PATTERNS OF GEOGRAPHICAL CONTENT CONTRIBUTIONS MADE BY THE GENERAL PUBLIC IN RESPONSE TO NATURAL DISASTER EVENTS

by

Stephanie M Parker
A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Masters of Science
Geoinformatics and Geospatial Intelligence

Committee:

___________________________________	Dr. Anthony Stefanidis, Thesis Director
___________________________________	Dr. Arie Croitoru, Committee Member
___________________________________	Dr. Andrew Crooks, Committee Member
___________________________________	Dr. Dieter Pfoser, Committee Member
___________________________________	Dr. Anthony Stefanidis, Department Chair
___________________________________	Dr. Donna M. Fox, Associate Dean for Academic and Student Affairs, College of Science
___________________________________	Dr. Peggy Agouris, Dean, College of Science
Date: ______________________________	Spring Semester 2015 George Mason University Fairfax, VA

Patterns of Geographical Content Contributions Made by the General Public in Response to Natural Disaster Events

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

by

Stephanie M Parker
Bachelor of Science
Virginia Polytechnic and State University, 2009

Director: Anthony Stefanidis, Professor
Department of Geoinformatics and Geospatial Intelligence

Spring Semester 2015
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my Mom and Dad and my dog Viva.

ACKNOWLEDGEMENTS

I would like to thank Drs. Stefanidis, Croitoru, Crooks, and Pfoser for their invaluable knowledge, guidance and most importantly their enduring patience with me during this entire process. I would also like to thank my friends who supported me during the year. Thank you to my parents who believed in me even when I had my doubts. Finally, thank you to my work for giving me this opportunity to pursue this degree.

TABLE OF CONTENTS

Page
List of Tables vii
List of Figures viii
List of Abbreviations x
Abstract xi
Chapter One 1
1.1 Introduction 1
1.2 Volunteered Geographic Information 3
1.3 Ambient Geospatial Information 6
1.4 Definitions 9
1.4.1 Definition of an OpenStreetMap Edit 9
1.4.2 Definition of an GDELT Report 9
1.4.3 Definition of a Natural Disaster 10
1.5 Scope 10
1.6 Contribution 11
1.7 Research and Objective Questions 12
1.8 Summary 14
Chapter Two 15
2.1 Background 15
Chapter Three 18
3.1 Introduction to Events 18
3.2 Oklahoma Background 19
3.3 Moore Tornado 19
3.4 Philippines Background 20
3.5 Typhoon Haiyan 21
Chapter Four 22
4.1 Data Sources 22

4.2 GDELT 22
4.3 OSM 23
Chapter Five 27
5.1 Analytical Methods and Processes 27
5.2 OSM 27
5.3 GDELT 30
5.4 Summary 32
Chapter SIx 33
6.1 Results 33
6.1.1 Display Approach 33
6.1.2 Results 33
6.2 Oklahoma Analysis Using GDELT 34
6.3 Oklahoma Analysis using OpenStreetMap 41
6.4 Typhoon Haiyan using GDELT 51
6.5 Typhoon Haiyan using OpenStreetMap 56
6.6 Population Density Comparisons 67
6.6.1 Population Density Comparisons for Typhoon Haiyan 67
6.6.2 Population Density Comparisons for Moore Tornado 74
6.7 Casualty Location within the Philippines 77
6.8 Philippines and Pre-Haiyan Event Editing Patterns 80
Chapter Seven 92
7.1 Discussion and Conclusion 92
7.2 Future Work 99
Chapter Eight 102
8.1 References 102

LIST OF TABLES

Table Page

Table 1: OpenStreetMap Edit Count for Moore Tornado 41

Table 2: OpenStreetMap Category Count for Moore Tornado 50

Table 3: OpenStreetMap Edit File Totals 56

Table 4: OpenStreetMap Edit Category Count 66

Table 5: Top Twenty Active Users and OSM Edit Count for each Feature Type 91

LIST OF FIGURES

Figure Page
Figure 1: Crowd Sourced Content Structure 8
Figure 2: OSM Data Processing Overview 29
Figure 3: GDELT Report Processing Overview 31
Figure 4: GDELT Theme Categories Present in regards to the Moore Tornado located in Oklahoma 37
Figure 5: GDELT Report Locations within Moore Tornado Track 38
Figure 6: GDELT Report Locations Referencing Moore Tornado but geo-located outside of Tornadoes Path 39
Figure 7: GDELT Report Daily Totals for Moore Tornado in Moore Oklahoma 40
Figure 8: OpenStreetMap Daily Edit Count located within Moore Tornado Track 45
Figure 9: OpenStreetMap Daily Edit Counts as Percentages located within Moore Tornado Track 46
Figure 10: OpenStreetMap Point Edit Density within Damage Track of Moore Tornado 47
Figure 11: OpenStreetMap Line Edit Density within Damage Track of Moore Tornado 48
Figure 12: OpenStreetMap Polygon Edit Density within Damage Track of Moore Tornado 49
Figure 13: GDELT Report Totals per Theme Category for Typhoon Haiyan 53
Figure 14: Daily GDELT Report Totals for Typhoon Haiyan in the Philippines 54
Figure 15: GDELT Report Density for Typhoon Haiyan 55
Figure 16: OpenStreetMap Edit Daily Totals for Typhoon Haiyan 61
Figure 17: OpenStreetMap Daily Edit Totals as Percentages for Typhoon Haiyan 62
Figure 18: OpenStreetMap Polygon Edit Density for Typhoon Haiyan 63
Figure 19: OpenStreetMap Line Edit Density for Typhoon Haiyan 64
Figure 20: OpenStreetMap Point Edit Density for Typhoon Haiyan 65
Figure 21: OSM Line Edit Density over laid on Population Density of the Philippines.. 69
Figure 22: OSM Point Edit Density overlaid over Population Density within the Philippines 70
Figure 23: OSM Polygon Edit Density overlaid on Population Density in the Philippines 71
Figure 24: Population Density Comparison in Philippines for Typhoon Haiyan 72
Figure 25: Population Density as Percentages in Philippines for Typhoon Haiyan 73
Figure 26: Population Density for Oklahoma during Moore Tornado 75
Figure 27: Population Density as percentages for Oklahoma during Moore Tornado 76
Figure 28: Provence's in the Philippines with Deaths caused by Typhoon Haiyan 78

Figure 29: Philippine Provence's with deaths directly related to Typhoon Haiyan and associated counts 79
Figure 30: Point Density Comparison for Year 2012 83
Figure 31: Point Density for Year 2014 84
Figure 32: Polygon Density Comparison for Year 2012 85
Figure 33: Polygon Density for Year 2014 86
Figure 34: Line Density for Year 2012 87
Figure 35: Line Density for Year 2014 88
Figure 36: Twitter Count for November 7-11, 2014 89
Figure 37: Top Twenty Active Users and Feature Count Totals 90

LIST OF ABBREVIATIONS

Ambient Geospatial Information .... AGI
Applications Programming Interface .... API
Comma Separated Values .... CSV
Enhanced Fujita Scale .... EF
Federal Emergency Management Agency .... FEMA
Global Database Risk Reduction and Management Council .... GDELT
Java OpenStreetMap Editor .... JOSM
National Disaster Risk Reduction and Management Council .... NDRRMC
Near Real Time .... NRT
OpenStreetMap .... OSM
Protocolbuffer Binary Format .... PBF
U.S Geological Survey .... USGS
United States Agency for International Development .... USAID
United States Dollar .... USD
Volunteered Geographic Information .... VGI
World Wide Web .... WWW

ABSTRACT

PATTERNS OF GEOGRAPHICAL CONTENT CONTRIBUTIONS MADE BY THE GENERAL PUBLIC IN RESPONSE TO NATURAL DISASTER EVENTS

Stephanie M Parker, M.A.

George Mason University, 2015

Thesis Director: Dr. Anthony Stefanidis

This thesis explores the updating patterns of Volunteered Geographic Information (VGI) in the direct aftermath of natural disasters. OpenStreetMap edits and GDELT reports are studied to explore the location, scope and types of features of such geospatial content contributions, comparing them to the impact of the events under study. The two disasters used for the study included the Moore Tornado in May 2013 and Typhoon Haiyan in November 2013. OpenStreetMap edits and GDELT reports were extracted and queried to answer the respective research questions. Our analysis shows specific patterns regarding each event separately, as well as overarching patterns regarding crowdsourced content. We observe that news reports (in the form of GDLET entries) provide initial indicators of the geographic location and extent of a natural disaster at a coarse resolution. VGI contributions are then emerging to provide a more detailed description of the scope and extent of the disaster impact area. VGI sourced data provides an insight into the infrastructure and locations regarded as either important or areas of greatest

impact. Further insight and the data analysis process used to conclude these findings and more are locate within the body of this the thesis.

CHAPTER ONE

1.1 Introduction

Within the past decade there has been a shift in how geospatial data are collected, analyzed, produced and distributed to users and applicable organizations. In the past, governmental organizations were in charge of producing maps and associated content. The role of surveyors, cartographers and geographers was to map the world and transcribe it on paper and these work roles have existed for centuries (Hakley et al 2008). However, there has been a recent change from maps being produced by official organizations to instead being produced and shared with the general public by individuals with little to no professional geography background. Geographic information created by amateur citizens is often known as volunteered geographic information (VGI). This rise in non-professional generated content has recently provided an interesting alternative to traditional authoritative information from mapping agencies and corporations (Goodchild et al 2010). Geographic information can be defined as information linking a property to a location on or near the earth's surface and possibly including a timestamp within the metadata (Goodchild et al 2010). The shift of production of professional produced maps to everyday citizens having the ability to produce VGI data led to the emergence of neogeography. Neogeography can be defined as the breaking down of traditional distinctions between expert and non-expert geographers (Goodchild et al 2010, Turner 2006).

A seminal event leading to the current popularity of VGI was the emergence of Web 2.0. The term Web 2.0 was first used 1999 but became a common reference term later on, in 2004. While some users of the internet may think that Web 2.0 is used to describe a new version of the World Wide Web (WWW), it is in fact referencing changes to how internet users are building and using World Wide Web pages. The term Web 2.0 actually is related to a new web based platform where users can customize their own applications on the WWW to meet their own design ideas, and functionality and most importantly, can create their own data or edit existing data (Neis et al 2012). One prime example of Web 2.0 as it relates to the geospatial community is the OpenStreetMap project which allows the general public to contribute geographical content to a global map database (Haklay and Weber, 2008; Zook et al. 2010)

Natural disasters are reoccurring events that occur in multiple forms across the globe. Their effects can range from minor inconveniences to residents to massive evacuation events that require local, state and even federal support. Currently, there are five phases associated with natural disasters: identification and planning, mitigation, preparedness, response and recovery. VGI has potential to be applicable in the five phases associated with natural's disaster events. Prior, to the rise in VGI produced data, maps relaying information about natural disasters including location, damage and predicted paths were slow in production since data and maps needed to be checked and verified before they could be released to the general public. However, with the rise of VGI technology this mapping or data production has shifted from being limited to only professional geographers to now the general public. There are multiple positive aspects to allowing public users to collect and produce VGI data. One benefit to this form of producing or distributing VGI data is that a greater number of maps can be produced in a shorter period of time allowing scarcer technical resources to be diverted else ware (Zook et all 2010). Instead of governments of affected areas of countries having to spend valuable resources on map production, the task of collecting information and creating maps from it can be crowd sourced. Peer prediction, crowd collaboration, or crowd sourcing refers to the ability of people from around the world to collaborate on projects that are often highly ambitious in both their scale and scope (Graham 2010) as present in VGI users trying to map large areas that have been affected by a natural disaster. An example of a natural disaster that had a wide affected area with very little data before the event was the earthquake in Haiti. The earthquake in Haiti affected thousands of citizens and most of the maps and content created for rescue personal and Haitian government was produced by VGI data and methods sourced and created from the general public.

1.2 Volunteered Geographic Information

VGI information is data that is updated or presented as changes in a geographic oriented environment. VGI is produced by volunteers through Web 2.0 applications rather than by traditional data producing methods (Goodchild 2007). VGI data can be considered geo-explicit information. In terms of being geo-explicit, VGI data provides geographical data that is clearly defined in relation to a specific location. An example of this would be a user updating the current road structure to include a new road or building. One downside observed about VGI data is it tends to only be updated in times of economic or personal distress such as a hurricane, flood or fire. This inconsistent updating cycle can present data consumers an updated locational situational awareness in times of distress but does not provide an original or ‘before’ overview of the affected area.

Traditional methods of map making by government agencies usually have an associated delay of map production and release since content produced normally has to go through revision process which in times of a disaster, could lead to outdated maps by time of release. However, VGI content can quickly capture changes in landscape as fast as they occur (Goodchild et al 2010) allowing for faster and more accurate maps.

In most cases VGI sources lack the mechanisms that accurately ensures that content is consistently updated to provide a valid baseline. This aspect plays a factor when VGI sources are used to try and complete change detection studies of geographic locations. One method to encourage social media users to use VGI sources to update ‘before” or baseline maps is for local or state governments to encourage data calls or fill in missing data locally. An example of an outreach program is the current OpenStreetMap Tasking manager hosted by the Humanitarian OpenStreetMap Team. This website is dedicated mapping areas that are not well represented in the context of mapping i.e. detailed topographic data is missing. The website posts current projects as well as percentage completed in order to encourage users to participate. Examples of current VGI sources include OpenStreetMaps (OSM), Google Map Maker, and Wikimapia. Unlike other platforms that rely on user contributions such as Wikipedia and Flicker, the collected information is not about a particular topic or image. Instead this collected info contains more specific details about elements such as streets, points of interest, or buildings which always include geographic reference: this is VGI or crowdsourcing geospatial data (Neis et al., 2012).

The types of forums where Volunteered Geographic Information content is created, stored and shared continues to grow and while the names and uses of these platforms have the potential to change and expand, the groups of foundation structure remains the same. Examples of these groups include forums, blogs, social networks, video and photo sharing.

The versatility that can be associated with Volunteered Geographic Information allows it to have the potential to be used in response to a natural disaster. Examples in a how Volunteered Geographic Information can be used include immediate access to information, familiarity in a frightening time, and real-time information and situational awareness as well as a way to reach rescuers and request help (Bowers 2014).

A question that has risen along with the interest and production of VGI content is why an individual would want to provide their time and services for little to no compensation. While there are multiple opinions on what a VGI contributor's motivations may be, the most common theory is that participants have a desire to gain cultural capital by sharing and creating information and also a widespread desire to help other people. Another, core motivation suggested for the reason behind the production of volunteered geographic information or VGI is likely inaccessibility and cost of accurate sources of geographic information (Zook et al., 2010).

While there are positives with VGI data there are also concerns and hesitations with using this publically derived data. One such concern is that contributors are acting independently and with loose coordination in regards to each other (Haklay 2010). Redundancy of data and maps can be a concern if collectors fail to communicate where they are working or where the final production content is located. However, the main concern is with the accuracy of the data. One prominent concern with accuracy is that the traditional means of mapping in disaster situations of will they contain flaws that would have been prevented by professional cartographers. However, while not a finite answer, VGI accuracy studies have shown that with enough people working together any errors by one individual can easily be corrected by the collected group (Zook et al., 2010)

1.3 Ambient Geospatial Information

Ambient Geospatial Information (AGI) is social media where the main objective of the information being shared publically is not directly based on geography. (Stefanidis et. al., 2013). Unlike VGI data which is geo-explicit, AGI information can contain geo-implicit indicators or information but this information is not always present or included within the metadata. AGI information expresses geographical information that is implied or the information that is shared does not directly state a specific location as to where the data is applicable. AGI provides useful information to both users and consumers, as it can provide valuable context about how information flows among users and provide an overview of how social networking responds to information that is constantly updated. An additional benefit of AGI is it can provide a general overview of locations that have active current events. Examples of these active events would be the protests in Missouri or current UN sanctions within North Korea.

The difficulty with using AGI derived information is linked to the aspect that most data is not directly geo-located. This means the consumer of the data must rely on references within the shared data or metadata so as to locate the geographic location of the information. Also, the publisher of the data may be providing data from an alternate location that is not the exact location represented by the posted location. This makes organizing and validating the quality and accuracy of the data difficult. Currently, there is no defined or concrete method to validating the accuracy of AGI presented information, however multiple methods and solutions have been proposed. Currently, there are more AGI dedicated source sites than VGI sources. Current social media platforms programs using AGI methods and algorithms include Twitter, Facebook, Flicker, and YouTube.

Figure 1 shows the structural breakdown of the crowd sourced content that is referenced within this thesis. The purpose of this chart is to help clarify differences between AGI and VGI content.

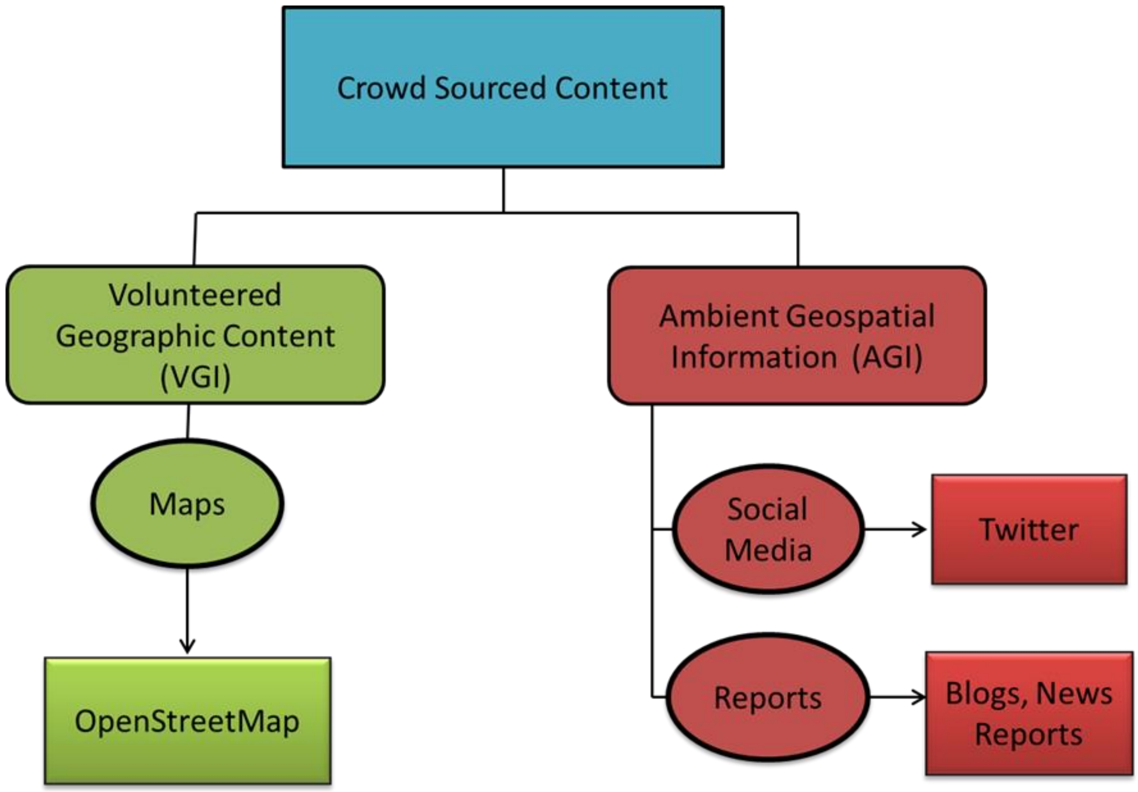


Figure 1: Crowd Sourced Content Structure

1.4 Definitions

1.4.1 Definition of an OpenStreetMap Edit

For the purpose of this research, an OpenStreetMap (OSM) edit is defined as a change to the current map architecture within the confines of a predetermined area. The OSM architecture allows for users to alter features within the interface. The research and edit area for each natural disaster covered in this study varies according to the extent of damage or path of the natural disaster. Each of these distances is defined further in each disasters introduction. For the purpose of this research three different types of edits were looked at and tabulated. The edits that are available to alter features within the OSM architecture are: point, line, and polygon. An example of a point edit is a new street light being installed. An example of a line edit is a new street and an example of a polygon edit is a new park or a new building constructed. In terms of editing OSM features, there are three main options that can be used: creating, modifying, or deleting a feature.

1.4.2 Definition of an GDELT Report

For the practice of this research period a GDLET report is defined as a change to the current architecture based on the knowledge that something is currently occurring at a specific location. An example of this is updating the location of a super storm cell,

The type of forums where Volunteered Geographic Information content is created, stored and shared continues to grow and while the names and uses of these platforms have the potential to expand, the group of foundations structure remains the same. Examples of these groups include forums, blogs, social networks, video and photo sharing.

1.4.3 Definition of a Natural Disaster

The definition of a natural disaster for this research is a naturally occurring, non-man-made event that causes negative economic and infrastructure impact. These events can include but are not limited to severe floods, hurricanes and earthquakes.

1.5 Scope

The scope of this study includes the updating patterns of OpenStreetMaps edits and GDELT user reports in response to natural disaster events. Research areas include what events create the largest response for edits and what type of edits are posted. The topics discussed include VGI and AGI social media sources occurring within the bounds of a natural disaster and what social media outlet shows the most edits or updates in response to a natural disaster. However, this thesis does not cover the accuracy that has been associated called into question and as a result explored through numerous studies, with VGI and AGI data. It is assumed that all edits and reports are relevant to the event meaning they have not been posted maliciously in an affect to provide false data.

The location of the study includes two natural disaster events, one of which occurring in the United States and the other one within the Philippians. The data extracted and analyzed was collected from the GDELT database and OpenStreetMaps associated interfaces. The study timeframe includes the months of May and November of year 2013.

1.6 Contribution

Volunteered Geographic Information content is generated about events from across the globe and posted through video, chat and image platforms on a daily basis. The use of Volunteered Geographic Information continues to grow, presenting an opportunity for further research and discovery into how VGI can be used to in regards to natural disasters. The earthquake occurring in Haiti in 2010 was an example of how Volunteered Geographic Information can be used to map areas of damage and provide information valuable to first responders and the Haitian citizens than what has been seen previously. This research helped explore user editing and updating patterns not only after a disaster has occurred but also posting habits or patterns before the disaster. Rather than having user provided information reported after the event, this research explored the Volunteered Geographic Information user cycle which began before an event and ends after a specified research timeframe.

This research is concerned with not only looking at how many user updates and edits were recorded in response to a natural disaster but also explored further into the Volunteered Geographic Information content that was posted. While the numbers of edits or reports were recorded, the location, type and date of the media content was also recorded and analyzed. This deeper look into the content that Volunteered Geographic Information users are publishing gives a deeper understanding into what users and consumers consider important. When Hurricane Sandy impacted the New Jersey Coast, one facet of the resulting Volunteered Geographic Information content showed homeowners posting information on damaged houses and their current locations (FEMA 2013). One interpretation of this posting is people are concerned with their personal property and where their surviving relatives are currently located. The research and results from this study provides further understanding into what edits and updates social media users and consumers consider important. This information or knowledge can be used in various ways. One example includes aiding authoritative organizations when trying to reach the general public concerning natural disasters updates and notices. During Hurricane Sandy and its aftermath, the Federal Emergency Management Agency (FEMA) posted updates via twitter regarding the status of the storm as well as providing information for help after the storm (FEMA 2014). Understanding the content and timeframe of user edits and updates improves the communication links between members of society.

1.7 Research and Objective Questions

When preliminary research for this thesis began, it was intriguing at the expansive amount of research that was being conducted regarding social media and terrorism. However, the topic of applying social media users' updates and edits in disaster response applications was less explored. One highly publicized event regarding social media users' edits and updates were the earthquake in Haiti in January of 2010. It is from this user contributed event that helped stem the research ideas for this thesis. The response from Volunteered Geographic Information users including OSM edits and GDELT user reports was impressive since a Volunteered Geographic Information response of this magnitude and variability had never been so widely publicized. With the Haiti response and event feedback, it became obvious to media users that Volunteered Geographic Information contributors located within both an affected area and outside, could contribute vital data content. This publically available content could potentially be used in multiple ways including but not limited to rescue and recovery efforts. It is with the Haiti disaster in mind and initial research that two overarching research questions were constructed.

The initial or primary hypothesis asks the questions of what are the updating patterns of VGI users with regards to natural disaster events. This initial hypothesis leads to three detailed questions: where is the geographic location of the edits in relation to the event, what type of map features were edited, and what is the scope of the edit. Or report. This paper postulates that although the patterns of contribution and access to OpenStreetMap may vary across the globe, in the aftermath of a natural disaster, users will edit the OpenStreetMap interface with similar levels of detail allowing direct correlation of damage location and extent. In regards to GDELT user updates, this paper proposes that user updates referencing the natural disaster will be almost non-existent before the event but will be present in higher capacity after the natural disaster occurrence. GDELT user updates will also be predominantly located within the area affected by the natural disaster.

As research progressed the realization emerged that there was another research area that warranted exploration. This second area of research proposes the question: Does the amount of ground surface area affected by a natural disaster event create a correspondingly larger increase in user updates and edits.

This thesis aims to answer and explain the results to the above proposed questions thru various scientific methods including data analysis, observation, and background research.

1.8 Summary

The remainder of the thesis is arranged as follows: section 1.5 present the scope, section 1.6 presents the contributions and section 1.7 presents the research objectives of this thesis. The chapter two presents prior work and research conducted within the field of Volunteered Geographic Information as it relates to the thesis topic. Background information on events and locations follows afterwards in chapter three and chapter four covers data sources. Chapter five leads into an overview of data preparation steps performed in order to process and analyze relevant data. The final chapters present the results and include a discussion of using OpenStreetMap Edit and GDELT Report data followed by conclusions and future work in chapter seven.

CHAPTER TWO

2.1 Background

The applicability and extent in which VGI and AGI can be used has not been clearly defined. The current applications and Web 2.0 platforms that utilize VGI continue to expand. However, the topic of VGI provides a vast research area that allows for extensive exploration regarding its uses and how the user generated data can be used to predict social behavior, location analysis and general awareness. Listed below are a sample of studies or articles that explore the sample of the applications and uses of VGI.

Multiple past studies have linked the growth of VGI with the increase of available technology of Web 2.0. VGI provides a unique data collection environment because the content of the information, the technology for collecting the shared VGI data, quality, and the social implications it can create (Elwood et. al., 2014) are not predefined. The interesting aspect of VGI data is that it is growing and how to accurately address each user's needs from the VGI data can be addressed in multiple ways. Currently, there are studies on how VGI can be used to answer specific users' needs such as water and soil conservation. A defining point is VGI is adaptable to every users needs but no concreate method or structure exists that explicitly dictates the outline of how a user must use VGI to answer their needs. (Werts et al. 2012).

Another topic being explored is the subject of crowdsourcing. Crowdsourcing relies on gaining information from more than one person. The term crowdsourcing originally was developed from a concept of outsourcing business related operations and tasks to cheaper and usually more remote locations. (Haklay et al 2008). One aspect of crowdsourcing that is currently being explored focuses on proposed methods to organize and gain intelligence from data provided by consumers or the community (Rice et al., 2013).

With the continued rise of available crowd sourced content available to consumers, a common issue of how to accurately map this data with minimal errors needs to be addressed. One method proposed to handle this data quality problem proposes a method that allows the collection of information relevant to a specific topic from feeds that most likely contain more data than what the end user is looking (Stefanidis et al. 2013). Essentially, the end product is a data mining process that is capable of collecting only relevant data. Detailed methods on how to handle crowd sourced information as well as emerging trends are also being presented to federal organizations, such as the Army Corps of Engineers, whom believe that crowdsourcing is a valuable way of collecting information (Rice, et al., 2014).

With user produced geospatial data, it is hard to gain a grasp on the extent of situations in which the associated the geospatial data is applicable. Proposed methods of how to handle the data include example systems that suggest examples of frameworks for harvesting geospatial data and gleaming both qualitative and quantitative data. (Croitoru et al., 2013).

It is not unusual for new users or consumers of social media to be confused about the differences between Volunteered Geographical Information and Ambient Geospatial

Information. However, with the increase in crowd sourced content including AGI and VGI sources the increase in articles and studies explaining the structure and applicable uses of this data have risen as well. (Stefanidis et al., 2013).

Currently 25.9% of the populations currently have access to internet but 67% have access to mobile phones (Zook et al 2010) and as a result the term OpenStreetMap has likely been introduced. However new users unfamiliar with the sources of VGI may be unfamiliar with the extent of which OpenStreetMap can be applied as well as what the support structure that makes up OpenStreetMap.

Along with journal articles exploring various aspects of social media types and uses, there are books focused on introducing the concept of Volunteered Geographic Information and how it is being used in todays' world. The first book reviewed in preparation for this study, was *Enabling Comprehensive Situational Awareness* by Susan Radke, Russ Johnson, and Jeff Baranyi. This text provided a general background on how social media can be used in planning, mitigation and recovery in reference to natural disasters. Another book dealing with social media integration was Disaster Communications in a Changing Media World by George Haddow and Kim Haddow. This updated volume includes examples of how social media including twitter can be used to create a two way communication between producers and consumers and the general patterns observed.

CHAPTER THREE

3.1 Introduction to Events

On order to test these hypotheses natural disaster events were required that fit certain criteria. The first criteria for these natural disaster events had to occur within the year 2013. This timeframe was chosen as an event occurring in 2013 would have leveraged fully matured Volunteered Geographic Information platforms and user participation models facilitated by the Web 2.0 paradigm. In 2013 both OSM and the GDELT project were being used worldwide providing users the ability to update and report on natural disaster events and responses in near real time. The second criterion the event must have met successfully was that it was considered a natural disaster.

Social media has the potential to cover a multitude of natural disaster occurrences due to its platform diversity and ease of use. However, the focus of this thesis is on natural disaster events which do not include manmade natural disasters events such as oil spills or chemical releases. The third criterion that an event had to have in order to be considered was in relation to test my second hypothesis. The second hypothesis of this thesis was in regards to the area affected or damaged by the event. In order to achieve this specific criterion, the two events selected needed to be of different categories and not located within the same continent.

In order to help locate natural disaster events that fit the above stated criteria, government and national disaster reference sites including the United States Geological

Survey (USGS), and USAID were consulted. After review, and two natural disaster events that occurred in 2013 fit the above stated criteria.

3.2 Oklahoma Background

Oklahoma received statehood on November 16, 1907 making it the forty-sixth state to join the United States. According to the US Census bureau in 2013 Oklahoma's population was around 3,850,568 citizens or about 1.22% of the total United States population and the 28th most populated state. Oklahoma combined land and water area is approximately 181,050 km^2 ranking it the 20th largest state.

Oklahoma is located in what is commonly known as tornado alley. Tornado alley is term given to southern plains of the central United States that regularly experiences a high rate of tornadoes occurrences each year (NOAA 2014). This area is where cold air from Canada, warm air from the Gulf of Mexico and humid/dry air coming off of the Rocky Mountains intersect. These intersections of different air temperatures and humidity lead to conditions where thunderstorms can occur that have the potential to produce tornadoes. The first recorded tornado occurred in 1893 and the most recent one was in 2013. Since 1893, Oklahoma has had over one hundred fifty six tornadoes. The prime month for a tornado to occur in Oklahoma is in May, more precisely early May for the eastern portion of the state and late May for the west part of the state. (Brooks et al 2013).

3.3 Moore Tornado

The first event selected was the Moore Tornado in Oklahoma. On May 20, 2013 at approximately 2:56 pm an EF1 tornado touched down on the outskirts of Newcastle

Oklahoma. Within the next thirty minutes the tornado would increases in intensity and top out as an EF 5 tornado with winds speeds ranging up to 210 mph. This tornado would be known as the Moore Tornado due to most of the infrastructure damage located within the city of Moore Oklahoma. The tornado was on the ground for approximately forty minutes and passed through the counties of Grady, McClain, and Cleveland. The tornado affected a 27.4 kilometer stretch of ground and in its aftermath twenty four people died with another 377 injured. It is estimated that the tornado was over 2.1 kilometer wide at its peak and caused over 2 billion dollars' worth of damage to the surrounding infrastructure.

3.4 Philippines Background

The Philippines are constructed of seventy regions split into eighty-two provinces. The Philippines are the seventy-third largest county with a total land mass of 298,170 km^2 and a coastline of 36,289 km. The estimated population in 2014 was 107,668,231 citizens making it the thirtieth most populated country.

The Philippines are located in what is has been coined as the typhoon belt (World Fact Book 2014). This area is located between the South China Sea and Pacific Ocean and is called so due to the amount of yearly typhoons and tropical storms that occurs in this swath of ocean. Typically, each year the Philippines are affected by fifteen typhoons or tropical storms but only directly impacted by five or six. In 2013 twenty five typhoons or tropical storms affected or directly impacted the Philippines.

3.5 Typhoon Haiyan

On November 8, 2013 Typhoon Haiyan or as it is commonly known in the rest of the world, Typhoon Yolanda, swept across the Indian Ocean and made landfall in the Philippian. Typhoon Haiyan crossed the Philippines approximately in the direction from Southeast to the Northwest. At the time of landfall, Typhoon Haiyan was considered a Category 5 typhoon according to the Saffir-Simpson scale and had a width of roughly 600 kilometers. The high Saffir-Simpson rating resulted from Haiyan having the sustained winds speeds of 195 mph and gusts ranging up to 378 kph and a maximum low pressure of 895 mbar. At the time of this writing Typhoon Haiyan is one of the strongest storms ever recorded and considered the deadliest typhoon to hit the Philippines, Southern China and Vietnam. The country that had the highest loss of life and infrastructure damage was the Philippines. After typhoon Haiyan dissipated on November 11, 2013 over 16 million people had been affected by the storm. Over 6,193 deaths have been attributed to the storm. Typhoon Haiyan displaced over 4.1 million people and destroyed or damaged over 1.1 million buildings. The estimated damage cost created by the storm ranges over 2.86 billion dollars (USD). Due to modern weather forecasting ability and technology, the typhoon was reported and status updates provided before it made landfall. However, the exact extent and path was not known and could only be predicted. Due to the social media outlets, knowledge of the storms likely approach through the Indian Ocean was broadcasted and updated across the globe.

CHAPTER FOUR

4.1 Data Sources

For the execution of this thesis, two main data sources were used. The first is the GDELT program which encompasses a wide range of media sources. The second data source used was OpenStreetMap (OSM) which encompasses users updating infrastructure aspects of a specific area via edits.

4.2 GDELT

The GDELT project has quickly gained attention since its initial launch due to its ability to collect and display news related content data in near real time to the requesting user. The GDELT project was as an initial quest to understand how society on a global scale interacts and connects with each other by communication and then the recourse and reaction in response to news posted via the Volunteered Geographic Information resources. The GDELT Project is considered a big data program that currently includes over a quarter of a billion event records categorized into three hundred separate categories to make retrieval and inquiry easier for the user. The overarching goal of the GDELT program is to help understand the connection between worldwide communication and behavior. The GDELT program archive currently covers event from 1979 to the present and covers the entire globe. The GDELT project gains or updates its vast data archive by using and collecting information from open source outlets including but not limited to broadcast, online, and print sources. It uses both foreign and domestic news feeds including the Washington Post, BBC News, and Agence France Presse to name a few. What can be considered unusual about the GDELT project is that is useable by users who have a range of skill levels. Novice to advanced users can benefit from the data located within the GDELT project. Novice users can query the analytical service network to use a set of predefined queries all the way to an advanced user who can download chunks of data and query it using their own queries. The GDELT project helps to connect users from across the globe via a common network that allows them to share their views and opinions to events, people and each other in a user friendly format.

4.3 OSM

OpenStreetMap (OSM) was founded by Steve Coast in 2004 after his frustration following the lack of publically available maps of the United Kingdom. Mr. Coast officially founded OSM on July1, 2004 and since then OSM use has grown exponentially around the globe. OSM was officially launched in 2004 but the age of the base layer that was the original foundation of the map can vary. The United Kingdom allows for historical maps to be used to define areas of the UK, however not all historical maps are loaded into the OSM database for use. In 2006 Yahoo! agreed to allow OSM to use its aerial photography to be the background for the OSM map. From this point on, the age and accuracy of the map depends on user updates. Users can update the map in a variety of way including aerial imagery, GPS traces, foot path, and vehicle such as car or boat etc. OSM is currently one of the largest free editable map interfaces that cover the entire globe. As of current, the map is visible to anyone who has internet access, but only registered users can make edits within the map interface. The map extent covers the entire globe and registered users can edit anywhere they wish. What makes OSM so appealing is that it can be used by individuals at all skill levels and there is no fee associated with an OSM account. Novice to professional cartographers can benefit and find information from OSM. OpenStreetMap has the potential to be an appealing source to both novice and professional users for a multitude of reasons. OSM is considered open data which allows for it to be downloaded and used easily. Since it is open data, minimal copyrights exist for it use and distribution. OSM is available to users across the globe which allows people located in different parts of the globe to be able to contribute to the same project. This communal effort was evident during the Haiti earthquake. Using OSM, users located within Haiti were able to express their needs and map damaged locations, still viable roads, and even trapped citizens to the world. Users located outside of Haiti were also able to contribute due to OSM interface and provide valuable information to authorities and rescue personnel including where help was needed from translated user posts to where emergency response teams could set up relief sites.

Founded at University College London its goal is to create free database with geographic information of the entire world and detailed introductions to the project have been published, a plethora of spatial data such as roads, buildings, land use areas, point of interest is entered into the projects database (Neis et al 2012) OSM is probably the most extensive and effective project currently under development (Haklay et al., 2008). Aims to create a map of data that's free to use, editable, and licensed under new copyright schemes (Haklay et al 2008). OSM unlike Wikimapia requires users to register for an account in order to edit the map interface. Wikimapia currently has an open-for-all approach does not require users to register for an account in order to edit (Haklay et al 2008).

The OSM interface relies on volunteers that do not necessarily have professional qualifications and background in geo-data collection or surveying (Neis et al 2011) Geographic information is the OSM database such as roads, land use information, or buildings is stored by the use of three object types: nodes, ways, relations, a node in the database contains location information of the point that form of latitude and longitude. Lines such as roads and polygons are stored as ways and relations define logical or geographic relationships between the objects. At the end of 2011 over half a million members had created accounts for OSM and by June 2012, OSM had over 640,000 registered members and counts as one of the most impressive sources of volunteered geographic information (VGI) on the internet. (Neis et al 2012).

Both space borne and airborne imagery allows OSM members to digitize data such as streets from the image. Positive aspects of applying the OSM interface to aerial imagery to derive VGI information is it is simple to add new data, but negative aspect is imagery that OSM can be dependent on for roads and ground features can be out dated and/or improperly georeferenced (Neis et al 2012). Aerial imagery does not provide metadata such as street or road names which can make labeling difficult.

User contributed geographical information is the primary backbone or core of OSM. It is how it grows and becomes more accurate to changing conditions on the earth's surface. OSM received a boost in data accuracy and location in regards to roads when the US Census Bureau loaded the Topologically Integrated Geographic Encoding and Referencing (TIGER) database for US streets and other primary linear features. (Neis et al 2012). TIGER files are files that contain features such as roads, railroads, rivers, as well as legal and statistical information for geographic areas (US Census 2014).

OSM can be considered a form of crowdsourcing due to the way data can be collected and implemented into the OSM infrastructure. Crowdsourcing developed from the concept of outsourcing in which business operations are transferred to remote many times cheaper locations. Crowdsourcing is how large groups of users can perform functions that are either difficult to automate or expensive to implement (Haklay et al 2008)

OSM also has the infrastructure that allows users to update and continue to be human sensors even when not connected to the internet. Java OpenStreetMap Editor (JOSM) an editing suite with an interface more akin to traditional GIS packages, lets users import, edit, and tag OSM data offline and allows bulk uploads of OSM updates through the OSM applications programming interface (API) (Haklay et al 2008). OSM provides the infrastructure that can render raw data provided by the user into cartographically acceptable street map content (Goodchild et al., 2010).

CHAPTER FIVE

5.1 Analytical Methods and Processes

For this thesis two main datasets were compiled and analyzed. The overview process for how the data was processes and analyzed was very similar for both OSM and GDELT sources. Although the process was similar each data type had specific steps that were a result of the nature of the data. Along with using GDELT and OSM data, two programs were critical to analyzing the data. The ESRI owned and designed ArcMap was used for displaying and executing spatial based queries and Microsoft Access provided the background database for storing and querying the database as a whole.

The timeframe for which data was collected was two weeks. Data collection started one week prior to and one week after the event. The logic for this was for two reasons. The first reason was the chosen natural disaster overall only lasted a few days. In the case of the Moore Tornado, the tornado was only on the ground for a little over forty minutes and Typhoon Yolanda was over the Philippines for approximately twenty hours. The second reason was due to the size of the generated GDELT and OSM files. A two week period allowed enough data to be collected without overwhelming the processing capabilities of the computer used for this study.

5.2 OSM

There were two file sets of OSM data, one for the Moore Tornado in Oklahoma and the other for Typhoon Haiyan. Each file contained three layer files of edits including a point, line, and polygon layer. Figure 1 gives a general overview of the OSM data download, manipulation and analyses of the data. The OSM main page allows for small area downloads however the research area for both case studies exceeded the allowing download limit. To mitigate this problem a third party company was used to download the required data. GeoFabrik is a German based company that provides OSM edit data for areas as small as states to the entire globe. GeoFabrik was used to download both required files of Oklahoma and the Philippines.

After the files were downloaded, each one was uncompressed by an exe file known as OSMconverter. The original downloaded files were in a .pbf format which and if the files were not changed into .osm format the analyzing software ArcMap would be unable to recognize the data. Once the OSM files were uncompressed they were imported into ArcGIS for display and analysis. Since .osm files cannot be recognized by ArcMap initially, a special extension was downloaded aptly called OpenStreetMap toolbox. Once the OSM file was rendered in ArcGIS, the data points and shape files were created to answer the research questions and hypothesis.

One research question relies on where OSM edits and GDELT user updates are located within relation of the natural disaster event. In order to answer this question a shape files of the natural disaster were downloaded from third party sources. The Moore Tornado track was sourced from NOAA and Typhoon Haiyan Path through the Philippines was provided from an open source cite dedicated to social media mapping of Typhoon Haiyan.

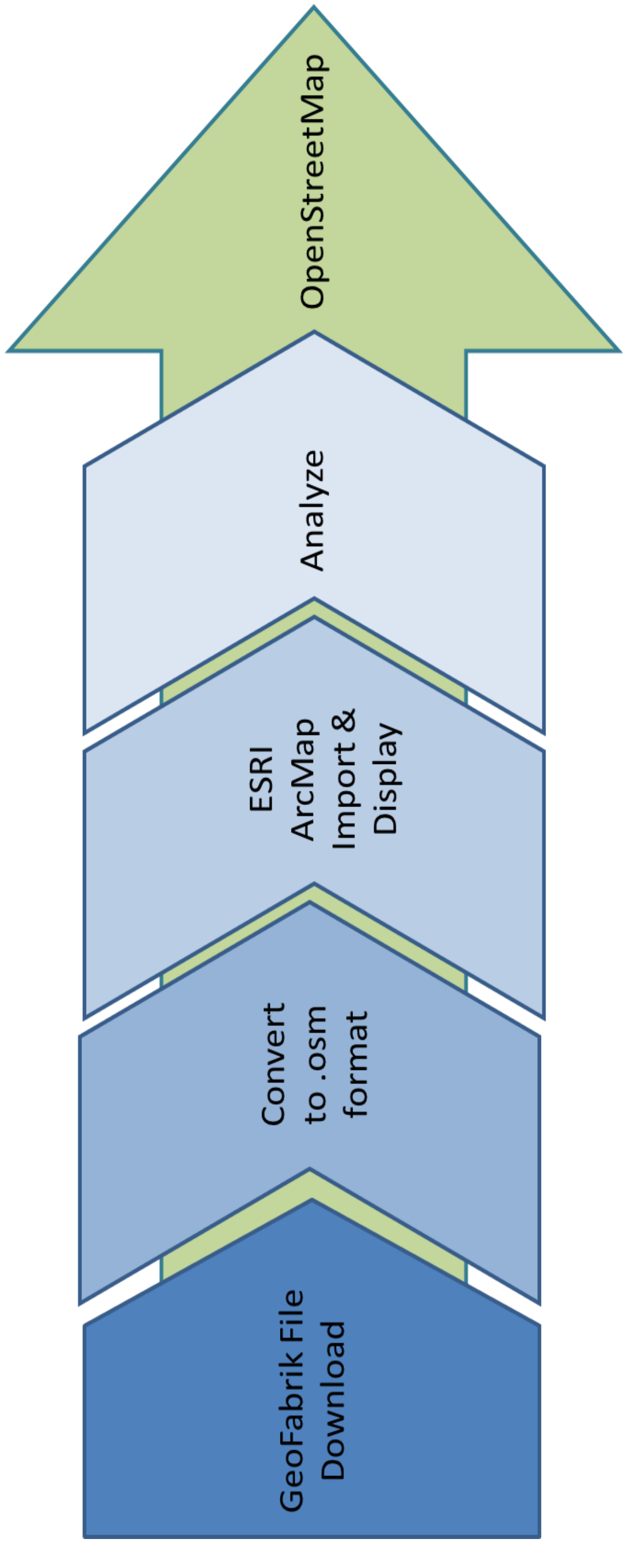


Figure 2: OSM Data Processing Overview

5.3 GDELT

In order to properly display the GDELT data, a different approach had to be taken than that was explained in regards to OSM data. This altered approach was due to the nature of the GDELT data. Figure 2 gives the general overview of the process used to analyze and process the GDELT data.

The GDELT data was in initially downloaded as a zip drive that contained a comma separated file. While ArcMap has the capabilities to display .csv files as coordinate points, the size of the file prevented the data from being properly displayed. To mitigate this problem Microsoft Access was used as a database in which all GDELT csv files could be stored and queried and the resulting data exported to ArcMap for display and analysis.

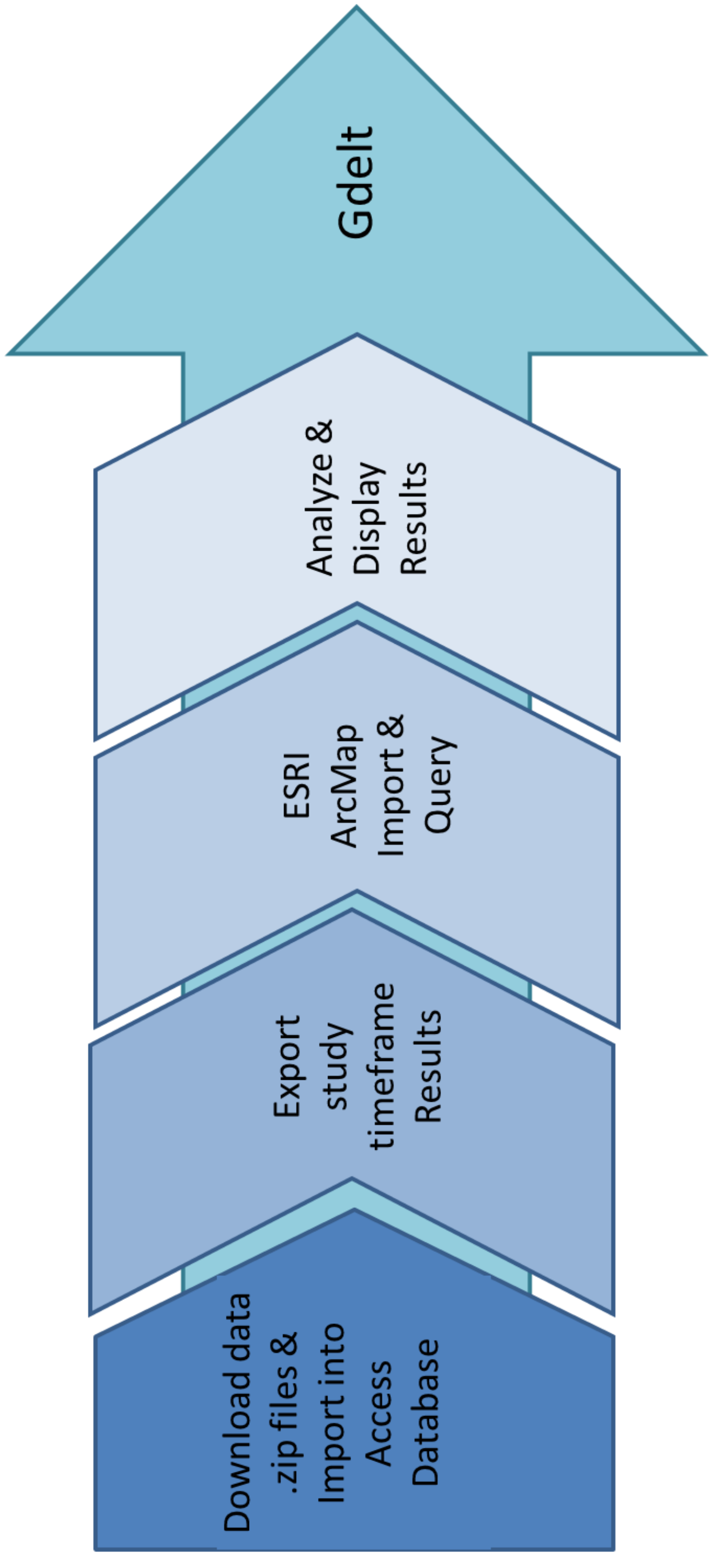


Figure 3: GDELT Report Processing Overview

5.4 Summary

While the overall process of how GDELT and OSM files were downloaded processes and analyzed is very similar, each process was tailored to the nature of the origin format of the data. Due to its size, OSM files needed to be converted into native .osm format while GDELT .csv files had to be converted in order to allow spatial reference for each update to be added. All of this processing was necessary in order to allow the data to be able to be analyzed and displayed. While the exact details and number of files and shape files will not be discussed in this thesis, associated ESRI shape files and .csv files were created in order to answer the research questions.

CHAPTER SIX

6.1 Results

6.1.1 Display Approach

The OpenStreetMap edit data and GDELT report data used for the analysis of this thesis provide a unique data environment in that both file sets contain a numerous amount data points that can be acquired and displayed to answer not only the focus of this thesis but as well as answer other VGI inquires. To make display of the results easier to display and present the rest of the results section of thesis will follow the following format where each the final results are split into four sections: GDELT report on Oklahoma, OSM edits in Oklahoma, GDELT report in the Philippines, and OSM edit results in the Philippines.

6.1.2 Results

The results used to support or answer the hypothesis and associated research questions, were displayed within the ArcMap interface if the original data file ended in an .osm file extension however if the file ended in a .csv file extension, the original data was stored in the microsoft access database and then exported to the ArcMap interface for display.

If the results stemmed from OSM data, then the results were feature layers that could contain point, line or polygon edits. If the original data was from the GDELT database the results were of GDELT reports that corresponded to either Typhoon Haiyan or the Moore Tornado and were within the set study timeframe.

The results calculated in this thesis were affected by the information that was originally contained in each data file. After the queries are applied to the initial data files, the information categories contained in the initial OSM and GDELT data files can be counted on to be present in the resulting query files. Knowing or understanding the categories types within the initial files, helped to design queries that allowed for only relevant OSM features and GDELT reports to be selected. If the information categories were ignored and the entire initial files were queried then irrelevant data not related to the events would be included and would have skewed the results. Also, removing irrelevant data prior to displaying the results shortened the processing and rendering time that was required for ArcMap to display the results. This efficient data analysis is important because the original data files were large and without understanding what is contained in the data, the results cannot be properly analyzed. To make the presentation of results more concise and easier to read, each location with regards to its analyses with OSM and GDELT data is presented separately.

6.2 Oklahoma Analysis Using GDELT

The preliminary GDELT file used for the Moore Tornado analysis had over nineteen thousand geo-located user edits. Of those updates only about 87% were used based on theme designation. GDELT data has a specified theme field that carries a natural disaster tag associated which was used to help determine whether a GDELT edit was about the Moore Tornado or another incident. Theme designation was taken into account based on the preliminary hypothesis to observe the scope of the update. Figure 3 shows the relevant theme categories present within the GDELT file. After removing the non-relevant categories such as snowfall or blizzard based reports, there were twenty-four relevant themes. The top five highest used themes in order of highest to lowest are: Tornado, natural Disaster, Twister, Severe Weather, and Floods.

The next visual representation of the GDLET reports data concerns the geolocation of report in regards to the damage path of the tornado. Figure 4 shows a map of the Moore Tornado in regards to its location, length and intensity. The tornado did pass through all six EF intensity levels; however it was considered an EF1 or EF2 for most of its ground time and damaged over 18.4 square kilometers. When the relevant GDELT reports were displayed only twenty-seven edits or about 0.2% were located within the Tornado disaster area and they are all centrally located around one area and located within a section where the Tornado was at a lower (EF0 to EF1) intensity.

With only 0.2% of GDELT reports within the Tornado disaster area, the question of where are the rest of the updates located was raised. Since GDELT reports are represented as points with the ArcMap interface, it was more beneficial to show the density of where the points were located versus displaying a map with over sixteen thousand points. To locate the remaining relevant GDELT report locations a map was created to show the density of the reaming reports in relation to the tornado damage track. Figure 5 shows the density locations of where update locations are in correlation the tornadoes damage track. The density results show two major and one minor location of user updates. The densest location is around forty-four miles from the tornado area north of the city of Shawnee Oklahoma. The second densest are was located over the state capital of Oklahoma, or Oklahoma City. The distance between the state capital and the city of Moore is around twelve miles. The third location that showed up as having a collection of GDEIT reports is Shawnee Oklahoma. Although it does not show the density level of the other two locations, it does have enough geo-located reports to make a presence on the map.

The Moore Tornado occurred on May 20, 2013. To help understand and visually see the updating patterns posted by GDELT users, the number of GDELT reports was calculated and graphed to see if any trends emerged. Figure 6 shows the number of GDELT reports per day for the research period. As the dates progress, the report count stays relatively low with two minor spikes at days two, four, and seven. After the tornadoes occurrence, the amount of reports spikes considerable and then begins to decline as days pass after the event. Overall, the GDELT report daily totals show a left skew pattern with minor date spikes preceding the event.

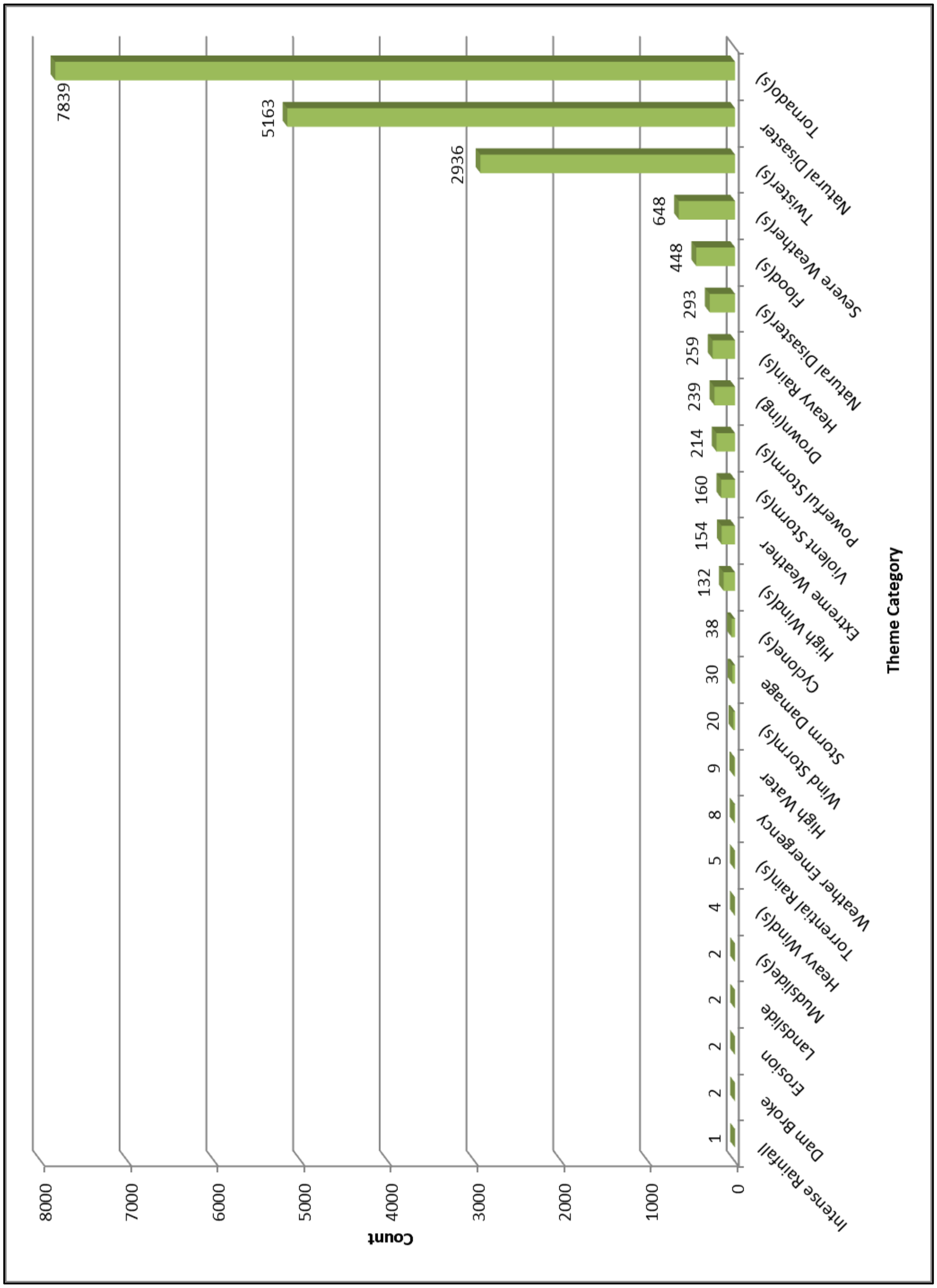


Figure 4: GDELT Theme Categories Present in regards to the Moore Tornado located in Oklahoma

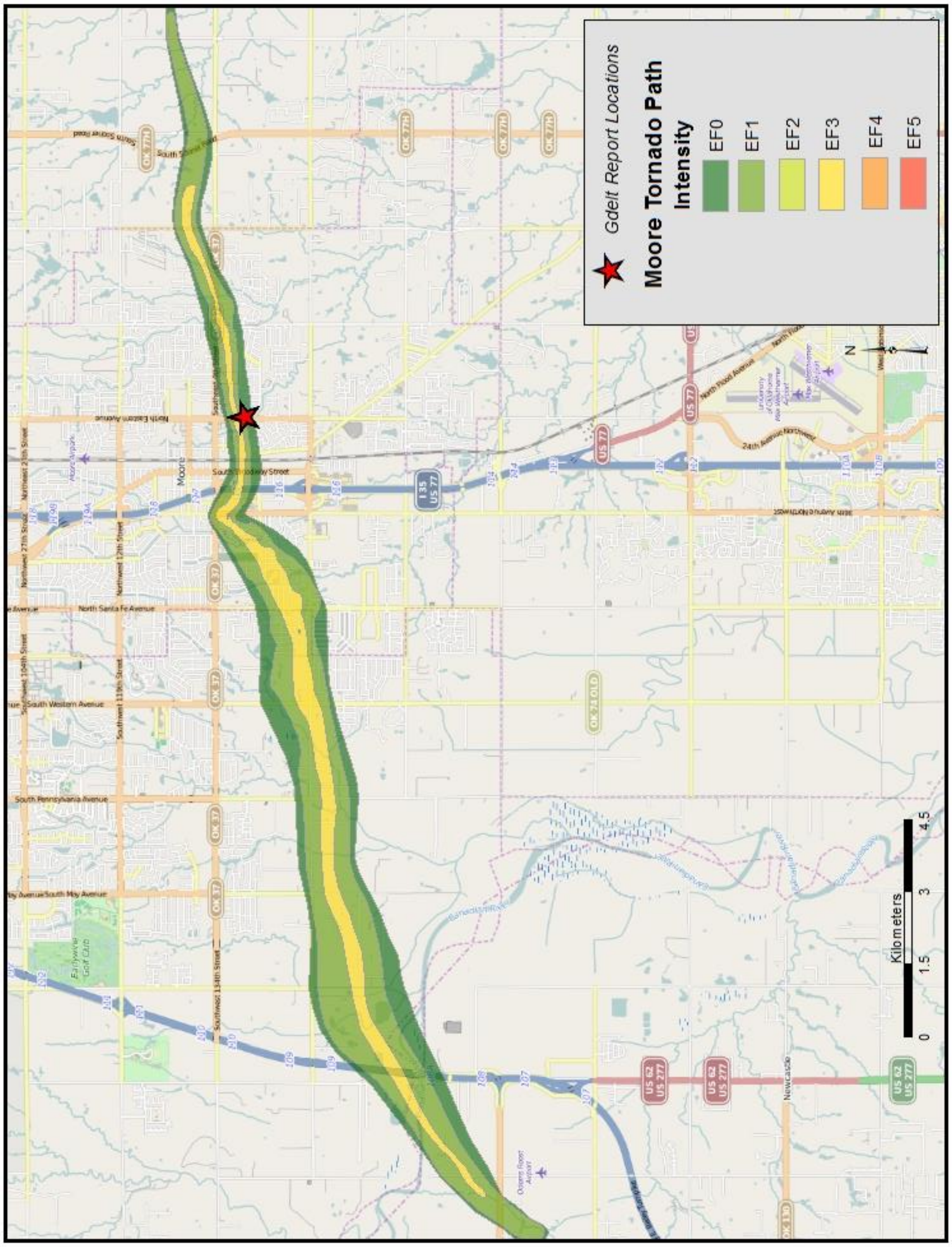


Figure 5: GDELT Report Locations within Moore Tornado Track

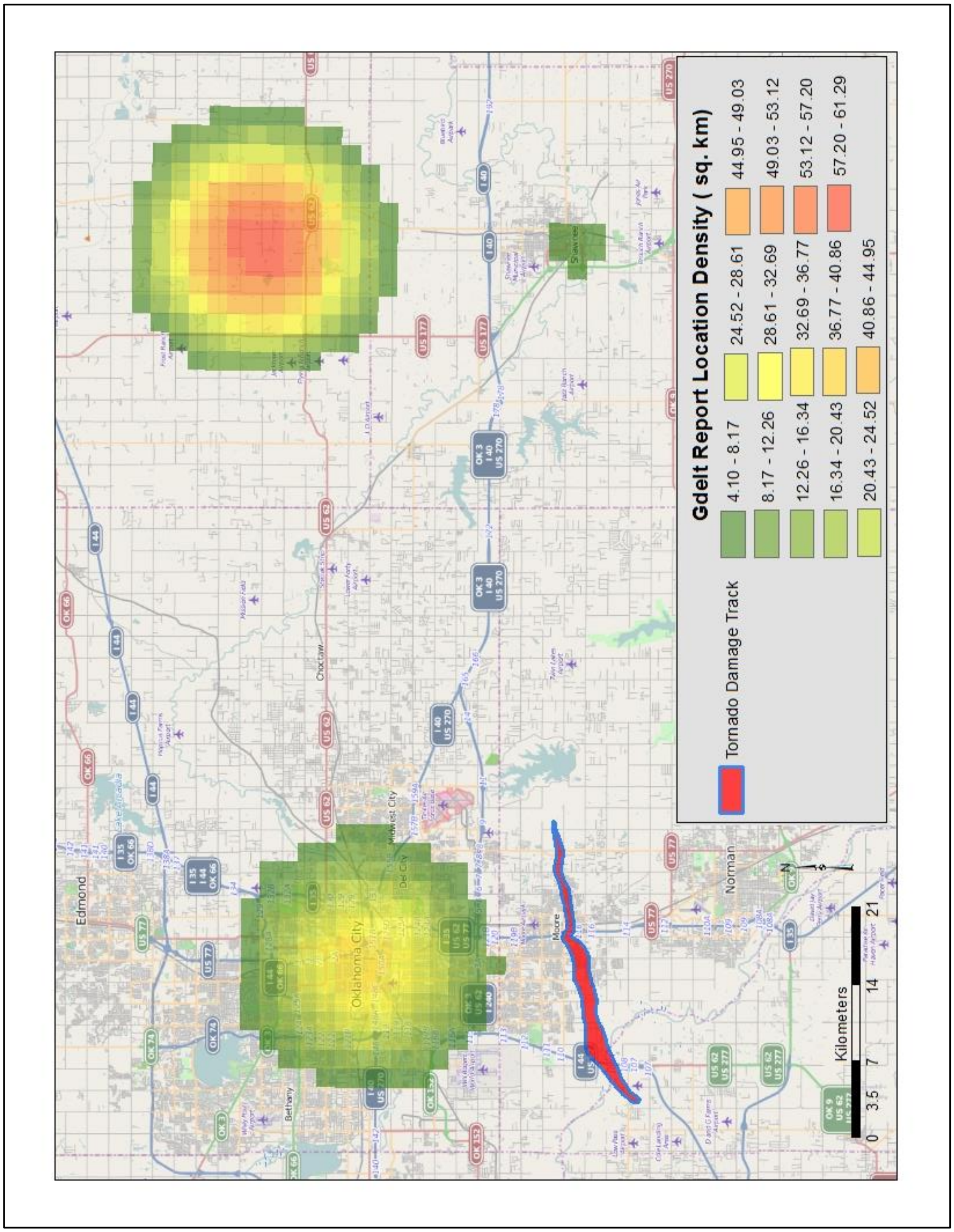


Figure 6: GDELT Report Locations Referencing Moore Tornado but geo-located outside of Tornadoes Path

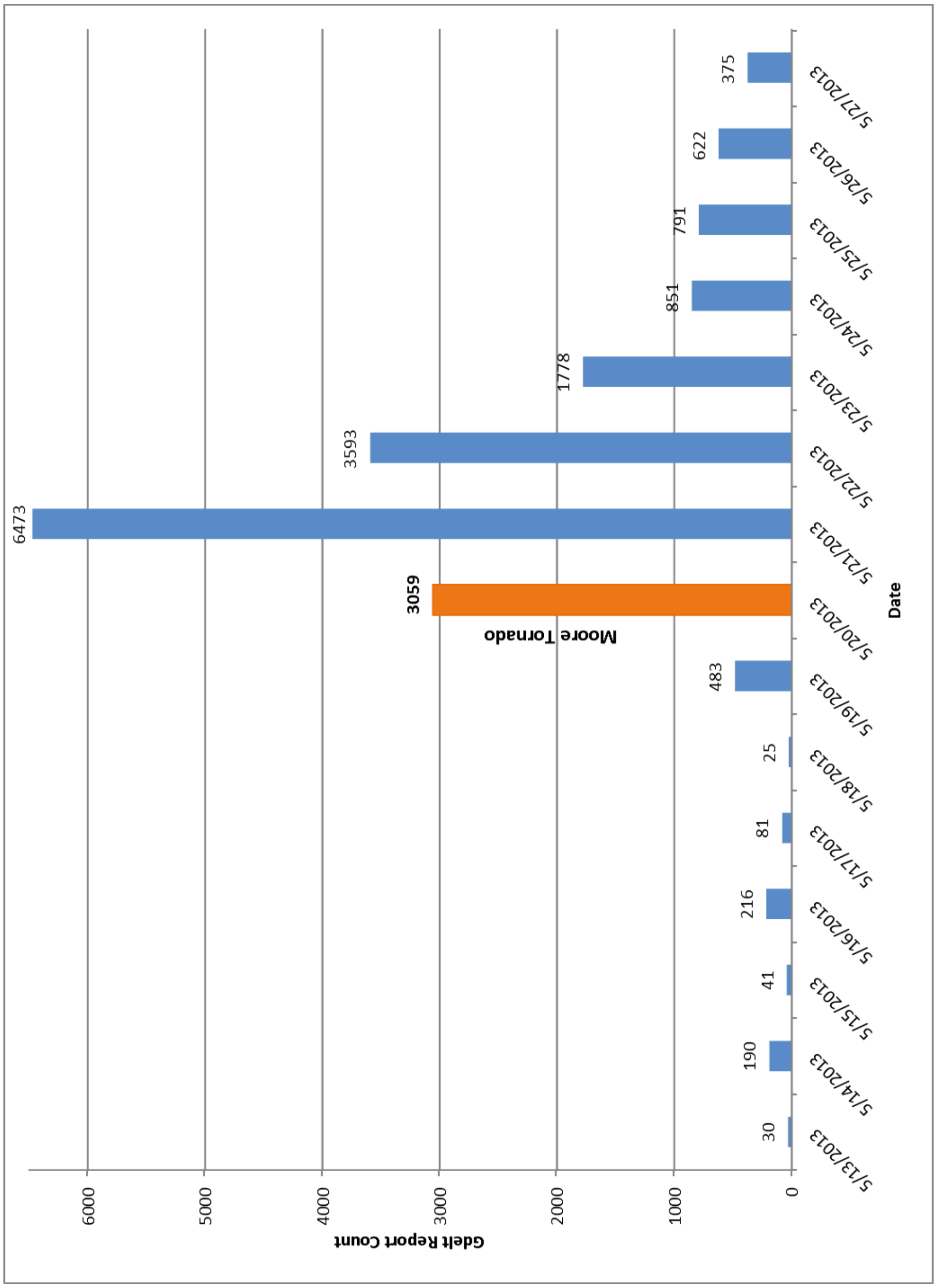


Figure 7: GDELT Report Daily Totals for Moore Tornado in Moore Oklahoma

6.3 Oklahoma Analysis using OpenStreetMap

The OpenStreetMap file downloaded for Oklahoma covered edits from 2006 through 2014. OSM edit file contains three types of edits: point, line, and polygon. Each of these edits was displayed as individual shape files within ArcMap in order to track location and calculate density. Table 1 shows an overview of the edit counts associated with the file. The initial OSM file contained edits that were outside the set study period and therefore only the relevant edits corresponding to the study dates were extracted. Overall only 2.03% of point, 59.04% of polygon, and 054.70% of line edits were within the research timeframe. Of the edits that fit the research period 52.5% of point, 77.3% of polygon, and 7.7% of line edits were located within the Tornado's impact period.

Table 1: OpenStreetMap Edit Count for Moore Tornado

	Total Number of Edits from Original File	Edit(s) Posted within May 2013	Edit(s) Within two week research period	Edit(s) Located within Tornado Track
Point Edit(s)	12,361,162	650,321	13,182	6,923
Polygon Edit(s)	172,037	1,731	1,022	790
Line Edit(s)	535,905	2,485	1,359	105

While the number and type of edits varied, the daily updating patterns of all three types of OSM edits followed a similar pattern. Figure 7 shows a detailed count of all three edit types throughout the two week research period. Overall, none of the three edits types were posted until the day of the Moore Tornado. After May 20, 2013 all three edit categories show an increase in edit counts but point and polygon edits exhibit a peak and valley pattern that occurs with two instances. New point and polygon edits are not submitted after May 24, 2013. Overall the edit type with the most updates is point edits. The highest date of point edits is on May 22nd, approximately two days after the event. In an effort to better understand the amount of OSM edits posted and when they are posted, the same data used to create figure 7 were turned into percentages of total to see what percentage of the total amount of edit and their types were posted each day. Figure 8 shows the percentages of each edit type per day during the research period. The resulting chart shows that approximately two days after the tornado hit Moore OK, is when the highest percentage of point and poly edits are present. Nearly fifty percent of the total amount of edits was posted during this day. However, the highest percentage of line edits; nearly forty-four percent was not present until the last day of the study.

Since OSM edit files have three edit layers associated with the main file, each layers edit's density was mapped separately. This was to locate where each edit type was occurring, and how many edits occurring represented by a density scale. In figure 9, point edit density present within the tornado tract is represented by a density range where the lowest areas are light blue and high areas are represented by red. The majority of point edits are located around the central portion of the damage track and the areas of highest density are located in portions where the Tornado was at a lesser EF rating. This shows that the areas where users posted the most point edits are where the tornado was not at its highest wind speed.

The line density edits were more numerous within the central portion of the tornadoes damage track. In the below map, Figure 9, areas of high intensity are represented by blue and areas of low density is represented by brown. The areas with the highest intensities are located in the southwest quadrant of the map where the tornado first touched down. The density shows that line edits were posted in portions of the tornadoes damage track that included areas where the tornado had low and high wind speeds.

Figure 10 shows the density levels of polygon edits with a range from low to high represented by blue to brown respectively. The majority of the polygon densities are located where the tornados wind intensity was lower but it is these areas where higher levels of building infrastructure were located. Residential, businesses and schools were located where the areas on the map are showing the highest polygon density.

One research question stemming from the main hypothesis explored within this thesis focuses on the scope or categories of the point, line, and polygon edits. OSM data producers can assign a category to their edits if they choose. Assigning a category to an edit is not required and as a result not all edits have a category designation or are uncategorized. For the natural disaster event of the Moore Tornado, the point, line, and polygon OSM layers were tallied in terms of how many edits were present and on what date but also how many of the edits had a category designation and what those designations were. Table 2 shows the count for each of the edit layers and how many of those edits are in each category. While the optional categories cover over nineteen different types, the assigned categories appear to be focused on three main designations. For the point layer the category with the highest count was highway, however only 0.3% of point edits were categorized. Within the line layer, the category with the highest count was the highway designation with over 99% of edits being categorized. The category with the highest count within the polygon layer was the building category designation with 98% of edits being categorized.

Overall, concerning OSM edits and the Moore Tornado, edits are primarily focused on three distinct locations of the Moore Tornado tract. OSM edits are located at the beginning and end of the tornado track and the central portion of the track. While there is an opportunity for contributors to categorize their edits, most edits are not categorized. Edits are also posted primarily after the event has occurred with high number two days after the tornado has occurred.

	5/13/2013	5/14/2013	5/15/2013	5/16/2013	5/17/2013	5/18/2013	5/19/2013	5/20/2013	5/21/2013	5/22/2013	5/23/2013	5/24/2013	5/25/2013	5/26/2013	5/27/2013
Line Edit	0	0	0	0	0	0	0	0	2	29	17	11	0	0	46
Point Edit	0	0	0	0	0	0	0	141	1894	3006	177	1705	0	0	0
Polygon Edit	0	0	0	0	0	0	0	7	229	366	0	188	0	0	0

Figure 8: OpenStreetMap Daily Edit Count located within Moore Tornado Track

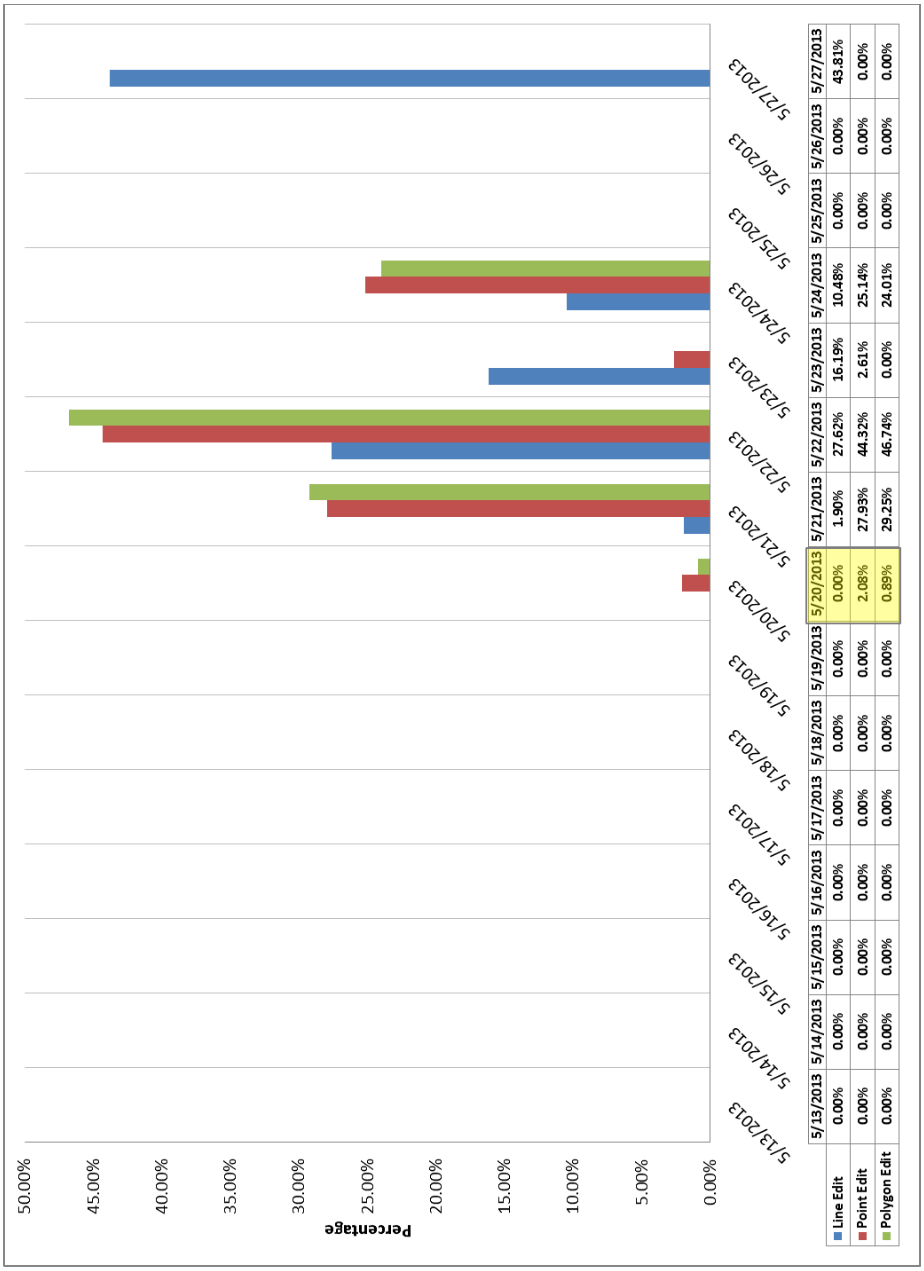


Figure 9: OpenStreetMap Daily Edit Counts as Percentages located within Moore Tornado Track

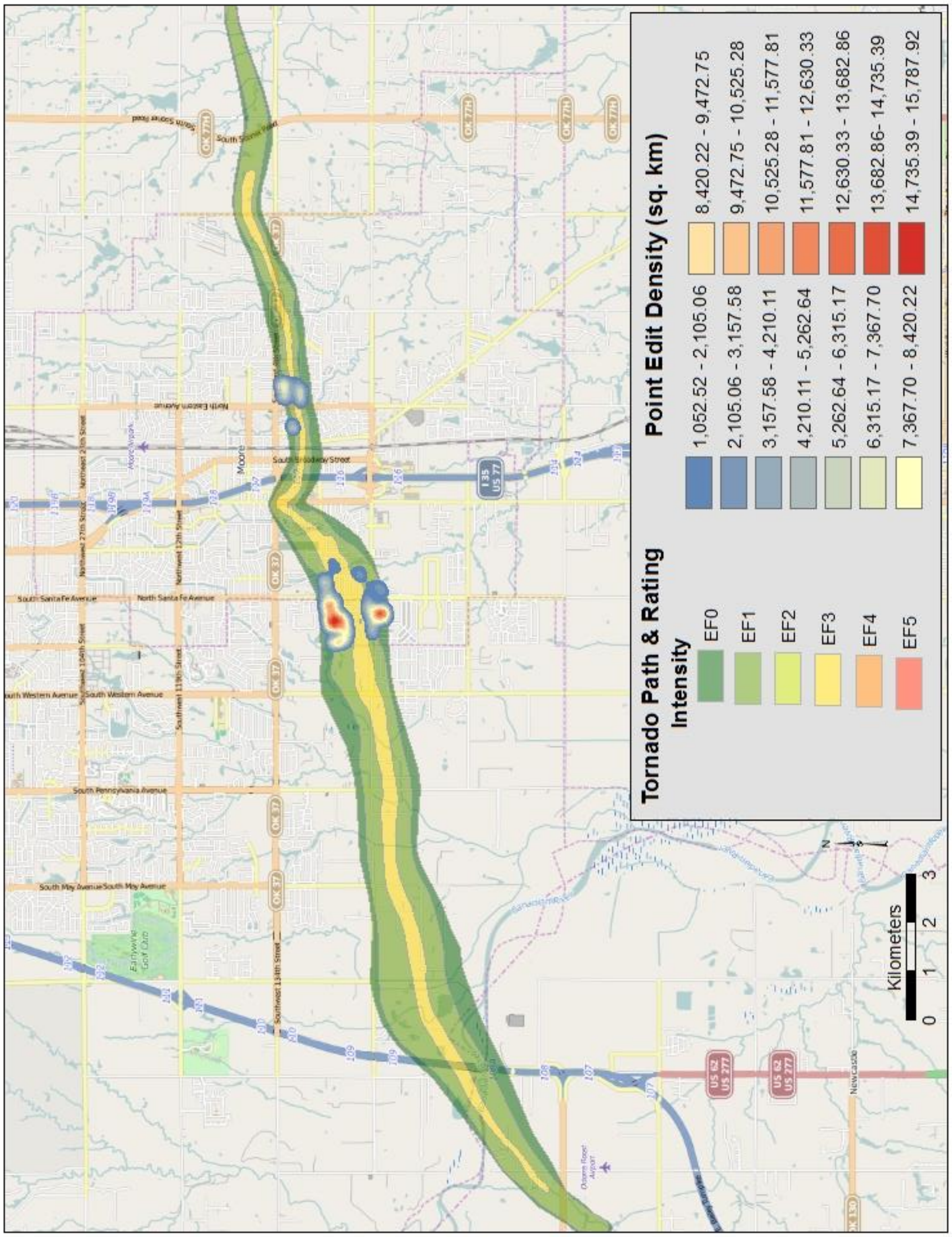


Figure 10: OpenStreetMap Point Edit Density within Damage Track of Moore Tornado

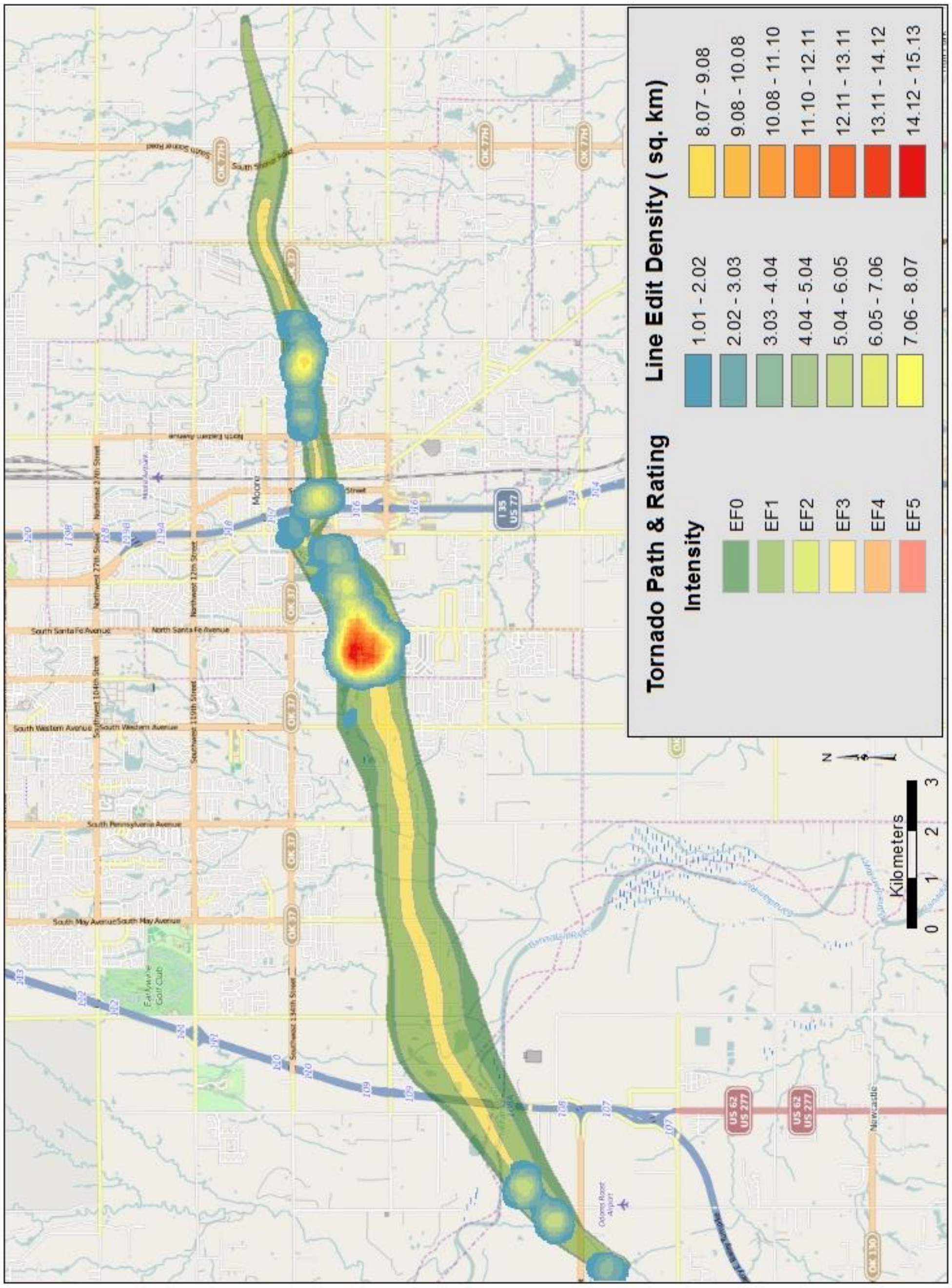


Figure 11: OpenStreetMap Line Edit Density within Damage Track of Moore Tornado

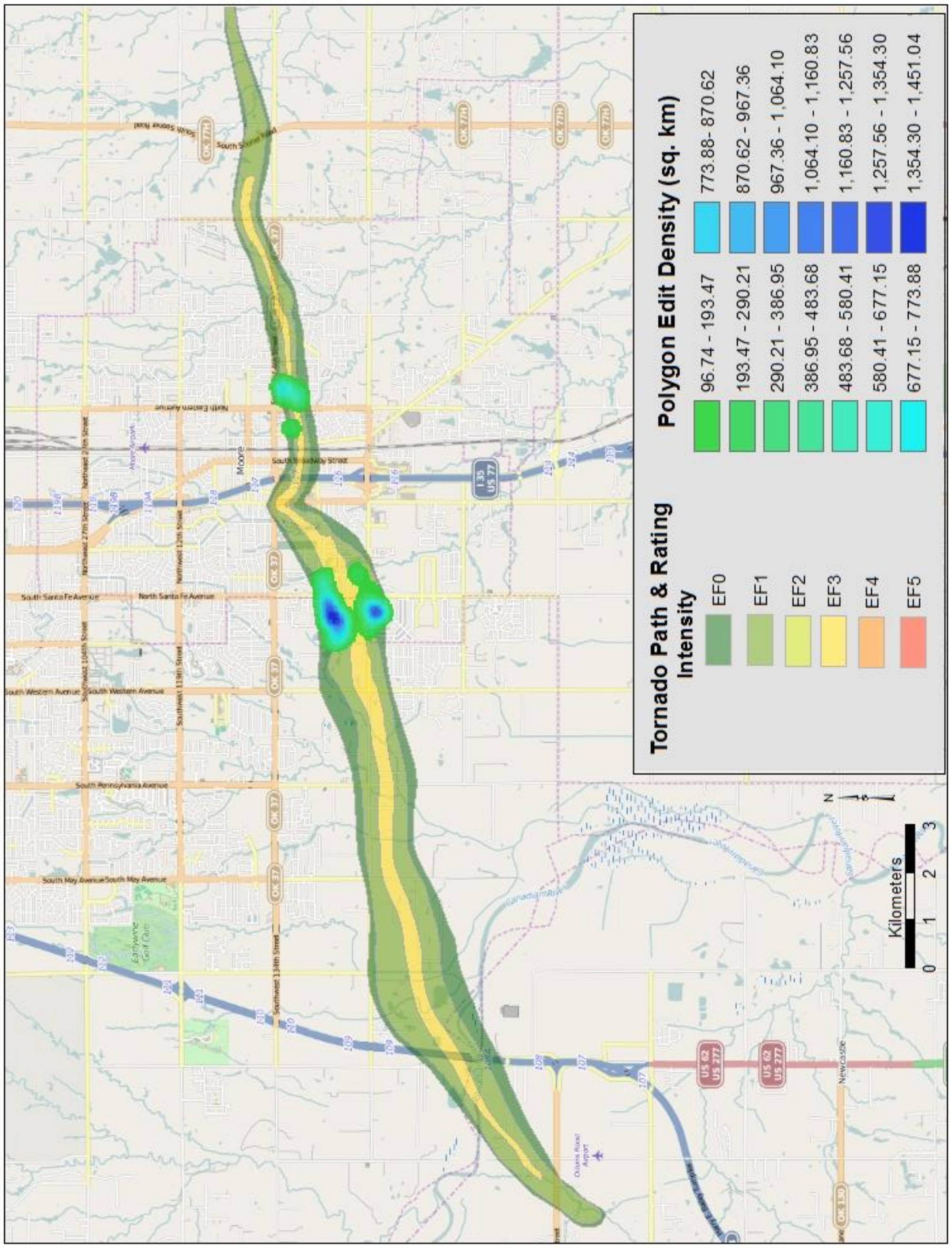


Figure 12: OpenStreetMap Polygon Edit Density within Damage Track of Moore Tornado

Table 2: OpenStreetMap Category Count for Moore Tornado

OSM Edit Category:	Number of Categorized Edits within Point Layer:	Number of Categorized Edits within Line Layer:	Number of Categorized Edits within Polygon Layer:
Aerial Way	0	0	0
Aero Way	0	0	0
Amenity	0	0	2
Barrier	1	0	0
Boundary	0	0	0
Building	0	0	738
Highway	17	103	0
Historic	0	0	0
Land Use	0	0	0
Leisure	0	0	43
Man Made	0	0	0
Natural	0	0	3
Place	0	0	0
Power:	0	0	0
Railway	0	0	0
Route	0	0	0
Shop	0	0	0
Tourism	0	0	0
Waterway	0	1	0
Uncategorized:	6,905	1	9
Total:	6,923	105	790

6.4 Typhoon Haiyan using GDELT

The prime GDELT file had over 300,000 user updates geo-located within the Philippines. Of the original updates 57% had theme tags related to Typhoon Haiyan and were posted during the two-week research period. Each GDELT update had an associated theme tag that allowed it to be associated with Typhoon Haiyan (i.e. high winds), or be discarded (i.e. snow storm). Figure 12 shows the twenty-eight theme categories that were considered associated with Typhoon Haiyan. Themes that were deemed not related to typhoon Haiyan, for example snow storm, and were removed from the master file.

The research period for Typhoon Haiyan was from November 1-15, 2013. Figure 13 shows the GDELT report count for each day. At the beginning of the research period, GDELT reports were lower in numbers but as time progressed closer to when Typhoon Haiyan made landfall, edit number increased at a high rate. After the Typhoon made initial landfall, GDELT user updates continue to be higher than before the event but continued to decrease as time progressed.

GDELT User Updates were recorded and were visually presented as points when displayed with the ArcMap interface. In order to see the majority of where updates were located, point density layers were created from the point data. Figure 14, shows Typhoon Haiyan's path and storm width as it progressed across the Philippines. Areas of high density are shown in red and areas of low densities are shown in as green. The resulting density layer shows five large locations with two locations showing high density patterns. The first location with a high density is located in the south-east portion of the track where the typhoon track first makes landfall. The other high density area is located on the capital city of Manila. Manila is located at the north-west location of Typhoon Haiyan tract. From looking at the density patterns of the GDELT reports the majority of reports are located where Typhoon Haiyan first made landfall.

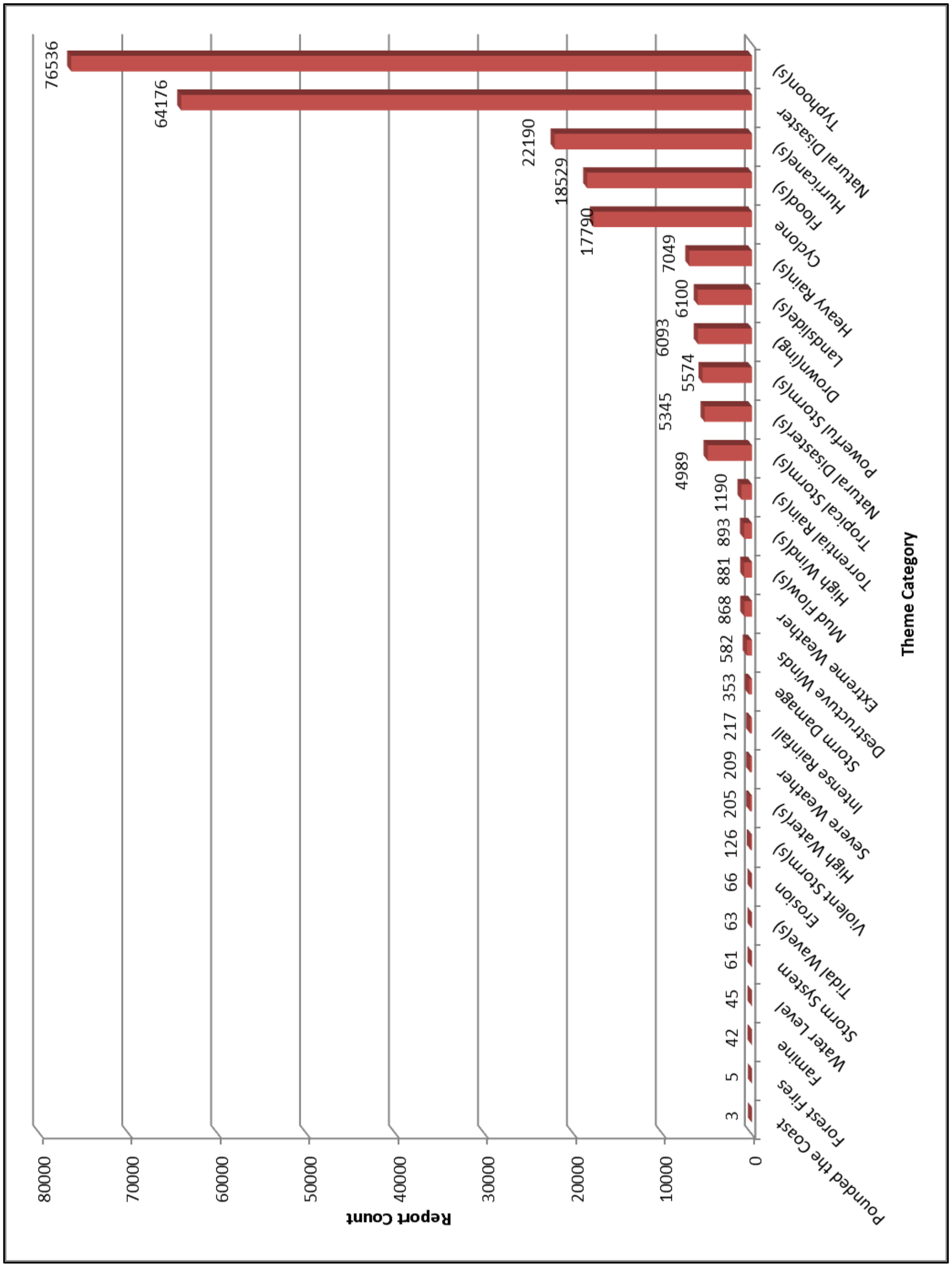


Figure 13: GDELT Report Totals per Theme Category for Typhoon Haiyan

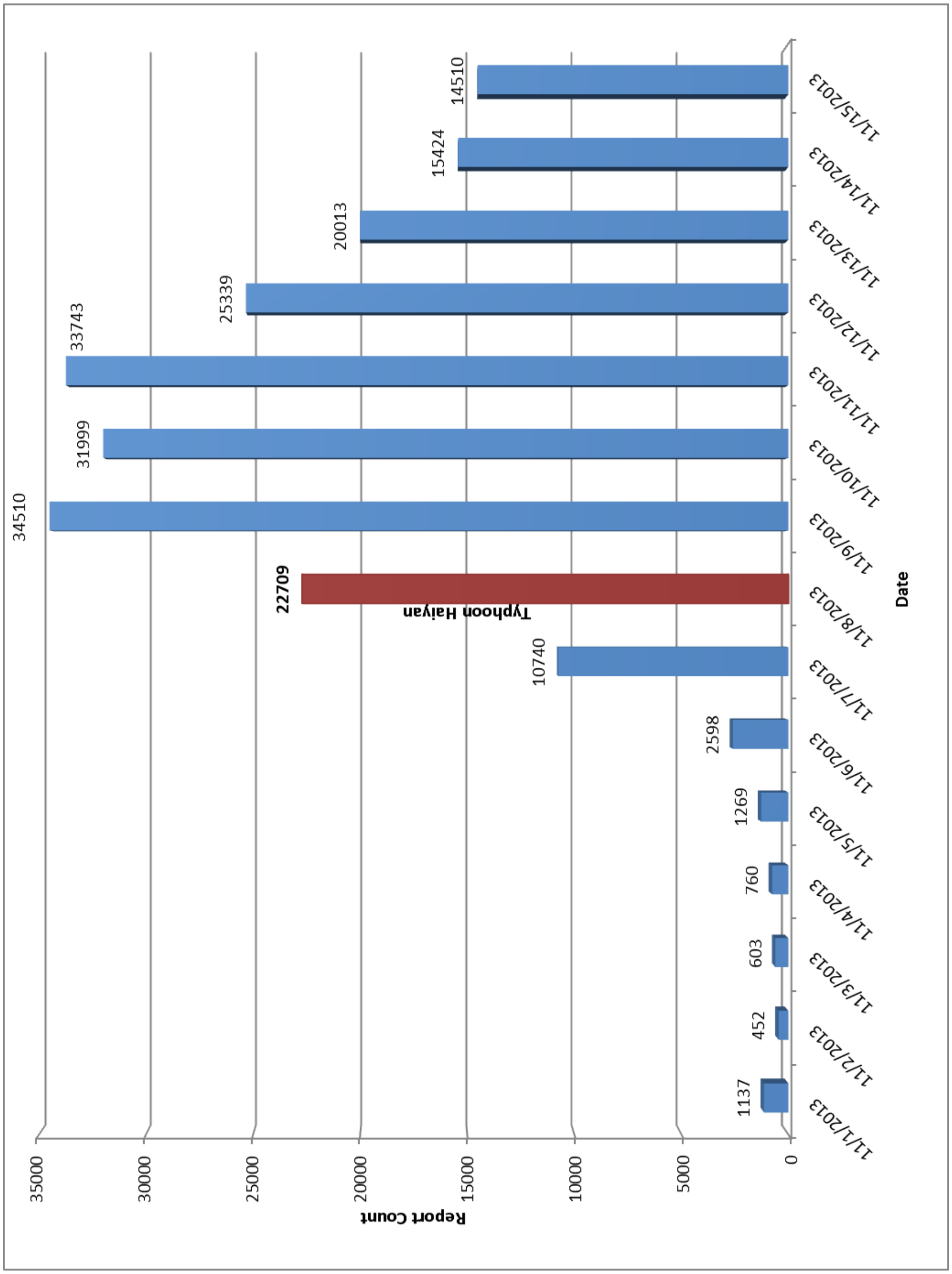


Figure 14: Daily GDELT Report Totals for Typhoon Haiyan in the Philippines

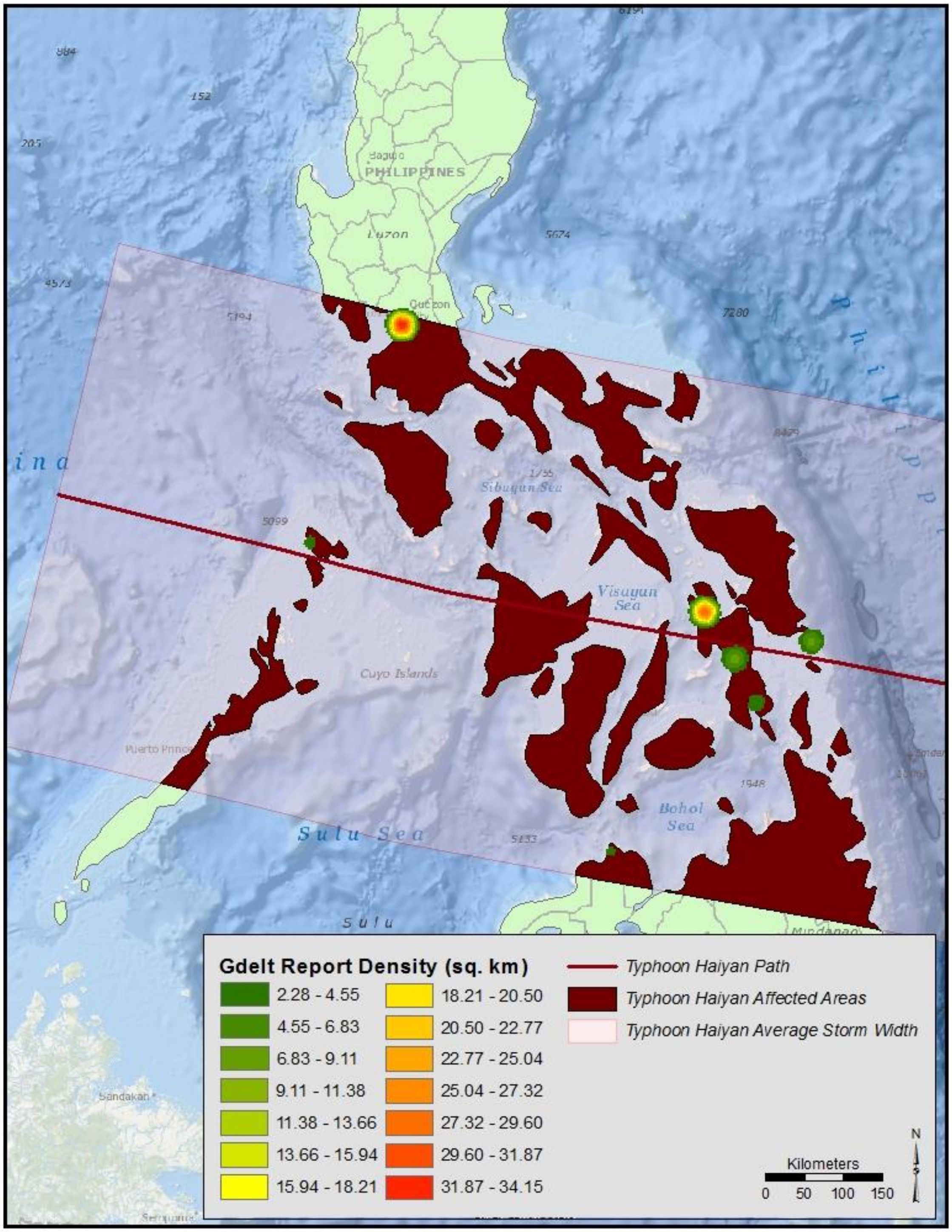


Figure 15: GDELT Report Density for Typhoon Haiyan

6.5 Typhoon Haiyan using OpenStreetMap

The OpenStreetMap file downloaded for the Philippines covered edits from August 29, 2006 to December 12, 2014. The primary OSM edit file contains three types of feature edits, including point, line, and polygon. Each of these feature edits was displayed as individual shape files within ArcMap in order to track location and calculate density. Table 3 shows an overview of the edit counts associated with the file. The initial OSM file contained edits that were outside the set study period and therefore only the relevant edits corresponding to the study dates were extracted. For the Philippines, only 49.51% of point, 45.32% of polygon, and 46.35% of line edits were posted within the month of November. Of the edits that were located within November, 70.84% of point, 98.65% of polygon, and 85.60% of line edits were located within the Typhoon Haiyan's impact timeframe.

Table 3: OpenStreetMap Edit File Totals

	Total Number from File	Edit(s) Posted within November 2013	Edit(s) Within Research Period Timeframe	Edit(s) within Typhoon Haiyan Path
Point Edit(s)	12,990,970	3,522,793	1,744,193	1,235,548
Polygon Edit(s)	1,340,353	513,040	232,486	229,341
Line Edit(s)	474,858	50,933	23,607	20,219

In the case of Typhoon Haiyan and OSM feature edits, the daily total of each edit type varied on a daily basis. The daily totals were calculated by querying the master OSM file to identify edits only occurring within a twenty-four hour period. As time progressed towards the date of the typhoons landfall, the daily updating patterns of all three types of

OSM increased. Figure 15 shows a detailed count of all three edit types throughout the research period. The amount of point edits demonstrates the largest increase in edits posted compared to polygon and line edits. It is interesting to note that line edit numbers increase after the typhoon landfall and continue to decline like the other two edit types but on a lesser scale. While edit counts is important for recognizing patterns, it was also beneficial to what percentage of edits are being posted each day. In order to observe this result, figure (16) was created in order to see what percentage of edits was posted each day. Typhoon Haiyan hit the coast of the Philippines on November 8th, however line edits starting being posted before the typhoon hit. For the two week study period, the majority of line, point and polygon edits occurred approximately five days after the typhoon first impacted the coast of the Philippines.

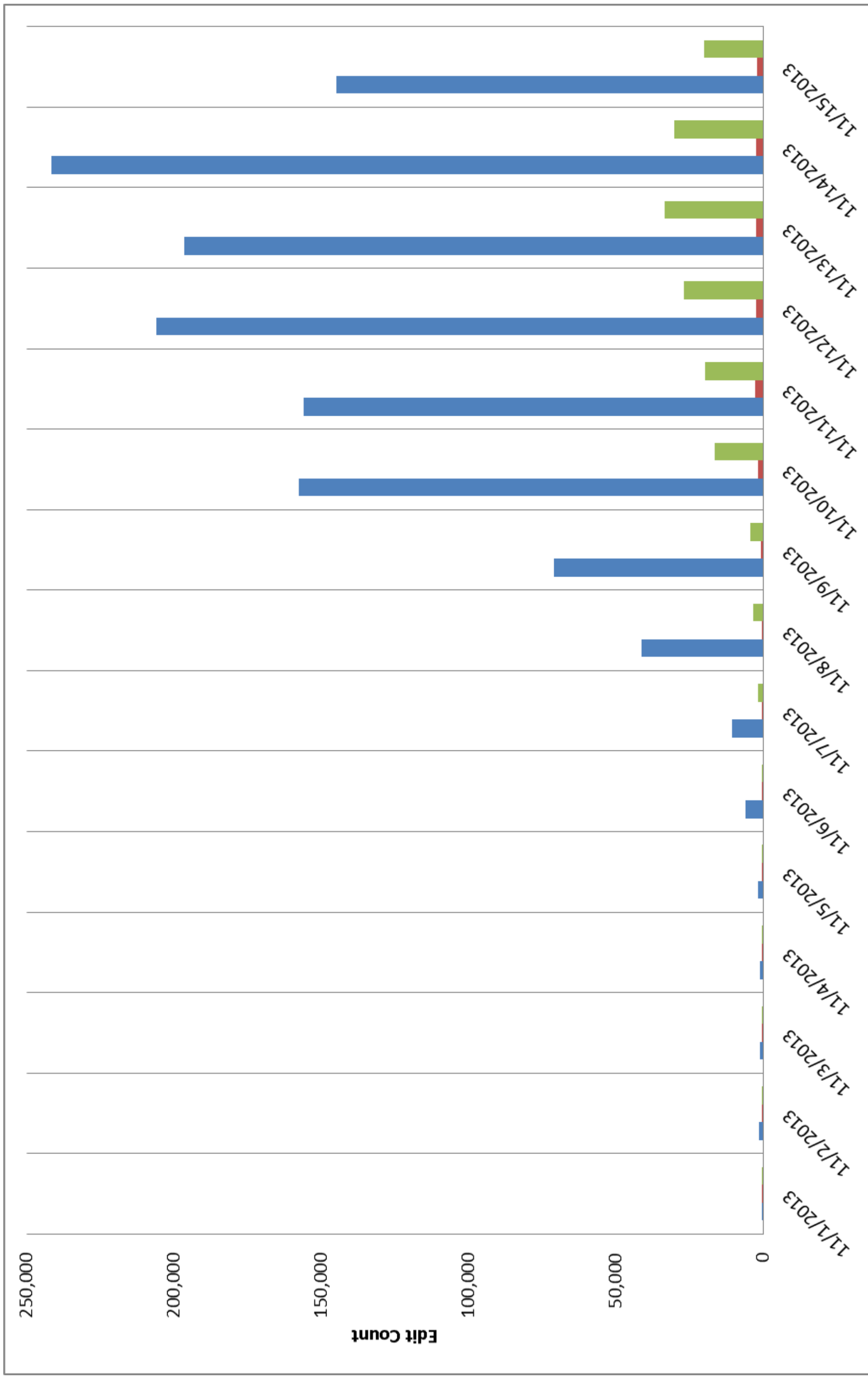
The following three density maps were created from the number and locations of edits posted of the polygon, line, and point layers located within Typhoon Haiyan's track. In these maps the blue line represents the path of the typhoon and the red overlay simulates the width of the storm. The parts of the Philippines that are colored red are areas that were directly affected by the storm.

The first density map presented, Figure 17, shows the density of polygon edits calculated within the storms track. The most significant or densest areas are located near or right on the path of Typhoon Yolanda. Figure 18 displays the density results of the line edits that were posted within the study period. Like the previous figure, areas of lower density are represented in green and areas of high density are within red. Line edits unlike the polygon edits are more spread out across the impact area affected by the storm. The densest areas are directly along the typhoons track or located near a major city. The majority of edit are along island coastlines however, line edits are present within the central portion of the larger islands.

Of the three OSM feature edit types, the edit with the largest response or presence was the point edits. Figure 19 shows the density layer results of point edits within Typhoon Haiyan's track. As previously seen with the density results of line and polygon edit densities, the majority of density edits are located along the typhoons track. Unlike the density layer of line edits, the location of the point edits are more centrally clustered around where the typhoon first made landfall.

One facet of the main hypothesis explored within this thesis focuses on the scope or categorized of edits. OSM has three types of edits and data producers can assign a category if they choose. Assigning a category to an edit is not required and as a result not all edits have a category designation. For Typhoon Haiyan, the point, line, and polygon OSM layers were not only tallied in terms of how many edits were present and on what date but also how many of the edits had a category designation. Table 4 shows the count for each of the edit layers and how many of those edits are in each category. While the optional categories cover over nineteen different types, the highest category for each edit layer was based on what a user could view as what that type of edit was representing. For example, the category with the highest count in the point layer was place. Point edit usually indicate specified locations and are represented by a single point. Within the line layer, the category with the highest count was highway. The category with the highest count within the polygon layer was the building category with over 94% of the total votes.

The creation of maps that show the location of the three types of OSM edits is important for locating where the edits are occurring but in order to fully answer the research question of where are edits occurring and the scope of the edits it is necessary to understand what is going on with the population of where the edits are occurring, Therefore the figures of 20, 21, and 22 show the results of line, polygon, and point edits overlaid over population density of the Philippines. The edit density layers are the same as what is presented in the previous maps but now edits can be compared as to where they are occurring in relation to population density. Figure 20 shows the line edit density and as a result the majority of edits are occurring in two places as before but the highest line edit density is occurring in population density with the highest concentration. This location is also the capital of the Philippines, Manila. The second density location is occurring in a lower population area of between 178.30 to 252.50 people per square kilometer. While this is not a higher populated area, it is near where Typhoon Haiyan first made landfall and falls along the line representing the eye of Typhoon Haiyan. Figure 21 shows the point edit density. From comparing the population density and the location of edits, the highest density of edits is occurring in areas first impacted by the storm. However, the majority of edits is occurring in areas where the population density is between 178.30 and 387.20 people per square kilometer and are located along the path of the eye of the storm. The next presented map, Figure 22 covers the location of polygon edit density compared to population density. Similar to the point edit density map, the location of polygon edit density compared to population is relatively the same. The location of polygon edits is mostly concentrated along medium to lower population density by concentrated along the eye of Typhoon Haiyan.



	11/1/2013	11/2/2013	11/3/2013	11/4/2013	11/5/2013	11/6/2013	11/7/2013	11/8/2013	11/9/2013	11/10/2013	11/11/2013	11/12/2013	11/13/2013	11/14/2013	11/15/2013
Point	97	1,340	1,032	1,044	1,860	5,818	10,567	41,195	70,913	157,409	155,776	205,822	196,308	241,568	144,799
Line	9	304	75	320	71	107	107	316	581	1,648	2,853	2,380	2,375	2,326	2,180
Polygon	9	181	346	35	49	481	1,643	3,433	4,298	16,383	19,708	26,796	33,354	30,122	20,043

Figure 16: OpenStreetMap Edit Daily Totals for Typhoon Haiyan

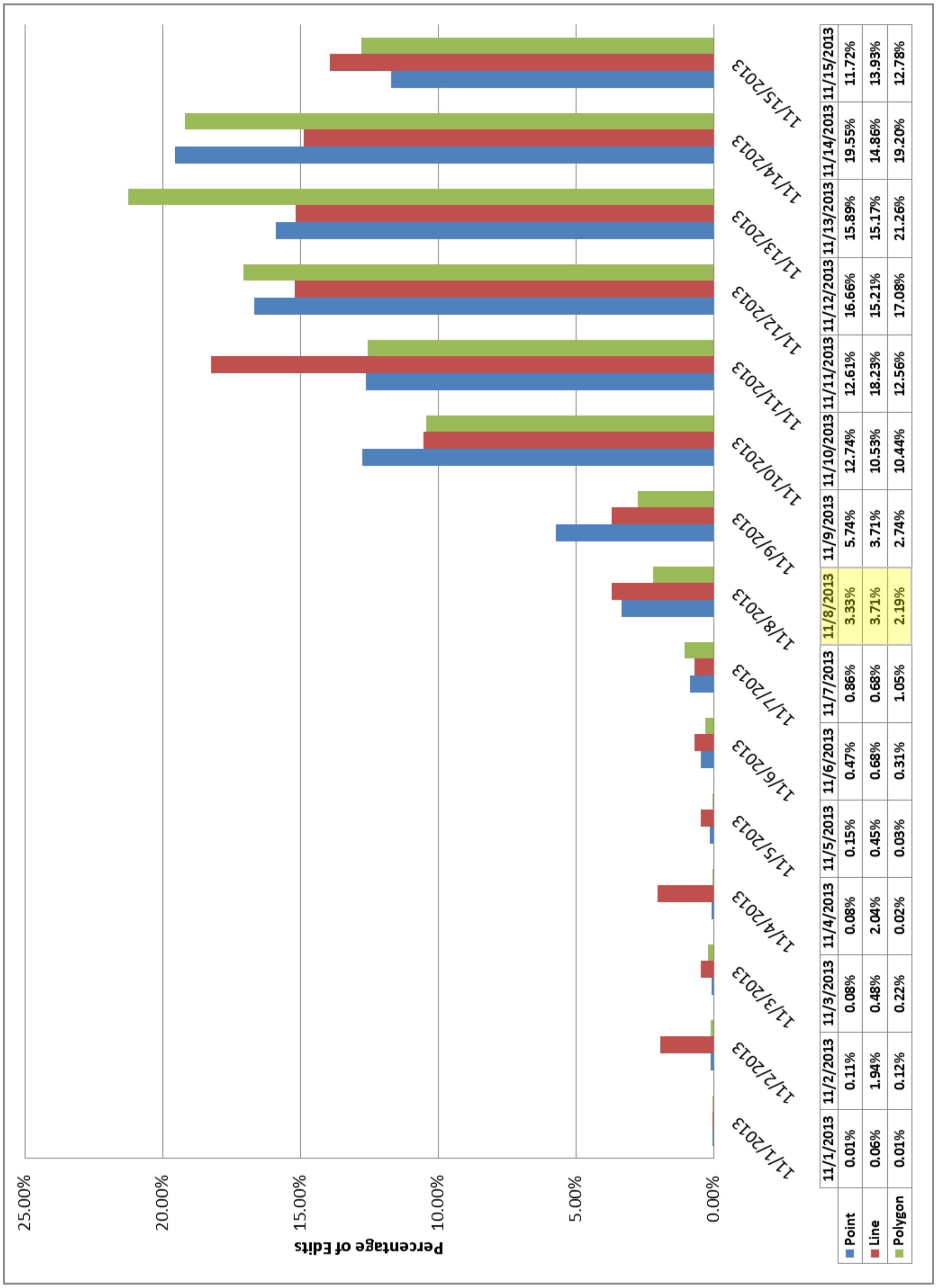


Figure 17: OpenStreetMap Daily Edit Totals as Percentages for Typhoon Haiyan

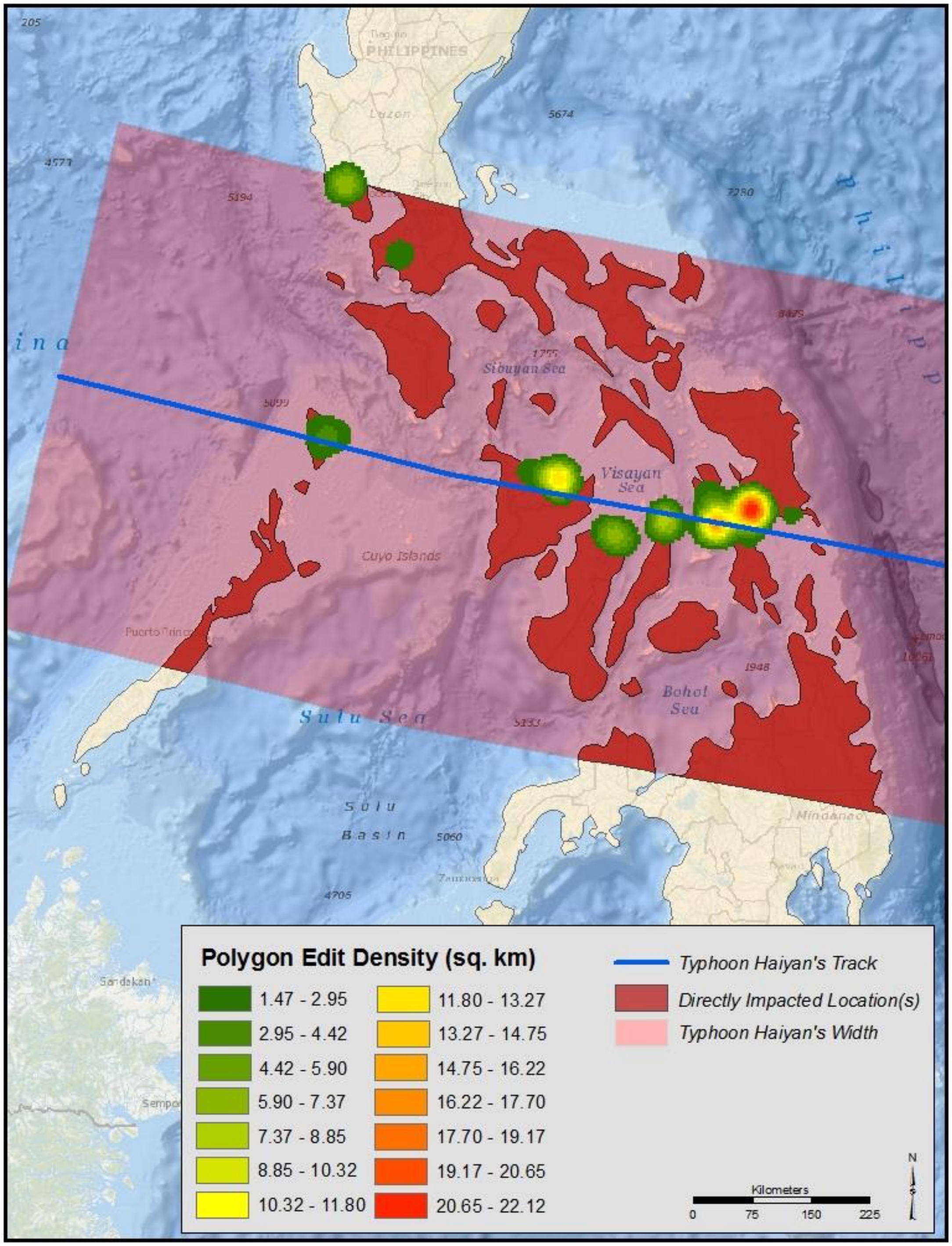


Figure 18: OpenStreetMap Polygon Edit Density for Typhoon Haiyan

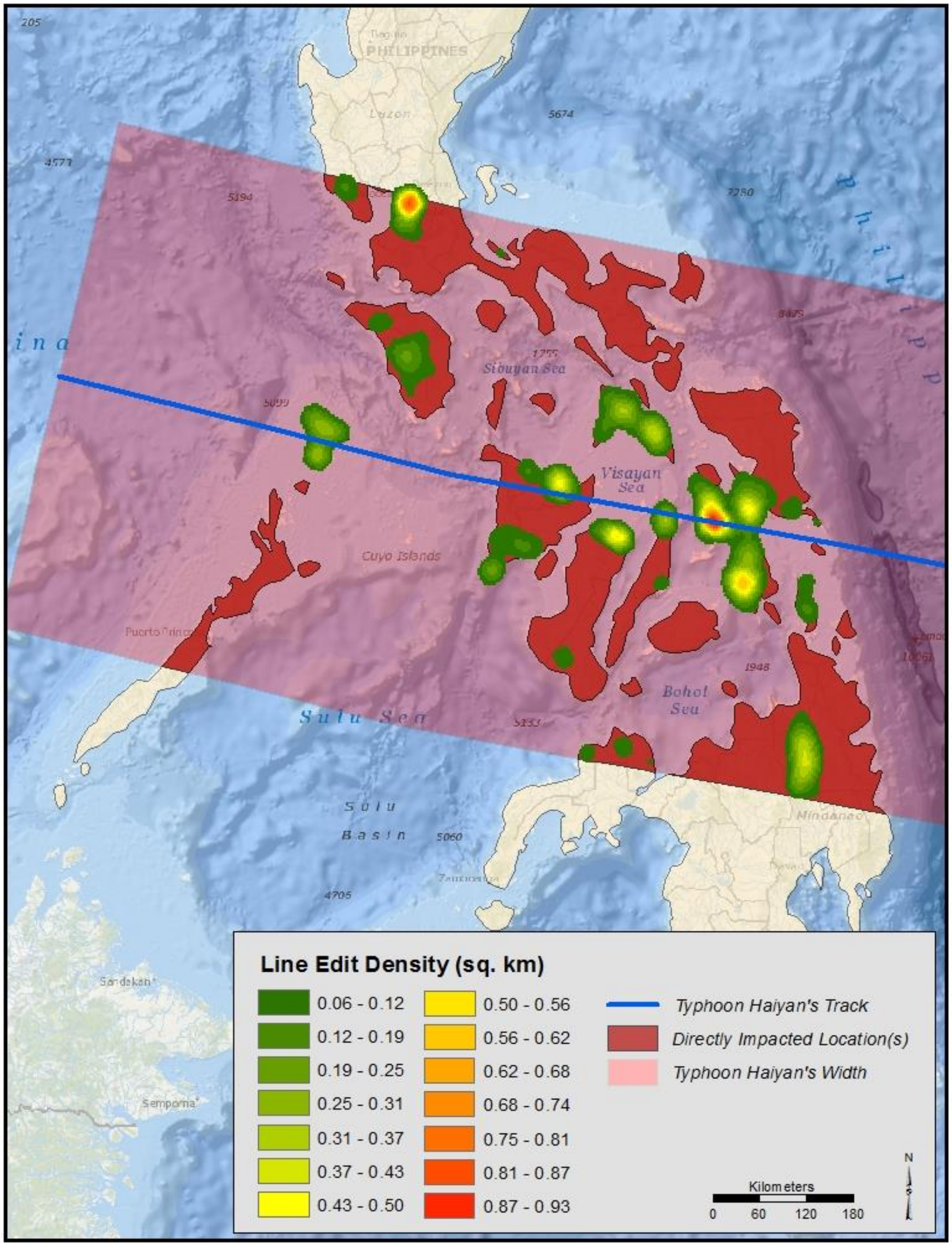


Figure 19: OpenStreetMap Line Edit Density for Typhoon Haiyan

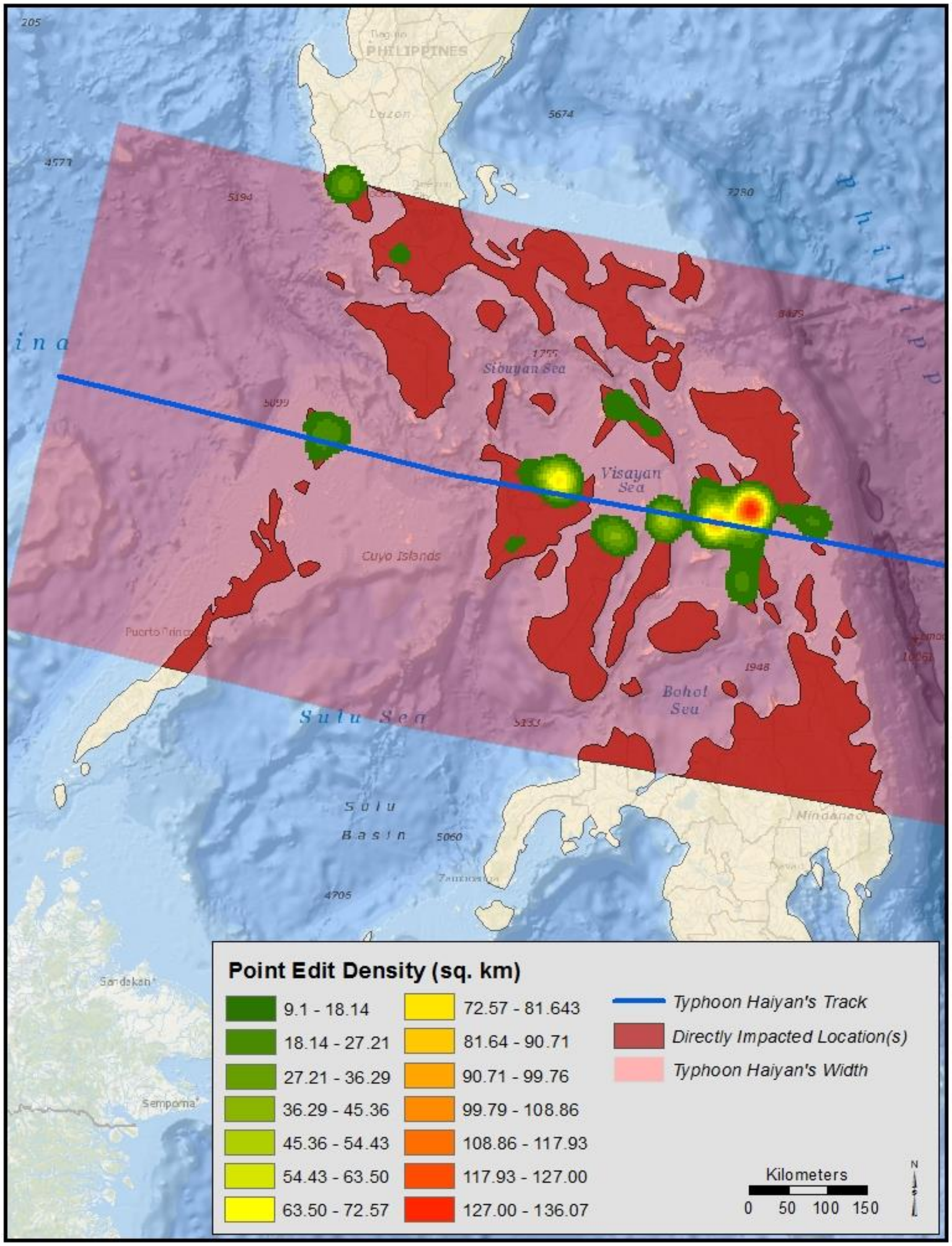


Figure 20: OpenStreetMap Point Edit Density for Typhoon Haiyan

Table 4: OpenStreetMap Edit Category Count

OSM Edit Category:	Number of Catorgorized Edits within Point Layer:	Number of Catogroized Edits within Line Layer:	Number of Catorgorized Edits within Polygon Layer:
Aerialway	0	1	0
Aeroway	4	6	3
Amenity	230	0	242
Barrier	86	58	16
Boundary	2	256	1
Building	481	5	149,811
Highway	49	13,585	3
Historic	7	0	3
Landuse	45	3	4,956
Leisure	16	0	201
Man Made	47	108	34
Natural	117	64	1,236
Place	4,244	22	53
Power:	811	19	7
Railway	15	5	0
Route	0	6	0
Shop	73	0	16
Tourism	25	0	11
Waterway	3	1,384	102
Uncatogorized:	1,229,293	130	186
Total:	1,235,548	15,652	156,881

6.6 Population Density Comparisons

6.6.1 Population Density Comparisons for Typhoon Haiyan

Although the edit density layers can provide valuable information about location and the intensity of edits with regards to population density, the amount of edits present per population density category is also valuable information. The first step in comparing population density and OSM editing density was to create three maps, each focused on a single OSM editing layer, compared to population density. Figures 20, 21, and 22 show population densities overlaid by OSM line, point, and polygon density layers respectively.

In order to see how many of each edit type is present within each population density category figures 23 and 24 were created. Figure 23 shows the amount of OSM line, polygon, and point edits that fall within each population density layer. Figure 24 shows the same information only presented as percentages. What is unique about both of these two graphs is GDELT Report counts are also included. This allows both the VGI information provided by OSM edits and GDELT reports to be included and compared to population density layers. For Figure 23, the highest concentration of point edits is within the population density of 178.30 to 252.50 people per square kilometer. The majority of polygon and line edits fall within the population density of 252.50 to 387.20 people per square kilometer however, line edits have almost an equal count per density category. Concerning the percentage of edits and GDELT reports posted within each category, line edits have almost equal percentage throughout all population layers but show a slight higher count within the 252.50 to 387.20 population per square kilometer. The highest percentage of polygon edits also appears within the 252.50 to 387.20 population per square kilometer category. OSM point edits have a higher percentage at the 178.30 to 252.50 population per square kilometer category. Although present through all categories of the population density, GDELT reports make a higher percentage appearance in the 252.50 to 387.20 population per square kilometer category.

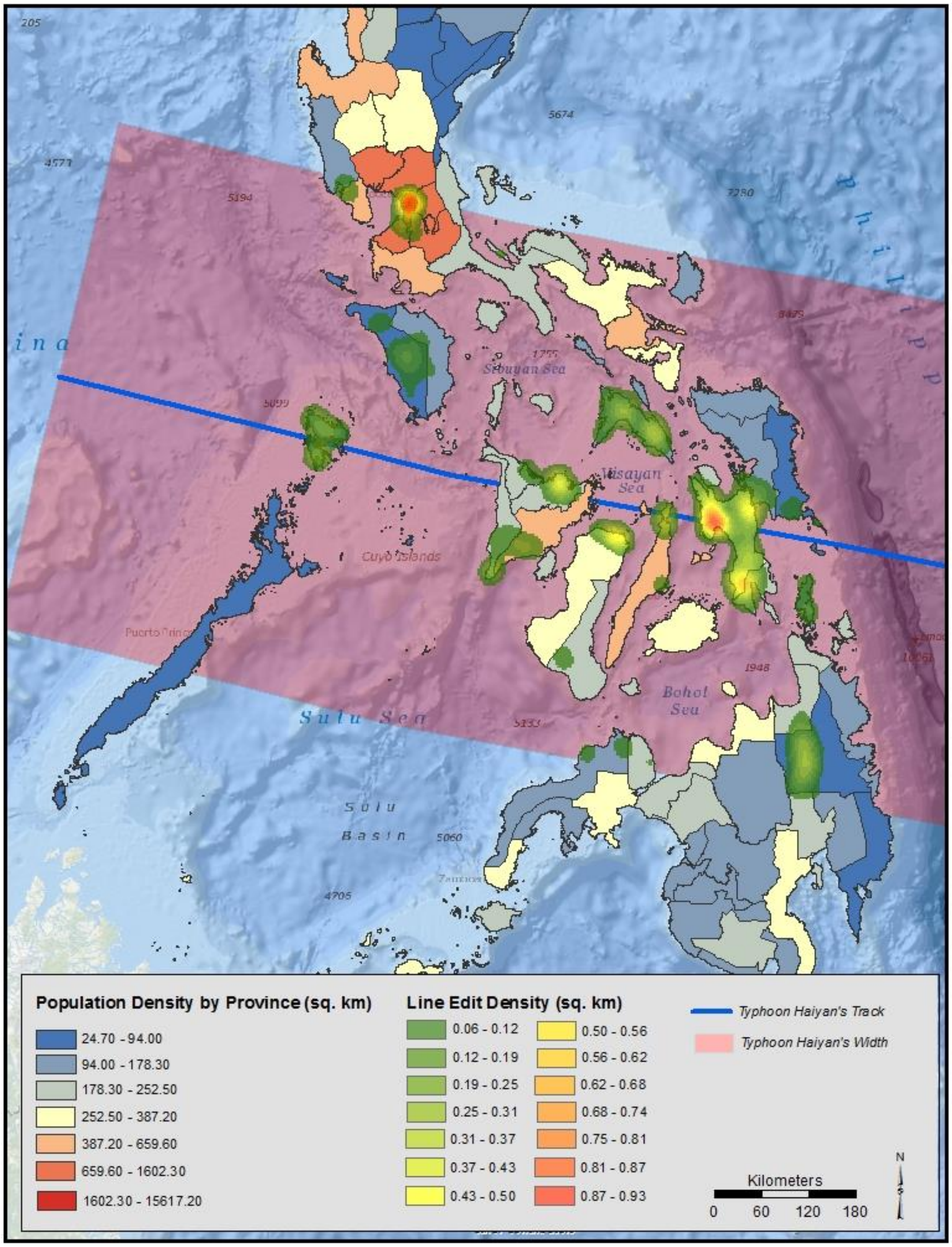


Figure 21: OSM Line Edit Density over laid on Population Density of the Philippines

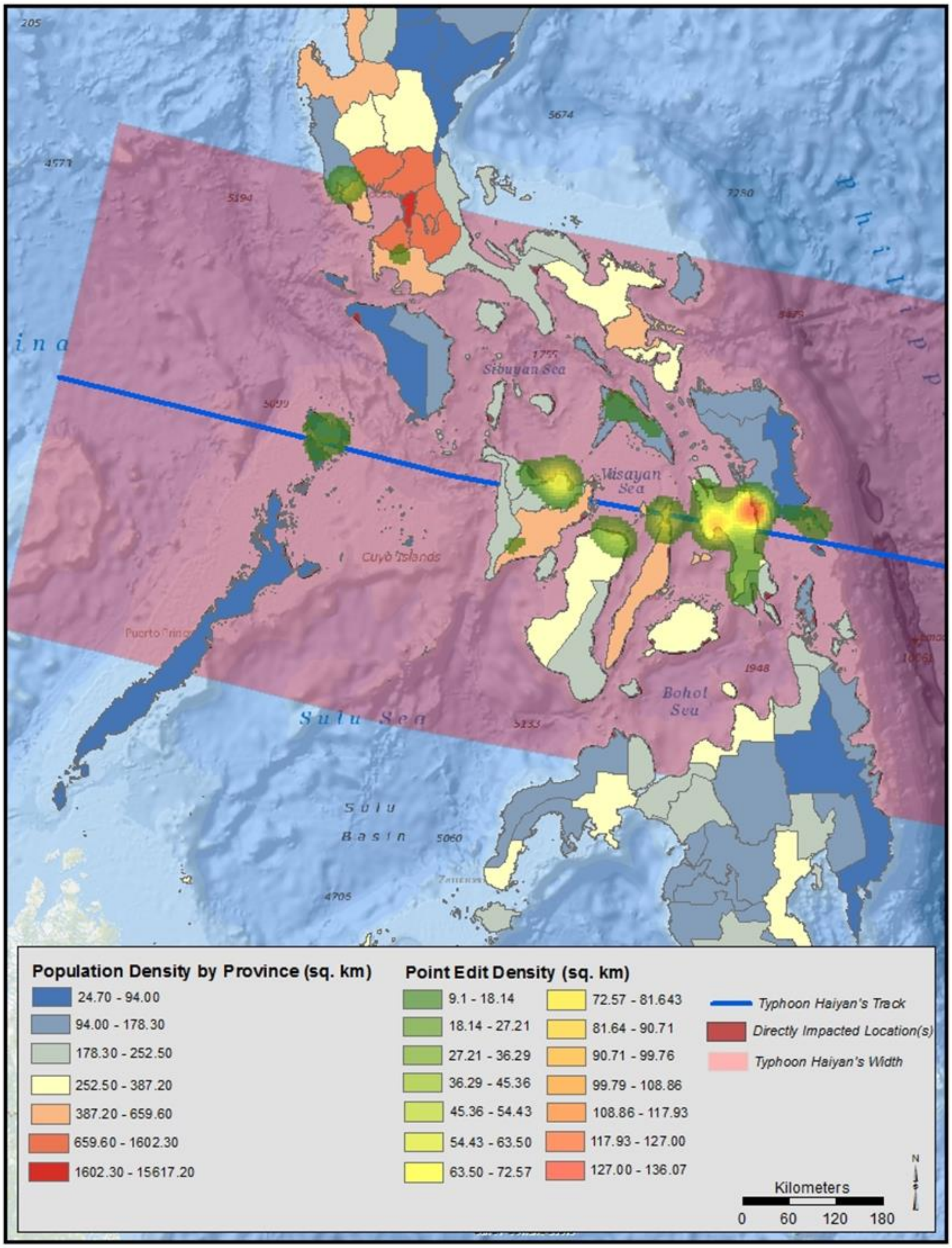


Figure 22: OSM Point Edit Density overlaid over Population Density within the Philippines

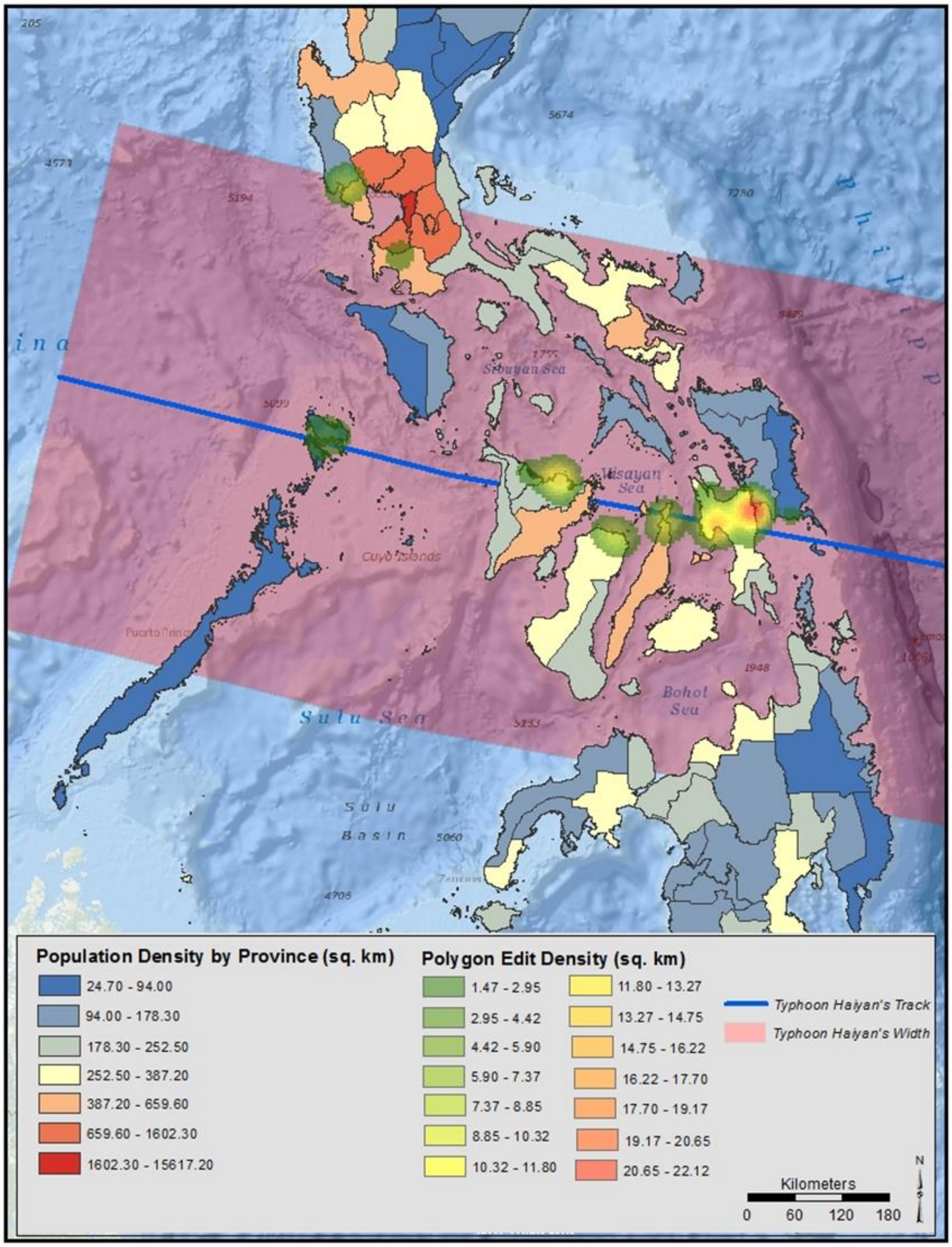


Figure 23: OSM Polygon Edit Density overlaid on Population Density in the Philippines

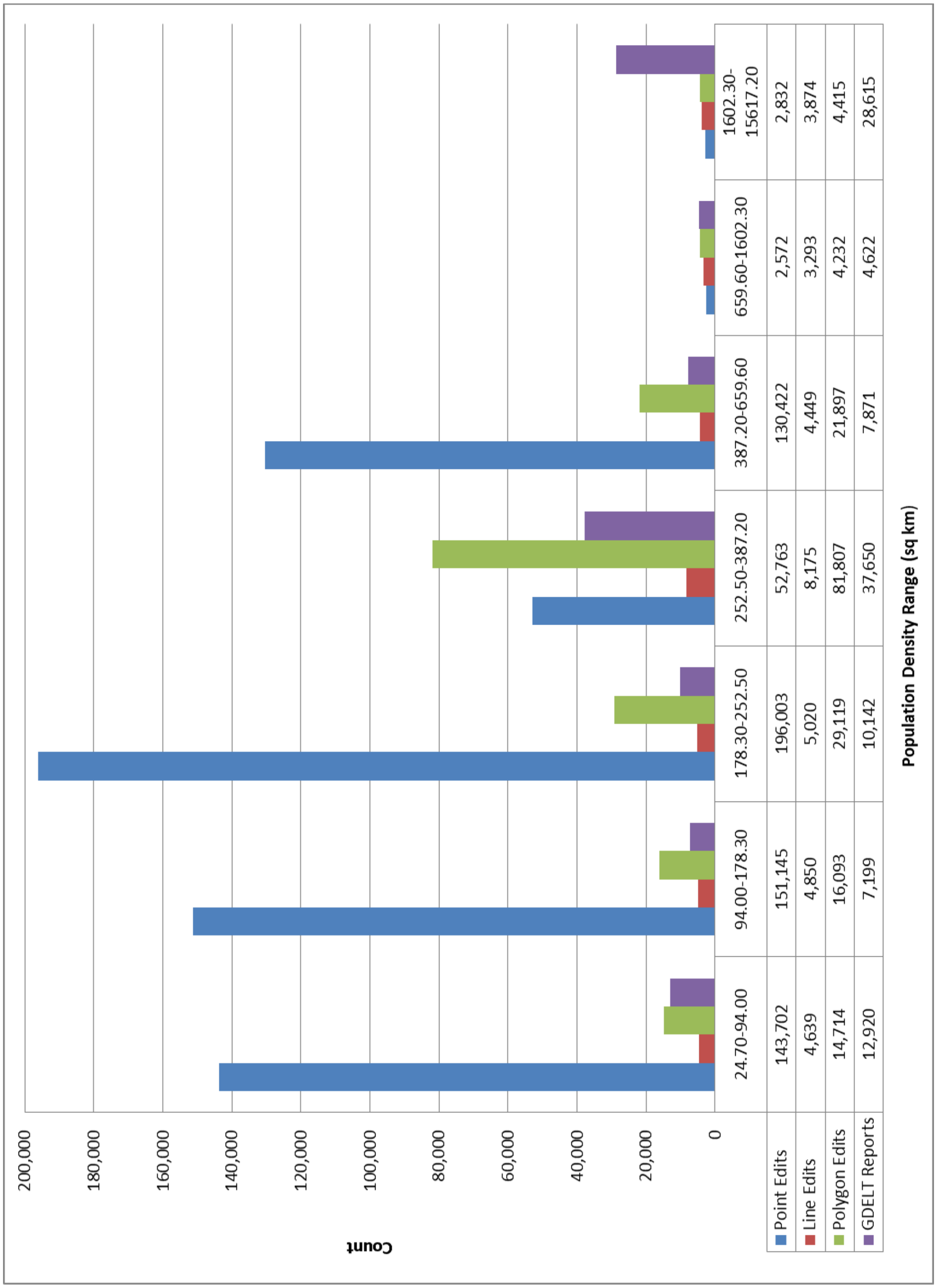


Figure 24: Population Density Comparison in Philippines for Typhoon Haiyan

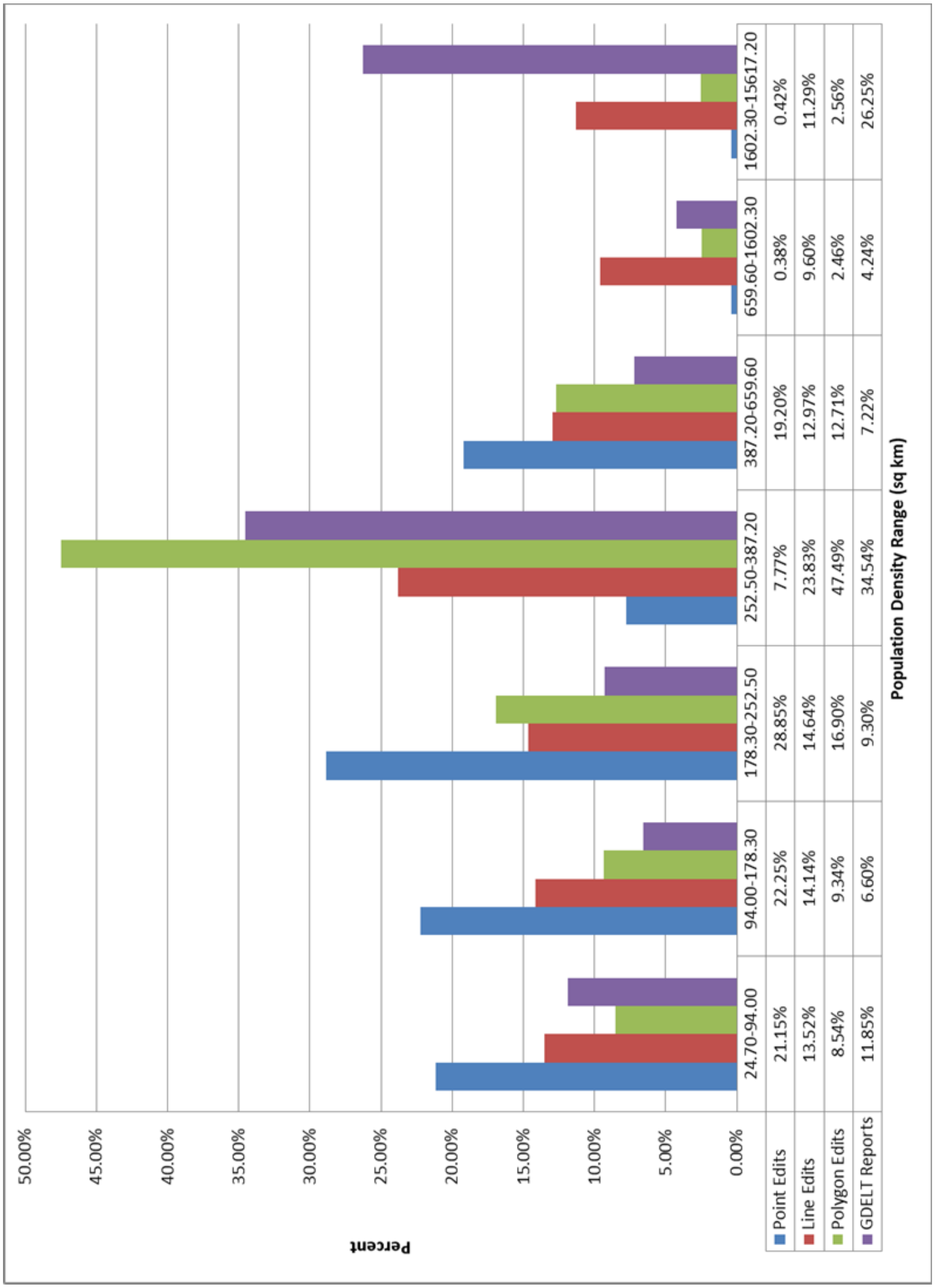


Figure 25: Population Density as Percentages in Philippines for Typhoon Haiyan

6.6.2 Population Density Comparisons for Moore Tornado

A population density layer was created based on the counties of Oklahoma. After creation the amount of OSM edits and Gdelt reports were calculated for each density category. Figure (25) shows the amount of OSM point, line, and polygon edits along with the amount of Gdelt reports present. The highest amount of OSM point edits was located within the medium populated layers of Oklahoma. Line edits are spread throughout all of the density layers. The highest concentration of Gdelt Reports is geo-referenced within the medium populated areas.

Although the numbers of OSM edits and Gdelt reports present within each population density category provides beneficial information, the percentage of OSM edits and Gdelt Reports within each category gives a glimpse into the pattern of updating and referencing in relation to the days leading, of and after an event. Figure (26) shows the percentage of OSM Edits and Gdelt reports. The highest percentage of OSM point edits are posted within the lower end of the medium populated category with about 71.34% of the point edits posted within this category. OSM line edits are present in similar percentages across all seven density categories however the highest percentage is in the lower populated density categories. OSM polygon is present within all categories as well however the spike in percentage is located within the medium populated entity categories. GDELT reports are also present within all areas of density but the largest percent of GDELT reports are presented within the high and low categories that make up the middle of the population density categories.

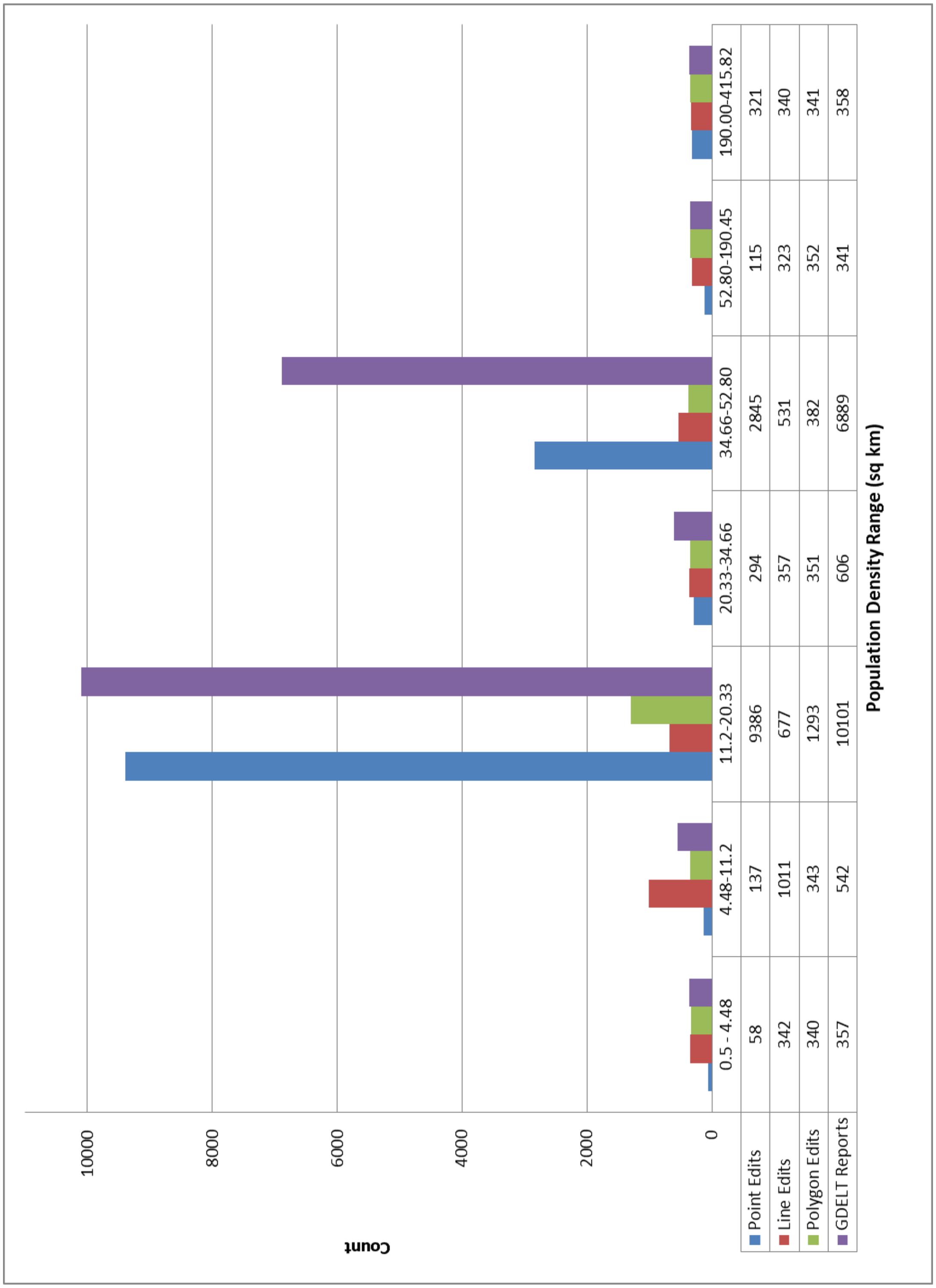


Figure 26: Population Density for Oklahoma during Moore Tornado

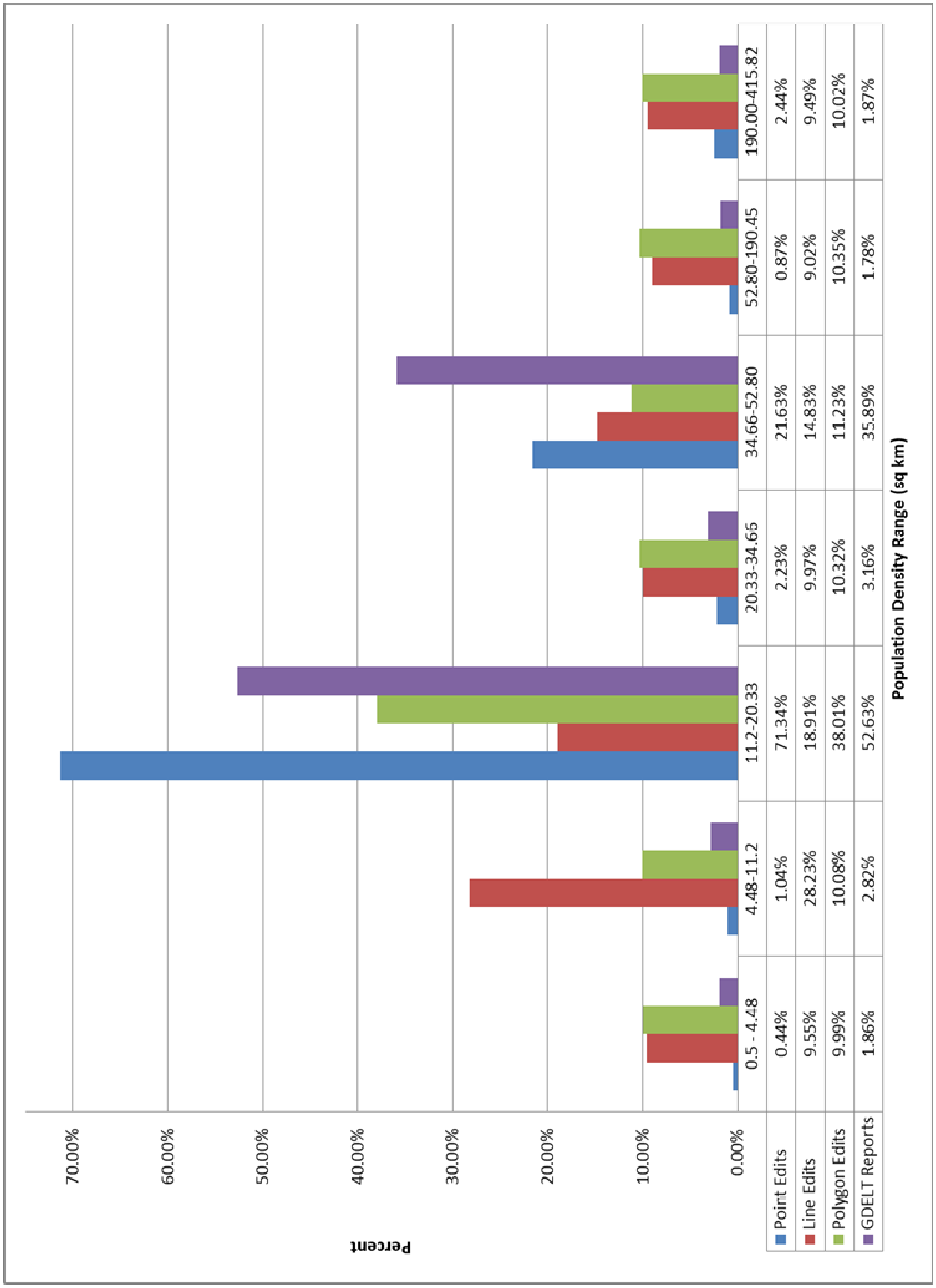


Figure 27: Population Density as percentages for Oklahoma during Moore Tornado

6.7 Casualty Location within the Philippines

Typhoon Haiyan was directly related to 6,193 deaths (The Official Gazette, 2013). The causes of these deaths ranged from drowning's to having bridges falling on individuals. Armed with the knowledge that Typhoon Haiyan caused significant loss of life the question was brought forth inquiry whether OSM edits and Gdelt Reports can be linked to areas linked to causalities. Figure (27) shows the provinces of the Philippines that experienced at least one death caused by Typhoon Haiyan. Along with highlighting the areas of causalities, the amount of casualties within each Provence also provides information regarding updating patterns as they can be compared to previously introduced maps and figures showing VGI positive areas. Figure (28) provides the number of casualties of each province with causalities. One province had a significantly higher amount of causalities than the other affected provinces. Leyte providence has 5,395 Typhoon Haiyan attributed causalities and was one of the first provinces hit by the storm. The rest of the affected provinces had single, double or triple digit associated deaths. The next province to experience the highest amount of casualties was Samar with 225 casualties. From a comparison of Figure 27 and Figure 22, the provinces with casualties fall with the breadth of population density, in that all of them do not fall into one or two population density categories.

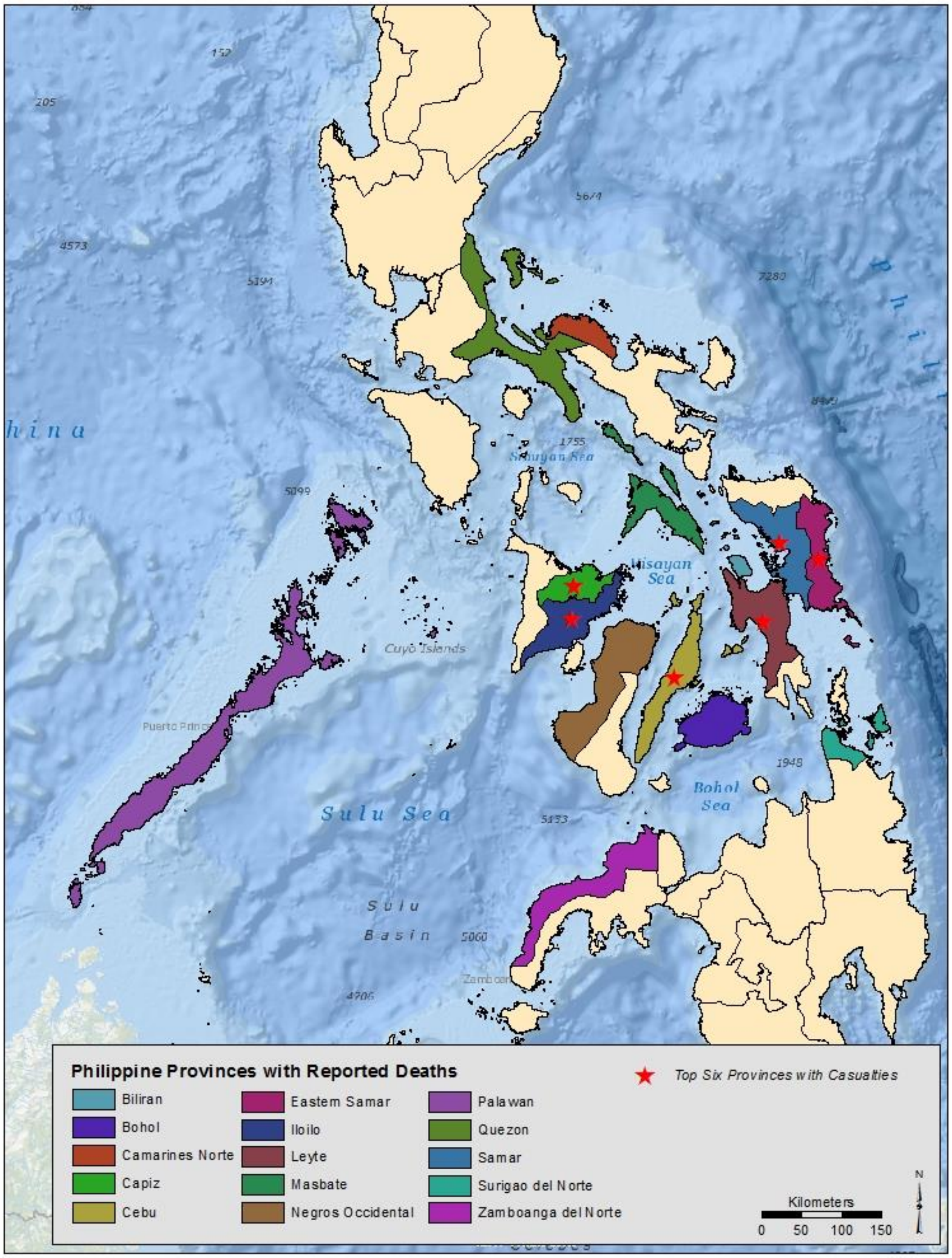


Figure 28: Provence's in the Philippines with Deaths caused by Typhoon Haiyan

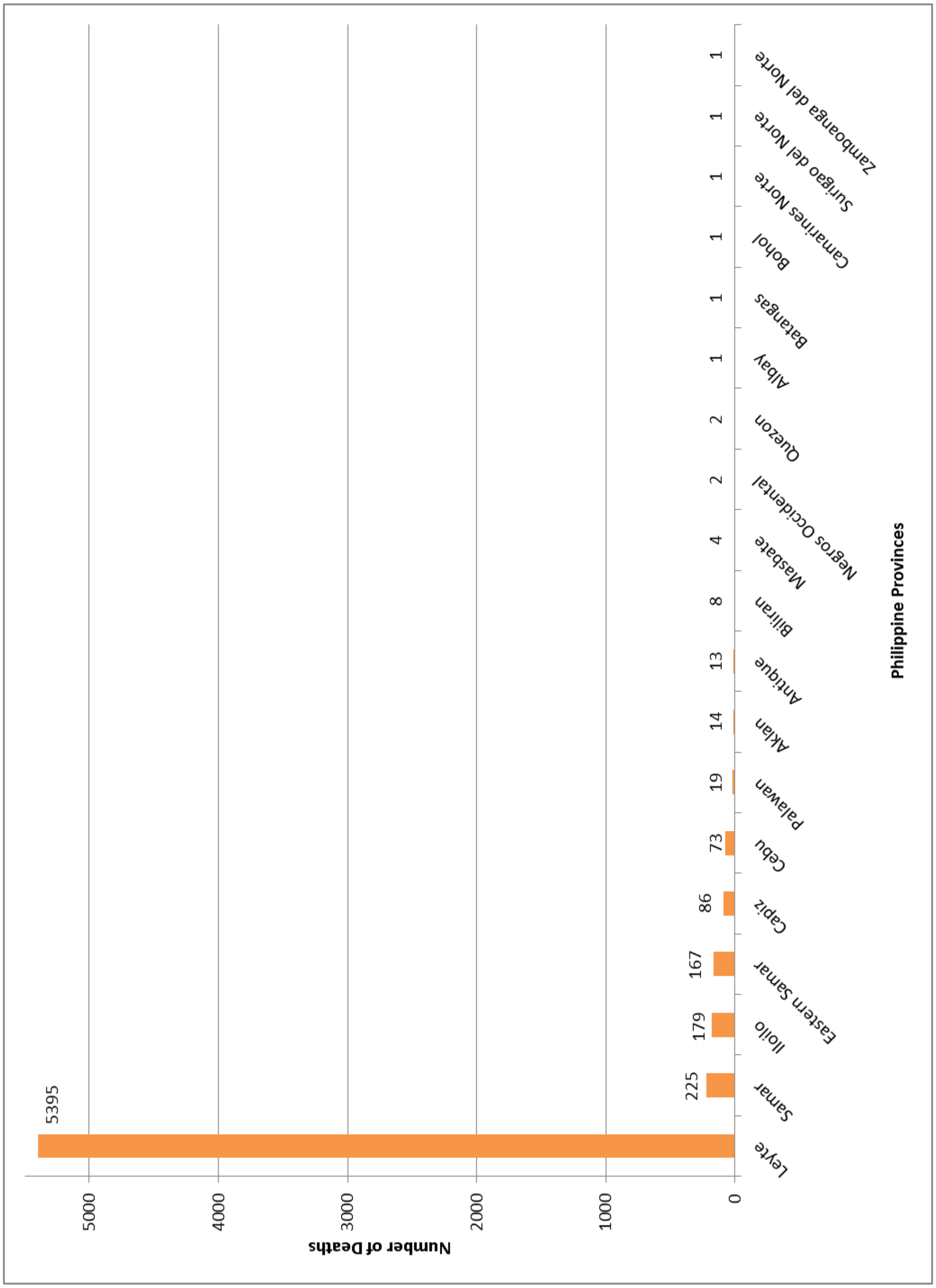


Figure 29: Philippine Provence's with deaths directly related to Typhoon Haiyan and associated counts

6.8 Philippines and Pre-Haiyan Event Editing Patterns

To understand the editing scope and patterns of VGI OSM edits it is beneficial to look at the editing patterns that exist before and after a natural disaster event. This comparison of before and after OS Edits and GDELT reports can help identify patterns and /or anomalies with VGI data. To help identify patterns within the Philippines, OSM point, line, and polygon edits were pulled for November 2012 and November 2014. This data provides a glimpse into the editing patterns that were present a year before and a year after Typhoon Haiyan. Figure 30 shows the point density for 2012 and Figure 31 shows the point density for 2014. The density patterns present in 2012 are located in both low and high density populated areas but restricted to the North West portion of the Philippines. The 2014 point density map shows a more broad spread. The highest density is located in the Philippines capital of Manila but also the 'backbone" of the Philippine islands. These 'backbone' islands are usually the first to be affected by Typhoons and were impacted by Typhoon Haiyan.

Moving on with 2012 and 2014 OSM comparisons, Figure 32 shows the polygon density of the Philippines for year 2012. The majority of polygon densities for 2012 are located within northern part of the Philippines provinces with the highest density located near Manila. Edits are also present along the lower portion of the Philippines providences. The density for year 2014 is shown in Figure 32 the polygon edits are still showing the highest density near the capital of Manila however edits are present in this year along the central backbone provinces that were directly hit first by Typhoon Haiyan.

The last OSM edit mapped and explored between years 2012 and 2014 was line edits. Figure 34 shows the line density for year 2012, the majority of edits are located near the country capital of manila and further up along the Northern provinces. For November of 2014, a massive outpouring of line edit occurred. In November, nearly all of the Philippine Regions or provinces experienced a higher level of line editing. As shown by figure (35), a moderate density of line edits exists across the entire Philippines. However, the highest density of line edits is located near the capital and surrounding lower than manila but still higher density populated provinces. Upon examination of the OSM line editing file, the result of what is seen in Figure 35 is due to boarders of all islands and regions within the Philippines having their boarders redrawn. This sudden and seemingly dramatic line editing of the entire Philippines can be traced to multiple events; including boundary disputes between China, Taiwan and the Philippines.

For a comparison check to validate and see if the OSM editing patterns within the previous three graphs were showing a pattern that was somewhat similar to other AGI sources, a three day collection of Twitter tweets was collected. The results of the Tweeter collection are shown in figure 32. Upon comparing the twitter graph to the other graphs, it can be seen that both twitter and OSM edits are concentrated around the capital of Manila. While the Twitter range is shorter than the collection range for OSM edits, the concentration near the capital provides insight that other AGI sources are showing the same updating patterns.

6.9 OSM User Activity in Philippines

Besides the locations and dates of when OSM features were posted, the contributions of OSM contributors were also explored. Figure 37 shows the top twenty active OSM contributors that provided content during Typhoon Haiyan. Table 5 shows the actual line, point, and polygon features that were posted by the contributors. Overall, each layer had over six hundred active users that actively participated, however in order to gain an overarching standard, only the top twenty were recorded and their features edits counted. Of the top twenty users only four contributed all three types of feature edits.

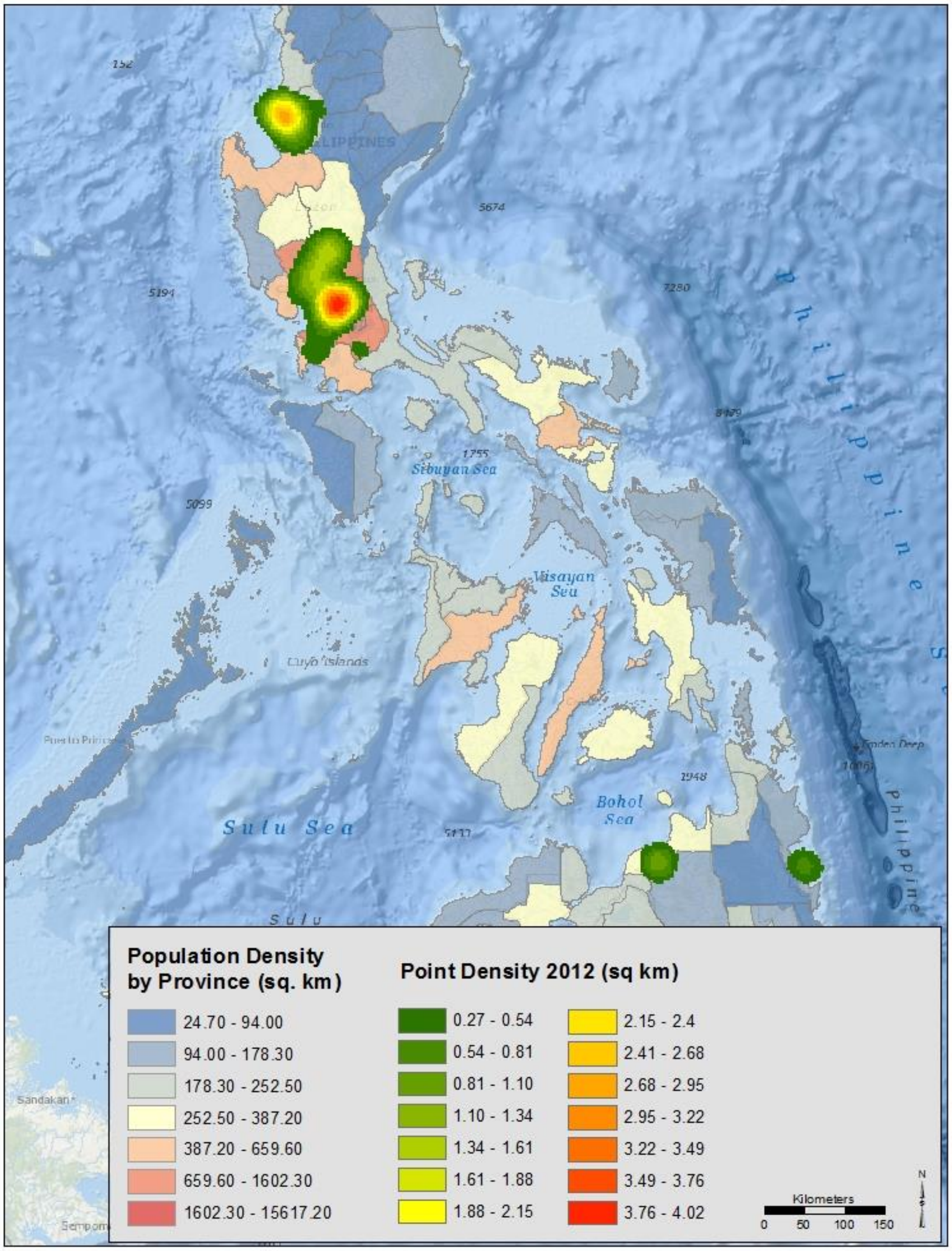


Figure 30: Point Density Comparison for Year 2012

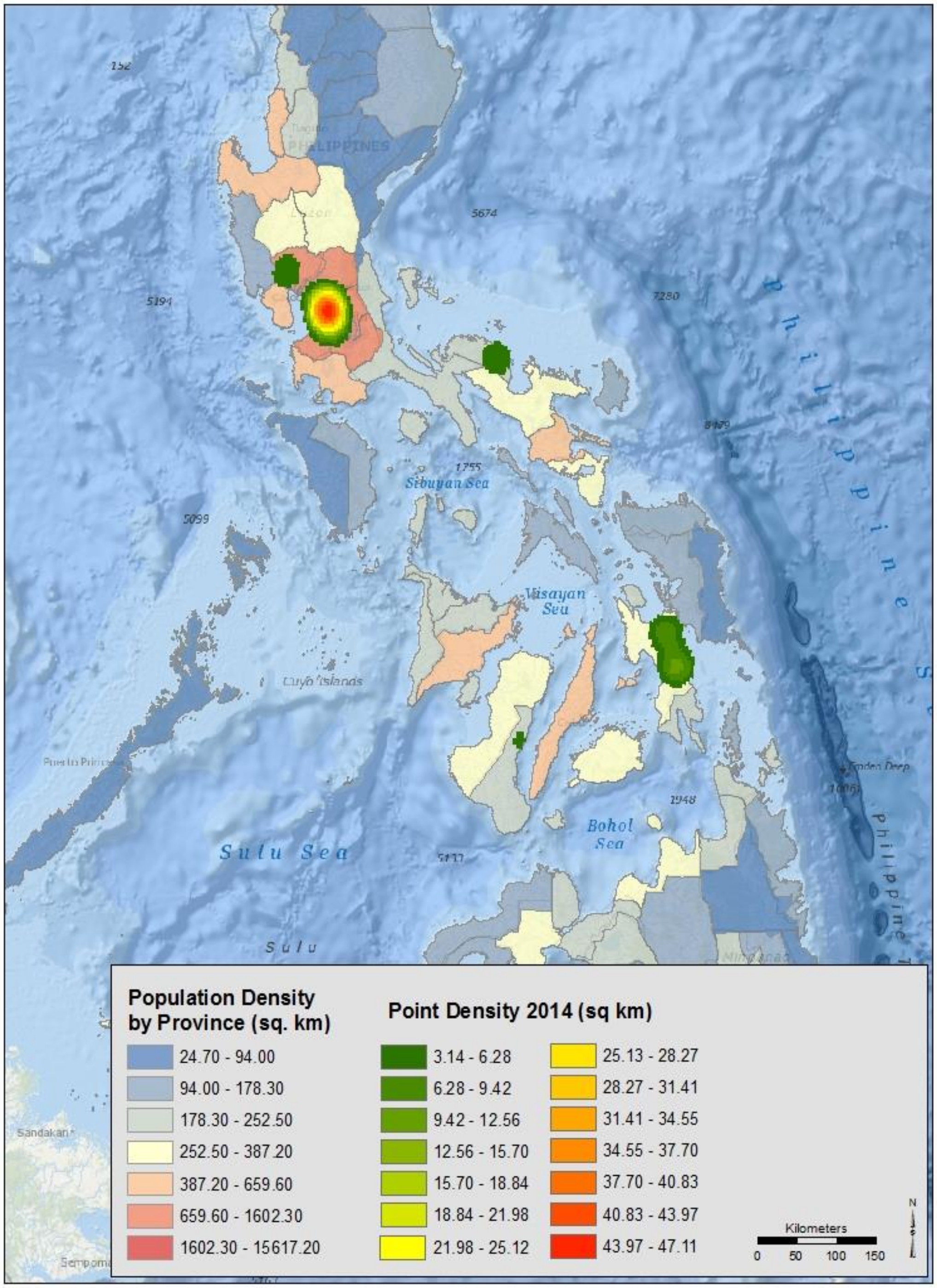


Figure 31: Point Density for Year 2014

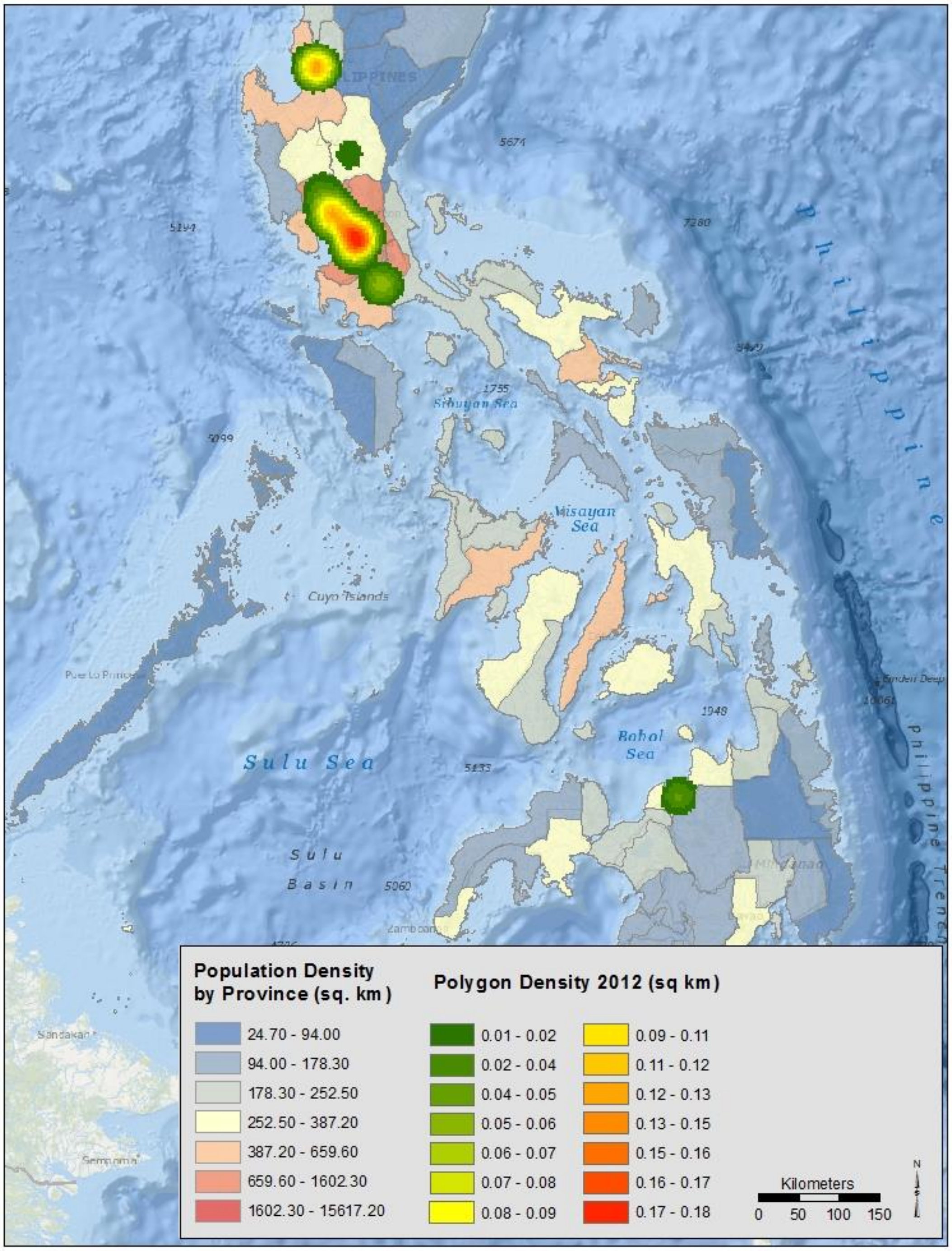


Figure 32: Polygon Density Comparison for Year 2012

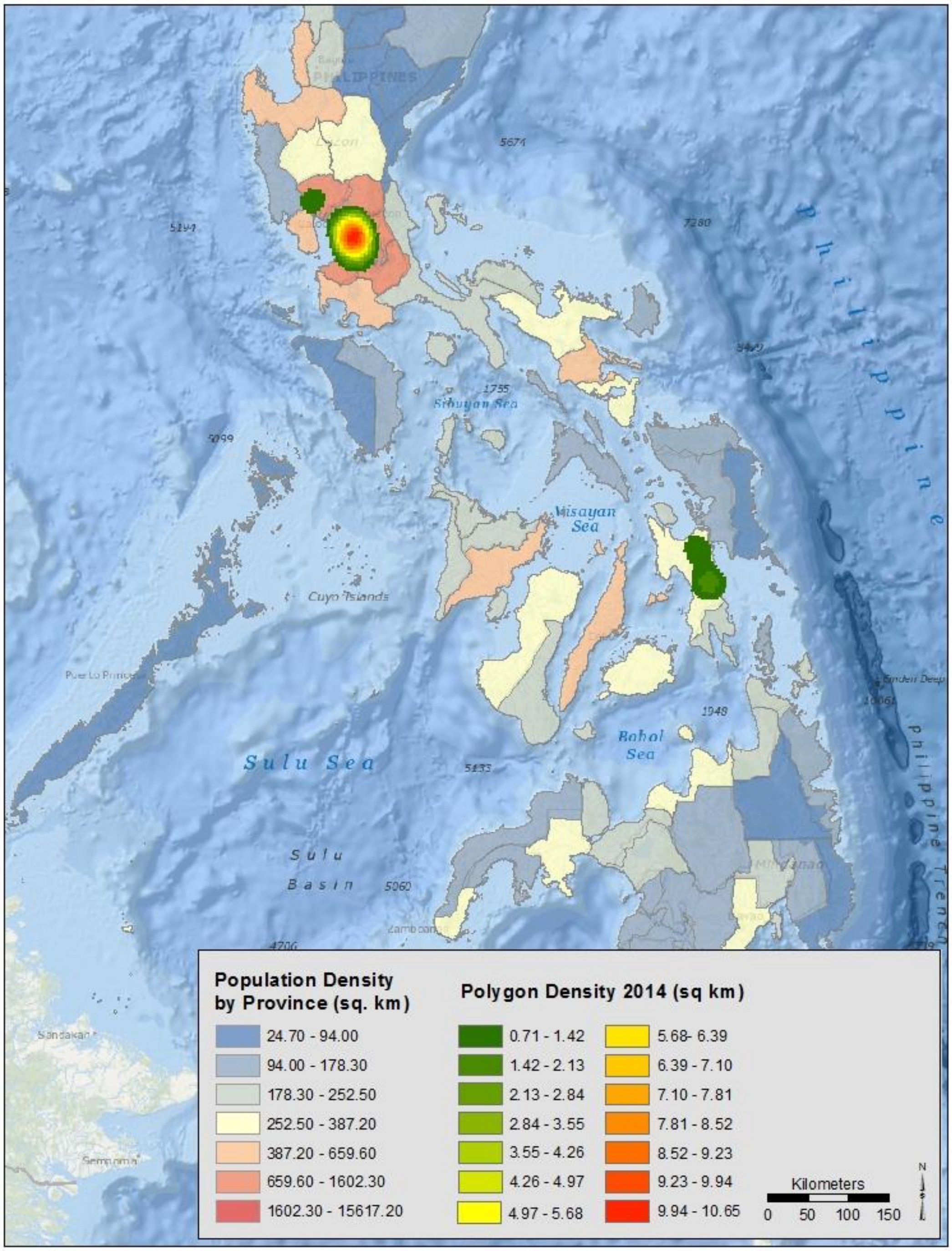


Figure 33: Polygon Density for Year 2014

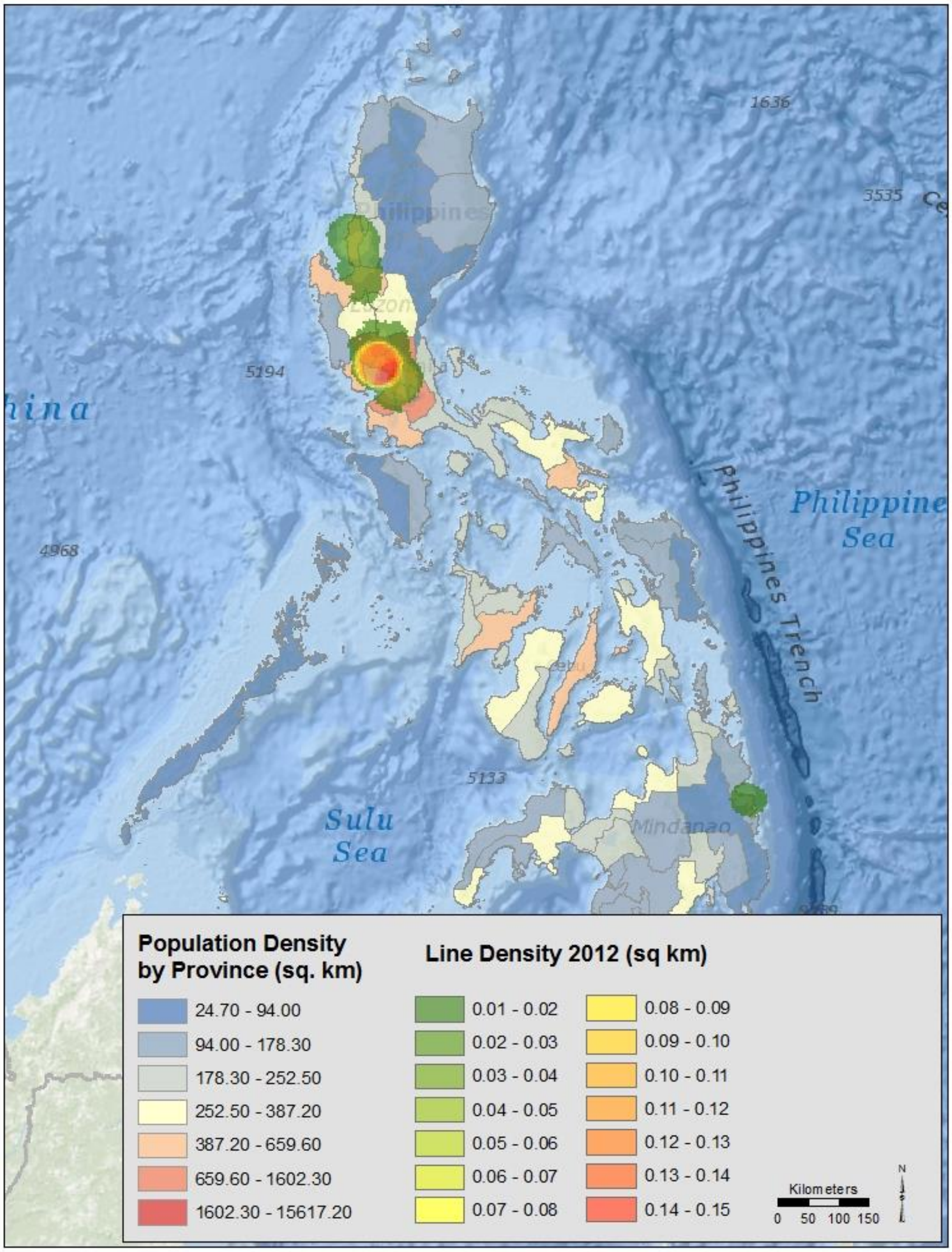


Figure 34: Line Density for Year 2012

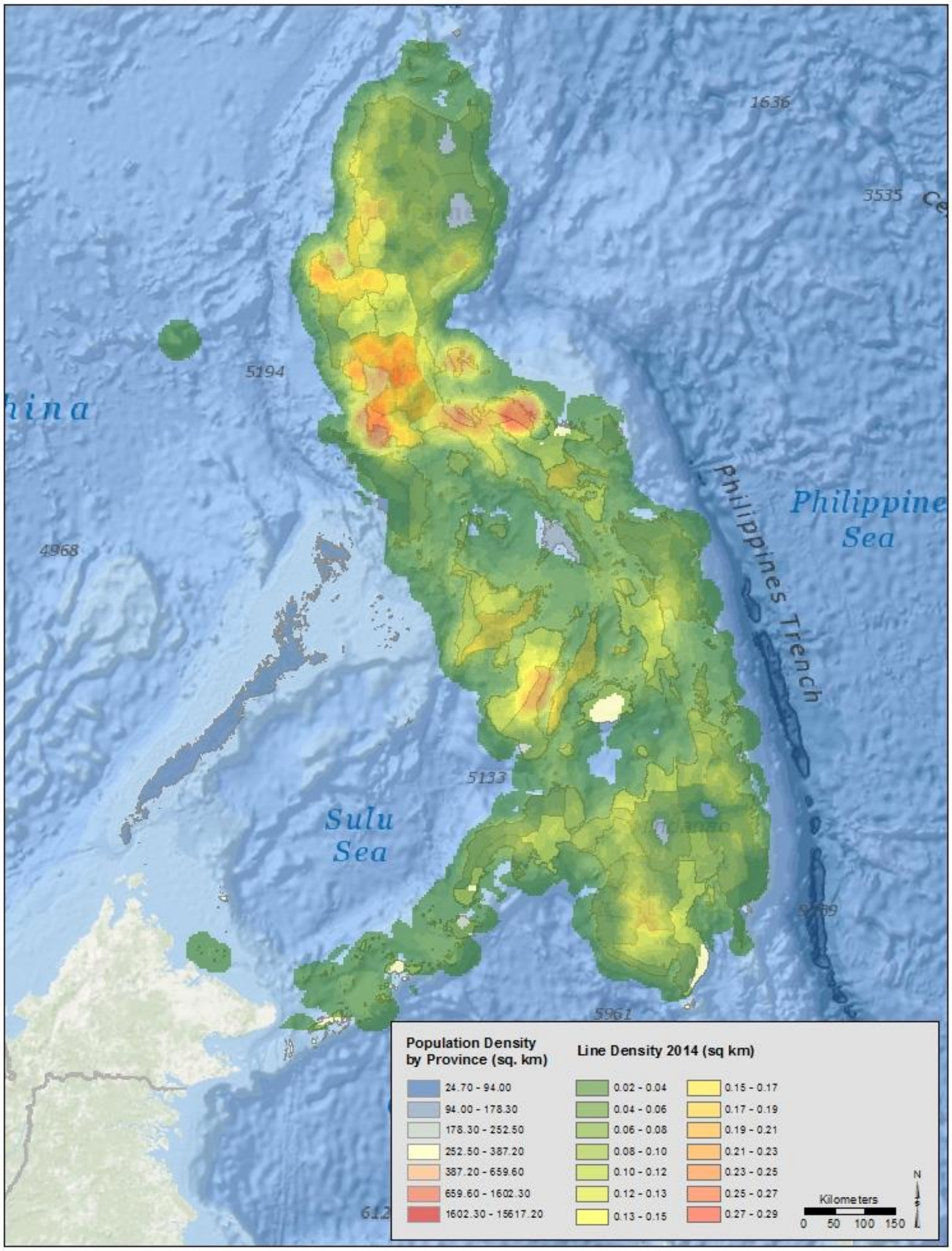


Figure 35: Line Density for Year 2014

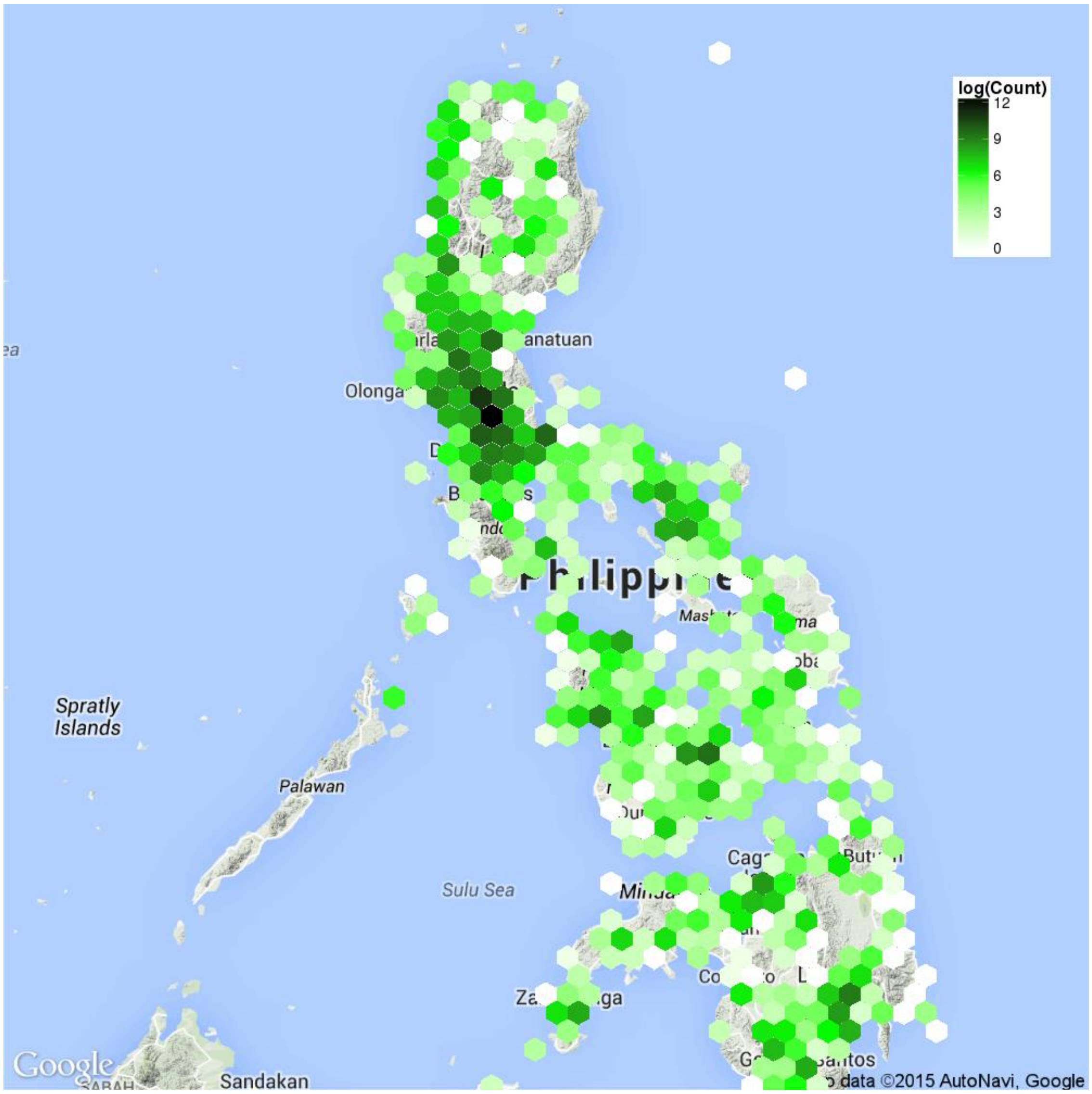


Figure 36: Twitter Count for November 7-11, 2014

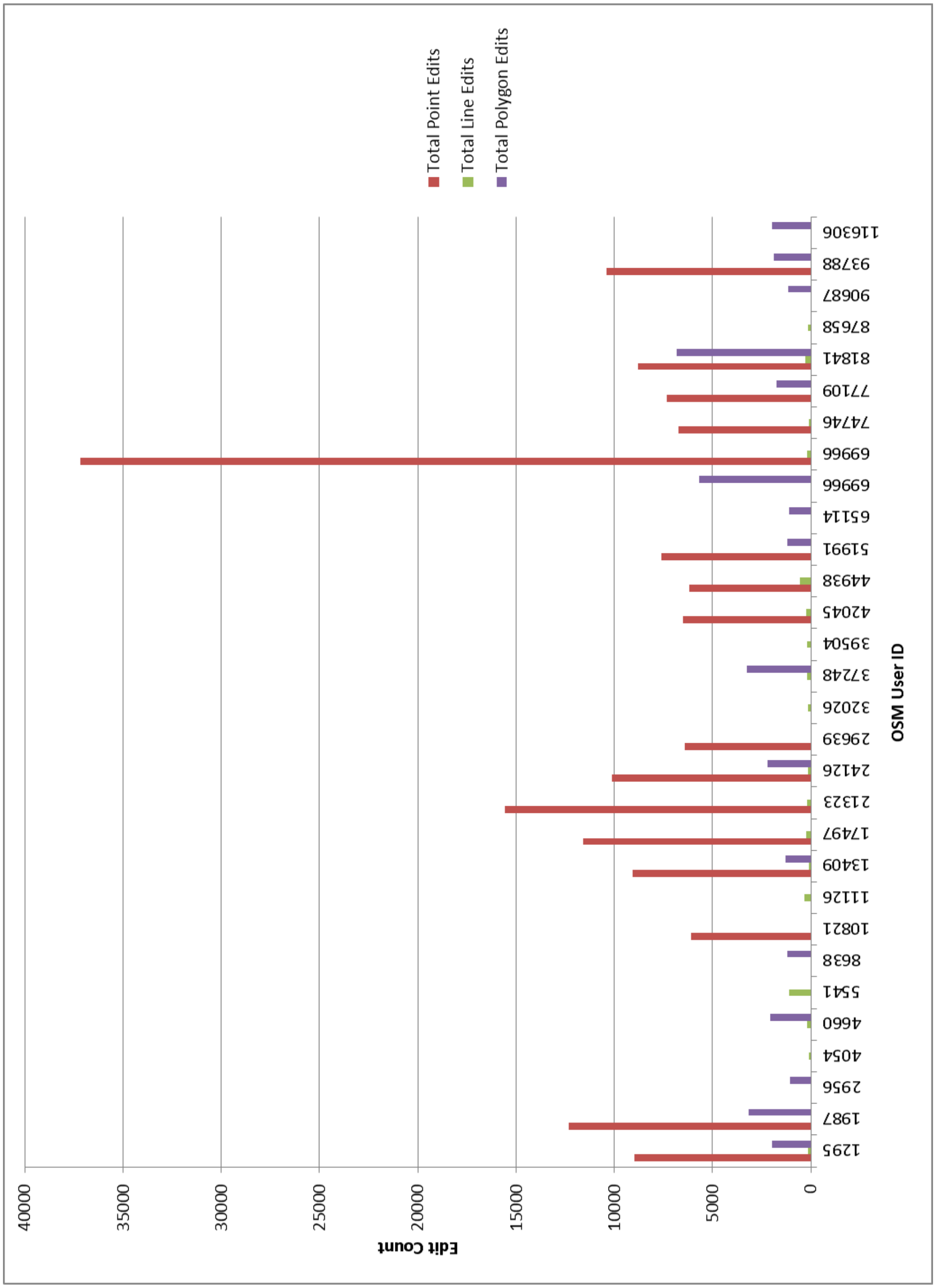


Figure 37: Top Twenty Active Users and Feature Count Totals

Table 5: Top Twenty Active Users and OSM Edit Count for each Feature Type

OSM User ID	Total Point Edits	Total Line Edits	Total Polygon Edits	Total Edits Per User
1295	8982	158	1970	11110
1987	12324	0	3172	15496
2956	0	0	1065	1065
4054	0	104	0	104
4660	0	195	2068	2263
5541	0	1097	0	1097
8638	0	0	1187	1187
10821	6095	0	0	6095
11126	0	343	0	343
13409	9093	117	1312	10522
17497	11576	233	0	11809
21323	15569	190	0	15759
24126	10109	126	2205	12440
29639	6409	0	0	6409
32026	0	140	0	140
37248	0	193	3263	3456
39504	0	196	0	196
42045	6488	251	0	6739
44938	6200	558	0	6758
51991	7604	0	1216	8820
65114	0	0	1100	1100
69966	0	0	5690	5690
69966	37179	176	0	37355
74746	6720	122	0	6842
77109	7340	0	1764	9104
81841	8798	281	6837	15916
87658	0	167	0	167
90687	0	0	1136	1136
93788	10420	0	1887	12307
116306	0	0	1991	1991

CHAPTER SEVEN

7.1 Discussion and Conclusion

The present thesis was designed and constructed to test a main hypothesis and four subsequent research questions. The main hypothesis and subsequent research questions aimed to explore the updating patterns of VGI contributed data including what OpenStreetMap edits and GDELT report updates were present in relation to a natural disaster event. Topics of investigation and research included update and edit location in relation to the disaster, what map features were edited, scope of edits and updates and if area affected is related to social media response. The results showed that social media users follow a similar pattern of updating VGI platforms or programs on the day of or right after the occurrence of a natural disaster. This updating pattern is not affected by the type of natural disaster. All three standard OpenStreetMap edits of point, line, and polygon were present in the OpenStreetMap associated natural disaster layers. The majority of OpenStreetMap edits are within the location of where the natural disaster occurred while GDELT reports are predominantly located in correlation to the largest city near the disaster affected location. The scope of an edit extends from general categories including geological, and boundary to more specific categories including tourism and leisure. Social Media update categories range based on what tags or key words are assigned based on the material of the update. These categories include weather related tags such as high winds and flash floods to more general tags including flood and storm damage. The amount of ground surface impacted or affected by the natural disaster event does impact the social media response. The larger the impacted area, the greater the social media edit and update response although other extenuating circumstances can play a part in the amount of social media updates and edits present.

The two data sources used within this study were edits originating from OpenStreetMap and user updates from the GDELT database. The OpenStreetMap edits and GDELT reports were recorded for two natural disaster events. The first event was the Moore Tornado in Oklahoma which occurred on May 20, 2013 and Typhoon Haiyan which made landfall in the Philippines on November 8, 2013. The social media generated by OSM and GDELT in regards to these listed natural disaster events followed similar posting patterns of only a small number of records were posted before an event but response and numbers increased rapidly after the event occurred.

GDELT update patterns varied from OSM edit counts on a daily basis in that no GDELT reports were posted until the day of or after the event occurred. The reason for this pattern is most likely linked to the nature of the GDELT database. GDELT derives its information from both foreign and domestic news sources. In some instances, media generated content takes longer to post due to verification of event and sources. On the other hand, OpenStreetMap edits corresponding to a related natural disaster event tend to appear before an event in limited quantities unlike GDELT which had no presence before an event.

For both the natural disasters studied in this thesis, each one had a corresponding area that was affected by the event. Not all related events were located within the affected track. For the Moore Tornado, over fifty percent of point and polygon edits and fewer than ten percent of line edits were geo-located within the damage track. One reason why the number of updates was below fifty percent is the social OSM editor could have associated damage caused to surround infrastructure as being caused by the tornado when it fact its origin stemmed from another source such as a previous windstorm. The thunderstorm that generated the Moore Tornado was not the first one thunderstorm to occur that week and previous storms had done damage to the surrounding counties. A reason why OSM line edits related to the storm but located outside of the damage path was higher than ninety three percent can be linked to two reasons the first being actually items that could have been marked as line edits such as roads or rivers were marked by point of polygon edits. Instead of an editor showing the section of demolished road by a line edit, they instead mark it with a point edit. Another reason for the low edit could have been not many significant or major linear features were affected by the storm. The density generated maps also show where edits were located within the damage tracks.

GDELT user updates percentages of the updates associated with the natural disaster events was below forty percent for Typhoon Haiyan and below one percent for the Moore Tornado. This low of can be linked to the structure of GDELT reports . GDELT reports take into account international new sources and as a result updates can be generated by users unfamiliar with the geography of the disaster location. Instead of linking updates to exact geo-coordinates, the update is geo-located as being located near the largest city. Of the two social media sources OpenStreetMap had more disaster related events edits within the damage track of the Moore Tornado and Typhoon Haiyan than GDELT reports.

GDELT reports are represented as point linked or geo-located to a specific point on the earth's surface. Although, GDELT does not break editing into categories such as OSM, GDELT does generate the number of updates associated with each geo-located coordinate. For example, a GDELT update located within the Oklahoma City is represented as one point but that one point has two media sources updates associated with it.

The scope of edits varied based on whether the source was VGI data from OpenStreetMap or AGI data from the GDELT database. The scope of the OpenStreetMap edits ranged from land cover categories such as geological and barrier to land use categories including roads and railway. The majority of OSM edits were categorized; however point edits were the least categorized. The reason for the lack of categorization can be linked to one of many reasons. Most likely the contributing factor for the lack of categorization is linked to user either posting edits are quickly as possible and neglecting to take the time to select a category or users assume the geo-location of the point is enough information to deduce the scope of the edit.

The scope extent associated with GDELT user updates for this thesis varied with over twenty-eight different categories. The GDELT scope extent is based on key words or tags that are broad enough to be applicable to multiple instances. An example is "natural disaster – tornado". In the context of itself this tag could refer to any tornado but when you add date and geolocation information it references a specific instance. Overall the GDELT database has over 159 tags that can be associated with user updates. In similarity with OpenStreetMap edits, GDELT reports do not have to have a specific category but must have at least the general “natural disaster” tag.

The Moore Tornado affected an area of around 18.39 km^2 within Oklahoma. Typhoon Haiyan affected around 147,880 km^2 of the Philippines. The OpenStreetMap edits and GDELT user updates were higher for Typhoon Yolanda than for the Moore Tornado. In theory, the answer for this research question is yes, the size of the affected area affects the amount of VGI and AGI media related to the event. However, the significant differences between the results of the Moore Tornado and Typhoon Haiyan may be contributed to social media outreach campaigns.

Due to its extensive use in Haiti during the earthquake in 2010, OSM has continually started outreach mapping campaigns to involve users of OSM to contribute data in response to natural disasters. Both OSM files used in this thesis contained a column of data that listed the edit as in support of an outreach campaign or not.

Specific conclusion on the updating patterns of users pertaining to a specific event such as the Moore Tornado and Typhoon Haiyan can be presented and explained but conclusions about how users are using crowd source content overall can be concluded from this thesis research. However, overarching patterns related to crowd sourced content and natural disasters emerged during the research for this thesis. OSM point edits were the most numerous due to the least amount of time to create them. Point edits are point and click and can be posted the fastest. Line edits were next followed by polygon edits, this order can be contributed to the time it takes to create such as edit. Users whether they are residents of the affected area ort citizens trying to help, tend to focus on areas where the infrastructure is more advanced or built up. Editors tend to focus more on damaged cities and residents rather than ripped up or damaged farmland. Editors tend to focus on areas that provide meaning for them. For example, in the case of the Moore Tornado, numerous edits were focused on school locations where children had perished. Most OSM edits are categorized but instead rely on their location and associated metadata to provide insight into what the edit is pertaining to.

What is interesting to note is that the Humanitarian OpenStreetMap Team (HOT) had an extensive outreach project in response to Typhoon Haiyan. Dedicated webpages were contrasted in order to help map the damaged areas affected by the typhoon. While the HOT website declares that all verified edits are imported into the main OSM architecture, there seems to be a disconnect between what the HOT website says and what was downloaded for the OSM database.. The amount of data or edits that HOT claims were posted in response to Typhoon Haiyan does not match the amount of edits that were downloaded from the GeoFabrik website. The reason for this disconnect could be due to multiple reasons including, not all edits were uploaded or edits are being removed once they become obsolete. Finding the reason for this disconnect could provide an area of further research.

This thesis explored crowd sourced content and how it is used in relation to natural disasters. Crowd sourced content encompasses both VGI and AGI media sources. Figure 32 provides a sample structure of crowd sourced content as well as examples of VGI and AGI sources. While having different characteristics, VGI and AGI can be used in conjunction with each other to identify potential natural disaster hot spots as well as track the response of users to the event. Natural disaster events have been occurring across the globe since the beginning of recorded time and they will continue to occur. Knowledge of natural disaster locations is of value if not only for awareness purposes. A single government or organization cannot maintain knowledge of natural disaster events that are occurring all over the world. However, with the surge in crowd source content, benefits from the collective knowledge of users can provide the needed insight. AGI source, such as GDELT reports, can provide the “heads up” or indication that there is an event either occurring or set to occur. After a general location have been identified using VGI sources, including but not limited to OSM, can provide more details about the extant location of the event. However, VGI can provide greater benefit after the event has occurred as residents will provide damage extent and unaffected areas of the location.

The study presented in this thesis provided a valid test of the stated hypothesis and associated research questions. The guidelines and methods were standardized and repeatable allowing the downloading and analyses to be repeated or updated if the need should arise.

7.2 Future Work

This thesis covered an aspect of Volunteered Geographic Information that has not been extensive explored. In retrospect this thesis topic only begins to scratch the surface of how Volunteered Geographic Content can benefit the intelligence gathered for and from a natural disaster. The methods and structure from this thesis can be applied to other Volunteered Geographic Information sources and natural disasters to see if the same results occur.

One way to further expand the research presented in this thesis is expand to other AGI and VGI sources besides the GDELT database and OSM. Recent events have shown that like most databases, GDELT data has obstacles that a user may face when using the database. Because GDELT relies partially on news feeds, news channels are rarely without some form of bias. This means that the data going into GDELT has the potential to be presented favoring one thought or another. The result of this is data that is more negative or positive than the general public opinion.

OSM has database issues occurring as well. The fact that anyone can register for account and then start editing leads to the quality and intent of the update or change to be called into question. Some would hope that the reason for the change in OSM is for the greater good; unfortunate in today's society that may not be the defined case. The biggest problem with OSM I currently see is that minimal quality controls are taken into effect.

A general problem faced with open source media is the quality and validity of the data that is being used. People also post irrelevant information that can lead one to believe more posts are occurring about an event when in reality nothing much more is happening then someone commenting on their lunch. Users also tend to post when a negative event has happened, and they either want help from a source or width to air their grievances to the user community, an example of this is a user posting a request to the national guard twitter website asking to evacuation assistance after a flood or someone posting a negative yelp review about their lunch server. While studies and research have been conducted on how to validate and control the quality of user generate content it would be interesting to focus on the users that are posting content including posting patterns, type of edits posted and locations a user is posting from and what locations they are editing.

The research period for this thesis was only two weeks. This short observation or research period only provides a 'soda straw' look at updating patterns. The methods and analysis queries conducted in this thesis could be applied to a longer time frame. This could lead to understanding the updating patterns related to a natural disaster over a longer time frame.

The exploration into how successful Volunteered Geographic Information outreach or grassroots programs are at generating useful data. As Volunteered Geographic Information use continues to rise, organizations are trying to gather relevant data corresponding to specific event or locations. One such example is the missing maps sponsored by the Humanitarian OSM team (HOT), whose goal is to map areas that have not been done so in the past.

The focus of this thesis was the updating patterns via Volunteered Geographic Information in regards to natural disasters. It would be interesting to apply the stated methods and analysis practices to events that are not considered natural disasters and see if the same patterns emerge. An example would be to apply the stated methods to Ebola epidemic in Serra Leone. Another non –natural disaster event that could be used could be the Boston Marathon bombings. Volunteered Geographic Information played a significant part in intelligence gathering and information sharing during the days following the bombings.

CHAPTER EIGHT

8.1 References

Arsanjani, J. J., Helbich, M., Bakillah, M., Hagenauer, J., & Zipf, A. (2013). Toward Mapping Land-use Patterns from Volunteered Geographic Information. International Journal of Geographical Information Science, 27(12), 264-278.

Bakillah, M., Liang, S., Mobasheri, A., Arsanjani, J, J., & Zipf. (2014). A., Fine-resolution Population Mapping Using OpenStreetMap Points-Of-Interest. International Journal of Geographical Information Science, 1-24.

Croitoru, A., Crooks, A., Radzikowski, J., & Stefanidis, A., (2013). Geosocial Gauge: A System Prototype for Knowledge Discovery from Social Media. International Journal of Geographical Information Science, 27(12), 483-508.

Curtis, A., & Mills, J., (2009). GIS, Human Geography, and Disasters. San Diego, CA: U Readers, Print.DOI: 10.1007/s10708-014-9564-8

Elwood, S., Goodchild, M., Sui, D., (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, & New Social Practice. Annals Of The Association Of American Geographers,102(3), 571-590.

Explore Layers: Yolanda Typhoon. GeoNode. Explore Layers - Yolanda Typhoon. Retrieved from http://yolandadata.org/layers/

Goodchild, M., F., & Glennon, J., A., (2010). Crowdsourcing Geographic Information for Disaster Response: A Research Frontier. International Journal of Digital Earth 3(3) 231-241.

Goodchild, M., F., & Li., L., (2012). Assuring the Quality of Volunteered Geographic Information. Spatial Statistics, 1(1), 110-120.

Goodchild, M., F., (2007). Citizens as Sensors: The World of Volunteered Geography. GeoJournal, 69(4), 211-221.

Greene, R., W., (2002) Confronting Catastrophe: A GIS Handbook. Redlands, ESRI Press, New York. Print

Haddow, G., D., & Haddow., K., (2014). Disaster Communications in a Changing Media World. 2nd ed. Waltham: Elsevier. Print.

Haklay, M. and Weber, P. (2008), Openstreetmap: User-generated Street Maps. IEEE Pervasive Computing, 7(4), 12-18.

Haklay, M., (2010). How Good Is Volunteered Geographical Information? A comparative Study of OpenStreetMap & Ordnance Survey Datasets. Environment and Planning B: Planning and Design, 37(4), 682-703.

Jackson, S., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., & Stefanidis, A., (2013). Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. ISPRS International Journal of Geo-Information, 2(2) 507-530.

Mountrakis, G., & Stefanidis., A., (2011). Moving towards Personalized Geospatial Queries. Journal of Geographic Information System, 3(4), 334-344.

Mullen, W., F., Jackson, S., P., Croitoru, A., Crooks, A., Stefanidis, A., & Agouris, P., (2014). Assessing the Impact of Demographic Characteristics on Spatial Error in Volunteered Geographic Information Features. GeoJournal,

National Weather Service: Weather Forecast Office, Table of Tornadoes Which Have Occurred in the Oklahoma City, Oklahoma Area Since 1890. (2014 January), Retrieved from http://www.srh.noaa.gov/oun/?n=tornadodata-okc

Neis, P., & Zipf, A., (2012). Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap. ISPRS International Journal of Geo-Information, 1(3), 146-165.

Neis, P., Zielstra, D., & Zipf, A., (2012). The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. Future Internet, 4(1), 1-21.

Oklahoma State Information - Symbols, Capital, Constitution, Flags, Maps. (n.d.) Retrieved from http://www.50states.com/oklahoma.htm#.VLhRdCvF-So

Oklahoma: Oklahoma's Official Travel & Tourism Site. (n.d.). Retrieved from http://www.travelok.com/oklahoma_city

OpenStreetMap: About. (n.d.) Open Street Map Information, Retrieved from http://www.openstreetmap.org/about

Prevatt, D., O., (2013). Post-Tornado Survey, Moore, OK. University of Florida

Radke, S., L., Johnson, R., & Baranyi, J., (2013). Enabling Comprehensive Situational Awareness, 1st ed. Redlands: ESRI, 2013. Print.
Retrieved from http://www.davidoprevatt.com/?s=engineer

Rice, M. T., Caldwell, D. R., Paez, F. I., Mulhollen, A., Shore, B. M. (2012) Crowdsourced Geospatial Data: A report on the emerging phenomena of crowdsourced and user-generated geospatial data". US Army Corps of Engineers, Engineer Research and Development Center, US Army Topographic Engineering Center Technical Report, Data Level Enterprise Tools Workgroup. 1-148.

Rice, M., T., Jacobson, D., Caldwell, D., R., Mcdermott, S., D., Fabiana, P., I., Aburizaiza, A., O., Curtin, K., M., Stefanidis, A., & Qin, H., (2013). Crowdsourcing Techniques for Augmenting Traditional Accessibility Maps with Transitory Obstacle Information. Cartography and Geographic Information Science, 40(3), 210-219.

Sester, M., Arsanjani, J., A., Klammer, R., Burghardt, D., & Haunert, J., (2014) Integrating and Generalizing Volunteered Geographic Information. Abstracting Geographic Information in a Data Rich World. Print. 119-155

Social Networking Fact Sheet: Pew Research Centers Internet American Life Project RSS. Pew Research Centers, Jan. 2014. Retrieved from http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

Stefanidis, A., Cotnoir, A., Croitoru, A., Crooks, A., Rice, M., & Radzikowski, J., (2013). Demarcating New Boundaries: Mapping Virtual Polycentric Communities through Social Media Content. Cartography and Geographic Information Science, 40(2) 116-129.

Stefanidis, A., Crooks, A., & Radzikowski, J., (2013). Harvesting Ambient Geospatial Information from Social Media Feeds. GeoJournal, 78(2) 319-338.

Tornado Alley: Tornado Alley. Retrieved from http://en.wikipedia.org/wiki/Tornado_Alley

About Twitter, Incorporated: Company. (n.d.) Retrieved by https://about.twitter.com/company

"Typhoon Haiyan/Yolanda Fact Sheet #22." Typhoon Haiyan/Yolanda Fact Sheet #22. USAID. (2014, April 21) Retrieved from: http://www.usaid.gov/haiyan/fy14/fs22

Typhoon Haiyan: A Crisis by the Numbers.: NBC News (2013, November 07), Retrieved from: http://www.nbcnews.com/news/other/typhoon-haiyan-crisis-numbers-f2D11609327

United States Census Bureau: Oklahoma Quick Facts from the US Census Bureau. (2014December 04), Retrieved from: http://quickfacts.census.gov/qfd/states/40000.html

Visser, N., (n.d.), Moore Tornado 2013 Path: Damage By The Numbers, The Huffington Post. Retrieved from http://tinyurl.com/njhku7c

Werts, J., D., Mikhailova, E., A., Post, C., J., and Sharp, J., L., (2012), An Integrated WebGIS Framework for Volunteered Geographic Information and Social Media in Soil and Water Conservation. Environmental Management 49(4), 816-832.

Zielstra, D., & Zipf, A., (n.d.), A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. 13th AGILE International Conference on Geographic Information Science, Portugal, Guimaraes.

Zook, M., Graham, M., Shelton, T., & Gorman, S., (2010), Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. World Medical & Health Policy 2(2) Article 2.

BIOGRAPHY

Stephanie M. Parker graduated from Chantilly High School, Chantilly, Virginia, in 2005. She received her Bachelor of Science from Virginia Polytechnic and State University in 2009. She is currently employed as a geospatial analyst for the Department of Defense.