Too big, too small, or just right? The influence of multispectral image size on mosquito population predictions in the Greater Toronto Area

Sydney A. DeMets[a], Amanda Ziemann[a], Carrie Manore[b], and Curtis Russell[c]

[a]Intelligence and Space Research Div., Los Alamos National Lab, Los Alamos, NM, USA
[b]Analytics, Intelligence & Technology Div., Los Alamos National Lab, Los Alamos, NM, USA
[c]Public Health Ontario, Toronto, Canada

ABSTRACT

Outbreaks of West Nile Virus (WNV) and St. Louis Encephalitis (SLE) are projected to increase in frequency and intensity with climate change, underlining the need to develop better mosquito borne disease (MBD) forecasting systems. Spread by *Culex*, WNV and SLE have seemingly random spatial and temporal outbreaks, making such outbreaks difficult to predict. However, recent studies have found that mosquito abundance and WNV/SLE transmission are strongly correlated, providing researchers with a foundation for the development MBD forecasting systems. Mosquito populations are impacted by several environmental variables, such as humidity, temperature, vegetation, and available breeding habitat. Mosquito-population forecasting models are beginning incorporate spectral data, such as the normalized difference vegetation index (NDVI). Vegetation is a crucial habitat for some mosquito species, and spectral data offers the best estimate of this habitat virtually anywhere on Earth. Additionally, vegetation offers a proxy for understanding how water flows across a landscape, a crucial consideration in urban areas with high landscape heterogeneity. This research explores how the spatial scale (extent) of multispectral imagery used in mosquito population prediction models influences mosquito population forecasts, specifically in the Greater Toronto Area. We derive three monthly time series of standard spectral indices from multispectral imagery over the Greater Toronto Area from 2004 to 2015; each time series is derived from images taken over the same locations, but using images taken over different spatial footprints. We then explore how spectral indices across the three spatial scales perform as predictors for combined *Cx. restuans* and *Cx. pipiens* populations.

Keywords: Multispectral imagery, mosquito population forecasting, spectral indices, NDVI

1. INTRODUCTION

The increasing burden of mosquito borne diseases (MBD) has underlined the importance of developing reliable disease forecasts. West Nile Virus (WNV) and St. Louis Encephalitis (SLE) are the two most common mosquito borne diseases (MBD) in North America. These MBDs are becoming more prevalent as climate change increases mean annual temperatures, making more areas habitable to MBD vectors. Similarly, urbanization is increasing temperatures in densely populated areas, furthering metropolitan susceptibility to MBD outbreaks. Recent research has found that the seemingly sporadic outbreaks of WNV/SLE correlate to fluctuations in mosquito populations; consequently, creating MBD forecasts requires the development of better mosquito population predictions.[1]

Mosquito breeding habitats – specifically their availability and productivity – control the size of mosquito populations. Temperature is commonly used to estimate the productivity of breeding habitats,[1,2] and in general, mosquitoes reproduce more quickly at warmer temperatures. For instance, *Culex*, the primary vectors of WNV and SLE, follow this pattern, and reproduce in temperatures of 16°C to 30°C.[3] Availability of breeding habitat is often estimated using precipitation: the more rain, the greater the likelihood of standing water. Most mosquitoes, including those in the *Culex* genera, deposit their eggs in standing water, making water's presence crucial to

Send correspondence to Sydney DeMets; E-mail: sdemets@lanl.gov

Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imagery XXVI,
edited by Miguel Velez-Reyes, David W. Messinger, Proc. of SPIE Vol. 11392, 113920Y
 · doi: 10.1117/12.2558128

their survival.[4] The wide availability of rainfall and temperature data has resulted in their regular incorporation into mosquito population prediction models.[2,5–7] However, precipitation data often come at spatial resolutions greater than 2 km, making it difficult to discern how water collects at finer spatial resolutions across an urban, heterogeneous landscape. While this comparatively coarse resolution can inform mosquito population predictions at large scales, the resolution exceeds the average lifetime range of *Culex* (anywhere from 500 m to 24,800 m), which limits our ability to forecast mosquito populations in varied landscapes.[8]

Many mosquito population models rely heavily on precipitation and temperature data. However, other data sources, especially multispectral imagery, are beginning to be more commonly incorporated in mosquito population prediction models . Although multispectral imagery does not have the ability to directly measure mosquito breeding habitats, measurements of types of vegetation can, for example, serve as indirect indicators of water availability.[9,10] In particular, spectral indices such as the Normalized Difference Vegetation Index (NDVI) and the short-wave infrared version of the Normalized Difference Water Index (sNDWI) provide us with quantifiable indicators of vegetation health or water content in leaves, respectively. Such indices also indicate available resting and feeding habitat for mosquitoes, as *Culex* rely on trees and shrubs as feeding and resting sites.[11] Depending on the source of the imagery, the resolution of this indices could range anywhere from 10 m (Sentinel-2) to 30 m (Landsat constellation) for global-coverage sensors, and even be as low as ∼2m in size for tasking-based sensors (*e.g.*, WorldView-3). Similarly, the frequent revisit rate of global-coverage multispectral satellites (Sentinel-2, Landsat constellation) provides data collected at a temporal and spatial extent that cannot be matched *in situ*.

Despite the comparatively high-resolution environmental information multispectral imagery offers, many mosquito population forecasts which incorporate multispectral data aggregate these data up to political regions of interest (*e.g.*, state or province level). Such aggregation may mask how mosquito populations *actually* behave within political boundaries, as many mosquito species, including *Culex*, will not travel more than 1 km to 2 km in their lifetime.[8] Previous work exploring forecasting mosquito populations in the Greater Toronto Area found that multispectral data were significant predictors of mosquito populations in regions with more complete mosquito trap coverage, whereas multispectral data could not produce meaningful mosquito population predictions in regions sparsely sampled for mosquitoes;[12] in other words, the completeness of the ground truth data had an impact on the ability of the multispectral data to predict that ground truth data. Understanding the impact of aggregated spectral data on mosquito population forecasts might pose a less urgent question if mosquito sampling sites were evenly distributed across study areas of interest. However, this is often not the case, as most mosquito research teams are limited by time, personnel, and resources. In short, testing against sub-regions that were not sampled for mosquitoes (resulting in insufficient ground truth data) will hinder evaluation of the ability of multispectral data to forecast mosquito populations, especially in urban areas with heterogeneous landscapes.

Culex restuans (*Cx. restuans*) and *Culex pipiens* (*Cx. pipiens*) are mosquito species found across Canada and the United States. Public Health Ontario has surveyed spring through fall *Cx. restuans* and *Cx. pipiens* populations across the Greater Toronto Area since 2004. Using these survey data and Landsat 5 data, we sought to discern if increases in the spatial scale of multispectral image data impact mosquito population predictions.

2. METHODS

2.1 Study area and mosquito data

The Greater Toronto Area (GTA), located in Ontario, Canada, comprises York, Durham, Peel, Toronto, and Halton Regions. Public Health Ontario has collected mosquito data across Ontario and the GTA every spring and summer since 2004. Mosquito count data were obtained from the Ontario mosquito population surveillance program. Sampling sites were distributed across Ontario every April through October from 2004-2011. One georeferenced mosquito trap was placed at each sampling site; the location of some sampling sites varied from 2004-2011, but only sites that fell within the GTA and were sampled every year were included in this analysis.

Traps were sampled at varying time intervals. When sampling sites were visited, the number of *Culex* in the given site's trap was recorded, and the traps were then emptied and reset. For this study we chose thirty sites based on sampling location and sampling frequency (Fig. 1; selected mosquito trap sites were sampled each year from 2004-2011 and had been sampled at least twice per month from June to September of each study

year). Sampling weights were added to mosquito-count data to account for differences in the number of sampling events for each trap. We then computed the mean of the weighted mosquito counts of all thirty traps for each month included in the study (*e.g.*, the average weighted mosquito count for July of 2006 was 124 mosquitoes). Additionally, we calculated the mean monthly air temperature for each month included in the study, using historical data from the Toronto City weather station.[13]

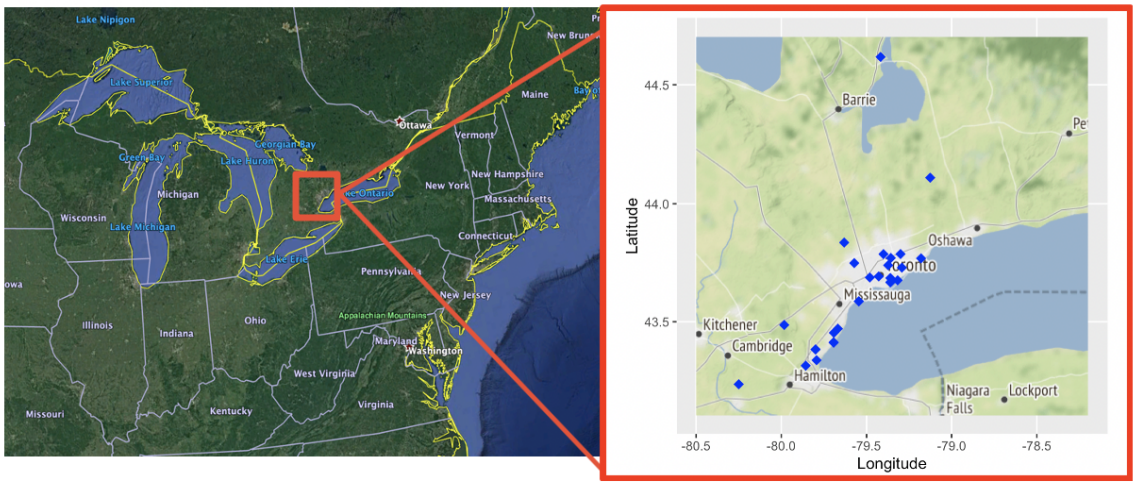

Figure 1. Location of Toronto (left) and selected trap sites, denoted by blue diamonds (right).

2.2 Multispectral data

We used Landsat 5 multispectral imagery for this analysis, as it was in operation for the entire time frame over which the mosquito trap data were collected. Landsat 5 was a multi-instrument global coverage satellite whose Thematic Mapper sensor had 7 bands covering the visible through the near infrared (NIR), short-wave infrared (SWIR), and thermal infrared (TIR) wavelength ranges, and was jointly managed by the U.S. Geological Survey (USGS) and the National Aeronautics and Space Administration (NASA). Landsat 5 had revisit rate of approximately 16 days and a spatial resolution of 30 m, allowing for the derivation of multiple time series of indices over the full seven-year range of this study.[14] We accessed the imagery through the Descartes Labs Platform.[15]

To test how increases in the area of spectral data influences mosquito population predictions, circles of radius 0.5 km, 1 km, and 2 km, respectively, were drawn around each trap previously selected for analysis (Fig. 2). These radii are based on the lifetime travel limits of mosquitoes discussed in Section 1. For each month and for each width of the concentric circles, we computed the following indicators from the Landsat 5 imagery: the Normalized Difference Vegetation Index (NDVI, Eq. (1)), Normalized Difference Built-up Index (NDBI, Eq. (3)), green-based Normalized Difference Water Index (gNDWI, Eq. (2)), SWIR-based Normalized Difference Water Index (sNDWI, Eq. (4)), and the Normalized Burn Ratio (NBR, Eq. (5)). We computed each index for every pixel in a given region, then aggregated the results to circle-level summary statistics, generating a total of three datasets (*i.e.*, datasets with aggregated 0.5 km radius data, 1 km radius data, and 2 km radius data, respectively). For each radius, we have a monthly time series of the following summary statistics for each of the pixel-level indices calculated within the region: mean, minimum, maximum, 2% minimum threshold, 98% maximum threshold, and standard deviation.

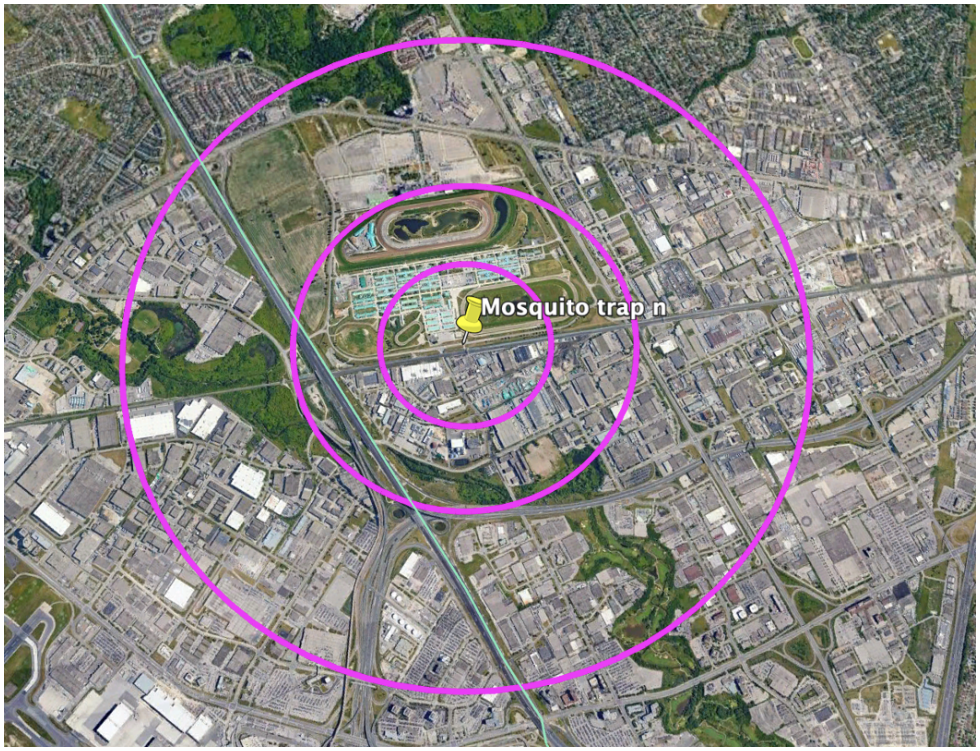

Figure 2. Example areas of interest across various radii. Image courtesy of Google Earth.

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

$$NDBI = \frac{SWIR_1 - NIR}{SWIR_1 + NIR} \tag{2}$$

$$gNDWI = \frac{Green - NIR}{Green + NIR} \tag{3}$$

$$sNDWI = \frac{NIR - SWIR_1}{NIR + SWIR_1} \tag{4}$$

$$NBR = \frac{NIR - SWIR_2}{NIR + SWIR_2} \tag{5}$$

2.3 Autoregressive distributive lag model

Distributive lag models (DLMs) are dynamic models in which the effect of n regressors (predictors) on y (the value of interest) occurs over time rather than all at once. Since mosquito populations are autoregressive (*i.e.*, current and past mosquito populations inform future mosquito populations), an autoregressive DLM (ARDL) was used for this study. Additionally, this allows us to assess how NDVI at timestep t_{n-2} influences the mosquito population at time t_n. Here, we gave the DLMs spectral data and mosquito data averaged across all circles for all three circle sizes. Mean monthly temperature and the summary statistics on the derived spectral indices (Eqs. (1)-(5)) were used as predictors, in addition to mosquito population. Lags for predictors were selected using the AIC criteria.[16] Alpha value was set at 0.05. All modeling and statistical analysis was programmed in R.

3. RESULTS

Multispectral data from each spatial footprint was able to significantly predict mosquito populations. sNDWI and NBR indices tended to be the most correlated with the weighted mosquito populations. The most accurate mosquito population predictions were derived from sNDWI data, though each model needed the average monthly temperature to be incorporated to produce significant results.

3.1 0.5 kilometer model

Mosquito population predictions developed using data from the 0.5 km circles were significant (Fig. 3, $p < 0.01$). The 98th percentile of sNDWI (sNDWI-98, 2 percent short of the sNDWI maximum value) and mean monthly temperature were the regressors which best predicted mosquito populations. Mean monthly temperature had a lag time of two months, and sNDWI-98 had a lag time of six months. sNDWI-98 was a significant predictor of mosquito populations at $t - 3$ months ($p < 0.05$); mean monthly temperature did not have a significant impact on mosquito populations at any particular time step. Overall, the model describes 93.5% of variation in mosquito populations ($R^2 = 0.9359$), with the independent variables (mean monthly temperature and sNDWI-98) describing 88.1% of variation in mosquito populations (adjusted $R^2 = 0.8805$).

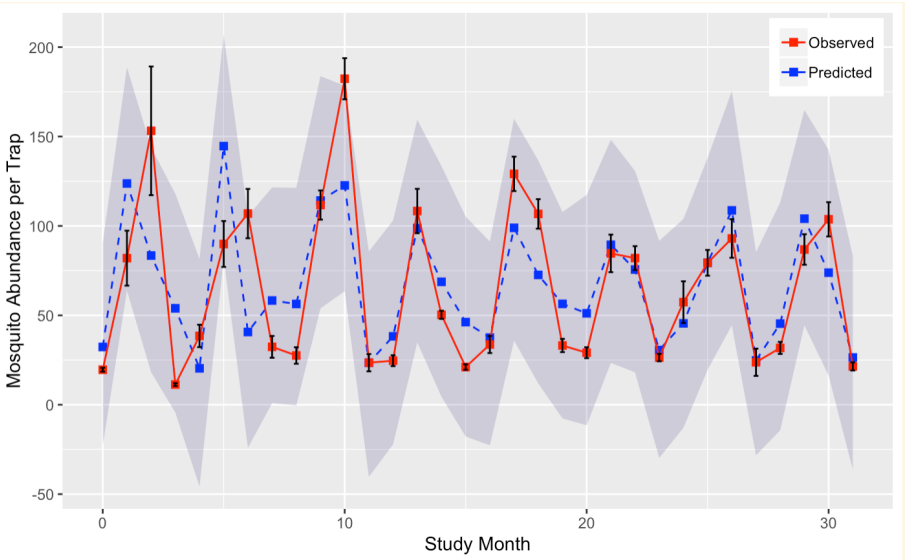

Figure 3. Predicted and observed mosquito counts generated from the 0.5 km radius dataset, where the predicted mosquito counts were generated using multispectral indices computing within the 2.0 km radius. The Study Month (x axis) is the nth month included in this study, and the shaded region marks the 95% confidence interval for predicted mosquito values. Error bars indicate ± standard error. $R^2 = 0.9359$, $p < 0.01$, and mean absolute scaled error (MASE) = 0.3517.

3.2 1.0 kilometer model

Mosquito population predictions developed using data from the 1.0 km circles were significant (Fig. 4, $p < 0.01$). sNDWI-98 and mean monthly temperature, again, best predicted mosquito populations. Mean monthly temperature had a lag time of two months; sNDWI-98 also had a lag time of two months (which was shorter than the lag time selected for sNDWI-98 when derived from the 0.5 km spectral data). Neither sNDWI-98 nor mean monthly temperature was a significant predictor of mosquito populations at any lag length, although sNDWI at time t was a marginally significant predictor of mosquito abundance ($p < 0.1$). In total, the model described 91.6% of variation in mosquito populations ($R^2 = 0.9159$). The independent variables (mean monthly temperature and sNDWI-98) described 89.8% of variation in mosquito populations (adjusted $R^2 = 0.8949$).

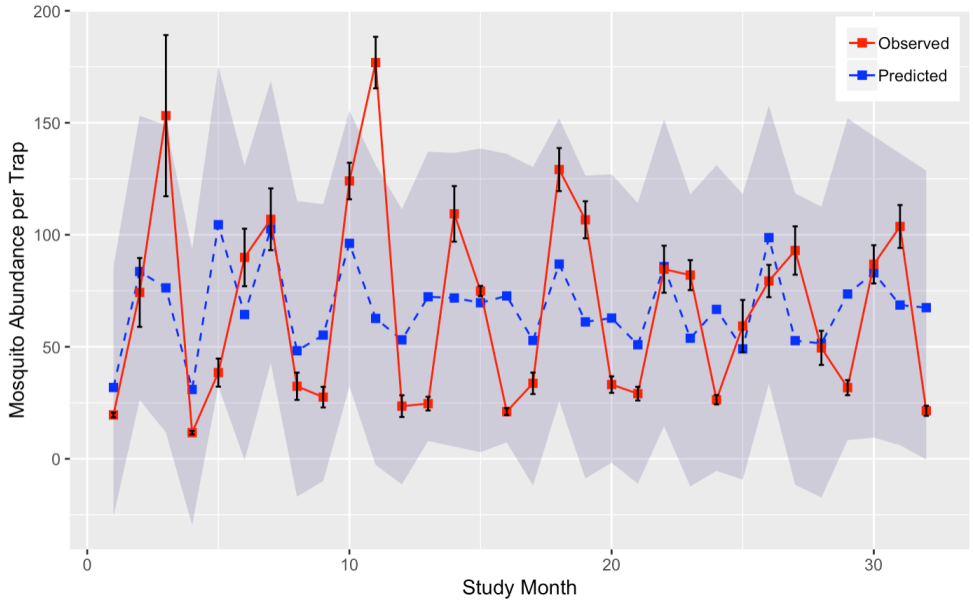

Figure 4. Predicted and observed mosquito counts generated from the 1.0 km radius dataset, where the predicted mosquito counts were generated using multispectral indices computing within the 2.0 km radius. The Study Month (x axis) is the nth month included in this study, and the shaded region marks the 95% confidence interval for predicted mosquito values. Error bars indicate ± standard error. $R^2 = 0.9159$, $p < 0.01$, and MASE = 0.4009.

3.3 2.0 kilometer model

Mosquito population predictions from the model developed using data from the 2.0 kilometer circles were significant (Fig. 5; $p < 0.01$). Mean sNDWI and mean monthly temperature were the strongest predictors of mosquito populations. Mean monthly temperature had a lag time of two months, and mean sNDWI had a lag time of one month. Neither mean sNDWI nor mean monthly temperature was a significant predictor of mosquito populations at any particular lag length, although mean monthly temperature at $t-2$ was a marginally significant predictor of mosquito populations at t ($p < 0.1$). The model describes 91% percent of variation in mosquito populations ($R^2 = 0.9096$), with the independent variables (mean monthly temperature and mean sNDWI) describing 88.2% of variation in mosquito populations (adjusted $R^2 = 0.8821$).

3.4 Spectral predictors

The strongest spectral predictors of mosquito abundance, along with their respective lag times, changed with each increase in the radius over which the spectral indices were computed. For the 0.5 km dataset, sNDWI-98, gNDWI-98, and gNDWI-max were the three most correlated spectral variables to mosquito population. Each index had correlation coefficients of 0.4587, 0.2772, and 0.2768, respectively. When the model was trained using mean monthly temperature and gNDWI-98 data, the model described 87.73% of variation in mosquito abundance; the model trained with mean monthly temperature and gNDWI-max data explained 87.60% of variation in mosquito abundance. Mosquito abundance was most correlated with sNDWI-98, sNDWI-mean, and NBR-mean derived from the 1.0 km dataset (correlation coefficients = 0.4577, 0.4563, 0.4245, respectively). sNDWI-mean and NBR-mean (along with mean monthly temperature) each accounted for 88.5% of variability in mosquito populations. Lastly, mosquito abundance was most closely correlated to sNDWI-mean, NBR-mean, and NDVI-98 derived from the 2.0 km spectral index (correlation coefficients = 0.4326, 0.3937, 0.3815, respectively). When the model was trained using mean monthly temperature and NBR mean data, the model described 89.8% of variation in mosquito abundance; replacing NBR-mean with NDVI-98 increased variance described to 90.1%.

4. DISCUSSION

The primary goal of this research was to examine how the spatial footprint of the imagery used to compute multispectral indices influenced mosquito population predictions. Significant population predictions were generated

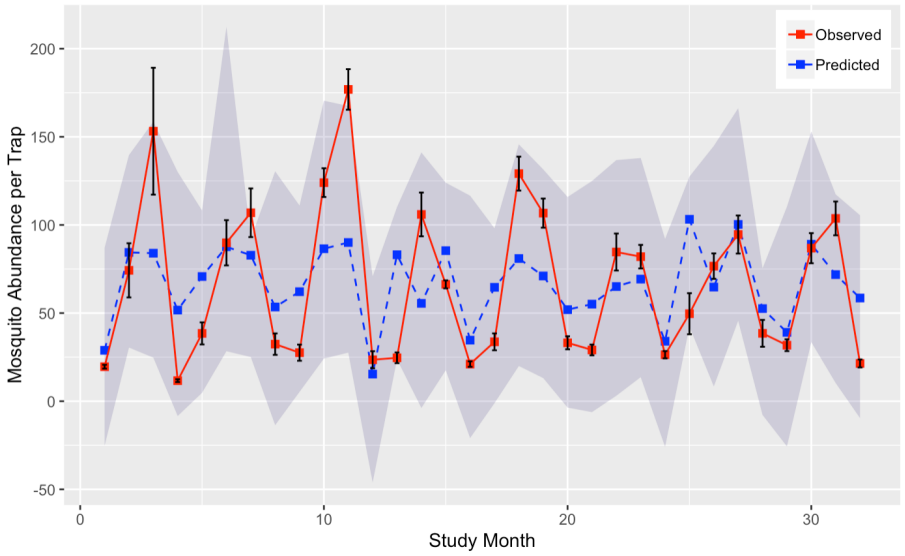

Figure 5. Predicted and observed mosquito counts generated from the 2.0 km radius dataset, where the predicted mosquito counts were generated using multispectral indices computing within the 2.0 km radius. The Study Month (x axis) is the nth month included in this study, and the shaded region marks the 95% confidence interval for predicted mosquito values. Error bars indicate ± standard error. $R^2 = 0.9159$, $p < 0.01$, and MASE = 0.3795.

using spectral data pulled from all three footprint sizes. However, mosquito population predictions derived from the 1 km and 2 km data had larger deviations from the observed mosquito populations than predictions made using 0.5 km radius spectral imagery. The model predictions from the 0.5 km radius spectral imagery better captured fluctuations in the amplitude of mosquito populations, as evidenced by the lower MASE score overall for the 0.5 km model. These results suggest that aggregating spectral data over large heterogeneous areas may negatively impact our ability to predict mosquito population. Still, a more thorough analysis will need to be completed in order to better estimate the error introduced into mosquito population models by using spectral data from areas significantly larger than the area over which mosquito traps are actually deployed.

It is also worth noting that, although the R^2 values of the mosquito population predictions are lower than those in DeMets et al., 2020, the variances in mosquito populations described in this study are consistently higher.[12] This suggests that spectral indices may do a better job of estimating mosquito abundance when using multispectral imagery taken from smaller areas around mosquito sampling sites. On a similar note, many of the sites included in this analysis fell within the Toronto region, one of the regions where spectral data failed to significantly predict mosquito populations in DeMets et al., 2020, despite the comparatively extensive sampling of the Toronto region. [12] This indicates that removing "excess" spectral image data, or data which exceed the (spatial) lifetime travel range of the mosquito species of interest, can help improve mosquito population predictions in areas with high landscape heterogeneity.

One unexpected finding was that sNDWI and NBR were most closely correlated to mosquito populations. sNDWI remained the most closely correlated to and the strongest spectral predictor of mosquito populations. sNDWI indicates the amount of water contained in tree canopies, thus providing a proxy measure for both the amount of vegetation and the amount of water across a landscape. Seeing sNDWI as a strong predictor of mosquito populations makes sense from an ecological standpoint, as *Culex* need water to breed, and rely on vegetation for resting and feeding. Interestingly, NBR (the normalized burn ratio) was also strongly correlated with mosquito abundance, and was a statistically significant predictor of mosquito abundance across all three of our data sets. The reason why NBR presents as a good predictor of mosquito abundance merits further investigation. Lastly, NDVI, the most commonly used multispectral index in mosquito population modeling and forecasting, did not perform as well; in fact, it only became the third most correlated spectral index with mosquito populations when using the multispectral indices computed from the 2.0 km data.

Future work will explore how mosquito population predictions are impacted by using multispectral imagery

with larger spatial footprints. Additionally, we will examine how successively larger spatial footprints impact which spectral indices emerge as the strongest predictors of mosquito abundance.

5. ACKNOWLEDGMENTS

We thank Descartes Labs for providing streamlined access to remote sensing imagery and we thank Public Health Ontario for mosquito data. We also thank Dr. Deborah Shutt for her help with cleaning and processing the mosquito data, and for her help with the models. This work was funded by the Los Alamos Laboratory Directed Research and Development (LDRD) Program.

REFERENCES

[1] Mulatti, P., Ferguson, H. M., Bonfanti, L., Montarsi, F., Capelli, G., and Marangon, S., "Determinants of the Population Growth of the West Nile virus Mosquito Vector Culex pipiens in a Repeatedly Affected Area in Italy," *Parasites & vectors* **7**(26) (2014).

[2] Larson, S. R., Degroot, J. P., Bartholomay, L. C., and Sugumaran, R., "Ecological Niche Modeling of Potential West Nile Virus vVctor Mosquito Species in Iowa," *Journal of Insect Science* **10**(110) (2010).

[3] Loetti, V., Schweigmann, N., and Burroni, N., "Development Rates, Larval Survivorship and Wing Length of Culex pipiens (Diptera: Culicidae) at Constant Temperatures," *Journal of Natural History* **45**(35-36), 2203–2213 (2011).

[4] Day, J. F., "Mosquito Oviposition Behavior and Vector Control," *Insects* **7.4**(65) (2016).

[5] Wang, J., Ogden, N. H., and Zhu, H., "The Impact of Weather Conditions on Culex pipiens and Culex restuans (Diptera: Culicidae) Abundance: A Case Study in Peel Region," *Journal of Medical Entomology* **48**(2), 468–475 (2011).

[6] Ruiz, M. O., Chaves, L. F., Hamer, G. L., Sun, T., Brown, W. M., Walker, E. D., Haramis, L., Goldberg, T. L., and Kitron, U. D., "Local Impact of Temperature and Precipitation on West Nile Virus Infection in Culex species Mosquitoes in Northeast Illinois, USA," *Parasites & Vectors* **3**(19) (2010).

[7] DeFelice, N. B., Schneider, Z. D., Little, E., Barker, C., Caillouet, K. A., Campbell, S. R., Damian, D., Irwin, P., Jones, H. M., Townsend, J., et al., "Use of Temperature to Improve West Nile Virus Forecasts," *PLoS Computational Biology* **14**(3), e1006047 (2018).

[8] Hamer, G. L., Anderson, T. K., Donovan, D. J., Brawn, J. D., Krebs, B. L., Gardner, A. M., Ruiz, M. O., Brown, W. M., Kitron, U. D., Newman, C. M., et al., "Dispersal of Adult Culex mosquitoes in an Urban West Nile Virus Hotspot: A mark-capture study incorporating stable isotope enrichment of natural larval habitats," *PLoS Neglected Tropical Diseases* **8**(3) (2014).

[9] Peters, A. J., Walter-Shea, E. A., Ji, L., Vina, A., Hayes, M., and Svoboda, M. D., "Drought Monitoring with NDVI-based Standardized Vegetation Index," *Photogrammetric Engineering and Remote Sensing* **68**(1), 71–75 (2002).

[10] Ziemann, A., Fairchild, G., Conrad, J., Manore, C., Parikh, N., Del Valle, S., and Generous, N., "Predicting dengue incidence in Brazil using broad-scale spectral remote sensing imagery," in [*Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS)*], IEEE (July 2018).

[11] Gardner, A. M., Anderson, T. K., Hamer, G. L., Johnson, D. E., Varela, K. E., Walker, E. D., and Ruiz, M. O., "Terrestrial vegetation and aquatic chemistry influence larval mosquito abundance in catch basins, chicago, usa," *Parasites & vectors* **6**(1), 9 (2013).

[12] DeMets, S., Ziemann, A., Manore, C., and Russell, C., "Improving mosquito population predictions in the greater toronto area using remote sensing imagery," in [*Proc. IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*], IEEE (March 2020).

[13] Meteorological Service of Canada, "Historical Climate Data."

[14] USGS, NASA, "Landsat Missions."

[15] Descartes Labs, "Descartes Labs Platform."

[16] Liew, V. K.-S., "Which lag length selection criteria should we employ?," *Economics bulletin* **3**(33), 1–9 (2004).