Predictive risk mapping of West Nile virus (WNV) infection in Saskatchewan horses

Tasha Y. Epp, Cheryl Waldner, Olaf Berke

Abstract

The objective of this study was to develop a model using equine data from geographically limited surveillance locations to predict risk categories for West Nile virus (WNV) infection in horses in all geographic locations across the province of Saskatchewan. The province was divided geographically into low-, medium-, or high-risk categories for WNV, based on available serology information from 923 horses obtained through 4 studies of WNV infection in horse populations in Saskatchewan. Discriminant analysis was used to build models using the observed risk of WNV in horses and geographic division-specific environmental data as well as to predict the risk category for all areas, including those beyond the surveillance zones. High-risk areas were indicated by relatively lower rainfall, higher temperatures, and a lower percentage of area covered in trees, water, and wetland. These conditions were most often identified in the southwest corner of the province. Environmental conditions can be used to identify those areas that are at highest risk for WNV. Public health managers could use prediction maps, which are based on animal or human information and developed from annual early season meteorological information, to guide ongoing decisions about when and where to focus intervention strategies for WNV.

Résumé

Cette étude avait comme objectif de développer un modèle utilisant les données provenant de chevaux de localisations géographiques limitées sous surveillance afin de prédire les catégories de risque pour l'infection par le virus du Nil occidental (WNV) chez les chevaux de toutes les localisations géographiques de la province de la Saskatchewan. La province était divisée géographiquement en trois catégories de risque pour le WNV (faible, moyen ou élevé), selon les informations sérologiques provenant de 923 chevaux ayant faits l'objet de 4 études portant sur l'infection par le WNV en Saskatchewan. Une analyse discriminante a été employée pour construire des modèles utilisant le risque observé de WVN chez les chevaux et les données environnementales spécifiques aux divisions géographiques ainsi que de prédire la catégorie de risque pour toutes les régions, incluant celles au-delà des zones de surveillance. Les régions à risque élevé étaient indiquées par des précipitations relativement faibles, des températures plus élevées et un pourcentage plus faible de superficie couverte par des arbres, de l'eau et des marais. Ces conditions étaient le plus souvent identifiées dans la portion sud-ouest de la province. Les conditions environnementales peuvent être utilisées pour identifier les régions qui sont plus à risque pour le WNV. Les gestionnaires de la santé publique pourraient utiliser les cartes de précipitation, qui sont basées sur des informations animales ou humaines et développées à partir d'informations météorologiques annuelles obtenues tôt en saison, pour aider dans la prise de décision continue sur le moment et l'endroit des stratégies d'intervention contre le WNV.

(Traduit par Docteur Serge Messier)

Introduction

The introduction of West Nile virus (WNV) into North America in 1999 sparked interest in predicting where and when the virus would appear next (1,2). New infections appeared to be geographically random, making it impossible to predict the location and timing of individual cases (1). It is possible, however, to identify areas of higher risk using geographical information systems (GIS), remotely sensed data (satellite imagery), ecological variables, and other spatial-analysis techniques (2,3). This approach has been useful in predicting the occurrence of other vector-borne diseases, such as Lyme disease and malaria (2,4).

Vector-borne diseases are particularly amenable to spatial and temporal analysis because they are highly influenced by annual seasonal variations in climate as well as unpredictable changes in climate and in the environment (3). Environmental conditions play a key role in determining the timing and intensity of the WNV cycle. Mosquito populations are especially sensitive to regular seasonal changes in climate and the environment, such as vegetation cover, rainfall, humidity, and temperature (5,6). The extrinsic incubation period, that is the time required from an infectious blood meal until transmission of the virus, is governed by temperature (7). Congregation of mosquitoes and birds, which is essential to the amplification cycle, is influenced by the availability of water sources

Department of Large Animal Clinical Sciences, Western College of Veterinary Medicine, University of Saskatchewan, 52 Campus Drive, Saskatoon, Saskatchewan S7N 5B4 (Epp, Waldner); Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, Ontario N1G 2W1 (Berke).

Address all correspondence to Dr. Tasha Y. Epp; telephone: (306) 966-6542; fax: (306) 966-7159; e-mail: tasha.epp@usask.ca

Received May 10, 2010. Accepted July 20, 2010.

(8). Environmental conditions affect the behavior of humans, which is particularly relevant when they spend time outdoors at peak periods of mosquito activity, such as dusk or dawn (9). These same conditions likely alter the behavior of horses and the humans who manage them. Although the source of WNV introduced in 1999 is not known, favorable conditions existed that allowed it to become established in the local mosquito and bird populations (2,10).

Defining the risk of acquiring infection with WNV is a key component of public health intervention strategies (11). Assessing WNV risk and deciding when and where to implement intervention strategies are largely based on surveillance of vector populations, suitable environmental conditions, and the presence of clinical disease in various host species. From the perspective of horse owners, knowing the risk of WNV in their area is a contributing factor to deciding whether or not to vaccinate. Surveillance of vectors and hosts over large geographic areas, however, can only produce maps with discontinuous patterns. Intervention decisions based on these maps alone require assumptions to be made about risk in the unsampled areas.

Predictive risk mapping is a 2-stage process in which components of the disease cycle (epidemiological, environmental, and/or entomological) are first used to build models. In the second step, predictions are made about the probability of membership within defined risk categories for all geographic sub-regions of the study area, including those without existing surveillance data (3,12). Methods have become more practical and are applied to a broader range of diseases and study locations because remote sensing can now provide environmental information at the necessary spatial and temporal resolution (13,14). It is critical, however, to use appropriate geographic resolution, variables of interest, and methodologies cautiously for sound investigation of the stated hypotheses (3,13).

In 2002, West Nile virus (WNV) was identified in birds and horses in Saskatchewan (SK). A series of 4 observational studies were undertaken in 2003 to monitor the progress of this emerging public health issue. This paper describes how we used these data to then define and model areas of low, medium, and high risk of WNV infection for horses in SK. We then assessed the model's predictive ability using 2005 data.

Materials and methods

Proportion of tested horses infected with West Nile virus

Information on the serological status of horses in Saskatchewan was compiled and summarized at the rural municipality (RM) level. The study area consisted of 297 RMs in total. In the summer of 2003, blood samples were collected from 923 horses as part of a series of studies to determine factors affecting the risk of infection, clinical disease, and case fatality. For the first study (case fatality), data from all 133 clinical cases were summarized, with the local laboratory reporting confirmed cases of WNV infection to the research team (15). A second study (risk of infection) randomly sampled horses from 20 herds across Saskatchewan at 2 separate time periods to determine serological status (16). Data were also obtained from a third case-control study of more than 300 horses. The 23 selected case farms reflected a subset of the herds with clinical cases identified in the first study (17). Finally, serological results were summarized from a fourth rapid-assessment study in which researchers sampled horses in the area surrounding 7 cities that were selected based on the locations of positive mosquito traps (18).

The blood samples were tested by both in-house immunoglobulin IgG and IgM enzyme-linked immunosorbent assay (ELISA) at Prairie Diagnostic Services (16). The IgM assay assesses recent infection/exposure (reported as a yes or no), while the IgG assay measures antibody concentrations at a point in time (reported as a sample-to-positive ratio). The horses were defined as infected or not infected based on the ELISA results and information from the accompanying studies (15–17). We determined the proportion of tested horses infected in each RM where 5 or more horses were sampled, then classified these RMs by category of risk of infection. The categories were established based on division at the 25th and 75th percentiles of WNV seroprevalences: low (0% to 29.9%), medium (30% to 77.7%), and high risk (77.8% to 100%).

In 2005, 196 horses were sampled and tested for WNV antibodies with an IgM ELISA only. Due to the lack of information on vaccination status, IgG data were not useful. Ten horses showed clinical signs of WNV. The rest of the horses were sampled through cooperation with veterinarians already visiting the farms for other reasons. As for the 2003 data, the proportion of infected horses in 2005 was estimated for each RM where 5 or more horses had been sampled. The same 3 categories of risk of infection (low, medium, and high) that were reported for 2003 were used to label the RMs sampled in 2005.

Predictor variables for WNV risk

Land surface temperature (LST) was obtained from the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) Gateway for the Moderate Resolution Imaging Spectrometer (MODIS) satellite. (These data are distributed by the Land Process Distributed Active Archive Center at the United States Geological Survey Center for Earth Resources Observation and Science and are available at http://lpdaac.usgs.gov). The images were joined together (mosaiced) and clipped to show only the province of SK (PCI Geomatica 9; PCI Geomatics, Richmond Hill, Ontario). Seventeen images were obtained as 8-day composites beginning on May 1 and ending on September 13 for 2003 and 2005. The images included maximum daytime and minimum night-time temperatures and were manipulated to give a mean LST, which was an average of the daytime and night-time images. The average value per RM was calculated for each time period for all 3 temperature variables for both 2003 and 2005.

Daily minimum (min), maximum (max), and mean temperature (°C) for both 2003 and 2005 were also obtained from Environment Canada climate stations. Eight-day composites (average min, mean, or max temperature per time period) were created to match the time periods of the remotely sensed data. For each time period, stations with data missing for ≥ 50% of the days were recorded as missing. Interpolation between stations was accomplished with inverse distance weighting (IDW) (ArcGIS 9; esri, Redlands, California, USA). Inverse distance weighting is a quick deterministic interpolation of the data with minimal input of parameters (19). For each time period in both 2003 and 2005, the average value per RM was calculated for all 3 variables.

Cumulative growing degree days (GDDs) are the sum of the positive daily differences between 14.3°C (the threshold for activity of *Culex tarsalis* and virus transmission) and the mean temperature for that day (20). The 9 time periods used in analysis were defined as the cumulative GDD from May 1 until the beginning or middle of each month, so that time period 1 was from May 1 until May 15, while time period 9 was from May 1 until September 15. Values for GDDs were recorded as unknown at and after a time period during which data were missing from that station. Values for GDDs were calculated for each climate station for 2003 and 2005. Interpolation between stations with complete data for each time point was accomplished using the IDW method (ArcGIS 9). The average value per RM was calculated for each time period in both 2003 and 2005.

Daily precipitation values (mm) for 2003 and 2005 were obtained from Environment Canada. Again, 8-d composites (total precipitation per time period) were created to match the remotely sensed time periods. Interpolation between stations was accomplished using the IDW method (ArcGIS 9). The average total precipitation value for each RM was calculated for each time period in both 2003 and 2005.

The normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) were obtained from EOS Gateway for the MODIS satellite. The NDVI is an index of vegetation cover based on near infrared and red reflectance bands of satellite imagery, which facilitates monitoring of seasonal and inter-annual changes in vegetation growth (21). The EVI is a modified NDVI that incorporates a soil-adjustment factor and correction for atmospheric scattering. It is more sensitive for areas of dense vegetation (21). Starting on April 23 and ending September 13 for both 2003 and 2005, 9 images were obtained as 16-d composites. The average value per RM was calculated for each time period (PCI Geomatica 9).

The South Digital Land Cover dataset (North Digital and South Digital Land Cover) based on satellite imagery from 2000 for the province of Saskatchewan was obtained from Information Services Corporation of Saskatchewan. Detailed classifications were aggregated to form the following categories: water, wetland (combined bog and marsh), treed (pine, spruce, hardwood, softwood), ground (rock, recent burn, cutovers, barren land), agriculture (cropland, pasture), and urban (settlements, roads). The percentage of each rural municipality (RM) covered by each of the categories was calculated (Geomatica 9).

The digital elevation model (DEM) obtained from the NASA Jet Propulsion Laboratory-Shuttle Radar Topography Mission was based on satellite images from 2000. The mean, majority, min, max, and range of elevation were calculated for each RM (Geomatica 9). The number of horses per rural municipality (RM) was obtained from Statistics Canada using 2001 Consolidated Census Subdivisions as a proxy for RM. The latitude and longitude of the centroid of each RM was extracted from RM shapefiles (ArcGIS 9). Ecoregions are a subdivision of ecological zones characterized by regional ecological factors such as vegetation, soil, climate, water, and fauna (22). There are 6 ecoregions within the geographic area covered by the RMs in the southern half of the province: Cypress Uplands, Mixed Grasslands, Moist Mixed Grasslands, Aspen Parkland, Boreal Transition, and Mid-boreal Uplands. Each RM was assigned the ecoregion that made up the largest proportion of land area (ArcGIS 9).

Statistical analysis (2003 data)

We used discriminant analysis to predict membership in the 3 mutually exclusive risk categories (low, medium, and high) for a large geographic area where surveillance information was neither consistent nor uniformly distributed (23) (SPSS 14.0; SPSS, Chicago, Illinois, USA). Discriminant analysis is a technique that can classify a set of individual observations into 2 or more predefined categories and determine which set of variables best predicts category membership (23). The process begins with a set of observations for which both the predefined category membership and the predictor variables are known. The end result is a model that predicts category membership for a set of observations for which only the predictor variables are known. Assessment of the final models provides the overall predictive ability of the model, by variable combinations, by category, or by individual observation.

The 2003 dataset was divided into a) a training dataset, which consisted of 53 RMs selected using random number generation from the 67 total sampled RMs, and b) a testing dataset, which consisted of the remaining 14 sampled RMs and the 231 unsampled RMs. The training data were used to develop the model by estimating the differences in environmental factors between areas sampled for West Nile virus (WNV). The testing data were used to test the accuracy of the model for prediction in unsampled areas.

Two groupings of models were created: early-season (evaluating environmental variables from May 1 to July 11) and whole-season (evaluating variables from May 1 to September 13). Early-season models were developed to assess the prediction accuracy of this approach during the season under analysis, while whole-season data were examined to assess the overall prediction accuracy of the model as well as to provide a baseline against which to compare the early-season model.

All variables and all time periods were introduced individually into the model for univariate analysis. Variables and time periods with a 2-sided $P > 0.10$ were excluded from any further analysis. To meet the assumption that variables are not highly correlated with each other, each variable with multiple statistically significant time periods by the univariate analysis was reduced using principal component analysis (SPSS 14.0). This condenses the multiple statistically significant time points into 1 or 2 principal components for consideration when developing final models (24). The temperature-related variables (GDD, temperature, and LST) were also highly correlated with each other and were, therefore, used individually as the basis for separate models for both early and whole season.

The final selection of variables for each model was determined by maximizing the percentage of RMs with correctly classified risk categories for both the training and testing datasets (23). In addition, the multivariate models were fit with unequal weighting schemes to account for the pre-model knowledge of the risk category of each sampled RM. For example, because > 50% of the sampled RM observations were in the medium-risk category, medium risk was given more weight in the analysis than the other 2 categories. Separate matrices (to account for unequal group covariance matrices) were used when Box's M test was significant ($P < 0.05$) and the model's

Table I. Variables (time periods where $P < 0.05$) tested in multivariable temperature-based models for both early- and whole-season predictions

Timeline	Model	Mean LST[a]	Mean temperature[a]	GDD	Precipitation	NDVI	Other time-constant variables
Early season	**1**	**June 26 to July 11**	NA	NA	**June 2 to 25**	**April 23 to May 8**	**DEM (mean), Latitude, Ecoregion, Land Cover[b]**
	2	NA	July 4 to 11	NA	June 2 to 25	April 23 to May 8	DEM (mean), Latitude, Ecoregion, Land Cover[b]
	3	NA	NA	May 1 to June 15	June 2 to 25	April 23 to May 8	DEM (mean), Latitude, Ecoregion, Land Cover[b]
Whole season	**1**	**June 26 to Sept 13**	NA	NA	**June 2 to 25 July 4 to 19**	**Aug 13 to Sept. 13**	**DEM (mean), Latitude, Ecoregion, Land Cover[b]**
	2	NA	July 4 to Sept. 5	NA	June 2 to 25 July 4 to 19	Aug 13 to Sept. 13	DEM (mean), Latitude, Ecoregion, Land Cover[b]
	3	NA	NA	May 1 to Sept. 15	June 2 to 25 July 4 to 19	Aug 13 to Sept. 13	DEM (mean), Latitude, Ecoregion, Land Cover[b]

Note: Entries in bold indicate the final prediction models for both the early and whole season.

[a] Remotely sensed time period (8-day composite). These underwent principal component analysis to be included in the model when more than 1 time period was significant.

[b] Land cover included percentages of trees, water, and wetland.

LST — Land surface temperature; GDD — Growing degree days; bimonthly time period. Cumulative GDD from May 1 until the end of the time period indicated; NDVI — Normalized difference vegetation index. Remotely sensed time period (16-d composites). These underwent principal component analysis to be included in the model when more than 1 time period was significant; DEM — Digital elevation model; NA — Not applicable (refers to variable not used in that model's analysis).

percentage of prediction accuracy changed substantially from a model that used a common matrix for all risk categories. Ultimately, the final multivariate models for both early and whole season were selected to produce the best overall classification accuracy with the least reclassification of risk category between high and low risk of the sampled RMs.

The discriminant analysis yielded 2 primary results. First, the analysis provided functions or sets of variables that distinguished among risk categories. The analysis also supplied information about the extent to which each of the functions contributed to the classification and which variables within the functions were the most powerful. Second, the analysis delivered a set of 3 probabilities that predicted the likelihood of membership in each of the 3 risk categories (23). Individual RMs were classified into 1 of the risk categories by determining the highest probability of risk category membership. For example, a RM was classified as high risk if the probability of being in the high-risk category as determined by the model was greater than for either the low- or medium-risk categories. The categorization strategy was most reliable for RMs where the most likely risk category had a maximum probability of $\geq 75\%$. When the probability was $< 75\%$, the next most likely category, based on the next highest probability, was assessed for each RM and an individual determination made regarding whether the categorization was valid.

The overall average probability of risk category membership for each risk category was maximized in the multivariate model selection process. Chloropleth maps were generated of the predicted risk categories from the best final early- and whole-season models for 2003 (ArcGIS 9). A chloropleth map of the probability associated with the predicted risk category membership was also generated for the early-season model for 2003 (ArcGIS 9).

Statistical analysis (2005 data)

To assess the model's prediction capability in a different year, the training dataset for 2003 (53 sampled RMs with the corresponding

Table II. Early- and whole-season final models
Model results from final multivariable early- and whole-season models for 2003 and 2005

		Year			
		2003		2005	
Model		Early-season	Whole-season	Early-season	Whole-season
RMs[a] sample[d]	Training	53 sampled RMs (observed categorization: 12 low-risk, 26 medium-risk, 15 high-risk)		Same as 2003	
	Testing	14 sampled RMs (observed categorization: 4 low-risk, 7 medium-risk, 3 high-risk) Plus remaining 231 unsampled RMs		10 sampled RMs (observed categorization: all low-risk) Plus remaining 287 unsampled RMs	
Model accuracy	Training	74% (39/53)	70% (37/53)	N/A[b]	NA
	Testing	64% (9/14)	71% (10/14)	10% (1/10)	30% (3/10)
Significant variablesa[ac]	Function 1	Precipitation Mean LST[d] Mean DEM[d]	Precipitation Water coverage Mean LST NDVI	Precipitation Mean LST Mean DEM	Precipitation Water coverage Mean LST NDVI
	Function 2	Water coverage Tree coverage Wetland coverage NDVI[d]	Wetland coverage Tree coverage	Water coverage Tree coverage Wetland coverage NDVI	Wetland coverage Tree coverage
Risk category membership probability	Low-risk	88%	85%	89%	78%
	Medium-risk	69%	67%	69%	68%
	High-risk	74%	71%	79%	74%

[a] RMs — rural municipalities.
[b] NA — not applicable.
[ac] Variables that were significant in the model are recorded by function in the order of importance for contributing to the function.
RMs — rural municipalities; NA — Not available; [d] LST — land surface temperature; DEM — digital elevation model; NDVI — normalized difference vegetation index.

2003 environmental information) was used to train the model in the known differences between sampled areas of different risk categories because the 2005 dataset only had representation from the low-risk category. The 2005 dataset (10 sampled RMs and 288 unsampled RMs) was then analyzed by the trained model using 2005 environmental data to produce both the 2005 early- and whole-season models. Chloropleth maps of the predicted risk categories from both the early- and whole-season models for 2005 were generated (ArcGIS 9).

Results

Distribution of infection in horses

In 2003, 133 of the 923 horses tested for WNV were clinical cases. The other 790 horses were asymptomatic, of which 395 (50%) were WNV seropositive. When aggregated by rural municipality (RM), horses were sampled in 117 RMs, of which 67 RMs had ≥ 5 horses sampled and were thus used in the analysis.

In 2005, only 15 of the 196 horses sampled in the province were seropositive. The seropositive horses were detected in 13 RMs, in which there were 7 seropositive horses in 2003, 5 were adjacent to RMs with seropositive horses in 2003, and 1 was not sampled in 2003.

Model building — 2003

The variables selected for consideration in the final multivariate models, based on $P < 0.10$ in univariate models, were: GDD, mean temperature, mean LST, land cover (treed, water, and wetland only), NDVI, mean DEM, and precipitation (Table I). The 67 sampled RMs used in the analysis were classified as: 16 low risk, 33 medium risk, and 18 high risk. Using random selection, 80% of these RMs were selected for the training dataset (53 RMs) and the rest were used as the testing dataset (Table II).

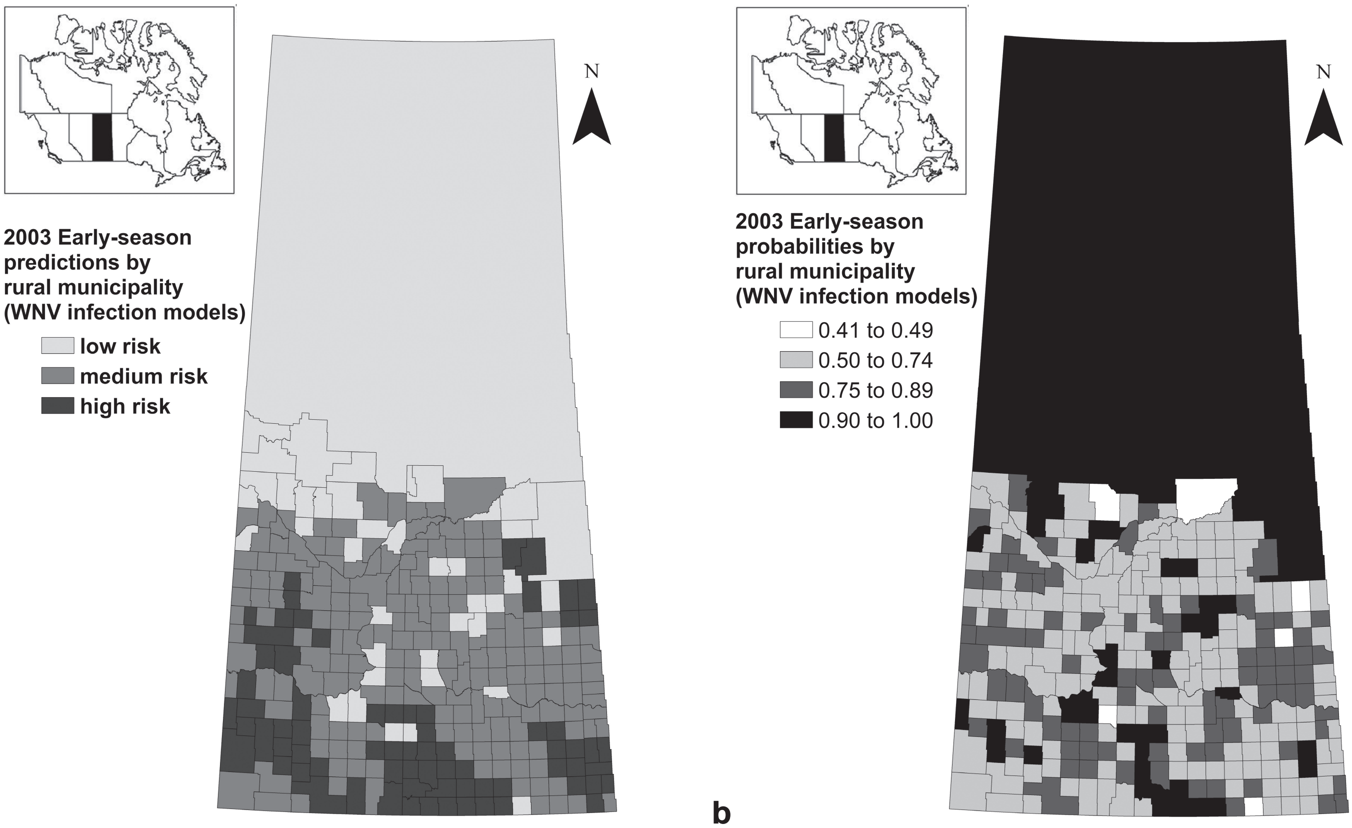

Figure 1. Results of the 2003 early-season model (April 23 to July 11). a — predicted risk category membership (low, medium, or high risk of WNV infection), and b — probability of risk category membership.

The best early-season multivariate model included mean LST, precipitation, land cover, mean DEM, and NDVI. Comparison of variables by risk category indicated that, for the first function, decreasing rainfall throughout June was the most important indicator for high-risk category membership, followed by higher mean LST in July, and higher mean DEM. In the second function, lower percentages of treed, water, and wetland areas in the RM and slightly lower NDVI-defined membership in the high-risk category. The risk category predictions and the probability associated with that predicted risk category by RM were mapped (Figures 1a and 1b).

The best whole-season model included mean LST, precipitation, land cover, and NDVI. Comparison of variables by risk category showed that, for the first function, higher rainfall in June with lower rainfall in July was the most important indicator of high-risk category membership followed by higher mean LST in July–September, lower NDVI at the end of August/beginning of September, and lower percentage of water area per RM. In the second function, lower percentages of treed and wetland areas per RM defined high-risk category membership. The risk category predictions by RM were mapped (Figure 2).

Predictive ability — 2005

The training dataset used for both the 2005 early-season and whole-season models was the training dataset for 2003 (53 sampled RMs). The testing dataset included 10 sampled RMs from 2005 and the remaining 287 unsampled RMs.

For the early-season model, the testing classification accuracy was 10%, indicating that 9 out of the 10 RMs sampled in 2005 were re-classified based on the trained model. When variables were compared by function and by risk category, the pattern was similar to the corresponding 2003 model. The risk category predictions by RM were mapped (Figure 3).

For the whole-season model, the testing classification accuracy was 30%. When variables were compared by function and by risk category, the pattern was similar to the corresponding 2003 model. The risk category predictions by RM were mapped (Figure 4).

Discussion

Predictive models were developed to maximize the accuracy of distinctions between high- and low-risk rural municipalities (RMs). These models were based on serological data and provided approximately 70% to 74% overall model accuracy in 2003. In the 2005 model process, the testing overall classification accuracy percentages were not considered indicative of overall model accuracy due to the small sample size, consisting of only low-risk RMs.

Our models used serological evidence and not clinical case status because the development of clinical signs could be affected by variations in the use of vaccination in different regions of the province (17). As expected, the complexity of the WNV cycle was not completely explained by the environmental variables available for use in these models. Factors such as human intervention, biodiversity,

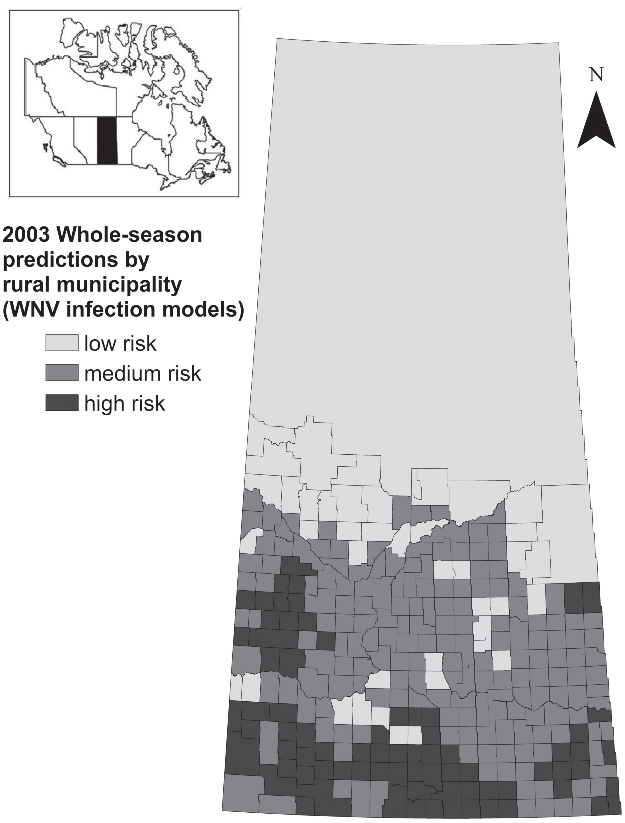

Figure 2. Predicted risk category membership (low, medium, or high risk of WNV infection) for the 2003 whole-season model (May 1 to September 13).

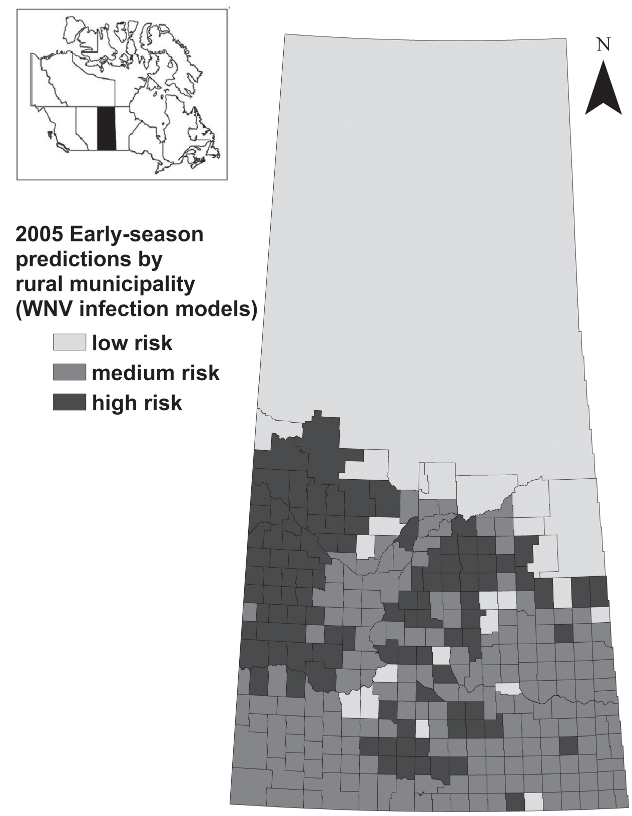

Figure 3. Predicted risk category membership (low, medium, or high risk of infection) for the 2005 early-season model (April 23 to July 11).

predators, parasites, food availability, and spatial resources will affect interactions between the vector and hosts. The immune status of the hosts will become more important the longer the virus remains in an area (11).

Precipitation was the most influential predictor of risk of infection. Precipitation in June was important in the early-season model and precipitation in June and July was important in the whole-season model. The life cycles of mosquitoes, encompassing measures of abundance and demographic composition, and breeding site habitat, encompassing the presence of suitable habitat, size, and persistence, are both influenced by rainfall (6,25). In Florida, USA and elsewhere, drought-like conditions in spring, followed by increased rainfall patterns in summer and fall make it easier for WNV to spread (26). Different species of mosquitoes will react differently to patterns of rainfall (25). *Culex tarsalis* prefers rainfall followed by hot and dry conditions because it uses standing water with increased organic content for oviposition (25). This water would be washed away by rainfall. Both prediction models emphasize that higher rainfall at the beginning of June, decreasing towards the end of June, with minimal rainfall in July is associated with high-risk, whereas lower rainfall in June than in July indicates low risk.

Another important variable was mean land surface temperature (LST) for July in the early-season model and for July and August in the whole-season model. Increasing temperatures are linked to increased survival of mosquitoes, increased biting habits, and a decreased extrinsic incubation period (6,7). The effect of temperature on the transmission of WNV was evident in the Central Red River Valley of North Dakota and Minnesota, USA, close to Saskatchewan, where substantially fewer human cases of WNV were reported in 2004 than in 2003; temperatures, summarized as the amount of thermal accumulations or degree days, were almost half of previous years (27,28). A similar pattern of no confirmed horse cases and a colder, wetter climate was observed in Saskatchewan during 2004 (29). In the current prediction models, higher rainfall in spring (June), followed by higher temperatures in July and August were predictive of high-risk areas in both the early- and whole-season models.

Different temperature-related variables were evaluated in developing these models. Mean LST was a slightly better predictor than either mean temperature or growing degree days (GDDs). This is likely because the data obtained from the weather station required interpolation to provide a continuous map from which average data for each RM could be calculated, whereas the satellite data were provided on a continuous 500-m scale. Studies show that LST (obtained on an 8×8 km or even 1×1 km basis) and ground-derived ambient temperatures are quite similar, especially if other variables such as NDVI or latitude were included in the analysis (30,31). Land surface temperature (LST), therefore, has several advantages over ground-derived temperature data for large areas (30).

The amount and type of vegetation in an area is influenced by both temperature and precipitation (5). Foci for WNV could potentially be identified based on the presence of favorable mosquito

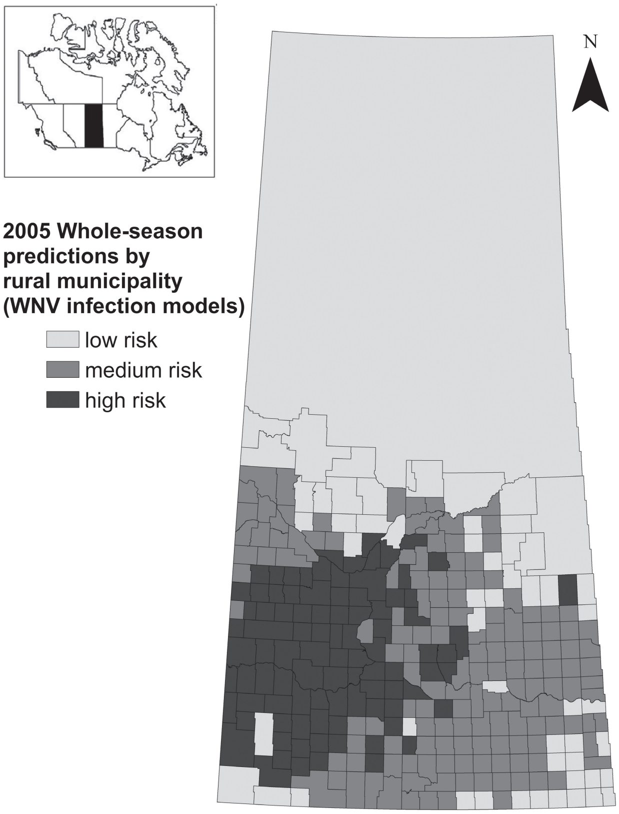

Figure 4. Predicted risk category membership (low, medium, or high risk of infection) for the 2005 whole-season model (May 1 to September 13).

habitat, which is specifically determined by water and vegetation (32). A study in northern Indiana examined the association between NDVI and the incidence of equine WNV cases and demonstrated that NDVI was potentially useful for predicting hot spots for implementation of mosquito-control measures, but its usefulness was limited by the amount of agriculture in the area (33). In our study, NDVI from the first week in May (early season) or from the end of August and beginning of September (whole season) were important predictors. The NDVI from the beginning of the season could indicate the amount of precipitation in the previous fall and winter, especially in a year when warmer spring conditions facilitated earlier growth. The NDVI from the end of August and beginning of September would indicate amounts of vegetation without a lot of agricultural influences because harvest would be well underway at this time in an average year. Ward et al. hypothesize that NDVI has a better ability to identify small areas where conditions are still suitable for vector habitat once harvest is under way (33).

Land cover data provided additional information for the final model. The percentage of land area covered by trees, water, or wetland areas differed between rural municipalities (RMs) that were high risk (low percentages) and low risk (high percentages). Based on the habitat preferences of *C. tarsalis,* RMs with high wetland and treed areas would not be considered ideal locations for WNV infection. Elevation was a useful variable in the early-season model, with municipalities with higher average elevation more likely to be classified as high risk. In south-central Saskatchewan in 2003, the human cases of WNV followed the higher elevation of the Missouri Couteau, where the wetlands are concentrated in the low-lying valleys and low-depression areas (Curry P. Saskatchewan Health, personal communication, 2005).

The predictive ability of the model for the 2005 data was compromised because all the RMs surveyed were initially classified as low risk. This was due both to a low number of horses sampled and because assessment of infection status was based solely on the IgM assay. This led to an inability to use the sampled 2005 RMs as a means to test the overall accuracy of the model (training or testing datasets) or to assess the importance of predictor variables. The risk predicted by environmental variables, however, as indicated by the maps seems to indicate a geographic shift in the high-risk areas from 2003. The predicted area of high risk in the western portion of the study area (in both early- and whole-season models) matches the area where the highest number of *C. tarsalis* mosquitoes was trapped in 2005 (Curry P. Saskatchewan Health, personal communication, 2005).

On the other hand, the RMs with positive cases in 2005 were the same as those in 2003. This indicates that, although the environmental risk of WNV infection changes on a yearly basis, WNV activity and the spatial distribution of cases might not. This has been observed with other spatial analyses of cases of WNV in birds and humans in Ontario (34). In addition, the area with the most positive pools of mosquitoes was in the southeastern corner, which did not correspond to the model predictions. The complex relationship between environmental risk, mosquito abundance and infectivity, and actual transmission of the virus to various host species cannot be adequately summarized by a predictive model based on 2 y of limited data from 1 species. The model is missing key information on the amplification cycle that would likely increase the accuracy of predicting the risk of infection.

There were only minor differences in the predictions made by the early-season versus the whole-season models. Environmental conditions during the early-season months (April, May, June, and early July) are the primary drivers for the development of the first and second generations of *C. tarsalis* in Saskatchewan (35). Therefore, in the absence of dramatic changes in environmental trends later in the season, model predictions for the early season should provide useful information within any given year.

Finally, the models were created with a training set of data consisting of a randomly selected portion of the RMs where horses were sampled in 2003. A different randomly selected training dataset could have produced slightly different results. In addition, these models relied on summarized and interpolated environmental data for each RM, which can produce a loss of inherent heterogeneity in the data and introduce errors. Caution should therefore be exercised when referring to predictions for individual RMs based on the maps. Instead, the maps will be useful to indicate general but larger areas of high risk of infection.

Our model might be useful as 1 component of an overall surveillance initiative. For example, an indication of environmental conditions conducive to WNV would be useful in determining sites for intensified surveillance or control of mosquitoes. By incorporating data from multiple sources of species, for example mosquito trapping or bird deaths, model predictions could become more stable and accurate. The distribution of cases in subsequent years indicates that

WNV infection follows a more spatially conservative and consistent distribution than one that solely follows environmental risk. More information on the connection between high environmental risk and WNV infection could be obtained by using the models in conjunction with other surveillance methods, particularly all mosquito data.

These models have identified time periods and variables to use when considering the contribution of environmental factors to WNV activity. The best predictors of high-risk areas on a yearly basis were precipitation and temperature, while the more static habitat variables (percentages of trees, water, or wetland and NDVI) were less important but still necessary for overall prediction accuracy. The geographical distribution of individual high-risk areas seems to change and does not necessarily follow the more conservative pattern of where clinical disease manifests itself in horses on a yearly basis. This model is a first step towards explaining a complex situation. The results would therefore be best used to augment other data collection. Future incursions of mosquito-borne diseases will benefit from the knowledge gained from ongoing targeted surveillance.

Acknowledgments

The authors gratefully acknowledge funding and support from a Western College of Veterinary Medicine Interprovincial Fellowship and from Wyeth Animal Health and the Equine Health Research Fund.

We also acknowledge Phillip Curry, Provincial West Nile Virus Coordinator at Saskatchewan Health, for his contributions to the study as well as for having the foresight to incorporate veterinary data into the provincial human surveillance initiative for West Nile virus.

References

1. West Nile Virus in Equids in Northeastern United States in 2000. United States Department of Agriculture. Available from http://www.aphis.usda.gov/vs/ceah/wnvreport.pdf Last accessed May 4, 2011.
2. Rogers DJ, Myers MF, Tucker CJ, et al. Predicting the distribution of West Nile fever in North America using satellite sensor data. Photogramm Eng Remote Sensing 2002;68:112–114.
3. Kitron U. Risk Maps: Transmission and burden of vector-borne diseases. Parasitology Today 2000;16:324–325.
4. Dister SW, Fish D, Bros SM, Frank DH, Wood BL. Landscape characterization of peridomestic risk for Lyme disease using satellite imagery. Am J Trop Med Hyg 1997;57:687–692.
5. Hay SI, Tucker CJ, Rogers DJ, Packer MJ. Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vectors of disease. Ann Trop Med Parasitol 1996;90:1–19.
6. Mellor PS, Leake CJ. Climatic and geographic influences on arboviral infections and vectors. Rev Sci Tech Off Int Epiz 2000; 19:41–54.
7. Reisen WK, Fang Y, Martinez VM. Effects of temperature on the transmission of West Nile virus by *Culex tarsalis* (Diptera: Culicidae). J Med Ent 2006;43:309–317.
8. Kramer LD, Bernard KA. West Nile virus in the western hemisphere. Curr Op Infect Dis 2001;14:519–525.
9. Loeb M, Elliott SJ, Gibson B, et al. Protective behavior and West Nile virus risk. Emerg Infec Dis 2005;11:1433–1436.
10. Epstein PR. West Nile virus and the climate. J Urban Health 2001;78:367–371.
11. Rainham DGC. Ecological complexity and West Nile virus: Perspectives on improving public health response. Can J Pub Health 2005;96:37–40.
12. Brooker S, Hay SI, Bundy DAP. Tools from ecology: Useful for evaluating infection risk models? Trends Parasitol 2002;18:70–74.
13. Beck LR, Lobitz BM, Wood BL. Remote sensing and human health: New sensors and new opportunities. Emerg Infect Dis 2000;6:217–226.
14. Rogers DJ, Randolph SE, Snow RW, Hay SI. Satellite imagery in the study and forecast of malaria. Nature 2002;415:710–715.
15. Epp T, Waldner C, West K, Townsend H. Factors associated with West Nile virus disease fatalities in horses. Can Vet J 2007; 48:1137–1145.
16. Epp T, Waldner C, Leighton FA, Berke O, Townsend HHG. Serologic prevalence and risk factors for infection with West Nile virus in Saskatchewan horses, 2003. Can J Vet Res 2007; 71:256–263.
17. Epp T, Waldner C, Townsend HHG. A case-control study of factors associated with development of clinical disease due to West Nile virus, Saskatchewan, 2003. Eq Vet J 2007;39:498–503.
18. Corrigan RLA. Prediction multi-species evaluation of human risk of West Nile virus infection, Saskatchewan, 2003 [Masters thesis]. Saskatoon, Saskatchewan: University of Saskatchewan, 2005.
19. Waller LA, Gotway CA. Interpolation and spatial prediction. In: Applied Spatial Statistics for Public Health Data. Hoboken, New Jersey: John Wiley and Sons, 2004:299–302.
20. Cornell Department of Horticulture Web site. (New York) Growing Degree-Day Tracker: Explanation of Growing Degree days. Available from http://www.entomology.cornell.edu/Extension/Woodys/GrowingDegreeDays.html Last accessed May 4, 2011.
21. Jensen JR. Vegetation indices. In: Introductory Digital Image Processing. 3rd ed. New Jersey: Pearson Prentice Hall, 2005: 310–322.
22. A National Ecological Framework for Canada: Overview. Agriculture and Agri-Food Canada Web site. Available from http://sis.agr.gc.ca/cansis/nsdb/ecostrat/intro.html Last accessed May 4, 2011.
23. Klecka WR. Discriminant Analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-019. Beverly Hills, California: Sage Publications,1980.
24. Dohoo I, Martin W, Stryhn H. Model-building strategies. In: Veterinary Epidemiologic Research. 1st ed. Charlottetown: AVC Inc., 2003:322.
25. Shaman J, Day JF. Achieving operational hydrologic monitoring of mosquito borne disease. Emerg Infect Dis 2005;11:1343–1350.
26. Shaman J, Day JF, Stieglitz M. Drought-induced amplification and epidemic transmission of West Nile virus in southern Florida. J Med Ent 2005;42:134–141.

27. Bell JA, Mickelson NJ, Vaughan JA. West Nile virus in host-seeking mosquitoes within a residential neighborhood in Grand Forks, North Dakota. Vect Borne Zoon Dis 2005;5:373–382.
28. Bell JA, Brewer CM, Mickelson NJ, Garman GW, Vaughan JA. West Nile virus epizootiology, Central Red River Valley, North Dakota and Minnesota, 2002–2005. Emerg Infect Dis 2006; 12:1245–1247.
29. Epp T, Waldner C, Corrigan R, Curry P. Public health use of surveillance for West Nile virus in horses: Saskatchewan, 2003–2005. Trans Emerg Dis 2008;55:411–416.
30. Hay SI, Lennon JJ. Deriving meteorological variables across Africa for the study and control of vector-borne disease: A comparison of remote sensing and spatial interpolation of climate. J Trop Med Hyg 1999;4:58–71.
31. Green RM, Hay SI. The potential of Pathfinder AVHRR data for providing surrogate climatic variables across Africa and Europe for epidemiological applications. Remote Sens Envir 2002; 79:166–175.
32. White DJ. Vector surveillance for West Nile virus. Ann NY Acad Sci 2001;951:74–82.
33. Ward MP, Ramsay BH, Gallo K. Rural cases of equine West Nile virus encephalomyelitis and the normalized difference vegetation index. Vect Borne Zoon Dis 2005;5:181–188.
34. Beroll H, Berke O, Wilson J, Barker IK. Investigating the spatial risk distribution of West Nile virus disease in birds and humans in southern Ontario from 2002 to 2005. Pop Health Metrics 2007;5:3.
35. Curry P. Saskatchewan mosquitoes and West Nile virus. Blue Jay 2004;62:104–111.