Assessing Ultra-Fine-Scale Factors to Improve Human West Nile Virus Disease Models in the Chicago Area

Johnny Uelmen (uelmen@illinois.edu)
University of Illinois at Urbana-Champaign https://orcid.org/0000-0003-3057-5107
Patrick Irwin
Northwest Mosquito Abatement District
William Brown
University of Illinois at Urbana-Champaign
Surendra Karki
University of Illinois at Urbana-Champaign
Marilyn O'Hara Ruiz
University of Illinois at Urbana-Champaign
Bo Li
University of Illinois at Urbana-Champaign
Rebecca Smith
University of Illinois at Urbana-Champaign

Research

Keywords: WNV, disease modeling, Culex, vector index, human illness

Posted Date: March 31st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-19812/v1

Assessing Ultra-Fine-Scale Factors to Improve Human West Nile Virus Disease Models in the Chicago Area

[*1]Uelmen, J.A., [2]Irwin, P., [1]Brown, W.M., [1,3]Karki, S., [1]Ruiz, M.O., [4]Li, B., and [1]Smith, R. L.

[1]University of Illinois at Urbana-Champaign, Department of Pathobiology, College of Veterinary Medicine, 2522 Veterinary Medicine Basic Sciences Building, 2001 South Lincoln Avenue, Urbana, IL 61802 USA

[2]Northwest Mosquito Abatement District, 147 West Hintz Road, Wheeling, IL 60090 USA

[3]Department of Epidemiology and Public Health, Himalayan College of Agricultural Sciences and Technology, Kirtipur, Kathmandu, Nepal

[4]University of Illinois at Urbana-Champaign, Department of Statistics, Illini Hall, Room 101, 725 South Wright Street, Champaign, IL 61820 USA

[*]Corresponding Author

Keywords: WNV, disease modeling, *Culex*, vector index, human illness

Abstract

Background: Since 1999, West Nile virus (WNV) has moved rapidly across the United States, resulting in tens of thousands of human cases. Both the number of human cases and the level of mosquito infection (MIR) vary across time and space and are related to numerous abiotic and biotic forces, ranging from differences in microclimates to socio-demographic factors. Because the interactions among these multiple factors affect the locally variable risk of WNV illness, it has been especially difficult to model human disease risk across varying spatial and temporal scales. Cook and DuPage Counties, comprising the city of Chicago and surrounding suburbs, are among the areas hardest hit by WNV in the United States. Despite active mosquito control efforts, there is consistent annual WNV presence, resulting in more than 285 confirmed WNV human cases and 20 deaths in the past 5 years in Cook County alone.

Methods: A previous WNV model for the greater Chicago area identified the fifty-five most high and low risk study areas in the Northwest Mosquito Abatement District (NWMAD), an enclave ¼ the size of the previous study area. In these locations, human WNV risk was stratified by strength of predictive success, as indicated by differences in studentized residuals. Within these areas, an additional two-years of field collections and data processing was added to a 10-year WNV dataset and assessed by an ultra-fine-scale multivariate logistic regression model.

Results: Multivariate statistical approaches revealed that this ultra-fine-scale model resulted in fewer explanatory variables while improving upon the fit of the existing model. Beyond mosquito infection rates and climatic factors, efforts to acquire additional covariates only slightly improve model predictive performance.

Conclusions: These results suggest human WNV illness in the Chicago area may be associated with fewer, but increasingly critical, key variables at finer scales. Given limited resources, this study suggests a large variation in the significance to model performance, and provides guidance in covariate selection for optimal WNV human illness modeling.

1. Introduction

In December of 1937 in Northern Uganda, a 37-year-old woman became ill with a fever of 100.6°F (Smithburn et al. 1940). She would later become the first documented human infected with the West Nile virus (WNV; Family *Flaviviridae*), a mosquito-borne disease originating from the West Nile region of Uganda. WNV first arrived to the United States (U.S., New York, NY) in 1999, most likely via a hitchhiking infected mosquito in an airline wheel well (Hadfield et al. 2019). This newly introduced WNV strain matched the Isr98 strain, isolated from a single goose in Israel in 1998 (McLean et al. 2002). Once arriving in New York, the virus took only three years to traverse the contiguous U.S., reaching California in 2002 (Sejvar 2003). The virus has now become one of the most widespread arboviruses in the world, and is present in every continent except Antarctica (Kramer et. al. 2008).

In the Midwestern U.S., mosquitoes of the *Culex* (*Cx.*) genus are the main vectors for transmitting WNV (Goddard et al. 2002). *Culex* mosquitoes are capable of feeding on several hosts to satisfy one blood meal, increasing the opportunity for multiple infections across species (Hamer et al. 2009). Although primarily ornithophilic, prior studies indicate that *Cx.* species may shift feeding preferences to humans later in the summer months (Russell and Hunter 2012); (Kilpatrick et al. 2006). Humans and other mammals, most notably horses, are considered "dead-end" hosts, not capable of producing sufficient levels of viremia to subsequently infect biting mosquitoes (Bowen and Nemeth 2007).

From 1999-2018, there have been a total of 50,830 human cases resulting in 2,330 deaths across the US (Centers for Disease Control (CDC) 2019). In many cities and states that experience high WNV incidence, there are efforts in place to control mosquito

65 populations. However, despite these methods, WNV continues its epizootic and enzootic
66 cycles year to year, and large-scale outbreaks have occurred in the years 2002, 2003,
67 2012, and 2018 across the United States (citation?). At local scales, drivers of human
68 disease, including WNV, vary in actual effect and magnitude from that observed on state,
69 regional, or national scales. Previous studies have identified common abiotic and biotic
70 factors associated with human WNV illness, including prior weather conditions (weekly
71 temperature and precipitation lags), mosquito infection and abundance, socio-
72 demographic characteristics of the local population, and level of public awareness and
73 education, but these were all at state or regional scales (Ruiz et al. 2004; Kilpatrick and
74 Pape 2013; Manore et al. 2014; Roiz et al. 2014; Rosà et al. 2014; Wimberly et al. 2014;
75 Hahn et al. 2015; Giordano et al. 2017).
76 Karki et al. (2019) is one of the few studies to evaluate weekly spatiotemporal factors
77 and their associations with human WNV illness at a smaller scale, in a highly urban 2-
78 county area (Cook/DuPage counties) that includes the greater Chicago, IL area. This
79 region consistently experiences one of the highest annual WNV incidences in the country.
80 The previous study incorporated among the finest temporal and spatial scales known to
81 date, using 1-km hexagon grids to minimize biases from political boundaries. While an
82 excellent overall model fit was achieved by using a large number of explanatory variables,
83 the relative importance of covariates and the resulting disease prediction across mico-
84 scales is still not understood.
85 The Northwest Mosquito Abatement District (NWMAD), occupying the northwest
86 corner of Cook County, is one of Chicago's four abatement districts responsible for
87 mosquito control, and has an excellent long-term mosquito abundance and testing data

88 throughout its jurisdiction. This study targeted fifty-five individual 1-km hexagons within
89 the NWMAD for an ultra-fine-scale (UFS) assessment of human WNV illness spatio-
90 temporal variability in suburban environments. Specifically, this study's main objectives
91 were to: (i) evaluate and contrast key variables in this study to the larger Cook/DuPage
92 model, (ii) assess the similarities and differences among locations that were predicted
93 accurately by the larger model and those that were predicted poorly, (iii) quantify the
94 impact of newly acquired data on prediction of human WNV illness, and (iv) determine if
95 vector index is a stronger predictor than the additive effects of MIR and mosquito
96 abundance. Ultimately, this study aims to highlight how WNV disease variance may be
97 better captured at finer spatio-temporal scales. The results of this study will provide
98 future researchers, public health agencies, and abatement districts essential details and
99 suggestions for improving WNV prediction and optimizing efficiency of targeted
100 mosquito control efforts.

101 **2. Methods**

102 This project was approved by the Institutional Review Board of the University of
103 Illinois at Urbana-Champaign, the Illinois Department of Public Health (IDPH), and the
104 University of Illinois Biosafety Committee. Human case data were provided by IDPH
105 without any personal identifying information.

106 *2.1 Study area*

107 This study was conducted within the NWMAD, a 605-km^2 area that comprises the
108 northwest suburbs of Chicago (Cook County, IL, Figure 1). As described in Karki et al.
109 (2019), all model data were summarized and processed within 1-km diameter hexagons,
110 as a neutral configuration in both size and shape, free of any political boundaries. Using

111 statistical selection processes (described below), fifty-five of the 1,019 hexagons within
112 the NWMAD were selected as the observational units for this study.

113 *2.2 Model covariates*

114 The Cook/DuPage model evaluated forty covariates derived from a variety of abiotic
115 and biotic factors, including climate and weather records, mosquito infection, socio-
116 demographic census data, and other biological conditions (described below). For this
117 study, additional data processing and field collections resulted in forty-two additional
118 non-collinear independent variables (Table 1). Each variable was independently
119 calculated by hexagon for CDC epidemiological weeks 18-38 (Sunday-Saturday) of the
120 years 2005 through 2016 (CDC 2019).

121 2.2.1 Previously existing data

122 *2.2.1a Human illness*

123 Human WNV cases in Illinois were classified as either confirmed[1] or probable[2], as
124 reported to the IDPH by public health or licensed medical professionals (mandatory
125 reporting of WNV cases is required in the state). We recognize that exposure to
126 mosquito-borne disease occurs often and in many locations. Confirming the moment an
127 infected mosquito inoculates a human is nearly impossible to document. Therefore, we
128 assumed human cases were exposed to WNV at their home addresses. The latitude and
129 longitude point locations were provided to the third decimal degree and aggregated to

[1] The case definition for a confirmed case of arboviral encephalitis in Illinois is a clinically compatible illness that is laboratory confirmed at a public health laboratory. The laboratory criteria are a fourfold or greater rise in serum antibody titer; or isolation of virus from, or demonstration of viral antigen in, tissue, blood, CSF or other body fluid; or specific IgM antibody in CSF.

[2] A probable case of arboviral encephalitis is a clinically compatible illness occurring during the season when arbovirus transmission is likely to occur and with the following supportive serology: a stable (twofold or smaller change) elevated antibody titer to an arbovirus, e.g., > 320 by hemagglutination inhibition, > 128 by complement fixation (CF), > 256 by IF, > 160 by neutralization, or a positive serologic result by enzyme immunoassay (EIA) or MAC ELISA.

130 the hexagon level for analytical and display purposes. Human cases were converted into
131 binary form (presence/absence of illness) and weekly case rate, controlling for human
132 population, for each hexagon. Use of human case data was approved by the University
133 of Illinois Institutional Review Board and the Illinois Department of Public Health.
134 *2.2.1b Abiotic Predictors*
135 Land Cover: The 2011 United States Geological Survey (USGS 2011) National Land
136 Cover Database (NLCD) provided 30 m. resolution classified raster data for the
137 NWMAD. The raster comprising NWMAD was clipped, extracted, and tabulated by
138 landscape code using the tabulate area tool in ArcGIS 10.5.1 (Environmental Systems
139 Research Insititute 2011). There were 15 unique land cover types: forests (deciduous,
140 evergreen, and mixed), urban (developed open space, developed low intensity, developed
141 medium intensity, and developed high intensity), open water, herbaceous wetlands,
142 cultivated crops, wetlands (woody and herbaceous), grassland, barren land, and shrubs.
143 Proportions of each type within each hexagon were calculated using the 30m. raster
144 resolution.
145 Weather: Daily mean temperature and precipitation were acquired from the PRISM
146 Climate Group (Oregon State University 2019), provided as 4-km resolution grids.
147 Weekly mean temperatures were calculated by taking the average of each of the seven
148 days of the week, whereas weekly precipitation totals were calculated as a sum of each of
149 the seven days of the week. As a proxy for winter temperature, the monthly average for
150 each January from 2005-2016 was also calculated. Using the zonal statistics as table
151 function in ArcGIS, each mean temperature and precipitation value was extracted for
152 each hexagon in this study.

153 *2.2.1c Biotic Predictors*

154 Mosquito infection: All mosquito infection data were acquired from the Illinois
155 Department of Public Health (IDPH), the state agency responsible for collecting and
156 maintaining standardized mosquito collection and testing data. Mosquito infection is
157 defined as the minimum infection rate (MIR), calculated by the following equation:

158 $$\frac{\#\ of\ positive\ mosquito\ pools}{total\ specimens\ tested} \text{ x } 1000,$$

159 where a mosquito pool in this analysis consisted of up to 50 female *Culex* mosquitoes
160 that were collected by the same trap. A vast majority of the tests used to identify the
161 presence of WNV was the Rapid Analyte Measurement Platform (RAMP), although
162 some mosquito pools were also tested by Real Time reverse transcriptase polymerase
163 chain reaction (RT-PCR) or VecTest.

164 Trap locations were provided by the IDPH. Whenever precise spatial locations were not
165 available, the existing address on file was used to generate a geocoded trap location. The
166 MIR values for each trap were calculated and interpolated across the NWMAD by
167 inverse distance weighting (IDW) in ArcGIS. The average MIR values were extracted for
168 each hexagon by using the zonal statistics as table function in ArcGIS.

169 Demographic: Total population and racial composition (White, African American,
170 Hispanic, and Asian) at the census block level were extracted from the 2010 U.S. Census
171 then converted as a percentage for each hexagon. Additionally, age of housing (built
172 before 1940, 1940-1969, 1970-1989, and post 1990) and income were averaged for each
173 hexagon using data provided by 2015 American Community Survey. These data were
174 processed in ArcGIS using the intersection tool.

175 2.2.2 Newly added data

176 *2.2.2a Abiotic Predictors*

177 Catch basin (e.g. sewer) density: The NWMAD provided point data for each catch basin
178 within its jurisdiction. All point data was then aggregated to each hexagon using the
179 spatial location join feature in ArcGIS. A combined total of 8,443 catch basins were
180 recorded among all hexagons (min = 1, max = 543).

181 Size and distribution of commercial and residential lots and buildings: High-resolution
182 (1 m.) aerial imagery from ArcGIS and USDA (2018) were used as a basemap for each
183 hexagon. Each permanent structure (e.g. residence, shed, garage, deck) was traced and
184 converted to polygons in ArcGIS. The area and perimeter of each polygon was calculated
185 and aggregated for each hexagon. Commercial and residential lots were provided by
186 Cook County Data Catalogy (2019), using 2016 tax appropriations. In total, there were a
187 combined 22,892 lots with 24,468 buildings or permanent structures.

188 Light pollution: Radiation from light pollution was provided by the New World Atlas of
189 Artificial Night Sky Brightness (Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M.,
190 Elvidge, C. D., Baugh, K., Portnov, B., Rybnikova, N. A., Furgoni 2016a, 2016b). Light
191 pollution was acquired from 2014 data of the VIIRS DNB sensor on the Suomi National
192 Polar-orbiting Partnership (NPP) satellite. Pixel resolution was 0.75 km; mean value for
193 each 1-km hexagon was calculated in ArcGIS.

194 *2.2.2b Biotic Predictors*

195 Historical mosquito abundance: The NWMAD consistently collected and diligently
196 maintained their mosquito trapping and identification data throughout the study period.
197 Once deployed, traps were usually checked at least twice a week. Over the 2005-2016
198 study period, there were a total of 59 traps used in the NWMAD, resulting in a total of

48,406 female *Cx.* spp. from 22 light traps, and 1,110,024 from 37 gravid traps. Weekly mosquito collections by trap were geocoded and interpolated across all hexagons via IDW and extracted using the zonal statistics as table function for each hexagon in ArcGIS 10.5.1. Mosquito abundance was calculated as the weekly cumulative number of captured female *Culex* spp. from each respective gravid trap (GT) and light trap (LT). Since early trap data did not reliably identify mosquitoes to species, all *Cx.* spp. values were pooled. However, prior studies from Chicago region indicated that *Cx. pipiens* and *Cx. restuans* are the major *Cx*. species present in this area. Normalized Difference Vegetation Index (NDVI): To evaluate the magnitude of all vegetation, NDVI was incorporated by hexagon, recorded as an average value at three timepoints of each year: CDC epidemiologic weeks 21 (3^{rd}-4^{th} week of May), 28 (2^{nd}-3^{rd} week of July), and 35 (4^{th} week of August-1^{st} week of September). These CDC epidemiologic weeks mark the center of each the three 8-week active WNV periods in the Midwest, represented as T1 = low WNV activity, T2 = high WNV activity, and T3 = moderate WNV activity. The best available Landsat 7 or 8 bands for each respective time period were acquired from EarthExplorer (USGS 2019) and processed in ArcGIS 10.5.1.

Human exposure during crepuscular time periods: Human activity observations were conducted in public spaces inside each hexagon, during the crepuscular hours between 6-9:30pm, the preferred feeding period for *Cx. pipiens/restuans*. Observations were conducted within each hexagon for a total of ten minutes per visit. Specifically, a researcher remained stationary for 2 minutes, walked 2 minutes, remained stationary in the new position for 2 minutes, walked back to origination point for 2 minutes, then remained stationary in the original position for 2 final minutes. Human exposure was

222 determined as any period in time a person was outside of any building, vehicle, or
223 enclosed dwelling during the observation period. Observations were classified by
224 apparent gender and age (child, adult, or senior citizen).
225 Human Landing Catches: During human observations, another researcher collected
226 human-seeking mosquitoes via the human landing catch (HLC) method at the same
227 location. Each HLC visit exposed the researcher for fifteen minutes. To mitigate actual
228 biting events, the researcher would expose only one limb (arm or leg) at a given time.
229 Any mosquitoes that landed would be collected via mechanical aspirator and transferred
230 to a collection vial. All collected mosquitoes were transported to the NWMAD within 2
231 hours and stored at -80°C. All mosquito specimens were identified to species within three
232 days. Any mosquitoes identified as *Culex* spp. were sent to the Fritz Lab at the University
233 of Maryland for species confirmation by *Cx. pipiens* group-specific primers via PCR.
234 Vector Index: The vector index (VI) was calculated as an estimate of the relative number
235 of WNV-infected mosquitoes. Specifically for this study, VI was calculated as the
236 average number of pooled *Culex* spp. collected per trap-week multiplied by the
237 proportion of mosquitoes infected with WNV. The following equation was modified from
238 the CDC (2013):
239 $$\text{VI} = \sum_{i=Culex\ \text{spp.(pooled)}} \overline{N}_i \widehat{P}_i,$$
240 where $\overline{N}_i$ = average density (number of mosquitoes per trap week) and $\widehat{P}_i$= estimated
241 MIR (proportion of mosquito pools testing positive for WNV). Calculated weekly VI for
242 each trap by week was then interpolated via IDW method for estimations across the
243 NWMAD.

Nuisance Factor and Human WNV Added Risk: Mosquitoes collected via HLC methods were categorized into one of two types: nuisance and WNV vectors. Since the majority of mosquitoes collected were non-*Culex*, a quantitative index, nuisance factor, was created to provide a risk spectrum of encountering nuisance mosquitoes in a given hexagon. The following equation defines the nuisance factor:

$$Nuisance\ Factor = \frac{\frac{Human\ Observations}{Hour} * \frac{Nuissance\ Mosquitoes}{Hour}}{100}$$

Nuisance factor values ranged from a low of 0 to a high of 32.3. To quantitatively estimate potential risk for exposure to disease within a given hexagon, the human WNV added risk factor was also created. This index is defined by the following equation:

$$Human\ WNV\ Added\ Risk = \frac{\frac{Human\ Observations}{Hour} * \frac{Culex\ \text{spp.}}{\text{Hour}}}{100}$$

Human WNV added risk ranged from a low of 0 to a high of 1.44.

2.3 Statistical methods

2.3.1 Location selection

Of the total 1019 hexagons within the NWMAD, we selected fifty-five (5.4%) as the maximum number of sites that our research team could visit for fifteen minutes each, weekly. The subset of fifty-five hexagons were selected based on two criteria: (1) the size of the human population was > 0, and (2) where the previous Cook/DuPage model either predicted human WNV extremely well or extremely poorly, as determined by the 2005-2016 average residual output. Furthermore, the residual output was stratified by those locations that had or had not experienced a human case during the 12-year period. These processes created a performance spectrum consisting of five categories of hexagons: negative residuals without a human case (NR0), low residuals without a case (LR0), low

residuals with a case (LR1), positive residuals without a case (PR0), and positive residuals with a case (PR1) (Table 2). No hexagons with negative residuals in the Cook/DuPage model had experienced a human case. The spatial arrangements of these hexagons provide adequate coverage of the NWMAD's jurisdiction (Figure 2).

2.3.2 Model Selection

Two seasons of field collections and processing of new data provided the UFS model with an additional 42 covariates not made available in the previous Cook/DuPage model. The generation of linear and logistic regression models began with a two-step selection process for the initial covariate inclusion: (1) conduct a univariate analysis with each predictor (independent variable) to the WNV disease outcome (binary = logistic, case rate = linear, dependent variable). Candidate variables for multivariate analysis were selected using slightly more conservative p-value than Bursac et al. (2008), p-value ≤ 0.20 vs. ≤ 0.25). Models that create cut-off values of p-value ≤ 0.1 for purposeful univariate covariate selection can erroneously prevent important variables from entering final models (Bendel and Afifi 1977; Greenland and Mickey 1989); (2) the final model, a generalized linear model personality with a Poisson distribution and probit link function, was selected using forward selection method, selecting the final model based on the Bayesian information criterion (BIC). Non-significant covariates were removed from the final model as a product of the iterative selection process. Secondarily, a receiver operating characteristic (ROC) curve was used to visualize overall model performance and Area Under the Curve (AUC) was calculated. All predictors were evaluated for multicollinearity using the PROC REG procedure (SAS Institute Inc. Cary, NC, USA). Regression analyses were analyzed using the Fit Model feature in JMP 14.2.0 (SAS

289 Institute Inc. Cary, NC, USA). Binary WNV case outcome was analyzed with as a
290 nominal logistic personality. The continuous WNV case rate outcome was analyzed as a
291 standard least squares personality.

292 2.3.3 Model Comparisons

293 Human WNV illness in the NWMAD was assessed under four model environments,
294 each expressing a defined set of specific parameters. The four model environments were:

295 1. MIR & Mosquito Abundance (contains no VI covariates),
296 2. Vector Index (contains no MIR or mosquito abundance covariates),
297 3. Best-Fit (best fit with all covariates in respective assessment), and
298 4. Global (all covariates made available in respective assessment)

299 As a comparison, the original Cook/DuPage model was fit using only the 40 covariates
300 included in the final model fit from Karki et al. (2019). Each of these four model
301 environments were assessed using four different covariate sets:

302 1. All covariates (82 available covariates),
303 2. Excluding HLC and human observations covariates (74 available covariates),
304 3. Force-fitting HLC and human observations covariates (8 forced covariates, 82
305 available covariates), and
306 4. Only the covariates made available to the Cook/DuPage 2019 model (control
307 model, 40 available covariates).

308 Under each model environment and covariate set, the outcome of human WNV illness
309 was analyzed using:

310 1. Logistic regression (presence/absence human WNV illness) and
311 2. Linear regression (WNV case rate) methods.

312 In total, there were 36 models assessed; models are named using the convention $E_xC_yO_z$,
313 where x is the model environment number (0-4, with number 0 assigned to the control
314 environment), C is the covariate set number (1-4), and O is the outcome number (1-2).
315 For both logistic and linear regression, each of the four model environments was fit using
316 each of the four covariate sets. In addition, the control models using only the covariates
317 from the final Cook/DuPage model applied to the UFS region were fit with and without
318 force fitting HLC and human observation covariates.
319 Half of the models were assessed under logistic and linear outcomes, respectively, and
320 based on the *# of Significant Covariates* (quantity of variables included in final model
321 with $p<0.05$) and *Degrees of Freedom* (the number of values in the final model that are
322 free to vary). Overall model performance was determined by BIC. While BIC and
323 Aikake's Information Criterion (AIC) are both maximum likelihood estimators, BIC was
324 chosen to determine model strength due to its stronger penalty term for covariate
325 inclusion (Schwarz 1978).
326 2.3.4 Covariate Performance
327 Similarly to the model performance index, to evaluate the performance for all covariates
328 across 18 logistic and 18 linear models, each of the 82 covariates were standardized by
329 creating the following index:
330 $$\bar{p}_{\text{Covariate}} = \frac{Significance\ Level}{Data\ Availability}$$
331 where: *Significance Level* = significance level of covariate in each of the 36 final models
332 ($p<0.001 = 4$, $p<0.01 = 3$, $p<0.05 = 2$, included in the final model = 1), and *Data*
333 *Availability* = tradeoff between resources required to acquire a respective covariate (level
334 1 = data widely available, no processing needed, level 2 = data available, requires

minimal to moderate processing/analyses, level 3 = data available, requires extensive processing/analyses, level 4 = data not available, needs to be collected, processed, and analyzed, Table S1). The final net prediction:availability tradeoff used to create the Data Availability variable are categorical and based on the authors' personal experiences with data used in this study.

3. Results

3.1 Location Description

The UFS study area contained a total of forty humans WNV cases from 2005-2016 (Table 2).

3.2 Model Fitting

With the exception of model $E_4C_30_1$, all models successfully converged (Tables 3 and 4), with AUC for the logistic models ranging from 0.84 to 0.97 and BIC values of 576 to 769, while BIC values for linear regression models ranged from -227444 to -181982. Despite converging, all global models (n=8) were excluded from the analysis due to statistical overfitting.

3.2 Model Comparison

The highest performing WNV human risk models were $E_3C_4O_3$ (Cook/DuPage Best Fit, df = 8, BIC = -227444) and $E_2C_4O_1$ (Cook/DuPage + VI, df = 14, BIC = 576.2), for linear and logistic regressions, respectively (Tables 3 & 4).

The top five models that predicted human WNV cases strongest were represented by the control (E_0, n=2), best-fit (E_3, n=2) and vector index (E_2, n=1) environments (Figures 3B, Table 5). These models' corresponding covariate sets were represented by variables only

available to the original Cook/DuPage models (C_4, n=4), and force-fitting HLC covariates (C_3, n=1) environments.

3.3 Covariate Performance

Of the 82 available covariates, 70 (85.4%) were included at least once among a given model, excluding the overfit global models. Of the 41 covariates (58.6%) that were greater than the mean covariate performance, seven were highly efficient (determined by natural break in the distribution), providing a crude estimation as most valuable variables for human WNV estimation (Figure 3A). These covariates are provided here in descending order of most importance: tempc (temperature (°C), $\bar{p}$ = 1.15), preci (precipitation (mm.), $\bar{p}$ = 1.14), Yr (year, $\bar{p}$ = 1.0), templag3 (temperature lagged by 3 weeks, $\bar{p}$ = 0.92), blpct (barren land (%), $\bar{p}$ = 0.92), precilag1 (precipitation lagged by 1 week, $\bar{p}$ = 0.90), and VIlag4 (vector index 4 weeks prior, $\bar{p}$ = 0.88). All eight HLC and human observation covariates were included in a final model, but none performed highly ($\bar{p}_{\text{each HLC Covariate}}$ = 0.25). Estimates and calculations for individual covariates are available in Additional file 1.

The eight HLC and human observation covariates provided significant differences in observations and mosquito collections by hexagon type (Figure 4A & 4B). The indices, nuisance mosquito exposure and human WNV added risk, significantly differed by hexagon type (Figure 4C). Hexagons designated as PR1 (positive residual (underpredicted actual cases) with a prior human WNV case) were found to have the most human observations and collected mosquitoes (from both *Culex* and non-*Culex* spp.) per visit. This combination of factors provides hexagons among the PR1 type as the

most “risky” in regard to human WNV added risk and increased nuisance mosquito exposure (Figure 5).

4. Discussion

With the exception of $E_0C_4O_2$, the Cook/DuPage control models, in conjunction with all other covariate and outcome sets, were consistently ranked moderate to low in WNV predictability and net value. Despite excellent prediction capabilities for the larger Cook/DuPage counties study area, this finding suggests that the UFS study areas have more variance from unaccounted sources that are missed or oversimplified in traditional, large-scale models.

In addition to model comparisons, this study evaluated the performance of the newly acquired VI in comparison to the previously used MIR in combination with mosquito abundance. The original Cook/DuPage model only used MIR and its associated 4-week lags and achieved very good prediction results over the 2-county area. Overall, when fit to the UFS study area, adding mosquito abundance and associated 4-week lags improved this model. When evaluating WNV prediction as a linear outcome, the best-fit model using only covariates available to the original Cook/DuPage model was the highest performing in WNV predictability. However, when evaluating WNV prediction as a binary outcome, VI (a product of MIR and abundance together) and its associated 4-week lags replaced MIR as the best predictor of human WNV. While no model emphasizing MIR and abundance was selected as one of the best predictive models, at least one of these variables (and their associated lags) were represented in 4 of the 5 best models (control and best-fit, n=2 for each model). On the contrary, VI, as an emphasized model environment, was selected as the best performing logistic model. Both MIR and VI are

402 critical components in predicting WNV. Deciding between the two biological indicators
403 will be largely dependent upon the data availability for each model of interest. However,
404 if resources are limited, the net model value leans in favor of using MIR.
405 The addition of 42 new covariates required a significant allocation of resources but
406 provided minimal benefits towards reducing variance in human WNV prediction.
407 Fortunately, this study suggests that excellent disease prediction models can be achieved
408 with conventional covariates that are publicly available, requiring little to no processing
409 and/or analyses (data availability scores ≤ 2, Figure 3B). However, any covariate used
410 should be adjusted and properly designed for the highest spatial and/or temporal
411 resolution possible, which may require additional efforts to accomplish.
412 Extensive review of literature indicated no other studies have evaluated covariate
413 strength given limited resources, particularly in the context of making decisions to
414 acquire data. Therefore, the categorizations of covariates by resource allocation (values
415 ranging from 1 (low) to 4 (high)) are based on the experiences of the authors during this
416 study. These values are subjective and may vary across institution or research group, but
417 they may be used as a general estimation in model selection and decision-making. For
418 example, variables related to building and lot size (avg bldg. area: avg lot area, bldg.
419 footprint area avg, bldg. footprint area total, bldg. footprint peri avg., bldg. footprint peri
420 total, and total bldg. area: total lot area) were all ranked a value of 4 because of extensive
421 data processing and review. The authors downloaded high resolution, cloud-free satellite
422 images that were used as a basemap for digital tracing of every building structure (houses,
423 businesses, sheds, detached garages, storage units, etc.) and lots (residential and
424 commercial). This resulted in >47,000 structures and lots digitally traced manually. On

425 the other hand, weather variables (e.g. preci, tempc) were ranked a value of 1 because
426 very little resources were devoted to have the data in a "ready" state. The source of these
427 data, PRISM Climate Group, allows for monthly summaries to be downloaded and
428 extracted with one quick geostatistic process.
429 This study also aimed to address a key missing index that few studies have evaluated:
430 the relationship of human activity, mosquito exposure, and WNV disease risk. While the
431 related variables did not greatly impact overall model strength, they did provide key
432 insight into a potential key in WNV ecology – the areas that were previously
433 underpredicted with recorded human WNV (hex type: PR1) were consistently found to
434 have the most human activity at crepuscular times, the most mosquitoes overall, and the
435 most *Culex* mosquitoes. However, our results appear to contradict the findings of Read et
436 al. (1994), who discovered that as reports of biting nuisance mosquitoes increased beyond
437 2 per minute, outdoor human activity rapidly declined. Our results indicate that as
438 mosquito collections increased, human observations also increased (Figure 4). Not only is
439 this a potentially dangerous combination that can foster environments ideal to mosquito-
440 human spillover, previous modeling efforts failed to capture these cases. Future
441 directions will target these highly susceptible locations and aim to capture any additional
442 unaccounted variance.
443 Like all disease modeling efforts, there are always reporting biases that directly affect
444 true case prevalence. Unfortunately, many vector-borne diseases are largely
445 underreported (Bowden et al. 2011, Nelson et al. 2015, Waterman et al. 2015, CDC 2018),
446 as human cases are vastly overlooked or misdiagnosed, largely due to low severity in
447 disease manifestation in a majority of cases (CDC 2015, Rosenberg et al. 2018). This

448 creates difficulties in predicting when and where VBD incidence will arise. In the
449 Chicago area, models in both the UFS and Cook/DuPage locations have very good
450 human WNV prediction capabilities. Despite having among the highest total number of
451 human WNV cases in the U.S. (CDC National arboviral surveillance system (ArboNET,
452 CDC 2020)), this region has more observational units denoted as non-cases than cases.
453 That has resulted in models with excellent accuracy in predicting where there are no
454 human cases, thus inflating the true accuracy of our models. Nonetheless, while our
455 models are able to reliably predict where human cases are present, the magnitude of
456 effect can be missed (e.g. "hot spots" with greater than 1 case may not be represented).
457 Disease modelers need to be cognizant of saturating their efforts, both statistically and
458 biologically. Statistically, additional and meaningful covariates will usually improve
459 model fit parameters. However, the inclusion of too many variables can result in
460 overfitting, resulting in models failing to converge (Babyak 2004, Hawkins 2004, Lever
461 et al. 2016). It is estimated that about 80% of human WNV infections are unreported, as
462 clinical signs are minor or asymptomatic (CDC 2010, Petersen, Brault, et al. 2013). The
463 remaining 20% of humans develop West Nile fever, and among this group, about 1% will
464 develop severe and sometimes fatal neuroinvasive disease. It is possible that no matter
465 the amount of effort to improve model fit, there is an element of variability attributed
466 with infected humans not seeking medical attention and thus, reducing true disease
467 prevalence (Petersen, Carson, et al. 2013).
468 Overall, when compared to the Cook/DuPage model, the best UFS models required
469 fewer predictors and produced a stronger overall fit using most, if not all, the same
470 covariates made available to both model types. Spending the resources (time, money,

471 human-power, processing, analyses, logistic, etc.) to acquire additional covariates may
472 not necessarily be worth the impact on improving human WNV modeling predictions.
473 Rather, fine-tuning the traditional covariates (climatic, weather, and MIR, for example),
474 to the highest spatiotemporal resolution possible may be the most efficient use of
475 resources to minimize variance in VBD prediction models.

Key Takeaways

1. The factors and their overall effect on the prediction of human WNV cases differs across scale. Although improved, in comparison to the control Cook/DuPage model applied to the same study region, the "best fit" UFS model AUC = 0.89, suggesting newly unaccounted variances are present.
2. Both vector index and MIR contribute to high performing human WNV prediction models under UFS study areas. In direct comparison, VI is favorable to MIR. However, given limited resources in acquiring and processing additional data, MIR is more efficient for predicting human WNV illness.
3. The effort and resources required to acquire additional covariates, most of which are not publicly available, demonstrate a slight improvement in model prediction and appear less important in reducing variance.
4. In addition to the conventional WNV covariates, namely weather and infection rates, land-use and land-cover and SES/demographic information are widely available with little to no processing or analyses required, and provide the breadth to develop excellent prediction models. However, any covariate utilized must be structured at the finest spatial and/or temporal resolution possible.
5. Human exposure to mosquito biting rates provided minimal benefits to model prediction. More importantly however, these two covariates provided potentially key insight to the susceptibility of humans in locations where WNV is prevalent. Additionally, where WNV is less of a concern, these results provide insight into nuisance mosquito exposure that may lead to improvements in targeted control efforts.

Declarations

Acknowledgements

The authors would like to thank Roger Nasci for providing expert insight in vector biology and modeling efforts, Dan Bartlett for providing geospatial data and detailed field information, the Megan Fritz lab for providing confirmatory *Culex* identification, and Chris Stone and Andrew Mackay from the Illinois Medical Entomology lab for providing guidance and expertise with human landing catch methodology.

Ethics Statement

All data collected from the IDPH were through a user agreement. The human activity observation protocol was approved by the University of Illinois Institutional Review Board. Field collections and any use of generated data were approved by the University of Illinois Biosafety Committee.

Availability of Data and Materials

The dataset supporting the conclusions of this article is available in the [repository name] repository, [unique identifier and hyperlink to dataset in http://format].

Authors' contributions

JAU conceived the presented idea, collected field samples and provided data analysis and processing. PI provided research assistance and expertise in mosquito collection and biology. SK provided expertise and additional datasets for analysis. WMB provided analytical assistance and provided data sources. BL provided statistical oversight and expertise. MOR provided the initial product idea, planning, and supervision. RLS provided oversight of all aspects throughout the study. All authors discussed the results and contributed to the final manuscript.

Funding

This publication was supported by Cooperative Agreement #U01 CK000505, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers of Disease Control and Prevention or the Department of Health and Human Services.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

Babyak, M. A. 2004. What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. Psychosom. Med. 66: 411–421.

Bendel, R. B., and A. A. Afifi. **1977**. Comparison of stopping rules in forward "stepwise" regression. J. Am. Stat. Assoc. 72: 46–53.

Bowden, S. E., K. Magori, and J. M. Drake. **2011**. Regional differences in the association between land cover and West Nile virus disease incidence in humans in the United States. Am. J. Trop. Med. Hyg. 84: 234–238.

Bowen, R. A., and N. M. Nemeth. **2007**. Experimental infections with West Nile virus. Curr. Opin. Infect. Dis.

Bursac, Z., C. H. Gauss, D. K. Williams, and D. W. Hosmer. **2008**. Purposeful selection of variables in logistic regression. Source Code Biol. Med. 3: 1–8.

CDC. **2013**. West Nile Virus in the United States: Guidelines for Surveillance, Prevention, and Control. Appendix 2: Calculation and Application of a Vector Index (VI) Reflecting the Number of WN Virus Infected Mosquitoes in a Population. 4th Revision.

CDC. **2015**. National Notifiable Disease Surveillance System (NNDSS): Arboviral Diseases, Neuroinvasive and Non-neuroinvasive 2015 Case Definition.

CDC. **2018**. West Nile virus Surveillance Resources: ArboNET.

CDC. **2019**. National Notifiable Diseases Surveillance Systems (NNDSS): MMWR Week Fact Sheet.

CDC. **2020**. National arboviral surveillance system (ArboNET), Arboviral Diseases Branch, Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention (CDC). **2010**. Morbidity and Mortality Weekly Report Surveillance for Human West Nile Virus Disease — United States , 1999 – 2008. MMWR Wkly. Rep. 59: 1999–2008.

County, C. 2019. Cook County Government Open Data. Cook Cty. Data Cat.

Environmental Systems Research Insititute. **2011**. ArcGIS Desktop. Version 10.5.1 Redlands, CA.

Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M., Elvidge, C. D., Baugh, K., Portnov, B., Rybnikova, N. A., Furgoni, R. 2016a. The new world atlas of artificial night sky brightness. Sci. Adv.

Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M., Elvidge, C. D., Baugh, K., Portnov, B., Rybnikova, N. A., Furgoni, R. 2016b. Supplement to the New World Atlas of Artificial Night Sky Brightness. GFZ Data Serv.

Giordano, B. V., S. Kaur, and F. F. Hunter. **2017**. West Nile virus in Ontario, Canada: A twelve-year analysis of human case prevalence, mosquito surveillance, and climate data. PLoS One. 12: 1–15.

Goddard, L. B., A. E. Roth, W. K. Reisen, and T. W. Scott. **2002**. Vector competence

of California mosquitoes for West Nile virus. Emerg. Infect. Dis. 8: 1385–1391.

Greenland, S., and R. M. Mickey. **1989**. The impact of confounder selection criteria on effect estimation. Am. J. Epidemiol. 130: 1066.

Hadfield, J., A. F. Brito, D. M. Swetnam, C. B. F. Vogels, R. E. Tokarz, K. G. Andersen, R. C. Smith, T. Bedford, and N. D. Grubaugh. **2019**. Twenty years of West Nile virus spread and evolution in the Americas visualized by Nextstrain. PLoS Pathog. 15: 1–18.

Hahn, M. B., A. J. Monaghan, M. H. Hayden, R. J. Eisen, M. J. Delorey, N. P. Lindsey, R. S. Nasci, and M. Fischer. **2015**. Meteorological conditions associated with increased incidence of west nile virus disease in the United States, 2004-2012. Am. J. Trop. Med. Hyg. 92: 1013–1022.

Hamer, G. L., U. D. Kitron, T. L. Goldberg, J. D. Brawn, S. R. Loss, M. O. Ruiz, D. B. Hayes, and E. D. Walker. **2009**. Host selection by Culex pipiens mosquitoes and west nile virus amplification. Am. J. Trop. Med. Hyg. 80: 268–278.

Hawkins, D. M. 2004. The Problem of Overfitting. J. Chem. Inf. Comput. Sci. 44: 1–12.

Karki, S., Brown, W. M., Uelmen, J. A., Ruiz, M. O., Smith, R. L. 2019. The drivers of West Nile virus human illness: fine scale dynamic effects of weather, mosquito infection, social, and biological conditions. bioRxiv.

Kilpatrick, A. M., L. D. Kramer, M. J. Jones, P. P. Marra, and P. Daszak. **2006**. West Nile virus epidemics in North America are driven by shifts in mosquito feeding behavior. PLoS Biol. 4: 606–610.

592 **Kilpatrick, A. M., and W. J. Pape**. **2013**. Predicting human west nile virus infections
593 with mosquito surveillance data. Am. J. Epidemiol. 178: 829–835.

594 **Lever, J., M. Krzywinski, and N. Altman**. **2016**. Points of Significance: Model
595 selection and overfitting. Nat. Methods. 13: 703–704.

596 **Manore, C. A., J. K. Davis, R. C. Christofferson, D. M. Wesson, J. M. Hyman, and C.**
597 **N. Mores**. **2014**. Towards an Early Warning System for Forecasting Human
598 West Nile Virus Incidence. PLoS Curr. 1–21.

599 **Nelson, C. A., S. Saha, K. J. Kugeler, M. J. Delorey, M. B. Shankar, A. F. Hinckley,**
600 **and P. S. Mead**. **2015**. Incidence of clinician-diagnosed lyme disease, United
601 States, 2005–2010. Emerg. Infect. Dis. 21: 1625–1631.

602 **Oregon State University**. **2019**. PRISM Climate Group.
603 (http://prism.oregonstate.edu).

604 **Petersen, L. R., A. C. Brault, and R. S. Nasci**. **2013**. West Nile virus: Review of the
605 literature. JAMA - J. Am. Med. Assoc. 310: 308–315.

606 **Petersen, L. R., P. J. Carson, B. J. Biggerstaff, B. Custer, S. M. Borchardt, and M. P.**
607 **Busch**. **2013**. Estimated cumulative incidence of West Nile virus infection in US
608 adults, 1999-2010. Epidemiol. Infect. 141: 591–595.

609 **Read, N. R., J. R. Rooker, and J. P. Gathman**. **1994**. Public perception of mosquito
610 annoyance measured by a survey and simultaneous mosquito sampling. J. Am.
611 Mosq. Control Assoc. 10: 79–87.

Roiz, D., S. Ruiz, R. Soriguer, and J. Figuerola. **2014**. Climatic effects on mosquito abundance in Mediterranean wetlands. Parasites and Vectors. 7: 1–13.

Rosà, R., G. Marini, L. Bolzoni, M. Neteler, M. Metz, L. Delucchi, E. A. Chadwick, L. Balbo, A. Mosca, M. Giacobini, L. Bertolotti, and A. Rizzoli. **2014**. Early warning of West Nile virus mosquito vector: Climate and land use models successfully explain phenology and abundance of Culex pipiens mosquitoes in north-western Italy. Parasites and Vectors. 7: 1–12.

Rosenberg, R., N. P. Lindsey, M. Fischer, C. J. Gregory, A. F. Hinckley, P. S. Mead, G. Paz-Bailey, S. H. Waterman, N. A. Drexler, G. J. Kersh, H. Hooks, S. K. Partridge, S. N. Visser, C. B. Beard, and L. R. Petersen. **2018**. Vital signs: Trends in reported vectorborne disease cases — United States and Territories, 2004-2016. Morb. Mortal. Wkly. Rep. 67: 496–501.

Ruiz, M. O., C. Tedesco, T. J. McTighe, C. Austin, and K. Uriel. **2004**. Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. Int. J. Health Geogr. 3.

Russell, C., and F. F. Hunter. **2012**. Culex pipiens (Culicidae) is attracted to humans in southern Ontario, but will it serve as a bridge vector of West Nile virus? Can. Entomol.

Schwarz, G. 1978. Estimating the Dimension of a Model. Ann. Stat. 6: 461–464.

Sejvar, J. J. 2003. West Nile virus: An historical overview. Ochsner J. 5: 6–10.

Smithburn, K. C., T. P. Hughes, A. W. Burke, and J. H. Paul. **1940**. A Neurotropic

633 Virus Isolated from the Blood of a Native of Uganda 1. Am. J. Trop. Med. Hyg. s1-
634 20: 471–492.

635 **USDA**. **2018**. National Agriculture Imagery Program (NAIP) Imagery.
636 (https://catalog.data.gov/dataset/national-geospatial-data-asset-ngda-naip-
637 imagery).

638 **USGS**. **2011**. National Geospatial Data Asset (NGDA) Land Use Land Cover.

639 **USGS**. **2019**. EarthExplorer. (https://earthexplorer.usgs.gov/).

640 **Waterman, S. H., H. S. Margolis, and J. J. Sejvar**. **2015**. Surveillance for dengue and
641 dengue-associated neurologic syndromes in the United States. Am. J. Trop. Med.
642 Hyg. 92: 996–998.

643 **Wimberly, M. C., A. Lamsal, P. Giacomo, and T. W. Chuang**. **2014**. Regional
644 variation of climatic influences on West Nile virus outbreaks in the United
645 States. Am. J. Trop. Med. Hyg. 91: 677–684.

646

647

Table 1. List of covariates used previously in Cook/DuPage Counties WNV model and those newly acquired variables used in newly revised 55 hexagon UFS model.

Covariate Information				Cook/DuPage Model	Ultra-fine-scale Model
Designation		**Description**	**Notation**		
Environmental	Land Cover	Proportion of developed open space	dospct	X	X
		Proportion of developed low intensity	dlipct	X	X
		Proportion of developed medium intensity	dmipct	X	X
		Proportion of developed high intensity	dhipct	X	X
		Proportion of deciduous forests	dfpct	X	X
		Proportion of evergreen forests	efpct	X	X
		Proportion of mixed forests	mfpct	X	X
		Proportion of barren land	blpct	X	X
		Proportion of shrubs	shrubspct	X	X
		Proportion of grassland	glandpct	X	X
		Proportion of pasture	pasturepct	X	X
		Proportion of cultivated land	clpct	X	X
		Proportion of woody wetlands	wwpct	X	X
		Proportion of herbaceous wetlands	hwpct	X	X
		Proportion of total forest	ftotpct		X
		Proportion of total wetlands	wtotpct		X
		Proportion of open water	owpct	X	X
		Normalized Difference Vegetation Index	NDVI		X
Biological	Minimum Infection Rate (MIR)	MIR one week before	mirlag1	X	X
		MIR two weeks before	mirlag2	X	X
		MIR three weeks before	mirlag3	X	X
		MIR four weeks before	mirlag4	X	X
		Average MIR current week	MIRmean		X
		Difference in weekly average MIR from 12-year average	MIRdiff		X
	Mosquito Abundance	Vector Index current week	Vector Index		X
		Vector Index one week before	VIlag1		X
		Vector Index two weeks before	VIlag2		X
		Vector Index three weeks before	VIlag3		X
		Vector Index four weeks before	VIlag4		X
	Mosquito Biting Rates (HLC)	Light and gravid trap collection mean current week	Trap_Mean		X
		Light and gravid trap collection mean one week before	Trap_Meanlag1		X
		Light and gravid trap collection mean two weeks before	Trap_Meanlag2		X
		Light and gravid trap collection mean three weeks before	Trap_Meanlag3		X
		Light and gravid trap collection mean four weeks before	Trap_Meanlag4		X
		Mosquitoes per visit	mosquitoes per visit		X
		Culex spp. per visit	Cx per visit		X
Weather	Temperature	Average temperature current week	tempc		X
		Average temperature of one week before	templag1	X	X
		Average temperature of two weeks before	templag2	X	X
		Average temperature of three weeks before	templag3	X	X
		Average temperature of four weeks before	templag4	X	X
	Precipitation	Mean January temperature	Jantemp	X	X
		Average precipitation current week	preci		X
		Average precipitation of one week before	precilag1	X	X
		Average precipitation of two weeks before	precilag2	X	X
		Average precipitation of three weeks before	precilag3	X	X
		Average precipitation of four weeks before	precilag4	X	X
Anthropogenic	Socio-demographic	Percentage of White population	whitepct	X	X
		Percentage of African American population	blackpct	X	X
		Percentage of Asian population	asianpct	X	X
		Percentage of Hispanic population	hispanicpct	X	X
		Median household income	Income	X	X
		Percentage of housing constructed before WWII	hpctpreww	X	X
		Percentage of housing constructed post WWII (1945-1969)	hpctpostww	X	X
		Percentage of housing constructed from 1970-1989	hpct7089	X	X
		Percentage of housing constructed in 1990 or later	hpctpost90	X	X
	Land change & manipulation	Catch basin density	CB		X
		Total area of building structures	bldg.._footprint_area_total		X
		Average area of building structures	bldg._footprint_area_avg		X
		Total perimeter of building structures	Building_Footprint_peri_total		X
		Average perimeter of building structures	Building_Footprint_peri_avg		X
		Total area of residential lot	Residential_lot_area_total		X
		Average area of residential lot	Residential_lot_area_avg		X
		Total perimeter of residential lot	Residential_lot_peri_total		X
		Average perimeter of residential lot	Residential_lot_peri_avg		X
		Ratio of total building area by total lot area	total_bldg._area/total_lot_area		X
		Ratio of average building area by average lot area	avg_bldg._area/avg_lot_area		X
		Ratio of total building perimeter by total lot area	total_bldg._peri/total_lot_area		X
		Ratio of average building perimeter by average lot area	avg_bldg._peri/avg_lot_area		X
		Number of buildings	buildings		X
		Building density per mi.^{2}	bldg._density		X
	Human population	Number of residents per building	persons_per_bldg.		X
		Total human population	totpop	X	X
	Activity Observations	Mean light pollution	lightpol		X
		Senior Citizen Observations per visit	Senior_obs per visit		X
		Adults Observations per visist	Adults_obs per visit		X
		Children Observations per visit	Child_obs per visit		X
		Male Observations per visit	Male_obs per visit		X
		Female Observations per visit	Female_obs per visit		X
		Total Observations per visit	Total_obs per visit		X
Other		Year	yr	X	X
		Hexagon Designation	hexid	X	X
			Total Covariates Evaluated	**40**	**82**

Table 2. Description of selected hexagons (n=55) by residual categorization (PR1, PR0, NR1, LR1, LR0) within the Northwest Mosquito Abatement District (NWMAD). Descriptions of residual categorizations are as follows: PR = positive residual (underprediction, residuals ≥ 1.0), NR = negative residual (great overprediction, residuals ≤ -1.0), LR = low residuals (prediction close to actual, residuals -1.0 < X < 1.0). Values following the residual categorizations designated as: 1 = at least one human WNV case between 2005-2016; 0 = no human WNV cases between 2005-2016.

HexID	Cases	Residual	Category	Field Season
4349	0	-1.001	NR0	1
4806	0	-1.001	NR0	1
4241	0	-1	LR0	1
4854	0	-1	LR0	1
5250	0	-1	LR0	1
4250	1	-0.271	LR1	1
4471	2	0.877	PR1	1
4183	2	0.902	PR1	1
4984	1	0.912	PR1	1
5188	0	1.134	PR0	1
4597	1	1.531	PR1	1

HexID	Cases	Residual	Category	Field Season
4014	0	-1.001	NR0	Both
4082	0	-1.001	NR0	Both
4217	0	-1.001	NR0	Both
4415	0	-1.001	NR0	Both
4467	0	-1	LR0	Both
5199	0	-1	LR0	Both
5286	0	-1	LR0	Both
4313	3	0.033	LR1	Both
5235	1	0.055	LR1	Both
4609	2	0.399	LR1	Both
4637	1	1.027	PR1	Both
4332	1	1.279	PR1	Both
4335	1	1.767	PR1	Both
4676	1	1.838	PR1	Both
4449	3	1.841	PR1	Both
5239	0	2.198	PR0	Both
4743	2	4.881	PR1	Both
5181	1	17.057	PR1	Both
4617	0	18.013	PR0	Both

HexID	Cases	Residual	Category	Field Season
4242	0	-1.001	NR0	2
4614	0	-1.001	NR0	2
4181	0	-1	NR0	2
4381	0	-1	NR0	2
4382	0	-1	NR0	2
4578	0	-1	NR0	2
4923	0	-1	NR0	2
5185	0	-1	NR0	2
5262	0	-1	NR0	2
5234	1	-0.338	LR1	2
4070	1	-0.32	LR1	2
4952	1	-0.193	LR1	2
4073	1	-0.038	LR1	2
4678	1	-0.014	LR1	2
4135	1	-0.002	LR1	2
4104	1	0	LR1	2
4334	1	0.024	LR1	2
4243	2	0.071	LR1	2
5126	1	0.11	LR1	2
4065	1	0.181	LR1	2
5265	1	0.221	LR1	2
4098	1	5.557	PR1	2
4636	1	6.481	PR1	2
4346	1	6.967	PR1	2
4310	1	17.856	PR1	2

A.

Regression Method[a]	Model Environment	Included Covariates	df	p-value	ROC[b]	BIC	ΔBIC
1. Logistic	1. MIR & Mosquito Abundance ($E_1C_2O_1$)	- tempc - preci + templag1 + templag3* - precilag1* - precilag3 - precilag4 + mir mean - mir diff + mirlag1 + mirlag2 + mirlag3** + mirlag4* + totpop + blackpct + asianpct + dmipct - dhipct + ccpct - hpctpreww - hpctpostww - hpct7089 + abund - abundlag1 + abundlag2 + abundlag3 + abundlag4 + bldg. footprint area avg - bldg. footprint peri avg + resi lot peri total - resi lot peri avg - avg bldg. area:avg lot area - total bldg. peri:total lot area	34	<0.0001	0.89	768.7	128.3
	2. Vector Index ($E_2C_2O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 + totpop + blackpct + asianpct + hispanicpct - Income + dospct - dmipct - hpctpreww - hpct7089 + CB + avg bldg. area:avg lot area*** + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4	23	<0.0001	0.86	661.1	20.7
	3. Best-Fit ($E_3C_2O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag3 + mirlag3** + totpop - whitepct + blackpct - Income + dospct + dmipct - hpctpreww - hpct7089 - abundlag1 - bldg. footprint area total + avg bldg. area:avg lot area** + VIlag4*	21	<0.0001	0.86	640.4	0
	4. Global ($E_4C_2O_1$)	- tempc* - preci* - yr + templag1* + templag2 + templag3 - templag4 – precilag1** - precilag2 - precilag3 – precilag4* + MIRmean - MIRdiff + mirlag1 + mirlag2 + mirlag3* + mirlag4 + totpop - whitepct - blackpct - asianpct - hispanicpct + income - owpct - dospct - dlipct - dmipct - dhipct - blpct - dfpct - mfpct - shrubpct - glandpct - pasturepct - wwpct - dtotpct - ftotpct + wtotpct - Jantemp + hpctpreww + hpctpostww + hpct7089 + hpctpost90 - abund - abundlag1 - abundlag2 + abundlag3 + abundlag4 + CB - bldg. footprint total area + bldg. footprint area avg - bldg. footprint peri total - bldg. footprint peri avg + resi lot area total + resi lot area avg + resi lot peri total - resi lot peri avg + total bldg. area:total lot area + VI - VIlag1 - VIlag2 - VIlag3 - VIlag4 + Light pol + NDVI	65	0.0009	0.97	833.5	193.1
2. Linear	1. MIR & Mosquito Abundance ($E_1C_2O_2$)	mir mean + mir diff + mirlag1 + mirlag2 + mirlag3 + mirlag4* + blpct** + abund - abundlag1 - abundlag2 + abundlag3 + abundlag4* + bldg. footprint area avg***	13	<0.0001	N/A	-182037	3358
	2. Vector Index ($E_2C_2O_2$)	bldg. footprint area avg*** + avg bldg. area:avg lot area*** + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4***	7	<0.0001		-185373	22
	3. Best-Fit ($E_3C_2O_2$)	- mirlag4 + bldg. footprint area avg***+ avg bldg. area:avg lot area** + VIlag4***	4	<0.0001		-185395	0
	4. Global ($E_4C_2O_2$)	- tempc - preci + yr - templag1 + templag2 + templag3 + templag4 - precilag1 + precilag2 - precilag3 - precilag4 + MIRmean - MIRdiff + mirlag1 - mirlag2 + mirlag3 - mirlag4** - totpop - whitepct + blackpct + asianpct + hispanicpct + income + owpct - dospct - dlipct - dmipct - dhipct - blpct - dfpct - mfpct + shrubpct - glandpct + pasturepct + wwpct - dtotpct - ftotpct + wtotpct + Jantemp - hpctpreww - hpctpostww - hpct7089 - hpctpost90 - abund - abundlag1 - abundlag2 + abundlag3 + abundlag4 + CB - bldg. footprint total area - bldg. footprint area avg + bldg. footprint peri total + bldg. footprint peri avg - resi lot area total - resi lot area avg + resi lot peri total - resi lot peri avg + total bldg. area:total lot area + - avg bldg. area:avg lot area - # blgds - VI - VIlag1 - VIlag2 - VIlag3 + VIlag4*** + Light pol +	68	0.0044		-94558.2	90836.8

Table 3. Model fit comparisons of the UFS hexagons, applying **(A)** newly acquired data (excluding HLC and human observations, covariate set 2), or **(B)** only the covariates made available to the previously published Cook/DuPage model (covariate set 4). Each model fit was assessed using logistic (presence/absence WNV human illness case) and generalized linear (WNV case rates, controlling for human population) methods.

B.

Regression Method	Model Environment	Included Covariates	df	p-value	ROC	BIC	ΔBIC
1. Logistic	0. Cook/DuPage ($E_0C_4O_1$)	1_Yr - templag2 + templag3* + templag4* - Jantemp + mirlag1 + mirlag2 + mirlag3 + mirlag4* + totpop - owpct - dlipct - dfpct - glandpct + hpctpost90	15	<0.0001	0.85	632.3	56.1
	1. MIR & Mosquito Abundance ($E_1C_4O_1$)	dhipct - mfpct - wwpct + mirlag1 + mirlag2 + mirlag3* + mirlag4 + templag3 + templag4* - precilag1* - precilag2 - precilag4* + asianpct** + totpop + mir mean - mir diff + abund – abundlag1* + abundlag2 + abundlag3 + abundlag4	21	<0.0001	0.88	653.3	77.1
	2. Vector Index ($E_2C_4O_1$)	dhipct - mfpct - wwpct + templag3* + templag4* - precilag1* - precilag2 – precilag4* + asianpct + VIlag1 + VIlag2 + VIlag3* + VIlag4	13	<0.0001	0.86	576.2	0
	3. Best-fit 55 hex ($E_3C_4O_1$)	dhipct - mfpct - wwpct + mirlag3* + mirlag4 + templag3* + templag4** - precilag1* - precilag2 – precilag4* + asianpct** + totpop	12	<0.0001	0.89	580.8	4.6

649

	4. Global ($E_4C_4O_1$)	Yr - dospct - dlipct - dmipct - dhipct - dfpct - mfpct + blpct - shrubpct - glandpct - pasturepct - ccpct - wwpct - owpct + mirlag1 + mirlag2 + mirlag3* + mirlag4 - templag1 - templag2 + templag3* + templag4* - precilag1* - precilag2 - precilag3 – precilag4* + whitepct - blackpct - asianpct + hispanicpct - Income - totpop + Jantemp	33	<0.0001	0.90	792.0	215.8
2. Linear	0. Cook/DuPage ($E_0C_4O_2$)	Yr - templag2 + templag3 + templag4 + Jantemp + mirlag1 + mirlag2 + mirlag3* + mirlag4** - totpop* - owpct* - dlipct – dfpct* - glandpct + hpctpost90	15	<0.0001	N/A[b]	-227354	90
	1. MIR & Mosquito Abundance ($E_1C_4O_2$)	Dmipct** + blpct** + mirlag1 + mirlag2 + mirlag3 + mirlag4* + templag3 - precilag1 - precilag4 – totpop* - mir mean + mir diff - abund - abundlag1 + abundlag2 - abundlag3 + abundlag4*	17	<0.0001		-182001	45443
	2. Vector Index ($E_2C_4O_2$)	dlipct + dmipct*** + templag3* - totpop** - VI - VIlag1 + VIlag2 + VIlag3* + VIlag4***	9	<0.0001		-185347	42097
	3. Best-fit 55 hex ($E_3C_4O_2$)	Dmipct** + blpct** + mirlag3** + mirlag4** + templag3* - precilag1 - precilag4 – totpop*	8	<0.0001		-227444	0
	4. Global ($E_4C_4O_2$)	Yr - dospct - dlipct - dmipct - dhipct - dfpct - mfpct + blpct - shrubpct - glandpct - pasturepct - ccpct - wwpct - owpct + mirlag1 + mirlag2 + mirlag3* + mirlag4* - templag1 - templag2 + templag3 + templag4 - precilag1 - precilag2 - precilag3 - precilag4 + whitepct - blackpct - asianpct + hispanicpct - Income – totpop* + Jantemp	33	<0.0001		-227199	245

[a]Logistic regression outcome = human WNV presence/absence per hexagon, per week; GLM outcome = WNV human case rate (per hexagon, per week).
[b]ROC applies to only logistic regression.
[c]As the final selected model in the Original Cook/DuPage paper (2019), this model environment was assessed only for the comparison to the Cook/DuPage models for this study and not applied to the UFS model. The original model covariates, eftpct and ehwpct, have 0 observations among the selected 55 hexagons and were removed.

Table 4. Model fit comparisons of the UFS hexagons, using best-fit models with additional human landing catch and human activity observations to incorporate added human risk. Human risk covariates were added to the UFS model by **(A)** best-fit integration (covariate set 1) and **(B)** force-fitting (covariate set 3).

A.

Regression Method	Model Environment	Included Covariates	df	p-value	AUC	BIC	ΔBIC
1. Logistic	1.MIR & Mosquito Abundance ($E_1C_1O_1$)	- tempc - preci + templag1 - templag2 + templag3* - precilag1* - precilag3 - precilag4 + mirmean - mirdiff + mirlag1 + mirlag2 + mirlag3** + mirlag4 - whitepct + blackpct + dospct - dmipct - hpctpreww - hpct7089 + abund - abundlag1 + abundlag2 + abundlag3 + abundlag4 + bldg. footprint area total - total bldg. area:total lot area + male obs per visit + female obs per visit + *Culex* per visit	30	<0.0001	0.87	742.5	107.9
	2. Vector Index ($E_2C_1O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 - whitepct + blackpct + dmipct - hpctpreww - hpctpostww - hpct7089 + male obs per visit + female obs per visit + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4	26	<0.0001	0.86	692.7	58.1
	3. Best-Fit ($E_3C_1O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 + mirlag3** + blackpct - dospct + dmipct - hpctpreww - hpct7089 + abund + Light pol + male obs per visit + female obs per visit + *Culex* per visit + VIlag4	19	<0.0001	0.84	634.6	0
	4. Global[a] ($E_4C_1O_1$)	- tempc* - preci* - yr + templag1* + templag2 + templag3 - templag4 - precilag1** - precilag2 - precilag3 - precilag4* + mirmean - mirdiff + mirlag1 + mirlag2 + mirlag3* + mirlag4 - totpop + whitepct + blackpct + asianpct + hispanicpct + income - owpct - dospct - dlipct - dmipct - dhipct - blpct - dfpct - mfpct - shrubpct - glandpct - pasturepct + wwpct - dtotpct - ftotpct + wtotpct - Jantemp + hpctpreww + hpctpostww + hpct7089 + hpctpost90 - abund - abundlag1 - abundlag2 - abundlag3 + abundlag4 + CB - bldg. footprint total area - bldg. footprint area avg + bldg. footprint peri total + bldg. footprint peri avg - resi lot area total - resi lot area avg - resi lot peri total + resi lot peri avg - total bldg. area:total lot area - # blgds + VI - VIlag1 - VIlag2 - VIlag3 - VIlag4 + Light pol + NDVI + mosquitoes per visit	65	0.0009	0.97	772.8	138.2
2. Linear	1. MIR & Mosquito Abundance ($E_1C_1O_2$)	mir mean + mir diff + mirlag1 + mirlag2 + mirlag3 + mirlag4* + blpct** + abund - abundlag1 - abundlag2 + abundlag3 + abundlag4* + bldg. footprint area avg***	13	<0.0001	N/A	-182037	3352
	2. Vector Index ($E_2C_1O_2$)	templag3* + bldg. footprint area avg*** - resi lot peri avg + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4***	8	<0.0001		-185362	27
	3. Best-Fit ($E_3C_1O_2$)	- mirlag4 + bldg. footprint area avg***- resi lot peri avg + VIlag4***	4	<0.0001		-185389	0
	4. Global[b] ($E_4C_1O_2$)	- tempc - preci + yr - templag1 + templag2 + templag3 + templag4 - precilag1 + precilag2 - precilag3 - precilag4 + mirmean - mirdiff + mirlag1 - mirlag2 + mirlag3 – mirlag4** - totpop + whitepct + blackpct + asianpct + hispanicpct + income - owpct - dospct - dlipct - dmipct - dhipct - blpct - dfpct - mfpct - shrubpct - glandpct - pasturepct + wwpct - dtotpct - ftotpct + wtotpct + Jantemp - hpctpreww - hpctpostww - hpct7089 - hpctpost90 - abund - abundlag1 - abundlag2 - abundlag3 + abundlag4 + CB - bldg. footprint total area - bldg. footprint area avg + bldg. footprint peri total + bldg. footprint peri avg - resi lot area total - resi lot area avg + resi lot peri total - resi lot peri avg - total bldg. area:total lot area - avg bldg. area:avg lot area - # blgds - VI - VIlag1 - VIlag2 - VIlag3 + VIlag4*** + Light pol + NDVI + senior obs per visit	64	0.0022		-94581.4	90797.6

B.

Regression Method	Model Environment	Included Covariates	df	p-value	ROC[b]	BIC	ΔBIC
1. Logistic	0. Cook/DuPage ($E_0C_3O_1$)	- Yr - templag2 + templag3* + templag4* - Jantemp + mirlag1 + mirlag2 + mirlag3* + mirlag4* - totpop - owpct - dlipct - dfpct - glandpct + hpctpost90 - senior obs per visit - adult obs per visit + child obs per visit + male obs per visit - mosquitoes per visit + *Culex* per visit	21	<0.0001	0.86	683.4	10.7
	1. MIR & Mosquito Abundance[c] ($E_1C_3O_1$)	- tempc - preci + templag1 + templag2 - precilag1* - precilag4* - mirmean* - mirdiff + mirlag1 + mirlag2 + mirlag3** - mirlag4* - totpop + blackpct + dospct + hpctpostww + hpct7089 + hpctpost90* + abund - abundlag1 + abundlag2 + abundlag3 + abundlag4 + bldg. footprint peri avg + resi lot area total + avg bldg. peri:avg lot area** + senior obs per visit + adult obs per visit + child obs per visit - male obs per visit + mosquitoes per visit + *Culex* per visit	32	<0.0001	0.87	757.7	61.1
	2. Vector Index ($E_2C_3O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 - whitepct + blackpct - dospct + dmipct + hpctpostww + hpct7089 + hpctpost90 + resi lot area total - senior obs per visit + adult obs per visit + child obs per visit - male obs per visit - mosquitoes per visit + *Culex* per visit + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4	26	<0.0001	0.85	696.6	23.9
	3. Best-fit 55 hex[d] ($E_3C_3O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 - whitepct + blackpct - dospct + dmipct + hpctpostww + hpct7089 + resi lot area total - senior obs per visit + adult obs per visit + child obs per visit - male obs per visit - mosquitoes per visit + *Culex* per visit + VIlag4 + mirlag3** - abund	23	<0.0001	0.84	672.7	0
	4. Global ($E_4C_3O_1$)	Model Failed to Converge	N/A	N/A	N/A	N/A	N/A
2. Linear	0. Cook/DuPage[e] ($E_0C_3O_2$)	Yr - templag2 + templag3 + templag4 + Jantemp + mirlag1 + mirlag2 + mirlag3* + mirlag4** - totpop* - owpct* - dlipct – dfpct* - glandpct + hpctpost90 - senior obs per visit + adult obs per visit + child obs per visit - male obs per visit + mosquitoes per visit - *Culex* per visit	21	<0.0001	N/A	-227300	0
	1. MIR & Mosquito Abundance[e] ($E_1C_3O_2$)	blpct** + bldg. footprint area avg*** - senior obs per visit + adult obs per visit + child obs per visit - male obs per visit - mosquitoes per visit + *Culex* per visit + mirmean + mirdiff + mirlag1 + mirlag2 + mirlag3 + mirlag4* + abund - abundlag1 - abundlag2 - abundlag3 + abundlag4*	19	<0.0001		-181982	45318
	2. Vector Index[f] ($E_2C_3O_2$)	bldg. footprint area avg*** + adult obs per visit + child obs per visit + female obs per visit - total obs per visit - mosquitoes per visit + *Culex* per visit + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4***	12	<0.0001		-185322	41978
	3. Best-fit 55 hex[g] ($E_3C_3O_2$)	bldg. footprint area avg*** - senior obs per visit - adult obs per visit + child obs per visit + female obs per visit - mosquitoes per visit + *Culex* per visit + VIlag4*** - mirlag4	9	<0.0001		-185344	41956
	4. Global[g] ($E_4C_3O_2$)	- preci + Yr - templag1 + templag2 + templag3 + templag4 - precilag1 + precilag2 - precilag3 - precilag4 - totpop + whitepct + blackpct + asianpct + hispanicpct - Income - owpct - dospct - dlipct - dmipct - dhipct + blpct - dfpct - mfpct - shrubpct + glandpct - pasturepct* - wwpct - dtotpct - ftotpct + wtotpct + Jantemp - hpctprewww - hpct7089 - hpctpost90 + CB - bldg. footprint area total - bldg. footprint area avg + bldg. footprint peri total + lightpol + NDVI + senior obs per visit - adult obs per visit - child obs per visit + female obs per visit - mosquitoes per visit + *Culex* per visit - VI - VIlag1 - VIlag2 - VIlag3 + VIlag4*** + mirmean - mirdiff + mirlag1 - mirlag2 + mirlag3 - mirlag4** - abund - abundlag1 - abundlag2 - abundlag3 + abundlag4	63	0.0013		-94601.5	132699

[a]avg bldg. peri:avg lot area, # bldg.s, Observations per visit: senior, adult, child, male, female, total human and Culex spp. were biased and/or zeroed and not assessed.

[b]Observations per minute: adult, child, male, female, total human, mosquito, and Culex spp. were biased and/or zeroed and not assessed.

[e]Observations per minute: female and total human were biased and/or zeroed and not assessed.

[f]Observations per minute: male and senior were biased and/or zeroed and not assessed.

[g]Observations per minute: male and total human were biased and/or zeroed and not assessed.

664 **Table 5.** Detailed assessment of each model evaluated in this study. Overall model strength was determined by BIC value (by linear and logistic regression types), with the following characteristics denoted as follows: Cumulative Significance Total, sum of each variable score, denoted as: $p<0.001 = 4$, $p<0.01 = 3$, $p<0.05 = 2$, included in model = 1; # of Significant Covariates = summation of included covariates with p-value<0.05; DF = degrees of freedom denoted in model; BIC value = overall model rank (best model = 1, worst model = 14)/14 for each logistic and linear model group, respectively.

Model[a]	Cumulative Significance Total	# of Significant Covariates	DF (lower is better)	BIC value	BIC value (lower is better) rank	Regression Type
E0C3O2	6	5	21	-227300	3	Linear
E0C4O2	6	5	15	-227354	2	
E1C1O2	7	4	13	-182037	11	
E1C2O2	7	4	13	-182037	12	
E1C3O2	7	4	19	-181982	14	
E1C4O2	7	5	17	-182001	13	
E2C1O2	8	4	8	-185362	7	
E2C2O2	10	4	7	-185373	6	
E2C3O2	7	3	12	-185322	10	
E2C4O2	10	5	9	-185347	8	
E3C1O2	6	2	4	-185389	5	
E3C2O2	8	3	4	-185395	4	
E3C3O2	6	2	9	-185344	9	
E3C4O2	10	6	8	-227444	1	
E4C1O2	Global Models Excluded					
E4C2O2						
E4C3O2						
E4C4O2						
E0C3O1	4	4	21	683.40	9	Logistic
E0C4O1	3	3	15	632.30	3	
E1C1O1	4	3	30	742.50	12	
E1C2O1	5	4	34	768.70	14	
E1C3O1	9	7	32	757.70	13	
E1C4O1	7	6	21	653.30	6	
E2C1O1	4	3	26	692.70	10	
E2C2O1	7	4	23	661.10	7	
E2C3O1	4	3	26	696.60	11	
E2C4O1	5	5	14	576.20	1	
E3C1O1	5	3	19	634.60	4	
E3C2O1	8	5	21	640.40	5	
E3C3O1	5	3	23	672.70	8	
E3C4O1	8	6	12	580.80	2	
E4C1O1	Global Models Excluded					
E4C2O1						
E4C3O1						
E4C4O1						

[a]All global models were excluded from analysis as they were all overfit and statistically biased

665 **Figure Captions**

666 **Figure 1.** The UFS study area, contained within the Northwest Mosquito Abatement District,
667 shown in relation to Cook & DuPage Counties. Overlaid are 1-km diameter hexagons, the
668 observational units used in this study. Northwest Mosquito Abatement District comprises 1,019
669 of the total 5,345 hexagons in all of Cook & DuPage Counties.

670 **Figure 2.** Location of the 55-hexagon study area within the Northwest Abatement District.
671 Hexagons are labeled by field season visited for mosquito collections and human activity
672 observations (color outline) and by total human cases from 2005-2016 (gray scale shaded
673 interior).

674 **Figure 3.** Mean performance of each of the 70 covariates used in the study (A). Covariates are
675 listed in alphabetic order by data availability/work load to acquire score (1-4). The contribution
676 of each covariate resulted in a net value performance for each linear and logistic model assessed
677 in the study (B). Means for each outcome ($\bar{x}_{\text{covariate}}$ = 0.48; $\bar{x}_{\text{linear}}$ = -193406; $\bar{x}_{\text{logistic}}$ = 670.9) are
678 designated by horizontal dashed lines. Details of scoring for each covariate and model are
679 provided in Tables 5 & S1

680 **Figure 4.** Relationship of hexagon type (LR = low residual, PR = positive residual, NR = large,
681 negative residual; 0 = no human case, 1 = human case) by human observations per visit (A),
682 mosquitoes collected per visit (B), and a product of the two former variables, nuisance factor and
683 WNV added risk (C). Letters above each box and whisker plot designate significantly different
684 groups by hexagon type, as calculated by Tukey's HSD.

685 **Figure 5.** Relationship of each of the 55 UFS study hexagons (y-axis = unique identification
686 number) by nuisance mosquito factor (A) and human WNV added risk (B). Letters above each

687 box and whisker plot designate significantly different groups by hexagon type, as calculated by
688 Tukey's HSD.

Figures

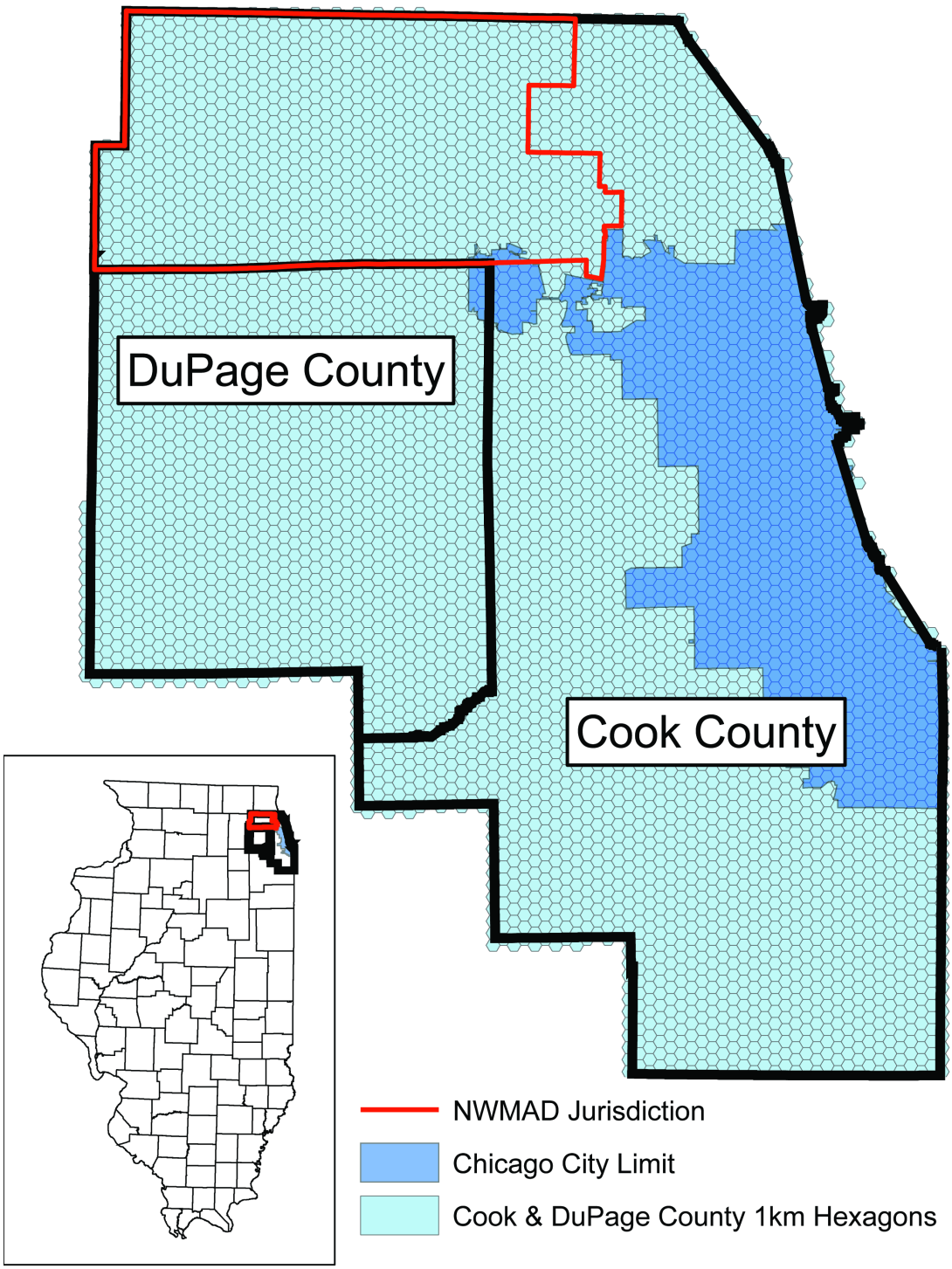

Figure 1

The UFS study area, contained within the Northwest Mosquito Abatement District, shown in relation to Cook & DuPage Counties. Overlaid are 1-km diameter hexagons, the observational units used in this

study. Northwest Mosquito Abatement District comprises 1,019 of the total 5,345 hexagons in all of Cook & DuPage Counties.

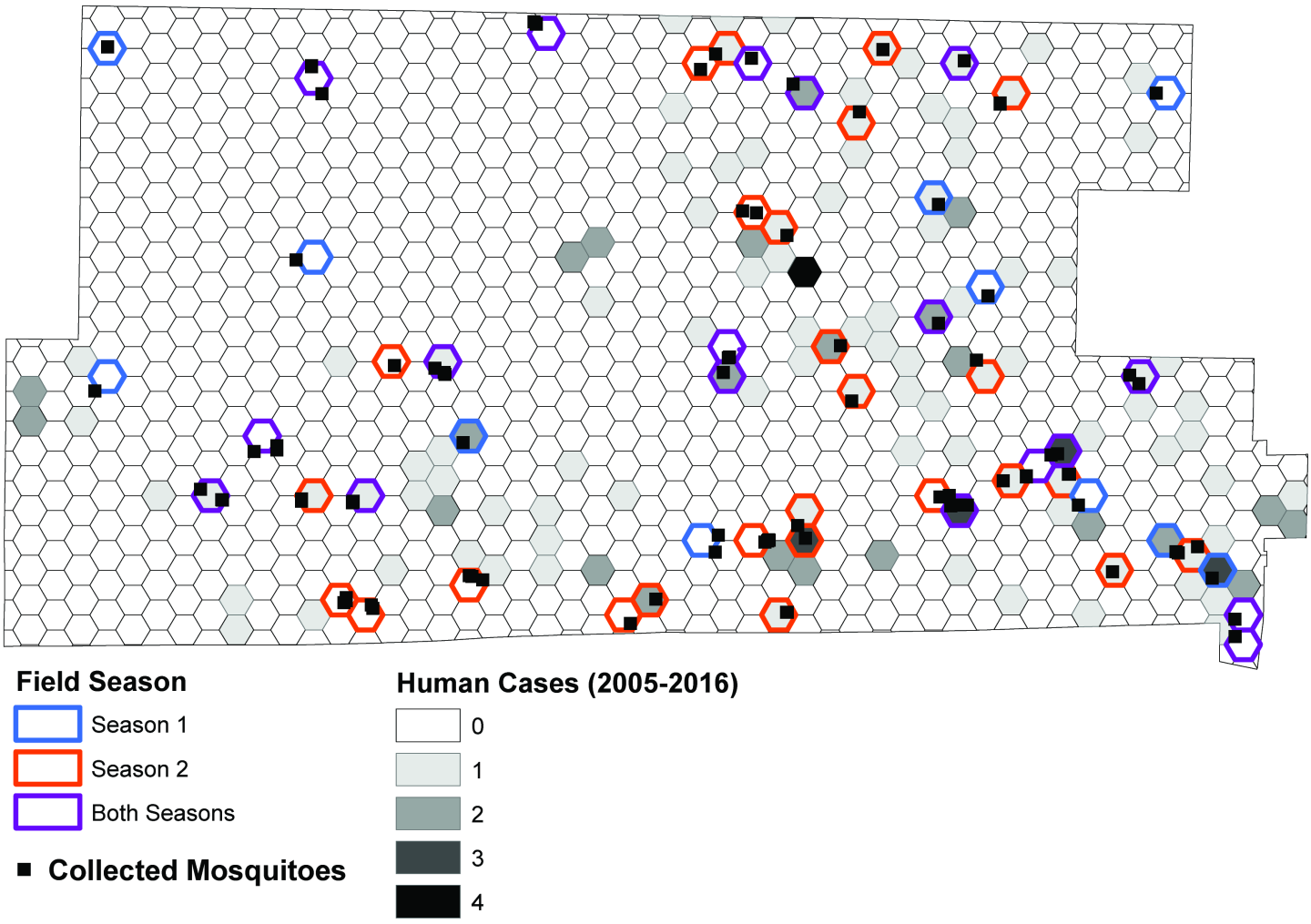

Figure 2

Location of the 55-hexagon study area within the Northwest Abatement District. Hexagons are labeled by field season visited for mosquito collections and human activity observations (color outline) and by total human cases from 2005-2016 (gray scale shaded interior).

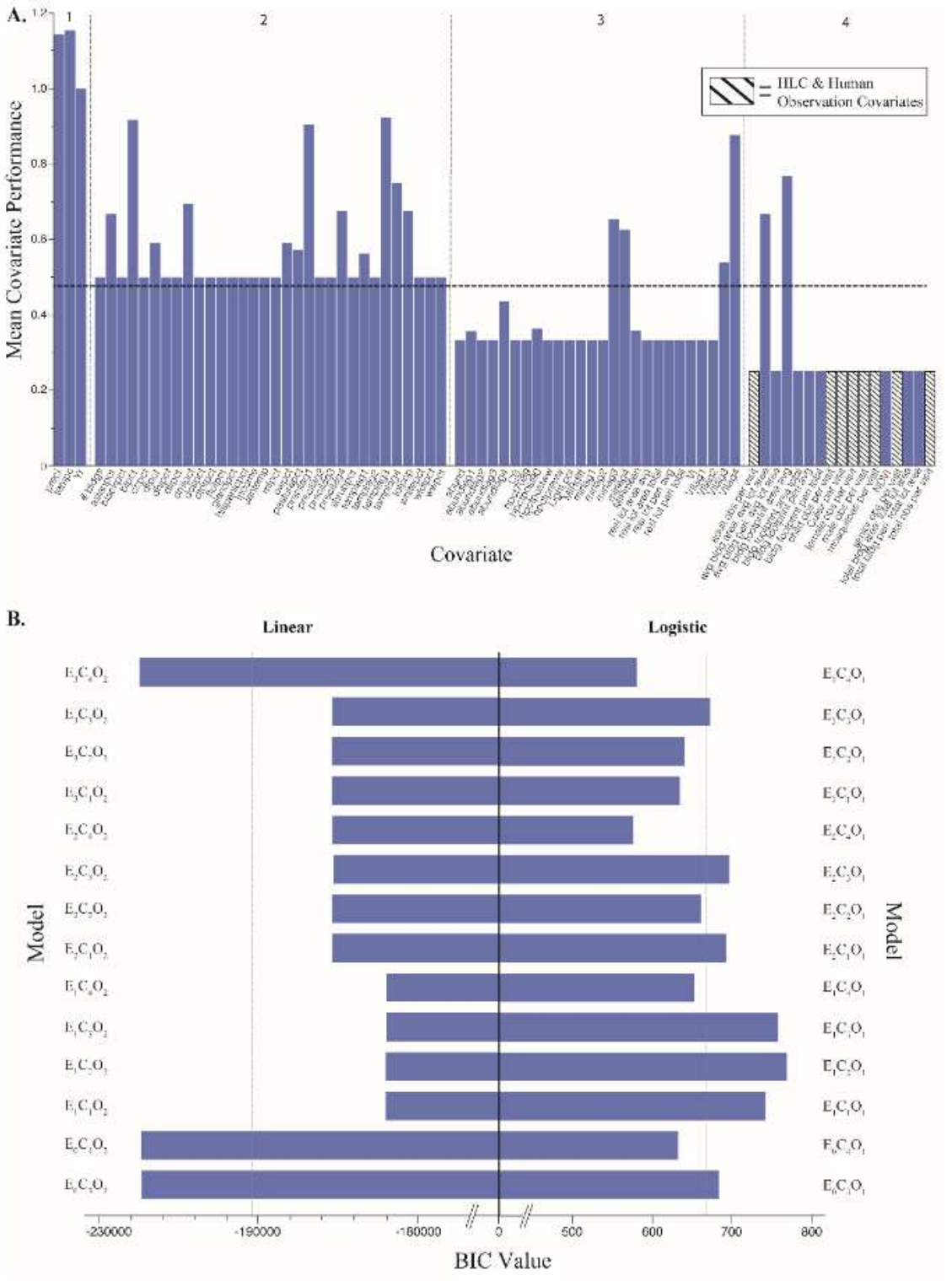

Figure 3

Mean performance of each of the 70 covariates used in the study (A). Covariates are listed in alphabetic order by data availability/work load to acquire score (1-4). The contribution of each covariate resulted in a net value performance for each linear and logistic model assessed in the study (B). Means for each outcome (x ̄covariate = 0.48; x ̄linear = -193406; x ̄logistic = 670.9) are designated by horizontal dashed lines. Details of scoring for each covariate and model are provided in Tables 5 & S1

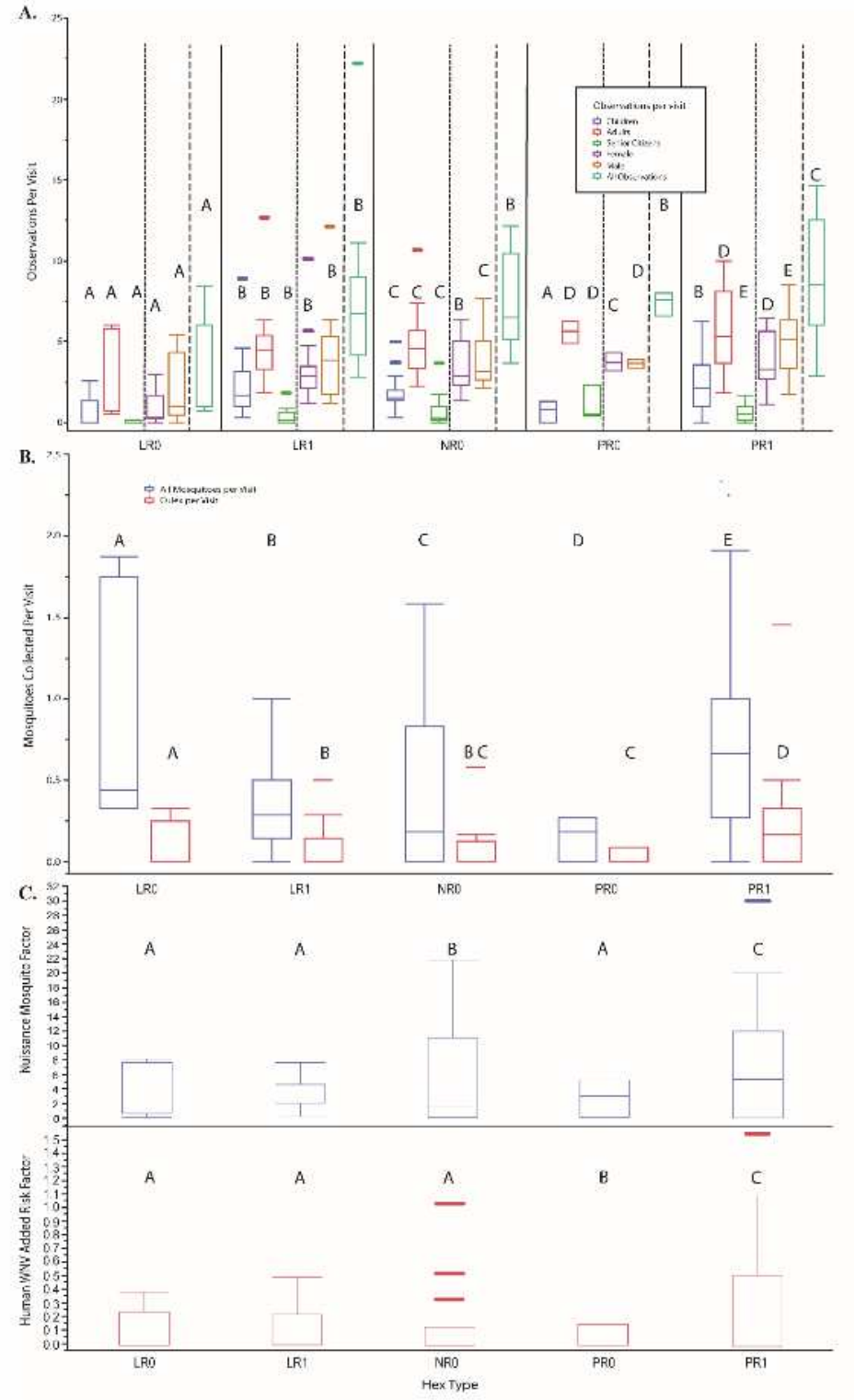

Figure 4

Relationship of hexagon type (LR = low residual, PR = positive residual, NR = large, negative residual; 0 = no human case, 1 = human case) by human observations per visit (A), mosquitoes collected per visit (B), and a product of the two former variables, nuisance factor and WNV added risk (C). Letters above each box and whisker plot designate significantly different groups by hexagon type, as calculated by Tukey's HSD.

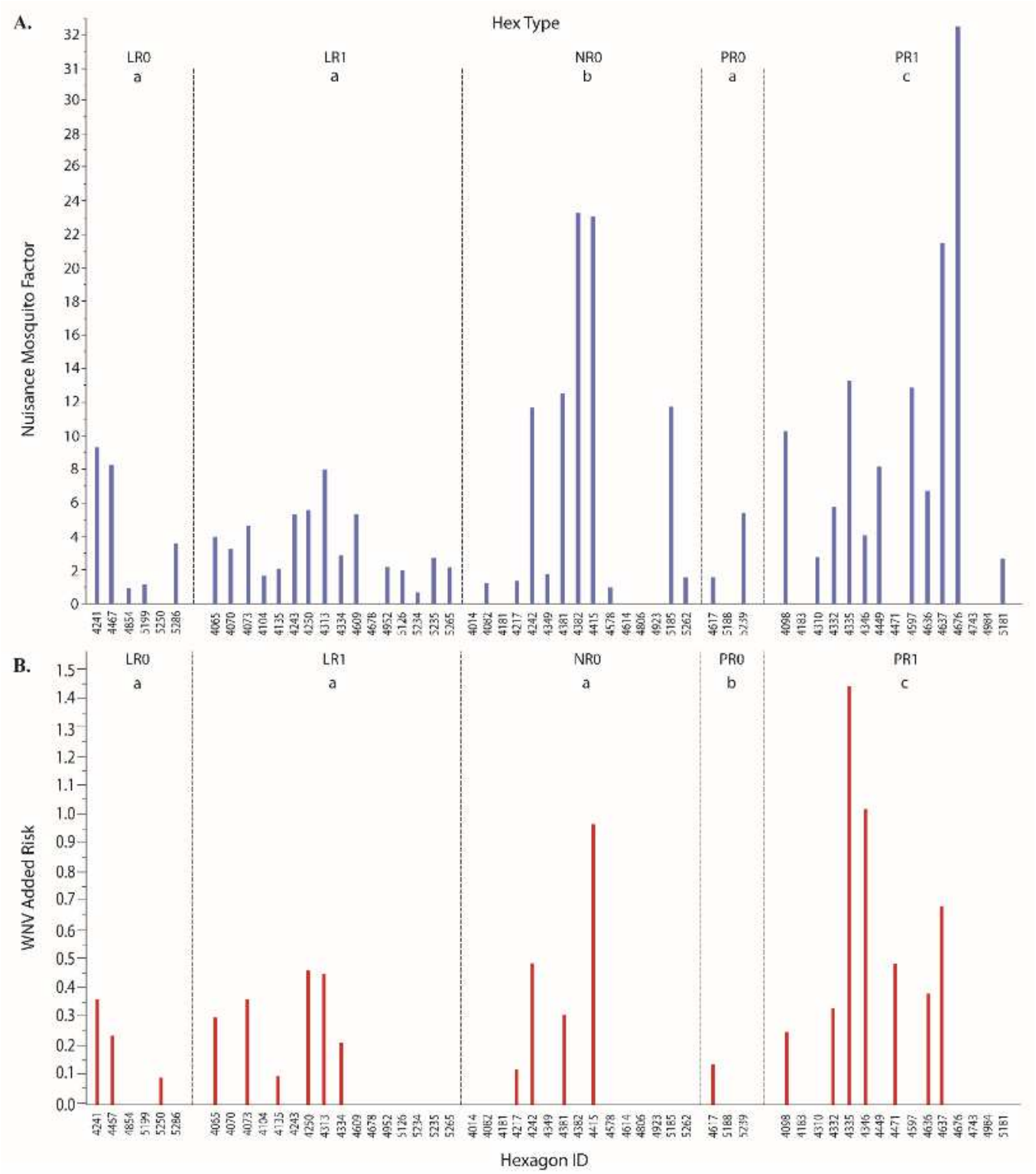

Figure 5

Relationship of each of the 55 UFS study hexagons (y-axis = unique identification number) by nuisance mosquito factor (A) and human WNV added risk (B). Letters above each box and whisker plot designate significantly different groups by hexagon type, as calculated by Tukey's HSD.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- graphicalabstract.tif
- Additionalfile1.docx