Research Paper

Modeling the Spatial Distribution of Mosquito Vectors for West Nile Virus in Connecticut, USA

MARIA A. DIUK-WASSER,[1] HEIDI E. BROWN,[1] THEODORE G. ANDREADIS,[2]
and DURLAND FISH[1]

ABSTRACT

The risk of transmission of West Nile virus (WNV) to humans is associated with the density of infected vector mosquitoes in a given area. Current technology for estimating vector distribution and abundance is primarily based on Centers for Disease Control and Prevention (CDC) light trap collections, which provide only point data. In order to estimate mosquito abundance in areas not sampled by traps, we developed logistic regression models for five mosquito species implicated as the most likely vectors of WNV in Connecticut. Using data from 32 traps in Fairfield County from 2001 to 2003, the models were developed to predict high and low abundance for every 30 × 30 m pixel in the County. They were then tested with an independent dataset from 16 traps in adjacent New Haven County. Environmental predictors of abundance were extracted from remotely sensed data. The best predictive models included non-forested areas for *Culex pipiens*, surface water and distance to estuaries for *Cx. salinarius*, surface water and grasslands/agriculture for *Aedes vexans* and seasonal difference in the normalized difference vegetation index and distance to palustrine habitats for *Culiseta melanura*. No significant predictors were found for *Cx. restuans*. The sensitivity of the models ranged from 75% to 87.5% and the specificity from 75% to 93.8%. In New Haven County, the models correctly classified 81.3% of the traps for *Cx. pipiens*, 75.0% for *Cx. salinarius*, 62.5% for *Ae. vexans*, and 75.0% for *Cs. melanura*. Continuous surface maps of habitat suitability were generated for each species for both counties, which could contribute to future surveillance and intervention activities. Key Words: West Nile virus—*Culex*—Spatial modeling—Light traps. Vector-Borne Zoonotic Dis. 6, 283–295.

INTRODUCTION

WEST NILE VIRUS (family Flaviviridae, genus *Flavivirus*; WNV) (Drebot et al. 2002) was introduced into North America in 1999 and was first reported during an epidemic in New York City (Nash et al. 2001). Within 5 years, the epidemic spread throughout the United States and southern Canada, causing more than 18,000 human cases and over 700 fatalities (CDC 2005). WNV seems now firmly established in North America, so effective mosquito control measures are necessary to reduce the risk of WNV transmission to humans. However, the efficacy of preventive measures depends largely upon the accuracy of the methods used to assess the risk of human exposure to potentially infectious mosquitoes.

The current technology available for assessing human risk from mosquito-borne pathogens rests primarily on information concerning vector distribution, abundance, and virus infection prevalence derived from mosquito collections made with Centers for Disease Control and Prevention (CDC) light and/or gravid traps, the standard tools for mosquito-borne disease sur-

---

[1]Department of Epidemiology and Public Health, Yale School of Medicine, New Haven, Connecticut.
[2]Connecticut Agricultural Experiment Station, New Haven, Connecticut.

veillance. The State of Connecticut has established and maintained a statewide network of 91 fixed mosquito collection sites employing CDC light traps that have been monitored from June through October each year since 1999. However, estimates of mosquito abundance from light trap data are limited to a relatively small sampling area. If we assumed that these light traps recruited mosquitoes from an average of 1 km radius, the total land area sampled by the light traps in Connecticut would be less than 0.02% of the total area of the state.

We developed a modeling approach designed to enhance the value of mosquito trapping data by estimating species abundance in areas not sampled by traps. The resulting continuous distribution maps of the abundance of the most likely WNV vector species in Connecticut (Andreadis et al. 2004) greatly improves the spatial resolution of existing vector surveillance data. These maps can be used to direct vector control operations in the event of an epidemic and to enhance WNV surveillance by identifying habitat and abundance of vector species in regions not monitored by conventional mosquito trapping methods.

Two main modeling approaches, interpolation and landscape-based, have been followed to convert point samples of arthropod abundance to surface data that could be used to produce continuous-surface maps. Interpolation techniques such as kriging have been used to model the distribution of mosquitoes (Jeffery et al. 2002, Ryan et al. 2004), and agricultural pests (Cocu et al. 2005, Nansen et al. 2003). These modeling approaches depend on the presence of spatial autocorrelation, which is the degree of interdependence between values of a variable at different geographic scales. An alternative approach is to use landscape variables derived from remote sensing satellites as predictors, with or without incorporating the effects of spatial dependence. Pertinent examples include vectors of Eastern equine encephalomyelitis (Moncayo et al. 2000), tick vectors of Lyme disease (Brownstein et al. 2003, Dister et al. 1997, Guerra et al. 2001, 2002, Kitron et al. 1996), sand fly vectors of leishmaniasis (Cross et al. 1996, Elnaiem et al. 2003, Miranda et al. 1998, Thomson et al. 1999), tse-tse fly vectors of African trypanosomiasis (Kitron

et al. 1996, Rogers 2000), and mosquito vectors of malaria (Beck et al. 1994, 1997, Diuk-Wasser et al. 2004, Thomson et al. 1996, 1997, Wood et al. 1991a,b, 1992). Of these models, however, only a few have been validated with an independent dataset (Beck et al. 1997, Brownstein et al. 2004).

The landscape-based modeling approach is more appropriate to model mosquito abundance at surveillance trap locales. The spatial distribution of mosquitoes is limited to landscapes that contain aquatic environments permissive of oviposition and the development of larvae, plus landscapes that determine adult habitat. Some landscape elements constitute potential breeding habitats (e.g., wetlands) while others do not constitute landscape elements themselves (e.g., artificial containers), but their occurrence is landscape dependent (e.g., urban areas). These landscapes can be identified by satellite imagery and by classified datasets developed for other purposes (land use/land cover, wetlands inventory, etc.). Another factor favoring the landscape rather than the interpolation approach is that traps are generally not arranged in an equidistant grid and are often too far from each other to be spatially autocorrelated. The intensity of sampling is rather arbitrary, usually dictated by the practical constraints of landscape, access, collection effort and processing time (Reisen and Lothrop 1999).

In Connecticut, *Culex pipiens* Linnaeus, *Cx. restuans* Theobald, *Cx. salinarius* Coquillett *Aedes vexans* (Meigen), and *Culiseta melanura* (Coquillett) have been implicated as the most likely vectors of WNV based on virus isolation data (Andreadis et al. 2004) and vector competence (Turell et al. 2005). Our objective was to develop models to predict abundance of these mosquito species in areas of Connecticut not covered by the existing surveillance system. We used remotely sensed data and classified datasets as explanatory variables.

METHODS

We developed logistic models to predict abundance of *Cx. pipiens*, *Cx. restuans*, *Cx. sali-*

narius, *Ae. vexans*, and *Cs. melanura* between 2001 and 2003 using mosquito collection data obtained with CDC light traps from 32 surveillance sites in Fairfield County, Connecticut. We validated the models using similar light trap data from 16 surveillance sites in New Haven County, Connecticut, collected during the same period. We focused our study on Fairfield and New Haven counties because the principal foci of WNV activity in the state were found in densely populated residential communities in coastal areas of these two counties (Andreadis et al. 2004).

The mosquito trapping data used in this study were part of the mosquito surveillance for West Nile virus that has been conducted in Connecticut since 1999 (Andreadis et al. 2004). Mosquito trapping data presented here was obtained from June to October 2001-2003. Trapping frequency was variable but was minimally made once every ten days at each trap site. Approximately one third of the sites were located in densely populated residential locales. Trap sites included parks, greenways, golf courses, undeveloped wood lots, sewage treatment plants, landfills and wetlands. For each species, the 16 traps with highest mosquito abundance in Fairfield County were classified as high, the remaining 16 were classified as low, resulting in a binary response variable. The same procedure was followed with the validation dataset from New Haven County, where eight traps were classified as high abundance and eight as low.

Logistic models are used to compare the habitat characteristics of positive versus negative samples in order to determine habitat suitability with a set of explanatory variables (Hosmer and Lemeshow 1989). In this study, we used high/low mosquito abundance rather than presence/absence as the response variable, since these mosquito species are widespread within the state and a few individuals were present in virtually all traps. The median was used to split the dataset into high and low abundance. We modeled the average of the 3 years in order to better characterize the "typical" landscape used by each of the species.

We extracted predictor variables from the following databases (for a description of the variables, see Table 1):

1. Connecticut Department of Environmental Protection 1990 land use/land cover classification (DEP LULC) (Civco et al. 1992). This classification was generated from Landsat Thematic Mapper (TM) and Multispectral Scanner (MSS) data from May 1987/1988 and August/September 1988/1990 scenes. We reduced the 23 initial classes to seven classes, following the Anderson Level I classification system (Anderson 1976), but we included three classes instead of one for urban areas.

2. U.S. Fish and Wildlife Service National Wetlands Inventory (NWI) (Cowardin et al. 1979). This classification was generated from the analysis of high altitude photography in conjunction with collateral data sources and field work. We included the palustrine and estuarine wetland classes in our models because of their relevance as larval mosquito habitats.

3. Human population density derived from the 2000 U.S. Census at the block level.

4. Landsat Enhanced Thematic Mapper (ETM+) data. We obtained level 1G Landsat ETM+ scenes for March 26, 2000 and September 8, 2002. The digital numbers were converted to planetary reflectance to reduce the in between-scene variability (Landsat Project Science Office 2005). The September scene was registered to the March scene by selecting 60 ground control points and using a first degree polynomial and nearest neighbor algorithm, obtaining a root mean square error of 0.83. From each of the scenes, we extracted the ETM+ middle infrared band 5, associated to the water content of plants and soil, and calculated the Normalized Difference Vegetation Index (NDVI). We also calculated the difference between September and March NDVI and the NDVI variance in a 3×3 moving window. High NDVI is related to the presence and condition of green vegetation (Lillesand and Kiefer 1994), and high NDVI difference is expected in deciduous forested areas, in which there is a marked contrast between the leaves-on (September) and leaves-off (March) scenes. High pixel variance is associated with urban areas, which typically have significant "texture" resulting from buildings, street grids, green areas,

TABLE 1.   VARIABLES EXTRACTED FROM THE DATABASES USED IN THE STUDY

Database	Code	Variable name	Description
DEP LULC	IMPE	Impervious	Inner city, high-density urban areas such as concrete surfaces, roofs, paved surfaces.
	RESI	Residential	Residential neighborhoods, houses with lawns, isolated houses with farming areas.
	COMM	Commercial	Fringe of inner city areas, commercial areas, dense housing.
	FORE	Forest	Typical southern New England mixed softwood forests.
	GRAS	Grasslands/agriculture	Agricultural land, bare soil and grasslands.
	WATE	Water	Open water bodies and watercourses with relatively shallow water; possibly including periodically flooded lands, such as the Connecticut River and some non-forested wetland habitats.
	WETL	Wetlands	Palustrine, emergent wetland (PEME) and palustrine, deciduous forested wetland (PFO1) in the U.S. Fish and Wildlife wetland and deep water habitat classification system.
NWI	PALU	Palustrine	Nontidal wetlands dominated by trees, shrubs, emergents, mosses or lichens or wetlands that occur in tidal areas where salinity is below 0.5%.
	DPAL	Distance to palustrine	Distance to the closest PALU pixel.
	ESTU	Estuarine	Deepwater tidal habitats and adjacent tidal wetlands that are usually semienclosed by land but have open, partly obstructed, or sporadic access to the open ocean.
	DEST	Distance to estuaries	Distance to the closest ESTU pixel.
2000 U.S. Census	POPU	Population	2000 U.S. Census mean population per 1000 m^2.
Landsat ETM+	NDVM	NDVI March	NDVI from ETM+ March 2000 scene.
	NDVS	NDVI September	NDVI from ETM+ September 2002 scene.
	NDIF	NDVI diference	Difference in NDVI between the September 2002 and March 2000 scenes.
	VARS	NDVI variance	Variance in NDVI from September 2000 scene in a 3×3 window.
	MIR5	Middle infrared	Reflectance from ETM+ band 5 (middle infrared), September 2002 scene.

LULC DEP, Connecticut Department of Environmental Protection 1990 land use/land cover classification; NWI, U.S. Fish and Wildlife Service National Wetlands Inventory; NDVI, normalized difference vegetation index.
Variable codes are those used to descrie the models in Table 3.

etc. (Stefanov et al. 2001). All areas classified as 'deep water' in the DEP LULC were masked out and not included in the calculations.

Variable extraction

We converted all databases to ArcMap (ESRI Inc., Redlands, CA) GRID format at 30×30 m resolution and re-projected them to Universal Transverse Mercator, zone 18N projection, World Geodetic System 1984 datum, to match the satellite imagery. We used a geographic information system (GIS) procedure called "buffering" (Longley et al. 1999) to extract the values of the variables within a circular area around the 32 study traps. For categorical variables (DEP LULC and NWI), we calculated the number of pixels within the buffers, while for continuous variables (census and remotely sensed data), we extracted the mean value. We repeated this procedure for 10 buffer sizes, every 100 m, from 100 to 1000 m, to explore the spatial scale at which each variable most affected mosquito abundance. We also calculated the Euclidean distance from each trap to the nearest palustrine or estuarine pixels from the NWI database.

Model development

We evaluated whether mosquito samples in Fairfield and New Haven counties were autocorrelated (samples from closer trap sites were more similar) by *Moran's I* using ArcMap Spatial Statistics Tools. We used Akaike's Infor-

mation Criterion (AIC; Akaike 1974) to rank candidate models. This technique identifies the most parsimonious model for the data by balancing the overall fit of the model with the number of parameters included in it. Because the ratio of the sample size to the number of parameters in our most complex model was lower than 40, we used AIC bias-corrected for small sample sizes (AICc) (Hurvich and Tsai 1989) as recommended by Burnham and Anderson (2004). Under this framework, the model with the smallest AICc value is interpreted as having the best fit to the data; models within two AICc units of the best-fitting model are considered to have substantial support, while those within seven AICc units have considerably less support. We calculated Akaike weights (Burnham and Anderson 2004) to show the relative support for competing models.

We first ran univariate logistic regressions for mosquito abundance and each of the predictor variables converted to deciles (tenth percentiles), using each of the 10 buffer sizes. Variables that resulted in a $p < 0.05$ for the logistic regression and the buffer sizes that resulted in the smallest AICc for each variable were included in multiple logistic regression models, which were fit for all combinations of two and three predictors. To minimize collinearity, we included in the same model only those variables with a Spearman rank correlation (Sokal and Rohlf 1995) lower than 0.75. We considered as competing all models that were within eight AICc units of the best model and for which all predictors were significant at $p < 0.1$. We assessed all models' goodness-of-fit using the Hosmer-Lemeshow test (Hosmer and Lemeshow 1989, Hosmer and Hjort 2002). Those models with a significant lack-of-fit ($p < 0.05$) were excluded from the list of candidates. Statistical analyses were performed in STATA 8.0 (Stata Corporation, College Station, TX).

Once we determined the set of competing logistic regression models for each species, we created a "smoothed" grid layer for all predictor variables, for both Fairfield and New Haven Counties, using the focal statistic procedure in ArcMap ArcToolbox. This procedure calculates a statistic on a raster layer over a specified neighborhood. For example, for forest in a 400-m buffer, the sum of forest pixels within a 400-m buffer was calculated for each pixel in the raster and the original value of the pixel was replaced with this "smoothed" value, which integrates the values of its neighbors.

Model validation

To obtain the predicted abundance for the traps in New Haven County, we ran the logistic regression equations derived for Fairfield County traps using the map calculator function in ArcMap Spatial Analyst. The environmental predictors were extracted from the smoothed layers in New Haven County as described in the previous section. The inverse logistic transformation:

Probability ($\times$ high abundance)
$$= \exp(Y)/(1+\exp(Y))$$

was applied to the linear predictors (Y in equation above) to transform them from the logit to the probability scale. We ranked all traps by this probability and classified the highest 50% as high and the lowest 50% as low abundance. We decided to split the validation dataset into two equally sized classes in order to be consistent with the choice of the median to split the Fairfield County dataset. We examined the percent that were correctly classified as either high or low abundance. This was used as the main criteria to select the best model. We also reported the kappa statistic (Landis and Koch 1977) for the best models, which is an indicator of the extent to which the percentage correct values are due to "true" agreement versus "chance" agreement.

RESULTS

Mosquito abundance was not spatially autocorrelated among the traps for *Cx. pipiens*, *Cx. salinarius*, *Ae. vexans*, and *Cx. restuans* and was only marginally significant for Cs. melanura (*Moran's I* = 0.13, z = 1.97). We therefore considered samples from all traps to be spatially independent.

Mosquito abundance varied across the years of the study (Table 2). However, the classification into high and low abundance resulted in

TABLE 2.   ABUNDANCE OF FIVE MOSQUITO SPECIES COLLECTED FROM JUNE TO OCTOBER
IN 48 LIGHT TRAPS IN FAIRFIELD AND NEW HAVEN COUNTIES

| | Mosquitoes collected (mean ± standard error) | | |
Species	2001	2002	2003
Cx. pipiens	2.76 ± 0.44	3.03 ± 0.32	2.42 ± 0.28
Cx. restuans	3.28 ± 0.39	2.77 ± 0.35	5.55 ± 0.67
Cx. salinarius	4.64 ± 0.63	4.72 ± 0.57	16.99 ± 2.23
Cs. melanura	3.86 ± 0.49	1.16 ± 0.14	10.30 ± 1.57
Ae. vexans	14.92 ± 2.27	10.24 ± 1.65	18.80 ± 3.07

The mean number of trap nights per site/year was 20.

high consistency, with 96.6% of the sites classified in the same abundance category in all 3 years. If the maximum of the 3 years was used to classify the traps into high and low abundance, consistency between years was very similar, but slightly lower (96.0%). Therefore, we decided to use the mean as a more consistent measure of abundance.

We successfully developed logistic regression models to predict *Cx. pipiens*, *Cx. salinarius*, *Ae. vexans*, and *Cs. melanura* abundance, but were unable to develop a model for *Cx. restuans*. Table 3 lists all candidate logistic regression models within eight AICc units of the model with the lowest AICc for each species, except a model for *Cs. melanura* with palustrine as the only predictor, because of lack of fit ($p < 0.05$). We considered the best model for each species to be the one with the highest predictive power for the New Haven County dataset and, if there were more than one, we selected that with the lowest AICc. The logistic regression equations for the selected candidate models were as follows:

$$\log \text{odds} (Cx.\ pipiens) = -0.70(\text{forest } 400) + 3.76$$

$$\log \text{odds} (Cx.\ salinarius) = 0.28(\text{water } 1000) - 0.43(\text{distance to estuaries}) + 0.92$$

$$\log \text{odds} (Ae.\ vexans) = 0.60(\text{water } 1000) + 0.47(\text{grasslands/agriculture } 200) - 5.38$$

$$\log \text{odds} (Cs.\ melanura) = 0.48(\text{NDVI difference}) - 0.45(\text{distance to palustrine}) - 0.43$$

where the selected variables and respective buffer sizes are between parentheses (Table 1).

After transformation to a probability scale, the selected models depict the modeled probability of high mosquito abundance within any given pixel in Fairfield and New Haven Counties (Fig. 1).

Cx. pipiens was most abundant in urbanized areas, as measured by a large number of impervious pixels, low NDVI, low NDVI difference and high NDVI variance. The best model showed an inverse relationship to forest, which made up 56.3% of the study area. In addition to having the highest predictive power in New Haven County—correctly classifying 81.3% of the traps (kappa = 0.63, $p < 0.01$), it had the lowest AICc, and sensitivity and specificity were both 75.0%. This model indicated an association with more suburban areas, since the majority of non-forested areas was either residential (18.4%) or grasslands/agriculture (14.1%). A model including high population density and grasslands/agriculture was also among the candidate models, further supporting the association with more suburban areas.

The best model for *Cx. salinarius* included water and distance to estuaries, correctly predicted 75.0% (kappa = 0.5, $p < 0.05$) of the traps in New Haven County and had sensitivity of 75% and specificity of 93.8%. In Fairfield County, *Cx. salinarius* model including distance to estuaries had a relatively high AICc. This is likely because there were two inland sites in Fairfield County with high abundance of *Cx. salinarius* in spite of the absence of brackish water in the area. When these two sites were excluded from the analysis, this model's AICc value dropped to 20.95, becoming the model with the lowest AICc. In New Haven County,

TABLE 3. CANDIDATE MODELS FOR EACH MOSQUITO SPECIES ABUNDANCE, ORDERED BY AICc

Species	Model	OR	log (L)	AICc	Δ_i	Wi	Sensitivity	Specificity	NH	H-L	p
Cx. pipiens	**FORE_400**	**0.49**	**−13.57**	**31.55**	**0**	**0.59**	**75.0**	**75.0**	**81.3**	**5.22**	**0.52**
	NDVS_600	0.55	−14.87	34.15	2.6	0.16	75.0	75.0	75.0	5.04	0.54
	GRAS_400+POPU_700	1.55/1.56	−14.03	34.92	3.37	0.11	87.5	68.8	75.0	9.12	0.33
	IMPE_500	1.51	−15.67	35.75	4.2	0.07	75.0	81.3	68.8	2.60	0.63
	MVAR_700	1.7	−16.00	36.42	4.87	0.05	75.0	75.0	68.8	6.44	0.60
	NDIF_500	0.63	−16.99	38.41	6.86	0.02	75.0	75.0	75.0	3.09	0.80
Cx. salinarius	WATE_1000+FORE_200	1.45/0.48	−11.56	29.99	0	0.48	81.3	81.3	50.0	3.45	0.90
	FORE_200	0.5	−13.57	31.55	1.57	0.22	81.3	81.3	50.0	3.64	0.73
	WATE_1000+GRAS_300	1.42/1.73	−13.12	33.09	3.1	0.10	87.5	87.5	62.5	10.04	0.26
	GRAS_300	1.78	−15.09	34.59	4.6	0.05	81.3	81.3	56.3	5.96	0.65
	WATE_1000+NDIF_200	1.39/0.58	−14.14	35.14	5.15	0.04	81.3	84.4	62.5	12.15	0.14
	MVAR_200	1.76	−15.46	35.32	5.34	0.03	75.0	75.0	50.0	5.46	0.70
	NDIF_200	0.59	−16.00	36.42	6.43	0.02	75.0	75.0	50.0	4.86	0.56
	WATE_1000+NDVS_500	1.34/0.63	−14.87	36.6	6.61	0.02	81.3	81.3	62.5	6.26	0.62
	WATE_1000+IMPE_500	1.37/1.40	−15.01	36.88	6.9	0.02	81.3	75.0	62.5	5.18	0.74
	NDVS_500	0.61	−16.52	37.45	7.46	0.01	75.0	75.0	62.5	2.36	0.88
	WATE_1000+DEST	**1.33/0.65**	**−15.46**	**37.77**	**7.78**	**0.01**	**75.0**	**93.8**	**75.0**	**10.20**	**0.25**
Ae. vexans	WATE_1000+DPAL	2.01/1.67	−11.37	29.59	0	0.32	81.3	87.5	50.0	4.10	0.85
	WATE_1000+GRAS_200	**1.82/1.60**	**−12.10**	**31.06**	**1.47**	**0.15**	**81.3**	**87.5**	**62.5**	**9.66**	**0.29**
	WATE_1000+PALU_300	1.75/0.61	−12.44	31.74	2.15	0.11	81.3	75.0	56.3	8.76	0.36
	WATE_1000+NDIF_500	1.79/0.61	−12.50	31.85	2.26	0.10	81.3	81.3	62.5	4.99	0.76
	WATE_1000+NDVS_200	1.65/0.60	−12.53	31.92	2.32	0.10	81.3	81.3	62.5	2.84	0.94
	WATE_1000+DEST	1.75/0.61	−12.61	32.07	2.48	0.09	75.0	87.5	62.5	6.39	0.60
	WATE_1000+MVAR_200	1.53/1.56	−13.11	33.08	3.49	0.06	75.0	75.0	62.5	6.60	0.58
	MVAR_200	1.7	−16.00	36.41	6.82	0.01	75.0	75.0	62.5	5.54	0.70
Cs. melanura	MIR5_500+NDVS_700+DPAL	0.42/2.24/0.38	−8.35	26.18	0	0.51	93.8	87.5	62.5	9.95	0.27
	NDIF_800+DPAL	**1.6/0.64**	**−10.00**	**28.85**	**2.67**	**0.13**	**87.5**	**81.3**	**75.0**	**5.56**	**0.70**
	FORE_700+DPAL	1.53/0.62	−11.26	29.37	3.19	0.10	87.5	81.3	62.5	5.08	0.75
	NDIF_800+MIR5_500	2.48/0.64	−11.28	29.43	3.25	0.10	87.5	75.0	75.0	9.96	0.27
	GRAS_200+DPAL	0.66/0.64	−11.52	29.89	3.71	0.08	75.0	75.0	75.0	8.48	0.29
	DPAL	0.53	−13.30	31.02	4.84	0.04	81.3	81.3	56.3	1.65	0.80
	NDIF_800	2.01	−13.57	31.55	5.37	0.03	75.0	75.0	75.0	5.80	0.67

All models within eight AICc are shown, and the selected model is in bold. Variable codes are listed in Table 1. OR, odds ratio for each of the variables included in the model; log(L), maximized log-likelihood values; AICc, bias-corrected Akaike Information Criterion; Δ_i, differences between each candidate model and the model with the lowest AICc; Wi, Akaike weights (Burnham and Anderson 2004); sensitivity, percent of correctly classified high abundance traps; specificity, percent of correctly classified low abundance traps; NH, percent of correctly classified high abundance traps in New Haven; H-L, Hosmer Lemeshow chi^2 statitic; p, p > chi^2.
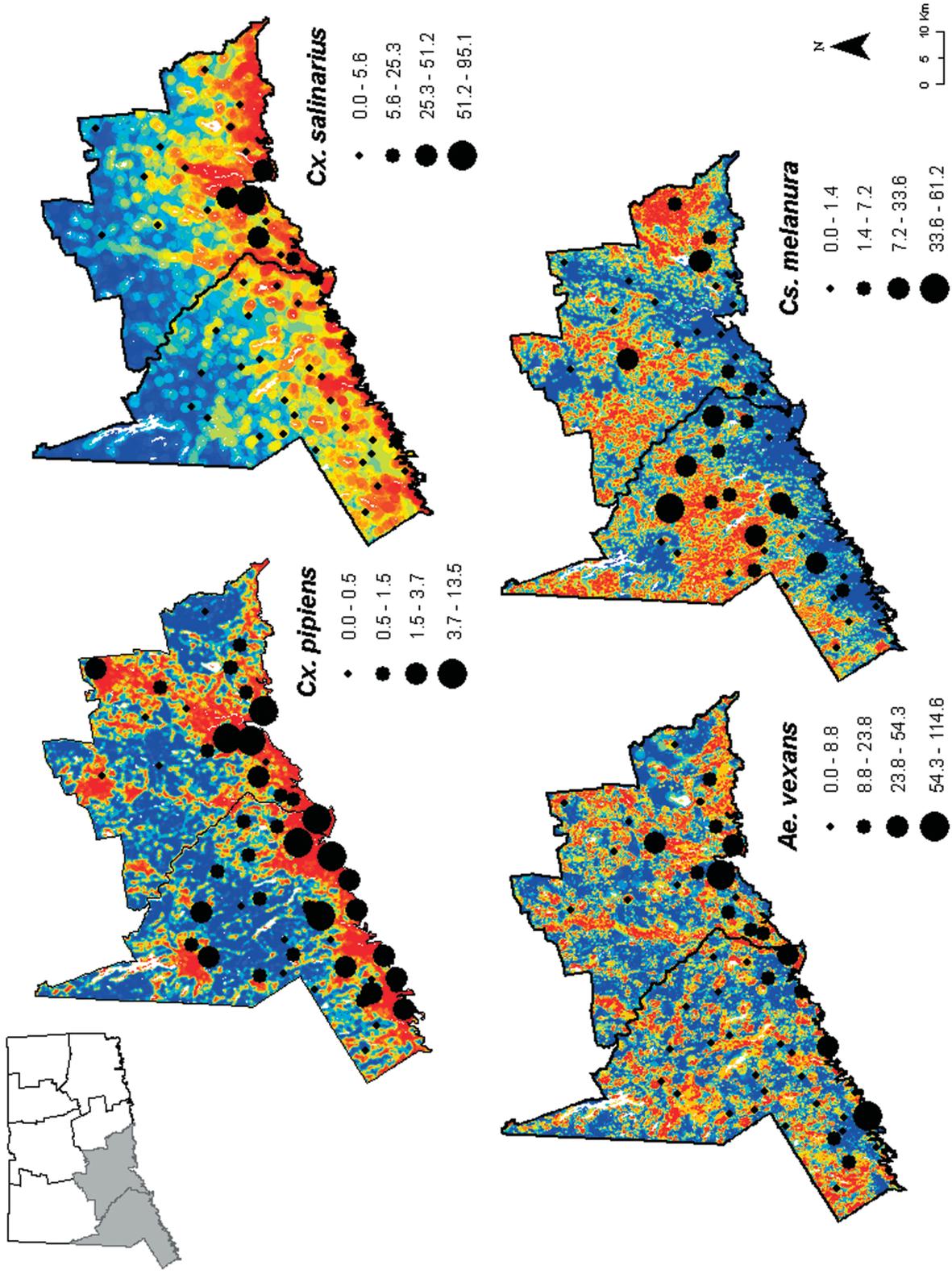
FIG. 1. Probability of high abundance for each mosquito species in Fairfield and New Haven Counties, Connecticut.. High values shown in red, low values in blue. The average number of mosquitoes/trap night in Centers for Disease Control and Prevention (CDC) light traps between 2001 and 2003, divided into quartiles, are shown in black.

Cx. salinarius showed a more typical concentration in coastal areas.

The selected model for *Ae. vexans* included water and grasslands/agriculture, and was only 1.47 AICc units from the one with the lowest AICc, so it was considered to have almost as much support. This model had high sensitivity (81.3%) and specificity (87.5%) and the highest predictive power in New Haven County (62.5%), although the kappa statistic was not significant.

The best model for *Cs. melanura* included a positive relationship with the NDVI difference and a negative one with the distance to palustrine pixels. This model had the highest predictive power in New Haven County (75.0%, kappa = 0.63, $p < 0.001$) and sensitivity and specificity were high (87.5 and 81.3%, respectively). An inverse relationship to distance to palustrine areas was included in several of the other candidate models.

DISCUSSION

We developed logistic regression models with sensitivities ranging from 75% to 87.5% and specificities from 75% to 93.8%, which correctly predicted high or low mosquito abundance in 63.0% to 81.0% of the traps of an independent dataset in New Haven County. That is, we were able to convert point-based data to surface data, allowing the estimation of mosquito abundance in areas not monitored by the light traps. To enhance the robustness of our model, we used one dataset to develop it (Fairfield County light trap samples) and another one to test it (New Haven County samples) and the final model selection was based on its predictive ability on the test dataset. Also, we used the novel approach of optimizing the size of buffers for each mosquito species/predictor variable, compared to the usual fixed 1-km buffer used in other mosquito studies (Beck et al. 1997, Moncayo et al. 2000). We used remotely sensed data, landcover and wetland classifications and census data as predictors, all of which resulted in highly predictive models. However, the variables that were most predictive were based on classified data. The variables in the selected models were consistent with known habitat preferences of each of the species, which indicates that these models were biologically sound.

The models selected for *Cx. pipiens* were consistent with the known "urban" nature of this mosquito, which breeds in ground water and artificial containers in and around domestic areas (Horsfall 1955), together with stagnant, temporary pools with a high organic content (Andreadis et al. 2005), which seems to be the attractant for oviposition (Madder et al. 1980). Andreadis et al. (2004) also reported a positive correlation between human population density and the mean number of adult female *Cx. pipiens* in 73 municipalities in Connecticut. Although *Cx. pipiens* abundance was also associated with high population density from census data, models including remotely sensed data identifying urbanized areas showed a better fit than those including population density. These two measures, however, usually coincide (Tran et al. 2002). The presence of grasslands/agriculture in the same models as variables linked to urbanization indicates that *Cx. pipiens* may be more abundant in suburban areas than is generally recognized.

We did not find any predictors for *Cx. restuans* abundance, likely because it utilizes a diverse range of larval habitats, from nearly clear water to grossly polluted (Means 1987, Andreadis et al. 2005). Information on the ecology of *Cx. restuans* is scarce, being often pooled with *Cx. pipiens* due to the difficulty in distinguishing adults and the assumption that they are sympatric and have similar feeding preferences (Apperson et al. 2002, 2004). However, Ebel et al. (2005) found that *Cx. restuans* tended to comprise a greater proportion of the total collections at rural sites compared with urban sites, which is consistent with our findings of a less tight link to urban areas.

Several candidate models for *Cx. salinarius* and *Ae. vexans* included the same variables, despite their different habitat preferences; *Cx. salinarius* is often referred to as a salt marsh-inhabiting mosquito, while and *Ae. vexans* commonly occurs in floodwaters. The similarity of the models could be due either to these two species using more diversified habitats or to the inclusion of several habitat types in some of the remotely-sensed environmental vari-

ables. There is support in the literature for the former argument since, according to Crans (2004), *Cx. salinarius* larvae rarely breed directly in tidal marshes, but rather occur more frequently in areas adjacent to salt marshes where fresh water from upland drains to coastal habitats producing a brackish water environment. They are also particularly abundant in freshwater impoundments, with their populations peaking immediately after flooding (Slaff and Crans 1982). *Ae. vexans* eggs are laid in sites subject to inundation by rain water, overflow, seepage or tidal water and can also be found in salt marsh impoundments, in addition to open rain pools, tire ruts, stormwater management facilities, dredge spoil sites, ditches and areas in which streams or creeks have flooded over their banks (O'Malley 1990). The models may also be similar because the "water" class includes open water bodies and watercourses with relatively shallow water, both permanent and periodically flooded lands, such as the margins of the Connecticut River, not normally perceived as water habitats (Civco et al. 1992). Other sources of information on surface water should be explored for their use in identification of mosquito habitats.

The high predictive power of NDVI difference—associated with deciduous forests, and closeness to palustrine areas for *Cs. melanura* abundance—is consistent with the bionomics of this species. Culiseta melanura is a widespread mosquito distributed throughout the eastern and central United States (Darsie and Ward 1989). It is a multivoltine species that has two to three overlapping generations a year in the northeastern United States (Mahmood and Crans 1998), and is among the most dominant mosquitoes found in densely wooded red maple and Atlantic white cedar swamps and sphagnum bogs in Connecticut (Andreadis et al. 2004, 2005). Eggs are laid in water in permanent subterranean habitats, and larvae develop in holes beneath mats of sphagnum and in deep shaded cavities around the roots of upturned trees (Mahmood and Crans 1998). The habitat is comparatively stable and is generally inundated with water throughout the year.

The effective sampling area of light traps is not known and it is likely to vary for different mosquito species and also for different land-cover types. This poses a challenge when relating environmental variables measured as a surface to mosquito abundance measured at a point (light trap). Previous studies estimating an environmental variable around a point (Moncayo et al. 2000, Beck et al. 1994, 1997) used buffers of 1 km, based on the typical flight range cited for *Anopheles albimanus* (Hobbs et al. 1974). However, a wide range of flight ranges have been reported for the mosquitoes studied here, 1–3.5 km for *Cx. pipiens* (Fussell 1964, Schreiber et al. 1988), 1–2 km for *Cx. salinarius* (Lasalle and Dakin 1982, Morris et al. 1991), 4–8 km for Cs. melanura (Howard et al. 1989), and 1–14 km for *Ae. vexans* (Brust 1980, Clarke 1943, Jensen and Washino 1994, MacCreary and Stearns 1937). Given the large variability in these flight ranges and the likely relationship with the specific landcovers in each area, we used the sampling area that resulted in the closest association between abundance of each mosquito species and each environmental variable. We found that buffer sizes smaller than 1 km resulted in closer associations for all species and environmental variables, except for the water variable. The larger buffer size for the water class may indicate that traps in these habitats recruited mosquitoes from a larger area and/or that *Cx. salinarius* and *Ae. vexans* dispersed greater distances than the other species.

Mosquito abundance was not spatially autocorrelated for four of the five species. Spatial autocorrelation would be expected if nearby traps recruited mosquitoes from common breeding sites, which would result in samples with similar abundance. The lack of spatial structure is likely due to the large mean distance between the traps (4.7 km) relative to the flight range reported for most mosquito species and the optimum buffer size identified in this study (<1 km). The lack of spatial autocorrelation precluded any conclusions on recommended distances for trap placements from this study. We therefore considered regression analysis using environmental variables as predictors as a more appropriate analytical approach than interpolation methods.

Logistic regression modeling has been extensively used to formalize the relationship between environmental factors and the pres-

ence/absence of vectors (Brownstein et al. 2003, Dumonteil and Gourbiere 2004, Thomson et al. 1999) and incidence of vector-borne diseases in humans (Brooker et al. 2001, Elnaiem et al. 2003, King et al. 2004, Lindsay and Thomas 2000, Thompson et al. 1996) and to quantify the amount of potential habitat. In this study, we did not use presence/absence but rather high/low mosquito abundance as the response variable. From an epidemiological viewpoint, some of these mosquito vectors may pose a health risk to humans once they have exceeded a certain threshold of abundance. If future research defines critical thresholds, these models could be easily adapted to represent the probability of being above a numerical threshold. However, defining areas of different levels of risk can still be useful to more efficiently allocated limited resources to those with the highest relative risk. Vector control officials could then use these models to make more informed decisions on the type, focus and intensity of mosquito abatement.

In summary, we have generated habitat suitability maps of four potential WNV vector species, which represent the probability of trapping high numbers of these mosquitoes in every 30 m $\times$ 30 m pixel of Fairfield and New Haven counties (3270 km^2). These surface maps constitute a significant improvement over traditional reliance on data derived from point samples from individual light traps. These validated predicted maps could help improve West Nile virus surveillance efforts by identifying all areas of expected high vector mosquito abundance across a large geographic area. They could therefore guide future trap placements to the most suitable habitats for particular mosquito species and aid in focusing disease prevention measures toward areas of high vector abundance during WNV epizootics.

ACKNOWLEDGMENTS

REFERENCES

Anderson, JR. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data.* Geological Survey Professional Paper 964. Washington, DC: U.S. Government Printing Office, 1976.

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

Andreadis, TG, Thomas, MC, Shepard, JJ. *Identification Guide to the Mosquitoes of Connecticut.* Bulletin No. 966. New Haven, CT: Connecticut Agricultural Experiment Station, 2005.

Beck, LR, Rodriguez, MH, Dister, SW, et al. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. Am J Trop Med Hyg 1994; 51:271–280.

Beck, LR, Rodriguez, MH, Dister, SW, et al. Assessment of a remote sensing–based model for predicting malaria transmission risk in villages of Chiapas, Mexico. Am J Trop Med Hyg 1997; 56:99–106.

Brooker, S, Hays, SI, Issae, W, et al. Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. Trop Med Int Health 2001; 6:998–1007.

Brownstein, JS, Rosen, H, Purdy, D, et al. Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. Vector-Borne Zoonotic Dis 2003; 3:155–155.

Brownstein, JS, Holford, TR, Fish, D. Enhancing West Nile virus surveillance, United States. Emerg Infect Dis 2004; 10:1129–1133.

Brust, RA. Dispersal behavior of adult *Aedes sticticus* and *Aedes vexans* (Diptera, Culicidae) in Manitoba. Can Entomol 1980; 112:31–42.

Burnham, KP, Anderson, DR. Multimodel inference—understanding AIC and BIC in model selection. Sociol Methods Res 2004; 33:261–304.

CDC (Centers for Disease Control and Prevention). *West Nile Virus Activity—United States.* Washington, DC: CDC, 2005.

Civco, DL, Miller, J, Hurd, J, et al. *Connecticut Statewide Land Use and Land Cover Mapping 1992. Project Comple-*

tion Report for U.S. EPA-Conn. DEP Joint Long Island Sound Research Project CWF 219-R.

Clarke, JL. Studies of the flight range of mosquitoes. J Econ Entomol 1943; 36:121–122.

Cocu, N, Conrad, K, Harrington, R, et al. Analysis of spatial patterns at a geographical scale over north-western Europe from point-referenced aphid count data. B Entomol Res 2005; 95:47–56.

Cowardin, L, Carter, V, Golet, F, et al. *Classification of Wetlands and Deepwater Habitats of the United States*. Washington DC: U.S. Department of the Interior Fish and Wildlife Services, 1979.

Crans, WJ. A classification system for mosquito life cycles: life cycle types for mosquitoes of the northeastern United States. J Vector Ecol 2004; 29:1–10.

Cross, ER, Newcomb, WW, Tucker, CJ. Use of weather data and remote sensing to predict the geographic and seasonal distribution of *Phlebotomus papatasi* in southwest Asia. Am J Trop Med Hyg 1996; 54:530–536.

Darsie, RF, Ward, RA. Review of new nearctic mosquito distributional records north of Mexico, with notes on additions and taxonomic changes of the fauna, 1982–89. J Am Mos Control Assoc 1989; 5:552–557.

Dister, SW, Fish D, Bros, SM, et al. Landscape characterization of peridomestic risk for Lyme disease using satellite imagery. Am J Trop Med Hyg 1997; 57:687–692.

Diuk-Wasser, MA, Bogayoko, M, Sogoba, N, et al. Mapping rice field anopheline breeding habitats in Mali, West Africa, using Landsat ETM+ sensor data. Int J Remote Sens 2004; 25:359–376.

Dumonteil, E, Gourbiere, S. Predicting *Triatoma dimidiata* abundance and infection rate: A risk map for natural transmission of Chagas disease in the Yucatan Peninsula of Mexico. Am J Trop Med Hyg 2004; 70:514–519.

Ebel, GD, Rochlin, I, Longacker, J, et al. *Culex restuans* (Diptera: Culicidae) relative abundance and vector competence for West Nile virus. J Med Entomol 2005; 42:838–843.

Elnaiem, DE, Schorschev, J, et al. Risk mapping of visceral leishmaniasis: the role of local variation in rainfall and altitude on the presence and incidence of kala-azar in eastern Sudan. Am J Trop Med Hyg 2003; 68:10–17.

Fussell, EM. Dispersal studies on radioactive-tagged *Culex quinquefasciatus* Say. Mosq News 1964; 24:422–426.

Guerra, M, Walker, E, Jones, C, et al. Predicting the risk of Lyme disease: habitat suitability for *Ixodes scapularis* in the north central United States. Emerg Infect Dis 2002; 8:289–297.

Guerra, MA, Walker, ED, Kitron, U. Canine surveillance system for Lyme borreliosis in Wisconsin and northern Illinois: geographic distribution and risk factor analysis. Am J Trop Med Hyg 2001; 65:546–552.

Hobbs, JH, Lowe, RE, Schreck, CE. Studies of flight range and survival of *Anopheles albimanus* Wiedemann in El Salvador. 1. Dispersal and survival during dry season. Mosq News 1974; 34:389–393.

Horsfall, WR. *Mosquitoes—Their Bionomics and Relation to Disease*. New York: Ronald Press, 1955.

Hosmer, DW, Lemeshow, S. *Applied Logistic Regression*. New York: Wiley; 1989.

Hosmer, DW, Hjort, NL. Goodness-of-fit processes for logistic regression: simulation results. Stat Med 2002; 21:2723–2738.

Howard, JJ, White, DJ, Muller, SL. Mark-recapture studies on the *Culiseta* (Diptera, Culicidae) vectors of Eastern Equine Encephalitis virus. J Med Entomol 1989; 26:190–199.

Hurvich, CM, Tsai, CL. Regression and time-series model selection in small samples. Biometrika 1989; 76:297–307.

Jeffery, JA, Ryan, PA, Lyons, SA, et al. Spatial distribution of vectors of Ross River virus and Barmah Forest virus on Russell Island, Moreton Bay, Queensland. Aust J Entomol 2002; 41:329–338.

Jensen, T, Washino, RK. Comparison of recapture patterns of marked and released *Aedes vexans* and *Ae. melanimon* (Diptera, Culicidae) in the Sacramento Valley of California. J Med Entomol 1994; 31:607–610.

King, RJ, Campbell-Lendrum, DH, Davies, CR. Predicting geographic variation in cutaneous leishmaniasis, Colombia. Emerg Infect Dis 2004; 10:598–607.

Kitron, U, Okeno, LH, Hungerford, LL, et al. Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS. J Anim Ecol 1996; 65:371–380.

Landis, JR, Koch, GG. Measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174.

Landsat Project Science Office. 2005. *Landsat 7 Science Data User's Handbook.* Available at: http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_htmls/chapter11/chapter11.html. Accessed July 1, 2006.

Lasalle, MW, Dakin, ME. Dispersal of *Culex salinarius* in Southwestern Louisiana. Mosq News 1982; 42:543–550.

Lillesand, TM, Kiefer, RW. *Remote Sensing and Image Interpretation*, 3rd ed. New York: Wiley, 1994.

Lindsay, SW, Thomas, CJ. Mapping and estimating the population at risk from lymphatic filariasis in Africa. Trans R Soc Trop Med Hyg 2000; 94:37–45.

Longley, PA, Goodchild, MF, Maguire, DJ, et al. *Geographical Information Systems: Principles, Techniques, Management and Applications*. New York: Wiley, 1999.

MacCreary, D, Stearns, LA. Mosquito migration across Delaware Bay. Proc NJ Mosq Exterm Assoc 1937; 188–197.

Madder, DJ, Macdonald, RS, Surgeoner, GA, et al. The use of oviposition activity to monitor populations of *Culex pipiens* and *Culex restuans* (Diptera, Culicidae). Can Entomol 1980; 112:1013–1018.

Miranda, C, Marques, CCA, Massa, JL. Satellite remote sensing as a tool for the analysis of the occurrence of American cutaneous leishmaniasis in Brazil. Rev Saude Publ 1998; 32:455–463.

Moncayo, AC, Edman, JD, Finn, JT. Application of geographic information technology in determining risk of eastern equine encephalomyelitis virus transmission. J Am Mosq Control Assoc 2000; 16:28–35.

Morris, CD, Larson, VL, Lounibos, LP. Measuring mosquito dispersal for control programs. J Am Mosq Control Assoc 1991; 7:608–615.

Nansen, C, Campbell, JF, Phillips, TW, et al. The impact of spatial structure on the accuracy of contour maps of small data sets. J Econ Entomol 2003; 96:1617–1625.

Nash, D, Mostashari, F, Fine, A, et al. The outbreak of West Nile virus infection in the New York City area in 1999. N Engl J Med 2001; 344:1807–1814.

O'Malley, C. *Aedes vexans* (Meigen): an old foe. Proc NJ Mosq Control Assoc 1990; 90–95.

Reisen, W, Lothrop, H. Effects of sampling design on the estimation of adult mosquito abundance. J Am Mosq Control Assoc 1999; 15:105–114.

Rogers, DJ. Satellites, space, time and the African trypanosomiases. Adv Parasitol 2000; 47:129–171.

Ryan, PA, Lyons, SA, Alsemgeest, D, et al. Spatial statistical analysis of adult mosquito (Diptera: Culicidae) counts: an example using light trap data, in Redland Shire, Southeastern Queensland, Australia. J Med Entomol 2004; 41:1143–1156.

Schreiber, ET, Mulla, MS, Chaney, JD, et al. Dispersal of *Culex quinquefasciatus* from a dairy in southern California. J Am Mosq Control Assoc 1988; 4:300–304.

Slaff, M, Crans, WJ. Impounded water as a major producer of *Culex salinarius* (Diptera, Culicidae) in coastal areas of New Jersey, USA. J Med Entomol 1982; 19:185–190.

Sokal, RR, Rohlf, FJ. *Biometry: The Principles and Practice of Statistics in Biological Research*, 3rd ed. New York: W.H. Freeman, 1995.

Stefanov, WL, Ramsey, MS, Christensen, PR. Monitoring urban land cover change: an expert system approach to land cover classification of semiarid to arid urban centers. Remote Sens Environ 2001; 77:173–185.

Thompson, DF, Malone, JB, Harb, M, et al. Bancroftian filariasis distribution and diurnal temperature differences in the southern Nile delta. Emerg Infect Dis 1996; 2:234–235.

Thomson, MC, Connor, SJ, Milligan, P, et al. The ecology of malaria—as seen from Earth-observation satellites. Ann Trop Med Parasitol 1996; 90:243–264.

Thomson, MC, Connor, CJ, Milligan, P, et al. Mapping malaria risk in Africa: what can satellite data contribute? Parasitol Today 1997; 13:313–318.

Thomson, MC, Elnaiem, DA, Ashford, RW et al. Towards a kala azar risk map for Sudan: mapping the potential distribution of *Phlebotomus orientalis* using digital data of environmental variables. Trop Med Int Health 1999; 4:105–113. [Erratum in: Trop Med Int Health 1999; 4:240–241.

Tran, A, Gardon, J, Weber, S, et al. Mapping disease incidence in suburban areas using remotely sensed data. Am J Epidemiol 2002; 156:662–668.

Wood, B, Washino, R, Beck, L, et al. Distinguishing high and low anopheline producing rice fields using remote sensing and GIS technologies. Prev Vet Med 1991a; 11:277–288.

Wood, BL, Washino, R, Palchick, S, et al. Spectral and spatial characterization of rice field mosquito habitat. Int J Remote Sens 1991b; 12:621–626.

Wood, BL, Beck, L, Washino, R, et al. Estimating high mosquito-producing rice fields using spectral and spatial data. Int J Remote Sens 1992; 13:2813–2826.

Address reprint requests to:
Dr. M.A. Diuk-Wasser
Department of Epidemiology and Public Health
Yale School of Medicine
60 College St.
New Haven, CT 06520-8034

E-mail: maria.diuk@yale.edu