ASSESSMENT OF SOCIAL WASP COMMUNITIES IN FRAGMENTED FOREST SYSTEMS OF NORTHWEST OHIO

Matthew C. Tobey

A Thesis
Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE

July 2025

Committee:

Karen V. Root, Committee Chair
Daniel M. Pavuk
Kevin E. McCluney

ABSTRACT

Karen V. Root, Committee Chair

Social wasps (*Vespidae*) are a group that maintain high connectivity within food webs, pollination networks, and nutrient cycling regimes, contributing to the resilience of natural systems. Previous research has indicated that a negative relationship exists between social wasp community diversity and the fragmentation of natural habitats. In Northwest Ohio, the Oak Openings Region (OOR) is a state biodiversity hotspot which has seen natural landcovers become increasingly fragmented over time. This research assessed social wasp community diversity across fragmented forest stands in various protected areas in an attempt to identify relationships between these native social wasp species and both their local habitat and their surrounding landscape features. Due to the unique life histories of social wasps, careful interpretation was needed for data collected at the level of individual wasps, rather than individual superorganisms. Assessments of sample independence and an exploration of how superorganismal structure introduces uncertainty, specifically with commonly used diversity metrics, demonstrated that seemingly precise results can become highly ambiguous. Results suggest that a negative relationship could exist between increasing fragmentation and social wasp community diversity in the Oak Openings Region. However, the treatment of individual wasps as independent with neglect for the role their colony plays in their behavior can spawn a series of errors that render standard diversity metrics and data analysis methods unreliable. This points to a much greater need for care in the interpretation of social wasp data moving forward. The implications from this study extend to other social insect species.

AGNOWLEDGEMENTS

I would like to thank my family and friends for supporting me throughout the course of this work. I would also like to thank the Bowling Green State University Department of Biological Sciences, and the funding acquired via the Barbara Long Masters award. Access and collection permits allocated by both Toledo Metroparks and the Ohio Department of Natural Resources were critical for conducting this research, and I am grateful for their generosity.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Background on Oak Openings Region and Social Wasp Function

Human induced habitat loss and fragmentation has been associated with declines in the biodiversity of natural systems. These impacts are seen across taxonomic groups throughout the world. The severity of these impacts differs situationally, but fragmentation represents a pressing threat to the conservation of biodiversity. In Northwest Ohio, the Oak Openings Region (OOR) is a globally rare oak savanna ecosystem, and a biodiversity hotspot for the state (Brewer and Vankat, 2004). The OOR has experienced intense development pressure from agriculture in the south, and residential expansion in the north. Fragmentation of natural landcovers has been observed over the 10-year period from 2006 to 2016 (Martin and Root, 2020), and has likely continued in the time since. Forested landcover in particular has seen a reduction in average patch size and an increasing distance between neighboring patches.

The extensive recovery time needed for forest systems highlights the potential fragility of wildlife communities reliant on them. Regional forests are subjected to colonization by non-native, injurious species including fungal pathogens like oak wilt (*Bretxiella fagacearum*) and invasive insects like the emerald ash borer (*Agrilus planipennis*) and elm zig-zag sawfly (*Aproceros leucopoda*) (Abella et al., 2021; Herms and McCullough, 2014; Oten et al., 2023). For insects, in particular, the connection between increased fragmentation and decreased populations of higher trophic level species, capable of limiting outbreaks at lower levels, has been documented (van Nouhuys, 2005). Though fragmentation-induced losses can decrease overall biodiversity, it is critical to assess the degree to which such loss impacts a system's resilience to future disturbance, and its capacity to maintain essential functions.

Research and conservation efforts targeting keystone species provide managers the opportunity to maximize often-limited resources allocated to their efforts (Johnson et al., 2017). Social wasps act as keystone species, and make a valuable research target for forested systems. They are a group of native species which maintain high connectivity within food webs, pollination networks, and nutrient cycling regimes, contributing to the resilience of natural systems by maintaining these essential functions (Brock et al., 2021). By looking at the diversity of social wasp species, which contribute to overall system function through the provisioning of ecosystem services, we can derive indications of the natural area's resilience to these persistent disturbances.

Social wasps are members of the paraphyletic group aculeata within the order Hymenoptera. In the state of Ohio, native species are classified in the family *Vespidae*, specifically within two sub-families *Vespinae* and *Polistinae*. They are all colony-forming, eusocial species exhibiting division into reproductive and non-reproductive castes, cooperative rearing of young, and overlapping generations. Throughout this paper the terms social wasp, eusocial wasp, and wasp may be used interchangeably when referring to the previously defined group. In the comprehensive review done by Brock et al. (2021), the contribution of social wasps to all defined categories of ecosystem services is rigorously supported. Most generally recognized is their ability as predators to suppress a variety of prey species. Their larvae also serve as a high protein food source for species brave enough to raid their nest. A growing body of work has also begun demonstrating their ability as compensatory pollinators, and their efficacy as pollinators in greenhouse settings. More minor are their roles as nutrient cyclers, seed dispersers, and bioindicators.

Social wasps are considered generalist species both as predators and pollinators, utilizing a diverse set of resources to meet the colony's needs (Richter, 2000). Specifically related to prey suppression, Greenop et al. (2018) determined that functional diversity, defined by diet-breadth among other features, positively affected the suppression of prey by invertebrate predators. Richmond et al. (2005) concluded that generalist species were dominant contributors to overall ecosystem stability and potential serve as the driving force in the relationship between ecosystem function and diversity. Similarly supporting the relationship between generalist species and ecosystem stability, Zografou et al. (2020) demonstrated that highly stable species were almost always classified as generalists, and that with a maintained core of generalist species, ecological networks can be robust to disturbances associated with climate change and invasive species introductions. Consequently, extinctions of these highly connected species tend to carry the greatest potential for secondary extinctions (Eklof and Ebenman, 2006). From what is known about the ecology of social wasps and what has been gained from theoretical work on the relationship between generalist species and ecosystem stability, it can be said that diverse assemblages of native social wasps may provide a stabilizing force in the forested systems of the OOR. Their stabilizing potential has been validated for pollination in greenhouse settings (Hallett et al., 2017) and as part of integrated pest management solutions for crop pests (Southon et al., 2019).

No species of social wasp was listed on the IUCN red list as threatened or endangered as of 2017 (Fisher et al., 2019), and have not appeared on the list to this point in 2025. While true, this lack of listing should not be interpreted as evidence of stability, but should rather be viewed skeptically. Social wasps are an understudied group that experiences both public and academic bias (Sumner et al., 2018). This bias potentially limits the data collection efforts needed to make such a determination of status. Social insects as a group maintain a set of features that allows them to be ecologically dominant, however these traits also maintain sensitivities that could be realized with increasing environmental change (Fisher et al., 2019; Chapman and Bourke, 2008). Despite their abundance on the landscape, common species are capable of significant losses in the face of environmental shifts (Gaston and Fuller, 2007). Studies in Brazil have found that fragments and greenspaces are capable of hosting diverse social wasp communities and alleviating some of the species loss generated by fragmentation (Clemente et al., 2020; Detoni et al., 2018). However, major dissimilarities in social wasp communities between continuous and fragmented forests have also been observed (Graca and Somavilla, 2019). Similarly, a study in Poland on the connection between social wasp diversity and deforestation indicated that declining forest habitat detrimentally impacts overall diversity (Pawlikowski et al., 2016). Independent of species designation, the evidence to this point suggests the potential for social wasps to experience negative effects of fragmentation, and that their study and conservation are needed independent of their current designation.

Social Behavior and Data Collection and Analysis

The behavior of these colony-forming, eusocial species presents unique challenges for identifying community level diversity metrics and habitat associations. These challenges are distinct from those experienced with solitary organisms. Biological individuality is a complex topic, particularly for social insects, where differentiation between organism and colony becomes unstable (Santelices, 1999). The superorganism concept was developed to alleviate this ambiguity (Wilson and Sober, 1989). It declared that natural selection acted on a colony as a social unit, rather than on each individual wasp. For effective conservation management, quantitative inferences about wildlife populations require that assessments be conducted at the level of individual, where selection pressures directly operate. These data may subsequently be aggregated to higher levels such as population and community. For studies on social wasps, and social insects in general, it is common practice for data to be collected at the level of workers and extrapolated to infer trends at these higher levels of organization (Clemente et al., 2021; Jacques et al., 2018; Rubene et al., 2015). This translation of worker abundance data, as a subset of the functional superorganism, directly into population counts without first aggregating counts into individual superorganisms, may spawn a unit of analysis error. This error can also be referred to as pseudoreplication, where data is analyzed at a level which is distinct from where it was collected (Hulbert, 2009; Smith-Ferguson and Beekman, 2019).

In Northern climates, only new queens overwinter into the following year, acting as the lone bridge between generations. As a result, queens are the dominant contributor to overall genetic diversity, and the effective population size is most directly related to queens. The sampling of queens can be challenging and have adverse impacts on population abundance when lethal means are employed. The number of colony nests can thus represent a population's reproductive potential, as it is the location in which current queens reside, and from which future queens will emerge. While reproductive success is highly variable (Archer, 1985; Lester et al., 2017) and the result of many interacting factors, colony nest totals can serve as a functional proxy for effective population size in the context of conservation management.

The number of colonies represented by captured workers is highly uncertain for superorganisms. Colonies of social wasps have been observed growing to enormous size. They also consistently replace workers over the course of a season, potentially allowing for one colony to be represented hundreds of times in collections throughout a season. There is also no determinate size a colony grows to. When workers are treated as equivalent to biological individuals, abundance totals can become inflated and alter the perception of community diversity (Pamilo and Crozier, 1997). With an explanatory gap preventing the consistent aggregation of worker counts into counts of individual superorganisms, the propagation of measurement error may increase concomitant with the complexity of analysis. Care must be taken when producing diversity metrics using relative or absolute worker abundance.

Similar concerns related to data reliability exist when analyzing relationships to environmental features. Native social wasp species are considered central-place foragers, returning to their nest overnight and intermittently throughout the day as they acquire resources. This behavior acts to restrict the movement of foraging workers, limiting the spatial range workers can occupy on the landscape. Restrictions on movement can bias associations to disproportionately represent those features near the nest. These nearby features may receive inflated counts due to occupation by numerous workers. It has been well established in the literature that social wasps, while not as effective in information sharing as other social species, are capable of recruiting nest-mates to forage (Taylor and Jeanne, 2018; Schueller et al., 2010; Wilson-Rankin, 2014; Reid et al., 1994). Essentially, data analysis may be overly sensitive to the influence of larger colonies because they can contribute disproportionately to the size of the data set, and represent a biased sub-set of features as a result of their nest location. If detected, worker recruitment and central-place foraging may preclude data points from being treated as independent. Therefore, colony location and worker interactions can function as omitted variables, creating potentially biased estimates (Wilms et al., 2021).

Objectives and Expectations

With an awareness of the complexity of social wasp life history and its implications, this study has three objectives, with each contingent on accomplishing those prior. The objectives are to determine 1) which social wasp species are represented within the OOR, 2) if communities occupying surveyed locations differ from each other, and 3) if any measured habitat and landscape features are associated with differences in community diversity. This will be the first study on social wasp communities in the protected areas of the Ohio portion of the OOR. Results will provide novel considerations for managers of Ohio's biodiversity, as well as create an initial point of comparison for future survey efforts. With the study of social wasps in the OOR being somewhat of a black box, emphasis will be placed on providing findings that are able to structure future research efforts geared towards social insects, specifically wasps, should anyone be inclined to peruse them.

Forest systems provide suitable nesting structure for all native species that could occupy the region. The existence of suitable nesting habitat for all species avoids bias against those nesting in above ground structures less likely to occupy earlier successional habitats. As such, in the sites visited across the Ohio portion of the OOR, we should expect to account for all species native to the region. As a generalist set of species, it is expected that all surveyed locations will be occupied to some degree by social wasps. Variability in species richness, community composition, and calculated diversity metrics between the disjunct protected areas is expected. Attempts will be made to suggest potential drivers of these differences at the habitat and landscape scale, if communities are indeed able to be differentiated from one another. Research to this point has suggested a positive relationship between forested area and overall community diversity exists. It is anticipated that a similar trend will be identified in this study as well.

MATERIALS AND METHODS

Study Sites

The Oak Openings Region spans 478 km^2 of Northwestern Ohio. Forests and woodlands are classified as either swamp forest, floodplain forest, upland deciduous forest, or coniferous forest (Martin and Root, 2020). The eight chosen survey sites were owned by either Metroparks Toledo or the Ohio Department of Natural Resources, all of which occurred in Lucas County, Ohio. Five selected locations fall within the OOR boundaries (Schetter and Root, 2011). The remaining sites are located adjacent to the OOR's eastern boundary. Ongoing forest loss has been recorded as a result of conifer removal efforts, invasive species related die-offs, and continued encroachment by development (Martin and Root, 2020). As a result, remnant forest stands have smaller patch sizes, making it challenging to find many spatially independent locations from which to collect data. Forest patches were chosen that had an area of greater than 250 m^2, and were adequately separated from each other to help ensure that workers collected at each site represented distinct colonies. The sites which met those criteria were: Oak Openings Preserve Metropark (OOMP), Secor Metropark (SEC), Fallen Timbers Battlefield (FTB), Swan Creek Preserve Metropark (SWCR), Wildwood Preserve Metropark (WW), Pearson Metropark (PEAR), Maumee State Forest (MSF), and a wildlife corridor research unit (COR-02). Foraging ranges surrounding social insect nests are unlikely to be known (Gotelli et al., 2011) and are the result of complex interactions between the colony's size and surrounding resource availability (Reynolds et al., 2009; Reynolds, 2009). There is little understanding of foraging ranges for social wasp species. It has been estimated that foragers can be found within 1,000 yards (914.4 meters) of their nest (Clemson University, n.d.). It was assumed that workers within 2,000 yards (1828.8 meters) of each other could occupy the same nest of origin.

Active Survey

Surveys were conducted in 2024, beginning in June and continuing until the end of August. Each of the eight chosen sites received six visits over the three-month time period. Surveys began at 09:00 (9 a.m.) and lasted until 16:00 (4 p.m.), unless weather or safety concerns halted collection activity prematurely. Days with high wind or sustained rain were not sampled, and collection was postponed until weather was more conducive to colony activity.

Sample effort was recorded using a stopwatch which ran uninterrupted unless activity was halted in order to record data on a collected specimen, during which the stopwatch was paused. A meander survey was conducted using an entomological sweep net for specimen collection. Upon capture, specimens were preserved in a 70% ethanol solution for later species identification. Total time spent searching was recorded at the end of each day.

Following specimen collection, habitat features in the immediate vicinity of the specimen were recorded along with GPS location and time of day. A 1 m^2 quadrat was used to record the local habitat composition. Composition was recorded as a percentage of the quadrat occupied by bare ground, vegetation, woody debris, leaf litter, rock, floral resources, or other. Structural complexity of the habitat was measured through canopy cover and vegetation density. Canopy cover was recorded as an average percent of four measurements, one taken facing each cardinal direction on the point of specimen collection. These were used as relative measures, and were acquired using the mobile app Canapeo. Local vegetation density was recorded using a Robel pole. Average measures of density were collected by observing vegetation coverage (cm) at a distance of one meter from the pole's location in each of the cardinal directions. Temperature and relative humidity (percent) were recorded using a handheld weather station.

Analysis

Worker Independence

In the absence of colony location data, indirect inference as to the spatial independence of worker samples was generated by analyzing the data's clustering properties. As the result of user error in the collection of GPS location data on specimens, some sample points were incorrectly recorded as being outside of the sample location, without the ability to retroactively plot their location. Due to this error, six location points were excluded from the samples in both tests of spatial independence. Points were also excluded if they occurred outside of the focal sampling area in the larger parks in a manner that could distort results. This was done to correct for the sampling approach. One example occurred in Oak Openings Preserve Metropark, where specimens were collected in land management units over 1,000 m away from 86% of the samples collected. Analysis was focused on areas that were sampled on multiple occasions during the three-month period.

Nearest-neighbor analysis was used to detect if captured specimens demonstrate a greater apparent affinity for members of their own species than would be expected by chance. This was analogous to a chi-squared test (McHugh, 2013). GPS coordinates and species identity of captured specimens were plotted onto a map in GoogleEarth, and the nearest neighboring specimen was identified using the measure tool. For individual species, each specimen whose nearest neighboring point was of the same species identity was counted. This value was then divided by the total number of representatives for that species collected within that particular site. This proportion was then compared to the chance that two specimens sharing a species identity would be selected contingently out of the total population sample. Calculated values were then assessed to detect if what was observed aligned with what was expected by chance.

A second method used in the detection of sample independence was the Ripley's K-function in the Spatial Statistics toolbox of ArcMap (Ripley, 1976). GPS locations were separated by site and species, and then plotted as X-Y coordinates in the ArcMap program. A rectangle delimiting the study area was plotted, representing the boundary within which samples were collected. One hundred distance bands beginning at 0m from each plotted point were recorded. Two trials were run, one where distance bands increased at the rate of 1m per iteration and one where the increase was 5m per iteration.

For each capture plotted in ArcMap, the K-function took the sum of neighboring points that occurred within a specified distance band, not including itself, and normalized the count to the total sampled area. The K-function is then transformed into the L-function for ease of visual interpretation. Output values were then plotted for the observed data and compared against what was expected under the assumption of uniformly distributed points. If the observed output falls above the line representing a uniform distribution, $L(d) = d$, it represents a clustered distribution on the landscape. If below the line, it represents a dispersed distribution of points.

The expectation of uniformly distributed samples serves to indicate clustering behavior, but the test for significance used a more biologically realistic binomial point process to generate simulated outputs using the same number of samples observed in the data set. 999 permutations of the sample were generated and analyzed under this assumption of complete spatial randomness (CSR) to produce a confidence envelope. The envelope was composed of the most extreme values obtained in the 999 simulations for each distance band. Sample and CSR points involved in the process were confined to the delimited study area, and no boundary corrections provided in the software were implemented. Results were plotted on a graph including the values for observed data along with the expectations under both a uniform and binomial point process.

Once the extreme values generated in the confidence envelope were plotted, any observation occurring at a more extreme value would be considered significantly different than expectation under CSR.

It is recommended that Ripley's K-function be computed with 30 points or more, so only the most abundant species, *V. maculifrons*, was reported. It is important to note, despite the limited sample size, that when fewer points are used in the generation of the 999 permutations under CSR, a wider and more unstable confidence envelope is produced. Significant differentiation above such an envelope with a small sample size would be relatively robust to Type I error, and it could be argued that given this volatility, significance would be more challenging to reach unless a strong signal were to exist.

Community Diversity

Alpha-diversity was compared within sites using the Shannon diversity index. Beta-diversity was calculated using the Bray-Curtis dissimilarity index. Shannon diversity provides more representation in the final metric for novel species (Bollarapu and Ramarao, 2021), and the use of this metric in other studies of social wasps allowed for greater commensurability of final results with those reported globally. Bray-Curtis dissimilarity is also a frequently employed measure when comparing community composition in social insects (Schuepp et al., 2012; Schwarz et al., 2024; Steiner et al., 2005; Wilson et al., 2020).

Each site's abundance data was separated by species. In order to account for the number of individual superorganisms within the area despite lacking knowledge of colony totals, a range of biologically realistic colony totals were created for each species. Due to the fact that no constant ratio of workers to individual superorganisms could be reliably produced, the range assumed all specimens represented one colony, they all represented unique colonies, and they represented each whole number between the previous two values. This was done for each species within a site, and Shannon diversity values for all possible combinations of a community's composition were calculated. Additionally, Shannon diversity values generated from the abundance data, as well as presence-absence counts, were calculated.

A slightly different protocol was used for obtaining the realistic range of values for Bray-Curtis dissimilarity. This assessment had a much greater computational burden than the Shannon diversity calculations because incorporating uncertainty across two communities, as opposed to only one, generated a much larger set of possible combinations as uncertainty compounded between them. Rather than assessing the possible dissimilarity values for all realistic community compositions, the biologically realistic ranges of colony totals were used to identify the maximum and minimum dissimilarity between all sampled communities. This was done by artificially selecting the observed potential colony totals that would maximize or minimize dissimilarity values to understand the range of uncertainty within the bounds of this equation.

In order to maximize dissimilarity produced by the Bray-Curtis equation, the larger abundance total for a species between two communities in question was assumed to represent unique colonies, while the lower value was reduced to the assumption that all captured workers represent the same colony. This inflated both the numerator and denominator of the equation producing the closest value to one (total dissimilarity) possible for a given data set. The inverse was done when calculating the minimum dissimilarity value. The lower of the two abundance totals for a particular species was held constant with the abundance total of the other community was reduced to match that lower value. This maximized contribution to the denominator of the equation while attempting to have values cancel to zero in the numerator. For comparisons involving a species present at one location but not at another, the minimum possible value for colonies, one, was selected in order to contribute the minimal possible weight to the equation's numerator. Ultimately, this protocol divided the pair of communities' smallest value attainable for difference in possible colony totals, by the largest value attainable under that precondition, for total observed specimens. Thus, it drove the value to its lower limit of zero. Zero was only obtained for sites in which the observed species were identical, and therefore capable of direct cancellation with no contribution to the numerator of the equation. Dissimilarity values based on the total abundance data and presence-absence data were also calculated in order to compare their position within the possible range defined between maximum and minimum values.

RESULTS

General Overview

Search time for each site averaged 26.37 hours, with overall search effort across sites totaling slightly over 211 hours (Table 1). 275 wasps were collected in total from all sampled locations, with values for individual sites ranging from 22 to 47 individuals (Table 2). From none of the sites were all 7 species collected; species collected per site ranged from 2 to 5 of the seven species. This survey spanned the months of June, July, and August. Over the course of these months both queens and workers were collected. The first worker was collected on June 26, 2024. The last queen was collected on July 27, 2024. Between those points in time, collected specimens transitioned from queen dominated samples to exclusively contain members of the worker caste.

The surveyed forested areas yielded seven of the eight expected native species. Surveys failed to detect any of the invasive wasp species reported to exist in Ohio. However, citizen science submissions through iNaturalist captured the invasive *Vespula germanica* and *Polistes dominula* in the surrounding areas, in addition to the lone undocumented native *Vespula squamosa* (Table 3). The habitat variables recorded upon specimen capture are displayed in Appendix A and show variability in species associations across all sites. Despite this variability, comparisons between species across multiple sites generated some significant differences in associated features commonly applied statistical tests. These results are omitted because of violations of a number of assumptions and a specific instance of significance testing's lack of reliability in this study is elaborated on in the discussion.

Worker Independence

Nearest-Neighbor Analysis

Testing for sample independence, through the calculation of nearest-neighbor proportions, showed all but two locations (FTB and COR-02) displayed large discrepancies between observed proportions and what was to be expected by chance (Table 4). FTB failed to display clustering above expectation for any species. However, two of the four species collected were only represented once, and consequently could not share an identity with a neighboring specimen, forcing both the observed and expected values to zero. Ultimately, the only species factored into the comparison at FTB were *V. maculiforns* and *V. flavopilosa.* The latter is capable of usurping *V. maculifrons* colonies and is not known to initiate nest building, suggesting it acts as an obligatory nest parasite (Borowski et al., 2021). The same pair of species were the only two represented for samples taken at the COR-02 property. For this location, as with FTB, assessing the potential structure of colony induced clustering by comparing species identity failed to realize its desired function. However, Table 4 is provided to facilitate a comparison of species proportions among sites, but with this caveat that the nearest-neighbor results are likely meaningless for both FTB and COR-02. Overall, this analysis, in conjunction with what is known about vespine social behavior, indicate that samples lack independence.

Ripley's K-Function

Vespula maculifrons was the most abundant species for all surveyed locations. Analyzing their spatial clustering pattern with the Ripley's K-function showed that the output values at local scales exceeded the line produced under the assumption uniform distribution for nearly all radius values. When distance bands increased at the interval of 1m per iteration, once radius values exceeded 15m, no subsequent output value in any location returned below the uniform distribution line. In other words, there was a clustered pattern, not uniform or random, for *V. maculifrons* at all surveyed locations at the local scale (0-100m).

999 permutations were generated under the assumption of complete spatial randomness (CSR), and the most extreme values were plotted to create a confidence envelope. The output values of *V. maculifrons* exceed this confidence envelope boundary in all sample locations. Output values across sites persist above the confidence envelope for variable durations of radii values. The observed clustering of observations at local scales, represented by our observed data's position above the confidence envelope, indicates significantly non-random spatial structure exists in the collection of worker specimens. The site-specific graphs are displayed in Appendix B.

Community Diversity

Shannon Diversity

The potential values for Shannon Diversity were explored using a combinatorial approach that generated closed intervals of varying lengths, rather than single point estimates typically produced by diversity metrics (Fig. 1). Community composition and the algorithmic properties of Shannon diversity determined the interval lengths recorded for each site. In other words, instead of a single value this approach simulates a range of plausible Shannon diversity values for each site based on the specimens collected and assumptions about the relationship between individuals captured and corresponding number of colonies. The maximum Shannon value produced in each community was created under the assumption of equal species proportions. The maximum possible value of Shannon diversity grew concomitantly with species richness. Consequently, the algorithm produced the same maximum value for sites of equal species richness. The minimum possible value of Shannon diversity diverged between locations as the result of site-specific abundance counts for individual species. As one of the species represented in a given community became dominant, the Shannon value approached zero. The more uneven a community became, the closer it's corresponding minimum value fell towards zero. The lowest minimum values result from species poor locations that achieved high abundance totals.

Only the COR-02 site generated a distinct interval from any of the other communities sampled (Fig. 2). The closed interval of Shannon values for COR-02 fell below the interval of four other locations (SWCR, SEC, PEAR, and MSF). Significant differentiation, in the context of these intervals, exists for any set of sites that occupy mutually exclusive stretches of the number line. With such a definition of significance in place, COR-02 is the only location occupying a significantly lower range of diversity values when compared to other sites in the OOR. Accounting for the ambiguity of social insect count data, the ability to differentiate locations from each other is greatly reduced. With the exception of the COR-02 property, surveyed locations are not mutually exclusive of one another. With a uniform distribution of likelihood for all colony totals, this overlap means sites cannot be differentiated from one another with any confidence using the Shannon diversity index.

Bray-Curtis Dissimilarity

Similarly, the range of biologically realistic Bray-Curtis dissimilarity values were explored for each site based on the specimens collected and assumptions about the relationship between individuals captured and corresponding number of colonies. The resulting values spanned nearly the entire range of possibilities from 0 to 1. Minimum dissimilarity occurred when communities composed of the same species had equal specimen totals in each category. For comparisons involving a novel species present at only one site in the pair of interest, the calculation yielded a non-zero minimum dissimilarity. Biologically realistic Bray-Curtis dissimilarity values spanned nearly the entire interval from 0 to 1. For all community comparisons, the minimum possible dissimilarity value fell below 0.1. Maximum values approached 1, generated by the largest possible difference in colony totals produced from worker count data. Maximum values always exceeded 0.767. 85.7% of comparisons had maximum values above 0.8, and 14.2% of comparisons had maximum values above 0.9 (Table 5). Presence-absence, where species occurrence was recorded as a binary outcome, and unaltered abundance data from worker counts failed to exhibit stable positioning relative to each other. The range of uncertainty produced for Bray-Curtis dissimilarity, when accounting for the ambiguity of extrapolating workers to individual superorganisms, was nearly entire and not capable of reliably differentiating communities. This mirrors the findings seen in calculations of alpha-diversity from the previous section. Figures for each individual site are displayed in Appendix C.

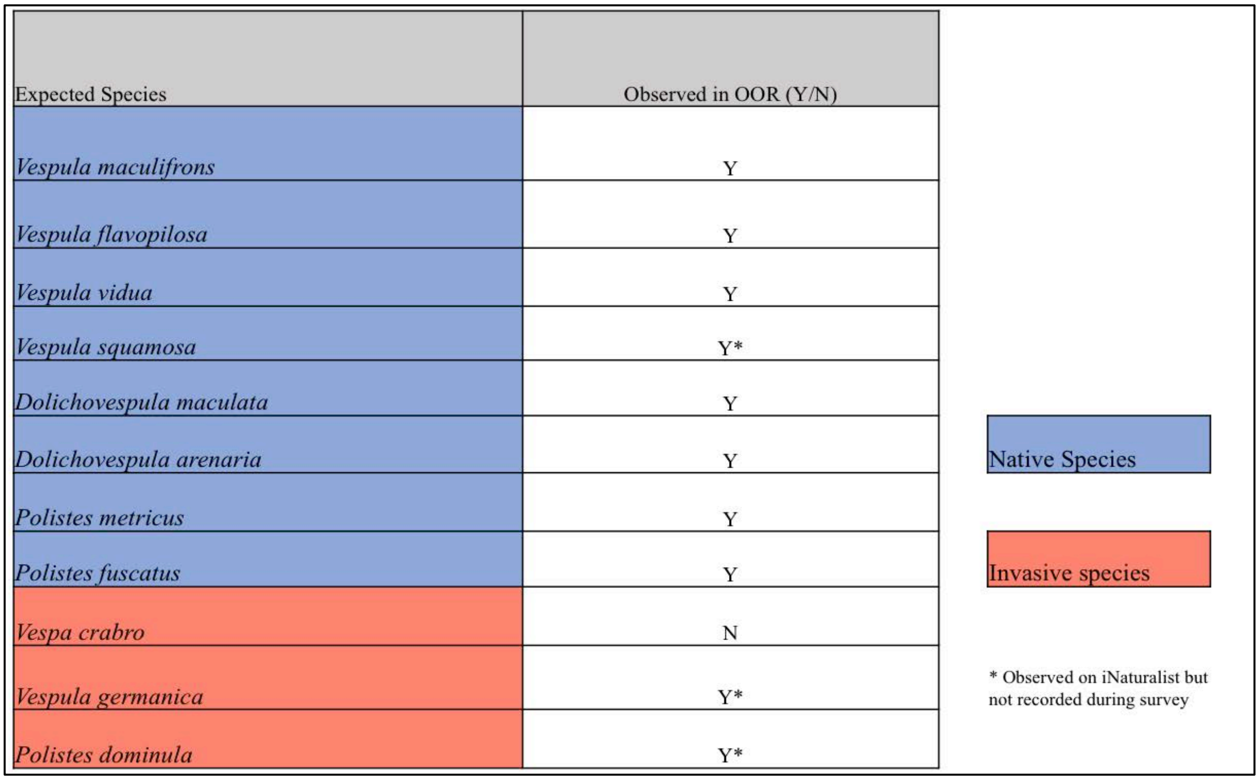
Table 1. Shown is the time spent actively searching for social wasps deconstructed by visit number and sample collection location. The bottom row indicates the total time spent searching over the entire sample period for each location.

Site Visit	Oak Openings Metropark	Fallen Timbers Battlefield	Pearson Metropark	Secor Metropark	Wildwood Preserve Metropark	Swan Creek Preserve Metropark	Maumee State Forest	COR-02
1	(6/1/24): 4h 45m	(6/3/24): 5h 43m	(6/4/24): 4h 14m	(6/6/24): 5h 29m	(6/7/24): 5h 58m	(6/9/24): 5h 43m	(6/9/24): 5h 55m	(6/12/24): 5h 15m
2	(6/15/24): 3h 54m	(6/14/24): 4h 58m	(6/16/24): 4h 42m	(6/16/24): 3h 54m	(6/21/24): 5h 22m	(6/20/24): 5h 30m	(6/24/24): 5h 22m	(6/22/24): 5h 19m
3	(7/1/24): 3h 56m	(6/30/24): 6h 10m	(6/28/24): 5h 09m	(6/27/24): 6h 28m	(7/6/24): 4h 32m	(7/4/24): 3h 54m	(7/3/24): 4h 20m	(7/2/24): 4h 32m
4	(7/9/24): 3h 40m	(7/11/24): 6h 06m	(7/7/24): 4h 42m	(7/13/24): 4h 29m	(7/15/24): 4h 44m	(7/22/24): 2h 44m	(7/12/24): 0h 53m	(7/14/24): 3h 48m
5	(7/27/24): 3h 28m	(7/28/24): 3h 44m	(7/29/24): 1h 59m	(7/31/24): 4h 27m	(8/1/24): 4h 03m	(8/2/24): 2h 54m	(8/4/23): 3h 15m	(8/5/24): 4h 00m
6	(8/9/24): 4h 46m	(8/7/24): 4h 24m	(8/21/24): 4h 25m	(8/11/24): 4h 37m	(8/12/24): 2h 44m	(8/22/24): 2h 23m	(8/23/24): 3h 41m	(8/24/24): 4h 02m
Total	24h 29m	31h 05m	25h 11m	29h 24m	27h 23m	23h 08m	23h 26m	26h 56m

Table 2. Shown are the total number of specimens collected for each species within the genera *Vespula*, *Dolichovespula*, and *Polistes* at each location over the duration of this study.

	COR-02	**Fallen Timbers Battlefield**	**Maumee State Forest**	**Oak Openings Preserve Metropark**	**Pearson Metropark**	**Secor Metropark**	**Swan Creek Preserve Metropark**	**Wildwood Preserve Metropark**
V. maculifrons	**23**	**15**	**19**	**25**	**17**	**15**	**16**	**23**
V. flavopilosa	**9**	**5**	**5**	**4**	**3**	**6**	**4**	**4**
V. vidua	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**
P. metricus	**0**	**1**	**1**	**2**	**19**	**3**	**1**	**1**
D. maculata	**0**	**1**	**3**	**0**	**7**	**2**	**4**	**0**
D. arenaria	**0**	**0**	**15**	**4**	**0**	**6**	**9**	**0**
P. fuscatus	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**

Table 3. Shown are the species within the genera *Vespula*, *Dolichovespula*, and *Polistes* that were observed in the Oak Openings Region (OOR). The left column lists the names of species that were expected to occur in the OOR. The adjacent column indicates if they were observed during the sample period in the region.



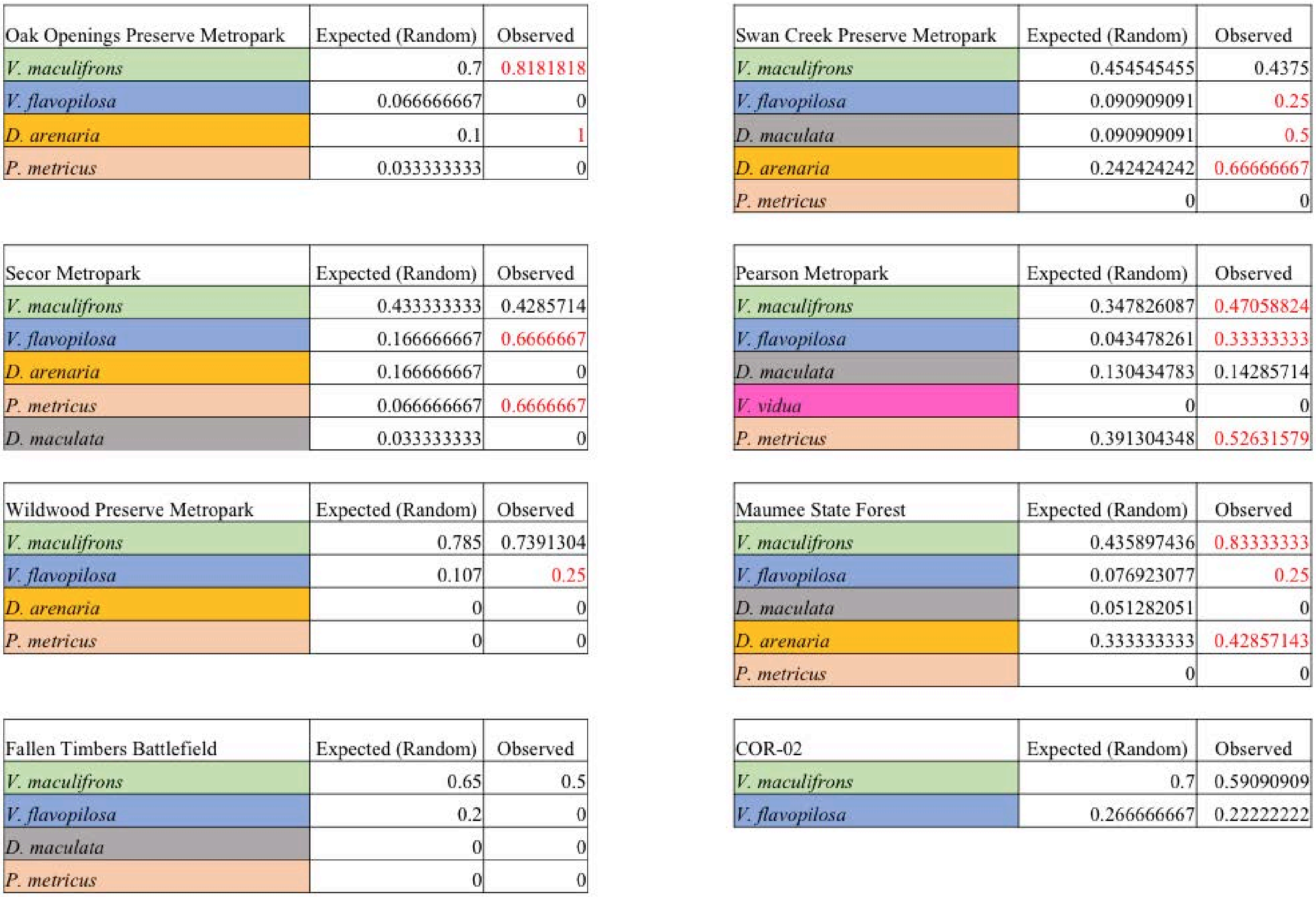
Expected Species	Observed in OOR (Y/N)
🟦 *Vespula maculifrons*	Y
🟦 *Vespula flavopilosa*	Y
🟦 *Vespula vidua*	Y
🟦 *Vespula squamosa*	Y*
🟦 *Dolichovespula maculata*	Y
🟦 *Dolichovespula arenaria*	Y
🟦 *Polistes metricus*	Y
🟦 *Polistes fuscatus*	Y
🟥 *Vespa crabro*	N
🟥 *Vespula germanica*	Y*
🟥 *Polistes dominula*	Y*

🟦 Native Species

🟥 Invasive species

* Observed on iNaturalist but not recorded during survey

Table 4. Nearest neighbor analysis for individual species within the genera *Vespula*, *Dolichovespula*, and *Polistes* at each sample location. Red text highlights when the observed percentage of nearest-neighboring samples sharing a species identity exceeded the expectation under random chance.



Oak Openings Preserve Metropark	Expected (Random)	Observed
V. maculifrons	0.7	0.8181818
V. flavopilosa	0.066666667	0
D. arenaria	0.1	1
P. metricus	0.033333333	0

Swan Creek Preserve Metropark	Expected (Random)	Observed
V. maculifrons	0.454545455	0.4375
V. flavopilosa	0.090909091	0.25
D. maculata	0.090909091	0.5
D. arenaria	0.242424242	0.66666667
P. metricus	0	0

Secor Metropark	Expected (Random)	Observed
V. maculifrons	0.433333333	0.4285714
V. flavopilosa	0.166666667	0.6666667
D. arenaria	0.166666667	0
P. metricus	0.066666667	0.6666667
D. maculata	0.033333333	0

Pearson Metropark	Expected (Random)	Observed
V. maculifrons	0.347826087	0.47058824
V. flavopilosa	0.043478261	0.33333333
D. maculata	0.130434783	0.14285714
V. vidua	0	0
P. metricus	0.391304348	0.52631579

Wildwood Preserve Metropark	Expected (Random)	Observed
V. maculifrons	0.785	0.7391304
V. flavopilosa	0.107	0.25
D. arenaria	0	0
P. metricus	0	0

Maumee State Forest	Expected (Random)	Observed
V. maculifrons	0.435897436	0.83333333
V. flavopilosa	0.076923077	0.25
D. maculata	0.051282051	0
D. arenaria	0.333333333	0.42857143
P. metricus	0	0

Fallen Timbers Battlefield	Expected (Random)	Observed
V. maculifrons	0.65	0.5
V. flavopilosa	0.2	0
D. maculata	0	0
P. metricus	0	0

COR-02	Expected (Random)	Observed
V. maculifrons	0.7	0.59090909
V. flavopilosa	0.266666667	0.22222222

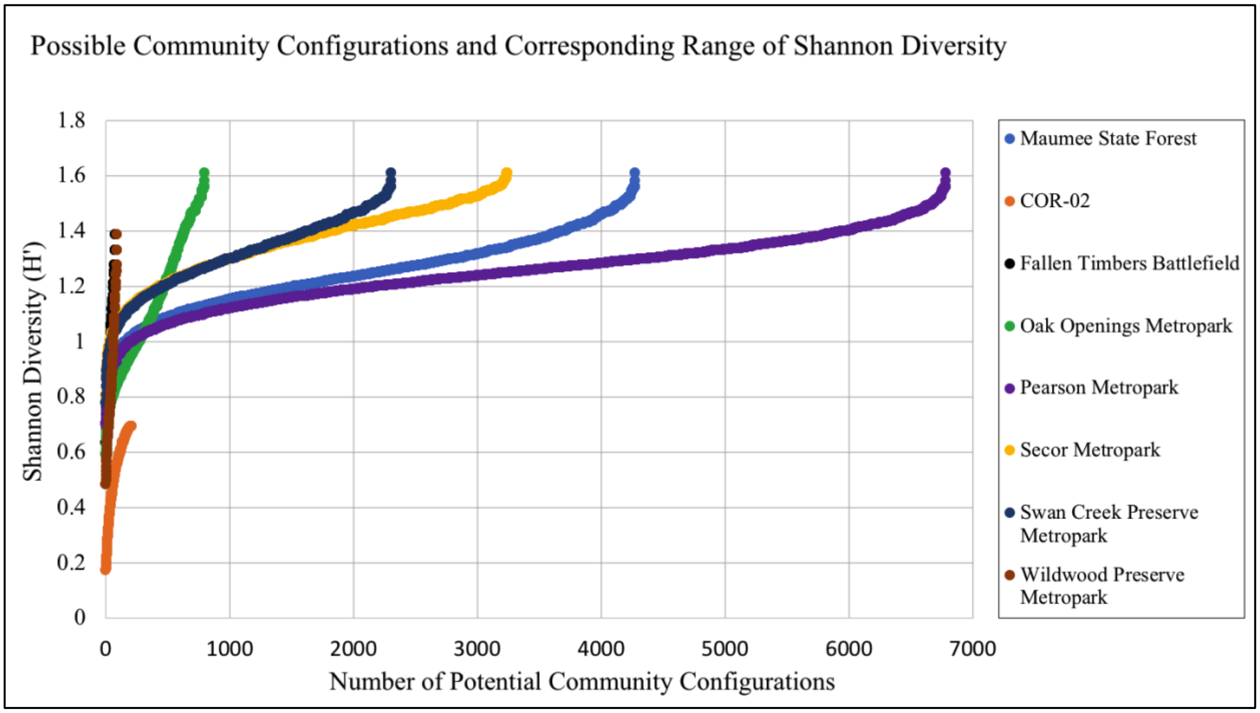


Figure 1. The graph shows all Shannon diversity values produced for each sample location plotted in ascending order. Each point on the graph represents the Shannon diversity value of one possible combination of a given sample location's community under colony count uncertainty. The horizontal length of a given location's point series across the x-axis is intended to display the relative size of each location's number of potential configurations. The vertical length of a given location's point series across the y-axis is intended to display the possible range within which a sample location's true Shannon diversity value would exist.

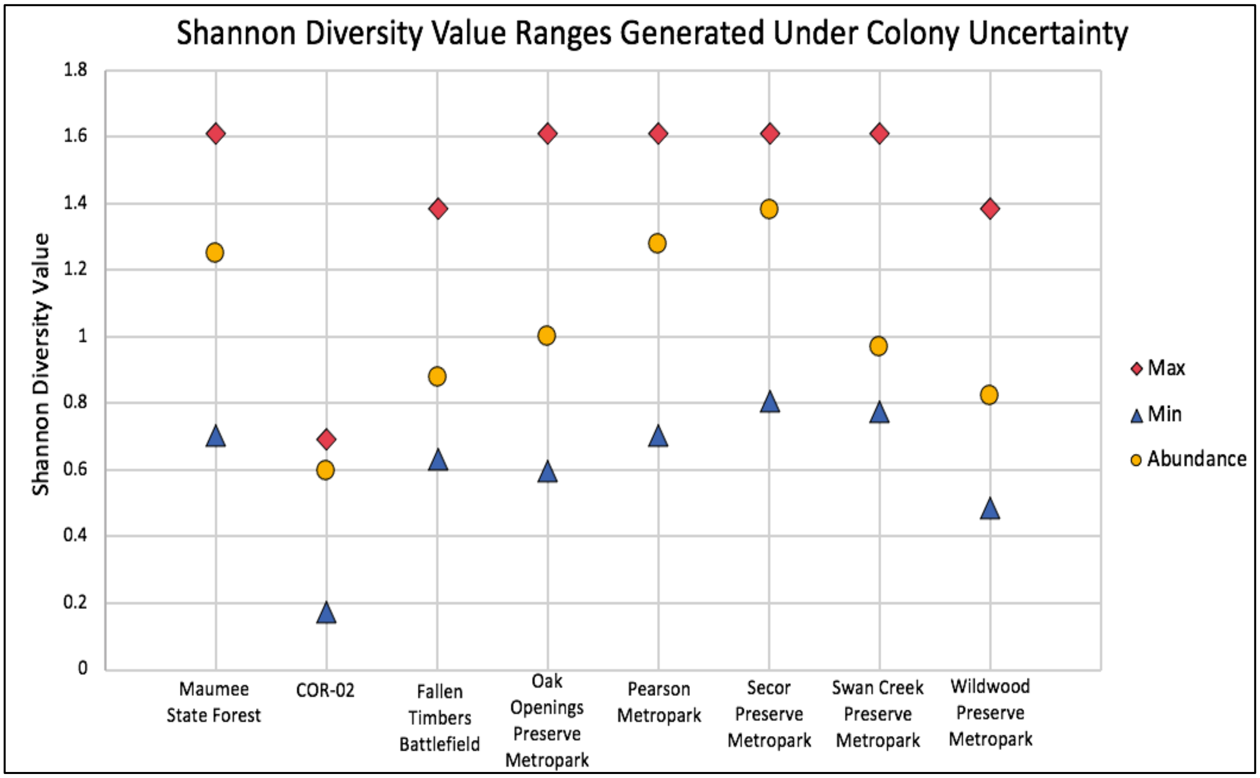


Figure 2. The ranges of Shannon diversity values generated under colony uncertainty separated by site, highlighting both the maximum (red) and minimum (blue) values produced by the combinatorial approach, in addition to the value produced by the unadjusted abundance (yellow) data.

DISCUSSION

All but one of the expected native social wasp species were identified across the OOR. *Vespula squamosa*, the lone uncaptured native species, was documented on iNaturalist in unsampled locations within the region during 2024 (grace927, 2024), as well as in sampled locations in previous years (Kimsmith, 2019; Kimsmith, 2020; Rickbarricklow, 2023). This confirms that a source exists for future potential occupation of forested habitats by all native species. This is encouraging from a conservation perspective as native species appear to be persisting within and spread throughout the region. While occurrence data indicates many species maintain a consistent presence across the region, occupying multiple distant protected areas, the difficulty associated with interpreting abundance data means no direct conclusion on the native wasp population's status can be reliably inferred at this time. Invasive species were not detected within surveyed protected areas; however, *P. dominula* and *V. germanica* were detected via iNaturalist records (Jmfarley, 2024; Jamwright, 2024). The European hornet, *V. crabro*, was the only species not to be observed in the field or recorded in the iNaturalist data base within the OOR as of Fall 2024. Similar to native species, the documented presence of invasive species in surrounding areas maintains their potential to colonize the surveyed forested locations. The current impact of these invasive species in the OOR is unknown, but their potential impact on populations of conservation concern has been acknowledged (Baker and Potter, 2020).

Survey data on species occupation was collected via active search, and specimens were removed lethally for proper identification to species level. Lethal sampling methods are standard protocol in research of this type, however, impacts on the system accrue as sampling extends throughout a season. Collecting flying insects in the field can be a difficult process due to variable environmental conditions and the rapid, erratic movement of the target. Despite a noticeable size discrepancy, differentiating queens from workers prior to collection, in a state of active movement, is challenging and potentially not feasible, especially when sample size is of concern and future collection opportunities are uncertain. This type of lethal sampling could result in the significant loss of individual queens, their subsequent future colonies, and ecosystem services provided within protected areas. Based on these concerns, sampling effort, aimed to mitigate the loss of queens to lethal collection methods, should take place after the critical transition where queens are unlikely to be encountered foraging outside of their nest. This study identified the time period following July 27th as the most resource conscious period for lethal surveys of wasps to occur. Moving collection permit windows to open in the month of August will likely reduce the impact of sampling on overall population totals tremendously. Lester et al. (2017) and Komonen et al. (2025) show a similar onset of rapid growth occurring around the start of August in the northern hemisphere. If this collection restriction is enforced, a greater number of workers are likely to be available for both specimen collection and the maintenance of overall colony function under this mortality pressure.

Communities displayed variability in both the total number of samples collected and their overall richness. Distinct point estimates for Shannon diversity created a definite rank-order of sites in their community diversity based on total abundance of worker counts. The order of these sites' diversity metric outputs aligns with what is believable from memory of sampling. The most diverse sites are remembered as having relatively continuous sightings for a wider array of species compared to those with lower values. Sites with lower diversity values are remembered as having long stretches of time between sightings and collections were often of the same species. Despite this alignment between the calculated results and the memory of the events as an observer conducting the survey, the complications brought about by social insects as a superorganism require analysis to extend past this point of logical consistency. Once the ambiguity of translating worker counts to the unit of analysis, individual superorganisms, is incorporated into the calculation of Shannon diversity, the well-defined rank-order of sites disappeared. Only the COR-02 property was mutually exclusive of the biologically realistic ranges generated in four sites across the OOR. The lack of disparity between communities prevented further association of diversity with landscape or habitat features. Their ordering is not justified past the COR-02 site being below at least four of the seven other locations. The remaining seven locations can themselves achieve any order with support for such a state's existence. Similarly, it is not possible to identify significant relationships between social wasps collected and their environment because of the challenge of identifying the number of colonies represented in the data collected, which should be the unit of comparison rather than individuals.

The problem mentioned in the previous paragraph is unique in that the standard solution for creating more reliable results, increasing the size of samples, not only fails to clarify results, but actively propagates the uncertainty surrounding those results. Even with major differences in community richness between locations, large samples of one species in a community can drive a potential Shannon diversity output towards zero by lacking evenness of samples. This range of Shannon diversity values moving towards zero as the collection of a single community species grows would eventually overlap the ranges of all less species rich communities along its path. Similarly, the Bray-Curtis dissimilarity output would see uncertainty in the range of potential values grow with sample size in the same manner identified above with Shannon diversity. Each additional sample collected brings about a disproportionately larger number of possible configurations as sample size grows. Based on these issues, to appropriately estimate diversity metrics of social wasps that require more than presence/absence data, careful attention should be paid to the unit of comparison (colony versus individual) and the consequences of the assumptions made.

While biases often lead to erroneous counts, it is essentially unavoidable in field studies. The question becomes whether these erroneous counts have the potential to support poor decision making (Elphick, 2008). The lack of consistent transferability between the level of observation (individual wasp specimen) and the level of analysis (superorganism), which will be subsequently aggregated into populations and communities, raises serious concerns about the defensibility of social wasp research findings based on such observations. Is the resulting analysis capable of providing unambiguous recommendations, or will this translation to higher levels of organization create uncertainty that obscures conclusions? It is critical this question be asked when applying common wildlife sampling techniques to social insects. Social insects globally have become species of concern, representing threatened species, invasive species, and species of public health concern. Ensuring appropriate conclusions are derived, given the methodologies used and their accompanying limitations, is crucial in maintaining transparency.

The correction of diversity metrics has been suggested in previous social insect research. Ferrari and Poldini (2022) used relatedness to lessen the impact of large individual superorganisms on the output Shannon diversity value. Such a correction could remove ambiguity and place social species on a commensurable scale with solitary species. Social wasps are however a group with highly volatile population totals from one year to the next, and it was suggested that the correction of Shannon diversity could be done using historic data. High variation and the lack of clear connections between current and future population structures, would require within-year data on the relatedness of workers to be considered reliable. Failing to incorporate information related to current colony totals would render these corrections empty mathematical workarounds incapable of resolving the root problem with interpreting superorganism community structure. This research recommends the calculation of diversity metrics be done using the colony totals within the surveyed area. This would not only improve metric reliability by accounting for superorganism totals directly, it could also allow for reduced mortality of workers as collection could be limited to a small number taken from the nest entrance to confirm the identity of its inhabitant. If such effort is deemed impractical, the other suggestion with acceptable levels of uncertainty is the direct comparison of richness between locations, avoiding the influence from abundance. Bastow-Wilson (2012) highlighted that communities can be effectively represented by species presence-absence data, and potentially generate more effective representations than those using abundance data.

While the analysis of community diversity between sample locations lacked obvious conclusions, the clear separation of the COR-02 property from other larger protected areas suggest that an influence on the community structure by higher-level landscape features should not be ruled out. A functional system, as a precondition, requires connectivity. With the continued development and application of remote sensing data, connectivity as an abstraction has become a measure for conservation planning and wildlife management (Kirk et al. 2023). Wildlife corridors and steppingstones are frequently implemented as conservation measures to better facilitate species movement across the landscape. However, the presence of corridors on the landscape is not sufficient by itself to ensure connectivity is achieved (Horskins et al. 2006). Differences between structural and functional connectivity may exist and can be variable between species.

The need for connectivity and the potential value of corridors in natural systems is undeniable, however there is substantial evidence that smaller fragments have a reduced capacity to host predator communities (Holt et al., 1999). This raises questions about the concomitant observation of reduced wasp community diversity at the COR-02 property, and the detection of a novel invasive forest pest (mct157106, 2024). There is potential for fragmentation to initiate predator release for lower-level consumers in these systems (Liao et al., 2017; Kruess and Tscharntke, 1994). This study reveals that connectivity via corridors may support the development and spread of phytophagous invasive populations, and pest insects in general, from smaller steppingstones into larger protected areas as a consequence of reduced suppression by predator communities. Reduced presence of social wasps, as a resilient generalist set of species, could signal a greater overall failing of predator communities composed of less resilient members. Populations are however volatile from year to year (Archer, 1985; Archer, 2014). Continued assessment of the variability of predatory insect communities in corridors and large protected areas would allow a more comprehensive analysis of the corridor's overall contribution to species transmission, both positive and negative.

Further complications stemming from social insect behavior arise when comparing the habitat associations of different species throughout the OOR. Central-place foraging restricts habitat use and subsequently the range of habitat features recorded in the data set. This restriction's influence on the final distribution of habitat feature values and statistical significance is not necessarily obvious in most cases, but undeniably creates a structure to the data not accounted for in many analyses. Both attempted methods for the detection of non-random structure, nearest-neighbor analysis and the Ripley's K-function, appeared to confirm that such structure exists in the data.

The impact of this hidden structure is most clearly identified in the comparison of canopy cover value distributions of *V. maculifrons* and *D. arenaria* in the MSF site (Appendix D). Clear segregation of the two species is apparent on the map of plotted location points. The western half was dominated by broadleaf trees and occupied by *V. maculifrons.* The eastern half was dominated by pine trees and occupied by *D. arenaria*. Both species displayed significant clustering within these distinct canopy types. Canopy cover values for pine trees were much lower than those for broadleaf trees. In this instance, the non-random positioning of collected workers surrounding their central-foraging location, and the non-random distribution of low canopy cover pine trees planted by humans, overlapped to generate a significant difference when compared to samples of *V. maculifrons.* The results of the comparison indicate the chance of a spurious relationship to be near zero (p-value=0.002) (Appendix D). However, if a colony is located within a stand of limited maximum canopy cover value, as *D. arenaria* was, a distinction could be created artificially if sample independence is assumed.

This amounts to pseudoreplication, where many collections of a superorganism are restricted within the pine stand with measurements treated as independent values distorting the final distribution of canopy cover measurements (Hurlbert, 2009). The importance of this result is not related to either canopy cover or these species specifically, but rather in the fact that the landscapes heterogeneity and clustering of similarly valued features can interact synergistically with colony-induced foraging range restrictions and nest-mate recruitment to produce false significance that is not easily recognized. The relationship between *D. arenaria* and pine trees seems to have no clear ecological justification, and is not upheld across other sites where pines are greatly reduced in their concentration. However, if a significant association is found between environmental features expected a priori to influence the distribution of workers, the potential role of chance could be ignored with findings going unscrutinized despite the clear potential for central-place foraging to distort distributions via pseudoreplication. As a result, observed habitat associations would not stem from preference alone. This is an essential requirement of modeling habitat associations, and without it, results may become uninformative or misleading (Hodgdon et al., 2021). This could lead to flawed management decisions based on belief in a trend rather than on clear evidence.

Research on the associations between foraging birds and the habitat surrounding active nest sites provides a framework for how to move forward in studying the associations of social wasp workers and their environment. The analysis of Rosenburg and McKelvey (1999) includes the nest as a central foraging location in the analysis as a feature capable of carrying explanatory power for the position of an individual on the landscape. Similar measures may be taken to assess social wasp foraging preferences in a more reliable manner, correcting for the highly influential variable omitted in many studies, including this one. By allowing a collection point's position to be explained by its distance from a nest, along with other habitat and landscape features, a more realistic interpretation of these relationships can emerge. This would require not only identifying the location of nests, but also identifying the nest of origin for each collected specimen. While clearly preferable when attempting to generate reliable conclusions, this amount of effort will likely be resource intensive.

The observed struggle to derive reliable count data at the population and community level is not unique to this study. Similar complications arise when attempting to quantify population abundance from eDNA readings (Danziger et al., 2022). The overlap in this struggle results from uncertainty about the state of the data-generating mechanism. The uncertainty's resolution is achieved by measuring totals at the organismal level, or super-organismal level in the case of wasps. Such a resolution would necessarily eliminate a need for quantification at lower levels of organization. Being the case, it is recommended that any attempted inference into the population status or community structure of social wasps be made at the colony level via active search. Tools for detecting colony locations are being developed and improved. Harmonic radar has specifically been applied in efforts to locate invasive wasp species (Sidehurst et al., 2025). Continued development and testing of these detection tools is one step that can be taken to improve the interpretability of studies such as this one.

The previous paragraph relates only to the measurement of diversity through abundance data, as this project was interested in system resilience at the population and community level using diversity as a proxy. Separate issues arise when output diversity metrics are translated into, and used as, proxies of social wasp function or impact within a system. These do not function as a reliable representation, as a colony's impact can change with its location and size. Potentially large, highly non-linear relationships between superorganism worker totals and ecosystem impacts exist. Given such complexity, in alignment with earlier statements, caution is urged when using diversity metrics to assess social wasps.

False baselines are one prominent concern when interpreting insect declines over time (Didham et al., 2020). There is obvious potential for proximity bias, where worker samples are driven by the overlap of survey location and colony foraging range around a central location, to create such false baselines. Superorganism totals could rise or fall significantly without a complementary reflection of significant variation in sample records at the worker level. An apt and ominous parallel was observed with the collapse of the Atlantic cod fisheries in the 1990s. Failure to account for biases in catch per unit effort (CPUE) resulting from the aggregating behavior of fish gave the appearance of stock stability and improvement until harvesting depleted the remaining aggregations (Rose and Kulka, 1999). This failure to analyze and skeptically probe the methods for obtaining counts carried disastrous economic consequences.

Similarly, the status of wasp populations is at best ambiguous following this study and surveys capable of accounting for genuine trends at the colony level will be required for the wasps and other social insects when determining conservation status. It is not difficult to imagine a long-term decline in colony totals being masked by few productive colonies until the loss of those remaining superorganisms leaves a species in ruin. It is recommended that the requirement to survey at the colony level for social wasps, or any social insect of management interest, be implemented immediately to establish better situational awareness for managers in the OOR.

REFERENCES

Abella, S. R., Sprow, L. A., Walters, T. W., & Schetter, T. A. (2021). Forest community structure and composition following containment treatments for the fungal pathogen oak wilt. *Biological Invasions*, *23*(12), 3733-3747.

Archer, M. E. (1985). Population dynamics of the social wasps Vespula vulgaris and Vespula germanica in England. *The Journal of Animal Ecology*, 473-485.

Archer, M. E., & Halstead, A. (2014). Population dynamics of social wasps (Hymenoptera: Vespidae) in the Royal Horticultural Society's garden at Wisley, Surrey. *Entomologist's Monthly Magazine*, *150*, 19-26.

Baker, A. M., & Potter, D. A. (2020). Invasive paper wasp turns urban pollinator gardens into ecological traps for monarch butterfly larvae. *Scientific Reports*, *10*(1), 9553.

Bastow Wilson, J. (2012). Species presence/absence sometimes represents a plant community as well as species abundances do, or better. *Journal of Vegetation Science*, *23*(6), 1013-1023.

Bollarapu, M. J., & Ramarao, K. V. S. N. (2021). Biodiversity measures-mathematical evaluation of various indices. *Oeconomia Copernican*, *12*(4), 46-59.

Borowski Jr, C. P., Daglio, A., Jacobson, R. S., Loope, K. J., & Walker, C. E. (2021). Interspecific Usurpation in the Downy Yellowjacket, Vespula flavopilosa (Hymenoptera: Vespidae). *Entomological News*, *130*(1), 101-107.

Brewer, L. G., & Vankat, J. L. (2004). Description of vegetation of the Oak Openings of northwestern Ohio at the time of Euro-American settlement.

Brock, R. E., Cini, A., & Sumner, S. (2021). Ecosystem services provided by aculeate wasps. *Biological Reviews*, *96*(4), 1645-1675.

Chapman, R. E., & Bourke, A. F. (2001). The influence of sociality on the conservation biology of social insects. *Ecology Letters*, *4*(6), 650-662.

Clemente, M. A., Guevara, R., Moleiro, H. R., Silveira, O. T., & Giannotti, E. (2020). Community structure and composition of social wasps (Hymenoptera: Vespidae) in different vegetation types in São Paulo, Brazil. *Sociobiology*, *67*(3), 449-461.

Clemente, M. A., Guevara, R., Moleiro, H. R., Silveira, O. T., de Souza, M. M., & Giannotti, E. (2021). Effectiveness of two sampling methods for social wasps in different ecosystems. *EntomoBrasilis*, *14*, e918-e918.

Clemson University. (n.d.). Yellow jackets. Clemson Cooperative Extension. Retrieved June 18, 2025, from https://www.clemson.edu/extension/beekeepers/fact-sheets-publications/yellow-jackets.html

Danziger, A. M., Olson, Z. H., & Frederich, M. (2022). Limitations of eDNA analysis for Carcinus maenas abundance estimations. *BMC Ecology and Evolution*, *22*(1), 14.

Detoni, M. F. D. F. S., Barbosa, B. C., Maciel, T. T., Dos Santos, S. J. L., & Prezoto, F. (2018). Long- and short-term changes in social wasp community structure in an urban area. *Sociobiology*, *65*(2), 305-311.

Didham, R. K., Basset, Y., Collins, C. M., Leather, S. R., Littlewood, N. A., Menz, M. H., ... & Hassall, C. (2020). Interpreting insect declines: seven challenges and a way forward. *Insect Conservation and Diversity*, *13*(2), 103-114.

Eklöf, A., & Ebenman, B. O. (2006). Species loss and secondary extinctions in simple and complex model communities. *Journal of animal ecology*, 239-246.

Elphick, C. S. (2008). How you count counts: the importance of methods research in applied ecology. *Journal of Applied Ecology*, *45*(5), 1313-1320.

Fisher, K., West, M., Lomeli, A. M., Woodard, S. H., & Purcell, J. (2019). Are societies resilient? Challenges faced by social insects in a changing world. *Insectes Sociaux*, *66*, 5-13.

Ferrari, A., & Polidori, C. (2023). Eusociality as a neglected aspect of wild bee α-diversity and its potential impact on diversity estimation. *Insect Conservation and Diversity*, *16*(1), 1-15.

Gaston, K. J., & Fuller, R. A. (2007). Biodiversity and extinction: losing the common and the widespread. *Progress in Physical Geography*, *31*(2), 213-225.

Gotelli, N. J., Ellison, A. M., Dunn, R. R., & Sanders, N. J. (2011). Counting ants (Hymenoptera: Formicidae): biodiversity sampling and statistical analysis for myrmecologists.

Graça, M. B., & Somavilla, A. (2019). Effects of forest fragmentation on community patterns of social wasps (Hymenoptera: Vespidae) in Central Amazon. *Austral Entomology*, *58*(3), 657-665.

Grace927 (2024). Observation of Vespula squamosa [Observation]. iNaturalist https://www.inaturalist.org/observations/213158830

Greenop, A., Woodcock, B. A., Wilby, A., Cook, S. M., & Pywell, R. F. (2018). Functional diversity positively affects prey suppression by invertebrate predators: a meta-analysis. *Ecology*, *99*(8), 1771-1782.

Hallett, A. C., Mitchell, R. J., Chamberlain, E. R., & Karron, J. D. (2017). Pollination success following loss of a frequent pollinator: the role of compensatory visitation by other effective pollinators. *AoB Plants*, *9*(3), plx020.

Herms, D. A., & McCullough, D. G. (2014). Emerald ash borer invasion of North America: history, biology, ecology, impacts, and management. *Annual review of entomology*, *59*(1), 13-30.

Hodgdon, C. T., Mazur, M. D., Friedland, K. D., Willse, N., & Chen, Y. (2021). Consequences of model assumptions when projecting habitat suitability: a caution of forecasting under uncertainties. *ICES Journal of Marine Science*, *78*(6), 2092-2108.

Holt, R. D., Lawton, J. H., Polis, G. A., & Martinez, N. D. (1999). Trophic rank and the species–area relationship. *Ecology*, *80*(5), 1495-1504.

Horskins, K., Mather, P. B., & Wilson, J. C. (2006). Corridors and connectivity: when use and function do not equate. *Landscape ecology*, *21*, 641-655.

Hurlbert, S. H. (2009). The ancient black art and transdisciplinary extent of pseudoreplication. *Journal of Comparative Psychology*, *123*(4), 434.

Jacques, G. C., Pires, E., Hermes, M. G., Faria, L. D. B., Souza, M. M., & Silveira, L. C. P. (2018). Evaluating the efficiency of different sampling methods to survey social wasps (Vespidae: Polistinae) in an anthropized environment. *Sociobiology*, *65*(3), 515-523.

Jamwright (2024). Observation of Vespula germanica [Observation]. iNaturalist https://www.inaturalist.org/observations/213749126

Jmfarley (2024). Observation of Polistes dominula [Observation]. iNaturalist https://www.inaturalist.org/observations/242894559

Johnson, S. A., Ober, H. K., & Adams, D. C. (2017). Are keystone species effective umbrellas for habitat conservation? A spatially explicit approach. *Journal for Nature Conservation*, *37*, 47-55.

Kimsmith (2019). Observation of Vespula squamosa [Observation]. iNaturalist https://www.inaturalist.org/observations/34835536

Kimsmith (2020). Observation of Vespula squamosa [Observation]. iNaturalist https://www.inaturalist.org/observations/47571400

Kirk, H., Soanes, K., Amati, M., Bekessy, S., Harrison, L., Parris, K., ... & Threlfall, C. (2023). Ecological connectivity as a planning tool for the conservation of wildlife in cities. *MethodsX*, *10*, 101989.

Komonen, A., Sorvari, J., & Torniainen, J. (2025, January). Population dynamics of the social wasps, Vespula vulgaris and Dolichovespula media (Vespidae), in Finland. In *Annales Zoologici Fennici* (Vol. 62, No. 1, pp. 1-11). Finnish Zoological and Botanical Publishing Board.

Kruess, A., & Tscharntke, T. (1994). Habitat fragmentation, species loss, and biological control. *Science*, *264*(5165), 1581-1584.

Lester, P. J., Haywood, J., Archer, M. E., & Shortall, C. R. (2017). The long-term population dynamics of common wasps in their native and invaded range. *Journal of Animal Ecology*, *86*(2), 337-347.

Liao, J., Bearup, D., & Blasius, B. (2017). Diverse responses of species to landscape fragmentation in a simple food chain. *Journal of Animal Ecology*, *86*(5), 1169-1178.

Martin, A. K., & Root, K. V. (2020). Examining land use changes to evaluate the effects of land management in a complex, dynamic landscape. *Environmental Management*, *66*(3), 333-347.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica*, *23*(2), 143-149.

mct157106 (2024). Observation of Vespula squamosa [Observation]. iNaturalist https://www.inaturalist.org/observations/249439765

Oten, K. L., Day, E., Dellinger, T., Disque, H. H., Barringer, L. E., Cancelliere, J., ... & Bertone, M. A. (2023). First records of elm zigzag sawfly (Hymenoptera: Argidae) in the United States. *Journal of Integrated Pest Management*, *14*(1), 12.

Pamilo, P., & Crozier, R. H. (1997). Population biology of social insect conservation. *Memoirs of the Museum of Victoria*, *56*(2), 411-419.

Pawlikowski, T., Olszewski, P., Piekarska-Boniecka, H., & Pawlikowski, K. (2016). Diversity of social wasp communities (Hymenoptera: Polistinae and Vespinae) in the agricultural landscape of Central Poland. *Acta zool. bulg*, *68*, 553-556.

Reid, B. L., MacDonald, J. F., & Ross, D. R. (1994). Foraging and spatial dispersion in protein-scavenging workers of Vespula germanica and V. maculifrons (Hymenoptera: Vespidae). *Journal of Insect Behavior*, *8*, 315-330.

Reynolds, A. M. (2009). Lévy flight patterns are predicted to be an emergent property of a bumblebees' foraging strategy. *Behavioral Ecology and Sociobiology*, *64*, 19-23.

Reynolds, A. M., Swain, J. L., Smith, A. D., Martin, A. P., & Osborne, J. L. (2009). Honeybees use a Lévy flight search strategy and odour-mediated anemotaxis to relocate food sources. *Behavioral Ecology and Sociobiology*, *64*, 115-123.

Richmond, C. E., Breitburg, D. L., & Rose, K. A. (2005). The role of environmental generalist species in ecosystem function. *Ecological modelling*, *188*(2-4), 279-295.

Richter, M. R. (2000). Social wasp (Hymenoptera: Vespidae) foraging behavior. *Annual review of entomology*, *45*(1), 121-150.

Rickbarricklow (2023). Observation of Vespula squamosa [Observation]. iNaturalist https://www.inaturalist.org/observations/91637961

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, *13*(2), 255-266.

Rose, G. A., & Kulka, D. W. (1999). Hyperaggregation of fish and fisheries: how catch-per-unit-effort increased as the northern cod (Gadus morhua) declined. *Canadian Journal of Fisheries and Aquatic Sciences*, *56*(S1), 118-127.

Rosenberg, D. K., & McKelvey, K. S. (1999). Estimation of habitat selection for central-place foraging animals. *The Journal of Wildlife Management*, 1028-1038.

Rubene, D., Schroeder, M., Ranius, T., & Didham, R. (2015). Estimating bee and wasp (Hymenoptera: Aculeata) diversity on clear-cuts in forest landscapes-an evaluation of sampling methods. *Insect Conservation & Diversity*, *8*(3).

Santelices, B. (1999). How many kinds of individual are there?. *Trends in ecology & evolution*, *14*(4), 152-155.

Schetter, T. A., & Root, K. V. (2011). Assessing an imperiled oak savanna landscape in northwestern Ohio using Landsat data. *Natural Areas Journal*, *31*(2), 118-130.

Schueller, T. I., Nordheim, E. V., Taylor, B. J., & Jeanne, R. L. (2010). The cues have it; nest-based, cue-mediated recruitment to carbohydrate resources in a swarm-founding social wasp. *Naturwissenschaften*, *97*, 1017-1022.

Schüepp, C., Rittiner, S., & Entling, M. H. (2012). High bee and wasp diversity in a heterogeneous tropical farming system compared to protected forest. *PLoS One*, *7*(12), e52109.

Schwarz, J., Habel, J. C., & Eberle, J. (2024). Effects of habitat biotic features on hymenopteran diversity in East Africa. *Journal of Insect Conservation*, *28*(4), 821-830.

Siderhurst, M. S., Fairbanks, K. E., Ladizinsky, N., Snyder, J., & Hurst, A. L. (2025). Tracking 3 wasp species (Hymenoptera: Vespidae) with harmonic radar: toward an accessible, inexpensive colony location tool. *Journal of Insect Science*, *25*(2), 19.

Smith-Ferguson, J., & Beekman, M. (2019). Can't see the colony for the bees: behavioural perspectives of biological individuality. *Biological Reviews*, *94*(6), 1935-1946.

Southon, R. J., Fernandes, O. A., Nascimento, F. S., & Sumner, S. (2019). Social wasps are effective biocontrol agents of key lepidopteran crop pests. *Proceedings of the Royal Society B*, *286*(1914), 20191676.

Steiner, F. M., Schlick-Steiner, B. C., Moder, K., Bruckner, A., & Christian, E. (2005). Congruence of data from different trapping periods of ant pitfall catches (Hymenoptera: Formicidae). *Sociobiology*, *46*(1), 105-116.

Sumner, S., Law, G., & Cini, A. (2018). Why we love bees and hate wasps. *Ecological Entomology*, *43*(6), 836-845.

Taylor, B. J., & Jeanne, R. L. (2018). Gastral drumming: a nest-based food-recruitment signal in a social wasp. *The Science of Nature*, *105*(3), 23.

van Nouhuys, S. (2005, January). Effects of habitat fragmentation at different trophic levels in insect communities. In *Annales Zoologici Fennici* (pp. 433-447). Finnish Zoological and Botanical Publishing Board.

Wilms, R., Mäthner, E., Winnen, L., & Lanwehr, R. (2021). *Omitted variable bias: A threat to estimating causal relationships. Methods in Psychology, 5, 100075*.

Wilson, D. S., & Sober, E. (1989). Reviving the superorganism. *Journal of theoretical Biology*, *136*(3), 337-356.

Wilson, R. S., Leonhardt, S. D., Burwell, C. J., Fuller, C., Smith, T. J., Kaluza, B. F., & Wallace, H. M. (2020). Landscape simplification modifies trap-nesting bee and wasp communities in the subtropics. *Insects*, *11*(12), 853.

Wilson-Rankin, E. E. (2014). Social context influences cue-mediated recruitment in an invasive social wasp. *Behavioral Ecology and Sociobiology*, *68*, 1151-1161.

Zografou, K., Swartz, M. T., Tilden, V. P., McKinney, E. N., Eckenrode, J. A., & Sewall, B. J. (2020). Stable generalist species anchor a dynamic pollination network. *Ecosphere*, *11*(8), e03225.

APPENDIX A: DISTRIBUTIONS OF ASSOCIATED HABITAT VARIABLES AT SPECIMEN CAPTURE LOCATION ACROSS ALL SAMPLING LOCATIONS IN THE REGION

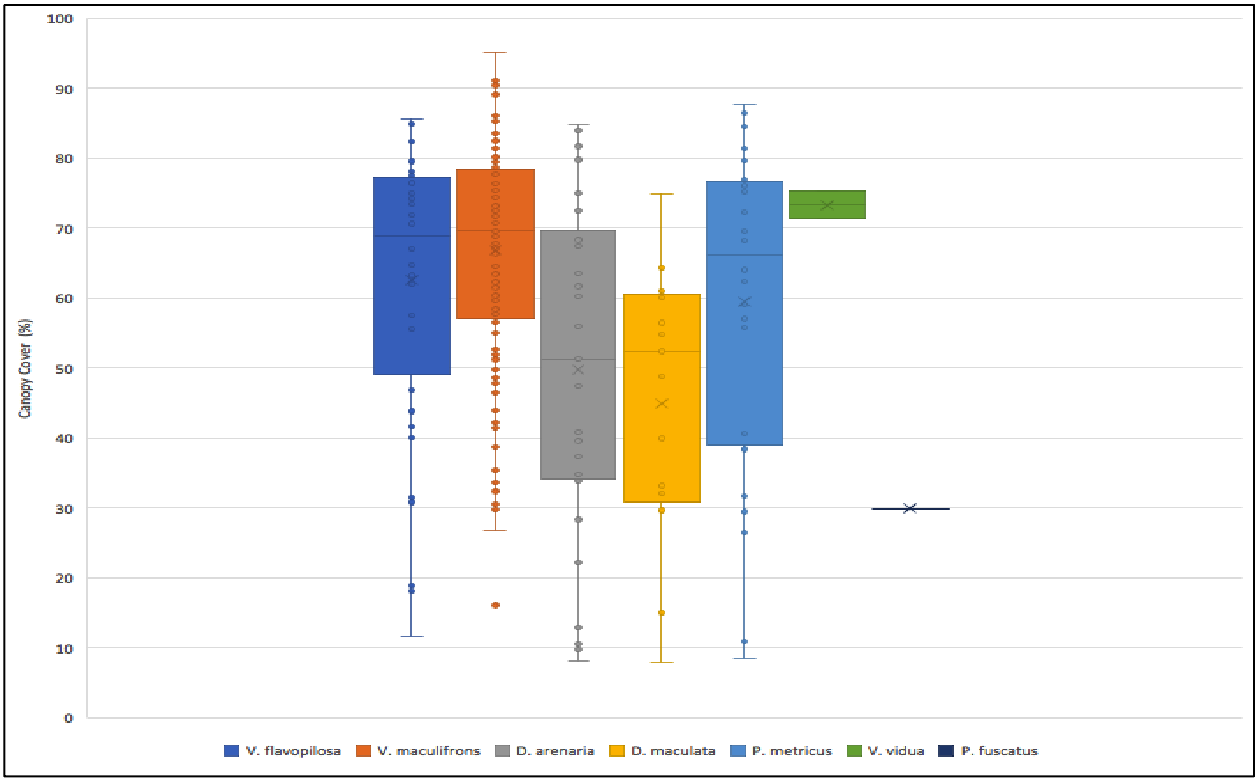


Figure A.1. Distribution of canopy cover by species. Readings were collected at specimen capture locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The distributions are summarized in box plots. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

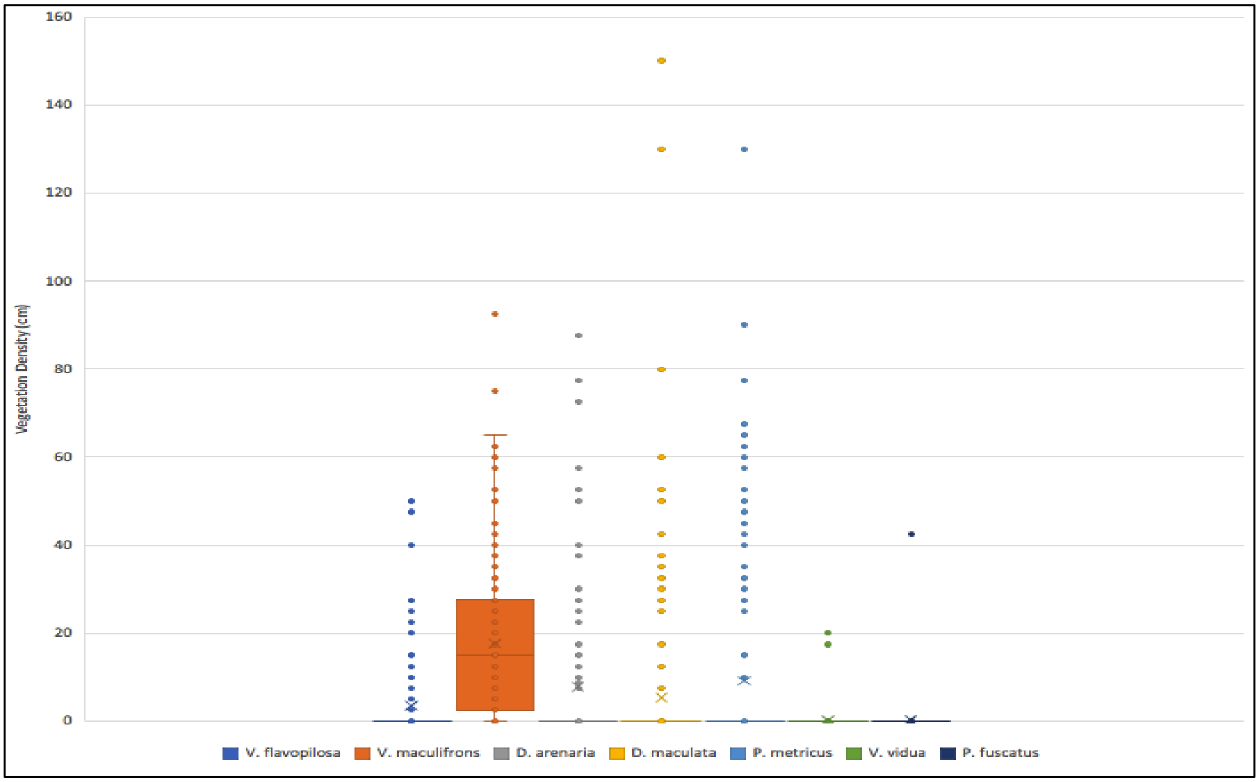


Figure A.2. Distribution of vegetation density by species. Measures were taken using a Robel Pole. Each band was spaced 10cm apart, and each band represented one count. Readings were collected at specimen capture locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The distributions are summarized in box plots. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

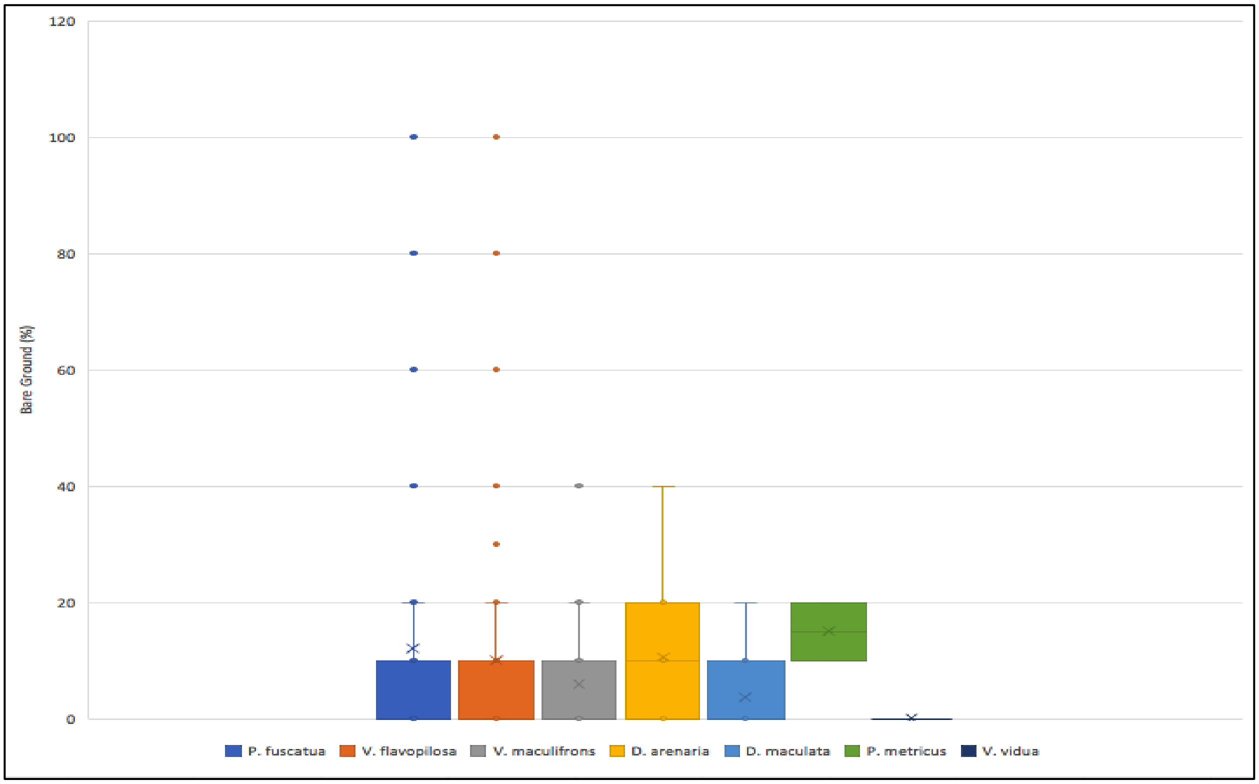


Figure A.3. Distributions of the percentage of bare ground contained in the 1x1 meter quadrat taken at specimen collection locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The distributions are summarized in box plots. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

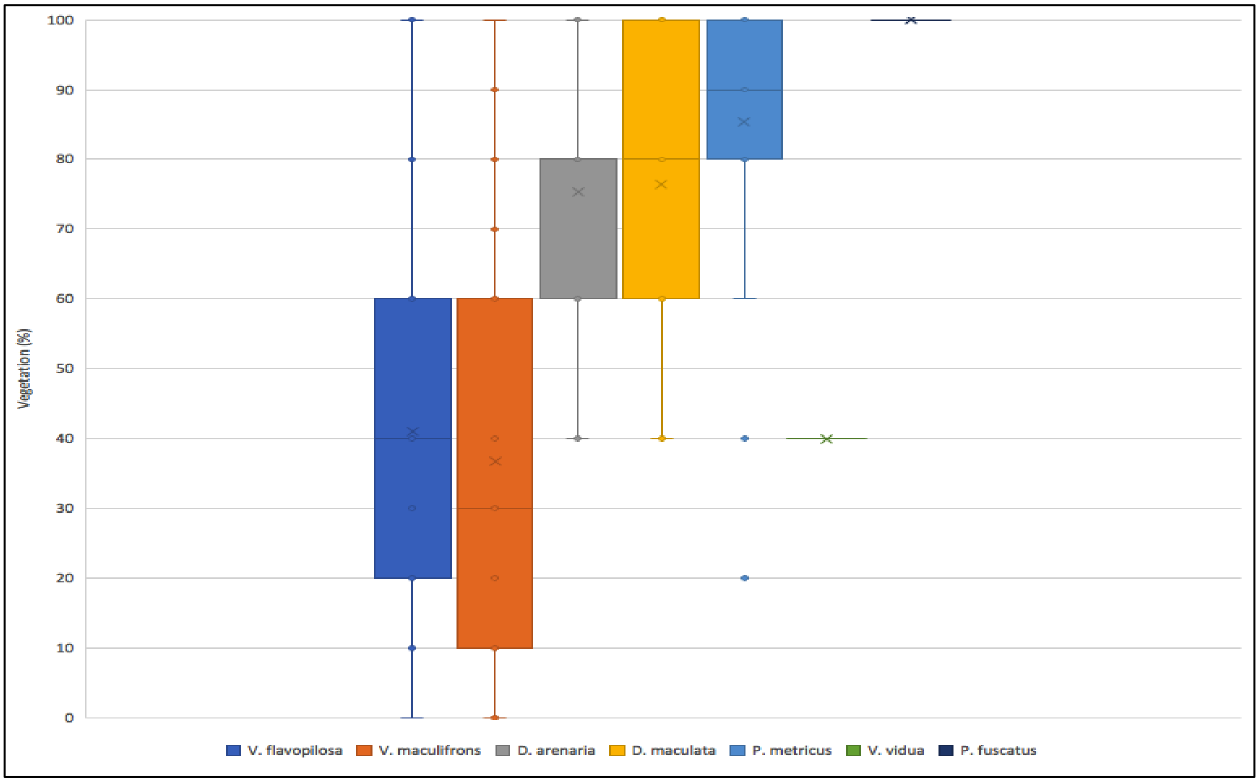


Figure A.4. Distributions of the percentage of vegetation contained in the 1x1 meter quadrat taken at specimen collection locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The distributions are summarized in box plots. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

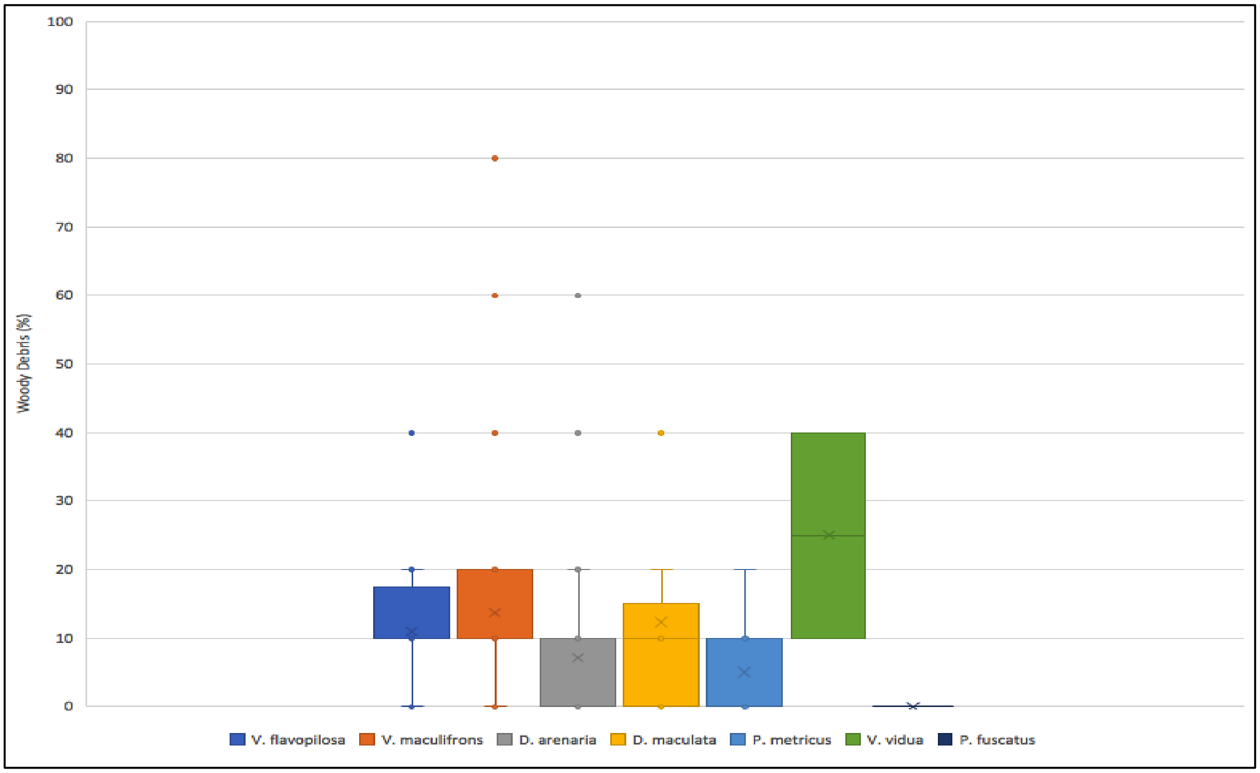


Figure A.5. Distributions of the percentage of woody debris contained in the 1x1 meter quadrat taken at specimen collection locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

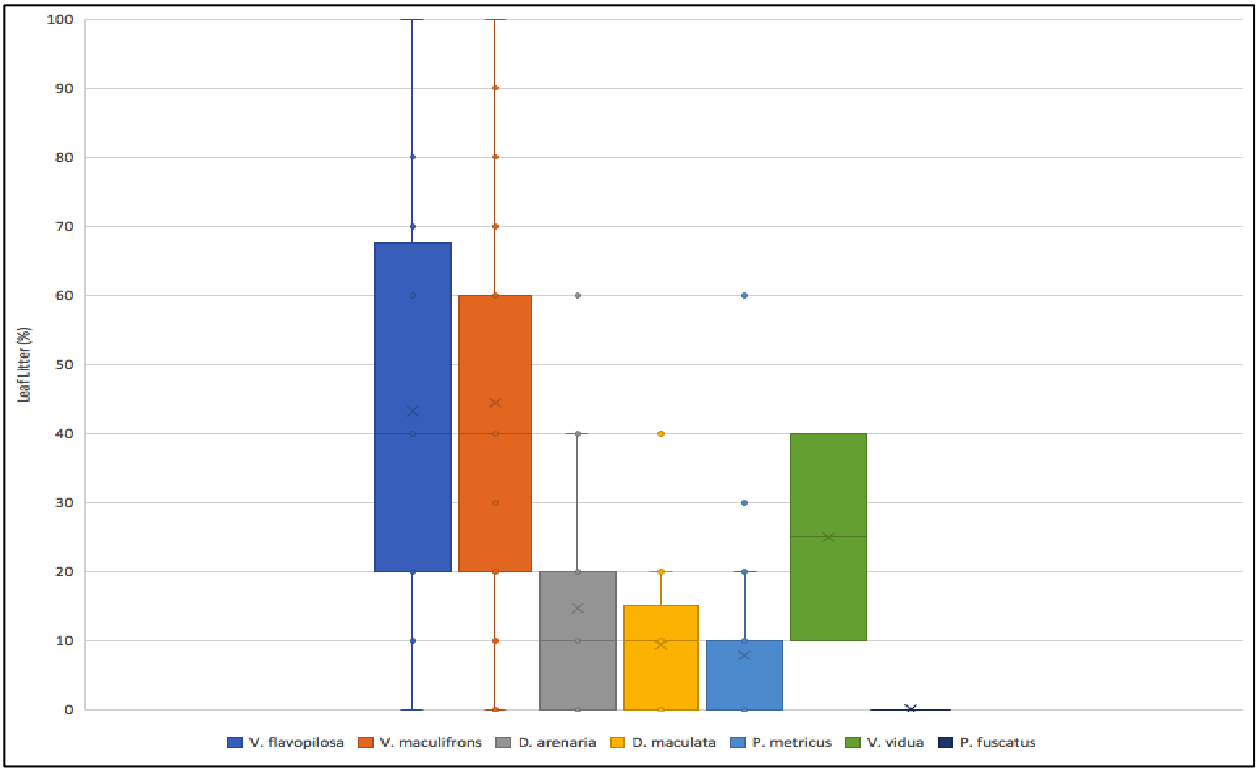


Figure A.6. Distributions of the percentage of leaf litter contained in the 1x1 meter quadrat taken at specimen collection locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

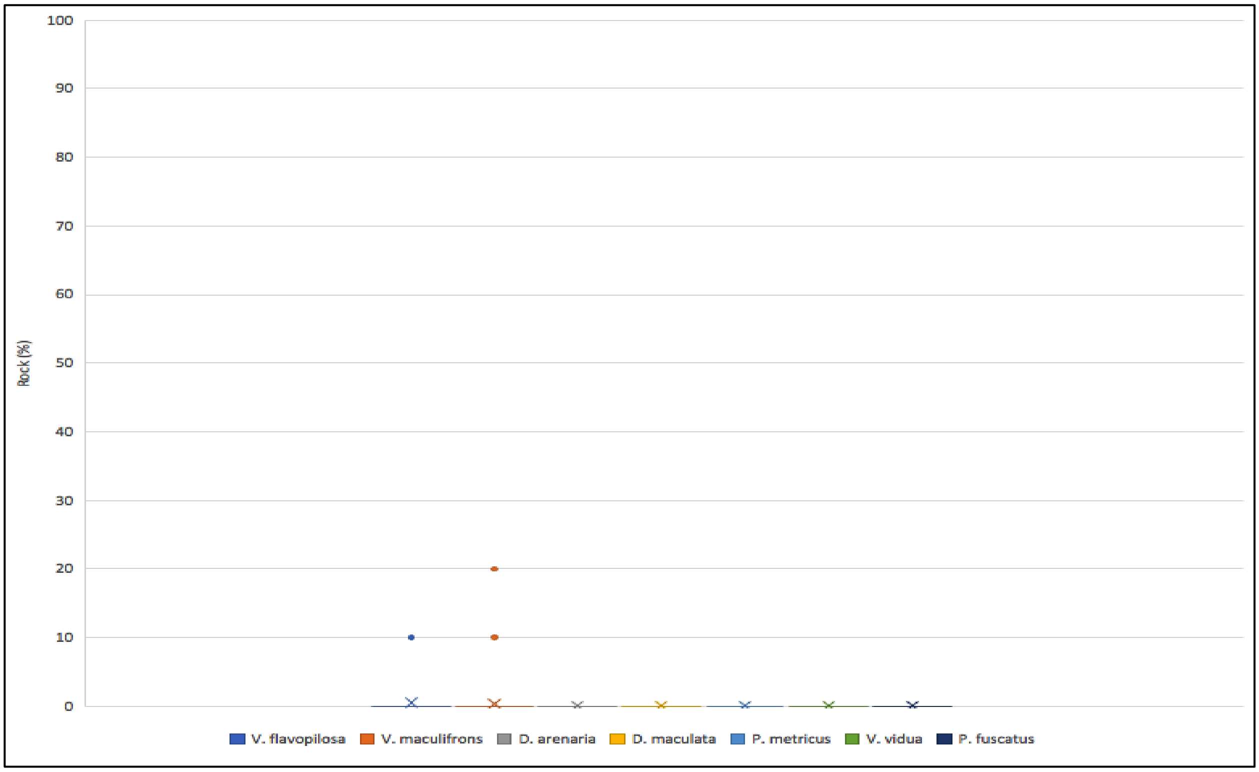


Figure A.7. Distributions of the percentage of rock contained in the 1x1 meter quadrat taken at specimen collection locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

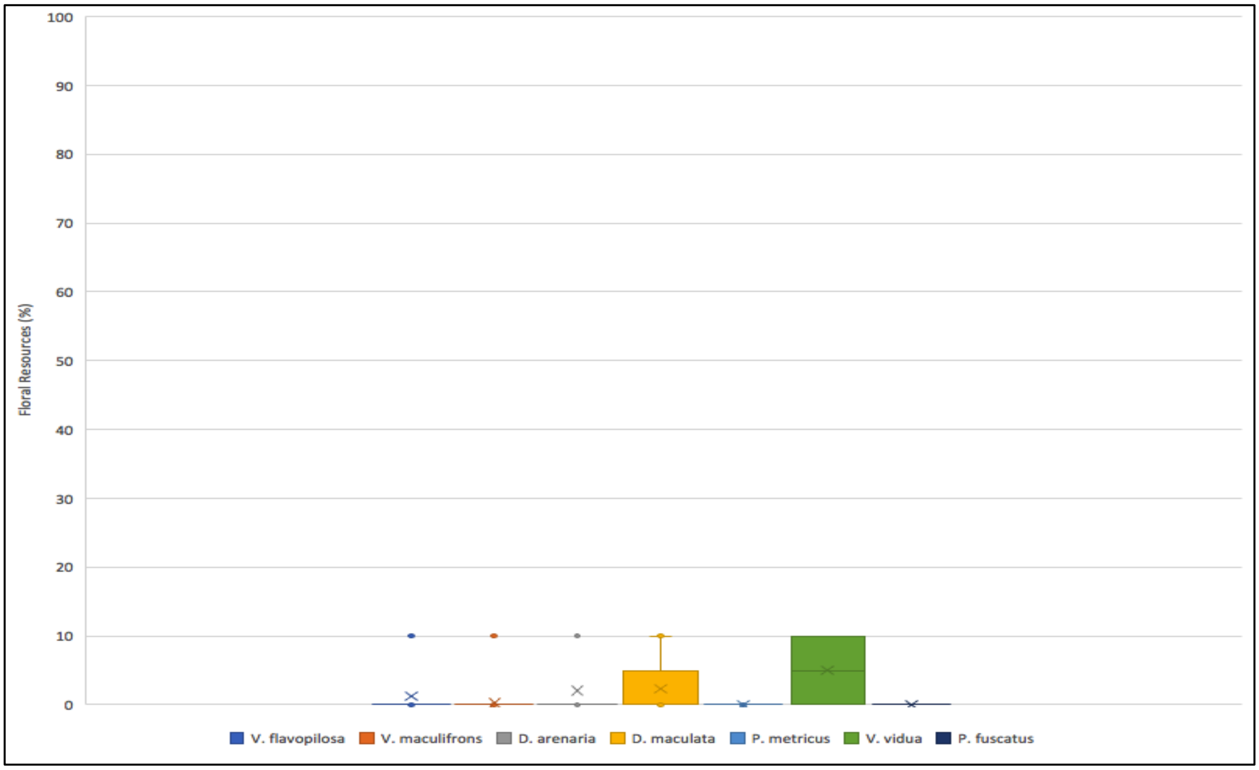


Figure A.8. Distributions of the percentage of floral resources contained in the 1x1 meter quadrat taken at specimen collection locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

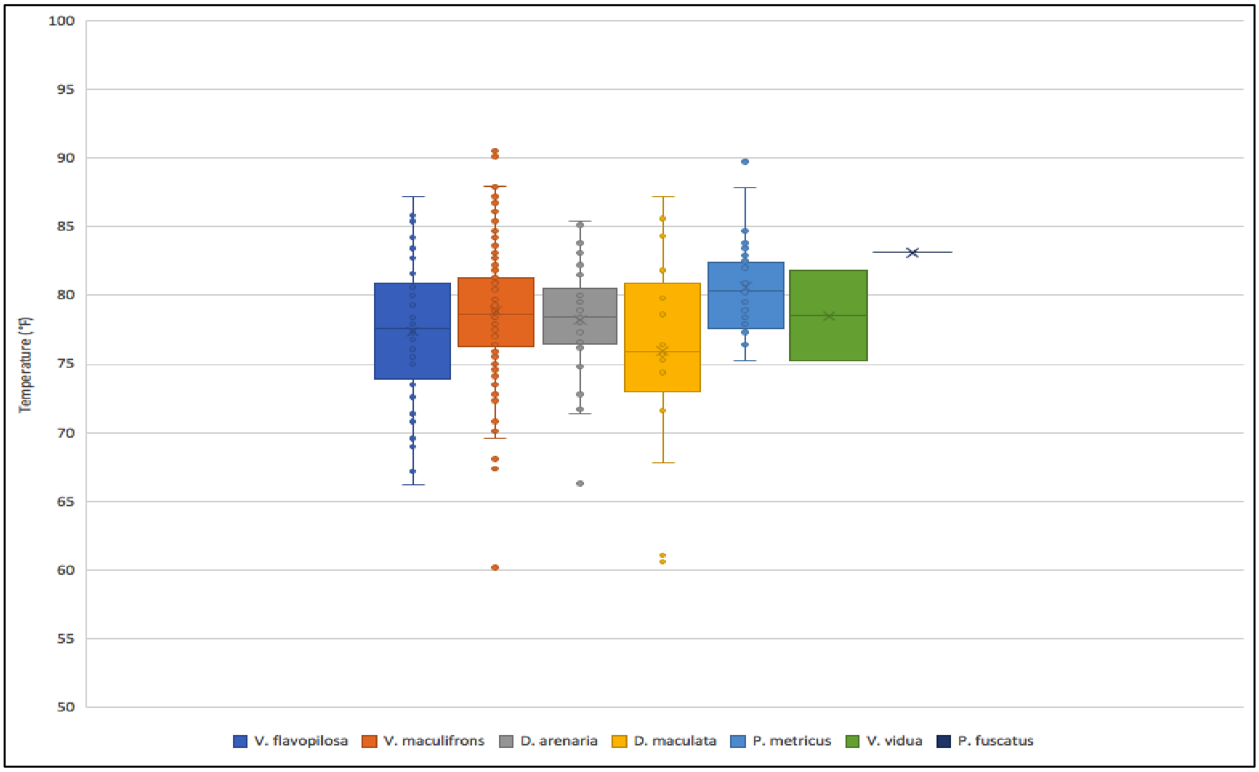


Figure A.9. Distributions of the temperature in degrees Fahrenheit recorded at specimen collection locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

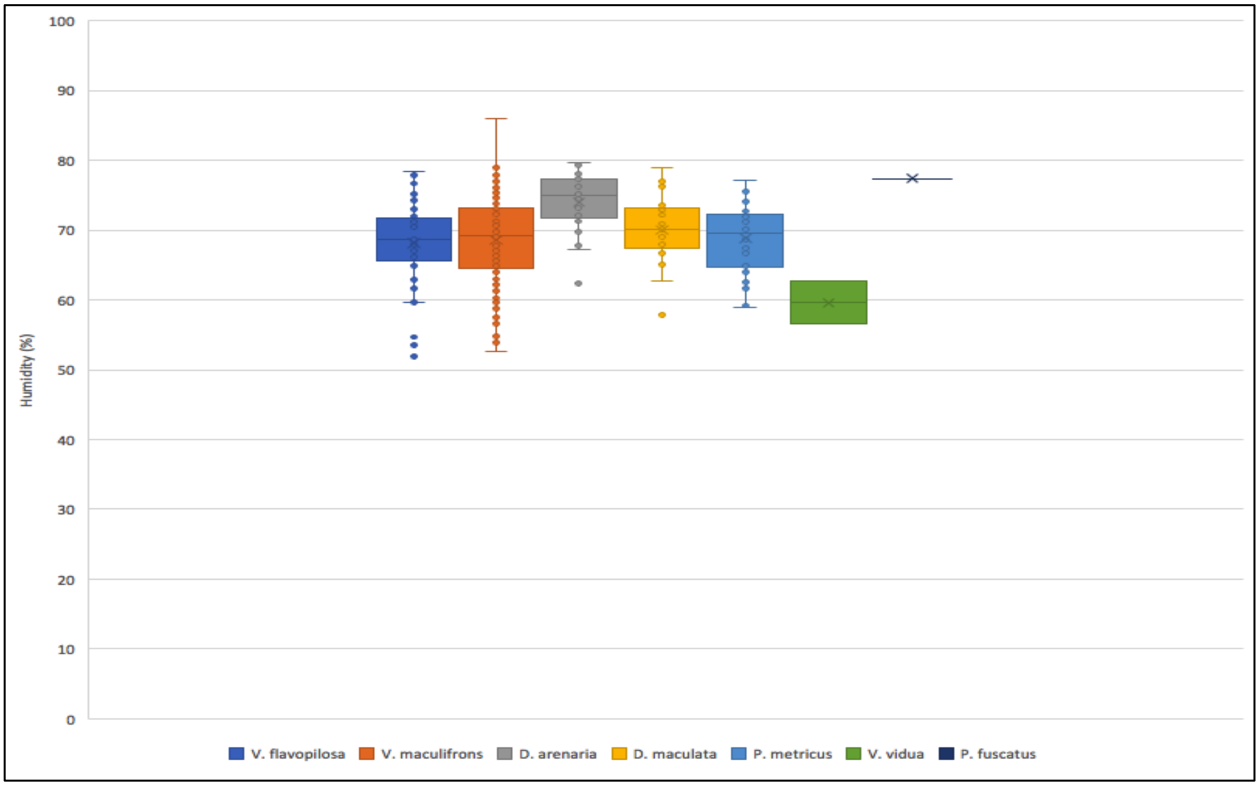


Figure A.10. Distributions of the relative humidity (percent) recorded at specimen collection locations. Values at points of collection are aggregated for species within the genera *Vespula*, *Dolichovespula*, and *Polistes* across all sample locations. The box represents the distance between the 25th percentile and the 75th percentile of each species data with the "x" denoting the samples average value. Whiskers extend from the distribution until reaching either 1.5 times the interquartile range or the maximum or minimum value. Points falling outside of the box plot represent outlier data points.

APPENDIX B: OUTPUT GRAPHS OF CLUSTERING ANALYSIS USING RIPLEY'S K-FUNCTION FOR VESPULA MACULIFRONS SPECIMENS AT EACH SAMPLE LOCATION

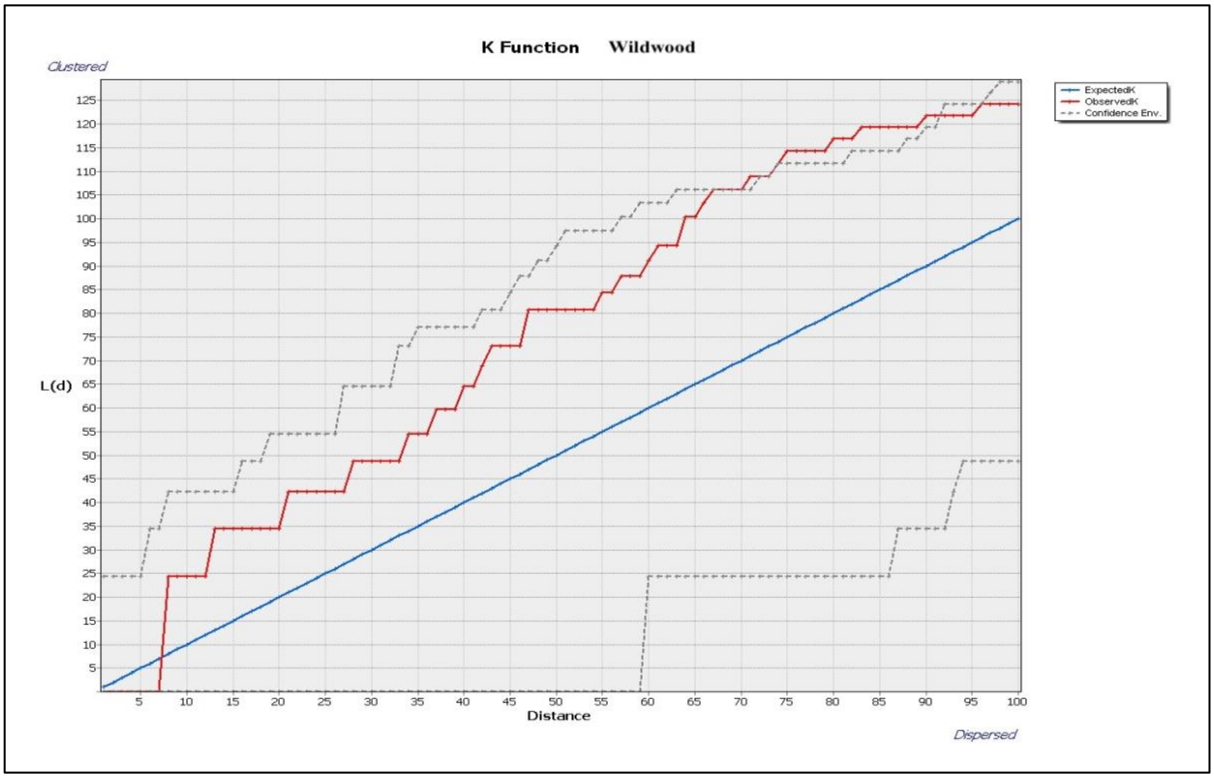


Figure B.1. Ripley's K-function (L-function transformation) showing significant clustering of observed values (red) at the Wildwood Preserve Metropark location. Observations exceed both the expectation of a uniform distribution (blue) and the confidence envelope (grey) created using 999 permutations with distance bands increasing from zero by one meter each iteration.

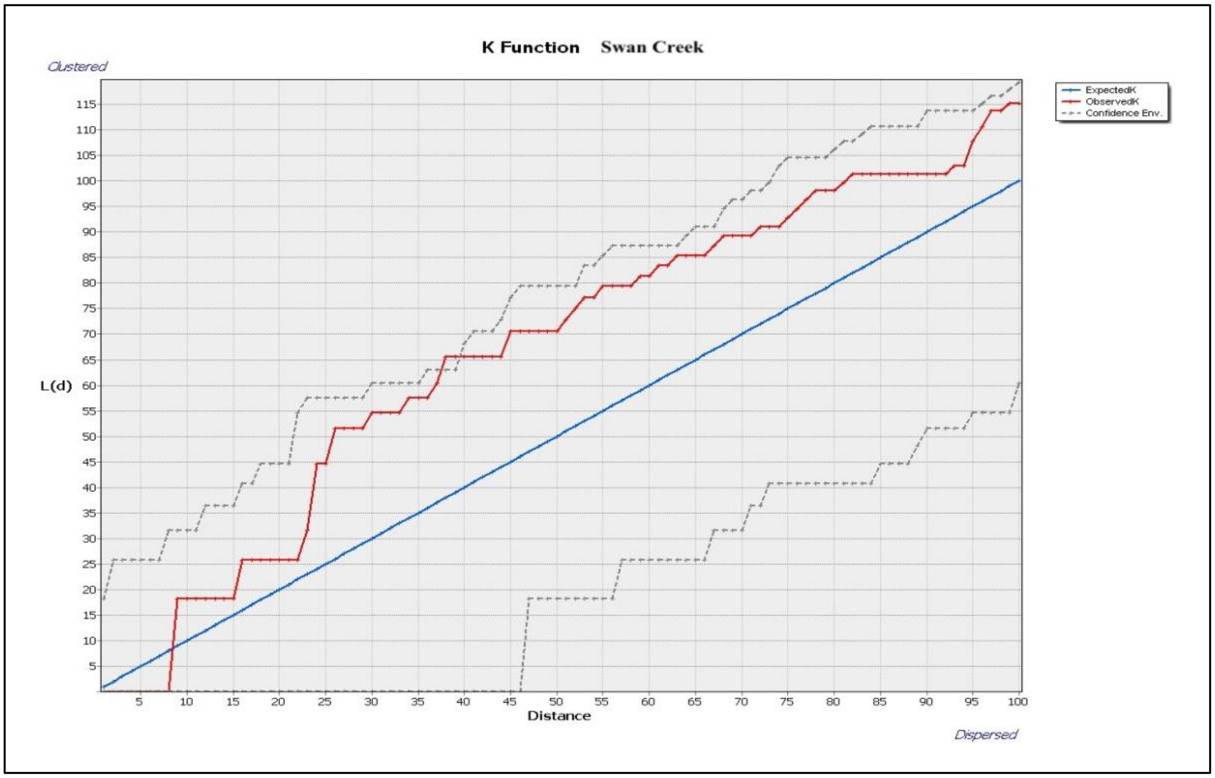


Figure B.2. Ripley's K-function (L-function transformation) showing significant clustering of observed values (red) at the Swan Creek Preserve Metropark location. Observations exceed both the expectation of a uniform distribution (blue) and the confidence envelope (grey) created using 999 permutations with distance bands increasing from zero by one meter each iteration.

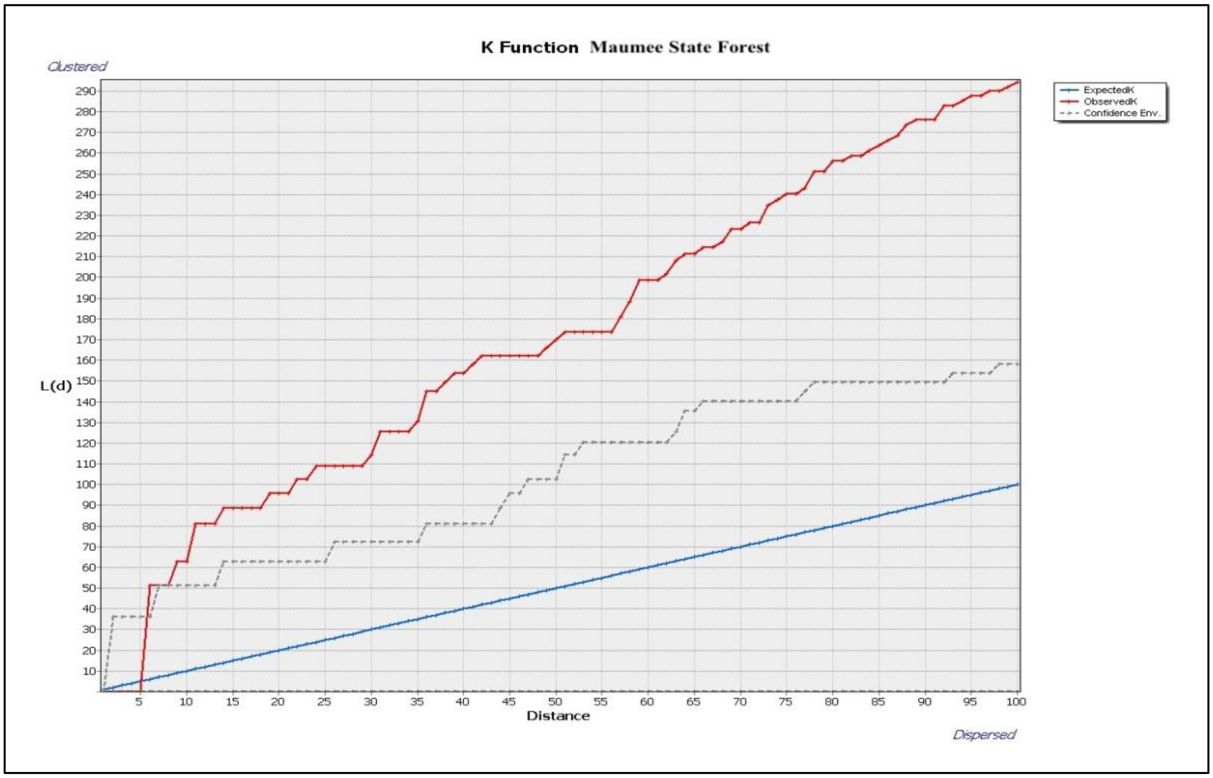


Figure B.3. Ripley's K-function (L-function transformation) showing significant clustering of observed values (red) at the Maumee State Forest location. Observations exceed both the expectation of a uniform distribution (blue) and the confidence envelope (grey) created using 999 permutations with distance bands increasing from zero by one meter each iteration.

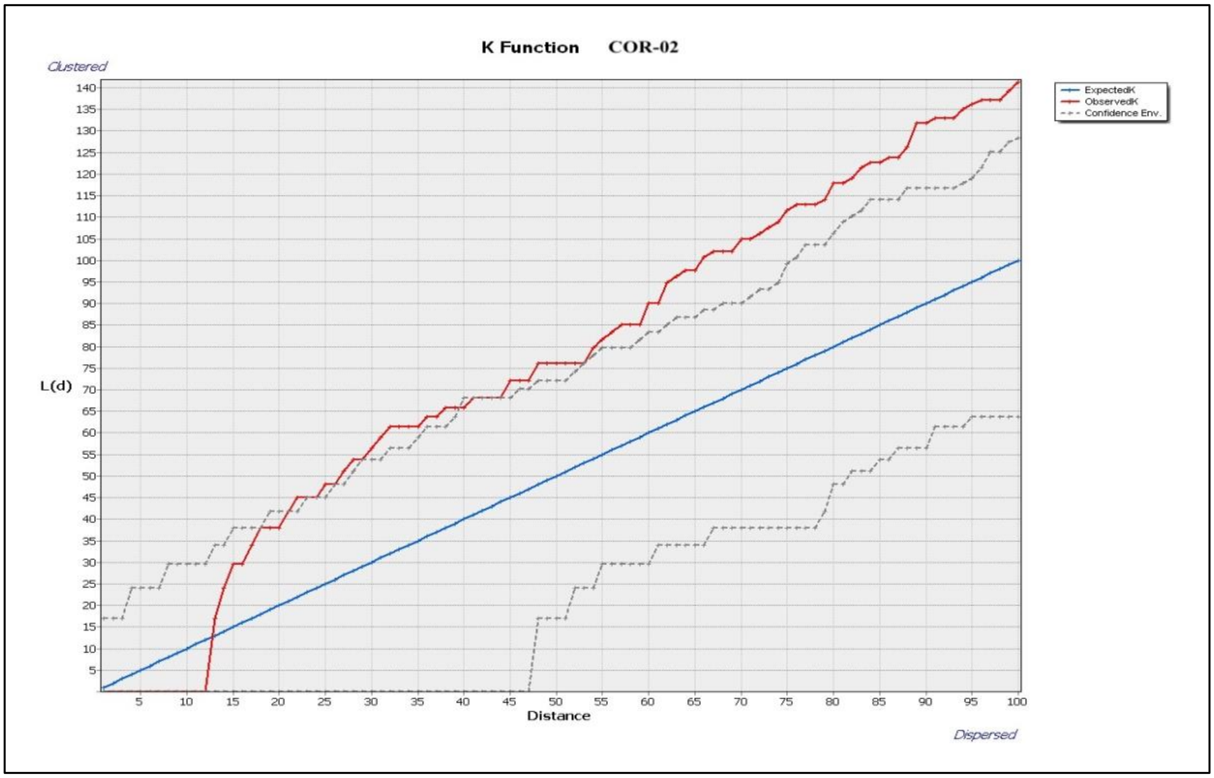


Figure B.4. Ripley's K-function (L-function transformation) showing significant clustering of observed values (red) at the COR-02 location. Observations exceed both the expectation of a uniform distribution (blue) and the confidence envelope (grey) created using 999 permutations with distance bands increasing from zero by one meter each iteration.

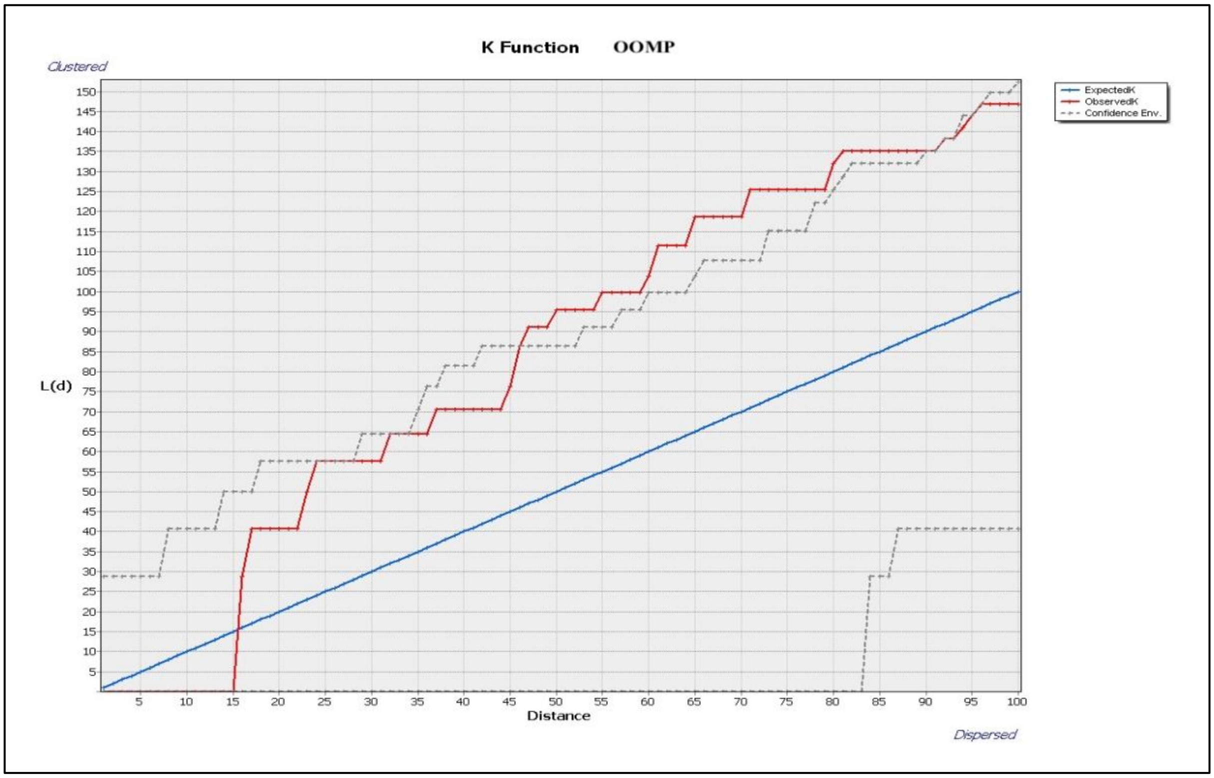


Figure B.5. Ripley's K-function (L-function transformation) showing significant clustering of observed values (red) at the Oak Openings Preserve Metropark location. Observations exceed both the expectation of a uniform distribution (blue) and the confidence envelope (grey) created using 999 permutations with distance bands increasing from zero by one meter each iteration.

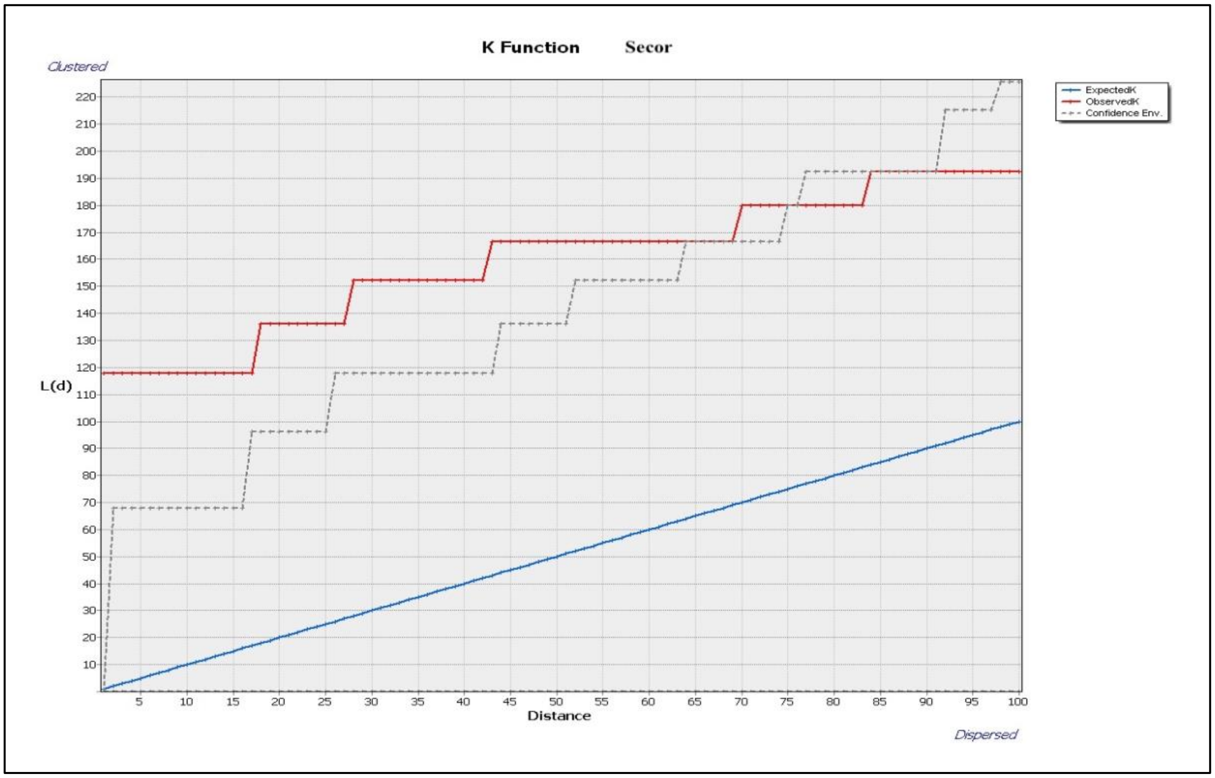


Figure B.6. Ripley's K-function (L-function transformation) showing significant clustering of observed values (red) at the Secor Metropark location. Observations exceed both the expectation of a uniform distribution (blue) and the confidence envelope (grey) created using 999 permutations with distance bands increasing from zero by one meter each iteration.

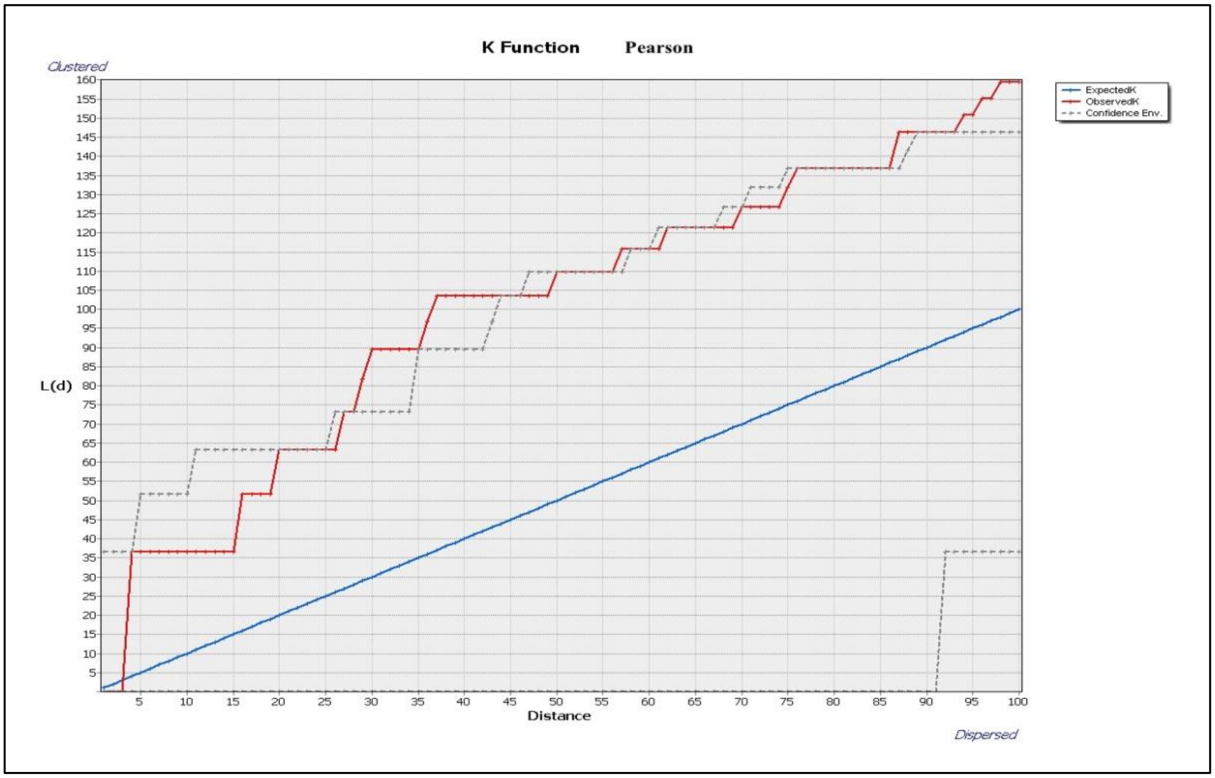


Figure B.7. Ripley's K-function (L-function transformation showing significant clustering of observed values (red) at the Pearson Metropark location. Observations exceed both the expectation of a uniform distribution (blue) and the confidence envelope (grey) created using 999 permutations with distance bands increasing from zero by one meter each iteration.

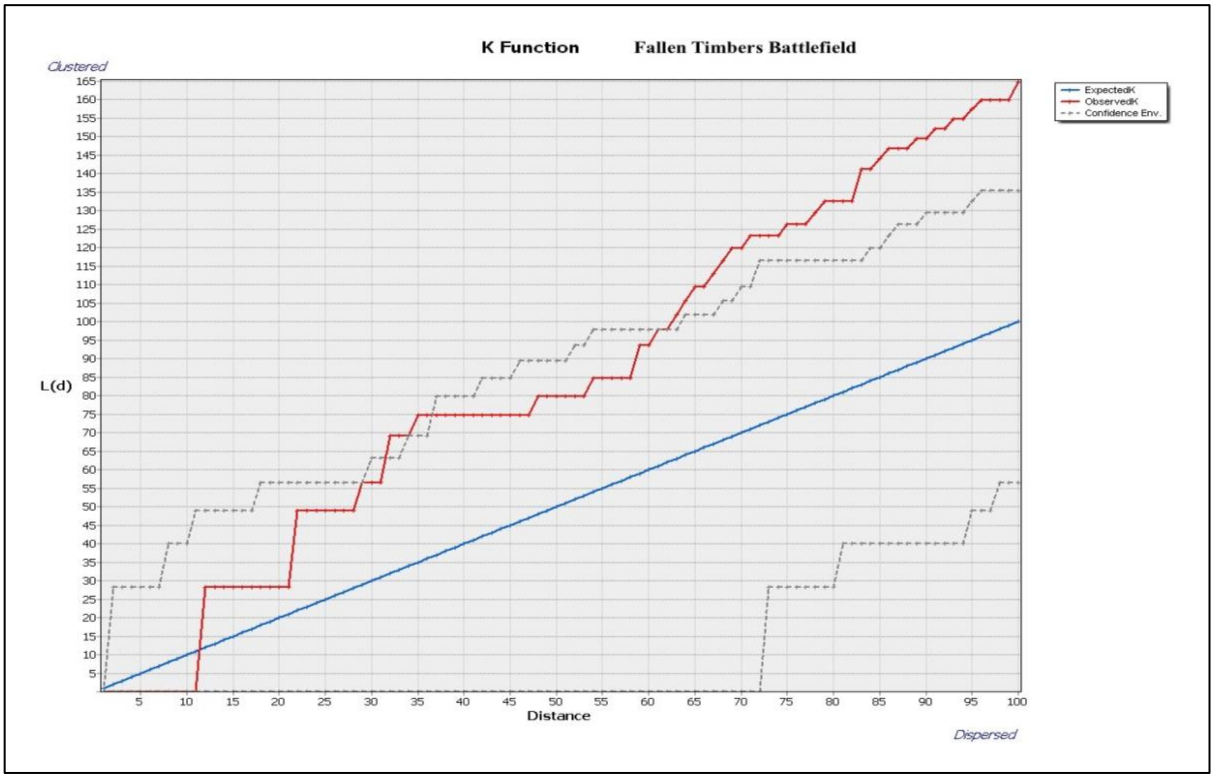


Figure B.8. Ripley's K-function (L-function transformation) showing significant clustering of observed values (red) at the Fallen Timbers Battlefield location. Observations exceed both the expectation of a uniform distribution (blue) and the confidence envelope (grey) created using 999 permutations with distance bands increasing from zero by one meter each iteration.

APPENDIX C: RANGE OF BRAY-CURTIS DISSIMILARITY VALUES PRODUCED UNDER COLONY UNCERTAINTY FOR EACH SAMPLE LOCATION

Figure C.1. Bray-Curtis dissimilarity outputs for COR-02 compared to the other seven locations in the region. Lines denote the output values for dissimilarity highlighting maximum dissimilarity obtained for a given set of communities (red), minimum dissimilarity obtained for a given set of communities (blue), dissimilarity using the unadjusted abundance data for a given set of communities (purple), and dissimilarity of sites using presence-absence alone (green). Site location names corresponding to the figure's acronyms can be found in the materials and methods section.

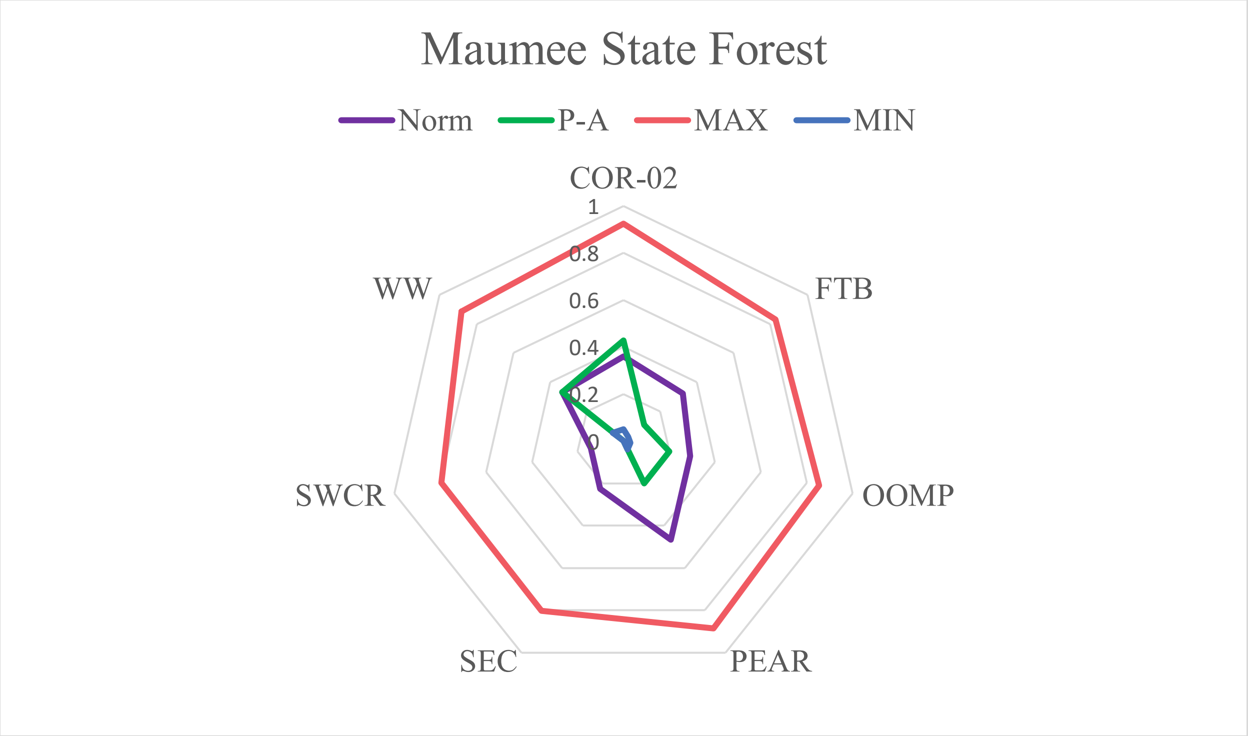


Figure C.2. Bray-Curtis dissimilarity outputs for Maumee State Forest compared to the other seven locations in the region. Lines denote the output values for dissimilarity highlighting maximum dissimilarity obtained for a given set of communities (red), minimum dissimilarity obtained for a given set of communities (blue), dissimilarity using the unadjusted abundance data for a given set of communities (purple), and dissimilarity of sites using presence-absence alone (green). Site location names corresponding to the figure’s acronyms can be found in the materials and methods section.

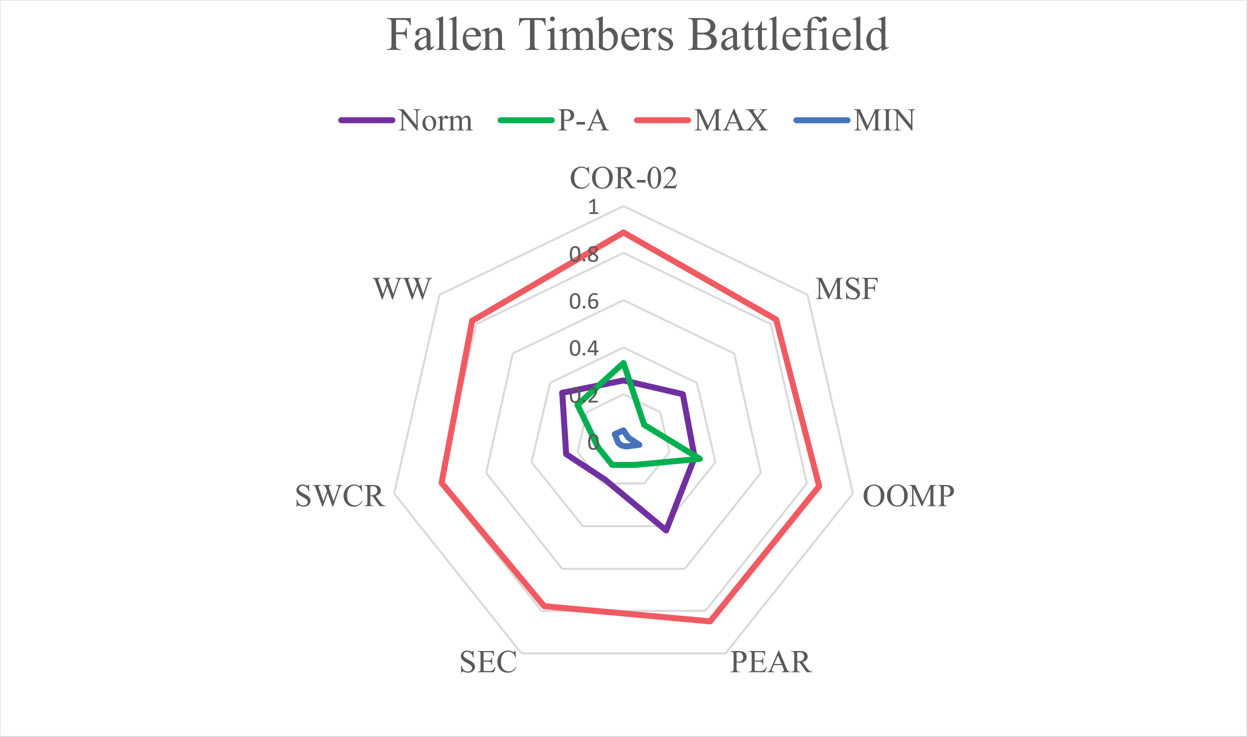


Figure C.3. Bray-Curtis dissimilarity outputs for Fallen Timbers Battlefield compared to the other seven locations in the region. Lines denote the output values for dissimilarity highlighting maximum dissimilarity obtained for a given set of communities (red), minimum dissimilarity obtained for a given set of communities (blue), dissimilarity using the unadjusted abundance data for a given set of communities (purple), and dissimilarity of sites using presence-absence alone (green). Site location names corresponding to the figure's acronyms can be found in the materials and methods section.

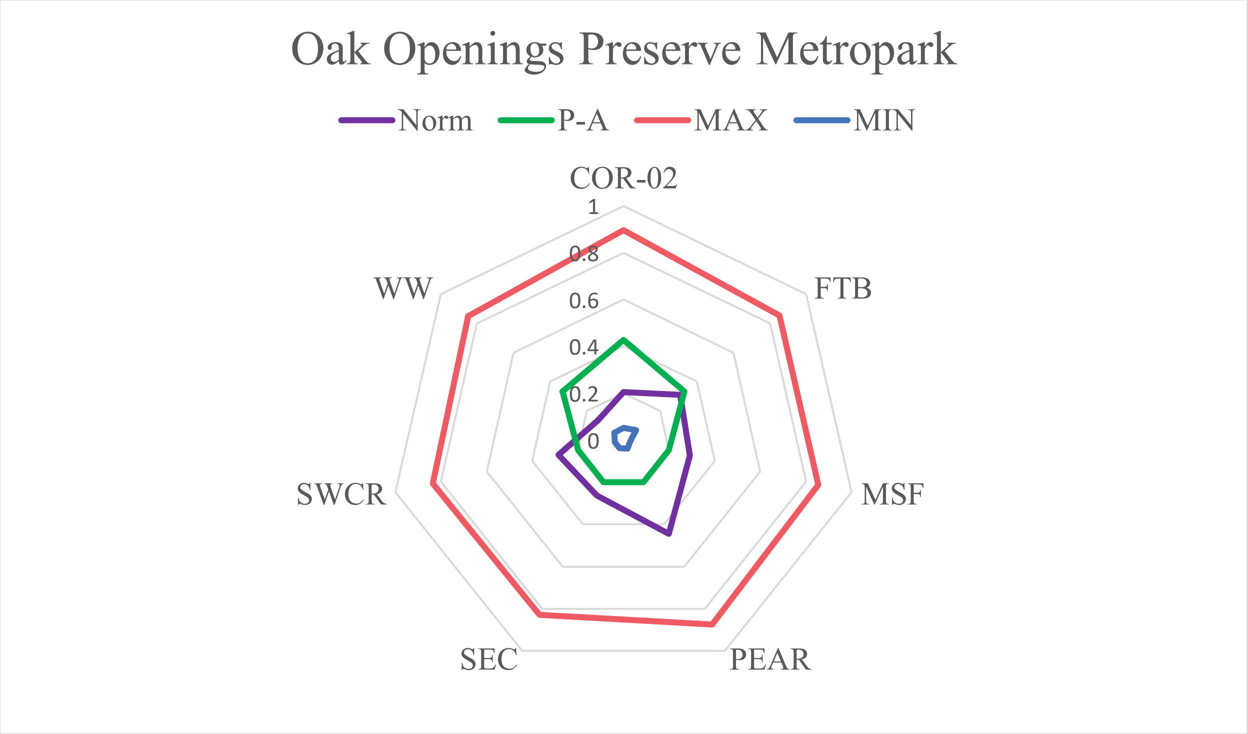


Figure C.4. Bray-Curtis dissimilarity outputs for Oak Openings Preserve Metropark compared to the other seven locations in the region. Lines denote the output values for dissimilarity highlighting maximum dissimilarity obtained for a given set of communities (red), minimum dissimilarity obtained for a given set of communities (blue), dissimilarity using the unadjusted abundance data for a given set of communities (purple), and dissimilarity of sites using presence-absence alone (green). Site location names corresponding to the figure's acronyms can be found in the materials and methods section.

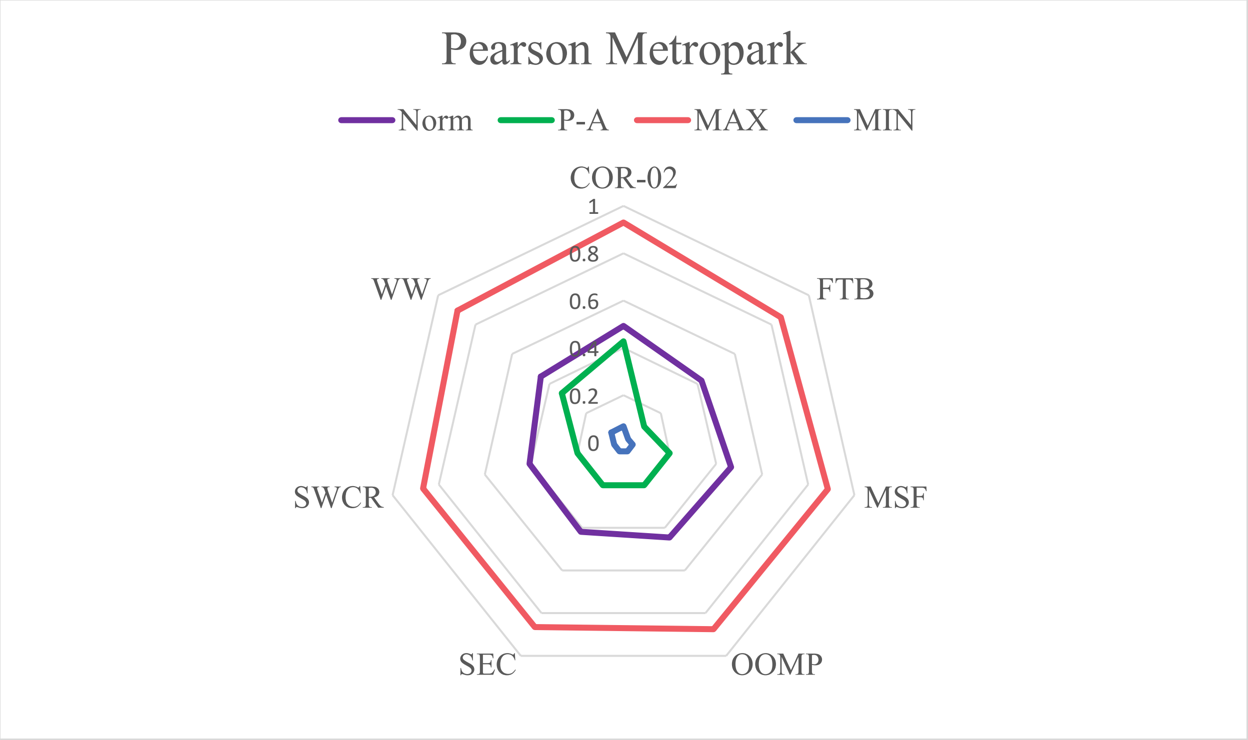


Figure C.5. Bray-Curtis dissimilarity outputs for Pearson Metropark compared to the other seven locations in the region. Lines denote the output values for dissimilarity highlighting maximum dissimilarity obtained for a given set of communities (red), minimum dissimilarity obtained for a given set of communities (blue), dissimilarity using the unadjusted abundance data for a given set of communities (purple), and dissimilarity of sites using presence-absence alone (green). Site location names corresponding to the figure's acronyms can be found in the materials and methods section.

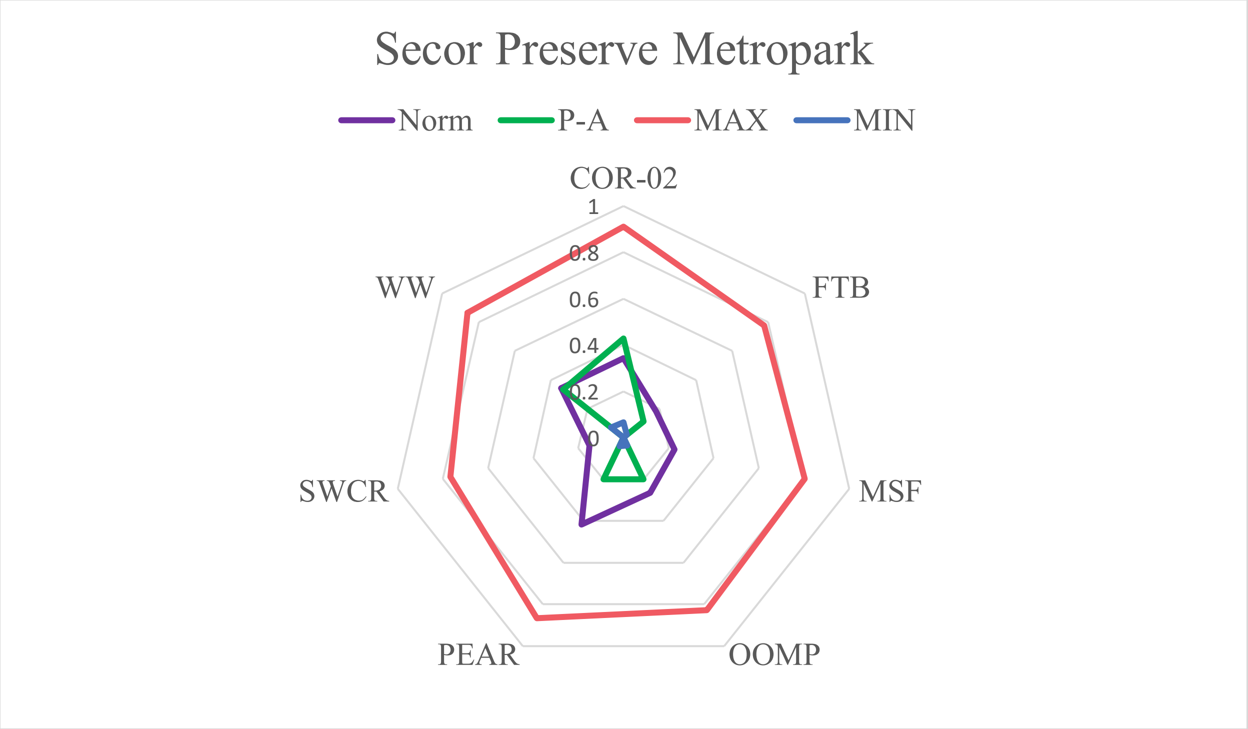


Figure C.6. Bray-Curtis dissimilarity outputs for Secor Preserve Metropark compared to the other seven locations in the region. Lines denote the output values for dissimilarity highlighting maximum dissimilarity obtained for a given set of communities (red), minimum dissimilarity obtained for a given set of communities (blue), dissimilarity using the unadjusted abundance data for a given set of communities (purple), and dissimilarity of sites using presence-absence alone (green). Site location names corresponding to the figure’s acronyms can be found in the materials and methods section.

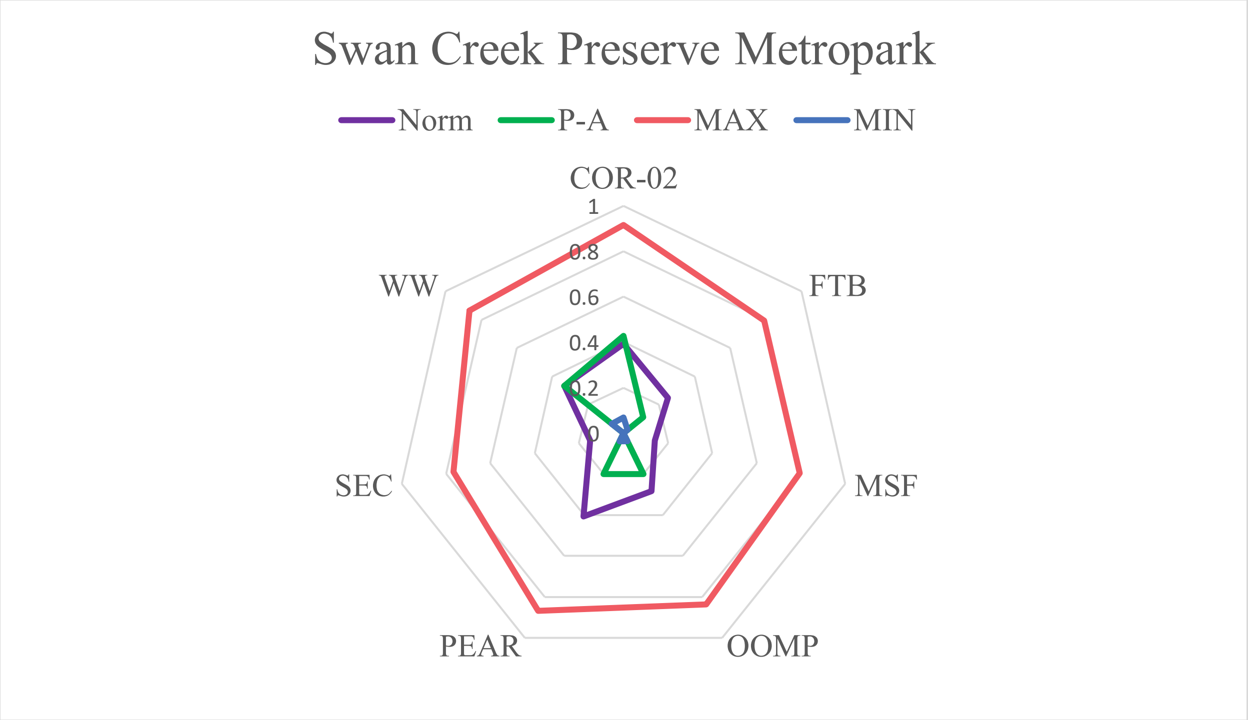


Figure C.7. Bray-Curtis dissimilarity outputs for Swan Creek Preserve Metropark compared to the other seven locations in the region. Lines denote the output values for dissimilarity highlighting maximum dissimilarity obtained for a given set of communities (red), minimum dissimilarity obtained for a given set of communities (blue), dissimilarity using the unadjusted abundance data for a given set of communities (purple), and dissimilarity of sites using presence-absence alone (green). Site location names corresponding to the figure's acronyms can be found in the materials and methods section.

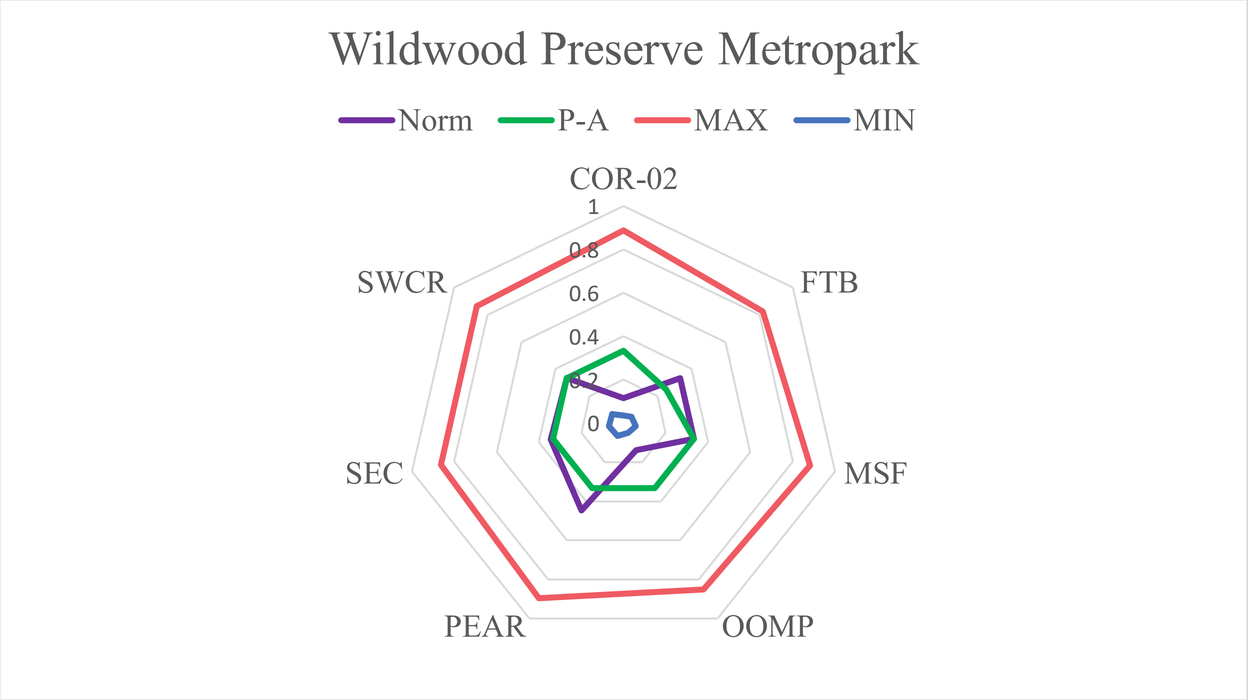


Figure C.8. Bray-Curtis dissimilarity outputs for Swan Creek Preserve Metropark compared to the other seven locations in the region. Lines denote the output values for dissimilarity highlighting maximum dissimilarity obtained for a given set of communities (red), minimum dissimilarity obtained for a given set of communities (blue), dissimilarity using the unadjusted abundance data for a given set of communities (purple), and dissimilarity of sites using presence-absence alone (green). Site location names corresponding to the figure's acronyms can be found in the materials and methods section.

APPENDIX D: CASE STUDY OF PSEUDOREPLICATION IN MAUMEE STATE FOREST

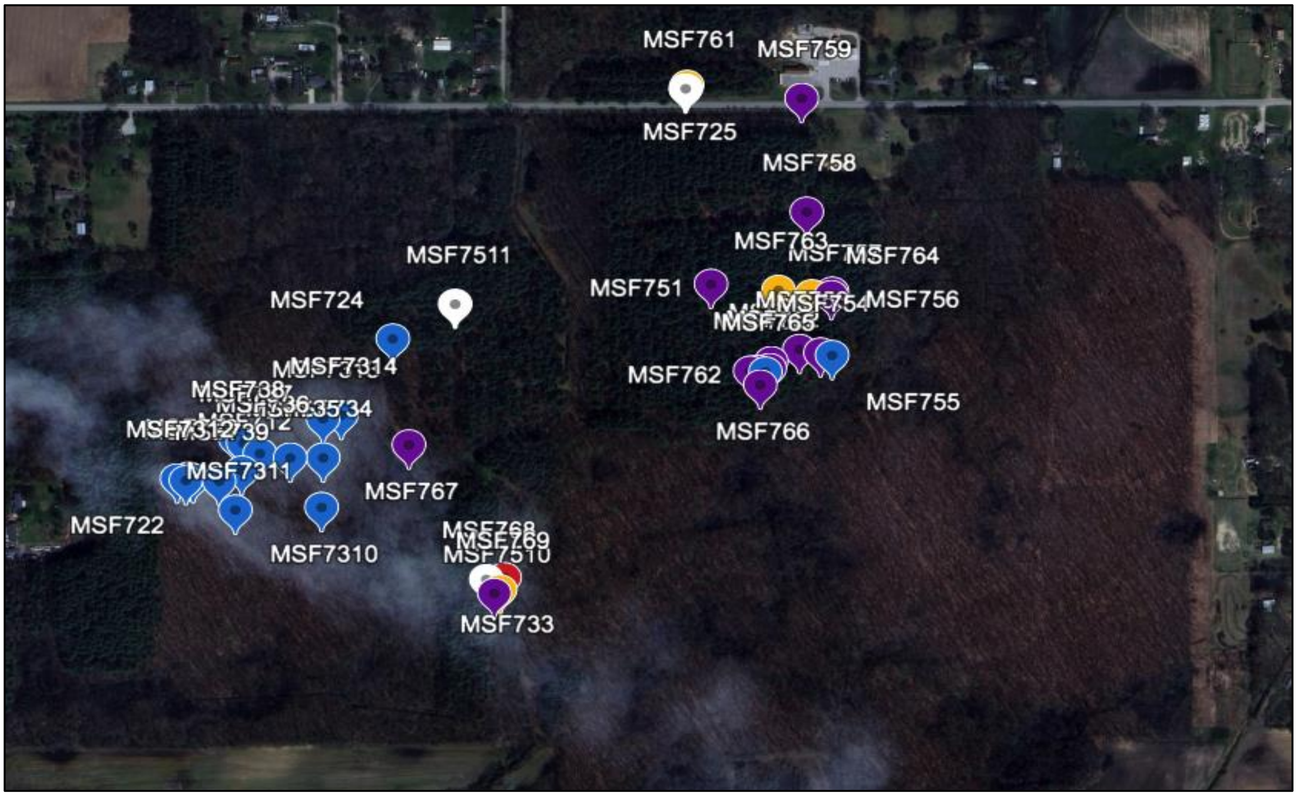
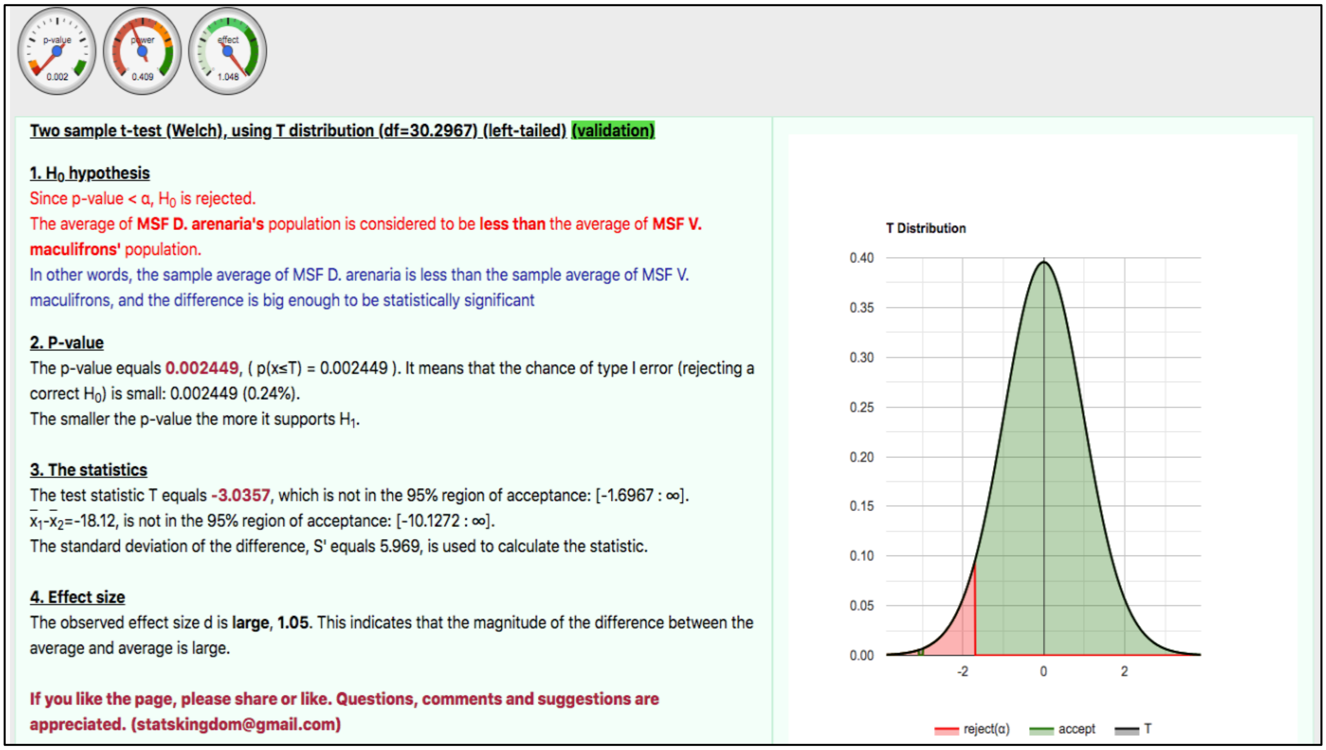


Figure D.1. Map of specimen locations at the MSF site. Blue markers indicate *V. maculifrons* specimens collected in broadleaf forest. Purple markers indicating *D. arenaria* specimens collected in pine forest. Points represented in other colors display the location of other species in the community.



Two sample t-test (Welch), using T distribution (df=30.2967) (left-tailed) (validation)

1. H_0 hypothesis

Since p-value < α, H_0 is rejected.

The average of **MSF D. arenaria's** population is considered to be **less than** the average of **MSF V. maculifrons'** population.

In other words, the sample average of MSF D. arenaria is less than the sample average of MSF V. maculifrons, and the difference is big enough to be statistically significant

2. P-value

The p-value equals **0.002449**, ($p(x \leq T) = 0.002449$). It means that the chance of type I error (rejecting a correct H_0) is small: 0.002449 (0.24%).

The smaller the p-value the more it supports H_1.

3. The statistics

The test statistic T equals **-3.0357**, which is not in the 95% region of acceptance: [-1.6967 : ∞].

$\bar{x}_1-\bar{x}_2=-18.12$, is not in the 95% region of acceptance: [-10.1272 : ∞].

The standard deviation of the difference, S' equals 5.969, is used to calculate the statistic.

4. Effect size

The observed effect size d is **large**, **1.05**. This indicates that the magnitude of the difference between the average and average is large.

If you like the page, please share or like. Questions, comments and suggestions are appreciated. (statskingdom@gmail.com)

Figure D.2. Output information for the comparison of canopy cover distributions in Maumee State Forest between significantly clustered species *V. maculifrons* and *D. arenaria.* Results are displayed using statskingdom.com's two sample t-test.