# Municipality Level Dengue Risk Prediction Modeling in Brazil and its Impacts for Future Public Health Interventions

by

**Meghan Matlack**

BS, Clemson University, 2015

MPH, Medical University of South Carolina, 2016

Submitted to the Graduate Faculty of the

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

**Meghan Matlack**

It was defended on

April 14, 2022

and approved by

Maureen Lichtveld, MD, MPH, Department of Environmental and Occupational Health

Harry Hochheiser, PhD, Department of Biomedical Informatics

Hawre Jalal, MD, MSc, PhD, School of Epidemiology and Public Health, University of Ottawa

Dissertation Director: Aaron Barchowsky, PhD, Department of Environmental and Occupational Health

# Municipality Level Dengue Risk Prediction Modeling in Brazil and its Impacts for Future Public Health Interventions

Meghan Matlack, PhD

University of Pittsburgh, 2022

Current prediction models for dengue risk are restricted to country-wide estimates or are insufficient at accounting for localized variations in outbreak risk. These models focus primarily on climate and large-scale factors that may not reflect true risk across municipalities or neighborhoods, and do not account for other determinants of health that have also been previously correlated with risk of dengue. We hypothesized that widespread municipality-level dengue outbreak forecasting would have the potential to better capture small-scale transmission dynamics and provide more accurate local outbreak predictions. We built several boosted regression tree (BRT) models to predict outbreak risk for 167 municipalities in Pernambuco, Brazil. Using training data from 2010-2015, we utilized climate, prior surveillance data, and social and environmental indicators of health as features for classification of regions as low or high risk for dengue in 2016. Our best model included real-time temperature and precipitation, lagged climate effects and prior surveillance data. This model gave a training AUC of 0.963 and a testing AUC of 0.939, with a total of 1,842 correct observations from 2,004 predictions. Additionally, this model successfully predicted 74.8% of high risk classifications, a marked improvement from previous iterations. Quantification of predictor associations through univariate and multivariate regression analyses revealed correlations fairly consistent with our BRT results. In general, we saw that inclusion of most socio-environmental predictors had minor influence over BRT predictions, and were not statistically significant in correlation with increased dengue risk when compared to other predictors. We can conclude from this dissertation research that a BRT approach is effective for modeling dengue transmission dynamics and can successfully predict high risk dengue regions using relevant climatic factors as well as prior case data, though more research is needed to establish strong socio-environmental patterns with small-scale dengue outbreak risk. Predictive models can serve as in-depth complementary tools to current dengue surveillance systems by providing useful information on upcoming outbreaks and by estimating where most cases are most likely to occur prior to peak transmission. Further development of these models can also provide the insight necessary to restructure current vector control policies and strengthen existing dengue intervention practices.

## Table of Contents

# List of Tables

# List of Figures

# Preface

I would first like to thank Dr. Aaron Barchowsky, who as my Dissertation Director has been a great source of support and encouragement. Dr. Barchowsky generously stepped in as my academic advisor to help me finish my PhD research after my previous advisor left the university last year. Despite having little knowledge about my dengue and climate work, Dr. Barchowsky has made every effort to ensure that I could graduate on time and has made the continuous effort to advocate for me as needed. I am very grateful for the time and energy that he has put forth into helping me achieve my academic goals.

I would like to specifically thank Dr. Maureen Lichtveld, Dean of the School of Public Health as well as a Professor in the Department of Environmental and Occupational Health who graciously offered to be a part of my dissertation committee. Dean Lichtveld has been instrumental in helping me to successfully complete this dissertation research and has been an incredible mentor to me during the short time she has been with the University of Pittsburgh. I greatly appreciate her in-depth insight as a fellow climate and health scientist, and I look forward to continuing my own climate and health research with her guidance as a mentor in my post-doctoral fellowship beginning this summer.

I also want to take the time to acknowledge my remaining committee members, Dr. Harry Hochheiser from the Department of Biomedical Informatics, and Dr. Hawre Jalal, formerly of the Department of Health Policy and Health Management, and now at the University of Ottawa as an Assistant Professor in the School of Epidemiology and Public Health. I am grateful to Dr. Hochheiser for the attention to detail he brings and for the the guidance he has provided throughout my dissertation work, especially with regards to his help with better understanding machine learning methods and for his coding expertise. Dr. Jalal has provided useful feedback on my dissertation research and I am thankful for his thorough reviews and suggestions to further improve my work.

Lastly, I would like to thank my family and acknowledge their support throughout my graduate studies. I am very proud to be the first in my family to obtain a doctoral degree, which would not have been possible without the immense love and emotional support I have received from all of them. To my parents, Chris and Rose Matlack, and Michelle Rodgers, thank you all for always pushing me to be the best I could be and supporting me in my goals throughout my masters and PhD programs. I will always be grateful to all of you for giving me the strength and independence to be successful in my academic endeavors and future research career. To my sister, Courtney, and my brother, Jacob, thank you for always being a source of happiness and comfort as my siblings. I love you both dearly and I hope to continue to be a positive role model for you.

# 1.0 Introduction

## 1.1 Overview of Dengue

### 1.1.1 Dengue Viruses and Structure

Dengue is a viral disease spread via the bite of an infected mosquito through four main serotypes, aptly named DENV-1, DENV-2, DENV-3, and DENV-4. These serotypes are distinct, but closely related and can all cause illness (World Health Organization 2022). Dengue viruses are a part of the genus Flavivirus, which also includes the yellow fever, West Nile, and Japanese encephalitis viruses (Scitable, 2014). Each serotype is genetically diverse, although they share about 60-75% of their genome (Guzman, 2014). Infection with one serotype is believed to confer immunity to that specific serotype, and cross-immunity to other serotypes is only temporary (World Health Organization, 2022). Subsequent dengue infections, called secondary infections, can also increase the likelihood of developing severe dengue illness (World Health Organization, 2022). Serotype co-circulation occurs frequently, though each serotype presents diverse epidemiological patterns. The first serotype, DENV-1, was initially reported in 1943 in both Japan and French Polynesia but was not consistently reported in other Asian countries until the 1950s (Messina, 2014). DENV-1 was reported in Sudan in 1984 and has since been intermittent across the rest of Africa (Messina, 2014). DENV-1 was first reported in the Americas in 1977, and has been persistently reported throughout Central and South America, peaking in 2005-2006 (Messina, 2014). DENV-2 occurrences have been consistent throughout Asia since it was first reported in Papua New Guinea and Indonesia in 1944 (Messina, 2014; Cucunawangsih, 2017). DENV-2 surfaced in Africa around 1964 in Nigeria, but has not been reported there since 1968 and has since had sporadic appearances over the rest of Africa (Messina, 2014). The first reported case of DENV-2 infection in the Americas was as early as 1953 in Trinidad and Tobago, although continuous reporting did not begin until the late 1960s (Messina, 2014). During the 1990s, a more virulent strain of DENV-2 was identified as severe dengue (DHF) cases surged throughout the Americas (Messina, 2014; Cucunawangsih, 2017). This dramatic increase in DHF cases is likely responsible for the rise in DENV-2 reporting as well (Messina, 2014). Both DENV-3 and DENV-4 have been circulating in Southeast Asia since 1953, with Thailand most frequently reporting occurrences of both strains annually (Messina, 2014; Cucunawangsih, 2017). DENV-3 in the Americas originated in Puerto Rico in 1963; the rest of the Americas did not report DENV-3 cases until the 1980s (Messina, 2014). DENV-4 did not appear in the Americas until 1981, and has been consistently reported since, and especially in Brazil, Peru, Nicaragua, Colombia, Venezuela and Puerto Rico (Messina, 2014). DENV co-circulation with two serotypes, DENV-1 and DENV-4, was first reported in 1982, in Puerto Rico (Andrade, 2016). Throughout the 1980s and into the 1990s, several countries became hyperendemic for dengue, particularly in the Americas (Loroño-Pino, 1999). By 1996, 11 countries within South and Central America reported multiple serotype co-circulation, including Brazil, Peru, Colombia, and Ecuador (Loroño-Pino, 1999). Among these recently hyperendemic regions, 2 countries (Mexico and Guatemala) reported co-circulation of all four viral DENV serotypes (Loroño-Pino, 1999).

The dengue virus is a single-strand, positive-sense RNA virus. Mature DENV virions are spherical and have a diameter of 500 Å, with an electron-dense core that is enclosed in a lipid bilayer (Kuhn 2002; Murugesan, 2020). The core of these DENV particles contains the nucleocapsid formed by the genome complexed with core (C) proteins (Murugesan, 2020). Two transmembrane viral proteins inserted into the lipid bilayer surrounding the core then forms a glycoprotein shell, also called the viral envelope (Murugesan, 2020). Around 180 copies of these membrane (prM and M) and envelope (E) proteins are embedded throughout the viral envelope (Murugesan, 2020), which forms a protective layer and controls entry of the virus into host cells. The embedded E protein lies flat against the bilayer in a continuous string of 90 homodimers lying head-to-tail to form a smooth surface, while the M protein, a fragment of its precursor prM protein after cleavage of the pr peptide during viral maturation, remains a transmembrane protein just underneath the shell formed by the E protein (Kuhn 2002; Murugesan, 2020). Along with the three structural proteins described (C, M, E), the dengue viral genome also encodes for seven non-structural (NS) proteins – NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5, all of which play a role in viral replication and assembly (Kuhn 2002, Lescar, 2018). Both NS1 and NS2 proteins are involved in the viral replication complex (Harapan, 2020). The NS3 protein has several enzymatic properties, which includes chymotrypsin-like serine protease activity and RNA helicase activity as well as RTPase/NTPase activity (Harapan, 2020). NS4A and NS4B are integral membrane proteins, and function in RNA replication by inducing membrane alterations during DENV replication (Harapan, 2020). NS5 is involved with RNA-dependent RNA polymerase activity, and functions as the DENV methyltransferase–polymerase (Harapan, 2020).

### 1.1.2 Viral Replication and Infectious Cycle

Once DENV is introduced into the human body, it can enter host cells via receptor-mediated endocytosis, where E proteins in the viral envelope bind to host cell-surface receptors (Back, 2013; Murugesan, 2020). Once bound to a target cell, DENV particles are internalized and transported to endosomes (Back, 2013; Lescar, 2018). The acidic pH of the endosome causes conformational changes in the E protein to mediate membrane fusion and release of the nucleocapsid into the cytoplasm (Back, 2013; Lescar, 2018). The viral genome is then released, and translated into a polypeptide containing C, prM, E and all seven NS proteins (Lescar, 2018, Murugesan, 2020). Following translation of the viral genome, DENV infection induces hypertrophy of cellular membranes to create vesicle packets where the viral replication complex resides (Back, 2013; Murugesan, 2020). Viral RNA synthesis occurs inside the vesicle packets, within the replication complexes (Murugesan, 2020). Signal sequences within the newly created viral polyprotein translocate NS1, E and prM proteins into the lumen of the endoplasmic reticulum, and the C, NS3 and NS5 proteins remain in the cytoplasm of the infected cell (Perera, 2008). The remaining proteins (NS2A, NS2B, NS4A, NS4B) remain as transmembrane proteins (Perera, 2008). Once processed, the polyprotein is then cleaved to create ten mature proteins via host proteases and NS2B and NS3 proteases (Lescar, 2018; Murugesan, 2020). Replication of the DENV genome starts with the creation of a negative-strand of viral RNA, complementary to the positive-input strand, which serves as a replication template (Back, 2013; Murugesan, 2020). Synthesis of this template strand is catalyzed by NS5 and NS3 along with other viral NS proteins and host factors (Murugesan, 2019). Viral RNA and proteins are then taken to the endoplasmic reticulum (ER), and are accumulated to create immature virions (Harapan, 2020). The newly formed virions are transported through the secretory pathway of the trans-Golgi network, where pr is cleaved from prM via furin-mediated proteolysis (Harapan, 2020; Murugesan, 2020). The M protein is kept in the mature viral particle as a transmembrane protein (Harapan, 2020). This process then triggers rearrangement and homodimerization of the E protein, and finally, creation of mature viral particles (Harapan, 2020; Murugesan, 2020).

DENV transmission occurs in a continuous infection cycle between humans and mosquitoes (Guzman, 2016; World Health Organization, 2022). DENV is fully adapted to human hosts, though it originated in sylvatic cycles between Aedes mosquitoes and non-human primates (Guzman, 2016). To begin the cycle, a female Aedes mosquito becomes infected after taking blood from an infected human during the febrile phase of dengue illness, where a person is highly viremic (Guzman, 2016). The virus first replicates in the mosquito's midgut cells before replicating in secondary tissues, and infects the salivary glands approximately 5-12 days later (Guzman, 2016; World Health Organization, 2022). Once the virus has infected the salivary glands, the mosquito is now infectious, and can spread DENV to another human host during feeding (Guzman, 2016). The time from ingesting the virus to infecting a new host is called the extrinsic incubation period, or EIP (World Health Organization, 2022). Time from initial infection to the onset of symptoms is between 3-14 days in humans and generally averages to around 4-7 days; this is known as the intrinsic incubation period (Gubler, 1998; Guzman, 2016). After symptom onset, the host becomes infectious and moves into the febrile phase, which can last between 2 and 10 days before recovery begins (Gubler, 1998). DENV can then be transmitted to a mosquito that bites an infected human during the febrile phase of illness and ingests the virus, thus continuing the cycle (Gubler, 1998).

### 1.1.3 Clinical Manifestations of Dengue

Dengue illness can be either symptomatic or asymptomatic. Most cases of dengue are asymptomatic, with approximately 1 in 4 infections resulting in illness (Centers for Disease Control and Prevention, 2021). Symptomatic dengue fever generally presents with a sudden onset of high fever of 40°C (104°F), along with at least two non-specific symptoms such as nausea/vomiting, joint pain, general malaise, swollen glands and rash (Gubler 1998, World Health

Organization, 2022). Symptom onset coincides with the febrile phase of dengue illness, which can last from 2 to 7 days (World Health Organization, 2022). Symptoms are usually mild and subside after this period, and the patient enters recovery phase.

On occasion, mild dengue illness can progress into severe dengue, also called dengue hemorrhagic fever (DHF). DHF manifests after a patient enters the critical phase of dengue illness about 3 to 7 days after symptoms begin (World Health Organization, 2022). The critical phase usually only lasts 24-48 hours, but can cause a sudden deterioration in patient condition (World Health Organization, 2022). Symptoms of DHF include severe abdominal pain, gastrointestinal bleeding, persistent vomiting, and bleeding from the gums and nose (Gubler, 1998; World Health Organization, 2022), though the most defining characteristic of DHF is plasma leakage in tissues due to increased capillary permeability (Guzman, 2004). In rare cases, DHF can be fatal in 5-10% of cases if a patient goes into shock, generally characterized by specific signals of circulatory failure such as weak, rapid pulse and hypotension (Guzman, 2004).

### 1.1.4 Diagnostics and Treatment

Dengue can be diagnosed during the febrile (viremic) phase of illness, and the viral load can be detected in blood, serum or plasma of the affected person (Guzman, 2016). Serology testing is the most widely used detection method for dengue primarily due to operational efficiency as well as low costs compared to other analyses (Kabir, 2021). Viral isolation methods such as reverse transcription PCR (RT-PCR) may also be performed (Guzman, 2016).

Anti-DENV IgM and IgG antibody detection using enzyme-linked immunosorbent assay (ELISA) is the most common serological method in current practice (Guzman, 2016). IgM antibodies are detectable in the blood approximately 1 week after initial infection and can remain at detectable levels for up to 3 months after recovery (World Health Organization, 2022). Anti-IgM titers are measured from sera of patients during the acute phase of dengue illness to serologically confirm infection (Guzman, 2016). Preemptive diagnosis in patients that are suspected to have dengue is done through detection of IgM antibodies in samples collected 6 days after acute symptoms begin (Guzman, 2016). IgG antibodies are typically detected during secondary infection, with a rapid IgG immune response to shared epitopes on viral proteins from both of the infective serotypes present (Muller, 2017). IgG antibodies can be detectable in as little as three days after symptom onset (Muller, 2017).

DENV can be isolated from the blood just a few days after initial infection (World Health Organization, 2022). Viral isolation can be done through RT-PCR methods, which are highly specific and can give a same- or next-day confirmation of diagnosis during acute illness (Guzman, 2016; Muller, 2017). However, PCR-based approaches require laboratory space, trained staff, and specialized equipment, which may not always be available (Muller, 2017; World Health Organization, 2022). Viral isolation of DENV can also be accomplished via NS1 antigen capture. NS1 viral protein can be detected in primary infections from symptom onset to at least 9 days after, and can be used as a surrogate marker for viremia (Muller, 2017). NS1 diagnostic tests are incredibly fast, with an approximately 20 minute turnaround time, and require no specialized equipment or techniques (Muller, 2017; World Health Organization, 2022).

There is no specific treatment for dengue illness. Patients are generally recommended to manage mild symptoms through rest and adequate hydration (World Health Organization, 2022). Painkillers, such as acetaminophen and paracetamol, may also be recommended to alleviate fever and muscle aches (World Health Organization, 2022). Symptoms of severe dengue may manifest 2-4 days after fever dissipates; it is essential to receive proper medical care during this time to avoid complications and potential death (Centers for Disease Control and Prevention, 2021; World Health Organization, 2022).

The first dengue vaccine, Dengvaxia, was licensed in December 2015, and is currently approved for use in 20 countries (Paz-Bailey, 2021; World Health Organization, 2022). Dengvaxia is a live-attenuated vaccine that is built upon a yellow fever 17D backbone and contains prM and E genes from all four of the DENV serotypes (Paz-Bailey, 2021). Two phase 3 randomized clinical trials (CYD14 and CYD15) have been conducted to evaluate vaccine efficacy (Hadinegoro, 2015; Paz-Bailey, 2021). CYD14 was conducted among children between 2-11 years of age from eleven study sites across Asia and the Western Pacific, totaling 10,275 participants (Paz-Bailey, 2021). CYD15 involved children from 22 study sites across Latin America between the ages of 9-16, with total sample size equaling 20,689 (Paz-Bailey, 2021). Both trials had approximately 2,000 participants that had their serostatus determined at baseline and had also received the vaccine (Paz-Bailey, 2021). The Dengvaxia vaccine was given in a total of 3 doses at 0, 6 and 12 months (Hadinegoro, 2015; Paz-Bailey, 2021). Efficacy of the vaccine against confirmed dengue infection and safety was assessed at 25 months after initial vaccination (13 months after final vaccination), after the active phase of the trial had ended (Hadinegoro, 2015; Paz-Bailey, 2021). Pooled serotype-specific vaccine efficacies against dengue had a range between 47.1% for DENV-2 to 83.2% for DENV-4 among participants who were 9 years or older (Hadinegoro, 2015). Nearly 80% of the participants in the CYD14 and CYD15 trials above 9 years of age tested seropositive for dengue at baseline, and tended to have higher vaccine efficacy when compared to similarly aged seronegative individuals, at 81.9% and 70.1%, respectively (Hadinegoro, 2015). Surprisingly, Dengvaxia also presents a higher risk of severe dengue among seronegative individuals that become naturally infected after vaccination (World Health Organization, 2018;

Paz-Bailey, 2021). The exact reasoning for this phenomenon is unknown, but it has been theorized that the vaccine may elicit a similar immune response to an actual primary dengue infection, and thus a first natural infection for a vaccinated seronegative individual may mimic a natural secondary infection along with greater risk for severe dengue illness (World Health Organization, 2018). As of January 22, 2020, the FDA has approved Dengvaxia for use in individuals between 9 and 16 years of age that are currently living in endemic regions and have previously had laboratory-confirmed dengue infection (Center for Biologics Evaluation and Research, 2020).

## 1.2 Climate and Dengue

Dengue is spread primarily by the Aedes aegypti mosquito, and to a lesser extent the Aedes albopictus mosquito (World Health Organization, 2022). Aedes mosquitoes thrive in tropical and subtropical environments, and climate conditions play a large role in directly influencing mosquito life cycle dynamics and indirectly increasing the risk for dengue and other mosquito-borne illnesses (Tran, 2020). Two of the most influential drivers of dengue transmission are temperature and precipitation.

### 1.2.1 Influence of Temperature

Temperature has a significant impact on dengue transmission and its effects on dengue incidence have been studied extensively. A. aegypti mosquitoes are highly sensitive to temperature, and even small changes can affect nearly every aspect of development. There is substantial evidence that increased temperatures can shorten both extrinsic incubation periods and viral replication rates, increase blood-feeding behaviors, and facilitate egg laying, larvae and pupal development (xiang, 2017). Prior studies have established the optimal maximum and minimum temperature ranges for dengue transmission to occur to be 21.6-32.9°C and 11.2-23.7°C, respectively (xiang, 2017).

Mosquito developmental cycles are significantly impacted by extreme temperature changes. Optimal temperature for immature developmental stages has been found to be approximately 30°C, though the minimum temperature required for Aedes development was between 8.3 and 14°C in laboratory settings (Tun-Lin, 2000; Xiang, 2017). At temperatures less than 20°C, fertilization declines (Christophers, 1960; Morin, 2013). Time to adulthood for mosquitoes from egg hatching has been shown to be inversely correlated with temperature, with ranges from 7-8 days at 35°C to nearly 40 days at 15°C (Tun-Lin, 2000). In laboratory settings, immature A. aegypti development rates were faster at temperatures up to 34°C (Morin, 2013; Rueda, 1990). Additionally, survival to adulthood was better at higher temperatures, peaking at 90% at 27°C (Rueda, 1990; Morin, 2013). Ideal Aedes survival rates, between 88 and 93%, have been reported to occur at temperatures between 20 and 30°C (Tun-Lin, 2000). Feeding behaviors and development activities are also limited or stopped completely at temperatures lower than 15°C, or higher than 36°C (Morin 2013; Xiang, 2017).

Dengue virus replication and transmission efficiency is also highly temperature-dependent, and typically functions best between temperatures of 18-35°C (Xiang, 2017). Watts et al. directly observed that the extrinsic incubation period (EIP) for DENV-2 in infected monkeys was as short as 7 days at temperatures between 32 and 35°C, and over 12 days at 30°C (1987). There was no viral transmission at 26°C during the time of this study (Watts, 1987). This work revealed that the EIP is also dependent on the amount of virus detected in the blood, as higher viral loads corresponded to shorter EIPs (Watts, 1987; Morin, 2013). Rohani et. al reported similar results for both DENV-2 and DENV-4 development in A. aegypti; they found that the EIP decreased from nine days to five days when the extrinsic incubation temperature increased from 26 to 30°C (2009). Diurnal temperature ranges (DTRs) are also known to have an effect on susceptibility of infection in mosquitoes. Lambrechts et al. reported there was greater susceptibility to dengue infection in mosquitoes exposed to larger DTRs (20°C) versus smaller DTRs (10°C), though EIPs did not change (2011). This study also found that DENV transmission increases as DTR increases at temperatures below 18°C, but found that the inverse was true at temperatures greater than 18°C (Lambrechts, 2011).

### 1.2.2 Influence of Precipitation

While temperature has a direct effect on mosquito life cycle dynamics, precipitation influences dengue transmission by providing highly suitable habitats for dengue vectors. Precipitation is necessary to create and maintain appropriate mosquito breeding sites, and significantly impacts vector distributions. Aedes mosquitoes typically breed in natural containers such as tree holes or rock pools, but have successfully adapted to breeding in urban environments with an abundance of man-made containers like buckets, storm drains and more (World Health Organization, 2022). Pupal surveys conducted by Barrera et al. during both dry and wet periods throughout 2007 and 2008 in San Juan city revealed that most mosquito pupae were produced in man-made containers, which included water storage vessels, plant pots, and water meters (2011). It has also been shown that high river levels and flooding due to increased precipitation during rainy seasons also contributes to higher availability of breeding sites and thus increased dengue case numbers (Hashizume, 2012). Additionally, there is substantial support for multiple relationships between rainfall and vector-borne disease dynamics, specifically with regards to time lag effects. Previous research has shown that excessive rainfall 1 month prior and drought 4-5 months prior was indicative of increased likelihood of dengue outbreaks in real time (Caldwell, 2021). Studies of delayed climate effects on dengue risk in India revealed that greater dengue transmission risk was associated with precipitation lags between 8 and 15 weeks (Kakarla, 2019). In Kaohsiung, Taiwan, 3-month lagged rainfall patterns were significantly correlated with higher dengue incidence using a Poisson regression analysis with a generalized estimating equations approach (Chen, 2010).

There is a limit between amount of precipitation and increased dengue transmission, however. Prior evidence has shown that heavy rainfall is counterproductive to dengue spread as it can flush out breeding sites and decimate mosquito populations (Morin, 2013; Xiang, 2017). Flushing, which refers to excessive water levels that surpass drainage thresholds of mosquito breeding sites, has been known to occur both in the field and in laboratory settings (Benedum, 2018). The degree of flushing is dependent on numerous factors, including intensity of rainfall, size of container, and age of mosquito larvae (Benedum, 2018). One particular study focused in Singapore revealed that dengue incidence was influenced by extensive rainfall in two specific ways (Seidahmed, 2016). They found that intense precipitation during the Northeast monsoon season initially flushes out main vector breeding sites, and that acute drying of these sites afterwards also impedes *Ae. aegypti* breeding (Seidahmed, 2016). At the other end of the spectrum, rainfall shortages may cause individuals to store greater amounts of water around households, which also can lead to an increase in dengue risk (Stewart Ibarra, 2013).

## 1.3 Dengue and Social Determinants of Health

Dengue case distributions within cities are often highly heterogeneous, and suggests that urban dengue transmission is greatly influenced by small-scale socio-environmental determinants (Zellweger, 2017). Some previous studies have suggested that higher dengue transmission is correlated with specific sociodemographic factors such as low socioeconomic status, low literacy rates, and poor household conditions (Zellweger, 2017). However, other research has established no distinct associations between dengue and socio-environmental variables (Heukelbach, 2001; Zellweger, 2017). The often contradictory results of these various experiments emphasize the importance for both better understanding setting-specific dengue transmission and for improvement in standardization of relevant socio-environmental indicators (Zellweger, 2017).

Results from a prior systematic literature review to elucidate the relationship between poverty and dengue incidence had mixed results (Mulligan, 2014). Mulligan et al. analyzed 12 English-language peer-reviewed articles concerning the relationship between dengue and poverty, and found that nine of them had demonstrated positive but inconsistent correlations, nine had null correlations, and five had negative correlations (2014). In this review, income and physical housing conditions were more consistently related to increased dengue risk over other measures of poverty (Mulligan, 2014). The authors of this systematic review acknowledge that the body of evidence for establishing a robust dengue-poverty correlation is at present incredibly weak, and that more research, especially with standardized measures for indicators and outcomes is necessary (Mulligan, 2014).

Household crowding has also been identified as a potential risk factor for increased dengue transmission, as higher density of people per room within a household is theorized to provide greater feeding opportunities to the mosquito vector (Heukelbach, 2001). One study focused specifically in Belo Horizonte, Brazil found that both high household density as well as higher proportions of elderly women and children in a household contributed to greater dengue risk (de Mattos Almeida, 2007). The association between elderly women and children to dengue has been justified as these are both vulnerable groups that tend to spend more time in and around the home – elderly women are generally retired, and children generally only spend 4 hours in school per day (de Mattos Almeida, 2007). In contrast, a study conducted by Costa and Natal in Brazil's Sao Jose do Rio Preto did not find a strong association between household density and risk of dengue (Costa, 1998).

Population density and high urban growth have also been established as potential drivers of increased dengue incidence. Within the past three decades, most urban growth has occurred in developing countries and has caused immense strain on vector control programs (Franklinos, 2019). Worldwide urbanization has positively impacted the spread of *Ae. aegypti*-transmitted diseases by influencing both climatic factors and resource availability (Franklinos, 2019). Arunachalam et. al found that higher population density within their designated study areas had higher vector abundance, and thus more opportunities for mosquitoes to feed (2010). Increased human travel also plays a role in spreading both *Ae. aegypti* vectors and related pathogens to areas that are otherwise unaffected, and large urban centers encourage continued propagation of illness (Franklinos, 2019).

### 1.3.1 Socio-Environmental Conditions

Other important variables of interest previously studied in correlation with increased dengue risk include piped water supply, access to sanitation services, and frequency of garbage collection. Costa and Natal found negative correlations between both households with direct garbage collection and households with a sanitary network connection and risk of dengue, however they did not establish an association between household access to piped water supply and dengue incidence (Costa, 1998). These associations were also noted in a spatial case-control study in Campinas, with lack of basic sanitation, low frequency of garbage collection and practice of water storage around the home having the highest odds of dengue incidence (Cordeiro, 2011). These correlations are well-established across other studies – one particular experiment conducted in a favela of Serviluz, located in the state of Fortaleza in Brazil found associations between lack of waste collection with high dengue risk, and noted that the presence of plants with temporary water pools in the garden or courtyard were connected with greater dengue incidence as well (Heukelbach, 2001). Additionally, Vasconcelos et. al found conflicting associations between increased socioeconomic status and increased dengue incidence (1998). What was interesting about these associations was that they were theorized to be in part due to the fact that people of higher socioeconomic classes tended to cultivate larger numbers of aquatic plants, and used more disposables compared to people of lower socioeconomic status and thus provided higher numbers of suitable breeding habitats for mosquito vectors (Vasconcelos, 1998).

Increased water sources in and around infested areas have been shown to consistently contribute to mosquito pupal populations, and can thus lead to increased dengue risk as well (Arunachalam, 2010). Unused and unprotected outdoor water storage containers were the largest contributors to mosquito pupal production across six different study sites located in India, Myanmar, Sri Lanka, Indonesia, Thailand and the Philippines (Arunachalam, 2010). *Ae. aegypti* mosquitoes also prefer natural water sources, and tend to gravitate towards rainwater-filled containers over tap water-filled sites (Arunachalam, 2010). The link between domestic water supply and mosquito pupal reproduction is multifaceted, however, as both lack of access to piped water and irregular piped water supply can lead to increased water storage (Arunachalam, 2010). In Myanmar, study sites with no piped water supply and study sites with irregular water supply had higher pupae per hectare indices (PHI), though there was no significant difference between them (Arunachalam, 2010).

### 1.3.2 Associations with Land Cover and Land Use

There has been recent evidence that land cover and deforestation practices may also contribute to increased risk of dengue in some regions. *Ae. aegypti* vector ecology is dependent on both abiotic and biotic factors, and changes to land cover or land use can affect mosquito populations through alteration of microclimates, increased competition for resources, and nutrient availability (Franklinos, 2019). Land cover variables including residential area per capita, construction areas, grasslands and watery areas have been significantly correlated with higher dengue risk (Akter, 2017). However, these results are not consistent – Akter et. al found overall increasing, but non-significant trends between intensive uses of, and production from dryland agriculture and plantations and dengue incidence but no other associations for land cover/land use variables in their spatial analysis of dengue incidence trends in Australia (Akter, 2017).

Moreover, deforestation can stimulate mosquito growth by increasing availability of sunlight in affected areas, which provides essential nutrients for mosquito larvae and by limiting fungal growth (Franklinos, 2019). An interesting finding made by Kalbus and colleagues in their study of dengue incidence using generalized linear models was that several municipalities within Amazonas state in Brazil experienced higher incidence of dengue two, three and five years after initial deforestation, but found no significant association between higher deforestation levels and increased dengue incidence when controlling for other variables (2021).

## 1.4 History of Dengue Transmission in the Americas

Dengue is the most rapidly spreading mosquito-borne illness in the world, and boasts a 30-fold increase in global incidence over the past several decades (World Health Organization, 2012). The number of dengue cases reported to WHO has dramatically increased from approximately 500 thousand cases in 2000 to over 5.9 million cases in 2019 (World Health Organization, 2022). In the Americas specifically, approximately 500 million people are currently at risk for dengue infection (Pan American Health Organization, n.d.). Over the past 40 years, dengue incidence in the Americas has skyrocketed from around 1.5 million cumulative cases in the 1980s to 16.2 million between 2010 and 2019 (Pan American Health Organization, n.d.). Dengue in the Americas follows an endemo-epidemic pattern, where outbreaks occur every three to five years (Salles, 2018). The first suspected dengue-like epidemics can be traced as far back as 1635 in both Martinique and Guadeloupe (Braithwaite Dick, 2013). Etiology of these initial reported outbreaks is unclear, though the first true documented case of dengue fever in the Americas occurred in Philadelphia in 1780 (Braithwaite Dick, 2013). Dengue epidemics in Caribbean countries and the southern United States can be separated into four distinct periods: introduction of dengue to the Americas (1600-1946), plan for eradication of *Ae. aegypti* (1947-1970), reinfestation with *Ae. aegypti* (1971-2000), and finally, increased dispersion of *Ae. aegypti* and dengue viral circulation between 2001 and 2010 (Guzman, 2003). Most epidemics during this first period were centered in and around port cities, and dengue outbreaks throughout the 19th century were mainly related to commercial activities (Guzman, 2003; Braithwaite Dick, 2013). A dengue-like disease outbreak in Peru in 1818 was responsible approximately 50,000 cases alone (Braithwaite Dick, 2013). Distribution of dengue-like illness throughout the end of the 19th century and into the 20th began to extend farther north into the United States and also as far south as Chile and Argentina in South

America (Braithwaite Dick, 2013). The first isolation of dengue viruses occurred between 1943 and 1944, and laboratory tests became available for diagnosis shortly after (Braithwaite Dick, 2013).

The first initiative to reduce *Ae. aegypti* populations was in 1901 in Havana, Cuba (Braithwaite Dick, 2013). Several similar programs followed in both Sao Paulo, Brazil in 1901 and Rio de Janeiro, Brazil in 1903 (Braithwaite Dick, 2013). These local programs were based on mosquito elimination through fumigation and destruction of abandoned containers (Braithwaite Dick, 2013). Beginning in the 1940s, the Pan American Health Organization (PAHO) emphasized the importance of *Ae. aegypti* prevention and control measures, first in Brazil, and later across the continent (Braithwaite Dick, 2013). These measures were strongly driven by the advent of DDT around this time (Braithwaite Dick, 2013). By 1947, PAHO approved a continental mosquito eradication program to combat urban yellow fever, which was largely successful due to intense promotion campaigns in several of the affected countries (Braithwaite Dick, 2013). Eradication of the *Ae. aegypti* vector was successful for 18 countries and several Caribbean islands by 1962 (Braithwaite Dick, 2013). *Ae. aegypti* was not eradicated however in the United States, Cuba, Venezuela and several other Caribbean countries (Braithwaite Dick, 2013). Between 1962 and 1972, eradication efforts began to quickly deteriorate, mainly due to loss of political importance, declines in surveillance and slow re-infestation response (Braithwaite Dick, 2013). The rapid decline in eradication efforts subsequently led to the re-infestation and global expansion of the Aedes mosquito, as well as the increase of serotype circulation in following years (Braithwaite Dick, 2013).

Both DENV-2 and DENV-3 circulated together in the Americas until 1977 (Braithwaite Dick, 2013). The DENV-1 serotype was first reported in 1977 and caused epidemics in Jamaica, Cuba, Puerto Rico and several Caribbean islands before expanding to northern areas of South America, Central America, and Mexico by 1978 (Braithwaite Dick, 2013). Between 1977 and 1980, the Americas reported over 700,000 total cases of dengue with DENV-1 as the predominant serotype in circulation (Braithwaite Dick, 2013). Throughout the 1980s, the number of reported dengue cases skyrocketed and many epidemics during this period were associated with DENV-2 circulation (Braithwaite Dick, 2013). In 1981, Cuba experienced the first major DHF epidemic in the region caused by DENV-2, with over 300,000 total cases and over 10,000 DHF cases in particular (Braithwaite Dick, 2013). 1981 was also the year that DENV-4 was introduced to the region, beginning in the eastern Caribbean and spreading across the rest of the Caribbean as well as Central and South America in areas where DENV-1 outbreaks were prevalent (Braithwaite Dick, 2013). Overall, DHF incidence in the Americas remained relatively low until 1989, when a second DHF epidemic occurred in Venezuela (Braithwaite Dick, 2013). This epidemic was traced to three (DENV-1, DENV-2, DENV-4) serotypes, and DENV-2 specifically was associated with fatal cases (Braithwaite Dick, 2013). Reintroduction of the DENV-3 virus to the Americas began in 1994, with outbreaks reported in Nicaragua and Panama (Braithwaite Dick, 2013). By 1998, the DENV-3 serotype had successfully infiltrated many Central American countries and Mexico, later spreading across South America (Braithwaite Dick, 2013). To combat increasing dengue cases and serotype co-circulation, PAHO developed and approved a continental plan to strengthen *Ae. aegypti* control measures (Braithwaite Dick, 2013). This plan was never fully implemented however and dengue epidemics across the Americas continued to intensify. In 1998, circulation of dengue within the region was widespread, culminating in extensive outbreaks across 16 countries, including Argentina, where dengue had not been reported for over 80 years (Braithwaite Dick, 2013).

The period between 2000 and 2010 saw the greatest number of dengue cases reported in a decade, in parallel with the co-circulation of all four viral serotypes in the Americas (Braithwaite Dick, 2013). Two major Pan American epidemics occurred in this time period, in 2002 and 2010 (Braithwaite Dick, 2013). In 2002, over one million cases were reported across the Americas, with over 75% of cases coming from Brazil (Braithwaite Dick, 2013). Nearly 15,000 cases were DHF cases, and the predominant serotype in circulation during this epidemic was DENV-3, followed by DENV-2 (Braithwaite Dick, 2013). The epidemic in 2010 affected over 1.7 million people and had over 50,000 cases classed as clinically severe (Braithwaite Dick, 2013). The dengue case fatality rate for this epidemic was calculated as 2.6% (Braithwaite Dick, 2013). To address the increasing number of cases in the region, PAHO implemented the Integrated Management Strategy for Dengue Prevention and Control (IMS-dengue) in 2003 (Braithwaite Dick, 2013). The IMS-dengue strategy is based on six different constituents: epidemiology, healthcare, laboratory, environment, entomology and social environment (Braithwaite Dick, 2013). By 2010, nineteen countries in the Americas had applied this system, and four additional sub-regional plans were developed for Central America, the Caribbean, the Andean subregion and MERCOSUR (Braithwaite Dick, 2013). Between 2011 and 2017, four major outbreaks occurred alongside four global sporting events: the 2011 Pan-American Games in Mexico and the 2013 Confederations Cup, the 2014 World Cup and the 2016 Olympics, all of which were held in Brazil (Salles, 2018). The period between 2011 and 2017 saw a 30% increase in dengue cases (over 10 million) compared to the previous decade, which reported a total number of 7.6 million cases (Salles, 2018). Compared to decades prior, the 2011-2017 timeframe was responsible for approximately 47% of total dengue cases reported (Salles, 2018).

## 1.5 History and Significance of Dengue in Brazil

Brazil is a country located in South America, and is the largest country in both South and Latin America. It is currently the fifth-largest country by area, and the sixth-largest by population with over 3.3 million square miles and 211 million people, respectively. Brazil is one of seventeen megadiverse countries in the world, and comprises a variety of tropical and subtropical environments, which include wetlands, plateaus, savannas and more. Brazil is also home to the Amazon River basin, the world's largest river system and including the world's most extensive rainforest. Temperatures average around 25°C with little seasonal variation in areas near the equator. Regions closer to the Tropic of Capricorn tend to be milder and get average temperatures between 13°C and 22°C. Most of Brazil sees approximately 1000 to 1500 mm of rainfall (39 to 59 in) per year, typically occurring during the summer months between December and April. The Amazon region, known for its incredibly humid tropical climate, usually gets over 2000 mm (79 in) of annual precipitation but can see over 3000 mm (118 in) in a year.

Brazil has experienced consistent urban growth for the last several decades and by 2020, 87% of the population lived within cities. Between 1940 and 1970, Brazil's major metropolitan areas increased at a rate of 4.5% per year (The Borgen Project, 2020), going from about 15% of the population living in urban areas to about 50% within this timeframe (Wagner, 1980). This rapid urbanization was led mainly by rural migrant workers in search of better opportunities (The Borgen Project, 2020; Wagner, 1980). Massive migration to urban city dwellings caused a noticeable disparity in living conditions, where rural workers settled around the city fringes (The Borgen Project 2020). The area around Brazilian cities, called "urban frontiers", were characterized by poor living conditions and led to a rise in slum growth and overall insufficient infrastructure (The Borgen Project, 2020), which has persisted today.

The largely hot and humid climate of Brazil, along with the high degree of population growth and poor urban organization, provides an incredibly suitable environment for several mosquito-borne illnesses, and particularly dengue fever. Global incidence of dengue has drastically increased in the past several decades, with over half of the world's population currently at risk (World Health Organization, 2022). Worldwide, an estimated 100-400 million dengue infections occur annually, though most (80%) of these infections are mild or asymptomatic (World Health Organization, 2022). Prior to 1970, there were only nine countries that had experienced major dengue epidemics. Today, over 100 countries across the world are endemic for dengue, with the Americas, Southeast Asia and the Western Pacific regions being the most affected.

The first recorded outbreak of dengue in South America occurred in Peru (1818), although the first known outbreak of dengue in Brazil did not occur until 1845, in the city of Rio de Janeiro (Schneider and Droll, 2001). Other epidemics in Brazil were reported between 1851-1853, as well as 1916-1923 (Schneider and Droll, 2001). By the 1940s, the Pan American Health Organization (PAHO) began to lead efforts in eradication of the Aedes aegypti mosquito, the primary dengue vector, from the Americas via an urban yellow fever eradication program (Braithwaite Dick, 2012; Fares, 2015). Beginning in 1947, the Pan American Sanitary Bureau (PASB) extensively promoted these efforts in infested countries, and successfully kept A. aegypti out of Brazil until 1976 (Braithwaite Dick 2012; Fares 2015; Schneider and Droll, 2001). Throughout this period however, eradication programs in many countries were quickly deteriorating, mainly due to insufficient surveillance, slow government response to re-infestations, and development of mosquito resistance to DDT (Braithwaite Dick, 2012). Dengue re-emergence in Brazil commenced in 1981 with an epidemic of both DENV-1 and DENV-4 circulation in Roraima (Braithwaite Dick, 2012). DENV-1 quickly became prominent throughout Brazil after introduction to Rio de Janeiro and subsequently the rest of the country in 1986 (Braithwaite Dick, 2012). DENV-2 was introduced to Brazil around 1990, which coincided with the first documented cases of severe dengue illness (Fares, 2015). Throughout the 1990s, both DENV-1 and DENV-2 spread intensified, causing major outbreaks across Brazil (Fares, 2015). DENV-3 first surfaced in Brazil around 2000, and became the predominant serotype from 2002 to 2006 (Fares, 2015). Additionally Brazil accounted for over 60% of the world’s reported dengue cases between 2000 and 2007 (Fares, 2015).

Historically, Brazil has not been truly successful in controlling the spread of A. aegypti vectors through traditional methods. Vector control is managed by all three government levels (federal, state, municipal), but the shared responsibility between these entities typically slows down epidemic response and efficiency of long-term vector control (Maciel-de-Freitas, 2014). Current vector control methods in Brazil rely heavily on surveillance of A. aegypti populations as well as low-volume space spraying during outbreaks, however little emphasis is placed on community awareness and education between epidemics (Maciel-de-Freitas, 2014) and dengue case numbers remain consistently high across the country. Thus, it is crucial to not only identify significant drivers of dengue and their contributions to future outbreak risk, but to also shift to a more localized approach that highlights the importance of individual and community contributions to sustained dengue prevention.

## 2.0 Significance for Public Health Research

Climate change has increasingly threatened to intensify and increase dengue epidemics across the world, as *Ae. aegypti* vectors rely heavily on high temperature and precipitation patterns for survival. The ideal temperature range for dengue transmission to occur is between 20 and 35°C, or 68 – 95°F (Islam, 2021). Temperatures between 20 and 27°C (68-82° F) are optimal for mosquito breeding and mosquitoes have been found to feed faster and more often in warmer temperatures, and higher temperatures also decrease larvae development rates and hatching times (Rueda, 1990; Johansson, 2009).

Additionally, Earth's temperature has risen by 0.08° C (0.14° F) per decade since 1880, and the rate of warming over the past 40 years is more than twice that: 0.18° C (0.32° F) per decade since 1981 (NOAA, 2021). 2020 was the second warmest year on record, and the ten warmest years have all occurred since 2005 (NOAA, 2021). Researchers have noted that over 60% of the world's population will be at risk for dengue by 2080 (Messina, 2019). Studies have also predicted that the range for *Ae. aegypti* vectors will shift northward to include parts of Europe and into the United States if climate change is left unmitigated (Kulkarni, 2022).

Dengue has quickly become a top global public health threat, and it is obvious it will continue to be if things are left unchanged. As such, it has become increasingly important to anticipate and plan for local and large-scale dengue outbreaks and epidemics, especially in the advent of climate change. The ability to accurately predict outbreaks prior to dengue season is crucial for rapid mobilization of resources as well as for directing future public interventions in terms of stronger vector control methods.

# 3.0 Hypothesis and Aims

To our knowledge, there is limited data on the utility of specific municipality-level predictions for upcoming dengue outbreaks. Current prediction models for dengue risk are generally restricted to country-wide estimates, and are ineffective at accounting for localized variations in outbreak risk. These models generally tend to focus primarily on climate and other large-scale factors that do not necessarily reflect true risk across municipalities or neighborhoods, and do not account for more specific determinants of health that have also been correlated with risk of dengue spread such as income levels, education, household density, and access to basic sanitation services.

Our work here seeks to determine the value in making municipality-level dengue risk predictions exploring both established climatic drivers of dengue as well as lesser studied socio-environmental health factors, and their implications for future public health interventions. Municipality level predictions have the potential to help restructure and improve current vector control strategies, as they can provide more specific details on dengue transmission dynamics across a smaller regional scale, and better assist in delivering targeted, real-time resource deployment to high risk areas during outbreaks.

Our major end goal with developing these models is to leverage these findings to expand into other dengue-endemic regions and improve upon current public health interventions for effective dengue control. In our analyses, we used reported monthly dengue case data per municipality as our outcome of interest and focused on climatic, sociodemographic and environmental indicators as relevant predictors using boosted regression tree (BRT) models. We used our predictions to then create monthly maps of potential outbreak risk across Pernambuco state, identifying key areas at higher risk for dengue outbreaks during 2016.

## 3.1 Central Hypothesis

Our central hypothesis states that utilizing municipality level data for dengue forecasting is advantageous for enhanced discrimination of outbreak-prone regions across Brazil and has the potential to improve current dengue control tactics. Development of a more focused model of dengue outbreak classification is important for prioritizing high-risk areas in terms of better vector control management, and also for optimizing deployment of essential resources prior to predicted outbreak occurrences. As part of this dissertation research, we developed three specific aims to investigate this hypothesis, as described below.

## 3.2 Aim 1: Building an Effective Boosted Regression Tree Model Based on Climatic Drivers of Dengue

Our first aim centered around building an effective predictive model for municipality-level dengue outbreak risk focusing on relevant climatic factors. We dedicated our efforts to creating a successful BRT model to predict outbreak risk for 129 municipalities in Pernambuco state using real-time temperature and precipitation values, lagged temperature and precipitation effects, and prior surveillance data as predictors. Outbreak risk was classified as either high or low based on a fixed incidence threshold of 100 cases per 100,000 people. With this aim, we hypothesized that forecasting dengue outbreaks at the municipality level would be superior to current national level surveillance systems and would better estimate and map monthly dengue risk using the previously mentioned predictors.

### 3.3 Aim 2: Incorporation of Additional Social and Environmental Predictors

We further explored our central hypothesis through our second aim, in which we incorporated social and environmental predictors to our model that had been previously correlated with increased dengue risk but were not as well-established in the literature. For our updated studies, we chose to include eight new predictors: population density per square kilometer, predominant land cover type, household access to piped water, household access to basic sanitation services, degree of urbanization, both primary and secondary suppression of vegetation (for deforestation), and recovery of secondary vegetation. We hypothesized that the inclusion of relevant social and environmental predictors would improve the accuracy of our predictions by accounting for more localized variance among dengue incidence across municipalities.

### 3.4 Aim 3: Implications for Public Health Research and Future Policy

Our final aim focused on examining our findings in terms of the implications for science, policy, and practice of future vector control. We hoped that our results would provide a greater push for more localized dengue control practices and more consistent vector control methods overall. Additionally, we planned to utilize our results to make further dengue forecasts for Brazil and provide a more localized representation of dengue risk. We expect that our results here will offer further insight into local transmission dynamics and initiate the push towards continued research for modeling outbreaks at the municipality level and for improved government vector control policies at all stages.

# 4.0 Overall Materials and Methods

## 4.1 Study Area of Interest

We focused specifically on Pernambuco, Brazil as our study area of interest for the purposes of this dissertation research. Pernambuco was chosen due to the fact that it contains 185 municipalities, and has three distinct climate zones that make it most suitable for analysis. The coastal region of Pernambuco is typically hot, humid and sees high precipitation (Encyclopedia Britannica, 2016). Pernambuco's capital, Recife is also located within this region and is a large urban hub (Encyclopedia Britannica, 2016). The agreste region is slightly drier and characterized by lighter vegetation, and separates the coast from the inner plateau region (Encyclopedia Britannica, 2016). Lastly, the sertão region of Pernambuco encompasses the large plateau which is generally rocky, cold and very dry (Encyclopedia Britannica, 2016). We expected that with these different climatic regions, we would see more pronounced differences among municipalities. Our timeframe of interest was constrained by the availability of our monthly case data, and spanned from 2010 to 2016 (see Section 4.1.3 'Imputations and Missing Data').

### 4.1.1 Data Collection

All of the data used for this research came from publicly available data sources. Case data was retrieved from the Brazilian Ministry of Health's DATASUS SINAN system, which is the most comprehensive surveillance system in Brazil for notifiable diseases, including dengue. We collected annual CSV files available for years 2010 up to 2017 for monthly dengue cases (classified by month of notification and municipality of residence). Temperature and precipitation values in degrees Celsius were contained in TIF files, which were downloaded from WorldClim's monthly historical climate data archive and uploaded as rasters using R version 3.6.3 (R Core Team, 2020). All raster climate data were correlated to a color gradient, ranging from light pink for lower values to yellow and bright green for higher values, at a spatial resolution of $21\text{km}^2$. The environmental data collected included information on primary and secondary suppression of vegetation, recovery of secondary vegetation, and land cover and use across Brazil, including degree of urbanization. All environmental data collected came from MapBiomas Collection 5, a collaborative SEEG initiative with land cover data available from 1985-2019. Environmental predictor data were contained in TIF files downloaded directly from Google Earth Engine and reformatted in R. Population density data were obtained from WorldPop, an open access resource containing high-resolution geospatial population data. Annual population density values at $1\text{km}^2$ resolution were obtained from available TIF files and reformatted with R. Population density values were in units of number of people per square kilometer. Finally, access to piped water and piped sanitation data were available as CSV files via the University of Washington's IHME website, from their LMIC Drinking Water and Sanitation Access Geospatial Estimates 2000-2017 dataset.

### 4.1.2 Data Cleaning and Pre-Processing

All datasets were cleaned and processed prior to further analysis. Our raster data were all uploaded, cropped and further processed to match specifically with associated Pernambuco municipalities with raster dimensions of 136 rows and 216 columns for a total of 29,376 pixels within each raster image. We began by downloading the most current shapefile of Brazil subset by administrative district level 2 from the Brazilian Institute of Geography and Statistics (IBGE), to reflect all municipal level boundaries. This shapefile was further subset to Pernambuco, containing a total of 185 municipalities. The resulting Pernambuco shapefile was then used to order and match all collected data to specific municipalities. Dengue case data was downloaded and imported into R as CSV files containing all monthly dengue cases reported by municipality. Case data was reformatted into a single column time-series format by first matching all values to the correct municipalities contained in our Pernambuco shapefile, and then re-ordering all values by chronological month and year. Population density, climate and environmental data were contained within TIF files which were converted to rasters in R to prepare for processing. Monthly minimum and maximum temperature values along with monthly average precipitation values were extracted from the corresponding rasters created for each month and year of analysis (January 2010-December 2016) with 'raster' package version 3.0-12. For each municipality of interest, temperature and precipitation data were extracted per pixel contained within the appropriate administrative boundaries, summed and averaged to get a final monthly value. Population density data were extracted similarly for each municipality between years 2010 to 2016. Population density estimates were then summed and averaged to get final values, however since this was an annual estimate, the same estimate was re-used per month across the same year.

All environmental variables included for this dissertation research were also contained in TIF files, collected directly from Google Earth Engine and uploaded into R as raster images. Due to the extensively large size of the environmental files collected, all environmental data used for analyses were extracted and processed via high throughput computing resources provided by the University of Pittsburgh's Center for Research Computing. Numeric classification codes corresponding to primary suppression were extracted from rasters and the number of pixels per municipality that corresponded to this classification were summed and then divided by the total number of pixels contained within the associated municipality to get a final percentage value. This process was repeated for our other environmental predictors of secondary suppression and recovery of vegetation. Land cover was classified as a categorical variable, with 3 classification levels based on the predominant land cover type per municipality: farming, forest, or urban. Land cover classifications were extracted similarly to the other environmental predictors, but our final values per municipality were calculated by adding the number of pixels within a municipality that was assigned each specific land cover classification, and the classification that had the greatest number was the classification used. Degree of urbanization was calculated from these existing land cover classifications by summing the number of pixels per municipality that already had an 'urban' classification and dividing this number by the total number of pixels per municipality and multiplying by 100 to get a final percentage. Annual household access to piped sanitation services and piped water were contained in CSV files as both percentages and numeric values. For the purposes of our specific analyses, we chose to work with percentages of households that had access for each municipality of interest. Piped sanitation and piped water access files were uploaded into R and specific values were matched to the correct municipality for each month and year included in our study period. Several previous studies have also noted that temperature lags of up to 4-5 weeks and precipitation lags between 1-3 months are most strongly correlated with higher dengue risk, and thus were also included. The entirety of predictor variables along with our response variable are listed in Appendix Table 1.

In previous literature, climate lagged effects of 1-, 2-, and 3-months were most influential for dengue risk predictions (Lowe, 2018), and so our major climate predictors were lagged at one, two and three months to see if any climate lagged effects would also be useful as covariates.

Dengue case data was lagged at both 1 month and 12 months based on previous studies showing that both short-term and long-term prior surveillance data are correlated to increased dengue transmission and models including both prior surveillance data and meteorological effects are most predictive of dengue outbreaks (Ramadona, 2016; Jain, 2019). After running several trial iterations, 2- and 3- month lags for minimum temperatures and precipitation as well as 1-month lags for dengue case data was kept as they had the best results. In total, sixteen predictors related to climate, environment and social determinants of health were included. Our response variable was cases per 100,000 people, which was calculated by dividing case numbers by municipality population (assuming everyone is at equal risk) and multiplying by 100,000 to get a final value.

### 4.1.3 Imputations and Missing Data

Due to the amount of missingness in our dengue case dataset, we decided to base our timeframe of interest on specific exclusion criteria to minimize the total number of missing observations included in our analysis. The specific timeline available to us based on the accessibility of dengue case data was initially 2001 to 2017. Municipalities were completely excluded if they had two or more years of missing data. Particular years of interest were excluded if they contained greater than 20% missing observations, and there had to be at least 4 consecutive years with less than 20% missing observations each for our timeframe to be considered. Based on the exclusion criteria set, the final timeframe selected was the period from January 2010 to December 2016, starting with a total of 128 municipalities that had complete non-missing data as defined by our exclusion criteria (69%). Data from 2000-2009 were not used for these analyses at all as they did not meet these exclusion criteria.

To increase the number of municipalities that could be effectively predicted by the model, data imputations were performed for municipalities that were only missing one year of data that did not occur at either the beginning or the end of our selected timeframe. Thirty-eight municipalities in the initial dataset were missing observations across an entire year, but not across specific months within that year. To mitigate this issue, we used a simple moving average imputation with the imputeTS 3.2 package in R to average the prior year and subsequent year's monthly observations to provide an estimated value for the missing observations. Once imputations were included, our final dataset consisted of 167 municipalities out of the 185 (90%) initially considered. In total, 167 municipalities with monthly observations from January 2010 to December 2016 gave 14,028 values for analysis. Appendix Figure 1 shows all dengue case data for January 2010 to December 2016 for all municipalities.

### 4.1.4 Building the Model

A boosted regression tree approach was utilized to make monthly predictions of dengue risk. We created a BRT model using a Bernoulli loss function to predict either high dengue outbreak risk, or low dengue outbreak risk. Our complete dataset was split into a training set containing monthly values from 2010 to 2015 for a total of 12,024 observations, and a test set containing monthly values from 2016 for a total of 2,004 observations. The 'gbm' version 2.1.5 and 'dismo' version 1.1-4 packages in R were used to fit and evaluate our model. We initially set parameters for our initial models using only climate and surveillance data at learning rate at 0.015, tree complexity at 3 and bag fraction at 0.75. These parameters were chosen because they collectively performed the best with our training data during trial iterations. All other parameters were kept at default and not specified in our call. For analyses that included the additional socio-environmental variables, learning rate was increased to 0.02 but all other parameters were kept the same.

The gbm.step function was initially used to determine the optimal number of trees by determining where the holdout deviance was minimized. Using 10-fold cross validation, the training data was randomly separated into 10 subsamples of equal length, with 9 unique samples of data and the last sample being used as a validation set. This process created 10 initial BRT models consisting of 50 trees each. Mean performance and standard error was recorded before the function stepped forward using the previously defined learning rate and increased the number of trees by a constant amount. The process was repeated for several iterations until minimum error was reached, and the optimal number of trees was determined. Optimal number of trees to include in a model can vary but for our purposes we focused on at least 1000 trees as a basic rule of thumb (Elith, 2008), though we limited total number of trees to under 2000 so as to not overfit our models to the training data. This combination of optimal trees along with our previously defined parameters allowed us to then fit several model iterations to our data and predict monthly transmission probabilities for the year 2016. Predictions of high and low dengue risk were further determined by setting a transmission threshold of 0.5 to separate probabilities by risk classification. Any probability less than the threshold was classified as a low risk prediction, while any probability greater than the threshold was classified as a high risk prediction.

### 4.1.5 Regression Analysis

To better quantify the associations seen with our BRT models, we performed univariate and multivariate logistic regressions on all variables included across all model iterations. Sixteen univariate regressions were completed for each variable vs. our response of low or high risk for dengue. Two multivariate regressions were completed for this research. The first multivariate regression included only climate, lagged climate effects and prior surveillance data, while the second included all climatic and socio-environmental predictors used for this dissertation work. Due to the fact that all variables had different scales and some had non-normal distributions, all predictors were normalized via log transformation prior to logistic regression. We operated on a significance level of $p < 0.05$ when determining which variables were statistically significant for each regression. For the multivariate regressions that were performed, all variables were initially input into the logistic regression model, and variables that were not statistically significant were removed in a sequential order, beginning with the highest non-significant p-value. Variables were removed until only statistically significant variables remained.

# 5.0 Assessing the Suitability of Climate and Associated Lagged Effects for Municipality-Level Dengue Outbreak Risk Modeling in Brazil

Manuscript to be submitted to the International Journal of Environmental Research and Public Health.

## 5.1 Abstract

Climate conditions have a huge impact on vector borne disease transmission, and temperature and precipitation particularly are some of the main drivers of global dengue epidemics. Current dengue outbreak prediction models do not account for intra-regional changes and thus are largely insufficient in providing fast and localized public health interventions. Here we built a simple boosted regression tree model using three main climate variables along with several lagged predictors of climate and case data to forecast monthly outbreak potential for 167 municipalities in Pernambuco, Brazil. Our results indicated that our final model was relatively accurate in discriminating between both areas of low and high outbreak risk, but only with incorporation of lag effect predictors. Although improvements can be made, predictive models such as the one created here have strong potential to restructure and redevelop current dengue prevention policies and provide better targeted public health solutions.

## 5.2 Introduction

Dengue is one of the most significant mosquito-borne illnesses in the world, and is an important public health challenge as we continue to see dramatic increases in dengue cases globally (World Health Organization, 2012). Recent evidence reports approximately 390 million cases annually, both in endemic regions as well as previously unaffected areas (Bhatt, 2013). Geographic dispersion of the virus and its mosquito vector has expanded extensively over the past several decades (Hernández-Ávila, 2013). Currently over 100 countries are estimated to be endemic for dengue, and are mostly concentrated in tropical and subtropical regions across the Americas, Asia and Africa (World Health Organization, 2012; Gubler, 1995). Dengue's primary mosquito vector, Aedes aegypti (*Ae. aegypti*), thrives in hot, humid environments and has proven to be highly adaptable to urban societies (Bhatt, 2013).

Mosquito-borne disease distributions tend to be highly heterogeneous in that they depend on the spatial distributions of both vector and host (Vanwambeke, 2010). These distributions are typically influenced by a variety of factors, most notably environmental conditions (Vanwambeke, 2010). The *Ae. aegypti* mosquito is highly sensitive to changes in both temperature and precipitation (Arboleda, 2009; Favier, 2005). Increases in temperatures can increase mosquito density through increased egg-hatching, decreased larval development rate, decreased reproductive cycles and shortened viral incubation periods (Johansson, 2009; Iguchi, 2018). Additionally, increased precipitation leads to a larger number of available mosquito breeding sites (Johansson, 2009). Climate change on a global scale has already expanded the current range of Aedes mosquitoes, thus exposing a greater number of human hosts to the virus and extending the dengue transmission season (Favier, 2005; Li, 2018).

Effective vector control strategies, such as mosquito surveillance and pesticide use in particular, rely greatly on access to reliable epidemiological data sources to facilitate central public health action (Hernández-Ávila, 2013). However, availability and accuracy of sub-national data is limited (Stevens, 2015; Wardrop, 2018). In addition, heterogeneity in environmental conditions across smaller administrative levels may not be captured completely within aggregated data (Stevens, 2015). The need for accurate, reliable spatial data has become increasingly important for future dengue research and public health decision making (Wardrop, 2018). Dengue risk predictions for upcoming epidemics have also become relevant in public health decision making as human populations continue to experience increased transmission rates in response to these changes in climate as well as increased interactions with mosquito vectors (Lauer, 2018; Stolerman, 2016; McGough, 2021). Current predictions are insufficient in illustrating the true distribution of dengue risk within smaller regions, and improvements continue to be necessary (Lauer, 2018; Stolerman, 2016). At present, there is limited research on the value of widespread municipality-level dengue predictions for outbreak risk. To date, only one other study has made use of focused municipality predictions for dengue outbreak risk, specifically utilizing boosted regression trees, or BRT with relevant climate predictors such as temperature, precipitation and elevation (Ashby, 2018). This research found that using a BRT modeling approach was adequate in providing more accurate representation of spatial dengue distribution for mapping of annual dengue incidence (Ashby, 2018). However, this study only predicted yearly dengue case counts, which do not provide essential information on real-time outbreaks. Monthly municipality-level predictions are more useful for public health interventions, specifically because they can offer improved insight on transmission dynamics throughout the course of an outbreak and can be used to improve vector control strategies and resource deployment.

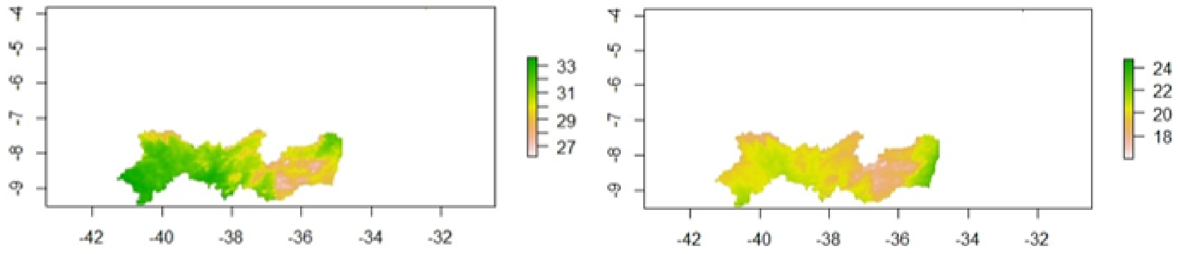
Additionally, the introduction of lagged variables in other models has proven much more successful in adequately creating forecasts than traditional climate factors on their own (Lowe, 2018). Lagged variables, or specifically climate lags, can be defined for our purposes as delayed effects of climate that cause changes in dengue incidence. For example, lagged times of 1 to 2 months for excess rainfall can increase current dengue risk (Lowe, 2018). Lowe et. al found that using bi-dimensional climate lag variables in a series of distributed lag non-linear models had overall accuracy of about 86%, compared to just 64% using their baseline model (Lowe, 2018). Duration of lagged temperature effects, most typically 1-2 months, have been shown to last 141-150 days in Guangzhou, China, which is equivalent to the entire dengue epidemic season (Xiang, 2017). These sustained lagged effects throughout the peak season can then trigger rapid growth of mosquito vector populations, leading to amplified transmission (Xiang, 2017). Another study found that precipitation lags can be highly predictive of incoming dengue epidemics if they occur closer to the end of dry season in India (Kakarla, 2019). It has also been noted that interestingly enough, lower amounts of rainfall with lags of 1-2 months can actually be indicative of higher dengue incidence in some areas (Chang, 2016), likely due to increased water storage in these areas. In terms of surveillance data, Jain et. al determined that both short-term (1-2 months) and long-term (12 months) prior dengue surveillance data was statistically significant in influencing real-time dengue predictions (Jain, 2019). Another analysis found that dengue data lagged at 2 months and 24 months were the most optimal predictors for dengue incidence using generalized linear models (Ramadona, 2016). Utilizing significant climatic and additional dengue surveillance lagged effects for outbreak forecasting is important because they are most likely to contribute to real-time increased monthly dengue incidence, accounting for both changes in mosquito population dynamics that can take weeks to manifest as well as high, sustained transmission rates among at-risk human populations.

Working with this information, we posited that by forecasting dengue outbreaks at the municipality level using both prominent climatic predictors and related lagged effects of climate and case data, we have the potential to better estimate and map monthly dengue outbreak risk in Brazil for future improvements and innovations to local vector control. We explored this hypothesis by building a BRT model using minimum and maximum monthly temperatures, as well as monthly average precipitation in Pernambuco as our main predictors, with a binary response corresponding to either high risk or low risk for dengue. We then built a final model based on our preliminary findings, integrating the most influential lagged effect variables for both climate and cases to determine potential for improvement in our forecasting accuracy.

## 5.3 Materials and Methods

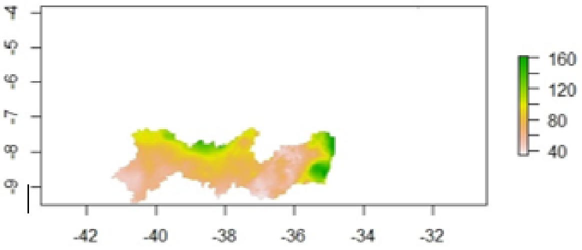
We focused solely on the state of Pernambuco for the purposes of this analysis. Pernambuco is a state in the Northeast region of Brazil, which borders the Atlantic Ocean. Northeast Brazil, along with Southeast Brazil, consistently reports the highest number of dengue cases annually compared to other local regions (Teixeira, 2013). Pernambuco's capital, Recife, is a large, densely populated municipality that encompasses several industrial towns (Britannica, 2016). Pernambuco as a whole currently comprises 185 municipalities, all with varying climate patterns, making it an excellent candidate for our initial study. More specifically, Pernambuco has three distinct climate regions. Its low coastal region narrowly spans approximately 116 miles along the coastline and typically experiences consistently hot and humid climates, with up to 1800 mm of rainfall annually (Britannica, 2016). This coastal zone is separated from the higher plateau region by the agreste zone, which is characterized by slightly drier climate and light vegetation (Britannica, 2016). Lastly, the sertão region makes up the inland plateau of Pernambuco and is cold, stony and very dry. This region is also prone to experiencing droughts (Britannica, 2016).

All data cleaning and processing was done using R version 3.6.3 and associated packages. Shapefiles containing the most recent configuration of municipalities for Brazil were imported directly into R from the Brazilian Institute of Geography and Statistics (IBGE). Our Brazil spatial object was then subset to the state of Pernambuco, containing 185 total municipalities. Case data were downloaded and imported into R as CSV files for each year of available data. Case data were formatted in a monthly timescale from January to December for each year and reformatted to include all monthly values in a single column chronological time-series format spanning across the specific time period chosen. The case data collected were matched to the previously defined administrative boundaries for each municipality and for each month of our analysis. Municipalities missing case data for a specific month were classified as ‘NA’. Data from our three primary climate variables were available as TIF files, which were downloaded and imported into R as rasters and cropped to the dimensions of Pernambuco using our Pernambuco shapefile for reference (Figure 1). Specific values per month and per municipality were extracted by pixel and averaged from each climate raster and formatted into a single column chronological time-series format to match our case data, shown in Figure 2. Monthly incidence rates per municipality were used as our outcome of interest. Incidence rates per 100,000 were calculated by dividing monthly case numbers by monthly population estimates and multiplying by 100,000 (see Appendix A.1.2 Calculations). Appendix Figure 2 depicts total monthly case counts across Pernambuco during our timeframe of interest.



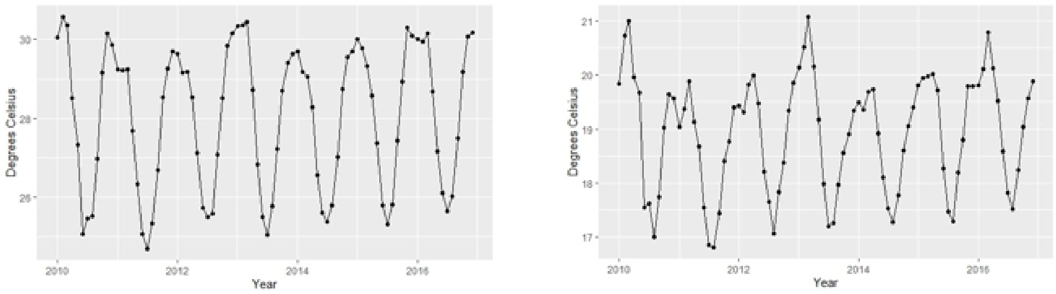
a) Maximum monthly temperatures (°C), January 2010.

b) Minimum monthly temperatures (°C), January 2010.



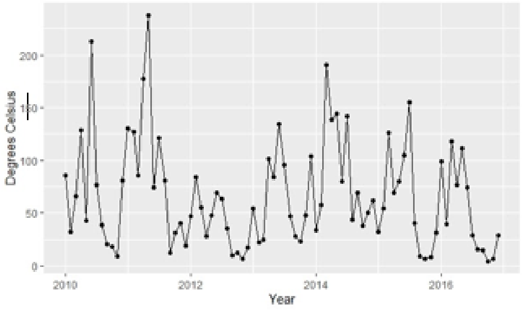
c) Average monthly precipitation (mm), January 2010.

**Figure 1 Climate data rasters cropped to Pernambuco state, January 2010.**



a) Average monthly maximum temperatures, January 2010 to December 2016.

b) Average monthly minimum temperatures, January 2010 to December 2016.



c) Average monthly precipitation, January 2010 to December 2016.

**Figure 2 Average monthly climate values in Pernambuco, 2010-2016.**

Specific municipalities were excluded from our final analyses if they had two or more years of missing data. Specific years of data had to have less than 20 percent of total data missing in order to be considered for inclusion, and there had to be at least 4 years of consecutive data to work with in order to select a comprehensive timeframe. Based on these exclusion criteria, our final timeframe for analysis was selected as the period from January 2010 to December 2016, with a starting total of 129 municipalities with complete non-missing data (69%). Municipalities missing only one year of case data that did not occur at either the beginning or the end of our time period were imputed; those that were missing only year of case data at the beginning or end of our timeframe were excluded from analysis. Thirty-eight of the municipalities evaluated in our initial dataset were missing annual observations (i.e. observations missing across an entire year), but none were missing across a monthly timescale. To mitigate this issue and to increase the number of observations available for our use, we imputed these values using a simple moving average imputation using the imputeTS version 3.2 package in R. Case observations from the season prior as well as the season after the missing observation were averaged to provide an imputed value (i.e. January 2010 and January 2012 averaged for an imputed January 2011 value). After imputations were included, our final model included data from 167 municipalities out of our original 185 (90%). Our initial workflow can be found in Appendix Figure 3.

With complete case data across our time period from 167 municipalities, we had a total of 14,028 observations to incorporate into our model with three initial predictors -- maximum monthly temperature, minimum monthly temperatures, and average monthly precipitation. Our initial response was cases per 100,000 people, or the incidence rate (IR). In accordance with prior literature guidelines (Brady, 2015), we set a fixed incidence threshold of 100 cases per 100,000 to classify our response values correspond with higher risk of transmission (greater than 100) or lower risk of transmission (less than 100).

To further explore these associations and determine whether seasonal lag effects significantly contribute to accuracy of predictions, we added lag variables of 1-3 months for each of our main predictors to our dataset as well as lags of 1-3 months and 1 year (prior season) for our case data. After several runs with these additional lags and thorough exploration of existing literature, we found that minimum temperature and precipitation lags of 2 and 3 months were the best climate lags for predictions. Lagged cases of up to 1 month were highly influential for future predictions and were also included. Prior seasonal case data did not have a significant effect on predictions, and was excluded from our models. Final predictors used for analysis are described in Table 1.

**Table 1 Variables collected for analysis. Population used only for calculation of incidence rates.***

| Type | Variable Name(s) | Source | Description |
|---|---|---|---|
| **Cases** | Month 1st symptoms by municipality of residence | DATASUS Brazilian Ministry of Health | Municipality level case estimates |
| **Climate** | tmin_2000-2009, tmin_2010-2018 | WorldClim | Global monthly minimum temperatures |
| **Climate** | tmax_2000-2009, tmax_2010-2018 | WorldClim | Global monthly maximum temperatures |
| **Climate** | prec_2000-2009; prec_2010-2018 | WorldClim | Global monthly average precipitation |
| **Population*** | Resident population estimates for TCU-Brazil | DATASUS Brazilian Ministry of Health | Annual residential population estimates |

We utilized a boosted regression tree (BRT) model for our predictions. Boosted regression trees are an ensemble learning method that involves fitting an initial regression tree and iteratively improving upon each subsequent tree created by accounting for the variation in the response that is not explained by the model and incorporating it into fitting of the next tree through gradient boosting (Elith, 2008; Messina, 2019). There were several strengths to using a BRT model for our classifications. A BRT approach was advantageous to use over other machine learning algorithms because it is able to model complex non-linear relationships and there is no need for prior data transformation (Elith, 2008; Messina, 2019). In addition, BRT models can also manage interaction effects between predictors, which is important when working with closely correlated variables, as such with our climate data here (Elith, 2008). Lastly, as an ensemble method, BRT algorithms generally tend to have higher predictive performance over most other modeling approaches (Elith, 2008). Here we employed a baseline BRT model using a Bernoulli loss function with a binary response to predict either high dengue outbreak risk, or low dengue outbreak risk. We then split our data into a training set containing case values from 2010-2015 to predict our test set of 2016 case values. In total, this gave us 12,024 training observations and 2,004 testing observations. We used the 'gbm' version 2.1.5 and 'dismo' version 1.1-4 packages in R to fit and evaluate our model with the following parameters: learning rate = 0.015, tree complexity = 3, family = bernoulli, and bag.fraction = 0.75. These parameters were chosen as they collectively performed best during trial runs with our training data. All other parameters were kept at default and not specified in our call.

The gbm.step function in R was first used to determine the optimal number of trees for our model by finding where the holdout deviance of the model was minimized. Using 10-fold cross validation, our training data was randomly subset into 10 equal subsamples, each comprising 9 unique subsets of data with the last subset being used as a validation set. This created 10 initial BRT models consisting of 50 trees each. Mean performance and standard error was recorded before stepping forward by our defined learning rate and increasing the number of trees by a constant amount for each model. This step was repeated for several iterations until minimum error was reached. The point at which this occurs then tells us the optimal number of trees. Optimal number of trees to include in a model can vary; we were aiming for at least 1000 trees as a basic rule of thumb (Elith, 2008), but were careful to not add so many trees so as to not overfit our model to our training data. This combination of optimal trees along with our previously defined parameters allowed us to then fit a final model to the available data and predict monthly transmission probabilities for the year 2016.

BRT predictor associations were first assessed using the ‘summary’ function in R on our fitted BRT model. This gave us the relative influence, or importance of each variable, which accounts for the percent reduction in the loss function given the features provided. After assessment of our BRT models was completed, we decided to further assess our predictor associations through univariate and multivariate logistic regression analyses using the glm function from the ‘stats’ base package in R on all variables included in our model iterations. Eight univariate regressions were completed for each single variable vs. our response of low or high risk for dengue. Our multivariate regression included all climate variables, lagged climate variables and prior surveillance data. Due to the fact that all variables had different scales and some had non-normal distributions, all predictors were normalized via log transformation prior to logistic regression. We operated on a significance level of $p < 0.05$ when determining which variables were statistically significant. For our multivariate regression, all variables were initially input into the logistic regression model. Variables that were not statistically significant were removed in a sequential order, beginning with the highest p-value. Variables were removed until only statistically significant variables remained.

## 5.4 Results

Our train/test split for predictions was roughly 80/20, because we decided to include 2015 into our training data. We did this because 2015 was a high outbreak year, especially compared to prior years, and overall we wanted to utilize this data to train our model to more accurately distinguish between outbreaks vs. non-outbreaks.  Using these parameters and our three initial predictors, we were able to fit a baseline model of 1300 trees to our training set with an AUC of 0.79, shown in Figure 3. Preliminary evaluation of our model using our testing set gave an AUC of 0.69. We further examined these results by setting a threshold of 0.5 to separate our prediction probabilities with any predicted value less than our threshold corresponding to the likelihood of a low risk prediction and any predicted value greater than our threshold corresponding to the likelihood of a high risk prediction. We then created a confusion matrix to assess the total number of correct vs. incorrect classifications produced.

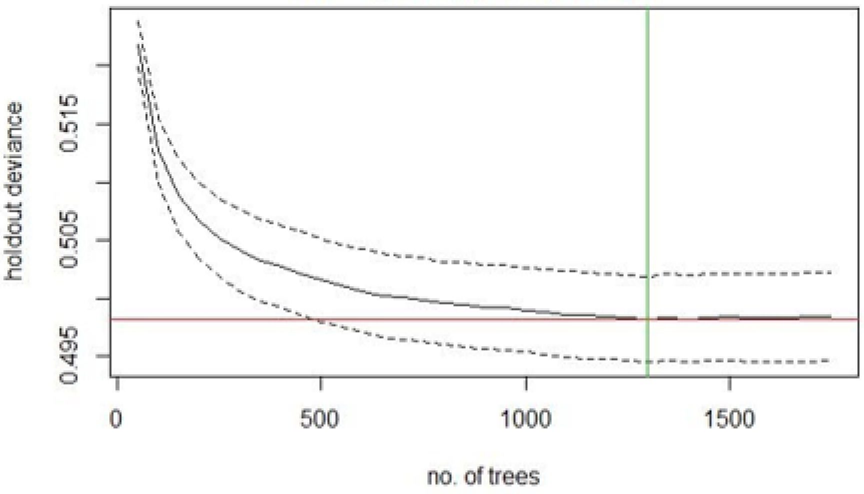


**Figure 3 Graph showing holdout deviance of our climate model. Holdout deviance, specified by the red line, was minimized at 1300 trees as indicated by the green line. The solid black line indicates mean deviance, and the dotted black lines indicate ± 1 standard deviation.**

Out of 2,004 total values, 1,751 were correctly predicted by our preliminary model (Table 2). Our model was highly sensitive (0.998) when detecting areas of low outbreak risk (1735 out of 1738 total low risk observations) but had greater difficulty determining areas of higher risk (specificity 0.060). Out of 266 total high risk observations, our model could only correctly predict sixteen observations. Overall accuracy of the model was 0.87 with a 95% confidence interval between 0.858 and 0.888. Our total misclassification rate was calculated as 12.6%.

**Table 2 Confusion matrix for climate model with outbreak risk predictions.**

| | **Reference** | |
|---|---|---|
| **Prediction** | low | high |
| low | 1735 | 250 |
| high | 3 | 16 |

Of the sixteen correctly predicted high risk classifications, all had minimum temperatures of at least 22 °C. Most of these classifications also had maximum temperatures greater than 29.5 °C, and/or heavy precipitation (> 150mm), with several municipalities averaging over 200 mm monthly. Several of the 'non-outbreak' classifications had either very low minimum temperatures (< 20 °C), low precipitation (<100mm), or both. There were numerous outliers within our data that did not follow these general patterns. Several incorrectly predicted high risk municipalities throughout 2016 had low monthly temperature ranges, typically between 17 and 27 °C. These municipalities also had much lower monthly precipitation rates. What was interesting about these patterns were that they were not just seen in off season months -- they were also noticed across peak dengue season in some areas as well, when temperatures and rainfall are most consistently high.

When looking at strength of our predictors, minimum temperatures had the greatest relative influence on initial predictions, while precipitation and maximum temperatures did not have as strong of an influence on initial outbreak classifications (Figure 4). Fitted values from our initial model in relation to our climate predictors can be seen in Appendix Figure 4. There appears to be a positive correlation between minimum temperatures and probability of a high risk prediction, as seen in the second panel, which was later tested through regression analyses. Both maximum temperatures and precipitation in relation to our fitted values had weak positive correlations, although there were several noticeable outliers (Appendix Figure 4, panels 1 and 3).

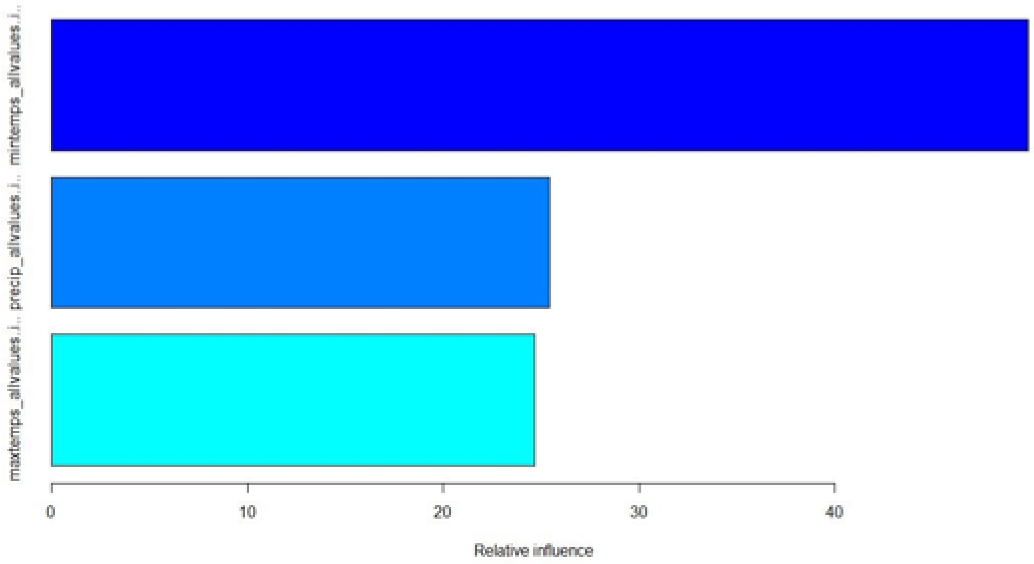


**Figure 4 Relative influence of the three primary climate variables over initial predictions.**

The univariate logistic regressions for minimum temperatures, maximum temperatures, and precipitation against our response of low or high risk revealed that all three variables included in the first part of this analysis were highly correlated to dengue risk. Out of these major climate predictors, minimum temperatures had the most statistically significant correlation with increased dengue risk and the highest beta coefficient (Table 3).

**Table 3 Univariate regression results for the primary climate predictors.**

| | Beta (β) coefficients | p-value |
|---|---|---|
| Maximum temperatures | 4.3025 | <0.001*** |
| Minimum temperatures | 8.4242 | <0.001*** |
| Precipitation | 0.4185 | <0.001*** |

Significance levels: p<0.05 = *, p<0.01 = **, p<0.001 = ***

With this information, we built our next model, which consisted of our three primary climate predictors, lag variables of 2 and 3 months for both monthly minimum temperature and precipitation, and a lag variable of 1 month for cases for a total of eight predictors and again using IR per 100,000 as our response. The gbm.step function fit a final model of 1600 trees using 10-fold cross validation at holdout deviance of 0.265 (Figure 5). Using the same function parameters as our initial run, re-running the model produced a training AUC of 0.96 and a testing AUC of 0.94. Our final confusion matrix (Table 4) yielded 1,842 correct predictions and 162 incorrect predictions out of 2,004 observations. This gave a total accuracy of 91.9% with 95% confidence interval between 0.906 and 0.931 and a misclassification rate of 8.1%.

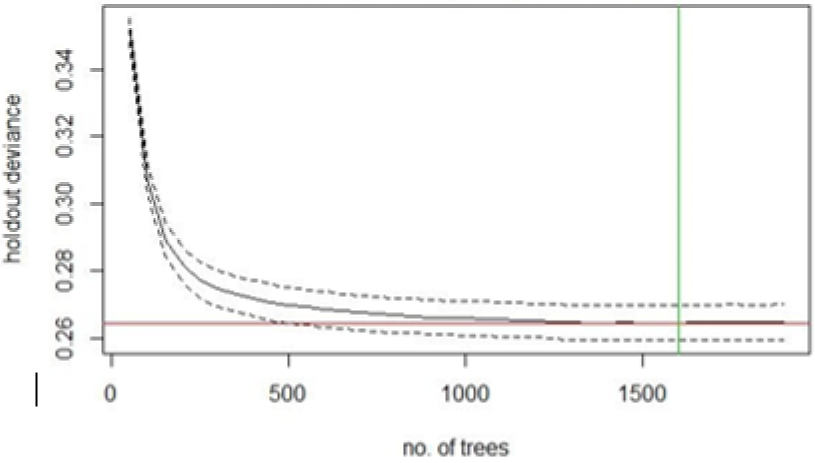


**Figure 5 Graph showing holdout deviance of our climate lag and cases model. Holdout deviance, specified by the red line, was minimized at 1600 trees as indicated by the green line. The solid black line indicates mean deviance, and the dotted black lines indicate ± 1 standard deviation.**

**Table 4 Confusion matrix for climate lag and case model with outbreak risk predictions.**

|  | Reference |  |
|---|---|---|
| **Prediction** | low | high |
| low | 1643 | 67 |
| high | 95 | 199 |

This final model was again highly sensitive in determining areas of low dengue risk (0.945), but improved greatly in discerning higher risk municipalities as well (specificity 0.748). In comparison to our previous run, which produced sixteen correct high risk predictions, the addition of specific lagged variables produced 199 correct high risk predictions. Adding prior case data greatly improved the model, and had the highest relative influence in making predictions. Both maximum and minimum temperatures also slightly impacted predictions, while lagged climate variables were less influential (Figure 6).

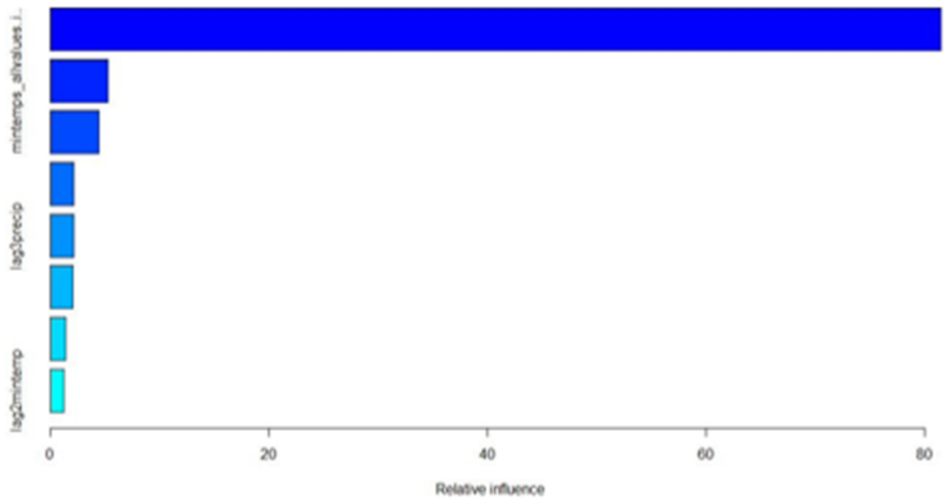


**Figure 6 Relative influence of all climate, lagged effect and prior case covariates on outbreak predictions.**

In looking at our fitted prediction values in relation to all of our original and newly included predictors for this model iteration (Appendix Figure 5), we saw a very strong positive association between 1-month lagged cases and probability of a high risk prediction, but no other distinct correlations between our fitted values and other predictors were established.

As with the first part of our analysis, univariate logistic regressions were completed for the rest of our predictors included in our updated BRT model. These regressions revealed that nearly all variables included in this part of the analysis were highly correlated to dengue risk. The only predictor included that did not have a statistically significant association with dengue risk was 3 month lagged precipitation, with a p-value of 0.429 (Table 5).

**Table 5 Univariate regression results for lagged climate effects and prior surveillance data.**

| | **Beta (β) coefficients** | **p-value** |
|---|---|---|
| **2-month lagged minimum temperature** | 9.5041 | <0.001*** |
| **3-month lagged minimum temperature** | 8.4083 | <0.001*** |
| **2-month lagged precipitation** | 0.1906 | <0.001*** |
| **3-month lagged precipitation** | 0.0209 | 0.429 |
| **1-month lagged cases** | 0.7357 | <0.001*** |

Significance level: p<0.05 = *, p<0.01 = **, p<0.001 = *

Multivariate regression analysis was then performed using all variables included, and revealed that the least significant predictors for dengue risk when compared to all other factors were average monthly precipitation as well as 2-month lagged monthly precipitation (Table 6). Both of these variables were removed from our final regression analysis. All other variables were statistically significant at level $p < 0.05$ and were kept in this additional regression model. Of the remaining predictors, 1 month lagged cases had the highest association with dengue risk, and 3 month lagged precipitation had the lowest association with dengue risk (Table 6).

**Table 6 Multivariate regression results with all significant climate and case predictors.**

| | **Beta (β) coefficients** | **p-values** |
|---|---|---|
| **Maximum temperatures** | 0.0802 | 0.0069** |
| **Minimum temperatures** | 0.1759 | 0.0015** |
| **2-month lagged minimum temperatures** | 0.2523 | 0.0015** |
| **3-month lagged minimum temperatures** | -0.2098 | 0.0015** |
| **3-month lagged precipitation** | -0.0021 | <0.001*** |
| **1-month lagged cases** | 0.7089 | <0.001*** |

Significance level: $p<0.05$ = *, $p<0.01$ = **, $p<0.001$ = ***

One last BRT iteration was run using only the statistically significant variables shown in Table 6 for outbreak risk prediction to determine if predictions would improve from prior models. Training and testing AUCs from this iteration were 0.942 and 0.938, respectively. This iteration was able to predict 1,837 observations (Appendix Table 2) correctly with a misclassification rate of 8.3%, and were comparable to results from our final model. However, the BRT model that included all eight variables studied was still the best-performing model for final predictions.

Overall, our final BRT model including climate lag data and prior dengue surveillance data dramatically improved in predicting to high risk regions over our initial model iterations. Our false positive rate for high risk predictions was significantly reduced from approximately 94% to only 25% after inclusion of relevant lag variables (Table 7). Mapped predictions from model iterations for climate predictors only and for all climate variables are shown in Figure 7. Mapped predictions for our statistically significant predictors model were not included, as they did not improve outbreak predictions. All model evaluation metrics are summarized in Table 7, and AUCS and predictions are summarized in Table 8.

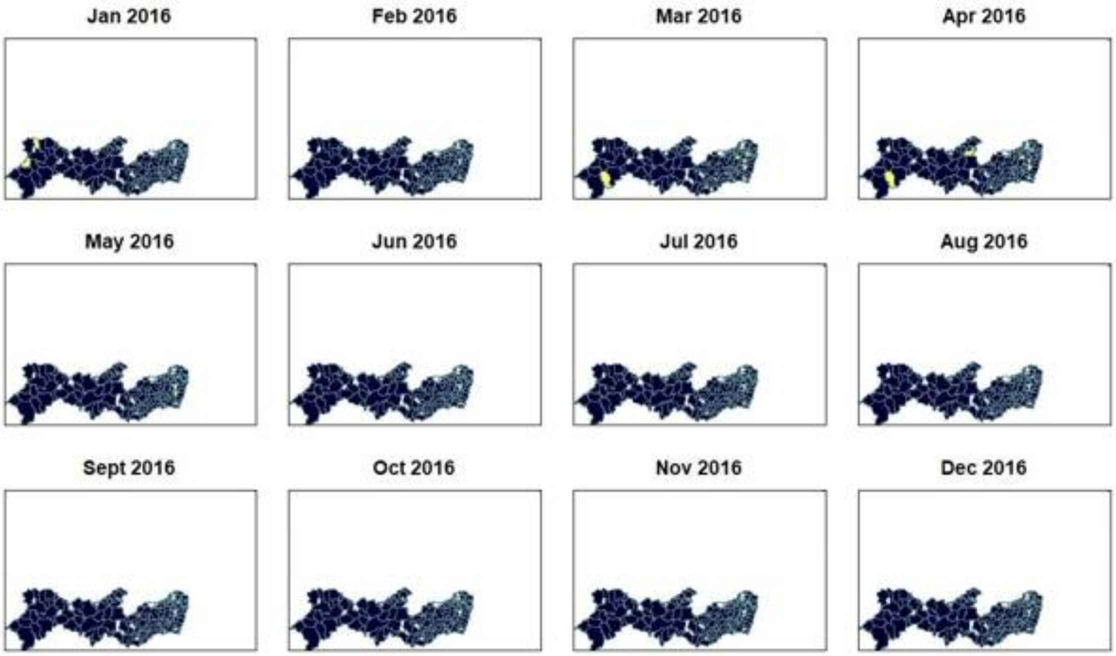
**Table 7 Comparison of evaluation metrics for all climate models. For overall prediction accuracy, the all predictors model was best.**

| | **Accuracy** | **Sensitivity** | **Specificity** | **FPR** | **FNR** | **p-value** |
|---|---|---|---|---|---|---|
| **Climate only model** | 0.8737 | 0.9983 | 0.0602 | 0.9398 | 0.0017 | 0.206 |
| **Significant predictors only model** | 0.9142 | 0.9396 | 0.7481 | 0.2519 | 0.0604 | 3.07 e-11 |
| **All predictors model** | **0.9192** | **0.9453** | **0.7481** | **0.2519** | **0.0546** | **1.9 e -13** |

**Table 8 Summary table of AUCs and predictions for all climate models.**

| | **Training AUC** | **Cross-validation AUC** | **Testing AUC** | **# correct predictions** | **# incorrect predictions** |
|---|---|---|---|---|---|
| **Climate only model** | 0.796 | 0.722 | 0.698 | 1735 | 253 |
| **Significant predictors only model** | 0.942 | 0.958 | 0.938 | 1837 | 167 |
| **All predictors model** | **0.964** | **0.945** | **0.938** | **1842** | **162** |

a) Predictions from our climate only model.



b) Predictions from our all climate predictors model.

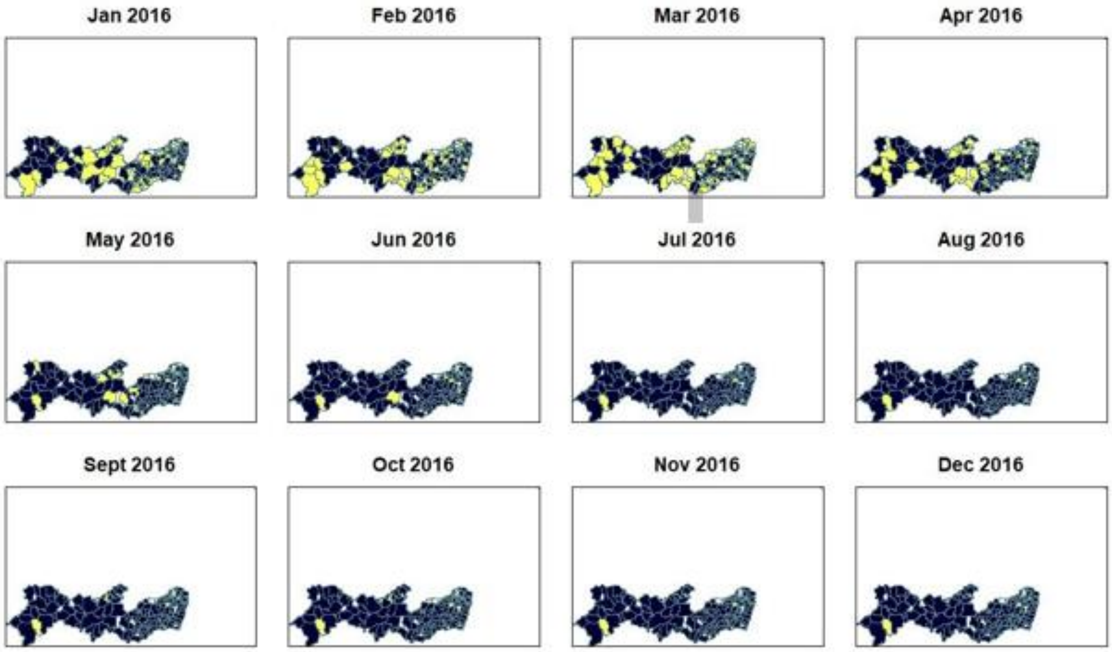


**Figure 7 Preliminary (top) and final (bottom) mapped outbreak predictions for 2016 from the climate only and all climate predictor models, respectively. Navy regions correspond to low risk predictions, yellow regions correspond to high risk predictions. Regions in white could not be predicted due to lack of data.**

## 5.5 Discussion

Our preliminary model was not satisfactory in predicting areas of potential outbreaks in Pernambuco based solely on our three primary climate predictors. Our initial model was good at predicting areas with low outbreak potential but could not effectively discern high outbreak areas when using traditional climate factors alone. From analyzing our testing set data, we found that most of our misclassified municipalities were also municipalities with relatively high population densities, leading us to believe that it could significantly contribute to improved accuracy of future risk classifications. Increased population density as well as poor urban infrastructure in tropical areas have previously been shown to contribute to increased dengue transmission (Wu, 2009). Additionally, urbanization around dengue-prone regions can increase the number of artificial water sources available, leading to increased mosquito breeding sites (Wu, 2009).

As expected, our results generally followed seasonal trends. Brazilian summers run from December to March, which also coincides with its primary rainy season from January to May. Average precipitation peaked for most areas in January and March, and tended to trend much lower between June and December. Overall temperature ranges for most municipalities from January to May fell between 20 and 31 ℃, with temperature ranges mostly between 17-27 ℃ during winter months. Municipalities residing within Pernambuco's coastal and surrounding regions tended to have higher overall temperatures and much more precipitation, while municipalities located in the sertão or agreste regions typically saw lower overall average temperatures and scarcer patterns of rainfall.

Most of the incorrect predictions by our initial model were areas that were predicted as low outbreak risk, although they were identified as true high risk observations. One noticeable trend seen in the raw data was that several of these incorrectly predicted areas had low minimum temperatures or temperature ranges and/or low precipitation rates during peak dengue season months. Out of 167 municipalities, five had 6 or more incorrect predictions over the course of a year. Most of these municipalities were incredibly population dense -- for example, Recife was consistently predicted incorrectly in our initial run in spite of being a hotspot for dengue outbreaks. One potential explanation as to why these areas were frequently incorrect is that outbreaks are also dependent on non-climate factors such as population density or poverty levels, which our model did not account for. Several of these incorrectly predicted regions were also located near or around the coastline/coastal zone. Typically this region is both hot and rainy during the summer months, however precipitation drops dramatically during the winter with temperature staying relatively consistent. This is also likely to have affected our predictions, as nearly all of the previously discussed municipalities had several incorrect predictions between June and December, with very little (<50mm) of precipitation recorded although temperature ranges remained moderate to high.

Minimum temperatures appear to have the most influential role in determining whether or not an area will be classified as having high outbreak risk based on our initial results. Typically we saw that higher minimum temperatures tended to trend towards producing high risk predictions in our initial BRT model, which was corroborated through univariate regression analysis which revealed a high and statistically significant correlation with increased dengue risk (see Table 5) . This association became stronger in conjunction with higher rates of precipitation, which was also statistically significant in univariate regression analysis, but only at baseline and 2 months. 3 month lagged precipitation rates showed no significant correlation for dengue risk, and had the lowest association to dengue risk overall in our multivariate regression. Maximum temperatures did not appear to have as strong of an association in our BRT models to predictions when compared to our other two climate variables, although there was a statistically significant and high correlation between maximum temperature and dengue risk in both univariate and multivariate regression analyses. This is generally consistent with previous literature, as temperature, and especially minimum temperature, is a key component of dengue transmission (Reinhold, 2018). Female A. aegypti mosquitoes were found to feed faster and more often at higher temperature ranges (between 26-35 ℃) versus temperature ranges between 19 and 25 degrees ℃ (Reinhold, 2018). Larvae development and hatching is optimal at around 32 ℃, with shorter development times associated with higher temperatures (30 ℃) (Reinhold, 2018).

Lagged effects on predictions influenced our model in different ways. The greatest influence on our monthly predictions in our BRT models was inclusion of lagged case data by 1 month, which was also highly statistically significant in regards to predicting dengue risk in both univariate and multivariate regressions (Figure 6). This could be potentially due to the fact that areas with consistently high cases also tend to be highly population dense – several of our previously incorrectly predicted areas contain large, urban cities that are more likely to sustain higher case numbers throughout an epidemic, and thus are more likely to be influenced by prior case data.

With additional lagged effects included in our model, we saw a significant improvement in testing accuracy for our high risk predictions. Specifically, we calculated an over 1100% increase in correct high risk predictions after running our final model (see Appendix A.1.2 -- Calculations). This increase in specificity for our high risk predictions did come at a cost, as our sensitivity in predicting for low risk areas had a 5.31% loss (see Appendix A.1.2 -- Calculations). However this was an insignificant loss, as we were able to still able to achieve higher overall prediction accuracy with our final model and were able to successfully predict to more high risk areas that are more likely to need vector control resources in general.

Of the 167 municipalities forecasted in our final run, only three had greater than 3 overall incorrect predictions, and no municipalities had 6 or more incorrect predictions for 2016 which was a marked improvement from our original model. Generally, most of these incorrect predictions tended to occur in the first five months of the year, coinciding with Brazil's rainy season. There were no major seasonal trends observed with our incorrect predictions that differed from our original model. We did see that most areas incorrectly predicted as low risk in this run tended to also have lower 1 month case lags (< 100 cases per 100,000), while the opposite was true for areas incorrectly predicted as higher risk. This is in concordance with the assumption that areas of high transmission see consistently higher case numbers across epidemic months, and vice versa.

A major advantage to modeling municipality level outbreaks is the potential for improvements in providing localized dengue treatments and interventions to specific high risk areas versus applying an all-inclusive approach to an entire state or country. Standard vector control methods primarily involve *Ae. aegypti* surveillance, as well as general pesticide spraying, though more community-centric measures, like educational campaigns, are infrequent (Maciel-de-Freitas, 2014). Community engagement practices such as encouragement and subsidy of individual repellants, coordination of neighborhood association surveillance and increased community reporting of breeding sites are all positive impacts that can be implemented through application of the research established here. Additionally, dengue disease burden is high in Brazil, and typically costs millions of dollars in direct and indirect costs per year. Between 2010 and 2015, the Brazilian Public Health System spent approximately 159 million USD and 10 million USD to treat dengue and severe dengue respectively, without accounting for patient self-treatment (Piassi Godói, 2018). Greater community level practices could help alleviate these high costs by providing more reliable local dengue control, which may decrease overall incidence as well.

Historically, another issue with controlling dengue transmission is the inability to effectively control the vector. Both Ae. aegypti and Ae. albopictus populations have remained stable or even increased populations in urban areas as well as in higher elevations (Fernández-Salas, 2015). This is in part due to poorly managed urban development/planning (i.e. lack of trash collection, poor water supply, and uncontrolled population growth) but also due to weak vector control strategies (Fernández-Salas, 2015). For example, poorly coordinated spatiotemporal vector control methods can cause previously mosquito-free regions to become re-infested from nearby untreated areas (Fernández-Salas, 2015). Additionally, these methods heavily rely on insecticide use as a primary means of control, which tends to lead to resistance during high transmission seasons (Maciel-de-Freitas, 2014).

Current vector control programs are designed and run by the federal government, and individuals are not typically held accountable to providing adequate dengue control measures for their own surroundings (Maciel-de-Freitas, 2014). Between epidemics, there is also little to no allocation of community-based resources focused on prevention awareness and very little education on the matter (Maciel-de-Freitas, 2014). Thus, large-scale solutions to controlling dengue outbreaks in Brazil have consistently proven to be unsuccessful (Maciel-de-Freitas, 2014). Previous endeavors in other countries have shown that there is great value in establishing smaller, community-based control programs to limit mosquito vector populations during peak transmission (Maciel-de-Freitas, 2014). For example, a prior study conducted in Singapore created a community initiative to eliminate mosquito breeding sites that led to a vast decrease in dengue outbreaks (Maciel-de-Freitas, 2014). With this in mind, small-scale engagement programs should be utilized more often to better prevent and reduce mosquito density, especially in higher-risk areas.

Based on this information and our own findings, we hope to encourage restructuring of current dengue control systems to place greater emphasis on local community-based dengue control strategies, including but not limited to, appropriate waste reduction, improved sanitation practices, and increased health education.

Although our model was able to produce good predictions, there are several noted limitations to this work. First and foremost, data collection was substandard for several reasons. In particular, the Brazilian Ministry of Health's DATASUS portal from which we collected our case numbers had vast amounts of missing data from municipalities without explanation despite being the most comprehensive dengue reporting system for the country. The proportion of missing municipality level data was also not consistent across years, and missing data was recorded as high as up to 45 percent for some years. These issues greatly constrained the size of both our timeframe and the amount of available data for use. This also meant that we were unable to predict to all areas of Pernambuco, as there was not enough viable data to produce predictions for 18 municipalities.

In addition, we imputed case values for 38 different municipalities across our dataset using a simple moving average to increase the number of available observations for analysis. These imputations were based on the assumption that both the seasons prior and seasons after our missing values for each municipality were likely to have similar climate profiles and thus similar case values, however this method does not account for annual variations in climate patterns or other factors such as serotype circulation dynamics, which have been linked to both local transmission rates and severity of disease (Teixeira, 2013; Andrioli, 2020). With that, our imputations may not completely reflect true dynamics of dengue transmission within these areas and may not provide the most accurate predictions for our imputed municipalities.

Finding a good baseline for outbreak definition can also be difficult. For dengue specifically, complex transmission dynamics and high variability in reporting strategies make it hard to pinpoint one specific definition (Brady, 2015). Common methods for determining dengue outbreaks in Brazil include calculation of a monthly mean as well as setting a fixed incidence threshold (typically 100 or 300 per 100,000 people for this purpose) (Brady, 2015). We chose to set a fixed incidence threshold based on this prior knowledge for consistency across our municipality level predictions, though we again acknowledge that this may not represent true dengue risk throughout all areas and could also contribute to the overall prediction inaccuracy.

Lastly, all data that was used here comes from public data sources, and reflects very broadly defined study areas. Our initial explorations into this work are useful, but may not be entirely relevant to finer scale areas in terms of intra- and inter-neighborhood dynamics and should be further investigated. For future analyses, private data sources with more localized (i.e. GIS) information may be able to help better identify specific neighborhoods within municipalities for resource distribution, and may provide traction for group engagement practices and increased dengue knowledge within communities.

Population density and urban growth rates appear to have the potential to moderately influence dengue outbreak predictions and is a prime consideration based on additional analysis of our raw data. Rapid urbanization and population growth have previously been attributed to worldwide dengue resurgence as it allows for greater human exposure to the Ae. aegypti vector (Matysiak, 2017). Uncontrolled population growth in endemic regions can also strain local resources, particularly water supply and waste removal, creating standing water sources where mosquito breeding sites can prosper (Matysiak, 2017).

Land use and deforestation rates are also important for future analyses. Other studies have found that reduction of forest cover is correlated with occupational migrations, leading to increased human-mosquito exposure and thus higher transmission risk (Husnina, 2019). Certain vegetation and land cover features can also impact dengue transmission via modification of mosquito habitat, causing Ae. aegypti vectors to colonize new areas and expand their range (Sarfraz, 2012).

From using Pernambuco as an example state for predictions, we can definitively conclude that climate, and more specifically temperature and rainfall, do play major roles in predicting small-scale outbreaks but are not adequate in making solid predictions on their own when using a BRT approach. Accounting for both prior seasonal lag effects and prior monthly case numbers greatly improved our forecasting accuracy, although there was an absolute trade-off between producing more correct high risk predictions vs. producing more low risk predictions. With this in mind, municipality level outbreak predictions can and should be improved with additional parameters that have also been noted to influence transmission, expanding upon both climatological attributes of an area as well as sociodemographic factors to paint a more comprehensive picture of localized dengue risk which can then be applied to direct future public health interventions.

# 6.0 Social and Environmental Predictors of Health in Dengue Outbreak Forecasting and Potential for Improving Future Vector Control Policy

Manuscript in preparation for submission.

## 6.1 Abstract

Social and environmental determinants of health have been previously associated with risk of dengue outbreaks, although specific correlations have been weak or understudied. Building upon our prior work in evaluating climate and dengue risk associations, we added eight new predictors to our models, which included access to piped water and sanitation services, population density, and percentage of urban space within municipalities. We hypothesized that the addition of previously correlated social and environmental risk factors into our boosted regression tree models would improve our dengue outbreak predictions when coupled with our previously defined climate variables. The best model iteration in this analysis was able to produce 1,830 correct predictions out of 2,004, with a misclassification rate of 8.7%. Training and testing AUCs for this model were 0.97 and 0.938, respectively. While we found that ultimately the covariates included here had higher specificity in correctly predicting high outbreak risk regions when they were classified as high risk compared to our previous models, sub-national dengue outbreak forecasting models should rely mainly on climate and prior surveillance data to make the most accurate overall outbreak predictions. Our investigation shows that while there is some value in utilizing social and environmental determinants of health in forecasting dengue outbreak potential for municipalities,

at such a small scale these effects are largely insignificant without further research. Moving forward, these results have the potential to redirect focus on improving real-time vector control epidemic response as well as further developing community-based engagement practices to mitigate future dengue transmission patterns.

## 6.2 Introduction

Dengue is a widely prevalent vector-borne illness that causes the highest burden on human populations out of any other arbovirus (Messina, 2019). Currently over half of the world's population is susceptible to dengue, with the disease being endemic to nearly 125 countries (Messina, 2019). Climate, and in particular, temperature and precipitation tend to have the greatest effect on dengue risk (Ebi, 2016; Xu, 2020) and will continue to impact dengue transmission as climate change remains a prevalent issue (Xu, 2020). The associations between climatic factors and dengue transmission have been strongly supported, although other predictors, and specifically environmental and social determinants of health, have not been as well explored.

There are several socio-environmental factors that can indirectly influence dengue transmission and risk by altering mosquito habitats and breeding patterns. Highly population dense areas provide mosquitoes with greater opportunities to feed, and global urbanization and travel patterns have contributed to increased dengue risk by influencing the spread of mosquito vectors between large city centers (Franklinos, 2019). Rapid and uncontrolled urban growth also leads to deficiencies in infrastructure, most notably in terms of supplying sufficient access to water supply and garbage collection (Freitas, 2010). Inadequate water storage and lack of piped water supply around rural households can significantly increase suitable mosquito breeding sites (Schmidt, 2011). Access to piped water has been previously negatively correlated with dengue incidence, likely due to the fact that lack of piped water infrastructure encourages water storage which can attract mosquitoes (Vargas, 2015). Frequency of garbage collection also plays a role in providing more suitable breeding areas. Higher dengue incidence has been correlated to places with garbage collection that occurs less than two times a week, as well as in places with inadequate sanitation practices (Cordeiro, 2011). In other studies, negative correlations between households with direct garbage collection or with a sanitary network connection and risk of dengue were discovered, however no significant associations between household access to piped water supply and dengue incidence were noted (Costa, 1998). In addition, the socio-environmental variables most consistently correlated with dengue risk are lack of basic sanitation, low frequency of garbage collection and practice of water storage (Cordeiro, 2011). Loss of forest cover can also lead to increases in urbanization and population density in some areas (Franklinos, 2019, Husnina, 2019). Deforestation practices can lead to increased local temperatures, enhancing dengue transmission through stimulation of mosquito growth, such as shortened incubation times (Olson, 2010; Husnina, 2019; Karuppusany, 2021). A recent study in Indonesia found that with a 1% increase in forest cover, there was a 9% decrease in dengue risk (Kalbus, 2021). Other studies have found that changes to land cover or land use in an area has the potential to increase dengue risk through alterations of microclimates and increased competition for resources among vector populations (Franklinos, 2019).

Sociodemographic risk factors are also important for further analysis in regards to the role they play in dengue transmission. Several social factors have been identified as having potential to influence dengue incidence, including income levels, access to health care, population density, and education levels (Charette, 2020). Prior studies have also determined associations between increased dengue risk and lower socioeconomic status, lower education, lack of knowledge about dengue illness, high unemployment, household density, and type of housing (Charette, 2020; Whiteman, 2020). In general, poverty rates and dengue risk have not been widely studied as of yet, and previous research has proven inconclusive in determining correlations (Mulligan, 2014; Whiteman, 2020). A systematic literature review by Mulligan et. al further explored poverty-dengue associations but had mixed results (2014). Of the twelve articles reviewed, nine of them showed positive but inconsistent predictor outcomes, while nine showed null outcomes, and five had negative outcomes (Mulligan, 2014). This review noted that both income levels and physical housing conditions more consistently associated with higher dengue risk compared to other poverty indicators, although these correlations are still weak (Mulligan, 2014). Higher dengue prevalence has also been correlated with overcrowding in more deprived Brazilian neighborhoods (Cordeiro, 2011). Household crowding has been previously studied as a potential factor for higher dengue risk, with greater density of people per room being hypothesized to provide more opportunities to mosquitoes for feeding (Heukelbach, 2001). Another study specifically found that the number of both elderly women and children in a single household was indicative of higher dengue risk, due to the fact that these groups tend to spend more time in and around the home and thus are more exposed to mosquito vectors (de Mattos Almeida, 2007). In contrast, there was no association found between household density and dengue risk in a different study completed by Costa and Natal (1998).

The associations between social and environmental determinants of health and dengue risk are still widely disputed. While some studies assert that there are meaningful associations between factors such as income levels, deforestation practices, and access to piped water supply in relation to dengue risk, these correlations are often inconsistent or contradictory across studies which make it difficult to isolate precise associations (Mulligan, 2014). The value in identifying and clarifying associations between important non-climate indicators and dengue risk is immense in that it can help better explain variations in outbreak risk among municipalities. Additionally, the inclusion of relevant social and environmental risk factors in dengue models, especially at the municipality level, has the potential to improve current outbreak predictions by accounting for more localized factors of dengue risk that can typically be missed in larger-scale modeling studies.

We were previously successful in creating boosted regression tree models to predict dengue outbreak potential across municipalities in Pernambuco based solely on temperature and precipitation effects. Based on some of the associations we found in the literature (Mulligan, 2014; Johansen, 2018; da Conceição Araújo, 2020), we believe that we can improve upon our prior models by including previously studied socio-environmental predictors and refine our ability to correctly identify high outbreak risk areas in Pernambuco. Moreover, our work in sub-national predictive modeling at the municipality scale has high potential to provide better insight into localized dengue transmission dynamics and help guide future vector control policy and implementation. Building upon our prior research, we hypothesize that introducing relevant non-climate variables into our methods will allow us not only to improve our current dengue risk predictions but to better understand underlying non-climate associations that have not been as well-studied.

## 6.3 Materials and Methods

As with our previous work, we focused on Pernambuco state as it contains 185 municipalities with near-complete case and environmental data, which gives us a considerable amount of total observations to work with. Dengue case data was available for 129 municipalities between the specified time period and thirty-eight municipalities had enough data to successfully impute values to bring our total study area to 167 municipalities as with our previous models. The time period of interest for this study was defined as the period from January 2010 and December 2016. As with our prior study, imputations were performed to increase the number of suitable municipalities that could be predicted with our BRT models. Any year that had greater than 20% of data missing were excluded, and our final timeframe had to have at least four years of consecutive data. Municipalities that were missing 2 or more consecutive annual observations or were missing observations at the start or end of our timeline were excluded. Municipalities missing only one annual observation that did not occur at the start or end of our time frame were imputed using a simple moving average with the imputeTS 3.2 package. Including the dengue case data that was imputed, there was a total of 14,028 monthly observations from 167 municipalities over the seven year timeframe.

Monthly dengue case data was downloaded directly from DATASUS's SINAN system for each municipality of interest for the specified time period. Available demographic and environmental data was collected from two main sources. Data on land cover and use, deforestation and recovery rates, and percent urbanization were obtained from MapBiomas Brasil (Collection 5), a SEEG initiative that has documented over 30 years of historical land data for all of Brazil from 1985-2019, and deforestation and recovery data from 1988 to 2017. The data was obtained directly from Google Earth Engine, and was subdivided by existing Pernambuco administrative district lines from shapefiles collected through the Brazilian Institute of Geography and Statistics (IBGE). Our population density data was collected from WorldPop, at a spatial scale of one km$^2$, and was available for each year of analysis. We obtained geospatial estimates of the percentages of households that had access to piped water and piped sanitation services for each year of analysis from the Institute for Health Metrics and Evaluation (IHME) via the University of Washington’s LMIC Drinking Water and Sanitation Access Geospatial Estimates 2000-2017 dataset. Other municipality level sociodemographic data for Brazil during our timeframe of interest was not available; the only reliable and publicly available source for this information was the Brazil Census through IBGE, which last occurred in 2010 and is not scheduled to occur again until 2022 due to lack of funds and governmental organization. Due to this, we were unable to include some of the other demographic variables of interest in this analysis, including poverty rates and income levels, education level, and household densities. All data processing was completed using R version 3.6.3 (R Core Team, 2020). Dengue case data were directly downloaded as and imported into R as CSV files. Case data were then formatted in a monthly time-series format in chronological order that spanned the entirety of our time frame of interest. The final included socio-environmental covariates are described in Table 10.

**Table 9 Additional social and environmental covariates included for analysis.**

| Type | Raw Variable Name | New Variable Name | Source | Description |
|---|---|---|---|---|
| **Demographic** | bra_pd_2010_1km; bra_pd_2011_1km; bra_pd_2012_1km; bra_pd_2013_1km; bra_pd_2014_1km; bra_pd_2015_1km; bra_pd_2016_1km | Population density (km$^2$) | WorldPop | Annual population density estimates for Brazil |
| **Demographic** | IHME_WASH_2000_2017_W_PIPED… | Access to piped water | IHME | Annual household access to piped water percentages |
| **Demographic** | IHME_WASH_2000_2017_S_PIPED... | Access to piped sanitation | IHME | Annual household access to sanitation service percentages |
| **Environmental** | Mapbiomas-brazil-collection-50 (2010-2016) | Land cover classification | MapBiomas | Land cover type across Pernambuco |
| **Environmental** | Mapbiomas-brazil-collection-50 (2010-2016) | Primary suppression | MapBiomas | Percent recovery of vegetation |
| **Environmental** | Mapbiomas-brazil-collection-50 (2010-2016) | Secondary suppression | MapBiomas | Primary suppression of vegetation |
| **Environmental** | Mapbiomas-brazil-collection-50 (2010-2016) | Percent recovery | MapBiomas | Secondary suppression of vegetation |
| **Demographic** | Mapbiomas-brazil-collection-50 (2010-2016) | Percent urbanization | MapBiomas | Percent urbanization per municipality |

Our land cover and deforestation data were uploaded into R as rasters from available TIF files downloaded from Google Earth Engine. The variables we chose to include were amount of primary and secondary suppression of vegetation (correlating to deforestation), percent recovery of vegetation, percent urbanization, and land cover classification with three levels – farming, forest, and urban. Primary suppression of vegetation was defined as native vegetation that had been undisturbed by human practices, while secondary vegetation was defined as vegetation that had been regenerated on land that had been cleared by natural or man-made sources. Using high performance computing resources available through the University of Pittsburgh, we extracted numeric class codes from each pixel in our rasters per municipality and per year of data in our timeframe that corresponded to the presence or absence of vegetation suppression and the type (primary or secondary) and/or the amount of vegetation recovery. Values within a municipality were then tabulated per pixel, summed, and multiplied by 100 to get a final percentage value for each specific variable per municipality. Land cover classification data were classified by numeric codes corresponding to certain land cover types from the satellite images collected from Google Earth Engine (MapBiomas). For each municipality of interest, numeric land cover codes were extracted by pixel in the same manner as the other available environmental data, and each municipality was then assigned a final land cover classification based on its highest recurring value (e.g. 3 = land for farming). Percent urbanization was calculated from land cover classification data by summing the number of pixels within a defined municipality that had an 'urban infrastructure' label and dividing it by the total number of pixels within that municipality, then multiplying by 100 to get a final percentage. Population density data for Brazil were uploaded into R as rasters from available TIF files from the WorldPop website for each year of analysis and subsequently pared down to municipality level for Pernambuco state. Population density values were extracted and averaged similarly to environmental predictors to get final values for each included municipality. Our water and sanitation data was downloaded as CSV files and then uploaded into R. These files contained global geospatial estimates in numbers and percentages for household drinking water and sanitation access estimates from 2000-2017. We subset these global files down to Brazil and further down to Pernambuco state. For the purposes of our analyses, we focused on the percentage of households with access to both sources per municipality for each of these predictors. Finally, all predictors for this analysis were transformed into an ordered, single column time-series format and matched to municipalities for each month and year of analysis to create a comprehensive dataset for analysis.

We built upon our previous models by incorporating the previously defined variables and re-running the model that we built with the following slightly updated parameters: lr = 0.02, tc = 3, bag fraction = 0.75. Learning rate was the only parameter changed, and was increased to ensure that we were not stepping down too slowly, and that we were not unnecessarily overfitting our models. These parameters were able to give us at least 1000 trees to fit to each model per run. With the addition of eight socio-environmental predictors, our updated models had sixteen total variables for analysis. We again set a fixed incidence threshold of 100 cases per 100,000 people, and used this to set a binary response corresponding to low risk for municipalities with fewer than 100 cases per 100,000 people, or high risk if a municipality had greater than 100 cases per 100,000 people. A rough 80/20 split of our data gave 12,024 of these observations to train and fit our models while 2,004 of these observations were used as our test set. 10-fold cross validation was applied to our training set and allowed for final model fitting. Once a model was fit to the training data, transmission probabilities were predicted using the test dataset. Probabilities were separated by a threshold value of 0.5, allowing us to classify municipalities as either low risk for a probability of less than 0.5 or high risk for a probability of greater than 0.5.

We then ran univariate and multivariate logistic regression analyses to further explore the associations found among our predictor variables in our BRT models. A significance level of 0.05 was set to determine statistically significant correlations to dengue risk. Since most variables in this analysis had different scales and some had non-normal distributions, all chosen variables were normalized via log transformation prior to logistic regression analysis. Eight univariate regressions for each socio-environmental predictor were run vs. our response variable of low or high risk for dengue. One multivariate regression was completed using all previously studied climate and surveillance variables (minimum and maximum temperatures, precipitation, 2- and 3-month lagged minimum temperatures and precipitation, 1-month lagged cases) along with all socio-environmental variables (primary and secondary suppression, percent recovery, percent urbanization, land cover, population density, access to piped water and piped sanitation) for a total of sixteen initial predictors tested. Variables that were not statistically significant were removed sequentially, beginning with the predictor with the highest non-significant p-value. All other non-significant variables were removed in this way until only significant predictors remained.

## 6.4 Results

Our first run with all sixteen predictors and the previously discussed parameters yielded an initial fitted model of 1500 trees with a training AUC of 0.97 and a testing AUC of 0.938. Initial predictions are shown in Table 11.

**Table 10 Confusion matrix with outbreak predictions for the model including all climate and socio-environmental predictors.**

| | Reference | |
|---|---|---|
| **Prediction** | **low** | **high** |
| **low** | 1625 | 62 |
| **high** | 113 | 204 |

Including both climatic and socio-environmental predictors, our fitted BRT model of 1500 trees (Figure 8) was able to make 1829 correct predictions out of 2,004 with accuracy of 91.3%. This run produced 1625 correct low risk predictions and 204 correct high risk predictions. Incorrect predictions totaled 175 observations for a misclassification rate of 8.7%. Sensitivity, or the ability of our model to correctly predict months when there was lower risk of outbreaks, was 0.935. Specificity, or the ability of our model to correctly predict months of higher outbreak risk, was 0.767. Assessment of the relative influence of all predictors in this run revealed that prior case data was again by far the greatest indicator of increased dengue transmission at a relative influence of 77.8%, followed by minimum and maximum temperatures at 4.5% and 3.8%, respectively (Figure 9). Looking at fitted prediction values against all predictors showed strong association with prior case data, but no other predictors (Appendix Figure 6). The factors with the least influence over predictions for this model were land cover, recovery of vegetation, primary and secondary suppression, and access to piped water (Figure 9).

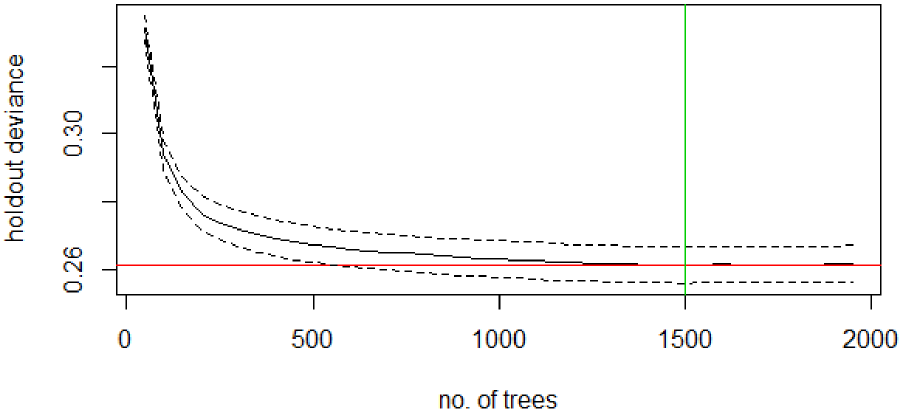


**Figure 8 Graph showing holdout deviance of the climate and socio-environmental predictor model. Holdout deviance, specified by the red line, was minimized at 1500 trees as indicated by the green line. The solid black line indicates mean deviance, and the dotted black lines indicate ± 1 standard deviation.**

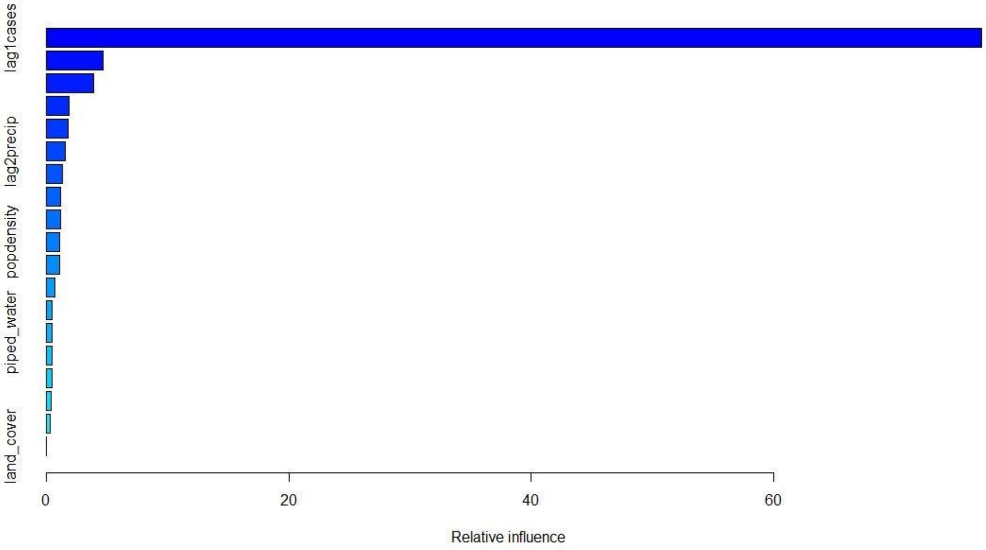


**Figure 9 Relative influence of all climate and socio-environmental predictors in this first model iteration.**

Quantification of associations between socio-environmental predictors and our response of low or high dengue risk were completed through both univariate and multivariate logistic regression analyses. In our univariate regressions, all socio-environmental predictors were highly statistically significant in predicting dengue risk except for household access to piped water, which had a p-value of 0.2343 (Table 12).

**Table 11 Univariate regression results for all included socio-environmental covariates.**

| | **Beta () coefficients** | **p-value** |
|---|---|---|
| **Population density** | 0.5036 | <0.001*** |
| **Percent urbanization** | 0.4612 | <0.001*** |
| **Percent recovery** | -0.3871 | <0.001*** |
| **Primary suppression** | -1.2867 | <0.001*** |
| **Secondary suppression** | -1.1026 | <0.001*** |
| **Access to piped sanitation** | 4.2716 | <0.001*** |
| **Access to piped water** | 1.0011 | 0.2343 |
| **Land cover (forest)** | 0.1925 | 0.0041** |
| **Land cover (urban)** | 4.2621 | <0.001*** |

Significance levels: p<0.5 = *, p<0.01 = **, p<0.001 = ***

With these associations in mind, we first removed all predictors from our model that had less than 1% influence on outbreak potential in our BRT summaries or were non-significant in our regressions and re-ran our model to determine if this would improve overall prediction accuracy. This second run consisted of eleven total predictors -- monthly minimum and maximum temperatures, monthly average precipitation, two and three month lag variables for minimum temperatures and precipitation, 1 month lag variable for cases, population density, access to piped sanitation, and percentage of urban space. Our training AUC from this iteration was 0.97 with a fitted model of 1500 trees (Figure 10) and a testing AUC of 0.939.

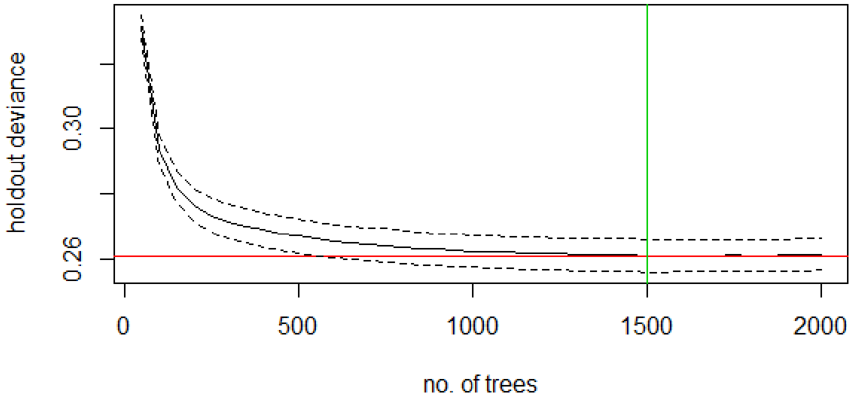


**Figure 10 Graph showing holdout deviance for influential predictors only model. Holdout deviance, specified by the red line, was minimized at 1500 trees as indicated by the green line. The solid black line indicates mean deviance, and the dotted black lines indicate ± 1 standard deviation.**

Predictions from this run are shown in Table 13. There was a total of 1832 correct predictions out of 2,004 observations for an accuracy of 91.4%. This second model iteration produced 1626 correct low risk predictions and 206 high risk predictions based on the eleven predictors that were kept. There were 172 total incorrect predictions for this model run with a misclassification rate of 8.6%. Sensitivity for this run was calculated as 0.936, with specificity of 0.774.

**Table 12 Confusion matrix with outbreak risk predictions for influential predictors only model.**

|  | **Reference** |  |
|---|---|---|
| **Prediction** | **low** | **high** |
| **low** | 1626 | 60 |
| **high** | 112 | 206 |

The predictors with the highest relative influence during this run were again prior lagged cases, minimum and maximum temperatures. The three socio-environmental predictors included in this run (percent urbanization, percentage of household access to piped sanitation, and population density) each only contributed about 1.5% of influence over our predictions, which was only a slight increase from the prior run with about 1.1% influence each (Figure 11). As with the previous model, the only strong association that could be seen between fitted prediction values and predictors was for 1-month lagged cases (Appendix Figure 7).

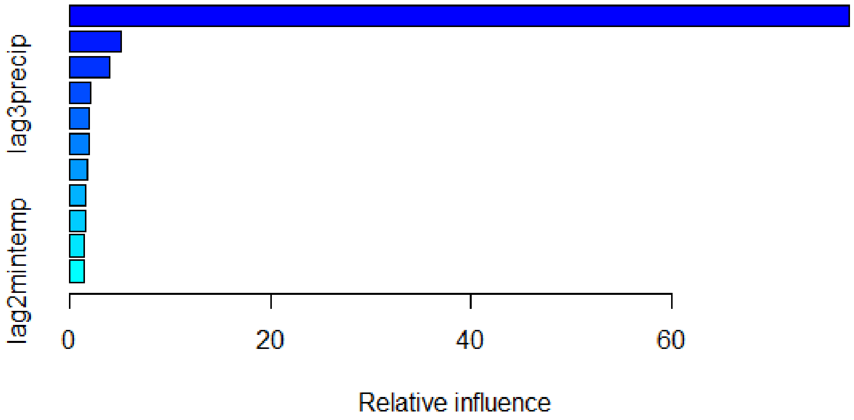


**Figure 11 Relative influence of all predictors included in influential predictor only model iteration.**

After running a multivariate regression that included both climate and socio-environmental predictors, only three socio-environmental variables were statistically significant at our pre-defined significance level: land cover (both forest and urban classifications in reference to farming), household access to piped sanitation, and primary suppression of vegetation (Table 14). The other variables that were statistically significant at $p < 0.05$ and included in this regression model were all related to climate and prior surveillance data: monthly minimum temperatures, precipitation, 1-month lagged cases, and 3-month lagged effects for both minimum temperatures and precipitation.

**Table 13 Multivariate regression results with significant climate and socio-environmental predictors.**

| | **Beta (β) coefficients** | **p-values** |
|---|---|---|
| **Minimum temperatures** | 6.3190 | <0.001*** |
| **Precipitation** | 0.1668 | <0.001*** |
| **3-month lagged minimum temperatures** | -1.6913 | 0.0128* |
| **3-month lagged precipitation** | -0.1628 | <0.001*** |
| **1-month lagged cases** | 0.6616 | <0.001*** |
| **Primary suppression** | 0.0226 | 0.0156* |
| **Access to piped sanitation** | 1.9870 | <0.001*** |
| **Land cover (forest)** | 0.2123 | 0.0169* |
| **Land cover (urban)** | 1.4802 | <0.001*** |

Significance levels: p<0.05 = *, p<0.01 = **, p<0.001 = ***

We then ran a BRT model that included only these statistically significant predictors to determine whether it would improve overall prediction accuracy over previous iterations. After fitting a model of 1,250 trees (Figure 12), this model produced 1,833 correct predictions, with a training AUC of 0.964 and a testing AUC of 0.937. Prediction accuracy was 91.4% with a misclassification rate of 8.5%. In this iteration, there were 1621 correct low risk classifications, and 202 correct high risk classifications (Appendix Table 3). This model was comparative in performance to the model that only included only predictors with greater than 1% influence (Table 15). The model that included only predictors that had greater than 1% influence over predictions from the model using all climate and socio-environmental predictors was best able to identify and predict high risk areas compared to current and previous model iterations (Table 16). Mapped predictions for the models that included all climate and socio-environmental variables and covariates with greater than 1% influence are shown in Figure 12.

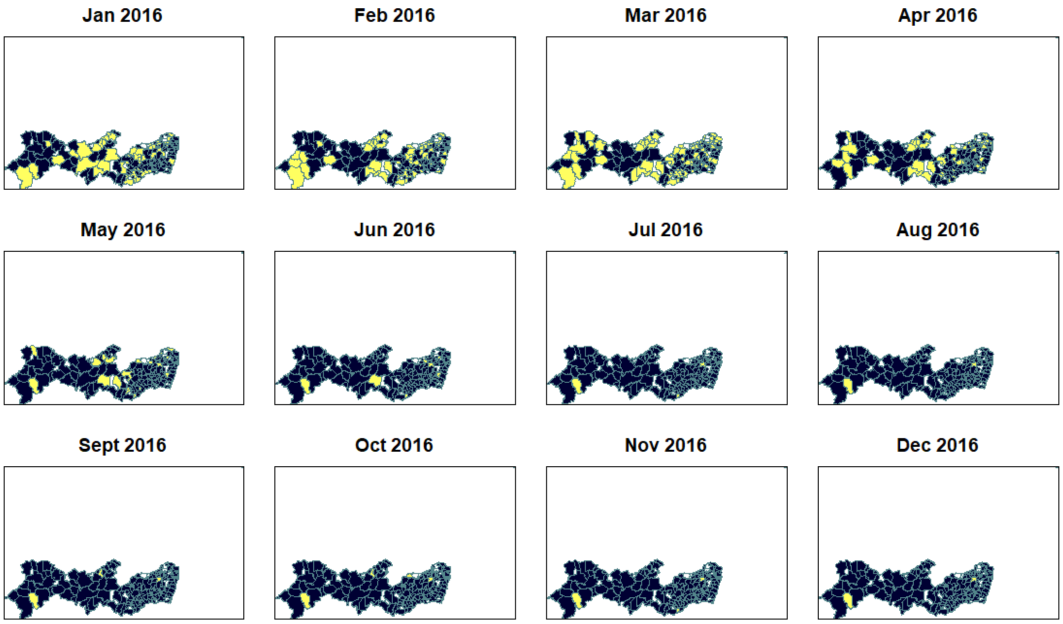
**Table 14 Summary table of AUCs and predictions for all models with socio-environmental covariates.**

| | Training AUC | Cross-validation AUC | Testing AUC | # correct predictions | # incorrect predictions |
|---|---|---|---|---|---|
| **Climate + socio-environmental model** | 0.97 | 0.95 | 0.939 | 1829 | 175 |
| **Climate + socio-environmental >1% influence model** | 0.97 | 0.949 | 0.939 | 1832 | 172 |
| **Significant all predictors model** | 0.964 | 0.946 | 0.937 | 1833 | 171 |

**Table 15 Evaluation metrics for all models with socio-environmental covariates.**

| | Accuracy | Sensitivity | Specificity | FPR | FNR | p-value |
|---|---|---|---|---|---|---|
| **Climate + socio-environmental model** | 0.9127 | 0.9350 | 0.7669 | 0.2331 | 0.0650 | 1.594 e-10 |
| **Predictors with >1% influence models** | 0.9142 | 0.9356 | 0.7740 | 0.226 | 0.0664 | 3.07 e-11 |
| **Significant all predictors model** | 0.9147 | 0.9384 | 0.7594 | 0.2406 | 0.0616 | 2.25 e-11 |

a) Predictions for 2016 from our climate and socio-environmental predictor model.



b) Predictions for 2016 from our influential predictors only model.

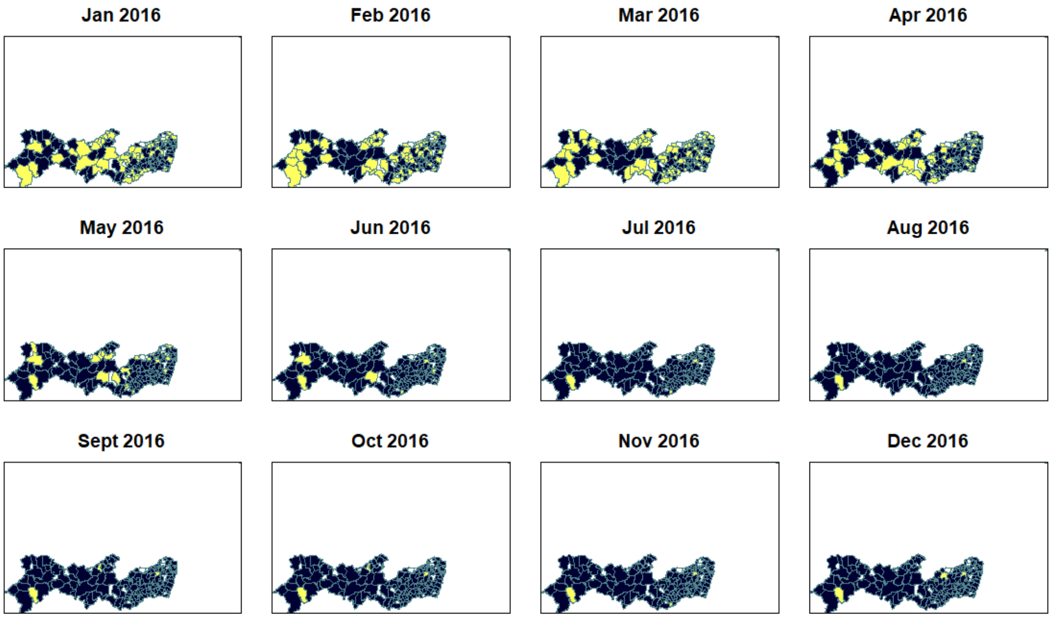


**Figure 12 Mapped monthly predictions for the a) climate and socio-environmental covariate model, and b) the covariates with >1% influence model. Navy regions correspond to low risk predictions, yellow regions correspond to high risk predictions. Regions in white could not be predicted due to lack of data.**

## 6.5 Discussion

The addition of relevant environmental and sociodemographic variables along with existing climate predictors did not significantly affect our predictions. The socio-environmental predictors we introduced showed no improvement on the overall number of correct predictions when compared to our original analyses that focused solely on climate and surveillance data predictors. What is meant by this is that our specificity in identifying high outbreak risk areas increased with the addition of the socio-environmental predictors described here, though our accuracy decreased in terms of overall outbreak risk predictions. In general, our model iterations that included all climate and socio-environmental predictors did not perform better than our original runs that strictly focused on climate and prior case data. In fact, the best overall model run we had previously was one that only considered climate (minimum and maximum temperatures, precipitation), associated 2- and 3- month lag variables for minimum temperatures and precipitation, and 1 month case lags. This iteration gave us 1,842 correct predictions and 162 incorrect predictions (91.9% accuracy) compared to our best run here, which gave 1,833 correct predictions with 171 incorrect predictions for an accuracy of 91.4% accuracy (Table 14). However, the best model iteration in these analyses that included previously studied climate and case predictors, as well as access to piped sanitation, percent urbanization, and population density was better able to identify and predict higher outbreak risk areas better than any of the current or previous models though overall prediction accuracy was not improved (Table 15).

In concordance with the associations found in our last study, cases in the month prior and minimum temperatures have the greatest impact on future predictions in both our BRT models and in regression analyses. In general, it appears that utilizing social and environmental data as predictors are not as essential to accurate municipality level predictions for overall dengue outbreak risk when compared to climate and prior surveillance data, but incorporation of these covariates into models do have value in better identifying and predicting to high outbreak risk areas specifically.

Out of all the new environmental variables that were included in these model runs, access to piped sanitation had the greatest additional influence on our predictions when included in our BRT models, although this influence was very slight compared to our initial predictors. This was consistent across our regression analyses as well -- access to piped sanitation was highly statistically significant to the level of $p < 0.001$ in both univariate and multivariate regressions and was highly correlated to increased dengue risk. Access to piped water had little influence on predictions compared to all other factors considered. Additionally, access to piped water was not statistically significant in either the univariate or multivariate regressions performed and was removed from our final models. Both population density and percent urbanization were statistically significant ($p < 0.001$) in correlation to dengue risk in univariate regressions, but interestingly were not associated with increased dengue risk in our multivariate regression. In looking further, these two variables were very highly correlated (Pearson's coefficient = 0.992; p-value < 0.001), which likely contributed to these results. Both of these predictors were included in our best model run, along with household access to piped sanitation. Primary and secondary suppression of vegetation cover, along with percent recovery of vegetation, all had statistically significant ($p< 0.001$) negative associations with dengue risk in univariate regressions. However, in our multivariate regression, only primary suppression of vegetation was statistically significant ($p< 0.05$), and was slightly positively correlated with dengue risk, though none of these predictors had much influence over final predictions in our BRT models. Besides access to piped water, our land cover associations were somewhat unexpected. We anticipated land cover to have a more consistently strong association with higher dengue incidence in that urban areas would be more susceptible to higher outbreak risk, but in looking at our relative influence summaries in our BRT models, we saw that land cover classifications were the least likely to influence predictions when compared to all other variables. However, land cover classifications of ‘forest’ and ‘urban’ were highly statistically significant at a level of 0.001 in our univariate regressions and specifically, ‘forest’ land cover classification was statistically significant at $p < 0.05$ in our multivariate regression, while an ‘urban’ classification was statistically significant at $p < 0.001$. Both classifications were positively correlated with dengue risk, with urban land cover having a stronger association with a beta coefficient of 1.4802.

Correlations between predictors were determined using Pearson’s correlation tests in R. A complete correlation matrix can be found in Appendix Figure 8. Land cover was not included as it was the only factor variable. In general, most of the climate predictors were weakly positively correlated with each other. We saw the lagged precipitation predictors were more strongly correlated to each other and real-time precipitation; lagged and real-time minimum temperatures also followed this trend, which was expected. Among the other predictors, only percent urbanization and population density were highly correlated to each other, while most of the social and environmental predictors showed weak negative correlations between them.

In relation to prior socio-environmental associations found in the literature (Heukelbach, 2001; Schmidt, 2011; Mulligan, 2014; Franklinos, 2019; Husnina, 2019; da Conceição Araújo, 2020; Gao, 2021), the correlations we have seen here were mixed. Population density and percent of urbanization within municipalities did not have statistically significant associations to increased dengue risk and did not greatly influence final predictions, though high population density and rapid urbanization have been previously theorized to increase dengue transmission (Franklinos, 2019). One interesting prior finding was the identification of an ideal density threshold for dengue risk, ranging from 3,000 to 7,000 people per square kilometer, typically representative of rural and peri-urban areas (Schmidt, 2011). These researchers asserted that study areas with a high population density were less likely to experience severe dengue outbreaks and that areas with low population densities had higher rates of dengue transmission (Schmidt, 2011). This study also theorized that higher urbanicity correlated to increased access to tap water, and that rural areas store greater amounts of water and provide mosquito vectors with and abundance of breeding sites (Schmidt, 2011). In our models, we established that household access to piped water was not essential for accurately predicting dengue risk, and was an ultimately non-significant association to dengue risk. Despite the assumption that access to piped water would decrease dengue risk due to decreased opportunities for storing water, our results generally reflected results found in the literature. Several other studies have found mostly null associations between access to water and increased dengue risk (Mulligan, 2014), which were corroborated with our results here. In contrast, household access to piped sanitation services was significantly correlated to higher dengue risk in both the univariate regression and multivariate regression analyses we performed, but overall had little influence over final predictions compared to our climate predictors. Previous correlations between piped sanitation and dengue risk have been inconsistent throughout the literature, and prior investigations have found either no association between sanitation and dengue risk (Heukelbach, 2001; Mulligan, 2014), or weak positive associations (da Conceição Araújo, 2020). For our models, using access to piped sanitation as a covariate did not improve overall prediction accuracy, but was able to improve prediction accuracy for higher outbreak risk areas when included in a model with all prior climate and case covariates, population density, and percent urbanization. Interestingly, land cover classifications of both ‘forest’ and ‘urban’ held very little influence in making outbreak predictions in our BRT models, but had statistically significant and positive associations in both univariate and multivariate regression analyses. In the literature, certain land cover types are more closely correlated to dengue incidence (Gao, 2021). For example, urban space and urbanization have been widely correlated to increased dengue risk by providing greater amounts of vector habitats and breeding sites (Gao, 2021). Studies have also suggested that urbanization changes microclimates by increasing surface temperatures, which can also attract mosquitoes (Husnina, 2019; Gao, 2021). Though the best model for total correct high risk predictions did not include land cover, it did include percent urbanization, which may be a better indicator of increased outbreak risk. The positive association we found between forest cover and dengue risk in both univariate and multivariate analyses were surprising, however. Most prior research studies have found that high forest cover leads to decreasing dengue incidence (Gao, 2021). We expected a forest cover classification in our models to correlate negatively to dengue risk based on these previous results, and because loss of forest cover leads to increased temperatures as well as promotion of high migration and urbanization rates (Husnina, 2019; Gao, 2021). We found negative correlations between primary suppression of vegetation, secondary suppression of vegetation, and percent recovery of vegetation in our univariate regressions. In our BRT models, none of these variables contributed meaningfully to prediction accuracy. In our multivariate regression, only primary suppression was significantly correlated to dengue risk, but had a slightly positive coefficient. These results were somewhat surprising, as recent research has increasingly linked deforestation to higher dengue risk through increased vector-human contact, increased local temperatures, and changes in the environment that favor vector replication (Husnina, 2019; Gao, 2021). Our results here suggest that only primary suppression of vegetation cover has a positive effect on dengue outbreak risk, though the association we found was weak and further studies are necessary.

Looking back at our raw data, we saw that most incorrect predictions during both model runs occurred within the first half of the year, coinciding with peak dengue season in Brazil. Seasonality was not included as a feature in our models here due to the low influence prior seasonality had on our previous climate models. Several of these predictions were in municipalities that were originally classed as low risk areas but were predicted as high risk areas by our model. These incorrectly predicted areas typically had high (>100 cases per 100,000 people) 1 month case lags, and also generally had higher maximum temperatures and higher rainfall (at time $t = 0$ as well as $t = 2$ and 3 months), which likely contributed to their incorrect prediction status as all three of these variables were highly significant in correlation to dengue risk in prior regression analyses and had strong influence over predictions in our BRT models. There were also several municipalities that were incorrectly predicted to be high risk areas by our BRT model but were actually classed as low risk areas. The main trends we saw within these areas were that they usually had much lower 1-month case lags and also less precipitation, and to a lesser extent somewhat lower temperature ranges. There were few common patterns on prediction status for our sociodemographic and environmental variables. In general, we saw that that if a municipality was highly population dense (> 500km$^2$) and/or very urbanized (> 50%), they were more likely to be at higher risk for outbreaks. This pattern mainly held true for municipality predictions that occurred in the first half of the year, but made no substantial difference in predictions occurring later in the year. However, this could also be due to the fact that our predictions were highly driven by climate and prior case data, which tended to peak during the first half of the year in which we saw a greater number of high risk predictions overall.

The greatest limitation to this work is the sparse availability of reliable, municipality level data. In terms of our sociodemographic variables of interest, we were unable to obtain the most up to date information, severely limiting the scope of our predictions. The most reliable public data source available that contained relevant municipality level sociodemographic information was the Brazilian Population Census, obtained from IBGE. However the most recent census occurred in 2010, and the next census is not scheduled until 2022. Other publicly available data sources did not have the appropriate level of sociodemographic and environmental data that we were looking for, and were not included in our analyses. The lack of relevant municipality level data for our timeframe meant that we were unable to capture and analyze several other potentially important social and environmental factors that may also influence regional dengue transmission, such as method of garbage collection, specifics on household numbers, income and education levels, and more. The land cover, deforestation, water, sanitation and population density data used for this work were all available during our time period of interest and also for all municipalities, however the data used is only available in an annual format. Although we made the assumption that areas remained somewhat static in terms of large scale changes on a monthly basis, we cannot truly discern whether this assumption is correct. Therefore our updated predictions may not be completely accurate based on our additional variables. Further, our land cover predictor was created as a factor variable with three general levels: forest, farming, and urban space. These levels did not account for distinct differences within these levels, however (e.g. an overall ‘forest’ classification for any given municipality could include individual land cover codes for ‘mangrove’, ‘savanna formation’ and ‘forest plantation’) and may misrepresent how dengue risk is truly affected by land cover type. Additionally, our piped water and piped sanitation predictors may not be fully representative of actual supply, as it is common for these services to not be regular although households do report having access to them (Vargas, 2015). This may also contribute to some of our incorrect predictions, as households may reserve large stores of water in between service accessibility, providing suitable breeding habitats for Aedes mosquitoes and increasing risk for disease (Husnina, 2019).

Lastly, our analyses were limited due to the vast computational complexity of some of our predictor datasets. The environmental data files we derived from Google Earth Engine through MapBiomas were so large that they required high performance computing resources provided through the University of Pittsburgh to fully extract all predictor values. These extractions, even with optimized code, took approximately two hours for each environmental predictor included. The extensive time and resources it took to extract and process these predictor values impeded our analyses in that we could not fully run our models until all predictors were extracted and processed.

Dengue control measures in Brazil are currently organized at three levels – federal, state, and municipality (Maciel-de-Freitas, 2014). The federal government provides general guidelines for vector control and purchases and provides states with pesticides and large-scale equipment (Maciel-de-Freitas, 2014). States then acquire smaller equipment and supplies, as well as assist and supervise municipalities in dengue control (Maciel-de-Freitas, 2014). Finally, municipalities themselves carry out vector control actions based on federal recommendations (Maciel-de-Freitas, 2014). However, these distinct levels of vector control in practice tend to be time-consuming and can lead to delayed epidemic response (Maciel-de-Freitas, 2014). Additionally, dengue epidemic response in Brazil is typically deployed after an initial outbreak has occurred, and becomes less effective at controlling transmission (Angelo, 2020). In general, low volume space spraying is the most widely recommended dengue control method used, and some regions also rely on adulticides

(23). However, these practices have led to high insecticide resistance and decreased effectiveness of chemical vector control during outbreaks (Maciel-de-Freitas, 2014).

More recently, community-level dengue interventions have been increasingly explored but not widely implemented. After the rise of the Zika virus in Brazil in 2016, government officials strived to develop more effective methods to control the Ae. aegypti vector, which also spreads chikungunya and yellow fever (do Nascimento, 2019). One of these measures included a national campaign that called for Brazilian citizens to submit their own ideas for vector control and shared experiences of successful individual vector control (do Nascimento, 2019). After completion of the campaign, the Ministry of Education prepared and made freely available a collection of best actions on dengue control via online platform (do Nascimento, 2019), however it is unclear to what extent this tool was used and successfully implemented by their target population. Another recent study conducted a qualitative investigation among 120 adults to better understand individual and community level experiences with Zika and other mosquito-borne illnesses in Brazil (Bancroft, 2022). This study highlighted several general obstacles to localized mosquito prevention strategies, most commonly lack of coordination and cooperation from communities and municipalities alike, and too much emphasis placed on individual participation in vector control (Bancroft, 2022). Significant personal barriers to dengue prevention were noted to be related to affordability of topical repellents and tools that were also perceived to be a sign of socioeconomic status (Bancroft, 2022).

There are several important applications that can be derived from using these predictive models. Most importantly, the development of a comprehensive, sub-national prediction system has immense potential to more quickly assemble and disseminate federal resources to municipalities for faster epidemic response implementation. Moreover, we see that the predictive models built here are highly accurate in determining monthly sub-national dengue risk and have the ability to provide localized outbreak predictions that can be utilized to provide necessary support or resources to anticipated higher risk areas prior to the advent of an outbreak or epidemic. The monthly predictions made here also have the ability to better track real-time changes in dengue risk throughout the course of an ongoing outbreak, allowing for policy makers to redirect support to areas that may urgently need it.

Our intention in developing these comprehensive dengue prediction models is to influence development of future vector control policies to focus more on localized public health interventions. Despite the increased interest in community-based vector control practices in Brazil in recent years, vector control at a community and even individual level are still not well coordinated or planned. The issues discussed here call attention to the need for more community-based and education-based programs in Brazil for dengue prevention and awareness, which have proven successful in other endemic countries. For example, community-based programs involving PSAs and school programs implemented by Puerto Rico's Department of Health made positive impacts on improving dengue knowledge and prevention behaviors among school-aged children and parents (Winch, 2002).

The inclusion of socio-environmental predictors did not make a significant impact on the dengue outbreak predictions shown here. The socio-environmental predictors least correlated with increased dengue incidence in our analyses were household access to piped water, secondary suppression of vegetation, recovery of vegetation, population density, and percent urbanization. In fact, the introduction of these predictors decreased overall prediction accuracy from prior models. We did notice however, that the inclusion of household access to piped sanitation, land cover, and primary suppression of vegetation as predictors increased the correct identification of high outbreak risk regions. Few of the additional predictors had significant correlations with dengue risk when compared to all prior covariates, though they did show significant associations through univariate regression analysis. We can conclude from this investigation that at a sub-national level, certain social and environmental determinants of health do not play a major role in predicting dengue risk. However, due to the limited scope of available sub-national data as well as inconsistent findings in previous literature, there is much more work to be done in terms of elucidating important socio-environmental correlations with dengue risk at a municipality scale. In general, there is great value in further developing sub-national predictive models such as the ones described here to identify and predict areas of higher outbreak risk. Furthermore, these results can provide better insight into local transmission dynamics and in turn can be utilized to develop future vector control interventions that shift focus to community-based and individual practices for improved dengue prevention.

# 7.0 Examination of Complete Findings and Implications for Policy and Practice

One major end goal in completing this dissertation research is to leverage these findings to highlight the need for new and improved vector control strategies. Current vector control methods in Brazil in particular are insufficient in controlling dengue and other Ae. aegypti transmitted diseases. We hope to apply these findings to not only improve current overarching vector control measures, but also to push for methods that are more localized to specific high risk areas as well as increase community engagement measures and awareness for dengue in preparation for future outbreaks.

Translation of these results into practice is the main objective in our studies. We propose that utilization of a Consolidated Framework for Implementation Research (CFIR) is important for dissemination of this work to public health practice through assessment of both barriers and facilitators to successful implementation of innovative vector control strategies (CFIR, 2022). The CFIR consists of 39 constructs organized into five different domains: intervention characteristics, inner setting, outer setting, characteristics of individuals, and process (CFIR, 2022). For Brazil in particular, the challenges to implementation are extensive. First, this aim was not completely met, since we were unable to fully make predictions across Brazil for all municipalities and thus our implementation strategies should focus currently on Pernambuco alone. Lack of municipality organization and poor enforcement of existing vector control methods is consistently a major barrier to implementation and practice in Brazil. Participants surveyed in recent interviews revealed that inconsistent household surveillance, mixed campaign messages, and chronic government underinvestment in vector control are primary factors to ineffective prevention strategies (Bancroft, 2022). Participants also reported lower individual protective behaviors or beliefs if other community members were not practicing individual prevention strategies (Bancroft, 2022). It is clear that vector control practices and regular vector maintenance need to be improved in order to successfully decrease dengue risk through community-based programs. Participatory models in prior research have been proven to be most effective at improving self-efficacy of novel repellants; these studies also tend to promote a sense of shared community responsibility (Bancroft, 2022). Based on these specific barriers to effective vector control, it is important to focus our efforts on constructs that center around the following: evaluation of strong leadership engagement and clear goals for implementation (inner setting), assessment of advantages for implementation of new vector control strategies versus current ones (intervention characteristics), potential for development of better-organized external policies and incentives (outer setting), likelihood of engagement of target audiences for participation in control measures (process), improvement in overall knowledge of dengue and beliefs about dengue prevention (individual characteristics), and increased access to dengue prevention information and resources (inner setting). Thorough evaluation of these constructs should give us the ability to apply our findings through implementation of targeted vector control measures for more efficient dengue prevention.

## 7.1 Gaps in Current Vector Control Methods

Brazil is consistently plagued with dengue epidemics, and while there are several control measures employed to contain and slow the spread of dengue, they ultimately prove to be largely unsuccessful (Vargas, 2015). Some of these measures include municipal sanitation surveillance, entomological survey indicators, and monitoring of Aedes resistance. In addition, new biological control measures, such as the release of genetically modified and Wolbachia-infected mosquitoes have been introduced in recent years as part of the World Mosquito Program, although with conflicting outcomes (O'Neill, 2018; Wermelinger, 2020). In spite of positive results in smaller trials as well as larger field sites (O'Neill, 2018), these results do not necessarily translate successfully to larger Brazilian cities such as Rio de Janeiro (Wermelinger, 2020).

There also tends to be a delay in surveillance activation despite Brazil's comprehensive and relatively successful dengue surveillance system (Angelo, 2020). Brazil's current dengue control infrastructure is partitioned into 3 levels: federal, state and municipal. The federal government is the most important, as they provide guidelines for vector control, resource allocation and acquisition of insecticides and necessary equipment (Maciel-de-Freitas, 2014). The states then provide supervision and assistance across municipalities, which then manage vector control interventions and ensure compliance with federal regulations. These measures in practice however, tend to reduce the overall effectiveness of vector control, as decision-making in regards to methods can be time-consuming and excessively bureaucratic (Maciel-de-Freitas, 2014). Typical vector control in Brazil centers around larval index surveillance, which monitors A. aegypti density throughout districts (Maciel-de-Freitas, 2014). Higher density districts then receive reinforced mosquito control actions. Low volume space spraying is typically recommended during outbreaks, although some municipalities use adulticides as their main source of vector control (Maciel-de-Freitas, 2014). This is problematic as it leads to high insecticide resistance and decreases chemical control effectiveness during outbreaks (Maciel-de-Freitas, 2014). Underreporting of dengue cases also plays a large role in delays of response action. Nearly 90% of patients reach the healthcare system after dengue illness, during the recovery phase (Angelo, 2020). In addition, long wait times also tend to deter patients from coming in for evaluation even when they are sick (Angelo, 2020).

Overburden of laboratories can lead to delays in test results, and can also contribute to the incompleteness of patient forms as providers forget to record results after receipt (Angelo, 2020). Moreover, implementation of response activities typically occur after an epidemic has already begun and can impede the success of an effective vector control response (Angelo, 2020). Prior literature also suggests that surveillance is subpar due to lack of investment, and more specifically due to high workloads during epidemic periods and lack of appropriate resources (Angelo, 2020).

Currently, Brazilian measures of both individual and community level practices for vector control have not been well established or reinforced. Other endemic regions have found moderate success with implementation of both community-based programs and increased individual awareness concerning vector control. The CDC in collaboration with the Puerto Rico Department of Health found that community-based programs involving PSAs, school programs, and a Children's Museum exhibit had a positive impact on improving knowledge and behaviors in parents and children relating to dengue prevention (Winch, 2002). In Hong Kong, individual protective behaviors against dengue, such as using mosquito repellent, wearing long sleeve clothing, and removing stagnant water sources increased after a local dengue case announcement in August 2018 (Chan, 2021). These shortages in both knowledge and resources in Brazil highlight the need for not only improvement in response implementation but also improved education and awareness of dengue throughout local infrastructure.

### 7.2 Potential for Upcoming Interventions

Community-based engagement programs and increased education on dengue prevention in general is a necessary next step. Efforts within the past several years following the Zika epidemic in 2016 have shown that post-epidemic engagement among the population, especially among disempowered groups, such as pregnant women and people of lower socioeconomic status, is limited (Bancroft, 2022). Most participants surveyed in this work reported community-level vector control engagement as reporting mosquito breeding sites to public health authorities, as well as observing and encouraging others to make behavior changes to decrease risk (Bancroft, 2022). Emphasis on individual protection behaviors is also important to more readily promote, as this study also revealed that there was low participant awareness of novel protection items, such insect repellant clothing (Bancroft, 2022). Interest in such items was also affected by an individual's community standing, which was linked to negative perceptions of mosquito-borne illnesses and socioeconomic status. Several participants, mainly women, also reported feeling unprepared for disease outbreaks with the practical knowledge obtained from public health campaigns, and felt that their messages for vector control were overly technical or not sufficient for household implementation (Bancroft, 2022). With the results from this dissertation research, we hope to push public health officials to place greater emphasis on the importance of both community- and education-based programs for improved prevention of dengue in Brazil.

## 8.0 Main Summary and Discussion

The best overall BRT model for monthly prediction of dengue risk in 2016 was able to correctly predict 1,842 out of 2,004 observations with a misclassification rate of 8.1%. This model included monthly minimum and maximum temperatures, monthly average precipitation, 2- and 3-month lagged values for both minimum temperature and precipitation, and 1-month lagged prior case data. Both training and testing AUCs for this model were high, at 0.963 and 0.939, respectively. The addition of prior case data was able to greatly improve the specificity of our model in correctly predicting high risk areas and was consistently one of the most significant predictors of dengue risk included in our models. In contrast, the worst overall model including only climate and lagged climate predictors was only able to correctly predict 1,724 out of 2,004 total observations, and had a misclassification rate of 13.97%. Training AUC for this model was high at 0.916, however testing AUC was low at 0.598. All model AUCs and prediction summaries can be found in Table 14. However, the model that was best able to correctly predict high risk areas for dengue risk was neither of these models – it was the model that included only covariates with greater than 1% influence on predictions, which included all of the climate predictors studied, prior surveillance data, household access to piped sanitation, population density and percent urbanization. Evaluation metrics for all models are summarized in Table 15, and show that this model had the highest specificity in terms of correctly predicting an area as high risk when it was in fact classified as higher risk.

We found that lagged climate effects for both minimum temperatures and precipitation, specifically at 3 months, had the lowest correlations with increased dengue risk compared to other predictors in multivariate regression analyses. Ultimately, the two best covariates for predicting dengue risk in our models were prior surveillance data and monthly minimum temperatures.

When investigating correlations between socio-environmental variables and dengue risk, we found that there were few factors that significantly contributed to increased dengue risk compared to climate and prior surveillance data. In assessing variable importance within our BRT models, we found that none of the socio-environmental variables included in our models had substantial influence over dengue risk predictions. Five of the eight additional socio-environmental predictors (primary and secondary suppression, piped water, land cover and recovery of vegetation) included in our models had less than one percent influence over outbreak predictions. No socio-environmental variable had over two percent influence over outbreak predictions when compared to all other climate and surveillance variables. Multivariate regression analysis with both climatic and socio-environmental variables revealed that three of the socio-environmental predictors included were statistically significant in predicting dengue risk: land cover (forest and urban classifications), access to piped sanitation, and primary suppression of vegetation, which was surprising given our prior BRT results. Access to piped sanitation was the only socio-environmental predictor included that was important for both BRT and logistic regression analyses.

One unexpected outcome of this work was that household access to piped water was not significant for predicting dengue risk in either our boosted regression tree models or our univariate and multivariate regression analyses. We expected household access to piped water to have a much greater influence on dengue risk, primarily because restricted or no access to piped water means that it is much more likely for individuals to store water around the household. In accordance with prior literature (Vargas, 2015), we expected that lower household access to piped water per municipality would correlate to higher dengue risk, although we could not definitively make this conclusion based on our results. In potential explanation of these results, it has been previously established that access to piped water does not necessarily reflect continuous supply, and that several households that have access to piped water still store water in containers that attract mosquito vectors (Arunachalam, 2010; Vargas, 2015).

There are several ways that this dissertation work can be valuable for future vector control methods and policy development. Most importantly, the use of a municipality-level dengue prediction system versus a larger, more general state-wide or national system, has the ability to better mobilize federal resources for faster epidemic response. Knowing exactly where resources should be concentrated prior to dengue season would be beneficial for future vector control response, allowing decision makers to concentrate resources in areas of higher risk prior to peak seasonal transmission to preemptively reduce outbreak and/or epidemic potential. Monthly predictions are also more useful in that they can better track real-time changes in dengue risk throughout the course of an ongoing epidemic or outbreak, and allow for government officials to quickly take action and redirect supplies or provide additional support to areas that may urgently need it. Furthermore, municipality level predictions provide essential information on general outbreak risk and important risk factors prior to dengue season. These results can provide public health authorities with reliable data to make better-informed decisions concerning vector control in Brazil, and in particular improve upon current municipal service coordination efforts, which have been reported as a priority need (Bancroft, 2022).

## 8.1 Conclusions

Our initial BRT model iterations showed that both climate and prior surveillance data are highly important factors to consider when predicting dengue risk. The best BRT model created was able to correctly predict 1,842 from a total of 2,004 monthly observations for 167 municipalities in 2016, and had a misclassification rate of 8.1%. These results were corroborated through our univariate regression analyses, which revealed that prior dengue case data, monthly minimum and maximum temperatures, and monthly average precipitation are all highly significant for predicting dengue risk at a significance level of $p < 0.001$. Lagged climate effects, specifically 2- and 3-month lagged minimum temperatures as well as 2-month lagged precipitation, were also highly significant ($p < 0.001$) in correlation with higher dengue risk in univariate regression analyses. 3-month lagged precipitation was the only non-significant factor ($p = 0.429$) detected in our univariate regressions. Further analysis of these predictors through multivariate regression showed that precipitation and 2-month lagged precipitation were not correlated to higher dengue risk when compared to the other climatic variables studied. Monthly minimum and maximum temperatures, 2-month lagged minimum temperature, and 1 month lagged cases were all statistically significant in predicting higher dengue risk at a significance level of $p < 0.05$. Surprisingly, 3-month lagged precipitation and minimum temperatures were also statistically significant in this regression, but were negatively correlated with dengue risk. 1 month lagged cases had the greatest association with dengue risk in this model, while 3 month lagged precipitation had the lowest correlation with dengue risk. Using only the predictors that were significant in regards to dengue risk from our multivariate regression, we further explored these associations by building an additional BRT model to include only these variables. The significant predictors only BRT model performed nearly as well as the previous BRT model built that included prior case data, climate and lagged climate effects, although the total number of correct dengue outbreak risk predictions was slightly lower in this model at 1,837 observations out of 2,004 with a misclassification rate of 8.3%.

Addition of both environmental and social determinants of health had little to no influence in making dengue risk predictions when included in further BRT model iterations. The BRT model created that included all predictors studied was only able to predict 1,830 observations correctly and had a misclassification rate of 8.7%. In looking at our univariate regressions for these predictors, percent urbanization, household access to piped sanitation, population density, and land cover all had statistically significant correlations with higher dengue risk. Primary and secondary suppression, and percent recovery of vegetation were negatively correlated with dengue risk, which was somewhat unexpected. The only variable not correlated with increased dengue risk in these univariate was household access to piped water supply, which was consistent across our BRT models and regressions. Multivariate regression of all variables studied revealed that minimum temperatures, precipitation, prior dengue case data, land cover (urban and forest classifications), primary suppression of vegetation, and household access to piped sanitation services had positive associations with dengue risk while lagged 3-month minimum temperatures and precipitation had negative associations with dengue risk. Compared to the other predictors included in this multivariate regression, maximum temperatures, 2- month lagged minimum temperatures and precipitation, secondary suppression of vegetation, percent recovery of vegetation, percent urbanization, population density and household access to piped sanitation had non-significant correlations to dengue risk and were not included in the final regression model. After running an additional BRT model that included only the significant climate and socio-environmental variables included in this multivariate regression analysis, we found that this iteration was able to correctly predict 1,831 observations out of a total of 2,004. Misclassification rate for this model was calculated as 8.6%.

Based on the results from this dissertation research, we can conclude that the most important indicators to include in monthly dengue outbreak risk prediction modeling are climatic factors, and more specifically minimum temperatures and precipitation, along with lagged effects up to 2 months. Social and environmental factors included in our BRT models did not have major influence over predictions, though several were significantly correlated with dengue risk through univariate regression analyses. We also found that including only significantly correlated covariates from multivariate regressions did not improve predictions, and in fact the best BRT model we built included only climate, lagged climate and prior case data covariates despite statistical significance in our multivariate regressions. This analysis provides a strong foundation for future municipality-level predictive models, and lays the framework for utilizing these models to provide a better representation of more localized dengue risk across endemic regions as compared to previous studies.

## 9.0 General Limitations

We noted several limitations to the scope of this dissertation work that restricted our final results and conclusions. First, collection of data was largely subpar for a variety of reasons. In regards to the dengue case data collected through the Brazilian Ministry of Health DATASUS system, there were several municipalities that were missing case values, and there was no consistent pattern in missing data across years of analysis that we could discern. The proportion of missing case data was as high as 45 percent in certain years, which additionally constrained our overall study time period to only seven years. The amount of missing data also restricted the total number of available municipalities that we were able to predict to, and as a result there were eighteen municipalities that were not included in our final models.

The amount of reliable, municipality level data that was available for study was also a major limitation to this dissertation research. Climate data was easily accessible as well as publicly available, but sociodemographic and environmental data was more difficult to collect. All environmental (primary and secondary suppression, percent recovery, land cover, and percent urbanization) and sociodemographic predictors (population density, household access to piped sanitation and piped water) used were available but only in an annual time format versus a monthly time format. The assumption we made was that changes in any of these variables was too slight to make a substantial difference in prediction accuracy and remained generally static over the course of a year. Thus, we used annual values for these specific predictors for each month over the course of a specific year (i.e. the final predictor value for 2010 was used for all months of 2010). Though these values did not change much between years, we cannot be certain that this assumption is entirely correct. In addition, several social determinants that we were interested in studying did not have any municipality level data available. The most reliable source of municipality level sociodemographic information, including income and education levels, was obtainable through IBGE via the Brazilian Population Census. These data, however, were not used for our analyses due to the fact that census data are only collected every ten years, and the last census occurred in 2010 with the next census not scheduled until later this year (IBGE, 2022). The lack of accessible sociodemographic data for municipality level analysis meant that we were unable to fully include these variables in our final models, and were unable to study and make definitive conclusions about potential associations with dengue risk.

Data complexity issues also restricted the ultimate scope of our analyses. Within the expected timeframe of this dissertation work, we planned to predict to Pernambuco state as a starting point for our studies prior to expanding these models to predict outbreak risk for all of Brazil. However, we were unable to do so in the amount of time remaining for completion of this work. This issue was mainly caused by the introduction of the specified environmental predictors into our model, as the extractions performed for these covariates were incredibly time-consuming at over two hours each even with optimization of our code, and so computationally complex that they required high performance computing resources. The sheer amount of time and resources needed to fully extract and analyze these predictors severely limited our final investigations to include only Pernambuco state. Moreover, the vast amount of municipalities for study (5,570 currently) made it difficult to attempt a complete analysis of Brazil within the scope of this dissertation research.

The three major climatic predictors we focused on with this dissertation work were monthly minimum temperatures, monthly maximum temperatures, and monthly average precipitation. Though temperature and precipitation are the most important drivers of dengue transmission, there are other climatic factors associated with dengue risk that were not considered for our studies for a variety of reasons. Monthly data concerning potential predictors such as elevation and humidity were not collected, though both have been previously associated with increased dengue risk (Lozano-Fuentes, 2012; Campbell, 2013). Elevations of 2,000 meters and below allow for higher vector proliferation (Lozano-Fuentes, 2012), while high humidity can amplify transmission rates within an ideal temperature range (Campbell, 2013). We did have access to elevation data through the WorldClim website, however the data were limited to the time period between 1970 and 2000, which was not useful for the analyses completed here and was not included. Moreover, the accessibility of other climatic data such as humidity was also an issue. Monthly humidity data for the time period specified was not readily available or in a format that was easily transformed, and was also not included in this dissertation analysis. In spite of the exclusion of these variables, the predictions we observed were overall highly accurate, though with further studies, it may be useful to introduce other relevant climatic variables into our prediction models.

Imputations were completed for thirty-eight municipalities that were only missing one year of monthly case data values within our main dataset. These were completed using a simple moving average calculation, with the assumption that missing monthly case values would be similar in climate profiles to the seasons prior and the seasons after the missing value occurred. However, this method does not account for seasonal variability or external factors such as current serotype co-circulation, which have been linked to both severity of dengue outbreaks and increased local transmission (Gubler, 1998; Fried, 2010). While the imputations allowed us to predict monthly outbreak risk for a greater overall number of municipalities, we cannot be certain that we were able to completely capture the true extent of dengue transmission dynamics in these areas, which may affect the predictions made.

Lastly, a standard baseline outbreak definition for dengue is difficult to establish and may contribute to incorrect classifications. A specific dengue outbreak definition is especially hard to distinguish due to complex transmission dynamics as well as high case reporting variability (Brady, 2015). We chose a fixed threshold incidence of 100 cases per 100,000 people to classify our monthly predictions based on previous literature guidelines (Brady, 2015) for overall consistency in making predictions, however this may not reflect true dengue transmission dynamics across municipalities with heterogeneous distributions. We also acknowledge that by using a fixed incidence threshold we could not account for variations in seasonality or even variations in current serotype distribution, which also play a role in patterns of dengue incidence.

# 10.0 Recommendations for Future Research

The conclusions of this work have great potential to drive future predictive models for sub-national level analyses. For the first part of our analysis, this dissertation work focused on the effects of temperature, precipitation, and associated lagged climate and prior case data as suitable predictors for municipality level dengue outbreak predictions. Though these associations were strong, we did not study any other climatic predictors or include them in the models we showed here due to lack of publicly available data and time constraints. Further studies on climate covariates such as humidity and elevation could be useful to include in future predictive modeling studies. To add to these recommendations, future studies should also incorporate spatial and seasonal variability as explicit predictors for dengue risk, which can assist our models in better identification of higher risk areas and also periods of higher transmission risk.

Throughout the course of this dissertation research, we also found that we were generally limited in the types of social and environmental predictors we could study based on the availability of usable data sources, as we relied solely on secondary data collection. Global climate and environmental data within our timeframe was widely available and easily accessible on several different data platforms, but sub-national sociodemographic information was much more difficult to procure and thus restricted our opportunity to adequate study these associations. The utilization of primary sources, or data from local state health departments would be important for further modeling studies in that social risk factors that were not studied here can be obtained and further analyzed using our existing model framework.

In addition, there are other sociodemographic risk factors that we did not consider in our models that require further research. For example, both race and age have been linked with dengue risk but were not included in our analyses. There is some evidence that people of African descent are more resistant to dengue infection (Sierra, 2007), while gene variants in people of Southeastern and Northeastern Asian descent have been identified that render these populations more susceptible to dengue infection (Oliviera, 2018). In the Americas specifically, people of all ages are equally susceptible to dengue, though the age group at highest risk for severe dengue is children under 10 years of age (Tantawichien, 2012; Burattini, 2016). Stratified dengue case data by both age and ethnicity, as well as by dengue classification and degree of illness, are available through the DATASUS SINAN system. Between 2000 and 2014, dengue complications occurred more frequently among children less than 10 years old and those who were infected with DENV-2 virus in Brazil (Burattini, 2016). Risk of hospitalization was also increased for children between 6 and 10 years of age (Burattini, 2016). Though it was outside of the scope of our research, it would be beneficial to see future modeling studies stratified by age or ethnicity to better observe and quantify differences in dengue risk between these groups, and additionally to better quantify and assess the risk for dengue hemorrhagic fever manifestations, which was not considered for the purposes of our research.

Skew of the data was also a concern. For our training and testing sets for BRT analysis, we had about 80% of the data for training and 20% of the data for testing. Typically a 70/30 split of the data is used for machine learning models, however, we decided to include 2015 as a training year in our models because it was a high outbreak year compared to earlier years and we did not want to skew our data to mainly predict only low risk regions. This problem ended up occurring anyway mainly due to the large disparity in the amount of low risk versus high risk classified areas within our data. Though our current models provided high overall prediction accuracy, it may be useful to further explore the possibility of upsampling as a way to better balance this large data skew and get better results using SMOTE, or Synthetic Minority Oversampling Technique. SMOTE is an oversampling technique where synthetic observations are created from the existing minority data (Chawla, 2002; Maheshwari, 2019). This oversampled data can then be incorporated into the original dataset and models can be applied to make predictions (Chawla, 2002; Maheshwari, 2019). Though this was not something considered during the course of this dissertation work, upsampling our classified high risk data through SMOTE is a future direction that should be further researched; overall there were very little high risk observations within our dataset and SMOTE may provide our models with a more balanced dataset and may offer improvement over our current predictions.

Due to issues with computational complexity in data extractions and general scope of this dissertation work, we were unable to predict outbreak risk for the rest of Brazil as previously planned. Though we did not complete this final part of our analysis as described in our last aim, we have provided an essential framework here for other investigators to expand upon this dissertation research and complete these predictions to provide a more comprehensive representation of overall dengue risk in Brazil by municipality. The code that was created during this dissertation work has been made reproducible and available on Github, and the datasets utilized for our predictor and response variables are all publicly available.

# Appendix A Supplementary Information

## Appendix A.1.1 Data Availability and Code

The code created during this dissertation work was done exclusively using R version 3.6.3. All code for data processing and extraction, model building, model implementation, model evaluation, further analysis, and additional figures and tables is publicly available on Github, under the repository ‘BRT Dengue Modeling’. Code is annotated accordingly and is reproducible to the best of our knowledge. The link to this repository is found here: https://github.com/matlackm/BRT-Dengue-Modeling.

All datasets used for this dissertation research were publicly available. A final comprehensive dataset including all relevant predictors as well as our binary response variable was created from these data and is also available on Github. Legend codes corresponding to specific classes of land cover can be found on the MapBiomas website, under 'Maps and Data', Collection 5. Specific suppression and recovery class codes are also available on the MapBiomas website at https://mapbiomas.org/en. Example raster containing land cover classifications can be found in Appendix Figure 9.

## Appendix A.1.2 Calculations

Calculations were performed for incidence per 100,000 people, and also for percent increase in specificity and percent decrease in sensitivity among our initial climate models. Example calculations for each of these values is shown below.

**Equation 1.** Example calculation of incidence rate per 100,000 people using monthly case data and current population.

Total number of monthly cases/total population x 100,000 = monthly incidence rate

55/10,000 x 100,000 = 550 cases per 100,000 people

**Equation 2.** Calculation of percent increase in specificity between models for correctly predicting to high risk areas.

Specificity for initial model: 0.0602

Specificity for final model: 0.7481

Difference (increase between numbers)/original number x 100 = percent increase

0.6879/0.0602 x 100 = 1142.69

**Equation 3.** Calculation of percent loss in sensitivity between models for correctly predicting to low risk areas.

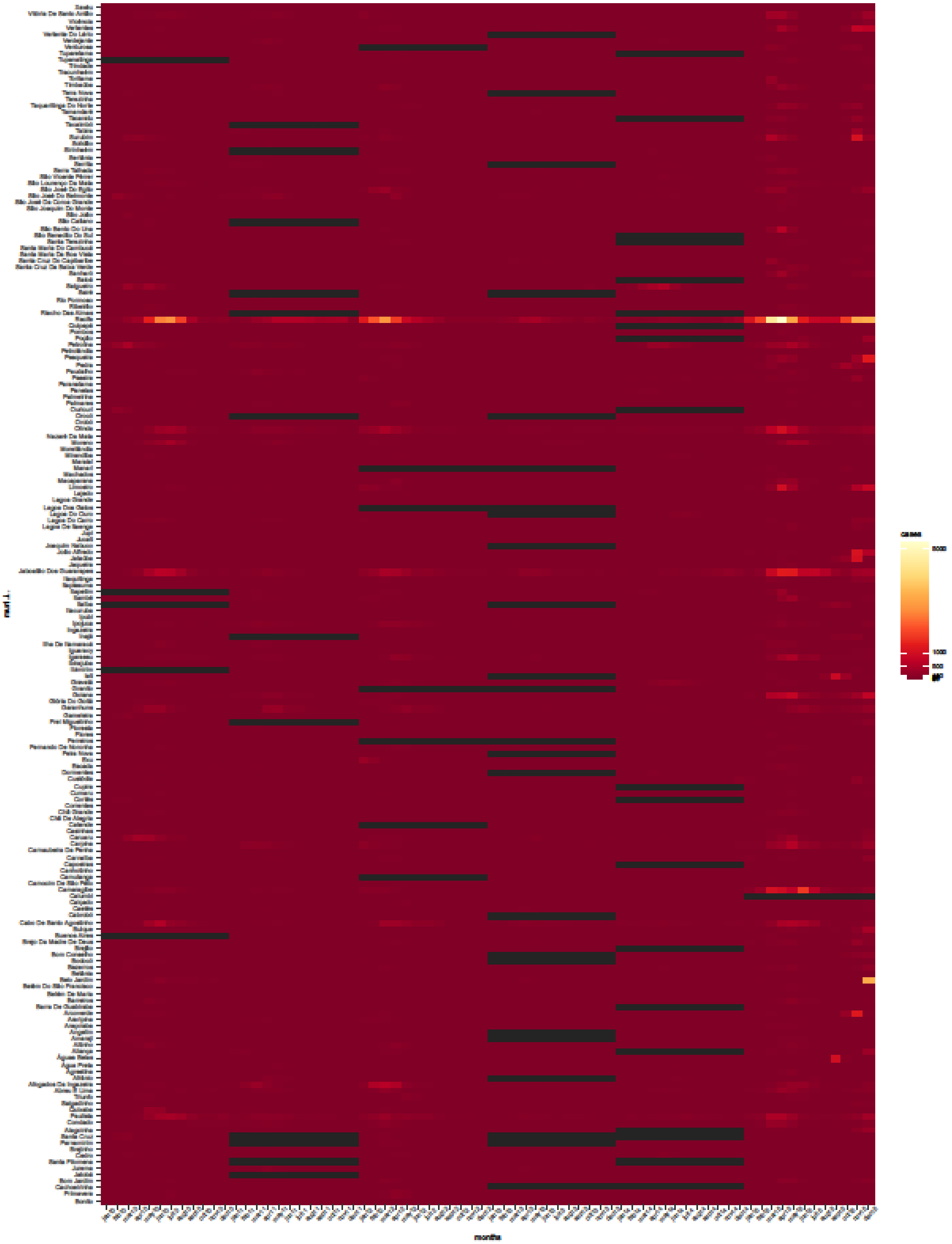
Sensitivity for initial model: 0.9983

Sensitivity for final model: 0.9453

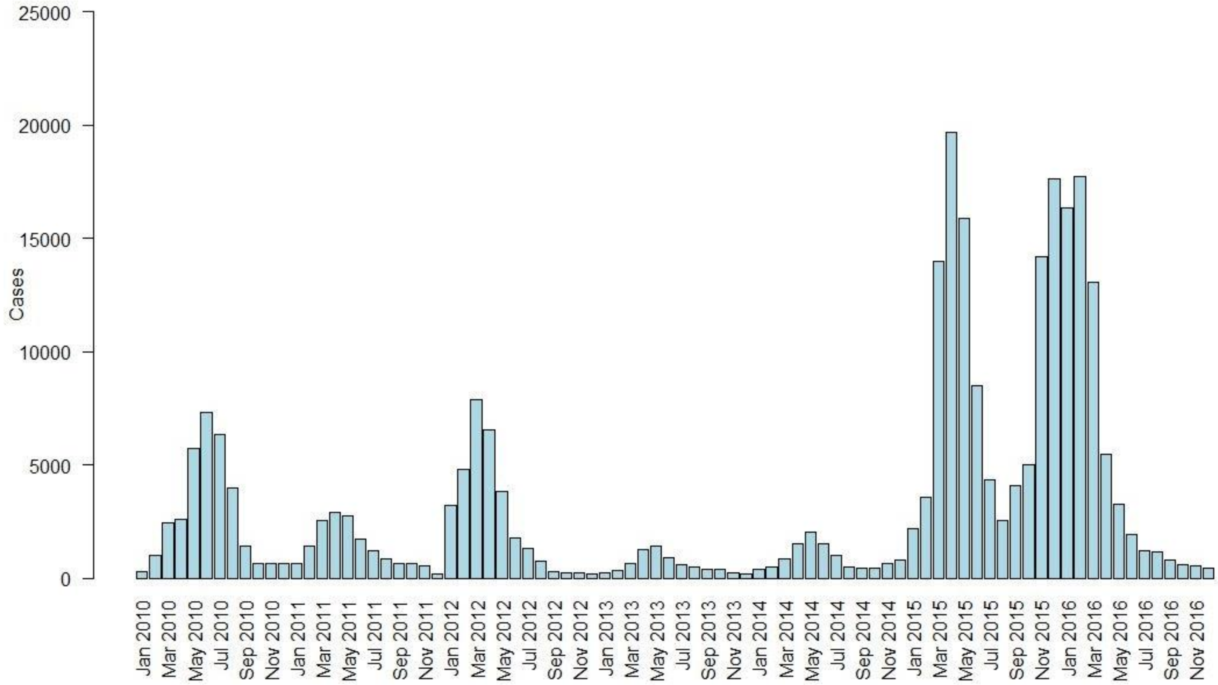
Difference (decrease between numbers)/original number x 100 = percent loss

0.053/0.9983 x 100 = 5.309

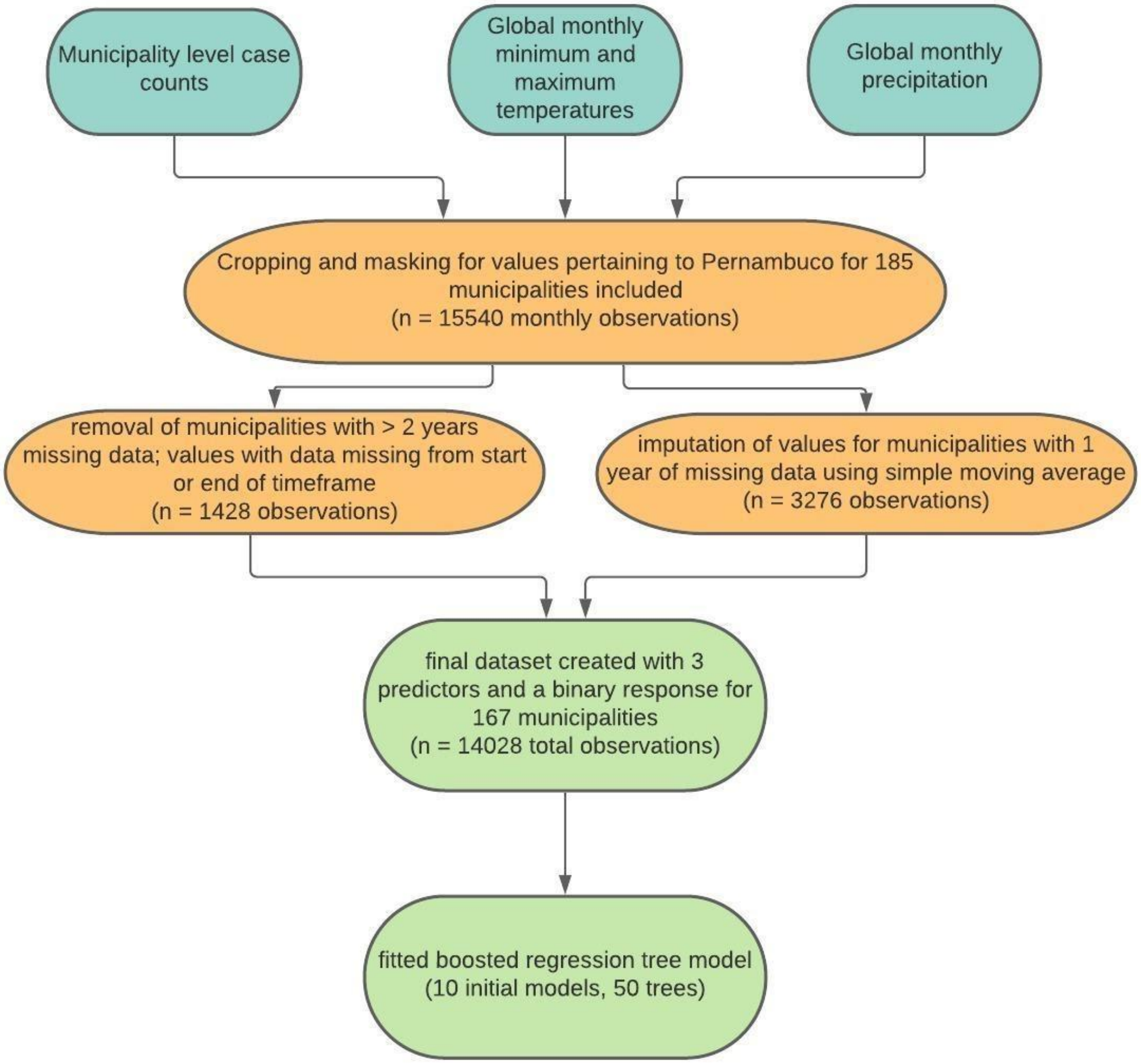
# Appendix B Supplementary Figures



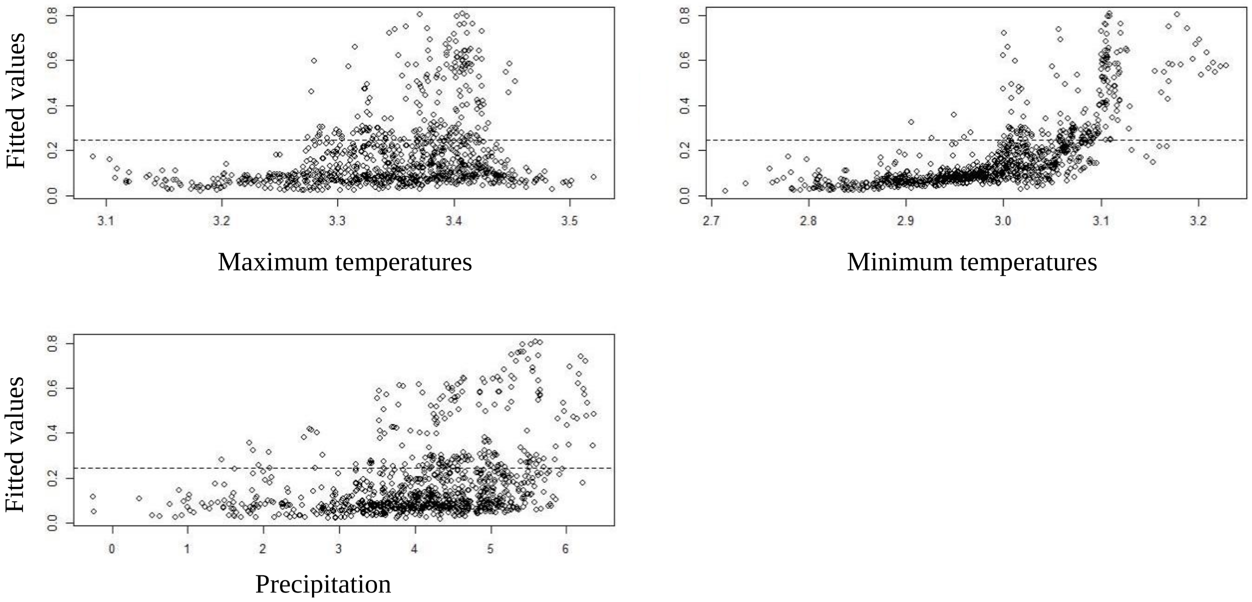
**Appendix Figure 1 Comprehensive heatmap of all dengue cases for all municipalities in Pernambuco from January 2010 to December 2016. Missing data is shown in black.**



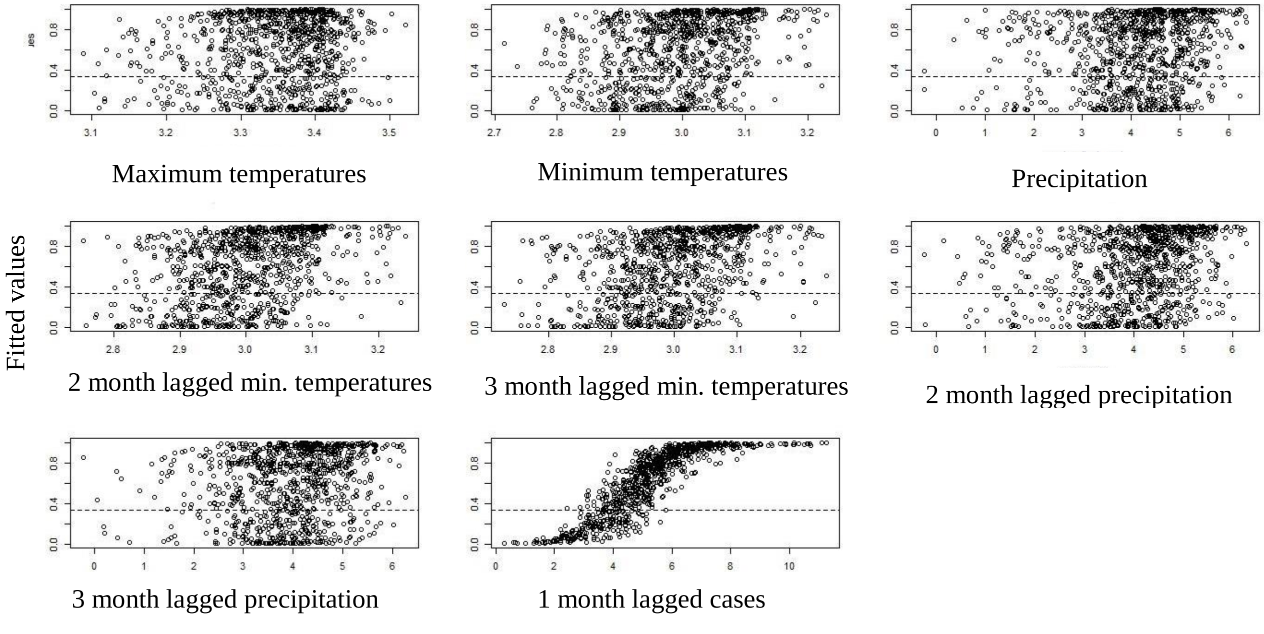
**Appendix Figure 2 Monthly dengue case counts for Pernambuco, 2010-2016.**



**Appendix Figure 3 Basic workflow of overall data collection, processing and model creation.**

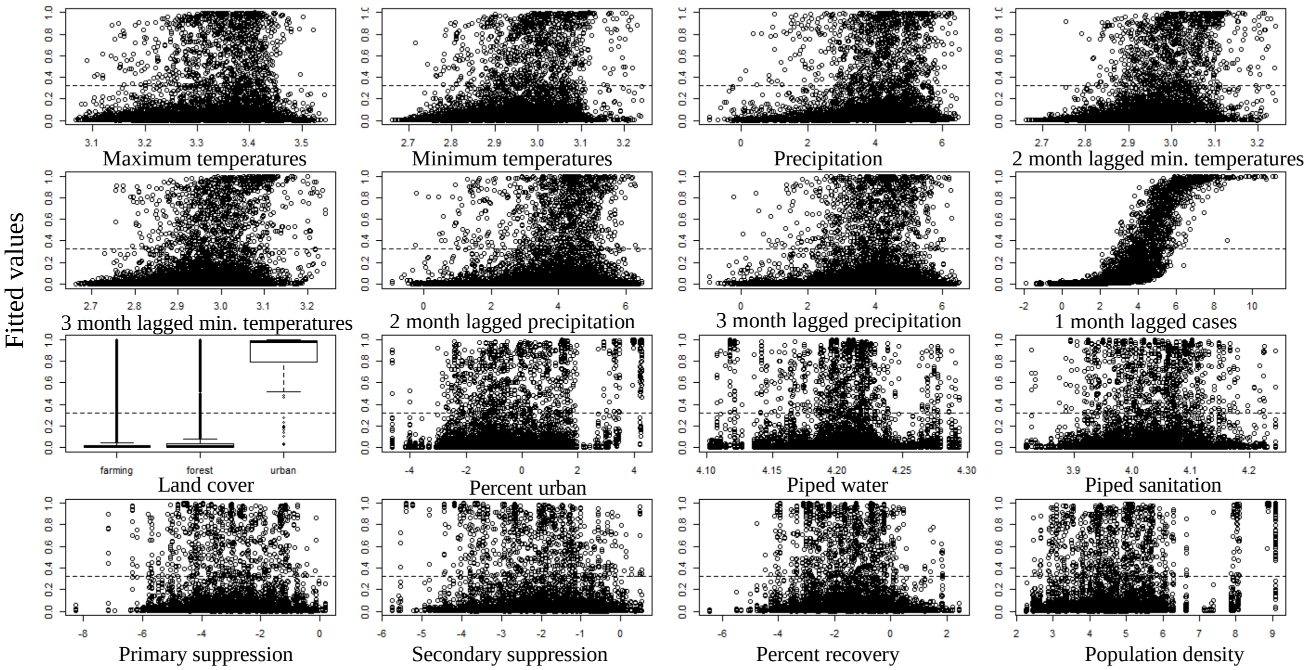


**Appendix Figure 4 Fitted values in relation to three initial climate predictors. These fitted values were the model's predictions of our mean response values when values of the predictors were input into the model. Predictor values were normalized via log transform prior to input and plotting, x axis is in log scale.**

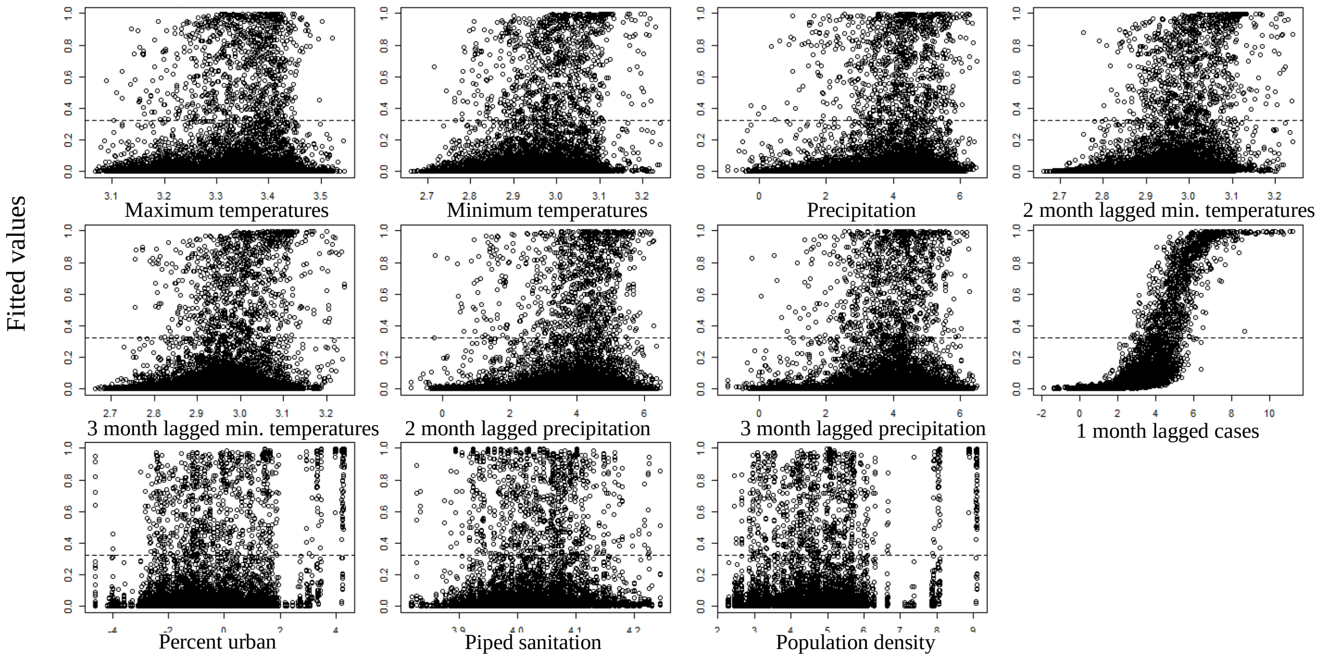


**Appendix Figure 5 Fitted values in relation to all climate, lagged climate and prior case data predictors. These fitted values were the model's predictions of mean response values when values of the predictors were**

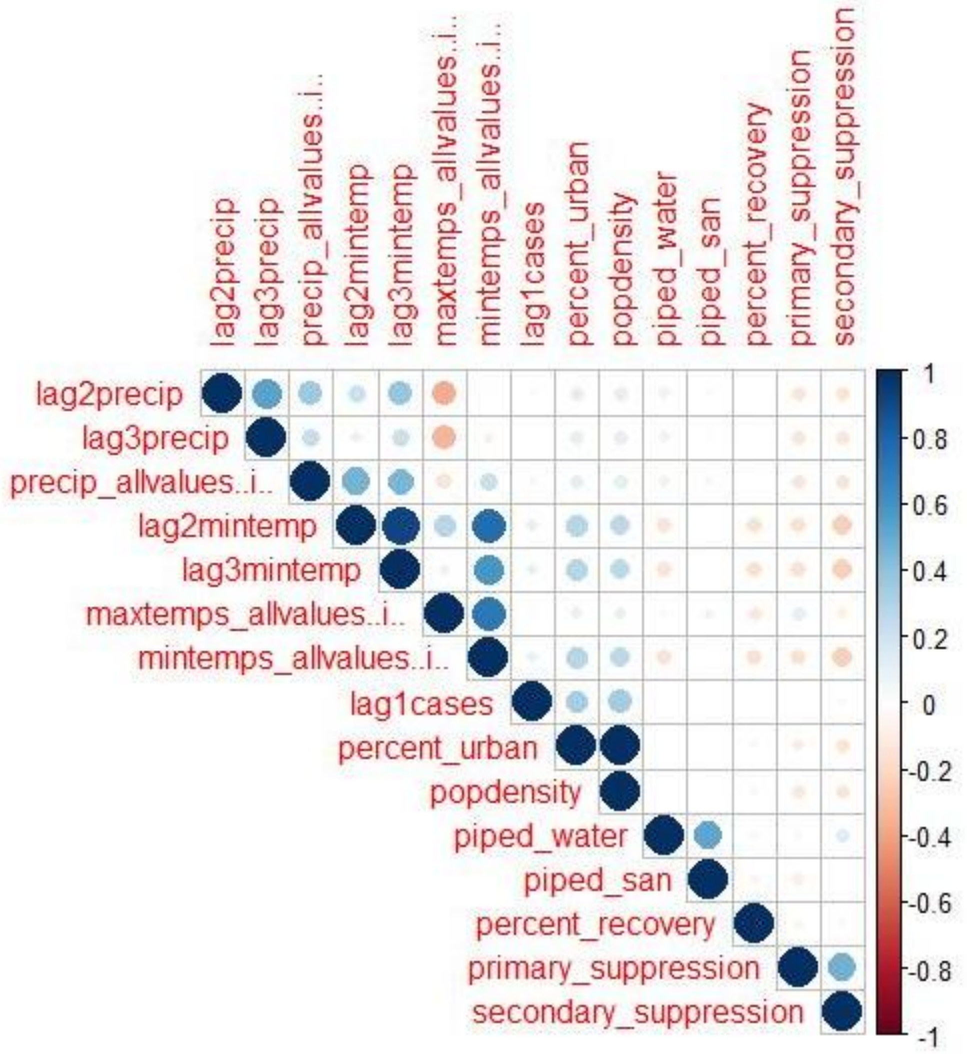
**input into the model. Predictor values were normalized via log transform prior to input and plotting. Predictor values were normalized via log transform prior to plotting, x axis is in log scale.**



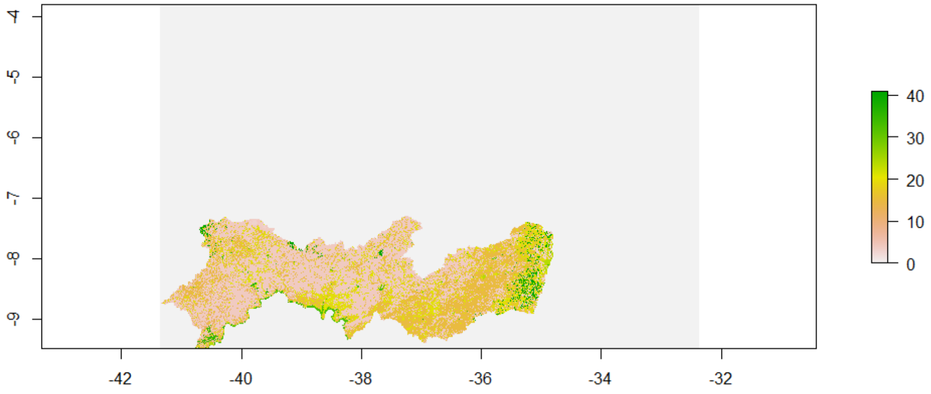
**Appendix Figure 6 Fitted values in relation to all climate and socio-environmental predictors. These fitted values were the model's predictions of mean response values when values of the predictors were input into the model. Predictor values were normalized via log transform prior to input and plotting. Predictor values were normalized via log transform prior to plotting, x axis is in log scale.**



**Appendix Figure 7 Fitted values in relation to climate and socio-environmental predictors that had greater than 1% influence. These fitted values were the model's predictions of  mean response values when values of the predictors were input into the model. Predictor values were normalized via log transform prior to input and plotting. Predictor values were normalized via log transform prior to plotting, x axis is in log scale.**



**Appendix Figure 8** Correlation matrix of all predictor variables, excluding land cover (as this was a factor variable and not numeric). Positive correlations are in blue, negative correlations are in red. Strength of the correlation is proportional to both the size and color of the circles. Boxes with correlations that had p-values > 0.05 were insignificant and left blank.



**Appendix Figure 9 Example raster of land cover classifications, January 2010. Different land cover types were mapped by their numeric classifications (e.g. "urban infrastructure" = numeric code 24).**

# Appendix C Supplementary Tables

**Appendix Table 1 All variables included for modeling. *Incidence was calculated from the number of dengue cases per municipality and the current population (assumed to all be at risk); this was our response variable.**

| Variable Name | Type | Units/Measure |
|---|---|---|
| **Incidence*** | Case | Cases/100,000 |
| **Maximum temperatures** | Climate | °C |
| **Minimum temperatures** | Climate | °C |
| **Precipitation** | Climate | mm |
| **2-month lagged minimum temperatures** | Climate | °C |
| **3-month lagged minimum temperatures** | Climate | °C |
| **2-month lagged precipitation** | Climate | mm |
| **3-month lagged precipitation** | Climate | mm |
| **1- month lagged cases** | Case | Cases/100,000 |
| **Population density** | Sociodemographic | People/km$^2$ |
| **Household access to piped water** | Sociodemographic | Percentage |
| **Household access to piped water** | Sociodemographic | Percentage |
| **Land cover** | Environmental | Factor (farming/forest/urban) |
| **Primary suppression of vegetation** | Environmental | Percentage |
| **Secondary suppression of vegetation** | Environmental | Percentage |
| **Percent recovery of vegetation** | Environmental | Percentage |
| **Percent urbanization** | Sociodemographic | Percentage |

**Appendix Table 2 Confusion matrix of outbreak risk predictions for model with only statistically significant climate predictors.**

| | Reference | |
|---|---|---|
| **Prediction** | **low** | **high** |
| **low** | 1639 | 68 |
| **high** | 99 | 198 |

**Appendix Table 3 Confusion matrix of outbreak risk predictions for model with only statistically significant climate and socio-environmental predictors.**

| | Reference | |
|---|---|---|
| **Prediction** | **Low** | **High** |
| **Low** | 1621 | 64 |
| **High** | 107 | 202 |

# Bibliography

Akter, R., Naish, S., Hu, W., & Tong, S. (2017). Socio-demographic, ecological factors and dengue infection trends in Australia. PloS one, 12(10), e0185551. https://doi.org/10.1371/journal.pone.0185551

Andrade, E. H., Figueiredo, L. B., Vilela, A. P., Rosa, J. C., Oliveira, J. G., Zibaoui, H. M., Araújo, V. E., Miranda, D. P., Ferreira, P. C., Abrahão, J. S., & Kroon, E. G. (2016). Spatial-Temporal Co-Circulation of Dengue Virus 1, 2, 3, and 4 Associated with Coinfection Cases in a Hyperendemic Area of Brazil: A 4-Week Survey. The American Journal of Tropical Medicine and Hygiene, 94(5), 1080–1084. https://doi.org/10.4269/ajtmh.15-0892

Andrioli, D.C., Busato, M.A., Lutinski, J.A. (2020). Spatial and Temporal Distribution of Dengue in Brazil, 1990 - 2017. PLoS ONE, 15(2): e0228346. https://doi.org/10.1371/journal.pone.0228346

Angelo, M., Ramalho, W. M., Gurgel, H., Belle, N., and Pilot, E. (2020). Dengue Surveillance System in Brazil: A Qualitative Study in the Federal District. International Journal of Environmental Research and Public Health, 17(6), 2062. https://doi.org/10.3390/ijerph17062062

Arboleda, S., Jaramillo-O., N., Peterson, A. T. (2009). Mapping Environmental Dimensions of Dengue Fever Transmission Risk in the Aburrá Valley, Colombia. International Journal of Environmental Research and Public Health, 6(12), 3040–3055. doi: 10.3390/ijerph6123040

Arunachalam, N., Tana, S., Espino, F., Kittayapong, P., Abeyewickrem, W., Wai, K. T., ... & Petzold, M. (2010). Eco-bio-social determinants of dengue vector breeding: a multicountry study in urban and periurban Asia. Bulletin of the world Health Organization, 88, 173-184.

Ashby, J. L., Moreno-Madriñán, M. J. (2018). Niche Modeling of Dengue Fever Using Remotely Sensed Environmental Factors and BRT. Online Journal of Public Health Informatics, 10(1). https://doi.org/10.5210/ojphi.v10i1.8336

Bäck, A. T., & Lundkvist, A. (2013). Dengue viruses - an overview. Infection ecology & epidemiology, 3, 10.3402/iee.v3i0.19839. https://doi.org/10.3402/iee.v3i0.19839

Bancroft, D., Power, G. M., Jones, R. T., Massad, E., Iriart, J. B., Preet, R., Kinsman, J., & Logan, J. G. (2022). Vector control strategies in Brazil: a qualitative investigation into community knowledge, attitudes and perceptions following the 2015–2016 Zika virus epidemic. BMJ Open, 12(1), e050991. https://doi.org/10.1136/bmjopen-2021-050991

Barrera, R., Amador, M., & MacKay, A. J. (2011). Population Dynamics of Aedes aegypti and Dengue as Influenced by Weather and Human Behavior in San Juan, Puerto Rico. PLoS Neglected Tropical Diseases, 5(12), e1378. https://doi.org/10.1371/journal.pntd.0001378

Benedum, C. M., Seidahmed, O. M. E., Eltahir, E. A. B., & Markuzon, N. (2018). Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore. PLOS Neglected Tropical Diseases, 12(12), e0006935. https://doi.org/10.1371/journal.pntd.0006935

Bhatt, S. et al. (2013). The Global Distribution and Burden of Dengue. Nature, 496, 504-507.

Brady, O. J., Smith, D. L., Scott, T. W., Hay, S. I. (2015). Dengue Disease Outbreak Definitions are Implicitly Variable. Epidemics, 11, 92–102. https://doi.org/10.1016/j.epidem.2015.03.002

Brathwaite Dick, O., San Martín, J. L., Montoya, R. H., del Diego, J., Zambrano, B., & Dayan, G. H. (2012). The History of Dengue Outbreaks in the Americas. The American Journal of Tropical Medicine and Hygiene, 87(4), 584–593. https://doi.org/10.4269/ajtmh.2012.11-0770

Buczak, A. L., Koshute, P. T., Babin, S. M., Feighner, B. H., & Lewis, S. H. (2012). A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data. BMC medical informatics and decision making, 12, 124. https://doi.org/10.1186/1472-6947-12-124

Burattini, M., Lopez, L., Coutinho, F., Siqueira-Jr, J., Homsani, S., Sarti, E., & Massad, E. (2016). Age and regional differences in clinical presentation and risk of hospitalization for dengue in Brazil, 2000–2014. Clinics, 71(8), 455–463. https://doi.org/10.6061/clinics/2016(08)08

Caldwell, J. M., LaBeaud, A. D., Lambin, E. F., Stewart-Ibarra, A. M., Ndenga, B. A., Mutuku, F. M., Krystosik, A. R., Ayala, E. B., Anyamba, A., Borbor-Cordova, M. J., Damoah, R., Grossi-Soyster, E. N., Heras, F. H., Ngugi, H. N., Ryan, S. J., Shah, M. M., Sippy, R., & Mordecai, E. A. (2021). Climate predicts geographic and temporal variation in mosquito-borne disease dynamics on two continents. Nature communications, 12(1), 1233. https://doi.org/10.1038/s41467-021-21496-7

Campbell, K. M., Lin, C. D., Iamsirithaworn, S., & Scott, T. W. (2013). The Complex Relationship between Weather and Dengue Virus Transmission in Thailand. The American Journal of Tropical Medicine and Hygiene, 89(6), 1066–1080. https://doi.org/10.4269/ajtmh.13-0321

Center for Biologics Evaluation and Research. (2020, January 22). DENGVAXIA. U.S. Food and Drug Administration. Retrieved February 8, 2022, from https://www.fda.gov/vaccines-blood-biologics/dengvaxia

Centers for Disease Control and Prevention. Symptoms and Treatment | Dengue | CDC. (2021, September 20). Retrieved January 17, 2022, from https://www.cdc.gov/dengue/symptoms/index.html

CFIR. The Consolidated Framework for Implementation Research – Technical Assistance for users of the CFIR framework. (2022). Retrieved April 11, 2022, from https://cfirguide.org

Chan, E., Lo, E., Huang, Z., Lam, H., Yeung, M. P., Kwok, K. O., Hung, K., and Tse, S. L. (2021). Sociodemographic predictors of knowledge, mosquito bite patterns and protective behaviors concerning vector borne disease: The case of dengue fever in Chinese subtropical city, Hong Kong. PLoS Neglected Tropical Diseases, 15(1), e0008993. https://doi.org/10.1371/journal.pntd.0008993

Chang, K., Chen, CD., Shih, CM. et al. (2016). Time-Lagging Interplay Effect and Excess Risk of Meteorological/Mosquito Parameters and Petrochemical Gas Explosion on Dengue Incidence. Scientific Reports, 6, 35028. https://doi.org/10.1038/srep35028

Charette, M., Berrang-Ford, L., Coomes, O., Llanos-Cuentas, E. A., Cárcamo, C., Kulkarni, M., Harper, S. L. (2020). Dengue Incidence and Sociodemographic Conditions in Pucallpa, Peruvian Amazon: What Role for Modification of the Dengue-Temperature Relationship?. The American Journal of Tropical Medicine and Hygiene, 102(1), 180–190. https://doi.org/10.4269/ajtmh.19-0033

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

Chen, S. C., Liao, C. M., Chio, C. P., Chou, H. H., You, S. H., & Cheng, Y. H. (2010). Lagged temperature effect with mosquito transmission potential explains dengue variability in southern Taiwan: Insights from a statistical analysis. Science of the Total Environment, 408(19), 4069–4075. https://doi.org/10.1016/j.scitotenv.2010.05.021

Christophers, R. (1960). Aëdes Aegypti (L.) The Yellow Fever Mosquito: Its Life History, Bionomics and Structure. Cambridge University Press.

Cordeiro, R., Donalisio, M.R., Andrade, V.R. et al. (2011). Spatial distribution of the risk of dengue fever in southeast Brazil, 2006-2007. BMC Public Health 11, 355. https://doi.org/10.1186/1471-2458-11-355

Costa, A.I.P, Natal, D. Distribuição espacial da dengue e determinantes socioeconômicos em localidade urbana no Sudeste do Brasil. Rev Saúde Pública. 1998;32(3):232–236.

Cucunawangsih, N. P. H. L. (2017). Trends of Dengue Disease Epidemiology. Virology: Research and Treatment, 8, 1178122X1769583. https://doi.org/10.1177/1178122x17695836 da Conceição Araújo, D., dos Santos, A. D., Lima, S. V. M. A., Vaez, A. C., Cunha, J. O., & Conceição Gomes Machado De Araújo, K. (2020). Determining the association between dengue and social inequality factors in north-eastern Brazil: A spatial modelling. Geospatial Health, 15(1). https://doi.org/10.4081/gh.2020.854

de Mattos Almeida, M. C., Caiaffa, W. T., Assunção, R. M., & Proietti, F. A. (2007). Spatial vulnerability to dengue in a Brazilian urban area during a 7-year surveillance. Journal of urban health : bulletin of the New York Academy of Medicine, 84(3), 334–345. https://doi.org/10.1007/s11524-006-9154-2

Dengue. Pan American Health Organization. (n.d.). Retrieved January 28, 2022, from https://www.paho.org/en/topics/dengue

Dengue Viruses. (2014). Scitable by Nature Education. Retrieved January 28, 2022, from https://www.nature.com/scitable/topicpage/dengue-viruses-22400925/?error=cookies_not_supported&code=c8f2276a-7cb7-4b97-a5a6-b8ea05a78307

do Nascimento, T.B.P. et al. (2019). e-Participation in the Exploration of Innovations: Confronting Aedes aegypti and Microcephaly in Brazil. In: Kő, A., Francesconi, E., Anderst-Kotsis, G., Tjoa, A., Khalil, I. (eds) Electronic Government and the Information Systems Perspective. EGOVIS 2019. Lecture Notes in Computer Science(), vol 11709. Springer, Cham. https://doi.org/10.1007/978-3-030-27523-5_8

Ebi, K. L., & Nealon, J. (2016). Dengue in a changing climate. Environmental Research, 151, 115–123. https://doi.org/10.1016/j.envres.2016.07.026

Editors of Encyclopaedia Britannica (2016, June 16). Pernambuco. Encyclopedia Britannica. https://www.britannica.com/place/Pernambuco

Ehrenkranz, N. J., Ventura, A. K., Cuadrado, R. R., Pond, W. L., & Porter, J. E. (1971). Pandemic Dengue in Caribbean Countries and the Southern United States — Past, Present and Potential Problems. New England Journal of Medicine, 285(26), 1460–1469. https://doi.org/10.1056/nejm197112232852606

Elith, J., Leathwick, J. R., Hastie, T. (2008). A Working Guide to Boosted Regression Trees. Journal of Animal Ecology, 77(4), 802-813. doi:10.1111/j.1365-2656.2008.01390.x

Fares, R. C., Souza, K. P., Añez, G., & Rios, M. (2015). Epidemiological Scenario of Dengue in Brazil. BioMed Research International, 2015, 321873. https://doi.org/10.1155/2015/321873

Favier, C., Schmit, D., Müller-Graf, C. D., Cazelles, B., Degallier, N., Mondet, B., Dubois, M. A. (2005). Influence of Spatial Heterogeneity on an Emerging Infectious Disease: the Case of Dengue Epidemics. Proceedings of the Royal Society B: Biological Sciences, 272(1568), 1171–1177. doi: 10.1098/rspb.2004.3020

Fernández-Salas, I., Danis-Lozano, R., Casas-Martínez, M., Ulloa, A., Bond, J. G., Marina, C. F., . . . Díaz-González, E. E. (2015). Historical Inability to Control Aedes Aegypti as a Main Contributor of Fast Dispersal of Chikungunya Outbreaks in Latin America. Antiviral Research, 124, 30-42. doi:10.1016/j.antiviral.2015.10.015

Franklinos, L. H. V., Jones, K. E., Redding, D. W., & Abubakar, I. (2019). The effect of global change on mosquito-borne disease. The Lancet Infectious Diseases, 19(9), e302–e312. https://doi.org/10.1016/s1473-3099(19)30161-6

Freitas, M. G., Tsouris, P., Reis, I. C., Magalhães, M. de Nascimento, T. F., and Honório, N. A. (2010). Dengue and Land Cover Heterogeneity in Rio De Janeiro. Oecologia Australis, 14(03), 641–667. https://doi.org/10.4257/oeco.2010.1403.04

Fried, J. R., Gibbons, R. V., Kalayanarooj, S., Thomas, S. J., Srikiatkhachorn, A., Yoon, I. K., Jarman, R. G., Green, S., Rothman, A. L., & Cummings, D. A. (2010). Serotype-specific differences in the risk of dengue hemorrhagic fever: an analysis of data collected in Bangkok, Thailand from 1994 to 2006. PLoS neglected tropical diseases, 4(3), e617. https://doi.org/10.1371/journal.pntd.0000617

Gao, P., Pilot, E., Rehbock, C., Gontariuk, M., Doreleijers, S., Wang, L., Krafft, T., Martens, P., & Liu, Q. (2021). Land use and land cover change and its impacts on dengue dynamics in China: A systematic review. PLOS Neglected Tropical Diseases, 15(10), e0009879. https://doi.org/10.1371/journal.pntd.0009879

Gubler, D. J., Clark, G. G. (1995). Dengue/Dengue Hemorrhagic Fever: The Emergence of a Global Health Problem. Emerging Infectious Diseases, 1(2), 55–57. doi: 10.3201/eid0102.952004

Gubler, D. J. (1998). Dengue and Dengue Hemorrhagic Fever. Clinical Microbiology Reviews, 11(3), 480–496. https://doi.org/10.1128/CMR.11.3.480

Guzman, M. (2003). Dengue and dengue hemorrhagic fever in the Americas: lessons and challenges. Journal of Clinical Virology, 27(1), 1–13. https://doi.org/10.1016/s1386-6532(03)00010-6

Guzmán, M. G., & Kourí, G. (2004). Dengue diagnosis, advances and challenges. International Journal of Infectious Diseases, 8(2), 69–80. https://doi.org/10.1016/j.ijid.2003.03.003

Guzman, M., Gubler, D., Izquierdo, A. et al. (2016). Dengue infection. Nat Rev Dis Primers 2, 16055. https://doi.org/10.1038/nrdp.2016.55

Hadinegoro, S. R., Arredondo-García, J. L., Capeding, M. R., Deseda, C., Chotpitayasunondh, T., Dietze, R., Hj Muhammad Ismail, H., Reynales, H., Limkittikul, K., Rivera-Medina, D. M., Tran, H. N., Bouckenooghe, A., Chansinghakul, D., Cortés, M., Fanouillere, K., Forrat, R., Frago, C., Gailhardou, S., Jackson, N., . . . Saville, M. (2015). Efficacy and Long-Term

Safety of a Dengue Vaccine in Regions of Endemic Disease. New England Journal of Medicine, 373(13), 1195–1206. https://doi.org/10.1056/nejmoa1506223

Harapan, H., Michie, A., Sasmono, R. T., & Imrie, A. (2020). Dengue: A Minireview. Viruses, 12(8), 829. https://doi.org/10.3390/v12080829

Hashizume, M., Dewan, A. M., Sunahara, T., Rahman, M. Z., & Yamamoto, T. (2012). Hydroclimatological variability and dengue transmission in Dhaka, Bangladesh: a time-series study. BMC Infectious Diseases, 12(1). https://doi.org/10.1186/1471-2334-12-98

Hernández-Ávila, J. E., Rodríguez, M.-H., Santos-Luna, R., Sánchez-Castañeda, V., Román-Pérez, S., Ríos-Salgado, V. H., Salas-Sarmiento, J. A. (2013). Nation-Wide, Web-Based, Geographic Information System for the Integrated Surveillance and Control of Dengue Fever in Mexico. PLoS ONE, 8(8). doi:10.1371/journal.pone.0070231

Heukelbach, J., de Oliveira, F. A. S., Kerr-Pontes, L. R. S., & Feldmeier, H. (2001). Risk factors associated with an outbreak of dengue fever in a favela in Fortaleza, north-east Brazil. Tropical Medicine and International Health, 6(8), 635–642. https://doi.org/10.1046/j.1365-3156.2001.00762.x

Husin, N.A, Salim, N. and Ahmad, A.R. (2008). Modeling of dengue outbreak prediction in Malaysia: A comparison of Neural Network and Nonlinear Regression Model. 2008 International Symposium on Information Technology, pp. 1-4, doi: 10.1109/ITSIM.2008.4632022.

Husnina, Z., Clements, A. C., and Wangdi, K. (2019). Forest Cover and Climate as Potential Drivers for Dengue Fever in Sumatra and Kalimantan 2006–2016: a Spatiotemporal Analysis. Tropical Medicine and International Health, 24(7), 888–898. https://doi.org/10.1111/tmi.13248

IBGE - Brazilian Institute of Geography and Statistics. (2022). Censo 2022. Brazilian Institute of Geography and Statistics. Retrieved April 10, 2022, from https://censo2022.ibge.gov.br/

Iguchi, J. A., Seposo, X. T., Honda, Y. (2018). Meteorological Factors Affecting Dengue Incidence in Davao, Philippines. BMC Public Health, 18(1). doi:10.1186/s12889-018-5532-4

Islam, S., Haque, C. E., Hossain, S., & Hanesiak, J. (2021). Climate Variability, Dengue Vector Abundance and Dengue Fever Cases in Dhaka, Bangladesh: A Time-Series Study. Atmosphere, 12(7), 905. https://doi.org/10.3390/atmos12070905

Jain, R., Sontisirikit, S., Iamsirithaworn, S., & Prendinger, H. (2019). Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. BMC infectious diseases, 19(1), 272.

Johansen, I. C., Carmo, R. L. D., Alves, L. C., & Dias Bueno, M. D. C. (2018). Environmental and demographic determinants of dengue incidence in Brazil. Revista de Salud Pública, 20(3), 346–353. https://doi.org/10.15446/rsap.v20n3.54315

Johansson, M. A., Dominici, F., Glass, G. E. (2009). Local and Global Effects of Climate on Dengue Transmission in Puerto Rico. PLoS Neglected Tropical Diseases, 3(2). doi:10.1371/journal.pntd.0000382

Kabir, M. A., Zilouchian, H., Younas, M. A., & Asghar, W. (2021). Dengue Detection: Advances in Diagnostic Tools from Conventional Technology to Point of Care. Biosensors, 11(7), 206. https://doi.org/10.3390/bios11070206

Kakarla, S. G., Caminade, C., Mutheneni, S. R., Morse, A. P., Upadhyayula, S. M., Kadiri, M. R., & Kumaraswamy, S. (2019). Lag effect of climatic variables on dengue burden in India. Epidemiology and infection, 147, e170. https://doi.org/10.1017/S0950268819000608

Kalbus A, de Souza Sampaio V, Boenecke J, Reintjes R. (2021). Exploring the influence of deforestation on dengue fever incidence in the Brazilian Amazonas state. PLoS ONE, 16(1): e0242685. https://doi.org/10.1371/journal.pone.0242685

Karuppusamy, B., Sarma, D. K., Lalmalsawma, P., Pautu, L., Karmodiya, K., & Balabaskaran Nina, P. (2021). Effect of climate change and deforestation on vector borne diseases in the North-Eastern Indian State of Mizoram bordering Myanmar. The Journal of Climate Change and Health, 2, 100015. https://doi.org/10.1016/j.joclim.2021.100015

Kuhn, R. J., Zhang, W., Rossmann, M. G., Pletnev, S. V., Corver, J., Lenches, E., Jones, C. T., Mukhopadhyay, S., Chipman, P. R., Strauss, E. G., Baker, T. S., & Strauss, J. H. (2002). Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. Cell, 108(5), 717–725. https://doi.org/10.1016/s0092-8674(02)00660-8

Kulkarni, M.A., Duguay, C. & Ost, K. Charting the evidence for climate change impacts on the global spread of malaria and dengue and adaptive responses: a scoping review of reviews. Global Health 18, 1 (2022). https://doi.org/10.1186/s12992-021-00793-2

Lambrechts, L., Paaijmans, K. P., Fansiri, T., Carrington, L. B., Kramer, L. D., Thomas, M. B., & Scott, T. W. (2011). Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti. Proceedings of the National Academy of Sciences, 108(18), 7460–7465. https://doi.org/10.1073/pnas.1101377108

Lauer, S. A. et al. (2018). Prospective Forecasts of Annual Dengue Hemorrhagic Fever Incidence in Thailand, 2010–2014. Proceedings of the National Academy of Sciences, 115, E2175–E2182

Lescar J., Soh S., Lee L.T., Vasudevan S.G., Kang C., Lim S.P. (2018). The Dengue Virus Replication Complex: From RNA Replication to Protein-Protein Interactions to Evasion of

Innate Immunity. In: Hilgenfeld R., Vasudevan S. (eds) Dengue and Zika: Control and Antiviral Treatment Strategies. Advances in Experimental Medicine and Biology, vol 1062. Springer, Singapore. https://doi.org/10.1007/978-981-10-8727-1_9

Li, C., Lu, Y., Liu, J., Wu, X. (2018). Climate Change and Dengue Fever Transmission in China: Evidences and Challenges. Science of The Total Environment, 622-623, 493–501. doi: 10.1016/j.scitotenv.2017.11.326

Loroño-Pino, M. A., Beaty, B. J., Farfán, J. A., Rosado-Paredes, E. P., Flores-Flores, L. F., Vorndam, A. V., Gubler, D. J., Cropp, C. B., & Rodríguez-Angulo, E. M. (1999). Common occurrence of concurrent infections by multiple dengue virus serotypes. The American Journal of Tropical Medicine and Hygiene, 61(5), 725–730. https://doi.org/10.4269/ajtmh.1999.61.725

Lowe, R., Gasparrini, A., Van Meerbeeck, C. J., Lippi, C. A., Mahon, R., Trotman, A. R., . . . Stewart-Ibarra, A. M. (2018). Nonlinear and Delayed Impacts of Climate on Dengue Risk In Barbados: A Modelling Study. PLOS Medicine, 15(7). doi:10.1371/journal.pmed.1002613

Lozano-Fuentes, S., Ochoa-Martinez, C., Tapia-Santos, B., Kobylinski, K. C., Uejio, C. K., Zielinski-Gutierrez, E., Monache, L. D., Steinhoff, D. F., Monaghan, A. J., Eisen, L., Hayden, M. H., & Welsh-Rodriguez, C. (2012). The Dengue Virus Mosquito Vector Aedes aegypti at High Elevation in México. The American Journal of Tropical Medicine and Hygiene, 87(5), 902–909. https://doi.org/10.4269/ajtmh.2012.12-0244

Maciel-de-Freitas, R., & Valle, D. (2014). Challenges Encountered Using Standard Vector Control Measures for Dengue in Boa Vista, Brazil. Bulletin of the World Health Organization, 92(9), 685–689.

Maheshwari, A. (2019, September 30). SMOTE for Imbalanced Dataset. OpenGenus IQ: Computing Expertise & Legacy. Retrieved April 22, 2022, from https://iq.opengenus.org/smote-for-imbalanced-dataset/

Matysiak, A., and Roess, A. (2017). Interrelationship Between Climatic, Ecologic, Social, and Cultural Determinants Affecting Dengue Emergence and Transmission in Puerto Rico and Their Implications for Zika Response. Journal of Tropical Medicine, 2017, 1–14. https://doi.org/10.1155/2017/8947067

McGough, S. F., Clemente, L., Kutz, J. N., & Santillana, M. (2021). A Dynamic, Ensemble Learning Approach to Forecast Dengue fever epidemic years in Brazil using weather and population susceptibility cycles. Journal of The Royal Society Interface, 18(179), 20201006. https://doi.org/10.1098/rsif.2020.1006

Messina, J. P., Brady, O. J., Scott, T. W., Zou, C., Pigott, D. M., Duda, K. A., Bhatt, S., Katzelnick, L., Howes, R. E., Battle, K. E., Simmons, C. P., & Hay, S. I. (2014). Global spread of dengue virus types: mapping the 70 year history. Trends in Microbiology, 22(3), 138–146. https://doi.org/10.1016/j.tim.2013.12.011

Messina, J. P., Brady, O. J., Golding, N., Kraemer, M. U., Wint, G. R., Ray, S. E., . . . Hay, S. I. (2019). The Current and Future Global Distribution and Population at Risk of Dengue. Nature Microbiology, 4(9), 1508-1515. doi:10.1038/s41564-019-0476-8

Morin, C. W., Comrie, A. C., & Ernst, K. (2013). Climate and Dengue Transmission: Evidence and Implications. Environmental Health Perspectives, 121(11–12), 1264–1272. https://doi.org/10.1289/ehp.1306556

Muller, D.A, Depelsenaire, A.C.I, Young, P.R. (2017).  Clinical and Laboratory Diagnosis of Dengue Virus Infection. The Journal of Infectious Diseases. 215(2), S89–S95. https://doi.org/10.1093/infdis/jiw649

Mulligan, K., Dixon, J., Joanna Sinn, C. L., & Elliott, S. J. (2014). Is dengue a disease of poverty? A systematic review. Pathogens and Global Health, 109(1), 10–18. https://doi.org/10.1179/2047773214y.0000000168

Murugesan, A. & M. Manoharan (2019). Chapter 16 -- Dengue Virus. In Emerging and Reemerging Viral Pathogens: Volume 1: Fundamental and Basic Virology Aspects of Human, Animal and Plant Pathogens (1st ed., Vol. 1, pp. 281–359). Academic Press. https://doi.org/10.1016/B978-0-12-819400-3.00016-8

Murugesan, A., & Manoharan, M. (2020). Dengue Virus. Emerging and Reemerging Viral Pathogens, 281–359. https://doi.org/10.1016/B978-0-12-819400-3.00016-8

NOAA. (2021, March 15). Climate Change: Global Temperature | NOAA Climate.gov. Retrieved February 25, 2022, from https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature

Oliveira, M., Lert-itthiporn, W., Cavadas, B., Fernandes, V., Chuansumrit, A., Anunciação, O., Casademont, I., Koeth, F., Penova, M., Tangnarararatchakit, K., Khor, C. C., Paul, R., Malasit, P., Matsuda, F., Simon-Lorière, E., Suriyaphol, P., Pereira, L., & Sakuntabhai, A. (2018). Joint ancestry and association test indicate two distinct pathogenic pathways involved in classical dengue fever and dengue shock syndrome. PLOS Neglected Tropical Diseases, 12(2), e0006202. https://doi.org/10.1371/journal.pntd.0006202

Olson, S. H., Gangnon, R., Silveira, G. A., & Patz, J. A. (2010). Deforestation and Malaria in Mâncio Lima County, Brazil. Emerging Infectious Diseases, 16(7), 1108–1115. https://doi.org/10.3201/eid1607.091785

O'Neill S.L. (2018). The Use of Wolbachia by the World Mosquito Program to Interrupt Transmission of Aedes aegypti Transmitted Viruses. In: Hilgenfeld R., Vasudevan S. (eds) Dengue and Zika: Control and Antiviral Treatment Strategies. Advances in Experimental Medicine and Biology, 1062, 355-360. Springer, Singapore.

Paz-Bailey, G. (2021, December 16). Dengue Vaccine: Recommendations of the Advisory Committee. . . Centers for Disease Control and Prevention. Retrieved February 7, 2022, from https://www.cdc.gov/mmwr/volumes/70/rr/rr7006a1.htm#:%7E:text=Dengvaxia%20is%20a%20safe%20and,first%20natural%20infection%20after%20vaccination.

Perera, R., & Kuhn, R. J. (2008). Structural proteomics of dengue virus. Current Opinion in Microbiology, 11(4), 369–377. https://doi.org/10.1016/j.mib.2008.06.004

Piassi Godói, I., Dias Da Silva, L.V., Razzaque Sarker, A., Megiddo, I., Morton, A., Godman, B., Alvarez-Madrazo, S., Bennie, M., Afonso Guerra-Junior, A. (2018). Economic and Epidemiological Impact of Dengue Illness Over 16 years From a Public Health System Perspective in Brazil to Inform Future Health Policies Including the Adoption of a Dengue Vaccine. Expert Review of Vaccines, 17:12, 1123-1133, doi: 10.1080/14760584.2018.1546581

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramadona, A. L., Lazuardi, L., Hii, Y. L., Holmner, S., Kusnanto, H., & Rocklöv, J. (2016). Prediction of Dengue Outbreaks Based on Disease Surveillance and Meteorological Data. PLOS ONE, 11(3), e0152688.

Reinhold, J. M., Lazzari, C. R., Lahondère, C. (2018). Effects of the Environmental Temperature on Aedes Aegypti and Aedes Albopictus Mosquitoes: A Review. Insects, 9(4), 158. https://doi.org/10.3390/insects9040158

Rohani A, Wong YC, Zamre I, Lee HL, Zurainee MN. 2009. The effect of extrinsic incubation temperature on development of dengue serotype 2 and 4 viruses in Aedes aegypti (L.).Southeast Asian J Trop Med Public Health,40:942-95019842378.

Rueda, L.M, Patel, K. J. ,Axtell, R., Stinner, C. R. E., Temperature-Dependent Development and Survival Rates of Culex quinquefasciatus and Aedes aegypti (Diptera: Culicidae), Journal of Medical Entomology, Volume 27, Issue 5, 1 September 1990, Pages 892–898, https://doi.org/10.1093/jmedent/27.5.892

Salles, T. S., da Encarnação Sá-Guimarães, T., de Alvarenga, E., Guimarães-Ribeiro, V., de Meneses, M., de Castro-Salles, P. F., Dos Santos, C. R., do Amaral Melo, A. C., Soares, M. R., Ferreira, D. F., & Moreira, M. F. (2018). History, epidemiology and diagnostics of dengue in the American and Brazilian contexts: a review. Parasites & vectors, 11(1), 264. https://doi.org/10.1186/s13071-018-2830-8

Sarfraz, M.S., Tripathi, N.K., Tipdecho, T. et al. (2012). Analyzing the Spatio-Temporal Relationship Between Dengue Vector Larval Density and Land-Use Using Factor Analysis and Spatial Ring Mapping. BMC Public Health, 12, 853. https://doi.org/10.1186/1471-2458-12-853

Schmidt, W. P., Suzuki, M., Thiem, V. D., White, R. G., Tsuzuki, A., Yoshida, L. M., Yanai, H., Haque, U., Tho, l., Anh, D. D., and Ariyoshi, K. (2011). Population density, water supply, and the risk of dengue fever in Vietnam: cohort study and spatial analysis. PLoS Medicine, 8(8), e1001082. https://doi.org/10.1371/journal.pmed.1001082

Schneider, J. and Droll, D. (2001, June). A Timeline for Dengue in the Americas to December 31, 2000 and Noted First Occurrences. Pan American Health Organization, Division of Disease Prevention and Control. Retrieved January 24, 2022, from https://www.paho.org/hq/dmdocuments/2010/A%20timeline%20for%20dengue.pdf

Seidahmed, O. M., & Eltahir, E. A. (2016). A Sequence of Flushing and Drying of Breeding Habitats of Aedes aegypti (L.) Prior to the Low Dengue Season in Singapore. PLoS neglected tropical diseases, 10(7), e0004842. https://doi.org/10.1371/journal.pntd.0004842

Sierra, B., Kourí, G. & Guzmán, M. (2007). Race: a risk factor for dengue hemorrhagic fever. Arch Virol 152, 533. https://doi.org/10.1007/s00705-006-0869-x

Stevens, F. R., Gaughan, A. E., Linard, C., Tatem, A. J. (2015). Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. Plos One, 10(2). doi: 10.1371/journal.pone.0107042

Stewart Ibarra, A. M., Ryan, S. J., Beltrán, E., Mejía, R., Silva, M., & Muñoz, A. (2013). Dengue vector dynamics (Aedes aegypti) influenced by climate and social factors in Ecuador: implications for targeted control. PloS one, 8(11), e78263. https://doi.org/10.1371/journal.pone.0078263

Stolerman, L., Maia, P. Kutz, J. N. (2016). Data-Driven Forecast of Dengue Outbreaks in Brazil: A Critical Assessment of Climate Conditions for Different Capitals. arXiv.org. Retrieved April 17, 2021, from https://arXiv.org/abs/1701.00166

Tantawichien T. (2012). Dengue fever and dengue haemorrhagic fever in adolescents and adults. Paediatrics and international child health, 32 Suppl 1(s1), 22–27. https://doi.org/10.1179/2046904712Z.00000000049

Teixeira M.G., Siqueira, J.B. Jr., Ferreira G.L.C., Bricks L., Joint G. (2013). Epidemiological Trends of Dengue Disease in Brazil (2000–2010): A Systematic Literature Search and Analysis. PLOS Neglected Tropical Diseases, 7(12). https://doi.org/10.1371/journal.pntd.0002520

The Borgen Project. (2020, July 10). Crossing streets: Social divides and urbanization in Brazil. Retrieved January 28, 2022, from https://borgenproject.org/crossing-streets-social-divides-urbanization-brazil/

Tran, B. L., Tseng, W. C., Chen, C. C., & Liao, S. Y. (2020). Estimating the Threshold Effects of Climate on Dengue: A Case Study of Taiwan. International journal of environmental research and public health, 17(4), 1392. https://doi.org/10.3390/ijerph17041392

Tun-Lin, W., Burkot, T. R., & Kay, B. H. (2000). Effects of temperature and larval diet on development rates and survival of the dengue vector Aedes aegypti in north Queensland, Australia. Medical and Veterinary Entomology, 14(1), 31–37. https://doi.org/10.1046/j.1365-2915.2000.00207.x

Vanwambeke, S. O., Bennett, S. N., Kapan, D. D. (2010). Spatially Disaggregated Disease Transmission Risk: Land Cover, Land Use and Risk of Dengue Transmission on the Island of Oahu. Tropical Medicine and International Health, 16(2), 174–185. doi: 10.1111/j.1365-3156.2010.02671.x

Vargas, W. P., Kawa, H., Sabroza, P. C., Soares, V. B., Honório, N. A., and de Almeida, A. S. (2015). Association among house infestation index, dengue incidence, and sociodemographic indicators: surveillance using geographic information system. BMC Public Health, 15, 746. https://doi.org/10.1186/s12889-015-2097-3

Vasconcelos P.F.C, Lima J.W.O, Rosa A.P.A.T, Timbó M.J., Rosa E.S.T, Lima H.R., et al. (1998). Dengue epidemic in Fortaleza, Ceará: random sero-epidemiological survey. Rev Saude Publica. 32(5):447-54.

Wagner, F. E. & Ward, J. O. (1980). Urbanization and migration in Brazil. American Journal of Economics and Sociology, 39(3), 249–259. https://doi.org/10.1111/j.1536-7150.1980.tb01276.x

Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., … Tatem, A. J. (2018). Spatially Disaggregated Population Estimates in the Absence of National Population and Housing Census Data. Proceedings of the National Academy of Sciences, 115(14), 3529–3537. doi: 10.1073/pnas.1715305115

Watts DM, Burke DS, Harrison BA, Whitmire RE, Nisalak A. (1987). Effect of temperature on the vector efficiency of Aedes aegypti for dengue 2 virus. Am J Trop Med Hyg 36:143-1523812879.

Winch, P. J., Rigau-Pérez, J. G., Clark, G. G., Ruiz-Pérez, M., Leontsini, E., and Gubler, D. J. (2002). Community-based dengue prevention programs in Puerto Rico: Impact on knowledge, behavior, and residential mosquito infestation. The American Journal of Tropical Medicine and Hygiene, 67(4), 363–370. https://doi.org/10.4269/ajtmh.2002.67.363

Whiteman, A., Loaiza, J. R., Yee, D. A., Poh, K. C., Watkins, A. S., Lucas, K. J., Rapp, T. J., Kline, L., Ahmed, A., Chen, S., Delmelle, E., and Oguzie, J. U. (2020). Do socioeconomic factors drive Aedes mosquito vectors and their arboviral diseases? A systematic review of dengue, chikungunya, yellow fever, and Zika Virus. One Health, 11, 100188. https://doi.org/10.1016/j.onehlt.2020.100188

World Health Organization. (2012). Global Strategy for Dengue Prevention and Control, 2012-2020. Retrieved January 15, 2022 from: https://apps.who.int/iris/bitstream/handle/10665/75303/9789241504034_eng.pdf

World Health Organization. (2018). Dengue vaccines: WHO position paper – September 2018. Weekly epidemiological record; No 36, 93, 457–476

World Health Organization. (2022, January 10). *Dengue and severe dengue*. Retrieved January 28, 2022, from https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

Wu, P., Lay, J., Guo, H., Lin, C., Lung, S., Su, H. (2009). Higher Temperature and Urbanization Affect the Spatial Patterns of Dengue Fever Transmission in Subtropical Taiwan. Science of The Total Environment, 407(7), 2224-2233.

Xiang, J., Hansen, A., Liu, Q., Liu, X., Tong, M. X., Sun, Y., Cameron, S., Hanson-Easey, S., Han, G. S., Williams, C., Weinstein, P., & Bi, P. (2017). Association between dengue fever incidence and meteorological factors in Guangzhou, China, 2005–2014. Environmental Research, 153, 17–26. https://doi.org/10.1016/j.envres.2016.11.009

Xu, Z., Bambrick, H., Frentiu, F. D., Devine, G., Yakob, L., Williams, G., & Hu, W. (2020). Projecting the future of dengue under climate change scenarios: Progress, uncertainties and research needs. PLOS Neglected Tropical Diseases, 14(3), e0008118. https://doi.org/10.1371/journal.pntd.0008118

Zellweger, R. M., Cano, J., Mangeas, M., Taglioni, F., Mercier, A., Despinoy, M., Menkès, C. E., Dupont-Rouzeyrol, M., Nikolay, B., & Teurlai, M. (2017). Socioeconomic and environmental determinants of dengue transmission in an urban setting: An ecological study in Nouméa, New Caledonia. PLoS neglected tropical diseases, 11(4), e0005471. https://doi.org/10.1371/journal.pntd.0005471