SPATIAL AND TEMPORAL MODELING OF IED EMPLACEMENTS AGAINST DISMOUNTED PATROLS

by

Arun Shankar
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Systems Engineering and Operations Research

Committee:

___________________________________ Dr. John Shortle, Dissertation Director

___________________________________ Dr. Andrew Loerch, Committee Member

___________________________________ Dr. David Schum, Committee Member

___________________________________ Dr. Kevin Curtin, Committee Member

___________________________________ Dr. Michael Bailey, Committee Member

___________________________________ Dr. Ariela Sofer, Department Chair

___________________________________ Dr. Kenneth S. Ball, Dean, Volgenau School of Engineering

Date: ______________________________ Fall Semester 2014
George Mason University
Fairfax, VA

Spatial and Temporal Modeling of IED Emplacements against Dismounted Patrols

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Arun Shankar
Master of Science
Naval Postgraduate School, 2008
Bachelor of Business Administration
University of Texas at Austin, 2002

Director: John Shortle, Professor
Department of Systems Engineering and Operations Research

Fall Semester 2014
George Mason University
Fairfax, VA

DEDICATION

This work is dedicated to LCpl Garrett Gamble. May he rest in peace.

ACKNOWLEDGEMENTS

This completion of this dissertation would not have been possible without the warm and gracious support of several key individuals. First and foremost, I am forever indebted to my Dissertation Chair, Dr. John Shortle, for taking six years of his time to help me to the finish line. With numerous deployments and personal distractions, I made this journey as difficult as possible on him. He nevertheless stuck by my side, kept me motivated, and held my hand to the end. Teachers like Dr. Shortle remind me why the path to learning is so rewarding.

I also thank the rest of my dissertation committee, Drs. Loerch, Schum, Curtin, and Bailey. I deliberately chose this "dream team" of professors to help me through a very difficult and unique research effort. Their genuine interest and advice were vital throughout this process.

I must thank the entire Marine Corps Operations Analysis Division leadership of 2008-2011, but particularly Mr. Doug Hoffman and Major John Bancroft. While I was stationed in Quantico, Mr. Hoffman made every effort to ensure I had the tools and the time to complete my research. Doug's support was endless in this regard, and I will never be able to fully repay him for his gratitude. Maj Bancroft was my boss in Afghanistan during the data collection effort described in this dissertation. This research absolutely would not have been possible without his invaluable support and leadership during that deployment. Everyone deserves a boss like Maj Bancroft.

Credit is due to Major Christopher "Junior" Cannon for originally directing me to this research idea. Additionally, I thank Dr. Robert Koyak for sharing his previous research and helping me to develop a roadmap for starting my dissertation. I also offer a heartfelt acknowledgment to Dr. David Alderson, my Masters thesis advisor, who never lost interest in my educational endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

ABSTRACT

SPATIAL AND TEMPORAL MODELING OF IED EMPLACEMENTS AGAINST DISMOUNTED PATROLS

Arun Shankar, Ph.D.

George Mason University, 2014

Dissertation Director: Dr. John Shortle

IED (Improvised Explosive Device) activity has been a concern for US and coalition troops in the Middle East for more than a decade. This dissertation describes a data collection effort in Afghanistan where IED event data and dismounted friendly force patrol movements were obtained. The IED event data is analyzed in time and space with no clear resemblance to a Poisson process in either domain. Consequently, a spatial clustering model is developed to model the collected data with high fidelity and few input parameters. Next, an IED emplacement model is developed to estimate emplacement times based on the interaction between the time and spatial dimensions of the friendly force data and the IED encounter data. Finally, simulated data is used to test the sensitivity of the model to a range of input parameters. From this we suggest improvements to the fidelity of the input data for the most accurate output results.

CHAPTER 1: INTRODUCTION

1.1 Motivation

Tactical-level commanders in Afghanistan face the near-daily challenge of estimating IED (Improvised Explosive Device) activity as they send out dismounted patrols. More accurate models to forecast likely times and locations of IED events are in high demand. Such models require the collection of unique data that has been difficult to acquire for a variety of reasons. This dissertation describes a data collection effort conducted by the first author to obtain friendly force movement of dismounted patrols in Afghanistan. From this data, models are developed to estimate IED emplacement times based on the interaction between spatial and temporal components of the data. The models provide a probabilistic characterization of the emplacement process in time and space.

While such a study of emplacements does not predict exactly when events will occur, it does provide a foundation for understanding the patterns of IED emplacements that are likely to occur in time and space. An appreciation of these trends can directly assist commanders in safely employing troops against the IED threat. In particular, initial versions of the models and data analyses in this dissertation were used by tactical units in Afghanistan to identify IED hotspots. Feedback from commanders for more sophisticated models was encouraging. Military staffs can also use this information to ideally allocate

IED clearance resources such as unmanned aerial vehicles or ground sweeping assets. Additionally, military intelligence sectors can immediately overlay IED emplacement data on existing analysis to directly influence enemy targeting efforts.

IED activity has been a prominent concern for U.S. military forces since the start of both the Iraq and Afghanistan campaigns. IEDs have accounted for over 70% of all U.S. casualties during these conflicts and continue to be the primary security concern for U.S. troops today (OAD, 2010). In 2010 alone, U.S. troops in Afghanistan encountered over 14,000 IEDs and suffered over 4,500 casualties due to IEDs (Cordesman, 2010). Between 2003 and 2010, there were almost 83,000 IED attacks (unexpected detonation in the presence of friendly forces) in Iraq (OAD, 2010). The Department of Defense has spent in excess of 15 billion dollars to mitigate this threat (OAD, 2010).

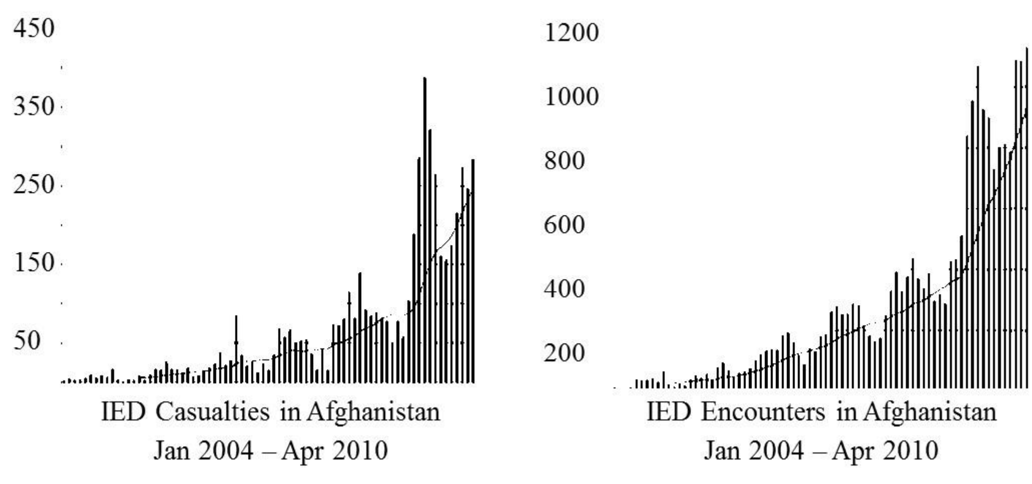

Figure 1: Number of IED-related Casualties and IED Encounters in Afghanistan

IED activity is not limited to Iraq and Afghanistan and is not a new threat to warfare. IEDs have been a strategic weapon in every major conflict since the American Revolution and continue to threaten lives all over the world. Besides Afghanistan and Iraq, much of the world's IED activity occurs in Albania, Pakistan, and Indonesia (OAD, 2010). Military strategists believe that as conventional warfare transforms into irregular warfare, IEDs will likely be the weapon of choice for any future counterinsurgency (OAD, 2010).

The use of IEDs remains resilient because of their simple construction and autonomous nature. IED materials can range from military grade munitions to a coffee can with household cleaning products. A simple IED can be just as devastating as a complex one. Insurgent forces are not hesitant to employ IEDs in any place they believe US troops will be traveling, whether on foot or in vehicles. IEDs are easy to activate and do not require manpower after the initial setup. Like a typical land mine, most IEDs are set off by the victim. The simplicity of IEDs has allowed even the crudest insurgencies to pose a significant threat against the world's most advanced militaries.

Commanders need information about the nature of the IED activity they can expect in their battlespaces. This includes the anticipated time of placement, quantity, lethality, location, and initiator. The information allows them to deter, defend, and reduce this threat. A commander that understands IED activity in his area can mitigate the threat against his troops and improve overall security.

1.2 Essential Background

Analysts have made considerable efforts over the last 10 years to understand, estimate, and predict IED activity. A general focus has been placed on statistical analysis, geospatial hotspot analysis, and future estimation of IED events. Statistical analysis includes summary statistics of IED activity over a given time period, significance tests to determine a change in activity, and trend comparisons to understand the causes of shifts in activity levels. Geospatial analysis considers the location of IED activity and identifies high activity areas with appropriate visual markers on a map. Future estimation can encompass a variety of techniques including mathematical model fitting or law enforcement investigative methods.

Together, the three techniques form a base for the majority of IED analyses since 2002. Such work is often presented in deployed environments or within organizations that support DoD research and procurement. Much of the analysis is comprised of routine products that are created at regular time intervals. Despite the amount of research that has been conducted in this area, shortfalls still remain. Commanders still need analysis that provides quick, actionable ways of mitigating the IED threat. They need a more exact understanding of where IEDs are located and when they were emplaced.

1.3 Proposed Research and Expected Contributions

Most attempts at IED analysis have focused on routine statistics explaining historical data. Some analysis is conducted on estimating future IED activity with limited success. The most sophisticated geospatial and predictive analysis on IED activity is constrained to IED events on road networks. Much of the analysis is based on friendly actions that remain in a perfect steady state, an assumption that is traditionally unrealistic. Additionally, the majority of casualties by IED events in Afghanistan occur against dismounted troops on foot, not vehicles. These IEDs are far more lethal than those targeted at vehicles because of the highly exposed nature of dismounted troops. A model focused on dismounted patrols could provide better utility for commanders.

Additionally, the use of friendly force data to develop such a model would greatly enhance its utility. A reliable estimation of IED emplacements in a two-dimensional space cannot be achieved without knowing friendly force travel patterns. No amount of simulated data can provide the fidelity needed for this endeavor. A real world data collection effort must be executed.

Four key research questions are examined in this dissertation. They are focused on the study of IED emplacements against dismounted patrols and the incorporation of friendly force patrol data in the study of IED emplacements.

- What spatial patterns do IED emplacements have with one another?
- Do IED emplacements conform to a particular spatial point process?
- How can friendly force patrol data be incorporated into the modeling of IED emplacements?
- What estimation benefits can be achieved over models that do not account for friendly force patrol data?

This dissertation describes a data collection effort of IED activity against dismounted patrols in Afghanistan and provides two models that estimate the times and locations of IED emplacements, given friendly force data in a two-dimensional space. There are three main contributions in this dissertation.

- The first is the data collection effort by the first author of friendly force data and associated IED encounters for dismounted patrols in Afghanistan (IED encounters are defined here as events in which an IED is unexpectedly found or detonated by friendly forces). A unique aspect of the collection effort is that the friendly force data is collected in a two-dimensional space, versus road-patrol data which can be viewed as one-dimensional. A spatial analysis shows that the IED encounter data are not well modeled by a spatial Poisson process. The second contribution is the development of a spatial clustering model to characterize the spatial traits of the data.
- The second contribution is the development of a clustering model to explain the data in two-dimensional space. While IED encounters can be directly measured, IED emplacements must be inferred.
- The third contribution is a model to estimate IED emplacement times based on the two-dimensional friendly force data (IED emplacements are defined as events in which an IED has been set up and armed, and is representing a threat to friendly forces). A sensitivity analysis of the model is conducted to understand the potential benefits of acquiring more accurate friendly force data.

1.4 Structure of the Dissertation

We structure the remainder of this dissertation into six additional chapters. Chapter 2 is a detailed background and literature review of existing research and an explanation of relevant stochastic and point processes. Chapter 3 describes the data collected and utilized for this research. Chapter 4 is focused on the temporal analysis of the collected IED event data. Chapter 5 is primarily dedicated to the spatial analysis of the collected data and the development of a spatial clustering estimation model. Chapter 6 describes the development of the emplacement calculation model. Chapter 7 explains the conclusions found from the research effort and possible future initiatives to remaining unanswered questions.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

The employment and effects of IEDs are similar to those of conventional landmines from previous military conflicts. The primary difference is in their construction. While landmines have historically been produced with military grade quality through large scale manufacturing processes, IEDs are generally made of homemade materials and constructed at a bomb maker's home residence. IEDs target people, buildings, and vehicles, just as landmines do. But unlike most landmines, the victim is not required to activate them. IEDs are often initiated by the enemy through either cellular phones or electric switches.

The majority of IEDs in Iraq and Afghanistan target vehicles patrolling urban districts or troops patrolling areas on foot. The urban terrain and established road networks in Iraq allowed troops to conduct most operations in vehicles. Hence, most IED events in Iraq targeted vehicles. In contrast, Afghanistan's rural and mountainous terrain does not easily provide freedom of maneuver for military vehicles. Military forces must often dismount from vehicles and traverse ground objectives on foot, leaving them highly exposed to IED threats with almost assuredly fatal consequences. Routine missions require troops to associate with local populations and positively influence US presence in the area. Consequently, military forces focus patrols in populated areas. IED emplacers are well aware of these objectives and use this information to target troops.

The immediate consequences of IED events range from minor equipment damage to multiple fatalities. However, less obvious consequences include a loss of confidence from the local population, restricted freedom of movement, and a degraded security perception. It is imperative that commanders understand the IED threat in their areas of operations in order to mitigate these effects.

We view IED activity as a game theoretical model involving the actions of both friendly and enemy forces. That is, the discovery of an IED requires emplacement by the enemy and discovery by friendly forces. We know when friendly forces encounter IEDs, but we do not know when they are emplaced. A useful estimation model might provide a solution to this dilemma.

Analysts throughout Afghanistan and Iraq focus significant efforts towards the study of IED activity. Particular interests include pattern exploration, statistical analysis, and predictive modeling. Of the most desired but least successful has been predictive modeling. Analysts are effective at exploring past IED trends and displaying them in a coherent manner to commanders. But it is usually up to the commanders to estimate future activity based on the historical trends. They are forced to integrate their qualitative understanding of the battlespace, past experiences, and gut instinct to prepare for future IED threats. Though this is a respected methodology, it lacks the analytical foundation that most decision makers prefer.

We divide the literature review section into six subsections. The first portion concerns examples of common statistical analysis focused on IED activity. The second section explores geospatial analysis conducted on IED activity, and the third section is focused on future estimation techniques. The fourth section investigates the concept of IED emplacement and research that has been completed in this area. The fifth section is a general overview of common point processes that could have application in the study of IED activity. The final section explores gaps in the research and opportunities for further study.

2.1 Statistical Analysis

Most of the analysis conducted in Iraq and Afghanistan regarding IED activity is rooted in the fundamental statistics of historical activity. This type of analysis is very popular for a few key reasons. First, historical data about enemy activity is readily available and easy to interpret. Almost all US commands in Iraq and Afghanistan have dedicated resources to collecting data on IED activity and share it through web-enabled systems around the world. Secondly, the analysis is simple for commanders to understand. Basic statistics like means, modes, and ranges are intuitive concepts that do not require advanced mathematical understanding. And even more advanced techniques that are explained properly can be understood by a variety of audiences. Third, statistics are a universal language that commanders can trust. Unlike optimization or simulation, statistics are often seen on TV in opinion polls or sports programs. Simple statistics are familiar to commanders, so they are more likely to adopt such analysis into their operations.

In the literature, the most straightforward approach to analyzing IED activity is a trend analysis. A classic analysis is a chart displaying IED encounter rates over time

(Huddleton, 2010; Shearer, 2011). The analysis may be augmented by moving averages or comparisons to other types of enemy activity.

Another approach is a statistical characterization. A common model is the Poisson process and/or the nonhomogeneous Poisson process (NHPP). Researchers have analyzed historical IED event data with Poisson control charts and other fitting mechanisms and have demonstrated that vehicular IED event activity across a broad enough time period can coincide with a NHPP (Kolesar 2008, Kolesar 2009). Though this research was only conducted on one dataset, it is believed that the results are transferable based on the nature of IED events and their relation to the Poisson distribution. A Poisson model is motivated by the hypothesis that IED encounters may be a result of a number of *independent* entities planting IEDs that contribute to the overall IED encounter process. The independence assumption may be weakened if the entities planting IEDs are coordinated or if they are affected by the movement of friendly forces. A partition of IED activity into emplacements and encounters provides a clearer understanding of the effects surrounding the process. Similar to other studies, no consideration is given to the effect of friendly force patrols or the analysis of IED activity against dismounted foot patrols.

If commanders are interested in estimating future activity over shorter time intervals with little historical data, the assumption that IED event data independently follows an NHPP may not be sufficient. The proper estimation of parameters for an NHPP often requires a large dataset, and the predictive power gains reliability as the future time period gets larger and loses granularity. Constraining the model with additional parameters (location, type, target) could provide more actionable results for commanders.

The approximate emplacement time of a future IED by an enemy is extremely valuable intelligence for a commander. Through survival analysis, the window of emplacement could be narrowed to a reasonable level (Koyak, 2009a). The period of "survival" could be defined as the period of time an IED remains emplaced in an area before it is encountered.

Within a given time period, some IEDs are encountered, and others are not. However, assuming that they are all eventually encountered, the time of encounter for the IEDs that are not encountered during this given time period is censored. Typical censoring techniques are used to incorporate this data into an IED emplacement model. Such a model would incorporate Type I censoring, where the point of completion of the study is not related to the number of IED events being analyzed (Klein, 1997). In practice, the completion time would be predetermined. Furthermore, the data could be right or left censored. Right censored data would incorporate emplaced IEDs that were never encountered during the time period. Left censored data would include IED events that were emplaced before the data collection began.

This survival experiment would require the construction of a likelihood function to complete the model. A likelihood function is defined by using maximum likelihood estimation (MLE), a process by which the probability of the existence of certain unknown variable values is maximized based on a set of given parameters. The likelihood requires the definition of a survival function as the underlying distribution for the data. The function need not be complex; something as simple as the Product-Limit Estimator or Nelson-Aalen Estimator (Klein, 1997) might be sufficient. Both estimators are step functions clearly defined by the given data.

Mine warfare has the closest parallel to present day IED threats, and models have been developed to analyze various aspects of this threat. Similar to IEDs, mines are stationary threats that explode when triggered by either the victim or the enemy. One model utilizes the Katz distribution to estimate the number of mines initially present prior to clearance operations (Washburn, 2006). Another model quantifies the risk of various types of mines towards friendly ships (Monarch, 2006). Neither model utilizes friendly force data or focuses on foot patrol movement.

Several statistical techniques exist to assist commanders with better understanding IED activity. Most are simple and easy for decision makers to understand. However, they generally do not incorporate all of the necessary variables, such as the movement of friendly forces, to provide a holistic understanding of IED activity in a complex environment.

2.2 Geospatial Analysis

Much of the statistical analysis conducted on IED activity data is complemented by geospatial techniques that portray historical IED activity on a two dimensional plane. This plane is usually a map of the area of interest, where the x and y axes represent the spatial position on the ground. Some analysts have begun using a z-axis to represent a relative point in time with "hovering" markers of various heights above the map.

Additionally, various clustering techniques have been used to group IED activity by similar characteristics such as time, location, or other traits.

For the most part, the geospatial techniques used to analyze IED activity are rudimentary in nature. In practice, dots are plotted on a map and intuitive conclusions are drawn from the portrayal. There is very little mathematical analysis that goes into the conclusions. However, thorough techniques for geospatial analysis do exist. The most applicable for IED analysis along road networks is known as linear referencing. This technique is popularly used in studying transportation networks and other fields related to networks in engineering (Curtin, 2009b).

Linear referencing offers the analyst an opportunity to shrink the problem space and focus on IED incidents along road networks. In the past, vehicular IED incidents were a matter of concern due to high casualty rates from lightly armored vehicles. Rather than studying an entire geographic area for IED activity, linear referencing allows an analyst to focus resources solely on the road networks (Curtin, 2007). A referencing system along the road networks is established to replace the need for a conventional Cartesian coordinate system or other grid reference platform.

Recent studies in IED analysis have incorporated linear referencing to analyze the density of IED events along a road network (Curtin, 2009a). This analysis allows for a visual and mathematical representation of IED events along road segments of any length. Such information could be useful for commanders, because it provides thorough, but simple, information about the battlespace that can be implemented in decision making. Additionally, the use of linear referencing is not confined to road networks. With the focus of IED activity now against dismounted foot patrols, a variation of the linear networks based on terrain features or other spatial characteristics might be useful in studying IED activity.

An extension of this research is the study of points along networks that are affected by infrastructure elements or the type of roads within the network (Okabe, 1995). For instance, the formation of points may have preference near certain landmarks or on particular road types. Models exist to account for these characteristics. Such a model might prove useful in the study of IED activity, because it would allow analysts to consider similar scenarios on the battlefield. Examples include roads that are more traversed, foot paths that have ideal target opportunities, and choke points that seem to attract more IED activity.

We have surveyed the literature for mathematical analysis techniques used on the estimation of landmine activity in previous wars. The few available articles related to spatial analysis were focused on search theory methods (Cooper, 2003). These models produce a probability heat map that portrays areas with high and low probabilities of success when searching for a landmine. The probabilities are calculated from a multitude of inputs, including terrain, weather, and enemy capabilities. The theory is that mine "sweepers" will start at the areas with the highest probabilities first, then proceed in an efficient way to the areas with the lowest probabilities last. Probabilities are refreshed with each unsuccessful search within the area. Bayesian methods are a critical part of this technique.

This methodology is still used in present day warfare to search for (or avoid) IEDs. The ideas of heat maps and IED estimation probabilities are all related to search theory. Information is provided about the potential IED event and translated into probabilities. However, the recent abundance of data and computing resources has trivialized the modeling that takes place to produce these probability maps. Prepackaged software has allowed for the creation of such briefs in a matter of seconds. Heat maps can be ideal, but the input probabilities to create them still need refinement. Most are based on historical IED encounter data with no incorporation of friendly force movement or other key variables.

The field of spatial statistics and spatial processes lends many ideas and techniques that have potential application in IED analysis. One of note is the use of quadrats to divide a spatial region and investigate the data within each quadrat. Several methods are possible, including the division of a region into equal squares and tallying the quantity of data points within each square (Ripley, 1981). The counts would then be fit to a particular statistical distribution. This use of quadrats can be modified to only include random squares within the grid, or squares of different sizes.

Additionally, metrics are available to study the relative “clustering” of data points within a region (Ripley, 1981). This data might yield conclusions on whether IED activity is homogenously distributed throughout a region, only focused in particular areas, or randomly grouped throughout the space. Many of these clustering techniques yield models that support the study of clustered Poisson processes and doubly stochastic Poisson processes (Ripley, 1981). Both processes promote an idea that the clusters follow one spatial distribution or process, while the data points within the cluster may fit a separate spatial process. A very popular cluster process is known as the Poisson Cluster Process, where cluster centers are spatially distributed according to a Poisson distribution, and cluster members are independently and identically distributed (Daley, 2003).

Many of the analysis techniques discussed so far are regularly found in a field known as crime analysis. Local law enforcement agencies use statistical methods, clustering analysis, and spatial statistics to study crime patterns (Boba, 2005). Data collection in this field usually has a much higher fidelity than within IED analysis, mainly because law enforcement systems are well-established. Queries are run on the incidents to detect trends, patterns, or relationships between data points. From a mathematical standpoint, most of the analysis is simple and performed by user-friendly software applications. Though crime analysis may sound very similar to IED analysis, it has one distinct difference. Crime analysis is generally performed in the United States in established cities that hold volumes of historical data linked together for thorough analysis. Whereas, IED analysis is conducted in unfamiliar war zones with little historical data and almost no prior understanding of local level geography or demographics. Additionally, crime analysis is conducted with the assistance of established data collection systems (Boba, 2005). The shortfalls of such systems that retain IED data will be discussed subsequently.

Geospatial analysis techniques provide decision makers with a visual understanding of IED activity in a given area. Several of the techniques could potentially assist IED analysis, but the necessary input data to study interactive variables is still lacking in most cases. More focus on understanding friendly force movement and applicable spatial processes would add significant value.

2.3 Future Estimation

The estimation of future IED activity is certainly the most beneficial method of analysis, but also the least utilized. Though commanders have interest in historical trends and clear visualizations of IED activity, they are most interested in threats against future operations. They want to understand where IEDs will be encountered when troops conduct patrols. Information should be precise, accurate, and actionable. To date, few analysts have met this requirement.

Some analysts have attempted predictive analysis of temporal observations. One researcher has produced a learning algorithm that better predicts the time of future events from historical data by utilizing a temporal clustering algorithm (Shenk, 2014). Change point detection is a method of detecting small changes in a process that may signal larger changes in the future. This method has been used in cumulative sum charting (CUSUM) to estimate future casualty trends in Iraq (Schneider, 2004). The technique has been used to detect when a significant rise or fall in IED events is about to occur. The CUSUM charting technique has heavy application in manufacturing industries and other processes of that nature, where events occur constantly and have little variation. IED events are typically more variable and unpredictable than a steady process. For that reason, limited research has been conducted by applying known statistical distributions to IED activity in hopes that predictable trends could be drawn from the analysis.

The Joint IED Defeat Organization (JIEDDO) houses a group known as the Crime Pattern Analysis Team (CPAT) that provides future IED activity estimates to commanders in Afghanistan (JIEDDO, 2010). The team contains four to six people with expertise in mathematics, law enforcement, psychology, and other related backgrounds. They use both quantitative and qualitative historical data as well as intelligence reporting to estimate when and where the next IED event will occur. Though an exact point in time and space is never presented, the prediction range provides commanders a range estimate to factor in future operations. In practice, this type of analysis is helpful, but it requires lead time, unique expertise, steady operations, and a volume of data that is not always available. It is also not strictly mathematical, which means it cannot be calculated with software applications and has a degree of subjectivity.

The JIEDDO Threatmapper is another predictive tool (JIEDDO, 2010). It is strictly a geospatial software tool that produces fast, objective results. It searches for similarities between geography and IED incidents within a given region. The results are only as good as the geographic data inputted into the system. For instance, the Threatmapper could conclude that four way intersections have a high probability of IED incidents based on historical data, and then highlight all of the four way intersections within a region as highly probable for future IED events. The results can have utility, but the requirement for detailed geographic and terrain data within a region is generally too cumbersome for the average tactical unit.

The Joint Warfare Analysis Center's (JWAC) Route Threat Assessment Tool similarly searches historical data and estimates the probability of encountering an IED along particular road segments within a network (JWAC, 2011). The analysis can be worthwhile to commanders if operations are steady and the friendly force traffic patterns are constant. Since the model only incorporates historical IED events as input parameters, it lacks the flexibility to integrate changes in future friendly force movement and those effects on IED activity.

Researchers have attempted prediction of US casualties in Afghanistan using equations known as progress curves (Johnson, 2011). Progress curves are used in biology and manufacturing processes to estimate the future production of goods or spread of diseases while considering the competition of the marketplace or medical advancements. In the case of US casualties, the curves attempt to estimate future days where casualties will occur while factoring improvements in US tactics to mitigate those casualties. The curves are general and demonstrate that in the long term, the push and pull between US and insurgent forces does indeed exist. The analysis also demonstrates the use of future estimation as a way to detect a change in US or enemy tactics. When overlaying a prediction on actual data, a deviation from the prediction may indicate a significant change on the battlefield. The technique is interesting to note and might provide a valuable methodology in an IED estimation model.

Significant research has been performed in optimizing the employment of vehicular convoys to reduce the threat of casualties due to IED encounters. One model inherently estimates the probability of encountering a casualty-causing IED in the future by analyzing historical IED casualty data (DeGregory, 2007). Another model optimizes the use of IED clearing assets to minimize future damage to military assets, but assumes all friendly and enemy interactions to be independent of one another (Washburn, 2009). Neither of these assumptions allows the consideration of IED emplacement within the models.

Future estimation is the preferred method of analysis for most commanders, this field still faces significant challenges. The reliability and utility of the predictions is still in development. In most cases, either an inordinate amount of data is necessary to compute an estimate, or the prediction is too broad for an actionable decision.

2.4 IED Emplacement

Statistical analysis, geospatial analysis, and the estimation of future IED activity were presented in the previous sections. In practice, most of this analysis is focused on IED encounters by friendly forces. An IED encounter, however, requires two opposing forces to meet at a single point – the friend, and the enemy. Models that estimate IED encounters are limited in value because they do not account for friendly troop movement. Because every IED encounter is necessarily preceded by an IED emplacement, battlefield commanders would be better served by a model that solely estimates emplacement. Consequently, a model dedicated to estimating IED *emplacements* might have more practical utility for commanders.

One researcher incorporated the probability of an IED emplacement on a vehicular route as part of a convoy scheduling problem (Marks, 2009). The model attempts to determine the optimal time to schedule patrols to clear IEDs from roadways and permit other military vehicles to pass freely. The objective is to clear the roads immediately after the IEDs are emplaced. The research is promising, but unfortunately was not utilized on real world military data. With the absence of this data analysis, the emplacement probability is determined from a standard exponential distribution (Marks, 2009).

One analyst in the community has attempted to build a model focused on IED emplacement that incorporates friendly force activity. The scarcity of research in this area is not unexpected. The idea of studying IED emplacement has long been the task of intelligence agencies and reconnaissance teams to simply obtain visual confirmation of such events and provide the information to commanders. Few analysts have considered building a mathematically based model that can aid in this mission. The obvious challenge is building a model, or set of models, to estimate the emplacement of the IED with the sparse data at hand. An analyst would need to quantify and bound the uncertainty of an IED emplacement, then build a model that could process these inputs and produce actionable results.

The one, existing IED emplacement model supports IED estimation along road networks (Koyak, 2009a). Emplacement dates of historical IED data are estimated and then fit to a statistical model to support the prediction of future emplacements. Various models under a similar framework are drawn up to estimate the emplacement date of a historical IED event. The underlying premise is that an IED could not have been emplaced any earlier than the last time friendly forces were in that area. This can be modified with stochastic or deterministic functions to allow flexibility in this assertion, but the fundamental concept still holds.

The model is further developed to include terms that directly accommodate Blue Force Tracker data (Koyak, 2009b). Almost every tactical vehicle in Iraq and Afghanistan is equipped with a Blue Force Tracker, a device that logs the movement of the vehicle and provides real time position information back to the command headquarters. The data from this device is archived by various agencies, including JIEDDO. Though it is not provided in a user-friendly format, the data is eventually brought into a spreadsheet where meaningful calculations can be accomplished. A final iteration of the research includes 32 functions in the statistical software package R that allows computations of emplacement probabilities and computes the chance of a vehicle convoy encountering an IED along a given road segment (Koyak, 2010).

In summary, IED emplacement is a critical segment of IED analysis that has had little focus to date. Only one researcher has devoted study to this area. Opportunities to broaden this research beyond vehicular IED incidents still exists.

2.5 Point Processes – Overview of Capabilities and Limitations

IED encounters and IED emplacements are best modeled mathematically by point processes. The most common point process used to model IED activity to date is the NHPP. Though this model is simple and flexible, other point processes could potentially have better results. Several point processes are introduced in this section, and the benefits and drawbacks of each are highlighted.

The stationary Poisson Process is the most fundamental point process and is found in many types of queuing models as well as manufacturing and industrial processes. The process is based on exponentially distributed interarrival times between events, or that the arrival and service rates follow a Poisson distribution (Gross, 2008). A Poisson distribution has stationary increments, meaning that the quantities of events in intervals of equal time are identically distributed. Events that follow a stationary Poisson process allow for simple calculations of probabilities and relevant statistics. However, in most circumstances, the assumptions of the stationary Poisson process are too constraining for real world application.

A slightly more flexible model is the NHPP. The NHPP is similar to the stationary Poisson Process except that it allows for nonstationary increments (Ross, 2000). Unlike the stationary Poisson process, the NHPP allows for events to be more likely to occur at some times than others. This characteristic relieves the requirement for exponentially distributed interarrival times. The mean arrival rate in an NHPP varies over time. Though the NHPP offers flexibility for modeling purposes, it still requires that the quantity of events over a fixed time period corresponds to a Poisson random variable, and that each time increment is independent. Most importantly, the nonstationary rate changes over time are deterministic. The changes in average arrival rates are predetermined by historical data or another assumption. There is no uncertainty in how and when the average arrival rate will change when designing an NHPP model. This requirement can be challenging to overcome in many applications.

One analyst has introduced a model where IED activity is portrayed as a Poisson arrival process. The basic model states that IED emplacements are arrivals, and IED encounters are departures. Using Little's Law, over a long enough period of time, the quantity of arrivals and departures is assumed to be equal (Woodaman, 2008). An additional assumption states that IED encounters follow an NHPP, so IED arrivals follow the same. This model allows for several probabilistic results concerning IED emplacement rates, encounter rates, and future IED activity. However, even if the assumptions were validated, the conclusions would only be applicable at a very high level. The results do not provide the granularity or specificity in actionable information for commanders.

The Batch Markovian Arrival Process (BMAP) is a flexible arrival process that does not require Poisson distributed events or deterministic rate changes (Chakravarthy, 2001). The process portrays arrivals in batches of more than one event. When the batch equals one event, the process is called the Markov Modulated Poisson Process (MMPP) (Chakravarthy, 2001). Both models incorporate stochastic rate changes of events, so rate changes can be determined by a separate probabilistic process. This might prove applicability in the study of IED activity if it is assumed that IED emplacements occur in independent "surges" that are not highly predictable. The model can serve a wide variety of applications, but is often underutilized because of the complicated techniques required to calculate final metrics. The fitting of all input parameters to the model is a difficult endeavor. Several of the techniques involve numerical solution methods that may not be practical.

Analysts have often compared IED activity to neighborhood crime activity. This is not a far leap when examined closely. IED activity and neighborhood crimes are both usually found in populated areas. Bad neighborhoods and combat zones are both kept under surveillance by policing authorities, whether civil or military. Analysis techniques for local crime are very similar to those of IED analysts, to include pattern and hotspot analysis (Boba, 2005). Clustering algorithms have also been used in studying auto theft (Kursun, 2005).

The Self Exciting point process (also known as the Hawkes process) has found application in crime analysis. In some scenarios, crimes in a local area spark more crimes in that same area. The Self Exciting point process models this assumption. The model was successfully utilized in describing local crime in Los Angeles in 2004 and civilian deaths within Iraq (Mohler, 2011; Lewis 2011). The validation of the models is performed through visual inspection of the observed and expected outputs as well as a quantitative comparison of the actual results. Other validation techniques include hypothesis testing, where the Self Exciting point process is compared to the stationary Poisson Process and a significance test is performed numerically (Dachian, 2006).

The Self Exciting point process is more complex than either of the earlier stated Poisson processes. Most importantly, it does not necessarily assume independent and identically distributed events. In the case of local crime, one incident causes more incidents to occur. Specifically, the Self-Exciting point process employs two different processes at once – a background event rate and a follow-on event rate. The background event rate is initially developed. Background events are the underlying events that eventually cause follow-on events to occur. Then a follow-on event rate is developed for each background event. Follow-on events do not spark their own follow-on events, unless specifically built into the model (Mohler, 2011). Such a model needs to be carefully tuned to accurately reflect real world circumstances.

Three parameters require estimation and identification to utilize the Self Exciting point process (Mino, 2001). The first is the standard background rate, the second is the rate of the follow-on events, and the third is the amount of time that the aftershock events occur. These parameters are estimated using either maximum likelihood estimation or the expectation-maximization (EM) algorithm (Mino, 2001). The estimation is solved numerically using computational software.

The processes described so far focus on analysis in time, but not space. Commanders need times and locations of IED events to help them make actionable decisions. There exist space-time point processes that track both spatial and temporal variables as events occur (Vere-Jones, 2008). Depending on the format, the space and time variables can be dependent upon each other, stationary, or nonstationary. The probability of an event occurring at a certain location and a given time can be written into the model (Daley, 2008). Such a feature is exceptionally useful for military decision makers.

Existing point processes can be modified to include a spatial component, but there is usually a cost. For instance, the BMAP described earlier contains a matrix that describes a point process. An additional spatial component requires another dimension to the matrix (Breuer, 2005). The modification requires every event in the process to be subject to a spatial restriction defined by a spatial distribution (Baum, 2001). Depending on the size of the state space, the computational requirements of such a change can grow exponentially.

Space-time point processes are a more specific version of a general class known as marked point processes. In short, each point in a marked point process is assigned a time and a mark. The mark can be any descriptive characteristic about the point, including location, value, or cost (Serefozo, 2009). A point can have any number of marks. In the case of IED analysis, a point can vary on time, location, mission of the unit, or the type of IED. An associated probability distribution would be assigned to each mark.

Cluster processes characterize points that form in groups positioned closely together. Input parameters determine the location of the points and the spread around the cluster centers. The Neyman-Scott cluster process assumes unobserved cluster centers with associated cluster members surrounding each center (Cowpertwait, 1997). The center points are formed according to a Poisson distribution and the cluster members are distributed randomly. Cluster processes have the potential to better explain IED activity in the spatial domain.

2.6 Gaps in the Research

The existing body of knowledge has exploited several areas in the study of IED activity. Considerable emphasis has been placed on spatial and temporal analysis with the limited data that has been available. The primary focus of research has been on the historical patterns of IED encounters in both space and time, particularly against vehicular convoys. We summarize key points of the existing research in Table 1 below.

Table 1: Literature Review Summary

Author	Title	Year	IED Encounter	IED Emplacement	Space	Time	Friendly Force	Patrols (General)	Foot Patrols
Kolesar, P.J.	Time Series Analysis of Improvised Explosive Device Incidence	2008	X			X		X	
Kolesar, P.J.	Poisson Trending of IED Event Frequencies	2009	X			X		X	
Huddleston, S.H.	The Warfighter's Guide to Counter-IED Analysis	2010	X			X		X	
Schneider, P.M.	Dynamic Incident Display and Change Point Detection and Counterinsurgency Operations	2004				X			
Curtin, K.M.	A Comprehensive Process for Linear Referencing	2007			X			X	
Curtin, K.M.	Road Network Analysis and Linear Referencing	2009			X			X	
JIEDDO	Threatmapper	2011	X		X			X	
JIEDDO	Crime Pattern Analysis Team	2011	X		X	X		X	
JWAC	Route Planning Tool	2011	X		X	X		X	
Koyak, R.A.	Risk on Roads: A Modeling Approach, Parts 1 & 2	2009		X	X	X	X	X	
Koyak, R.A.	Risk on Roads: A Modeling Approach, Part 3	2010		X	X	X	X	X	

Several of the IED incidents in Afghanistan have occurred against dismounted patrols, not vehicles. In the past, commanders have been able to mitigate much of the IED threat against vehicles with heavier armor and mine resistant designs. Such solutions cannot be implemented for dismounted troops. In most circumstances, the consequences of IED incidents that target dismounted troops are far greater than those against vehicular patrols.

As discussed before, analysts have attempted to study IED activity as an independent phenomenon. That is, one that is not affected by friendly force actions. Though this is known to be a flawed methodology, past researchers have had few choices otherwise. With the exception of BFT data along vehicular routes, there has been no known acceptable record of friendly force movement that can be applied in quantitative analysis.

Additionally, a study of IED encounters provides limited use for commanders since the probabilistic portion of an encounter is the emplacement by the enemy, not the encounter by the friendly forces. Rather, a study of IED emplacements and estimates of future IED emplacement activity could serve a tactical commander with actionable information.

Consequently, models that estimate the *time or location of IED emplacements against dismounted foot patrols* have high utility for commanders. But the creation of such models comes with many challenges. The first is collecting or acquiring data with enough fidelity to build a useful model for tactical commanders. As trivial as the task may sound, it has never been done, mainly because of the nature of the data and the obstacles to collecting it in a war zone. The second is the challenge of building the framework for an IED emplacement model in a two dimensional space. The previously discussed emplacement model that incorporates BFT data and vehicular patrols was built for one dimensional, linear roads. Foot patrols are not just restricted to roads, so such a model would need more flexibility. And the third challenge is selecting and developing the correct point process with proper parameters to describe IED emplacement activity. This is arguably the most daunting of the three challenges. These obstacles lend credence to why no such research exists to date on this topic.

CHAPTER 3: DATA COLLECTION

During a recent visit to Afghanistan by the author, a data collection effort regarding the movement of friendly force patrols on foot and associated IED encounter data was conducted. At the conclusion of the deployment, data was obtained for 9,550 foot patrols and 714 IED encounters over almost a 90 day period for a particular area of operations in Afghanistan. The data was documented in an anonymous manner that was deemed unclassified. Of the 9,550 foot patrol records, only 4,526 entries had all fields completely entered. These 4,526 entries comprise the data set used in this research.

The locations of the IED events and friendly patrols were provided in Geodetic format with familiar latitude and longitude coordinates. In order to present the data in an unclassified format, the coordinates were converted to Universal Trans Mercator format and the points were shifted to another location in the world. A similar technique was performed on the date and time of each event. Consequently, the data is unclassified but still maintains the spatial and temporal relationships necessary for analysis.

3.1 Data Collection Challenges in the Military

The military often has difficulty evaluating trends and effectiveness over time because of a lack of complete and accurate data. In many instances, lower level leadership does not make an emphasis on the requirement for deliberate data collection strategies. But there are valid reasons for this dilemma. Commanders and military staff members that are deployed focus on future planning. This includes everything from movement strategies at the highest levels to food and ammunition readiness at the lowest levels. An additional burden to analyze the past is usually not in the interest of decision makers, particularly if the benefits of such collection efforts are not obvious.

Additionally, for those few commanders that are interested in data collection at tactical levels, the systems necessary to make the collection possible are often broken or nonexistent. These systems include computers, networks, trained personnel, operating procedures, quality control, and final analysis required to make data collection efforts meaningful and worthwhile. Maintenance of these systems puts a strain on competing requirements focused on present and future operations.

Furthermore, for much of the data that has been gathered meticulously over the years, it is believed by many commanders that the analyst community has often not produced relevant, timely, and actionable solutions that leaders need to make rapid decisions. Commanders have often dedicated resources to data collection only to be provided with abstract, theoretical results that do not provide immediate solutions. Such results have brought about reputations that shun analysis efforts all together, where commanders would rather dedicate energies to assured results like field patrols and intelligence operations. In summary, the incentive for dedicated data analysis has come up short for many tactical commanders in recent history.

3.2 Patrol Data Collection Effort

As previously noted, data regarding *enemy activity* is maintained in a sufficiently complete and accurate manner across nearly all battlespaces in Afghanistan. But data regarding the specific movement of *friendly forces* is almost nowhere to be found in Afghanistan. It might be maintained by local commanders in Powerpoint briefs and written logbooks, but it is not stored in a format that can be analyzed in a practical manner. The main exception to this is Blue Force Tracker (BFT) data. This data shows the movement of US and some coalition troops in vehicles. But even this data is dependent upon whether or not the vehicles moving are carrying a BFT device. And it does not include data for troops moving on foot that are not in vehicles.

Research analysts in Afghanistan were surveyed for information about existing databases on foot patrol movement. Surprisingly, US units had little to provide, but one coalition partner had a skeleton spreadsheet that was being tested on their subordinate units. With this inspiration, the author and his fellow analyst began a similar process with some tactical units in the Marine Corps. The energy motivated other analysts to consider similar collection efforts in their areas. It also recursively generated increased emphasis by commanders in the Marine Corps for more thorough and complete collection of friendly force data.

The efforts for collection of such data highlighted several important lessons for field analysts. First, matters discussed earlier about the overall challenges of collecting data in the military were validated to be true. Convincing a commander of the importance of data collection required a high level of patience and effort that was not initially expected. Commanders were not willing to support a requirement for data until the benefits were thoroughly explained and sample results were produced. They were also adamant about fast answers to questions, and 48 hour deadlines for the delivery of statistical products. They also demanded a user-friendly spreadsheet and user training to collect the data efficiently.

These requirements obviously necessitated face to face interaction with the tactical units on a regular basis. Routine visits to forward positions were conducted to train Marines on inputting data, ensure that all questions were answered, and make revisions to the statistical products that were developed with the data. The author utilized Visual Basic for Applications (VBA) behind the spreadsheet tracker to automate real-time statistical reports for the users and develop error control messages so that erroneous data was minimized. Such tools allowed the troops in the field to log the patrol data directly after the mission was completed. This face to face interaction at the tactical level made the collection effort a success.

Data were obtained for 4,526 foot patrols and 714 IED encounters over an approximate three month period for a particular area of operations in Afghanistan. Times and locations were shifted to remain unclassified, but necessary spatial and temporal relationships remained intact for analysis. For each foot patrol, a start time, a start point, and a farthest point were collected. Ideally, a continuous track would have been preferred, but such data was not available due to a lack of Global Positioning System (GPS) tracking devices and time constraints on data collection efforts. With only a start point and a farthest point, an assumption must be made regarding the area or path of the patrol. This will be discussed in Chapter 6. Table 2 shows the set of fields collected for the patrol data. These data fields are defined as follows:

- Start day – start day of the patrol, where decimal indicates a portion of the day.
- Start x and y (meters) – the x and y coordinates of the patrol start point.
- Farthest x and y (meters) – the x and y coordinates of the farthest point from the start that was traveled by the patrol.

Table 2: Sample Patrol Data

Friendly Force Patrols				
Start Day	Start x (m)	Start y (m)	Farthest x (m)	Farthest y (m)
26.80	602400	3491300	601500	3491000
26.80	602400	3491300	601500	3491000
26.99	610400	3477300	610070	3477360
…	…	…	…	…

The foot patrol data is far from perfect, as only a start point and farthest point are provided. A set of checkpoints for each foot patrol or even an electronic track that accurately reflects where troops have traversed would have been preferred. Furthermore, there is reason to believe that the farthest point was not always filled in accurately. The finding of several duplicate values highlights that either patrols were continuously turning around at the same point, or the more likely scenario that patrol leaders were not remembering to document farthest points until they returned to base. The data gives some understanding of where the troops patrolled, but a level of estimation is required for this data to feed an IED emplacement model.

3.3 Available IED Data

This section describes an existing database of IED events. These events are linked to the foot patrol data described previously. IED events fall under the category of enemy actions and are recorded in the Combined Information Data Network Exchange (CIDNE). Quality control measures are taken to ensure that records are complete and accurate in CIDNE. Thus, unlike the foot patrol data collected by the author, all of the data fields are complete. Table 3 shows the fields for the IED encounter data used in this research. These include the day of the IED event (decimal indicates a portion of the day), and the *x* and *y* coordinates of the location of the IED encounter.

Table 3: Sample IED Encounter Data

IED Encounter		
Day	*x* (m)	*y* (m)
0.00	654420	3520550
0.33	617140	3494130
1.38	597378	3478780
2.00	647732	3506634
…	…	…

Because the encounter and patrol data come from separate sources, each IED encounter date does not necessarily match with a patrol start date. To link these data sets, it is assumed that the most recent patrol whose patrol area (described later) overlaps the IED encounter is the patrol that encountered the IED. A single patrol may be matched with zero, one, or several encounters. An IED is matched with at most one patrol, and some IEDs are not matched with any patrols. Figure 2 shows the relationship between these two datasets.

Friendly Force Patrols				
Start Day	Start x (m)	Start y (m)	Farthest x (m)	Farthest y (m)
t_1	x_{11}	y_{11}	x_{21}	y_{21}
t_2	x_{12}	y_{12}	x_{22}	y_{22}
t_3	x_{13}	y_{13}	x_{23}	y_{23}
t_4	x_{14}	y_{14}	x_{24}	y_{24}
…	…	…	…	…

IED Encounter		
Day	x (m)	y (m)
t_5	x_1	y_1
t_6	x_2	y_2
t_7	x_3	y_3
t_8	x_4	y_4
…	…	…

Figure 2: Data Source Relationships

Despite the lack of useful data, improvements to data collection have seen leaps and bounds since the beginning of the Iraq War in 2003. Many new web-based systems archive data in useable forms that can be accessed by US and coalition military forces across the globe. Much of the data is utilized by intelligence shops in Afghanistan to help produce quantitative and qualitative reports that support operations. The data is stored in a combination of web-enabled databases along with several locally owned spreadsheets maintained within individual units.

CIDNE is the primary database of record in Iraq and Afghanistan. Theoretically, a plethora of information about patrols, enemy activity, friendly activity, and humanitarian aid missions can be stored in this database. However, as discussed earlier, data collection is driven by commanders. If commanders are not convinced that a collection effort will help them to make decisions, they likely will not support it. As a result, enemy activity data is the most complete and accurate of all the available repositories. This includes information regarding IED events as well any other enemy engagements with enemy weapons. The one dimensional analysis of enemy activity has been a long standing metric of progress for commanders in both wars since 2003 as well as justification for the purchase of new gear and equipment. Consequently, it has remained the main data collection effort.

Other repositories in CIDNE receive intermittent attention whenever a commander considers the need for the data. But the efforts usually fade when it is realized that a considerable amount of data is usually required before any noteworthy analysis can be produced. Additionally, lateral collection efforts by adjacent units are required to compare trends and understand relationships within an area of operations. Therefore, if one unit is collecting data but another is not, the existing collection effort falls short of its potential and may lose inertia.

The completeness of data is also a matter of concern. If Commander X collects data Y and Commander Z collects data YY, it may not be possible to combine Y and YY in a global way to do analysis. Thus the inconsistency of data across multiple collection efforts hinders analyses. In general, the military lacks the overarching emphasis to consistently collect relevant data in high fidelity across all levels.

Furthermore, many of the less used and dormant data repositories in CIDNE lack the database structure to allow rudimentary analysis of trends and statistics. Several of the repositories are laden with free text entries that cannot be analyzed without sophisticated word search algorithms that must be customized for each analysis product. Such circumstances deter analysts from utilizing CIDNE and instead rely on other databases for support.

There exist other data repositories that hold various types of information about military operations and their outcomes. The Combat Operations Center, the central hub for the coordination of all operations within a battlespace, usually maintains several spreadsheet trackers and classified chat logs that can be mined by analysts. The various intelligence shops within a unit also maintain spreadsheets and web based systems that interact to provide a qualitative understanding of the battlespace. Unfortunately, most intelligence products are stored in PowerPoint files that are extremely difficult to mine for quantitative analysis.

3.4 Overview of the Data

The data is initially examined by location with an expectation that a spatial correlation exists between the location of patrol start points, patrol farthest points, and IED events. The results are shown in Figure 3 and substantiate this premise.

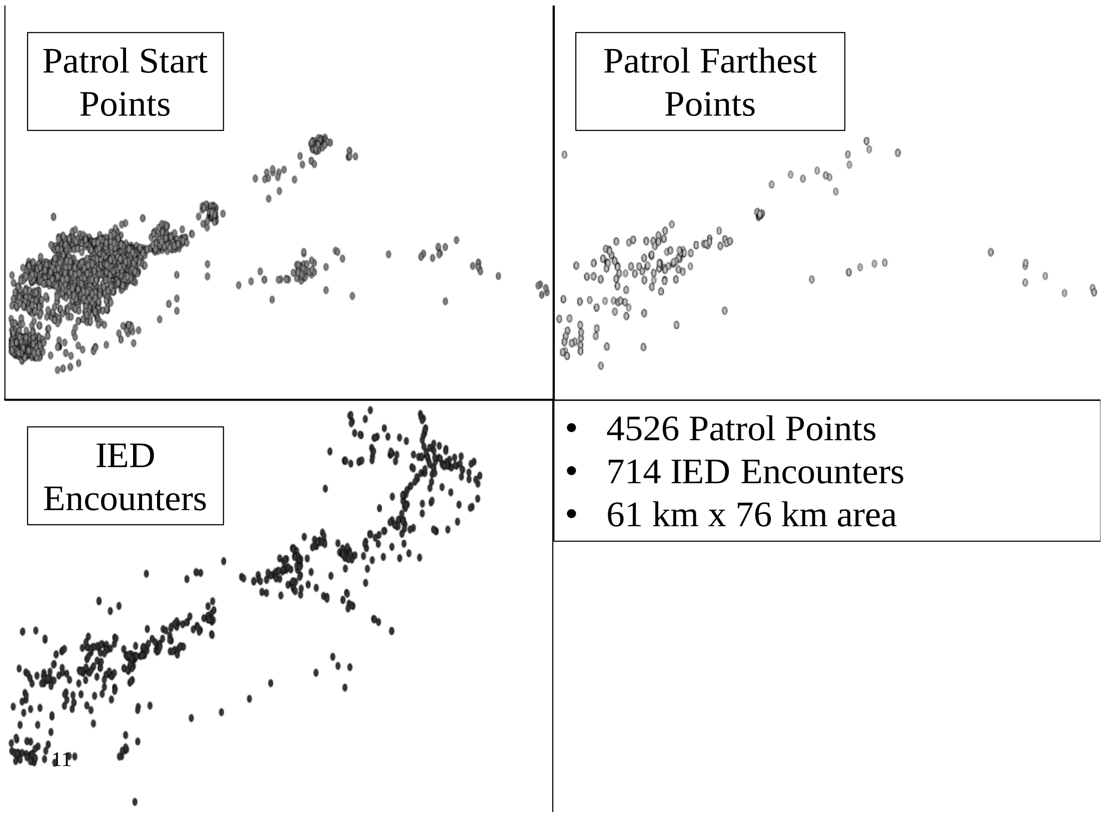

Figure 3: IED and Patrol Data Locations

There appear to be fewer farthest points than start points, but this portrayal is misleading. The dataset includes one start point and one farthest point for every entry. However, many of the farthest points are identical to each other. The patrol leaders probably patrolled to the vicinity of the same farthest point regularly. But rather than logging the entry accurately with a GPS, they almost certainly estimated the farthest point upon their return, resulting in the same response after numerous patrols. This is unlike the determination of start points, where the patrol leader deliberately used a GPS prior to departure.

Figure 4 portrays the distribution of distances traveled by each patrol to their farthest points. Of the 4526 patrols, 3357 traveled up to 1600 meters to their farthest points. In other words, most patrols did not travel more than one mile from their point of origin.

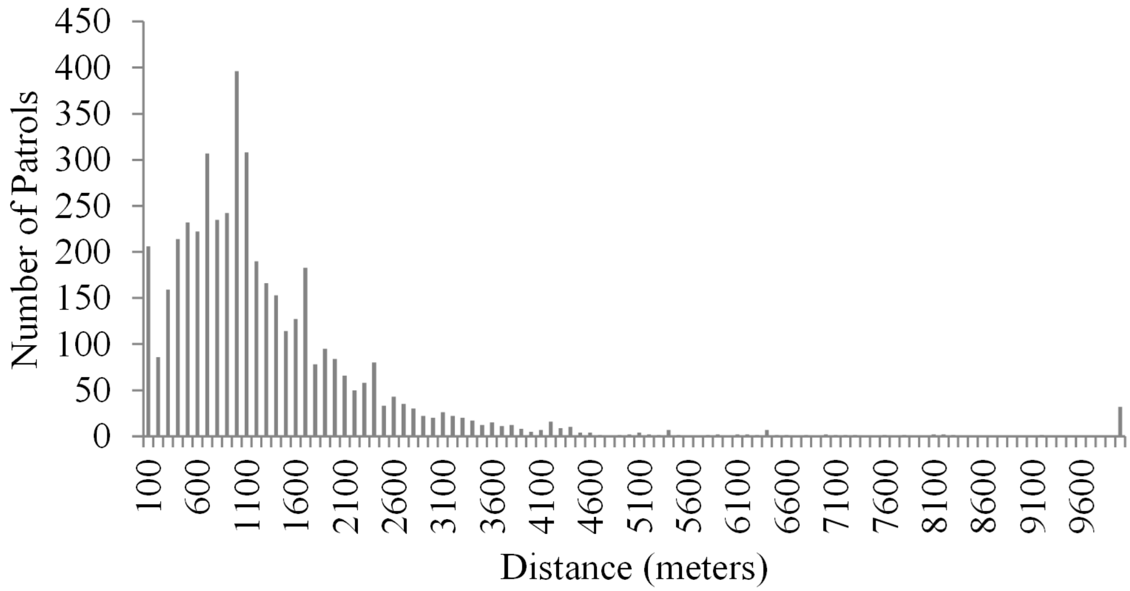

Figure 4: Distance from Patrol Start to Farthest Point

The data collection endeavor was a success on many terms. First, the author and his fellow analyst were able to immediately utilize the data to produce statistical products that summarized the data. Particular focus was placed on showing patterns of patrolling and IED encounters. At the time, a model was not yet created to estimate IED emplacements, though leaders were certainly interested.

The statistical products were provided directly to the local commander and his troops within 48 hours of the analysts receiving the data. The rapid response time combined with the customized analysis thrilled the unit and energized the tactical level leaders to continue to collect more accurate data for this cause. In practice, such analysis is usually held at the operational and strategic level for large scale decisions by senior officers. However, it was refreshing for analysts to make a direct impact with infantryman at the tactical level.

CHAPTER 4: ANALYSIS OF EVENTS IN TEMPORAL DOMAIN

This section provides an analysis of the IED encounter data in the time domain. First, we compare the encounter data to a stationary Poisson process with an equivalent encounter rate λ. The Poisson distribution is the benchmark for comparison throughout this dissertation because of its positive, discrete nature and simple mathematical structure. In the data, there are 714 encounters over 153 days, so the sample encounter rate is λ = 4.67 per day. The theoretical probability of observing *k* IED encounters in one day from an equivalent stationary Poisson process is $p_k = \lambda^k e^{-\lambda} / k!$. Figure 5 compares the observed data with the theoretical fit. The sample mean is 4.67 encounters per day, and the sample variance is 9.48. For a Poisson process, the mean and variance would be equal. Thus, the data do not support a Poisson fit. This is reinforced by the chi-squared test results $\left(\chi^2 = 88.63, \text{df} = 14, \text{p-value} = 0\right)$ that reject the null hypothesis of the observed and expected values coming from the same distribution.

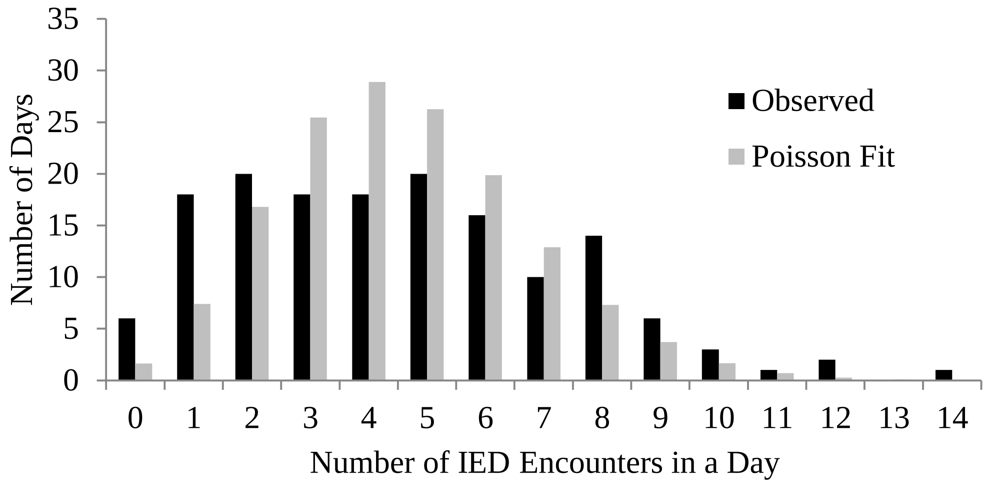

Figure 5: IED Encounters Fit to Stationary Poisson Process

Next, we compare the encounter data to a NHPP. Figure 6 shows the 3-day, 10-day, 20-day, and 30-day moving averages of the IED encounters over the time period. Three distinct periods of activity are identified, based on visual inspection of the encounter rates. Each period is then fit to an individual stationary Poisson process. The results are displayed in Figures 7-9.

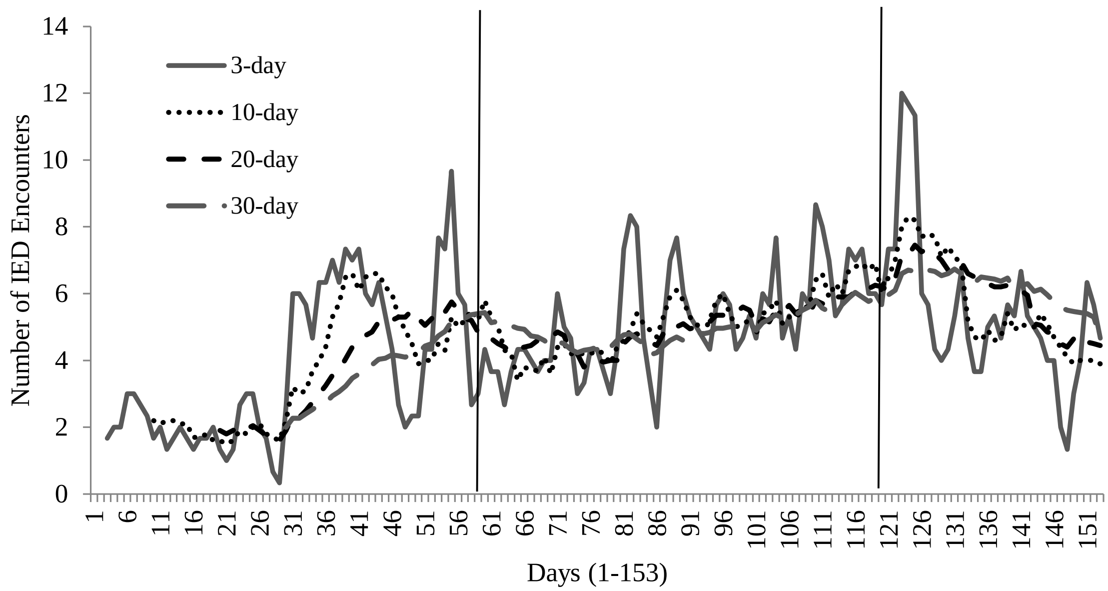

Figure 6: Moving Average of IED Encounters over Time

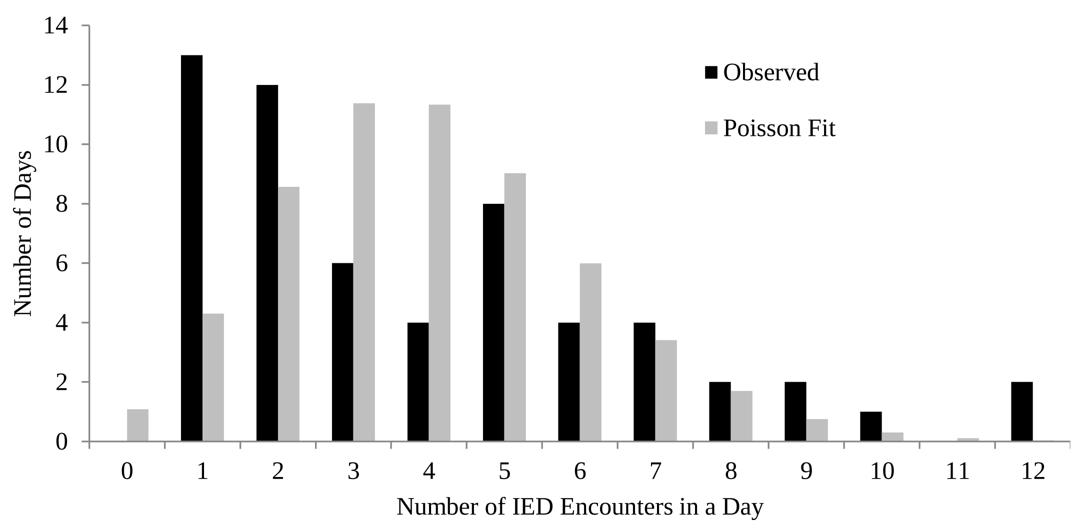

Figure 7: IED Encounters Fit to a Stationary Poisson Process (Days 1-58)

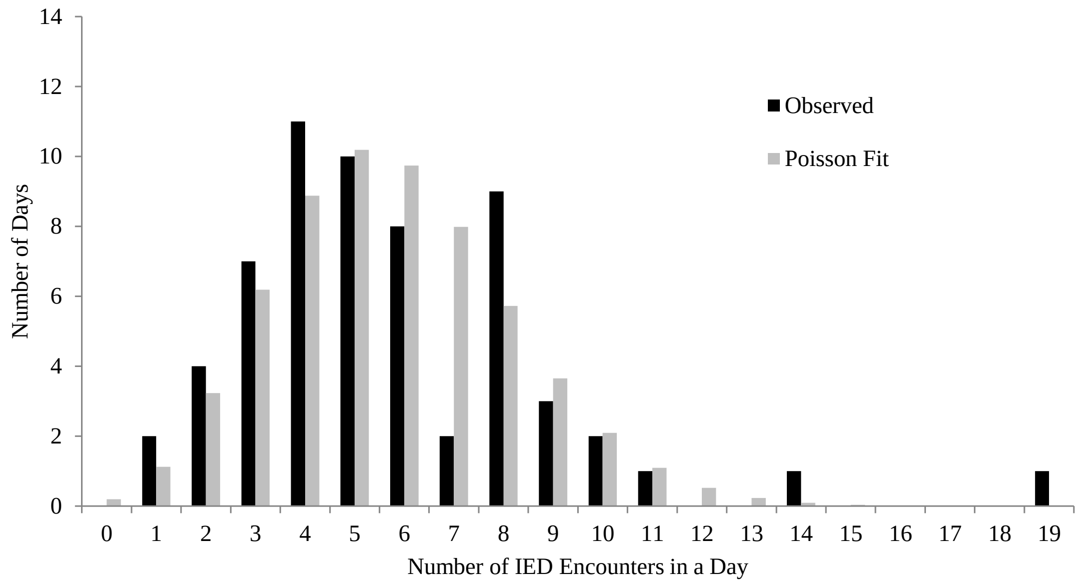

Figure 8: IED Encounters Fit to a Stationary Poisson Process (Days 59-119)

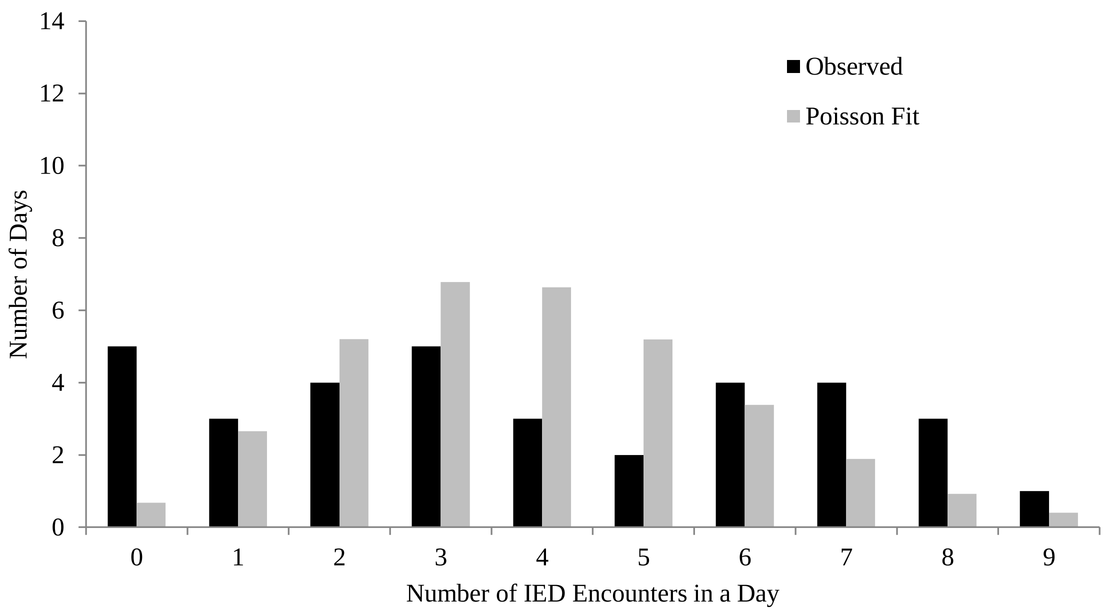

Figure 9: IED Encounters Fit to a Stationary Poisson Process (Days 120-153)

The fit for the middle period (days 59-119) is the best, but still do not follow a Poisson distribution. The sample mean (5.74) and the sample variance (9.70) do not match. Also, the result of the chi-squared test $\left(\chi^2 = 139.39,\ \text{df} = 12, \text{p-value} = 0\right)$ rejects the null hypothesis that the observed and expected values come from the same distribution.

Comparisons of IED encounter processes with NHPPs have also been conducted by other researchers such as Kolesar (2008, 2009). In these studies, it was concluded that IED encounter processes can reflect an NHPP. A key difference between our work and this work is that Kolesar (2008, 2009) used data from Iraq where the majority of IED incidents are targeted against vehicles on common roadways. Our study considers IED events against foot patrols in a two dimensional space. Foot patrols likely encounter IEDs at different rates and have more uncertain patrol routes than vehicular patrols due to the two dimensional nature of their movement. Also, encounters are subject to the timings and routes of patrols which may have less regularity in a two-dimensional space. Furthermore, none of the results were found to reflect a Poisson process. We further analyze the IED encounter data in the spatial domain for patterns and statistical characterizations.

CHAPTER 5: ANALYSIS OF EVENTS IN SPATIAL DOMAIN

This section provides an analysis of the IED encounter data in the spatial domain. One common metric used in geospatial analysis is the nearest neighbor metric – that is, the Euclidean distance between a given event and its closest neighboring event. This metric can be used to determine if spatial clustering is evident (Ripley, 1981). For a spatial Poisson process, the expected Euclidean distance from a given event to its nearest neighbor is $\frac{1}{2\sqrt{\rho}}$, where ρ is the expected number of points per unit of area (Dixon, 2013). The CDF of this distance is $F(x) = 1 - e^{-\rho\pi x^2}$. (For a spatial Poisson process, the number of events in a region of area A is a Poisson random variable with a mean of ρA; the numbers of events in disjoint regions are independent.) Figure 10 shows the nearest neighbor distribution observed in the data and the corresponding distribution for a spatial Poisson process. The value of ρ is estimated by taking the total number of IED events and dividing by the total area associated with square kilometer regions containing at least one IED event. The value of ρ is found to be 2.03 IEDs/km^2. A chi-squared test result $\left(\chi^2 = 1.12*10^6, \text{df} = 20, \text{p-value} = 0\right)$ rejects the null hypothesis that the observed data and Poisson results come from the same distribution. Furthermore, we perform the Clark-Evans test (p-value = 0) by assuming a rectangular region around the friendly force travel area and also reject the null hypothesis that the observed and expected nearest neighbor distances are equal (Dixon, 2013). Additionally, a large number of nearest neighbor values in the observed data are small. We conclude that clustering is evident, and a model that considers this characteristic is necessary to reflect the spatial patterns of the data.

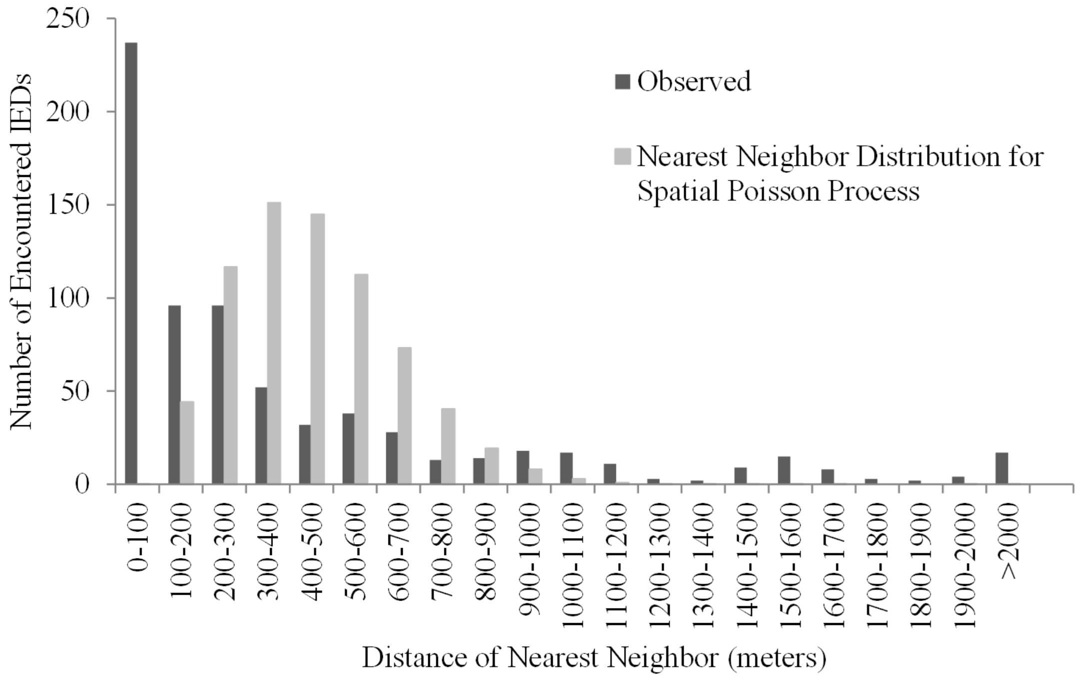

Figure 10: IED Encounters – Nearest Neighbor in Space

5.1 Spatial Clustering Model

We further investigate the degree of clustering in the data using a generalized version of the Neyman-Scott cluster process (Cowpertwait, 1997). In the Neyman-Scott model, a set of unobserved cluster centers are produced spatially according to a Poisson process. Then associated cluster members are formed randomly about the cluster centers.

Here we consider a non-stationary variant in which the expected number of events per cluster varies by location. Cluster members are assumed to be located a distance from the cluster center according to a Rayleigh distribution (Figure 11). We specifically make the following assumptions:

- The number of clusters in grid square k is a Poisson random variable with mean h. The locations of the cluster centers within grid square k are chosen according to a uniform spatial distribution. (That is, the cluster centers are chosen according to a stationary spatial Poisson process.)
- The number of events in cluster i (in grid cell k) is a Poisson random variable with mean λ_k/h . Thus, the average number of events in grid square k is λ_k.
- The location of event j relative to cluster center i in grid cell k is a random variable in polar coordinates (r, θ), where r follows a Rayleigh distribution with mean μ, and θ is uniformly distributed between 0 and 360 degrees. This is equivalent to a bivariate normal distribution in Cartesian coordinates.

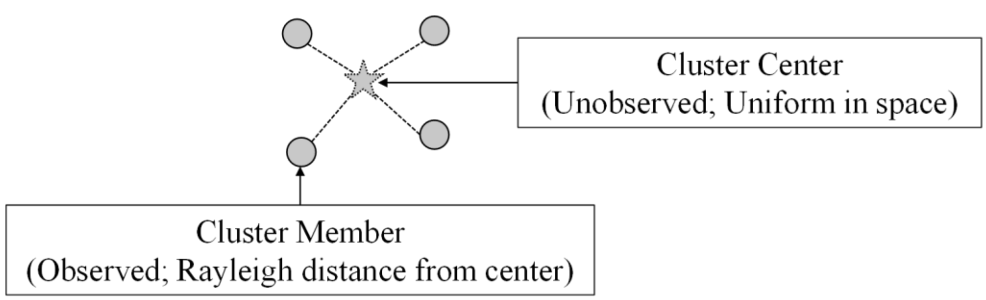

Figure 11: Cluster Structure

Table 4: Outputs to Spatial Clustering Model

Variable	Random Distribution
Number of clusters in grid square *k*	$Poisson(h)$
Position of cluster center *i* in grid square *k*	*Uniform* (in grid square)
Number of events in cluster *i* in grid square *k*	$Poisson\left(\frac{\lambda_k}{h}\right)$
Distance of event *j* from cluster center *i* in grid square *k*	$Rayleigh\,(\mu)$
Angle of event *j* from cluster center *i* in grid square *k*	$Uniform(0,2\pi)$

The model is specified by the parameters λ_k, *h*, and μ. The average number of events λ_k in grid square *k* is obtained from the data. The average number *h* of clusters in each grid square and the mean distance μ of cluster events from each cluster center are determined using the *k*-means algorithm. The *k*-means algorithm attempts to assign points to *k* clusters in order to minimize the sum of the squared distances of points from their cluster centers (the center of a cluster is defined as the centroid of points assigned to the cluster). Since the number *k* (in our case, *h*) must be pre-specified, a simple heuristic to choose *h* is to run the *k*-means algorithm for different cluster numbers *h* and then identify a point at which there are diminishing improvements in an error metric (Pham, 2004). For each value of *h*, the algorithm attempts to optimize the location of these cluster centers by minimizing the aggregate distance of the cluster centers to the cluster members. A global optimization is not guaranteed, and the initial seeding of the algorithm is random. The routine is run 10 times for each value of *h* ranging from 1 to 15 in each 5x5 grid square. Each run produces an output metric specifying the average distance of cluster events from each cluster center (μ) across all grid squares in the region. The output metrics of the 10 runs are averaged for each value of *h*. This is plotted in Figure 12 along with the difference in averages between each cluster quantity *h*. We aim to find the smallest *h* that, when increased further, shows little improvement (knee in the curve). Such a choice might be in the range of three to five, with a corresponding value of μ near 500m.

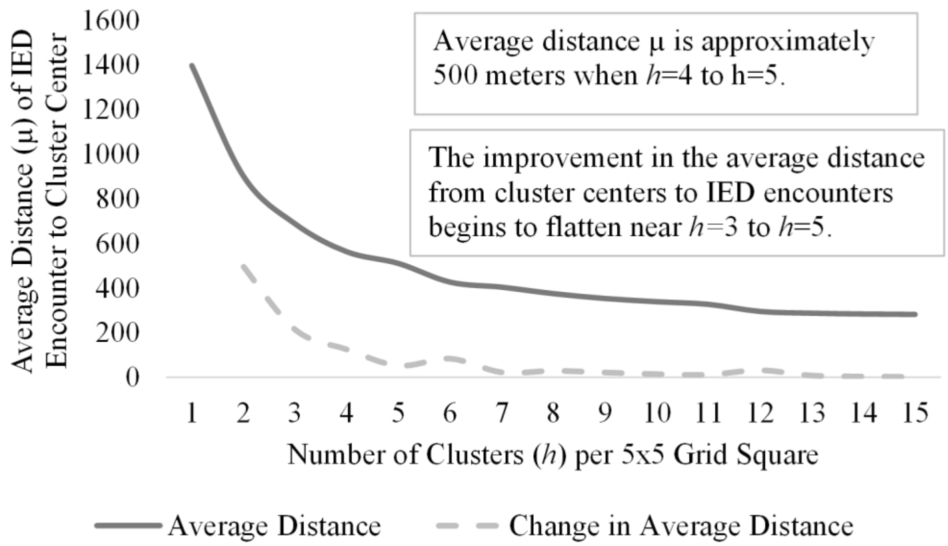

Figure 12: Optimal Number of Clusters as a Result of K-Means Computation

Figure 13 shows sample output from one run of the simulation compared to the actual data, using $h = 4$. The parameter input value agrees with the results of the earlier *k*-means calculations.

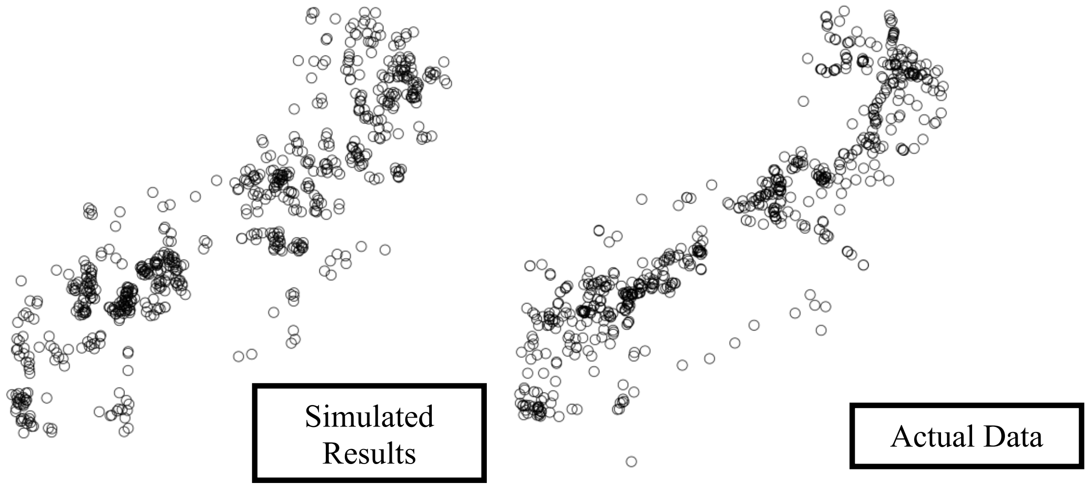

Figure 13: Sample Results of Spatial Clustering Model

To evaluate the performance of the model more rigorously, we use a chi-squared test. The simulation is run using five-kilometer grid squares, and the distribution of the quantity of events across each of the 25, nested, one-kilometer squares is recorded for each simulation (Figure 14). A measure of success is for a simulation methodology to output a distribution of events in one-kilometer squares that is similar to the collected IED encounter data. This similarity is quantified using the chi-squared test.

A simulation of the clustering model is run for five different values of the clustering parameter *h*. Each simulation is replicated ten times utilizing five-kilometer grid squares and an average 500-meter (μ) distance from the cluster centers. Table 5 shows the average results of seven different clustering scenarios. The quantity of events in each 1x1 square is totaled over all simulations and divided by 10. These averages are compared to the actual data in the far right column of Table 5, and the chi-squared p-value result for each scenario is calculated on the last row. The results indicate that the spatial cluster model outperforms a traditional NHPP or a homogenous Poisson Process. The model performs best when we assume an average of three or four clusters per grid square (*h*=3, *h*=4), as noted by the higher p-values. These parameter values agree with the results of the earlier *k*-means computations.

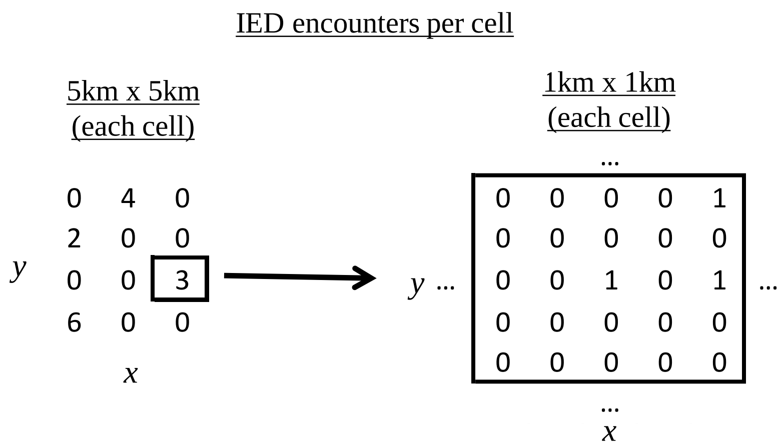

Figure 14: Breakout of IED Encounters from 5x5 to 1x1

Table 5: Quantitative Results of Spatial Clustering Model

Average of 10 Simulations								
5km x 5km Grid, μ=500m								
Events per 1x1	h=1	h=2	h=3	h=4	h=5	NHPP	PP	Actual
1	86.40	151.30	192.20	226.40	244.30	269.50	349.70	214
2	35.80	53.80	67.60	69.10	79.50	100.20	108.20	60
3	17.60	26.20	31.50	32.00	33.30	40.20	30.60	27
4	10.70	18.40	14.80	16.20	13.90	16.30	7.10	22
5	7.00	8.40	8.50	8.90	7.80	4.50	2.00	9
6	5.50	5.50	5.00	4.70	3.80	2.40	0.80	5
>6	11.00	13.00	11.90	8.50	8.00	1.40	0.10	14
P-value	0.00	0.00	**0.36**	**0.35**	0.01	0.00	0.00	

The previous analysis characterized the spatial attributes of the IED encounter data. The estimation of IED locations is an essential requirement for commanders to plan safer and more effective patrolling efforts. However, more robust conclusions can be drawn if the time of enemy IED emplacements were known. The following section describes such a model.

CHAPTER 6: EMPLACEMENT CALCULATION MODEL

The most direct way to model IED activity is by observing IED encounters. While emplacement times may be of more interest, they can only be inferred indirectly. This section gives an emplacement time model that uses both IED encounter data and friendly force patrol data to calculate the approximate time that an encountered IED was emplaced. Sample results are given applying the model to the collected data set. Finally, a sensitivity analysis is applied to understand the sensitivity of the model to input parameters and the quality of the input data.

When studying IEDs along road networks, a common assumption is that an IED could not have been emplaced any earlier than the last time a patrol was in that area – in other words, a patrol that passes by an emplaced IED is guaranteed to encounter it (Koyak, 2009a). This assumption may be reasonable in a road network since vehicles travel along a linear route with little room for deviation. However, foot mobile troops patrol two dimensional areas, particularly in urban locations. Depending on the scenario, foot patrols can either begin their travel from their base of departure or dismount from a vehicle. Since the patrol paths are not exact, it cannot be guaranteed that a patrol that was in the vicinity of an IED would have encountered it. This assumption is relaxed in this dissertation, so that each patrol has a certain probability of encountering an IED over some area of coverage.

The patrol data in this dissertation include a start point and farthest point. Since the precise path is unknown, each patrol is assumed to cover an area between the start and the farthest point in the shape of an ellipse with some specified width (Figure 15). An IED that is emplaced within the ellipse is assumed to be encountered with some probability. The actual track may extend beyond the boundary of the ellipse.

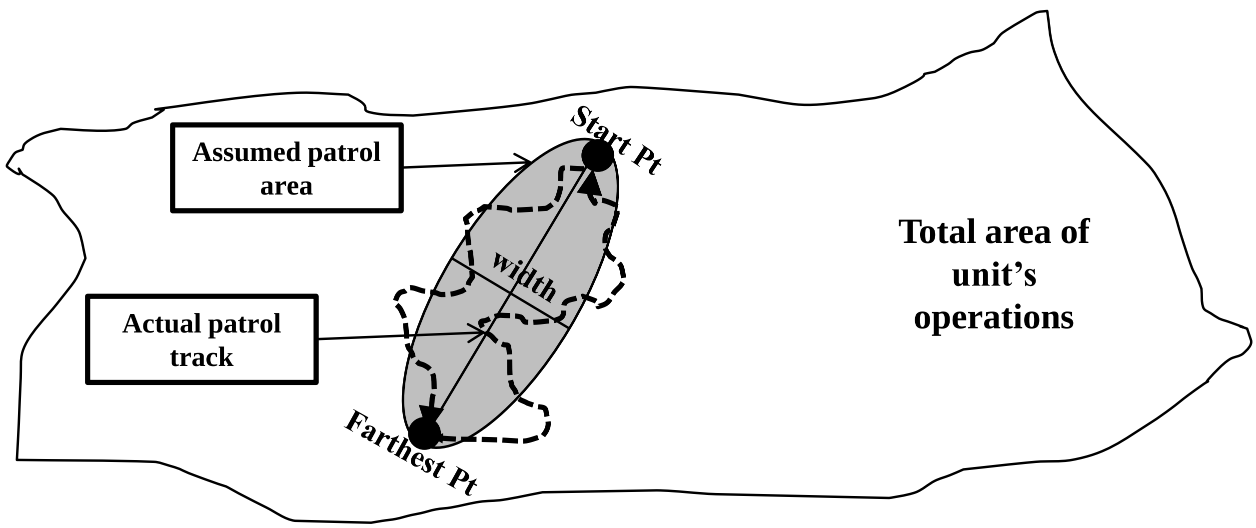

Figure 15: Notional Patrol Track and Assumed Patrol Ellipse

The following assumptions are used in the IED emplacement model: First, each IED encounter must be assigned to a patrol. Since the encounter data and patrol data are separate, the most recent patrol whose ellipse overlaps the location of the IED encounter is assumed to be the patrol that encountered the IED. Second, if zero or one patrol traversed the area of an IED encounter, the IED encounter is excluded from the analysis and is not assigned an emplacement date. This is because a second patrol is needed to provide a non-arbitrary lower bound on the emplacement time. Third, it is assumed that a patrol i, whose ellipse overlaps an IED j, encounters that IED with probability $p(i, j)$. (This is assuming that the IED was emplaced prior to the start time of the patrol and that the IED has not been previously discovered by an earlier patrol.) The probability is assumed to be a linear function of distance from the straight-line path between the start and farthest points; see equation (1) below. This implicitly assumes that patrols are more likely to travel along a direct path between the start and farthest points. Finally, it is assumed that the underlying emplacement process is a Poisson Process. Simplicity in mathematical structure is the primary driving force behind this assumption.

To identify IED emplacement times, the following steps are conducted for each IED encounter j:

1) Identify the set of patrols $R = \{1, 2, \ldots, m\}$ whose ellipses overlap the IED encounter and whose start times occurred prior to the IED encounter. The patrols are assumed to be indexed in chronological order, and the most recent patrol that overlapped the IED encounter event is assumed to be the patrol that encountered the IED. Let A_{ij} be the event that IED j was *emplaced* between patrols i and i+1; $i = \{1,2,\ldots, m-1\}$. Let B_j be the event that IED j was encountered by the most recent patrol m; $m > 1$. Also, let d_{ij} be the distance of IED j from the axis formed between the start and farthest point of patrol ellipse i, and let w be the width of the patrol ellipse. Then

$$p(i, j) = 1 - \frac{d_{ij}}{w/2} \qquad (1)$$

2) Calculate the probability that IED j was encountered by patrol m, given that the IED was emplaced between patrols i and i+1; i ={1,2,…, m–1}:

$$\mathrm{P}(\mathrm{B}_j \mid \mathrm{A}_{ij}) = \begin{cases} p(m, j) * \prod_{k=i+1}^{m-1} \left(1 - p(k, j)\right), & \text{if } i < m - 1 \\ p(m, j), & \text{if } i = m - 1 \\ 0, & \text{if } \mathrm{i} > m - 1 \end{cases} \qquad (2)$$

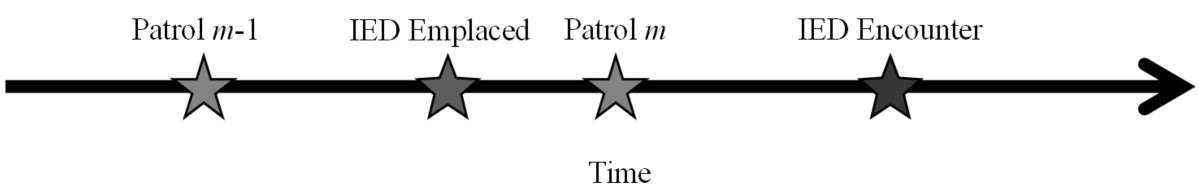

Figure 16: Emplacement Timeline

3) Calculate the probability that IED j was emplaced between patrols i and i+1:

$$\mathrm{P}(\mathrm{A}_{ij}) = 1 - e^{-\lambda(t_{i+1} - t_i)}, \qquad i = \{1,2,\ldots, m\text{–}1\}. \qquad (3)$$

where λ is the average daily rate of IED events and t_i is the start time of patrol i. Since the number of emplaced IEDs is fixed based on spatial characteristics, altering λ does not change the emplacement calculation results.

4) Calculate the probability that IED j was encountered by the most recent patrol (Law of Total Probability):

$$P(B_j) = \sum_{i=1}^{m-1} P(B_j \mid A_{ij}) * P(A_{ij}) \tag{4}$$

5) Calculate the probability that IED j was emplaced between patrols i and $i+1$, given that it was encountered by the most recent patrol m (Bayes Rule):

$$P(A_{ij} \mid B_j) = P(B_j \mid A_{ij}) * P(A_{ij}) / P(B_j), \qquad i = \{1, 2, \ldots, m-1\}. \tag{5}$$

6) Input all m-1 probabilities calculated in step 5 into a random sampling algorithm and draw one sample that represents a single patrol interval [i, i+1] in which the IED was emplaced. Select a uniform random point within the interval to declare the exact estimated emplacement time.

Numerical Results

To apply the model to the collected data, the first step is to identify patrols whose ellipses overlap the IED encounter locations. In the original data set, there are 4,526 foot patrols and 714 IED encounters. Of these, 1,381 patrols have ellipses that directly overlap at least one IED event (and where the start time of the patrol is before the IED event time). Thus, a lower bound on the probability that a patrol encounters an IED that has been emplaced within its ellipse is 714 / 1381 ≈ 0.52. (The lower bound is realized in a scenario in which all IEDs are emplaced prior to the start of the data collection.) The model only considers IED events that are overlapped by two or more patrols that departed previous to the IED encounter. Of the 714 IED events, there are only 173 such events. Of these 173 events, 24 are overlapped by two patrols, 23 are overlapped by three patrols, and the rest are overlapped by four or more patrols. It is expected that IED encounters occurring at the beginning of the time period are overlapped by fewer patrols than those occurring at the end of the time period.

Figure 17 shows an application of the model to data. The figure shows IED encounters and patrols over time (from historical data) as well as inferred IED emplacements (from the model). Emplacements are shown as long-term averages obtained by adding up the long term probabilities of IED emplacements in each day from step 5 in the algorithm.

IED events require an enemy emplacement and a friendly force encounter. Thus, emplacements, friendly patrols, and encounters are all related. The hypothesized correlations appear to hold true in Figure 17. The rise and fall of IED events correlates with the rise and fall of friendly patrols, as shown during days 40-90 and 91-120. The variation in IED emplacements is related to the future variation in IED encounters. This can be seen with the rise in IED emplacements at Days 20 and 56 along with the rise in encounters at Days 40 and 70.

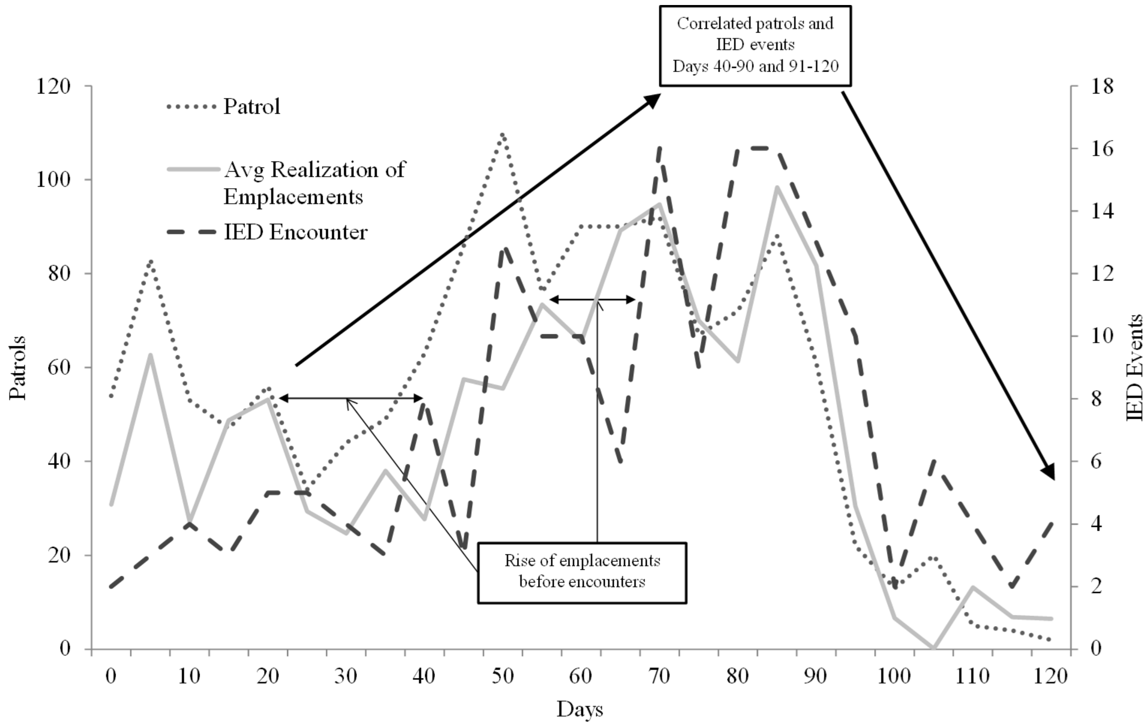

Figure 17: Patrols and IED Events by Day

The IED emplacements do not fit a stationary Poisson process over time (Figure 18). We can see visually that the observed and expected emplacements generally do not equal each other. The results of the chi-squared goodness of fit test agree with this assessment, noting a probability of almost zero that the observed and expected values come from the same distribution. The mean and variance are also not equal.

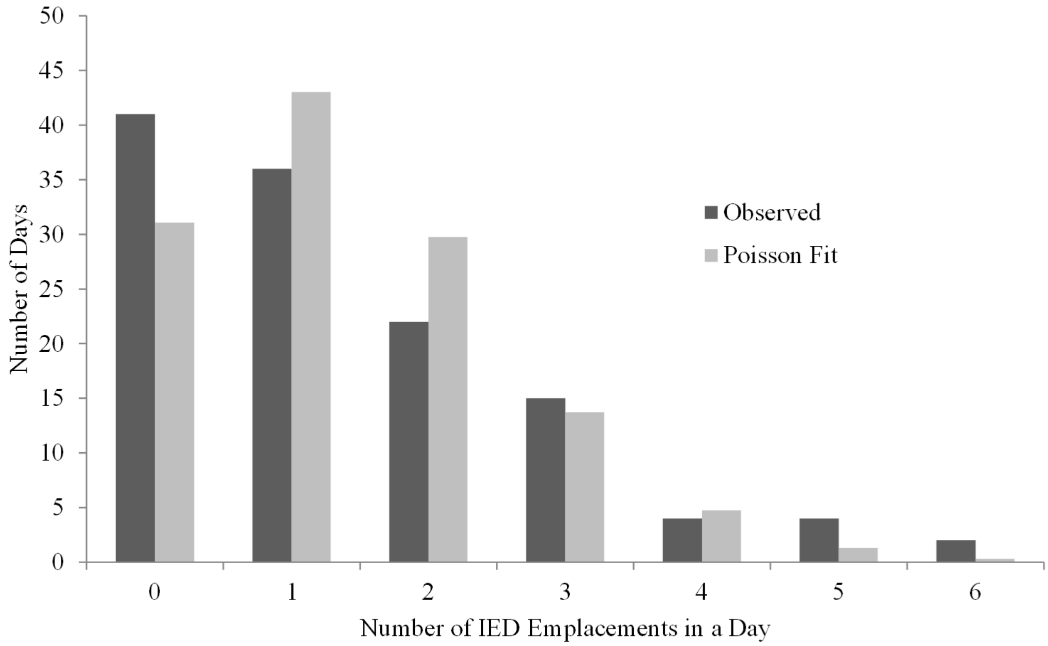

Figure 18: IED Encounters Fit to Stationary Poisson Process

Next, the emplacement results from one simulation are fit to a NHPP utilizing the method stated earlier for IED encounters. The moving average of IED emplacements is displayed in Figure 19. In viewing the 10, 20, and 30 day moving averages, there appears to be three distinct periods in which IED emplacement rates remain constant. Data in each of the following time periods are fit to Poisson distributions – days 1-55, 56-95, and 96-123. The results are shown in Figures 20-22.

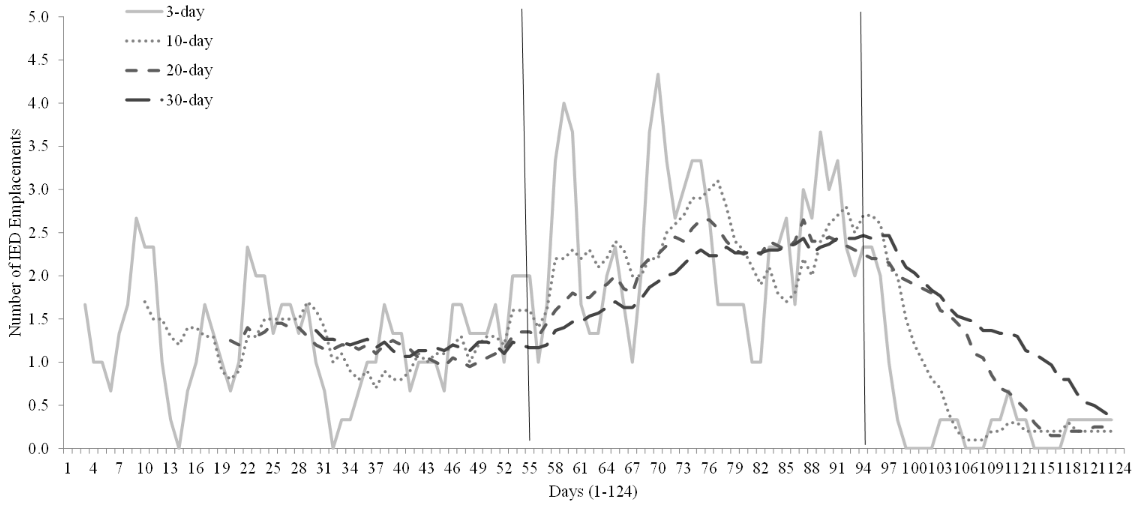

Figure 19: Moving Average of IED Emplacements over Time

Each period is fit to a separate Poisson distribution to determine if the aggregate forms an NHPP. Chi-squared tests were performed on each period to determine the probability of the data resembling a Poisson distribution. The results from the first period ($\chi^2 = 1.48$, df=4, p-value=0.83), the second period $\left(\chi^2 = 5.68, \text{df} = 6, \text{p-value} = 0.46\right)$, and the third period $\left(\chi^2 = 0.16, \text{df} = 1, \text{p-value} = 0.69\right)$ do not allow us to reject the null hypothesis that the data reflect a NHPP. We assume the underlying emplacement process to be a Poisson process in Equation 3, so this result is to be expected and further validates the model.

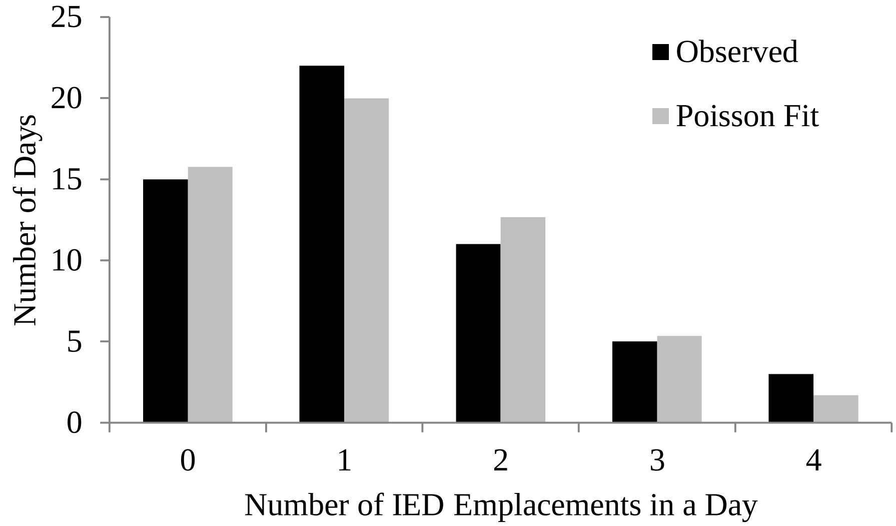

Figure 20: IED Emplacements fit to a Stationary Poisson Process (Days 1-55)

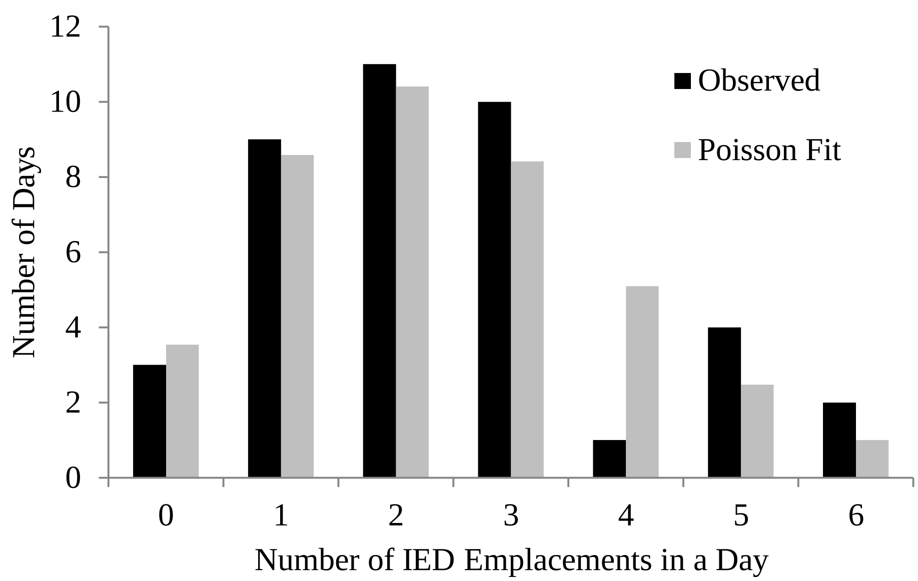

Figure 21: IED Emplacements Fit to a Stationary Poisson Process (Days 56-95)

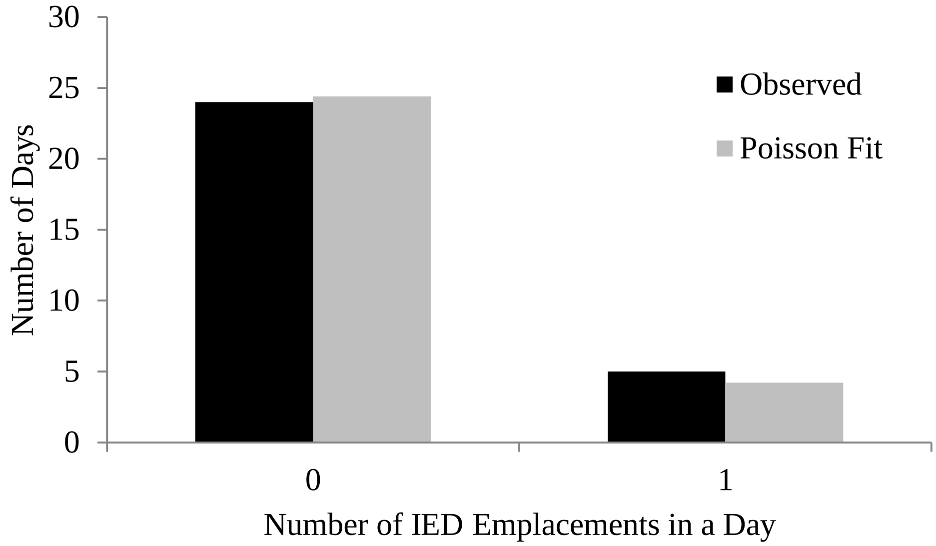

Figure 22: IED Emplacements Fit to a Stationary Poisson Process (Days 96-124)

The number of IED emplacements over time in one simulation of the model is analyzed to determine if temporal clustering is present. The quantity of IED emplacements within 10 and 20 days of each IED emplacement is compared to a simulated Poisson distributed dataset in Figures 23-24. The actual data should exhibit the Poisson distributed simulated dataset if there is no clustering. However, the results show no relationship between the actual and simulated datasets, implying the possibility of clustering in time.

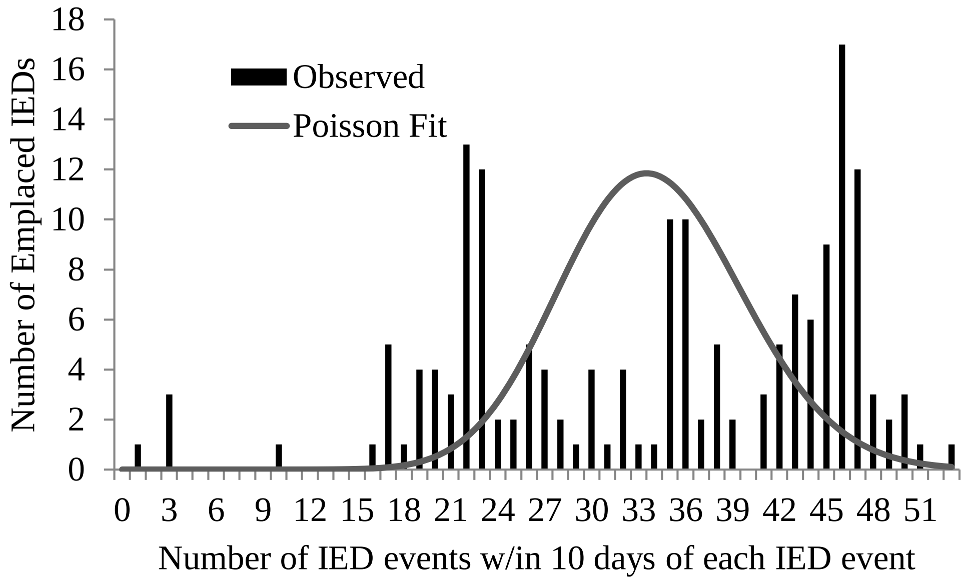

Figure 23: Number of IED Emplacements within 10 Days of Each IED Emplacement

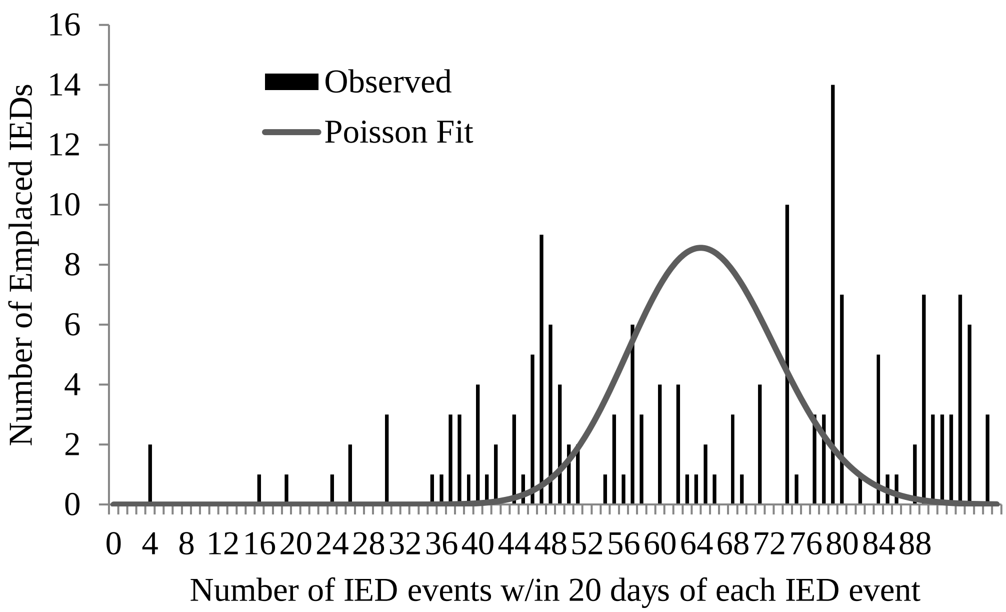

Figure 24: Number of IED Emplacements within 20 Days of Each IED Emplacement

The calculated time lapse between the emplacement of an IED and its encounter can add validity to the emplacement algorithm. The commanders from the region of the collected dataset implied that a fairly constant patrolling effort occurred throughout the area of operations, with key areas being patrolled daily and most other areas weekly or bi-weekly. This implies that the lag between most IED emplacements and encounters ranged between 1 and 14 days. Figure 25 shows the difference between the IED encounter time and the IED emplacement time (as determined by the emplacement model). The figure shows that 77% of the IED events were emplaced in this range and are therefore consistent with this observation.

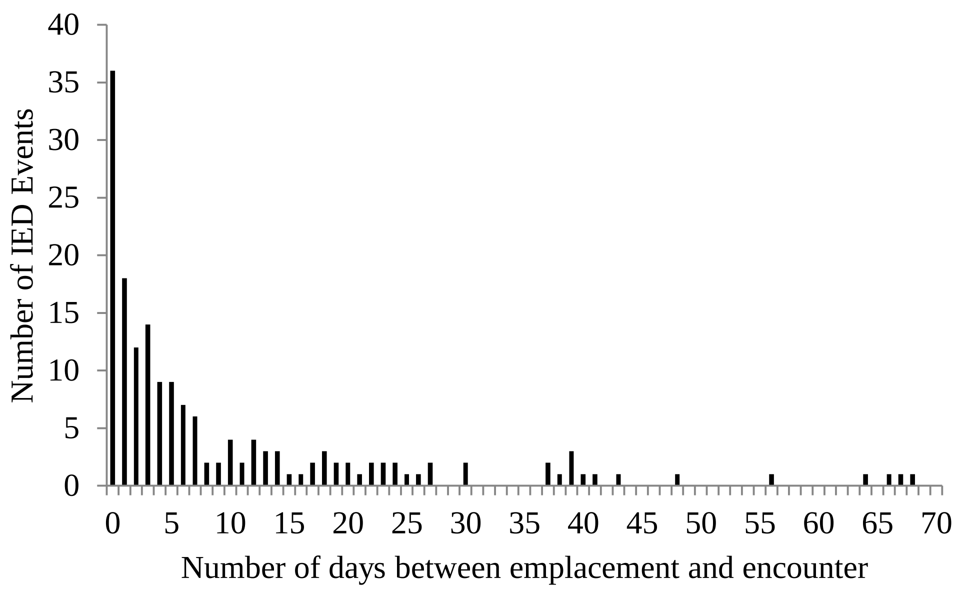

Figure 25: Time lapse between modeled emplacement time and encounter

A consideration for the sensitivity of changes to key input parameters against the output of the model brings clarity to the flexibility and robustness of the model.

Specifically, earlier described shortfalls within the collected data prompt a consideration for the effect of more precise data on the output of the model. These points are explored in the following section.

6.1 Sensitivity of the Model to Hypothetical Data

The previous emplacement results were calculated using a non-uniform probability of encounter that is linearly related to a given IED's distance from the axis formed between the start point and the farthest point of the overlapping patrol ellipse. Figure 26 depicts additional results using fixed probabilities of 1.0, 0.50, and 0.05. The results of the chi-squared goodness of fit test for all four values of *p* in Figure 26 show no conclusive evidence that the observed values come from a Poisson distribution.

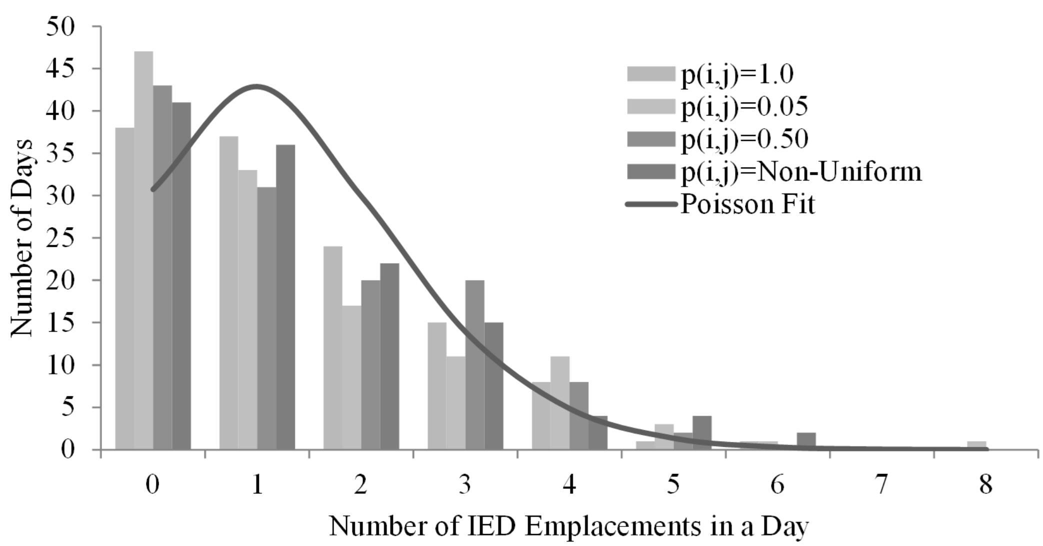

Figure 26: IED Emplacements Fit to Stationary Poisson Process

Figure 27 shows a sensitivity analysis based on varying the width of the patrol ellipse. A larger sized ellipse places more patrols in each IED event area and therefore causes more IEDs, particularly in the early portion of the time period, to be assigned emplacement dates. The overall temporal trends of IED emplacements remain the same regardless of the ellipse size. Since subject matter experts agree that 500 meters is the proper estimate, so there is no evidence from the sensitivity analysis to conclude otherwise.

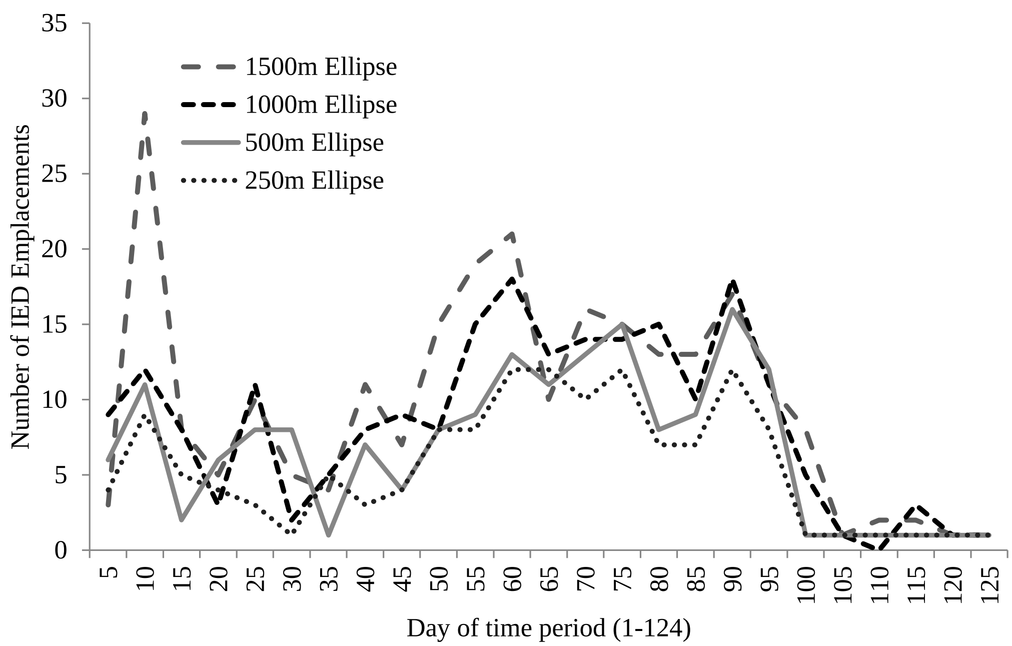

Figure 27: Ellipse Width Sensitivity

Finally, we conduct a sensitivity analysis on the quality of the patrol data. The collected friendly force data does not define patrol routes with high fidelity. Only the start points and farthest points are known, so there is considerable uncertainty in the paths. The objective here is to determine whether precise track information would yield significantly improved results via a similar emplacement model. To do this, a set of hypothetical IED and patrol data are created. We allow for sensitivity in the number of patrols to determine if such variation would have a significant effect on the results.

Figure 28 shows the overall methodology. First, hypothetical patrol tracks and IED emplacement data are generated as "truth" data. Second, IED encounter times are inferred from the track and emplacement data. Next, the emplacement model is run to estimate IED emplacement times. Lastly, the estimated and actual emplacement times are compared.

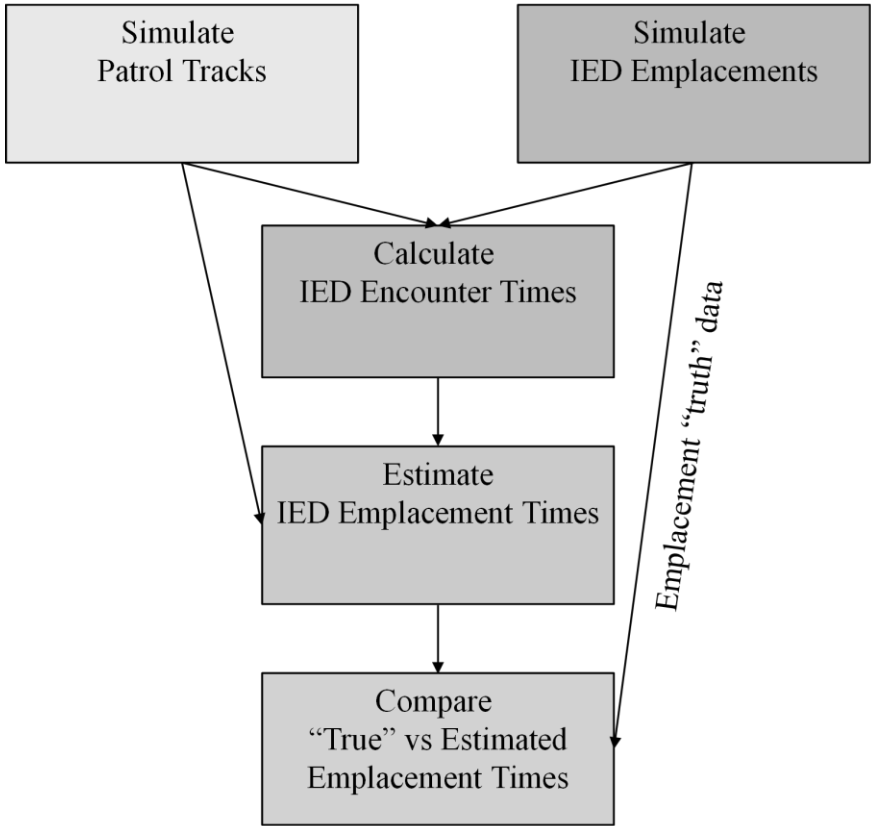

Figure 28: Improved Emplacement Calculation Model Workflow

Hypothetical patrol tracks are generated as follows. We assume that patrol tracks consist of multiple straight-line segments. A patrol track is established by joining five segments of lengths that range from 0 to 500 meters, randomly chosen from a uniform distribution. Segments are joined at angles between -90° and 90° relative to the previous segment, to model a general direction of travel by the patrols. Patrols are assumed to retrace their steps in the reverse direction. Start points are uniformly distributed in space across the study region. A commonly known mathematical formulation for this path is the random walk with retrace, since the path is a sequence of random steps and the return route of the patrol matches the outbound route. Figure 29 portrays an example of a segmented patrol track, and Table 6 outlines the data parameters for the track data. If a track exceeds the region's initially specified boundary as stated in Table 6, the region is simply expanded to accommodate the new data.

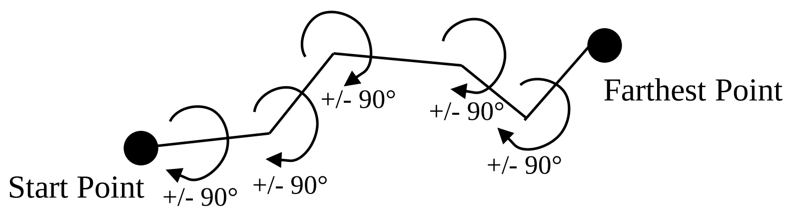

Figure 29: Patrol Track Example

Table 6: Patrol Track Data Parameters

Hypothetical Patrol Track Data	
Start Times	Uniformly distributed over a period of 121 days
Start locations	Uniformly distributed within an approximate 7.62 km x 6.13 km space.
Number of segments	5
Length of each segment	Uniformly distributed from 0m to 500m
Angle of each segment	Uniformly distributed from +/- 90 degrees, where angles are relative to previous track assignment
Number of patrols	5,000 or 10,000 - based on sensitivity analysis

The creation of IED emplacement data is also required for the model. A quantity of IED events is chosen and "true" emplacement times and locations are assigned. Each event location is uniformly assigned within the study space, and each emplacement time is uniformly assigned within a predetermined time period. This data would be unavailable in a real-world scenario, but the information is used here as a reference for comparison of the calculated emplacement times. Parameters for the data are outlined in Table 7.

Table 7: IED Emplacement Data Parameters

Hypothetical IED Emplacement Data	
Emplacement times	Uniformly distributed over a period of 121 days
Locations	Uniformly distributed within an approximate 7.62km x 6.13km space
Number of events	A constant value

IED encounter times are calculated with the underlying assumption that patrols can encounter more than one IED, and IEDs can only be encountered by one patrol. For each patrol, all IEDs are identified that are within 100m of the segmented patrol track and emplaced before the start of the patrol. An actual encounter probability is provided to determine whether or not a patrol encounters an IED on its path. Once all of the IEDs are encountered or all of the patrols are complete, the algorithm is complete. It is likely that not all emplaced IEDs are encountered.

The original model used patrol data that only had start points and farthest points. IED emplacement calculations relied upon an assumption that the area patrolled by a given patrol was within an ellipse. Here, we also consider patrol tracks for which the entire path is known. The improved track data presented in this section employs a 100m buffer around each segment of the patrol that forms the area assumed to be traveled by the patrol. Figure 30 displays a comparison of the two concepts.

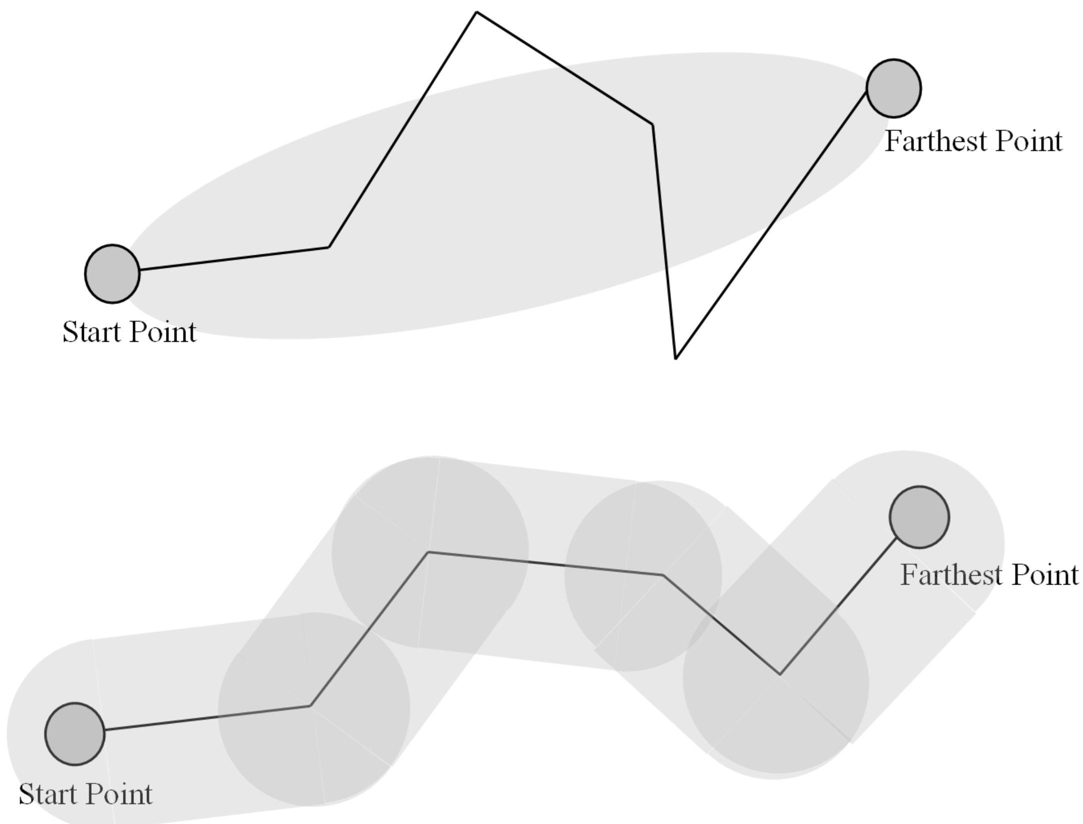

Figure 30: Comparison of Ellipse and Segmented Patrol Track Boundaries

Finally, a comparison of the difference between the “true” and calculated emplacement times is provided as a metric for the accuracy of the model. The model is run and sensitivity is performed on two parameters – patrol area shape (ellipse or segmented line) and number of patrols. The *assumed* encounter probability ($p(i, j)$ in Equation 1) is the probability *within the emplacement model* that an IED is encountered by a given patrol. The *actual* encounter probability is the probability used to generate the truth encounter data in Figure 28. We assume both to be described by the same non-uniform probability function in Equation 1.

Table 8 shows the full factorial design and resulting performance of the model, as measured by the average lapse between the calculated and actual emplacement times. As might be expected, the best results are achieved using a larger number of patrols with full track information (segmented line tracks).

Table 8: Full Factorial

Patrol Data	Qty of Patrols	Average Lapse b/w calculated and actual emplacement time
Estimated Region	1250	24.01
Estimated Region	5000	15.63
Estimated Region	10000	9.54
Actual Track	1250	26.08
Actual Track	5000	11.72
Actual Track	10000	6.17

Table 9 shows the main effects observed from the full-factorial design. Though increasing the assumed encounter probability generally lowers the lapse time, this parameter should still be chosen without bias for the most accurate results. As stated before, increasing the number of patrols and utilizing the segmented track data greatly increase the effect of the model. Data collection efforts within the field should be targeted at these factors for the best calculation results.

Table 9: Main Effects

Parameter	Value	Average Lapse (days)	Main Effect (days)
Patrol Shape	Estimated Region	16.39	-1.74
	Actual Track	14.66	
Number of Patrols	1250	25.04	-11.37
	5000	13.68	
	5000	13.68	-5.82
	10000	7.85	

A commitment to increasing patrols or better defining patrol tracks has a tradeoff for commanders. In the case of the first, leaders must either increase the number of troops to raise the patrolling effort or further tax the troops that are presently on the ground. However, better patrol track data simply requires carrying a portable GPS device. This may be an added expense to wartime efforts, but several units already possess such devices. They simply need to be employed in a way that is focused on archiving useful data. Table 9 shows that doubling the patrolling effort or improving the patrol tracks both have nearly the same benefit. Commanders should initially focus on the data collection and then raise patrolling efforts if conditions allow.

CHAPTER 7: CONCLUSION

The data collection effort was a key factor to the success of this research. During a deployment to Afghanistan, the first author and a fellow analyst developed the necessary relationships, spreadsheet tools, and data collection procedures to aggregate an event repository of over 4,000 military foot patrols and over 700 IED events. This unique endeavor provided real-world data from a recent military campaign that has not been known to exist in such a form until now.

The IED encounter data used in this research did not clearly resemble a Poisson Process or NHPP in either space or time. Consequently, a spatial cluster model was developed that “recreated” the original IED encounter data with high fidelity and few input parameters. Additionally, an IED emplacement time calculation model was established and exploited the collected data from Afghanistan to produce results. Greater amounts of more precise data were simulated and run against the existing emplacement model and benefits were determined. Results were markedly improved when the granularity of the patrol track data was improved and the number of foot patrol entries was increased.

The models developed in this dissertation will be provided to military analysts in Afghanistan in hopes to eventually assist commanders with estimating the threat of IEDs against their troops before they depart friendly lines. Direct applications during steady state operations are most likely. In this scenario, analysts can use historical data to estimate future enemy emplacement activity with confidence. Unmanned aerial overflights, route clearance scheduling, and intelligence targeting can be improved with this trend information. We also expect the research to encourage the cooperation of commanders towards more precise and consistent patrol data collection.

Future research should examine the potential relationship between IED events in space and time. Attempts at decreasing the lapse between actual and calculated emplacement times should also remain in focus. Presently, the models produce results strictly based on inputted historical data, but modifications to the inputs can greatly increase the forecasting power of these models in a variety of scenarios. Other applications of the spatial and temporal models developed in this dissertation should be examined against air and naval military operations, as the study of friendly and enemy movement is not limited to ground forces alone.

The application of these models need not be limited to IED activity. The IED emplacement methodology has particular application in activities within science, engineering, and sociology that have unobservable origins but observable effects. For instance, the contraction of a disease is usually only known after the symptoms arise, but the time of the infection is usually not observed. Similarly, the seeding of plants, origins of crime activity, changes in consumer behavior, and even assembly line process errors all have unobserved origins that can potentially be calculated using a methodology similar to that used in this dissertation to calculate IED emplacement times. The spatial clustering model has specific benefits over a traditional spatial NHPP in that it more accurately characterizes the spatial relationships between IED events. The model correctly considers that IED activity does not occur uniformly in space. When inputs are modified properly, the spatial clustering model can forecast IED activity in a given region that can greatly assist intelligence targeting and IED clearance efforts. The development of both the temporal and spatial models combine to form a unique advancement in the research of IED activity.

REFERENCES

Baum, D., "Spatial Generalization of BMAPs with Finite State Space", Journal of Mathematical Sciences, Vol. 105, No. 6. 2001.

Boba, R., "Crime Analysis and Crime Mapping", Sage Publications. 2005.

Breuer, L., "An Introduction to Queuing Theory", Springer. 2005.

Chakravarthy, S.R., "The Batch Markovian Arrival Process: A Review and Future Work", Department of Industrial and Manufacturing Engineering and Business, Kettering University.

Cooper, D.C., "Compatibility of Land Search and Rescue Procedures with Search Theory", Potomac Management Group. December 2003.

Cordesman, A.H., "IED Metrics for Afghanistan, January 2004 - May 2010", Center for Strategic and International Studies. July 2010.

Cowpertwait, P.S., "A Regionalised Neyman-Scott Model of Rainfall with Convective and Stratiform Cells", Hydrology and Earth Systems Sciences. 1997.

Curtin, K.M., "A Comprehensive Process for Linear Referencing", URISA Journal, Vol. 19, No. 2. 2007.

Curtin, K.M., "Road Network Analysis and Linear Referencing", GMU/JIEDDO Annual Technical Report. Sept 2009.

Curtin, K.M., "Linear Referencing for Network Analysis of IED Events", Department of Geography and GeoInformation Science, George Mason University. Sept 2009.

Dachian, S., "Hypothesis Testing: Poisson vs Self Exciting", Scandinavian Journal of Statistics. 2006.

Daley, D.J., "An Introduction to the Theory of Point Processes, Volume 1: Elementary Theory and Methods", Springer. 2003.

Daley, D.J., “An Introduction to the Theory of Point Processes, Volume 2: General Theory and Structure”, Springer. 2008.

DeGregory, K.W., “Optimization-Based Allocation of Force Protection Resources in an Asymmetric Environment,” Sloan School of Management. June 2007.

Dixon, P.M., “Nearest Neighbor Methods”, Encyclopedia of Environmetrics. Jan 2013.

Gross, D., Shortle, J.F., Thompson, J.M., Harris, C.M., “Fundamentals of Queuing Theory, Fourth Edition”, Wiley. 2008.

Johnson, N., “Dynamic Red Queen Explains Patterns in Fatal Insurgent Attacks”, The MITRE Corporation, Complex Systems Center. Mar 2011.

Joint IED Defeat Organization (JIEDDO). Reston, VA. Deployed Analyst Products, Archives. 2010.

Joint Warfare Analysis Center (JWAC). Dahlgren, VA. Deployed Support Products, Archives. 2011.

Klein, J.P., “Survival Analysis, Techniques for Censored and Truncated Data”, Springer. 1997.

Kolesar, P.J., “Poisson Trending of IED Event Frequencies”, GMU/JIEDDO Annual Technical Report. Sept 2009.

Kolesar, P.J., “Time Series Analysis of Improvised Explosive Device Incidence”, GMU/JIEDDO Annual Technical Report. Sept 2008.

Koyak, R.A., “Risk on Roads: A Modeling Approach, Part 1: Estimation Models”, Naval Postgraduate School. Mar 2009.

Koyak, R.A., “Risk on Roads: A Modeling Approach, Part 2: Model Refinement and Implementation”, Naval Postgraduate School. Sept 2009.

Koyak, R.A., “Risk on Roads: A Modeling Approach, Part 3: Application Development”, Naval Postgraduate School . Sept 2010.

Kursun, O., “Development of an Artificial Intelligence System for Detection and Visualization of Auto Theft Recovery Patterns”, IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety. Apr 2005.

Lewis, E., “Self-Exciting Point Process Models of Civilian Deaths in Iraq”, Security Journal, Vol. 25(3). 2011.

Mikosch, T., “Non-Life Insurance Mathematics”, Ch.11, Springer, 2009.

Marks, C.E., “Optimization-Based Routing and Scheduling of IED-Detection Assets in Contemporary Military Operations”, Sloan School of Management. May 2009.

Mino, H., “Parameter Estimation of the Intensity Process of Self Exciting Point Processes Using the EM Algorithm”, IEEE Transactions on Instrumentation and Measurement, Vol. 50, No. 3. June 2001.

Mohler, G.O., “Self-Exciting Point Process Modeling of Crime”, Journal of the American Statistical Association. March 2011.

Monarch, W.R., “Estimating Risk to Transiting Ships Due to Multiple Threat Mine Types”, Military Operations Research, Volume 4, Number 3. 2006

Pham, D.T., “Selection of K in K-Means Clustering”, Manufacturing Engineering Centre, Cardiff University. Sept 2004.

Okabe, A, “Statistical Analysis of the Distribution of Points on a Network”, Geographical Analysis, Vol. 27, No. 2. Apr 1995.

Operations Analysis Division (OAD). Quantico, VA. United States Marine Corps (USMC), Archives. 2010.

Ripley, B.D., “Spatial Statistics”, John Wiley and Sons. 1981.

Ross, S.M., “Introduction to Probability Models, Seventh Edition”, Academic Press. 2000.

Schneider, P.M., “Dynamic Incident Display and Change Point Detection and Counterinsurgency Operations”, Naval Postgraduate School. 2004.

Serefozo, R.F. “Conditional Poisson Processes”, Journal of Applied Probability, Vol. 9, No. 2. Jun 1972.

Shearer, R., “Operations Analysis in Iraq: Sifting Through the Fog of War”, Military Operations Research, Volume 16, Number 2. 2011.

Shenk, K, “Semi-Supervised Approach to Predictive Analysis Using Temporal Data”, Military Operations Research, Volume 19, Number 1. 2014.

Vere-Jones, D., “Some Models and Procedures for Space-Time Point Processes”, Springer. Mar 2008.

Washburn, A., “Katz Distributions of Minefield Clearance”, Military Operations Research, Volume 11, Number 3. 2006

Washburn, A., “Allocation of Clearance Assets in IED Warfare”, Wiley Periodicals. Aug 2009.

Woodaman, R.F., “Modeling IEDs and Compound Poisson Arrival Process on Arcs of the MSR Network”, GMU/JIEDDO Annual Technical Report. Sept 2008.

BIOGRAPHY

Arun Shankar graduated from Stephen F. Austin High School, Sugar Land, Texas, in 1999. He received his Bachelor of Arts from The University of Texas at Austin in 2002 and Masters of Science in Operations Research and Applied Mathematics from the Naval Postgraduate School in 2008. He spent a combined 28 months in both Iraq and Afghanistan as a Marine Corps captain in 2009, 2010, and 2012 and remains on active duty today.