Mapping and Modeling Neglected Tropical Diseases and Poverty in Brazil, Bolivia and Colombia

JB Malone, P Nieto, P Mischler, M Martins, JC McCarroll
Louisiana State University, USA

Penelope Vounatsou, Ronaldo Scholte
Swiss TPH, Switzerland

ME Bavia
Universidade Federal da Bahia, Brazil

International Society for Photogrammetry and Remote Sensing
2nd Symposium on Advances in Geospatial Technologies for Health
Objectives

• Data Portal – A resource data base accessible by FTP was developed for 6 NTD in Brazil, Bolivia and Colombia (Chagas disease, Leishmaniasis, Schistosomiasis, Leprosy, Lymphatic Filariasis and Soil-Transmitted Helminths), with relevant climatic, environmental, population and poverty data

• Risk Modeling – Maximum Entropy, Bayesian and GIS methodologies were used to map and model environmental and socioeconomic risk of 6 NTD

• Course Development – A 4-day short course was developed for training use by PAHO on data portal access and geospatial analysis using ArcGIS 9.3.1, Maximum Entropy (Maxent) and Bayesian modeling
Data Portal

All data clipped to the country boundary; WGS84 projection, 1 km spatial resolution; in ASCII format for Maxent or Bayesian modeling

This example shows the data available for Colombia

Worldclim (global coverage, 1km resolution) used for ecological Niche modeling and by the climate change community

MODIS EVI, LST annual composites for 2005-2009

Socioeconomic Data at the Municipality level
Worldclim Global Climate Data

Tmin, Tmax, Precip, SRTM, Bioclim – 1 km resolution

Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These are often used in ecological niche modeling (e.g., BIOCLIM, GARP).

BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (P2/P7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
BIO7 = Temperature Annual Range (P5-P6)
BIO8 = Mean Temperature of Wettest Quarter
BIO9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter
Contents of Data Portal/FTP Site

MODIS
Mean annual composites for 2005-2009:
Enhanced Vegetation index (EVI), Normalized difference Vegetation Index (NDVI)
Land surface temperature (LST) day and night and dT

Climate GRID
Long term normal (LTN) climate grid (18x18 km cell size) – Precip, Tmax, Tmin, PET, PPE

Environmental
World Wildlife Fund Ecoregions
Locations of springs, dams, rivers, small streams

Health Data
*Bolivia*: Ministerio de Salud y Deportes/ Sistema Nacional de información en Salud
*Brazil*: Ministerio da Saude, SINAN
*Colombia*: Instituto Nacional de salud/Estadísticas de la Vigilancia en Salud Pública
Ministerios de la protección Social (SIVIGILA), literature reports.

Infrastructure
Roads, airfields/airports, rail road lines layer, utility lines

Political Boundaries
Counties, major cities, States/Departments, Municipalities
## Socioeconomical Variables at Municipality Level Used for Risk Analysis of NTDs in Colombia

<table>
<thead>
<tr>
<th>Area of municipality</th>
<th>Floors: carpet, marmol, hardwood, tablet</th>
<th>Garbage: in the river, stream, lake, lagoon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Displacement (just COL)</td>
<td>Floors: carpet, brick, vinyl,</td>
<td>Garbage: in another way</td>
</tr>
<tr>
<td>Population</td>
<td>Floors: cement</td>
<td>Drinking water from: running water service</td>
</tr>
<tr>
<td>Extension Km2</td>
<td>Floors: tough wood, other vegetal material</td>
<td>Drinking water from: well, pump</td>
</tr>
<tr>
<td>Human development index</td>
<td>Floors: soil, sand</td>
<td>Drinking water: rain fall</td>
</tr>
<tr>
<td>Unsatisfied basic needs * UBN</td>
<td>Walls: block, brick, stones, hardwood</td>
<td>Drinking water: public tank</td>
</tr>
<tr>
<td>Miseria (2 or more *UBN)</td>
<td>Walls: adobe, bahareque</td>
<td>Drinking water: car-tank</td>
</tr>
<tr>
<td>Un adequate housing * UBN</td>
<td>Walls: rough wood</td>
<td>Drinking water from: river, stream, lake, lagoon</td>
</tr>
<tr>
<td>Unsatisfied services* UBN</td>
<td>Wall: pre fabricated walls</td>
<td>Drinking water from: bottles, bag</td>
</tr>
<tr>
<td>Overcrowding * UBN</td>
<td>Walls: cane, bamboo, vegetal material</td>
<td>Infant mortality</td>
</tr>
<tr>
<td>Educational needs* UBN</td>
<td>Walls: zinc, fabric, cardboard, plastic</td>
<td>Life expectancy</td>
</tr>
<tr>
<td>Economical dependency*UBN</td>
<td>No walls</td>
<td>Attendance educational institution YES</td>
</tr>
<tr>
<td>Sewage</td>
<td>Electricity: yes</td>
<td>Attendance /educational institution NO</td>
</tr>
<tr>
<td>Running water</td>
<td>Electricity: no</td>
<td></td>
</tr>
<tr>
<td>Toilet connected to sewage</td>
<td>Garbage collection services</td>
<td></td>
</tr>
<tr>
<td>Toilet connected to septic tank</td>
<td>Burrow the garbage</td>
<td></td>
</tr>
<tr>
<td>Latrine</td>
<td>Burn the garbage</td>
<td></td>
</tr>
<tr>
<td>No sanitary service</td>
<td>Garbage: patio, back yard, ditch</td>
<td></td>
</tr>
</tbody>
</table>

### Table 1. Socioeconomical variables (47) selected for risk analysis of NTDs in Colombia

OpenNLP.maxent package is a mature Java package for training and using maximum entropy models.

Check out the Sourceforge page for Maxent for the latest news. You can also ask questions and join in discussions on the forums. Download the latest version of Maxent.

Environmental Models

Sivigila (disease reports)
29 environmental variables

Multiple regression

Significant variables

Logistic regression

Variance Inflation factor

Variables VIF<10

Maxent

Re run Maxent

Variable selection Pearson’s

Final Model
Chagas Disease

*Trypanosoma cruzi* - 20 million infected in the Americas - Chronic Cardiomyopathy

Circulating Trypomastigote and Tissue Amastigote forms in mammals

Triatomid ‘kissing’ bug vectors

Romana’s Sign

Tissue amastigote form
Chagas Vector Distribution

Rhodnius prolixus Environmental Model

Triatoma dimidiata Environmental Model
Chagas vectors - Environmental Niche model

Jackknife of AUC for R._prolixus

Jackknife of AUC for T._dimidiata
Chagas Environmental Niche Model
Socio-Economical Model

41 socio economical variables divided in 8 groups

Multiple Regression and VIF

Choose variables for weighted models

Weighted model:

- SocioEc 1
  - Re-classify
  - SocioEc Final model
  - Reclassify
  - weighted

- SocioEc 2
  - Re-classify
  - weighted

- SocioEc 3
  - Re-classify
  - weighted

Maxent Environ Model

Combined (Socio economical – environmental) final model
Socioeconomic Factors – Municipality level

Chagas Disease

Combined Model

9.6
.1

Chagas Disease
SocioEconomical Model

9.02
1

Chagas Disease
Combined Model

9.6
.1
### Variable Contribution

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>prec02_brazil</td>
<td>75.3</td>
</tr>
<tr>
<td>bio14_brazil</td>
<td>13.1</td>
</tr>
<tr>
<td>alt01_brazil</td>
<td>5.4</td>
</tr>
<tr>
<td>lstnight_2008_brazil</td>
<td>4.5</td>
</tr>
<tr>
<td>brazil_ubn24</td>
<td>1.1</td>
</tr>
<tr>
<td>brazil_gdp1</td>
<td>0.7</td>
</tr>
</tbody>
</table>
Visceral Leishmaniasis

Caused by protozoans of the genus *Leishmania*

- Amastigote form – mammals
- Promastigote form – arthropod vector

Sandfly vector (*Lutzomyia*)
Leishmaniasis – Visceral and Cutaneous

Maxent Environmental Model using Worldclim data
Cutaneous Leishmaniasis

Visceral Leishmaniasis

- VL: precipitation of October (11.6%); mean temperature of warmest quarter (14.5%) (AUC 0.948)

- CL: precipitation of September (26.2%); annual precipitation (17.3%) (AUC 0.80)
Leprosy in Brazil

Maxent predictive model showing the distribution probability of leprosy occurrence. Red indicates a higher probability of occurrence, while blue indicates a low probability of occurrence.

The predicted risk map of leprosy overlaid with 2010 leprosy occurrence data.
Schistosomiasis
Hookworm in Bolivia

Worldclim Model

AUC = 0.924

Variable
   | % Contribution
----|-----------------|
tmax06 | 28.2
prec01 | 18.4
prec08 | 13.3
tmax04 | 9.5
xmin05 | 8.9
tmin04 | 7.1
tmin03 | 3.9
prec03 | 3.0
prec04 | 2.0
prec05 | 1.8
prec09 | 1.7

Bioclim Model

AUC = 0.815

Variable
   | % Contribution
----|-----------------|
Bio13 - Precipitation wettest month | 37.2
Bio15 - Precipitation Seasonality | 25.4
Bio03 - Isothermality | 13.0
Bio09 - Mean temperature driest quarter | 11.5
Bio01 - Annual mean temperature | 6.3
Bio02 - Mean diurnal range | 2.1
Bio07 - Temperature annual range | 0.3
Bio12 - Annual precipitation | 0.2

MODIS Model

AUC = 0.7

Variable
   | % Contribution
----|-----------------|
GDD/WB Gradient | 44.3
LST day | 21.8
LST night | 20.3
Temperature Difference (dT) | 12.8
Enhanced Vegetation Index | 0.7
Conclusions and Recommendations

1. Maxent Ecological Niche Modeling is a useful tool to guide surveillance and control programs for NTD, particularly where health surveillance data are scarce.

2. Extrapolation of risk surfaces is of limited validity where representative survey data are absent in a given ecosystem.

3. Socioeconomic data or poverty indicators should be at the census tract level; Municipality level data is typically too heterogeneous.

4. Results of Maxent ecologic niche mapping and modeling should be validated by alternative methods eg. biology based GDDxWB climate models.
Maxent generated risk surfaces extracted for Bahia from national scale maps on visceral leishmaniasis (a) and cutaneous leishmaniasis (b) using MODIS environmental satellite annual composite data on vegetation index (EVI) and land surface temperature (LST).
Local Intervention Scenarios

Environmental (15-30 m²)
- Climate
- Hydrology

Vulnerability (census block)
- Landuse
- Poverty
- Population #/Density/migration
- Exposure/occupation
- Reservoir Hosts
- Vectors

Community Profile modeling System

Select High, Medium, Low Risk municipalities (5 each) using SINAN case reports, vector records

Local Intervention Scenarios
- Vector Control
- Reservoir control
- Surveillance Planning