



Jul 1st, 12:00 AM

A Computer System for Forecasting Malaria Epidemic Risk Using Remotely- Sensed Environmental Data

Michael C. Wimberly

Ting-Wu Chuang

Geoffrey M. Henebry

Yi Liu

Alemayehu Midekisa

See next page for additional authors

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Wimberly, Michael C.; Chuang, Ting-Wu; Henebry, Geoffrey M.; Liu, Yi; Midekisa, Alemayehu; Semuniguse, Paulos; and Senay, Gabriel, "A Computer System for Forecasting Malaria Epidemic Risk Using Remotely- Sensed Environmental Data" (2012).
International Congress on Environmental Modelling and Software. 36.
<https://scholarsarchive.byu.edu/iemssconference/2012/Stream-B/36>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Presenter/Author Information

Michael C. Wimberly, Ting-Wu Chuang, Geoffrey M. Henebry, Yi Liu, Alemayehu Midekisa, Paulos Semuniguse, and Gabriel Senay

A Computer System for Forecasting Malaria Epidemic Risk Using Remotely-Sensed Environmental Data

Michael C. Wimberly¹, Ting-Wu Chuang¹, Geoffrey M. Henebry¹, Yi Liu², Alemayehu Midekisa¹, Paulos Semuniguse³, Gabriel Senay⁴

¹South Dakota State University, Geographic Information Science Center of Excellence, Brookings, SD USA

²South Dakota State University, Department of Electrical Engineering and Computer Science, Brookings SD USA

³Health, Development, and Anti-Malaria Association, Addis Ababa, Ethiopia

⁴United States Geological Survey, Earth Resource Observation and Science (EROS) Center, Sioux Falls, SD USA

Abstract: Epidemic malaria is a major public health problem in the highlands of East Africa. Identifying the climatic triggers that increase malaria risk affords a basis for developing environmentally-driven early warning systems. Satellite remote sensing provides a wide range of environmental metrics that are sensitive to temperature, rainfall, and other climatic variables. The goals of this study were to develop a computer application for automatically acquiring and processing remote sensing data, and to test the utility of these data for modelling and forecasting malaria epidemics in the Amhara region of Ethiopia. The application was programmed using JAVA for user interface development and overall system control. Spatial analyses were carried out using Python scripts to call ArcGIS geoprocessing functions, and PostgreSQL was used to store and manipulate the resulting data summaries. Remotely-sensed variables included land surface temperature from MODIS/Terra, vegetation indices computed using MODIS nadir BRDF-adjusted reflectance, precipitation estimates from the Tropical Rainfall Measuring Mission, and actual evapotranspiration modelled using the simplified surface energy balance method. Historical remote sensing data from 2000-2010 were summarized at the district level by 8-day MODIS composite periods and transformed to deviations from their 11-year means. Time series of monthly malaria outpatient cases were collected for 19 districts in the Amhara region and used to compute risk indices for the main epidemic season from September-December. Malaria epidemics during this season were associated with a higher-than-normal number of malaria cases in May-June, higher-than normal rainfall in January-May, and warmer-than-normal temperatures in May-June. A cross-validated statistical model containing these variables predicted more than 50% of the variability in malaria relative risk. Continued environmental monitoring using satellite remote sensing will allow us to forecast the environmental risk of malaria epidemics in future years and validate these initial results.

Keywords: Malaria; early warning; remote sensing; geoinformatics.

1 INTRODUCTION

Malaria is a globally important disease that imposes an immense public health burden throughout the developing world, particularly in sub-Saharan Africa. Marginal environments that support unstable transmission, such as highland and semi-arid regions, pose a particular challenge for malaria control and prevention [Abeku, 2007]. Although malaria transmission in these areas is low and often seasonal, interannual fluctuations in rainfall and temperature can increase transmission rates and trigger devastating epidemics. Because of the low immunity

of populations living in areas of unstable transmission, malaria epidemics tend to cause severe morbidity and mortality in people of all ages. Knowing the timing and locations of malaria epidemics in advance would allow for more efficient targeting of resources for malaria prevention, control, and treatment [Guintran et al. 2006]. In response to this need, several approaches to malaria epidemic forecasting have been proposed, including early detection of epidemics using malaria surveillance data and early warning systems that use environmental risk factors to predict epidemics based on seasonal forecasts and environmental monitoring data [DaSilva et al., 2004; Thomson and Connor, 2001].

Satellite remote sensing has been identified as a key source of environmental monitoring data on climatic variability that can support early warning efforts by providing consistent, repeatable measurements across nearly the entire surface of the Earth [Ford et al. 2009]. Although there have been calls to expand the use of remote sensing in the health sciences, there are still relatively few examples of operational systems with demonstrated impacts on public health. One of the reasons for this gap is that many health scientists and public health professionals lack the time, technical skills, and computational tools required to acquire, process, and analyse satellite remote sensing data in a way that is relevant to their needs. In response, the objective of our research is to narrow the divide between remote sensing and public health by developing tools to provide health scientists and practitioners with remotely-sensed environmental data to support early warning of malaria and other mosquito-borne diseases.

To achieve this goal, we have developed a computer system for automatically acquiring and processing satellite data to create environmental indices that can be used for ecological forecasting of mosquito-borne disease risk. This paper provides an outline of the overall system design and presents the results of a case study where the system was used to model malaria epidemic risk in the Amhara region of Ethiopia. Our results emphasize the potential for using remote sensing data to monitor environmental risk factors for malaria epidemics, but also highlight several important challenges in system design and implementation.

2 SYSTEM DESIGN

The Epidemiological Applications of Spatial Technologies (EASTWeb) system is designed to provide GIS analysts and spatial epidemiologists with automated access to remote sensing data from online archives (Figure 1). The core of the system is the EASTWeb software application, which is programmed using JAVA for user interface development and overall system control. Spatial analyses are carried out using Python scripts to call ArcGIS 10 geoprocessing functions. A PostgreSQL database is used to store and manipulate the resulting data summaries. Two versions of the application have been produced: one as a stand-alone application that runs on a single computer, and the other as a distributed application that runs on a server and can be accessed by multiple users. Additional components of the system include R scripts for data analysis, modelling, and map generation, along with a software framework for generating web atlas applications to disseminate and visualize the resulting geovisualization products.

2.1 Remote Sensing Data

The system was designed to automatically acquire data from the following sources: (1) The daily Tropical Rainfall Measuring Mission (TRMM) product (3B42) from the Goddard Earth Sciences Data and Information Services Center (GES DISC), (2) the MODIS Terra land surface temperature (LST) and emissivity product (MOD11A2) from the Land Processes Distributed Active Archive Center (LP DAAC), (3) the MODIS nadir BRDF-adjusted reflectance product (MCD43B4) from the LP DAAC, and (4) the potential evapotranspiration product (ETo) from the Famine Early Warning Systems Network (FEWS NET) data portal.

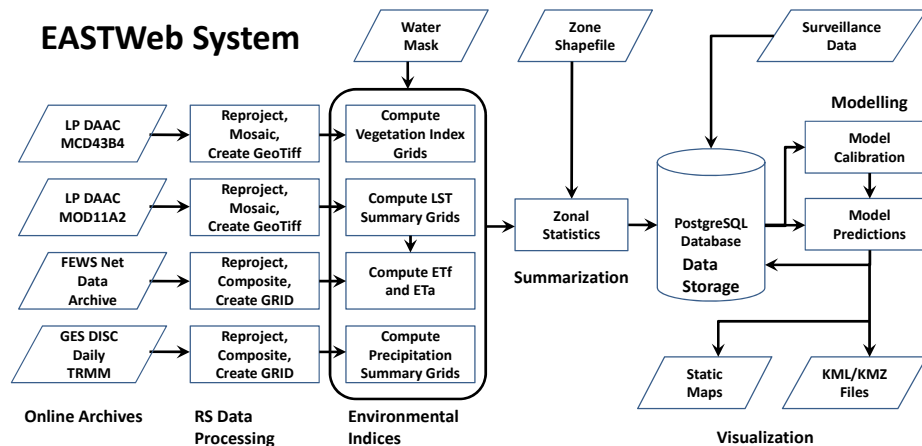


Figure 1: Flowchart outlining the main components of the Epidemiological Applications of Spatial Technologies (EASTWeb) system. The system integrates a variety of software tools for acquiring, processing, and managing remote sensing data for spatial modelling and visualization of mosquito-borne disease risk.

2.2 Data Processing and Management

The EASTWeb software tool is controlled via a graphical user interface (Figure 2). The user first defines a project by selecting a geographic area and a time period for analysis, choosing the remote sensing variables of interest, and providing GIS layers for generating zonal summaries of the environmental indices. The software then connects to the online archives and automatically acquires, reprojects, mosaics, and masks the remote sensing data. After historical data have been acquired, the software continues to automatically collect and process new data as it becomes available.

The raw remote sensing data are used to automatically calculate a variety of environmental metrics. Reflectance data, for example, is used to compute vegetation indices such as the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI) as well as other types of spectral indices such as the Normalized Difference Water Index (NDWI). Day and night LST retrievals are summarized as a daily mean. The system also incorporates a model for computing actual evapotranspiration (ETa) using the simplified surface energy balance method [Senay et al., 2007; 2011].

To maintain consistency across the different data sources and facilitate linking with epidemiological datasets, the environmental metrics are standardized to a common spatial and temporal resolution. Temporal summaries are calculated using standard 8-day MODIS composite periods. Spatial summaries are generated by computing zonal statistics based on user supplied datasets. Zones typically represent counties, census tracts, postal codes, or other areal units for which epidemiological data summaries are available. The summaries are automatically stored in the PostgreSQL database, which can then be queried to access the environmental metrics and link them to epidemiological datasets for analysis and forecasting.

2.3 Analysis and Modelling

A set of scripts and functions in the R environment (R Core Development Team 2012) are used to extract data from the PostgreSQL database and conduct

statistical analyses. These scripts implement a variety of techniques for exploratory data analysis, calibration of statistical models of malaria risk based on historical epidemiological and environmental data, and application of the models to generate early warning predictions of malaria epidemics based on current environmental conditions. Predictions are generated in the form of malaria risk maps, which are output as static maps in image formats such as portable network graphics (PNG) or dynamic maps implemented using keyhole markup language (KML).

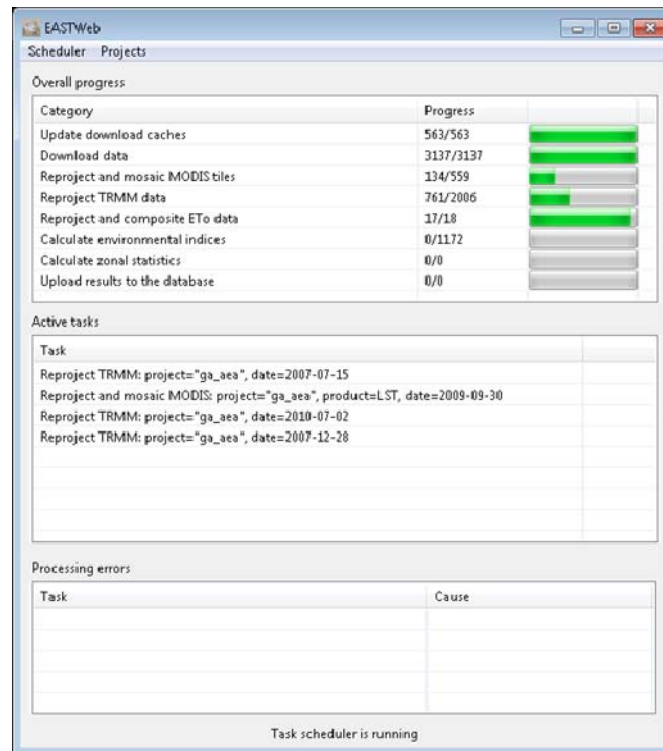


Figure 2: Graphical user interface of the EASTWeb software application. The main interface page tracks the progress of multiple data processing tasks.

2.4 Dissemination and Visualization of Results

We developed the Framework of Web Atlas (FWA), a software framework to support the organization, management, and dissemination of large libraries of geospatial data [Rop et al. 2011]. An XML schema is used to manage the FWA document content and support the use of semantic search. Web atlas content is updated through a producer's interface that uses both AJAX and PHP scripting to dynamically generate new web pages. The FWA supports viewing and downloading of static maps in a variety of formats (e.g., PNG, PDF) and also facilitates dynamic visualization of KML files via a custom Web GIS interface implemented using the Google Earth plugin.

3 CASE STUDY – THE AMHARA REGION OF ETHIOPIA

3.1 Methods

We carried out a case study in which we applied the EASTWeb system to identify environmental risk factors and develop early warning models of malaria risk for the Amhara region of Ethiopia. The Amhara region is home to more than 17 million people and encompasses 157,000 km² with elevations ranging from 506 m to

4533 m. A large percentage of the population (87.4%) lives in rural areas and practices subsistence agriculture. The region is characterized by a monsoon climate with a dry season extending from December through March, and a rainy season that typically begins in April and increases to a peak from June through August. This work was conducted in collaboration with the Amhara Regional Health Bureau and the Health, Development, and Anti-Malaria Association, an Ethiopian non-governmental organization based in Addis Ababa.

A database of remotely-sensed environmental indices over the time period from 2000-2011 was generated using the techniques described previously. Our analysis considered total precipitation, LST (average of day and night retrievals), NDVI, and ETa. These indices were summarized spatially as zonal means at the *woreda* (district) level, summarized temporally by computing running means over six 8-day MODIS composite periods (48 days), and transformed to percent deviations from their 11-year means. This transformation factored out underlying spatial variability in the environmental variables, resulting in temporal deviations that were more sensitive to climatic anomalies.

The Anti-Malaria association created a historical surveillance database for the Amhara region that included 19 districts with long-term data on malaria cases between 2001 and 2009. These data were used to compute logarithm of relative risk (LRR) indices that quantified interannual variability in malaria cases while controlling for seasonal trends and spatial variability in the population at risk. LRR_p quantified interannual variability during the peak season (September-December) when malaria case numbers are highest and epidemics typically occur, and LRR_e quantified interannual variability during the early season (May-June) before the onset of the long rains. We conducted statistical analyses to explore the influences of LRR_e and remotely sensed environmental variables measured earlier in the year (January-August) on LRR_p . We first screened the environmental metrics using Spearman's rank correlation coefficient to identify the times of the year when each metric exhibited the strongest correlation of LRR_p , and then used these seasonal metrics to develop multivariate linear models.

3.2 Results

Malaria relative risk in May-June (LRR_e) was correlated with malaria relative risk during the peak September-December epidemic season (LRR_p) (Spearman's rank correlation $\rho=0.60$, $P < 0.0001$). January- March TRMM rainfall deviations were positively correlated with LRR_p . May-July, LST deviations were positively correlated with LRR_p and May-July NDVI, TRMM rainfall, and ETa deviations were negatively correlated with LRR_p . September-October TRMM rainfall deviations were weakly positively correlated with LRR_p and September-October ETa deviations were negatively correlated with LRR_p .

The best-fitting multivariate model included January-February TRMM deviations, May-July LST deviations, and LRR_e as independent variables and had a cross-validated R^2 of 0.53 (Figure 3A). The best fitting model containing only remotely-sensed environmental variables included January-February TRMM deviations, May-July LST deviations, and April-May TRMM deviations as independent variables and had a cross-validated R^2 of 0.39 (Figure 3B).

The model containing only remotely-sensed environmental variables was used to generate predictive maps of malaria relative risk from 2000-2011 (Figure 4). Large portions of the region were classified as high risk during the major malaria epidemic year in 2003. In contrast most of the region was classified as a mixture of low and moderate risk categories in 2010 and 2011.

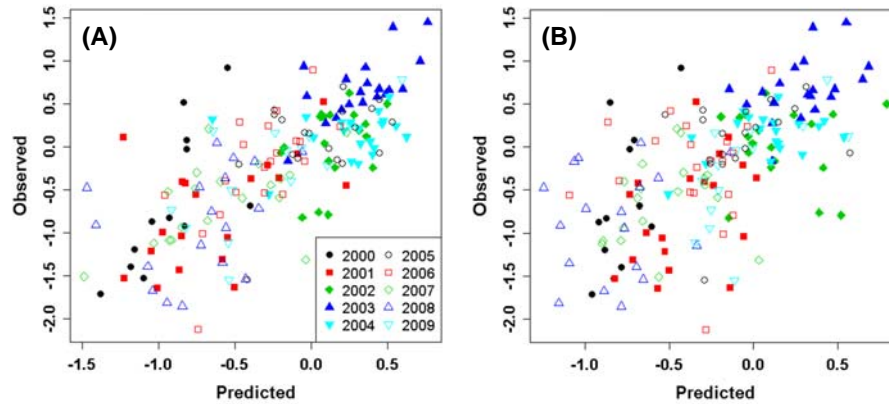


Figure 3: Cross-validation of a model of relative risk of malaria epidemics during September-December (LRR_p). (A) Predictions based on remotely sensed environmental variables and malaria relative risk during May-June (LRR_e). (B) Predictions based on remotely sensed environmental variables alone.

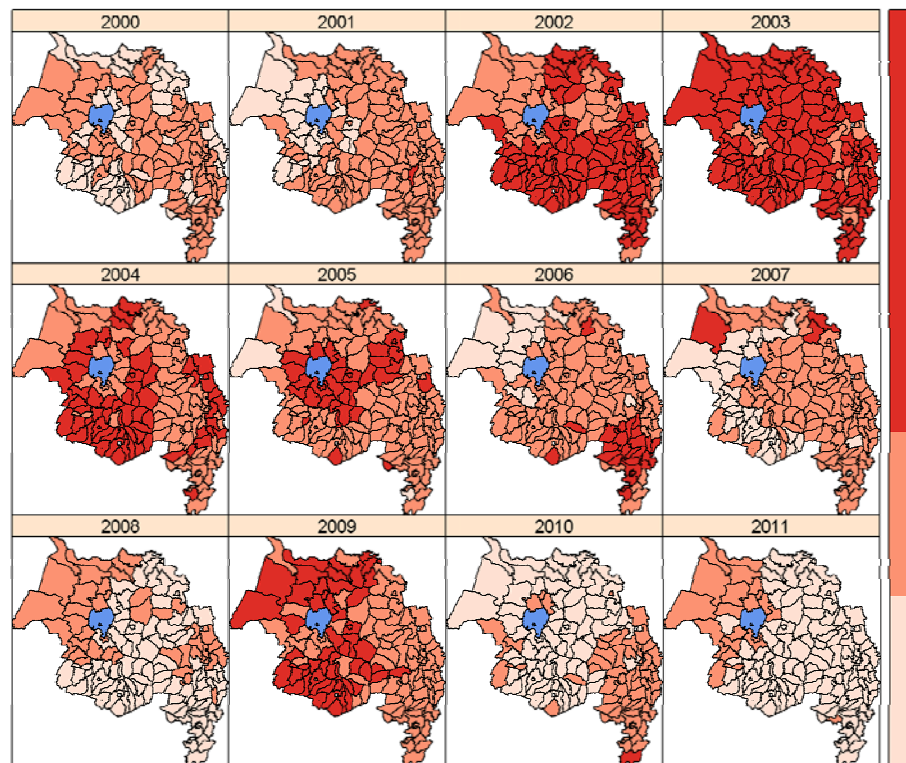


Figure 4: Categories of malaria relative risk (darker red = higher risk) for September-December based on a model with January-February precipitation deviations, May-July land surface temperature deviations, and April-May precipitation deviations as independent variables. Three categories were based on quartiles of predicted relative risk (low=0-25%, medium=25-75%, high=75-100%).

4 DISCUSSION

The automated nature of the EASTWeb system will allow us to continue providing maps of environmental risk factors and modelled risk categories (e.g., Figure 4) in future years. Public health officials can use these products, in combination with

other information sources, to project the risk of a malaria epidemic occurring during the upcoming peak season from September-December. In particular, we recommend that reports of malaria case numbers in the early season (May-June) be examined in conjunction with environmental risk factors. In situations where the early season case numbers and environmental risk factors are both high, this information can be interpreted as high risk for an epidemic in September-December. If one of these two indicators (spring case numbers or environmental risk factors) is high, but the other is low, this information can be interpreted as moderate risk for an epidemic in September-December. For example, even if environmental risk factors are high, the number of malaria cases may still remain low because of malaria interventions or other environmental phenomena not captured in the model. If both the early-season case numbers and environmental risk factors are low, then there is stronger evidence for the low risk of an epidemic in the upcoming peak season.

The environmental risk factors that we have identified are interpretable from an epidemiological and ecological standpoint, and have been locally calibrated using actual data on malaria case numbers. However, several caveats and limitations should be taken into account when using this information to support public health efforts. First, our results are based on the analysis of relative risk that characterizes temporal variability in malaria case numbers relative to their long-term baseline. Additional work is needed to further characterize and map the underlying geographic patterns of baseline malaria risk. Other factors not included in our analyses, including regional migrations of human populations, malaria prevention and control efforts, land cover/land use, soils, and hydrology, will also influence the spatial and temporal patterns of malaria risk. Our analysis was conducted at a district level because the epidemiological data that we used were available at this resolution; thus, we were not able to consider finer-scale heterogeneity in environmental drivers. Finally, although we were fortunate to have nine years of historical malaria surveillance data, this dataset only encompassed one major malaria epidemic period (2002-2004). We are confident that we have identified the major climatic anomalies associated with these outbreaks, but it is possible that different sets of climatic factors could trigger future epidemics, and it is also possible that the environmental factors that we identified will not trigger future epidemics if effective malaria prevention activities are occurring.

Although the development of software systems like EASTWeb can facilitate the application remote sensing data for malaria early warning, the limited availability of malaria surveillance data still limits the development of effective forecasting models in many areas. Expanding the EASTWeb system to include tools for easily capturing and managing public health data in addition to remote sensing data could thus greatly expand the utility of the system. For example, additional modules could be added to allow regular upload of surveillance data via mobile phone or email messages. Timely acquisition of surveillance data could enhance epidemic forecasting as described previously, and consistent archiving and data management would facilitate the development of long-term databases to enhance model calibration and validation. Limited bandwidth is another challenge that constrains the ability of our Ethiopian partners to access large remote sensing databases or use the high-bandwidth tools needed for dynamic geovisualization. The EASTWeb system addresses this problem by automatically processing terabytes of remote sensing imagery and reducing this large data volume to a much smaller amount of information that is specifically targeted for malaria early warning. To facilitate access to this information, we have also developed a low-bandwidth version of our web atlas to facilitate data access.

5 CONCLUSIONS

The development of the integrated EASTWeb system has facilitated the application of satellite remote sensing data malaria early warning in the Amhara region of Ethiopia. The EASTWeb software application expedited the acquisition,

processing, and management of historical remote sensing data and has continued to incorporate new data as it becomes available. The resulting database of environmental metrics was linked with a set of R scripts for modelling malaria epidemic risk and a web atlas application for visualization and dissemination of malaria risk maps. Using the database generated with EASTWeb, we were able to develop statistical models that explained between 39% and 53% of the variability in historical malaria epidemic risk. Key areas for future research include expanding the scope of the system from climate to encompass other social and environmental determinants of malaria, and developing additional tools to better integrate disease surveillance with environmental monitoring.

ACKNOWLEDGMENTS

We thank Isaiah Snell-Feikema and Michael VanBemmel for their work designing and programming the EASTWeb application, and Aashis Lamsal for developing the initial prototype of the software. We also thank Peiyi Xiao and Mark Rop for developing the web atlas software framework. We thank Abere Mihretie of the Health, Development, and Anti-Malaria Association for coordinating data collection activities and facilitating communication with the public health sector. We also acknowledge the cooperation and support of the Amhara Regional Health Bureau and the Federal Democratic Republic of Ethiopia Ministry of Health. This study was supported by National Institute of Health/National Institute of Allergy and Infectious Diseases Grant R01-AI079411. More information can be found on the EASTWeb project website: <http://globalmonitoring.sdstate.edu/eastweb/>.

REFERENCES

- Abeku, T. A., Response to malaria epidemics in Africa. *Emerging Infectious Diseases*, 13, 681-686, 2007.
- DaSilva, J., B. Garanganga, V. Teveredzi, S. M. Marx, S. J. Mason, and S. J. Connor, Improving epidemic malaria planning, preparedness and response in Southern Africa, *Malaria Journal*, 3, 37, 2004.
- Ford, T. E., R. R. Colwell, J. B. Rose, S. S. Morse, D. J. Rogers, and T. L. Yates, Using Satellite Images of Environmental Changes to Predict Infectious Disease Outbreaks, *Emerging Infectious Diseases*, 15, 1341-1346, 2009.
- Guintran J, Delacollette C and Trigg P, 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, 2006.
- R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2012.
- Rop, M., Y. Liu, and M. C. Wimberly, FWA - A Framework for Developing Web-Atlas Applications, Proceedings of the 2011 International Conference on Software Engineering Research and Practice (SERP'11), Las Vegas, NV, July 18-21, 2011.
- Senay, G. B., M. Buddle, J. P. Verdin, and A. M. Melesse, A coupled remote sensing and simplified surface energy balance approach to estimate actual evapotranspiration from irrigated field, *Sensors*, 7, 979-1000, 2007.
- Senay, G. B., M. E. Budde, and J. P. Verdin, Enhancing the Simplified Surface Energy Balance (SSEB) approach for estimating landscape ET: Validation with the METRIC model, *Agricultural Water Management*, 98, 606-618, 2011.
- Thomson, M. C., and S. J. Connor, The development of malaria early warning systems for Africa, *TRENDS in Parasitology*, 17, 438-445, 2001.