A Novel Sampling Method to Measure Socioeconomic Drivers of Aedes Albopictus Distribution in Mecklenburg County, North Carolina

Ari Whiteman 1,2,*, Eric Delmelle 1, Tyler Rapp 2, Shi Chen 2, Gang Chen 1 and Michael Dulin 3,4

1. Department of Geography and Earth Sciences and Center for Applied Geographic Information Science, University of North Carolina at Charlotte, 2901 University City Blvd, Charlotte, NC 28223, USA; delmelle@gmail.com (E.D.); gang.chen@uncc.edu (G.C.)
2. Academy for Population Health Innovation, 9201 University City Blvd, Charlotte, NC, 28223, USA; trapp1@uncc.edu (T.R.); Schen56@uncc.edu (S.C.); mdulin3@uncc.edu (M.D.)
3. Department of Public Health Sciences, University of North Carolina at Charlotte, 2901 University City Blvd, Charlotte, NC 28223, USA
4. Academy for Public Health Innovation, University of North Carolina at Charlotte, 2901 University City Blvd, Charlotte, NC 28223, USA

* Corresponding: awhitem1@uncc.edu

Received: 22 August 2018; Accepted: 1 October 2018; Published: 5 October 2018

Abstract: Climate change, urbanization, and globalization have facilitated the spread of Aedes mosquitoes into regions that were previously unsuitable, causing an increased threat of arbovirus transmission on a global scale. While numerous studies have addressed the urban ecology of Ae. albopictus, few have accounted for socioeconomic factors that affect their range in urban regions. Here we introduce an original sampling design for Ae. albopictus, that uses a spatial optimization process to identify urban collection sites based on both geographic parameters as well as the gradient of socioeconomic variables present in Mecklenburg County, North Carolina, encompassing the city of Charlotte, a rapidly growing urban environment. We collected 3,645 specimens of Ae. albopictus (87% of total samples) across 12 weeks at the 90 optimized site locations and modelled the relationships between the abundance of gravid Ae. albopictus and a variety of neighborhood socioeconomic attributes as well as land cover characteristics. Our results demonstrate that the abundance of gravid Ae. albopictus is inversely related to the socioeconomic status of the neighborhood and directly related to both landscape heterogeneity as well as proportions of particular resident races/ethnicities. We present our results alongside a description of our novel sampling scheme and its usefulness as an approach to urban vector epidemiology. Additionally, we supply recommendations for future investigations into the socioeconomic determinants of vector-borne disease risk.

Keywords: vector-borne disease; health disparities; optimization; social determinants of health

1. Introduction

As a result of the proliferation of container-breeding Aedes mosquitoes across much of the world’s urban regions, rates of arbovirus infections have increased globally over the last several decades [1–3]. Subsequently, research has been initiated to better understand the ecology of the urban Aedes mosquitoes that act as vectors for the most commonly transmitted viruses in these urban regions, including: dengue virus, yellow fever virus, chikungunya virus, and Zika virus. Studies of the most abundant urban Aedes species, Ae. albopictus and Ae. aegypti, have investigated breeding behavior [4,5], feeding behavior [6,7], and habitat preferences [8,9]. This work has produced vital information that has advanced vector control efforts and reduced disease transmission. However,
with climate change [10], globalization [11], and urbanization [12] threatening to spur further increases in arbovirus transmission rates, it is important to both evaluate prior methods of vector surveillance as well as test novel sampling designs that will improve current practices of designing mosquito surveillance systems.

There have been numerous studies of mosquito trapping techniques aimed at determining the sampling methods that most accurately account for known aspects of mosquito ecology [13–23]. Additionally, the accuracy of commonly used entomological indexes have also been examined [24–27]. To date, the majority of these studies focus on differences in trap types rather than arrangement of the traps in the geographic space. In most studies on Aeles population distribution, survey site selection is completed by identifying multiple focal areas in a given region, typically urban in nature, which sharply contrasts with the study’s independent variables (e.g., high versus low category groups). These contrasting areas are used as either independent groups containing multiple collection sites or a single collection site where sampling then occurs using the same collection methods at each location so that the results can be appropriately compared between sites. This traditional study design has been used to successfully identify effects of land cover type [28], socioeconomic conditions [29,30], degree of urbanization [31], seasonal change [32], precipitation [33], and interspecific competition [34] on Aeles population sizes.

While these studies have inarguably provided useful conclusions, there are limitations in their applicability to regional vector control programs. Namely, the sites sampled in these types of studies are often chosen to maximize differences in particular conditions rather than assessing across the range of conditions between extremes. This approach facilitates the interpreting of results, but even if there are intermediate groups or sites chosen, this technique presents an overly simplistic view of the heterogeneity of urban landscapes. For example, urban regions are not collections of high, medium, and low category neighborhoods, but are instead continuous and complex gradients of environmental, social, economic, and physical characteristics [35–37]. Thus, drawing conclusions from sites that divide these gradients into neatly differential groups make it difficult to recommend vector control strategies for areas that do not fit the characteristics of any one group. This limitation is also present in the other approaches that have been applied to vector surveillance including: (i) creation of a grid and surveying properties at each intersect [38,39]; (ii) using randomly selected grid cells [40]; (iii) sampling randomly at selected points within a spectral range [41]; (iv) using intervals along roads [42]; and (v) using prior knowledge of presumed vector hotspots to inform sample site selection [43]. Therefore, the development of alternative sampling schemes that consider the heterogeneous nature of urban landscapes will provide a pragmatic tool that improves the efficacy of surveillance.

The goal of this study was to create and evaluate a novel sampling design for Aeles mosquitoes which accounts for the complexity of their urban habitats, namely the socioeconomic and environmental variation present in a large urban region. Understanding Aeles ecology within the context of social determinants of health will aid in the identification of risk pools within a diverse urban population, allowing vector control programs to maximize their effectiveness by targeting regions or neighborhoods where vector abundance and virus transmission risk is predicted to be highest. Thus, the specific objective of our study is to design a site selection scheme, which objectively and quantifiably targets the widest and most accurate range of socioeconomic and environmental conditions along a continuous covariate gradient. Additionally, we have maximized the distance between each site allowing for the widest geographic coverage of observation and maximizing resource utilization. We present this design as an alternative surveillance method to existing strategies through description of a case study for Ae.albopictus surveillance in Mecklenburg County, North Carolina.

2. Methods

2.1. Study Site
Charlotte, North Carolina, located in Mecklenburg County (Figure 1), sits in a humid subtropical climate zone. It totals an average of 1100 mL of precipitation annually, fairly evenly distributed across all seasons, though slightly higher in the summer. Temperatures also peak in the summer months between 25 and 30 degrees Celsius [44]. The temperature and precipitation conditions from July to August are highly conducive to mosquito population growth [45]. Charlotte is one of the fastest growing cities in the United States, with a growth rate of 59.6% over the previous decade [46] and the sixth highest increase in population of any city in the country between 2000 and 2012, and the second highest increase between 2010 and 2013 [47]. Socioeconomically, Charlotte has one of the highest rates of poverty and has the single lowest rate of upward mobility of any city in the U.S., with only 4.4% of children raised in a low income bracket likely to transition to a higher income bracket in their lifetime [48]. A high growth rate combined with the lowest rate of upward mobility indicates that economic disparity in Charlotte is high and likely to continue to increase. As evidence, the Gini Index (0.479) indicates that Charlotte has the tenth highest level of income inequality of any city in U.S. and that the percentage of residents in the 95th percentile and 20th percentile of income are growing at the ninth highest rate in the country, while middle income earners are less common in the metropolitan area [49]. Furthermore, as a result of segregation Charlotte has a unique “crescent and wedge” growth pattern, with high earners tending to occupy a single hyper-concentrated “wedge” of the city while lower income residents are distributed throughout the remaining “crescent.” This has led to considerable differences in the infrastructure and zoning of neighborhoods depending on the income of their residents, which in turn has led to considerable variation in the composition and configuration of the landscape present in those neighborhoods [50]. How strong these differences are and their impact on urban mosquito ecology is still unknown. Thus, while neighborhood-level health outcomes and their relationships to resident socioeconomic attributes have been well-documented [51], it is unclear to what degree local ecological factors contribute to social determinants of health.

General and broad scale mosquito surveillance efforts have occurred in Charlotte throughout the previous three decades. The most recent surveys in the summer of 2017 found *Ae. albopictus* larvae in 25 neighborhoods spread across Mecklenburg County. *Aedes aegypti* has not been seen in the county since the mid-1980s, having since been completely displaced by the invasive and superior competitor *Ae. albopictus*. There have been no assessments aimed at comparing neighborhood socioeconomic characteristics with local mosquito populations in the region.
2.2. Trap Location Selection

We identified 90 sample sites, a number recommended by an a priori G*Power analysis [52], using a two-step optimization procedure similar to the ones typically used in soil sampling designs [53]. Specifically, the main objective was to maximize the spatial spreading of mosquito sample sites across the study region, while the secondary objective was to sample at locations that would reflect a large range of socioeconomic conditions. The unit for both phases was the neighborhood planning area (NPA), a geographical delineation originally based on the 1990 U.S. Census tracts and updated following the 2010 U.S. Census. The NPA was used by the Charlotte-Mecklenburg Planning Commission to more accurately fit the geographical boundaries of the county’s neighborhoods than the 1990 U.S. Census tracts did (n = 462, see Figure 1). The data used in the optimization procedure were extracted from the Charlotte Quality of Life Study (CQOLS; [50]), a resource which has been used for over 20 years in Mecklenburg County to illustrate and describe the quality of life in the county at the NPA scale based on 80 variables. The data describing each variable are drawn from a variety of local sources including: state, county, and municipal governments. The data are categorized under the following structures: economic, environmental, education, engagement, health, housing, safety, and transportation. More information on the specific methods of the CQOLS can be found at: https://mcmap.org/qol/.

We utilized nine variables from the CQOLS, each with a demonstrated or hypothetical relationship with Aedes distribution [39,54–55]. These nine variables were selected following a correlation matrix that identified collinearity between one of more of the initially chosen 21 variables. The nine final variables were: socioeconomic status (an index of the normalized weighted average of three common socioeconomic variables—percent with bachelor’s degree, household income, and home sales price); population density; employment rate; total area covered by tree canopy; foreclosure rate; violent crime rate; Hispanic population rate; African-American population rate; and proximity to a park (Figure 2). All variables were standardized from 0 to 1.

2.3. Optimization Phase 1: P-dispersion

The first phase of the optimization was aimed at dispersing sampling sites to guarantee wide spatial coverage of study. The sites used for the optimization were the centroids of each NPA. One model that attempts at spreading sites as from one another is the p-dispersion model [56], which essentially maximizes the distance that separates any two sites. Four NPAs were excluded from the analysis: two which comprise an airport, and two which did not contain any residential units. The p-dispersion model is formulated as follows:

$$\text{Max } D$$  \hspace{1cm} (1)

Subject to:

$$D + (M - d_{ij})X_i + (M - d_{ij})X_j \leq 2M - d_{ij} \forall i, j > i$$  \hspace{1cm} (2)

$$\sum_i X_i = p$$  \hspace{1cm} (3)

$$X_i \in \{0, 1\} \quad \forall i \in I, j \in J$$  \hspace{1cm} (4)

With $X_i$ a decision variable equal to 1 when we selected a sample site, located in NPA $i$, and 0 otherwise. The term $d_{ij}$ is the distance that separates two sample site $i$ and $j$ and can be calculated prior to optimization. In our model, Equation (1) maximizes the distance $D$ between the closest pair of NPAs $i$ and $j$. Constraint (2) tracks the distance between NPA centroids, when both are selected (If either NPA $i$ or $j$ has not been selected, then $D$ is forced to be less than or equal to $d_{ij} + M$, where $M$ is a very large number). Constraint (3) stipulates that the total number of sample sites must be equal to...
Constraints (4) are binary integer constraints. The advantage of this model is that inter-site distances are tracked.

2.4. Optimization Phase 2: Maximal Coverage Approach

For this phase, we added 60 sample sites to the existing set, following a maximal coverage approach [57]. This second phase of the optimization was designed to locate samples in highly populated neighborhoods, and across the widest range of neighborhood input variables. This was done to maximize the potential host population in the study region, because the species of interest in the research, *A. albopictus*, is a peri-domestic urbanized organism that thrives in areas of high human habitation [27, 58]. In addition to selecting NPAs with the highest populations, the model is constrained so that the selected NPAs are representative of the data distribution of nine input variables from the CQOLS. For each of the nine aforementioned variables, we identified the 20th, 40th, 60th, 80th, and 100th percentiles (Figure 3). We then constrained the model so that in the final set of 90 neighborhoods (60 + 30), there would be at least 5 NPAs from the 0 to 20th percentile range of each variable, four neighborhoods from the 20th to 40th percentile range for each variable, and so on (Figures 2–4; Table 1). Table 2 indicates the results of a Kolmogorov-Smirnov test comparing the distribution of the nine variables with exhaustive and optimized sample, respectively. The coverage approach model is formulated as follows:

\[ \text{Max} \sum_{j \in J} h_j Y_j \]  

Subject to:

\[ \sum_{i \in I} X_i \geq Y_j \]  

\[ \sum_{i} X_i = p \]  

\[ \sum_{i} X_i \geq 5 \ \forall q \in Q, k \in K \]  

\[ X_i \in \{0, 1\} \quad \forall i \in I \]  

\[ Y_j \in \{0, 1\} \quad \forall j \in J \]  

Whereas \( Y_j \) is a decision variables equal to 1 when a NPA centroid is ‘covered’ by a sampling site \( i \) (Figure 5). Whether a sampling site is selected is unknown, hence similar to the \( p \)-dispersion model, \( X_i \) needs to be determined through the optimization method. The objective of the model, constraint (5), is to maximize the population \( h_j \) within NPA \( j \) that is covered (represented) by the sampling location at \( i \) (we assume that a population is covered if it is within a “service” distance of 1000 meters from the sample site; this parameter can be modified, but reflects the average diameter of NPAs; Figure 5). As such, samples were located in regions with higher population counts.

Constraint (6) stipulates that an NPA \( j \) is covered only when there is at least one sample location \( i \) in the vicinity of \( j \). The latter is defined by imposing a radius around sampling unit \( i \) and computing the set of NPAs \( N_j \) potentially covered by sampling unit \( i \) (Figure 5). Constraint (7) restricts the number of sample sites to be equal to \( p \) (\( p = 60 \) in this case). Constraint (8) stipulates that a minimum number of sample locations (here 5) should be selected for each quintile \( q \) (\(|Q| = 5\)) of each of the standardized variables \( K \) with \(|K| = 9\) (a higher number of samples per quintile made the problem infeasible). Finally, constraints (9) and 10 are standard integrality constraints; sampling sites are located or not in (9) and NPAs are covered or not in (10). Both models were solved using the optimal solver CPLEX [59].

Following the selection of 90 samples sites (Figure 6; \( n = 30 \) from the \( p \)-dispersion phase and \( n = 60 \) from the coverage approach phase), we attempted to establish a trap site as close to the NPA’s centroid as possible. If the centroid was located on private property, we asked the property landowner for permission, and if denied, we continued asking the owners of adjacent properties until
we were granted access. Once placed, traps were given a unique ID code and georeferenced using a handheld GPS unit. Traps were also marked with a notice dissuading tampering.

Figure 2. These nine variables, broken into quintiles, were used in the optimization process to identify NPAs suitable for surveying.
Figure 3. Histogram for each of nine variables used in the optimization process, using all NPAs. The red lines indicate the limits of each quintile.

Figure 4. Histogram for each of nine variables using the 90 sites selected in the optimization process (x-axis: standardized variable value; y-axis: count).
Figure 5. Illustration of Phase 2 mechanism in the optimization procedure. If a sampling unit at \( i = 12 \) is selected, it will ‘cover’ neighborhoods \( j = 7, 8, 9, 12 \) and \( 13 \).

Table 1. Quintile ranges for each of the nine variables used in the optimization process and number of sample points within each range.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socioeconomic Status</td>
<td>[0–0.305)</td>
<td>[0.305–0.401)</td>
<td>[0.401–0.491)</td>
<td>[0.491–0.603)</td>
<td>[0.603–1)</td>
</tr>
<tr>
<td>Population density</td>
<td>[0–0.0645)</td>
<td>[0.0645–0.129)</td>
<td>[0.129–0.161)</td>
<td>[0.161–0.225)</td>
<td>[0.225–1)</td>
</tr>
<tr>
<td>Employment rate</td>
<td>[0–0.708)</td>
<td>[0.708–0.810)</td>
<td>[0.810–0.856)</td>
<td>[0.856–0.916)</td>
<td>[0.916–1)</td>
</tr>
<tr>
<td>Tree canopy cover</td>
<td>[0–0.388)</td>
<td>[0.388–0.508)</td>
<td>[0.508–0.579)</td>
<td>[0.579–0.674)</td>
<td>[0.674–1)</td>
</tr>
<tr>
<td>Foreclosure rate</td>
<td>[0–0.016)</td>
<td>[0.016–0.032)</td>
<td>[0.032–0.072)</td>
<td>[0.072–1)</td>
<td></td>
</tr>
<tr>
<td>Violent crime rate</td>
<td>[0–0.005)</td>
<td>[0.005–0.017)</td>
<td>[0.017–0.043)</td>
<td>[0.043–0.103)</td>
<td>[0.103–1)</td>
</tr>
<tr>
<td>African Americans (%)</td>
<td>[0–0.062)</td>
<td>[0.062–0.172)</td>
<td>[0.172–0.349)</td>
<td>[0.349–0.519)</td>
<td>[0.519–1)</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>[0–0.047)</td>
<td>[0.047–0.083)</td>
<td>[0.083–0.127)</td>
<td>[0.127–0.239)</td>
<td>[0.239–1)</td>
</tr>
<tr>
<td>Proximity to park</td>
<td>[0–0.08)</td>
<td>[0.08–0.36)</td>
<td>[0.36–0.68)</td>
<td>[0.68–0.97)</td>
<td>[0.97–1)</td>
</tr>
</tbody>
</table>
Table 2. Results from the Kolmogorov-Smirnov test comparing the distribution of the nine variables with exhaustive and optimized sample, respectively. The p-value indicates the significance level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socioeconomic Status</td>
<td>0</td>
<td>0.4493</td>
</tr>
<tr>
<td>Population Density</td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>Employment Rate</td>
<td>0</td>
<td>0.9074</td>
</tr>
<tr>
<td>Tree Canopy</td>
<td>0</td>
<td>0.2081</td>
</tr>
<tr>
<td>Foreclosure Rate</td>
<td>0</td>
<td>0.9950</td>
</tr>
<tr>
<td>Violent Crime Rate</td>
<td>0</td>
<td>0.4321</td>
</tr>
<tr>
<td>Percent African American</td>
<td>0</td>
<td>0.4374</td>
</tr>
<tr>
<td>Percent Latino</td>
<td>0</td>
<td>0.0488</td>
</tr>
<tr>
<td>Proximity to Park</td>
<td>0</td>
<td>0.9652</td>
</tr>
</tbody>
</table>

Figure 6. Location of selected NPAs.

2.5. Entomological Surveys

Gravid *Aedes* Traps (GAT) were placed at each site selected using the optimization algorithm outlined above. Traps were located in shaded conditions protected from precipitation and direct sunlight. The GAT traps the gravid female as she arrives to lay her eggs in the water-filled container located below a mesh lining [13]. This is counter to host-seeking traps, such as the BG Sentinel [60-62],
that attract females by mimicking host-characteristics rather than oviposition sites. Host-seeking traps are highly effective yet provide a less direct link to virus transmission potential than gravid traps, which only trap females that have already had a bloodmeal. Additionally, at less than $20 USD per unit, GATs are cost-effective and allow for wider distribution without sacrificing efficacy, compared to host-seeking traps which can cost hundreds of dollars. Additionally, GATs have outperformed other gravid ovitraps, including sticky traps [17,63]. We infused the traps with hay and replaced the infusion at the four and eight week mark across the 12-weeks sample period. We emptied the traps’ contents and identified them [64] in the lab on a weekly basis from May 26, 2017 to August 21, 2017. Highly degraded or rubbed samples were verified to species using PCR techniques conducted at the Walter Reed Biosystematic Unit. This time frame represents the summer months in Charlotte and high season for mosquito activity, where the average daytime high temperature is approximately 29 degrees and the average precipitation is around 90mm per month [65]. Data was recorded as the number of Ae. albopictus females in each trap for each respective week.

2.6. Data Analysis

We used a generalized linear model (GLM) with a Poisson distribution and log link, treating weeks as a random effect in a mixed model (GLM; [66]) as well as cross-validation to validate the relationship between the abundance of gravid Ae. albopictus and the socioeconomic attributes of the NPA. GLMs have been used numerous times in vector epidemiological research and represent a robust method of variable association for presumed linear relationships [38,67–70]. We also generated land cover variables based on the heterogeneity of the landscape around each trap site to include in the model as predictors. Specifically, we used land cover maps and the software FRAGSTATS [71] to determine the Shannon land cover diversity and percent covered by each land cover type (tree canopy, grass/shrub, building, road/railroad, other paved surface, and water) within a 30 m radius of each trap site, chosen based on previous research into the average radius of urban Aedes hotspots [22,72,73]. Because these values were computed after the optimization process and required us to know the exact location of each trap, we ran an initial GLM model only containing the nine variables used for the second-phase of the optimization process and followed that with a model which included the land cover variables as well, comparing them using Akaike Information Criterion (AIC) values and residuals.

Thus, independent variables included in the second model were the variables at the NPA scale included in the optimization, plus the Shannon land cover diversity and percentage of each land cover type at the 30m radius scale, while the dependent variable for both models was represented by the abundance of gravid Ae. albopictus caught each week at each trap. We ran and validated the model using the Crossfold module in Stata [74], where k-fold cross-validation is performed to determine a model’s ability to fit out-of-sample data. This involves splitting the data randomly into k partitions (five being the default), then for each partition fitting the specified model using the other k-1 groups. The resulting parameters are used to predict the dependent variable in the unused group. Finally, the module reports the root mean squared error (RMSE) for each attempt, with the specified model being validated when the RMSE variation across attempts is minimal.

3. Results

Following 12 weeks of sampling that took place in each selected NPA, a total of 3,645 gravid female Ae. albopictus were collected throughout the length of the study period (Figure 7), with 72% of the traps across all locations and weeks being positive for Ae. albopictus. The Ae. albopictus represented 86% of the total mosquitoes collected, with the remainder divided between Ae. trisariatus (n = 203), Ae. vexans (n = 41), Ae. japonicus (n=39), Culex restuans (n = 41), and Cx. pipoens (n = 16). These are all known to be common species in the Southeastern U.S. The complete absence of Ae. aegypti is not surprising given the recent invasion of Ae. albopictus [34]. This result validates Mecklenburg County government surveillance efforts from previous years, as Ae. aegypti has not been found locally since
the 1980s. There was also a seasonal component, with the number of samples increasing slowly from late May to the height in late July before dropping sharply into August.

Figure 7. Average number of gravid *Ae. albopictus* caught each week from May 26, 2017 to August 21, 2017.

The first model (Table 3; AIC = 8.008), containing only the significant variables found among the nine used in optimization, was slightly weaker than the second model (Table 4; AIC = 7.858), which included the land cover variables around each trap, though is less than the generally required value of 2 to signify meaningful difference. This indicates that despite being significant predictors, the land cover variables do not add sufficient additional information to the model. The results of the second model illustrate that population, socioeconomic percentile, and percent Hispanic residents have a positive effect while foreclosure rate and violent crime rate have a positive effect on the abundance of gravid *Ae. albopictus* (Table 3). Figure 8 illustrates the sum of all gravid *Ae. albopictus* caught at each site. Finally, the residuals of the models did not exhibit any spatial autocorrelation (Moran’s I z-score: 0.002) and the models had consistent RSME and pseudo-$R^2$ values across five partitions (Table 5).
Figure 8. Sum of gravid *A. albopictus* caught over 12 weeks in each selected NPA (small circle: fewer total samples; large circle: more total samples)

Table 3. Results of the first GLM predicting the abundance of gravid *Ae. Albopictus*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Source</th>
<th>Coefficient</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing Density</td>
<td>CQOLS</td>
<td>-0.021</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Socioeconomic percentile</td>
<td>CQOLS</td>
<td>-0.731</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Foreclosure rate</td>
<td>CQOLS</td>
<td>0.050</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Violent crime rate</td>
<td>CQOLS</td>
<td>0.005</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Percent of residents Hispanic</td>
<td>CQOLS</td>
<td>0.007</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 4. Results of the second GLM predicting the abundance of gravid *Ae. Albopictus*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Source</th>
<th>Coefficient</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing Density</td>
<td>CQOLS</td>
<td>-0.107</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Socioeconomic percentile</td>
<td>CQOLS</td>
<td>-2.274</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Foreclosure rate</td>
<td>CQOLS</td>
<td>0.241</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Violent crime rate</td>
<td>CQOLS</td>
<td>0.039</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Percent of residents Hispanic</td>
<td>CQOLS</td>
<td>0.033</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Percent of land covered by buildings</td>
<td>Mecklenburg County GIS</td>
<td>-4.938</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Percent of land covered by roads/railroads</td>
<td>Mecklenburg County GIS</td>
<td>-11.98</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Percent of land covered by grass/shrubs</td>
<td>Mecklenburg County GIS</td>
<td>-5.681</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Percent of land covered by tree canopy</td>
<td>Mecklenburg County GIS</td>
<td>-2.698</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Shannon diversity index: land cover types</td>
<td>Mecklenburg County GIS</td>
<td>1.718</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 5. RMSE as calculated by Crossfold model validation.

<table>
<thead>
<tr>
<th>Model Run</th>
<th>RMSE</th>
<th>Rsquared</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.354</td>
<td>0.075</td>
</tr>
<tr>
<td>2</td>
<td>4.744</td>
<td>0.048</td>
</tr>
<tr>
<td>3</td>
<td>4.142</td>
<td>0.053</td>
</tr>
<tr>
<td>4</td>
<td>5.112</td>
<td>0.062</td>
</tr>
<tr>
<td>5</td>
<td>4.891</td>
<td>0.044</td>
</tr>
</tbody>
</table>

4. Discussion

Our overall goal was to create and evaluate a novel sampling design for *Aedes* mosquitoes which matched the complexity and heterogeneity of their urban habitats. We were successful in this objective, establishing a list of potential sample sites that accounts for the true distribution of socioeconomic and land cover based values that exist across a major urban region, while also promoting the spatial spreading of selected sites. This study design is beneficial in that it treats the potential predictors as a continuous gradient, which is more accurately represents how they exist in reality, rather than in a collection of groupings. We conceived the design with the intention creating a flexible approach that could be adapted to other urban regions and study systems. With the stipulation being that there is available data on neighborhood-scale attributes, the sample size in both optimization phases can be manipulated based on the desired parameters or size of the region. Additionally, following the optimization procedure, any relevant dependent variable can be measured along the devised sample site gradient.

One of the primary limitations of such a study design is the sacrificing of within-group sample size in exchange for maximizing the number of overall sample groups. With only one trap site per NPA, there is certainly potential for error, though we attempted to offset this by including a minimum of five samples per quintile per input variable. Although it was not possible to set a higher minimum number of samples per quintile due to model infeasibility, increasing the number of explanatory variables in the optimization model (|K|) would force us to select a smaller number of samples per percentile. Overall, this design provided a direct contrast to studies where high numbers of traps are distributed across a small number of sample groups. Both designs have inherent flaws, and while resource limitations prevent studies from placing a trap every 30 m across an entire urban
region, surveillance authorities should be cognizant of the costs and benefits to both approaches before designing their own programs.

In our study, we created a pool of study sites that incorporated the widest possible range of neighborhood-scale attributes with potential relationships to *Ae. albopictus* distribution, while concurrently maximizing the distance between sites to increase the spatial coverage of the study. In doing so, we were able to create predictions of gravid *Ae. albopictus* abundance across the distribution of values in the input variables. The results indicate that various metrics associated with socioeconomic status are related to the abundance of gravid *Ae. albopictus*. However, our validation metrics indicate that overall, the model is not highly predictive. This is not unexpected, as mosquito ecology is highly complex, with environmental components such as vegetation, microclimate, and breeding habitat availability likely playing greater roles in the distribution of individuals than neighborhood scale metrics. Still, while other studies have found that *Ae. albopictus* infestation is related to urban decay, resident income, and local resident knowledge [54,75–77], this is the first time that the additional socioeconomic metrics of foreclosure rate, violent crime rate, and an index incorporating education, home sales price, and household income have been connected to gravid vector abundance. However, it is unclear how exactly some of these variables are linked to mosquito ecology. While we did not find correlations between independent variables, there may be certain ecological features important to *Ae. albopictus* that link these socioeconomic variables to each other. For example, future studies may investigate relationships between violent crime rate and local abundance of abandoned or dilapidated structures, which may subsequently harbor mosquito breeding habitat. This can aid public health risk assessments by isolating relationships between socioeconomic variables that may directly influence disease risk in a community. Since *Ae. albopictus* breeds in water-filled man-made containers, these socioeconomic variables would seemingly have some positive relationship on breeding habitat availability. This was found in Bakersfield, California [78], where housing delinquency following economic downturn led to neglected swimming pools and other unmanaged water-collecting containers. Similarly, the accumulation of waste in India [79] and Texas [80] as well as the presence of dilapidated urban structures in low-income city blocks in Washington D.C. [81] have been found to be associated with increased vector breeding habitat. However, the relationship is not consistent. For example, in Baltimore in 2014, Becker et al. found that mosquito production was higher in high SES neighborhoods than low SES neighborhoods, hypothesized to be related to containers being regularly yet artificially supplied with water, while unmanaged rain-filled containers in low SES neighborhoods dry up too quickly to be utilized for breeding. In general, while vector ecology has been well-studied, we suggest that future studies focus more diligently on conditions in the host community that may support ecological underpinnings. This not only includes examinations of socioeconomic drivers of vector abundance, but can also involve studies of resident knowledge of risk or variations in vector control practices.

Overall, these studies hold important implications for public health, as exposure to vector-borne diseases can be viewed under the lens of health disparities. Communities with lower relative educational attainment [82], poor economic stability [83], and lower property values [84] are known to be at a disproportionate risk of a myriad of health concerns. Our findings indicate that higher exposure to vector-borne diseases is an additional risk factor that can further erode the health of these already disadvantaged community members. Indeed, relationships between socioeconomic or demographic predictors and mosquito-borne disease risk have been established for West Nile virus [85–87], St. Louis encephalitis [88], and dengue virus [89,90].

As previous studies did not utilize GATs, our results provide indication that while gravid *Ae. albopictus* abundance is higher in neighborhoods of lowered socioeconomic status and high landscape diversity, human biting rate may also be higher in such neighborhoods. Our traps solely captured females who have consumed a bloodmeal, and while *Ae. albopictus* host preference is more diverse than *Ae. aegypti* [91], studies in urban areas indicate that humans still comprise 80–100% of *Ae. albopictus* host targets [7,9]. Further studies specifically focused on biting rates and host preferences would be needed to confirm this hypothesis.
Additionally, the determination that gravid *Ae. albopictus* abundance is higher in neighborhoods with high proportions of residents that identify as Hispanic is the first known published relationship between gravid vector abundance and local ethnicity. Additional studies would be needed to determine how variables correlated with particular ethnicities are related to vector ecology. Knowledge, attitude, and practice surveys [92–94] in potentially at-risk communities may be a useful method of understanding how cultural practices can lead to increased contact with host-seeking *Ae. albopictus*. Regardless, with higher gravid *Ae. albopictus* abundance than sites with other predominant ethnicities, we suggest that public health authorities take particular care to address mosquito-biting risk in the Hispanic community, including designing specialized bilingual public education or information campaigns as-needed.

The relationship between vector abundance and land cover has been studied in diverse regions including Thailand [95], Hawaii [96], South Dakota [97], Peru [98], and Chicago [99]. Our demonstration that land cover variables have no meaningful effect on a model which already includes socioeconomic variables may indicate the reduced scale of impact of landscape metrics on gravid *Ae. albopictus* abundance compared to other neighborhood characteristics. However, as land use change continues into suburban and rural areas, increased vector suitability and potential for virus transmission may be a consequence of such development [100,101] and municipal vector control efforts should increase proportionally.

It is important to note that while we have indeed illustrated key relationships between socioeconomic variables and gravid *Ae. albopictus* abundance, there are numerous other variables we did not examine with hypothesized or proven links to *Ae. albopictus* ecology. In future surveys, combining socioeconomic attributes with other meaningful independent variables such as climate, interspecific-competition, predation, and vector control would reduce missing variable bias and aid in building a more comprehensive picture of vector distribution and potential virus transmission risk.

5. Conclusion

Our study provides a novel vector surveillance approach which accounts for the true distribution of socioeconomic attribute values in an urban region. We intend for this design to be compared and contrasted with existing surveillance approaches so that authorities can maximize the validity of their results. In our study, we used socioeconomic variation across the urban region as the central focus of the optimization, though there is potential to use any other data that may be available at the neighborhood scale to identify ideal sample sites, including environmental, physical, social, or cultural neighborhood attributes. We found that gravid *Ae. albopictus* abundance is negatively related to several neighborhood socioeconomic characteristics and demonstrate the importance of understanding vector-borne disease risk within the context of social determinants of health. In general, the body of literature examining *Aedes* ecology and urban socioeconomic variation is small yet growing, and further investigation will be required to better understand the nature and implications of this important relationship.

**Authors Contribution:** A.W. conceived of the study concept in addition to supervising and conducting the study, analysis and writing. E.D. was responsible for developing core aspects of the methodology as well as data visualization, and manuscript editing. T.R. assisted with methodology and also provided manuscript editing. S.C. and G.C. contributed methodological assistance along with manuscript editing. M.D. provided overall supervision funding acquisition, and manuscript editing.

**Funding:** This study was funded by the Academy for Population Health Innovation, a collaboration between Mecklenburg County Public Health and The University of North Carolina at Charlotte

**Conflict of Interest:** The authors declare no conflicts of interest.

**Reference;**


46. United States Census Bureau Population and Housing Unit Estimates. **2010**.


84. Portney, P.R. Housing prices, health effects, and valuing reductions in risk of death. *J. Environ. Econ. Manage.* 1981, 8, 72–78.


87. Liu, H.; Weng, Q.; Gaines, D. Geographic incidence of human West Nile virus in northern Virginia, USA,


