Geospatial Analysis of Unintended Casualties during Combat Training: Fort Drum, New York

by

Reina Golda Kahn

A Thesis Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

May 2015

ACKNOWLEDGEMENTS

I want to thank John Wilson for his continued guidance and patience throughout all stages of this thesis. I would also like to thank Karen Kemp for her expert advice and input. I would also like to thank the following people at SRI International for their help in obtaining permission to access the thesis dataset, as well as their thoughts and guidance during the research process itself: Isobel Wadsworth, Pratik Mehta, Kipp Peppel and Brett Heliker. Without your help, I would have no data or inspiration. Finally, I would like to thank my husband for his continued support and belief in my abilities. His continued support and words of encouragement were my saving grace not just during this thesis but in life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAR	After Action Review
ARNG	Army Reserve National Guard
COB	Civilian on Battlefield
CTC	Combat Training Center
EDI	Enhanced Dismount Instrumentation
FOB	Forward Operating Base
GIS	Geographic Information Science
GPS	Global Positioning System
IED	Improvised Explosive Device
MILES	Multiple Integrated Laser Engagement System
MOUT	Military Operations on Urban Terrain
NTC	National Training Center
OPFOR	Opposing Force
PEO STRI	Program Executive Office for Simulation and Instrumentation
RFID	Radio-Frequency Identification
SAT	System Approach to Training
SQL	Structured Query Language
TRADOC	Training and Doctrine Command
UIR	Unique Identifying Record
XCTC	Exportable Combat Training Capability

ABSTRACT

Training soldiers for combat is necessary to mitigate casualties of civilians and soldiers in the field during wartime. An advanced system of training has been developed that prepares soldiers for war by simulating combat scenarios and tracking a soldier's location and if they are shot. The data acquired from these training scenarios has the potential to inform training doctrine and improve combat performance. The use of Geographic Information Systems (GIS) to analyze fatalities in the training exercise has not been implemented to explore ways performance might be improved. This study used data acquired at an Army National Guard Exportable Combat Training Capability (XCTC) training event at Fort Drum in New York on the 7^{th} August 2013 to visualize the numbers of unique persons travelling through a cell during the day as well as the average number of people in a grid cell within 30 seconds of an engagement, hot spots of engagements on a linear network, and how the number of people and engagements changed across the field site at 15 minute intervals throughout the day. The output can then be used in the daily After Action Review (AAR) in conjunction with the training playbook and mission objectives to assist soldiers and commanding officers in clarifying what factors are contributing to the hot spots. The results might then be used to require training iterations under specific scenarios to improve training performance.

CHAPTER 1: INTRODUCTION

In 2003, when the war in Iraq began, the need for military forces to effectively and efficiently train US soldiers for deployment increased significantly. Both active-duty Army and Army National Guard (ARNG) troops needed to simultaneously train soldiers at designated National Training Centers (NTC) which resulted in limited training time for ARNG units because active-duty Army received training priority at these locations. To solve this problem the ARNG developed the Exportable Combat Training Capability (XCTC), which was designed to fulfill the training requirements for National Guard troops at military bases other than the designated NTC location (Milbrodt 2009). The ARNG's XCTC program is a series of field training exercises that provides realistic training for every participating soldier, and provides the required company level certification (Carpenter 2009) during pre-mobilization training. Geospatial analysis is not currently incorporated in the training technology, and doing so will provide additional insights and possibly better training.

1.1 Military Training

Prior to the mid-1970s, Army training was mostly held in the classroom and performance feedback was seldom provided (Anonymous 1998). Since then, Combat Training Centers (CTCs) have been developed to facilitate combat training consisting of two opposing forces engaged in a simulated combat environment. Over the years, this simulated training environment has evolved to mirror many of the possible scenarios that soldiers will encounter in the battlefield. In addition to the Army, the US Air Force has also developed training techniques that utilize simulated war fighting scenarios. In 1997, they developed the Distributed Mission Operations, which provide a synthesized training environment by linking the fighter to command and control simulators that monitor performance in the battle-space (Chapman and Colegrove

2013). The Army National Guard has achieved this same capability partly through the XCTC program, which provides measured and simulated training by linking the soldier with a weapon using the Multiple Integrated Laser Engagement System (MILES). Military training with MILES involves the soldier using a rifle stock that is mounted with a laser transmitter. Each soldier carries optical detectors on his or her helmet and on a body harness adapted to detect a laser “bullet” hit (Healey and Parikh 1995). The soldier pulls the trigger of the rifle to fire a blank cartridge to simulate the firing of an actual round and a sensor triggers the laser. Use of the MILES system during an XCTC training allows the player identification and weapon type to be encoded on the laser beam using a MILES code.

These XCTC training exercises were designed to fill the gap in training set forth by the Army. Taking the soldiers out of the classroom while maintaining consistency with the training criteria has been paramount in honing soldier’s skills by completely immersing them in an environment that looks, sounds and smells like what they will experience upon deployment.

1.2 Combat Training with XCTC

Before soldiers are deployed into a combat situation overseas, they endure a rigorous training regime that not only includes fitness training but also combat readiness preparation. This full-immersion military training often encompasses an instrumented battlefield training exercise, in which soldiers are given a scenario representing a situation similar to what they would see in combat. The soldiers engage in a battle scenario complete with objectives and targets, improvised explosive devices (IED), opposing forces and actors playing the roles of civilians. The soldiers and role players are equipped with Global Positioning System (GPS) devices that allow their movements to be tracked and recorded. The MILES unit is linked to the GPS device on the participant. This allows for the horizontal positioning as well as the weapons engagements to be recorded for each participant. The entire training scenario is then played back to them during the After Action Review (AAR) in a 3D virtual terrain that includes every participant's unique avatar. The AAR is a process, developed by the Army, designed to provide feedback to the military units on their individual and collective performance during a combat training exercise (Morrison and Meliza 1999). The AAR typically occurs several hours after the training event when the participating soldiers reconvene in a specified location to review the training exercise with their commanding officer and training coordinator. During the AAR, the units' individual and collective performance is reviewed by answering the following questions:

1. What happened during the exercise?
    - Participants specify the actions and outcomes of the simulated battle.
2. Why did it happen?
    - Participants explain the important actions and outcomes.
3. How can the unit improve their performance?
    - Participants determine appropriate actions to solve problems identified in their performance (Morrison and Meliza 1999).

Prior to 2006, this type of training was only available at a few select NTCs, and most National Guard units did not have the resources to attend (Milbrodt 2009). The XCTC training allows for National Guard units to train at any military installation area, where training scenarios can be adjusted according to the terrain and environmental conditions of the selected geographic location. A typical day of training during an XCTC exercise includes several designated units, typically squads and platoons (as identified in Table 1) that depart from a staging area known as the Forward Operating Base (FOB).

Table 1: Army unit size

Unit	**Approximate Personnel Size**	**Composition**	**Typical Commander**
Army	100,000	2+ corps	General
Corps	30,000+	2+ divisions	Lt. General
Division	15,000+	3+ brigades	Maj. General
Brigade	4,500+	3+ regiments	Brig. General
Regiment	1,500+	2+ battalions	Colonel
Battalion	700	4+ companies	Lt. Colonel
Company	175	4 platoons	Captain
Platoon	40	4 squads	Lieutenant
Squad	10		Staff Sergeant

Source: http://www.mirecc.va.gov/docs/visn6/8_us_military_unit_size.pdf

The units moves through designated routes and are engaged by opposing forces that are also given specific attack roles by the training officer. An objective or target is given to the units as they move through a route to a village. An objective during a training scenario might include the following: The unit must travel from the FOB to village "x" where they are to meet the mayor of the town (played by a participating role player). The unit receives intelligence that an explosive device in located in a specific target building that they must locate and sequester. During the exercise scenario, opposing force participants are given the task of engaging the unit in which potential gun fire ensues, while role players acting as civilians are in close proximity as well. How the individual soldiers in the various units respond to different attack situations is reviewed during the AAR.

The field exercises are developed by the commanding officers in conjunction with the US Army Training and Doctrine Command (TRADOC). TRADOC is charged with overseeing training of Army forces and the development of operational doctrine. When TRADOC was developed in the 1960s, it employed the Systems Approach to Training (SAT) that only focused on the behavior and skills soldiers need in order to successfully perform a task (Perez et al., 1992). With the deployment of the XCTC program, ARNG training now employs technology so soldiers can visualize and see the training in a virtual environment. While virtual technology has been used to benefit military training (Anonymous 2011), geospatial analysis has not been used to evaluate the data collected during these XCTC exercises.

1.3 Performance Reviews

The importance of enhanced training capabilities is highlighted by the statistics of US military fatalities in Iraq and Afghanistan over the past decade. During the first two years of war in Iraq, the US suffered more than 13,000 casualties, both fatal and non-fatal (Kutler 2005). While research has been conducted looking at physical geography and other environmental factors to explain how daily weather impact combat fatalities (Swann 1999), research is still lacking on how training techniques impact combat fatalities. Environmental factors can be used to explain only a certain percentage of combat casualties: therefore, human error and lack of preparedness should be analyzed to gauge their impact on wartime fatalities.

The US Army's Program Executive Office for Simulation, Training, Research, and Instrumentation (PEO STRI) has grown significantly in an attempt to develop training to mitigate combat casualties. Their budget increased from $744 million in 2000 to $1.767 billion in 2005 (Kemp 2006) in an effort to enhance combat training. The difficulty lies in being able to quantify where and why human errors are occurring that may lead to unintended deaths, both military and civilian. While soldiers are being immersed in a simulated combat environment, their actions and engagements can be used to identify factors on the battlefield that could be causing additional fatalities or near misses. Spatial analytics may help to improve these simulated training programs that continue to proliferate in the armed forces. Typically, the XCTC exercise is the last training exercise a soldier will encounter prior to deployment and therefore, some analysis of the effectiveness of this training on a soldier's performance should be undertaken. Such analysis is currently lacking in the XCTC program, making it difficult to augment the training according to individual battalion unit needs.

1.4 Spatial Analysis in Military Training

Information gleaned from geospatial analysis of the XCTC training data could help the officials in charge of developing the daily training agenda to determine which scenarios, if any, require additional iterations to improve training tactics and soldier readiness. Throughout the day during the training exercise, soldiers are moving through the installation area on roads and through mock villages. Attack-events are set up to engage the soldiers in battle. If it could be determined that a statistically significant occurrence of unintended casualties or engagements occur in certain environments or with identifiable circumstances, the commanding officers might then be able to adjust training to improve the unit's performance.

The results of this analysis are not intended to explain why more engagements occur at specific locations, but rather to identify where statistical hot spots of engagements occur. Officials can then use this information during an AAR to determine what factors might cause a greater fatality rate at certain locations and require more training iterations using similar scenarios to give soldiers more practice and experience in an effort to mitigate such fatalities during actual warfare.

Spatial analysis can be used to augment military training and adjust for the individual needs of the battalion units. An individual unit may be required to participate in additional training iterations around a large mock village that has many participants in the area due to high occurrence of errors there, while another may require additional training on the lanes with only a few people, en route to a location when attacked by opposing forces. Ultimately, the goal is to improve the survival rate of soldiers at war and to mitigate unintended casualties of both soldiers and civilians on the battlefield.

Using a series of GIS and Microsoft Access tools as well as Structured Query Language (SQL), this study aimed to identify locations with a statistically higher occurrence of various types of engagements during a military training exercise during the course of one day. The types of engagements that were considered in this study included: total engagements and red-on-blue force, which refers to a member of the opposing force (i.e. red force) engaging a member of the blue force (American force). For the purposes of this military training exercise, an engagement is considered a kill or a near miss, which is defined as a non-lethal hit. The objective of this study was to explore hypotheses about the occurrence of statistical hot spots of casualties during military combat training.

Three hypotheses were explored as follows:

1. Statistical hot spots of total engagements and red-on-blue engagements can be identified along routes using off-the-shelf spatial analysis tools.
2. The hot spots for total engagements and red on blue engagements occur in the same locations or at least similar types of locations.
3. An increase in the number of engagements in a grid cell is directly correlated to an increase in the number of unique participants passing through that grid cell.

1.5 Organization of the Thesis

The remainder of this thesis is divided into four chapters. Chapter 2 describes previous GIS work in military applications as well as current trends in training data analysis. Chapter 3 describes the methods and data sources used for the hot spot analysis. Chapter 4 presents the results of the analysis. Chapter 5 summarizes the major findings and considers how the results might be used to improve daily training performance.

CHAPTER 2: RELATED WORK

The majority of GIS work in military applications has historically been centered on cartographic visualization and more recently, on the environmental management of military bases. The use of GIS during wartime efforts is focused mainly on developing situational awareness for the soldiers on the ground in active combat as well as advancing remote sensing technologies to improve imagery-based target detection. The spatial analysis of simulated field training exercises is undertaken sparingly and there is little in the way of academic research on the subject.

2.1 History of GIS Work in Military Applications

The use of geotechnology in the military has been a significant contributor to military effectiveness since before World War I, when topographic maps were used to enhance artillery effectiveness and to show both landforms and elevation through contour lines (Corson and Palka 2004). World Wars I and II both prompted a number of advances in cartography and remote sensing. For the first time, topographic mapping was utilized as a tool for intelligence, planning, movement, logistics resupply, artillery bombardment and command and control (Corson and Palka 2004). While the Cold War era introduced a number of other innovations, including laser range finding, digital mapping and GPS, it was not until the war in Afghanistan that several geotechnologies developed by private corporations and defense contractors were deployed **(**Corson and Palka 2004**)**. During this time, virtual 3D maps as well as traditional paper maps were used to help plan tactical missions.

2.2 Analysis of Soldier Performance in Military Training

While current applications of GIS in the military have centered primarily on cartographic production and battlefield terrain analysis, research has been conducted using statistical models looking at the relationship between individual soldier performance and their impact on group outcomes (Semmens 2013). For example, Semmens (2013) developed a statistical model to show that marksmanship accuracy and firefight outcome are positively correlated and that from the standpoint of readiness, it is useful to track a soldiers' weapon accuracy. This study suggests that weapon accuracy can continuously be improved upon, providing better battle readiness and support and mitigating civilian casualties from target misses.

Although statistical models such as these can speak to soldier performance as it relates to weapon accuracy, overall soldier performance must also be looked at as a function of situational awareness. Recent technological developments by for profit companies and government contractors have been used to better understand the environmental factors in which the soldier is fighting while using the soldier's location and status to improve situational awareness during a training exercise (Copley and Wagner 2008). A system using situational awareness to improve combat training performance was implemented at Fort Benning's McKenna Military Operations on Urban Terrain (MOUT) site. This system used GIS, GPS, mesh-networking and radio-frequency identification (RFID) to track soldiers both in and outdoors while engagements were captured in full motion video. Soldier performance was then critiqued during the AAR. As the leading engineers who helped develop this system explain:

> The system is used to give commanders and their counterparts a tool for improving situational awareness during exercises and experimentation. Not only can they see their units, status and movements, they can track live fire, soldiers, vehicles, other entities and assets in real time. They can easily distinguish BLUEFOR [blue force or domestic soldiers], REDFOR [red force or opposing forces] and gain insight on mission progress. They can locate 'Hazard' areas, see live video and move to viewpoints within the 3D application (Copley and Wagner 2008, p. 4).

2.3 Methodologies

While various methodologies can be used for analysis of combat training data, three distinct ones were selected for this thesis: Choropleth mapping, hot spot analysis, and time-interval animation. Each of these methodologies is widely used across sectors; however, implementing them for combat training analysis is a novel application in all three instances. Choropleth maps are one of the most common thematic mapping techniques used today (Indie Mapper, 2014), and visualize how a measurement varies across a geographic area. Choropleth mapping continues to proliferate because the majority of geodata is now reported by enumeration units (such as census data). As a result, professionals are now accustomed to thinking about the world as divided into spatial units defined by boundaries. Because of this, choropleth mapping is commonly used to visualize census data, map percentages of a certain event, or show percentage changes over time.

Hot spot analysis identifies statistically significant clusters of high values of a particular phenomenon. Hot spot analysis is commonly used in the medical field looking for areas of statistically significant rates of a specific disease. This method of analysis is also commonly used in law enforcement identifying locations of high rates of crimes.

Time-lapsed animations are used to visualize a particular event over time and space. Animations are most commonly used to show the path and movement of hurricanes or forest fires.

These three methods of analysis all have appropriate application with a dataset provided by a combat training event. Because these methods of analysis are novel for the unique nature of the dataset discussed in this thesis, the results can fill a void that is currently lacking in this field.

Considering the financial and political investment in preparing US soldiers for combat, the lack of research on the usefulness of spatial analytics in analyzing the scenario-based training exercises is striking. The US military has an enormous potential to leverage spatial analysis for battlefield use at various scales (Corson and Palka 2004). While not entirely a comprehensive analysis of all aspects of training, this study demonstrates the possibility that spatial analytics can be used as a tool during military training events and provides a framework for such analysis. As Vince Lombardi once stated, “Practice does not make perfect. Only perfect practice makes perfect.” This concept underpins the attempt in this study to use GIS as a tool for understanding combat training events, and ultimately helping soldiers ‘practice’ better.

CHAPTER 3: METHODS AND DATA SOURCES

This study used GIS tools and Microsoft Access SQL to explore the relationship between GPS tracking data of individual participants and engagements during a training event. This initial investigation focused on one day of training during a three week-long exercise. The objective of this study was to explore several hypotheses surrounding the location of potential engagement hot spots through the following three questions:

1. Is there a presence of statistical hot spots along the linear network for engagements representing the total number of casualties and blue force casualties?
2. Are the hot spots for all engagements the same as the hot spots for red-on-blue force engagements?
3. Do the number of engagements increase as the number of unique participants in a grid cell increases?

Three methodologies were chosen to analyze and visualize the data: choropleth mapping, hot route analysis and time-lapse animation. The diagram in Figure 1 shows the workflow and how these methods were coupled or linked with one another.

A single day of training was chosen as the unit of analysis due to the size and complexity of the source dataset. There were 1,139 participants involved in the training. Each participant was equipped with a personal Enhanced Dismount Instrumentation (EDI) pack recording its GPS location every three seconds. This resulted in the size of the whole dataset for the entire three-week training period being too large for processing. The whole dataset would have required a larger and more sophisticated hardware platform than was available for this analysis; therefore, one day of training was extracted for analysis.

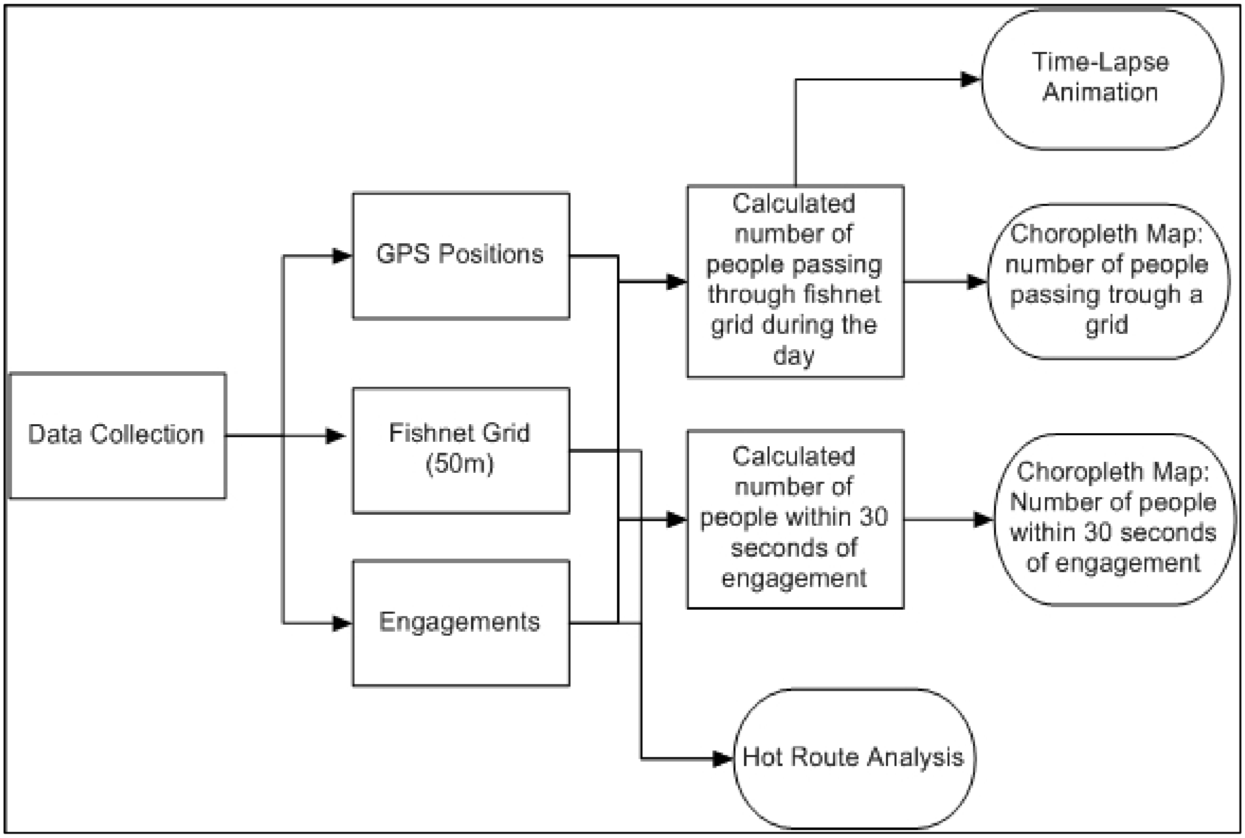

Figure 1: Schematic showing methods and data used

3.1 Data Collection

Training occurs along the road network and around mock villages that range in size from small hut villages to large, city-like towns. A small village might include 4-6 structures built out of conexes, which are repurposed shipping containers made of corrugated steel. The individual structures might include doublewide conexes or stacked conexes with stairs built in. A training area might also include a town-like area built up of many large cinder block structures representing hotels, banks, schools and hospitals. The purpose of these villages is to simulate what a town might look like in the country of deployment. During a training exercise, role players, as seen in Figure 2, are used to represent civilians on the battlefield (COB) as well as the opposing forces.

Figure 2: Role player representing a COB

The soldiers representing various combat training units, as pictured in Figure 3, are equipped with the EDI GPS pack that is connected to a MILES unit which records and logs when a soldier is shot and who shot him/her. The whole training exercise is meant to mimic a day in country during deployment and prepare soldiers for the threats that occur during wartime on a daily basis.

Figure 3: Soldier equipped with EDI pack

The data from one day of training on 7th August 2013 during an XCTC exercise at Fort Drum, NY was obtained and used for this thesis project. Before a day's training begins, all participating soldiers are given a role that is associated with a color, either blue force for the American military forces, or red for the opposing force. Role players are placed on the field as well representing COBs. Each person in the field of play during the day's exercise is equipped with an EDI device that is linked to their MILES unit that contains the individual's military identification information. The EDI device is responsible for recording and logging the latitude and longitude at 3-second intervals and has a horizontal accuracy of 2 to 4 m. The MILES unit records if soldiers are shot and who shot them, time of day, and type of weapon used. The MILES device will also record whether the shot is considered a "catastrophic kill" or a "near miss", meaning that the person was shot but on an extremity or some other non-fatal part of the body. Because the EDI is linked to the military ID information, name, unit, and military rank are also recorded. For individuals representing a COB, the simple moniker "COB" is used to identify them. Throughout the course of the day, different training scenarios are assigned to the units that involve traveling on the roads and possibly encountering an IED or responding to a threat on the road or in a mock village.

Upon completing the day's training, all of the data from the EDI and MILES devices are saved to a log file. The data from the EDI device is saved to a separate CSV file. An algorithm written by software engineers at SRI International then aggregates and compiles the data from the EDI and MILES units into one CSV file that contains the information summarized in Table 2.

Table 2: EDI engagement log attribute data

Attribute	**Description**	**Unit of Measurement**
ENGAGEMENT_TYPE	Indicates if the engagement is Red on Blue, Blue on Red, or Blue on Purple	Text
DATE	Month/Day/Year	Date
TIME	Time of Day in 24-hour format	Time
TIME-sec	Time of day in Unix format	Time
ENGAGEMENT	Indicates if the engagement is a near miss or catastrophic kill	Text
WEAPON	Type of weapon used in engagement	Text
VICTIM_ID	ID for the EDI Issued	Number
VICTIM_PID	ID for the MILES unit issued	Number
VICTIM_CALL	Rank and Name of Victim	Text
VICTIM_FORCE	Indicates whether victim is Red, Blue or Purple	Text
LIFE_STATUS	Indicates if the victim is alive or dead	Text
VIC_LAT	Latitude of victim	Decimal Degrees
VIC_LONG	Longitude of victim	Decimal Degrees
VICTIM_ALT	Altitude of victim	Meters
SHOT_RANGE	Distance between victim and shooter	Meters
SHOOTER	Shooter EDI ID	Text
SHOOTER_CALL	Shooter rank and name	Text
SHOOTER_FORCE	Indicates if shooter is Red, Blue or Purple force	Text
SHOOTER_LAT	Shooter latitude	Decimal Degrees
SHOOTER_LON	Shooter Longitude	Decimal Degrees
SHOOTER_ALT	Shooter Altitude	Meters

3.2 Description of Study Area

The Fort Drum military base is located in the northern part of the State of New York near the Canadian border. The boundary of the training area used on 8/7/13 was approximately 37 km by 26 km. The total length of the roads at Fort Drum measures 1,029.45 km and site itself covers a total of 434 square km. The elevation ranges from 52.8 to 646.9 m. The training site also includes an Impact Area where live fire ordnance will detonate. This area measures 81.8 km^2. Movement in this area was limited during training due to live fire simulations that occur daily. Only six positions were recorded in this area during the day as part of the training exercise. During the training exercise 14 mock village locations were used with a total of 141 buildings spread out between them. These villages range from conex or small wooden structures to a complex Military Operations on Urban Terrain (MOUT) site, pictured in Figure 4.

Figure 4: Fort Drum MOUT site

Figure 5 shows the study area including the impact area, roads and base boundary as well as all of the GPS locations recorded from the EDI devices. The training throughout the day occurs on roads en route to the training villages. Many of the roads in and around the villages are small dirt roads leading through forested areas. Of all the roads in the Fort Drum training facility, 9.8% are primary paved roads, 38.7% are secondary paved roads, 30.7% are unpaved but maintained roads and 20.8% are unpaved and not maintained roads. The vast majority of movement on the field during the training day occurred on unpaved maintained roads. There are approximately 301 square km of forestland on Fort Drum (Niver 2009). Most of the training occurred in and around deciduous forests. A summary of the distribution of vegetation is listed in Table 3.

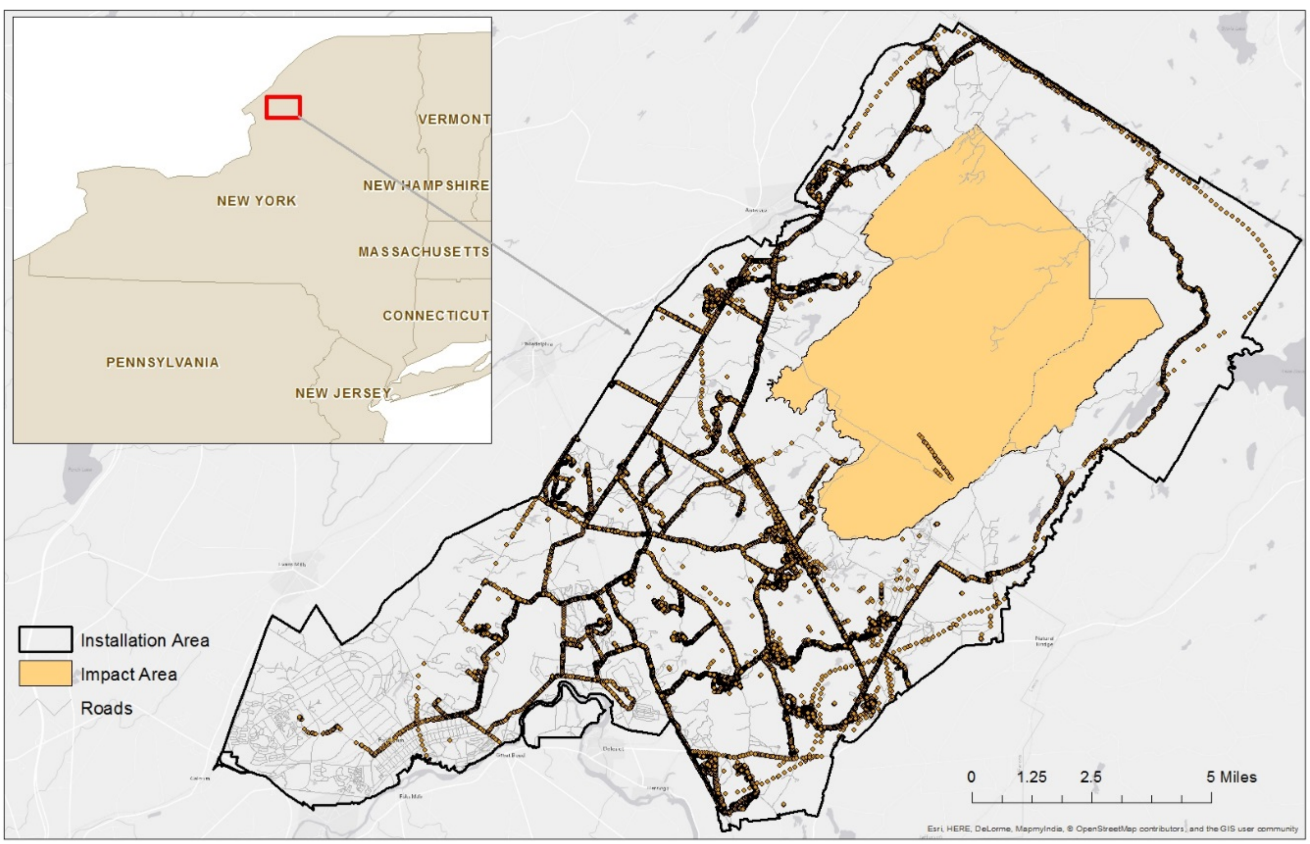

Figure 5: Fort Drum study area with EDI GPS recordings

Table 3: Distribution of vegetation at Fort Drum, NY

Vegetation Type	Area (km^2)
Built Up	2
Conifer Forest, Closed Canopy, Pine	28
Conifer Forest, Open Canopy, Pine	3
Conifer Plantation, Closed Canopy	5
Deciduous Forest, Closed Canopy	117
Deciduous Forest, Open Canopy	38
Developed Road	2
Disturbed/Developed	12
Grassland, Medium-Tall Bunch	11
Grassland, Medium-Tall Sod	2
Grassland, Medium-Tall, Sparse Deciduous Shrubs	20
Grassland, Medium-tall, Sparse Deciduous Trees	1
Grassland, Medium-tall, Sparse Trees	1
Grassland, Short Bunch	3
Grassland, Short, Sparse Conifer Trees	1
Grassland, Short, Sparse Deciduous Trees	3
Lacustrine Wetland	3
Landscaped Yard	12
Mixed Forest, Closed Canopy	80
Mixed Forest, Open Canopy	9
Palustrine Wetland, Conifer Forest, Closed Canopy	2
Palustrine Wetland, Conifer Forest, Open Canopy	1
Palustrine Wetland, Deciduous Forest, Closed Canopy	10
Palustrine Wetland, Deciduous Forest, Open Canopy	10
Palustrine Wetland, Deciduous Shrub land	9
Palustrine Wetland, Grassland	5
Palustrine Wetland, Grassland, Sparse Shrubs	4
Palustrine Wetland, Hydromorphic Vegetation	2
Palustrine Wetland, Mixed Forest, Closed Canopy	5
Palustrine Wetland, Mixed Forest, Open Canopy	3
Palustrine Wetland, Mixed Shrub land	1
Palustrine Wetland, Small Drainage	11
Rangeland, Forbs	3
Riverine Wetland	3
Rocky Area, Sparse Grasses	9
Sand Dunes, Sparse Grasses	1
Shrub land, Deciduous	15
TOTAL	447

3.3 Data Preparation

On 7th August 2013, there were 1,139 participants on the training field. Training began at 9:13:07 GMT on 7th August 2013 and was completed at 3:35:17 GMT on 8th August 2013, spanning a total of 18 hours and 22 minutes. The location of every individual as well as every vehicle was recorded approximately every 3 seconds, leading to a file with 5,940,971 records. Due to technological problems on the EDI device, occasionally the data collected had faulty GPS information that placed the individual and/or vehicle carrying them erroneously outside of the Fort Drum installation area. To correct for this, the data records that did not fall within the Fort Drum boundary were removed. This process reduced the total number of EDI GPS records from 5,940,971 to 5,837,881, removing approximately 1.1% of the EDI data. There were 418 engagements that resulted in catastrophic kills or near misses recorded by the MILES devices on 7th August 2013. Of these engagements, 212 also properly recorded complete shooter information including latitude and longitude, which meant that 50.1% of the engagement data lacked the expected shooter information. Due to the incomplete shooter information, analysis involving proximity between victim and shooter was not included in this study.

During the exercise, the soldiers' movements were limited mostly to the roads and villages. To represent every area in which there was movement during the day, a fishnet (grid cell) measuring 50 m on a side was created and the grids that intersected with a GPS recording were saved to the output file. The resulting fishnet used 9,031 grid cells to summarize all of the GPS recorded movements throughout the day. Figure 6 shows a sample of this fishnet with the GPS coordinates recorded at 3 second intervals overlaid on the individual grid cells.

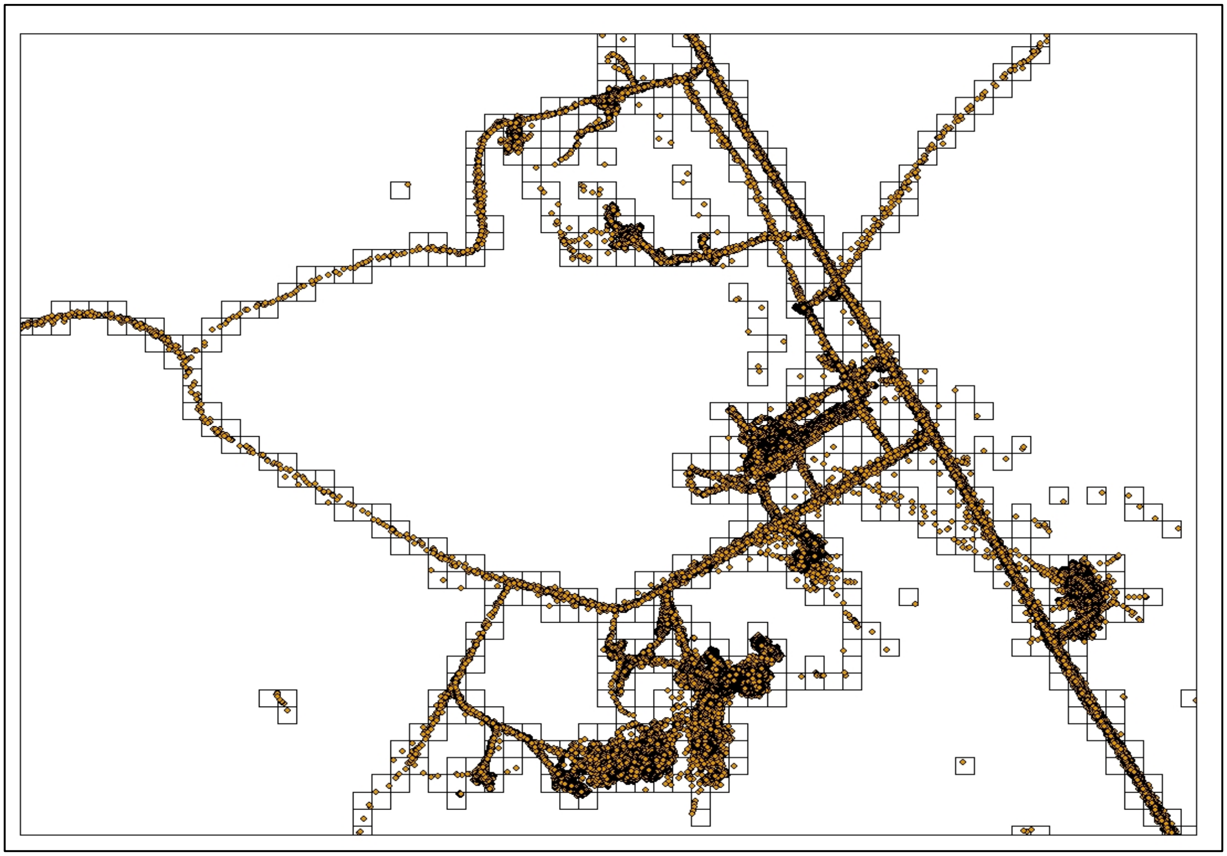

Figure 6: Fishnet with intersecting EDI coordinates at 3-second intervals

3.3.1 Timestamp Analysis

The time recorded on the EDI devices is logged in EPOCH UNIX format. To prepare the data for analysis the EPOCH time had to be converted into GMT format. This was achieved with the following calculation:

$$GMT\_TIME = (EPOCH\_TIME/86400) + 25569 \quad (1)$$

For analysis purposes the GMT time was then rounded to the nearest 15-minute interval. This aggregation was used to determine the total number of people passing through a cell in each 15-minute interval throughout the day.

3.4 Choropleth Mapping Analysis

A choropleth map is a thematic map in which areas are distinctly colored to represent classed values of a particular phenomenon (Esri 2014b). A choropleth map is appropriate when values of a phenomenon change abruptly or when the reader should focus on 'typical' values for individual enumeration units (Slocum et al. 2009). Choropleth mapping was used in the study to demonstrate how one grid cell on the training area compares with another grid cell in terms of participant movement and density.

3.4.1 Counting the Numbers of Unique Individuals Passing through Each Grid Cell

A choropleth map was constructed to visualize the number of unique individuals that passed through each grid cell during the day. Considering the data consisted of every participant's location recorded every 3 seconds, a simple intersect of the GPS data could not be done. The result of performing this operation would have indicated that each cell had hundreds of participants. To correct for this, the feature class containing all the GPS recordings for every participant was joined to the fishnet grid cell feature class. This resulted in a new point feature class containing all the GPS locations recorded every 3 seconds for each individual with an additional attribute of the grid cell ID that the GPS recording fell in. This table was then imported into Microsoft Excel 2013. Using the Pivot Table tools in Excel, a new table was created showing each fishnet grid cell ID in one column and in the adjacent column, the Distinct Count of each participant ID for the corresponding grid ID was summed. Using this pivot table method, the Distinct Count for each participant in a grid cell was also determined for each force type: Blue Force, Red Force and COB. The resulting Excel table was imported back into ArcMap and joined to the fishnet grid cell feature class. The new grid cell feature class now contained an attribute summarizing the total number of unique participants that were recorded in each cell during the day.

The Equal Interval classification method with five break values was used because it emphasizes the amount of an attribute relative to other values (Esri 2014a). For the purposes of this study, the goal was to emphasize the value of each grid cell relative to all the other grid cells in terms of foot traffic.

3.4.2 Counting the Average Number of People in Cells within 30 seconds of an Engagement

To determine the average number of people in a grid cell within 30 seconds of an engagement, the data was imported into Microsoft Access 2013 in order to utilize the available Structured Query Language (SQL) library within Access. For each grid cell, the average number of unique persons that were present within 30 seconds of a kill was calculated using the following equation:

$$K = \sum_{i=1}^{x} Pn \qquad (2)$$

where $\boldsymbol{x}$ = number of engagements in the grid cell, P*n* = number of unique persons per grid square *n*, and *k* = average number of unique persons per engagement per grid cell. This equation calculates the average number of unique individuals per engagement in a grid cell.

An SQL query was used to determine the average number of people that were in a grid cell within 30 seconds of an engagement by dividing the total number of engagements that occurred a grid cell by the average number of people in the same grid cell within 30 seconds (the *k* –value calculated above). The feature class representing the GPS recordings and the associated fishnet grid cell ID were imported into Microsoft Access 2013 and a series of three SQL statements were used to determine the average number of people that were in a 250 m^2 grid cell within 30 seconds of an engagement. The resulting table was imported back into ArcGIS and joined to the fishnet grid cell feature class based on grid cell ID so that each grid cell contained the *k* value. A thematic map was then created using the Equal Interval classification with five break values.

3.5 Hot Spots on a Linear Network

For a day's training event, the identification of hot spots for engagements is beneficial for locating where statistical clusters of near misses or kills are occurring in relation to the total number of events and movement during the day. The challenge with a dataset like the one used for this thesis project is that the data is distributed along a linear network. This poses several challenges. Most hot spot analyses assume events are situated across a homogenous environment or a continuous Euclidean space where events can occur at any point (Tompson, Partridge, and Shepherd 2009). The Euclidean distance measurements take the shortest path between points and use the hypotenuse and the minimum bounding rectangle to identify hot spots. This methodology is inappropriate for the dataset used here because the points are constrained by a linear network (mainly roads) such that the engagements occur along a one-dimensional subset of Euclidean space (Miller 1994). To compensate for this, a methodology for determining hot spots developed by Tompson, Partridge, and Shepherd (2009) was used to determine hot spots of engagements along a linear network. The methodology entitled Hot Routes is better suited for determining linear clusters of events compared to other techniques because it provides a more localized view of incidents which is better suited for small-scale analysis (Tompson, Partridge, and Shepherd 2009).

The movement throughout the day by the participants was mainly on a road network, and therefore the road centerline shapefile was used as the linear network layer. This network layer represents a collection of varying length line segments that are delineated by nodes. The varying lengths influence the precision when calculating the Hot Routes (e.g. longer lines will have a greater chance of events occurring on them than shorter lines). To reduce the impact of this problem, a fishnet of 250 m^2 was created, as seen in Figure 7, and used to split the road network layer so that a road segment spanning three grid cells would be split into three separate features based on the grid cell it was located in.

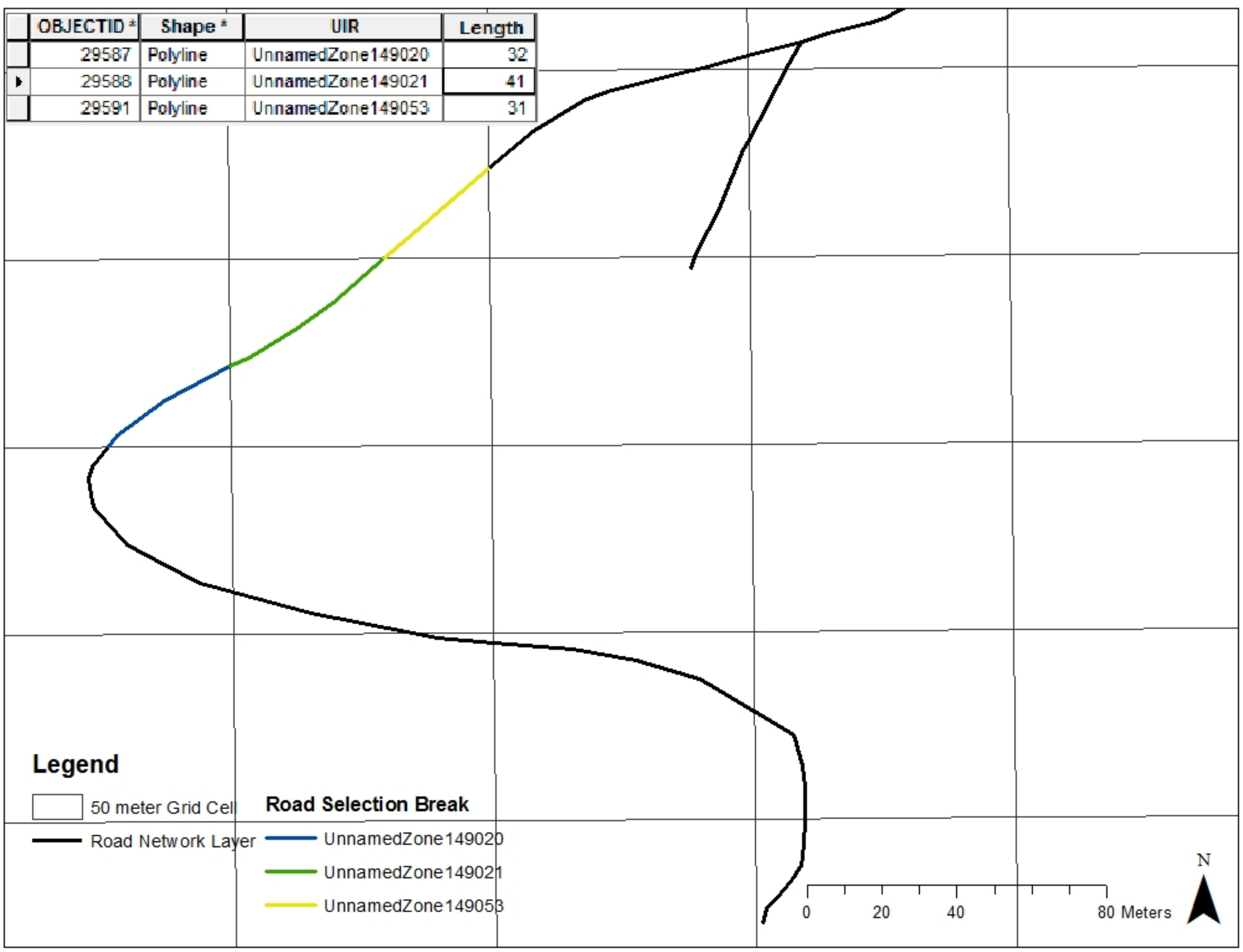

Figure 7: A road spanning three grid cells, with road network layer in black and 250 m^2 grid cells in grey

Each road segment within a grid cell was given a Unique Identifying Record (UIR), therefor, each road segment within a cell that contains multiple road segments contains a unique name as is the case in the third cell shown in Figure 7. This results in some grid cells having multiple roads, each with a UIR associated with it. Occasionally, a road was split within the same grid cell due to the nature of the original road layer. To fix this, the roads were merged together using the Dissolve tool to merge road segments of the same name within the same grid cell. This workaround ensured that each road name within a grid cell had one feature. The length in meters of each road segment was then calculated.

A UIR was created by concatenating the road name with the grid cell zone ID number. The road network layer was then spatially joined to the engagement point layer based on closest proximity so that each engagement was given the UIR and length of the road segment it fell closest to.

In the engagement point layer, a new field was added in the attribute table titled “Event Rate per Meter” that was used to calculate the weighted rate of engagements per linear meter. This was calculated by taking the value of 1 and dividing it by the length of the line segment that the point occurred on. For example, if an engagement occurred on a line that was 10 m long, the “Event Rate per Meter” would be 1/10 or 0.1—reflecting the rate per meter of events (Tompson, Partridge, and Shepherd 2009).

A field was then added to the road network layer with the sum of the event rate per meter field from the engagement point layer for each UIR. This value, reproduced in Figure 8, represents the hot spots along the linear route with various colors used to signify varying rates and red used to indicate a high rate of incidents. This methodology calculated the rate of engagements per linear meter of a road feature in order to represent the risk distribution along a linear network.

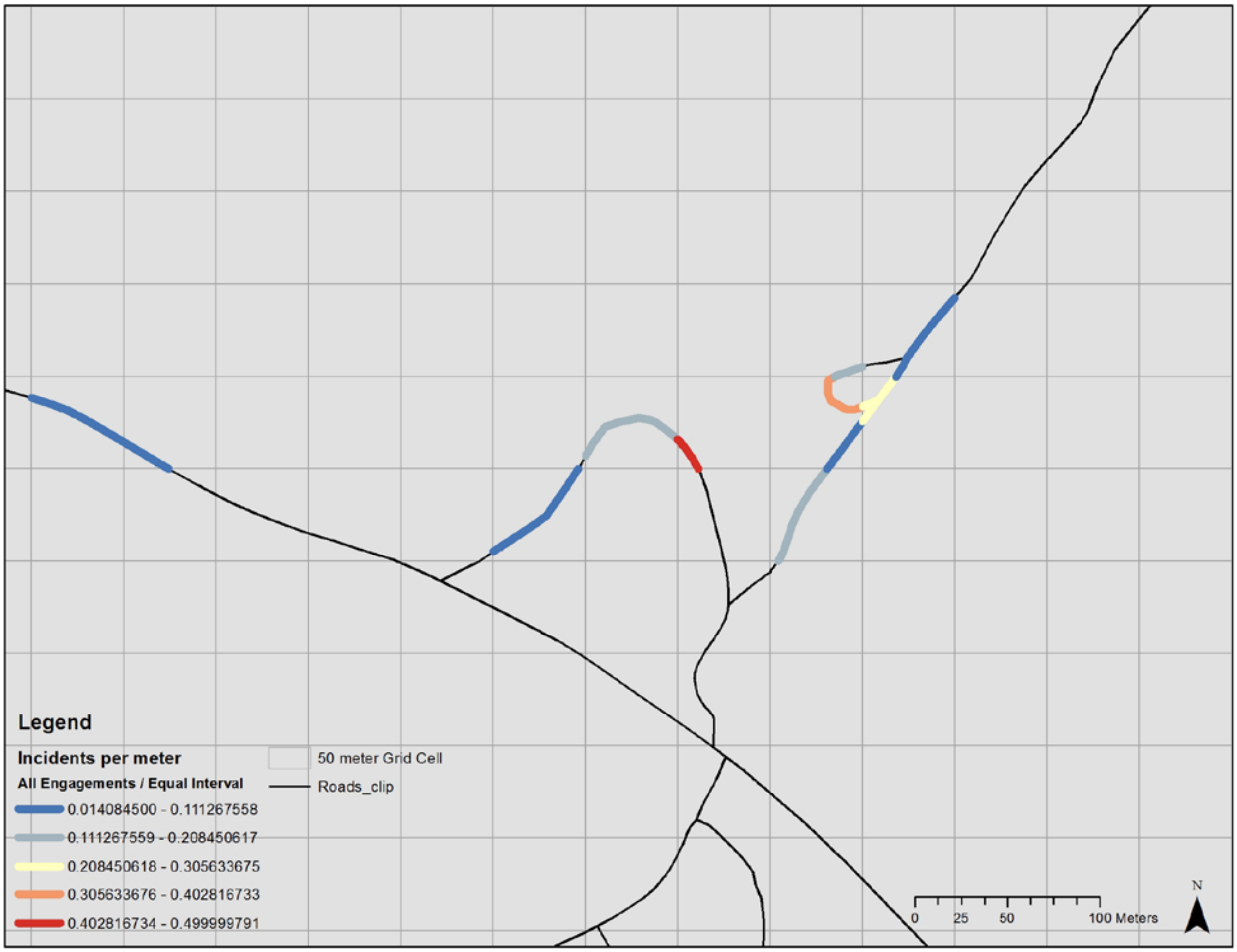

Figure 8: Map showing Hot Routes based on the rate of incidents per linear meter for ALL events

The same analysis was completed, but only for Blue Force engagements along the routes as well. Figure 9 shows the hot routes for the Blue Force engagements in the same area as Figure 8. Comparing the results in Figures 8 and 9, there is one area that differed in terms of hot spot delineation. The area on the looping road is a hot spot when all engagements are taken into consideration, but not a hot spot for Blue Force engagements. It is important to distinguish a hot spot for all engagements from a hot spot for blue force engagements because the purpose of the training is to mitigate engagements targeting the American forces. Engagements on the red force would not necessarily be considered erroneous, although they are still worthy of noting during the AAR for review and training purposes.

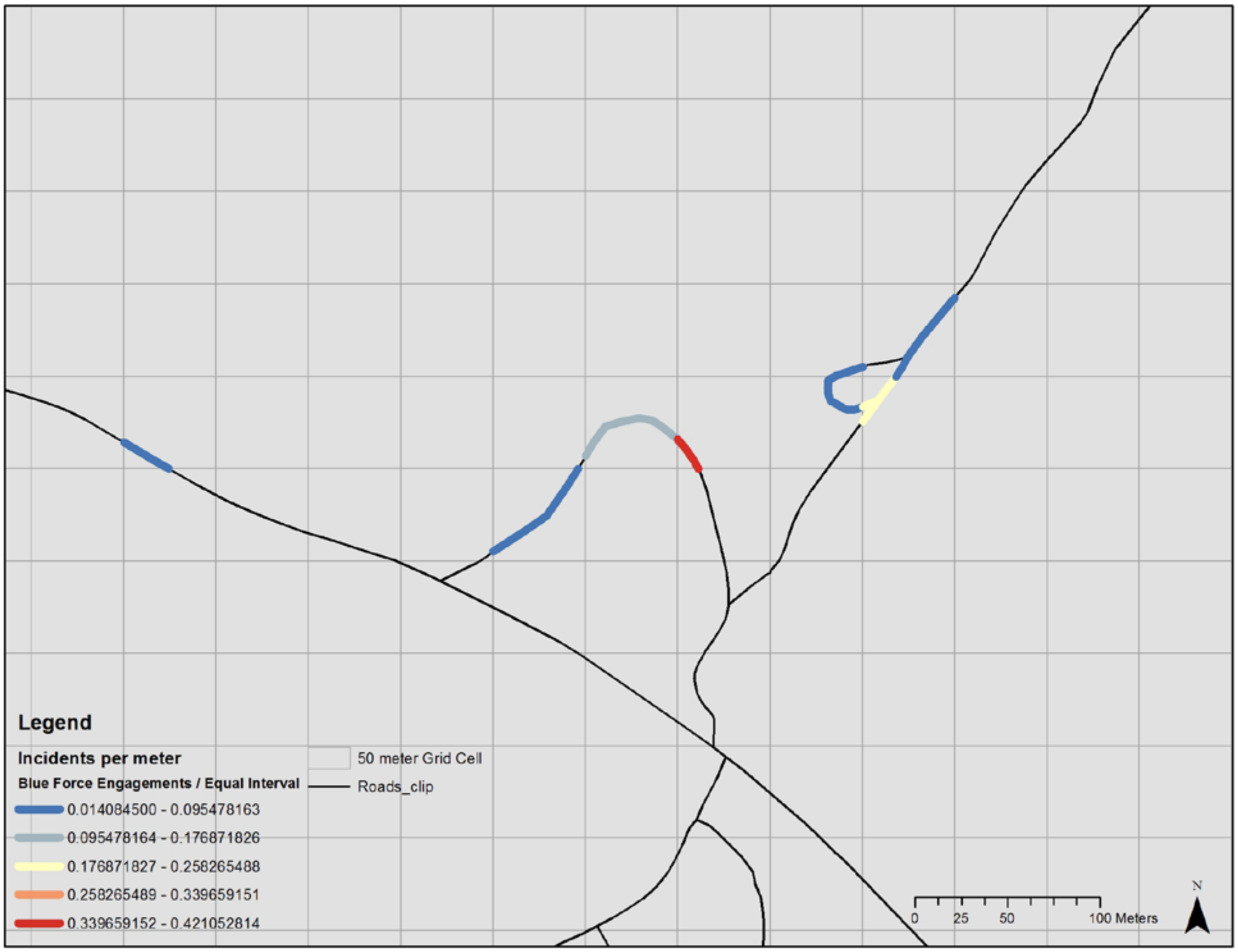

Figure 9: Map showing Hot Routes for blue force engagements

3.6 Time Interval Analysis

While the previous analysis informs us how many people are moving throughout the area during the whole day, it is also important to understand the timeframe in which the movement occurred. A table was created representing the unique number of participants in every grid cell at 15 minute intervals throughout the day. First, each GPS point was spatially joined to the grid cell layer. Then, every GPS point for every individual was rounded to the nearest 15 minute interval and unique participant IDs were counted for each time interval for each cell. A sample from the resulting table is shown in Table 4.

Table 4: Part of table showing number of unique EDI IDs counted in each cell at 15 minute intervals

Grid cell ID	Rounded (15 min) Time Interval (GMT)	Count of Unique EDI ID
Zone100000	14:00	1
Zone100000	14:30	1
Zone100000	15:15	2
Zone100011	15:00	1
Zone100012	14:00	3
Zone100012	14:15	2
Zone100012	14:30	3
Zone100012	14:45	2
Zone100012	15:00	2
Zone100012	15:15	2
Zone100023	15:30	1
Zone100032	14:00	1
Zone100032	14:30	1
Zone100032	15:15	2
Zone100037	15:15	1
Zone100038	14:00	1
Zone100038	15:15	2
Zone100039	14:00	1
Zone100039	14:30	1
Zone100039	15:15	1
Zone100040	14:00	1
Zone100040	14:30	1
Zone100040	15:15	2
Zone100041	14:00	1
Zone100041	14:30	1
Zone100041	15:15	2

The TIME_GMT field was converted into the format of MM/DD/YYYY HH:MM:SS to represent time and visualize the variability of the locations of the participants over time. Using the time field, the above analysis was able to provide a time lapse of the movement in the field throughout the day.

CHAPTER 4: RESULTS

This chapter describes the results of the choropleth, hot routes and time interval analysis for the selected day of training at Fort Drum. The Fort Drum training site shown in Figure 10 was selected due to the accessibility and permission granted to use the data by the ARNG as well as the number of engagements observed relative to other locations. This map shows 418 engagements on 7th August 2013. Most occurred south of the (live ordinance) impact area with the remaining events to the west of the impact area.

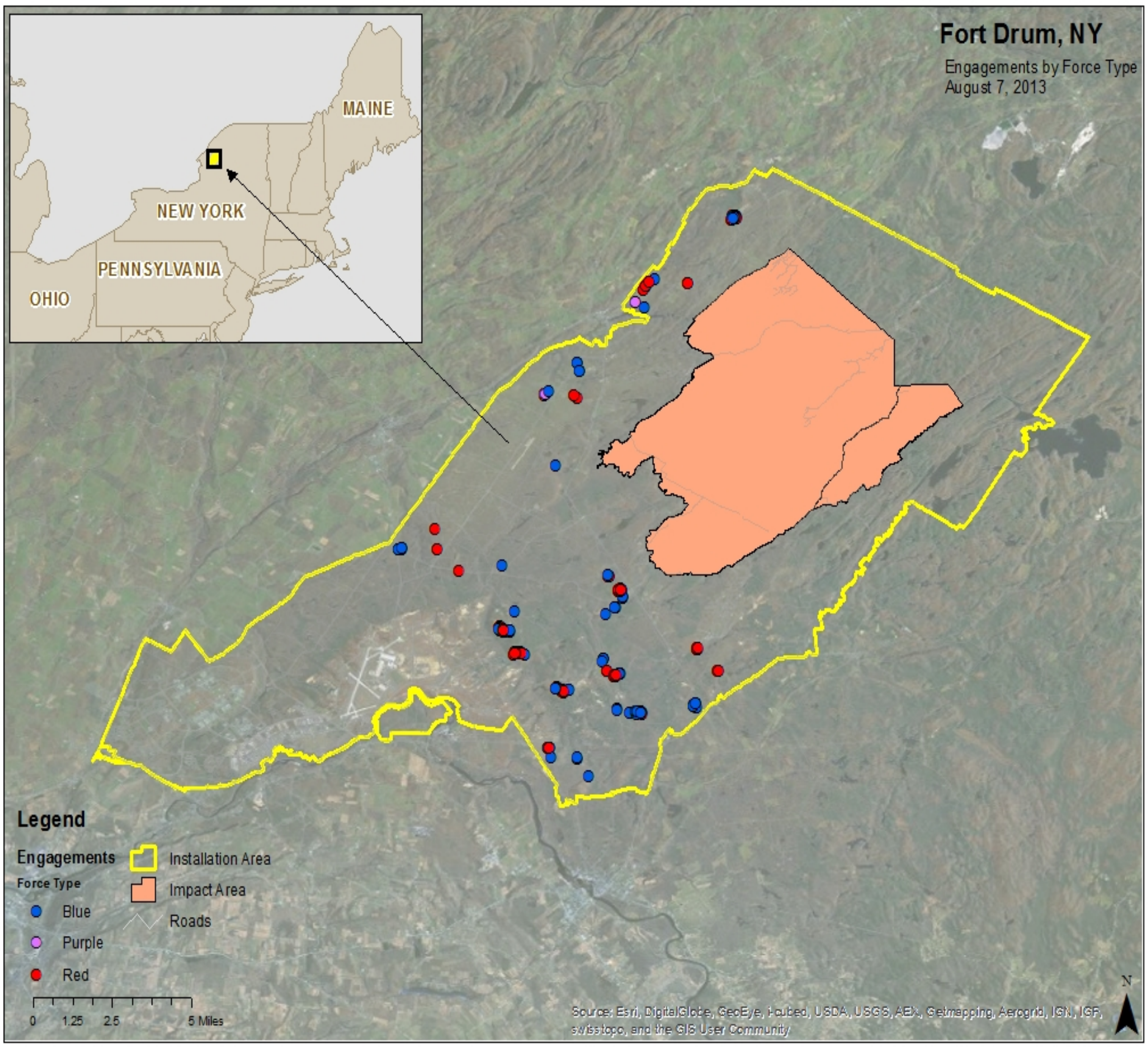

Figure 10: Map showing engagements by individual force type on 8/7/13 at Fort Drum training site.

Two choropleth maps were created showing the number of unique participants passing through a grid cell as well as the average number of participants in a grid cell within 30 seconds of an engagement. The results of this analysis indicate the inverse of what was originally hypothesized: that there is no evidence indicating that an increase in participants passing through a grid cell leads to an increase in the number of engagements. Moreover, large numbers of engagements occur in grid cells with a limited number of participants passing through. The results are similar for the analysis of participants in a grid cell within 30 seconds of an engagement. The number of engagements in a grid cell does not increase when the number of participants within 30 seconds of an engagement increases.

The hot routes analysis method developed by Thompson, Partridge, and Shepherd (2009) was used to generate areas of hot spots along a linear network. This analysis was completed twice: once for all engagements along the linear road network and once for only blue force engagements along the road network using the Equal Interval classification method. The results indicate that there is no area on the map with a greater presence of blue force engagements compared with all engagements. In addition, only a small segment of the road network could be classified as a hot route indicating a statistically higher of presence of engagements there.

The time interval analysis not only visualized the movements of the participants through each grid cell by 15 minute intervals, but also, a table and graph showing the breakdown of each participant type and each interval on the whole field was calculated. The detailed results of this analysis can be found in Appendices A and B. These analyses can be used to the locations of the largest numbers of participants on the field as a whole.

These three methods of analysis were chosen for their potential use during an AAR due to their ability to inform commanding officers of results not otherwise seen without visual analysis. The results produced with the three methods are presented in more detail in the three sections that follow.

4.1 Choropleth Maps

Two choropleth maps were created to visualize the movement of unique participants during the day on the battle field as well as the number of people in a grid cell within 30 seconds of an engagement.

A choropleth map was constructed using the Equal Interval classification method with five classes to visualize the total number of unique participants passing through a grid cell (Figure 11). The red grid cells recorded on the whole area and inset (small area) maps indicate a higher density of people passing through a grid cell during the day.

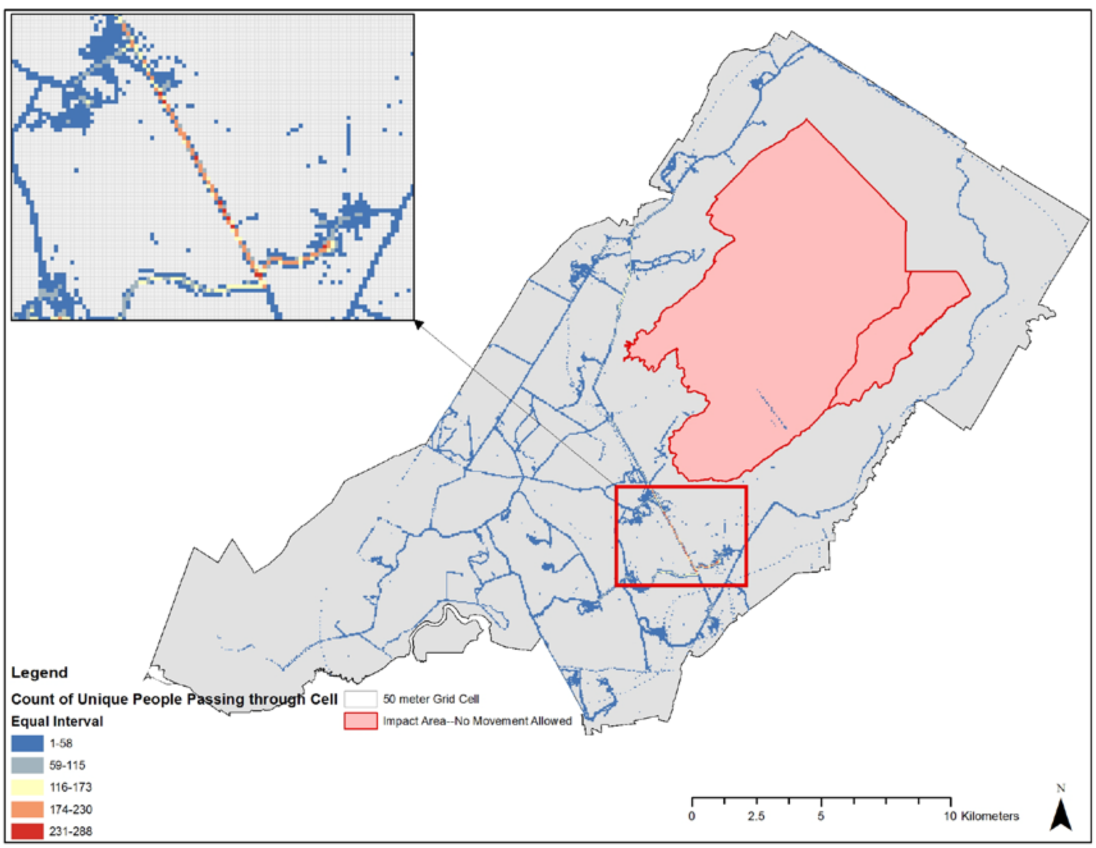

Figure 11: Choropleth map depicting total number of unique participants moving through each grid cell on 7th August, 2013

The map reproduced in Figure 11 show the number of participants traveling on the linear network throughout the day whether an engagement occurred in the cell or not. During the day, the participants mainly followed convoys on roads en route to a village location that would be either on the road or several meters off the road. The participants used several roads to the south of the impact area and one main road that circled the impact area. At the start of the day the participants would begin from their FOB, which acts as a staging area for the soldiers, the activities would end at the location where the training objective was being held (typically a village or location on a road). The map in Figure 11 show that the largest numbers of people passed through the area just to the southeast of the impact area. With a large amount of activity occurring in this area, we would expect to see more engagements along this road. This however is not the case. The stretch of road that had the greatest number of people passing through did not have any engagements. This is important for the training officers to know because they can emphasize what the soldiers did well in high density areas compared to what errors were evident in low density areas with more engagements.

The number of unique participants moving through grid cells in which at least one participant was recorded ranged from 1 to 288 with the mean being 15. Figure 12 shows the frequency distribution of the number of unique individuals in a cell regardless of the force type. The graph in Figure 12 indicates that the majority of cells had just a few participants and there were very few cells that had large numbers of individual participants. One hundred and twenty seven engagements occurred in cells with 15 people or less – which means that approximately 30% of the engagements occurred in cells with less than 15 people passing through them, which indicates that small numbers of people passing through grid cells is correlated with the occurrence of one or more engagements.

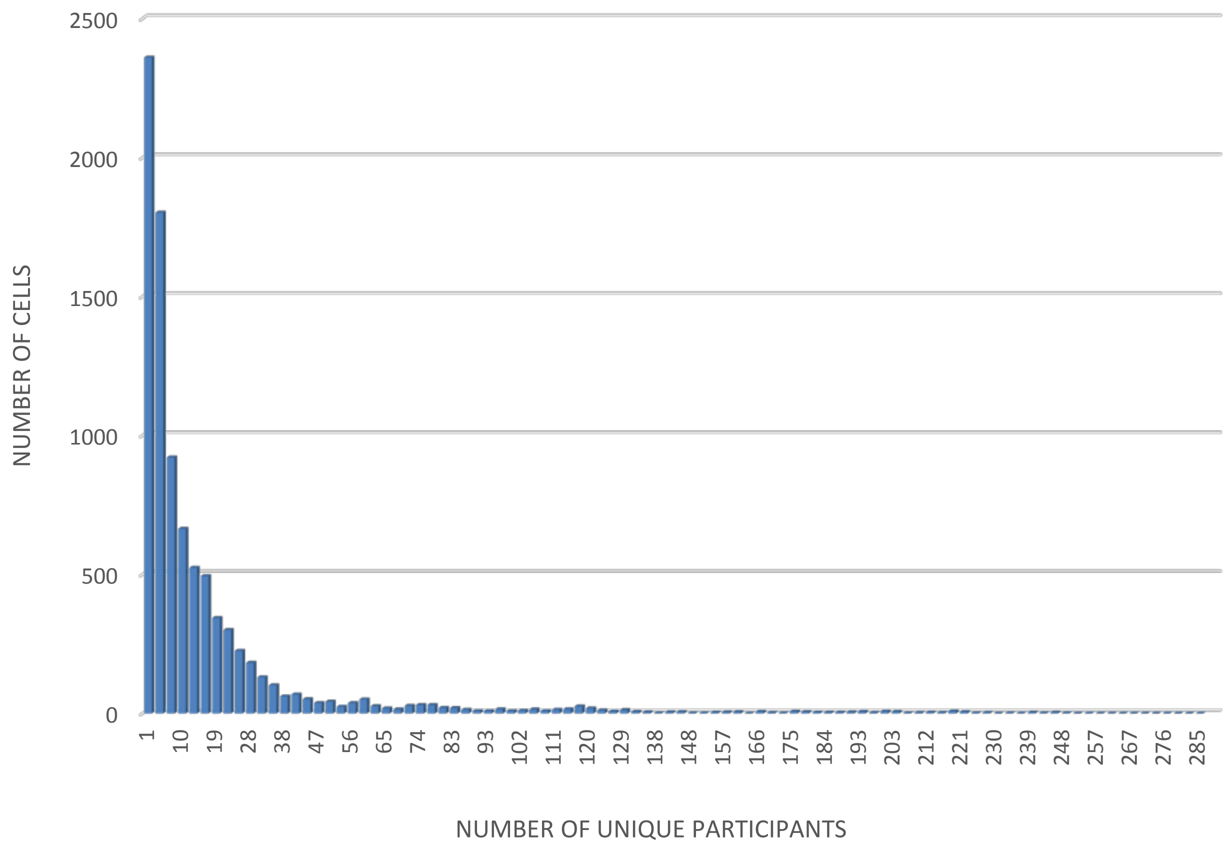

Figure 12: Graph showing frequency numbers of unique participants by numbers of grid cells

The engagements were then overlaid onto the choropleth map to determine if there is a correlation between the number of people passing through a cell and the likelihood of an engagement happening in that cell. Figure 13 shows the locations of the engagements color coded by force type with the numbers of individuals passing through the corresponding cell.

Of the 418 engagements during the day, 317 occurred in 126 grid cells with between 1 and 30 unique people passing through them, whereas the 40 grid cells that had more than 200 unique participants passing through them had only one engagement.

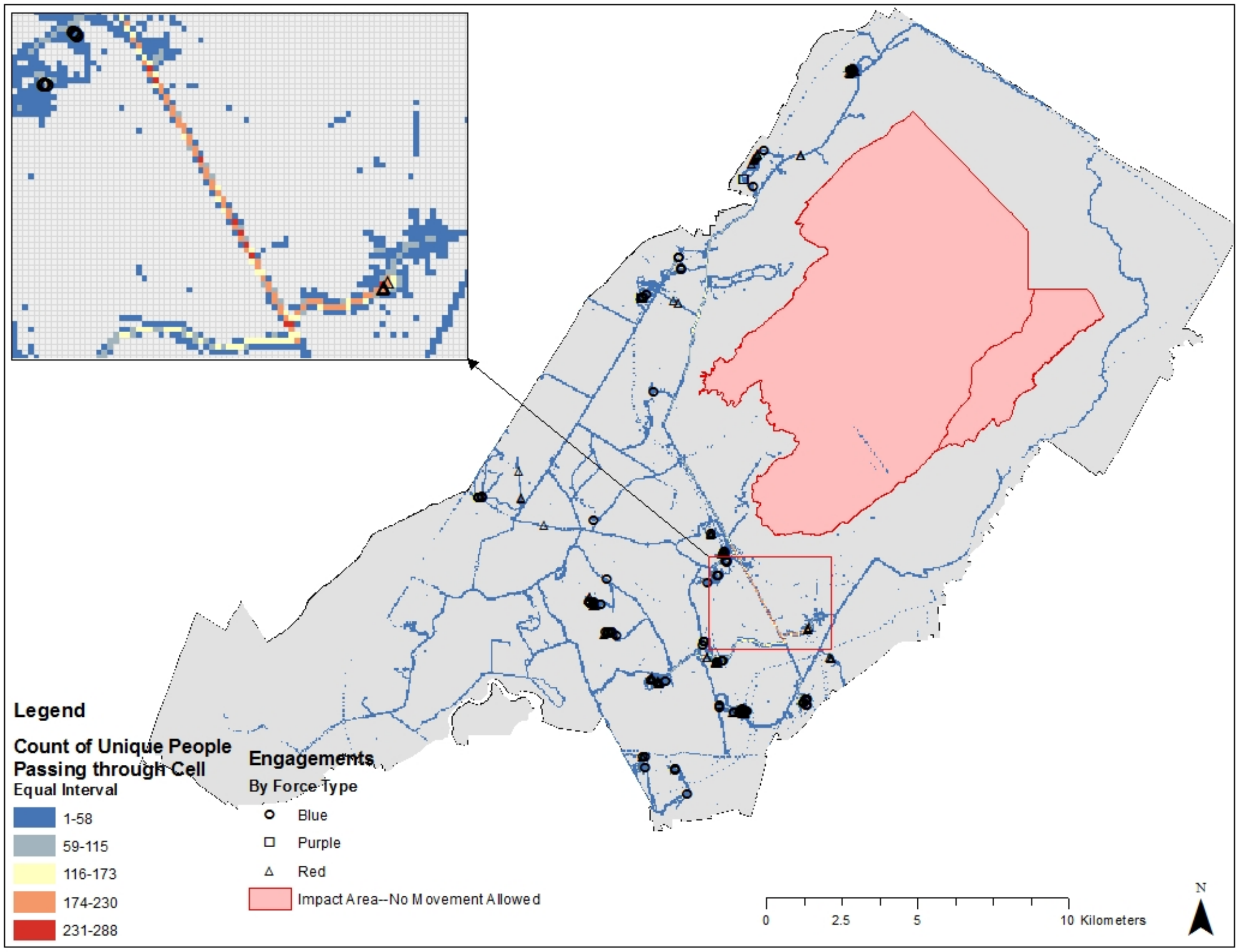

Figure 13: Map showing engagements by number of unique individuals passing through the corresponding cells

The graph reproduced in Figure 14 shows the number of engagements that occurred in grid cells with varying numbers of unique participants and confirms the pattern evident on the map reproduced in Figure 13; namely, that most of the engagements occurred in cells with relatively small numbers of participants passing through them.

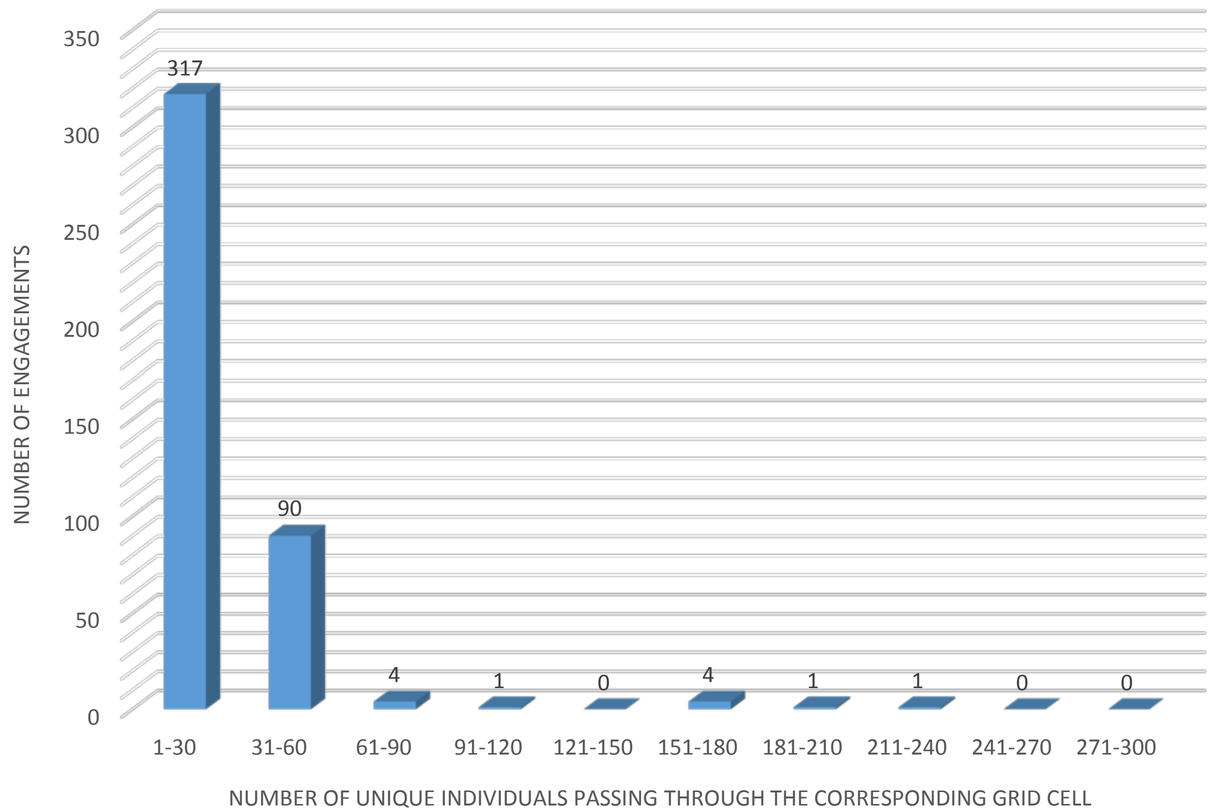

Figure 14: Graph showing number of engagements by number of individuals passing through the corresponding grid cell

4.1.1 Average Number of People in Cell within 30 seconds of an Engagement

To determine the average number of participants in a cell within 30 seconds of an engagement, the data was imported into Microsoft Access and an SQL statement was used to determine the distinct count of individuals in a grid cell within 30 seconds of an engagement event. The summary statistics reproduced in Figure 15 indicate that the average number of people in a grid cell within 30 seconds of an engagement was six, with the minimum being one person and the maximum being 68 people. The graph results reproduced in Figure 15 indicate that there were not many cells with a high count of participants per engagement. Of the 126 grid cells in which an engagement occurred, 83% had less than an average of 10 participants in the grid cell within

30 seconds of the engagement. This confirms that for this particular day of training, a large number of participants in a grid cell does not correlate to an increase in engagements.

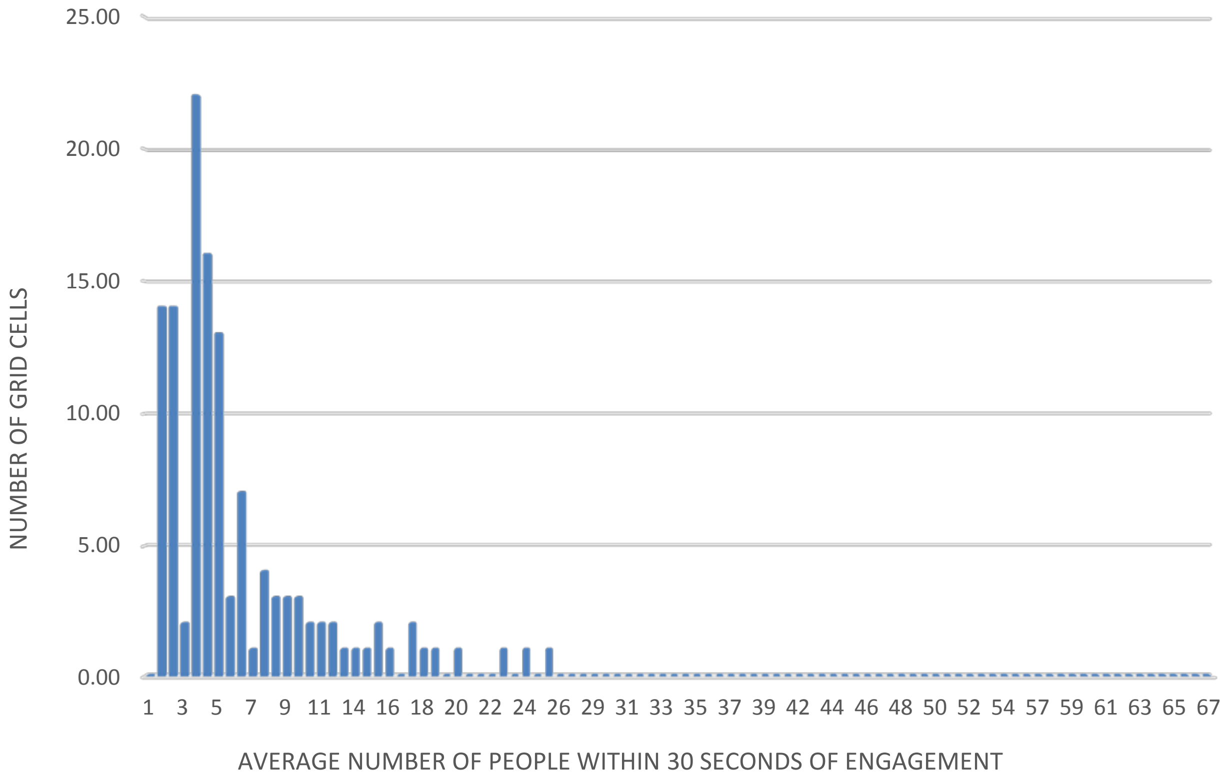

Figure 15: Graph showing number of unique individuals in a grid cell within 30 seconds of an engagement per cell

The choropleth map reproduced in Figure 16 shows the average number of people in a grid cell within 30 seconds of each of an engagement. The area highlighted in the inset map indicates that only one location had a large number of people (55-68 people) in the immediate area within 30 seconds of an engagement. Most of the other locations in which an engagement occurred had between 1-14 people in the grid cell within 30 seconds of an engagement. Considering that the average number of participants that passed through a grid cell during the day was 15, the average number of people present during an engagement was typically less than the average number of people that passed through the cells. The fact that more engagements occurred with a small number of people present is significant and there are several potential factors for why this might be the case. The participants may have been in closer proximity with fewer people around then when there are more people present and had a better view of the target or they may have been more apt to shoot knowing there were relatively few people around. To get a clearer understanding of why this occurred, the training officer would have to look at the engagements along with the training objectives in the area of interest and analyze the scenario as a whole.

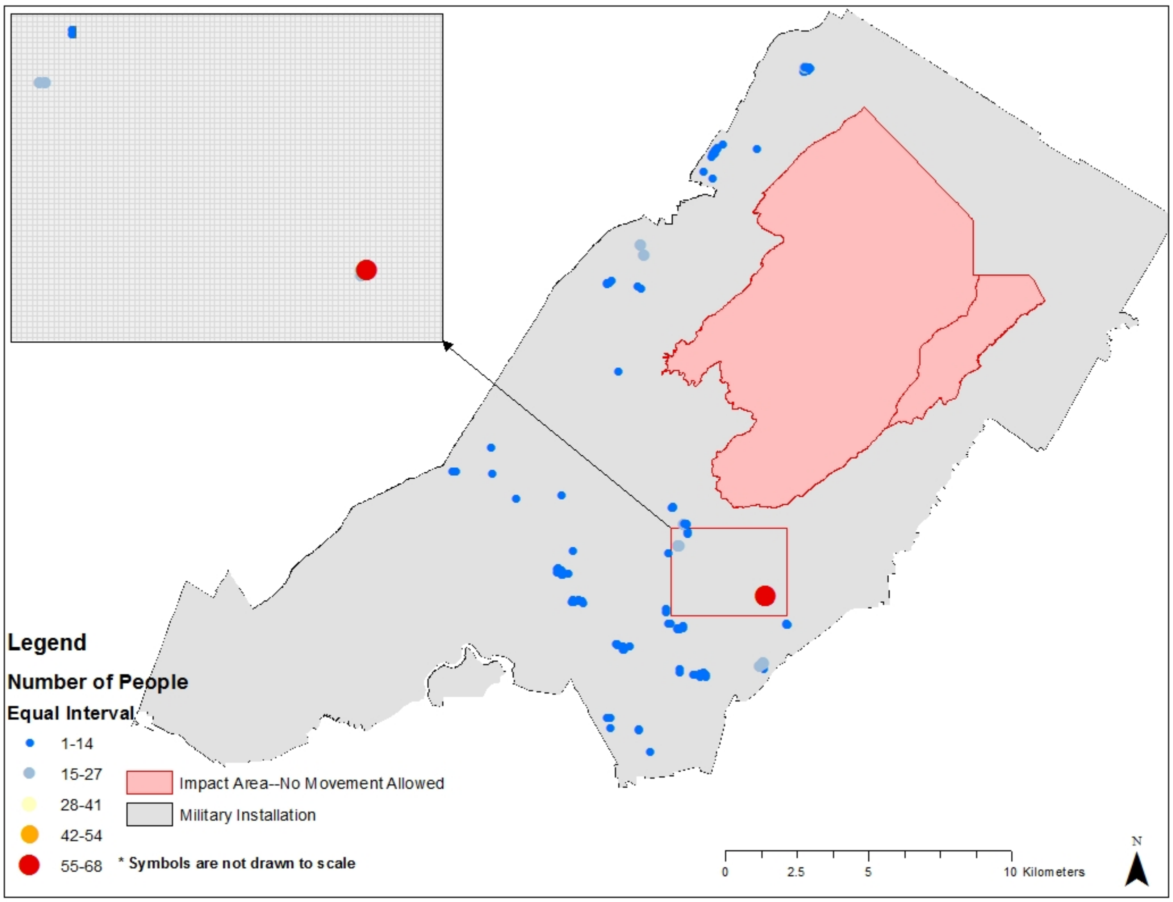

Figure 16: Average number of participants present in each grid cell within 30 seconds of an engagement

4.2 Hot Routes Analysis Results

The hot routes analysis technique developed by Tompson, Partridge, and Shepheld (2009) was used to generate hot spots along a linear network. This analysis was completed twice: once for all engagements along the linear road network and once for the blue force engagements along the road network.

Using the quantile method of classification with five classes, 69% of the road network in which an engagement occurred could be viewed as a hot route (69% of the routes are represented in red, indicating hot routes). This could be misleading because the reader may see many routes in red indicating a hot route with this classification method.

When the classification method was changed to Equal Interval, only 11% of the road network would be considered hot routes (only 11% of the roads are represented in red). While this analysis informs us where potential hot spots of engagements are occurring, how the data is classified is crucial to whether a specific area is determined to be hot route or not. Figure 17 shows the hot routes analysis using the Equal Interval classification method. The results in Figure 17 show that using the Equal Interval classification method ensured that 97.9 m of road fell under the "high incident per meter" category of ≥ 0.41 numbers of incidents per meter rate threshold used to calculate hot routes for all engagements. The next category of 'orange' routes with an incident rate of 0.3-0.4 can still be considered areas of semi-hot routes and this combined with the previous category gave 147.7 m of roads that might be considered as medium or high risk locations.

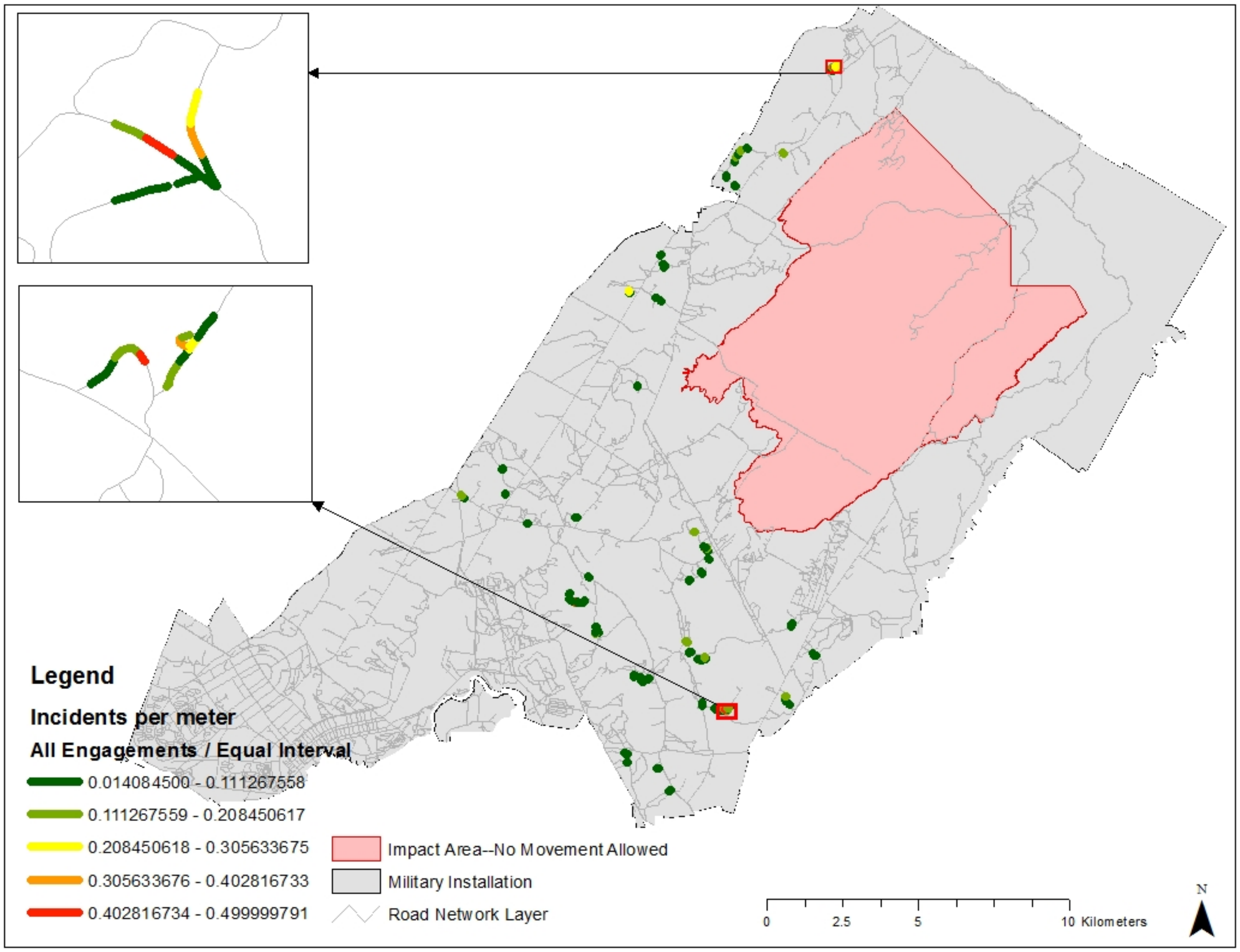

Figure 17: Map showing locations of hot routes for all engagements based on equal interval classification method

The same analysis was completed only using the same Equal Interval classification method for hot routes for the blue force participants, the results show that only 3% of the linear network in which there was engagement activity would be considered a hot route (Figure 18). Only 69.5 m of the road falls under the red class indicating a hot route. In addition, there were no areas on the map that indicated a hot route for blue force engagements that were not also a hot route when all engagements for every participant type were taken into consideration.

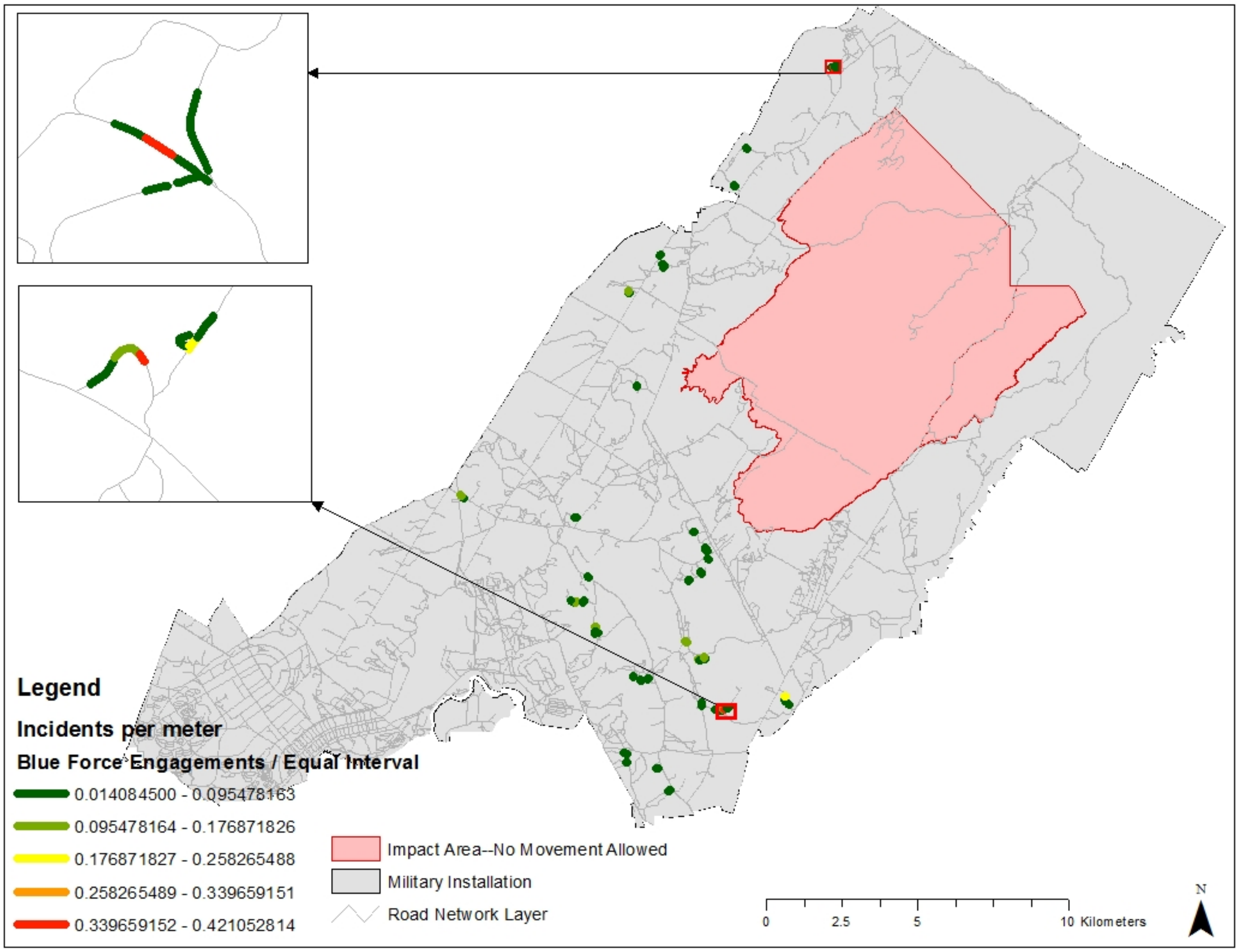

Figure 18: Map showing locations of hot routes for blue force engagements based on Equal Interval classification methods

Figure 19 shows the maps for the blue force engagements and all engagements using the Equal Interval classification method next to one another. As indicated in Figure 19, there are more roads representing a higher incident per meter when all engagements are considered. There is one area in the north that represents a hot route for all engagements but not blue force engagements. This would indicate that this area is a hot route for either red force engagements or COB engagements. There is one area to the southeast that indicates it is more of a hot route for blue force engagements than for all engagements because there was a higher incident rate there.

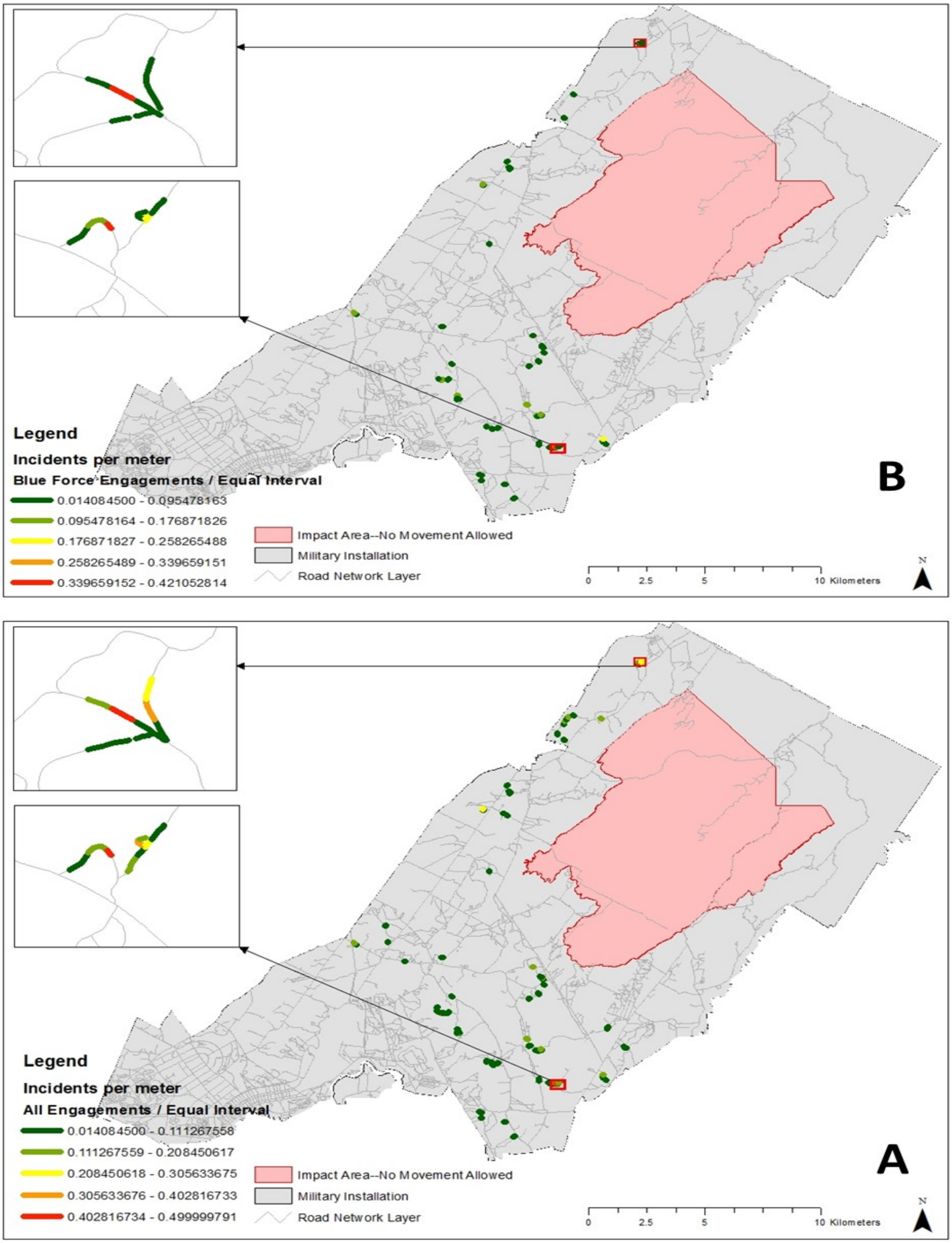

Figure 19: Comparison of all engagements (a) and blue force engagements (b) using Equal Interval classification

It can be concluded that there was not any location that had significantly more hot routes for blue force engagements than for all engagements meaning that when engagements occurred, they were distributed fairly evenly among all three force types. During an AAR, this map could be looked at closely and alongside the daily training objectives to see what may have caused a higher incident per meter rate for the roads in the north and southeast of map A, potentially giving insight into errors that were made.

4.3 Time Interval Analysis

The time interval analysis not only visualized the movements of the participants through each grid cell in 15 minute intervals, but also a table and graph showing the breakdown of each participant type and each interval on the whole field was calculated. The detailed results of this analysis can be found in Appendices A and B. These results provide a larger scale view of participants as a whole on the field. As the graph in Appendix A (Figure A1) shows, the number of participants on the field during the day remained fairly consistent with a slight increase occurring between 10:00 p.m. and 12:00 a.m. GMT. At 12:45 p.m. GMT, the largest number of blue force engagements occurred. At the same time, there was a fairly small number of blue force participants on the field. During the AAR, this time interval should be reviewed closely to determine exactly what errors were being made to account for such a large casualty rate at this time given the small number of participants. The table in Appendix A (Table A1) shows that the largest number of blue force engagements occurred between 12:45 p.m. and 2:30 p.m. GMT. Considering that this was closer to the beginning of the training day, the assumption could be made that the soldiers would be alert and not very tired. The increase in blue force casualties at the beginning of the training exercise would require the commanding training officer to focus on this time of day to determine what might be causing higher numbers of casualties at this time. It could be because the soldiers have not gotten into a rhythm and are not as sure of themselves. To determine if this is an anomaly, the same analysis would need to be to be conducted for every day of the training exercise to see if a similar trend exists from one day to the next. If so, the training officer would need to work with the soldiers during this time of day to increase their situational awareness.

The dataset that included the number of participants passing through a cell was rounded to the nearest 15-minute interval. This dataset along with the engagements dataset were both time enabled in order to visualize the change in movement and engagements throughout the day. A time-lapse video was uploaded to http://spatial.usc.edu/ReinaKahn/Time2.avi visualizing the movement of people throughout the day. A sequential map of each time interval was saved. A sample set is shown in Appendix B showing intervals occurring at the beginning of the day, in the middle of the day and towards the end of the day. The movement throughout the day was mostly linear along the road network. The results indicate that the time that had the greatest number of participants on the field at once (8/7/13 11:30 p.m. GMT) with 805 participants had no engagements during that time interval. Conversely, the time interval that had the most blue force engagements was 8/7/13 12:45 p.m. GMT with 19 blue force engagements. The grid cell that had the most number of participants at any one given time with 103 at 8/7/13 9:30 a.m. GMT had only one red force kill in that grid cell during that time. These results suggest that the number of people in a cell or on the field does not necessarily correlate to an increase in engagements. It is important to note that these results are only for one day. As such, these results could change day by day or between various training exercises at different training locations. The participating soldiers in various units may contribute to different results. Considering the training is dynamic, it cannot be concluded that the number of participants in a cell or on the field will never correlate to an increase in engagements. It can only be concluded that for this day of training, a large number of participants does not cause an increase in engagements.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

The primary purpose of the XCTC training is to prepare soldiers for military conflict in a realistic setting by simulating combat in a controlled environment. Current Army training doctrine does not offer specification for training data analytics. This research has examined the potential benefit of providing daily analysis of training data to inform commanding officers of locations of hot spots of engagements. In addition, the results can also be used to identify if there is a correlation between the number of participants in a given area and the number of engagements. If a correlation does exist, then training iterations could be added at such locations to allow for more focused training. The overall goal of this analysis is to improve the training to overcome identified weaknesses discovered in the analysis. This study provides a framework that training officers can use within their units to provide meaningful feedback following an exercise. By evaluating locations of hot routes on the linear network and providing analysis of events throughout the day at specified intervals, subsequent AARs can include more meaningful feedback.

Each method of analysis for this study was selected to assist stakeholders to efficiently summarize the training day's events and breakdown the data into meaningful results. This training capability uses state of the art proprietary GPS technology to record movement and capture data. The task of processing and organizing the raw output dataset from the GPS equipment proved difficult and timely due to its size (over 5 million records) and complexity. The analyses performed in this thesis were the first attempted for an XCTC exercise. Moreover, there is no known spatial analysis available for any other military training exercise. This made the initial data processing and workflow fairly intricate and time-consuming due to the lack of any available methodological precedents. The final workflow was established by first determining how the data is distributed over space, then modifying and creating a series of query tables that best summarize and quantify the movement of people through time and space. Thematic mapping techniques were then used to visualize the movement and density of people in such a way so that the average laymen could understand and make sense of the results.

5.1 Study Limitations

Due to the unique dataset, a substantial amount of time was spent removing bad data records and clarifying which records were a result of error. The EDI equipment records the first record when the machine is turned on. This results in thousands of erroneous recordings that need to be removed. This problem could be mitigated with improved technology on the EDI equipment. Additionally, the EDI occasionally records a coordinate at the -90^{o}, 0^{o} latitude and longitude. These records also had to be removed leaving the dataset not entirely complete. Given the large number of participants on the training field throughout the day and the large dataset that results from one day of training, only one day of training was examined for this thesis project. Most XCTC exercises last approximately three weeks. Ideally, analysis would occur on a daily basis for the duration of the exercise and then all exercise data would be aggregated and analyzed as well. Unfortunately, the data set was too large for analysis with the current hardware that was available to the author. To perform such analysis on "big data" would require a more advanced processing system and software that could handle big data.

5.2 Needs and Opportunities

In order for the analyses completed in this thesis to be replicated and used during various exercises and at different locations, several additional steps would need to occur in order to ensure repeatability. An Esri ModelBuilder template could be used with specified parameters for the unique input files and to select a specific day or days. This would automate the queries and generate the initial output tables required for the choropleth mapping. Custom programming scripts would be required along with ModelBuilder in order to automate the hot route analysis on a daily basis. The creation of the programming and ModelBuilder templates would be necessary to ensure timely analysis at the end of each day's exercise and before the start of the daily AAR.

The analysis performed in this thesis could be taken further by incorporating terrain analysis as well. Prevalence of engagements could be analyzed in conjunction with slope or elevation. Additionally, looking at shooter information could improve the analytical results as well. Much of the shooter information was not recorded for the engagements during this exercise due to faulty MILES equipment. If a dataset is available in the future that contains adequate shooter information, then analysis of line of site and elevation differences between shooters and victims could be compared.

The hot routes analysis in this study looked at hot spots on a linear network for all engagements and blue force engagements. It may also be beneficial to look at hot routes on the network for fratricide during the whole exercise to determine if it was occurring in the same locations at each iteration. This same analysis could also be performed for engagements based on the weapon(s) used and military rank. This could help to clarify if a certain weapon was more responsible for engagements than other types or if there were hot spots for engagements by rank. To make this analysis more meaningful, enhancements to the EDI equipment could be made so that data regarding orientation is also recorded and then more information regarding the engagements, such as if the participant was facing the sun during the kill or if they were facing towards or away from the opposing force.

A beneficial addition to this analysis would be the use the training objective book for the day's training. This would allow the results of the choropleth map analysis as well as the hot routes analysis to be reviewed taking the objective of the daily mission into account. If this analysis were incorporated into the training standard operating procedure, then this would provide an opportunity for the training officer to speak to the objectives of the day while simultaneously looking at the analytical results.

Future XCTC military training should have a spatially analytical approach to the data being collected. As tracking technology improves and the ability to process results becomes faster, analysis can be performed in the field to ensure that the soldiers are getting the most out of their training and a better understanding of what needs to be improved and how.

REFERENCES

Anonymous. 1998. "Combat Training Centres: A Quantum Leap in Training." *Military Technology* 22.10: 38-42.

Anonymous. 2011. "Video system enables detailed analysis for military training." *SDM* 41(3), 167-168.

Carpenter, Raymond. 2009. "Army National Guard: Integrated Missions." *Army* 59:123-126,128.

Chapman, Robert and Charles Colegrove. 2013. "Transforming operational training in the Combat Air Forces." *Military Psychology* 25:177-190.

Copley, Rhonda and Eric Wagner. 2008. "Improved Situational Awareness through GIS and RFID in Military Exercises." Accessed 26 January 2014 http://proceedings.esri.com/library/userconf/proc06/papers/papers/pap_2350.pdf.

Corson, Mark and Eugene Palka, *Geography and Technology*. (Springer, 2004), 401-427.

Dark, Shawna and Danielle Bram. 2007. "The modifiable areal unit problem (MAUP) in physical geography." *Progress in Physical Geography* 31:471-479.

Esri. 2014a. "ArcGIS 10.1 Resource Center. Classifying numeric fields for graduated symbology." Accessed 23 September 2014. http://resources.arcgis.com/en/help/main/10.1/index.html#/Classifying_numerical_fields_for_graduated_symbology/00s50000001r000000/.

Esri. 2014b. "ArcGIS 10.1 Resource Center. Drawing Thematic Raster Datasets Using Continuous Data." Accessed 16 February 2014. http://resources.arcgis.com/en/help/main/10.1/index.html#//009t00000075000000.

Healey, Fritz W. and Himanshu N. Parikh. 1995. U.S. Patent 5426295 A. Multiple integrated laser engagement system employing fiber optic detection signal transmission.

Indie Mapper. 2015. " Choropleth Maps." Accessed 19 January 2015 http://indiemapper.com/app/learnmore.php?l=choropleth.

Kemp, Ian. 2006. "Training for Close Combat." *Armada International* 30:34-36, 38.

Kutler, Glenn. 2005. "U.S. Military Fatalities in Iraq: A Two-Year Retrospective." *Orbis* 49:529-544.

Milbrodt, Christopher. 2009. "Realistic Training Gets Soldiers Ready for Iraq." Report, Federal Information and News Dispatch, Inc., 2009. ProQuest (190342143).

Miller, H.J. 1994. "Market area delineation with networks using geographic information systems." *Journal of Geographic Systems* 1(2): 157-173.

Morrison, John. E. and Larry L. Meliza. "Foundations of the after action review process." Special Report, U.S Army Research Institute for the Behavioral and Social Sciences, 1999. ProQuest (ADA368651).

Niver, Robyn. "Biological Opinion of the Proposed Activities of the Fort Drum Military Installation." (Biological Opinion, U.S Fish and Wildlife Service, 2009).

Perez, Ray. S., Mark. R. Gregory, and David. P. Minionis. 1992. "Tools and decision aids for training development in the U.S. Army." *Instructional Science* 21:225-237.

Semmens, Robert. "Do you Shoot Well Enough to Save Your Buddy's Life? A Model of Team Firefight Performance." (Research thesis, AAALab. Stanford University, 2013).

Slocum, Terry, Robert. McMaster, Fritz. Kessler, and Hugh Howard. *Thematic Cartography and Geovisualization*. (Prentice Hall, 2009), 134.

Swann, D. 1999. "Military applications of GIS." *International Journal of Geographical Information Systems* 2:889-899.

Tompson, Lisa, Henry Partridge, and Naomi Shepherd. 2009. "Hot Routes: Developing a New Technique for the Spatial Analysis of Crime." *Crime Mapping: A Journal of Research and Practice* 1:77-96.

U.S. Department of Veteran Affairs. 2014. "Unit Size (US Army in the Late 20th Century)." Accessed 13 November 2014. http://www.mirecc.va.gov/docs/visn6/8_us_military_unit_size.pdf.

APPENDIX A: NUMBER OF PARTICIPANTS AND ENGAGEMENTS BY FORCE TYPE BY INTERVAL

Table A1: Count of unique participants (P) and engagements (E) by force type in 15-minute intervals

TIME_15min_Interval (GMT)	P_Blue	P_Purple	P_Red	E_Blue	E_Purple	E_Red
8/7/13 9:00 AM	342	90	163			
8/7/13 9:15 AM	371	91	165			
8/7/13 9:30 AM	377	91	160	1		5
8/7/13 9:45 AM	398	91	158	2		1
8/7/13 10:00 AM	413	91	157			
8/7/13 10:15 AM	426	90	158			
8/7/13 10:30 AM	430	90	158			
8/7/13 10:45 AM	428	88	156	4		
8/7/13 11:00 AM	443	88	156	8		5
8/7/13 11:15 AM	449	88	155	5	4	
8/7/13 11:30 AM	478	87	155			
8/7/13 11:45 AM	478	88	157	2		
8/7/13 12:00 PM	478	88	157			
8/7/13 12:15 PM	467	88	154			
8/7/13 12:30 PM	462	88	155			
8/7/13 12:45 PM	470	91	159	19		
8/7/13 1:00 PM	472	90	159			
8/7/13 1:15 PM	485	89	159	1		2
8/7/13 1:30 PM	488	89	160			2
8/7/13 1:45 PM	489	89	159	11		3
8/7/13 2:00 PM	496	89	157	7		10
8/7/13 2:15 PM	491	87	153	2		1
8/7/13 2:30 PM	497	86	153	11		17
8/7/13 2:45 PM	488	87	154	2		4
8/7/13 3:00 PM	509	89	154			1
8/7/13 3:15 PM	503	86	150	1		6
8/7/13 3:30 PM	502	86	151			
8/7/13 3:45 PM	498	87	152	7		1
8/7/13 4:00 PM	503	85	153			8
8/7/13 4:15 PM	499	85	153	2		1
8/7/13 4:30 PM	495	87	155			1
8/7/13 4:45 PM	495	90	155	1		
8/7/13 5:00 PM	497	94	154	3		4
8/7/13 5:15 PM	495	97	153			
8/7/13 5:30 PM	498	95	152	9		7
8/7/13 5:45 PM	499	95	155	6		5
8/7/13 6:00 PM	497	95	151	1		
8/7/13 6:15 PM	505	96	151			

Table A1 (Cont.)

TIME_15min_Interval (GMT)	P_Blue	P_Purple	P_Red	E_Blue	E_Purple	E_Red
8/7/13 6:30 PM	501	96	150			4
8/7/13 6:45 PM	505	96	147	2		8
8/7/13 7:00 PM	508	93	146	1		11
8/7/13 7:15 PM	490	93	146	2		
8/7/13 7:30 PM	510	93	146	7		3
8/7/13 7:45 PM	509	96	147	3		1
8/7/13 8:00 PM	513	95	147	4		9
8/7/13 8:15 PM	503	92	149			4
8/7/13 8:30 PM	497	91	149			1
8/7/13 8:45 PM	497	92	147			1
8/7/13 9:00 PM	500	92	149			
8/7/13 9:15 PM	510	93	149	1		4
8/7/13 9:30 PM	506	91	148	8		22
8/7/13 9:45 PM	523	90	151	2		1
8/7/13 10:00 PM	533	93	149	2		
8/7/13 10:15 PM	537	93	149	6		4
8/7/13 10:30 PM	535	92	147			3
8/7/13 10:45 PM	541	92	149	10		26
8/7/13 11:00 PM	542	95	152	5		3
8/7/13 11:15 PM	548	97	155			2
8/7/13 11:30 PM	552	97	156			
8/7/13 11:45 PM	547	96	155			1
8/8/13 12:00 AM	541	96	157	5	4	2
8/8/13 12:15 AM	540	95	157	1		1
8/8/13 12:30 AM	521	96	157	13		15
8/8/13 12:45 AM	532	95	156	1		2
8/8/13 1:00 AM	524	94	155	1		2
8/8/13 1:15 AM	523	95	153	2		1
8/8/13 1:30 AM	518	95	157			
8/8/13 1:45 AM	502	91	158			
8/8/13 2:00 AM	514	88	161	1		
8/8/13 2:15 AM	513	86	159			
8/8/13 2:30 AM	510	82	160	1		
8/8/13 2:45 AM	506	82	160	2		
8/8/13 3:00 AM	490	83	159			
8/8/13 3:15 AM	490	83	158			
8/8/13 3:30 AM	488	83	159			

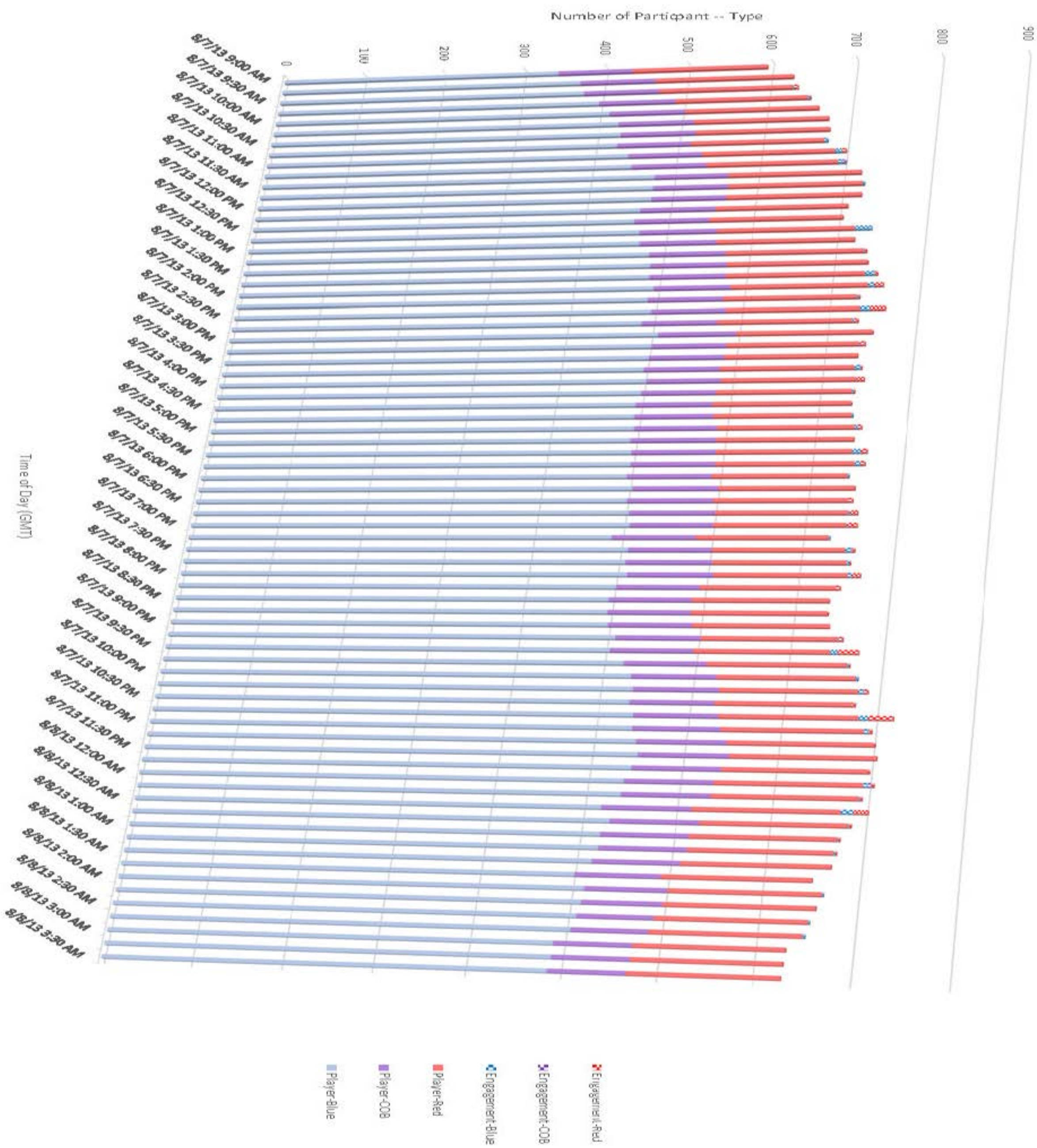

Figure A1: Graph of counts of unique participants and engagements by force type in 15-minute intervals

APPENDIX B: SAMPLE SET OF UNIQUE INDIVIDUALS PER CELL AT 15 MIN INTERVALS REPRESENTING BEGINING, MIDDLE AND END OF THE DAY

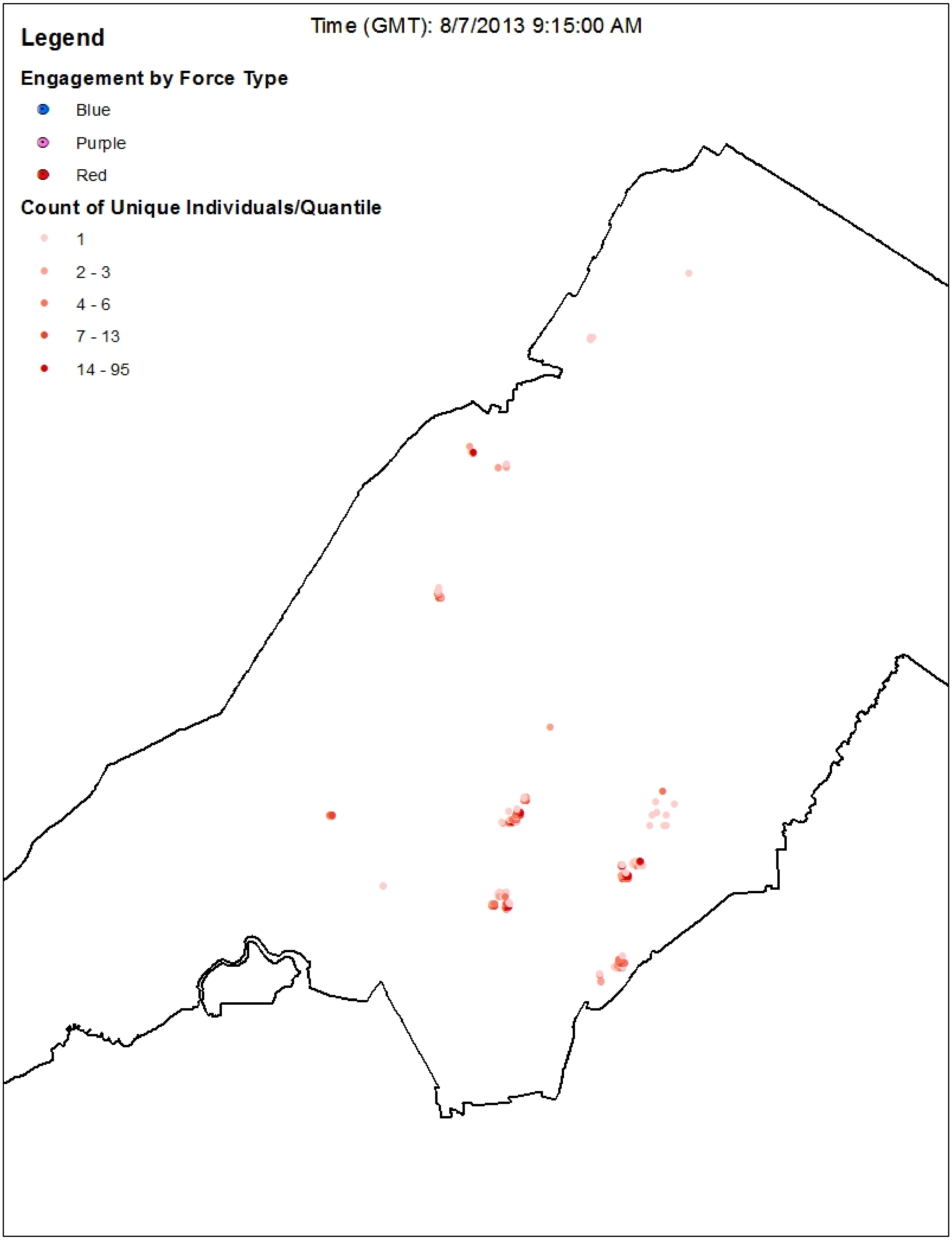

Figure B1: Density of unique participants at 8/7/13 9:15am GMT

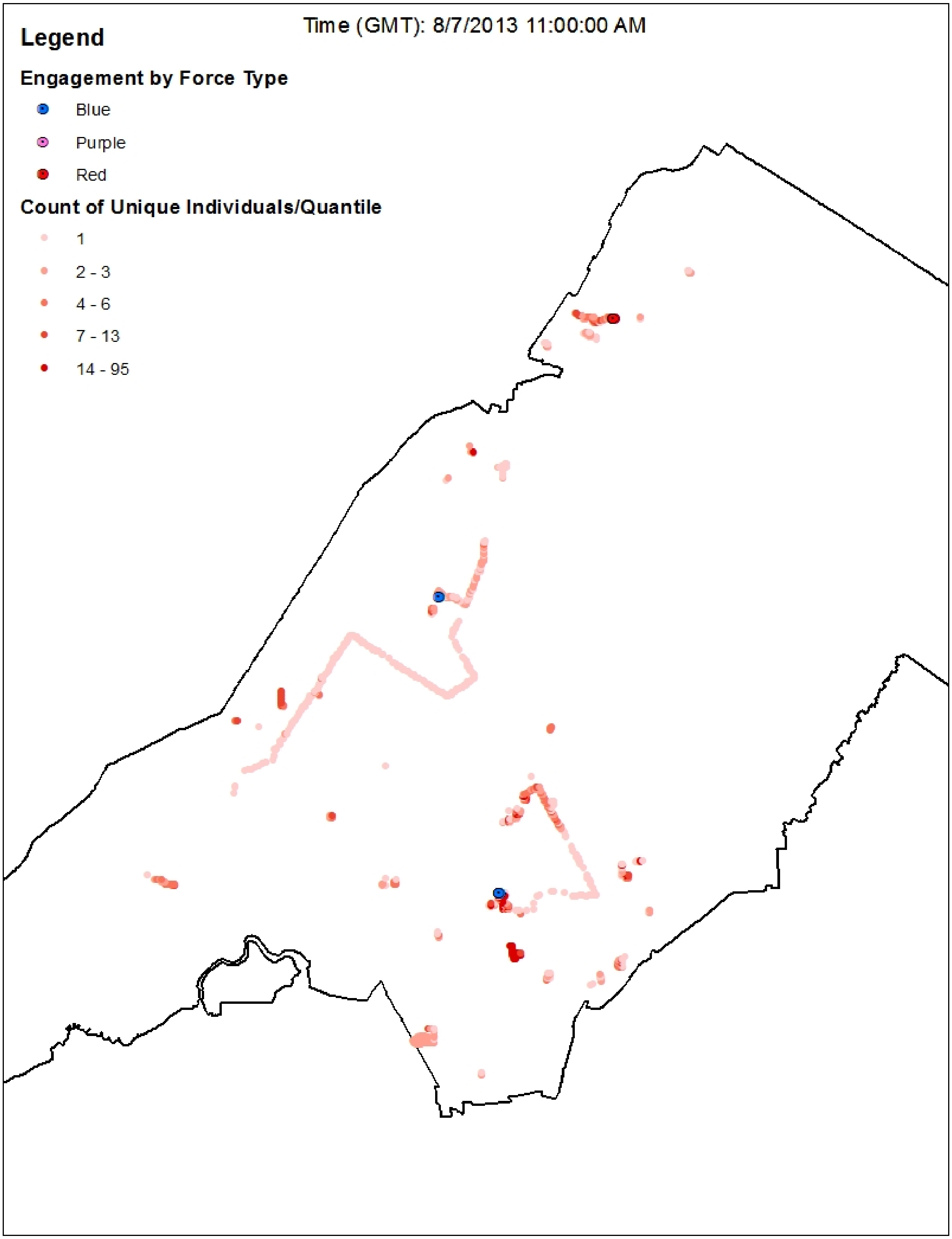

Figure B2: Density of unique participants at 8/7/13 11:00am GMT

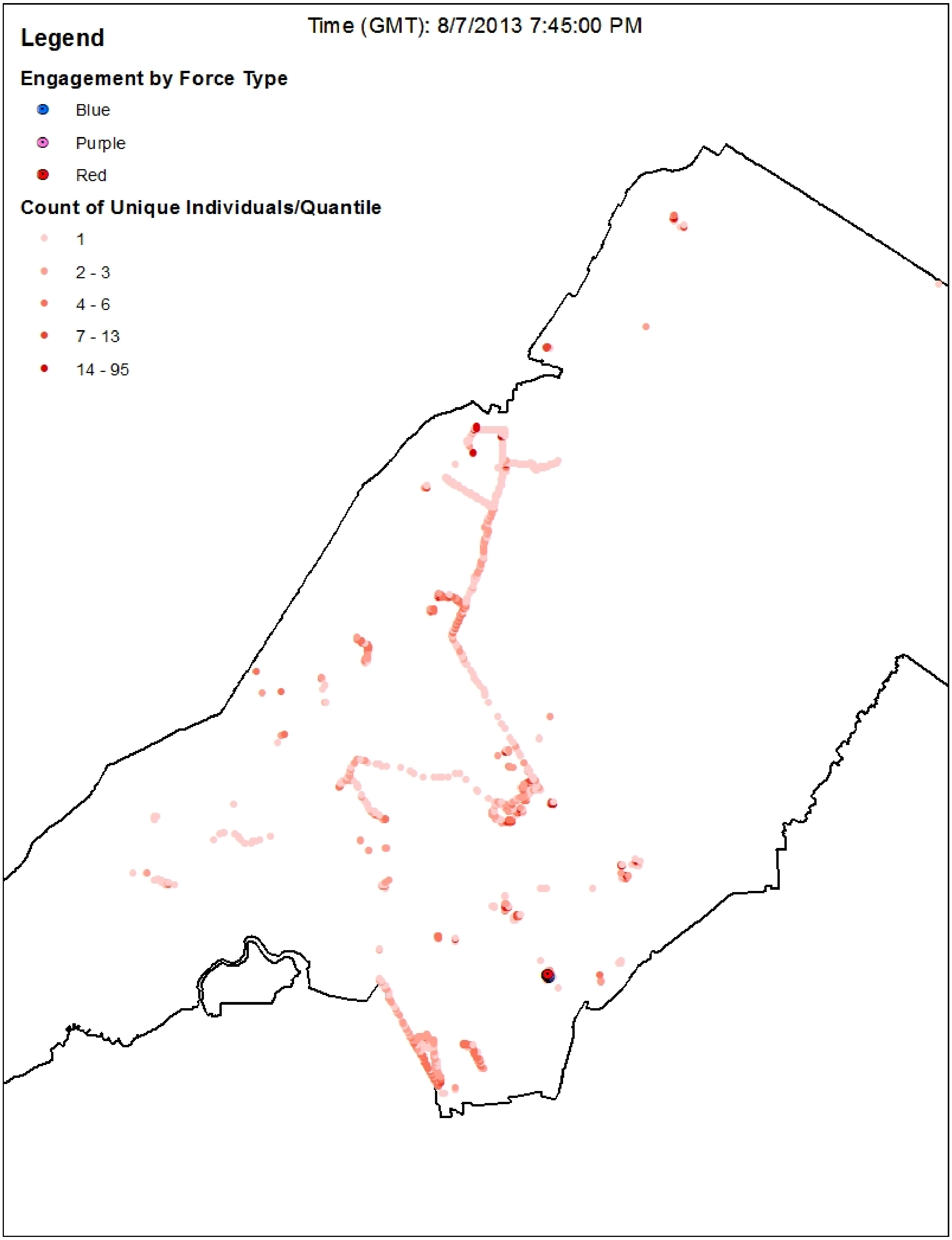

Figure B3: Density of unique participants at 8/7/13 7:45pm GMT

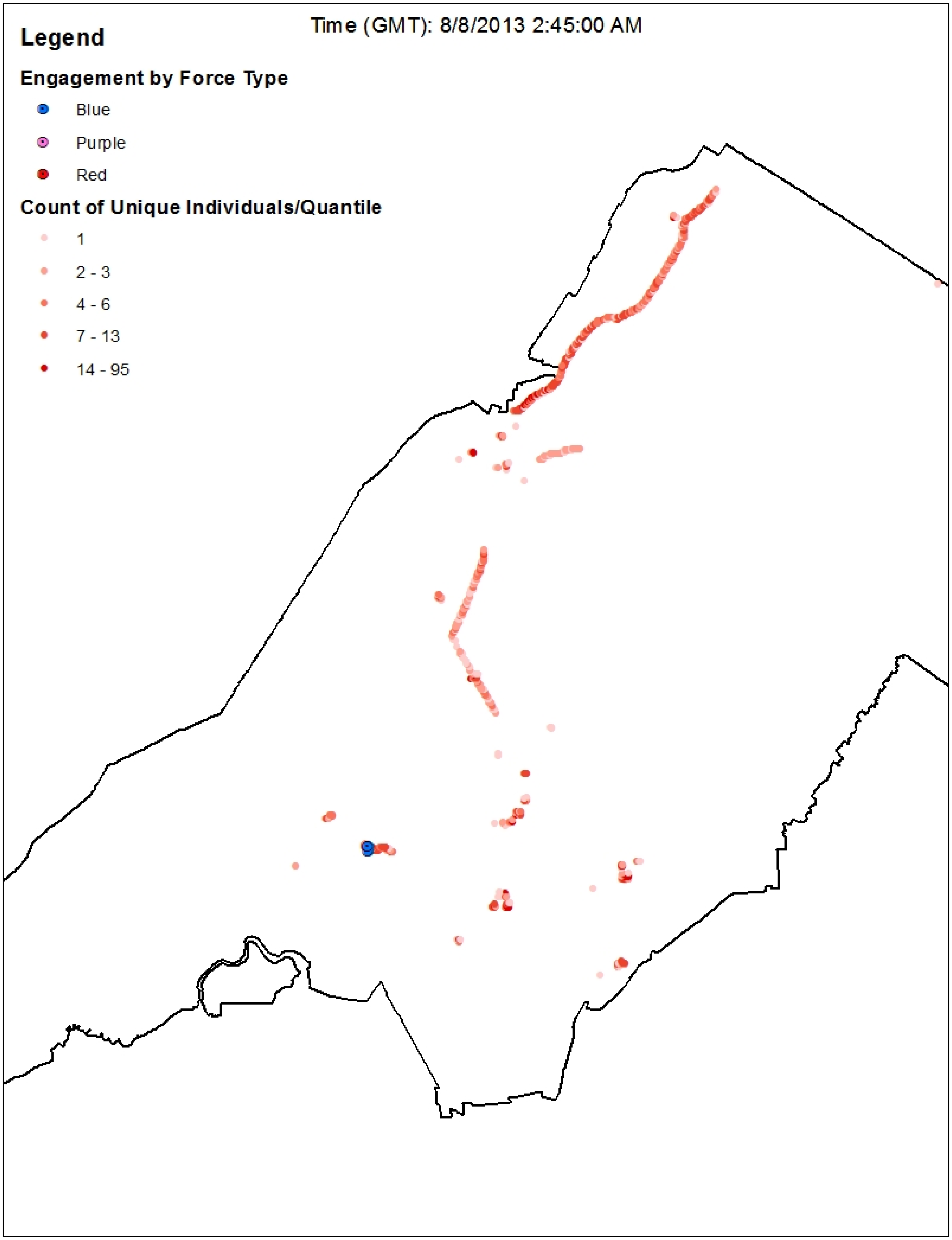

Figure B4: Density of unique participants at 8/8/13 2:45am GMT