DengAI - Analysing the Effects of Climate on the Spread of Dengue

Aditi Killedar
Computer Science and Engineering
PES University
Bengaluru, India
aditi.s.killedar@gmail.com

Chandradhar Rao
Computer Science and Engineering
PES University
Bengaluru, India
chandradhar.rao@gmail.com

Dhruval PB
Computer Science and Engineering
PES University
Bengaluru, India
dhruvalpb@pesu.pes.edu

***Abstract*—An accurate model to predict possible outbreaks of Dengue Fever enables a nation to equip itself in the fight against the disease. In this study, we present a model capable of forecasting the emergence of Dengue cases. With the help of statistical and analytical techniques we gain insight into how the weather conditions and vegetation of a region affects the number of Dengue cases reported, enabling our model to perform accurate predictions about the outbreaks. These predictions would serve as an early warning estimate.**

***Index Terms*—Dengue Fever, Correlation Analysis, Time Series Analysis, Climate**

I. Introduction

Dengue Fever is a vector-borne viral disease occurring in tropical and subtropical regions. The primary symptoms of Dengue include high fever and joint pain. In severe cases, Dengue can also lead to life-threatening symptoms like serious bleeding and shock. It is spread by the female *aedes aegypti* mosquito. These mosquitoes thrive in warm and humid conditions. Putting in place a warning system that can predict the intensity of the outbreak of cases is exceedingly useful while tackling the disease. Without suitable vaccines, the only way of controlling its spread lies in controlling the population of the mosquitoes that spread it. Using a predictive model to forecast the next Dengue outbreak months in advance will help put in place effective practices for vector control. Health officials will have more insight and this will enable them to be well-prepared in the event of rising Dengue Fever cases.

The authors of [1] present reports showing an increase in Dengue Fever occurrences during summers, owing to the abundance of breeding places for mosquitoes in ample sunshine and precipitation. Recent studies have shown global warming has aggravated the number of Dengue Fever cases as it has made the climate during spring and winter warmer and wetter making conditions favourable for both the Dengue virus as well as its vector throughout the year.

In this study we find the association between weather variability, vegetation and the occurrence of Dengue Fever, and hence predict the number of cases of Dengue under the given weather conditions using a time-series model. This model could be integrated into a warning system used by officials to estimate the number of cases that may emerge in the future.

II. Related Work

A. *Climate change and Dengue Fever transmission in China: evidences and challenges*

The authors of [1] have compiled a comprehensive review of 81 publications regarding Dengue Fever and how climatic conditions affect its spread in China. The impact of climate on Dengue Fever is three fold, affecting the virus, the vector and the transmission environment. They present a framework named "Tri-model Supported Prediction of Dengue Fever Distribution and Patterns" in order to model the three fold effect of climate on Dengue Fever spread. It is a culmination of three models, each describing one aspect of the connection, weather, mosquito distribution/population and climate. The Climate model estimates climate patterns and the dynamics of weather conditions as well as provide crucial inputs to the mosquito model which would produce outlooks of the vectorial capacity across space and through time providing a holistic spatio-temporal view of the problem. There are studies showing Dengue Virus is very sensitive to changes in temperature as its vector's body temperature is directly dependent on the ambient temperature owing to its incapacity to perform thermo-regulation. Rise in temperature accelerates the reproduction while shortening the extrinsic incubation period (EIP) of Dengue Virus. One of the studies the authors reviewed found that the EIP of Dengue Virus reduces from 12 days at 30 °C to 7 days at 34 °C. Multiple studies have shown that precipitation, wind velocity, temperature and humidity impact the survival and reproduction of mosquitoes.

Temperature affects mosquitoes in multiple aspects. The incubation period in mosquitoes is longer in colder environments. Warmer conditions increase mosquito density (however, temperatures in excess of 32 °C cause mosquito density to drop significantly). The biting behaviour of female mosquitoes (which need a blood meal for ovarian development) is influenced by temperature. The optimal temperature for high biting activity is between 25 °C 30 °C. Spatio-temporal range of mosquitoes is increased due to global warming as they are active throughout the year owing to warmer conditions and have extended their habitats to previously colder places as well. There have also been numerous studies showing the impact of temperature on the flying distance of mosquitoes.

Colder temperatures reduce the range of mosquitoes to only a few metres while optimal temperature conditions of 20 °C to 27 °C boost the range to hundreds of metres. Using the classification and regression tree (CART) model, [2] found that temperature at a lag of 2 months was positively associated with Dengue Fever incidence, whereas diurnal temperature range had an inverse association with Dengue Fever transmission at a lag of 1 month. [3] showed that temperature, humidity and precipitation were positively associated with Dengue Fever incidence, while air pressure had a negative effect on it at a lag of 3 months.

Precipitation plays a vital role in providing mosquitoes with essential breeding sites. Although mosquitoes often lay eggs in shallow water that evaporates before their eggs hatch, deep puddles and stagnant swamps are a mosquito's favorite place to breed, which are directly affected by the amount of precipitation the place sees. Studies have also shown that humid conditions alter the mosquitoes behaviour. A mosquito is more likely to bite in warm and humid conditions compared to cold and dry conditions. Wind velocity impacts how far the mosquito can fly. Through their extensive survey, a major conclusion drawn is that the results are not always consistent. While one study shows a positive correlation between some weather factors, other studies from a different region show the opposite. The authors present several such studies. Hence, what we can conclude from this is that the spread of Dengue Fever is hyper-local and one cannot generalise the effects weather conditions have on it.

The authors also discuss different models based on different distributional assumptions. These include the likes of Poisson regression model for time series analysis to determine the association between weather variables (minimum temperature, precipitation) and Dengue Fever, structural equation models to explore the direct and indirect effects of temperature and precipitation on Dengue Fever occurrences, generalized additive model (GAM) and zero-inflated GAM for an ecological model of Dengue Fever, a Dengue Fever model using GAM with Poisson link function by considering key weather factors, negative binomial regression model with a log link function to analyze the relationship between weekly Dengue Fever cases and climate factors and a lot more interesting approaches researchers have used to far.

B. *Weather as an effective predictor for occurrence of Dengue Fever in Taiwan*

The authors of [4] explored the effects of weather variability on the occurrence of Dengue Fever in the Kaohsiung City of Taiwan. Due to global warming, a warming trend has been observed in Taiwan, which can extend an explanation as to why Dengue Fever cases in Taiwan's major city are emerging. These mosquitoes thrive in warm and humid conditions, making Taiwan's weather ideal for the mosquitoes to breed. The re-emergence of Dengue Fever can be credited to the re-emergence of the vector, increased contact with vulnerable subjects, the rising severity of global warming and the vectors' growing resistance to insecticides. The authors explore the effects of climate change on the cases of Dengue Fever. Since there exist differences between sites in terms of immunity, severity of vector's presence and more, a localised analysis is done using site-specific data. The ARIMA (auto-regressive integrated moving average) model was used to analyse weather variation. Overall increasing temperature in different regions of the world might allow these vectors to survive over winter and help to extend into regions previously free of disease or exacerbation of transmission in endemic regions and also change the transmission seasons. This study concluded that temperature and relative humidity were statistically associated with the incidence of Dengue Fever, with a lag of 2 months time.

C. *Forecast of Dengue Incidence Using Temperature and Rainfall*

The authors of [5] used a Poisson time-series multivariate model and analyzed various lag times between Dengue and weather variables to identify the optimal Dengue forecasting period. They conclude that weather can be a crucial factor for providing early warning of Dengue epidemics, and the long term sustainability of forecast precision is challenging, considering the complex dynamics of disease transmission, which include a lot of other factors. As temperature increases, Aedes mosquitoes display shorter periods of development in all stages of the life cycle leading to increased population growth; the mosquito feeding rate also increases; and Dengue viruses in Aedes adult mosquitoes require shorter incubation periods to migrate to salivary glands. Conversely, high temperatures above 35°C or heavy rainfall possibly lower Dengue transmission by reducing the survival rate of Aedes, and thus reducing the cases of Dengue Fever in the local area. Autocorrelation was included in the model to account for the serial relationship between past and current cases. A Poisson multivariate regression model was composed combining equations that account for various factors like the meteorological cycle, lag terms, and the serial correlation of Dengue cases. The study then evaluated the model using the ROC curve, which revealed a good fit of the model to the actual predictions. Some setbacks involved here are the fact that the Dengue Fever epidemic depends on a lot of other dynamic factors like the movement of people, the immunity of the community, which have not been taken into account in this predictor.

D. *Lag effect of climatic variables on Dengue burden in India*

The authors of [6] explore the relationship between El Niño Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD) and Dengue cases in India. Additionally, a distributed lag non-linear model was used to assess the delayed effects of climatic factors on Dengue cases. The study shows that Dengue cases usually follow a seasonal pattern, with most cases reported in August and September. Both temperature and rainfall were positively associated with the number of Dengue cases. The precipitation shows the higher transmission risk of Dengue was observed between 8 and 15 weeks of lag. The highest relative risk (RR) of Dengue was observed at 60 mm rainfall with a 12-week lag period when compared with 40 and 80 mm rainfall. The RR of Dengue tends to increase with increasing mean temperature above 24 °C. The largest transmission risk of Dengue was observed at 30 °C with a 0–3 weeks of lag. Similarly, the transmission risk increases more than twofold when the minimum temperature reaches 26 °C with a 2-week lag period. The Dengue cases and El Niño were positively correlated with a 3–6 months lag period. The significant correlation observed between the IOD and Dengue cases was shown for a 0–2 months lag period.

E. Climatic factors influencing Dengue cases in Dhaka city: A model for Dengue prediction

The authors of [7] aim to predict yearly Dengue cases in Dhaka using climatic factors. They perform a log transformation on the dependent variable to normalize the data so that multiple linear regression is viable. The minimum temperature column was ignored due to high multicollinearity with the maximum temperature. Variance Inflation Factor (VIF) method which uses tolerance, was used to test the collinearity between independent variables while Durbin-Watson Test was used for testing the autocorrelation. The authors developed three models, one with no lag month, one with one lag month and one with two lag months. Lag was used as the mosquitoes take 45 days to mature and turn into adults. The best result was observed for the third model using area under ROC curve (AUC) to check accuracy, with the normal probability plot of residuals also satisfied the normality assumption. The authors concluded that humidity has the highest influence on the number of Dengue cases but in contrary to paper [n+1], Dengue case outbreak coincided with monsoon period but also states that other factors also contribute. Here too, data on bugology, vegetation and immune system of people need to be taken into account in order to establish correlations.

F. A Weather-Based Prediction Model of Malaria Prevalence in Amenfi West District, Ghana

The authors of [8] aimed to predict the incidence of malaria using variables like rainfall and temperature by performing time series analysis. The authors used monthly data of rainfall and temperature from 2002 to 2015 from hospital and meteorological office. Box Jenkins model was employed which takes into account past observations, errors and a random term to explain dependent variable. As ARIMAX couldn't capture the seasonality in malaria cases, the authors use SARIMA. Ljung Box Test was performed to test for the absence of serial autocorrelation and ARCH-LM was performed for homoscedasticity check. Log linear trend model was used based on the nature of the trend. This confirmed the upward trend. To investigate seasonality, additional seasonal dummy variables were used in the linear regression trend. Predictors included in the final model by the authors were min and max temp (excluding rainfall) which lagged at 3 months i.e, unit increase in max temp will cause increase in malaria incidence 3 months later in accordance with the ARIMA model. In conclusion, the researchers inferred that high rainfall did not result in increased malaria instantly contrary to other works. This lack of consistency in weather variables indicates complex interaction between weather variable and disease transmission. Moreover, there is non linear relationship between EIP and temperature. The paper hasn't incorporated vegetation (NDengue VirusI) to help in predicting malaria.

III. Problem Statement

In this study, we aim to analyse the effects of climate on the spread of Dengue Fever by employing various statistical techniques. Additionally, we aim to build a model that can forecast the number of Dengue Fever cases that will emerge in the following months, based on the climatic conditions and vegetation of the local region, thus enabling the local authorities to be well equipped. We model different regions separately as Dengue occurrences are hyper-local to the region and cannot be generalised.

IV. Dataset

The dataset of choice is from the DengAI competition hosted on www.drivendata.org [9]. The data for this competition comes from multiple sources aimed at supporting the "Predict the Next Pandemic Initiative". Dengue surveillance data is provided by the U.S. Centers for Disease Control and prevention [10], as well as the Department of Defense's Naval Medical Research Unit 6 [11] and the Armed Forces Health Surveillance Center [12], in collaboration with the Peruvian government [13] and U.S. universities. Environmental and climate data is provided by the National Oceanic and Atmospheric Administration (NOAA) [14], an agency of the U.S. Department of Commerce.

The Dataset consist of readings for the cities of San Juan and Iquitos. The features of the dataset are from multiple satellites as follows:

- NOAA's GHCN daily climate data weather station measurements
  - Maximum temperature
  - Minimum temperature
  - Average temperature
  - Total precipitation
  - Diurnal temperature range
- PERSIANN satellite precipitation measurements (0.25x0.25 degree scale)
  - Total precipitation
- NOAA's NCEP Climate Forecast System Reanalysis measurements (0.5x0.5 degree scale)
  - Total precipitation
  - Mean dew point temperature
  - Mean air temperature
  - Mean relative humidity
  - Mean specific humidity
  - Total precipitation
  - Maximum air temperature
  - Minimum air temperature

  - Average air temperature
  - Diurnal temperature range
- Satellite vegetation - Normalized difference vegetation index (NDVI) - NOAA's CDR Normalized Difference Vegetation Index (0.5x0.5 degree scale) measurements
  - Pixel southeast of city centroid
  - Pixel southwest of city centroid
  - Pixel northeast of city centroid
  - Pixel northwest of city centroid

V. Our Approach

The papers that we researched tried to predict the number of Dengue cases due to climatic variables only and did not consider the vegetation of the demography under consideration. We believe that the vegetation has a vital effect on the mosquito population and have decided to include it as one of the key predictors in our models. We noticed that most of the papers stuck to a single model without giving a lot of details about the comparative study between different models and the reason behind choosing a specific model. We present multiple models and compare and contrast their strengths and weaknesses. Based on the results obtained, we would be able to choose the optimal model for performing prediction.

```
city                                      0
year                                      0
weekofyear                                0
week_start_date                           0
ndvi_ne                                 194
ndvi_nw                                  52
ndvi_se                                  22
ndvi_sw                                  22
precipitation_amt_mm                     13
reanalysis_air_temp_k                    10
reanalysis_avg_temp_k                    10
reanalysis_dew_point_temp_k              10
reanalysis_max_air_temp_k                10
reanalysis_min_air_temp_k                10
reanalysis_precip_amt_kg_per_m2          10
reanalysis_relative_humidity_percent     10
reanalysis_sat_precip_amt_mm             13
reanalysis_specific_humidity_g_per_kg    10
reanalysis_tdtr_k                        10
station_avg_temp_c                       43
station_diur_temp_rng_c                  43
station_max_temp_c                       20
station_min_temp_c                       14
station_precip_mm                        22
dtype: int64
```

Fig. 1. Null values in the dataset

VI. Exploratory Data Analysis

From Figure 1. we see that there are a few null values. Handling Null values in a time series dataset is tricky and will be addressed during the modelling of the data.

Line plots of number of Dengue cases, Humidity and Temperature against time were plotted for both the cities as shown in Figure 2.

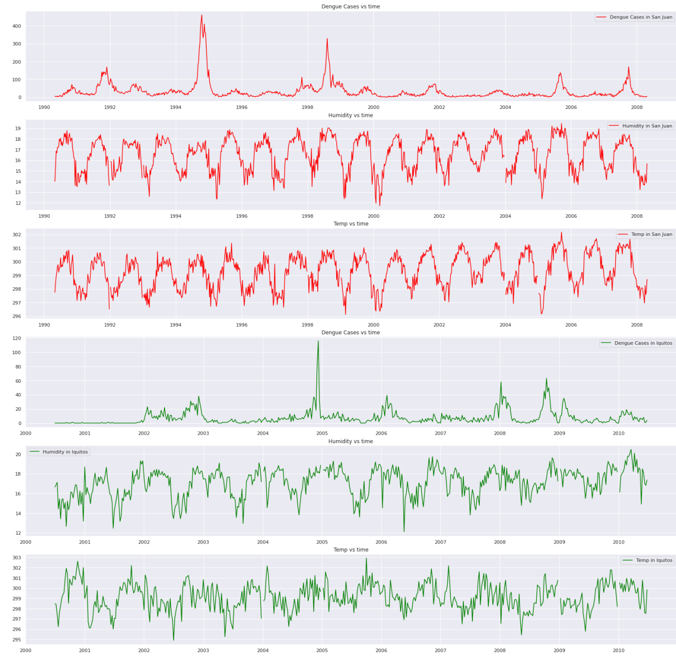

Fig. 2. Plots for various attributes against time

Figure 3. shows the distribution of various attributes such as total number of cases, specific humidity and average temperature.

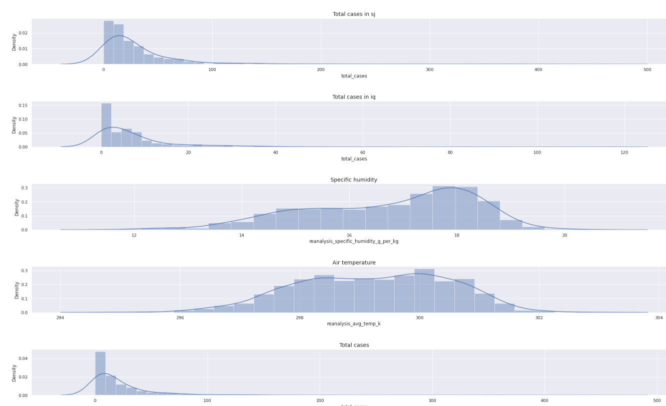

Fig. 3. Distribution of various attributes

Box plots of yearly Dengue cases in both the cities are shown in Figure 4. We see there are quite a few outliers. We will be considering these during our modelling phase.

Acknowledgment

We would like to express out profound gratitude to Dr. Gowri Srinivasa and the entire Data Analytics team, for encouraging and providing us with this opportunity to get hands-on experience in the field, and guiding us along the way. We would also like to thank the Computer Science and Engineering department at PES University, for always inspiring us to conduct frequent research and inculcating a problem-solving discipline in us.

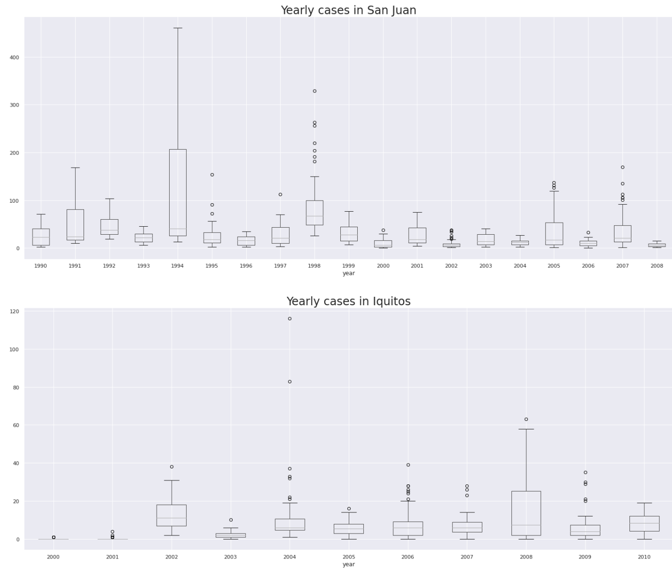

Fig. 4. Box plots of yearly Dengue cases

References

[1] C. Li, Y. Lu, J. Liu, and X. Wu, "Climate change and dengue fever transmission in china: evidences and challenges," *Science of the Total Environment*, vol. 622, pp. 493–501, 2018.

[2] C. Li, X. Wang, X. Wu, J. Liu, D. Ji, and J. Du, "Modeling and projection of dengue fever cases in guangzhou based on variation of weather factors," *Science of The Total Environment*, vol. 605, pp. 867–873, 2017.

[3] S. Sang, W. Yin, P. Bi, H. Zhang, C. Wang, X. Liu, B. Chen, W. Yang, and Q. Liu, "Predicting local dengue transmission in guangzhou, china, through the influence of imported cases, mosquito density and climate variability," *PloS one*, vol. 9, no. 7, p. e102755, 2014.

[4] P.-C. Wu, H.-R. Guo, S.-C. Lung, C.-Y. Lin, and H.-J. Su, "Weather as an effective predictor for occurrence of dengue fever in taiwan," *Acta tropica*, vol. 103, no. 1, pp. 50–57, 2007.

[5] Y. L. Hii, H. Zhu, N. Ng, L. C. Ng, and J. Rocklöv, "Forecast of dengue incidence using temperature and rainfall," *PLoS neglected tropical diseases*, vol. 6, no. 11, p. e1908, 2012.

[6] S. G. Kakarla, C. Caminade, S. R. Mutheneni, A. P. Morse, S. M. Upadhyayula, M. R. Kadiri, and S. Kumaraswamy, "Lag effect of climatic variables on dengue burden in india," *Epidemiology & Infection*, vol. 147, 2019.

[7] M. N. Karim, S. U. Munshi, N. Anwar, and M. S. Alam, "Climatic factors influencing dengue cases in dhaka city: a model for dengue prediction," *The Indian journal of medical research*, vol. 136, no. 1, p. 32, 2012.

[8] E. L. Darkoh, J. A. Larbi, and E. A. Lawer, "A weather-based prediction model of malaria prevalence in amenfi west district, ghana," *Malaria research and treatment*, vol. 2017, 2017.

[9] P. Bull, I. Slavitt, and G. Lipstein, "Harnessing the power of the crowd to increase capacity for data science in the social sector," *arXiv preprint arXiv:1606.07781*, 2016.

[10] "Centers for disease control and prevention." http://www.cdc.gov/.

[11] "US department of defense." https://www.defense.gov/.

[12] "Armed forces health surveillance center." https://health.mil/Military-Health-Topics/Combat-Support/Armed-Forces-Health-Surveillance-Branch.

[13] "Gobierno del perú." https://www.gob.pe/.

[14] "National oceanic and atmospheric administration." http://www.noaa.gov/.