HIT: A GIS-Based Hotspot Identification Taxonomy

Dana Steil and Allen Parrish
University of Alabama, Tuscaloosa, AL USA

Abstract

The purpose of this GIS-based hotspot identification taxonomy is to bring together strategies for planning and evaluation of selective traffic enforcement programs (STEPs). The HIT model is specifically designed for enforcement programs that focus on events and countermeasures that can be spatially and temporally described. The HIT model encompasses four core activities that are tightly integrated: data-collection, linear hotspot identification (HISD), presentation and assessment. Event location data is preprocessed using geographic information systems (GIS) then three types of hotspots are identified. First order hotspots locate road segments with a high frequency of a specific type of event. Second order hotspots locate road segments for which the ratio of event counts to countermeasure duration is minimized or maximized. Third order hotspots locate road segments with a high frequency of events and where a countermeasure is historically effective.

Key Words: Hotspot identification, GIS, traffic law enforcement, traffic safety, STEP, TACT.

1 Introduction

In recent years concentrated traffic law enforcement efforts have become a popular tactic because of their high visibility and positive impact. These patrol efforts often concentrate on specific offenses, times, locations or any combination thereof. A time concentrated effort, commonly referred to as a blitz, focuses on events or holidays that are associated with high crash frequencies. Space concentrated efforts focus patrols on roadway segments with higher than expected crash frequencies called hotspots. Concentrated patrol efforts such as these are implemented using a STEP (Selective Traffic Enforcement Program) which typically aims to reduce a specific category of crashes or discourage a dangerous behavior [13]. One of the most popular STEP programs in recent years was “Click it or Ticket” sponsored by the National Highway Safety Administration. Many of the STEP programs use a combination of publicity and increased law enforcement to educate and motivate the public.

This effort focuses on building GIS enabled software systems to support two of the major components of STEP: hotspot identification and program evaluation. This software best fits STEP with a strong emphasis on spatial and temporal factors. One recent STEP that fits this bill is TACT (Targeting Aggressive Cars and Trucks) sponsored by FMCSA (Federal Motor Carriers Safety Administration) [13]. TACT was piloted in the State of Washington and has now expanded to six states. The FMCSA is encouraging the participating states to build on Washington’s success and identify additional law enforcement and publicity strategies that will deter aggressive driving.

This paper presents a hotspot identification taxonomy (HIT) for selective traffic enforcement programs. The HIT model is used as a strategic framework for STEP planning and evaluation software. In this taxonomy, hotspot identification techniques are used in both the planning and evaluation stages. Furthermore, HIT integrates the planning, evaluation and feedback loop to the extent that the separation between stages is blurred when used in conjunction with continuous enforcement programs. HIT is designed to support both short term and continuous STEP. In fact HIT relies on data collected during daily patrols in conjunction with other continuously collected data (such as crash reports) to inform its recommendations for both short term and continuous STEP.

The body of this paper is structured as follows. First, background literature in hotspot identification (HSID) methods, HSID used in law enforcement tools, and traffic law enforcement efficacy measures is reviewed. Second, the HIT model is presented in detail including descriptions of data collection, first, second and third order hotspots, effectiveness assessment, and presentation issues. Third, future plans are addressed.

2 Background

2.1 Hotspot Definitions

While there is an abundance of literature addressing HSID, a single definition of a hotspot does not exist [3]. The umbrella term hotspot is used for a variety of location-targeting strategies. Hotspots are also known as black-spots, hotbeds, hot-points, sites with promise, high risk locations, and Local Indicators of Spatial Association (LISA) [7, 21]. The following will archetypal hotspot definitions are used as a foundation.

1) In 1978 Hakkert defined hotspots as “Those sites whose
accident frequency is significantly higher than expected at some prescribed level of significance” [15].

2) In 1981 McGuigan defined hotspots as “Sites with a potential for accident reduction, measured as the difference between the observed and expected number of crashes at a site” [20].

3) In 2005 Cheng defined hotspots as “transportation system locations (road segments, intersections, interchanges, ramps, etc.) that possess underlying correctable safety problems, and whose effect will be revealed through elevated crash frequencies relative to similar locations” [7].

2.2 (HSID) Methods

In general, HSID (Hot Spot Identification) methods are specific strategies for identifying areas with a high concentration of some metric. The following survey of HSID methods offers a brief introduction to various strategies.

High Frequency. The most common method for hotspot identification is to choose the locations with the highest frequency of events [3]. This method is sometimes called simple ranking [7]. Practitioners generally refer to the top $n$ locations from each of a variety of road categories as hotspots. Roads are typically categorized by factors such as number of lanes, average daily traffic, rural or urban, etc.

Statistical Confidence Intervals. The statistical confidence interval method for identifying hotspots is very similar to the High Frequency method. Rather than choosing the top $n$ segments, all of the segments whose frequencies are above the mean by a prescribed level of confidence are identified as hotspots. Much effort has focused on choosing proper confidence level.

Bayesian Methods. Empirical Bayesian methods for hotspot identification reduce the number of false positives and negatives when compared to the simple ranking and statistical confidence interval methods [7]. Bayesian methods use crash history and the expected crashes for similar sites to distinguish between clusters of random crashes and true patterns. This method assumes that crash occurrences obey the Poisson probability law and the expected number of crashes is gamma distributed. Many have recommended using a Poisson probability for crash modeling over the common multiple linear regression model (MLR) [8]. Chin reported that “Since accident occurrences are necessarily discrete, often sporadic and more likely random events, the Poisson regression models appear to be more suitable than the MLR models” [8]. The benefits of the Bayesian methods depend on good traffic and crash data as well as high crash counts. Mirand-Moreno showed that two Bayesian testing procedures for hotspot identification, false discovery rate (FDR) and false negative rate (FNR), were effective for ranking hotspots for treatment selection [22].

Grid Scanning. The use of event counts within the cells of 2-dimensional grids has long been used to identify hotspots [21]. The grid scanning process simply sums the number of events within the space then hotspots are identified. Hotspots can be identified by sorting the list of cells by the count and then selecting the top $m$ cells. Alternatively, the mean and the standard deviation of the cell counts can be used to select all of the cells above a predetermined threshold. McCullagh illustrated the use of temporal grid scanning by weighting the cells with deltas of two time periods in epidemiology studies.

Kernel Estimation Method. The Kernel Estimation Method uses a fixed sized window to view one portion of a 2-dimensional world at a time [21]. As the window is moved to each of the possible positions in space, a density is calculated for the center point of the window. The kernel is a “three-dimensional function which weights events within its sphere of influence according to their distance from the point at which the intensity is being estimated” [11]. The kernel density method is particularly useful when used in areas of dense road networks and when multiple events occur at the same location [17, 27]. Anderson held that the Kernel Estimation method provides a “realistic continuous model of accident hotspot patterns reflecting the changes in density which are often difficult to represent” [3]. Krishnakumar used the kernel estimation method to locate pedestrian crash hotspots in Las Vegas. The practitioner is left to choose an appropriate kernel size. Erdogan used a .5km radius for a kernel density analysis of crashes in Afyonkarahisar, Turkey [9]. The Kernel Density method in Section 3.2.2 explores a similar approach using events on a 1-dimensional road network.

Local Indicators of Spatial Association. Local Indicators of Spatial Association (LISA) HSID methods look for “significantly different areas detected by the applied statistics” [4, 21]. LISA methods identify circular, elliptical or amoeboid regions that contain values that are well above or below the overall random distribution in space. Steenberghen described LISA as “an indicator of the extent to which the value of an observation is similar to or different from its neighboring observations” [27]. Spatial and Temporal Analysis of Crime (STAC), a tool developed at the Illinois Department of Justice, uses a LISA method [24]. STAC is now part of Crimestat [18]. A common criticism of the LISA method is that crime, disease and other events that are studied seldom fit the circular or elliptical pattern [21, 24].

Time Lag Between Events. The “time lag between events” hotspot method looks for locations/segments for which events are frequent and regular [2]. These spots are found by looking for a low mean time-between-events along with a low standard deviation. Ratcliffe used this approach when using a GIS approach to look for “repeat victimization” [24]. Little work has been done to include temporal aspects in commercially available hotspot tools.

K-Means Clustering. K-Means clustering was created by MacQueen in 1967 [19]. Many clustering techniques begin with all of the elements in their own cluster and then groups them until there is one cluster. “The goal of the K-Means clustering algorithm is to divide (or partition) the objects into $k$ clusters such that a value relative to the centroids of the clusters is minimized” [16]. Specifically, K-Means minimizes the sum of the distance from every point to the K centers. The value of K must be selected in advance (this is one of the
greatest challenges to using K-Means clustering). Kim recommends using K-Means as an exploratory tool, testing multiple K values to find clusters. Since the objects in Kim’s study were auto crashes, distances were measured along the road network instead of using Euclidean distances. As with many clustering techniques, it is possible for the centroid of a K-Means cluster to be located where there are in fact no objects.

**Network Distance Weighted Clustering.** Network Distance Weighted Clustering (NDWC) distinguishes itself from other GIS methods by using distances along the road network instead of Euclidean distance. NDWC looks for areas of high concentration by measuring the distance between many random locations along the road network and the events within a fixed travel distance. The measurement from random locations “guarantees independence between the results and the actual accident locations” [1]. The randomness also accounts for the typical rounding issues associated with crash location reporting. Aerts refers to the NDWC’s output as a “dangerousness map”. The dangerousness map is built using Aerts’ dangerousness index which is “a measure for the dangerousness of a point of measurement and is calculated from the weighted number of accidents that occurred in a certain distance along the road network from the point of measurement” [1]. Aerts tested three weighting functions (Distance Band, Inverse Distance, and Linear Decrease). He reported that the Distance Band method overrated the dangerousness index, the Inverse Distance method made underestimations and the Linear Decrease method gave the most homogeneous results.

### 2.3 HSID and Law Enforcement Tools

**Geographical Analysis Machine.** The Geographical Analysis Machine (GAM), developed by Stan Openshaw in 1987, provided a “new approach to the analysis of point pattern data based on a fully automated process whereby a point data set is explored for evidence of pattern without being unduly affected by predefined areal units or data error” [21]. GAM was the “first of the new generation of spatial analytical technology and was based on a fusion of statistical, GIS and computational techniques intended to detect deviations from the Poisson distribution of rare events” [23]. It was initially used in epidemiological studies. GAM’s strategy was to repeatedly define circular areas of varying radii in the target area and checking for a significant difference between events inside and outside the circles. When a significant difference between the areas inside and outside a circle was observed, the circle was drawn on the resulting map. GAM first had limited use because it required the use of a supercomputer. Over the years GAM’s strategies have changed and it has become more widely used.

**Spatial and Temporal Analysis of Crime.** Spatial and Temporal Analysis of Crime (STAC) was developed at the Illinois Department of Justice (IDJ) in 1988. Ratcliffe gives credit to STAC as “one of the first crime mapping packages to be wildly available” [24]. The Illinois Criminal Justice Information Authority’s (ICJIA) website says that STAC is “a tool to find and examine Hotspot Areas on the map”. They freely offer both the STAC Spatial Analyzer and STAC Time Analyzer. Both of these tools are also integrated in CrimeStatII, which is also freely available from the National Archive of Criminal Justice Data. As noted earlier, STAC uses the LISA hotspot identification method.

**Spatio-Temporal Visualization.** Spatio-Temporal Visualization (STV) was developed at the AI Lab at the University of Arizona for the Tucson Police Department. “The STV enables crime analysts to examine the same data from three different views simultaneously and to identify crime patterns” [5]. These views include a GIS view, a time-line view and a pattern view. STV uses data extracted from COPLINK where the crime reports are stored with the needed location and time information. COPLINK is another tool developed by the AI Lab at the University of Arizona and funded by the National Institute of Justice [6]. STV’s GIS view uses ESRI’s tools to display a map of the city and their analysis results. They developed their own time-line and pattern views which are controlled by a “central time-slider”. The central time slider is tied to the three views and allows the user to select a time window size and move forward and backward through time. According to Chen, the Tucson police have been pleased with STV and report increased investigation effectiveness and a 20-time speed improvement.

**Desktop Hypercube.** The Desktop Hypercube, officer deployment simulation tool, was developed and evaluated under the leadership of Stephen Sacks at the University of Connecticut. The Desktop Hypercube software “allows police planners to draw patrol districts on a map of their city and to evaluate performance” [26]. The system measures the quality of police services using four measures (response time, workload balance among the cars, patrol frequency and the number of interdistrict dispatches). Sacks found that the police planners and academics are often poor predictors of the impact of police care deployment changes. At the time of his writing in 2000, Sacks reported application of the Desktop Hypercube in Chapel Hill, North Carolina; Orlando, Florida; and Hartford, Connecticut. He noted that “significant changes in population, crime rates, and number of police officers make it impossible to do meaningful before-and-after comparisons of performance”.

### 2.4 Citation Effectiveness

Accurate evaluation of traffic law enforcement efforts are at the core of the HIT model. The mission statements of many of the highway patrol departments in the United States reflect the belief that issuing citations is an effective auto crash countermeasure. Furthermore, the positive impact of citations on crashes is an important premise for patrol advisory tools. “The fact that a law is on the books will, by itself, change the behavior of some drivers who believe that obeying the law is a canon of good citizenship. Other drivers change their behavior to avoid the penalties specified by law” [10]. This section reviews several studies, each with its own way of measuring citation effectiveness.

On a micro level Redelmeier found that “traffic-law enforcement effectively reduces the frequency of fatal motor-
vehicle crashes in countries with high rates of motor-vehicle use” [25]. His study in Ontario focused on the impact of a citation on a specific driver. In the month after receiving a citation the drivers were 35 percent less likely to be involved in a fatal crash than in comparable months without a citation. The influence of the citation lessened as time passed. Within 3-4 months the influence was not significant. It was also found that the influence of the citation was greater when it carried “points” that negatively impacted the drivers record.

Gebers specifically studied how traffic convictions and their related points can be used as predictors of accident-risk drivers in California [12]. He found that traffic conviction history is nearly as powerful as crash history when predicting ones accident-risk. Gebers recommended the use of a combination of these histories to obtain the best predictive results. It is important to note that Gerber did not suggest that a driver’s receipt of a citation caused their increased likelihood of involvement in an auto-crash. Rather he is showed the correlation between a driver’s citation history and their subsequent likelihood of accident involvement.

Hakkert’s study in Israel took a macro approach to the relationship between citations and crashes [14]. Hakkert reported on the impact of a 1997 National Traffic Police (NTP) initiative named the 700-project. The 700-project increased law enforcement on 700kms of Israeli interurban roads (20 percent of the interurban roads). Their goal was to reduce serious injury accidents by 10 percent in the treatment areas. In the first month of the treatment period, amongst high publicity, success looked promising. When the 700-project was complete the results were mixed. By some measures of traffic law compliance, such as safety belt usage, the project was successful. Unfortunately, during the study period the average frequency of serious injury accidents on interurban roads increased, the 700 km study area included. The study found that 4 of the 5 treatment segments did experience a significantly less serious injury increase than average. Hakkert concluded that “although the enforcement considered did not reduce the accident numbers during the project year, it seemed to be a deterrent factor for the increasing accident trend that appeared this year on the interurban roads.”

Truls Vaa’s work also took a macro approach to the relationship between police enforcement and driving speed on a 35 km study area in Norway [28]. Vaa’s results showed that the average driving speed was reduced between 0.9-4.8 km/h in 60-80 km/h zones during the 6 weeks of increased enforcement. The speed reduction was observed during all time periods except the “morning commute”. In some areas, the speed reductions associated with the increased enforcement continued for a short period afterward.

The supposition that citations are generally an effective countermeasure for crashes will be evaluated with several further studies. First, a repeat Redelmeier’s micro approach will be performed using Alabama’s citation and crash data. Second, a temporal and spatial comparison of citation and crash location data will attempt to find their relationship (a macro approach). See Section 3.4 below for more details about the macro approach.

3 HIT Model

The model in Figure 1 represents the hotspot identification taxonomy (HIT). The HIT model is a cycle of data collection, hotspot definitions, presentation and feedback. Each of these categories will be described in this section.

3.1 Location Data Collection

Accurate and timely collection of event data is essential to the success of a HSID system. This section surveys HIT’s

![Figure 1: HIT - STEP model](image)
core data sources, crashes and citations. Several complementary data sources are also described.

3.1.1 Crashes. In Alabama the Department of Public Safety (DPS) is responsible for collecting crash data and forwarding summary information to other state and Federal agencies. For many years transportation safety engineers at the Alabama Department of Transportation (ALDOT) have used the crash data to locate road segments in need of improvement. One of the tools that ALDOT uses to analyze the crash data is CARE (Critical Analysis and Reporting Environment) [28]. Alabama’s crashes are currently first documented on paper by police officers then later recorded in an electronic format for storage and analysis. Complete electronic crash data is available in CARE approximately three months behind the crash date. While CARE was originally created for crash analysis, other datasets such as citations can be used. To date in Alabama, only the crashes that occur on mile-posted routes, the state routes and interstates, can be geo-located (mapped). Non-mappable crashes are referred to as “off system”. Early in 2009 a new crash form will be used in Alabama that requires lat/long for each crash. After discussions with other states, which have had procedural difficulties reading accurate GPS locations directly input into their crash forms, it was decided to use a combination of direct GPS reading, manual entry, and map clicking to geo-locate crashes.

3.1.2 Citations and Patrol Coverage. E-Cite is an application used by officers in their patrol cars to issue and record citations on a mobile computer (Figure 2). E-Cite was first deployed five years ago to a subset of Alabama’s Troopers that are specifically target commercial vehicles violations. In the past three years its usage steadily increased. E-Cite is currently used by most Alabama State Troopers and more than 15 municipalities totaling more than 600 officers in Alabama. Officers periodically upload the citations to a central server, typically within a day. These citations include the spatial and temporal data needed by HIT.

3.1.3 Other Data Sources. Some related data sources are currently under development. First, weight logs from weigh stations, WIM (weigh in motion) devices, and periodic measurements of road rutting can be used to define additional target zones. Second, potential trouble areas can be identified by building an inventory of skid marks via analysis of roadway images. Third, crime locations, driver’s license addresses, and auto registration addresses are used to identify non-traffic related target areas.

3.2 First Order Hotspots

First order hot spots give a high level view of events along the roadways. Hakkert’s hotspot definition is appropriate for first order hot spots, “Those sites whose accident frequency is significantly higher than expected at some prescribed level of significance” [15]. McGuigan and Cheng’s definitions of roadway hotspots also include the notion of identifying road segments that have high event frequencies as compared to similar locations [7, 20]. In this section a few key parameters and six first order hotspot designation methods are described.

3.2.1 First Order Hotspot Factors. The CARE system relies on the user to make some choices when defining first order hotspots. The user first selects a subset of events by building a filter using any of the associated event variables. In the case of Alabama crashes there are more than 200 variables to work with. Counts of crashes and citations for a given road segment vary greatly by time of day. Figure 3 shows the time of day when crashes and citations occurred on an 8 mile road segment in Alabama from January to July 2007. Another important consideration is how to segment the road into either fixed or dynamic length segments. Currently most of CARE’s HSID methods use fixed segmentation (FS), where all identified hotspot segments are exactly length s and give the user the power to choose s. FS methods often identify contiguous hotspots that are only separated because s is too small. When the user chooses larger values of s, FS hotspot segments often have large portions without events. There are two variations of dynamic segmentation (DS) strategies, neither rely on the user to choose s. The first DS method programatically selects a fixed s by trying all of the possibilities in a range and choosing the one that minimizes contiguous hotspots and hotspot portions without events. The second DS method...
locates hotspots of varying length depending on how the events are clustered. The Kernel Density method described later is an example of the variable length DS method.

**3.2.2 Designation Methods.** Section 2.2 reviewed several popular techniques for identifying first order hotspots. This section describes six specific techniques available in the CARE environment. The results from each of CARE’s HSID tools can be viewed as a GIS layer on a traditional map or linearly on a strip-map. Recall, these HSID tools are used for any roadway events, most commonly crashes and citations.

*Frequency-Display* graphically represents roadway event frequencies on a designated output view (map or strip map). This method leaves the actual hotspot identification to the viewer. When using a map view, graduated lines are superimposed over segmented roadways according to event counts or event rates. Event rates are event counts normalized by ADT (average daily traffic) or any other roadway characteristics such as number of lanes or road type. The map view method also allows the user to weigh events using any combination of their variables (such as accident severity or alcohol involvement). These weights in turn alter the line thickness of each segment’s overlay. Figure 4 illustrates a map view frequency display of crash events. When using a strip-map view, only one segment of a single road is displayed at a time. The strip map allows details, such as crash severity, to be shown. Figure 5 illustrates a strip-map view of crash events along with a map to help orient the user.

Second, the *High-Frequency* HSID algorithm looks for hotspots using a user defined frequency threshold \( t \) and segments size \( s \). The method examines each road segment in tenth mile increments \( i \) from beginning to end. For each \( i \) if the next \( s \) miles contain \( t \) or more events the \( s \) length segment is marked as a hotspot and \( i \) is incremented by \( s \). Figures 6 and 7 illustrate the map and strip-map views of *High-Frequency* and CARE’s other fixed segment length HSID methods.

Third, CARE can define hotspots using *Trend-Detection* (aka early-warning and before-after). The trend-detection hotspots alert the user to changes in event frequencies over time, possibly due to road or officer patrolling changes. Like high-frequency, trend-detection uses a user defined frequency threshold \( t \) and a segments size \( s \). Additionally, the user must define \( t_1 \) and \( t_2 \) as before and after time periods to compare. The road segments are again examined using tenth mile increments \( i \). For each \( i \), if the next \( s \) miles during \( t_2 \) contain \( t \) or more events than \( t_1 \) the \( s \) length segment is marked as a hotspot and \( i \) is incremented by \( s \). When \( t_1 \) and \( t_2 \) span different amounts of time the frequencies are scaled proportionally before comparisons are made.

Fourth, two subsets of events are compared with one another using the *Comparison* HSID method. While Trend-Detection compares temporally partitioned event subsets, the Comparison method compares any two filtered subsets of events \( f_1 \) and \( f_2 \). Following the same pattern as High-Frequency and Trend-Detection, segments of length \( s \) are flagged as hotspots when the ratio of event frequencies between \( f_1 \) and \( f_2 \) is greater than or equal to a user defined threshold \( t \).

Fifth, *Rate-QC* finds hotspots by comparing each road segment with similar segments in the same dataset. Similarity between segments is determined by comparing characteristics such as ADT and number of lanes. Rate-QC starts by finding the mean event frequency \( m \) and variance \( v \) for each road segment type. For each tenth mile increment \( i \), the segment of size \( s \) is flagged as a hotspot if its event frequency is above a threshold \( t \) derived from a combination of \( m \) and \( v \).
Lastly, Kernel-Density is a linear version of the 2 dimensional HSID method discussed in the Background section. Kernel-Density is CARE’s only variable length HSID method. In this linear implementation each tenth mile segment at increment \( i \) is assigned a weight \( w \) using a kernel function \( k \) which makes use of other frequency values near segment \( i \). The kernel function \( k \) defines the number of segments near \( i \) to be used and the influence each segment has on \( w \). \( k \) may also specify that some event types exert more influence on \( w \) than others (fatal crashes may have more influence on \( w \) than property damage crashes). A combination of the mean \( m \) and variance \( v \) of all of the \( w \)’s is used to define a hotspot threshold \( t \). Given that the kernel uses the frequencies from neighboring segments when calculating \( w \), adjacent \( w \)’s will vary little. Hotspots are identified by grouping adjacent tenth mile segments that have \( w \)’s greater than \( t \). This variable length HSID method results in fewer contiguous hotspots that really belong together. The kernel’s weights also help compensate for the rounding error often associated with event reporting. In Figure 8 Kernel-Density hotspots are shown together with a curve that represents the weights created using a .7 mile kernel.
3.3 Second Order Hotspots

Second-order hotspots recommend where to position countermeasures to maximize or minimize something. For example, when a department decides to try to maximize alcohol related citations, second order hotspots would be used to identify the locations and times when alcohol citations have historically been the greatest per time period. Second order hotspots are characterized by their “per time period” nature. Second order hotspots could also be used to minimize a measure such as response time by positioning officers in historically “optimal” locations. Second order HSID methods require two logical data sources. The first data source, the event-log, stores the time and location of discrete events such as citations or crashes. The second data source, the cm-log, forms a continuous representation of countermeasures using their time spans and locations. Records within a cm-log contain periodic time and location samples. To illustrate, assume there is an event-log that contains alcohol related citations and a cm-log of patrols for the same time period. The cm-log can be used to find the total time spent patrolling each road segment and the event-log to find the number of alcohol related citations issued in the same segments. In this case, the second order hotspots are the segments with the greatest citation to time ratio. Data can be collected for the cm-log manually if historic patrol assignments are kept, or automatically by periodically logging officer’s locations while on patrol. For many countermeasures manual cm-log creation is the only option (traffic safety bill boards or roadside speed indicators).

3.4 Third Order Hotspots

Third order hotspot identification focuses on countermeasure assessment and feedback into the target selection process. Before locating third order hotspots, an effectiveness measure \( e \) must be calculated for each road segment. \( e \) represents the historical relationships between a specific countermeasure and a set of undesirable events. Third order hotspots are created by using \( e \) to narrow the results of a first or second order HSID method. Using this process, third order hotspots identify trouble areas where countermeasures are historically effective. Figure 1 lists three examples of third order hotspots. The first example, Citations vs. Crashes, identifies locations where crash frequencies are high and where citations are known to be an effective countermeasure. When designing a STEP, third-order hotspots can be used for both planning and evaluation. A STEP planner first provides a third order hotspot tool, a specific countermeasure and event target type. A TACT SETP planner may use reckless driving citations and CMV crashes. After collecting data from the study period there are two third order hotspot measures to evaluate. First, STEP is evaluated by looking for a change in CMV crash frequencies or other related events. Second, the hotspot selection is evaluated by calculating the effectiveness measure during the study period \( e_2 \) and comparing it to the historical \( e \).

3.5 Presentation

The appropriate presentation of hotspot varies by user. The HIT model identifies four distinct groups of users (researchers, planners, supervisors, officers). Each group of users will have one or more tools geared toward their goals. First, extensions of the current CARE interface will play a dominant role for researchers. It provides the most power and freedom but at the cost of complexity. Figures 4-8 illustrate CARE’s hotspot presentation. Second, officials need a planning tool to be used when orchestrating STEPs. Third, supervisors such as post commanders need a tool to plan patrols for a fixed area on a daily basis. Supervisors can choose to select patrol routes that target hotspots in proportion to their rank or rotate intense treatment to each of the hotspots.

The last, and largest, set of tools is designed for officers. First, CARE-Web is a simplified web-interface that uses the CARE engine and allows officers to map events and hot-spots using sets of predefined filters. Figure 9 shows a CARE Web map of alcohol related crashes. Each map or chart result in the CARE Web Application can be saved for repeat use as a gadget in Google, Yahoo, or a portal designed for E-Citation users, E-Citation Dashboard. Figure 10 shows a second officer tool, E-Citation Dashboard, with a map of crashes per county. These gadgets will allow the officer to quickly view events that occurred during the previous month, week, day or shift. The content of the maps are sent to the viewing tool via a GeoRSS feed. Third, is a hot-spot navigation system called Tour Guide. Tour Guide will provide audio commentary of approaching hot-spots as the officer drives or can be used to...
virtually navigate the roads while parked. Figure 11 shows a five mile segment of Interstate 59. The user can navigate to a road segment by clicking on the map in the upper left corner.

Figure 11: Tour guide

3.6 Assessment

A plan must be put in place to assess the tools’ effectiveness. The first metrics will address the following questions:

1) How often are the tools used (measured per tool)?
2) When the tools are used, are the recommendations followed?
3) When the tools are not used, how close is the actual patrol behavior to what would have been recommended?
4) How does the user rank the recommendations they receive?
5) How do event rates change (in and out of the target areas)?
6) How effective are countermeasures in each road segment? This is addressed by third order hotspots.

Lessons learned from these questions will be fed back into the process of defining first, second, and third order hotspots. The answers will also inform the design of the hotspot presentation.

4 Future Work

The HIT model will evolve as open issues are addressed and STEP software tools continue to be developed. Some of the open issues are:

- If shifts in targeted enforcement just move crashes from place to place, which hotspots should be targeted?
- What is the best means to calculate the effectiveness measure for third order hotspots?

Additionally, a route planning tool will be developed for the officer that includes approximate times and places to “sit”. The tool will use the officer’s current location, the current time, a destination-location and time, the patrol history, and sets of predefined hot-spots to recommend a route. When an officer’s plans change the tool will be able to update the suggested route accordingly.

5 Conclusions

The HIT model is designed to support activities of STEP program planners and evaluators. It relies on location and temporal data about historical events and countermeasures. First order hotspots identify locations with a high frequency of events. Second order hotspots identify locations with a high event count to countermeasure time ratio. Third order hotspots identify locations with a high frequency of events and with a history of good countermeasure effectiveness. The HIT model recognizes four types of users and the need for specific tools for each. Lastly, the model integrates assessment and feedback for future recommendations.

References


Dana Steil received his B.S. in Computer Science (1997) and MSE (2004) from Harding University, Searcy, Arkansas. He received his M.S. in Computer Science (2008) from the University of Alabama, Tuscaloosa. From 1997-2000 he worked as a software engineer at Alltel Information Services. Dana is currently an Assistant Professor of Computer Science at Harding University and a Ph.D. candidate at the University of Alabama. His primary area of research includes spatial analysis and hotspot identification.

Allen Parrish is a Professor of Computer Science at The University of Alabama. He is also the Director of the CARE Research and Development Laboratory (CRDL), a 25-person R&D organization dedicated to information technology applications to homeland security, public safety, traffic safety and court automation. CRDL has active projects in homeland security information sharing, public safety information sharing, traffic safety data analysis and electronic traffic citations. Dr. Parrish received a Ph.D. in Computer Science from The Ohio State University, Columbus, Ohio.