Inventorying and Monitoring Urban Insect Biodiversity for the Natural History Museum of Los Angeles County’s BioSCAN Program

Amanda Russell, Jennifer Truong, Joanna Shen, Jonathan Kalmanson, Rachel Chung, Sarah Toback, Trevor Johnson

Advisors: Dr. Thomas Gillespie (UCLA Geography Dept.), Dr. Dean Pentcheff (Natural History Museum), and Dr. Travis Longcore (USC Spatial Sciences Institute)

Environmental Science Senior Practicum 2015, UCLA
Abstract

The Natural History Museum of Los Angeles County initiated their “Biodiversity Science: City and Nature” (BioSCAN) project in 2012, as a way to quantify insect biodiversity and how it varies across Los Angeles, CA. Within an urban setting, there are many landscape parameters that may have an effect on species richness. In this report, we investigate the number of plant species surrounding BioSCAN malaise traps, land cover and land use, surface temperature, tree canopy cover, annual average precipitation, photosynthetic activity, elevation, and night light intensity, and determine how each of these parameters is related to insect biodiversity. We found that land use (including open space and medium intensity of development), tree canopy cover, impervious surfaces, mean temperature, mean precipitation, normalized difference vegetation index (mean and center), elevation, were all significantly correlated with species richness and generic richness. One tree canopy cover dataset alone can explain 41.2% of the species richness variance (R-squared = 0.412). The metrics with the strongest correlations (NDVI center, percent open space, mean precipitation, percent impervious surface, percent tree canopy cover, and mean temperature) were combined into a multiple regression model that is able to explain 50.1% of the species richness variance. We hope that these landscape metrics can be used to help predict biodiversity patterns and be scaled up to urban areas across the country in order to draw broader conclusions about biodiversity in cities.

Introduction

Traditionally, landmark conservation and ecology studies have examined animal populations in pristine environments, but with accelerating habitat disturbances and urbanization around the world, it has become increasingly important to also examine how biodiversity is affected by the abiotic and biotic components of man-made landscapes. Habitat loss, alteration, and fragmentation are all consequences of urbanization that may influence local diversity. Many studies have been conducted researching the impacts of urbanization on various taxa, and while these impacts vary greatly, the consensus of the majority of research is that urbanization has an overall negative impact on the diversity of species including insects, birds, amphibians, and mammals.

In some studies, biodiversity is assessed using manual field work. The labor is often expensive as well as time-consuming, and can vary in feasibility depending on the target area and subjects (Duro 2007). Remote sensing can alleviate many of these difficulties and provide more convenient access to larger data sets. Satellite imagery and other types of remote sensing can provide spatial data for vast areas of land, and give researchers the ability to remotely confirm field observations or assess inconvenient locations. Most satellites can provide recurring images that analyzed to detect changes in the environment over time, including parameters that are relevant to ecology and biodiversity. However, there is a very large gap in knowledge regarding urban biodiversity and its relationship with remote sensing techniques for assessment, especially in regards to insects (Du 2014, Muller 2009). The purpose of our research is to bridge this gap by assessing the relationship between various environmental factors and the biodiversity data provided by the BioSCAN project.
Methods

We obtained species and genus richness data for the BioSCAN sites from B. Brown, curator of the Entomology section at the Natural History Museum.

<table>
<thead>
<tr>
<th>Site</th>
<th>Species Richness</th>
<th>Genera</th>
<th>Site</th>
<th>Species Richness</th>
<th>Genera</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38</td>
<td>7</td>
<td>11</td>
<td>37</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Excluded</td>
<td></td>
<td>12</td>
<td>38</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>9</td>
<td>13</td>
<td>77</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>10</td>
<td>14</td>
<td>44</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>11</td>
<td>15</td>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>46</td>
<td>12</td>
<td>16</td>
<td>43</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>41</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>24</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>18</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>37</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>38</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>37</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>44</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>27</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>43</td>
<td>10</td>
<td>32</td>
<td>Excluded</td>
<td>Excluded</td>
</tr>
</tbody>
</table>

Table 1. Species richness and genera richness data for each of the BioSCAN sites, excluding sites 2 and 32.

We identified landscape factors that might be related to insect biodiversity, and gathered data on these metrics using a combination of remote sensing techniques, GIS, and field measurements.

Study Area

The study area consisted of 30 sites in Los Angeles County. We excluded sites 2 and 32 for all of our analysis, as BioSCAN species and genera-level richness data were not provided for these sites. When analyzing our field data, we also excluded sites 22 and 30, as the raw data from this step of our research is still being processed. Finally, as the McPherson land cover dataset only included the city of Los Angeles, we excluded sites 17 and 21, which were just outside the city boundary.

Field Data

First, we assessed plant diversity surrounding each site, in order to test if plant diversity has a positive correlation with insect diversity. Past studies showed that higher plant diversity meant more resources for the insects to use (Pardee and Philpott 2014). To measure the plant diversity in the BioSCAN sites, we collaborated with Dean Pentcheff and Lisa Gonzales from the Natural History Museum to organize site visits to collect data. From April until the end of May, we visited two to three sites every Friday and Saturday on average. We first developed a method where we measured radii of 2, 5, 10, and 20 meters from the malaise trap found at each site (Figure 1). We took pictures of the different plants found at each radii and stored these pictures in a central Google Drive folder. We then began to identify the plant species with the help of our
advisor, Dr. Tom Gillespie. Some plants that were not identified easily were collected from the sites and put in an herbarium press, where Dr. Gillespie helped us identify them later. Obstacles that disrupted the radii such as fences and pathways were noted. Once the plants were identified, each species was accounted for and the number of species was recorded for later analysis.

Figure 1. A layout of a typical observational study at a BioSCAN field site.

Land Cover and Land Use

For datasets on land cover and land use, data was extracted for analysis in 1 hectare square plots (100 m x 100 m) centered on each BioSCAN site. This included LARIAC tree canopy cover, McPherson land cover derived from Quickbird satellite imagery, and land use from the USGS Gap Analysis Program. These square plots were created as a layer in ArcGIS by drawing circular buffer polygons around each site location, each with radii of 50 meters, and then applying the “Feature Envelope to Polygon” tool to create square polygons with dimensions of approximately 100 m x 100 m. It should be noted that, using this method, the square buffers produced were slightly below one hectare in area; instead, each plot was within about 1 square meter of 9980 m² in total area.

USGS Gap Analysis Program (GAP)

Land use data (resolution of ~30 m) for the region was acquired from the USGS National Gap Analysis Program, or GAP (http://gapanalysis.usgs.gov/gaplandcover/data/). “Gap analysis” is a method of biodiversity conservation that maps suitable habitat and species richness across a region, in order to find ‘gaps’ where species need additional protection and are not covered by existing “conservation networks” (Allen, Pearlstine, and Kitchens 2001). While GAP was not necessarily created for the assessment of urban biodiversity, we believe that it is worth looking into how well the data predicts diversity, because it is both easy to access and covers the entire United States; therefore, if successful, our methods can easily be applied to almost any urban area in the country.
The highest level of detail provided by GAP is the “Ecological System or Land Use Class” attribute, at a spatial resolution of 30 m. The ecological systems were based on NatureServe’s Ecological System Classification for “natural and semi-natural vegetated areas” (Comer et al. 2003; Lowry et al. 2005). Land Use Class refers mainly to developed areas not included in the Ecological System Classification. These classes were derived from the National Land Cover Dataset (NLCD), which sorts developed areas into four categories. These categories are defined by Homer et al. (2004) as follows: (1) “Developed, Open Space” is described as “a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses,” where the land cover consists of less than 20% impervious surfaces; (2) “Developed, Low Intensity” includes areas in which impervious surfaces make up between 20 and 49% of the land cover; (3) “Developed, Medium Intensity” includes areas with between 50 and 79% impervious surfaces; and (4) “Developed, High Intensity” regions are “areas where people reside or work in high numbers… [such as] apartment complexes, row houses, and commercial/industrial,” and are from 80 to 100% impervious.

The GAP land cover dataset for California was downloaded as an ESRI raster file, and imported to ArcGIS. A portion of the grid covering the total extent of the BioSCAN site locations was extracted and then projected using the North America Albers Equal-Area Conic projection. The extracted portion of the GAP data was then converted to a vector file using the “Raster to Polygon” ArcGIS tool, selecting “ECOLSYS_LU” as the output feature class. The new polygon features were then intersected with the shapefile of square buffers, in order to extract the land use data for the 100 m x 100 m area surrounding each BioSCAN site. Within each 100 m x 100 m site, the area covered by each land use type was calculated. Any other portions of the plot were combined into one “Undeveloped/Other” category. Finally, the attribute table was exported and the calculated areas were converted into percentages of the total area within each plot. The expectation is that, the higher the overall “intensity” of development within a given plot, the lower the insect diversity is likely to be.

**LARIAC Tree Canopy**

The Los Angeles Regional Imagery Acquisition Consortium (LARIAC) produced Tree Canopy data ([http://egis3.lacounty.gov/dataportal/2010/12/23/tree-canopy-raster-2006-data/](http://egis3.lacounty.gov/dataportal/2010/12/23/tree-canopy-raster-2006-data/)) - with resolution of 5 feet, or ~1.5 meters - during the development of the L.A. County Solar Map and Green Planning Tool ([http://solarmap.lacounty.gov](http://solarmap.lacounty.gov)). They did this by developing a Digital Surface Model (derived from LARIAC lidar data) and Digital Elevation Model (elevation, if trees, buildings, and other features were not present), both with resolution of 5 feet, in combination with an NDVI layer calculated using 4-inch resolution red and infrared-band images. (source: [http://egis3.lacounty.gov/dataportal/wp-content/uploads/2010/02/Poster-Developing-the-Solar-Model-for-the-LA-County-Solar-Mapping-Portal.pdf](http://egis3.lacounty.gov/dataportal/wp-content/uploads/2010/02/Poster-Developing-the-Solar-Model-for-the-LA-County-Solar-Mapping-Portal.pdf)). Tree canopy was defined as any portions of the map in which the Surface Height Model - the difference between the digital elevation model and the digital surface model - was greater than 8 feet, and the
normalized difference vegetation index (NDVI) value was greater than 0.1 (to distinguish trees from other tall objects, such as buildings) (source: http://egis3.lacounty.gov/dataportal/2010/12/23/tree-canopy-raster-2006-data/)

As with the GAP data, the LARIAC tree canopy data was extracted in 1 hectare plots, for each BioSCAN site. After determining the area of each plot that was covered by tree canopy, this was used to calculate the area as a percentage of the total area of the plot.

**McPherson Land Cover**

We also acquired land cover data (resolution of ~1 m) for the Los Angeles region, developed by McPherson et al. (2010). This data was created using 82 images - captured between 2002 and 2005 - from the Quickbird satellite; most were taken during the warmer months, when deciduous trees are “in leaf” (McPherson et al. 2010).

The McPherson land cover data contains four categories of land cover: “Trees” (which included trees and shrubs), “Irrigated Grass” (green grass and ground cover), “Dry Grass/Bare Soil,” and “Impervious Surface/Other.” Pixels were classified based on radiance (brightness) rather than reflectance, and the data was refined using NDVI-based “masks” that filter out specific categories of land cover: vegetation, nonvegetation, and dry/unirrigated vegetation (McPherson et al. 2010). Classification of trees/shrubs was human-assisted, to distinguish trees from irrigated grass.

The dataset was imported to ArcGIS in order to extract land cover features within the 1 hectare plot at each BioSCAN site. Within each plot (minus two sites that were not covered by the dataset, as they were outside the Los Angeles city limits), the area was calculated for each land cover type. Finally, the attribute table was exported, and the area values were used to calculate each land cover area as a percent of the total area within the plot.

**WorldClim Mean Annual Temperature and Precipitation**

To evaluate the effect of precipitation and temperature on urban insect biodiversity, we acquired data from Worldclim (http://www.worldclim.org/current) in an ESRI grid format (resolution of 30 arc-seconds, or ~1 km). This dataset was created by spatially interpolating precipitation and temperature values based on data from tens of thousands of weather stations located around the world, with a model that included longitude, latitude, and elevation as independent variables (Hijmans et al. 2005). Worldclim temperature data was in units of °C*10 - a temperature of 23.1°C is displayed as 231 - and precipitation is in millimeters (source: http://www.worldclim.org/formats).
This data, along with the BioSCAN site locations, was then imported and mapped using ArcGIS. The grids were cropped to include the full extent of the BioSCAN sites, using the “Extract by Rectangle” ArcGIS tool. Finally, we recorded the mean temperature and precipitation for each site using the pixel values that most closely overlap with each site.

**Landsat 8: Temperature**

To begin this portion of our analysis, the Landsat 8 satellite data (resolution of ~30 m) was downloaded from the USGS EarthExplorer application. Landsat 8 has collected data every 16 days since its launch in 2013 (USGS, 2013). The imagery used was captured on August 14, 2014, during the peak of the dry season. Once this data was accessed, ENVI Classic was used to display the data. We used “Band 10”, one of two thermal infrared bands provided by the Landsat 8 program, as a measure of surface temperature. It has a resolution of 30 meters, and units are brightness values for the 10.60 to 11.19 micrometer wavelength range (http://landsat.usgs.gov/band_designations_landsat_satellites.php). The BioSCAN site locations were then imported as a vector file and overlaid on the map.

Using the pixel grid boxes on ENVI Classic (Figure 2), data was recorded for the pixel underlying each site location, as well as the top left, top center, top right, center left, center right, bottom left, bottom center, and bottom right pixels surrounding each site.

![Figure 2. In ENVI Classic, we can distinguish between the different pixels which each have a unique value. This image shows the nine-pixel ‘window’ that we looked at when determining the values of the remote sensing metrics for each site; this included Landsat 8 NDVI and Temperature, as well as Suomi NPP Night Light Intensity.](image-url)
**Landsat 8: Normalized Difference Vegetation Index (NDVI)**

The images obtained from the Landsat 8 satellite (resolution of 30 m) previously used in analyzing temperature were also used to calculate the Normalized Difference Vegetation Index (NDVI). The images downloaded from USGS EarthExplorer (source: [http://earthexplorer.usgs.gov/](http://earthexplorer.usgs.gov/)) were taken on August 14th, 2014. Bands 4 and 5 were imported into ENVI Classic, and the data was transformed by ENVI Classic into NDVI. The resulting image was imported as a vector file and overlaid with the BioSCAN site locations.

The pixel indicator identified the GPS coordinates of each individual site, and the spatial pixel editor measured the NDVI value for individual pixels. Values were recorded for the pixel containing the malaise trap as well as the surrounding pixels, following the same procedure as what was used for the Landsat temperature data. The first variable defined was “NDVI mean”, which averaged the NDVI values recorded from each of the nine pixels surrounding the malaise trip for each site. The second variable, “NDVI standard deviation” calculated the standard deviation from NDVI mean for each site. Finally, the variable “NDVI center” measured the value of the center pixel alone, which contains the malaise trap.

**Elevation**

The data for elevation was recorded using Google Earth. After entering the coordinates for each BioSCAN site location, the “elevation profile” tool was selected. This feature produced the exact elevation (in feet) at the site which was then converted into meters.

**Suomi NPP Night Lights**

The Night Light images were obtained from NASA’s Suomi National Polar-Orbiting Partnership VIIRS Sensor as Large GeoTIFF composite image of the United States (Source: [http://earthobservatory.nasa.gov/IOTD/view.php?id=79800](http://earthobservatory.nasa.gov/IOTD/view.php?id=79800)) from April 18 - October 23, 2012. The Defense Meteorological Satellite Program provided a secondary source of night light images, but the resolution of Suomi NPP (~742m) is much greater than DMSP (3km), making the DMSP less suitable given the proximity of some BioSCAN sites (Source: [http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-87-2-191](http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-87-2-191)).

Images were taken using Visible Infrared Imaging Radiometer Suite (VIIRS) day-night, low light band to show ambient light from anthropogenic sources on both land and water. BioSCAN site locations were then overlaid on the GeoTIFF image using ENVI Classic. Values were recorded for each pixel surrounding the site location. This included the top left, top center, top right, middle left, center, middle right, lower left, lower center, and lower right pixels. The center pixel contained the coordinates of the malaise trap. Georectified information included three values per pixel, which correlates with the RGB values for each pixel. The RGB values indicate
the brightness of ambient night light. Because of the relatively coarse spatial resolution (~742 m), we chose to conduct statistical analysis only on the center pixel brightness values.

**Statistical Analysis**

All species and landscape metrics were analyzed in SPSS. We examined all variables for a normal distribution and examined correlations with a Pearson-moment correlation coefficient. We also examined a relationship of highly correlated landscape metrics and species richness using a simple linear regression and multiple regressions. R-squared values were then calculated using a regression to determine the statistical relationship between the variables. The most highly correlated landscape metrics were combined using multiple regression analysis.

**Results**

<table>
<thead>
<tr>
<th>Linear Regression</th>
<th>R</th>
<th>R²</th>
<th>Std. Error of Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable: Species Richness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDVI Mean</td>
<td>0.384</td>
<td>0.147</td>
<td>11.8685</td>
</tr>
<tr>
<td>% Developed, Medium Intensity</td>
<td>0.419</td>
<td>0.176</td>
<td>11.6708</td>
</tr>
<tr>
<td>NDVI Center</td>
<td>0.426</td>
<td>0.182</td>
<td>11.6278</td>
</tr>
<tr>
<td>Mean Temp</td>
<td>0.516</td>
<td>0.240</td>
<td>11.0090</td>
</tr>
<tr>
<td>% Impervious/Other</td>
<td>0.497</td>
<td>0.247</td>
<td>11.2834</td>
</tr>
<tr>
<td>% Developed, Open Space</td>
<td>0.520</td>
<td>0.270</td>
<td>10.9820</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>0.583</td>
<td>0.340</td>
<td>10.4411</td>
</tr>
<tr>
<td>Mean Precipitation</td>
<td>0.624</td>
<td>0.389</td>
<td>10.0448</td>
</tr>
<tr>
<td>% Trees (McPherson)</td>
<td>0.642</td>
<td>0.412</td>
<td>9.9700</td>
</tr>
<tr>
<td>Multiple Regression:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-NDVI Center</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- % Open Space</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Mean Precipitation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- % Impervious/Other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- % Tree Canopy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Mean Temperature</td>
<td>0.708</td>
<td>0.501</td>
<td>10.2208</td>
</tr>
</tbody>
</table>

*Table 2. Linear regression results of each variable from SPSS*
Using SPSS, we combined several landscape metrics into a linear model, with the goal of explaining as much of the variance in biodiversity as possible. The model includes NDVI Center, Percent Open Space, Mean Precipitation, Percent Impervious Surfaces, tree canopy cover, and mean temperature. When combined, these metrics have an R-squared value of 0.501, meaning that 50.1% of the biodiversity variance can be explained with these variables.

**Field Data**

<table>
<thead>
<tr>
<th></th>
<th>Plants (0 to 2 m)</th>
<th>Plants (2 to 5 m)</th>
<th>Plants (0 to 5 m)</th>
<th>Plants (5 to 10 m)</th>
<th>Plants (0 to 10 m)</th>
<th>Plants (10 to 20 m)</th>
<th>Plants (0 to 20 m)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>-0.168</td>
<td>0.055</td>
<td>-0.011</td>
<td>0.157</td>
<td>0.101</td>
<td>0.304</td>
<td>0.193</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.466</td>
<td>0.814</td>
<td>0.963</td>
<td>0.497</td>
<td>0.662</td>
<td>0.181</td>
<td>0.402</td>
</tr>
<tr>
<td><strong>Genera</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>-0.081</td>
<td>0.086</td>
<td>0.041</td>
<td>0.1</td>
<td>0.086</td>
<td>0.301</td>
<td>0.183</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.726</td>
<td>0.71</td>
<td>0.859</td>
<td>0.665</td>
<td>0.711</td>
<td>0.184</td>
<td>0.427</td>
</tr>
</tbody>
</table>

*Table 3. Statistical analysis of field data collected from the 30 BioSCAN sites.*

The correlations for all radii in relation to species richness changed from negative to positive depending on the distance from the malaise trap, although none were statistically significant. For plant species richness in the 5 meter circle surrounding the malaise trap, there was a weak negative correlation with insect diversity ($r=-0.168$ for zero to 2 meters, $r=-0.011$ for 0 to 5 m). As we reached 10 meters away from the trap, the correlation began to be more positive ($r=0.101$). It continued to be more positive as we reached 20 meters ($r=0.193$). The correlations in relation to genera richness followed the same pattern as the correlations for species richness, except on a smaller scale.
**GAP Land Use**

The correlation results on the GAP land use dataset were mixed. As predicted, sites with higher overall “intensity” of development tended to have lower biodiversity levels. For example, when tested against our species richness data, the proportion of “Developed, Open Space” area appeared to have a strong positive correlation with insect species richness ($r=0.520, p<0.01$). The “Developed, Medium Intensity” category stood out as well, having a relatively strong negative correlation ($r=-0.419, p<0.05$). While the “Low Intensity” category had a weaker correlation ($r=0.204$), it could still be interpreted as supportive of the same trend. Just as interesting were the two extreme cases - “Developed, High Intensity” and “Undeveloped/Other” - both of which are almost completely uncorrelated with the biodiversity metrics.

<table>
<thead>
<tr>
<th>Species Richness</th>
<th>% Developed, High Intensity</th>
<th>% Medium Intensity</th>
<th>% Low Intensity</th>
<th>% Open Space</th>
<th>% Undeveloped/ Other</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pearson Correlation</strong></td>
<td>-0.022</td>
<td>-0.419</td>
<td>0.204</td>
<td>0.520</td>
<td>0.052</td>
</tr>
<tr>
<td><strong>Sig. (2-tailed)</strong></td>
<td>0.908</td>
<td>0.021</td>
<td>0.281</td>
<td>0.003</td>
<td>0.784</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genera</th>
<th>% Developed, High Intensity</th>
<th>% Medium Intensity</th>
<th>% Low Intensity</th>
<th>% Open Space</th>
<th>% Undeveloped/ Other</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pearson Correlation</strong></td>
<td>-0.085</td>
<td>-0.383</td>
<td>0.155</td>
<td>0.567</td>
<td>0.088</td>
</tr>
<tr>
<td><strong>Sig. (2-tailed)</strong></td>
<td>0.655</td>
<td>0.037</td>
<td>0.415</td>
<td>0.001</td>
<td>0.644</td>
</tr>
</tbody>
</table>

*Table 4. Statistical analysis of land use from the 30 BioCAN sites.*

Given the somewhat coarse spatial resolution of the GAP data, the range in values for each land use metric was quite large. In the most extreme case, “Developed, Medium Intensity” ranged from 0% all the way to 100% (as a fraction of the standard 1 hectare plot surrounding each site). The mean was 52.0%, and the standard deviation was 39.6% The distribution of the “Developed, Open Space” variable was not quite as severe, being from 0% to 61.7%. The mean was 11.2%, and the standard deviation was 19.5%.

**LARIAC Tree Canopy**

The percent tree canopy calculated from the LARIAC dataset was not significantly correlated with either species richness ($r=0.312, p=0.093$) or genera ($r=0.337, p=0.069$).
Table 5. Statistical analysis of tree canopy cover from the 30 BioSCAN sites.

The tree canopy estimates (as a percentage of the 1 hectare region surrounding each site) varied significantly, ranging from ~6.23% (622 m² of tree canopy) to 57.6% (5747 m²). The mean was 22.8%, and the standard deviation was 13.3%.

**McPherson Land Cover**

Table 6. Statistical analysis of land cover from the 30 BioSCAN sites.

The mean value for the “Impervious/Other” classification was 49.0%, with a minimum of 4.26%, maximum of 83.6%, and a standard deviation of 22.4%. This variable was strongly correlated with species richness, with a significance of 0.007 (r=-0.497, p<0.01; R-squared=0.247). The correlation of “Impervious/Other” and genera was even stronger, and once again had a significance of 0.007 (r=-0.501, p<0.01). This statistically significant negative correlation suggests that as the proportion of “Impervious/Other”-type land cover increases, species richness and genera decreases.

The mean for tree canopy cover was 31.8%, with a minimum of 7.61%, maximum of 89.3%, and a standard deviation of 21.1%. Of all the variables we tested, tree canopy cover was most strongly correlated with species richness (r=0.642, p<0.01; R-squared 0.412) as well as genera (r=0.607, p<0.01). This statistically significant positive correlation indicates that as tree canopy cover area increases as a proportion of total area, insect species richness and genera increase as well.
Interestingly, the “Irrigated Grass” variable was negatively correlated with species richness and genera, though the results were not statistically significant. In addition, the results for “Dry Grass or Bare Soil” were actually positively correlated with species richness and genera, but not to a significant degree.

**WorldClim Mean Annual Temperature and Precipitation**

<table>
<thead>
<tr>
<th></th>
<th>Mean Temp</th>
<th>Mean Precipitation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species Richness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>-0.516</td>
<td>0.624</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.004</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Genera</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>-0.472</td>
<td>0.584</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.009</td>
<td>0.001</td>
</tr>
</tbody>
</table>

*Table 7. Statistical analysis of mean annual temperature and precipitation from the 30 BioSCAN sites.*

The mean, across all the BioSCAN sites, for the annual temperature data was 13.52 °C, with standard deviation of 0.539°C, and a range of 12.6 to 14.2 °C. There was a significant negative correlation between mean temperature and species richness \( r=-0.516, p<0.01; \text{R-squared}=0.240 \), as well as mean temperature and genera \( r=-0.472, p<0.01 \), indicating that an area with higher mean temperature would yield less diversity in genera and species richness.

The mean for the annual precipitation data was 87.5 mm, the standard deviation was 7.82 mm, and the range was 73 to 104 mm. The \( R^2 \) value for mean precipitation is 0.389. There are significant positive correlations between mean precipitation and species richness \( r=0.624, p<0.01; \text{R-squared}=0.389 \), and mean precipitation and genera \( r=0.584, p<0.01 \). This indicates that an area with higher mean precipitation is likely to have greater insect biodiversity, at both the genera and species level. The high Pearson correlation values between mean precipitation and species richness and genera show that mean precipitation is a crucial factor in the overall variance in urban insect biodiversity.
**Landsat 8: Temperature**

<table>
<thead>
<tr>
<th>Species Richness</th>
<th>Temperature (Mean of the 9 pixels)</th>
<th>Stand. Dev.</th>
<th>Temperature (Center pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>-0.32</td>
<td>0.123</td>
<td>-0.323</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.085</td>
<td>0.518</td>
<td>0.082</td>
</tr>
</tbody>
</table>

| Genera | Pearson Correlation | -0.351 | 0.163 | -0.351 |
| Sig. (2-tailed) | 0.058 | 0.391 | 0.057 |

Table 8. Statistical analysis of temperature from the 30 BioSCAN sites.

The brightness for the “center” pixels had an average value of 135.5, with standard deviation of 3.54, and a range from 107 to 183. The center pixel values were not significantly correlated with species richness ($r=-0.323, p=0.082$), or genera ($r=-0.351, p=0.058$). This was similar for mean brightness for the nine-pixel window, in that the correlation was not quite significant for either species richness ($r=-0.32, p=0.085$) or genera ($r=-0.351, p=0.058$).

The standard deviation of values in each nine-pixel window had an even weaker correlation with species richness ($r=0.123, p=0.518$) and genera ($r=0.163, p=0.391$), suggesting that insect diversity is not dependent on local variation in surface temperatures.

**Landsat 8: NDVI**

<table>
<thead>
<tr>
<th>Species Richness</th>
<th>NDVI Mean</th>
<th>NDVI Stand. Dev.</th>
<th>NDVI Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>.384</td>
<td>0.248</td>
<td>0.426</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.036</td>
<td>0.186</td>
<td>0.019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genera</th>
<th>NDVI Mean</th>
<th>NDVI Stand. Dev.</th>
<th>NDVI Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>.373</td>
<td>0.295</td>
<td>0.385</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.042</td>
<td>0.114</td>
<td>0.036</td>
</tr>
</tbody>
</table>

Table 9. Statistical analysis of NDVI from the 30 BioSCAN sites.

The average value and standard deviation of the NDVI mean (nine-pixel window) variable was 0.226 and 0.123 respectively. The highest value was 0.471, and the lowest value was 0.054. The mean and standard deviation were also calculated for the NDVI standard deviation variable, and were 0.072 and 0.028 respectively. The highest value of this variable was 0.132, and the lowest was 0.023. Finally, the mean and standard deviation were calculated for the NDVI center variable, and were 0.238 and 0.124 respectively. The highest value was 0.521, and the lowest value was 0.101.
Species richness and NDVI mean were positively correlated \( (r=0.384; p<0.05) \), and species richness and NDVI center were also positively correlated \( (r=0.426; p<0.05) \). However, species richness was not correlated with NDVI standard deviation. Genera and NDVI mean was positively correlated \( (r=0.373; p<0.05) \), as well as genera and NDVI center \( (r=0.385; p<0.05) \). Again, there was no correlation between genera and NDVI standard deviation. The R-squared value of species richness and NDVI mean center was 0.147, and 0.182 for species richness and NDVI center.

**Elevation**

The mean for the elevation data from Google Earth was 158.3 meters, the standard deviation was 146.3 meters, and the range was 9.144 to 261.823 meters. The Pearson correlation between species richness and elevation was 0.34. The r value for genera richness and elevation was 0.321.

**Night Light Intensity**

<table>
<thead>
<tr>
<th></th>
<th>Center R</th>
<th>Center G</th>
<th>Center B</th>
<th>St. Dev. (Intensity)</th>
<th>Mean (Intensity)</th>
<th>Center (Intensity)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species Richness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>0.084</td>
<td>0.007</td>
<td>-0.078</td>
<td>-0.051</td>
<td>0.1</td>
<td>0.083</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.659</td>
<td>0.971</td>
<td>0.68</td>
<td>0.789</td>
<td>0.598</td>
<td>0.662</td>
</tr>
<tr>
<td><strong>Genera</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>0.112</td>
<td>0.055</td>
<td>-0.044</td>
<td>-0.082</td>
<td>0.14</td>
<td>0.111</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.556</td>
<td>0.771</td>
<td>0.816</td>
<td>0.666</td>
<td>0.461</td>
<td>0.558</td>
</tr>
</tbody>
</table>

*Table 10. Statistical analysis of night light intensity from the 30 BioSCAN sites.*

The mean Red/Green/Blue values were calculated for the center pixels for all sites. The mean values were [255, 249, 211]. The standard deviations between RGB values in the center pixels are [0.751, 7.779, 24.006]. The highest values (of all sites cumulatively) were [255, 254, 249], and the lowest values were [251, 214, 139]. Species richness and night light RGB “color” were marginally positively correlated \( (r=0.084, p=0.659) \) for the red band, marginally positively correlated \( (r=0.007, p=0.971) \) for the green band, and marginally negatively correlated \( (r=-0.078, p=0.68) \) for the blue band. Genera and night light RGB values were slightly positively correlated \( (r=0.112, p=0.556) \) for the red band, marginally positively correlated \( (r=0.005, p=0.771) \) for the green band, and marginally negatively correlated \( (r=-0.044, p=0.816) \) for the blue band. None of the calculated r or p-values were statistically significant, though this might be a result of the RGB values losing significance as standalone values, which is a limitation to this data set.


Discussion

Field Data

The correlations for our field data were surprisingly lower and weaker than expected, as we expected high plant diversity to be correlated with high insect diversity according to past literature (Bates et al. 2014; Pardee and Philpott 2014). However, since thirty sites were a relatively small sample scattered over a large area, the weaker correlations might not have been that surprising. Our measurements of the plant diversity of each site have been rough estimates, so a more precise method of measuring plant diversity besides identifying plants based on pictures would be more helpful and can be done in the future.

During our research, we counted the number of different plant species seen instead of identifying each plant as native or non-native, as we did not have the proper resources nor the knowledge to identify all plants that were observed. If we had more access to current technology, such as iNaturalist (http://www.inaturalist.org/), we would have been able to identify each plant species relatively quickly. This method of measuring plant diversity could be something to consider for future research.

GAP Land Use

In the GAP Land Use data, the low correlations of “Developed, High Intensity” and “Undeveloped/Other” with the biodiversity metrics make it difficult to reject the null hypothesis (that intensity of development and insect diversity are unrelated) outright. While additional research is still needed, it may simply be that biodiversity in the most highly developed urban areas - insect diversity, at least - is not as severely impacted as one might expect. Alternatively, it could be a limitation of the GAP classification system - or, more specifically, the National Land Cover Data that the “Land Use” classes were based on - which groups a potentially very broad range of land uses together into somewhat vague categories. For this type of research, it may be simpler - and provide more meaningful and consistent results - if the data that these classification schemes are based on (NDVI, proportion of impervious surfaces, etc.) is focused on instead.

Since the Gap Analysis Program covers the entire contiguous United States, as well as Alaska, Hawaii, and Puerto Rico (source: http://gapanalysis.usgs.gov/gaplandcover/data/), we hoped that this data could become a valuable resource for future research on urban biodiversity anywhere in the country. However, for now we recommend that more work be done to determine why some levels of development have the expected correlation with diversity, while others do not.
**LARIAC Tree Canopy**

While the LARIAC tree canopy data is not highly correlated with species richness or genera by the standard significance values (p<0.05 or p<0.01), it is certainly not completely uncorrelated. It is hard to say at this point why the Quickbird (McPherson) tree canopy is so much more highly correlated; while this still needs to be investigated, it is most likely due to the differences in resolution of the two datasets, different methods used for distinguishing tree canopy from other features, and/or the different times (possibly capturing seasonal differences) that the two tree canopy layers (or the satellite and lidar data they were derived from) were created.

**McPherson Land Cover**

The weak negative correlation between “Irrigated Grass” and species richness and genera may be attributed to the overall uniformity of a bed of grass, inhibiting a diversity of habitats for various insects. The weak positive correlation between “Dry Grass or Bare Soil” and species richness and genera may be due to the limitation of the 1 meter resolution, as areas classified as “Bare Soil” may actually contain small sprouts, growths, and sporadic flowering plants, which may provide habitat types for various insects. “Dry Grass” could also include decaying grass, which may also attract additional species of insects.

The strong positive correlation between tree canopy and species richness and genera as well as the strong negative correlation of “%Impervious/ Other” and species richness and genera indicates that these two factors are decisive predictors of urban insect biodiversity.

**WorldClim Mean Temperature and Precipitation**

There was a strong positive correlation between mean precipitation and species richness and genera, indicating that precipitation was an important factor in determining insect biodiversity in an urban area. This correlation can be explained by the effects precipitation on vegetation growth and plant diversity. The more humid an area is, the more plant growth there will be, resulting in an increase of plant diversity. Thus, precipitation affects urban insect biodiversity through the growth and development of plant species.

The strong negative correlation between mean temperature and species richness and genera can also be explained by the effect temperature has on vegetation growth and plant diversity. The hotter an area is, the less plant growth there will be, which results in the decrease of plant diversity. Temperature is a significant factor in limiting insect species richness and genera of an urban area.
**Landsat Temperature**

There was a fairly weak negative correlation between Landsat 8 (band 10) temperature data and species richness, as well as Landsat 8 (band 10) temperature data and genera. The Landsat 8 temperature data gave measurements of brightness, not actual temperature. The difference in statistical significance between this set of data and the data from WorldClim could be due to the difference in units used to measure temperature.

**NDVI**

NDVI measures chlorophyll content and subsequently plant productivity and vegetation density ([http://phenology.cr.usgs.gov/ndvi_foundation.php](http://phenology.cr.usgs.gov/ndvi_foundation.php)). The data analyzed indicates that this is correlated with species richness. Of the three NDVI variables analyzed (NDVI mean, NDVI standard deviation, and NDVI center), NDVI mean and NDVI center were both positively correlated with species richness. NDVI center showed the strongest correlation between the three variables. This indicates that the plant productivity within 30m x 30m of the malaise trap, the size of one pixel, was more strongly correlated with species richness than the mean plant productivity of the immediate surrounding area of approximately 90m x 90m. The high significance associated with the positive correlation and R-squared value of NDVI center indicates the importance of this metric in the multiple regression model. While NDVI mean was also shown to be correlated, the R-squared value and significance were relatively lower than NDVI center, but was still included in the multiple regression.

**Elevation**

There was a positive correlation between elevation (meters) and species richness, as well as elevation (meters) and genera. Additionally, the data for elevation and climate were significantly related, as well as elevation and tree canopy cover. These results were expected because as elevation increases, so does precipitation, which in turn increase vegetation cover.

**Night Light Intensity**

There was no statistically significant correlation between night light intensity (measured as three separate RGB values) and either species richness or genera. Because night lights are saturated in the Los Angeles metropolitan area, the VIIRS sensor was not able to differentiate between sites at a high enough resolution to be of any use. Another limitation may be that the separate RGB values (ex: [254, 250, 140]) had no intrinsic value as three comma separate values. Our statistical analysis treated each band as a separate value (i.e. r-value was calculated for red, green, and blue separately), and this methodology yielded no significant results. Potential future studies in night lights could investigate different methods of quantifying the intensity of light, but the Suomi NPP VIIRS sensor’s images provide georectified data only in this format. Even more so, the
biological mechanism which drives insects’ reactions to artificial lights in cities (ex: disrupting circadian rhythms) is still relatively unknown, and this metric will be interesting to investigate as the field of urban ecology becomes more understood.

**Mapping L.A. Biodiversity**

As a preview of the type of product that might be designed from this type of analysis, we’ve created a map of insect diversity for the entire city of Los Angeles (Figure 6). This map is based on a linear regression of species richness with McPherson tree canopy (in one-hectare plots centered on each BioSCAN site) as the independent variable, as this version of tree canopy had the highest R-squared value of any single metric we analyzed. A multiple regression model, such as the one we made - using NDVI, tree canopy, open space, impervious surfaces, temperature, and precipitation as explanatory variables - could be applied in a similar manner, but with potentially more accurate biodiversity predictions.

![Figure 5. A regression of percent canopy cover (McPherson) and species richness. The difference from Figure 4 is due to the difference in extraction methods for obtaining the McPherson data set.](image)

![Figure 6. A map of Los Angeles showing species per hectare, derived from data shown in Figure 5.](image)
Conclusion

The NHM BioSCAN program is one of the first of its kind - research and concrete data on biodiversity in urban environments is very scarce. As a result, the program is engaging in cutting-edge science with no set methodology as to how to collect and assess urban biodiversity metrics. Part of the goal of this paper was to test and develop ideal practices in analyzing the diversity and abundance data that BioSCAN has accumulated since 2012. For example, we looked at which distances from the malaise trap yield the most correlated results; this may assist future studies in determining what level of spatial resolution to prioritize when identifying and analyzing other parameters that might influence biodiversity. Techniques such as this, which can apply to both fieldwork and remote sensing techniques, will be applicable even as landscapes change over time. We also identified methods and metrics that were not effective (e.g. night lights as an RGB metric did not yield any interesting results), and some metrics that would be interesting to test in the future, but were not within the scope of our study (native vs. non-native plants, proximity to parks/wilderness, etc.).

Some limitations to our study included the extent of available data. For one, the BioSCAN program only has complete species and genera richness datasets for the phorid fly family, thus we only considered phorid fly diversity in our analysis. As a result, our models and conclusions are not applicable to other insects, and we suspect that urban diversity and abundances will vary widely in other families. The use of malaise traps presented another physical limitation to the study because they preferentially catch flying insects, and ground-dwelling species are generally excluded. We expect that total species diversity of ground-dwelling insects will greatly differ from flying insects because of their different dispersal mechanisms. Third, the placement of the traps within yards widely varied relative to barriers (proximity to buildings, fences, etc.) and other environmental features, which were difficult to account for in our analysis. Finally, some remote sensing images were not taken in recent years (e.g. Suomi NPP nightlights data was from 2012), therefore remote sensing data may not reflect current conditions that we noted in other metrics or in our fieldwork data. For instance, landscaping around homes - and around malaise traps - can change dramatically and quickly over time. These changes could have a real impact on the diversity of insects collected by the BioSCAN project, but they would not be captured in remote sensing images.

Acknowledgements

We would like to thank Travis Longcore for providing the LARIAC Tree Canopy Cover, and Brian Brown for providing the preliminary BioSCAN insect diversity data used in this study. We also thank Dean Pentcheff and Lisa Gonzales for coordinating site visits for collection of plant diversity data, and the 28 site hosts for allowing us to survey their vegetation. Finally, we would
also like to thank Tom Gillespie for providing the McPherson land cover data for the city of Los Angeles, and for his assistance and input through the entire project as our advisor.


WorldClim data available from worldclim.org (http://www.worldclim.org/current).


References


