UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

EARTH OBSERVATION AND MOSQUITO-BORNE DISEASES: ASSESSING ENVIRONMENTAL RISK FACTORS FOR DISEASE TRANSMISSION VIA REMOTE SENSING DATA

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ANDREA MCMAHON

Norman, Oklahoma

2021

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/330179

EARTH OBSERVATION AND MOSQUITO-BORNE DISEASES: ASSESSING ENVIRONMENTAL RISK FACTORS FOR DISEASE TRANSMISSION VIA REMOTE SENSING DATA

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SUSTAINABILITY

BY THE COMMITTEE CONSISTING OF

Dr. Michael C. Wimberly, Chair

Dr. Kirsten de Beurs

Dr. John S. Greene

Dr. Katherine Hirschfeld

© Copyright by ANDREA MCMAHON 2021
All Rights Reserved.

To all the women who have come before me and have trodden the path on which I now walk.

Acknowledgements

Over the course of the last five years, I have received tremendous support from so many people, and I want to express my deepest gratitude to all of them.

First and foremost, I would like to thank my advisor Dr. Michael Wimberly. They say an advisor can make or break the PhD experience. I sincerely want to thank you for giving me the opportunity to do this exciting research, and for your patience and ongoing support without which I would not have made it to this point.

I would also like to thank my committee members, each of whom has supported me on different segments of my PhD journey. I would like to thank Dr. Geoffrey Henebry and Dr. Michael Hildreth who served on my committee during my time at South Dakota State University and who have been supporting in framing my research ideas. I would also like to thank all of my committee members at the University of Oklahoma: Drs. Kirsten de Beurs, Katherine Hirschfeld, Scott Greene, Hernan Moreno, and Matthew Miller. I thank you all for your time, your support, and your insightful advice. Special shout out to those who served on my committee during my last year. Writing a dissertation is hard and writing one during a global pandemic is even harder. I am tremendously grateful for how flexible and supportive you all were during this time.

Any of this work in this dissertation would have been impossible without all those who were involved as co-authors in each of the case studies, and I wholeheartedly want to thank them as well. This includes Dr. Justin Davis who always had an open ear for questions of a statistical nature, Dr. Caio França who helped me learn how to identify mosquitoes, and

our colleagues in Ethiopia who made our malaria research possible: Dr. Abere Mihretie, Adem Agmas Ahmed, Mastewal Lake, and Worku Awoke.

A very big thank you also goes to past and present EcoGRAPH Lab members who gave valuable feedback during lab meetings and were always there when I needed help. Special thanks go to Dr. Dawn Nekorchuk who has been endlessly supportive and has become a dear friend.

And since I transferred half-way through my studies, I would like to thank everybody at the Geospatial Sciences Center of Excellence at SDSU and at the Department of Geography and Environmental Sustainability at OU. This goes to faculty members for teaching such interesting classes, to department staff who relentlessly help with every problem graduate students come up with, and to fellow students who have become great friends along the way.

Last but not least, I want to thank my biggest champions: my parents who are always there for me even when I'm far away, my PhD support puppy Hazel whose cold snout is what gets me up in the morning, and of course my husband Kenny who quit his job to be with me while I worked on my degree and who was crazy enough to marry me during a pandemic.

Contents

Acknowledgements ..... **v**

List of Tables ..... **x**

List of Figures ..... **xi**

Abstract ..... **xiv**

1 Introduction **1**

1.1 Challenges in combatting mosquito-borne diseases 2

1.2 Ecology of mosquito-borne diseases on a landscape scale 4

1.3 The use of satellite data to study mosquito-borne diseases 7

1.4 Research overview 9

2 Remote sensing of environmental risk factors for malaria in different geographic contexts **11**

2.1 Introduction 12

2.2 Methods 16

2.2.1 Study area 16

2.2.2 Malaria data 19

2.2.3 Environmental variables 21

2.2.4 Statistical analysis 26

2.3 Results 27

2.3.1 Spatio-temporal patterns in malaria cases 27

2.3.2 Association between malaria and environmental variables 32

2.4 Discussion 34

2.5 Conclusion 40

2.6 Supplemental material 41

3 Identifying environmental risk factors and mapping the distribution of West Nile virus in an endemic region of North America 45

3.1 Introduction 47

3.2 Methods 51

3.2.1 West Nile virus case and control data 52

3.2.2 Exploratory analysis 53

3.2.3 Environmental variables 54

3.2.4 Statistical modelling 58

3.3 Results 63

3.3.1 Spatial patterns 63

3.3.2 Environmental data analysis 67

3.4 Discussion 69

3.5 Acknowledgments 74

4 Comparing satellite data and *in-situ* measurements in studying vector mosquito habitats in an urban environment 75

4.1 Introduction 77

4.2 Materials and Methods 79

4.2.1 Study area 79

4.2.2 Mosquito trapping and processing 81

4.2.3 Environmental data collection 81

4.2.4 Analysis 88

4.3 Results 89

4.4 Discussion 98

5 Synthesis 104

5.1 Summary 104

5.2 General considerations 107

5.2.1 Remotely-sensed variables influencing disease risk 108

5.2.2 The importance of landscape context 110

5.2.3 The case for remote sensing data in disease risk applications 111

5.3 Conclusion and future outlook 112

Bibliography 116

List of Tables

Table 2-1: List of static and dynamic variables used to predict malaria proportion..... 21

Table 2-2: Total malaria case numbers, as well as case number by species, from 2014-2017. The number of traveled cases represents all diagnosed malaria cases that indicated a travel history. Outpatient numbers were the total numbers of patients seeking care at health facilities. Outpatient numbers were used to calculate the proportion of outpatients diagnosed with malaria (the malaria proportion). 28

Table 2-3: Mecha and Bahir Dar Zuryia variable importance measures. Shown are all variables used to train a Boosted Regression Tree Model, and their variable importance measure (%). Variables are ranked in descending order, according to their contribution in fitting the model. 44

Table 3-1: Environmental datasets used as explanatory variables. 55

Table 3-2: List of environmental variables and their abbreviations. 56

Table 4-1: Dynamic variables summarized for the three temporal data sources: microclimate data, weather station data, satellite data 83

Table 4-2: Mosquito abundances by species for 2019 and 2020. Displayed are abundances for all species, as well as their relative abundance. 90

Table 4-3: Comparison of negative binomial fixed effect models for each variable type individually: Landcover data from NLCD data set, microclimate loggers, Mesonet weather station data, and satellite time series data. 94

Table 4-4: Comparison of negative binomial fixed effect models for each temporal variable type in combination with land cover data 95

List of Figures

Figure 2-1: Study area in the Amhara region of Ethiopia. The left map shows the two study areas within the greater geographic area. The right map shows the two study areas and the topography of the area. Blue lines indicate woredas (district) boundaries, green lines indicate the kebeles within those woredas. 18

Figure 2-2: Maps displaying the variation in environmental conditions within the two study areas. Variables include NDVI for rainy season (NDVI_rn), NDVI for transition season (NDVI_tr), percent woody vegetation (WOODY), NDMI for dry season (NDMI_dr), NDMI for transition season (NDMI_tr), land surface temperature during transition season (LST_tr), maximum settlement index (SETMX), mean settlement index (SETMN), as well as precipitation during rainy season (PREC_rn). Note that the positions of the two study areas do not represent their true proximity. See the map in Figure 2-1 for their actual locations. 26

Figure 2-3: Total malaria case numbers from Jan 2014 – Dec 2017 by study area. a) Monthly case numbers shown over entire study period. Lines indicate LOESS smoothed trend. b) Seasonality of total malaria cases. Lines represent mean case numbers over all kebeles and years. Error bars indicate the standard error of the observed mean. 29

Figure 2-4: Spatial distribution of malaria proportion. Malaria proportion is the number out of 1,000 outpatients that were diagnosed with malaria for total malaria cases, P. falciparum + mixed cases, and P. vivax cases. The upper maps are the woredas Mecha & Bahir Dar Zuria, and the lower maps are the woredas Aneded & Awabel. Blue dots indicate the travel proportion, which is the total malaria patients with travel history per 1,000 outpatients. 30

Figure 2-5: Number of years each kebele was identified as part of a case cluster by the SaTScan analyses. 31

Figure 2-6: Variable importance of the five most important environmental variables as identified by boosted regression tree models. Different colored bars represent different malaria species (green = total species, dark blue = P. falciparum and mixed infections, light blue = P. vivax). Curves represent the fitted partial dependence curve for each variable. 34

Figure 2-7: Major land cover classes and hillshade in Mecha and Bahir Dar Zuria. Settlement density is derived based on a classification of buildings via high-resolution PlanetScope imagery. A settlement density value above 0.1 can be considered an agglomeration of buildings. All other land cover classes are taken from Midekisa et al. (Midekisa et al., 2014) 42

Figure 2-8: Major land cover classes and hillshade in Aneded and Awabel. Settlement density is derived based on a classification of buildings via high-resolution

PlanetScope imagery. A settlement density value above 0.1 can be considered an agglomeration of buildings. All other land cover classes are taken from Midekisa et al. (Midekisa et al., 2014) 43

Figure 3-1: Workflow for West Nile virus (WNV) modeling and risk mapping. The analysis is divided into four major sections. The processing of case and control data, an exploratory analysis of case and control data, the processing of environmental variables, and the statistical modeling to produce WNV risk maps. Bold text represents the four main sections; italic text represents data processing steps, boxes, and arrows represent the main data flows. BRT = boosted regression tree..... 52

Figure 3-2: . Environmental variables mapped across South Dakota. (a) Elevation, (b) 14-year standard deviation of relative humidity in July/August, (c) 14-year mean for relative humidity in July/August, (d) 14-year standard deviation of tasseled cap wetness in May/June, (e) percent pasture, (f) 14-year standard deviation of precipitation in May/June, (g) 14-year mean of precipitation for July/August, (h) ponding frequency. 60

Figure 3-3: (a) Spatially smoothed West Nile virus (WNV) case points, (b) population weighted control points, and (c) incidence rates of human WNV disease in South Dakota for 2004-2017. 62

Figure 3-4: West Nile virus risk patterns based on three different environmental datasets: (a) land cover and physiography data, (b) climate data, and (c) spectral data..... 64

Figure 3-5: West Nile virus risk patterns based on BRT model predictions using all available environmental variables..... 66

Figure 3-6: Relative importance of environmental variables in predicting the relative risk of West Nile virus. We used four sets of environmental variables to fit a BRT model: (a) land cover and physiography data, (b) climate data, (c) spectral data, and (d) all datasets combined. Explanations of the variable codes are provided in Table 3-2. 68

Figure 3-7: Boosted regression tree (BRT) partial dependence of risk on the top ranked environmental variables from the combined model containing all data sources. Black lines represent the fitted function for environmental variables and relative West Nile virus risk. Blue lines represent a smoothed approximation. Variables are explained in Table 3-2..... 69

Figure 4-1: Land cover and location of trap sites within the study area with the Mesonet weather station (triangle), as well as the proportion of mosquito species collected at each trap site (pie charts). The size of each circle is proportional to the total mosquito abundance. Land cover data are derived from the 2016 NLCD dataset. 80

Figure 4-2: Environmental variables for temperature, spectral indices, precipitation, and relative humidity from microclimate loggers (top row), satellite remote sensing (middle row), and weather station data (bottom row). For microclimate data, the solid lines represent the median value and ribbons the minimum and maximum across all sites. For satellite data, the graphs show daily interpolated value for a 1km buffer zone around all trap sites with the solid lines representing the median value and the ribbons the minimum and maximum across all sites. For weather station data, the solid lines show the daily values measured at the location of the weather station. 87

Figure 4-3: Mosquito counts per trap date for *Ae. albopictus* and *Cx. quinquefasciatus*. Box plots display the distribution of counts across trap sites for each trap date. . 92

Figure 4-4: Coefficient estimates of multivariate models for each data type (spectral indices, landcover, precipitation, relative humidity, temperature), and species (Ae.albopictus and Cx. quinquefasciatus) 93

Figure 4-5: Fitted values for all models including models based on land cover data, microclimate, satellite, and weather station data for *Ae. albopictus* and *Cx. quinquefasciatus*. Model based on microclimate, satellite, and weather station data were fitted with land cover data (with LC) and without land cover data (no LC). Different lines represent spatial variation among trap sites and each line represents variation through time. Blue lines represent fitted values for 2020, green lines for 2019. The grey shaded areas represent the range of mosquito counts across all traps for each year. 97

Abstract

Despite global intervention efforts, mosquito-borne diseases remain a major public health concern in many parts of the world. New strategies to target interventions rely on spatially explicit information about disease transmission risk. Because the transmission of mosquito borne diseases is influenced by environmental conditions, environmental data are often used to predict disease risk. However, the relationships between environmental conditions and such diseases are not homogeneous across different landscapes and requires a context-dependent understanding. The research presented in this dissertation consists of three case studies that used remote sensing data to identify environmental risk factors for mosquito-borne diseases in different geographic settings.

In the first project, the distribution of malaria cases in two study areas in the Amhara region of Ethiopia was analyzed with the help of remote sensing data on land surface temperature, precipitation, spectral indices, as well as land cover and water availability. Environmental variables were derived from remote sensing data and their relationships with spatial and temporal patterns of malaria occurrence were investigated. Settlement structure played an important role in malaria occurrence in both study areas. Climatic factors were also important, with relative risk following a precipitation gradient in the area by lake Tana and a temperature gradient along the Blue Nile River escarpment. This research suggests that studies aiming to understand malaria-environmental relationships should be geographically context specific so they can account for such differences.

Second, the spatial distribution of West Nile virus (WNV) risk in South Dakota was studied via different geospatial environmental datasets. We compared the effectiveness of 1) land

cover and physiography data, 2) climate data, and 3) spectral data for mapping the risk of WNV transmission. The combination of all data sources resulted in the most accurate predictions. Elevation, late season (July/August) humidity, and early-season (May/June) surface moisture were the most important predictors of disease distribution. Indices that quantified inter-annual variability of climatic conditions and land surface moisture were better predictors than inter-annual means. These results suggest that combining measures of inter-annual environmental variability with static land cover and physiography variables can help to improve spatial predictions of arbovirus transmission risk.

Third, mosquito populations in Norman, Oklahoma, were analyzed to investigate the influences of land cover and microclimate on the abundance of vector mosquitoes in a heterogeneous urban environment. Remotely-sensed variables, microclimate measurements, and weather station data were used to study patterns of mosquito abundances. Spatial distributions of the two vector species *Ae. albopictus* and *Cx. quinquefasciatus* were strongly associated with land cover variables. Impervious surface area positively affected the abundance of both species. Canopy cover was positively associated with the abundance of *Cx. quinquefasciatus* but negatively with *Ae. albopictus* abundance. Among all models based on time-varying environmental data, those based on remotely-sensed variables performed best in predicting species abundances. Abundances of both species were positively associated with high temperature and high relative humidity on the trap day, but negatively associated with precipitation two weeks prior to trapping. These results emphasize the great potential for including satellite imagery in habitat analyses of different vector mosquitoes.

The results presented in this dissertation contribute to the understanding of how land cover and geographic context influence the transmission of mosquito-borne diseases. Particularly remote sensing variables capturing static land cover conditions and dynamic measures of vegetation greenness and moisture can explain spatial variation in disease transmission. as well as vector mosquito distribution. Whereas remotely sensed climatic variables like temperature and precipitation influenced gradients in malaria cases at a regional scale, they explained mostly seasonal variation in mosquito distribution at a city scale. Over-all, freely available remote sensing data can help us understand the environmental determinants of disease distribution and can be a valuable tool for predicting disease dynamics on a landscape scale.

1 Introduction

Mosquito-borne diseases are of global relevance and a great concern among health officials and the public. Particularly important diseases include malaria, which infects more than 200 million people every year (WHO, 2020), and dengue which causes over 90 million symptomatic cases annually (Bhatt et al., 2013). The United Nations Sustainable Development Goals (SDGs) include goals on combatting diseases and on ending epidemics of malaria and neglected tropical diseases by 2030 (United Nations, 2015). However, global financial resources to achieve these goals are limited (Pigott et al., 2012), making efficient use of resources through scientifically informed policy essential. To control these diseases effectively and efficiently, we need to be able to predict spatial and temporal patterns of disease risk and mosquito abundances. Understanding how environmental conditions affect disease transmission can help us map disease risk and forecast disease outbreaks. In recent years, the use of satellite imagery has proven to be helpful in this context, as it allows us to retrieve measurements that can be related to environmental conditions on the earth surface that are relevant for the life cycle of mosquito-borne diseases. For example, optical satellite imagery can provide information on land cover, thermal imagery can be used to measure land surface temperature, and radar measurements can give information on precipitation and soil moisture. These data can help researchers address challenges that still persist in efforts to control, eliminate and eradicate mosquito-borne diseases.

1.1 Challenges in combatting mosquito-borne diseases

There are multiple challenges in preventing mosquito-borne diseases across the globe. Each mosquito-transmitted disease has its own unique characteristics that pose barriers to prevention, control, and elimination. Malaria for example has seen large reductions in cases, but remains the deadliest mosquito-borne disease globally. Global annual deaths have gone down by more than half, from 839,000 in 2000 (World Health Organization, 2016) to 409,000 in 2019 (WHO, 2020). This reduction is undoubtedly a success story. However, progress towards eradication has slowed in recent years as new challenges arise (Feachem et al., 2010; Rosenthal et al., 2019; World Health Organization, 2016). A major challenge is the emergence of drug-resistant parasites and insecticide-resistant malaria vectors, which requires the development of new tools for vector control and drugs for treatment and prevention (Rosenthal et al., 2019). Some challenges in reducing the burden even further are due to the changed transmission setting that has resulted from successful interventions. For example, successful reduction of malaria cases caused by the pathogen *Plasmodium falciparum* has led to an increasing share of *Plasmodium vivax* cases, which have a much higher rate of asymptomatic infection and a different life cycle, and are therefore more challenging to study and control (Cotter et al., 2013; Howes et al., 2016). In some locations, successful use of indoor residual spraying and insecticide treated bed nets is altering the risk demographic from children who primarily get infected indoors while asleep to adult men who get infected during outdoor activities (Cotter et al., 2013). This shift of transmission risk from indoors to outdoors makes exposure to malaria dependent on land use activities linked to different types of land cover (Hoffman-Hall et al., 2020).

These changes in disease transmission settings could lead to a shift in areas of highest malaria transmission risk. Different thermal responses of *P. vivax* and *P. falciparum* could influence spatial patterns of malaria transmission. Similarly, the increased relevance of outdoor infections could lead to landscape configurations playing a larger role in driving disease risk than previously.

Also globally relevant are arboviral diseases transmitted by mosquito species in the genus *Aedes*. *Ae. aegypti* and *Ae. albopictus* in particular have increased their global distributions considerably in the past decades (Johnson et al., 2017; Kraemer et al., 2019). Both species originate from the Asian continent and have spread globally. *Ae. aegypti* was introduced to the Americas 350 years ago (Tabachnick and Powell, 1979), and *Ae. albopictus* through international trade of tires in the 1970s and 1980s (Reiter and Sprenger, 1987). A combination of global trade and human movement and a changing climate that increases climatic suitability in formerly unsuitable areas leads to continued expansion of the range of these invasive species (Benedict et al., 2007; Tatem et al., 2006). These range expansions of vector mosquitoes ultimately lead to exposure of new populations to mosquito-transmitted pathogens and disease emergence in new areas. Recent research has focused on predicting the possible future range of these two highly aggressive invaders (Kraemer et al., 2019; Metelmann et al., 2019; Ryan et al., 2019) and the diseases they carry (Fischer et al., 2013; Leta et al., 2018; Lwande et al., 2019). However, to effectively predict future areas under risk, we not only need to understand the influence of a changing climate, but also learn what role landscape ecology plays in this process.

In the United States of America West Nile virus (WNV) is the mosquito-borne disease that causes most cases of human disease. It arrived in New York City in 1999 (Jones, 1999),

likely introduced through air travel (Reisen, 2013), and has since spread across the continent as the climate was suitable for transmission and susceptible bird populations and mosquito vectors were present (Kramer et al., 2019; Peterson et al., 2003; Reisen, 2019, 2013). One of the major challenges in reducing the burden of West Nile virus cases is the fact that the disease transmission cycle of this zoonotic disease is influenced by the interplay between host bird and vector mosquito populations, as well as that of mosquitos and humans, who are considered dead-end hosts. All these interactions are dependent on the environmental conditions and create disease risk patterns that vary across the landscape.

As every disease has its own unique transmission cycle, we face unique challenges in addressing each of them. The common ground is that there are many ecological intricacies in how each disease manifests on a landscape scale. This makes it challenging to predict when and where the risk of mosquito-borne disease transmission will be high.

1.2 Ecology of mosquito-borne diseases on a landscape scale

To understand better how remote sensing technologies can help us address scientific challenges regarding the spatial distribution of mosquito-borne diseases, we first have to understand the environmental factors that contribute to the transmission of mosquito-borne diseases. For mosquitoes to cause disease outbreaks, the environmental conditions need to be suitable for mosquitoes to breed, become infected, live long enough to be infectious, and finally transmit the virus to a susceptible host.

For successful breeding, suitable habitat needs to be available. For certain mosquito species, such as anopheline vectors of malaria and culicine vectors of West Nile virus, these breeding sites include ephemeral waterbodies (Walker and Lynch, 2007; Walton et

al., 1990). Mosquitoes like *Ae. aegypti* and *Ae. albopictus*, which are vectors of several diseases including dengue, chikungunya and Zika, are container breeders and are mostly found in close proximity to human populations (McBride et al., 2014). Rearing success of mosquitoes is also environmentally driven, as water temperature affects pupal and larval development rates (Mamai et al., 2018) and humidity affects the likelihood of larval survival and adult emergence (Murdock et al., 2017). Adult life history traits are known to be influenced by humidity and temperature as well, since humidity is known to influence mosquito activity (Rowley and Graham, 1968) and mosquito population dynamics (Yamana and Eltahir, 2013), and temperature influences multiple life history traits, such as biting rate, relative fecundity and lifespan (Ciota et al., 2014; Mordecai et al., 2019).

Not only the mosquitoes, but also the disease-causing pathogens are influenced by the environment. Since mosquitoes are small exothermic organisms, pathogen development rates within the mosquito are highly dependent on temperature conditions (Mordecai et al., 2017, 2013). Higher temperature leads to a faster development of a pathogen within a mosquito, up to a threshold (Mordecai et al., 2017, 2013). This period of time that determines how long it takes for a pathogen from being ingested by a mosquito to then travel from a mosquito's mid gut into its salivary glands from where it can be injected into a new host is often referred to as the extrinsic incubation time (EIT) or extrinsic incubation period (EIP). The shorter this period is, the sooner a mosquito can infect new humans. This makes it more likely for a mosquito to become infectious before it dies, and it also makes it more likely for it to infect multiple humans within its life span.

In diseases that are not only transmitted between mosquitoes and humans but also involve a zoonotic host or reservoir species, environmental conditions that influence host habitat

also need to be considered. West Nile virus is primarily a bird disease and only causes diseases in humans when so-called “bridge vector” mosquitoes feed on both, birds and humans (Hamer et al., 2009; Molaei et al., 2006). Vector-host contact has shown to be dependent on climatic conditions and land cover. For example, temperature conditions that have shown to be favorable for bird nesting and roosting have shown to also influence WNV transmission (Sallam et al., 2017). The presence of wetlands can be related to host community composition and increases the chances for vector-host contact (Sallam et al., 2017) and therefore also the risk of WNV transmission (Chuang et al., 2012; Ezenwa et al., 2006).

Mosquito-borne diseases also vary within and across seasons, as climatic conditions change throughout the year. Mosquito population dynamics are influenced by fluctuations in sub-seasonal flooding and drying of breeding sites (Bicout et al., 2015), which makes disease transmission equally sensitive to these variations (Chuang and Wimberly, 2012). Seasonal climatic variability has also been shown to influence disease transmission potential (Murdock et al., 2017). These seasonal and sub-seasonal effects on mosquito-borne disease transmission can additionally be affected by land cover.

Land cover influences the climate as built-up areas, vegetated areas and water bodies influence latent and sensible heat fluxes, as well as water flow. The interaction between disease relevant species is heterogeneous across a landscape with varying environmental conditions. Lambin et al. (2010) have coined the term “pathogenic landscapes” to describe how landscape heterogeneity influences spatial variation in disease transmission. Environmental conditions can be a limiting factor for disease transmission on a landscape, as factors like land cover and hydrology can determine whether mosquitoes find suitable

habitat that includes breeding sites to lay eggs or vegetated spots to rest after taking a blood meal. As land cover and physiography also influence microclimate, their interplay with seasonal or even sub-seasonal climatic variation may influence disease risk across a landscape. A next step to advance mosquito habitat and disease risk mapping is to expand studies to measure spatial and temporal patterns of environmental drivers, as there have been few studies to date investigating the influence of climatic variability while taking the landscape composition into account.

1.3 The use of satellite data to study mosquito-borne diseases

Geospatial environmental data help us understand environmental drivers of mosquito-borne diseases across landscapes. The data for such studies can come from different sources, including national survey data, interpolated data from weather stations and more recently also from satellite remote sensing. These can then be combined with mosquito or disease data to predict disease risk across a given landscape (Johnson et al., 2019; Peterson, 2014). Such models have been applied for predictive mapping of vector mosquito species (Uusitalo et al., 2019), and also to map the transmission risk of diseases like malaria (Solano-Villarreal et al., 2019), West Nile virus (Epp et al., 2009), and other environmentally driven infectious diseases (Mayfield et al., 2018).

Satellite data are particularly useful in this context, as they are recorded in a systematic manner and provide reliable and regular observations across the globe, including over areas where environmental data from ground measurement are not available. The decision of governments to make federally funded earth observation data publicly available has shown to be tremendously useful for science supported policy (Zhu et al., 2019) and makes the implementation of low-cost environmental health applications possible.

The application of remotely sensed data for mosquito-borne diseases is not a new concept. It goes back to the first aerial photographs that were used to map mosquito habitats and has gained much more popularity with the launch of the first Landsat satellites in the 1970s (Hay et al., 1997). Today, we have unprecedented access to long-term archives of high quality satellite imagery (Hansen and Loveland, 2012; Wimberly et al., 2021; Wulder et al., 2016). Examples of these freely available high quality long-term archives include the Landsat archive with its 30m resolution imagery at a 16-day return interval that is continuously available since the 1980s, as well as the MODIS sensors that provide daily imagery at a resolution of 250 – 1000m, depending on the wavelength spectrum, and are available since the launch of the Terra satellite in 1999 and the Aqua satellite in 2002. For precipitation data, a major milestone was the Global Precipitation Measurement (GPM) mission and the creation of the Integrated Multi-satellitE Retrievals for GPM (IMERG) dataset, which combines data from the satellite missions TRMM and GPM to provide continuous precipitation data since 1997.

Additionally, an increasing amount of commercially acquired high spatial and high temporal resolution data are available. Just in the last 4 years, the company Planet Labs has launched its fleet of CubeSats and now provides at least daily imagery at approximately 3m spatial resolution (Houborg and McCabe, 2018). Multi-decadal satellite data archives, as well as these new high-resolution data open new doors to a variety of applications and are particularly helpful in tackling current global challenges associated with environmental change, and to answer ecological questions that that could not be answered in the past. The availability of high-quality long-term data archives, as well as new data sources of high spatial resolution, offer new opportunities to use different data sources synergistically and

leverage their unique strengths to study long term, high spatial and high temporal resolution environmental processes.

In recent years, the application of high-quality satellite data that allow researchers to measure environmental conditions over large extents has contributed to great progress towards mapping disease (Benali et al., 2014; Kazansky et al., 2016; Rogers et al., 2002; Wimberly et al., 2021). The use of satellite remote sensing derived environmental data for disease modeling is increasing, and variables such as precipitation, land surface temperature, vegetation indices and other land cover and climatic variables are widely used in studies of mosquito-borne diseases (Hoffman-Hall et al., 2020; Kazansky et al., 2016; Marcantonio et al., 2015; Marti et al., 2020; Shand et al., 2016). However, challenges in identifying and mapping specific environmental drivers of mosquito-borne diseases from satellite earth observation data still remain. For example, even high-resolution satellite imagery will not suffice to capture breeding sites of mosquitoes with a preference for small waterbodies. Researchers need to develop ways to relate satellite measurements to ecologically meaningful information that can predict when and where disease risk may be high.

1.4 Research overview

The following research aims to improve the understanding of environmental spatio-temporal variability and its influence on mosquito-borne diseases, as well as vector-mosquito distribution. This research assesses ways of using remotely-sensed environmental information and to identifying remotely-sensed variables that are relevant to the landscape ecology of mosquito-borne diseases. This will ultimately help us understand spatio-temporal environmental processes that influence mosquito distribution

and the risk of transmission of mosquito-borne diseases. Within this general objective, the goal of this research is to answer the following research questions:

1. What remotely-sensed variables are most associated with the distribution of mosquito-borne diseases?
2. How do these associations depend on the geographic setting?
3. How do remotely-sensed environmental variables compare to other environmental data sources in understanding disease transmission patterns?

The following research is structured in three research chapters and a final summary of the main findings, as well as a general conclusion. Chapters 2, 3, and 4 are standalone manuscripts of which chapter 2 and 3 are published in peer-reviewed journals, and chapter 4 is ready for submission. Each of those chapters showcases a specific use case of remote sensing technology application and presents research to address unique challenges of different mosquito-borne diseases, their ecology, the geographic context, and different ways that public health data are reported and stored. Chapter 2 investigates the association of remotely-sensed static and dynamic environmental variables and malaria occurrence in two geographically distinct areas in Ethiopia. Chapter 3 explores the use of different geospatial environmental data for predicting spatial patterns of West Nile virus disease risk in South Dakota, USA. Chapter 4 compares remotely-sensed environmental variables with other environmental data sources to study the abundance of vector mosquitoes in Norman, Oklahoma. Chapter 5 provides a synthesis of research findings and draws general conclusions.

2 Remote sensing of environmental risk factors for malaria in different geographic contexts

Abstract

Despite global intervention efforts, malaria remains a major public health concern in many parts of the world. Understanding geographic variation in malaria patterns and their environmental determinants can support targeting of malaria control and development of elimination strategies.

We used remotely sensed environmental data to analyze the influences of environmental risk factors on malaria cases caused by *Plasmodium falciparum* and *Plasmodium vivax* from 2014-2017 in two geographic settings in Ethiopia. Geospatial datasets were derived from multiple sources and characterized climate, vegetation, land use, topography, and surface water. All data were summarized annually at the sub-district (*kebele*) level for each of the two study areas. We analyzed the associations between environmental data and malaria cases with Boosted Regression Tree (BRT) models.

We found considerable spatial variation in malaria occurrence. Spectral indices related to land cover greenness (NDVI) and moisture (NDWI) showed a negative association with malaria, as the highest malaria rates were found in landscapes with low vegetation cover and moisture during the months that follow the rainy season. Climatic factors, including precipitation and land surface temperature, had positive associations with malaria. Settlement structure also played an important role, with different effects in the two study areas. Variables related to surface water, such as irrigated agriculture, wetlands, seasonally

flooded waterbodies, and height above nearest drainage did not have strong influences on malaria proportion.

We found different relationships between malaria and environmental conditions in two geographically distinctive areas. These results emphasize that studies of malaria-environmental relationships and predictive models of malaria occurrence should be context specific to account for such differences.

2.1 Introduction

According to the United Nations Sustainable Development Goals (SDGs), combatting diseases, including mosquito borne diseases such as malaria, is a high priority. In particular, malaria is the focus of ongoing efforts toward control and elimination (Dhiman, 2019; Newby et al., 2016; Talapko et al., 2019; Tanner et al., 2015). There has been significant progress in reducing the burden of malaria, but it remains a major public health concern with 229 million malaria cases and 409,000 malaria deaths globally in 2019 (WHO, 2020). The goal is to reduce these numbers by enabling access to prevention, diagnostic testing and treatment for all people (United Nations, 2013). However, global funding to achieve these goals is limited (Patouillard et al., 2017; Pigott et al., 2012). It is essential to use available resources efficiently by spatially targeting prevention, control, and elimination efforts. Therefore, identifying areas with high risk of disease transmission is crucial when responding to disease outbreaks, and knowing what environmental factors drive spatial and temporal patterns of disease risk can aid in identifying disease hotspots (Bousema et al., 2012; Cohen et al., 2017; Hay et al., 2013; Pigott et al., 2015).

Mosquito borne diseases, such as malaria, are highly sensitive to environmental conditions. Mosquito life history traits like longevity, fecundity, and biting rates are highly influenced by temperature (Mordecai et al., 2019) and humidity (Lunde et al., 2013; Yamana and Eltahir, 2013). The rate of pathogen development inside the mosquito (extrinsic incubation period) and transmission probabilities between human and mosquito are also influenced by temperature (Mordecai et al., 2019). Larval habitats, and therefore mosquito abundance, are influenced by land cover (Debebe et al., 2018; Munga et al., 2006), hydrological setting (Hardy et al., 2013) and water management for irrigation (Kibret et al., 2014). All of these factors determine the *receptivity* of an ecosystem, defined as its potential to support vector mosquitoes and malaria transmission cycles (WHO and World Health Organization, 2017). Receptivity is also affected by the *susceptibility* of humans to malaria, which depends on their exposure to mosquito bites and access to resources for malaria prevention and treatment. In addition, the *vulnerability* of a location to malaria is defined as the rate at which parasites are imported through movement of humans or mosquitoes from endemic areas (WHO and World Health Organization, 2017).

The increasing availability of very-high resolution satellite data and long-term satellite records provide new opportunities for malaria research (Wimberly et al., 2021). The use of geospatial environmental data to study the risk of diseases, including those transmitted by mosquitoes, has greatly expanded in recent decades (Kazansky et al., 2016; Parselia et al., 2019). Satellite imagery enables the continuous monitoring of environmental conditions over large areas. The large number of available sensors allows us to measure a wide range of environmental factors that influence malaria receptivity, including meteorological factors such as temperature, humidity, and precipitation as well as landscape features

related to land use, vegetation, surface water, and terrain (Wimberly et al., 2021). Long-term records from moderate resolution sensors such as MODIS (1 km spatial resolution) have been used to study the influences of climate variation on seasonal and interannual variations in malaria (Davis et al., 2019; Sewe et al., 2017). Data from high-resolution sensors such as Landsat (30 m spatial resolution) have been used to assess the effects of water, land cover, and land use on spatial patterns of mosquito vectors (Adeola et al., 2017) and malaria cases (Bui et al., 2019; Midekisa et al., 2014; Rakotoarison et al., 2020). Very-high resolution imagery from commercial satellites such as GeoEye (0.5 m spatial resolution) has been used to map individual households (Shields et al., 2016), and SPOT imagery (1.5 m spatial resolution) has been used to map larval habitats (Kabaria et al., 2016) in support of malaria research and control efforts. Thus, remote sensing is a useful tool for studying the effects of environmental conditions on mosquito borne diseases like malaria (Davis et al., 2019; Ebhuoma and Gebreslasie, 2016; Machault et al., 2011; Midekisa et al., 2012; Solano-Villarreal et al., 2019). In Ethiopia, such studies have established relationships between malaria risk and remotely-sensed environmental factors, such as land surface temperature (Endo and Eltahir, 2020; Midekisa et al., 2012; Minale and Alemu, 2018), precipitation (Belay et al., 2017; Minale and Alemu, 2018), greenness and moisture indices (Midekisa et al., 2012), wetland cover (Midekisa et al., 2014), and distance to water bodies (Minale and Alemu, 2018).

Identifying the environmental risk factors for malaria is challenging because malaria-environment relationships are heterogeneous across different geographic settings (Stresman, 2010). For example, the influence of rainfall on malaria cases in the Amazon varied between wetlands and upland regions (Olson et al., 2009), and the seasonal effects

of rainfall on malaria vector larvae in Tanzania depended on the type of water body and its geomorphological setting (Hardy et al., 2013). In the Amhara region of Ethiopia, yearly fluctuations in malaria cases were influenced more by temperature in wetter climates and by precipitation and other variables sensitive to soil moisture in drier locations (Midekisa et al., 2015). The accuracy of climate-driven models of malaria can be improved by identifying groups of locations that have similar climate sensitivities data and fitting a separate model for each group (Davis et al., 2019).

Malaria-environment relationships are also sensitive to spatial scale. In Ethiopia, spatial analyses of malaria often use case data that are aggregated by district (*woreda*) (Lankir et al., 2020; Lemma, 2020), including several previous studies relating malaria cases to environmental risk factors (Davis et al., 2019; Merkord et al., 2017). However, woredas vary considerably in size (from < 1 km^2 in cities to > 10,000 km^2 in some rural areas) and encompass varied climate, topography, and population density. Research conducted at the individual household level shows that overall malaria prevalence varies considerably between households (Esayas et al., 2020; Zemene et al., 2018), as do the distributions of *P. falciparum* and *P. vivax* (Yalew et al., 2017). However, these household-based studies are typically limited to a few villages or to a sample of households across a larger region.

The goal of this study was to assess the potential for using satellite remote sensing data to quantify malaria receptivity across two study sites in Ethiopia. We considered a wide range of variables derived from moderate, high, and very high-resolution satellite imagery that measured climate variation, land cover and land use, surface water dynamics, and human settlements. The study used a malaria surveillance dataset that was aggregated at the scale of sub-district administrative units called *kebeles.* These are the smallest administrative

units in Ethiopia and are important for the planning and implementation of malaria interventions. Previous research has shown that there is considerable kebele-level variation of malaria prevalence within the same district in the Amhara (Scott et al., 2016) and Oromia (Zemene et al., 2018) regions of Ethiopia. Specific objectives were to identify the remotely sensed variables that were most effective at predicting spatial and interannual variability of malaria cases at the kebele level and to determine whether these relationships varied between two distinctive geographic settings.

2.2 Methods

2.2.1 Study area

Our study area covers four woredas in the Amhara region of Ethiopia that were selected because they were part of a malaria elimination demonstration project to enhance surveillance at sub-district levels (Figure 2-1). The Amhara region is located in northwestern Ethiopia and is geographically very diverse. Elevation ranges from 500 m to 4500 m. Climate is very seasonal with a pronounced dry season (January – April), rainy season (May – August) and a transition season (September – December). Average monthly rainfalls in Ethiopia ranges from 15 mm in January to 200 mm in July (Lemma et al., 2019). The four woredas are located in two separate areas with distinctive environmental characteristics. To explore how different environment-malaria relationships depended on geographic context, we studied each area individually.

The first study area, encompassing Mecha and Bahir Dar Zuria, is located south of Lake Tana and the city of Bahir Dar in central Amhara. Elevations range from 1,500 m to 3,200 m, mean daily temperatures range from 16 °C to 26 °C, and annual precipitation is between

1,000 mm and 2,000 mm. The northern section is relatively flat with large expanses of mixed agriculture and extensive wetlands, transitioning to the foothills of the Choke Mountains in the southern section. This study area contains a substantial area of irrigated agriculture within the Koga irrigation development project. The Koga project encompasses two dams with respective water detention ponds and water storage reservoirs. It irrigates 7,004 ha of farmland via a system of canals (Kassie et al., 2018).

The second study area, encompassing Aneded and Awabel, is located on the southern edge of the Amhara region between the Choke Mountains to the north and the Blue Nile River to the south. Elevations range from 1,000 m to 3,600 m, mean daily temperatures range from 15 °C to 28 °C, and annual precipitation is between 900 mm and 1300 mm. This northern section has gradually sloping terrain and is covered by a mosaic of croplands and wetlands, while the southern section drops steeply into the Blue Nile Gorge and has some croplands with large areas of bare soil and sparse vegetation. There is no large-scale irrigation scheme in this area.

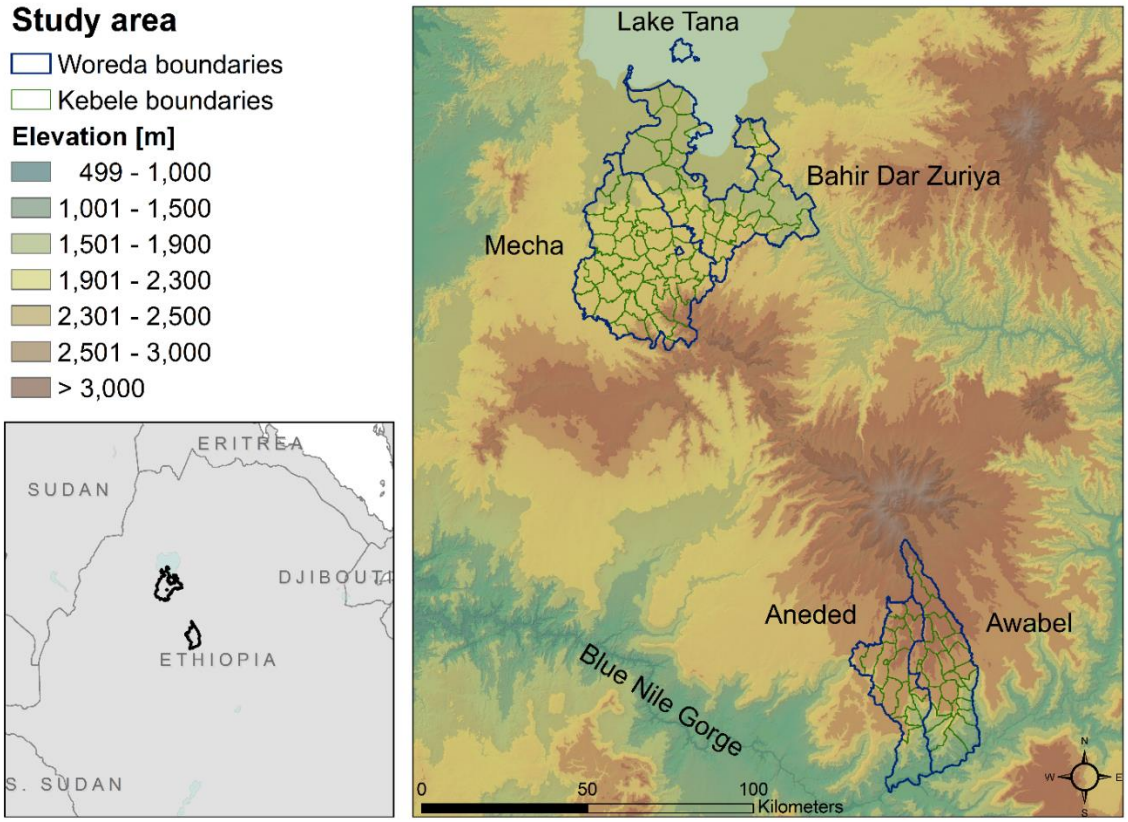

Figure 2-1: Study area in the Amhara region of Ethiopia. The left map shows the two study areas within the greater geographic area. The right map shows the two study areas and the topography of the area. Blue lines indicate woredas (district) boundaries, green lines indicate the kebeles within those woredas.

The Amhara region has unstable malaria transmission that results in sporadic localized and regional outbreaks (Abeku et al., 2002). Outbreaks occur primarily between September and December. The malaria parasites in this region are *Plasmodium falciparum* and *Plasmodium vivax* (Yalew et al., 2017) and the primary vector species in this region is the mosquito *Anopheles arabiensis* (Animut et al., 2013). The Ethiopian government is committed to move towards nationwide malaria elimination by 2030 (Federal Ministry of Health, 2020). Due to successes of malaria control efforts, malaria incidence and mortality rates have been declining in the Amhara region (Lankir et al., 2020) and in other parts of

Ethiopia (WHO, 2020). In Ethiopia, malaria surveillance data are usually reported from health facilities and then aggregated by woredas. As part of a malaria elimination pilot project, malaria data from the four woredas were tracked at the kebele level from 2013-2018. Kebeles are the smallest administrative unit in Ethiopia, with areas of kebeles for the entire country ranging from 0.1 km^2 – 9,500 km^2. The average size is 50 km^2 with a population of typically a few thousand inhabitants (Central Statistical Agency, 2012). Each kebele has at least one health post that serves up to 5,000 people. Larger kebeles have health centers that serve 20,000 people. Our study area includes a total of 122 kebeles from the four study woredas: Mecha, Bahir Dar Zuria, Aneded and Awabel.

2.2.2 Malaria data

Malaria case data were collected at local health posts and health centers, summarized by week for each kebele, and reported to the Amhara Regional Health Bureau. The data included malaria cases of patients who sought care at a health post or health center. As per national malaria guidelines, all suspected malaria cases at all health facilities were confirmed by multi species rapid diagnostic tests (at rural health posts) or Giemsa based microscopy (at hospitals and health centers). Weekly summaries included information on the malaria-causing pathogen (total *P. falciparum* cases, *P. vivax* cases, and mixed infections), age of the patients (above or below 5 years of age), the number of malaria patients with a travel history, and the number of total outpatients seeking care during a given week.

Because no recent population data were available at the kebele level, we calculated the proportion of outpatients diagnosed with malaria (hereafter referred to as malaria proportion). This ratio is considered a reliable indicator of malaria burden because it

controls for temporal variation in health facility attendance, can be calculated in situations where accurate population data are not available (Guintran et al., 2006), and has been used as a measure of malaria burden in previous studies . Although global population datasets like WorldPop (Tatem, 2017) and LandScan (Dobson et al., 2000) are also available, we decided not to use them because we were not able to validate their accuracy at the kebele level within our study area.

The case data ranged from September 2013 to July 2018. However, we only used data from 2014-2017, as these were available for the entire year. Out of 472 reporting health posts or health centers, we did not have reliable location data for six. Together, these six reported only 23 malaria cases for the entire time frame and their exclusion was not expected to influence the results.

Epidemiological data were summarized for all health posts within a kebele and for each year to produce a kebele-wide annual tally of total malaria cases. For treatment and reporting purposes, malaria cases are grouped into two categories: 1) *P. falciparum* plus mixed infections, 2) *P. vivax* only. We additionally summarized total malaria cases. To detect statistically significant spatial clusters in malaria occurrence, we performed a scan statistic using SaTScan software version 9.6 (Kulldorff, 2018). We ran a retrospective purely spatial discrete Poisson model with total outpatients as the population at risk. The scan statistic was performed with an elliptical window for each year and each malaria pathogen group (total malaria cases, *P. falciparum* + mixed cases, and *P. vivax* cases) separately. After determining the spatial clusters for each year, we then identified stable hot spots (areas with recurring clusters in three or four years), unstable hot spots (clusters

in one or two years), and areas that were never identified as clusters. More details on the SaTScan analysis can be found in chapter 2.6.

2.2.3 Environmental variables

Two types of explanatory environmental variables were used to investigate malaria case patterns: dynamic and static variables (Table 2-1). *Dynamic variables* included environmental conditions that were expected to vary between and within years, such as land surface temperature and remotely sensed greenness and moisture indices. *Static* variables included variables related to land cover, land use, and physiography that were not expected to vary substantially from year to year. All satellite data were reprojected to UTM zone 37 before we calculated zonal summaries for each kebele. Maps of summaries of selected variables are shown in Figure 2-2. Additional baseline maps of land cover, settlements, and topography can be found in figures 2.7 and 2.8.

Table 2-1: List of static and dynamic variables used to predict malaria proportion.

Type	Description	Source	Name	Units
Dynamic variables	Daytime temperature	Terra MODIS	LST	°C
	Annual rainfall	IMERG	PREC	mm
	Normalized difference vegetation index	MODIS NBAR	NDVI	Index
	Normalized difference moisture index	MODIS NBAR	NDMI	Index
	Distance to seasonal waterbodies	Landsat OLI	DISTSW	m
Static variables	Settlement mean density	PlanetScope	SETME	Index
	Settlement max density	PlanetScope	SETMX	Index
	Area below 2 m above nearest drainage	DEM, Stream network	HAND	%
	Wetland cover	Midekisa et al. [27]	WETL	%
	Woody vegetation cover	Midekisa et al. [27]	WOODY	%
	Cropland cover	Midekisa et al. [27]	CROP	%
	Open water cover	Midekisa et al. [27]	WATER	%
	Sparse vegetation cover	Midekisa et al. [27]	SPVEG	%
	Irrigation cover	Digitized from Google Earth	IRRI	%

Dynamic variables were summarized for each year for the dry season (_dr), rainy season (_rn), and transition season (_tr)

Temperature data were derived from the MODIS Terra Land Surface Temperature and Emissivity Product (MOD11A2) (Wan et al., 2015). These data have a spatial resolution

of 1km and are provided as 8-day composites. We used daytime observations to reduce the problem of missing data from nighttime clouds. Data were filtered using the quality assurance flags to only include observations with an average LST error of below 2 °K. Temperature values were then converted to °C.

Spectral indices measuring vegetation greenness and surface moisture were derived from daily 500 m spatial resolution MODIS Nadir BRDF-Adjusted Reflectance data (MCD43A4) (Schaaf and Wang, 2015). We applied a data quality filter using the BRDF/Albedo quality product (MCD43A2) and only included observations that were flagged as land and were "good" or "best" quality. We then calculated the Normalized Difference Vegetation Index (NDVI) (Tucker, 1979), as well as a Normalized Difference Moisture Index (NDMI) (Gao, 1996).

We imputed missing values and replaced outliers from imperfect cloud-screening using a robust linear regression model from the R MASS library (Venables and Ripley, 2002). We fitted a robust linear regression model on our temporal data for NDVI, NDMI, and LST using cyclical splines and estimated outliers with a z-score above 3 or below -3. We then removed observations that were identified as outliers and replaced them, as well as missing values, with predicted values from the robust linear regression model.

Precipitation data were derived from the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement product (IMERG) (Huffman et al., 2020). IMERG has a spatial resolution of 0.1° (approximately 11km) and a temporal resolution of 30 minutes. The dynamic variables for land surface temperature, spectral indices, and precipitation were summarized for each kebele and for each season of the year: dry season (January – April),

rainy season (May – August), and the transition season (September – December). For each kebele we then calculated the mean seasonal value of each dynamic variable for each year.

Because distance to potential breeding habitat has been identified as risk factor for anopheline mosquito abundance (Mushinzimana et al., 2006), we mapped seasonally flooded areas. We used 30 m Landsat 8 Operational Land Imager (OLI) imagery for the dry season and the rainy season of each year. We removed pixels that were flagged as cloud or cloud shadow and calculated the Normalized Difference Water Index (NDWI) (McFeeters, 1996) for each Landsat scene. To identify seasonally flooded areas, we extracted areas where the median NDWI during the end of the wet season (September - October) was above zero and the median NDWI during the dry season (January – April) was below zero. We then calculated the distance of each pixel to the nearest seasonally flooded pixel, expressed as cumulative cost distance of the shortest path measured in meters from the neatest water source. This was done for each year separately to account for inter-annual variation in flooding extents. Additionally, we added "year" as a continuous variable, to capture inter-annual trends of malaria proportion that are not related to environmental conditions.

S*tatic variables* measured environmental conditions that were not expected to change substantially between years. These included data on settlement structures, land cover, and the hydrological setting of the landscape. We created two settlement density indices from 3 m spatial resolution PlanetScope imagery (Planet Team, 2017). We acquired images from November 2016 and classified each tile into building and non-building areas with a Random Forest model using a bag fraction of 0.67 over 500 classification trees. Training data consisted of sampling 7,500 points collected from 2,000 manually digitized polygons

in each study area. The overall accuracy of the building classification based on out-of-bag data was 0.98 for both areas. In Mecha and Bahir Dar Zuria we reached a sensitivity of 0.97 and a specificity of 0.98. In Aneded and Awabel sensitivity and specificity were both 0.98. To create variables for settlement density, we resampled the classification to 100 m pixels via a majority filter and then applied a Kernel density estimator using a gaussian kernel with a radius of 100 m and a sigma of 50 m. For each kebele we summarized the mean and maximum settlement density. Settlement classification and density estimation were performed in Google Earth Engine (Gorelick et al., 2017).

To study the impact of different land cover types on sub-district malaria transmission, we acquired land cover maps from Midekisa et al. (Midekisa et al., 2014). This dataset provides Landsat based classifications of major land cover types in the Amhara region at a 30 m resolution. It is locally calibrated, and its overall accuracy is reported at 93.5%. We calculated the percentage of each kebele covered by the following land cover classes: open water, herbaceous wetlands, woody vegetation, cropland, and sparse vegetation.

To map areas that are likely to be flooded during larger rain events, we calculated the height above nearest drainage (HAND). The height above nearest drainage measures the vertical distance between a given point and the nearest stream. Low values indicate floodplains and other low-lying areas that are likely to be inundated during and after the rainy season when flow rates are high. We used topographic and stream network data to calculate the HAND using Topography Tools for ArcGIS (Dilts, 2015, p. 10). We then identified areas less than 2 m above the nearest drainage and calculated the proportion of each kebele that falls into this category.

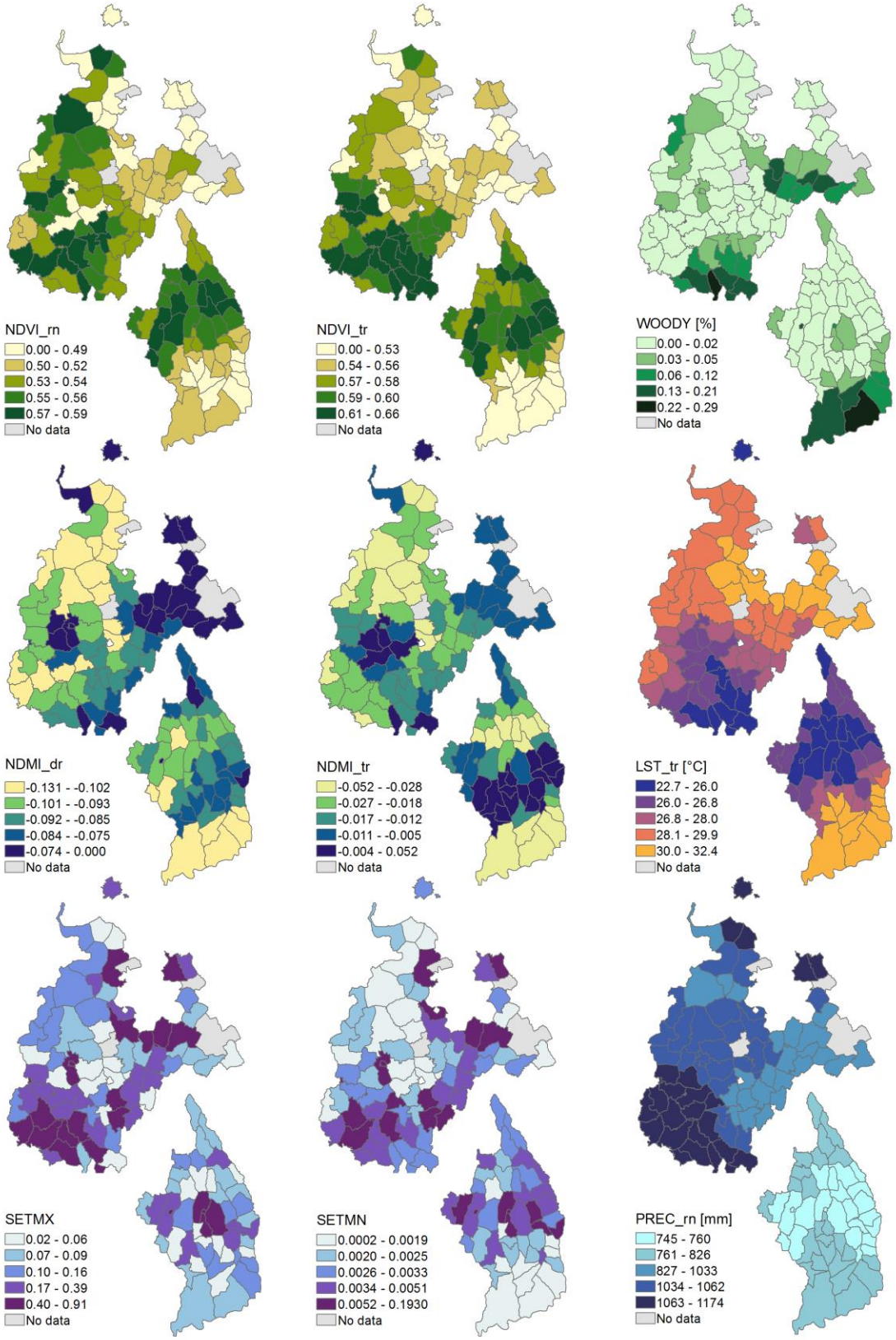
NDVI_rn
0.00 - 0.49
0.50 - 0.52
0.53 - 0.54
0.55 - 0.56
0.57 - 0.59
No data
NDVI_tr
0.00 - 0.53
0.54 - 0.56
0.57 - 0.58
0.59 - 0.60
0.61 - 0.66
No data
WOODY [%]
0.00 - 0.02
0.03 - 0.05
0.06 - 0.12
0.13 - 0.21
0.22 - 0.29
No data
NDMI_dr
-0.131 - -0.102
-0.101 - -0.093
-0.092 - -0.085
-0.084 - -0.075
-0.074 - 0.000
No data
NDMI_tr
-0.052 - -0.028
-0.027 - -0.018
-0.017 - -0.012
-0.011 - -0.005
-0.004 - 0.052
No data
LST_tr [°C]
22.7 - 26.0
26.0 - 26.8
26.8 - 28.0
28.1 - 29.9
30.0 - 32.4
No data
SETMX
0.02 - 0.06
0.07 - 0.09
0.10 - 0.16
0.17 - 0.39
0.40 - 0.91
No data
SETMN
0.0002 - 0.0019
0.0020 - 0.0025
0.0026 - 0.0033
0.0034 - 0.0051
0.0052 - 0.1930
No data
PREC_rn [mm]
745 - 760
761 - 826
827 - 1033
1034 - 1062
1063 - 1174
No data

Figure 2-2: Maps displaying the variation in environmental conditions within the two study areas. Variables include NDVI for rainy season (NDVI_rn), NDVI for transition season (NDVI_tr), percent woody vegetation (WOODY), NDMI for dry season (NDMI_dr), NDMI for transition season (NDMI_tr), land surface temperature during transition season (LST_tr), maximum settlement index (SETMX), mean settlement index (SETMN), as well as precipitation during rainy season (PREC_rn). Note that the positions of the two study areas do not represent their true proximity. See the map in Figure 2-1 for their actual locations.

2.2.4 Statistical analysis

We used Boosted Regression Trees (BRT) to model the relationships between malaria cases and environmental variables. BRT models are used in species distribution modeling (Elith et al., 2008; Elith and Graham, 2009), in epidemiology (Friedman and Meulman, 2003), and to study mosquito borne diseases (Bhatt et al., 2013; Hess et al., 2018; Messina et al., 2019; Solano-Villarreal et al., 2019). Advantages of the BRT method are that it yields high accuracy while requiring little tuning (Friedman and Meulman, 2003), performs well under moderate collinearity of predictor variables (Dormann et al., 2013), and captures non-linear relationships between predictor and response variables, and allows for interactions between predictor variables. Poisson BRT models were created using the R library gbm (Ridgeway et al., 2013).

For each study area, separate models were fitted for *P. falciparum* + mixed cases, *P. vivax* cases, and total cases. Outpatient numbers were included as offset term. The models were fitted with a learning rate of 0.01, a tree complexity of 3, and a bag fraction of 0.5. Several parameter combinations were tested and the combination that yielded the best R^2 from a 5-fold cross validation was selected.

Relative importance of predictor variables was estimated to determine which variables had the strongest influences on malaria patterns. Variable importance is a measure of how often

a variable is used to create a split, normalized by the improvement in squared error resulting from the corresponding split. A high variable importance of a variable indicates that the BRT model relied heavily on a specific variable during the model fitting process. It is a common metric to compare the relative influence of predictor variables (Elith et al., 2008), and has been used to study environmental effects on malaria transmission (Solano-Villarreal et al., 2019). We then ranked the variables by their importance value and identified the five most important variables for total malaria cases. To visualize the relationships between these variables and malaria proportion, we fitted partial dependence plots that show how a response variable depends on the predictor variable after taking the average effects of all other variables into account.

2.3 Results

2.3.1 Spatio-temporal patterns in malaria cases

Between January 2014 and December 2017, a total of 22,584 malaria cases were reported (Table 2-2). Fifty-nine percent of all cases were attributed to *P. falciparum* + mixed, and 41% of cases were due to *P. vivax*. In Bahir Dar Zuria and Mecha, *P. falciparum* + mixed cases made up the largest share of infections, whereas *P. vivax* was more dominant in Aneded and Awabel. Of the four woredas, the highest malaria proportion was in Awabel with 62 malaria cases per 1,000 outpatients. A total of 2,108 cases had a travel history, meaning they left the village for at least one night within the last 30 days. The highest proportion of traveled cases was in Awabel with 6 traveled cases per 1,000 outpatients.

Table 2-2: Total malaria case numbers, as well as case number by species, from 2014-2017. The number of traveled cases represents all diagnosed malaria cases that indicated a travel history. Outpatient numbers were the total numbers of patients seeking care at health facilities. Outpatient numbers were used to calculate the proportion of outpatients diagnosed with malaria (the malaria proportion).

Woredas	Total malaria cases	*P. falciparum* + mixed cases		*P. vivax* cases		Traveled cases		Outpatients
	Count	Count	% of total cases	Count	% of total cases	Count	% of total cases	Count
Aneded	3993	2069	51.8	1924	48.2	201	5.0	79,394
Awabel	6110	3436	56.2	2674	43.8	592	9.6	97,052
Bahir Dar Zuria	6145	3767	61.3	2378	38.7	257	4.1	324,769
Mecha	6336	3975	62.7	2361	37.3	1058	16.7	416,009

From 2014 to 2017, we identified a downward trend in total malaria cases for Mecha and Bahir Dar Zuria (Figure 2-3). Case numbers in Aneded and Awabel remained more stable over the study period. This led to case numbers in Mecha and Bahir Dar Zuria being higher than those in Aneded and Awabel during 2014 and 2015, but lower in 2016 and 2017. Both study areas showed strong seasonality with low case occurrence between January – April, a small case peak during May – August and most cases occurring during September – December.

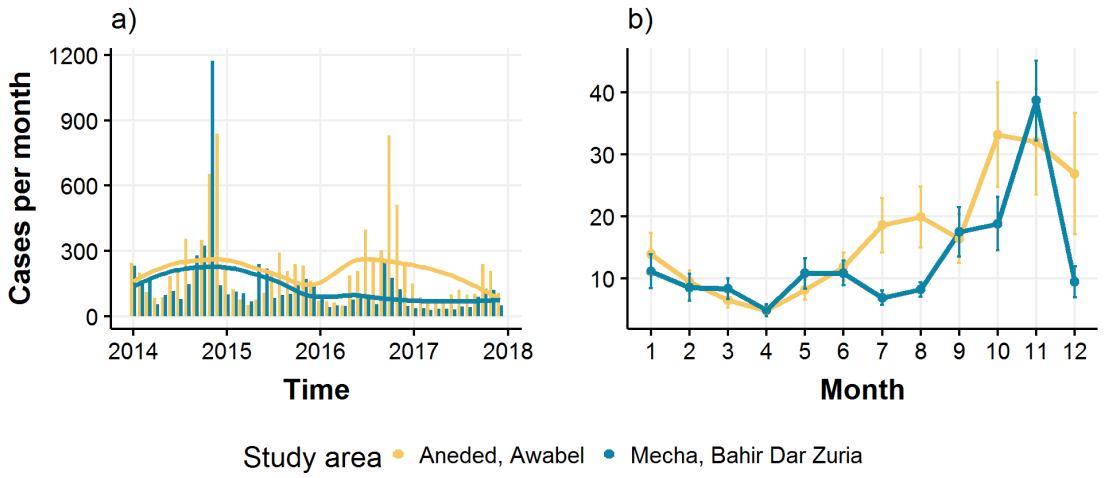

Figure 2-3: Total malaria case numbers from Jan 2014 – Dec 2017 by study area. a) Monthly case numbers shown over entire study period. Lines indicate LOESS smoothed trend. b) Seasonality of total malaria cases. Lines represent mean case numbers over all kebeles and years. Error bars indicate the standard error of the observed mean.

There was considerable variation in total malaria proportion within woredas, as well as for cases broken up into the two pathogen groups *P. falciparum* + mixed and *P. vivax* (Figure 2-4). In Mecha and Bahir Dar Zuria there was a high malaria proportion in the hilly southern part of the study area, as well as in the flatter central area. In Aneded and Awabel there was a much stronger spatial gradient in malaria proportion with southern kebeles on the Blue Nile escarpment showing a higher malaria proportion. In Mecha and Bahir Dar Zuria, *P. vivax* was more confined to the kebeles in the southern hills, whereas *P. falciparum* cases were largely responsible for the central cluster. In Aneded and Awabel, *P. falciparum* malaria proportion was highest in the kebeles along the Blue Nile escarpment in the south, whereas infections from *P. vivax* cases had additional clusters in the northern and western kebeles. Most cases with a travel history were recorded in the southern hills of Mecha and on the Blue Nile escarpment in Aneded and Awabel. High malaria proportion was correlated with a high proportion of cases with a travel history. The Spearman rank

correlation coefficient between travel proportion and *P. falciparum* proportion was 0.6 for Mecha and Bahir Dar Zuria, and 0.62 for Aneded and Awabel. The correlation between travel proportion and *P. vivax* proportion was 0.58 for Mecha and Bahir Dar and 0.54 for Aneded and Awabel.

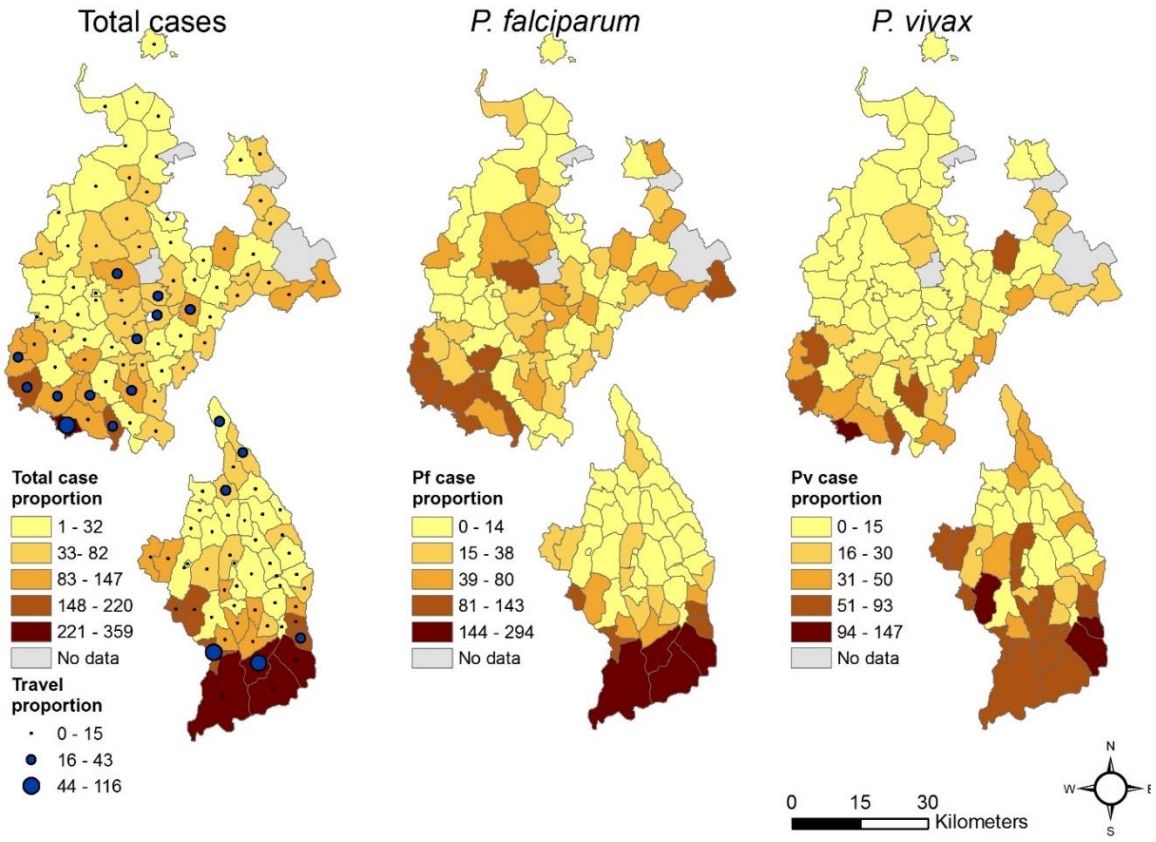

Figure 2-4: Spatial distribution of malaria proportion. Malaria proportion is the number out of 1,000 outpatients that were diagnosed with malaria for total malaria cases, P. falciparum + mixed cases, and P. vivax cases. The upper maps are the woredas Mecha & Bahir Dar Zuria, and the lower maps are the woredas Aneded & Awabel. Blue dots indicate the travel proportion, which is the total malaria patients with travel history per 1,000 outpatients.

These general patterns were consistent with the results from the SaTScan analysis (Figure 2-5). The kebeles in southern and central Mecha and Bahir Dar Zuria that showed a high malaria proportion over the entire time period were identified as stable malaria hotspots with malaria clusters in at least 3 years. Similarly, in Aneded and Awabel, areas with

overall high malaria proportion were found to be stable malaria hotspots with clusters identified for every year. The SaTScan results showed that the high malaria proportion in the southern kebeles of both zones were due to consistent annually recurring outbreaks and not due to individual large outbreaks. Notably, in Aneded and Awabel most kebeles were either identified as stable hotspots for *P. falciparum* with clusters in three or four years, or as areas that were never identified as hotspots. Only very few kebeles were identified as unstable *P. falciparum* hotspots with clusters in one or two years. *P. vivax* case clusters in Aneded and Awabel were less stable with a considerable number of kebeles being identified as clusters in one or two years.

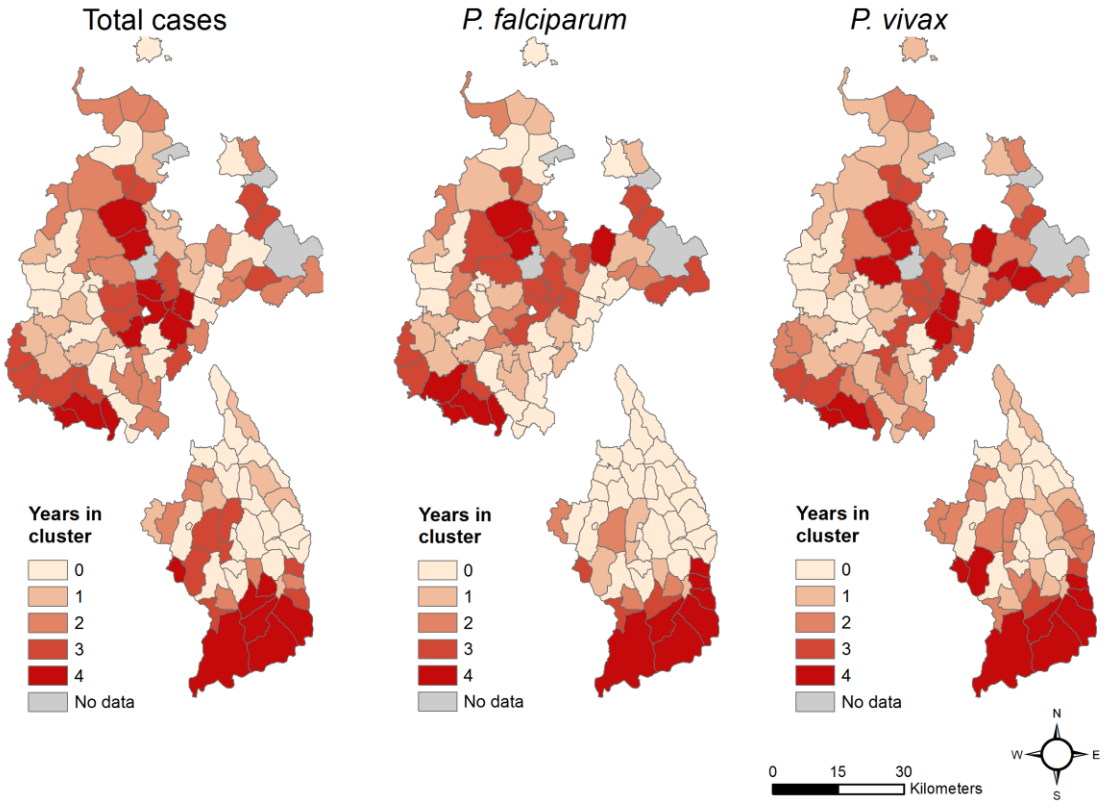

Figure 2-5: Number of years each kebele was identified as part of a case cluster by the SaTScan analyses.

2.3.2 Association between malaria and environmental variables

We compared model performance metrics for BRT models for each study area and malaria species, as well as total cases. The cross-validated R^2 values were higher in Aneded and Awabel with 0.90 for total cases, 0.91 for *P. falciparum* and mixed cases, and 0.78 for *P. vivax* cases. In Mecha and Bahir Dar Zuria, the cross-validated R^2 values were 0.6 for total cases, 0.54 for *P. falciparum* and mixed cases and 0.68 for *P. vivax* cases. In Mecha and Bahir Dar Zuria, the cross-validated R^2 for *P. vivax* was higher than for *P. falciparum*, whereas in Aneded and Awabel the cross-validated R^2 for *P. falciparum* was higher than for *P. vivax*.

We quantified the relative importance of all variables in the different BRT models and visualized a subset of variables that represent the five most important variables for total malaria cases, as well as the respective partial dependence plots to interpret the influence of environmental variables on malaria proportion (Figure 2-6). Tables with all variables and their relative importance scores can be found in the supplemental chapter 2.6. The most important variables differed between the two study areas with only NDVI during the transition season appearing in the top five of both study areas and both malaria species. In both study areas at least one of the two seasonal vegetation, or moisture indices was important, as well as was at least one settlement density index. Except for percent Woody vegetation, none of the land cover variables contributed substantially to any of the models. None of the hydrological variables, including the distance to seasonal water, height above nearest drainage, or percent of land within the Koga irrigation scheme, were among the most important variables.

In Mecha and Bahir Dar Zuria, all variable importance values were below 20%, indicating that the BRT model relied relatively evenly on a larger number of variables and that no variable stood out in explaining most of the variation on its own. *P. falciparum* was most strongly influenced by NDVI during the transition season (10 %), with higher malaria proportion coinciding with NDVI values below 0.5. Similarly, NDMI values below -0.01 during the transition season were also associated with a higher malaria proportion. Precipitation above 1,200 mm during the rainy season was associated with a higher malaria proportion. *P. vivax* in Mecha and Bahir Dar Zuria was mostly influenced by NDVI during transition season (17 %) with a higher malaria proportion related to lower NDVI values, and maximum settlement index (13 %) with a higher malaria proportion in kebeles with the highest aggregations of buildings. Year was an important variable for both malaria pathogens in Mecha and Bahir Dar Zuria, indicating that the downward trend over the entire study period was driven by processes not related to environmental factors included in this model.

In Aneded and Awabel, NDMI during the dry season was by far the most important variable for both malaria species (41 % for *P. falciparum* and 36 % for *P. vivax*). Additionally, NDVI values below 0.55 during transition season were also associated with more malaria for both species (both < 10%). Both spectral indices showed associations of drier or less green conditions with malaria occurrence. Higher malaria proportion was also positively associated with percent cover of woody vegetation for *P. falciparum* (11%), and *P. vivax* (6%). Mean LST above 25°C during transition season had a positive effect on malaria proportion for both species. However, the relative importance of LST for *P. falciparum* and mixed cases (14%) was larger than for *P. vivax* (6%). For both species, a mean

settlement index of close to zero was associated with a higher malaria proportion (both <10%).

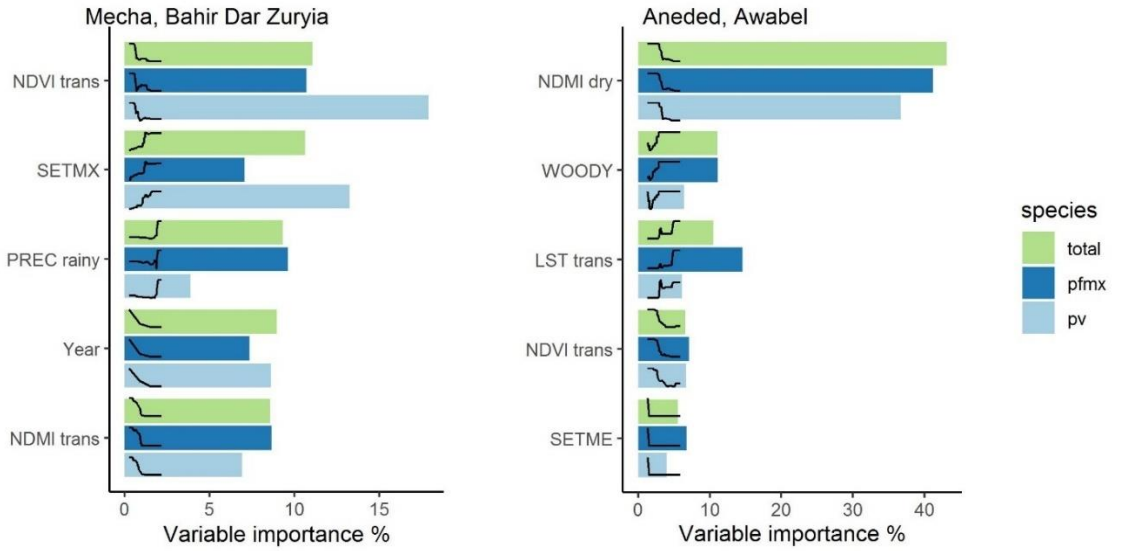

Figure 2-6: Variable importance of the five most important environmental variables as identified by boosted regression tree models. Different colored bars represent different malaria species (green = total species, dark blue = P. falciparum and mixed infections, light blue = P. vivax). Curves represent the fitted partial dependence curve for each variable.

2.4 Discussion

This study presents a kebele-level analysis of malaria occurrence in the Amhara region in Ethiopia between 2014 and 2017. Remotely sensed environmental variables were used to explore the influences of environmental risk factors. We found considerable small-scale variation in the proportion of outpatients with malaria and malaria case clusters among the kebeles of each study area. Climatic variables, settlement structure, and spectral indices were associated with malaria risk, whereas variables related to seasonal waterbodies, flood plains, and irrigation had weaker relationships. However, the relative importance of the variables, as well as the association between those variables and malaria, varied between the two study areas, as well as between the different *Plasmodium* species. These results

support the findings of previous studies that there is geographic variation in the relationships between environmental conditions and malaria .(Davis et al., 2019; Hardy et al., 2013; Midekisa et al., 2015; Olson et al., 2009) and highlights the importance of stratifying disease risk models into zones with similar geographic context when studying large, heterogeneous areas (Davis et al., 2019).

In Mecha and Bahir Dar Zuria, the malaria cases varied along a precipitation gradient, with the highest malaria risk in areas with the most rainfall. This effect is not surprising, as the association of malaria with climatic factors, including rainfall, has been well established in the area (Belay et al., 2017; Midekisa et al., 2012; Minale and Alemu, 2018). The negative associations with NDVI and NDMI values during transition season could be an indicator of available breeding sites for *An. arabiensis*. Areas that received heavy rainfall during the rainy season, but had drier conditions during the transition season, could have more isolated pools as surface water recedes. Such areas have been found to be favorable breeding sites for *An. arabiensis* (Hamza and Rayah, 2016). Additionally, a high vegetation cover can create unfavorable conditions for *An. arabiensis*, which prefers sun-lit pools with low vegetation cover (Gone et al., 2014). The positive influence of maximum settlement index indicates that kebeles where buildings are aggregated into one or more settlements are at higher risk than kebeles where dwellings are dispersed throughout the countryside.

In Aneded and Awabel, the malaria proportion was primarily associated with low NDMI values during the dry season. These dry areas are also located on the Blue Nile escarpment, which is the most isolated and sparsely settled part of this study area. In addition to the mosquito habitat associations discussed for Mecha and Bahir Dar Zuria, the negative influences of NDMI on malaria in Aneded and Awabel could also reflect underlying socio-

economic factors. For example, there is a relationship between the risk of getting infected with malaria and the main source of water of a household, with households that need less time to access water being at lower risk than households that require more time to access water (Ayele et al., 2012). This effect may be amplified in drier environments with less water availability. Additionally, the driest kebeles in this region are sparsely settled, as reflected in the negative association between mean settlement index and malaria. Further, the terrain of the escarpment is steep and rugged and is isolated from major highways and cities. Households on the escarpment may have difficulties in accessing resources known to reduce infection risk, such as information about malaria treatment (Hwang et al., 2010), bed nets (Zerdo et al., 2020), and other malaria prevention tools (Paulander et al., 2009), resulting in higher susceptibility to malaria infections.

Land surface temperature was found to be the second strongest predictor of *P. falciparum* in Aneded and Awabel. However, it was less important in predicting *P. vivax* distribution. Previous research has found that due to the different biology of the two parasites, *P. falciparum* is limited to lower elevations with warmer temperatures, whereas *P. vivax* is more tolerant of lower temperatures and therefore occurs in higher elevations in the Amhara region (Lyon et al., 2017; Yalew et al., 2017). The effects of temperature on *P. falciparum* in Aneded and Awabel may therefore be due to a temperature threshold that is reached in the higher elevations of this study area. Additionally, in other parts of Ethiopia it has been reported that *An. arabiensis* is mostly prevalent in lower elevations whereas *An. cinereus* dominates at higher elevations (W. Lemma et al., 2019). Because the temperature gradient in Aneded and Awabel follows an elevation gradient, change in the dominant

vector species may be an additional explanation of why we see a temperature effect in this region.

In both study areas we found that malaria cases with recorded travel histories were concentrated in certain kebeles, and these kebeles tended to have high *P. falciparum* cases. Travel has been shown to be a significant risk factor for malaria transmission, particularly for *P. falciparum* related infections (Yukich et al., 2013). In rural areas in Amhara seasonal migrants often leave their village to seek an additional income elsewhere (Asfaw et al., 2010). Migrant workers are often found to have increased exposure to malaria risk due to low bed net utilization, as well as sleeping outside (Aschale et al., 2018). Migrant workers from rural areas could therefore be a source of repeated pathogen introduction into their home communities. From the available data we could not tell where and for how long they traveled, nor could we derive which *Plasmodium* species the patients with travel history were infected with. Travel history was therefore not included in our formal analysis. However, these data suggest that vulnerability to imported cases is an important factor that must be considered in addition to local receptivity for malaria transmission. Some of the observed relationships with remotely sensed environmental variables may reflect environmental settings with marginal agriculture where residents are more likely to travel for seasonal employment.

Surface water-related variables did not strongly influence malaria proportion in our model in either of the two study areas. These included the coverage of open water, herbaceous wetlands, percent of land within the Koga irrigation scheme, the average distance to the nearest seasonal waterbody, as well as areas likely to flood due to their height above the nearest drainage. These findings contrast with the general expectation of relationships

between malaria cases and temporary water bodies. Other studies from different parts of Ethiopia have found that woredas with high wetland cover had high malaria incidence (Midekisa et al., 2014) and household surveys showed that the distance to breeding sites for samples of individual households influences malaria risk (Alemu et al., 2011; Belay et al., 2017; Nissen et al., 2017). However, those previous studies have been performed at different spatial scales, and water-related variables were much less important at the kebele level within the woredas that we studied. Additionally, the water variables were derived from 30 m Landsat imagery and a 30 m digital terrain model, and this spatial resolution is unable to capture smaller water bodies that can act as larval habitat.

The presence of the Koga irrigation scheme did not influence malaria cases. The consistent source of water in irrigated areas creates breeding habitat for anopheline mosquitoes, as observed for a comparable large dam project in Ethiopia (Kibret et al., 2017). The lack of an irrigation effect on malaria in our study may be an example of what Ijumba and Lindsay describe as the "paddies paradox" (Ijumba and Lindsay, 2001). There are two possible explanations for this phenomenon in this area. The first factor may be the separation between agricultural land and settlements. The kebeles that are dominated by the Koga irrigation scheme do not have large settlements, as those are located further away in neighboring kebeles that do not have irrigated land cover. Secondly, the Koga irrigation scheme reportedly increased wealth in irrigator households (Kassie et al., 2018). Increased wealth, better access to health care and education may contribute to a low burden of malaria in this region. Despite high receptivity to malaria because of readily available breeding sites, the spatial separation between irrigation and settlements, in addition to improvements of the socio-economic conditions in the area may have reduced susceptibility and

vulnerability of the human population, leading to low malaria cases despite available mosquito habitats.

The utilization of remotely sensed satellite data to study malaria infection patterns comes with limitations. The nature of satellite earth observation data allows us to study the physical properties of the earth surface and measure a variety of factors that influence malaria receptivity. We could not directly measure socio-economic factors like housing, access to health care, and effectiveness of malaria interventions that affect the susceptibility of human populations. In addition, we could not quantify the vulnerability of communities to imported cases. We also could not relate the observed patterns to different species of vector mosquitoes, as entomological data was not available at such a large extent. Particularly in Mecha and Bahir Dar Zuria, the cross-validated R^2 values from 0.54-0.68 indicate that environmental variables alone cannot explain all of the spatial and interannual variation in malaria cases. The importance of non-environmental factors in this study area is also reflected in the strong downward trend of malaria cases due to the impact of recent intervention programs. Because this study focused on assessing different sources of remotely-sensed environmental data, incorporating these socio-economic factors was outside our scope.

We faced additional limitations in classifying settlements via PlanetScope imagery. The spatial resolution of 3 m was sufficient to capture agglomerations of buildings, as well as most of the isolated buildings. However, it was not sufficient to capture some smaller huts with thatched roofs. To capture these buildings, imagery with an even higher spatial resolution would be necessary. Additional research on high resolution settlement mapping could also help to improve downscaled population estimates such as those produced by

WorldPop (Tatem, 2017) and LandScan (Dobson et al., 2000), which would allow local estimates of malaria incidence in addition to the proportion of outpatients diagnosed with malaria.

2.5 Conclusion

Relationships between spatio-temporal patterns of malaria proportion and environmental variables derived from satellite imagery varied in two landscapes in Ethiopia and were different from results of previous malaria-environment studies conducted at coarser resolution (data summarized at woreda level) and finer (sample households at village level) resolutions. Associations between climate variables and malaria followed the expected pattern with higher temperatures and more rainfall leading to more malaria cases. In both study areas, we found that kebeles with lower vegetation greenness and moisture during the malaria transmission season had the most malaria cases. Malaria was associated with concentrated settlement patterns in one study area, and with low settlement density in the other study area. To some degree, the relationships of these environmental variables with malaria likely reflected indirect relationships with aspects of the social environment such as seasonal migration, water management, and access to health care that affect malaria risk. Recent interventions have lowered malaria incidence and may also have modified some of these malaria-environment associations. These results emphasize that malaria-environment relationships based on remotely sensed environmental indices are contingent on the scale of analysis as well as the geographic setting. Knowledge of the local malaria epidemiology and its connections with physical and human geography is therefore essential for understanding these relationships and applying them to support risk assessment for public health applications.

2.6 Supplemental material

To study whether spatial clustering of malaria cases occurs in our study area, we ran spatial scan statistics using SaTScan software. SatScan is designed to detect statistically significant spatial, temporal, or space-time clusters (Kulldorff, 2018). For our study we used a retrospective, purely spatial Poisson-based scan statistic with an elliptical scan window, which we then ran for each year individually. The SatScan method searches for clusters by exploring multiple elliptical windows of varying location, size, and shapes across the study area.

For each sample ellipse, SaTScan determines the number of observed versus the expected observations, where the number of expected cases is proportional to the population under risk in each area. As no recent population data is available at kebele level, we used annualized outpatient numbers from health facilities within each kebele as indicators for population at risk.

The observed and expected values within each ellipse are then compared to those outside of the ellipse via a likelihood ratio. Scan clusters with significantly higher than expected prevalence, are labeled as high clusters. We refer to these statistically significant clusters of higher than expected malaria cases, as “hotspots”.

The maximum spatial cluster size identifiable was defined as a percentage of the population used in the analysis. We used an upper limit of 50% of the population at risk, as it was recommended by the user guide, and as it is the highest percentage the software will allow (Kulldorff, 2018). The advantage of choosing a high percentage is that it allows the

software to look for clusters of varying size more freely and without cluster size pre-selection bias.

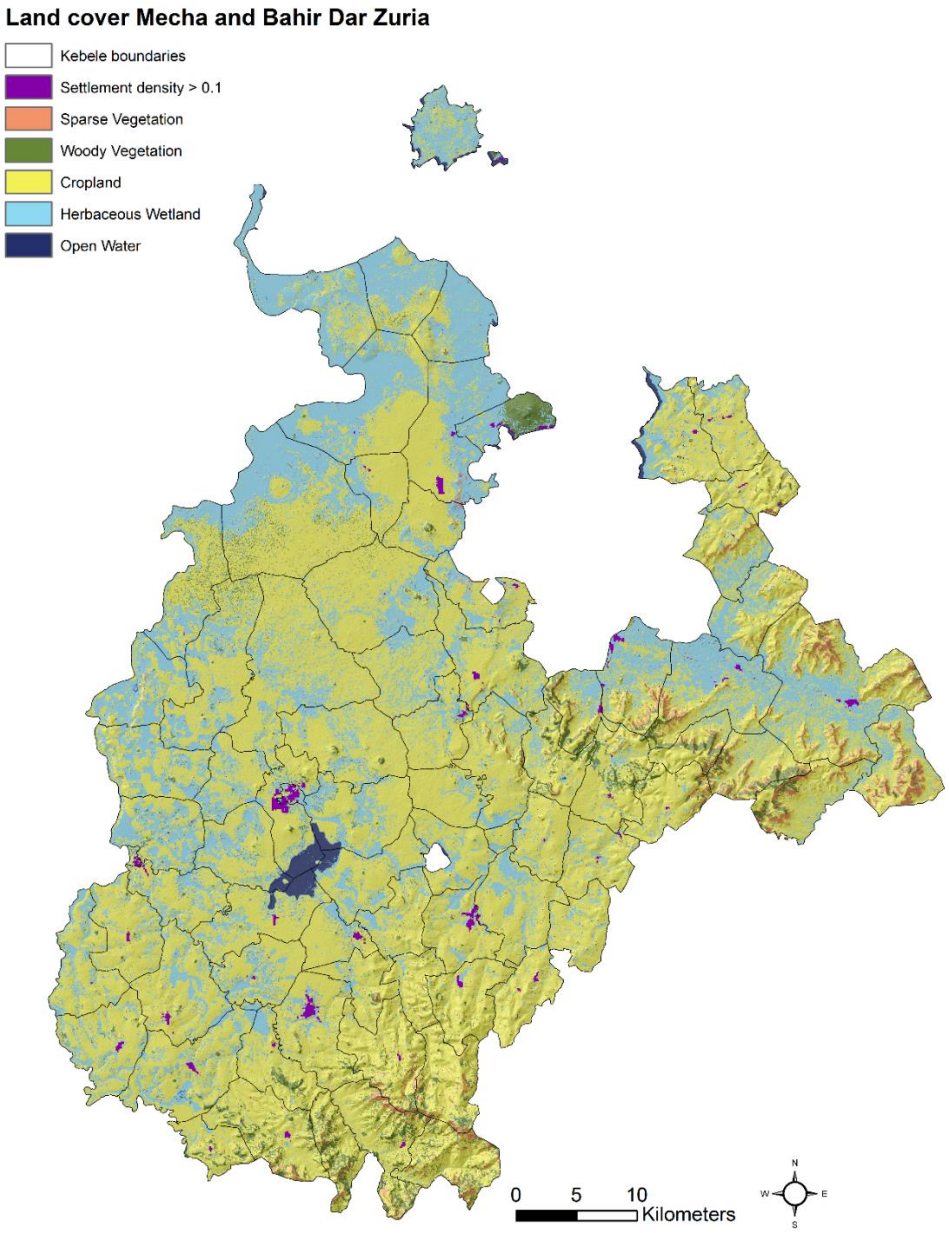

Figure 2-7: Major land cover classes and hillshade in Mecha and Bahir Dar Zuria. Settlement density is derived based on a classification of buildings via high-resolution PlanetScope imagery. A settlement density value above 0.1 can be considered an agglomeration of buildings. All other land cover classes are taken from Midekisa et al. (Midekisa et al., 2014)

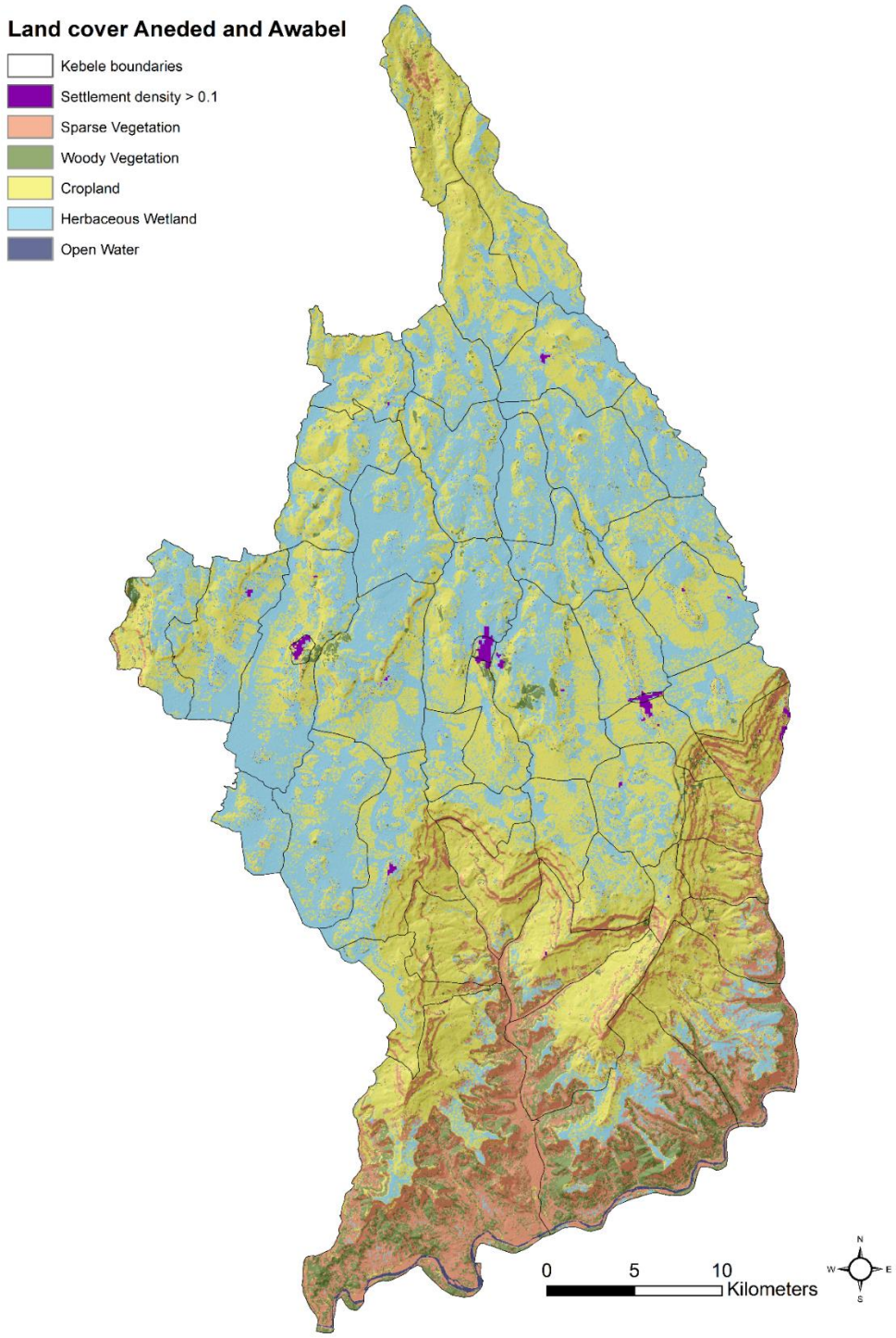

Figure 2-8: Major land cover classes and hillshade in Aneded and Awabel. Settlement density is derived based on a classification of buildings via high-resolution PlanetScope imagery. A settlement density value above 0.1 can be considered an agglomeration of buildings. All other land cover classes are taken from Midekisa et al. (Midekisa et al., 2014)

Table 2-3: Mecha and Bahir Dar Zuryia variable importance measures. Shown are all variables used to train a Boosted Regression Tree Model, and their variable importance measure (%). Variables are ranked in descending order, according to their contribution in fitting the model.

Total Malaria		P. falciparum		P. vivax	
Variable	Variable importance %	Variable	Variable importance %	Variable	Variable importance %
NDVI trans	11.07	NDVI trans	10.73	NDVI trans	17.91
SETMX	10.64	PREC rainy	9.63	SETMX	13.25
PREC rainy	9.30	NDMI trans	8.66	Year	8.62
Year	8.97	Year	7.35	NDMI trans	6.92
NDMI trans	8.55	PREC trans	7.26	PREC trans	5.07
PREC trans	8.34	SETMX	7.06	NDVI dry	4.90
SPVEG	5.09	LST trans	5.80	SETME	4.87
WETL	4.91	WETL	4.74	NDMI rainy	4.75
LST trans	4.39	SPVEG	4.62	NDMI dry	4.65
NDMI rainy	3.53	NDMI rainy	4.12	NDVI rainy	4.08
HAND	3.17	NDVI rainy	3.57	PREC rainy	3.86
NDVI dry	3.09	PREC dry	3.52	CROP	3.26
NDVI rainy	2.48	WOODY	3.40	SPVEG	3.16
SETME	2.48	NDVI dry	3.14	LST trans	3.01
PREC dry	2.45	DISTSW	3.09	HAND	2.60
WOODY	2.41	NDMI dry	2.41	PREC dry	1.92
NDMI dry	2.33	LST dry	2.26	WETL	1.51
LST dry	1.55	HAND	2.24	LST dry	1.41
WATER	1.52	LST rainy	2.02	WATER	1.22
LST rainy	1.50	SETME	1.76	WOODY	1.15
DISTSW	1.25	CROP	1.72	DISTSW	1.02
CROP	0.94	WATER	0.88	LST rainy	0.84
IRRI	0.04	IRRI	0.01	IRRI	0.04

3 Identifying environmental risk factors and mapping the distribution of West Nile virus in an endemic region of North America

Abstract

Understanding the geographic distribution of mosquito-borne disease and mapping disease risk are important for prevention and control efforts. Mosquito-borne viruses (arboviruses), such as West Nile virus (WNV), are highly dependent on environmental conditions. Therefore, the use of environmental data can help in making spatial predictions of disease distribution. We used geocoded human case data for 2004-2017 and population-weighted control points in combination with multiple geospatial environmental datasets to assess the environmental drivers of WNV cases and to map relative infection risk in South Dakota, USA. We compared the effectiveness of 1) land cover and physiography data, 2) climate data, and 3) spectral data for mapping the risk of WNV in South Dakota. A final model combining all datasets was used to predict spatial patterns of disease transmission and characterize the associations between environmental factors and WNV risk.

We used a boosted regression tree (BRT) model to identify the most important variables driving WNV risk and generated risk maps by applying this model across the entire state. We found that combining multiple sources of environmental data resulted in the most accurate predictions. Elevation, late season humidity, and early-season surface moisture were the most important predictors of disease distribution. Indices that quantified inter-annual variability of climatic conditions and land surface moisture were better predictors

than inter-annual mean. We suggest that combining measures of inter-annual environmental variability with static land cover and physiography variables can help to improve spatial predictions of arbovirus transmission risk.

3.1 Introduction

Understanding the geographic distribution of disease and mapping disease risk are important in supporting public health efforts to reduce the burden of infectious disease (Hay et al., 2013; Ostfeld et al., 2005; Pigott et al., 2015). Locating high-risk areas and identifying populations at risk can allow health officials to implement targeted disease prevention, control, and elimination efforts. In the case of mosquito-borne diseases, targeted interventions can improve the effectiveness of disease prevention and vector control campaigns. By knowing where disease hot spots are located, public authorities can apply insecticides for mosquito control primarily in high-risk areas, and therefore efficiently decrease disease transmission (Bousema et al., 2012) while reducing concerns associated with the development of resistance against insecticides (Hemingway and Ranson, 2000). Additionally, targeting community outreach and public education encourages communities at risk to engage in prevention behaviors, as people who are aware of the risk in their neighborhood are more likely to eliminate potential breeding sites in their homes, apply insect repellent, dress appropriately to avoid bites, and avoid the outdoors during mosquito feeding hours (Mitchell et al. 2018). Such spatially-targeted strategies at the community level are only possible if we can identify risk areas at a sufficiently high spatial resolution.

In this study, we focus on understanding and predicting fine-grained spatial patterns of human West Nile virus (WNV) risk. WNV is a mosquito-borne pathogen of global importance, as it is the most widely distributed encephalitic flavivirus (May et al., 2011) and the most widespread cause of arboviral neurological disease in the world (Chancey et al., 2015). It is primarily a zoonotic disease transmitted between birds as the main reservoir

hosts and mosquitoes as vectors. Infections in horses and humans can occur as spillovers, but these organisms are dead-end hosts. First isolated in Uganda, the virus has spread across the globe (Chancey et al., 2015), including to the United States. After its first introduction to the United States in 1999, it dispersed rapidly and now covers a large geographical area and a wide ecological range (Reisen, 2013). In the United States, WNV has caused more than 45,000 reported human cases, including 22,913 cases of severe neuroinvasive disease and 2,138 deaths (CDC, 2017, 2016). Surprisingly, the highest incidence rates of WNV neuroinvasive disease within the country are reported in the Northern Great Plains, with South Dakota having the highest disease rate (Burakoff et al., 2018). Detailed WNV risk maps for high-risk states such as South Dakota will help public health authorities to implement strategies that can reduce transmission rates by allowing them to target those communities at highest risk.

Environmental conditions can have a large influence on the ecology of mosquito-borne diseases, such as WNV. For WNV to occur in humans, the environmental conditions need to be suitable for vector mosquitoes to breed, become infected from suitable avian hosts, survive long enough to become infectious, and finally transmit the virus to a susceptible human host. Vector species, host species, and the virus itself are sensitive to fluctuations in climate and landscape features. High temperature has been found to increase the risk of WNV transmission (Davis et al., 2017; Lockaby et al., 2016; Marcantonio et al., 2015), whereas the effects of humidity (Lebl et al., 2013) and precipitation (Wimberly et al., 2014) are more variable. Increasing temperature generally leads to higher mosquito growth rates and shortens the gonotrophic cycle (the time required to produce eggs after a blood meal) and the extrinsic incubation period (the time required for an infected mosquito to become

infectious) (Hartley et al., 2012; Kilpatrick et al., 2008; Reisen et al., 2006). However, high temperature can also decrease the longevity of female mosquitoes (Ciota et al., 2014). In addition to these temperature effects, humidity influences the survival of adult mosquitoes and their activity pattern, and precipitation influences the availability of breeding sites. Landscape features, such as topography, soils, and vegetation also affect WNV transmission, as these variables determine the available habitats for mosquitoes and birds, and WNV risk is highest in areas with suitable habitat for mosquitoes and birds to complete the transmission cycle (Chuang et al., 2012). Depending on the geographic location and the particular vector and host species, these high-risk zones can include rural agricultural areas (Chuang et al., 2012; Reisen et al., 2013), wetlands (Sánchez-Gómez et al., 2017), and urban areas with small forest patches (Lockaby et al., 2016).

Because mosquito-borne disease transmission is sensitive to variations in habitat and climate, geospatial environmental data can be used to develop disease risk maps. In this approach, data on disease occurrence at specific locations are combined with predictor variables from geospatial datasets to develop predictive models (Peterson, 2014). These models are then applied across the entire geographic domain to produce predictions of disease risk at unsampled locations. Considerable advances in using environmental data for assessing disease risk have been made in recent years (Kraemer et al., 2016). The increasing availability of geospatial datasets such as gridded meteorological data, land cover maps, and digital elevation models now facilitates the analysis and prediction of disease risk across large areas ranging from landscapes to the globe (Ostfeld et al., 2005; Peterson, 2014). Numerous studies have used geospatial datasets and spatial modelling techniques to predict spatial patterns of vector borne diseases, including tick-borne diseases

(Wimberly et al., 2008), West Nile virus (Chuang et al., 2012; Peterson et al., 2008; Tran et al., 2014; Young et al., 2013), malaria (Benali et al., 2014; Midekisa et al., 2014) and chikungunya (Fischer et al., 2013).

One important source of environmental data is satellite remote sensing. Sensors carried by Earth-observing satellites provide information on reflected, emitted, and backscattered radiation that can be used to identify objects and phenomena on the Earth's surface. Today, we have unprecedented access to freely-available, long-term archives of high-quality satellite images with global coverage at spatial resolutions from tens to thousands of meters (Hansen and Loveland, 2012; Wulder et al., 2016). As a result, there has been increasing usage of satellite remote sensing data in large-scale environmental monitoring applications, including disease modelling. Satellite data can be used to compute a variety of spectral indices that are sensitive to environmental factors such as vegetation greenness, water, and land surface temperature. These indices characterize a wide range of environmental conditions, and thus may be more effective at identifying the specific vector and host habitats associated with disease transmission than other data sources such as classified land cover maps and coarser-grained meteorological datasets. Additionally, historical satellite archives allow us to observe long-term changes in land surface characteristics and measure their inter- and intra-annual variability, as opposed to viewing land cover as a static condition. Thus, the use of satellite-derived spectral information can help to improve disease risk maps by providing novel sources of environmental information that are relevant to vector and host ecology.

With the ever-increasing availability of high-quality geospatial data, researchers face the challenge of deciding what data to include for their specific applications. The objective of

this study is to explore the use of different types of geospatial environmental data for predicting the spatial patterns WNV disease risk in South Dakota. We compared three broad classes of geospatial data commonly used in previous studies: 1) land cover and physiography data, describing the physical structure, vegetation, and hydrology of the land surface at a single point in time; 2) climate data, derived from gridded and interpolated meteorological data; and 3) spectral indices derived from satellite sensors, measuring a variety of environmental conditions on the land surface. We sought to determine which of these data types produced the best fitting spatial model of human WNV cases and which specific environmental variables had the strongest influence on WNV patterns when all data types were included in a combined model. This approach allowed us to identify the best model for predicting the spatial distribution of WNV risk in South Dakota and yielded broader insights that can help to inform similar mapping efforts for different diseases in other locations.

3.2 Methods

The overall workflow involved acquiring and processing human WNV case data, exploratory analysis of the WNV case data, acquiring and processing environmental data, fitting boosted regression tree models of WNV cases using the environmental variables, and applying these models to generate risk maps (Figure 3-1). These steps are described in detail in the following subsections.

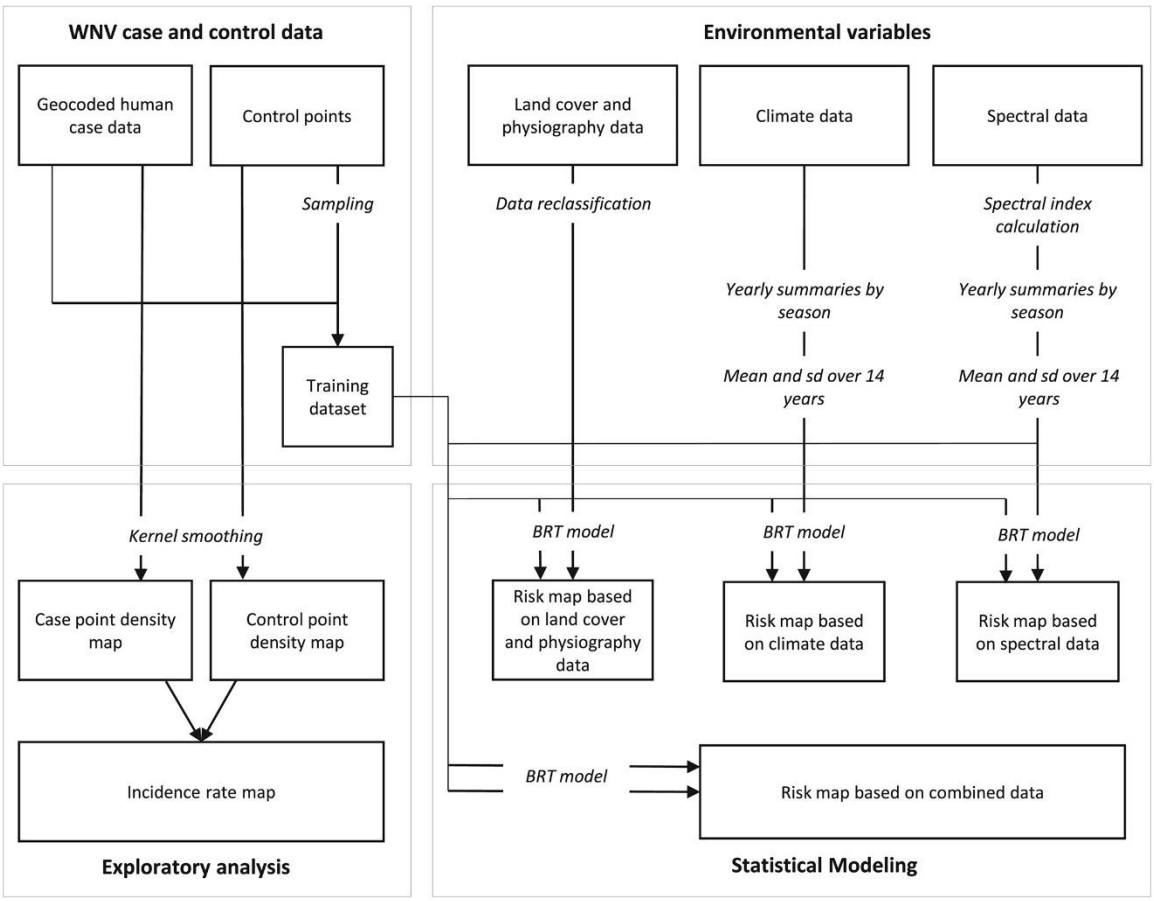

Figure 3-1: Workflow for West Nile virus (WNV) modeling and risk mapping. The analysis is divided into four major sections. The processing of case and control data, an exploratory analysis of case and control data, the processing of environmental variables, and the statistical modeling to produce WNV risk maps. Bold text represents the four main sections; italic text represents data processing steps, boxes, and arrows represent the main data flows. BRT = boosted regression tree.

3.2.1 West Nile virus case and control data

From the South Dakota Department of Health (SDDOH) we received 1381 serologically-confirmed case records of West Nile fever, West Nile neuroinvasive disease, and WNV infected blood donor detections with spatial information for the years 2004 – 2017. Cases of West Nile fever and West Nile neuroinvasive diseases were diagnosed by clinicians and confirmed by laboratory techniques and meet the CDC case definition (CDC, 2015). In

addition to symptomatic WNV cases, we also included cases of WNV-positive blood donors reported to the SDDOH. After the first two years of introduction in 2002-2003, the annual number of WNV cases declined and the geographic distribution of cases shifted (Wimberly et al., 2013). These first two years were excluded from our analysis as the disease was not yet endemic, and the patterns were atypical from the following years.

Of the case records between the years 2004-2017, 1257 contained geocoded coordinates of the home addresses of confirmed cases, an additional 121 data points contained at least zip code information, and only three cases had no spatial information. For those cases where only zip-code information was available, we assigned the case data point to a random location within the given zip code tabulation area (ZCTA). The availability of spatial coordinates at the household level allowed us to assign a precise spatial location to most cases and avoided problems associated with summarizing cases within arbitrary spatial boundaries such as ZCTAs and counties.

We used the technique of Chuang et al. (2012) to generate a set of 50,000 control points to represent the background distribution of the human population. We used census-block level population information from the 2010 census to generate control points proportional to the population of census block. We assigned these control points to random locations within each block. This population-based weighting controlled for the uneven population distribution and prevented a bias that would otherwise have skewed our results towards over-predicting risk in highly populated places.

3.2.2 Exploratory analysis

To study the distribution of WNV cases in relation to the population distribution, we created a map of smoothed incidence rates. We applied a kernel smoothing technique to

create a gridded layer where every grid cell contained density estimates for case and control points. The density estimates for each cell were based on the numbers of case and control points within a set radius given by a smoothing parameter, and their distances to the cell. As smoothing parameter, we chose a bandwidth of 0.5 degree. We used the R package kernSmooth (Wand & Jones 1995). Then we created an approximate smoothed incidence rate by dividing the pixel values of the smoothed case grid by the control grid.

3.2.3 Environmental variables

We compiled a variety of datasets that contain environmental information relevant to disease amplification and transmission. A list of the data used for the disease risk models is provided in Table 3-1. We grouped the data into three main types. *Land cover and physiography data* included a variety of standard national-level data products characterizing land cover and land use, topography, soils, and wetlands. *Climate data* included indices derived from a gridded meteorological dataset. *Spectral data* included indices calculated from high temporal resolution remotely sensed surface reflectance data in the optical and infrared wavelengths. The specific variables derived from these data sets are listed in Table 3-2.

We retrieved land cover and physiography data for South Dakota from various national datasets. Land cover variables were derived from the 2011 National Land Cover Database (NLCD) at a 30m resolution. From the initial NLCD land cover gridded dataset, we created new datasets containing presence/absence information for each of the five land cover classes: developed areas (developed low intensity, developed medium, and developed high), forest (deciduous forest, evergreen forest, and mixed forest), grassland (grassland/herbaceous), pasture (pasture/hay), wetland (woody wetlands, emergent

herbaceous wetland) and cropland (cultivated crops). Data on wetland distribution was also downloaded from the National Wetlands Inventory (NWI). We first converted the shapefiles into raster format and then created individual presence/absence layers for each of five wetland types: freshwater emergent wetland, freshwater pond, riverine, freshwater forested/shrub wetland and lakes. We extracted data on ponding frequency from the Soil Survey Geographic (SSURGO) Database. Ponding frequency is reported as "percentage of the map unit that is subject to water ponded on the soil surface" Lastly, we downloaded data from the National Elevation Dataset (NED) at a 30-meter resolution.

Table 3-1: Environmental datasets used as explanatory variables.

Data set	Spatial resolution / map scale	Temporal resolution	Time frame	Source
MODIS MCD43A4 (Collection 6)	500 m	daily, based on 16-day moving window	2004-2017	https://urs.earthdata.nasa.gov/
NLDAS	1/8 degree (at 40° north, 13.9 km latitude and 10.7 longitude)	hourly	2004-2017	https://daac.gsfc.nasa.gov
NLCD	30 m	NA	2011	https://datagateway.nrcs.usda.gov/
NWI	1:24,000	NA	2017	https://www.fws.gov/wetlands/Data/ Data-Download.html
SSURGO	10m	NA	2018	https://datagateway.nrcs.usda.gov/
NED	1 arcsec (approx 30 m)	NA	n.d	https://datagateway.nrcs.usda.gov/

Climatic variables were derived from the North American Land Data Assimilation System (NLDAS) atmospheric forcing data (Mitchell et al., 2004). We retrieved hourly temperature, precipitation, and specific humidity data for the years 2004-2017 from the

Goddard Earth Sciences Data and Information Services Center. We calculated daily mean values for temperature [°C], relative humidity [%], and vapour pressure deficit [kPa], as well as daily total values for precipitation [mm] from the initial hourly data. Then we calculated yearly mean values for two parts of the WNV transmission season, as defined by Wimberly et al. (2008b). The early WNV season (May & June) encompasses the period of virus amplification in the mosquito-bird system prior to the occurrence of most human WNV cases. The late WNV season is when most mosquito-human transmission takes place, and the majority of human cases occur. We calculated the inter-annual mean and standard deviation of each early-season and late season variable.

Table 3-2: List of environmental variables and their abbreviations.

Variable	Description	Data source
NDVI_*	Normalized Difference Vegetation Index (Bands: red, nir)	MODIS MCD43A4 Collection 6
NDWI_*	Normalized Difference Water Index (Bands: green, nir)	MODIS MCD43A4 Collection 6
NDWI2_*	Normalized Difference Water Index 2 (Bands: nir, swir2)	MODIS MCD43A4 Collection 6
MNDWI_*	Modified Normalized Difference Water Index (Bands: green, swir2)	MODIS MCD43A4 Collection 6
TcWet_*	Tasseled cap transformation wetness	MODIS MCD43A4 Collection 6
TcGreen_*	Tasseled cap transformation greenness	MODIS MCD43A4 Collection 6
TcBright_*	Tasseled cap transformation brightness	MODIS MCD43A4 Collection 6
Prec_*	Total precipitation	NLDAS
RH_*	Relative humidity	NLDAS
VPD_*	Vapour pressure deficit	NLDAS
Temp_*	Daily mean temperature	NLDAS
Crop	Cultivated cropland	NLCD
Pasture	Pasture/Hay	NLCD
Develop	Developed land	NLCD
Forest	Forest	NLCD
Grass	Grassland	NLCD
Lakes	Lakes	NWI
Ponds	Ponds	NWI
River	Riverine	NWI
Emergent	Emergent wetlands	NWI
Elevation	Elevation	NED
PondFr	Ponding frequency	SSURGO

* We calculated four seasonal summaries for each of these variables. EarlyMean: May/June inter-annual mean, EarlySD: May/June inter-annual standard deviation, LateMean: May/June inter-annual mean, LateSD: July/August inter-annual standard deviation.

We acquired spectral data from the MODIS MCD43A4 Collection 6 nadir bidirectional reflectance distribution function (BRDF) adjusted surface reflectance data set for the years 2004-2017. The imagery has a 500m spatial resolution and comes as daily data summarized over a 16-day moving window. Only pixels using the full BRDF inversion were used. On a pixel basis, we first calculated different environmental indices presenting different aspects of environmental conditions on the land surface. We calculated the normalized difference vegetation index (NDVI) (Tucker, 1979) to study the impact of vegetation on disease risk. Additionally, we calculated three water indices, the normalized difference water index (NDWI) which is sensitive to open surface water (McFeeters, 1996), a different normalized difference water index (NDWI2) which is sensitive to vegetation liquid water (Gao, 1996), and the modified normalized difference water index (MNDWI) which is useful for identifying open water in areas with high background noise (Xu, 2006). Additionally, we used the tasseled cap transformation, which converts spectral layers into interpretable bands conveying information on vegetation greenness, vegetation and soil wetness, and land surface brightness (Lobser and Cohen, 2007). The spectral indices and tasseled cap transformation were calculated in the R-package RStoolbox. We then calculated the median of these indices in each year for the early and late West Nile transmission seasons defined previously. Finally, we calculated the inter-annual mean and standard deviation for these yearly indices.

We converted all datasets to the same spatial resolution. We chose 300 m as the common resolution, as a multiple of the NLCD and NED data at 30 m resolution. For rescaling continuous variables, we used bilinear interpolation. For rescaling categorical variables, we aggregated pixels using a majority filter. To account for the uncertainty of the exact

location of infection as given by the patient's home address, and to include environmental information in further vicinity of the case locations, we applied a moving window of 5x5 pixels. Examples of the gridded environmental layers are shown in Figure 3-2.

3.2.4 Statistical modelling

We fitted an ensemble of boosted regression tree **(**BRT) models to predict the spatial distribution of West Nile virus transmission risk and to study the influence of the different environmental factors, using the gbm package in the statistical software R, version 3.4.4 (R Core Team, 2016).

BRT models are widely used in applications to predict species distribution (Elith et al., 2008; Elith and Graham, 2009) and increasingly also to predict spatial patterns in disease transmission (Hay et al., 2013). BRT models are useful in these applications due to their high predictive accuracy and their ability to capture non-linear relationships between predictor and response variables and interactions among predictor variables. We created four different models based on different sets of environmental data, 1) land cover and physiography data, 2) climate data, 3) spectral data, and 4) a combination of all datasets.

To fit the BRT models, we created a dataset of all 1,374 disease case points and a set of 12,366 control points (nine times the number of all case points) that were randomly sampled from our previously generated control dataset. Pixel information was extracted from the environmental datasets, at each sampled case and control point location. We initially fit the BRT model using the same approach as Elith et al. (2008) with a tree complexity of 5, a learning rate of 0.01 and a bag fraction of 0.5. We compared the cross-validated predictive performance of the model under different parameter combinations

with tree complexity ranging from 3-6, learning rate from 0.01-0.001 and bag fraction from 0.5-0.7 to select the optimal parameters for each model.

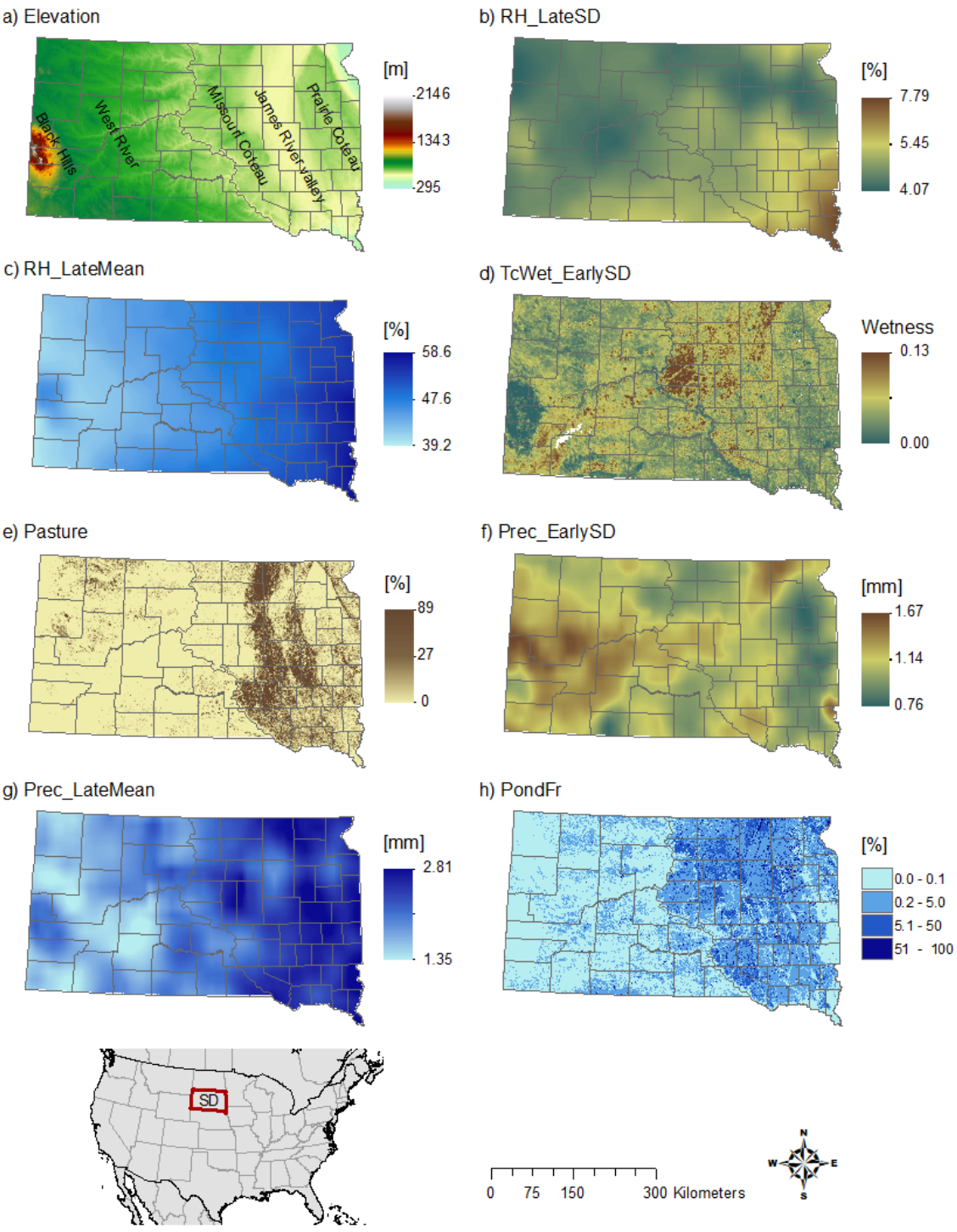

Figure 3-2: . Environmental variables mapped across South Dakota. (a) Elevation, (b) 14-year standard deviation of relative humidity in July/August, (c) 14-year mean for relative humidity in July/August, (d) 14-year standard deviation of tasseled cap wetness in May/June, (e) percent pasture, (f) 14-year standard deviation of precipitation in May/June, (g) 14-year mean of precipitation for July/August, (h) ponding frequency.

To evaluate the predictive accuracy of each model, the data were split into subsets for training (80%) and validation (20%). To account for variation in the model results that arise from control point sampling, we fitted each model 20 times, using a different subset of control points and a different random assignment of training and validation subsets each time. For each model run, we used all available case points, and sampled new control points (with replacement) from the complete control dataset. We then used the fitted BRT model object to produce spatial predictions across the entire state. We calculated the mean predicted value for each grid cell from all 20 predictions for each pixel to characterize the spatial distribution of WNV transmission risk. We converted the case probabilities to percentiles to visualize the relative risk of WNV transmission across the state. To quantify the relative influence of different environmental variables, we ranked their relative importance during the model fitting process. Relative importance was measured as the normalized reduction of squared error attributable to each variable. We then created partial dependence plots for the most important variables to examine the relationship between environmental variables and disease risk. For accuracy assessment, we used the predicted values and the validation dataset to calculate AUC values for each model run and calculated a mean AUC for the 20 model runs.

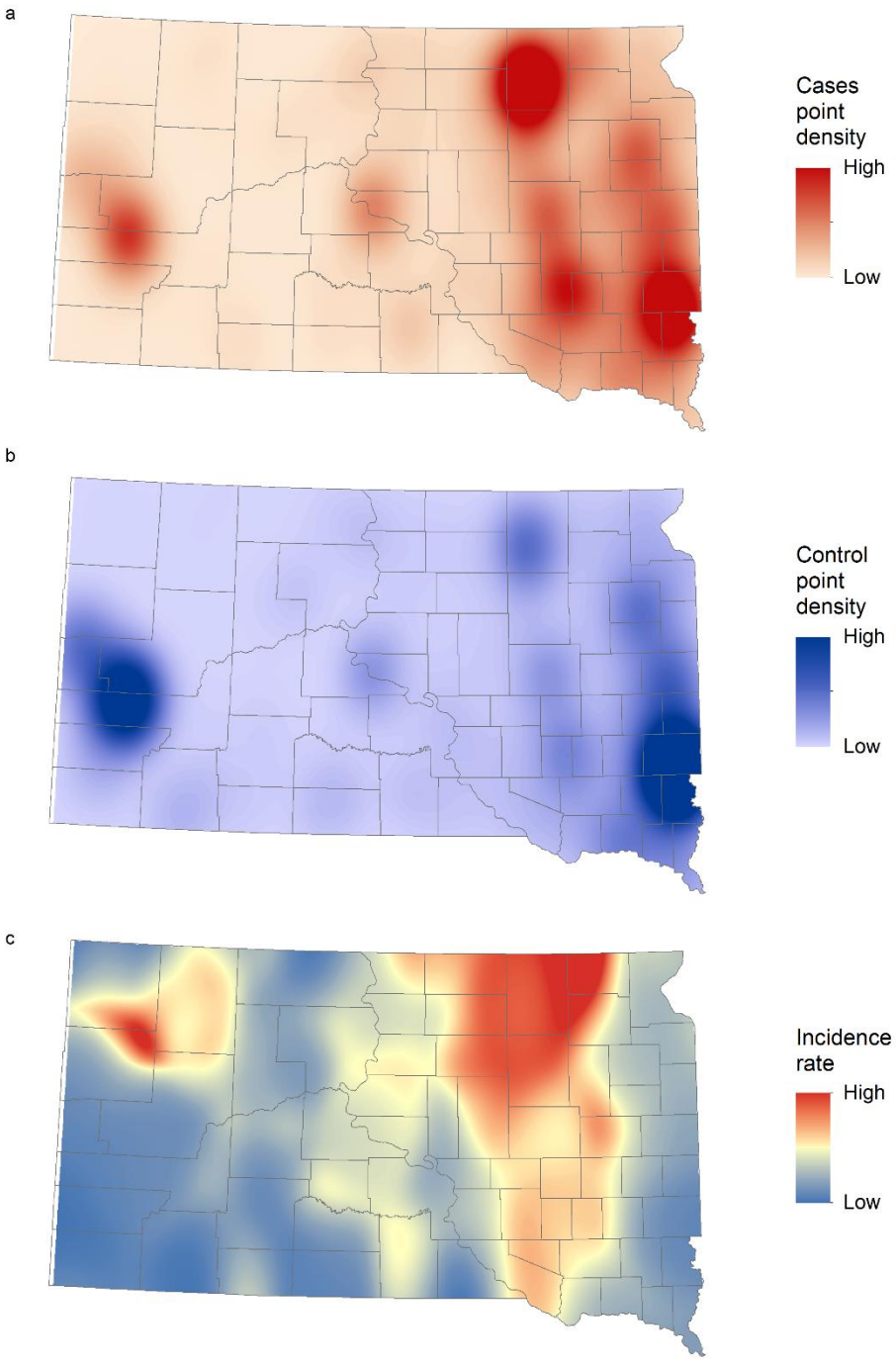

Figure 3-3: (a) Spatially smoothed West Nile virus (WNV) case points, (b) population weighted control points, and (c) incidence rates of human WNV disease in South Dakota for 2004-2017.

3.3 Results

3.3.1 Spatial patterns

Figure 3-3 shows the spatially-smoothed distribution of WNV cases, control points, and incidence rates for South Dakota. We observed the largest number of cases (Figure 3-3a) in areas with high population densities (Figure 3-3b). However, the incidence rate, calculated as cases divided by population, was relatively low in most urban areas and highest in the northern James River Valley located in northeastern South Dakota. There was another, smaller area of high incidence just north of the Black Hills in northwestern South Dakota.

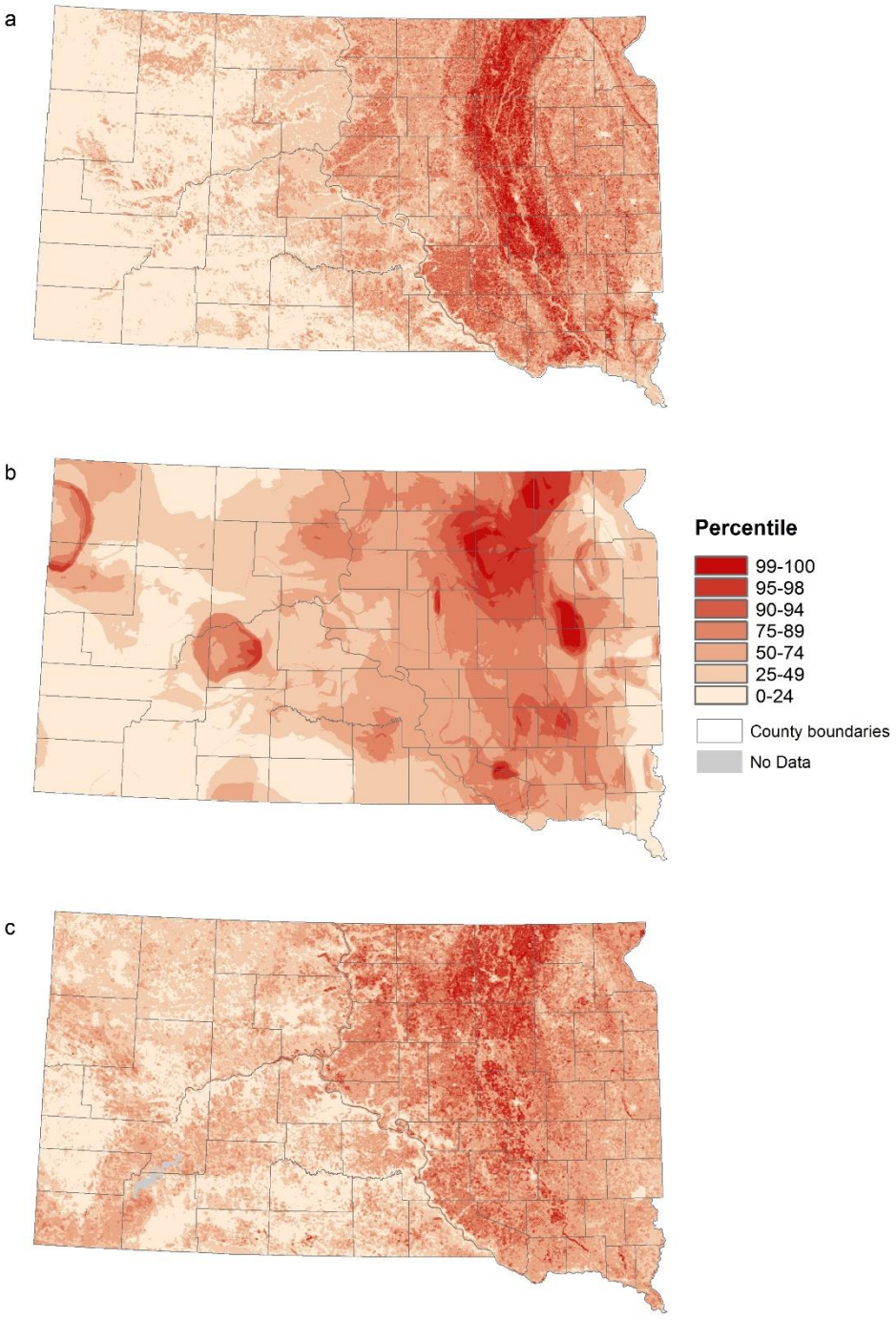

Figure 3-4: West Nile virus risk patterns based on three different environmental datasets: (a) land cover and physiography data, (b) climate data, and (c) spectral data.

All three risk maps in Figure 3-4 showed a similar pattern in highlighting the James River Valley stretching from northeastern to southeastern South Dakota, and area of high risk. Also, all models showed an east-west gradient with areas east of the Missouri River having higher risk (Figure 3-4).

The model based on the land cover and physiography data (Figure 3-4a) clearly highlighted the whole James River Valley the areas of highest relative risk all located in the valley. The risk gradually decreased with increasing elevation above and distance from the river. The areas west of the Missouri River predominantly had relative risk below the median. The climate-driven model (Figure 3-4b) showed a patchy pattern, with high risk occurring mostly in the northern parts of the James River Valley and the plateaus to the east and west of the valley, as well as a few isolated hotspots in western South Dakota. The model based on spectral data (Figure 3-4c) mostly highlighted the northern James River Valley and the parts of the Missouri Coteau region located to the west. Also, as a result of the different resolution of the environmental datasets, risk patterns are resolved at different scales. The models based on land cover and physiography data and the spectral data show finer grained spatial patterns, whereas the climate data driven model presents a comparatively smooth pattern.

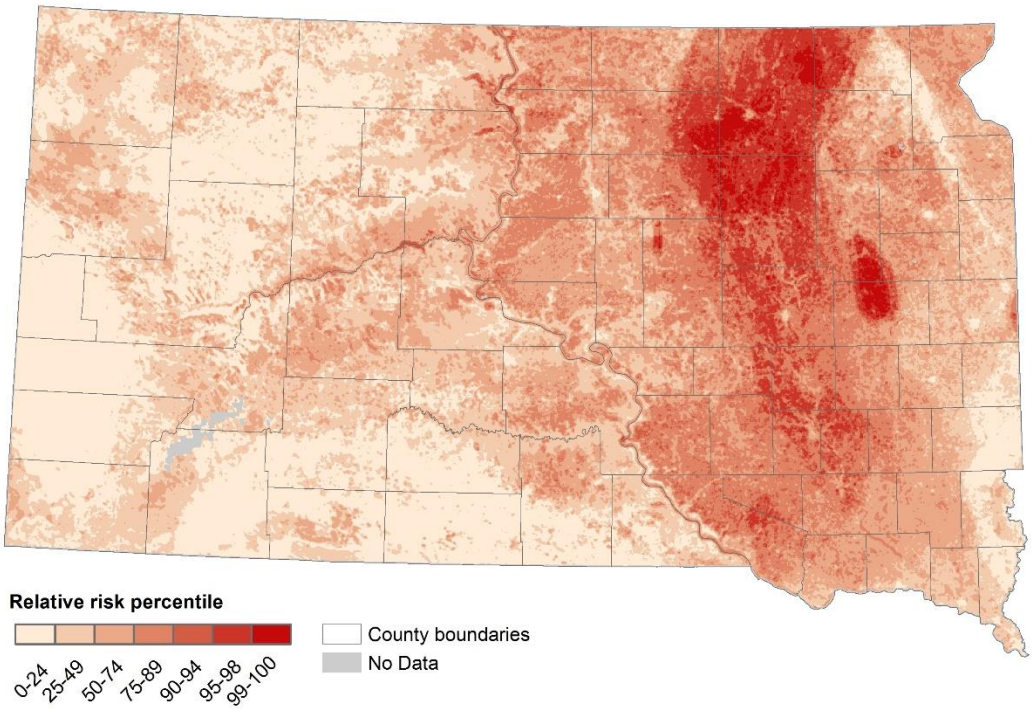

Figure 3-5: West Nile virus risk patterns based on BRT model predictions using all available environmental variables.

The spatial pattern of relative WNV risk in South Dakota predicted by the model including all available environmental variables is presented in Figure 3-5. Based on this model, the relative risk of WNV was highest in the northern part of the James River Valley, with more areas of high risk in eastern South Dakota and one additional hot spot at the edge of the Prairie Coteau.

The combined model performed best, with the highest mean AUC value of 0.727. The combined model's AUC was followed by lower AUC values from the models driven by land cover and physiography data (0.711), climate data (0.719) and spectral data (0.679). Despite the relatively narrow range of AUC values for the best three models, the maps

produced by the models showed considerable differences in spatial patterns of relative WNV risk as described previously.

3.3.2 Environmental data analysis

We interpreted the influences of environmental variables by quantifying the relative importance of each variable in the different BRT models (Figure 3-6) and by visualizing the partial dependence between relative risk and the most influential variables (Figure 3-7). Elevation was the most important variable in the model based on land cover and physiography data (Figure 3-6a), with low elevations more favorable for disease transmission. Of the climatic variables, the standard deviation of precipitation during early mosquito season (May/June) had the strongest impact on disease risk (Figure 3-6b). We observed the highest risk at very high and very low inter-annual variation of early-season precipitation. Inter-annual variation in moisture-related metrics, including the tasseled cap wetness index, NDVI, and MNDWI, were the most important predictors in the model based on spectral data (Figure 3-6c). Higher inter-annual variation of these moisture-sensitive indices was associated with higher disease risk.

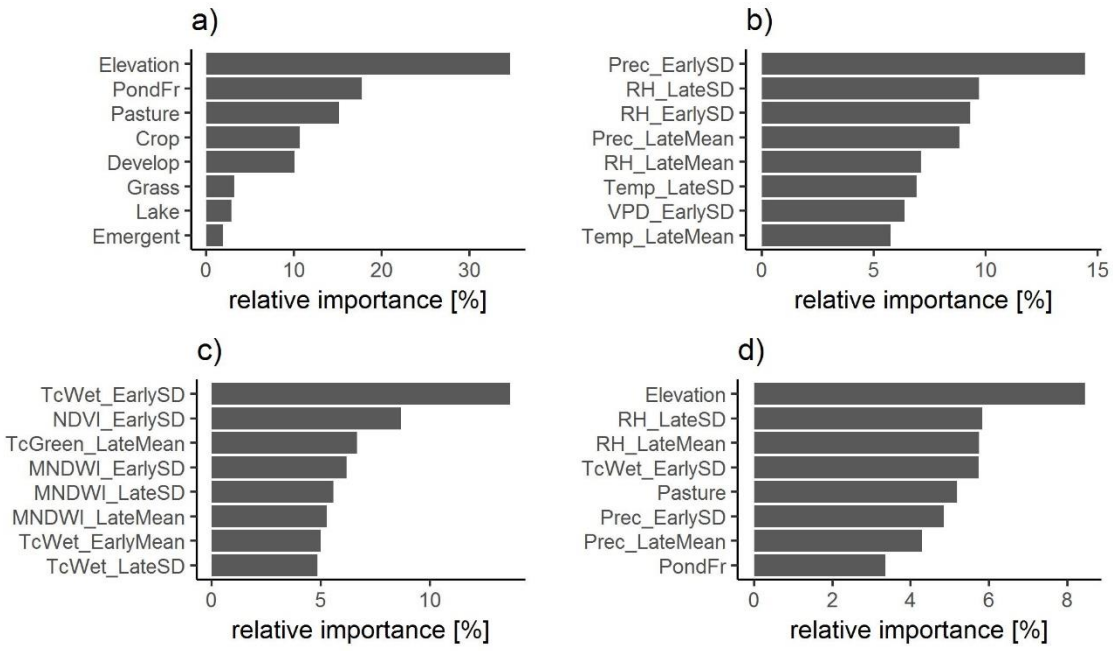

Figure 3-6: Relative importance of environmental variables in predicting the relative risk of West Nile virus. We used four sets of environmental variables to fit a BRT model: (a) land cover and physiography data, (b) climate data, (c) spectral data, and (d) all datasets combined. Explanations of the variable codes are provided in Table 3-2.

The combined model was mainly driven by elevation, followed by metrics of relative humidity and precipitation in the late season (RH_LateSD, RH_LateMean, and Prec_LateMean), inter-annual variation in precipitation and surface moisture availability in early-season (TcWet_EarlySD, Prec_EarlySD), and percent pasture, and as soil conditions associated with surface water ponding (PondFr) (Figure 3-6d). In general, climatic variables proved to be more influential than spectral variables. Of all spectral variables, inter-annual variability in surface moisture proved to be of highest importance.

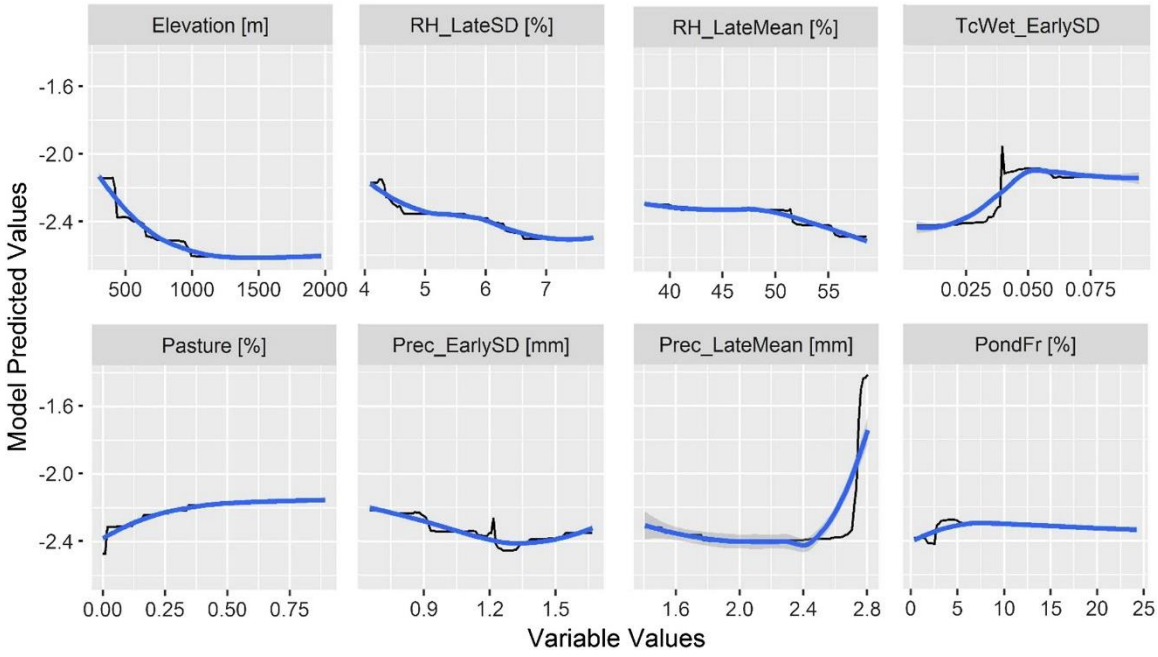

Figure 3-7: Boosted regression tree (BRT) partial dependence of risk on the top ranked environmental variables from the combined model containing all data sources. Black lines represent the fitted function for environmental variables and relative West Nile virus risk. Blue lines represent a smoothed approximation. Variables are explained in Table 3-2

3.4 Discussion

This study applied classified land cover and physiography data, climate data, and spectral data to map the relative risk of WNV transmission in South Dakota. All models highlighted particular aspects of environmentally-driven spatial heterogeneity in WNV risk, and they generated predicted risk maps with more spatial detail than the smoothed incidence rates. The models also showed some broad-scale similarities when at the state level. These similarities reflect broad-scale associations between the land cover, physiography, climate, vegetation, and surface moisture variables that were measured in the various datasets. In particular, there is a strong east-west variation across the state reflected in an elevation gradient, a climatic gradient, and a transition from predominantly cropland and pasture in the east to predominantly grasslands in the west. Despite these large-scale collinearities,

each of the datasets provided unique information and captured a different aspect of the environmental influences on WNV risk distribution.

Increasing the complexity of the model by combining all three datasets produced the best-fitting model and helped the quality of the spatial prediction. The combined model identified spatial differences in transmission risk and highlighted a major high-risk area in the northern James River Valley with one distinct hotspot on the Prairie Coteau. We found that these high-risk areas were characterized by distinctive hydrological conditions during the early WNV season and climatological conditions during late WNV season. The combined model also helped to lessen the influence of spatial artifacts visible in the other models, such as the "rings" in the climate model that are a by-product of processing the climate data and the very strong signature of elevation in the land cover and physiography model. It is notable that the four most important variables in the combined model came from all three of the data sources, emphasizing that these different types of data provide complementary information to support disease predictions.

The model based on land cover and physiography data, which was driven primarily by elevation, highlighted the whole James River Valley as a high-risk area, with lower risk on the plateaus to the east and west of the valley. The James River Valley is a glacial valley that is dominated by relatively flat lowlands with poorly drained soils, shallow water tables, and extensive wetlands. In South Dakota, the main mosquito vector is *Culex tarsalis* Coquillett (Vincent, 2018). This mosquito is associated with small unconnected semi-permanent ponds with standing water (Mercer et al., 2005; Skaff and Cheruvelil, 2016), which are abundant in the northern James River Valley. Our study also found that such catchment areas with temporary ponding are associated with high densities of WNV cases.

In contrast, the highest elevation in the state is in the Black Hills in western South Dakota, which have a different ecosystem than in the rest of the state, with lower temperatures, higher forest cover, and few suitable breeding sites for *Culex tarsalis*. Therefore, we suggest that geospatial data on land cover, physiography, and soils can identify locations that are likely to contain the temporary water bodies that serve as breeding sites for the main vector species, *Culex tarsalis*.

The model based on climatic variables highlighted the northeastern part of the state, emphasizing the distinct climatology in the northern part of the James River Valley and adjacent uplands. Both late season inter-annual variability and the long-term mean of relative humidity were important, as well as the late season long term mean of precipitation. Mosquitoes are small, ectothermic animals, and are sensitive to ambient climatic conditions, especially humidity. Mosquitoes desiccate at low humidity and avoid high humidity (Thomson, 1938), and it is not surprising that humidity is associated with disease transmission. In addition, drought conditions negatively affect the nesting success of bird communities, including passerine birds, in the Northern Great Plains (George at al., 1992). Variability in precipitation and relative humidity can therefore also affect the abundance of avian hosts. Climatic conditions were important mostly during late season, when the vector mosquito *Culex tarsalis* is feeding primarily on humans (Kent et al., 2009) and we observe the majority of the human WNV cases. We assume that climatic conditions are related to increased WNV cases in late season because during this time of increased feeding on humans, a longer life span of mosquitoes and their increased host-seeking activity will increase the chances of transmitting the disease. However, late-season transmission to

humans is also contingent upon high levels of mosquito-bird transmission earlier in the season that amplify the virus and increase the mosquito infection rate.

The spectrally-driven model highlights the same north-eastern cluster as the climate-driven model. However, the distinctive land surface dynamics of this area were characterized by inter-annual variability in early-season surface moisture availability. In contrast to the importance of climate during the late season, during the early-season the availability of surface moisture was most important. As mentioned earlier, *Culex tarsalis* mosquitoes depend on small, temporary water bodies. Mosquito-bird interaction and the potential for virus transmission are influenced by the availability of species-rich environments, such as wetlands (Reisen et al., 2013). We suggest that surface moisture conditions may be most important for early-season mosquito population growth and disease amplification in the mosquito-bird system, setting the stage for large outbreaks later in the season if conditions are favorable for mosquito survival and increased feeding on humans. We found that areas with high inter-annual variability of surface moisture were at highest risk. This result does not mean that the risk in those areas is consistently high, but rather that those areas have the potential to produce abundant breeding sites in some years, which can lead to high mosquito numbers and high rates of disease amplification.

Contrary to our expectation, temperature did not have a strong influence in our models. Temperature does have a major effect on year-to-year variation in case occurrence (Davis et al., 2017), with WNV outbreaks tending to occur in warmer years, but it does not explain the spatial patterns of disease risk. Within South Dakota, the areas with the highest incidence rates are located in the north, where the mean annual temperature is lowest. Even on a larger scale, areas of the United States with warmer temperature are not the ones most

prone to WNV. South Dakota is one of the coldest states, but has the highest incidence rate in the country (Burakoff et al., 2018). We found that hydrology and humidity related variables were more strongly associated with spatial variation in WNV risk. This finding is relevant to projections of disease distribution under different climate change scenarios. Increasing temperatures might lead to increasing frequency and severity of WNV transmission in areas already prone to outbreaks. However, other climatic variables such as precipitation and relative humidity, as well as changing land cover and hydrological conditions may be more important predictors of future shifts in the geographic pattern of WNV transmission.

This study had several important limitations. First, the locations where the WNV transmission took place are unknown, and our data only include the home residence of each case. Therefore, our map is based on the assumption that most people become infected in close proximity to their homes. Additionally, we did not include other non-environmental factors such as access to health care, underreporting of cases with weak symptoms, and underlying variability in the blood donation rate. These variables most likely explain some of the spatial distribution of WNV in South Dakota, but they were not available as geospatial datasets. Instead, this study focused on predicting WNV using environmental datasets, and the results therefore highlight the specific and environmental conditions for which local residents are at highest risk of WNV. These various sources of uncertainty are one reason why the AUC values we reported did not exceed 0.73. However, our aim was not to deliver an accurate classification of where every WNV case occurred within the state, but rather to present the distribution of relative risk across the state given environmental differences. Despite these limitations, our study was able to identify the

underlying environmental factors that influenced risk, and to map the environmentally-driven relative risk across the state.

Our research suggests that combining multiple sources of data that characterize the hydrological and climatological characteristics and their variability over long time frame can improve estimates of the spatial dimension of disease risk across larger scales. Although models with only spectral indices did not demonstrate the best model performance, adding spectral variables to the combined model improved the spatial predictions. Our recommendation for future research is to include spectral data in combination with additional environmental data to identify areas of high risk. Locating these areas will assist public health education efforts to reach the local audiences at highest risk and enable more targeted and efficient disease prevention activities. Also, targeted vector control strategies can also help to reduce the usage of insecticides, which would have financial benefits, reduce public concerns about insecticide use, and help slow the development of insecticide resistance.

3.5 Acknowledgments

This work was supported by the Grant NNX15AF74G from the NASA Applied Sciences Health and Air Quality Program. Human case data were obtained through a data-sharing agreement with the South Dakota Department of Health (SDOOH). We thank Lon Kightlinger, Nick Hill and Joshua Clayton of the SDDOH for facilitating access to these data. The underlying data for this article will be deposited in University of Oklahoma's public access institutional repository and information exchange https://shareok.org/.

4 Comparing satellite data and *in-situ* measurements in studying vector mosquito habitats in an urban environment

Abstract

Exposure to mosquito-borne diseases is influenced by landscape patterns as well as microclimates associated with different land covers. With increasing urbanization, it becomes important to understand the distribution of mosquitoes in urban landscapes. We investigated how land cover and climate influenced the abundance of the vector mosquitoes *Ae. albopictus* and *Cx. quinquefasciatus* in Norman, Oklahoma (USA).

From May-September 2019 and June-October 2020 we sampled abundances of vector mosquitoes along an urban-rural gradient using CO_2 baited BG Sentinel-2 traps. Microclimatic conditions at different sites across our study area were measured with sensors that continually recorded temperature and humidity. We mapped environmental patterns in and around Norman, Oklahoma using from Landsat, Sentinel-2, VIIRS, and CHIRPS data. We also obtained meteorological data from the closest weather station. We compared statistical models of mosquito abundance based on microclimate, remote sensing, and weather station data.

For both species, abundance models based on remotely sensed variables performed best in predicting species abundances. Both species were more abundant on trap days with higher temperature and higher relative humidity or vegetation moisture. Rainfall two weeks prior to the trap day negatively affected the abundance of both species. Both species were

positively associated with impervious surface cover surrounding a trap site. However, tree cover was negatively associated with *Ae. albopictus* abundance and positively associated with *Cx. quinquefasciatus* abundance. These results demonstrate the potential of including satellite imagery into small-scale mosquito habitat analyses as well as the need to target vector control within urban areas.

4.1 Introduction

For effective and efficient control of disease transmitting mosquitoes, we need to understand the distribution of mosquito species (Bousema et al., 2012; Hay et al., 2013; Pigott et al., 2015). Information about when and where mosquito abundance is highest can be used to target mosquito control efforts and disease prevention activities. Spatial and temporal variation of mosquito abundance is influenced by habitat availability and climate factors that affect their life cycles. Models can help us understand how different environmental variables influence species abundance and distribution. These environmental relations can also be applied to predict species occurrence and abundance for locations where mosquito observations are not available (Franklin, 2010; Guisan and Zimmermann, 2000).

In central Oklahoma, the main medically important mosquito species are in the genus *Culex* as well as *Aedes albopictus*. *Culex* mosquitoes are the primary vectors of West Nile virus in Oklahoma (Noden et al., 2015), where up to 176 human cases have been reported annually (Oklahoma State Department Of Health, 2019). *Culex quinquefasciatus* is a particularly important vector species in urban areas because it is prevalent in peridomestic environments and likely to seek blood meals from humans (Reisen et al. 1992 b). *Aedes albopictus* is a competent vector for a variety of arboviruses, including yellow fever, dengue, chikungunya, Zika and Eastern Equine encephalitis. It has also been found to transmit West Nile virus (Holick et al., 2002) and La Crosse virus (Gerhardt et al., 2001). *Ae. albopictus* is expanding its range into temperate areas (Farajollahi and Nelder, 2009; Tjaden et al., 2021) and is likely to be responsible for disease outbreaks in these areas (Manore et al., 2017) as it has already been in Europe (Vega-Rua et al., 2013). This species

occurs at high densities in urban areas, which puts densely populated areas at risk of disease outbreaks (Manore et al., 2017). Mosquito communities are understudied in Oklahoma and more research is needed to understand the abundances of disease vector mosquitoes throughout space and time. Previous studies from central Oklahoma have identified mosquito communities (Noden et al., 2015; Rozeboom, 1942) and seasonal fluctuations in abundance (Bradt et al., 2019). However, to our knowledge no studies have linked species abundances to environmental factors in this area.

The problem posed by the concentration of these species in urban areas is compounded by the difficulty of tracking these species in heterogeneous urban environments (Murdock et al., 2017), where there can be a large range of microhabitats within the landscape (Pincebourde et al., 2016). Highly heterogeneous landscapes can influence biologically relevant factors in the disease transmission cycle, such as vector density, survivorship, or biting rate (Hamer et al., 2011; LaDeau et al., 2015). Hence, it is crucial to study aspects of the transmission cycle that are meaningful for species occurrence at the relevant scale. Detailed microclimate measurements can provide comprehensive information at specific sampling sites. However, such studies may fail to capture the full range of microhabitats within a heterogeneous urban landscape.

Entomological surveys that include detailed microclimate data are collected at a limited number of sampling sites and times within a study period. Remote sensing data offers an alternative, since satellites provide environmental measurements across large landscapes at consistent time intervals, and over longer time frames. Satellite sensors have drastically improved over the last decades, and long-term data archives of high-quality satellite imagery are increasingly expanding (Wulder et al., 2016). These improvements have led

to the common use of satellite data as a source to study spatiotemporal risk factors of a variety of vector borne diseases (Dlamini et al., 2019; Malone et al., 2019; Wimberly et al., 2021). Additionally, earth observation data from satellites provide standardized measurements that can be converted into ecologically meaningful variables that are relevant for mosquito ecology including land cover, land surface temperature, and rainfall. However, there is often a mismatch between the scales at which satellite data are retrieved and the scales at which entomological surveys are performed. Previous research addressed this issue by combining remotely-sensed data with *in-situ* microclimatic measurements (Wimberly et al., 2020). Here, we compare different environmental datasets to determine if satellite data can be an adequate substitute for climate loggers or weather station data in hard to access locations.

The goal of this study is to compare different sources of environmental data to examine vector mosquito distribution in an urban environment. Our study objectives were to: 1) identify environmental factors that influence abundance of the vector species *Cx. quinquefasciatus* and *Ae. albopictus,* and 2) evaluate how remotely-sensed environmental data compares to *in-situ* measurements in studying mosquito abundance. We achieved this goal by comparing models of mosquito abundance based on environmental data from microclimate loggers, local weather station data, and variables derived from satellite data.

4.2 Materials and Methods

4.2.1 Study area

The study area consisted of urban and rural areas within the city of Norman, OK. Norman has a population of 122,837 and a total area of approximately 460 km^2 (U.S. Census

Bureau, 2019). The municipal boundary encompasses two Level III Ecoregions. The densely populated western part is within the Central Great Plains ecoregion, which is dominated by mixed grass prairie and riparian woodlands. The less populated eastern part is located within the Cross Timbers ecoregion, which consists of dense oak forests with open woodland areas (Woods et al., 2016). Norman, OK has a subtropical, semi-arid climate with average annual minimum temperature of 9.5 ° C, average annual maximum temperature of 21.8 ° C, and annual average precipitation of 987 mm (Oklahoma Climatological Survey, 2021).

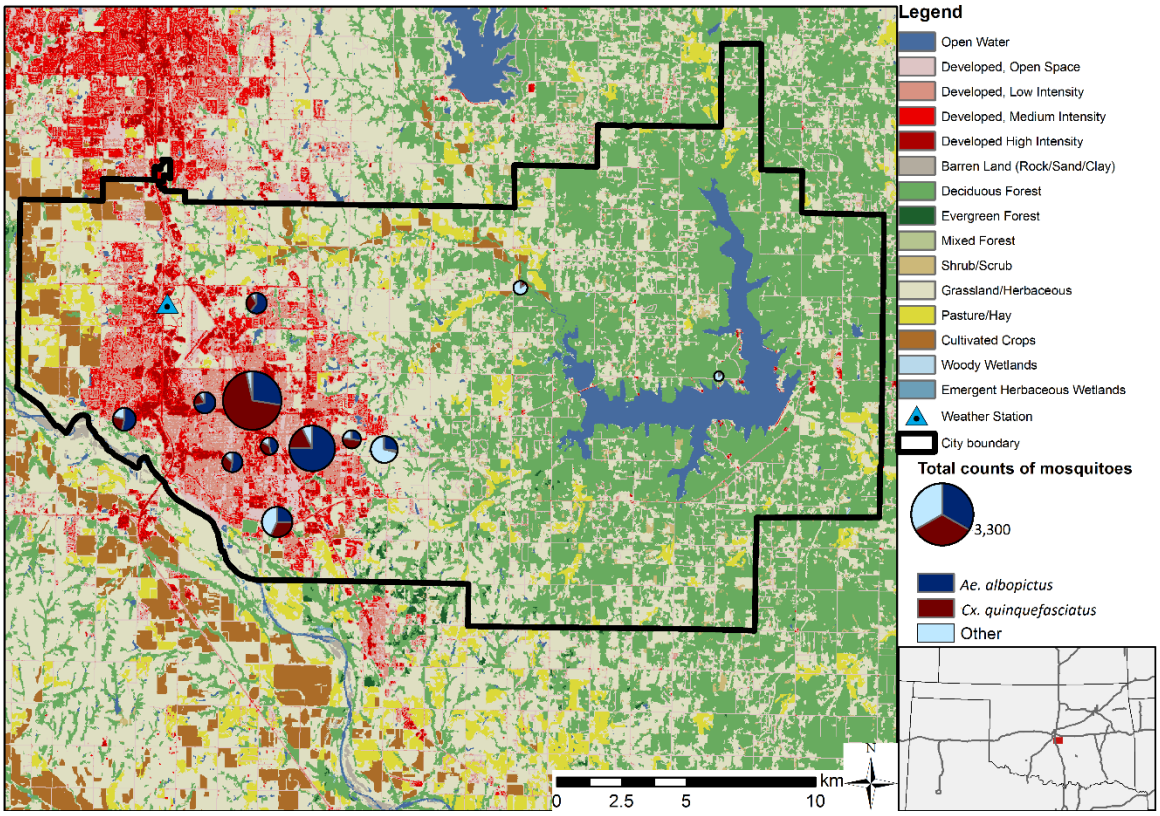

Figure 4-1: Land cover and location of trap sites within the study area with the Mesonet weather station (triangle), as well as the proportion of mosquito species collected at each trap site (pie charts). The size of each circle is proportional to the total mosquito abundance. Land cover data are derived from the 2016 NLCD dataset.

4.2.2 Mosquito trapping and processing

Mosquitoes were sampled between June – October in 2019, and May – November in 2020. We sampled at 11 consistent trap sites in both years, with a 12th site added in 2020. Trap locations were chosen to span a gradient from urban locations dominated by impervious surfaces to rural locations dominated by natural vegetation. Sites were located in public parks, on residential property, and on the campus of the University of Oklahoma. Trapping was performed with BG Sentinel 2 traps (Biogents AG, Regensburg, Germany). Because the octenol based lure has been shown to underestimate mosquito species richness when compared to CO_2- baited traps (Giordano et al., 2021; Wilke et al., 2019), we added dry ice as a CO_2 source to attract more mosquitoes. Mosquitoes were sampled bi-weekly, with traps deployed between 8 am and 11 am and collected 24 hours later. Mosquitoes were collected, immobilized, transported on dry ice, and stored at -80 °C. Adult female mosquitoes were examined under a dissecting microscope and morphologically identified using the taxonomic keys from Darsie and Ward (2005).

4.2.3 Environmental data collection

We obtained static land cover data and dynamic data on relative humidity, precipitation, and temperature. Land cover data was retrieved from the National Land Cover Database (NLCD). Data sources for dynamic climate variables included microclimate loggers that were deployed in the field next to mosquito traps, satellite data from different sensors, and local weather station data from the Mesonet network (Table 4-1).

Landcover data

We retrieved land cover data on urban imperviousness and canopy cover from the 2016 National Land Cover Database (NLCD) (Jin et al., 2019; Yang et al., 2018, 2003). Urban imperviousness and canopy cover are measured as a percentage of every 30-meter pixel. The two land cover variables were summarized as focal means within a 1 km buffer around each trap site.

Microclimate Data Loggers

Table 4-1: Dynamic variables summarized for the three temporal data sources: microclimate data, weather station data, satellite data

Temporal variables	Unit	Microclimate	Satellite	Weather Station
Tmin	C	daily min	VIIRS nighttime LST (1km)	daily min
Tmean	C	average of all 10-minute averaged temperature observations each day	NA	average of all 5-minute averaged temperature observations each day
Tmax	C	daily max	VIIRS daytime LST (1km)	daily max
RHmin	%	daily min	NA	daily min
RHmean	%	daily mean	NA	daily mean
RHmax	%	daily max	NA	daily max
Prec	mm	NA	CHIRPS (4.5km)	liquid precipitation measured each day
NDVI	index	NA	Sentinel 2 and Landsat OLI (30m)	NA
NDMI	index	NA	Sentinel 2 and Landsat OLI (30m)	NA

Microclimate loggers (RFID Track-it, Monarch Instruments) were deployed in close proximity to trap sites to measure temperature and relative humidity. The loggers were equipped with radiation shields to reduce measurement errors (Tarara and Hoheisel, 2007). We deployed two microclimate loggers at each residential trap site to measure conditions

in the front and back yards. We deployed between 4 and 6 loggers within a radius of 100 m at each trap site on the university campus and in public parks. Loggers were placed at heights between 1 m and 1.5 m above ground and were located in areas with and without tree canopies to capture small-scale variability in microclimate. The loggers recorded data every 10 minutes from mid-May to the end-November for 2019 and 2020. Of the 41 loggers used in this study, ten malfunctioned and had to be replaced. In four instances this led to missing data for the 4-week period prior to the malfunction date. For each trap location, we calculated the daily minimum, mean, and maximum temperature and relative humidity. Values from all loggers around each trap site were interpolated via inverse distance weighting (Pebesma, 2004). Daily values for all environmental data sources are displayed in Figure 4-2.

Remote-sensing data

To assess the value of satellite data for mosquito-borne disease applications, we gathered data from products derived from satellite observations. We used data from different satellite sensors to be able to compare remote sensing variables with weather station data and microclimate logger measurements**.** Daily rainfall measurements were derived from the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) dataset (Funk et al., 2015). This product provides daily disaggregated rainfall data at a spatial resolution of 0.05° × 0.05° (4.5 km at a Latitude of 35◦ N). Land surface temperature data were obtained from the Visible Infrared Imaging Radiometer Suite sensor (VIIRS) (Hulley and Hook, 2018). We downloaded the VNP21A1D data product for daytime temperatures with observations at 1:30 pm local solar time and VNP21A2N for nighttime temperatures with observations at 1:30 am local solar time. Daytime and nighttime land surface temperature

data are available at daily intervals at a spatial resolution of 1km. We converted temperature values from Kelvin to Celsius. Data tiles were masked via the products' internal quality control layers to exclude cloud contaminated pixels. Vegetation and moisture indices were calculated from harmonized Landsat and Sentinel-2 (HLS) data (Claverie et al., 2018). The dataset consists of Landsat 8, Sentinel 2A and Sentinel 2B imagery, which are combined into a harmonized, analysis-ready surface reflectance dataset. Sentinel images are resampled to match the 30m spatial resolution of Landsat 8 images. With observations from all satellites combined, the HLS dataset has a return interval of two to three days. From these HLS observations, we calculated a Normalized Difference Vegetation Index (NDVI) using the red and near infrared reflectance bands (Tucker, 1979), as well as the Normalized Difference Moisture Index (NDMI), using the near infrared and the shortwave infrared reflectance bands (Gao, 1996).

Data were summarized in a standardized way for all the satellite variables. For days with missing data, we used a linear regression model to impute those missing values for each of the satellite time series. We fit a linear model by robust regression using cyclical splines (Venables and Ripley, 2002). We then replaced missing values with predicted values from the regression model. For each of the satellite variables, we summarized the values as focal means within a 1 km buffer around each trap site to have a standardized measure across all satellite products. This 1 km buffer reflects the maximum flight range of adult mosquitoes of the two target species, reported at approximately 700 m and 1 km for *Ae. albopictus* and *Cx. quinquefasciatus*, respectively (Verdonschot and Besse-Lototskaya, 2014).

Meteorological data

We obtained local meteorological data from the local weather station operated by the Oklahoma Mesonet network (Brock et al., 1995; McPherson et al., 2007). Variables included daily minimum, mean, and maximum values for temperature and relative humidity, as well as daily rainfall. The weather station is located at 35.2556° N, 97.4836° W in the north-western part of Norman (Figure 4-1). The distances between the weather station and the mosquito traps were between 2 and 20 km. Analyses based on weather station data assigned the same time series of meteorological variables to each trap site.

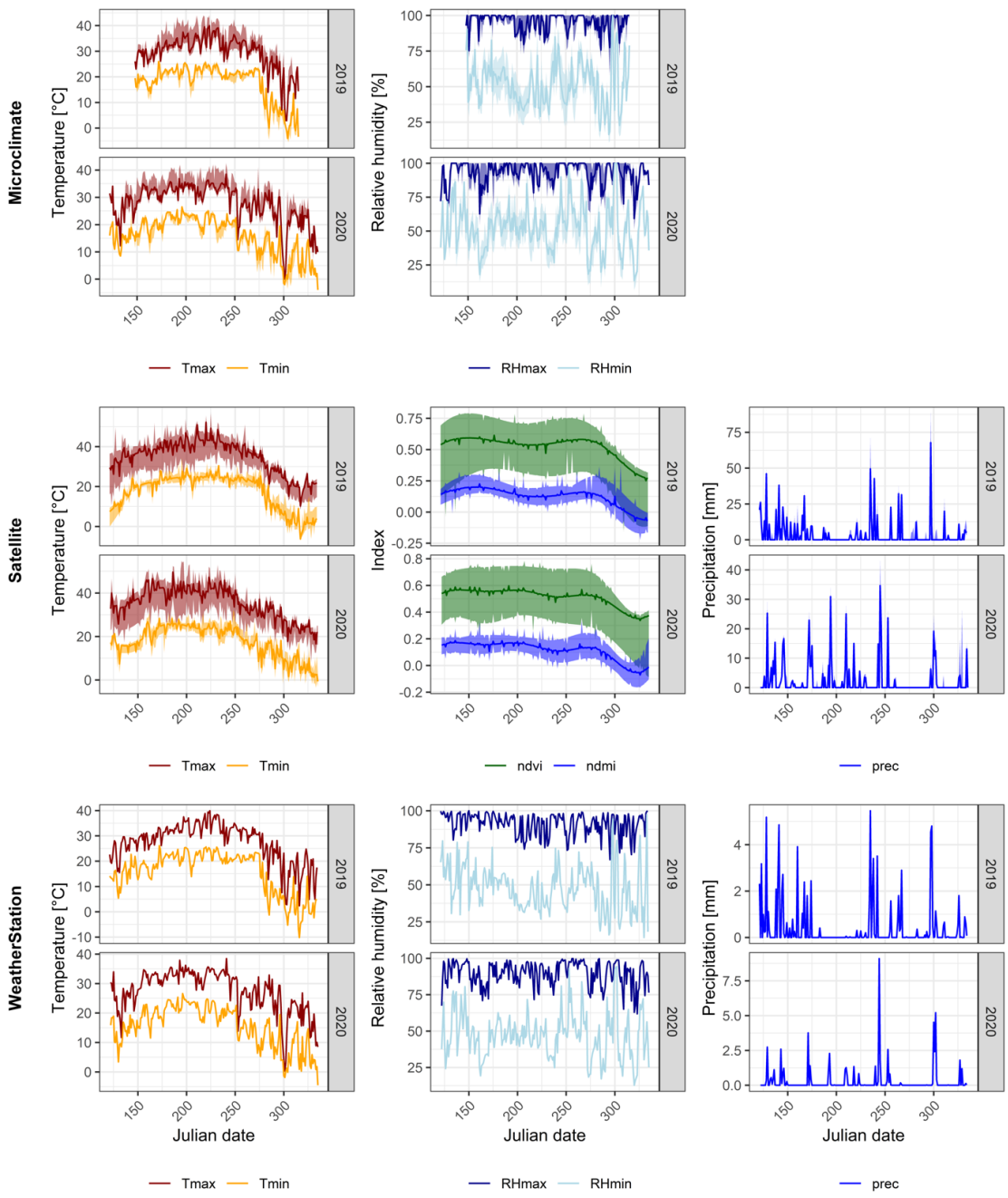

Figure 4-2: Environmental variables for temperature, spectral indices, precipitation, and relative humidity from microclimate loggers (top row), satellite remote sensing (middle row), and weather station data (bottom row). For microclimate data, the solid lines represent the median value and ribbons the minimum and maximum across all sites. For satellite data, the graphs show daily interpolated value for a 1km buffer zone around all trap sites with the solid lines representing the median value and the ribbons the minimum and maximum across all sites. For weather station data, the solid lines show the daily values measured at the location of the weather station.

4.2.4 Analysis

To study the relationships between mosquito abundance and environmental variables, we first used linear models to select the most important predictors for each type of variable and species. We then trained models with selected variables to compare the explanatory power of data from the different sources.

All environmental variables were summarized for each trap day as well as for the one, two, three and four weeks leading up to the trapping day. We then fitted linear models using maximum likelihood estimation (Ripley et al., 2021). To correct for overdispersion, all models were fit using a negative binomial distribution on untransformed count data (O'Hara and Kotze, 2010). Variable selection was done via univariate linear models that were fitted for every combination of variable type (temperature, humidity, precipitation, spectral index) and temporal summary (1-day, 1 week, 2 weeks, 3 weeks, 4 weeks). For each data source (Microclimate, Weather Station, Satellite), we ranked these models by their Akaike information criterion (AIC) value. The AIC is a commonly used information criterion for model selection, that penalizes overparameterization to pick the simplest of the best fitting models (Akaike, 1998; Bozdogan, 1987). We selected the top four variables that were either a daily or weekly summary from each variable type.

We then fitted three multi-variable models with the best four predictors from the microclimate data, weather station data, and the satellite data, as well as an additional multivariate model with only land cover variables. Three additional models were fitted where we added the land cover variables to account for purely spatial differences between trap sites. To compare the effect sizes of the variables, we standardized them by rescaling to a mean of zero and a standard deviation of one. Lastly, we calculated AIC statistics and

Akaike weights (Wagenmakers and Farrell, 2004) to compare the multivariate models for each mosquito species. Furthermore, the Nagelkerke pseudo-R^2 (Nagelkerke, 1991) was calculated to allow comparison of models for different species.

4.3 Results

We collected a total of 8,368 adult female mosquitoes across 12 trap sites in 2019 and 2020 (Table 4-1). *Ae. albopictus* was the most abundant species over the entire time frame, with 3,592 adult female mosquitoes (42.9% of total counts). *Cx. quinquefasciatus* was the second most abundant species with 3,256 adult female mosquitoes (38.9% of total counts). *Ps. ferox* made up 3.7% of total abundance for both years, and *Ae. trivittatus* 2.72%. All other species combined made up less than 2% of the total mosquito abundance. When broken down by year, *Ae. albopictus* was most abundant in 2019 with 1,534 mosquitoes (46.5% of 2019 counts). *Cx. quinquefasciatus* was most abundant in 2020 with 2,438 mosquitoes collected (48.1% of 2020 counts).

Table 4-2: Mosquito abundances by species for 2019 and 2020. Displayed are abundances for all species, as well as their relative abundance.

Species	2019	%	2020	%	total	%
Aedes albopictus	1,534	46.47	2058	40.62	3592	42.93
Culex quinquefasciatus	818	24.78	2438	48.12	3256	38.91
Psorophora ferox	224	6.79	85	1.68	309	3.69
Aedes trivittatus	183	5.54	45	0.89	228	2.72
Culiseta inornata	60	1.82	106	2.09	166	1.98
Culex salinarius	74	2.24	36	0.71	110	1.31
Aedes vexans	41	1.24	67	1.32	108	1.29
Anopheles quadrimaculatus	64	1.94	44	0.87	108	1.29
Aedes triseriatus	81	2.45	24	0.47	105	1.25
Culex nigripalpus	57	1.73	46	0.91	103	1.23
Psorophora columbiae	37	1.12	31	0.61	68	0.81
Anopheles pseudopunctipennis	25	0.76	12	0.24	37	0.44
Culex tarsalis	21	0.64	13	0.26	34	0.41
Anopheles punctipennis	7	0.21	13	0.26	20	0.24
Psorophora cyanescens	7	0.21	5	0.10	12	0.14
Orthopodomyia signifera	1	0.03	1	0.02	2	0.02
Culex spp	22	0.67	23	0.45	45	0.54
Aedes spp	20	0.61	2	0.04	22	0.26
Anopheles spp	1	0.03	1	0.02	2	0.02
Unknown	24	0.73	17	0.34	41	0.49
Total	3,301	100.00	5,067	100.00	8,368	100.00

Species composition varied considerably by trap site and throughout the season (Figure 4-3). *Ae. albopictus* had higher counts per trap week for almost the entire study period except for June 2019, August 2020, and October and November 2020, when we observed higher *Cx. quinquefasciatus* counts. Particularly in 2020, one trap site had consistently high counts of *Cx. quinquefasciatus*, that outnumbered *Ae. albopictus* even in the warmer summer months. However, the average counts of *Cx. quinquefasciatus* across all sites were lower than those of *Ae. albopictus* in most months. Seasonal patterns were different in the two years studied. In 2019, *Cx. quinquefasciatus* had the highest count in June, followed lower numbers later in the summer and a spike in numbers in September. In 2020, *Cx. quinquefasciatus* numbers increased steadily between June and August, decreased in September, and increased again in October. *Ae. albopictus* counts had a comparable pattern in 2019, with highest case numbers in September, and in 2020 with rising case numbers between May and August and lower numbers in September followed by a late peak in early October.

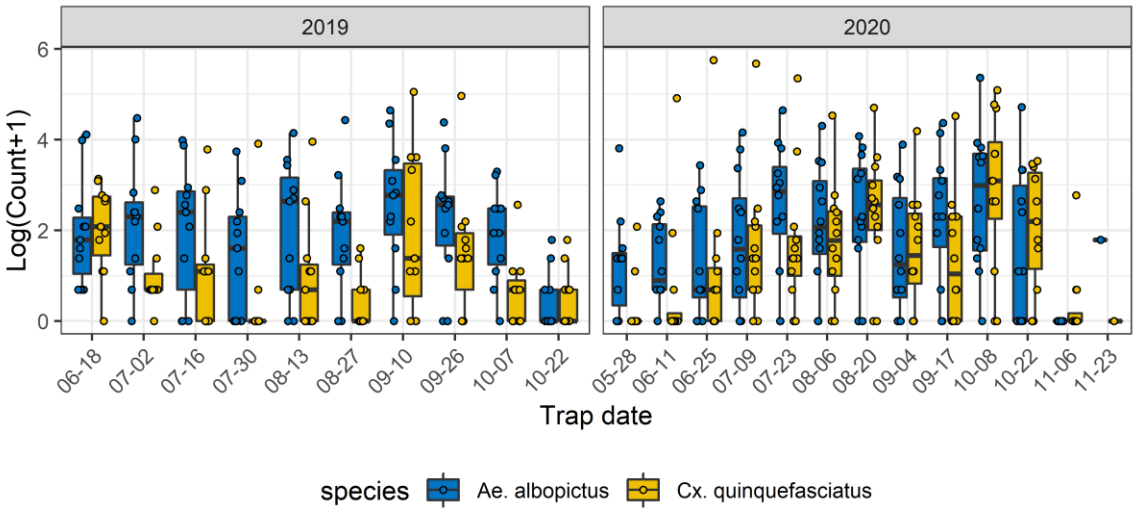

Figure 4-3: Mosquito counts per trap date for *Ae. albopictus* and *Cx. quinquefasciatus*. Box plots display the distribution of counts across trap sites for each trap date.

Temperature and relative humidity from microclimate and weather station measurements on the trap day had a positive effect on mosquito abundances (Figure 4-4). However, weekly relative humidity had a negative association with mosquito abundance with *Ae albopictus* responding to relative humidity a week prior to trap date and *Cx quinquefasciatus* responding to humidity three weeks prior to the trap date. Additionally, both species had higher mosquito counts when NDMI values in the preceding week were high and when precipitation was low two weeks prior to the trapping date. For both species, counts were positively associated with % impervious surface. *Ae. albopictus* counts were also negatively associated with % canopy cover, whereas *Cx. quinquefasciatus* counts were positively associated with % canopy cover.

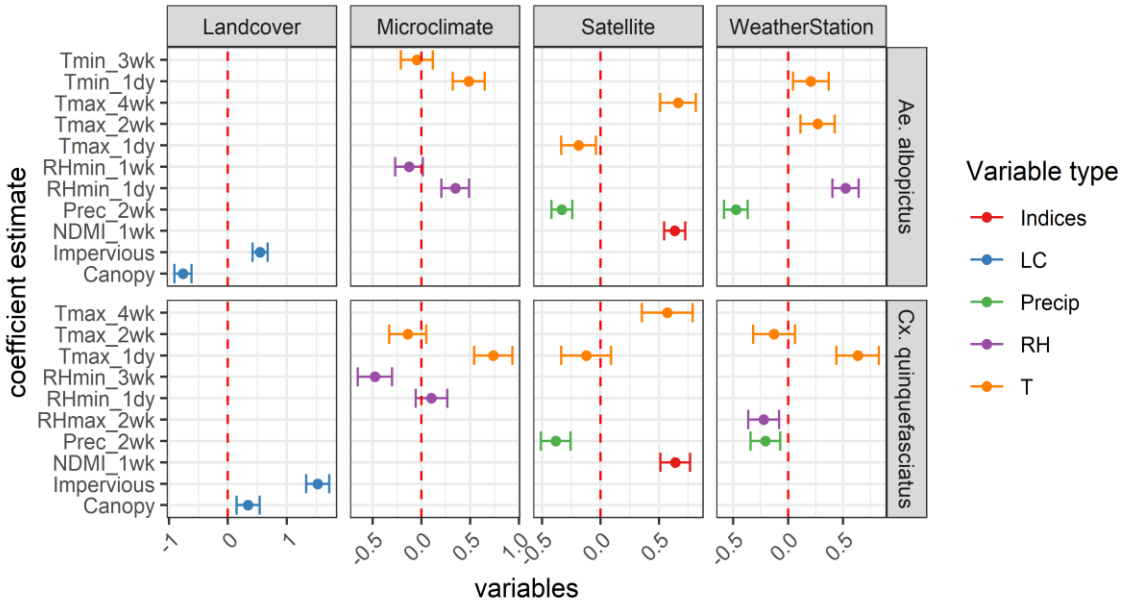

Figure 4-4: Coefficient estimates of multivariate models for each data type (spectral indices, landcover, precipitation, relative humidity, temperature), and species (Ae.albopictus and Cx. quinquefasciatus)

When comparing models based on different data sources, the land cover model was the most parsimonious models with lowest AIC and highest pseudo R^2 values for each species (Table 4-2). For *Ae. albopictus* this model yielded an Akaike weight of close to one and a pseudo R^2 of 0.49. For *Cx. quinquefasciatus*, this model also yielded an Akaike weight of close to one and a pseudo R^2 of 0.41. Satellite time series data yielded the second-best model for both species with an Akaike weight of 9.69e-08 and a pseudo-R^2 of 0.37 for *Ae. albopictus*, and an Akaike weight of 1.46e-11 and a pseudo-R^2 of 0.22 for *Cx. quinquefasciatus*. For *Ae. albopictus*, the third best model was from Mesonet weather station data, leaving the microclimate logger data as the worst model. For *Cx. quinquefasciatus*, the opposite was observed with microclimate logger data yielding a better fitting model than Mesonet weather station data.

Table 4-3: Comparison of negative binomial fixed effect models for each variable type individually: Landcover data from NLCD data set, microclimate loggers, Mesonet weather station data, and satellite time series data.

	Date source	AIC	delta AIC	Akaike weight	Pseudo R^2	Model
Aedes	Land cover	1581	0	1.00E+00	0.49	canopy + impervious
	Satellite	1661	32.3	9.69E-08	0.37	ndmi_1wk + Tmax_4wk + Tmax_1dy + prec_2wk
	Weather station	1652	70.6	4.67E-16	0.17	Tmin_1dy + Tmax_2wk + RHmin_1dy + prec_2wk
	Microclimate	1662	80.8	2.85E-18	0.11	Tmin_1dy + Tmin_3wk + RHmin_1dy + RHmin_1wk
Culex	Land cover	1359	0	1.00E+00	0.41	canopy + impervious
	Satellite	1400	40.6	1.53E-09	0.22	ndmi_1wk + Tmax_4wk + Tmax_1dy + prec_2wk
	Microclimate	1409	49.9	1.46E-11	0.17	Tmax_1dy + RHmin_3wk + Tmax_2wk + RHmin_1dy
	Weather station	1421	62.5	2.68E-14	0.09	Tmax_1dy + prec_2wk + RHmax_2wk + Tmax_2wk

Adding land cover data on canopy cover and impervious surface cover to the time series summary models led to lower AIC values and higher pseudo R^2 values for all models. Satellite time series data in combination with land cover data yielded an Akaike weight of close to one and the highest R^2 for both species. However, the difference between the best model and the second-best model was much larger for *Ae. albopictus* than for *Cx.*

quinquefasciatus. The over-all highest pseudo-R^2 values were reached by models with *Ae. albopictus* as response variable.

Table 4-4: Comparison of negative binomial fixed effect models for each temporal variable type in combination with land cover data

	Data source	AIC	delta AIC	Akaike weight	Pseudo R^2	Model
Aedes	Satellite	1527	0	1.00E+00	0.7	ndmi_1wk + Tmax_4wk + Tmax_1dy + prec_2wk + canopy + impervious
	Weather station	1555	28.3	7.16E-07	0.61	Tmin_1dy + Tmax_2wk + RHmin_1dy + prec_2wk + canopy + impervious
	Microclimate	1571	43.8	3.08E-10	0.56	Tmin_1dy + Tmin_3wk + RHmin_1dy + RHmin_1wk + canopy + impervious
Culex	Satellite	1344	0	9.96E-01	0.51	ndmi_1wk + Tmax_4wk + Tmax_1dy + prec_2wk + impervious + canopy
	Microclimate	1356	11.2	3.68E-03	0.46	Tmax_1dy + RHmin_3wk + Tmax_2wk + RHmin_1dy +impervious + canopy
	Weather station	1360	15.6	4.08E-04	0.44	Tmax_1dy + prec_2wk + RHmax_2wk + Tmax_2wk + impervious + canopy

Land cover data captured spatial variability of both mosquito species, but no temporal variation. Microclimate logger data captured almost no spatial variability of *Ae. albopictus* counts but did capture spatial variability in *Cx. quinquefasciatus* counts (Figure 4-5). Mesonet variables, having been measured at a single location, captured temporal variability, but no spatial variability. Variables included in the satellite time series model were able to capture temporal, as well as spatial variability for both species. When land cover variables were added to each model, all models were able to capture more spatial variability than comparable models without land cover variables.

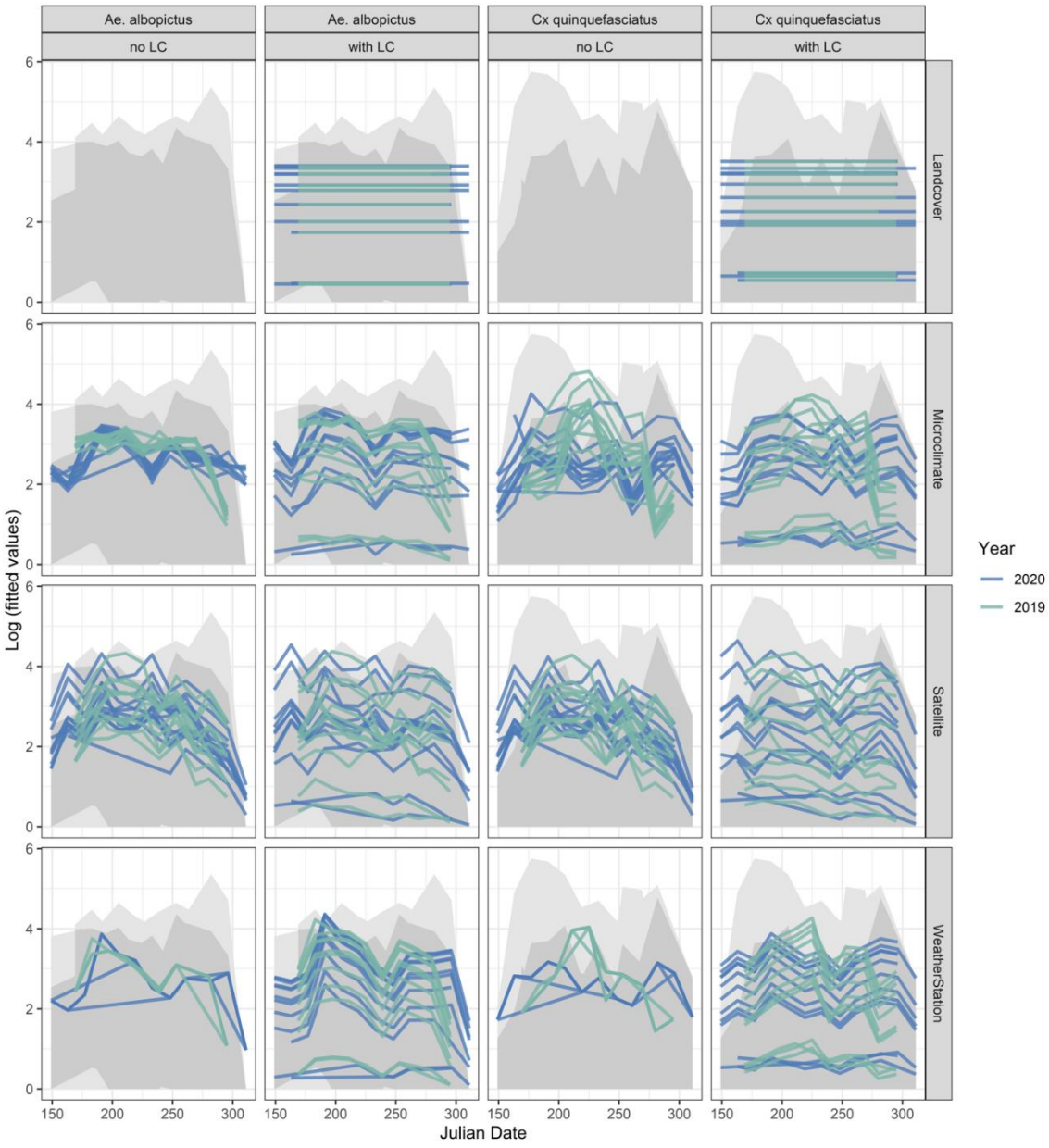

Figure 4-5: Fitted values for all models including models based on land cover data, microclimate, satellite, and weather station data for *Ae. albopictus* and *Cx. quinquefasciatus*. Model based on microclimate, satellite, and weather station data were fitted with land cover data (with LC) and without land cover data (no LC). Different lines represent spatial variation among trap sites and each line represents variation through time. Blue lines represent fitted values for 2020, green lines for 2019. The grey shaded areas represent the range of mosquito counts across all traps for each year.

4.4 Discussion

This study aimed to understand mosquito abundances for the two primary arboviral vector mosquitoes (*Ae. albopictus* and *Cx. quinquefasciatus*) in an urban area in central Oklahoma. Our models based on *in-situ* and satellite derived environmental variables and mosquito abundances for *Ae. albopictus* and *Cx. quinquefasciatus* showed that the two species have similar habitat requirements in our study area. Both species showed a positive association with temperature and relative humidity on the trap day. We attribute this relationship to influences of temperature (Nance et al., 2018; Veronesi et al., 2012) and humidity (Veronesi et al., 2012) on host seeking behavior. Both species also had stronger relationships with summaries of daily minimum and maximum temperatures than with summaries of mean temperatures. This result is in accordance with previous findings that show how mosquito species react to ranges between minimum and maximum temperatures, as opposed to mean values (Carrington et al., 2013; Metelmann et al., 2019).

Both species had negative associations between species abundance and precipitation at a 2-week time lag. This 2-week time lag is possibly driven by the effect of breeding site flushing and ovipositional repellency from heavy rainfall events (Dieng et al., 2012). The flushing effect of breeding sites of anthropophilic urban mosquitoes was found to be stronger in cities with more build-up area, (Rydzanicz et al., 2016), as these tend to be areas where there is less infiltration and runoff is more extreme (Walsh et al., 2012). Particularly in central Oklahoma, where rainfall often occurs in extreme weather events, this effect may be exacerbated so that mosquito populations breeding is interrupted by flooding events, which may lead to decreased numbers of adult mosquitoes two weeks after a heavy rainfall event. However, both species are however positively associated with impervious surfaces.

This effect has been previously observed for both, *Cx quinquefasciatus* (Landau and van Leeuwen, 2012), and for *Ae albopictus* (Shragai and Harrington, 2019). The main difference in microhabitats between the two species was *Ae. albopictus*' negative association with tree cover, and *Cx. quinquefasciatus*' positive association with tree cover.

We were also able to compare different sources of environmental data commonly used in entomological surveys. All four data source types vary in what variables they measure, how much spatial and/or temporal variation they measure, and how much spatial information they provide.

All three dynamic data source types provide very comparable results once paired with land cover data on impervious surface and canopy cover. We found this common effect, despite all data sources varying in what environmental variables they measure. Satellite time series data performed best, likely because spectral indices like NDVI and NDMI capture aspects of land cover that have been shown to drive site-to-site variation and are additionally able to capture temporal variation. The NDVI measures vegetation greenness, and the NDMI measures vegetation moisture, both of which are positively correlated with canopy cover and negatively associated with impervious surfaces. However, even with spectral indices present in the satellite model, adding additional land cover variables was still beneficial. Additionally, both satellite time series data and the weather station provided information on precipitation, which revealed that precipitation two weeks prior to the trap date negatively influenced mosquito abundance. For models based on microclimate logger data, this relationship was reflected in a negative association between relative humidity one week and three weeks prior to the trap date.

The second major difference between the data types (land cover, microclimate, satellite, and weather station data), was the varying degree of spatial versus temporal information they provided. This ultimately determines how well they capture spatial and temporal variability in species abundance. Even though land cover data on canopy cover and impervious surfaces only captured spatial variability in species abundance it did so comparatively well for both species, making them the best stand-alone models. This shows the strong importance of land cover variables as predictors of the geographic pattern of mosquito abundance. Satellite data were able to capture spatial, as well as temporal variation in species abundance for both species. This did not come as a surprise, as satellite variables capture spatially explicit information of multiple variables for areas around trap sites representing mosquito activity ranges. Microclimate loggers captured some spatial variability for *Cx. quinquefasciatus*, but not for *Ae. albopictus*. This means that there was a spatial component of temperature and relative humidity that influenced *Cx. quinquefasciatus*, but not *Ae. albopictus*. Since *Cx. quinquefasciatus* was positively related to canopy cover, which is known to affect microclimatic conditions (Howe et al., 2017; Wimberly et al., 2020), we hypothesize that the spatial component explained by microclimate logger data could be related to land cover. Unsurprisingly, weather station data by itself, measured at a single location in town, could not explain any of the spatial variability of abundance for either species. However, weather station data for mapping variation in mosquito abundance over time can be combined with land cover data to predict variation in both space and time.

All data sources measured a different scale of spatial information. The microclimate measurements were taken directly at the mosquito trap sites and were better predictors than

weather station data for *Culex quinquefasciatus*, but not *Aedes albopictus*. However, the satellite data always had the best fits for both species even though they were summarized over a radius of 1km. Our results are in accordance with those of Hopperstad et al. (2021) who found that species distribution of *Ae. albopictus* were better predicted by coarser scale remotely sensed land cover variables than by finer scale microclimate data (Hopperstad et al., 2021). We hypothesize that microclimate loggers placed in a small radius around mosquito traps may not be capturing the full range of available microhabitats within the activity range of mosquitoes. Satellite data on the other hand provided a coarser-brained view on the landscape, but were able to capture a larger extent that may better reflect the activity range of a mosquito. Considering that *Culex* species have been found to travel up to 3km during their lifetime (Hamer et al., 2014), and *Ae. albopictus* up to 1 km (Vavassori et al., 2019), the scales at which medium resolution satellite missions like VIIRS and MODIS gather data (250-1000m), appear to be suitable for predicting mosquito abundance.

There are several limitations to this study that we need to take into consideration. Constraints regarding the study design arise from the traps used, the time frame sampled, and the number of traps deployed. We used BG Sentinel traps, which are designed for artificial container breeding *Aedes* species. However, we added dry ice as additional bait, which has been shown to increase abundance and diversity of mosquitoes caught (Wilke et al., 2019). Furthermore, our analysis was only conducted over a two-year period. Further research over longer time frame is desirable to confirm our results and to investigate how variation between years affects the results. We also had a limited number of traps, which we placed primarily in state parks and on private properties. A larger number of traps over a wider range of microhabitats would be desirable. The microclimate measurements were

taken using small, inexpensive data loggers that have an accuracy of up to ± 0.5˚C (Wimberly et al., 2020). Some microclimate loggers failed and had to be replaced mid-season which could have introduce non-random errors into the analysis.

In addition, we have not considered socio-economic factors that may influence mosquito abundance. It is known that factors such as the age of a neighborhood (Spence Beaulieu et al., 2019), household income and education level (Little et al., 2017; Whiteman et al., 2019), as well as ownership and use status (Lambin et al., 2010) can influence mosquito distribution. These socio-economic differences are related to persistence and quality of vector habitat within heterogeneous urban areas, which can lead to distinct patterns of species composition and abundance between neighborhoods of differing socio-economic status (LaDeau et al., 2013). For example, the abundance of uncovered trash containers and old car tires as breeding sites are more likely to be found in low-income areas, whereas containers for recreational use are the main breeding sites in high income areas (LaDeau et al., 2013). Differences in water accumulation in rain-fed unmanaged containers and containers that are purposefully watered can also lead to seasonal differences in mosquito abundance between neighborhoods of socio-economic status (Becker et al., 2014). Socio-economic factors may also influence how likely it is that households or neighborhood associations control for mosquito populations. Because the scope of this research was to identify environmental drivers, and compare sources of environmental data, we did not include socioeconomic data. Future studies could include additional data on irrigation, mosquito control practices, and artificial containers at household and neighborhood levels.

In conclusion, this study showed the strong influence of impervious surface and forest cover on species abundance of *Ae. albopictus* and *Cx. quinquefasciatus* in Norman,

Oklahoma. Our results also indicate that satellite-based models produced the best fitting models of mosquito abundance and were able to capture spatial as well as temporal variation. The fit of models based on microclimate and weather station data improved substantially when they were combined with spatially explicit land cover information. Remotely sensed variables act as useful data source to study species abundance across space and time, are important predictors in areas where *in-situ* climate data are not available, and in some instances may even be better predictors than *in-situ* environmental variables.

5 Synthesis

5.1 Summary

When and where humans are at risk of contracting mosquito-borne diseases is influenced by environmental conditions. Climatic aspects, such as rainfall, temperature and humidity, as well as landscape composition including land cover, topography, and hydrology, can all influence disease transmission cycles. However, interactions between landscape, climate and disease transmission are complex and can be context-dependent. The use of remote sensing technology to study these associations can help to disentangle some of these effects in order to predict disease risk across space and time. The research presented in this dissertation aims at utilizing satellite remote sensing technology to explore ways how satellite data can be converted into ecologically meaningful variables that help us understand what drives disease risk spatially and temporally. Within this context, this research aimed at answering three main questions:

1. What remotely-sensed variables are most associated with the distribution of mosquito-borne diseases?
2. How do these associations depend on the geographic setting?
3. How do remotely-sensed environmental variables compare to other environmental data sources in understanding disease transmission patterns?

Chapter 2 integrated the use of static and dynamic remotely sensed variables to investigate associations between malaria cases and environmental factors that drive disease risk in two study areas in Ethiopia. We used multiple geospatial environmental datasets, including products from the satellite missions MODIS, Landsat, GPM, and PlanetScope. All

environmental data were summarized for each *kebele*, which is the smallest administrative level in Ethiopia at which malaria case data are reported. We also evaluated environmental impact on malaria transmission for two geographically distinct study areas separately. This approach allowed us to evaluate the influence of each environmental variable at a scale relevant for public health decision makers, while elucidating regional differences.

By using a wide range of environmental data, we identified different variables that influenced malaria transmission in these two different geographic areas. Malaria cases followed a strong temperature gradient in one study area, and a precipitation gradient in another study area. Additionally, settlement structure influenced malaria cases differently in the two areas, possibly due to the differences in population density and how households and agricultural areas are distributed across the landscape. We were thus able to identify how spatio-temporal associations between disease and environmental conditions can vary between different geographies.

Chapter 3 compared the use of three different types of geospatial environmental data in mapping transmission risk of West Nile virus in South Dakota, USA. The data types were grouped as 1) land cover and physiography data including from the National Land Cover Database (NLCD), the National Wetlands Inventory (NWI), the Soil Survey Geographic Database (SSURGO), as well as the National Elevation Dataset (NED), 2) climate data from the North American Land Data Assimilation System atmospheric forcing dataset (NLDAS), and 3) spectral data derived from the MODIS MCD43A4 product. Georeferenced human case data were available, which allowed us to study disease-environment relationships without having to aggregate over larger areas. Relying on

remotely-sensed data also allowed us to retrospectively access environmental information within a of 750 m radius of a positive WNV patient's home.

This research revealed the importance of inter-annual variability of climatic and surface moisture variables for WNV transmission risk. Among the most important variables explaining transmission risk were the inter-annual variability of relative humidity during late transmission season, and of Tasseled Cap wetness and precipitation during early transmission season. Contrary to our expectations, temperature did not drive the spatial pattern of disease risk. This shows how landscape configuration and hydrology play vital roles in disease transmission cycles. Climatic variables, such as temperature, which are known to be important drivers of drive temporal patterns of WNV risk, do not necessarily influence spatial patterns.

Our combined WNV model, which included gridded climate data, spectral indices from remote sensing data, as well as land cover and physiography data, resulted in a higher cross-validated accuracy than any of the models that were based only one variable type. These findings show that integration different sources of geospatial data can improve predictions, as each of these data sources captures a different aspect of the disease transmission cycle.

Chapter 4 examined how vector mosquito abundance varies in space and time in an urban environment in central Oklahoma. For this study mosquitos were collected at study sites along an urban-rural gradient between June and October of 2019 and 2020. In contrast to studies based on retrospective reports of human cases, this also allowed us to take *in-situ* measurements of microclimatic conditions via loggers that were located within a 100 m radius around trap sites and measured temperature and relative humidity every 10 minutes throughout the trapping season.

This study also shows how satellite remote sensing data compare to *in-situ* microclimate logger data, as well as local weather station data in explaining mosquito abundances. Remotely sensed data on land surface temperature, precipitation and spectral indices were derived from VIIRS, CHIRPS and harmonized Landsat-Sentinel 2 data products respectively.

This study revealed a clear spatial pattern in the abundance of the vector mosquitoes *Ae. albopictus* and *Cx. quinquefasciatus*, which was largely driven by the degree of urbanization and tree cover surrounding the trap sites. Remote sensing data produced models with the highest penalized fit to the historical data. They provided environmental information that explained spatial and temporal variation in vector mosquito abundances, measured environmental variables at a broader scale, and measured variables that can't easily be obtained *in situ*, such as seasonal variation in vegetation water content and precipitation.

5.2 General considerations

The research in chapters 2–4 illustrated ways in which remote sensing data can help us understand landscape-scale disease risk and vector mosquito habitat. It showed how environmental variables derived from satellite data can help us improve our understanding of disease risk in different geographic settings, and how well remote sensing data can predict disease risk and vector mosquito abundance, compared to other data sources. This research adds to the existing body of knowledge surrounding Earth observation technology for human health applications and highlights how remotely sensed environmental variables can help us understand the distribution of different diseases in different landscape contexts.

5.2.1 Remotely-sensed variables influencing disease risk

One of the major findings in all three case studies was the importance of land cover and landscape composition in disease transmission. Each study identified specific land cover characteristics that provide favorable conditions for disease transmission. In Ethiopia, areas with low vegetation cover or low surface moisture index values were identified as stable malaria hot spots in the two study areas. In South Dakota, low lying pastoral areas with a high inter-annual variability in surface water within the James River valley were identified as high-risk areas for WNV. In Norman, Oklahoma, impervious surface area and canopy cover explained a large portion of the spatial variability of vector mosquito abundances.

One of the challenges in working with remote sensing data is that satellites measure a variety of characteristics of the earth surface, but these measurements often do not directly relate to the biology of mosquitoes and the life cycle of mosquito-borne diseases. For example, satellite data can tell us how green an area is, or how much precipitation is recorded. However, these satellite measurements typically do not capture the specific features relevant to mosquitoes, including the exact habitat characteristics such as the extent or quantity of larval habitat within a landscape. Most current satellite missions are not able detect small waterbodies like unconnected semi-permanent ponds, which act as breeding sites for the WNV vector mosquito *Cx. tarsalis* (Mercer et al., 2005; Skaff et al., 2017), isolated pools that remain in the landscape as surface water recedes after flooding events and are preferred breeding sites of the malaria vector mosquito *An. arabiensis* (Hamza and Rayah, 2016), or the artificial breeding sites of container breeders such as *Ae. albopictus*.

However, satellites provide data on land surface characteristics that can be related to these conditions that directly influence the life cycle of mosquitoes and the diseases they carry. For example, the Normalized Difference Moisture Index (NDMI), sometimes also referred to as the Modified Normalized Water Index (MNDWI), was important in all three studies. Calculated from reflectance in the near infrared and short-wave infrared wavelengths, NDMI is sensitive to vegetation liquid water (Gao, 1996). Based on the results presented here, it is a useful metric to relate moisture available across landscapes to mosquito-borne diseases. Inter-annual variability of spring-time NDMI predicted WNV cases in South Dakota, NDMI during dry season and transition season was related to malaria in Ethiopia, and NDMI values a week prior to the trap date explained differences in vector mosquito abundances in Norman, Oklahoma. In all of these examples, this measure of vegetation moisture was able to explain aspects of the observed patterns, even though it only indirectly captures water availability in the landscape and does not directly provide information on available breeding sites.

Also notable were significant difference in the influence of settlement structure or presence of developed urban land between the three case studies. Whether settlements were clustered or dispersed across the landscape influenced malaria transmission differently in the two study areas in Ethiopia. The risk of WNV in South Dakota was higher in rural areas with pastoral land than in urban areas. In addition, the coverage of impervious surfaces around mosquito traps had a positive effect on the abundance of the two vector species *Ae. albopictus* and *Cx. quinquefasciatus* in Norman, Oklahoma. These relationships are relevant as we expect global land cover and land use change to continue (Pielke, 2005), particularly with further population growth and the expansion of urban areas (UNDESA,

2019). Knowing how urbanization influences disease transmission and how these relationships depend on the ecology of the mosquito species as well as the social and economic context will help us identify future areas at risk of mosquito-borne diseases.

5.2.2 The importance of landscape context

Climatic variables, most notably temperature, precipitation, and relative humidity, are important factors in variation of disease cases or mosquito abundances. However, it is worth noting that their influence must be evaluated in the context of a given landscape. On a regional scale in the Amhara region of Ethiopia, the transmission risk was mostly influenced by seasonal vegetation and surface moisture conditions, as well as settlement structure. Although temperature and precipitation also influenced disease risk, those effects were only secondary after the effect of the landscape setting. At the state-level in South Dakota, the combined model showed that elevation was the most important variable. Even though climatic variables such as relative humidity and precipitation were identified as important variables, the disease risk patterns across the landscape strongly followed the pattern of low-lying rural areas within the James River valley which are dominated by pastoral landscapes with high ponding frequency. On a municipal scale in Norman, Oklahoma, the spatial patterns of mosquito abundance followed land cover differences between trap sites, whereas climatic variables mostly explained inter-annual variation. These findings suggest that in areas where climatic conditions are not a limiting factor, the composition of a landscape is important in explaining potential disease risk.

The knowledge of the relevance of land cover and physiography for mosquito-borne disease transmission is also crucial when we try to evaluate potential future spread of vector mosquitoes and the diseases they can transmit. Even though the changing range of

mosquito populations and areas of potential disease transmission was outside the scope of this dissertation, it is important to point out the implications of these results for research that projects future risk of mosquito-borne diseases or future habitat ranges of vector mosquitoes. Many studies are already taking on the challenging task of predicting the potential spread of vector borne diseases or disease vectors in response to climate change (Brown et al., 2015; Fischer et al., 2013; Paz, 2015; Ryan et al., 2019). However, there is increasing evidence that the influence of climate factors on disease transmission depends on the geographic context (Olson et al., 2009; Skaff and Cheruvelil, 2016), indicating that future studies would benefit from taking non-climatic factors into account when projecting shifts in ranges of mosquito-borne diseases (Rocklöv and Dubrow, 2020). From the research presented here, these non-climatic factors include land cover, topography and hydrology. We suggest that projections on future spread of mosquitoes and mosquito-borne diseases should take these factors into account to create more accurate predictions of where we can expect future disease outbreaks.

5.2.3 The case for remote sensing data in disease risk applications

In addition to adding to the knowledge of complex associations between environmental conditions and disease risk, this study has highlighted some unique advantages of using remote sensing technology for this field of application. Remote sensing data has shown to be a valuable source of environmental information that can help us understand the risk of mosquito-borne diseases better.

This research has shown how much environmental information can be gained from publicly accessible satellite data when direct measurements such as direct field observations or weather station data are not available or difficult to obtain. For example, we were able to

gather environmental information for remote areas in Ethiopia where continuous and systematic ground-based environmental monitoring is difficult to conduct. The availability of spatially continuous remote sensing data also made it possible to undertake predictive mapping of West Nile virus risk in South Dakota.

We could also show that even when we collected environmental data via microclimate loggers to study vector mosquito abundances, satellite derived variables produced better models. This finding shows that satellite data a viable alternative when studying mosquito populations over larger areas where *in-situ* environment data collection becomes costly and time-consuming, as well as in hard-to-access locations that may lack resources for systematic environmental monitoring.

This research also highlighted how flexible the use of remotely sensed environmental data can be. Data can be aggregated to specific administrative units like the malaria data in Ethiopia, be associated with point locations of disease cases as West Nile virus data in South Dakota was reported, or represent the approximate time and location of vector mosquito activity as in the Norman mosquito study. In all of those cases, remotely sensed environmental variables could be summarized to reflect the environmental conditions within the given administrative unit or within a buffer surrounding where a patient was diagnosed or a vector mosquito was caught.

5.3 Conclusion and future outlook

This work has contributed to our understanding of how land cover and geographic context influence the transmission of mosquito-borne diseases and how remote sensing data can help us study these associations. Remotely sensed variables that were most associated with

the distribution of mosquito-borne diseases depended on the disease and the vector mosquito involved. On a local scale, remote sensing variables that characterized land cover, vegetation conditions or surface moisture were important explanatory variables in all case studies. The geographic setting played a role in Ethiopia, where different variables explained malaria case distribution in two distinct study areas. In comparison with alternative environmental data sources, remote sensing variables added information that couldn't be captured by ground data in all three case studies. However, further research is needed to understand the complex interactions between climatic factors and heterogeneous landscapes across different spatial scales and how these relate to disease transmission.

There will be potential for improved environmental mapping and disease risk applications as new remote sensing data products are becoming available and easier to use. Very high-resolution optical imagery is currently available from commercial earth observation systems like WorldView, Quickbird or GeoEye. Optical imagery at meter or even sub-meter resolutions can help to map medium sized breeding sites such as roadside ditches. Availability of imagery with shorter return intervals can also prove beneficiary to mapping of mosquito-borne disease risk. For example, daily revisit times from the PlanetScope constellation increase the chance of obtaining cloud free imagery. Currently, commercial satellite data products come at high cost. However, with ongoing sensor development, such data may become available at lower costs, which will make them more accessible for public health applications.

The development of new methods to combine publicly available datasets from different sensors can also help public health applications. These methods include the merging of imagery obtained from different sensors at similar wavelengths to create combined data

products of higher temporal resolution or longer data continuity. For example, the harmonized Landsat-Sentinel dataset combines data from two optical sensors into one data product (Claverie et al., 2018). This merging of datasets increases chances of obtaining cloud free imagery while also reducing the need for temporal interpolation to fill gaps between observation dates. A similar approach applied to different sensors could improve the temporal resolution of data available to researchers even further. The integration of data from comparable sensors into one product is additionally being used to create continuous datasets that exceed the life span of individual sensors, as done with the IMERG product that combines two decades of precipitation data (Huffman et al., 2020). With the development of more such datasets and with more time passing, researchers will be able to study trends and temporal anomalies in environmental conditions and how they relate to disease outbreaks in much longer time frames than what is currently possible.

Additionally, sensor fusion techniques can be used to generate finer spatial resolution data from coarser satellite imagery. For example, very high-resolution optical imagery can be used to downscale satellite products that are only available at coarser resolutions, such as thermal imagery or multispectral imagery. Data fusion and downscaling techniques already allow researchers to study environmental conditions at a finer scale (Peng et al., 2017; Wimberly et al., 2020). Such techniques could be applied to entire data archives to make downscaled environmental information more accessible for professionals without a remote sensing background so they could be used for broader public health applications. Future research on disease risk mapping should leverage such new data products and techniques to study complex environment-disease associations and synthesize the information to make

it usable and relevant for public health officials so we can understand, control, and ultimately prevent the transmission of mosquito-borne diseases.

Bibliography

Abeku, T.A., Vlas, S.J.D., Borsboom, G., Teklehaimanot, A., Kebede, A., Olana, D., Oortmarssen, G.J.V., Habbema, J.D.F., 2002. Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best. Trop. Med. Int. Health 7, 851–857.

Adeola, A.M., Olwoch, J.M., Botai, J.O., Rautenbach, C.J. deW, Kalumba, A.M., Tsela, P.L., Adisa, O.M., Nsubuga, F.W.N., 2017. Landsat satellite derived environmental metric for mapping mosquitoes breeding habitats in the Nkomazi municipality, Mpumalanga Province, South Africa. South Afr. Geogr. J. 99, 14–28.

Akaike, H., 1998. Information Theory and an Extension of the Maximum Likelihood Principle, in: Parzen, E., Tanabe, K., Kitagawa, G. (Eds.), Selected Papers of Hirotugu Akaike, Springer Series in Statistics. Springer, New York, NY, pp. 199–213.

Alemu, A., Tsegaye, W., Golassa, L., Abebe, G., 2011. Urban malaria and associated risk factors in Jimma town, south-west Ethiopia. Malar. J. 10, 173.

Animut, A., Balkew, M., Lindtjørn, B., 2013. Impact of housing condition on indoor-biting and indoor-resting Anopheles arabiensis density in a highland area, central Ethiopia. Malar. J. 12, 393.

Aschale, Y., Mengist, A., Bitew, A., Kassie, B., Talie, A., 2018. Prevalence of malaria and associated risk factors among asymptomatic migrant laborers in West Armachiho District, Northwest Ethiopia. Res. Rep. Trop. Med. 9, 95–101.

Asfaw, W., Tolossa, D., Zeleke, G., 2010. Causes and impacts of seasonal migration on rural livelihoods: Case studies from Amhara Region in Ethiopia. Nor. Geogr. Tidsskr. 14.

Ayele, D.G., Zewotir, T.T., Mwambi, H.G., 2012. Prevalence and risk factors of malaria in Ethiopia. Malar. J. 11, 195–195.

Becker, B., Leisnham, P.T., LaDeau, S.L., 2014. A Tale of Two City Blocks: Differences in Immature and Adult Mosquito Abundances between Socioeconomically Different Urban Blocks in Baltimore (Maryland, USA). Int. J. Environ. Res. Public Health Basel 11, 3256–70.

Belay, D.B., Kifle, Y.G., Goshu, A.T., Gran, J.M., Yewhalaw, D., Duchateau, L., Frigessi, A., 2017. Joint Bayesian modeling of time to malaria and mosquito abundance in Ethiopia. BMC Infect. Dis. 17, 415–415.

Benali, A., Nunes, J.P., Freitas, F.B., Sousa, C.A., Novo, M.T., Lourenço, P.M., Lima, J.C., Seixas, J., Almeida, A.P.G., 2014. Satellite-derived estimation of environmental suitability for malaria vector development in Portugal. Remote Sens. Environ. 145, 116–130.

Benedict, M.Q., Levine, R.S., Hawley, W.A., Lounibos, L.P., 2007. Spread of The Tiger: Global Risk of Invasion by The Mosquito Aedes albopictus. Vector-Borne Zoonotic Dis. 7, 76–85.

Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., Myers, M.F., George, D.B.,

Jaenisch, T., Wint, G.R.W., 2013. The global distribution and burden of dengue. Nature 496, 504–504.
Bicout, D.J., Vautrin, M., Vignolles, C., Sabatier, P., 2015. Modeling the dynamics of mosquito breeding sites vs rainfall in Barkedji area, Senegal. Ecol. Model. 317, 41–49.
Bousema, T., Griffin, J.T., Sauerwein, R.W., Smith, D.L., Churcher, T.S., Takken, W., Ghani, A., Drakeley, C., Gosling, R., 2012. Hitting hotspots: spatial targeting of malaria for control and elimination. PLoS Med. 9, e1001165–e1001165.
Bozdogan, H., 1987. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika 52, 345–370.
Bradt, D., Wormington, J.D., Long, J.M., Hoback, W.W., Noden, B.H., 2019. Differences in Mosquito Communities in Six Cities in Oklahoma. J. Med. Entomol.
Brock, F.V., Crawford, K.C., Elliott, R.L., Cuperus, G.W., Stadler, S.J., Johnson, H.L., Eilts, M.D., 1995. The Oklahoma Mesonet: A Technical Overview. J. Atmospheric Ocean. Technol. 12, 5–19.
Brown, H.E., Young, A., Lega, J., Andreadis, T.G., Schurich, J., Comrie, A., 2015. Projection of climate change influences on U.S. West Nile virus vectors. Earth Interact. 19, 1–18.
Bui, Q.-T., Nguyen, Q.-H., Pham, V.M., Pham, M.H., Tran, A.T., 2019. Understanding spatial variations of malaria in Vietnam using remotely sensed data integrated into GIS and machine learning classifiers. Geocarto Int. 34, 1300–1314.
Burakoff, A., Lehman, J., Fischer, M., Staples, J.E., Lindsey, N.P., 2018. West Nile virus and other nationally notifiable arboviral diseases - United States, 2016. Morb. Mortal. Wkly. Rep. 67, 13–17.
Carrington, L.B., Seifert, S.N., Willits, N.H., Lambrechts, L., Scott, T.W., 2013. Large diurnal temperature fluctuations negatively influence Aedes aegypti (Diptera: Culicidae) life-history traits. J. Med. Entomol. 50, 43–51.
CDC, 2017. West Nile virus disease cases by state [WWW Document]. URL https://www.cdc.gov/westnile/statsmaps/preliminarymapsdata2017/disease-cases-state.html (accessed 6.4.18).
CDC, 2016. Final cumulative maps & data for 1999–2016 [WWW Document]. URL https://www.cdc.gov/westnile/statsmaps/cumMapsData.html (accessed 5.24.18).
CDC, 2015. Arboviral diseases, neuroinvasive and non-neuroinvasive | 2015 case definition [WWW Document]. URL https://wwwn.cdc.gov/nndss/conditions/arboviral-diseases-neuroinvasive-and-non-neuroinvasive/case-definition/2015/ (accessed 5.24.18).
Central Statistical Agency, 2012. 2007 population and housing census of Ethiopia. Administrative Report. Addis Ababa.
Chancey, C., Grinev, A., Volkova, E., Rios, M., 2015. The global ecology and epidemiology of West Nile virus. BioMed Res. Int. 1–20.
Chuang, T.W., Hockett, C.W., Kightlinger, L., Wimberly, M.C., 2012. Landscape-level spatial patterns of West Nile virus risk in the northern Great Plains. Am. J. Trop. Med. Hyg. 86, 724–31.
Chuang, T.-W., Wimberly, M.C., 2012. Remote Sensing of Climatic Anomalies and West Nile Virus Incidence in the Northern Great Plains of the United States. PLoS ONE 7, e46882–e46882.

Ciota, A.T., Matacchiero, A.C., Kilpatrick, A.M., Kramer, L.D., 2014. The effect of temperature on life history traits of Culex mosquitoes. J. Med. Entomol. 51, 55–62.
Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161.
Cohen, J.M., Le Menach, A., Pothin, E., Eisele, T.P., Gething, P.W., Eckhoff, P.A., Moonen, B., Schapira, A., Smith, D.L., 2017. Mapping multiple components of malaria risk for improved targeting of elimination interventions. Malar. J. 16, 459.
Cotter, C., Sturrock, H.J., Hsiang, M.S., Liu, J., Phillips, A.A., Hwang, J., Gueye, C.S., Fullman, N., Gosling, R.D., Feachem, R.G., 2013. The changing epidemiology of malaria elimination: new strategies for new challenges. The Lancet 382, 900–911.
Davis, J.K., Gebrehiwot, T., Worku, M., Awoke, W., Mihretie, A., Nekorchuk, D., Wimberly, M.C., 2019. A genetic algorithm for identifying spatially-varying environmental drivers in a malaria time series model. Environ. Model. Softw. 119, 275–284.
Davis, J.K., Vincent, G., Hildreth, M.B., Kightlinger, L., Carlson, C., Wimberly, M.C., 2017. Integrating environmental monitoring and mosquito surveillance to predict vector-borne disease: Prospective forecasts of a West Nile virus outbreak. PLoS Curr. Outbreaks 9, 1–11.
Debebe, Y., Hill, S.R., Tekie, H., Ignell, R., Hopkins, R.J., 2018. Shady business: understanding the spatial ecology of exophilic Anopheles mosquitoes. Malar. J. 17, 351.
Dhiman, S., 2019. Are malaria elimination efforts on right track? An analysis of gains achieved and challenges ahead. Infect. Dis. Poverty 8, 14.
Dieng, H., Rahman, G.M.S., Abu Hassan, A., Che Salmah, M.R., Satho, T., Miake, F., Boots, M., Sazaly, A., 2012. The effects of simulated rainfall on immature population dynamics of Aedes albopictus and female oviposition. Int. J. Biometeorol. 56, 113–120.
Dilts, T.E., 2015. Topography Tools for ArcGIS 10.1. University of Nevada Reno.
Dlamini, S.N., Beloconi, A., Mabaso, S., Vounatsou, P., Impouma, B., Fall, I.S., 2019. Review of remotely sensed data products for disease mapping and epidemiology. Remote Sens. Appl. Soc. Environ.
Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley, B.A., 2000. LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Remote Sens. 66, 849–857.
Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46.
Ebhuoma, O., Gebreslasie, M., 2016. Remote sensing-driven climatic/environmental variables for modelling malaria transmission in sub-saharan Africa. Int. J. Environ. Res. Public. Health 13, 584.
Elith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32, 66–77.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813.
Endo, N., Eltahir, E.A.B., 2020. Increased risk of malaria transmission with warming temperature in the Ethiopian Highlands. Environ. Res. Lett.
Epp, T.Y., Wakdber Cgeryl L., Berke, O., 2009. Predicting Geographical Human Risk of West Nile Virus - Saskatchewan, 2003 and 2007. Can. J. Public Health. 100, 344–348.
Esayas, E., Woyessa, A., Massebo, F., 2020. Malaria infection clustered into small residential areas in lowlands of southern Ethiopia. Parasite Epidemiol. Control 10, e00149.
Ezenwa, V.O., Godsey, M.S., King, R.J., Guptill, S.C., 2006. Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. Proc. Biol. Sci. 273, 109–17.
Farajollahi, A., Nelder, M.P., 2009. Changes in Aedes albopictus (Diptera: Culicidae) populations in New Jersey and implications for arbovirus transmission. J. Med. Entomol. 46, 1220–1224.
Feachem, R.G., Phillips, A.A., Hwang, J., Cotter, C., Wielgosz, B., Greenwood, B.M., Sabot, O., Rodriguez, M.H., Abeyasinghe, R.R., Ghebreyesus, T.A., Snow, R.W., 2010. Shrinking the malaria map: progress and prospects. The Lancet 376, 1566–1578.
Federal Ministry of Health, 2020. National malaria elimination strategic plan: 2021-2025. Addis Ababa.
Fischer, D., Thomas, S.M., Suk, J.E., Sudre, B., Hess, A., Tjaden, N.B., Beierkuhnlein, C., Semenza, J.C., 2013. Climate change effects on Chikungunya transmission in Europe: Geospatial analysis of vector's climatic suitability and virus' temperature requirements. Int. J. Health Geogr. 12, 1–12.
Franklin, J., 2010. Mapping Species Distributions: Spatial Inference and Prediction, Ecology, Biodiversity and Conservation. Cambridge University Press, Cambridge.
Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Stat. Med. 22, 1365–1381.
Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. Data 2, 150066.
Gao, B.C., 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266.
Gerhardt, R.R., Gottfried, K.L., Apperson, C.S., Davis, B.S., Erwin, P.C., Smith, A.B., Panella, N.A., Powell, E.E., Nasci, R.S., 2001. First isolation of La Crosse virus from naturally infected Aedes albopictus. Emerg. Infect. Dis. 7, 807–811.
Giordano, B.V., Allen, B.T., Wishard, R., Xue, R.-D., Campbell, L.P., 2021. Light Trap Collections of Mosquitoes (Diptera: Culicidae) Using Dry Ice and Octenol Attractants in Adjacent Mosquito Control Programs. Fla. Entomol. 103, 499–504.
Gone, T., Balkew, M., Gebre-Michael, T., 2014. Comparative entomological study on ecology and behaviour of Anopheles mosquitoes in highland and lowland localities of Derashe District, southern Ethiopia. Parasit. Vectors 7, 483.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ., Big Remotely Sensed Data: tools, applications and experiences 202, 18–27.

Guintran, J.-O., Delacollette, C., Trigg, P., 2006. Systems for the early detection of malaria epidemics in Africa: an analysis of current practices and future priorities. World Health Organization, Geneva.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186.

Hamer, G.L., Anderson, T.K., Donovan, D.J., Brawn, J.D., Krebs, B.L., Gardner, A.M., Ruiz, M.O., Brown, W.M., Kitron, U.D., Newman, C.M., Goldberg, T.L., Walker, E.D., 2014. Dispersal of Adult Culex Mosquitoes in an Urban West Nile Virus Hotspot: A Mark-Capture Study Incorporating Stable Isotope Enrichment of Natural Larval Habitats. PLoS Negl. Trop. Dis. 8, e2768–e2768.

Hamer, G.L., Chaves, L.F., Anderson, T.K., Kitron, U.D., Brawn, J.D., Ruiz, M.O., Loss, S.R., Walker, E.D., Goldberg, T.L., 2011. Fine-scale variation in vector host use and force of infection drive localized patterns of West Nile virus transmission. PloS One 6, e23767.

Hamer, G.L., Kitron, U.D., Goldberg, T.L., Brawn, J.D., Loss, S.R., Ruiz, M.O., Hayes, D.B., Walker, E.D., 2009. Host Selection by Culex pipiens Mosquitoes and West Nile Virus Amplification. Am J Trop Med Hyg 80, 268–278.

Hamza, A.M., Rayah, E.A.E., 2016. A Qualitative Evidence of the Breeding Sites of Anopheles arabiensis Patton (Diptera: Culicidae) in and Around Kassala Town, Eastern Sudan. Int. J. Insect Sci. 8, IJIS.S40071.

Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sens. Environ. 122, 66–74.

Hardy, A.J., Gamarra, J.G.P., Cross, D.E., Macklin, M.G., Smith, M.W., Kihonda, J., Killeen, G.F., Ling'ala, G.N., Thomas, C.J., 2013. Habitat hydrology and geomorphology control the distribution of malaria vector larvae in rural Africa. PLOS ONE 8, e81931.

Hartley, D.M., Barker, C.M., Le Menach, A., Niu, T., Gaff, H.D., Reisen, W.K., 2012. Effects of temperature on emergence and seasonality of West Nile virus in California. Am. J. Trop. Med. Hyg. 86, 884–94.

Hay, S.I., Battle, K.E., Pigott, D.M., Smith, D.L., Moyes, C.L., Bhatt, S., Brownstein, J.S., Collier, N., Myers, M.F., George, D.B., Gething, P.W., 2013. Global mapping of infectious disease. Philos. Trans. R. Soc. B Biol. Sci. 368, 20120250–20120250.

Hay, S.I., Packer, M.J., Rogers, D.J., 1997. Review article - The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. Int. J. Remote Sens. 18, 2899–2930.

Hemingway, J., Ranson, H., 2000. Insecticide resistance in insect vectors of human disease. Annu. Rev. Entomol. 45, 371–391.

Hess, A., Davis, J.K., Wimberly, M.C., 2018. Identifying environmental risk factors and mapping the distribution of West Nile virus in an endemic region of North America. GeoHealth 2, 395–409.

Hoffman-Hall, A., Puett, R., Silva, J.A., Chen, D., Baer, A., Han, K.T., Han, Z.Y., Thi, A., Htay, T., Thein, Z.W., Aung, P.P., Plowe, C.V., Nyunt, M.M., Loboda, T.V., 2020.

Malaria Exposure in Ann Township, Myanmar, as a Function of Land Cover and Land Use: Combining Satellite Earth Observations and Field Surveys. GeoHealth 4, e2020GH000299.
Holick, J., Kyle, A., Ferraro, W., Delaney, R.R., Iwaseczko, M., 2002. Discovery of Aedes albopictus infected with west nile virus in southeastern Pennsylvania. J. Am. Mosq. Control Assoc. 18, 131.
Hopperstad, K.A., Sallam, M.F., Reiskind, M.H., 2021. Estimations of Fine-Scale Species Distributions of Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in Eastern Florida. J. Med. Entomol. 58, 699–707.
Houborg, R., McCabe, M.F., 2018. A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. Remote Sens. Environ. 209, 211–226.
Howe, D.A., Hathaway, J.M., Ellis, K.N., Mason, L.R., 2017. Spatial and temporal variability of air temperature across urban neighborhoods with varying amounts of tree canopy. Urban For. Urban Green. 27, 109–116.
Howes, R.E., Battle, K.E., Mendis, K.N., Smith, D.L., Cibulskis, R.E., Baird, J.K., Hay, S.I., 2016. Global Epidemiology of Plasmodium vivax. Am. J. Trop. Med. Hyg. 95, 15–34.
Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., Yoo, S.H., 2020. NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG).
Hulley, G., Hook, S., 2018. VIIRS/NPP Land Surface Temperature 6-Min L2 Swath 750m V001.
Hwang, J., Graves, P.M., Jima, D., Reithinger, R., Kachur, S.P., Group, and the E.M.W., 2010. Knowledge of malaria and its association with malaria-related behaviors—results from the malaria indicator survey, Ethiopia, 2007. PLOS ONE 5, e11692.
Ijumba, J.N., Lindsay, S.W., 2001. Impact of irrigation on malaria in Africa: paddies paradox. Med. Vet. Entomol. 15, 1–11.
Jin, S., Homer, C., Yang, L., Danielson, P., Dewitz, J., Li, C., Zhu, Z., Xian, G., Howard, D., 2019. Overall Methodology Design for the United States National Land Cover Database 2016 Products. Remote Sens. 11, 2971.
Johnson, E.E., Escobar, L.E., Zambrana-Torrelio, C., 2019. An Ecological Framework for Modeling the Geography of Disease Transmission. Trends Ecol. Evol.
Johnson, T.L., Haque, U., Monaghan, A.J., Eisen, L., Hahn, M.B., Hayden, M.H., Savage, H.M., McAllister, J., Mutebi, J.-P., Eisen, R.J., 2017. Modeling the Environmental Suitability for Aedes (Stegomyia) aegypti and Aedes (Stegomyia) albopictus (Diptera: Culicidae) in the Contiguous United States. J. Med. Entomol. 54, 1605–1614.
Jones, W.E., 1999. West Nile Virus New To North America. J. Equine Vet. Sci. 19, 680–683.
Kabaria, C.W., Molteni, F., Mandike, R., Chacky, F., Noor, A.M., Snow, R.W., Linard, C., 2016. Mapping intra-urban malaria risk using high resolution satellite imagery: a case study of Dar es Salaam. Int. J. Health Geogr. 15, 26.
Kassie, K.E., Alemu, B.A., Wedajoo, A.S., 2018. Impact of irrigation on household multidimensional poverty reduction in the Koga irrigation development project, northern Ethiopia. Asian Dev. Perspect. 29.

Kazansky, Y., Wood, D., Sutherlun, J., 2016. The current and potential role of satellite remote sensing in the campaign against malaria. Acta Astronaut. 121, 292–305.
Kent, R., Juliusson, L., Weissmann, M., Evans, S., Komar, N., 2009. Seasonal blood-feeding behavior of Culex tarsalis (Diptera: Culicidae) in Weld county, Colorado, 2007. J. Med. Entomol. 46, 380–390.
Kibret, S., Wilson, G.G., Ryder, D., Tekie, H., Petros, B., 2017. Malaria impact of large dams at different eco-epidemiological settings in Ethiopia. Trop. Med. Health 45, 4–4.
Kibret, S., Wilson, G.G., Tekie, H., Petros, B., 2014. Increased malaria transmission around irrigation schemes in Ethiopia and the potential of canal water management for malaria vector control. Malar. J. 13.
Kilpatrick, A.M., Meola, M.A., Moudy, R.M., Kramer, L.D., 2008. Temperature, Viral Genetics, and the Transmission of West Nile Virus by Culex pipiens Mosquitoes. PLoS Pathog. 4, e1000092–e1000092.
Kraemer, M., Messina, J., Faria, N., Pybus, O., Wint, W., et al, 2019. Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. Nat. Microbiol.
Kraemer, M.U.G., Hay, S.I., Pigott, D.M., Smith, D.L., Wint, G.R.W., Golding, N., 2016. Progress and Challenges in Infectious Disease Cartography. Trends Parasitol. 32, 19–29.
Kramer, L.D., Ciota, A.T., Kilpatrick, A.M., 2019. Introduction, Spread, and Establishment of West Nile Virus in the Americas. J. Med. Entomol.
Kulldorff, M., 2018. SaTScan TM User Guide.
LaDeau, S.L., Allan, B.F., Leisnham, P.T., Levy, M.Z., 2015. The ecological foundations of transmission potential and vector-borne disease in urban landscapes. Funct. Ecol. 29, 889–901.
LaDeau, S.L., Leisnham, P.T., Biehler, D., Bodner, D., 2013. Higher Mosquito Production in Low-Income Neighborhoods of Baltimore and Washington, DC: Understanding Ecological Drivers and Mosquito-Borne Disease Risk in Temperate Cities. Int. J. Environ. Res. Public. Health 10, 1505–1526. https://doi.org/10.3390/ijerph10041505
Lambin, E.F., Tran, A., Vanwambeke, S.O., Linard, C., Soti, V., 2010. Pathogenic landscapes: Interactions between land, people, disease vectors, and their animal hosts. Int. J. Health Geogr. 9, 54.
Landau, K.I., van Leeuwen, W.J.D., 2012. Fine scale spatial urban land cover factors associated with adult mosquito abundance and risk in Tucson, Arizona. J. Vector Ecol. J. Soc. Vector Ecol. 37, 407–418.
Lankir, D., Solomon, S., Gize, A., 2020. A five-year trend analysis of malaria surveillance data in selected zones of Amhara region, Northwest Ethiopia. BMC Public Health 20, 1175.
Lebl, K., Brugger, K., Rubel, F., 2013. Predicting Culex pipiens/restuans population dynamics by interval lagged weather data. Parasit. Vectors 6, 1–11.
Lemma, E., Upadhyaya, S., Ramsankaran, R., 2019. Investigating the performance of satellite and reanalysis rainfall products at monthly timescales across different rainfall regimes of Ethiopia. Int. J. Remote Sens. 40, 4019–4042.

Lemma, W., 2020. Impact of high malaria incidence in seasonal migrant and permanent adult male laborers in mechanized agricultural farms in Metema – Humera lowlands on malaria elimination program in Ethiopia. BMC Public Health 20, 320.
Lemma, W., Alemu, K., Birhanie, M., Worku, L., Niedbalski, J., McDowell, M.A., Lobo, N.F., 2019. Anopheles cinereus implicated as a vector of malaria transmission in the highlands of north-west Ethiopia. Parasit. Vectors 12, 557.
Leta, S., Beyene, T.J., De Clercq, E.M., Amenu, K., Kraemer, M.U.G., Revie, C.W., 2018. Global risk mapping for major diseases transmitted by Aedes aegypti and Aedes albopictus. Int. J. Infect. Dis. 67, 25–35.
Little, E., Biehler, D., Leisnham, P.T., Jordan, R., Wilson, S., LaDeau, S.L., 2017. Socio-Ecological Mechanisms Supporting High Densities of Aedes albopictus (Diptera: Culicidae) in Baltimore, MD. J. Med. Entomol. 54, 1183–1192.
Lobser, S.E., Cohen, W.B., 2007. MODIS tasselled cap: land cover characteristics expressed through transformed MODIS data. Int. J. Remote Sens. 28, 5079–5101.
Lockaby, G., Noori, N., Morse, W., Zipperer, W., Kalin, L., Governo, R., Sawant, R., Ricker, M., 2016. Climatic, ecological, and socioeconomic factors associated with West Nile virus incidence in Atlanta, Georgia, U.S.A. J. Vector Ecol. 41, 232–243.
Lunde, T.M., Korecha, D., Loha, E., Sorteberg, A., Lindtjørn, B., 2013. A dynamic model of some malaria-transmitting anopheline mosquitoes of the Afrotropical region. I. Model description and sensitivity analysis. Malar. J. 12, 28.
Lwande, O.W., Obanda, V., Lindström, A., Ahlm, C., Evander, M., Näslund, J., Bucht, G., 2019. Globe-Trotting *Aedes aegypti* and *Aedes albopictus*: Risk Factors for Arbovirus Pandemics. Vector-Borne Zoonotic Dis.
Lyon, B., Dinku, T., Raman, A., Thomson, M.C., 2017. Temperature suitability for malaria climbing the Ethiopian Highlands. Environ. Res. Lett. 12, 064015–064015.
Machault, V., Vignolles, C., Borchi, F., Vounatsou, P., Pages, F., Briolant, S., Lacaux, J.P., Rogier, C., 2011. The use of remotely sensed environmental data in the study of malaria. Geospatial Health 5, 151–168.
Malone, J.B., Bergquist, R., Martins, M., Luvall, J.C., 2019. Use of Geospatial Surveillance and Response Systems for Vector-Borne Diseases in the Elimination Phase. Trop. Med. Infect. Dis. 4, 15.
Mamai, W., Lobb, L.N., Bimbilé Somda, N.S., Maiga, H., Yamada, H., Lees, R.S., Bouyer, J., Gilles, J.R.L., 2018. Optimization of Mass-Rearing Methods for Anopheles arabiensis Larval Stages: Effects of Rearing Water Temperature and Larval Density on Mosquito Life-History Traits. J. Econ. Entomol. 111, 2383–2390.
Manore, C.A., Ostfeld, R.S., Agusto, F.B., Gaff, H., LaDeau, S.L., 2017. Defining the Risk of Zika and Chikungunya Virus Transmission in Human Population Centers of the Eastern United States. PLoS Negl. Trop. Dis. 11, e0005255.
Marcantonio, M., Rizzoli, A., Metz, M., Rosà, R., Marini, G., Chadwick, E., Neteler, M., 2015. Identifying the environmental conditions favouring West Nile virus outbreaks in Europe. PLOS ONE 10, e0121158–e0121158.
Marti, R., Li, Z., Catry, T., Roux, E., Mangeas, M., Handschumacher, P., Gaudart, J., Tran, A., Demagistri, L., Faure, J.-F., Carvajal, J.J., Drumond, B., Xu, L., Herbreteau, V., Gurgel, H., Dessay, N., Gong, P., 2020. A Mapping Review on Urban Landscape Factors of Dengue Retrieved from Earth Observation Data, GIS Techniques, and Survey Questionnaires. Remote Sens. 12, 932.

May, F.J., Davis, C.T., Tesh, R.B., Barrett, A.D.T., 2011. Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. J. Virol. 85, 2964–74.

Mayfield, H.J., Smith, C.S., Lowry, J.H., Watson, C.H., Baker, M.G., Kama, M., Nilles, E.J., Lau, C.L., 2018. Predictive risk mapping of an environmentally-driven infectious disease using spatial Bayesian networks: A case study of leptospirosis in Fiji. PLoS Negl. Trop. Dis. 12, e0006857.

McBride, C.S., Baier, F., Omondi, A.B., Spitzer, S.A., Lutomiah, J., Sang, R., Ignell, R., Vosshall, L.B., 2014. Evolution of mosquito preference for humans linked to an odorant receptor. Nature 515, 222–227.

McFeeters, S.K., 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.

McPherson, R.A., Fiebrich, C., Crawford, K.C., Elliott, R.L., Kilby, J.R., Grimsley, D.L., Basara, Illston, B.G., Morris, Kloesel, K.A., Stadler, S.J., Melvin, A.D., Sutherland, A.J., Shrivastava, H., 2007. Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. J Atmos Ocean. Technol.

Mercer, D.R., Sheeley, S.L., Brown, E.J., 2005. Mosquito (Diptera : Culicidae) development within microhabitats of an Iowa wetland. Entomol. Soc. Am. 42, 685–693.

Merkord, C.L., Liu, Y., Mihretie, A., Gebrehiwot, T., Awoke, W., Bayabil, E., Henebry, G.M., Kassa, G.T., Lake, M., Wimberly, M.C., 2017. Integrating malaria surveillance with climate data for outbreak detection and forecasting: the EPIDEMIA system. Malar. J. 16, 89.

Messina, J.P., Brady, O.J., Golding, N., Kraemer, M.U.G., Wint, G.R.W., Ray, S.E., Pigott, D.M., Shearer, F.M., Johnson, K., Earl, L., Marczak, L.B., Shirude, S., Weaver, N.D., Gilbert, M., Velayudhan, R., Jones, P., Jaenisch, T., Scott, T.W., Reiner, R.C., Hay, S.I., 2019. The current and future global distribution and population at risk of dengue. Nat. Microbiol. 1.

Metelmann, S., Caminade, C., Jones, A.E., Medlock, J.M., Baylis, M., Morse, A.P., 2019. The UK's suitability for *Aedes albopictus* in current and future climates. J. R. Soc. Interface 16, 20180761.

Midekisa, A., Beyene, B., Mihretie, A., Bayabil, E., Wimberly, M.C., 2015. Seasonal associations of climatic drivers and malaria in the highlands of Ethiopia. Parasit. Vectors 8, 339.

Midekisa, A., Senay, G., Henebry, G.M., Semuniguse, P., Wimberly, M.C., 2012. Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. Malar. J. 11, 165–165.

Midekisa, A., Senay, G.B., Wimberly, M.C., 2014. Multisensor earth observations to characterize wetlands and malaria epidemiology in Ethiopia. Water Resour. Res. 50, 8791–8806.

Minale, A.S., Alemu, K., 2018. Malaria-risk assessment using geographical information system and remote sensing in Mecha District, West Gojjam, Ethiopia. Geospatial Health 13, 157–163.

Molaei, G., Andreadis, T.G., Armstrong, P.M., Anderson, J.F., Vossbrinck, C.R., 2006. Host Feeding Patterns of Culex Mosquitoes and West Nile Virus Transmission, Northeastern United States. Emerg. Infect. Dis. 12, 468–474.

Mordecai, E.A., Caldwell, J.M., Grossman, M.K., Lippi, C.A., Johnson, L.R., Neira, M., Rohr, J.R., Ryan, S.J., Savage, V., Shocket, M.S., Sippy, R., Ibarra, A.M.S., Thomas, M.B., Villena, O., 2019. Thermal biology of mosquito-borne disease. Ecol. Lett. 22, 1690–1708.

Mordecai, E.A., Cohen, J.M., Evans, M.V., Gudapati, P., Johnson, L.R., Lippi, C.A., Miazgowicz, K., Murdock, C.C., Rohr, J.R., Ryan, S.J., Savage, V., Shocket, M.S., Ibarra, A.S., Thomas, M.B., Weikel, D.P., 2017. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. PLoS Negl. Trop. Dis. 11, e0005568.

Mordecai, E.A., Paaijmans, K.P., Johnson, L.R., Balzer, C., Ben-Horin, T., de Moor, E., McNally, A., Pawar, S., Ryan, S.J., Smith, T.C., Lafferty, K.D., 2013. Optimal temperature for malaria transmission is dramatically lower than previously predicted. Ecol. Lett. 16, 22–30.

Munga, S., Minakawa, N., Zhou, G., Mushinzimana, E., Barrack, O.-O.J., Githeko, A.K., Yan, G., 2006. Association between land cover and habitat productivity of malaria vectors in western Kenyan highlands. Am. J. Trop. Med. Hyg. 74, 69–75.

Murdock, C.C., Evans, M.V., Mcclanahan, T.D., Miazgowicz, K.L., Tesla, B., 2017. Fine-scale variation in microclimate across an urban landscape shapes variation in mosquito population dynamics and the potential of Aedes albopictus to transmit arboviral disease. PLoS Negl. Trop. Dis. 11, e0005640–e0005640.

Mushinzimana, E., Munga, S., Minakawa, N., Li, L., Feng, C., Bian, L., Kitron, U., Schmidt, C., Beck, L., Zhou, G., Githeko, A.K., Yan, G., 2006. Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands. Malar. J. 5, 13.

Nagelkerke, N.J.D., 1991. A note on a general definition of the coefficient of determination. Biometrika 78, 691–692.

Nance, J., Fryxell, R.T., Lenhart, S., 2018. Modeling a single season of Aedes albopictus populations based on host-seeking data in response to temperature and precipitation in eastern Tennessee. J. Vector Ecol. 43, 138–147.

Newby, G., Bennett, A., Larson, E., Cotter, C., Shretta, R., Phillips, A.A., Feachem, R.G.A., 2016. The path to eradication: a progress report on the malaria-eliminating countries. Lancet 387, 1775–1784.

Nissen, A., Cook, J., Loha, E., Lindtjørn, B., 2017. Proximity to vector breeding site and risk of Plasmodium vivax infection: a prospective cohort study in rural Ethiopia. Malar. J. 16, 380.

Noden, B.H., Coburn, L., Wright, R., Bradley, K., 2015. An Updated Checklist of the Mosquitoes of Oklahoma Including New State Records and West Nile Virus Vectors, 2003-06. J. Am. Mosq. Control Assoc. Mt. Laurel 31, 336–345.

O'Hara, R.B., Kotze, D.J., 2010. Do not log-transform count data. Methods Ecol. Evol. 1, 118–122.

Oklahoma Climatological Survey, 2021. Oklahoma climate - Long term averages and extremes [WWW Document]. URL http://climate.ok.gov/index.php/climate (accessed 5.23.21).

Oklahoma State Department Of Health, 2019. West Nile Virus Disease and Deaths by Year Oklahoma, 2002-2019.

Olson, S.H., Gangnon, R., Elguero, E., Durieux, L., Guégan, J.F., Foley, J.A., Patz, J.A., 2009. Links between climate, malaria, and wetlands in the amazon basin. Emerg. Infect. Dis. 15, 659–662.
Ostfeld, R.S., Glass, G.E., Keesing, F., 2005. Spatial epidemiology: An emerging (or re-emerging) discipline. Trends Ecol. Evol. 20, 328–336.
Parselia, E., Kontoes, C., Tsouni, A., Hadjichristodoulou, C., Kioutsioukis, I., Magiorkinis, G., Stilianakis, N.I., 2019. Satellite earth observation data in epidemiological modeling of malaria, dengue and West Nile virus: a scoping review. Remote Sens. 11, 1862.
Patouillard, E., Griffin, J., Bhatt, S., Ghani, A., Cibulskis, R., 2017. Global investment targets for malaria control and elimination between 2016 and 2030. BMJ Glob. Health 2, e000176.
Paulander, J., Olsson, H., Lemma, H., Getachew, A., Sebastian, M.S., 2009. Knowledge, attitudes and practice about malaria in rural Tigray, Ethiopia. Glob. Health Action 2, 1839.
Paz, S., 2015. Climate change impacts on West Nile virus transmission in a global context. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 370, 20130561–20130561.
Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691.
Peng, J., Loew, A., Merlin, O., Verhoest, N.E.C., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. Rev. Geophys. 55, 341–366.
Peterson, A.. T., Vieglais, D. a, Andreasen, J.K., 2003. Migratory birds modeled as critical transport agents for West Nile Virus in North America. Vector Borne Zoonotic Dis. Larchmt. N 3, 27–37.
Peterson, A.T., 2014. Mapping Disease Transmission Risk. The Johns Hopkins University Press, Baltimore.
Peterson, A.T., Robbing, A., Restifo, R., Howell, J., Nasci, R., 2008. Predictable ecology and geography of West Nile virus transmission in the central United States. J. Vector Ecol. 33, 342–352.
Pielke, R.A., 2005. Land Use and Climate Change. Science 310, 1625–1626.
Pigott, D.M., Atun, R., Moyes, C.L., Hay, S.I., Gething, P.W., 2012. Funding for malaria control 2006–2010: A comprehensive global assessment. Malar. J. 11.
Pigott, D.M., Howes, R.E., Wiebe, A., Battle, K.E., Golding, N., Gething, P.W., Dowell, S.F., Farag, T.H., Garcia, A.J., Kimball, A.M., Krause, L.K., Smith, C.H., Brooker, S.J., Kyu, H.H., Vos, T., Murray, C.J.L., Moyes, C.L., Hay, S.I., 2015. Prioritising infectious disease mapping. PLoS Neglected Trop. Dis. 9, e0003756–e0003756.
Pincebourde, S., Murdock, C.C., Vickers, M., Sears, M.W., 2016. Fine-Scale Microclimatic Variation Can Shape the Responses of Organisms to Global Change in Both Natural and Urban Environments. Integr. Comp. Biol. 56, 45–61.
Planet Team, 2017. Planet application program interface: in space for life on earth. San Francisco, CA.
R Core Team, 2016. A language and environment for statistical computing. R Foundation for Statistical Computing.
Rakotoarison, H.A., Rasamimalala, M., Rakotondramanga, J.M., Ramiranirina, B., Franchard, T., Kapesa, L., Razafindrakoto, J., Guis, H., Tantely, L.M., Girod, R., Rakotoniaina, S., Baril, L., Piola, P., Rakotomanana, F., 2020. Remote Sensing and

Multi-Criteria Evaluation for Malaria Risk Mapping to Support Indoor Residual Spraying Prioritization in the Central Highlands of Madagascar. Remote Sens. 12, 1585.
Reisen, W.K., 2019. Twenty Years of West Nile Virus in the United States: Introduction. J. Med. Entomol. 56, 1447.
Reisen, W.K., 2013. Ecology of West Nile Virus in North America. Viruses 5, 2079–2105.
Reisen, W.K., Fang, Y., Martinez, V.M., 2006. Effects of Temperature on the Transmission of West Nile Virus by *Culex tarsalis* (Diptera: Culicidae). J. Med. Entomol. 43, 309–317.
Reisen, W.K., Lothrop, H.D., Thiemann, T., 2013. Host Selection Patterns of Culex tarsalis (Diptera: Culicidae) at Wetlands Near the Salton Sea, Coachella Valley, California, 1998–2002. J. Med. Entomol. 50, 1071–1076.
Reiter, P., Sprenger, D., 1987. The used tire trade: a mechanism for the worldwide dispersal of container breeding mosquitoes. J. Am. Mosq. Control Assoc. 3, 494–501.
Ridgeway, G., Southworth, M., RUnit, 2013. Package 'gbm.' CRAN.
Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., 2021. Package "MASS."
Rocklöv, J., Dubrow, R., 2020. Climate change: an enduring challenge for vector-borne disease prevention and control. Nat. Immunol. 21, 479–483.
Rogers, D.J., Randolph, S.E., Snow, R.W., Hay, S.I., 2002. Satellite imagery in the study ad forecast of malaria. Nature 415, 710–715.
Rosenthal, P.J., John, C.C., Rabinovich, N.R., 2019. Malaria: How Are We Doing and How Can We Do Better? Am. J. Trop. Med. Hyg. 100, 239–241.
Rowley, W.A., Graham, C.L., 1968. The effect of temperature and relative humidity on the flight performance of female Aedes aegypti. J. Insect Physiol. 14, 1251–1257.
Rozeboom, E., 1942. THE MOSQUITOES OF OKLAHOMA 56.
Ryan, S.J., Carlson, C.J., Mordecai, E.A., Johnson, L.R., 2019. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. PLoS Negl. Trop. Dis. 13, e0007213.
Rydzanicz, K., Jawień, P., Lonc, E., Modelska, M., 2016. Assessment of productivity of Culex spp. larvae (Diptera: Culicidae) in urban storm water catch basin system in Wrocław (SW Poland). Parasitol. Res. 115, 1711–1720.
Sallam, M., Michaels, S., Riegel, C., Pereira, R., Zipperer, W., Lockaby, B., Koehler, P., 2017. Spatio-Temporal Distribution of Vector-Host Contact (VHC) Ratios and Ecological Niche Modeling of the West Nile Virus Mosquito Vector, Culex quinquefasciatus, in the City of New Orleans, LA, USA. Int. J. Environ. Res. Public. Health 14, 892–892.
Sánchez-Gómez, A., Amela, C., Fernández-Carrión, E., Martínez-Avilés, M., Sánchez-Vizcaíno, J.M., Sierra-Moros, M.J., 2017. Risk mapping of West Nile virus circulation in Spain, 2015. Acta Trop. 169, 163–169.
Schaaf, C., Wang, Z., 2015. MCD43C4 MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF-Adjusted Ref Daily L3 Global 0.05Deg CMG V006 [Data set].
Scott, C.A., Yeshiwondim, A.K., Serda, B., Guinovart, C., Tesfay, B.H., Agmas, A., Zeleke, M.T., Guesses, G.S., Ayenew, A.L., Workie, W.M., Steketee, R.W., Earle, D., Bezabih, B., Getachew, A., 2016. Mass testing and treatment for malaria in low transmission areas in Amhara Region, Ethiopia. Malar. J. 15, 305.

Sewe, M.O., Tozan, Y., Ahlm, C., Rocklöv, J., 2017. Using remote sensing environmental data to forecast malaria incidence at a rural district hospital in Western Kenya. Sci. Rep. 7, 2589.

Shand, L., Brown, W.M., Chaves, L.F., Goldberg, T.L., Hamer, G.L., Haramis, L., Kitron, U., Walker, E.D., Ruiz, M.O., 2016. Predicting West Nile Virus Infection Risk From the Synergistic Effects of Rainfall and Temperature. J. Med. Entomol. 53, 935–944.

Shields, T., Pinchoff, J., Lubinda, J., Hamapumbu, H., Searle, K., Kobayashi, T., Thuma, P.E., Moss, W.J., Curriero, F.C., 2016. Spatial and temporal changes in household structure locations using high-resolution satellite imagery for population assessment: an analysis of household locations in southern Zambia between 2006 and 2011. Geospatial Health 11, 410.

Shragai, T., Harrington, L.C., 2019. Aedes albopictus (Diptera: Culicidae) on an Invasive Edge: Abundance, Spatial Distribution, and Habitat Usage of Larvae and Pupae Across Urban and Socioeconomic Environmental Gradients. J. Med. Entomol. 56, 472–482.

Skaff, N.K., Armstrong, P.M., Andreadis, T.G., Cheruvelil, K.S., 2017. Wetland characteristics linked to broad-scale patterns in *Culiseta melanura* abundance and eastern equine encephalitis virus infection. Parasit. Vectors 10, 501–501.

Skaff, N.K., Cheruvelil, K.S., 2016. Fine-scale wetland features mediate vector and climate-dependent macroscale patterns in human West Nile virus incidence. Landsc. Ecol. 31, 1615–1628.

Solano-Villarreal, E., Valdivia, W., Pearcy, M., Linard, C., Pasapera-Gonzales, J., Moreno-Gutierrez, D., Lejeune, P., Llanos-Cuentas, A., Speybroeck, N., Hayette, M.-P., Rosas-Aguirre, A., 2019. Malaria risk assessment and mapping using satellite imagery and boosted regression trees in the Peruvian Amazon. Sci. Rep. 9, 15173.

Spence Beaulieu, M.R., Hopperstad, K., Dunn, R.R., Reiskind, M.H., 2019. Simplification of vector communities during suburban succession. PLOS ONE 14, e0215485.

Stresman, G.H., 2010. Beyond temperature and precipitation: Ecological risk factors that modify malaria transmission. Acta Trop. 116, 167–172.

Tabachnick, W.J., Powell, J.R., 1979. A world-wide survey of genetic variation in the yellow fever mosquito, Aedes aegypti*. Genet. Res. 34, 215–229.

Talapko, J., Škrlec, I., Alebić, T., Jukić, M., Včev, A., 2019. Malaria: The Past and the Present. Microorganisms 7, 179.

Tanner, M., Greenwood, B., Whitty, C.J.M., Ansah, E.K., Price, R.N., Dondorp, A.M., Seidlein, L.V., Baird, J.K., Beeson, J.G., 2015. Malaria eradication and elimination: views on how to translate a vision into reality. BMC Med. 13, 1–22.

Tarara, J.M., Hoheisel, G.-A., 2007. Low-cost Shielding to Minimize Radiation Errors of Temperature Sensors in the Field. HortScience 42, 1372–1379.

Tatem, A.J., 2017. WorldPop, open data for spatial demography. Sci. Data 4, 170004.

Tatem, A.J., Hay, S.I., Rogers, D.J., 2006. Global traffic and disease vector dispersal. Proc. Natl. Acad. Sci. 103, 6242–6247.

Thomson, R.C.M., 1938. The reactions of mosquitoes to temperature and humidity. Bull. Entomol. Res. 29, 125–140.

Tjaden, N.B., Cheng, Y., Beierkuhnlein, C., Thomas, S.M., 2021. Chikungunya Beyond the Tropics: Where and When Do We Expect Disease Transmission in Europe? Viruses 13, 1024.
Tran, A., Sudre, B., Paz, S., Rossi, M., Desbrosse, A., Chevalier, V., Semenza, J.C., 2014. Environmental predictors of West Nile Fever risk in Europe. Int. J. Health Geogr. 13, 26.
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.
UNDESA, 2019. World urbanization prospects: the 2018 revision.
United Nations, 2015. Transforming our world: the 2030 Agenda for Sustainable Development.
United Nations, 2013. The Millennium Development Goals Report. New York.
U.S. Census Bureau, 2019. Census - Geography Profile [WWW Document]. URL https://data.census.gov/cedsci/profile?g=1600000US4052500 (accessed 4.22.21).
Uusitalo, R., Siljander, M., Culverwell, C.L., Mutai, N.C., Forbes, K.M., Vapalahti, O., Pellikka, P.K.E., 2019. Predictive mapping of mosquito distribution based on environmental and anthropogenic factors in Taita Hills, Kenya. Int. J. Appl. Earth Obs. Geoinformation 76, 84–92.
Vavassori, L., Saddler, A., Müller, P., 2019. Active dispersal of Aedes albopictus: a mark-release-recapture study using self-marking units. Parasit. Vectors 12, 583.
Vega-Rua, A., Zouache, K., Caro, V., Diancourt, L., Delaunay, P., Grandadam, M., Failloux, A.-B., 2013. High efficiency of temperate *Aedes albopictus* to transmit chikungunya and dengue viruses in the Southeast of France. PloS One 8, e59716.
Venables, W., Ripley, B., 2002. Modern Applied Statistics with S, 4th ed. Springer, New York.
Verdonschot, P.F.M., Besse-Lototskaya, A.A., 2014. Flight distance of mosquitoes (Culicidae): A metadata analysis to support the management of barrier zones around rewetted and newly constructed wetlands. Limnologica 45, 69–79.
Veronesi, R., Gentile, G., Carrieri, M., Maccagnani, B., Stermieri, L., Bellini, R., 2012. Seasonal pattern of daily activity of Aedes caspius, Aedes detritus, Culex modestus, and Culex pipiens in the Po Delta of northern Italy and significance for vector-borne disease risk assessment. J. Vector Ecol. 37, 49–61.
Vincent, G.P., 2018. Surveillance of South Dakota mosquito abundance, infection rate, and insecticide susceptibility. (Doctoral dissertation).
Wagenmakers, E.-J., Farrell, S., 2004. AIC model selection using Akaike weights. Psychon. Bull. Rev. 11, 192–196.
Walker, K., Lynch, M., 2007. Contributions of Anopheles larval control to malaria suppression in tropical Africa: review of achievements and potential. Med. Vet. Entomol. 21, 2–21.
Walsh, C.J., Fletcher, T.D., Burns, M.J., 2012. Urban stormwater runoff: a new class of environmental flow problem. PloS One 7, e45814.
Walton, W.E., Tietze, N.S., Mulla, M.S., 1990. Ecology of Culex tarsalis (Diptera: Culicidae): Factors Influencing Larval Abundance in Mesocosms in Southern California. J. Med. Entomol. 27, 57–67.

Wan, Z., Hook, S., Hulley, G., 2015. MOD11A2 MODIS/Terra land surface temperature/emissivity 8-day L3 global 1km SIN grid V006. NASA EOSDIS Land Processes DAAC.

Whiteman, A., Gomez, C., Rovira, J., Chen, G., McMillan, W.O., Loaiza, J., 2019. Aedes Mosquito Infestation in Socioeconomically Contrasting Neighborhoods of Panama City. EcoHealth.

WHO, 2020. World malaria report 2020. Geneva.

WHO, World Health Organization, 2017. A framework for malaria elimination.

Wilke, A.B.B., Carvajal, A., Medina, J., Anderson, M., Nieves, V.J., Ramirez, M., Vasquez, C., Petrie, W., Cardenas, G., Beier, J.C., 2019. Assessment of the effectiveness of BG-Sentinel traps baited with CO2 and BG-Lure for the surveillance of vector mosquitoes in Miami-Dade County, Florida. PLOS ONE 14, e0212688.

Wimberly, M.C., Baer, A.D., Yabsley, M.J., 2008. Enhanced spatial models for predicting the geographic distributions of tick-borne pathogens. Int. J. Health Geogr. 7, 15–15.

Wimberly, M.C., Davis, J.K., Evans, M.V., Hess, A., Newberry, P.M., Solano-Asamoah, N., Murdock, C.C., 2020. Land cover affects microclimate and temperature suitability for arbovirus transmission in an urban landscape. PLoS Negl. Trop. Dis. 14, e0008614.

Wimberly, M.C., de Beurs, K.M., Loboda, T.V., Pan, W.K., 2021. Satellite Observations and Malaria: New Opportunities for Research and Applications. Trends Parasitol.

Wimberly, M.C., Giacomo, P., Kightlinger, L., Hildreth, M., 2013. Spatio-temporal epidemiology of human West Nile virus disease in South Dakota. Int. J. Environ. Res. Public. Health 10, 5584–5602.

Wimberly, M.C., Lamsal, A., Giacomo, P., Chuang, T.W., 2014. Regional variation of climatic influences on West Nile virus outbreaks in the United States. Am. J. Trop. Med. Hyg. 91, 677–684.

Woods, A.J., Omernik, J.M., Butler, D.R., Ford, J.G., Henley, J.G., Hoagland, B.W., Arndt, D.S., 2016. Ecoregions of Oklahoma: Reston, Virginia, U.S. Geological Survey, map scale 1:1,250,000.

World Health Organization, 2016. World Malaria Report 2015. WHO.

Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Fosnight, E.A., Shaw, J., Masek, J.G., Roy, D.P., 2016. The global Landsat archive: Status, consolidation, and direction. Remote Sens. Environ. 185, 271–283.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033.

Yalew, W.G., Pal, S., Bansil, P., Dabbs, R., Tetteh, K., Guinovart, C., Kalnoky, M., Serda, B.A., Tesfay, B.H., Beyene, B.B., Seneviratne, C., Littrell, M., Yokobe, L., Noland, G.S., Domingo, G.J., Getachew, A., Drakeley, C., Steketee, R.W., 2017. Current and cumulative malaria infections in a setting embarking on elimination: Amhara, Ethiopia. Malar. J. 16, 242.

Yamana, T.K., Eltahir, E.A.B., 2013. Incorporating the effects of humidity in a mechanistic model of Anopheles gambiae mosquito population dynamics in the Sahel region of Africa. Parasit. Vectors 6, 235.

Yang, L., Huang, C., Homer, C.G., Wylie, B.K., Coan, M.J., 2003. An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. Can. J. Remote Sens. 29, 230–240.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., Xian, G., 2018. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. ISPRS J. Photogramm. Remote Sens. 146, 108–123.

Young, S.G., Tullis, J.A., Cothren, J., 2013. A remote sensing and GIS-assisted landscape epidemiology approach to West Nile virus. Appl. Geogr. 45, 241–249.

Yukich, J.O., Taylor, C., Eisele, T.P., Reithinger, R., Nauhassenay, H., Berhane, Y., Keating, J., 2013. Travel history and malaria infection risk in a low-transmission setting in Ethiopia: a case control study. Malar. J. 12, 33.

Zemene, E., Koepfli, C., Tiruneh, A., Yeshiwondim, A.K., Seyoum, D., Lee, M.-C., Yan, G., Yewhalaw, D., 2018. Detection of foci of residual malaria transmission through reactive case detection in Ethiopia. Malar. J. 17, 390.

Zerdo, Z., Bastiaens, H., Anthierens, S., Massebo, F., Masne, M., Biresaw, G., Shewangizaw, M., Tunje, A., Chisha, Y., Yohannes, T., Van Geertruyden, J.-P., 2020. Long-lasting insecticide-treated bed net ownership, utilization and associated factors among school-age children in Dara Mallo and Uba Debretsehay districts, Southern Ethiopia. Malar. J. 19, 369.

Zhu, Z., Wulder, M.A., Roy, D.P., Woodcock, C.E., Hansen, M.C., Radeloff, V.C., Healey, S.P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J.-F., Lymburner, L., Pahlevan, N., Scambos, T.A., 2019. Benefits of the free and open Landsat data policy. Remote Sens. Environ. 224, 382–385.