

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b>		<b>2. REPORT TYPE</b>	<b>3. DATES COVERED (From - To)</b>		
<b>4. TITLE AND SUBTITLE</b>			<b>5a. CONTRACT NUMBER</b>		
			<b>5b. GRANT NUMBER</b>		
			<b>5c. PROGRAM ELEMENT NUMBER</b>		
<b>6. AUTHOR(S)</b>			<b>5d. PROJECT NUMBER</b>		
			<b>5e. TASK NUMBER</b>		
			<b>5f. WORK UNIT NUMBER</b>		
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>		
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>		
			<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>		
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b>					
<b>15. SUBJECT TERMS</b>					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>19b. TELEPHONE NUMBER (include area code)</b>

**MECH: ALGORITHMS AND TOOLS FOR AUTOMATED ASSESSMENT OF  
POTENTIAL ATTACK LOCATIONS**

Final Report

**Technical point of contact:**

Dr. J. C. Liu

Department of Computer Science and Engineering

Texas A&M University

H. R. Bright Building, Room 502B

3112 TAMU

College Station, TX 77843-3112

979-845-8739

[liu@cse.tamu.edu](mailto:liu@cse.tamu.edu)

Date of Submission: October 6, 2015

## Table of Contents

<b>EXECUTIVE SUMMARY</b> .....	1
<b>PROJECT OUTCOMES</b> .....	3
<b>The Strategic View</b> .....	11
<b>The Tactical View</b> .....	12
<b>MECH Overview</b> .....	12
<b>2. MODELING OF ASYMMETRIC CONFLICT EVENTS</b> .....	14
<b>Anatomy of a Conflict Event</b> .....	15
<b>The MECH Model</b> .....	16
Emplacement Modeling .....	17
Monitor/Control Modeling .....	17
<b>MECH and Tactics</b> .....	19
<b>3. FEATURE EXTRACTION OF ASYMMETRIC CONFLICT EVENTS</b> .....	21
<b>Visibility-related Features</b> .....	24
Line-of-Sight and Viewshed .....	24
Line of Sight .....	25
Viewshed .....	26
Viewshed Feature: Visibility Index.....	30
Viewshed Feature: Discrete Shape Complexity Index.....	31
Risk Aversion and Escape Adjacency.....	33
Cumulative Escape Adjacency .....	34
Viewshed Feature: Median Cumulative Escape Adjacency.....	37
Optimal Cumulative Escape Adjacency .....	38
Viewshed Feature: Route Visibility .....	39
Sparse Viewshed .....	41
Viewshed Feature: Shortest Radial .....	44
Viewshed Feature: Longest Radial.....	45
Viewshed Feature: Local Openness .....	46
<b>Geomorphometric Features</b> .....	47
Feature: Slope.....	48
Feature: Texture .....	49
Feature: Local Convexity .....	50
Feature: Elevation Range .....	52
Feature: Roughness .....	53
<b>Social/Cultural Features</b> .....	54
Social/Cultural Feature: Proximity to Populated Areas .....	54
<b>Comparison of Features across Classes</b> .....	55

<b>Summary of Features</b> .....	<b>56</b>
<b>4. PREDICTIVE ANALYSIS OF ASYMMETRIC CONFLICT EVENTS</b> .....	<b>59</b>
<b>The MECH Classification Algorithm and Evaluation Criteria</b> .....	<b>59</b>
Data Source, Pre-processing and Quality assessment .....	61
<b>Classification Algorithm Introduction and Baseline Results</b> .....	<b>65</b>
Set selection.....	65
Data Normalization .....	66
Feature Reduction.....	67
Classifier Training Algorithm .....	67
Baseline Results .....	68
<b>Parameter Tuning</b> .....	<b>72</b>
Feature reduction: <i>penier</i> and the cumulative variance.....	72
Estimation of k for the kNN Classifier.....	75
Box Constraint Estimation for an SVM Classifier with a Linear Kernel.....	77
Parameter Estimation for an SVM Classifier with an RBF kernel.....	79
Dynamic Geographical Constraints.....	80
Summary of Parameter and Hyperparameter Estimation.....	83
<b>The Impact of Unbalanced Classes</b> .....	<b>83</b>
<b>Contribution of Each Feature</b> .....	<b>85</b>
<b>Human and Machine Expert Feature Selection</b> .....	<b>87</b>
<b>Ensemble of Classifiers</b> .....	<b>90</b>
Ensembles Constructed of Single Algorithm Base Classifiers .....	92
Ensembles Constructed of Varied Algorithm Base Classifiers.....	92
Stacking.....	95
<b>Decision Tree (DT) Learning Algorithm</b> .....	<b>97</b>
<b>5. MECH BASED SITUATIONAL SIMULATION</b> .....	<b>103</b>
<b>6. CONCLUSION</b> .....	<b>108</b>
REFERENCES.....	110
APPENDIX A. DATA SET USED FOR THIS PROJECT .....	113
<b>A.1 Global Digital Elevation Model</b> .....	<b>113</b>
<b>A.2 Asymmetric Conflict Events</b> .....	<b>114</b>
Collocated Events.....	118
Dataset Problems.....	119
<b>A.3 Afghanistan Roads</b> .....	<b>120</b>
<b>A.4 Population</b> .....	<b>121</b>
APPENDIX B. ADDITIONAL FEATURES .....	123

APPENDIX C. SUMMARY OF FEATURES .....	145
<b>VOCABULARY GLOSSARY</b> .....	152
<b>FEATURE GLOSSARY</b> .....	160

## EXECUTIVE SUMMARY

This report presents major findings and lessons learned from the modeling and simulation of behaviors associated with insurgent attacks, and their relationship with geographic locations and temporal windows. The research methodology is based on quantification of insurgent actor risk aversion in the site planning of hostile actions. The *Monitor, Emplacement, and Control in a Halo* (MECH) model provides a constrained environment and decision space for describing these hostile actions. The planning and emplacement of improvised explosive devices (IED) and direct fire (DF) attacks are transformed into the balance between acceptable risk and desired security. (Section 2)

The first step in development of the MECH model aimed to assess usefulness of the basic statistics of common geomorphometric measures. This test sought significant indicators that could be readily used to identify geographic features used in site selection. While certain measures of historical attack locations did show noticeable statistical differences, but they were not significant enough to support the design of analytics algorithms with acceptable error rates. Next, tactical operations were abstracted in terms of *interobservability*, *distances*, and *logistics/shelter distance* which were added with the geomorphometric measures to form the feature set. (Section 3)

Both anecdotal and empirical evidence suggested the existence of hidden patterns and common actor roles across classes of attacks. This insight led to the formulation of supervised Machine Learning (ML) algorithms for classification of attack locations. Initial performance of individual learning algorithms reached the range of error rate of 20%-30% in select cases, with common issues such as conditioning (normalization) of

features, sizing of analysis windows, etc. taken into account. Further expansion of the training architecture to an ensemble of multiple algorithms pushed performance to single-digit error rates. The system produced consistent performance outcomes from numerous experiments based on different data sets and time spans, albeit from a constrained and noisy unclassified dataset spanning only 19 months. A series of analyses were constructed to identify leading contributors of the 77 features further confirmed risk-averse behavioral features as the most relevant features in site selection, a conclusion consistent with that of Section 3. The list of key features picked by the algorithms was found to be highly consistent with that hand-picked by three military personnel with extensive deployment experience. (Section 4)

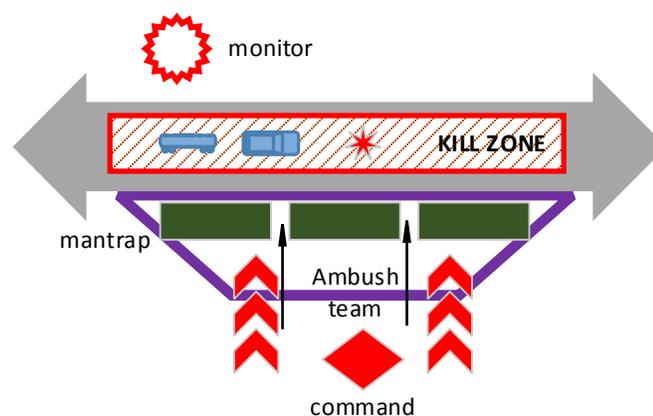
The statistical pattern analysis method in Section 4 is limited to analysis of emplacement locations of attacks. The MECH model is used to simulate site selection for the monitoring and control locations around a potential emplacement area. A general model was developed to characterize different levels of acceptable risk vs. security. (Section 5)

To explore the practicality of the MECH-based modeling methodology, a software prototype was developed for a user to use an Android App to access MECH analytics algorithms that run on a server. The prototype demonstrates the effectiveness of fusion of statistical pattern analysis, simulation, and human interpretation of military doctrines within the context of the two modeling approaches. It shows the feasibility of self-guided situational analysis informed by MECH-based situational awareness analytics.

## PROJECT OUTCOMES

For most end users, MECH analytics produces three types of map overlays: (a) past Emplacement locations; (b) potential Emplacement locations along a route whose features are statistically similar to that of past IED and DF events; and (c) locations near the route that are predicted by the MECH model to have high utility for Emplacement, Monitor and Control functions. When they are displayed together with the past event sites, the resulting graphic offers a composite view of past events and possible future actions. Whenever possible, they should be used together to gain a full understanding of the environmental situations based on three different methods.

MECH demonstrates the effectiveness of combining human intuitions, geographical structures, and behavior dynamics into computing abstractions to predict the likelihood of locations being used for attacks. The following three use cases discuss how to use MECH to perform “what if” style of tactical analysis. We adopt an idealized linear ambush model shown in Figure 1 for a human expert to interpret and annotate MECH-produced overlays for select scenarios at different scales.



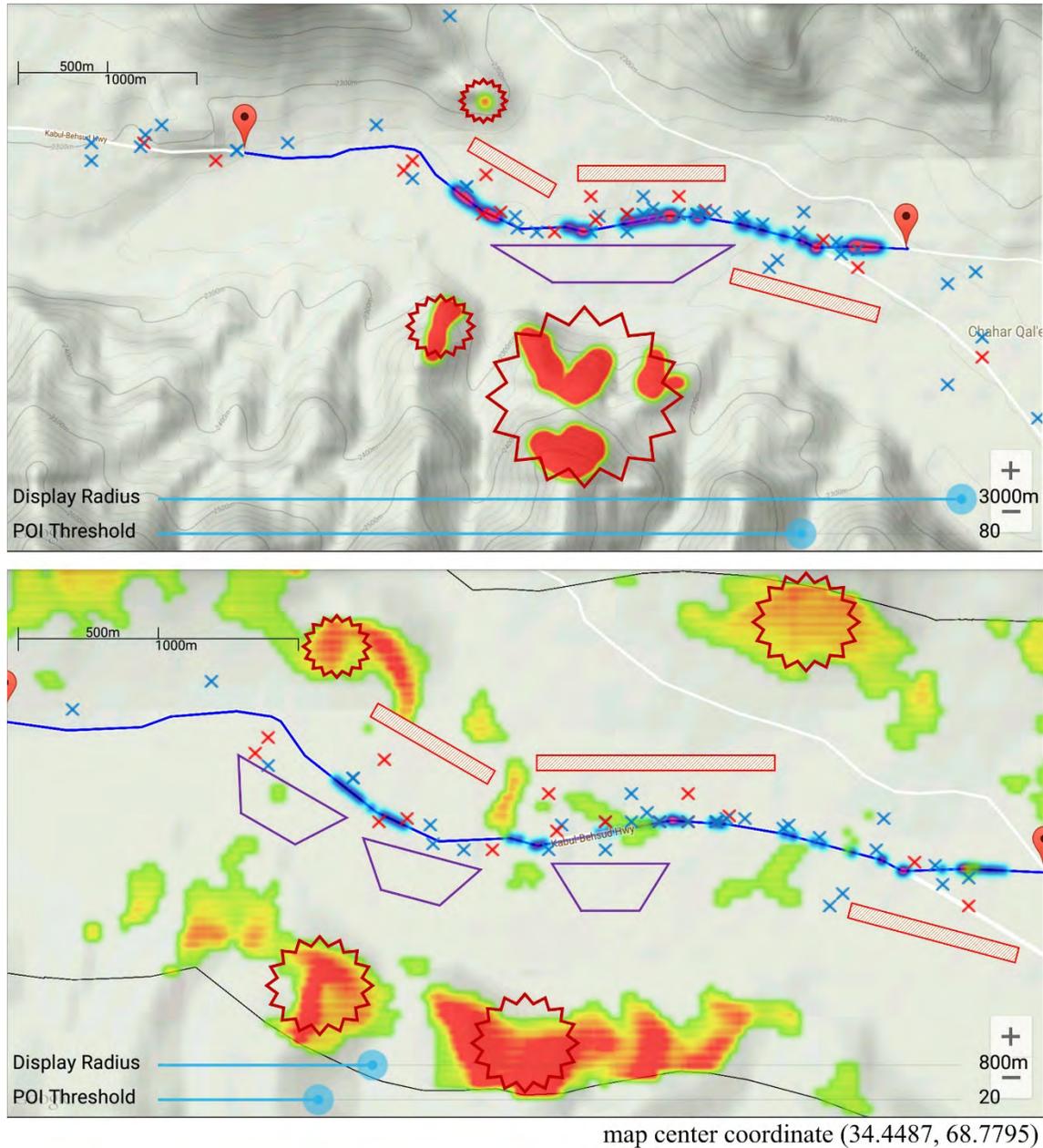
**Figure 1: An idealized ambushed model of the U.S. Army Field Manual 7-85. The kill zone is marked with a red box. A mantrap is designated with a purple trapezoid; and a monitoring/overwatch site for the kill zone is designated with a red starburst.**

### **Use Case I: Proximity and Threshold Control at Different Scales**

In this use case, a heavily attacked road segment shown in Figure 2 is displayed at two different scales. The objective of the analysis is to assess 1) if the road segment is naturally a hot zone, and 2) what might be the good watching spots for the scout of the attackers, and the hiding places for the ambush/control team. Using the MECH App, the road segment and the area around it are processed at different display ranges (display radius) and user-chosen thresholds (POI threshold). The resulting view defines the area under analysis and constrains the floor values of locations' tactical features.

In these two views, the user first selected a road segment by simply touching its beginning and end points on a MECH App running on an Android device. The selected segment is marked as a blue line with two pins marking its terminal points. The user chose to display locations of past events, which are displayed as blue crosses (IED events) and red crosses (DF events). The user asked the server to predict high risk potential emplacement locations, which were displayed as heatmaps (blue boundary, shaded from light blue to purple) along the road. Then, the user asked the server to perform observability analysis around the chosen road segment, and locations with highest observability toward the road segment displayed also as heatmaps (green boundary, shaded from green to red). The map on the App was downloaded to a desktop

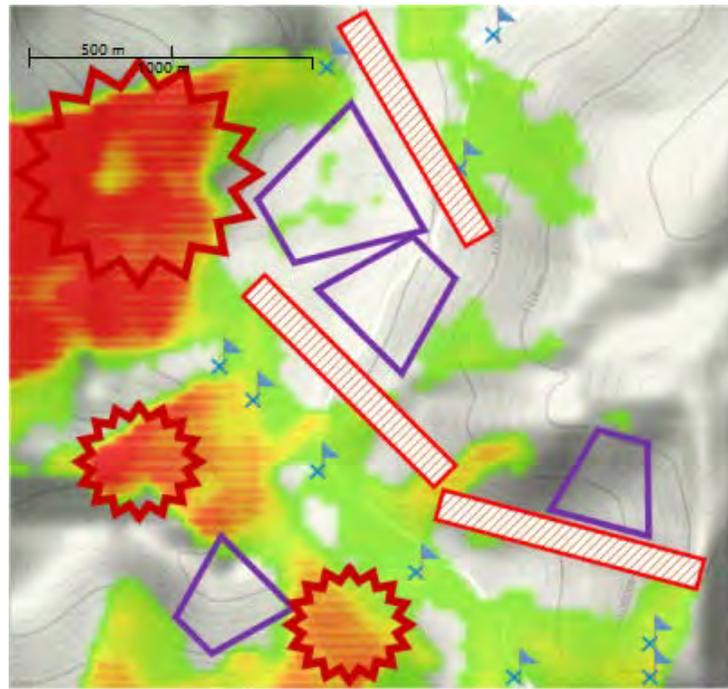
computer. Then, idealized ambush model elements from Figure 1 were manually marked to call out possible tactical plans by hostile actors.



**Figure 2: An example of route-level assessment in the valley along Kabul-Behsud Hwy: the red crosses mark the IED events, and the blue crosses mark the DF events.**

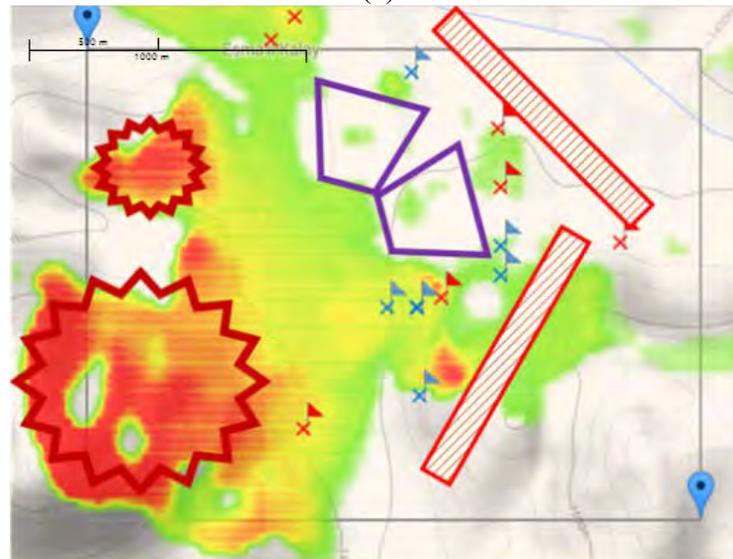
## **Use Case II: Rugged Terrain**

MECH based situational analysis can be constrained to a road segment, as shown in Use Case I, or to a rectangle shape marked by the user. The latter option is better suited for analysis of rugged terrains, which may or may not have obvious pathways. After walking through similar operational steps as in Use Case I, Figure 3 (a) shows a location in the Nuristan Forest National Reserve between Jalalabad and Asadabad. The attacks appear to follow a road/trail with good visibility to its surrounding area that offers overwatch positions on one side of the road/trail, and a favorable, concealed location for an attack team on the other side. The second example shown in Figure 3 (b) represents a similar situation, in which two vantage overwatch positions can observe movement along an area with good observability, and the attack team can station at areas with virtually no visibility. Notably for all cases, the attacks mostly occurred near the edge of areas with and without visibility. This is a strong evidence supporting the argument that actors prefer to stay near the edge of observability to execute “hit and hide” or “hit and run” tactics under the threat of return fire.



map center coordinate (34.7878, 69.8607)

(a)



map center coordinate (32.1026, 66.2843)

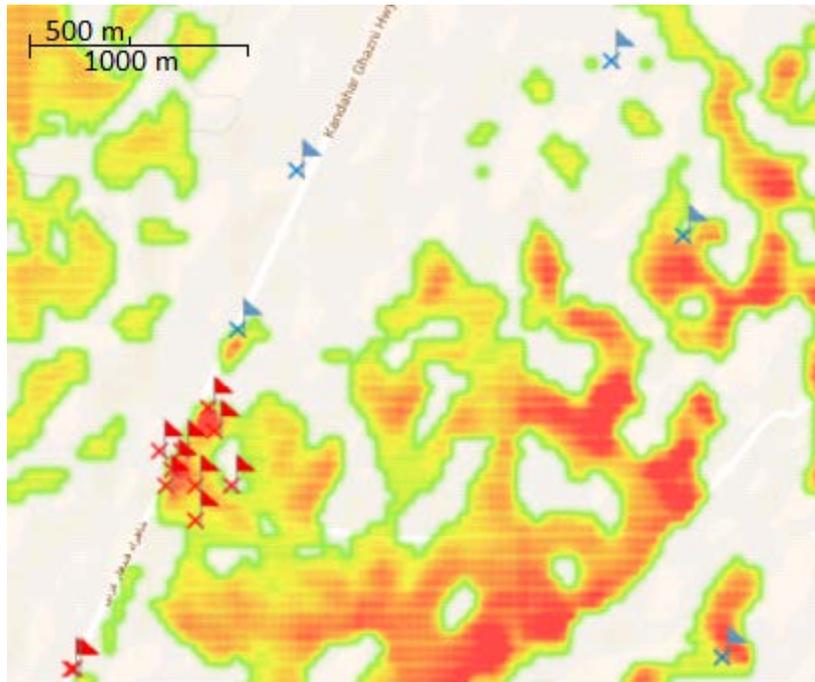
(b)

**Figure 3: (a) Mountain area in Nuristan Forest National Reserve between Jalalabad and Asadabad; and (b) Mountain area near Esma'il Kalay**

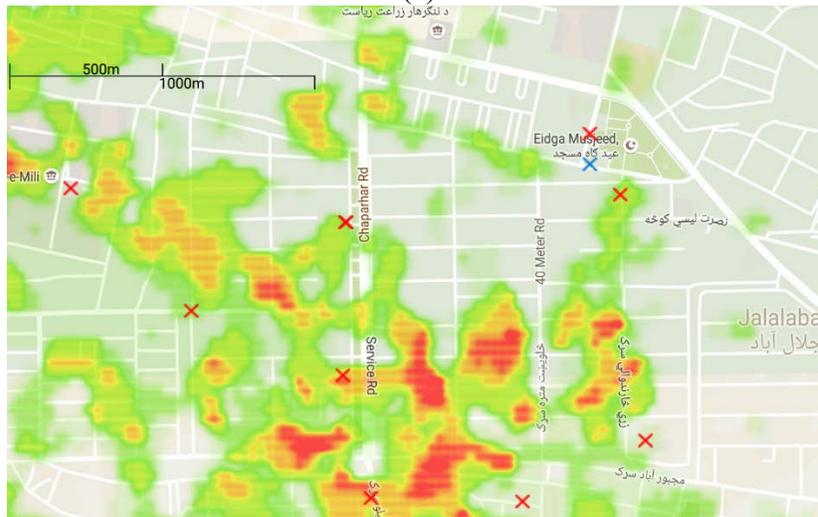
### **Use Case III: Viewshed vs. attack density**

Figure 4 shows two examples which represent distinct distributions of attack locations. Figure 4 (a) presents an attack hot spot on the Kandahar Ghazni Highway located on/around the boundary of a region of limited visibility. On the other hand, Figure 4 (b) shows the northwest downtown area near Jalalabad Airport where visibility is mostly unconstrained, attack locations appear to be fairly dispersed. The heat maps were produced through iterative steps between the road locations and their surrounding area.

Overall, it is interesting to observe that many DF locations, especially those in rugged terrain, are located at boundaries of large viewsheds. An anecdotal interpretation of this situation is that the aggressor can watch target movements from the safety of concealed locations. Then, from these covered locations, they are able to launch an attack at will, perhaps when the target is close enough for accurate aiming. On the other hand, IED attacks tend to be placed close to the viewshed center. This suggests that the attacker may choose terrain that allows better estimation of target speed or movement. This also places the attacker at a greater distance from the target when triggering the IED device. Technical insights on the design of the MECH models and algorithms are discussed in the rest of this report.



(a)



(b)

**Figure 4: The overwatch vantage locations for (a) attack hot spot on the Kandahar Ghazni Hwy; (b) dispersed attack locations in urban area northwest of Jalalabad Airport. Blue cross marks are DF attacks. Red cross marks are IED attacks.**

## 1. INTRODUCTION

Situational awareness across a battlespace is difficult to achieve. Patrols deploy frequently and react to battlefield events as situations dictate. Convoys move men and material across hazardous routes while avoiding conflict. Long distances, combined with low-bandwidth and occasionally disrupted communications complicate the issue. For the strategist, focused on resource allocation and threat detection, the level of details available on the modern battlefield is staggering but difficult to integrate. For the tactician, this wealth of information is largely useless as soon as his patrol is outside the fence. In fact, the problems with situational awareness are asymmetric across a deployed force. Awareness of recent events at the level of a brigade may be fairly comprehensive, with accurate knowledge of deployed elements, recent assessments of their status, and current disposition of known threats in the area. However, real-time awareness of battlefield conditions is usually unavailable and real-time support to mobile elements is difficult. At the same time, for the patrol on the move, the picture has simultaneously more or less details. The patrol has a rich level of details about their immediate surroundings but little awareness of what is over the next ridgeline or beyond an upcoming curve. Providing a patrol with the appropriate level of details is difficult for a number of reasons. Security concerns may limit the total amount of sensitive data deployed, while space and power impact the availability and uptime of systems collocated with the patrol. Recent advances in aerial surveillance capabilities try to meet this need but still require the full-time attention of trained operators. Training can be a

problem if the systems are too complex but overly simplistic systems may offer little information of practical use.

This research effort provides a set of situational awareness algorithms and tools, named MECH, designed to deliver appropriate intelligence derived from a common set of data to both strategic and tactical users. For the strategic user, MECH provides output and analysis that support resource allocation decisions and enhances collection management. For the tactical user, MECH provides tools that enhance situational awareness of the immediate area and adjacent terrain. This includes geographic and social analysis integrated with predictive modeling based on past events and informed by friendly and adversary tactics.

## **The Strategic View**

Strategic users face the challenge of maintaining situational awareness in a highly dynamic environment that includes patrols and convoys, friendly forces from various nations and NGOs, an active and aggressive enemy, and a relatively mobile native population. Responsible for oversight of a large geographic region, the strategic user has to coordinate the collection efforts of a sophisticated but finite suite of sensors and the analytical capacity of a very limited set of humans. Priorities and focus may have to shift quickly in response to emerging situations. Given these needs, two important requirements emerge:

1. The strategic user must be able to identify the geographic areas of most import to ongoing and planned missions. This allows efficient sensor tasking and control.

2. Within a set of collected data, the strategic user must be able to prioritize analytical efforts to maximize the utility of scarce human analysts.

## **The Tactical View**

Tactical users face the challenge of maintaining situational awareness as they move across a hostile region for extended periods of time. Although they may depart their bases with current and accurate intelligence, the modern battlefield is a fluid environment and may change quickly. The volume of communications necessary to remain fully synchronized with brigade-level echelons is impractical and potentially hazardous to a tactical patrol. Additionally, the tactical environment itself may vary dramatically as patrols transition from mounted to dismounted operations and move between rural and urban environments. Training levels and experience vary. Given these constraints, two capabilities become important:

1. The tactical user must have the ability to refine analysis in real-time.
2. Tools must accommodate users with varying levels of expertise using flexible input controls and providing intuitive visual outputs.

## **MECH Overview**

The components of MECH include algorithms and tools. The algorithms include spatial analysis, temporal analysis, predictive modeling, and route planning. These algorithms are designed to be tuned for the platform where they are deployed. For spatial analysis, we use the MECH model to identify and evaluate the usefulness of geographic

locations for conflict-related activities. The MECH model describes potential Monitor (M), Emplacement (E), and Control (C) locations within a Halo-shaped (H) space. By fitting attack strategies into a mathematical optimization framework, MECH provides automated reasoning capabilities about the utility of locations for insurgent attacks. Designed to transform various enemy-relevant factors into a unified representation, MECH identifies high threat locations and associated observation points related to insurgent objectives. MECH supports pre-trip planning, en route guidance, and post-trip model adjustment. It can accommodate a range of insurgent behaviors including intelligent and risk-averse, suicidal, random, and opportunistic behaviors through simple change of parameters.

Core MECH model components include algorithms for grading M-E-C locations based on attack/protection attributes like Line of Sight (LOS) to the potential attack site or 'X', LOS to target approach corridors, insurgent mobility and escape routes, and cover and concealment. For temporal analysis, we identify and correlate patterns associated with known historical enemy activities. These patterns may be fixed, like proximity to specific significant dates, or relative to some other activities, like the poppy harvest in Afghanistan. The patterns may also be relative to a particular triggering event. Additional MECH components include capturing of regional modifiers like population distribution, and incorporation of human expert input.

## 2. MODELING OF ASYMMETRIC CONFLICT EVENTS

Central to this research effort is the development of a system model that captures human decisions and the interactions between attackers and terrain in the siting and execution of a conflict event. The model describes the combination of terrain and tactics that make a conflict event possible, including characterization of useful terrain and the use of terrain by attackers.

In military terms, tactics cover all aspects of the employment of units in combat. This includes the movement and arrangement of the personnel and resources involved with respect to terrain and opposing forces [1]. In this research, tactic is similarly defined for asymmetric conflict events. A successful tactic maximizes the probability of success by making employment decisions optimized for the local terrain, the capabilities of the attackers, the vulnerabilities of the target, and the goals of the conflict event. In other words, the siting and execution of an event at a specific geographic location is constrained by tactics suited to the local terrain, appropriate for the opposing force, and within the abilities of the attacker.

Each conflict event is unique. However, all conflict events are planned to some extent and executed by humans. Many of these humans share common training and similar experiences. These humans are likely to make similar decisions when faced with similar choices. The conflict environment also imposes constraints. At a given time and place in a particular conflict, attacker access to conflict tools and weaponry is likely to be similar. Terrain constrains tactic choice similarly across the conflict area. Target capabilities and vulnerabilities will also tend to constrain tactics, especially as

countermeasures emerge for classes of conflict events. These shared elements enable predictive analysis of conflict events.

## **Anatomy of a Conflict Event**

The planning and execution of a conflict event involves a series of choices made over a period of time. These decisions are primarily concerned with selecting a location that supports execution of some particular tactic and addresses three conflict event elements. The Emplacement site is the place where the event occurs. For an IED, this is the location where the device is concealed. For a direct fire event, the Emplacement site is the center of the targeted force. The Monitor location is used for overwatch and early warning and will typically have good visibility of terrain along the approaches to the Emplacement site. The Control location is used to initiate execution of the conflict event and will typically have good visibility of the Emplacement site and adjacent terrain.

For this research effort, we believe that the planning and execution of a conflict event is accomplished in series of steps. First, the conflict event planner selects a particular class of event to execute, like Improvised Explosive Device (IED) or direct fire (DF), and a general geographic area that is likely to be well-suited for the event being planned. Factors involved in the selection of the area probably include the availability of targets, known or suspected availability of useful conflict event sites, and proximity to necessary support structures like population centers and communications networks.

At this point in the conflict event planning effort, a class of event has been selected. However, availability of specific supporting features in the general area may

constrain the choice of tactics. Next, a planner travels to the general area and selects a specific site. Site selection starts with analysis of potential Emplacement locations. This analysis primarily addresses the utility of the terrain at the site for the type of event chosen and the general embedding of the site in the local terrain. Useful locations are further analyzed for availability of Control and Monitor locations. This utility of these locations is primarily a function of their intervisibility with terrain at and adjacent to the Emplacement site. Adequate Emplacement sites with adequate Control and Monitor locations are identified. One of these sites is selected and the Emplacement occurs. Notably, it seems unlikely that the choice of Emplacement site is globally optimal. Planners select sites that meet all required criteria but do not exhaustively analyze every possible combination of Emplacement sites and Monitor and Control locations in order to make an optimal choice.

Once a conflict event location has been selected, the Emplacement site is prepared and human actors are placed where needed. Actors at Monitor sites provide overwatch and early warning. The conflict event is initiated by the Control actor when a suitable target reaches the Emplacement site.

### **The MECH Model**

The MECH model is composed of conflict event features that represent Emplacement and Monitor/Control locations. These features capture the outcome of complex decisions made in the planning and execution of tactics. The features are collected into tactical patterns at various resolutions: at the Emplacement site, immediately adjacent to the Emplacement site; and within the Halo, the annular area

centered on the Emplacement site that can be used to perform Monitor and Control functions.

### Emplacement Modeling

As previously described, the selection of a site and the execution of a conflict event is often the result of a carefully planned process. Emplacement at a site includes all activities required to select and prepare the location for the event. It also includes the relationship of the site to nearby and surrounding terrain. Let  $R = \{r_1, r_2, \dots, r_n\}$  be a set of past conflict event locations. Let  $X^E(r_x) = \{x_1, x_2, \dots, x_m\}$  be a set of  $m$  geomorphometric and other measurable features at the location  $r_x \in R$ . Then the tactical pattern  $\tau_E$  of the Emplacement site  $r_x$  is defined as the vector

$$\tau_E(r_x) = [c_j f_j(X_j^E(r_x))] \quad \forall j \text{ in } X^E(r_x) \quad (1)$$

, where  $c_j$  is a weight coefficient for feature  $j$  and  $f_j(X_j^E(r_x))$  is the score of feature  $j$  for location  $r_x$ . Feature  $j$  is at or adjacent to the Emplacement site.

### Monitor/Control Modeling

Actors play important roles in the execution of conflict events. For example, a carefully timed ambush only succeeds if the triggerman can observe his target's movements without being detected as an attacker before the attack is launched. Two roles common to many conflict events are Monitor and Control. The Monitor observes the target at a distance, provides overwatch, and reports to the Control. The Control observes the target and directs the execution of the conflict event. Note that in some

cases, a single actor may perform both Monitor and Control functions from one or more locations.

Let  $H$  be an annulus, or Halo, with a variable inner boundary that may approach zero and a variable outer boundary that may approach the maximum limit of intervisibility with  $r_x \in R$ . This maximum limit may be the absolute physical limit of aided or unaided human eyesight. More typically, this maximum limit will be related to details of the conflict event task being performed. In this research effort, we assume that an actor at a Monitor location must be able to distinguish between targets and non-targets and report target activities while observing from the outer bound of  $H$  using unaided eyesight. We estimate this distance to be approximately 1500 meters and define its maximum value as 2500 meters for the case of Afghanistan. Define  $X^H(r_x) = \{x_1, x, \dots, x_m\}$  as a set of  $m$  Monitor and Control features measured over or extracted from terrain within  $H$ . Then the tactical pattern  $\tau_H$  of the area surrounding  $r_x$  encompassed by  $H$  is defined as the vector

$$\tau_H(r_x) = [c_j f_j(X_j^H(r_x))] \quad \forall j \text{ in } X^H(r_x) \quad (2)$$

, where  $c_j$  is a weight coefficient for feature  $j$  and  $f_j(X_j^H(r_x))$  is the score of feature  $j$  for location  $r_x$ . Feature  $j$  is measured over or extracted from terrain within  $H$ .

Once the Emplacement and Monitor/Control models are complete, it is necessary to model their interaction in order to accurately characterize their relationship in the execution of specific conflict events. The tactical pattern of the conflict event  $T(r_x)$ , is the vector

$$T(r_x) = [\tau_E(r_x) \quad \tau_H(r_x)] \quad (3)$$

## MECH and Tactics

Attackers are faced with scarce resources and have a desire to optimize outcomes. Thus, an attacker with a specific goal, e.g. targeting opposing forces with an IED, will try to maximize success by making good choices. This does not mean that every choice is optimal. However, in the eyes of the attacker, the attack configuration for each specific conflict event is good enough to succeed given the resources, training and time available to the attacker.

Attackers are assumed to have some level of training and experience. They are also assumed to be familiar with the area local to the attack, although the familiarity may be cursory or limited. Components of a successful attack, particularly Control and Monitor locations, may be reused. Likewise, successful attacks may be replicated on distant but similar terrain. Replication may also be a function of training, where successful tactics and adaptations are communicated to distant groups [2], [3].

The concepts of cursory familiarity and attack replication expose MECH's underlying assumptions about attacker methodology and abilities. MECH does not assume that attackers use detailed geographic maps or perform exhaustive analysis of local terrain prior to the placement of an attack. Instead, as described by Gladwell [12], they rely on experience and instinct. Attack emplacement is done at a place that 'feels right' or 'looks right'. This feeling is the result of both conscious and subconscious processing of the geometric structure of the local terrain, sight lines to prominent or useful terrain features, proximity to necessary logistical support, similarity to past successful conflict event sites, and, most abstractly, similarity to a mental model of a

‘good’ site. A location with the right ‘feel’ is further assessed for critical attack support structures, like adequate concealment for an ambush or acceptable IED Control sites within range of the available command detonation wire. A successful attack confirms the ‘feels right’ analysis and solidifies the attacker’s intuition. Thus, the general shape and configuration of an attack, a tactical pattern captured by Equation 3, may be mapped onto new locations (roughly replicated) in order to duplicate previous successes.

Pattern drift is a side effect of replication. New locations are never exactly the same as previous locations. Attack parameters must be shifted to make the old pattern fit the new location. When these adjustments are made and a successful attack occurs, the pattern grows or shifts. The result is a change in tactics over time.

Occasionally, a pattern will lose effectiveness. This may occur due to countermeasures, like new IED detection equipment, or due to a lack of critical attack components, like a particular type of IED detonator. When this happens, an abrupt shift in tactics may be seen.

### 3. FEATURE EXTRACTION OF ASYMMETRIC CONFLICT

#### EVENTS

A principal component of MECH-based analytics is extraction of features relevant to the siting and execution of a conflict event. These features are drawn from three general classes: visibility-related features based on characteristics of the Emplacement site's viewshed; geomorphometric features based on characteristics of the land's surface; and social/cultural features related to proximity of human population centers.

Visibility-based analysis attempts to use human factors and limitations to constrain areas under consideration. For example, if an IED uses a command-detonated trigger, then it is likely that the Control site has direct line-of-sight (LOS) to the Emplacement site. Potential Control sites that cannot see the Emplacement site are probably less useful. Conversely, an ambush that relies on attacker concealment probably requires a relatively large area near the Emplacement site that is concealed from target view. A potential Emplacement site without nearby concealment is less useful in this scenario. Visibility-based analysis also attempts to summarize the impression of a location gained by a trained attacker during site assessment. Visible areas and inferred hidden areas are assessed and mentally summarized by the attacker at various scales related to the planned attack. Local viewshed and related features attempt to capture this assessment process.

Geomorphometry is the science of quantitative land-surface analysis [4]. For MECH, geomorphometric features are drawn from statistical analysis of the ASTER

Global Digital Elevation Map (DEM) (is a product of METI and NASA) [5]. Various morphologic features describing terrain structure and capturing terrain surface information are collected using elevation data with a resolution of 30 meters. The features are collected at a variety of windows sizes.

For both geomorphometric and visibility-related features, analysis at various window sizes is necessary. Window size determination is an open problem in geomorphometry [6]. Although several automated and semi-automated approaches have been advanced (variograms are frequently proposed [7]), the most common solution is to manually assign fixed window sizes large enough to contain the features and activities being analyzed [8]. For MECH, a total of six window sizes have been assigned, based on known or estimated attacker requirements. Table 1 lists and describes these window sizes.

For each of the collected features, we show a boxplot and a smoothed histogram in which three different Afghanistan data sets are compared: roads and two classes of events: IED and DF. The IED and DF datasets are composed of conflict events that occurred in Afghanistan between early 2011 and mid-2012. See Appendix A.2 for more information. The roads set is composed of discrete points sampled at 30 meter intervals from paved and improved roads across all of Afghanistan. See Appendix A.3 for more information.

**Table 1: Window sizes used for geomorphometric and visibility-based analysis**

Radius (meters)	Rationale
25	The radius of the area considered immediately adjacent to the Emplacement site. Also, estimated to be the maximum range of a typical IED blast.
100	The radius of the area surrounding the Emplacement site that is useful for a near ambush or similar direct fire event.
350	The radius of the area surrounding the Emplacement site that is useful for a far ambush or similar direct fire event. Also, a typical limit of long rifle suppressive fire.
500	The radius of an area surrounding the Emplacement site that is useful for Control functions, like command wire detonation of an IED. (Estimated anecdotally.)
1000	The radius of an area surrounding the Emplacement site estimated to be most useful for Monitor functions. Also, a typical limit of crew-served weapon suppressive fire. (Estimated anecdotally.)
2500	The maximum sight line considered in this analysis.

In the following analysis, box plots and histograms are used to concisely describe the features. For the boxplots, the whiskers extend to the most extreme data points that are not considered outliers. Outliers are not displayed. The notches on either side of the median can be used to understand similarities between samples. Two samples are probably drawn from different populations (significantly different at  $\alpha=0.05$ ) if their intervals (the width of the opening of the notch) do not overlap. In addition to conventional boxplot information, there is an additional symbol in each boxplot (diamond, square, and circle) located at the mean of the data.

The histograms were generated using either 100 bins or bins at 30-meter intervals (to coincide with resolution of some of the data). Since the number of samples varies greatly between sets, e.g. ~3.4 million road points vs. ~13,000 IED events, the resulting bin magnitudes are represented as a percentage of the total sample. Thus, the magnitude of a bin containing 41,000 road points would be  $100 * (\frac{41k}{3.4M})$ . Bins are represented by a point at [bin center, magnitude] and adjacent points are joined with a line segment to facilitate visual interpretation.

The following sections detail an illustrative subset of features used in the MECH model. For conciseness in the body of this dissertation, other features collected and used in MECH can be found in Appendix B.

## Visibility-related Features

### Line-of-Sight and Viewshed

Line-of-sight (LOS) describes the intervisibility between two points: if the points are visible to each other, they have LOS. Intervisibility is a common requirement for many conflict event activities. The Control actor intending to accurately trigger an IED needs intervisibility with the Emplacement site. The sniper needs intervisibility with the target in order to fire accurately. The addition of an LOS constraint to geomorphometric features allows the interpretation of those features in a new way. In some cases, LOS may be interrupted by nearby terrain. In these cases, activities that require LOS cannot occur beyond the interruption or break in intervisibility. Thus, terrain beyond the break can be excluded from analysis related to activities that require LOS. Multiple LOS may

be combined into a viewshed, which may be exhaustive or sparse. Features can be collected directly from both types of viewsheds. Viewsheds may also be used as a constraint or mask for the collection of other features. Finally, multiple viewsheds may be combined and features then collected from the result.

As presented below, this analysis offers a greedy assessment of viewshed. The impact of DEM error, surface irradiance, precipitation, dust, and vegetation are not assessed and would tend to degrade or change the estimate of visibility. Also, the viewsheds are estimated using elevation data at a resolution of 30 meters. The resolution is probably insufficient to capture some significant viewshed details.

### Line of Sight

Denote as  $r_x$  a location of interest and let  $P = [p_0 p_1 \dots p_m]$  be a vector of  $m$  points evenly distributed along a line extending from  $r_x$  to a distant point  $DEM_{m,n}$  such that  $p_0 = r_x$ ,  $p_m = DEM_{m,n}$  where  $DEM$  is a digital elevation map. Let  $L$  denote the vector of elevations for  $P$ . These elevations may be adjusted for observer height,  $h$ .

$$L = [l_0 l_1 \dots l_m], \quad l_0 = \text{elevation}(r_x), \quad l_j = \text{elevation}(p_j) \text{ for } j = 1, \dots, m \quad (4)$$

Then, LOS between  $r_x$  and all points in  $P$  is the vector

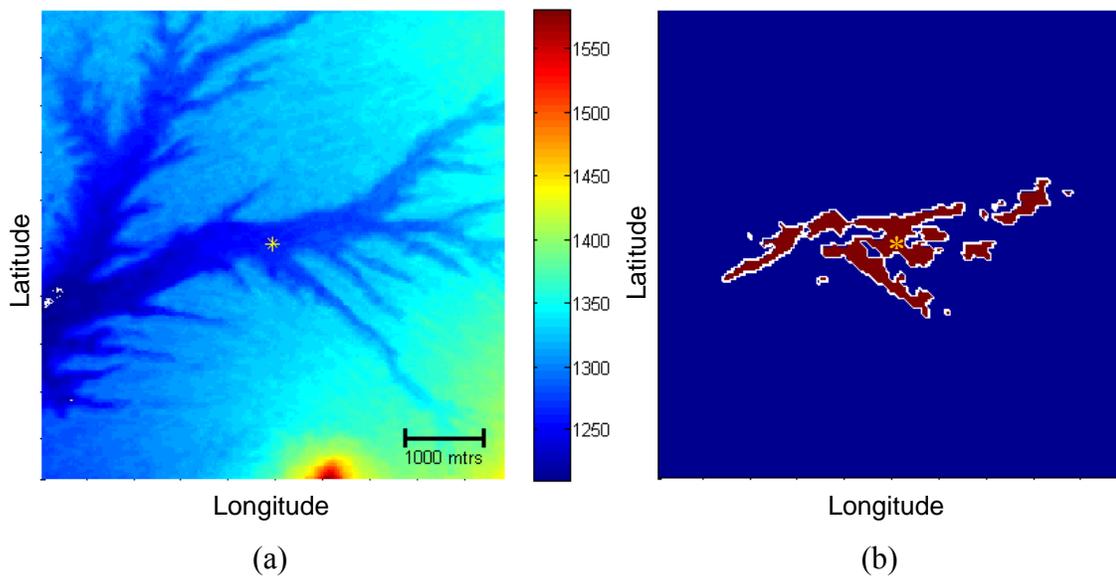
$$LOS(L) = \left[ \text{slope}(l_0 + h, l_j + h) \geq \max \left( \text{slope}(l_0 + h, l_{1, \dots, j-1}) \right) \right] \quad (5)$$

$$\text{for } j = 1, \dots, m$$

$LOS(L)$  is a vector of Boolean values that describes the intervisibility between  $r_x$  and each of the points in  $P$ . This is similar to the approach adopted by Izraelevitz in [9].

## Viewshed

A viewshed describes the portion of a geographic area visible from a single point. Viewshed is calculated by determining LOS between a site and a set of surrounding points. Viewshed is a useful way to see terrain through the eyes of an attacker. Hidden or revealed terrain and openness or exposure of a site are examples of



**Figure 5: The relationship between elevation and viewshed; (a) An elevation map and (b) its associated viewshed. A yellow asterisk denotes  $r_x$ .**

information visible in a viewshed that is difficult to see in a conventional elevation map.

Figure 5(a) offers an elevation map and its associated viewshed (Figure 5(b)). The location of interest,  $r_x$ , is marked with a yellow asterisk in the center and the viewshed is calculated with respect to this location. In the viewshed, locations with an LOS to  $r_x$  are marked in red. Locations without LOS to  $r_x$  are marked in blue. Locations

marked in white are at the edge of intervisibility. Interestingly, it is difficult to gain an understanding of the viewshed from visual inspection of the elevation map alone.

Calculation of a viewshed provides an understanding of the intervisibility of  $r_x$  with the surrounding terrain. This understanding contributes to remote assessment of the site regarding its vulnerability to certain types of attacks or its potential exposure to hostile observers. Any action directed against  $r_x$  that requires intervisibility must originate or be triggered from a location within the  $r_x$  viewshed.

There is a wide variety of viewshed algorithms. A cross section of these algorithms can be found in [9], [10], [11]. In this research, a radial sweep algorithm similar to [10] is used to determine viewsheds. Two drivers inform this choice: optimization and feature extraction. Since radials can be processed independently or in batches, a high degree of parallelization may be achieved and the parallelization can be tailored to the number of threads or processes available. This allows more optimal use of available computing resources. This optimality is important because exhaustive viewshed determination is computationally expensive. The second driver is feature extraction. Subsets of radials can be used to extract features that summarize terrain geometry at various degrees of compression to succinctly capture visibility-related characteristics.

In order to determine viewshed using a radial sampling algorithm, denote the location of interest as  $r_x$ . Define radius *rad* to be the length of the maximum possible sightline of interest. Define  $N_S$  as the number of radials that will be used to determine the viewshed. As a rule of thumb, for an exhaustive viewshed where LOS is determined

between  $r_x$  and every pixel within radius  $rad$ ,  $N_S$  should be the number of pixels on a circle of radius  $rad$ ,

$$N_S = \lceil 2\pi * rad \rceil \quad (6)$$

Note that  $rad$  constrains the area considered to be part of the viewshed. There may be points at distances greater than  $rad$  that have LOS with  $r_x$ . Also note that for some features,  $N_S$  may be set to sample radials much more sparsely. The impact of this choice will be explored for some sparse viewshed visibility-related features.

Let  $\Theta$  be a vector of size  $N_S$  consisting of angles, evenly spaced between  $2\pi/N_S$  and  $2\pi$ , where  $\Theta = [\theta_1, \theta_2, \dots, \theta_{N_S}]$ , and  $\theta_i = i * 360/N_S^\circ$ . Then,  $P_{\theta_i}$  is a vector of points distributed along a radial line extending outward from  $r_x$  at angle  $\theta_i$  to a distance of  $rad$  and the elevations of the points are denoted as  $S = [L_1, L_2, \dots, L_{N_S}]$  where  $L_i$  is the elevation vector of  $P_{\theta_i}$  and given by

$$L_i = [l_{i,0} \ l_{i,1} \ \dots \ l_{i,m}], \quad l_{i,0} = \text{elevation}(r_x), \quad l_{i,j} = \text{elevation}(p_j) \quad (7)$$

for  $j = 1, \dots, m$ ,

and  $p_j$  is the coordinates of a point in  $P_{\theta_i}$  defined by  $\left( \left[ \frac{rad}{m} * j * \cos \theta_i \right], \left[ \frac{rad}{m} * j * \sin \theta_i \right] \right)$ . Then the sampling matrix  $S$  of a circular area around  $r_x$  is translated into matrix form, where the  $i$ th row of  $S$  represents elevations for the points in  $P_{\theta_i}$  and the  $j$ th column of  $S$  represents the pixel distance from  $r_x$  given by  $\frac{rad}{m} * j$ .

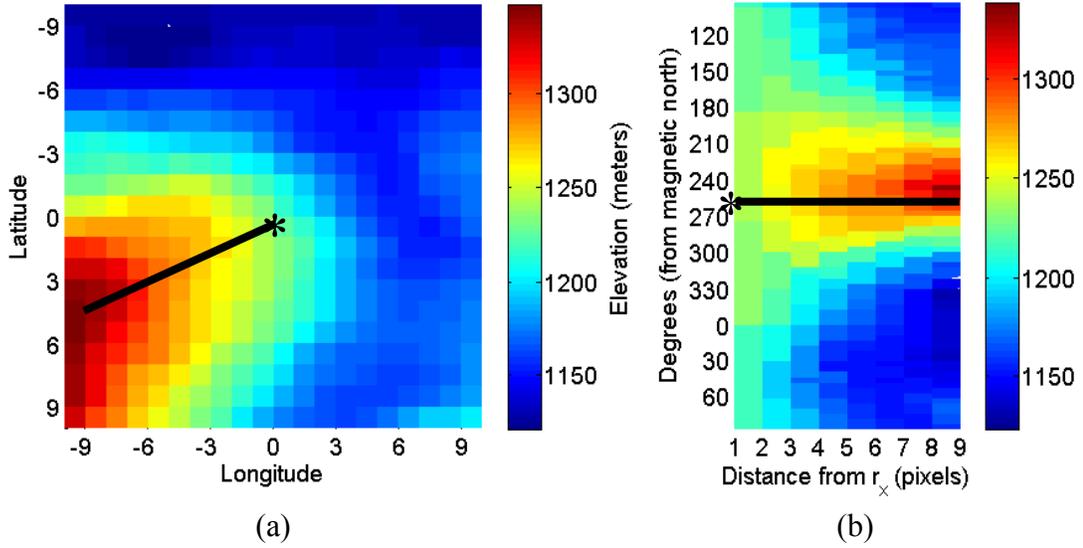
LOS along radial  $P_{\theta_i}$  is described similar to Equation 5:

$$LOS(L_i) = \left[ \text{slope}(l_{i,0} + h, l_{i,j} + h) \geq \max \left( \text{slope}(l_{i,0} + h, l_{(i,1), \dots, (i,j-1)}) \right) \right] \quad (8)$$

for  $j = 1, \dots, m$ , where a value of TRUE indicates that the point has LOS with  $r_x$ .

Viewshed is determined by finding  $LOS(L_i)$  for all desired radials

$$VS(r_x) = LOS(L_i) \quad \text{for } i = 1, \dots, N_S \quad (9)$$

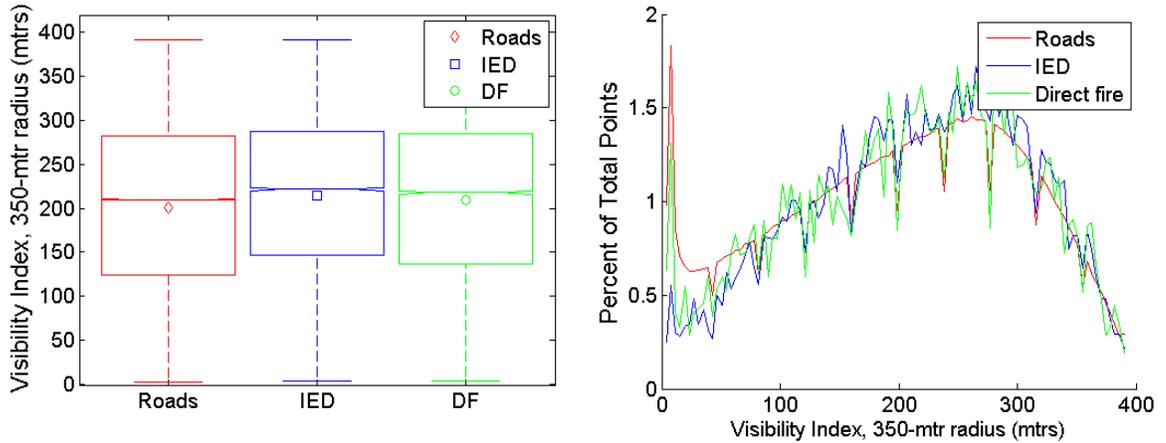


**Figure 6: Conventional and radial elevation maps; (a) A conventional elevation map; (b) the same elevation map, presented in radial format. The black line denotes the same radial in both plots.**

Figure 6(a) displays a conventional elevation map. The x-axis and y-axis represent longitude and latitude, respectively, and the colors indicate elevation. The asterisk at coordinate  $[0,0]$  represents  $r_x$ . Figure 6(b) is the same elevation map displayed in radial form. Each row contains the elevation vector  $L_i$  for a single  $P_{\theta_i}$ . Each column contains the  $N_S$  elevations of points on the circumference of a circle centered on  $r_x$ . In Figure 6(a), the black line starting at  $r_x$  and extending outward along an azimuth of approximately  $250^\circ$  magnetic represents  $P_{250}$ . The black line in Figure 6(b) is also  $P_{250}$ .

## Viewshed Feature: Visibility Index

Visibility index describes the total amount of terrain within a specified radius or window that is visible from the center of the viewshed. Visibility index provides a scalar assessment of the total intervisibility of  $r_x$  and the surrounding terrain and gives an



**Figure 7: Visibility index at a radius of 350 meters.**

indication of the degree of exposure or concealment of the site. Given a conventional exhaustive viewshed  $VS(r_x)$ , as depicted in Figure 5(b), and a radius  $w$ ,

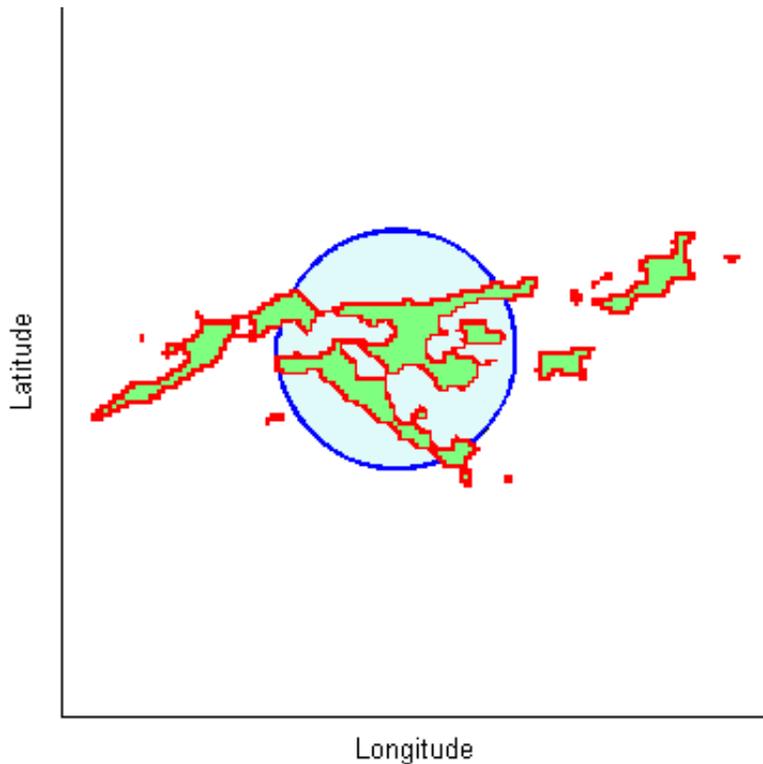
$$VisIndex_w(r_x) = \sum_{i,j} VS(r_x)_{i,j} \mid distance(r_x, VS(r_x)_{i,j}) < w \quad (10)$$

In Figure 7, visibility index is calculated from an exhaustive viewshed over a radius of 350 meters. Notably, there is a significant degree of overlap between roads and all classes of events. Visibility indices for other window sizes are in Appendix B.1.

### Viewshed Feature: Discrete Shape Complexity Index

The Discrete Shape Complexity Index ( $SCI_D$ ) describes the general complexity of a viewshed by capturing how dispersed it is.  $SCI_D$  is derived as a perimeter-to-boundary ratio. The perimeter  $P$  is the actual count of pixels with LOS that are adjacent to pixels without LOS. In terms of the exhaustive viewshed, this means that at least one of the eight adjacent pixels has no LOS with  $r_x$ . The boundary is a circumference of the smallest circle whose area equals the total count of pixels with LOS to  $r_x$ .

In Figure 8, points within the viewshed are colored green and red. The red pixels denote the perimeter of the actual viewshed. Each red pixel is adjacent to at least one pixel that does not have intervisibility with  $r_x$ . The dark blue circle encloses an area



**Figure 8: Viewshed and discrete shape complexity index.**

equal in size to the total area represented by the red and green pixels.

$SCI_D$  is found as

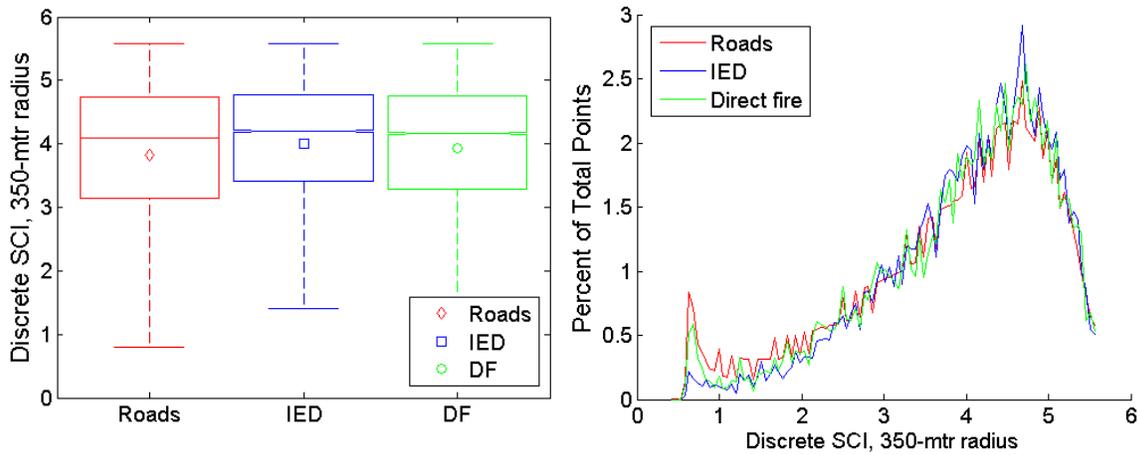
$$SCI_D(r_x) = \frac{P}{2 * \pi * r}, \quad r = \sqrt{\frac{\sum_{i,j} VS(r_x)_{i,j}}{\pi}} \quad \forall i, j \text{ in } VS(r_x) \quad (11)$$

$$P = \sum_{i,j} \left( VS(r_x)_{i,j} \equiv 1 \quad \& \quad \left( \sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} VS(r_x)_{m,n} \right) < 9 \right) \quad \forall i, j \text{ in } VS(r_x) \quad (12)$$

Note that in Equation 12, if  $VS(r_x)_{i,j}$  has a value of 1 (has LOS with  $r_x$ ) AND at least one surrounding pixel does not (no LOS with  $r_x$ ), then the result is a 1.

Figure 9 shows  $SCI_D$  calculated across a window with a radius of 350 meters.

The distribution of road points and events appears to be similar.  $SCI_D$  for other window sizes can be found in Appendix B.2.



**Figure 9: Discrete Shape Complexity Index at a radius of 350 meters.**

## Risk Aversion and Escape Adjacency

Many types of conflict events rely on concealment, camouflage, and the element of surprise. The success of IED attacks frequently depends on the ability of the actors to remain hidden until the target is correctly positioned and the attack is launched. Similarly, direct fire attacks may last longer or be more effective if the shooters can fire from a protected location. Success for both types of events requires the actors to avoid risk as much as possible before the attack. Thus, an important element of conflict event site selection is the identification of locations around the conflict event that offer cover and concealment to a risk averse attacker. In this case, the concept of escape adjacency may provide insight into the tolerance of risk by an actor. A location with escape adjacency has LOS with  $r_x$  but is adjacent to a location without LOS to  $r_x$ . These locations lie on the perimeter of the viewshed and are marked in red in Figure 4. When situated at an escape adjacent location, an actor can move quickly from visible to hidden with regard to  $r_x$ . Similarly, an actor can position himself at the edge of intervisibility in an effort to jointly maximize visibility of the target and concealment from the target. A risk averse attacker might prefer locations with escape adjacency.

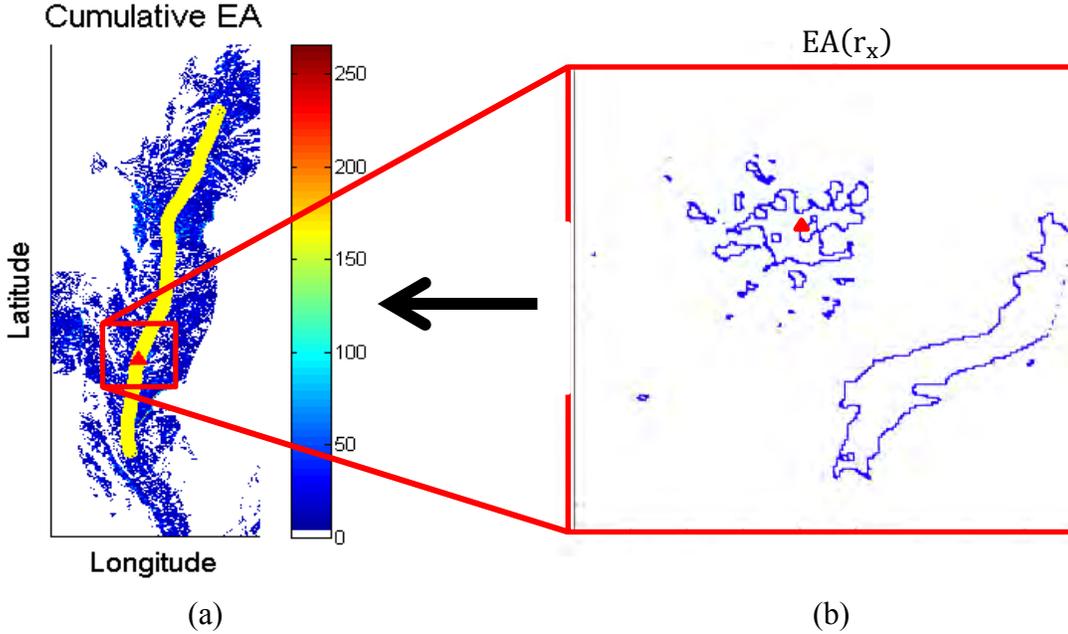
Let  $VS(r_x)$  be a viewshed, organized as an  $n \times n$  binary matrix with  $r_x$  at its center, as depicted in Figure 1.b. Then the escape adjacency matrix  $EA(r_x)$  defined over  $VS(r_x)$  is

$$EA(r_x) = \left[ \begin{array}{c} (VS(r_x)_{i,j} \equiv 1) \quad \& \quad \left( \left( \sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} VS(r_x)_{m,n} \right) < 9 \right) \\ \forall i, j \text{ in } VS(r_x) \end{array} \right] \quad (13)$$

In  $EA(r_x)$ , the resulting  $n \times n$  matrix, escape adjacent pixels are those that (1) have LOS with  $r_x$ ; and (2) have at least one neighboring that does not have LOS with  $r_x$ . Condition (1) is satisfied when  $VS(r_x)_{i,j}$  has a value of 1, indicating that the location has intervisibility with  $r_x$ . Condition (2) is satisfied when the sum of  $VS(r_x)_{i,j}$  and all adjacent pixels is less than nine, indicating that at least one of the neighbors does not have intervisibility with  $r_x$ . These are the same conditions used in Equation 12.

### Cumulative Escape Adjacency

Once escape adjacent locations have been defined for all  $r_x \in R$ , cumulative escape adjacency (CEA) can be calculated. CEA is determined by overlaying onto a single map the escape adjacent locations for all  $r_x$  along a road or route.



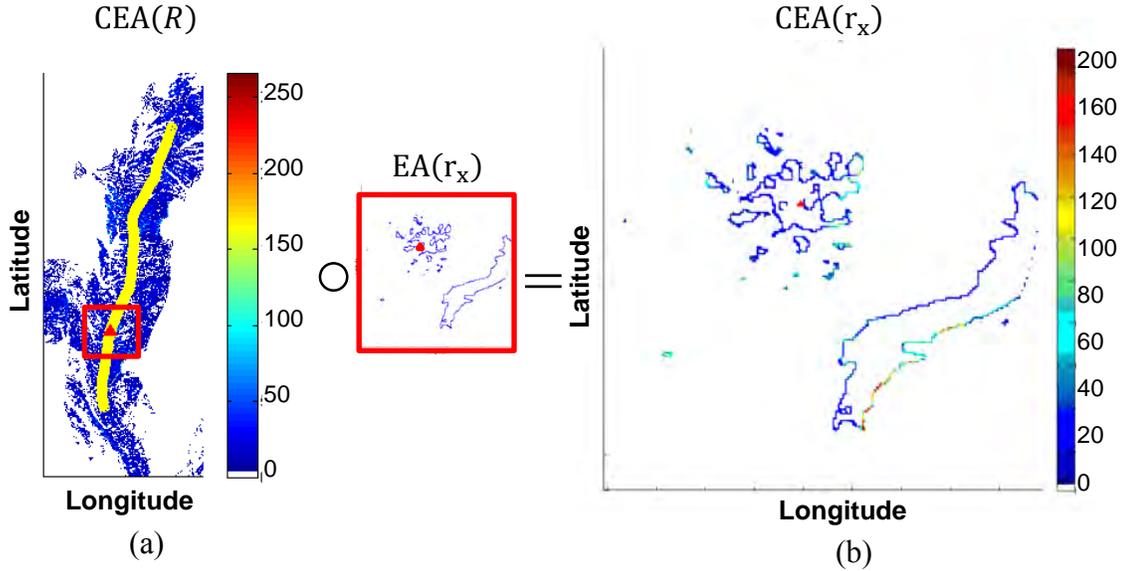
**Figure 10: Cumulative escape adjacency (CEA); (a) CEA along route  $R$ ; (b) Escape adjacency for a single  $r_x \in R$ .**

Let  $R = [r_1, r_2, \dots, r_n]$  be a route or roadway of interest divided into  $n$  discrete points at a constant interval. Let  $CEA(R)$  be a geographically localized two-dimensional zero matrix sufficiently large to encompass all terrain within some specified distance of every point in  $R$ . Then the cumulative escape adjacency map for the route  $R$ ,  $CEA(R)$ , is the summation of the individual  $EA(r_x)$  maps for each  $r_x \in R$ .

$$CEA(R) = CEA(R) + EA(r_x) \quad \forall r_x \in R \quad (14)$$

Figure 10(a) shows a cumulative escape adjacency map for an 800-point route. The callout, outlined in red in Figure 10(b), is the escape adjacency map for a single  $r_x$ .

Interpretation of  $CEA(R)$  is straightforward. The value of each pixel in the  $CEA(R)$  map is the total number of  $r_x \in R$  for which that pixel is escape adjacent. In Figure 10(a), the maximum radial length used to calculate the viewshed was 2500



**Figure 11:  $CEA(r_x)$  is the Hadamard product of  $CEA(R)$  and  $EA(r_x)$ .**

meters. So, a point (or pixel)  $p$  with a value of 80 is escape adjacent for 80 points ( $r_x \in R$ ) along the route ( $R$ ), all of which are within 2500 meters of  $p$ .

Once  $CEA(R)$  has been assembled, cumulative escape adjacency for individual points along the road,  $CEA(r_x)$ , can be determined.  $CEA(r_x)$  is found by taking the Hadamard product (entrywise product) of  $EA(r_x)$  and  $CEA(R)$ .

$$CEA(r_x) = CEA(R) \circ EA(r_x) \quad (15)$$

Figure 11 shows how  $CEA(R)$  and  $EA(r_x)$  are multiplied to get  $CEA(r_x)$ . The result is a false-color map centered on  $r_x$  showing a set of points that are escape adjacent with  $r_x$  and with other points in  $R$ , where the color of a pixel depicts the number of  $r_x \in R$  for which that pixel is escape adjacent. In the figure, the escape adjacent points colored red can see  $r_x$  and approximately 200 other points along  $R$ .

## Viewshed Feature: Median Cumulative Escape Adjacency

$CEA(r_x)$  provides a mechanism for describing the visibility and escape adjacency for a particular  $r_x$ . This allows an attacker or a target to determine the points that are likely to provide good support to a conflict event. Locations with visibility and a high CEA value are very useful for monitoring and overwatch. As previously noted, locations (pixels) with a value greater than one are escape adjacent both for  $r_x$  as well as for other points. Median cumulative escape adjacency provides some indication of how visible  $r_x$  and other points in  $R$  are to surrounding escape adjacent locations and provides an ability to roughly assess the degree of conflict event support available to  $r_x$ .

$$\widetilde{CEA}(r_x) = \text{median}(CEA(r_x)) \quad (16)$$

Figure 12 shows the distribution of  $\widetilde{CEA}(r_x)$ . It appears that both types of events and roads are all drawn from similar distributions. Maximum and minimum  $\widetilde{CEA}(r_x)$  can be found in Appendix B.3.

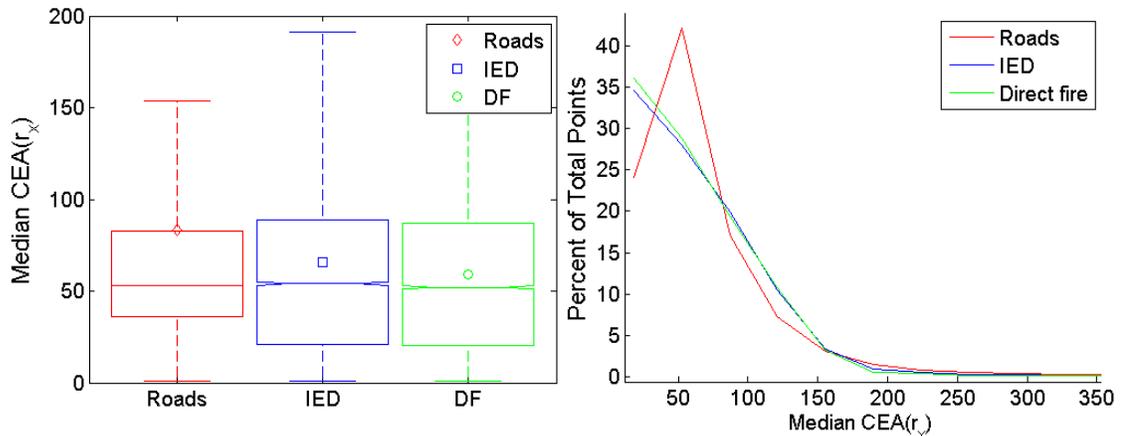


Figure 12: Median  $CEA(r_x)$ .

## Optimal Cumulative Escape Adjacency

$CEA(r_x)$  provides the set of all points that are escape adjacent for some particular  $r_x$  and may be escape adjacent for other  $r_x \in R$ . However, perusal of this exhaustive list of potential Monitor/Control locations is probably not common for risk averse actors and probably not representative of actual human behavior. Instead, a human seeking a good Monitor/Control location probably selects a general area with good potential sites and then selects the optimal location within or near that general area. This fits with our understanding of “thin-slicing” as proposed by Gladwell [12] where decisions made are strongly influenced by intuition, instinct. This intuition or instinct is the outcome of an unconscious or subconscious integration of available facts and impressions.

Thus, in the search for a good location to support a conflict event, attackers may follow, consciously or unconsciously, a three-step process: (1) identify a general area that appears to meet Monitor/Control criteria; (2) move towards and around the selected general area; and (3) choose the locally optimal site at or near the selected general area for use as a Monitor/Control location. Optimal Cumulative Escape Adjacency attempts to capture the notion that humans are often willing to make some level of effort in order to improve their position or outcome. Assuming that a ‘better’ location has greater cumulative escape adjacency, a reduced, more optimal set of  $CEA(r_x)$  locations can be selected by discarding points that have nearby neighbors with greater escape adjacency.

Let  $w$  be the maximum distance that a human actor is willing to move in order to improve a position. Then,

$$CEA'(r_x)_{i,j} = \begin{cases} CEA(r_x)_{i,j} & | \ CEA(r_x)_{i,j} \equiv \max(mat\_w_{i,j}) \\ 0 & \end{cases} \quad \forall i,j \in CEA(r_x) \quad (17)$$

, where  $mat\_w_{i,j}$  is a circular window with radius  $w$  centered on  $CEA(r_x)_{i,j}$ . Each location in the resulting reduced set of locations can be considered locally optimal within a radius of  $w$  with respect to maximum cumulative escape adjacency.

### Viewshed Feature: Route Visibility

For the tactics used in many conflict events, simple intervisibility with  $r_x$  is not enough. Visibility of the approaches to  $r_x$  is also important and the total extent of the visible area needed for a particular tactic varies with terrain and tactics. In the case where the target is mobile, an actor at a Control location may need sufficient visible extent to estimate vehicle speed accurately in order to trigger the IED or fire a weapon at a preselected site. Better roads and faster targets increase the total visibility required. Intervisibility over larger extents may also be required for the Control actor to ensure that the target is appropriate for the attack being planned. For example, an ambush using light shoulder-fired weapons should not engage a heavily armed patrol. In this situation, a Control actor might want to see all of the vehicles in a patrol before choosing to initiate the attack. Attack scale may also play an important role. A visible stretch of road may be required for a large-scale ambush. The Control actor is likely to want to have the entire target patrol within the kill zone before initiation of an attack.

Let  $w$  be the maximum extent of the approaches under consideration. Then  $R'_w(r_x)$  is the subset of  $R$  within distance  $w$  of  $r_x$ .

$$R'_w(r_x) = R \ | \ distance(r_x, r_i) < w \quad \forall i \ in \ R \quad (18)$$

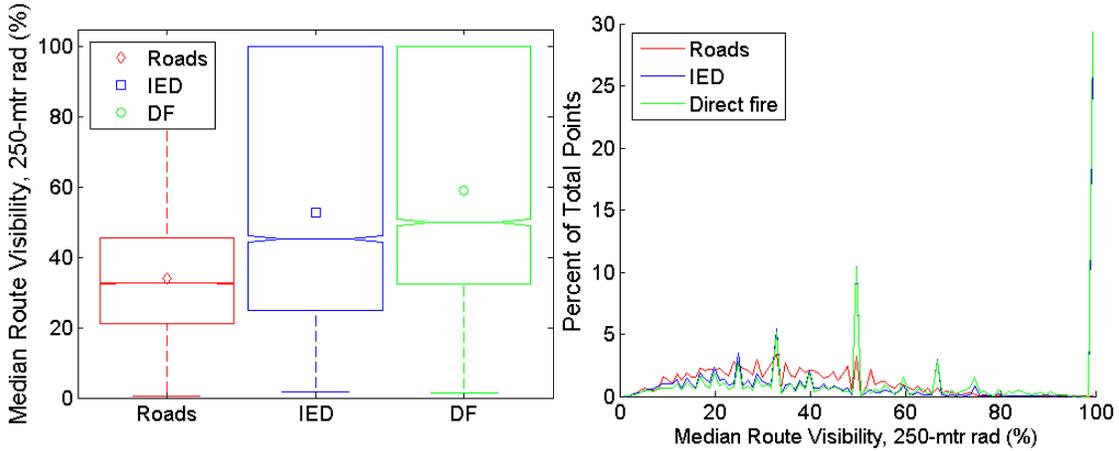
Then the route visibility of the approaches to  $r_x$  for  $CEA'(r_x)_{i,j}$  is the fraction of total points in  $R'_w(r_x)$  that have intervisibility with  $CEA'(r_x)_{i,j}$ .

$$Vis_w(CEA'(r_x)_{i,j}) = \frac{\sum_{i,j} (LOS(CEA'(r_x)_{i,j}, R'_w(r_x)_k) \equiv 1)}{|R'_w(r_x)|} \quad \forall k \text{ in } R'_w(r_x) \quad (19)$$

, and the route visibility for all points in  $CEA'(r_x)$  is the matrix

$$Vis_w(CEA'(r_x)) = [Vis_w(CEA'(r_x)_{i,j})] \quad i, j \in CEA'(r_x) \quad (20)$$

Once  $Vis_w(CEA'(r_x))$  has been calculated, then the median route visibility for the  $CEA$  points surrounding  $r_x$  is the median of all values in  $Vis_w(CEA'(r_x))$



**Figure 13: Median Route Visibility at 250 meters.**

$$Vis_w(\widetilde{CEA'(r_x)}) = \text{median}(Vis_w(CEA'(r_x))) \quad (21)$$

Figure 13 shows the importance of route visibility for approaches of 250 meters. In the figure, conflict events are significantly more likely to have better median route visibility than typical road points. In fact, the boxplot shows that a majority of IED and

DF sites have approaches that are very exposed or visible (> 50%). In the histogram, note that approximately 30% of all conflict events have greater than 95% route visibility for all road points within 250 meters.

Appendix B.4 shows median route visibility at radii of 100, 500, and 1000 meters. Appendices B.5 and B.6 show maximum and minimum route visibility at radii of 100, 250, 500, and 1000 meters.

### **Sparse Viewshed**

In some cases, analysis of the exhaustive viewshed fails to capture salient features. The noisiness of discrete data may hide general trends over larger areas. Also, the use of symmetric windows centered on  $r_x$  may conceal or wash out interesting asymmetric features. In these cases, sparse viewshed provides a mechanism for feature extraction that summarizes or constrains key viewshed features at scales appropriate to the feature being analyzed.

Sparse viewshed models terrain in a way that might be similar to the mental model constructed by a human assessing terrain. Humans tend to assess terrain by taking notice of major features, like hilltops, ridgelines, and running water. A mental model is constructed that locates these features in relation to each other. When a specific task needs to be accomplished, a human might also notice smaller or more specific features. For example, a hiker will notice the slope and roughness of possible routes. Sparse viewshed provides a mechanism to simplistically model the limits of visibility. These limits can be used to estimate viewshed and to build viewshed-constrained versions of several common geomorphometric features.

To build a sparse viewshed, calculate viewshed as described in Equations 7-9. However, modify Equation 8 to include a stopping criteria,  $tol$ , and select an  $N_s$  appropriate for the feature being collected. Define  $tol$  to be the maximum number of consecutive points along a radial that may have LOS=0. The end of the radial is set to be the last visible point before  $tol$  is exceeded. Once  $tol$  is exceeded, all more distant points are set to zero.

$$LOS(L_i) = \left[ slope(l_{i,0}, l_{i,j}) \geq \max \left( slope(l_{i,0}, l_{(i,1)}, \dots, l_{(i,j-1)}) \right) \right], \text{ for } j = 1, \dots, m \quad (22)$$

$$\text{while} \left( \sum_{k=j-tol}^j LOS(L_{i,k}) \right) > 0$$

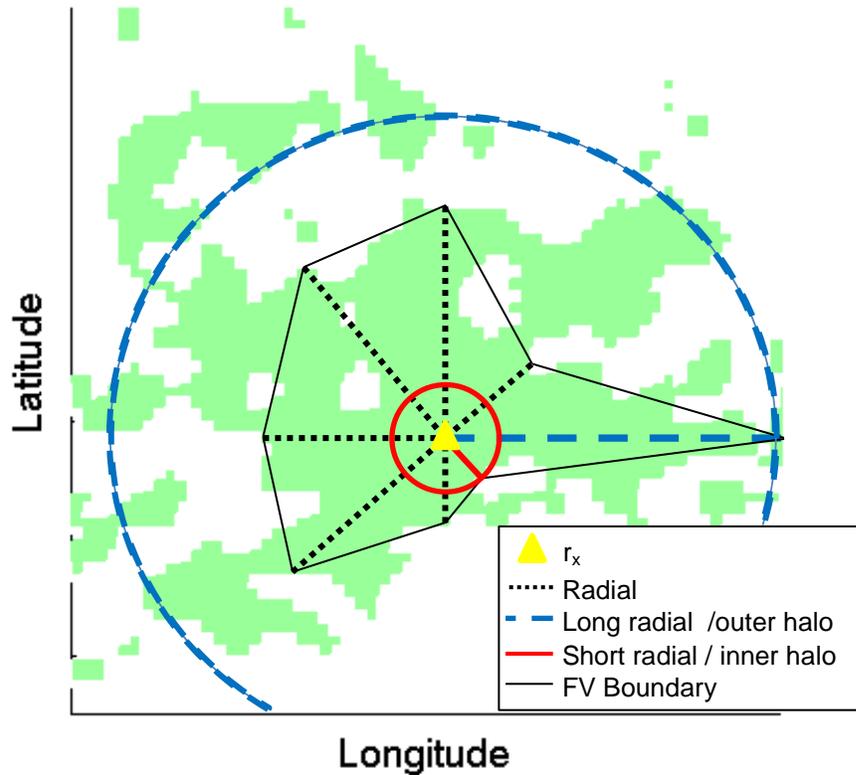


Figure 14: Example of a sparse viewshed using eight radials.

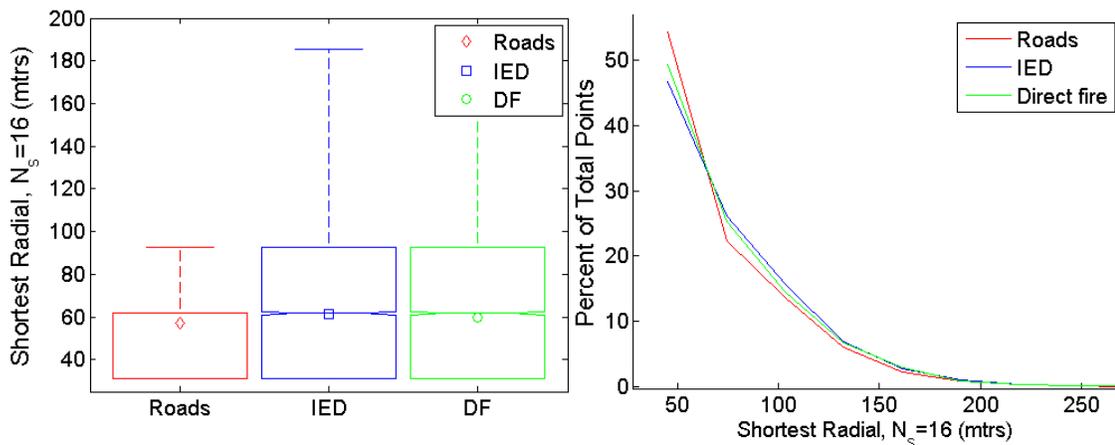
The inclusion of the stopping criteria provides a mechanism to determine the length of a mostly uninterrupted sightline along a particular azimuth. In other words, all or almost all points between  $r_x$  and the end of a radial are visible. The first significant gap in intervisibility occurs beyond the end of the radial. The *tol* variable is used to specify the width of a gap considered significant. Once the stopping criterion has been incorporated, sparse viewshed is calculated as described in Equation 9. The choice of  $N_s$  is a tunable parameter that varies with the degree of summarization desired.

Figure 14 shows a sparse viewshed overlaying an exhaustive viewshed. The green pixels in the figure are an exhaustive viewshed calculated using a radial sweep algorithm with large  $N_s$  as described in Equations 7-9. A sparse viewshed is formed using  $N_s = 8$  evenly spaced radials with the ends of the radials being determined as described in Equation 22. The *tol* criterion was set to a value of 2, so the radials stop when a gap of three or more pixels is encountered.

The thin black line joining the ends of the radials denotes the sparse viewshed boundary. In the figure, the longest radial is marked with a heavy-dash blue line. A heavy-dash line marks a circle whose radius is the longest radial. A solid red line marks a circle whose radius is the shortest radial. These two circles form the outer and inner boundary of a Halo or annulus. Key elements of conflict events planned by risk-averse attackers will occur within this Halo. Elements closer to  $r_x$  than the red circle are likely to be exposed to view by the target. Elements further away from  $r_x$  than the blue circle are likely to have poor or no visibility of  $r_x$ . Thus, analysis of terrain within this Halo may provide key insights in attacker tactics and potential use of terrain.

## Viewshed Feature: Shortest Radial

The length of the shortest radial denotes the nearest location along a selected radial where there is a significant gap in intervisibility. It is also an estimate of the length of the shortest sightline. For some types of events, the shortest radial may describe the closest place to  $r_x$  where attackers can conceal themselves. An interruption in intervisibility captured by the shortest radial may also be an indication that nearby terrain changes abruptly.



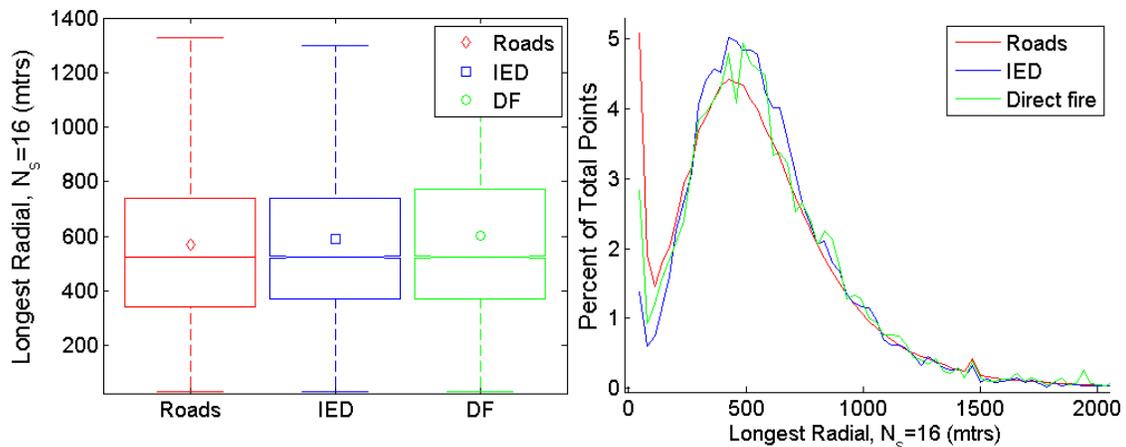
**Figure 15: Distribution of shortest radial, for sparse viewshed with  $N_s = 16$ .**

Figure 15 shows the distribution of the length of the shortest radial when  $N_s = 16$ . Although all conflict events seem to share a common distribution, they are clearly distinct from a majority of road points. The shortest radial for various sizes of  $N_s$  (4, 8, 32, and 64 radials) can be found in Appendices B.7.

## Viewshed Feature: Longest Radial

The length of the longest radial is the length of the longest uninterrupted or mostly uninterrupted sightline. For some types of conflict events, the longest radial may describe the direction in which Monitor actors may possibly be found. A long radial may often highlight linear features that lie along a radial, like a river valley or ridge, or long gentle slopes, where intervisibility is not interrupted. Also, long radials tend to indicate that the terrain along that radial tends to be relatively smooth.

Figure 16 shows the distribution of the longest radial. It appears that conflict events and roads share a common distribution. The longest radial for various sizes of  $N_s$  (4, 8, 32, and 64 radials) can be found in Appendices B.8.



**Figure 16: Distribution of longest radial, for sparse viewshed with  $N_s = 16$ .**

### Viewshed Feature: Local Openness

Local openness quantifies the general lay of the land as described by the radials of a sparse viewshed. Defined as the mean of the slopes of the radials, upward openness provides some insight into the general shape of the terrain. Smaller values are found on flatter terrain, which tends to be more open, while larger values are found in more rugged terrain.

Let  $abs(slope_i(r_x))$  be the absolute value of the slope between  $r_x$  and the end of radial  $i$ . Then local openness for a sparse viewshed is calculated as

$$\mu_{slope}(r_x) = \frac{1}{N_s} \sum abs(slope_i(r_x)) \quad \text{for } i = 1, \dots, N_s \quad (23)$$

Figure 17 uses eight radials to depict a sparse viewshed in three dimensions. The colored triangles estimate the terrain's surface between adjacent radials. The gray area

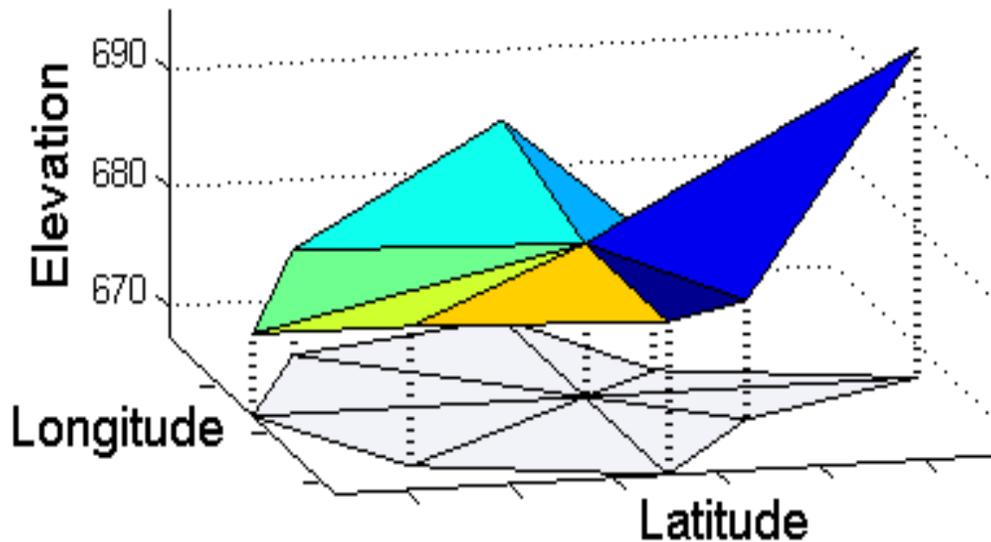
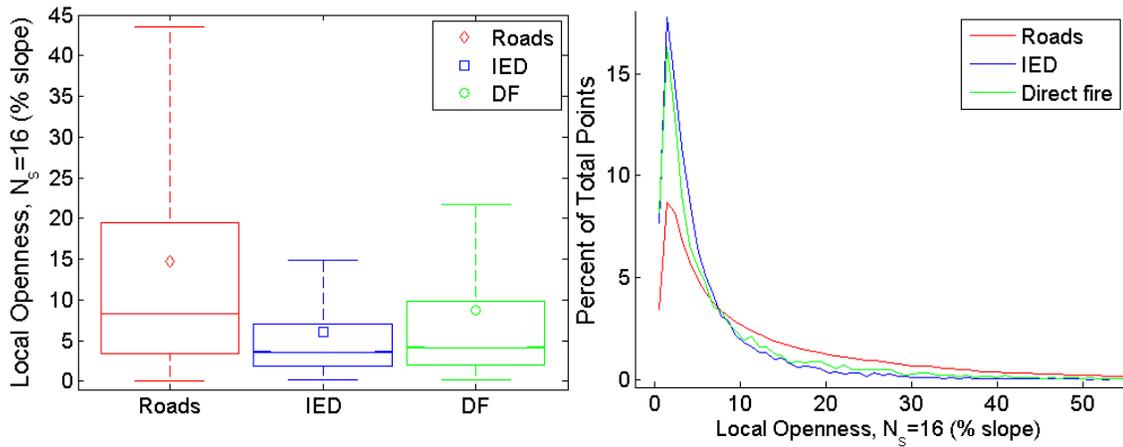


Figure 17: Sparse viewshed portrayed in three dimensions.



**Figure 18: Distribution of local openness, for sparse viewshed with  $N_s = 16$ .**

underneath represents the planimetric area described by the radials. Local openness for this eight-radial sparse viewshed is the average of the absolute values of the slopes of the radials.

Figure 18 depicts local openness calculated using a sparse viewshed with 16 radials. Interestingly, conflict event sites tend to be more open than most road sites with IED sites distributed across the smallest range of openness. Local openness for various sizes of  $N_s$  (4, 8, 32, and 64 radials) can be found in Appendices B.9.

## Geomorphometric Features

There is a wide variety of geomorphometric parameters that describe the underlying morphographic structure of terrain. For a baseline, we use the three part geometric pattern proposed by Iwahashi and Pike in [13] based in part on the work of Horn [14]. The pattern, consisting of slope gradient, texture and local convexity, is

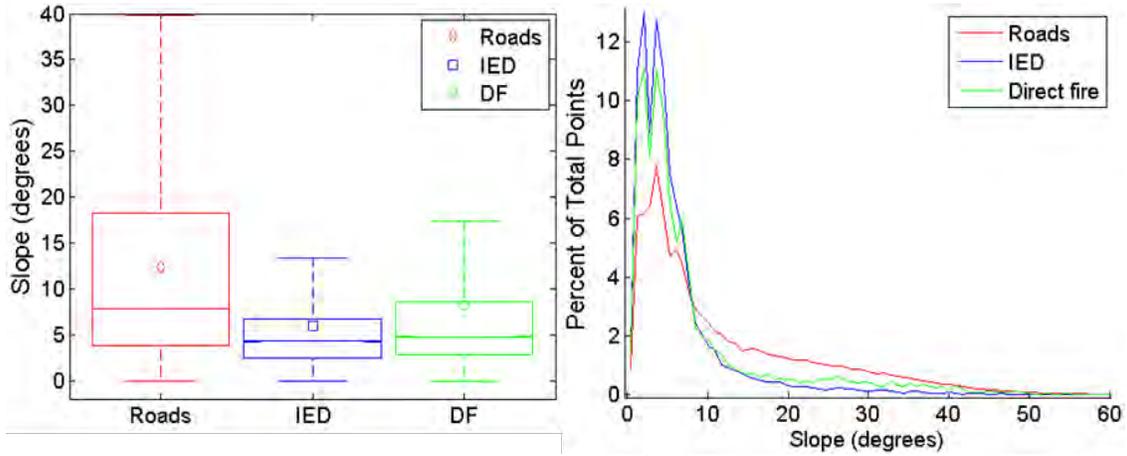
designed to capture key differentiating features of different landscapes. Other geomorphometric features are also collected.

Three important facts should be noted regarding the collection of geomorphometric features in this research. First, the resolution of the elevation data is fixed and limited to approximately 30 meters. Since terrain features are strongly scale dependent, this resolution may be insufficient to capture larger features and vice versa. Second, the geomorphometric features selected are representative and commonly found in the literature. However, they may not be optimal for this resolution or terrain type. Finally, some parameterization, especially for window sizes, is strongly based on anecdotal estimation of the distances required for certain asymmetric warfare activities. These estimates are likely to change if field-based analysis becomes available.

### **Feature: Slope**

Slope is defined as the change in elevation per meter of distance along the path of steepest ascent or descent. It is calculated using a 3-pixel x 3-pixel (3x3) analysis window centered on the elevation map pixel containing the location of interest,  $r_x$ . Matlab provides slope as an output of the *gradient* function and calculates it as:

$$\nabla \bar{z} = \left( \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y} \right), \quad SLOPE = \arctan(|\nabla \bar{z}|) \quad (24)$$



**Figure 19: Distribution of Slope.**

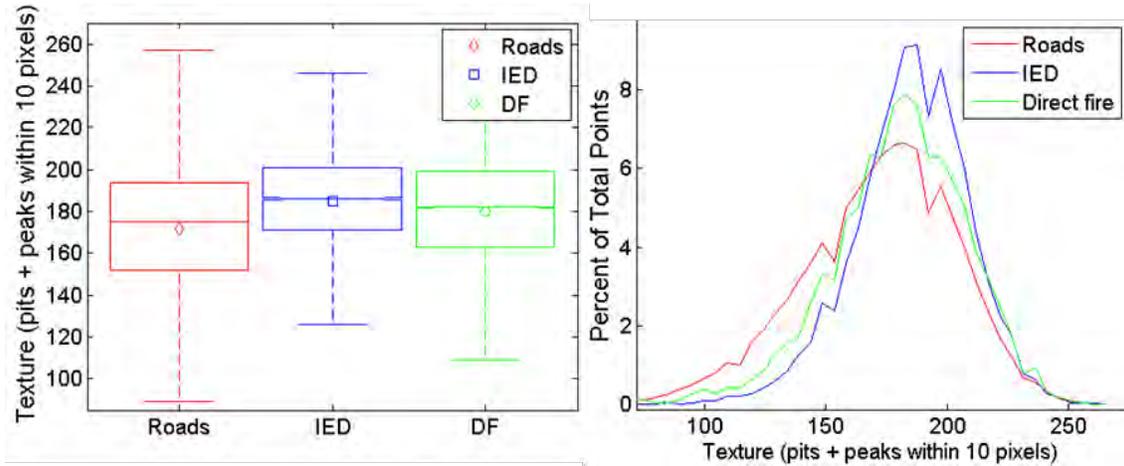
Figure 19 compares the distributions of slope for roads, IED, and DF. Notably, events tend to be on flatter sites than most roads. This is similar to the observations from the local openness feature. Additionally, conflict events and roads appear to be from different populations, based on the width and positioning of the boxplot notches.

### Feature: Texture

Texture is defined by Iwahashi and Pike as the total number of pits and peaks within a ten pixel radius of a point [13]. A 3-pixel x 3-pixel (3x3) median filter is used to smooth the original DEM. The output of the filter is a smoothed DEM that is subtracted from the original DEM and examined for magnitude. Magnitudes greater than zero indicate peaks and magnitudes less than zero indicate pits.

$$T = |DEM - f(DEM)| > 0, \quad f \text{ is a median filter} \quad (25)$$

$$texture(r_x) = \sum_{m,n} T_{m,n} \mid distance(r_x, T_{m,n}) \leq 10 \text{ pixels} \quad (26)$$



**Figure 20: Distribution of texture, as defined by Iwahashi and Pike.**

For context, in the case of Afghanistan, where the publically available DEMs have a resolution of approximately 30 meters, individual textures are calculated over a circular area encompassing approximately 280,000 meters<sup>2</sup>. Figure 20 compares the distributions of texture for roads, IED, and DF events. While all three classes follow a similar distribution, conflict events tend to have greater texture than roads. The boxplots indicate that all three classes are probably drawn from different populations.

### Feature: Local Convexity

The convexity of individual map pixels is found by calculating the surface curvature of a 3x3 DEM subgrid using a Laplacian filter. Matlab calculates convexity,  $Conv$ , using a convolution kernel  $K$ ,

$$K = \begin{bmatrix} 0.1667 & 0.6667 & 0.1667 \\ 0.6667 & -3.3333 & 0.6667 \\ 0.1667 & 0.6667 & 0.1667 \end{bmatrix} \quad (27)$$

$$Conv_{i,j} = \sum_{k=1}^3 \sum_{l=1}^3 DEM(i+2-k, j+2-l)K(k,l) \quad \forall i,j \in DEM \quad (28)$$

Since local convexity, as defined by Iwahashi and Pike, only counts pixels with positive values, all values less than or equal to zero can be set to zero.

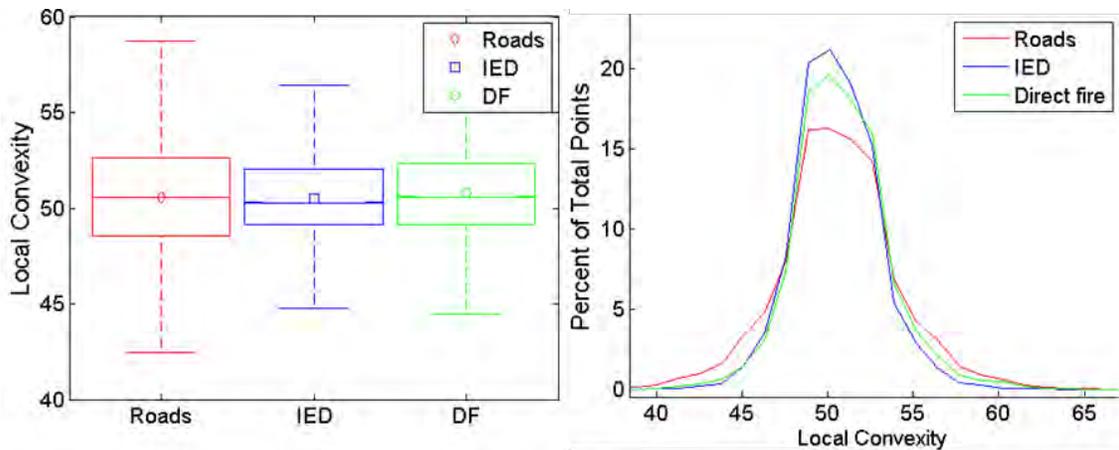
$$C = Conv > 0 \quad (29)$$

Then, local convexity is defined as the percentage of convex upward (positive) pixels within a ten pixel radius of a point [13].

$$LocalConvexity(r_x) = \frac{100}{N_p} \sum_{m,n} C_{m,n} \mid distance(r_x, C_{m,n}) \leq 10 \text{ pixels} \quad (30)$$

, where  $N_p$  is the number of pixels within a ten pixel radius of  $(i,j)$ .

As shown in Figure 21, for all types of conflict events, the distribution of local convexity varies little and the range of convexity values is relatively narrow.



**Figure 21: Distribution of local convexity, as defined by Iwahashi and Pike.**

## Feature: Elevation Range

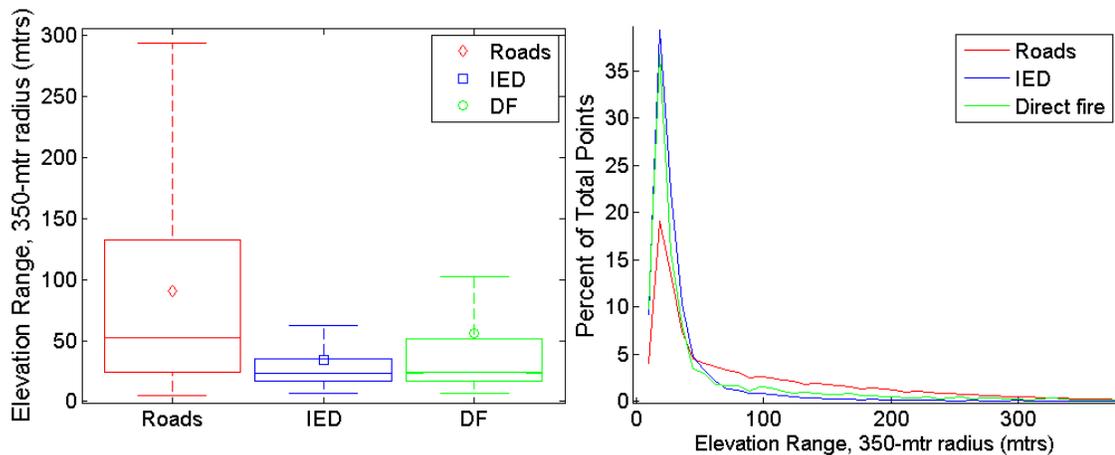
Elevation range is the difference between the highest and lowest elevation in a window [4]. For this analysis, we examine the difference at a radius of 350 meters. Note that in Equations 31 and 22,  $G(r_x)_{rad}$  is a matrix of elevations within  $rad$  meters of  $r_x$ .

$$G(r_x)_{rad} = DEM \mid distance(r_x, DEM_{m,n}) < rad \quad \forall m, n \in DEM \quad (31)$$

$$range(r_x)_{rad} = max(G(r_x)_{rad}) - min(G(r_x)_{rad}) \quad (32)$$

Figure 22 shows that the elevation range is markedly different between roads and conflict events using a window size of 350 meters. At this window size, roads have a larger range of values while conflict events are seen on terrain with smaller ranges. This small range of values may indicate that attackers prefer flatter or smoother ground in the vicinity of an attack site.

Graphs showing elevation range at window radii of 50, 100, 500 and 1000 meters can be found in Appendix B.10.



**Figure 22: Distribution of elevation range at a radius of 350 meters.**

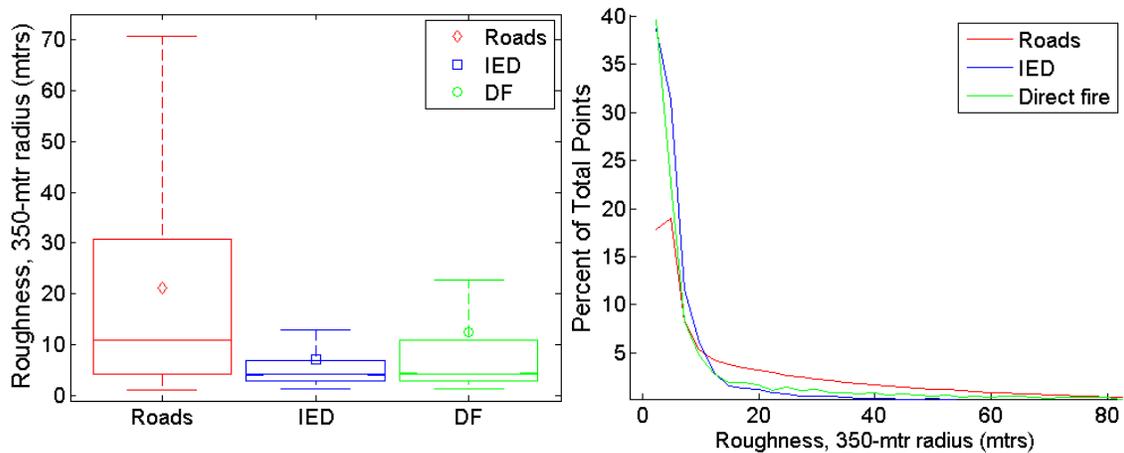
## Feature: Roughness

Roughness uses the standard deviation of elevation across a window to estimate the texture of a surface. Large standard deviations are an indication of a more undulating or rougher surface. Determine  $G(r_x)_{rad}$  as in Equation 31. Then the standard deviation of elevation across  $G(r_x)_{rad}$  is

$$\sigma(r_x)_{rad} = \sigma(G(r_x)_{rad}) \quad (33)$$

, where  $\sigma$  is the standard deviation function.

Figure 23 compares the distribution of roughness across the three classes for a window size of 350 meters. The classes of events tend to be on smoother ground than roads. Notably, the distribution of roughness closely resembles the distribution of elevation range. Appendix B.11 contains figures for roughness across other window sizes.



**Figure 23: Distribution of roughness over a radius of 350 meters.**

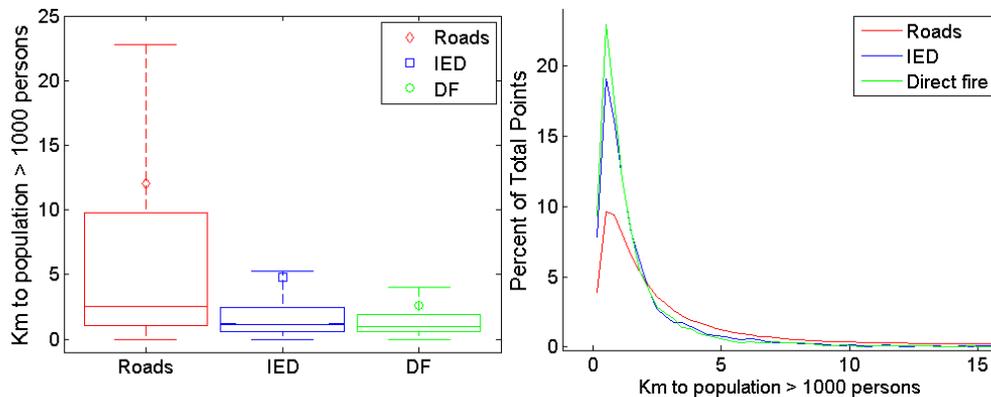
## Social/Cultural Features

Some features are not directly linked to geomorphometry or viewshed. These features, often related to social or cultural factors, capture aspects of site selection not related to the land itself. Proximity to populated areas is explored below.

### Social/Cultural Feature: Proximity to Populated Areas

In some cases, attackers may require access to populated areas. Access may be for logistical reasons, e.g. attackers need access to communications, lodging, etc., or for cover and concealment, e.g. attackers may be able to hide within the local populace.

Figure 24 shows the distribution of the distance from conflict event sites to the nearest population center with more than 1000 inhabitants. Notably, conflict events tend to be much closer to inhabited areas than points along roads with a median value of approximately 1 km. Appendix B.13 shows the distributions for distances to populated areas ranging in size from 1 to 1 million inhabitants.



**Figure 24: Distribution of distance to nearest populated area with greater than 1000 inhabitants.**

## Comparison of Features across Classes

In the previous sections, a total of 20 different features were collected. Some of the features were collected at various geographic window sizes or at various resolutions. The resulting feature set consists of 77 measurements. Table 2 summarizes the results of multiple Kruskal-Wallis tests to determine if the observations for distinct classes (roads, IED and DF) come from different populations.

The Kruskal-Wallis (KW) test is a non-parametric statistical test designed to assess if the measurements for two or more classes come from the same population. It tests the null hypothesis—that all measurements are drawn from the same population—by comparing the medians for each class. As a non-parametric test, KW makes no assumptions about distributions of the measurements or residuals. The test only assumes that all measurements are independent and that they are all drawn from the same continuous distribution. The output of the KW test is the  $p$ -value for the null hypothesis. Small  $p$ -values call into question the null hypothesis and indicate that, in the measurements used, at least one class median is significantly different from the others.

Table 2 provides the output of multiple KW tests. For each of the 77 measurements in the two-column table, three different KW tests were run comparing (1) roads with IED events, (2) roads with DF events, and (3) IED events with DF events. The  $p$ -values are captured in the table and significant values ( $p \leq 0.05$ ) are highlighted in yellow. As shown, roads and events (of any type) are very likely to have been drawn from different populations. These results hold for IED versus roads in 73 of 77 features. DF versus roads shows similar results. Interestingly, IED and DF events appear to come

from the same population for 55 of 77 measures. For IED and DF events, the 22 measures assessed by the KW test to come from different classes actually include 11 measures of route visibility and five measures of shape complexity. In other words, IED and DF events are assessed to be from different populations in only eight of 20 features.

Appendix C contains tables summarizing key statistics for datasets used in this research (mean, standard deviation, skewness, kurtosis, maximum and minimum).

### **Summary of Features**

A total of 20 features were collected at various resolutions or using various geographic windows. These 20 features produced a total of 77 distinct measures. The following measures can be used to describe the Emplacement site:

- Elevation, slope, convexity, texture, elevation range at 50 meters, roughness at 50 meters, local openness (all resolutions), distance to populated areas, and route visibility at 100 meters.

Monitor/control sites can be described using

- Elevation range and roughness at radii greater than 50 meters, visibility index, discrete shape complexity, long radial, short radial, mean radial, planimetric area, rugosity, sparse viewshed shape complexity, cumulative escape adjacency and route visibility at radii greater than 100 meters.

As seen by visual inspection of the boxplots, several of the measures are clearly different for different classes. Table 2 offers one way to quantify this difference and support the intuition gained from the visual inspection. Given this difference, it is probably possible to use these features for predictive analysis.

**Table 2: Kruskal-Wallis test results**

Feature	Rds : IED	Rds: DF	IED : DF	Feature	Rds : IED	Rds: DF	IED : DF
Elevation	1.11E-13	1.32E-12	0.00392 8	Local open (16)	9.07E-239	3.29E-147	0.1321
Slope	5.43E-255	3.05E-122	1.08E-06	Planimet area (16)	6.31E-18	1.55E-14	0.8096
Convexity (IW)	0.5576	0.06056	0.04094	Rugosity (16)	5.23E-18	6.57E-11	0.5024
Texture (IW)	2.96E-108	4.44E-56	0.00538 7	Shape cmplx (16)	2.45E-20	1.63E-06	0.00960 4
Elev. Rng (50)	1.13E-201	2.94E-146	0.7953	Short radial (32)	1.12E-08	1.52E-07	0.8531
Elev. Rng (100)	8.10E-239	8.89E-150	0.0926	Long radial (32)	1.65E-20	1.02E-15	0.9818
Elev. Rng (350)	5.935e-315	4.60E-188	0.1276	Mean radial (32)	2.86E-19	1.04E-16	0.6169
Elev. Rng (500)	4.397e-322	1.88E-198	0.4111	Local open (32)	2.49E-240	2.44E-147	0.1145
Elev. Rng (1000)	1.808e-316	5.96E-192	0.5589	Planimet area (32)	1.38E-18	1.64E-16	0.5914
Roughness (50)	9.85E-196	1.53E-143	0.6819	Rugosity (32)	1.23E-21	6.36E-11	0.1907
Roughness (100)	1.07E-223	5.28E-144	0.2267	Shape cmplx (32)	3.12E-14	0.001562	0.00719
Roughness (350)	8.66E-300	1.43E-179	0.1431	Short radial (64)	7.34E-10	7.76E-09	0.7676
Roughness (500)	0.00E+00	1.14E-191	0.5112	Long radial (64)	2.74E-21	5.90E-18	0.7203
Roughness (1000)	0.00E+00	2.12E-196	0.7572	Mean radial (64)	1.40E-19	9.99E-17	0.6435
Vis Idx(100_350)	1.54E-05	4.80E-08	0.1539	Local open (64)	8.24E-241	3.98E-148	0.1272
Vis Index (350)	9.44E-06	3.38E-08	0.1591	Planimet area (64)	6.62E-19	8.63E-17	0.5935
Vis Index (500)	0.000149 1	1.93E-08	0.07718	Rugosity (64)	5.19E-18	2.69E-14	0.9133
Vis Index (1000)	0.1004	2.67E-06	0.02136	Shape cmplx (64)	3.88E-11	0.00224	0.03802
SCID (100_350)	0.01643	1.11E-05	0.06298	Dist. to ppl (1)	0	5.17E-246	0.6226

SCID (350)	9.44E-06	3.38E-08	0.1591	Dist. to ppl (1k)	0	1.61E-293	0.4416
SCID (500)	0.0001491	1.93E-08	0.07718	Dist. to ppl (10k)	3.62E-182	6.51E-163	0.5654
SCID (1000)	0.1004	2.67E-06	0.02136	Dist. to ppl (50k)	1.39E-246	1.16E-179	0.05796
Short radial (4)	1.47E-05	1.78E-05	0.6582	Dist. to ppl (100k)	1.61E-23	1.58E-39	2.73E-08
Long radial (4)	1.18E-07	7.73E-07	0.7779	CEA min	2.23E-79	1.17E-74	0.2858
Mean radial (4)	1.07E-09	2.67E-10	0.3949	CEA max	3.25E-168	2.18E-116	0.07424
Local open (4)	2.03E-237	1.02E-147	0.1425	CEA med	1.92E-248	4.56E-189	0.945
Planimet area (4)	1.65E-09	5.09E-10	0.4079	Rte Vis. (min 1k)	3.26E-127	4.17E-06	5.95E-26
Rugosity (4)	4.60E-07	0.0004824	0.4863	Rte Vis. (max 1k)	0.0009066	0.7642	0.03552
Shape cmplx (4)	7.81E-25	1.86E-10	0.03742	Rte Vis. (med 1k)	0.3002	9.45E-09	0.000248
Short radial (8)	7.21E-07	2.55E-07	0.5359	Rte Vis. (min 500)	1.04E-192	3.68E-17	3.47E-27
Long radial (8)	8.77E-13	1.46E-09	0.8917	Rte Vis. (max 500)	5.61E-25	4.75E-09	0.01279
Mean radial (8)	2.04E-15	2.87E-12	0.8934	Rte Vis. (med 500)	7.91E-11	0.98	2.87E-05
Local open (8)	2.55E-241	9.74E-149	0.1198	Rte Vis. (min 250)	1.16E-190	1.54E-23	5.01E-20
Planimet. area (8)	4.49E-14	1.09E-11	0.8191	Rte Vis. (max 250)	6.38E-54	3.41E-22	0.00094
Rugosity (8)	2.86E-10	3.06E-09	0.7138	Rte Vis. (med 250)	3.61E-41	1.21E-08	2.81E-05
Shape cmplx (8)	2.02E-20	1.08E-09	0.1107	Rte Vis. (min 100)	0.004383	1.49E-30	1.56E-11
Short radial (16)	1.28E-07	0.000156	0.5149	Rte Vis. (max 100)	1.61E-15	5.41E-06	0.06075
Long radial (16)	3.78E-19	6.75E-13	0.6864	Rte Vis. (med 100)	8.42E-40	0.05053	1.30E-10
Mean radial (16)	1.22E-18	1.20E-14	0.8745				

## **4. PREDICTIVE ANALYSIS OF ASYMMETRIC CONFLICT**

### **EVENTS**

Predictive analysis uses a variety of techniques to analyze historical data for the purpose of making predictions about the future or about unvisited locations. In this section, we propose algorithms to populate and use the MECH model. The purpose of this study is to design an accurate and robust classification algorithm that learns from available data under realistic constraints. In the development of the algorithm, subset selection and principal component analysis are compared as mechanism to reduce dimensionality. Classification is performed on the resulting reduced data using supervised parametric and non-parametric techniques including Support Vector Machines (SVM), Discriminant Analysis (DA) and k Nearest Neighbor (kNN) classifiers. Different geographic-temporal constraints are applied to take advantage of the locality property.

### **The MECH Classification Algorithm and Evaluation Criteria**

Conflict events are rare. When overlaid on a tokenized road map of Afghanistan, conflict events only occupy 0.8% of the total number of points comprising the roads. Conflict events are notably different than average road points. Route visibility, shortest radial, local openness, elevation range, and distance to populated areas are all features where the distribution of conflict events and road points are clearly different. Knowing that this difference exists, we can design an effective classification algorithm that

adaptively learns from the subset of MECH features most relevant to the area and time under consideration.

The main procedure of MECH classification algorithm is summarized in Table 3 and is explained below. A constrained set of points is identified that include both recent conflict event sites and non-conflict event sites in the local area. Then, features are collected from these sites based on the previous description in Section 2. The MECH model is comprised of two tactical patterns: one composed of Emplacement features,  $\tau_E(r_x)$ , and one composed of Monitor/Control features,  $\tau_H(r_x)$ . Together, these features form the tactical pattern of the conflict event  $T(r_x) = [\tau_E(r_x) \ \tau_H(r_x)]$ . This pattern is the core of the MECH classification algorithm.

Once features have been collected, normalization factors are dynamically determined from the data. This ensures that scaling is determined from the data set and appropriate for the terrain and tactics described by the data. Next, relevant features are determined from the same local data. The resulting set of features identified as relevant varies with tactics, terrain, and time. Finally, parameterization of the model is derived from local data. Thus, each time the model is used for predictive analysis, it is uniquely tuned to past events, terrain, and the tactics in use in the local area.

Two main criteria will be used in evaluation step: Percent error and event classification error. *Percent error* is the percent of all classifications that are not correct. Cast in terms of the conventional confusion matrix, *percent error* or *error rate* is

$$Percent\ Error = 100 * \frac{False\ Positive + False\ Negative}{All\ Positive + All\ Negative}$$

A second measure is *event classification error*, which is the percent of misclassified events. Commonly described as the *complement to the precision*, it is found as

$$\text{Event Classification Error} = 100 * \frac{\text{False Positive}}{\text{All Positive}}$$

**Table 3: The MECH Classification Algorithm**

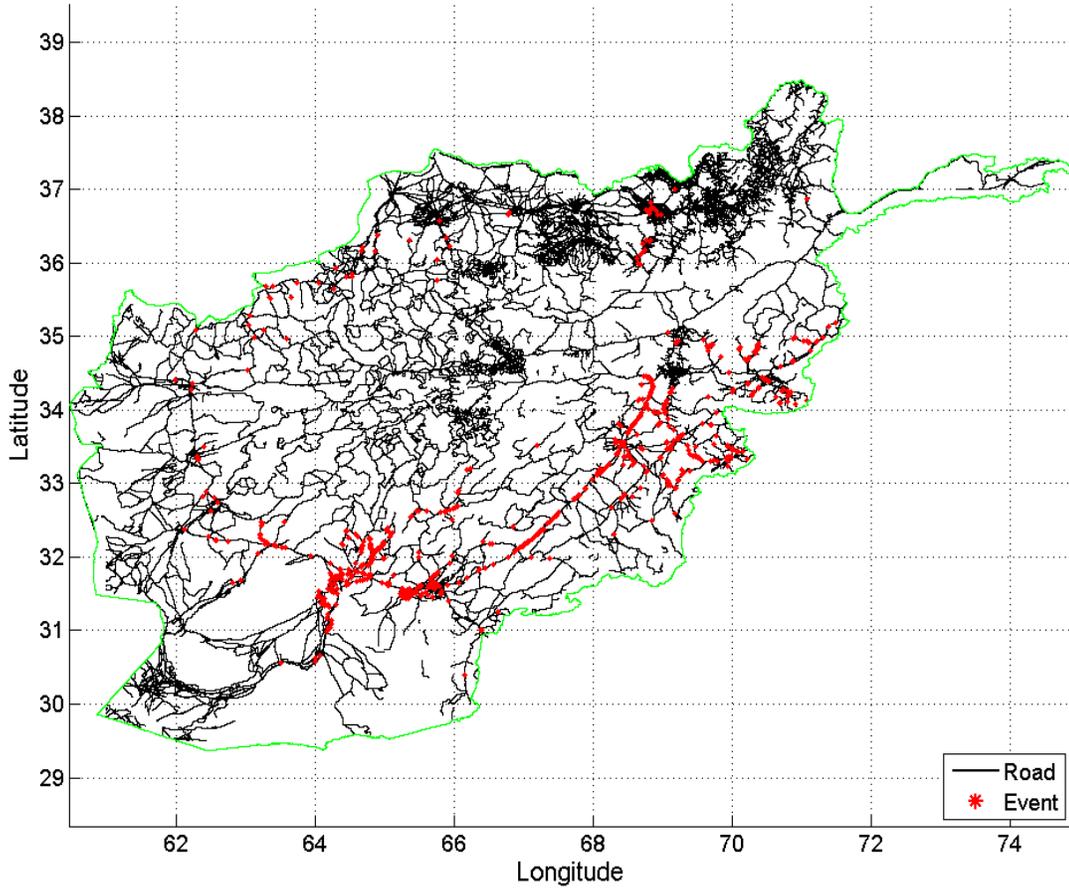
<p><b><u>Objective</u></b> Classify Points Along Roads</p>
<p><b><u>Algorithm</u></b>  Step 1: Select a dataset.  Given a set of locations collected from conflict events and road points, select the conflict events and road points to include in the training set  (i) Apply geographic constraints  (ii) Apply temporal constraints  If there are enough events in the resulting constrained sample  (iii) Divide the data into training and test sets  (iv) Collect Emplacement and Monitor/Control (E and M/C) features    Step 2: Prepare the data  (v) Normalize the training set  (vi) Normalize the test set using scaling factors from the training set    Step 3: Train and assess the model  (vii) Select relevant features  (viii) Determine model parameters and/or hyperparameters  (ix) Learn the classification rules for the training set  (x) Estimate classification accuracy  Step 4: Classify  (xi) Apply the classification rules</p>

### **Data Source, Pre-processing and Quality assessment**

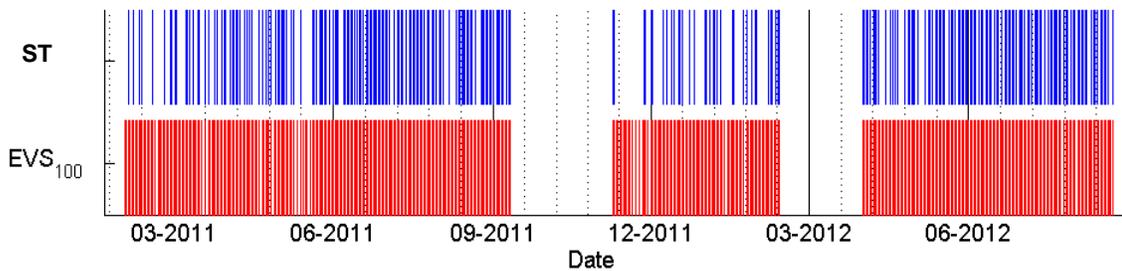
A detailed description of the raw data for conflict events and the road is in Appendix A.2. Here we present the pre-filtering on the raw data, analyze the quality of the data and infer the possible consequence of noisy data.

Two past event classes are formed by IED events within 100 meters of a known road ( $IED_{100}$ ) and direct fire events within 100 meters of a known road ( $DF_{100}$ ). Collectively, these two sets comprise  $EVS_{100}$ , the set of all events within 100 meters of a road. Roads are tokenized into discrete points at an interval of 30 meters, which coincides approximately with the elevation map resolution. Points along the road that are at least 250 meters from any known conflict event ( $RD_{250}$ ) are used as the non-event class. Many of the figures displayed are based on  $ST$ , a dataset consisting of 250 events drawn randomly from  $IED_{100}$  and 250 events drawn randomly from  $DF_{100}$ . The dates and locations of the  $ST$  events provide a consistent set of geographic and temporal coordinates that are used to test and compare algorithms. Figure 25 shows the temporal and geographic distribution of  $ST$ .

The restricted conflict event sets used for learning are a reflection of an incomplete road dataset. This is evidenced by the fact that many conflict event sites are located away from known roads. Appendix A.3 illustrates this problem. The  $IED_{100}$  and  $DF_{100}$  datasets were chosen to ensure that the features collected from conflict event locations used for training were comparable to features collected from points known to lie along roads. The inclusionary radius of 100 meters ensures that conflict event points used for training lie sufficiently close to known roads that all features may be collected. This criterion is particularly applicable to features based on visibility and cumulative escape adjacency. Even so, we inevitably introduce a bias between events and non-event road locations when calculating the route visibility features.



(a)



(b)

**Figure 25: Distribution of random sample of past events,  $ST$ , used for analysis and parameter discovery of the MECH classification algorithm; (a) geographic distribution of  $ST$  locations and roads; (b) temporal distribution of  $ST$  compared to the temporal distribution of  $EVS_{100}$ .**

The  $RD_{250}$  dataset was chosen to minimize suspected problems with the conflict event data. First, it is not clear that the location of the conflict event is identified in the original data using consistent criteria. For example, the actual location of an exploded or discovered IED is very clear from physical observation. Blast marks, craters or the actual device can be seen and the location measured accurately. However, the reported location is not always so accurate. It may be the actual site of the explosion, an estimate of the location made from a distance, or the location of the person reporting the event. In the case of a military patrol, a person reporting the event may be a hundred or more meters away since patrols typically maintain 25-meter or greater spacing between vehicles. The location chosen to represent a direct fire event is similarly unclear. The reported location may be the first vehicle in the patrol, the first vehicle to come under fire, the vehicle of the person reporting the event, or an estimated ‘center’ of the kill zone or ambush site. The 250-meter-radius exclusionary zone around known conflict event sites provides a buffer that attempts to mitigate these problems.

It is important to note another significant issue with the  $RD_{250}$  dataset.  $IED_{100}$  and  $DF_{100}$  contain single classes of data: locations and dates where known conflict events occurred. The  $RD_{250}$  dataset, although labeled as a single class, actually contains at least three classes:

1. Locations that are not useful for conflict events and will never be used;
2. Locations that are useful for a conflict event but have not been used yet; and
3. Locations that have already been used for a conflict event but this event occurred before or after the time period covered by the available conflict event data.

Thus, when  $RD_{250}$  is used as the non-event class in a two-class classification algorithm, it is likely that some misclassified non-events are members of (2) or (3). One impact of this issue is that the overall misclassification rate may not be a good indicator of performance. Instead, it is necessary to examine both the overall misclassification rate as well as the individual misclassification rates of each class.

## Classification Algorithm Introduction and Baseline Results

In this sub-section, we first introduce underlying details of the classification algorithm. Then, we will present some baseline performance.

### Set selection

Attackers may vary their attacks with geography and over time. Attacks may vary with geography for some reasons including the shape and structure of the terrain, availability of critical tactical elements like overwatch sites, proximity to attacker safe zones, and political and tribal boundaries. Over time, attacks may vary due to the deployment of countermeasures, availability of war materials, the experience level of the attackers, and sophistication of the target. This spatial and temporal variation requires us to apply both geographic and temporal constraints when producing training and evaluation tests sets. The training set produced by these constraints contains roads and events that are within *radius* of a specified *loc*. Training events are further constrained to be within a specified timespan before or on *date*. Let *loc* and *date* describe the geographic location and date of some event and let *radius*, *span<sub>1</sub>* and *span<sub>2</sub>* describe the

geographic window radius, training timespan and test timespan. Then the training sample is found by

$$event\_obs = events \mid distance(loc, events) < radius \quad (34)$$

$$trn\_event\_obs = event\_obs \mid -span_1 \leq timespan(date, event\_obs) \leq 0 \quad (35)$$

$$roads\_obs = roads \mid distance(loc, roads) < radius \quad (36)$$

$$trn\_sample = [trn\_event\_obs; rand\_select(roads\_obs, k)] \quad (37)$$

$$k = |trn\_event\_obs|$$

The test set is found similarly, using  $span_2$  as the timespan.

Although the use of geographic and temporal constraints produces sets of events and terrain that are probably more homogeneous, a side effect is that the training and test sets are temporally disjoint. In some parts of the year, this means that the test events will occur in a completely different season from the training events. Another problem is that smaller windows tend to produce smaller data sets. For some machine learning algorithms, small training sets may be difficult to use or may produce unreliable results.

## Data Normalization

Once constraints have been applied, the resulting datasets need to be prepared for use in machine learning techniques. The raw data must be centered and scaled so that the resulting features share a common mean at or near zero and approximately equal ranges. Z-score is probably the most widely used normalization technique. Calculated using mean for location and standard deviation for scale, the z-score for a single feature or measure  $s$  for an instance  $i$  is found by  $s'_i = \frac{s_i - \mu(s)}{\sigma(s)}$

## Feature Reduction

We conduct feature reduction to remove noise, remove correlation among features, and reduce dimension due to the curse of dimensionality. Two dimension reduction methods are used here. Unsupervised principal component analysis (PCA) [15] transforms the data from high dimensional space into low dimensional orthogonal space that conserve most of its variance. Supervised stepwise feature selection (STP) [16] is a regression-based iterative greedy algorithm. It evaluates the importance of the feature based on coefficients of the linear regression model.

## Classifier Training Algorithm

The outcome of feature reduction is a reduced, labeled training set and the parameters needed to normalize and reduce the test set. The following sections examine the classification of this data using three machine learning algorithms based on different heuristics [16].

- kNN is based on the heuristic of density estimation. The density function for each class at each location in high dimensional feature space is estimated by the number of instances of the current class in unit space volume around the current location. The label of a new instance is assigned to be the class with largest density value at the location of this new instance. This method is sensitive to rescale of the feature space, and density estimation is not accurate at high dimensionality.
- Discriminant analysis finds a projection that maximizes between-class variance and minimizes within-class variance. This method assumes that each class has a Gaussian

distribution and that mean value is the main difference between different classes. The assumption is questionable for this problem.

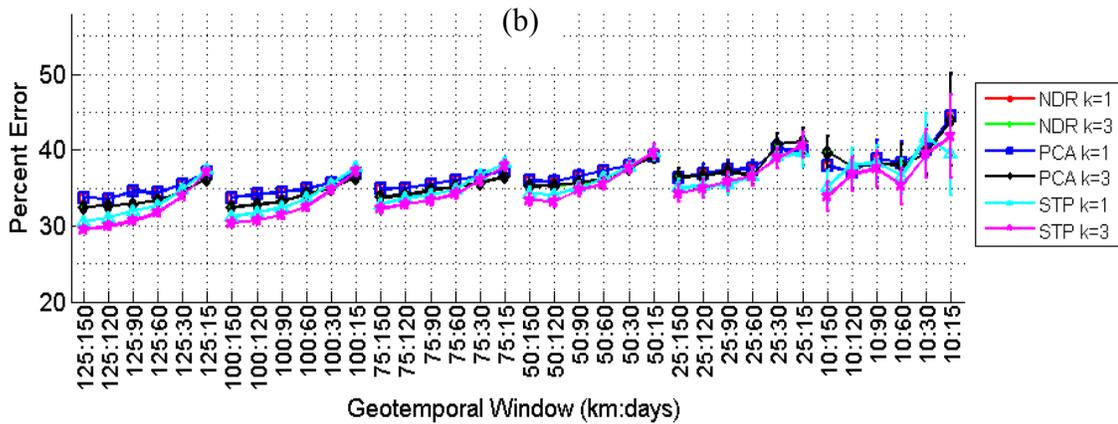
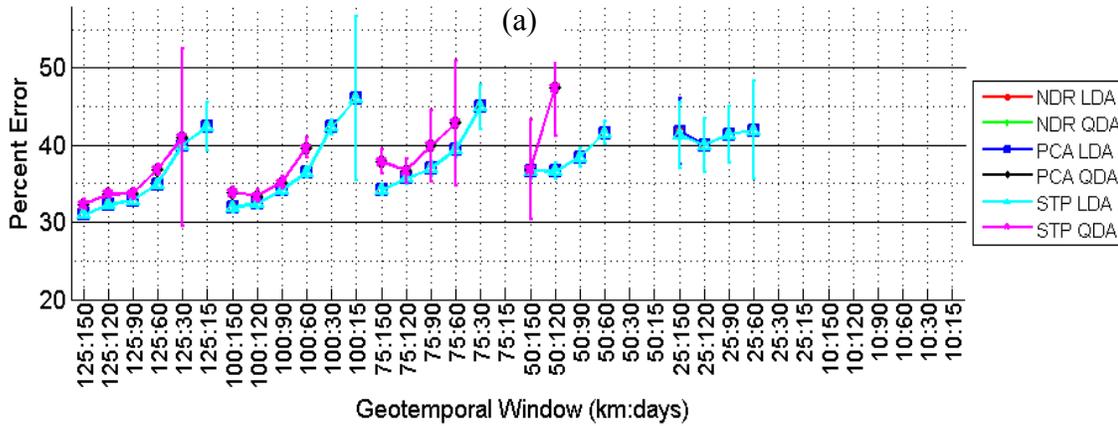
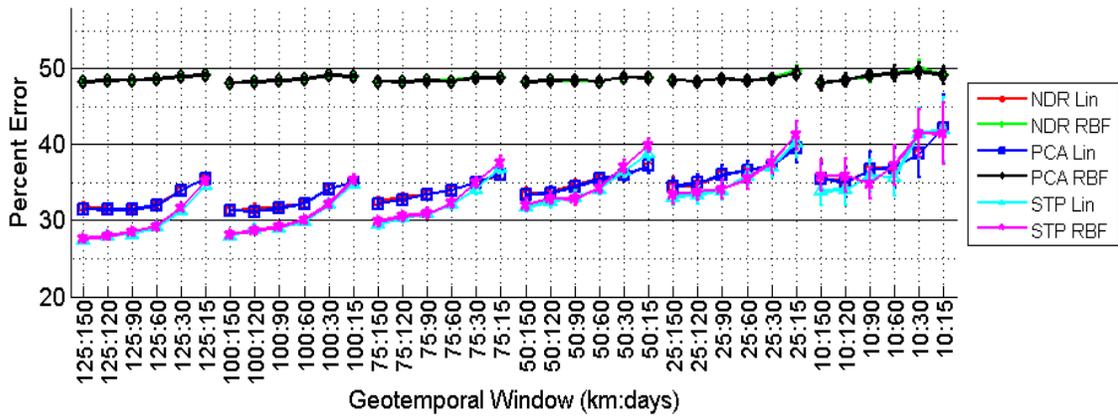
- Support vector machine performs structural risk minimization. This theory shows that an algorithm can achieve the minimal risk of the linear model by maximizing the margin which is described as the minimum distance of an example to the decision hyper plane. This method overcomes the curse of high dimension, and the kernel trick can map features into a high dimensional space that is more separable.

### Baseline Results

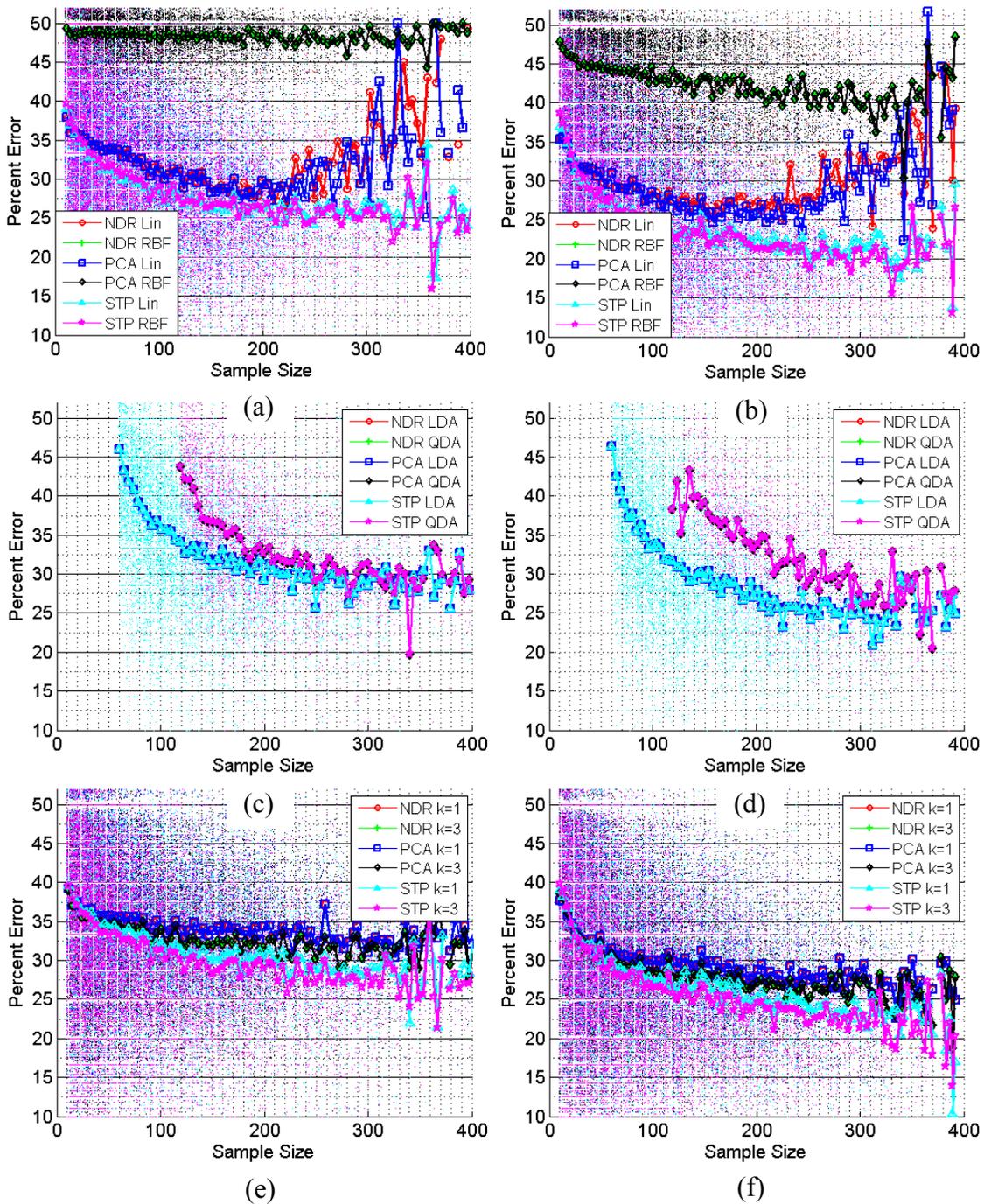
Figure 26 present a brief comparison of set filtering, feature reduction, and machine learning methods on classification performance for IED events (Similar results are obtained for DF events). For each event in  $ST$  and for various durations, when the total number of IED events within the combined geographic and temporal training windows  $b$  exceeds 10, three dimensionality reduction schemes, PCA, STP and NDR(No dimension reduction), are combined with SVM (using linear and RBF kernels), discriminant analysis (linear and quadratic) and kNN (1-NN and 3-NN) to estimate the classification error, which is calculated as the mean of the percent error. The duration of the test set is universally constrained to 60 days. Three-way cross-validation is used and error bars indicate a 95% confidence interval. In the figure, the x-axis is the <geographic window: temporal window > combination used to select the training set for the machine learning algorithm. Note that the plots are grouped by geographic constraint, with a break in the connecting line signifying the jump to the next geographic group. (The lines connecting data points are provided to increase the readability of the figure.)

In Figure 26, classification error is between 30-40% for both SVM and kNN with SVM performing slightly better. As before, SVM with linear kernels performs significantly better than SVM with RBF kernels. Note that these are still using the default parameters for the SVM box constraint and RBF scaling factor. Both SVM and kNN show fairly constant accuracy within each geographic constraint group with an upwards trend in error of approximately 3-5% as the temporal window shrinks. The increase in classification error is more significant for DA and shows significant problems with small datasets due to the limitation of matrix inversion, which is reflected in the missing error values for classification attempts using NDR and PCA.

Figure 27 shows the impact of sample size on classification accuracy under geographic and temporal constraints. In general, classification accuracy improves as the sample size increases. SVM shows the best performance at sample size around 200 while kNN has its best performance at the largest sizes. DA's classification accuracy is similar to that of SVM and kNN for large samples. Small samples continue to be problematic for DA.



**Figure 26: The impact of combined geographic and temporal windows on classification accuracy of IED events; (a) IED SVM; (b) IED DA; (c) IED kNN.**



**Figure 27: The impact of sample size on classification accuracy under combined geographic and temporal constraints; (a) IED SVM; (b) DF SVM; (c) IED DA; (d) DF DA; (e) IED kNN; (f) DF kNN.**

## Parameter Tuning

All statistical methods are based on some assumptions on the model structure, which consists of many parameters (variables) and operations on these variables. The success of these methods on real-world problem depends on correct estimation of the parameters. We conducted a comparative empirical evaluation of the effect of parameters and explore some potential automation mechanisms in parameter estimation.

### Feature reduction: *penter* and the cumulative variance

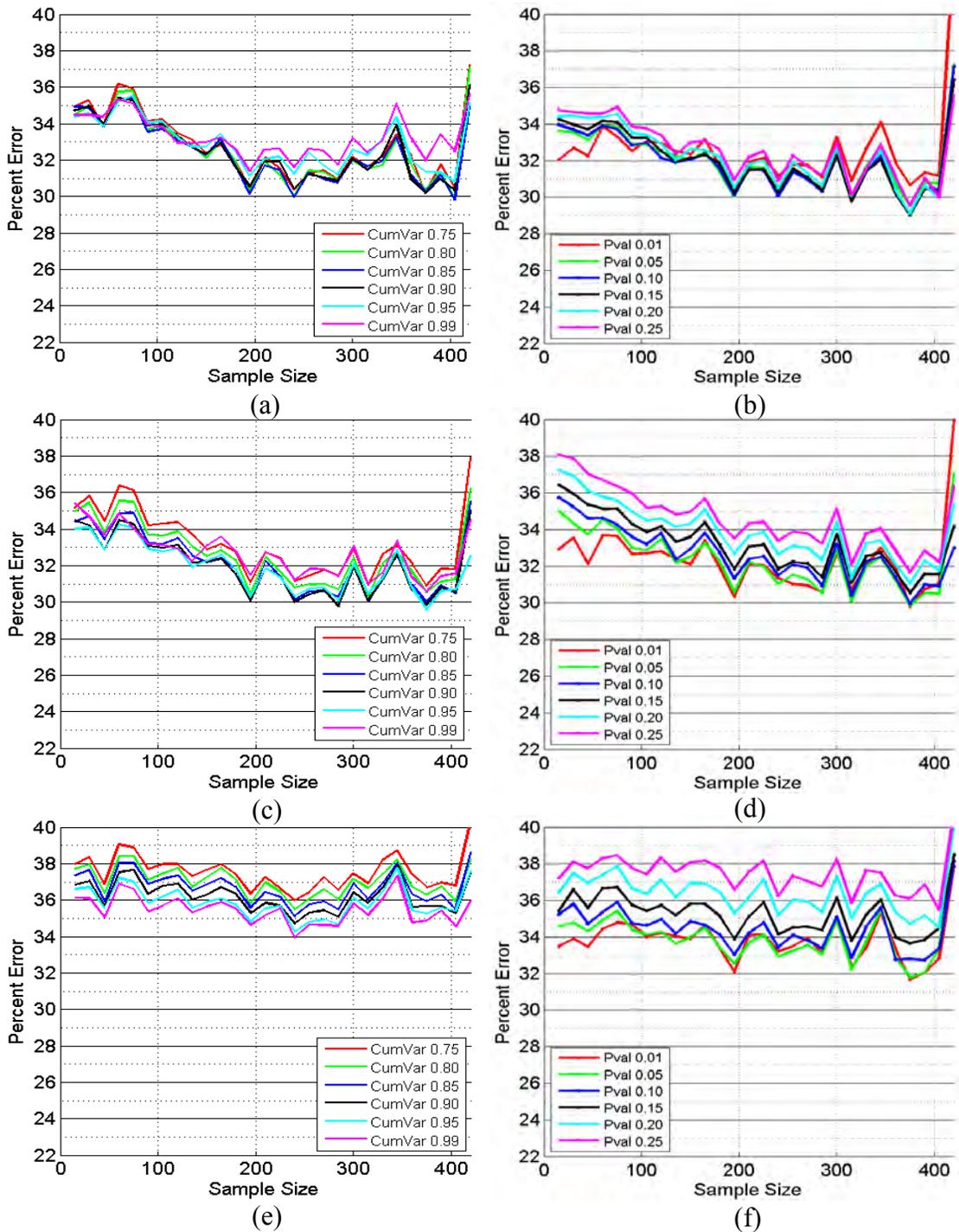
In the previous experiments involving geographic, temporal and geotemporal constraints, default settings were used for PCA and stepwise feature selection. In the case of PCA, all principal components were used, regardless of their contribution. Similarly, for stepwise feature selection, all weighted features were used. However, it may be possible to reduce error by reducing dimensionality. With PCA, one method of reducing dimensionality is by assessing the amount of variance accounted for in the reduced model. The number of principal components in the final model is controlled by limiting the total cumulative variance explained by these components. For stepwise feature selection, dimensionality may be managed by varying the  $p$ -value threshold (*penter* parameter). Smaller  $p$ -values lead to smaller models

Figure 28 examines the impact of varying cumulative variance (for PCA) and the *penter* parameter (for STP). For conciseness, only three combinations of the learning algorithm and dimensionality reduction are applied to the IED data and shown here. Both PCA and STP are combined with SVM using a linear kernel, LDA, and kNN with

$k=1$ . For each combination, the error rate produced by cumulative variances between 0.75 (75%) and 0.99 (99%) and *penter* parameters between 0.25 and 0.01 are shown.

For PCA, changing the cumulative variance appears to have little effect using SVM, LDA or kNN. For all three of these learners, a cumulative variance of 95% showed consistently good performance across the entire range of sample sizes. For kNN, a cumulative variance of 0.99 performed best but this performance was not shared by SVM and LDA.

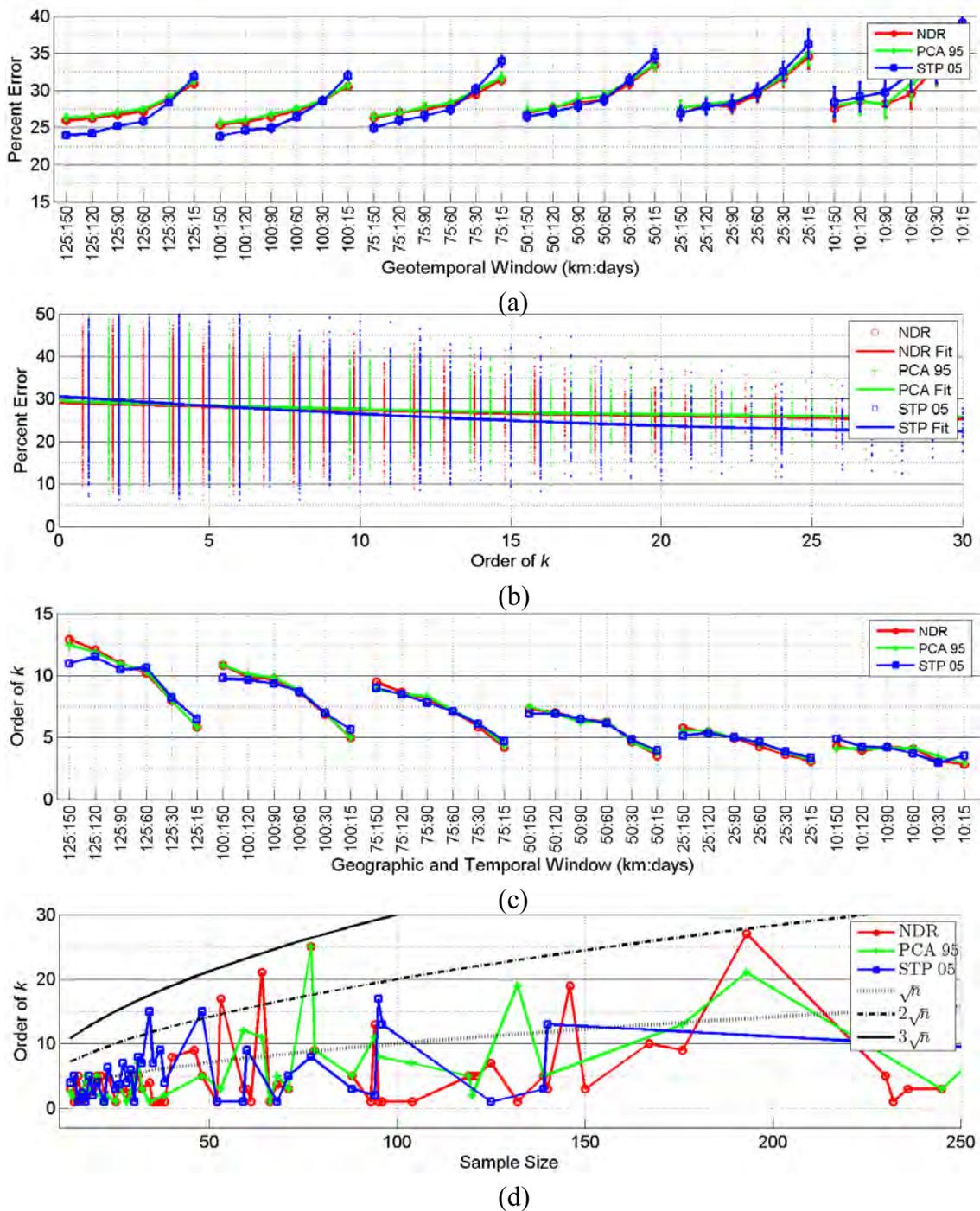
For STP, a *penter* parameter of 0.01 consistently produces the lowest error rates at small sample sizes but performs less well at large sample sizes. However, a *penter* value of 0.05 performs well across the entire range of samples sizes. Error rate differences as high as 5% were seen, with larger *penter* values tending to produce higher error rates, especially at smaller sample sizes.



**Figure 28: The impact of varying cumulative variance and *penter* parameters on IED classification; (a) PCA with SVM linear kernel; (b) STP with SVM linear kernel; (c) PCA with LDA; (d) STP with LDA; (e) PCA with kNN ( $k=1$ ); (f) STP with kNN ( $k=1$ ).**

### Estimation of $k$ for the kNN Classifier

The only parameter for kNN method is the count  $k$  of neighbors and  $k$  is usually determined by choosing the one with the best cross-validation performance on training set using exhaustive enumeration of all possible  $k$  in certain range. As described by Ghosh [17], we bound the upper value of  $k$  to be  $3\sqrt{n}$ , where  $n$  is the size of training data set, and further constrain this upper bound to be no larger than the size of the smallest class. Its optimal value is determined by exhaustively testing all odd  $k$  in the range  $[1, 3\sqrt{n}]$ . Figure 29 examines the impact of dynamically selecting  $k$  on the classification of IED events. A similar figure was obtained for DF events. In Figure 29(a), the dynamic selection of  $k$  decreases the misclassification rate to approximately 25%. The misclassification rate climbs as the window gets smaller, likely a reflection of smaller sample size. Stepwise feature selection slightly outperforms the other dimensionality reduction schemes. Figure 29(b) examines the impact of the order of  $k$  on classification accuracy. The solid lines are a second-order polynomial fit to the available data for each classification method. As  $k$  increases, the classification rate improves slightly. Once again, this is likely to be a reflection of sample size. Figure 29(c) shows how the order of  $k$  changes with window size. In the figure, the median value of  $k$  is presented for each geotemporal window. Interestingly, optimal values of  $k$  tend to be small. Figure 29(d) examines the impact of sample size on the order of  $k$ . Each point is the average of all classification attempts at that sample size, regardless of window. The results are in line with the results reported by Ghosh regarding the upper bound and choice of  $C$ , with most  $k$  at or below  $2\sqrt{n}$  and all  $k$  below  $3\sqrt{n}$ .

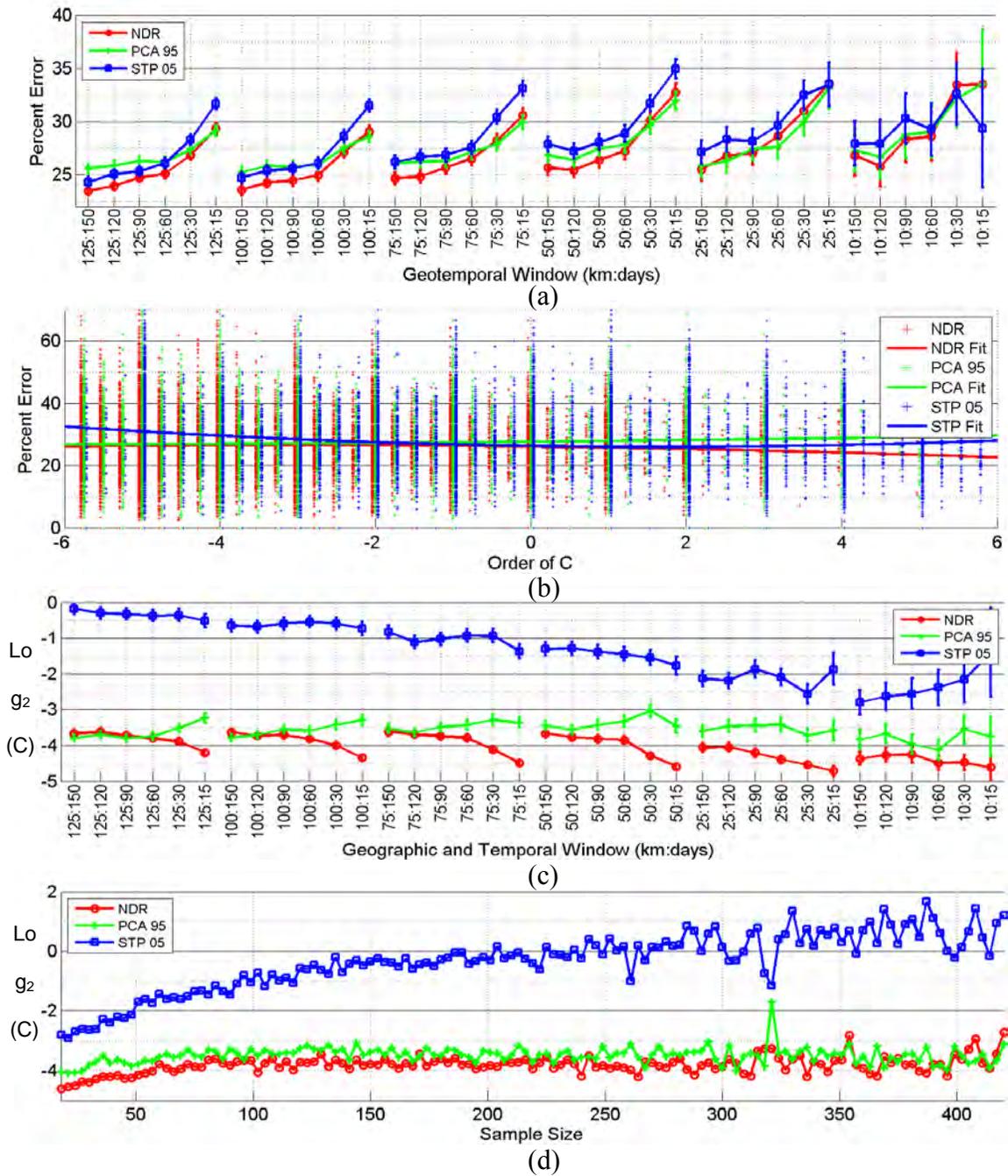


**Figure 29: The impact of dynamically varying the order of  $k$  on the classification of IED events using kNN; (a) Classification accuracy under combined geographic and temporal constraints using varying  $k$ ; (b) Classification accuracy at various values of  $k$ ; (c) Mean order of  $k$  at various window sizes; (d) Mean order of  $k$  at various sample sizes.**

## Box Constraint Estimation for an SVM Classifier with a Linear Kernel

For the case of SVM using a linear kernel, the only parameter is the box constraint,  $C$ . A larger value of  $C$  assesses a larger penalty on samples violating the label assignment by the hyperplane. We implement an exponential search in the range  $C = 2^i$  for  $i \in [-5, 5]$ , in 0.25 increments. This approach is implemented as a two-step process employing a coarse grid using increments of 1 and a fine grid with increments of 0.25 for optimization.

Figure 30 examines the impact of the size of the box constraint  $C$  on classification accuracy when using SVM with a linear kernel. By selecting a more optimal  $C$  than the default of 1, classification error approaching 23% can be achieved for IED events, as noted in Figure 30(a). Figure 30(b) shows that the order of  $C$  has little impact on classification accuracy. Figure 30(c) shows that  $C$  varies little for NDR and PCA with respect to the window size. For STP, the trend is to decrease as the window shrinks. A similar outcome can be seen in Figure 30(d), where the order of  $C$  is fairly constant for NDR and PCA. The order of  $C$  using STP ends to increase with increasing sample size. Note that Figure 30(c) has no  $\log(C)$  values greater than 0 while Figure 30(d) has  $\log(C)$  values greater than zero. This is an impact of sample size. Larger windows tend to produce a larger sample, but not always. Some of the samples are smaller and tend to produce smaller values of  $C$ . A similar results could be got for DF events.



**Figure 30: The impact of dynamically varying the box constraint  $C$  on the classification of IED events using SVM with a linear kernel; (a) Classification accuracy under combined geographic and temporal constraints using varying  $C$ ; (b) Classification accuracy at various values of  $C$ ; (c) Mean order of  $C$  at various window sizes; (d) Mean order of  $C$  at various sample sizes.**

## Parameter Estimation for an SVM Classifier with an RBF kernel

While SVM with a linear kernel has a single parameter to estimate, changing to a radial basis function kernel adds one more parameter,  $\sigma$ . In this research, we examine the classification accuracy obtained by selecting  $\sigma$  using two methods: grid search across a set of fixed values and direct estimation based on class separability in the kernel space as proposed by [18]. For the grid search method, the search range of  $\sigma$  is constrained to  $\sigma \in [0.25, 30]$ , bounds that were empirically determined from analysis of the data. The grid search was implemented as a two-step process using a coarse grid first and then refining the results using a fine grid. The outcome of the grid search for a given set of samples is the  $[C, \sigma]$  pair that produces the lowest classification error. For the direct estimation, Liu and Zuo propose an estimate of  $\sigma$  defined as  $\hat{\sigma} = \sqrt{(B' - W') / (4 \cdot \log(B'/W'))}$ .

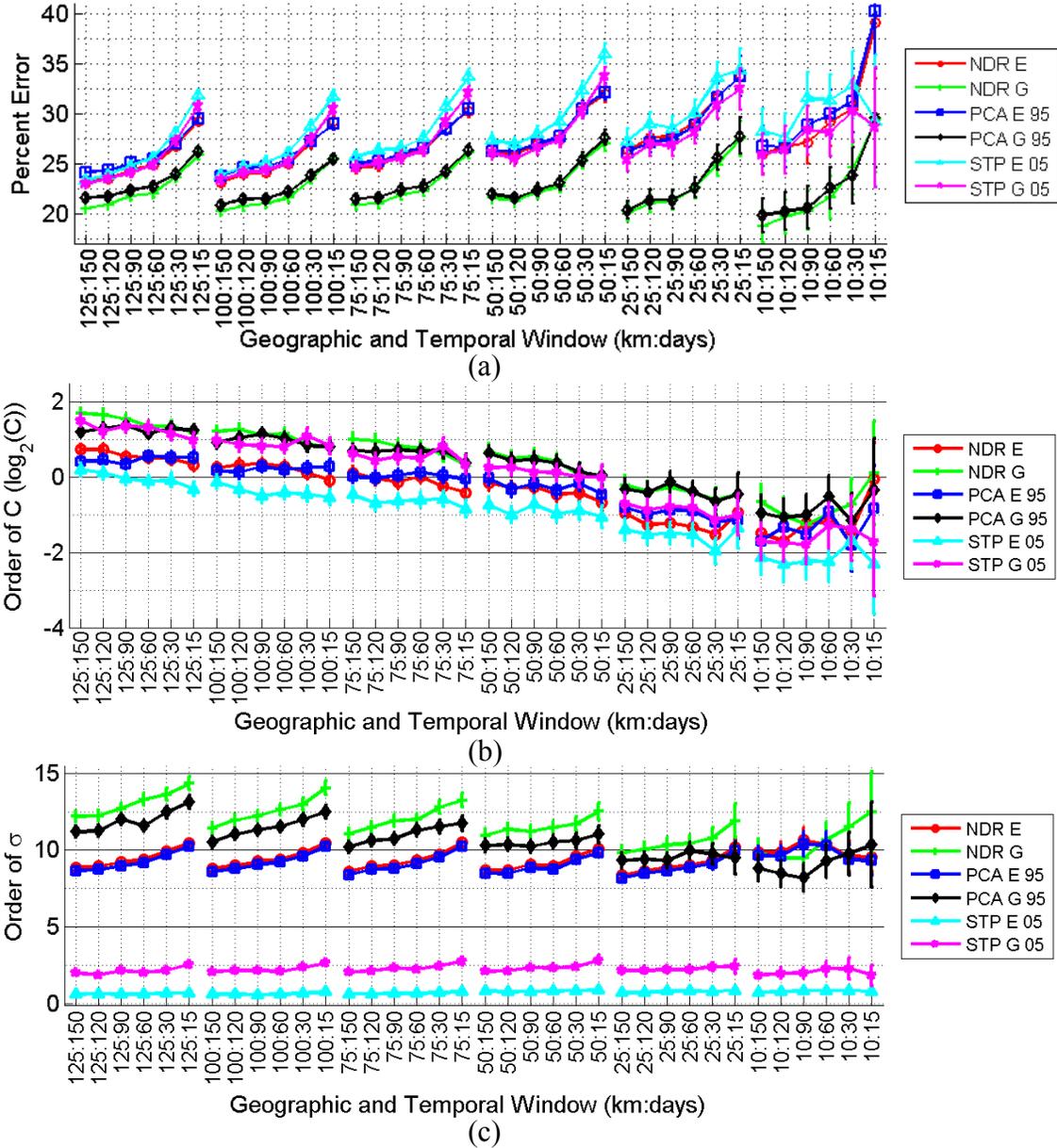
Figure 31 notes the impact of varying  $C$  and  $\sigma$  on the classification of IED events across a variety of geotemporal windows. In the figure, the cumulative variance of the PCA components is constrained to be  $\leq 95\%$  of total variance and the maximum  $p$ -value of the stepwise-selected features is constrained to 0.05. Results found using hyperparameters generated using grid search are marked with a 'G' in the legend. Results found using an estimated  $\sigma$  are marked in the legend with an 'E'. At each window size, geographic and temporal filters were applied to the dataset. Three-way cross-validation was used on the resulting subset to produce the performance data. As shown in Figure 31(a), grid search outperforms  $\sigma$  estimation at every window size (labeled as NDR G and PCA G). Interestingly, the best classification performance is seen in the smallest geographic windows, using NDR and PCA, with error rates approaching 20%. Figure

31(b) shows that typical mean values of  $C$  tend to be small. Mean values of  $\sigma$  tend to show little change with geographical window in Figure 31(c). The results for DF events are similar.

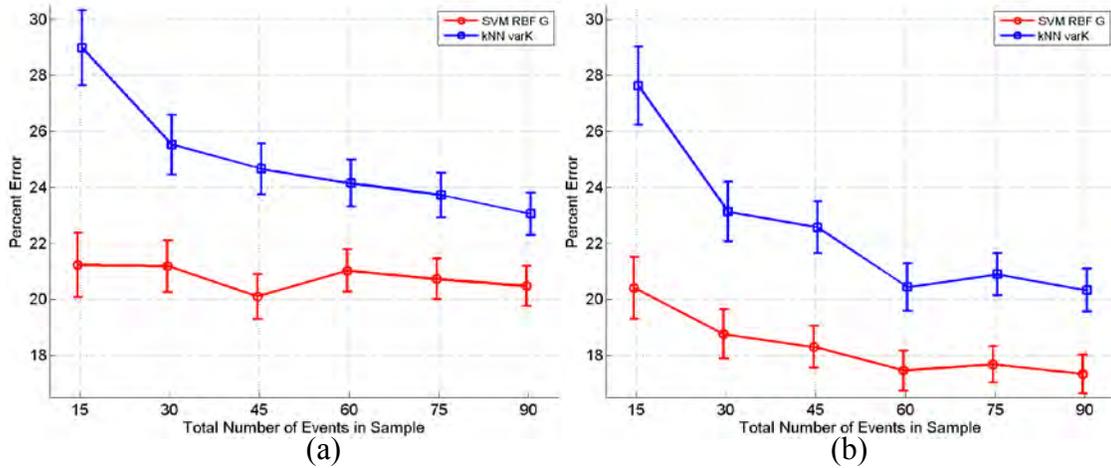
### **Dynamic Geographical Constraints**

Up to this point, fixed geographic and temporal windows have been used to constrain the data used for training and testing the classifiers. Small windows result in inefficient events to support the classifier's training while large windows will lose the benefit of the locality. Rather than fix the window, we choose to fix the number of events and dynamically determine the minimal feasible geographic constraint. In the following figures, the size of the event class is constrained to a set of fixed values:  $n = [15\ 30\ 45\ 60\ 75\ 90]$ . These number are selected based on the performance of two best performing algorithms at this point:

- SVM RBF kernel using PCA with hyperparameters estimated by grid search; and
- kNN using PCA with k estimated dynamically.



**Figure 31: The impact of dynamically varying the box constraint  $C$  and the RBF shape parameter  $\sigma$  on the classification of IED events using SVM with an RBF kernel; (a) Classification accuracy under combined geographic and temporal constraints using varying  $C$  and  $\sigma$ ; (b) Mean order of  $C$  at various window sizes; (c) Mean order of  $\sigma$  at various window sizes.**



**Figure 32: Classification error using fixed sample sizes and dynamic geographic constraints; (a) IED classification error; (b) DF classification error.**

Figure 32 examines the impact of using a constant sample size to constrain the geographic window. In the experiment, the training sets were produced by first temporally constraining the data to a fixed window of 120 days before each *ST* event. From this temporally constrained set, the 15-90 events geographically nearest to the center (*ST* location) were selected. The distance of the most distant event from the center (*ST* location) was used as the radius of the geographic window and non-event points ( $RD_{250}$ ) were selected from within this window. The test set was constructed by temporally constraining events to a fixed window of 60 days after each *ST* event and geographically constraining events to the same radius used to select training data. Non-events points for the test were also selected using the same geographic constraint as the training set. Note that no cross-validation was used. Figure 32(a) shows that the classification error rate using IED events approaches 20 % when using SVM, regardless of sample size. DF events produce even lower error rates in Figure 32(b). The consistent

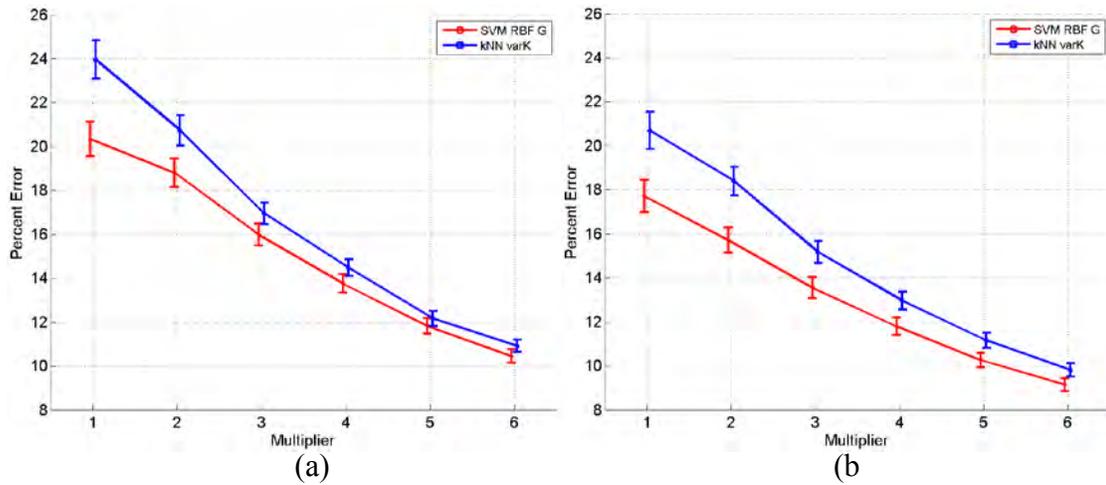
performance of SVM at small samples sizes using a dynamic geographic constraint is an interesting outcome. The results shown here are generally better and definitely more consistent than those seen using fixed windows.

### **Summary of Parameter and Hyperparameter Estimation**

Parameter and hyperparameter selection are an effective way to improve effects of feature reduction and classification accuracy for both SVM and kNN. In the case of SVM, an RBF kernel with hyperparameters selected by grid search produced classification errors on the order to 20%. The lowest error occurred at the smallest windows and the smallest sample sizes, indicating that this combination may be the best overall in cases where the conflict event data is sparse. kNN also showed improvement by dynamically estimating  $k$ . Classification error of approximately 25% was consistent across all window sizes using this method.

### **The Impact of Unbalanced Classes**

So far, all of the experiments shown in the research have used balanced classes. In other words, the number of events and the number of non-events were equal in the training and test sets. However, when large geographic and temporal constraints are used, the number of non-event points tends to outnumber events by a factor of 10,000 or more. As a result, when balanced classes are used, the number of non-events points selected may be very small compared with the total number of non-event points found in the window. In comparison, all conflict events found in the window end up in either the training or set sets. In this case, it can be argued that the selection of balanced classes



**Figure 33: The impact of unbalanced classes; (a) IED classification error; (b) DF classification error;**

causes the event class to be over-represented in the training and test sets. To compensate, it is possible to intentionally create unbalanced training and test sets. However, this violates an often un-noticed assumption for many classifiers that the classes are equally represented. The resulting problems, often described as between-class imbalance, tends to grow worse as the imbalance gets more pronounced, potentially impacting the way in which  $k$  is chosen for kNN or requiring sophisticated sampling methods to compensate.

Figure 33 examines the impact of unbalanced classes on predictive analysis using SVM and kNN. In the experiment, for each point in  $ST$ , the nearest 60 events are chosen for the training sample as described in the previous sub-section. Six different sets of non-event points are chosen ranging in size from 60 to 360, in 60-point increments. The resulting event:non-event ratios go from 1:1 to 1:6. Six samples are produced from each entry in  $ST$  and used for training and testing with SVM and kNN. Note that no cross-validation was used. In Figure 33(a) and Figure 33(b), accuracy apparently improves,

with classification error trending as low as 10% for both IED and DF events. However, this low error rate hides a serious problem: the misclassification rate of events increases greatly as the class imbalance grows.

### **Contribution of Each Feature**

For a given tactic or type of attack, some features are more relevant and tend to be selected more often by stepwise feature selection. Figure 34 shows the selection rate of features used in the experiment that generated the data in Figure 28. Of the 77 features, only 13 are selected more than ten percent of the time. An additional 20 features are selected between 5-10% of the time. The two most frequently selected features are the elevation (1) and proximity to a human population center of any size (58).

Two observations emerge. First, the features contributing to the classification of IED and DF events are almost always selected at roughly the same rate. This may mean that both types of events can be treated as a single class. Also, the 33 features most frequently selected by STP fall into a few categories. Fourteen of the features are related to CEA (denoted as 'cea' in Figure 34) and a visibility metric related to CEA. Five of the features are related to the distance from human population centers. Radial-based shape complexity and radial-based rugosity both appear at four different resolutions.

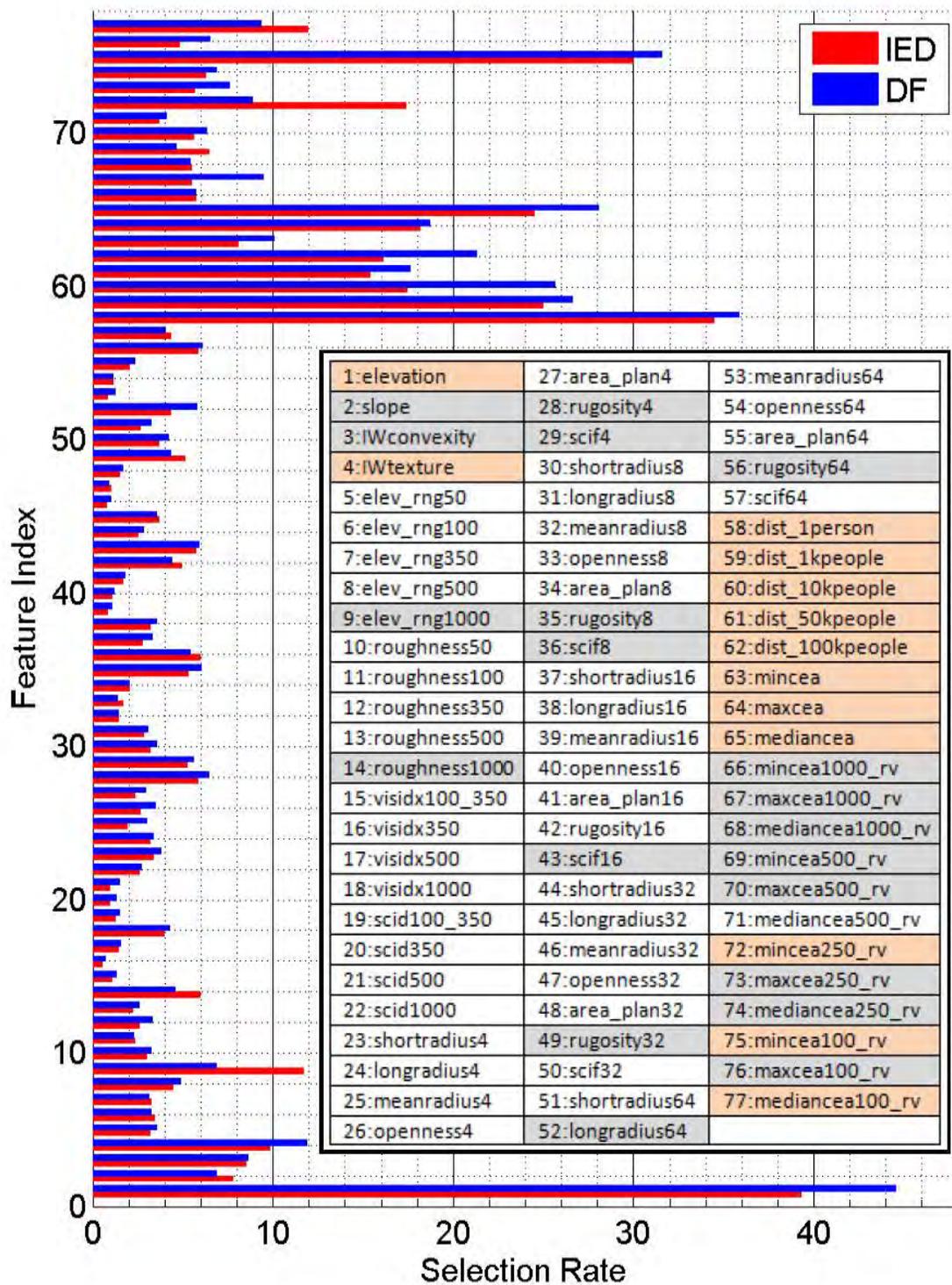


Figure 34: Features selected using stepwise selection for dimensionality reduction. The inset lists the features by name and highlights the largest contributors: light gray boxes are selected in 5-10% of all classification attempts, light peach boxes are selected in > 10% of all classification attempts.

## Human and Machine Expert Feature Selection

Recent conflicts in Iraq and Afghanistan have produced a generation of military personnel trained to detect and react to attacks. Many of these soldiers, sailors and Marines have convoyed and patrolled extensively in Afghanistan and developed a “sixth sense” or intuition about attack sites. Two of these experts, an Army Ranger, GP, and two Army Special Forces soldiers, TN and SG, were asked to identify features and distances that described a likely conflict event site. In their responses, the experts emphasized the importance of the field of view for the attackers. TN focused on the importance of communications (no current feature captures this information), cover, concealment and escape adjacency for attackers. He also discussed the importance of terrain that restricts target movement, (possibly captured when a very short radial is found). GP and SG focused on the characteristics of terrain required by attackers to support different types of attack. Their responses were used to select features from the existing set of 77.

- TN: Short, long and median radial, local openness, planimetric area, rugosity, sparse viewshed shape complexity, maximum and median cumulative escape adjacency, maximum and median route visibility at 100 meters
- GP: Short, long and median radial, local openness, planimetric area, rugosity, sparse viewshed shape complexity, elevation, slope convexity, texture, roughness at 350 m., discrete shape complexity at 1000 m., distance to populated area with at least one person

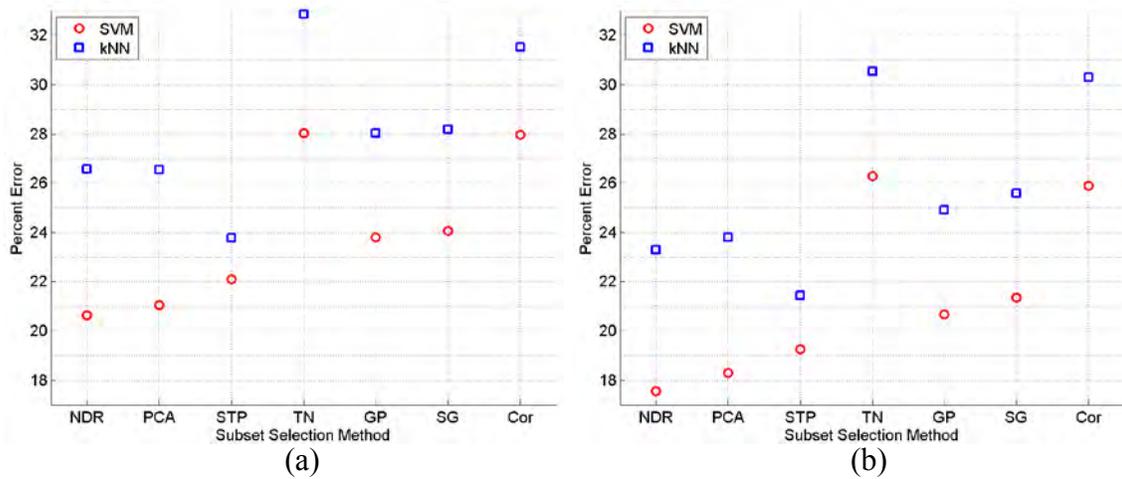
- SG: Short, long and median radial, local openness, planimetric area, rugosity, sparse viewshed shape complexity, slope convexity, texture, elevation range at 350 meters, distance to populated area with at least 1000 people, max and median route visibility at 100 m., max cumulative escape adjacency at 250 meters

Next, an automated subset selection method or blind expert uses feature correlation as a discriminator. Pearson's correlation coefficient was used to calculate the dependence between event features and non-event features,  $\rho_{X,Y} = \frac{\text{covariance}(X,Y)}{\sigma_X \sigma_Y}$

, where X and Y are an event feature and a non-event feature, respectively, and  $\sigma$  is the standard deviation of the feature. Then a correlation matrix,  $M$ , is built for the 77 features used in this research,  $M_{i,j} = \rho_{event_i, non-event_j}$  for  $i, j = 1, \dots, 77$ . The sum of the absolute value of each column of  $M$  is inspected. The ten features with the lowest cumulative correlations are selected.

The experiment to test these different subset selection methods used fixed sets of 60 events and 60 non-events chosen using

- fixed temporal training and test windows of 120 and 60 days, respectively;
- dynamic geographic windows, with a radii equal to the distance from the most distant event to the  $ST$  location.
- Feature selection: three experts (GP, TN, SG), the blind expert (COR), and the more conventional feature selection approaches of NDR, STP and PCA.
- No cross-validation



**Figure 35: Alternate methods for subset selection; (a) IED classification error; (b) DF classification error.**

**Table 4: Event classification error, from experts (% of total sample).**

	IED (%).		DF (%).	
	SVM	kNN	SVM	kNN
NDR	9.7	10.1	8.9	9.5
PCA	9.7	10.0	9.1	9.5
STP	10.2	11.4	9.1	10.1
TN	14.3	16.3	14.4	15.8
GP	10.3	10.8	9.7	10.5
SG	10.0	11.5	9.2	10.3
Cor	12.8	14.6	13.3	14.2

Figure 35 shows the results of the experiment where experts select the best features. Across the board, SVM produces lower overall classification error rates for both IED and DF.

Table 4 shows the actual event classification error rate, or the percent of the total sample that consists of misclassified events. As shown in the table, the best result for IED classification is SVM NDR and PCA where 9.7% of the total sample consists of misclassified events.

## **Ensemble of Classifiers**

The best individual classifiers used in this research produce classification errors near 20%. Both SVM and kNN perform well although SVM does better with small training sets. For both SVM and kNN there seems to be a tradeoff: decreased overall classification error comes at the cost of increased event classification error. Since a misclassified event site has a large real-world cost—an IED explodes or an ambush is not anticipated—we prefer to find classifiers that reduce misclassified events while keeping total classification error as small as possible. Ensemble-based classifiers offer a potential way to achieve this goal. Ensemble-based classifiers make classification decisions by combining the output of multiple individual classifiers according to some rule or algorithm.

In this report, we examine three cases: majority vote rule, cost-sensitive rule, and stacking using SVM and kNN algorithms. The majority vote rule assigns a class label by simply counting the number of ‘votes’ for each class from individual classifiers. Ensemble-based classifier  $C$  is composed of the output of multiple individual classifiers.

Let  $c_i$  be the output of an individual classifier containing  $n$  classifications and let  $c_{ij}$  be its  $j$ -th output. Then, the classification output of ensemble-based classifier  $C$  is assigned by a majority vote of its individual classifiers,

$$C_{i,j} = \begin{cases} event & | \sum_i (c_{i,j} \equiv event) \geq \sum_i (c_{i,j} \equiv nonev) \\ nonev & | \sum_i (c_{i,j} \equiv nonev) > \sum_i (c_{i,j} \equiv event) \end{cases} \text{ for } j = 1, \dots, n \quad (38)$$

The cost-sensitive rule takes into account the real-world cost of a misclassification and prefers to misclassify non-events. If any member of the ensemble classifies a location as an event, then the ensemble classifies it as an event,

$$C_{i,j} = \begin{cases} event & | \sum_i (c_{i,j} \equiv event) > 0 \\ nonev & | \sum_i (c_{i,j} \equiv event) \equiv 0 \end{cases} \text{ for } j = 1, \dots, n \quad (39)$$

Note that when the ensemble consists of two classifiers, both of these rules produce the same result.

Voting algorithms simply combine the output of existing classification outcomes according to some rule. Stacking, on the other hand, takes the same combinations of classification outcomes and uses them as input to a classification algorithm. Instead of simply counting outcomes, stacking combines the output of multiple base classifiers to form a new dataset. This new dataset, composed of binary classification algorithm results, is used to train a new classifier. In this research, several different base classifier combinations, or stacks, are used with SVM and kNN.

## Ensembles Constructed of Single Algorithm Base Classifiers

Figure examines the classification error using SVM and kNN of three ensembles using two different rules. In each ensemble, all classifiers use a common, single algorithm, either SVM or kNN. The outcome of the ensembles is compared to several individual classifiers, including:

- NDR, PCA and STP;
- Emp: a classifier that uses the Emplacement features from Appendix Table 7;
- MC: a classifier that uses the Monitor/Control features from Appendix Table 8.

The ensembles include

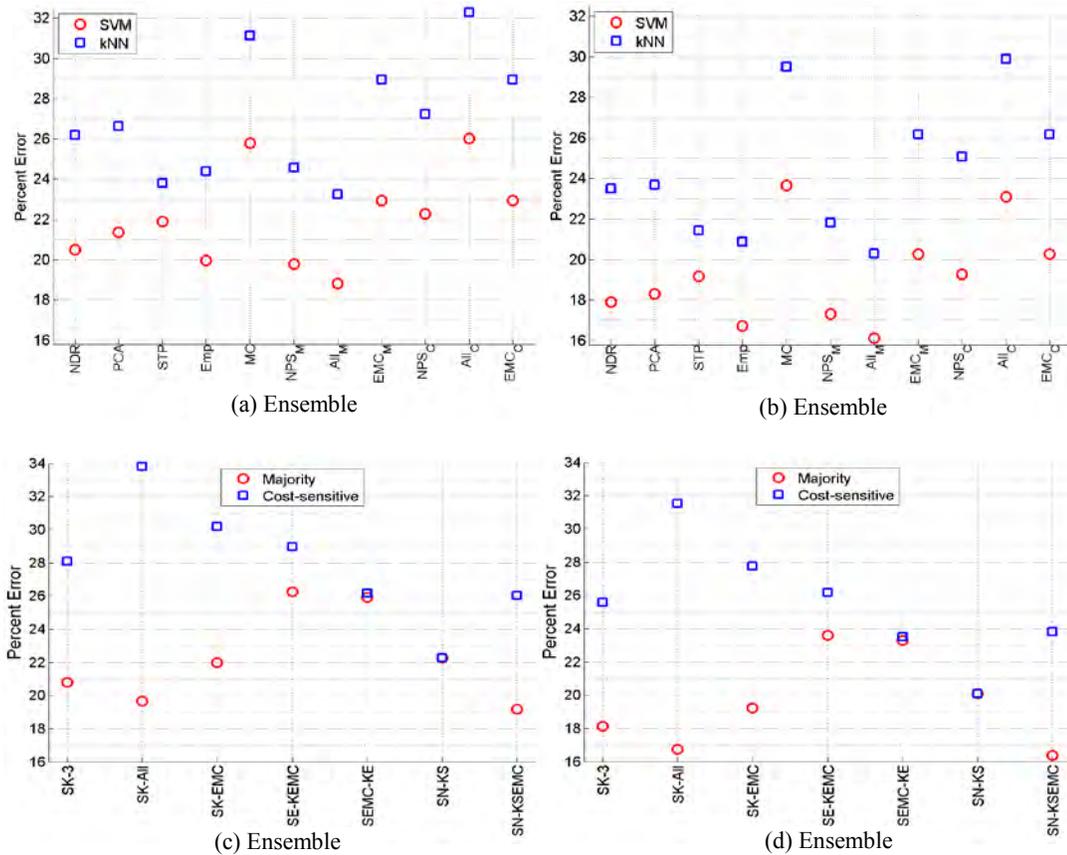
- NPS: the classification outcomes of {NDR, PCA, STP};
- All: the classification outcomes of {NDR, PCA, STP, Emp and MC}; and
- EMC: the classification outcomes of {Emp and MC}.

## Ensembles Constructed of Varied Algorithm Base Classifiers

The experiment settings above only addressed ensembles composed of outcomes using the same base classifier, kNN or SVM. However, ensembles composed of varied algorithm base classifiers offer some improvement by increasing diversity. We also propose following varied algorithm ensembles:

- SK3: [ SVM-NDR, SVM-PCA, SVM-STP, kNN-NDR, kNN-PCA kNN-STP];
- SK-All: [SVM-NDR, SVM-PCA, SVM-STP, SVM-Emp, SVM-MC, kNN-NDR, kNN-PCA, kNN-STP, kNN-Emp, kNN-MC];
- SK-EMC: [SVM-Emp, SVM-MC, kNN-Emp, kNN-MC];
- SE-KEMC: [SVM-Emp, kNN-Emp, kNN-MC];

- SEMC-KE: [SVM-Emp, SVM-MC, kNN-Emp];
- SN-KS: [SVM-NDR, kNN-STP];
- SN-KSEMC: [SVM-NDR, kNN-STP, kNN-Emp, kNN-MC].



**Figure 36: The classification accuracy of classifier ensembles; (a,c) IED classification error; (b,d) DF classification error.**

Figure 36 examines the classification accuracy of the ensembles for vote based methods. For single type ensemble, All<sub>M</sub> performs slightly better than others in single type classifier ensembles, but the difference is not pronounced. All of the ensembles using the cost-sensitive rule show a significant decrease in event classification error.

Note that  $EMC_M$  and  $EMC_C$  show the same performance because the two rules, majority and cost-sensitive, produce the same results when there are only two ensemble members. For the mixed-type classifier ensembles, the overall error rates show a slight improvement in some cases.

Table 5 shows the actual event classification error rate. Combining classifiers can significantly reduce misclassified events. Although the best result for IED classification is  $All_C$  at 2.1%, a close second is the MECH model-based EMC at 3.8%. For the mixed-algorithm part, of particular note is the performance of the mixed classifiers SE-KEMC and SEMC-KE. Both of these MECH model-based classifiers perform well using both majority and cost-sensitive rules. This performance supports the MECH model concept of dividing the analysis into two spaces: features collected from the conflict event Emplacement site and features collected from the conflict event Monitor/control area surrounding the Emplacement site.

**Table 5: Event classification error for vote based ensemble of classifiers (% of total sample).**

	IED (%)		DF (%)	
	SVM	kNN	SVM	kNN
NDR	9.6	9.5	9.2	9.2
PCA	9.9	9.5	9.1	9.3
STP	10.1	11.1	9.1	10.1
Emp	9.0	9.1	8.5	8.2
MC	13.0	14.3	12.9	14.4
$NPS_M$	9.1	8.8	8.7	8.6
$All_M$	8.6	8.4	8.4	8.3
$EMC_M$	3.8	3.7	4.5	3.9

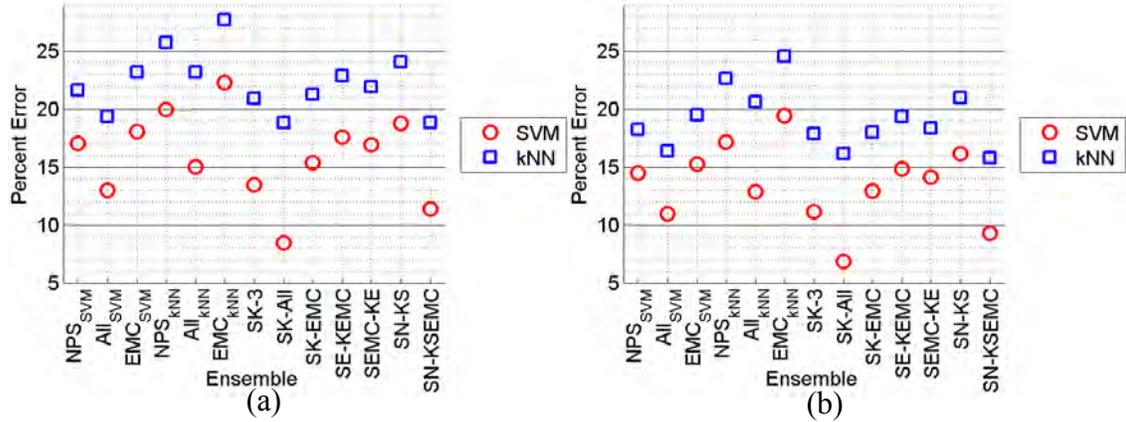
NPS <sub>C</sub>	4.6	3.7	4.6	4.1
All <sub>C</sub>	2.1	1.6	2.5	1.9
EMC <sub>C</sub>	3.8	3.7	4.5	3.9
SK-3	6.8	2.4	6.9	2.7
SK-All	6.9	0.8	6.7	1.1
SK-EMC	4.8	1.7	5.2	2.1
SE-KEMC	3.7	2.5	4.0	2.7
SEMC-KE	3.6	2.4	3.9	2.9
SN-KS	5.3	5.3	5.2	5.2
SN-KSEMC	6.0	2.1	6.1	2.5

## Stacking

The final experiment in this category examines the impact of stacking on classification error. Stacking, or stacked generalization, uses an ensemble to train a learning algorithm. The ensemble is created by combining the classification decisions of two or more individual classifiers into a new dataset. In this dataset, each feature is the outcome of an individual classifier that was trained on original data. This new, composite dataset is used to train a new classifier, which then makes a final classification decision or prediction. In this experiment, various combinations of base classifiers are used to create 13 different stacks. Each different stack is used to train and test SVM and kNN classifiers.

Figure 37 examines the impact of stacking. In the figure, the 13 ensembles are listed on the x-axis. The first six ensembles use subscripts to describe the algorithm used by the original base classifiers. So, the NPS<sub>SVM</sub> ensemble is composed of three classification outcomes produced by using SVM on original data with dimensionality reduction schemes of NDR, PCA, and STP. In the figure, the overall classification error is reduced across all ensembles, especially for outcomes predicted using SVM. The

improvement is particularly pronounced for SK-ALL using SVM, which has an overall error rate for IEDs of around 8.5 % and an event classification rate of 4.1%.



**Figure 37: The classification accuracy of stacking; (a) IED classification error; (b) DF classification error;**

**Table 6: Event classification error, from stacking using SVM and kNN (% of total sample).**

	IED (%)		DF (%)	
	SVM	kNN	SVM	kNN
NPS-SVM	8.1	7.0	7.5	7.1
All-SVM	5.9	6.5	5.8	6.6
EMC-SVM	8.7	4.7	8.5	5.5
NPS-kNN	9.0	6.8	8.4	6.7
All-kNN	6.6	6.0	6.2	6.2
EMC-kNN	9.9	5.4	9.1	5.5
SK-3	5.9	6.4	5.6	6.4
SK-All	4.1	6.3	3.8	6.5
SK-EMC	7.2	5.3	6.8	5.6
SE-KEMC	8.0	5.5	7.5	5.8
SEMC-KE	8.0	5.5	7.6	6.0
SN-KS	8.7	5.4	8.1	5.3
SN-KSEMC	5.2	5.9	4.9	6.0

Table 6 shows that stacking does not perform as well as single algorithm and mixed algorithm base classifiers when event classification error is the primary concern. However, SK-All<sub>SVM</sub> performs better than any other classification scheme when total error or event error are used to make the decision. The best performance comes with a relatively high computational cost: a dataset consisting of 60 events and 60 non-events is classified five times with SVM and five times with kNN. The ten sets of classification outcomes are used to create a new feature set that is then classified using SVM. The decision to use this relatively heavy-weight process will depend on the computational resources and time available to the user.

A common challenge of the training algorithms discussed so far is lack of intuitive casualty interpretation. A kNN may produce clusters that can be visually reasoned, but it is often difficult to comprehend the underlying structure in the data that led to formation of the cluster. Discriminant analysis and SVM face similar challenges. To support human in the loop pattern mining, the decision tree learning algorithm can be considered. In this reasoning process, a set of location data (with labelled classes and locations' features) is being mined to produce a list of most relevant features and their numerical thresholds in classification of event vs. non-event locations. The decision tree can then be used to create prediction rules for the area being analyzed.

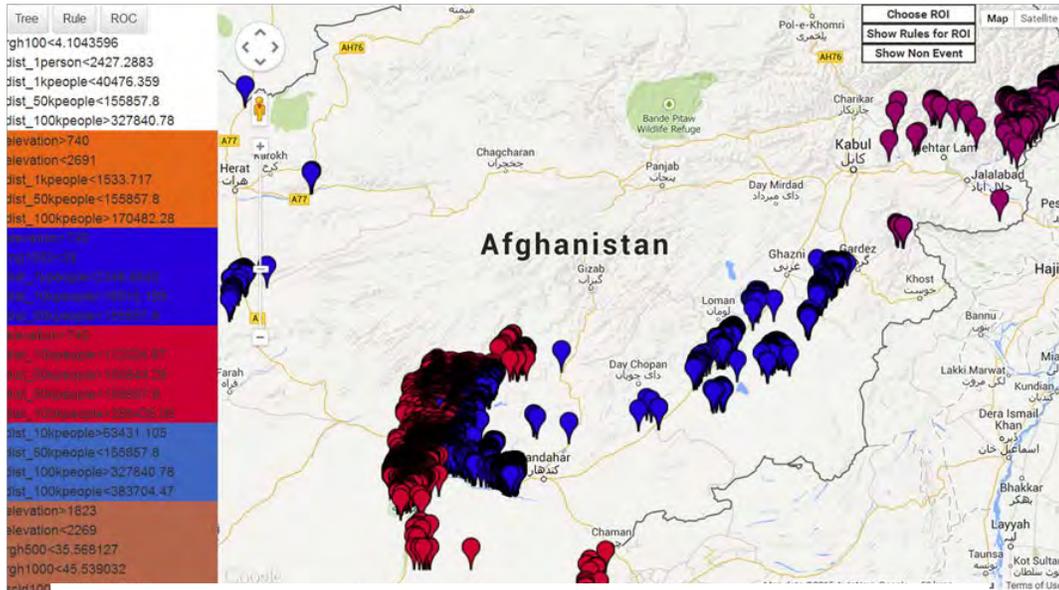
### **Decision Tree (DT) Learning Algorithm**

The Decision Tree (DT) learning algorithm iteratively searches the feature  $f$  and its corresponding threshold  $T$  that splits a data set  $D$  into 2 subsets  $\rightarrow \{D_T, D \setminus D_T\}$ , with the largest *information gain*, which is defined as the difference between the entropies of

the dataset before and after the separation [19], where  $D_T$  represents the subset of data with value of feature  $f$  larger than  $T$ , and  $D \setminus D_T$  otherwise. The two-tuple  $(f, T)$  represents a decision node  $N$ . The splitting process is applied to both  $D_T$  and  $D \setminus D_T$  iteratively to generate children nodes of  $N$ , based on different features and corresponding thresholds until a predetermined termination condition is satisfied. Each leaf node is assigned a *class label* based on the majority of class labels of data in the node. By traversing from a leaf node upward to the root, we can produce a *prediction rule* to assert which of the two classes that an unknown sample belongs to.

The 77 features of the Afghanistan attack data set are trained by the open source DT package C5.0 to produce a decision tree. In a standalone prototype, two different views of the tree, the *tree* view, and *rule* view can be visualized with data points associated with the displayed view. The two different views of the DT for the entire Afghanistan are illustrated in Figure 35(a) and (b), which display screen is limited to the area west of Kandahar.

The left hand side of Figure 35(a) is a long list of rules that has been automatically generated by the DT algorithm. Each rule is assigned a unique color, and the locations of events that are associated with each clicked rule are displayed in the rule color on the map at the right hand side.



(a) Rule View for the whole Afghanistan



(b) Decision Tree view for the rule  $DR_{may}$ , events satisfying rule  $DR_{may}$

**Figure 38: A DT based user interface for human in the loop pattern classification.**

For the tree view shown in Figure 38(b), which is partially obscured due to the size of the feature set, the optimal prediction rule  $DR_{may}$ ★ corresponding to the leaf node marked with a red star in the left column of Figure 38(b), which marks a cluster of attacks located 15 km north of Maymana, capital of Faryab province. When the features of the nodes and their threshold values to reach the leaf node are tabulated in Table 7, the analyst can translate the thresholded features into the qualitative tactical interpretation about the class label. The rank of the conditions is equal to the depth of the corresponding decision node on the decision tree. In this example, the most dominant nodes in the decision tree are the population features that describe the distance to the nearest city with a population of a certain size.

**Table 7: Conditions and tactical interpretation of decision rule  $DR_{may}$**

Rank on the tree	Rule condition (Quantitative)	Tactical interpretation (qualitative)
1	dist_50kpeople > 155857.8	Away from larger cities
2	elevation <= 2094	Relatively low elevation grounds
3	dist_50kpeople >= 254555.8	Not too far from larger cities
4	Rng1000 <= 92	Maximum elevation change within 1 km < 92 meters
5	dist_1person <= 1892.959	Less than 2 km from nearest populated area
6	dist_10kpeople <= 70686.7	Less than 70 km from nearest small town
7	rng500 > 52	Maximum elevation change within 500 m > 52 meters

At the upper right corner of the map in Figure 38(b) has a *region of interest* (ROI) button. By clicking on this button, the user can point and drag an ROI area on the map. In this example, the ROI is a region north of Maymana (highlighted on google map in Figure 38(b)). Then, for the selected area, the user can click again the “Select Rule from ROI” to visualize the rule associated with it. As a result, we get a major rule marked with a red star on the decision tree in left hand side of Figure 39. Condition 'longrad16 <= 494.2' means that the attacker tends to select areas where the longest sight line from a potential Emplacement is less than 500 meters. Conditions 'rgh100 <= 11.374' and 'rgh500 > 6.0728' mean the roughness of the terrain surrounding a potential Emplacement is relatively high but not extreme. Roughness is an indicator of texture. A value of zero indicates a perfectly flat surface and the increasing value indicate terrain that is increasing uneven, rough or rugged.

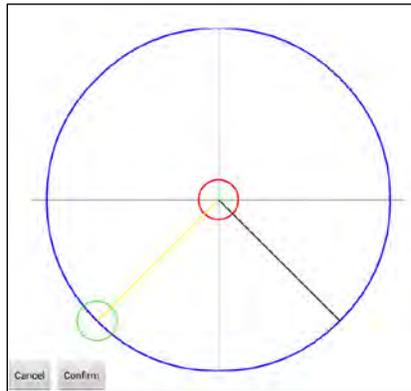
The red markers representing events satisfying the rule in the selected ROI will show up after clicking on the leaf node marked by a red star. By clicking the 'Show Non Event' in the panel on top right corner, non-attack locations will also be displayed on the map. The purple marks with stars are non-event locations that should be considered as high risk because they satisfy all rules in the rule set learned from historical data. The green locations have a lower risk of being attacked, at least using tactics similar to those in the past event set.



**Figure 39: Refined analysis by constraining ROI**

## 5. MECH BASED SITUATIONAL SIMULATION

MECH based situational simulation aims to mimic the mental process of an actor in planning of an attack. As shown in Figure 40, the Halo model characterizes the risk averse behavior model by a few parameters, each of which can be adjusted by the user to reflect individual differences.



**Figure 40: Halo Parameters for MECH Models**

Legend

Blue circle: sight range

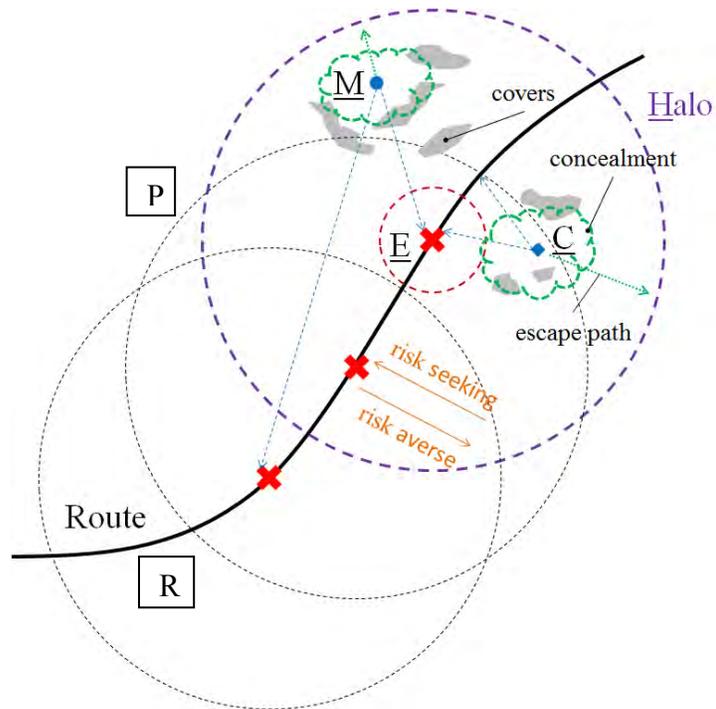
Red circle: blast range

Green circle: reachable range for cover

Black line: return fire range

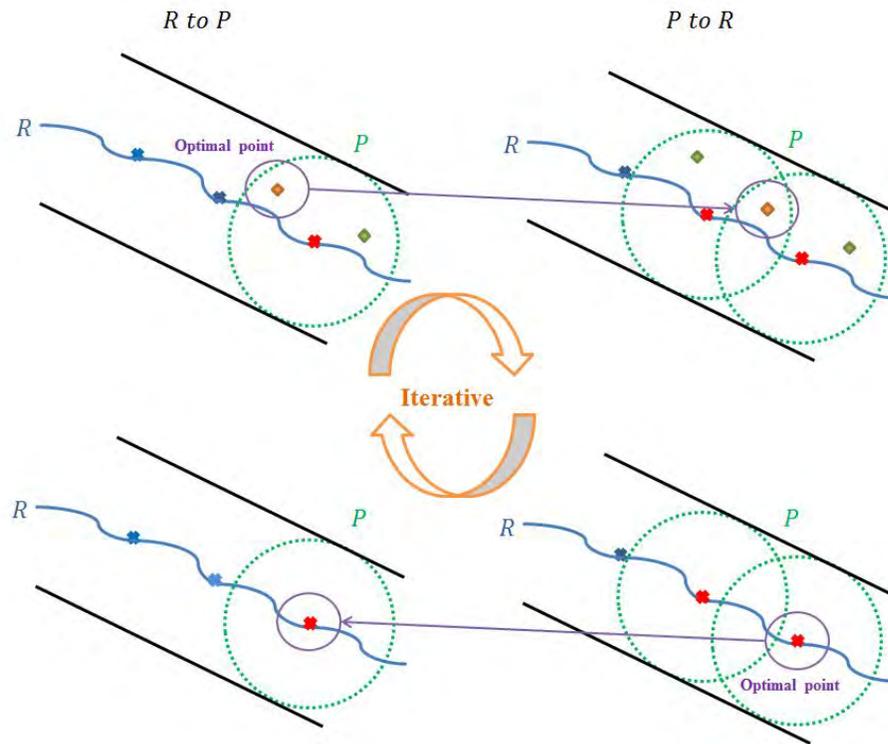
Yellow line: effective device trigger range, or DF range to target

The HALO model characterizes the distance constraints between the M, E and C functions, with the E location placed at the center. The line of sight (LOS) or no line of sight (NLOS) between the actor and target can be derived from DEM. As shown in Figure 41, by applying the analysis to every point on a route R and its surrounding area P, and sum up the results from all points, we can generate simulated vintage points, which can then be represented in the heat map format.



**Figure 41: The MECH based risk averse behavior model in tactical planning [31]**

As shown in Figure 42, the simulation process can start with selection of high value E positions on R, which have high exposure to all P points. Then, the user can examine top portion of P points that have the highest observability to these R points [28]. Similarly, one can start with locating P points that have high observability to all R points. After eliminating P points with low observability, the remaining P points (with highest observability) can be used to produce R points that have the highest exposure as the vintage E positions. Different levels details on tactical actions (monitoring, control, hiding, firing, etc.) can be incorporated in the simulation process.

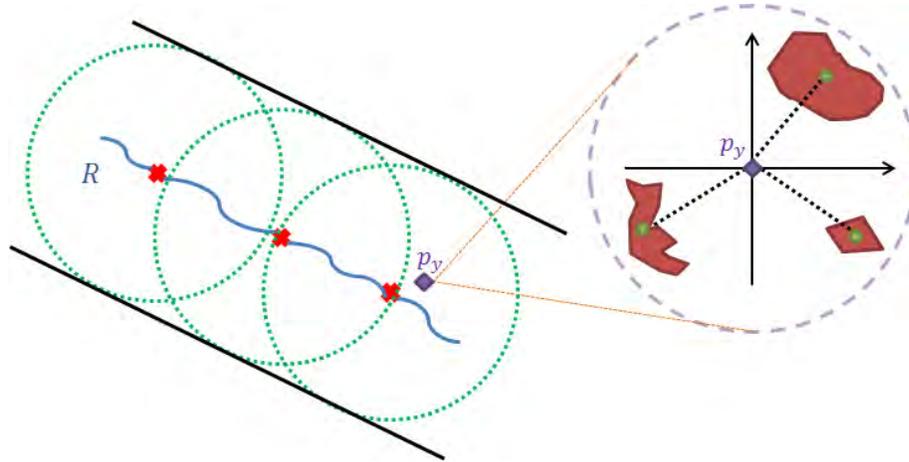


**Figure 42: The interactive threat assessment process of route points (R) and the proximity points (P)**

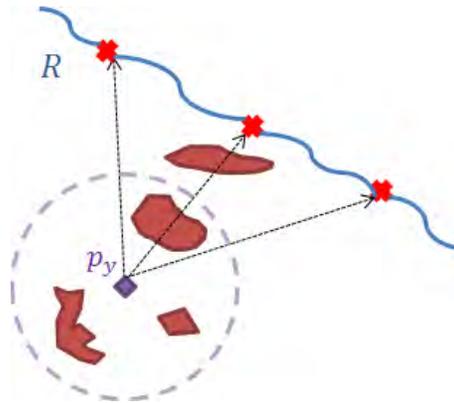
Decision options in planning of M, E, and C locations is formulated into an optimization problem subject to a set of environmental and behavioral constraints defined by the MECH model.

The tactical value of a location for the M, E, or C action is based on a compound assessment of the action effectiveness against the target and environmental protection for the actor. Within the weapon's effective range, a well-covered position with good visibility to the target is a good Control position. But when one or both of the two factors are less than ideal, individual actors may make very different choices based on their own reasons [20], [21]. An aggressive actor may weight a lot more on the ability to

execute an attack, while other actors may weight more on the protection power of locations.



(a)



(b)

**Figure 43: (a) The environ concealment, and (b) the cover from a target.**

To support a broad spectrum of users, a general reward-risk tradeoff function is designed for the simulator. The composite optimization function  $f(U_O, U_D)$  is based on the weighted sum (+) or product ( $\times$ ) of the offense utility  $U_O$  and defense utility  $U_D$ .

$f(U_O, U_D)$  can be represented as  $f(U_O, U_D) = (\omega_O \cdot U_O \Delta \omega_D \cdot U_D)$ , where  $\Delta$  represents a sum/multiplication or other operations, and  $\omega_O$  and  $\omega_D$  weights for offense and defense utilities, respectively, with  $\omega_O + \omega_D = 1$ . A location cannot be considered for action if either  $U_O$  or  $U_D$  is below thresholds  $\tau_O$  or  $\tau_D$ , respectively. This constraints can be expressed as a switch function  $s_O = \begin{cases} 1, & \text{if } U_O > \tau_O \\ 0, & \text{otherwise} \end{cases}$ , and  $s_D = \begin{cases} 1, & \text{if } U_D > \tau_D \\ 0, & \text{otherwise} \end{cases}$ . A general utility function can be expressed as  $f(U_O, U_D) = s_O \cdot s_D (\omega_O \cdot U_O \Delta \omega_D \cdot U_D)$ .

The utility function can then be measured based on various tactical activities, e.g., aiming, observability, monitoring, concealment, cover, etc. These physical measurements are normalized from their very different dynamic ranges to a scale of 0-100 before they are used in the optimization process.

Despite the simplicity of the MECH model, the simulation model proved to offer highly consistent, complementary results on the M/E/C locations with respect to the E locations produced by the statistical pattern analysis approach. The use cases to illustrate this point is given in the Project Outcomes of the Executive Summary. Details on how to set parameters for the simulation are described in the user guide of the MECH software prototype.

## 6. CONCLUSION

This investigation, which emerged from curiosity about the nature of seemingly random locations chosen for asymmetric conflict events, demonstrates the effectiveness of statistical pattern modeling of human behaviors under geographic constraints.

Although purely geomorphometry-based models produce fairly weak indicators for design of prediction algorithms, features that capture the conflicting motivations of aggression and risk aversion provide a strong signal of potential attacker intent. Unlike most existing behavior models where key features are of qualitative nature, the MECH model successfully fused geomorphometry and human behaviors into a single quantitative model based on intervisibility and distance constraint functions. It transforms and captures geographic features, human behaviors and logistic needs into risk factors. The results is a situational awareness solution that predicts the likelihood and utility of locations in future attacks and identifies locations for associated staging operations.

Feature selection is the most important step of the whole process. It requires extensive analysis of the system dynamics based on doctrine, past experience, literature, as well as heuristic steps to balance performance goals with computing costs. The set of features presented in this report is optimized for our available data set and, as expected, the contributions of individual features to the model were found to be quite different. Interestingly, computer-based analysis independently confirms the intuition of three human experts with regards to useful features. An alternative view is that, in a limited way, the algorithms are able to detect subtle patterns previously only intuited by battle-

trained, experienced soldiers. Given that it is practically impossible to acquire reliable data about monitoring and control locations, analytic results are focused on historical and potential emplacement locations and their environs. The MECH model provides useful simulation-based assessments in the form of heat maps and overlays that can be understood by technology novice users.

The overall effort demonstrates the consistency of human behaviors and the viability of algorithm-based modeling of such behaviors in the development of next-generation situational awareness analytics tools. That being said, given the probabilistic nature of such modeling, correct understanding of the semantics of the analysis outputs is important, especially for users. Blind adherence to statistical models may ignore the impact of adaptive adversaries, hasty attacks, and other battlefield influences not captured in the current feature set. Adaptive, automated and continuous assessment of the adversary and their risk tolerance remains a critical issue in the formulation and design of computer-based tools for such purposes.

As a final note, we emphasize the strength of the model and outputs presented here. The past event dataset is limited to 19 months of data containing [latitude, longitude, date, time, and event class]. Geographic data comes from publically available sources at a resolution of approximately 30 meters, a resolution potentially large enough to conceal tactically-significant terrain features. Even with these suboptimal and limited data sources, the results offer unusual and potentially unique insight into adversary tactics and risk tolerance. More research will be required to minimize or eliminate hidden bias in the experimental data. Higher resolution digital elevation maps (on the

order of 5-meter intervals) and more detailed event information (perhaps including IED subclasses like VBIED, PBIED, etc. and trigger mechanisms like RF, command wire, etc.) would probably provide increased overlay and more refined tactical constraints.

## References

1. Army Field Manual: Tactics (FM 3-90). Headquarters, US Army Training and Doctrine Command, 2001.
2. Apolloni, Andrea, et al. "A study of information diffusion over a realistic social network model." Computational Science and Engineering, 2009. CSE'09. International Conference on. Vol. 4. IEEE, 2009.
3. Norwitz, Jeffrey H. Armed Groups: Studies in National Security, Counterterrorism, and Counterinsurgency. Government Printing Office, 2008.
4. Reuter, H. I., and A. Nelson. "Geomorphometry: concepts, software, applications." Developments in soil science series. Elsevier, SBN-13 (2008): 978-0.
5. NASA Land Processes Distributed Active Archive Center, "ASTER GDEM Data Download Site." [Online]. Available: [https://lpdaac.usgs.gov/get\\_data](https://lpdaac.usgs.gov/get_data). [Accessed: 12-Nov-2013].
6. MacMillan, R. A., and P. A. Shary. "Landforms and landform elements in geomorphometry." Developments in soil science 33 (2009): 227-254.
7. Moran, Christopher J., and Elisabeth N. Bui. "Spatial data mining for enhanced soil map modelling." International Journal of Geographical Information Science 16.6 (2002): 533-549.
8. Guth, Peter L. "Quantifying terrain fabric in digital elevation models." Reviews in Engineering Geology 14 (2001): 13-26.
9. Izraelevitz, David. "A fast algorithm for approximate viewshed computation." Photogrammetric Engineering & Remote Sensing 69.7 (2003): 767-774.
10. Kreveld, Marc Van. "Variations on sweep algorithms: Efficient computation of extended viewsheds and classifications." In Proc. 7th Int. Symp. on Spatial Data Handling. 1996.
11. Shen, Y. I. N. G., et al. "Viewshed computation based on LOS scanning." Computer Science and Software Engineering, 2008 International Conference on. Vol. 2. IEEE, 2008.

12. Gladwell, Malcolm. *Blink: The power of thinking without thinking*. Back Bay Books, 2007.
13. Iwahashi, Junko, and Richard J. Pike. "Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature." *Geomorphology* 86.3 (2007): 409-440.
14. Horn, Berthold KP. "Hill shading and the reflectance map." *Proceedings of the IEEE* 69.1 (1981): 14-47.
15. Jolliffe, Ian. *Principal component analysis*. John Wiley & Sons, Ltd, 2002.
16. Hastie, Trevor, et al. "The elements of statistical learning: data mining, inference and prediction." *The Mathematical Intelligencer* 27.2 (2005): 83-85.
17. Ghosh, Anil K. "On nearest neighbor classification using adaptive choice of k." *Journal of computational and graphical statistics* 16.2 (2007): 482-502.
18. Liu, Zhiliang, Ming J. Zuo, and Hongbing Xu. "Parameter selection for Gaussian radial basis function in support vector machine classification." *Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2012 International Conference on*. IEEE, 2012.
19. Quinlan, J. Ross. "Induction of decision trees." *Machine learning* 1.1 (1986): 81-106.
20. Kahneman, Daniel, and Amos Tversky. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the Econometric Society* (1979): 263-291.
21. Roos, Patrick, J. Ryan Carr, and Dana S. Nau. "Evolution of state-dependent risk preferences." *ACM Transactions on Intelligent Systems and Technology (TIST)* 1.1 (2010): 6.
22. Lanchester, Frederick William. "Mathematics in warfare." *The world of mathematics* 4 (1956): 2138-2157.
23. Deitchman, Seymour J. "A Lanchester model of guerrilla warfare." *Operations Research* 10.6 (1962): 818-827.
24. Richbourg, Robert, and Warren K. Olson. "A hybrid expert system that combines technologies to address the problem of military terrain analysis." *Expert Systems with Applications* 11.2 (1996): 207-225.
25. Janlov, M., et al. "Developing military situation picture by spatial analysis and visualization." *ScanGIS*. 2005.
26. Okada, Isamu, and Hitoshi Yamamoto. "Mathematical description and analysis of adaptive risk choice behavior." *ACM Transactions on Intelligent Systems and Technology (TIST)* 4.1 (2013): 17.
27. Shakarian, Paulo, John P. Dickerson, and V. S. Subrahmanian. "Adversarial geospatial abduction problems." *ACM Transactions on Intelligent Systems and Technology (TIST)* 3.2 (2012): 34.

28. George, Stephen, Xing Wang, and Jyh-Charn Liu. "MECH: A model for predictive analysis of human choices in asymmetric conflicts." *Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer International Publishing, 2015. 302-307.
29. Steinbach, Marc C. "Markowitz revisited: Mean-variance models in financial portfolio analysis." *SIAM review* 43.1 (2001): 31-85.
30. Krokmal, Pavlo, Michael Zabrankin, and Stan Uryasev. "Modeling and optimization of risk." *Surveys in Operations Research and Management Science* 16.2 (2011): 49-66.
31. Lin, Jason, et al. "Risk management in asymmetric conflict: using predictive route reconnaissance to assess and mitigate threats." *Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer International Publishing, 2015. 350-355.
32. Ranger Training Brigade, *Ranger Handbook*. Fort Benning, GA: Department of the Army, 2006
33. Guevara, Che. *Guerrilla warfare*. Rowman & Littlefield Publishers, 2002.
34. Corps, US Marine. "Mao Tse-tung on Guerrilla Warfare." *Fleet Marine Force Reference Publication*. Last accessed 24 (1989): 2014.

## APPENDIX A. Data set used for this project

### A.1 Global Digital Elevation Model

Elevation maps were obtained from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 [127]. Dated October 2011, these maps offer digital elevations with a horizontal resolution of approximately 30 meters (1/3 arc second). The data uses the WGS84 geoid and is stored in GeoTIFF format in 1° x 1° tiles. The ASTER L1B data were obtained through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota ([https://lpdaac.usgs.gov/get\\_data](https://lpdaac.usgs.gov/get_data)) .

The absolute vertical error for this product is estimated to be  $\pm 17$  meters. However, the relative error (between adjacent pixels) is much smaller. At the distances used for MECH analysis, on the order of 3 km, it is unlikely that this error is problematic.

A more serious problem is the resolution. Each pixel in these elevation maps covers a  $\sim 30 \times 30$  meter square. There are probably features of interest that are small enough to hide within the large pixels. We believe that a more appropriate resolution is on the order of 5 meters.

## **A.2 Asymmetric Conflict Events**

Asymmetric warfare events were obtained from the ISAF-NATO Civilian Integration Team. The events are provided as a service to contractors and non-governmental agencies that operate or may operate in Afghanistan. An UNCLASSIFIED//FOUO extract of the Afghanistan SIGACTS database, this dataset consists of a variety of events including IEDs, direct fire, indirect fire, surface-to-air fire and more. The data is provided with at least one week delay and occasional outages and missing data occur.

The dataset used for this analysis includes 33,140 events that occurred between February 01, 2011 and August 23, 2012. Of these, 13,295 were classified as IED and 16610 were classified as direct fire.

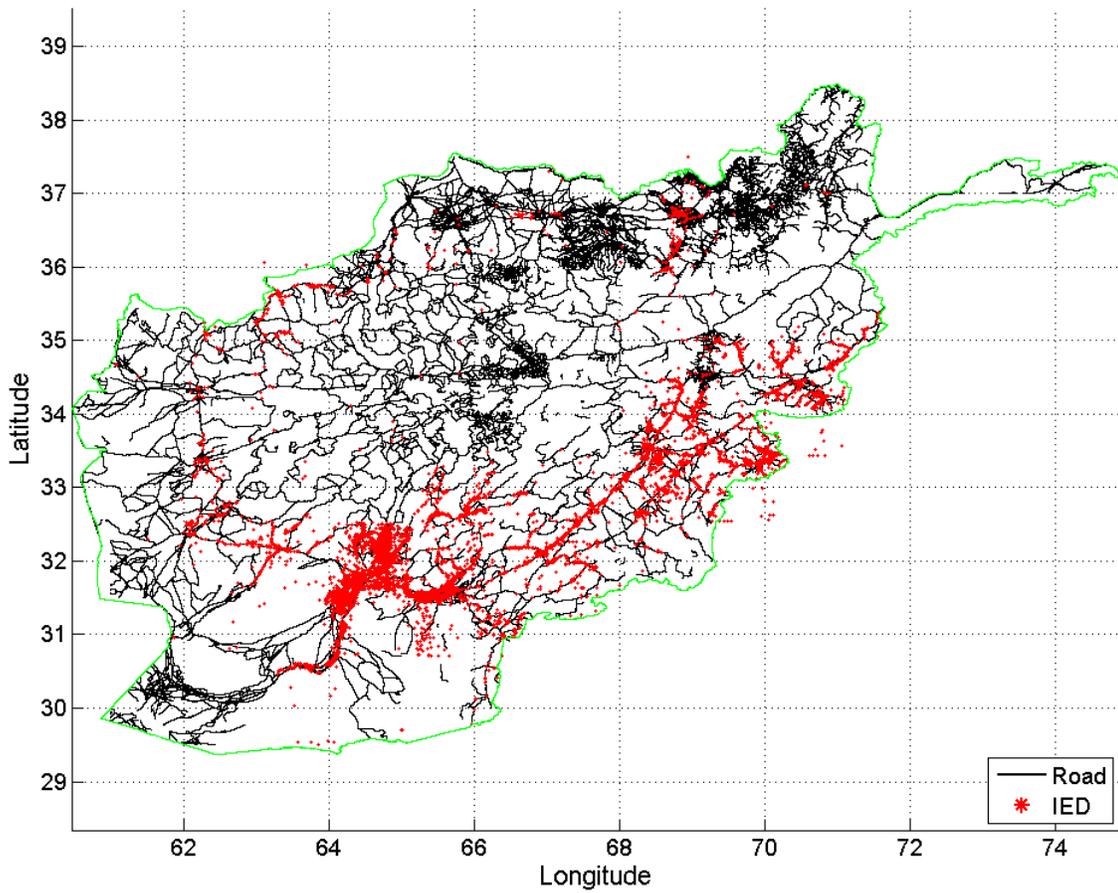
Table 10 shows the dates covered by the current dataset.

**Table 8: Date coverage of asymmetric warfare events in the ISAF-NATO Civilian Integration Team Unclassified Dataset.**

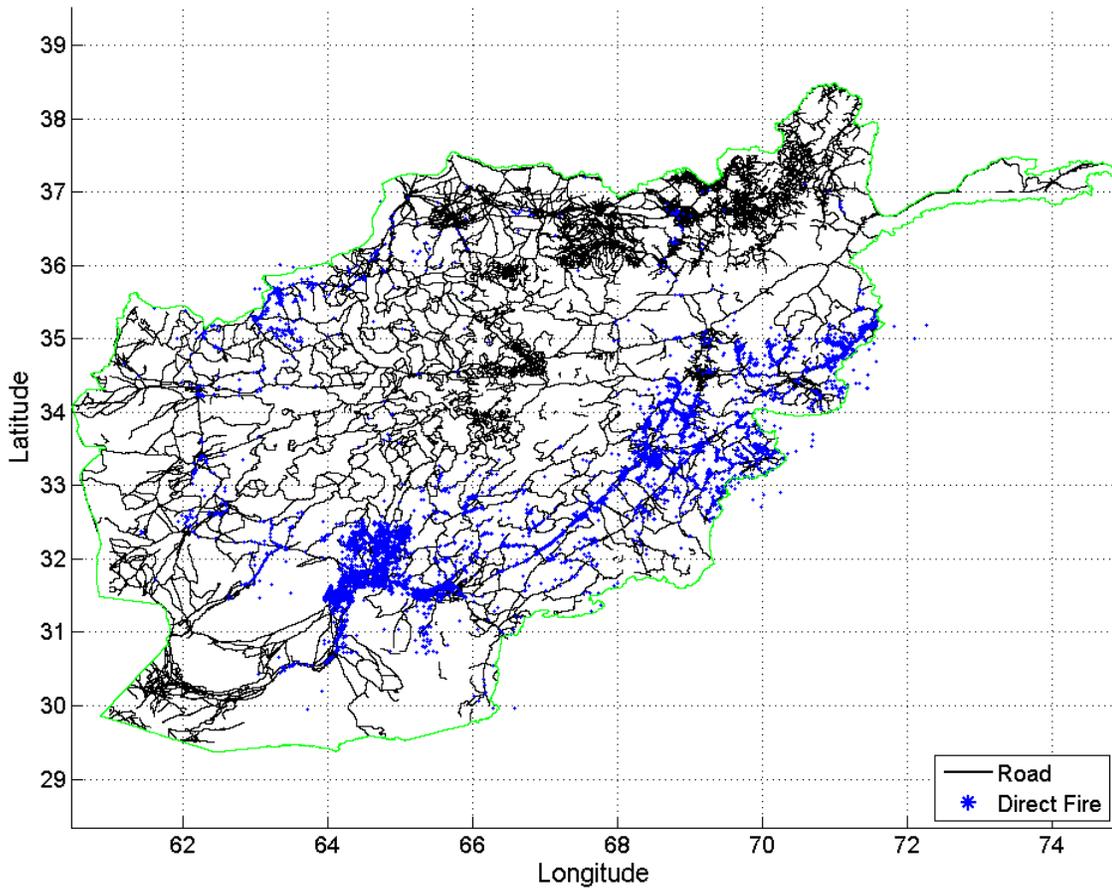
		2011											2012									
Day of the month		J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13
	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16
	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17
	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18
	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19
	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23
	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26
	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28
	29		29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29
	30		30	30	30	30	30	30	30	30	30	30	30	30		30	30	30	30	30	30	30
31		31		31		31	31		31		31	31		31		31		31		31	31	

 Data not available from the CIT website.

The following figures offer an overview of the geographic distribution of IED and direct fire events in Afghanistan.

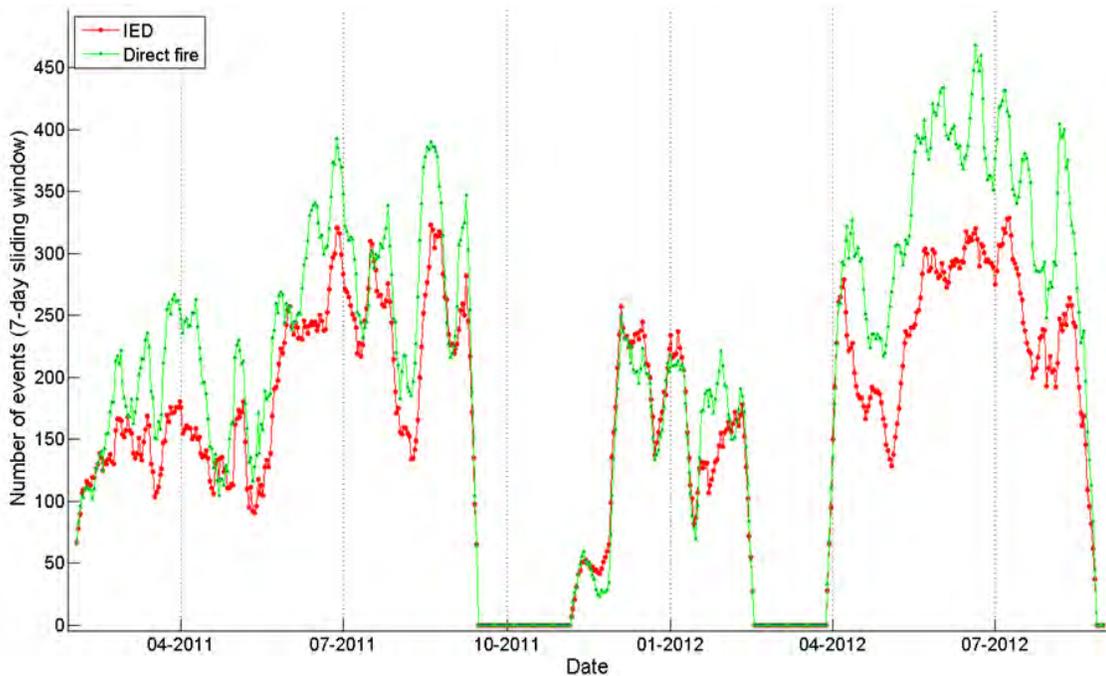


**Figure 44: IED attacks in Afghanistan, February 2011 - August 2012.**



**Figure 45: Direct fire attacks in Afghanistan, February 2011 - August 2012.**

Figure 46 shows the distribution of events by date and type using a 7-day sliding window. All of the events show the same general trends. Note that the trend to zero in August 2012 is an edge effect based on availability of data. The dips in Sep-Oct 2011 and Feb-Mar 2012 are also due to missing data.



**Figure 46: Distribution of events by date, using a 7-day sliding window.**

### *Collocated Events*

Some locations also lend themselves to multiple events, sometimes of the same type and sometimes not. In the following analysis, collocated events are defined as successive events that occur within 250 meters and 1.5 hours of each other. The threshold of 250 meters was selected to account for typical patrol configurations, where vehicles are separated by 25 meters or more. The time window was selected based on anecdotal information about typical patrol behavior following an attack. A total of 894 initiating events meet these criteria.

**Table 9: Collocated events**

Initiating Event	Following Event	Count
IED	IED	319
IED	Direct fire	63
Direct fire	IED	42
Direct fire	Direct fire	470

### *Dataset Problems*

A principle problem with this conflict event dataset is data quality. In particular, the exact coordinates of events seem to be collected in a variety of ways and using a variety of datums. No information is provided to assess or normalize these inputs. Thus, the dataset is likely to contain locations that are erroneous due to estimation errors, datum translation errors, and simple manual data entry error.

Another problem is the lack of descriptor specificity. All IED events are classified with the same descriptor. Thus, command-detonated and victim-detonated devices are labeled with the same descriptor. Similarly, all direct fire events share the same label. So, a company-sized ambush and an individual sniper attack are marked as being part of the same class.

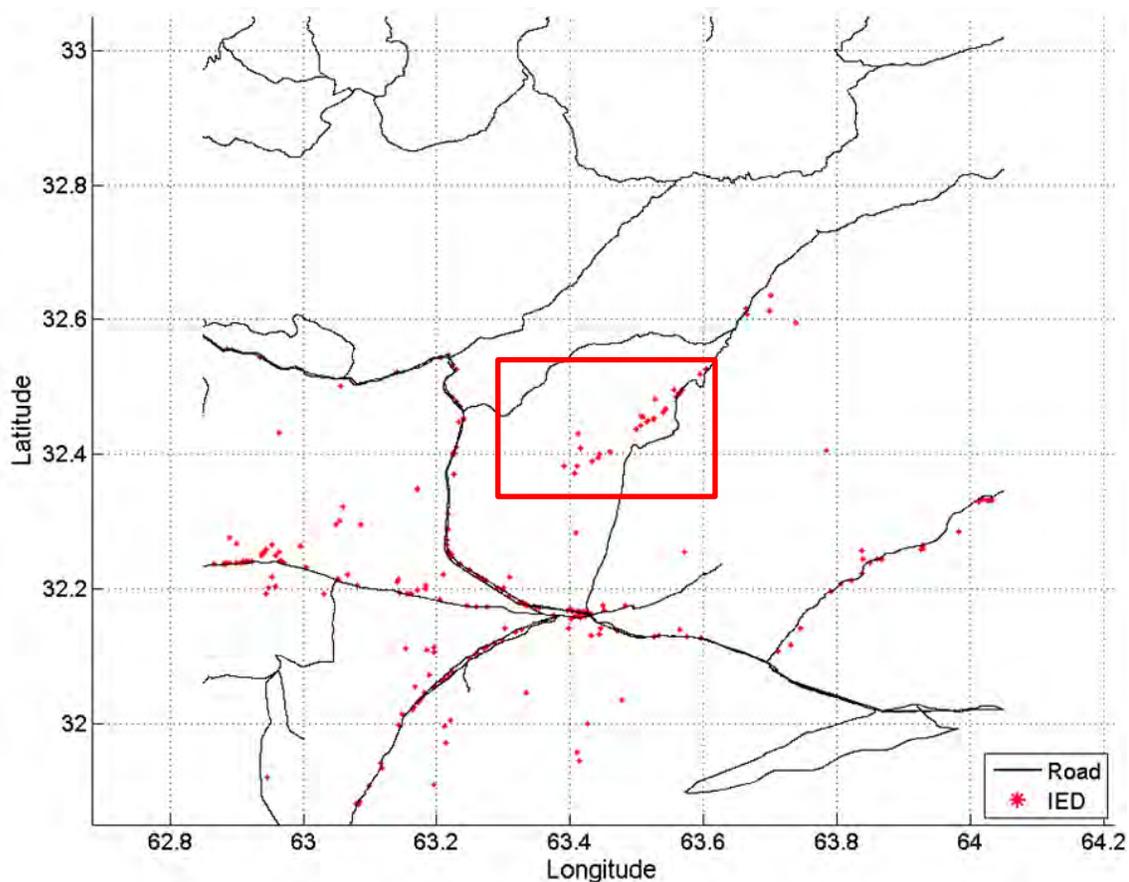
A final problem is the lack of consistency in the measurements. For an ambush, the specified location is likely to be the geographic coordinate of the person reporting the event when it started. Depending on the size of the convoy or patrol, this location could be tens to hundreds of meters from the actual place where the attack actually occurred. If the patrol was moving during the attack, the location may be estimated or may be the place where the patrol stopped. Similar problems exist for IED events.

### A.3 Afghanistan Roads

Road data is collected and maintained by the Afghanistan Information Management Service (<http://www.aims.org.af>) and distributed by mapcruzin.com at <http://www.mapcruzin.com/afghanistan-shapefiles/roads.zip>. Three types of roads are identified, including all weather primary, all weather secondary and tracks.

The roads are stored as polyline objects in a shapefile. In order to use the roads for this project, each road segment was split into discrete points at 30 meter intervals (the resolution of the elevation maps). A total of 3,306,680 discrete points were produced in this way.

Principle problems with this dataset include its age and apparent incompleteness. As far as can be determined, this map of roads was produced in the early 2000's using data from Russian and U.S. maps published in the 1980's. The age of the data suggests that some current roads may be missing from the map, particularly after post-war reconstruction efforts by the U.S. and others. Figure 70 illustrates the problem. A red square draws attention to a number of IED events that seemed to occur away from roads. Visual analysis of Google Earth imagery reveals the presence of a road and a number of villages along a watercourse. It seems likely that this road did not exist or was not surveyed when the Russian and U.S. maps were originally created.

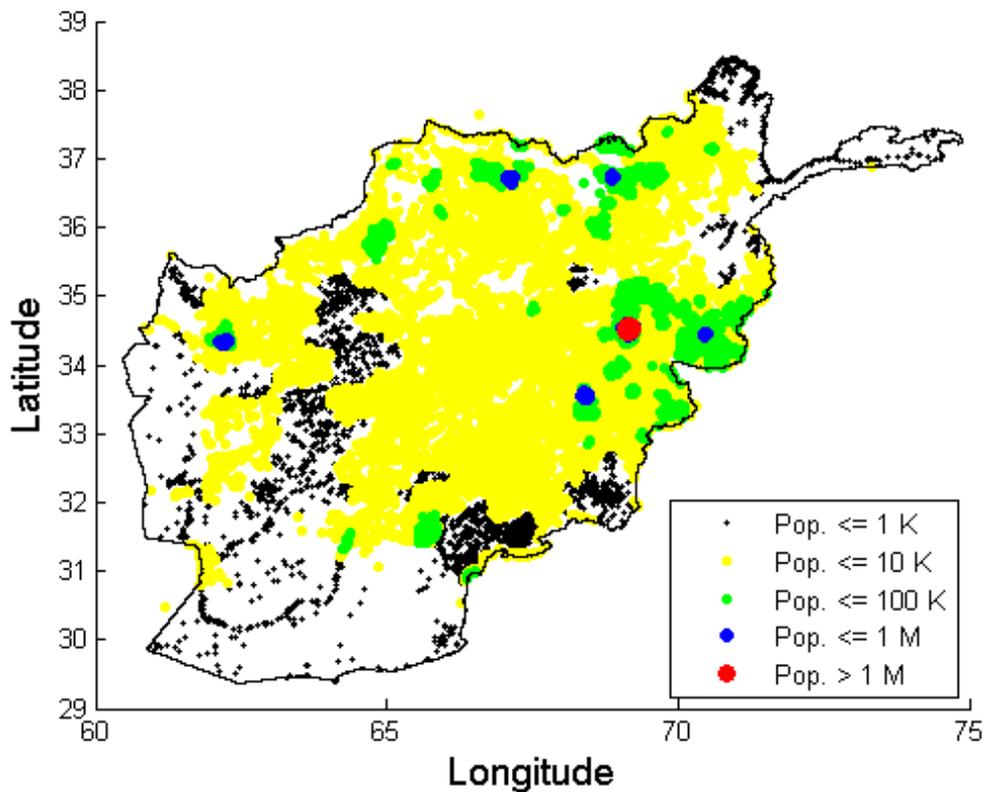


**Figure 47: Example of area where IED events occurred away from known roads.**

#### **A.4 Population**

Population estimates were scraped from the site <http://www.fallingrain.com/world/AF/>. For each known, fixed, and named populated place (village, town, city) the total population within 7 km is estimated. Figure 71 gives an idea of the population distribution in Afghanistan.

The source of the raw population data is unknown. Therefore, the validity of the estimated populations is unknown. Additionally, as a largely rural and tribal society, Afghan participation in a national census is likely to be less than complete. Informally, the fallingrain.com estimates for the area surrounding the center of Kabul, Kandahar and Mazar-i-sharif appear to be roughly consistent with Wikipedia estimates for population for the same towns.

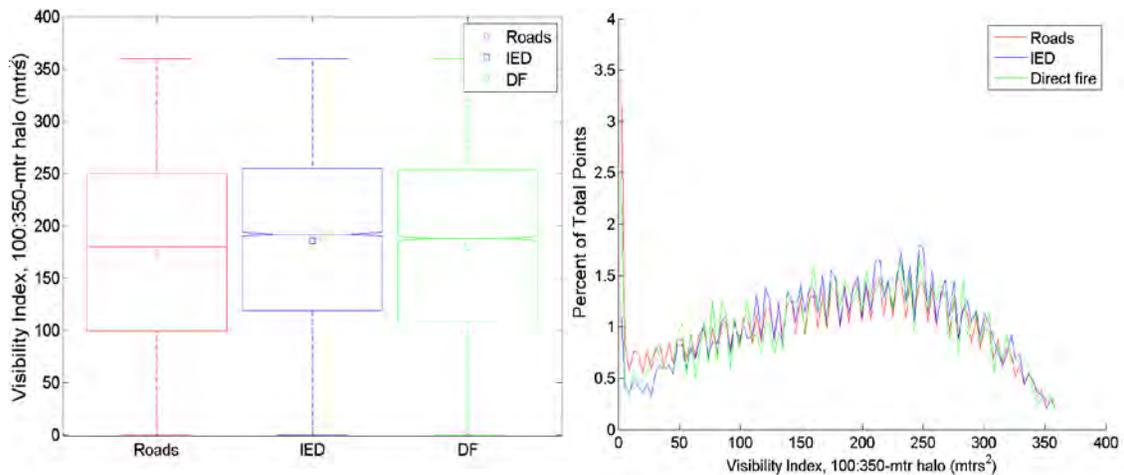


**Figure 48: Estimated population at locations throughout Afghanistan.**

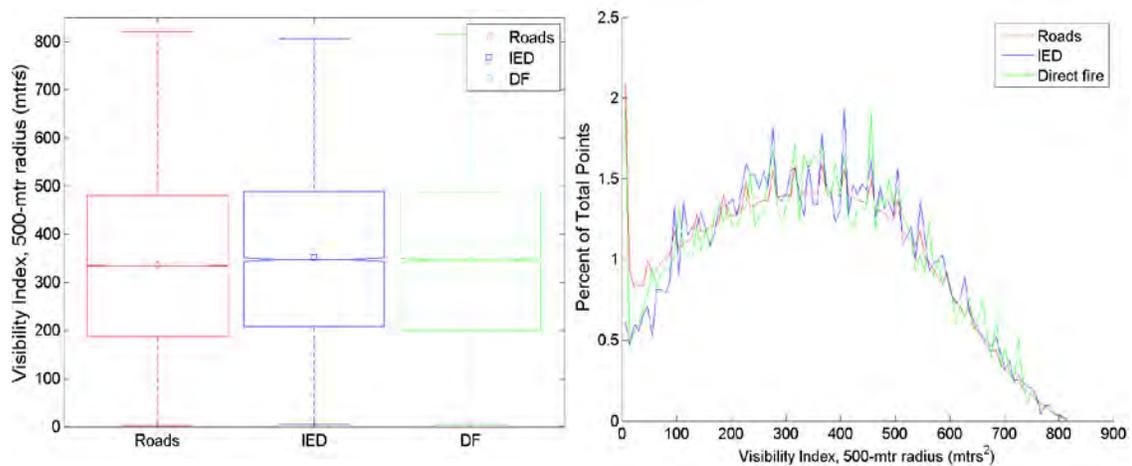
## Appendix B. Additional Features

This appendix contains additional features. For the most part, these features are similar to those already presented in the body of the dissertation and differ only in window size, radius, or number of radials. They are included here for completeness.

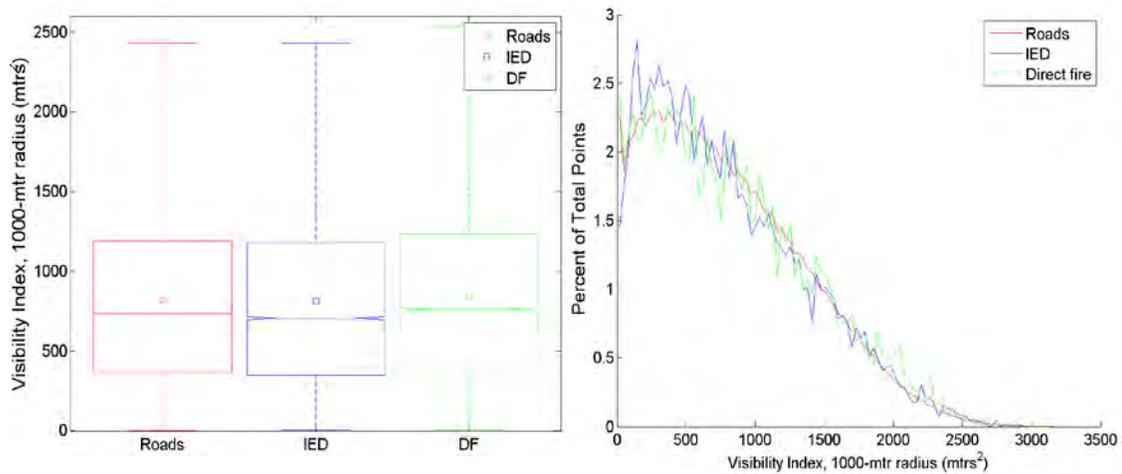
### Appendix B.1 Visibility Index



**Figure 49: Visibility Index inside a halo with an inner radius of 100 meters and an outer radius of 350 meters.**

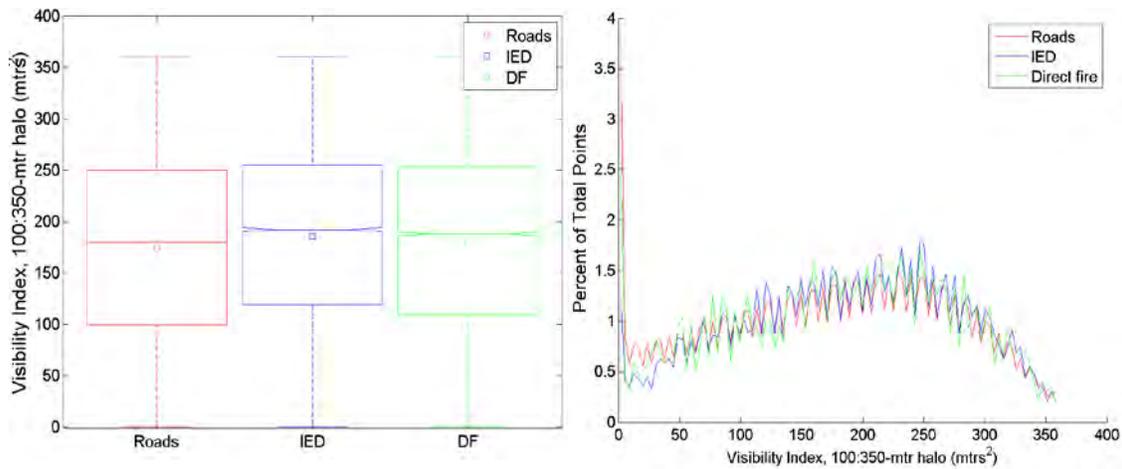


**Figure 50: Visibility Index at a radius of 500 meters.**

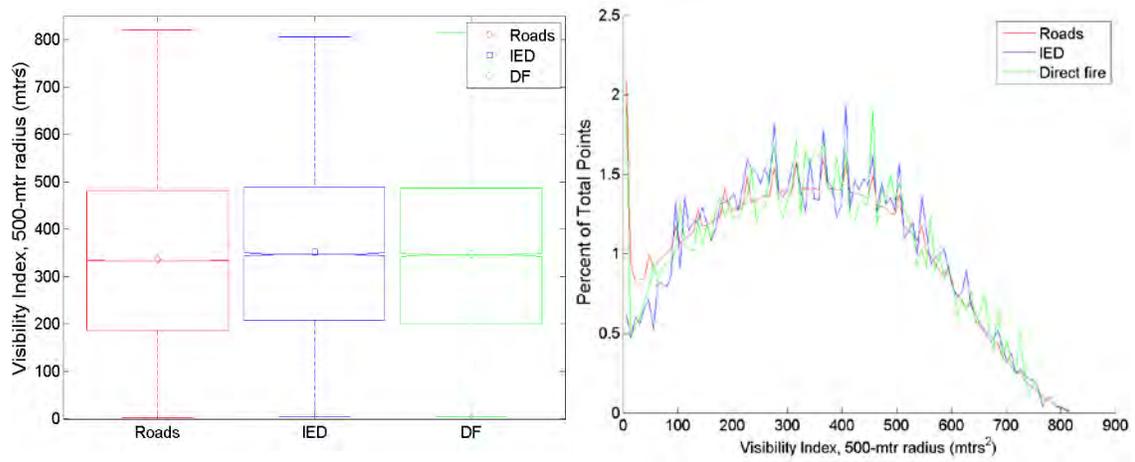


**Figure 51: Visibility Index at a radius of 1000 meters.**

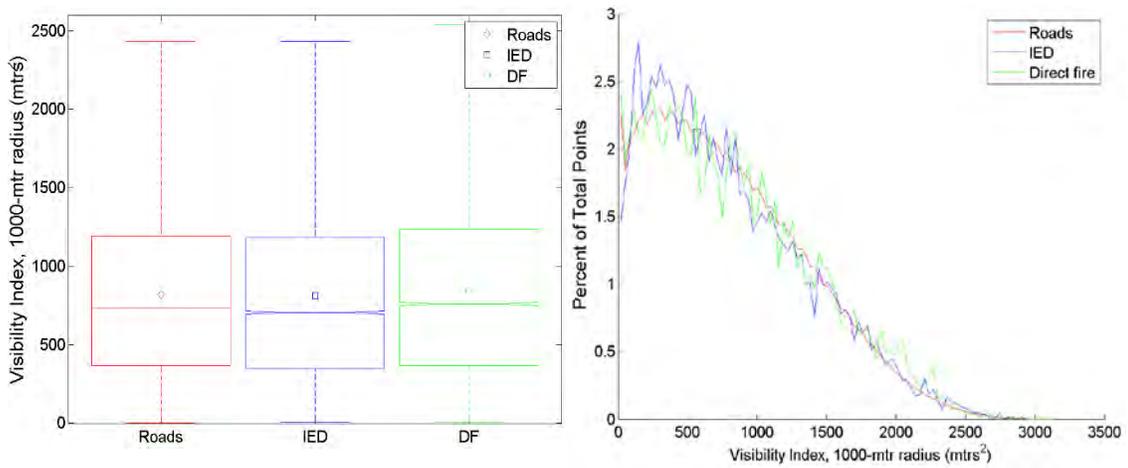
### Appendix B.2 Discrete Shape Complexity Index



**Figure 52: Discrete Shape Complexity Index in a halo with an inner radius of 100 meters and an outer radius of 350 meters.**

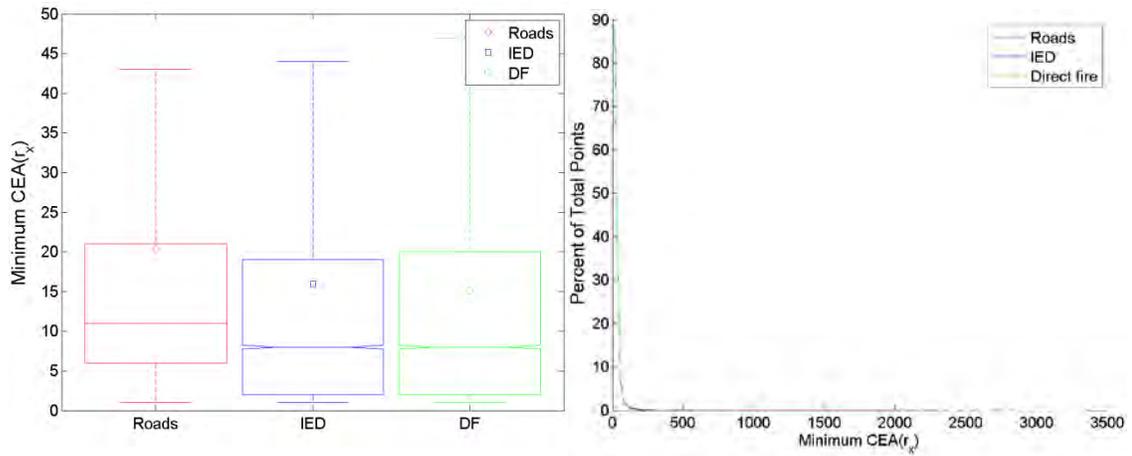


**Figure 53: Discrete Shape Complexity Index at a radius of 500 meters.**

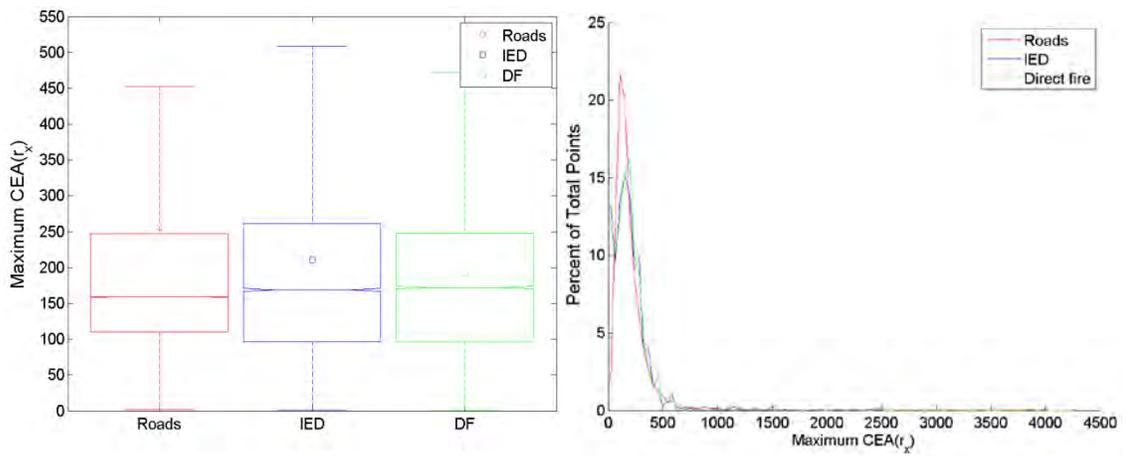


**Figure 54: Discrete Shape Complexity Index at a radius of 1000 meters.**

### Appendix B.3 Cumulative Escape Adjacency for a Single Point



**Figure 55: Minimum CEA( $r_x$ ).**



**Figure 56: Maximum CEA( $r_x$ ).**

## Appendix B.4 Median Route Visibility

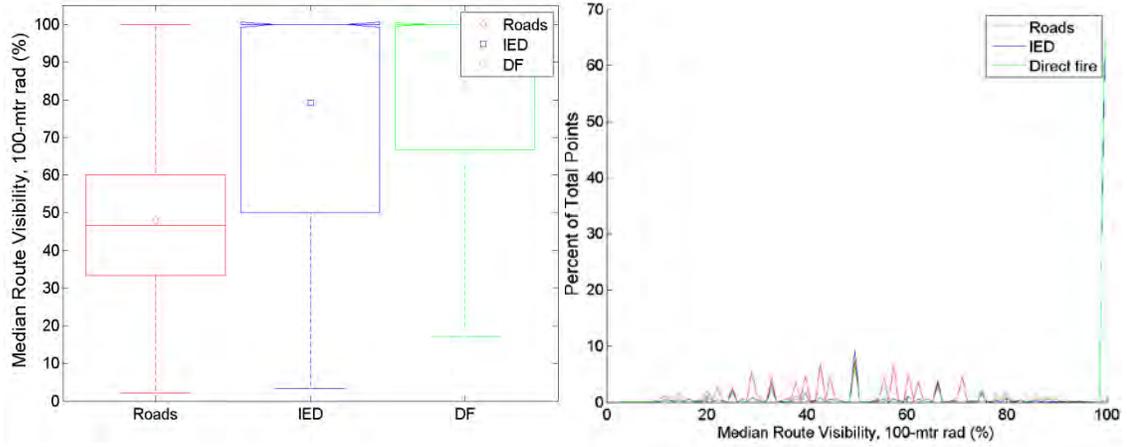


Figure 57: Median route visibility at 100 meters.

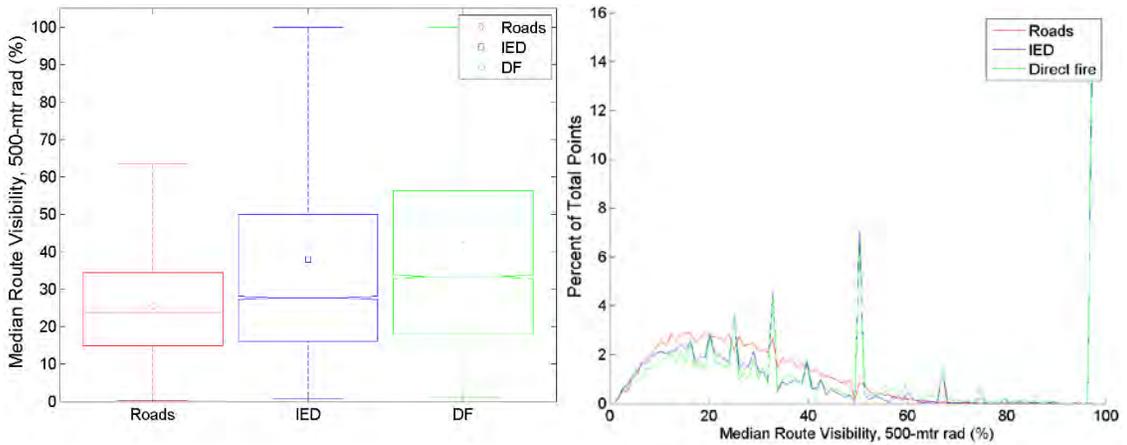


Figure 58: Median route visibility at 500 meters.

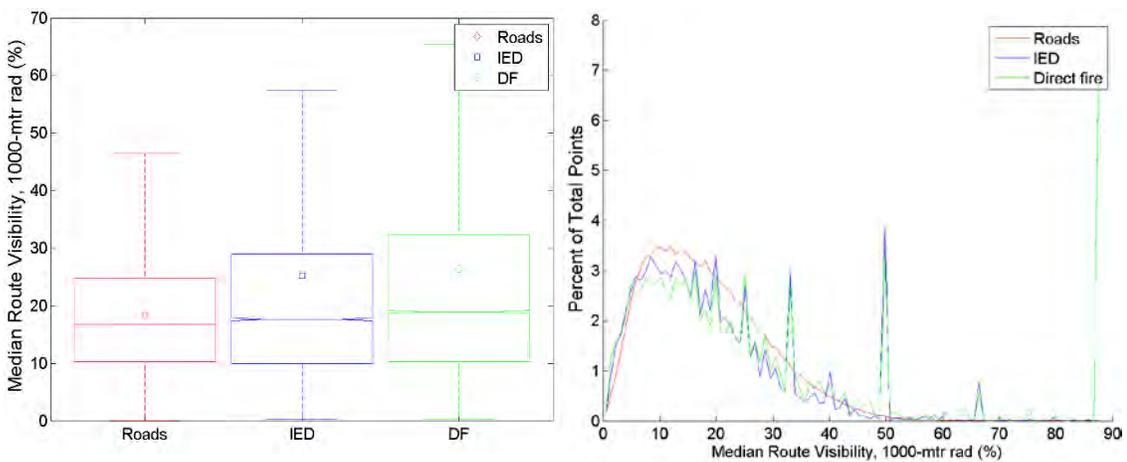
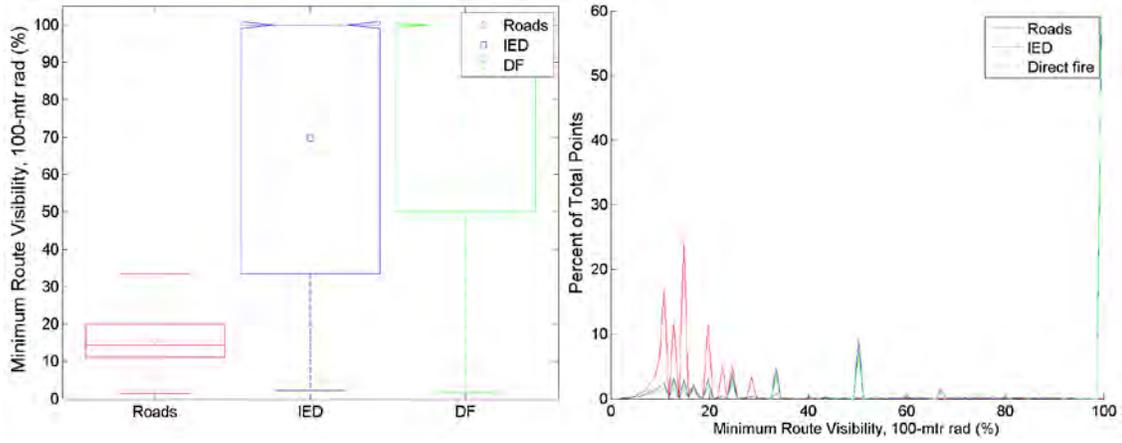
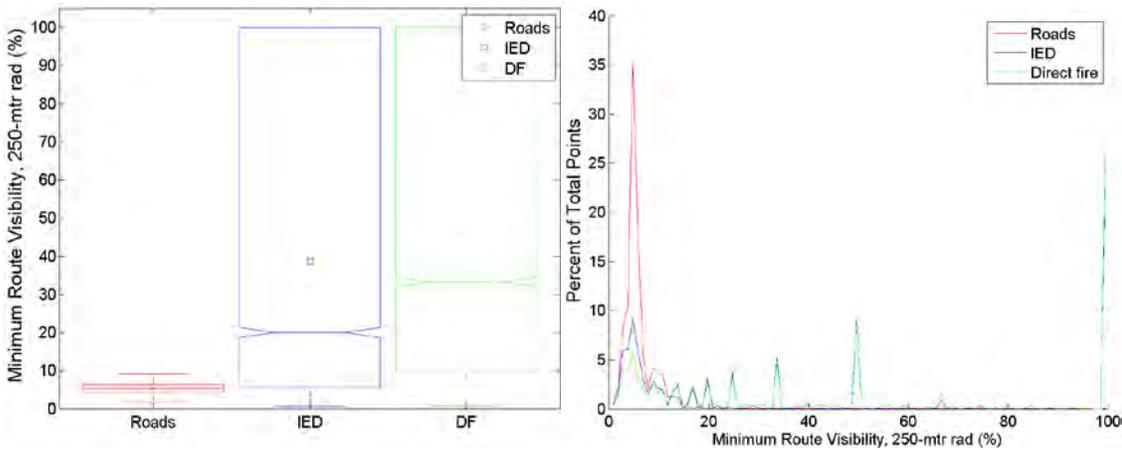


Figure 59: Median route visibility at 1000 meters.

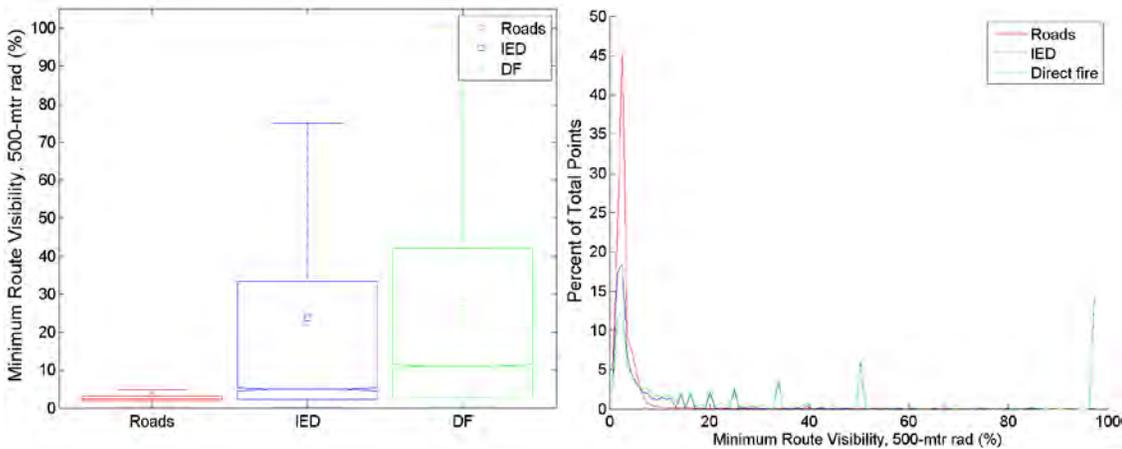
## Appendix B.5 Minimum Route Visibility



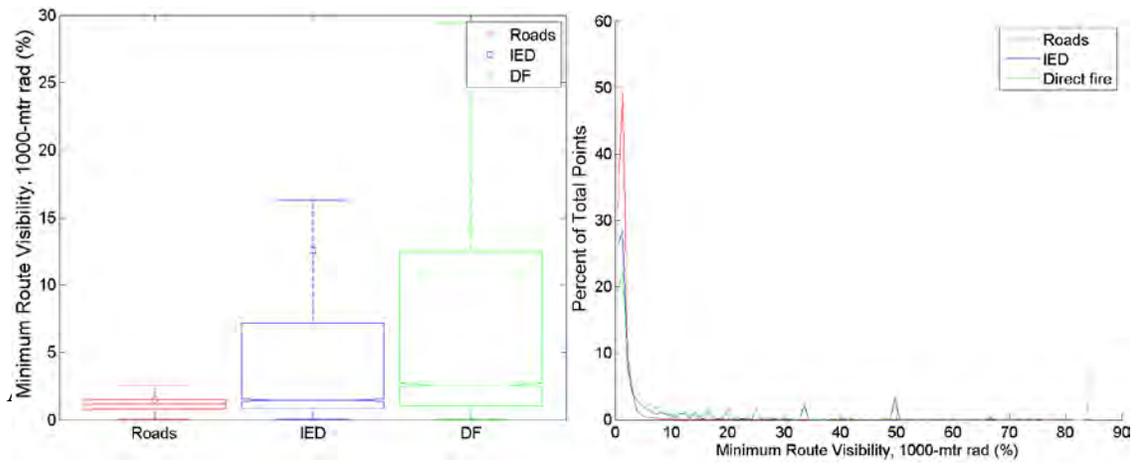
**Figure 60: Minimum route visibility at 100 meters.**



**Figure 61: Minimum route visibility at 250 meters.**

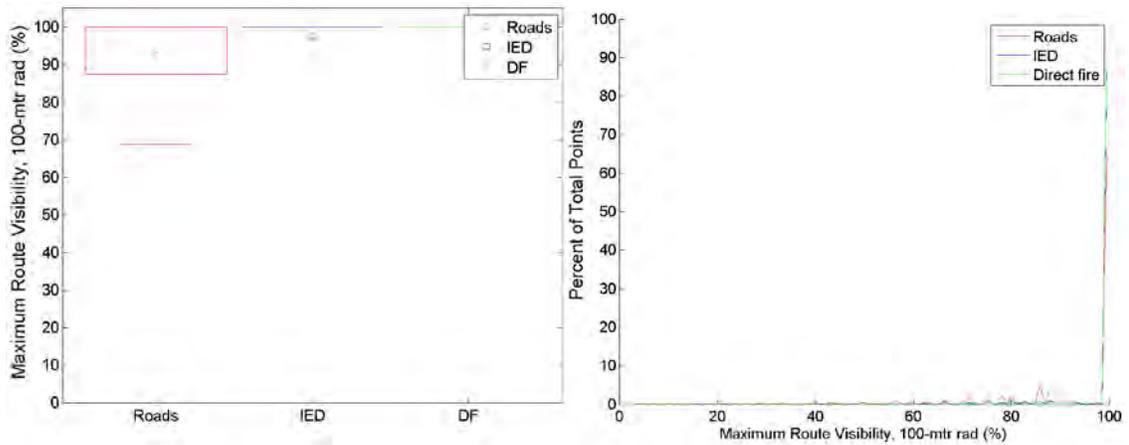


**Figure 62: Minimum route visibility at 500 meters.**

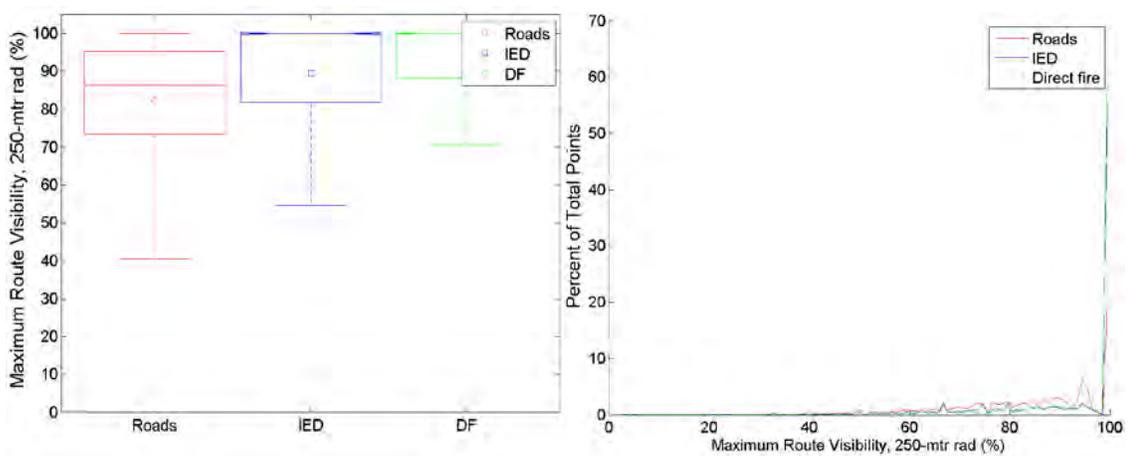


**Figure 63: Minimum route visibility at 1000 meters.**

### Appendix B.6 Maximum Route Visibility



**Figure 64: Maximum route visibility at 100 meters.**



**Figure 65: Maximum route visibility at 250 meters.**

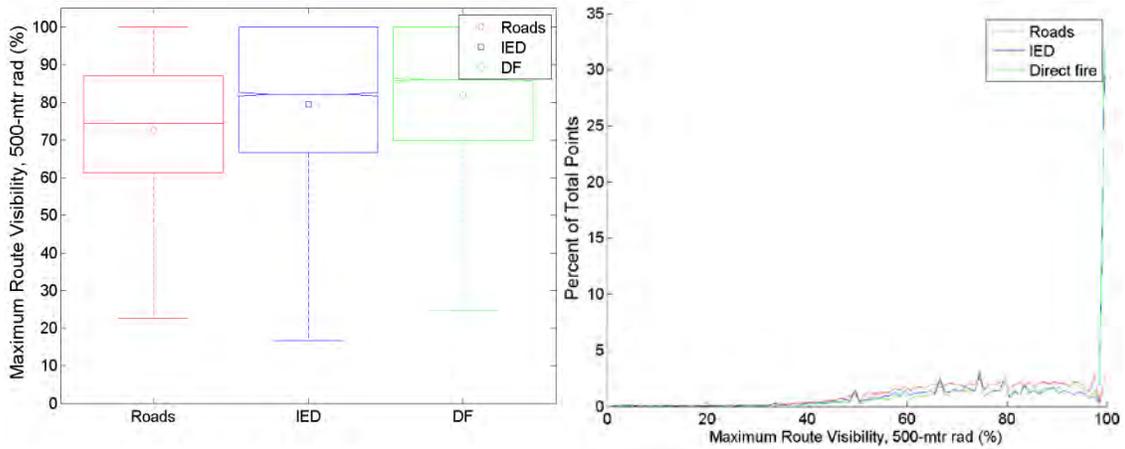


Figure 66: Maximum route visibility at 500 meters.

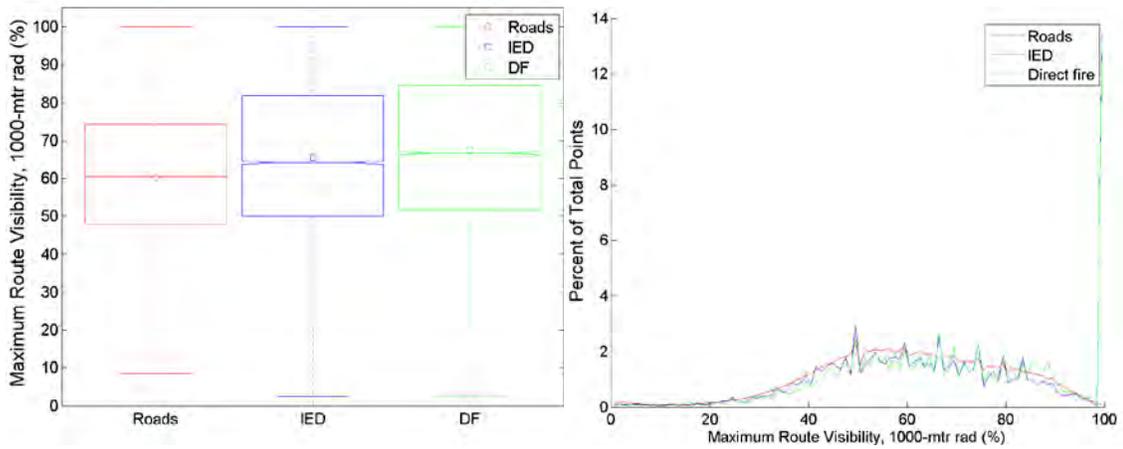


Figure 67: Maximum route visibility at 1000 meters.

### Appendix B.7 Sparse Viewshed Shortest Radial

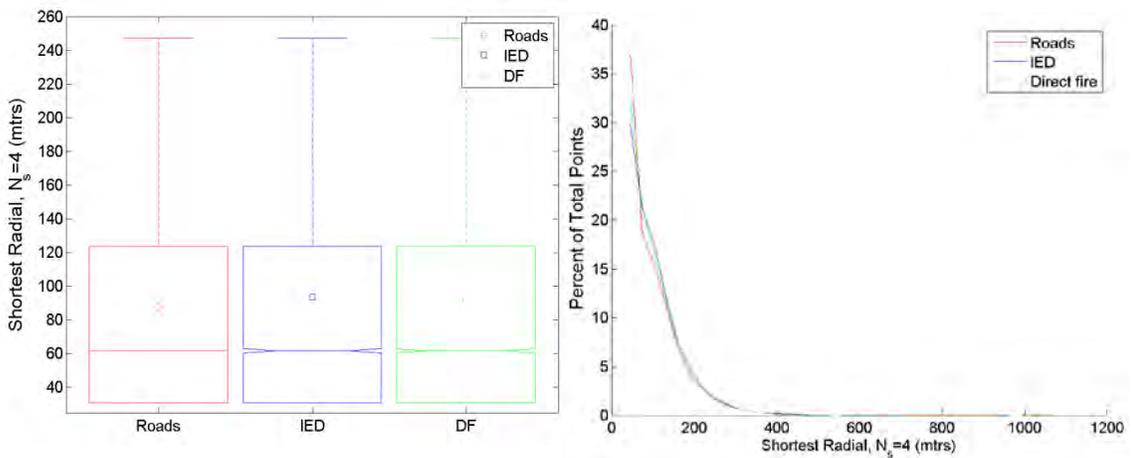
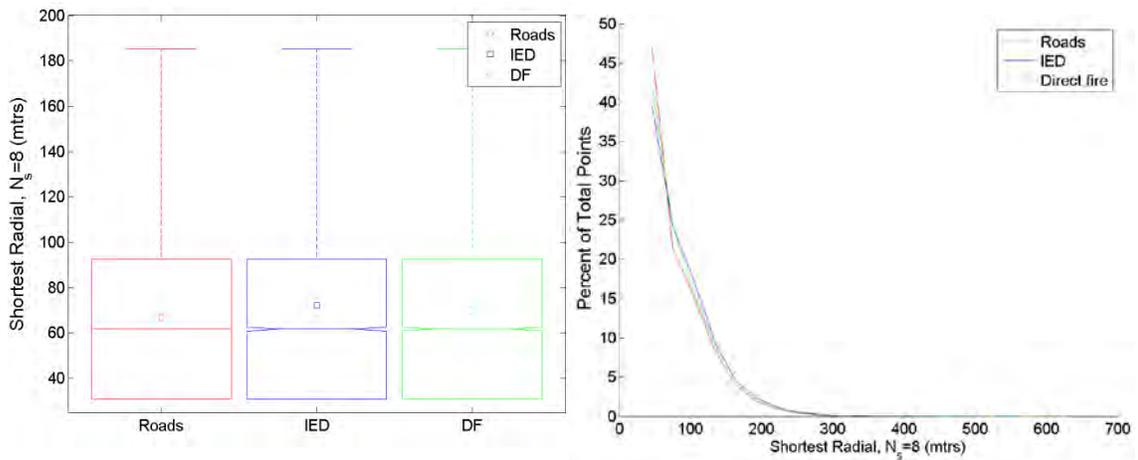
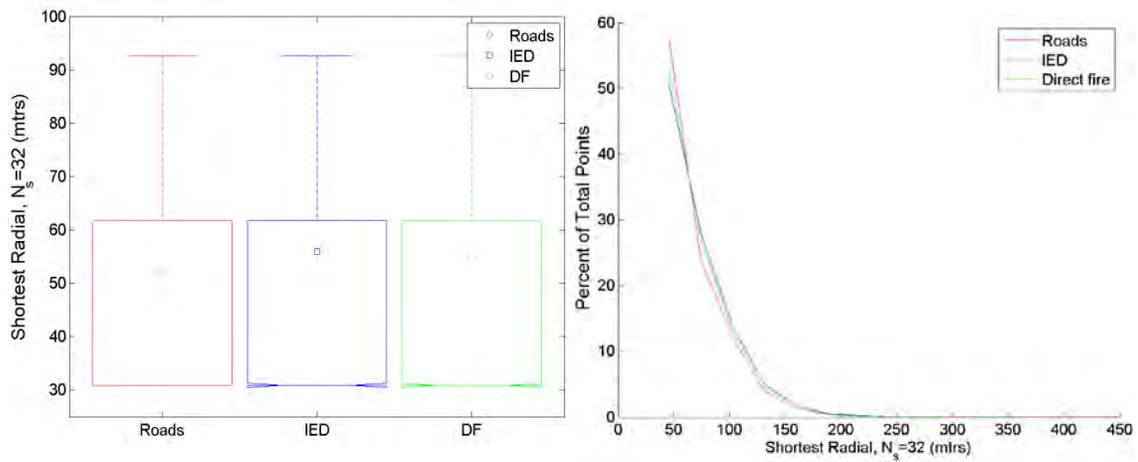


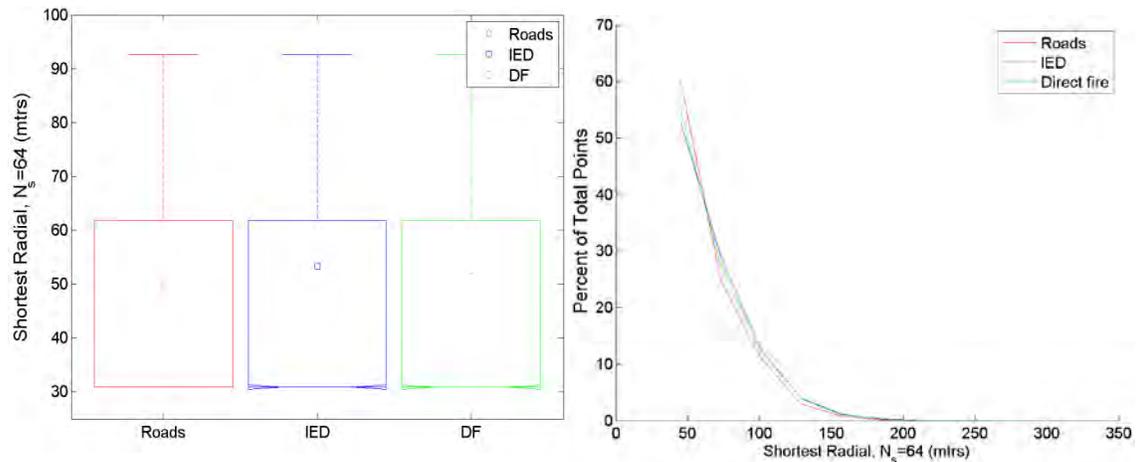
Figure 68: Sparse viewshed shortest radial ( $N_s = 4$ ).



**Figure 69: Sparse viewshed shortest radial ( $N_s = 8$ ).**



**Figure 70: Sparse viewshed shortest radial ( $N_s = 32$ ).**



**Figure 71: Sparse viewshed shortest radial ( $N_s = 64$ ).**

### Appendix B.8 Sparse Viewshed Longest Radial

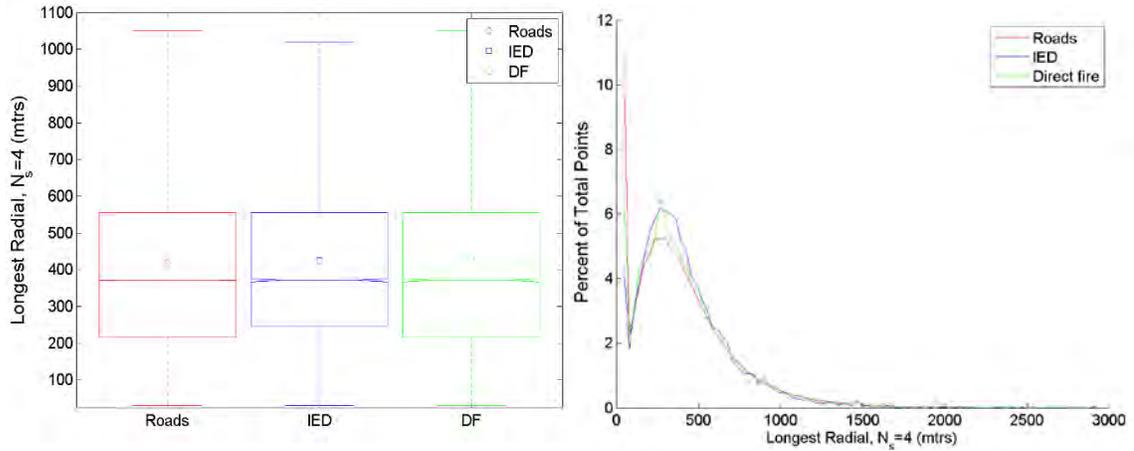


Figure 72: Sparse viewshed longest radial ( $N_s = 4$ ).

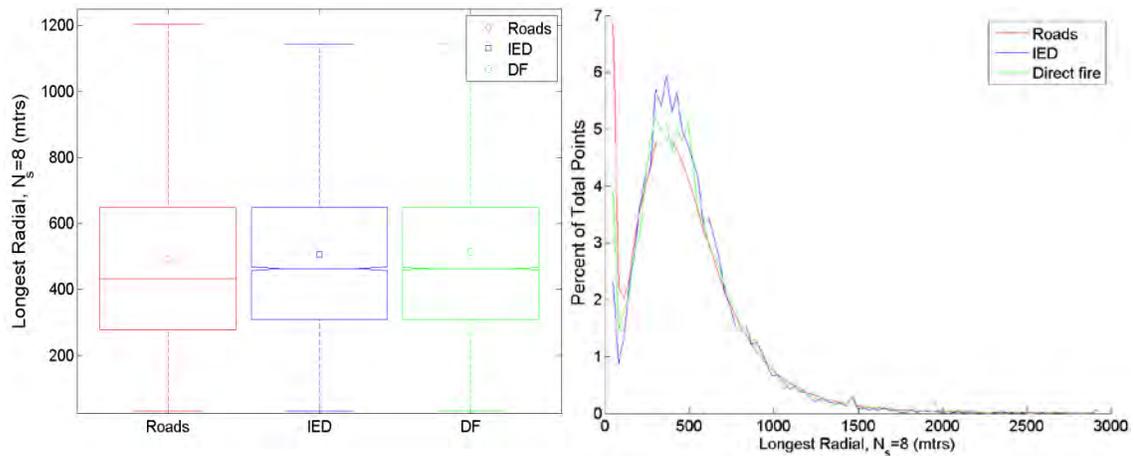


Figure 73: Sparse viewshed longest radial ( $N_s = 8$ ).

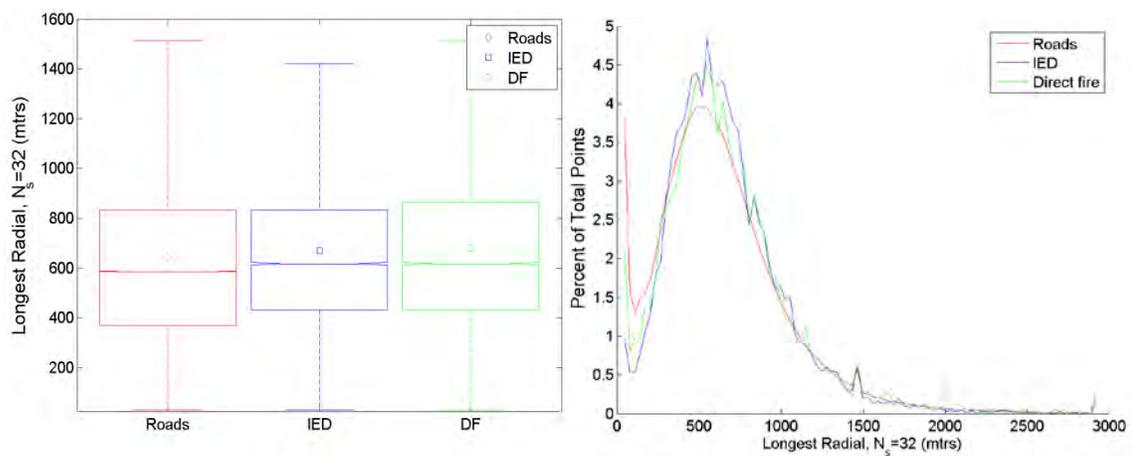
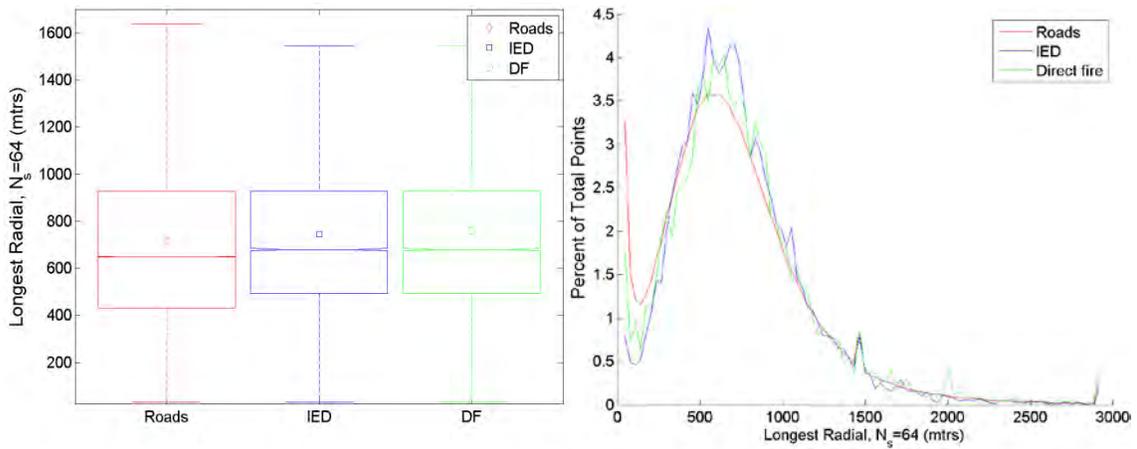
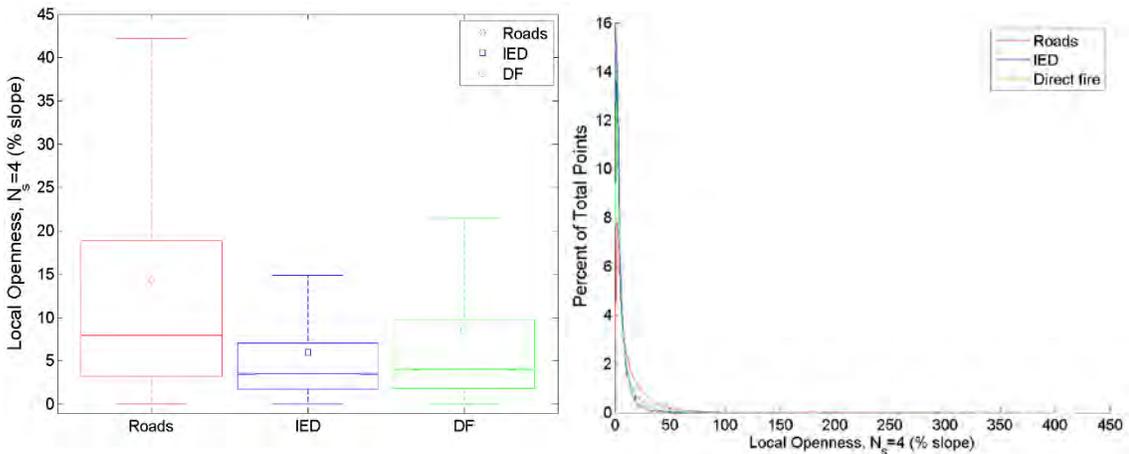


Figure 74: Sparse viewshed longest radial ( $N_s = 32$ ).

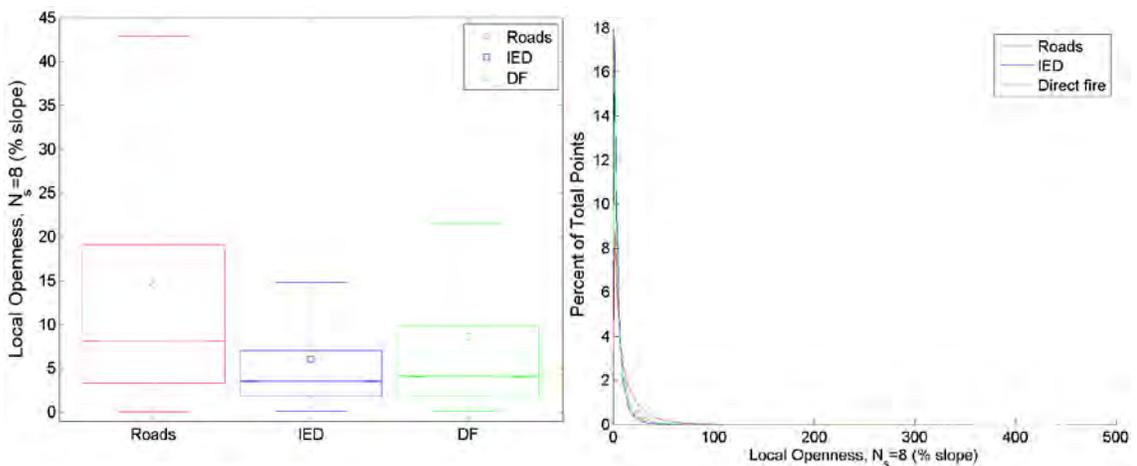


**Figure 75: Sparse viewshed longest radial ( $N_s = 64$ ).**

### Appendix B.9 Sparse Viewshed Local Openness



**Figure 76: Sparse viewshed local openness ( $N_s = 4$ ).**



**Figure 77: Sparse viewshed local openness ( $N_s = 8$ ).**

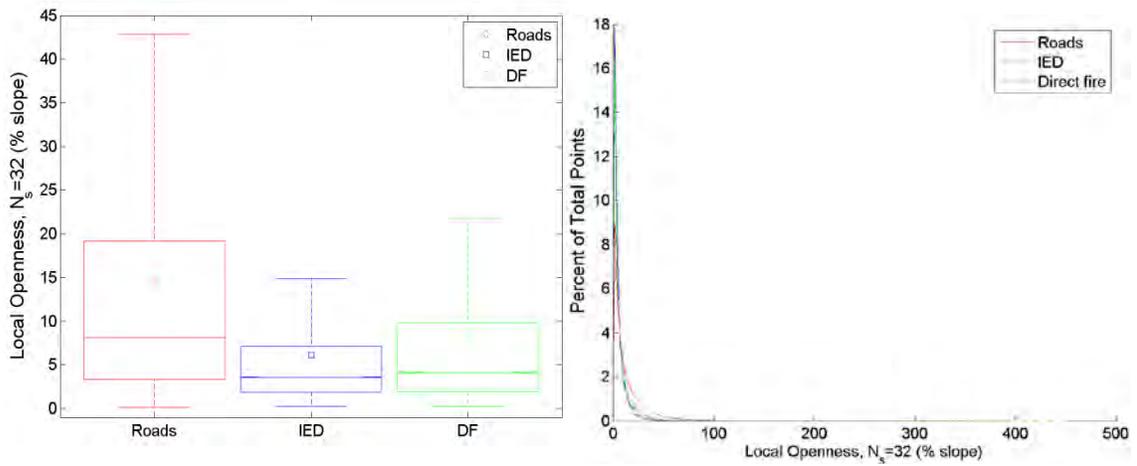


Figure 78: Sparse viewshed local openness ( $N_s = 32$ ).

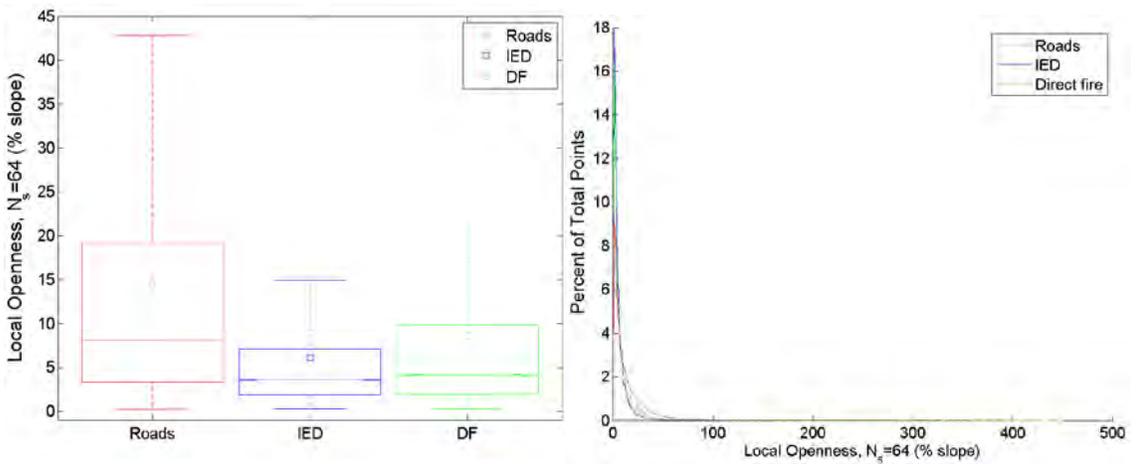


Figure 1: Sparse viewshed local openness ( $N_s = 64$ ).

Appendix B.10 Elevation Range

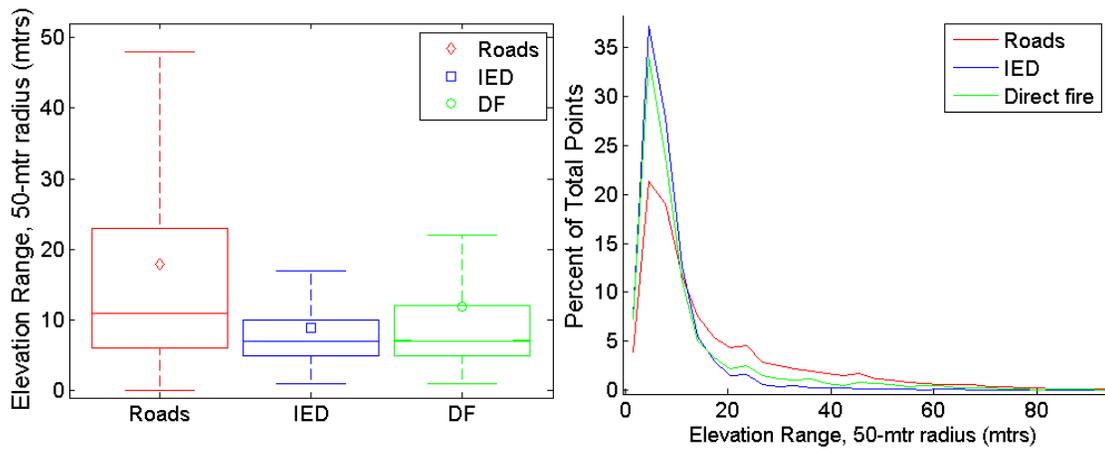
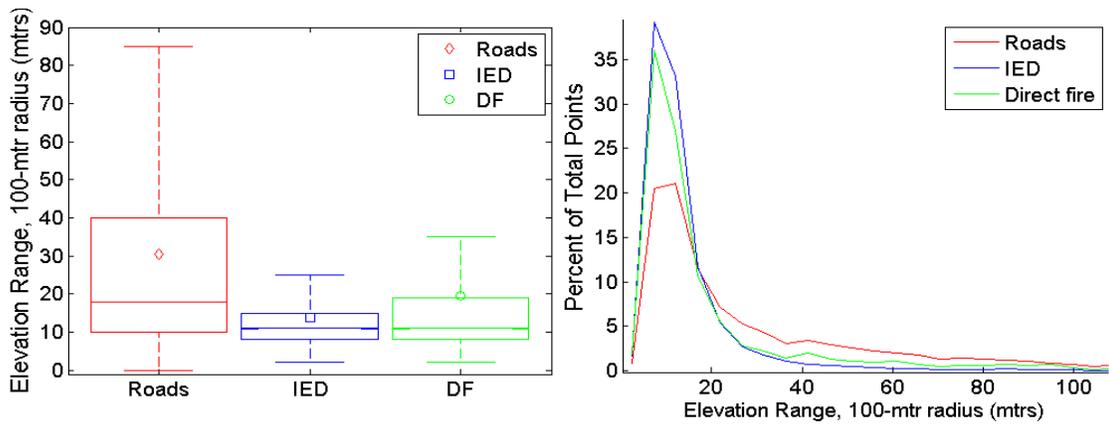
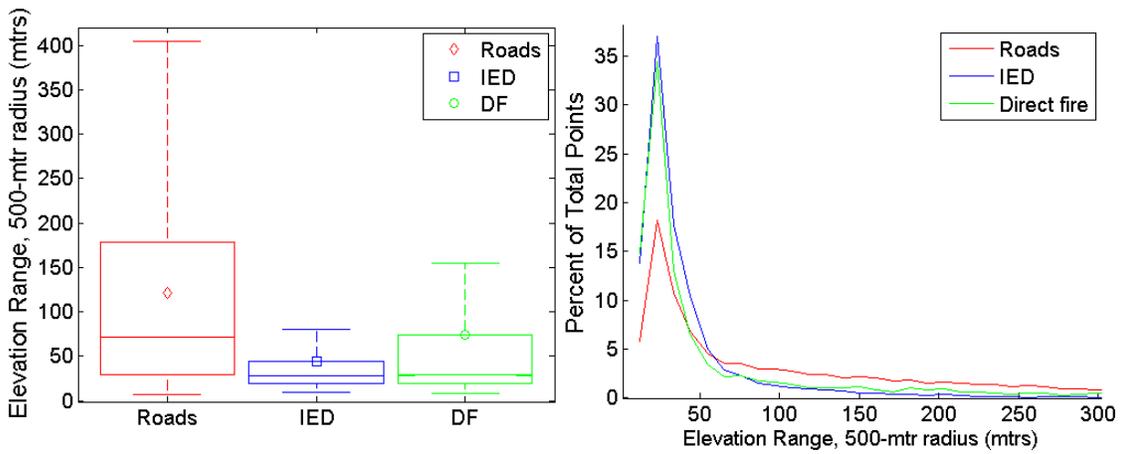


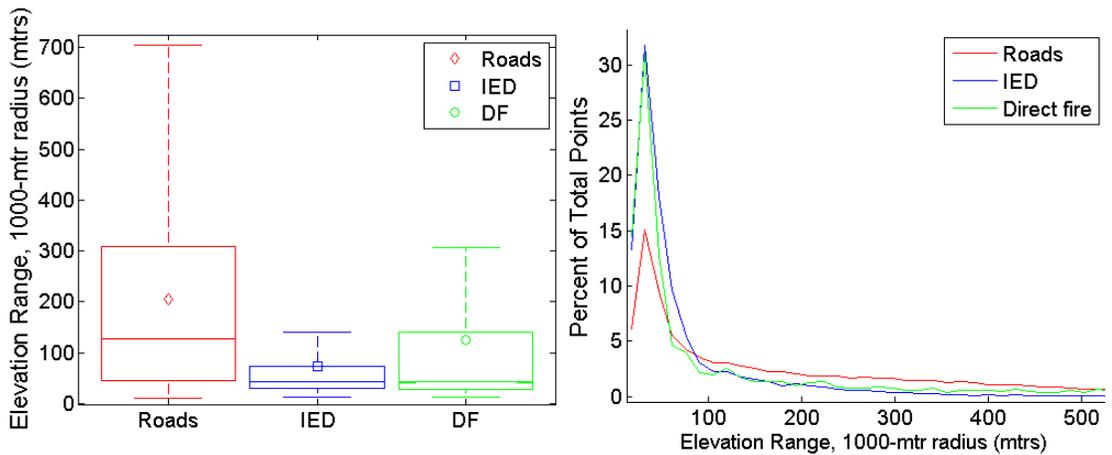
Figure 80: Elevation range at a radius of 50 meters.



**Figure 81: Elevation range at a radius of 100 meters.**

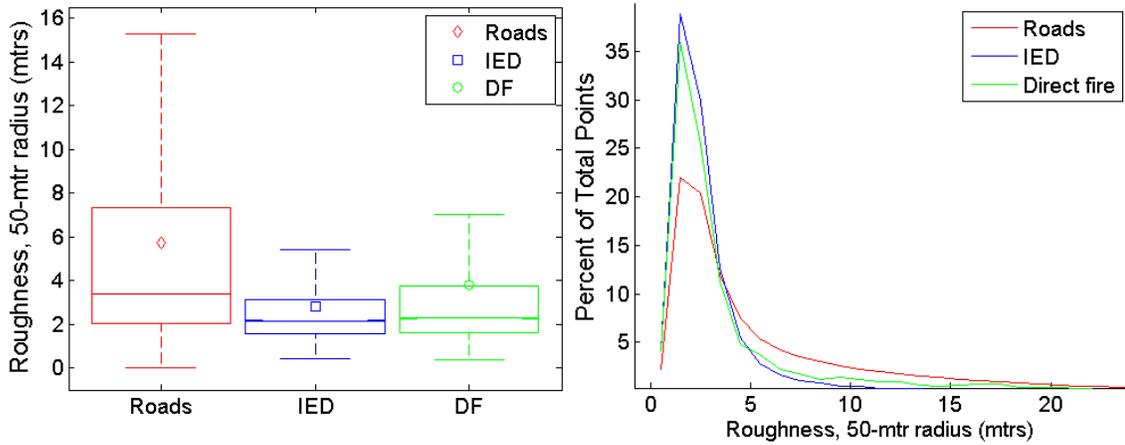


**Figure 82: Elevation range at a radius of 500 meters.**

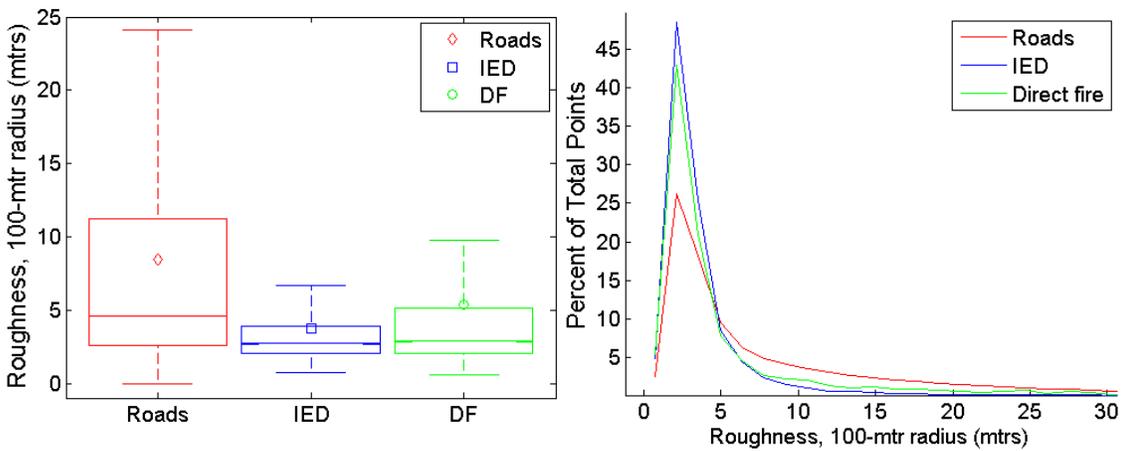


**Figure 83: Elevation range at a radius of 1000 meters.**

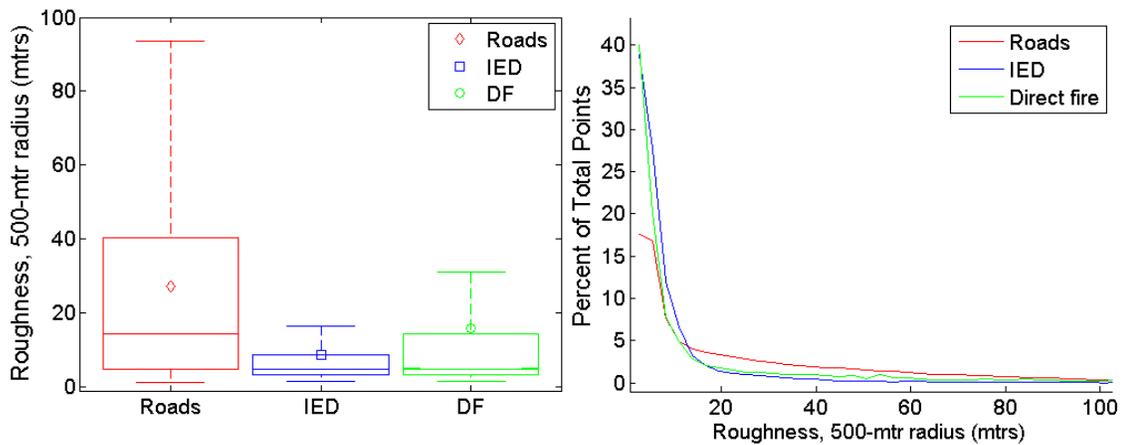
### Appendix B.11 Roughness (Standard Deviation of Elevation)



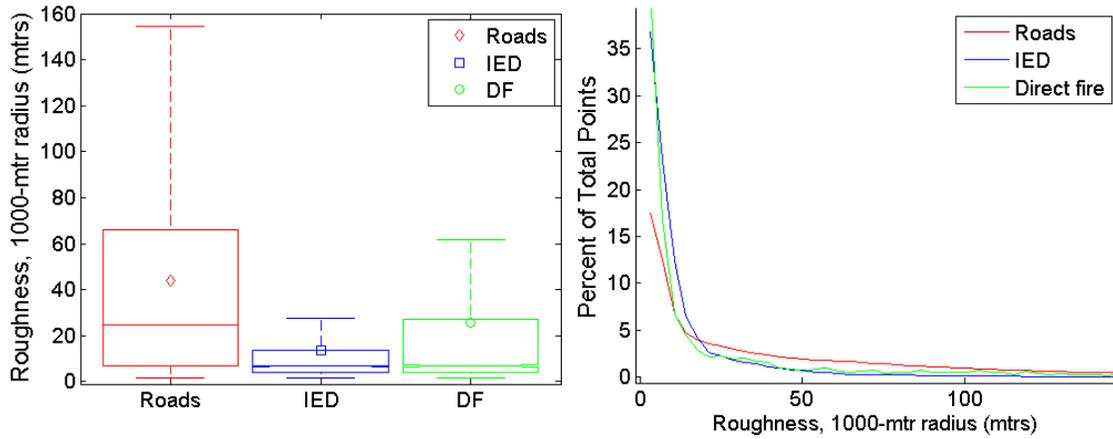
**Figure 84: Roughness at a radius of 50 meters.**



**Figure 85: Roughness at a radius of 100 meters.**

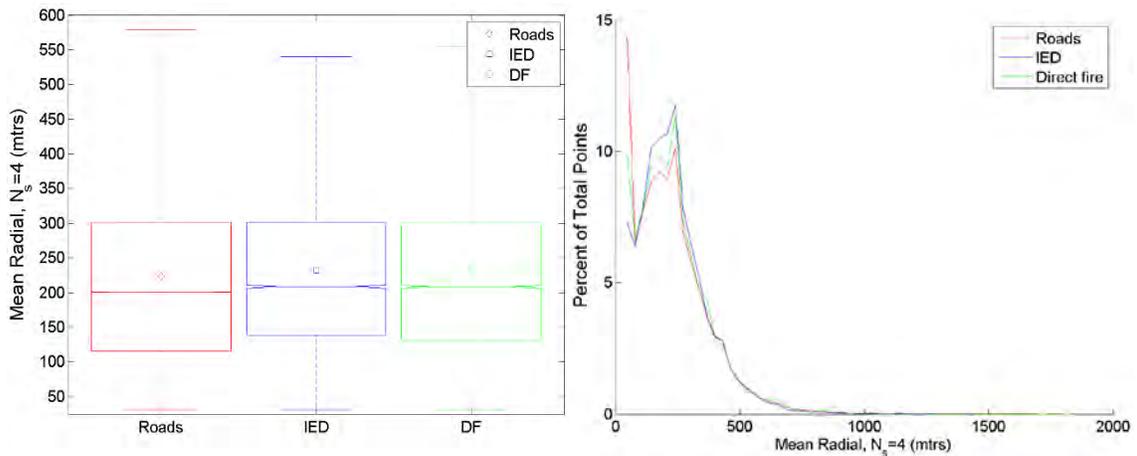


**Figure 86: Roughness at a radius of 500 meters.**

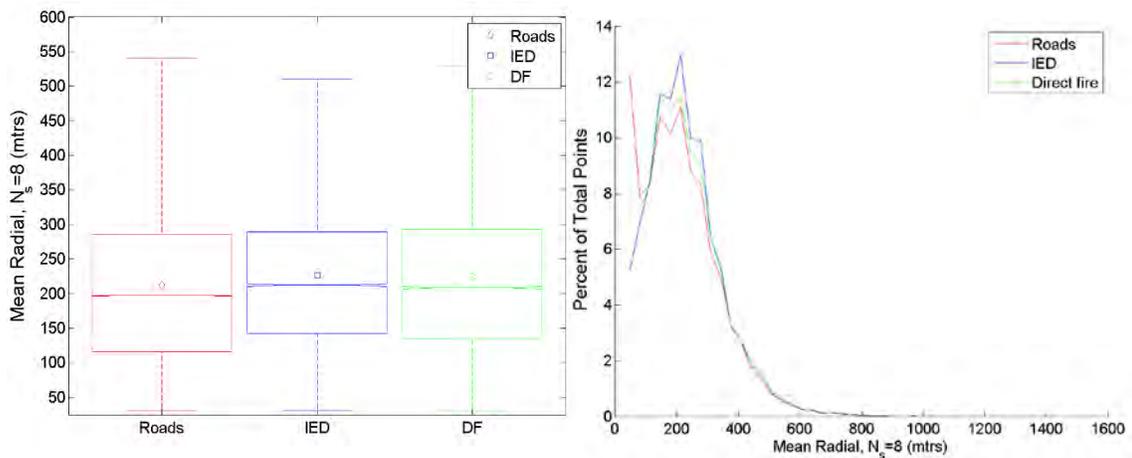


**Figure 87: Roughness at a radius of 1000 meters.**

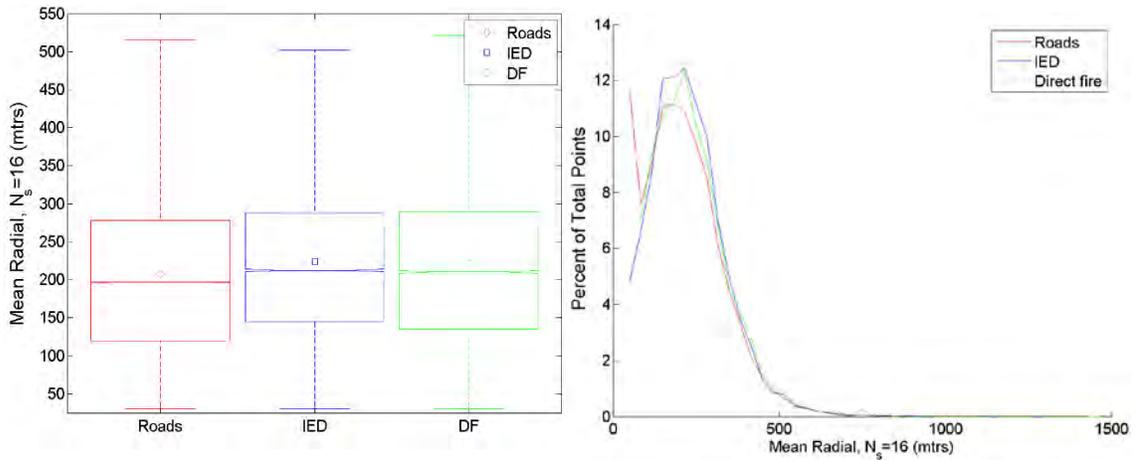
**Appendix B.12 Sparse Viewshed Mean Radial**



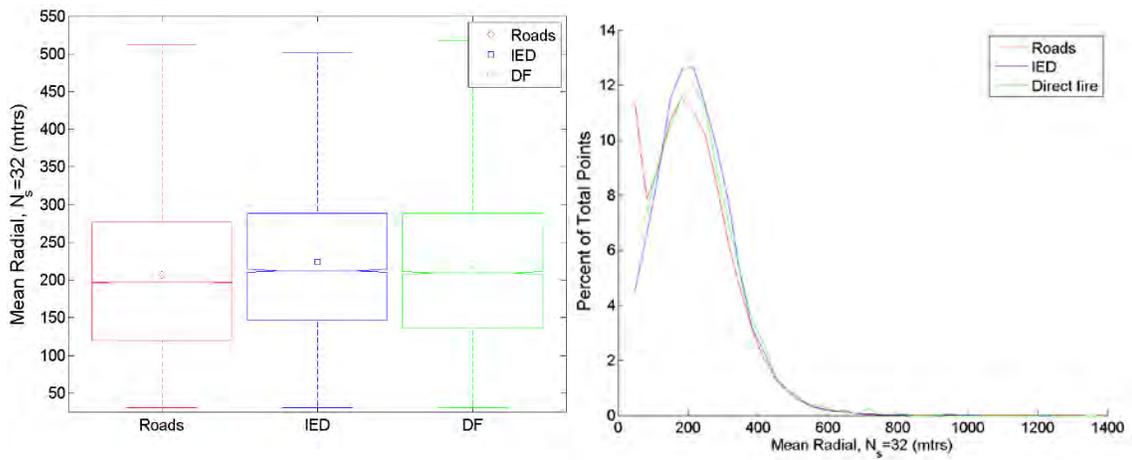
**Figure 88: Sparse viewshed mean radial ( $N_s = 4$ ).**



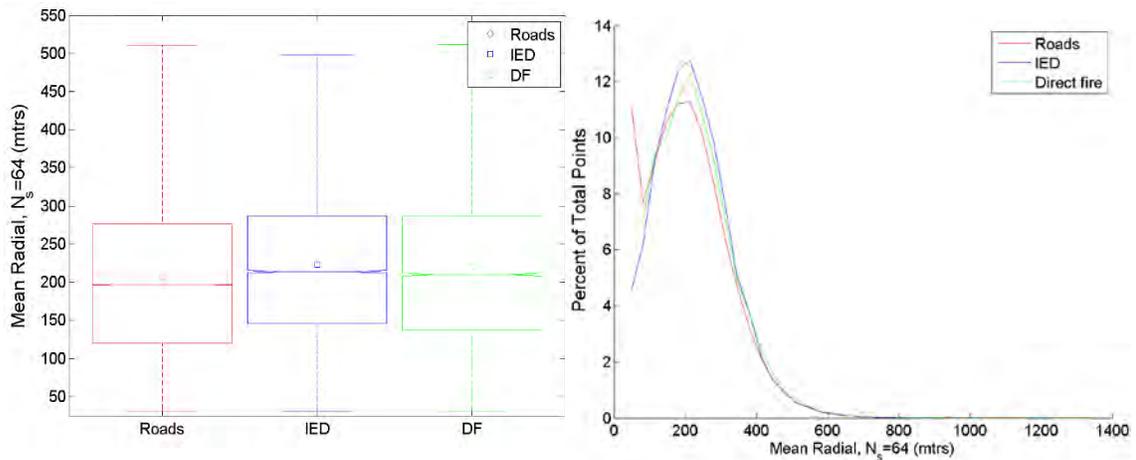
**Figure 89: Sparse viewshed mean radial ( $N_s = 8$ ).**



**Figure 90: Sparse viewshed mean radial ( $N_s = 16$ ).**

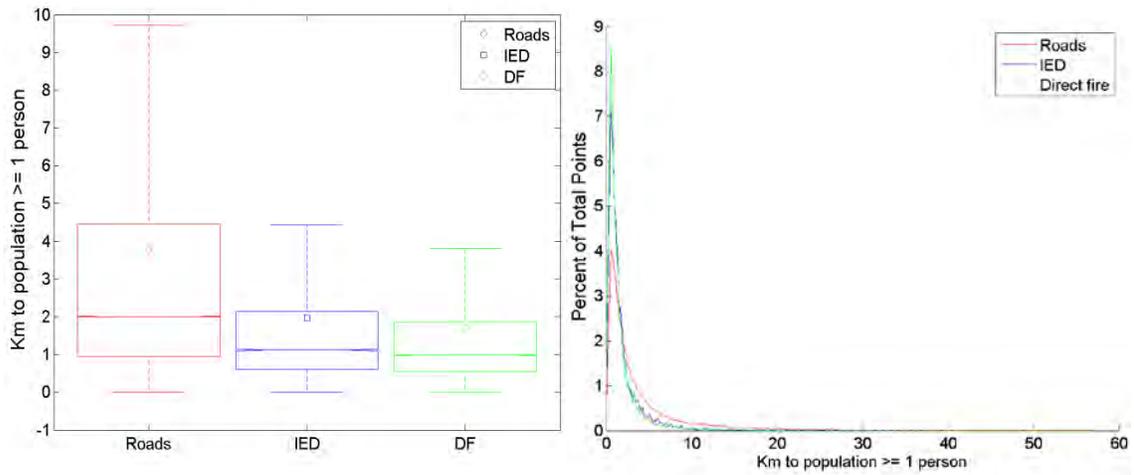


**Figure 91: Sparse viewshed mean radial ( $N_s = 32$ ).**

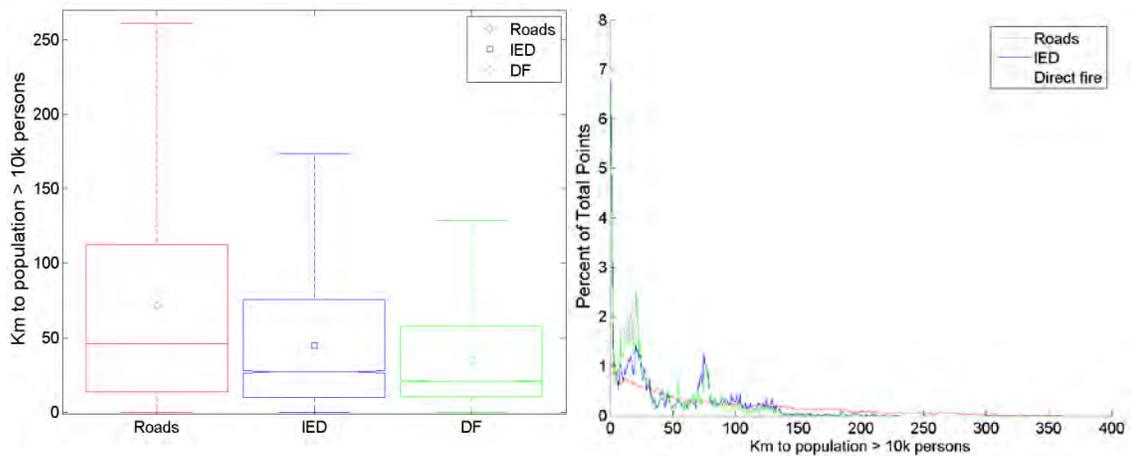


**Figure 92: Sparse viewshed mean radial ( $N_s = 64$ ).**

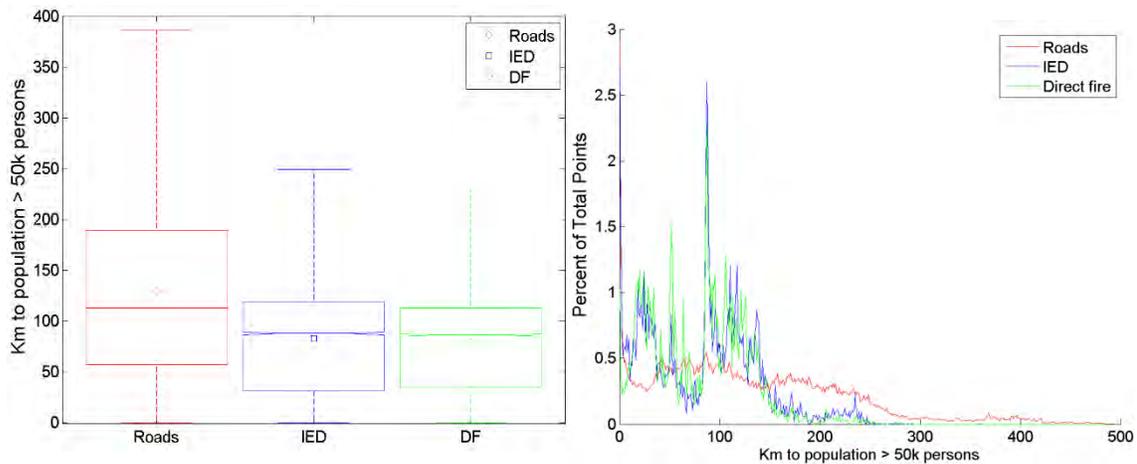
### Appendix B.13 Distance to Population Centers



**Figure 93: Distance to nearest population center with more than 1 person.**



**Figure 94: Distance to nearest population center with more than 10,000 people.**



**Figure 95: Distance to nearest population center with more than 50,000 people.**

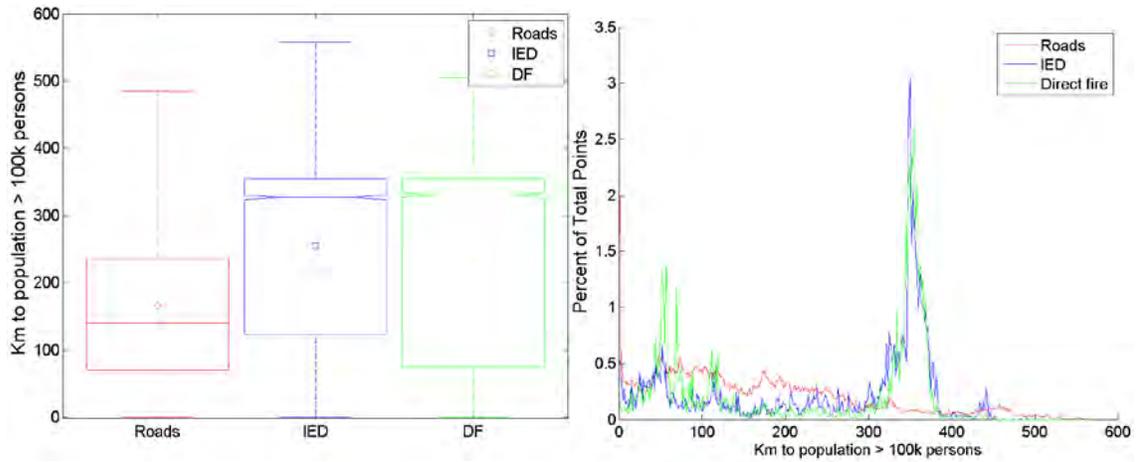


Figure 96: Distance to nearest population center with more than 100,000 people.

### Appendix B.14 Sparse Viewshed Planimetric Area

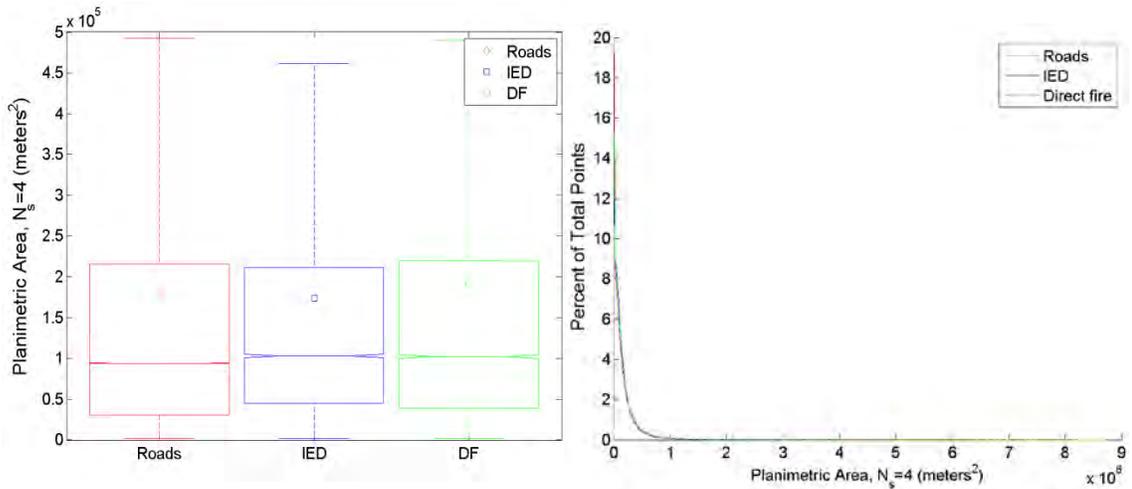


Figure 97: sparse viewshed planimetric area ( $N_s = 4$ ).

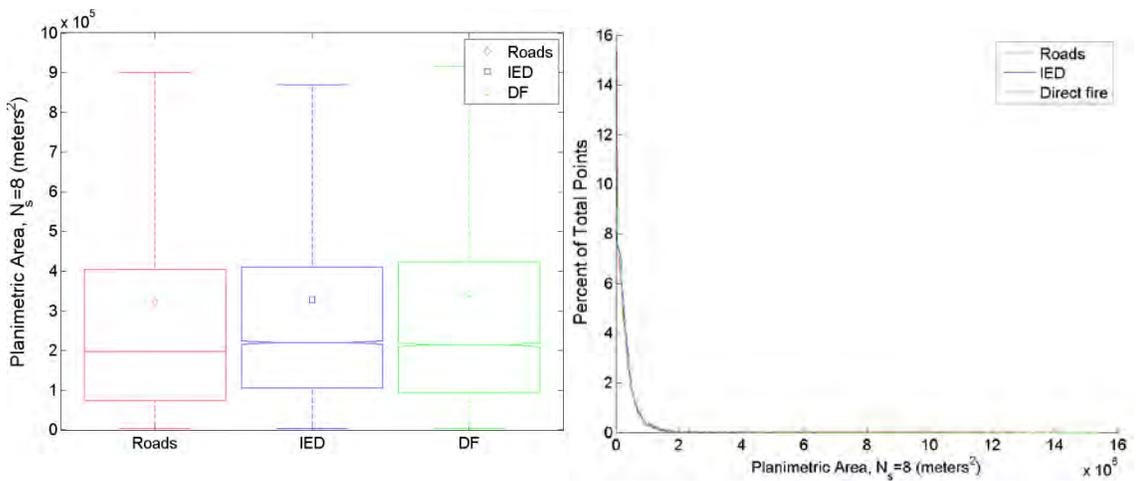
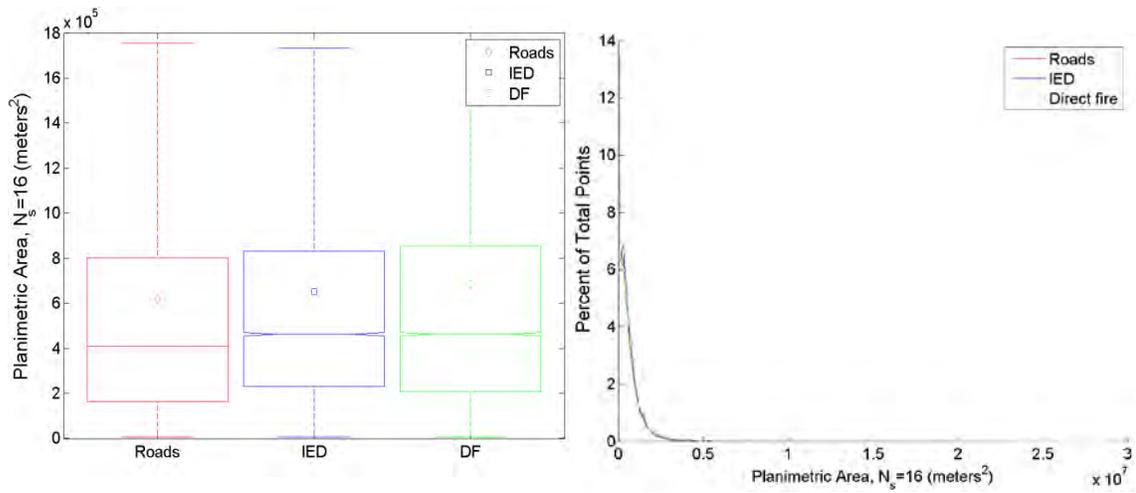
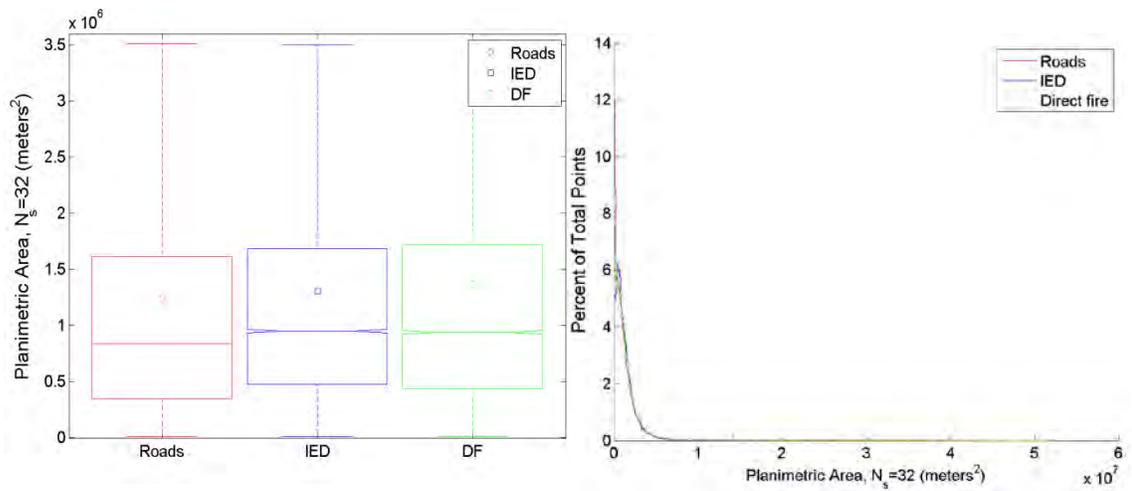


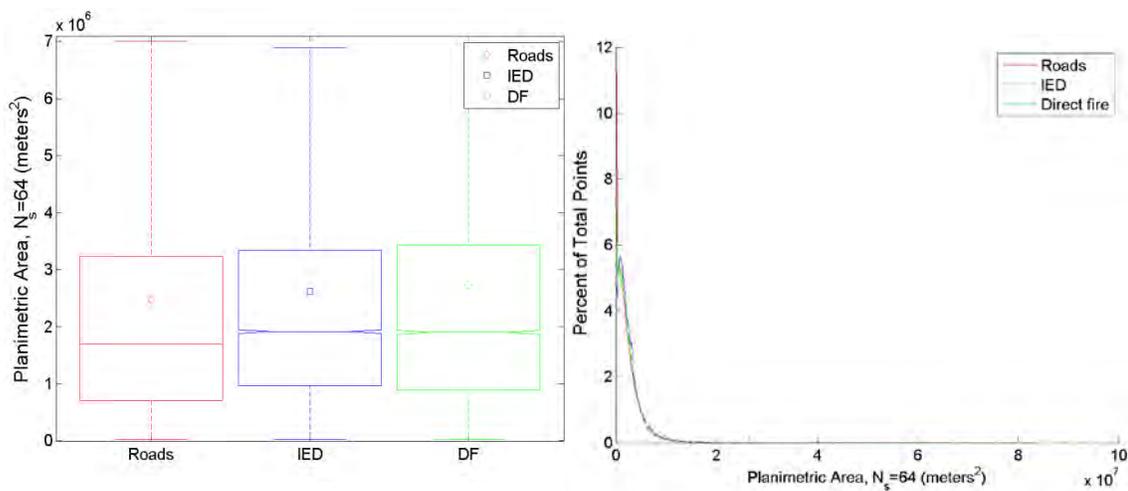
Figure 98: Sparse viewshed planimetric area ( $N_s = 8$ ).



**Figure 99: Sparse viewshed planimetric area ( $N_s = 16$ ).**



**Figure 100: Sparse viewshed planimetric area ( $N_s = 32$ ).**



**Figure 101: Sparse viewshed planimetric area ( $N_s = 64$ ).**

## Appendix B.15 Sparse Viewshed Rugosity

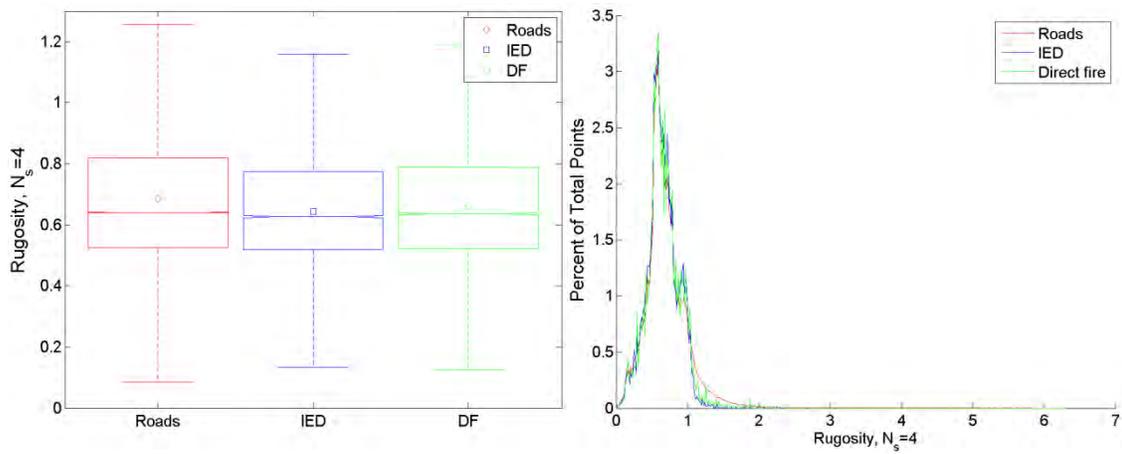


Figure 102: Sparse viewshed rugosity ( $N_s = 4$ ).

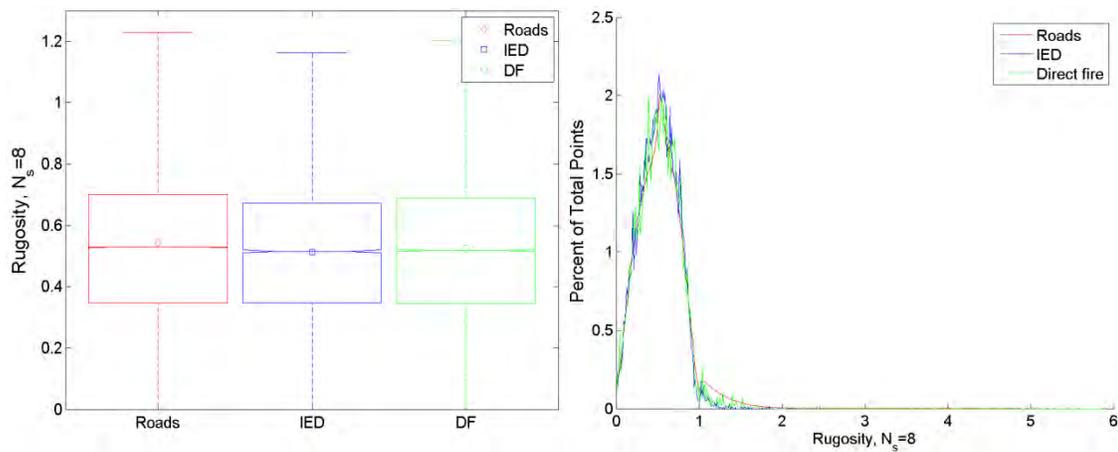


Figure 103: Sparse viewshed rugosity ( $N_s = 8$ ).

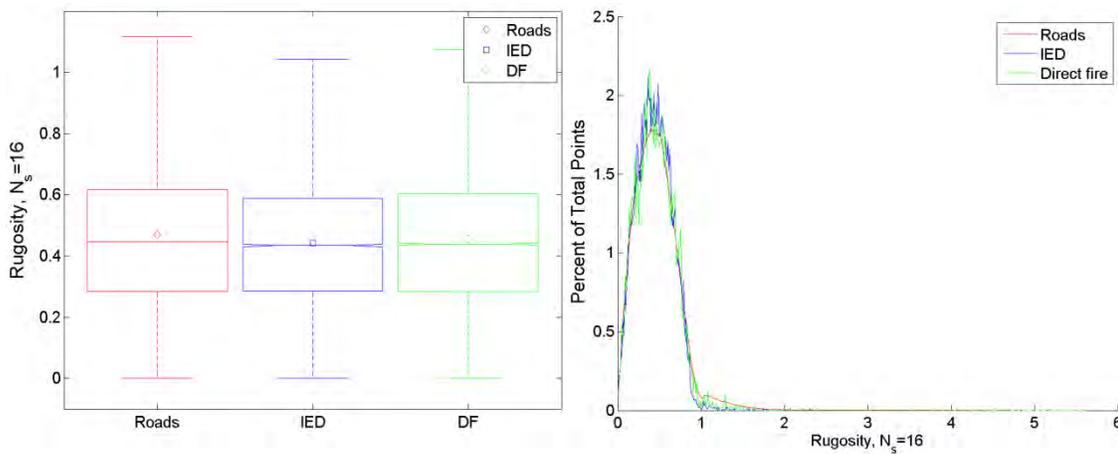
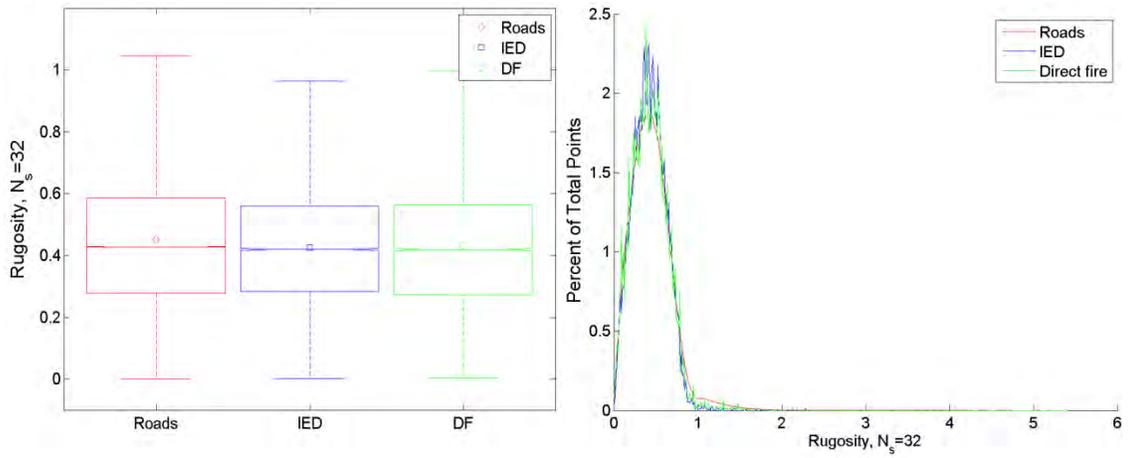
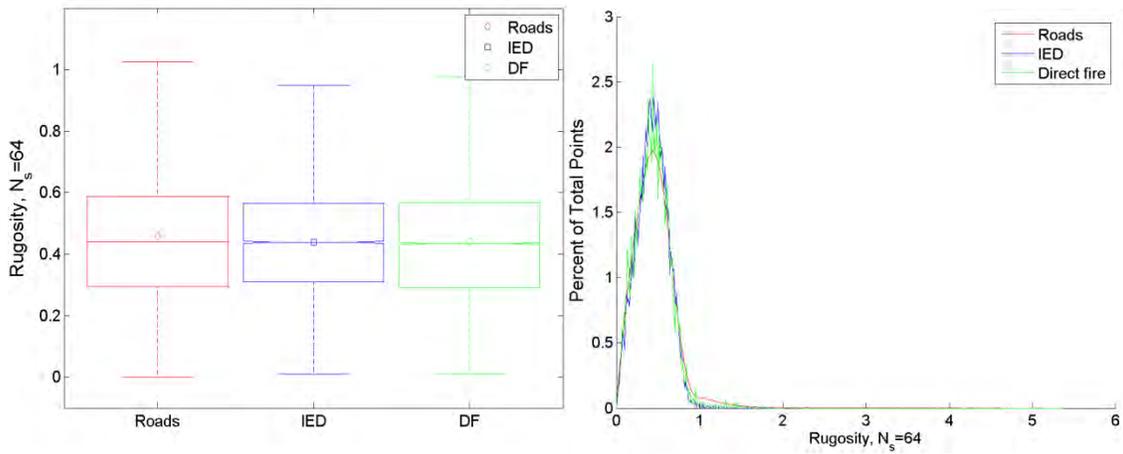


Figure 104: Sparse viewshed rugosity ( $N_s = 16$ ).

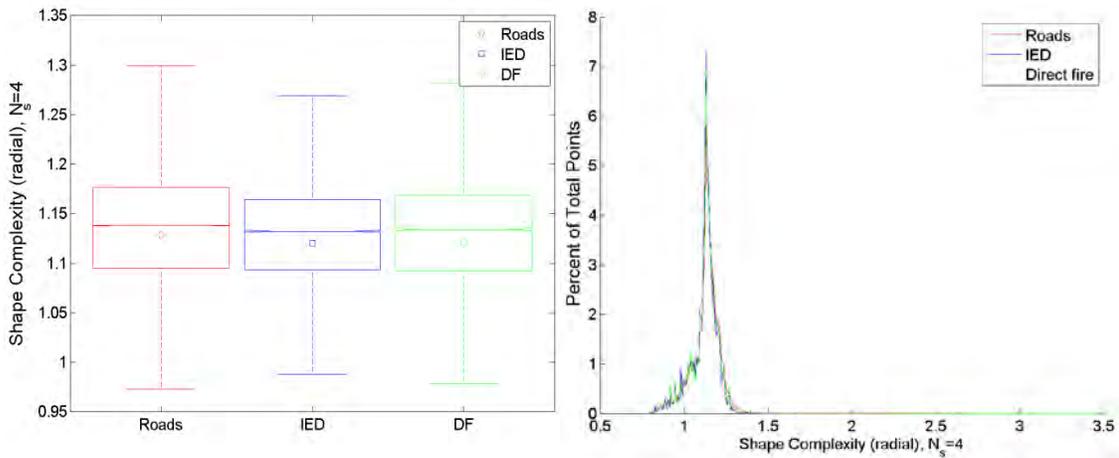


**Figure 105: Sparse viewshed rugosity ( $N_s = 32$ ).**



**Figure 106: Sparse viewshed rugosity ( $N_s = 64$ ).**

**Appendix B.16 Sparse Viewshed Shape Complexity Index**



**Figure 107: Sparse viewshed shape complexity index ( $N_s = 4$ ).**

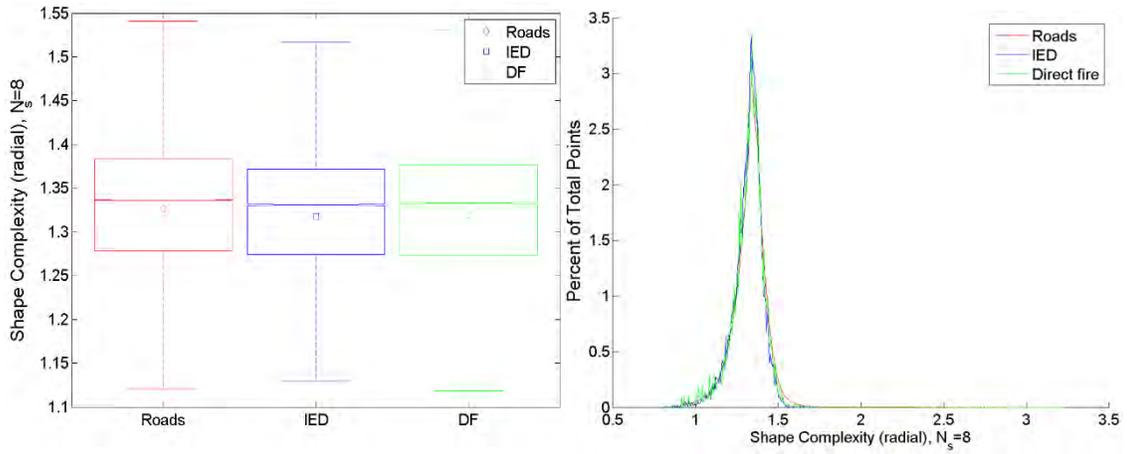


Figure 108: Sparse viewshed shape complexity index ( $N_s = 8$ ).

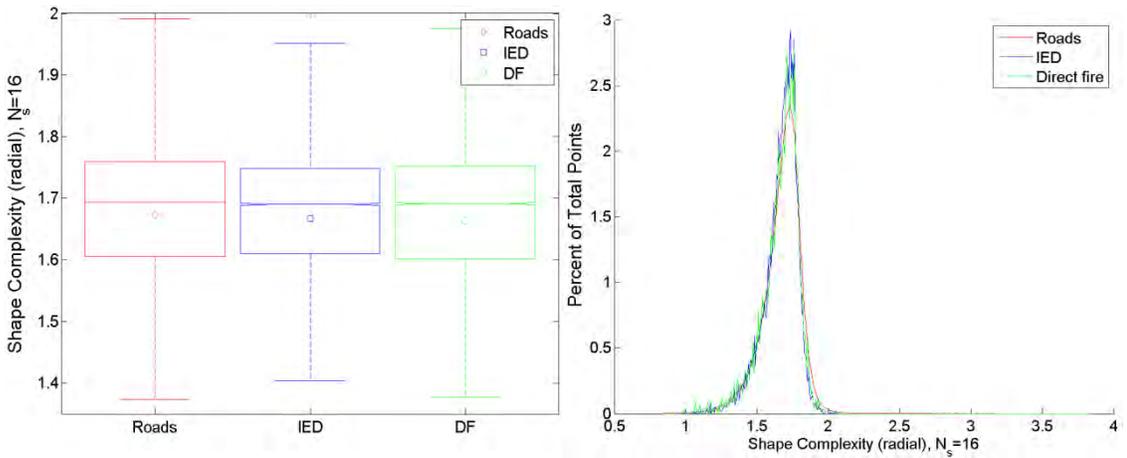


Figure 2 Sparse viewshed shape complexity index ( $N_s = 16$ ).

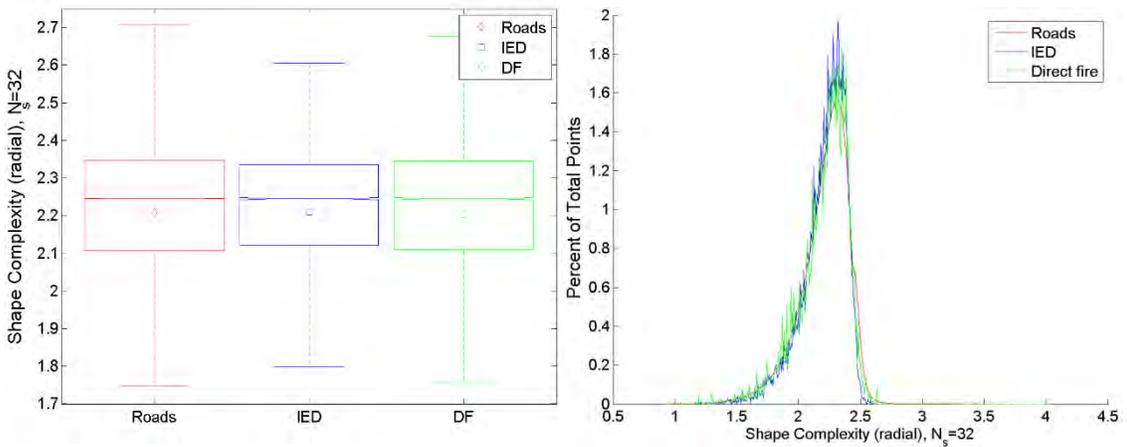
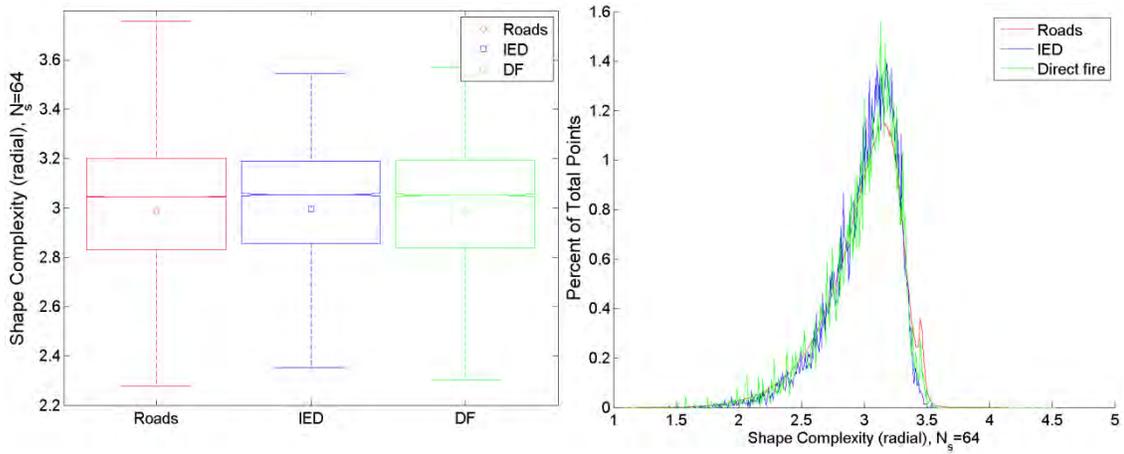


Figure 110: Sparse viewshed shape complexity index ( $N_s = 32$ ).



**Figure 111: Sparse viewshed shape complexity index ( $N_s = 64$ ).**

## Appendix C. Summary of Features

The following tables summarize key statistics of the conflict event and road datasets.

**Table 10: Statistics for Road Points**

Road Features	Mean	Std Dev.	Skewness	Kurtosis	Max	Min
Elevation	1500.567	894.5486	0.543728	2.420101	5059	245
Slope	12.29649	11.22895	1.307544	4.18051	80.8767	0
IW_convexity	50.58728	3.467145	0.27602	7.149377	100	30.52326
IW_texture	171.7688	31.52016	-0.47859	3.251345	268	0
Elv_rng50	17.84263	17.63961	2.342669	11.46453	314	0
Elv_rng100	30.36374	30.16361	2.123798	9.377894	486	0
Elv_rng350	90.69964	90.76243	1.63301	5.771874	884	5
Elv_rng500	120.7624	120.3921	1.524657	5.201717	1067	7
Elv_rng1000	204.5278	202.0049	1.388893	4.580955	1480	10
Rough_50	5.734752	5.809725	2.415465	12.15683	99.49794	0
Rough_100	8.43356	8.84612	2.235514	10.32687	141.1226	0
Rough_350	21.28197	23.7253	1.842704	7.022945	251.019	1.071471
Rough_500	27.38109	30.47396	1.731273	6.292839	282.6358	1.182403
Rough_1000	43.7797	48.01696	1.566073	5.328774	362.9134	1.575278

Visidx100-350	173.6378	94.9768	-0.13214	2.01938	360	0
Visidx_350	200.8412	100.623	-0.22429	2.080677	392	2
Visidx_500	336.1799	187.3552	0.108157	2.140409	820	2
Visidx_1000	820.063	557.0581	0.643432	2.798222	3172	2
SCID100-350	NaN	NaN	-0.75397	2.926884	5.352372	0.282095
SCID_350	3.817166	1.188141	-0.87124	3.083763	5.585192	0.398942
SCID_500	4.902237	1.649369	-0.59101	2.726625	8.077966	0.398942
SCID_1000	7.522361	2.944933	-0.13269	2.383551	15.88772	0.398942
Short_rad_4	87.85215	68.96896	1.87485	8.553442	1081.062	30.88748
Long_rad_4	418.7539	325.2703	1.950796	10.20155	2934.311	30.88748
Mean_rad_4	222.8605	150.8685	1.300956	6.354747	1860.971	30.88748
Local_op_4	0.145411	0.178123	3.137949	21.55073	4.192637	0
Planimtrc_4	181005.4	299234.4	6.403355	78.60991	8741126	1908.073
Rugosity_4	0.687045	0.291456	1.559916	11.25873	6.303492	0
SCIF_4	1.128786	0.089746	-0.1246	8.936473	3.486189	0.78235
Short_rad_8	66.55558	46.47777	1.73964	7.199426	648.6371	30.88748
Long_rad_8	492.7444	341.9749	1.766337	9.106218	2934.311	30.88748
Mean_rad_8	211.2556	127.8879	1.020124	5.198145	1579.122	30.88748
Road Features	Mean	Std Dev.	Skewness	Kurtosis	Max	Min
Local_op_8	0.147636	0.180521	3.202525	22.41808	4.528534	0.000458
Planimtrc_8	321310.9	448361.5	5.669126	69.59668	15413320	3569.677
Rugosity_8	0.545856	0.284494	1.308529	8.992094	5.920185	0
SCIF_8	1.327031	0.099617	-0.36451	6.428349	3.232216	0.804348
Short_rad_16	56.8759	36.33826	1.705347	6.641047	494.1997	30.88748
Long_rad_16	568.9883	367.5516	1.600696	8.075665	2934.311	30.88748
Mean_rad_16	207.2524	119.6326	0.887791	4.712483	1498.043	30.88748
Local_op_16	0.14788	0.181114	3.247771	22.90394	4.492111	0.001172
Planimtrc_16	620615.5	795392.7	5.303083	63.93605	28862852	6769.165
Rugosity_16	0.471404	0.265345	1.471321	9.785589	5.613903	0
SCIF_16	1.672743	0.134303	-0.84879	5.229183	3.840335	0.837967
Short_rad_32	51.82218	30.52004	1.712476	6.511612	463.3122	30.88748
Long_rad_32	646.2164	399.4636	1.455157	7.140774	2934.311	30.88748
Mean_rad_32	206.5036	117.1211	0.833915	4.517314	1402.485	30.88748
Local_op_32	0.148107	0.181391	3.257947	23.0757	4.51437	0.001517
Planimtrc_32	1238309	1527504	5.130433	61.27438	51675859	13303.11
Rugosity_32	0.451506	0.249126	1.556002	10.71974	5.414795	0
SCIF_32	2.209487	0.198408	-1.04606	4.735505	4.030713	0.919574
Short_rad_64	49.37544	27.3026	1.670685	6.23432	339.7623	30.88748
Long_rad_64	716.3153	432.0182	1.331562	6.378148	2934.311	30.88748

Mean_rad_64	206.3157	116.2885	0.812963	4.443949	1358.566	30.88748
Local_op_64	0.148148	0.181436	3.259188	23.07553	4.519934	0.001954
Planimtrc_64	2473714	3004681	5.068907	60.3886	95729354	26481.33
Rugosity_64	0.458616	0.241175	1.545595	10.99351	5.373364	0.000588
SCIF_64	2.988666	0.29893	-1.03653	4.406598	4.528361	1.07449
Dist_pop_1	3866.659	5055.309	3.280635	18.8791	56877.3	0.799611
Dist_pop_1k	12018.52	22347.75	2.873837	11.86605	165470.1	0.799611
Dist_pop_10k	72976.86	73195.13	1.194508	3.810034	360925.8	0.799611
Dist_pop_50k	131591.1	94707.11	0.848839	3.589145	494683.7	0.799611
Dist_pop100k	165769.2	121818.8	0.882078	3.216328	566862.7	0.799611
Min_CEA	19.97321	39.10452	12.92167	393.3227	3394	1
Max_CEA	251.9178	416.94	6.272908	48.01058	4297	2
Med_CEA	83.38715	132.3518	6.841458	62.75816	3431	1
RtVisMin_1k	0.014877	0.021871	11.84692	200.5709	0.849057	0.000453
RtVisMax_1k	0.603916	0.184814	-0.32498	2.982343	1	0.000582
RtVisMed_1k	0.182582	0.102707	0.729848	3.292173	0.88	0.000582
RtVisMin_500	0.03024	0.03052	9.437664	136.5116	0.977273	0.001196
RtVisMax500	0.726549	0.182017	-0.84383	3.937267	1	0.001848
Road Features	Mean	Std Dev.	Skewness	Kurtosis	Max	Min
RtVisMed500	0.255945	0.133704	0.56446	2.837612	0.977273	0.001848
RtVisMin_250	0.061097	0.041276	6.2992	72.69977	1	0.003472
RtVisMax250	0.824756	0.170423	-1.43312	5.704653	1	0.006098
RtVisMed250	0.339879	0.161213	0.38499	2.535758	1	0.006098
RtVisMin_100	0.15594	0.068637	2.642588	18.16312	1	0.014493
RtVisMax100	0.924422	0.139425	-2.53312	10.72765	1	0.026316
RtVisMed100	0.480261	0.189432	0.076296	2.37194	1	0.020833

**Table 11: Statistics for IED Events**

IED Features	Mean	Std Dev.	Skewness	Kurtosis	Max	Min
Elevation	1189.741	516.1805	1.195342	3.562144	4231	284
Slope	5.960902	6.007035	2.954775	14.20707	57.00436	0
IW_convexity	50.50021	2.387105	0.214614	4.469146	66.27907	38.0814
IW_texture	185.0368	23.28743	-0.40387	3.523254	261	69
Elv_rng50	8.860173	7.52035	4.608615	37.40607	131	1
Elv_rng100	13.89868	12.3034	4.693826	36.62869	204	2
Elv_rng350	34.13291	36.99049	4.682419	34.68556	547	7
Elv_rng500	44.05115	49.79456	4.384505	30.32021	663	9

Elv_rng1000	73.97255	86.47978	3.587551	21.02291	1056	12
Rough_50	2.810664	2.463042	4.6916	37.65946	41.45635	0.426401
Rough_100	3.711215	3.581271	4.951779	40.19897	63.20649	0.700662
Rough_350	7.02624	9.215696	5.019333	38.89898	154.6467	1.366325
Rough_500	8.655027	12.0059	4.845296	36.23133	187.591	1.520224
Rough_1000	13.58592	19.17193	4.150777	27.41496	245.6013	1.839131
Visidx100-350	185.6878	87.87364	-0.18106	2.140015	360	0
Visidx_350	214.4752	91.72991	-0.26132	2.227028	392	4
Visidx_500	351.4775	180.5928	0.120343	2.1702	806	4
Visidx_1000	811.8062	567.0791	0.73444	2.8867	2943	4
SCID100-350	NaN	NaN	-0.77079	3.112951	5.352372	0.282095
SCID_350	4.006368	1.008209	-0.8829	3.44103	5.585192	0.56419
SCID_500	5.069513	1.506619	-0.51472	2.722967	8.00871	0.56419
SCID_1000	7.480773	2.939417	0.00766	2.292969	15.30348	0.56419
Short_rad_4	93.87285	70.00596	1.909031	9.484288	957.5119	30.88748
Long_rad_4	424.1586	280.8768	1.994782	11.18061	2934.311	30.88748
Mean_rad_4	232.2845	134.2474	1.205886	5.800196	1258.665	30.88748
Local_op_4	0.059758	0.080886	5.322349	52.79569	1.359774	0
IED Features	Mean	Std Dev.	Skewness	Kurtosis	Max	Min
Planimtrc_4	174157.1	244022.5	5.429662	55.67555	4540737	1908.073
Rugosity_4	0.642722	0.216575	0.354551	4.804443	2.282296	0
SCIF_4	1.120155	0.077516	-1.01414	4.691009	1.410832	0.798868
Short_rad_8	72.11338	48.61465	1.698557	7.617441	555.9746	30.88748
Long_rad_8	506.9798	299.1997	1.835823	10.03663	2934.311	30.88748
Mean_rad_8	226.2865	117.4495	0.991791	5.003164	1077.201	30.88748
Local_op_8	0.060769	0.082075	5.63435	61.95781	1.663295	0.001457
Planimtrc_8	328125.1	389897.4	4.877909	56.08882	9702172	3569.677
Rugosity_8	0.512435	0.227316	0.309122	4.177166	2.425573	0
SCIF_8	1.317845	0.086136	-0.92441	5.105927	1.653756	0.852437
Short_rad_16	61.38465	38.18321	1.594601	6.427533	339.7623	30.88748
Long_rad_16	590.6188	322.0858	1.664434	8.798545	2934.311	30.88748
Mean_rad_16	224.1002	110.6749	0.908346	4.969063	1104.227	30.88748
Local_op_16	0.061132	0.082036	5.615595	61.46527	1.675436	0.002342
Planimtrc_16	651005.1	727405.5	4.99593	55.56582	14577227	6769.165
Rugosity_16	0.441361	0.209275	0.46621	4.412503	2.308768	0
SCIF_16	1.667039	0.118685	-1.1348	5.096488	2.062202	1.042637
Short_rad_32	55.92736	32.65441	1.610797	6.399642	277.9873	30.88748
Long_rad_32	670.5452	348.6348	1.519722	7.758947	2934.311	30.88748
Mean_rad_32	223.7752	108.2676	0.856616	4.84667	1008.669	30.88748

Local_op_32	0.061157	0.082141	5.712708	64.01138	1.691624	0.002632
Planimtrc_32	1307310	1412254	5.038128	58.74511	30304717	13303.11
Rugosity_32	0.4264	0.195578	0.530602	5.117283	2.28104	0.002348
SCIF_32	2.208976	0.176938	-1.21226	4.989893	2.605695	1.162805
Short_rad_64	53.34624	29.65709	1.571263	6.196626	277.9873	30.88748
Long_rad_64	744.8215	383.6695	1.462154	7.206687	2934.311	30.88748
Mean_rad_64	223.5386	107.386	0.831561	4.755836	1006.256	30.88748
Local_op_64	0.061176	0.082221	5.720145	64.28753	1.711353	0.002826
Planimtrc_64	2616312	2791769	4.85154	53.25151	57114787	26481.33
Rugosity_64	0.438592	0.18791	0.462449	5.409028	2.278868	0.009078
SCIF_64	2.997456	0.264928	-1.15723	4.735597	3.545446	1.447101
Dist_pop_1	1970.09	3025.814	6.955584	82.21161	62357.3	8.145166
Dist_pop_1k	4775.628	12941.26	5.415877	41.72045	192750.7	8.145166
Dist_pop_10k	44790.97	42667.25	1.00429	3.476587	230029.5	8.145166
Dist_pop_50k	82942.68	55757.04	0.484679	2.966735	360583	38.99919
Dist_pop100k	255149.6	129590.4	-0.71614	1.960563	557783.2	139.1335
Min_CEA	15.98684	28.32625	7.408402	102.6326	769	1
Max_CEA	210.9615	246.4451	5.496054	49.01561	4297	1
IED Features	Mean	Std Dev.	Skewness	Kurtosis	Max	Min
Med_CEA	65.95197	76.60394	4.774067	40.36239	1357	1
RtVisMin_1k	0.126199	0.261115	2.616346	8.707304	1	0.000518
RtVisMax_1k	0.655776	0.216014	-0.06175	2.469475	1	0.004367
RtVisMed_1k	0.251898	0.245152	2.030872	6.560558	1	0.003802
RtVisMin_500	0.236852	0.340464	1.509931	3.722207	1	0.001678
RtVisMax500	0.794972	0.195926	-0.81358	3.290479	1	0.007692
RtVisMed500	0.379401	0.300473	1.125179	3.029845	1	0.006289
RtVisMin_250	0.38808	0.388385	0.718764	1.822657	1	0.006803
RtVisMax250	0.894392	0.159392	-1.80274	6.542154	1	0.032258
RtVisMed250	0.52769	0.326781	0.409904	1.690512	1	0.016667
RtVisMin_100	0.696318	0.371742	-0.55682	1.525536	1	0.023256
RtVisMax100	0.974347	0.089017	-4.57539	28.23482	1	0.083333
RtVisMed100	0.792028	0.283912	-0.90026	2.263288	1	0.033333

**Table 12: Statistics for Direct Fire Events**

DF Features	Mean	Std Dev.	Skewness	Kurtosis	Max	Min
Elevation	1239.97	557.3591	1.172477	3.269075	4151	275
Slope	8.178474	9.209225	2.202922	7.689192	64.3739	0

IW_convexity	50.78444	2.767828	0.341509	4.972919	69.18605	35.75581
IW_texture	179.9506	28.21021	-0.62173	4.025256	255	47
Elv_rng50	11.94124	13.05931	2.970686	13.34795	116	1
Elv_rng100	19.69348	22.53343	2.863105	12.18684	188	2
Elv_rng350	56.02511	76.16447	2.742375	11.17331	571	7
Elv_rng500	74.25974	103.065	2.605197	10.1242	769	8
Elv_rng1000	126.0677	171.9339	2.249472	7.89859	1351	12
Rough_50	3.811751	4.2688	3.014123	13.6756	40.07682	0.389249
Rough_100	5.365653	6.525197	2.941597	12.74078	55.62539	0.552669
Rough_350	12.44763	19.00739	2.874951	12.27379	158.8052	1.306791
Rough_500	15.99117	25.3049	2.824251	11.9456	218.367	1.480848
Rough_1000	25.8792	40.7347	2.487302	9.286029	337.7308	1.844114
Visidx100-350	180.6762	91.50847	-0.18309	2.100827	360	0
Visidx_350	208.9338	96.00504	-0.26845	2.184552	392	4
Visidx_500	347.2455	185.3923	0.08659	2.167492	816	4
Visidx_1000	845.3988	585.3677	0.632782	2.67354	3152	4
SCID100-350	NaN	NaN	-0.7774	3.048465	5.352372	0.282095
SCID_350	3.927529	1.095906	-0.93006	3.422014	5.585192	0.56419
DF Features	Mean	Std Dev.	Skewness	Kurtosis	Max	Min
SCID_500	5.008353	1.596708	-0.61598	2.859334	8.058239	0.56419
SCID_1000	7.61272	3.053154	-0.12254	2.32551	15.83756	0.56419
Short_rad_4	91.54477	70.65889	1.916574	8.422847	679.5246	30.88748
Long_rad_4	432.9138	323.5822	2.21322	11.68228	2934.311	30.88748
Mean_rad_4	232.4906	147.4341	1.328615	6.070495	1413.102	30.88748
Local_op_4	0.084963	0.12842	4.37544	32.41407	1.853502	0.000506
Planimtrc_4	191101.9	311498.5	5.611846	54.29834	7054350	1908.073
Rugosity_4	0.661678	0.246279	1.091447	8.373619	3.08116	0
SCIF_4	1.12125	0.084467	-0.78846	4.821488	1.659523	0.793605
Short_rad_8	70.05427	48.26989	1.768018	7.634148	494.1997	30.88748
Long_rad_8	515.6846	339.143	1.95246	9.925631	2934.311	30.88748
Mean_rad_8	223.782	124.5285	0.973981	4.845281	1343.605	30.88748
Local_op_8	0.086389	0.131173	4.532036	34.4163	1.82922	0.001515
Planimtrc_8	342874.2	443987.1	5.006058	59.0276	11788326	3569.677
Rugosity_8	0.524474	0.254098	0.883732	6.585551	2.599429	0
SCIF_8	1.319373	0.095203	-0.93913	5.47292	1.738768	0.818671
Short_rad_16	59.66794	37.1889	1.564908	6.062438	339.7623	30.88748
Long_rad_16	601.1742	370.5594	1.822419	9.073959	2934.311	30.88748
Mean_rad_16	221.924	118.4034	0.940168	4.981722	1270.248	30.88748
Local_op_16	0.086577	0.131033	4.520773	34.05139	1.865643	0.002342

Planimtrc_16	685665.5	852809.9	4.978382	52.65761	20257471	6769.165
Rugosity_16	0.45431	0.237426	1.042668	7.099692	2.523818	0
SCIF_16	1.663809	0.132212	-1.23675	5.690666	2.191069	0.926427
Short_rad_32	54.86252	32.02801	1.574108	6.080916	339.7623	30.88748
Long_rad_32	682.435	402.3281	1.695075	8.127751	2934.311	30.88748
Mean_rad_32	221.0894	115.7138	0.897515	4.936341	1230.673	30.88748
Local_op_32	0.086496	0.130627	4.503245	33.54321	1.795833	0.002536
Planimtrc_32	1367396	1639130	4.997501	55.40848	38216401	13303.11
Rugosity_32	0.431424	0.220303	1.105502	7.612239	2.183659	0.003546
SCIF_32	2.203323	0.197384	-1.22504	5.221941	2.72501	1.044676
Short_rad_64	52.61044	29.49649	1.55894	6.011126	339.7623	30.88748
Long_rad_64	759.4668	438.0354	1.558668	7.210818	2934.311	30.88748
Mean_rad_64	220.8066	115.0348	0.875594	4.827426	1207.99	30.88748
Local_op_64	0.086506	0.130681	4.512222	33.68811	1.819103	0.002668
Planimtrc_64	2730956	3228337	4.916075	54.12197	73489390	26481.33
Rugosity_64	0.4401	0.212687	1.071789	7.815886	2.198489	0.009
SCIF_64	2.985627	0.293502	-1.19028	4.828701	3.570024	1.327981
Dist_pop_1	1700.359	2487.212	6.552423	79.11162	60470.71	6.711227
DF Features	Mean	Std Dev.	Skewness	Kurtosis	Max	Min
Dist_pop_1k	2617.723	7444.843	8.182352	87.77123	148677.1	6.711227
Dist_pop_10k	35394.62	35869.29	1.364658	4.598799	250508.8	20.60933
Dist_pop_50k	79040.08	47634.61	0.439541	3.084528	290180.5	45.84043
Dist_pop100k	235779.4	137015.1	-0.41086	1.371092	504834.1	214.2961
Min_CEA	15.15311	21.41143	4.59485	40.23002	367	1
Max_CEA	192.7419	168.7335	5.057386	57.53948	3849	1
Med_CEA	59.22232	55.58704	3.862144	40.28346	1166	1
RtVisMin_1k	0.141893	0.260763	2.488321	8.165724	1	0.000726
RtVisMax_1k	0.674881	0.214215	-0.22523	2.539017	1	0.007463
RtVisMed_1k	0.264446	0.247878	1.862538	5.937132	1	0.003846
RtVisMin_500	0.277799	0.340138	1.282869	3.182472	1	0.002632
RtVisMax500	0.818447	0.189101	-1.05744	3.910993	1	0.018519
RtVisMed500	0.41733	0.304879	0.860691	2.505724	1	0.011494
RtVisMin_250	0.455908	0.379136	0.4267	1.576591	1	0.00885
RtVisMax250	0.919258	0.141838	-2.28398	9.050088	1	0.045455
RtVisMed250	0.589679	0.319247	0.081273	1.573079	1	0.015152
RtVisMin_100	0.750182	0.339886	-0.87639	2.106256	1	0.015625
RtVisMax100	0.983039	0.07348	-6.07345	48.09473	1	0.058824
RtVisMed100	0.838916	0.254278	-1.33644	3.416514	1	0.02381

## VOCABULARY GLOSSARY

Symbol	Term	Definition
	Conflict Event	A lethal or non-lethal encounter between forces where the encounter is planned and executed in order to obtain some specific outcome such as casualties.
	Tactic	The movement and arrangement of actors as well as resources involved in relation to geography and opposing forces.
E	Emplacement Location	The location where a Conflict Event occurs.
M	Monitor Location	The location used for supervision and early warning for a Conflict Event. This location typically has good visibility of terrain along the approaches to the Emplacement site.
C	Control Location	The location where a conflict event's execution is initiated and will typically have good visibility of the Emplacement site and adjacent terrain.
H	Halo	The annular area centered on the Emplacement site that can be used to perform Monitor and Control functions.
MECH	Monitor, Emplacement, Control, Halo	Monitoring of victim movements, Emplacement of device or attack, and Control of the device or attack within a Halo. An analytical abstraction to model the locational relationships between victims and attackers in asymmetric Conflict Events.
	Pattern Shift	The effect within the evolution of tactics that as locations change slightly, attack parameters must be shifted to match old patterns to new locations. When these adjustments are made and a new successful attack is made, the pattern grows or changes.
	Visibility-based Analysis	An analysis that focuses on a class of features that attempts to use human factors and limitations to limit areas under consideration.

	Geomorphometry Analysis	An analysis that focuses on a class of features that attempts to use topology and elevation data to limit areas under consideration.
	Social/Cultural Analysis	An analysis that focuses on a class of features that attempts to use aspects of site selection separate from the land itself and associated to the interrelations of people to limit areas under consideration.
IED	Improvised Explosive Device	A class of attack utilizing a homemade bomb using any materials conveniently accessible. IEDs are typically deployed in unconventional configurations that attempt to maximize concealment and lethality.
DF	Direct Fire	A class of attack utilizing weapons pointed directly at a target. These attacks are typically performed using rifles and pistols, but may include any type of weapon that can be pointed directly at the target. Notably, artillery and mortars are not direct fire weapons.
LOS	Line of Sight	The intervisibility between two points such that the points are visible to each other.
DEM	Digital Elevation Model	The representation of a terrain's surface created from topological data.
	Viewshed	The collection of points visible from a specific location using line of sight.
	Feature Extraction	The process of reducing factors that describe a data set.
	Kill Zone	The area of a conflict event that has a high concentration of fatalities.
	Escape Adjacency	A position that is within the line of sight to a conflict event, but also directly adjacent to a position that is not in the line of sight to the same conflict event.
○	Hadamard Product	A matrix that is the result of two equal dimensioned matrices where each position in the resultant matrix is the multiplicative product of the same position in each of the original two matrices. The result is of the same dimensions.
SVM	Support Vector	A non-parametric, supervised learning method used to

	Machine	perform binary classification.
DA	Discriminant Analysis	A statistical analysis used to predict a categorical dependent variable by one or more continuous or binary independent variables.
kNN	k Nearest Neighbor	A non-parametric method of classifying a sample or observation with an unknown classification.
KW	Kruskal-Wallis Test	A non-parametric statistical test designed to assess if the measurements for two or more classes come from the same population.
PCA	Unsupervised Principal Component Analysis	An unsupervised method of converting a set of possibly correlated features into set of linearly uncorrelated variables.
STP	Supervised Stepwise Function Selection	A method of reducing dimensionality that iteratively adds individual features, typically based on some statistical measure such as the F-test.
NDR	No Dimension Reduction	An experiment setting where there is no processing within the feature reduction stage of classifier training.
RBF	Gaussian Radical Basis Function	A real-valued function whose value depends only on the distance from some other point.
LDA	Linear Discriminant Analysis	A classification algorithm that finds a projection that maximizes between-class variance and minimizes within-class variance.
	Percent Error	The percent of all classifications that are not correct with respect to the MECH Classification Algorithm.
	Event Classification Error	The percentage of misclassified events out of the total number of classifications with respect to the MECH Classification Algorithm.
si'	Z-Score	A standard score that represents how far from the mean a sample is using the dispersion of that data.
penter	p-value enter	A threshold for the stepwise feature selection where

	threshold	hypotheses are tested with each feature and any p-value result that is above this threshold is added into statistically significant features.
	Balanced Class	A case of machine learning where the number of events and non-events are equal in training and test sets.
	Ensemble-Based Classifier	Classification scheme where output of multiple individual classifiers are combined according to some algorithm or rule.
COR	Blind Expert	An automated subset selection method that uses feature correlation as an identifier for event features.
	Majority Vote Rule	An Ensemble-based classifier that takes multiple individual classifiers and combines results by adding features that are labeled as events more than non-events.
	Cost-Sensitive Rule	An Ensemble-based classifier that accounts for real-world cost of a misclassification and prefers to misclassify non-events.
	Stacking	An Ensemble-based classifier that uses multiple individual classifiers and inputs and processes the output into a classification algorithm.
DT	Decision Tree	A supervised classification system that uses a divide-and-conquer computing approach to solve complex statistical decision making problems.
ROI	Region of Interest	An area marked by an operator to minimize noise of large spatial samples that allows the area to be marked as a relevant study area.
HITL	Human in the Loop	A model system that requires human interaction.
NSA	Not-Suitable-for-Attack(Non-Event)	A location that is labeled as not being the location of a conflict event.
	Sample(Instance)	A single object of the world from which a model will be learned, or on which a model will be used.
	Feature(Attribute)	A quantity describing an instance. An attribute has a domain defined by the attribute type, which denotes the

		values that can be taken by an attribute.									
	Feature Normalization	A method used to standardize the range of feature.									
	Feature Reduction (Dimensionality Reduction)	The process of removing noise and correlation among features.									
	Classifier	A mapping from unlabeled instances to (discrete) classes.									
ML	Machine Learning	The application of induction algorithms.									
	cross-validation	A method for estimating the accuracy (or error) of an inducer by dividing the data into k mutually exclusive subsets of approximately equal size.									
	Confusion Matrix	<p>A matrix showing the predicted and actual classifications. A confusion matrix is of size LxL, where L is the number of different label values. The following confusion matrix is for L=2:</p> <table border="0" style="margin-left: auto; margin-right: auto;"> <tr> <td style="padding-right: 10px;">actual \ predicted</td> <td style="padding-right: 20px;">negative</td> <td>positive</td> </tr> <tr> <td>Negative</td> <td style="text-align: center;">a</td> <td style="text-align: center;">b</td> </tr> <tr> <td>Positive</td> <td style="text-align: center;">c</td> <td style="text-align: center;">d</td> </tr> </table>	actual \ predicted	negative	positive	Negative	a	b	Positive	c	d
actual \ predicted	negative	positive									
Negative	a	b									
Positive	c	d									
R->P	Environmental Behavior module	A module which determines the tactical value of a location to get the potential Monitor and Control locations in within the Halo									
P->R	Route Based Behavior Module	A module which determines the highest risk positions of being attacked along a route, or inside an isolated location or area.									
	Sight Range	The outer radius of the Halo based on assumed human sight range.									
	Blast Range	Blast radius of an IED or extremely vulnerable distance from small arms fire.									
	Aiming Range	A range for Monitor and Control points to see the target continuously move along a route to the attack									

		emplacement location E.
	Device Triggering Range	The assumed maximum range of a trigger for an IED.
	Return Fire Range	The range of return fire by the victims from the Emplacement location to the Monitor and Control location.
	Retreat Distance To Cover	The distance to the nearest cover.
	Visibility	A measure of how easily from a location any point inside an isolated location, area, or route can be seen.
	cover	A patch of area which as no line of site to a location large enough for an attack to retreat behind and hide as needed.
	concealment	The measure of the extent of terrain near a potential attack site that does not have visibility to the attack site.
	Aiming	The measure of the extent of an isolated area, location, or route within the immediate vicinity of a potential attack that is in continuous visibility from a potential control or monitor location.
	Observability	The measure of an isolated area, location, or route that is visible from a location in the proximity of the isolated location.
	Route Exposure	The measure of an estimate of the total visibility of a potential attack site with suitable control and monitor sites surrounding it, and estimate of the exposure of the target at that location.
	Route Curvature	The measure of an estimate degree in curvature of a route that approaches the attack site within an isolated area, location, or route.
MECH-APP	Android MECH Application	The front-end interface for users to request tactical risk assessments of a study area.
MECH-WPS	MECH Web Portal Server	The back-end processing engine which computes in real-time studies created by users.

MECH-BM	MECH Behavior modeling	
MECH-CTS	MECH Classifier Training System	The back-end classifier training system which allow for importation of raw data and feature generation, configuration of classifier training experiments and performance reports.
CA1	Category 1:	Basic Measurements, several measurements derive from line of sight on the perimeter and location of an isolated point, area, or route.
CA2	Category 2:	Behavior Modeling, Probabilistic reasoning of locations for Monitor, Emplacement, and Control activities, Tactical parameters can be defined for risk seeking or aversion behaviors.
CA3	Category 3:	Machine Learning, Classification of R points by using one of the trained classifiers stored on MECH-WPS
CA4	Category 4:	Radio Activity, the power-frequency surveillance results of a software defined radio placed at a location.
CA5	Category 5:	Past Events, past events are placed on map area of the APP.
FG	Feature Generation	The module that imports raw data, and generates features which are then uploaded to the database. Feature Generation then extracts the features and stores in the specified database.
TC	Classifier Training	The module that preprocesses which consists of feature normalization and feature reduction, and running the training algorithms of classifiers. Classifier training also evaluates performance on data sets and uploads trained classifier to MECH-WPS.
CE	Classifier Ensemble	The module responsible for creation of ensembles of trained classifiers, which are uploaded to MECH-WPS.
	Mantrap	A small room with an entry door on one wall and an exit door on the opposite wall. The opening of the doors are mutually exclusive.

	Route-level Assessment	An assessment that focuses on possible attack locations along an isolated route to determine the most vulnerable or likely conflict event locations.
	Event-level Assessment	An assessment that focuses on features around an isolated location or area such as determining likely Monitor and Control locations given an Emplacement location.
POI	Point of Interest	A location input as a study area to be processed.

## FEATURE GLOSSARY

Class of Feature	Monitor/ Control/ Emplacement	Feature Name	Description
Visibility-Based	M/C	Visibility Index	The number of visible points within the viewshed of a point.
Visibility-Based	M/C	Discrete Shape Complexity Index	A discrete shape complexity index to characterize the evenness of radii along different directions in a (full) viewshed.
Visibility-Based	M/C	Minimum Cumulative Escape Adjacency	Minimum of the cumulative escape adjacency (CEA)
Visibility-Based	M/C	Maximum Cumulative Escape Adjacency	Maximum of the cumulative escape adjacency (CEA)
Visibility-Based	M/C/E	Route Visibility	Minimal (Min), Median (Med), and Maximum (Max) visibility from 100 meters to the route.
Visibility-Based	M/C	Median Cumulative Escape Adjacency	Median of the cumulative escape adjacency (CEA)
Visibility-Based	M/C	Shortest Radial	The shortest distance from the center to an invisible point along the n directions.
Visibility-Based	M/C	Longest Radial	The longest distance from the center to an invisible point along the n directions.

Visibility-Based	M/C	Mean Radical	The average distance from the center to an invisible point along the n directions.
Visibility-Based	E	Local Openness	Derived from sparse viewshed, a summarized viewshed along n equally spaced directions; an indicator of flatness or openness of the terrain. Smaller values imply flatter or more open terrain.
Geomorphometric	M/C	Planimetric Area	The area of a sparse viewshed based on its pixel count along its n directions.
Geomorphometric	M/C	Rugosity	The surface area (which considers the elevations of points) of a viewshed divided by its planimetric area along its n directions.
Geomorphometric	E	Shape Complexity	The surface curvature of a circle area (radius =10 pixels). (Smaller values imply smoother areas.)
Geomorphometric	E	Slope	The absolute value of the change rate in elevation along steepest path.
Geomorphometric	E	Texture	The number of pits divided by the number of pits and peaks in a circle area (radius =10 pixels, or 334 meters).
Geomorphometric	E	Local Convexity	The number of pits divided by the number of pits and peaks in a circle area (radius =10 pixels, or 334 meters).

Geomorphometric	M/C/E	Elevation Range	The difference between largest and smallest elevations with 50 meters.
Geomorphometric	M/C/E	Roughness	The standard deviation of elevations with 50 meters of a location.
Geomorphometric	E	Elevation	The height above or below sea level.
Social/Cultural	E	Proximity to Populated Areas	The nearest distance to a city/village with the population size of n.