THE GEOGRAPHY OF CLOUD-TO-GROUND LIGHTNING IN YELLOWSTONE NATIONAL PARK

by

Ed Amrhein

A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Geography and Geoinformation Science

Committee:

_____	Dr. Paul Delamater, Thesis Director
_____	Dr. Timothy Leslie, Committee Member
_____	Dr. Paul Houser, Committee Member
_____	Dr. Anthony Stefanidis, Department Chair
_____	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
_____	Dr. Peggy Agouris, Dean, College of Science
Date: _____	Fall Semester 2016 George Mason University Fairfax, VA

The Geography of Cloud-To-Ground Lightning in Yellowstone National Park

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Ed Amrhein
Master of Science
Texas A&M University, 2001
Bachelor of Science
Texas A&M University, 1999

Director: Paul Delamater, Professor
Department of Geography and Geoinformation Science

Fall Semester 2016
George Mason University
Fairfax, VA

DEDICATION

To the Spirit of Yellowstone

ACKNOWLEDGEMENTS

Very few things in life can be accomplished alone, especially a task such as a Master's thesis. This thesis could not have been accomplished without the love, help, support and encouragement of several people.

First I would like to express my deepest gratitude to my thesis advisor Dr. Paul Delamater for his willingness and eagerness in mentoring me during this process. He showed a deep and genuine interest in my research and was a great source of motivation. His leadership and extensive knowledge contributed greatly to the success of this research.

I would also like to express my gratitude to my committee members Dr. Timothy Leslie and Dr. Paul Houser for their support and instrumental comments which enhanced this research. I'm also thankful for Melissa Hayes who helped me navigate the extension process and obtain the approvals for me to continue and complete my work.

I am also indebted to my fellow Airman at the U.S. Air Force's 14th Weather Squadron and to Vaisala for graciously providing the lightning data used to complete this research.

I am also grateful to have an extensive team of coworkers and close friends who were extremely supportive during this process: Rita, Mike, Chris, Leanna, Jason, Beth, Neal, Tim, and Brian.

Furthermore, I could not have accomplished this goal without the love and encouragement of my parents and my daughter Tara who sacrificed a lot of family time so that I may continue my studies and complete this research.

Finally, I owe a special thanks to my beautiful bride Cyndy who has faithfully stood by me through 27 years of marriage that included 24 years of faithful service to our country while in the Air Force, one Bachelor's degree, and now two Masters degrees. She's always been extremely supportive allowing me to successfully reach my goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF EQUATIONS

Equation **Page**

LIST OF ABBREVIATIONS

Cloud-To-Ground ..... CG
Cloud-To-Ground Lightning Density ..... Ng
Comma Separated Values ..... CSV
Complete Spatial Randomness ..... CSR
Coordinated Universal Time ..... UTC
Geographical Information System ..... GIS
Kiloamperes ..... kA
Kilohertz ..... kHz
Kilometer ..... km
Lightning Location System ..... LLS
Local Standard Time ..... LST
Microseconds ..... ms
Mountain Standard Time ..... MST
National Aeronautics and Space Administration ..... NASA
National Geospatial Intelligence Agency ..... NGA
National Land Cover Dataset ..... NLCD
National Lightning Detection Network ..... NLDN
Ordinary Least Squares ..... OLS
Shuttle Radar Topography Mission ..... SRTM
Time Of Arrival ..... TOA
United States Geological Survey ..... USGS
Yellowstone National Park ..... YNP

ABSTRACT

THE GEOGRAPHY OF CLOUD-TO-GROUND LIGHTNING IN YELLOWSTONE NATIONAL PARK

Ed Amrhein, M.S.

George Mason University, 2016

Thesis Director: Dr. Paul Delamater

Yellowstone National Park is well known for its vivid and diverse landscape, its abundance of wildlife, and its wildfires such as the 1988 “Summer of Fire”. Yellowstone is also well known for its volcanic activity-the cause of several geothermal hot spots that cover the landscape in the form of geysers, hot springs, mud pots, and fumaroles. In this study we learn that Yellowstone has other hot spots that are also important to the ecology of the park: clusters of cloud-to-ground (CG) lightning flashes.

With the use of the latest GIS technology available, a 10-year National Lightning Detection Network (NLDN) dataset from 1995 to 2004 was analyzed to better understand the spatial and temporal pattern of CG lightning in Yellowstone. Graphs and maps visualizing lightning strikes and flash density reveal the seasonal and diurnal behavior of CG lightning in the park. Global spatial statistics reveal the spatial pattern of CG lightning is more of a random pattern while local spatial statistics indicate CG lightning is

locally clustered. Maps visualizing results from local spatial statistics show hot spots of CG lightning activity over the mountain regions of the park and cold spots of CG lightning activity over the western and central plateau of Yellowstone. Finally, spatial regression analysis using the physical terrain properties of slope, aspect, elevation, and land cover resulted in statistically significant models that at most explained 17 percent of the variability in CG lightning flash density. Of the variables tested, only elevation appears to have a statistically significant relationship with observed spatial pattern of lightning flash density.

1. INTRODUCTION

The summer of 2016 marked the 28th anniversary of Yellowstone's "Summer of Fire" (as well as the 100th anniversary of the U.S. National Park Service). During the summer of 1988, there were 51 fires with 42 ignited by lightning (National Park Service, 2007). The first major fire, the Storm Creek fire, began on June 14 while the last fire, the North Fork fire, started 22 July; all fires were declared out on 18 November (Franke, 2000). These fires scorched about 1.4 million acres in the Greater Yellowstone Area, burning more than 36% of the park (793,880 acres burned within the park's boundaries) making it the worst fire in the history of the park (Franke, 2000). Approximately 300 large mammals perished in the fires and 67 physical structures were destroyed including 18 backcountry patrol cabins. The total property damage was estimated at $3 million. More than 25,000 people and $120 million in logistical support were required to suppress these fires (National Park Service, 2007; Franke, 2000). While lightning provided the spark for most of the fires in the Greater Yellowstone Area, it was the combination of the fire management policy and extreme weather conditions that helped shape an environment conducive to wildfires ignited by cloud-to-ground lightning.

The massive and highly destructive wildfires experienced during Yellowstone's "Summer of Fire" highlighted the importance and benefits of fire in maintaining a healthy forest and revealed significant shortfalls in the forest fire management policy.

Since 1988, fire management policy has evolved from that of extinguishing all forest fires immediately upon ignition to an approach that allows small, naturally occurring fires to burn out with limited human intervention. Forest fire management policy continues to evolve as scientists learn more about the forest, the environment and the role of fire. The more recent plan includes factors such as the fire ignition type, the succession state of the forest, weather, location, and availability of fire suppression resources (National Park Service, 2014, 2004).

This research project focuses on analyzing the geography of cloud-to-ground lightning as a step toward a better understanding of lightning-ignited wildfires in Yellowstone National Park. A better understanding of lightning-ignited wildfires begins with an examination of the spatial and temporal lightning patterns within the park in order to better understand their relationship with the physical environment. Lightning plays a role in the ecology of the park, while at the same time presents a hazard to the physical structures used to operate and manage the park as well as to the safety of its visitors. A better understanding of the geography of lightning in the park can enable park managers to refine the fire management program in a way that improves the park's ability to "allow fire to continue to play its ecological role in the park while protecting human life, developments, and sensitive cultural and natural resources" (National Park Service, 2014, 2004). Park managers would also be able to use lightning geography to point out lightning high risk areas to visitors so they can safely experience the park's natural resources.

A better understanding of Yellowstone's lightning geography can be obtained by leveraging geographic information systems and science (GIS). The latest advancements in GIS have provided methods and techniques to perform a more robust spatial analysis. We can now conduct point pattern or cluster analysis and surface analysis with relative ease within a GIS framework. We are also able to more readily apply various geospatial statistics-based methods that have been used in ecology for several years now. This research employs GIS and spatial analysis techniques to assess the spatial patterns, relationships, and trends in cloud-to-ground lightning. The overall goal of this research is to advance the understanding of the geography of cloud-to-ground lightning in Yellowstone National Park through the use of a spatial analytic framework. This goal is accomplished by answering the following research questions:

1. **Does cloud-to-ground lightning occur randomly across space or does it manifest in an observable spatial pattern?** *Hypothesis: The physical properties of lightning coupled with the topography of Yellowstone National Park interact to create a spatially clustered cloud-to-ground lightning pattern.*
2. **Is the spatial pattern of cloud-to-ground lightning consistent throughout the year or does it vary by month or season?** *Hypothesis: The overall spatial pattern of cloud-to-ground lightning is not consistent throughout the year due to seasonal changes in regional and local scale weather patterns.*
3. **Are specific areas or regions of Yellowstone Park more prone to experience cloud-to-ground lightning than others?** *Hypothesis: The spatial pattern of cloud-to-ground lightning manifests as localized clusters of high and low lightning activity within the park.*
4. **Is cloud-to-ground lightning activity associated with Yellowstone Park's topography?** *Hypothesis: The flash density pattern of cloud-to-ground lightning activity is related with features, such as land cover, and terrain variables (i.e. elevation, slope, and aspect).*

Knowledge about lightning cluster locations may help park managers enhance their fire monitoring programs by allowing them to determine where additional resources

and monitoring may be needed. Park managers may also use the information gained from this study in contingency planning for fire related events. Knowledge of potential problem areas due to increased levels of lightning activity could also be used to determine locations where park managers should consider pre-positioning firefighting equipment.

Mapping Yellowstone's cloud-to-ground lightning activity may also provide benefits for those who visit the park. These maps could be used by park rangers as part of the park's public safety education program or by hikers when planning day hikes and back country trips. Park rangers can better inform visitors about the dangers of lightning and the areas of the park where increased awareness of the local weather conditions is warranted. Information from this study could be incorporated into the mandatory bear and animal safety briefing back country campers receive before they obtain the required back country permit at the ranger's office.

This thesis is organized as follows. Chapter two presents information on the cloud-to-ground lightning process and the basics of the lightning detection and measurement system used in the United States (US). A review of the previous lightning and wildfire related research as well as the use of GIS in studying this phenomenon is also presented in Chapter two. Chapter three describes the study area while Chapter four discusses the datasets used in the research. Chapter five contains the methods used to conduct the analysis. The results and discussion are presented in Chapter six. Chapter seven contains potential future research and Chapter eight presents a summary of the research and findings.

2. LITERATURE REVIEW

2.1 The Lightning Flash

Lightning is defined as a transient, high current electric discharge resulting from the separation of charged particles within a cloud, usually a thunderstorm cloud or cumulonimbus cloud (Uman, 1987). The total lightning discharge, as an entity, is called a lightning "flash". A flash that makes contact with an object on the ground is often called a "lightning strike" and defines cloud-to-ground lightning. The flash is a complex process involving one or more flows of current. In cloud-to-ground discharges this current flow is called a "stroke" (Telesca, Bernardi and Rovellie, 2005; Rakov and Uman, 2003). Each stroke is composed of a downward leader from the cloud followed by a return stroke from the ground. A flash can contain a single stroke or multiple strokes and the number of strokes in a flash is known as the multiplicity. The time scale of the lightning flash process is on the order of a half second while the strokes have a duration on the order of tens of microseconds (ms) that typically occur within 20 to 100 ms of each other (Orville et al., 2011; Telesca, Bernardi and Rovelli, 2005). The first downward leader of the flash is started as a "stepped" leader while the subsequent downward leaders of the flash are started as "dart" leaders (Rakov and Uman, 2003). The stepped leader-return stroke creates the initial channel for the movement of charge. Cloud-to-ground flashes can transfer either negative or positive electrical charge although negative flashes are considered more common occurring, 90% of the time (Rakov and Uman, 2003). The percentage of flashes transferring positive charge to the ground have been found to vary both seasonally and regionally (Orville and Huffines, 2001).

The return stroke is sometimes followed by a continuing current. While most charge is transferred during the sequence of leader-return strokes, there are cases where most of the charge is transferred as a continuing current (Orville et al., 2011). In one study, 75% of positive cloud-to-ground lightning flashes contained a continuing current while only 30% of negative cloud-to-ground flashes had a continuing current (Saba et al., 2010). The presence of continuing current within a flash is considered a critical factor for wildfire ignition (Latham and Williams, 2001). Positive flashes often contain a single return stroke lasting more than 40 ms, with the greatest peak currents producing the largest charge transfer to ground. Negative flashes rarely contain continuing currents longer than 40 ms with peak currents greater than 20 kiloamperes (kA). Some evidence suggests that the combination of the long continuing and high peak current in the positive flashes are responsible for increased physical damage and wildfire ignition (Rudolsky and Fuelberg, 2011; Rakov, 2003; Rakov and Uman, 2003; Latham and Williams, 2001).

2.2 Lightning Detection and Measurement

Only within the last 40 years have detailed studies on the spatial distribution of lightning events in the U.S. been conducted and published (Zajac and Rutledge, 2001). This was largely due to the lack of detailed lightning data until the implementation of the National Lightning Detection Network (NLDN). The NLDN is a modern day lightning location system (LLS) that is able to determine location, intensity, and movement of thunderstorms in real time (Cummins and Murphy, 2009). An LLS employs a networked array of broadband (~1 to 400 kilohertz [kHz]) sensors that detect the electromagnetic fields radiated by lightning flashes (Krider et al., 1980). It is the distinct electric and

magnetic waveforms produced by the return strokes of cloud-to-ground lightning that makes detection possible (Cummins et al., 2010). The LLS obtains time of arrival (TOA) information from the distinct waveform generated by the fast-rising "onset" or "rise-time" of these fields from the lightning discharges and uses the ratio of the peak field values measured by a pair of north-south and east-west (orthogonal) magnetic-loop antennas to obtain the direction of arrival (Cummins et al., 2010; Krider et al., 1980). The NLDN uses the Improved Accuracy through Combined Technology (IMPACT) algorithm to derive the latitude, longitude, and discharge time of the cloud-to-ground lightning based on the directional and arrival-time information (Cummins and Murphy, 2009). The NLDN has been providing lightning data at the continental scale in the U.S. since 1989 (Cummins and Murphy, 2009). The flash detection efficiency of the NLDN has been found to be between 90-95% with positional errors (median error) less than 500 meters (m) (Mallick et al., 2014; Nag et al., 2014; Cummins et al., 2006).

2.3 Lightning Studies

Data collected by the NLDN, as well as data collected by comparable systems in other countries, have been used in a variety of studies analyzing lightning behavior. Most studies use the collected lightning data to analyze lightning activity at the national or country scale. The analyzed data is often presented in the form of maps of flash density. One of the first papers published using the NLDN data at the national scale mapped the annual lightning flash density for the contiguous U.S. for 1989 (Orville, 1991). As the archive of NLDN data for the U.S has grown researchers have been able to map the flash density using longer periods of records to build national level lightning climatologies

(Orville and Huffines, 2001, 1999; Huffines and Orville, 1999; Lyons, Uliasz, Nelson, 1998; Orville and Silver, 1997). These national climatologies often show the mean annual flash density across the country as well as other parameters such as mean annual positive flash density, mean annual negative flash density, percentages of positive flashes, mean multiplicities for both positive and negative flashes, median peak currents, and mean annual flash rates (Orville and Huffines, 2001; Huffines and Orville, 1999).

While these lightning climatology studies conducted at the national scale are useful for identifying the overall spatial distribution of lightning activity across the area of study, they do little to address the relationships that may exist between the distribution of lightning flashes and the local physical environment. These studies may mask important relationships between the physical environment and lightning activity at the regional and local levels. In the U.S., studies conducted at the national level resulted in lightning flash density maps showing no apparent relationship between lightning strikes and the terrain (Orville and Huffines, 2001; Zajac and Rutledge, 2001). The lightning maps in these studies show the highest concentration of lightning in Florida and other parts of the southeast—regions that are relatively flat compared to other parts of the country.

More research on lightning at local scales is necessary in order to gain a better understanding on how lightning flashes are related to land cover and terrain variables such as elevation, slope, and aspect. In recent years, studies conducted at regional scales have produced mixed results. A lightning study of the Northern Territory of Australia found noticeable differences between lightning strike density and vegetation type and

terrain height (Kilinc and Beringer, 2007). In this study, strike density appeared to be more related to elevation in the southern region while more dependent on vegetation cover in northern region of the study area. Weber et al. (2003) found a negative relationship between elevation and lightning frequency in the foothills of mountains in the Snake River Plains of southeastern Idaho. In Colorado, Hodanish and Wolyn (2006) found the highest flash densities occurred where the mountains and the plains intersect instead of at the peaks. In southeast Pennsylvania, DeCaria and Babij (2003) found a weak negative but statistically significant relationship between lightning density and terrain height.

Research conducted on lightning strikes in the interior of Alaska revealed a dual relationship between lightning density and elevation (Dissing and Verbya, 2003). The relationship appears to be positive up to a maximum elevation of 1100-1200 m. However, above this elevation, the relationship is negative. Regression analysis further revealed that vegetation and elevation characteristics accounted for approximately 66% of the variation seen in lightning strike data.

Reap (1986) found a pronounced magnitude of lightning activity with increase in terrain height while studying lightning activity in the Western U.S. during the 1983-84 Summer seasons. The study also revealed lightning activity minimums between 0700 and 0900 Mountain Standard Time (MST) and rapid increases in lightning activity between 1500 and 1700 MST. In addition to the diurnal activity, the study found a high correlation between time of day and elevation. Maximum lightning activity in the higher elevations occurred in the early afternoon hours while the early evening and nighttime hours

appeared to favor the lower elevations. Reap (1986) concluded that the geographical and seasonal distributions of lightning activity appear to be the result of local topography, timing, and location.

The NLDN not only made it possible to study both lightning location and timing, but it has also made it possible to analyze some of the physical characteristics about cloud-to-ground lightning such as strike polarity, peak current, and multiplicity. A study by Saraiva et al. (2008) reveals relatively low multiplicity (flashes per strike) for negatively charged strikes in Tucson, AZ. Lyons et al. (1998) found significant geographic differences in multiplicity with respect to both polarity and peak current. The authors found higher concentrations of cloud-to-ground lightning strikes with positive charge and high peak currents in the Midwestern high plains region and high concentrations of lightning strikes with negative charge and large peak currents over the Gulf of Mexico and along the Gulf Coast states.

2.4 Lightning Ignited Wildfire Studies

The spatial distribution of wildfire is often determined by a variety of structural and environmental factors. Vegetation, topography, and land use are structural factors that help shape the spatial distribution of wildland fires (Vadrevu, Badarinath, Anuradha, 2008; Chou, 1992). Environmental factors include wind speed and direction, precipitation, and temperature (Westerling, 2003; Chou, 1992). In combination, these factors help determine the fuel characteristics of the area which is key in the spatial distribution of wildland fires (Amatulli, Perez-Cabello, de la Riva, 2007). While fuel characteristics are important at the point of ignition, they are more important in the

surrounding regions. The spatial distribution of wildland fires is greatly impacted by neighborhood effects and those from adjacent areas (Chou, 1992). The surrounding neighborhoods or adjacent areas influence the spread of wildfires as well as the level of fire suppression applied (Vadrevu, Badarinath, Anuradha, 2008; Chou, 1992;).

Although these are important factors in the general category of wildfires, lightning ignited wildfires have other contributing factors. Lightning-caused fires are also dependent upon the meteorological conditions that lead to atmospheric convection including both large-scale synoptic and regional scale climatological conditions that vary each year (Bartlein et. al., 2008). Fires ignited by lightning are also generally collocated with fuels and high lightning frequencies (Rorig and Ferguson, 1999) and depend on fuel moisture conditions, atmospheric moisture conditions, concurrent rainfall amounts and duration, and fire suppression efforts (Rorig and Ferguson, 2002).

Due to these relationships, a number of studies have found that fires caused by lightning are spatially clustered (see Amatulli, Perez-Cabello, de la Riva, 2007). Some have concluded that lightning caused fires are the most spatially clustered of all wildland fire types (Genton et. al., 2006). In mountainous regions, lightning-caused fires appear to be more clustered at higher elevations (Bartlein, 2008; Vazquez and Moreno, 1998).

In addition to demonstrating spatial clustering, lightning-ignited wildfires also appear to be clustered in time, manifesting as a distinct fire season. For example, wildfires caused by lightning in the western U.S. are extensive in summer and have spatial and temporal structures that reflect controlling factors such as climate, weather, and regional distributions of vegetation (Bartlein et. al., 2008; Westerling et. al., 2003). In

addition, this work showed that lightning ignited fires display a definite seasonal distribution with annual variations. Seasonally, lightning fires peak in the July-August timeframe as vegetation is more flammable due to deficits in soil moisture. Furthermore, the location of the largest lightning-caused fires changes from year to year which reflect the inter-annual variation in climate (Bartlein et. al., 2008). Likewise, Vazquez and Moreno (1998) showed that the pattern of lightning fires coincides with the local weather patterns. This research found that fires are clustered more towards mid-year and are greatly affected by vegetation. Fire frequency has also shown to be related to the growing season temperature; precipitation, lightning strike frequency, elevation, aspect, and level of forest cover (Kasischke, Williams and Barry, 2002).

In studying the physical properties of lightning strikes and their relationship to wildfires, Fuquay (1982, 1980) found that a relationship exists between flash polarity and fire ignition. Meisner (1993) correlated fire ignition with total flash counts. However, Rorig and Ferguson (1999) found no relationship between polarity and fire ignition by lightning.

The role of fire in the ecosystem of national forests and parklands is known and it forms the basis of fire management plans. Forest fires are important issues from both a social perspective as well as a scientific perspective (Lee, Park and Chung, 2006). Both the National Park Service and U.S. Forest Service have to constantly balance the needs of the ecosystem with the needs of the people. Finding the balance point that enables successful planning and execution of any fire management plan requires considerable knowledge about wildland fires. Knowledge regarding the spatial and temporal patterns

of all types of wildfire can be effective in shaping a fire management plan (Chou, 1992) as it can lead to a better understanding of the fire regimes (Lee, Park and Chung, 2006). Fire regimes are used to study forest fires in terms of fire type, intensity, severity, size, return interval, seasonality, and spatial pattern (Agee 1993; Christensen and Abbott, 1989). This type of knowledge can be obtained through the application of spatial analysis and GIS as well as spatio-temporal techniques. The results of spatial pattern analysis can provide managers with improved information used in determining how to enhance fire management policies (Vadrevu, Badarinath and Anuradha, 2007).

2.5 GIS Applications

A GIS is commonly defined as an integrated computer system that can capture, store, query, analyze, and display geographical information. The most well-known integrated system in use today is the ArcGIS platform, developed by the Environmental Systems Research Institute, Inc (ESRI) located in Redlands, CA. ESRI (2012) defines a GIS as a system with integrated hardware, software, and data allowing users to "view, understand, question, interpret, and visualize" data to reveal spatial relationships, patterns, and trends through maps, globes, reports, and charts. GIS allows users to analyze spatial and temporal patterns to identify, describe, and measure the shape, arrangement, location, configuration, trend, and relationships of spatial data (Scott, 2015).

As GIS technology evolves, more disciplines are leveraging its power. GIS has proven to be a great asset in fire research and park management (van Wagtendonk, 2002). Spatial analysis techniques have also been proven to be extremely valuable in analyzing

other incident data, including crime analysis (Chainey and Ratcliffe, 2005; Hirschfield and Bowers, 2001) and ecology (Johnston, 1998). Some of the techniques developed for research in ecology are now used in other environmental fields like forestry (Chuvieco and Salas,1996; Jaiswal et al., 2002; Kasischke, Williams and Barry 2002). Spatial pattern analytic approaches were shown to be very useful in detecting fire "hotspots" using a number of cluster analysis methods often employed in crime analysis such as quadrat analysis, nearest neighbor analysis, and the Ripley's K-statistic (Vadrevu, Badarinath and Anuradha, 2008).

Quadrat analysis evaluates point distributions by determining the change in density over space. It can help determine whether or not clustering of points is more or less than in a random pattern (Vadrevu, Badarinath and Anuradha, 2008). The nearest neighbor analysis gives a single value index that provides a more of a quantitative measure of a pattern. This index is simply the observed mean nearest neighbor distance divided by the expected mean nearest neighbor distance of a random pattern (Vadrevu, Badarinath and Anuradha, 2008). Ripley's K-statistic can be used to describe the degree of spatial dependence in the pattern and to identify significance of clustering (Vadrevu, Badarinath and Anuradha, 2008).

Another technique used to study forest fires is density estimation. Density estimation is the modeling of a density surface given a finite amount of data points recorded with x-y coordinates. The most common method simply counts the number of points within predefined aerial units of space using units of the same size. However, one drawback is that the resulting density surface can be affected by grid size and grid

orientation (Amatulli, Perez-Cabello, and de la Riva, 2007). Kernel density, which is a non-parametric technique, has several adjustable parameters such as the kernel function used, the bandwidth, and the amount of smoothing. Amatulli, Perez-Cabello, and de la Riva (2007) focused on tuning the bandwidth based on the nearest neighbors to convert fire point data into a continuous density surface to investigate wildfire occurrence in Aragón, Spain. This adaptive technique resulted in the application of a different bandwidth for each point. Wierzchowski, Heathcott and Flannigan (2002) also used density maps to analyze lightning-caused fires. Density maps were created for three grid cell resolutions: 1 kilometer (km), 5 km, and 10 km. The study found the 10 km resolution to show the most pronounced spatial patterns.

Most research that has examined the spatial and temporal distribution of cloud-to-ground lightning activity are based on the use of flash density maps. Average annual flash density maps are often generated along with monthly and seasonal flash density maps. While the nature of atmospheric science studies has always required the use of spatial thinking, the field itself has been slow to adopt the GIS framework for conducting analysis. Only over the previous ~10 years has the field moved to research using other spatial techniques to analyze lightning activity. Wagner et al. (2006) used a GIS based approach to lightning studies to create a series of flash density maps for studying lightning activity in West Texas. Vogt (2014) leveraged a GIS framework in studying the summertime lightning patterns on the Colorado "Fourteeners" (peaks ≥ 14,000 feet [ft]). In this study GIS was used to perform buffer operations to extract lightning data based on the terrain features of interest. Vogt (2011) used a GIS workflow to derive the local high

points from a digital elevation model within a study area in Colorado. These local high points fell along peaks, ridges, and steep convex slopes. A buffer operation was then applied to these local high spots to select lightning events that fell within the buffer allowing the author to focus on lightning-high point interaction. In Neuwirth, Spitzer, and Prinz (2012), the authors used a GIS to generate a series of maps that allowed them to determine high and low cluster regions based on flash density. They first generated the standard flash density map as seen in many other studies. These maps were then subjected to a 3 x 3 moving window to calculate focal mean to create a focal flash density. The focal flash density maps where then used to visualize “hot” spots, regions of high density and “cold” spots, regions of low density.

Applying spatial analysis techniques within a GIS framework can help determine the variables that help shape the environment conducive for lightning ignited wildfires within Yellowstone National Park. Evaluating the spatial distribution of these lightning wildfire events can highlight areas within the park where variables such as topography and vegetation play a more favorable role as suggested by other similar studies. Knowledge about the temporal distribution of these events can also help identify factors favoring lightning ignition such as environmental factors. The use of GIS and spatial analysis has the potential to not only reaffirm previous knowledge but also may lead to the discovery of other variables that are important in this wildfire process. More importantly is what happens to the knowledge gained from these analytical techniques. Knowing the environment is only the first step in managing the process. These same

techniques can also be applied in the second step to help make the decisions needed to effectively manage wildfires in the park.

3. STUDY AREA

Yellowstone National Park was established in 1872 as the first national park in the U.S. and is often proclaimed as the world's first national park. The park occupies 3,472 square miles and is primarily located in the extreme northwestern part of Wyoming (96% of the park's area) but extends westward into Idaho (1%) and northward into Montana (3%) (Figure 1). It is home to a variety of wildlife including grizzly and black bear, elk, wolves, and bison. The park is considered one of the most thermally active regions in the world containing several types of geothermal features, hot spots in the form of geyser basins, hot springs, mud pots, and fumaroles. Yellowstone contains approximately 1,000 miles of hiking trails with 300 backcountry campsites. Over four million people visited the park in 2015 (National Park Service, 2016).

Yellowstone's terrain is very rugged (Figure 2 and Figure 3) with elevation ranging from a low of 5,282 ft at Reese Creek to a high of 11,385 ft at Eagle Peak. More than 90% of Yellowstone National Park is above 7,000 ft in elevation as it is located on top of a thermal bulge created through the uplift of the Yellowstone hotspot, an area of long-lived volcanism. Most of the park's volcanic central plateau has an elevation between 8,000 and 9,000 ft. Flanking this plateau on three sides are mountainous regions; the Absaroka Range to the east, the Gallatin Range to the northwest, and the Beartooth Range to the north and northeast. The low lying Snake River Plain is located to the west and southwest of Yellowstone (Marcus, Meacham and Rodman, 2012).

The land cover in Yellowstone is 80% forest, 15% meadow, and 5% water and is home to more than 10,000 geothermal features. These features include mud pots, hot springs, fumaroles, pools, and geysers, the most famous being Old Faithful (Marcus, Meacham and Rodman, 2012). Major valleys include Lamar and Haden and major rivers include the Yellowstone River, the Madison River, and the Lamar River. Lake Yellowstone is the largest lake in the park as well as the largest high altitude lake in Northern America covering 286 square miles (Marcus, Meacham and Rodman, 2012).

The weather regime in the Greater Yellowstone Area is driven by a combination of the synoptic scale atmospheric circulations and the regional influence of the mountainous topography. During the winter and spring months, the synoptic level flow funnels Pacific moisture through the Snake River valley and onto the Yellowstone plateau resulting in the highest precipitation and snowfall amounts of the year with south through southwestern part of the park receiving the most precipitation. During the summer and fall months, moisture from the Gulf of Mexico fuels the thunderstorms that often form due to the localized differences in surface heating and local wind patterns resulting from the mountainous terrain. Average temperatures can range from slightly below 0° Fahrenheit (F) during the winter months to the 70s and 80s (°F) during the summer months (Marcus, Meacham and Rodman, 2012).

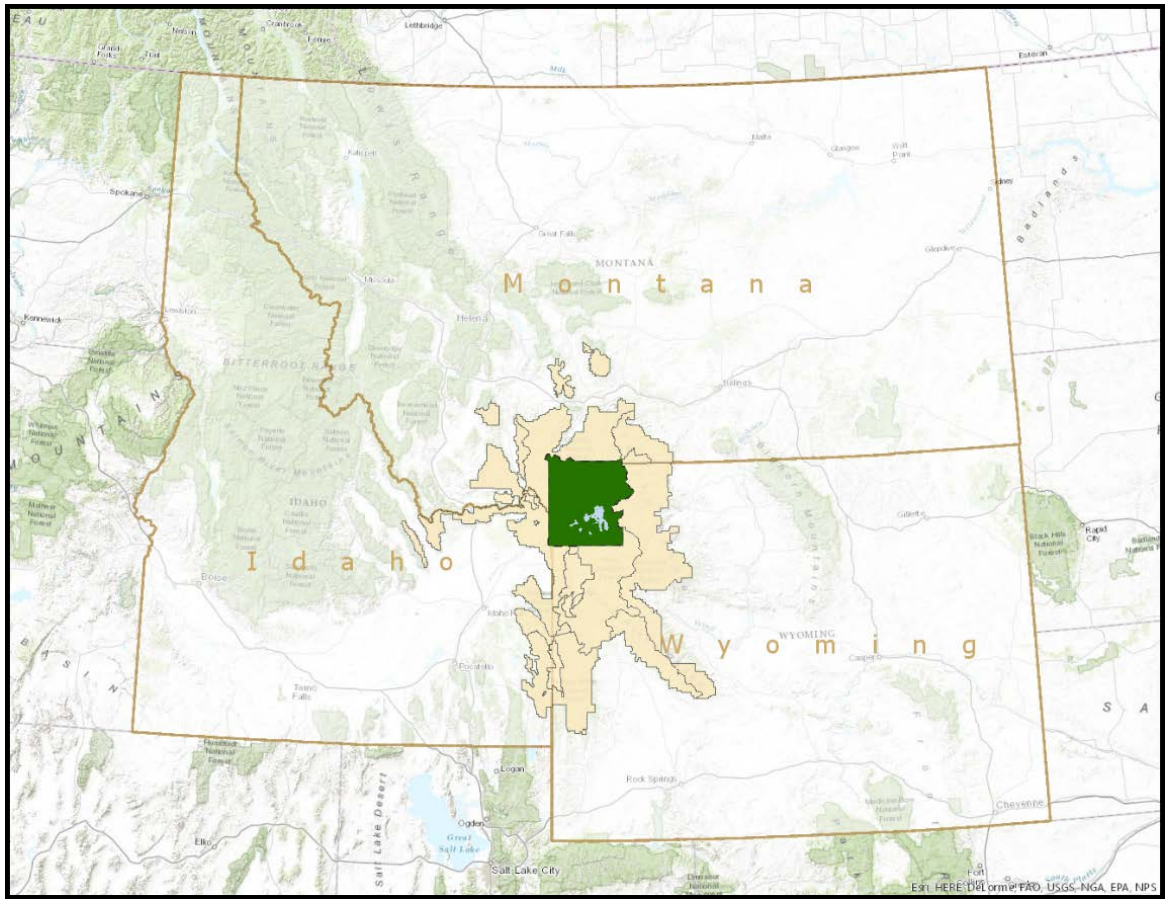

Figure 1. Map depicting location of Yellowstone National Park (shaded in green) and the Greater Yellowstone ecosystem (shaded in tan). Base map courtesy of the Environmental Systems Research Institute (ESRI). State, park, and ecosystem boundaries courtesy of the U.S. National Park Service (NPS).

Figure 2. Map of Yellowstone National Park depicting topography, main roads, points of interest, lakes and rivers. Map data obtained from ESRI and NPS.

Figure 3. Visualization of the terrain in Yellowstone using data collected under the NASA Shuttle Radar Topography Mission (SRTM) and obtained from the U.S. Geological Survey (USGS) in DTED level 2 format.

4. DATA

The primary datasets used to complete this project include the lightning data, a digital terrain elevation model and the National Land Cover Dataset (NLCD). Data preparation and pre-processing tasks were accomplished using ArcGIS Pro (version 1.3), python (version 3.4), and PostgreSQL (version 9.3).

4.1 Lightning Data

Lightning data used in this study were collected by the National Lightning Detection Network (NLDN), which is operated and maintained by Vaisala Inc.,Tucson, AZ (Orville, 2008) and obtained from the U.S. Air Force, 14th Weather Squadron, Asheville, NC. The dataset contains lightning data for a 20-year period of record ending in 2008 for a region bounded between 40°N and 46°N latitude and 105°W and 115°W longitude. Due to a series of sensor and network upgrades in the NLDN, a subset of the data covering a 10-year period of record of lightning events occurring within the park boundary was extracted for this study in order to have data points with consistent accuracy and to decrease data processing and analysis time. The period of record for this study covers from January 1995 to December 2004. The data were received in comma delimited (csv) file format, one csv file per year, with each row in the files representing a single lightning flash event. The dataset contains the following fields or attributes: Month, Day, Hour, Minutes, Seconds, Polarity, Strength, Multiplicity, Detectors, and Type.

In combination, the date- and time- related fields identify the precise time the lightning event occurred. Time is recorded using coordinated universal time (UTC) in the 24-hour clock format. UTC is the internationally accepted time standard used to regulate clocks and time throughout the world. The 24-hour clock format does not use the day (am) and night (pm) designators used in the 12-hour clock format. A time of 19 UTC in 24-hour clock format equates to 7PM UTC in 12-hour clock format and a time of 07 UTC in 24-hour clock format is 07AM UTC in the 12-hour clock format. Yellowstone falls in the MST time zone and the conversion from UTC to MST is -7 (MST = UTC – 7) during standard time and -6 during daylight savings.

The remaining fields describe the physical properties of CG lightning. The type field indicates the type of lightning detected by the NLDN detectors. This analysis only considered the cloud-to-ground lightning type events. The polarity attribute indicates the charge of the lightning stroke; recorded as either positive or negative. The strength attribute is derived from the conversion of the signal strength to electrical current and represents the electrical strength of the lightning event (Lyons, Uliasz and Nelson, 1998). Only cloud-to-ground lightning events with strength values greater than 10 kA were used in order to alleviate noise being counted as false lightning detections (Cummins et al., 2006). The count field, also called multiplicity, is an integer field indicating the number of flashes that occurred during the cloud-to-ground event (Grogan, 2004). The detectors field identifies how many of the NLDN sensors detected the lightning event. Only events detected by one or more detectors were used in this study.

The ten csv files containing data for the period of study were loaded as tables into a geospatially enabled PostgreSQL database to build a lightning data point cloud cube for processing and analysis. The 10-year lightning data point cloud cube was “sliced” by aggregating the data by month and by hour for use in analyzing the data at the month and hour time scales. Aggregating by month resulted in 12 slices while aggregating by hour of day resulted in 24 slices. A graphical representation of this slicing process is shown below (Figure 4) where the x-axis represents the year, the y-axis represents the month and the z-axis represents the hour. The monthly slices contain all of the data for the specified month from each year and hour in the study period while the hourly slices contain all of the data for the specified hour.

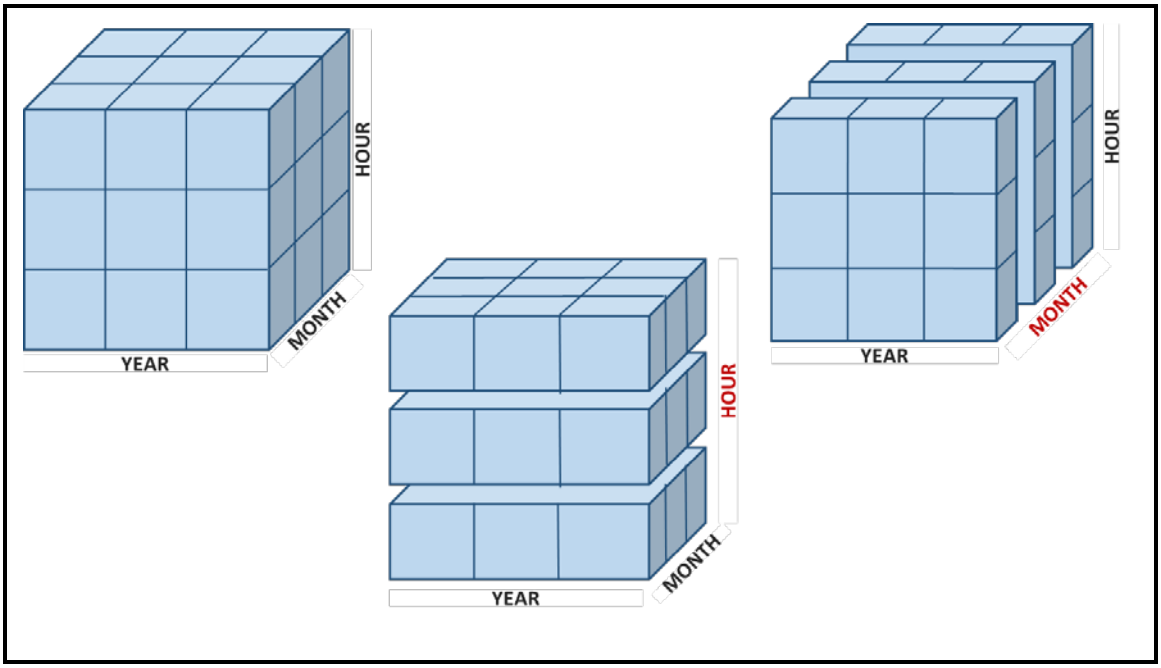

Figure 4. Conceptual drawing of the 10-year lightning data cube and temporal slices.

The lightning data cube was assigned a projected spatial reference of UTM Zone 12. A projected spatial reference is required when performing GIS and spatial analytic methods based on distance such as in clustering algorithms. A copy of the data cube was written out as a series of shapefiles.

Python 3.4 code snippets were used to process the data. Python 3.4 provides an environment to script various geoprocessing tasks using the ESRI *arcpy* python module. The latest version of ArcGIS Pro (version 1.3) was used for all geoprocessing tasks to take full advantage of the software's ability to implement multi-threaded, 64-bit processing on modern computer architectures.

4.2 Terrain Data

Terrain data used in this project was obtained online from the U.S. Geological Survey (USGS). The USGS EarthExplorer web application (http://earthexplorer.usgs.gov) was used to select and download elevation data collected during the National Aeronautics and Space Administration (NASA) space shuttle *Endeavor's* Shuttle Radar Topography Mission (SRTM) occurring February 11-22, 2000. The data collected by the *Endeavor* was post processed by the National Geospatial Intelligence Agency (NGA) to fill in void areas using special terrain interpolation algorithms and other elevation datasets. The final void filled dataset for the U.S. has a grid resolution of 1 arc-second (~30 m). For this project, nine tiles of SRTM elevation data in DTED 2 format covering the Greater Yellowstone region were downloaded from the EarthExplorer web application and mosaicked into a single elevation file. A Yellowstone National Park boundary shapefile was used to extract a clipped copy of the terrain mosaic file. The extracted Yellowstone Park terrain was used to generate terrain aspect and slope files for Yellowstone National Park.

4.3 National Land Cover Data

Land cover data was obtained online from the USGS Multi-Resolution Land Characteristics Consortium (MRLC) website (www.mrlc.gov). The National Land Cover Database (NLCD) is considered the definitive Landsat-based, 30 m resolution, land cover data base for the US (Homer et al., 2015). The 2011 edition (amended in 2014) for the Greater Yellowstone region was downloaded for this analysis. The data is in raster image format where each cell is assigned a thematic category code that describes the underlying

land surface. This project only used the land cover classification codes for forest (code 42) and the grasslands/shrublands (codes 52 and 71).

5. ANALYSIS METHODOLOGY

This study of cloud-to-ground (CG) lightning activity in Yellowstone began with a series of graphs to explore the temporal distribution of lightning flash events by year, by month (aggregated), and by hour (aggregated). To explore the spatial pattern and distribution of lightning activity in the park, a series of maps visualizing the locations of the individual CG lightning flashes as points was created as well as a series of maps displaying the flash density of the lightning activity. These point maps and flash density maps provided an initial visual analysis of the spatio-temporal pattern of lightning flashes within the park and were used as a first step in determining whether or not lightning activity is clustered or random (*Research Question 1*) and whether the activity varies by month or season (*Research Question 2*).

While the point maps and density maps are useful in displaying the spatial distribution of lightning activity and may show areas where clustering may exist, they are qualitative in nature and can't be used to measure the spatial autocorrelation of the pattern. According to Tobler's "first law of geography", "everything is related to everything else, but near things are more related than distant things" (Tobler, 1979). This relationship between "near things" is the foundation of spatial autocorrelation where Legendre defines spatial autocorrelation as pairs of observation a certain distance apart having more similar (positive autocorrelation) or less similar (negative autocorrelation) values than expected for randomly associated pairs of observations (1993).

Global spatial statistics were used to measure the degree of spatial autocorrelation to determine whether or not there is clustering in CG lightning flashes within the given study area (*Research Question 1*). Local spatial statistics were used to determine locations of clustering and the scale of clustering within the study area to identify areas in the park that statistically experience more or less CG lightning activity than other parts of the park (*Research Question 3*). The statistical tests were applied to each of the monthly and hourly data slices as well as to the entire data cube as a whole. Some of the statistical tests used the CG lightning point data as input while others used a spatially aggregated version of the data where the points were aggregated to a grid of hexagon polygon cells.

The investigation of the influence of land formations, land cover, and terrain on the lightning pattern (*Research Question 4*) was based on ordinary linear regression and spatial regression analysis. Test results are presented as a series of maps, graphs, and tables as appropriate. The following text details the analytical workflow and summarizes the spatial statistics and tests used to answer the research questions.

5.1 Visualizing the Spatial Pattern of CG Lightning Events

5.1.1 CG Lightning Event Maps

An essential element in conducting and communicating geospatial analysis is the map. This study used a series of maps to visualize the geospatial distribution of CG lightning activity within Yellowstone National Park. These maps displayed the individual lightning events as points using one point to represent a single CG lightning flash event. A map series was created visualizing the 10-year activity aggregated by month and another series was created showing the activity aggregated by hour of day. A map

visualizing the entire 10-year period of lightning activity was also created. These maps show *where* lightning activity occurred within the park over the 10-year period as well as the monthly and hourly spatial distributions of the lightning activity in an effort to show *when* the lightning activity occurred.

5.1.2 CG Lightning Flash Density Maps

When large numbers of point events are visualized on a map, it can be difficult to evaluate the spatial pattern for clustering. Mapping the density of a variable can be helpful in visualizing a variable's spatial pattern. This study used a hexagonal binning technique to spatially aggregate the lightning flash data. A series of maps depicting the density of the lightning flashes were created as a first step in analyzing the spatio-temporal pattern. Binning is considered a grouping and data reduction method often used when dealing with large datasets. While using squares to tessellate an area may be the more popular method, hexagonal tessellation is thought to be more advantageous since hexagons are considered to be more visually appealing, have better symmetry of nearest neighbors than square or rectangular bins, and are more efficient in spatially covering the study area (Genton et al., 2006; Carr, Olsen and White, 1992, Carr et al., 1987).

Hex binning was accomplished by spatially joining each of the lightning data tables to a hexagon grid feature class. The spatial join process resulted in a new feature table containing a count field to store the number of flash events in each hex cell (labeled as "hexCounts"). A lightning flash density attribute was also added to the new feature table and calculated by dividing the hex counts by the hex cell area and then by the number of years spanning the dataset resulting in a count per km^2 per year (N_g). For each

hex cell, the mean and standard deviation of the lightning flash density were also calculated and mapped using the aggregated monthly data to show the mean flash density and the flash density spatial variance across the study area.

To determine an appropriate hex grid cell size to use for this study, several factors were considered including the size of the study area and the number of points within the study area. The size of the study area and number of events were used in the following equation (Equation 1) from Oyana and Margai (2015) for determining an appropriate cell size for gridding a study area:

$$\sqrt{2}\frac{A}{n}$$

Equation 1. Cell size for tessellating a study area.

where *A* is the area of the region of interest and *n* is the number of points or events in the study area. Applying this equation to this study resulted in cell sizes ranging from 0.66 to 2.5 km^2 for *n* ranging from 5,000 to 20,000 points. For this study, it was desirable to have a single hex grid for use in conducting the analysis at the monthly and hourly time scales as well as for the entire data cube. Using a single grid instead of one that changes throughout the analysis aided in comparing the spatial patterns of lightning activity through time. Furthermore, several of the spatial statistical techniques used in the study required the input data to have a variety of cell values and very few empty grid cells. In addition, as noted by Yin et al. (2007), meteorologists use grid cell sizes of 1 ~ 5 km to study micro-scale patterns that may be influenced by topography, cell sizes of 10 ~ 20 km

to study meso-scale or regional patterns, and 40 ~ 50 km grids to analyze synoptic or planetary scale patterns.

After considering all of these factors and after reviewing the frequency distributions of the resulting spatial join counts when cell sizes ranged from 0.5 to 22 km^2, a cell size of 1 km^2 was chosen to tessellate the study area as this grid size met the analytical needs of this study. Tessellating the study area using 1 km^2 hex cells resulted in a gridded feature layer containing 9,128 hexagonal polygon bins. Many of the hex cells along the park's boundary had areas less than 1 km^2 and therefor were not used in the analysis. This resulted in a total of 8,624 usable hex cells.

5.2 Temporal Analysis of CG Lightning Activity

This study used basic bar and line charts to visualize and explore the temporal distribution of CG lightning events within Yellowstone. Bar charts plotting the number of events versus year, events versus month, and events versus hour of day were created. A table was also created showing event counts aggregated by year and month along with max, min, and average events counts.

5.3 Testing the Spatial Pattern of CG Lightning Activity for Clusters

5.3.1 Nearest Neighbor and K-order Nearest Neighbor Analysis

Nearest Neighbor Analysis (NNA) was used to determine if the spatial pattern of the CG lightning flashes is clustered, dispersed, or random. Using the raw point data, the distance between each point feature and its nearest neighbor is calculated and these nearest neighbor distances are then averaged over all of the points in the study area (Equation 2). The expected average nearest neighbor distance based on a hypothetical

random distribution of points is also calculated (Equation 3). The observed average nearest neighbor distance is divided by the expected average nearest neighbor distance to form a nearest neighbor index (NNI) *R* (Equation 4). When *R* is equal to one (*R*=1), the overall distribution of point features is considered to be perfectly random. When *R* is less than one (*R*<1), the point distribution is considered to be exhibiting clustering with *R* equaling zero (*R*=0) indicating a completely clustered pattern. When *R* is greater than one (*R*>1), the pattern is considered to be tending toward dispersion. A Z-score (Equation 5) and p-value are also calculated for testing the null hypothesis of complete spatial randomness (CSR) (Clark and Evans, 1954).

$$r_o = \frac{\sum_{i=1}^{N} r_i}{N}$$

Equation 2. Observed mean distance (r_o) where *N* is the total number of points and r_i is the nearest neighbor distance for point *i*.

$$r_e = \frac{1}{2\sqrt{\rho}}$$

Equation 3. Expected mean distance (r_e) where ρ is the study area point density calculated as the total number of points divided by the total area of the region under study.

$$R = \frac{r_o}{r_e}$$

Equation 4. Average nearest neighbor index (*R*).

$$z = \frac{r_0 - r_e}{0.26136 \Big/ \sqrt{N\rho}}$$

Equation 5. Average nearest neighbor Z-score (*Z*).

With a modification in how the expected mean distance is calculated, a distribution of *K*-order nearest neighbor indices can be calculated for neighbors beyond the first nearest neighbor or the first order (K = 1). Plotting the K-order NNI versus K can show larger scale clustering than what can be seen when using only the first neighbor. The plot can also show the order(s) where spatial clustering may be more intense or where clustering may cease to exist or begin (Bailey and Gatrell, 1995). The *K*-order expected mean distance is calculated according to Equation 6 (Levine, 2015; Thompson, 1956).

$$r_e(K) = \frac{K(2K)!}{(2^K K!)^2 \sqrt{\frac{N}{A}}}$$

Equation 6. K-order expected mean distance (r_e). *K* is the second, third, fourth, etc. nearest neighbor. *N* and *A* are the total number of points and the size of the study area.

Nearest neighbor analysis was performed using the lightning flash event point data as input. First order nearest neighbor indices were calculated using ArcGIS Pro while CrimeStat was used to calculate the *K*-order nearest neighbor indices. *K*-order NNI were calculated out to the 100th order both with and without the application of a border correction. For the border correction case, the rectangular border correction was used to help minimize study area boundary edge effects. Results of NNA were displayed as a series of graphs.

5.3.2 Getis-Ord General G Statistic (High/Low Clustering)

The Getis-Ord General *G* statistic, *G(d)*, is a global measure of spatial association measuring the concentration or the degree of high and low clustering within a study area (Getis and Ord, 1992). The *G(d)* statistic is calculated using Equation 7.

$$G(d) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(d)x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, j \neq i$$

Equation 7. Getis-Ord General *G* statistic (*G(d)*).

where $w_{ij}(d)$ is a spatial weights matrix of ones and zeros where 1 indicates points *i* and *j* are within the specified distance (*d*) of each other. The attribute values for points *i* and *j* are represented as x_i and x_j and *n* is the total number of points.

A Z-score and p-value are also calculated for the statistic for use in evaluating the null hypothesis of no spatial clustering of high or low feature values (complete spatial randomness) within the study area. For statistically significant positive Z-scores, the null hypothesis is rejected and the spatial distribution of the high values is considered to be more clustered than expected under random conditions. Likewise, for statistically significant negative Z-scores, the null hypothesis is rejected and the spatial distribution of the low values is considered to be more clustered than expected under random conditions.

The General *G* statistic was calculated using ArcGIS Pro. The calculated lightning flash density was used as input and the statistic was calculated using fixed distance bands (*d*) of 5, 10, 15, 20, and 25 km. These distances represent the micro and meso spatial scales previously mentioned (section 5.1.2). The results from the series of monthly and

hourly tests were compiled in a spreadsheet and graphs of *G(d)* statistic Z-score versus distance were generated.

5.3.3 Global Moran's I (Spatial Autocorrelation)

Moran's *I* is another global measure of the degree of spatial autocorrelation and clustering. Whereas the Getis-Ord General *G* statistic is a measure of the overall concentration of high and low values, Moran's *I* is a measure of the correlation where negative correlation indicates dispersion of similar values, positive correlation indicates clustering of similar values (either high or low) and zero correlation indicates complete spatial randomness. Moran's *I* is based on the spatial covariation divided by the total variation as shown in Equation 8 (Oyana and Margai, 2015; Lloyd, 2011; Moran, 1950).

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^{n}(y_i - \bar{y})^2)\,(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij})}$$

Equation 8. Moran's *I*. The total number of points is given by *n* while w_{ij} represents the spatial weights matrix between points *i* and *j*. The variable *y* is the points attribute value.

Values for Moran's *I* range from -1 (dispersion) to +1 (clustered). A Z-score and p-value are also calculated for Moran's *I* statistic. For a statistically significant positive Z-scores, the null hypothesis of spatial randomness is rejected and the high or low values in the dataset are considered to be more clustered than expected. For a statistically significant negative Z-scores, the spatial distribution of high and low values is considered dispersed and the null hypothesis is rejected (Oyana and Margai, 2015, Lloyd, 2011; Moran, 1950).

The ArcGIS Pro software was used to calculate Global Moran's *I* for lightning flash density using 30 distance bands beginning at 1500 m with distance increments of 500 m. Results were plotted as graphs of Z-score versus distance.

5.3.4 Ripley's K (Multi-Distance Spatial Cluster Analysis)

Ripley's *K*, or the *K* function, is a measure used to describe the extent of spatial dependence in the distribution of events by analyzing the pattern at several distance scales. It differs from the nearest neighbor technique in that the nearest neighbor technique is only interested in the distance to the nearest features. The *K* function, on the other hand, is a second-order analysis and uses all inter-event distances between the points within the study area. Ripley's *K* can show the impact of neighborhood size on spatial clustering or dispersion (Oyana and Margai, 2015; Bailey and Gatrell, 1995; Ripley, 1981). Ripley's *K* is estimated using Equation 9.

$$K(h) = \frac{A}{n^2} \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} \frac{I_h(d_{ij})}{w_{ij}}$$

Equation 9. Ripley's *K*.

where *K(h)* is the expected number of points within radius *h*. The area is given by *A* and *n* represents the total number of observed points. The spatial weights matrix represented as w_{ij} contains the weights associated with edge correction. The variable d_{ij} is the distance between points *i* and *j*. When the distance between points *i* and *j* are less than or equal to the radius (*h*), then $I_h(d_{ij})$ is 1 and 0 otherwise. *K(h)* is often transformed (Equation 10) to make it linear, faster to compute and easier to interpret (Oyana and Margai, 2015).

$$L(d) = \sqrt{\frac{K(h)}{\pi}} - h$$

Equation 10. Transformed *K* function to the linear function *L(d)*.

In addition to calculating the observed and expected *K*, a simulated confidence interval (envelope) for the desired α level is calculated under the null hypothesis of complete randomness by using a simulation process to determine a minimum and maximum *L(d)*. *L(d)* is plotted along with the confidence envelope to determine areas of clustering and dispersion. Statistically significant clustering occurs where *L(d)* lies outside of the confidence envelope and above the max *L(d)* while statistically significant dispersion is indicated where *L(d)* lies outside the confidence envelope and below the *L(d)* min (Oyana and Margai, 2015).

Ripley's *K* analysis was performed on the lightning point data using CrimeStat. The analysis was conducted with the simulation number set to 100 and performed with and without the rectangular border correction. In the CrimeStat simulation process, the total number of point observations are randomly distributed in a minimum bounding rectangle equal to the specified study area and the *L* statistics are calculated. This process was repeated 100 times, each time using a different randomly distributed point pattern (Levine, 2015). The results were tabulated and graphed.

5.3.5 Getis-Ord G_i^* (Hot Spot Analysis)

The Getis-Ord G_i^* is a local version of the Getis-Ord General *G* statistic previously discussed. Whereas the General *G* determines an overall measure of spatial

autocorrelation, the G_i^* evaluates spatial autocorrelation at the local scale and identifies local "pockets" of dependence. The General *G* measures the concentration of the high and low values; the G_i^* statistic attempts to show the location of clusters through the identification of spatially significant patterns of high and low clusters by examining each feature within the context of its neighborhood; comparing local averages to global averages. The G_i^* statistic ranges in value from -3 to +3 and is calculated for each feature and produces a Z-score indicating the intensity of the high or low clustering with respect to its neighborhood depending on the sign of the Z-score. A statistically significant "hot spot" is one where a feature with a high value is surrounded by other features with high values (positive Z-score). Likewise, a "cold spot" is one where a feature with low value is surrounded by other features with low values (negative Z-score). (Ord and Getis, 1995, Getis and Ord, 1992). The G_i^* statistic is calculated using Equation 11.

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j - \bar{x}\sum_{j=1}^{n} w_{ij}}{s\left\{\left[\left(n\sum_{j=1}^{n} w_{ij}^2\right) - \left(\sum_{j=1}^{n} w_{ij}\right)^2\right]/(n-1)\right\}^{1/2}}, all\ j$$

Equation 11. Getis-Ord G_i^* statistic. The variable *s* is the variance. All other variables are same as in the *G(d)* statistic previously discussed.

The G_i^* statistic was calculated for lightning flash density using a fixed distance band of 3 km *(d)* in ArcGIS Pro. The results were visualized as a series of maps.

5.3.6 Anselin's Local Moran's I (Cluster and Outlier Analysis)

Anselin's Local Moran's I is a local test statistic for spatial autocorrelation. It is used to find statistically significant local clusters of similar high or low values as well as outliers (Anselin, 1995). Clusters of similar high values are areas of high values that are

surrounded by similar high values and are labeled as “H-H”. Likewise, clusters of similar low values are areas of low values surrounded by similar low values and are labeled as “L-L”. These two categories of clusters represent positive association. Two categories representing negative association are clusters of high values surround by low values, labeled as “H-L”, and clusters of low values surrounded by high values, labeled as “L-H” (Lloyd, 2011). Anselin’s local Moran’s I is calculated using Equation 12.

$$I_i = (z_i - \bar{z}) \frac{\sum_{j=1}^{n} w_{ij}(z_j - \bar{z})}{\sum_{j=1}^{n} w_{ij}}, j \neq i$$

Equation 12. Anselin's Local Moran's I statistic.

where *z* are the attribute values for points *i* and *j*, *n* is the total number of points and *w* is the spatial weights matrix. Positive values of I_i indicate features with neighboring features of similar values while negative I_i values indicate features with neighboring features that have different values (outliers). A Z-score and p-value are also calculated as part of the test to determine statistical significance. The Local Moran’s I (I_i) statistic was calculated for lightning flash density a fixed distance band of 3 km in ArcGIS Pro. The results were visualized as a series of maps.

5.4 Analyzing the CG Lightning-Topography Relationship

Regression analysis was used to investigate the relationship among physical terrain properties and the observed lightning flash density (*Research Question 4*). The regression analysis process produces a set of equations or models that can be used to examine, explore, and explain whether a statistical relationship exists between the physical terrain and the observed spatial patterns in the CG lightning data.

In preparing the data for regression analysis, since the lightning data is in point data format aggregated to hexagonal polygons and the terrain and land cover data are in raster form, an extraction procedure was used to assign terrain elevation, slope, aspect, and land cover values as attributes for each of the 1 km^2 hexagon polygons. The raster terrain dataset is stored as a grid with a much higher resolution than the 1 km^2 hexagon bins used to aggregate the lightning data. Therefore, a majority filter was used to extract and assign the terrain parameters to the hexagon polygons. In addition, since aspect is a circular variable ranging from 0 to 360 degrees, it was decomposed into its north-south and east-west components for use in regression analysis. The land cover data was extracted to the hexagon polygons by calculating the percentage of forest (NLDC 42) and the percentage of grass/shrub (NLDC 52 and 71) occupying each hexagon. The flash density variable was transformed using the square root function to better approximate a normal distribution for this variable.

Regression was first performed using the entire data. Initially, an ordinary least squares (OLS) linear regression analysis was performed on the 10-year period of record. Initial results showed that very little of the variance in the lightning density could be explained by terrain and land cover using general linear regression. Furthermore, a visual

inspection of the resulting residual standard deviation maps showed evidence of spatial autocorrelation, violating the assumptions of an OLS regression. Spatial regression using the GeoDa software package (Anselin, 2016; Anselin, Syabri & Kho, 2010) was also performed on the 10-year dataset following the spatial regression decision tree outlined by Anselin in the GeoDa workbook (Anselin, 2005) to determine whether spatial lag or spatial error model should be applied. During the initial regression, the spatial relationships were defined by generating the spatial weights matrix (w_{ij}) based on the first-order rook neighbors, neighbors (j) who directly share a boundary or edge with i (contiguous neighbors). In the case of a grid with rectangular cells, a rook contiguity order of 1 means the cells located immediately to the north, south, east, and west are considered neighbors resulting in 4 total neighbors. For a grid of hexagonal cells, first-order rook contiguity results in all 6 of the immediate cells being identified as neighbors.

Once it was determined that spatial error regression should be applied to this dataset, a series of spatial error regression models were generated by varying the order of the spatial weights matrices from 1 to 5, essentially increasing the radius of influence. While the first-order neighbors are those directly sharing a boundary or edge with i, the second-order neighbors are those that share boundaries with the first-order neighbors of i and the third-order neighbors share boundaries with the second-order neighbors, etc. The results were compared to see which model had the highest R^2 value. In this case, the model using the third-order rook contiguity spatial weights matrix provided the highest R^2 value. Regression analysis was then performed at the monthly scale using the spatial weights matrix based on the third-order rook contiguity. The results were aggregated and

summarized in a series of tables. To account for spatial effects, the spatial error regression models include a spatial autoregressive error term on the right hand side of the model and is an indicator of missing information in the model.

To further explore the relationship between terrain parameters and CG lightning flash density, OLS regression was applied on the cells that were identified as either hot or cold spots for the entire dataset. The hex cells that the Getis-Ord G_i^* statistic identified as clusters of either high or low values for the entire dataset were selected and joined with the CG lightning flash density calculated for the entire 10-years, aspect, slope, elevation and land cover variables. Two separate OLS regression models were generated, one using the cells selected as hot spots and the other using the cells selected as cold spots mainly to see if there was any difference in the relationships between elevation and CG lightning flash density in areas of hot and cold spots.

6. RESULTS AND DISCUSSION

This section presents the visualizations and analysis beginning with the dot maps of the individual lightning events and the lightning flash density maps. No lightning activity was observed during the winter months of the 10-year period of study; therefore, only maps showing activity between March and November are shown. The temporal distribution of the aggregated monthly and hourly data is shown next in the form of bar graphs and a table. These results are followed by charts and maps highlighting the results from the various geospatial statistical tests used to test the spatial pattern for evidence of clustering. For the aggregated monthly data, only the results from May through September are presented since the bulk of lightning activity occurred during these months. Likewise, for the aggregated hourly data, only the results for the 19 UTC through 02 UTC hours will be presented. The last part of this section presents the results of the regression analysis.

6.1 Visualizing the Spatial Pattern and Analyzing for Global Clustering

6.1.1 CG Lightning Event Maps

Between January 1995 and December 2004, there were 64,721 cloud-to-ground lightning flashes within Yellowstone National Park. In the map displaying the entire 10-year data set (Figure 5), no large areas void of CG lightning activity appear. Lightning flashes occurred in all parts of the park. Furthermore, visual inspection suggests higher density regions or clusters of lightning activity in some parts of the park.

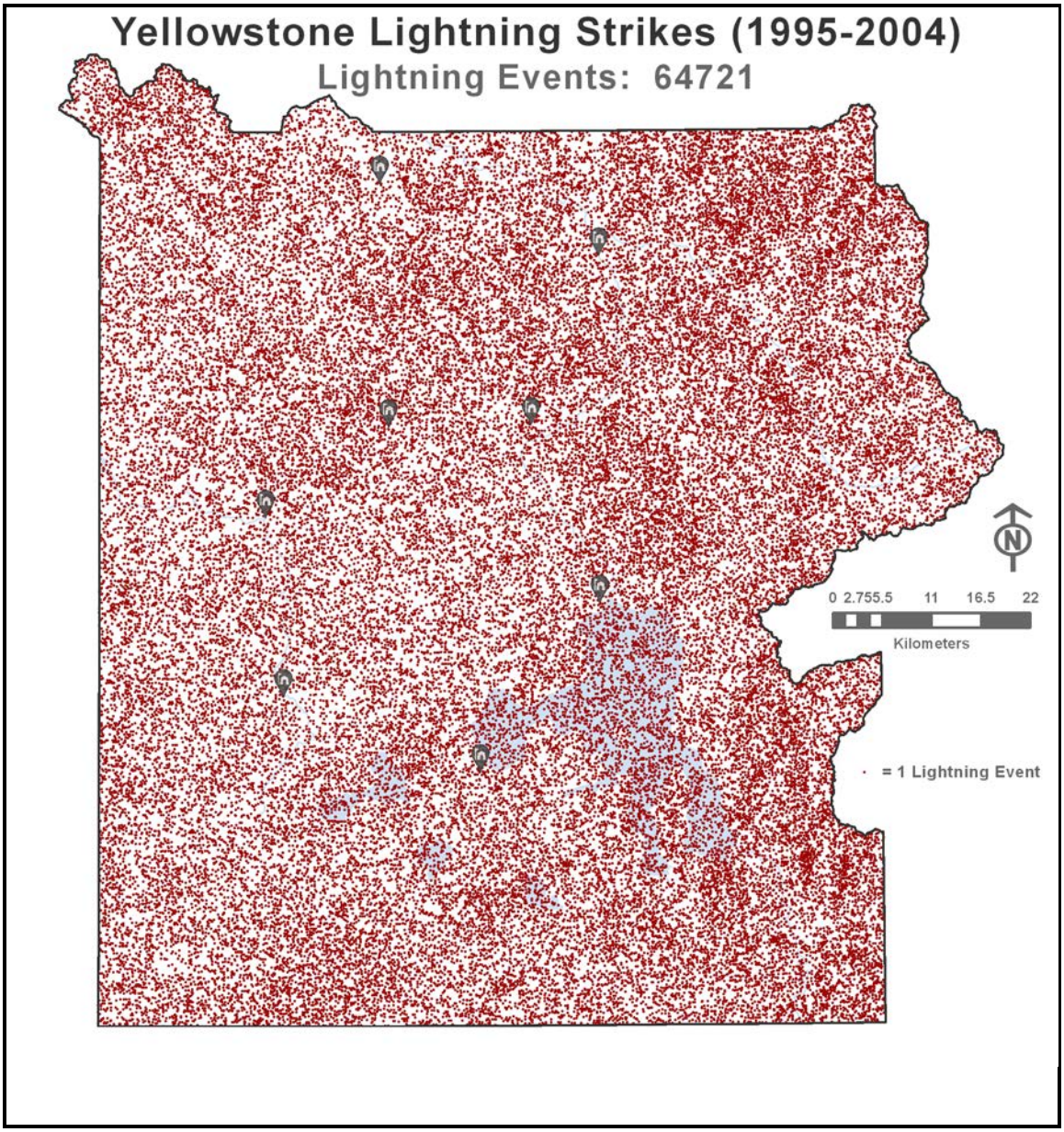

Figure 5. Map of all cloud to ground lightning strikes in Yellowstone National Park between 1995 and 2004. Each red cross represents a single CG lightning event. During the 10-year period of study, 64,721 lightning events occurred within the park.

The series of maps (Figure 6) depicting the 10-year lightning data set aggregated on the monthly scale shows the number of lightning strikes increased as the seasons progressed from spring to summer and declined from summer to fall with no lightning strikes recorded during the winter months (December to February) of the period of study. The least amount of lightning strikes occurred during March (22 strikes) and November (28 strikes). The highest number of lightning strikes occurred in July (19,936 CG lightning events) followed by August (19,756 CG lightning events) and June (14,417 CG lightning events). Both July and August experience more than twice the number of lightning strikes than experienced during the March through May period and the September through November period combined. As in the previous map, visual inspection suggests that some regions may have a higher density or clusters of lightning activity.

In the series of maps showing the dataset aggregated by hour of day (Figure 7 and Figure 8), the highest number of lightning strikes take place during the 19 UTC to 02 UTC period of time with more than 4,000 lightning flash events occurring each hour. The 03 UTC to 18 UTC time period experiences less than 3,000 strikes each hour with the 06 UTC to 16 UTC time period showing less than 1,000 strikes each hour during this study period. It's difficult to visually discern spatial clustering during the 03 UTC to 18 UTC time period. During the hours of peak activity between 19 UTC and 02 UTC, areas of possible clusters are more discernable.

This series of maps shows that CG lightning activity is both seasonal and diurnal. This is an indicator that CG lightning activity is *not* consistent throughout the year.

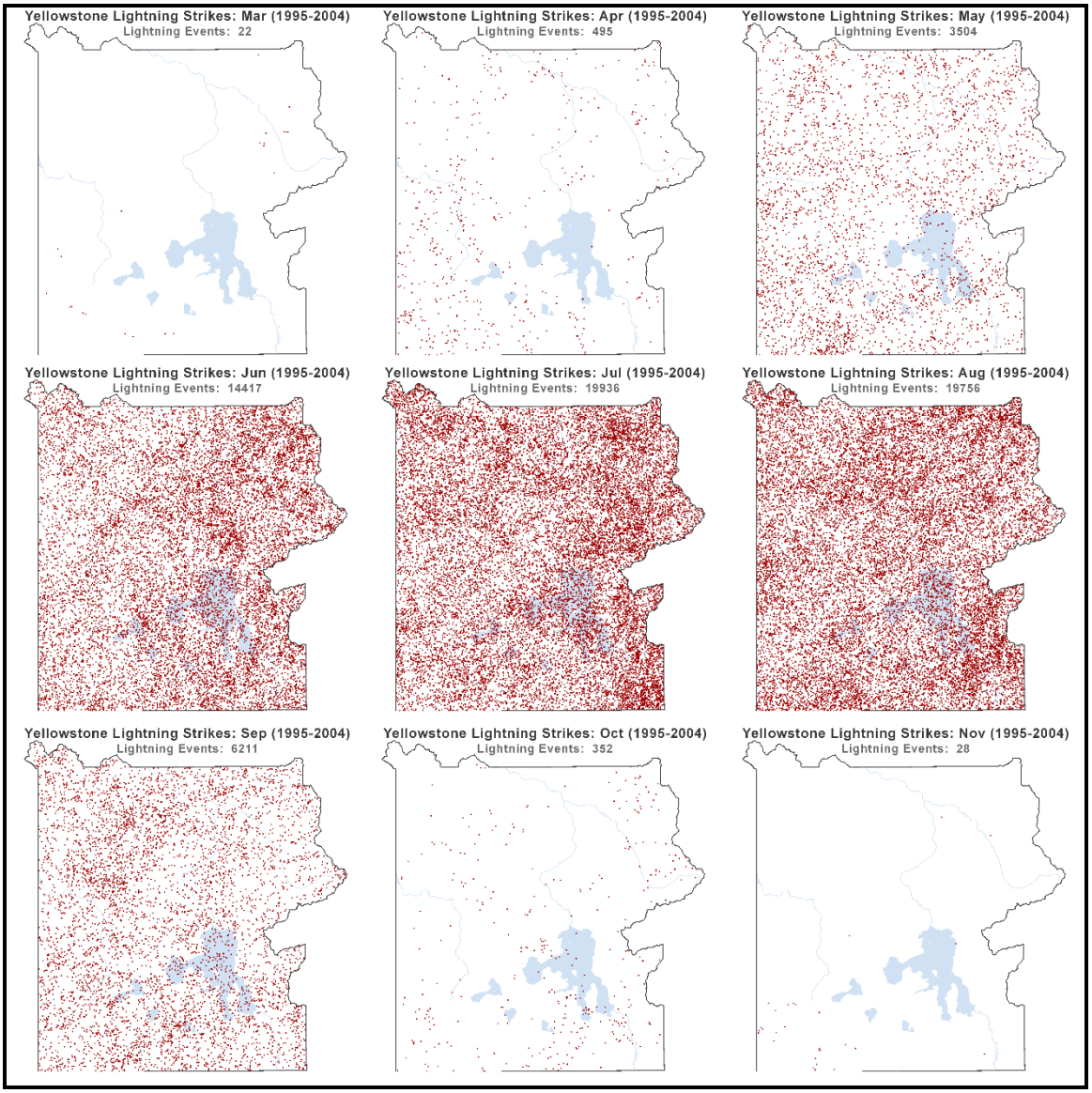

Figure 6. Maps series of the aggregated monthly lightning activity between 1995 and 2004 in Yellowstone National Park. Each red cross represents a single CG lightning event.

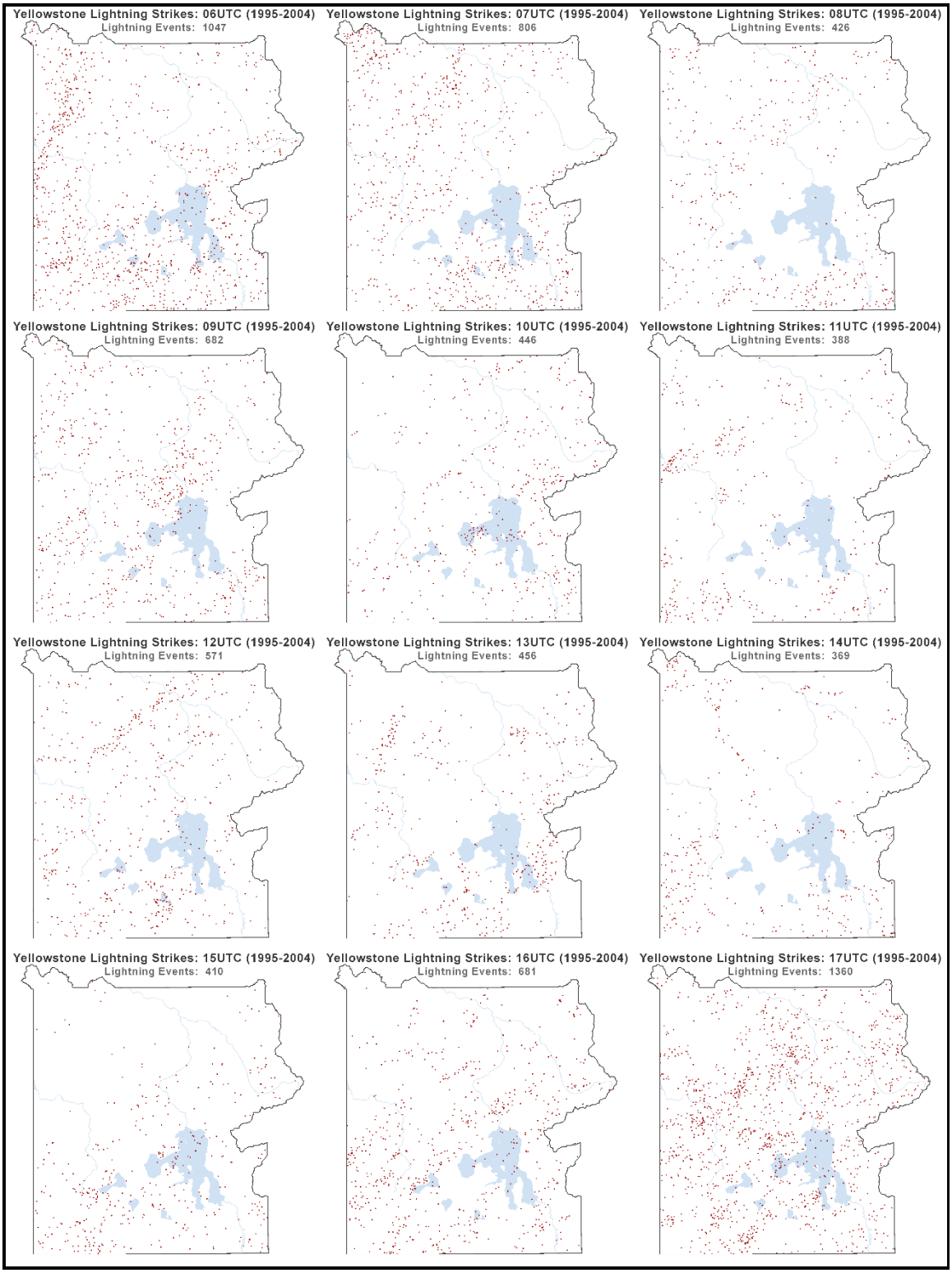

Figure 7. Map series of lightning strikes aggregated by hour for the hours 06UTC to 17UTC. Each red cross represents a single CG lightning event.

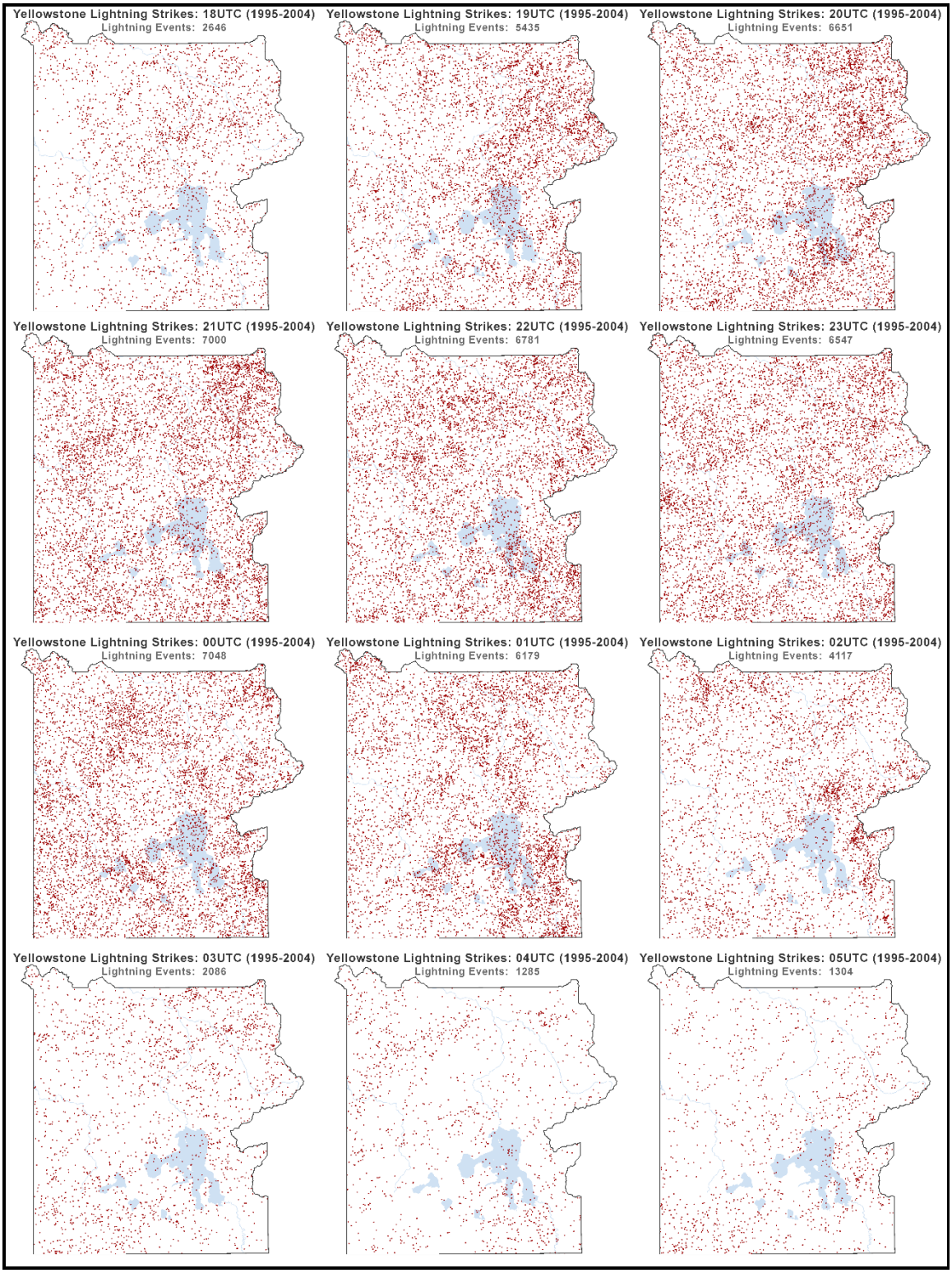

Figure 8. Map sereis of lightning strikes aggregated by hour for the hours 18UTC to 05UTC. Each red cross represents a single CG lightning event.

6.1.2 CG Lightning Flash Density Maps

Maps depicting the individual cloud-to-ground lightning point events are somewhat difficult to visually interpret. Clustering of CG lightning activity in Yellowstone is easier to detect in the map series of lightning flash density on the hexagon polygon grid. In the map depicting the CG lightning flash density using the entire 10-year dataset (Figure 9), the density of lightning flashes is not uniformly distributed across the study region. Areas of higher density values can be seen throughout the park especially in the more mountainous northeastern and southeastern regions.

The map series depicting the aggregated monthly flash density of CG lightning (Figure 10) shows the density increases in coverage and value between March and August and then decreases in spatial coverage and value. Lightning flash density appears to be greatest during the summer months (Jun – Aug) with more spatial coverage and areas of higher density values during July and August, especially in the mountainous eastern part of the park.

Lightning flash event density aggregated at the hourly scale is displayed as a map series in Figure 11 and Figure 12. The flash density pattern does not appear to be as intense between 06 UTC and 17UTC. Between 18UTC and 05UTC, there is an increase in spatial coverage and the flash density pattern appears more intense. Specifically, flash density is not consistent through time.

Maps of the mean monthly flash density (Figure 13) and monthly flash density standard deviation (Figure 14) also show evidence that the spatial pattern of CG lightning activity is not consistent through time. These maps show several areas of the park where the flash density of lightning activity has a higher mean, including the mountainous areas

such as in the northeast and southeastern parts of park. Likewise, the spatial pattern of the monthly flash density standard deviation is not consistent across the study area, areas of low and high standard deviation values can be seen throughout the park.

The spatial patterns of CG lightning events displayed in these series of maps show the spatial pattern is not consistent through time. The spatial pattern of lightning events changes along with the seasons on the monthly scale and diurnally on the hourly scale. These observed seasonal and diurnal changes in the CG lightning pattern are consistent with the changes observed as the weather regime in Yellowstone becomes influenced more by the mountain induced daytime convection during the summer months when surface heating from the sun is strongest.

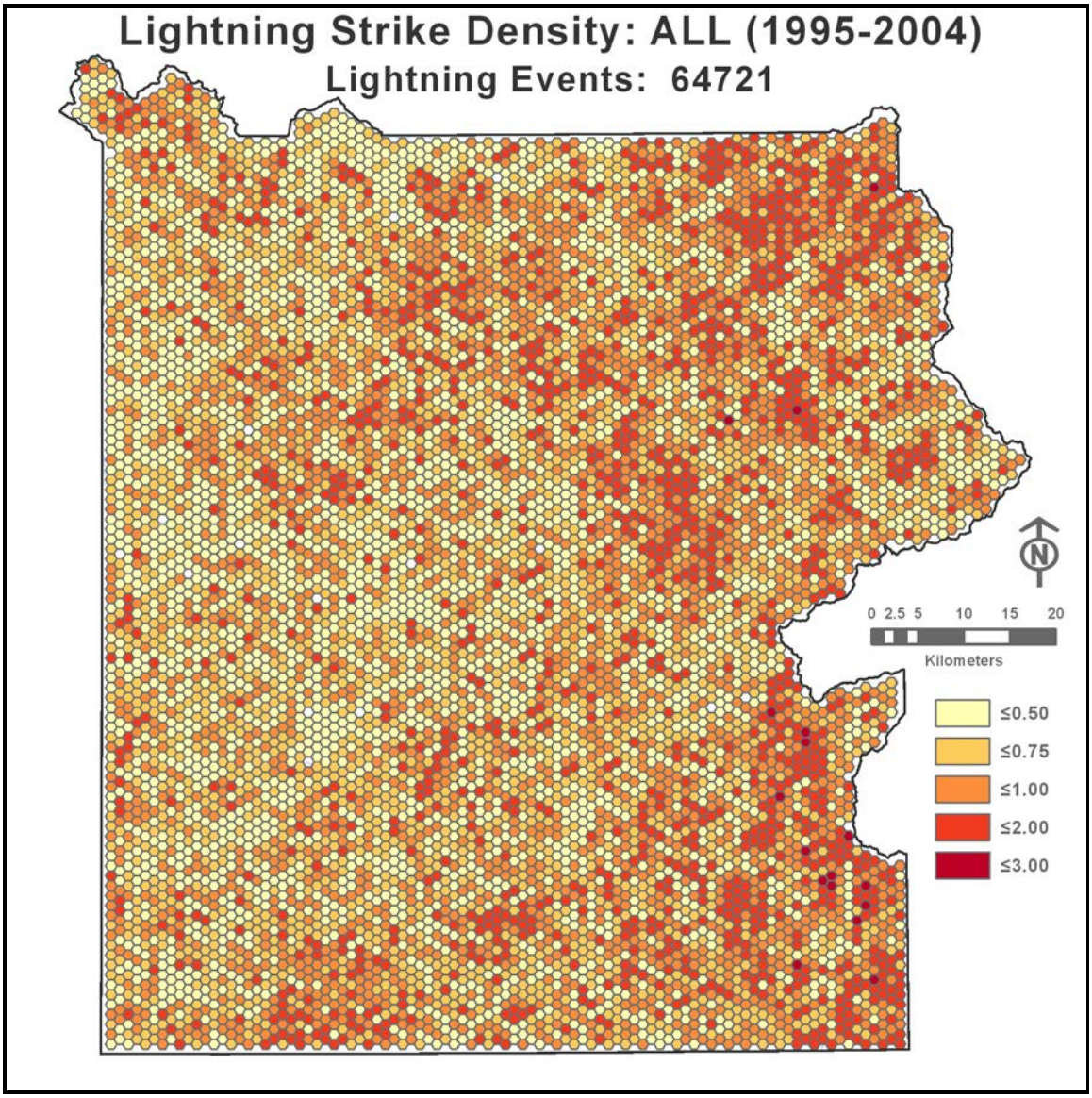

Figure 9. Map depicting spatial density of cloud-to-ground lightning flashes within Yellowstone National Park between 1995 and 2004. Lightning density (*Ng*) displayed as the number of CG lightning events per km^2 per year.

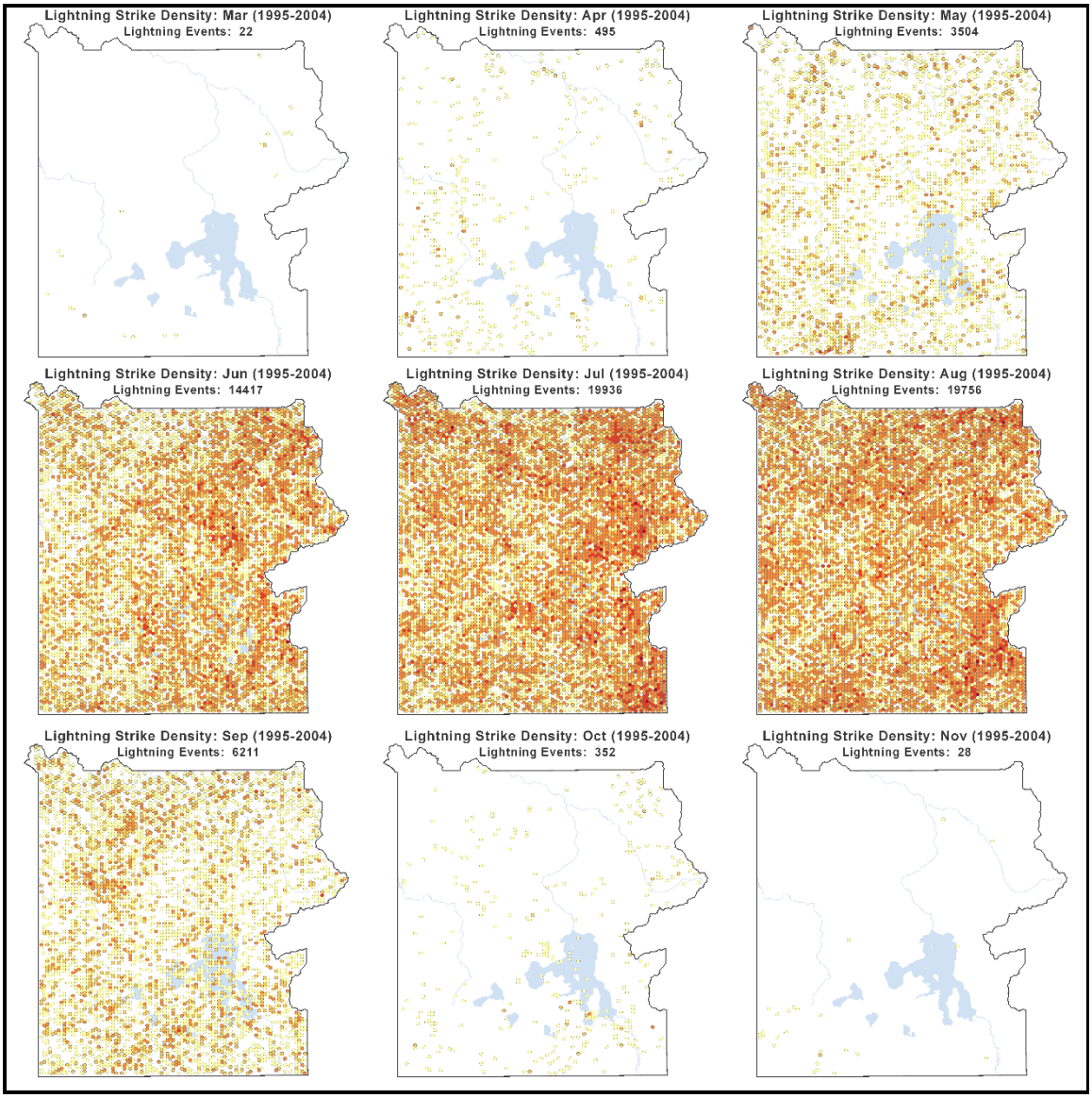

Figure 10. Map series depicting CG lightning flash density (*Ng*) between 1995 and 2004 aggregated by month.

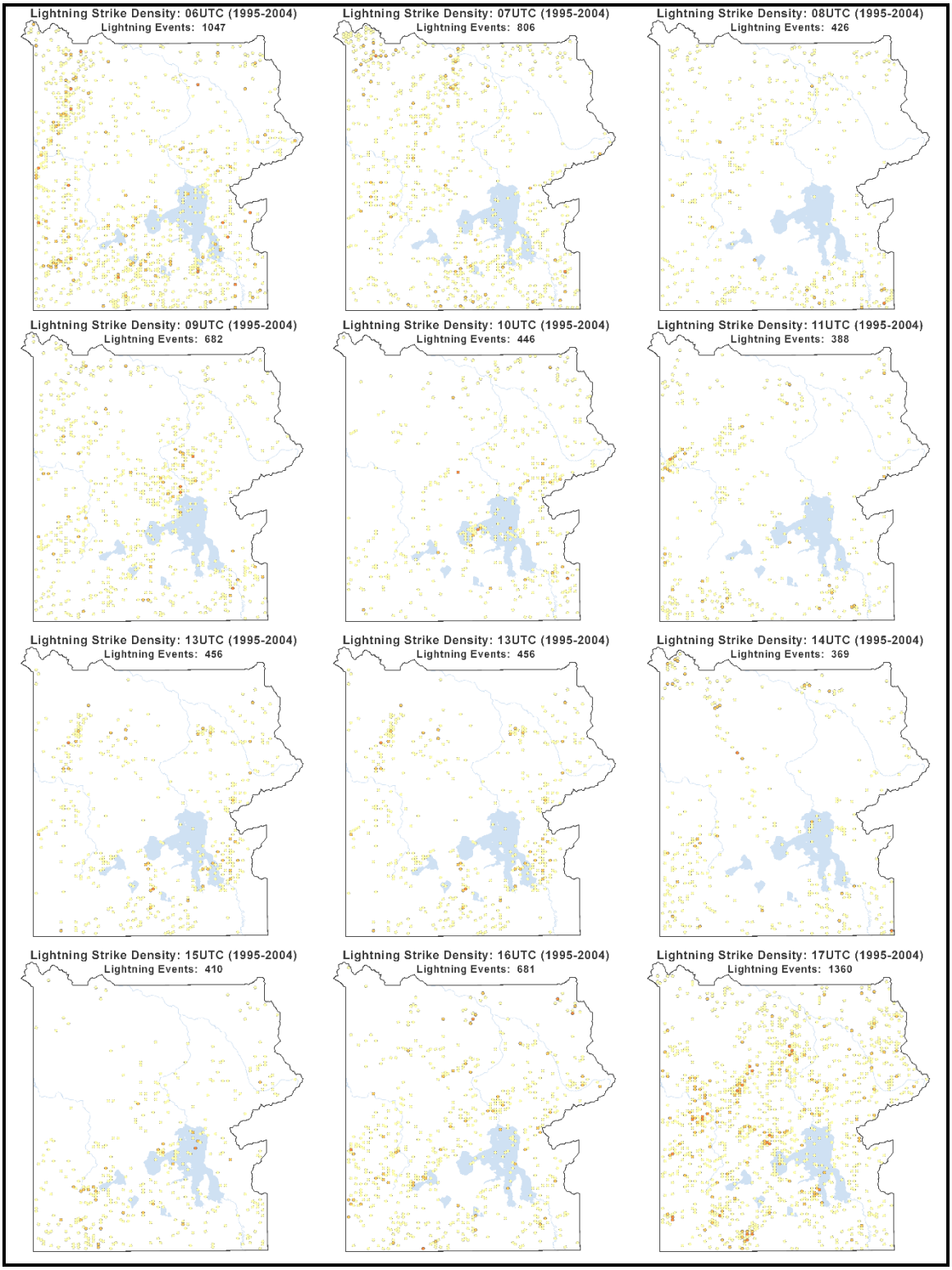

Figure 11. Map series depicting CG lightning flash density (*Ng*) between 1995 and 2004 aggregated by hour (06UTC to 17UTC).

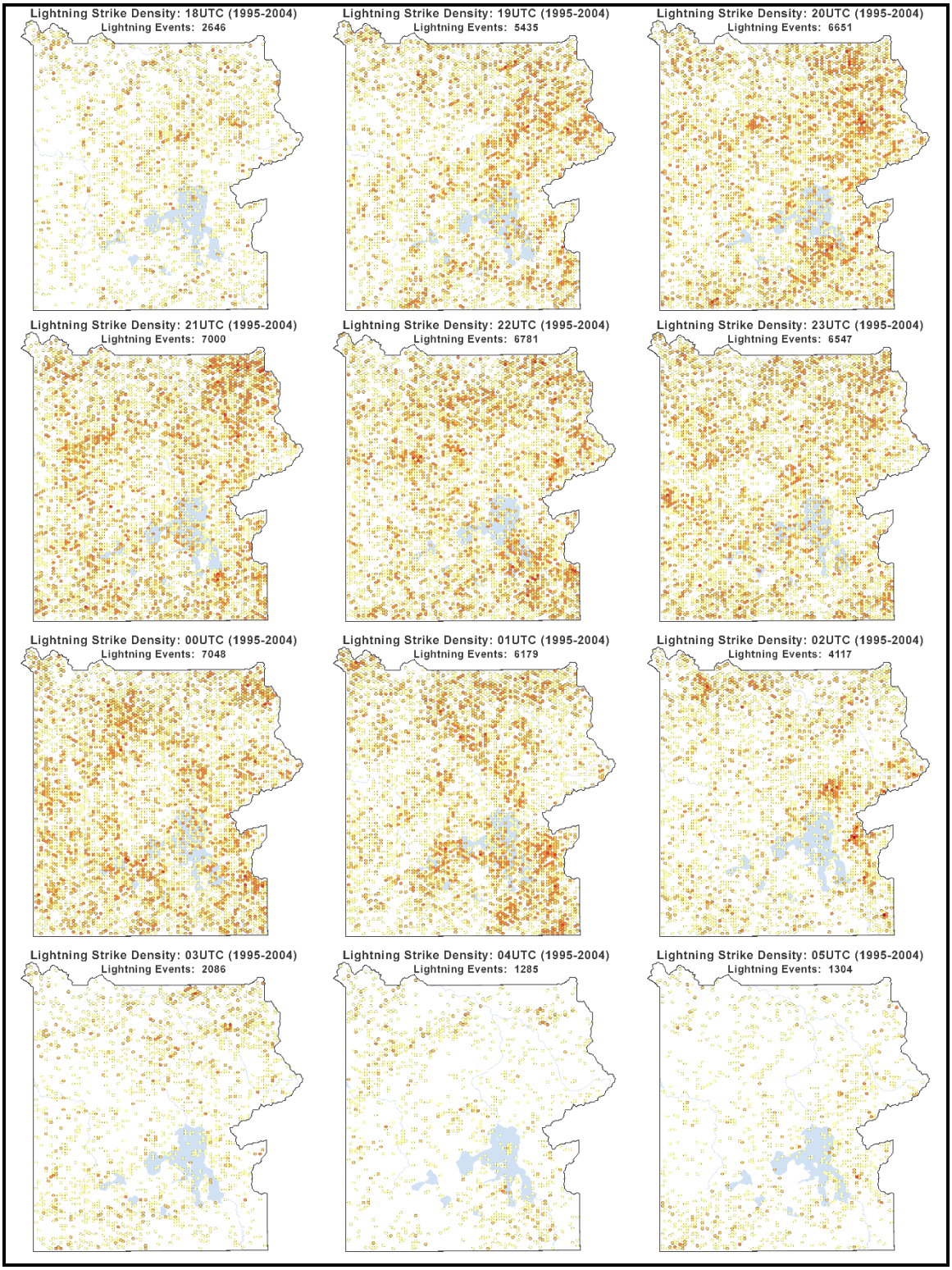

Figure 12. Map series depicting CG lightning flash density (*Ng*) between 1995 and 2004 aggregated by hour (18UTC - 05UTC).

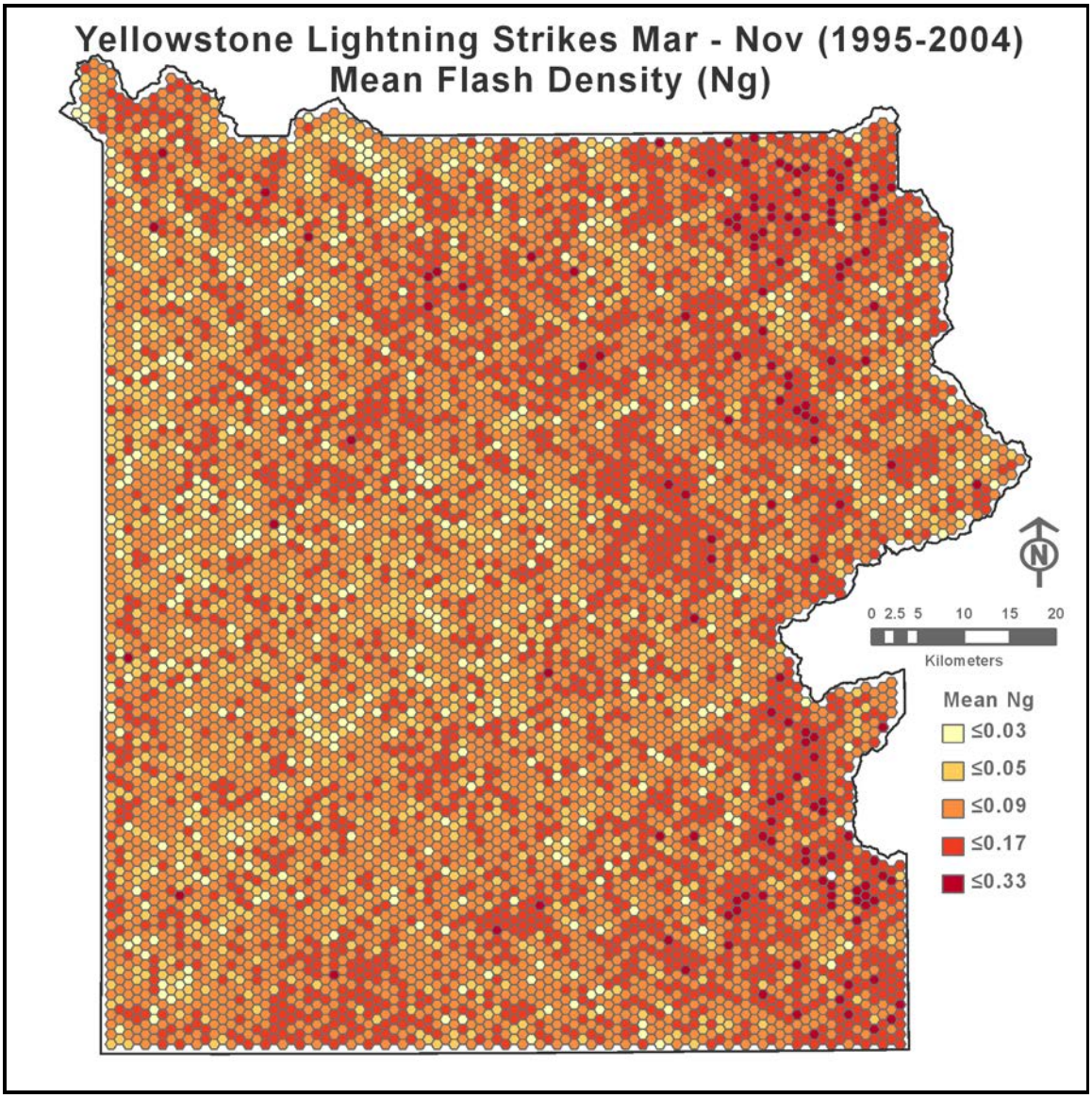

Figure 13. Map of the mean monthly (Mar- Nov) lightning flash density (*Ng*) between 1995 and 2004.

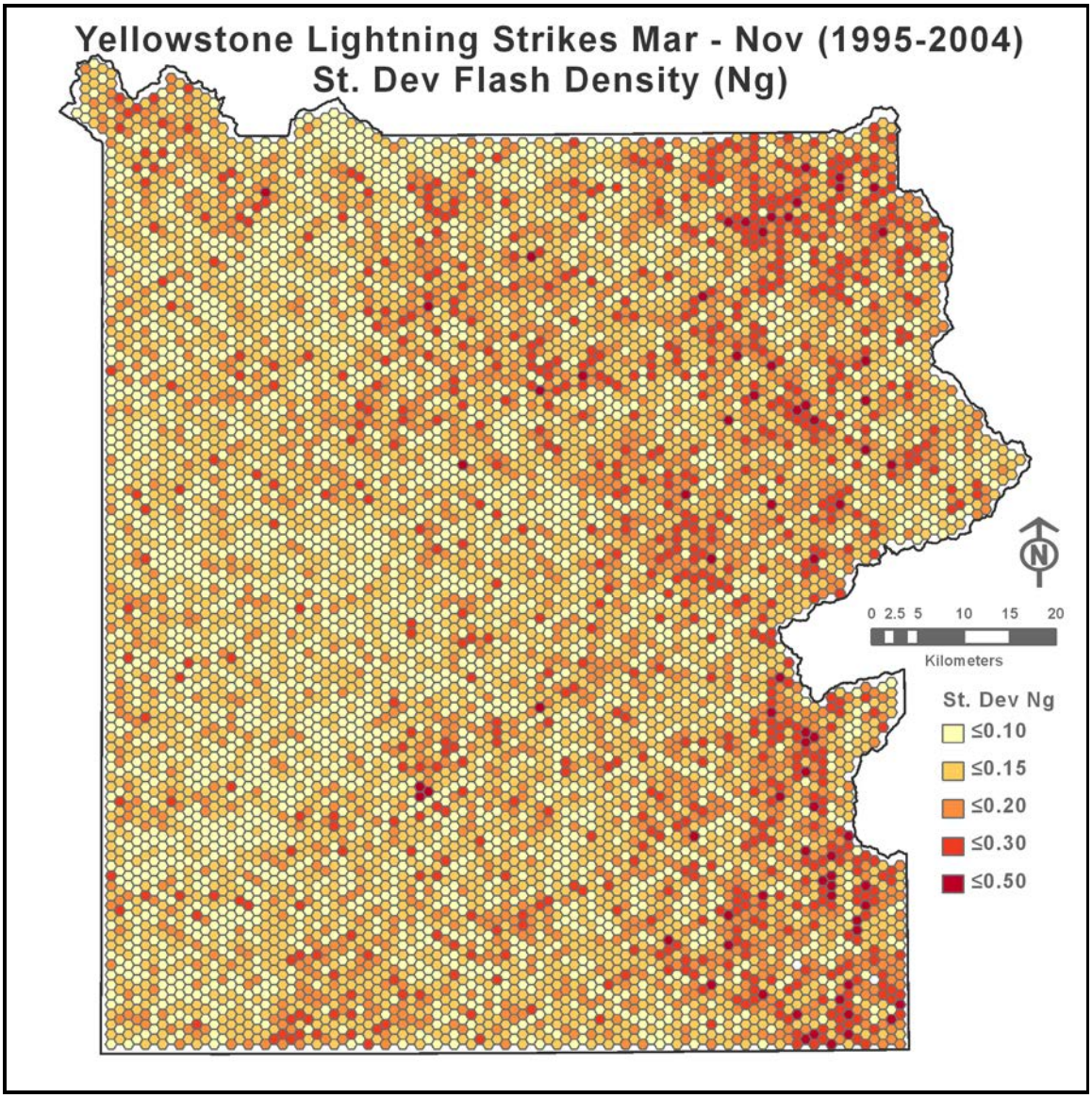

Figure 14. Map of the standard deviation of the monthly CG lightning flash density (Mar - Nov) between 1995 and 2004.

6.1.3 Global Spatial Autocorrelation

In the previous section, maps of CG lightning activity demonstrated periods of peak lightning activity between May and September when evaluating the aggregated monthly patterns and between 19 UTC and 02 UTC when analyzing the aggregated hourly data. The focus of this section is the results of the clustering analysis during the periods of peak CG lightning activity. Table 1 summarizes the spatial test statistics used for analyzing the pattern and testing whether or not the CG lightning activity is spatially clustered (*Research Question 1*).

Table 1. Table summarizing the spatial statistics used to test for global clustering of CG lightning.

TEST STATISTIC	REMARKS	INPUT	TOOL/ RESULTS
NNA and K-Order NNA	Determines if the distribution of points is clustered, random, or dispersed using a nearest neighbor index (NNI) comparing the average nearest neighbor distance for the given points to the expected average nearest neighbor distance for a random set of points. K-NNA is calculated over a range of nearest neighbors to evaluate clustering across spatial scales.	Points	CrimeStat (Tables)
Getis-Ord General G	Determines if similar values are clustered, either high values or low values clustered.	Hex Cells	ArcGISPro (Graphs)
Global Moran's I	Measures spatial autocorrelation to determine if the pattern is clustered, dispersed, or random. Tested across a range of distances to identify scale of clustering.	Hex Cells	ArcGISPro (Graphs)
Ripley's K	Identifies statistically significant clustering or dispersion over a range of distances.	Points	CrimeStat (Graphs)

6.1.3.1 Nearest Neighbor Analysis and K-order Nearest Neighbor Analysis

Output from the Nearest Neighbor Index (NNI) analysis reveals that lightning activity is clustered both at the monthly scale (Table 2) and the hourly scale (Table 3). NNI values below 1 indicate spatial clustering while values greater than 1 indicate dispersion. When NNI equals 1 the pattern is considered to be completely random. The null hypothesis in a nearest neighbor analysis states the spatial pattern is random. In Table 2, the NNI values are all less than 1 ranging from a low of 0.44 to a high of 0.98. It is during the summer months when the spatial pattern is the least clustered (NNI nearest to 1) followed by May and September. The observed mean distance between events is less than 1000 m between May and September with the smallest mean distance occurring during July and August (326 m and 330 m). The observed mean distance is 183 m with an NNI of 0.98 when performed on the entire dataset. While the table also shows statistically significant results with very low Z-scores and p-values due to the large number of observations, the NNI values show the spatial pattern of CG lightning is closer to a random pattern than a clustered or dispersed pattern between May and September when CG lightning activity is the highest.

Similar results are seen when NNI is calculated on the aggregated hourly data. The observed NNI is less than 1 for all hours of the day and the observed mean distance is less than 1000 m between the hours of 19 UTC and 02 UTC. Between 03 UTC and 18 UTC, the mean distance is between 1000 m and 2000 m. The NNI values are the highest and closest to 1 between 18 UTC and 01 UTC suggesting a nearly random pattern of CG lightning activity. The spatial pattern has the highest degree of clustering with the lowest NNI values between 10 and 16 UTC with NNI values between 0.76 and 0.83. Once again

the table also shows statistically significant results with very low Z-scores and p-values due to the large number of observations.

Table 2. Table of Nearest Neighbor Index (NNI) values for lightning point event data collected between 1995 and 2004 and aggregated by month. Columns labeled as "Observed" and "Expected" refer to distance (m) between point events, "N" is the number of point events and "C/R/D" indicates whether the point pattern is considered as clustered, dispersed, or random. The row labeled "All" shows the results when applying nearest neighbor analysis to the entire 10-year dataset as a whole.

	N	Observed (m)	Expected (m)	NNI	Z-score	p-value	C/R/D
Mar	22	4462	10106	0.44	-5.01	<0.001	Clustered
Apr	495	1767	2130	0.83	-7.26	<0.001	Clustered
May	3504	740	801	0.92	-8.63	<0.001	Clustered
Jun	14417	386	395	0.98	-4.82	<0.001	Clustered
Jul	19936	326	336	0.97	-7.76	<0.001	Clustered
Aug	19756	330	337	0.98	-5.53	<0.001	Clustered
Sep	6211	576	601	0.96	-6.38	<0.001	Clustered
Oct	352	2083	2526	0.82	-6.31	<0.001	Clustered
Nov	28	6559	8958	0.73	-2.71	0.0067	Clustered
ALL	64721	183	186	0.98	-8.54	<0.001	Clustered

Table 3. Table of Nearest Neighbor Index (NNI) values for lightning point event data collected between 1995 and 2004 and aggregated by hour (UTC). Columns labeled as "Observed" and "Expected" refer to distance between point events, "N" is the number of point events and "C/R/D" indicates whether the point pattern is considered as clustered, dispersed, or random. The row labeled "All" shows the results when applying nearest neighbor analysis to the entire 10-year dataset as a whole.

	N	Observed (m)	Expected (m)	NNI	Z-score	p-value	C/R/D
0	7048	554	565	0.98	-3.05	0.0022	Clustered
1	6179	570	603	0.94	-8.28	<0.001	Clustered
2	4117	676	739	0.91	-10.50	<0.001	Clustered
3	2086	949	1038	0.91	-7.50	<0.001	Clustered
4	1285	1192	1322	0.90	-6.74	<0.001	Clustered
5	1304	1209	1313	0.92	-5.43	<0.001	Clustered
6	1047	1322	1465	0.90	-6.03	<0.001	Clustered
7	806	1521	1670	0.91	-4.84	<0.001	Clustered
8	426	1994	2297	0.87	-5.21	<0.001	Clustered
9	682	1614	1815	0.89	-5.52	<0.001	Clustered
10	446	1858	2244	0.83	-6.96	<0.001	Clustered
11	388	1839	2406	0.76	-8.89	<0.001	Clustered
12	571	1637	1984	0.83	-7.98	<0.001	Clustered
13	456	1709	2220	0.77	-9.41	<0.001	Clustered
14	369	1920	2468	0.78	-8.16	<0.001	Clustered
15	410	1897	2341	0.81	-7.34	<0.001	Clustered
16	681	1488	1816	0.82	-9.03	<0.001	Clustered
17	1360	1155	1285	0.90	-7.15	<0.001	Clustered
18	2646	882	921	0.96	-4.23	<0.001	Clustered
19	5435	607	643	0.94	-7.91	<0.001	Clustered
20	6651	552	581	0.95	-7.73	<0.001	Clustered
21	7000	545	567	0.96	-6.14	<0.001	Clustered
22	6781	551	576	0.96	-6.73	<0.001	Clustered
23	6547	571	586	0.98	-3.80	<0.001	Clustered

The K-order Nearest Neighbor Analysis calculates the NNI for a specified range of nearest neighbors. In this study, nearest neighbor distances and indices were calculated for the first 100 neighbors (K = 100). Graphs of the K-order NNI plotted over the calculated average nearest neighbor distance for each order (K^{th} nearest neighbor) can show the scale of clustering. Distances where the NNI graph lies below 1 indicate the distances where clustering occurs while distances where the NNI graph lies above 1 indicate the distances where dispersion exists in the pattern.

In the graphs for the monthly aggregated CG lightning event data (Figure 15), the NNI increases as K (and distance) increases. The K-order NNI is less than 1 at distances less than 6 km in May. During the summer months, the distance where the NNI becomes larger than 1 decreases to less than 2 km by August and then increases to 4 km during September. Table 4 summarizes the results of the K-order NNI analysis for the monthly data showing the K^{th} neighbor and distance where the NNI first becomes greater than or equal to 1 as well as the minimum and maximum NNI. During the summer months when CG lightning is at its peak, the NNI ranges from a low of 0.97 to a high of 1.02 for K ranging from 1 to 100. While these values indicate a clustered pattern for distances up to 3.28 km in June to 1.76 km in August, the degree of clustering is relatively small and the pattern of CG lightning activity is actually closer to a random pattern than it is to a clustered pattern. This is also true for May and September although the NNI is slightly smaller than seen during the summer months. Excluding March and November due to an extremely low number of CG lightning events, the magnitude of clustering is highest during April and October.

As seen in the graphs for the aggregated monthly data, the K-order NNI graph for the hourly data generally increases with the distance (Figure 16) as *K* increases from 1 to 100. Between 19 and 02 UTC, the distance where the spatial pattern of CG lightning changes from a slightly clustered to a slightly dispersed pattern is between 4 and 8 km when CG lightning activity is highest. During these hours, the minimum NNI ranges from 0.93 to 0.97 and the maximum NNI ranges from 1.02 to 1.04 (Table 5). These values are very close to 1 suggesting the spatial pattern of CG lightning activity is closer to a random pattern than either a dispersed or clustered pattern during this time period. Between 03 and 18 UTC, the distance where the pattern changes from clustered to dispersed is between 6 and 23 km. During this time period the minimum NNI is between 0.72 and 0.90 and the maximum NNI is between 1.05 to 1.31. The range of NNI values (highest maximum NNI – lowest minimum NNI) is much smaller for the 19 to 02 UTC time period, when CG lightning activity is highest, than it is for the 03 to 18 UTC time period. This suggests the spatial pattern of CG lightning is consistently closer to a random pattern between 19 and 02 UTC across spatial scales. During the 03 to 01 UTC time period, the spatial pattern changes from more clustered to more dispersed across spatial scales.

When K-order NNI analysis is applied to the entire dataset, the average nearest neighbor distance where the NNI becomes greater or equal to 1 is 1.23 km. As *K* increase from 1 to 100, the NNI changes from 0.98 to 1.01. Once again these values are very close to 1 suggesting the spatial pattern of CG lightning is consistently closer to being a random pattern across spatial scales.

Overall, the results from the NNA and K-order NNA suggests the spatial pattern of CG lightning activity is only slightly clustered for the spatial scale ranging in distance from 1 to 23 km. During times of peak CG lightning activity, the spatial pattern exhibits the least amount of statistically significant clustering with the pattern actually being closer to a random pattern.

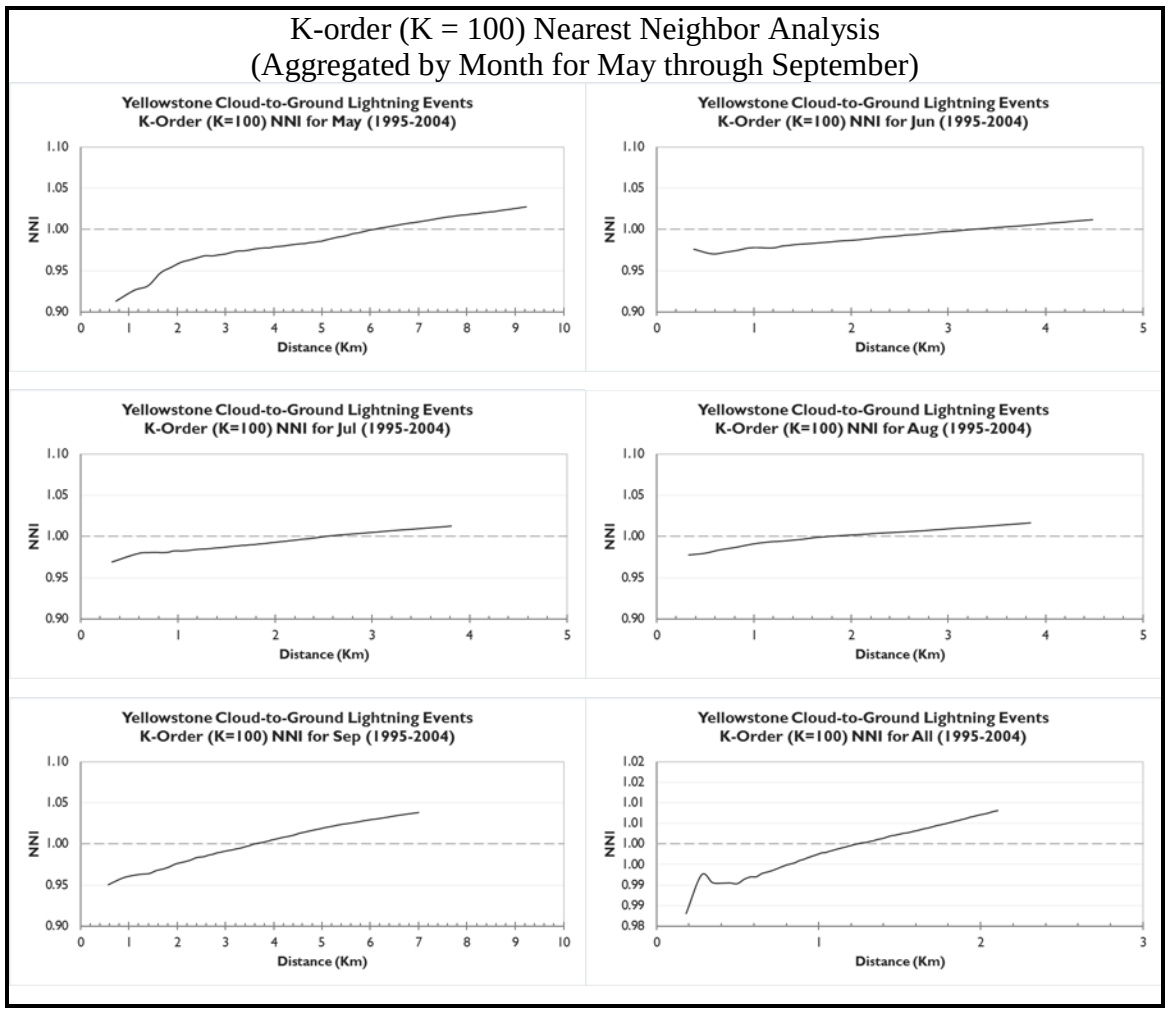

Figure 15. Series of graphs showing the K-order (K = 100) Nearest Neighbor Index (NNI) versus distance for Yellowstone Cloud-to-Ground lightning events aggregated by month for May - Sep.

Table 4. Summary results of K-order Nearest Neighbor analysis for the monthly data. N is the total number of events. K is the K^{th} nearest neighbor where the Nearest Neighbor Index (NNI) first becomes greater than or equal to 1. D is the calculated average nearest neighbor distance for K (km). NNI min and NNI max are the lowest and highest NNI for K = 1 to 100.

	N	K	D (km)	NNI min	NNI max
Mar	22	11	46.79	0.39	1.84
Apr	495	26	12.16	0.78	1.12
May	3504	46	6.08	0.91	1.02
Jun	14417	55	3.28	0.97	1.01
Jul	19936	45	2.52	0.97	1.01
Aug	19756	22	1.76	0.98	1.02
Sep	6211	29	3.62	0.95	1.04
Oct	352	35	16.74	0.75	1.20
Nov	28	25	61.44	0.58	1.68
ALL	64721	35	1.23	0.98	1.01

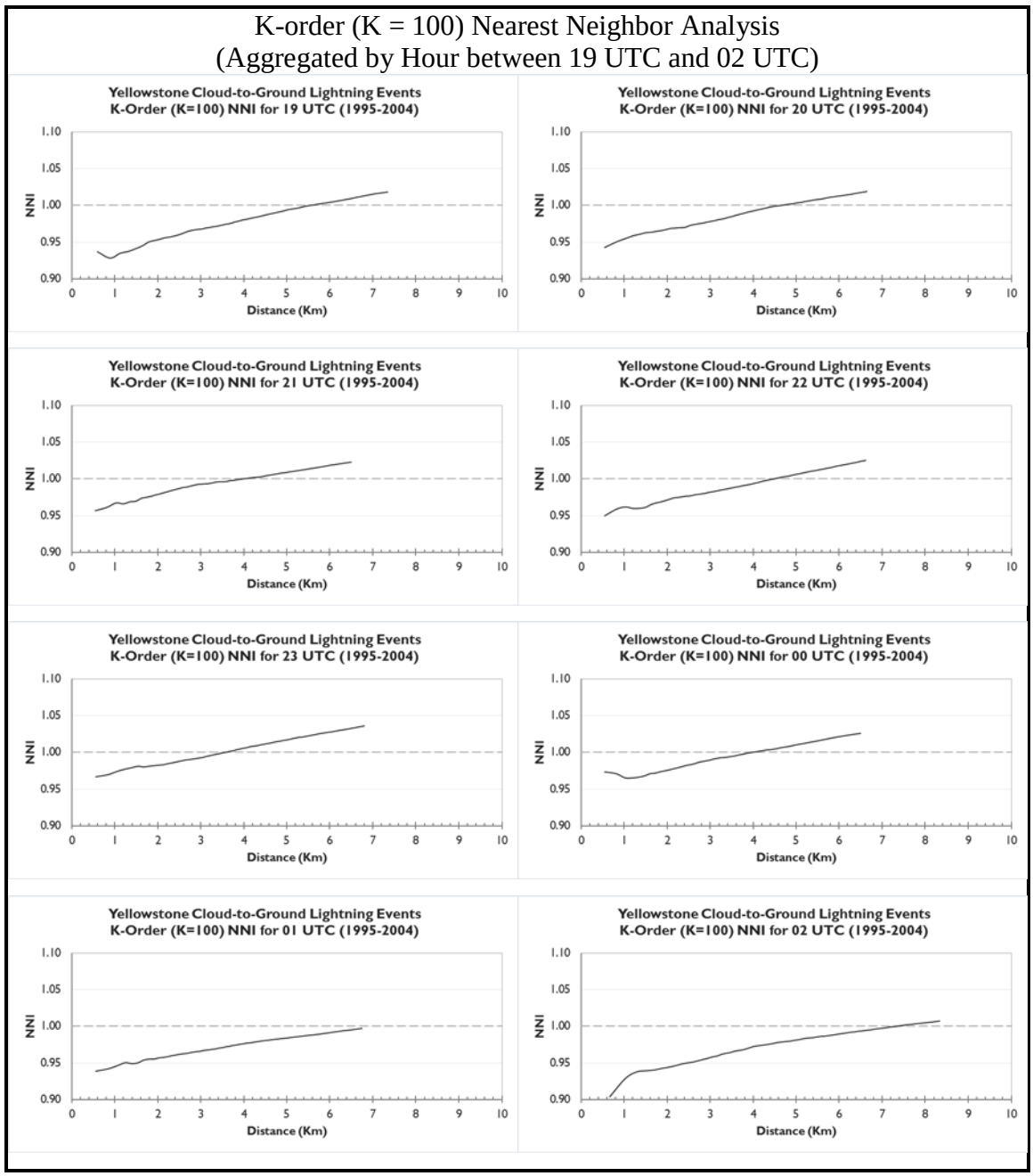

Figure 16. A series of graphs plotting the NNI versus distance for lightning events aggregated by hour of day.

Table 5. Summary results of K-order Nearest Neighbor analysis for the hourly data. N is the total number of events. K is the Kth nearest neighbor where the Nearest Neighbor Index (NNI) first becomes greater than or equal to 1. D is the calculated average nearest neighbor distance for K (km). NNI min and NNI max are the lowest and highest NNI for K = 1 to 100.

	N	K	D (km)	NNI min	NNI max
0	7048	39	3.94	0.97	1.02
1	6179	N/A	N/A	0.94	0.99
2	4117	79	7.36	0.90	1.01
3	2086	32	6.57	0.90	1.06
4	1285	49	10.37	0.88	1.06
5	1304	26	7.50	0.89	1.07
6	1047	31	9.17	0.87	1.06
7	806	25	9.34	0.88	1.10
8	426	34	15.00	0.82	1.19
9	682	52	14.67	0.85	1.09
10	446	32	14.24	0.79	1.12
11	388	30	14.75	0.73	1.25
12	571	49	15.56	0.78	1.05
13	456	81	22.43	0.74	1.07
14	369	25	13.80	0.72	1.31
15	410	48	18.17	0.77	1.04
16	681	32	11.51	0.79	1.08
17	1360	38	8.87	0.87	1.05
18	2646	36	6.19	0.95	1.04
19	5435	61	5.63	0.93	1.02
20	6651	51	4.65	0.94	1.02
21	7000	39	3.96	0.96	1.02
22	6781	49	4.52	0.95	1.03
23	6547	30	3.59	0.97	1.03

6.1.3.2 Getis-Ord General G Statistic (High/Low Clustering)

The Getis-Ord General G statistic, *G(d)*, measures the overall concentration, or lack of concentration, of all pairs of x_i and x_j that are within a specified distance (d) of each other by taking the sum of the products of each x_i with all the $x_j s$ within the specified distance. Evaluating and comparing *G(d)* results are accomplished by using their equivalent Z-scores of the approximate normal distribution. When statistically significant, positive Z-scores indicate clusters of high values and the higher the Z-score the more intense the clustering. Negative Z-scores indicate clusters of low values and the lower the Z-score the more intense the clustering of the low values. Z-scores near zero indicate a random spatial pattern (Getis-Ord, 1992).

The results of the Getis-Ord General G statistic are presented as a series of "spaghetti" charts displaying the *G(d)* index, the difference between the observed and the expected *G(d)*, and the Z-score versus distance for the aggregated monthly (Figure 17) and aggregated hourly (Figure 18) CG lightning events. For the months of May through September, the *G(d)* has extremely small values. The values are near 0.001 at distance *(d)* of 3 km and increases to near 0.22 at 25 km. These values are statistically significant at the 99% confidence level for distances out to 25 km. The Z-scores remain positive as distance increases with the exception of August where the Z-score becomes negative near 19 km. Z-score values for these months are highest for distances less than about 5 km ranging in value from 25 to 40. Sharp decreases in Z-score values are seen up to about 15 km where the decreases in Z-score are more gradual. This suggests the intensity of clustering of high values of CG Lightning is strongest for the shorter distances ($d \leq 5$). For d = 25 km, high values of CG lightning are still clustered during May, June, and

September. During July and August, the Z-scores are near zero starting at 20 km indicating a random spatial pattern at this distance. Similar results are seen when applied to the entire dataset. Graphs of Z-scores for April and October are similar to the graphs for May and September showing a higher intensity of clustering of high values at the shorter distances and the intensity of clustering decreasing by 25 km.

Similar observations are noted in the graphs for the CG lightning events aggregated by hour. The *G(d)* index starts out very small at the shorter distances (near 0.003) and increases with distance to values less than 0.26. These values are also statistically significant at the 99% confidence level for distances out to 25 km. Likewise, the intensity of the clustering of high values is more intense for distances less than 5 km based on the higher Z-scores at these distances. The Z-scores generally remain above zero except for 23 and 00 UTC where they fall slightly below zero between 15 and 20 km suggesting a random spatial pattern at these distances. The Z-score values are also close to zero for 21, 22, and 02 UTC at 25 km also indicating a random spatial pattern of CG lightning at this distance during these times.

Overall, the results from the General G tests were statistically significant at the 99 percent confidence level (high Z-scores and low p-values were noted). The Z-scores show that the magnitude of the clustering of high values varies with distance with the intensity of clustering much higher at shorter distances ($d \leq 5$ km) and lower at 25 km. In some cases, the Z-scores suggest the spatial pattern is more random between 20 and 25 km. With statistically significant results, one could normally reject the null hypothesis of a random spatial pattern of events and conclude that CG lightning is clustered when

tested using this global statistic. However, the difference between the observed and expected *G(d)* is very small. This is especially true for the results when the CG lightning is at peak activity where the differences between the observed and expected *G(d)* are less than 0.005. Differences this small makes it very difficult to reject the null hypothesis of a random spatial pattern to conclude that CG lightning is spatially clustered globally (*Research Question 1*).

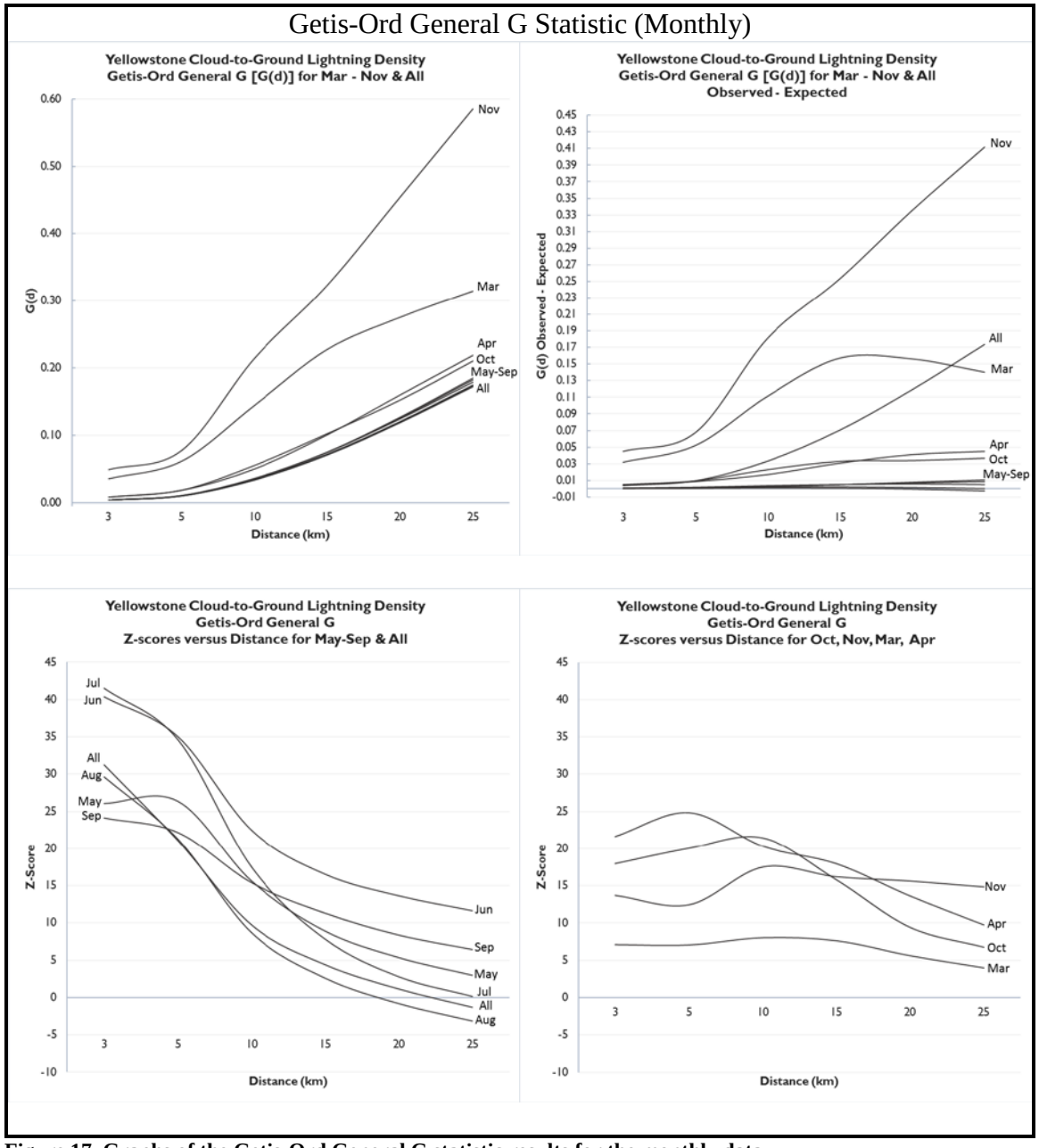

Figure 17. Graphs of the Getis-Ord General G statistic results for the monthly data.

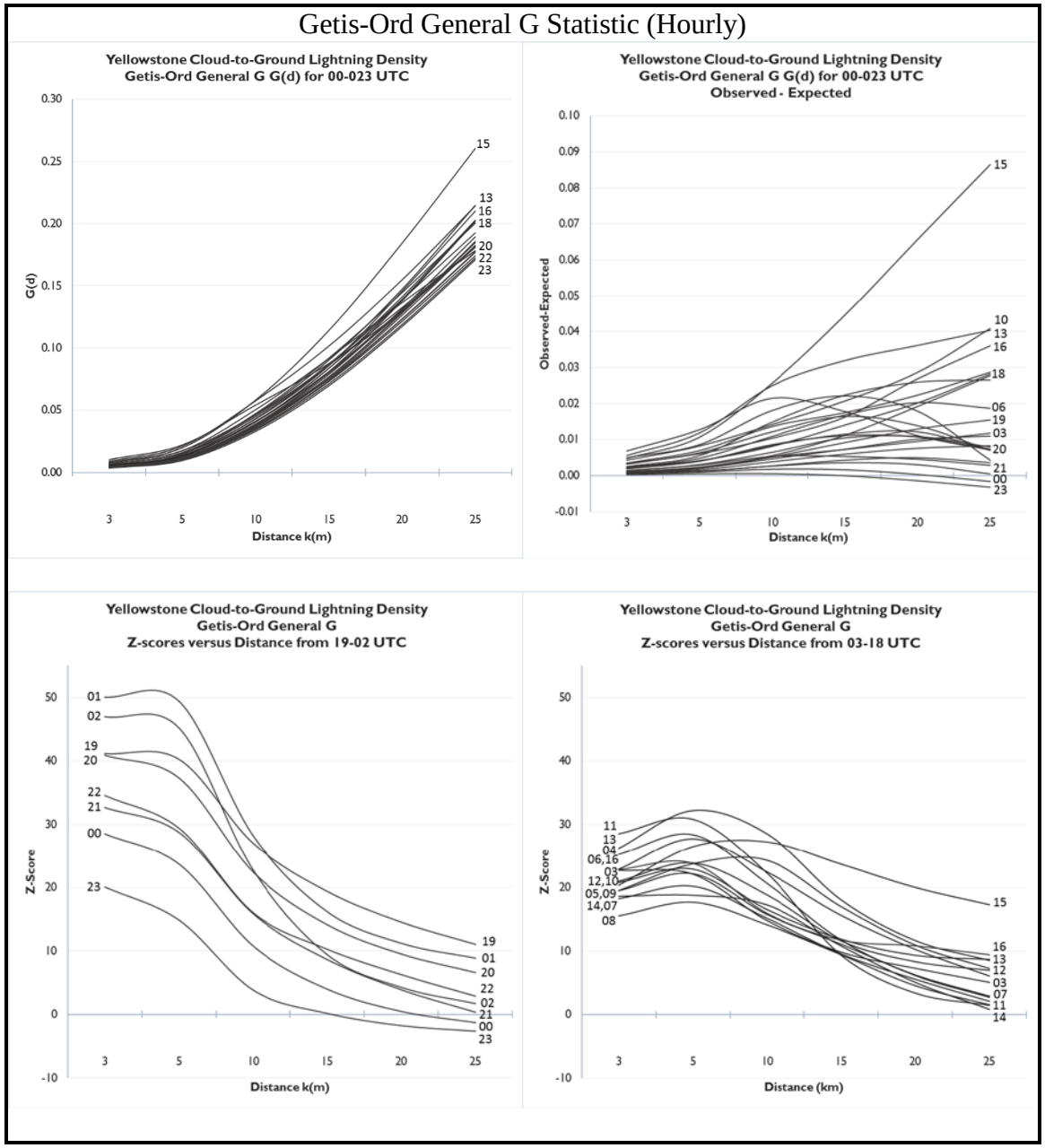

Figure 18. Graphs of the Getis-Ord General G statistic results for the hourly data.

6.1.3.3 Global Moran's I (Spatial Autocorrelation)

Whereas the Getis-Ord General G statistic measures the overall concentration of all pairs of x_i and x_j that are within a specified distance *(d)* of each other, Moran's I, *I(d)*, measures the correlation of each x_i with all x_js that are within a specified distance of each other (Getis-Ord, 1992). In this section, the results from the Global Moran's I test are presented in the same way that the results for the Getis-Ord General G statistic were presented in the previous section. A series of charts are presented that display the *I(d)* index, the difference between the observed and the expected *I(d)*, and the Z-score versus distance for the aggregated monthly (Figure 19) and aggregated hourly (Figure 20) CG lightning density.

For the aggregated monthly CG lightning data, Moran's I was greater than zero for all months tested, ranging in value from 0.06 to 0.21, indicating slight positive spatial autocorrelation or clustering of either high or low values of CG lightning density. The values of *I(d)* are greatest at 1.5 km *(d)* and decreases with distance where *I(d)* decreases to less than 0.08 indicating the spatial pattern is closer to a random pattern of CG lightning for the neighborhoods with longer distance *(d)*. Clustering appears to be the greatest at all distances during the summer months as well as for the case when the entire dataset is tested. The Z-scores are well above zero and p-values are very low indicating the results are statistically significant. The graphs of Z-scores versus distance generally increase with distance with the greatest increase seen during the summer months, especially for June and July. Z-scores are highest during June and July where they approach 150 and are lowest during May and September where they stay below 50. The Z-score curves show no distinct peaks but some minor perturbations in the curves are

seen during the summer months between 1.5 and 12 km suggesting complex spatial processes are at work. Z-scores reach their highest value where the values of *I(d)* reach their lowest suggesting the spatial pattern is much closer to a random pattern at the farther neighborhood distances. Graphs of the difference between the observed and expected *I(d)* are not that useful here. These graphs are similar to the graphs of the observed *I(d)* since the expected *I(d)* are all very near to zero.

The results for the aggregated hourly data are similar to those seen for the monthly data. The *I(d)* curves all decrease with distance ranging from 0.06 to 0.21 at 1.5 km to values less than 0.06 at 15 km indicating the spatial pattern is slightly clustered at 1.5 km and becomes closer to random as the neighborhood distance increases. The *I(d)* curves for 19 to 02 UTC all lie above the curves for the remaining hours suggesting the pattern is slightly more clustered during these times when lightning activity is the greatest. The Z-score graphs generally increase with distance with values in the 20 to 40 range at 1.5 km and reaching values between 80 and 120 during the peak hours of activity. The curves have slightly distinct peaks in Z-scores for the hours between 23 and 02 UTC where peaks in Z-scores occur between 5 and 10 km suggesting clustering is greatest between these distances during these times. Minor perturbations are also seen in the curves between 1.5 and 10 km suggesting complex spatial processes are involved.

Overall, the results from this test indicate there is very slight global clustering of either high or low values of CG lightning density. The results are statistically significant to allow one to reject the null hypothesis of a random spatial pattern, however, the values of Moran's I are generally less than 0.20 suggesting pattern of CG lightning density in

Yellowstone is actually closer to a random pattern for this spatial scale, especially when using neighborhoods with longer distances where the the Z-scores are the highest yet the Moran's I values are the lowest. (*Research Question 1*).

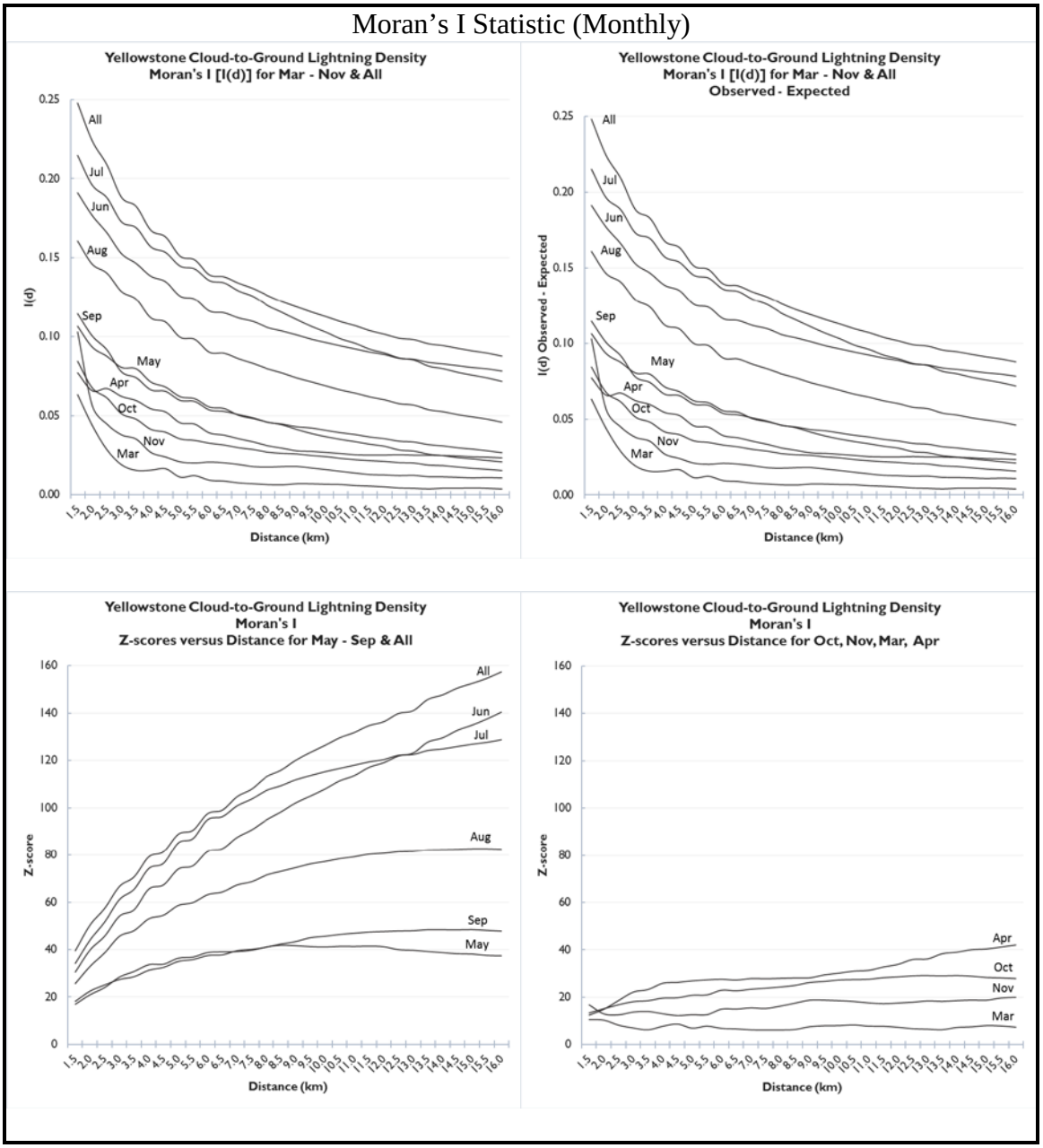

Figure 19. Graphs of the Moran's I statistic results for the monthly data.

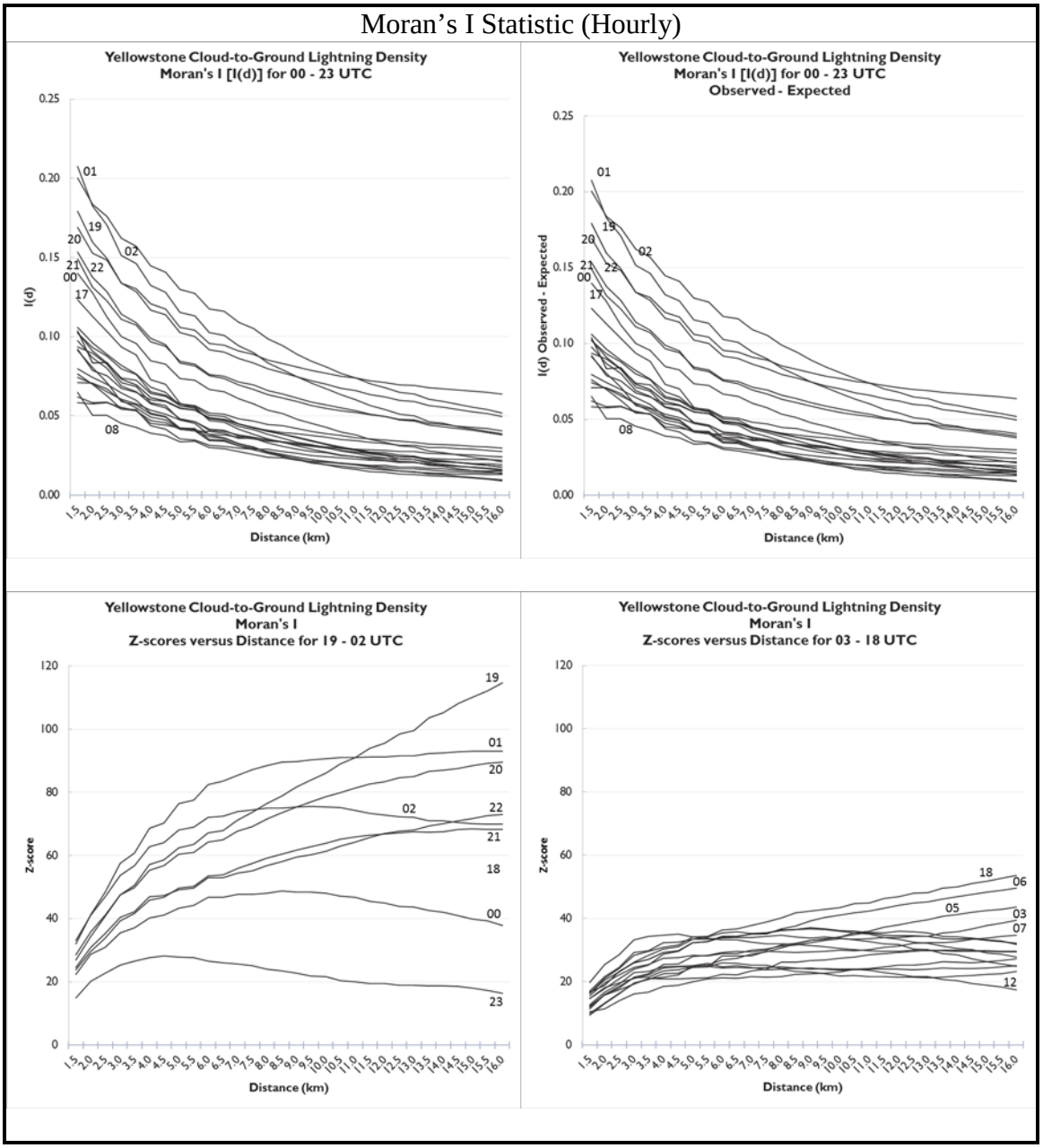

Figure 20. Graphs of the Moran's I statistic results for the hourly data.

6.1.3.4 Ripley's K (Multi-Distance Spatial Cluster Analysis):

The Ripley's K test results are visualized as series of line graphs using the transformed Ripley's K function, *L(d)*. *L(d)* is plotted along with the simulated confidence envelope with an upper bound of 97.5% and a lower bound of 2.5%. Statistically significant clustering is indicated when the transformed Ripley's K lies outside and above the envelope while statistically significant dispersion is indicated when it lies outside and below the confidence envelope. When Ripley's K falls within the envelope, the pattern is considered random.

The monthly plots of *L(d)* all increase with distance and indicate statistically significant clustering of CG lightning events for distances out to 35 km (Figure 21). Similar results are seen for the hourly (Figure 22) plots for the 19 to 02 UTC time period. The Ripley's K test shows evidence that CG lightning events are not random but are globally clustered across the Yellowstone study area (*Research Question 1*).

At first glance this may seem to counter the results seen in the NNI and K-order NNI tests. The NNI and K-order NNI are first-order tests that characterize the point pattern at the specified nearest neighbor; they only use the specified nearest neighbor for calculating the average neighbor distance. On the other hand, Ripley's K is a second-order, cumulative test based on the average distance of all neighbors that fall within the specified distance (Perry, Miller and Enright, 2006; Fortin, Dale and ver Hoef, 2002). In this case, the weak clustering seen in the NNI and K-order NNI is probably a reflection of the individual thunderstorms, as given by the shorter distances, whereas Ripley's K is probably more of a reflection of the individual thunderstorms located close to each other resulting in clustering of lightning activity at larger distances.

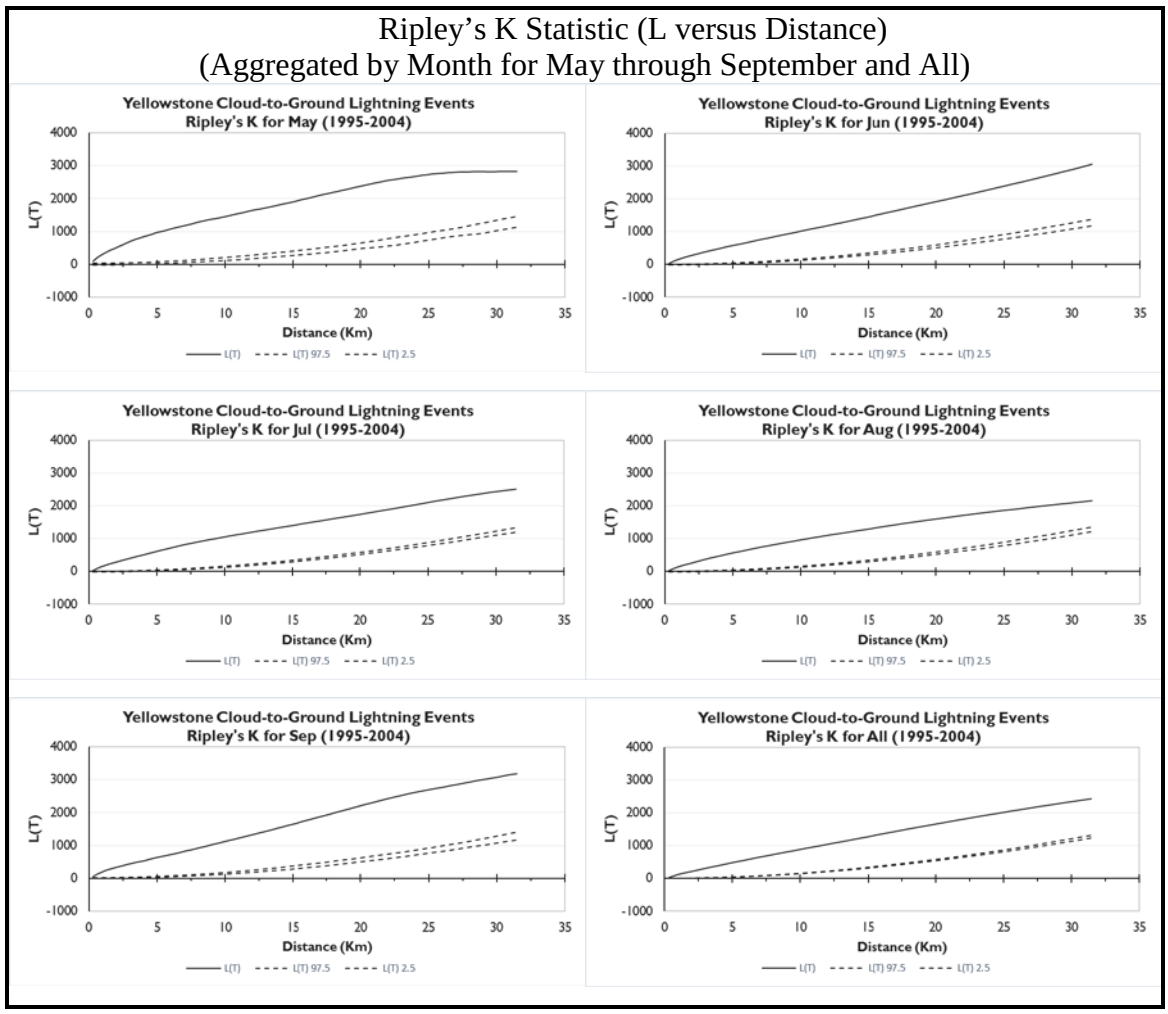

Figure 21. Graphs of *L(d)* (solid line) for CG lightning events aggregated by month. The confidence envelope of 2.5% to 97.5% is plotted as dashed lines.

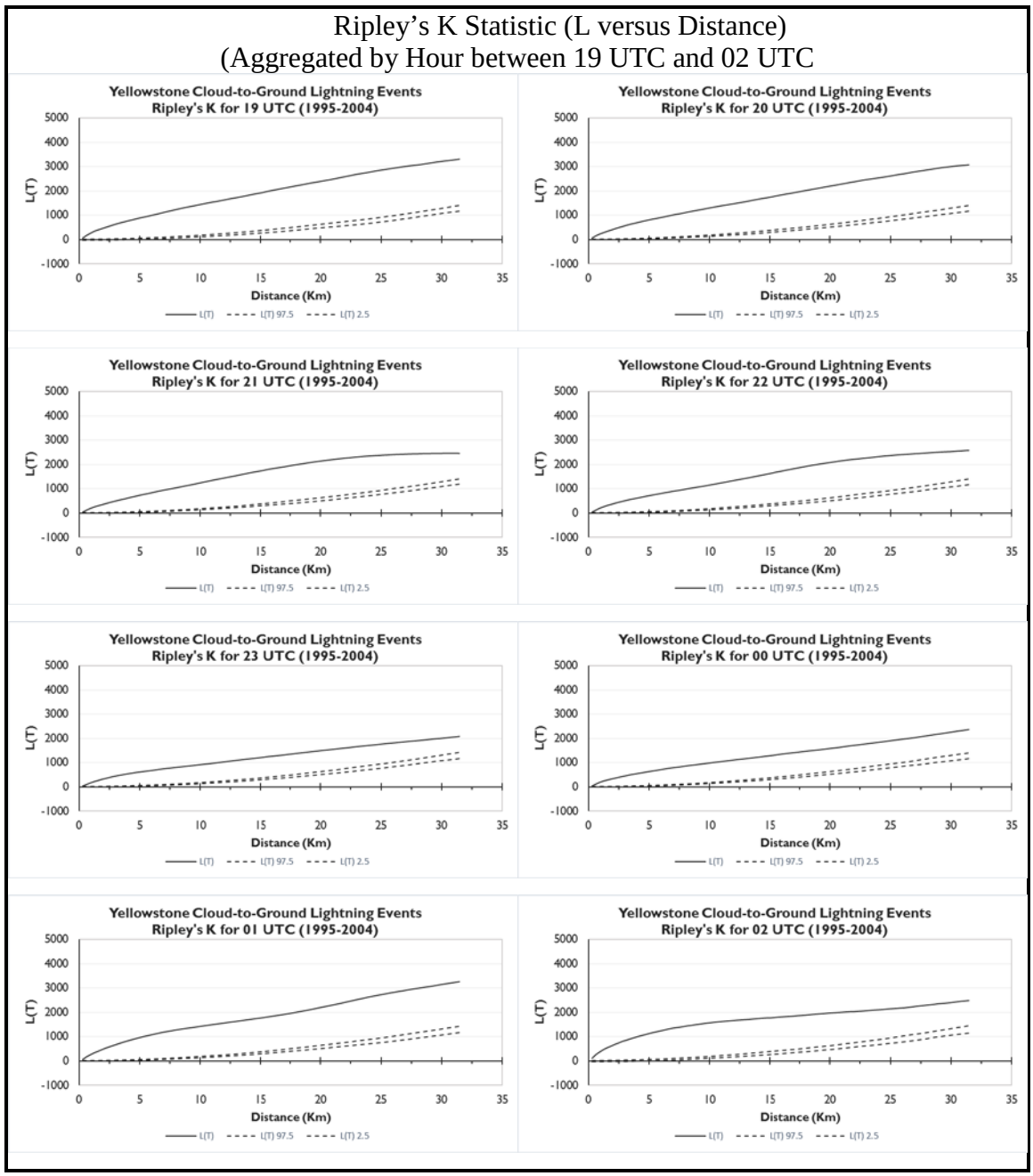

Figure 22. Graphs of *L(d)* (solid line) for CG lightning events aggregated by hour. The confidence envelope of 2.5% to 97.5% is plotted as dashed lines.

6.2 Temporal Analysis

The series of maps of the aggregated monthly and hourly lightning events and flash density presented in the previous section highlighted some aspects of the spatial distribution of these events in the park at these time scales. Figure 23 highlights the temporal distribution of the CG lightning data during the study period. Between 1995 and 2004, Yellowstone experienced the highest number of lightning strikes during 2004 (11,683 events). The lowest number of events occurred in Yellowstone during 2000 when only 3,675 events were recorded.

As shown in the lightning event and the hex flash density maps, the aggregated monthly total number of lightning flashes increases sharply as the year progresses from spring to summer. The total CG lightning events in the park peaks during July and August and then sharply declines during the fall season. During the summer months, July and August experienced an average of nearly 2,000 flashes each month (Table 6). No lightning strikes were recorded during the winter months between 1995 and 2004.

When slicing the data on the hourly time scale, as demonstrated in the previous maps, the aggregated hourly data shows a distinct diurnal pattern. The total lightning events is at its lowest during the nighttime hours between 08 UTC and 15 UTC. The highest number of events are seen between 19 UTC and 02 UTC, in the early afternoon when daytime surface heating is at its strongest and continues through to the late evening hours as surface heating wanes with the setting sun. This same general trend can be seen throughout the spring, summer, and fall seasons although it’s not as evident during the spring and fall.

In combination with the series of maps presented in the previous sections, these graphs show that CG lightning activity in Yellowstone Park is not consistent through time (*Research Question 2*). Interannual variability can be seen in the graph of total lightning events by year. This interannual variability is probably a reflection of the variability seen in the two large scale weather patterns impacting this part of the country. The first is known as the North American Monsoon (NAM) bringing moisture from the Gulf of California and the Gulf of Mexico. The extreme northern extent of the NAM just reaches the Greater Yellowstone Area (Adams and Comrie, 1997). The second climatic control is known as the Pacific-North American (PNA) teleconnection pattern that streams moisture from the Pacific Ocean through the Snake River Gap and up the western and central plateau regions of Yellowstone (Stewart et al., 2002; Mock, 1996). The variability of these two large scale weather patterns is also probably a contributing factor in the monthly variability of lightning activity although the monthly variability is most likely more of a reflection of the variability in the amount of solar heating received at the earth's surface throughout the year as the sun's position with respect to earth changes throughout the year. The diurnal variability seen is also a reflection of the daily solar heating cycle as mountain storms are driven by mountain induced convection, once the sun sets the convection ceases.

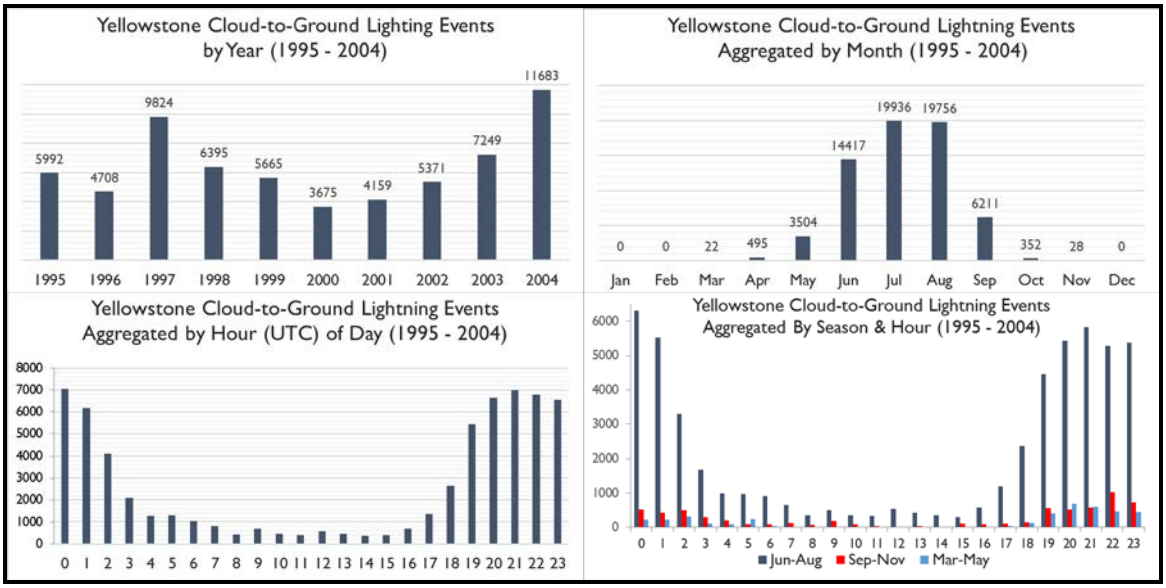

Figure 23. Graphs of lightning events by year, aggregated by month, aggregated by hour, and by season.

Table 6. Table showing the number of lightning events by year, month, and average, minimum, and maximum events. No lightning events were recorded during the 10-year period for Dec, Jan, and Feb.

	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
1995	0	0	64	469	2437	1601	1412	7	2
1996	3	2	120	1243	1153	1869	315	0	3
1997	0	62	322	2400	2308	2537	2109	81	5
1998	1	3	153	1050	2433	2053	621	81	0
1999	0	189	149	1250	1449	2378	222	28	0
2000	0	47	544	522	982	1250	219	111	0
2001	0	73	384	144	2072	915	564	1	6
2002	14	61	119	1095	2171	1791	114	1	5
2003	4	50	1312	2362	541	2772	169	32	7
2004	0	8	337	3882	4390	2590	466	10	0
Total	**22**	**495**	**3504**	**14417**	**19936**	**19756**	**6211**	**352**	**28**
Min	0	0	64	144	541	915	114	0	0
Avg	2	50	350	1442	1994	1976	621	35	3
Max	14	189	1312	3882	4390	2772	2109	111	7

6.3 Analyzing the Spatial Pattern for Local Clustering

The previous analysis demonstrated that CG lightning activity within Yellowstone National Park was at its peak during the summer months (June, July, and August) between 1995 and 2004. During the summer months, the Yellowstone weather regime is more under the influence of mountain induced convection resulting from surface heating that is strongest during the summer months. The analysis found that lightning activity is also noticeable during the seasonal transition months of May and September and that no activity to very little activity occurred during the remaining months.

The remainder of this chapter will primarily focus on presenting the spatial pattern analysis results for the peak periods of CG lightning activity, using May through September when evaluating the aggregated monthly patterns and 19 – 02UTC when analyzing the aggregated hourly data. Table 7 summarizes the spatial test statistics used for analyzing the pattern and testing whether or not the CG lightning activity is spatially clustered (*Research Question 1*) and identifying those areas that may be hot or cold spots for lightning activity in the park (*Research Question 3*).

Table 7. Table summarizing the spatial statistics used to test for local clustering in CG lightning.

TEST STATISTIC	REMARKS	INPUT	TOOL/ RESULTS
Getis-Ord Gi*	Finds locations of clusters of high values (hot spots) or low values (cold spots).	Hex Cells	ArcGISPro (Maps)
Anselin's Local Moran's I	Local indicator of spatial autocorrelation; identifies clusters of high and low values and outliers, strength of patterns.	Hex Cells	ArcGISPro (Maps)

6.3.1 Getis-Ord G_i^* (Hot Spot Analysis)

As a first step in applying the Getis-Ord G_i^* test statistic, an optimization routine was executed using ArcGIS Pro to determine an optimized set of parameters that can be used while testing both the monthly and hourly data as well as the entire dataset. This optimization routine included a test for determining an appropriate scale and the application of a false discovery rate (FDR) correction. The output from the optimization routines is provided in Table 8 and Table 9. The tables contain several columns displaying the descriptive statistics as well as the number of outliers for the flash density for the monthly or hourly data. The column labeled "Distance" indicates the optimal scale distance for that hour or month. The "D-Method" column indicates the method used by the routine to determine an optimal distance scale. The "PeakCluster" method determines an optimal distance by calculating Moran's I over a range of distances to determine distance where peak clustering occurs. The "AvgD30" method indicates the average distance to 30 nearest neighbors. The number of statistically significant cells identified as part of a cluster are shown in the "FDRSigOut" column. Of particular interest in the two tables are the distance, the distance calculation method and the number of significant resulting features. The goal was to pick one set of parameters to apply when performing hot spot analysis to aid in comparing the monthly and hourly results. The distance that occurs most in the monthly and hourly tables is 2962 m. This is also the distance that was calculated when the optimization routine was run on the entire dataset. Based on these results, a fixed distance band of 3000 m was used to perform the Getis-Ord G_i^* hot spot analysis.

Table 8. Results from the ArcGIS Pro optimization routine applied to the monthly data. Column labeled as "Distance" shows the calculated optimal scale distance for hot spot analysis. The "D-Method" column indicates how the optimal distance was determined. The "FDRSigOut" column shows how many clusters were identified as significant clusters. The remaining columns are the descriptive statistics for the flash density for that month.

MM	Min	Max	Mean	StDev	Outliers	Distance	D Method	FDRSigOut
Mar	0.0000	0.1998	0.0003	0.0055	0	4300	PeakCluster	338
Apr	0.0000	0.2998	0.0055	0.0253	0	5375	PeakCluster	1215
May	0.0000	0.6994	0.0391	0.0688	0	2962	Avg D30	974
Jun	0.0000	1.0991	0.1623	0.1483	0	2962	Avg D30	3094
Jul	0.0000	1.6986	0.2236	0.1776	0	2962	Avg D30	2889
Aug	0.0000	1.3989	0.2209	0.1672	0	2962	Avg D30	2093
Sep	0.0000	0.7994	0.0700	0.0915	0	2962	Avg D30	891
Oct	0.0000	0.3997	0.0039	0.0210	0	2962	Avg D30	480
Nov	0.0000	0.1998	0.0003	0.0058	0	3225	PeakCluster	203
All	0.0000	0.29976	0.7260	0.3303	0	2962	Avg D30	3355

Table 9. Results from the ArcGIS Pro optimization routine applied to the hourly data. Column labeled as "Distance" shows the calculated optimal scale distance for hot spot analysis. The "D-Method" column indicates how the optimal distance was determined. The "FDRSigOut" column shows how many clusters were identified as significant clusters. The remaining columns are the descriptive statistics for the flash density for that hour.

HH	Min	Max	Mean	StDev	Outliers	Distance	D Method	FDRSigOut
00	0.0000	0.9992	0.0791	0.0977	0	8600	PeakCluster	4381
01	0.0000	0.8993	0.0694	0.0960	0	2962	Avg D30	2943
02	0.0000	1.1990	0.0463	0.0800	0	9675	PeakCluster	5708
03	0.0000	0.5995	0.0232	0.0517	0	2962	Avg D30	596
04	0.0000	0.3997	0.0114	0.0403	0	4190	PeakCluster	4190
05	0.0000	0.5995	0.0146	0.0404	0	2962	Avg D30	565
06	0.0000	0.3997	0.0117	0.0365	0	2962	Avg D30	722
07	0.0000	0.2998	0.0090	0.0308	0	7525	PeakCluster	3388
08	0.0000	0.2998	0.0048	0.0223	0	2962	Avg D30	600
09	0.0000	0.2998	0.0077	0.0286	0	2962	Avg D30	602
10	0.0000	0.2998	0.0050	0.0232	0	6450	PeakCluster	750
11	0.0000	0.2998	0.0044	0.0221	0	7525	PeakCluster	1933
12	0.0000	0.3997	0.0065	0.0271	0	4300	PeakCluster	698
13	0.0000	0.2998	0.0051	0.0240	0	6450	PeakCluster	2042
14	0.0000	0.3997	0.0041	0.0219	0	2962	Avg D30	604
15	0.0000	0.2998	0.0046	0.0224	0	2962	Avg D30	645
16	0.0000	0.2998	0.0077	0.0295	0	5375	PeakCluster	1554
17	0.0000	0.4996	0.0154	0.0421	0	4300	PeakCluster	1350
18	0.0000	0.4996	0.0298	0.0579	0	2962	Avg D30	647
19	0.0000	0.6994	0.0606	0.0876	0	2962	Avg D30	2524
20	0.0000	0.8993	0.0745	0.0976	0	2962	Avg D30	1816
21	0.0000	0.9992	0.0784	0.0989	0	2962	Avg D30	1433
22	0.0000	0.7994	0.0762	0.0974	0	2962	Avg D30	1729
23	0.0000	0.5995	0.0733	0.0908	0	4300	PeakCluster	1696

Results of the Getis-Ord G_i^* statistical analysis on the entire 10-year CG lightning event dataset are mapped in Figure 24. The map graphic highlights regions of statistically significant clusters of high and low values (hot and cold spots) in shades of red and blue, depending on the calculated Z-score and p-values. Most of the larger clusters identified as hot spots are found in the more mountainous regions of the park such as in the northeastern through southeastern parts of the park. Clusters of statistically significant low values, or cold spots, are mainly located in the plateau regions in the western and central parts of the park as well as in the extreme northern section in the area of the park's north entrance. The Getis-Ord G_i^* statistic identified 2,363 cells as statistically significant clusters at the 99 percent confidence level and 1,317 of these cells were identified as clusters of high values.

The series of map graphics depicting the results of the Getis-Ord G_i^* statistic for the aggregated monthly data shows similar results during the summer months (Figure 25). The number of hex cells identified as clusters of either high or low values is much smaller during the season transitional months of May and September and more cells are identified as clusters of high values for all months between May and September. Table 10 shows the count and percentage of hex cells identified as clusters. The count and percentage of high and low clusters are also shown. June and July have the most cells identified as clusters and the most clusters identified as hot spots. The hot spots are not as extensive and are located in the extreme southern portion of the park during May and more central over the plateau region in September. During the summer months when lightning activity is at its peak, the larger hot spots are located in the more mountainous

region of the park along the Absoroka range extending from the northeast to the southeastern part of the park. The extreme northeastern and southeastern parts of the park consistently show high values of lightning flash density. Cold spots of lightning flash density are larger in coverage during June and July. Of the summer months, August appears to have the smallest coverage in both hot and cold spots of flash density.

During the most active hours of the day (19 UTC to 02 UTC), the northeastern mountains have consistent high levels of lightning flash activity during this time period through 02 UTC where it's almost non-existent. The hot spot coverage in this region of the park is largest between 19 UTC and 21 UTC (Figure 26). In the mountains of southeastern Yellowstone, the hot spot activity is consistent but relatively smaller in coverage except during 01 UTC where coverage is at its greatest in this region. Cold spot coverage appears to be at its greatest during 19 UTC and 01 UTC and appears to occur mostly over the plateau region of Yellowstone.

Overall, graphics mapping the results of the Getis-Ord G_i^* test static reveal localized hot and cold spots of CG lightning activity within Yellowstone National Park (*Research Question 3*). The localized hot and cold spots vary with the seasons on the monthly time scale and varies diurnally with hour of day. The hot spots are generally found in or near the mountains while most of the cold spots of CG lightning flash density are found in the plateaus in the west and central part of Yellowstone.

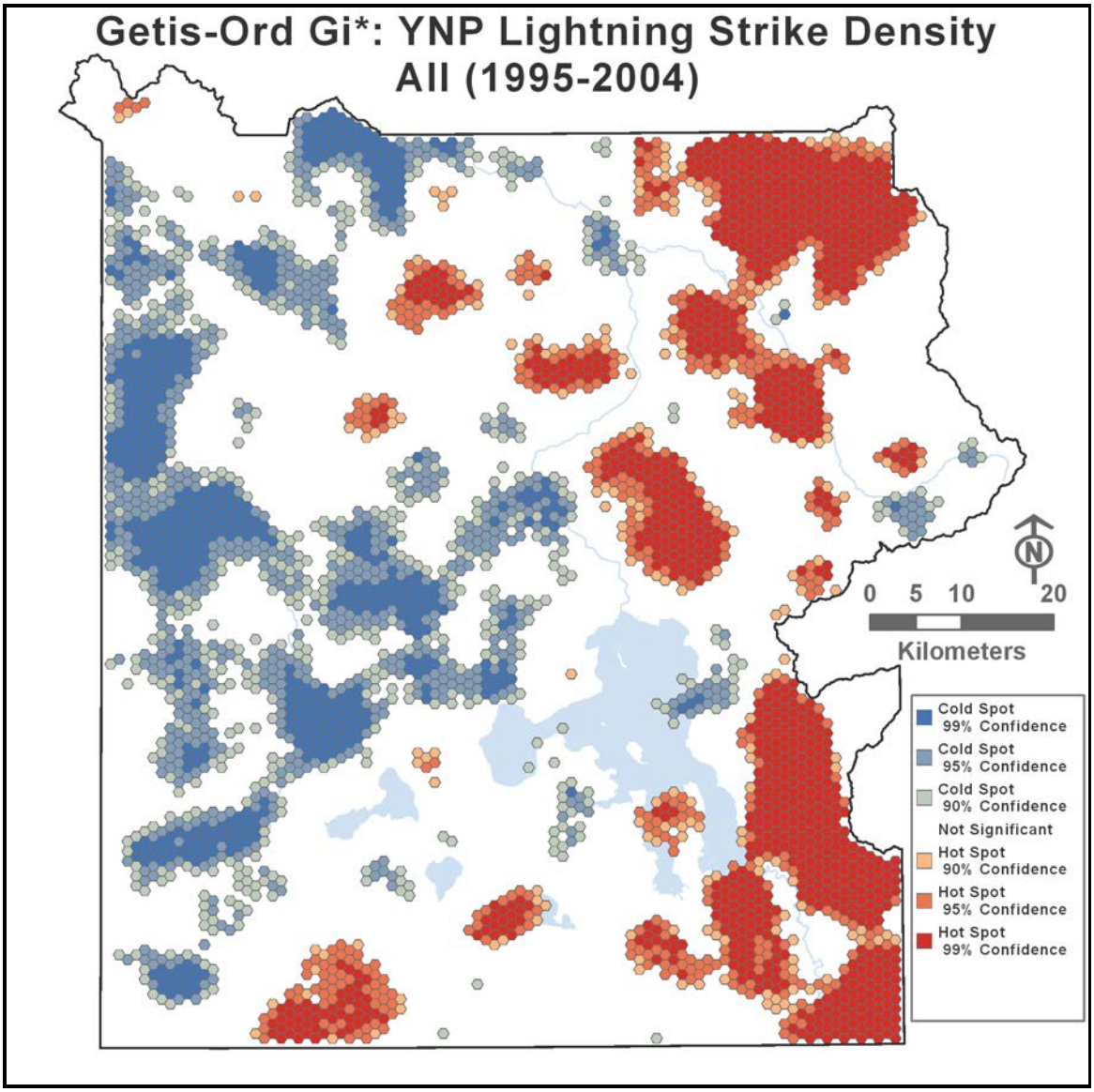

Figure 24. Results from the Getis-Ord Gi* statistic tested on the entire CG lightning event dataset showing flash density hot and cold spots.

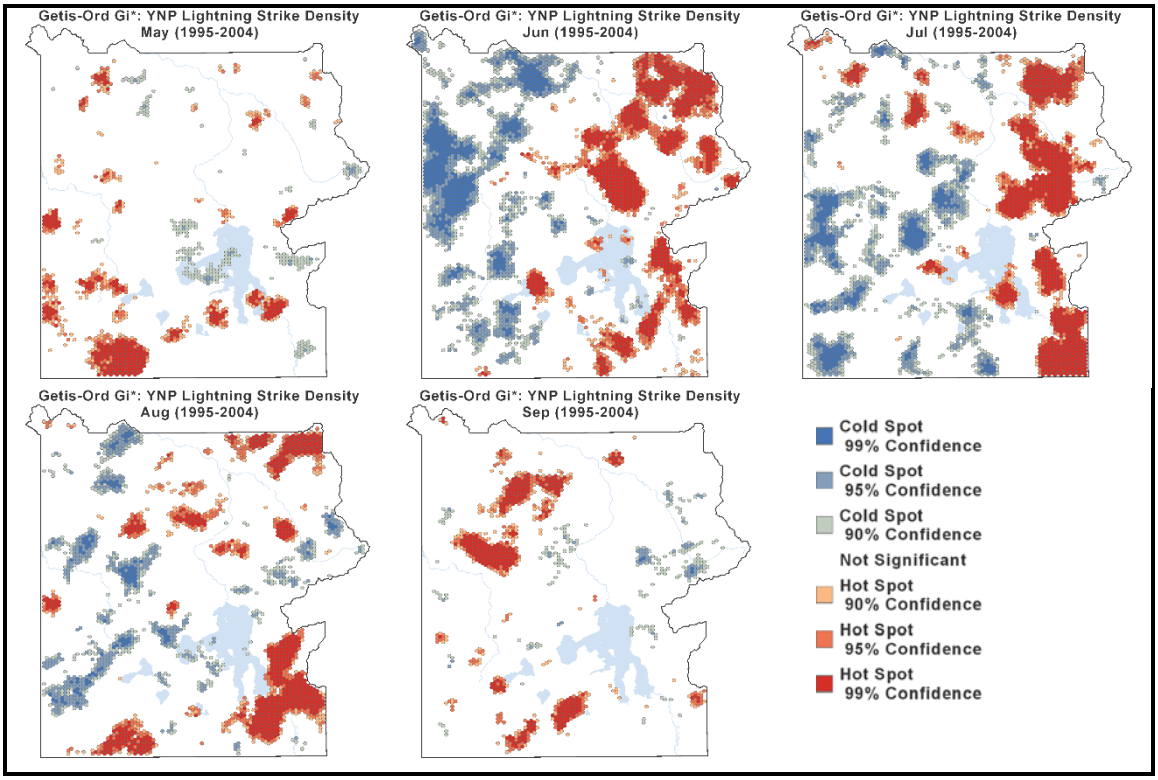

Figure 25. Results from the Getis-Ord Gi* statistic applied to the aggregated monthly CG lightning event data showing hot and cold spots of lightning flash density by month (May – Sep).

Table 10. Summary table of the monthly and hourly Getis-Ord Gi* results. N is the total number (percentage) of the 8,624 hex cells that were identified as a cluster. High shows the number (percentage) of identified clusters labeled as clusters of high values. Low shows the number (percentage) of identified clusters labeled as clusters of low values.

	N	High	Low
May	855 (10)	615 (72)	240 (28)
Jun	2120 (25)	1161 (55)	959 (45)
Jul	1947 (23)	1202 (62)	745 (38)
Aug	1555 (18)	965 (62)	590 (38)
Sep	891 (10)	664 (75)	227 (25)
All	2363 (27)	1317 (56)	1045 (44)
19 UTC	1747 (20)	1102 (63)	645 (37)
20 UTC	1415 (16)	875 (62)	540 (38)
21 UTC	1157 (13)	805 (70)	352 (30)
22 UTC	1251 (15)	880 (70)	371 (30)
23 UTC	886 (10)	587 (66)	299 (34)
00 UTC	1191 (14)	803 (67)	388 (33)
01 UTC	1713 (20)	1173 (68)	540 (32)
02 UTC	857 (10)	745 (87)	112 (13)

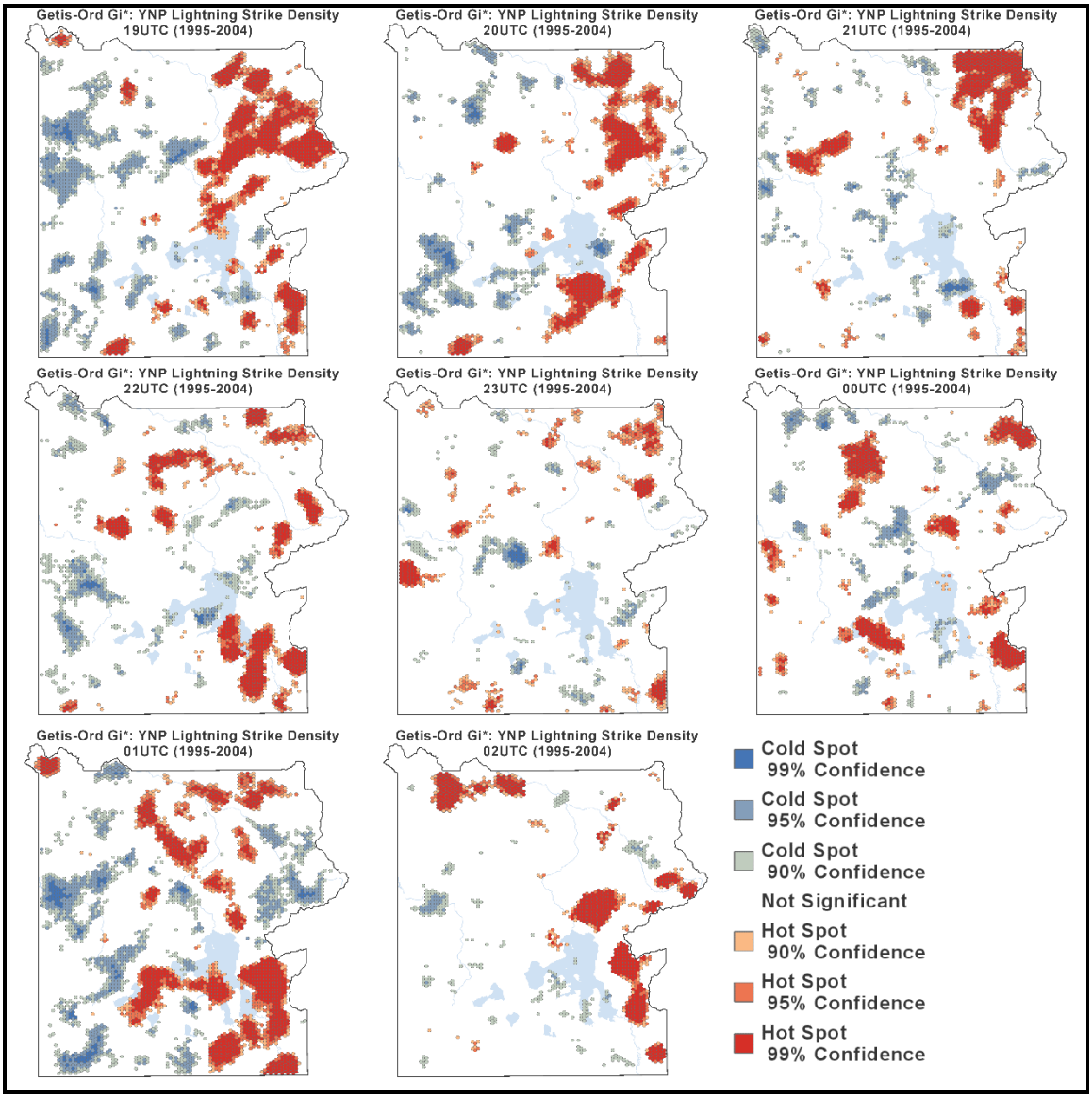

Figure 26. Results from the Getis-Ord Gi* statistic tested on the aggregated hourly CG lightning event data showing hot and cold spots of lightning flash density by hour (19UTC – 02UTC).

6.3.2 Anselin Local Moran's I (Cluster and Outlier Analysis)

Clustering of high and low CG lightning activity can also be seen in the maps of the Anselin Local Moran's I spatial statistic (Figure 27). Results using the entire dataset identified clusters of high values mainly over the eastern part of the park where the Absaroka mountain range is located and clusters of low values mainly in the plateau regions of Yellowstone. Several hexagon cells are identified as outliers, either as a low value surrounded by high values or as a high value surrounded by low values. Table 11 shows the count and percentage of hex cells identified as clusters. The count and percentage of clusters identified as high values surrounded by high values or low values surrounded by low values are also shown. The table also includes counts of cells identified as outliers, either low values surrounded by high values or high values surrounded by low values. The Local Moran's I highlighted several hundred more hex cells than did the Getis-Ord G_i^*. However, a comparison of the "HH, LL" column from Table 11 with the total identified cells from Table 10 shows the Getis-Ord G_i^* identified more hex cells as clusters than Local Moran's I. These differences in numbers are probably due to the slight difference in the structure of the two statistics, sum of products versus sum of covariances.

As seen in the results for the Getis-Ord G_i^*, a few clusters of high values are indicated by the Local Moran's I statistic during May and September (Figure 28). During May, a cluster of high values is located in the area of the southern border of Yellowstone. During June and July, the larger clusters of high values are located in the mountainous regions in the northeastern through southeastern parts of Yellowstone. A few clusters of low values are seen during July and August, although not as extensive as seen in the map

of the results for the entire dataset. Low value outliers and high value outliers are seen throughout the map series. However, these outliers appear more dispersed throughout the study area.

The clustering pattern for the aggregated hourly data is similar to that observed in the hourly Getis-Ord G_i^* results (Figure 29). Clusters of high values are found in the mountainous northeastern part of Yellowstone between 19 UTC and 21 UTC. The clusters of high values decrease in coverage in this area of the park by 01 UTC. High value clustering is at its greatest in southeastern Yellowstone at 01 UTC. Clustering of low values is at its greatest extent during 19 UTC and 01 UTC in the western plateau region of the park. During 01 UTC, a relatively large cluster of low values appears in the extreme east-central part of Yellowstone.

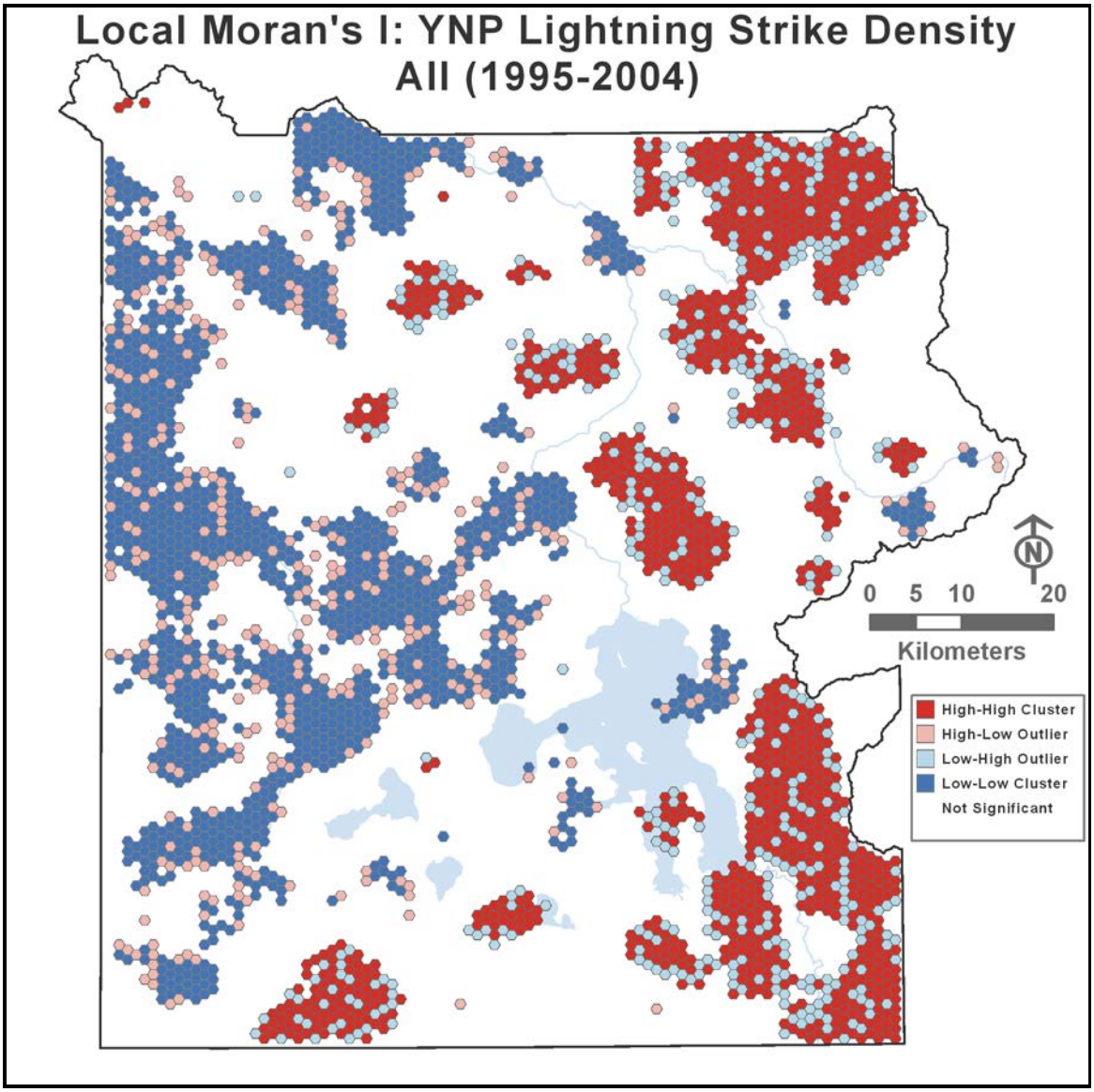

Figure 27. Results of Local Moran's I applied to the entire 10-year lightning dataset showing clusters of high and low flash density and outliers.

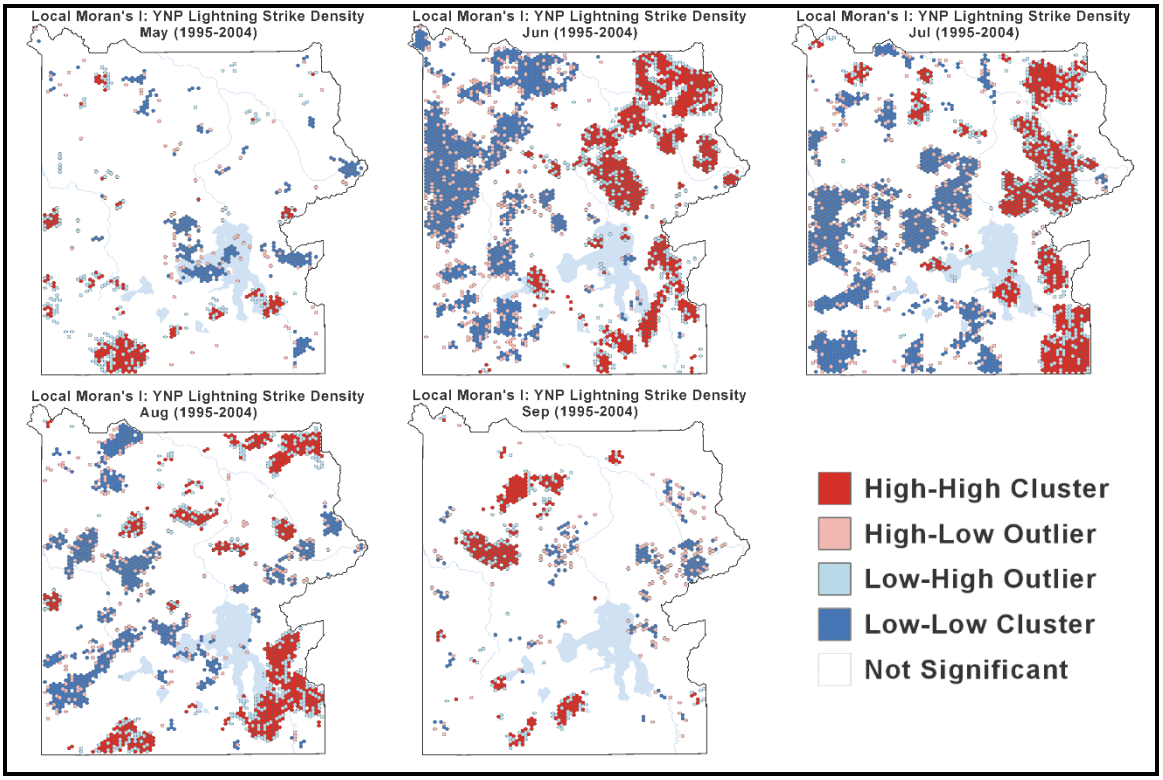

Figure 28. Local Moran's I results applied to the 10-year lightning dataset aggregated by month (May – Sep) showing clusters of high and low flash density and outliers.

Table 11. Summary table of the monthly and hourly Local Moran's I results. N is the total number (percentage) of the 8,624 hex cells that were identified as a cluster. HH, LL shows the number (percentage) of identified clusters labeled as clusters of either high values surrounded by high values or low values surrounded by low values. HL, LH shows the number (percentage) of identified clusters labeled as clusters of high values surrounded by low values or low values surrounded by high values.

	N	HH, LL	HL, LH
May	1226 (14)	796 (65)	430 (35)
Jun	2517 (29)	1793 (71)	724 (29)
Jul	2428 (28)	1768 (73)	660 (27)
Aug	1912 (22)	1337 (70)	575 (30)
Sep	1208 (14)	748 (62)	460 (38)
All	2755 (32)	2055 (75)	700 (25)
19 UTC	2201 (26)	1522 (69)	679 (31)
20 UTC	1814 (21)	1237 (68)	577 (32)
21 UTC	1522 (18)	1014 (67)	508 (33)
22 UTC	1678 (19)	1098 (65)	580 (35)
23 UTC	1168 (14)	720 (62)	448 (38)
00 UTC	1602 (19)	1041 (65)	561 (35)
01 UTC	2355 (27)	1619 (69)	736 (31)
02 UTC	1340 (16)	952 (71)	388 (29)

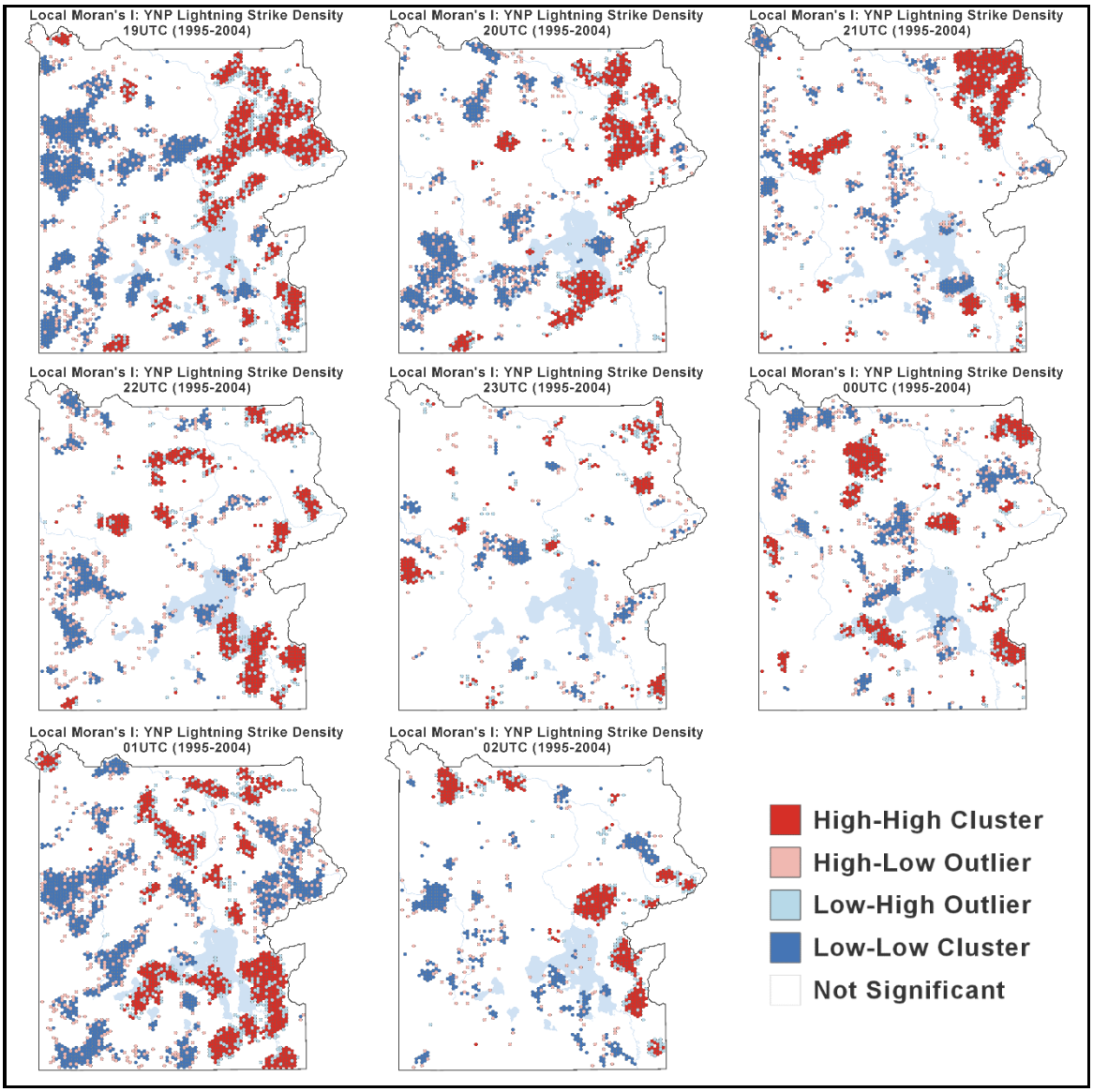

Figure 29. Local Moran's I results applied to the 10-year lightning dataset aggregated by hourly (19UTC – 02UTC) showing clusters of high and low flash density and outliers.

Maps of the Getis-Ord G_i^* and local Moran's I output are displayed side by side for the entire 10-year dataset (Figure 30) as well as for the monthly results (Figure 31). In the side-by-side maps it is apparent that the two spatial techniques identified similar clusters of hot and cold spots (or clusters of high and low values) and the clusters identified also have similar spatial extents.

Overall, graphics mapping the results of the local Moran's I test statistic reveal localized statistically significant clusters of high and low values of CG lightning activity within Yellowstone National Park (*Research Question 3*). The localized clusters vary monthly with the seasons and diurnally with hour of day.

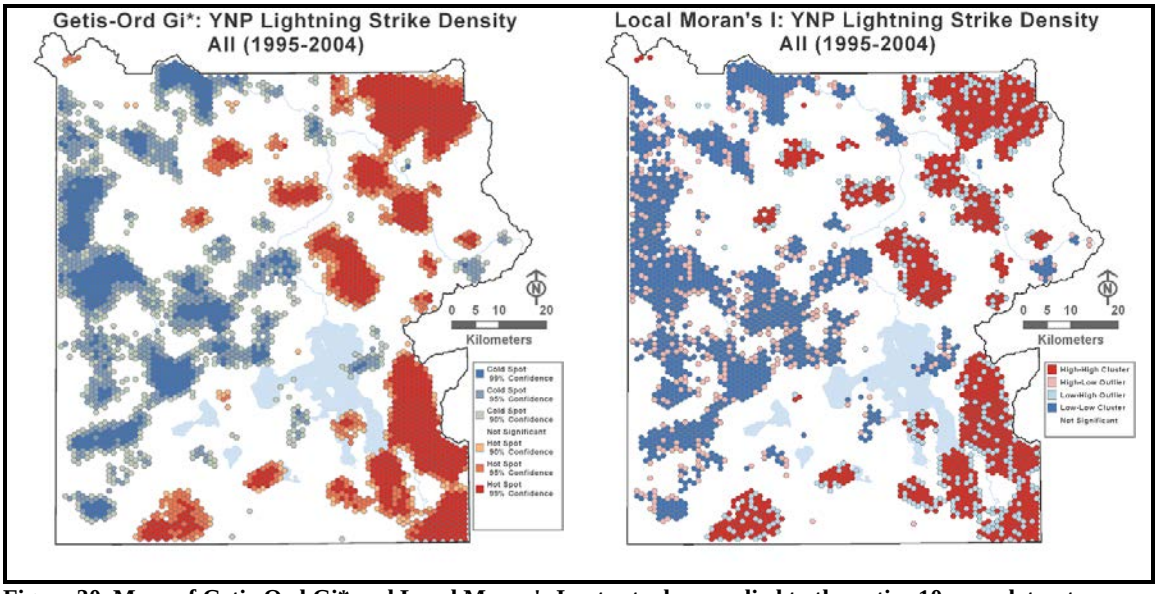

Figure 30. Maps of Getis-Ord Gi* and Local Moran's I output when applied to the entire 10-year dataset displayed side by side for visual comparison of identified hot/cold spots and high/low clusters of flash density.

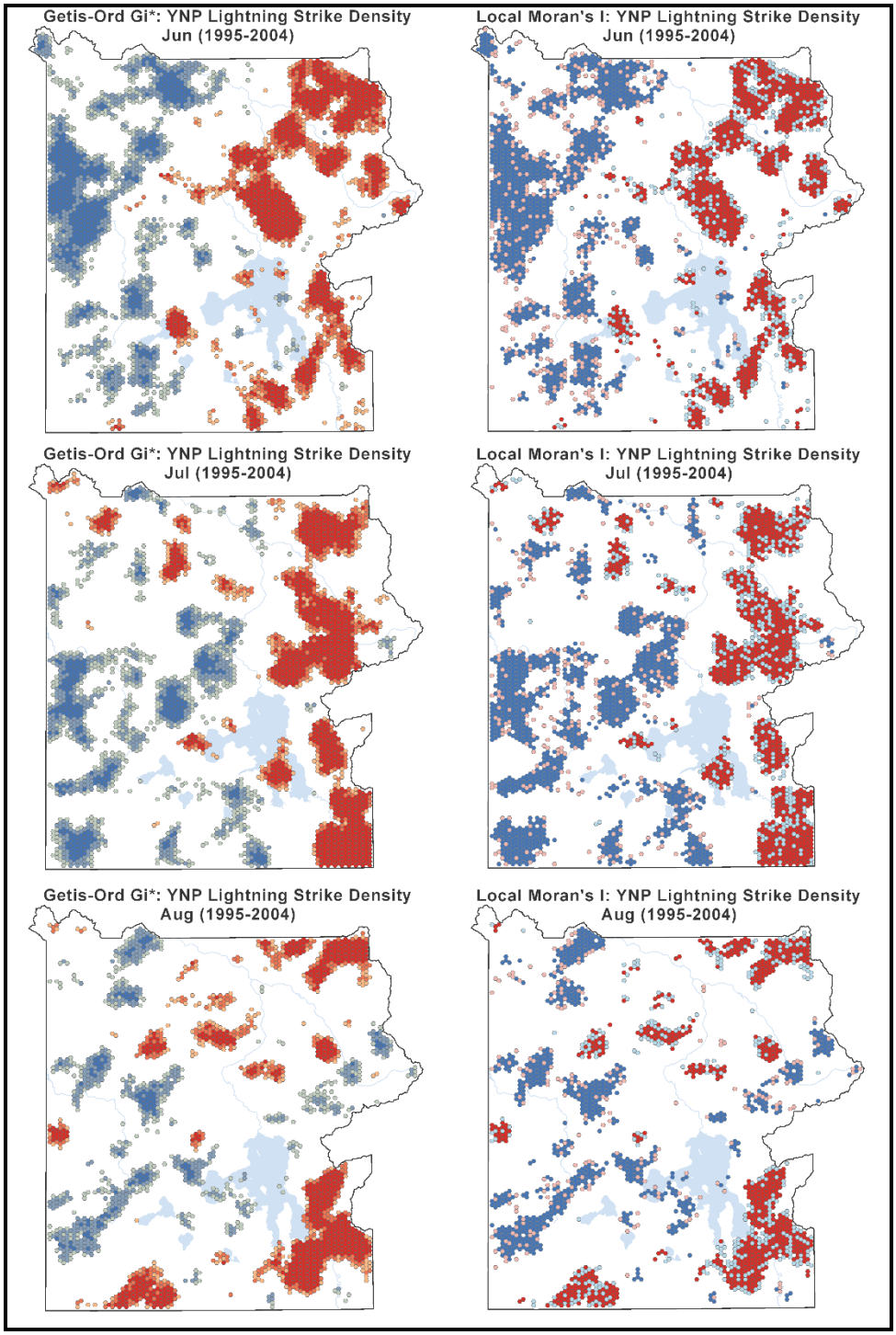

Figure 31. Maps of Getis-Ord Gi* and Local Moran's I output when applied to the aggregated monthly data displayed side by side for visual comparison of identified hot/cold spots and high/low clusters of flash density.

6.4 Analyzing the Relationship Between CG Lightning and Topography

Overall, the regression analysis did provide some information regarding the relationship between the transformed flash density variable and the terrain parameters-land cover, aspect, elevation, and slope. Both the transformed flash density variable and the elevation variable have an approximately normal distribution while the aspect variables had close to a uniform distribution. The slope and land cover variables were skewed slightly to the right. OLS regression (Table 12 and Table 13) yielded models for the aggregated monthly data that could only explain 1 to 4 percent of the variation in CG lightning flash density during the peak activity months based on the resulting R^2 values. During the non-peak months, less than 1 percent of the variation in lightning flash density could be explained by the physical terrain properties. For the case where the entire dataset was used in regression analysis, the resulting model explains 8 percent of the variability.

The predictor variables used to predict flash density all have very low variance inflation factors (VIF) indicating very little redundancy among the explanatory variables. The terrain aspect variables do not appear to be statistically significant in predicting lightning flash density. The land cover variables were statistically significant ($p \leq 0.05$) for most of the months as well as in the model based on the entire dataset and were mostly a positive relationship suggesting cells with higher percentage of forest or shrub had higher flash density (except during July where the relationship is slightly negative). Slope is mainly statistically significant during the summer months also having a positive relationship where hex cells with higher slope values had higher CG lightning flash density values. Elevation is also statistically significant ($p \leq 0.05$) for most months as

well as in the model using the entire dataset. The elevation coefficients are positive suggesting that hex cells with higher elevation values have higher CG lightning flash density.

Each of the models were identified as statistically significant with very low p values and the models for the high CG lightning activity months have very high F values meaning the model explains a non-zero amount of the variation in lightning density. However, these models have issues limiting their use in explaining the relationship between terrain and CG lightning flash density. The Jarque-Bera test statistic for each of the models were all statistically significant at the 99 percent confidence level indicating the regression residuals were not normally distributed. The histogram of the standardized residuals (not shown) approximate a normal curve but the scatter plot of standardized residuals versus predicted (not shown) shows a pattern with structure. Furthermore, the Breusch-Pagan (B-P) and the Koenker-Bassett (K-B) were also statistically significant with p-values much smaller than 0.01 allowing one to reject a null hypothesis of homeoskedasticity. The diagnostics for spatial dependence (not shown) showed evidence of spatial autocorrelation with statistically significant Moran's I for the residuals with p-values much smaller than 0.01.

Table 12. Classic OLS regression diagnostics for the data aggregated by month.

Classic OLS Regression Diagnostics							
	R^2 F-stat p-value	LL	Akaike	Schwarz	Jarque-Bera p-value	B-P p-value	K-B p-value
Apr	0.010259 14.8869 <0.001	10431	-20849	-20800	125959 <0.001	391813833 <0.001	41857188 <0.001
May	0.007871 11.3931 <0.001	3230	-6446	-6397	1731 <0.001	726234 <0.001	880295 <0.001
Jun	0.033186 49.2971 <0.001	916	-1819	-1770	261 <0.001	14660177 <0.001	22238500 <0.001
Jul	0.042335 63.4879 <0.001	930	-1846	-1796	310 <0.001	40636790 <0.001	42887062 <0.001
Aug	0.017072 24.9443 <0.001	1118	-2222	-2173	413 <0.001	15155024 <0.001	15564236 <0.001
Sep	0.002195 3.15879 0.004272	1792	-3570	-3520	848 <0.001	120156673 <0.001	320526945 <0.001
Oct	0.004810 6.94078 <0.001	11818	-23622	-23572	246811 <0.001	1270107980 <0.001	96923997 <0.001
ALL	0.080835 126.302 <0.001	2099	-4184	-4134	107 <0.001	711118 <0.001	588241 <0.001

Table 13. Classic OLS regression model parameters for data aggregated by month.

Classic OLS Regression Model Parameters								
		Constant	**Slope**	**Elevation**	**NSAspect**	**EWAspect**	**%Forest**	**%ShrubGrass**
Apr	Coefficient	0.05336	1.41e-005	1.84e-005	-0.00194	-0.00101	0.00041	0.00026
	t-statistic	6.08352	0.15512	-5.05921	-1.73659	-0.92919	7.34732	4.48401
	p-value	<0.001	0.87653	<0.001	0.08248	0.35280	<0.001	<0.001
	VIF		1.04175	1.08086	1.00699	1.00738	1.04553	1.09074
May	Coefficient	0.04706	1.39e-005	1.69e-005	-0.00184	0.00211	0.00096	0.00048
	t-statistic	2.32735	0.06655	2.01913	-0.71476	0.84091	7.41191	3.49506
	p-value	0.01997	0.94580	0.04352	0.47469	0.40036	<0.001	<0.001
	VIF		1.04175	1.08086	1.00699	1.00738	1.04553	1.09074
Jun	Coefficient	-0.06722	0.00164	0.00015	0.00324	-0.00170	0.00014	0.00013
	t-statistic	-2.54247	6.01805	14.3295	0.961511	-0.51761	0.82904	0.76655
	p-value	0.01103	<0.001	<0.001	0.33630	0.60461	0.40712	0.44336
	VIF		1.04175	1.08086	1.00699	1.00738	1.04553	1.09074
Jul	Coefficient	-0.00796	0.00178	0.00017	0.00366	-0.00711	-0.00089	-0.00048
	t-statistic	-0.301738	6.52659	15.8385	1.08808	-2.16506	-5.23382	-2.69201
	p-value	0.76303	<0.001	<0.001	0.27657	0.03041	<0.001	0.00712
	VIF		1.04175	1.08086	1.00699	1.00738	1.04553	1.09074
Aug	Coefficient	0.15981	0.00127	9.57e-005	0.00040	0.00249	0.00067	0.00040
	t-statistic	6.18693	4.75617	8.90260	0.12235	0.77647	4.07568	2.30866
	p-value	<0.001	<0.001	<0.001	0.90287	0.43745	<0.001	0.02098
	VIF		1.04175	1.08086	1.00699	1.00738	1.04553	1.09074
Sep	Coefficient	0.09435	-0.00018	3.15e-005	0.00296	-0.00183	0.00031	0.00025
	t-statistic	3.94969	-0.75514	3.17055	0.97203	-0.61667	2.01611	1.58354
	p-value	<0.001	0.45022	0.00153	0.33112	0.53757	0.04382	0.11333
	VIF		1.04175	1.08086	1.00699	1.00738	1.04553	1.09074
Oct	Coefficient	-0.00795	0.00020	5.62e-006	0.00180	-0.00210	0.00014	0.00019
	t-statistic	-1.06529	2.58572	1.81073	1.89128	-2.26698	3.11253	3.87742
	p-value	0.28680	0.00973	0.07023	0.05864	0.02342	0.00186	<0.001
	VIF		1.04175	1.08086	1.00699	1.00738	1.04553	1.09074
ALL	Coefficient	0.26290	0.00221	0.00021	0.00211	-0.00219	0.00048	0.00036
	t-statistic	11.4035	9.26592	22.8258	0.71804	-0.76541	3.24055	2.35903
	p-value	<0.001	<0.001	<0.001	0.47270	0.44396	0.00120	0.01834
	VIF		1.04175	1.08086	1.00699	1.00738	1.04553	1.09074

According to the diagnostics for spatial dependence, the spatial error model was appropriate to address the problems with spatial autocorrelation in the OLS models. Spatial error regression demonstrated some improvement in the ability to explain the variance in lightning flash density. During the peak months, spatial error regression resulted in models that could explain roughly 10 percent of the variability (Table 14). During non-peak months, spatial error models explained less than 4 percent of the

lightning flash density variation. When applied to the entire dataset, the resulting model explains 17 percent of the variability in flash density. The Akaike Information Criterion (AIC) is smaller for the spatial error models than the OLS models, indicating the error model is a better fit than the linear model. The coefficient for the lambda variable representing the spatially autocorrelated error terms is positive influence and is statistically significant in each of the monthly models and the model for the entire dataset. The likelihood ratio test (LRT) is statistically significant with very small p values (< 0.01) hinting that spatial dependence is still a problem.

During the summer months of peak activity, slope is statistically significant ($p \leq 0.05$) in the spatial error models and has a positive relationship indicating hex cells with higher slopes have higher CG lightning flash density (Table 15). Elevation is also statistically significant for all months except October with a positive relationship meaning that hex cells with higher elevation values have higher CG lightning flash density except in April where it is slightly negative. The land cover variables are also statistically significant ($p \leq 0.05$) for all months except June and September. The relationships are positive except during June where it is slightly negative. Similar to previous models, the aspect variables do not appear to be associated with lightning flash density. For the model based on the entire dataset, slope, elevation, and the land cover variable %forest appear to be statistically significant factors ($p \leq 0.05$) followed by the land cover variable %shrubgrass ($p \leq 0.10$).

Table 14. Spatial Error regression diagnostics for the data aggregated by month.

Spatial Error Regression Diagnostics								
	R^2	L-L	Akaike	Schwarz	Likelihood Ratio Test	LRT p-val	Breusch-Pagan Test	BP p-val
Apr	0.024921	10477.40	-20940.8	-20891.4	91.27	<0.001	N/A	N/A
May	0.041013	3337.15	-6660.31	-6610.88	213.45	<0.001	N/A	N/A
Jun	0.093827	1133.28	-2252.58	-2203.14	432.61	<0.001	N/A	N/A
Jul	0.103505	1148.71	-2283.44	-2234	437.02	<0.001	N/A	N/A
Aug	0.067737	1290.10	-2566.22	-2516.78	343.34	<0.001	N/A	N/A
Sep	0.022921	1858.05	-3702.11	-3652.67	131.98	<0.001	N/A	N/A
Oct	0.013796	11846.10	-23678.2	-23628.8	55.87	<0.001	120067821	<0.001
ALL	0.173819	2456.35	-4898.71	-4849.27	714.68	<0.001	N/A	N/A

Table 15. Spatial Error regression model parameters for the data aggregated by month.

Spatial Error Regression Model Parameters (coefficient, z-stat, p-value)								
	Lambda	Constant	Slope	Elevation	NSAspect	EWAspect	%Forest	%ShrubGrass
Apr	0.26730	0.04362	-1.17e-005	-1.40e-005	-0.00156	-0.00150	0.00037	0.00023
	9.04136	4.18175	-0.12204	-3.24351	-1.41031	-1.38629	5.35203	3.21590
	<0.001	<0.001	0.90287	0.00118	0.15845	0.16566	<0.001	0.00130
May	0.38597	-0.01846	0.00022	4.36e-005	-0.00100	0.00174	0.00078	0.00053
	13.6522	-0.71237	1.00850	4.06903	-0.42300	0.70481	4.39764	2.84409
	<0.001	0.47623	0.31321	<0.001	0.67229	0.48092	<0.001	0.00445
Jun	0.48112	-0.09904	0.00114	0.00017	0.00410	0.00050	5.74e-005	0.00016
	17.8993	-2.77132	3.85333	11.7265	1.26210	0.15960	0.22634	0.60294
	<0.001	0.00558	0.00012	<0.001	0.20691	0.87319	0.82093	0.54655
Jul	0.49220	-0.10598	0.00072	0.00021	0.00261	-0.00522	-0.00052	-0.00049
	18.4399	-2.94816	2.45682	14.5993	0.80576	-1.64055	-2.06234	-1.82547
	<0.001	0.00320	0.01402	<0.001	0.42038	0.10089	0.03917	0.06793
Aug	0.45689	0.04611	0.00086	0.00014	0.00144	0.00414	0.00064	0.00047
	16.7552	1.33598	2.98237	10.0676	0.45297	1.31925	2.65178	1.85900
	<0.001	0.18155	0.00286	<0.001	0.65057	0.18709	0.00801	0.06303
Sep	0.30506	0.02530	-0.00017	5.96e-005	0.00320	-0.00233	0.00030	0.00025
	10.4475	0.86990	-0.64406	4.95330	1.06379	-0.78990	1.58830	1.23627
	<0.001	0.38435	0.51953	<0.001	0.28742	0.42958	0.11222	0.21636
Oct	0.20735	-0.00919	0.00020	6.20e-006	0.00178	-0.00190	0.00014	0.00019
	6.89636	-1.07653	2.49384	1.74943	1.87908	-2.05206	2.55575	3.20686
	<0.001	0.28169	0.01264	0.08021	0.06023	0.04016	0.01060	0.00134
ALL	0.60416	0.00236	0.00110	0.00032	0.00275	0.00061	0.00061	0.00047
	24.7215	0.07110	4.30364	24.0163	0.99456	0.22742	2.41848	1.79540
	<0.001	0.94331	<0.001	<0.001	0.31995	0.82009	0.01559	0.07259

To further investigate the relationship between terrain parameters and CG lightning flash density, regression was applied separately on the cells identified as either hot or cold spots for the entire dataset (Table 16 and Table 17). The hot spot model was

able to explain 8.6 percent of the flash density variability while the cold spot model was only able to explain 1.3 percent. The models are statistically significant having very low p-values. The hot spot model shows a normal distribution in the standardized errors (not shown) and some structure in the scatter plot (not shown). The standardized errors in the cold spot model are skewed and the scatter plots shows more structure (not shown). In the hot spot models only the slope and elevation are statistically significant having very low p-values while elevation is the only statistically significant variable in the cold spot model. In both cases elevation has a positive coefficient meaning cells with higher elevation values have higher CG lightning flash density. Neither of the models showed evidence of statistically significant spatial dependence so spatial regression was not performed for these set of observations.

Table 16. OLS regression analysis model diagnostics using the cells designated as either hot or cold spots.

Classic OLS Regression Diagnostics								
	R^2 F-stat p-valuel	**LL**	**Akaike**	**Schwarz**	**Jarque-Bera** p-value	**B-P** p-value	**K-B** p-value	**M-C-N**
Hot Spots	0.086266 25.4436 <0.001	517	-1021	-983	4.2443 0.11977	-1071405 N/A	-970835 N/A	32.793394
Cold Spots	0.013565 3.95135 <0.001	520	-1027	-989	99.3534 <0.001	-11092573 N/A	-7909273 N/A	27.972654

Table 17. OLS Regression analysis model parameters using the cells designated as either hot or cold spots.

Classic OLS Regression Model Parameters								
		Constant	**Slope**	**Elevation**	**NSAspect**	**EWAspect**	**%Forest**	**%ShrubGrass**
Hot Spots	Coefficient	0.46515	0.00214	0.00019	0.00156	-0.00444	-0.00019	-0.00032
	t-statistic	8.97709	4.59193	9.180997	0.25201	-0.71824	-0.63863	-1.20236
	p-value	<0.001	<0.001	<0.001	0.80105	0.47270	0.52313	0.22940
	VIF		1.124873	1.211906	1.000795	1.003682	1.041779	1.123380
Cold Spots	Coefficient	0.50716	0.00027	8.61e-005	0.00849	-0.00298	-7.26e-005	-8.87e-005
	t-statistic	11.3853	0.485364	4.53113	1.38889	-0.48718	-0.150631	-0.1363446
	p-value	<0.001	0.62750	<0.001	0.16505	0.62622	0.88023	0.89154
	VIF		1.053150	1.047945	1.005689	1.002574	1.407169	1.386593

Overall, while the models were statistically significant, they only explain a small percentage of the variability of CG lightning flash density in Yellowstone National Park (*Research Question 4*). In this study the regression models only explained at most 17 percent of the variability in CG lightning flash density which is slightly better than a previous study by DeCaria and Babij where only 3 percent of the variability was explained when modeling the relationship between flash density and terrain (2003). Of the variables tested in this study, it appears that elevation has the greatest role in the observed spatial pattern of CG lightning activity where the hex cells with higher elevation values have higher CG lightning flash density. This is followed by slope during the summer months where hex cells with higher slope values have higher CG lightning flash density. Areas with higher slope values are probably found mostly in the mountainous regions so a positive relationship here makes sense as lightning generally favors the high points of a landscape. Land cover also appears to be a positive factor in CG lightning flash density except during June. The hex cells with higher percentage of forest or shrub and grass have higher CG lightning flash density. This is important when

considering the role that CG lightning plays in the wildfire process. Lightning provides the ignition while forest, shrub, and grass lands provide the fuel source for wildfires.

While these variables are important aspects to the study, some variables appear to be missing from the regression modeling. One variable that could be important and rather easy to include is terrain difference. For each hex cell, one could easily calculate the difference between the elevation of the hex cells and the difference in elevation between its neighboring hex cells. The elevation differences could be calculated as the difference with the average elevation of its neighbors. One could also include a difference between the hex cell and the one with the highest elevation (maximum difference) as well as with the cell and its neighbor with the lowest elevation (minimum difference). Furthermore, the hex cells neighborhood could be extended to include second or third order neighbors to form a terrain gradient variable that may also prove useful. Other variables that may be included is daily solar radiation received in each hex cell as well as weather related variables such as surface heat flux, wind direction and speed, upward vertical motion, and atmospheric stability to name a few. Since Yellowstone does not have a robust weather observing network, these variables would have to be modeled using numerical weather prediction techniques.

7. FUTURE RESEARCH

This research analyzed the spatial pattern of cloud-to-ground lightning activity to better understand *where* lightning activity occurred in Yellowstone National Park between 1995 and 2004. Mapping where activity occurred (and did not occur) and identifying what areas of the park experienced significant clusters of activity is a step toward new information regarding the geography of CG lightning in the park. While this research provided important results, there is much that remains unknown regarding the physical processes influencing CG lightning activity and its expression on the landscape.

In the introductory remarks of this paper, the original research questions hypothesized about the relationship between the physical geography of Yellowstone and the physical characteristics of CG lightning in Yellowstone. This research evaluated the physical aspects of Yellowstone with respect to CG lightning. Spatial analysis was performed to test for both global and local clustering based strictly on the location of CG lightning events and regression analysis was performed testing the relationship between CG lightning flash density and the physical properties of Yellowstone's terrain. This research did not consider the physical properties of the lightning itself such as multiplicity (number of flashes that occurred during the cloud-to-ground event), polarity, and strength all of which could be important factors in the ignition of wildfires. This analysis should be extended to include these lightning parameters to determine locations of clusters of CG lightning activity with high (or low) multiplicity, clusters with positive (or negative) polarity, and clusters based on the strength of the current delivered by the

lightning. Adding the locations of lightning ignited wildfires to maps showing the location of these types of clusters could greatly improve our understanding in lightning-ignited wildfires. This could help in determining whether or not lightning-ignited wildfires are more likely from lightning flashes with single strikes or multiple strikes and whether or not lightning strikes with high electrical currents are needed for wildfire ignition.

The inclusion of the physical properties of lightning could also be used to possibly improve upon the regression results by using the lightning point data. The multiplicity property contains the number of flashes within each lightning event. This is basically a count variable that can be used as the dependent variable. Likewise, the strength (electrical current) of the lightning flash could also be used as a dependent variable in the regression analysis.

Improvements in the regression analysis using lightning flash density as the dependent variable might also be realized by modifying the method used to downscale the terrain and land cover variables to the hex cells. This study used a majority filter to assign terrain and land cover values to the hex cells. Using another filter such as maximum, mean, or median may improve the regression results. Another method that might prove useful is to assign the terrain variables using categories that are based on bands of elevation values. One could also analyze the terrain for high points and use the distance between the lightning event and the nearest high point as a regression variable.

Further evaluation of the cold spots or clusters of low CG lightning activity may prove interesting as well. Most of the cold spots of activity were located in the western and central plateau regions of the park, which is where most of the basins of geothermal activity are located. These geothermal hot spots release gases in the atmosphere. The concentration of releases in this part of the park may be impacting the weather process in a way that leads to the clusters of low CG lightning activity in this region.

Another potentially fruitful extension of this research is analyzing CG lightning events using spatio-temporal point pattern analysis. This study sliced and aggregated 10-years of CG lightning data using monthly and hourly time units to determine whether spatial clustering was present. Space-time clustering approaches evaluate the entire dataset as a whole to simultaneously find events that cluster in both space and time. This type of analysis may reveal trends in the processes leading to CG lightning occurrence that can't be seen when evaluating an aggregated 10-year dataset such as in this research. Trends revealed in space-time clustering could be used to help compare and predict fire seasons from year to year.

In addition to further investigating the temporal aggregation of CG lightning activity, the spatial aggregation of the study area should also be investigated further. While this study did an initial, quick test to determine an optimal cell size for tessellating the study area to a hexagonal grid, a more robust test to determine the sensitivity of grid cell size should be completed. Different cell sizes may result in different means and variances of CG lightning flash density which may in turn impact the resulting local and global statistics used in this study to test for spatial autocorrelation and clustering.

Along with performing sensitivity testing on cell size and clustering, the size of the study area could also be expanded to include the entire Greater Yellowstone Area. This would include the entire Absaroka Range and Gallatin Range instead of just parts of the ranges that lie within Yellowstone as well as the Teton Range to the south of Yellowstone. These ranges have several peaks that are higher than the peaks that lie within the Yellowstone. Including more of the mountain ranges should reveal more localized clusters of CG lightning activity may provide more clues about the CG lightning in these regions and a better understanding of the local weather patterns that lead to the spatial distribution of lightning activity in the region.

While this research provided some initial results, there remains much to be discovered regarding the physical process leading to the observed spatial pattern of CG lightning activity. Integrating the regional meteorological conditions, as well as the local weather patterns resulting from the physical geography of Yellowstone, may provide an improved understanding of why CG lightning clusters in some regions while absent in other regions of the park. Knowledge of where and why clusters occur in these regions can be used to refine wildfire models used to predict and monitor the environment for lightning ignited wildfires.

8. SUMMARY

Yellowstone National Park is well known for its vivid and diverse landscape, its abundance of wildlife, and its wildfires such as the 1988 “Summer of Fire”. Yellowstone is also well known for its volcanic activity which is the source of the geothermal hot spots that cover the landscape in the form of geysers like Old Faithful, hot springs, mud pots, and fumaroles. This research demonstrated that Yellowstone National Park has other hot spots that are also important to the ecology of the park: clusters of cloud-to-ground lightning flashes.

Using GIS and spatial analysis techniques, this research evaluated a 10-year CG lightning dataset from 1995 to 2004 to examine the spatial patterns of CG lightning over different time scales and their relationships with physical characteristics of the park such as terrain elevation, slope, aspect, and land cover. CG lightning point data was visualized as a series of maps with each point representing a single CG lightning flash event. CG lightning flash density was also calculated and visualized after tessellating the Yellowstone study area into a hexagon grid. The CG lightning points were aggregated to the hex grid temporally by month and by hour such that each grid cell contained a calculated flash density (*Ng*) value for each of the 12 months as well as a value for each of the 24 hours. The overall CG lightning flash density based on the entire 10-year dataset along with its mean and standard deviation were also calculated for each grid cell. Spatial analysis techniques were applied to this data focusing on 4 research questions.

Research question 1 asked “Does cloud-to-ground lightning occur randomly across space or does it manifest in an observable spatial pattern?” The hypothesis stated that the physical properties of CG lightning along with Yellowstone’s topography interact to create a spatially clustered pattern of CG lightning activity. Visual inspection of the spatial patterns in the point and flash density maps suggested that lightning activity exhibited spatial clustering. To formally test for global clustering, Nearest Neighbor Index (NNI), K-order Nearest Neighbor Index, and Ripley’s K (*L(d)*) were calculated for the study area using the lightning flash event data and the Getis-Ord General G statistic and the Global Moran’s I were calculated using the lightning flash density. Based strictly on the location of the CG lightning, these spatial statistics suggest there is weak evidence of global clustering and the spatial pattern of CG lightning activity is closer to a random pattern. Further tests integrating the physical properties of lightning and Yellowstone’s topography needs to be completed to fully investigate clustering resulting from the interaction of these properties.

Research question 2 asked “Is the spatial pattern of cloud-to-ground lightning consistent throughout the year or does it vary by month or season?” The hypothesis asserted the overall spatial pattern of CG lightning varies in time due to regional and local scale weather patterns. The series of bar charts and line graphs of the lightning events provided evidence of the temporal variability of the lightning activity with the data showing definitive seasonal and diurnal trends. While these charts and graphs along with the maps of the lightning point data and lightning flash density all reveal that CG

lightning does vary with time, more research is needed to determine the impact of regional and local weather patterns on the spatial pattern of CG lightning.

The third research question asked "Are specific areas or regions of Yellowstone Park more prone to experience cloud-to-ground lightning than others?" It was anticipated that the spatial pattern of CG lightning would exhibit localized clusters of high and low lightning values of activity within the park. To test for local clustering in the spatial pattern, the Getis-Ord G_i^* and Anselin's Local Moran's I were used to identify clusters of high and low activity. These tests revealed several clusters of low CG lightning flash density located mainly in the western and central plateau regions of the park and several clusters of high CG lightning flash density mainly in the mountainous regions of the park.

The spatial scale of clustering appears to be consistent with the spatial scale of the physical process associated with the life cycle of convection and thunderstorm activity in mountain regions. The initiation of the convection process can take place on a scale of less than 1 km. As the process evolves to form thunderstorms, the scale of the process also grows resulting in thunderstorm cells that can be 1 to 10 km in diameter. The downdrafts of cold air in mature thunderstorms will often lead to the initiation of new thunderstorms adjacent to the mature storm thus forming a multicellular thunderstorm cluster. The scale of these thunderstorm clusters can range from 10 to 25 km (Vogt and Hodanish, 2016; Mason and Mason, 2003; Ziegler, Ray and MacGorman, 1986; Raymond and Wilkening, 1980). The results of the tests for global clustering showed weak evidence of clustering at distances reflecting both the single cell and multicellular thunderstorm clusters. The NNI and K-order NNI appear to reflect the clustering of

lightning activity resulting from single cell thunderstorm activity with weak clustering indicated between 1 and 5 km. Results from the Global Moran's I and Getis-Ord General G statistic appear to reflect clustering of lightning activity associated with both single cell and multicellular thunderstorm activity as weak clustering was identified for distances up to about 20 km with peak clustering in the 5 to 10 km range. Likewise, the tests for local clustering appear to be a reflection of the lightning associated with single and multicellular thunderstorm activity. These tests, Getis-Ord G_i^* and Anselin's Local Moran's I, were conducted using a 3 km neighborhood and resulted in clusters of CG lightning activity that were generally 5 to 25 km wide.

Finally, the last research question asked "Is cloud-to-ground lightning activity associated with Yellowstone Park's topography?" The lightning flash density pattern was expected to be related to land cover and the terrain variables elevation, slope, and aspect. The relationship between CG lightning flash density, land cover, and the terrain variables was analyzed using OLS and spatial error regression to account for spatial autocorrelation. While the resulting regression models were statistically significant, overall the regression only explained at most 17 percent of the variability in lightning flash density. Of the variables tested, elevation appears to play the greatest role in the observed spatial pattern of CG lightning activity, especially during the months of peak activity. Slope also appears to be a factor as well as the land cover variables forest and shrub-grass. For the most part, the variables have a positive relationship with lightning flash density such that hex cells with a higher percentage of forest or shrub-grass had

higher CG lightning flash density and hex cells with higher slope also had higher values of CG lightning flash density.

While the regression models generated in this study may not have revealed as strong of a relationship with the terrain properties as expected, mapping the results of the tests for local clustering do show clusters of high CG lightning activity that are predominantly located in the mountain regions and clusters of low CG lightning predominantly in the western and central plateau of Yellowstone. This becomes more visually apparent when the results are visualized on a 3D map like the ones shown in Figure 32 and Figure 33.

CG lightning studies found in scholarly publications typically only focus on visualizing the distribution of CG lightning activity as flash density maps for the simple purpose of identifying areas of high flash density rates. These types of studies often do little in explaining the observed spatial pattern of flash density. As possibly the first attempt at applying spatial statistics when analyzing CG lightning, this study appears to show the value in leveraging these techniques to gain new information about CG lightning. Spatial statistics revealed the statistically significant clustered nature of CG lightning in Yellowstone National Park during the 10-year period of study, identifying clusters of high and low CG lightning activity by evaluating these clusters within the context of its neighborhood. Maps of flash density do not tell us anything about areas of high or low flash density values within the context of its neighborhood. The spatial statistics implemented in this study also revealed that CG lightning in Yellowstone takes

place over spatial scales ranging from 1 to 20 km indicating complex processes are involved, information that can't be readily obtained from flash density maps.

Analyzing CG lightning activity in this manner allowed us to discover that clustering of CG lightning occurs over a range of spatial scales and helped to locate areas of the park that are more prone to localized clusters of high and low CG lightning activity. This naturally leads to the question of "why". Why do localized clusters occur where they do? Is there something unique about the physical properties of the CG lightning that allows clusters to occur where they do? Is there something unique about the underlying terrain that plays a role in the process?

In conclusion, this study clearly benefited from leveraging the power of GIS and spatial statistics to analyze the CG lightning within Yellowstone National Park. The spatial statistics used in this study allowed for the discovery of Yellowstone's other "hot spots". These hot spots are important not only to the ecology of the park but also to those who manage and visit the park.

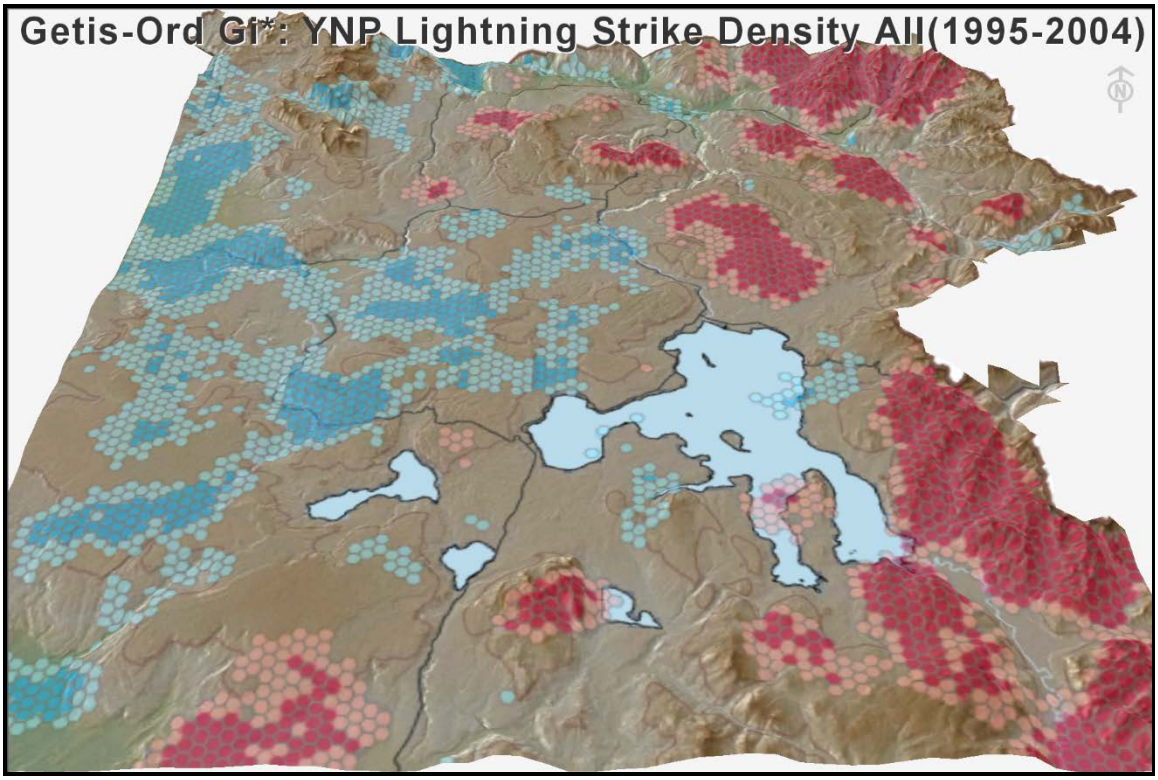

Figure 32. A 3D visualization of hot and cold spots of CG lightning flash density in Yellowstone National Park.

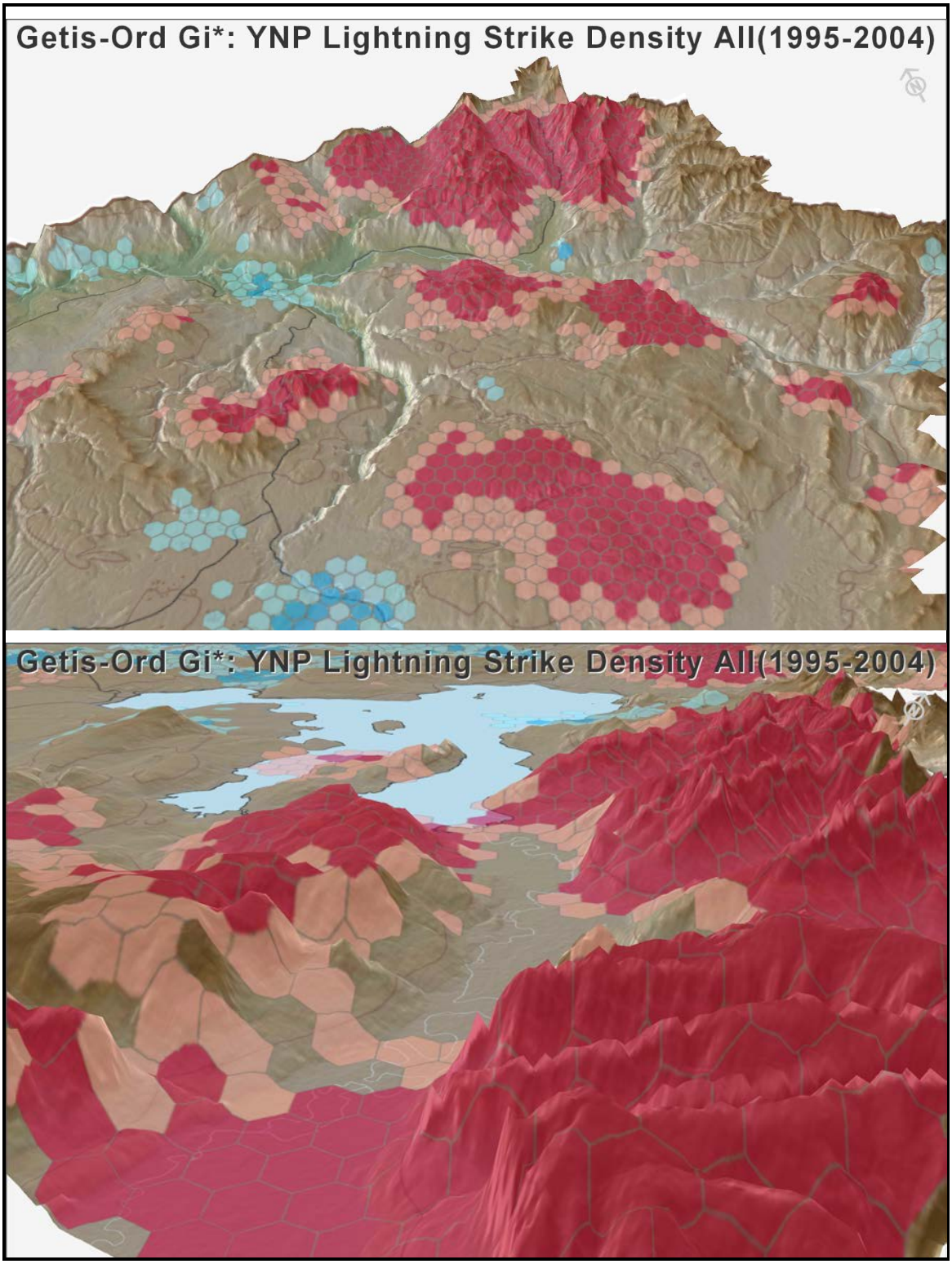

Figure 33. A 3D visualization of hot and cold spots of CG lightning flash density zoomed in to select regions.

REFERENCES

Adams, D.K., & Comrie, A.C. (1997). The North American Monsoon. *Bulletin of the American Meteorological Society, 78*(10), 2197–2213.

Agee, J. K. (1993). *Fire Ecology of Pacific Northwest Forests*. Washington, DC; Covelo CA: Island Press.

Amatulli, G., Peréz-Cabello, F., & de la Riva, J. (2007). Mapping Lightning/Human-Caused Wildfires Occurrence Under Ignition Point Location Uncertainty. *Ecological Modelling, 200*(3–4), 321–333. http://doi.org/10.1016/j.ecolmodel.2006.08.001

Anselin, A. (2016). GeoDa: Release 1.8 [computer software]. Center for Spatial Data Science, University of Chicago.

Anselin, L., Syabri, I., & Kho, Y. (2010). GeoDa: An Introduction to Spatial Data Analysis. In M. M. Fischer & A. Getis (Eds) *Handbook of Applied Spatial Analysis* (pp. 73-89). Berlin, Heidelberg: Springer.

Anselin, L. (2005). Exploring Spatial Data With GeoDa: A Workbook. Retrieved from http://la1.rcc.uchicago.edu/media/geoda_files/docs/geodaworkbook.pdf (Accessed 9 July 2016).

Anselin, L. (1995). Local Indicators of Spatial Association (LISA). *Geographical Analysis, 27*, 92–115. http://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Harlow Essex, England; New York, NY: Longman Scientific & Technical; J. Wiley.

Bartlein P., Hostetler, S., Shafer, S., Holman, J., & Solomon, A. (2008). Temporal and Spatial Structure in Daily Wildfire-Start Data Set from the Western United States (1986–96). *International Journal of Wildland Fire, 17*(1), 8–17. http://dx.doi.org/10.1071/wf07022

Carr, D. B., Olsen, A. R., & White, D. (1992). Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data. *Cartography and Geographic Information Systems, 19*(4), 228–236. http://doi.org/10.1559/152304092783721231

Carr, D. B., Littlefield, R. J., Nicholson, W. L., & Littlefield, J. S. (1987). Scatterplot Matrix Techniques for Large N. *Journal of the American Statistical Association, 82*(398), 424–436. http://doi.org/10.1080/01621459.1987.10478445

Chainey, S. & Ratcliffe, J. (2005). *GIS and Crime Mapping*. Chichester, West Sussex, England: J. Wiley.

Chou, Y. (1992). Spatial Autocorrelation Analysis and Weighting Functions in the Distribution of Wildland Fires. *International Journal of Wildland Fire*, *2*(4), 169–176. http://dx.doi.org/10.1071/wf9920169

Christensen, P., & Abbott, I. (1989). Impact of Fire in the Eucalypt Forest Ecosystem of Southern Western Australia: A Critical Review. *Australian Forestry*, *52*(2), 103–121. http://doi.org/10.1080/00049158.1989.10674542

Chuvieco, E., & Salas, J. (1996). Mapping the Spatial Distribution of Forest Fire Danger Using GIS. *International Journal of Geographical Information Systems*, *10*(3), 333–345. http://doi.org/10.1080/02693799608902082

Clark, P. J., & Evans, F. C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, *35*(4), 445–453. http://doi.org/10.2307/1931034

Cummins, K., Murphy, M., Cramer, J., Scheftic, W., Demtriades, N. & Nag, A. (2010) Location Accuracy Improvements Using Propagation Corrections: A Case Study of the U.S. National Lightning Detection Network. *21st International Lightning Detection Conference & Third International Lightning Meteorology Conference*, Orlando, FL 19-22 April 2010.

Cummins, K. & Murphy, M. (2009). An Overview of Lightning Locations Systems: History, Techniques and Data Uses, With an In-depth Look at the U.S. NLDN. *IEEE Transactions On Electromagnetic Compatibility, 51*(3), 499–517. http://dx.doi.org/10.1109/temc.2009.2023450

Cummins, K., Cramer, J., Biagi, C., Krider, E., Jerauld, J., Uman, M. & Rakov, V. (2006). The US National Lightning Detection Network: Post-upgrade status. In *Preprints, Second Conference on Meteorological Applications of Lightning Data*, Atlanta, GA.

DeCaria, A. & Babij, M. (2003). A Map of Lightning Strike Density for Southeastern Penssylvania and Correlation with Terrain Elevation. *USGS Digital Mapping Techniques '03 Workshop*, Millersville, PA, 1-4 June 2003

Despain, D. (1990). *Yellowstone Vegetation: Consequences of Environment and History in a Natural Setting.* Santa Barbara, CA: Roberts Rinehart Publishers.

Dissing, D., & Verbyla, D. L. (2003). Spatial Patterns of Lightning Strikes in Interior Alaska and Their Relations to Elevation and Vegetation. *Canadian Journal of Forest Research*, *33*(5), 770–782. http://doi.org/10.1139/x02-214

Environmental Systems Research Institute (ESRI). (2016). ArcGIS Pro Online Help Documentation, http://pro.arcgis.com/en/pro-app/help/main/welcome-to-the-arcgis-pro-app-help.htm (Accessed 9 July 2016).

Environmental Systems Research Institute. (2016). ArcGIS Pro: Release 1.2 [computer software]. Redlands, CA

Fortin, M., Dale, M. R. T., & ver Hoef, J., (2002). Spatial Analysis in Ecology. *Encyclopedia of Environmetrics*, *4*, 2051–2058.

Franke, M., & Yellowstone Center for Resources. (2000). *Yellowstone in the Afterglow : Lessons From the Fires*. Mammoth Hot Springs, Wyo.: Yellowstone Center for Resources, Yellowstone National Park.

Fuquay, D. M. (1982). Positive Cloud-to-Ground Lightning in Summer Thunderstorms. *Journal of Geophysical Research: Oceans*, *87*(C9), 7131–7140. http://doi.org/10.1029/JC087iC09p07131

Fuquay, D. M. (1980). Lightning That Ignites Forest Fires. *Sixth Conference of Fire and Forest Meteorology*, Seattle, WA, 22-24 April 1980.

Genton, M. G., Butry, D. T., Gumpertz, M. L., & Prestemon, J. P. (2006). Spatio-temporal Analysis of Wildfire Ignitions in the St Johns River Water Management District, Florida. *International Journal of Wildland Fire*, *15*(1), 87–97. http://dx.doi.org/10.1071/WF04034

Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, *24*(3), 189–206. http://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Grogan, M. (2004). Report on the 2002-2003 U.S. NLDN System Wide Upgrade, Vaisala, Inc. http://www.vaisala.com/Vaisala%20Documents/Vaisala%20News%20Articles/VN165/VN165_Report_on_the_2002-2003_U.S._NLDN_System-wide_Upgrade.pdf (Accessed 9 July 2016).

Hirschfield, A., & Bowers, K. (2001). *Mapping and Analysing Crime Data: Lessons from Research and Practice*. London; New York: Taylor & Francis.

Hodanish, S. & Wolyn, P. (2006). Lightning Climatology for the State of Colorado. *Fourth International Lightning Meteorology Conference*, Broomfield, CO 22-27 January 2011.

Huang, S., Rich, P. M., Crabtree, R. L., Potter, C. S., & Fu, P. (2008). Modeling Monthly Near-Surface Air Temperature from Solar Radiation and Lapse Rate: Application over Complex Terrain in Yellowstone National Park. *Physical Geography*, *29*(2), 158–178. http://doi.org/10.2747/0272-3646.29.2.158

Huffines, G. R., & Orville, R. E. (1999). Lightning Gound Flash Density and Thunderstorm Duration in the Continental United States: 1989-96. *Journal of Applied Meteorology*, *38*(7), 1013–1019.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D. & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering & Remote Sensing*, *81*(5), 345–354.

Jaiswal, R. K., Mukherjee, S., Raju, K. D., & Saxena, R. (2002). Forest Fire Risk Zone Mapping from Satellite Imagery and GIS. *International Journal of Applied Earth Observation and Geoinformation*, *4*(1), 1–10. http://doi.org/10.1016/S0303-2434(02)00006-5

Johnston, C. A. (1998). *Geographic Information Systems in Ecology*. Oxford; Malden, MA: Blackwell Science.

Kasischke, E. S., Williams, D., & Barry, D. (2002). Analysis of the Patterns of Large Fires in the Boreal Forest Region of Alaska. *International Journal of Wildland Fire*, *11*(2), 131–144. http://doi.org/10.1071/WF02023

Kharuk, V. I., Kasischke, E. S., & Yakubailik, O. E. (2007). The Spatial and Temporal Distribution of Fires on Sakhalin Island, Russia. *International Journal of Wildland Fire*, *16*(5), 556–562. http://doi.org/10.1071/WF05009

Kilinc, M., & Beringer, J. (2007). The Spatial and Temporal Distribution of Lightning Strikes and Their Relationship with Vegetation Type, Elevation, and Fire Scars in the Northern Territory. *Journal of Climate*, *20*(7), 1161–1173. http://doi.org/10.1175/JCLI4039.1

Krider, E. P., Noggle, R. C., Pifer, A. E., & Vance, D. L. (1980). Lightning Direction-Finding Systems for Forest Fire Detection. *Bulletin of the American Meteorological Society*, *61*(9), 980–986. http://doi.org/10.1175/1520-0477(1980)061<0980:LDFSFF>2.0.CO;2

Latham, D., & Williams, E. (2001). Lightning and forest fires. In E. A. Johnson & K. Mayanishi (Eds) *Forest Fires: Behavior and Ecological Effects* (pp 375-418). San Diego, CA: Academic Press.

Lee, B., Park, P. S., & Chung, J. (2006). Temporal and Spatial Characteristics of Forest Fires in South Korea Between 1970 and 2003. *International Journal of Wildland Fire*, *15*(3), 389–396. http://doi.org/10.1071/WF05090

Legendre, P. (1993). Spatial Autocorrelation: Trouble or New Paradigm? *Ecolog*, *74*(6), 1659–1673.

Levine, N. (2015). CrimeStat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 4.0.2) [computer software]. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC.

Lloyd, C. D. (2011). *Local Models for Spatial Analysis*. Boca Raton, FL: CRC Press.

Lyons, W. A., Uliasz, M., & Nelson, T. E. (1998). Large Peak Current Cloud-to-Ground Lightning Flashes during the Summer Months in the Contiguous United States. *Monthly Weather Review*, *126*(8), 2217–2233. http://doi.org/10.1175/1520-0493(1998)126<2217:LPCCTG>2.0.CO;2

Mallick, S., Rakov, V. A., Ngin, T., Gamerota, W. R. , Pilkey, J. T. , Hill, J. T., Uman, M. A., Jordan, D. M. , Cramer, J. A. & Nag, A. (2014). An Update on the Performance Characteristics of the NLDN. *23rd International Lightning Detection Conference & Fifth International Lightning Meteorology Conference*, Tuscon, AZ 18-21 March 2014.

Marcus, W. A., Meacham, J. E., & Rodman, A. W. (2012). Atlas of Yellowstone. Berkeley; [Eugene, Ore.]: University of California Press; University of Oregon.

Mason, J. & Mason, N. (2003). The Physics of a Thunderstorm. *European Journal of Physics*, *24*, S99–S110.

Meisner, B. N. (1993). A Lightning Fire Ignition Assessment Model. *Twelfth Conference on Fire and Forest Meteorology*, Jekyll Island, GA, 26-28 October 1993.

Mitchell A (2005) *The ESRI Guide to GIS Analysis Volume 2: Spatial Measurements and Statistics.* Redlands, CA: ESRI Press.

Mock, C.J. (1996). Climatic Controls and Spatial Variations of Precipitation in the Western United States. *Journal of Climate*, *9*, 1111–1125.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, *37*(1/2), 17–23. http://doi.org/10.2307/2332142

Nag, A., Murphy, M. J., Cummins, K. L., Pifer, A. E. & Cramer, J.A. (2014). Recent Evolution of the U.S. National Lightning Detection Network. *23rd International Lightning Detection Conference & Fifth International Lightning Meteorology Conference, Tuscon, AZ 18-21 March 2014.*

Neuwirth, C., Spitzer, W. & Prinz, T. (2012). Lightning Density Distribution and Hazard in an Alpine Region. *Journal of Lightning Research*, 4, 166–172. http://dx.doi.org/10.2174/1652803401204010166

Ord, J., & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, *27*(4), 286–306. http://doi.org/10.1111/j.1538-4632.1995.tb00912.x

Orville, R. E., Huffines, G. R., Burrows, W. R., & Cummins, K. L. (2011). The North American Lightning Detection Network (NALDN)—Analysis of Flash Data: 2001–09. *Monthly Weather Review*, *139*(5), 1305–1322. http://doi.org/10.1175/2010MWR3452.1

Orville, R. E. (2008). Development of the National Lightning Detection Network. *Bulletin of the American Meteorological Society*, *89*(2), 180–190. http://doi.org/10.1175/BAMS-89-2-180

Orville, R. E., & Huffines, G. R. (2001). Cloud-to-Ground Lightning in the United States: NLDN Results in the First Decade, 1989–98. *Monthly Weather Review*, *129*(5), 1179–1193. http://doi.org/10.1175/1520-0493(2001)129<1179:CTGLIT>2.0.CO;2

Orville, R. E., & Huffines, G. R. (1999). Lightning Ground Flash Measurements over the Contiguous United States: 1995–97. *Monthly Weather Review*, *127*(11), 2693–2703. http://doi.org/10.1175/1520-0493(1999)127<2693:LGFMOT>2.0.CO;2

Orville, R. E., & Silver, A. C. (1997). Lightning Ground Flash Density in the Contiguous United States: 1992–95. *Monthly Weather Review*, *125*(4), 631–638. http://doi.org/10.1175/1520-0493(1997)125<0631:LGFDIT>2.0.CO;2

Orville, R. E. Lightning Ground Flash Density in the Contiguous United States–1989. (1991). *Monthly Weather Review,* 119(2), 573–577.

Oyana, T. J., & Margai, F. M. (2015). *Spatial Analysis: Statistics, Visualization, and Computational Methods*. CRC Press.

Perry, G. L. W., Miller, B. P., & Enright, N. J. (2006). A Comparison of Methods for the Statistical Analysis of Spatial Point Patterns in Plant Ecology. *Plant Ecology, 187*, 59–82.

Rakov, V. A. (2003). A Review of Positive and Bipolar Lightning Discharges. *Bulletin of the American Meteorological Society*, *84*(6), 767–776. http://doi.org/10.1175/BAMS-84-6-767

Rakov, V., & Uman, M. A. (2003). *Lightning: Physics and Effects*. Cambridge, U.K.; New York: Cambridge University Press.

Raymond, D., & Wilkening, M. (1980). Mountain-Induced Convection under Fair Weather Conditions. *Journal of the Atmospheric Sciences*, *37*(12), 2693–2706. http://doi.org/10.1175/1520-0469(1980)037<2693:MICUFW>2.0.CO;2

Reap, R. M. (1986). Evaluation of Cloud-to-Ground Lightning Data from the Western United States for the 1983–84 Summer Seasons. *Journal of Climate and Applied Meteorology*, *25*(6), 785–799. http://doi.org/10.1175/1520-0450(1986)025<0785:EOCTGL>2.0.CO;2

Ripley, B. D. (1981). *Spatial Statistics*. New York: Wiley.

Rorig, M. L., & Ferguson, S. A. (2002). The 2000 Fire Season: Lightning-Caused Fires. *Journal of Applied Meteorology*, *41*(7), 786–791. http://doi.org/10.1175/1520-0450(2002)041<0786:TFSLCF>2.0.CO;2

Rorig, M. L., & Ferguson, S. A. (1999). Characteristics of Lightning and Wildland Fire Ignition in the Pacific Northwest. *Journal of Applied Meteorology*, *38*(11), 1565–1575. http://doi.org/10.1175/1520-0450(1999)038<1565:COLAWF>2.0.CO;2

Rudlosky, S. D., & Fuelberg, H. E. (2011). Seasonal, Regional, and Storm-Scale Variability of Cloud-to-Ground Lightning Characteristics in Florida. *Monthly Weather Review*, *139*(6), 1826–1843. http://doi.org/10.1175/2010MWR3585.1

Saba, M. M.,F., Schulz, W., Warner, T. A., Campos, L. Z.,S., Schumann, C., Krider, E. P., Cummins, K. & Orville, R. E. (2010). High-speed Video Observations of Positive Lightning Flashes to Ground. *Journal of Geophysical Research Atmospheres*, *115*(24) doi:http://dx.doi.org.mutex.gmu.edu/10.1029/2010JD014330

Saraiva, A. C. V., Saba, M. M. F., Campos, L. Z. S., Pinto Jr., O., Cummins, K. L. & Krider, E. R. (2008). Properties of Negative Cloud-to-Ground Lightning from High-Speed Video Observations in Arizona, U.S.A., and Sao Paulo, Brazil. *20th International Lightning Detection Conference*¸ Tucson, AZ 21-23 April 2008.

Scott, L.M. (2015). Analysis of Spatial Patterns. In J.D. Wright (Ed) *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 178-184). Oxford, Elsevier.

Stewart, J.Q., Whiteman, C.D., Steenburgh, W.J., & Bian, X. (2002). A Climatological Study of Thermally Driven Wind Systems of the U.S. Intermountain West. *Bulletin of the American Meteorological Society*,*83*(5), 699–708.

Telesca, L., Bernardi, M., & Rovelli, C. (2005). Intra-cluster and Inter-cluster Time Correlations in Lightning Sequences. *Physica A: Statistical Mechanics and Its Applications*, *356*(2–4), 655–661. http://doi.org/10.1016/j.physa.2005.02.090

Thompson, H. R. (1956). Distribution of Distance to Nth Neighbour in a Population of Randomly Distributed Individuals. *Ecology*, *37*(2), 391–394. http://doi.org/10.2307/1933159

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *47*, 234–240.

Uman, M. A. (1987). *The lightning discharge*. Orlando, FL: Academic Press.

U.S. National Park Service. (2016). Facts About Yellowstone. Retrieved from https://www.nps.gov/yell/planyourvisit/parkfacts.htm (Accessed 9 July 2016).

U.S. National Park Service. (2016). Yellowstone Resources and Issues Handbook. Retrieved from https://www.nps.gov/yell/planyourvisit/upload/RI_2016_FINAL_web.pdf (Accessed 9 July 2016).

U.S. National Park Service, Yellowstone National Park. (2014). 2014 Fire Management Plan. Retrieved from https://www.nps.gov/yell/learn/management/upload/YELL-2014-FMP-Final_sm.pdf (Accessed 9 Jul 2016).

U.S. National Park Service. (2012). Yellowstone Resources and Issues Handbook. Retrieved from http://s3.documentcloud.org/documents/357352/yellowstone-resources-issues-2012.pdf (Accessed 9 July 2016).

U.S. National Park Service. (2007). Fire as a Natural Force. Retrieved from http://www.nps.gov/yell/planyourvisit/upload/Yell288.pdf (Accessed 9 July 2016).

U.S. National Park Service, Yellowstone National Park. (2004). 2004 Update of the Yellowstone National Park Wildland Fire Management Plan. Retrieved from http://gisedu.colostate.edu/webcontent/nr505/nps07/team7/Maps/Fire_Plans_pdf/Yellowstone.pdf (Accessed 9 July 2016)

Vadrevu, K. P., Badarinath, K. V. S., & Anuradha, E. (2007). Spatial Patterns in Vegetation Fires in the Indian Region. *Environmental Monitoring and Assessment*, *147*(1), 1–13. http://doi.org/10.1007/s10661-007-0092-6

van Wagtendonk, J.W., van Wagtendonk, K.A., Meyer, J.B. & Paintner, K.J. (2002). The Use of Geographic Information For Fire Management Planning In Yosemite National Park. *Applied Geography*, *19*(1), 19–39.

Vazquez, A., & Moreno, J. M. (1998). Patterns of Lightning-, and People-caused Fires in Peninsular Spain. *International Journal of Wildland Fire*, *8*(2), 103–115. http://doi.org/10.1071/WF9980103

Vogt, B. J., & Hodanish, S. J. (2016). A Geographical Analysis of Warm Season Lightning/Landscape Interactions Across Colorado, USA. *Applied geography, 75*, 93–103. http://dx.doi.org/10.1016/j.apgeog.2016.08.006

Vogt, B. J., & Hodanish, S. J.. (2014). A High-Resolution Lightning Map of the State of Colorado. *Monthly Weather Review*, *142*(7), 2353–2360. http://doi.org/10.1175/MWR-D-13-00334.1

Vogt, B. J. (2014). Visualizing Summertime Lightning Patterns on Colorado Fourteeners. *Professional Geographer*, *66*(1), 41–57.

Vogt, B. J. (2011). Exploring Cloud-to-Ground Lightning Earth Highpoint Attachment Geography by Peak Current. *Earth Interactions*, *15*(8), 1–16.

Wagner, G., Fuelberg, H.E., Kann, D., Wynne, R. & Cobb, S. (2006). A GIS-Based Approach to Lightning Studies for West Texas and New Mexico. *Second Conference on Meteorological Applications of Lightning Data.* Atlanta, GA 27 January - 3 February 2006.

Weber, K., McMahan, B., Johnson, P., Russell, G. (2003). Modeling Lightning as an Ignition Source Of Rangeleand Wildfire In Southeastern Idaho. Retrieved from http://giscenter.isu.edu/Research/Techpg/nasa_wildfire/pdf/Chapter6.pdf (Accessed 9 July 2016)

Westerling, A. L., Gershunov, A., Brown, T. J., Cayan, D. R., & Dettinger, M. D. (2003). Climate and Wildfire in the Western United States. *Bulletin of the American Meteorological Society*, *84*(5), 595.

Wierzchowski, J., Heathcott, M., & Flannigan, M. D. (2002). Lightning and Lightning Fire, Central Cordillera, Canada. *International Journal of Wildland Fire*, *11*(1), 41–51. http://doi.org/10.1071/WF01048

Yin, Z. Y., Estberg, J., Hallissey, E. J., & Cayan, D. R. (2007). Spatial patterns of lightning at different spatial scales in the Western United States during August of 1990 – A case study using geographic information systems technology. *Journal of Environmental Informatics*, *9*(1), 4–17.

Zajac, B. A., & Rutledge, S. A. (2001). Cloud-to-Ground Lightning Activity in the Contiguous United States from 1995 to 1999. *Monthly Weather Review*, *129*(5), 999–1019. http://doi.org/10.1175/1520-0493(2001)129<0999:CTGLAI>2.0.CO;2

Ziegler, C. L., Ray, P. S., & MacGorman, D. R. (1986). Relations of Kinimatics, Microphysics and Electrification in an Isolated Mountain Thunderstorm. *Journal of Atmospheric Sciences*, *43*(19), 2098–2114.

BIOGRAPHY

Ed Amrhein graduated from Bishop McNamara High School, Forestville, Maryland, in 1985. Upon graduation, he enlisted in the United States Air Force as a Weather Technician. After returning from OPERATION DESERT STORM, Mr. Amrhein began his college studies at Central Texas College while stationed at Fort Hood, Texas. He was soon accepted into the highly competitive Air Force Airman Education and Commissioning Program and received a scholarship to attend Texas A&M University. He received his Bachelor of Science in Meteorology from Texas A&M University in 1999 and was commissioned as a 2nd Lieutenant after attending Air Force Officer Training School at Maxwell Air Force Base, Alabama. He received his Master of Science in Atmospheric Sciences from Texas A&M University in 2001. Mr. Amrhein retired from the Air Force in 2010 and began his second career as a Geospatial Information Scientist working for Kinsey Technical Services Incorporated (KTSI).