GMU-AFCEA Symposium 2009
Linear Referencing for Network Analysis of IED Events

Kevin M. Curtin, PhD
Department of Geography and Geoinformation Science
George Mason University
05/20/09
curtin@gmu.edu

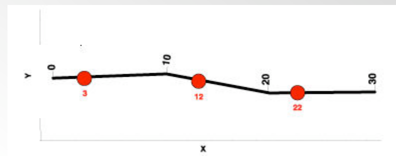
All data shown in this presentation is unclassified, publicly available

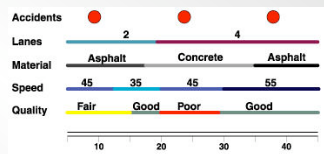
Presentation Outline

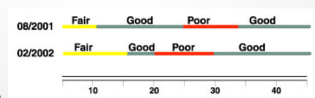
- Linear Referencing Background
  - Historical Uses
  - Justification for using Linear Referencing for IED Analysis
- Implementation
  - Network Representation
  - Generating the Linearly Referenced Network Database
- Outputs
  - Visualizations
  - Network-based spatial statistics
  - Measures of incident intensity on the network
- Ongoing/Future Research
  - Linearly Referencing Human/Physical Terrain Characteristics
    - Generate measures of Risk or Demand for Route Clearance Team (RCT) services for segments of the road network
  - Additional Network-based spatial statistics
  - Optimization of RCT services based on Linearly Referenced Demand

Linear Referencing Background – Historical

- Linear Referencing (a.k.a Dynamic Segmentation) is defined as:
  - GIS tools where point or linear geometry for database record events are referenced by their position along a linear feature
- Why use Linear Referencing?
  - Coordinate systems are unintuitive
    - Can a person in the field easily find:
      - 45°38'50" N 108°22'2" W
      - (12) 705152 5058224
    - How about:
      - Main Street, 30 yards north of Milepost 41
  - Attribute values do not match network topology
    - Traffic flow volumes
    - Pavement quality
    - Accident Locations, etc.
  - Attribute values change over time
  - Higher accuracy measurements exist outside GIS
  - **The network is the appropriate domain for analysis**

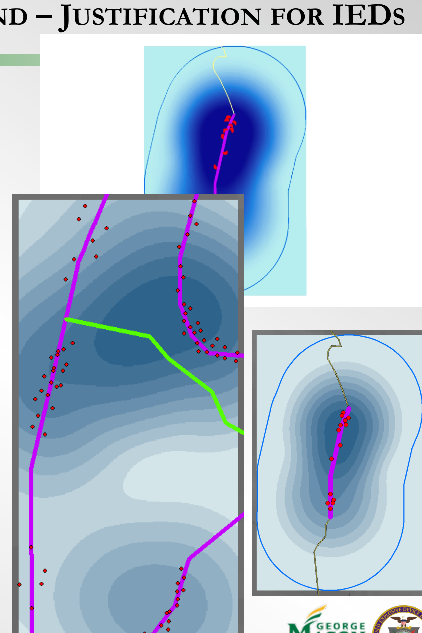






Linear Referencing Background – Justification for IEDs

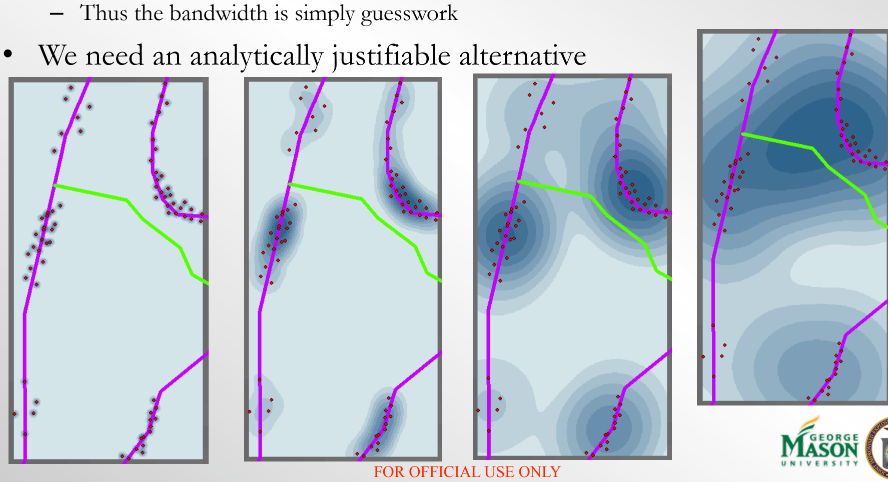
- Blue assets traveling on roads are a primary target for incidents/attacks
  - Road network ought to be the spatial platform for analysis
- Two dimensional kernel density plots are apparently useful in the field
  - Give a general sense of the "hot spots" of activity
- They are not defensible analytically
  - They assume that the process is acting uniformly across space in all directions
    - The greater the intensity on the roads, the greater the intensity off the roads
  - They can lead to "false alarms" when identifying road segments



FOR OFFICIAL USE ONLY

Linear Referencing Background – Justification for IEDs

- Density values are a function of the kernel and the bandwidth
  - The bandwidth is the representation of the spatial interaction among points
  - There is no known (empirically derived) spatial relationship among these points
  - Thus the bandwidth is simply guesswork
- We need an analytically justifiable alternative

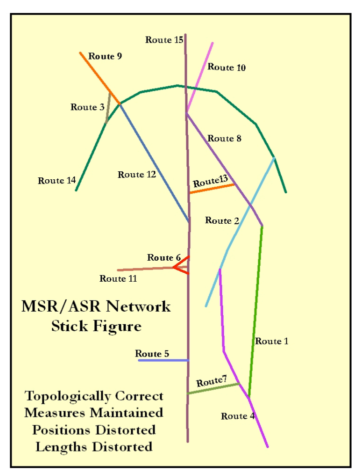


FOR OFFICIAL USE ONLY

FOR OFFICIAL USE ONLY

Implementing Linear Referencing for IEDs - Network

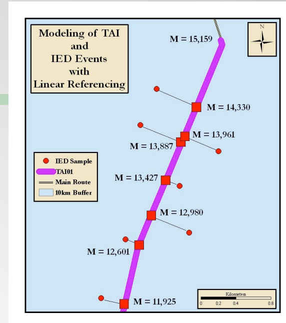
- Blue assets traveling on roads are the primary targets for incidents/attacks
  - Road network must be the spatial platform
- MSR/ASR Route Network
  - "Stick Figure" unclassified representation
    - Topologically correct and measures preserved through LR
    - Shapes changed, locations moved, lengths distorted
- Many spatial representations can be generated from these routes:
  - Network to be divided into 20 km segments
    - For the purpose of RCT scheduling
    - Current TAIs for RCT activity are 10 - 20 km
- Additional modifications/representations can be generated as needed
  - Blue Force monitoring (1 km segments)
  - New ASR definitions



FOR OFFICIAL USE ONLY

FOR OFFICIAL USE ONLY

Implementing Linear Referencing for IEDs – IED Events



	A	B	C	D	E	F	G	H
6979	Year	Route	RouteSeg	Begin	End	Target	Outcome	Count
6980	2008	R15	R1508	140000	160000	A	F&C	0
6981	2008	R15	R1508	140000	160000	A	INEFF	3
6982	2008	R15	R1508	140000	160000	B	EFF	16
6983	2008	R15	R1508	140000	160000	B	F&C	34
6984	2008	R15	R1508	140000	160000	B	INEFF	17
6985	2008	R15	R1508	140000	160000	C	EFF	14
6986	2008	R15	R1508	140000	160000	C	F&C	26
6987	2008	R15	R1508	140000	160000	C	INEFF	33
6988	2008	R15	R1508	140000	160000	D	EFF	0
6989	2008	R15	R1508	140000	160000	D	F&C	11
6990	2008	R15	R1508	140000	160000	D	INEFF	5
6991	2008	R15	R1509	160000	180000	A	EFF	6
6992	2008	R15	R1509	160000	180000	A	F&C	7
6993	2008	R15	R1509	160000	180000	A	INEFF	13
6994	2008	R15	R1509	160000	180000	B	EFF	2
6995	2008	R15	R1509	160000	180000	B	F&C	8
6996	2008	R15	R1509	160000	180000	B	INEFF	5
6997	2008	R15	R1509	160000	180000	C	EFF	1
6998	2008	R15	R1509	160000	180000	C	F&C	10
6999	2008	R15	R1509	160000	180000	C	INEFF	4
7000	2008	R15	R1509	160000	180000	D	EFF	0
7001	2008	R15	R1509	160000	180000	D	F&C	19
7002	2008	R15	R1509	160000	180000	D	INEFF	10
7003	2008	R15	R1510	180000	200000	A	EFF	1

- Linear Referencing of IED Events
  - Push incidents onto the network
  - Give each incident a "measure" location
    - Can retain offset measure
  - Assumptions
    - Accuracy of incident locations from SIGACTS
    - Accuracy of underlying road locations
    - "Road" has some non-zero width
      - Any attack within 100 m of the road is directed at the road
  - **Deliverable Linearly Referenced Incident Database**
- Derived Incident Data Tables
  - Variables:
    - Count of Incidents
    - Route and Segment along each route (20km)
    - Begin and End Measure of each Segment
    - Year (2004-2008)
    - Target Type
    - Outcome
  - Allows multiple characterizations of Incidents on the Road Network

FOR OFFICIAL USE ONLY

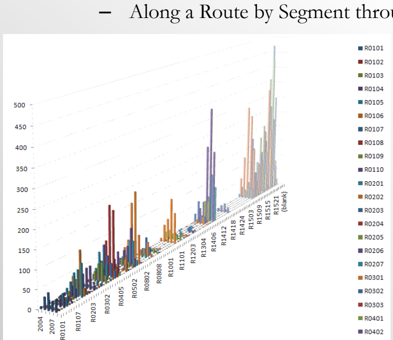
FOR OFFICIAL USE ONLY

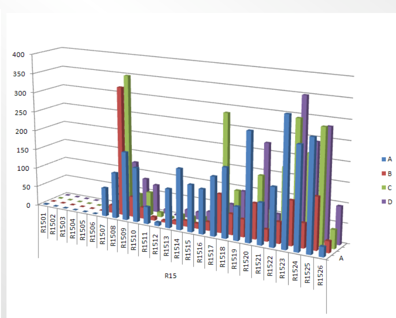
Outputs – Linearly Referenced IED Database

- Permits Spatial Summary by Route or Segment

Sum of Count	Route															
Year	R01	R02	R03	R04	R05	R06	R07	R08	R09	R10	R11	R12	R13	R14	R15	(blank)	Grand Total
2004	44	39	30	208	9	22	18	60	1	10	22	2	7	82	568		1122
2005	102	183	51	362	22	62	24	101	8	51	44	16	1	195	1055		2277
2006	137	175	7	408	4	75	67	255	24	71	16	16	11	445	2500		4211
2007	140	282	17	418	34	122	68	287	28	131	19	34	36	645	3345		5606
2008	93	142	14	91	47	34	15	99	7	92	15	19	15	356	822		1861
(blank)																	
Grand Total	516	821	119	1487	116	315	192	802	68	355	116	87	70	1723	8290		15077

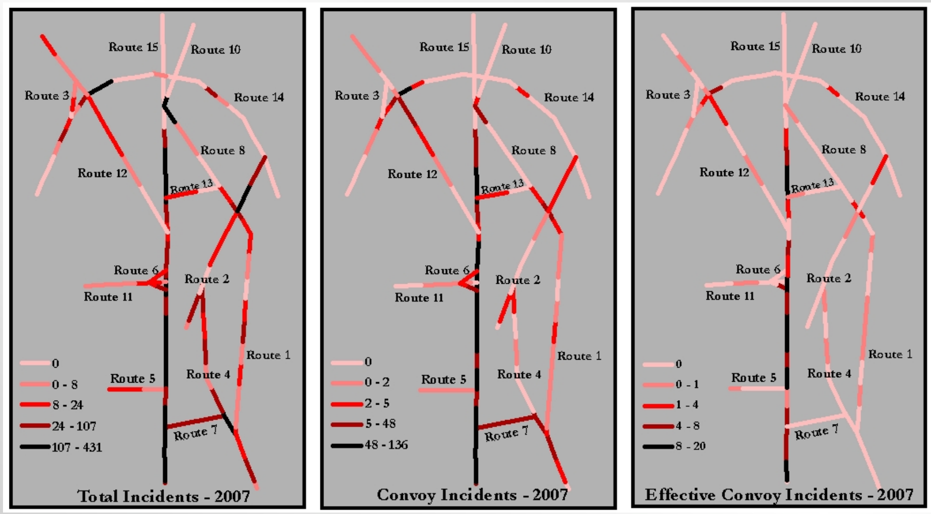
- Permits visualization:
  - Along a Route by Segment through Time; By Target along a Route





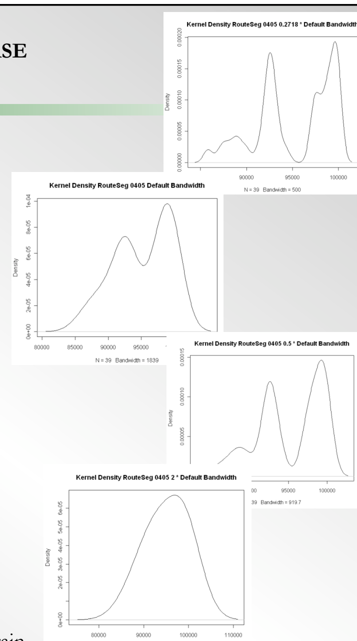
FOR OFFICIAL USE ONLY

Outputs – Linearly Referenced IED Database



Outputs – Linearly Referenced IED Database Used for Network-Based Spatial Statistics

- Network density can be:
  - # of incidents per unit length
  - Kernel density estimations
    - Bandwidth and kernel function vary
  - Density values can be linearly referenced
- Network based clustering statistics
  - Initial effort Linear Nearest Neighbor Clustering Statistic
  - Based on Okabe, Yomono, and Kitamura (1995)
    - Is there a significant level of clustering (or dispersion) along a line
    - Null Hypothesis: the events are distributed randomly
    - Euclidean distance replaced with Shortest Path distance
    - Difference between incident measure values
  - We have run the statistic on the 20 km segments
- Questions we can Ask/Answer with this Statistic
  - Where are the significant clusters along the network?
    - Should these areas be designated as TAIs?
    - Do current TAIs correspond to significant clusters?
    - Are significant clusters consistent through time?
  - Are clusters associated with particular human/physical terrain characteristics…and if so, how can we use these to forecast future incident activity?



FOR OFFICIAL USE ONLY

Outputs – Linearly Referenced IED Database Used for Network-based Spatial Statistics

- Point Pattern Analysis first used by botanists and ecologists to explain the distribution of plant species
  - It has since spread to many different fields
  - The basis for “Hot-Spot” analysis in criminology
- Point Pattern analysis asks
  - Is there a statistically significant clustering (dispersion) of activity?
- Compare the mean nearest neighbor distance of your data set to the values of theoretical distributions
  - Observed pattern must fall between:
    - Theoretically most dispersed pattern and
    - Theoretically most clustered pattern
    - The random pattern is somewhere in-between
  - Use the Nearest-neighbor index
    - Observed distance divided by random distance
    - A nearest neighbor index of 1.0 means your pattern is random
    - Less than 1 means clustered, greater than 1 means dispersed
- Test for Significance

$$R_l = \frac{\bar{t}_{obs}}{E(t)}$$

$$E(t) = \frac{(n+2)l}{2n(n+1)}$$

$$Var(t) = \frac{(n^3 + 8n^2 - 8)l^2}{4n^2(n+1)^2(n+2)}$$

FOR OFFICIAL USE ONLY

FOR OFFICIAL USE ONLY

Outputs – Linearly Referenced IED Database Used for Network-based Spatial Statistics

- Results by 20-km segments:
  - Red – Significant Clustering
  - Green – Significant Dispersion
- Allows us to determine departures from spatial randomness
  - Through time
  - Across space
- Allows us to choose areas to target for mitigation efforts
- Allows us to associate clusters with other human and physical terrain characteristics for forecasting or planning efforts
- It is the Linear Referencing that allows us to generate these results

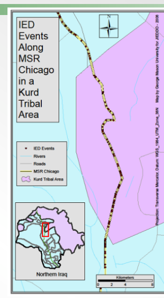
Route 15					
Segment	2004	2005	2006	2007	2008
7					
8			298(0.001)		163(0.001)
9				195(0.05)	
10			41(0.1)		55(0.05)
11			9(0.1)		
12				10(0.05)	
13	11 (0.1)				6(0.1)
14				99(0.05)	
15				126(0.001)	8(0.1)
16		30(0.05)			11(0.01)
17			180(0.1)		
18	21(0.1)				
19		38(0.005)			
20			195(0.05)		
21				83(0.05)	42(0.05)
22				197(0.001)	
23	76(0.1)	155(0.01)	365(0.001)		
24			272(0.001)		70(0.1)
25		133(0.005)	474(0.001)		23(0.1)
26		21(0.001)	83(0.001)	81(0.001)	

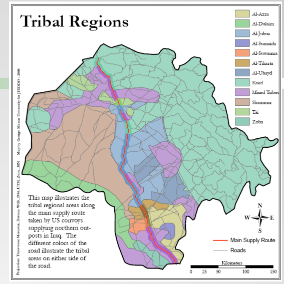
FOR OFFICIAL USE ONLY

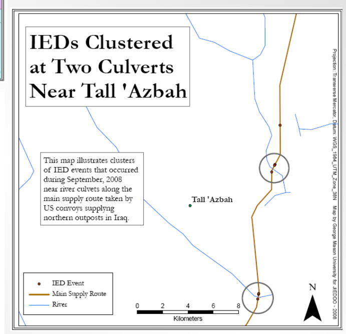
FOR OFFICIAL USE ONLY

Ongoing/Future Research – Linearly Referencing Terrain Characteristics

- Linear referencing allows spatial comparisons of incident locations to other road characteristics
  - Physical Terrain characteristics
    - Streams
    - Culverts
    - Road Intersections
    - Land Cover
    - Visibility
  - Human Terrain characteristics
    - Province boundaries
    - District boundaries
    - Built-up Areas
    - Road Type
  - Incident Density from Task 3
- Purpose: Spatial Association of Variables





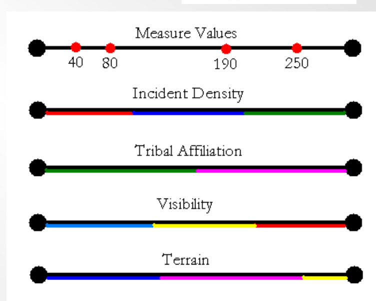


FOR OFFICIAL USE ONLY

FOR OFFICIAL USE ONLY

Ongoing/Future Research – Linearly Referencing Terrain Characteristics

- Combine Linearly Referenced characteristics to generate:
  - Overall Measures of Risk
  - Potential for Safe Travel
  - Demand for RCT Services
- Identify measures of incident risk to:
  - Forecast (predict) most likely areas for incidents/ attacks
  - Identify areas most in need of RCT services
- Precise methodology for combining all elements into a single risk measure not yet formulated
  - Will rely on complementary efforts of other sub-tasks of the Mason JIEDDO group
- Intended for use as an input to RCT Allocation Optimization



FOR OFFICIAL USE ONLY

Ongoing/Future Research – Optimization of RCT Locations

- Combine (through optimization):
  - Ability to locate events (IEDs, terrain, etc.) through linear referencing
  - Goal of optimizing RCT activities
- Initial formulation - Maximal Covering Location Problem
  - Measure of priority for covering any section of the network is a function of:
    - Risk of IED, or demand for RCT services
    - Generated in Task 4
  - May be possible to add temporal component
    - Model time patterns observed in the data
    - To meet scheduling needs of the RCTs
  - Future work must combine Blue, Red, and Green activities

i, j = the indices of nodes in the network
E = the set of edges (i, j) in the network
k, K = index and set of base locations for RCTs
P = the number of RCTs available to patrol the network
S = the acceptable service distance (covering distance)
$N_{ij} = [k \text{ in } K \mid \min(d_{ki}, d_{kj}) \le S]$
y_{ij} = the number of RCTs covering the arc from i to j
x_k = the number of RCTs assigned to base location k
a_{ij} = the measure of priority for covering the arc from i to j

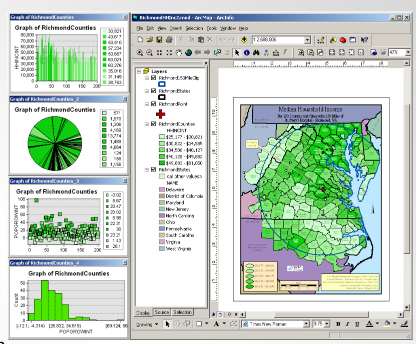
$$\text{Max} \sum_{(i,j)\in E} a_{ij} y_{ij}$$

$$\sum_{k\in K} x_k = P$$

$$\sum_{k\in N_{ij}} x_k \ge y_{ij} \quad \forall\, (i,j) \in E$$

Ongoing/Future Research – GIS-Based Tool Development

- Research tasks above could be combined into a functional tool for the generation of efficient solutions to operational problems
  - Since
    - GIS is well-accepted in military contexts
    - A call has been made to provide access to mapping software to the ORSAs in the field who are collecting and maintaining the databases of IED locations
  - Next logical step is to develop a set of GIS-based tools to implement the research advances in an operational environment
  - Our research group has extensive experience in developing such tools



Ongoing/Future Research – Additional Network Based Spatial Statistics

- The Linear Nearest Neighbor method:
  - Uses only the nearest neighbor distance (what about 2nd nearest, 3rd nearest etc.)
  - Describes only if clustering occurs globally on the study area (the segment in our case)
  - Cannot specify the locations and sizes of the clusters
    - If we can identify which incidents are in which cluster we can identify which types of incidents are clustering
- Next steps are to explore:
  - Network Based K-functions
    - Multi-Distance Spatial Cluster Analysis
    - Summarizes spatial dependence over a range of distances
  - Local K-function to identify clusters explicitly
- End