Environmental Modelling & Software xxx (2015) 1–11

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Software to facilitate remote sensing data access for disease early warning systems

Yi Liu [a], Jiameng Hu [a], Isaiah Snell-Feikema [a], Michael S. VanBemmel [a], Aashis Lamsal [b], Michael C. Wimberly [b,*]

[a] Department of Electrical Engineering and Computer Science, South Dakota State University, Brookings, SD, USA
[b] Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD, USA

A R T I C L E   I N F O

Article history:
Received 12 August 2014
Received in revised form
29 May 2015
Accepted 12 July 2015
Available online xxx

Keywords:
Software architecture
Workflows
Remote sensing
Early warning
Malaria
West Nile virus

A B S T R A C T

Satellite remote sensing produces an abundance of environmental data that can be used in the study of human health. To support the development of early warning systems for mosquito-borne diseases, we developed an open-source, client-based software application to enable the Epidemiological Applications of Spatial Technologies (EASTWeb). Two major design decisions were full automation of the discovery, retrieval and processing of remote sensing data from multiple sources, and making the system easily modifiable in response to changes in data availability and user needs. Key innovations that helped to achieve these goals were the implementation of a software framework for data downloading and the design of a scheduler that tracks the complex dependencies among multiple data processing tasks and makes the system resilient to external errors. EASTWeb has been successfully applied to support forecasting of West Nile virus outbreaks in the United States and malaria epidemics in the Ethiopian highlands.

Software availability

Software Name: EASTWeb 1.0
Software Requirements: Java JDK6 or above, PostgreSQL 9.3 or above
Programming Language: Java
Availability and Cost: Free, https://epidemia.sdstate.edu/eastweb/
Contact Person: Yi Liu, email: yi.liu@sdstate.edu

1. Introduction

Satellite remote sensing technologies produce an abundance of environmental data that can be used in the study of human health (Ford et al., 2009). Measurements of reflected, emitted, and backscattered radiation provide information about patterns of vegetation, moisture, temperature, and other spatially and temporally variable characteristics of the Earth's surface. These data can be applied to map environmental conditions that influence the development and transmission of pathogens, habitats for vectors and hosts, and human exposures (Estrada-Pena, 2002; Brown et al., 2008; Chuang et al., 2012a; Baeza et al., 2013). Remote sensing has also been used monitor environmental health risks such as extreme heat events (Johnson et al., 2014) and air pollution (Martin, 2008), and to map human populations at risk of disease (Linard et al., 2010). These applications are highly relevant to new interdisciplinary fields such as One Health (Coker et al., 2011) and EcoHealth (Charron, 2012), which have expanded the scope of research on large-scale connections between ecosystems, global change, and human health. At the same time, the continued development of health geography (Kearns and Moon, 2002) and spatial epidemiology (Elliott and Wartenberg, 2004) has provided new conceptual models and analytical tools for the spatially explicit study of environmental health. The current constellation of earth observing satellites thus offers enormous potential for supporting transformational public health research and applications at scales ranging from local to global. In particular, the planetary scope and temporal consistency of most remote sensing missions can provide high-quality environmental monitoring data in developing countries where the health risks from environmentally-mediated infectious diseases are highest and data availability is the most limited.

There has been broad interest in using earth observation data for

* Corresponding author.
E-mail address: michael.wimberly@sdstate.edu (M.C. Wimberly).

http://dx.doi.org/10.1016/j.envsoft.2015.07.006

mapping and forecasting the risk of malaria and other mosquito-borne diseases (Kalluri et al., 2007; Anyamba et al., 2009; Machault et al., 2011). Satellite measurements of rainfall, vegetation greenness, soil moisture, and surface water can serve as indicators of potential mosquito breeding locations, and land surface and air temperature retrievals can provide information about rates of mosquito and pathogen development as well as the frequency of mosquito bites. Mosquito-borne disease outbreaks exhibit lagged responses to these environmental fluctuations because mosquito population sizes and infection rates require time to increase to levels that can trigger an outbreak (Teklehaimanot et al., 2004). Therefore, environmental monitoring using satellite remote sensing has been proposed for developing early warning systems to forecast future epidemics (Rogers et al., 2002; Ceccato et al., 2007; Midekisa et al., 2012). However, this type of early warning application requires long-term historical datasets for model development as well as consistent and timely updates to support operational forecasting.

In the present research, our overarching goal was to implement and test early warning systems for case studies of two mosquito-borne diseases in different social and ecological contexts. West Nile virus (WNV) is a vector-borne and zoonotic disease for which birds are the primary zoonotic reservoir. Spatial and temporal variability in human disease occurrence has been associated with climatic fluctuations that influence vector and host populations (Soverow et al., 2009; Hartley et al., 2012; Kwan et al., 2012; Wimberly et al., 2014b). Our focal area for WNV encompassed South Dakota and the surrounding states in the Northern Great Plains of the United States, a region of high WNV risk (Wimberly et al., 2013). Malaria is a vector-borne disease caused by *Plasmodium* parasites that are transmitted between human hosts by mosquitoes. In many environmental settings, the transmission of malaria is highly sensitive to seasonal and interannual fluctuations in air temperature, rainfall, and other environmental factors that influence the population dynamics of the mosquito vector and development rate of the parasite (Stresman, 2010). Our focal area for malaria was the Amhara region of Ethiopia, a highland region with highly seasonal rainfall and strong elevation-driven temperature gradients where epidemics are associated with climatic fluctuations (Wimberly et al., 2012b).

In both case studies, the specific objectives were to identify environmental risk factors associated with high numbers of human disease cases at lead times ranging from weeks to months, and to use these risk factors to generate dynamic risk maps for use by the public health sector. To meet these objectives, we utilized environmental information from a variety of earth science data products. These products are produced by processing raw data acquired by sensors on board earth-observing satellites, and are then stored in online archives hosted by discipline-oriented data centers. Despite the fundamental similarities of many of the data products, they also have a diversity of source characteristics that include different access methods, data formats, grid sizes, time step lengths, and frequency and latency of updates. Multiple additional steps are therefore required before these “big data” can be linked with smaller epidemiological or entomological datasets for analysis and modeling, and the necessary workflows can be complex (Nativi et al., 2015; Vitolo et al., 2015). To achieve our goal of developing and implementing disease early warning systems, we therefore needed to (1) automatically acquire and process large volumes of historical remote sensing data from online archives, (2) process environmental data from disparate sources into a unified database format that could be linked to epidemiological data for model development, and (3) automatically update the environmental database to make forecasts as new data became available.

To address these needs, we developed the Epidemiological Applications of Spatial Technologies (EASTWeb) open-source software tool to facilitate access to earth science datasets for public health research and application. EASTWeb is a client-based desktop application that allows users to specify an area and time period of interest and then automatically carries out multiple data acquisition and processing steps to generate environmental variables that can be readily linked with epidemiological data. The application generates historical databases from archived data and will continue to update them in real time as new data become available. Although the software was developed to support disease early warning in our case study areas, it also has the potential to be used more broadly by GIS analysts, biostatisticians, epidemiologists, and others who need access to streams of environmental information that can be immediately connected with health outcome data to conduct analyses and generate risk maps, forecasts, and other products. In this paper we describe the design of the EASTWeb software, highlight key elements of the system architecture that were developed to achieve our project objectives, present examples of how the software has been applied to support the forecasting of mosquito-borne diseases, and discuss lessons learned and future development plans.

2. System overview

EASTWeb is an open-source client-based application that automatically connects to earth science data archives and acquires, processes, and summarizes selected remote sensing datasets. The user defines a project by specifying a geographic extent and historical time period and providing geospatial data layers that mask out water bodies and characterize the extent and minimum mapping unit of the epidemiological data (e.g., counties, census tracts, or buffered plots) that will be integrated with the remote sensing data stream. After first building a historical database, EASTWeb continues to add information as new data become available on the earth science data archives. Environmental data summaries are stored in a relational database that can be easily queried and linked to ecological and epidemiological datasets for analysis and forecasting in R and other software environments. The software is programmed using JAVA for overall system control and use the Standard Widget Toolkit (SWT), (https://www.eclipse.org/swt/) for user interface development. Spatial analyses are carried out using the open source Geospatial Data Abstraction Library (http://www.gdal.org). PostgreSQL (http://www.postgresql.org) is used to store and manipulate the resulting data summaries. EASTWeb system has been tested and run in Windows 7 and Windows 8. The software package provided on the EASTWeb website (https://epidemia.sdstate.edu/eastweb/) is for Windows.

The current version of EASTWeb incorporates information streams from multiple earth science data archives. MODIS land surface temperature (MOD11A2) and nadir BRDF-adjusted reflectance (MCD43B4) products are obtained from the Land Processes Distributed Active Archive Center (LP DAAC, https://lpdaac.usgs.gov). TRMM rainfall products (3B42 and 3B42RT) are obtained from the Goddard Earth Sciences Data and Information Services Center (GES DISC, http://disc.sci.gsfc.nasa.gov/). The daily potential evapotranspiration (ETo) product is obtained from the USGS Early Warning and Environmental Monitoring Program (https://earlywarning.usgs.gov/) (Table 1).

The EASTWeb system encompasses four main types of data processing (Fig. 1): *downloading* data from online archives, *remote sensing data processing*, *computing environmental indices*, and *summarizing data*. The *downloading* step acquires data for the selected study area and time period. Downloaded data come in a variety of formats including scientific data libraries such as NetCDF and HDF as well as georeferenced image formats such as GeoTIFF.

Please cite this article in press as: Liu, Y., et al., Software to facilitate remote sensing data access for disease early warning systems, Environmental Modelling & Software (2015), http://dx.doi.org/10.1016/j.envsoft.2015.07.006