VISUALIZING THE SPREAD OF WEST NILE VIRUS

Jürgen Symanzik*, Utah State University,
Samson Gebreab, Robert Gillies, Utah State University,
James Wilson, Georgetown University Medical Center.
*Department of Mathematics and Statistics, Logan, UT 84322–3900,
e–mail: `symanzik@math.usu.edu`

Key Words: Choropleth Map, Dead Birds, Infectious Disease, Micromaps, Mosquitoes, Remote Sensing.

Abstract

The West Nile Virus was first recognized in the United States in 1999 as the cause of severe and fatal human illness in metropolitan New York City. Since 1999, the West Nile Virus has rapidly spread westward. By the end of 2002, there were only a few states in the Western United States where the virus had not been observed. However, indications are that the virus will be observed in every state of the contiguous United States by the end of 2003. The West Nile Virus transmission cycle involves different species of mosquitos and birds and can be deadly to humans, horses, birds, and other animals.

In this paper, we visualize the spread and occurrence of the West Nile Virus and graphically relate these to various climatic and environmental conditions. Part of the data that is being presented originates from remote sensing satellites as well as from different medical centers. Our visual exploration is based on micromaps and standard choropleth maps.

1. Introduction

In 1999, West Nile (WN) fever, a mosquito–transmitted viral infectious disease, was identified for the first time in the Western Hemisphere in New York City (NYC). This was an unprecedented event, as West Nile Virus (WNV) had been considered endemic to the Middle East, Africa, and Asia. The first human cases of this outbreak were identified on August 2, 1999, with additional cases reported until September 22, 1999. Investigators retrospectively identified 61 human cases, 55 of which were hospitalized with infection involving brain tissue ("meningo–encephalitis"), resulting in 7 deaths (Nash, Mostashari, Fine, Miller, O'Leary, Murray, Huang, Rosenberg, Greenberg, Sherman, Wong & Layton 2001). Further investigation revealed that a WN epidemic in birds had preceded the human phase of the epidemic by nearly a month. Moreover, the virus was found to be capable of over–wintering in local mosquitoes and thus had gained permanent ecological establishment (CDC 2000). This and further events caught the attention of the media.

The transmission cycle of WNV is complicated. An amplification cycle is the initial event where bird–preferential mosquito vectors (e.g., *Culex* species) transmit the virus within local avian populations. Once the amplification cycle has started, bridging mosquito vectors (e.g., *Aedes albopictus*) feed on infected birds and then bite animals and humans, and so infect them. In humans, the vast majority of West Nile cases present as a mild viral illness. In some patients older than 55 years of age, the virus has a higher tendency to infect the brain tissue, with associated higher death rates. Presently, there is no cure and no vaccine for humans. However, there exists a vaccine for animals, in particular horses. Preventive measures for humans include spraying for mosquitoes and wearing long sleeved shirts and applying insect repellent (`http://www.cdc.gov/ncidod/dvbid/westnile/ecology.htm`).

The purpose of this paper is two–fold: the first is to demonstrate the appropriateness of statistical visualization techniques to identify WNV spatially (Section 3); the second is the application of similar statistical techniques to investigate the 2002 outbreak of WNV in the District of Columbia (DC) in a spatial and temporal context (Section 4). We start with a summary of our data and methods in Section 2 and finish with our conclusions in Section 5.

2. Materials and Methods

2.1. Data

WNV data at the US level were obtained from several Web sites maintained by the Centers for Disease Control and Prevention (CDC), in partiuclar, `http://www.cdc.gov/ncidod/dvbid/westnile/surv&controlCaseCount02.htm` for the 2002 counts of human WNV infections and linked Web pages for human data prior to 2002 and animal data.

Avian, mosquito, and human WNV surveillance

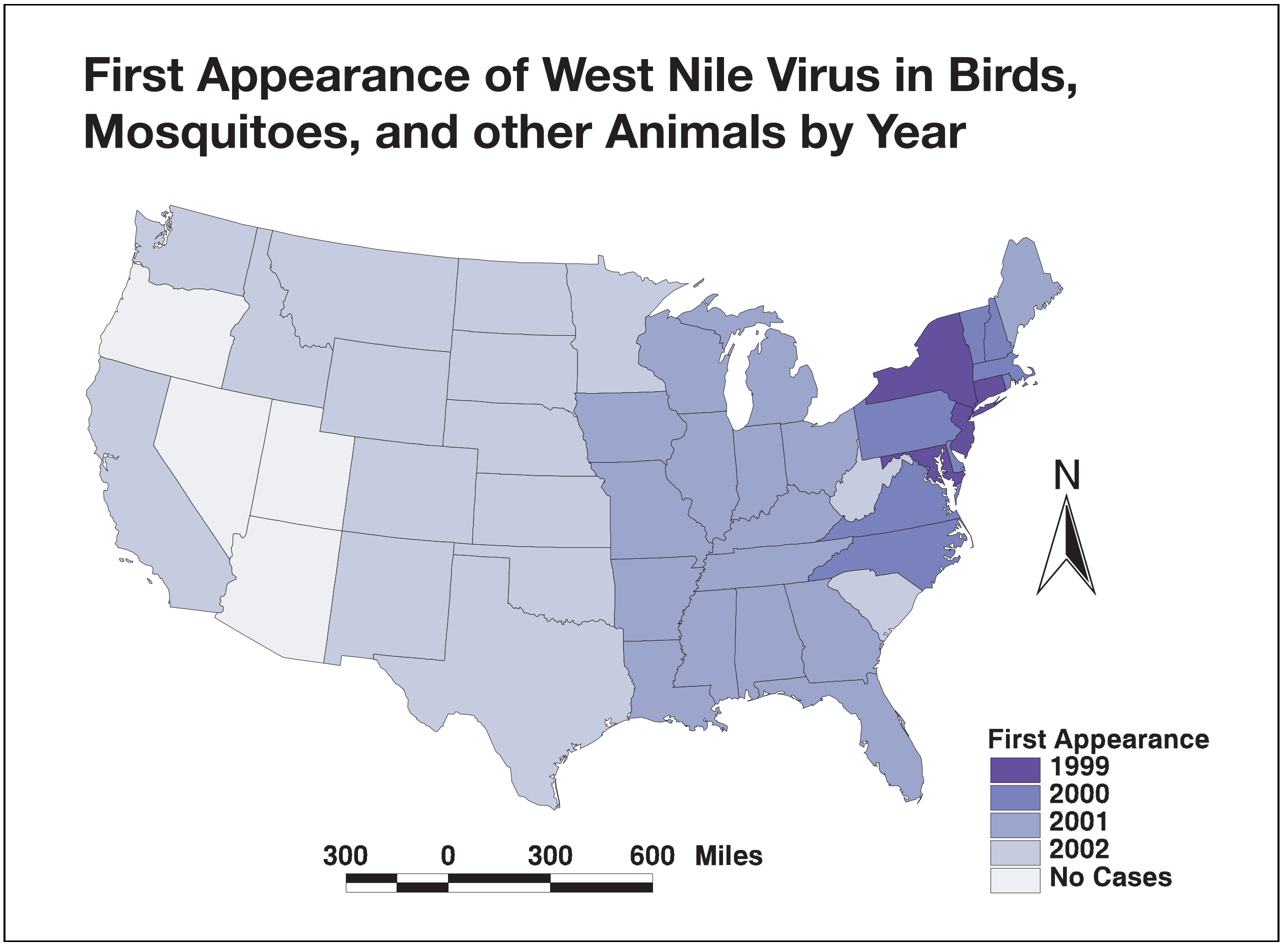

Figure 1: Spread of WNV in birds, mosquitoes, and other animals since 1999. The initial North – South spread along the East Coast can be explained by the migratory pathways of birds. One possible explanation for the quick spread of WNV from the East to the West can be the incidental transportation of infected mosquitoes via airplanes.

data within the boundaries of Washington, DC were collected by the District of Columbia Department of Health (DC DOH), US Army Center for Health Promotion and Preventive Medicine (CHPPM), and the US Park Service. Geotemporal information was recorded with each data entry, however differences in how geotemporal information was recorded among the three agencies varied. For instance, CHPPM recorded GPS coordinates for all mosquito trap sites whereas DC DOH recorded street addresses. Avian surveillance was based on citizens calling the DC DOH to report dead birds from May to August 2002. Once a collected bird was determined WNV–positive, no further collections were conducted for the neighborhood in question. A total of 190 birds were collected, 160 of which were tested positive for WNV. Mosquitoes were collected continually at the same sites from April to October 2002. For all three agencies combined, a total of 1245 pools of mosquitoes (10584 total mosquitoes) were collected, speciated, and tested for presence of WNV. Of these, 84 positive pools of mosquitoes were found. Human cases were identified from August to October 2002 on the basis of individuals presenting to physicians who proceeded to test them for WNV. A total of 25 individuals were identified, which is less than the published number by CDC of 34 (`http://www.cdc.gov/ncidod/dvbid/westnile/surv&controlCaseCount02.htm`). The number of human cases reported to CDC by DC DOH reflected suspected WNV positives and cases that did not have geotemporal information recorded. We opted to include only those cases that were laboratory–confirmed positive and had both temporal and geographic information recorded for the entry.

Population data estimates for 2001 (the most recent available population data at the time of the analysis) were obtained from the U.S. Census Bureau Web site (`http://www.census.gov`), specifically the Web pages for each state

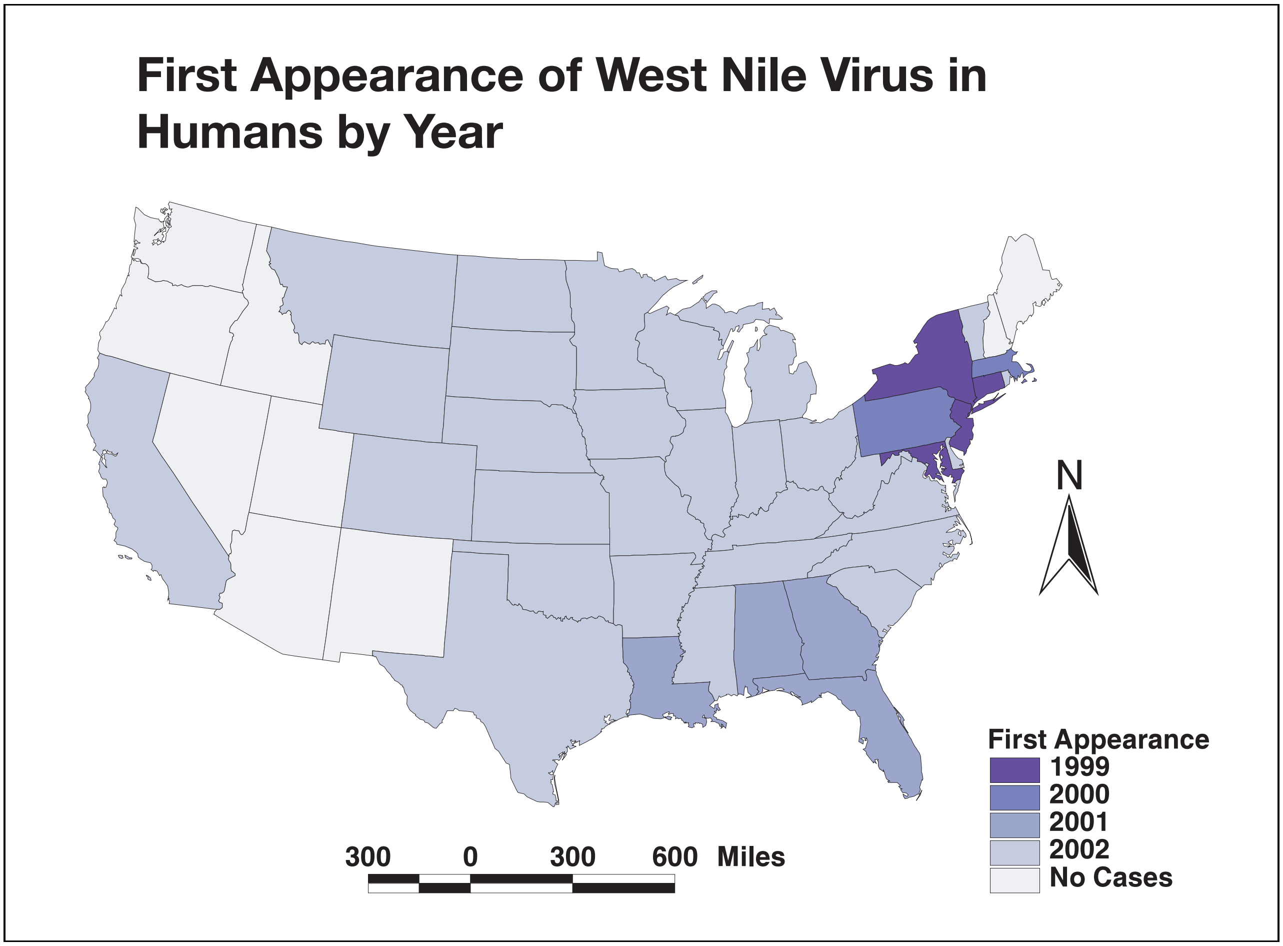

Figure 2: Spread of WNV in humans since 1999. The first appearance of WNV in humans in a state usually lags the appearance of WNV in birds, mosquitoes, and other animals by a few months up to 1 or 2 years. By the end of 2002, only 11 States had no human WN cases.

(`http://quickfacts.census.gov/qfd/states/01000.html` through `http://quickfacts.census.gov/qfd/states/56000.html`). This population data has been used to calculate the (non–age–adjusted) human WN rates for each state in 2002.

Climate data for the study was obtained from the Utah State University Interactive Climate & Weather Server (`http://climateweb.ser.usu.edu`). For four locations within the Washington, DC neighborhood, the daily minimum and maximum temperatures (T_{min} and T_{max}) and daily precipitation were extracted from January 2002 through May 2003. T_{min} and T_{max} were subsequently averaged for the month while daily precipitation was summed for the month.

Normalized Difference Vegetation Index (NDVI) data was derived from SPOT (Le Système Pour l'Observation de la Terre) satellite data courtesy of the NASA–Goddard Space Flight Center. NDVI (at a spatial resolution of 1km × 1km pixels) was available as 10–day composites. These composites were averaged over the pixels within the wards (i.e., geographics subregions of Washington, DC) where NDVI was greater than zero — this was to exclude those areas which were built–up, i.e., contain no vegetation.

2.2. Visualization Techniques

The main graphical presentation techniques used in the following sections are choropleth maps, linked micromap plots, and time series (graphs).

According to Harris (1999), "a choropleth map is a variation of a statistical map that displays area data by means of shading, color, or patterns. ... The data is generally in terms of ratios, percents or rates as opposed to absolute units." Further details on choropleth maps can be found in MacEachren (1995). In this paper we display the spread of WNV over time and the human WNV infection rate by state via choropleth maps. The color schemes used

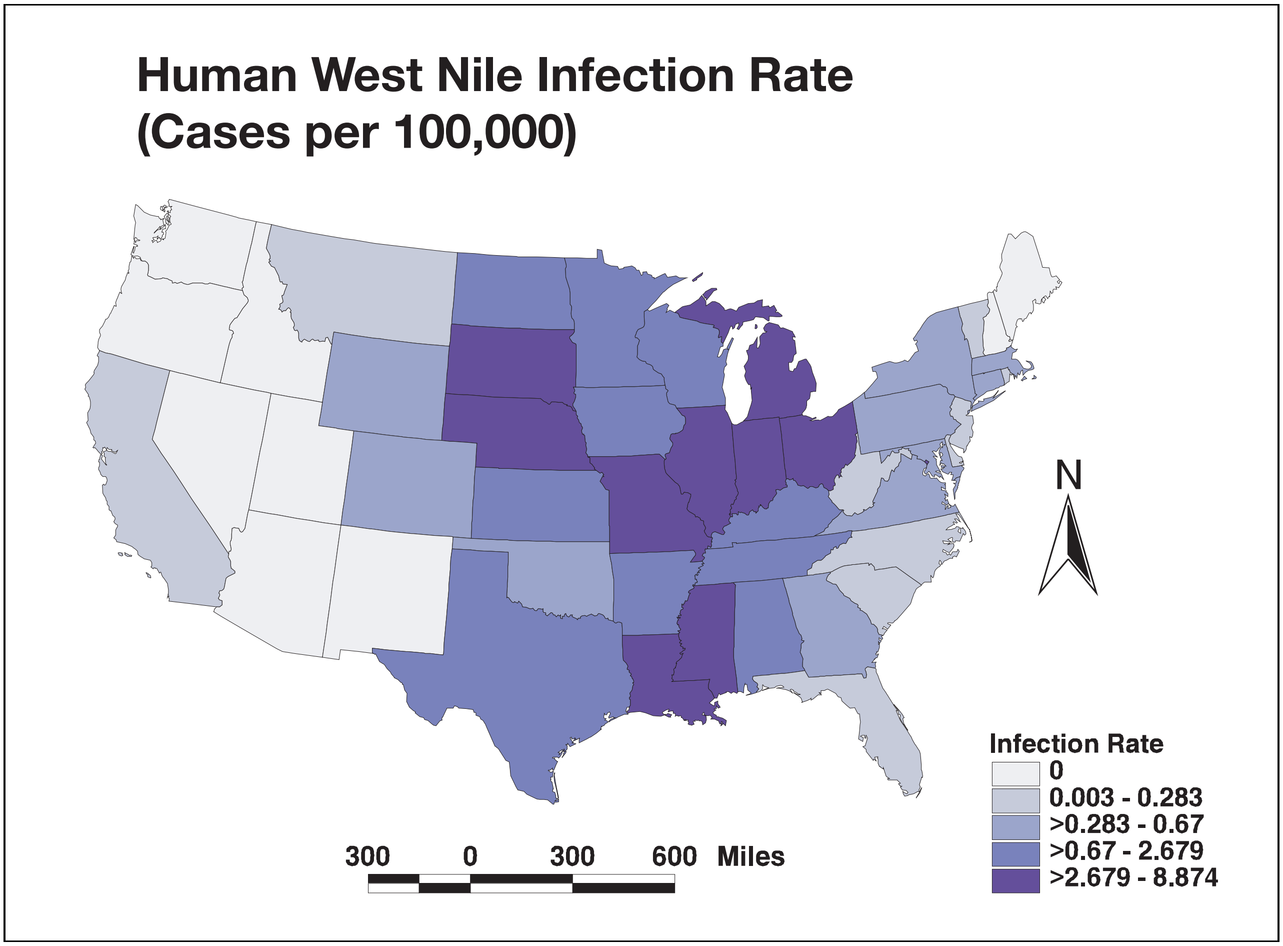

Figure 3: Human WN infection rates by state, based on data sources from CDC and the U.S. Census Bureau. Highest rates were observed in the central states. Due to the lack of available data, the rates on this map are not age adjusted.

for these choropleth maps have been obtained from Cindy Brewer's `http://colorbrewer.com` Web site (see Leslie (2002) for a description of this Web site). The choropleth maps were produced with the Geographic Information System (GIS) ArcView 3.2.

Linked micromap (LM) plots, often simply called micromaps, provide a new statistical paradigm for the viewing of geographically referenced statistical summaries in the corresponding spatial context. The main idea behind LM plots is to focus the viewer's attention on the statistical information presented in a graphical display. Multiple small maps are used to provide the appropriate geographic reference for the statistical data. LM plots were first presented at the 1996 Joint Statistical Meetings in Chicago. More details on the history of LM plots and their connection to other research can be found in early references on micromaps (Carr & Pierson 1996, Carr, Olsen, Courbois, Pierson & Carr 1998). Recent references on LM plots (Symanzik, Carr, Axelrad, Wang, Wong & Woodruff 1999, Carr, Wallin & Carr 2000, Carr 2001, Symanzik & Jones 2001, Symanzik, Hurst & Gunter 2002) focus on their use for communicating summary data from health and environmental studies. Wang, Chen, Carr, Bell & Pickle (2002) and Carr, Chen, Bell, Pickle & Zhang (2002) provide more details on the design of the National Cancer Institute (NCI) Web site, accessible at `http://www.statecancerprofiles.cancer.gov/micromaps`, that provides interactive access to NCI cancer data via micromaps. The LM plot in this article is based on sample S–Plus code that can be obtained from Dan Carr's ftp site at `ftp://galaxy.gmu.edu/pub/dcarr/newsletter/micromap/`.

According to Harris (1999), "a time series graph ... has a time series (frequently abbreviated to time) scale on the horizontal axis and a quantitative scale on the vertical axis." In this paper we make use of time series plots, produced with S–Plus, to

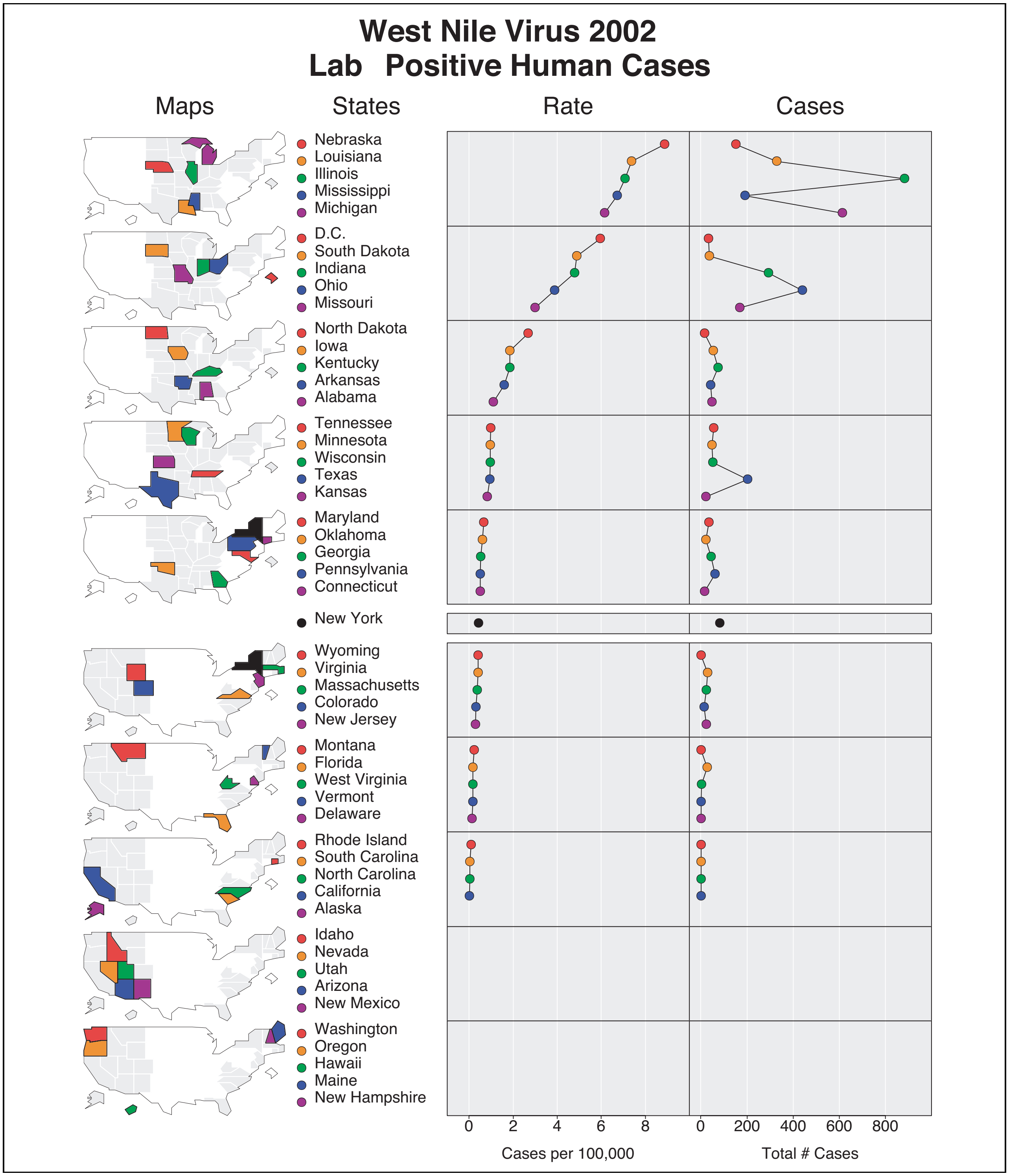

Figure 4: Linked micromap plots showing the same WN infection rates as in Figure 3 and total number of human WN cases for states. The micromap plots highlight spatial structures (highest WN rates in the central states, lowest WN rates along the East Coast), spatial outliers (notable DC), and regions commonly overlooked when referencing a choropleth representation as in Figure 3 — i.e., Nebraska with only 152 cases, but highest WN rate. No WN cases have been observed in 2002 in the eleven states listed at the bottom. New York has the median WN rate.

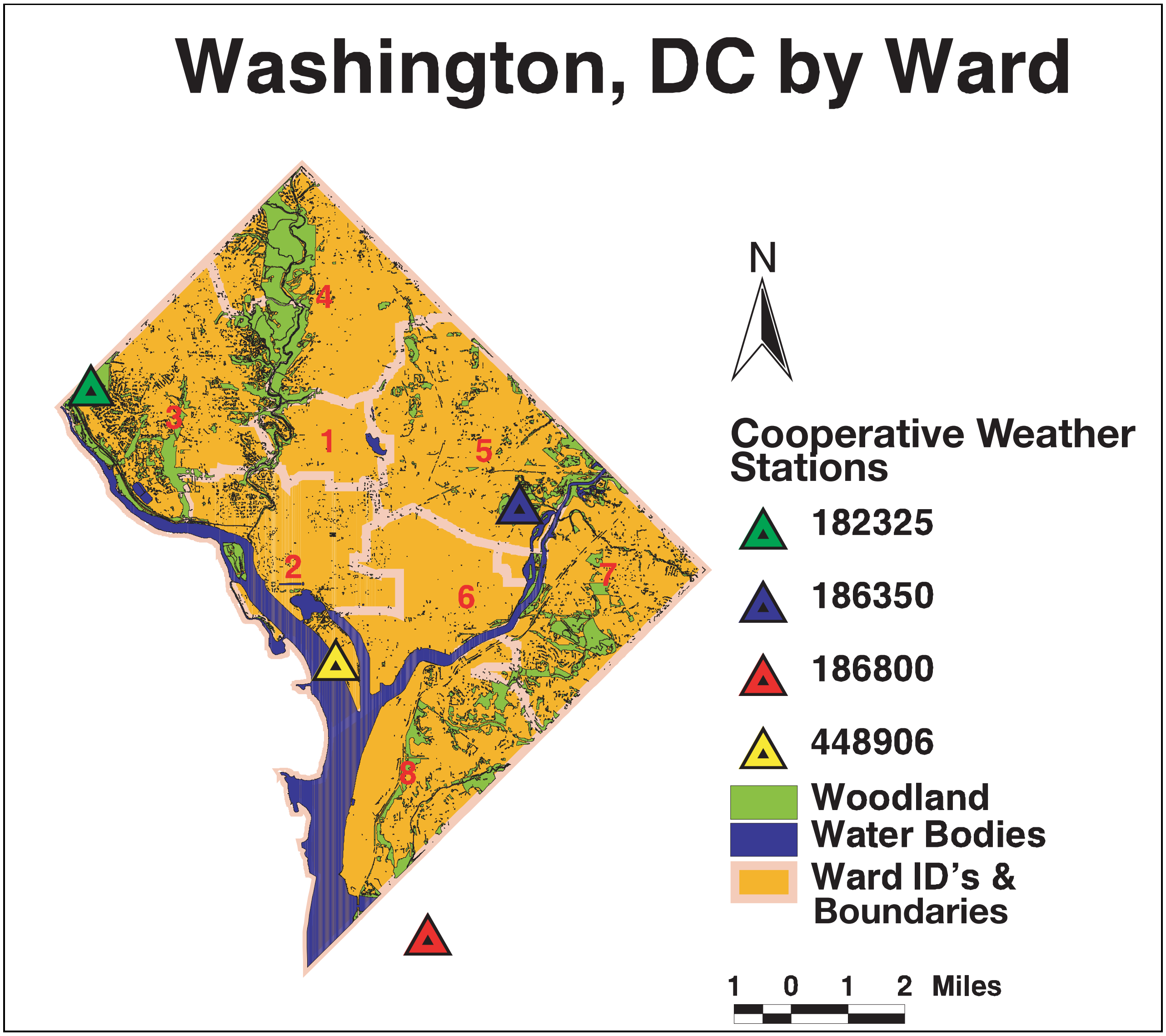

Figure 5: Washington, DC, map, highlighting administrative units (wards), woodland areas and water bodies, and COOP weather stations.

display the change of environmental conditions such as temperature and precipitation throughout the year.

In addition, we present a geographic time–series plot (via choropleth maps), also produced with S–Plus, that is used to display the change of WNV rates in different geographic regions over time.

3. US Analysis

In 2002, 4,156 human cases of WN fever were identified in the US, with 284 deaths. This was the largest insect–vectored meningo–encephalitis epidemic in the known history of the Western Hemisphere and the largest WNV–related meningo–encephalitis epidemic worldwide recorded to–date. Ecological damage has been an unforeseen consequence, with over 140 species of birds, reptiles, and mammals infected and killed by WNV in the US. Over 36 species of mosquitoes are known to be able to transmit the virus. Moreover, 14,000 horses were killed due to WNV since the epidemic began in 1999. WNV has also posed a major threat to endangered species and zoological park animals, with over 100 US zoos reporting cases. The spread of WNV is exemplified by the choropleth maps of the US as shown in Figures 1 and 2.

Figure 3, a similar representation as Figures 1 and 2, showing infection rate by state, demonstrates the limitation of choropleth representations.

However, a micromap representation of the same

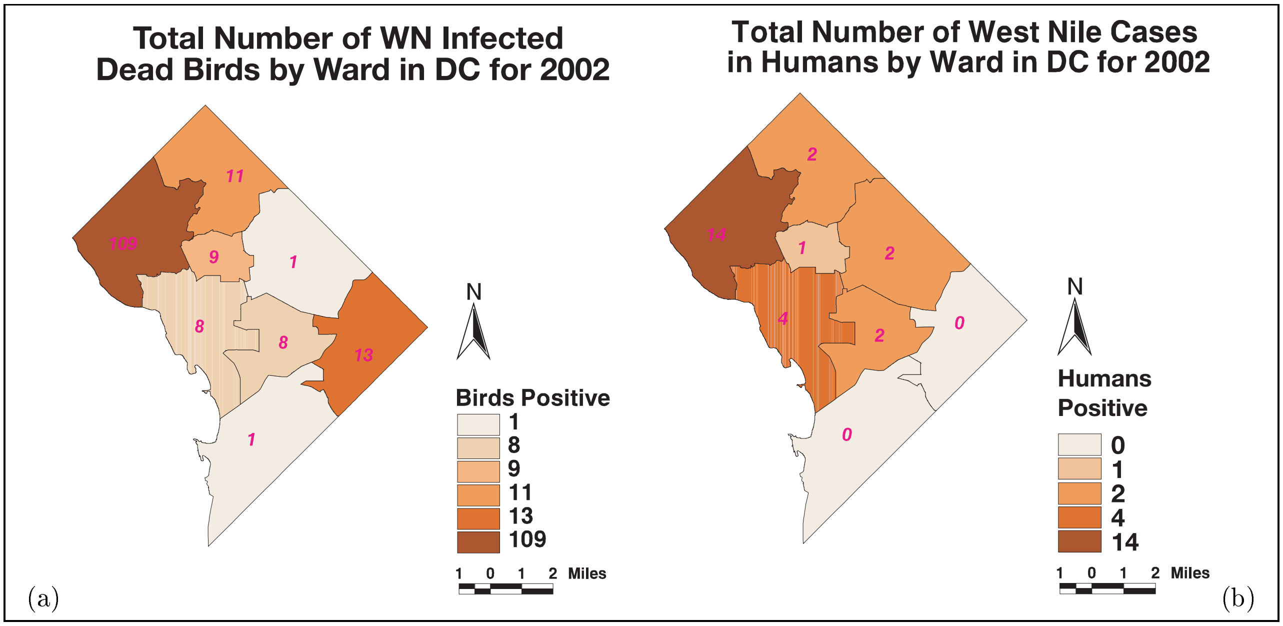

Figure 6: (a) Total number of WN infected dead birds in DC in the summer of 2002. The highest number was found in Ward 3. The map is based on a convenience sample based on people reporting dead birds. Moreover, only about 75% of reported dead birds could be tested for WNV. (b) Total number of WN infections in humans in DC in the summer of 2002. Similar to the bird WN infections in Figure 6(a), the highest number was found in Ward 3. The total number of human cases reported on this map is less than the numbers for DC in Figures 3 and 4. Some of the human cases initially reported to the CDC eventually turned out to be WN negative.

data (see Figure 4) illustrates the individual infection rates (i.e., exact rates for each state) somewhat differently. In particular, ranking of states with respect to WN rates is possible in micromap plots. Of note, in Figure 4, is the high national profile of DC in respect to infection rate.

4. DC Analysis

The first evidence of WNV in the District of Columbia was documented in 2000 with the discovery of 5 WNV positive birds; this number increased to 360 in 2001 with the discovery of 3 positive pools of vector mosquitoes and no other animal or human cases. In 2002, the first positive bird was identified on May 2nd, first positive mosquito vectors were found in July, and the first human case was identified in late July. Full ecological establishment of WNV in the District of Columbia occurred in 2002, as reported on the DC DOH West Nile Virus and Malaria Update Web site (`http://dchealth.dc.gov/information/wnv/wnv&malaria_update110402.shtm#program_overview`).

For future reference, Figure 5 is a representative map of the Washington, DC area highlighting the eight wards that comprise the administrative units used by the health department. In addition, both woodland and water bodies are identified for reference, along with the cooperative weather stations (COOP). Data from the COOP stations are part of our enviro–climatic analysis for the DC area.

An analysis for DC similar to that for the US is portrayed in Figures 6(a) and 6(b). Figure 6(a) is quite striking in that it shows that Ward 3 is associated with a ten–fold elevation in WNV–infected birds as compared to the next highest ward. Figure 6(b), the number of WNV cases in humans, mirrors the statistics for Ward 3 in that the highest number of human cases occurs in this ward. A further analysis of DC is given by a geographic time series representation (via choropleth maps) of the weekly mosquito WNV–positive rates (Figure 7). Note that Ward 3 again stands out as an area of high level of infection. Figure 7 shows a time series of 12 choropleth maps. This visualization provides the benefit of noting the timing of mosquito positive rates reaching a critical value. Figures 6(a) and 6(b) could be shown in a similar manner, and, together with Figure 7, these maps would show the temporal progression of the disease as would be expected (i.e., first birds, then mosquitoes, then humans).

A further analysis of the enviro–climatic set is given in Figure 8, panels (a) to (c). The spring seasonal transition in the DC area was noted based on air temperature (min and max) and remotely sensed derived NDVI data: NDVI being an indicator of vegetation cover over the ground. NDVI is calculated

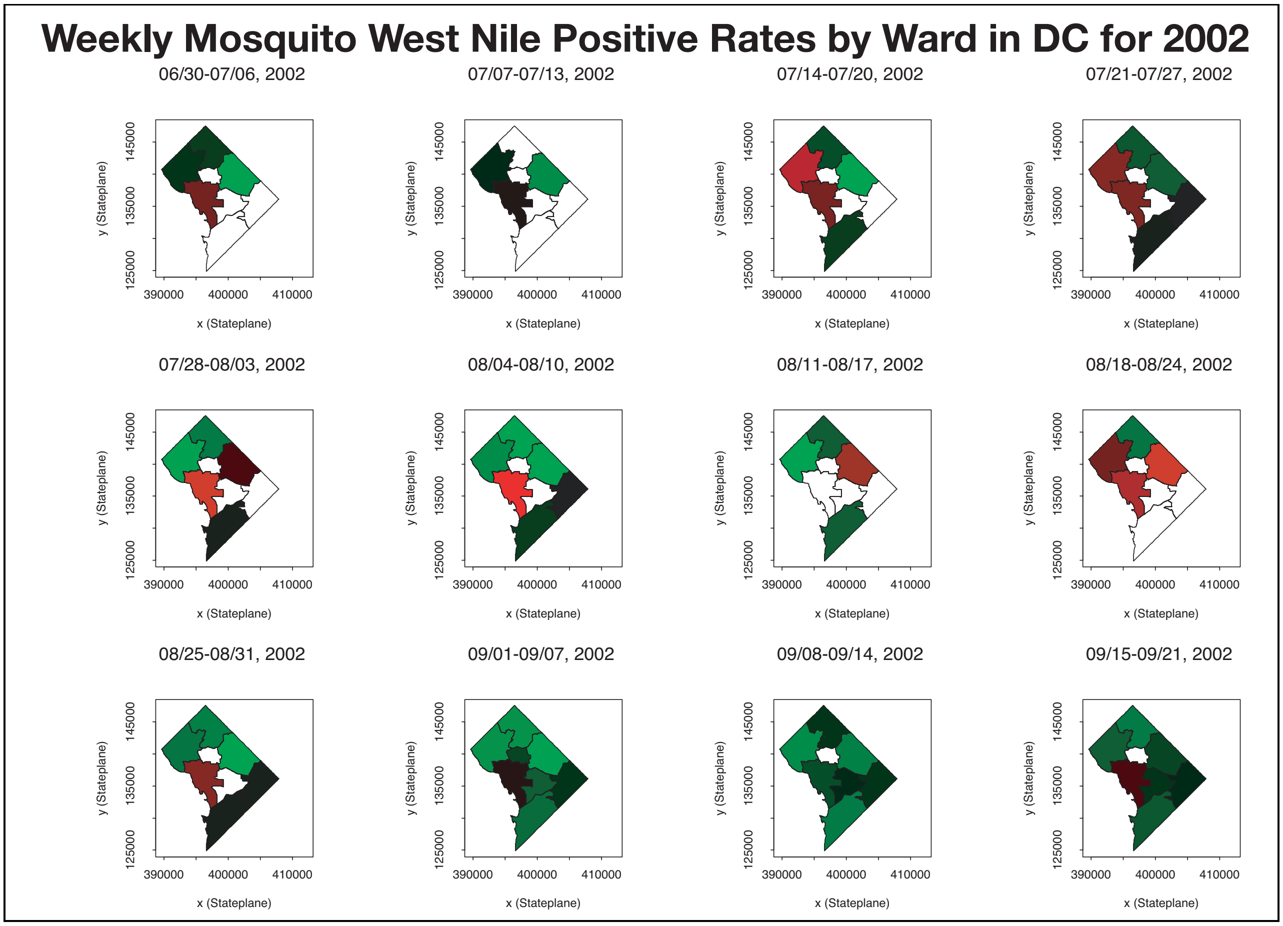

Figure 7: Geographic time–series (via choropleth maps) of the spread of WN in mosquitoes in the summer of 2002. Red is used to display regions where sampled pools of mosquitoes tested positive. The brighter the red, the higher the percentage of positive sampled pools. The brightest red is used, for example, in week 8/4 – 8/10 for Ward 2 where 42% of the sampled pools (11 out of 26) tested positive. Green is used to display regions where all sampled pools of mosquitoes tested negative. The brighter the green, the higher the number of negative samples. The brightest green is used in week 8/4 – 8/10 for Ward 4 where all 33 sampled pools tested negative. A ward is not color–coded if no mosquitoes were sampled in that ward during a particular week. In week 9/1 – 9/7, mosquito samples were collected in all wards. No positive mosquito samples were detected before week 6/30 – 7/6 or after week 9/15 – 9/21.

as $NDVI = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1}$, where $-1 \leq NDVI \leq 1$ and ρ_1 and ρ_2 represent the visible red band and near–infrared band, respectively. Highly vegetated surfaces have NDVI values around 0.8 while bare soil surfaces have NDVI values around 0.1.

The minor rainy season (see Figure 8(b)) in DC from March to June likely contributed to hydration and subsequent hatching of mosquito eggs. Of note was the observation of a temperature (see Figure 8(a)) and NDVI (see Figure 8(c)) peak over July and August. This is an important observation because the ability of a mosquito to transmit WNV is maximized with increases in temperature, specifically 26 to 30 degrees Centigrade. Maximized transmission occurs when, in the context of optimized temperature, a mosquito feeds on a WNV–positive host and the virus gains entry into the mosquito and disseminates to the salivary glands, thus enabling transmission. Optimized temperature (i.e., the 60% and 100% lines indicated in Figure 8(a)) implies acceleration of time to transmission competency and enhancement of epidemic propagation (Turell 2003). Temperatures around 30 degrees Centigrade occurred within the time frame of infection as observed in the bird, mosquito, and human analyses provided earlier. We believe the discovery of positive mosquitoes followed by positive human cases in July, August, and September is directly related to this temperature increase and raises important implications for future mosquito control efforts in DC. The NDVI analysis is supportive in the sense that it is driven by the climate (particularly

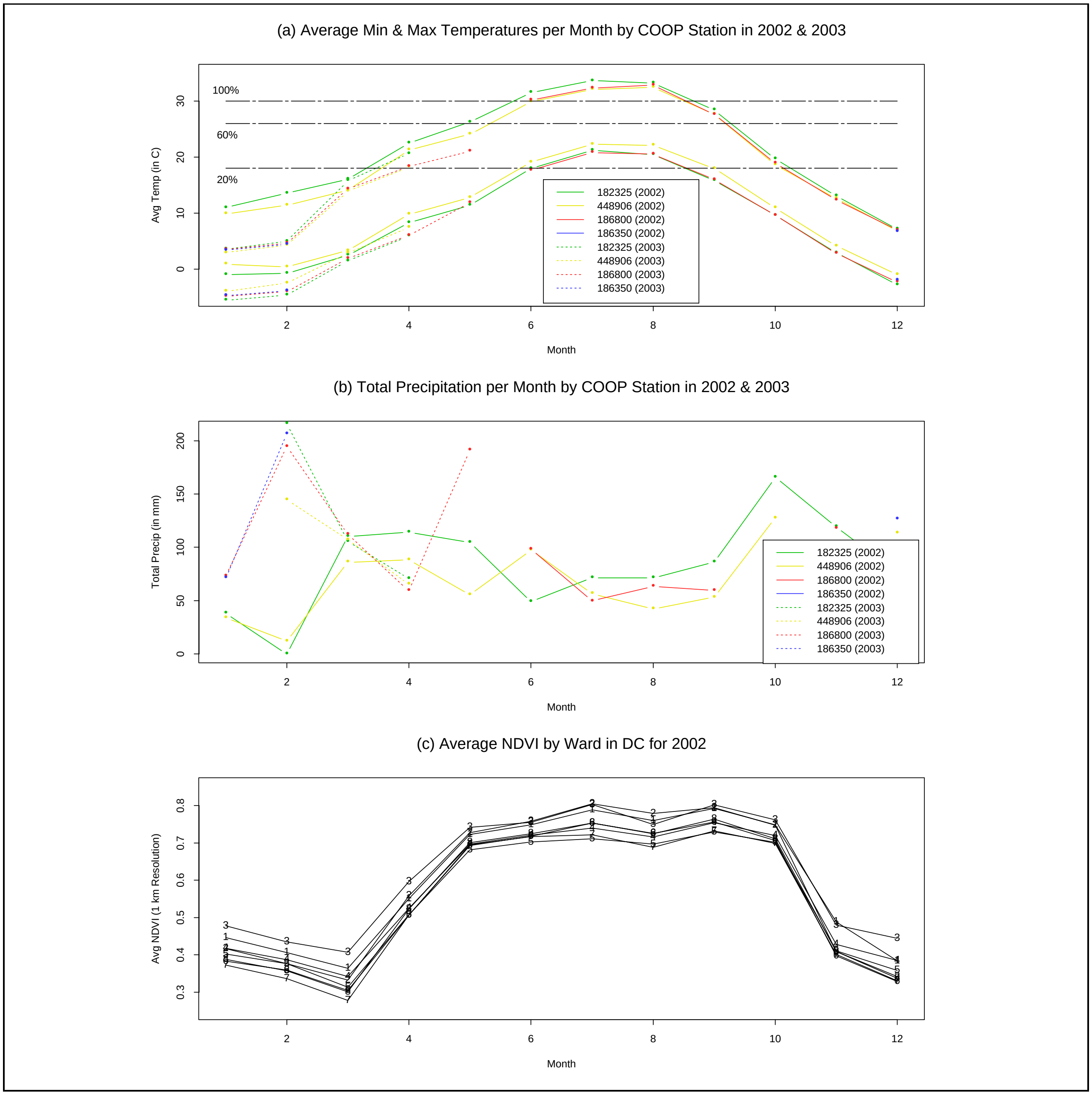

Figure 8: (a) Average min / max temperatures per month, revealing a continuous increase of temperatures from January through July/August. The 2003 min / max temperatures are consistently lower than the 2002 temperatures for the same time period (until May 2003). No further 2003 data is available yet. The dotted percentage lines indicate transmission competency of *Culex pipiens* for WNV based on laboratory data provided by Michael Turell, US Army Medical Research Institute for Infectious Disease. A higher percentage indicates a higher efficiency of the mosquito to transmit the virus. (b) Total precipitation per month for the entire year of 2002 and part of 2003. There is consistent precipitation each month while the monthly precipitation data for 2003 is markedly higher as far as the record goes. (c) Average Normalized Difference Vegetation Index (NDVI) by ward for 2002. The average is computed from the second 10 month composite of SPOT reflectance data. The data is consistent with seasonal green–up and green–down and shows that for certain wards (e.g., Ward 3) the green–up is more elevated, hence the possible establishment of more expansive mosquito breeding habitats.

temperature). Temperature itself is a point measurement provided by cooperative weather stations of which there are only four in the DC area (see Figure 5). NDVI is a spatially distributed variable that reflects the spatial complexity of the temperature data more fully and it can be easily obtained via remote sensing satellites. Quite distinct temperatures and precipitation patterns for the first five months in 2003 may be a possible indicator for a different development of WNV later in 2003.

5. Conclusion

It has been shown that WNV, a spatially and temporally complex phenomenon, can be described well and summarized using geographical and statistical visualization techniques. We have presented various representations of the data that have provided insight into the spread and ecological establishment of WNV. Choropleth maps such as in Figures 1 and 2 effectively show the westward spread of WNV, first in birds, mosquitoes, and other animals, followed by human WNV cases at a lag of a few months up to one or two years. When comparing detailed geographically referenced multivariate statistical information, however micromap plots (see Figure 4) have a clear advantage over choropleth maps (see Figure 3) since micromaps can display more than just one variable and provide exact values and therefore allow a ranking of different geographic regions. The choropleth maps in Figures 6(a) and 6(b) are effective in providing a quick association between regions of large numbers of (WNV–positive) dead birds and the number of human WNV cases. Similar associations between dead birds (not necessarily WNV–positive) and human WNV cases have been reported for New York City (Eidson, Kramer, Stone, Hagiwara, Schmit & The New York State West Nile Virus Avian Surveillance Team 2001, Eidson, Miller, Kramer, Cherry, Hagiwara & The West Nile Virus Bird Mortality Analysis Group 2001, Theophilides, Ahearn, Grady & Merlino 2003). The geographic time series via choropleth maps in Figure 7 allows to display the increase and decrease of WNV–positive pools of mosquitoes over a time span of 12 weeks. The time series in Figure 8 permit an assessment as to which environmental conditions (temperature, precipitation, and NDVI as an aggregate) may effect the outbreak of WNV in different geographic regions. The visualization techniques presented here could be similarly applied in the study of other vectorborne pathogens.

Acknowledgements

We wish to gratefully acknowledge the efforts of Jane Winchester (ISIS Center); Dan Dansereau (Utah State University Climate Database); John O. Davies–Cole, Peggy Keller, and Jamie Hinson (DC Department of Health); Benedict Pagac and Karl Neidhardt (US Army Center for Health Promotion and Preventive Medicine); Michael Turell (US Army Medical Research Institute for Infectious Disease); Compton Tucker and Bob Mahoney (NASA–Goddard Space Flight Center); and Gerald Dittberner (NOAA). We also acknowledge the Centers for Disease Control and Prevention (CDC) for the provision of national WNV data (e.g., `http://www.cdc.gov/ncidod/dvbid/westnile/surv&controlCaseCount02.htm` for the 2002 human data) and the Washington, DC GIS Consortium for the provision of some of the GIS data layers. Population data was obtained from the U.S. Census Bureau Web site (`http://www.census.gov`). The color coding used for the choropleth maps in Figures 1, 2, 3, and 6(a) and 6(b) was obtained from Cindy Brewer's ColorBrewer Web site (`http://colorbrewer.com`). The micromap in Figure 4 is based on a template from Dan Carr (George Mason University).

This work was supported through the Department of Defense Telemedicine and Advanced Technology Research Center (DOD–TATRC) grant DAMD17–94–V–4015. In addition, Symanzik's work was supported in part by the NSF "Digital Government" (NSF 99–103) grant #EIA–9983461 and by a New Faculty Research Grant from the Vice President for Research Office from Utah State University.

References

Carr, D. B. (2001), 'Designing Linked Micromap Plots for States with Many Counties', *Statistics in Medicine* **20**(9–10), 1331–1339.

Carr, D. B., Chen, J., Bell, B. S., Pickle, L. & Zhang, Y. (2002), Interactive Linked Micromap Plots and Dynamically Conditioned Choropleth Maps, *in* 'dg.o2002 Proceedings', Digital Government Research Center (DGRC). `http://www.dgrc.org/conferences/2002_proceedings.jsp`.

Carr, D. B., Olsen, A. R., Courbois, J. P., Pierson, S. M. & Carr, D. A. (1998), 'Linked Micromap Plots: Named and Described', *Statistical Computing and Statistical Graphics Newsletter* **9**(1), 24–32.

Carr, D. B. & Pierson, S. M. (1996), 'Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps', *Statistical Computing and Statistical Graphics Newsletter* **7**(3), 16–23.

Carr, D. B., Wallin, J. F. & Carr, D. A. (2000), 'Two New Templates for Epidemiology Applications: Linked Micromap Plots and Conditioned Choropleth Maps', *Statistics in Medicine* **19**(17–18), 2521–2538.

CDC (2000), 'Notice to Readers: Update: West Nile Virus Isolated from Mosquitoes – New York, 2000', *Morbidity and Mortality Weekly Report (MMWR), Mar 17, 2000,* **49**(10), 211. `http://www.cdc.gov/mmwr/preview/mmwrhtml/mm4910a4.htm`.

Eidson, M., Kramer, L., Stone, W., Hagiwara, Y., Schmit, K. & The New York State West Nile Virus Avian Surveillance Team (2001), 'Dead Bird Surveillance as an Early Warning

System for West Nile Virus', *Emerging Infectious Diseases* **7**(4), 631–635.

Eidson, M., Miller, J., Kramer, L., Cherry, B., Hagiwara, Y. & The West Nile Virus Bird Mortality Analysis Group (2001), 'Dead Crow Densities and Human Cases of West Nile Virus, New York State, 2000', *Emerging Infectious Diseases* **7**(4), 662–664.

Harris, R. L. (1999), *Information Graphics — A Comprehensive Illustrated Reference*, Oxford University Press, New York, NY.

Leslie, M. (2002), 'Tools: A Site for Sore Eyes', *Science* **296**(5567), 435.

MacEachren, A. M. (1995), *How Maps Work — Representation, Visualization, and Design*, The Guilford Press, New York, NY.

Nash, D., Mostashari, F., Fine, A., Miller, J., O'Leary, D., Murray, K., Huang, A., Rosenberg, A., Greenberg, A., Sherman, M., Wong, S. & Layton, M. (2001), 'The Outbreak of West Nile Virus Infection in the New York City Area in 1999', *New England Journal of Medicine* **3441**, 1807–1814.

Symanzik, J., Carr, D. B., Axelrad, D. A., Wang, J., Wong, D. & Woodruff, T. J. (1999), Interactive Tables and Maps — A Glance at EPA's Cumulative Exposure Project Web Page, *in* '1999 Proceedings of the Section on Statistical Graphics', American Statistical Association, Alexandria, VA, pp. 94–99.

Symanzik, J., Hurst, J. & Gunter, L. (2002), Recent Developments for Interactive Statistical Graphics on the Web Using "nViZn", *in* '2002 Proceedings', American Statistical Association, Alexandria, VA. (CD).

Symanzik, J. & Jones, L. (2001), "nViZn" Federal Statistical Data on the Web, *in* '2001 Proceedings', American Statistical Association, Alexandria, VA. (CD).

Theophilides, C. N., Ahearn, S. C., Grady, S. & Merlino, M. (2003), 'Identifying West Nile Virus Risk Areas: The Dynamic Continuous–Area Space–Time System', *American Journal of Epidemiology* **157**(9), 843–854.

Turell, M. (2003), 'Personal communication, Unpublished Data'.

Wang, X., Chen, J. X., Carr, D. B., Bell, B. S. & Pickle, L. W. (2002), 'Geographic Statistics Visualization: Web–based Linked Micromap Plots', *Computing in Science & Engineering* **4**(3), 90–94.