Kernel Density Analysis of Traffic Incident Data

Patrick W. Perkins

GISC3200K Programming for GIS Fall 2020

Dr. Huidae Cho

University of North Georgia, IESA

UNG

UNIVERSITY of NORTH GEORGIA™

LEWIS F. ROGERS INSTITUTE FOR ENVIRONMENTAL AND SPATIAL ANALYSIS

Abstract: The purpose of this analysis is to simultaneously evaluate motor vehicle accident density and determine correlation point density to a road network, while developing a custom arcpy tool for motor vehicle accident density analysis. The resultant code will allow for ease of processing multiple point datasets delineated by 'Year', while producing a kernel density raster output raster for efficient analysis. In the data for this analysis, individual fire department emergency calls, or fire calls, represent each individual dispatch of emergency personnel for non-medical purposes or for medical emergency response. Specific call codes have been defined in this analysis to evaluate the frequency of non-medical and medical calls responding to motor vehicle accidents. For these MVA calls, it is necessary to evaluate the location and frequency of call data to determine if a correlation exists due to road network design. Call data ranging from 2010 to 2018 will be evaluated using a custom python tool for use in ESRI ArcGIS. The literature review portion of the Introduction will discuss the methodology The output displays a kernel density heatmap showing traffic accident density.

KEYWORDS: "**spatial analysis"**, **"kernel density**", "**motor vehicle accident**", "**Python**", "**Arcpy**".

1. Introduction

This analysis uses accident data to perform a kernel density analysis to evaluate traffic accident density in Forsyth County, Georgia. This analysis attempts to answer two questions: How can motor vehicle accident data be more efficiently analyzed, and where are the highest motor vehicle incident locations in Forsyth County, Georgia? With a heatmap kernel density analysis, accident density can be spatially interpolated to show clustering or random dispersion. Accident call data is extracted from all emergency response calls to produce a single ArcGIS feature class that displays the spatial location, call incident type, alarm dispatch date, and call by year fields. Kernel Density calculates the density of point features around each output raster cell, by calculating a magnitude per unit area from point or polyline features using a kernel function to fit a smoothly tapered surface to each point or polyline (Greibe, 2003). Kernel density estimation involves placing a symmetrical surface over each point, evaluating the distance from the point to a reference location based on a mathematical function, and summing the value of all the surfaces for that reference location. This procedure is repeated for all reference locations (Levine, 2010). The advantage of the kernel density method is its ability to represent multivariate relationships and to visualize kernel densities in 1D, 2D and 3D, of which 3D offers the greatest potential in the analysis of traffic violation data. In other words, if univariate kernel density (i.e., histogram) is used for analyzing the temporal distribution of traffic violations, the results would only show the time of the highest frequency of violations within a day. If binary variable kernel density (i.e., hot spot map) is used for analyzing the traffic violations' spatial distribution, the results only show the hot spot on a fixed surface (Silverman, 1986). For interpretive analysis of the resultant kernel density raster, a heatmap proves to be the proper analysis output. A heatmap is a data matrix visualizing values in the cells using a color gradient. This gives a good overview of the largest and smallest values in the matrix. Rows and/or columns of the matrix are often clustered so that users can interpret sets of rows or columns rather than individual ones (Haghighi, Cathy Liu, Zhang, & Porter, 2018).

MVA point data is processed by use of a Python script tool utilized in ESRI's ArcGIS software. The script performs this function by first importing computer-aided dispatch call history. This data contains call incident data that has been geolocated. Using this call incident feature layer, an analysis can be performed using the Kernel Density ArcGIS tool (Allen, David W., 2013). This tool creates a heatmap of traffic incidents can be created to show density by location, allowing for quick identification of statistically significant spatial clusters of high values (hot spots) and low values (cold spots). It creates a new Output Feature Class with a z-score, p-value, and confidence level bin for each feature in the Input Feature Class (Allen, David W.; Coffey, Jeffery M., 2011). Conceptually, a smoothly curved surface is fitted over each point. The surface value is highest at the location of the point and diminishes with increasing distance from the point, reaching zero at the Search radius distance from the point. Only a circular neighborhood is possible. The volume under the surface equals the Population field value for the point, or 1 if '**NONE**' is specified. The density at each output raster cell is calculated by adding the values of all the kernel surfaces where they overlay the raster cell center (Allen, David W., 2013). The kernel function is based on the quartic kernel function described by Silverman (Maldonado, Weber, & Famili, 2014).

Several models have been developed to analyze the cause of motor vehicle accidents at locations based on several factors, such as time of day, road type, road conditions, shoulder width, etc. For instance, Castro et al. (Castro, Paleti, & Bhat, 2012) studied the spatio-temporal incidence of accident counts at urban intersections. It proved to be advisable to consider both the spatial and the temporal effect, and a significant effect was found for roadway configuration, approach roadway typology and traffic flow, among other factors. Kamalov (Kamalov, 2020) also developed several modelling approaches to analyze accident occurrence at intersections. The consideration of a hierarchical spatial model accounting for the effects produced at intersections by contiguous segments (corridor-level) clearly outperformed the rest of the models applied. Maldonado et al (Maldonado, Weber, & Famili, 2014) analyzed accident counts at road intersections considering types of users (pedestrians, bicycles or motor vehicles) involved in accidents with a multivariate Poisson lognormal regression model. Moreover, Alvaro et al. (Á lvaro Briz-Redón, 2019) used a mixed effects negative binomial model accounting for macro-level and micro-level factors to study accident counts at road intersections. Several covariates constructed at both levels of spatial resolution were found to be associated with more accidents at intersections.

2. Materials

The analysis uses an installation of ESRI ArcGIS Pro with advanced license for the Spatial Analyst Extension. This project should be tested on several CAD tabular data sources to ensure proper functioning of the python script toolset.

- ArcGIS Pro v. 2.x with spatial analysis license
- Microsoft Excel
- Python v. 2.x/3.x software installation
- Data: All fire call data provided by Forsyth County, Georgia Fire Department and base data provided by Forsyth County GIS Department

3. Methods

Methods for performing this analysis will follow a step-by-step process to calculate kernel density of motor vehicle accident points. Data points were provided by the Forsyth County Fire Department and will be processed to show a heatmap for visual analysis. Figure 1, shown below, displays the Python script developed as an ArcGIS python tool, to automate kernel density analysis of MVA point data. The model shown by Figure 2 displays the workflow process graphically.

```
MVA Heatmap Development Tool Pseudocode:

Developed in Arc Python Toolbox:
        self.label = "Kernel Density MVA Toolbox"
        self.alias = "Kernel Density MVA Toolbox"

"This tool allows the user display motor vehicle accidents in a
kernel density raster.

    Def user input parameters by user input parameter
definitions.
        """Define parameter definitions"""

# Define Save Folder for Input Data
# Define Input GDB
# Define Forsyth County Boundary
# Define Template Layer File

# Set workspace extent to defined boundary

# Batch Import Data: Iterates and Imports all features from list
into defined Output_Data geodatabase

# Import Feature Layer Template: Feature Layer used to populate
all calls

# Create Data List for Fire Lists

# Make Feature Layer: This an all records feature layer to save
in intermediate data by type, using SQL selection:
     'IncidentTyp' IN ('322', '323', '324', '352')

# Split by Attribute and dump to Project Data
  Split function performs split by ['Year']

# List Split Data for MVA by Year
# Run KD on Split Data List
```

Figure 1. MVA Kernel Density pseudocode.

Below, Figure 2 displays a graphical representation of the script tool logic, developed in ArcGIS Pro Model Builder. Model Builder was not used to develop this script, instead it represents a basic logical flow path for script development. The python script features a modification of this logic.

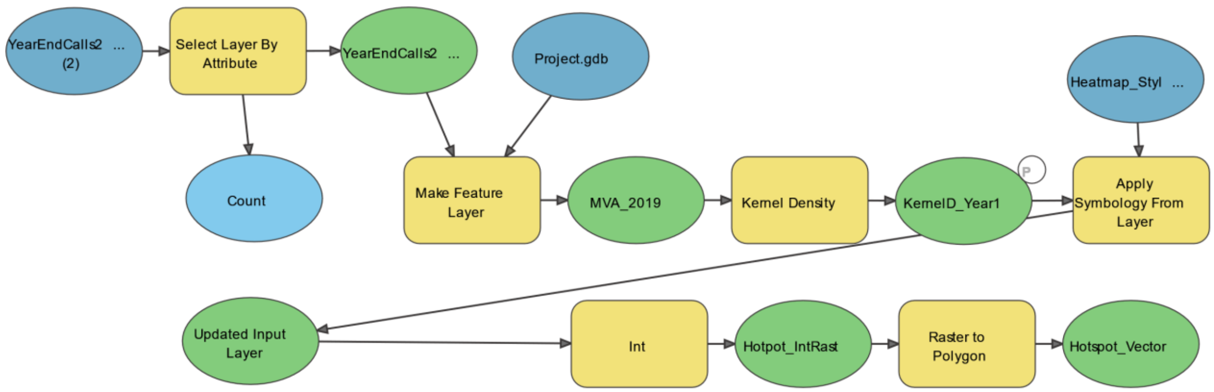

Figure 2. ArcGIS Model Builder Model Design for MVA Heatmap Analysis.

Methods Defined

1. The script will allow the user to define 4 parameter types, including the data folder workspace, the input folder geodatabase, and templates for boundary and out feature class.
2. The script will run the Create Geodatabase tool to create geodatabases that store temporary data in a defined workspace, input data to avoid manipulation of existing data, an output feature class geodatabase, and an output kernel density raster geodatabase for use in analysis.
3. The script performs a batch import of all feature class input data for all 911 fire department incident calls. In this study, that includes data through the years 2010-2018. The input environment is designed to allow addition or removal of sequential years of fire call data at the discretion of the analyst. The batch import performs iteration in the geodatabase to import all feature classes that exist in the geodatabase. This can be set using a wildcard '*' statement if other data should be present in the input geodatabase.
4. The tool will import a template layer used for appending all features. In doing so, it will automate a query selection of incident calls through a SQL expression: ['`IncidenTyp`']

`IN ('322', '323', '324', '352')`. These call values represent motor vehicle accidents, where ‘`322`’ is classified motor vehicle accident w/injuries, ‘`323`’ is classified as motor vehicle accident involving a pedestrian, ‘`324`’ is classified motor vehicle accident without injuries, and ‘`352`’ is classified extrication of victim from a vehicle. This tool outputs feature class layers and can display a count of records that match the selection query.

5. The script will perform a split list function by year to produce a year selection to reference in the kernel density tool.
6. Lastly, the ArcGIS arcpy tool ‘Kernel Density’ processes the split list stored in the project data geodatabase, by iterating through ‘Year’ field values and creating an output kernel density raster for each year input. An example of the kernel density mathematical formula is shown in Figure 3. Figure 5 will display an example of final output.

$$Density = \frac{1}{(radius)^2} \sum_{i=1}^{n} \left[\frac{3}{\pi} \cdot pop_i \left(1 - \left(\frac{dist_i}{radius} \right)^2 \right)^2 \right]$$

$$\text{For } dist_i < radius$$

Figure 3. Kernel Density as a mathematical formula.

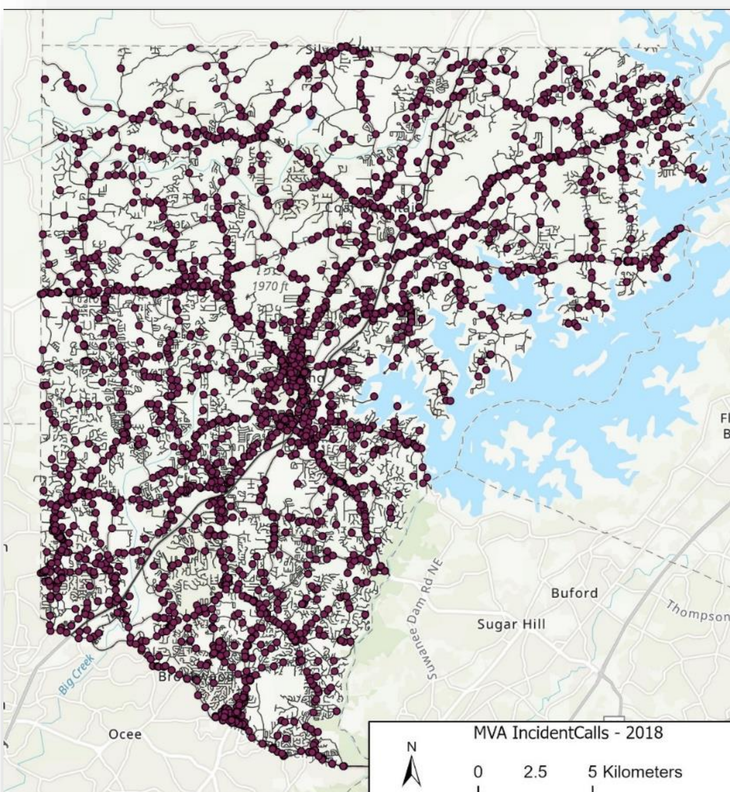

Figure 4. Example point density data for MVA calls. Shown: 2018.

4. Results

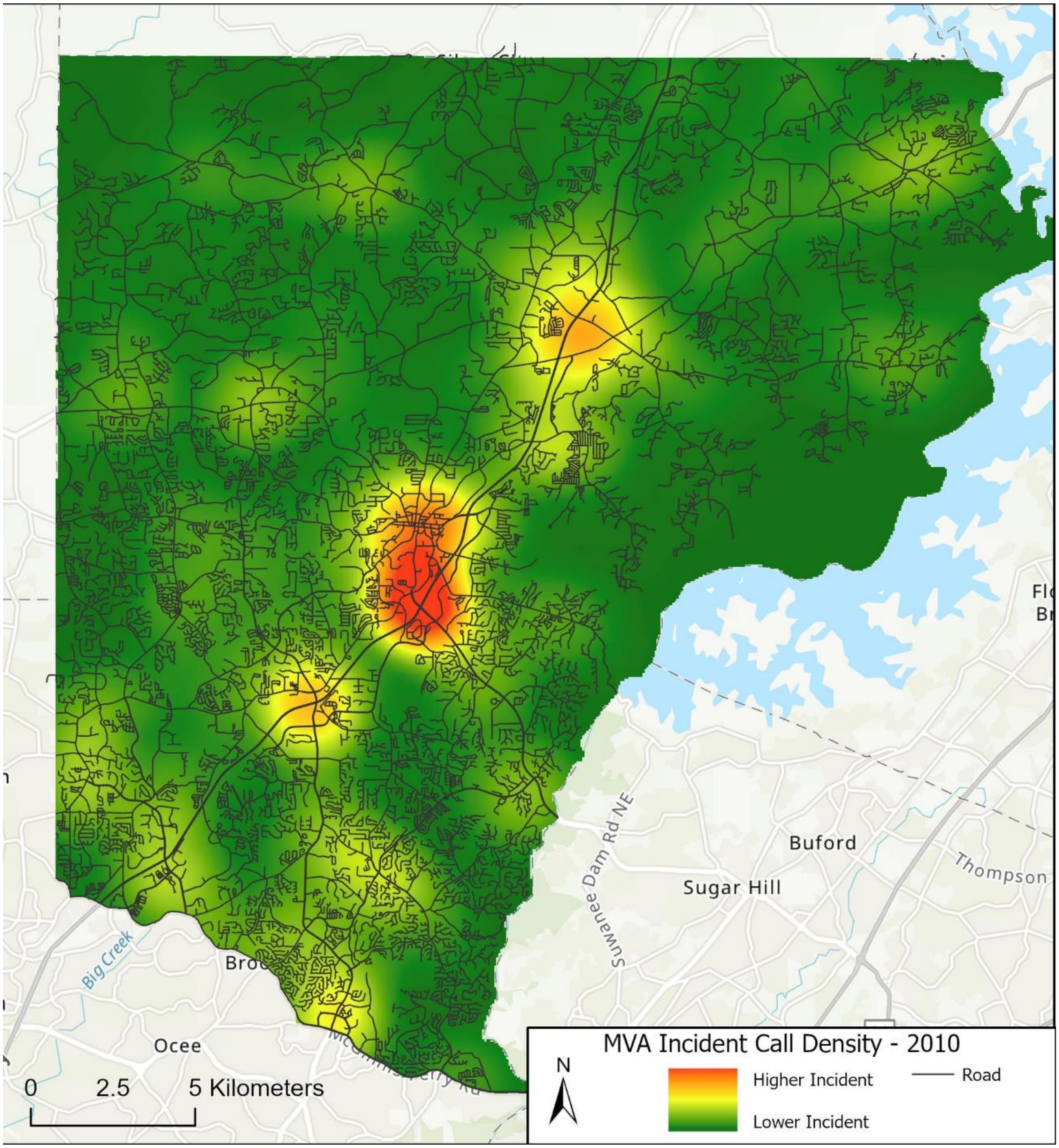

Figure 5. Kernel Density of motor vehicle accidents during 2010 in Forsyth County, GA. Shown with road centerlines.

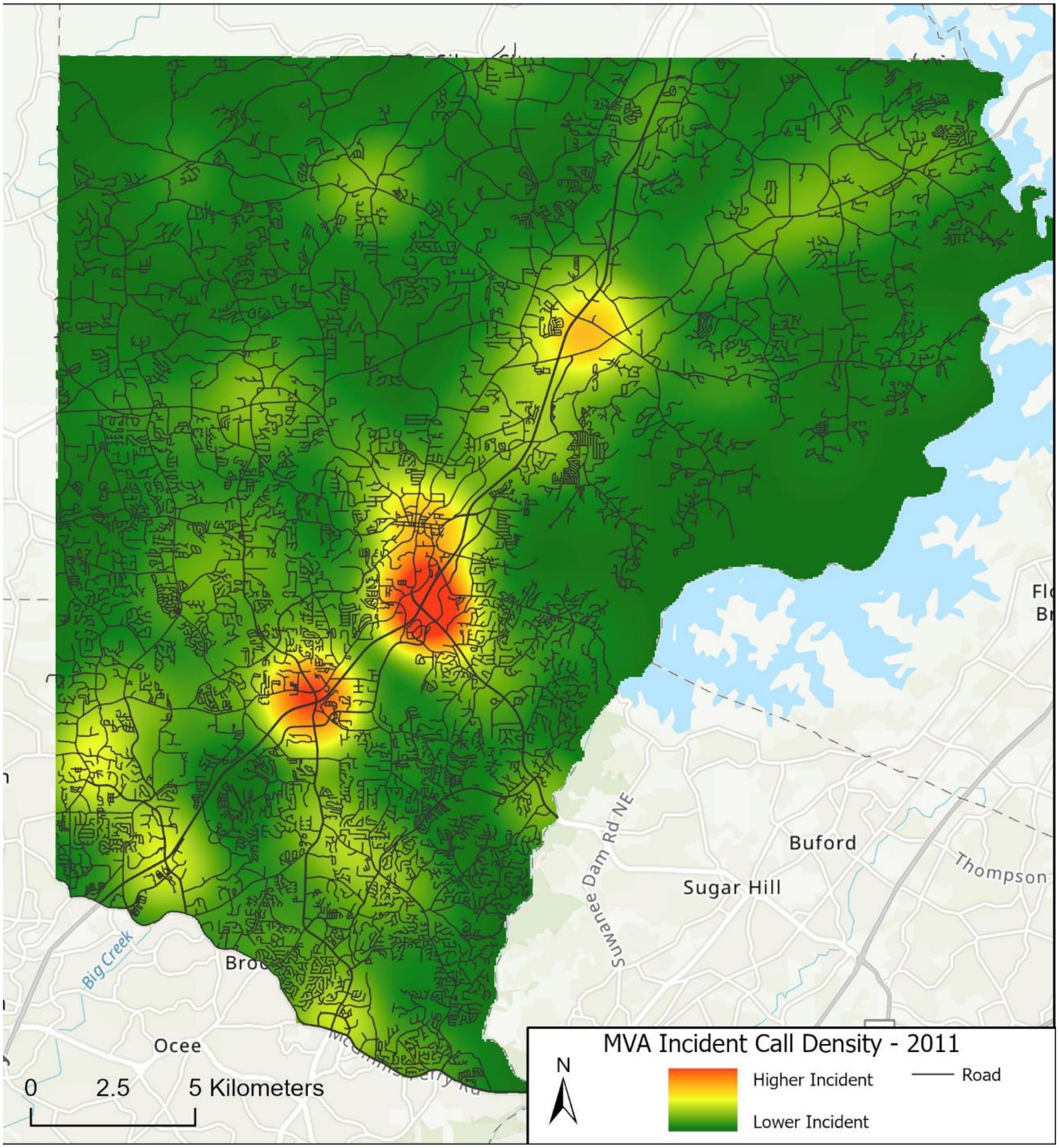

Figure 6. Kernel Density of motor vehicle accidents during 2011 in Forsyth County, GA. Shown with road centerlines.

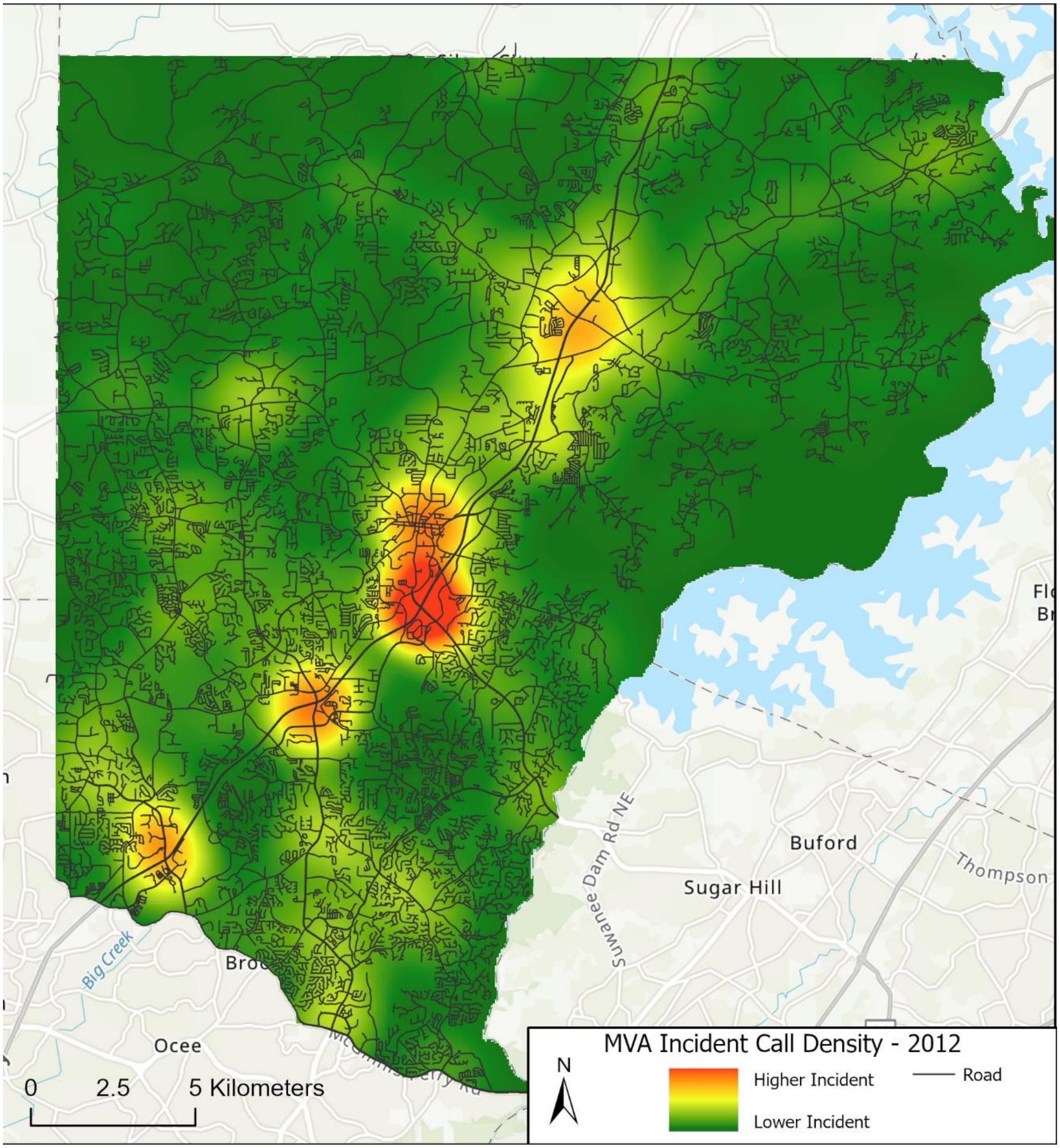

Figure 7. Kernel Density of motor vehicle accidents during 2012 in Forsyth County, GA. Shown with road centerlines.

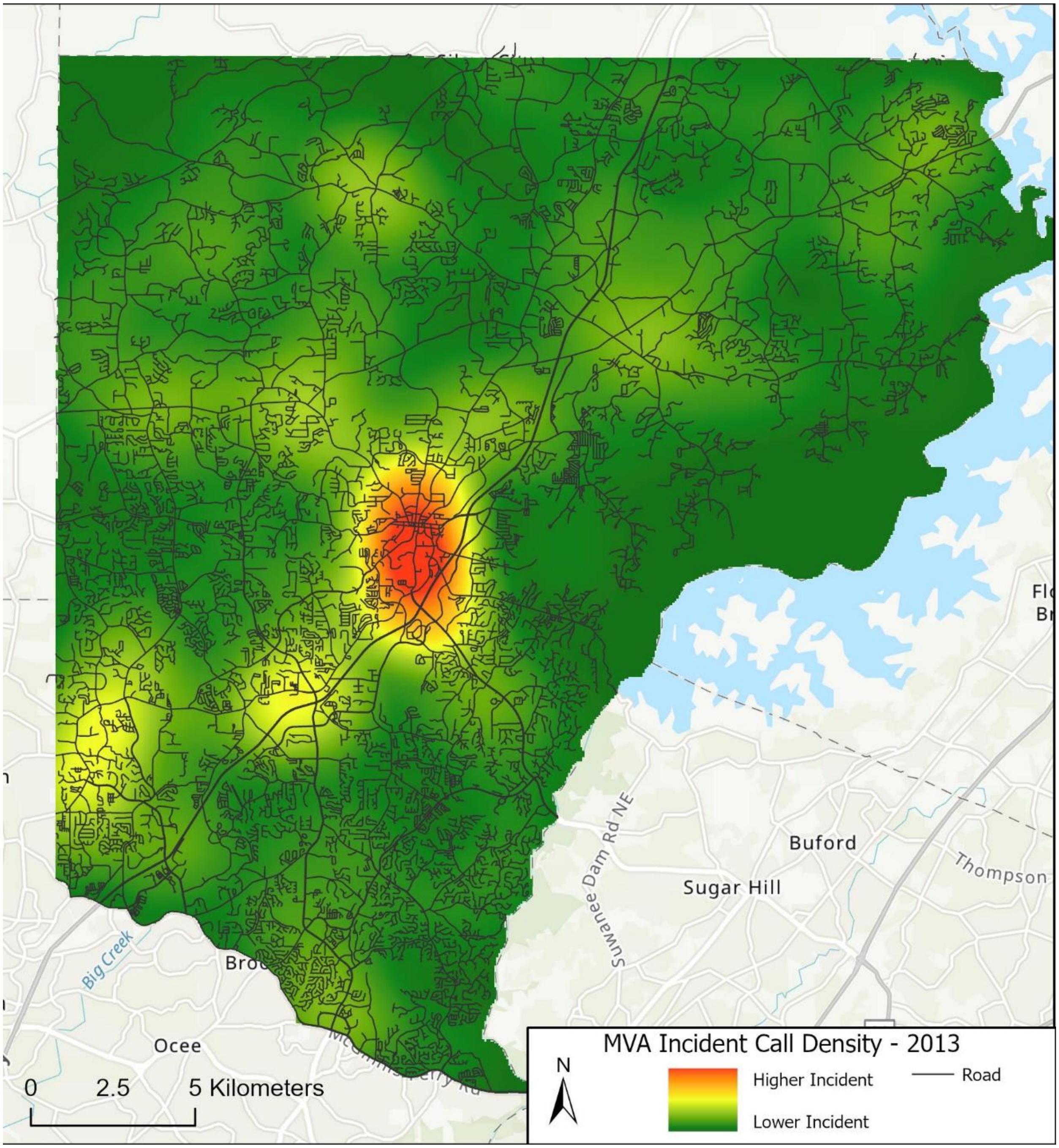

Figure 8. Kernel Density of motor vehicle accidents during 2013 in Forsyth County, GA. Shown with road centerlines.

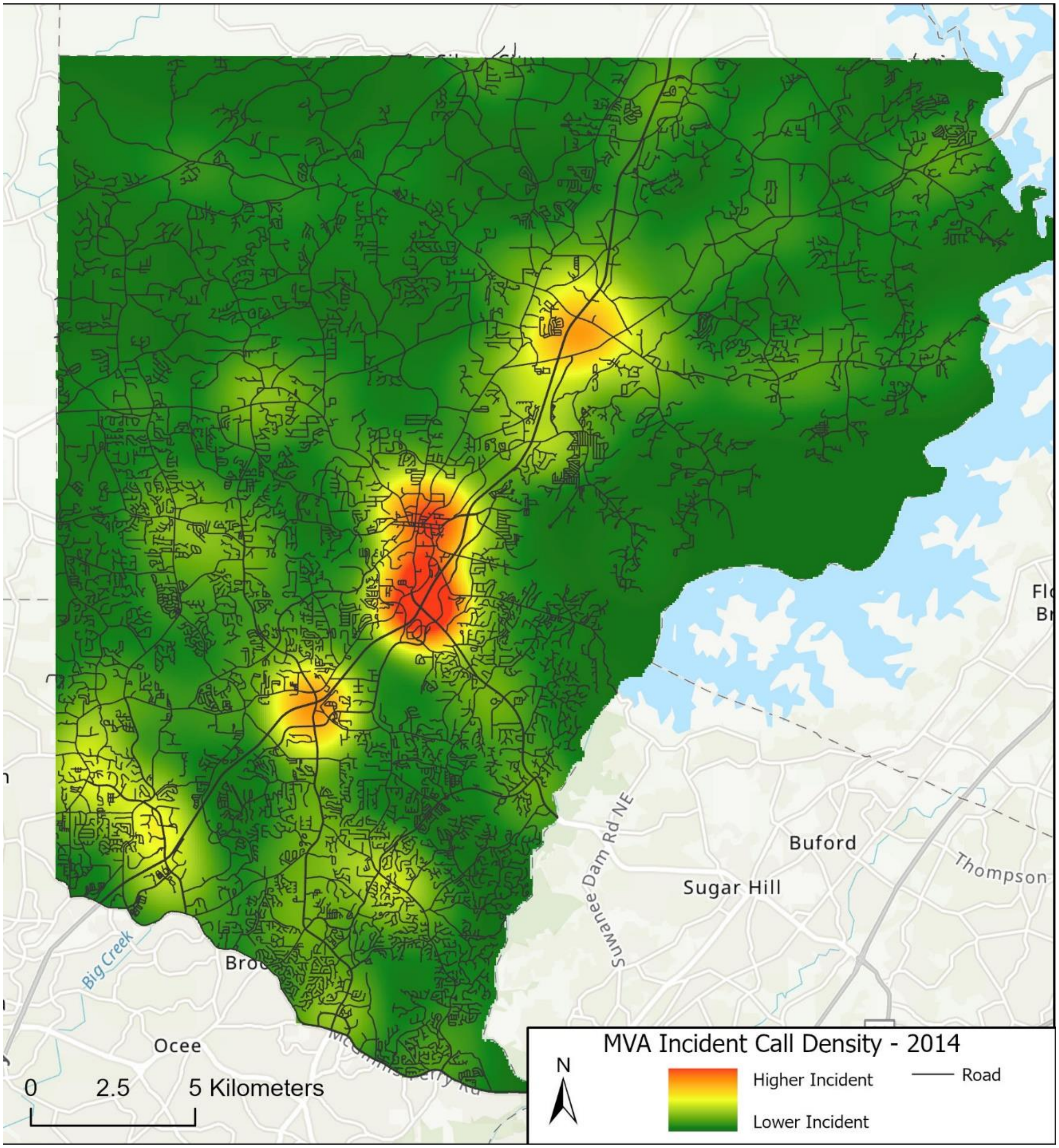

Figure 9. Kernel Density of motor vehicle accidents during 2014 in Forsyth County, GA. Shown with road centerlines.

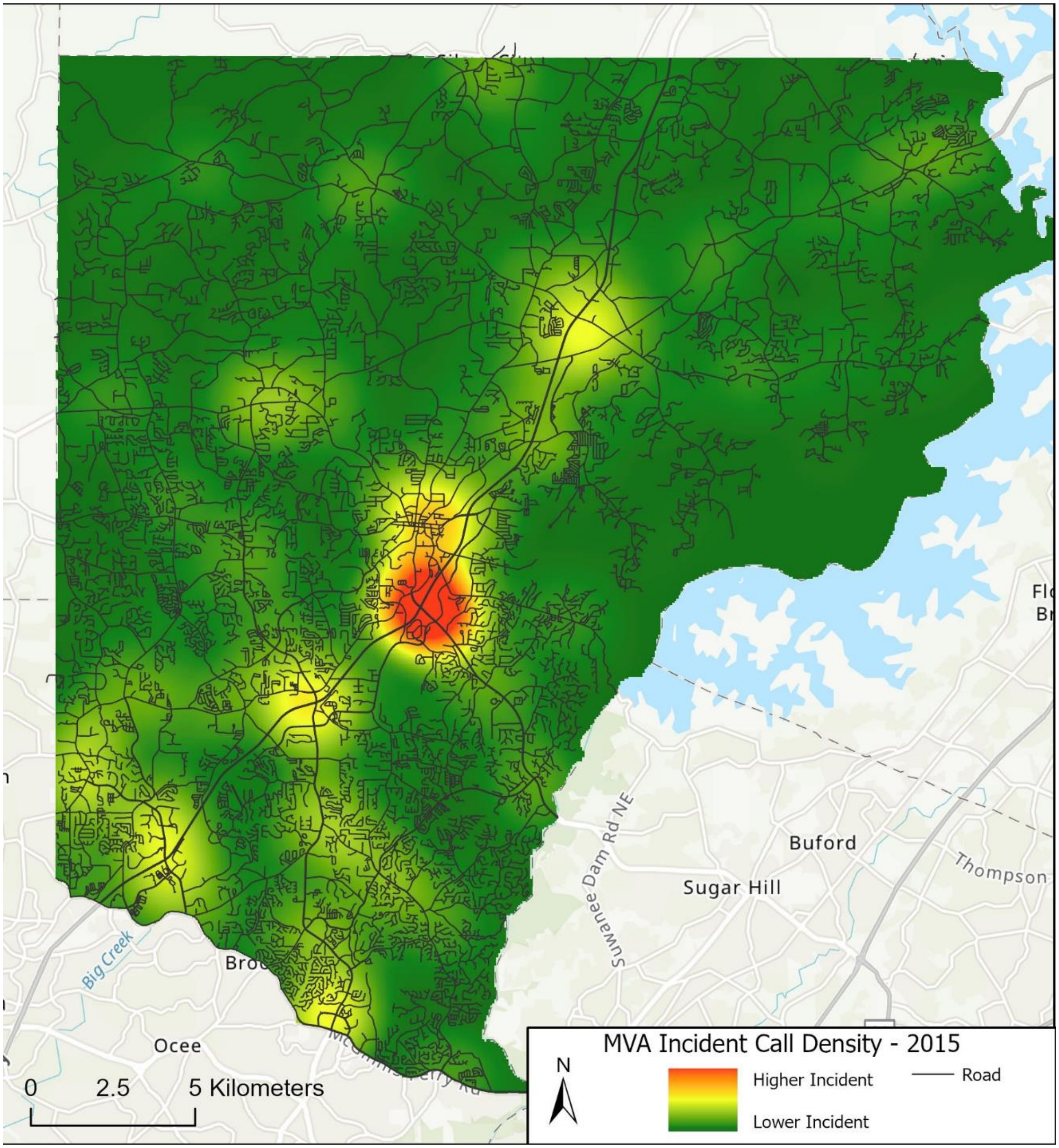

Figure 10. Kernel Density of motor vehicle accidents during 2015 in Forsyth County, GA. Shown with road centerlines.

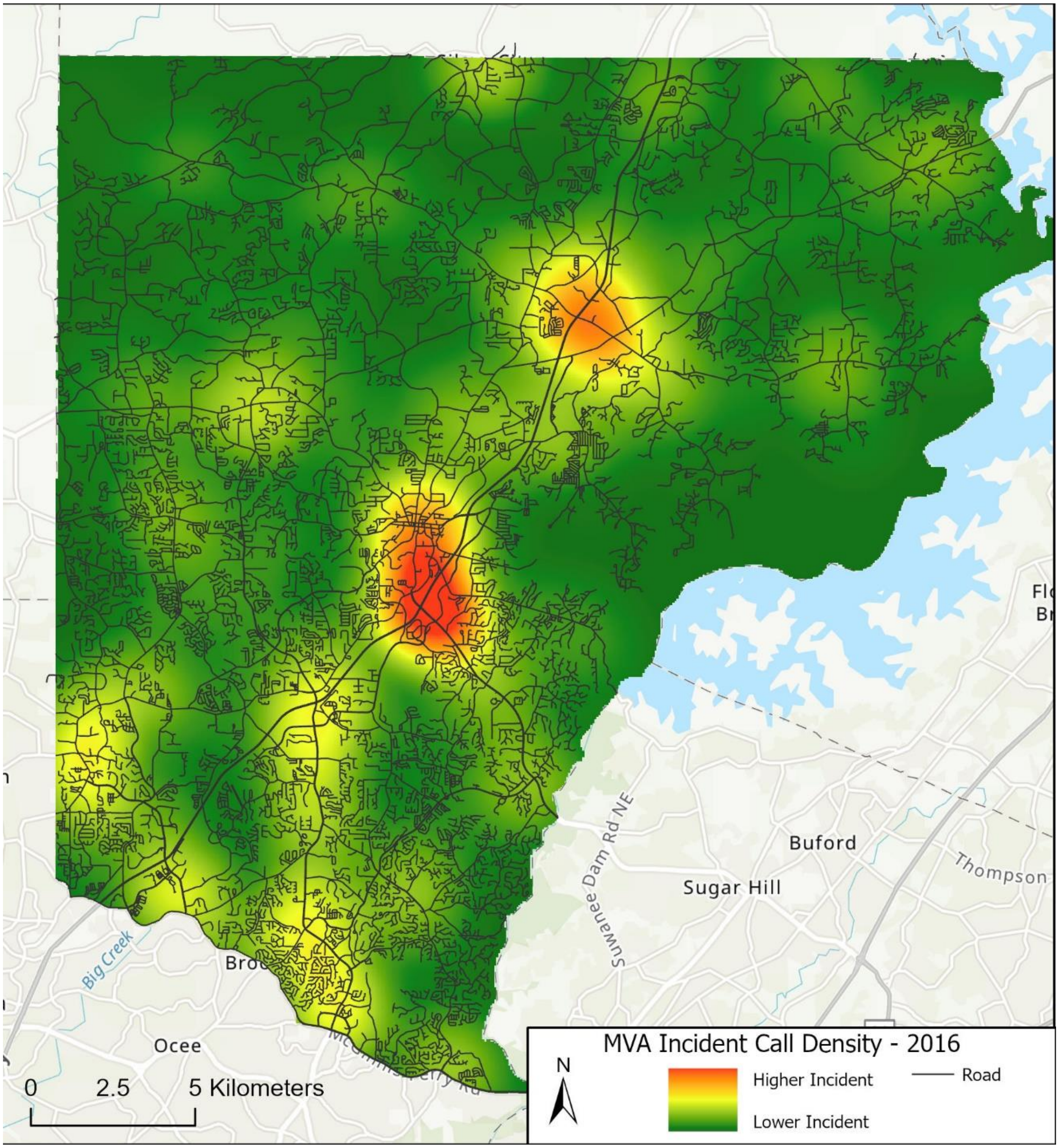

Figure 11. Kernel Density of motor vehicle accidents during 2016 in Forsyth County, GA. Shown with road centerlines.

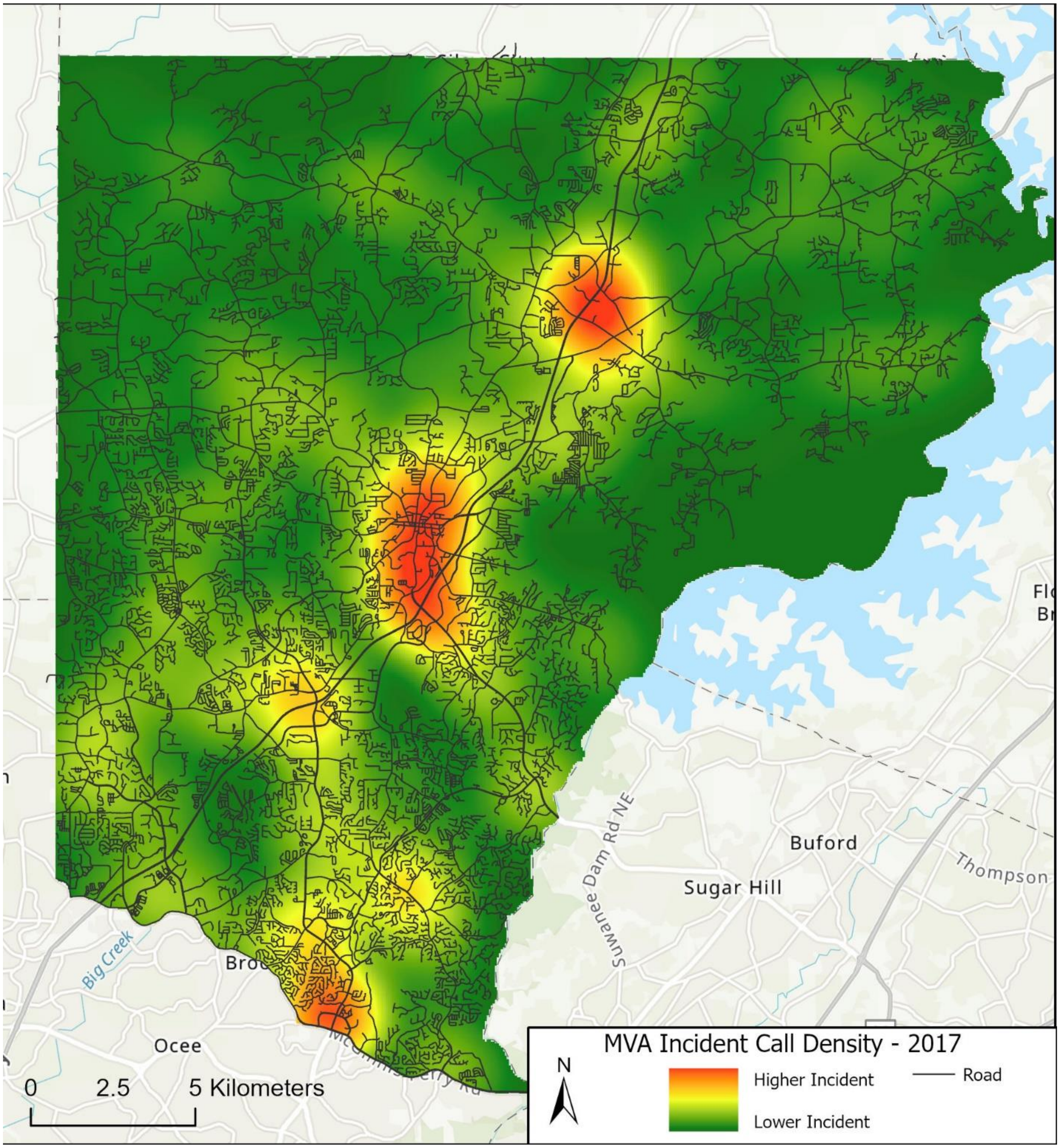

Figure 12. Kernel Density of motor vehicle accidents during 2017 in Forsyth County, GA. Shown with road centerlines.

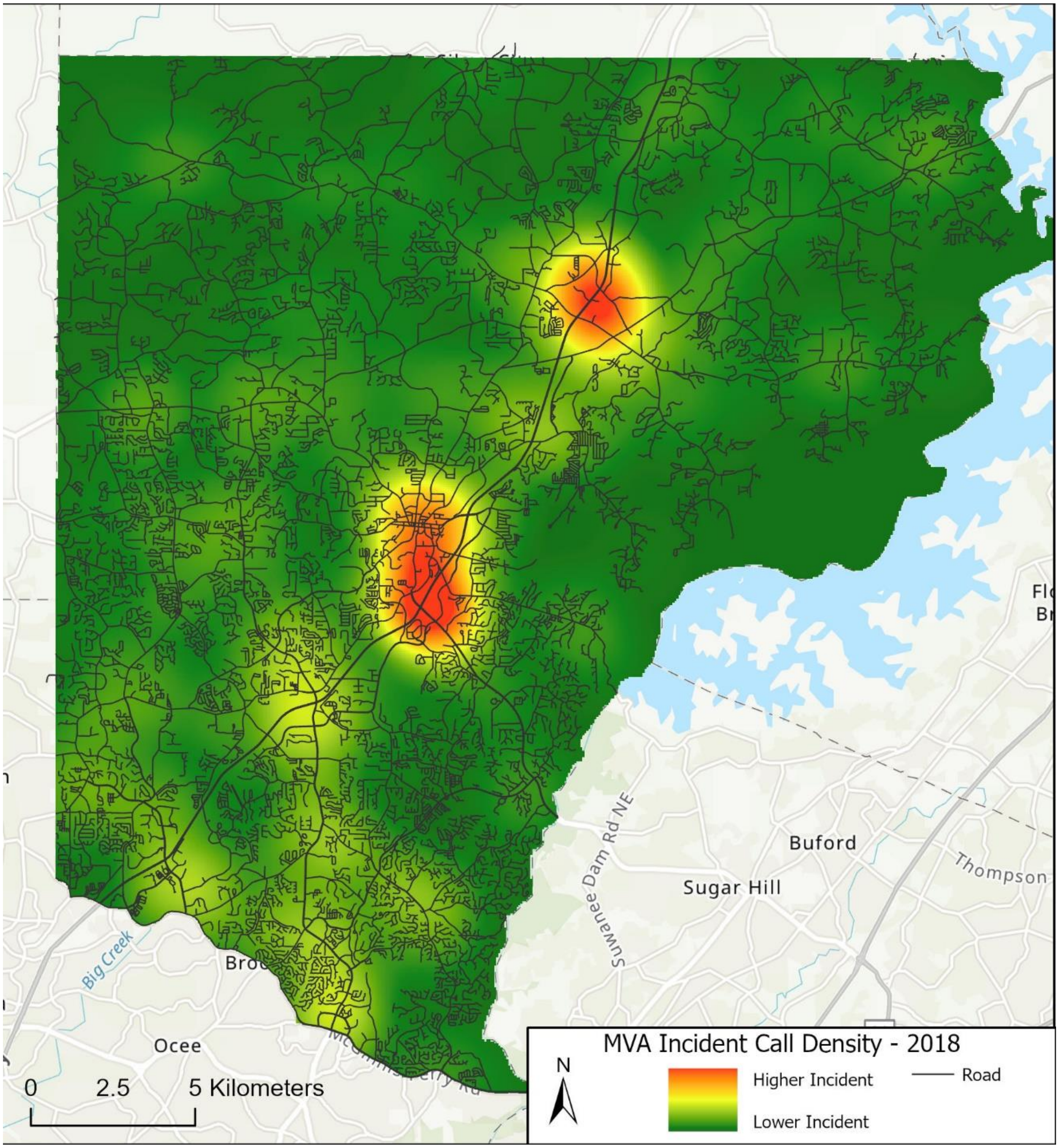

Figure 13. Kernel Density of motor vehicle accidents during 2018 in Forsyth County, GA. Shown with road centerlines.

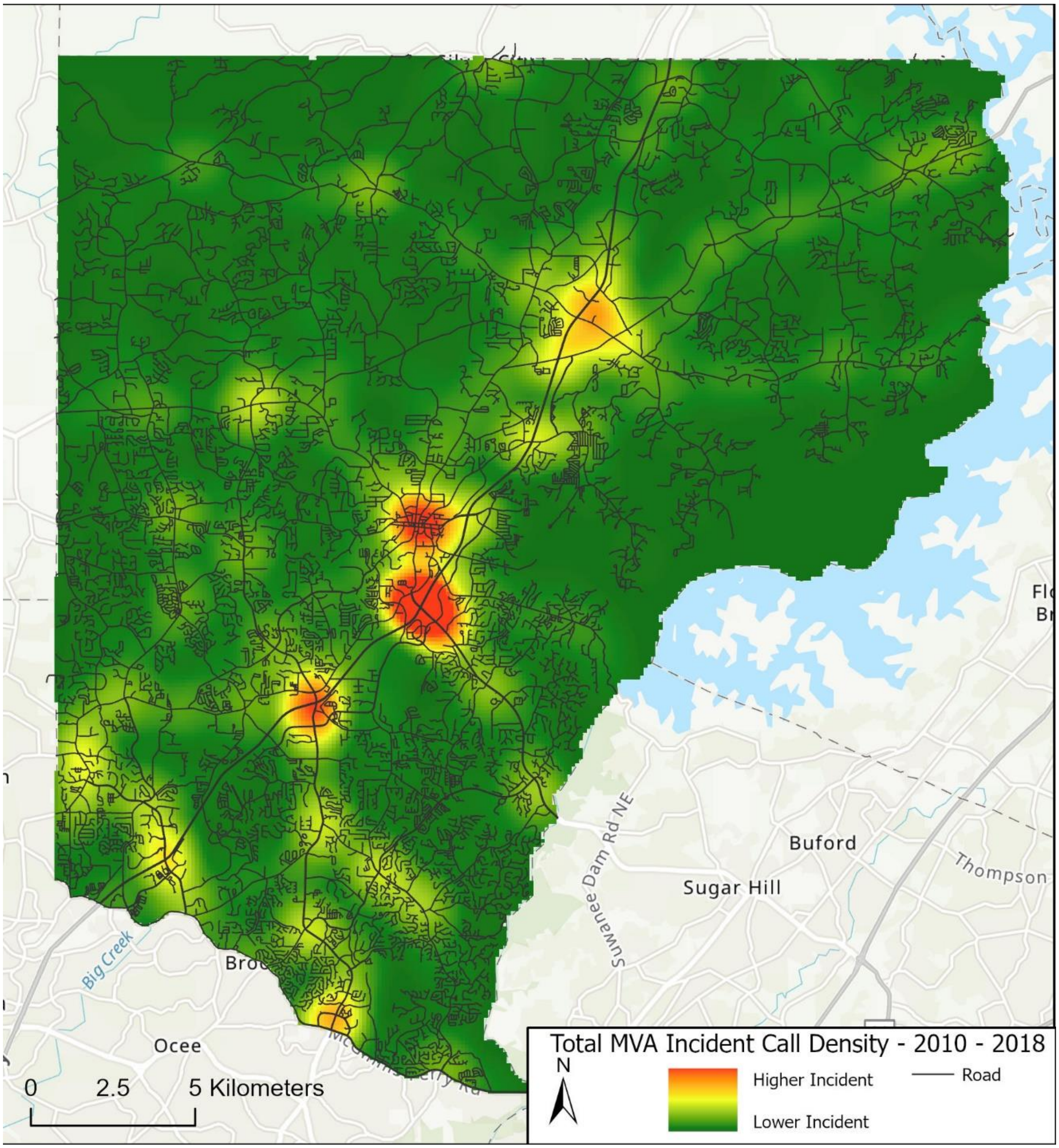

Figure 14. Total Kernel Density of motor vehicle accidents from 2010 - 2018 in Forsyth County, GA. Shown with road centerlines.

5. Discussion of Results

The objective results of this study, as seen above in Figures 5 - 14, show that the tool was successfully run and properly developed. Each year's output kernel density raster has road centerlines overlaid to help visualize where accident frequency occurs, and point density can be objectively located as it correlates to a road network, showing that the highest incident frequency occurs at major intersections. This analysis did not take into account accident frequency direct correlation to road class type, but from visual analysis, high incident point density directly correlates to intersections of major roads with other major roads, and dense ground features such as buildings, areas of high pedestrian occurrence such as shopping and municipal buildings, and with one-way streets. In particular, high incidence areas occurred in the downtown region of Cumming, Georgia where pedestrian foot-traffic is frequent. Other areas of high incidence, such as the major thoroughfare in the county, motor vehicle accident occurrence may be attributed to high-speed vehicle traffic along this route. Weather condition data could be evaluated for data of alarm dispatch. For the subjective portion of this analysis, the correlation between road class can be made easily as road class can be defined as local road, collector highway, minor arterial, major arterial/multi-lane/state highway, and closed-access freeway. In this study, road class type is not defined and a continued analysis where road class is defined would show those results. The intent of this study was not to necessarily establish the statistics of motor vehicle accidents by their location, but to establish an efficient and capable tool for beginning analysis.

One specific focus of the analysis is the temporal attribute of the data. Call data can be evaluated for call dispatch time. A separate analysis can show correlation of accident frequency by time of day, which may have more meaningful results than a stand-alone kernel density heatmap. From Lu, et al, 'the results show that the STKDE space-time cube made it easier to detect the spatio-temporal patterns of traffic violations than did the traditional hotspots map. An interesting finding was that traffic sign violations and traffic marking violations were primarily concentrated not in regular peak hours, but during the time period of 14:00-16:00, which indicates that these intersections were the most congested during this period' (Lu, Cheng, Yuan, Mohamed, & Li, 2020). Evaluation of MVA call locations and their frequency in a location is crucial for infrastructure planning as well. Fire station locations should be placed according to call frequency. These objectives are equalizing coverage area and minimizing total travel time. The solution to this dilemma typically involves making equipment and personnel allocations that are compromises between the two extremes. By placing more companies in high hazard and demand areas, yet still assuring that all regions receive an acceptable level of coverage, communities are finding the balance between these two objectives (Barr & Caputo, 1997).

6. Conclusion

In summary, kernel density analysis proved to be a useful tool for performing spatial analysis of point data. The project's goal was to develop an efficient means of analyzing point data through a python-based script tool. The development of the ESRI Python Arcpy script tool proved to be successful, as the tool was intended to function. After defining a workspace data folder, all temporary processing geodatabases were created correctly, followed by batch file import, append with SQL query, split list function, and exported kernel density output raster files by respective year. Interpreting data results using a heatmap raster is not necessarily intended to be scientific but is more of a quick reference tool for informed decision making. An end-user's objective to perform objective analysis of motor vehicle accident density was achieved. The scripted tool allows quick visual analysis of MVA point density as it correlates to a road network. It should be noted that analysis of a single kernel density raster product is insufficient for interpolating results. Statistics are not interpolated from this python tool. Therefore, continued analysis must be performed to develop objective results. This is possible with continued script development, model builder environment, or by use of step-by-step geoprocessing tools in the ESRI Arc environment. Several methods could be pursued to obtain statistical results, such as the addition of polygon zone features (e.g.: by each fire station service area) and point density analysis using these zones as processing boundaries. While it is important to note that a vast majority of motor vehicle accidents will occur at some point along an established road network, the intent is to find clustering/high incidence to ensure that proper planning for intersection upgrades.

References

Allen, David W. (2013). *GIS Tutorial 2: Spatial Analysis Workbook.* Redlands, CA: ESRI Press.

Allen, David W.; Coffey, Jeffery M. (2011). *GIS Tutorial 3: Advanced Workbook.* Redlands: ESRI Press.

Á lvaro Briz-Redón, F. M.-R. (2019, November). Spatial analysis of traffic accidents near and between road intersections in a directed linear network. *Accident Analysis & Prevention*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0001457519306098?via%3Dihub

Barr, R. C., & Caputo, A. P. (1997). *Planning Fire Station Locations* (18th ed.). Quincy, MA: National Fire Protection Association.

Castro, M., Paleti, R., & Bhat, C. R. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B: Methodological, 46*(1), 253-272. Retrieved from https://www.sciencedirect.com/science/article/pii/S0191261511001366

Greibe, P. (2003). Accident prediction models for urban roads. *Accident Analysis & Prevention*, 273-285. Retrieved from http://www.sciencedirect.com/science/article/pii/S0001457502000052

Haghighi, N., Cathy Liu, X., Zhang, G., & Porter, R. J. (2018). Impact of roadway geometric features on crash severity on rural two-lane highways. *Accident Analysis & Prevention, 111*, 34-42. Retrieved from http://www.sciencedirect.com/science/article/pii/S0001457517304013

Kamalov, F. (2020, February). Kernel density estimation based sampling for imbalanced class distribution. *Information Sciences, 512*, 1192-1201. Retrieved from https://www.sciencedirect.com/science/article/pii/S0020025519309740

Law, Michael; Collins, Amy. (2013). *Getting To Know ArcGIS* (3rd ed.). Redlands, CA: ESRI Press.

Levine, N. (2010). *CrimeStat III: Spatial Modeling.* Retrieved from CrimeStat: https://www.icpsr.umich.edu/CrimeStat/files/CrimeStatChapter.8.pdf

Lu, J., Cheng, Z., Yuan, J., Mohamed, A.-A., & Li, Y. (2020). Analyzing traffic violation behavior at urban intersections: A spatio-temporal kernel density estimation approach using automated enforcement system data. *Accident Analysis & Prevention, 141*. Retrieved from http://www.sciencedirect.com/science/article/pii/S000145751931485X

Maldonado, S., Weber, R., & Famili, F. (2014). Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines. *Information Sciences*, 228-246. Retrieved from https://doi.org/10.1016/j.ins.2014.07.015

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* New York: Chapman and Hall.

Acknowledgements to Dr. Huidae Cho, UNG/IESA, and Forsyth County Fire Department. Special thanks to Dr. Cho for continued support, and to Forsyth County Fire for providing a data source to perform this study.