Meteorological, environmental remote sensing and neural network analysis of the epidemiology of malaria transmission in Thailand

Richard Kiang[1], Farida Adimi[2†], Valerii Soika[2†], Joseph Nigro[2], Pratap Singhasivanon[3], Jeeraphat Sirichaisinthop[4], Somjai Leemingsawat[3], Chamnarn Apiwathnasorn[3], Sornchai Looareesuwan[3]

[1]Code 610.2, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA; [2]Science Systems Applications, Inc., Lanham, Maryland 20706, USA; [3]Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand; [4]Vector Borne Disease Training Center, Ministry of Public Health, Saraburi, Thailand; [†]Current address: RS Information Systems, Inc., Lanham, Maryland 20706, USA

Abstract. In many malarious regions malaria transmission roughly coincides with rainy seasons, which provide for more abundant larval habitats. In addition to precipitation, other meteorological and environmental factors may also influence malaria transmission. These factors can be remotely sensed using earth observing environmental satellites and estimated with seasonal climate forecasts. The use of remote sensing usage as an early warning tool for malaria epidemics have been broadly studied in recent years, especially for Africa, where the majority of the world's malaria occurs. Although the Greater Mekong Subregion (GMS), which includes Thailand and the surrounding countries, is an epicenter of multidrug resistant falciparum malaria, the meteorological and environmental factors affecting malaria transmissions in the GMS have not been examined in detail. In this study, the parasitological data used consisted of the monthly malaria epidemiology data at the provincial level compiled by the Thai Ministry of Public Health. Precipitation, temperature, relative humidity, and vegetation index obtained from both climate time series and satellite measurements were used as independent variables to model malaria. We used neural network methods, an artificial-intelligence technique, to model the dependency of malaria transmission on these variables. The average training accuracy of the neural network analysis for three provinces (Kanchanaburi, Mae Hong Son, and Tak) which are among the provinces most endemic for malaria, is 72.8% and the average testing accuracy is 62.9% based on the 1994-1999 data. A more complex neural network architecture resulted in higher training accuracy but also lower testing accuracy. Taking into account of the uncertainty regarding reported malaria cases, we divided the malaria cases into bands (classes) to compute training accuracy. Using the same neural network architecture on the 19 most endemic provinces for years 1994 to 2000, the mean training accuracy weighted by provincial malaria cases was 73%. Prediction of malaria cases for 2001 using neural networks trained for 1994-2000 gave a weighted accuracy of 53%. Because there was a significant decrease (31%) in the number of malaria cases in the 19 provinces from 2000 to 2001, the networks overestimated malaria transmissions. The decrease in transmission was not due to climatic or environmental changes. Thailand is a country with long borders. Migrant populations from the neighboring countries enlarge the human malaria reservoir because these populations have more limited access to health care. This issue also confounds the complexity of modeling malaria based on meteorological and environmental variables alone. In spite of the relatively low resolution of the data and the impact of migrant populations, we have uncovered a reasonably clear dependency of malaria on meteorological and environmental remote sensing variables. When other contextual determinants do not vary significantly, using neural network analysis along with remote sensing variables to predict malaria endemicity should be feasible.

Keywords: malaria, Thailand, remote sensing, neural networks.

Corresponding author:
Richard Kiang
Code 610.2, NASA Goddard Space Flight Center
Greenbelt, Maryland 20771, USA
Tel. +1 301 614 5375; Fax +1 301 614 5268
E-mail: richard.kiang@nasa.gov

Introduction

The transmission of malaria is influenced by a myriad of factors. Environmental, climatic, social, economic, public health, political, and wartime conditions have all been shown to contribute to malaria occurrence and outbreaks. Among these, the environmental conditions, especially rainfall, appears to be the most recognizable determinant. The intensity of malaria transmission has long been associated with rainy seasons based on informal long-term observations in endemic areas. The excessive rain or droughts brought about by climatic events like El Niño Southern Oscillation (ENSO) have also been shown to enhance the occurrence of malaria epidemics in affected regions (Bouma and van der Kaay, 1996; Poveda et al., 2001; Githeko and Ndegwa, 2001; Gagnon et al., 2002; Kovats et al., 2003). Remote sensing is considered an important technologic tool for predicting, preventing, and containing malaria epidemics (MARA/ARMA, 1998; WHO, 2001; WHO, 2004a; WHO, 2004b) because environmental variables can be remotely sensed from earth observing satellites, and the influence of ENSO events may be forecast using satellite-measured parameters. In recent years, researchers have reported various methods and techniques for using meteorological data or remotely sensed measurements for forecasting malaria epidemics, in particular for Africa (Thomson et al., 1996; Hay et al., 1998; Kleinschmidt et al., 2000; Rogers et al., 2002; Nalim et al., 2002; Small et al., 2003; Abeku et al., 2004; Teklehaimanot et al., 2004a,b; Omumbo et al., 2004; Thomson et al., 2006). Some of these forecasting techniques have already been used in operational control programs (Grover-Kopec et al., 2005). Advances in Geographic Information Systems (GIS) have made it possible to integrate remote sensing measurements, epidemiological data, other information important to malaria transmission, and risk modeling results (Albert et al., 2000).

Since rainfall provides vector breeding sites and prolongs the vector life span by increasing water availability, precipitation or precipitation anomalies are the attributes most frequently used for predicting malaria epidemics. It has also been shown, however, that rainfall, or the lack of it, has a complex effect on malaria transmission in various parts of the world (Kovats et al., 2003). For example, although moderate rainfall may promote malaria transmission, intense and prolonged rainfall may flush away larval habitats and thus reduce transmission. Similarly, lack of rainfall does not always reduce larval populations. On the contrary, lack of rainfall may create new habitats, such as pools and puddles, in some regions and therefore increase larval population. In addition, droughts may be deleterious to predator populations or may cause human populations with no immunity to move to areas endemic for malaria (Kovats et al., 2003). These factors may indirectly increase overall malaria transmission. For regions where regular, yearly malaria infections contribute to partial immunity, a reduced transmission in certain years may increase human population vulnerability in later years.

Another meteorological variable that is often used for predicting malaria transmission is temperature. Warmer temperatures hasten larval and vector development (Craig et al., 1999) and shorten the sporogonic cycle thus prolonging the risk of infection. Warmer air holds more moisture and therefore encourages mosquito survival. The range of rainfall and temperature needed to maintain stable malaria transmission is called climate suitability (MARA/ARMA, 1998; Craig et al., 1999; Small et al., 2003; Omunbo et al., 2004; Hay et al., 2004). For example, in African regions where the *Anopheles gambiae* complex is the dominant factor in the transmission of falciparum malaria, climate suitability is associated with a temperature between 18º and 32ºC and a rainfall exceeding 80 mm per month for at least 3 to 5 months (MARA/ARMA, 1998; Craig et al., 1999).

Naturally, climate suitability depends on the ecology of the dominant malaria vector species, and

therefore varies with geographic region (WHO and UNICEF, 2005). Climate suitability indicates how favorable the regional climate is for stable malaria transmission, but how much this potential materializes into malaria endemicity depends on other contextual determinants. Socioeconomic conditions, public health infrastructure, herd immunity, irrigation and transportation projects, natural disasters, and military conflicts, have overriding effects on malaria transmission. When these contextual determinants are relatively unchanged, environmental determinants like rainfall and temperature are indeed the essential predictors for estimating the intensity of malaria transmission.

In previous studies, rainfall surrogates were often used for modeling when no suitable remotely sensed or ground based measurements were available (Hay et al., 1996; Thomson et al., 1996; Hay et al., 1998). The most frequently used surrogates include the Cold Cloud Duration (CCD) derived from Meteosat remote sensing measurements (Snijders, 1991) and the Normalized Difference Vegetation Index (NDVI) derived from Advanced Very High Resolution Radiometer (AVHRR) measurements. The NDVI is not a measure of the precipitation at the time of the satellite overpass, but an increase in NDVI over nonirrigated area indicates the greening of vegetation and therefore implies that rainfall was recently received. The NDVI is also good for estimating the extent of vegetation cover on the ground. It is therefore a useful indicator for differentiating between urban and rural areas. Direct, space-based rainfall measurement capabilities started with NASA's Tropical Rainfall Measuring Mission (TRMM) in 1999 (Kummerow et al., 1998). TRMM is expected to last through 2009. The successor of the TRMM is the Global Precipitation Measurement (GPM) mission, an international collaboration involving a constellation of satellites (Flaming, 2005; Smith et al., 2006).

Most malaria early-warning capabilities developed to date are for Africa. It is generally agreed that rainfall excess is the main determinant for malaria epidemics in lowland and in warm, semi-arid and desert-fringe areas. For highland areas, temperature or temperature and rainfall together are the main predictors (Hay et al., 2001; Thomson and Connor, 2001; Grover-Kopec et al., 2005). The Greater Mekong Subregion (GMS), which consists of Thailand, Myanmar, Laos, Cambodia, Vietnam, and the Yunnan Province of China, is an epicenter of multidrug resistant falciparum malaria (Kidson et al., 1999). The Mekong Roll Back Malaria Program has identified remote sensing and GIS as important elements for malaria prevention and control in this region (Thimasarn, 2003). Remote sensing has been shown to be useful in Thailand for detecting potential larval habitats of malaria vectors and for estimating the associated malaria risks in the vicinity (Sithiprasasna et al., 2005; Zollner et al., 2006). Early-warning capabilities similar to those available in Africa with respect to malaria do not yet exist in the GMS.

The objectives of this study are to examine the meteorological and environmental dependency of malaria transmission in Thailand at the provincial level. Because some satellite measurements were not yet available during the years in which the malaria epidemiological data were taken, meteorological data based on both ground observed and satellite remotely sensed measurements are used in this study. Therefore, another objective of this study is to examine the feasibility of using meteorological and environmental data of mixed origins and resolution to estimate malaria endemicity at provincial resolution scales.

Materials and Methods

Several categories of data are used in this study, including malaria cases, climate time series constructed from ground-based measurements, and satellite measurements. Meteorological and environmental parameters from these measurements will be used as independent variables to statistically model malaria cases.

Malaria data

All four human malaria species are present in Thailand. There are approximately equal numbers of *Plasmodium falciparum* and *P. vivax* malaria cases. Together, they account for approximately 99% of all cases. *P. malariae* malaria constitutes less than 1% of all cases and *P. ovale* malaria is rare (Thai Ministry of Public Health, 2003). Through concerted efforts in surveillance and treatment, and prevention and control, malaria morbidity and mortality in Thailand have declined significantly in the last three decades. The current annual parasite incidence is less than 1 per 1000 population. Foreign workers and migrant, displaced populations from neighboring countries (Myanmar, Cambodia, Laos, and Malaysia) contribute significantly to malaria transmission in Thailand. Implementing positive health care policy for the non-Thai population in recent years has helped lower the overall malaria prevalence.

The monthly, provincial malaria data compiled by the Epidemiology Division, Department of Disease Control, Thai Ministry of Public Health were used in this study. These data are based on passive detection, mainly confirmed malaria cases reported by hospitals and clinics. The data do not provide information on parasite species. Annual (but not monthly) statistics with breakdowns into age groups and Thai or foreigner groups are also provided. Since it is not known whether the cases are new, due to recrudescence, or relapses, the incidence rate cannot be directly calculated from the compiled data. In our analysis, we used the total number of monthly provincial malaria cases which groups parasite species and Thai or non-Thai populations together. Malaria data with higher spatial resolution (at district, village, and hamlet levels) and more epidemiological details (parasite species, mixed infection, ages, and nationality) are archived at the Department of Disease Control.

Understandably, the data only include symptomatic cases. In Thailand, there may be a significant number of asymptomatic cases among repeatedly infected adults but the distribution may be geographically dependent (Coleman et al., 2004; Pethleart et al., 2004). In addition, there are an unknown number of symptomatic cases among the migrant and displaced people who may not have sought or received treatment from public health organizations for a variety of reasons. The malaria cases used in the analyses therefore reflect the lower bound of the true prevalence.

Meteorological and environmental data

The malaria epidemiological data used for this study span from 1994 to 2001. For modeling, a variety of data sources is needed to provide the meteorological and environmental data.

Air temperature and precipitation data from 1994 to the end of 1999 are based on the Seasonal-to-Interannual Earth Science Information Partner (SIESIP) data set compiled by the Center for Climate Research of the University of Delaware USA. SIESIP is one of the Earth Science Information Partner (ESIP) projects funded by the National Aeronautics and Space Administration (NASA) to compile and develop customized Earth science data sets.

This data set was produced from the Global Historical Climatology Network (GHCN version 2) and Legates and Willmott's station records of monthly and annual mean air temperature and total precipitation. Using a spherical distance-weighting algorithm, station averages of monthly values were interpolated to a 0.5° x 0.5° latitude-longitude grid, with nodes centered on 0.25°. The number of nearby stations influencing grid node estimates was 20 on average. Both Digital Elevation Model-assisted interpolation and Climatologically Aided Interpolation were employed to estimate the monthly fields. This data set spans the time period from 1950 to 1999 (Vose et al., 1992; Easterling et al., 1996; Peterson and Vose, 1997).

From the beginning of 2000, we extracted monthly temperature data from the Moderate Resolution Imaging Spectroradiometer (MODIS) data set.

MODIS has two bands at 250 m resolution, five bands at 500 m, and 29 bands at 1,000 m, with its spectral region ranging from visible to thermal wavelengths. MODIS is a key instrument on board the Terra Earth Observing System AM platform (EOS AM) and Aqua (EOS PM) satellites. Data from MODIS improve our understanding of global dynamics and processes occurring on the land, in the oceans, and in the lower atmosphere. A wide variety of geophysical parameters can be derived from MODIS measurements. To be precise, the temperature parameter in the MODIS product is land surface temperature instead of air temperature. However, the average monthly air temperature can be approximated by the average monthly land surface temperature, since these two parameters exhibit similar seasonal trends.

Also, from the beginning of 2000 we extracted monthly precipitation data from rainfall data sets measured by the instruments on board the Tropical Rainfall Measuring Mission (TRMM) spacecraft (Kummerow et al., 1998). TRMM is a joint mission between NASA and the Japan Aerospace Exploration Agency designed to monitor and study tropical rainfall and to help our understanding of the water cycle in the climate system. Of the five instruments carried by TRMM, the Precipitation Radar and the TRMM Microwave Imager are most directly related to rain measurements. The TRMM precipitation data has a resolution of approximately 5 km at nadir.

When more than one data source is used, there may be valid and intrinsic differences due to the conditions under which the data were obtained. In these cases, a linear transformation was performed on each data stream to match up the statistical properties of the first data stream.

Relative humidity data were extracted from the National Centers for Environmental Prediction's (NCEP) Reanalysis Monthly Means and Other Derived Variables data set. The NCEP/National Center for Atmospheric Research (NCAR) Reanalysis Project uses a state-of-the-art analysis–forecast system to perform data assimilation using past data from 1948 to the present. A subset of this data was processed to create monthly means of a subset of the original data. These variables are instantaneous values at the reference time and are averages of instantaneous values at the four reference times - 0, 6, 12, and 18 Z -over the averaging period. Spatial resolution of the data set is a 2.5° by 2.5° latitude/longitude global grid. Alternatively, we can compute relative humidity from water vapor, which is one of the geophysical parameters available in the MODIS atmospheric profile product (http://modis-atmos.gsfc.nasa.gov).

Vegetation plays an important role in vector breeding, feeding, and resting sites. A number of vegetation indices have been used in remote sensing and Earth science disciplines. The most widely used index is the Normalized Difference Vegetation Index (NDVI) (Tucker, 1979). It is simply defined as the difference between the red and the near infrared bands normalized by twice the mean of these two bands. For green vegetation, the reflectance in the red band is low because of chlorophyll absorption, and the reflectance in the near infra-red band is high because of the spongy mesophyll leave structure. Therefore, the more vigorous and denser the vegetation, the higher the NDVI.

NDVI has also been used as a surrogate for rainfall estimate. However, although it is an effective measure for arid or semi-arid regions, vegetation index may be a less sensitive measure for estimating rainfall for tropical regions where ample rainfall is normally received.

The mean vegetation index over a region does reflect the degree of urbanization or lack of vegetation. In this sense, NDVI in a grid cell is used as an indicator for the mean level of vegetation present in the cell in question.

Any satellite instrument with red and infrared bands can be used to compute NDVI. However, because of the difference in band definitions, spatial resolutions, and satellite passing time, NDVI computed from different sensors must first be calibrated before the

NDVI from different sensors can be compared.

The NDVI data are processed and distributed by the NASA Goddard Space Flight Center's Goddard Earth Sciences Data and Information Services Center (GES DISC). These data are 8 km resolution monthly NDVI maximum value composite images from the GES DISC's Global Land Biosphere Data and Resources web site (http://disc.sci.gsfc.nasa.gov /guides/GSFC/guide/avhrr_dataset.gd.shtml). The original data set was produced as part of the National Oceanic and Atmospheric Administration (NOAA)/NASA Pathfinder Advanced Very High Resolution Radiometer (AVHRR) Land Program. The data set spans July 1981 through December 2000, with the exception of September through December 1994. We extracted the NDVI data for 2000 and beyond from MODIS measurements (http://modis-atmos.gsfc.nasa.gov).

Neural network methodology

We use the neural network (NN) method to approximate the dependency of malaria cases on the meteorological and environmental variables. This method has been successfully used in many applications, including classification, regression, time series analysis, and handwritten character recognition (Nelson and Illingworth, 1990). In this approach, the probability density of the data is not assumed to follow any particular functional form. Rather, the characteristics of the probability density are determined entirely by the distribution in the data, hence, it is a data driven approach. This method is most suitable for problems that are too complex to be expressed in a closed, analytical form. For problems in which there are hidden, implicit variables, this approach is particularly suitable, as it is difficult to either specify the variables properly or sufficiently account for their effects mathematically.

This method is called neural network because it resembles how biological neurons function (Gardner, 1993). Nodes in a neural network are analogous to neurons; the connections between the nodes are analogous to synapses. The behavior of

the activation function corresponds to the firing of a neuron. The weights of the connections can be trained to give the aggregate of neurons a specific functionality. A network may accommodate complicated geometries in multidimensional space by incorporating hidden layers. Without hidden layers, the neural network method will be equivalent to the generalized linear model.

To train our neural network model, we feed observed or measured parameters from the past into the network. The input parameters may consist of meteorological, environmental, and other variables and the output parameter is the corresponding malaria cases for that specific location and time. Once trained, the network will be able to estimate the cases at some other time period using the parameters corresponding to that time period.

The neural network used in this study is in the class of multi-layer perceptron (Rumelhart and McClelland, 1986; Haykin, 1994; Bishop, 1996). The general network architecture is composed of an input layer, one or more hidden layers, and an output layer. Each layer consists of a number of nodes. In this study, meteorological and environmental data are the main parameters fed into the input layer; and the malaria cases or other data indicating malaria prevalence are the parameters generated from the output layers. A hidden layer consists of one or more hidden nodes. The function of the hidden layers in a neural network is to map the data structure into a new representation that facilitates the optimization of the objective function. For example, if the objective function is to maximize classification accuracy, hidden layers will transform the input parameters into functions of the parameters to make the classes more readily separable. Without hidden layers, a neural network may only differentiate linearly separable classes. Because the complexity of the data structure and the objective function drive the construction of hidden layers, trial and error is the usual approach to determine the numbers of hidden layers (HL) and hidden nodes (HN) to be used. In fully interconnected networks, weight decay (Bishop, 1996) can be used to

eliminate nodes and links that are insensitive to the optimization of the objective function.

In the hindcasting (or retrospective forecasting) mode, the model is used to estimate historical cases. The model's estimation accuracy can then be determined by comparing the model output with the events that actually took place. Although not a topic of this paper, future malaria cases can be predicted by using forecast parameters as input in the forecasting mode. Once a model is trained with past epidemiological data for a region, estimates on current malaria endemicity for that region can be obtained by feeding current meteorological and environmental data into the trained model.

For the remainder of this paper, we denote the average temperature by T, the precipitation amount by P, the precipitation amount in the previous month by P_{-1}, the relative humidity by H, and vegetation index by V. The major malaria vector species in Thailand include *Anopheles dirus*, *An. minimus*, and *An. maculatus*.

Various neural network architectures were used in this analysis. The most suitable architecture may vary from province to province. For the ease of discussion, four configurations are reported in the following – networks with one hidden layer (HL) imbedded with one, two, or three hidden nodes (HN). The input variables include P, P_{-1}, T, H, V. Time (t) is also used as an input parameter to account for trends that are independent of meteorological and environmental variables. The trend can be linear or nonlinear. This time factor may reflect the advances in malaria detection and treatment methods, improvement in public health support, establishment of more cultivated crop areas, construction of transportation routes and irrigation projects, and changes in the influx of refugees and migrant populations. In general, the time factor helps to account for the effects of the changes in non-meteorological and non-environmental contextual parameters on malaria transmission during the time period under study.

We selected three provinces (Kanchanaburi, Mae Hong Son, and Tak) for more intensive study. These provinces are among the provinces with the most intense transmission. The locations of these three provinces are shown in Fig. 1. To examine how well networks with these configurations perform, the six years of data are divided into six groups. Each group consists of five years of data for training and one year of data for testing. The average over the six groups of the root-mean-square error between the real cases and the fitted cases normalized by the real cases is used as an accuracy measure. The network for each input data combination was trained using backward propagation (Haykin, 1994; Bishop, 1996) for a million epochs or until the training errors converged. An epoch is a complete round of training over all the input samples. Although the training might not have completely converged after a million epochs, the decrease in the value of the objective function and the changes in the network parameters at this point were negligibly small from one epoch to the next.

Software

We developed the majority of the processing, modeling, and analysis software in IDL and C, including a neural network code in C. Commercial software used in this study include ENVI/IDL 4.2 (ITT Visual Information Solutions, Boulder, CO, USA), Matlab 7.0.4 (MathWorks, Natick, MA, USA), NeuroSolution 4.3 (NeuroDimension, Gainesville, FL), and ArcGIS 9.0 (ESRI, Redlands, CA, USA).

Results

Table 1 shows the training and testing results from the neural network. Configuration 1 in the table shows the simplest architecture. For example, the average training accuracy is 75±9%, and testing accuracy is 67±10% for Kanchanaburi. Like in other classification techniques, the training accuracy is normally higher than the testing accuracy. This is due to the differences in statistical characteristics
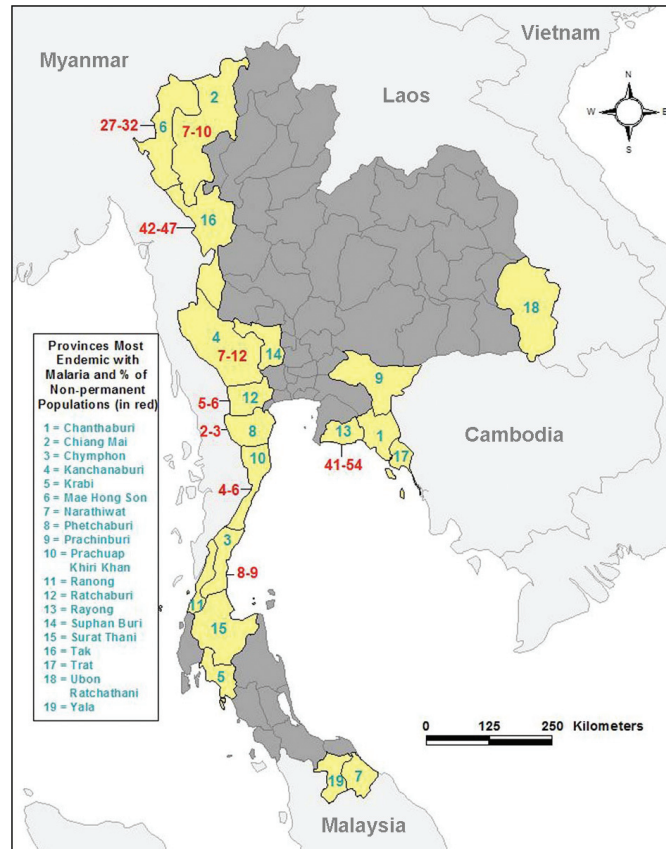
Fig. 1. The 19 provinces for which malaria transmission were analyzed. The percentage of non-permanent populations is shown for some provinces.

between the training and testing samples. In the context of this study, it implies that malaria transmission in the training and testing samples may respond differently to changes in meteorological and environmental parameters. When the temperature parameter is removed from the input (Configuration 2), the accuracy for training and testing in Kanchanaburi is reduced to 74±9% and 62±12%, respectively. When another hidden node is included (Configuration 3), more complex geometries can be introduced to assure better classification. For example, the average training accuracy becomes 83±6% for Kanchanaburi. The testing accuracy, however, reduces to 57±16%. This indicates that the more complex geometry might have been constructed to accommodate the noise components in the training

samples, and thus worsened the testing accuracy. When one more hidden node is included (Configuration 4), the training accuracy increases further, while the testing accuracy decreases and becomes erratic for Mae Hong Son indicating that the network is over-trained.

As described above, training accuracy is expressed as normalized root-mean-square error by comparing network output with malaria cases. Because of the uncertainty regarding the reported malaria cases, accuracy measures may be defined less stringently in the following way. We divided the malaria cases into 20 bands (classes) – from zero to 1.5 times of the historical maximum. As long as the fitted data falls into the correct band or one of the two adjacent bands, the classification will be considered

Table 1. Training and testing accuracy for modeling provincial malaria cases using neural networks in four specific configurations.

	Configuration 1 t, T, P, P -1 , H, V I HL 1 HN		Configuration 2 t, P, P -1 , H, V I HL 1 HN		Configuration 3 t, T, P, P-1, H, V I HL 2 HN		Configuration 4 t, T, P, P-1, H, V I HL 3 HN	
Province	training	testing	training	testing	training	testing	training	testing
Kanchanaburi	75±9%	67±10%	74±9%	62±12%	83±6%	57±16%	88±4%	58±17%
Mae Hong Son	71±10%	57±11%	69±11%	56±6%	77±8%	56±7%	84±6%	10±50%
Tak	72±10%	64±6%	70±11%	63±8%	78±8%	55±20%	84±6%	48±22%

correct. The upper limit of the highest band was set to infinity. The architecture used for Configuration 1 in Table 1 was then applied to the malaria data for the 19 provinces that were most endemic for malaria between 1994 and 2000. Aside from the Kanchanaburi, Mae Hong Song, and Tak provinces that were selected before, the other 16 provinces are, in alphabetical order, Chanthaburi, Chiang Mai, Chymphon, Krabi, Narathiwat, Phetchaburi, Prachinburi, Prachuap Khiri Khan, Ranong, Ratchaburi, Rayong, Suphan Buri, Surat Thani, Trat, Ubon Ratchathani, and Yala (Fig. 1). The malaria endemicity in the other 57 provinces in Thailand ranges from very low to none. The training accuracy results from these 19 provinces and the width of each band are shown in Table 2. The aver-

Table 2. Training and testing accuracy for the 19 Thailand provinces most endemic with malaria. Malaria cases are grouped into 20 band classes to account for the uncertainty regarding the reported malaria cases. The band width is both expressed in absolute numbers of malaria cases per 100,000 population. The training accuracy is based on 1994-2000 data and the testing accuracy on 2001.

| Province	Training Accuracy	Testing Accuracy	Band Width	Band Width per 100,000 Population
Chanthaburi	0.55	0.08	180	37
Chiang Mai	0.57	0.00	84	5
Chymphon	0.70	0.58	140	30
Kanchanaburi	0.73	0.58	503	64
Krabi	0.76	0.00	59	16
Mae Hong Son	0.88	0.75	435	187
Narathiwat	0.81	0.42	83	12
Phetchaburi	0.74	0.33	92	20
Prachinburi	0.64	0.00	23	5
Prachuap Khiri Khan	0.95	0.50	287	60
Ranong	0.77	0.92	126	79
Ratchaburi	0.79	0.42	182	22
Rayong	0.61	0.33	35	7
Suphan Buri	0.61	0.33	27	3
Surat Thani	0.76	0.67	304	34
Tak	0.64	0.67	408	84
Trat	0.45	0.17	203	91
Ubon Ratchathani	0.52	0.42	91	5
Yala	0.86	0.92	112	26
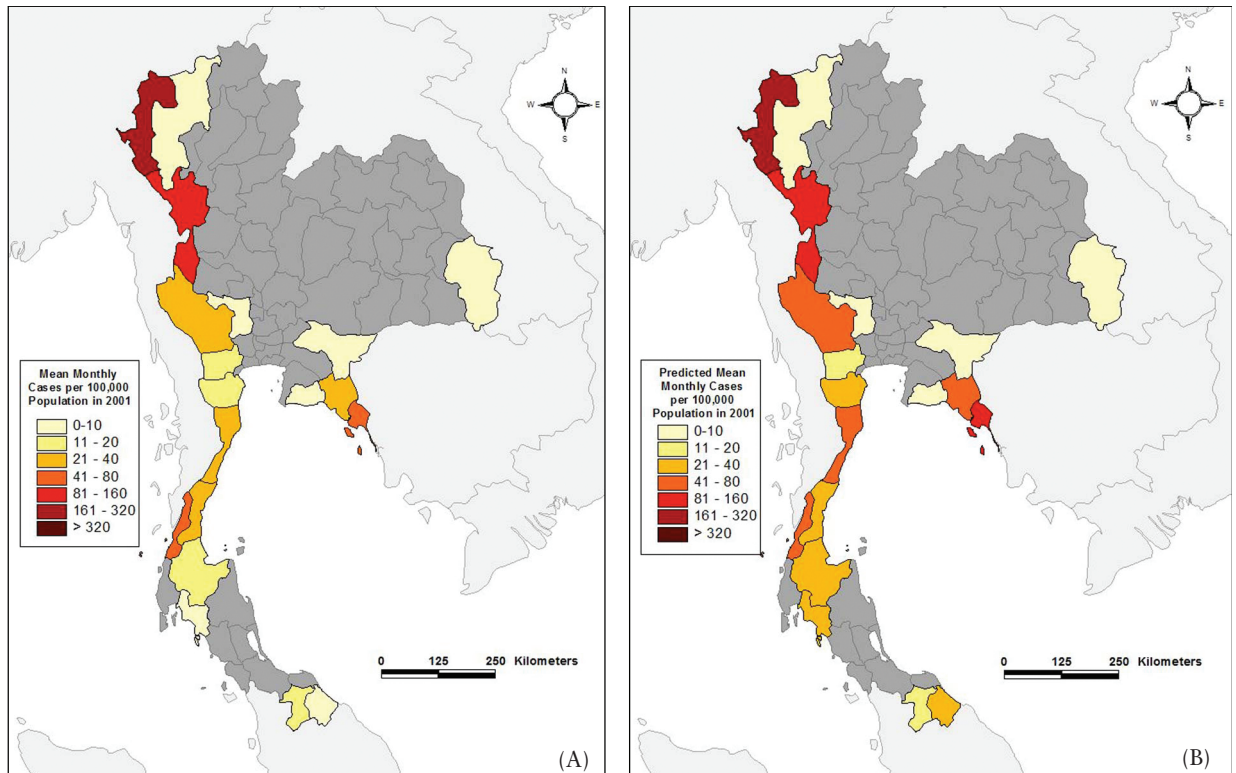
Fig. 2. (A) Mean monthly malaria cases per 100,000 population in 2001. (B) Predicted (hindcast) mean monthly malaria cases per 100,000 population in 2001 by neural networks trained with 1994-2000 data. Because of the significant reduction in malaria transmission in 2001, the networks overestimated malaria transmission in some provinces.

age training accuracy of the 19 provinces, weighed by the malaria cases, is 73%. We then used the neural networks trained with 1994-2000 data to predict (hindcast) the malaria cases in 2001. The testing or prediction accuracy was computed in the same way as the training accuracy and is given in Table 2.

The average testing accuracy weighted by malaria cases is 53%. When weighted by the number of cases per 100,000 population, the training and testing accuracy became 74% and 57%, respectively. Note that testing accuracy is the probability for the prediction to fall into the correct band. In 2001, the malaria cases in these 19 provinces experienced a 31% decrease. Because there was no apparent climatic or environmental change, this decrease was most likely due to changes in other contextual determinants, and was generally attributed to the effectiveness of malaria control efforts in Thailand. The

prediction or testing therefore overestimated the malaria cases for 2001. The actual average monthly malaria cases per 100,000 population in 2001 are shown in Fig. 2A.

The predicted cases are shown in Fig. 2B. In this representation, 11 provinces were correctly predicted, and 8 provinces were over predicted. Therefore, the testing or prediction accuracy at province level is 58%, which largely agrees with the accuracy measure in the band representation.

Discussion

Malaria transmission is dependent on the diverse factors that influence the vectors, parasites, human hosts, and the interactions among them. These factors may include, among others, meteorological and

environmental condition, the innate and adapted immunity of the human hosts, public health system, housing construction, vector control, construction of transportation networks, irrigation projects, and population movements. The couplings among these factors may be so complex that it is difficult to isolate the key factors that promote or sustain malaria transmission in an area.

The most apparent determinants are meteorological and environmental factors, including the parameters evaluated here: rainfall, temperature, humidity, and vegetation. Experience has shown that malaria is correlated with the rainy season, and that ENSO events may either increase or decrease malaria transmission. When other factors remain more or less constant, the meteorological and environmental conditions can indeed be considered the driving factors. These conditions can be remotely sensed using satellites that regularly cover extensive geographical areas. Therefore, remote sensing has been used in recent years for developing malaria early-warning systems, particularly for Africa. For the Greater Mekong Subregion, an epicenter of multi-drug resistant malaria, few studies have examined the dependency of malaria cases on these factors.

In this study, we show that neural network techniques are useful for modeling the dependency of malaria cases on meteorological and environmental parameters. Neural network analysis is a vital part of machine or artificial intelligence, which is a discipline to study a machine's ability for learning and adaptation, and exhibition of intelligent behaviors. It has found broad application involving estimation, classification, and time series analysis, and therefore this methodology overlaps with traditional statistical methods.

Generally speaking, the neural network methods excel in real-life problems that are too complex to be expressed in analytical form. The disadvantage of these methods, as has been demonstrated in various applications, is the long computation required in the training process relative to other statistical methods. In addition, the trained network is not in a form as compact as the regression techniques when using the

training results on application samples or for prediction. This, however, may be the price to pay to accommodate more complex representations. We have yet to extensively compare the performance of neural network methods with respect to the more traditional statistical approaches with the data used in this study. This may be a topic for future study.

The example for predicting (hindcasting) malaria cases in 2001 demonstrated that the use of meteorological and environmental data is only feasible when other contextual determinants do not vary significantly. The success of malaria control effort in Thailand led to a significant reduction of malaria transmission in 2001. Because a neural network learns from examples, and this reduction differs from what it learned from the 1994-2000 seasons, the networks overestimated the malaria transmission in 2001. Likewise, for countries in which malaria control efforts are disrupted, either due to economic crises or to military conflicts, neural networks trained with normal transmission conditions would likely underestimate future malaria transmission. This example underscores the importance of other contextual determinants on malaria transmission.

Thailand has long borders, nearly 3,200 km over land, with Myanmar, Laos, Cambodia, and Malaysia as its neighboring countries. Significant populations from the neighboring countries have come into Thailand or stayed near the border over the last two decades. Along the Thai-Myanmar border, there were approximately 1.3 million registered migrants and 120,000 registered refugees living in camps at the end of 2004 (WHO Thailand, 2006). There were at least a million unregistered migrants in Thailand. The estimated percentage of non-permanent populations are available for some of the border provinces (WHO Thailand, 2006) and are shown in Fig. 1.

Taking the Tak province for example, it is estimated that 42-47% of its population are refugees, migrants, or displaced populations. Because of the large border-crossing population movement, it may not be surprising that Tak is one of the most

malaria-endemic provinces in Thailand. Overall, approximately 70% of all malaria cases in Thailand occur in the 10 border provinces (WHO Thailand, 2006). Due to the limited accessibility of health care, transient and migrant populations expand the human reservoir for malaria transmission. In turn, these populations escalate the endemicity among the native Thai populations. The movement of migrant population is therefore an important contextual determinant that contributes to malaria transmission. In addition, it confounds the complexity for the prediction of malaria transmission intensity based on meteorological and environmental parameters.

In this study, we have used remotely-sensed and on-site data to examine the meteorological and environmental factors in malaria transmission. The malaria cases are at provincial resolution, and the spatial resolution of the meteorological and environmental data are from medium to coarse. Malaria transmission is known to be spatially heterogeneous. In spite of the relative low resolution of the data and the potential impact of migrant populations, it appears that it is still possible to extract a reasonable amount of dependency for malaria transmission on the meteorological and environmental parameters. As the trend for satellite measurements is toward higher spatial resolution, more and more geophysical parameters relevant to malaria transmission are becoming available. By coupling these parameters with malaria data at lower administrative levels (e.g., district or village), it is reasonable to believe that a more precise dependency will be found between the malaria cases and the meteorological and environmental parameters.

Weight decay (Bishop, 1996) may reduce the effective number of synapses and nodes in a fully connected network. This technique was also used in the present study. But the result with weight decays in training has less bearing on the objective of this paper. We will describe weight decay and network pruning in a follow-on paper in which these techniques may improve prediction accuracy.

References

Abeku TA, De Vlas SJ, Borsboom GJJM, Tadege A, Gebreyesus Y, Gebreyohannes H, Alamirew D, Seifu A, Nagelkerke NJD, Habbema JDF, 2004. Effects of meteorological factors on epidemic malaria in Ethiopia: a statistical modelling approach based on theoretical reasoning. Parasitology 128, 585–593.

Albert DP, Gesler WM, Levergood B, 2000. Spatial Analysis, GIS, and Remote Sensing Applications in the Health Sciences, Ann Arbor Press, 217pp.

Bishop CM, 1996. Neural Networks for Pattern Recognition, Oxford University Press, 482pp.

Bouma M, van der Kaay H, 1996. The El Niño Southern Oscillation and the historic malaria epidemics on the Indian subcontinent and Sri Lanka: an early warning system. Trop Med Int Health 1, 86-96.

Coleman RE, Kumpitak C, Ponlawat A, Maneechai N, Phunkitchar V, Rachapaew N, Zollner G, Sattabongkot J, 2004. Infectivity of asymptomatic Plasmodium-infected human populations to *Anopheles dirus* mosquitoes in western Thailand. J Med Entomol 41, 201-8.

Craig MH, Snow RW, le Sueur D, 1999. A climate-based distribution model of malaria transmission in sub-Saharan Africa. Parasitol Today 15, 105-11.

Easterling DR, Peterson TC, Karl TR, 1996. On the development and use of homogenized climate data sets. J Clim 9, 1429-1434.

Flaming GM, 2005. Global precipitation measurement update. In International Geoscience and Remote Sensing Symposium proceedings, Seoul, Korea, 25-29 July 2005, 1, 79-82.

Gagnon A, Smoyer-Tomic K, Bush A, 2002. The El Niño southern oscillation and malaria epidemics in South America. Int J Biometeorol 46, 81-89.

Gardner D, eds., 1993. The neurobiology of neural networks, MIT Press, 227pp.

Githeko AK, Ndegwa W, 2001. Predicting malaria epidemics in the Kenyan highlands using climate data: a tool for decision makers. Global Change & Human Health 2, 54-63.

Grover-Kopec E, Kawano M, Klaver RW, Blumenthal B, Ceccato P, Connor SJ, 2005. An online operational rainfall-monitoring resource for epidemic malaria early warning systems in Africa. Malar J 21(4), 6.

Hay SI, Guerra CA, Snow RW, 2004. Report on Agreement to Perform Work (APW): Determination of populations at malaria risk. Ref. M50/370/19, TALA Research Group, Department of Zoology, University of Oxford, 50pp.

Hay SI, Rogers DJ, Shanks GD, Myers MF, Snow RW, 2001. Malaria early warning in Kenya. Trends Parasitol 17, 95-99.

Hay SI, Snow RW, Rogers DJ, 1998. From Predicting Mosquito Habitat to Malaria Seasons Using Remotely Sensed Data: Practice, Problems and Perspectives. Parasitol Today 14, 306-313.

Hay SI, Tucker CJ, Rogers DJ, Packer MJ, 1996. Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vectors of disease. Ann Trop Med Parasitol 90, 1-19.

Haykin S, 1994. Neural networks – a comprehensive foundation, IEEE Press, 696pp.

Kidson C, Singhasivanon P, Supavej S, eds., 1999. Mekong malaria. Malaria, multi-drug resistance and economic development in the greater Mekong subregion of Southeast Asia. Southeast Asian J Trop Med Public Health 30 Suppl 4, 1-101.

Kleinschmidt I, Bagayoko M, Clarke G, Craig M, Le Sueur D, 2000. A spatial statistical approach to malaria mapping. Int J Epidemiol 29, 355-361.

Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A, 2003. El Niño and health. Lancet 362, 1481-89.

Kummerow C, Barnes W, Kozu T, Shiue J, Simpson J, 1998. The Tropical Rainfall Measuring Mission (TRMM) Sensor Package. Journal of Atmospheric and Oceanic Technology 15, 809-817.

MARA/ARMA, 1998. Towards an Atlas of Malaria Risk in Africa — First Technical Report of the MARA/ARMA Collaboration. Mapping Malaria Risk in Africa/Atlas du Risque de la Malaria en Afrique, 45pp.

MODIS atmosphere web site. http://modis-atmos.gsfc.nasa.gov

Nalim S, Hartono, Sugeng, Bogh C, Bos R, 2002. Rapid assessment of correlations between remotely sensed data and malaria prevalence in the Menoreh Hills area of Central Java, Indonesia. Final report. WHO/SDE/WSH/02.06, WHO, 25pp.

Nelson MM, Illingworth WT, 1990. A practical guide to neural nets, Addison-Wesley, 344pp.

Omumbo JA, Hay SI, Guerra CA, Snow RW, 2004. The relationship between the *Plasmodium falciparum* parasite ratio in childhood and climate estimates of malaria transmission in Kenya. Mal J 17(3), 17.

Peterson TC, Vose RS, 1997. An overview of the Global Historical Climatology Network temperature data base. Bulletin of the American Meteorological Society 78, 2837-2849.

Pethleart A, Prajakwong S, Suwonkerd W, Corthong B, Webber R, Curtis C, 2004. Infectious reservoir of *Plasmodium* infection in Mae Hong Son Province, northwest Thailand. Mal J 3

Poveda G, Rojas W, Quiñones ML, Vélez ID, Mantilla RI, Ruiz D, Zuluaga JS, Rua GL, 2001. Coupling between Annual and ENSO Timescales in the Malaria–Climate Association in Colombia. Environ Health Prospect 109, 489-493.

Rogers DJ, Randolph SE, Snow RW, Hay SI, 2002. Satellite imagery in the study and forecast of malaria. Nature 415, 710-715.

Rumelhart DE, McClelland JL, eds., 1986. Parallel distributed processing: exploring in the Microstructure of cognition, Vol 1. Foundations, MIT Press, 517pp.

Sithiprasasna R, Ugsang D, Honda K, Jones J, Singhasivanon P, 2005. Ikonos-derived malaria transmission risk in northwestern Thailand. Southeast Asian J Trop Med Public Health 36, 14-22.

Small J, Goetz SJ, Hay SI, 2003. Climatic suitability for malaria transmission in Africa, 1911-1995. Proc Natl Acad Sci USA 100, 15341-15345.

Smith EA, Asrar G, Furuhama Y, Ginati A, Adler R, Casse V, Durning J, Entin J, Houser P, Iguchi T, Kakar R, Kaye J, Kojima M, Kummerow C, Levizzani V, Luther M, Mehta A, Morel P, Mugnai A, Nakamura K, Nakazawa T, Neeck S, Oki R, Raju G, Shepherd M, Simpson J, Stocker E,

Testud J, 2006. In Measuring Precipitation from Space: EURAINSAT and the Future (eds. Levizzani V, Bauer P, Turk FJ), Kluwer Publishers (in press).

Snijders F, 1991. Rainfall monitoring based on Meteosat data – a comparison of techniques applied to the western Sahel. Int J Remote Sens 12, 1331-1347.

Teklehaimanot HD, Lipsitch M, Teklehaimanot A, Schwartz J, 2004a. Weather-based prediction of *Plasmodium falciparum* malaria in epidemic-prone regions of Ethiopia I. Patterns of lagged weather effects reflect biological mechanisms. Mal J 12(3), 41.

Teklehaimanot HD, Schwartz J, Teklehaimanot A, Lipsitch M, 2004b. Weather-based prediction of *Plasmodium falciparum* malaria in epidemic-prone regions of Ethiopia II. Weather-based prediction systems perform comparably to early detection systems in identifying times for interventions. Mal J 19(3), 44.

Thai Ministry of Public Health, 2003. Malaria control programme in Thailand.

Thimasarn K, 2003. A Strategic Framework for Rolling Back Malaria in the Mekong Region. WHO RBM Mekong Office, 51pp.

Thomson MC, Connor SJ, 2001. The development of Malaria Early Warning Systems for Africa. Trends Parasitol 17, 438-445.

Thomson MC, Connor SJ, Milligan PJ, Flasse SP, 1996. The ecology of malaria — as seen from Earth-observation satellites. Ann Trop Med Parasitol 90, 243-264.

Thomson MC, Doblas-Reyes FJ, Mason SJ, Hagedorn R, Connor SJ, Phindela T, Morse AP, Palmer TN, 2006. Malaria early warnings based on seasonal climate forecasts

from multi-model ensembles. Nature 439, 576-579.

Tucker CJ, 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8, 127-150.

Vose RS, Schmoyer RL, Steurer PM, Peterson TC, Heim R, Karl TR, Eischeid J, 1992. The Global Historical Climatology Network: long-term monthly temperature, precipitation, sea level pressure, and station pressure data. ORNL/CDIAC-53, NDP-041, Oak Ridge National Laboratory,

WHO, 2001. Malaria Early Warning Systems. WHO/CDS/RBM/2001.32, World Health Organization, 80pp.

WHO, 2004a. Malaria epidemics: forecasting, prevention, early detection and control. From policy to practice. WHO/HTM/MAL/2004.1098, World Health Organization, 52pp.

WHO, 2004b. Using Climate to Predict Infectious Disease Outbreaks: A Review. WHO/SDE/OEH/04.01, World Health Organization, 55pp.

WHO, UNICEF, 2005. World Malaria Report 2005. WHO/HTM/MAL/2005.1102, World Health Organization.

World Health Organization Thailand, 2005. Overview of Thai-Myanmar border health situation 2005, 210pp.

World Health Organization Thailand, 2006. http://w3.whothai.org/EN/Section3/Section39.htm

Zollner GE, Nigro JD, Sattabongkot J, Sithiprasana R, Khuntirat B, Maneechai N, Kankaew P, Masuoka PM, Robert LL, Roberts DW, Thimasarn K, Vaughan JA, Coleman RE, 2006. Longitudinal evaluation of malaria epidemiology in an isolated village in western Thailand I: Study site and human demographics, manuscript in preparation.