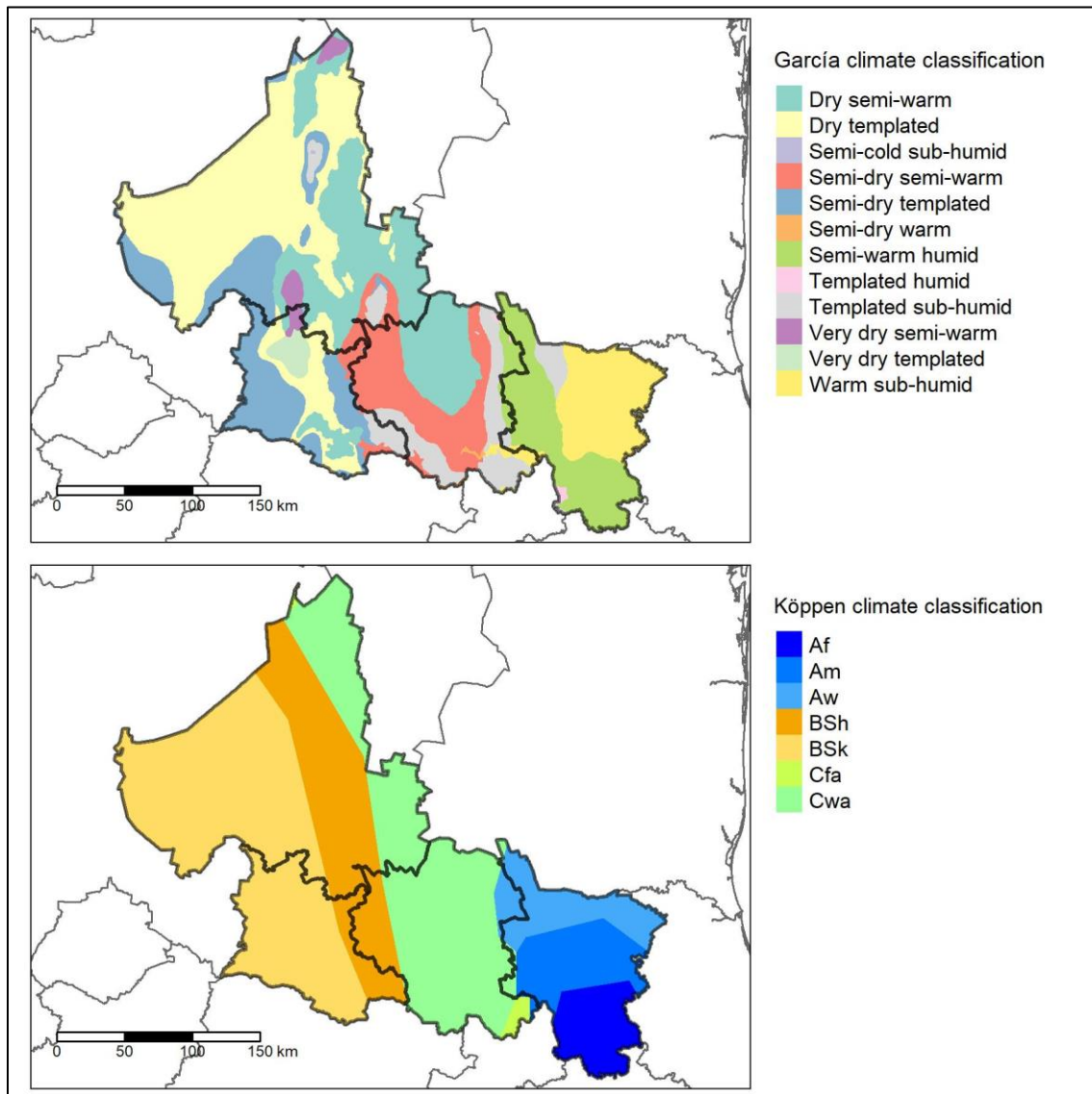


Figure 4: Climates of San Luis Potosi. On top, climate map from INEGI based on the modified Köppen classification by García (1998), on bottom, Köppen climate map updated by Peel et al. (2007).

4.2 Methods overview

The methods followed to achieve the objectives are divided in three parts. Firstly, important vector and reservoir species in San Luis Potosi were chosen. Secondly, species distribution models were made for the selected vector and reservoir species. Lastly, binary predictions were generated for vectors and reservoir and overlaid to identify risk areas. An overview of the methods used is shown in Figure 5.

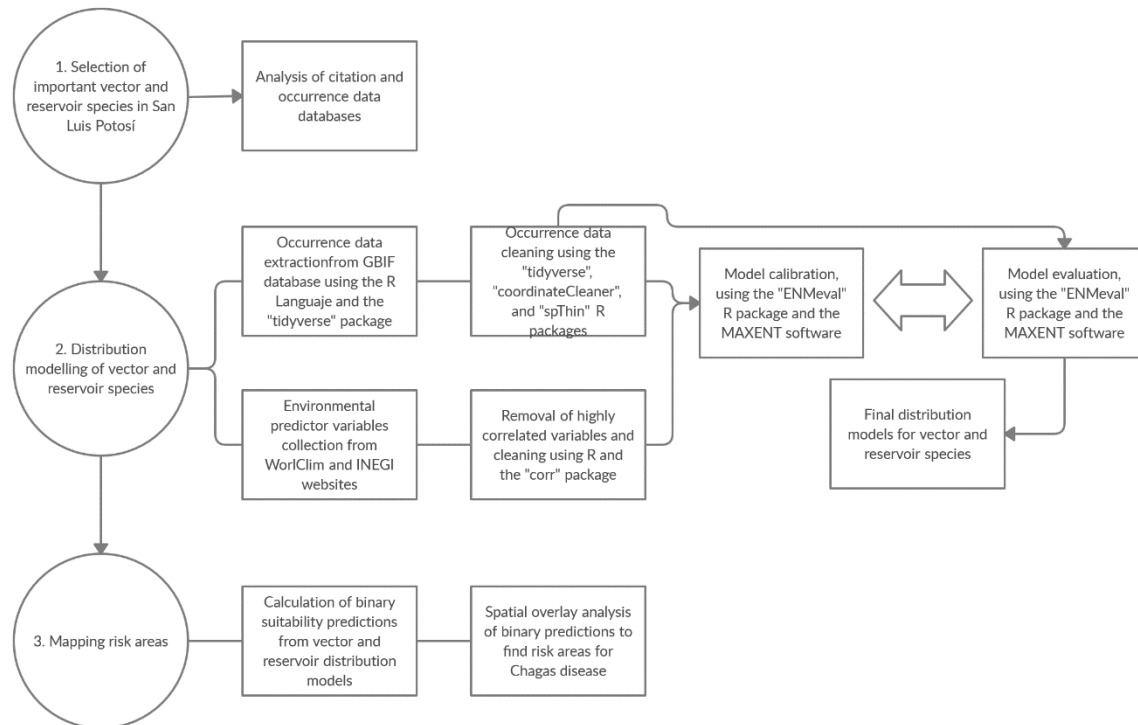


Figure 5: Graphic overview of methods used.

4.3 Selection of important vector and reservoir species

4.3.1 *Vector species*

For the selection of important triatomine species, previous publications and databases that contained occurrence data were searched for. The selected sources for triatomine data were the study from Ramsey *et al.* (2015), and the Global Biodiversity Information Facility (GBIF) database. Only species with more than 50 presence points in San Luis Potosi were considered as important triatomine species in the state, thus one species, *Triatoma barberi* Usinger 1939, was dropped. The triatomine species *Triatoma dimidiata* Latreille 1811, *Triatoma gerstaeckeri* Stal 1859 and *Triatoma Mexicana* Herrich-Schaeffer 1848 were chosen as important vectors because there was enough data from these species and they had been previously studied in the state of San Luis Potosí (Medina-Garza *et al.*, 2016; Ramsey *et al.*, 2015).

4.3.2 *Reservoir species*

To select the important reservoir species, a spreadsheet with relevant information about reservoirs of *T. cruzi* in Mexico was made. The spreadsheet was filled using data from articles about Chagas disease reservoirs found in the citation database Scopus. Articles which objective was only to seek the parasite in dogs or cats were not included, otherwise, every article found reporting an animal's infection with *T. cruzi* was included. To identify probable reservoir species in San Luis Potosí, the datasheet was compared with a checklist of mammals of San Luis Potosí (Martínez de la Vega *et al.*, 2016). Animals present in San Luis Potosí were compared in importance by their reported infection and number of times reported in the recollected literature. Only species with more than 30 reports were included.

4.4 **Distribution modelling of vector and reservoir species**

4.4.1 *Species niche description*

A niche description was made for each vector species to model. Papers related to the ecology and behavior of each species were searched for in the citation database Scopus. The search terms were the scientific names of each species and all results were read. The information was collected and organized in a table.

4.4.2 *Species occurrence data collection and cleaning*

Species Occurrence data was downloaded from the Global Biodiversity Information Facility (GBIF) online database for *Triatoma dimidiata* (GBIF.org, 2020b), *Triatoma gerstaeckeri* (GBIF.org, 2020a), *Triatoma Mexicana* (GBIF.org, 2020c), and *Didelphis virginiana* (GBIF.org, 2020d). Additionally, data used in Ramsey *et al.* (2015) for triatomine distribution modelling was also obtained online. The data was explored and manipulated in R 3.6.3 using the “tidyverse” package. Occurrence data was explored and selected using the decimal coordinates, year, and institution information available in the GBIF files.

Firstly, all records with missing longitude or latitude coordinates were removed. Additionally, the data was cropped using a layer of the states of Mexico to remove occurrences outside of the country. The dataset from (Ramsey *et al.*, 2015), was already contained in the GBIF database, so the GBIF dataset was mainly used. After that, the data was filtered by responsible institution. Only research centers and museums were accepted.

Occurrences were checked using the R “coordinateCleaner” package (Zizka *et al.*, 2018), which checks for common mistakes in species occurrence database records. Then, as an additional filter the institution information from GBIF was checked and only records coming from research or government institutions were used for the triatomine occurrences, however, data from iNaturalist was also used for the reservoir species occurrences. Only occurrences between the years 1995 and 2015 were used since most of the occurrences were from that time interval.

Lastly, a spatial thinning was applied to the occurrence data using the R package "spThin" (Aiello-Lammens *et al.*, 2015). Occurrences were filtered according to the distance to their nearest neighbor. For triatomine and opossum data, distances of 1000 m were used with the objective of reducing observations too near of each other and avoid errors due to spatial autocorrelation.

4.4.3 *Environmental predictor variables collection and cleaning*

Historical climate data (years 1970 to 2000) were downloaded from the WorldClim website (Fick and Hijmans, 2017). Multicollinearity was dealt with by dropping variables using a Pearson correlation matrix as described in Dormann *et al.* (2013). The correlation matrix was done using R 3.6.3 and the "raster" and "corr" packages. All the variables were gathered into a raster stack and then transformed into a data frame. A correlation matrix was calculated from the data frame and all variable pairs with correlation coefficients higher than 0.7 or lower than -0.7 were extracted (Annex 1). Variables derived from other variables were eliminated first, then, the least relevant and most correlated variables for species niche were dropped. The remaining uncorrelated variables were Max Temperature of Warmest Month (BIO 5), Min Temperature of Coldest Month (BIO 6), Precipitation Seasonality (BIO 15), Precipitation of Warmest Quarter (BIO 18) and Precipitation of Coldest Quarter (BIO 19). The elevation dataset from the WorldClim variables was also included and considered to have low collinearity with the other variables (Raedig, Ho Dac Thai and Nehren, 2017).

Additionally, the land use map "Carta de Uso del Suelo y Vegetación escala 1:250 000 Serie III" published in 2005 by INEGI was downloaded from their website. The map was generated by INEGI using LandSat images of 25 and 30 m of resolution. The map focus on vegetation types but includes information about human settlements and urban areas in the metadata. A new attribute was created in R from the metadata, to include the categories of "Areas devoid of vegetation", "Human settlements", "Water bodies" and "Urban zones".

4.4.4 *Modelling*

The species distribution modelling was performed using the Maxent algorithm implemented in the software by Phillips, Dudík and Schapire (2017). To determine the input parameters that produced the best results, many models were trained and evaluated in R using the package "ENMeval" (Muscarella *et al.*, 2014) as interface for the maxent software. This package automatizes many common tasks of model training and evaluation and organizes the results in an easy to compare way.

Model training and evaluation was done taking into account considerations described by Merow, Smith and Silander (2013), and as described by Radosavljevic and Anderson (2014). A k-fold cross validation approach (Franklin, 2012; Peterson *et al.*, 2011) was used for model evaluation. The parameters used can be found in Table 1. The best models were chosen using the Akaike information criterion (AIC) which penalizes for model overfitting and number of

parameters. Lowest values of AIC indicate better, less overfitted models (Warren and Seifert, 2011). The best model's significance was assessed using the area under the curve (AUC) statistic, which can take values between 0.5 and 1. Values of 0.5 for AUC indicate models no better than a random prediction while values of 1 indicate robust models (Franklin, 2012; Peterson *et al.*, 2011). The performance was assessed using the Minimum Training Presence Test Omission Rate (MTPOR) and the 10 Percentile Training Presence Test Omission Rate (10PTPOR), where values of 0 and 0.10 are expected; higher values than those indicate model overfitting, or, omission error (Peterson *et al.*, 2011).

Table 1: list of parameters used in model training. Models varied between the combination of regularization multipliers (nine possibilities) and features used (either LQH or LQHP).

Maxent parameter	Values
Fixed values	
Number of k-folds for evaluation	<i>Triatoma dimidiata</i> : 10 <i>Triatoma gerstaeckeri</i> : 5 <i>Triatoma mexicana</i> : 10 Reservoir species: 10
Background points	10,000 taken randomly from each species' known range in Mexico.
Parameter combinations evaluated per species	
Regularization multiplier	0.25, 1, 3, 5, 6, 7, 8, 9, 10
Features allowed: linear (L), quadratic (Q), product (P), hinge (H)**	Features combinations modelled LQH and LQHP

4.5 Mapping risk areas

Binary species distribution predictions were generated by classifying the continuous model raster outputs into present and absent (represented by 1 and 0 respectively in the raster). The classification was done by using the Maximum test Sensitivity Plus Specificity (MTSPS) which is the value at which sensibility and specificity are maximized (Liu, White and Newell, 2013). Afterwards, areas of overlap between binary species distribution predictions were identified and quantified through an overlay spatial analysis in ArcGIS 10.4.1. The human population figure residing in the overlapping areas was calculated using shapefiles from the 2010 Mexican census data provided by INEGI.

5 Results

5.1 Selection of important vector and reservoir species

5.1.1 Important reservoir species identification

A total of 58 mammal individuals have been reported to be infected with *T. cruzi* in 16 different studies from 1997 to 2018 in Mexico (Table 2). These individuals belong to 7 orders (Figure 6) and 16 families. In total of 38 different species have been reported (Table 3).

The studies were from seven different Mexican states (Figure 7), of which Yucatan and Morelos where the states with the most reports. No reports of reservoirs were found from the state of San Luis Potosí, however, 23 of the 38 reservoir species are known to habit San Luis Potosí (Table 3). To compare importance of the reservoirs, the reported infection rates and number of times seen in the literature were considered. *Rattus rattus*, *Mus musculus* and *Didelphis virginiana* were selected as the most important reservoirs present in San Luis Potosí (Figure 8).

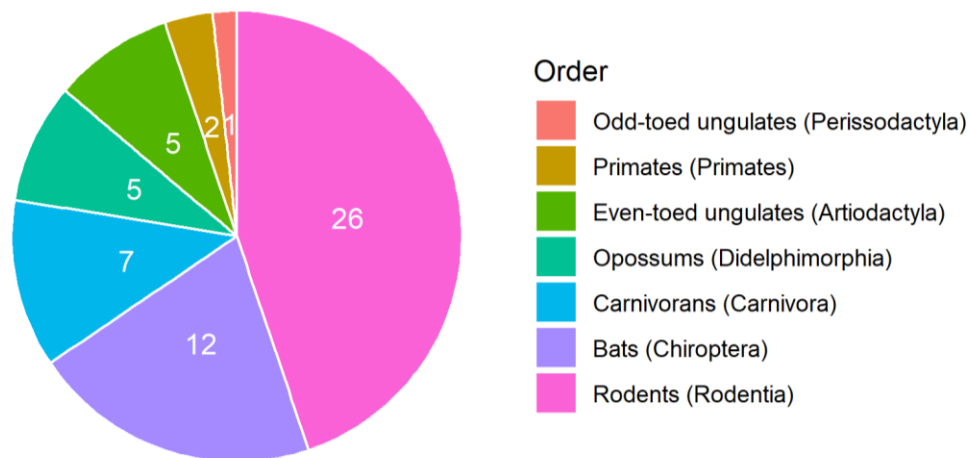


Figure 6: Mammal Individuals reported infected with *T. cruzi* by order. The labels inside the sections of the pie chart represent total of individuals.

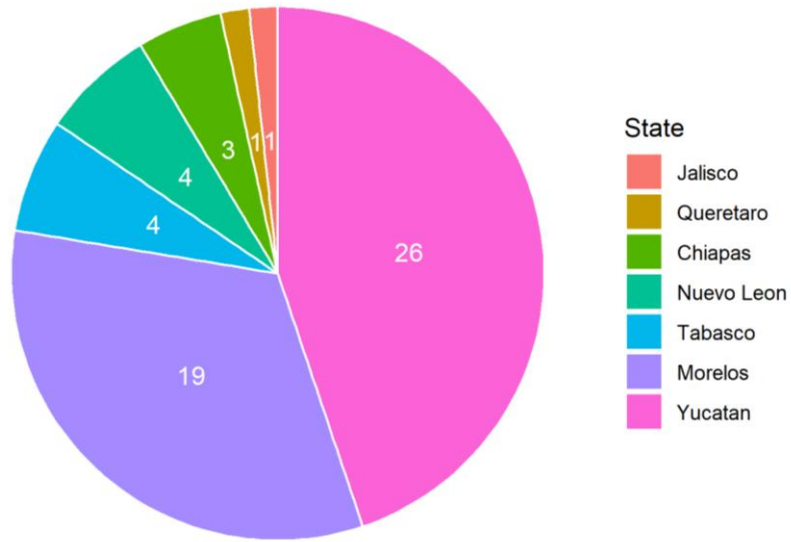


Figure 7: Mammal Individuals reported infected with *T. cruzi* by 7 states of the 32 states of Mexico. The labels inside the sections represent total of individuals.

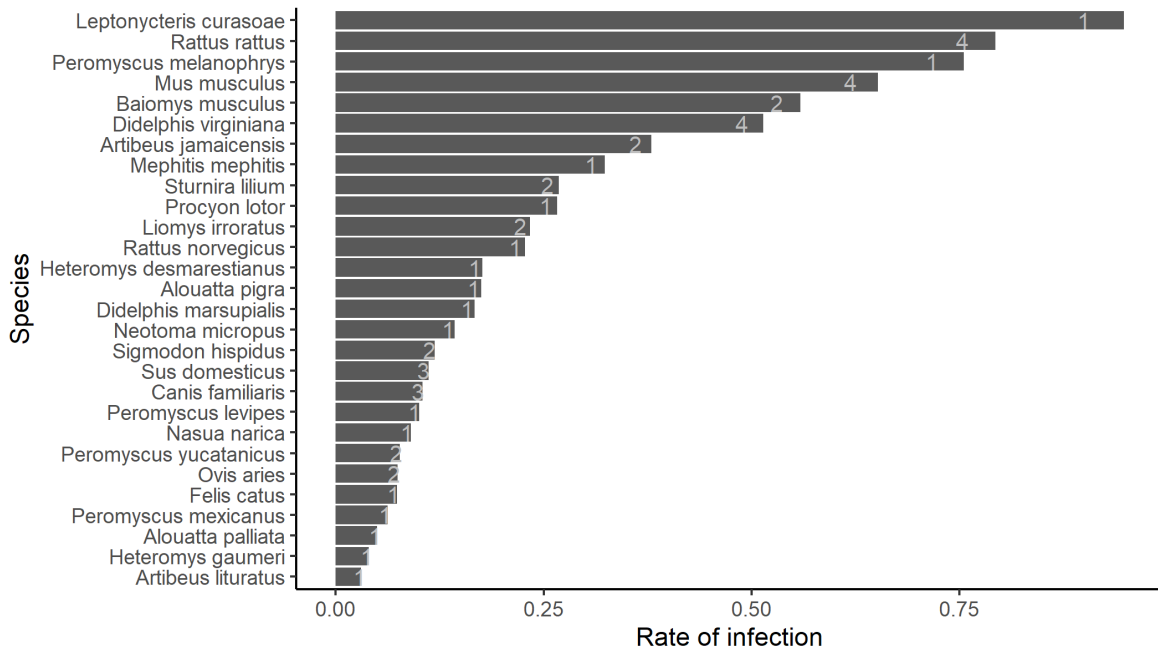


Figure 8: Rate of reported infections with *T. cruzi* by species. The label at the end of each bar represents the number of papers that reported the species.

Table 2: List of mammal individuals found infected with *T. cruzi* in studies about reservoirs of Chagas disease in Mexico.

Order	Family	Species	State	Seroprevalence ¹	Capture habitat ²	Reference
Artiodactyla	Bovidae	<i>Ovis aries</i>	Yucatán	6/7 (86%)	p	(Ruiz-Piña <i>et al.</i> , 2018b)
Artiodactyla	Bovidae	<i>Ovis aries</i>	Yucatán	6/155 (4%)	d/p	(López-Cancino <i>et al.</i> , 2015)
Artiodactyla	Suidae	<i>Sus domesticus</i>	Yucatán	15/273 (5%)	p	(Jiménez-Coello <i>et al.</i> , 2012)
Artiodactyla	Suidae	<i>Sus domesticus</i>	Yucatán	21/28 (75%)	p	(Ruiz-Piña <i>et al.</i> , 2018b)
Artiodactyla	Suidae	<i>Sus domesticus</i>	Yucatán	1/31 (3%)	d	(López-Cancino <i>et al.</i> , 2015)
Carnivora	Canidae	<i>Canis familiaris</i>	Querétaro	10/14 (71%)	d	(Zamora-Ledesma <i>et al.</i> , 2016)
Carnivora	Canidae	<i>Canis familiaris</i>	Nuevo León	13/136 (10%)	d	(Galaviz-Silva <i>et al.</i> , 2017)
Carnivora	Canidae	<i>Canis familiaris</i>	Yucatán	11/176 (6%)	d/p	(López-Cancino <i>et al.</i> , 2015)
Carnivora	Felidae	<i>Felis catus</i>	Yucatán	7/95 (7%)	p	(Longoni <i>et al.</i> , 2012)
Carnivora	Procyonidae	<i>Nasua narica</i>	Tabasco	19/210 (9%)	s	(Martínez-Hernández <i>et al.</i> , 2014)
Carnivora	Mephitidae	<i>Mephitis mephitis</i>	Nuevo León	11/34 (32%)	s	(Galaviz-Silva <i>et al.</i> , 2017)
Carnivora	Procyonidae	<i>Procyon lotor</i>	Tabasco	29/109 (27%)	s	(Martínez-Hernández <i>et al.</i> , 2014)
Chiroptera	Phyllostomidae	<i>Artibeus jamaicensis</i>	Yucatán	2/62 (3%)	p/s	(López-Cancino <i>et al.</i> , 2015)
Chiroptera	Phyllostomidae	<i>Artibeus jamaicensis</i>	Morelos	39/46 (85%)	s	(Villegas-García and Santillán-Alarcón, 2004)
Chiroptera	Phyllostomidae	<i>Carollia brevicauda</i>	Yucatán	2/3 (67%)	s	(López-Cancino <i>et al.</i> , 2015)
Chiroptera	Phyllostomidae	<i>Desmodus rotundus</i>	Morelos	9/12 (75%)	s	(Villegas-García and Santillán-Alarcón, 2004)
Chiroptera	Phyllostomidae	<i>Leptonycteris curasoae</i>	Morelos	36/38 (95%)	s	(Villegas-García and Santillán-Alarcón, 2004)
Chiroptera	Vespertilionidae	<i>Myotis keaysi</i>	Yucatán	1/7 (14%)	p/s	(López-Cancino <i>et al.</i> , 2015)
Chiroptera	Mormoopidae	<i>Pteronotus parnellii</i>	Morelos	10/13 (77%)	s	(Villegas-García and Santillán-Alarcón, 2004)
Chiroptera	Phyllostomidae	<i>Sturnira ludovici</i>	Yucatán	1/18 (6%)	d/p/s	(López-Cancino <i>et al.</i> , 2015)

Chiroptera	Phyllostomidae	<i>Artibeus lituratus</i>	Yucatán	1/33 (3%)	d	(López-Cancino <i>et al.</i> , 2015)
Chiroptera	Phyllostomidae	<i>Dermanura phaeotis</i>	Yucatán	1/10 (10%)	d/p/s	(López-Cancino <i>et al.</i> , 2015)
Chiroptera	Phyllostomidae	<i>Sturnira lilium</i>	Yucatán	3/51 (6%)	d/p/s	(López-Cancino <i>et al.</i> , 2015)
Chiroptera	Phyllostomidae	<i>Sturnira lilium</i>	Morelos	19/31 (61%)	s	(Villegas-García and Santillán-Alarcón, 2004)
Didelphimorphia	Didelphidae	<i>Didelphis marsupialis</i>	Nuevo León	6/36 (17%)	s	(Galaviz-Silva <i>et al.</i> , 2017)
Didelphimorphia	Didelphidae	<i>Didelphis virginiana</i>	Chiapas	2/12 (17%)	s	(Solís-Franco, Romo-Zapata and Martínez-Ibarra, 1997)
Didelphimorphia	Didelphidae	<i>Didelphis virginiana</i>	Yucatán	55/102 (54%)	d/p	(Ruiz-Piña and Cruz-Reyes, 2002)
Didelphimorphia	Didelphidae	<i>Didelphis virginiana</i>	Yucatán	3/9 (33%)	d/s	(López-Cancino <i>et al.</i> , 2015)
Didelphimorphia	Didelphidae	<i>Didelphis virginiana</i>	Morelos	13/19 (68%)	s	(Villegas-García and Santillán-Alarcón, 2004)
Perissodactyla	Equidae	<i>Equus ferus</i>	Yucatán	8/8 (100%)	p	(Ruiz-Piña <i>et al.</i> , 2018b)
Primates	Atelidae	<i>Alouatta palliata</i>	Tabasco	2/40 (5%)	s	(Roviroso-Hernández <i>et al.</i> , 2013)
Primates	Atelidae	<i>Alouatta pigra</i>	Tabasco	7/40 (18%)	s	(Roviroso-Hernández <i>et al.</i> , 2013)
Rodentia	Cricetidae	<i>Baiomys musculus</i>	Morelos	8/30 (27%)	d/p/s	(Ramsey <i>et al.</i> , 2012)
Rodentia	Cricetidae	<i>Baiomys musculus</i>	Morelos	35/47 (74%)	s	(Villegas-García and Santillán-Alarcón, 2004)
Rodentia	Heteromyidae	<i>Liomys irroratus</i>	Morelos	4/62 (6%)	p/s	(Ramsey <i>et al.</i> , 2012)
Rodentia	Heteromyidae	<i>Liomys irroratus</i>	Morelos	14/15 (93%)	s	(Villegas-García and Santillán-Alarcón, 2004)
Rodentia	Muridae	<i>Mus musculus</i>	Yucatán	233/302 (77%)	d	(Panti-May <i>et al.</i> , 2017)
Rodentia	Muridae	<i>Mus musculus</i>	Morelos	4/65 (6%)	d/p	(Ramsey <i>et al.</i> , 2012)
Rodentia	Muridae	<i>Mus musculus</i>	Yucatán	1/9 (11%)	d/p	(López-Cancino <i>et al.</i> , 2015)
Rodentia	Muridae	<i>Mus musculus</i>	Morelos	26/29 (90%)	p	(Villegas-García and Santillán-Alarcón, 2004)
Rodentia	Cricetidae	<i>Neotoma mexicana</i>	Morelos	5/19 (26%)	s	(Ramsey <i>et al.</i> , 2012)
Rodentia	Cricetidae	<i>Neotoma micropus</i>	Nuevo León	5/35 (14%)	s	(Galaviz-Silva <i>et al.</i> , 2017)
Rodentia	Cricetidae	<i>Peromyscus levipes</i>	Morelos	4/40 (10%)	d/p/s	(Ramsey <i>et al.</i> , 2012)

Rodentia	Cricetidae	<i>Peromyscus mexicanus</i>	Chiapas	5/80 (6%)	s	(Solís-Franco, Romo-Zapata and Martínez-Ibarra, 1997)
Rodentia	Cricetidae	<i>Sigmodon hispidus</i>	Morelos	8/91 (9%)	d/p/s	(Ramsey <i>et al.</i> , 2012)
Rodentia	Cricetidae	<i>Sigmodon hispidus</i>	Yucatán	9/52 (17%)	d/p	(López-Cancino <i>et al.</i> , 2015)
Rodentia	Heteromyidae	<i>Heteromys desmarestianus</i>	Chiapas	6/34 (18%)	s	(Solís-Franco, Romo-Zapata and Martínez-Ibarra, 1997)
Rodentia	Heteromyidae	<i>Heteromys gaumeri</i>	Yucatán	3/76 (4%)	d/p/s	(López-Cancino <i>et al.</i> , 2015)
Rodentia	Sciuridae	<i>Otospermophilus variegatus</i>	Morelos	12/12 (100%)	p	(Villegas-García and Santillán-Alarcón, 2004)
Rodentia	Cricetidae	<i>Ototylomys phyllotis</i>	Yucatán	3/7 (43%)	s	(Hernández-Cortazar <i>et al.</i> , 2018)
Rodentia	Cricetidae	<i>Peromyscus melanophrys</i>	Morelos	37/49 (76%)	p/s	(Villegas-García and Santillán-Alarcón, 2004)
Rodentia	Cricetidae	<i>Peromyscus yucatanicus</i>	Yucatán	2/47 (4%)	p/s	(López-Cancino <i>et al.</i> , 2015)
Rodentia	Cricetidae	<i>Peromyscus yucatanicus</i>	Yucatán	2/5 (40%)	s	(Hernández-Cortazar <i>et al.</i> , 2018)
Rodentia	Muridae	<i>Rattus norvegicus</i>	Jalisco	71/312 (23%)	d/p	(Rosal <i>et al.</i> , 2018)
Rodentia	Muridae	<i>Rattus rattus</i>	Yucatán	145/161 (90%)	d	(Panti-May <i>et al.</i> , 2017)
Rodentia	Muridae	<i>Rattus rattus</i>	Morelos	2/7 (29%)	d/p/s	(Ramsey <i>et al.</i> , 2012)
Rodentia	Muridae	<i>Rattus rattus</i>	Morelos	19/22 (86%)	p	(Villegas-García and Santillán-Alarcón, 2004)
Rodentia	Muridae	<i>Rattus rattus</i>	Yucatán	3/23 (13%)	d/p	(Hernández-Cortazar <i>et al.</i> , 2018)

1. Calculated as reported infection and total captures in the referenced article.
2. The following abbreviations are used: d: domicile, p: peridomicile, s: sylvatic

Table 3: List of species reported to be infected by *T. cruzi* in the literature in Mexico, with reporting status in San Luis Potosí.

Order	Family	Species	Reports in SLP
Artiodactyla	Bovidae	<i>Ovis aries</i>	Reported
Artiodactyla	Suidae	<i>Sus domesticus</i>	Reported
Carnivora	Canidae	<i>Canis familiaris</i>	Reported
Carnivora	Felidae	<i>Felis catus</i>	Reported
Carnivora	Procyonidae	<i>Nasua narica</i>	Reported
Chiroptera	Mormoopidae	<i>Pteronotus parnellii</i>	Reported
Chiroptera	Phyllostomidae	<i>Artibeus jamaicensis</i>	Reported
Chiroptera	Phyllostomidae	<i>Carollia brevicauda</i>	Reported
Chiroptera	Phyllostomidae	<i>Desmodus rotundus</i>	Reported
Chiroptera	Phyllostomidae	<i>Leptonycteris curasoae</i>	Reported
Chiroptera	Phyllostomidae	<i>Sturnira ludovici</i>	Reported
Chiroptera	Vespertilionidae	<i>Myotis keaysi</i>	Reported
Didelphimorphia	Didelphidae	<i>Didelphis marsupialis</i>	Reported
Didelphimorphia	Didelphidae	<i>Didelphis virginiana</i>	Reported
Perissodactyla	Equidae	<i>Equus ferus</i>	Reported
Rodentia	Cricetidae	<i>Baiomys musculus</i>	Reported
Rodentia	Cricetidae	<i>Neotoma mexicana</i>	Reported
Rodentia	Cricetidae	<i>Neotoma micropus</i>	Reported
Rodentia	Cricetidae	<i>Peromyscus levipes</i>	Reported
Rodentia	Cricetidae	<i>Peromyscus mexicanus</i>	Reported
Rodentia	Cricetidae	<i>Sigmodon hispidus</i>	Reported
Rodentia	Heteromyidae	<i>Liomys irroratus</i>	Reported
Rodentia	Muridae	<i>Mus musculus</i>	Reported
Carnivora	Mephitidae	<i>Mephitis mephitis</i>	Not Reported
Carnivora	Procyonidae	<i>Procyon lotor</i>	Not Reported
Chiroptera	Phyllostomidae	<i>Artibeus lituratus</i>	Not Reported
Chiroptera	Phyllostomidae	<i>Dermanura phaeotis</i>	Not Reported
Chiroptera	Phyllostomidae	<i>Sturnira lilium</i>	Not Reported
Primates	Atelidae	<i>Alouatta palliata</i>	Not Reported
Primates	Atelidae	<i>Alouatta pigra</i>	Not Reported
Rodentia	Cricetidae	<i>Otodylomys phyllotis</i>	Not Reported
Rodentia	Cricetidae	<i>Peromyscus melanophrys</i>	Not Reported
Rodentia	Cricetidae	<i>Peromyscus yucatanicus</i>	Not Reported
Rodentia	Heteromyidae	<i>Heteromys desmarestianus</i>	Not Reported
Rodentia	Heteromyidae	<i>Heteromys gaumeri</i>	Not Reported
Rodentia	Muridae	<i>Rattus norvegicus</i>	Not Reported
Rodentia	Muridae	<i>Rattus rattus</i>	Not Reported
Rodentia	Sciuridae	<i>Otospermophilus variegatus</i>	Not Reported

5.2 Distribution modelling of vector and reservoir species

5.2.1 *Niche description of species to model*

The niche of *Triatoma dimidiata* is described in Table 4. Temperature and seasonality seem to affect *T. dimidiata* directly by influencing the duration of its life cycle and its abundance. It uses the burrows of other animals, like opossums as shelters, however, it can also find shelter inside human houses and feed from domestic animals inside. It has also been found where cattle or poultry sleep.

Table 4: Niche description of *Triatoma dimidiata*

Related variable	Statement	Reference
Life cycle	To complete its life cycle, <i>T. dimidiata</i> requires blood meals from vertebrates between molts. Life cycle lasts approximately one year and decreases when temperature is increased. The immature forms of these bugs (nymphs) can survive long periods of starvation.	(Zeledón <i>et al.</i> , 1970)
Outdoor shelters	It has been found in holes on trees, or nests inhabited by animals, specially opossums, they have been found near houses. Bugs that had fed from opossum where frequently infected with <i>T. cruzi</i> .	(Zeledón <i>et al.</i> , 1973)
Indoor shelters	Nymphs can use debris from the soil floor to camouflage. Around houses, bugs where found hiding in piles of objects like firewood, lumber, tiles, stones, sugar cane, peelings, and corn. There appears to be movement of nymphs and adults in and out dwellings because of the diverse bloodmeals found inside, especially on nymphs.	(Zeledón <i>et al.</i> , 1973)
Blood sources	The bugs bit more often dogs and rats inside households, and chickens and opossum outside.	(Zeledón <i>et al.</i> , 1973)
Seasonal activity patterns	<i>T. dimidiata</i> is more frequently found inside households in dry season, than in rainy season. This might be due to migration from outside.	(DUMONTEIL <i>et al.</i> , 2002)
Climate Land use	Some individuals were found in a dry climate on an extensive pastureland.	(Villagrán <i>et al.</i> , 2008)
Outdoor shelters	Houses located near public light poles are at higher risk for being infested by <i>T. dimidiata</i> .	(Pacheco-Tucuch <i>et al.</i> , 2012)
Indoor shelters Outdoor shelters	House made of bahareque are at higher risk of infestation	(Parra-Henao <i>et al.</i> , 2015)
Outdoor shelters	For peridomestic infestation, the presence of dogs, cats, goats and cows was important.	(Parra-Henao <i>et al.</i> , 2015)
Outdoor shelter	The presence of a chicken coops in the backyard was associated with Infestation with <i>T. dimidiata</i> .	(Koyoc-Cardeña <i>et al.</i> , 2015)

Generalities	They can live in natural and artificial habitats. Depend on shelter availability and nearby food sources, dogs and chickens are the most frequent. They are attracted to artificial light.	(Quirós-Gómez <i>et al.</i> , 2017)
Indoor shelters	Other factors associated with infestation are soil floors, and houses with more than 3 rooms,	(Quirós-Gómez <i>et al.</i> , 2017)
Outdoor shelters	chicken coops and streetlights are important factors for peridomicile infestation.	(Quirós-Gómez <i>et al.</i> , 2017)
Blood sources	From 12 blood meal sources identified, the most important are human, dog and cow. Other sources are cats, pigs and poultry for domestic animals, and mouse, dove and bats for synanthropic animals.	(Moo-Millan <i>et al.</i> , 2019)

The niche of *Triatoma gerstaeckeri* is described in Table 5. It has been shown to be directly affected by temperature and seasonality. Additionally, it is associated to woodrats and opuntia cacti species. Its shelters include animal burrows and nests, and opuntia roots. It can find shelters in human dwellings and nearby areas as well. Furthermore, it has been found in dog kennels in North America.

Table 5: Niche description of *Triatoma gerstaeckeri*

Related variable	Statement	Reference
Seasonal activity patterns	They are hardly found in January, February, and March because they reduce their activity and their life cycle on winter. The number of instars found during the year varies according to season	(Thurman, 1945)
Life cycle	Life cycle in laboratory conditions is from six to eleven months.	(Thurman, 1945)
Host symbiosis	Associated to “cactus” <i>Opuntia</i> and “woodrats” <i>Neotoma micropus</i> . In the experiments they were found in cracks and cervices in tunnels leading to the nests, or in debris covering the nests. They could also be found as far as 90m from the nests of known hosts, This under logs and debris.	(Pippin, 1970)
Sylvatic shelters	They could also be found as far as 90m from the nests of known hosts, under logs and debris.	(Pippin, 1970)
Outdoor shelters	Has been found under cement slabs and in a garage away from the main home.	(Beard <i>et al.</i> , 2003)
Climate	Has been found in thick vegetation and dry climate, found north of Queretaro.	(Villagrán <i>et al.</i> , 2008)
Land use and vegetation	<i>T. gerstaeckeri</i> was mostly detected in Texas in the following ecoregions: Southern Texas Plains, Edwards Plateau, Western Gulf Coastal Plain.	(Kjos, Snowden and Olson, 2009)

Outdoor shelters	It has been found inside human dwellings, dog kennels in Texas. It has bitten humans.	(Kjos, Snowden and Olson, 2009)
Outdoor shelters	The authors examined hollow or rotten logs, cactus plants, dead yucca plants, and debris piles near burrows. They were also mostly found in dog kennels in the military base. authors believe they flew there because no instars were found.	(McPhatter <i>et al.</i> , 2012)
Outdoor shelters Sylvatic shelters	Found in woodrat nests, dog kennels and around houses	(Kjos <i>et al.</i> , 2013)
Blood meals	Here it was determined that they fed on <i>Neotoma micropus</i> , <i>Felis catus</i> , human, dogs, cows, <i>Procyon lotor</i> (raccoon), <i>Sus scrofa</i> (pig), <i>Coragyps atratus</i> (black vulture), <i>Cathartes aura</i> (turkey vulture), <i>Gryllus texensis</i> .	(Kjos <i>et al.</i> , 2013)
Land use and vegetation	The authors collected <i>T. gerstaeckeri</i> in domiciliary, peridomiciliar with mesquite vegetation and pine forests.	(MARTÍNEZ-TOVAR <i>et al.</i> , 2013)
Temperature and climate	They used temperature and climate variables to predict the distribution of <i>T. gerstaeckeri</i> using MAXENT. They predict in all scenarios tested, that there might be range change due to climate change (higher temperatures).	(Garza <i>et al.</i> , 2014)
Seasonal activity patterns Daily activity patterns	April and may months with peak activity, with collection from April to July. Activity at night, after midnight and some activity at dawn.	(Flores <i>et al.</i> , 2017)
Seasonal activity patterns Outdoor shelters	In this study most captures were performed from April to September, peaking in June/July. Found in residencies, hunting cabins, dog kennels, peridomicile, barns, chicken coops	(Curtis-Robles <i>et al.</i> , 2018)
Daily activity patterns	During the night, <i>T. gerstaeckeri</i> was more active around midnight, around 3:00 AM and around 7:00 AM at dawn.	(Wormington <i>et al.</i> , 2018)

The niche of *Triatoma mexicana* is described in Table 6. Temperature and humidity have been shown to affect its life cycle, also, it has more specific requirements of those factors than other triatomines. It has been found inside and outside human dwellings, but primarily outside.

Table 6: Niche description of *Triatoma mexicana*

Related variable	Statement	Reference
Indoor shelters Outdoor shelters	Inside hoses, <i>T. Mexicana</i> was found on beds, bedroom walls, and kitchen lamps, being bedroom walls the most frequent place. In outdoors, they were found on walls, patios, light posts, under stone walls, in lamps, windows, and water tanks, being the most frequent, outdoor walls. The bugs were found primarily outdoors. It may accomplish its entire lifecycle outdoors. They were collected in all risky and not risky housing materials considered, but either way, it was primarily collected outside the house.	(Salazar Schettino <i>et al.</i> , 2007)
Land use and vegetation	Has been found near agricultural land and scrubby vegetation. Has been found in and outside dwellings.	(Villagrán <i>et al.</i> , 2008)
Indoor shelters Outdoor shelters	Has been found in the typical places where one should expect to find the bugs.	(Villagrán <i>et al.</i> , 2008)
Life cycle	Egg to adult, 8.5 months on average, high temperatures and relative humidity shortens the development of the species	(Martínez-Ibarra <i>et al.</i> , 2008)
Climate	Compared the temperature and humidity conditions of <i>T. mexicana</i> and <i>T. barberi</i> , <i>T. Mexicana</i> had more specific requirements of temperature and relative humidity than <i>T. barberi</i> . There were places where both could be found.	(Becerril <i>et al.</i> , 2010)
Species distribution modelling	To model the species, they used four topographic variables, elevation, slope, aspect, and topographic index, one water layer and nine BIOCLIM variables.	(Ramsey <i>et al.</i> , 2015)
Species distribution modelling	Only WorldClim variables were used for the model	(Carmona-Castro, Moo-Llanes and Ramsey, 2018)

The type factors frequently described in the niche descriptions can be classified into local climate factors, vegetation and land use factors, and anthropogenic factors. In local climate factors, humidity and temperature were found to be important for the species presence in a region, while seasonality was important for the abundance of some species. In the vegetation and land use category, the presence of vertebrate hosts and shelters nearby them were important, as well as the vegetation and land use of the locality. Some triatomines had symbiotic relationships with specific host species like *T. gerstaeckeri* with *Neotoma micropus* wood rats and *Opuntia* cacti species. Inside anthropogenic factors, house construction materials, artificial light, and presence of synanthropic and domestic animals influenced the presence of triatomines.

Temperature and seasonality directly affected the physiology and life cycle of the three triatomine species. They were also found related to different land use types. Consequently, climatic factors and land use are good candidates of predictor variables to represent the niches of these species in species distribution modelling.

5.2.2 *Triatomine occurrence data collection and cleaning*

A total of 10,134 triatomine occurrences were downloaded from the GBIF database. A number of occurrences were marked as having problems; 1 record was marked as having the same coordinates as Mexico’s capital centroid, 18 as having coordinates in the sea, and 4 as being geographic outliers, 1 as having the same coordinates as a natural history museum. From these, the coordinates on the sea, and the ones on centroids and on biodiversity institutions were dropped. A total of 3,033 occurrences remained. From these occurrences, 2,022 belonged to *T. dimidiata*, 240 to *T. gerstaeckeri*, and 771 to *T. Mexicana*.

Occurrence data was also classified by the institution responsible into trusted and untrusted data as seen in Table 7. Only trusted data was used.

From the 3,033 records, only 1,290 had year data which was mostly distributed between 1995 and 2015 (Figure 9, and Figure 10). Only data from 1995 to 2015 was used, and after the spatial thinning only 1033 triatomine occurrences remained (Table 8).

Figure 9: Distribution of occurrences by species and year between 2000 and 2015 with outliers removed

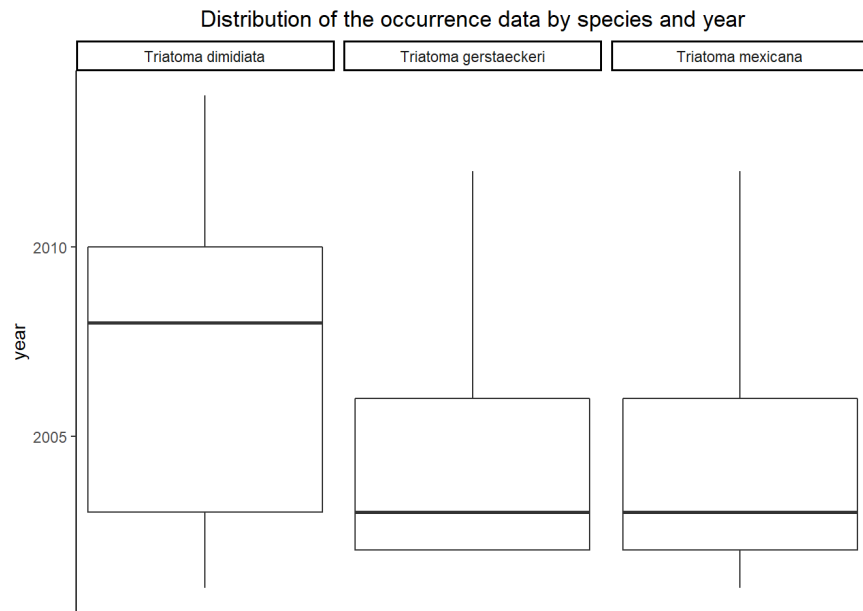


Figure 10: Distribution of all triatomine occurrences by year between 1990 and 2015 with outliers removed.

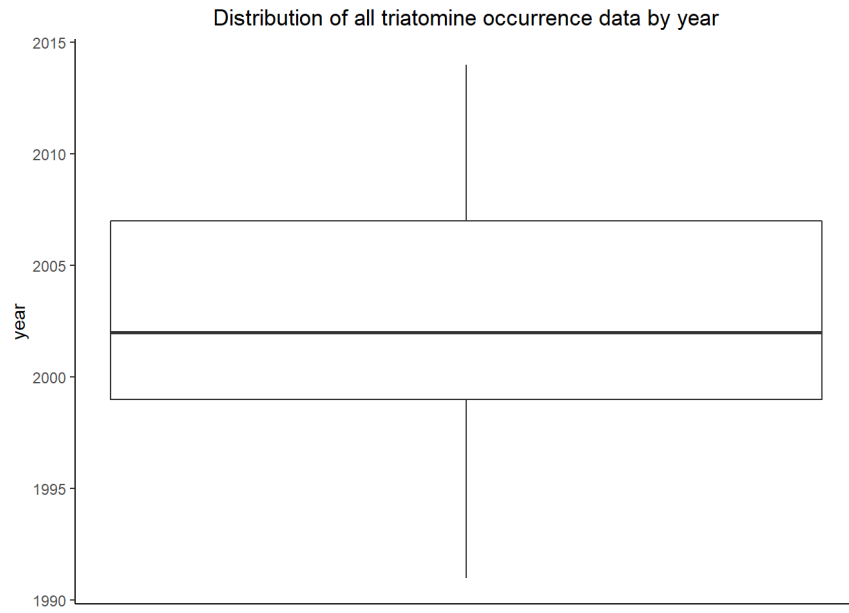


Table 7: Institution general data and classification by trusted source.

Database name	Institution name	n	Website	Description	Trusted source
CEPAVE.edu.ar	CEPAVE	2145	https://www.cepave.edu.ar/	A research center with its own triatomine collection and database.	TRUE
CP	CONABIO	7	https://www.gob.mx/conabio	Data published by CONABIO in GBIF	TRUE
ECOSUR	ECOSUR	26	https://www.ecosur.mx/	Colegio de la frontera del Sur	TRUE
FCB-UANL	FCB-UANL	21	http://www.fcb.uanl.mx/nw/es/	Biology faculty of UANL	TRUE
IBUNAM	IBUNAM	165	http://www.ib.unam.mx/	Biology institute of UNAM	TRUE
IIZD-UASLP	IIZD-UASLP	1	http://www.iizd.uaslp.mx/	Research institute in San Luis Potosi.	TRUE
iNaturalist	iNaturalist	35	https://www.inaturalist.org/	A database of citizen science	FALSE
InDRE-SSA	InDRE-SSA	4456	https://www.gob.mx/salud/acciones-y-programas/instituto-de-diagnostico-y-referencia-epidemiologicos-indre	Instituto de Diagnóstico y Referencia Epidemiológicos, an institution from the Mexican government.	TRUE
MHNCM	MHNCM	4	http://data.sedema.cdmx.gob.mx/museodehistorianatural/?view=featured	Museo de Historia Natural de la Ciudad de México, the natural history museum of Mexico city.	TRUE
NO DISPONIBLE	CONABIO	16	https://www.gob.mx/conabio	Published by CONABIO, in GBIF	TRUE
SSG	SSG	1014	https://salud.guanajuato.gob.mx/	Secretaría de Salud de Guanajuato	TRUE

5.2.3 Reservoir occurrence data collection and cleaning

A total of 18,888 occurrences were downloaded from the GBIF database. In all, 95 occurrences were flagged, 47 occurrences were removed because they had sea coordinates, 1 because it was detected as an outlier and 39 because the coordinates corresponded to biodiversity institutions like museums or research centers. In total 87 coordinates were removed. Occurrences on cities were not removed.

1,934 remained after the cleaning. However, the data was not classified into trusted and untrusted because of the high number of institutions and time constraints. Data comes from 60 different institutions, including from the iNaturalist database.

Overall, 1,875 occurrences have year data, which is distributed mostly between years 2015 and 2020, with a mean of 2015 and a median of 2018 (Figure 11). Only data between 1995 and 2015 was used, and after spatial thinning, only 250 records remained (Table 8).

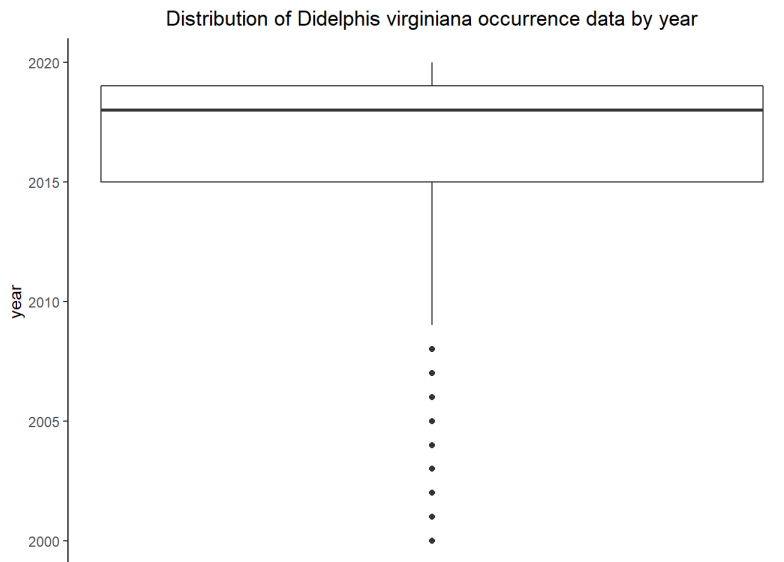


Figure 11: Distribution of *Didelphis virginiana* occurrence data from 2000 to 2020

Table 8: Number of occurrence points used for modelling

Species	Number of occurrences after cleaning
<i>Didelphis virginiana</i>	250
<i>Triatoma dimidiata</i>	646
<i>Triatoma gerstaeckeri</i>	53
<i>Triatoma mexicana</i>	334

The cleaned vector and reservoir occurrence data is mapped in Figure 12.

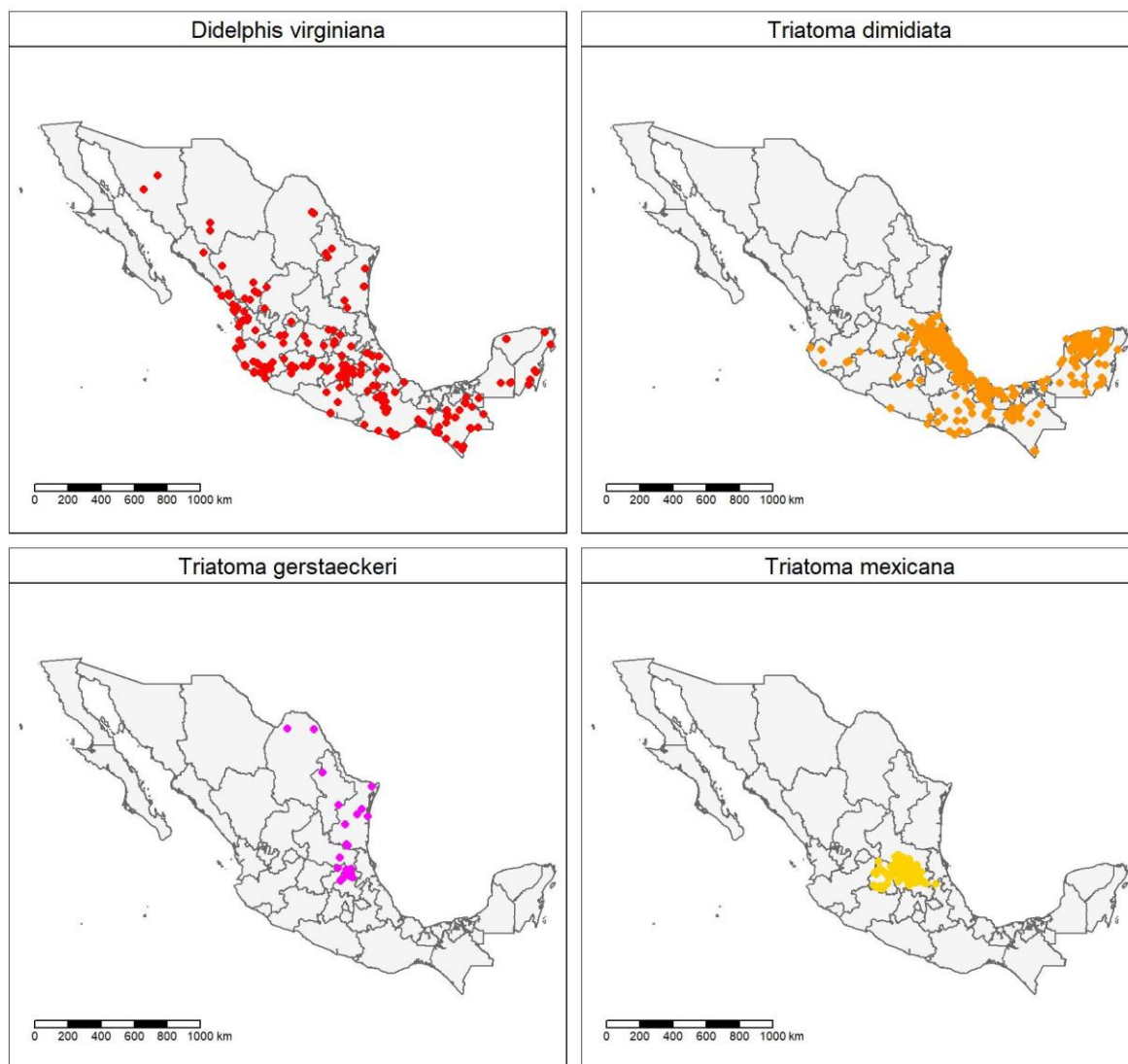


Figure 12: Map of cleaned occurrence points of vectors and reservoir chosen to be used in modelling

5.2.4 Model training and evaluation

Many models were ran and evaluated to choose the best combination of parameters. Forty of the models ran to do this selection are shown in Annex 2. The three model parameter combinations which yielded the lowest AIC values for each species are shown in Table 9.

The settings, average significance and performance statistics of the chosen models can be seen in Table 10, Table 11, Table 12, and Table 13. All distribution models were significant, with AUC test values from 0.80 to 0.93 which indicate robust models.

Table 9: Model parameters ran and corresponding AIC values. Delta AIC marks zero for the lowest value of AIC in a group of models.

Species	Features	Regularization multiplier	AICc	Delta AICc
<i>T. dimidiata</i>	LQHP	1	17119.53	40.04
<i>T. dimidiata</i>	LQH	3	17079.48	0
<i>T. dimidiata</i>	LQHP	3	17095.38	15.89
<i>T. gerstaeckeri</i>	LQHP	1	1414.00	68.90
<i>T. gerstaeckeri</i>	LQH	3	1345.09	0
<i>T. gerstaeckeri</i>	LQHP	3	1397.41	52.32
<i>T. mexicana</i>	LQH	3	7572.69	23.35
<i>T. mexicana</i>	LQHP	3	7549.34	0
<i>T. mexicana</i>	LQH	5	7682.37	133.03
<i>D. virginiana</i>	LQH	5	6894.71	8.30
<i>D. virginiana</i>	LQHP	5	6886.41	0
<i>D. virginiana</i>	LQH	10	6930.78	44.37

The model for *T. dimidiata* has a minimum training presence test omission rate (MTPTOR) and 10% percentile training test omission rate (10PTPTOR) of 0.0031 and 0.1102, respectively (Table 10). The expected values of MTPTOR and 10PTPTOR are 0 and 0.1 respectively, consequently both models are not overfitted.

Table 10: Settings and average evaluation statistics of the species distribution models for *T. dimidiata*.

Statistic	Value
Regularization multiplier:	2
Features:	LQH
#Training samples	579.6
Training AUC	0.8645
#Test samples	64.4
Test AUC	0.8559
Minimum training presence test omission	0.0031
Minimum training presence binomial probability	0.0077
10 percentile training presence test omission	0.1102
10 percentile training presence binomial probability	< 0.05
Maximum test sensitivity plus specificity Logistic threshold	0.3417
Maximum test sensitivity plus specificity test omission	0.1366
Maximum test sensitivity plus specificity binomial probability	< 0.05

For the *T. dimidiata* model prediction (Figure 13), greater output values are predicted in the Huasteca region, at the east of the state. Inside the Huasteca region, more suitable conditions are predicted at southern municipalities.

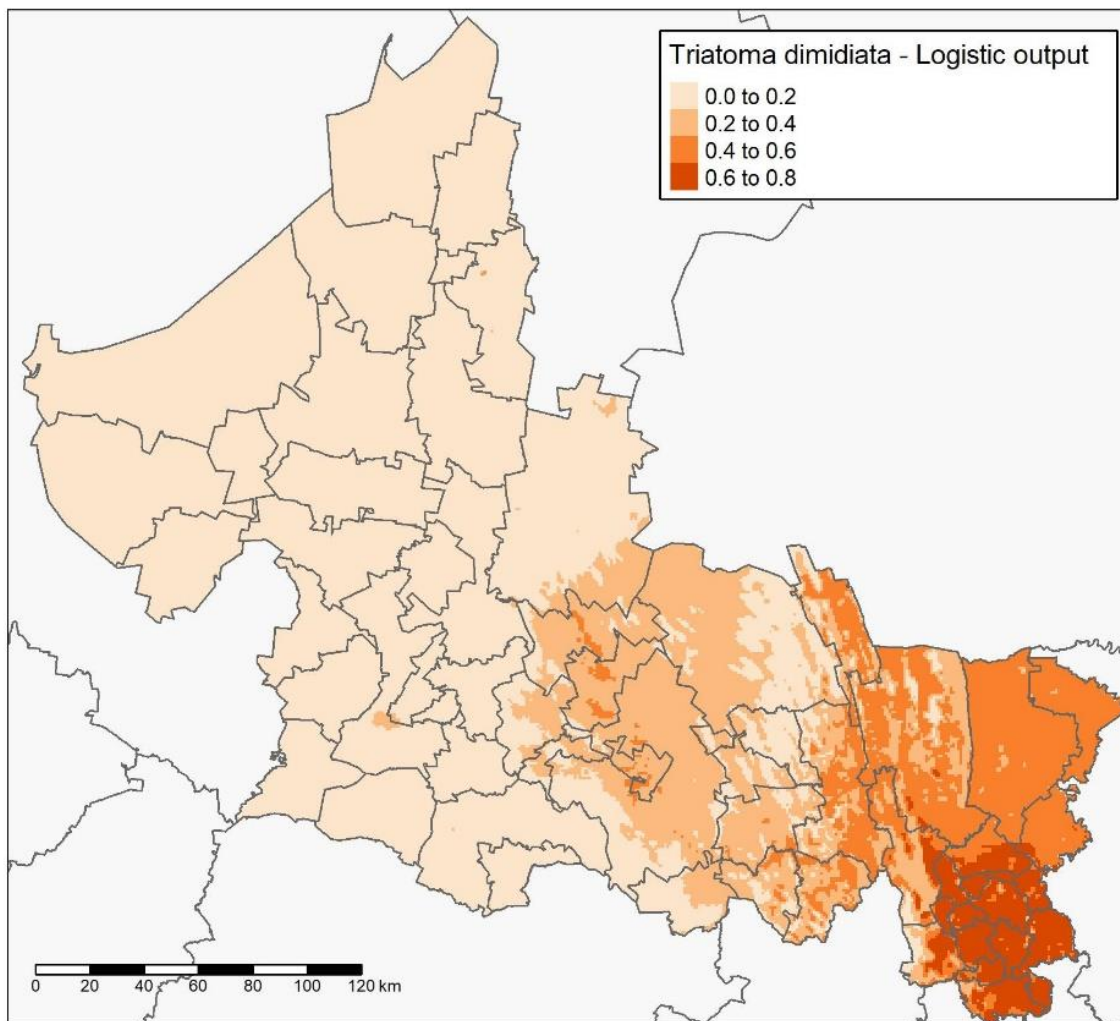


Figure 13: *Triatoma dimidiata* Maxent distribution model, logistic output.

Regarding the model for *T. gesrtaeckeri*, MTPTOR and 10PTPTOR are 0.03 and 0.18, respectively (Table 11). The MTPTOR shows that the model is a little over fitted, and the 10PTPTOR, which is less sensitive to outlier occurrences, shows a higher omission rate than the 10% expected.

Table 11: Settings and average evaluation statistics of the species distribution models for *T. gerstaeckeri*.

Statistic	Value
Regularization multiplier:	2
Features:	LQH
#Training samples	42.4
Training AUC	0.9576
#Test samples	10.6
Test AUC	0.9112
Minimum training presence test omission	0.0382
Minimum training presence binomial probability	0.0056
10 percentile training presence test omission	0.1891
10 percentile training presence binomial probability	< 0.05
Maximum test sensitivity plus specificity Logistic threshold	0.1934
Maximum test sensitivity plus specificity test omission	0.1327
Maximum test sensitivity plus specificity binomial probability	< 0.05

The prediction (Figure 14) for *T. gerstaeckeri* distribution shows more suitable conditions at the west of the Huasteca region, in the Sierra Madre Oriental mountain range.

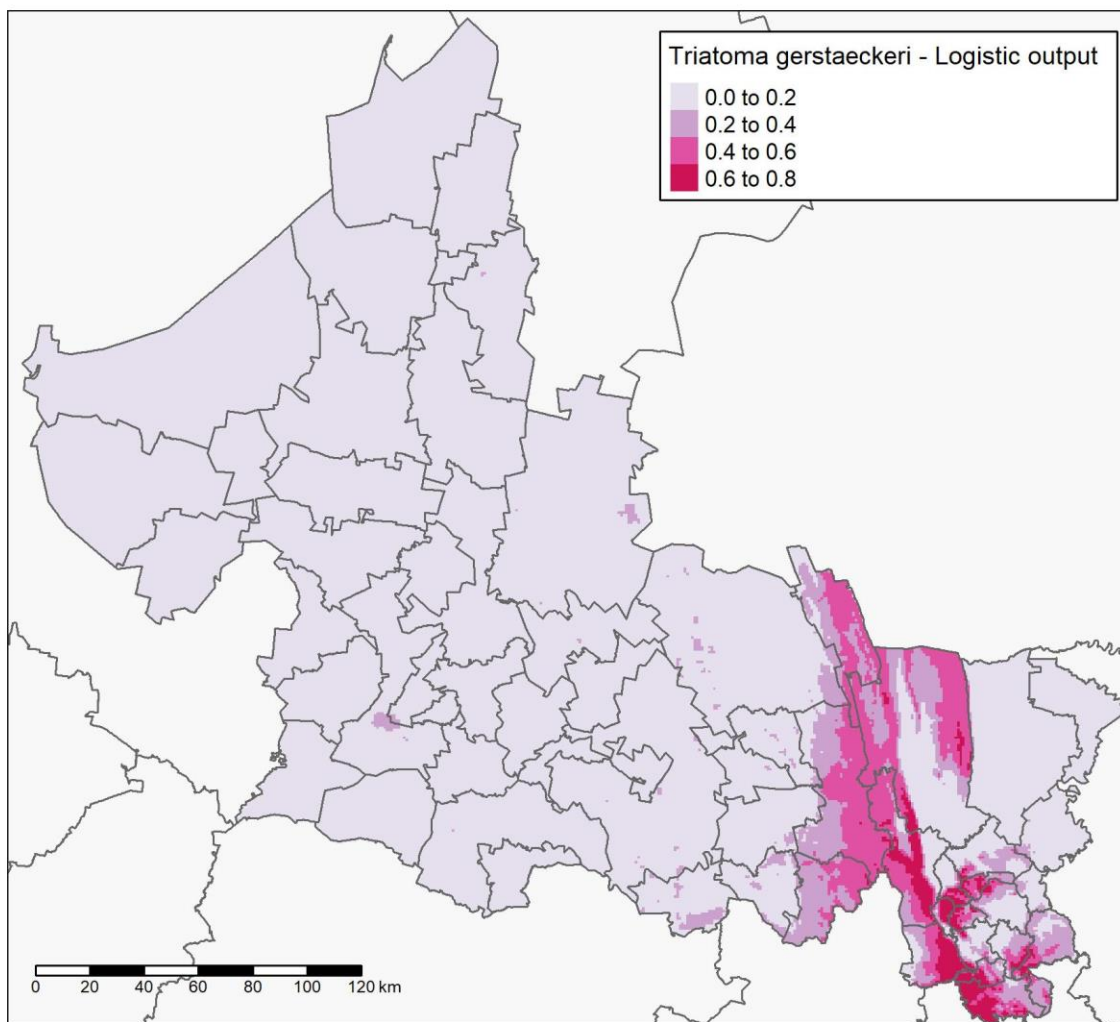


Figure 14: *Triatoma gerstaeckeri* Maxent distribution model, logistic output.

The model for *T. Mexicana* had a MTPTOR of 0.006 (Table 12) which indicates good performance, but a 10PTPTOR of 0.21 which indicates 0.1 more than expected. By this statistic, this model has the most omission error.

Table 12: Settings and average evaluation statistics of the species distribution models for *T. mexicana*.

Statistic	Value
Regularization multiplier:	3
Features:	LQHP
#Training samples	300.6
Training AUC	0.9498
#Test samples	33.4
Test AUC	0.9307
Minimum training presence test omission	0.006
Minimum training presence binomial probability	0.0003
10 percentile training presence test omission	0.219
10 percentile training presence binomial probability	< 0.05
Maximum test sensitivity plus specificity Logistic threshold	0.2265
Maximum test sensitivity plus specificity test omission	0.0803
Maximum test sensitivity plus specificity binomial probability	< 0.05

The prediction (Figure 15) shows the most suitable areas in the south of the media region of the state and in the east of the central region.

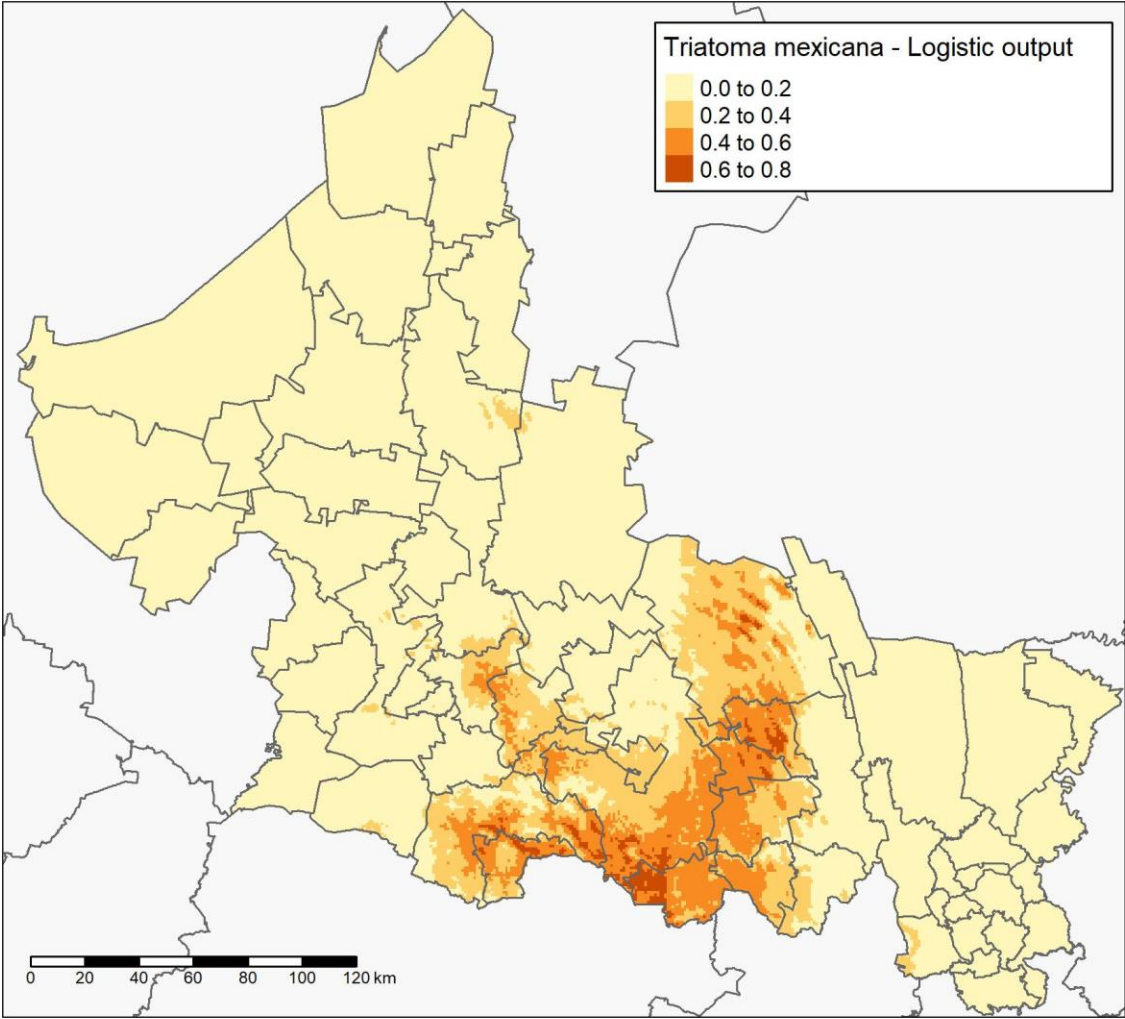


Figure 15: *Triatoma mexicana* Maxent distribution model, logistic output.

The MTPTOR is not significant in the model for *D. virginiana* (Table 13), therefore it cannot be interpreted. However, 10PTPTOR is a better estimator for omission error in this case. 10PTPTOR shows a low omission error.

Table 13: Settings and average evaluation statistics of the species distribution models for *D. virginiana*.

Statistic	Value
Regularization multiplier:	5
Features:	LQHP
#Training samples	224.1
Training AUC	0.8183
#Test samples	24.9
Test AUC	0.8058
Minimum training presence test omission	0.004
Minimum training presence binomial probability	0.2799
10 percentile training presence test omission	0.1165
10 percentile training presence binomial probability	< 0.05
Maximum test sensitivity plus specificity Logistic threshold	0.4109
Maximum test sensitivity plus specificity test omission	0.2408
Maximum test sensitivity plus specificity binomial probability	< 0.05

In the model prediction (Figure 16), the most suitable areas are in the Huasteca and Central regions, specially the Sierra Madre Oriental mountain range and the San Luis Potosi and Ciudad Valles cities.

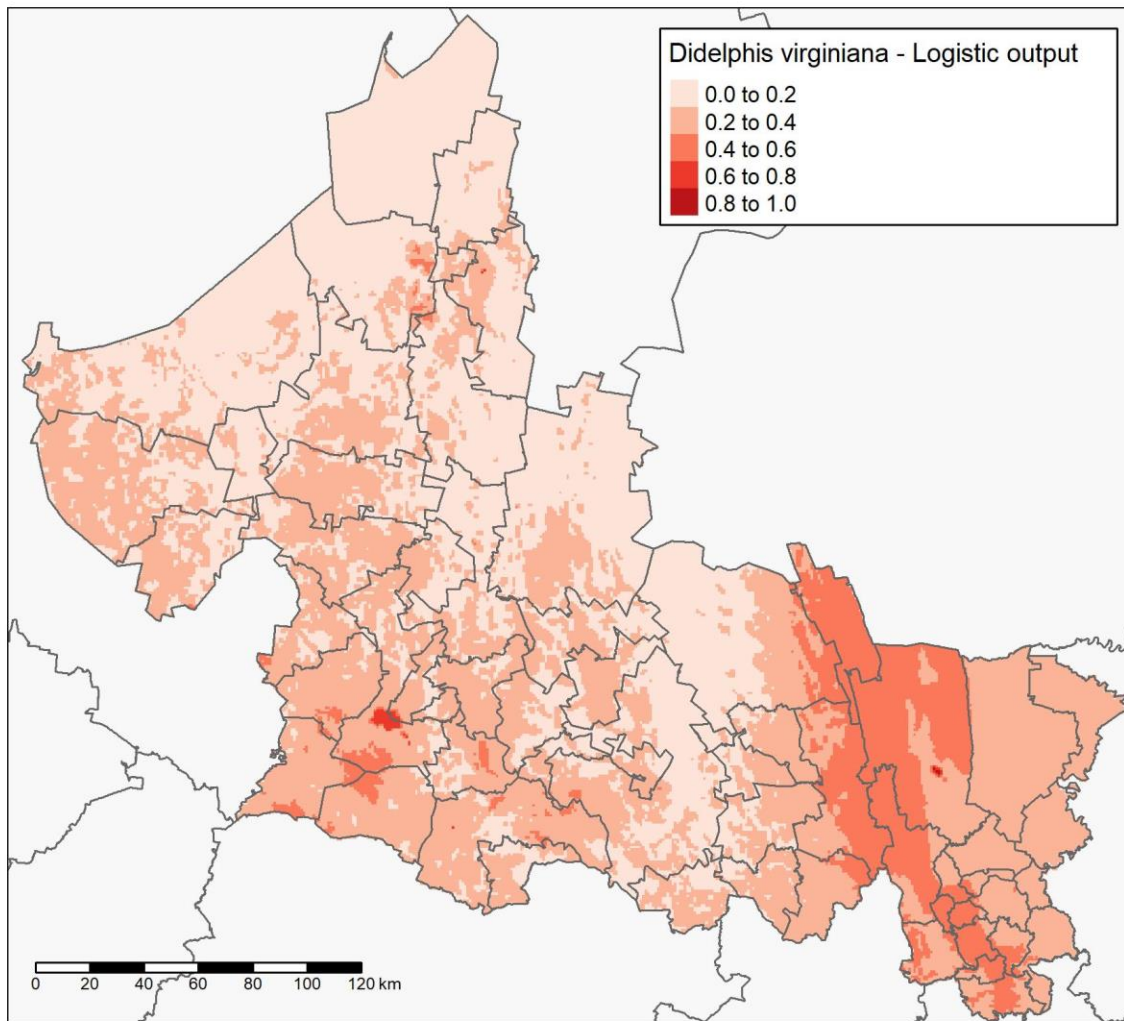


Figure 16: *Didelphis virginiana* Maxent distribution model, logistic output.

5.3 Mapping risk areas

Firstly, the Maxent's logistic output of all species models was converted into binary predictions of presence through setting the Maximum test sensitivity plus specificity (MTSPS) as a threshold for the logistic output (Table 10, Table 11, Table 12, and Table 13). For *T. dimidiata*, the logistic threshold used was 0.3417 which had an omission rate of 0.13. The total area predicted to be suitable for *T. dimidiata* is 10,717.37 km² which corresponds to 17.53% of the total area of the state of San Luis Potosi. Most of the area predicted (Figure 17) is at the east of Huasteca region, with less area in the east. There is also some area predicted in Media region.

For *T. gerstaeckeri*, a threshold of 0.19 was used which resulted in an omission rate of 0.13. The total predicted area to be suitable for the species was 7,506.15 km² which corresponds to 12.27% of the total area of the state. The prediction (Figure 17) shows presences areas mostly in the Sierra Madre Oriental mountain range in the Huasteca region, with smaller predicted areas across Media and Central regions, including San Luis Potosi city.

In the models for *T. mexicana*, a threshold of 0.2265 was used., this gave omission rates of 0.08. The area predicted to be suitable for the species is 9,605.80 km² which corresponds to 15.71 % of the study area. The model prediction (Figure 17) shows suitable area mostly in media region and in the eastern part of central region.

Lastly, for *D. virginiana*, a threshold of 0.4109 was used and omission rates of 0.24 were obtained. The suitable predicted area is of 4,976.06 km² Which corresponds to 8.13% of the state's territory. The prediction (Figure 17) shows suitable area in the Sierra Madre Oriental mountain range and in some localities in central region including San Luis Potosi City.

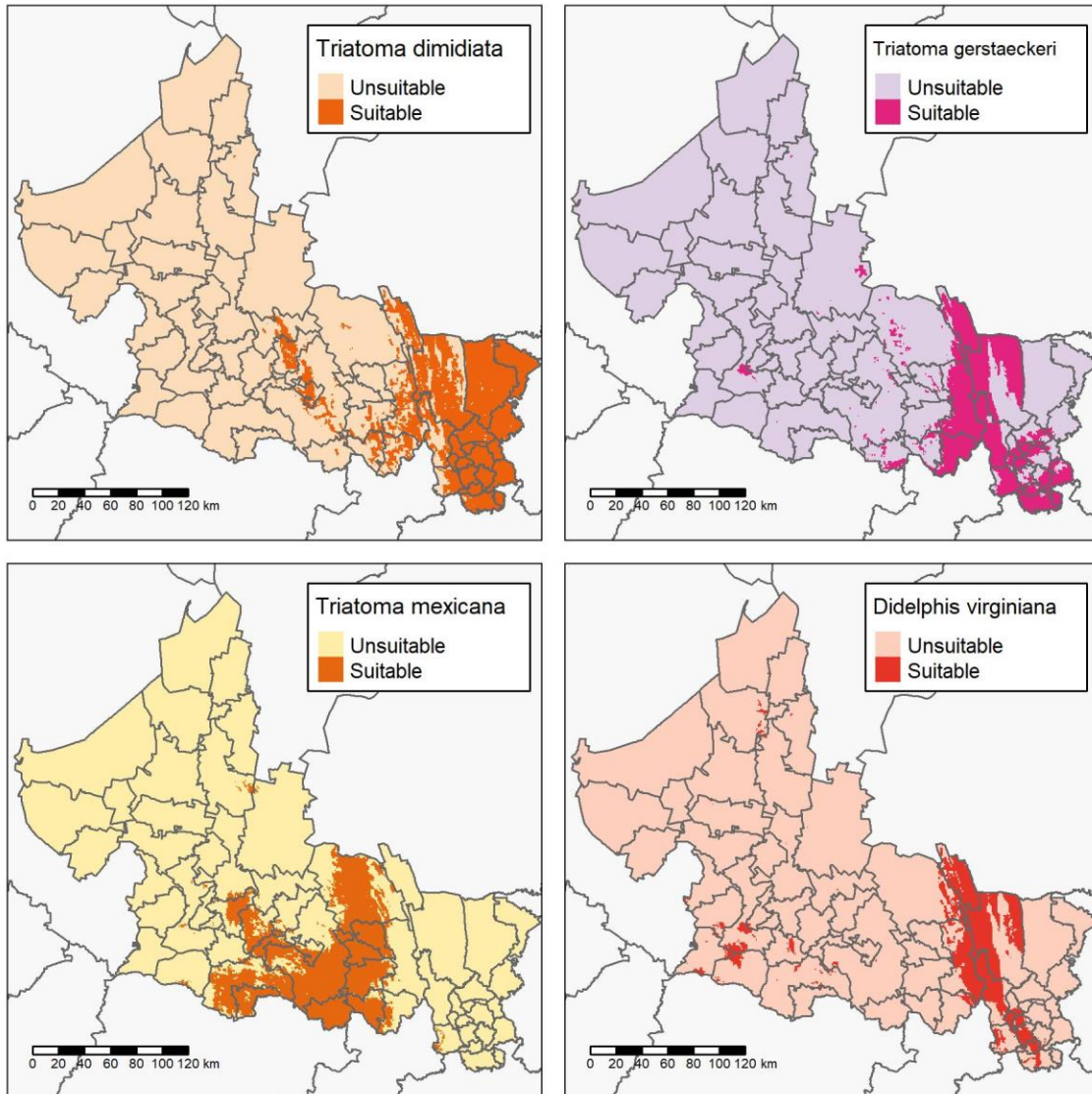


Figure 17: Binary prediction of species suitability for vector species *T. dimidiata*, *T. gerstaeckeri*, *T. mexicana* and reservoir species *D. virginiana*.

The overlap of the binary predictions of vectors and reservoir is shown in Figure 18. The total area with at least one predicted vector presence is 22,061.67 km² which corresponds to 36.08% of the state's territory. Moreover, the overlapping area between vectors and reservoirs was 4,306.13 km² (7.04% of the state's area). The overlapping area is mostly located in the Sierra Madre Oriental mountain range; however, vectors and reservoir are also predicted for the state's capital (Figure 18).

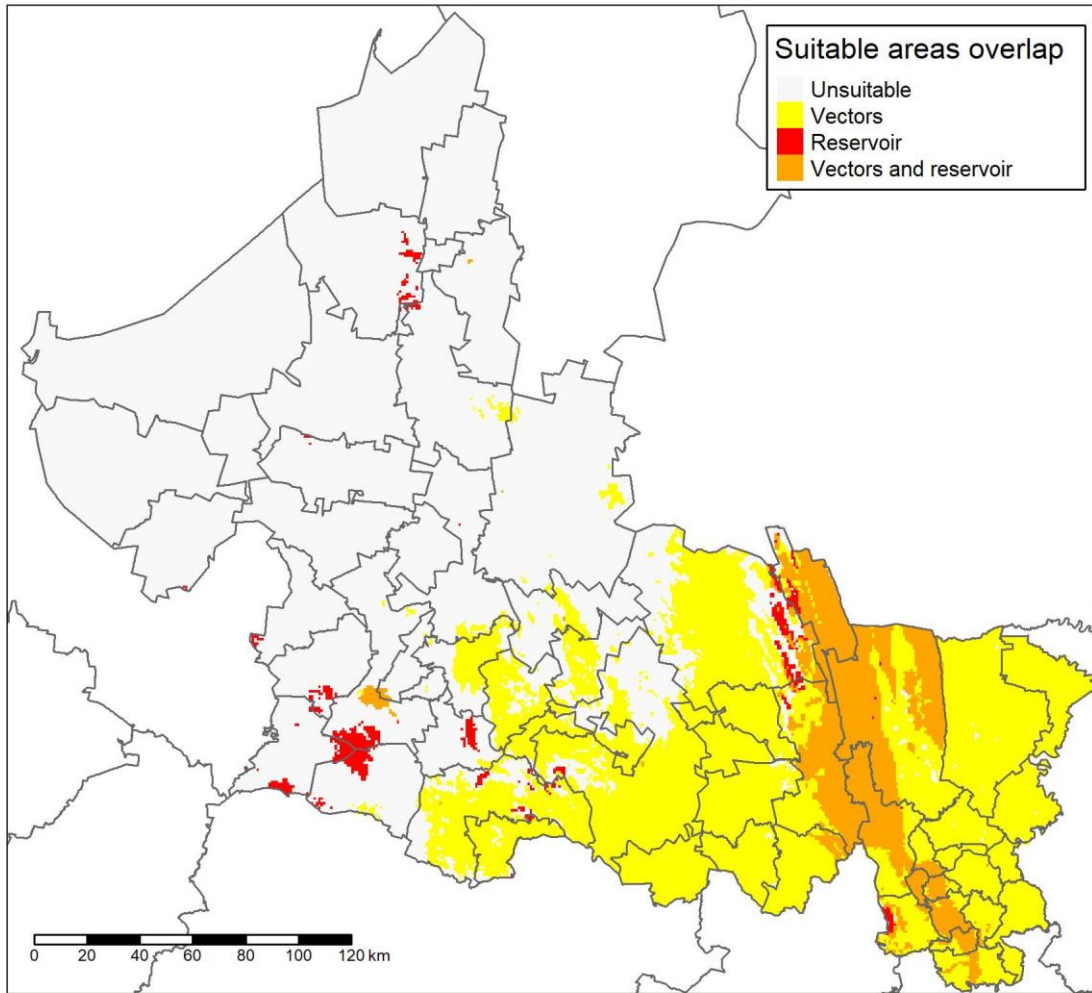


Figure 18: Vector and reservoir overlapping areas

The total human population living in areas with predicted vector or reservoir suitability are shown in Table 14.

Table 14: Total of rural and urban population living in areas suitable for Chagas disease vectors and reservoirs

	Type of host	Population	%
Rural	Reservoir	2,072	0.37
	Vector	392,420	69.32
	Vector-Reservoir	171,595	30.31
	Total	566,087	100.00
Urban	Reservoir	722,772	16.39
	Vector	1,359,839	30.83
	Vector-Reservoir	2,327,792	52.78
	Total	4410403	100.00

Between vectors, most overlapping areas were predicted in the west of Huasteca region and east of Media region (Figure 19). *T. dimidiata* and *T. gerstaeckeri*, both were predicted to occur in southern Huasteca region and in Sierra Madre Oriental. The “*T. dimidiata* - *T. mexicana*”, and “*T. gerstaeckeri* - *T. mexicana*” pairs were predicted in patches in east Media region. A small quantity of pixels suitable for all vector species were predicted at the south east of Media region (Figure 19). Regarding the vector species, *D. virginiana* had overlapping suitable areas with all species of vectors, mostly with *T. gerstaeckeri* and *T. dimidiata* in the Sierra Madre Oriental (Figure 19).

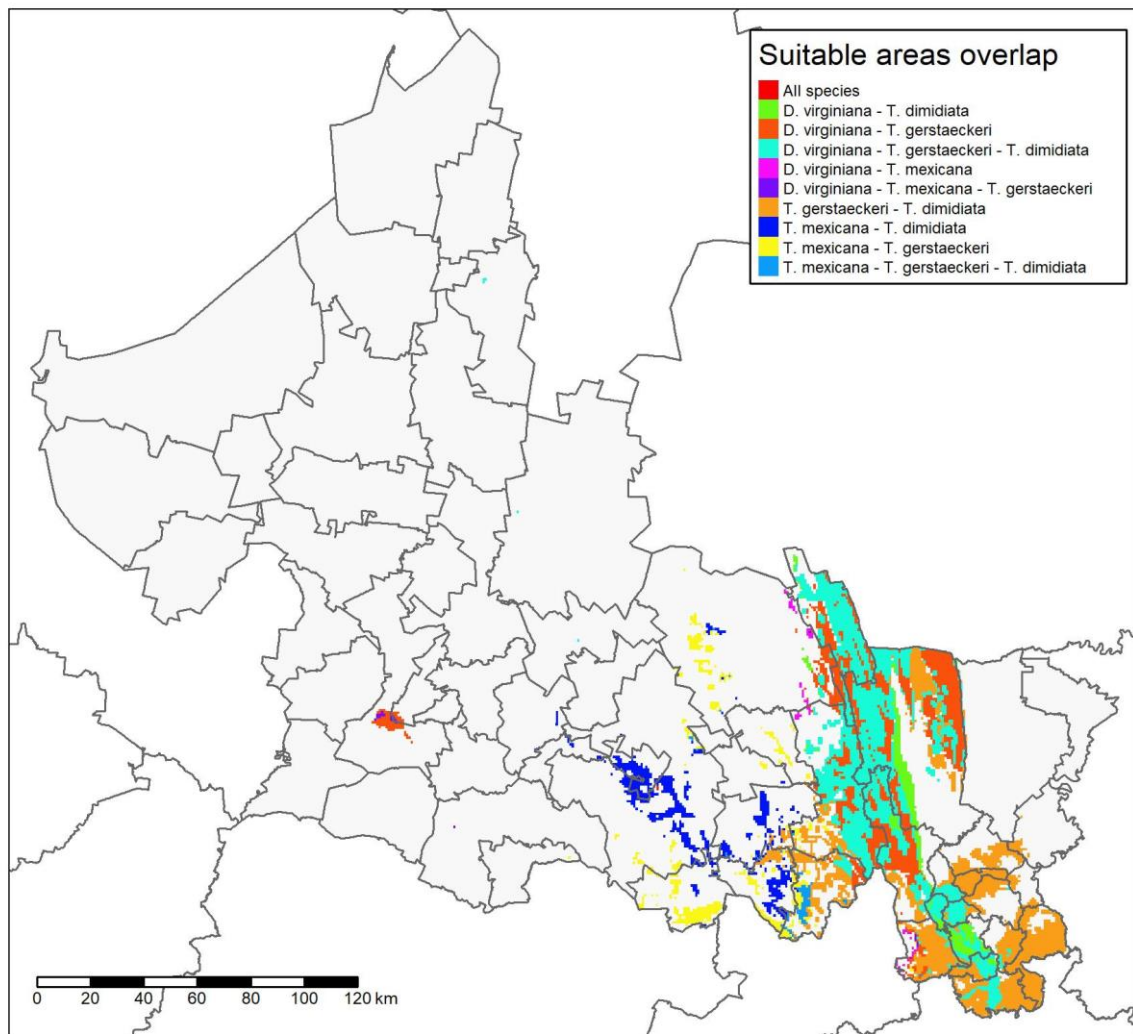


Figure 19: Identified overlaps suitable areas between vectors and reservoir of *T. cruzi*.

The extension of the overlap areas for all species can be seen in Table 15.

Table 15: Total predicted presence areas per species in San Luis Potosi

Species present	Area (km ²)
D. virginiana	703.93
T. mexicana	8399.70
T. gerstaeckeri	949.47
T. dimidiata	5109.89
D. virginiana - T. mexicana	48.39
D. virginiana - T. gerstaeckeri	1405.15
D. virginiana - T. dimidiata	332.09
T. mexicana - T. gerstaeckeri	457.68
T. mexicana - T. dimidiata	599.01
T. gerstaeckeri - T. dimidiata	2125.40
D. virginiana - T. mexicana - T. gerstaeckeri	17.47
D. virginiana - T. gerstaeckeri - T. dimidiata	2467.43
T. mexicana - T. gerstaeckeri - T. dimidiata	81.95
All species	1.59
Total	22699.18

6 Discussion and conclusions

6.1 Selection of important reservoir species

The most important reservoir species in this study was *Didelphis virginiana* because of its reported infection rates and its numbers of reports in literature. Some similarities can be found between these results and with the ones of Jansen, Xavier and Roque (2018), who in their 20 yearlong longitudinal study of Brazilian *T. cruzi* reservoirs, reported that the most important orders for *T. cruzi* transmission in the wild were Didelphimorphia, Primates Carnivora and Chiroptera. In their results, *Didelphis spp.* and *Philander Spp.* (Didelphimorphia) have been reported to be the most infective in some of the studied biomes. The same mammal orders and the genus *Didelphis* seem to be important, both in Brazil and in Mexico. Evidence that suggest that there is association between *T. cruzi* and certain taxa of mammals, for example, specific genotypes of *T. cruzi* have been found to be associated to opossums and armadillos (YEO *et al.*, 2005), moreover, some opossum species can have special immunological responses to the parasite and be more infective depending on the specific parasite's genotype they are infected with (Legey *et al.*, 2003).

Other important reservoir species were the house mouse (*Mus musculus*) and rat (*Rattus rattus*), which were not used in the models because those species are closely related to human populations and are virtually everywhere. However, they have epidemiological importance because they can live in the household and the peridomicile and could be a nexus for the parasite to make its way into the household (Rosal *et al.*, 2018).

An interesting rodent species was the woodrat *Neotoma micropus*, which had an infection rate of less than 0.25 and was only mentioned in one article. However, it was mentioned in the niche descriptions of *T. gerstaeckeri* and *T. mexicana* (Kjos *et al.*, 2013; Pippin, 1970), where the woodrat provided shelter for the bugs in its burrows. This species may allow the parasite to exist in dryer climates where shrub and cacti vegetation are present, and at the same time, favor the transmission to the vector species. More studies about rodent species' role as *T. cruzi* reservoirs in SLP are needed.

Yucatán and Morelos had the most reservoirs reported by state. Chagas disease has been historically studied more in the south of Mexico, this has led to a difference in research output by state and also, a difference in Chagas disease reported cases (Cruz-Reyes and Pickering-López, 2006; Ibáñez-Cervantes *et al.*, 2018). Nevertheless, there are still some central and northern states like Morelos and Nuevo Leon, that have carried Chagas disease reservoir research. Ibáñez-Cervantes *et al.* (2018), discussed that there has been an increase in reported cases of Chagas disease in Mexico due to better testing capabilities, but there are still gaps in the knowledge of reservoir of *T. cruzi* in Mexico.

One limitation of the important reservoir selection method used in this study was that the reporting of the infection rates of species could have been made clearer using standard meta-analysis methods (Forero *et al.*, 2019). Another limitation was that only two variables were used to select important species, the sum of the reported infection rates and the number of times a species was reported as infected in different articles. Other relevant criteria could be used in future studies like symbiotic relationships with vectors or potential for infection to vectors and humans.

This study is based on Chagas disease reservoir research from other states, *T. cruzi*'s reservoirs have not been studied in the state of San Luis Potosi, however, the disease has been reported in the state (Ibáñez-Cervantes *et al.*, 2018; Medina-Garza *et al.*, 2016; Oscar Aldana Cruz *et al.*, 2009). A better knowledge of Chagas disease reservoirs could give new insights over the epidemiology of the disease in the state.

Thus, the opossum *D. virginiana* was found to be possibly one of the most important reservoirs of *T. cruzi* in San Luis Potosi, however these results are based on research from other states of Mexico. Another finding from this section was that domestic and wild rodents could have a role in Chagas disease epidemiology and should also be studied. To confirm these hypotheses, research on *T. cruzi* reservoirs in San Luis Potosi is needed, this could help to better comprehend the disease epidemiology in the state.

6.2 Distribution modelling of vectors and reservoir species

Potential distributions of Chagas disease vectors and one reservoir were modelled in San Luis Potosí. The species distribution modelling works by taking species occurrence points with environmental variables associated and then making prediction (using an algorithm) in unexplored places with similar environmental conditions (Franklin, 2012). The models in this study are based mostly in environmental variables and the only biotic variable taken into account was the land use map which is a proxy for human activity. The model does not consider biotic interactions such as intraspecific competition or the target species mobility, consequently the models on the study are best interpreted as suitability maps or potential distributions (Araújo and Guisan, 2006; Franklin, 2012; Soberón, Osorio-Olvera and Peterson, 2017). Since the models were trained using background points from the current known species range in Mexico and not just the state of San Luis Potosí, the predictions should be robust, and do well at interpolating presences at unsampled regions in the state (Merow, Smith and Silander, 2013).

Regarding the selection of predictor variables to use in the models, since climatic and anthropogenic factors were frequently mentioned in the niche descriptions, the WorldClim and Land use maps were used as proxies for these. However, the presence of specific species of woodrats and *Opuntia* cacti were also mentioned frequently in the literature. Araújo, Marcondes-Machado and Costa (2014), observed that the inclusion of biotic factors (as

species distribution models) in the predictor variables significantly improved the model predictions. In the future, distribution models or abundance data of woodrats and *Opuntia* cacti could be used to make better models for vectors of Chagas disease.

The final predictions of suitability for each species are explained by the climates present in the state, as well as elevation and land use. Rezedowski (1965), reports that San Luis Potosí has various degrees of arid climate in 70% of its territory, with very humid climates at the south east of the state. However, sudden changes in topography make possible for very different climates to be adjacent to each other. Therefore, the topography of the state explains the climatic variation in the state. The Model predictions can be associated to climatic and topographic features of the state. Firstly, *T. dimidiata* has suitable habitat mostly at the east, in the coastal plains, while *T. gerstaeckeri* is mostly associated with the Sierra Madre Oriental mountain range and *T. mexicana*'s suitable habitat occurs west of the Sierra Madre Oriental where semi-arid climate occurs. Similarly, to *T. gerstaeckeri*, *D. virginiana* suitability was predicted in the Sierra Madre Oriental area and in some patches at the west of the state. Patches of high suitability outside the main predicted areas can be explained by the presence of isolated mountains in the San Luis Potosí plateau, which can determine local climates. Another factor is the presence of urban areas across the state because the land use map used as a predictor variable also factored urban areas.

The statistics chosen for model evaluation were Test AUC for model significance and Minimum Training Presence Omission Rates (MTPTOR), and 10 Percentile Training Presence Test Omission Rate (10PTPTOR) (Peterson *et al.*, 2011). No validation step was considered in this study because the objective of the modelling was a descriptive one (Araújo and Guisan, 2006), consequently the evaluation was used to assess significance and describe omission error, which is the only type of error that can be estimated in presence only distribution models (Franklin, 2012; Peterson *et al.*, 2011). The AUC for all the models indicated they were robust (higher than 0.8), the high values of 10PTPTOR for *T. gerstaeckeri* and *T. mexicana* indicate a high omission error (false negatives) in the predictions of suitability of these species. For *T. gerstaeckeri* this could be explained by the low number of presence data available for training and evaluation. *T. mexicana* had a much higher number of training data, consequently, error could come from other sources like the predictor variables uses, bias on the data, or the endemic distribution of the species.

Ramsey *et al.* (2015), produced an atlas of the geographic distributions of the triatomines of Mexico. In their work they modeled the three species of vectors modeled here, together with other 16 species of triatomines. They reported that the Sierra Madre Oriental was one of the places with the most predicted triatomine richness in México. The present study builds up from and improves on this previous work by using higher resolution predictor variables, focusing only on predicting three species of triatomines in one state, and taking into

consideration more criteria for data cleaning. The resulting models of both studies show differences because of these reasons. Additionally, triatomine distributions have also been mapped in the Mexican states of Puebla, Guanajuato and Veracruz using different algorithms than Maxent (López-Cárdenas *et al.*, 2005; Sandoval-ruiz, Zumaquero-rios and Rojas-soto, 2008; SANDOVAL-RUIZ *et al.*, 2012). The advantages of the models in this thesis over these last studies is that the models considered the whole known species range of the vectors. Also, a species of reservoir was considered, and year data was considered in data cleaning.

One limitation of the modelling is that data cleaning for *D. virginiana* was not done as thoroughly as with the triatomine species because of time constraints. As a result, the institution information in GBIF metadata was not examined. This may lead to misidentifications present in the occurrence data. Another limitation is that only occurrence points in Mexican territory were used for species with range outside of the country. This was due the lack of comparable land use maps for Central America and the United States.

An additional consideration is that triatomines respond to seasonal temperature changes which can increase bug abundance and trigger migrations (Di Iorio and Gürtler, 2017; DUMONTEIL *et al.*, 2002). In future studies about triatomine distribution in San Luis Potosí, models with higher temporal resolution could be used to see if the suitable areas change depending on seasonality. Another direction that future studies could take is visiting areas previously predicted with “high habitat suitability” and perform random sampling. Using this data it is possible to further evaluate the models in this work. Franklin (2012) calls this methodology prospective sampling.

As a last recommendation, it is important to consider that other types of distribution modelling exist. Peterson, Papeş and Soberón (2015) distinguish between correlative models (Maxent being one of them) and mechanistic models. They argue that correlative models do not calculate the full fundamental niche, but a subset of this, depending on the occurrence points. For the purposes of this thesis, this is useful because the models predict more reduced areas that are more likely to be occupied by the animals. On the other hand, mechanistic models, on the other hand, are based on the physiological requirements of a species and could produce accurate prediction of the fundamental niche of a species. However, they require detailed measurements of physiological requirements and characteristics of the species. With more knowledge about the fundamental niche of the species, better predictor variables could be chosen for correlative models, and risk areas could be identified more accurately. Consequently, more studies about the physiological limits for the most important Chagas vectors are still needed.

In conclusion, using species distribution modeling to predict the distribution of vectors and reservoirs could allow us to understand the geography of Chagas diseases in San Luis Potosí. data of Chagas disease vectors and mammal reservoirs is essential. knowledge about Chagas disease vectors and the state's mammal fauna could be generated and used to feed these models.

6.3 Mapping risk areas

In this study, areas with potential human contact with Chagas disease vectors and reservoir were mapped and then overlaid with human population spatial demographic data. One third of all the rural human population and half of the urban population of San Luis Potosi were calculated to live in areas suitable for both vectors and reservoir. These results could be used by health authorities to identify new locations where transmission through contact with vectors still occurs.

Although, urban localities traditionally are not the focus of surveillance of the disease, half of the urban population was calculated at risk here. Urdaneta-Morales (2014), that Chagas disease is becoming an emergent urban zoonosis (transmission from animals to humans), due to the increase in reports of infected vectors and reservoirs (mice and opossums) in urban localities in Caracas, the capital city of Venezuela. The results of this thesis could be used in the identification of on going urban transmission cycles of the disease as well as prevention in not yet affected localities.

The presence of vectors and reservoirs in an area allow mainly two routes of transmission, vectorial and oral (Pérez-Molina and Molina, 2018). The vectorial route is the most frequent and studied way of transmission, where people have direct contact with the feces of an infected triatomine bug (Jurberg and Galvao, 2006), however, oral transmission has received more attention in recent years. Filigheddu, Górgolas and Ramos (2017), found that many cases of oral transmission have been reported in the literature. These have been caused mainly by contaminated food with triatomine or opossum feces or other secretions, and by the ingestion of undercooked meat of infected animals. Oral transmission is still a possibility in Mexico wherever triatomine bugs can be found, especially since animals like sheep and pigs have been found infected with *T. cruzi*, in some cases with a high infection rate (Jiménez-Coello *et al.*, 2012; López-Cancino *et al.*, 2015; Ruiz-Piña *et al.*, 2018b). Taking into account vector and reservoir distributions in disease surveillance could be effective for the identification of risk areas for oral transmission as well. Cattle farms inside risk areas for Chagas disease could make it possible for the parasite to infect humans through the consumption of contaminated cattle meat.

It has been argued before that the predictions made by distribution models that use occurrence data are better interpreted as potential distributions, habitat suitability, or under certain conditions, realized niche (Araújo and Guisan, 2006; Franklin, 2012; Soberón, Osorio-Olvera

and Peterson, 2017). However, it can also be interpreted as predicting a segment of the fundamental niche (Peterson, Papeş and Soberón, 2015). By considering important vectors and reservoirs for transmission to humans, regions of the realized niche of the disease, where transmission is higher, should be predicted. By including land use into the models, the human biotic interaction is being considered in all distribution models, and existing bias on the collection of triatomine occurrence data should allow the models to select suitable areas similar to where Chagas disease has already been found. Triatomines are collected and identified most of the time in the context of Chagas disease surveillance and research, thus available occurrence data is biased.

In this thesis a components based approach is being used to estimate disease distribution (Johnson, Escobar and Zambrana-Torrelío, 2019; Peterson, 2014), where the abiotic requirements of vectors and reservoir were represented by the WorldClim climatic variables and Biotic interactions are represented by the distribution models of vectors and reservoirs and the land use map as proxy for human activity. The idea of including vector and reservoir distributions and interactions into disease niche modelling has been done before. Peterson (2008), recognized how fundamental are biotic interactions for disease biogeography and how traditional niche modelling and species distribution modelling strategies have to address the specific case of disease distribution. Johnson, Escobar and Zambrana-Torrelío (2019), proposed a framework for modelling the geography of diseases. In this framework they recognize the importance of considering host-parasite interactions in the niche modelling of diseases and propose to include in the model; abiotic conditions required by hosts, host interactions, and current parasite populations. The modelling done in this thesis is a first step to an understanding of the distribution of Chagas disease in San Luis Potosi that could help inform decision makers.

There have been many studies before which model different species of vector, to understand various disease systems. Baak-Baak *et al.* (2017), modelled the distribution of various species of mosquitoes, vectors of dengue, in Yucatan, Mexico, to determine the human population at risk. Samy, Campbell and Peterson (2014), Modelled the niches of two species of sandflies and one species of *Leishmania* (the parasite that causes leishmaniasis), and compared the niches to better understand the geography of the transmission of cutaneous leishmaniasis. There have also been studies that take into account vectors and reservoirs together, but they are less frequent. Peterson *et al.* (2002), modelled many species of triatomines and woodrats (*Neotoma spp.*) to try to understand association between triatomines and woodrat species. The thesis presented here considers the distribution of vectors and reservoirs of Chagas disease for risk mapping, as well as human interactions in the system, represented by the land use map included in the predictor variables. This combination of factors has not been considered for Chagas disease distribution modelling

before thus, this research is a step forward in the understanding of multi-host disease systems and Chagas disease.

Regarding the conversion the models from a continuous output to a binary one (suitable/unsuitable), A threshold for each species to define the limit between a suitable and unsuitable cell was needed. There are many methods for doing this, from the fixed method where the researcher sets an arbitrary thresholds (usually 0.5), to methods based in the confusion matrix where it is possible to decide which kind of errors to minimize (Franklin, 2012; Peterson *et al.*, 2011). However, only a handful of methods are available for presence only methods like Maxent because most of the thresholds are based on the confusion matrix which is tricky to consider in presence only methods (Peterson *et al.*, 2011). For this thesis, the MTSPS was used as a threshold because it has shown good performance for presence only methods and it also minimizes both omission and commission errors by maximizing sensitivity and specificity (Liu *et al.*, 2005; Liu, White and Newell, 2013). The MTP approach was considered (Annex 4) but it has shown to be very sensitive to outliers in environmental space (Peterson *et al.*, 2011). Additionally, the fixed threshold approach produced very narrow areas of suitability (Annex 5) and has been proven to be the worst performing thresholding method (Liu *et al.*, 2005).

Omission rates between 0.08 and 0.24 were calculated from the binary predictions, being the lowest for *T. mexicana* and the highest for *D. virginiana*. This difference could be because normally it is more difficult to predict distributions of widely distributed generalist species, like *D. virginiana* than distributions of endemic species like *T. mexicana* (Franklin, 2012). Moreover, the high omission rates for *D. virginiana* could be caused by the occurrences used which were not cleaned as thoroughly as the data for triatomines due to time constraints. In contrast, data cleaning for triatomines was cleaned as planned and year data was considered for the selection of the final occurrence points to use in the models.

Regarding the technical aspects of data extraction from databases, transformation, cleaning and modelling, I consider important to highlight the fact that the R language and R Studio software, as well as the available packages related to environmental niche modelling (Aiello-Lammens *et al.*, 2015; Muscarella *et al.*, 2014; Zizka *et al.*, 2018) and data management (Wickham and Grolemond, 2017) were instrumental to the completion of this research. The occurrence points used in the modelling were chosen from 28,000 occurrence records of vectors and reservoirs. Selecting through this high quantity of records manually using spreadsheet software would have been a time consuming and prone to human error task. Additionally, R is an open source project free of cost and with an active Environmental Niche Modelling community that offers lectures, tutorials, and advice on the technical aspects of distribution modelling.

A limitation of this risk map is that, even if there exist vectors and reservoirs in an area, *T. cruzi* could still not be present due different biotic interactions not taken into account into the models like host immunity, or coinfection with other microorganisms which can affect the geographic distribution of a parasite (Johnson, Escobar and Zambrana-Torrelío, 2019). Nevertheless, this kind of predictions can still be a great tool for decision making and community intervention planning.

On the other hand, the contrary can still be possible. It may be possible to encounter vector species in unsuitable areas thanks to human assisted dispersal, like reported by López-Cárdenas *et al.* (2005). This kind of movement could allow species to stablish populations where previously absent. This event could be a possibility with the suitable area for vectors and reservoir predicted for the city of San Luis Potosi (Figure 18), in case there are not already vector and parasite populations. This kind of risk mapping could help identify new areas where vector or parasite populations may stablish.

Finally, this type of risk mapping could have important applications. Ibáñez-Cervantes *et al.* (2018), reported that between 2007 and 2016, reported Chagas disease cases in Mexico increased thanks to improvements in surveillance and diagnosis. Moreover, Chagas disease has been classified as an emergent urban zoonosis (Urdaneta-Morales, 2014). Species distribution model-based risk maps could be a useful tool to detect and prevent new cases in previously unexplored locations.

In conclusion, risk maps based on species distribution models are often used for predicting transmission risk due to vector exposure, however maps based on vector and reservoirs in multi-host disease systems are less frequent. In this thesis a relevant reservoir species was chosen, then suitable habitats for vectors and reservoirs were predicted, and then the predictions were combined into a risk map. These methods can prove to be useful for health authorities to better understand the distribution of Chagas disease in San Luis Potosí. These maps can also be used to prioritize Chagas disease surveillance and education.

Reference list

- Aiello-Lammens, M.E. *et al.* (2015) ‘spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models’, *Ecography*, 38(5), pp. 541–545. doi: 10.1111/ecog.01132
- Araújo, C.B. de, Marcondes-Machado, L.O. and Costa, G.C. (2014) ‘The importance of biotic interactions in species distribution models: a test of the Eltonian noise hypothesis using parrots’, *Journal of Biogeography*, 41(3), pp. 513–523. doi: 10.1111/jbi.12234
- Araújo, M.B. and Guisan, A. (2006) ‘Five (or so) challenges for species distribution modelling’, *Journal of Biogeography*, 33(10), pp. 1677–1688. doi: 10.1111/j.1365-2699.2006.01584.x
- Arnal, A. *et al.* (2019) ‘Estimating the current burden of Chagas disease in Mexico: A systematic review and meta-analysis of epidemiological surveys from 2006 to 2017’, *PLOS Neglected Tropical Diseases*, 13(4), e0006859. doi: 10.1371/journal.pntd.0006859
- Baak-Baak, C.M. *et al.* (2017) ‘Ecological Niche Model for Predicting Distribution of Disease-Vector Mosquitoes in Yucatán State, México’, *Journal of Medical Entomology*, 54(4), pp. 854–861. doi: 10.1093/jme/tjw243
- Beard, C.B. *et al.* (2003) ‘Chagas disease in a domestic transmission cycle, southern Texas, USA’, *Emerging Infectious Diseases*, 9(1), pp. 103–105. doi: 10.3201/eid0901.020217
- Becerril, M.A. *et al.* (2010) ‘Riesgo de transmisión de *Trypanosoma cruzi* en el municipio de Metztlán, estado de Hidalgo, México, mediante la caracterización de unidades domiciliarias y sus índices entomológicos’, *Neotropical Entomology*, 39(5), pp. 810–817. doi: 10.1590/s1519-566x2010000500021
- Carmona-Castro, O., Moo-Llanes, D.A. and Ramsey, J.M. (2018) ‘Impact of climate change on vector transmission of *Trypanosoma cruzi* (Chagas, 1909) in North America’, *Medical and Veterinary Entomology*, 32(1), pp. 84–101. doi: 10.1111/mve.12269
- Catalá, S.S., Noireau, F. and Dujardin, J.-P. (2017) ‘Biology of Triatominae’, in Telleria, J. and Tibayrenc, M. (eds.) *American trypanosomiasis Chagas disease: One hundred years of research*. Amsterdam Netherlands: Elsevier, pp. 145–167.
- Coura, J.R. (2015) ‘The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions--a comprehensive review’, *Memorias Do Instituto Oswaldo Cruz*, 110(3), pp. 277–282. doi: 10.1590/0074-0276140362
- Cruz-Reyes, A. and Pickering-López, J.M. (2006) ‘Chagas disease in Mexico: an analysis of geographical distribution during the past 76 years - A review’, *Memórias do Instituto Oswaldo Cruz*, 101(4), pp. 345–354. doi: 10.1590/S0074-02762006000400001
- Curtis-Robles, R. *et al.* (2018) ‘Bionomics and Spatial Distribution of Triatomine Vectors of *Trypanosoma cruzi* in Texas and Other Southern States, USA’, *The American Journal of Tropical Medicine and Hygiene*, 98(1), pp. 113–121. doi: 10.4269/ajtmh.17-0526

- Di Iorio, O. and Gürtler, R.E. (2017) ‘Seasonality and Temperature-Dependent Flight Dispersal of *Triatoma infestans* (Hemiptera: Reduviidae) and Other Vectors of Chagas Disease in Western Argentina’, *Journal of Medical Entomology*, 54(5), pp. 1285–1292. doi: 10.1093/jme/tjx109
- Dormann, C.F. *et al.* (2013) ‘Collinearity: a review of methods to deal with it and a simulation study evaluating their performance’, *Ecography*, 36(1), pp. 27–46. doi: 10.1111/j.1600-0587.2012.07348.x
- DUMONTEIL, E. *et al.* (2002) ‘Geographic distribution of *Triatoma dimidiata* and transmission dynamics of *Trypanosoma cruzi* in the Yucatan peninsula of Mexico’, *American Journal of Tropical Medicine and Hygiene*, 67(2), pp. 176–183. doi: 10.4269/ajtmh.2002.67.176
- Echeverria, L.E. and Morillo, C.A. (2019) ‘American Trypanosomiasis (Chagas Disease)’, *Infectious Disease Clinics of North America*, 33(1), pp. 119–134. doi: 10.1016/j.idc.2018.10.015
- Elith, J. *et al.* (2011) ‘A statistical explanation of MaxEnt for ecologists’, *Diversity and Distributions*, 17(1), pp. 43–57. doi: 10.1111/j.1472-4642.2010.00725.x
- Fick, S.E. and Hijmans, R.J. (2017) ‘WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas’, *International Journal of Climatology*, 37(12), pp. 4302–4315. doi: 10.1002/joc.5086
- Filigheddu, M.T., Górgolas, M. and Ramos, J.M. (2017) ‘Enfermedad de Chagas de transmisión oral’, *Medicina Clínica*, 148(3), pp. 125–131. doi: 10.1016/j.medcli.2016.10.038
- Flores, A. *et al.* (2017) ‘Temporal Variation in the Abundance and Timing of Daily Activity of Chagas Disease Vector *Triatoma gerstaeckeri* (Stål, 1859) in a Natural Habitat in the Lower Rio Grande Valley, South Texas’, *The Journal of Parasitology*, 103(5), pp. 574–578. doi: 10.1645/17-50
- Forero, D.A. *et al.* (2019) ‘Ten simple rules for carrying out and writing meta-analyses’, *PLoS Computational Biology*, 15(5), e1006922. doi: 10.1371/journal.pcbi.1006922
- Franklin, J. (2012) *Mapping species distributions: Spatial inference and prediction*. (Ecology, biodiversity and conservation). Cambridge: Cambridge University Press.
- Galaviz-Silva, L. *et al.* (2017) ‘Prevalence of *Trypanosoma cruzi* infection in dogs and small mammals in Nuevo León, Mexico’, *Revista Argentina De Microbiología*, 49(3), pp. 216–223. doi: 10.1016/j.ram.2016.11.006
- Galvão, C. and Justi, S.A. (2015) ‘An overview on the ecology of Triatominae (Hemiptera:Reduviidae)’, *Acta Tropica*, 151, pp. 116–125. doi: 10.1016/j.actatropica.2015.06.006
- García, E. (1998) *Modificaciones al sistema de clasificación climática de Köppen*. Mexico: CONABIO.

- Garza, M. *et al.* (2014) 'Projected future distributions of vectors of *Trypanosoma cruzi* in North America under climate change scenarios', *PLOS Neglected Tropical Diseases*, 8(5), e2818. doi: 10.1371/journal.pntd.0002818
- GBIF.org (2020a) 'Occurrence Download'. doi: 10.15468/DL.V4G7UM
- GBIF.org (2020b) 'Occurrence Download'. doi: 10.15468/DL.DX7FUQ
- GBIF.org (2020c) 'Occurrence Download'. doi: 10.15468/DL.9KY97N
- GBIF.org (2020d) 'Occurrence Download'. doi: 10.15468/DL.9U3RXY
- González, A.M. (2020) 'Photograph of *Triatoma* individuals.'
- Guisan, A. and Thuiller, W. (2005) 'Predicting species distribution: offering more than simple habitat models', *Ecology Letters*, 8(9), pp. 993–1009. doi: 10.1111/j.1461-0248.2005.00792.x
- Hagenlocher, M. *et al.* (2014) 'Implications of Spatial Scales and Reporting Units for the Spatial Modelling of Vulnerability to Vector-borne Diseases', *Geospatial innovation for society, GI_Forum 2014*. Geoinformatics Forum, Salzburg. Berlin: Wichmann; Verl. der Österr. Akad. der Wiss., ÖAW, pp. 197–206. doi: 10.1553/giscience2014s197
- Hammer, C.C. *et al.* (2019) '(Re-) conceptualising vulnerability as a part of risk in global health emergency response: Updating the pressure and release model for global health emergencies', *Emerging Themes in Epidemiology*, 16, p. 2. doi: 10.1186/s12982-019-0084-3
- Hernández-Cortazar, I. *et al.* (2018) 'Frequency of *Trypanosoma cruzi* Infection in Synanthropic and Wild Rodents Captured in a Rural Community in Southeast of Mexico', *Veterinary Medicine International*, 2018, p. 8059613. doi: 10.1155/2018/8059613
- Herrera, L. (2010) 'Una revisión sobre reservorios de *Trypanosoma* (*Schizotrypanum*) *cruzi* (Chagas, 1909), agente etiológico de la Enfermedad de Chagas', *Boletín de Malariología y Salud Ambiental*, 50(1), pp. 3–15.
- Herrera, L. (2014) '*Trypanosoma cruzi*, the Causal Agent of Chagas Disease: Boundaries between Wild and Domestic Cycles in Venezuela', *Frontiers in public health*, 2, p. 259. doi: 10.3389/fpubh.2014.00259
- Hosseini, P.R. *et al.* (2017) 'Does the impact of biodiversity differ between emerging and endemic pathogens? The need to separate the concepts of hazard and risk', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 372(1722). doi: 10.1098/rstb.2016.0129
- Hotez, P.J. *et al.* (2008) 'The Neglected Tropical Diseases of Latin America and the Caribbean: A Review of Disease Burden and Distribution and a Roadmap for Control and Elimination', *PLOS Neglected Tropical Diseases*, 2(9), e300. doi: 10.1371/journal.pntd.0000300
- Hutchinson, G.E. (1957) 'Concluding Remarks', *Cold Spring Harbor Symposia on Quantitative Biology* (22), pp. 415–427.

- Ibáñez-Cervantes, G. *et al.* (2018) 'Evolution of incidence and geographical distribution of Chagas disease in Mexico during a decade (2007-2016)', *Epidemiology and Infection*, pp. 1–7. doi: 10.1017/S0950268818002984
- Jansen, A.M., Xavier, S.C.d.C. and Roque, A.L.R. (2018) 'Trypanosoma cruzi transmission in the wild and its most important reservoir hosts in Brazil', *Parasites & Vectors*, 11(1), p. 502. doi: 10.1186/s13071-018-3067-2
- Jiménez-Coello, M. *et al.* (2012) 'American trypanosomiasis infection in fattening pigs from the south-east of Mexico', *Zoonoses and Public Health*, 59 Suppl 2, pp. 166–169. doi: 10.1111/zph.12016
- Johnson, E.E., Escobar, L.E. and Zambrana-Torrel, C. (2019) 'An Ecological Framework for Modeling the Geography of Disease Transmission', *Trends in Ecology & Evolution*, 34(7), pp. 655–668. doi: 10.1016/j.tree.2019.03.004
- Johnson, P.T.J., Roode, J.C. de and Fenton, A. (2015) 'Why infectious disease research needs community ecology', *Science*, 349(6252), p. 1259504. doi: 10.1126/science.1259504
- Juarez-Tobias, S. *et al.* (2009) 'Seroprevalence of Trypanosoma cruzi Among Teenek Amerindian Residents of the Huasteca Region in San Luis Potosi, Mexico', *The American Journal of Tropical Medicine and Hygiene*, 81(2), pp. 219–222. doi: 10.4269/ajtmh.2009.81.219
- Jurberg, J. and Galvao, C. (2006) 'Biology, ecology, and systematics of Triatominae (Heteroptera, Reduviidae), vectors of Chagas disease, and implications for human health', in *Hug The Bug*, p. 1184.
- Kjos, S.A. *et al.* (2013) 'Identification of bloodmeal sources and Trypanosoma cruzi infection in triatomine bugs (Hemiptera: Reduviidae) from residential settings in Texas, the United States', *Journal of Medical Entomology*, 50(5), pp. 1126–1139. doi: 10.1603/me12242
- Kjos, S.A., Snowden, K.F. and Olson, J.K. (2009) 'Biogeography and Trypanosoma cruzi infection prevalence of Chagas disease vectors in Texas, USA', *Vector Borne and Zoonotic Diseases (Larchmont, N.Y.)*, 9(1), pp. 41–50. doi: 10.1089/vbz.2008.0026
- Koyoc-Cardena, E. *et al.* (2015) 'CHICKEN COOPS, Triatoma dimidiata INFESTATION AND ITS INFECTION WITH Trypanosoma cruzi IN A RURAL VILLAGE OF YUCATAN, MEXICO', *Revista Do Instituto De Medicina Tropical De Sao Paulo*, 57(3), pp. 269–272. doi: 10.1590/S0036-46652015000300015
- Legey, A.P. *et al.* (2003) 'Trypanosoma cruzi in marsupial didelphids (Philander frenata and Didelphis marsupialis): differences in the humoral immune response in natural and experimental infections', *Revista Da Sociedade Brasileira De Medicina Tropical*, 36(2), pp. 241–248. doi: 10.1590/s0037-86822003000200008
- Lent, H. and Wygodzinsky, P. (1979) 'Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease', *Bulletin of the American museum of natural history*, 163(3), pp. 123–520.

- Liu, C. *et al.* (2005) 'Selecting thresholds of occurrence in the prediction of species distributions', *Ecography*, 28(3), pp. 385–393. doi: 10.1111/j.0906-7590.2005.03957.x
- Liu, C., White, M. and Newell, G. (2013) 'Selecting thresholds for the prediction of species occurrence with presence-only data', *Journal of Biogeography*, 40(4), pp. 778–789. doi: 10.1111/jbi.12058
- Longoni, S.S. *et al.* (2012) 'Detection of different *Leishmania* spp. and *Trypanosoma cruzi* antibodies in cats from the Yucatan Peninsula (Mexico) using an iron superoxide dismutase excreted as antigen', *Comparative Immunology, Microbiology and Infectious Diseases*, 35(5), pp. 469–476. doi: 10.1016/j.cimid.2012.04.003
- López-Cancino, S.A. *et al.* (2015) 'Landscape ecology of *Trypanosoma cruzi* in the southern Yucatan Peninsula', *Acta Tropica*, 151, pp. 58–72. doi: 10.1016/j.actatropica.2015.07.021
- López-Cárdenas, J. *et al.* (2005) 'Fine-Scale Predictions of Distributions of Chagas Disease Vectors in the State of Guanajuato, Mexico', *Journal of Medical Entomology*, 42(6), pp. 1068–1081. doi: 10.1603/0022-2585(2005)042[1068:FPODOC]2.0.CO;2
- Manne, J.M. *et al.* (2013) 'Barriers to treatment access for Chagas disease in Mexico', *PLOS Neglected Tropical Diseases*, 7(10), e2488. doi: 10.1371/journal.pntd.0002488
- Martínez de la Vega, G. *et al.* (2016) 'Mamíferos de San Luis Potosí, México', in Briones-Salas, M. *et al.* (eds.) *Riqueza y Conservación de los Mamíferos en México a Nivel Estatal*. Mexico City, pp. 367–404.
- Martínez-Hernández, F. *et al.* (2014) 'Follow up of natural infection with *Trypanosoma cruzi* in two mammals species, *Nasua narica* and *Procyon lotor* (Carnivora: Procyonidae): evidence of infection control?' *Parasites & Vectors*, 7, p. 405. doi: 10.1186/1756-3305-7-405
- Martínez-Ibarra, J.A. *et al.* (2008) 'Influence of temperature and humidity on the biology of *Triatoma mexicana* (Hemiptera: Reduviidae: Triatominae) under laboratory conditions', *Memórias do Instituto Oswaldo Cruz*, 103(7), pp. 719–723. doi: 10.1590/S0074-02762008000700015
- MARTÍNEZ-TOVAR, J.G. *et al.* (2013) 'Nuevos registros geográficos y notas de infección de *Triatoma gerstaeckeri* (Stål) y *Triatoma rubida* (Uhler) (Hemiptera: Reduviidae: Triatominae) en Nuevo León y Coahuila, México', *Acta Zoológica Mexicana*, 29(1), pp. 227–233. Available at: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0065-17372013000100012.
- McPhatter, L. *et al.* (2012) *Vector Surveillance to Determine Species Composition and Occurrence of Trypanosoma cruzi Infection at Three Military Installations in San Antonio, Texas*. Available at: <https://apps.dtic.mil/dtic/tr/fulltext/u2/a568322.pdf>.
- Medina-Garza, H. *et al.* (2016) 'Patrón de distribución espacial de la enfermedad de Chagas en la Huasteca Potosina', *Entomología Médica y Forense* (3), pp. 682–688.

- Merow, C., Smith, M.J. and Silander, J.A. (2013) 'A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter', *Ecography*, 36(10), pp. 1058–1069. doi: 10.1111/j.1600-0587.2013.07872.x
- Moo-Millan, J.I. *et al.* (2019) 'Disentangling *Trypanosoma cruzi* transmission cycle dynamics through the identification of blood meal sources of natural populations of *Triatoma dimidiata* in Yucatán, Mexico', *Parasites & Vectors*, 12(1), p. 572. doi: 10.1186/s13071-019-3819-7
- Muscarella, R. *et al.* (2014) 'ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models', *Methods in Ecology and Evolution*, 5(11), pp. 1198–1205. doi: 10.1111/2041-210X.12261
- Oscar Aldana Cruz *et al.* (2009) 'Seroprevalencia de la enfermedad de Chagas en Tamazunchale San Luis Potosí', *Enfermedades Infecciosas y Microbiología*, 29(3), pp. 107–110. Available at: <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=26750>.
- Pacheco-Tucuch, F.S. *et al.* (2012) 'Public street lights increase house infestation by the Chagas disease vector *Triatoma dimidiata*', *PLOS ONE*, 7(4), e36207. doi: 10.1371/journal.pone.0036207
- Panti-May, J.A. *et al.* (2017) 'A survey of zoonotic pathogens carried by house mouse and black rat populations in Yucatan, Mexico', *Epidemiology and Infection*, 145(11), pp. 2287–2295. doi: 10.1017/S0950268817001352
- Parra-Henao, G. *et al.* (2015) 'House-level risk factors for *Triatoma dimidiata* infestation in Colombia', *The American Journal of Tropical Medicine and Hygiene*, 92(1), pp. 193–200. doi: 10.4269/ajtmh.14-0273
- Peel, M.C., Finlayson, B.L. and McMahon, T.A. (2007) 'Updated world map of the Köppen-Geiger climate classification', *Hydrology and Earth System Sciences*, 11(5), pp. 1633–1644. doi: 10.5194/hess-11-1633-2007
- Pérez-Molina, J.A. and Molina, I. (2018) 'Chagas disease', *The Lancet*, 391(10115), pp. 82–94. doi: 10.1016/S0140-6736(17)31612-4
- Peterson, A.T. *et al.* (2002) 'Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico', *Emerging Infectious Diseases*, 8(7), pp. 662–667. doi: 10.3201/eid0807.010454
- Peterson, A.T. (2008) 'Biogeography of diseases: a framework for analysis', *Die Naturwissenschaften*, 95(6), pp. 483–491. doi: 10.1007/s00114-008-0352-5
- Peterson, A.T. *et al.* (2011) *Ecological niches and geographic distributions*. (Monographs in population biology, 49). Princeton N.J.: Princeton University Press.
- Peterson, A.T. (2014) *Mapping disease transmission risk: Enriching models using biogeography and ecology*. Baltimore: Johns Hopkins University Press.

- Peterson, A.T., Papeş, M. and Soberón, J. (2015) ‘Mechanistic and Correlative Models of Ecological Niches’, *European Journal of Ecology*, 1(2), pp. 28–38. doi: 10.1515/eje-2015-0014
- Peterson, A.T. and Soberón, J. (2012) ‘Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right’, *Natureza a Conservacao*, 10(2), pp. 102–107. doi: 10.4322/natcon.2012.019
- Phillips, S.J., Anderson, R.P. and Schapire, R.E. (2006) ‘Maximum entropy modeling of species geographic distributions’, *Ecological Modelling*, 190(3-4), pp. 231–259. doi: 10.1016/j.ecolmodel.2005.03.026
- Phillips, S.J., Dudík, M. and Schapire, R.E. (2017) *Maxent software for modeling species niches and distributions*. Available at: http://biodiversityinformatics.amnh.org/open_source/maxent/ (Accessed: 22 July 2020).
- Pippin, W.F. (1970) ‘The biology and vector capability of *Triatoma sanguisuga texana* usinger and *Triatoma gerstaeckeri* (Stål) compared with *Rhodnius prolixus* (Stål) (Hemiptera: Triatominae)’, *Journal of Medical Entomology*, 7(1), pp. 30–45. doi: 10.1093/jmedent/7.1.30
- Porta, M. (2014) *A dictionary of Epidemiology*. 6th edn. New York: Oxford University Press.
- Quirós-Gómez, Ó. *et al.* (2017) ‘*Triatoma dimidiata* in Colombia: Distribution, ecology and epidemiological importance’, *Biomedica : revista del Instituto Nacional de Salud*, 37(2), pp. 274–285. doi: 10.7705/biomedica.v37i2.2893
- Radosavljevic, A. and Anderson, R.P. (2014) ‘Making better Maxent models of species distributions: complexity, overfitting and evaluation’, *Journal of Biogeography*, 41(4), pp. 629–643. doi: 10.1111/jbi.12227
- Raedig, C. (2010) *Distribution patterns of Neotropical angiosperm species: tackling difficulties of large-scale research*. Doctoral dissertation. Universität Leipzig.
- Raedig, C., Ho Dac Thai, H. and Nehren, U. (2017) ‘Connectivity Conservation Management: A Biodiversity Corridor for Central Vietnam’, in Nauditt, A. and Ribbe, L. (eds.) *Land Use and Climate Change Interactions in Central Vietnam: LUCCi*. (Water Resources Development and Management). Singapore: Springer Singapore, pp. 69–81.
- Ramsey, J.M. *et al.* (2012) ‘Ecological connectivity of *Trypanosoma cruzi* reservoirs and *Triatoma pallidipennis* hosts in an anthropogenic landscape with endemic Chagas disease’, *PLOS ONE*, 7(9), e46013. doi: 10.1371/journal.pone.0046013
- Ramsey, J.M. *et al.* (2015) ‘Atlas of Mexican Triatominae (Reduviidae: Hemiptera) and vector transmission of Chagas disease’, *Memórias do Instituto Oswaldo Cruz*, 110(3), pp. 339–352. doi: 10.1590/0074-02760140404
- Rezedowski, J. (1965) ‘Vegetación del Estado de San Luis Potosí’, *Acta Científica Potosina*, 5(1, 2), pp. 1–227.

- Roque, A.L.R. *et al.* (2008) 'Trypanosoma cruzi transmission cycle among wild and domestic mammals in three areas of orally transmitted Chagas disease outbreaks', *The American Journal of Tropical Medicine and Hygiene*, 79(5), pp. 742–749.
- Rosal, G.G. *et al.* (2018) 'Chagas disease: Importance of rats as reservoir hosts of Trypanosoma cruzi (Chagas, 1909) in western Mexico', *Journal of Infection and Public Health*, 11(2), pp. 230–233. doi: 10.1016/j.jiph.2017.07.017
- Rovirosa-Hernández, M.d.J. *et al.* (2013) 'Seroprevalence of Trypanosoma cruzi and Leishmania mexicana in free-ranging howler monkeys in southeastern Mexico', *American Journal of Primatology*, 75(2), pp. 161–169. doi: 10.1002/ajp.22094
- Ruiz-Piña, H.A. *et al.* (2018a) 'Prevalence of Trypanosoma cruzi in backyard mammals from a rural community of Yucatan, Mexico', *Tropical and Subtropical Agroecosystems*, 21(3). Available at: <http://www.revista.ccba.uady.mx/ojs/index.php/TSA/article/download/1586/1207>.
- Ruiz-Piña, H.A. *et al.* (2018b) 'Prevalence of Trypanosoma cruzi, in farm animals kept in the backyard system in the community of Molas, Yucatan, Mexico', *Tropical and Subtropical Agroecosystems*, 21(3). Available at: <http://www.revista.ccba.uady.mx/ojs/index.php/TSA/article/download/1586/1207>.
- Ruiz-Piña, H.A. and Cruz-Reyes, A. (2002) 'The opossum Didelphis virginiana as a synanthropic reservoir of Trypanosoma cruzi in Dzidzilché, Yucatán, México', *Memórias do Instituto Oswaldo Cruz*, 97(5), pp. 613–620. doi: 10.1590/S0074-02762002000500003
- Salazar Schettino, P.M. *et al.* (2007) 'Triatoma mexicana (Hemiptera: Reduviidae) in Guanajuato, Mexico: house infestation and seasonal variation', *Memórias do Instituto Oswaldo Cruz*, 102(7), pp. 803–807. doi: 10.1590/S0074-02762007005000118
- Samy, A.M., Campbell, L.P. and Peterson, A.T. (2014) 'Leishmaniasis transmission: distribution and coarse-resolution ecology of two vectors and two parasites in Egypt', *Revista Da Sociedade Brasileira De Medicina Tropical*, 47(1), pp. 57–62. doi: 10.1590/0037-8682-0189-2013
- Sandoval-ruiz, C.A., Zumaquero-rios, J.L. and Rojas-soto, O.R. (2008) 'Predicting Geographic and Ecological Distributions of Triatomine Species in the Southern Mexican State of Puebla Using Ecological Niche Modeling', *Journal of Medical Entomology*, 45(3), pp. 540–546. doi: 10.1603/0022-2585(2008)45[540:PGAEDO]2.0.CO;2
- SANDOVAL-RUIZ, C.A. *et al.* (2012) 'The Triatominae (Hemiptera: Heteroptera: Reduviidae) of Veracruz, Mexico: geographic distribution, taxonomic redescriptions, and a key', *Zootaxa*, 3487(1), p. 1. doi: 10.11646/zootaxa.3487.1.1
- Schofield, C.J. and Galvão, C. (2009) 'Classification, evolution, and species groups within the Triatominae', *Acta Tropica*, 110(2-3), pp. 88–100. doi: 10.1016/j.actatropica.2009.01.010

Secretaría de Salud (2015) 'Norma Oficial Mexicana para la vigilancia epidemiológica, promoción, prevención y control de las enfermedades transmitidas por vectores: NOM-032-SSA2-2014', 16 April.

Soberón, J., Osorio-Olvera, L. and Peterson, T. (2017) 'Diferencias conceptuales entre modelación de nichos y modelación de áreas de distribución', *Revista Mexicana de Biodiversidad*, 88(2), pp. 437–441. doi: 10.1016/j.rmb.2017.03.011

Solís-Franco, R.R., Romo-Zapata, J.A. and Martínez-Ibarra, J.A. (1997) 'Wild Reservoirs Infected by *Trypanosoma cruzi* in the Ecological Park "El Zapotal", Tuxtla Gutiérrez, Chiapas, México', *Memórias do Instituto Oswaldo Cruz*, 92(2), pp. 163–164. doi: 10.1590/S0074-02761997000200006

Thurman, D.C. (1945) 'The Biology of *Triatoma gerstaeckeri*', *Journal of Economic Entomology*, 38(5), pp. 597–598. doi: 10.1093/jee/38.5.597a

Urdaneta-Morales, S. (2014) 'Chagas' disease: An emergent urban zoonosis. The caracas valley (Venezuela) as an epidemiological model', *Frontiers in Public Health*, 2, p. 265. doi: 10.3389/fpubh.2014.00265

Villagrán, M.E. *et al.* (2008) 'Natural infection and distribution of triatomines (Hemiptera: Reduviidae) in the state of Querétaro, Mexico', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(8), pp. 833–838. doi: 10.1016/j.trstmh.2008.05.005

Villegas-García, J.C. and Santillán-Alarcón, S. (2004) 'American trypanosomiasis in central Mexico: *Trypanosoma cruzi* infection in triatomine bugs and mammals from the municipality of Jiutepec in the state of Morelos', *Annals of Tropical Medicine and Parasitology*, 98(5), pp. 529–532. doi: 10.1179/000349804225003497

Warren, D.L. and Seifert, S.N. (2011) 'Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria', *Ecological Applications : a Publication of the Ecological Society of America*, 21(2), pp. 335–342. doi: 10.1890/10-1171.1

WHO (2018) *Chagas disease (American trypanosomiasis)*, 1 February. Available at: <https://www.who.int/chagas/disease/> (Accessed: 23 January 2019).

Wickham, H. and Grolemund, G. (2017) *R for data science: Import, tidy, transform, visualize, and model data*. Sebastopol CA: O'Reilly.

Wisz, M.S. *et al.* (2008) 'Effects of sample size on the performance of species distribution models', *Diversity and Distributions*, 14(5), pp. 763–773. doi: 10.1111/j.1472-4642.2008.00482.x

Wormington, J.D. *et al.* (2018) 'Daily activity patterns of movement and refuge use in *Triatoma gerstaeckeri* and *Rhodnius prolixus* (Hemiptera: Reduviidae), vectors of the Chagas disease parasite', *Acta Tropica*, 185, pp. 301–306. doi: 10.1016/j.actatropica.2018.06.012

YEO, M. *et al.* (2005) 'Origins of Chagas disease: *Didelphis* species are natural hosts of *Trypanosoma cruzi* I and armadillos hosts of *Trypanosoma cruzi* II, including hybrids',

International Journal for Parasitology, 35(2), pp. 225–233.
doi: 10.1016/j.ijpara.2004.10.024

Zamora-Ledesma, S. *et al.* (2016) ‘Presence of trypanosomatid antibodies in gray foxes (*Urocyon cinereoargenteus*) and domestic and feral dogs (*Canis lupus familiaris*) in Queretaro, Mexico’, *Veterinary Parasitology: Regional Studies and Reports*, 5, pp. 25–30.
doi: 10.1016/j.vprsr.2016.08.005

Zeledón, R. *et al.* (1970) ‘Biology and ethology of *Triatoma dimidiata* (Latreille, 1811). I. Life cycle, amount of blood ingested, resistance of starvation, and size of adults: I. Life cycle, amount of blood ingested, resistance to starvation, and size of adults¹’, *Journal of Medical Entomology*, 7(3), pp. 313–319. doi: 10.1093/jmedent/7.3.313

Zeledón, R. *et al.* (1973) ‘Biology and ethology of *Triatoma dimidiata* (Latreille, 1811). 3. Habitat and blood sources’, *Journal of Medical Entomology*, 10(4), pp. 363–370.
doi: 10.1093/jmedent/10.4.363

Zizka, A. *et al.* (2018) ‘CoordinateCleaner : Standardized cleaning of occurrence records from biological collection databases’, *Methods in Ecology and Evolution*, 10(5), pp. 744–751. doi: 10.1111/2041-210X.13152

Annexes

Annex 1: Pearson correlation matrix for WorldClim variables in Mexican territory. Only pairs with correlation coefficients higher than 0.7 and lower than -0.7 are shown.

Row names	Column names	Pearson coefficient	Complete name of row variable	Complete name of column
BIO4	BIO3	-0.90	Temperature Seasonality (standard deviation x100)	Isothermality (BIO2/BIO7) (x100)
BIO6	BIO1	0.88	Min Temperature of Coldest Month	Annual Mean Temperature
BIO6	BIO2	-0.77	Min Temperature of Coldest Month	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO7	BIO2	0.82	Temperature Annual Range (BIO5-BIO6)	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO7	BIO3	-0.72	Temperature Annual Range (BIO5-BIO6)	Isothermality (BIO2/BIO7) (x100)
BIO7	BIO4	0.91	Temperature Annual Range (BIO5-BIO6)	Temperature Seasonality (standard deviation x100)
BIO7	BIO6	-0.81	Temperature Annual Range (BIO5-BIO6)	Min Temperature of Coldest Month
BIO8	BIO1	0.78	Mean Temperature of Wettest Quarter	Annual Mean Temperature
BIO8	BIO5	0.81	Mean Temperature of Wettest Quarter	Max Temperature of Warmest Month
BIO9	BIO1	0.88	Mean Temperature of Driest Quarter	Annual Mean Temperature

BIO9	BIO6	0.83	Mean Temperature of Driest Quarter	Min Temperature of Coldest Month
BIO10	BIO1	0.83	Mean Temperature of Warmest Quarter	Annual Mean Temperature
BIO10	BIO5	0.94	Mean Temperature of Warmest Quarter	Max Temperature of Warmest Month
BIO10	BIO8	0.88	Mean Temperature of Warmest Quarter	Mean Temperature of Wettest Quarter
BIO10	BIO9	0.70	Mean Temperature of Warmest Quarter	Mean Temperature of Driest Quarter
BIO11	BIO1	0.90	Mean Temperature of Coldest Quarter	Annual Mean Temperature
BIO11	BIO6	0.98	Mean Temperature of Coldest Quarter	Min Temperature of Coldest Month
BIO11	BIO7	-0.75	Mean Temperature of Coldest Quarter	Temperature Annual Range (BIO5-BIO6)
BIO11	BIO9	0.85	Mean Temperature of Coldest Quarter	Mean Temperature of Driest Quarter
BIO12	BIO7	-0.71	Annual Precipitation	Temperature Annual Range (BIO5-BIO6)
BIO13	BIO12	0.95	Precipitation of Wettest Month	Annual Precipitation
BIO14	BIO12	0.76	Precipitation of Driest Month	Annual Precipitation
BIO16	BIO12	0.95	Precipitation of Wettest Quarter	Annual Precipitation

BIO16	BIO13	0.99	Precipitation of Wettest Quarter	Precipitation of Wettest Month
BIO17	BIO12	0.77	Precipitation of Driest Quarter	Annual Precipitation
BIO17	BIO14	0.99	Precipitation of Driest Quarter	Precipitation of Driest Month
BIO18	BIO12	0.79	Precipitation of Warmest Quarter	Annual Precipitation
BIO18	BIO13	0.85	Precipitation of Warmest Quarter	Precipitation of Wettest Month
BIO18	BIO16	0.84	Precipitation of Warmest Quarter	Precipitation of Wettest Quarter
BIO19	BIO12	0.71	Precipitation of Coldest Quarter	Annual Precipitation
BIO19	BIO14	0.84	Precipitation of Coldest Quarter	Precipitation of Driest Month
BIO19	BIO17	0.88	Precipitation of Coldest Quarter	Precipitation of Driest Quarter

Annex 2: Parameters and evaluation statistics of all models ran in the “ENMeval” package.

#	Species	Features	Regularization multiplier	Train AUC	Average test AUC	Average test omission rate MTP	Average test omission rate 10%	AICc	Delta AICc
1	<i>T. dimidiata</i>	LQH	0.25	0.9027	0.8865919	0.00776442	0.11629808	17205.2033	125.716038
2	<i>T. dimidiata</i>	LQHP	0.25	0.9056	0.89028372	0.00466346	0.11790865	17160.1846	80.6973044
3	<i>T. dimidiata</i>	LQH	1	0.8967	0.88747018	0.00463942	0.11480769	17126.6857	47.1984419
4	<i>T. dimidiata</i>	LQHP	1	0.9001	0.89176603	0.00310096	0.11170673	17119.5341	40.0468355
5	<i>T. dimidiata</i>	LQH	3	0.8877	0.88189383	0.0015625	0.11326923	17079.4873	0
6	<i>T. dimidiata</i>	LQHP	3	0.8908	0.8846326	0.0015625	0.11168269	17095.3862	15.8988493
7	<i>T. dimidiata</i>	LQH	5	0.8836	0.87801789	0.0015625	0.10862981	17191.6023	112.114945
8	<i>T. dimidiata</i>	LQHP	5	0.8859	0.87932664	0.003125	0.11173077	17231.9348	152.447521
9	<i>T. dimidiata</i>	LQH	10	0.8743	0.86750851	0.0015625	0.11329327	17368.9399	289.452635
10	<i>T. dimidiata</i>	LQHP	10	0.8751	0.86876292	0.0015625	0.10711538	17548.9822	469.494852
11	<i>T. gerstaeckeri</i>	LQH	0.25	0.9598	0.92383437	0.03666667	0.16333333	NA	NA
12	<i>T. gerstaeckeri</i>	LQHP	0.25	0.966	0.92492923	0.03666667	0.22	NA	NA
13	<i>T. gerstaeckeri</i>	LQH	1	0.9398	0.92326424	0.03666667	0.14666667	1716.78423	371.688325
14	<i>T. gerstaeckeri</i>	LQHP	1	0.9433	0.92839242	0.01666667	0.14666667	1414.00386	68.9079529
15	<i>T. gerstaeckeri</i>	LQH	3	0.9343	0.92494259	0.03666667	0.13	1345.0959	0
16	<i>T. gerstaeckeri</i>	LQHP	3	0.9337	0.92420522	0.01666667	0.13	1397.41755	52.32164
17	<i>T. gerstaeckeri</i>	LQH	5	0.9291	0.91879131	0.01666667	0.13	1381.45659	36.3606872
18	<i>T. gerstaeckeri</i>	LQHP	5	0.9282	0.91679952	0.01666667	0.11333333	1422.80923	77.7133202
19	<i>T. gerstaeckeri</i>	LQH	10	0.9045	0.89175179	0.01666667	0.09333333	1501.35158	156.255676
20	<i>T. gerstaeckeri</i>	LQHP	10	0.9169	0.90203752	0.01666667	0.09333333	1481.72644	136.630532
21	<i>T. mexicana</i>	LQH	0.25	0.9624	0.95178901	0.01194296	0.12593583	9224.59194	1675.24777
22	<i>T. mexicana</i>	LQHP	0.25	0.9623	0.9516982	0.01194296	0.12290553	9191.15987	1641.8157
23	<i>T. mexicana</i>	LQH	1	0.9599	0.95247448	0.00900178	0.11693405	7900.66725	351.323084
24	<i>T. mexicana</i>	LQHP	1	0.9597	0.95231067	0.00900178	0.11693405	7868.60625	319.262083

25	<i>T. mexicana</i>	LQH	3	0.9544	0.94970434	0.00900178	0.12290553	7572.6998	23.35563
26	<i>T. mexicana</i>	LQHP	3	0.9542	0.94987723	0.00900178	0.12290553	7549.34417	0
27	<i>T. mexicana</i>	LQH	5	0.9517	0.94836305	0.00597148	0.10766488	7682.37902	133.034855
28	<i>T. mexicana</i>	LQHP	5	0.9516	0.9482164	0.0030303	0.10766488	7655.38507	106.0409
29	<i>T. mexicana</i>	LQH	10	0.9475	0.94546215	0.0030303	0.10178253	7855.79539	306.451225
30	<i>T. mexicana</i>	LQHP	10	0.947	0.94396167	0.0030303	0.10178253	7832.27757	282.933401
31	<i>D. virginiana</i>	LQH	0.25	0.5076	0.76799567	0.02	0.16033333	NA	NA
32	<i>D. virginiana</i>	LQHP	0.25	0.5989	0.75180664	0.02	0.15633333	NA	NA
33	<i>D. virginiana</i>	LQH	1	0.8506	0.8281279	0.008	0.132	7048.70123	162.287068
34	<i>D. virginiana</i>	LQHP	1	0.8491	0.82701622	0.008	0.136	7005.66105	119.246889
35	<i>D. virginiana</i>	LQH	3	0.8324	0.81725908	0.008	0.132	6900.38087	13.9667045
36	<i>D. virginiana</i>	LQHP	3	0.8319	0.81654378	0.008	0.14	6896.89118	10.4770174
37	<i>D. virginiana</i>	LQH	5	0.8267	0.8157162	0.008	0.116	6894.71543	8.30126426
38	<i>D. virginiana</i>	LQHP	5	0.8276	0.81355938	0.008	0.128	6886.41416	0
39	<i>D. virginiana</i>	LQH	10	0.8188	0.81232923	0.008	0.112	6930.78524	44.3710751
40	<i>D. virginiana</i>	LQHP	10	0.819	0.80797056	0.008	0.124	6934.43818	48.024013

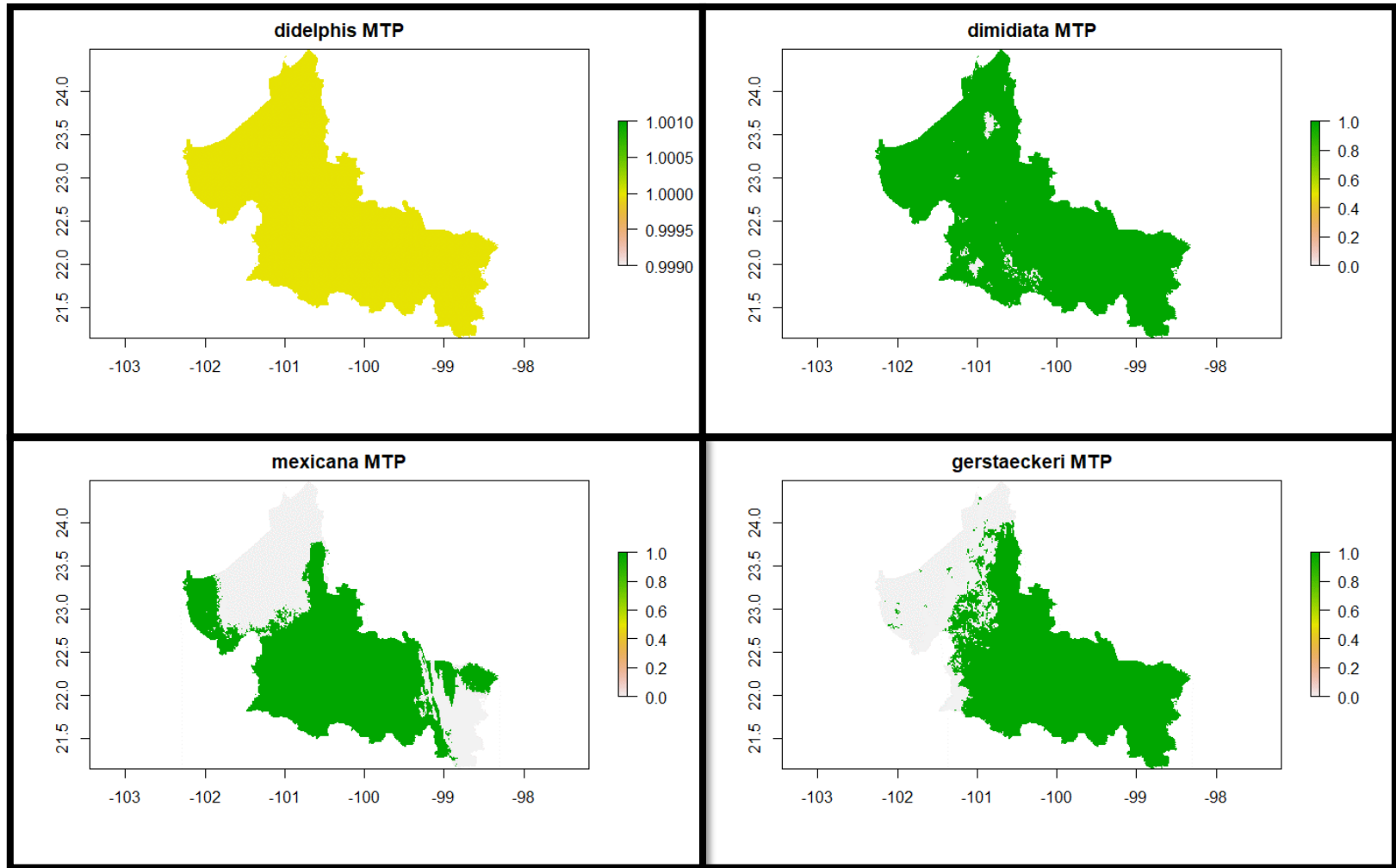
Annex 3: Evaluation statistics for all species distribution models made.

Species	#Training samples	Training AUC	#Test samples	Test AUC	MTP Logistic threshold	MTP test omission	MTP binomial probability	10PTP Logistic threshold	10PTP test omission	10PTP binomial probability	Max test SSS Logistic threshold	Max test SSS test omission	Max test SSS probability
Triatoma_dimidiata_0	579	0.8654	65	0.8455	0.0147	0.0154	0.007019	0.2688	0.0923	5.741E-26	0.3721	0.1385	8.958E-35
Triatoma_dimidiata_1	579	0.8643	65	0.8531	0.0125	0	0.01459	0.2719	0.1538	1.078E-21	0.364	0.2	4.53E-28
Triatoma_dimidiata_2	579	0.8614	65	0.8877	0.0192	0	0.003561	0.2752	0.0615	1.023E-28	0.3234	0.0615	1.059E-35
Triatoma_dimidiata_3	579	0.8651	65	0.856	0.0131	0	0.009421	0.2757	0.1385	8.295E-24	0.3573	0.1538	1.523E-31
Triatoma_dimidiata_4	580	0.868	64	0.8171	0.0148	0.0156	0.01151	0.2699	0.1406	3.046E-23	0.3692	0.2031	2.658E-29
Triatoma_dimidiata_5	580	0.8634	64	0.8633	0.0155	0	0.008198	0.2723	0.0938	1.171E-26	0.3171	0.125	4.107E-29
Triatoma_dimidiata_6	580	0.8669	64	0.833	0.0116	0	0.01179	0.2777	0.1406	1.13E-23	0.3039	0.1406	5.87E-27
Triatoma_dimidiata_7	580	0.8654	64	0.8477	0.0181	0	0.00285	0.2687	0.1719	8.673E-21	0.3957	0.2188	1.524E-29
Triatoma_dimidiata_8	580	0.8615	64	0.8914	0.0205	0	0.003012	0.2675	0.0156	6.972E-31	0.2908	0.0312	2.835E-32
Triatoma_dimidiata_9	580	0.8633	64	0.8643	0.0178	0	0.004882	0.264	0.0938	1.011E-25	0.3237	0.0938	2.44E-33
Triatoma_gerstaeckeri_0	42	0.945	11	0.9914	0.0183	0	0.007702	0.0891	0	1.049E-07	0.4505	0	1.122E-17
Triatoma_gerstaeckeri_1	42	0.969	11	0.8813	0.0154	0.0909	0.002297	0.1835	0.2727	1.441E-07	0.267	0.2727	6.557E-09
Triatoma_gerstaeckeri_2	42	0.9531	11	0.9059	0.011	0	0.006222	0.1492	0.2727	0.00000144	0.0476	0.0909	0.0000113
Triatoma_gerstaeckeri_3	43	0.9706	10	0.8679	0.0142	0.1	0.005348	0.1365	0.3	2.034E-06	0.1033	0.2	7.703E-07

Triatoma_gerstaeckeri_4	43	0.9505	10	0.9093	0.0128	0	0.006409	0.0546	0.1	0.00001518	0.0984	0.1	4.427E-07
Triatoma_mexicana_0	300	0.95	34	0.9221	0.0089	0	6.635E-06	0.3696	0.2647	2.428E-41	0.2103	0.1176	2.943E-34
Triatoma_mexicana_1	300	0.9494	34	0.9246	0.004	0	0.0005045	0.3423	0.2059	3.039E-43	0.2543	0.1176	6.289E-39
Triatoma_mexicana_2	300	0.9488	34	0.957	0.0056	0	0.0002722	0.3407	0.1176	2.83E-51	0.4168	0.1176	0
Triatoma_mexicana_3	300	0.9476	34	0.9289	0.0042	0.0294	0.001877	0.3339	0.1471	7.666E-46	0.4755	0.1471	0
Triatoma_mexicana_4	301	0.9506	33	0.9221	0.0616	0.0303	9.807E-16	0.3557	0.1818	5.124E-46	0.141	0.0303	2.991E-30
Triatoma_mexicana_5	301	0.9511	33	0.9259	0.0091	0	0.00001064	0.3432	0.2424	4E-38	0.1756	0.1212	1.881E-29
Triatoma_mexicana_6	301	0.9482	33	0.9495	0.0061	0	0.0001875	0.3331	0.1515	1.489E-44	0.2288	0.0303	5.638E-43
Triatoma_mexicana_7	301	0.9505	33	0.9346	0.0087	0	0.00002242	0.3498	0.303	2.074E-33	0.1591	0.0303	2.361E-33
Triatoma_mexicana_8	301	0.9514	33	0.9147	0.0101	0	0.00000439	0.3434	0.3636	1.024E-27	0.1305	0.0909	1.265E-25
Triatoma_mexicana_9	301	0.95	33	0.9274	0.0067	0	0.00007387	0.3375	0.2121	4.909E-40	0.0728	0	1.281E-19
Didelphis_virginiana_0	224	0.8196	25	0.7897	0.0244	0	0.3262	0.2738	0.12	0.0000511	0.4332	0.28	1.605E-08
Didelphis_virginiana_1	224	0.8148	25	0.8284	0.0383	0	0.2023	0.2643	0.04	2.875E-06	0.4091	0.28	1.603E-07
Didelphis_virginiana_2	224	0.8161	25	0.8305	0.0381	0	0.129	0.2679	0.08	0.00002086	0.3486	0.08	2.799E-09
Didelphis_virginiana_3	224	0.8148	25	0.8447	0.0357	0	0.1981	0.2731	0.08	0.00001088	0.4621	0.24	7.15E-11
Didelphis_virginiana_4	224	0.8149	25	0.8414	0.0328	0	0.3604	0.2774	0.04	9.443E-07	0.522	0.32	5.41E-11

Didelphis_virginiana _5	224	0.8227	25	0.7636	0.0323	0	0.3115	0.2586	0.12	0.000067 64	0.294	0.16	0.000023 73
Didelphis_virginiana _6	224	0.828	25	0.7135	0.0279	0.04	0.4222	0.2601	0.28	0.01562	0.3268	0.28	0.000285 9
Didelphis_virginiana _7	224	0.8192	25	0.8006	0.0343	0	0.2257	0.2719	0.16	0.000494 7	0.4438	0.32	6.568E- 08
Didelphis_virginiana _8	224	0.816	25	0.8176	0.0352	0	0.2862	0.262	0.12	0.000114 6	0.4318	0.24	1.766E- 09
Didelphis_virginiana _9	225	0.8168	24	0.8284	0.0259	0	0.337	0.2806	0.125	0.000111 2	0.4374	0.2083	5.627E- 10

Annex 4: Binary predictions of suitability of vectors and reservoirs using their respective Minimum Training Presence value as threshold. Values of one indicate suitable habitat while zero indicates unsuitable habitat.



Annex 5: Binary predictions of suitability of vectors and reservoirs using a fixed threshold value of 0.5 Presence value as threshold. Values of one indicate suitable habitat while zero indicates unsuitable habitat.

