TOWARD MALARIA RISK PREDICTION IN AFGHANISTAN USING REMOTE SENSING

N. Safi [a], F. Adimi [b, c], R. P. Soebiyanto [b, d], R. K. Kiang [b]*

[a] National Malaria and Leishmaniasis Control Programme, Afghan Ministry of Public Health, Sanatorium Road, Darulaman, Kabul, Afghanistan – safinajibullah2000@yahoo.com

[b] Global Change Data Center, NASA Goddard Space Flight Center, Code 610.2, Greenbelt, Maryland USA - (farida.adimi, radina.soebiyanto, richard.kiang)@nasa.gov

[c] Wyle Information Systems, McLean, Virginia, USA

[d] Goddard Earth Sciences and Technology Center, University of Maryland Baltimore County, Maryland USA

Commission VIII, WG 2

KEY WORDS: Malaria, risk prediction, Afghanistan, remote sensing

ABSTRACT:

Malaria causes more than one million deaths every year worldwide, with most of the mortality in Sub-Saharan Africa. It is also a significant public health concern in Afghanistan, with approximately 60% of the population, or nearly 14 million people, living in a malaria-endemic area. Malaria transmission has been shown to be dependent on a number of environmental and meteorological variables. For countries in the tropics and the subtropics, rainfall is normally the most important variable, except for regions with high altitude where temperature may also be important. Afghanistan's diverse landscape contributes to the heterogeneous malaria distribution. Understanding the environmental effects on malaria transmission is essential to the effective control of malaria in Afghanistan. Provincial malaria data gathered by Health Posts in 23 provinces during 2004-2007 are used in this study. Remotely sensed geophysical parameters, including precipitation from TRMM, and surface temperature and vegetation index from MODIS are used to derive the empirical relationship between malaria cases and these geophysical parameters. Both neural network methods and regression analyses are used to examine the environmental dependency of malaria transmission. And the trained models are used for predicting future transmission. While neural network methods are intrinsically more adaptive for nonlinear relationship, the regression approach lends itself in providing statistical significance measures. Our results indicate that NDVI is the strongest predictor. This reflects the role of irrigation, instead of precipitation, in Afghanistan for agricultural production. The second strongest prediction is surface temperature. Precipitation is not shown as a significant predictor, contrary to other malarious countries in the tropics or subtropics. With the regression approach, the malaria time series are modelled well, with average R^2 of 0.845. For cumulative 6-month prediction of malaria cases, the average provincial accuracy reaches 91%. The developed predictive and early warning capabilities support the Third Strategic Approach of the WHO EMRO Malaria Control and Elimination Plan.

1. INTRODUCTION

Malaria causes more than one million deaths every year, with most of the mortality in Sub-Saharan Africa. Normally children are the most effected because they lack the partial immunity that will only come later in life through repeated infections. In fact, on average more than a thousand children may die of malaria every day. The spread of multi-drug resistant malaria has also greatly compounded the problem. For example, chloroquine and sulfadoxinepyrimethamine (commonly called SP) are now ineffective for treating falciparum malaria in many parts of the world. Over the past decade, however, the artemisinin-based combination therapy (ACT) has emerged as a new form of treatment that has proven effective in countries particularly hard hit by epidemics of malaria. In spite of its effectiveness, the cost difference between ACT and chloroquine prevents an even broader use.

Malaria is a significant public health concern in Afghanistan. Control efforts were largely effective in the 1970s. During the last three decades with continuous social disruption and military conflicts, the malaria burden has continued to rise and spread to regions where malaria was about to be eliminated. After the establishment of interim government in 2001, National Malaria Control Program was reorganized and new strategies were adopted for malaria control. These strategies include prompt diagnosis and effective treatment, integrated vector management, behavioural change communication, and health system strengthening. A National Strategic Plan was developed for the year 2006-2010. Global Fund has also provided significant support for malaria control in Afghanistan. All these contributed to the significant decline in vivax and falciparum cases between 2002 and 2008.

Although using bednet is not a traditional practice, insecticide-treated bed nets (ITN) have been gradually accepted through public awareness campaigns and mobile, subsidized sales. Furthermore, free distribution of long-lasting insecticide-impregnated nets (LLINs) have been adopted in 2008 for high-risk populations. It is generally recognized that effective use of ACT and LLINs could further reduce the infection rate. Many important efforts and outcomes of the National Malaria and Leishmaniasis Control Program concerning malaria are

* Corresponding author.

documented in the first issue of the Afghanistan Annual Malaria Journal (2009).

Presently, around 60% of the population in Afghanistan is at risk of malaria infection (Asha, May 2003). It is estimated that the highest incidence (23%) is in the Northeast region of Takhar province, followed by Kundoz province (13%). Current malaria strata for all 31 provinces in Afghanistan are shown in Figure 1 (Safi et al., 2009).



Figure 1 Provincial malaria risk strata. Source: Afghan Ministry of public health (Safi et al., 2009)

2. MATERIALS AND METHODS

2.1 Malaria Data

Malaria surveillance data was obtained from Afghan Ministry of Public Health. The surveillance data was collected through village-level, community-based health posts which are part of the primary health care facilities that provide limited services including malaria clinical diagnosis and treatment for uncomplicated case (Islamic Republic of Afghanistan-Ministry of Public Health, 2005). The surveillance was based on passive case detection, and the parasite species was not differentiated. Out the 31 provinces in Afghanistan, we used data from 23 provinces since they have more complete time series. This provincial monthly malaria data spans from 2003 to 2007, though some provinces have shorter period (two years or less). For each province, the last 6 months of data is reserved for prediction, while the remaining is used in calibrating the model – or in other words in estimating the model parameters.

2.2 Environmental Data

This study uses 3 satellite-derived data: precipitation, Land Surface Temperature (LST) and Normalized Difference Vegetation Index (NDVI). Precipitation data was measured from the Tropical Rainfall Measuring Mission (TRMM), which is a joint mission between NASA and the Japan Aerospace Exploration Agency (NASA GSFC, 2010). We retrieved TRMM precipitation with monthly resolution and at 0.25° from the GIOVANNI interface (Acker and Leptoukh, 2007).

Both LST and NDVI were measured by Moderate Resolution Imaging Spectroradiometer (MODIS) aboard Terra and Aqua Spacecraft (NASA, 2009). The MODIS datasets are distributed by the Land Processing Distributed Active Archive Center (LPDAAC) of the U.S. Geological Survey (USGS). The retrieved 8-day LST data was at 1 km spatial resolution with a sinusoidal projection, where NDVI is at a monthly resolution. These two datasets were re-projected to a geographical projection compatible with the other datasets. In addition, the 8-day LST data was subsequently aggregated to monthly so as to synchronize with the malaria data temporal resolution.

For each of the environmental variable, we created the lagged variables. The lags were from the previous 1 month up to 4 months. As we previously mentioned, all data was divided into two: fit and prediction dataset. The fit dataset was used in estimating the parameters, and the prediction dataset is reserved for forecast.

2.3 Modelling Approaches

In this study we used two modelling approaches: Neural Network (NN) and General Linear Model (GLM), specifically poisson regression. NN is capable of capturing the nonlinear relationship. However, interpretation of predictors' contribution is not as obvious as in linear regression models. In the following we will briefly describes the two methods.

2.3.1 Neural Network

This information processing method is inspired by the functioning of the brain. NN typically consists of interconnected nodes arranged in 3 major layers: input, hidden and output. In this study we limited the number of nodes in the input layer into 3 in order to reduce the model complexity. The input nodes are the environmental variables as previously described. The number of nodes in the hidden layer is limited to 2 for the same reason. There is only 1 node in the output layer that represents the level of monthly malaria case. More specifically, we employed feed-forward, back-propagation neural network. The NN performance was measured by the Root Mean Square Error (RMSE) of both the fit and prediction dataset.

2.3.2 Linear Regression

Linear regression is a widely used method to predict the risk of infectious diseases. We also employed stepwise regression method in order to eliminate insignificant environmental variable predictors. In each iteration, any environmental variable that has *p*-value greater than 0.05 is further excluded from the analysis.

As is the case for any infectious disease, the number of malaria cases in any given month depends on the number of cases in the previous months. Thus, the autocorrelation terms for the preceding four months were also used as input in the stepwise regression. In order to account for any possible components not represented in the environmental input parameters, we have also included sinusoidal terms which consist of one factor for the trigonometric functions *SINE* and another for the *COSINE* component. The monthly malaria case for each province can be written as:

$$C(t) = \sum_{\tau=0}^{4} EV(t-\tau) + \sum_{\tau=1}^{4} AR(t-\tau) + S(t) + f(t)$$

Where,

$$EV(t) = T(t) + NDVI(t) + P(t)$$

$$AR(t) = C(t)$$

$$S(t) = \sin(\frac{2\pi t}{T}) + \cos(\frac{2\pi t}{T})$$

$T(t)$ = Average temperature in month t

$NDVI(t)$ = Average NDVI in month t

$P(t)$ = Total precipitation in month t

$C(t)$ = Total cases in month t

f(t) is the general trend accounting for factors other than environmental in nature such as improvement in public health support or population movements. The regression model is thus:

$$C(t) = \sum_j \beta_j X_j(t)$$

Where β_j is the weight for the predictor X_j that was selected using stepwise regression.

Again, the regression model is further evaluated using the RMSE, as well as the correlation coefficient, R^2, for both fit and prediction dataset.

3. RESULTS

Neural Networks were developed for each of the selected 23 provinces. All the environmental input combinations, including their lags, were explored. The predicted malaria cases in general show a good agreement with the data. As an example is Takhar province in which the environmental time series profile is shown in Figure 2A. The NN outputs, fitted and predicted dataset, shows a good agreement with the actual data (Figure 2B). Here, the inputs to the NN are the previous month case, previous month LST and previous month NDVI.

Similarly, linear regression methods can forecast the malaria dynamics within acceptable accuracy. Regression output for Takhar province is shown in Figure 2C. Selected environmental inputs for this model are the current month LST and previous month NDVI. As it is the case with other provinces, our results showed that precipitation is not a significant predictor for malaria. NDVI, on the other hand, seemed to be the stronger indicator for malaria in most provinces. This result implies that malaria risk in Afghanistan is driven by irrigation, not rainfall.

We show in Table 1 the R^2 values for both fitted and predicted datasets from the linear regression. As we can see, the R^2 values for predicted dataset have larger variation compare to the fitted dataset. This could be due to the short time series for some provinces. If we exclude those provinces with time series shorter than 24 months (Kunar, Laghman and Nangarhar), then the remaining half (10 provinces) has $R^2 > 0.5$ and the other half has $R^2 < 0.5$. Previously we have shown that for cumulative 6-month prediction of malaria cases, the average provincial accuracy reaches 91% (Adimi et al., 2010). The developed predictive and early warning capabilities support the Third Strategic Approach of the WHO EMRO Malaria Control and Elimination Plan.

4. CONCLUSIONS

We have shown that using remote sensing for malaria risk prediction is an achievable goal even in a resource-constrained country like Afghanistan. Assuming the epidemiological data gathered by the health posts are reliable, the models can predict 6-month cumulative cases with high accuracy. This capability can undoubtedly help public health organizations for more effective malaria prevention and control effort. Because all the remote sensing data used in this study is free and freely available, and inexpensive personal computers can be used for data processing and modeling, it only takes minimum investment to gain such a capability. It is difficult to foretell when all the conflicts in Afghanistan will end, but it should not stop the public health organizations from further developing and implementing malaria risk prediction systems.

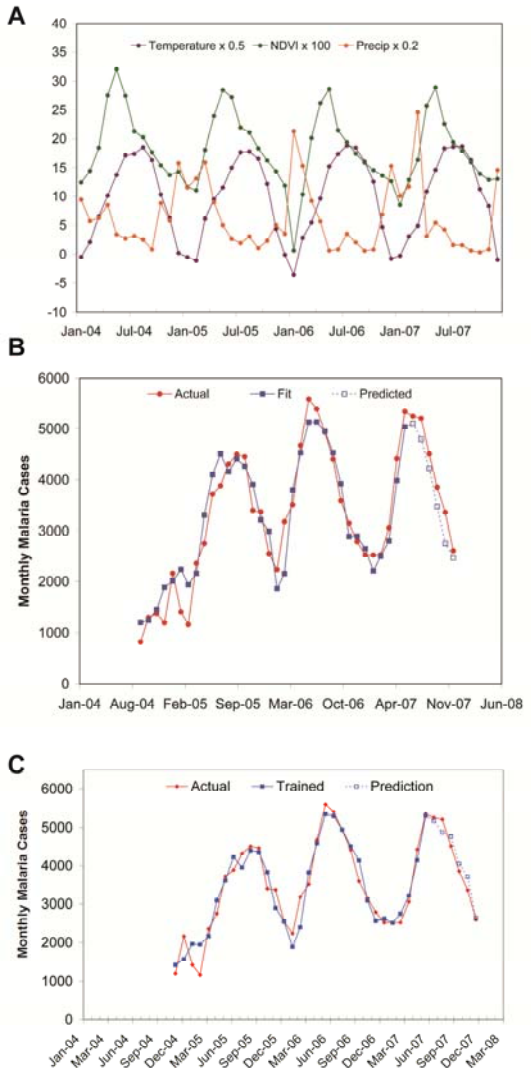


Figure 2 Takhar Province: (A) Environmental profiles, (B) NN output, and (C) linear regression output

Province	Months of Data	Fit R^2	Pred. R^2
Badakhshan	38	0.83	0.74
Baghlan	33	0.73	0.30
Balkh	39	0.71	0.04
Bamyan	36	0.72	0.31
Farah	32	0.73	0.72
Faryab	38	0.79	0.03
Ghazni	34	0.80	0.95
Ghor	33	0.77	0.67
Hirat	41	0.88	0.95
Jawzjan	28	0.64	0.18
Kabul	33	0.86	0.74
Kandahar	41	0.89	0.003
Khost	29	0.85	0.85
Kunar	20	0.87	0.07
Laghman	21	0.64	0.03
Nangarhar	17	0.86	0.02
Paktika	31	0.94	0.17
Paktya	28	0.76	0.74
Samangan	36	0.53	0.01
Sari Pul	27	0.77	0.13
Takhar	37	0.92	0.95
Uruzgan	41	0.07	0.43
Wardak	30	0.42	0.78

Table 1 Regression model performance

5. REFERENCE

Acker, J.G. and Leptoukh, G., 2007. Online analysis enhances use of NASA Earth Science Data. Transactions of the American Geophysical Union, 88(2): 14-17.

Adimi, F., Soebiyanto, R.P., Safi, N. and Kiang, R., 2010. Towards malaria risk prediction in Afghanistan using remote sensing. Malaria Journal, 9:125.

Asha, A.W., May 2003. Rolling back malaria in Afghanistan.

Islamic Republic of Afghanistan-Ministry of Public Health, 2005. A basic package of health services for Afghanistan, 2005/1834.

Islamic Republic of Afghanistan-Ministry of Public Health, 2009. Afghanistan Annual Malaria Journal(1).

NASA, 2009. MODIS. http://modis.gsfc.nasa.gov (accessed January 2009)

NASA GSFC, 2010. TRMM: Tropical Rainfall Measuring Mission. http://trmm.gsfc.nasa.gov/ (accessed June 2010)

Safi, N., Hameed, H., Sediqi, W. and Himmat, E., 2009. NMLCP Annual Report 2008. Afghanistan Annual Malaria Journal(1): 8-14.

6. ACKNOWLEDGEMENTS

This study is supported by NASA Applied Sciences Public Health Program (http://nasascience.nasa.gov/earth-science/applied-sciences).