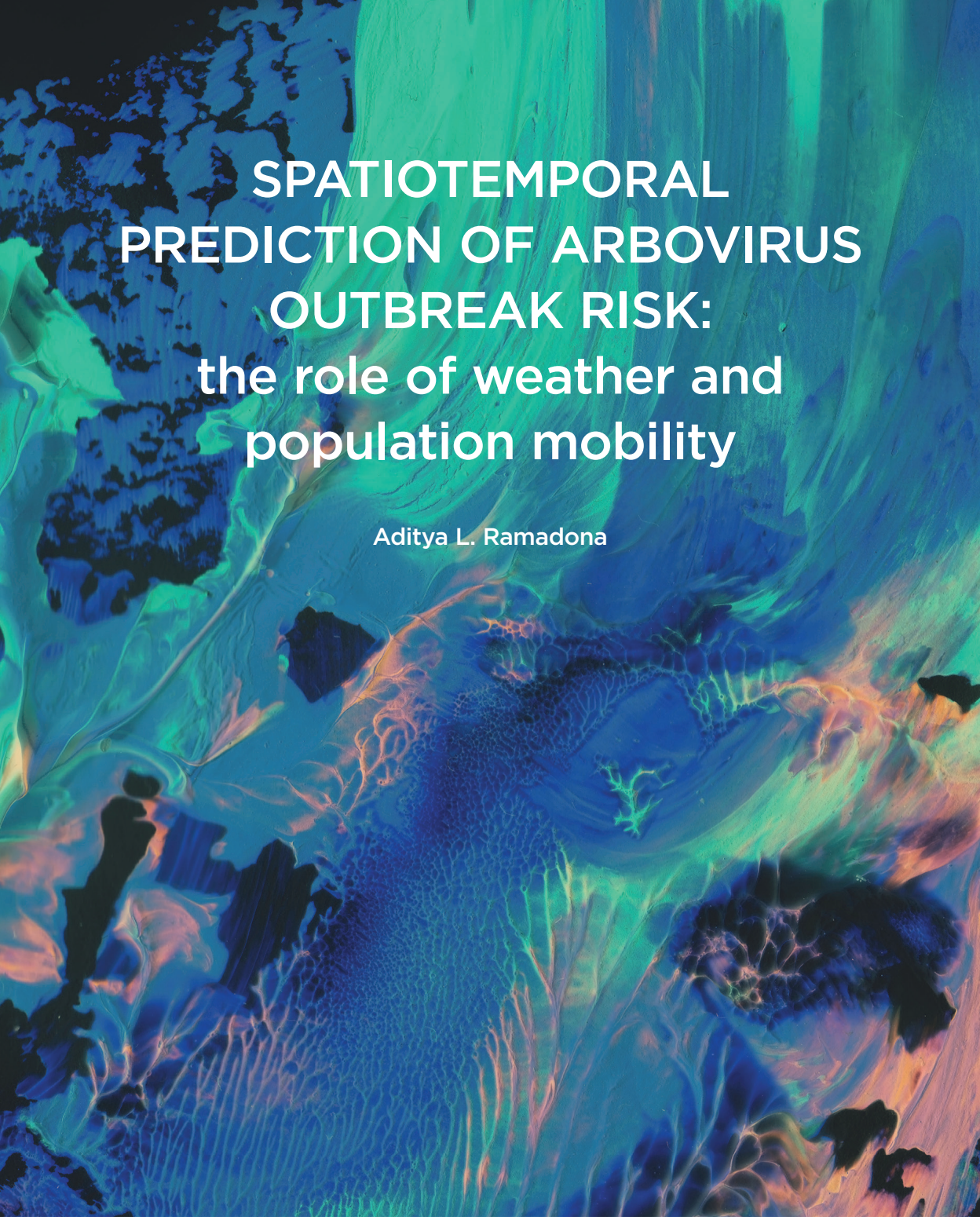


SPATIOTEMPORAL PREDICTION OF ARBOVIRUS OUTBREAK RISK: the role of weather and population mobility

Aditya L. Ramadona

UMEÅ UNIVERSITY

UMEÅ UNIVERSITY

Spatiotemporal Prediction of Arbovirus Outbreak Risk

- the role of weather and population mobility

Aditya Lia Ramadona

Department of Epidemiology and Global Health
Umeå 2021

This work is protected by the Swedish Copyright Legislation (Act 1960:729)
Dissertation for PhD
ISBN: 978-91-7855-687-8 (print)
ISBN: 978-91-7855-688-5 (pdf)
ISSN: 0346-6612
Information about cover design: Inhousebyrån, Umeå University
Electronic version available at: http://umu.diva-portal.org/
Printed by: Cityprint i Norr AB
Umeå, Sweden 2021

"You can't connect the dots looking forward;
you can only connect them looking backward."

— Steve Jobs

Table of Contents

Abstract v

Sammanfattning på Svenska vii

Contributing Papers ix

Introduction 1

- Dengue Infection 1
- Dengue Vector and Weather 2
- Population Mobility 3
- Spatiotemporal Model and Early Warning Systems 4
- Arbovirus Outbreak Risk 5

Aims and Objectives 6

Methods 7

- Summary 7
- Study Location 8
- Ethical Considerations 9
- Data Collection 9
  - Dengue and population data 9
  - Meteorological data 12
  - Chikungunya data 15
  - Mobility data 16

Analysis 16
Influence of temperature and rainfall 16
Influence of past dengue counts 16
Influence of population mobility 17
Fine-scale (Paper 2) 17
Coarse-scale (Paper 4) 19
Areas spatial dependency (Paper 3) 19
Adjacency Matrix based on the administrative map 20
Adjacency Matrix based on geotagged social media data 20
Spatiotemporal model of dengue outbreak risk 20
Prediction 21
Results 23
Influence of Temperature and Rainfall 23
Influence of Past Dengue Counts 24
Influence of Population Mobility 25
Fine-scale (Paper 2) 25
Coarse-scale (Paper 4) 29
Spatiotemporal Model of Dengue Outbreak Risk 30
Predictive Performance of the Models 32
Discussion 36
Influence of Climate Variability 36
Influence of Population Mobility on Disease Spread 37
The Implementation of Dengue Early Warning Systems 38

Limitations 41

Conclusion 42

Acknowledgement 44

References 46

Abstract

Background. Arboviruses such as dengue and chikungunya have been a significant public health burden globally for several decades. In Indonesia, all four dengue serotypes are circulating. Considering that Indonesian children are exposed to dengue early in life, and secondary infection is more likely to cause severe dengue, the population of Indonesia is confronting a high potential risk of severe dengue. Severe complications such as hemorrhage can develop and lead to fatal outcomes. There exists no specific treatment for dengue infection, but symptomatic treatment can be effective to prevent deaths. Consequently, vector control has become a critical component for controlling dengue transmission, but it is currently often triggered as a reactive response to observed outbreak clusters. Based on disease surveillance, it thus remains challenging to implement vector control efficiently to prevent outbreaks. While meteorological conditions have shown to be predictive of dengue incidence over space and time, it has rarely been used to predict outbreaks at a fine-scale intra-urban level. Further, as the propagation of dengue outbreaks and the introduction of viruses has been found to be associated with human mobility, predictive models combining meteorological conditions with granular mobility data hold promise to provide more predictive models. The objectives in this thesis were to 1) describe the influence of temperature, rainfall, and past dengue cases, and population mobility on dengue risk; 2) develop and validate spatiotemporal models of dengue outbreak risk at fine-scale at the intra-urban level; 3) to utilize new data to assess the emergence and spread of chikungunya in an outbreak situation.

Methods. Initially, multivariate time series regression models were established to analyze the risk of dengue corresponding to monthly mean temperature, cumulative rainfall, and past dengue case. Following that, we investigated the potential use of geotagged social media data as a proxy of population mobility to estimate the effect of dengue virus importation pressure in urban villages. Subsequently, we employed distributed lag non-linear models with a Spatiotemporal Bayesian hierarchical model framework to determine the exposure-lag-response association between the risk of dengue and meteorological data while allowing the spatial covariance to be informed by mobility flows. Finally, we validated the selected best-fitted model by its predictive ability using an unseen dataset to mimic an actual situation of an early warning system in use.

Results. We found that an optimal combination of meteorology and autoregressive lag terms of past dengue cases was predictive of dengue

incidence and the occurrence of dengue epidemics. Subsequently, when we integrated mobility data our results suggested that population mobility was an essential driver of the spread of dengue within cities when combined with information on the local circulation of the dengue virus. The geotagged Twitter data was found to provide important information on presumably local population mobility patterns which were predictive and can improve our understanding of the direction and the risk of spread.

Conclusions. A spatiotemporal prediction model was developed that predicted a prognosis of dengue at fine spatial and temporal resolution. Subsequently, such a prognosis can be used as the foundation for developing an early warning system to more effectively deploy vector control prior to the establishment of local outbreak clusters . These findings have implications for targeting dengue control activities at the intra-urban villages level, especially in the light of ever increasing population growth, mobility and climate change.

Keywords: arbovirus, dengue, temperature, rainfall, extreme weather, climate variability, population mobility, twitter data, social media, forecasting model, early warning, epidemic, big data, INLA, spatiotemporal model, climate services

Sammanfattning på Svenska

Bakgrund. Arbovirus som dengue och chikungunya har i ett globalt perspektiv varit en betydande folkhälsobelastning i flera decennier. I Indonesien cirkulerar alla fyra dengue-serotyperna och sjukdomen drabbar främst barn. Med tanke på att indonesiska barn utsätts för dengue tidigt i livet, och att sekundär infektion är mer benägen att orsaka svår dengue, står befolkningen i Indonesien inför en hög potentiell risk för ökad risk för epidemier med svår sjukdom. Allvarliga komplikationer som blödning kan utvecklas och leda till dödliga utfall. Det finns ingen specifik behandling för dengueinfektion, men symtomatisk behandling kan vara effektiv för att förhindra dödsfall. Istället för behandling eller vaccin haråtgärder mot myggor blivit en viktig komponent för att kontrollera smittspridningen av dengue. Dock mobiliseras för närvarande mygginterventioner ofta som ett reaktivt svar på obserververade utbrottskluster då det delvis redan är för sent. Baserat på sjukdomsövervakning är det därför fortfarande en utmaning att implementera vektorkontroll effektivt för att förhindra utbrott i ett tidigare skede. Även om meteorologiska förhållanden har visat sig kunna föregå dengueutbrott över tid och rum, har det sällan använts för att förutsäga utbrott på en finskalig nivå inom städer. Eftersom spridningen av dengue-utbrott och införandet av virus har visat sig vara associerad med mänsklig rörlighet, kan prediktiva modeller som kombinerar meteorologiska förhållanden med mobilitetsdata att ge bättre prognosmodeller. Målen i denna avhandling har varit att 1) beskriva inverkan av temperatur, nederbörd och tidigare fall av denguefeber och befolkningsrörlighet på risken för denguefeber; 2) utveckla och validera spatiotemporala modeller för risk för dengueutbrott i fin skala inom städer; 3) att använda nya data för att bedöma uppkomsten och spridningen av chikungunya i en utbrottssituation.

Metoder. Initialt etablerades multivariata tidsserieregressionsmodeller för att analysera risken för denguefeber motsvarande månadsmedeltemperatur, kumulativ nederbörd och tidigare denguefall. Därefter undersökte vi den hur geotaggade sociala medier-data kan användas som en proxy för befolkningsrörlighet för att uppskatta effekten av importer av denguevirus i urbana byar. Vi gick sedan vidare genom att använda distribuerade icke-linjära lagmodeller med en Spatiotemporal Bayesian hierarkisk modellram för att bestämma sambandet för risken för dengue och meteorologiska data. Detta samtidigt som den rumsliga kovariansen utformades efter mobilitetsflöden. Slutligen validerade vi den utvalda bäst anpassade modellen genom dess prediktiva förmåga med hjälp av en osynlig datauppsättning för att efterlikna en faktisk situation med ett tidigt varningssystem i bruk.

Resultat. Vi fann att en optimal kombination av meteorologi och autoregressiva fördröjningstermer från tidigare denguefall var förutsägande för förekomsten av denguefeber och förekomsten av dengueepidemier. När vi sedan integrerade mobilitetsdata visade våra resultat att befolkningsrörlighet var en viktig drivkraft för spridningen av denguefeber i städer i kombination med information om den lokala cirkulationen av dengueviruset. Den geotaggade Twitter-datan visade sig ge viktig information om förmodligen lokala mobilitetsmönster som var förutsägande och kan förbättra vår förståelse av riktningen och risken för spridning.

Slutsatser. En spatiotemporal prediktionsmodell utvecklades som förutspådde dengue med fin rumslig och tidsmässig upplösning. En sådan prognosmodell kan användas tillsammans med väderinformation för att utveckla ett system för tidig varning. En tidig varning kan användas för att mer effektivt distribuera vektorkontroll före etableringen av lokala utbrottskluster. Dessa resultat har implikationer för inriktningen av denguekontrollaktiviteter inom städer, särskilt mot bakgrund av ökande befolkningstillväxt, rörlighet och klimatförändringar.

Contributing Papers

This thesis is based on the following four individual papers. The first, second, and fourth papers were published in open access journals, and therefore, no permission was required to reprint.

1. **Ramadona A**, Lazuardi L, Hii Y, Holmner Å, Kusnanto H, Rocklöv J. Prediction of dengue outbreaks based on disease surveillance and meteorological data. PLOS One. 2016;11(3):e0152688. doi:10.1371/journal.pone.0152688
2. **Ramadona A**, Tozan Y, Lazuardi L, Rocklöv J. A combination of incidence data and mobility proxies from social media predicts the intra-urban spread of dengue in Yogyakarta, Indonesia. PLOS Neglected Tropical Diseases. 2019;13(4):e0007298. doi:10.1371/journal.pntd.0007298
3. **Ramadona A**, Tozan Y, Wallin J, Lazuardi L, Utarini A, Rocklöv J. Prediction of dengue virus spatiotemporal outbreak cluster dynamics in Yogyakarta, Indonesia. (manuscript).
4. Rocklöv J, Tozan Y, **Ramadona A**, Sewe M, Sudre B, Garrido J, Lary C, Lohr W, Semenza J. Using big data to monitor the introduction and spread of Chikungunya, Europe, 2017. Emerg Infect Dis. 2019;25(6):1041-1049. doi:10.3201/eid2506.180138

Introduction

Dengue Infection

Dengue infection is the most important arthropod-borne viral infection of humans, an ongoing global threat, and an endemic or epidemic disease in almost every country located in the tropics.[1] Severe dengue occurs in only a small proportion of dengue infections, but early identification of high-risk patients is difficult.[2] Accordingly, patients with uncomplicated infections are frequently hospitalized for observation, and as a result, it burdens health care systems.

A study to estimate the public health burden of dengue in 2010 put the toll at 390 million dengue infections per year,[3] or more than three times the dengue burden estimate released by the World Health Organization.[4] Meanwhile, in Indonesia, actual dengue infections were estimated to be 2.3 to 11.5 times higher than the data reported from the routine disease surveillance system.[5–7] Most dengue burden in Indonesia was due to non-severe cases and these patients do not seek treatment, and therefore leading to substantial underreporting of cases.[8] Although the incidence of dengue has been increasing over the years, the case fatality rate in Indonesia has declined, from an estimated 41% when first reported in 1968 to 0.7% in 2013.[9]

A rough estimate of the cost of the dengue outbreak in Indonesia in 2011 outbreak costs was 6.75 million USD.[10] A more recent study in 2015, using data from public and private hospitals and health centers, revealed that the total direct and indirect costs per dengue case assessed at three provinces in Indonesia, namely, Yogyakarta, Bali, and Jakarta, were US$791, US$1,241, and US$1,250, respectively.[11] Considering the estimated total number of cases and the unit costs per dengue episode, Indonesia was considered to be the country with the highest economic burden of dengue in the Southeast Asia region.[12]

Dengue can be caused by four different serotypes (DENV 1-4) of the genus flavivirus (family Flaviviridae).[2] Human infection with one serotype confers long-lived serotype-specific immunity but only short-lived cross-immunity between serotypes.[2,13] Nevertheless, secondary infection, in the form of two sequential infections by different serotypes, is a risk factor for severe illness.[2,14,15] In Indonesia, all four serotypes are circulating, and secondary infection by DENV-1 serotype is more likely to cause severe dengue.[16,17]

Dengue infection has a broad clinical illness spectrum, ranging from asymptomatic or undifferentiated febrile illness (viral syndrome), dengue fever (DF), dengue hemorrhagic fever (DHF), to dengue shock syndrome (DSS).[18] Until now, there has been no specific treatment for dengue infection. The treatment is symptomatic and supportive aiming to overcome the loss of plasma fluid due to increased capillary permeability and bleeding.[19] The diagnosis of dengue was based on clinical and laboratory criteria.[20] However, in Indonesia at present, most dengue diagnoses are based on clinical criteria. One study in Yogyakarta conducted in November-December 2016 found that of 230 dengue cases reported, only about 45% of cases were accompanied by laboratory confirmation.[19]

Indonesian children are exposed to dengue early in life. A national serosurvey in urban environments reported frequencies of seropositivity: 33.8% in 1- to 4-year-olds and 65.4% in the 5- to 9-year-olds.[17,21] In addition, the findings of a DENV seroprevalence survey in Yogyakarta indicated that most children were infected by at least one DENV serotype and frequently ≥ 1 DENV serotype before the age of 10 years.[22] Furthermore, this study also showed that the overall median age of dengue cases was 11 years (interquartile range [IQR] 6–20 years), and this remained broadly consistent over the 9 years of the study period.[22]

Dengue Vector and Weather

Aedes aegypti and Aedes albopictus are considered the primary vectors of dengue.[23] These two vectors adapt well to human environments and may sustain dengue transmission among humans.[23,24] Indoor transmission events mainly trigger dengue infections, and urbanized areas have a greater risk.[25] Aedes aegypti and Aedes albopictus exist in Yogyakarta, and an earlier study here showed that Aedes aegypti and Aedes albopictus had a negative association. Aedes albopictus numbers were positively correlated with vegetation around areas, while Aedes aegypti were negatively correlated with such.[25]

Dengue vector dynamics are sensitive to the changes in weather patterns, and many studies indicated that weather patterns were associated with dengue incidence.[26–30] Aedes larvae are primarily found in the rainy season because rainfall generates vector-breeding sites.[31–33] Numerous artificial and natural containers can reserve water from rain, and it then becomes potential breeding habitat for Aedes mosquitoes. However, the relationship between rainfall and dengue is non-linear, and heavy rain can flush breeding sites and potentially destroy the immature vector, which in turn affects vector abundance downwards.[32,34] Temperature influences vector biology

and vector-virus interactions through vector longevity, mating, dispersal, feeding behavior, oviposition, and more rapid virus replication.[26,35] In a temperature-controlled experiment, optimal mean temperature ranges for the survival of female adult mosquitoes and immature stages are 15-30 °C and 15-35 °C, respectively.[36]

Population Mobility

Spatial connectivity plays a vital role in mosquito-borne disease transmission, driven by human mobility.[37] Mobility allows individuals to transport virus between localities causing infections among local mosquitoes.[38] Therefore, human mobility data has the potential to inform where and when new dengue cases are likely to occur within a region and could be used to improve surveillance, intervention, and disease prevention.[39–41] Earlier studies have shown that human mobility influences the spatial spread of vector-borne diseases even on the short spatial scale corresponding to intra-city distances.[42,43]

One obstacle to studying the interactions among vectors, humans, and pathogens has been the limited availability of datasets for analyzing human mobility patterns.[44] In the past, human mobility patterns were mainly estimated based on the traveling history of survey respondents. These estimations can potentially have inaccuracies due to respondents' errors in recalling the information, are labor-intensive, and tend to be quite time consuming and costly. More recent studies have relied on low-resolution mobile phone records, such as call and messaging logs from mobile phone networks, but the use of mobile phone data for tracking human mobility is likely to be fraught with privacy concerns and data access restrictions.[45]

Twitter is a social media platform providing free samples of their user data that could be used as novel data sources to complement traditional infectious disease surveillance.[44,46] Several studies have explored using near-real-time streaming data from Twitter to investigate public health issues.[47–49] Although health-related studies using Twitter data have mainly focused on analyzing Twitter content using text mining; nowadays, studies using geotagged Twitter data are emerging.[45,50–54] Given that social media platforms, especially Twitter, are in widespread use in Indonesia,[55] the use of geolocation data as a proxy to monitor human mobility patterns could be applicable here.

Spatiotemporal Model and Early Warning Systems

Since there is no specific treatment for dengue infection, vector control has become a critical component of curbing dengue transmission. However, vector control is currently often triggered as a reactive response to observed outbreak clusters,[56] and it remains challenging to know when and where to use vector control efficiently. Descriptive maps showing dengue hotspots could help in identifying high-risk areas. However, a large number of unknowns, including effective entomological predictors, genetic diversity of circulating viruses, population serological profile, and human mobility, continue to pose challenges and limit the ability to produce accurate and effective risk maps.[57]

Spatiotemporal models with random-effects could potentially help to account for unknown or unobserved factors of importance, and therefore developing spatiotemporal random-effect models with covariates would help to better understand dengue transmission patterns.[58] Temperature and rainfall have been shown to often influence the risk of dengue,[58] but these data have rarely been used to predict outbreaks at a fine spatial level.[59] Further, as the propagation of dengue outbreaks and the introduction of virus have been found to be related to human mobility, predictive models combining meteorological conditions with granular mobility data hold promise to provide better models.[59]

As an outbreak-prone disease, timely and accurate reporting and forecasting of dengue cases is the pillar of its control. A spatiotemporal model could provide risk maps to identify high-risk areas and, subsequently, to support an early warning system.[57] For Indonesia, an initiative has been taken to implement dengue early warning systems locally. Still, this system has its limitations as it only tries to estimate the nature of the following year's epidemic[60] and requires data collection on a set of predictors at the household level.[61] These limitations render this early warning system impractical, labor-intensive, time consuming, costly, and unsustainable.

Despite the widespread use of social media data in Indonesia, and the many models and early warning systems developed for dengue, we believe that no attempt has been made to integrate surveillance and meteorological data, as well as population mobility proxied from social media data, to predict dengue outbreak risk and spread at a fine-scale at intra-urban village level over time.

Arbovirus Outbreak Risk

An unprecedented emergence of epidemic arboviral diseases has been seen in the past five decades, resulting from urbanisation, globalisation, and international mobility,[62] accounting for more than 17% of infectious diseases worldwide.[63] This thesis discusses two arboviral diseases, dengue and chikungunya, for which Aedes aegypti and Aedes albopictus are the primary vectors.

Globally, the geographic distribution of Aedes aegypti and Aedes albopictus is rapidly expanding and they have become the most significant disease vectors.[64] Both vectors have a clear seasonal pattern and are mainly clustered around urban areas.[65] The abundance of Aedes aegypti and Aedes albopictus was estimated to increase 9.5% over the last century, and a potential further increase of 20 or 30% is expected by the end of this century.[65]

There have not been many studies investigating the epidemiologic consequences of mobility patterns at different spatial scales, fine and coarse, within the context of arboviral diseases. Combining interventions known to be effective against multiple arboviral diseases will offer the most cost-effective and sustainable strategy for disease reduction.[62] Therefore, in this study, we also aim to utilize newly available data and methods to assess the emergence and spread of chikungunya in an outbreak situation, besides dengue, which is the main focus of this thesis, at a coarser and larger spatial scale.

An effective response to outbreaks relies on timely intervention, but the collection and analysis of outbreak data are becoming increasingly complex.[66] Considering this, here, we took an extra step by exploring social media datasets from Twitter as a proxy for population mobility to understand mobility patterns and, subsequently, to inform intervention strategies to prevent the rapid geographical spread of emerging infectious diseases. We investigated the 2017 chikungunya outbreaks in Europe by analyzing the patterns of population mobility in Italy and France to assess the risk of virus transmission, importation, and short-range dispersion. This assessment becomes essential since outbreak response planning have very limited time and resources for the response on the ground.[66]

Aims and Objectives

This study aims to explore the possibility of developing spatiotemporal prediction models of arbovirus outbreak risks that can be used as the foundation for a reliable early warning system and allow sufficient time and information for more efficient outbreak management.

Objectives

The objectives of this study are to address the following issues:

1. To describe the influence of temperature, rainfall, past dengue cases, and population mobility on dengue risk.
2. To develop and validate spatiotemporal models of dengue outbreak risk at fine-scale at the intra-urban level.
3. To utilize data from various data sources to demonstrate the feasibility to assess the onward spread of a chikungunya outbreak in Europe.

The objectives are discussed sequentially in a series of four papers, as shown in **Figure 1**.

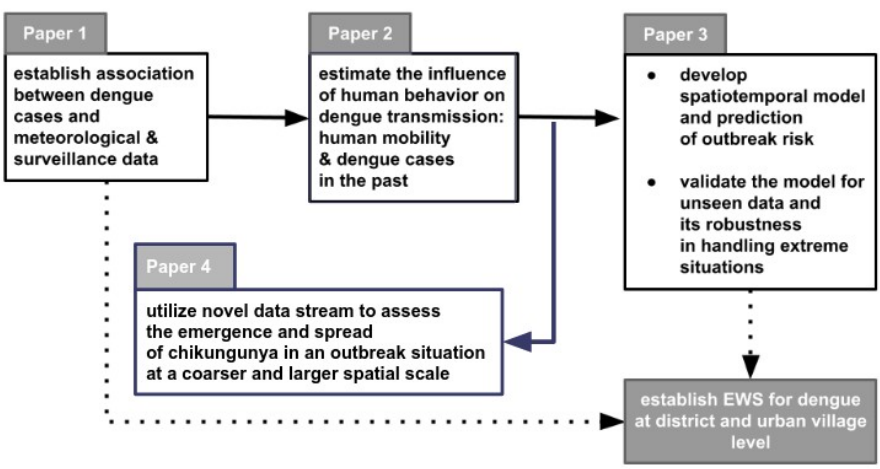


Figure 1. Description of research and research progression in the papers that are included in the thesis.

Methods

Summary

This dissertation presents an analysis of meteorological influences and people's mobility on disease transmission through the use of traditional and novel data streams. The analysis incorporates temperature, rainfall, and dengue cases from surveillance data, as well as population mobility using geotagged social media data as a proxy.

We strive to study various prediction models of dengue outbreak risk and how well and what type of model is most suited for developing an early warning system of arbovirus outbreak risk. Initially, we investigated the association of temperature, rainfall, and past dengue cases on dengue risk. Later, we also included population mobility in the models. We then developed and validated a spatiotemporal model for dengue outbreak risk prediction. This study's results include the number of predicted cases at the city level and at the urban village level, including the spatial pattern of dengue outbreak risk. We also assessed the spatiotemporal zones of risk of chikungunya spreading from a local outbreak zone in Europe.

In paper 1, we investigated the association of temperature, rainfall, and disease counts using the generalized additive modeling framework. We used the meteorological and disease data to capture the expected baseline dynamics of dengue and quantify the effects of temperature and rainfall, as well as past dengue cases, on disease risk. Paper 2 assessed the influence of population mobility and past dengue cases on dengue risk, using the integrated nested Laplace approximation (INLA) framework. We investigated the potential use of geotagged tweets from publicly available social media datasets as a proxy for population mobility. In Paper 3, we further investigated the findings of Paper 1 and Paper 2 as the basis for developing a spatiotemporal model to predict dengue outbreak risk at the fine-scale intra-urban village level. We also explored the nonlinear and delayed associations between dengue risk and meteorological variables using distributed lag nonlinear models (DLNMs). Following that, we validated the highest-performance model, using an unseen dataset.

In Paper 4, we took an extra step by applying part of the methods developed in Paper 2 to a different arboviral disease - chikungunya- at a coarse spatial resolution of a much larger region. There are not many studies investigating the epidemiological consequences of mobility patterns at different spatial scales. Considering this, we applied social media datasets from Twitter as a proxy for population mobility in Italy and France to assess the risks of virus

transmission, virus importation, and its short- and medium range dispersion during the 2017 chikungunya outbreaks in Europe.

Study Location

Our study was conducted in two different locations. The study location for the first three papers was Yogyakarta Municipality. Yogyakarta Municipality is one of the five districts and the capital of Yogyakarta Province in Indonesia. Yogyakarta municipality is a densely populated urban area spread over 32.5 square kilometers. The municipality has 14 sub-districts and 45 urban villages (**Figure 2**). The mean urban village area is 0.73 km², with an interquartile range of 0.47 km² and 0.87 km².

Figure 2. The administrative map of the 45 study urban villages in Yogyakarta municipality, Indonesia.

The study area for the fourth paper included parts of France and Italy. In this study, we based our analysis on freely available administrative country maps provided by DIVA-GIS (www.diva-gis.org).

Figure 3. The administrative map of the study areas delineated by geographical confirmed presence of the Aedes aegypti vector transmitting chikungunya. Figure from Paper 4.[44]

Ethical Considerations

Ethical approval was not required for this study. This ecological or population-based study used published and open datasets from official institutions in the study areas and did not include any private or identifiable information on human subjects. The disease surveillance and meteorological data were extracted from the regular reports of the Health Office of Yogyakarta Municipality, the Department of Population and Civil Registration of Yogyakarta Municipality, the Indonesian Agency for Meteorology and Climatology and Geophysics, and the European Centre for Disease Prevention and Control. Location data were obtained from Twitter through the Academic research product track agreement (ref:00DA0000000K0A8.5004w000025vF9p).

Data Collection

The data used in this study include dengue and chikungunya case counts, temperature, rainfall, and geotagged tweet data. We collected these data continuously over time as each of the four papers was written. This permitted the development and training of a forecasting model and risk assessment using updated data.

Dengue and population data

We obtained the total monthly notified dengue cases for Yogyakarta municipality since January 2001, and for each of the 45 urban villages of Yogyakarta between February 2013 and January 2020, from the Dengue Surveillance Report prepared by the Yogyakarta Municipality Health Office (**Figure 4,5**). Notified dengue cases included only those classified as DHF or DSS. Dengue is a notifiable disease with passive case reporting of DHF and DSS cases from government and private hospitals to the national surveillance system.[9] In addition, we complemented dengue surveillance data with monthly population data during the same period from Yogyakarta Municipality's Department of Population and Civil Registration (**Figure 6**).

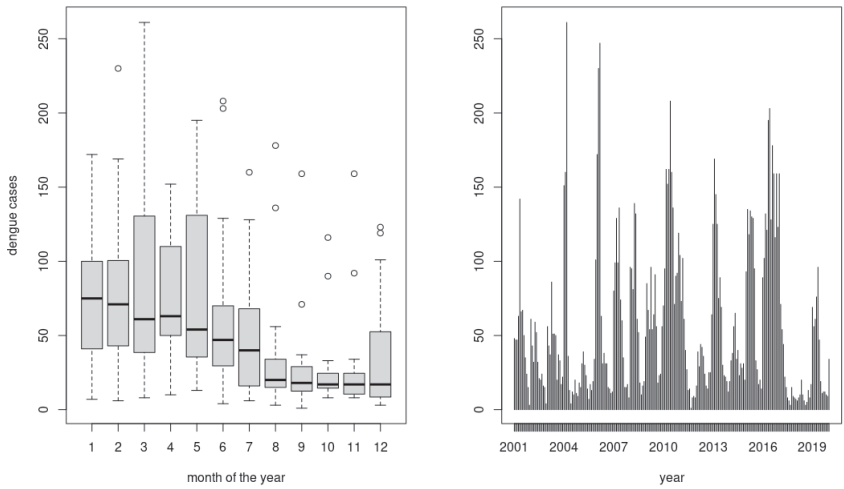


Figure 4. The time-series aggregated number of dengue cases at the district level.

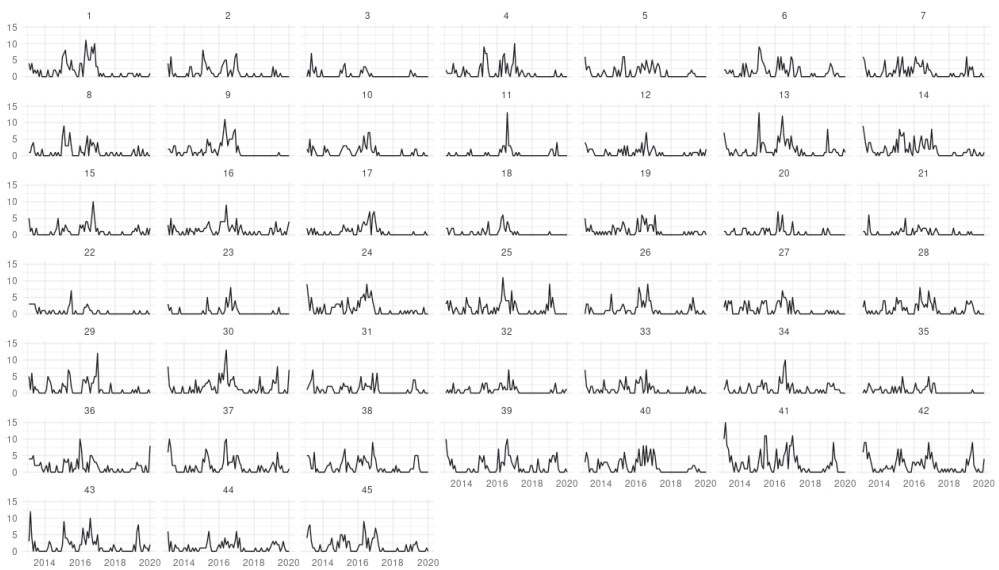


Figure 5. Notified dengue cases for the 45 urban villages in Yogyakarta, Indonesia, between February 2013 and January 2020.

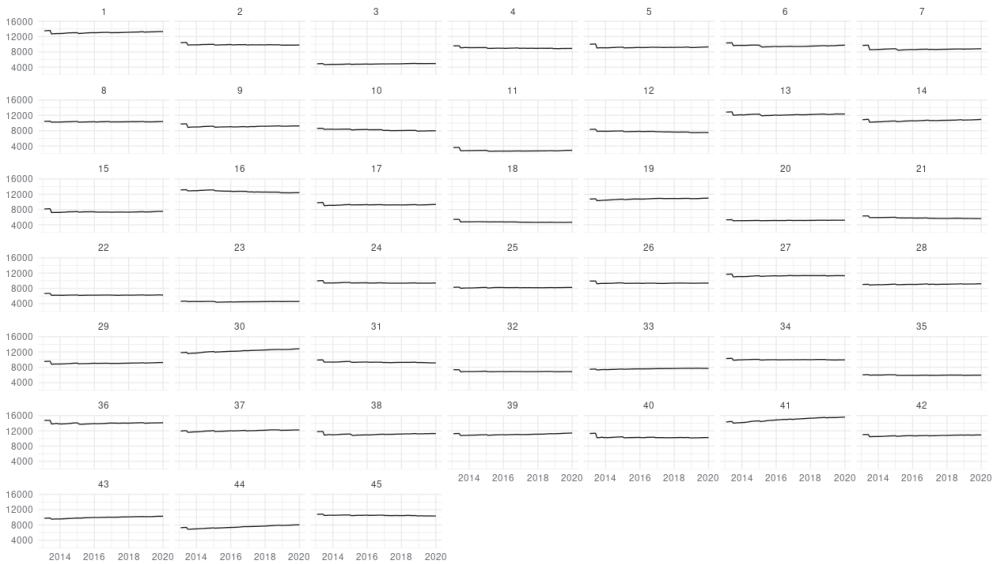


Figure 6. Population for the 45 urban villages in Yogyakarta, Indonesia, between February 2013 and January 2020.

Meteorological data

Meteorological data from January 2001 to January 2020 were obtained from the Indonesian Agency for Meteorology, Climatology, and Geophysic, namely records of cumulative monthly rainfall (mm) and mean monthly temperature readings (°C). There were missing observations; To address missingness, we replaced missing observations with the monthly mean values of the observations of the month of the year for each variable (**Figure 8**).

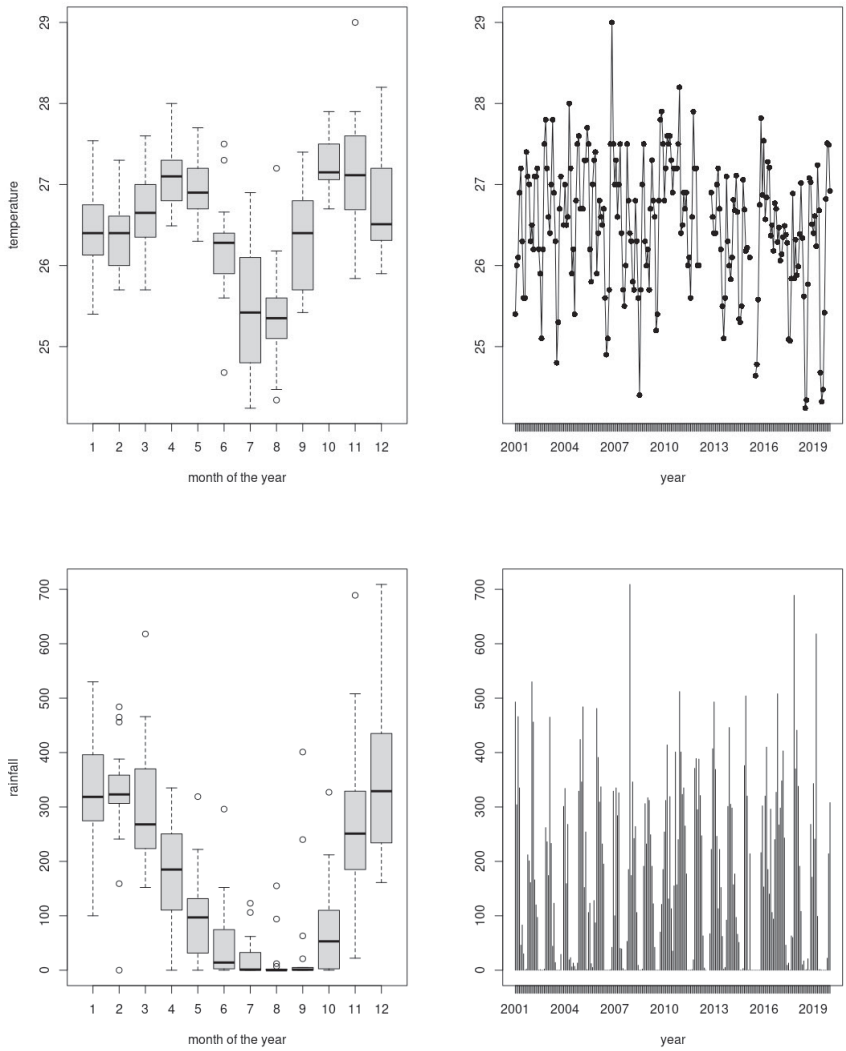


Figure 7. Time-series graphs of meteorological data.

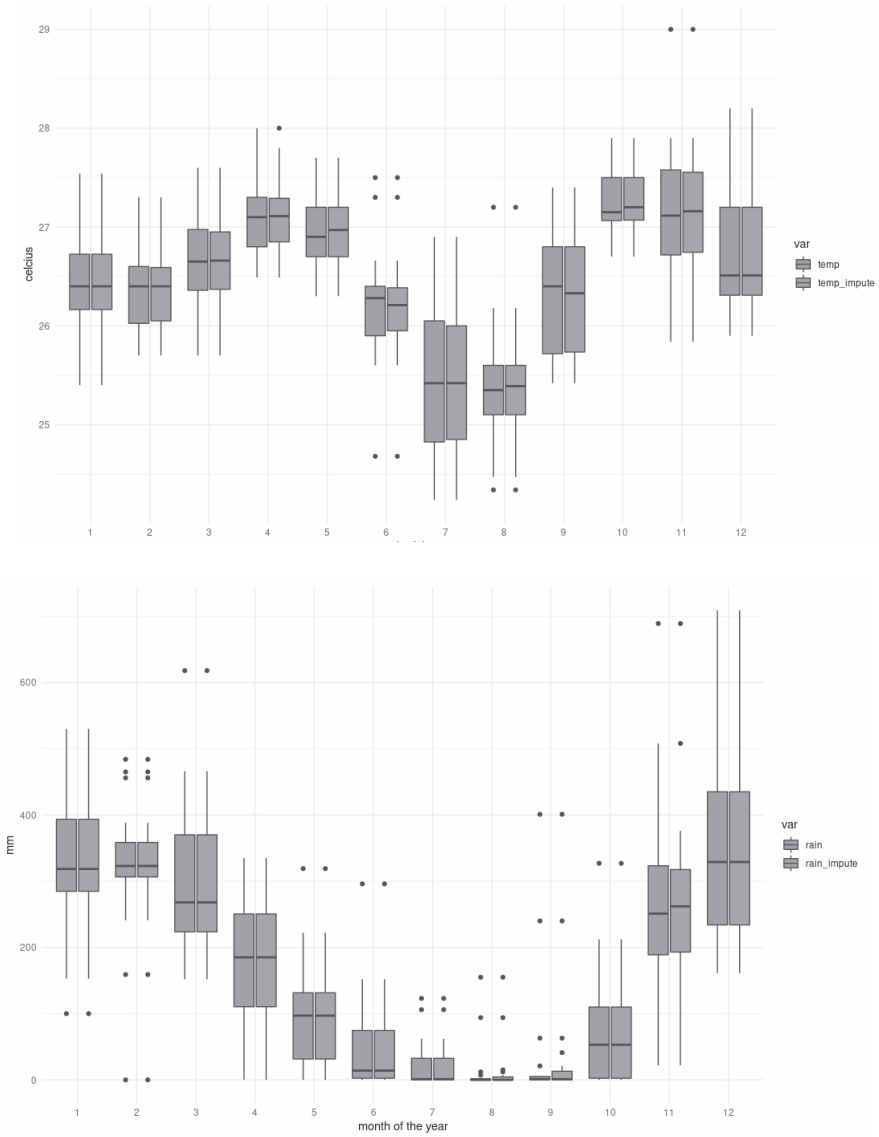


Figure 8. The comparison of monthly weather observations between variables with and without missing observations replacement.

Chikungunya data

Chikungunya outbreak reports were compiled by the Epidemic Intelligence team at the European Centre for Disease Prevention and Control. The reports were compiled through data mining from the World Health Organization, Ministries of Health, and other official and non-official sources, such as media reports. We used these reports to survey the current worldwide chikungunya situation (www.ecdc.europa.eu/en/all-topics-z/threats-and-outbreaks/weekly-threats-reports-cdtr).

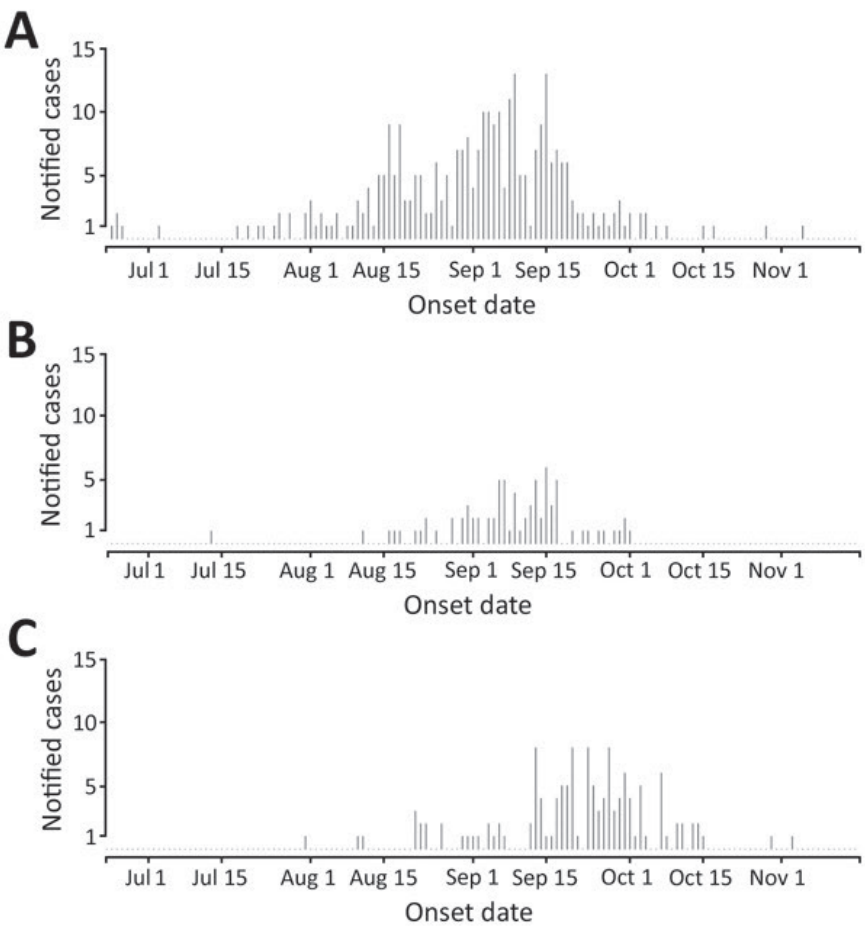


Figure 9. Notified chikungunya cases in the Anzio (A), Rome (B), and Calabria (C) regions, June–October 2017. Figure from Paper 4.[44]

Mobility data

We employed Twitter's Application Programming Interface and selected Tweets within the study areas for analysis. We only extracted the user identification string, timestamp, and longitude and latitude of the user's location in the Tweets. Subsequently, the geotagged tweets were overlaid on the administrative map of the study areas. We exchanged the geocode to the region's identification number (**Figure 2,3**).

Analysis

Influence of temperature and rainfall

The initial analyses of the meteorological variables and their association to the dengue incidence patterns over time were established using all lags of 0–3 months for temperature and rainfall in the model at the same time. This approach was taken to disentangle the contributions of various lag influences of the same meteorological variable. The patterns were studied in relation to disease counts, assuming a quasi-Poisson distribution and a log link function of dengue count data allowing for over-dispersion. The generalized linear regression models fitted allowed non-linear relationships between the weather variables and the dengue outcome variable by using natural cubic spline functions with 3 degrees of freedom per variable. The regression model can be expressed as follows:

$$\log(D_{0,t}) \sim \alpha + \sum_{l=0}^{3} ns(TMP_{lt}, df = 3) + \sum_{l=0}^{3} ns(PRE_{lt}, df = 3) \quad (1)$$

log() denotes the natural logarithm, t corresponds to the time in months, D corresponds to the dengue count, the first index, l, denotes the lag variable, *ns()* denotes a natural cubic spline, and TMP_t and PRE_t correspond to the readings of temperature and rainfall, and df denotes the degrees of freedom for the natural cubic spline function.

Influence of past dengue counts

In order to build models that could use recent and historic disease surveillance data in the predictions, we assessed first how lags of past dengue counts affected dengue transmission. Such relationships can be seen as a variant of autoregressive patterns in time series. This part of the analysis aimed at estimating how the most recent transmission influenced the dengue counts using the same lags as for the meteorological variables of 1–3 months. We excluded lag 0 as it is, in fact, the outcome variable.

Prior studies have shown that the recurrence of dengue heterotypic cases in a spatial location may be reduced up to years.[67] In a similar way, our interpretation of the reverse relationship between dengue up to 24 months back and current transmission risk may indicate a similar mechanism on a population level through cross-immunity. In general, this pattern may also explain and mediate some of the complex inter-annual cyclic patterns of dengue observed in endemic settings. These analyses set out to study if longer lag times up to four years prior might explain the variability in disease transmission dynamics. The purpose of this analysis was to assess and estimate the change in disease transmission risk based on the transmission rate in the past 1 to 48 months. We hypothesized that this would function as a proxy for the susceptibility of the population to dengue infections. To assess this, we used distributed nonlinear lag models in the DLNM[68] package of the R[69] statistical software. The variable component of the DLNM function used the natural cubic spline function, while the lag relationships were estimated using polynomials. The model established was a generalized linear regression model with a Poisson log link function that allowed for over-dispersion. The model is described below:

$$\log(D_{0,t}) \sim \alpha + \sum_{l=1}^{47} DLNM(D_{l,t}, df_{variable} = 4, df_{lag} = 4) \quad (2)$$

log() denotes the natural logarithm, *t* corresponds to the time in months, *D* corresponds to the dengue count, the first index, *l*, denotes the lag of the dengue count variable, DLNM() denotes a distributed nonlinear lag model, and *df* denotes the degrees of freedom for the *D* and the lag function.

Influence of population mobility

Fine-scale (Paper 2)

In Paper 2, we formulated an algorithm to estimate a dynamic mobility-weighted incidence (MI) index, quantifying the level of exposure to dengue virus in any given urban village due to importation from other urban villages in the study area. The MI index was calculated based on Twitter users' mobility patterns between pairs of urban villages. The mobility patterns were computed by estimating the rate with which a Twitter user in one urban village re-tweeted in another urban village within the same month. Based on this information, we generated a monthly matrix (I_t) measuring the cumulative number of mobility events between each pair of urban villages at time *t*, in months.

We created the monthly mobility network ($N_{i,j,t}$) of urban villages by multiplying (I_t) with its transpose (I_t^T). We set the diagonal of the 45×45 matrix of the affiliation to zero. Following that, we standardized the monthly mobility matrix ($N_{i,j,t}$) by dividing it by the total number of mobility events observed at time t (N_t) for all the urban villages. This ensured that, at a fixed time, the mobility matrix would always sum to 1. We referred to the standardized mobility matrix as $\check{N}_t$. To capture the total exposure due to incoming mobility into each urban village, j, we aggregated the standardized mobility from all the 45 urban villages over one month and referred to this as the TW_{jt}. The TW index is thus a time-dependent vector of length 45.

We further constructed a new matrix by multiplying the standardized mobility, $\check{N}_{ijt}$, by the vector of the number of dengue cases reported in each urban village i (of outgoing mobility), and we referred to this index as importations and computed it as $I_{ijt} = \check{N}_{ijt} \times Den_{it}$. Lastly, to capture the total exposure to the dengue virus imported due to human mobility into each urban village, j, from all the other urban villages, i, we aggregated the importations, I_{ijt}, from all the 45 urban villages over one month and referred to this as the MI_{jt}. The MI index is thus a time-dependent vector of length 45.

We investigated the association between dengue incidence and the Den, TW, and MI variables using a Bayesian spatiotemporal modeling framework assuming a Poisson distribution of the monthly counts in each neighborhood. In the model, we estimated and adjusted for the spatial covariance between neighborhoods using an unstructured spatial covariance matrix. We further adjusted for the influence of the variability in population size across neighborhoods by offsetting population size. Thus, the regression analysis assessed the predictors of the incidence of dengue. We implemented the models using the Integrated Nested Laplace Approximation (R-INLA) package.[70–72]

In the regression model, we started out by investigating how much of the variability in the dengue counts could be explained by the spatial covariance and intercept model only (the null model), leaving out all predictor variables. Subsequently, we included the MI, TW, and Den variables one lag at a time (crude), and then all lags 1 to 6 months simultaneously, but only one variable at a time. For variables showing important prediction skills, we also analyzed their combined predictive ability. The models were fitted with all lags in the same model, but with only one of the MI, TW, and Den variables at a time. The model structure can be described as:

$$y_{it} = Poisson(\lambda_{it});\ \lambda_{it} = E_{it}\,\rho_{it} \qquad (3)$$

$$log(\rho_{it}) = \eta_{it} \quad (4)$$

$$\eta_{it} = bo + \Sigma \beta_k z_{i(t-k)} + u_i + v_i + log(p_i) \quad (5)$$

The terms u_i and v_i are the spatial effects, representing unspecified features of neighborhood i that display and do not display spatial structure,[72] respectively. The k indicates the lag in months and takes values from 1 to 6. The z corresponds to the variables MI, TW, and Den. The coefficient β_k represents the regression coefficients for the variable z at lag k. The p_i variable offsets the population size of neighborhood i.

Coarse-scale (Paper 4)

In this fourth paper, we quantified the potential risk for onward transmission and spread of the chikungunya virus from the initial outbreak foci in Europe to other areas during July, August, and September, by computing users' mobility patterns. We filtered the collected tweets based on location using geocodes and the administrative map (**Figure 3**). We extracted only those originating from the study area in July, August, and up to September 19, 2017. We longitudinally analyzed these collected Tweets. When Tweets from the same users could be followed by geographic coordinates, we obtained users' individual track.

We analyzed the unidirectional mobility of Twitter users by estimating the frequency of a user being observed in a specific geographic region within the administrative map and later being observed in any other region within the same month. To compute a rate, we aggregated the total number of movements in a month between any two regions and divided this by the total number of movements across all the regions. The range of all between-regions mobility values was 0–1 and added up to 1 when summarized across the regions for inbound and outbound movements. We derived this quantity as a proxy for mobility proximity between any two regions, and computed it for each month.

Areas spatial dependency (Paper 3)

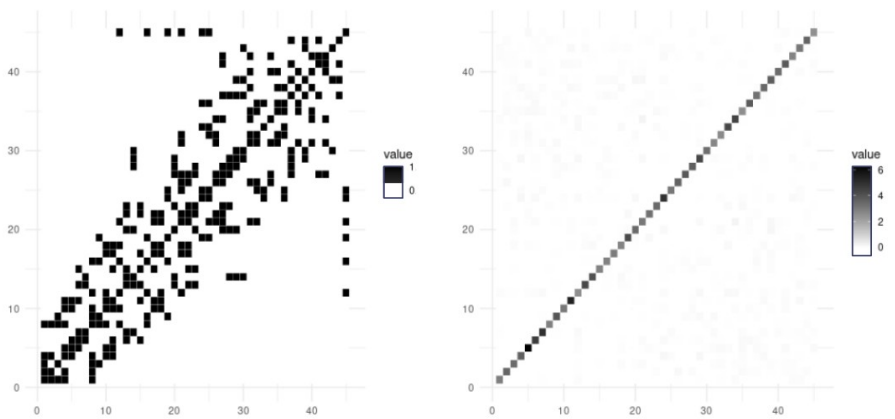
In Paper 3, we investigated two different constructions for the baseline models to estimate the most predictive spatiotemporal random effects structure. We investigated two types of adjacency matrices to represent the study area's spatial dependency among urban villages. The adjacency matrices are symmetric matrices H and $H_{ij} \neq 0$ if and only if i and j areas are affiliated. The two types of adjacency matrices are described below.

Adjacency Matrix based on the administrative map

This is an adjacency matrix based on the position of an urban village to the other urban villages in the administrative map (**Figure 2**). Here we only take into account the first-order neighbors, namely $H_{ij} = 1$ when urban villages share borders, and $H_{ij} = 0$ when urban villages do not share borders (**Figure 10a**).

Adjacency Matrix based on geotagged social media data

For this, we initially created the affiliation matrix by specifying $A_{un} = 1$ if a Twitter user u posted tweets in an urban village n; otherwise, $A_{un} = 0$. The adjacency matrix was obtained by multiplying the affiliation matrix transpose with the affiliation matrix itself, and then being standardized (**Figure 10b**).



(a) (b)

Figure 10. (a) Adjacency Matrix based on the administrative map; (b) Adjacency Matrix based on geotagged social media data. Rows and columns identify urban villages; squares identify neighbors.

Spatiotemporal model of dengue outbreak risk

We specified a spatiotemporal Bayesian hierarchical mixed model where the response consists of monthly counts of confirmed dengue cases for all 45 urban villages in Yogyakarta, over a 4-year period, from February 2013 and January 2017. Counts of dengue cases, y_{it} ($i = 1, \ldots, 45$, $t = 1, \ldots, 48$), were assumed to follow Negative Binomial distribution.

$$y_{it} \mid \mu_{it}, \kappa \sim NegBin(\mu_{it}, \kappa) \quad (6)$$

$$log(\mu_{it}) = log(p_{ia(t)}) + log(p_{it}) \quad (7)$$

where μ_{it} is the corresponding distribution mean, which is equal to the monthly population per 10 000 $p_{ia(t)}$, where $a(t)$ = 1, ..., 48, multiplied by the unknown dengue incidence rate estimate ρ_{it} for urban village i at time t, and κ is the scale parameter.[73] Population effects were accounted for by including log population per 10 000 in the model as an offset at the linear predictor scale. We investigated two different constructions for the baseline model to estimate the most predictive spatiotemporal random effects structure.

First, we investigated the Besag-York-Mollie (BYM) model for the spatial covariance based on the adjacency structure presented in **Figure 10a**. It comprises an unstructured spatial random effect to account for independent area-specific noise and a conditional autoregressive model, where the spatial effect of a particular region depends on the effects of all neighboring areas.[74] The unstructured component helps account for unknown or unobserved confounding factors, such as population immunity and local vector control interventions. Meanwhile, the structured component assumes that spatial dependency exists if villages share a border, acting as a surrogate for spatial autocorrelation that arises between nearby areas due to shared environmental or socio-economic characteristics, such as land cover and land use.

Second, we investigated a model for the spatial covariance based on the adjacency matrix informed by human mobility patterns. This rationale was based on the hypothesis that social connections shape disease transmission because of re-occuring movement patterns among the same places, such as markets, offices, schools, or homes of relatives and friends.[75] The mobility patterns were computed by estimating the rate with which a Twitter user in one study urban village re-tweeted in another urban village within the study period. Based on this information, we generated a matrix measuring the cumulative number of mobility events between each pair of urban villages.[76] In addition, we also investigated the influence of structured effects across the study period to account for time trends.

Prediction

We explored possible nonlinear and delayed associations between dengue risk and meteorological variables using distributed lag nonlinear models (DLNMs).[68] The delayed effects of the meteorological variables were assessed in one-month increments, starting from a lag of one month up to a lag of seven months to allow at least a lead time of one month to respond to an incipient outbreak. Model parameters were estimated in a Bayesian framework, using Integrated Nested Laplace Approximation (R-INLA).[69,70,73]

Subsequently, we calculated goodness of fit measures to select the best fitting model using the mean cross-validated log score. The log score measures the model's predictive power when excluding one data point at a time, and smaller values indicate better fitting models. It will therefore be aligned with the final objective of this study to find the model as the basis of an early warning system.

We then established predictions using an unseen testing dataset for the selected best performance model from February 2017 to January 2020. For this purpose, we ran the model 36 times to get monthly predictions for the 36 months of the unseen testing dataset. We compared the predicted values with observations collected for the same period. For the sake of completeness, we also calculated the difference in mean absolute error (MAE) between the baseline model and the more complex model to identify which more complex model had a better fit but also was predictive in the unseen testing dataset.

Results

The incidence of dengue was observed to be higher during the first four months of the year during the study period. The average number of cases for each month from January to April, was higher than the average monthly case over the study period (**Figure 4**). The monthly mean temperatures ranged from 24.24 to 29.02 °C. Yogyakarta experienced the lowest monthly mean temperature in July 2018 and the highest monthly mean temperature in November 2006. The highest amount of monthly cumulative rainfall was recorded at 709 mm in December 2007, with an average of 171.15 mm.

Influence of Temperature and Rainfall

Temperature showed a weak association with dengue cases at lags 0, 1 and 2 months, but at lag 3 months the relative risk increases linearly when temperature increases. Rainfall at lag 0 month is associated with a slight increase in dengue transmission at lower rainfall levels, but indicates a strong drop when the monthly rainfall is higher than 300 mm. At lags 1, 2 and 3 months, increasing rainfall is associated with a linear increase in the relative risk of dengue. But also here, the increasing relative risk of dengue is lower when the rainfall is very high, especially at shorter lag distances (**Figure 11**).

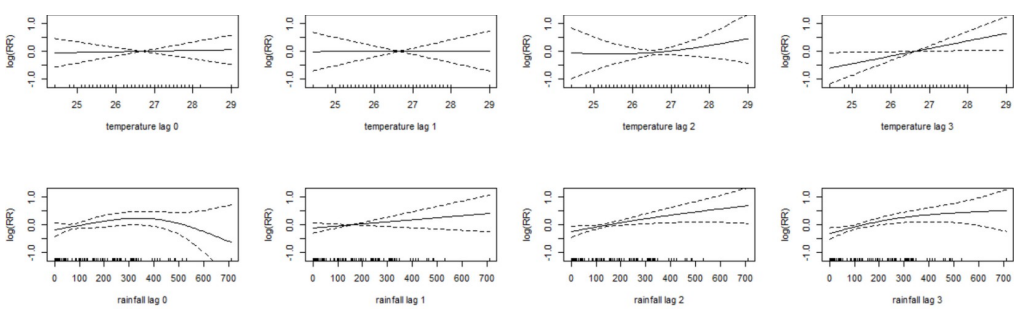


Figure 11. Association between temperature and rainfall and dengue over Lag 0–3. Figure from Paper 1.[77]

Influence of Past Dengue Counts

Dengue cases at lag 2 months show associations to increase dengue risk with lower values, but to decrease when the dengue cases is more than 150 (**Figure 12a**). The non-linear distributed lag models showed a peak of the risk around lag 24 months (**Figure 12b**) suggesting a negative feedback cyclic pattern with lower relative risks of transmission up to two years following a large outbreak in around lag 24. An increase in dengue cases counts in a specific month will increase the dengue risk in each of the following months with a peak at approximately 24 months (**Figure 12c**).

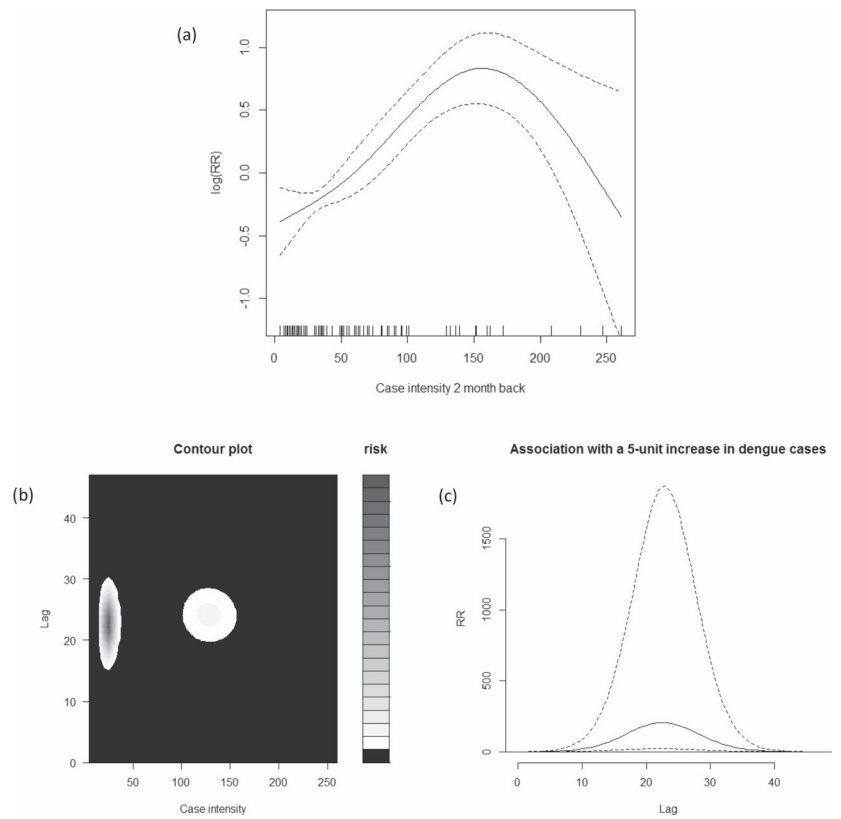


Figure 12. Relationship between autoregressive lags and dengue case counts. *Upper panel shows (a) relative risks of dengue cases as a function of dengue surveillance at a 2-month lag. Lower panel shows (b) the relationship between case intensity and dengue risk categories at all lag months considered in the analysis; and (c) the relative risk of dengue in each future exemplified here for an increase of 5 dengue cases in a specific month*. Figure from Paper 1.[77]

Influence of Population Mobility

Fine-scale (Paper 2)

The monthly mobility patterns for the 45 urban villages in Yogyakarta appeared to be relatively consistent over the study period, except that a slight increase in the number of mobility events was observed over the same period. The mobility patterns varied considerably across different neighborhoods (**Figure 13**).

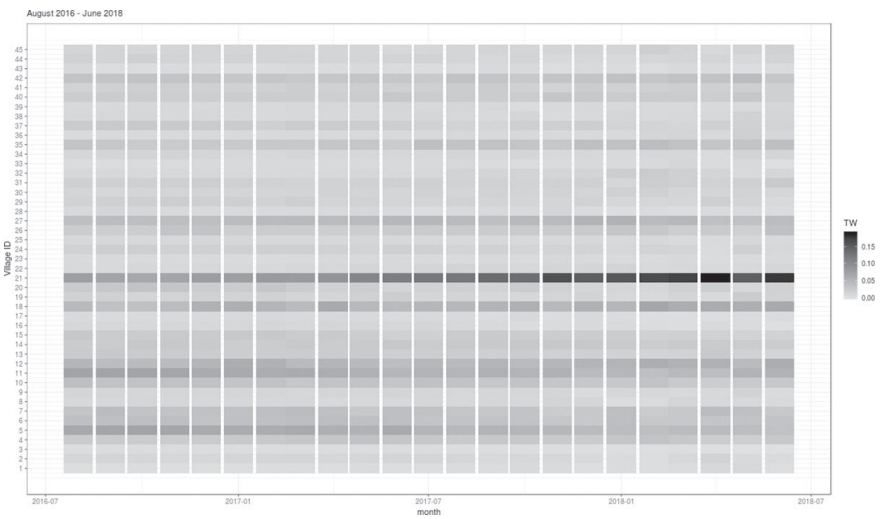


Figure 13. The TW index captures the temporal pattern of the aggregated total monthly mobility into each of the 45 urban villages in Yogyakarta, Indonesia. Figure from Paper 2.[76]

The MI index (**Figure 14**) for each neighborhood reflected a combination of the mobility estimates and the disease counts. In general, we found that the MI index was not only higher for the neighborhoods with relatively higher mobility to other neighborhoods, but also reflected the decreasing trend in the disease counts over time.

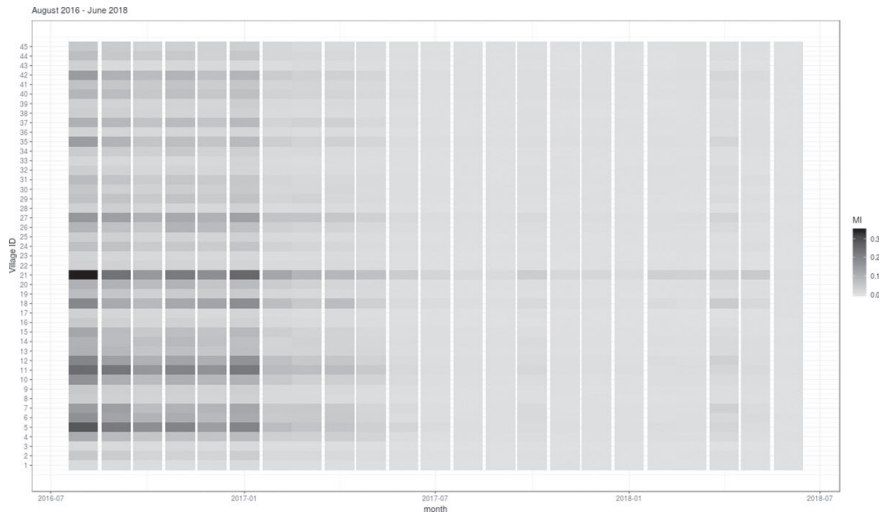


Figure 14. The MI index estimating the temporal pattern of the aggregated importations into each of the 45 urban villages in Yogyakarta, Indonesia. Figure from Paper 2.[76]

Table 1 describes the crude and adjusted model estimates of the lag effects of Den, TW and MI using the Bayesian spatiotemporal regression model. We found that the mobility and the centrality of a neighborhood proved not to be important for predicting the incidence of dengue on its own. This is shown by comparing the model fit of the TW lag variables (crude and adjusted) to the null model and their observed lack of difference in the R-square, BIC and SRMSE. In contrast, the Den and MI variables provided important information for predicting the incidence of dengue at lag times 1 to 6 months based on the crude and adjusted estimates of the model.

The coefficients from the crude and adjusted models are graphically presented in **Figure 15**. The uncertainty and confidence intervals for the coefficient estimates increased in the lag adjusted models compared to the crude single lag models. Overall, the coefficients were smaller in the adjusted models. This is because of the similarity of information carried over in lags of a specific variable, i.e. due to temporal covariance. In the adjusted models, we observed a decreasing pattern in the association to the Den and MI variables with increasing lags, with the exception that both peaked at lag 3 months. While most lags associated with the Den variable showed statistically significant associations, the associations with the MI variable were more uncertain, with the exception of at lag 3 months. However, since the SRMSE was lower for the MI model, it appeared that this variable still

included more vital information for predicting dengue incidence in the neighborhoods.

Table 1. Model fitting statistics and coefficients. Table from Paper 2.[76]

Variable	R-sq	BIC	SRMSE	fixed effects			
				coefficient mean	coefficient sd	2.5 percentile	97.5 percentile
Null model	0.064	1337.855	0.882	-10.0739	0.0668	-10.2084	-9.9463
Crude lag estimates							
Den lag1	0.098	1246.184	0.865	0.2472	0.0196	0.2075	0.2844
Den lag2	0.123	1244.557	0.854	0.2385	0.0194	0.1993	0.2757
Den lag3	0.164	1213.810	0.833	0.2410	0.0178	0.2054	0.2754
Den lag4	0.090	1262.592	0.869	0.1996	0.0191	0.1612	0.2363
Den lag5	0.105	1270.408	0.862	0.1847	0.0191	0.1465	0.2215
Den lag6	0.080	1278.772	0.874	0.1675	0.0186	0.1304	0.2034
TW lag1	0.065	1337.276	0.881	4.2158	2.9655	-1.8731	9.8080
TW lag2	0.065	1337.581	0.881	4.0453	3.0282	-2.1836	9.7437
TW lag3	0.065	1337.596	0.881	4.2636	3.1498	-2.1978	10.2092
TW lag4	0.066	1337.292	0.881	4.7574	3.2264	-1.8396	10.8690
TW lag5	0.066	1336.294	0.881	5.7825	3.2885	-0.9000	12.0587
TW lag6	0.066	1336.181	0.881	6.0791	3.3850	-0.7870	12.5534
MI lag1	0.086	1237.947	0.871	14.7771	1.2940	12.2140	17.3077
MI lag2	0.132	1224.506	0.849	15.1857	1.3731	12.5236	17.9179
MI lag3	0.197	1198.824	0.817	15.6868	1.4773	12.8975	18.7013
MI lag4	0.174	1221.191	0.828	14.4542	1.4724	11.6631	17.4517
MI lag5	0.172	1220.364	0.829	13.8227	1.4964	11.0159	16.8935
MI lag6	0.137	1230.597	0.847	11.1021	1.1871	8.8499	13.5165
Adjusted lag estimates							
Den lag1	0.201	1168.257	0.815	0.1080	0.0288	0.0503	0.1634
Den lag2				0.0689	0.0300	0.0090	0.1267
Den lag3				0.1136	0.0282	0.0574	0.1681
Den lag4				0.0544	0.0266	0.0015	0.1058
Den lag5				0.0326	0.0263	-0.0193	0.0839
Den lag6				0.0526	0.0247	0.0035	0.1005
TW lag1	0.068	1340.723	0.880	0.1794	12.6758	-24.7466	25.0110
TW lag2				-9.0344	14.7070	-38.1347	19.6128
TW lag3				-10.4439	15.8923	-41.8313	20.5637
TW lag4				-4.4227	15.8500	-35.7277	26.5013
TW lag5				15.4022	15.9844	-16.1796	46.5789
TW lag6				15.9111	14.9473	-13.5140	45.1656
MI lag1	0.217	1194.667	0.807	5.6002	3.7329	-1.6199	13.0294
MI lag2				4.6889	3.3112	-1.8357	11.1655
MI lag3				8.4452	3.6902	1.2679	15.7461
MI lag4				6.1895	3.8815	-1.6790	13.5752
MI lag5				2.0293	4.3635	-6.9640	10.1850
MI lag6				-4.8919	4.0121	-13.1124	2.6493
Combined variable model							
Den lag1	0.271	1140.838	0.778	0.0816	0.0302	0.0213	0.1398
Den lag2				0.0485	0.0302	-0.0118	0.1068
Den lag3				0.0964	0.0279	0.0408	0.1505
Den lag4				0.0510	0.0273	-0.0034	0.1040
Den lag5				0.0265	0.0269	-0.0267	0.0788
Den lag6				0.0445	0.0246	-0.0045	0.0921
MI lag1				3.2332	3.9101	-4.3982	10.9449
MI lag2				3.8463	3.6771	-3.4612	10.9795
MI lag3				7.0018	3.9220	-0.7220	14.6739
MI lag4				2.9156	4.1978	-5.6567	10.8263
MI lag5				0.5181	4.3102	-8.3993	8.5290
MI lag6				-4.5618	3.9001	-12.5642	2.7519

R-sq = R square, BIC = Bayesian Information Criterion,
SRMSE = standardized root mean square error, coefficient mean = log(relative risk),
coefficient sd = standard error, 2.5 percentile = lower end of 95% credible interval of coefficient,
97.5 percentile = higher end of 95% credible interval of coefficient

The MI variable at lag 3 months had clearly the best predictive ability and almost the same predictive ability as the adjusted models with all lags, considering the R-square, BIC and SRMSE values (**Table 1**). The Den variable at lag 3 months did not show a similar good performance with significantly lower predictive ability, R-square, BIC, and SRMSE.

The model including both the Den and MI variables at lags 1 to 6 months estimated an R-square, BIC and SRMSE of 0.271, 1140.8, and 0.778, respectively, and showed considerably higher predictive ability compared to the adjusted models of the Den and MI variables alone (**Table 1**). These findings support the fact that these variables contributed different information to the predictive ability of the model. Looking at the coefficients in this combined model, the estimates were not very different from those obtained from the adjusted single variable model estimates, confirming the exclusive unique contribution of these two variables to the predictive ability.

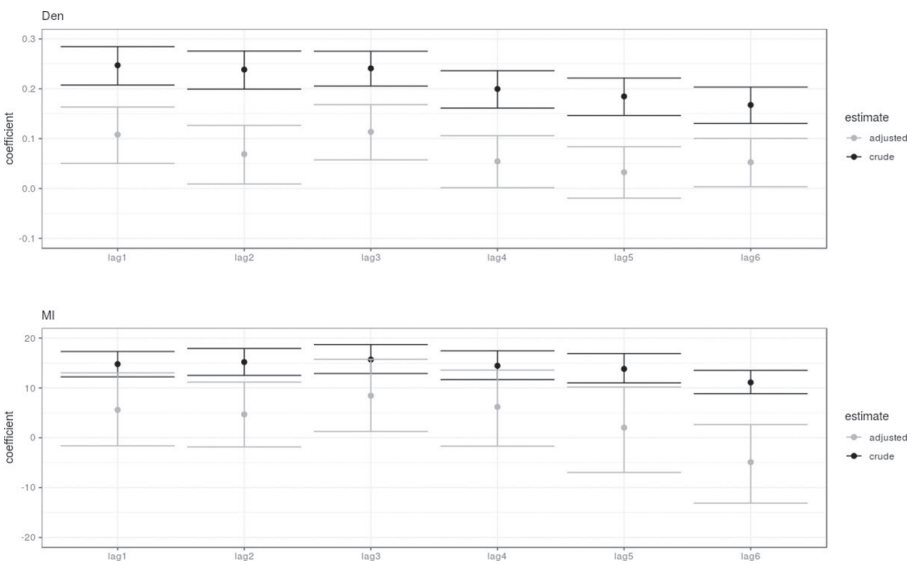


Figure 15. The crude and adjusted coefficients for the Den and MI models for lag times 1 to 6. Figure from Paper 2.[76]

Coarse-scale (Paper 4)

The spatiotemporal analysis of geocoded Twitter data showed strong population mobility from Lazio Region in Italy and Var department in France towards several larger cities where Ae. albopictus mosquitoes are present. The top 10 estimates of mobility out of the 2 outbreak zones of Var and Lazio showed the strongest pattern for potential dispersion of the chikungunya virus not only into the areas geographically close to the outbreak zones but also to several relatively large cities in Italy, France, and Spain (**Table 2**).

Table 2. Top 10 areas where mobility proximity to the 2 chikungunya outbreak zones was highest, Europe, August 2017. Table from Paper 4.[44]

	Southern Europe		Lazio region	
Rank	From Var department, France	From Lazio region, Italy	From Anzio	From Rome
1	Alpes-Maritimes	Florence	Roma	Vatican
2	Bouches-du-Rhône	Milano	Nettuno	Fiumicino
3	Torino	Napoli	Sabaudia	Sabaudia
4	Paris	Venezia	Ardea	Civitavecchia
5	Alpes-de-Haute-Provence	Paris	Civitavecchia	Santa Marinella
6	Rhone	Barcelona	Pomezia	Tivoli
7	Hérault	Perugia	Aprilia	Anzio
8	Vaucluse	Latina	Cisterna Di Latina	Ladispoli
9	Barcelona	Siena	Fondi	Pomezia
10	Baeleares	Salerno	Amatrice	Valmontone

The monthly mobility patterns during the study period varied; for example, the vacation month of August showed a stronger mobility pattern out of Var to areas not in direct connectivity, most notably to Rome. When we contrasted the mobility proximities between the 2 outbreak zones, we observed the highest proximities within the two countries. In Italy, cases were first notified in Anzio at the end of June, followed by notifications in Rome later in July, and in Calabria in early August in order of temporal appearance (**Figure 9**).

A closer look at the Lazio outbreak zone in Italy revealed strong connectivity between Anzio (where the first cases in Italy were confirmed) and Rome

(where a higher number of cases were notified). We compiled the top 10 mobility proximity areas from the outbreak zones of Anzio and Rome in August and September (**Table 2**). Although the highest mobility proximity from Anzio was to Rome in August and September, the mobility proximity from Rome to Anzio was also found among the top 10 destinations. Overall, Rome had higher connectivity to many more areas than Anzio.

Spatiotemporal Model of Dengue Outbreak Risk

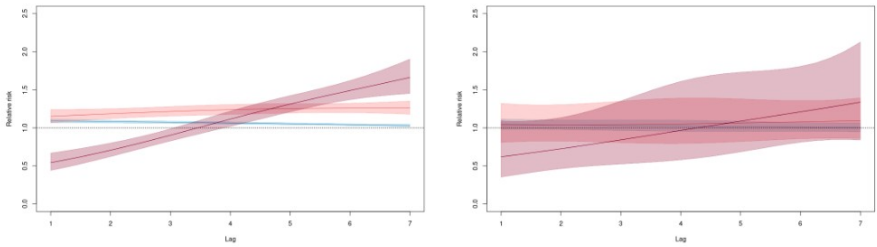
The models including spatio-temporal random effects using the adjacency matrix based on the geographical location of urban villages (BYM model), showed poorer performance when compared with the models using the adjacency matrix based on human mobility patterns proxied from the geolocated tweet data (**Tabel 3**). The latter model is henceforth referred to as the Base Model.

Expanding the Base Model by including population data, as well as temperature in DLNMs, resulted in a further reduction in the log score compared with the Base Model (**Table 3**). Therefore, we found that the best-fitting model assessed by the log score was the model with spatio-temporal random effects using the adjacency matrix based on human mobility patterns, population, and temperature lag 1 to 7 months in DLNMs.

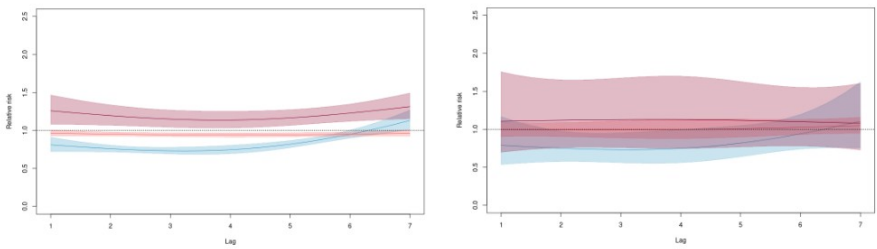
Table 3. The performance of the spatiotemporal models.
Table from Paper 3.

Model	Description	log score
BYM	Adjacency matrix based on geographical location of urban villages	1.590
MOB (Base Model)	Adjacency matrix based on the human mobility patterns proxied from the geolocated location of Twitter users	1.577
DLNM Temp	Distributed lag non-linear models for temperature including lags for 1-7 months. The model also included population size.	1.651
DLNM Rain	Distributed lag non-linear models for rainfall including lags for 1-7 months. The model also included population size.	1.733
Base Model + pop	Adjacency matrix based on the human mobility patterns proxied from the geolocated location of Twitter users. The model also included population size.	1.713
Base Model + pop + DLNM Temp	Adjacency matrix based on the human mobility patterns proxied from the geolocated location of Twitter users in combination with distributed lag non-linear models for temperature including lags for 1-7 months. The model also included population size.	1.568
Base Model + pop + DLNM Rain	Adjacency matrix based on the human mobility patterns proxied from the geolocated location of Twitter users in combination with distributed lag non-linear models for rainfall including lags for 1-7 months. The model also included population size.	1.584

In the crude model, the median temperature values influenced the whole lag period. In contrast, higher temperatures showed a strong linear association with dengue risk: lower risk at the short lag, and higher risk at the longer lag, starting after around four months lag. Meanwhile, dengue risk dropped gradually after high rainfall and the relative risk of dengue was lowest at lag 4 months. Adjusting for the mobility covariance in the model altered the relationship between dengue risk and rainfall. The risk rose slowly after high rainfall and reached the peak at 4- months lag. In addition, the median rainfall values also seemed to influence the whole lag period, but a higher risk was observed when adjusting for the mobility covariance in the model (**Figure 16**).



(a) Temperature crude model (b) Temperature adjusted model

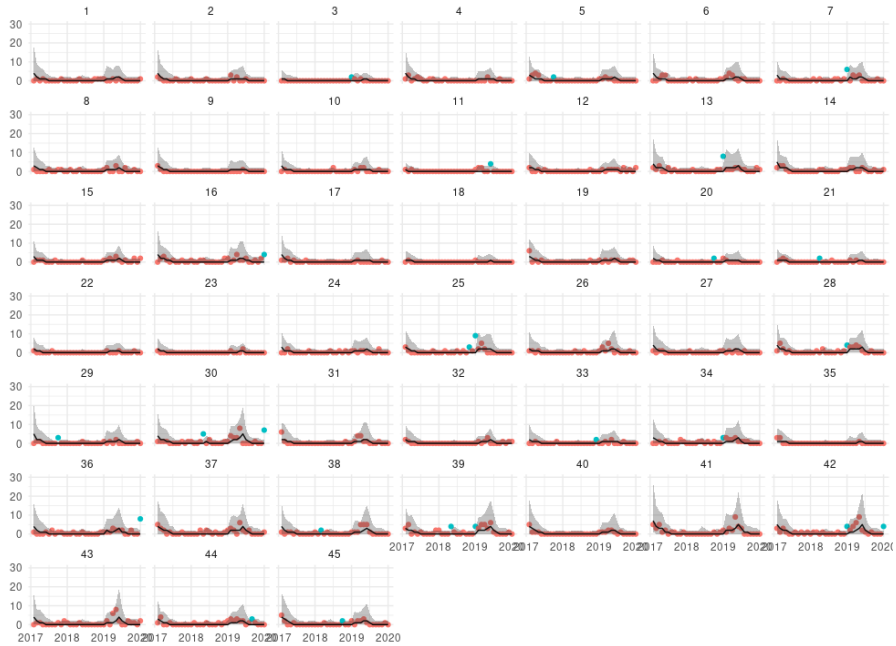


(c) Rainfall crude model (d) Rainfall adjusted model

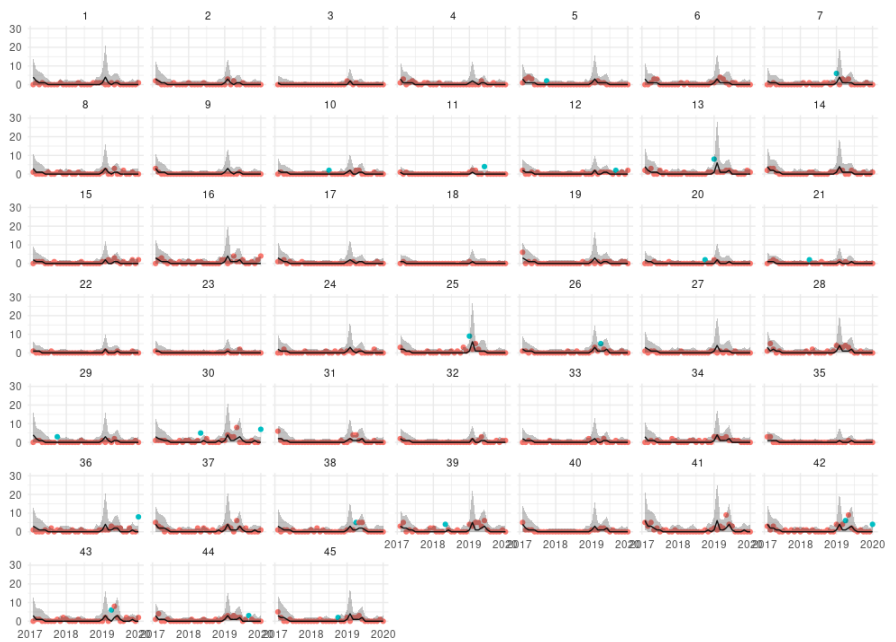
Figure 16. Association between dengue risk, temperature, and rainfall at different lag times: *the blue line represents Q1 value, the orange line represents Q2 value, the red line represents Q3 value.* Figure from Paper 3.

Predictive Performance of the Models

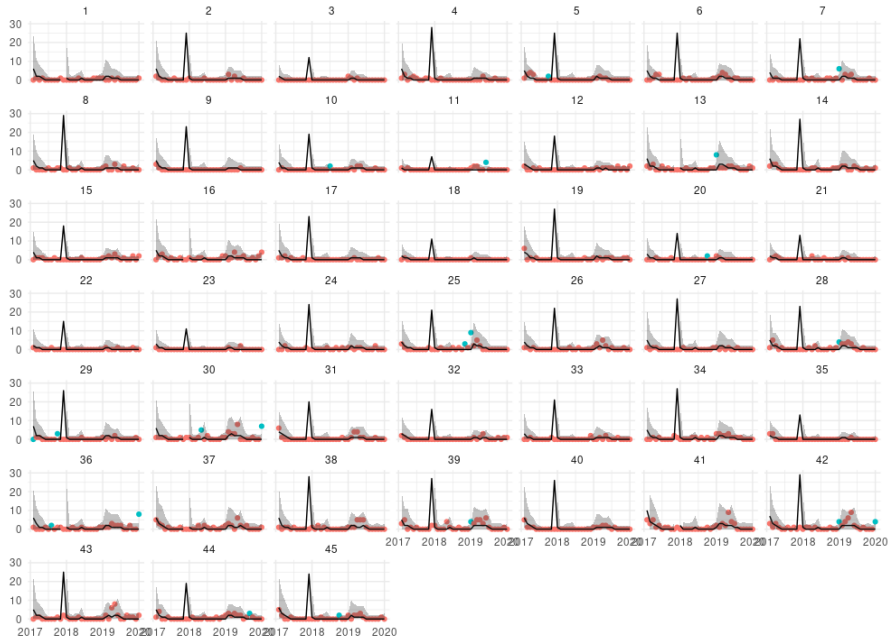
The estimated relationships in the best-fitted model (Base Model + pop + Temp) were used for the forecast unseen data from February 2017 to January 2020 to mimic an actual situation of an early warning system in use. We ran the model 36 times to predict the number of dengue cases one month ahead at the village level each month. We observed that very few observations lay outside the credible interval of predictions (**Table 4**). The model using rainfall, actually, had more observations within the credible interval of predictions compared to the other models. However, this model seemed sensitive and unstable in handling climate variability. At the end of November 2017, Tropical Cyclone Cempaka caused extreme rainfall in the study area and provided extreme predictor values for rainfall. It subsequently yielded unrealistic predicted values (**Figure 17c**) and severely impacted the MAE (**Table 4**). Nevertheless, the best-fitted model provided predicted values with a more narrow credible interval than the credible interval provided by the other models (**Figure 17b**). Furthermore, the best-fitted model yielded more urban villages with the lowest MAE (**Figure 18**).



(a) Base Model + pop



(b) Base Model + pop + Temp



(c) Base Model + pop + Rain

Figure 17. The comparison of posterior predictive median and 95% credible interval to observed cases. *The black line represents median values and the gray shading represents 95 % credible intervals. The dots represent the observed values: red means the values lie between credible intervals, and blue indicates otherwise.*
Figure from Paper 3.

Table 4. The predictive performance of the spatiotemporal models.
Table from Paper 3.

Model	**% within the CI***	**MAE**
Base Model + pop	98.5 (1596/1620)	0.535
Base Model + pop + Temp	98.7 (1599/1620)	0.510
Base Model + pop + Rain	98.8 (1600/1620)	1.197

*percentage of times the observed incidence is within the credible interval

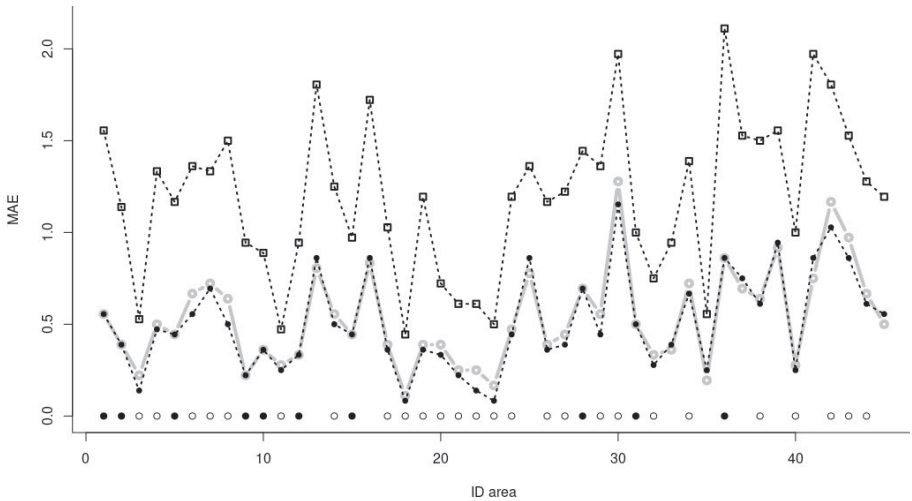


Figure 18. MAE for each urban village.

Greyline with white circle points represents Base Model; The dashed line with black circle points represents Base Model + pop + Temp; The dashed line with white square points represents Base Model + pop + Rain; On the bottom, white circle points mean Base Model + pop + Temp has the lowest MAE values, while black circle points mean Base Model + pop + Temp has the same MAE values with Base Model.

Figure from Paper 3

Discussion

Influence of Climate Variability

The observed relationship between weather and dengue risk in this study is consistent with previous research studies suggesting a 4-month delay.[78,79] Our study shows that temperature at 3-month lead-time and rainfall at 2 and 3-month lead-time are the best predictors of dengue transmission patterns in Yogyakarta Municipality.[77] A similar lag between rainfall and temperature and dengue risk has been shown in other provinces of Indonesia and neighboring countries.[27,30,78]

Weather conditions in the study area appeared to be highly suitable for maintaining dengue transmission, but climate variability resulting in periods of drought or extreme rainfall can potentially boost or limit transmission.[80] In the context of vector-borne diseases, climate variability greatly influences disease transmission patterns through the path of vector biology. A higher or longer rainy season is likely to generate more vector-breeding sites and sustain them,[31–33] while higher temperatures can influence vector-virus relationships through vector longevity, mating, dispersal, feeding behavior, oviposition, and more rapid virus replication.[26,35]

Climate variability events also affect human behavior, such as use of additional containers to reserve water in the dry season or reduction in mobility during the rainy season. Numerous artificial containers can become potential breeding sites for Aedes mosquitoes. On the hand, heavy rain can flush breeding sites and potentially destroy immature mosquitoes and, in turn, negatively affect vector abundance.[32,34] In addition, Aedes aegypti and Aedes albopictus, as the primary vectors of dengue,[23] adapt well to human environments, and indoor activities could create conducive conditions for vector proliferation and hence dengue transmission.[25]

Extreme weather events can potentially affect the model's predictive performance by providing extreme input values for the meteorological predictors that can invalidate the underlying assumptions of the model. It, subsequently, may yield extreme or unrealistic predicted values. At the end of November 2017, a tropical cyclone caused extreme rainfall and floods in the study area and its surroundings. Subsequently, in terms of dengue forecasting, the model predicted unrealistically high dengue cases when this excessive rainfall was inputted as the predictor.

Influence of Population Mobility on Disease Spread

Our study revealed insights into how the intra-urban outbreak risk relates to the combination of population mobility and the size of local outbreaks. Population mobility has been shown to influence the spread of dengue. The transmission appears to be shaped by social connections because of routine movements among the same places, such as commuting and visiting family and friends.[75]

The MI indicator proposed in this thesis integrated population mobility proxies, derived from analyzing geolocated twitter data, with disease surveillance data. We demonstrated its ability to predict dengue incidence at the intra-urban village level and found that the impact of population mobility on disease spread cannot be effectively studied without combining mobility information with disease incidence data. Mobility does not necessarily translate into a greater exposure to the circulating virus unless it is combined with disease incidence information.

An earlier study in the same study area found no evidence of spatial autocorrelation in models with dengue incidence at the village-level.[22] This was possibly because the study was mainly focused on children, who spend most of their time at or near home. In addition, in this earlier study, dengue data were aggregated annually.[22] However, another study found that even though individuals preserved their number of interactions when mobility enhanced, the epidemic could still emerge.[81] As measured by population mobility, human behavior emerges as a key factor in understanding the spread of disease, and it is one of the many aspects of vulnerability infectious disease modeling.[82]

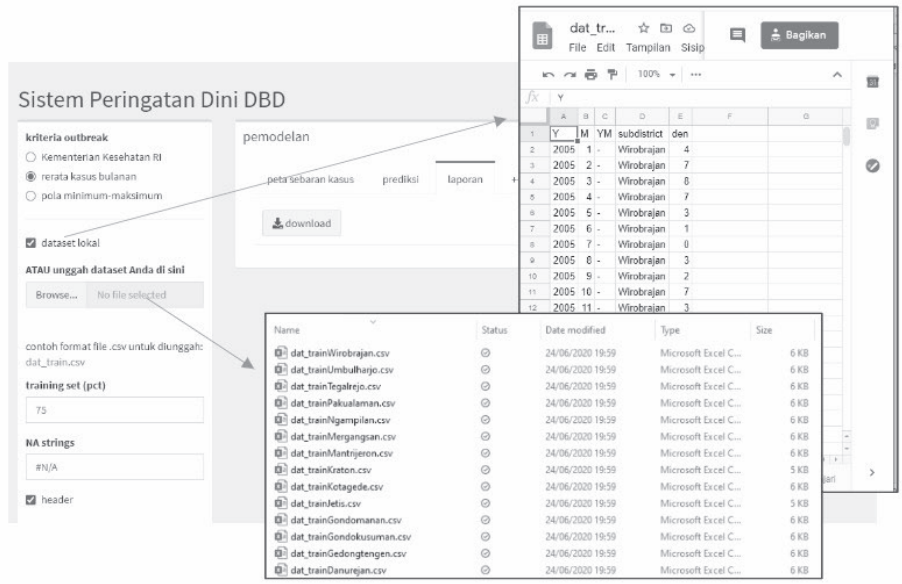
Establishing proper adjacency relationships between a set of spatial units under investigation is essential when conducting spatial or spatiotemporal analysis. In many cases, researchers have chosen some of the most typical neighborhood structures, such as the first-order contiguity matrix based on shared borders, that are suitable for many situations.[83] In this study, we developed a spatiotemporal model using an adjacency matrix based on human mobility patterns proxied from geotagged Twitter data. As a novel data source, geotagged Twitter data can provide important information on mobility patterns and estimate the level of exposure to virus importation in any given neighborhood.[76] One of the strengths of this data source is the near real-time availability, which are timelier and more accessible, and less costly than mobile phone or air passenger data available from commercial providers. This approach can identify areas of heightened mobility that are potentially at risk for onward transmission.

The Implementation of Dengue Early Warning Systems

Surveillance systems are widely used to support public health efforts, but they fail to promote efficiency since spatial prediction remains a limitation, with no tools to map high-transmission areas at a fine spatial level.[59] The method introduced here maximizes the use of surveillance information collected. In addition, novel data streams, such as web search data or social media updates, hold promise for enhancing the capabilities of public health surveillance since they can provide robust, long-term surveillance solutions.[84] While this research focused on dengue, we see a significant potential in expanding the method to similar urban areas and climate-sensitive vector-borne diseases.

Previous studies have shown that the implementation of Dengue Control Programs has not produced satisfactory results due to, among others, dengue surveillance program implementation in Indonesia which relies on passive surveillance,[9] the difference in communication intensity between authorities responsible for the Dengue Control Programs, and the different levels of public knowledge about dengue in the community.[85–87] In addition, dengue vector, Aedes aegypti, has developed resistance to insecticides from the group of organophosphates, and pyrethroids carbamic.[88] Early warning based on this model might be used as a part of the advocacy process for developing cross-sectorial networking for the surveillance and control of dengue. The models allowed some recognition, although not perfect, of outbreak periods at an early stage. Thus, the development of an early warning system could benefit from these predictions since one main reason for vector control being less successful is related to the poor and reactive management and delayed deployment of interventions during epidemics.[56] Interventions may target high-risk periods, during which public health interventions can effectively prepare communities and potentially curb epidemics. Given constrained public health budgets, the methods we present are critical to the future reliability and sustainability of infectious disease surveillance systems.

Our model forecasted dengue cases up to 2 months ahead and showed a consistent ability to differentiate between months with epidemic and non-epidemic transmission.[76,77] The model presented here was initiated and developed as the foundation of Dengue Early Warning Systems in Yogyakarta Municipality's District Health Office in 2016 but has only been used operationally since 2020.



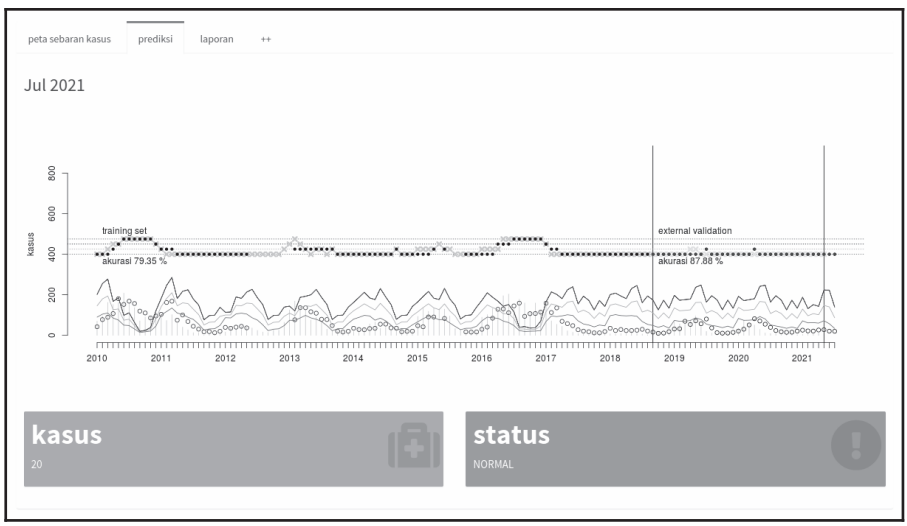


Figure 18. Web-based dashboard of Dengue Early Warning Systems used by District Health Office Yogyakarta Municipality

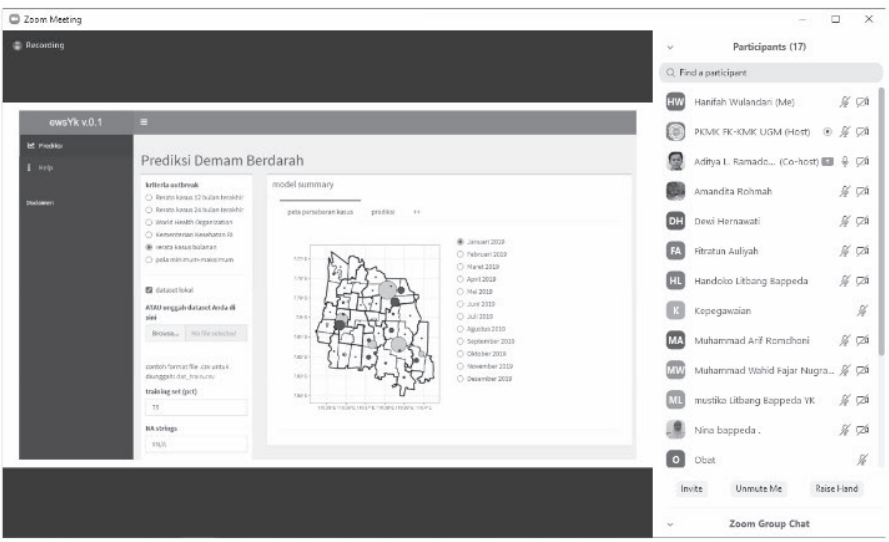


Figure 19. Workshop and dissemination of Dengue Early Warning Systems for stakeholders in Yogyakarta Municipality

Limitations

In this study, underreporting might have occurred due to asymptomatic infections, self-treated dengue cases, or cases managed at private clinics.[9] However, since the national surveillance system only detects hospitalized dengue cases, asymptomatic and non-severe dengue will not be notified. Since the asymptomatic and non-severe dengue are not included in the official dengue case reports, they do not describe all the disease's manifestations and the actual dengue burden, and our study has only based the analysis on notified severe dengue in the study area.

We used the province-level meteorological data obtained from the Indonesian Agency for Meteorology, Climatology, and Geophysics. The meteorological data were measured from a single ground weather station and are officially used to represent the weather condition for Yogyakarta Province and its surrounding. This is a limitation since it is not capturing local variability in time and space. However, an earlier study in the neighboring area that compared the total accumulated daily rainfall each week in villages at the southeast and northwest of our study site showed that the total rainfall among these tends to be highly correlated, with Spearman rank correlations varied between 0.52 - 0.88.[89] In addition, meteorological data provided by the Indonesian Agency for Meteorology, Climatology, and Geophysics is suggested reliable for the Yogyakarta setting.

Dengue reporting data at the monthly level is a limitation as a finer time resolution could have revealed more granular patterns, especially when considering the analysis of lag distributions of the associations. A finer temporal resolution could have revealed better the potential underlying processes.

We assume that user movements between consecutive tweets were representative of the overall population mobility, while in fact Twitter users represent a selected group of individuals. It is, however, important to note that the use of Twitter and other social media platforms is very common in Indonesia.[55] Nevertheless, prior studies have validated Twitter as a viable data source to study human mobility.[45,52,90] Using mobile phone data with geo-tags would have been a better alternative, although the downside is that such data are harder to acquire and use prospectively over time. Yet, human mobility patterns extracted from geotagged tweets have been reported to have similar overall features with those proxied from mobile phone records.[45]

Finally, the Wolbachia trial has been implemented in several areas in the study site. The entire Wolbachia trial site was an area of 26 square kilometers[22] or approximately 80% of the Yogyakarta municipality area. The Wolbachia intervention was started gradually from March through December 2017,[22] or inside our testing period, February 2017 to January 2020, using an unseen dataset. We believe that our currently developed spatiotemporal model is adjusting for the Wolbachia intervention in the modelled time trend functions. In fact, this is supported by the predictive ability of the models.

Conclusion

This study shows that 1) past dengue cases showed a stronger association with current dengue risk at short lags, but a cyclic pattern was also observed with longer lag times; 2) temperature in the study area is optimal for dengue transmission, and was, therefore, influential over the whole lag period; 3) higher amount of monthly cumulative rainfall was overall associated with increasing dengue risk, but the influence of population mobility could change the time-lag patterns in reaching the peak risk. Furthermore, this study also found that 4) the outbreak risk relates to a combination of population mobility and the size of local outbreaks; 5) it is possible to predict the potential spread of arboviral outbreak risk.

We predicted dengue risk patterns using disease surveillance and meteorological data. The best model for dengue prediction was one which combined meteorological and surveillance data as predictor variables. The results support that temperature and rainfall, as well as short-term and longer-term retrospective lags of disease counts, are associated with dengue incidence and epidemics.

Our study also revealed insights into how the intra-urban outbreak risk relates to a combination of population mobility and the size of local outbreaks. The indicator integrated human mobility proxies derived from an analysis of geolocation data with disease surveillance data. It demonstrated its ability to predict dengue risk at the intra-urban village level.

This study takes advantage of available surveillance data routinely collected by official institutions and novel data sources from social media platforms. In addition, this study also uses minimal capital investment to create a prediction tool. Our software is based on open-source software and is freely available to the public at no cost. The use of our findings as the foundation for a forecasting tool has the potential to help local governments reduce the burden of arboviral diseases by enhancing the effectiveness and efficiency of

disease surveillance and control efforts. Spatiotemporal prediction can facilitate the targeting of interventions to higher risk periods and locations and thereby minimize the overall costs of operations.

Acknowledgment

I have been extremely fortunate to be supervised by Joacim Rocklöv and Yesim Tozan.

I am grateful to Joacim for his flexibility, encouragement, and enthusiasm in my work throughout my doctoral project, especially for his understanding and tolerance when I was slowly making progress. I sincerely appreciate your willingness to include me in your team's projects.

I would like to thank Yesim for always carefully showing holes in my thinking and for her concise writing style. Thank you for allowing me the opportunity to study under your guidance.

I would also like to express my gratitude to my UGM co-supervisor, Lutfan Lazuardi and Hari Kusnanto. I sincerely appreciate your confidence in me.

I would like to thank Henrik Sjödin, Junwen Guo, Prasad Liyanage, and Sulistyawati for helpful discussions during and great feedback after pre-defense.

I would also like to express my deep appreciation to Maria Nilsson, my examiner; and Paula Moraga for her kind acceptance to be my opponent.

I would like to thank all my colleagues, former and current members in Joacim Rocklöv's Lab, for the great discussions, for all their help and their friendship: Yien Ling, Vijendra Ingole, Jing Helmersson, Mikkel Quam, Wolfgang Lohr, Sewe Maquines, Zia Farooq, Chaibo Armando, and Mikael Ashorn.

I have been fortunate to meet Nawi Ng, Ailiana Santosa, Åsa Holmner, Birgitta Åström, and Camilla Andersson during my time in Umeå. They welcomed me and made me feel at home from the first day. Thank you all for your generosity, care, and help.

Thank you to the administrators in the Department of Epidemiology and Global Health who have made my traveling time to Umeå very smooth and enjoyable: Ulrika Harju and Ulrika Järvholm.

DHO Yogyakarta colleagues' invaluable knowledge and data contributions have greatly benefited this project: Anwar Cahyono, Williamto, Arif Romadhoni, and Solikhin Ramtana.

I have received excellent support from my colleagues in the Department of Health Behaviour, Environment, and Social Medicine, UGM; and Center for Environmental Studies, UGM.

I would also like to express my gratitude to Mubasysyir Hasanbasri for mentoring me since the early days and always asking provocative questions.

My doctoral journey has been challenging, and there were many times when I thought that giving up seemed to be reasonable. Nevertheless, given my starting point, I am not very surprised at how many obstacles I encountered during my 2×4-year study period. I am more surprised at how much kindness of people out there and the serendipity I have experienced along the way. To my unnamed colleagues, co-authors, friends, relatives, and wonderful strangers sporadically I met, thank you for your kindness.

Finally, I am indescribably thankful to my wife, mainly for becoming the beloved mother of my daughters. To my daughters, I am very grateful to have you.

"Segala puji bagi Allah yang dengan nikmat-Nya
semua kebaikan menjadi sempurna."

References

1. Guzman MG, Halstead SB, Artsob H, et al. Dengue: a continuing global threat. *Nat Rev Microbiol*. 2010;8(12 Suppl):S7-16. doi:10.1038/nrmicro2460
2. Simmons CP, Farrar JJ, Chau N van V, Wills B. Dengue: current concepts. *N Engl J Med*. 2012;366:1423-1432
3. Bhatt S, Gething PW, Brady OJ, et al. The global distribution and burden of dengue. *Nature*. 2013;496(7446):504-507. doi:10.1038/nature12060
4. World Health Organization. Dengue guidelines for diagnosis, treatment, prevention and control: new edition. Published online 2009:WHO/HTM/NTD/DEN/2009.1.
5. Wahyono TYM, Nealon J, Beucher S, et al. Indonesian dengue burden estimates: review of evidence by an expert panel. *Epidemiol Infect*. doi:10.1017/S0950268817001030
6. Nealon J, Taurel AF, Capeding MR, et al. Symptomatic Dengue Disease in Five Southeast Asian Countries: Epidemiological Evidence from a Dengue Vaccine Trial. *PLoS Negl Trop Dis*. 2016;10(8):e0004918. doi:10.1371/journal.pntd.0004918
7. Undurraga EA, Halasa YA, Shepard DS. Use of expansion factors to estimate the burden of dengue in Southeast Asia: a systematic analysis. *PLoS Negl Trop Dis*. 2013;7(2):e2056. doi:10.1371/journal.pntd.0002056
8. O'Reilly KM, Hendrickx E, Kharisma DD, et al. Estimating the burden of dengue and the impact of release of wMel Wolbachia-infected mosquitoes in Indonesia: a modelling study. *BMC Med*. 2019;17(1):172. doi:10.1186/s12916-019-1396-4
9. Karyanti MR, Uiterwaal CS, Kusriastuti R, et al. The changing incidence of Dengue Haemorrhagic Fever in Indonesia: a 45-year registry-based analysis. *BMC Infect Dis*. 2014;14(1):412. doi:doi.org/10.1186/1471-2334-14-412
10. Stahl HC, Butenschoen VM, Tran HT, et al. Cost of dengue outbreaks: literature review and country case studies. *BMC Public Health*. 2013;13(1):1048. doi:10.1186/1471-2458-13-1048
11. Nadjib M, Setiawan E, Putri S, et al. Economic burden of dengue in Indonesia. *PLoS Negl Trop Dis*. 2019;13(1):e0007038. doi:10.1371/journal.pntd.0007038
12. Shepard DS, Undurraga EA, Halasa YA. Economic and Disease Burden of Dengue in Southeast Asia. *PLoS Negl Trop Dis*. 2013;7(2):e2055. doi:10.1371/journal.pntd.0002055
13. Reich NG, Shrestha S, King AA, et al. Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *J R Soc Interface R Soc*. 2013;10(86):20130414. doi:10.1098/rsif.2013.0414

14. Guzman MG, Gubler DJ, Izquierdo A, Martinez E, Halstead SB. Dengue infection. *Nat Rev Dis Primer*. 2016;2(1):16055. doi:10.1038/nrdp.2016.55
15. OhAinle M, Balmaseda A, Macalalad R, et al. Dynamics of Dengue Disease Severity Determined by the Interplay Between Viral Genetics and Serotype-Specific Immunity. *Sci Transl Med*. 2011;3(114):114ra128-114ra128. doi:10.1126/scitranslmed.3003084
16. Messina JP, Brady OJ, Scott TW, et al. Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol*. 2014;22(3):138-146. doi:10.1016/j.tim.2013.12.011
17. Utama IMS, Lukman N, Sukmawati DD, et al. Dengue viral infection in Indonesia: Epidemiology, diagnostic challenges, and mutations from an observational cohort study. *PLoS Negl Trop Dis*. 2019;13(10):e0007785. doi:10.1371/journal.pntd.0007785
18. World Health Organization RO for SEA. *Comprehensive Guidelines for Prevention and Control of Dengue and Dengue Haemorrhagic Fever: Revised and Expanded Edition*.; 2011.
19. *Strategi Nasional Penanggulangan Dengue 2021-2025*. Kementerian Kesehatan RI https://ptvz.kemkes.go.id/download-media/strategi-nasional-penanggulangan-dengue-tahun-2021-2025
20. *Pedoman Pencegahan dan Pengendalian Demam Berdarah Dengue*. Kementerian Kesehatan RI; 2017.
21. Prayitno A, Taurel A frieda, Nealon J, et al. Dengue seroprevalence and force of primary infection in a representative population of urban dwelling Indonesian children. *PLoS Negl Trop Dis*. 2017;11(6):e0005621. doi:10.1371/journal.pntd.0005621
22. Indriani C, Ahmad RA, Wiratama BS, et al. Baseline Characterization of Dengue Epidemiology in Yogyakarta City, Indonesia, before a Randomized Controlled Trial of Wolbachia for Arboviral Disease Control. *Am J Trop Med Hyg*. 2018;99(5):1299-1307. doi:10.4269/ajtmh.18-0315
23. Lwande OW, Obanda V, Lindström A, et al. Globe-Trotting *Aedes aegypti* and *Aedes albopictus*: Risk Factors for Arbovirus Pandemics. *Vector-Borne Zoonotic Dis*. 2020;20(2):71-81. doi:10.1089/vbz.2019.2486
24. Näslund J, Ahlm C, Islam K, Evander M, Bucht G, Lwande OW. Emerging Mosquito-Borne Viruses Linked to *Aedes aegypti* and *Aedes albopictus*: Global Status and Preventive Strategies. *Vector-Borne Zoonotic Dis*. Published online August 23, 2021:vbz.2020.2762. doi:10.1089/vbz.2020.2762
25. Wijayanti SPM, Porphyre T, Chase-Topping M, et al. The Importance of Socio-Economic Versus Environmental Risk Factors for Reported Dengue Cases in Java, Indonesia. *PLoS Negl Trop Dis*. 2016;10(9):e0004964. doi:10.1371/journal.pntd.0004964
26. Promprou S, Jaroensutasinee M, Jaroensutasinee K. Climatic Factors Affecting Dengue Haemorrhagic Fever Incidence in Southern Thailand. *Dengue Bull*. 2005;29.

27. Arcari P, Tapper N, Pfueller S. Regional variability in relationships between climate and dengue/DHF in Indonesia. *Singap J Trop Geogr*. 2007;28(3):251-272. doi:10.1111/j.1467-9493.2007.00300.x
28. Hii YL, Rocklöv J, Ng N, Tang CS, Pang FY, Sauerborn R. Climate variability and increase in intensity and magnitude of dengue incidence in Singapore. *Glob Health Action*. 2009;2(1):1-9. doi:10.3402/gha.v2i0.2036
29. Johansson M, Cummings DT, Glass GE. Multiyear climate variability and dengue–El Niño southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis. *PLoS Med*. 2009;6(11):e1000168. doi:10.1371/journal.pmed.1000168
30. Wongkoon S, Jaroensutasinee M, Jaroensutasinee K. Weather factors influencing the occurrence of dengue fever in Nakhon Si Thammarat, Thailand. *Trop Biomed*. 2013;30(4):631-641.
31. Wongkoon S, Jaroensutasinee M, Jaroensutasinee K. Distribution, seasonal variation & dengue transmission prediction in Sisaket, Thailand. *Indian J Med Res*. 2013;138(3):347-353.
32. Jeelani S, Sabesan S. Aedes vector population dynamics and occurrence of dengue fever in relation to climate variables in Puducherry, South India. *Int J Curr Microbiol Appl Sci*. 2013;2(12):313-322.
33. Barrera R, Amador M, MacKay AJ. Population Dynamics of Aedes aegypti and Dengue as Influenced by Weather and Human Behavior in San Juan, Puerto Rico. *PLoS Negl Trop Dis*. 2011;5(12):e1378. doi:10.1371/journal.pntd.0001378
34. Hii YL, Zhu H, Ng N, Ng LC, Rocklöv J. Forecast of dengue incidence using temperature and rainfall. *PLoS Negl Trop Dis*. 2012;6(11):e1908. doi:10.1371/journal.pntd.0001908
35. Carrington LB, Seifert SN, Armijos MV, Lambrechts L, Scott TW. Reduction of Aedes aegypti vector competence for dengue virus under large temperature fluctuations. *Am J Trop Med Hyg*. 2013;88(4):689-697. doi:10.4269/ajtmh.12-0488
36. Yang HM, Macoris MLG, Galvani KC, Andrighetti MTM, Wanderley DMV. Assessing the effects of temperature on the population of Aedes aegypti, the vector of dengue. *Epidemiol Infect*. 2009;137(8):1188-1202. doi:10.1017/S0950268809002040
37. Lee SA, Jarvis CI, Edmunds WJ, Economou T, Lowe R. Spatial connectivity in mosquito-borne disease models: a systematic review of methods and assumptions. Published online 2021. doi:10.1098/rsif.2021.0096
38. Schaber KL, Perkins TA, Lloyd AL, et al. Disease-driven reduction in human mobility influences human-mosquito contacts and dengue transmission dynamics. *PLOS Comput Biol*. 2021;17(1):e1008627. doi:10.1371/journal.pcbi.1008627
39. Kiang MV, Santillana M, Chen JT, et al. Incorporating human mobility data improves forecasts of Dengue fever in Thailand. *Sci Rep*. 2021;11(1):923. doi:10.1038/s41598-020-79438-0

40. Zhang Y, Riera J, Ostrow K, et al. Modeling the relative role of human mobility, land-use and climate factors on dengue outbreak emergence in Sri Lanka. *BMC Infect Dis*. 2020;20(1):649. doi:10.1186/s12879-020-05369-w
41. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, et al. The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl Trop Dis*. 2009;3(7):e481. doi:10.1371/journal.pntd.0000481
42. Massaro E, Kondor D, Ratti C. Assessing the interplay between human mobility and mosquito borne diseases in urban environments. *Sci Rep*. 2019;9(1):16911. doi:10.1038/s41598-019-53127-z
43. Soriano-Paños D, Arias-Castro JH, Reyna-Lara A, Martínez HJ, Meloni S, Gómez-Gardeñes J. Vector-borne epidemics driven by human mobility. *Phys Rev Res*. 2020;2(1):013312. doi:10.1103/PhysRevResearch.2.013312
44. Rocklöv J, Tozan Y, Ramadona A, et al. Using Big Data to Monitor the Introduction and Spread of Chikungunya, Europe, 2017. *Emerg Infect Dis*. 2019;25(6):1041-1049. doi:10.3201/eid2506.180138
45. Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D. Understanding human mobility from Twitter. *PLoS ONE*. 2015;10(7):1-16. doi:10.1371/journal.pone.0131469
46. Lee EC, Asher JM, Goldlust S, Kraemer JD, Lawson AB, Bansal S. Mind the scales: Harnessing spatial big data for infectious disease surveillance and inference. 2016;1:10.
47. Paul MJ, Sarker A, Brownstein JS, et al. Social Media Mining for Public Health Monitoring and Surveillance. *Pac Symp Biocomput Pac Symp Biocomput*. 2016;21:468-479.
48. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. *Am J Public Health*. 2017;107(1):143-143. doi:10.2105/AJPH.2016.303512a
49. Marques-Toledo C de A, Degener CM, Vinhal L, et al. Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level. *PLoS Negl Trop Dis*. 2017;11(7):e0005729. doi:10.1371/journal.pntd.0005729
50. Gomide J, Veloso A, Meira W, et al. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In: *Proceedings of the ACM WebSci'11, June 14-17 2011, Koblenz, Germany*. ; 2011:1-8. doi:10.1145/2527031.2527049
51. Cuttone A, Lehmann S, González MC. Understanding predictability and exploration in human mobility. *EPJ Data Sci*. 2018;7(1):2. doi:10.1140/epjds/s13688-017-0129-1
52. McNeill G, Bright J, Hale SA. Estimating local commuting patterns from geolocated Twitter data. *EPJ Data Sci*. 2017;6(1):24. doi:10.1140/epjds/s13688-017-0120-x
53. Manca M, Boratto L, Morell Roman V, Martori i Gallissà O, Kaltenbrunner A. Using social media to characterize urban mobility patterns: State-of-the-

art survey and case-study. *Online Soc Netw Media*. 2017;1:56-69. doi:10.1016/j.osnem.2017.04.002

54. Kraemer MUG, Bisanzio D, Reiner RC, et al. Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about human mobility: a case study using geolocated tweets from Lahore, Pakistan. *EPJ Data Sci*. 2018;7(1):16. doi:10.1140/epjds/s13688-018-0144-x
55. Carley KM, Malik M, Kowalchuk M, Pfeffer J, Landwehr P. *Twitter Usage in Indonesia*. Institute for Software Research, School of Computer Science, Carnegie Mellon University; 2015.
56. Dirjen Pengendalian Penyakit dan Penyehatan Lingkungan. *Modul Pengendalian Demam Berdarah Dengue*. (Handoko D, Prasetyowati EB, Hartoyo S, eds.). Direktorat Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan; 2011.
57. Louis VR, Phalkey R, Horstick O, et al. Modeling tools for dengue risk mapping - a systematic review. *Int J Health Geogr*. 2014;13(1):50. doi:10.1186/1476-072X-13-50
58. Aswi A, Cramb SM, Moraga P, Mengersen K. Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review. *Epidemiol Infect*. 2019;147:e33. doi:10.1017/S0950268818002807
59. Hussain-Alkhateeb L, Rivera Ramírez T, Kroeger A, Gozzer E, Runge-Ranzinger S. Early warning systems (EWSs) for chikungunya, dengue, malaria, yellow fever, and Zika outbreaks: What is the evidence? A scoping review. *PLoS Negl Trop Dis*. 2021;15(9):e0009686. doi:10.1371/journal.pntd.0009686
60. Juffrie M, Focks DA. Early Warning System (EWS) for Dengue in Indonesia dan Thailand. *Berk Ilmu Kedokt*. 2009;41:134-142.
61. Tunggul Satoto TB, Dwiputro AH, Risdwiyanto RN, Hakim AUF, Pascawati NA, Diptyanusa A. Prediction model of dengue hemorrhagic fever transmission to enhance early warning system in Gergunung Village, Klaten District, Central Java. *J Thee Med Sci Berk Ilmu Kedokt*. 2019;51(03). doi:10.19106/JMedSci005103201909
62. Wilder-Smith A, Gubler DJ, Weaver SC, Monath TP, Heymann DL, Scott TW. Epidemic arboviral diseases: priorities for research and public health. *Lancet Infect Dis*. 2017;17(3):e101-e106. doi:10.1016/S1473-3099(16)30518-7
63. Kading RC, Brault AC, Beckham JD. Global Perspectives on Arbovirus Outbreaks: A 2020 Snapshot. *Trop Med Infect Dis*. 2020;5(3):142. doi:10.3390/tropicalmed5030142
64. Kraemer MUG, Sinka ME, Duda K a, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *eLife*. 2015;4. doi:10.7554/eLife.08347
65. Liu-Helmersson J, Brännström Ake, Sewe MO, Semenza JC, Rocklöv J. Estimating Past, Present, and Future Trends in the Global Distribution and

Abundance of the Arbovirus Vector Aedes aegypti Under Climate Change Scenarios. *Front Public Health*. 2019;7(JUN):1-10. doi:10.3389/fpubh.2019.00148

66. Polonsky JA, Baidjoe A, Kamvar ZN, et al. Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philos Trans R Soc B Biol Sci*. 2019;374(1776):20180276. doi:10.1098/rstb.2018.0276
67. Salje H, Lessler J, Endy TP, et al. Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. *Proc Natl Acad Sci*. 2012;109(24):9535-9538. doi:10.1073/pnas.1120621109
68. Gasparrini A. Distributed lag linear and non-linear models in R: the package DLNM. *J Stat Softw*. 2011;43(8):1-20.
69. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021. https://www.r-project.org/
70. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol*. 2009;71(2):319-392. doi:10.1111/j.1467-9868.2008.00700.x
71. Martins TG, Simpson D, Lindgren F, Rue H. Bayesian computing with INLA: New features. *Comput Stat Data Anal*. 2013;67:68-83. doi:10.1016/j.csda.2013.04.014
72. Blangiardo M, Cameletti M. *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. Wiley; 2015.
73. Lowe R, Lee SA, O'Reilly KM, et al. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *Lancet Planet Health*. 2021;5(4):e209-e219. doi:10.1016/S2542-5196(20)30292-8
74. Besag J, York J, Mollie a. Bayesian image-restoration, with 2 applications in spatial statistics. *Ann Inst Stat Math*. 1991;43(1):1-20. doi:10.1007/BF00116466
75. Stoddard ST, Forshey BM, Morrison AC, et al. House-to-house human movement drives dengue virus transmission. *Proc Natl Acad Sci U S A*. 2013;110(3):994-999. doi:10.1073/pnas.1213349110
76. Ramadona AL, Tozan Y, Lazuardi L, Rocklöv J. A combination of incidence data and mobility proxies from social media predicts the intra-urban spread of dengue in Yogyakarta, Indonesia. *PLoS Negl Trop Dis*. 2019;13(4):e0007298. doi:10.1371/journal.pntd.0007298
77. Ramadona AL, Lazuardi L, Hii YL, Holmner Asa, Kusnanto H, Rocklöv J. Prediction of Dengue Outbreaks Based on Disease Surveillance and Meteorological Data. *PLOS ONE*. 2016;11(3):e0152688. doi:10.1371/journal.pone.0152688
78. Hii YL, Rocklöv J, Wall S, Ng LC, Tang CS, Ng N. Optimal Lead Time for Dengue Forecast. *PLoS Negl Trop Dis*. 2012;6(10):e1848. doi:10.1371/journal.pntd.0001848

79. Lowe R, Gasparrini A, Van Meerbeeck CJ, et al. Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. Thomson M, ed. *PLOS Med*. 2018;15(7):e1002613. doi:10.1371/journal.pmed.1002613
80. Rocque RJ, Beaudoin C, Ndjaboue R, et al. Health effects of climate change: an overview of systematic reviews. *BMJ Open*. 2021;11(6):e046333. doi:10.1136/bmjopen-2020-046333
81. Cota W, Soriano-Paños D, Arenas A, Ferreira SC, Gómez-Gardeñes J. Infectious disease dynamics in metapopulations with heterogeneous transmission and recurrent mobility. *New J Phys*. 2021;23(7):073019. doi:10.1088/1367-2630/ac0c99
82. Carroll R, Prentice CR. Community vulnerability and mobility: What matters most in spatio-temporal modeling of the COVID-19 pandemic? *Soc Sci Med*. 2021;287:114395. doi:10.1016/j.socscimed.2021.114395
83. Briz-Redón Á, Iftimi A, Correcher JF, De Andrés J, Lozano M, Romero-García C. A comparison of multiple neighborhood matrix specifications for spatio-temporal model fitting: a case study on COVID-19 data. *Stoch Environ Res Risk Assess*. doi:10.1007/s00477-021-02077-y
84. Althouse BM, Scarpino SV, Meyers LA, et al. Enhancing Disease Surveillance with Novel Data Streams: Challenges and Opportunities. *N Engl J Med*. 2015;4:1-10. doi:10.1140/epjds/s13688-015-0054-0
85. Anastuti RA. Implementasi Program Pengendalian Demam Berdarah Dengue (DBD) di Kelurahan Sorosutan dan Kelurahan Semaki Kota Yogyakarta. Published online 2013.
86. Sulistyawati S, Dwi Astuti F, Rahmah Umniyati S, et al. Dengue Vector Control through Community Empowerment: Lessons Learned from a Community-Based Study in Yogyakarta, Indonesia. *Int J Environ Res Public Health*. 2019;16(6):1013. doi:10.3390/ijerph16061013
87. Sulistyawati S, Nilsson M, Ekasari MP, et al. Untapped Potential: A Qualitative Study of a Hospital-Based Dengue Surveillance System. *Am J Trop Med Hyg*. 2020;103(1):120-131. doi:10.4269/ajtmh.19-0719
88. Widiarti, Heriyanto B, Boewono DT, et al. Peta Resistensi Vektor Demam Berdarah Dengue Aedes aegypti terhadap Insektisida Kelompok Organofosfat, Karbamat dan Pyrethroid di Provinsi Jawa tengah dan Daerah Istimewa Yogyakarta. *Bul Penelit Kesehat*. 2011;39(4):176-189.
89. Tantowijoyo W, Arguni E, Johnson P, et al. Spatial and Temporal Variation in Aedes aegypti and Aedes albopictus (Diptera: Culicidae) Numbers in the Yogyakarta Area of Java, Indonesia, With Implications for Wolbachia Releases. *J Med Entomol*. 2016;53(1):188-198. doi:10.1093/jme/tjv180
90. Hasan S, Zhan X, Ukkusuri SV. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing - UrbComp '13*. ACM Press; 2013:1. doi:10.1145/2505821.2505823

Department of Epidemiology and Global Health
Umeå University, SE-901 87 Umeå, Sweden
www.umu.se/english

UMEÅ UNIVERSITY

ISBN: 978-91-7855-687-8 (print)
ISBN: 978-91-7855-688-5 (pdf)
ISSN 0346-6612