

Identifying and predicting climate change impact on vector-borne disease using machine learning: Case study of *Plasmodium falciparum* from Africa

Priyanka Singh^{1,*}, Sameer Saran²

¹ UPES, School of Computer Science, Dehradun-248007, Uttarakhand, India – priyanka.iirs@gmail.com

²Regional Remote Sensing Center-North, ISRO, Dept. of Space, New Delhi- 110049, India – saran.iirs@gmail.com

Abstract

Vector-borne diseases pose a significant threat to human health, particularly in regions vulnerable to climate change. Among these diseases, malaria, caused by the parasite *Plasmodium falciparum* and transmitted through the Anopheles mosquito, remains a major global health concern, particularly in sub-Saharan Africa. This study explores the use of machine learning techniques to identify and predict the impact of climate change on the transmission dynamics of *P. falciparum* malaria in Africa.

The research utilizes a combination of climate data, epidemiological records, and machine learning algorithms to analyze historical patterns and project future trends in malaria transmission. Key climate variables such as temperature, precipitation, humidity, and vegetation cover are integrated into predictive models to assess their influence on the abundance and distribution of mosquito vectors and the parasite's lifecycle. Through the application of machine learning models such as Maximum Entropy, this study aims to uncover complex relationships between climatic factors and malaria transmission dynamics. By training these models on historical data, they can accurately predict future scenarios under various climate change scenarios. The findings of this research will provide valuable insights into the potential impact of climate change on the spatial and temporal distribution of *P. falciparum* malaria in Africa. Such insights are crucial for designing targeted interventions and adaptation strategies to mitigate the anticipated rise in malaria cases and associated morbidity and mortality in the region. Moreover, the methodology developed in this study can serve as a framework for assessing and addressing the impact of climate change on other vector-borne diseases globally.

Keywords: Maximum Entropy, Malaria, Machine Learning, Geospatial, Climate Change, Healthcare

1. Introduction

Vector-borne diseases, such as malaria, dengue fever, Zika virus, and Lyme disease, pose significant public health challenges worldwide, particularly in regions where environmental conditions are favorable for the proliferation of disease vectors. Among these diseases, malaria remains a major global health concern, with the majority of cases occurring in sub-Saharan Africa. Climate change is increasingly recognized as a key driver influencing the distribution, abundance, and transmission dynamics of vector-borne diseases. In recent years, there has been growing interest in utilizing machine learning techniques to better understand and predict the impact of climate change on these diseases, particularly focusing on malaria caused by the parasite *Plasmodium falciparum*. Studies examining the relationship between climate change and vector-borne diseases have a rich history dating back several decades. Early research primarily focused on statistical modeling approaches to assess the association between climatic variables and disease incidence. These studies laid the groundwork for understanding the complex interactions between climate, vectors, hosts, and pathogens in disease transmission cycles. However, traditional statistical methods often have limitations in capturing nonlinear relationships and complex interactions within large and heterogeneous datasets.

Seasonal variations in vector abundance have a significant impact on the seasonal dynamics and geographic distributions of vector borne parasites abundance (Chavasse et al., 1999;

Emerson, Bailey, Mahdi, Walraven, & Lindsay, 2000; Mabaso, Craig, Vounatsou, & Smith, 2005). For example, malaria is one of the vector-borne diseases that often shows seasonal abundance to annual rainfall and temperature in various regions, such as Kenyan Highlands (Hay et al. 2002). These diseases have become increasingly common in recent years in many parts of the world, including Saudi Arabia (dengue) (Alkhalidy and Barnett, 2021), Senegal (chikungunya) (Dieng et al., 2022), Brazil (high incidence of dengue, Zika, and chikungunya) (Lisboa et al., 2022), and China (dengue fever (FD) outbreaks in epidemic areas) (Zhang et al., 2022). Generally, their survival depends upon food, water and shelter at certain geographic locations, but now they are surpassing their geographical barriers and becoming endemic in other regions of the globe. For example, the *Aedes aegypti* species is the main vector known for the predominance of malaria in the Americas (WHO, 2023) and now it is prevalent in entire world. The five protozoa that cause malaria are *Plasmodium falciparum*, *P. vivax*, *P. malariae*, *P. ovale*, and most recently *P. knowlesi* are spread by mosquitoes. More than 90% of malaria-related deaths worldwide are attributed to *P. falciparum* infection, which means that the disease still poses a serious threat to public health on a global scale (Snow, 2015). According to the World Health Organization's (WHO) 2019 World Malaria Report, there were 228 million cases of malaria worldwide in 2018, which resulted in 405 000 deaths, many of them children under the age of 5. Approximately 40% of the world's population is

affected by malaria, which is endemic in more than 90 nations (Garcia, 2010).

Therefore, when it comes to vector-borne disease epidemiology, temporal variation along with its spatial distribution is most concentrated aspect in infectious disease transmission. As evident from various studies on vector-borne diseases, the association of environmental variables were explored, and this topic is gaining recognition in spread of diseases in rural and diverse urban environments (Abd Majid et al., 2019; Almeida et al., 2020; Cordeiro et al., 2011; Khalid & Ghaffar, 2015; Maftai et al., 2021; Rose et al., 2020; Tunali et al., 2021). The knowledge about the variables which can influence the spreading of the mosquito throughout the world is of utmost importance, as it can be used in the proposition of preventive models (Sriklın et al., 2021). Therefore, it is necessary to know about the predictor variables that can imply the existence of *P. falciparum*, so that it becomes possible to apply more assertive measures for the sake of epidemiologic vigilance, prevention of diseases and health promotion. It is crucial to be aware of the environmental factors that can affect how widely a mosquito breeds since this information can be utilised to develop preventive models. Recently, various machine learning (ML) algorithms such as maximum entropy modelling (Maxent), random forests, and artificial neural networks have all been used successfully to assess the risk of invasive species (Abdulkareem et al., 2021; Hu et al., 2020; Kaur et al., 2022; Mbunge et al., 2022; Nkiruka et al., 2021; Peters et al., 2020).

The knowledge of predictor variables that can suggest the existence of *P. falciparum*. In this study, the environmental conditions for the occurrence of *P. falciparum* in Benin are evaluated and its potential distribution are predicted for risk assessment through commonly used machine learning algorithm, i.e. maximum entropy (maxent), to help in establishing the surveillance and preventive programs. Therefore, this study aims to propose methodology for the purposes of epidemiologic vigilance, preventing vector-borne diseases, and promoting public health.

In recent years, the emergence of machine learning techniques has provided new opportunities to overcome the limitations of traditional statistical approaches in modeling the impact of climate change on vector-borne diseases. Machine learning algorithms, such as Random Forest, Support Vector Machines, Gradient Boosting, and Neural Networks, offer powerful tools for analyzing large and multidimensional datasets, identifying complex patterns, and making accurate predictions.

Several studies have applied machine learning techniques to assess the impact of climate change on malaria transmission dynamics, with a particular focus on *P. falciparum* in Africa. These studies have utilized diverse datasets, including climate data from remote sensing satellites, epidemiological records, land cover data, and entomological surveys. By integrating these datasets and applying machine learning algorithms, researchers have been able to identify key environmental factors influencing mosquito abundance, parasite development, and disease transmission. While machine learning holds promise for improving our understanding of the impact of climate change on vector-borne diseases, several challenges remain. These include the need for high-quality and spatially explicit data, addressing

issues of data bias and uncertainty, and ensuring the interpretability and generalizability of machine learning models. Additionally, there is a need for interdisciplinary collaboration between climatologists, epidemiologists, entomologists, and data scientists to effectively leverage machine learning techniques for disease modeling and prediction.

2. Research Impact

The discussion surrounding the identification and prediction of climate change impacts on vector-borne diseases, focusing on *Plasmodium falciparum* malaria in Africa and employing machine learning techniques, encompasses several key points:

Methodological Advances and Contributions: The utilization of machine learning techniques represents a significant advancement in the field of disease ecology and epidemiology. By leveraging sophisticated algorithms, researchers can analyze large and complex datasets to uncover intricate relationships between climatic variables and disease transmission dynamics. The case study focusing on *P. falciparum* malaria in Africa serves as an illustrative example of how machine learning can be applied to address pressing public health challenges in regions highly susceptible to climate change.

Insights into Climate-Vector-Pathogen Interactions: Through the integration of climate data, epidemiological records, and entomological surveys, machine learning models can provide valuable insights into the intricate interactions between climate, vectors (e.g., *Anopheles* mosquitoes), and pathogens (e.g., *Plasmodium* parasites). These models can identify key environmental factors influencing vector abundance, parasite development, and disease transmission, thereby elucidating the underlying mechanisms driving malaria dynamics in Africa.

Predictive Capabilities and Future Projections: One of the primary strengths of machine learning approaches is their ability to make accurate predictions based on historical data. By training models on past climate and disease data, researchers can project future scenarios under various climate change scenarios. These projections are essential for anticipating the potential impact of climate change on malaria transmission dynamics, guiding the development of targeted intervention strategies, and informing public health policies in endemic regions.

Challenges and Limitations: Despite their promise, machine learning techniques also pose certain challenges and limitations. These include the need for high-quality and spatially explicit data, issues related to data bias and uncertainty, and concerns regarding the interpretability and generalizability of models. Addressing these challenges requires interdisciplinary collaboration between climatologists, epidemiologists, entomologists, and data scientists, as well as ongoing efforts to improve data collection, validation, and model validation processes.

Implications for Public Health and Adaptation Strategies:

The findings derived from machine learning models have significant implications for public health and adaptation strategies in malaria-endemic regions of Africa. By identifying areas at high risk of malaria transmission under different climate change scenarios, policymakers can prioritize resource allocation, implement targeted vector control measures, and strengthen healthcare infrastructure to mitigate the anticipated rise in malaria cases and associated morbidity and mortality. Additionally, machine learning can facilitate the development of early warning systems for malaria outbreaks, enabling proactive responses to emerging threats.

3. Methodology

This study is carried in four succeeding steps: 1) identification of endemic site and data collection on occurrences of *P. falciparum* at the site, 2) evaluation of the environmental elements that influence the incidence of *P. falciparum*, 3) construction of predictive model using ML, and 4) analysis of contribution of each environmental factors to *P. falciparum* occurrence in model.

2.1. Data Collection and Preprocessing

The targeted area is Benin country in West Africa where *P. falciparum* is the deadliest malaria parasite posing the greatest threat to African continent. The current environmental data of African continent are obtained from the Worldclim (<http://worldclim.org>) project, which has gathered high resolution worldwide climate coverages. For this study, the environmental data of bioclimatic variables at spatial resolution of 2.5 arc seconds. The next step is to obtain the occurrence of *P. falciparum* from Global Biodiversity Information Facility (GBIF) repository (GBIF.org, 2023). The GBIF is a global repository that holds compiled records of millions of species worldwide. The occurrences are then pre-processed to remove duplicate points and NA values.

2.2. Model Construction

The occurrence and environmental predictors are ready for model construction and predicting the potential range of *P. falciparum*. Before this, it is required to discover how well the model can perform. Therefore, some of the occurrences are reserved for testing of model as *test* data and remaining are considered as *train* data used in model training. In this study, 20% of the observations will be randomly withheld as test data while the remaining 80% would be kept as training data.

4. Results

3.1. Model Assessment

The Area Under the Curve (AUC) or sensitivity vs. 1-specificity graph generated by Maxent model explains the precision and fit of the predicted model. If AUC values are close to 1.0, this indicates the improved model performance, whereas 0.5 suggested no better performance than random model. However, if the species has a wide

range of distribution, AUC values may be lower because of the greater commission. The AUC value of this Maxent model is 0.871 signifying that this model is better than the random, shown in Figure 1.

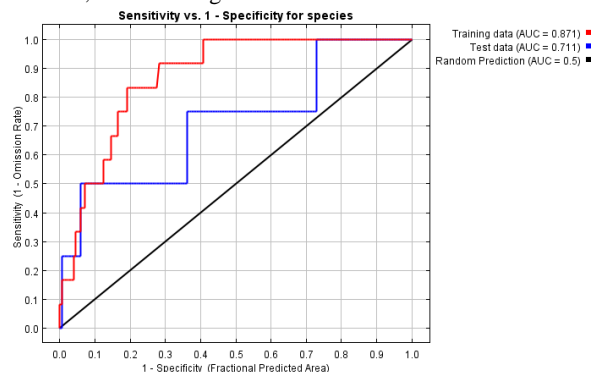


Figure 1. Sensitivity vs. 1-specificity graph of Maxent model

3.2. Important predictors evaluation

The environmental predictors contributed to model are evaluated by comparing the relative importance of the given predictors in the final model. The three main environmental predictors that contributed significantly to predicting the preferred prevalence site of *P. falciparum* out of the 19 variables were precipitation seasonality as bio15 (68.7%), precipitation of coldest quarter as bio19 (18.1%), and precipitation of warmest quarter as bio18 (8.8%), shown in Figure 2. The relative contribution of 19 variables is depicted in Jackknife graph (Figure 3). The response curves, depicted in Figure 4, from the Maxent output depicts the variations in the logistic value imparted by the changes in each predictor when all other variables are kept at its mean value. These plots demonstrate the probability of species occurrence at Y-axes which each climatic predictors in X-axes.

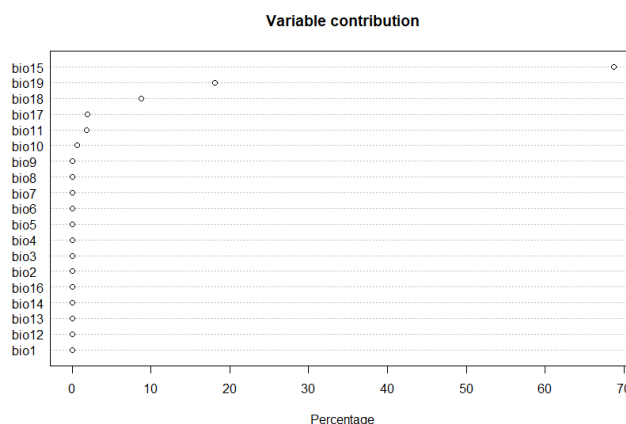


Figure 2. Variables contribution in modelling

3.3. Projecting Model for Site Prediction

The model determined the present appropriate existence site for *P. falciparum* in south coast of Benin, in West Africa based on the Maxent model (Figure 5). Based on the 'max SSS' logistic threshold, the logistic distribution was transformed to a binary raster. The Maxent model with se-

lected predictors is then projected on Africa extent to predict the potential zones for this malaria parasite.

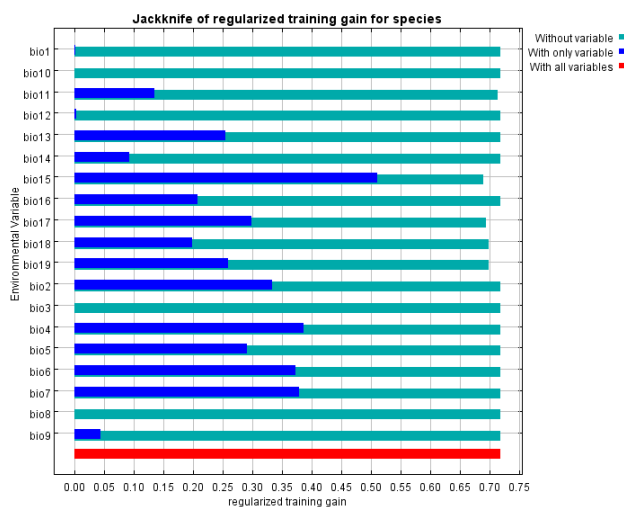


Figure 3. Jackknife graph

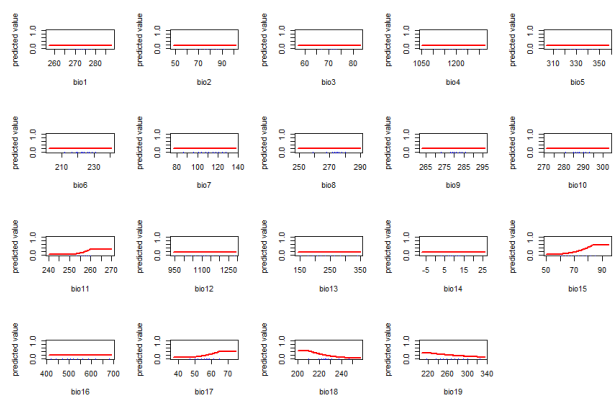


Figure 4. Response curves of predictors

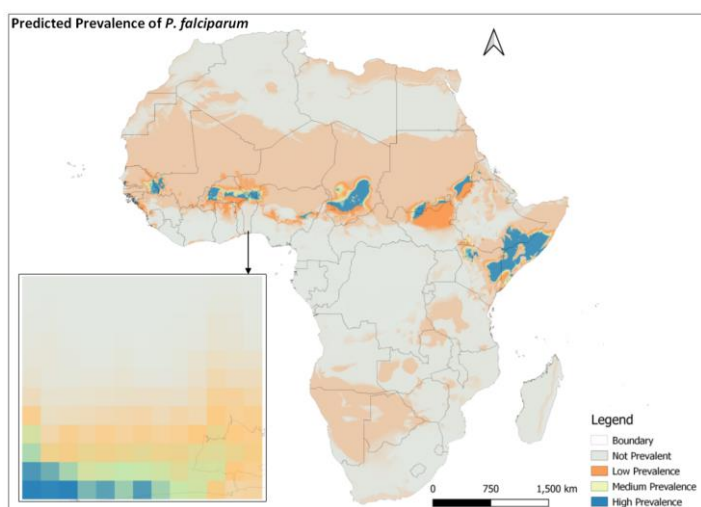


Figure 5. Predicted potential zone for *P. falciparum*.

4. Conclusions

In conclusion, the application of machine learning techniques in identifying and predicting the impact of climate change on vector-borne diseases, such as malaria caused by *P. falciparum* in Africa, represents a promising avenue for future research. By integrating diverse datasets and employing advanced modeling

approaches, researchers can gain valuable insights into the complex interactions between climate, vectors, and pathogens, thereby informing evidence-based strategies for disease control and adaptation in the face of changing environmental conditions. These modelling approach suggested to understand the widespread prevalence of malaria parasite, and this could be used to have preventive models. Through the study of variable importance and partial dependence plots, it is discovered that the most important climatic parameters, that influenced the prevalence of malaria parasite, *P. falciparum*. The model's predicted risk map for malaria parasite incidence closely tracked actual field data, demonstrating its strong capacity for prediction and management. The model also demonstrated the expanded incidence of parasite in African continent under the same climatic space. Therefore, when developing management and control strategies for malaria, this information could be considered.

Despite the challenges and limitations associated with these approaches, continued research efforts and interdisciplinary collaboration are essential for harnessing the full benefits of machine learning in the fight against vector-borne diseases in the context of a changing climate.

References

- Abd Majid, N., Muhamad Nazi, N., & Mohamed, A. F. (2019). Distribution and spatial pattern analysis on dengue cases in Seremban District, Negeri Sembilan, Malaysia. *Sustainability*, 11(13), 3572.
- Abdulkareem, A. B., Sani, N. S., Sahran, S., Alyessari, Z. A. A., Adam, A., Abd Rahman, A. H., & Abdulkareem, A. B. (2021). Predicting COVID-19 based on environmental factors with machine learning. *Intell. Autom. Soft Comput*, 28, 305-320.
- Alkhalidy, I., & Barnett, R. (2021). Explaining neighbourhood variations in the incidence of dengue fever in Jeddah City, Saudi Arabia. *International Journal of Environmental Research and Public Health*, 18(24), 13220.
- Almeida, L. S., Cota, A. L. S., & Rodrigues, D. F. (2020). Sanitation, Arboviruses, and Environmental Determinants of Disease: impacts on urban health. *Ciencia & saude coletiva*, 25, 3857-3868.
- Chavasse, D. C., Shier, R. P., Murphy, O. A., Huttly, S. R. A., Cousens, S. N., & Akhtar, T. (1999). Impact of fly control on childhood diarrhoea in Pakistan: community-randomised trial. *The Lancet*, 353(9146), 22-25.
- Cordeiro, R., Donalisio, M. R., Andrade, V. R., Mafra, A. C., Nucci, L. B., Brown, J. C., & Stephan, C. (2011). Spatial distribution of the risk of dengue fever in southeast Brazil, 2006-2007. *BMC Public Health*, 11(1), 1-10.
- Dieng, I., Fall, C., Barry, M. A., Gaye, A., Dia, N., Ndione, M. H. D., ... & Sall, A. A. (2022). Re-Emergence of Dengue Serotype 3 in the Context of a Large Religious Gathering

- Event in Touba, Senegal. *International Journal of Environmental Research and Public Health*, 19(24), 16912.
- Emerson, P. M., Bailey, R. L., Mahdi, O. S., Walraven, G. E., & Lindsay, S. W. (2000). Transmission ecology of the fly *Musca sorbens*, a putative vector of trachoma. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94(1), 28-32.
- Garcia, L. S. *Malaria Clin Lab Med*, 30 (1)(2010). *View PDF View article View in Scopus*, 93-129.
- GBIF.org (03 July 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.e4qzqu>
- Hay, S. I., Cox, J., Rogers, D. J., Randolph, S. E., Stern, D. I., Shanks, G. D., ... & Snow, R. W. (2002). Climate change and the resurgence of malaria in the East African highlands. *Nature*, 415(6874), 905-909.
- Hu, L., Liu, B., & Li, Y. (2020). Ranking sociodemographic, health behavior, prevention, and environmental factors in predicting neighborhood cardiovascular health: A Bayesian machine learning approach. *Preventive medicine*, 141, 106240.
- Kaur, I., Sandhu, A. K., & Kumar, Y. (2022). Artificial intelligence techniques for predictive modeling of vector-borne diseases and its pathogens: a systematic review. *Archives of Computational Methods in Engineering*, 29(6), 3741-3771.
- Khalid, B., & Ghaffar, A. (2015). Dengue transmission based on urban environmental gradients in different cities of Pakistan. *International journal of biometeorology*, 59, 267-283.
- Lisboa, T. R., Serafim, I. B. M., Serafim, J. C. M., Ramos, A. C., do Nascimento, R. M., & Roner, M. N. B. (2022). Relação entre incidência de casos de arboviroses e a pandemia da COVID-19. *Interdisciplinary Journal of Applied Science*, 6(10), 31-36.
- Mabaso, M. L. H., Craig, M., Vounatsou, P., & Smith, T. (2005). Towards empirical description of malaria seasonality in southern Africa: the example of Zimbabwe. *Tropical Medicine & International Health*, 10(9), 909-918.
- Maftai, C., Bărbulescu, A., Rugina, S., Nastac, C. D., & Dumitru, I. M. (2021). Analysis of the arbovirose potential occurrence in Dobrogea, Romania. *Water*, 13(3), 374.
- Mbunge, E., Millham, R. C., Sibiya, M. N., & Takavarasha, S. (2022, March). Application of machine learning models to predict malaria using malaria cases and environmental risk factors. In *2022 Conference on Information Communications Technology and Society (ICTAS)* (pp. 1-5). IEEE.
- Nkiruka, O., Prasad, R., & Clement, O. (2021). Prediction of malaria incidence using climate variability and machine learning. *Informatics in Medicine Unlocked*, 22, 100508.
- Peters, D. P., McVey, D. S., Elias, E. H., Pelzel-McCluskey, A. M., Derner, J. D., Burruss, N. D., ... & Rodriguez, L. L. (2020). Big data–model integration and AI for vector-borne disease prediction. *Ecosphere*, 11(6), e03157.
- Rose, N. H., Sylla, M., Badolo, A., Lutomia, J., Ayala, D., Aribodor, O. B., ... & McBride, C. S. (2020). Climate and urbanization drive mosquito preference for humans. *Current Biology*, 30(18), 3570-3579.
- Sriklin, T., Kajornkasirat, S., & Puttinaovarat, S. (2021). Dengue transmission mapping with weather-based predictive model in three southernmost provinces of Thailand. *Sustainability*, 13(12), 6754.
- Snow, R. W. (2015). Global malaria eradication and the importance of *Plasmodium falciparum* epidemiology in Africa. *BMC medicine*, 13(1), 1-3.
- Tunali, M., Radin, A. A., Başbüyük, S., Musah, A., Borges, I. V. G., Yenigun, O., ... & Campos, L. C. (2021). A review exploring the overarching burden of Zika virus with emphasis on epidemiological case studies from Brazil. *Environmental Science and Pollution Research*, 1-15.
- WHO (World Health Organization). Dengue and Severe Dengue. 2022. Available online: <https://www.who.int/news-room/q-a-detail/dengue-and-severe-dengue> (accessed on 04 August 2023).
- Zhang, Y., Ren, H., & Shi, R. (2022). Influences of Differentiated Residence and Workplace Location on the Identification of Spatiotemporal Patterns of Dengue Epidemics: A Case Study in Guangzhou, China. *International Journal of Environmental Research and Public Health*, 19(20), 13393.