**NASA DEVELOP National Program**



Mobile County Health Department

*Spring 2014*

**Chagas Transmission Risk in Alabama**

*Habitat Suitability Modeling of Triatoma sanguisuga, the Expected Local Vector for Chagas Disease in the South Eastern United States*

**Technical Report**

April 4, 2014

James Pickett, University of South Alabama (Project Lead)

Brandi Stewart, University of South Alabama

Shikher Mishra, University of South Alabama

Dr. Bernard Eichold, Mobile Public Health Officer

Joe Spruce, Computer Sciences Corporation

Dr. Kenton Ross, DEVELOP National Science Advisor

# Abstract

Chagas Disease is caused by the parasitic protozoa *Trypanosoma cruzi,* and is transmitted by members of Triatominae, a hematophagous subfamily of the Reduviidae family of insects, commonly known as “kissing bugs”. Triatomines transmit *T. cruzi* to hosts when they defecate after taking a blood meal. Long term symptoms of the disease are primarily characterized by afflictions of the cardiovascular and gastrointestinal systems where the protozoa reproduce, resulting in the accumulation of internal scar tissue. Until recently, both vector and parasite were considered endemic only to Latin America. However, since the 1960s, triatomines and infected reservoirs have been present in the Southern United States. *Triatoma sanguisuga,* the most adaptable triatomine in the U.S., is becoming increasingly widespread throughout Louisiana and has been documented extensively in the southern portion of the state. Despite this, little attention is given to Chagas Disease in the United States. After a suspected case emerged in Dothan, Alabama, the Mobile County Health Department (MCHD) DEVELOP node began conducting research to determine triatomine habitat suitability in the Southeast, and particularly in Mobile, AL.

To facilitate collection and testing of *T. sanguisuga*, ecological niche modeling was conducted using data provided by NASA Earth observing sensors. For example, Shuttle Radar Topography Mission (SRTM) data were used to derive terrain elevation and slope, and Total Rainfall Measuring Mission (TRMM) data were used to calculate rainfall accumulations on a seasonal basis. These and other remotely sensed data products were inputted into an ecological niche model formulated by David Stockwell and David Peters known as the Genetic Algorithm for Rule-set Production (GARP). When the project began in the fall 2013 DEVELOP term, the objective was to process the primary layers identified in literature as needed for the model: temperature, elevation, slope and altitude, rainfall, and land cover. The goals of continuing research in the spring 2014 DEVELOP term were to implement additional satellite-based data products such as Normalized Difference Vegetation and Water Indices (NDVI, NDWI), calculated from observations made by Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA’s *Aqua* and *Terra* satellites, and to screen the model results with Wildland-Urban Interface (WUI) and Socioeconomic data.

The majority of the Southeastern United States showed high habitat suitability for *T. sanguisuga*. These results demonstrate the ability of NASA Earth Observing data to aid in ecological niche modeling for the insect vector of Chagas disease.

**Keywords**

Ecological Niche Modeling, GARP, *Triatoma sanguisuga,* Chagas Disease

# I. Introduction

**Background**

In the fall of 2013, the DEVELOP MCHD team began a project aimed at aiding public health officials and entomologists in locating the local insect vector for Chagas Disease, *Triatoma sanguisuga.* In order to do this, the team implemented the GARP ecological niche model. GARP makes correlations between known species location coordinates and various environmental parameters. This term, the team aimed to expand on the previous study by acquiring additional known species locations, scaling to a finer spatial resolution, and by incorporating additional environmental parameters to more accurately indicate species habitat constraints. Additionally, a secondary study was conducted wherein model outputs were compared with WUI and socioeconomic data in order to map areas where *T. sanguisuga* would most likely be captured.

**The Insect Vector and Chagas**

American trypanosomiasis was first documented by Carlos Chagas in 1909 as a zoonotic disease caused by the hemoflagellate parasite, Trypanosoma cruzi. T. cruzi’s lifecycle includes a pathway through invertebrates and vertebrates with multiple differentiation stages in its life cycle (Bern, et al. 2007). Its heteroxenous life cycle initiates with the reproductive period using binary fission in the intestine of the subfamily Triatominae, becoming epimastigotes (Ley, V, et al. 1988). The epimastigotes migrate to the hindgut and develop into metacyclic trypomastigotes (Bern, et al. 2007). The metacyclic trypomastigotes (infective period) mature in the rectal cell walls awaiting transmission through vector defecation (Ley, V, et al. 1988). Initial stages of the disease present symptoms similar to the flu, while a lack of treatment leads to a chronic phase in which the protozoa multiplies in soft tissues, causing irreparable damage to the heart and intestines. Vector-borne transmission can occur in two ways, infected fecal contamination of open wounds such as the bite site or fecal contamination of mucous membranes (Bern, et al. 2007). Other transmission routes include, congenital, blood-borne, and organ derived (CDC 2012). However, vector-borne transmission through triatomines is the most common form of transmission.

Triatomines seek blood meals, often from humans, and transmit the disease through defecation, soon after feeding. Being attracted to carbon dioxide, the insect displays a preference for the face, which offers ample infection routes through the eyes or mouth. Until recently, both vector and parasite associated with Chagas Disease were considered endemic only to Latin America. Cases seen in the United States were believed to be a result of emigration from these endemic regions (Zeledón et al 2012). Through the work of Rodrigo Zeledón, et al. (2012) however, scientists now agree Chagas Disease has long been a disease endemic to all of the Americas. Although it is hypothesized that better housing within the U.S. prevents rampant transmission, increases in the wild urban interface and climate change may open up new habitats for various species of Triatominae and increase the transmission potential for Chagas, serving to further justify concern. As suggested by Zeledón, et al., two primary reasons for the continued belief that Chagas Disease does not pose a risk in the United States is because of the lack of domestic dwelling places suitable for habitation by vectors and the delayed defecation of the triatomid species dominant in this region. It is important to note here that the effectiveness of a given species as a vector for Chagas is attributed to how soon after feeding the insect defecates, as the average time interval between feeding and defecation varies between the many triatomine species and ranges anywhere from thirty seconds to over an hour. Currently Chagas is not a reportable disease in the state of Alabama, thus the goal of the project was to aid Dr. Mujica and Dr. McCreadie in their efforts to collect and test a local *T. sanguisuga* specimen.

**The GARP Model**

In the previous term, the Desktop GARP interface was used, however with remote access to the Alabama Supercomputer, the C language based Command Line S/W version of GARP was attempted this term. GARP analyzes a given area’s extent, and based on known species occurrence locations, climatic, and environmental inputs, finds correlations which are expressed in the form of a series of decision rules used to predict suitable habitats for the species to maintain a population. Using GARP, an individual can both derive an expected native distribution range for a species, and also use GARP to rank environmental and climatic parameters by GARP’s interpretation of their significance for the species’ survival. This is particularly useful for species collections and for performing future studies with ecological niche modeling of the given species. Ranked parameters can indicate if a particular type of data appeared to be more influential, such as temperature or rainfall. This provides insight for future parameter choices. Because it was necessary to use a larger area as input for GARP, and because known species location coordinates are scarcely available locally (within Alabama), the entire Southeastern United States was modeled, but only the results found in Alabama were assessed. All parameters were taken as recently as possible.

In the previous DEVELOP term; the parameters used to run the model included the nineteen bioclimatic data layer inputs found in literature. These were expanded upon and more derived indices such as land cover, altitude, slope and NDVI were used this term. Prior ecological niche modeling studies have indicated that for suitability modeling done with plants, soil moisture yielded significantly more accurate results than bioclimatic parameters relating to precipitation. This is because the amount of moisture actually stored in the ground is a more telling factor as to the limitations of various plants. Additionally, soil temperature and moisture are known to be important limiting factors for insects that live in or upon the ground. Because *T. sanguisuga* is most commonly found sharing the burrows of animals such as armadillos or possums, it was decided to include both of these as parameters in this study.

It was intended that the model would be run with the inputs having a finer resolution of 100 by 100 meters. Even with access to the Alabama Supercomputer, the allotted diskspace quota of 40 gigabytes was not large enough to contain all of our parameters and run GARP, as at least 80 gigabytes would be required. To overcome this, it was decided that a principal component analysis (PCA) would be conducted on the 28 parameters to create a subset of parameters to run the model with.

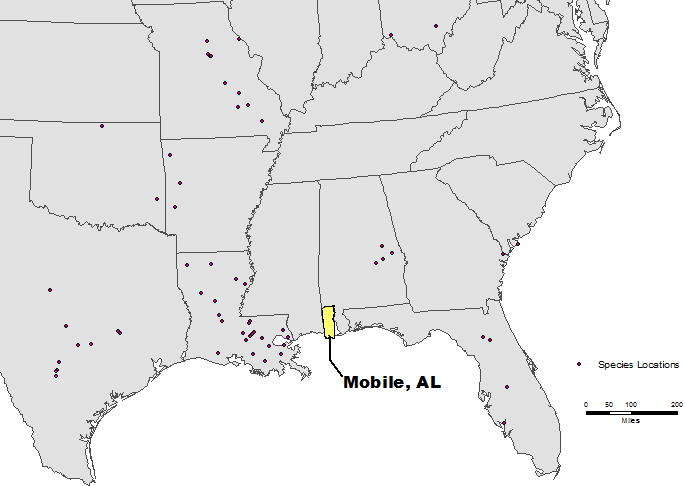
# II. Methodology

**Data Acquisition and Processing**

For this implementation of GARP, nineteen bioclimatic variables along with Land Cover, NDVI, NDWI, Slope, Altitude, Soil Moisture (for winter and summer) and Land Surface Temperature (for winter and summer) constituted the total 28 environmental parameters that were used. Details about the environmental parameters are given in **Appendix 1** as are the data sources. Each environmental layer was clipped according to the area of interest, which ranged from -116.003 to -72.995 longitude, and 44.003 to 23.996 latitude. These were then resampled to approximately ninety meters and 500 meters resolution using ERDAS to create two sets of environmental parameters. After this these layers, which were used as an input to GARP, were converted into ASCII format using ArcGIS.

As mentioned, because of computational diskspace limitations a PCA was performed on both sets of environmental parameters, which is further detailed later in this paper (**Figure 2**). In order to carry out this procedure, the datasets were divided into six categories, each one containing related parameters, and performing individual PCAs on each category. This process rendered five data layers that included the principal component for eleven temperature variables, eight precipitation variables, NDVI & NDWI, LST & SM, Slope & Elevation. A PCA was not performed on Land Cover. Finally, four sets of environmental layers were obtained. These included the 28 parameters with 90-m resolution, the 28 parameters with 500-m resolution, the six parameters derived using PCA at 90-m resolution, and the six parameters derived using PCA at 500-m resolution.

In all, 67 distinct specimen locations were acquired. Details about specimen location source are given in **Appendix1**. The collected sites were used to create an excel file with latitude and longitude values which was used as an input to GARP.



**Figure 1** The 67 distinct specimen locations

Socioeconomic data and WUI/intermix data were used to derive poorer/rural areas (refer to **Appendix 2** for sources of WUI and socioeconomic data). Socioeconomic data were used to distinguish poorer/rural areas where substandard housing could be found. The indicators used were median household income, households below poverty line, and housing values below $39, 999 USD. The socioeconomic data were further enhanced with WUI/intermix data. Wildland Urban Interface (WUI) refers to the zone of transition between unoccupied land and human development. This was used as an additional indicator to find the peridomestic habitat. Socioeconomic data and WUI were combined with the 100 percentile areas in the habitat suitability output given by GARP. This was done to return a more precise area for possible collection sites.

**GARP Implementation and Output Results**

Processing the data was completed in two phases. The initial phase was the implementation of GARP. The secondary phase was to mask the GARP habitat suitability using the socioeconomic and WUI data to screen the model outputs. Bioclimatic and elevation data were downloaded as four TIFF files that covered the United States. These were merged into one layer using ERDAS Image. A raster layer for slope was created with the help of elevation data using ArcGIS. Land Cover and NDVI data were downloaded as TIFF files for the entire USA. Data were clipped to cover the area of interest (i.e., Texas, Alabama, Louisiana, Mississippi, Florida and Georgia) using ArcGIS. Data were then resampled to an approximate spatial resolution of 100-m and finally converted into ASCII files using ArcGIS. When implementing GARP, the four different environmental layer combinations previously described were tried as input parameters. In the end, the combination that was used was: all bioclimatic variables, land cover, elevation, slope, NDVI, and NDWI at 500-m resolution. These combinations of inputs, along with the specimen locations, were the inputs used when running GARP. The model was set to run fifty times for each dataset with convergence limit of 0.01 and maximum iterations of 1000. Fifty risk maps were obtained from each of the datasets with niche prediction probability of either one or zero, where one represented presence of species and zero represented absence of species.

The plan was to run GARP with 28, 90-m resolution layers which resulted in a huge dataset that was impossible to simulate using Desktop GARP. A possible solution was to use CLI version of GARP on the supercomputer but diskspace limitation would not allow for all 28 parameters to be used. To limit diskspace requirements and run GARP at a finer spatial resolution, it was decided that a principal component analysis would be conducted for the 28 layers and the top components would be used in place of the initially intended parameters.

To justify the use of the top PCA components as surrogate layers for original parameters, a PCA at 500-m resolution was performed and simulated using Desktop GARP with and without PCA derived layers. Outputs from both simulations were filtered with a threshold test probability of greater than 0.5 and were added to produce probability maps separately. These probability maps were compared by producing their difference map. The analysis of their difference map suggested that the PCA derived layers should not be used for the simulation. It was also observed that the very coarse source size of the soil moisture (SM) and land surface temperature (LST) layers were strongly impacting the model outputs for the run, and that using these with the original parameters could be the cause of such a drastic difference. This was noticed due to the large square shaped patterns running throughout the stacked model outputs of this implementation. Because of this the exact same methodology was enacted, except without the SM and LST datasets. The two resulting maps, one of GARP ran with the categorical PCA excluding LST and SM and one of GARP ran with all original parameters except SM and LST still proved to be significantly different. Furthermore the outputs corresponding to the PCA technique and with the original parameters, had patterns strongly resembling the previous two model output stacked maps, respectively. It could not be justified to run GARP on the supercomputer with the PCA approach, as these observations indicate that using alternate parameters derived from the top components of a PCA performed on the intended parameters, is not suitable as a replacement for the originally intended parameters. Furthermore, the prevalence of the huge square tiling found throughout the model outputs from the implementation that included soil moisture indicates that source size is extremely relevant. It appears that data source size plays a very large role in the model’s rule making decisions and that layers with drastically different original resolutions should not be used together. Because of this it is suspected that a finer resolution run of GARP- with these datasets- on the Alabama Supercomputer would not have yielded better results.

O/Ps with test probability > 0.5 filtered.

(Simulation 1:- 86 O/Ps were filtered.

Simulation 2:- 57 O/Ps were filtered)

Clipped & Resampled

GARP (Desktop Version)

Environmental Parameters

(Simulation 1:- 28 Parameters

Simulation 2:- 24 Parameters)

28 Parameters

28 Environmental Layers

500 m

PCA (Categorical) derived Parameters

(Simulation 1:- 6 Parameters

Simulation 2:- 5 Parameters)

GARP (Desktop Version)

PCA (Categorical) derived Parameters

(Simulation 1:- 6 Parameters

Simulation 2:- 5 Parameters)

Clipped & Resampled

90 m

90 m

O/Ps with test probability > 0.5 filtered.

(Simulation 1:- 200 O/Ps were filtered.

Simulation 2:- 100 O/Ps were filtered)

All O/Ps were combined to produce composite probability map.

All Parameters

All O/Ps were combined to produce composite probability map.

GARP (CLI Version)

GARP (CLI Version)

Simulation w/ PCA derived layers was not performed.

Image

Differencing

Simulation w/o PCA derived layers couldn’t be acquired due to disk space limitation of supercomputer and large data size.

Justification?

Masking with Wild-Urban Interface & Socio-Economic Data.

Simulation 2 results were chosen as final after analysis.

Analysis showed that both O/Ps were significantly different.

**Final Probability Map**

**Figure 2** Methodology of GARP implementation using resampled datasets and principle component analysis

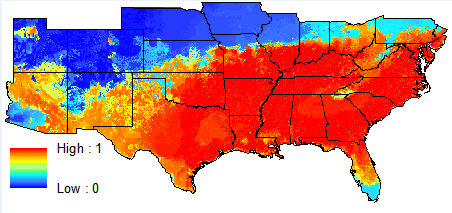
**Post GARP processing details**

The secondary phase of masking the GARP habitat suitability with the WUI & Socioeconomic data was done in ArcMap 10. The model outputs were clipped using a Mobile County, AL shapefile and masked with the socioeconomic data and WUI. Socioeconomic indicators (median household income, housing values and households below the poverty line) and WUI were represented as layers in ArcMap10. The median house income and housing values were cross-referenced with the WUI to create a layer representing the intersection of these values. A conversion of the model outputs from a TIFF format to a shapefile allowed the optimal habitat suitability to be represented by the percentage of households below the poverty line. This was done by using the clipping tool in ArcMap 10. In addition, a shapefile representing Mobile County’s major road network was added for perspective and location.

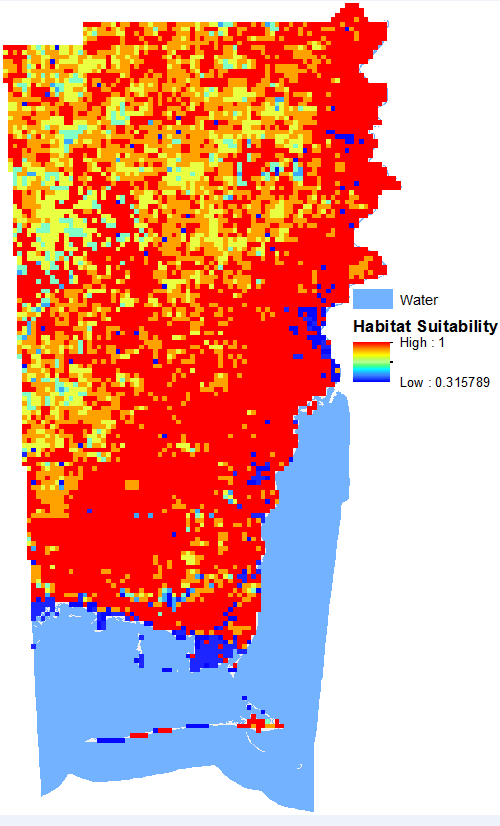
# III. Results & Discussion

The model outputs showed that most of the southeast U.S. is at risk; however, the project’s aim was to identify specific areas with comparatively higher risk so that partners at the University of South Alabama can place traps to collect the insects for testing. Many interesting observations were deduced from the final probability maps. Large square patterns appear throughout the final risk map of the run performed with SM and LST. These images can be seen in **Appendix 3**. This was caused by the large cell sizes of the soil moisture layer. Because of how GARP creates rules, using a layer with such a noticeable difference in source resolutions distorted the results. It can be concluded that resampling data to a drastically finer resolution is not a good approach for GARP modeling. This also suggested that going to finer resolution from relatively coarser source data (1000m -> 90m or 25km -> 500m) may have produced distorted results. Upcoming NASA Satellite Soil Moisture Active Passive (SMAP) would have allowed the model to make better correlations with the vector’s preferred habitat.

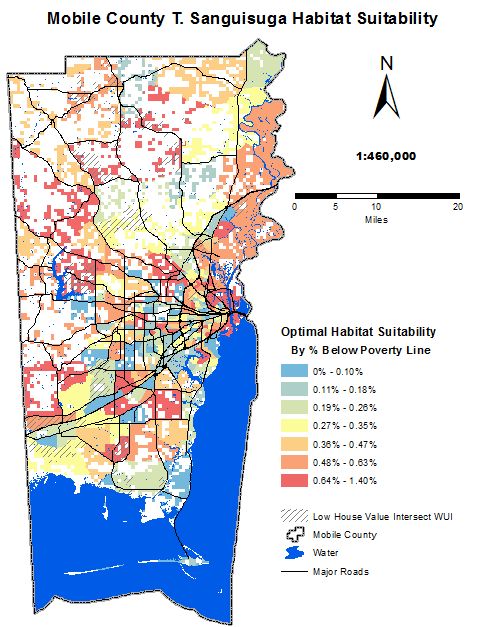
In contrast, for the local level for Mobile County, the patterns between the stacked model outputs of the run with and without SM and LST were actually quite similar. Both do conform to expectations, considering the known species location coordinates that were used. From the known species locations input, there was a high density of points in southern Louisiana where swampy areas are commonly found. In both GARP outputs a pattern can be observed showing high suitability in the southern portions of Mobile county where wetlands are prevalent. The northern part of the county, which is dominated by forested areas, also contained regions of comparatively higher suitability.

These higher suitability ranges were masked with the socioeconomic and WUI data to represent the highest probability of risk within Mobile County. A map was created with all elements (GARP outputs, socioeconomic data, and WUI) including a major road network. The final map shows the optimal sites for collection within the county limits of Mobile, AL (**Figure 5**).

**Figure 3** Final GARP output map created by stacking the results of all fifty runs modeled with 24 distinct, satellite derived environmental parameters, (excluding LST and SM)

****

**Figure 4** Same as feature 3 except clipped to Mobile County



**Figure 5** Final, comprehensive habitat suitability map created by combining final GARP outputs with WUI and socioeconomic maps found in **Appendix 4**

# IV. Conclusions

# The majority of the Southeastern United States shows high habitat suitability for *T. sanguisuga.* These results demonstrate the ability of NASA Earth Observing data to aid in ecological niche modeling for the insect vector of Chagas Disease. The goal was to compile a habitat suitability map for the Chagas vector using the GARP model. Stacking all outputs reveals areas of Mobile County with higher probabilities of encountering *T. sanguisuga*, specifically the southern portion of Mobile County, which contains large areas of wetlands and are dominated by coastline. Maskings performed with WUI/intermix and socio-economic data further refined target trap setting locations. This helped to determine optimal collection sites for insect species using post-analysis data. An improvement on this study could be implemented by collecting additional coordinates for *T. sanguisuga*. Specifically acquiring *T. sanguisuga* location coordinates from areas with greater environmental variation could improve the results. This is mentioned because, due to the accumulation of points in southern Louisiana, it appears as if the model outputs could be biased towards wetlands. Although it is not necessarily the case, acquiring a greater amount of points from environmentally diverse locations would make this distinction clear.

# In addition to utilizing the SMAP data, VIIRS nightlights could also be included as an emissive light source for post model processing as another way to aid in locating specimens, as triatomines are attracted to lights and many that have been collected are documented to have been found near light sources. Running the model at a finer resolution would provide finer resolution at the local level, but only if the source data is originally finer. The project could be reformatted and adapted to adequately meet the prerequisites of the NASA DEVELOP ecological forecasting application areas, where GARP outputs would be used to forecast the disease distribution of a possible outbreak.

# V. Acknowledgments

We would like to extend our thanks and gratitude to everyone who contributed in seeing this project through, including those that enable a venue for such projects to be conducted, and to those that ensured it see the light of day. Special thanks to the NASA mission scientists and principal investigators who provided data for this research; Stockwell and Peters, Genetic Algorithm for Rule set Production; Dr. Bernard Eichold with the Mobile County Health Department for his excellent mentorship and providing a place for us to conduct our research; Mr. Joe Spruce of Computer Sciences Corporation for his patience, diligence, understanding, insight and most of all scientific integrity; Dr. Kenton Ross, NASA Langley Research Center; Mike Ruiz our DEVELOP Program Manager, without whom none of this would be possible; Lauren Childs, DEVELOP National Science Lead; and Jamie Favors, DEVELOP Deputy National Science Lead for delegating time and resources to visit our location and advise us at a critical point in the term. We would also like to express our appreciation of our project partners, Drs. Mujica and McCreadie for their contributions to the project and their continuation in pursuing research that they believe will benefit the public health of our county and state.

# VI. References

Batista, T.A., and R. Gurgel-Goncalves. "Ecological niche modelling and differentiation between Rhodnius neglectus Lent, 1954 and Rhodnius nasutus Stål, 1859 (Hemiptera: Reduviidae: Triatominae) in Brazil." *Mem Inst Oswaldo Cruz, Rio de Janeiro*. 104.8 (2009): 1165-1170.

Bern, C., S. Kjos, M.J. Yabsley, and S.P. Montgomery. "Trypanosoma cruzi and Chagas Disease in the United States." *Clinical Microbiology Reviews*. (2011): 655-681. Doi:10.1128/CMR.00005-11

Bodbyl-Roels, Sarah, Townsend A. Peterson and Xiangming Xiao. "Comparative analysis of remotely-sensed data products via ecological niche modeling of avian influenza case occurences in Middle Eastern poultry." *International Journal of Health Geographics* (2011): 10-21.

Hernandez, J., I. Nunez, A. Bacigulapo, and P.E. Cattan. "Modeling the spatial distribution of Chagas disease vectors using environmental variables and peoples' knowledge." *International Journal of Health Geographics*. 12.29 (2013).

Huete, A.R., et al. "A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS." *Remote Sensing of Environment* (1997): 440-451.

Jurberg, J., and C. Galvao. "Biology, ecology, and systematics of Triatominae (Heteroptera, Reduviidae), vectors of Chagas disease, and implications for human health." *Denisia*. (2006): 1096-1116.

Klotz, J. H., et al. 2010. “Kissing bugs”: potential disease vectors and cause of anaphylaxis. Clin. Infect. Dis. 50:1629–1634.

Ley, V, et al. "Amastigotes of Trypansoma Cruzi sustain an Infective Cycle in Mammalian Cells." *The Journal of Experimental Medicine* (1988): 649-659.

Maguire, J. H. 2004. Trypanosoma, p. 2327–2334. *In* S. Gorbach, J. Bartlett,

and N. Blacklow (ed.), Infectious diseases, 2nd ed. Lippincott, Williams &

Wilkins, Philadelphia, PA.

Peterson, A.T., V. Sanchez-Cordero, C.B. Beard, and J.M. Ramsey. "Ecological Niche Modeling and Potential Reservoirs for Chagas Disease, Mexico."*Emerging Infectious Disease Journal - CDC*. 8.7 (2002).

Piedallu, Christian, et al. "Soil water balence performs better than climatic variables in tree species distribution modeling." *Global Ecology and Biogeography* (2013): 470-482.

Radeloff, V.C., et al. "The Wildland-Urban Interface in the United States." *Ecological Applications* (2005): 799-805.

Rassi, A., Jr., A. Rassi, and J. A. Marin-Neto. 2010. Chagas disease. Lancet 375:1388–1402.

Sarkar S., Strutz S.E., Frank D.M., Rivaldi C., Sissel B., et al. (2010) “Chagas Disease Risk in Texas.” PLOS Neglected Tropical Diseases 4.10 (2010): e386. doi: 10.1371/journal.ptnd.0000836

Shapiro, T. A and P.T. Englund. "The Structure and Prelication of Kinetoplastid DNA." *Annual Review of Microbiology* (1995): 117-143.

Tyler, K.M., and D.M. Engman. "The life cycle of Trypanosoma cruzi revisited." *International Journal for Parasitology*. 31. (2001): 472-481.

WHO Expert Committee. 2002. Control of Chagas disease. WHO technical report series number 905. World Health Organization, Geneva, Switzerland

Zeledon, Rodrigo, et al. *An Apprasial of the Status of Chagas Disease in the United States*. Waltham: Elsevier, 2012.

# 

# VII. Appendices

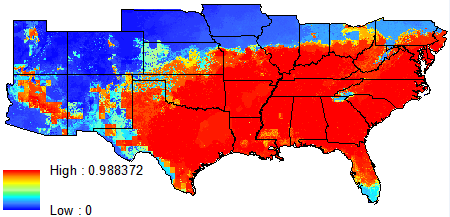
**Appendix 1** Satellite derived environmental parameters inputted into GARP

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Sensor/Source | Spatial Resolution | Temporal Coverage | EOS? |
| Annual Mean Temperature | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Mean Diurnal Range | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Isothermality | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Temperature Seasonality | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Temperature Seasonality | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Maximum Temperature of Warmest Month | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Minimum Temperature of Coldest Month | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Temperature Annual Range | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Mean Temperature of Wettest Quarter | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Mean Temperature of Driest Quarter | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Mean Temperature of Warmest Quarter | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Mean Temperature of Coldest Quarter | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Annual Precipitation | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Precipitation of Wettest Month | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Precipitation of Driest Month | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Precipitation Seasonality (Coefficient of Variation) | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Precipitation of Wettest Quarter | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Precipitation of Driest Quarter | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Precipitation of Warmest Quarter | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Precipitation of Coldest Quarter | www.worldclim.org1 | ~(1 KM x 1 KM) | 1950 – 2000 (Average) | May Be2 |
| Land cover | National Land Cover Database (NLCD). Based primarily on the unsupervised classification of [Landsat Enhanced Thematic Mapper+](http://landsat.usgs.gov/) (ETM+) circa data | ~ (30m x 30m) | 2006 | Yes |
| NDVI | Global Land Cover Facility (GLCF). Based on MODIS data. | ~(250m x 250m) | 2006 | Yes |
| NDWI | www.geoland2.eu. Based on VGT data. | ~(1 KM x 1 KM) | 2009 - Present | No |
| Slope | www.worldclim.org1. Based onShuttle Radar Topography Mission (SRTM) data | ~(1 KM x 1 KM) | Unknown | Yes |
| Altitude | www.worldclim.org1. Based onShuttle Radar Topography Mission (SRTM) data | ~(1 KM x 1 KM) | Unknown | Yes |
| Land Surface Temperature | http://neo.sci.gsfc.nasa.gov. Based on MODIS data | 0.1 degreex 0.1 degree | 2007 – 2013 (Average for alternate years) | Yes |
| Soil Moisture | http://www.esrl.noaa.gov. Modal calculated | 0.5 degreex 0.5 degree | 2005 – 2013 (Average for alternate years) | No |
| Species Location | Global Biodiversity Information Facility, Louisiana State Arthropod Museum, and the Illinois Natural History Museum databases | N/A | N/A | No |
| 1Worldclim.org created interpolated bioclimatic layers using climate databases by the Global Historical Climatology Network ([GHCN](http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=ghcn.html)), [Food](http://www.fao.org/) and Agriculture Organization (FAO), [World](http://www.wmo.ch/) Meteorological Organization (WMO), International Centre for Tropical Agriculture ([CIAT](http://www.ciat.cgiar.org/)), Regional, Electronic Hydrometeorological Data Network (R-HydroNET), and a number of additional minor databases.  2 Because of the diverse collection of sources in preparation of final data, it can’t be said that NASA EOS data were used for sure. | | | | |

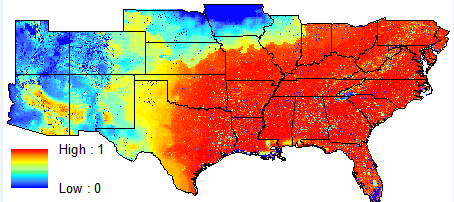
**Appendix 2** Ancillary data sources

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Socioeconomic | http://sedac.ciesin.columbia.edu/, https://www.census.gov/ | N/A | 2010 | Yes |
| Wild-Urban Interface | Census, National Land Cover Database (NLCD) and the protected Areas Database version 1.1 | N/A | 2010 (Census), 2006 (NLCD) | Yes |
| *Triatoma sanguisuga* specimen coordinates | Louisiana State Arthropod Museum and the Global Biodiversity Information Facility | X, Y data | 1960 - Present | N/A |

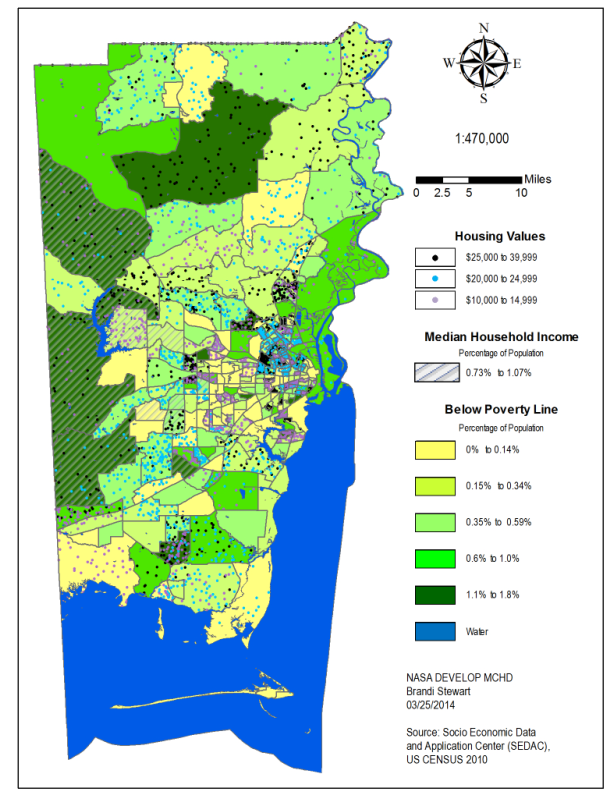
**Appendix 3** Alternative model outputs resulting from experimental runs at resampled resolution and with or without PCA analysis

****

**Figure A** All 28 parameters ran on Desktop GARPwithout resampling or PCA methods applied

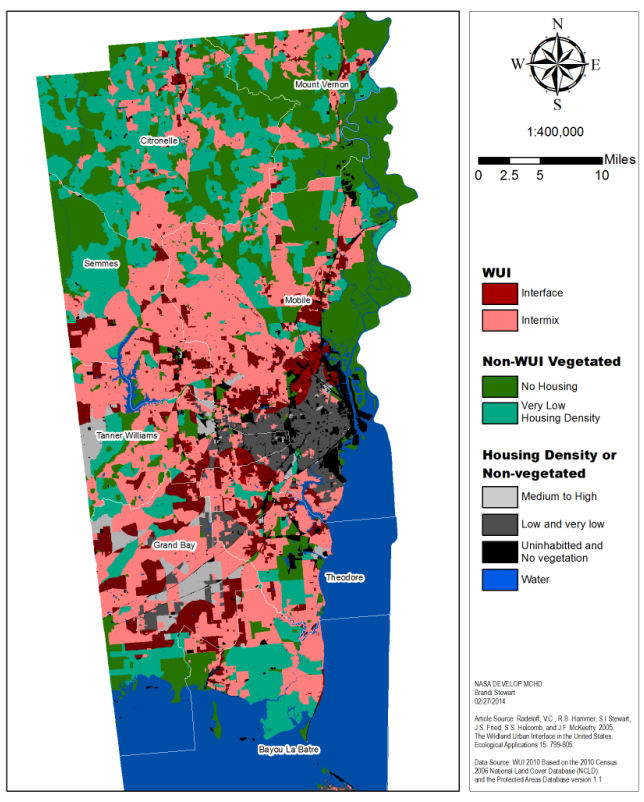
****

**Figure B** Resampled input parameters derived with PCA, from original parameters excluding Land Surface Temperature and Soil Moisture

**Appendix 4** WUI/intermix and Socioeconomic maps used in secondary studies

**-urban Interface Map**

**Figure C** Socioeconomic map

**-urban Interface Map **

**Figure D**

Wildland-Urban Interface and Intermix map