PREDICTING DENGUE INCIDENCE IN BRAZIL USING BROAD-SCALE SPECTRAL REMOTE SENSING IMAGERY

Amanda Ziemann, Geoffrey Fairchild, Jessica Conrad, Carrie Manore, Nidhi Parikh, Sara Del Valle, and Nicholas Generous

Los Alamos National Laboratory, Los Alamos, NM, 87545 USA

ABSTRACT

Infectious disease burden is continuing to increase around the globe. These diseases have increased, in part, due to globalization, human behavior, and environmental changes. There is an urgent need for improved prediction of their spread so that mitigation techniques and treatments can be administered proactively rather than just reactively. One of the challenges is that many regions of interest are in hard-to-reach locations, and as such, clinical surveillance data (reliant upon self-reporting) can be both sparse and lagging. Remote sensing imagery is an attractive data source to exploit for this application as it provides real-time information without having to physically be on the ground. Here, we derive standard indices from multispectral imagery, and explore how predictive they are for forecasting dengue incidence in Brazil. This is done on broad spatial and temporal scales, covering all of Brazil for multiple years. Results will show that the normalized difference vegetation index is a leading predictor for dengue incidence.

Index Terms— multispectral, remote sensing, mosquito-borne disease, forecasting, dengue, Brazil

1. INTRODUCTION

Traditionally, epidemiology has focused on the study of public health data and the assessments of health programs in order to devise optimal prevention as well as educational programs. However, these approaches lack the ability to predict and forecast disease trends based on complex links among humans, technology, and the environment [1, 2]. For instance, environmental changes can play a major role in public health, but are not always captured in disease surveillance systems [3, 4]. These environmental changes may be due to climate change (*e.g.*, droughts, rise in temperature, wildfires) or human-driven activities (*e.g.*, construction, pollution, deforestation). Multispectral remote sensing imagery (Fig. 1) provides a means for remotely characterizing many of these changes, albeit some more straightforwardly than others [5–8]. In this study we leverage clinical surveillance data for dengue in Brazil, as well as access to a centralized image repository for different satellites [9], in order to analyze the effectiveness of satellite-derived indices for predicting dengue incidence. This is part of a larger effort that integrates Internet data (*e.g.*, social media, online searches) [10], climatological data, demographic data, and remote sensing data, with the aim of developing a heterogeneous data forecasting tool for dengue incidence in Brazil.

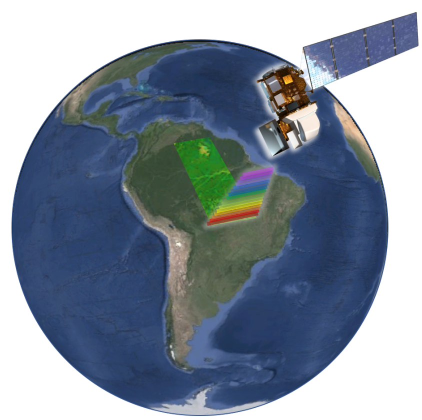

Fig. 1: Illustration of Landsat 8 collecting spectral image swaths over Brazil. We leverage this broad-scale, repeatable image collection for time series analysis of spectral indices.

2. THE ROLE OF REMOTE SENSING

When integrating multiple data sources into a predictive model, subject matter expertise is critical in order to provide context to the different data sources. The primary contributor to the spread of mosquito-borne diseases is the presence of standing water. In terms of remote sensing, there are various secondary indicators that—while they may not directly measure standing water—can be correlated to standing water and thus serve as proxy measurements. We looked at the normalized difference vegetation index (NDVI, Eq. 1) [11], two versions of the normalized difference water index (NDWI, Eqs. 2 & 3) [12], and the normalized burn ratio (NBR, Eq. 4) [13], as well as the percentage of cloud cover within

This work was supported by Los Alamos National Laboratory internal funding. Send correspondence to: *ziemann@lanl.gov*

the given area of interest. NDVI is an indicator of healthy green vegetation, $NDWI_1$ is an indicator of water content in leaves, $NDWI_2$ is an indicator of water content in large water bodies, and NBR is an indicator of burned areas and fire severity. High NDVI, NDWI, and percentage of cloud cover can all be linked to high rainfall, which in turn correlates to increased standing water. Brazil has also had wildfires within the last decade, making NBR important in order to account for that change in environment.

$$\text{NDVI} = \frac{\text{NIR - Red}}{\text{NIR + Red}} \tag{1}$$

$$\text{NDWI}_1 = \frac{\text{Green - NIR}}{\text{Green + NIR}} \tag{2}$$

$$\text{NDWI}_2 = \frac{\text{NIR - SWIR}_1}{\text{NIR + SWIR}_1} \tag{3}$$

$$\text{NBR} = \frac{\text{SWIR}_1\text{ - SWIR}_2}{\text{SWIR}_1 + \text{SWIR}_2} \tag{4}$$

3. EXPERIMENT AND RESULTS

Our clinical surveillance data for dengue in Brazil covers 2010-2017 on a weekly basis (specifically epidemiological weeks), and is at the municipality level. At present, Brazil has 5,570 municipalities, which are combined into 136 meso-regions. Fig. 2 illustrates dengue incidence in 2015 when combined at the meso-region level.

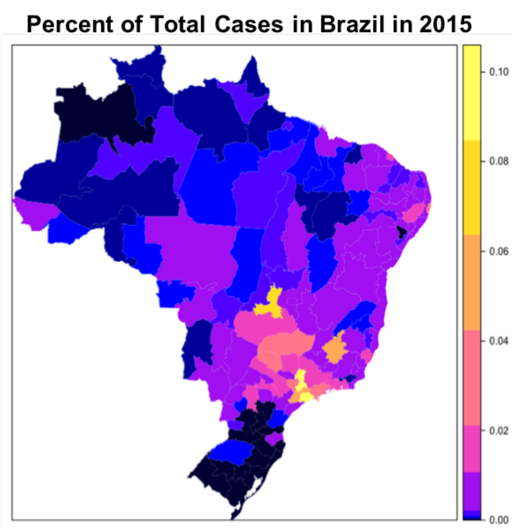

Fig. 2: Total cases in Brazil in 2015 at the meso-region level.

For each week, for each municipality, for seven years, we computed the NDVI, $NDWI_1$, $NDWI_2$, and NBR for every pixel in the municipality, and then aggregated the results to municipality-level statistics. So, for each municipality, on a weekly basis we have: the mean, min, max, and standard deviation of all of the pixel-level NDVI, $NDWI_1$, $NDWI_2$, and NBR values, as well as the percentage of cloudy pixels, for a total of 17 different values; this resulted in an overall computation of over 32,000,000 satellite-derived metrics. We also leveraged four multispectral satellites in this process: Landsat 5, Landsat 7, Landsat 8, and Sentinel-2. We used Descartes Labs' imagery platform to access the imagery in a centralized manner, completing all of the computations in python while leveraging their python-based API. This study utilized every Landsat 5, 7, 8 and Sentinel-2 image that covered Brazil over the seven year period, for a total of over 26 terabytes of imagery. Our analysis showed that NDVI was the most predictive for dengue incidence, and as such is the focus of the results presented here (Fig. 3).

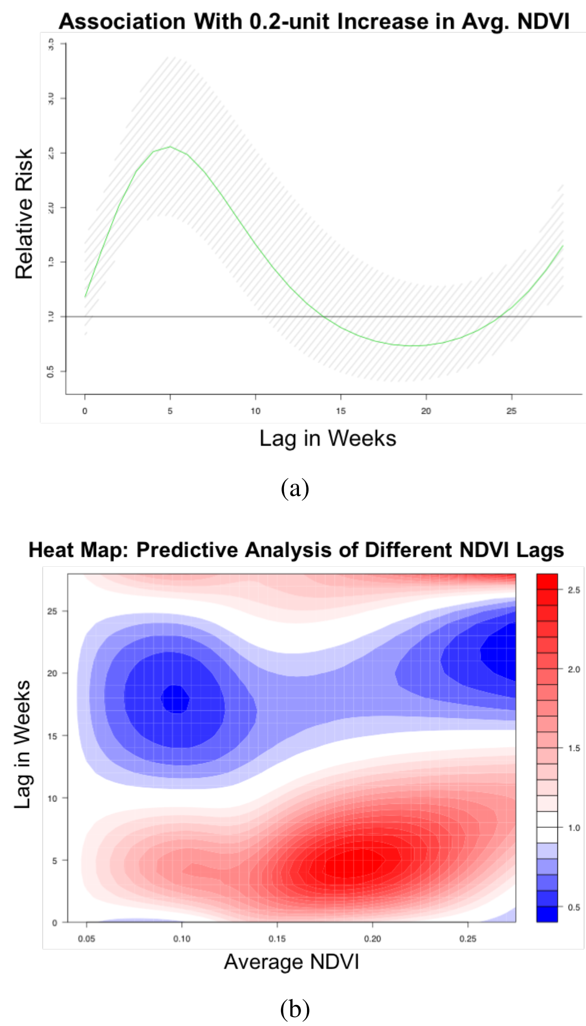

Fig. 3: Distributed lag nonlinear model for (a) average NDVI = 0.2 (indicating relative risk), and (b) multiple values of average NDVI. The heat map in (b) is used to find the optimal lag for correlating average NDVI with dengue incidence.

Distributed lag nonlinear models were used for this analysis. They allow for quantification of how a particular data source contributes to the prediction of dengue incidence as either a leading or lagging indicator. Results here suggest that high NDVI is associated with a 5-week leading indicator, or in other words, if NDVI is high, then in about 5 weeks there will be an uptick in dengue incidence. This is consistent with mosquito breeding rates, as the lag time from standing water to mosquito breeding is about 5 weeks. In the risk map shown here, if NDVI is low, then there will be fewer dengue cases in the near future (the blue portion of the risk map), and if NDVI is high, there will be more dengue cases in the near future (the red portion of the risk map). These results are being integrated into a larger heterogeneous forecasting model, outside of the scope of this paper.

4. CONCLUSIONS

The spread of infectious and mosquito-borne diseases has important implications for public health as well as regional stability, in turn affecting national security. The implications of vector-borne diseases are far-reaching. As such, a capability for better predicting and forecasting the spread of these diseases would be of high utility, as it would enable mitigating and preventative action to take place. This paper explores the utility of using multispectral remote sensing imagery for predicting dengue incidence in Brazil, as remote sensing is an appealing data source to leverage for hard-to-reach areas. Results of this study demonstrate that out of the considered indices, NDVI is the most consistently predictive, and in particular is a 5-week leading indicator for increased dengue incidence. Our study was as comprehensive as possible from a remote sensing standpoint, leveraging every moderate-resolution global-coverage satellite that imaged Brazil over a seven year period, ultimately utilizing over 26 terabytes of imagery. The results of this study are being integrated into a heterogeneous forecasting model that—in addition to remote sensing—utilizes climatological data, demographic data, Internet data, and clinical surveillance data.

5. REFERENCES

[1] C. Manore, K. Hickmann, S. Xu, H. Wearing, and J. Hyman, "Comparing dengue and chikungunya emergence and endemic transmission in A. aegypti and A. albopictus," *Journal of Theoretical Biology*, vol. 356, pp. 174–191, 2014.

[2] C. Manore, J. David, R. Christofferson, D. Wesson, J. Hyman, and C. Mores, "Towards an early warning system for forecasting human West Nile virus incidence," *PLoS Currents*, vol. 6, 2014.

[3] L. Eisen and S. Lozano-Fuentes, "Use of mapping and spatial and space-time modeling approaches in operational control of Aedes aegypti and dengue," *PLoS Neglected Tropical Diseases*, vol. 3, no. 4, 2009.

[4] V. R. Louis, R. Phalkey, O. Horstick, P. Ratanawong, A. Wilder-Smith, Y. Tozan, and P. Dambach, "Modeling tools for dengue risk mapping - a systematic review," *International Journal of Health Geophysics*, vol. 13, no. 50, 2014.

[5] A.L. Buczak, P.T. Koshute, S.M. Babin, B.H. Feighner, and S.H. Lewis, "A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data," *BMC Medical Informatics and Decision Making*, vol. 12, no. 124, November 2012.

[6] V. Machault, A. Yebakima, M. Etienne, C. Vignolles, P. Palany, Y.M. Tourre, M. Guerecheau, and J.-P. Lacaux, "Mapping entomological dengue risk levels in martinique using high-resolution remote-sensing environmental data," *ISPRS Int. J. Geo-Inf.*, vol. 3, no. 4, pp. 1352–1371, December 2014.

[7] S.I. Hay, R.W. Snow, and D.J. Rogers, "From predicting mosquito habitat to malaria seasons using remotely sensed data: practice, problems and perspectives," *Parasitology Today*, vol. 14, no. 8, pp. 306–313, 1998.

[8] T.R. Allen and D.W. Wong, "Exploring GIS, spatial statistics and remote sensing for risk assessment of vector-borne diseases: a West Nile virus example," *Int. J. of Risk Assessment and Management*, vol. 6, no. 4-6, 2006.

[9] Descartes Labs: a platform for geospatial science, "https://descarteslabs.com/platform.html," 2018.

[10] N. Generous, G. Fairchild, A. Deshpande, S. Del Valle, and R. Priedhorsky, "Global disease monitoring and forecasting with wikipedia," *PLoS Computational Biology*, vol. 10, no. 11, 2014.

[11] J.W. Rouse, R.H. Haas, J.A. Scheel, and D.W. Deering, "Monitoring vegetation systems in the great plains with ERTS," *Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium*, vol. 1, pp. 48–62, 1974.

[12] B.-C. Gao, "NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sensing of Environment*, vol. 58, no. 3, pp. 257–266, 1996.

[13] M.J. Lopez Garcia and V. Caselles, "Mapping burns and natural reforestation using thematic Mapper data," *Geocarto International*, vol. 6, no. 1, pp. 31–37, 1991.