University of Northern Iowa

UNI ScholarWorks

2014

Landscape, demographic and climatic associations with human West Nile virus occurrence regionally in 2012 in the United States of America

John P. DeGroote
Univeristy of Northern Iowa

Ramanathan Sugumaran
University of Iowa

See next page for additional authors

Let us know how access to this document benefits you

Follow this and additional works at: https://scholarworks.uni.edu/geo_facpub

Authors

John P. DeGroote, Ramanathan Sugumaran, and Mark Ecker

Landscape, demographic and climatic associations with human West Nile virus occurrence regionally in 2012 in the United States of America

John P. DeGroote[1], Ramanathan Sugumaran[2], Mark Ecker[3]

[1]*GeoInformatics, Training, Research, Education and Extension Center, Department of Geography, University of Northern Iowa, Iowa, United States of America;* [2]*Department of Geography, University of Iowa, Iowa City, Iowa, United States of America;* [3]*Department of Mathematics, University of Northern Iowa, Iowa, United States of America*

Abstract. After several years of low West Nile virus (WNV) occurrence in the United States of America (USA), 2012 witnessed large outbreaks in several parts of the country. In order to understand the outbreak dynamics, spatial clustering and landscape, demographic and climatic associations with WNV occurrence were investigated at a regional level in the USA. Previous research has demonstrated that there are a handful of prominent WNV mosquito vectors with varying ecological requirements responsible for WNV transmission in the USA. Published range maps of these important vectors were georeferenced and used to define eight functional ecological regions in the coterminous USA. The number of human WNV cases and human populations by county were attained in order to calculate a WNV rate for each county in 2012. Additionally, a binary value (high/low) was calculated for each county based on whether the county WNV rate was above or below the rate for the region it fell in. Global Moran's *I* and Anselin Local Moran's *I* statistics of spatial association were used per region to examine and visualize clustering of the WNV rate and the high/low rating. Spatial data on landscape, demographic and climatic variables were compiled and derived from a variety of sources and then investigated in relation to human WNV using both Spearman rho correlation coefficients and Poisson regression models. Findings demonstrated significant spatial clustering of WNV and substantial inter-regional differences in relationships between WNV occurrence and landscape, demographic and climatically related variables. The regional associations were consistent with the ecologies of the dominant vectors for those regions. The large outbreak in the Southeast region was preceded by higher than normal winter and spring precipitation followed by dry and hot conditions in the summer.

Keywords: West Nile virus, climate, *culex*, geographical information systems, United States of America.

Introduction

Vector-borne diseases continue to emerge in new areas as serious public health concerns as has been witnessed by the recent outbreak of Chikungunya virus in the Caribbean region and the spread of West Nile virus (WNV) in North America over the last 15 years. WNV was introduced to the United States of America (USA) in the late 1990s and by 2004 had spread across most of the coterminous USA. WNV is an arbovirus maintained primarily in a bird-mosquito-bird transmission cycle with larger mammals, such as humans, being considered incidental, or dead-end hosts as they do not sufficiently amplify the virus for further transmission

(Gould and Fikrig, 2004). Although the virus has been detected in up to 65 different mosquito species (Peterson et al., 2013), a much smaller number of mosquito species are considered the primary vectors responsible for transmission to humans. The distributions of these vectors vary across the country due to biotic and abiotic requirements of the species. From 2001-2004, *Culex pipiens*, *Culex restuans*, *Culex salinarius*, *Culex quinquefasciatus* and *Culex tarsalis* were in the top three infected mosquitoes in at least one year in the USA (Hayes et al., 2005). Studies from Northeast and Great Lakes regions of the USA have often implicated *Cx. pipiens* and *Cx. restuans* as crucial vectors (Andreadis et al., 2004; Ruiz et al., 2010). In the southeastern portion of the USA, *Cx. quinquefasciatus* (Gibbs et al., 2006) and *Culex nigripalpus* (Ezenwa et al., 2006) are considered to be the primary WNV vectors. In states west of the Mississippi, *Cx. tarsalis* has been shown to be the prominent WNV vector (Reisen et al., 2004; Bell et al., 2006). In addition to *Cx. tarsalis*, *Cx. quinquefasciatus* plays an

Corresponding author:
John P. DeGroote
GeoTREE Center ITTC 240
University of Northern Iowa, Cedar Falls, IA 50614-0406, USA
Tel. +1 319 273-6158; Fax: +1 319 273-7103
E-mail: john.degroote@uni.edu

important role in WNV transmission in the southwest USA (Reisen et al., 2008). The risk of WNV transmission to humans generally is greatest from mid to late summer in much of the coterminous USA (Petersen et al., 2013).

Annual WNV occurrence has varied spatially since its introduction in 1999. After large outbreaks in 2002, 2003 and 2006, there were relatively few WNV cases from 2007 to 2011. In 2009, 2010 and 2011 there were 720, 1,021 and 712 cases annually in the USA, while in 2012 there were 5,674 cases with 2,734 of the more severe neuroinvasive variety (CDC, 2014). The 2012 outbreak emphasised the fact that there is still considerable uncertainty in regards to the spatial and temporal occurrence of WNV and in the identification of ecological factors controlling vector-borne disease incidence in general (Petersen and Fischer, 2012).

The influence of landscape, demographic and climate factors in relation to infectious disease in general, and WNV specifically, has been recognised (Epstein and Defilippo, 2001; Kuhn et al., 2005; DeGroote et al., 2012). In the USA, there have been numerous national (Landesman et al., 2007; Soverow et al., 2009), state/regional (Day and Shaman, 2008; DeGroote et al., 2008; Wimeberly et al., 2008; Liu et al., 2009), and local (Brown et al., 2008; Nielsen et al., 2008; Ghosh and Guha, 2010) studies that have investigated or discussed the relationships between WNV infections and at least one landscape, demographic or climatic-related variable. However, to date there have been a lack of studies comprehensively examining regional patterns of WNV occurrence in relation to a range of landscape, demographic and climatically related variables based on ranges of WNV mosquito vector species in the coterminous USA.

The relationships between human WNV occurrence and climatically related variables has varied based on the scale and geographical area of the study. In the USA, warmer temperatures, elevated humidity and heavy precipitation in the weeks preceding infection were associated with the rate of WNV infection (Soverow et al., 2009). It was found that human WNV occurrence was associated with above-average precipitation in the preceding year in the eastern USA and below-average precipitation in the western USA (Landesman et al., 2007). In Florida, Day and Shaman (2008) discussed the idea that a sequence of drought and wetting may provide optimal conditions for WNV outbreaks. In Iowa, inconsistent inter-annual climatic relationships were found in relation to WNV incidence (DeGroote et al., 2008). In the Great Plains region, it

was suggested that a May-July precipitation of approximately 200 mm reflects optimum conditions for WNV amplification (Wimberly et al., 2008). In Mississippi, an inverse relationship between county-level WNV risk and total annual rainfall during the previous year was demonstrated (Wang et al., 2010). In a long term study of WNV in South Africa, summer precipitation and change in summer precipitation from previous to current summer were strongly associated with WNV human outbreaks (Uejio et al., 2012). Around Davis, California, warm nightly temperatures and positive WNV *Cx. pipiens* mosquito pools indicated high risk for human WNV incidence (Nielsen et al., 2008). In the Chicago, Illinois area cumulative high temperature differences were a key factor for distinguishing years with higher mosquito infection and higher human illness rates (Ruiz et al., 2010).

Studies at state, regional or national level have demonstrated varying landscape and demographic associations with WNV incidence. For example, in Connecticut population density was positively associated with WNV occurrence (Liu et al., 2009), while in Iowa there was a negative association (DeGroote et al., 2008). Specific landscape conditions have been identified as being important in certain locations, including age of housing in Chicago (Ruiz et al., 2004) and orchards in the northwest (Crowder et al., 2013). Generally, the authors of many of these studies conclude that spatial associations between the occurrence of WNV and some landscape, demographic or climatic conditions logically matched the ecologies of the dominant vector(s) of that area.

There have been several studies that have examined WNV occurrence nationally, either by investigating data from all counties or certain counties from throughout the coterminous USA. Tackett et al. (2006) investigated association between WNV fatality rates and demographic and climatic variables from 13 states spread throughout the USA, finding that poverty rates and average temperature were related to probability of a fatal outcome. Gates and Boston (2009) found strong associations between irrigated land and WNV incidence using a sample of 92% of counties in the coterminous USA. These studies gave no consideration to how these associations might vary regionally. Landesman et al. (2007) demonstrated differing relationships in the eastern USA as compared to the western USA, using the Mississippi River to divide the country, between precipitation and human WNV incidence. While that study usefully identified broad regional differences, the authors pointed out it should

be expected that transmission dynamics will vary between finer geographical regions. Two more recent studies have utilised a regional framework to examine nationwide WNV incidence (Bowden et al., 2011; DeGroote and Sugumaran, 2012) in relation to landscape conditions with both studies showing regionally varying associations.

There have been few studies which have investigated climatic, landscape and demographic variables regionally based on potential vector distributions for the coterminous USA (DeGroote and Sugumaran, 2012) and no studies that have done so for the outbreak of 2012. This study attempts to provide insight into the drivers of large outbreaks of WNV incidence regionally in the USA in 2012 that came as a bit of a surprise after numerous years of low WNV occurrence. In addition, this study builds on the results of DeGroote and Sugumaran (2012) by including climatic variables in addition to the landscape and demographic variables and provides insight into the large outbreaks of 2012.

Materials and methods

Region definition

The methodology for defining functional ecological regions follows from that described in DeGroote and Sugumaran (2012). In that study, six regions were defined based on species distribution maps for *Cx. nigripalpus*, *Cx. quinquefasciatus*, *Cx. pipiens*, *Cx. restuans* and *Cx. tarsalis* (Darsie and Ward, 2005) and prominent geographical features. These species were selected as they were considered to be the most prominent vectors based on the literature. The Southeast region was approximately based on the overlap of *Cx. nigripalpus* and *Cx. quinquefasciatus* distributions as far north as the southern edge of the *Cx. pipiens* distribution. Originally, a large Northeast region was defined by the overlap in the distribution of *Cx. pipiens* and *Cx. restuans* east of the Mississippi River. The Great Plains region was based on the intersection of *Cx. pipiens*, *Cx. restuans* and *Cx. tarsalis* west of the Mississippi River and east of the Rocky Mountains. The Southwest region was roughly set as the intersection of *Cx. tarsalis* and *Cx. quinquefasciatus* ranges west of the *Cx. nigripalpus* range and south of the *Cx. pipiens* range. The Northwest region was characterised as the intersection of *Cx. pipiens* and *Cx. tarsalis* ranges west of the Rocky Mountains. Finally, the Rocky Mountain range was considered to be the intersection of the *Cx. pipiens* and *Cx. tarsalis*

range but within the Rocky Mountain region as defined by physiographic divisions of the coterminous USA. For the present study, the original Great Plains (Eastern Great Plains and Great Plains) and Northeast (North Central and Northeast) regions were split into two regions (Fig. 1). This was done as these regions were very large and encompassed fairly different land use patterns.

West Nile virus cases

Human WNV disease cases were attained from the Centers for Disease Control and Prevention (CDC) through the United States Geological Survey Disease Maps website (USGS, 2012) for the year 2012. This web mapping application displays the cumulative number of WNV cases annually by county in map and tabular form. A total number of cases of 4,986 were digitised from this website in the second week of December 2012. The annual total number of cases per county was joined to county boundary spatial data (United States Department of Interior, 2013). Annual population estimates for each county were attained from the US Census Bureau (2013). The WNV rate for 2012 for each county in the coterminous USA was calculated by dividing the number of WNV cases in 2012 by the population of the county. For each of the eight regions in the coterminous USA, a regional 2012 rate was calculated by dividing the total number of cases in 2012 in each region by the total population of that region. Subsequently, for each county, a binary value of zero (lower than regional 2012 rate) or one (higher than regional 2012 rate) was assigned after comparing the county WNV rate to the regional rate that the county fell in. The WNV rate for 2012 (cases in 2012/population of county), as well as the binary value of zero (lower than regional 2012 rate) and one (higher than regional 2012 rate) were used as dependent variables in statistical tests described in sections below. Similar methodologies using binary values have been used previously to examine high/low WNV occurrence statewide in Colorado (Winters et al., 2008) and nationally/regionally (Sugumaran et al., 2009; DeGroote and Sugumaran, 2012) in the coterminous USA.

Spatio-temporal variables

Climatic, landscape, demographic and land use metrics per county were processed from a variety of sources and used as independent variables as described in sections below. Land cover class percentages per
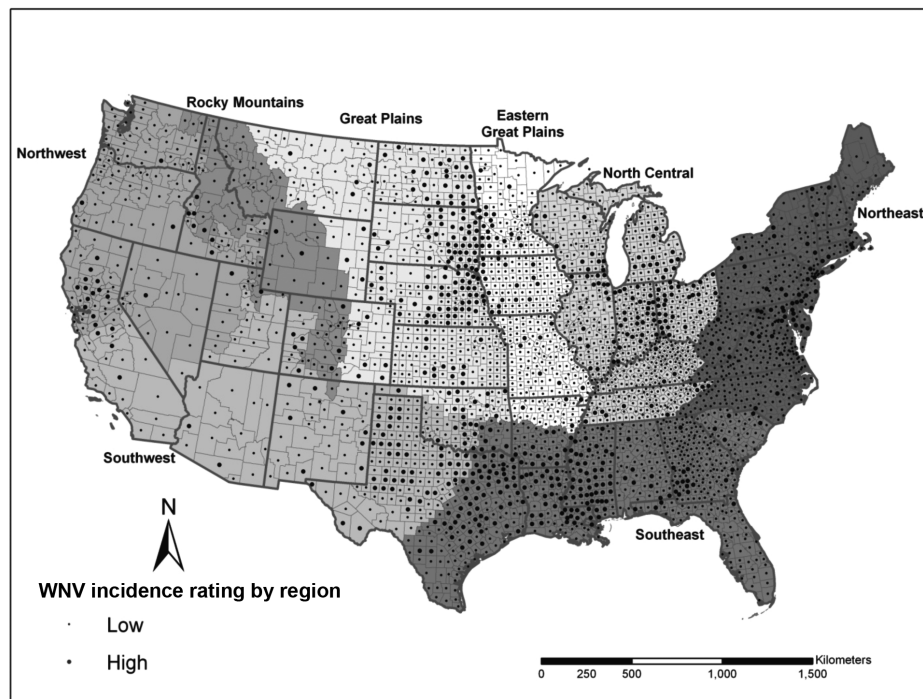
Fig. 1. West Nile virus regional rating (high/low) in the USA as calculated for eight regions defined primarily by vector ranges.

county were derived from the National Land Cover Database (Homer et al., 2007). An average hydric soils score (higher score represents a higher proportion of hydric or wet soils) as well as an average available water storage score for soils in each county was calculated based on the US General Soil Map downloaded from the Soil Data Mart (USDA, 2011). More detail on the derivation of the other demographic and land use variables is described in DeGroote and Sugumaran (2012). Raster datasets (~4 km spatial resolution) from 2011 and 2012 representing total monthly precipitation, and average monthly minimum and maximum temperature for the coterminous USA were downloaded from the Oregon State University PRISM Climate Group (Oregon State University PRISM Group, 2013). Precipitation and temperature anomalies were calculated based on their departure from long-term average raster datasets from 1981-2010 from PRISM. The PRISM Climate Group produces these data using a climate-elevation regression method based on data points from more than 10,000 stations (Daly et al., 2008). In addition to calculating monthly anomalies, aggregated time period anomalies were calculated from January to June, May to July, November 2011 to February 2012, and the entire year for 2011. Climatic parameters from 2011 were included as it has been demonstrated

that antecedent climatic conditions can contribute to vector population levels and WNV dynamics in the following year (Chase and Knight, 2003; Landesman et al., 2007; DeGroote et al., 2008). Climatic anomalies were calculated per county and per time period mentioned above and included in addition to the landscape, demographic and land use independent variables in statistical tests described below. This is the first time, as known to the authors, these types of data have been examined regionally for the entire coterminous USA in relation to WNV occurrence.

Statistical analysis

In order to test whether there was statistically significant spatial autocorrelation present within each region, a Global Moran's *I* statistic (Mitchell, 2005) was calculated for each of the eight regions based on the WNV rate for 2012. In order to visualise the pattern of clustering in WNV occurrence in 2012, two separate Anselin Local Moran's *I* tests (Anselin, 1995) were calculated. In the first instance, the WNV rate for each county was used in calculating the Anselin Local Moran's *I*. In the second Anselin Local Moran's *I* statistic calculation, the high/low binary WNV ranking per county in each of the eight regions were used as the variable of interest for the Anselin Local Moran's *I*.

For both the global and local spatial autocorrelation operations, a contiguous neighbouring county spatial relationship was used as the conceptualisation of spatial relationship. This means that each county was compared to all counties that shared at least a partial border. Due to the evidence of spatial clustering within regions and for inclusion in the Poisson regression, an autoregressive variable was calculated for each county within each region. The autoregressive variable was calculated using the distance-weighted average method described in the autocovariate models section in Dormann et al. (2007). A spatial weights matrix for each county was produced for each set of regional counties based on a contiguity edges conceptualisation of spatial relationships (Mitchell, 2005). Thus, the autoregressive term was calculated for each county based on the WNV rates of neighbouring counties and based on the spatial weights matrix. If a county's neighbours had higher WNV rates, than that county would have a higher autoregressive term value. The autoregressive variable was considered as an independent variable in the Poisson regressions.

Two main sets of statistical tests were used to examine landscape, demographic and climatic relations with WNV occurrence in 2012. Spearman's rho correlation coefficient tests were used to examine a range of landscape, demographic and climatic variables and to identify variables for further consideration in building Poisson multi-variable regression regional models. Spearman rho correlation coefficients were calculated to examine the association between the binary regional rankings (per county whether above or below respective regional 2012 WNV rate) and the landscape, land use, demographic and climatic related metrics for each county. The 2012 WNV rates per county were not normally distributed with many counties having values of zero, ruling out the use of Pearson correlation coefficients. The marginal correlations were explored to determine which variables most significantly impact the WNV rate in each region. Spearman's rho is a rank correlation statistic, meaning that it calculates the usual (Pearson) correlation after separately ranking each of the two variables from one to the total number of observations. It is employed when one is concerned about the lack of normality of two variables, or more specifically, when outliers are present in the dataset.

Prior to creating regional Poisson regression models, Pearson correlation coefficients were calculated between the independent landscape, demographic and climatic variables in order to identify and eliminate collinear variables. The Pearson correlation coeffi-

cients were calculated between each pair of independent (climate, landscape, demographic) variables and those independent variables with coefficients greater than 0.75 were removed from the list of variables being further analysed in the Poisson regression. Previous studies have used a similar methodology (Diuk-Wasser et al., 2006; DeGroote and Sugumaran, 2012) to eliminate potentially redundant independent variables. After eliminating potentially collinear variables, the five climatic and five landscape, land use, and demographic variables with the highest Spearman rho correlation coefficients were organised for each region. Next, the impact of the collection of variables identified by the Spearman's rho correlation coefficients, were investigated jointly using a Poisson regression. A Poisson regression is similar to a normal ordinary least squares (OLS) regression but is appropriate for count data, especially when normality is lacking or more specifically when outliers are present in the data (Agresti, 2007). Since the WNV rates are per capita counts, a Poisson distribution is the appropriate distribution for modelling these non-negative integers. Furthermore, a Poisson regression, which incorporates independent variables with a Poisson dependent variable, typically outperforms a normal OLS regression with count data (Agresti, 2007). All statistical analyses were carried out using SPSS, SAS or the *stats* module from *scipy* software.

Results

The total number of counties, WNV cases, the WNV rates as well as the populations for each region are shown in Table 1. The Southeast region had the highest number of cases while the Great Plains region had the highest WNV rate. The Northeast and Northwest regions had the lowest rates of WNV. Table 2 displays the results for each of the regional Global Moran's *I* statistic calculations based on the WNV rate. All regions demonstrated significant spatial autocorrelation except for the Rocky Mountain and Southwest regions. The results from the Anselin Local Moran's *I* calculations for the WNV rate carried out for the whole country is shown in Fig. 2. The high rates in the western Great Plains region dominated with one large cluster spilling into the Eastern Great Plains region. The results created from running the regional Anselin Local Moran's *I* calculations for the regional high/low WNV rating is displayed in Fig. 3. Distinct clusters of higher WNV ratings are visible in the Southeast region in eastern Texas, Louisiana and Mississippi, in the Great Plains and Eastern Great Plains in the eastern

Table 1. Regional descriptive statistics of county populations and WNV rates in the eight regions of USA.

Region	Number of counties	Number of cases	Region population	Regional WNV rate per 100,000
Eastern Great Plains (EGP)	326	126	15,921,356	0.79
Great Plains (GP)	409	636	11,793,290	5.39
North Central (NC)	652	742	57,291,793	1.30
Northeast (NE)	532	337	82,483,387	0.41
Northwest (NW)	159	116	25,537,343	0.45
Rocky Mountains (RM)	91	26	3,139,145	0.83
Southeast (SE)	669	1,888	67,517,403	2.80
Southwest (SW)	246	1,109	45,810,672	2.42

Dakotas, Nebraska and western Minnesota. Smaller clusters are visible in northern California in the Northwest region, on the borders of Indiana and Ohio and around Chicago in the North Central region, and in Maryland in the Northeast region.

Spearman rho correlation coefficients for the landscape, demographic and climatic variables in relation to high/low WNV ratings by region are presented in Tables 3-6. Table 3 demonstrates the correlation coefficients between high/low WNV occurrence and landscape, land use and demographic conditions. Only the 10 variables with the highest correlation coefficients and those that were significant ($P \leq 0.05$) for each region are presented. In the Eastern Great Plains, forested areas were negatively associated with higher WNV incidence while row crops and minimum elevation were positively associated. Row crops, hydric soils and wetlands were associated with higher WNV occurrence in the Great Plains while grasslands and small grains were negatively associated. In the North Central region, high WNV ratings were positively associated with urban areas (population density, residential, all developed land), higher median income and Hispanic population percentage, while there were negative associations with mixed forest and higher medi-

an ages. In the Northeast region, there were negative associations with forested areas and higher elevations, while there were positive associations with higher incomes, higher population density, and wetlands/water land cover classes. In the Northwest region, irrigation, agricultural, grasslands, Hispanic populations and orchards were positively associated with higher WNV while forests were negatively associated. In the Rocky Mountains regions only five variables were significant, with irrigated crops, dairy cows, range in elevation and Hispanic populations being positively associated, while white population proportions were negatively associated. In the Southeast region pasture, higher available soil water, and younger, poorer and higher black populations were positively associated with higher WNV occurrence while irrigated crops, orchards and higher male population proportion were negatively associated. In the Southwest region populated areas, higher proportion black populations, soils that can hold more water, as well as row crops and pasture were positively associated with higher WNV.

The strength and direction of associations between precipitation anomalies and high/low WNV ratings varied between regions. In the Great Plains and

Table 2. Results of Global Moran's statistic calculations for the WNV rates for each of the eight regions of USA.

Region	Number of counties	Regional WNV rate per 100,000	Moran's *I* index	z-score	P-value
Eastern Great Plains (EGP)	326	0.76	0.4324	13.64	<0.001
Great Plains (GP)	409	5.42	0.3193	11.31	<0.001
North Central (NC)	652	1.30	0.2230	9.73	<0.001
Northeast (NE)	532	0.40	0.0513	2.14	0.032
Northwest (NW)	159	0.45	0.2522	5.95	<0.001
Rocky Mountains (RM)	91	0.83	-0.0322	-0.40	0.686
Southeast (SE)	669	2.80	0.3646	15.83	<0.001
Southwest (SW)	246	2.42	0.0467	1.45	0.146

High z-scores accompanied by low P-values indicate that there is statistically significant spatial clustering of WNV rates.
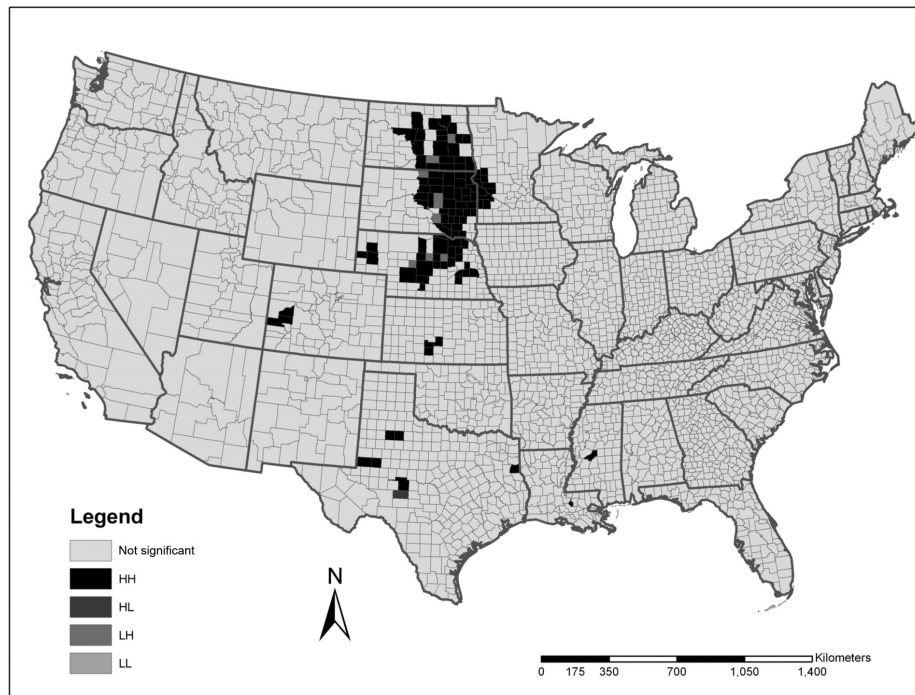
Fig. 2. Results of the Anselin Local Moran's *I* analysis carried out for the WNV incidence rate in 2012. HH indicates counties of higher WNV rates with neighbouring counties having higher WNV rates, HL indicates outlier counties with higher WNV rates, LH indicate outlier counties with lower WNV rates and LL indicates clusters of lower WNV rates.
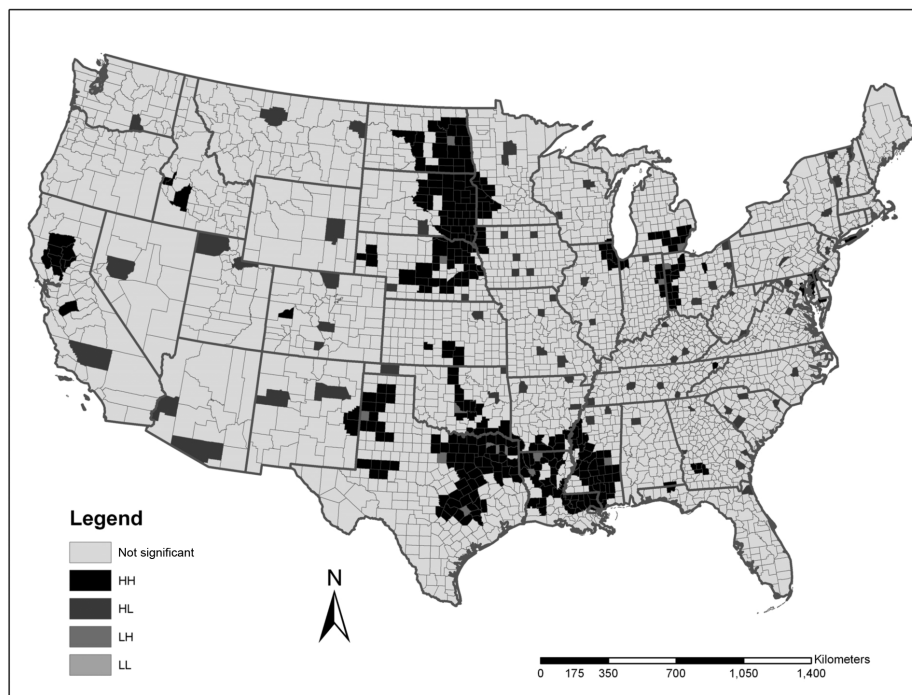

Fig. 3. Results of the Anselin Local Moran's *I* analysis carried out based on the regional high/low WNV incidence rating.

Table 3. Statistically significant* correlation coefficients resulting from comparison of landscape, land use and demographic variables to high/low WNV regional ratings.

Eastern Great Plains (EGP) variable	EGP Spearman coefficient	Great Plains (GP) variable	GP Spearman coefficient	North Central (NC) variable	NC Spearman coefficient
All forest	-0.31	Row crops	0.42	Population density	0.36
Deciduous forest	-0.30	Hydric soils	0.41	Low-intensity residential	0.36
Row crops	0.27	Wetlands dominated by herbaceous vegetation	0.38	Developed urban land cover	0.35
Mixed forest	-0.26	All wetlands	0.33	High-intensity residential	0.33
Minimum elevation	0.20	Grasslands	-0.32	Commercial/industrial area	0.29
Woody wetlands	-0.17	Shrub land	0.26	Mixed forest	-0.27
Number of cattle	-0.17	Small grains	-0.25	Median income	0.26
Wetlands dominated by herbaceous vegetation	0.16	Pasture	0.22	Urban grass (parks, golf courses, etc.)	0.25
Mean elevation	0.16	Low-intensity residential	0.21	Hispanic population (%)	0.24
Pasture	-0.15	Soil water storage capacity (top 25 cm)	0.20	Median age	-0.23

Northeast (NE) variable	NE Spearman coefficient	Northwest (NW) variable	NW Spearman coefficient	Rocky Mountains (RM) variable	RM Spearman coefficient
Minimum elevation	-0.19	Irrigation percentage	0.39	Range of elevations	0.33
Median income	0.19	Total irrigated area	0.39	White population (%)	-0.25
Evergreen forest	-0.18	Grasslands	0.36	Hispanic population (%)	0.22
All forest	-0.18	Orchards	0.35	Irrigated crops	0.24
Population density	0.16	Irrigated crops	0.32	Agricultural census of dairy cattle per county	0.26
Low-intensity residential area	0.15	Small grains	0.28		
Water	0.15	Row crops	0.26		
Mean elevation	-0.15	Evergreen forest	-0.26		
Developed urban land cover	0.14	Hispanic population (%)	0.25		
Wetlands dominated by herbaceous vegetation	0.14	All forest	-0.23		

Southeast (SE) variable	SE Spearman coefficient	Southwest (SW) variable	GP Spearman coefficient
Pasture	0.32	Low-intensity residential	0.34
Soil water storage capacity (top 25 cm)	0.30	Population density	0.32
Population <18 years old	0.16	Developed urban land cover	0.31
Irrigated crops	-0.15	Range elevation	-0.25
Male/female ratio	-0.15	Soil water storage capacity (top 25 cm)	0.25
Orchards	-0.14	Black population (%)	0.24
Number of cattle	0.12	Urban grass (parks, golf courses, etc.)	0.23
Population living in poverty (%)	0.12	Commercial/industrial area	0.23
White population (%)	-0.11	Row crops	0.22
Black population (%)	0.11	Pasture	0.22

*Spearman rho correlation (P <0.05); Only the top 10 variables are shown in the table for each region; In the Rocky Mountains region there were only five significant variables.

Table 4. Spearman rho correlation coefficients resulting from comparison of precipitation anomalies to high/low WNV regional rating.

Variable	EGP	GP	NC	NE	NW	RM	SE	SW
January	0.01	**0.21**	0.13	-0.02	-0.09	-0.01	**0.29**	0.13
January-June	**0.19**	**0.30**	0.00	**-0.21**	**-0.20**	-0.01	**0.13**	**0.17**
February	**0.23**	**0.24**	0.03	-0.05	**-0.36**	-0.01	**0.29**	0.06
March	**-0.21**	**-0.22**	0.06	**-0.21**	-0.09	-0.09	**0.41**	0.12
April	0.08	0.04	-0.02	-0.04	**0.19**	-0.07	**0.20**	-0.09
May	**0.18**	**0.33**	-0.09	-0.13	**-0.28**	0.01	**-0.35**	0.11
May-July	**0.19**	**0.19**	-0.14	-0.12	**-0.28**	-0.03	**-0.23**	-0.01
June	0.08	-0.03	0.00	**0.16**	**-0.20**	-0.06	**-0.18**	**0.19**
July	-0.09	**-0.18**	-0.14	-0.15	**-0.16**	-0.04	0.16	**-0.20**
Nov 2011-Feb 2012	0.02	**-0.11**	**0.11**	-0.01	**-0.28**	0.03	**0.34**	**0.14**
2011	-0.01	-0.05	**0.08**	**0.11**	**-0.15**	0.00	**-0.21**	**-0.11**

Coefficients with a P-value <0.05 in bold.

Eastern Great Plains regions, above average precipitation in May, May-July, February, and January-June were positively associated with higher WNV occurrence. In the North Central region, the strongest associations were with below average precipitation in July and May-July and above average precipitation in the preceding winter. In the Northeast region, there were generally negative associations with precipitation, indicating WNV was higher in areas with lower than average precipitation. This was reversed for the precipitation from the previous year and June of 2012. In the Northwest region all time periods, excluding April 2012, had negative associations indicating the cases of WNV occurred in drier areas. In the Southeast region, which had the largest number of WNV cases, early year above average precipitation was strongly positively associated with the higher WNV while drier than average conditions in May and June were associated with higher WNV. In the Rocky Mountains region, there were no strong associations, while in the Southwest region, higher than above precipitation in November 2011 through February 2012 and in the first half of 2012 were associated with WNV occurrence.

In the Eastern Great Plains and Great Plains region, warmer than average maximum temperatures in the winter months (November 2011-February 2012) and early months of 2012 (January through April) were positively associated with higher WNV occurrence while higher minimum temperature anomalies in summer months were positively associated. In the North Central region, higher than average maximum and minimum temperatures in most time periods were associated with higher WNV occurrence especially in preceding winter and late spring/early summer months. In the Northeast region, warmer winter maximum temperatures and late spring/early summer higher minimum temperatures were significantly associated with the higher WNV ratings. There were strong associations

Table 5. Spearman rho correlation coefficients resulting from comparison of maximum temperature anomalies to high/low WNV regional rating.

Variable	EGP	GP	NC	NE	NW	RM	SE	SW
January	**0.29**	**0.22**	-0.04	**0.12**	**0.29**	0.01	**0.27**	0.07
January-June	**0.19**	**0.20**	**0.14**	0.05	**0.21**	0.01	**0.15**	0.10
February	**0.15**	**0.22**	0.09	**0.16**	**0.38**	-0.08	**-0.28**	0.04
March	**0.18**	**0.35**	**0.18**	-0.07	0.00	0.06	-0.04	**0.14**
April	**0.21**	0.01	**-0.11**	0.09	-0.07	0.07	0.08	0.11
May	**-0.15**	**-0.13**	**0.18**	0.01	**0.23**	0.06	**0.26**	-0.01
May-July	**-0.16**	**-0.18**	**0.18**	0.00	**0.21**	0.06	**0.23**	0.05
June	-0.06	**-0.21**	**0.16**	0.07	**0.19**	0.07	**0.33**	**-0.13**
July	**-0.13**	0.00	**0.14**	-0.05	0.10	0.03	**-0.16**	0.11
Nov 2011-Feb 2012	**0.22**	**0.32**	**0.19**	**0.17**	**0.36**	-0.08	**-0.24**	0.01
2011	**-0.23**	**-0.15**	0.04	0.02	**0.19**	0.06	**0.17**	0.09

Coefficients with a P-value <0.05 in bold.

Table 6. Spearman rho correlation coefficients resulting from comparison of minimum temperature anomalies to high/low WNV regional rates.

Variable	EGP	GP	NC	NE	NW	RM	SE	SW
January	**0.19**	**0.29**	**0.11**	0.06	-0.11	0.05	**0.22**	**-0.13**
January-June	0.07	**0.24**	0.07	**0.09**	0.01	0.04	**0.19**	0.05
February	-0.08	**0.12**	**0.12**	**0.13**	-0.15	-0.13	**0.15**	0.06
March	**-0.13**	**0.26**	**0.12**	-0.03	0.02	0.05	**0.14**	0.12
April	**0.15**	**-0.12**	**-0.11**	-0.03	-0.02	0.04	**0.28**	**0.13**
May	0.04	0.09	0.06	0.08	**0.17**	0.09	**-0.12**	-0.06
May-July	**0.26**	**0.20**	**0.12**	**0.12**	0.09	0.04	**0.10**	**-0.11**
June	**0.28**	**0.11**	**0.10**	**0.14**	0.08	-0.03	**0.27**	**-0.11**
July	**0.22**	**0.20**	**0.14**	0.01	0.00	0.07	-0.07	-0.07
Nov 2011-Feb 2012	0.03	**0.17**	**0.18**	**0.10**	**-0.16**	**-0.14**	**0.14**	-0.05
2011	**-0.21**	**0.10**	0.05	0.08	-0.06	-0.04	**0.14**	0.00

Coefficients with a P-value <0.05 in bold.

Table 7. Poisson regression results by region with an estimate of the deviance explained by the model.

Region	Number of counties	Regional WNV rate per 100,000	PCT (%) model explains	Positive significant variables	Negative significant variables
EGP	326	0.76	55.0%	January-t_{max}, July-t_{min} Mean elevation	Annual t_{max} % emergent wetlands
GP	409	5.42	36.9%	March-t_{max}, autoregressive average neighbour WNV rate Hydric soils score Shrub land % row crops % emergent wetlands	
NC	652	1.30	22.8%	Nov 2011-Feb 2012-t_{min} autoregressive average neighbour WNV rate Population density	Nov 2011-Feb 2012-t_{max}
NE	532	0.40	14.3%	Autoregressive average neighbour WNV rate	July precipitation anomaly % evergreen forest
NW	159	0.45	56.2%	February-t_{max} % irrigated crops Autoregressive average neighbour WNV rate	May-July precipitation anomaly
RM	91	0.83	62.2%	March precipitation anomaly April precipitation anomaly February-t_{min} May-t_{min} % irrigated crops No. of dairy cattle Range of elevation values	Nov 2011-Feb 2012-t_{min} % white population
SE	669	2.80	36.1%	March precipitation anomaly Soil available water storage % pasture land Population <18 Autoregressive average neighbour WNV rate	May precipitation anomaly
SW	246	2.42	12.0%	Soil available water storage	

Only variables significant at the P-value <0.05 level are displayed.

between above average maximum winter temperatures and higher WNV occurrence in the Northwest region. The Rocky Mountains showed only statistically insignificant associations to climatic variables. Higher than average May and June maximum temperatures and higher than average winter minimum temperatures were associated with high WNV ratings in the Southeast region. No strong climatic associations were seen in the Southwest region.

The Poisson regression analyses were used to consider landscape, demographic and climatic variables in single models in relation to the WNV rates per region. Table 7 presents results of the Poisson regression analysis, highlighting the significant contributing variables in each model. As seen in Table 7, the largest amount of variation explained were: the Rocky Mountains, the Northwest, the Eastern Great Plains, the Great Plains, and the Southeast regions with 62.2%, 56.2%, 55.0%, 39.6%, and 36.1% of the deviance (variation) explained, respectively. The Northeast and the Southwest region had the weakest amount of deviance (14.3% and 12.0% respectively) accounted for. The spatial autoregressive variable was a statistically significant variable in five out of the eight regional Poisson regression models, indicating the importance of spatial autocorrelation. In general, the landscape, demographic and climatic associations were consistent with the results seen from the Spearman's rho correlation coefficients. Temperature variables were important in the Eastern Great Plain while in the Great Plains region row crops, shrubland, soils which could hold more water and emergent wetlands were all significantly positive terms. In the Southeast pasture, younger populations, and soils which could hold more water and high March precipitation were all positive terms while low precipitation in May was a negative term.

Discussion

Based on global and local clustering tests carried out nationally and regionally, there was evidence of spatial clustering of WNV in 2012. In Fig. 2, the results from the Anselin Local Moran's *I* test (Mitchell, 2005) calculated for the national WNV rates demonstrated one large cluster centered in the eastern portion of the Northern Great Plains region. This region had the highest WNV rate (approximately twice as high as the next highest regional rate). The high WNV rates in the Northern Great Plains region likely dampened potential significant clusters from other regions as the Anselin Local Moran's *I* test considers the global pop-

ulation mean in its local calculations. States in the northern Great Plains have been shown to have high WNV rates with consistent spatial clustering, especially in eastern Colorado and throughout Nebraska, South Dakota and North Dakota (Wimberly et al., 2008; Sugumaran et al., 2009; DeGroote and Sugumaran, 2012). South Dakota, in the Northern Great Plains region, has had the highest average incidence of WNV neuroinvasive disease in the USA since 2004 (Wimberly et al., 2013). Using the high/low WNV rating per region, statistically significant spatial autocorrelation was demonstrated in six out of eight regions based on the eight regional Global Moran's *I* tests (Table 2). In addition, numerous clusters of high WNV occurrence were demonstrated by the Anselin Local Moran's *I* tests calculated with the high/low WNV ratings that were calculated on a region by region basis (Fig. 3). Clustering of high WNV has varied geographically east to west and north to south during individual years (Wimberly et al., 2008; Sugumaran et al., 2009; Chuang and Wimberly, 2012; Wimberly et al., 2013). This geographical variation might suggest that annual climatic patterns could affect WNV transmission. Landesman et al. (2007) showed that annual precipitation from the previous year was positively associated with WNV in the eastern USA and negatively associated with WNV in the western USA from 2002-2004. Chuang and Wimberly (2012) showed that WNV occurrence from 2004-2010 in the Northern Great Plains was sensitive to temperature variability in late spring and early summer as well as moisture availability from late spring to early summer. In 2012, higher late spring and early summer rainfall and higher minimum temperatures were associated with higher WNV in the Great Plains and Eastern Great Plains region. In a recent paper, Duggal et al. (2013) concluded that it was more likely that climate and other environmental or ecological factors drove the outbreak in Texas as opposed to potential viral genetic changes. The Southeast region had the highest number of cases in 2012 with large number of cases in Dallas, Denton and tarrant counties in Texas. A large cluster, centered on these counties in eastern Texas, and stretching through Louisiana and into Mississippi was evident. Sugumaran et al. (2009) demonstrated consistent clustering of high WNV occurrence in Louisiana and Mississippi. However, the large number of WNV cases and clustering in eastern Texas has not been seen previously. Cook County and counties around Chicago, as well as the area around Detroit, were shown as a cluster of high WNV incidence for the North Central region in Fig. 3. These areas experienced high levels of

WNV previously, especially in 2002 (Ruiz et al., 2007). The clusters evident in western Texas and northern California are in counties with fairly low populations that have had sporadic higher WNV rates (Sugumaran et al., 2009).

An analysis of regional landscape and demographic associations with WNV occurrence were generally consistent with the likely dominant vectors for each of those individual regions and also with other studies from those regions. The Great Plains region was shown to have significantly positive associations between row crop areas, wetlands and hydric soils while being negatively associated with grasslands and small grains. Similar to findings in this paper, Chuang et al. (2012) found, in South Dakota, that risk for WNV from 2003-2007 was generally higher in rural areas that were also close to wetlands or soils with higher ponding frequency. DeGroote and Sugumaran (2012) similarly found positive associations with emergent wetlands and weaker positive associations with cropped areas for high WNV incidence from 2002-2009. However, they found a strong positive association with herbaceous (grasslands) which was not found in this study. The clustering of higher WNV in the Great Plains in 2012 was primarily eastern part of the region where row crop agriculture is more prominent and this was likely driven by climatic patterns like wetter late spring in this part of the region and warmer winter temperatures. The Great Plains region defined in DeGroote and Sugumaran (2012) was equal to the Great Plains and Eastern Great Plains region used in this study. The Eastern Great Plains region had positive associations with row crops and higher elevations while also showing negative associations with forested areas, wooded wetlands, and higher number of cattle. The findings for the two Plains regions are consistent with the biology of the likely dominant vector, *Cx. tarsalis*, whose larval habitats include sunlit surface pools in open areas often near agricultural areas (Reisen, 1993). DeGroote et al. (2008) demonstrated that the proportion of *Cx. tarsalis* as a proportion of all *Culex* mosquitoes grew in an east to west direction in Iowa. This supports the idea of the importance of this vector in the western portion of the Eastern Great Plains and into the Great Plains region. Higher rates of WNV occurred in western Iowa and Minnesota in 2012. Counties with Hispanic populations and with irrigated crops and dairy cattle were associated with higher WNV incidence in the Rocky Mountains which follows findings from DeGroote and Sugumaran (2012). Higher WNV ratings were associated with rural irrigated agricultural conditions in the Northwest region in 2012 which was also consistent with DeGroote and Sugumaran (2012). Interestingly, and supported by findings in this study, Crowder et al. (2013) recently reported that WNV infection in vectors (both *Cx. pipiens* and *Cx. tarsalis)* and hosts in the Northwest region were associated with orchards.

In the North Central region, positive associations with high WNV occurrence were seen with developed areas with higher incomes, while negative associations were seen with forested areas. In Detroit and Chicago, Ruiz et al. (2007) found WNV to be located in the inner suburbs with 1940-1960 era housing and moderate housing density. Although none of the associations were strong in the Northeast region, they were similar to the North Central region with positive associations with higher income developed areas and wetland areas and negatively associated with forested areas. Liu et al. (2009) reported, for a multi-year analysis in Connecticut, that residential and commercial areas were associated with WNV. In both the North Central and Northeast regions, *Cx. pipiens*, an urban/suburban mosquito, is considered the primary vector (Andreadis et al., 2004; Ruiz et al., 2004).

In the Southeast region, pasture, soils with high available water storage, cattle, and younger, poorer, and black populations were positively associated with high WNV occurrence. Similarly, DeGroote and Sugumaran (2012) found more poorly drained soils, higher poverty percentages, higher black percentages, and higher under 18% were associated with WNV incidence in this region. The regional association with higher poverty percentages might be explained by the fact that poorer households are likely to have less secure homes potentially without air-conditioning which would make them more susceptible to indoor biting from *Cx. quinquefasciatus* mosquitoes. Other studies have examined socioeconomic vulnerability factors to mosquito-borne diseases, such as Hagenlocher et al. (2013) for dengue fever in Cali, Columbia. At a larger or more detailed scale, Chuang et al. (2013) found that around Dallas, Texas that census tracts with higher property values, housing density, and unoccupied homes were positively associated with WNV occurrence in 2012. Nolan et al. (2012) found living near slow moving water sources was statistically associated with increased odds for human infection in the Houston, Texas area. Eastern Texas, which was part of a distinct cluster in 2012, has a significant amount of pasture land, which might have organically enriched standing water, outside of urban areas which might help to explain the strong positive association in

2012. The likely dominant vector in this area, *Cx. quinquefasciatus*, are found in habitats that include water with high organic content found in urban and peridomestic habitats and tend to bite indoors (Hribar et al., 2001; Reisen et al., 2008). The large number of unoccupied properties in higher income areas in and around Dallas possibly, as mentioned by Chuang et al. (2013), have unmaintained swimming pools that could serve as high organic water breeding sites for *Cx. quinquefasciatus*.

Climatic associations varied across regions and time lags. In the Southeast region, winter to spring precipitation (November 2011 to February 2012 and March and April 2012) were positively associated with higher WNV while precipitation from May to July and the previous year's precipitation were negatively associated. Higher maximum temperatures from January to June and May to July were also associated with counties with high WNV ratings. The Poisson regression model for the Southeast included the abnormally high March precipitation as a strong predictor of higher WNV rates, indicating that climatic patterns contributed to the surprisingly strong outbreak in parts of this region. Similar to this study, Chuang et al. (2012) showed that high precipitation from January-March and higher January to June temperatures were associated with the two largest WNV outbreaks (2006 and 2012) in Dallas county, Texas. These findings imply that early winter and spring rains followed by high temperatures in early summer led to conditions favourable to WNV transmission in this region. In 2012, the January to April higher than normal rainfall could have created standing water that was still present when the hotter temperatures came. The longer these standing water bodies lasted, the more enriched with organic material they would have become, providing excellent larval habitat for *Cx. quinquefasciatus* mosquitoes and potentially greater interaction between vectors and bird hosts. Similar to the present study, Wang et al. (2010) discovered an inverse relationship between human WNV incidence and rainfall from the previous year in Mississippi. In the Great Plains and Eastern Great Plains, greater precipitation in January through June (exception March) and higher average maximum temperature anomalies through April were positively associated with high county WNV ratings. Similarly, Chuang and Wimberly (2012) found that early spring temperatures may be particularly important for virus amplification in the Northern Great Plains in addition to moisture availability in late spring and early summer. Bell et al. (2006) reported that low temperatures in spring and summer led to

insufficient time for virus amplification cycles and low WNV incidence in the Red River Valley in North Dakota and Minnesota. In the North Central region, generally higher winter (November 2011-February 2012), spring and early summer average minimum and maximum temperatures were associated with higher WNV ratings while precipitation in early summer (May-July) was negatively associated with higher WNV. In Indianapolis from 2002-2007, low precipitation and warm temperatures were associated with more WNV cases (Liu et al., 2008). In the Northeast region, precipitation was generally negatively associated with WNV ratings while early summer average minimum temperatures were positively associated. Liu et al. (2009) found preceding 30 day temperature to be a significant predictor of human infection risk in Connecticut. In the Northwest region, precipitation was negatively associated with higher WNV occurrence in all time periods except April. Average maximum temperatures in the preceding winter, February, and early summer time periods were positively associated with higher WNV. Crowder et al. (2013) found precipitation to be associated with fewer WNV infections in mosquitoes in Washington, Oregon, and southwestern Idaho, while temperature did not have any significant associations. The western part of the Northwest region can be very wet and it is likely that the drier eastern agricultural parts of the region favour transmission by *Cx. tarsalis* mosquitoes.

The WNV outbreak of 2012 came as a surprise after several years of relatively few cases throughout the USA. This study demonstrated that there were clear clusters of high WNV occurrence with two large clusters being seen in eastern Nebraska, South Dakota, and North Dakota and western Minnesota and in eastern Texas, Louisiana, and Mississippi. This study highlighted potential drivers of these outbreaks, including the large outbreak in the Southeast region being possibly driven by high winter and spring precipitation followed by drier and hotter than normal conditions in the summer. The regional climatic, landscape, and demographic associations with WNV ratings in 2012 were consistent with previous studies for individual regions and also consistent with the ecology of the potentially most important mosquito vector for each region. The regions used in this study were fairly large in area and the ecological dynamics likely vary within these regions. It might be useful to further subdivide the regions for more detailed analysis. However, that type of study would require more detailed spatial data on WNV incidence that is not generally available. Also there are many areas of the country in which two vec-

tor species with somewhat conflicting habitat preferences are present. In a study such as this it is difficult to considering local vector dynamics and how they might change in a given year. In addition, there were important factors that might help to explain variation in WNV occurrence that could not be considered in this study due to lack of available data. These factors include, but are not limited to, levels of immunity in host and human populations and spatial patterns of the presence and abundance of potential hosts. For the coterminous USA, our regional study has provided valuable insight into the influence of climate, landscape, and demographic conditions on occurrence of WNV in 2012 which can help inform future studies at various scales.

Acknowledgements

References

Agresti A, 2007. An introduction to categorical data analysis. (second edition). New Jersey: John Wiley and Sons, 372 pp.

Andreadis TG, Anderson JF, Vossbrinck CR, Main AJ, 2004. Epidemiology of West Nile virus in Connecticut: a five-year analysis of mosquito data 1999-2003. Vector Borne Zoonotic Dis 4, 360-378.

Anselin L, 1995. Local indicators of spatial association - LISA. Geogr Anal 27, 93-115.

Bell JA, Brewer CM, Mickelson NJ, Garman GW, Vaughan JA, 2006. West Nile virus epizootiology, Central Red River Valley, North Dakota and Minnesota, 2002-2005. Emerg Infect Dis 12, 1245-1247.

Bowden SE, Magori K, Drake JM, 2011. Regional differences in the association between land cover and West Nile virus disease incidence in humans in the United States. Am J Trop Med Hyg 84, 234-238.

Brown H, Diuk-Wasser M, Andreadis T, Fish D, 2008. Remotely-sensed vegetation indices identify mosquito clusters of West Nile virus vectors in an urban landscape in the Northeastern United States. Vector Borne Zoonotic Dis 8, 197-206.

CDC, 2014. West Nile virus. Available at: http://www.cdc.gov/westnile/resources/pdfs/cummulative/99_2013_CasesAndDeathsClinicalPresentationHumanCases.pdf (accessed on June 2014).

Chase JM, Knight TM, 2003. Drought-induced mosquito outbreaks in wetlands. Ecol Lett 6, 1017-1024.

Chuang TW, Hockett CW, Kightlinger L, Wimberly MC, 2012. Landscape-level spatial patterns of West Nile virus risk in Northern Great Plains. Am J Trop Med Hyg 86, 724-731.

Chuang TW, Wimberly MC, 2012. Remote sensing of climatic anomalies and West Nile virus incidence in the Northern Great Plains of the United States. PLoS One 7, e46882.

Chuang WM, Buseman CM, Joyner SN, Hughes SM, Fomby TB, Luby JP, Haley RW, 2013. The 2012 West Nile encephalitis epidemic in Dallas, Texas. JAMA 310, 297-307.

Crowder DW, Dykstra EA, Brauner JM, Duffy A, Reed C, Martin E, Peterson W, Carriére Y, Dutilleul P, Owen JP, 2013. West Nile virus prevalence across landscapes is mediated by local effects of agriculture on vector and host communities. PLoS One 8, e55006.

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP, 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the coterminous United States. Int J Climatol 28, 2031-2064.

Darsie RF Jr, Ward RA, 2005. Identification and geographical distribution of the mosquitoes of North America, north of Mexico. Gainesville: University Press of Florida, 383 pp.

Day JF, Shaman J, 2008. Using hydrologic conditions to forecast the risk of focal and epidemic arboviral transmission in peninsular Florida. J Med Entomol 45, 458-465.

DeGroote JP, Sugumaran R, Brend SM, Tucker BJ, Bartholomay LC, 2008. Landscape, demographic, entomological, and climatic associations with human disease incidence of West Nile virus in the state of Iowa, USA. Int J Health Geogr 7, 197-206.

DeGroote JP, Sugumaran R, 2012. National and regional associations between human West Nile virus incidence and demographic, landscape, and land use conditions in the coterminous United States. Vector Borne Zoonotic Dis 12, 657-665.

Diuk-Wasser MA, Brown HE, Andreadis TG, Fish D, 2006. Modeling the spatial distribution of mosquito vectors for West Nile virus in Connecticut, USA. Vector Borne Zoonotic Dis 6, 283-295.

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling et al., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609-628.

Duggal NK, D'Anton M, Xiang J, Seiferth R, Day J, Nasci R, Brault AC, 2013. Sequence analysis of 2012 West Nile virus isolates from Texas fail to associate viral genetic factors with outbreak magnitude. Am J Trop Med Hyg 89, 205-210.

Epstein PR, Defilippo C, 2001. West Nile virus and drought. Global Change Human Health 2, 105-107.

Ezenwa VO, Godsey MS, King RJ, Guptill SC, 2006. Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. Proc Biol Sci 7, 109-117.

Gates MC, Boston RC, 2009. Irrigation linked to a greater incidence of human and veterinary West Nile virus cases in the United States from 2004 to 2006. Prev Vet Med 1, 134-137.

Gibbs SEJ, Wimberly MC, Madden M, Masour J, Yabsley MJ, Stallknecht DE, 2006. Factors affecting the geographic distribution of West Nile virus in Georgia, USA: 2002-2004. Vector Borne Zoonotic Dis 6, 73-82.

Ghosh D, Guha R, 2010. Use of genetic algorithm and neural network approaches for risk factor selection: a case study of West Nile virus dynamics in an urban environment. Computers Environ Urban 34, 189-203.

Gould LH, Fikrig E, 2004. West Nile virus: a growing concern? J Clin Invest 113, 1102-1107.

Hagenlocher M, Delmelle E, Casas I, Kienberger S. Assessing socioeconomic vulnerability to dengue fever in Cali, Columbia: statistical vs. expert-based modeling. Int J Health Geogr 12, 36.

Hayes EB, Komar N, Nasci RS, Montgomery SP, O'Leary DR, Campbell GL, 2005. Epidemiology and transmission dynamics of West Nile virus disease. Emerg Infect Dis 11, 1167-1173.

Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C, Herold N, McKerrow A, VanDriel JN, Wickham J, 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogramm Eng Remote Sensing 73, 337-341.

Hribar LJ, Smith JM, Vlach JJ, Verna TN, 2001. Survey of container-breeding mosquitoes from the Florida Keys, Monroe County, Florida. J Am Mosq Control Assoc 17, 245-248.

Kuhn K, Campbell-Lendrum D, Haines A, Cox J, 2005. Using climate to predict infectious disease epidemics. Geneva: World Health Organization, 54 pp.

Landesman WJ, Allan BF, Langerhans RB, Knight TM, Chase JM, 2007. Inter-annual associations between precipitation and human incidence of West Nile virus in the United States. Vector Borne Zoonotic Dis 7, 337-343.

Liu A, Lee V, Galusha D, Slade M, Diuk-Wasser M, Andreadis T, Scotch M, Rabinowitz PM, 2009. Risk factors for human infection with West Nile virus in Connecticut: a multi-year analysis. Int J Health Geogr 8, 67.

Liu H, Weng Q, Gaines D, 2008. Spatio-temporal analysis of the relationship between WNV dissemination and environmental variables in Indianapolis, USA. Int J Health Geogr 7, 66.

Mitchell A, 2005. The ESRI guide to spatial analysis. ESRI Press.

Nielsen CF, Armijos MV, Wheeler S, Carpenter TE, Boyce WM, Kelley K, Brown D, Scott TW, Reisen WK, 2008. Risk factors associated with human infection during the 2006 West Nile virus outbreak in Davis, a residential community in Northern California. Am J Trop Med Hyg 78, 53-62.

Nolan MS, Zangeneh A, Khuwaja SA, Martinez D, Rossman SN, Cardenas V, Murray KO, 2012. Proximity of residence to bodies of water and risk for West Nile virus infection: a case-control study in Houston, Texas. J Biomed Biotechnol, article ID 159578.

Oregon State University PRISM Climatic Group. PRISM Products Matrix, 2013. Available at: http://www.prism.oregonstate.edu/products/matrix.phtml?view=data (accessed on March 2013).

Petersen LR, Brault AC, Nasci RS, 2013. West Nile virus: review of the literature. JAMA 310, 308-315.

Petersen LR, Fischer M, 2012. Unpredictable and difficult to control - the adolescence of West Nile virus. N Engl J Med 367, 1281-1284.

Reisen W, 1993. The western encephalitis mosquito, *Culex tarsalis*. Wing Beats 4, 16.

Reisen W, Lothrop H, Chiles R, Madon M, Cossen C, Woods L, Husted S, Kramer V, Edman J, 2004. West Nile virus in California. Emerg Infect Dis 10, 1369-1378.

Reisen WK, Lothrop HD, Wheeler SS, Kennsington M, Gutierrez A, Fang Y, Garcia S, Lothrop B, 2008. Persistent West Nile virus transmission and the apparent displacement St. Louis encephalitis virus in southeastern California, 2003-2006. J Med Entomol 45, 494-508.

Ruiz MO, Chaves LF, Hamer GL, Sun T, Brown WM, Walker ED, Haramis, L, Goldberg TL, Kirtron 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, Kitron U, 2004. Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. Int J Health Geogr 3, 8.

Ruiz MO, Walker ED, Foster ES, Haramis LD, Kitron UD. 2007. Association of West Nile virus illness and urban landscapes in Chicago and Detroit. Int J Health Geogr 6, 10.

Soverow JE, Wellenius GA, Fisman DN, Mittleman MA, 2009. Infectious diseases in a warming world: how weather influenced West Nile virus in the United States (2001-2005). Environ Health Perspect 117, 1049-1052.

Sugumaran R, Larson SR, DeGroote JP, 2009. Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. Int J Health Geogr 8, 43.

Tackett J, Charnigo R, Caldwell G, 2006. Relating West Nile virus case fatality rates to demographic and surveillance variables. Public Health Rep 121, 666-673.

Uejio CK, Kemp A, Comrie AC, 2012. Climatic controls on West Nile virus and sindbis virus transmission and outbreaks in South Africa. Vector Borne Zoonotic Dis 12, 117-125.

United States Census Bureau. Population Estimates, 2013. Available at: http://www.census.gov/popest/data/counties/totals/2011/index.html (accessed on March 2013).

United States Department of Agriculture Natural Resource Conservation Service. Soil Data Mart, 2011. Available at: http://soildatamart.nrcs.usda.gov/ (accessed on May 2011).

United States Department of Interior, 2013. Available at: http://www.nationalatlas.gov/index.html (acessed on March 2013).

United States Geologic Survey, 2012. Disease Maps Home. Available at: http://diseasemaps.usgs.gov/index.html (accessed on December 2012).

Wang G, Minnis RB, Belant JL, Wax CL, 2010. Dry weather induces outbreaks of human West Nile virus infections. BMC Infect Dis 10, 38.

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

Wimberly MC, Hildreth MB, Boyte SP, Lindquist E, Kightlinger L, 2008. Ecological niche of the 2003 West Nile virus epidemic in the Northern Great Plains of the United States. PLoS One 3, e12.

Winters AM, Eisen RJ, Lozano-Fuentes S, Moore CG, Pape WJ, Eisen L, 2008. Predictive models of risk of West Nile virus exposure in eastern and western Colorado. Am J Trop Med Hyg 79, 581-590.