International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

EPIDEMIA – An EcoHealth Informatics System for Integrated Forecasting of Malaria Epidemics

Michael C. Wimberly[1], Geoffrey M. Henebry[1], Yi Liu[2], and Gabriel B. Senay[3]
[1]*South Dakota State University, Geospatial Sciences Center of Excellence, Brookings, SD USA, michael.wimberly@sdstate.edu*
[2]*South Dakota State University, Department of Electrical Engineering and Computer Science, Brookings SD USA*
[3]*United States Geological Survey, Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD USA*

Abstract: Advance information about the timing and locations of malaria epidemics allows more effective targeting of resources for prevention, control, and treatment. However, these predictions must be accurate to ensure that potential outbreaks are not missed and resources are not wasted responding to predicted outbreaks that do not occur. Early warning systems based on environmental monitoring can identify critical risk factors before an epidemic actually starts, but their accuracy is constrained by the complex interrelationships of climatic variability, mosquito population dynamics, malaria transmission, and the resulting risk of human infection. In contrast, early detection of malaria epidemics based on epidemiological surveillance can be more reliable because it is conditioned on direct observations, but it offers limited lead time and is dependent on timely, accurate surveillance data. Ideally, a malaria forecasting system that integrates elements of both early warning and early detection should be able to leverage the strengths and minimize the limitations of each approach. One reason for this lack of integration is a dearth of suitable tools and techniques. In response, we have developed a conceptual framework for the Epidemic Prognosis Incorporating Disease and Environmental Monitoring for Integrated Assessment (EPIDEMIA) computer system. This framework includes definitions of system users and their associated use cases, identification of critical input data, and a scientific workflow for integrating these spatially and temporally heterogeneous data streams to yield predictions of malaria risk. The system will be tested by implementing it in the Amhara Region of Ethiopia in collaboration with local stakeholders in the NGO and public health sectors. This innovative translational approach to health informatics will enable us to assess the practical effectiveness of the tools and continually upgrade and improve the technologies.

***Keywords**:* Malaria; early warning; early detection; remote sensing; health informatics.

1. INTRODUCTION

Malaria is one of the most common infectious diseases in the world and a major public health problem throughout sub-Saharan Africa. Malaria epidemics occur most frequently in highland and semi-arid regions where marginal environments support unstable transmission. In these areas, malaria outbreaks are often associated with interannual fluctuations in rainfall and temperature. These epidemics can be particularly devastating because they occur in areas where large portions of the population lack immunity to malaria (Abeku, 2007). Therefore, there is strong agreement about the potential value of predicting future epidemic risk and detecting epidemics at an early stage (Thomson and Connor, 2001; DaSilva et al., 2004; Abeku, 2007; Ford et al., 2009). In particular, advance information about the timing and locations of malaria epidemics would facilitate more effective targeting of resources for prevention, control, and treatment. However, these predictions must be accurate to ensure that potential outbreaks are not missed and resources are not wasted responding to predicted outbreaks that do not occur.

Techniques for malaria epidemic forecasting can be grouped into two main categories: early detection and early warning (World Health Organization, 2001; DaSilva et al., 2004). Early detection monitors temporal patterns of malaria indicators, such as clinical or confirmed malaria cases (Figure 1) and identifies anomalies that signify the initial stages of an epidemic (Guintran et al., 2006). In contrast, early warning predicts future risk of an epidemic based on environmental risk factors (World Health Organization, 2001). To develop effective malaria early warning systems, it is necessary to have a strong understanding of the local environmental factors that influence the ecology and epidemiology of malaria. Furthermore, it is essential to have sufficient data on temporal patterns of both malaria occurrence and environmental risk factors to support model development, hypothesis testing, prediction, and accuracy assessment.

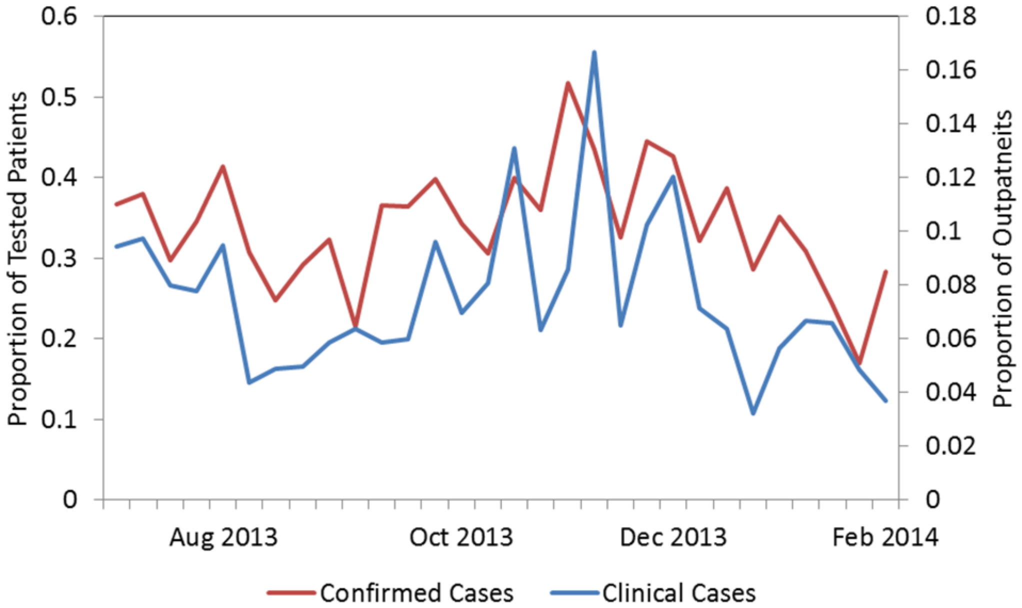

Figure 1: Weekly time series of malaria indicators from the South Achefer district in the Amhara region of Ethiopia. Indicators include the proportion of tested patients with confirmed malaria (red line, left axis) and the proportion of outpatients clinically diagnosed with malaria.

There is an inherent trade-off among disease forecasting approaches in which both accuracy and precision decrease with increasing lead time (DaSilva et al., 2004). Although environmental monitoring can identify critical risk factors before an epidemic actually starts, the accuracy of early warning is constrained by the complex interrelationships of climatic variability, mosquito population dynamics, malaria transmission, and the resulting risk of human infection. In contrast, early detection of malaria epidemics can be more reliable because it is conditioned on actual observations, but offers less lead time for response and is dependent on timely, accurate surveillance data. Ideally, malaria forecasts that combine elements of early detection and early warning should be more accurate than predictions based on either environmental or surveillance data alone. However, to date there have been few attempts to explicitly use multiple sources of information to improve the effectiveness of malaria forecasting.

One reason for this lack of integration is a dearth of established techniques and available health informatics tools for linking malaria surveillance and environmental monitoring. In response to this need, we have developed a conceptual framework for the Epidemic Prognosis Incorporating Disease and Environmental Monitoring for Integrated Assessment (EPIDEMIA) computer system. This system will extend our existing EASTWeb software for processing satellite remote sensing data (Wimberly et al., 2012a) to run as a cloud-based application and incorporate new functionality for uploading and managing surveillance data, making forecasts using advanced modeling and decision support approaches, and providing access to data and forecasts using a low-bandwidth interface. We will test the effectiveness of these tools by implementing them across a network of sentinel locations in the Amhara region of Ethiopia. This innovative translational approach to health informatics will enable us to assess the practical effectiveness of the tools that we develop and to upgrade and improve the technologies throughout the project.

2. SYSTEM DESIGN

The EPIDEMIA system facilitates acquisition, processing, storage, modeling, and dissemination of remotely sensed environmental monitoring data and epidemiological surveillance data. The core of the system is a set of databases that integrate these data for modeling and forecasting malaria risk. The functions for acquiring, processing, and managing remotely-sensed environmental monitoring data will be adapted from the existing EASTWeb software (Wimberly et al., 2012a), which encompasses the *Researcher Interface*, *Remote Sensing Data Acquisition Subsystem, Remote Sensing Data Processing Subsystem* and *Environmental Database* (Figure 2). The *Forecasting Subsystem* is being updated to automate modeling and forecasting steps that are now performed interactively by the researchers. The *Surveillance Data Acquisition Subsystem, Environmental Database,* and *Reporting Subsystem* are being developed to automatically capture, store, and manage disease surveillance data and produce summary maps and reports. We are also developing a new *Public Health Interface* that will allow end users to directly interact with the EPIDEMIA system.

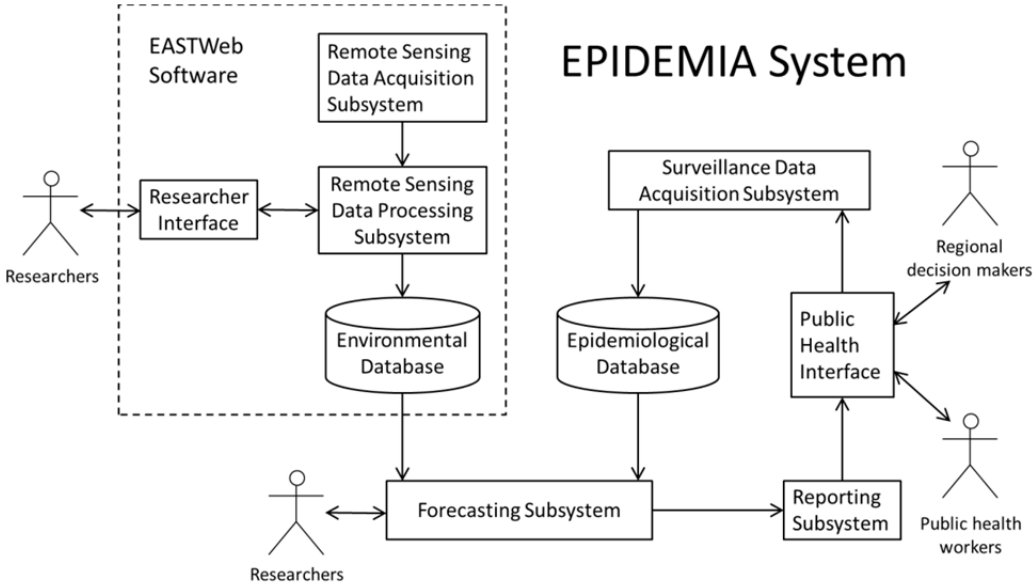

Figure 2: Block diagram depicting the major subsystems and users of the EPIDEMIA system

The main classes of system users and their associated use cases are displayed in Figure 2. Public Health Workers are responsible for uploading disease surveillance data into the system. Regional Decision Makers are the primary users of the malaria forecasting information because they have the primary responsibility for coordinating the regional response to malaria epidemics. The Regional Decision Makers and the Public Health Workers both have the ability to query summary information, including reports and maps of both epidemiological data and environmental indicators. Researchers use the system to query malaria surveillance data and environmental monitoring data for developing, refining, and implementing the malaria risk forecasting models. The Researchers are also responsible for general system control and management.

A general narrative of system function is as follows. Weekly malaria surveillance data are uploaded by Public Health Workers via the *Public Health Interface*. These data include a variety of malaria indicators summarized by political units such as districts and zones (Table 1). The data are checked for consistency and then incorporated into the *Epidemiological Database* via the *Surveillance Data Acquisition Subsystem*. Simultaneously, environmental monitoring data from multiple sources is automatically acquired and processed through the *Remote Sensing Data Acquisition* and *Processing Systems* and stored in the *Environmental Database* (Table 2). The *Forecasting Subsystem* links these two databases to create a unified data stream in which environmental indices and malaria case data summaries are linked based on time and location. The Researchers use this unified data stream to develop and test forecasting models. Model development and testing primarily use the R environment for statistical computing, which has functions for querying the PostgreSQL databases using the RPostgreSQL library.

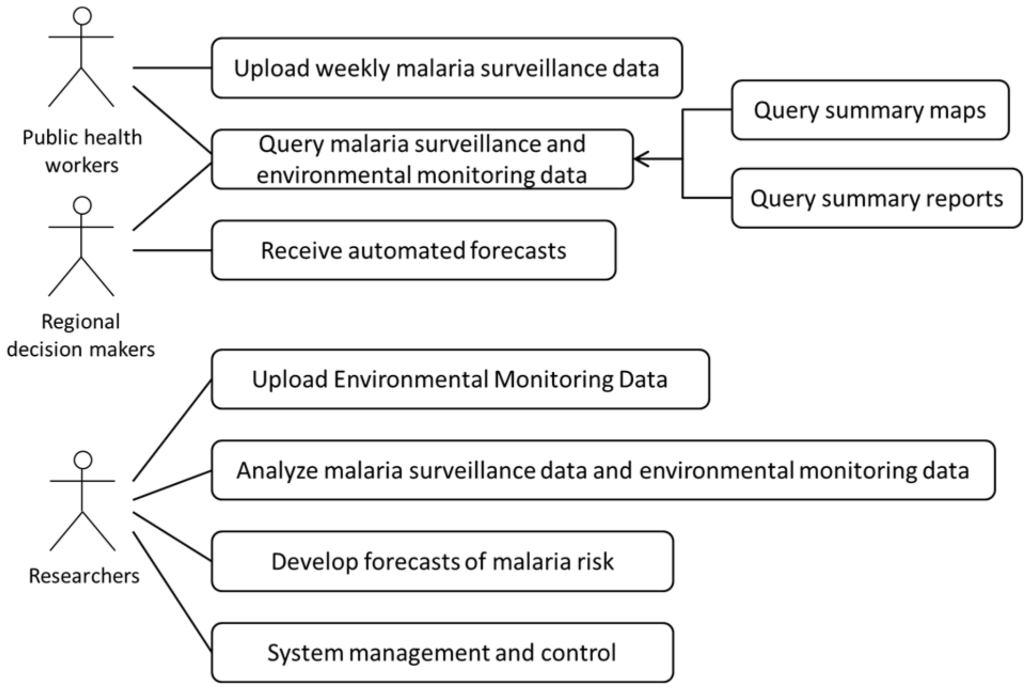

Figure 3: Use case diagram for the EPIDEMIA system

Table 1: Examples of malaria surveillance variables used in the EPIDEMIA system.

Malaria Surveillance Variable	**Notes**
Total number of outpatients	
Total number of malaria cases	Includes both confirmed and unconfirmed cases Stratified by age and sex
Blood films examined for malaria	
Blood films confirmed for malaria	By parasite species
RDTs tested for malaria	
RDTs confirmed for malaria	By parasite species
Hospitalized malaria cases	Stratified by age and sex
Malaria deaths	Stratified by age and sex

Table 2: Examples of remotely sensed environmental data sources used in the EPIDEMIA system

Environmental Data Source	**Spatial Resolution**	**Temporal Resolution**
MODIS Vegetation Indices	1 km	8 d
MODIS Land Surface Temperature	1 km	8 d
TRMM Precipitation	0.25 degree	3 hr
SSEBop Evapotranspiration	1 km	8 d
GLDAS Climate Reanalysis	0.25 degree	1 hr
SMAP Soil Moisture	9 km	2 d (L3) & 7 d (L4)

Models are used to produce both long-term and short-term forecasts. Long-term seasonal forecasting is based on the hypothesis that climatic anomalies throughout the year have cumulative effects of climate on the seasonal development of mosquito populations and the level of parasitemia in the human population. For example, environmental conditions that lead to higher-than-average mosquito populations and malaria prevalence prior to the main rainy season can set the stage for rapid increases in transmission rates as the rains subside (Pascual et al., 2008). In the Amhara region, temperature and precipitation anomalies prior to the monsoon season are associated with interannual variability in malaria risk during the main epidemic season following the monsoon (Wimberly et al., 2012a) and similar associations with seasonal anomalies have been found in other parts of East Africa (Thomson et al., 2005; Pascual et al., 2008) as well as in other regions of the world (Rahman et al., 2010; Rahman et al., 2011). Short-term forecasting models typically include lagged environmental variables as well as autoregressive terms that capture temporal autocorrelation in the pattern of malaria cases. These models typically operate at lags of 1-3 months, similar to the time that it takes for mosquitoes to develop from eggs to adults, acquire a malaria infection, and retransmit the infection to another human (Teklehaimanot et al., 2004a). To implement the short-term forecasts, we use data assimilation methods including the Kalman Filter and related techniques that can automatically handle missing data by substituting predicted values and provide a framework for allowing parameters to vary through time as more recent data is assimilated (Stöckli et al., 2008; Batt and Carpenter, 2012).

Once models have been developed, they are implemented using the *Forecasting Subsystem*. This subsystem includes controller functions that query data from the *Environmental and Epidemiological Databases*, use them to make predictions with the forecasting models in R, and pass the results to the *Reporting Subsystem*. Finally, the *Reporting System* packages the results into a report format and makes them available to the Public Health Workers and Regional Decision Makers via the *Public Health Interface*. Queries for specific types of map and tabular data summaries will be executed via the low-bandwidth *Public Health Interface* and processed by the *Reporting Subsystem*, which will in turn display the requested information for the end users. Java is the main programming language for the EASTWeb application as well as the other EPIDEMIA system components. The GDAL library is used for geoprocessing, R is used for statistical modeling and graphics, and PostgreSQL with the PostGIS spatial database extension is used for data storage and management.

3. DEVELOPMENT AND IMPLEMENTATION

The EPIDEMIA system will be implemented for malaria risk mapping and forecasting in the Amhara region of Ethiopia. We previously analyzed associations between climatic variability and malaria epidemics in this region using a historical malaria surveillance dataset assembled by our collaborators at the Health, Development, and Anti-Malaria Association (HDAMA). The finding of significant spatial synchrony in the temporal patterns of malaria epidemics supported the hypothesis that interannual variability in malaria risk is associated with broad-scale climatic anomalies (Wimberly et al., 2012b). An environmental model based on January-February precipitation anomalies, May-June temperature anomalies, and April-May precipitation anomalies predicted the relative rate of malaria during the peak September-December epidemic season with a cross-validated R^2 of 0.39. Incorporating malaria cases during the preceding May-June as an additional predictor variable increased the cross-validated R^2 to 0.53, highlighting the value of integrated predictions based on lagged environmental and epidemiological variables (Wimberly et al., 2012a). Similarly, a time series analysis of monthly variability in malaria cases found that models including lagged functions of both malaria cases and environmental variables (remotely sensed land surface temperature, precipitation estimates, vegetation indices, and actual evapotranspiration) fit the data better and made more accurate predictions than models based on case data alone (Midekisa et al., 2012).

EPIDEMIA is being developed using a scenario-based participatory design approach in collaboration with public health partners from the HDAMA and the Amhara Regional Health Bureau (Sutcliffe et al., 2010; Reeder and Turner, 2011) . This approach is particularly appropriate for the Amhara region where our end users (Regional Decision Makers and Public Health Workers) have limited experience with web-based decision support tools. In particular, we utilize narrative descriptions of envisioned usage scenarios to educate the end users as to the possibilities of what can be done with the system while allowing the researchers to gain feedback about user requirements and potential barriers to implementation. At the initial design workshop to be held in July 2014, researchers will present an initial vision for the system, conduct a requirements analysis to produce a set of problem scenarios, and formulate an initial design proposal.

Subsequent workshops will involve training on how to use the system and will continue the iterative process of scenario-based design by conducting usability analysis of the system, refining usage scenarios based on new or changed user requirements, and updating the design proposal as necessary. Usability testing includes objective tests of the time required for users to perform specific tasks (such as entering and uploading new surveillance data) and subjective assessment of user experiences based on a survey with responses ranked on a Likert scale. Evaluation instruments are derived from existing usability surveys (Tullis and Stetson, 2004), and are modified to be more specific to the objectives of the EPIDEMIA system. Another objective of these workshops is to identify specific actionable decision steps that can be linked to various malaria epidemic alerts produced by the EPIDEMIA system over a range of different lead times.

4. CONCLUSIONS AND RECOMMENDATIONS

To our knowledge, EPIDEMIA is the first major effort to facilitate integrated malaria early detection, early warning, and decision support through the development of novel health informatics tools. Previous work has treated these approaches as alternatives and has focused on comparing their effectiveness (Hay et al., 2003; Teklehaimanot et al., 2004b; Cox and Abeku, 2007) rather than combining them in a unified system. Accomplishing this integration requires innovative translational research to develop and implement new concepts, methodologies, and tools. In public health informatics, information systems have been designed to support syndromic surveillance by collecting and processing health indicators in near-real time and using this information to generate early alerts of potential disease outbreaks or bioterrorism threats (May et al., 2009; Randrianasolo et al., 2010). In the environmental sciences, the field of ecological informatics has emerged to in response to the large volumes of data collected through ecological monitoring networks and the consequent need to effectively manage and utilize this information (Hale and Hollister, 2009; Michener and Jones, 2012). Understanding the ecological determinants of human health has become increasingly important in light of recent trends in human population growth, climate change, and land use change (Moore et al., 2003; Norris, 2004; Patz et al., 2005). EPIDEMIA provides a framework for coalescing these disparate fields to facilitate a truly interdisciplinary approach to malaria forecasting and decision support that will ultimately help to support disease prevention, control, and elimination in Ethiopia and in other vulnerable areas across the globe.

5. ACKNOWLEDGMENTS

We acknowledge Abere Mihretie, Estifanos Bayabil, Paulos Semuniguse, Hiwot Teka, and other current and former employees of the Health, Development, and Anti-Malaria Association who have worked with us to build the collaborative foundation for the EPIDEMIA system. This work is supported by Grant Number R01AI079411 from the National Institute of Allergy and Infectious Diseases.

6. REFERENCES

Abeku, T. A. 2007. Response to malaria epidemics in Africa. Emerg. Infect. Dis. 13 (5):681-686.

Batt, R. D., and Carpenter, S. R. 2012. Free-water lake metabolism: addressing noisy time series with a Kalman filter. Limnol. Oceanogr. Methods 10:20-30.

Cox, J., and Abeku, T. A. 2007. Early warning systems for malaria in Africa: from blueprint to practice. Trends Paratisol. 23 (6):243-246.

DaSilva, J., Garanganga, B., Teveredzi, V., Marx, S. M., Mason, S. J., and Connor, S. J. 2004. Improving epidemic malaria planning, preparedness and response in Southern Africa. Malaria J. 3:37.

Ford, T. E., Colwell, R. R., Rose, J. B., Morse, S. S., Rogers, D. J., and Yates, T. L. 2009. Using Satellite Images of Environmental Changes to Predict Infectious Disease Outbreaks. Emerg. Infect. Dis. 15 (9):1341-1346.

Guintran, J., Delacollette, C., and Trigg, P. 2006. Systems for the early detection of malaria epidemics in Africa: An analysis of current practices and future priorities. WHO/HTM/MAL/2006.1115, World Health Organization, Geneva.

Hale, S. S., and Hollister, J. W. 2009. Beyond data management: how ecoinformatics can benefit environmental monitoring programs. Environ. Monit. Assess. 150 (1-4):227-235.

Hay, S. I., Were, E. C., Renshaw, M., Noor, A. M., Ochola, S. A., Olusanmi, L., Alipui, N., and Snow, R. W. 2003. Forecasting, warning, and detection of malaria epidemics: a case study. Lancet 361 (9370):1705-1706.

May, L., Chretien, J.-P., and Pavlin, J. 2009. Beyond traditional surveillance: applying syndromic surveillance to developing settings - opportunities and challenges. BMC Public Health 9 (1):242.

Michener, W. K., and Jones, M. B. 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends Ecol. Evol. 27 (2):85-93.

Midekisa, A., Senay, G., Henebry, G. M., Semuniguse, P., and Wimberly, M. C. 2012. Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. Malaria J. 11:165.

Moore, M., Gould, P., and Keary, B. S. 2003. Global urbanization and impact on health. Int. J. Hyg. Envir. Heal. 206 (4-5):269-278.

Norris, D. 2004. Mosquito-borne Diseases as a Consequence of Land Use Change. EcoHealth 1 (1):19-24.

Pascual, M., Cazelles, B., Bouma, M. J., Chaves, L. F., and Koelle, K. 2008. Shifting patterns: malaria dynamics and rainfall variability in an African highland. P. Roy. Soc. B 275 (1631):123-132.

Patz, J. A., Campbell-Lendrum, D., Holloway, T., and Foley, J. A. 2005. Impact of regional climate change on human health. Nature 438 (7066):310-317.

Rahman, A., Kogan, F., Roytman, L., Goldberg, M., and Guo, W. 2011. Modelling and prediction of malaria vector distribution in Bangladesh from remote-sensing data. Int. J. Remote Sens. 32 (5):1233-1251.

Rahman, A., Krakauer, N., Roytman, L., Goldberg, M., and Kogan, F. 2010. Application of Advanced Very High Resolution Radiometer (AVHRR)-based Vegetation Health Indices for Estimation of Malaria Cases. Am. J. Trop. Med. Hyg. 82 (6):1004-1009.

Randrianasolo, L., Raoelina, Y., Ratsitorahina, M., Ravolomanana, L., Andriamandimby, S., Heraud, J.-M., Rakotomanana, F., Ramanjato, R., Randrianarivo-Solofoniaina, A., and Richard, V. 2010. Sentinel surveillance system for early outbreak detection in Madagascar. BMC Public Health 10 (1):31.

Reeder, B., and Turner, A. M. 2011. Scenario-based design: A method for connecting information system design with public health operations and emergency management. Journal of Biomedical Informatics 44 (6):978-988.

Stöckli, R., Rutishauser, T., Dragoni, D., O'keefe, J., Thornton, P., Jolly, M., Lu, L., and Denning, A. 2008. Remote sensing data assimilation for a prognostic phenology model. J. Geophys. Res. 113:G04021.

Sutcliffe, A., Thew, S., De Bruijn, O., Buchan, I., Jarvis, P., McNaught, J., and Procter, R. 2010. User engagement by user-centred design in e-Health. Philos. T. R. Soc. A. 368 (1926):4209-4224.

Teklehaimanot, H. D., Lipsitch, M., Teklehaimanot, A., and 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. Malaria J. 3:41.

Teklehaimanot, H. D., Schwartz, J., Teklehaimanot, A., and 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. Malaria J. 3:44.

Thomson, M. C., and Connor, S. J. 2001. The development of malaria early warning systems for Africa. Trends Paratisol. 17 (9):438-445.

Thomson, M. C., Mason, S. J., Phindela, T., and Connor, S. J. 2005. Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana. Am. J. Trop. Med. Hyg. 73 (1):214-221.

Tullis, T., and Stetson, J. 2004. A comparison of questionnaires for assessing website usability.*in* Usability Professionals Association (UPA) 2004 Conference, Minneapolis, MN, June 7-11.

Wimberly, M. C., Chuang, T.-W., Henebry, G. M., Liu, Y., Midekisa, A., Semuniguse, P., and Senay, G. 2012a. A computer system for forecasting malaria epidemic risk using remotely-sensed environmental data.*in* Proceedings: 6th International Congress on Environmental Modelling and Software. International Environmental Modelling and Software Society (iEMSs), Leipzig, Germany.

Wimberly, M. C., Midekisa, A., Semuniguse, P., Teka, H., Henebry, G. M., Chuang, T. W., and Senay, G. B. 2012b. Spatial synchrony of malaria outbreaks in a highland region of Ethiopia. Trop. Med. Intl. Health 17 (10):1192-1201.

World Health Organization. 2001. Malaria Early Warning Systems - Concepts, Indicators, and Partners - A Framework For Field Research in Africa. WHO/CDS/RBM/2001.32, World Health Organization, Roll Back Malaria Cabinet Project, Geneva.