A geospatial risk assessment model for leprosy in Ethiopia based on environmental thermal-hydrological regime analysis

Azeb Tadesse Argaw[1,2,3], E.J. Shannon[1,3], Abraham Assefa[2], Fekade Silassie Mikru[4], Berhane Kidane Mariam[4], John B. Malone[1]

[1]Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana States University, Baton Rouge, Louisiana, USA; [2]Armauer Hansen Research Institute, Addis Ababa, Ethiopia; [3]Gillis W. Long Hansen's Disease Center, Lab Research Branch, Louisiana State University, Baton Rouge, Louisiana, USA; [4]Ethiopian Ministry of Health, Addis Ababa, Ethiopia

Abstract. Geospatial methods were used to study the associations of the environmental thermal-hydrological regime with leprosy prevalence in the Oromia and Amhara regions of Ethiopia. Prediction models were developed that indicated leprosy prevalence was related to: (i) long-term normal climate grid data on temperature and moisture balance (rain/potential evapo-transpiration); (ii) satellite surveillance data on the Normalized Difference Vegetation Index (NDVI) and daytime earth surface temperature (Tmax) from the Advanced Very High Resolution Radiometer (AVHRR); and (iii) a Genetic Algorithm Rule-Set Prediction (GARP) model based on NDVI and Tmax data in relation to leprosy prevalence data. Our results suggest that vertical transmission is not the only means of acquiring leprosy and support earlier reports that a major factor that governs transmission of leprosy is the viability of *Mycobacterium leprae* outside the human body which is related to the thermal-hydrologic regime of the environment.

Keywords: leprosy, epidemiology, Geographic Information Systems, remote sensing, climate, Genetic Algorithm Rule-Set Prediction (GARP).

Introduction

Two decades of the multi-drug treatment approach of the World Health Organization (WHO) Global Leprosy Elimination Campaign has reduced the global prevalence of leprosy by 90% resulting in a current prevalence of 1.4 per 10,000 people (The Global situation of leprosy, 2002). However, this reduction in prevalence does not appear to be successful in interrupting transmission. For the past 10-15 years, stable numbers of

new cases have been recorded yearly, and in the six countries that account for approximately 90% of all new cases, the incidence is actually rising (Durrheim and Speare, 2003; Lockwood, 2002). Some health agencies explain this increase as better detection due to improved case-detection services and special education campaigns. Although this might be true for the incidence among patients with advanced disease or disability, it does not explain why children make up 13% of the new cases (Lockwood, 2002).

The relationships between *Mycobacterium leprae*, its human host and way of transmission are not clear. A preferential route of entry of the leprosy bacillus is the nose (Chehl et al., 1985) and multibacillary leprosy (MB) patients, who harbor viable bacilli in their nasal secretions, are thought to be the main source of infection (Hastings, 1993). However,

Corresponding author:
Azeb Tadesse Argaw
Mount Sinai School of Medicine
One Gustave L. Levy Place
10029 New York, USA
Tel. +1 212 241 6128; Fax +1 212 348 1310
E-mail: Azeb.tadesse@mssm.edu

little correlation exists between the prevalence of lepromatous leprosy (LL) and total new cases registered. The association of new cases with the incidence of conjugal leprosy or frequent contact between individuals is also insignificant (Jopling, 1988). The rate of infection by *M. leprae* is much greater than the rate of disease development, indicating the presence of sources of infection other than leprosy patients (Fine et al., 1997). Alternative sources of *M. leprae* infection could be sub-clinically infected individuals, a vector (vertebrate or insect) or environmental contamination.

Although, according to WHO, Ethiopia reached the leprosy elimination target of 1 case/10,000 population in 1999, the incidence has not changed appreciably (Tuberculosis and Leprosy control of Ethiopia 10th Annual Review meeting, 2002). As in other endemic countries, about 5,000 new cases are detected yearly. In 2002, clusters of endemicity with prevalence rates higher than the elimination target were recorded in four of the 14 administrative regions in the country (Tuberculosis and Leprosy control of Ethiopia 10th Annual Review meeting, 2002). Environmental sources have long been suspected to play a role in the spread of leprosy, especially due to the geographical differences in leprosy distribution and local clustering. Previous beliefs that *M. leprae* cannot survive once outside the human body are now being challenged. Under favorable conditions, such in a hot humid climate, *M. leprae* is capable of surviving for months (Desikan and Sreevatsa, 1995).

The objective of the work presented here is to use Geographic Information Systems (GIS) methods to assess the influence of environmental factors on the prevalence of leprosy in Ethiopia.

Materials and methods

The environmental attributes selected for analysis were the Normalized Difference Vegetation Index (NDVI), daytime earth surface temperature (Tmax), and a climate surface grid (5x5 km cell size) interpolated from long-term-normal climate station data (Table 1). These attributes were analyzed for their suitability for leprosy prevalence in the Oromia region in Ethiopia. ArcView GIS 3.3 software (ESRI, Redlands, CA, USA) and Genetic Algorithm for Rule-set Prediction (GARP) software (DesktopGarp 1.1.3, University of Kansas Biodiversity Research Center, Ka, USA) were used to build models and to create maps showing the distribution of leprosy in the region. An independent leprosy data set from the Amhara region was used to validate the leprosy prediction model.

Baseline data

Leprosy case data, prepared by the Amhara and Oromia Regional Bureaus of Health and the Ethiopian Ministry of Health, included both multi-bacillary leprosy (MB) and pauci-bacillary leprosy (PB). As of June 2001, the total number of PB (N=1367) and MB (318) cases recorded in 285 zonal health institutions in the Amhara region was used in this study. The total PB (N=136) and MB (N=2402) cases for the Oromia region collected as of August 2003, comprised data collected from 124 health institutions.

Table 1. Data used for the analysis.

Parameter	Feature type	Source
NDVI	8-bit Image	United States Geological Survey (USGS), http://www.gnosis.org
TMax	16-bit Image	USGS, http://www.gnosis.org
ACT Climate	Grid	Blackland Research and Extension Center, Texas Agricultural experiment station
NDVI, TMax for GARP	Image	Clipped and modified with GARP Dataset Manager

Determination of geographical coordinates

The latitude and longitude of each clinic were obtained in geographic decimal degree format from the website of the National Imagery and Mapping Agency (NIMA) (http://gnps.www.nima). Coordinates not found were obtained from Encarta (Microsoft Encarta, 2002). Digital maps based on the point coordinates for each clinic were prepared using ArcView GIS 3.3 (ESRI, Redlands, CA, USA) software.

Database

All parameters tested in this study were taken from the Minimum Medical Database (MMDb) (Malone et al., 2001) which is a collection of data gathered by collaborating health workers and scientists interested in the application of GIS to control program management and research for schistosomiasis and other snail-borne diseases. This database contains information for the Intergovernmental Authority of Development/Nile Basin (IGAD/Nile) region of East Africa. It includes maps and compiled data on environmental attributes such as NDVI, Tmax, elevation, climate, land-use, soil types, infrastructure, political boundaries and population data.

Prevalence calculation and normalization of data

To avoid confounding factors like clinic size and the number of people living around each clinic and to obtain comparable data, we calculated the prevalence of leprosy per 10,000 people for both the Amhara and Oromia regions. For Oromia, we used the population census provided at administrative zone levels with the health data. For the Amhara region the total population per each zone was estimated by creating a 10 km diameter buffer zone surrounding each clinic and extracting data from a population density grid map.

Extraction of environmental data

All environmental feature data, including

Advanced Very High Resolution Radiometer (AVHRR) satellite sensor data, were based on a 10-km diameter buffer zone around each clinic.

NDVI (1992-1995)

AVHRR data on NDVI were obtained from the United States Geological Survey (USGS) global 1km^2 website (Huh and Malone, 2001). NDVI is an indicator of plant photosynthetic activity related to environmental moisture; it is an index that ranges from 0 to 200 depending on the amount of vegetation at a site. The AVHRR data, which consisted of 10-day composites of daytime imagery, was processed by the USGS to minimize cloud cover and atmospheric attenuation of the sensor signal. All images were calibrated and geo-referenced to a geographic decimal degree latitude and longitude coordinate system using ERDAS Imagine 8.6 image processing software. The images were averaged to create monthly composites, which were averaged together to create maps for dry (October-March) and wet (April-September) seasons and an annual composite map, which included both seasons. The NDVI image files were converted into a grid file and used for extraction of mean NDVI values for each buffer zone. The extracted data were exported as Dbase files into Microsoft Excel datasheets. Scatter diagrams of cases/10,000 versus mean NDVI were plotted using Microsoft Excel. Outliers such as buffer zones falling fully or partially in water were excluded. A range of NDVI that includes all prevalence data points was determined from the plot and used as the value range specified to perform queries in ArcView to create map overlays showing areas suitable for leprosy prevalence.

Tmax (1992-1995)

Similar to the NDVI data, the image files for Tmax were prepared for the dry, wet and annual seasons. The same extraction procedures were applied to obtain a range of Tmax suitable for the

prevalence of leprosy in the regions. As for NDVI, scatter plots were prepared using Microsoft Excel and the range was used as criterium for queries in Arc view 3.3 GIS to create map overlays. The Amhara prevalence data points were overlaid on the model and points falling inside and outside the model were recorded for calculation of positive and negative predictive values.

Almanac Characterization Tool (ACT) climate data

A climate grid file with 5x5 km^2 cell size was used. It contained long-term normal data on precipitation (PRE), potential evapotranspiration (PET), the ratio of precipitation and potential evapo-transpiration (PPE) plus the minimum temperature (IT) and maximum temperature (XT) for January to December of each year. PPE, an indicator of the availability of environmental moisture or 'water budget', was included to validate the previously reported association of *M. leprae* with environmental moisture (Sterne et al., 1995). The leprosy prevalence point table was joined to the climate data table to extract climate data values for the grid cell in which each prevalence data point fell. The linked tables were exported as a DBase file and mean temperature ([IT + XT]/2) and mean PPE were calculated for the dry season (October to March) and the wet season (April to September) to give dry temperature, wet temperature, dry PPE and wet PPE values. Scatter plots of these values versus the stratified prevalence/10,000 (0-0.1, 0.1-0.7 and >0.7) were used to define ranges that included 95% of all points (Table 2). These data ranges were used as the extrapolation criteria to create map overlays using ArcView 3.3 GIS.

Ecologic niche-modeling using GARP

GARP is a high precision, computer-based genetic algorithm system used in ecologic niche modeling. It has a high predictive ability of a species' potential distributions based on user selected environmental layers. GARP incorporates different procedures that involve powerful rules including atomic rules, logistic regression analysis and frequency distribution-based bioclimatic rules. The GARP model has a minimum of errors of omissions and commissions making it an important tool in understanding the geographical distribution of the species of interest (Towsend and Vieglais, 2001; Costa et al., 2002). It equally divides the loaded prevalence data into training and test data sets performing multiple iterations based on rule selection, evaluation, testing and incorporation or rejection. The training data sets are subjected to randomly selected statistical evaluation methods like logistic regression and bioclimatic rules to develop a procedure to be used to evaluate predictive accuracy based on the test data sets and points randomly sampled from the whole region as a whole (Towsend and Vieglais, 2001).

The annual composite data for Tmax and NDVI were used as environmental dimensions for ecologic niche-modeling. Environmental layers for the analysis were prepared by converting the image files to the grid and processing of these files using the GARP module within ArcView GIS 3.3. These files were transformed to GARP-compatible format with the GARP Dataset Manager. The Amhara and Oromia leprosy prevalence data were combined to generate a total of 317 occurrence points. These distributional points were loaded into Desktop

Table 2. Temperature and potential evapo-transpiration ranges used in ACT climate query.

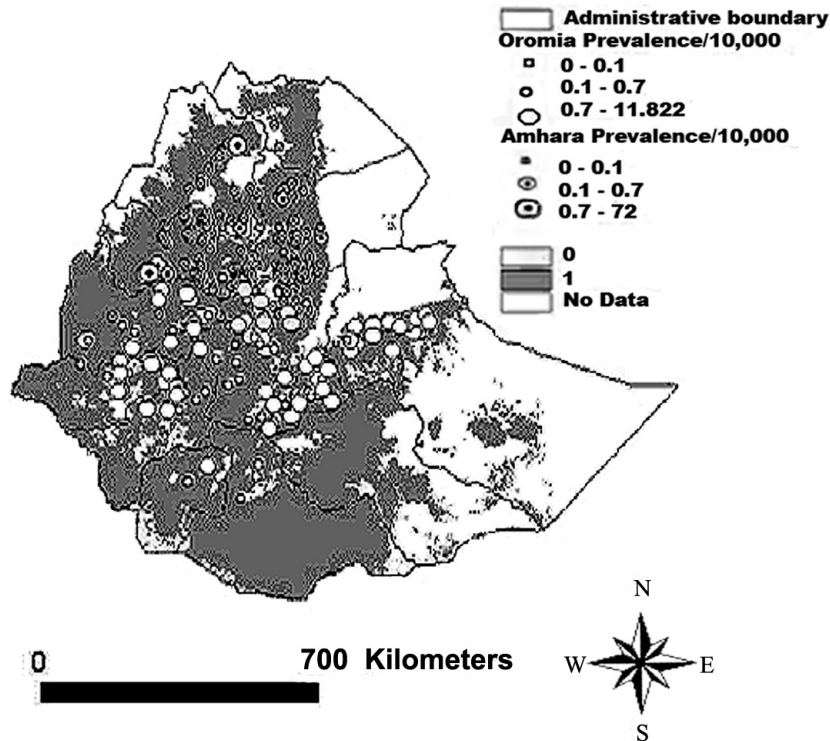| Prevalence	Dry season T (°C)	Wet season T (°C)	Dry season PPE	Wet season PPE
0 - 0.1	11.6 - 17.9	11.6 - 21.0	0.12 - 0.42	1.03 - 2.61
0.1 - 0.7	9.8 - 24.3	9.6 - 25.3	0.12 - 0.6	0.59 - 2.9
> 0.7	8.9 - 21.8	9.7 - 24.5	0.13 - 0.76	0.6 - 3.05

Fig. 1. GIS annual composite model for leprosy in Ethiopia. Query that met the ranges of NDVI (116.2 - 154.46) and TMax (9.4 - 28.82) in relation to leprosy prevalence in the Oromia region was overlaid on the political boundary map of Ethiopia. Area in grey represents the "best fit" model. The positive predictive value calculated by using data from the Amhara region was 90.476%.

GARP and used to model ecologic niches and potential geographic distribution of *M. leprae*. For each set of data used, GARP generates 20 models. Three criteria were used for the selection of the best model:

(i)   elimination of models with high number of training points omitted and, when possible, including only those models with 0 omission;
(ii) the average of total area predicted for species to be present was calculated and models falling within 5% range below or above this average were selected; and
(iii) models that passed the two criteria were compared based on the p value for Chi-square analysis of the probability of a random sample predictions being similar to the one generated by GARP.

Results

The AVHRR model using NDVI and Tmax

The ranges that included 95% of all leprosy prevalence points were defined from the scatter plots of Oromia prevalence data versus NDVI or Tmax data. The NDVI value ranges were 116-154.5 for the annual composite, 116-161 for the dry-season map and 115-155 for the wet-season map. The corresponding Tmax ranges were 9.4-28.8°C, 11.7-32.5°C and 7.7-25.8°C. When these ranges were used to prepare query-based map overlays, showing all areas that met both the NDVI and Tmax criteria ranges in the country, with the annual composite model it became clear that whole country, except for areas in the South East (Somali) and North East (Afar), are indeed suitable for leprosy (Fig. 1).
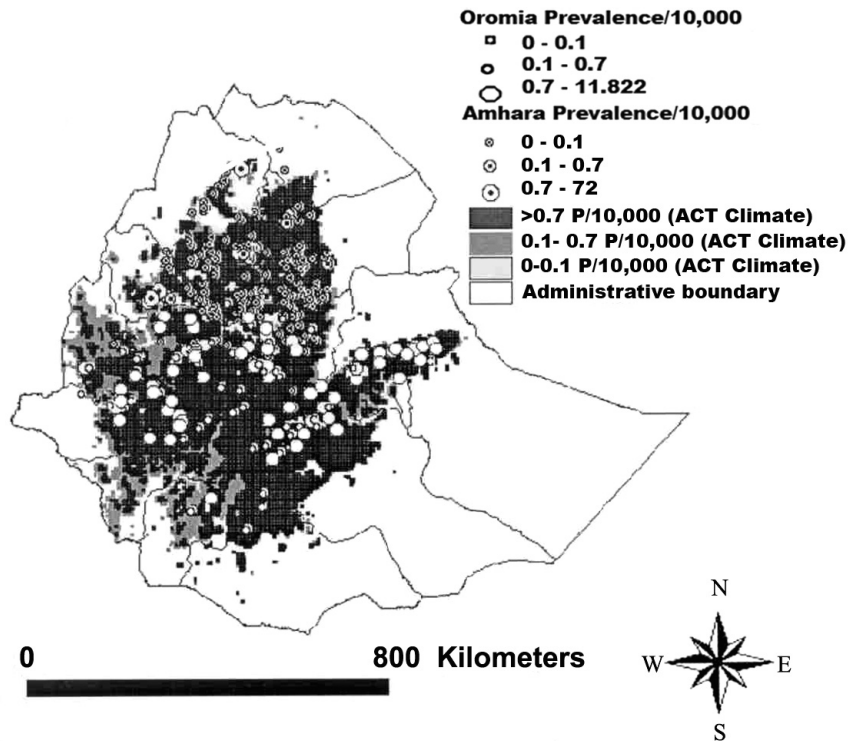
Fig. 2. Climate Prediction Model for Leprosy in Ethiopia. Point potential evapo-transpiration (PPE), average maximum and minimum temperatures for dry and wet seasons were used to study the influence of thermal-hydrologic regime on the prevalence of leprosy in Ethiopia. The ACT 5x5 climate grid was used for extraction. The Oromia prevalence table was joined to the ACT data and exported as a Dbase file.

In addition, the positive predictive value of the model developed in the Oromia region was 90.5% when applied to Amhara region data. All of the prevalence points that fell outside the predicted area were in the lowest ranges of prevalence/10,000 (0-0.1).

The climate-based model

The ACT climate grid was used to determine the influence of thermal-hydrologic regime, i.e. rain/potential evapo-transpiration (PPE) and mean temperature, in relation to leprosy distribution. Table 2 shows the defined ranges of PPE and minimum and maximum temperatures for the dry and wet seasons used for model development queries. This model predicted most of the Amhara and the Oromia region as suitable for leprosy. The excep-

tions were the southern part, most of the Southern Nations, Nationality and People Region (SNNPR), Harari and DireDawa (Fig. 2). The model excluded Northern Tigray, Western Gambela, Somali and most parts of Afar, especially the Eastern part. The positive predictive value generated when using the model parameters based on Oromia data applied to the Amhara data was 87.8%. Except for one health clinic (Chagnie) with a prevalence/10,000 of 1.32, all the prevalence points not predicted by the model had a prevalence/10,000 in the 0-0.1 range.

The GARP model

The GARP *M. leprae* distribution model showed itself as the most powerful of the models tested due to restrictions and statistical methods employed. The GARP model prediction area was based on data
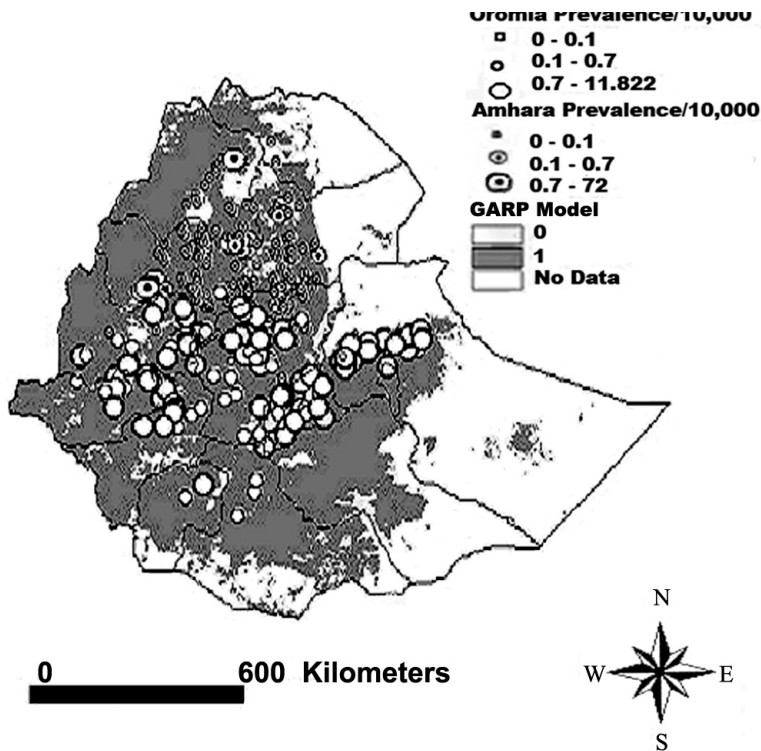
Fig. 3. GARP predictive model for the distribution of *M. Leprae* in Ethiopia. The environmental layers used are the annual TMax and NDVI. The model with minimum percentage of omission and within 5% range of the average presence area was selected.

points for all of Amhara and Oromia while the two models based on query of climate grid data and AVHRR data ranges for Tmax and NDVI were based on the Oromia data only. The prediction area pattern of the GARP ecologic niche-model resembled the AVHRR annual composite prediction model. It included most of the Amhara, Oromia, SNNP, Tigray and Gambela regions and excluded the Afar and Somali regions (Fig. 3).

Discussion

Understanding of the natural history of a disease, its ecologic distribution and the essential host-pathogen interactions are necessary in order to establish a successful control program. It is important to understand that, inspite of its success, the

WHO leprosy elimination campaign, mainly based on case finding, treatment and follow-up of household contacts, has not reduced the incidence of leprosy. Enhanced detection of hidden cases cannot explain the new occurrence of leprosy among young children, an indicator of ongoing active transmission (The Global situation of leprosy, 2002).

Several countries that have reached the elimination target still have localities with clusters of leprosy cases. These high endemicity regions must be eliminated in order to achieve sustained low transmission and reduced incidence rates. Ethiopia is one example where leprosy elimination has been achieved according to WHO criteria, but districts with leprosy prevalence above the 1/10,000 cut-off limit still exist. The transmission rate in children below 15 years of age is about 6% (Tuberculosis and Leprosy control of Ethiopia 10th Annual

Review meeting, 2002). Out of the total number of leprosy cases recorded in Ethiopia in 2002, 46% occurred in Oromia, the largest administrative region in the country, followed by the Amhara region with 31.5% of the total new cases (Durrheim and Speare, 2003).

M. leprae has been shown to be capable of withstanding severe adverse environmental conditions. Hot and humid weather, wet soils and water have all been proposed as factors that favor survival of the bacilli for a few months (Tuberculosis and Leprosy control of Ethiopia 10th Annual Review meeting, 2002; Desikan, 1977). *M. leprae* DNA has been detected in water in Indonesia at locations where a high proportion of individuals utilizing this water were affected by leprosy (Matsuoka et al., 1999). A well-controlled study in Karonga district in Malawi has shown water and soil moisture to be associated with high prevalence of leprosy (Sterne et al., 1995). Indirect transmission of leprosy is further supported by the natural occurrence of leprosy in animals like the armadillo (Job et al., 1986) and reports on the potential of insects to transmit leprosy (Geater, 1975; Banerjee et al., 1990).

Our study indicates that leprosy may be associated with specific environmental features. The models we developed are in concordance with leprosy distribution in Ethiopia and support the conclusion that: (i) certain thermal-hydrological regimes favor survival of leprosy in the environment, and that (ii) NDVI and Tmax environmental satellite data may be incorporated into a predictive model of leprosy that could be used to guide control programs. Thermal-hydrological regime risk factors for leprosy were measured by conventional climate station data or by satellite-sensor data on NDVI and Tmax as surrogates of moisture and temperature, respectively.

The climate-based model with stratified prevalences (0-0.1, 0.1-0.7 and >0.7 cases/10,000 population) predicted that the highest risk in the country occurred in Amhara and Oromia, the two regions with highest leprosy incidence (Tuberculosis and Leprosy control of Ethiopia 10th Annual Review meeting, 2002). The GARP analysis of NDVI and Tmax annual composite values as input variables confirmed what we observed in the annual AVHRR Tmax/NDVI model. It predicted most of Amhara and Oromia regions as the two major leprosy areas in Ethiopia. The ecologic niche of a species is the suitable ecologic space within which a species is potentially able to maintain population without immigration. All three models excluded the dry and hot Afar and Somali area, the two regions with the least number of leprosy cases in Ethiopia in 2002 (Tuberculosis and Leprosy control of Ethiopia 10th Annual Review meeting, 2002).

Various parameters have been identified as risk factors in acquiring leprosy disease. These include demographic factors such as age and sex, and immunological factors such as the absence of BCG vaccination (Ponnighaus et al., 1993). A significant negative correlation between the incidence of leprosy and higher economic indices has been reported (Di et al., 1992). Rural endemic settings with poor housing and sanitation conditions have been described as important factors in the transmission of leprosy (Mani, 1996). For any of these factors to have a role in the development of leprosy, the sustained presence of *M. leprae* in the environment is necessary. We propose that leprosy occurs most frequently when a suitable micro-environment such as moist soil coexists with other known or unknown predisposing factors.

This study provides evidence that environmental factors are of importance in the prevalence of leprosy. Such factors must be taken into consideration when planning a control program. Further studies that incorporate field-collected validation data (ground truth) may shed more light into the precise thermal-hydrological regime factors or other risk factors associated with the environmental risk of leprosy.

Acknowledgements

gather data, and students and staff at Dr J.B. Malone's lab for assisting in GIS work.

References

Banerjee R, Chaudhury S, Hati AK, 1990. Transmission of *Mycobacterium leprae* from lepromatous leprosy patients to the skin of mice through intermittent feeding. Trop Geog Med 42, 97-99.

Chehl S, Job CK, Hastings RC, 1985. Transmission of leprosy in nude mice. Am J Trop Med Hyg 34, 1161-1166.

Costa J, Towsend AP, Beard CB, 2002. Ecologic niche modeling and differentiation of populations of *Triatoma brasiliensis* Neiva, 1911, the most important Chagas' disease vector in Northeastern Brazil. Am J Trop Med Hyg 67, 516-520.

Desikan KV, Sreevatsa, 1995. Extended studies on the viability of *M. leprae* outside the human body. Lep Rev 66, 287-295.

Desikan KV, 1977. Viability of *Mycobacterium leprae* outside the human body. Lep Rev 48, 231-235.

Di Z, Yuewen N, Jingzeng Z, 1992. Leprosy-economy-environment; might we predict leprosy incidence from this point of view? Int J Lep 61, 631-632.

Durrheim DN, Speare R, 2003. Global leprosy elimination: time to change more than the elimination target. J Epidemiol Comm Health 57, 316-317.

Fine PE, Sterne JA, Ponnighaus JM, Bliss L, Saui J, Chihana A, Munthali M, Warndorff DK, 1997. Household and dwelling contact as risk factors for leprosy in Northern Malawi. A J Epidemiol 146, 91-102.

Geater JG, 1975. The fly as potential vector in the transmission of leprosy. Lep Rev 46, 279-286.

Hastings RC, 1993. Leprosy. Edinburgh: Churchill Livingston.

Huh OK, Malone JB, 2001. New tools: potential medical applications of data from new and old environmental satellites. Acta Trop 79, 35-47.

Job CK, Harris EB, Allen JL, Hastings RC, 1986. A random survey of leprosy in wild nine-banded armadillos in Louisiana. Int J Lep 54, 453-457.

Jopling WH, 1988. Handbook of leprosy. Heinemann Professional Publishing.

Lockwood DNJ, 2002. Leprosy elimination- a virtual phenomenon or a reality. BMJ 324, 1516-1518.

Malone JB, McCarroll JC, Kristensen TK, Yilma JM, Erko B, El-Bahy MM, Corbett JD, 2001. Minimum Medical Database spatial decision support system for the Authority on development-Nile Basin Region (IGAD-Nile), 49 pp.

Mani MZ, 1996. Ecologic factors in transmission of leprosy. Ind J Lep 68, 375-376.

Matsuoka M, Izumi S, Budiawan T, Natak N, Saeki K, 1999. *M. leprae* DNA in daily using water as a possible source of leprosy infection. Indian J Lep 71, 61-67.

Ponnighaus JM, Fine PE, Sterne JA, Bliss L, Wilson JR, Malema SS, Kileta S, 1993. Incidence rates of leprosy in Karonga district, Northern Malawi: patterns by age, sex, BCG status and classification. Int J Lep 62, 10-23.

Sterne JAC, Ponnighaus JM, Fine PEM, Malema SS, 1995. Geographic determinants of leprosy in Karonga district, Northern Malawi. Int J Epidemiol 24, 1211-1222.

The Global situation of leprosy control at the beginning of the 21st century, 2002. Int J Lep 70, S21-S22.

Townsend AP, Vieglais AD, 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. Bioscience 51, 363-371.

Tuberculosis and Leprosy control of Ethiopia. Tenth Annual Review Meeting, 18 to 20 September, 2002.