

**SPATIALLY EXPLICIT MODELING OF WEST NILE VIRUS RISK USING
ENVIRONMENTAL DATA**

Abhishek K. Kala, M.Sc.,

Dissertation Prepared for the Degree of
DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2015

APPROVED:

Sam Atkinson, Major Professor
Chetan Tiwari, Committee member
Armin Mikler, Committee member
Joseph Oppong, Committee member
Bruce Hunter, Committee member
Chair of the Department of Biology
Program Coordinator
Costas Tsatsoulis, Dean of the Toulouse Graduate School

Kala, Abhishek K. Spatially explicit modeling of West Nile virus risk using environmental data. Doctor of Philosophy (Environmental Science), December 2015, 99 pp., 7 tables, 26 figures, 120 references.

West Nile virus (WNV) is an emerging infectious disease that has widespread implications for public health practitioners across the world. Within a few years of its arrival in the United States the virus had spread across the North American continent. This research focuses on the development of a spatially explicit GIS-based predictive epidemiological model based on suitable environmental factors. We examined eleven commonly mapped environmental factors using both ordinary least squares regression (OLS) and geographically weighted regression (GWR). The GWR model was utilized to ascertain the impact of environmental factors on WNV risk patterns without the confounding effects of spatial non-stationarity that exist between place and health. It identifies the important underlying environmental factors related to suitable mosquito habitat conditions to make meaningful and spatially explicit predictions. Our model represents a multi-criteria decision analysis approach to create disease risk maps under data sparse situations. The best fitting model with an adjusted R^2 of 0.71 revealed a strong association between WNV infection risk and a subset of environmental risk factors including road density, stream density, and land surface temperature.

This research also postulates that understanding the underlying place characteristics and population composition for the occurrence of WNV infection is important for mitigating future outbreaks. While many spatial and aspatial models have attempted to predict the risk of WNV transmission, efforts to link these factors within a GIS framework are limited. One of the major challenges for such integration is the high

dimensionality and large volumes typically associated with such models and data. This research uses a spatially explicit, multivariate geovisualization framework to integrate an environmental model of mosquito habitat with human risk factors derived from socio-economic and demographic variables. Our results show that such an integrated approach facilitates the exploratory analysis of complex data and supports reasoning about the underlying spatial processes that result in differential risks for WNV.

This research provides different tools and techniques for predicting the WNV epidemic and provides more insights into targeting specific areas for controlling WNV outbreaks.

Copyright 2015
by
Abhishek Kumar Kala

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor Dr. Sam Atkinson for his expert guidance, encouragement, and providing me with an excellent research environment. I am fully indebted to him for his understanding and for providing me with every possible support that I needed throughout my research. My appreciation also extends to Dr. Chetan Tiwari, Dr. Armin Mikler, Dr. Joseph Oppong, and Dr. Bruce Hunter for accepting to be in my research committee and for mentoring my research with their interdisciplinary knowledge. Special thanks to Dr. Tiwari for his early insights about my research and for helping me to sustain a positive atmosphere. I would also like to take this opportunity to thank the Biology Department for all the financial support that I received as a Beth Baird scholar and by giving me an opportunity of teaching in Microbiology lab. Sincere thanks to all the faculty and staff members of Institute of Applied Sciences, staff members from Graduate School and International Advising at University of North Texas for their help and support.

I am eternally grateful to my family. I would like to acknowledge with gratitude the support and love of my family - late grandparents, parents, in-laws, my lovely wife, and my darling daughter who is still in the minor league. They all kept me going, and this research would not have been possible without them. This Ph.D. is a testament to your belief in me.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES.....	viii
 CHAPTER 1	
INTRODUCTION.....	1
1.1 Introduction	1
1.2 Research questions	2
1.3 Environmental model	4
1.4 Spatial Heterogeneity	6
1.5 Geovisualization framework	8
 CHAPTER 2	
SPATIALLY EXPLICIT MODELING OF WEST NILE VIRUS RISK USING ENVIRONMENTAL DATA	11
2.1 Abstract.....	11
2.2 Introduction	12
2.3 West Nile Virus.....	13
2.4 Putative environmental risk factors	16
2.5 Spatial-epidemiology modeling for risk prediction	19
2.6 Materials and Methods.....	23
2.6.1 Study area	23
2.6.2 Study design overview	24

2.6.3 Data sources.....	25
2.7 Data processing and analysis	28
2.8 Results and discussion	36
2.9 Acknowledgments	40
CHAPTER 3	
USING LOCAL MODELING APPROACH TO EXPLORE SPATIAL HETEROGENEITY IN WEST NILE VIRUS RISK	41
3.1 Abstract.....	41
3.2 Introduction	42
3.3 Modeling WNV risk in a local environment: Transmission and risk factors	44
3.4 Modeling WNV risk in a local environment: statistical considerations	47
3.5 Study area and data.....	48
3.5.1 Study area	48
3.5.2 Environmental factors and data sources.....	49
3.6 Methods	51
3.6.1 OLS modeling	51
3.6.2 GWR model specification.....	53
3.7 Modeling results	56
3.8 Discussion.....	63
3.9 Conclusion	66
CHAPTER 4	
USING MULTIVARIATE GEOVISUALIZATION TO MODEL THE RELATIONSHIPS BETWEEN PLACE VULNERABILITY AND WEST NILE VIRUS RISK	68
4.1 Abstract.....	68

4.2 Introduction	69
4.3 Materials and methods.....	71
4.3.1 Contextual and compositional factors affecting WNV	71
4.3.2 Study area and data	71
4.3.3 Mosquito habitat model.....	73
4.3.4 Geovisualization techniques	74
4.4 Results and discussion	78
4.5 Conclusions	81
CHAPTER 5	
CONCLUSION	82
LIST OF REFERENCES	88

LIST OF TABLES

Table 1-1 Environmental factors associated with WNV vector habitat.....	6
Table 2-1 Epidemiologic studies utilizing multi-criteria evaluation approach	21
Table 2-2 Sources of data and their spatial resolution.....	27
Table 2-3 Association of environmental factors with WNV vector habitat.....	32
Table 3-1 Environmental conditions related with WNV human transmission risk.....	46
Table 3-2 Data sources.....	50
Table 4-1 Population and place characteristics associated with WNV risk.....	72

LIST OF FIGURES

Figure 1-1 West Nile Virus transmission cycle.....	2
Figure 1-2 Modeling approach and research questions.....	4
Figure 2-1 West Nile virus transmission cycle and mosquito life cycle.....	16
Figure 2.2 World physical map with inset map showing study area.....	23
Figure 2-3 West Nile virus model skeleton.....	25
Figure 2-4 Map showing distribution of different land-cover classes in California.....	26
Figure 2-5 Graph showing WNV disease onset week for California, 2004-2009.....	29
Figure 2-7 Observed Vs Predicted infected dead birds.....	34
Figure 2-8 Maps of WNV human incidence rate and WNV infected dead birds.....	37
Figure 2-9 WNV human infection risk map for 2005.....	39
Figure 2-10 Counties with minimum threshold of WNV infected dead birds (at least 40 dead birds infected with WNV; $r^2=0.82$).....	40
Figure 3-1 Distribution of model's standardized residuals (line indicates normal curve).....	53
Figure 3-2 Trendline plot for global OLS model.....	56

Figure 3-3 Trendline plot for local GWR model.....	57
Figure 3-4 Map showing distribution of standardized residuals.....	58
Figure 3-5 Maps showing distribution of land surface temperature coefficients.....	60
Figure 3-6 Maps showing distribution of road density coefficients.....	61
Figure 3-7 Maps showing distribution of stream density coefficients.....	62
Figure 4-1 Geovisualization model framework.....	73
Figure 4-2 WNV infection risk map.....	76
Figure 4-3 Census tracts enclosed within standard deviation ellipse.....	76
Figure 4-4 SOM clusters representing a valid combination of variables.....	76
Figure 4-5 PCP cluster showing population group of highest median age.....	77
Figure 4-6 PCP cluster showing the highest mosquito habitat model risk.....	77
Figure 4-7 Geomap showing the spread of census tracts associated with SOM clusters.....	78

CHAPTER 1

INTRODUCTION

1.1 Introduction

West Nile virus (WNV) is a globally emerging infectious disease that first appeared in United States in 1999. It poses a considerable public health concern and a continuing challenge as currently no medications or vaccines exist that can treat WNV infection in human population (De Filette et al. 2012). Therefore it becomes important to develop efficient methods for disease prediction and modeling as it can help to identify areas of risk, improve the specificity of exposure prediction, and, therefore, help develop better public health intervention strategies.

The distribution of WNV is dependent on the occurrence of susceptible avian reservoir hosts and competent mosquito vectors, mosquito host preference, and availability of hosts. The WNV transmission cycle (Figure 1-1, JAMA, 2012) commences with the deposition of eggs by a female mosquito. A competent female mosquito bites an infected bird reservoir host (Blair 2009) while seeking a blood meal to obtain nutrients necessary for egg development. The infected mosquito can now potentially transmit the virus to another bird or animal host when it feeds again. Upon migration of these infected birds the virus is once again transmitted to susceptible mosquitoes. Ultimately the disease is transmitted by mosquitoes to humans or other mammals that act as incidental hosts.

WNV has been detected in dead birds of at least 326 species. Although birds, particularly crows and jays, infected with WNV can die or become ill, most infected birds do survive.

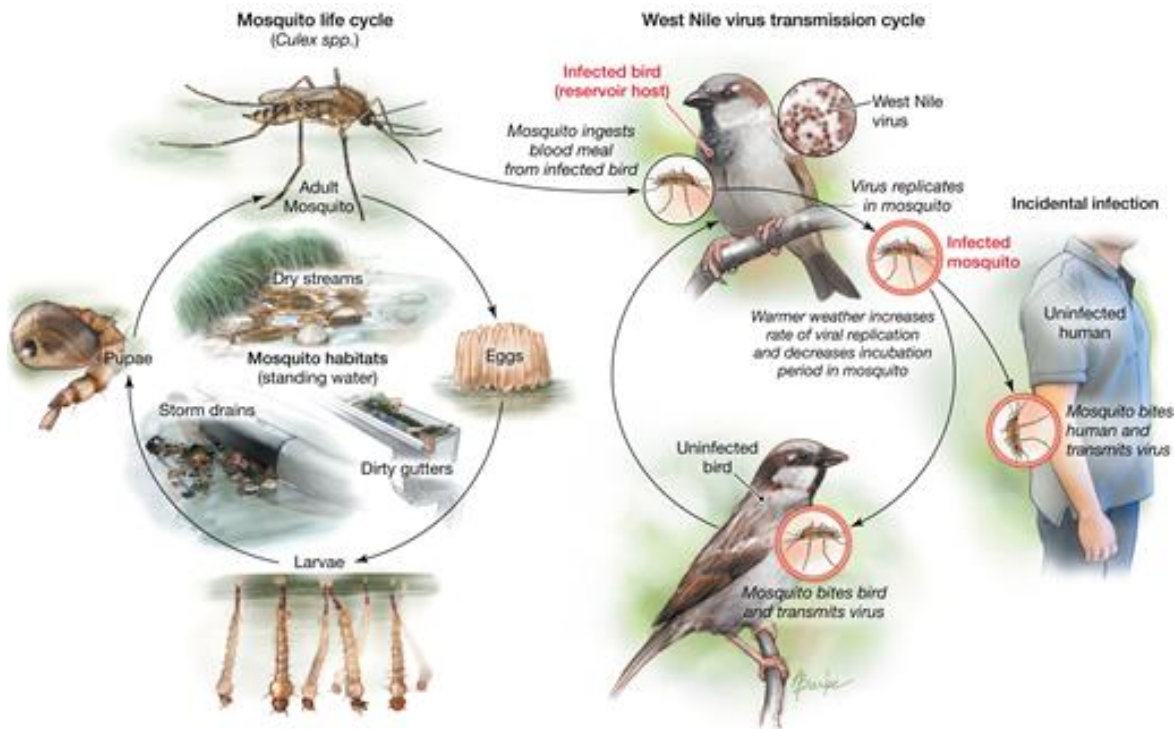


Figure 1-1 West Nile Virus transmission cycle, Photo. JAMA. 2012/jamanetwork.com (Kuehn 2012) (accessed March 30 2013).

1.2 Research questions

Across the research continuum several methods have been applied for developing spatial epidemiological models. Some of the traditional spatio-temporal prediction methods include general and generalized linear models (GLM), generalized additive models (GAM) and Bayesian estimation methods that require both presence and absence data of the disease. Newer spatial modeling methods like maximum entropy (MAXENT) and the genetic algorithm for rule set production (GARP) require only

disease presence data. These methods have been used widely in the fields of ecology, species distribution modeling and habitat suitability. For the research reported in this dissertation a knowledge driven approach, multi-criteria decision analysis (MCDA) that uses knowledge of the causal factors of disease occurrence was used to identify areas potentially suitable for harboring the disease. Literature (Hongoh et al. 2011) also recommend this approach for developing spatially explicit decision support tools for vector borne diseases. The objective of this research was to develop a spatially explicit methodological approach for predicting areas vulnerable to WNV risk.

Three questions were defined (Figure 1-2) related to spatially explicit modeling of WNV risk in human population:

- 1) In the absence of case data how can environmental factors be used to predict WNV risk among humans?
- 2) How can spatial variations in the WNV model be determined?
- 3) How can human population characteristics be integrated to further improve the assessment of WNV risk in human population?

Each of these questions is addressed as a chapter in this dissertation and it is written in a manuscript format for publishing purpose.

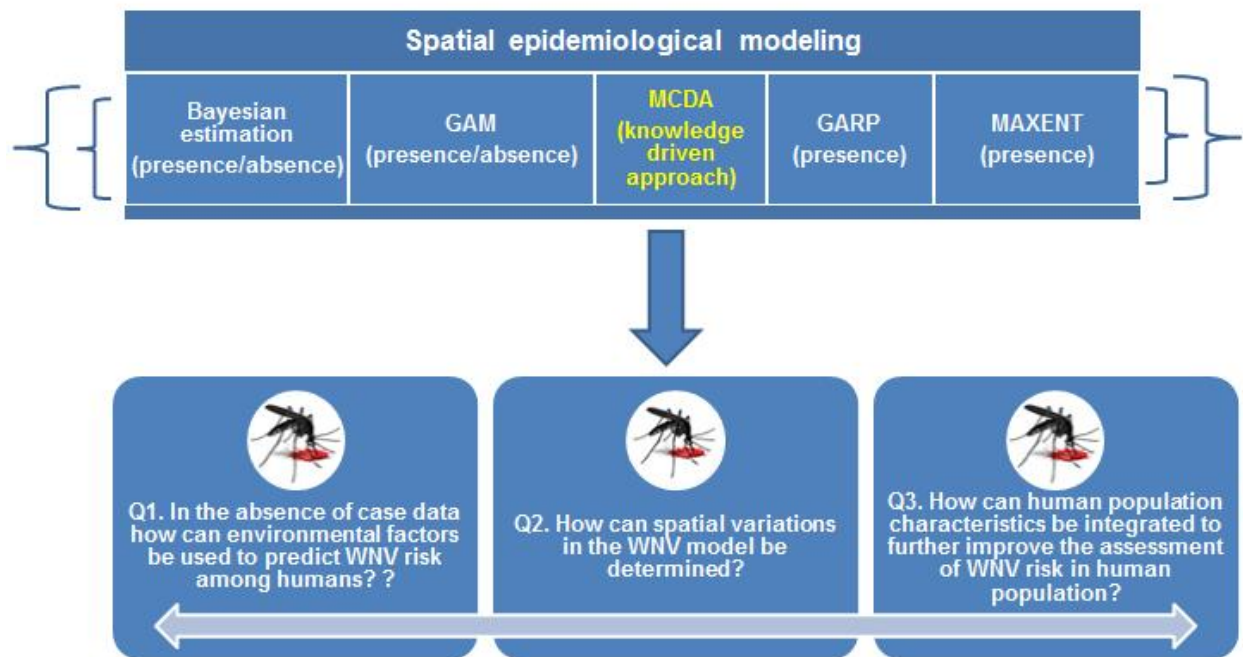


Figure 1-2 Modeling approach and research questions.

1.3 Environmental model

Due to the unavailability of reliable and complete data, developing models of WNV risk pose additional challenges. Human case data is lacking due to issues of under-reporting and limited surveillance. Animal sentinels have proved to be useful for disease prediction modeling and identifying areas for human infection risk (Ann et al. 2009). Several studies (Chaintoutis et al. 2014; Eidson et al. 2001a; Eidson et al. 2001b; Eidson et al. 2001c; Guptill et al. 2003; Johnson et al. 2006; Mostashari et al. 2003; Nielsen & Reisen 2007; Patnaik et al. 2007) have suggested links between infected dead birds and WNV human infection rates. Since wild birds are also primary reservoir hosts (Cooke et al. 2006) for WNV, we utilized this association and determined its utility by correlating the infected dead bird data with human infection rate ($r^2 = 0.409$, p-value

= 0.01). This allowed to the use of infected birds data as a proxy for incomplete human case data. Further, mosquito habitat suitability has been used as a surrogate for estimating WNV infection risk for humans (Cooke et al. 2006). Research suggests that a model based solely on bird data does not determine the severity of the disease nor describes the coincidence of bird exposure to mosquito habitats because bird habitats cover vast areas. In this research a model of mosquito habitat suitability conditions as a predictor of the spatial distributions of infected birds was developed, which in turn can be used to inform WNV disease risk among human populations. The habitats for potential WNV mosquito vectors can be modeled using environmental variables as vectors and pathogen reservoirs overlap under favorable conditions of environmental factors (Rochlin et al. 2011). GIS, geostatistical, and spatial analysis techniques were utilized in this study to evaluate habitat suitability of WNV-carrying mosquitoes. Table 1-1 provides an overview of the environmental factors that are associated with WNV transmission. These include characteristics of a place such as the WNV mosquito species habitat; climatic conditions, topography, and land use/land cover classes such as vegetation, water, and urbanized areas. Spectral indices acquired from satellite imagery provide information about environmental characteristics like land surface temperature, vegetation cover, and moisture (Rodgers & Mather 2006). Additionally, these environmental variables also play an important role in influencing WNV transmission by affecting the gonotrophic cycle of mosquito vectors (Reisen et al. 2006b).

Table 1-1 Environmental factors associated with WNV vector habitat

Factors	Relation to WNV risk	Factors	Relation to WNV risk
Stream (Cooke et al. 2006)	Sites for breeding and resting	Road (Cooke et al. 2006)	Sites for breeding and resting
Temperature (Kuehn 2012; Srivastava et al. 2001; Wimberly et al. 2008b)	Increases growth rate of vector, decreases egg development cycle and shortens extrinsic incubation period of vector	Vegetation (Brownstein et al. 2002; Cooke et al. 2006; Ruiz et al. 2004a; Srivastava et al. 2001)	Sites for breeding and resting
Surface slope (Cooke et al. 2006; Ozdenerol et al. 2008b; Srivastava et al. 2001)	Water stagnation creating mosquito breeding ground	Evapotranspiration (Liu & Weng 2012a; Trawinski & Mackay 2008)	reflects the amount of surface wetness that is related to mosquito abundance
Cultivated land (Kilpatrick 2011)	Preferred natural ground pools	Developed land (Kilpatrick 2011)	warmer micro-climates

1.4 Spatial Heterogeneity

The environmental variables considered in this study are known to vary across space and therefore require an adjustment for spatial autocorrelation and non-stationarity (DeGroot et al. 2008). Other studies suggest that if variables vary across geographic space, then global models may lead to inaccurate predictions and lead to inaccurate decision-making. According to Miller (2012) there are two important

considerations that must be examined when developing spatially explicit environmental disease risk models (Miller 2012). First, spatial autocorrelation considers that disease rates as well as corresponding environmental factors in geographically proximate areas are more related than the geographically distant areas. Second, due to variability in the underlying risk factors for a disease, the model must account for any evidence of spatial non-stationarity (Fotheringham 2009a; Miller 2012).

Geographically weighted regression (GWR) is commonly used to account for these considerations and to produce improved models that enable better spatial inference and prediction. Recent studies have applied GWR modeling to drug-resistant tuberculosis versus risk factors (Liu et al. 2011); environmental factors versus typhoid fever (Dewan et al. 2013); local climate and population distribution versus hand, foot, and mouth disease (Hu et al. 2012); and environmental factors and tick-borne disease (Atkinson et al. 2012; Atkinson et al. 2014; Wimberly et al. 2008a; Wimberly et al. 2008b), all showing that predictor variables varied spatially across large geographic regions, implying that studies examining diseases across geographic space are improved using GWR.

The spatially explicit model that is discussed in this research uses GWR to account for spatial heterogeneity for two reasons: (a) disease risk observed over space are covered by similar environmental conditions that are related to vector habitat suitability and thus to disease transmission and (b) environmental variables that influence WNV risk are not uniformly distributed across the study area. Although many epidemiological models of WNV risk have been developed, few have explicitly considered techniques that account for spatial heterogeneity as discussed above. Such

models assume that the impact of various environmental factors are constant across the study region, which is unrealistic as larger areas display substantial variations in distribution of environmental, socio-economic, and demographic conditions (Goovaerts 2008).

Traditional global regression models such as ordinary least square regression are not designed to address issues of spatial heterogeneity discussed above. This research presents an analysis that utilizes a local regression approach to explicitly link WNV human infection and environmental risk factors. The spatial heterogeneity in these relationships was analyzed using GWR, a local regression analysis that accounts for the spatial non-stationarity present in the independent variables (Fotheringham et al. 2003). The significance of such an approach in predicting WNV risk is that it is likely to yield more accurate predictions. The spatial granularity resulting from this approach provides detailed information that can be used to identify locales where suitable conditions for WNV mosquito habitat exists, thereby allowing public health planners to design better interventions. The broader implication of this research is that it allows public health authorities to design locally specific strategies for places with limited resources.

1.5 Geovisualization framework

Another challenge with spatially explicit disease modeling is the understanding of spatial variations in disease risk across different regions as well as population groups. Oppong & Harold (2009) argue that vulnerability to disease is inevitably tied to specific places whether it involves risk behaviors, population, or place characteristics. Understanding these factors is useful for hypothesis-generation in public health

research, policy-making, and for understanding the underlying spatial processes that produce differential risks. In this research a framework for integrating population characteristics and place characteristics for assessing WNV risk was provided.

While many spatial and non-spatial models have attempted to predict the risk of WNV transmission, efforts to link these factors within a GIS framework are limited. However, the complex relationships that exist between WNV human infection risk and any contextual or compositional factors are poorly understood. In this research, a geovisualization technique was used to explore these relationships. A major challenge for such integration is the high dimensionality of the data making it difficult to tease out patterns of interest.

The lack of any integrated approaches in understanding the effect of these variables on WNV limits the ability to discover potential hidden patterns in a large number of possible combinations from such multivariate datasets. Guo et al. (2005) asserted that even in a selected subset of the data it is still a challenge to discover hidden relationships as potential patterns may take various forms - linear or non-linear, spatial or non-spatial. It is also important that analyses of such complex datasets are performed using methods that are computationally efficient. Tools developed for geovisualization can be used to support multivariate analysis of geospatial data. In this study, contextual environmental factors affecting WNV risk were modeled using geostatistical and spatial analysis techniques. The compositional factors affecting WNV risk are then combined with the contextual model output using geovisualization techniques such as the self-organizing map (SOM) and the parallel coordinate plot (PCP) approaches. When coupled with multivariate visualization, these techniques can

reveal how interactions between the contextual and compositional factors vary locally across geographic space. The clustering and mapping of these factors is then analyzed to identify the places vulnerable to WNV risk. Our study postulates that understanding the underlying place characteristics and population composition for the occurrence of WNV is important for mitigating future outbreaks.

CHAPTER 2

SPATIALLY EXPLICIT MODELING OF WEST NILE VIRUS RISK USING ENVIRONMENTAL DATA

2.1 Abstract

West Nile virus (WNV) is an emerging infectious disease that has widespread implications for public health practitioners across the world. Within a few years of its arrival in the United States the virus had spread across the North American continent. This research focuses on the development of a spatially explicit GIS-based predictive epidemiological model that operates at the local level and enables the study of spatial distribution of WNV. The study attempts to construct a WNV risk model based on suitable environmental factors for predicting mosquito habitats. GIS data processing, geo-statistical measures, and spatial analysis techniques were utilized for detailed investigation. Our model represents a multi-criteria decision analysis approach to create spatially explicit disease risk maps. It also considers the non-stationarity of epidemiologic processes by utilizing GWR technique. The best fitting model with an adjusted R^2 of 0.709 ($p < 0.05$) revealed a strong association between WNV infection risk and a subset of environmental risk factors including road density, stream density, and land surface temperature. This study postulates that understanding the conditions underlying the occurrence of WNV provides tools for detecting, tracking, and predicting the epidemic.

2.2 Introduction

WNV is an emerging infectious disease that has widespread implications for public health practitioners across the world. The disease was first identified in the Western Hemisphere in New York City in 1999. Understanding the spatial distribution of the disease is important as it can help identify areas of risk, improve the specificity of exposure prediction, and therefore help develop better public health intervention strategies. While studies have previously modeled the relationship between environmental factors and WNV risk (Ann et al. 2009; Brown et al. 2008; Gibbs et al. 2006; Ward et al. 2009), they often assume that the effects of various environmental factors are constant across the study region. However, such assumptions are unrealistic as larger areas are known to have substantial local variations in the distribution of environmental factors (Goovaerts 2008). While several spatial and aspatial models have attempted to model the risk of WNV transmission, they typically do not incorporate such localized differences. This research focuses on the development of a spatially explicit GIS-based predictive epidemiological model that operates at the local level and enables the study of spatial distribution of WNV. Our approach incorporates fine-scale environmental variables for modeling habitat suitability of disease-carrying mosquitoes. It also represents a multi-criteria decision analysis technique to create disease risk maps while considering the non-stationary relationship between place and health.

The model introduced in this paper focuses on the habitat suitability for the transmission vector using a variety of spatial and geostatistical techniques. Further, it utilizes spatially indexed environmental data on land use, land surface temperature, vegetation, slope, road density, stream density, and evapotranspiration. Geographically

weighted regression is a spatially explicit technique for diagnosing spatial heterogeneity of the predicted risk and has been utilized in our model. The best fitting model with an adjusted R^2 of 0.709 ($p < 0.05$) revealed a strong association between WNV-infected dead birds and a subset of environmental risk factors including road density, stream density, and land surface temperature, which emphasizes the importance of environmental factors in modeling habitat suitability for disease-carrying mosquitoes.

2.3 West Nile Virus

WNV is a mosquito transmitted zoonotic arbovirus that belongs to the genus *Flavivirus* in the family *Flaviviridae*. It is a vector-borne pathogen that was first discovered in Uganda in 1937 (Smithburn et al. 1940). Since then, its geographic range has expanded throughout Africa and to Asia, Europe, North America, Central and South America, and the Caribbean (Cruz et al. 2005; Hubálek & Halouzka 1999; Komar & Clark 2006; Malkinson & Banet 2002; McIntosh et al. 1969; Morales-Betoulle et al. 2006; Quirin et al. 2004; Steele et al. 2000). Since its arrival in the United States, within a few years the virus had spread across the North American continent (Hayes et al. 2005b). WNV has led to a decline in numerous bird species (CDC); however, an increased morbidity and mortality of humans is of more concern. Although there are licensed vaccines to treat horses infected with WNV, there is currently no immunization available for WNV in humans, making it a continuing public health challenge in the United States for the indefinite future (De Filette et al. 2012). With the spread of this infection across North America, public health strategies have been developed to identify places of high risk of infection to humans, with the goal of prevention.

The distribution of WNV is dependent on the occurrence of susceptible avian reservoir hosts and competent mosquito vectors, mosquito host preference, and availability of hosts. The mosquito life cycle commences with the deposition of eggs by a female mosquito. These eggs develop into larvae which mature to pupae and finally into adults. This development requires the availability of standing water. While seeking a blood meal for egg development, female mosquitoes become infected by acquiring the WNV from infected birds (Blair 2009). A female mosquito may go through a number of egg development cycles during her lifetime. Some mosquito species take more than one blood meal per cycle (Kramer & Ebel 2003). After taking an infectious blood meal, a mosquito can pick up a permanent infection and then has the potential to transmit the virus to multiple hosts while seeking a blood meal. Multiple blood meals increase the host contact as well as the fertility and life-span of these mosquitoes, thereby enhancing viral transmission (Kramer & Ebel 2003). Transmission of the virus is caused primarily through the bite of an infected mosquito when feeding on a susceptible avian host. Infected birds then fly to different locations where the virus is transmitted to other susceptible mosquitoes. Ultimately the disease is transmitted by mosquitoes to humans or other mammals that act as incidental hosts. A study conducted in California has shown vertical transmission of WNV in infected *Culex* mosquitoes that were collected during late summer and fall. The study showed that females having high titered infections were capable of passing the virus onto their offspring destined for overwintering (Nelms et al. 2013). The incubation period of WNV is the time between exposure to WNV and the first appearance of disease symptoms. In humans WNV incubation period ranges from 2-14 days (Campbell et al. 2002). Figure 2-1 gives a

diagrammatic representation of the WNV transmission cycle and life cycle of a mosquito. There are more than 300 bird species (Gubler 2007) that act as reservoirs for WNV. WNV has been detected in more than 60 mosquito species (CDC), though it appears that only a few *Culex* species are implicated in the transmission of WNV infection in North America. WNV-infected dead birds are currently a primary indicator for the presence of disease in a geographic region. Reisen et al. (Reisen et al. 2006a) studied the role of corvids in epidemiology of WNV in Southern California. Their study showed that spatially significant clusters of dead corvids were congruent with clusters of WNV human cases, indicating their importance in viral amplification and as a risk factor for human infection. Because animal sentinels have proved to be useful for disease prediction modeling and identifying areas for human infection risk (Ann et al. 2009), our study focuses on using environmental factors to predict where dead birds infected with the virus are most likely to be found. Furthermore, due to the unavailability of reliable and complete data, developing models of WNV risk pose additional challenges. Human case data is lacking due to issues of under-reporting and limited surveillance. WNV infected birds have been used as a proxy for incomplete human case data (Cooke et al. 2006). However, this data is also spotty and sparsely available as it is not routinely collected and is generally tabulated using user-reported observations of dead birds. Further, mosquito habitat suitability has been used as a surrogate for estimating WNV infection risk for humans (Cooke et al. 2006). In this paper, we follow a similar approach and use a model of mosquito habitat suitability conditions as a predictor of the spatial distributions of infected birds, which in turn can be used to inform WNV disease risk among human populations.

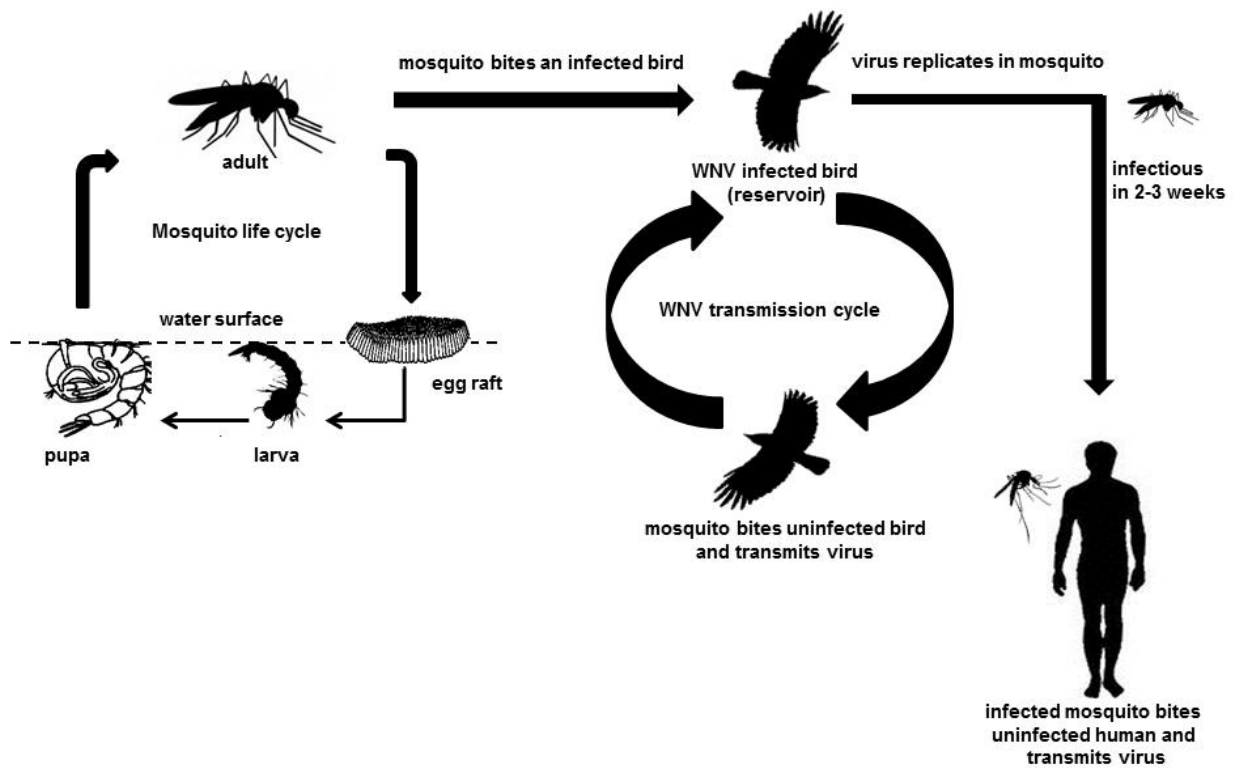


Figure 2-1 West Nile virus transmission cycle and mosquito life cycle.

2.4 Putative environmental risk factors

This research postulates that understanding the conditions underlying the occurrence of WNV is important for mitigating future outbreaks. Vectors and pathogen reservoirs overlap under favorable conditions of environmental factors, distinct landscape features, and ecological settings (Rochlin et al. 2011). The distribution of WNV appears to be dependent upon the presence of both susceptible bird reservoir hosts and the mosquito vectors. Previous studies linking environmental variables to WNV are typically undertaken at a coarse scale and have suggested a wide range of

environmental correlates, including: long-term temperature patterns (Wimberly et al. 2008b), precipitation (Shaman et al. 2005), land-use changes and urbanization (Ezenwa et al. 2007), host biodiversity (Kilpatrick et al. 2006), and global climate change (Harrigan et al. 2014). Areas of high-vegetation abundance and vegetated areas along the rivers are notably associated with WNV risk (Brownstein et al. 2002; Cooke et al. 2006). Low, flat areas with poor drainage have also been considered as preferred habitats for mosquitoes (Cooke et al. 2006; Ozdenerol et al. 2008c). Land use classes of cultivated and developed lands are also associated as preferred vector breeding sites (Bowden et al. 2011; Bradley et al. 2008; Eisen et al. 2010; Kilpatrick 2011). Cooke, Grala, and Wallis (Cooke et al. 2006) investigated environmental conditions favorable for mosquito habitats by including variables such as road density, surface slope, stream density, vegetation, and seasonal water budget in the risk model they developed. They concluded that WNV risk was associated with road density being high, surface slope being gentle, and stream density being low. Srivastava et al. (Srivastava et al. 2001) research on predictive habitat modeling for malaria risk included forest cover, altitude, rainfall, and temperature variables predicting areas of occurrence of *Anopheles* at the micro-level geographic scale. Ruiz et al. (Ruiz et al. 2010) showed consistent and significant impact of Spatio-temporal patterns of air temperature and precipitation on increased mosquito infection. Brown et al. (Brown et al. 2008) systematically determined a significant association of WNV risk with urbanization, suggesting that urbanization enhances environmental conditions for virus transmission. As of the writing of this paper, the most recent study (Valiakos et al. 2014b) relating environmental parameters to WNV found that altitude and distance to water were significant correlates.

Among these environmental variables, ambient temperature correlates with WNV amplification because increased temperatures can result in increased mosquito population, increased maturation of mosquitoes, frequency of mosquito feeding behavior, and decreased extrinsic incubation period of the virus (Wimberly et al. 2008b). This may trigger a shift from WNV amplification within avian communities to human transmission. Kuehn (Kuehn 2012) reported that the 2012 WNV season had 5 times the number of cases that were reported in 2011 and was one of the worst since the virus emerged in the United States. Kuehn's research suggests that the increase in WNV cases concentrated in Texas, Mississippi, Michigan, South Dakota, Louisiana, Oklahoma, and California in 2012 could be attributed to ecological factors, including higher-than-normal temperatures, that may have influenced the abundance of mosquitoes and birds, viral replication in host mosquitoes, and the interactions between birds and mosquitoes. Liu and Weng (Liu & Weng 2012b) in a study on WNV risk in southern California found that one of the main factors contributing to the WNV propagation included land surface temperature. They related higher temperature to viral replication in mosquitoes for WNV to be disseminated throughout the year. The results also show that areas with lower elevations tended to be more susceptible to WNV invasion as mosquito population propagates in the plain habitats with warmer temperatures compared to areas with higher elevation that have lower temperatures.

These studies have resulted in the practice of using dead bird counts or mosquito-abundance data as an indicator of WNV risk. Research suggests that a model based solely on bird data does not determine the severity of the disease nor does it describe the coincidence of bird exposure to mosquito habitats because bird habitats

cover vast areas of land (Cooke et al. 2006). For these reasons, bird data alone may not correctly predict the pattern of WNV infection, and therefore the research reported here involved development of a WNV risk model based on environmental factors related to mosquito habitats.

2.5 Spatial-epidemiology modeling for risk prediction

Traditional spatial-temporal methods for disease prediction include general and generalized linear models (GLM), generalized additive models (GAM), and Bayesian estimation methods. These methodologies require both presence and absence data of the disease. Recent spatial modeling methods like maximum entropy (MAXENT) and the genetic algorithm for rule set production (GARP) require only disease presence data.

Multi-criteria decision analysis (MCDA) modeling uses knowledge of the causal risk factors of disease occurrence to identify potential risk areas. Because of various environmental determinants, the risk for vector-borne diseases can be considered to be heterogeneously distributed in space. The MCDA-based approach is useful to the development of geospatial models and spatially explicit decision support tools for the management of vector-borne diseases (Hongoh et al. 2011). In data-sparse situations, disease risk maps can be produced using a GIS-based MCDA approach. An important element of this approach is professional judgment, which provides a framework for interaction between formal statistical methods and experience-based qualitative knowledge. The MCDA approach can include both quantitative and qualitative variables, and it is able to accommodate the non-linear relationships that are generally

encountered between disease organisms, vectors, reservoirs, and the environment (Stevens & Pfeiffer 2011). Table 2-1 provides a list of epidemiologic studies that have utilized the MCDA approach for modeling disease risk as seen from Table 1. MCDA modeling includes four steps, the first two of which are: 1) establish the factors and constraints and 2) standardize the factors and constraints. Standardization of input factors is important as it incorporates different scales of each factor in the model and also the factors' relationships to the dependent variable (Steele et al. 2009). For standardization, there are different transformation techniques, but within MCDA, the linear transformation method is most commonly used (Malczewski 2000). The final two steps of the MCDA approach include: 3) establish relative weights of each factor and 4) conduct the final evaluation. There are different ways of defining factor weights (Malczewski 2000; Steele et al. 2009), including an analytical hierarchy process (Saaty 1990). The final evaluation is achieved by combining the factors using methods such as weighted linear combination (WLC), ordered weighted averaging (OWA), simple additive weighting (SAW), or Dempster-Shafer theory. (Rakotomanana et al. 2007). Our research utilized the MCDA approach for identifying mosquito habitat suitability areas and creating the WNV risk map.

Table 2-1 Epidemiologic studies utilizing multi-criteria evaluation approach

Epidemiologic study	Publication(s)	Multi-criteria evaluation approach	Method for defining factor weights
Avian influenza virus H5N1	(Stevens & Pfeiffer 2011)	Weighted linear combination	Analytical Hierarchy Process
West Nile Virus	(Cooke et al. 2006)	Simple additive weighting	Rank sum
Malaria	(Hanafi-Bojd et al. 2012; Rakotomanana et al. 2007)	Weighted linear combination, Simple additive weighting	Analytical Hierarchy Process, Evidence based weighting
Trypanosomiasis	(Robinson et al. 2002)	Weighted linear combination	Established by experienced professionals
Rift valley fever	(Clements et al. 2006)	Weighted linear combination, Ordered weighted averages, Dempster-Shafer theory	Relative frequency of reporting in literature, Rank of suitability scores within pixel

Research on spatial epidemiology tends to focus on people and time, with little consideration of the implications of spatial dimensions of disease processes (Rezaeian et al. 2007). Many epidemiological analyses implicitly or explicitly assume the impact of various factors to be constant across the study region, which may be unrealistic for large areas that might display substantial variations in distribution of environmental, socio-economic, and demographic conditions (Goovaerts 2008). The consideration of

place in public health studies is useful for formulating etiological hypotheses and for the efficient distribution of resources for prevention and treatment (Rezaeian et al. 2007). There are various spatial statistical techniques that have improved specificity of exposure and disease relationships and increased our understanding of the relationships between place and health (Auchincloss et al. 2012). Spatial statistics such as Local Indicators of Spatial Associations (LISA) (Anselin 1995), Geographical Weighted Regression (GWR) (Fotheringham et al. 2002), the G statistics (Getis & Ord 1992), local Moran's I, and Geary's C have been developed to measure spatial dependency within local space. These techniques highlight local variation in the study area to identify "hot" or "cold" spots, allowing investigations of non-stationary relationships across space. Spatial autocorrelation results from the lack of independence of data from neighboring areas, which implies that disease rates for geographically proximate areas are more highly related than the geographically distant areas. Simultaneous autoregressive models are methods designed to address spatial autocorrelation, thus accounting for spatial dependency of the data (Rezaeian et al. 2007). The GWR technique is a local regression method that can be used for diagnosing spatial heterogeneity between dependent and explanatory variables over space (Auchincloss et al. 2012). It is performed within local windows centered on the nodes of a regular grid; each observation is weighted based on its proximity to the center of the window. This technique avoids abrupt changes in the local statistics of adjacent windows, helps visualize spatial variability within the geographic entity, and allows analyses of regionally aggregated data of different factors (Goovaerts 2008). Herein, we utilize GWR to emphasize the role that environmental factors play in

modeling habitat suitability for disease-carrying mosquitoes and demonstrate an approach to create disease risk maps under data sparse situations while considering the non-stationary relationship between place and health.

2.6 Materials and Methods

2.6.1 Study area

The study was conducted in the state of California at the county level. California is the third largest state in area in the United States and is made up of 58 counties (Figure 2-2). WNV was first detected in California in July 2003 (Reisen et al. 2004) and the state became the national epicenter of WNV activity in 2004 and 2005 (Jean et al. 2007). California has the highest population in the U.S., but the population is unevenly distributed across the state, due in part to its variable landscape with a central valley bounded by mountain ranges.

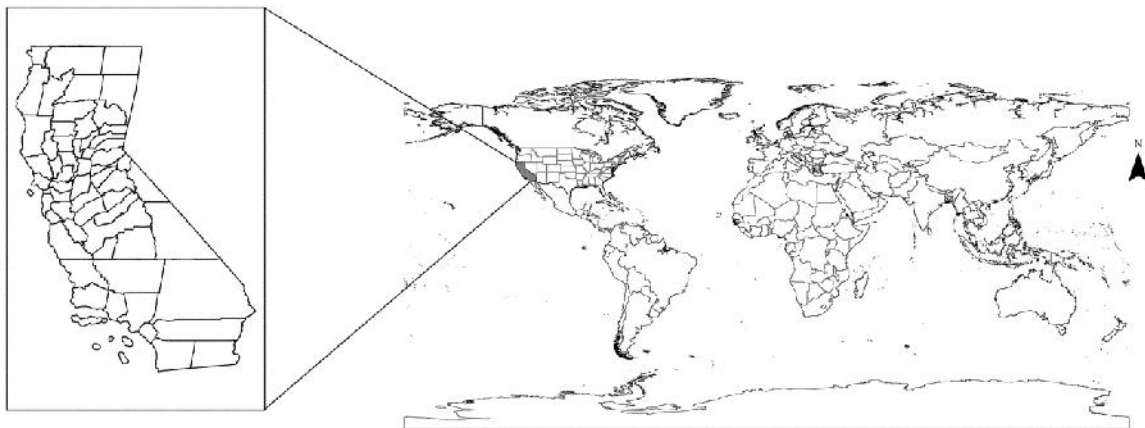


Figure 2.2 World physical map with inset map showing study area.

2.6.2 Study design overview

Clinical data of WNV human incidence cases, infected dead bird counts, and spatially explicit data on environmental factors from 2004 to 2010 were utilized in this study to develop a WNV risk map for California. Our study demonstrates a spatial modeling approach in which temporal data is strengthening the results. Figure 2-3 summarizes the different steps of the WNV risk mapping methodology. The selected environmental variables were categorized in two groups: static environmental factors, which show slow or no change over the time frame of our study, and dynamic environmental factors, which do. GIS data processing, geo-statistical measures, and spatial analysis, along with MCDA and GWR techniques, were utilized in this study to evaluate habitat suitability of WNV-carrying mosquitoes. Finally, a map of disease risk for California was created.

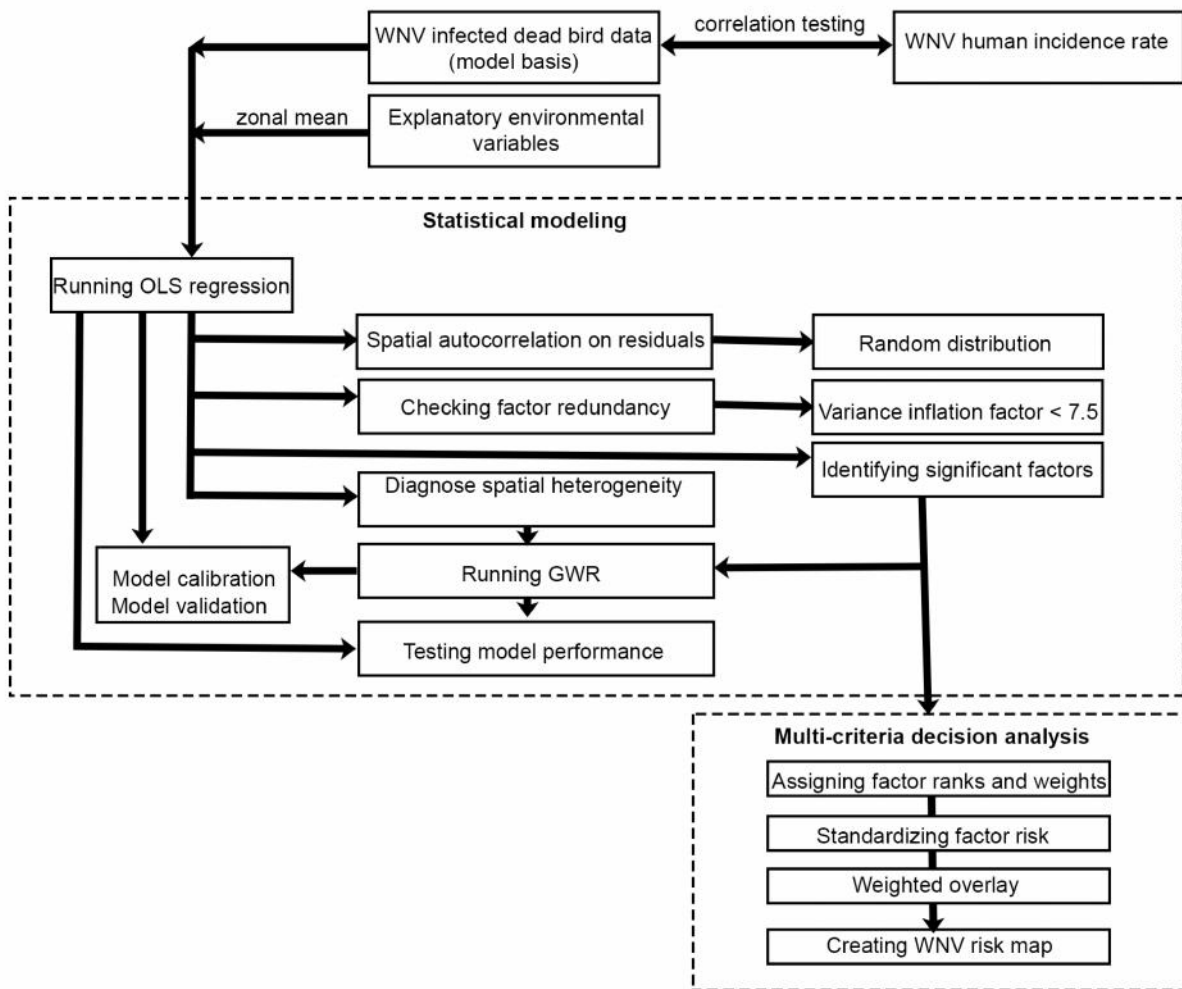


Figure 2-3 West Nile virus model skeleton.

2.6.3 Data sources

The Moderate Resolution Imaging Spectro-radiometer (MODIS) toolbox was downloaded from ArcGIS® Resource Center (<http://resources.arcgis.com>) and incorporated into the ArcGIS 10.0® environment for rapid download of imagery from MODIS obtained from Terra and Aqua satellites. This toolbox also supports the download of various dynamic environmental data products that are required for this study including Normalized Difference Vegetation Index (NDVI), Land surface temperature (LST), and Evapotranspiration (ET). Table 2-2 lists the different datasets

and acquisition sources utilized for this research. Figure 2-4 provides a visual reference to the land cover data that were acquired as an example.

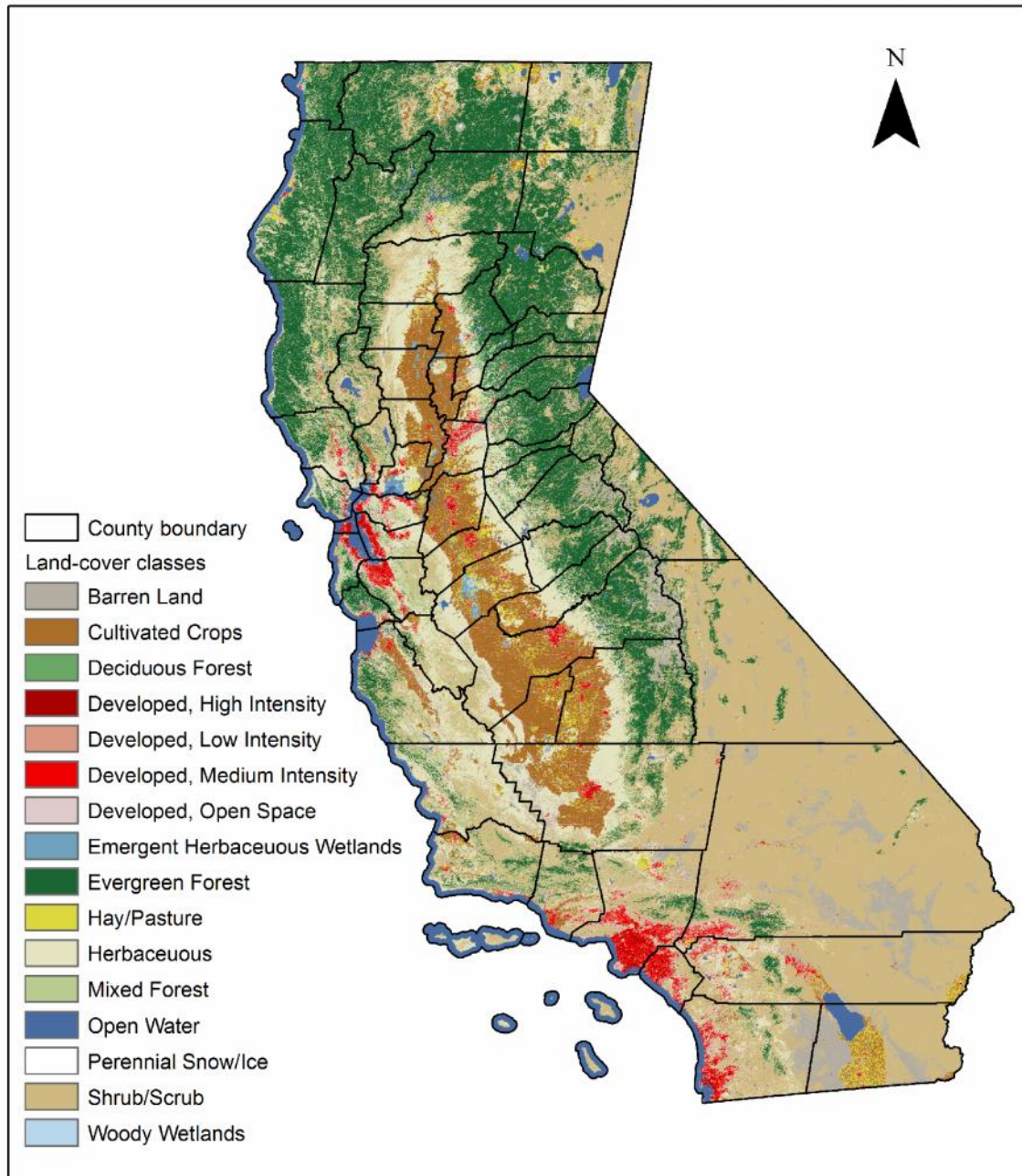


Figure 2-4 Map showing distribution of different land-cover classes in California.

Table 2-2 Sources of data and their spatial resolution

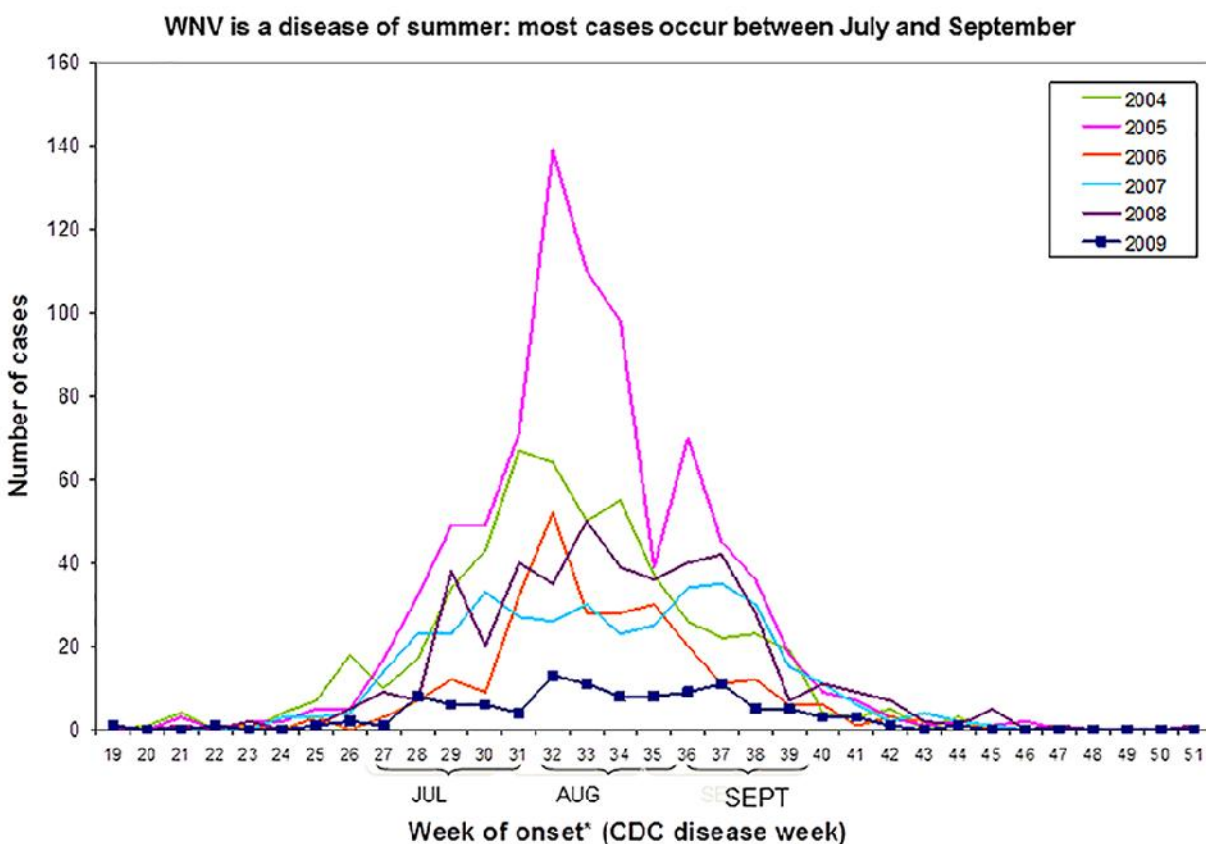
Data	Source	Spatial resolution
Elevation	National Elevation Dataset (NED)	10m
Land Surface Temperature (LST)	MODIS aboard the Terra and Aqua satellites	1 Km
Normalized Difference Vegetation Index (NDVI)	MODIS aboard the Terra and Aqua satellites	250 m
Evapotranspiration (ET)	MODIS aboard the Terra and Aqua satellites	1 Km
Streams	U.S. Bureau of Reclamation	available in vector format for the State of California
Roads	U.S. Census bureau	available in vector format for the State of California
Cultivated land	National Land Cover Database	30 m
Developed land	National Land Cover Database	30 m
WNV infected dead birds count	U.S.G.S. National wildlife health center	County scale
WNV human incidence cases	U.S.G.S. National wildlife health center	County scale
Human population	U.S. Census bureau (2010 census)	County scale

2.7 Data processing and analysis

The data for human incidence cases and WNV-infected dead bird counts were compiled for analysis with other spatial datasets. Human incidence cases were normalized for every 100,000 population, whereas infected dead bird counts were averaged for the study's 7 year period (2004-2010) for each county.

WNV vector-mosquitoes have specific habitat requirements, which were included as factors in this study. Factors selected for this study were based on their established relationship with WNV human risk as determined from previous research. Table 2-3 shows the association of environmental factors with WNV vector habitats. While spatial epidemiological models are useful for disease surveillance and control program if it uses predictors that are easily available and interpretable (King et al. 2004); in practice it is not possible to include each factor that affects habitat suitability. Here, the environmental factors considered most important to mosquito habitat prediction and for which data are easy to acquire in georeferenced format were utilized. Road density, stream density, slope, percent of developed land, and percent of cultivated land were selected as static variables, while land surface temperature, normalized difference vegetation index, and evapotranspiration were selected as dynamic variables for modeling suitable vector habitats. There is a time elapse of 2-14 days between the infection of a human by an infected mosquito and the appearance of disease symptoms. The incubation period of WNV in mosquitoes was taken into account and the climate data used in this study was for the month of July – the month prior to peak WNV human incidence cases. Figure 2-5 presents a graph displaying WNV incidence with maximum number of cases in the month of August. Data for each of the dynamic variables were

downloaded in the form of GIS tiles. The tiles were mosaicked and clipped to the outline of California, the study area boundary, in ArcGIS. The modeling method utilized in this study was based on analyzing data in raster format, and therefore road and stream vector data were converted to raster format using the 'Kernel Density Estimation' tool in ArcGIS to create road density and stream density grid files. The tool assumes a Gaussian distribution and thus assigns more importance towards the center of kernel in comparison to the features that are further apart. Figure 2-6 shows kernel density maps of major roads and streams for California.



* Onset dates known for 2,518 (88%) cases

Figure 2-5 Graph showing WNV disease onset week for California, 2004-2009 (CDPH)

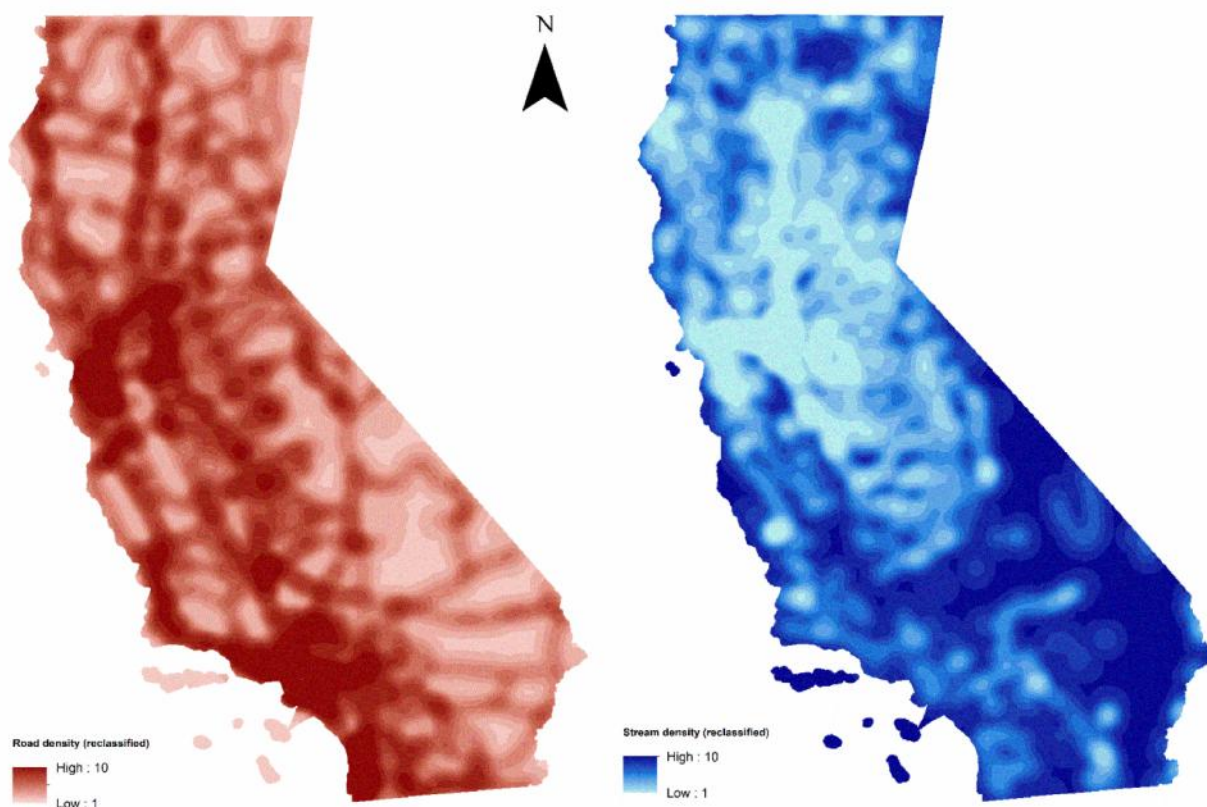


Figure 2-6 Kernel density maps of major roads and streams in California.

To initiate modeling, correlation analysis was performed on infected dead bird counts versus human incidence rates to confirm that dead birds can be used as a surrogate of WNV human risk. A positive moderate correlation ($r^2 = 0.409$; $p = 0.01$) was found, and, therefore, infected dead bird data were utilized as the dependent variable in the model. Zonal Statistics tool in ArcGIS was used to compute the statistical mean values for the independent variables at county scale. This was required for regression analysis, also performed in ArcGIS, for the purpose of exploring spatial relationships and understanding the factors responsible for observed patterns. For this research, the Exploratory Regression tool in ArcGIS, similar to stepwise regression, was used to evaluate possible combinations of the variables, in order to choose the model

that best explains the dependent variable under specified criteria. The ordinary least squares (OLS) regression model was run on the dataset with infected dead bird counts as the dependent variable and environmental factors as independent variables. Moran's Index of spatial autocorrelation was used to check if the residuals from the model output exhibited a random spatial pattern. In order to reduce over-estimation bias in the results, redundancy among the explanatory variables was checked to examine overlap among variables that might contribute to multi-collinearity. The OLS model allowed identification of statistically significant explanatory variables using adjusted R^2 values and the corrected Akaike's Information Criterion (AICc) (Hurvich et al. 1998). The Jarque-Bera test statistic was also employed to test for model bias (Jarque & Bera 1980).

Table 2-3 Association of environmental factors with WNV vector habitat

Environmental factors considered	Publication(s)	Relation to WNV vector habitat
Slope	(Ozdenerol et al. 2008c), (Cooke et al. 2006),(Srivastava et al. 2001)	Water stagnation/ outflow rate
Stream	(Cooke et al. 2006)	Water used for breeding
Road	(Cooke et al. 2006), (LaBeaud et al. 2008)	Sites for breeding along the roads
NDVI	(Ozdenerol et al. 2008a),(Cooke et al. 2006), (Brownstein et al. 2002),(Srivastava et al. 2001), (Ruiz et al. 2004a)	Vegetation areas are used for breeding and resting
LST	(Srivastava et al. 2001),(Wimberly et al. 2008b),(Kuehn 2012)	Increase growth rates of vector, decrease the length of gonotrophic cycle, shorten the extrinsic incubation period of the virus in the vector, increase the rate of virus evolution
Evapo-transpiration	(Trawinski & Mackay 2008), (Liu & Weng 2012a)	Reflects the amount of surface wetness that is related to mosquito abundance
Cultivated land	(Kilpatrick 2011), (Eisen et al. 2010)	Natural ground pools preferred by some vector species
Developed land	(Kilpatrick 2011), (Bowden et al. 2011), (Bradley et al. 2008)	increasing breeding sites or warmer microclimates

The OLS assumes that the variation in the coefficients are the same for every area (spatially stationary), but many spatial epidemiologic processes are non-stationary in nature. The Koenker statistic was used to determine if explanatory variables have a consistent relationship with our dependent variable across the study area. Finding conditions of non-stationarity model results were improved by use of GWR that assumes that relationships can be stronger in one area than in another. The GWR model was executed on the dataset with infected dead bird counts as the dependent variable and independent variables comprising the statistically significant environmental factors as determined by the OLS regression model results. Similar to the OLS model, spatial autocorrelation analyses were performed on the model residuals to check if over and under predictions were randomly distributed. Model performance was again measured using adjusted R^2 values and ultimately the model was calibrated using the AICc method. The AICc method compared GWR model performance to the OLS model and indicates improvement in model performance (a decrease of more than 3 points in AICc value indicates a real improvement in model performance) (Charlton et al. 2009; Krivoruchko 2011). The model was developed using two-thirds of California (randomly selected). Those results were validated on the remaining one-third of the counties. Figure 2-7 shows a map that represents the observed count of infected dead birds and the predicted dead birds for each county in California. Comparing the predicted results for each county, it appears that the model is relatively erratic when dead bird count is low. This indicates that a more reliable model can be built when the dead bird count reaches a minimum threshold.

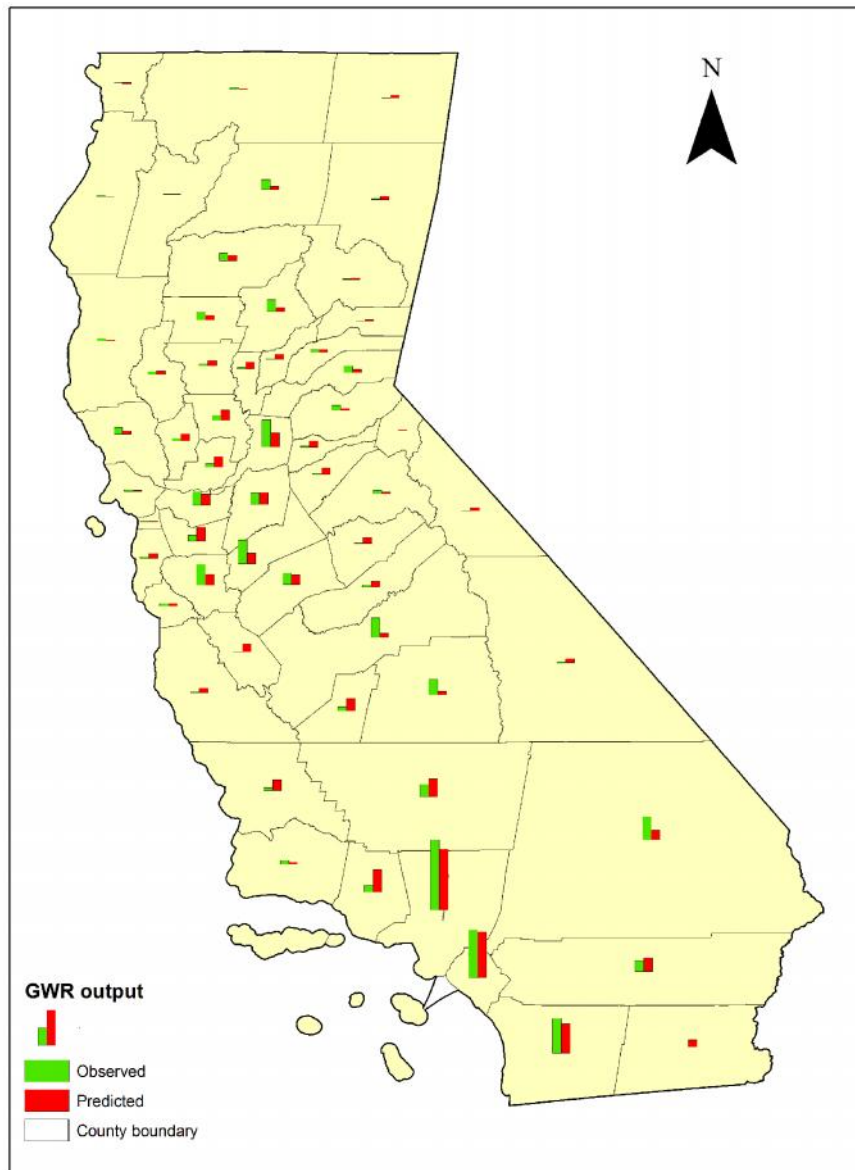


Figure 2-7 Observed Vs Predicted infected dead birds.

The MCDA approach was utilized to create a WNV disease risk map. This modeling process was based upon four steps: 1) establishment of the environmental factors, 2) standardization of the factors, 3) establishment of relative weights for each factor, and 4) conduction of the final evaluation (Rakotomanana et al. 2007).

Determination of relative factor weight was a crucial step in the modeling process. The

significant environmental variables obtained from the spatial regression output were ranked in order of their statistical significance, with more significant factors being assigned a higher rank. Numerical weights were then calculated based on the Rank Sum method that involved normalized summation of weights derived for each criterion and is considered useful when the number of criteria being evaluated is relatively small (Liu & Mason 2009), as follows:

$$w_i = \frac{n - r_i + 1}{\sum (n - r_i + 1)}$$

where w_i is the normalized weight for i^{th} criterion, n is the number of criteria being evaluated, and r_i is the rank position of the i^{th} criterion.

Once the factors were ranked and weighted, each factor layer was standardized for the “risk scale” of 1 to 10, where 1 represents least risk and 10 represents highest risk. This step was executed using the spatial analyst extension in ArcGIS. Quantile classification was used for standardization so that each risk class contained an equal number of pixels. An important consideration during the risk standardization assignment process was the coefficient sign since a negative coefficient creates an inverse risk scale where 10 represents the least risk and 1 represents the highest risk. Finally, a Simple Additive Weighting (SAW) method was applied to construct the disease risk surface. For this method, each layer was multiplied with its weight before overlaying and summing all the factor layers together to produce the output of the disease risk map.

2.8 Results and discussion

Correlation analysis performed on infected dead bird counts and human incidence rate showed a positive significant correlation ($r^2 = 0.409$; $p = 0.01$) (see Figure 2-8). Quantile classification was used for standardization and comparison of these two maps. Because of this relationship, we suggest that a model that can predict dead bird counts may be useful for planning intervention and response plans to reduce human risk of WNV disease. The analysis demonstrated that a number of environmental variables showed predictive value for WNV disease risk over the seven-year time period our data covered. Significant associations were determined between infected dead bird counts and environmental risk factors, including high road density values, low stream density values, and land surface temperature (LST). The resulting environmental risk factors are in accordance with various research publications recommending that environmental data is effective in WNV risk prediction by focusing on vector habitat requirements (Table 2-3). For the static environmental variables, road density and lower stream density were significant predictors. Ditches along roadways are likely places for stagnant water, thus optimal for mosquito breeding. Contrary to our initial thoughts, stream density demonstrated a negative relation to disease risk. This may reflect that flowing water is normally not suitable for larval development of the various species of mosquitos that commonly transmit WNV or that rasterizing the stream database into stream density introduces a component that is not yet fully understood. For our dynamic environmental variables, the analysis determined that LST is a good predictor of the abundance of infected dead bird counts over the seven year period; temperature increases the growth rates of mosquito populations, decreases the

length of the gonotrophic cycle (interval between blood meals), shortens the extrinsic incubation period of the virus in the vector, and increases the rate of virus evolution.

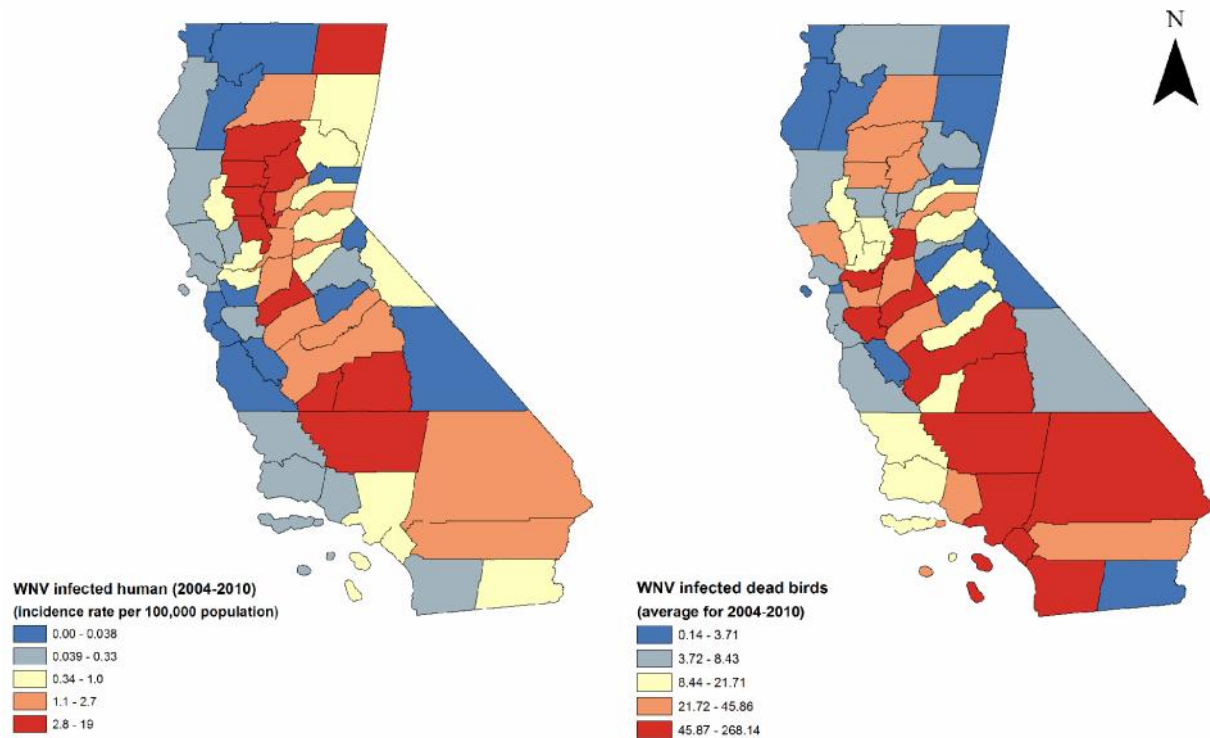


Figure 2-8 Maps of WNV human incidence rate and WNV infected dead birds.

Results of model performance from OLS regression gave a significant adjusted R^2 value of 0.617 ($p < 0.05$) and AICc value of 567.70. Moran's index of spatial autocorrelation exhibited a complete spatial randomness of residuals from OLS regression model thereby confirming that our model assumptions were met. The Jarque-Bera test statistic was not significant, implying that the model was not biased and that the regression residuals were normally distributed. Multi-collinearity amongst explanatory variables was checked using the variance inflation factor that indicated that the factors did not exhibit redundancy. The Koenker test statistic was found to be

statistically significant indicating non-stationary relationships across the study area for our dependent variables. Moving to GWR resulted in model improvement, increasing the adjusted R^2 value to 0.709 ($p < 0.05$). Also, upon running the GWR model, the AICc value decreased to 551.40, a drop of nearly 16 points, showing it to be a much better model than the OLS regression model. Figure 2-9 displays our WNV human infection risk map for 2005 using significant variables in MCDA model. Based on a threshold count of at least forty infected dead birds, fifteen counties were selected. Executing the model on this dataset gave an adjusted R^2 of 0.82 ($p < 0.05$). Figure 2-10 graphically compares the observed and predicted dead bird counts for the selected counties.

There are several assumptions upon which this study is based, most importantly that factors suitable for mosquito habitat leads to higher mosquito populations, increasing the likelihood that WNV is spread to human populations. It was also assumed that the probability of human infection is higher in counties with multiple confirmed WNV bird cases. In reality, a potential bias in this assumption may be that human population density, variations in level of public concern, and resource availability might bias the reporting of dead birds. Another assumption in this study is that people are infected within the county of their residence, ignoring the possibility of contracting an infection while traveling outside their county of residence. This could result in an overestimation of locally occurring WNV human infection risk. On the other hand, human WNV cases may be under reported because some people may not have access to healthcare services or the illness presents with mild symptoms resulting in misdiagnosis. Additionally, mosquito control effort information was not integrated into the model even though active control measures would likely influence the model results.

Finally, data aggregation and smoothing processes were likely to introduce some uncertainty into the model. Despite the assumptions, limitations and technical caveats, we suggest that the model has enough explanatory power that can be potentially useful at a coarse geographic scale for public health surveillance, intervention, and response for WNV risk.

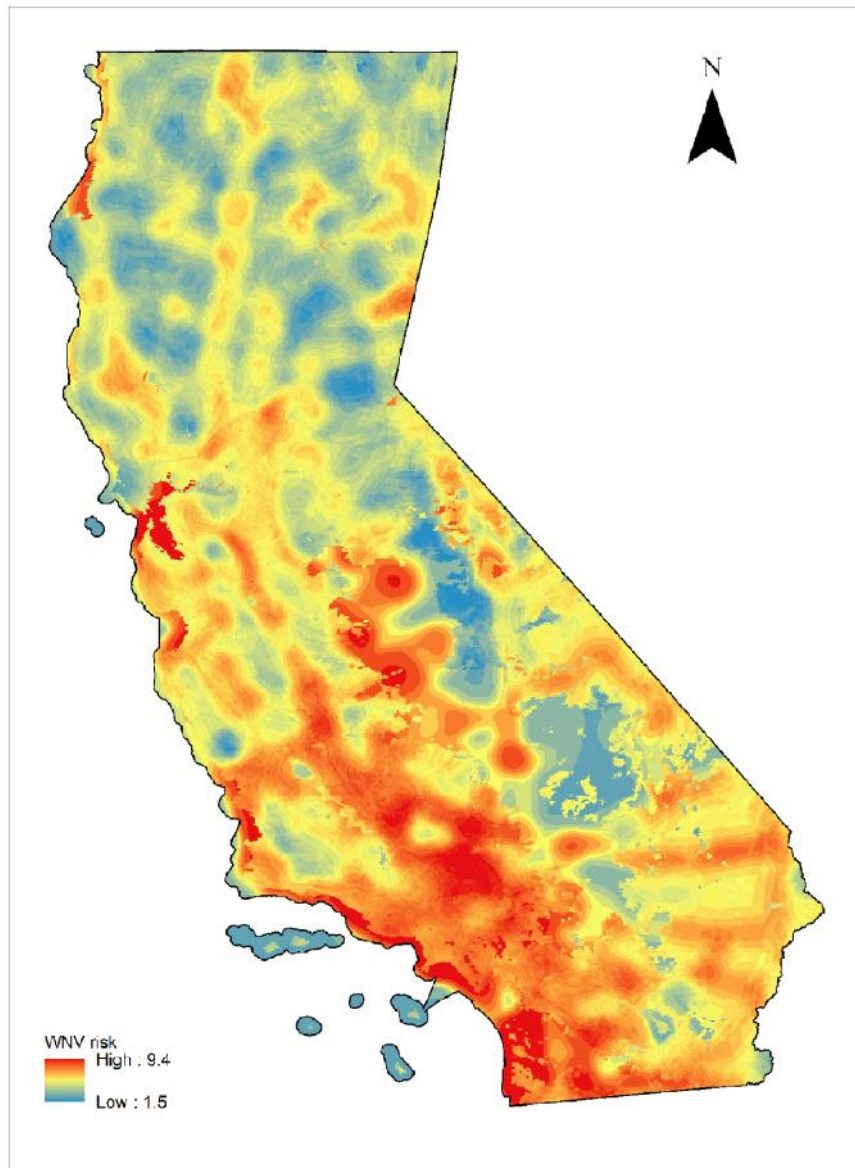


Figure 2-9 WNV human infection risk map for 2005.

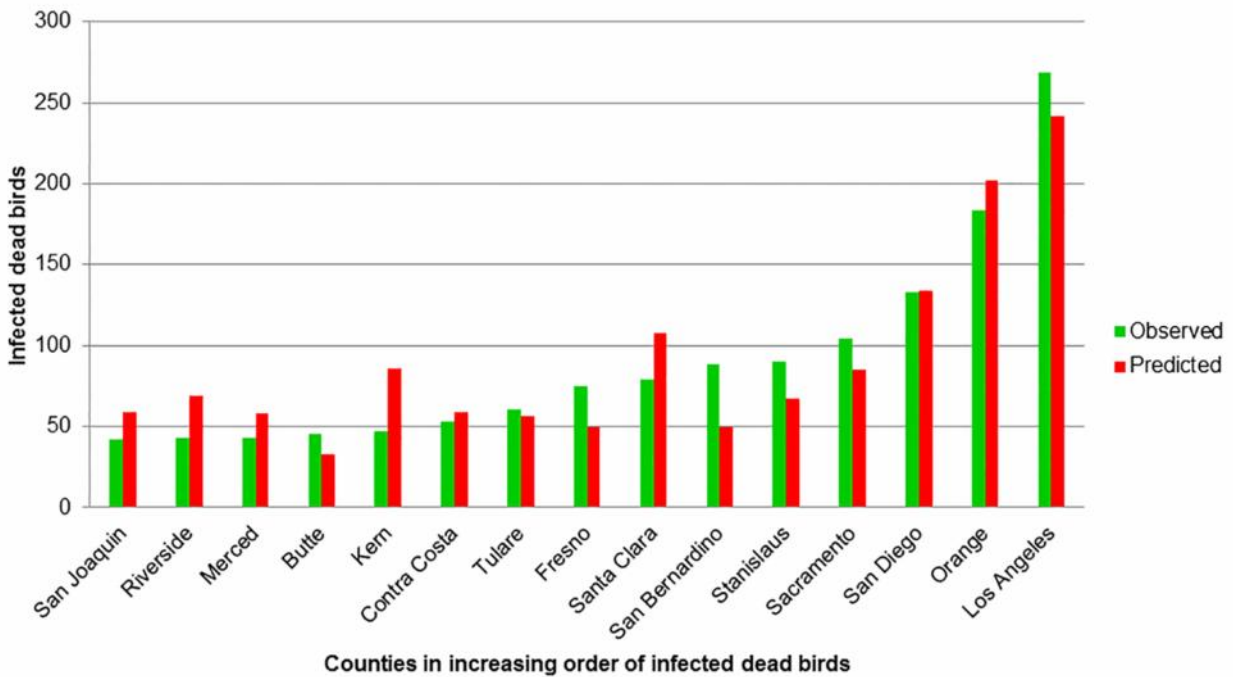


Figure 2-10 Counties with minimum threshold of WNV infected dead birds (at least 40 dead birds infected with WNV; $r^2=0.82$)

2.9 Acknowledgments

The authors wish to thank the Center for Computational Epidemiology and Response Analysis, the Center for Remote Sensing, and the Institute of Applied Science at the University of North Texas for technical advice and access to software and hardware to undertake this project.

CHAPTER 3

USING LOCAL MODELING APPROACH TO EXPLORE SPATIAL HETEROGENEITY IN WEST NILE VIRUS RISK

3.1 Abstract

The primary aim of the study reported here was to determine local spatial variations in the statistical relationships between West Nile Virus (WNV) transmission and environmental risk factors. Because standard regression methods do not account for autocorrelation and non-stationarity of the type of spatial data analyzed for this study, and because they often violate the model parameters, we hypothesized that a geographically weighted regression model would help us better understand whether environmental factors are related to WNV risk patterns without the confounding effects of spatial non-stationarity. We examined eleven commonly mapped environmental factors using both ordinary least squares regression and geographically weighted regression. Ordinary least squares regression efforts lead to identifying three environmental variables that were statistically significantly related to WNV risk (adjusted $R^2=0.61$): stream density, road density, and land surface temperature. Geographically weighted regression helps us improve the explanatory value of these three environmental variables with better spatial preciseness (adjusted $R^2 = 0.71$, $p<0.5$). Public health professionals may benefit from this understanding of environmental spatial heterogeneity in relation to WNV transmission risk to human populations.

3.2 Introduction

WNV is an emerging vector-borne disease that was first detected in the United States in 1999 (Nash et al. 2001). Within a few years the virus had spread across the North American continent (Hayes et al. 2005b). WNV has had important environmental and human impacts, including a decline in numerous bird species (CDC) and increased morbidity and mortality among humans. This has also resulted in increased economic burdens (Barrett 2014). We became interested in developing a spatially explicit model using environmental factors to predict WNV risk.

There are two important considerations that should be examined when developing spatially explicit environmental disease risk models (Miller 2012). First, spatial autocorrelation (the degree to which a set of spatial features and their associated data values tend to be clustered together in space) should be considered. This will account for whether environmental factors and the corresponding disease rates in geographically proximate areas are more or less clustered together than they are in geographically distant areas. Second, data non-stationarity (changing means, variances and covariances in data across space) should be controlled (Fotheringham 2009a; Miller 2012). Geographically weighted regression (GWR) is commonly used to account for these considerations and to produce improved models that enable better spatial inference and prediction. Recent studies have applied GWR modeling to drug-resistant tuberculosis versus risk factors (Liu et al. 2011); environmental factors versus typhoid fever (Dewan et al. 2013); local climate and population distribution versus hand, foot, and mouth disease (Hu et al. 2012); and environmental factors and tick-borne disease (Atkinson et al. 2012; Atkinson et al. 2014; Wimberly et al. 2008a; Wimberly et al.

2008b), all showing that predictor variables varied spatially across large geographic regions, implying that studies examining diseases across geographic space are improved using GWR.

The spatially explicit model that is discussed in this paper uses GWR to account for spatial heterogeneity for two reasons: (a) WNV disease risk observed across space may be related to similar environmental variables that increase vector habitat suitability and (b) environmental variables that influence WNV risk are not uniformly distributed across the study area. Although many epidemiological models of WNV risk have been developed, there has been little research that has explicitly examined techniques that account for spatial heterogeneity. Most models assume that the impact of various environmental factors are constant across the study region, which is unrealistic as larger areas display substantial variations in distribution of environmental, socio-economic, and demographic conditions (Goovaerts 2008).

Due to the unavailability of reliable and complete data, developing models of WNV risk pose additional challenges. Human case data is lacking due to issues of under-reporting and limited surveillance. WNV-infected birds have been used as a proxy for incomplete human case data (Cooke et al. 2006). However, these data are also spotty and sparsely available as it is not routinely collected and is generally tabulated from user-reported observations of dead birds. Further, mosquito habitat suitability has been used as a surrogate for estimating WNV infection risk for humans (Cooke et al. 2006). For our study, we followed a similar approach and used a model of mosquito habitat suitability condition as a predictor of the spatial distributions of infected birds, this in turn can be used to inform WNV disease risk among human populations. Further,

because the environmental variables considered in this study are known to vary across space, we account for spatial autocorrelation and non-stationarity using Geographically Weighted Regression (DeGroot et al. 2008) in order to improve the predictability of a model.

The spatial granularity resulting from this approach provides detailed information that can be used to identify locales where suitable conditions for a WNV mosquito habitat exists, thereby allowing public health planners to design better interventions. The broader implication of this research is that it may encourage public health authorities to design more effective and efficient locally specific strategies for places with limited resources.

3.3 Modeling WNV risk in a local environment: Transmission and risk factors

The WNV transmission cycle was an important component of the model discussed in this paper. The first step in the WNV transmission cycle primarily occurs when a competent female mosquito vector bites an infected bird reservoir host, which in turn results in the virus being transmitted to the mosquito (Blair 2009). This occurs when the female mosquito is seeking a blood meal to obtain nutrients necessary for egg development. After taking an infectious blood meal, a mosquito may pick up a permanent infection. The infected mosquito now has the potential to transmit the virus to another bird or animal when it feeds again. Infected birds then fly to different locations where the virus is transmitted to susceptible mosquitoes. Ultimately the disease is transmitted by mosquitoes to humans or other mammals that act as incidental hosts. Dead birds found to be infected with WNV are currently the primary

indicators for the presence of disease in a geographic region and have proven to be useful for disease prediction modeling and identifying areas for human infection risk (Cooke et al. 2006; Ruiz et al. 2004b; Valiakos et al. 2014a).

Vector and pathogen reservoirs overlap when certain environmental conditions are present (Rochlin et al. 2011). Table 3-1 provides an overview of the environmental conditions that are associated with WNV transmission, which were utilized for our research. These include characteristics of a place such as the mosquito species habitat: climatic conditions, topography and land use/land cover classes such as vegetation, water, and urbanized areas. Spectral indices acquired from satellite imagery provide information about environmental characteristics like temperature, vegetation cover, and moisture (Rodgers & Mather 2006).

Table 3-1 Environmental conditions related with WNV human transmission risk

Factors studied (reference)	Relation to WNV risk	Factors studied (reference)	Relation to WNV risk
Stream (Cooke et al. 2006; Curtis et al. 2014; Schurich et al. 2014)	Sites for breeding and resting	Road (Cooke et al. 2006)	Sites for breeding and resting
Temperature (DeGroote et al. 2014; Kuehn 2012; Srivastava et al. 2001; Wimberly et al. 2008b)	Increases growth rate of vector, decreases egg development cycle and shortens extrinsic incubation period of vector	Vegetation (Brownstein et al. 2002; Cooke et al. 2006; DeGroote et al. 2014; Ruiz et al. 2004a; Schurich et al. 2014; Srivastava et al. 2001)	Sites for breeding and resting.
Surface slope (Cooke et al. 2006; Ozdenerol et al. 2008b; Schurich et al. 2014; Srivastava et al. 2001)	Water stagnation creating mosquito breeding ground	Evapotranspiration (Liu & Weng 2012a; Trawinski & Mackay 2008)	Reflects the amount of surface wetness that is related to mosquito abundance
Cultivated land (Kilpatrick 2011)	Preferred natural ground pools	Developed land (Kilpatrick 2011)	Warmer micro-climates

3.4 Modeling WNV risk in a local environment: statistical considerations

Miller (2012) suggests that a 'global' model is the one that assumes that the parameters (commonly mean and variance) of some process are constant across geographic space. This is referred to as the spatial stationarity of a process. Miller suggests that in case these parameters vary across geographic space (spatial heterogeneity), then such models may lead to inaccurate predictions and incorrect decision-making. In an ecological context, spatial heterogeneity usually results from the interaction of various environmental processes that operate at different scales (Legendre 1993). Fotheringham (Fotheringham 2009b) used local statistics for linking the concepts of spatial autocorrelation and heterogeneity that are deemed important when developing spatial models. Local statistics disaggregate a global mean value into locally computed values for each spatial unit. It is based on a conceptualization of Tobler's first law in Geography (Tobler 1970) that specifies that observations that are located closer together are more related than those situated further apart. Spatial autocorrelation is a commonly used measure of the degree of spatial heterogeneity.

GWR is a local regression method that can be used for diagnosing spatial heterogeneity between dependent and explanatory variables over space (Fotheringham et al. 2003). It is performed within local windows centered on the nodes of a regular grid. Each observation within the local window is weighted based on its proximity to the center of that window. This approach has several advantages: it avoids abrupt changes in the local statistics computed for adjacent windows, helps visualize spatial variability within the geographic entity, and allows analysis of regionally aggregated data of different factors (Goovaerts 2008). A model's predictive ability, particularly in ecological

modeling is influenced not only by the strength of relationships between the species and its environment, but whether the model recognizes if the relationships are operating at multiple spatial scales. GWR provides a framework for exploring scale-dependent effects. It tests the effect on model's predictive ability by systematically increasing the local window (Miller 2012).

GWR has been used for mapping the spatial distribution of the model's coefficient values in order to identify potential missing variables or to suggest other underlying factors associated with the observed non-stationarity (Miller 2012). GWR is also useful for exploratory data analysis and visualization. Kupfer and Farris (2007) (Kupfer & Farris 2007) used a 'leave-one-out' (jackknifing) methodology to compare residuals from GWR and ordinary least squares regression. They found that GWR often had more accurate predictions for sites that were difficult to predict (where both models had overall higher residuals), which is why we used a GWR framework for explicitly modeling the spatial relationships between WNV and its environmental risk factors.

3.5 Study area and data

3.5.1 Study area

The model is built for the state of California, which was the national epicenter of WNV activity in 2004 and 2005 (Jean et al. 2007). WNV was first detected there in July 2003 (Reisen et al. 2004). It is the third largest state by area in the United States and is made up of 58 counties. California has the largest population in the U.S., but it is

unevenly distributed across the state. The state also has a variable landscape with a large valley in the middle, bounded by mountain ranges.

3.5.2 Environmental factors and data sources

Our model considers various environmental factors (Table 3-1) that have been suggested as descriptive in local WNV risk distribution: surface slope, density of roads, density of streams, monthly mean temperature, monthly mean evapotranspiration, and land cover classes like vegetation, developed land, cultivated land, and open surface water. All environmental parameters except roads and streams (Table 3-2) were acquired in grid format and resampled to 120 meter resolution as suggested by Cooke et al. (2006).

Various dynamic environmental data including Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), and Evapotranspiration (ET) were downloaded from Moderate Resolution Imaging Spectro-radiometer (MODIS) toolbox incorporated in ArcGIS®. The time lapse of 2-14 days between the infection of a human by an infected mosquito and the appearance of WNV disease symptoms, referred to as incubation period (Campbell et al. 2002), was taken into account and the environmental data used for this study was for the month of July, the month prior to peak WNV human incidence cases.

Table 3-2 Data sources

Data	Spatial resolution	Source
Elevation	10m	National Elevation Dataset (NED)
LST	1 Km	MODIS aboard the Terra and Aqua satellites
NDVI	250 m	MODIS aboard the Terra and Aqua satellites
Evapotranspiration (ET)	1 Km	MODIS aboard the Terra and Aqua satellites
Streams	available in vector format	U.S. bureau of reclamation
Roads	available in vector format	U.S. Census bureau
Cultivated land	30 m	National Land Cover Database
Developed land	30 m	National Land Cover Database
WNV infected dead birds count	County scale	U.S.G.S. National wildlife health center
WNV human incidence cases	County scale	U.S.G.S. National wildlife health center
Human population	County scale	U.S. Census bureau

3.6 Methods

3.6.1 OLS modeling

WNV disease annual incidence rate (cases per 100,000 populations) was used as the measure of disease severity for purposes of this study. Annual WNV-infected dead birds sentinel data, averaged for 2004-2010, was used as a surrogate of WNV risk, which was used as the dependent variable for modeling purposes. Several studies (Chaintoutis et al. 2014; Eidson et al. 2001a; Eidson et al. 2001b; Eidson et al. 2001c; Guptill et al. 2003; Johnson et al. 2006; Mostashari et al. 2003; Nielsen & Reisen 2007; Patnaik et al. 2007; Ruiz et al. 2004b) have suggested links between infected dead birds and WNV human infection rates. Since wild birds are the primary reservoir hosts for WNV and indicate human infection risk, we utilized this association to develop the prediction model and to examine the extent to which infected dead birds are associated with WNV infection risk. We determined the utility of the infected dead bird data by correlating ($r^2 = 0.409$; $p = 0.01$) it with the human WNV infection rate. We explored all environmental factors listed in Table 3-1 as potential independent variables in our OLS regression modeling.

Interpretations of OLS model performance were based on assessing multi-collinearity, robust probability, adjusted R^2 , and Akaike's information criterion (AIC) (Akaike 1974). Multi-collinearity was assessed through the variance inflation factor (VIF) statistic, which measures redundancy among explanatory variables. Explanatory variables associated with VIF values larger than about 7.5 indicate that these variables are providing similar information, and they were removed one at a time from the model based on VIF value until the model became unbiased. Robust probability indicates the

statistically significant variables that are important to the regression model. Upon examining the VIF values and the robust probability, we re-ran OLS model on the non-redundant and significant variables, which are LST, stream density, and road density. Akaike's information criterion approach lead to our best OLS model, which identified land surface temperature (VIF = 1.046), stream density (VIF = 1.177), and road density (VIF = 1.143) as statistically significant ($p < 0.5$) variables related to WNV risk:

$$\text{WNV risk} = -75.87 + 595.60 (RD) + 1.89 (LST) - 146.89 (SD) \quad (1)$$

Where:

WNV risk = average infected dead bird count

RD = road density

LST = land surface temperature

SD = stream density

Spatial autocorrelation (Global Moran's I) was utilized to assess whether the environmental factors exhibited a random spatial pattern (Goodchild 1986) where a good model has a random distribution of the residuals (Mitchell 2005). Figure 1 indicates that the histogram of the OLS model's residuals approximates that of a normal curve, the non-significant (0.134, $p < 0.05$) Jarque-Bera statistic (Jarque & Bera 1980), and the Moran's I Index Z-score (1.23) all imply that the model is unbiased and significantly different than random.

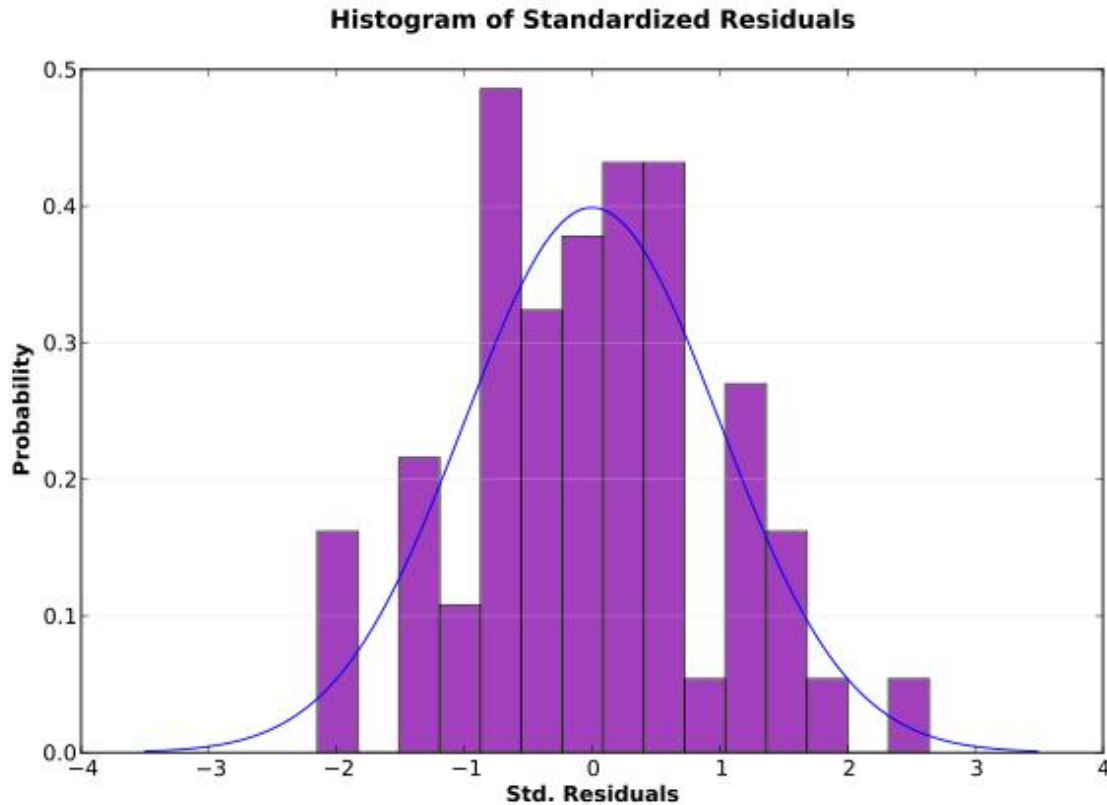


Figure 3-1 Distribution of model's standardized residuals (line indicates normal curve).

However, the Koenker statistic (0.000007^* , $p < 0.05$) confirmed non-stationarity in the OLS model indicating that there is not a consistent relationship between the explanatory variables and WNV risk across the study area, indicating that the OLS model is stable but non-stationary, confirming that proceeding with GWR model is valid.

3.6.2 GWR model specification

Finding conditions of non-stationarity in our ordinary least squares (OLS) model, we explored geographically weighted regression (GWR) in an attempt to improve model results. Using the same dependent and explanatory variables as identified in the OLS

modeling, we proceeded to explore the spatial variability of local regression coefficients to determine whether the underlying process exhibited spatial heterogeneity (Fotheringham et al. 2003). A GWR local model was applied to analyze how the relationship between infected dead bird counts and environmental factors changed from one county to another. Unlike conventional OLS regression modeling, which produces a single regression equation to summarize global relationships among the independent and dependent variables, GWR detects spatial variation of relationships in a model and produces information useful for exploring and interpreting spatial non-stationarity (Fotheringham et al. 2003). The GWR model in this study was implemented using ESRI®ArcGIS10.1 and can be rewritten as:

$$WNV\ risk_{(i)} = \beta_0 + \beta_{(i1)} RD_{(i)} + \beta_{(i2)} LST_{(i)} - \beta_{(i3)} SD_{(i)} + \epsilon_{(i)} \quad (2)$$

Where β coefficients are county specific.

Instead of remaining the same everywhere, model coefficients now vary in terms of locations (i), which in this study represents individual counties. Individual parameters are estimated at each county in a GWR model.

A spatial kernel was used to provide geographic weighting for the local window centered on the grid nodes used in our model. There are two possible choices for the spatial kernel: fixed or adaptive, and bandwidth, which is a key coefficient that controls the size of the kernel (Akaike 1974). These kernel tend to be Gaussian or Gaussian-like which implies that distant samples are weighed lesser than the proximal ones. An optimal bandwidth can be found by minimizing the corrected Akaike information criterion

(AICc). There are three choices for the bandwidth method: AICc, cross validation (CV), and bandwidth parameter. Both AICc and CV are automated methods in ArcGIS for finding the bandwidth that gives the best predictions, whereas the 'bandwidth parameter' function allows you to specify a bandwidth. The AICc method finds the bandwidth that minimizes the AICc value, and it is computed from (a) a measure of the divergence between the observed and fitted values and (b) a measure of the complexity of the model that depends on the number of variables and the bandwidth in the model. This interaction between the bandwidth and the complexity of the model is the basis for our preference for the AICc method over the CV score. In the GWR model, the adaptive kernel with AICc estimated bandwidth was chosen as the distribution of infected dead birds was inhomogeneous in the study area. Initially we selected the following settings in our ArcGIS GWR: Bandwidth method = AICc and Kernel type = Adaptive. We also tested various settings by keeping fixed bandwidth parameters to each unit of analysis.

Finally, we also examined local collinearity, independency, normality of residuals, and local collinearity of our GWR model to evaluate the fit of the model. If a county's local collinearity, the square root of the largest eigenvalue divided by the smallest eigenvalue, was greater than 30, the county was removed from analysis before proceeding. The adjusted coefficient of determination (Adjusted R^2) was used for comparing OLS and GWR models. The Akaike Information Criterion (AIC) generated for OLS and corrected Akaike Information Criterion (AICc) calculated for GWR were also used to help us determine which approach would provide a better understanding of the

relationship between environmental conditions and West Nile Virus risk (Fotheringham et al. 2003).

3.7 Modeling results

By comparing the fit of the global OLS and local GWR models, we found that the global adjusted R^2 is 0.61 (R^2 is 0.65, $P < 0.5$, Figure 3-2) and the local adjusted R^2 is 0.71 (R^2 is 0.74, $p < 0.5$, Figure 3-3), which suggests that there has been some improvement in model performance. Our preferred measure of model fit, which is the AICc, gave a value of 567.70 for the global model and 551.40 for the local model. The difference of 16.30 is a strong evidence of an improvement in the model fit to the data.

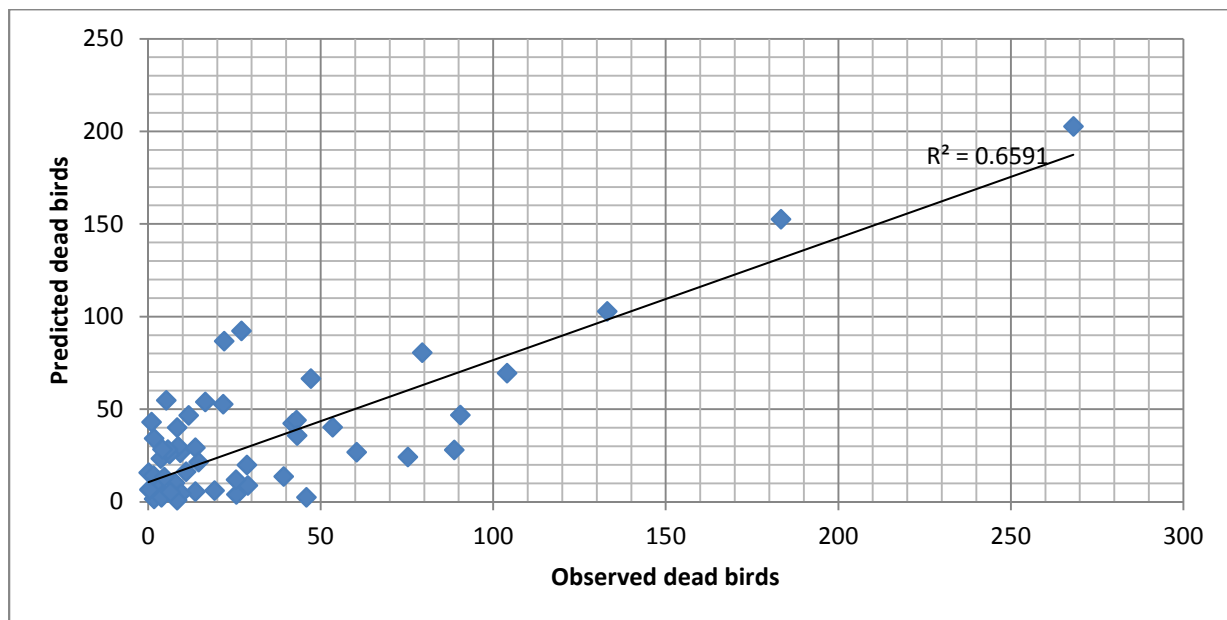


Figure 3-2 Trendline plot for global OLS model

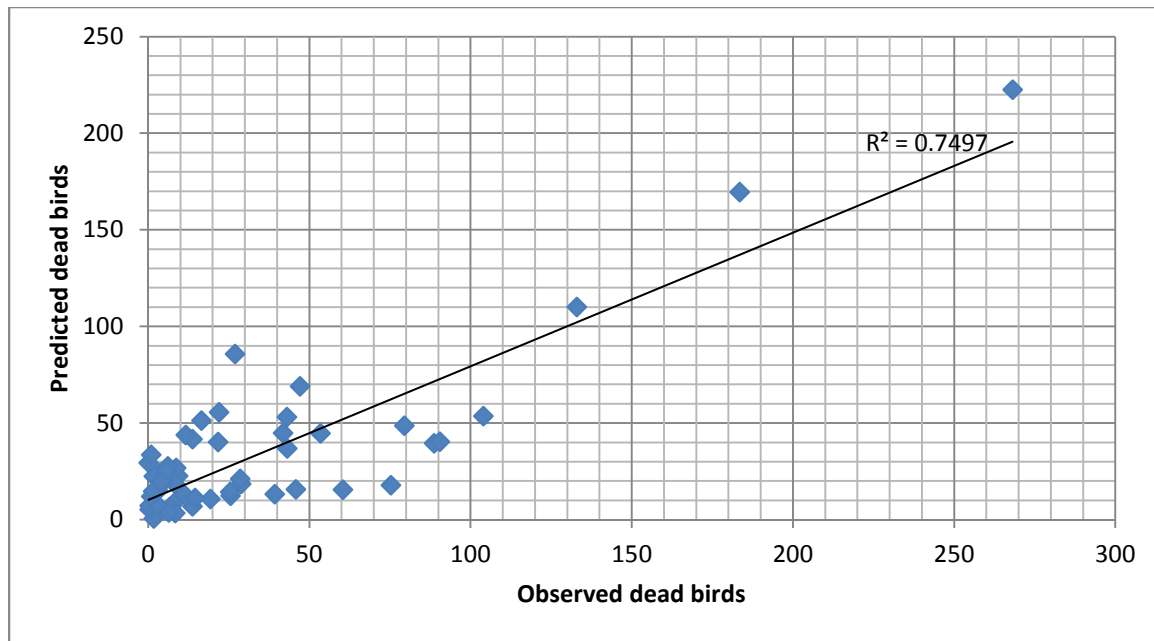


Figure 3-3 Trendline plot for local GWR model

In this study the number of nearest neighbors (NN) or bandwidth used in the estimation of each set of coefficients is 54. This bandwidth is large in comparison with the total number of observations in the dataset (58). It may be noted that although the bandwidth criteria is large but it is still Gaussian like distribution. During analysis we observed that although a smaller band-width criterion gave an improved combination of AICc and adjusted R^2 values, it also compromised the model diagnostics. Thus, it is better to have a larger band-width rather than violating model assumptions and to avoid the unstable prediction (Charlton & Fotheringham ; Nakaya 2014).

Mapping the values of the standardized residual (Figure 3-4) provided a representation of: (a) areas with unusually high or low residuals and (b) whether the residuals were spatially autocorrelated. Counties with excessively large positive residuals would under-predict the WNV infection risk, and counties with excessively

large negative residuals would over-predict the WNV infection risk. The spatial autocorrelation of GWR residuals for our model resulted in a Moran's I value of -0.11 ($p=0.18$), implying little evidence of any autocorrelation in them.

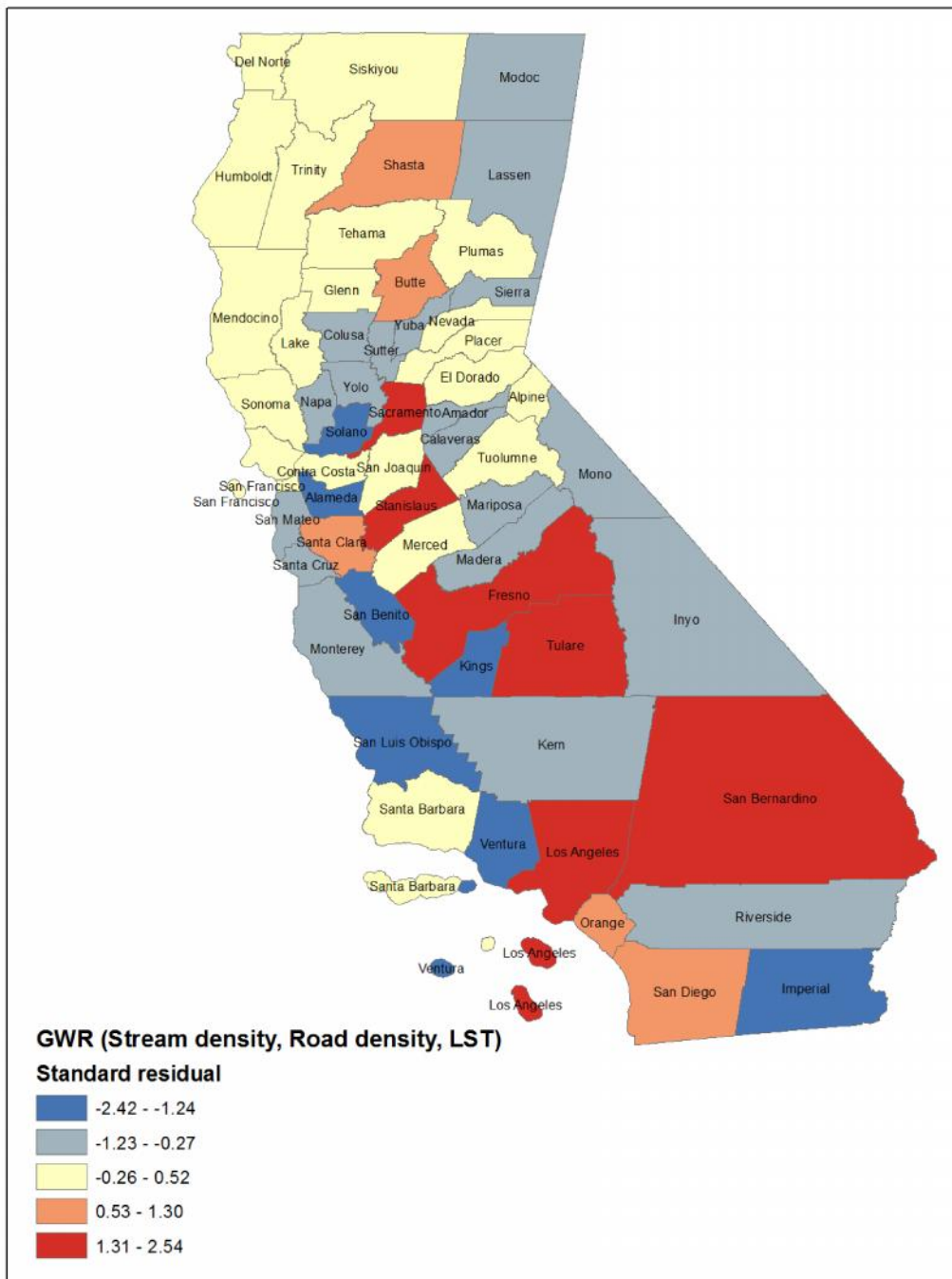


Figure 3-4 Map showing distribution of standardized residuals

Local coefficient estimates were also mapped. Figure 3-5 shows the variation in the coefficient estimates for the LST variable. The map for the local coefficients reveals that the influence of this variable in the model varies considerably over California, with a strong north-south direction. The range of the local coefficient is from 1.26 in the northernmost counties to 3.06 in the southernmost counties – evidence that points to heterogeneity in the model structure within California. The global coefficient and all the local coefficients for this variable are positive – there is agreement between the two models on the direction of the influence of this variable. Figure 3-6 shows a similar distribution in north-south direction of positive road density coefficient. Figure 3-7 reveals the opposite for stream density coefficients, with larger values in the north and smaller values in south.

Our best ordinary least squares model, the global OLS model (Equation 1) produced an adjusted R^2 of 0.61 ($p < 0.5$) with a corresponding corrected AIC of 567.70. Utilizing the same environmental variables, our best model using geographically weighted regression, the local GWR model, produced an adjusted R^2 of 0.71 ($p < 0.5$) with a corresponding corrected AIC of 551.4. This 16 points decrease in the AICc and approximately 16% improvement in the model performance suggest that incorporating spatial data improves the predictive ability of WNV risk in human population.

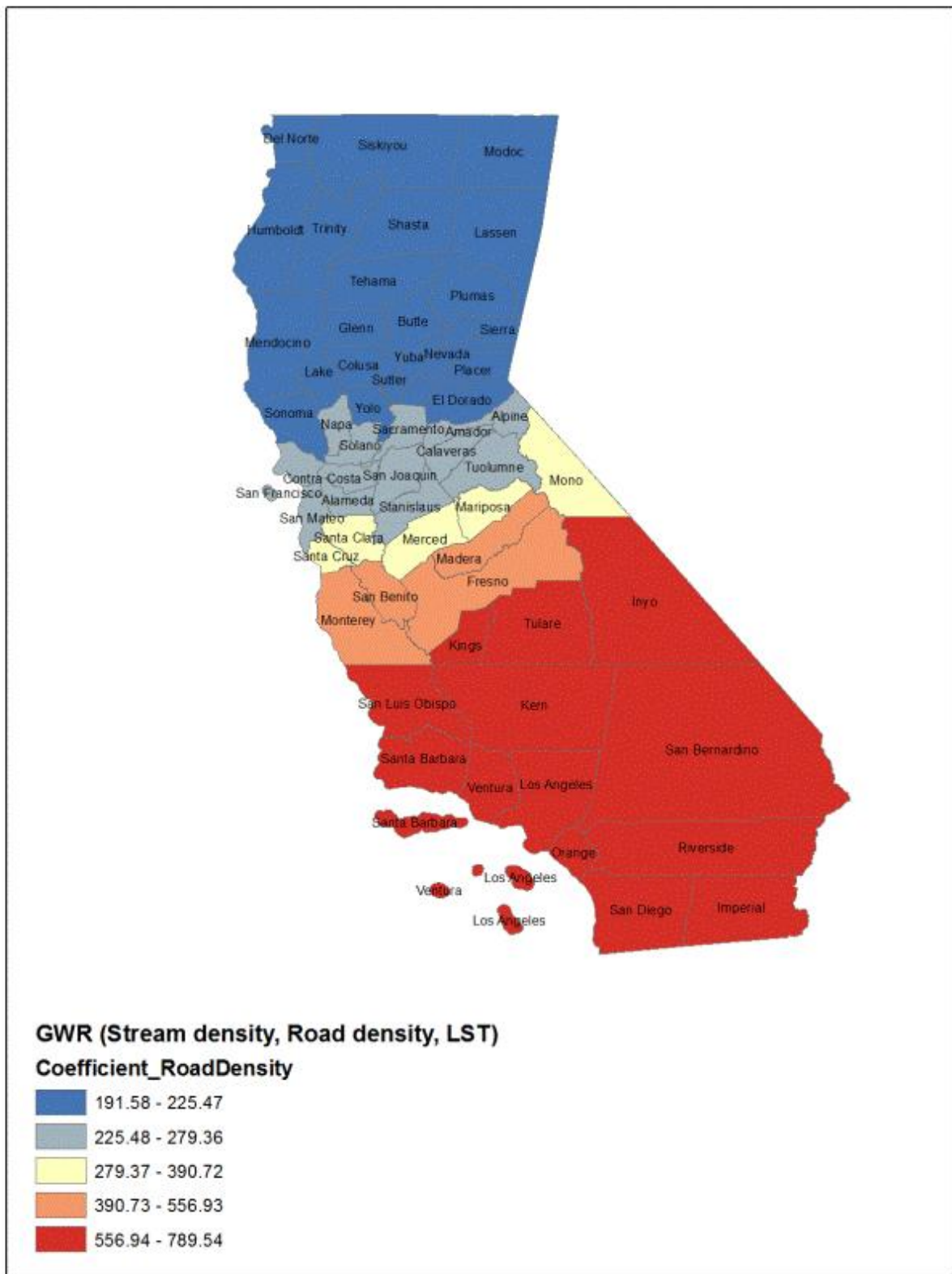


Figure 3-6 Maps showing distribution of road density coefficients

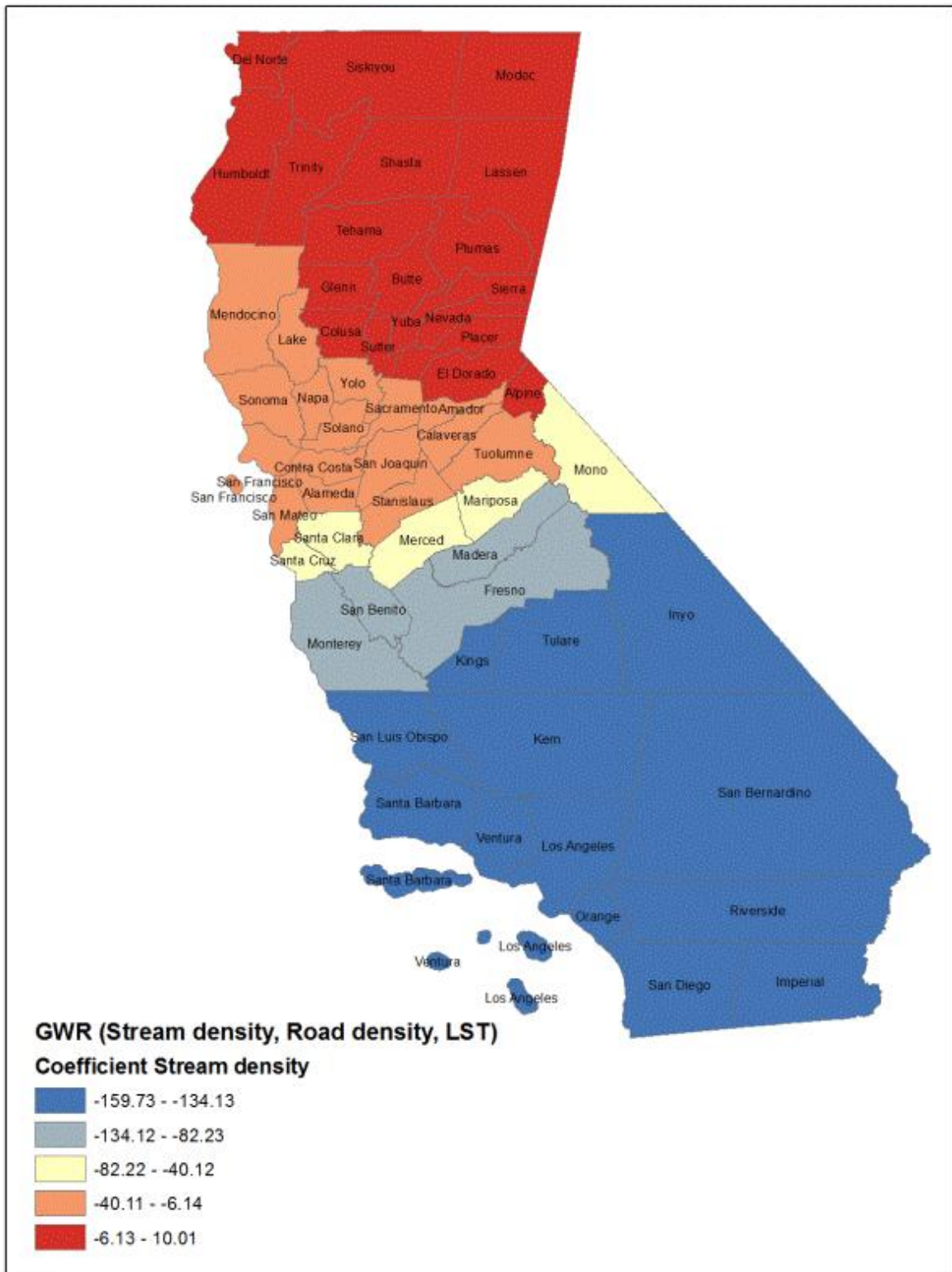


Figure 3-7 Maps showing distribution of stream density coefficients

3.8 Discussion

One of the frequent technical issues in modeling disease risk is to incorporate the local rather than global associations in these models (Foley et al. 2009). In spatial regression models, a global model can be used to examine the relationship between disease risk and potential explanatory factors, which are based on the assumption that the relationship is a stationary spatial process (Miller 2012). For a small and homogenous region of interest, it is reasonable to assume that the explanatory factors would not change significantly across the whole region, and the relationship between WNV incidence and the potential factors would also be unchanged. However, the topography, climate, and population distribution change greatly when it comes to a large region like California with a territory over 163,000 square miles. California is geographically diverse and is equally varied in its range of climates with several climatic sub-regions recognized. It would be unexpected to find that the spatial stationarity assumption holds in such large areas having a substantial range of climatic conditions. Our results concur that WNV epidemiological models with improved predictive ability for risk for exposure to vectors can be achieved through consideration of spatial heterogeneity (Beck et al. 1994). Besides improving prediction accuracy, spatial heterogeneity can also provide insights into the underlying ecological processes controlling the distributions of zoonotic pathogens (Wimberly et al. 2008a) and vector population because GWR models consider spatial heterogeneity by separating the large heterogeneous region into smaller, more homogeneous local regions. Fotheringham, 2009 (Fotheringham 2009a), stated that an advantage of using GWR is that it accounts for much of the spatial autocorrelation in the residuals that is usually found in global

modeling. It should also be realized that a variable that is insignificant at the global level might be important locally.

Environmental factors serve as correlates of the microhabitats that are the proximal influences on vector and host populations (Wimberly et al. 2008a). In this study, a GWR model was successfully used to explore the effects of local environmental factors on WNV incidence at the county level, which demonstrates that GWR models can be used to geographically differentiate the relationships between diseases and their explanatory factors. The local GWR models' statistical tests run on 58 counties show that factors like road density, stream density, and land surface temperature are significantly related to the WNV risk in all counties. We found that all three significant factors together could explain about 71% with bandwidth criterion of 54. The road density and land surface temperature exhibited a positive relation to the WNV infection risk in all counties. In contrast, stream density had a negative relationship in 76% of counties.

There are several limitations of this study. First, it is assumed that factors suitable for mosquito habitat increase the likelihood of WNV spread in human populations. On the surface this seems to be reasonably sound; however, we do not have specific evidence that this is true. Second, it is also assumed that the probability of human infection is higher in counties with multiple confirmed WNV bird cases. Again, a sound presumption with several references in the previous section, but we do not have specific evidence. A potential problem with this assumption is that human population density, variations in level of public concern, and resource availability might bias the reporting of dead birds. Thus, proper surveillance methods that take into consideration

these limitations while collecting infected dead bird data will contribute to more meaningful results. Third, the study considers that people are infected within the county of their residence, ignoring the possibility of contracting an infection while traveling outside the county limits. Unfortunately, we have no way of gathering any data to confirm or refute this assumption; we recognize that this can result in an overestimation of locally occurring WNV infection rate. Potential future research may include the travel habits of WNV-infected individuals, i.e. the distance from home and travel frequency. Such information can be used to filter model inputs to those human cases that occurred in close proximity to their homes or workplaces. Fourth, human WNV cases may be under reported because some people may not have access to healthcare services or the illness presents with mild symptoms (or they are asymptomatic) resulting in misdiagnosis. Additionally, mosquito control effort information was not integrated into our model even though active control measures would likely influence the model results. The unaccounted conditions such as population immunity to WNV risk, public health measures taken by local health departments, and personal and environmental hygiene, may also contribute to the occurrence, transmission, and spread of WNV among the community (Hu et al. 2012) that are not accounted for in our modeling efforts. These factors are potential covariates and can be included in future research. Fifth, while environmental data used in this research was of fine resolution, WNV disease human incidence data and infected dead bird data that is used is available only at coarse county level. This presented a spatial scale problem that could have been avoided by using fine resolution or location-specific data; however, due to patient confidentiality and data reporting issues, these data were unavailable. We had to assume that aggregating

the environmental data up to the county adequately represented the environmental conditions presented in the county, but we knew that data aggregation and smoothing processes were likely to introduce some uncertainty into the model.

3.9 Conclusion

The potential for amplification of mosquito vectors and for humans to become infected may differ greatly from one geographic area to the next depending on the environmental conditions of the area. Thus, the intrinsic potential for a human outbreak of WNV may be very different in areas having similar population density. These results demonstrate that mosquito habitat modeling provides a valuable public health tool for assessing the risk of human arboviral infections and justifying the allocation of limited public health resources.

The model described in this paper is a spatially explicit model that used GWR to adjust for spatial autocorrelation and non-stationarity and produced continuous estimates of mosquito habitat suitability. The model output may be useful for predicting WNV risk in human populations, particularly in those areas where outcome data are unavailable. GWR is a good example of a spatial statistical method that uses both the locational and the attribute information. It employs a spatial weighting function assuming that near places are more similar than distant ones, producing location-specific outputs that may be used to identify sites for further investigation (Fotheringham 2009a).

The results of our study could contribute to the understanding of spatial variability of disease risk burden at the local level. This spatially explicit modeling technique may

be useful in policy-making and decision-making depending on the granularity and resolution of available data. Identifying the spatial variations in relationships by estimating local regression parameters allows the spatial distribution and interaction of predictor variables to be explored. Analyzing local variations in relationships provides those concerned with public health policy with the ability to target resources and to achieve improved outcomes through location-specific activities (Comber et al. 2011). Predictive modelling of disease risk can be enhanced using spatially explicit methods that account for either spatial autocorrelation (the tendency for pathogen distributions to be clustered in space) or spatial heterogeneity (the potential for environmental influences on pathogens to vary predictably in space). Predictions based on spatial autocorrelation can be very effective when key environmental variables are unknown or unavailable as geospatial datasets. Spatial heterogeneity can improve predictions by capturing geographic shifts in the predominant ecological drivers (Wimberly et al. 2008a).

In summary, WNV, a globally emerging infectious disease, was found to be heterogeneously related to environmental factors at the county level throughout California during the time that our data were collected. Our findings may assist in the risk assessment for WNV transmission in local areas and guide local public health institutes to rationally allocate public health resources and improve their preparedness for an outbreak according to region-specific conditions.

CHAPTER 4

USING MULTIVARIATE GEOVISUALIZATION TO MODEL THE RELATIONSHIPS BETWEEN PLACE VULNERABILITY AND WEST NILE VIRUS RISK

4.1 Abstract

Understanding spatial variations in disease risk across different regions as well as population groups remains a difficult challenge. This study postulates that understanding the underlying place characteristics and population composition for the occurrence of West Nile Virus (WNV) infection is important for mitigating future outbreaks. While many spatial and aspatial models have attempted to predict the risk of WNV transmission, efforts to link these factors within a GIS framework are limited. One of the major challenges for such integration is the high dimensionality and large volumes typically associated with such models and data. This study uses a spatially explicit, multivariate geovisualization framework to integrate an environmental model of mosquito habitat with human risk factors derived from socio-economic and demographic variables. Our results show that such an integrated approach facilitates the exploratory analysis of complex data and supports reasoning about the underlying spatial processes that result in differential risks for WNV.

4.2 Introduction

Understanding spatial variations in disease risk across different regions as well as population groups remains a difficult challenge. Oppong & Harold (2009) argue that vulnerability to disease is inevitably tied to specific places whether it involves risk behaviors, population, or place characteristics. Further, Diez Roux & Mair (2010) argue that these factors operate at different spatial scales and attribute differential disease risks to individual- and group-level characteristics. Individual characteristics can be summarized as the biological attributes of individuals while group-level characteristics refer to the context of places to which those populations belong. We stratify group-level characteristics into contextual and compositional factors. Contextual factors describe the place characteristics, whereas compositional factors include population characteristics. Understanding these factors is useful for hypothesis-generation in public health research, policy-making, and understanding the underlying spatial processes that produce differential risks. In this study we provide a framework for integrating compositional and contextual factors for assessing West Nile Virus (WNV) risks.

WNV is an emerging vector-borne disease that was first detected in the United States in 1999, according to Nash et al. (2001). Cooke et al. (2006) focused on tracing the connections between WNV human infection risk and contextual settings of environmental conditions including the presence of streams, vegetation, and roads. Ruiz et al. (2004) demonstrated that composition factors such as age, income, and race/ethnicity of the human population can be important predictors of WNV infection risk in humans. While many spatial and non-spatial models have attempted to predict the risk of WNV transmission, efforts to link these factors within a GIS framework are

limited. Thus, the complex relationships existing between WNV human infection risk and contextual and compositional factors are poorly understood. In this paper, we use geovisualization techniques to unravel these relationships. A major challenge for such integration is high dimensionality of the data that makes it difficult to tease out patterns of interest.

The lack of an integrated approach in understanding the effect of these variables on WNV limits our ability to discover potential hidden patterns in a large number of possible combinations from such multivariate datasets. Guo et al. (2005) asserted that even in a selected subset of the data it is still a challenge to discover hidden relationships as potential patterns may take various forms - linear or non-linear, spatial or non-spatial. It is also important that analysis of such complex datasets is performed using methods that are computationally efficient. Tools developed for geovisualization can be used to support multivariate analysis of geospatial data. In this particular study contextual factors affecting WNV risk are modeled using geostatistical and spatial analysis techniques. The compositional factors affecting WNV risk are then combined with the contextual model output using self-organizing map (SOM) and parallel coordinate plot (PCP). When coupled with multivariate visualization, these techniques can reveal how interactions between the contextual and compositional factors vary locally across geographic space. The clustering and mapping of these factors is then analyzed to identify the places vulnerable to WNV risk. This study postulates that understanding the underlying place characteristics and population composition for the occurrence of WNV is important for mitigating future outbreaks.

4.3 Materials and methods

4.3.1 Contextual and compositional factors affecting WNV

Vectors and pathogen reservoirs overlap under favorable conditions of environmental factors (Rochlin et al. 2011). Contextual factors describe the place characteristics such as the mosquito species habitat, climatic conditions, topography, and land-use cover classes such as vegetation, water, and urbanized areas. Composition factors associated with risk of WNV infection among humans include population characteristics such as age-sex structure and susceptibility based on socio-economic status. This study builds on contextual and compositional factors that have been identified by previous research studies (Table 4-1).

4.3.2 Study area and data

The study area was the state of California, which is the third largest state in area in the United States and is made up of 58 counties and 8,057 census tracts. WNV was first detected in California in July 2003 (Reisen et al. 2004) and the state became the national epicenter of WNV activity in 2004 and 2005 (Jean et al. 2007). This study utilized coarse-scale data (county level) of WNV human incidence cases and infected dead bird counts. Fine scale environmental and demographic data (census tract level) was used to define contextual and compositional factors derived from Table 4-1. The study was conducted in two phases: (1) mosquito habitat model based on contextual factors and (2) geovisualization techniques based on compositional factors.

Table 4-1 Population and place characteristics associated with WNV risk.

Human population characteristics (composition)		Place characteristics (context: environmental factors related with vector habitat)	
Factors studied (reference)	Relation to WNV risk	Factors studied (reference)	Relation to WNV risk
Old age (Jean et al. 2007; Ruiz et al. 2004b)	Weakened immune system	Stream , Vegetation, Road (Cooke et al. 2006)	Sites for breeding and resting.
Male sex (Murray et al. 2006)	Social history or lifestyle.	Temperature (Wimberly et al. 2008b)	Increases growth rate of vector, decreases egg development cycle and shortens extrinsic incubation period of vector.
Race/Ethnicity (Ruiz et al. 2004b)	Increased risk from behaviors linked to their lifestyle.	Surface slope (Ozdenerol et al. 2008b)	Water stagnation creating mosquito breeding ground.
Income (Ruiz et al. 2004b)	Increased risk from behaviors linked to their lifestyle.	Cultivated land, Developed land (Kilpatrick 2011)	Preferred natural ground pools in cultivated land and warmer micro- climates in developed lands.

4.3.3 Mosquito habitat model

The spatial pattern of WNV risk among human populations was predicted using a mosquito habitat model. This model was developed using statistically significant contextual variables such as streams, vegetation, roads, and temperature. GIS, geostatistical, and spatial analysis techniques were utilized in this study to evaluate habitat suitability of WNV-carrying mosquitoes. WNV risk is modeled using a geographically weighted regression (GWR), and WNV risk map is created using a multi-criteria decision analysis (MCDA) approach. GWR helps to reveal patterns that are local, and MCDA creates a spatially explicit risk map. Figure 4-1 illustrates the model framework including the advantages of using this approach.

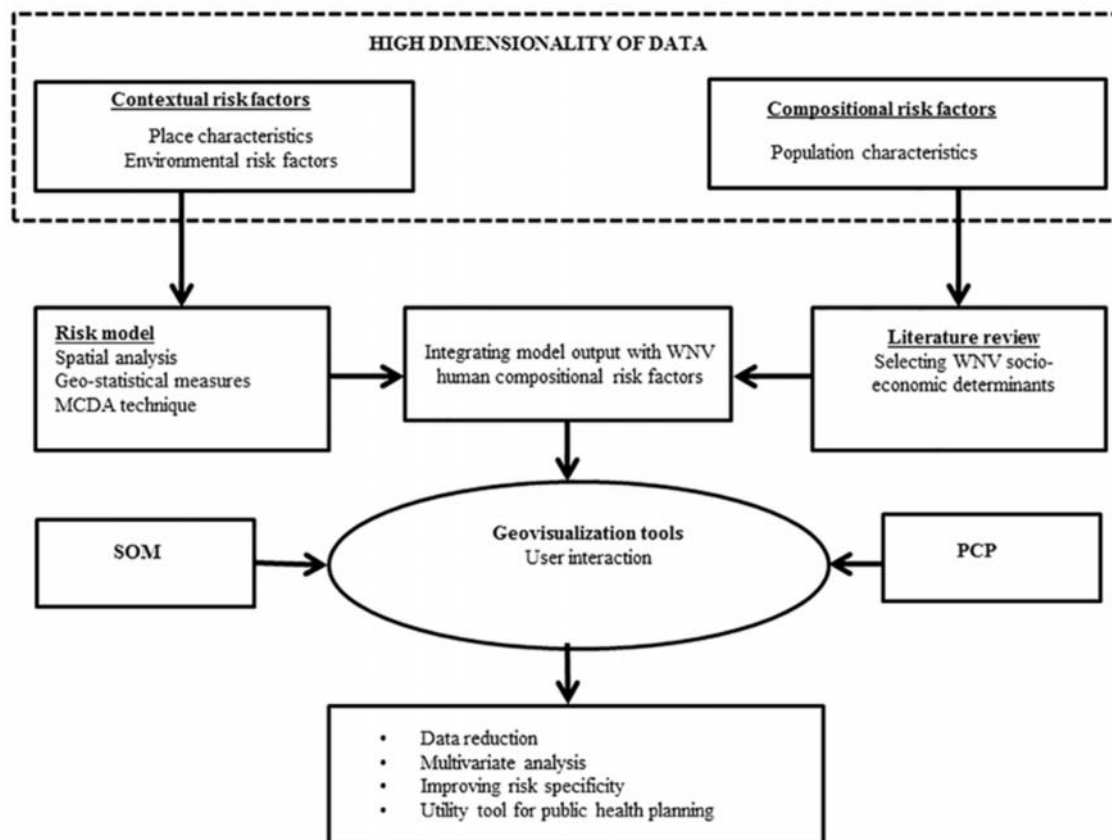


Figure 4-1 Geovisualization model framework

4.3.4 Geovisualization techniques

This study utilizes a spatially explicit exploratory approach for identifying the interaction between different contextual and compositional factors. Such an approach consists of a combination of risk maps and multivariate visualization techniques to effectively facilitate the exploration and understanding of complex patterns within data. The risk surface from mosquito habitat model was further explored for compositional variables using this approach. The analysis was done using SomVis, which is an integrated software tool consisting of three interactively linked visualizations that support human interactions: (1) Self-Organizing Map (SOM) (Kohonen 2001) to perform multivariate analysis, dimensional reduction, and data reduction; (2) Parallel Coordinate Plot (PCP) (Inselberg 2002) to visualize the multivariate patterns with display; and (3) Geographic map (GeoMap).

SOM is a unique partitioning clustering method, which segments multivariate data into non-overlapping clusters and projects them in a two-dimensional layout. Etien L Koua (2004) indicated that SOM is an unsupervised neural clustering technique that is useful in situations where the data volumes are large and the relationships are unclear. Each cluster is represented with a node (circle) whose size is linearly scaled according to the number of data items that it contains. Nodes are equally spaced in a two-dimensional space. Behind the nodes, there is a layer of hexagons, which are shaded to show the multivariate dissimilarity between neighboring nodes. Clusters falling on bright-tone hexagons are more similar to each other than those in darker tones of these hexagons.

PCP maps the n dimensional space onto a two-dimensional layout by using n equidistant parallel vertical axes. Each axis represents one variable and is linearly scaled using its minimum and maximum values. Each data element is displayed as a polyline intersecting each of the axes at the point that corresponds to the respective attribute value for this data element. The PCP can visualize the data either using combinations of variables (cluster level) or for each individual variable (data item level). At the cluster level, the PCP shows each cluster as a polyline, and it has the same color as the cluster in the SOM. The thickness of this polyline is proportional to the cluster size. At the data item level, each polyline in the PCP represents an individual data item with the same color as its containing cluster. The trajectory of the lines in the PCP serves as a legend for interpreting the meaning of different-colored clusters.

The Geomap is a sketch of multivariate information. The colors in a Geomap (Figure 4-7) represent the spatial distribution of multivariate patterns. The Geomap provides a spatial perspective to clusters of similar variables identified using PCPs. The three visual components allow an array of user interactions such as selection-based brushing and linking. We used the PCP in conjunction with a selection of clusters in the SOM to summarize the spatial context and enhance exploratory analysis of WNV risk.

Our contextual mosquito habitat model in conjunction with the compositional geovisualization approach represents a novel multivariate analysis technique to create a composite disease risk map that addresses the spatial heterogeneity of epidemiologic processes.

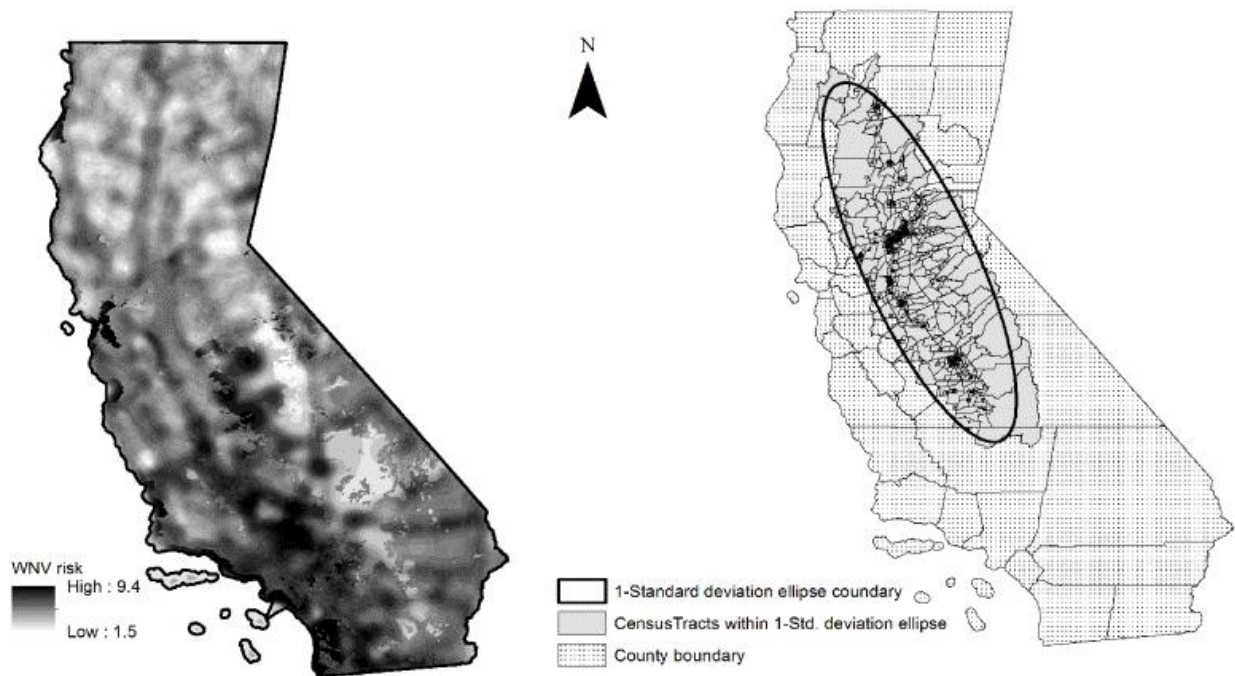


Figure 4-2 (left) WNV infection risk map

Figure 4-3 (right) Census tracts enclosed within standard deviation ellipse

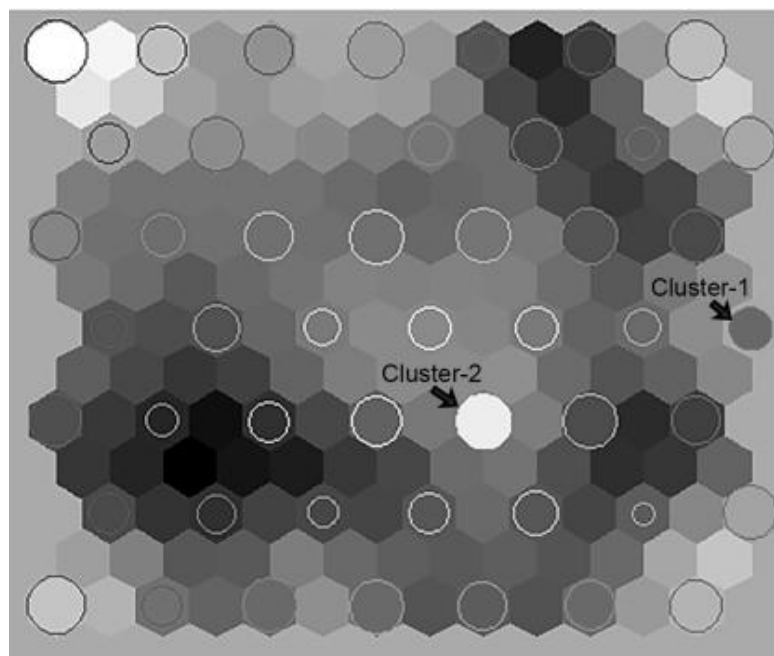


Figure 4-4 SOM clusters representing a valid combination of variables

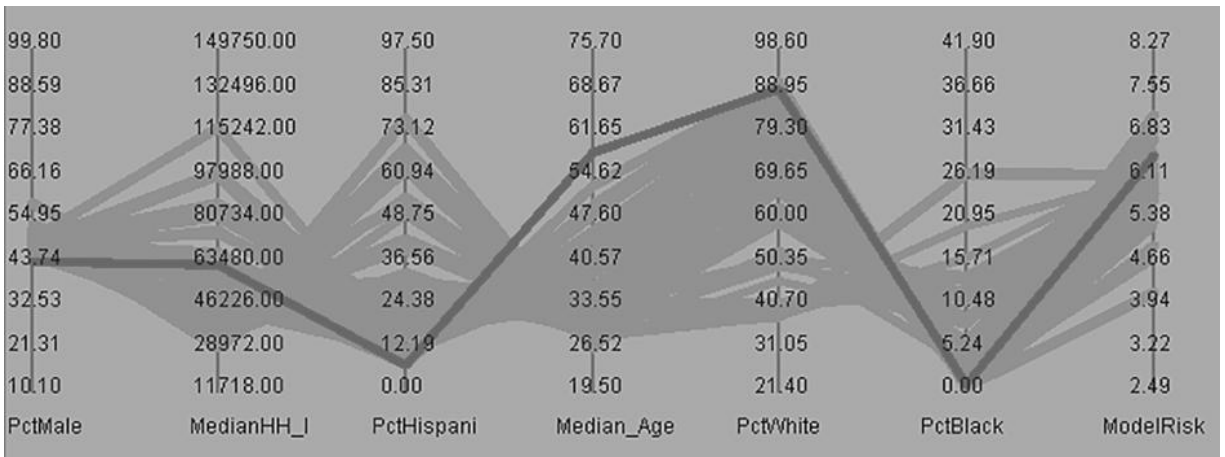


Figure 4-5 PCP cluster showing population group of highest median age

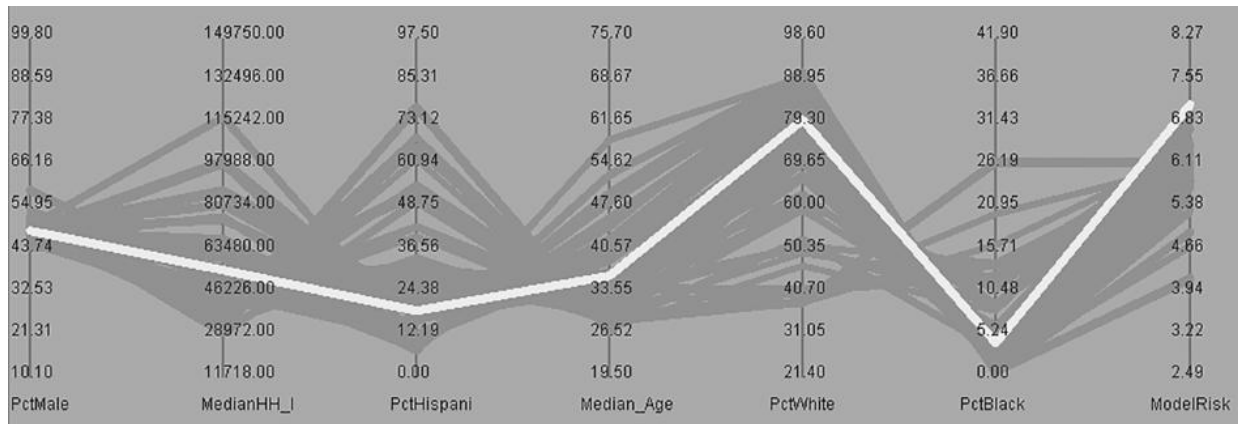


Figure 4-6 PCP cluster showing the highest mosquito habitat model risk

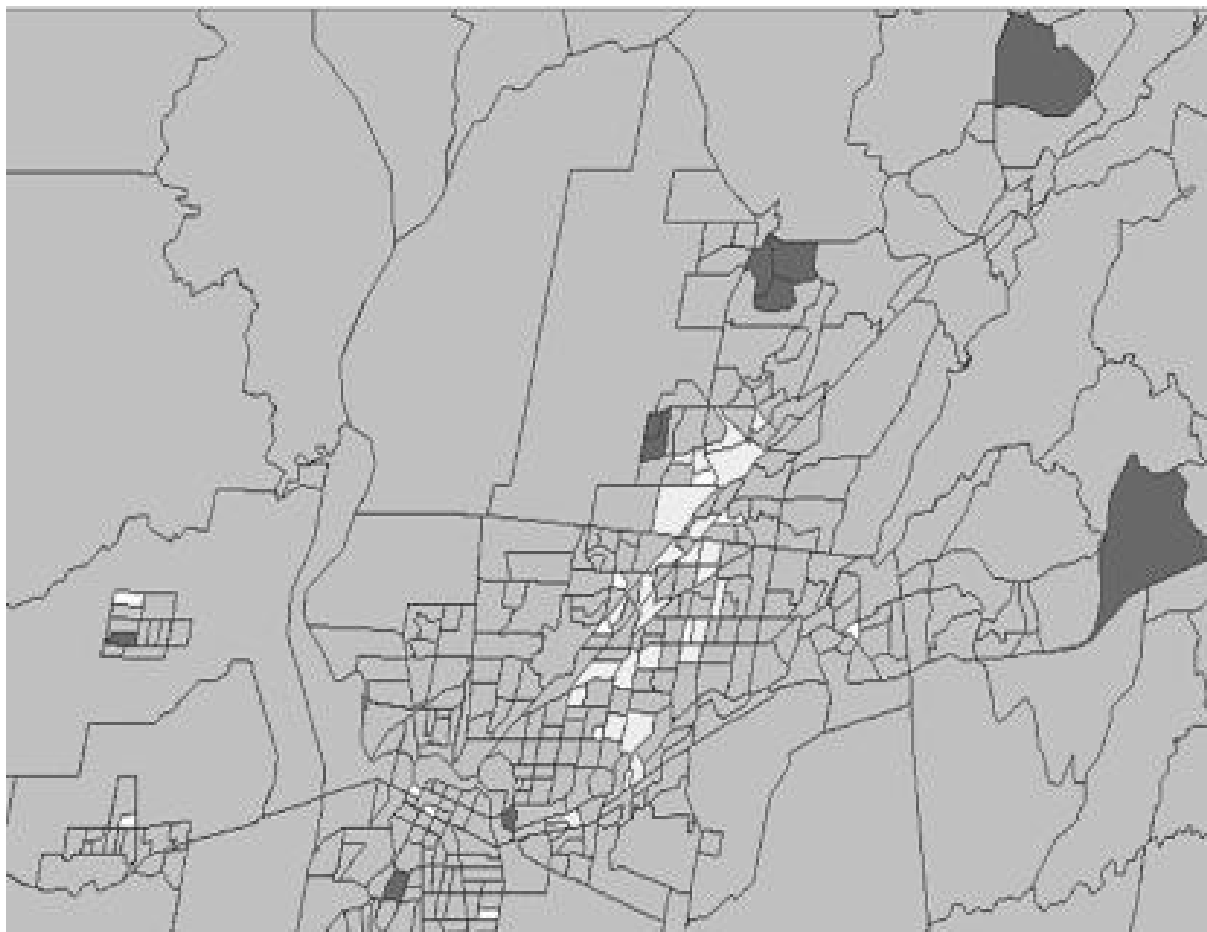


Figure 4-7 Geomap showing the spread of census tracts associated with SOM clusters

4.4 Results and discussion

The best-fitting mosquito habitat model for the whole state had an adjusted R^2 of 0.709 ($p < 0.05$). Figure 4-2 shows WNV infection risk map based on the mosquito habitat model. In this study, we concentrated on a smaller subset of California counties that contained the highest human WNV incidence rate (mosquito habitat model adjusted $R^2 = 0.82$, $p < 0.5$). This area was identified using a 1-standard deviation ellipse constructed with human WNV rates. Compositional variables were extracted for all census tracts within this area (Figure 4-3). SOM and PCP were used for identifying the

interaction between different contextual and compositional risk factors. As seen in Figures 4-5 and 4-6, all compositional factors included in Table 4-1 are represented using the first 6 vertical axes. The contextual model is represented using the last vertical axis in the two figures (i.e. ModelRisk). Each SOM cluster shown in Figure 4-4 represents a valid combination of variables that warrant further investigation.

In Figure 4-4, we focus on the node (labelled as cluster 1) that corresponds to the census tracts with the highest median age; this is the population group that has been reported as being most vulnerable to WNV risk (Campbell et al. 2002). This cluster is represented by the highlighted line in Figure 4-5. The trajectory of this line represents the relationship between attribute values of factors contained in this cluster. This line suggests that high median age (~58 years) is associated with moderately high mosquito habitat model risk (~6.4). This relationship has been suggested by other studies including the one conducted by Hayes et al. (2005a). It may be noted this cluster is also associated with the population group that belongs to low median household income. Further, in Figure 4-7, the non-contiguous spread of the census tracts associated with this cluster (n=19) indicates that they are only similar based on their attribute characteristics rather than geographical space.

We then selected the cluster with the highest mosquito habitat model risk (labelled as cluster-2 in Figure 4-4). In comparison to the previous analysis, we found that the highest levels of model risk were associated with low median household income and substantially lower median age. Also, census tracts associated with this cluster tend to be concentrated in geographic space (Figure 4-7). Study demonstrated that the integration of factors provides alternate views on the relationship between WNV and its

risk factors (median age) and suggests that further investigation is required to fully understand the relationship between age and WNV risk. Carson et al. (2012) shows that WNV infection was greatest for the younger population. The approach proposed in this study provides an exploratory framework for comparing relationships between a variety of compositional and contextual variables. However, this approach is not intended to derive causal relationships. Another way to look at this analysis from a public health perspective is that this approach provides different options to choose from when resources are limited and helps to design better intervention strategies.

The major challenges for multivariate mapping include large data volumes, high dimensionality, and the perception of complex patterns (Guo 2009). This research utilizes a spatially explicit exploratory approach that combines geovisualization, spatial analysis, and computational methods for identifying the interaction between different contextual and compositional factors. Our results show that the visualization of similarity clustering of multivariate attributes facilitates the analysis of complex data. It also supports reasoning about the underlying spatial processes that result in differential risks. Another advantage of this approach is that patterns found in voluminous and complex epidemiological data can provide more focused opportunities for analysis and interpretation by experts in that field. With an interactive user platform, geovisualization techniques can efficiently obtain new knowledge from the data and become an important hypothesis-generating tool in public health research. Understanding the underlying place characteristics and population composition for the occurrence of WNV is important for mitigating future outbreaks.

4.5 Conclusions

The major challenges for multivariate mapping include large data volume, high dimensionality, the perception of complex patterns, and the availability of data at multiple spatial resolutions. Geovisualization techniques assist in interpreting the results in such a multivariate context and are useful for identifying multivariate spatial and non-spatial patterns, thus providing useful insights into the causes of underlying spatial processes. We have demonstrated that the integration of both contextual and compositional factors using geovisualization techniques provides alternate views on the relationship between WNV and its associated risk factors. Such exploratory analysis supports hypothesis-generation in public health research and policy-making.

CHAPTER 5

CONCLUSION

This research postulates that understanding the conditions underlying the occurrence of WNV is important for mitigating future outbreaks. Environmental factors serve as correlates of the microhabitats that are the proximal influences on vector and host populations. The potential for amplification of mosquito vectors and for humans to become infected may differ greatly from one geographic area to the next depending on the environmental conditions of the area. Thus, the intrinsic potential for a human outbreak of WNV may be very different in areas with similar population density.

One of the frequent technical issues in modeling disease risk is to incorporate the local rather than global associations in these models. In spatial regression models, a global model can be used to examine the relationship between disease risk and potential explanatory factors, which are based on the assumption that the relationship is a stationary spatial process. For a small and homogenous region of interest, it is reasonable to assume that the explanatory factors would not change significantly across the whole region, and the relationship between WNV incidence and the potential factors would also be unchanged. However, the topography, climate, and population distribution change greatly when it comes to a large region like California with a territory over 163,000 square miles. California is geographically diverse and is equally varied in its range of climates with several climatic sub regions recognized. It would be unexpected to find that the spatial stationarity assumption holds in such large areas having a substantial range of climatic conditions. Our results concur that WNV

epidemiological models with improved predictive ability for risk for exposure to vectors can be achieved through consideration of spatial heterogeneity. Besides improving prediction accuracy, spatial heterogeneity can also provide insights into the underlying ecological processes controlling the distributions of zoonotic pathogens and vector population.

Predictive modelling of disease risk can be enhanced using spatially explicit methods that account for either spatial autocorrelation (the tendency for pathogen distributions to be clustered in space) or spatial heterogeneity (the potential for environmental influences on pathogens to vary predictably in space). Predictions based on spatial autocorrelation can be very effective when key environmental variables are unknown or unavailable as geospatial datasets. Spatial heterogeneity can improve predictions by capturing geographic shifts in the predominant ecological drivers. The model described in this paper is a spatially explicit model that used GWR to adjust for spatial autocorrelation and non-stationarity and produced continuous estimates of mosquito habitat suitability. The model output may be useful for predicting WNV risk in human populations, particularly in those areas where outcome data are unavailable. GWR is a good example of a spatial statistical method that uses both the locational and the attribute information. It employs a spatial weighting function assuming that near places are more similar than distant ones, producing location-specific outputs that may be used to identify sites for further investigation.

This study also postulates that understanding the underlying place characteristics and population composition for the occurrence of WNV infection is important for mitigating future outbreaks. The efforts to link these factors within a GIS framework are

limited. One of the major challenges for such integration is the high dimensionality, large volumes, and the availability of data at multiple spatial resolutions typically associated with such models. This research utilizes a spatially explicit exploratory approach that combines geovisualization, spatial analysis, and computational methods to integrate an environmental model of mosquito habitat with human risk factors derived from socio-economic and demographic variables. Geovisualization techniques assist in interpreting the results in such a multivariate context and are useful for identifying multivariate spatial and non-spatial patterns, thus providing useful insights into the causes of underlying spatial processes. We have demonstrated that the integration of both contextual and compositional factors using geovisualization techniques provides alternate views on the relationship between WNV and its associated risk factors. Such exploratory analysis supports hypothesis-generation in public health research and policy-making.

There are several limitations of this study. First, it is assumed that factors suitable for mosquito habitat increases the likelihood of WNV spread in human populations. On the surface this seems to be reasonably sound; however, we do not have specific evidence that this is true. Second, it is also assumed that the probability of human infection is higher in counties with multiple confirmed WNV bird cases. Again, a sound presumption with several references in previous section, but we do not have specific evidence. A potential problem with this assumption is that human population density, variations in level of public concern, and resource availability might bias the reporting of dead birds. Thus, proper surveillance methods that take into consideration these limitations while collecting infected dead bird data will contribute to more

meaningful results. Third, the study considers that people are infected within the county of their residence, ignoring the possibility of contracting an infection while traveling outside the county limits. Unfortunately, we have no way of gathering any data to confirm or refute this assumption; we recognize that this can result in an overestimation of locally occurring WNV infection rate. Potential future research may include the travel habits of WNV-infected individuals, i.e. the distance from home and travel frequency. Such information can be used to filter model inputs to those human cases that occurred in close proximity to their homes or workplaces. Fourth, human WNV cases may be under reported because some people may not have access to healthcare services or the illness presents with mild symptoms (or they are asymptomatic) resulting in misdiagnosis. Additionally, mosquito control effort information was not integrated into our model even though active control measures would likely influence the model results. The unaccounted conditions such as population immunity to WNV risk, public health measures taken by local health departments, and personal and environmental hygiene may also contribute to the occurrence, transmission, and spread of WNV among the community that are not accounted for in our modeling efforts. These factors are potential covariates and can be included in future research. Fifth, while environmental data used in this research was of fine resolution, WNV disease human incidence data and infected dead bird data that is used is available only at coarse county level. This presented a spatial scale problem that could have been avoided by using fine resolution or location-specific data; however, due to patient confidentiality and data reporting issues, these data were unavailable. We had to assume that aggregating the environmental data up to the county adequately represented the environmental

conditions presented in the county, but we know that data aggregation and smoothing processes were likely to introduce some uncertainty into the model.

In this research, WNV was found to be heterogeneously related to environmental factors at the county level throughout California during the time that our data were collected. The results of our study could contribute to the understanding of spatial variability of disease risk burden at local level. This spatially explicit modeling technique may be useful in policy-making and decision-making depending on the granularity and resolution of available data. Identifying the spatial variations in relationships, by estimating local regression parameters, allows the spatial distribution and interaction of predictor variables to be explored.

Our results also show that an integrated approach with geovisualization framework facilitates the exploratory analysis of complex data and supports reasoning about the underlying spatial processes that result in differential risks for WNV. Understanding the underlying place characteristics and population composition for the occurrence of WNV is important for mitigating future outbreaks. Our results show that the visualization of similarity clustering of multivariate attributes facilitates the analysis of complex data. Another advantage of this approach is that patterns found in voluminous and complex epidemiological data can provide more focused opportunities for analysis and interpretation by experts in that field. With an interactive user platform, geovisualization techniques can efficiently obtain new knowledge from the data and become an important hypothesis-generating tool in public health research. The approach proposed in this study provides an exploratory framework for comparing

relationships between a variety of compositional and contextual variables. However, this approach is not intended to derive causal relationships.

These results demonstrate that mosquito habitat modeling provides a valuable public health tool for assessing the risk of human arboviral infections. Our integrated approach may assist in the risk assessment for WNV transmission in local areas and guide local public health institutes to rationally allocate public health resources and improve their preparedness for an outbreak according to region-specific conditions.

Our methods have made several improvements over the existing research on WNV risk modeling: 1) we developed a novel technique to model WNV human risk under data sparse conditions using publically available data; 2) we used MCDA approach to create spatially explicit risk map by assigning different ranks and weights to the risk factors; 3) we also incorporated spatial heterogeneity into environmental model to improve its performance by 16 percent; 4) we developed a geovisualization framework for integrating compositional and contextual factors that supports multivariate analysis and hypothesis generation; and 5) our methods improved the specificity of exposure prediction by improving the spatial resolution of the model results.

LIST OF REFERENCES

- Akaike H. 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19:716-723.
- Ann L, Vivian L, and Deron G. 2009. Risk factors for human infection with West Nile Virus in Connecticut: a multi-year analysis. *International Journal of Health Geographics*.
- Anselin L. 1995. Local indicators of spatial association—LISA. *Geographical analysis* 27:93-115.
- Atkinson SF, Sarkar S, Aviña A, Schuermann JA, and Williamson P. 2012. Modelling spatial concordance between Rocky Mountain spotted fever disease incidence and habitat probability of its vector *Dermacentor variabilis* (American dog tick). *Geospatial health* 7:91-100.
- Atkinson SF, Sarkar S, Aviña A, Schuermann JA, and Williamson P. 2014. A determination of the spatial concordance between Lyme disease incidence and habitat probability of its primary vector *Ixodes scapularis* (black-legged tick). *Geospatial health* 9:203-212.
- Auchincloss AH, Gebreab SY, Mair C, and Roux AVD. 2012. A review of spatial methods in epidemiology, 2000–2010. *Annual review of public health* 33:107.
- Barrett AD. 2014. Economic Burden of West Nile Virus in the United States. *The American journal of tropical medicine and hygiene* 90:389-390.
- Beck LR, Rodriguez MH, Dister SW, Rodriguez AD, Rejmankova E, Ulloa A, Meza RA, Roberts DR, Paris JF, and Spanner MA. 1994. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *The American journal of tropical medicine and hygiene* 51:271-280.
- Blair C. 2009. Vector biology and West Nile Virus. In: Diamond M, ed. *West Nile Encephalitis Virus Infection: Viral Pathogenesis and the host immune response*: Springer, 45-61.

- Bowden SE, Magori K, and Drake JM. 2011. Regional differences in the association between land cover and West Nile virus disease incidence in humans in the United States. *The American Journal of Tropical Medicine and Hygiene* 84:234.
- Bradley CA, Gibbs SE, and Altizer S. 2008. Urban land use predicts West Nile virus exposure in songbirds. *Ecological Applications* 18:1083-1092.
- Brown HE, Childs JE, Diuk-Wasser MA, and Fish D. 2008. Ecologic factors associated with West Nile virus transmission, northeastern United States. *Emerging infectious diseases* 14:1539.
- Brownstein JS, Rosen H, Purdy D, Miller JR, Merlino M, Mostashari F, and Fish D. 2002. Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector Borne and Zoonotic Diseases* 2:157-164.
- Campbell GL, Marfin AA, Lanciotti RS, and Gubler DJ. 2002. West nile virus. *The Lancet infectious diseases* 2:519-529.
- Carson PJ, Borchardt SM, Custer B, Prince HE, Dunn-Williams J, Winkelman V, Tobler L, Biggerstaff BJ, Lanciotti R, and Petersen LR. 2012. Neuroinvasive disease and West Nile virus infection, North Dakota, USA, 1999–2008. *Emerging infectious diseases* 18:684.
- CDC. Mosquito species in which West Nile virus has been detected, United States, 1999-2012 Available at <http://www.cdc.gov/westnile/resources/pdfs/Mosquito%20Species%201999-2012.pdf> (accessed 01/21 2014).
- CDC. Species of dead birds in which West Nile virus has been detected, United States, 1999-2012. Available at <http://www.cdc.gov/westnile/resources/pdfs/Bird%20Species%201999-2012.pdf> (accessed 01/15 2014).
- CDPH. West Nile Virus: Keep it on your Radar.
- Chaintoutis SC, Dovas CI, Papanastassopoulou M, Gewehr S, Danis K, Beck C, Lecollinet S, Antalis V, Kalaitzopoulou S, and Panagiotopoulos T. 2014. Evaluation of a West Nile virus surveillance and early warning system in Greece, based on domestic pigeons. *Comparative immunology, microbiology and infectious diseases* 37:131-141.

- Charlton M, and Fotheringham A. nd Geographically Weighted Regression: A Tutorial on Using GWR in ArcGIS 9.3. National Centre for Geocomputation. National University of Ireland.
- Charlton M, Fotheringham S, and Brunsdon C. 2009. Geographically weighted regression. *White paper National Centre for Geocomputation National University of Ireland Maynooth*.
- Clements AC, Pfeiffer DU, and Martin V. 2006. Application of knowledge-driven spatial modelling approaches and uncertainty management to a study of Rift Valley fever in Africa. *International Journal of Health Geographics* 5:57.
- Comber AJ, Brunsdon C, and Radburn R. 2011. A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International Journal of Health Geographics* 10:44-55.
- Cooke WH, Grala K, and Wallis RC. 2006. Avian GIS models signal human risk for West Nile virus in Mississippi. *International Journal of Health Geographics* 5:36.
- Cruz L, Cardenas VM, Abarca M, Rodriguez T, Reyna RF, Serpas MV, Fontaine RE, Beasley DW, DA ROSA APT, and Weaver SC. 2005. Serological evidence of West Nile virus activity in El Salvador. *The American Journal of Tropical Medicine and Hygiene* 72:612-615.
- Curtis A, Ye X, Heob E, Targhetta J, Salvato V, Reyna M, Bueno R, and Holmes L. 2014. A comparison of three approaches to identify West Nile Virus mosquito space-time hotspots in the Houston Vicinity for the period 2002–2011. *Applied Geography* 51:58-64.
- De Filette M, Ulbert S, Diamond M, and Sanders NN. 2012. Recent progress in West Nile virus diagnosis and vaccination. *Vet Res* 43:16.
- DeGroote J, Sugumaran R, Brend S, Tucker B, and Bartholomay L. 2008. Landscape, demographic, entomological, and climatic associations with human disease incidence of West Nile virus in the state of Iowa, USA. *International Journal of Health Geographics* 7:19.
- DeGroote JP, Sugumaran R, and Ecker M. 2014. Landscape, demographic and climatic associations with human West Nile virus occurrence regionally in 2012 in the United States of America. *Geospatial health* 9:153-168.

- Dewan AM, Corner R, Hashizume M, and Ongee ET. 2013. Typhoid fever and its association with environmental factors in the Dhaka Metropolitan Area of Bangladesh: a spatial and time-series approach. *PLoS neglected tropical diseases* 7:e1998.
- Diez Roux AV, and Mair C. 2010. Neighborhoods and health. *Annals of the New York Academy of Sciences* 1186:125-145.
- Eidson M, Komar N, Sorhage F, Nelson R, Talbot T, Mostashari F, McLean R, and Group WNVAMS. 2001a. Crow deaths as a sentinel surveillance system for West Nile virus in the northeastern United States, 1999. *Emerging infectious diseases* 7:615.
- Eidson M, Kramer L, Stone W, Hagiwara Y, Schmit K, and Team NYSWNVAS. 2001b. Dead bird surveillance as an early warning system for West Nile virus. *Emerging infectious diseases* 7:631.
- Eidson M, Miller J, Kramer L, Cherry B, Hagiwara Y, and Group WNVBMA. 2001c. Dead crow densities and human cases of West Nile virus, New York State, 2000. *Emerging infectious diseases* 7:662.
- Eisen L, Barker CM, Moore CG, Pape WJ, Winters AM, and Cheronis N. 2010. Irrigated agriculture is an important risk factor for West Nile virus disease in the hyperendemic Larimer-Boulder-Weld area of north central Colorado. *Journal of Medical Entomology* 47:939-951.
- Etien L Koua M-JK. 2004. Geovisualization to support the exploration of large health and demographic survey data. *Int J Health Geogr* 3:1-13.
- Ezenwa VO, Milheim LE, Coffey MF, Godsey MS, King RJ, and Guptill SC. 2007. Land cover variation and West Nile virus prevalence: patterns, processes, and implications for disease control. *Vector-Borne and Zoonotic Diseases* 7:173-180.
- Foley R, Charlton MC, and Fotheringham AS. 2009. GIS in Health and Social Care Planning.
- Fotheringham AS. 2009a. Geographically weighted regression. *The Sage handbook of spatial analysis*:243-254.

- Fotheringham AS. 2009b. "The problem of spatial autocorrelation" and local spatial statistics. *Geographical analysis* 41:398-403.
- Fotheringham AS, Brunsdon C, and Charlton M. 2002. *Geographically weighted regression*: Wiley New York.
- Fotheringham AS, Brunsdon C, and Charlton M. 2003. *Geographically weighted regression: the analysis of spatially varying relationships*: John Wiley & Sons.
- Getis A, and Ord JK. 1992. The analysis of spatial association by use of distance statistics. *Geographical analysis* 24:189-206.
- Gibbs SE, Wimberly MC, Madden M, Masour J, Yabsley MJ, and Stallknecht DE. 2006. Factors affecting the geographic distribution of West Nile virus in Georgia, USA: 2002-2004. *Vector-Borne & Zoonotic Diseases* 6:73-82.
- Goodchild MF. 1986. *Spatial autocorrelation*: Geo Books Norwich.
- Goovaerts P. 2008. Geostatistical analysis of health data: State-of-the-art and perspectives. *geoENV VI—Geostatistics for Environmental Applications*: Springer, 3-22.
- Gubler DJ. 2007. The continuing spread of West Nile virus in the western hemisphere. *Clinical Infectious Diseases* 45:1039-1046.
- Guo D. 2009. Multivariate Spatial Clustering and Geovisualization. *Geographic Data Mining and Knowledge Discovery*:325.
- Guo D, Gahegan M, MacEachren AM, and Zhou B. 2005. Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Science* 32:113-132.
- Guptill SC, Julian KG, Campbell GL, Price SD, and Marfin AA. 2003. Early-season avian deaths from West Nile virus as warnings of human infection. *Emerging infectious diseases* 9:483.
- Hanafi-Bojd A, Vatandoost H, Oshaghi M, Charrahy Z, Haghdoost A, Zamani G, Abedi F, Sedaghat M, Soltani M, and Shahi M. 2012. Spatial analysis and mapping of

- malaria risk in an endemic area, south of Iran: A GIS based decision making for planning of control. *Acta tropica* 122:132-137.
- Harrigan RJ, Thomassen HA, Buermann W, and Smith TB. 2014. A continental risk assessment of West Nile virus under climate change. *Global Change Biology*.
- Hayes EB, Komar N, Nasci RS, Montgomery SP, O'Leary DR, and Campbell GL. 2005a. Epidemiology and Transmission Dynamics of West Nile Virus. *Emerging infectious diseases* 11.
- Hayes EB, Komar N, Nasci RS, Montgomery SP, O'Leary DR, and Campbell GL. 2005b. Epidemiology and transmission dynamics of West Nile virus disease. *Emerging infectious diseases* 11:1167.
- Hongoh V, Hoen AG, Aenishaenslin C, Waaub J-P, Bélanger D, and Michel P. 2011. *Spatially explicit multi-criteria decision analysis for managing vector-borne diseases*: Groupe d'études et de recherche en analyse des décisions.
- Hu M, Li Z, Wang J, Jia L, Liao Y, Lai S, Guo Y, Zhao D, and Yang W. 2012. Determinants of the incidence of hand, foot and mouth disease in China using geographically weighted regression models. *PloS one* 7:e38978.
- Hubálek Z, and Halouzka J. 1999. West Nile fever--a reemerging mosquito-borne viral disease in Europe. *Emerging infectious diseases* 5:643.
- Hurvich CM, Simonoff JS, and Tsai C-L. 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60:271-293. 10.1111/1467-9868.00125
- Inselberg A. 2002. Visualization and data mining of high-dimensional data. *Chemometrics and intelligent laboratory systems* 60:147-159.
- Jarque CM, and Bera AK. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6:255-259.
- Jean CM, Honarmand S, Louie JK, and Glaser CA. 2007. Risk factors for West Nile virus neuroinvasive disease, California, 2005. *Emerging infectious diseases* 13:1918.

- Johnson GD, Eidson M, Schmit K, Ellis A, and Kulldorff M. 2006. Geographic prediction of human onset of West Nile virus using dead crow clusters: an evaluation of year 2002 data in New York State. *American journal of epidemiology* 163:171-180.
- Kilpatrick AM. 2011. Globalization, land use, and the invasion of West Nile virus. *Science* 334:323-327.
- Kilpatrick AM, Daszak P, Jones MJ, Marra PP, and Kramer LD. 2006. Host heterogeneity dominates West Nile virus transmission. *Proceedings of the Royal Society B: Biological Sciences* 273:2327-2333.
- King RJ, Campbell-Lendrum DH, and Davies CR. 2004. Predicting geographic variation in cutaneous leishmaniasis, Colombia. *Emerging infectious diseases* 10:598.
- Komar N, and Clark GG. 2006. West Nile virus activity in Latin America and the Caribbean. *Revista Panamericana de Salud Pública* 19:112-117.
- Kramer LD, and Ebel GD. 2003. Dynamics of flavivirus infection in mosquitoes. *Advances in virus research* 60:187-232.
- Krivoruchko K. 2011. *Spatial statistical data analysis for GIS users*: Esri Press.
- Kuehn BM. 2012. Record Heat May Have Contributed to a Banner Year for West Nile Virus. *JAMA: The Journal of the American Medical Association* 308:1846-1848.
- Kupfer JA, and Farris CA. 2007. Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. *Landscape Ecology* 22:837-852.
- LaBeaud AD, Gorman A-M, Koonce J, Kippes C, McLeod J, Lynch J, Gallagher T, King CH, and Mandalakas AM. 2008. Rapid GIS-based profiling of West Nile virus transmission: defining environmental factors associated with an urban-suburban outbreak in Northeast Ohio, USA. *Geospatial health* 2:215.
- Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659-1673.

- Liu H, and Weng Q. 2012a. Enhancing temporal resolution of satellite imagery for public health studies: a case study of West Nile Virus outbreak in Los Angeles in 2007. *Remote Sensing of Environment* 117:57-71.
- Liu H, and Weng Q. 2012b. Environmental factors and risk areas of West Nile Virus in southern California, 2007–2009. *Environmental Modeling & Assessment* 17:441-452.
- Liu JG, and Mason P. 2009. *Essential image processing and GIS for remote sensing*: Wiley. com.
- Liu Y, Jiang S, Wang R, Li X, Yuan Z, Wang L, and Xue F. 2011. Spatial epidemiology and spatial ecology study of worldwide drug-resistant tuberculosis. *Int J Health Geogr* 10:10.1186.
- Malczewski J. 2000. On the use of weighted linear combination method in GIS: common and best practice approaches. *Transactions in GIS* 4:5-22.
- Malkinson M, and Banet C. 2002. The role of birds in the ecology of West Nile virus in Europe and Africa. *Japanese Encephalitis and West Nile Viruses*: Springer, 309-322.
- McIntosh B, Dickinson D, and McGillivray G. 1969. Ecological studies on Sindbis and West Nile viruses in South Africa. V. The response of birds to inoculation of virus. *The South African journal of medical sciences* 34:77.
- Miller JA. 2012. Species distribution models Spatial autocorrelation and non-stationarity. *Progress in Physical Geography* 36:681-692.
- Mitchell A. 2005. *The ESRI Guide to GIS Analysis: Spatial Measurements and Statistics*. Vol 2. Redlands. CA: ESRI Press.
- Morales-Betoulle ME, Morales H, Blitvich BJ, Powers AM, Davis EA, Klein R, and Córdón-Rosales C. 2006. West Nile virus in horses, Guatemala. *Emerging infectious diseases* 12:1038.
- Mostashari F, Kulldorff M, Hartman JJ, Miller JR, and Kulasekera V. 2003. Dead bird clusters as an early warning system for West Nile virus activity. *Emerging infectious diseases* 9:641.

- Murray K, Baraniuk S, Resnick M, Arafat R, Kilborn C, Cain K, Shallenberger R, York T, Martinez D, and Hellums J. 2006. Risk factors for encephalitis and death from West Nile virus infection. *Epidemiology and infection* 134:1325-1332.
- Nakaya T. 2014. GWR4 user manual. *WWW document*, https://geodacenterasu.edu/drupal_files/gwr/GWR4manualpdf.
- Nash D, Mostashari F, Fine A, Miller J, O'Leary D, Murray K, Huang A, Rosenberg A, Greenberg A, and Sherman M. 2001. The outbreak of West Nile virus infection in the New York City area in 1999. *New England Journal of Medicine* 344:1807-1814.
- Nelms BM, Fechter-Leggett E, Carroll BD, Macedo P, Kluh S, and Reisen WK. 2013. Experimental and Natural Vertical Transmission of West Nile Virus by California Culex (Diptera: Culicidae) Mosquitoes. *Journal of Medical Entomology* 50:371-378.
- Nielsen CF, and Reisen WK. 2007. West Nile virus-infected dead corvids increase the risk of infection in Culex mosquitoes (Diptera: Culicidae) in domestic landscapes. *Journal of medical entomology* 44:1067-1073.
- Oppong JR, and Harold A. 2009. Disease, ecology, and environment. *A companion to health and medical geography*:81-95.
- Ozdenerol E, Bialkowska-Jelinska E, and Taff GN. 2008a. International Journal of Health Geographics. *International Journal of Health Geographics* 7:12.
- Ozdenerol E, Bialkowska-Jelinska E, and Taff GN. 2008b. Locating suitable habitats for West Nile Virus-infected mosquitoes through association of environmental characteristics with infected mosquito locations: a case study in Shelby County, Tennessee. *International Journal of Health Geographics* 7:12.
- Ozdenerol E, Bialkowska-Jelinska E, and Taff GN. 2008c. Locating Suitable Habitats for West Nile Virus-Infected Mosquitoes through Association of Environmental Characteristics with Infected Mosquito Locations: A Case Study in Shelby County, Tennessee. *International Journal of Health Geographics* 7:12.
- Patnaik JL, Juliusson L, and Vogt RL. 2007. Environmental predictors of human West Nile virus infections, Colorado. *Emerging infectious diseases* 13:1788.

- Quirin R, Salas M, Zientara S, Zeller H, Labie J, Murri S, Lefrançois T, Petitsclerc M, and Martinez D. 2004. West Nile virus, Guadeloupe. *Emerging infectious diseases* 10:706.
- Rakotomanana F, Randremanana R, Rabarijaona L, Duchemin J, Ratovonjato J, Arieu F, Rudant J, and Jeanne I. 2007. Determining areas that require indoor insecticide spraying using Multi Criteria Evaluation, a decision-support tool for malaria vector control programmes in the Central Highlands of Madagascar. *International Journal of Health Geographics* 6:2.
- Reisen W, Lothrop H, Chiles R, Madon M, Cossen C, Woods L, Husted S, Kramer V, and Edman J. 2004. West Nile virus in California. *Emerging infectious diseases* 10:1369.
- Reisen WK, Barker CM, Carney R, Lothrop HD, Wheeler SS, Wilson JL, Madon MB, Takahashi R, Carroll B, and Garcia S. 2006a. Role of corvids in epidemiology of West Nile virus in southern California. *Journal of Medical Entomology* 43:356-367.
- Reisen WK, Fang Y, and Martinez VM. 2006b. Effects of temperature on the transmission of West Nile virus by *Culex tarsalis* (Diptera: Culicidae). *Journal of medical entomology* 43:309-317.
- Rezaeian M, Dunn G, St Leger S, and Appleby L. 2007. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *Journal of epidemiology and community health* 61:98-102.
- Robinson T, Harris R, Hopkins J, and Williams B. 2002. An example of decision support for trypanosomiasis control using a geographical information system in eastern Zambia. *International Journal of Geographical Information Science* 16:345-360.
- Rochlin I, Turbow D, Gomez F, Ninivaggi DV, and Campbell SR. 2011. Predictive mapping of human risk for West Nile virus (WNV) based on environmental and socioeconomic factors. *PloS one* 6:e23280.
- Rodgers SE, and Mather TN. 2006. Evaluating satellite sensor-derived indices for Lyme disease risk prediction. *Journal of medical entomology* 43:337-343.
- Ruiz M, Wolf A, Tedesco C, Kitron U, Durr P, and Martin S. 2004a. West Nile virus risk in Illinois. GISVET'04: second international conference on the applications of GIS

and spatial analysis to veterinary science, University of Guelph, Ontario, Canada, 23rd-25th June: Veterinary Laboratories Agency. p 59-61.

Ruiz MO, Chaves LF, Hamer GL, Sun T, Brown WM, Walker ED, Haramis L, Goldberg TL, and Kitron UD. 2010. Local impact of temperature and precipitation on West Nile virus infection in *Culex* species mosquitoes in northeast Illinois, USA. *Parasit Vectors* 3:19.

Ruiz MO, Tedesco C, McTighe TJ, Austin C, and Kitron U. 2004b. Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. *International Journal of Health Geographics* 3:8.

Saaty TL. 1990. How to make a decision: the analytic hierarchy process. *European journal of operational research* 48:9-26.

Schurich JA, Kumar S, Eisen L, and Moore CG. 2014. Modeling *Culex tarsalis* Abundance on the Northern Colorado Front Range Using a Landscape-Level Approach. *Journal of the American Mosquito Control Association* 30:7-20.

Shaman J, Day JF, and Stieglitz M. 2005. Drought-induced amplification and epidemic transmission of West Nile virus in southern Florida. *J Med Entomol* 42:134-141.

Smithburn K, Hughes T, Burke A, and Paul J. 1940. A neurotropic virus isolated from the blood of a native of Uganda. *The American Journal of Tropical Medicine and Hygiene* 1:471-492.

Srivastava A, Nagpal B, Saxena R, and Subbarao S. 2001. Predictive habitat modelling for forest malaria vector species *An. dirus* in India-A GIS-based approach. *Current Science-BanGALORE* 80:1129-1134.

Steele K, Carmel Y, Cross J, and Wilcox C. 2009. Uses and misuses of multicriteria decision analysis (MCDA) in environmental decision making. *Risk analysis* 29:26-33.

Steele K, Linn M, Schoepp R, Komar N, Geisbert T, Manduca R, Calle P, Raphael B, Clippinger T, and Larsen T. 2000. Pathology of fatal West Nile virus infections in native and exotic birds during the 1999 outbreak in New York City, New York. *Veterinary Pathology Online* 37:208-224.

- Stevens KB, and Pfeiffer DU. 2011. Spatial modelling of disease using data-and knowledge-driven approaches. *Spatial and Spatio-temporal Epidemiology* 2:125-133.
- Tobler WR. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography*:234-240.
- Trawinski P, and Mackay D. 2008. Meteorologically conditioned time-series predictions of West Nile virus vector mosquitoes. *Vector-Borne and Zoonotic Diseases* 8:505-522.
- Valiakos G, Papaspyropoulos K, Giannakopoulos A, Birtsas P, Tsiodras S, Hutchings MR, Spyrou V, Pervanidou D, Athanasiou LV, and Papadopoulos N. 2014a. Use of Wild Bird Surveillance, Human Case Data and GIS Spatial Analysis for Predicting Spatial Distributions of West Nile Virus in Greece. *PloS one* 9:e96935.
- Valiakos G, Papaspyropoulos K, Giannakopoulos A, Birtsas P, Tsiodras S, Hutchings MR, Spyrou V, Pervanidou D, Athanasiou LV, Papadopoulos N, Tsokana C, Baka A, Manolakou K, Chatzopoulos D, Artois M, Yon L, Hannant D, Petrovska L, Hadjichristodoulou C, and Billinis C. 2014b. Use of wild bird surveillance, human case data and GIS spatial analysis for predicting spatial distributions of West Nile virus in Greece. *PloS one* 9:e96935. doi:96910.91371/journal.pone.0096935.
- Ward MP, Wittich CA, Fosgate G, and Srinivasan R. 2009. Environmental risk factors for equine West Nile virus disease cases in Texas. *Veterinary research communications* 33:461-471.
- Wimberly MC, Baer AD, and Yabsley MJ. 2008a. Enhanced spatial models for predicting the geographic distributions of tick-borne pathogens. *International Journal of Health Geographics* 7:15.
- Wimberly MC, Hildreth MB, Boyte SP, Lindquist E, and Kightlinger L. 2008b. Ecological niche of the 2003 west nile virus epidemic in the northern great plains of the United States. *PloS one* 3:e3744.