Development and Use of a Reproducible Framework for Spatiotemporal Climatic Risk Assessment and its Association with Decadal Trend of Dengue in India

Gurpreet Singh, Arun Mitra, Biju Soman

Achutha Menon Centre for Health Science Studies, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Thiruvananthapuram, Kerala, India

Abstract

Introduction: The study aimed to develop a reproducible, open-source, and scalable framework for extracting climate data from satellite imagery, understanding dengue's decadal trend in India, and estimating the relationship between dengue occurrence and climatic factors. **Materials and Methods:** A framework was developed in the Open Source Software, and it was empirically tested using reported annual dengue occurrence data in India during 2010–2019. Census 2011 and population projections were used to calculate incidence rates. Zonal statistics were performed to extract climate parameters. Correlation coefficients were calculated to estimate the relationship of dengue with the annual average of daily mean and minimum temperature and rainy days. **Results:** Total 818,973 dengue cases were reported from India, with median annual incidence of 6.57 per lakh population; it was high in 2019 and 2017 (11.80 and 11.55 per lakh) and the Southern region (8.18 per lakh). The highest median annual dengue incidence was observed in Punjab (24.49 per lakh). Daily climatic data were extracted from 1164 coordinate locations across the country for the decadal period (4,249,734 observations). The annual average of daily temperature and rainy days positively correlated with dengue in India ($r = 0.31$ and 0.06, at $P < 0.01$ and 0.30, respectively). **Conclusion:** The study provides a reproducible algorithm for bulk climatic data extraction from research-level satellite imagery. Infectious disease models can be used to understand disease epidemiology and strengthen disease surveillance in the country.

Keywords: Climate risk, dengue, public health, remote sensing, reproducible approach, satellite imagery, spatiotemporal

Introduction

Spatiotemporal and machine learning approaches are increasingly used to understand the epidemiology of infectious diseases.[1] The epidemiological understanding gained using these approaches has been instrumental in developing decision support tools, early warning systems, aberration detection algorithms, disease forecasting models, and evidence-informed public health decision-making.[2-4] Implementation of Integrated Health Information Portal, deregulation of geospatial data by Department of Science and Technology, National Digital Health Mission, and other digital health initiatives will generate high-resolution geocoded big data on health-related events in India in the coming years.[5-7] Existing routine datasets have also been used to understand micro-climatic determinants using algorithms that can extract spatiotemporal parameters associated with disease occurrence.[2]

The development of infectious disease models in low-and middle-income countries is faced with challenges of obtaining high-resolution data on climatic risk variation from on-ground meteorological stations. Global and National intersectoral initiatives provide satellite imagery-based Analysis Ready Datasets (ARDs) and global climatic models through multiple sources.[8-11] The use of these ARDs will enable public health managers and epidemiologists to obtain high-resolution climatic data,

Address of the correspondence: Dr. Biju Soman,
Achutha Menon Centre for Health Science Studies, Sree
Chitra Tirunal Institute for Medical Sciences and Technology,
Thiruvananthapuram, Kerala, India.
E-mail: bijusoman@sctimst.ac.in

Access this article online

Quick Response Code:

Website:
www.ijcm.org.in

DOI:
10.4103/ijcm.ijcm_862_21

providing a future opportunity to strengthen existing disease surveillance.

Dengue is hyperendemic in India, and resultant economic losses have surpassed other vector-borne diseases.[12] The occurrence of dengue is critically determined by the microclimatic conditions.[13,14] Satellite imagery ARDs and preprocessed climatic models are routine data sources on microclimatic conditions which can be modeled for dengue analytics.[15] The incorporation of lagged climatic variables and spatial characteristics in such models establishes temporality as defined in Hill's criteria and adheres to Tobler's law in geography.

Satellite imagery ARDs are large datasets commonly available in Hierarchical Data Formats (HDF), network Common Data Form (NetCDF), and other data formats (Application Programming Interface [API] based).[9-11,16] Moderate Resolution Imaging Spectroradiometer provides ARDs in HDF; Integrated Multi-satellitE Retrievals for Global Precipitation Mission (IMERG) datasets, and Indian Meteorological Department (IMD) in NetCDF format; and Modern-Era Retrospective analysis for Research and Applications, Version 2, Meteorological and Oceanographic Satellite Data Archival Centre, Bhuvan web portal, and Open Government Data Platform India are API-based routine geospatial data sources. Handling large datasets in a reproducible environment increases the grade of evidence, reduces manual errors, and is computationally efficient.[17] Thus, the present study was conducted to explore and develop a reproducible framework for extracting spatiotemporal climatic risk parameters from satellite imagery ARDs, understand the decadal trend of dengue in India, and estimate the relationship between dengue occurrence and climatic factors in India.

MATERIALS AND METHODS

Study design

The study was carried out in two phases. The first phase included exploring and developing a reproducible framework for research-level satellite imagery bulk preprocessing. The second phase included ecological analysis of publicly available dengue occurrence data and climatic variables obtained using the developed framework.

Exploration and reproducible framework development

Algorithms provided by Level-1 and Atmosphere Archive and Distribution System Distributed Active Archive Center, IMD Gridded datasets archive, Global Precipitation Mission, R package archives, GitHub, and other code repositories were explored. Proprietary software-based algorithms and algorithms for platforms other than the R environment were excluded. Framework for HDF, NetCDF, and API-based satellite imagery ARDs extraction into analyzable tidy data formats was developed.

Secondary data sources

Annual state-wise dengue occurrence data for the decadal period from January 01, 2010 to December 31, 2019, was extracted from the National Health Profile reports and National Vector Borne Disease Control Programme, India website.[18,19] Population estimates from Census 2011 and population projections for the year 2012–2019 provided population denominators for calculating dengue incidence per lakh population.[20] Climatic variables (temperature (mean and minimum) and cumulative precipitation) for daily timestamps were extracted using the "*nasapower*" package.[21]

Data analysis and interpretation

The National-, regional-, and state-level decadal trend of dengue was calculated. For regional level analysis, the zonal councils as defined by the Ministry of Home Affairs were adopted.[22] Zonal statistics were performed to calculate climate parameters. Descriptive measures were calculated for climatic variables. Data visualization using the GIS environment in an open-source platform was carried out. Correlation coefficients were calculated to estimate the relationship of dengue with mean annual temperature and rainy days. A $P < 0.05$ was considered statistically significant. The framework development and statistical analysis were carried out using R version 4.0.3 (R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria).[23]

Ethics statement

The present study is part of a larger research project culminating in the Ph.D. program of the first author. Institutional Ethics Committee (IEC/IEC-1653; IEC Reg. No. ECR/189/Inst/KL/2013/RR-16) clearance obtained vide letter SCT/IEC/IEC-1653/DECEMBER-2020 dated 19/12/2020.

RESULTS

Framework

The algorithm undertakes bulk data extraction of climate parameters for a multi-polygon from stored NetCDF/HDF files. The files should be downloaded according to the instructions given on respective websites and stored in a file directory. All the steps for data extraction are automated in the algorithm based on user inputs on the path of the directory where NetCDF/HDF files are stored. The researcher/user should provide identification of desired sub dataset, scale factor, and offsets if any. For API-based extraction, a local grid with spatial resolution as of the data source is constructed for data extraction. The framework for data extraction from satellite imagery ARDs (NetCDF, HDF, and API-based) can be accessed on the GitHub repository.

Epidemiological trend of dengue in India

During the decadal period, 8,18,973 dengue cases were reported with a mean (standard deviation) annual incidence of 6.36 (3.60) per lakh population. The median annual dengue incidence for India was 6.57 per lakh population. Nationally, dengue incidence was maximum in 2019 followed by 2017 (11.80 and 11.55 per lakh, respectively), and minimum dengue incidence was in 2011 (1.56 per lakh).

Regionally, the highest median annual dengue incidence was observed in the South, followed by the West, North, North East, Central, and East region (8.18, 8.05, 4.5, 1.89, 1.62, and 1.6 per lakh, respectively). Among the states, the highest median annual dengue incidence was observed in Punjab, Goa, Kerala, and Odisha (24.49, 14.41, 12.13, and 9.1 per lakh, respectively). The Union Territories showed higher dengue incidence rates with the highest median annual incidence reported from Dadar and Nagar Haveli (126.22 per lakh), followed by Puducherry (77.45 per lakh). The national capital, Delhi, reported a median annual incidence of 28.70 per lakh population. Lakshadweep was the only state/UT with zero reported cases during the decadal period. Further, among the states, the outbreak years, as indicated by unusually high (more than 50 per lakh) dengue incidence were reported from the states of Arunachal Pradesh (134 per lakh in 2015), Uttarakhand (95 per lakh in 2019), Sikkim (66 per lakh in 2019), Goa (64 per lakh in 2019), Himachal Pradesh (64 per lakh in 2018), Kerala (57 per lakh in 2017), and Punjab (52 per lakh in 2017). The highest dengue incidence among union territories was reported from Dadar and Nagar Haveli (921 per lakh in 2016 and 427 per lakh in 2017), followed by Puducherry (318 per lakh in 2017 and 274 per lakh in 2012).

Climatic trends in India

Daily climatic data were extracted from 1164 coordinate locations across the country for the decadal study period (4,249,734 observations). The regional summary of decadal temperature is represented in Figure 1. The West, South, Central, and East regions of the country were warmer (decadal mean temperature of 26.31, 26.22, 26.31, and 25.41°C respectively) compared to North and Northeast regions (decadal mean temperature of 18.71 and 19.47°C, respectively). The temperature variation was maximum in the North region (IQR 17.37) and minimum in the South (IQR 3.78). The highest decadal mean rainfall was present in the Northeast region, followed by the South and East regions (75.53, 67.04, and 62.66 mm, respectively).

Correlation between dengue occurrence and climatic variables

The correlation between climatic variables and dengue is represented in Table 1. The annual average daily mean temperature was positively correlated with dengue at the national level ($r = 0.31$, $P < 0.01$). At the regional level, the correlation between mean temperature and dengue was maximum in West, North, and Central regions ($r = 0.43$, 0.37, and 0.35, $P = 0.02$, < 0.01, and 0.13 respectively). The annual average of daily minimum temperature was significantly correlated with dengue in the East and Northeast regions ($r = 0.33$ and-0.32, $P = 0.04$ and < 0.01, respectively). The precipitation days were positively correlated with the dengue at the national level ($r = 0.06$, $P = 0.30$). At the regional level, the East and Northeast regions had a statistically significant relationship between precipitation days and dengue ($r = 0.38$ and 0.28, respectively, $P = 0.02$).

DISCUSSION

The present study documents availability of high-resolution satellite imagery research-level datasets and provide a reproducible algorithm for bulk data extraction and preprocessing of these datasets. Availability of micro-climatic data enables the development of models for understanding knowledge gaps in infectious disease epidemiology.[1,13,14,24] Advances in technology and increasing geocoded health data generation provide a challenge and an opportunity for the growth of epidemiological theories. Digital healthcare epidemiology, as compared to conventional epidemiology, is based on routine unstructured big datasets and requires a data science approach.[25] Research with reproducible open-source algorithms facilitates understanding of the research pathways and enables future expansion of existing frameworks.[17,26]

Satellite remote sensing has increased manifold in the past few decades in technology and application potential. High-resolution and multi-frequency satellite sensors can capture data on multiple climatic and environmental parameters, among others.[27] A validation study of the IMERG rainfall dataset with IMD gridded data showed a correlation of $+ 0.88$ in India.[28] It is also essential to understand that raw satellite imagery datasets have inherent data quality issues and require technical proficiency for preprocessing. Thus, the availability of research-level datasets from domain expert teams helps public health professionals and epidemiologists to estimate the spatiotemporal variation of risk factors in disease causation.

The decadal dengue trend in India showed an increase across the country. This may be attributed to an actual increase over the decadal period and enhanced diagnostics, surveillance, and reporting mechanisms in the country. The correlation of climatic factors was found to be varying across regions in the country. It may be attributed to the large geographical extent and presence of multiple climatic zones. Temperature between 16-30 degrees Celsius is optimal for dengue transmission.[29] Precipitation provides water habitat for immature stages in the mosquito life cycle; however, high precipitation leading to flushing of immature stages is likely to have a negative association with dengue occurrence. In a study carried out to assess climatic factors and dengue occurrence in Thailand, different climatic factors were found to be associated with dengue incidence in coastal areas and plains.[30] Further studies at a more granular level (district/sub-district) are required to understand micro-climatic risk variation and its association with dengue in India.

The limitations in the present study include the lack of availability of granular dengue occurrence data. Data with a higher spatial and temporal resolution of disease occurrence would have further enhanced the understanding of the spatiotemporal epidemiology of dengue and its microclimatic associations. Furthermore, higher resolution data is required to understand the variance in these associations as per topography. These were beyond the scope of the present study. The role

of bio-eco-social determinants on the association of climatic factors with dengue occurrence was not studied in the present study. Incorporation of the same will enable the development of forecasting models to strengthen disease surveillance. The strength of the present study was the novel approach of using satellite imagery data to estimate the association between climatic factors and decadal dengue trends at national, regional, and state levels in India and the ability of the reproducible algorithm to process 4.2 million observations capturing daily climatic variables over a decade in a reproducible manner. The algorithms developed can be utilized in understanding the epidemiology of diseases affected by climatic conditions. The algorithm, being open-source and scalable, can be expanded to include additional satellite datasets in the future.

Collaborative studies between health departments and academic institutions with granular dengue surveillance data need to be conducted for understanding micro-climatic associations of dengue. Further, additional covariates such as climatic, environmental, sociodemographic, behavioral, and health system characteristics should be incorporated to understand the complex interplay of factors associated with dengue transmission. This understanding will enable us to develop efficient disease prevention and control strategies in the country.

CONCLUSION
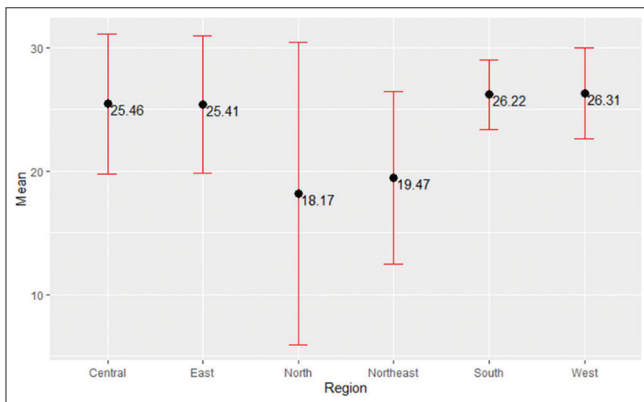
The present study documents and provides a reproducible,

Figure 1: Region-wise decadal temperature distribution in India

systematic algorithm for spatiotemporal climatic risk assessment using research-level satellite imagery datasets. Further, the study highlights heterogenous high dengue burden in the country associated with climatic factors. The data science approach for spatiotemporal modelling of dengue incorporating climatic variables has the potential to develop forecasting models for strengthening routine surveillance in the country.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. An Overview of GeoAI Applications in Health and Healthcare | International Journal of Health Geographics | Full Text. Available from: https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-019-0171-2. [Last accessed on 2021 Feb 20].
2. Hung YW, Hoxha K, Irwin BR, Law MR, Grépin KA. Using routine health information data for research in low- and middle-income countries: A systematic review. BMC Health Serv Res 2020;20:790.
3. Carvajal TM, Viacrusis KM, Hernandez LF, Ho HT, Amalin DM, Watanabe K. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. BMC Infect Dis 2018;18:183.
4. Shi Y, Liu X, Kok SY, Rajarethinam J, Liang S, Yap G, *et al.* Three-month real-time dengue forecast models: An early warning system for outbreak alerts and policy decision support in Singapore. Environ Health Perspect 2016;124:1369-75.
5. IHIP-Integrated Health Information Platform. Available from: https://idsp.nhp.gov.in/#!/. [Last accessed on 2021 Feb 20].
6. Department of Science and Technology. Guidelines for Acquiring and Producing Geospatial Data and Geospatial Data Services including Maps; DST F.No.SM/25/02/2020. Available from https://dst.gov.in/sites/default/files/Final%20Approved%20Guidelines%20on%20Geospatial%20Data.pdf [Last accessed on 2021 Feb 20].
7. NDHM. Available from: https://ndhm.gov.in/. [Last accessed on 2021 Feb 19].
8. Open Government Data (OGD) Platform India. Open Government Data (OGD) Platform India. Available from: https://data.gov.in/. [Last accessed on 2021 Feb 20].
9. IMERG: Integrated Multi-satellitE Retrievals for GPM | NASA Global Precipitation Measurement Mission. Available from: https://gpm.nasa.gov/data/imerg. [Last accessed on 2021 Feb 19].
10. MODIS Web. Available from: https://modis.gsfc.nasa.gov/data/dataprod/. [Last accessed on 2021 Feb 19].
11. National Remote Sensing Centre. Bhuvan. Indian Geo-Platform of

Table 1: Correlation between dengue occurrence and climatic factors

Region	Mean temperature (°C)		Minimum temperature (°C)		Precipitation (days)	
	Correlation coefficient (95% CI)	P	Correlation coefficient (95% CI)	P	Correlation coefficient (95% CI)	P
India	0.31 (0.20-0.41)	<0.01	−0.07 (−0.19-0.04)	0.22	0.06 (−0.05-0.18)	0.30
North	0.37 (0.17-0.55)	<0.01	0.14 (−0.08-0.35)	0.21	−0.04 (−0.26-0.18)	0.71
South	0.12 (−0.19-0.42)	0.44	0.06 (−0.25-0.36)	0.71	0.14 (−0.17-0.44)	0.37
East	0.19 (−0.14-0.49)	0.26	0.33 (0.0-0.6)	0.04	0.38 (0.05-0.64)	0.02
West	0.43 (0.09-0.69)	0.02	0.17 (−0.20-0.50)	0.37	−0.16 (−0.49-0.21)	0.39
Central	0.35 (−0.10-0.68)	0.13	0.03 (−0.42-0.46)	0.91	−0.38 (−0.7-0.07)	0.09
Northeast	0.13 (−0.12-0.36)	0.31	−0.32 (−0.53-−0.08)	<0.01	0.28 (0.03-0.49)	0.02

CI: Confidence interval

ISRO; 2020. Available from: https://bhuvan-app3.nrsc.gov.in/data/download/#. [Last accessed on 2020 Dec 27].

12. World Health Organization. Dengue and Severe Dengue; June 23, 2020. Available from: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. [Last accessed on 2021 Jan 29].

13. Fan J, Wei W, Bai Z, Fan C, Li S, Liu Q, *et al.* A systematic review and meta-analysis of dengue risk with temperature change. Int J Environ Res Public Health 2014;12:1-15.

14. Morin CW, Comrie AC, Ernst K. Climate and dengue transmission: Evidence and implications. Environ Health Perspect 2013;121:1264-72.

15. Louis VR, Phalkey R, Horstick O, Ratanawong P, Wilder-Smith A, Tozan Y, *et al.* Modeling tools for dengue risk mapping – A systematic review. Int J Health Geogr 2014;13:50.

16. Rosenzweig C, Horton RM, Bader DA, Brown ME, DeYoung R, Dominguez O. *et al*. Enhancing climate resilience at NASA centers: A collaboration between science and stewardship. Bull. Amer. Meteorol. Soc., 2014;95:1351-63, doi:10.1175/BAMS-D-12-00169.1.

17. Peng RD. Reproducible research and Biostatistics. Biostatistics 2009;10:405-8.

18. National Health Profile: Central Bureau of Health Intelligence. Available from: https://www.cbhidghs.nic.in/index1.php?lang=1&level=1&sublinkid=75&lid=1135. [Last accessed on 2021 Feb 20].

19. Directorate General of Health Services, Ministry of Health and Family Welfare, Government of India. DENGUE/DHF SITUATION IN INDIA. National Vector Borne Disease Control Programme; 2021. Available from: https://nvbdcp.gov.in/index4.php?lang=1&level=0&linkid=431&lid=3715. [Last accessed on 2021 Jan 30].

20. National Commission on Population. Population Projections for India and States 2011 – 2036. Ministry of Health and Family Welfare, Govt of India; 2019. Available from: https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf. [Last accessed on 2021 Feb 20].

21. Sparks A. Nasapower: A NASA POWER global meteorology, surface solar energy and climatology data client for R. JOSS 2018;3:1035.

22. Ministry of Home Affairs, Government of India. Zonal Council. Ministry of Home Affairs; 2017. Available from: https://www.mha.gov.in/zonal-council. [Last accessed on 2021 Feb 18].

23. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/. [Last accessed on 2021 Feb 18].

24. Wang X, Tang S, Wu J, Xiao Y, Cheke RA. A combination of climatic conditions determines major within-season dengue outbreaks in Guangdong Province, China. Parasit Vectors 2019;12:45.

25. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol 2019;188:2222-39.

26. Hemant P. Reproducible Machine Learning. Medium, 2020. Available from: https://towardsdatascience.com/reproducible-machine-learning-cf1841606805. [Last accessed on 2021 Feb 20].

27. Manikiam B. Satellite based climate change study. Vayu Mandal 2015;41:9.

28. Kumar TV, Barbosa HA, Thakur MK, Paredes-Trejo F. Validation of Satellite (TMPA and IMERG) Rainfall Products with the IMD Gridded Data Sets over Monsoon Core Region of India. In: B. Rustamov R, ed. Satellite Information Classification and Interpretation. IntechOpen, 2019. DOI:10.5772/intechopen.84999.

29. Farrar J, Manson P, editors. Manson's Tropical Diseases. 23rd ed. Edinburgh: Elsevier Saunders; 2014.

30. Promprou, S, Jaroensutasinee, M & Jaroensutasinee, K. (2005). Climatic Factors Affecting Dengue Haemorrhagic Fever Incidence in Southern Thailand. WHO Regional Office for South-East Asia. Available from https://apps.who.int/iris/handle/10665/164135 [Last accessed on 2021 Feb 20].