Mapping dengue vulnerability in Peru

Charles Godfrey Kelsey Maass Connor Sawaske Saumya Sinha

***Abstract*—Mosquito-borne illnesses pose a severe health risk to the people of Peru, and to ensure efficient deployment of preventive resources it is necessary to identify high-risk regions. Data-driven analytical tools can aid healthcare providers and policy-makers with this task. To this effect, we develop an interactive visualization tool displaying the incidence of dengue in Peru from 2010 to 2014. The interaction is designed to allow a user to observe temporal as well as geographical trends - reported cases are classified by department (the largest administrative sub-unit of Peru) and month. In particular, the tool highlights high-incidence regions and establishes a visual connection between dengue outbreaks and climatic factors such as temperature and precipitation.**

I. Introduction

Infections such as Zika, West Nile, dengue, and malaria wreak havoc on populations with limited access to medical services. Identifying communities that are especially susceptible to infections and outbreaks is essential for the prevention and containment of these diseases. Putting data-driven analytical tools in the hands of healthcare providers in the most at-risk locations may help them to deploy limited resources as efficiently as possible. For example, predicting optimal times and locations to spray for mosquitos, distribute nets or re-allocate medical services could enhance the efficacy of these measures. Unfortunately, there are few existing resources of this sort targeted towards dengue prevention in Peru. Inspired by the visualizations for investigating infectious diseases in Africa provided by [11], we teamed up with The UW MetaCenter for Global Disease Preparedness to build tools for this purpose.

Research suggests that environmental factors play an influential role in the spread of dengue; for instance, [6] demonstrates a strong correlation between dengue infections and precipitation in Singapore. A compilation of such studies for different parts of the world is available in [9]. With this motivation, a core component of our tools was to allow users to compare the number of dengue cases reported (both geographically and temporally) against environmental variables like temperature, precipitation and the Enhanced Vegetation Index (EVI).

Related work

The Peruvian Centro Nacional de Epidemiologia provides a basic visualization of time series of dengue cases [3]. In [8], the spread of dengue in Peru is analyzed at varying levels, and its relationship with climatic variables is considered in [4]. The visualizations developed in [11] to study epidemic viral hemorrhagic fevers in Africa inspired our design choices (and the overarching goal of our project).

II. An interactive visualization tool

A. Methods

The UW MetaCenter for Global Disease Preparedness provided the bulk of our case data; we supplemented this with the dataset [3] from the Centro Nacional de Epidemiologia. Exploratory analysis showed interesting geographic trends across the different departments (administrative divisions/regions) of Peru, and this motivated us to display case data on a map. To do so, we obtained GeoJSON data converted from shapefiles provided by the Ministerio del Ambiente del Peru in [1]. As previously mentioned, existing research has shown correlation between climate variables and dengue incidence. To incorporate weather data, we used GeoTIFF files obtained from the Metacenter, and averaged over the departments according to our GeoJSON perimeter files. Supplementary weather data was obtained from the Climatic Research Unit of the University of East Anglia [7].

Exploratory analysis guided many of the implementation details of our visualizations. For instance, after looking at distributions of infection and precipitation counts (Figure 1), it was clear that a log scale would be more effective than a linear scale.

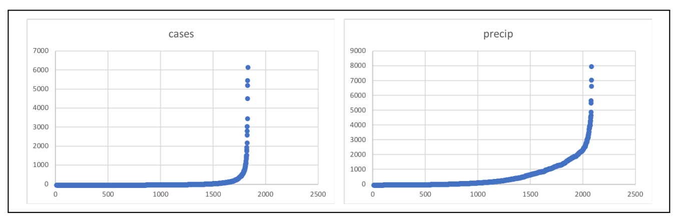

Fig. 1: Initial scatter plots of dengue cases and precipitation. The sharp change in scale indicates that log-scale plots will give better visual resolution.

The case data for each year was available as a cumulative count through the months, and monthly incidence was calculated using successive differences. Figure 2 shows the total number of cases for each monthly period from the years 2010 to 2014. The figure suggests a large outbreak in August of 2012. However, we recently learned from our collaborators at the Metacenter that the case-reporting system changed three months prior, and the data has missing values for the months May-July, 2012. Therefore we hypothesize that this spike should actually be spread out over the months May-August, 2012; we have left the provided data as-is for the moment.

Before explaining our choices of visualizations, we mention some approaches that we attempted but ultimately deemed unsuccessful. Given the seasonal nature of dengue infection (apparent from our initial analysis), we hoped a connected scatter plot would be an effective encoding. In practice, however, it did not produce anything meaningful (one team member summed up the result as “a mess of tangled wires”). In a similar vein, time-series presentations of climate data in each province proved to be uninteresting - for instance, the equatorial climate of Peru produces few significant temperature variations in any given department. Our final product consists of a browser-based interactive visualization implemented using d3. We split the design into three main components, the details of which are discussed in the next section.

B. Results

Our objective is to allow a user to explore temporal and geographic trends in dengue incidence in Peru over the years 2010 to 2014, and to visually investigate relationships between the reported number of cases and three environmental variables: mean temperature (°C), mean precipitation (mm) and Enhanced Vegetation Index (a dimensionless score on a scale of 0 to 100). With these goals in mind, the interface is divided into three tabs labeled “Temporal,” “Maps” and “Correlations,” each dedicated to one of the primary-use cases communicated to us by our collaborators.

The Temporal tab focusses on seasonal and annual variation in the rate of dengue. The choice of encoding for these trends is a bar chart; Figure 2 gives a snapshot of this tab, with its interactive components highlighted. The default view shows reported cases by month from January 2010 to December 2014. Using a dropdown menu, a user can aggregate the data by month or year, and filter to a specified department.

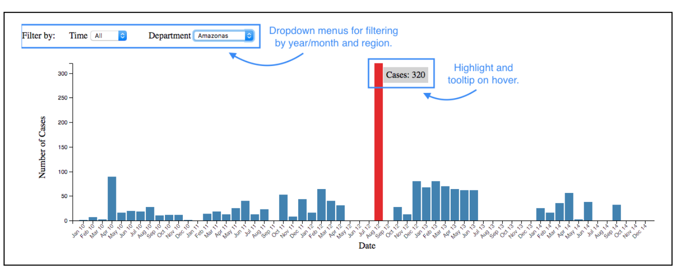

Fig. 2: Incidence of dengue in Amazonas, Peru, from 2010 to 2014.

The Maps tab widens the scope, enabling analysis of spatial trends and dependence on environmental measures. Using small multiples ([12, Chapter 4]), it displays four heat-maps overlaid on separate maps of Peru, with departments saturated according to the number of dengue cases reported, mean temperature, mean precipitation and EVI respectively. Figure 3 shows a snapshot of this tab.

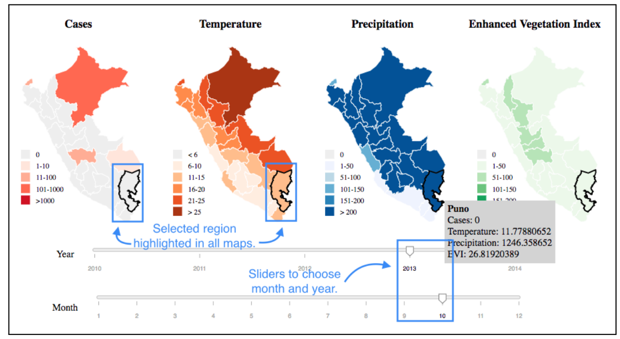

Fig. 3: Maps showing dengue cases, mean temperature, precipitation and EVI for Peru for October, 2013.

At any given instant, the Maps tab illustrates a slice of the data corresponding to a given year and month, both of which can be varied using two sliders. Hovering over a department on any of the maps highlights it on all four, for greater contextual reference. A tooltip displays precise values of each of the four measures. Each map uses a different hue; these hues are chosen to naturally reflect what the map represents (orange for heat, blue for water, green for vegetation), and also to be easily distinguishable ([5]). Saturation levels vary by (binned) values being plotted, making identification of gradients a perceptual task rather than a cognitive one. We remark that the maps are all plotted at once and juxtaposed, rather than e.g. cases versus one environmental variable at a time, allowing a user to examine correlations (or lack therof) between all variables at once.

Since inference of quantitative relationships from saturation gradients is limited, we investigate these relationships numerically in the Correlations tab. This tab contains scatter plots of dengue cases (in log scale) versus the three climatic measures. Once again, we used the method of small multiples; see Figure 4.

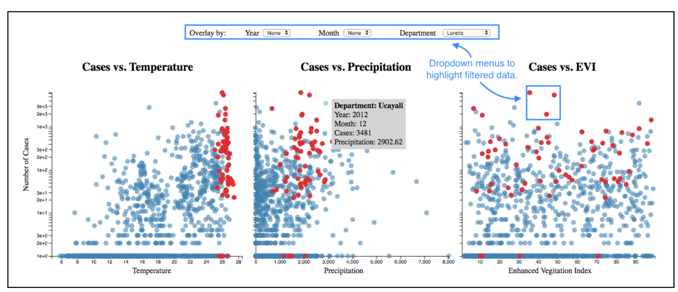

Fig. 4: Cases of dengue (in log-scale) versus temperature, precipitation and EVI. The data-points corresponding to Loreto are highlighted.

Each point in each plot corresponds to a year-month-department tuple, and all the data is plotted at once. Transparency is used to limit occlusion due to over-plotting. Using drop-down menus a user can select a particular year, month and/or department. As opposed to the bar-chart in Figure 2, we retain the original data and only highlight the filtered data in a different color. This gives context for the filtered data within the space of all observations.

C. Discussion

The time series tools in our Temporal tab permit some extrapolation of when certain departments are more vulnerable to dengue. This may inform the deployment of preventative measures. For instance, a community education approach could be more effective shortly before the month when the surrounding region has historically had more cases, as opposed to its typical down-season.

The Maps tab highlights departments that suffer most drastically from dengue, which may inform dengue prevention strategy for the entirety of Peru. Users interested in regional climatic dependencies can see how dengue infections relate to temperature, precipitation, and EVI.

Further analysis of dengue's correlation (or lack thereof) with these chosen measures can be carried out in the Correlations tab, which provides an alternate view of the data. We find that the overlay technique employed in these scatter plots illustrates how a single region or time period fits into the bigger picture - the forest is not lost for the trees, if you will.

III. Future Work

The ultimate goal is to prevent dengue infections; with this in mind it will be important to incorporate predictive methods. In particular, the climate of Peru is expected to change drastically in the coming years. Research predicts that over the next 100 years, temperature will increase by up to 3-3.5°C across Peru ([10]). Given that our tools show a correlation between the rate of dengue cases and temperature, the effect of climate change could make dengue an even more severe threat. One direction for improvement of our toolkit would be a display of projected future infection rates based on climate change predictions.

Modeling the effectiveness of various methods of dengue prevention would make our visualization tools more useful to healthcare providers and policy makers. Some research has been done with this goal in mind (for instance, [2] suggests that screening houses provides a large reduction in dengue outbreaks, while spraying insecticides actually has an adverse effect). An understanding of the relative effectiveness of existing preventative measures would be invaluable to organizations fighting mosquito-borne illnesses.

Finally, the use of data of some higher fidelity could be of great benefit to our toolkit. Being able to change the view to individual municipalities in a department or to cases reported per-week would allow for even more trends to be analyzed. It would also be useful to incorporate reporting-uncertainty in the dengue incidence being visualized. Moreover, we would like to incorporate visualizations which allow simultaneous quantitative comparison between multiple departments, relying less on the user's memory than they currently do.

Acknowledgements

Our sincere thanks to Prof. Peter Rabinowitz and Dr. David Pigott at the University of Washington for suggesting this project and providing us with invaluable contextual information to help define our goals.

References

[1] https://github.com/juaneladio/peru-geojson. Accessed: 2018-05.

[2] Leigh R. Bowman, Sarah Donegan, and Philip J. McCall. Is dengue vector control deficient in effectiveness or evidence?: Systematic review and meta-analysis. *PLOS Neglected Tropical Diseases*, 10(3):1–24, 03 2016.

[3] Prevención y Control de Enfermedades Centro Nacional de Epidemiología. http://www.dge.gob.pe/salasituacional/sala/index/3_GraficoxSE/84. Accessed: 2018-05.

[4] Gerardo Chowell, Bernard Cazelles, Hélène Broutin, and Cesar V. Munayco. The influence of geographic and climate factors on the timing of dengue epidemics in Peru, 1994-2008. *BMC Infectious Diseases*, 11(1):164, Jun 2011.

[5] Connor C. Gramazio, David H. Laidlaw, and Karen B. Schloss. Colorgorical: Creating discriminable and preferable color palettes for information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 23(1):521–530, January 2017.

[6] Yien Ling Hii, Huaiping Zhu, Nawi Ng, Lee Ching Ng, and Joacim Rocklöv. Forecast of dengue incidence using temperature and rainfall (forecast of dengue incidence in singapore). *PLOS Neglected Tropical Diseases*, 6(11), November 2012.

[7] Harris I., Jones P.D., Osborn T.J., and Lister D.H. Updated high-resolution grids of monthly climatic observations – the cru ts3.10 dataset. *International Journal of Climatology*, 34(3):623–642.

[8] Amy C. Morrison et al. Epidemiology of dengue virus in iquitos, peru 1999 to 2005: Interepidemic and epidemic patterns of transmission. *PLOS Neglected Tropical Diseases*, 4(5):1–17, 05 2010.

[9] Suchithra Naish, Pat Dale, John S. Mackenzie, John McBride, Kerrie Mengersen, and Shilu Tong. Climate change and dengue: a critical and systematic review of quantitative modelling approaches. *BMC Infectious Diseases*, 14(1), Mar 2014.

[10] (UKMO) Met Office, Simon N. Gosling, et al. *Climate: observations, projections and impacts*. Met Office, 2011.

[11] David M. Pigott et al. Local, national, and regional viral haemorrhagic fever pandemic potential in africa: a multistage analysis. *The Lancet*, 390(10113):2662–2672, December 2017.

[12] Edward R. Tufte. Envisioning information, 1990.