Data Driven Dengue Dynamics via Satellite Data

Rafi Magdon-Ismail
The Albany Academy
Albany, NY, USA
rafimagdon@gmail.com

Thilanka Munasinghe
Information Technology and Web Science
Rensselaer Polytechnic Institute (RPI)
Troy, NY, USA
munast@rpi.edu

Jennifer C Wei
NASA Goddard Space Flight Center
Greenbelt, Maryland, USA
jennifer.c.wei@nasa.gov

Abstract—Vector-borne diseases like dengue remain a major global public health concern, with transmission strongly shaped by environmental factors. We examine how precipitation (rainfall), vegetation density, and temperature relate to dengue incidence across Sri Lanka using satellite-based data from NASA's Earth Observing System including the Global Precipitation Measurement (GPM) and Moderate Resolution Imaging Spectroradiometer (MODIS) derived Normalized Difference Vegetation Index (NDVI). We explored spatial and related statistical associations between weather conditions and dengue spread across the country. Results show substantial regional variability in dengue incidence, with urban districts such as Colombo reporting higher case counts despite low vegetation level. A negative association was observed between NDVI values and dengue cases, while rainfall in both daytime and nighttime land surface temperatures (average temperature) had a weaker relationship with dengue across districts. These findings demonstrate the potential of using satellite-derived environmental data for large-scale dengue outbreak surveillance and highlight the integration of climate indicators with public health datasets to support predictive modeling, as well as the implementation of Early Warning Systems (EWS) to improve dengue prevention strategies in Sri Lanka and other countries.

Index Terms—Dengue, NDVI, precipitation, temperature, NASA GPM, Earth observation, climate data, regression analysis, Sri Lanka, public health informatics.

I. INTRODUCTION

Vector-borne diseases, such as dengue fever, remain a persistent public health threat across many tropical regions [1]. This study explores how environmental variables, such as vegetation density and precipitation levels, affect dengue presence and spread in Sri Lanka. Using NASA Earth Observation datasets, we examine how NDVI, precipitation, and temperature correlate with reported dengue cases at the district level [2]. Because dengue outbreaks often follow environmental factors and weather patterns [1], understanding these relationships can provide a strong foundation for predictive modeling of outbreaks and increased public health awareness.

II. RELATED WORK

Research by Abbasi shows that changes in weather and environmental factors contribute to the spread of dengue fever caused by Aedes aegypti, a mosquito species [3]. Studies by Caminade et al. further explore this phenomenon, discussing not only the effects of weather and climate variations on infectious diseases but also the significant role they play in affecting human dynamics across the globe [4]. A study by

Sheela et al. discusses the importance of land use and land cover in the transmission of vector-borne diseases such as malaria and dengue [5]. Sirisena et al. discuss the climatic factors that impact the epidemiological patterns of dengue and the public health implications [1], [6]. Prabodanie et al. study the impact of wet zones (Colombo district) and dry zones (Batticaloa district) on dengue incidence [7]. Faruk et al. studied the seasonal and spatial variation in dengue outbreaks [8].

Earth observational environmental, weather, and climate data such as precipitation and land surface temperature, obtained from NASA, have been used in many other applications, such as those relating to wildfires, floods, and landslides, vector-borne disease spread, as well as energy infrastructure vulnerability, using statistical methods and machine learning techniques across diverse sectors, including agriculture, urban environments, and human mobility [9]–[20]. Similarly, in our exploratory study, we utilized NASA EO data to study the relationship between the dengue cases and environmental and weather-related variables.

III. METHODOLOGY

A. Data Sources

District-level dengue case data were obtained from the Sri Lanka Ministry of Health [21]. Environmental variables such as NDVI, precipitation, and Land Surface Temperature (LST) were obtained from the NASA Goddard Earth Science Data Information Services Center (GES DISC) [2], [22].

B. Data Processing and Integration

Data processing was done using Python libraries such as Pandas, GeoPandas, and NumPy during the preliminary data cleaning process. Environmental data was aggregated into monthly averages to ensure consistency with the dengue cases. Then, the monthly district-level environmental data was combined with the monthly dengue case counts and the dataframe was constructed for visulization and Exploratory Data Analysis (EDA). All visualizations, including heatmaps, scatterplots, and simple regressions, were produced using Python libraries such as Matplotlib [23].

C. Statistical and Computational Methods

We started the EDA process to check any outlier values and explore the correlations between variables, then conducted

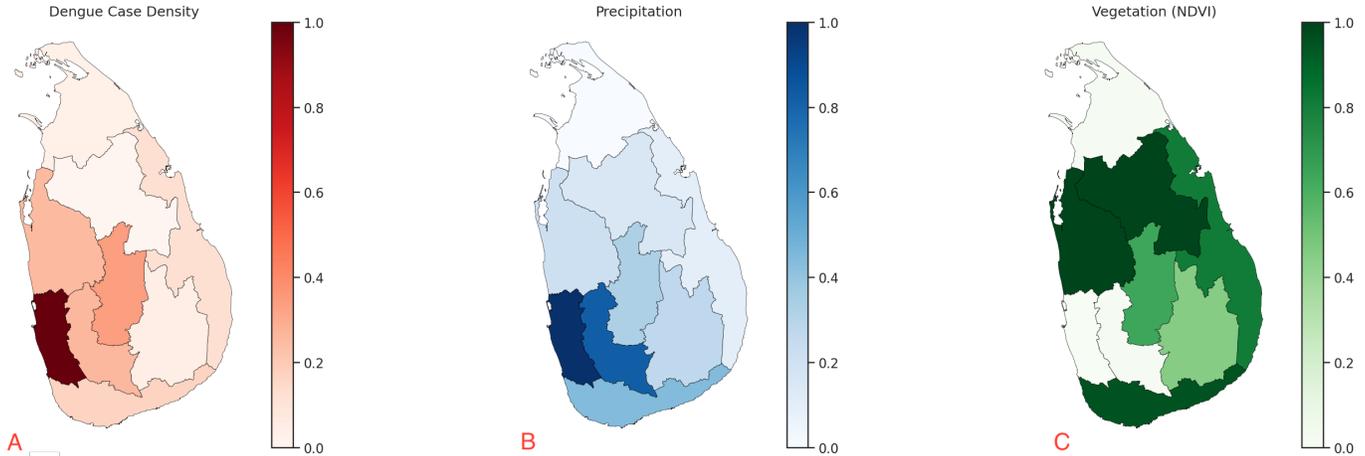**Sri Lanka — Spatial Heatmaps of Dengue and Environmental Variables**

Fig. 1. Spatial heatmaps of dengue case density (2013-2022) (A), precipitation (B), and vegetation density (NDVI) (C) across Sri Lanka. Color bars represent normalized values for each variable; darker color indicates higher values. Data were plotted using GeoPandas and Matplotlib.

linear regression [24]. Spearman correlation coefficients were calculated as shown in the Figure 4. We paid attention to descriptive and summary statistics to obtain insights from this exploratory analysis work [25].

IV. RESULTS

Our exploratory data analysis revealed substantial regional variation in dengue incidence across Sri Lanka. Using district-level dengue data obtained from the Sri Lanka Ministry of Health website (available for public domain access) and environmental weather data from NASA's MODIS and GPM, we regressed the dengue case with NDVI, precipitation, and temperature values to examine the relationships among these environmental factors.

Figure 1 (A) presents the case densities scaled from 0 to 1. (A) shows the dengue case density, (B) the precipitation, and (C) NDVI values over the nine administrative provinces of Sri Lanka. The subplot (A) showing the dengue densities indicates that the highest case densities occurred in the Western Province (particularly around Colombo, the capital of Sri Lanka) and its neighboring provinces, which aligns with the highest urban density levels that have reported the highest dengue transmission cases [26], [27]. The precipitation (rainfall) subplot (B) shows that the highest average rainfall during 2013-2022 occurred in the south-western region of Sri Lanka. Typically, this south-western part of Sri Lanka is subject to "South-Western" severe monsoon rain [28]. We hypothesize that excessive rainfall may contribute to mosquito outbreaks by creating more breeding sites, and we plan to test this hypothesis in future studies. The subplot (C) shows the vegetation across 9 provinces. The vegetation value and dengue cases over the provinces show a negative correlation. Our exploratory analysis shows that highly urban provinces such as the Western Province have the highest number of dengue cases with less vegetation compared to the central

and eastern provinces of Sri Lanka. We generated the scatter plots shown in the Figure 3 and 2 for Average Dengue Cases VS precipitation and vegetation. We observed a weaker correlation between average temperature and dengue incidence than between vegetation and precipitation (rainfall) values.
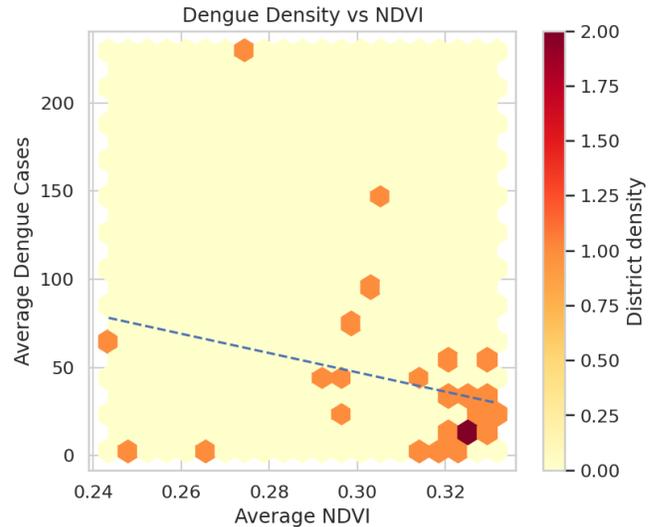

Fig. 2. Scatterplot of average NDVI versus average dengue incidence (2013-2022) across Sri Lankan districts. Hexagons represent individual district densities.

The scatterplot in Figure 2 shows a clear negative relationship between average vegetation and dengue incidence. Each district is represented by a hexagonal marker on the plot, with darker shades indicating higher population density in that district. Districts with lower vegetation (NDVI values) consistently reported higher dengue case counts, leading us to conduct further investigation to find out the potential link between vegetation loss and urban growth versus increased
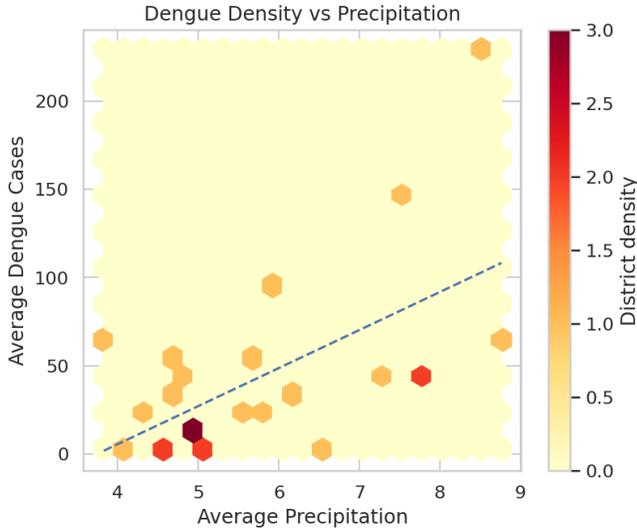
dengue presence in those districts.

Fig. 3. Scatterplot of average precipitation versus average dengue incidence (2013-2022) across Sri Lankan districts.

Figure 3 shows a positive association between precipitation and dengue incidence across 25 districts in Sri Lanka. Our exploratory analysis shows that, during the period 2013 to 2022, the highest average rainfall was reported in districts in the southwestern region. Also, the Western Province has the highest population of Sri Lanka [29].

To provide an overview of the relationships among dengue incidence and environmental factor variables, a correlation matrix was computed using Spearman's correlation coefficient method [30]. We observed weaker positive and negative correlations coefficient values as shown in Figure 4.
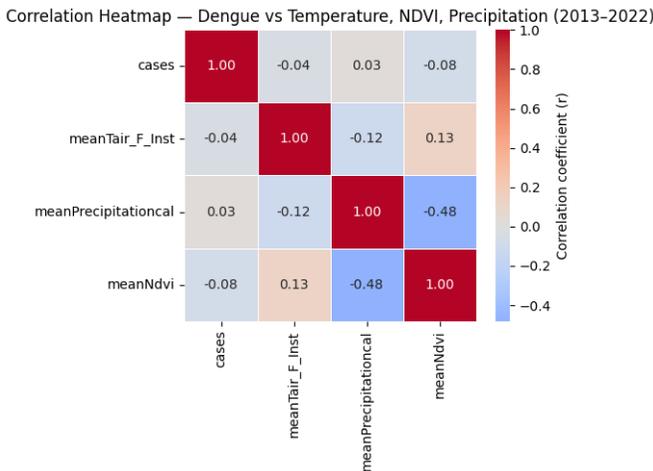

Fig. 4. Spearman correlation coefficient-based heatmap showing relationships among dengue incidence, average temperature, precipitation, and NDVI across Sri Lankan districts over the period of 2013 to 2022.

This study revealed weaker, monotonic correlations with satellite-derived environmental variables.

As we observed a non-linear relationship among these environmental variables, indicating that other hidden variables (latent variables) we did not use in this study, such as human mobility and migration over a decade across districts, may be playing a significant role in changes in dengue incidence across districts.

Therefore, while the exploratory study captures the interaction between weather and environmental variables, it does not fully capture the complex nature of dengue outbreak patterns in Sri Lanka. Additionally, these nonlinear relationships are much more evident when examined using Locally-Weighted-Regression Scatter-Plot Smoothing (LOWESS) methods [31], such as the one shown in Figure 5.
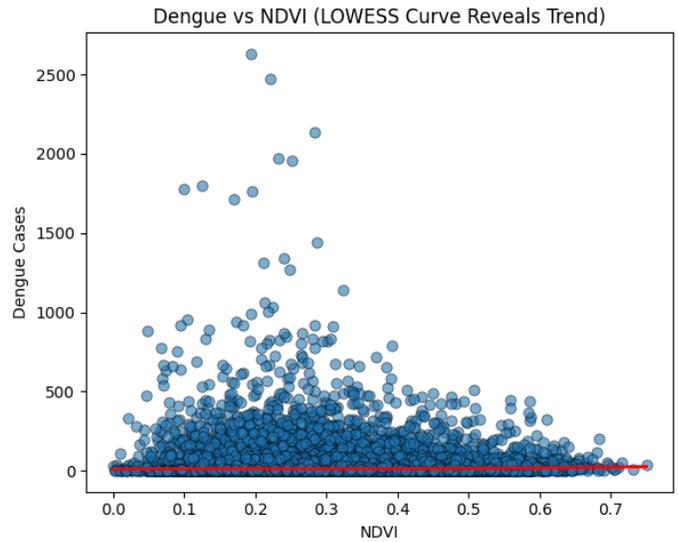

Fig. 5. LOWESS-smoothed relationship between NDVI and dengue incidence (2013-2022) across Sri Lankan districts. LOWESS smoothing followed Cleveland's method.

Figure 5 shows the nonlinear relationship between NDVI and dengue incidence using a LOWESS method. Additional exploratory data analysis-related plots for this study are available in our GitHub repository for reference.

V. CONCLUSION

This study demonstrates the potential of combining epidemiological data with satellite-derived environmental and weather indicators to explore dengue incidence across the provinces of Sri Lanka [17]. By combining district-level dengue case counts obtained from the Sri Lanka Ministry of Health, we used vegetation, rainfall, and temperature data from the NASA during this exploratory study. We observed that the averaged environmental variables show weaker correlations with dengue incidence.

Our study shows substantial regional variability in dengue incidence across the provinces of Sri Lanka, with urban districts such as Colombo, which have the highest population density, reporting higher dengue case counts despite low vegetation levels. The study revealed a negative association between NDVI values and dengue cases. At the same time,

rainfall showed a weak positive correlation, and average temperature had a weaker relationship with dengue across the districts of Sri Lanka. These findings demonstrate the potential of using satellite-derived environmental data for large-scale dengue outbreak surveillance and highlight the potential of integrating climate and weather-related indicators with public health datasets to implement predictive models, as well as the use of Early Warning Systems (EWS) to improve dengue prevention strategies in Sri Lanka.

REFERENCES

[1] P. Sirisena, F. Noordeen, H. Kurukulasuriya, T. A. Romesh, and L. Fernando, "Effect of climatic factors and population density on the distribution of dengue in sri lanka: a gis based evaluation for prediction of outbreaks," *PloS one*, vol. 12, no. 1, p. e0166806, 2017.

[2] Z. Liu, D. Ostrenga, B. Vollmer, B. Deshong, K. Macritchie, M. Greene, and S. Kempler, "Global precipitation measurement mission products and services at the nasa ges disc," *Bulletin of the American Meteorological Society*, vol. 98, no. 3, pp. 437–444, 2017.

[3] E. Abbasi, "The impact of climate change on travel-related vector-borne diseases: a case study on dengue virus transmission," *Travel Medicine and Infectious Disease*, p. 102841, 2025.

[4] C. Caminade, K. M. McIntyre, and A. E. Jones, "Impact of recent and future climate change on vector-borne diseases," *Annals of the New York Academy of Sciences*, vol. 1436, no. 1, pp. 157–173, 2019.

[5] A. Sheela, A. Ghermandi, P. Vineetha, R. Sheeja, J. Justus, and K. Ajayakrishna, "Assessment of relation of land use characteristics with vector-borne diseases in tropical areas," *Land Use Policy*, vol. 63, pp. 369–380, 2017.

[6] P. Sirisena and F. Noordeen, "Evolution of dengue in sri lanka—changes in the virus, vector, and climate," *International Journal of Infectious Diseases*, vol. 19, pp. 6–12, 2014.

[7] R. R. Prabodanie, S. Schreider, B. Cazelles, and L. Stone, "Coherence of dengue incidence and climate in the wet and dry zones of sri lanka," *Science of the Total Environment*, vol. 724, p. 138269, 2020.

[8] M. Faruk, S. Jannat, and M. S. Rahman, "Impact of environmental factors on the spread of dengue fever in sri lanka," *International Journal of Environmental Science and Technology*, vol. 19, no. 11, pp. 10 637–10 648, 2022.

[9] V. Acharya, A. Ghosh, I. Kang, T. Munasinghe, and K. Binita, "Landslide likelihood prediction using machine learning algorithms," in *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2022, pp. 5395–5403.

[10] Z. Sun, T. ten Brink, W. Carande, G. Koren, N. Cristea, C. Jorgenson, B. Janga, G. P. Asamani, S. Achan, M. Mahoney *et al.*, "Towards practical artificial intelligence in earth sciences," *Computational Geosciences*, vol. 28, no. 6, pp. 1305–1329, 2024.

[11] S. Nangia, T. Munasinghe, H. Tubbs, and A. Anyamba, "From satellites to fields: Machine learning applications for prediction of corn production using ndvi, precipitation and land surface temperature for large producer countries," in *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 2023, pp. 3479–3488.

[12] T. Zhu, J. Wu, T. Tan, S. Wang, T. Munasinghe, H. Tubbs, and A. Anyamba, "Integrating climate variable data in machine learning models for predictive analytics of tomato yields in california," in *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 2023, pp. 3538–3546.

[13] N. Thelasingha and T. Munasinghe, "Energy infrastructure risk modeling using quantum and classical machine learning," in *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 2024, pp. 4611–4620.

[14] B. Hegyi, P. W. Stackhouse, P. Taylor, and F. Patadia, "Nasa power: providing present and future climate services based on nasa data for the energy, agricultural, and sustainable buildings communities," in *104th American Meteorological Society (AMS) Annual Meeting*, 2024.

[15] T. Munasinghe, K. A. Cornell, J. C. Wei, G. Berg, and J. Hendler, "A knowledge graph framework for organizing heterogeneous datasets for utilization in classical and quantum computing: Current challenges and future directions," in *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 2024, pp. 8781–8785.

[16] T. Munasinghe and A. Svirsky, "Engaging students in data literacy: Lessons learned from data intensive classrooms," in *Companion Publication of the 13th ACM Web Science Conference 2021*, 2021, pp. 40–43.

[17] J. Weng, D. Qiu, E. Cruz, M. Magdon-Ismail, T. Munasinghe, J. C. Wei, A. Pathirana, and M. Pasarkar, "Graph representation learning for dengue forecasting," in *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 2024, pp. 4448–4456.

[18] K. Bhanot, D. Schroeder, I. Llewellyn, N. Luczak, and T. Munasinghe, "Dengue spread information system (dsis)," in *Proceedings of the 4th International Conference on Medical and Health Informatics*, 2020, pp. 150–159.

[19] M. Pasarkar, J. Kim, E. O'Gara, A. Zhang, M. Magdon-Ismail, T. Munasinghe, J. Weng, D. Qiu, E. Cruz, J. C. Wei *et al.*, "Natural language processing for extracting rich disease data aligned to satellite meteorological data," in *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 2024, pp. 4394–4401.

[20] L. Ngweta, K. Bhanot, A. Maharaj, I. Bogle, and T. Munasinghe, "Identifying the relationship between precipitation and zika outbreaks in argentina," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 6163–6165.

[21] Sri Lanka Ministry of Health, Epidemiology Unit, "Weekly epidemiological reports: Dengue surveillance data," https://www.epid.gov.lk, 2025, accessed: 2025-11-23.

[22] J. Acker, R. Soebiyanto, R. Kiang, and S. Kempler, "Use of the nasa giovanni data system for geospatial public health research: example of weather-influenza connection," *ISPRS International Journal of Geo-Information*, vol. 3, no. 4, pp. 1372–1386, 2014.

[23] N. Ari and M. Ustazhanov, "Matplotlib in python," in *2014 11th International Conference on Electronics, Computer and Computation (ICECCO)*. IEEE, 2014, pp. 1–6.

[24] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American statistical association*, vol. 74, no. 368, pp. 829–836, 1979.

[25] J. W. Tukey, F. Mosteller, and D. C. Hoaglin, *Fundamentals of exploratory analysis of variance*. Wiley Online Library, 1991.

[26] G. N. Malavige, C. Jeewandara, A. Ghouse, G. Somathilake, and H. Tissera, "Changing epidemiology of dengue in sri lanka—challenges for the future," *PLoS neglected tropical diseases*, vol. 15, no. 8, p. e0009624, 2021.

[27] H. A. Tissera, B. D. Jayamanne, R. Raut, S. M. Janaki, Y. Tozan, P. C. Samaraweera, P. Liyanage, A. Ghouse, C. Rodrigo, A. M. de Silva *et al.*, "Severe dengue epidemic, sri lanka, 2017," *Emerging infectious diseases*, vol. 26, no. 4, p. 682, 2020.

[28] E. Ranatunge, B. Malmgren, Y. Hayashi, T. Mikami, W. Morishima, M. Yokozawa, and M. Nishimori, "Changes in the southwest monsoon mean daily rainfall intensity in sri lanka: relationship to the el niño–southern oscillation," *Palaeogeography, Palaeoclimatology, Palaeoecology*, vol. 197, no. 1-2, pp. 1–14, 2003.

[29] K. Weerakoon, "Overview of urbanization in geodemographic approach; colombo urban area, sri lanka," *European Academic Research*, vol. 1, no. 9, pp. 2819–31, 2013.

[30] J. C. De Winter, S. D. Gosling, and J. Potter, "Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data." *Psychological methods*, vol. 21, no. 3, p. 273, 2016.

[31] G. W. Moran, "Locally-weighted-regression scatter-plot smoothing (lowess): a graphical exploratory data analysis technique." Ph.D. dissertation, Monterey, California. Naval Postgraduate School, 1984.