

# Surveillance and Control of Malaria Transmission Using Remotely Sensed Meteorological and Environmental Parameters

R. Kiang<sup>a,\*</sup>, F. Adimi<sup>b</sup>, J. Nigro<sup>c</sup>

<sup>a</sup> Code 610.2, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA – richard.kiang@nasa.gov

<sup>b</sup> RS Information Systems, Inc., Lanham, MD 20706, USA – farida.adimi@gsfc.nasa.gov

<sup>c</sup> Science Systems & Applications, Inc., Lanham, MD 20706, USA – jnigro@ltpmail.gsfc.nasa.gov

**Abstract** – Meteorological and environmental parameters important to malaria transmission include temperature, relative humidity, precipitation, and vegetation conditions. These parameters can most conveniently be obtained using remote sensing. Selected provinces and districts in Thailand and Indonesia are used to illustrate how remotely sensed meteorological and environmental parameters may enhance the capabilities for malaria surveillance and control. Hindcastings based on these environmental parameters have shown good agreement to epidemiological records.

**Keywords:** malaria, remote sensing, neural network, Thailand, Indonesia

## 1. INTRODUCTION

Malaria has been with the human race since ancient times. Worldwide, there are approximately 300–500 million cases and at least 1 million deaths in any given year. The advances of biomedical research, and the completion of genomic mappings for *Plasmodium falciparum* and *Anopheles gambiae* give hope for a reduced malaria burden in the future. Before any effective vaccines become available, however, approximately 40% of the world's population is at risk.

In the Malaria Modeling and Surveillance Project, we have been developing techniques to enhance public health's decision capability for malaria risk assessments and controls. The main objectives are: 1) identification of the potential breeding sites for major vector species; 2) implementation of a malaria transmission model to identify the key factors that sustain or intensify malaria transmission; and 3) implementation of a risk algorithm to predict the occurrence of malaria and its transmission intensity.

In the following, we use selected provinces and districts in Thailand and Indonesia to illustrate how remotely sensed meteorological and environmental parameters may enhance public health organizations' capabilities for malaria surveillance and control.

The Mekong River is the tenth longest river in the world. It directly and indirectly influences the lives of hundreds of millions of inhabitants in its basin. The riparian countries – Thailand, Myanmar, Cambodia, Laos, Vietnam, and a small part of China – form the Greater Mekong Subregion (GMS). This geographical region is the world's epicenter of falciparum malaria (Kidson *et al.*, 1999) which is the most severe form of malaria caused by

*Plasmodium falciparum*. Depending on the country, approximately 50 to 90% of all malaria cases are due to this species.

Indonesia is the fourth most populous nation in the world. It has the third highest malaria endemicity in Southeast Asia after Myanmar and India. Approximately 40% of its population lives in malarious regions. The distribution of malaria in Indonesia is highly heterogeneous. On Java and Bali, the two islands where about 70% of the population concentrates, malaria is hypoendemic. But on the Outer Islands, which include the rest of the archipelago, malaria ranges from hypo- to hyperendemic.

## 2. DATA

### 2.1 Environmental Data

The malaria epidemiological data used for this study span from 1994 to 2001 for Thailand, and from 2001 to 2002 for Indonesia. We have used a variety of meteorological and environmental data for modeling.

Environmental parameters important to malaria transmission include temperature, relative humidity, precipitation, and vegetation conditions. The National Aeronautics and Space Administration (NASA) Earth science data sets that have been used for malaria surveillance and risk assessment include AVHRR Pathfinder, TRMM, MODIS, NSIPP, and SIESIP.

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 website). SIESIP is one of the NASA Earth Science Information Partner (ESIP) projects to compile and develop customized Earth science data sets.

From the beginning of 2000, we extracted monthly temperature data from the Moderate Resolution Imaging Spectroradiometer (MODIS) data set (MODIS web site, 2007). 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

\* Corresponding author.

\*\* Supported by NASA Applied Sciences Public Health Applications Program.

instruments on board the Tropical Rainfall Measuring Mission (TRMM) spacecraft (Kummerow et al., 1998; TRMM web site, 2007). 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.

Relative humidity data were extracted from the National Centers for Environmental Prediction's (NCEP) Reanalysis Monthly Means and Other Derived Variables data set. Alternatively, we can compute relative humidity from water vapor, which is one of the geophysical parameters available in the MODIS atmospheric profile product.

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. 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 NDVI data used in this project are from the Advanced Very High Resolution Radiometer (AVHRR) and MODIS data products.

Precipitation, surface temperature, NDVI and relative humidity time series for Tak Province, Thailand are shown in Fig. 1. Tak is one of the provinces most endemic with malaria in Thailand.

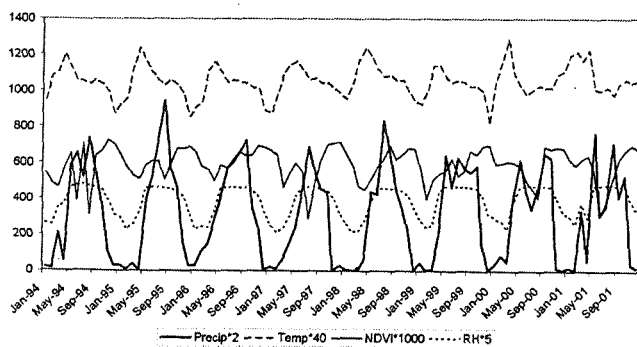


Figure 1. Precipitation (mm/month), temperature (°C), NDVI and relative humidity time series for Tak Province, Thailand.

## 2.2 Epidemiological Data

Thailand 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.

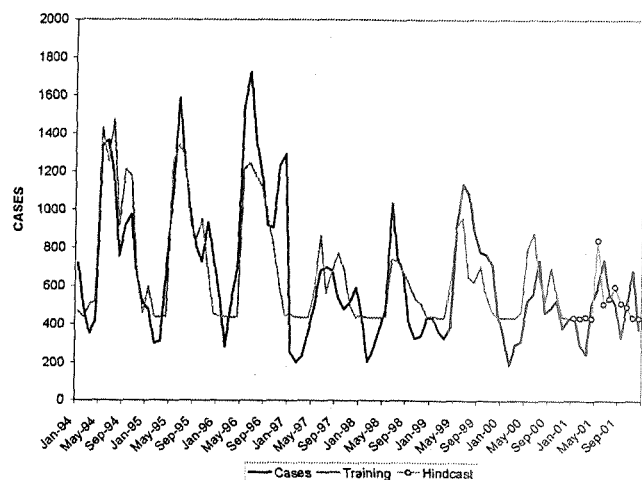


Figure 2. Actual, fitted and hindcast malaria cases for Tak Province, Thailand.

Indonesia has the third highest malaria endemicity in Southeast Asia after Myanmar and India (WHO SEARO website, 2006). The Malaria Subdirector in the Ministry of Health's (MOH) Center for Disease Control and Environment Health has the general responsibility for malaria control. Since 2001, as part of the overall decentralization efforts, the implementation of malaria control has been relegated to the district level.

The malaria control efforts include passive case detection, clinical diagnosis and treatment, and vector control. Only the districts on Java-Bali, where 70% of the total population concentrates, are equipped to provide also active case detection and laboratory diagnosis. Obtaining reliable malaria epidemiological data is a concern.

We have obtained the data from the 7-month Menoreh Hills malaria project (2001-2002) as well as a 24-month malaria time series (2000-2001) used by the project (Indonesian MOH report; Indonesian MOH, 2002). Menoreh Hills is an area in Central Java (Jawa Tengah) with persistent malaria transmission. Geographically, it spans parts of three districts – Purworejo, Kulon Progo, and Magelang. This project was a MOH-WHO Roll Back Malaria (RBM) collaboration with funding provided by USAID. Passive Case Detection (PCD), Mass Blood Survey (MBS), and

Mass Fever Survey (MFS) were used in the project. Because the latter part of the malaria time series may include both PCD and MBS/MFS data, it is difficult to express the time series in case rates. Based on the MBS and MFS results quoted in the report, the approximate endemicity for vivax and falciparum together is 20% for Purworejo, 10% for Kulon Progo, and 10% for Magelang.

In the Menoreh Hills region, normally there are two annual transmission peaks – one during the dry season (June-August), and another during the rainy season (November-January). There are hypotheses that different malaria vectors are responsible for the two transmission peaks. The 24-month time series of malaria cases through MBS are shown in Figure 3. The two transmission peaks may merge if there are meteorological abnormalities.

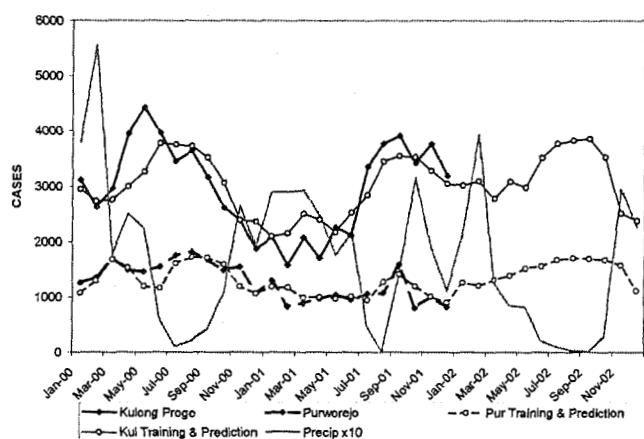


Figure 3. Actual, fitted and hindcast malaria cases for Kulon Progo and Purworejo, Indonesia. Precipitation (in mm/month) is also shown.

### 3. METHODS AND RESULTS

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.

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.

The actual malaria cases, training results and hindcast cases for Tak Province, Thailand are shown in Fig. 2. Reasonably good agreement with the actual malaria cases time series can be seen.

Fig. 3 shows the same set of parameters as well as the precipitation time series for Kulon Progo and Purworejo, the two districts in Central Java with persistent malaria transmission.

Again, we can see the neural network methods can model malaria cases well with remotely sensed meteorological and environmental parameters.

#### 4. CONCLUSIONS

Using selected provinces and districts in Thailand and Indonesia, we have shown that NASA data and results are useful for assessing malaria risks and for epidemic prevention and containment. The potential benefits are: 1) increased warning time for public health organizations to respond to malaria outbreaks; 2) optimized utilization of pesticide and chemoprophylaxis; 3) reduced likelihood of pesticide and drug resistance; and 4) reduced damage to environment. Application of our models, however, is not restricted to Southeast Asia. The model and techniques are equally applicable to other regions of the world, such as Central and South Americas and Africa, when appropriate epidemiological and vector ecological parameters are used as input.

#### 5. REFERENCES

##### 5.1 References from Journals

C. M. Bishop, *Neural Networks for Pattern Recognition*. London: Oxford University Press, 1996.

R. E. Coleman, C. Kumpitak, et al., "Infectivity of asymptomatic Plasmodium-infected human populations to *Anopheles dirus* mosquitoes in western Thailand." *J Med Entomol*, Vol. 41, p.p. 201-208, 2004.

D. Gardner and eds., *The neurobiology of neural networks*, MIT Press, 1993.

C. Kidson, P. Singhasivanon, S. Supavej, Eds., *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, Supplement 4, 1999.

C. Kummerow, W. Barnes, et al., "The Tropical Rainfall Measuring Mission (TRMM) Sensor Package." *Journal of Atmospheric and Oceanic Technology*, Vol. 15, p.p. 809-817, 1998.

M. M. Nelson and W. T. Illingworth, *A practical guide to neural nets*, Addison-Wesley, 1990.

A. Pethleart, S. Prajakwong, et al., "Infectious reservoir of Plasmodium infection in Mae Hong Son Province, north-west Thailand." *Malaria Journal* 3, 2004.

D. E. Rumelhart, J. L. McClelland, et al., Eds. *Parallel distributed processing: exploring in the Microstructure of cognition*, Vol. 1. *Foundations*. Cambridge, MA, MIT Press, 1986.

C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation." *Remote Sensing of Environment*, Vol. 8, p.p. 127-150, 1979.

##### 5.2 References from Websites

MODIS web site, 2007. <http://modis.gsfc.nasa.gov>

SIESIP web site, 2007. <http://climate.geog.udel.edu/~climate>, and <http://www.siesip.gmu.edu>

TRMM web site, 2007. <http://trmm.gsfc.nasa.gov>

WHO SEARO web site, 2007. <http://www.searo.who.int>

##### 5.3 Acknowledgments

This work is supported by the NASA Applied Sciences Public Health Application Program. We appreciate the discussions with Drs. Pratap Singhasivanon, Jeeraphat Sirichaisinthop, Ferdinand Laihad, Krongthong Thimasarn, and Rakesh Rastogi on malaria transmission in Thailand and Indonesia.