PREDICTING RISK FOR DEER-VEHICLE COLLISIONS USING A SOCIAL MEDIA BASED GEOGRAPHIC INFORMATION SYSTEM

G. Kent Webb, San Jose State University, Email: g.webb@sjsu.edu

ABSTRACT

As an experiment investigating social media as a data source for making management decisions, photo sharing websites were searched for data on deer sightings. Data about deer density and location are important factors in decisions related to herd management and transportation safety, but such data are often limited or not available. Results indicate that when combined with simple rules, data from photo sharing websites reliably predicted the location of road segments with high risk for deer-vehicle collisions as reported by volunteers to an internet site tracking roadkill. Use of Google Maps as the GIS platform was helpful in plotting and sharing data, measuring road segments and other distances, and overlaying geographical data. The ability to view satellite images and panoramic street views proved to be a particularly useful. As a general conclusion, the two independently collected sets of data from social media provided consistent information, suggesting investigative value to this data source. Overlaying two independently collected data sets can be a useful step in evaluating or mitigating reporting bias and human error in data taken from social media.

Keywords: Social Media, Geographical Information System (GIS), Decision Support System (DSS)

INTRODUCTION

Social media encourage the sharing of information that may be useful to support decision making. Many decisions suffer from the lack of good supporting data when the collection cost is prohibitively high. The use of social media as a data source, however, brings with it many questions about relevance and reliability. In addition to providing a potential source of low-cost data, social data collection can be used to engage the public. An example is the annual Christmas bird count conducted by the Audubon Society, the longest running Citizen Science survey in the world [1]. What was once a social activity involving the shooting of birds for sport was transformed into a social data collection project used by researchers studying bird populations. The bird count now relies on Facebook as one of its organizational tools. The experiment described in this paper was designed to investigate the usefulness of data collected from social networks in an application to a transportation safety and wildlife management issue: deer-vehicle collisions. Deer location and population density are an important, but hard to acquire data element of the problem. Flickr, a photo sharing site, was the source for most of the data on deer locations used in this study. The California Roadkill Observation System [4], relying on citizens to report the locations of deer-vehicle collisions into a collaborative format supported by Google Maps, was the source of data on deer-vehicle collisions.

There are up to 1.5 million deer-vehicle collisions in the United States every year, resulting in about 29,000 human injuries, 200 human fatalities, 1.3 million deer fatalities, and over a billion dollars in damages. Using its claims data, State Farm insurance estimated that there were 1.09 million deer-vehicle collisions in the United States between July 1, 2010 and June 30, 2011, a decline over the past few years [31]. Less than 0.02 percent of the collisions result in human fatalities and most of these do not directly result from the deer collision, but from collisions with other vehicles or loss of control in swerving to avoid the deer collision. Motorcyclists face a higher risk of fatality while other motorists face a lower risk given the relative size of their vehicle compared to the size of a deer. In spite of these significant societal costs, a recent study concludes that there are “relatively few well-designed studies that have evaluated techniques that can be used to reduce DVCs (deer-vehicle collisions)” [24, p. 80].
Developments in Social Networks and Geographical Information Systems (GIS)

The explosive increase in traffic on social networks has generated many new research and application opportunities. Liu [22] and others find that photo sharing websites provide support in disaster recovery, particularly Flickr which was launched in 2004. Kennedy and others experiment with procedures that combine Flickr photos, converting unstructured data into location-based knowledge [16]. Noting the significantly increased availability of self-reported geographical information available through network applications, Flanagin and Metzger [8] consider strategies to determine the credibility of volunteered geographical information, such as: peer-to-peer credibility assessments, social computing environments, and providing the name of the information author. Another study examines the use of Google maps to support volunteers collecting geographical information, finding support for what is known as ‘Linus’s Law: as the number of contributors increases, so does the quality [11].

Hudson-Smith and Crooks argue that “location and space are becoming increasingly important in the information technology revolution” so have created software tools to support online visualization based on Google Map Creator which can be used to create thematic maps that can be overlaid on Google Maps (MapTube) [13, p.118]. Li and Shao suggest that Google Earth and supporting technologies will create a new era for geo-information [21]. Boulos enthusiastically describes the use of Google Maps in developing an interactive GIS for England’s health care service [3].

Gu and Wang examine the use of Google Earth as a web based geographic information system decision support system, noting that “the rapid growth of Internet has significantly increased the quantity of data available … [9, p. 210].” Jankowski discusses the advantages of creating a simple GIS in ArcView™ enabling community decision making for environmental issues [15]. In an application to the Mojave desert tortoise, Darst [6] used a fuzzy logic based set of rules to develop a risk map as part of a decision support system to assist with environmental evaluation of development of solar energy facilities. In a 2011 study of animal-vehicle collisions in Spain, the researchers focus on analysis using Geographic Information Systems by inputting the coordinates of the crashes into a GIS that allows them to overlay other GIS data and so investigate contributing factors [7]. The next section provides a review of the literature on factors contributing to deer-vehicle collisions that were considered in building a deer-vehicle risk prediction map for the study area.

Risk Factors For Deer Vehicle Collisions, Literature Review

A Canadian study of deer-vehicle collisions using random locations in a geographic information system and a multivariate logistic regression found the highest risk road segments were those with high speed limits and load densities [25]. Using remotely sensed landscape data, a Swedish group of researchers found traffic volume, vehicle speed, and fences were the dominate factors in determining moose-vehicle collisions. Moose density measured by harvest statistics and distance to forest cover were also significant [30]. A study of Yellowstone National Park [10] concluded that vehicle speed was the primary factor contributing to vehicle-wildlife collisions. Sullivan examined 19 years of animal-vehicle collisions, finding that fatal crashes in darkness increased “by 2.3% for each mile-per-hour increase in speed” with a smaller but similar relation for non-fatal crashes [32, p. 9]. However, noting the widespread consensus that speed is correlated with deer-vehicle collisions, Bissonette and Kassar examine the relationship between annual average daily traffic, posted speed limit, and deer-vehicle collisions, finding no relationship. The authors suggest that while these may be useful explanatory variables, average traffic and posted speed limit, may be poor surrogate variables for actual road conditions so that the resulting risk predictions will be unreliable [2].

Deer population density and location may seem like obviously important factors influencing the likelihood of deer vehicle collisions, but detailed data is typically not available so surrogate variables are often used in research studies. For example, Schwabe, Schuhmann, Tonkovich, and Wu [29] used the county buck gun harvest per square mile and the doe gun harvest per square mile. They find statistical significance for bucks, more bucks more collisions, but not for does. In a later, similar study, however, several of the same authors (Schwabe, Schuman, and Tonkovich) report a positive relation between deer-vehicle collisions and the number of bucks harvested per square
Use of a county-wide surrogate variable for geographical data that will likely vary substantially over the region is one explanation for the contradictory results. Another problem is that the surrogate variables, the number of bucks and the number of does harvested per year, are likely to be very collinear. A common symptom of collinearity is that while both independent variables will have a similar relationship to the dependent variable, one of the independent variables will have the sign flipped (such as positive to negative), resulting in what appears to be a significant relationship, but which in fact disguises the true underlying relationship. In any event, having deer density and location data at a more localized level than at the county can be expected to improve analysis. Tappe and Enderle carefully eliminated all intercorrelated variables in their study of deer-vehicle collisions in Arkansas, finding that the most important factor was “the density of pasture edge within 1200 m of a location” [33, p. 496]. However, they did not investigate deer density or vehicle speed.

In developing a decision support system for management of deer-vehicle collisions, Huijser and others conclude that “even though the average costs for large ungulate-vehicle collisions per kilometer per year may not meet the thresholds of many of the mitigation measures, specific locations on a road can still exceed thresholds [14, p. 1].” Noting the importance of good predictive models to predict the best location for mitigation measures, Malo, Suarez, Diez divide highways into 1 km segments to investigate the relation between animal-vehicle collisions and found high collision rates were associated with areas having high forest cover or woodlands near the road [23]. A Virginia examining 228 small road segments of 500 meters found that aggregate measures of deer density and deer harvest levels were not correlated with the location of deer-vehicle collisions, but that traffic volume and road type were highly predictive. The authors measured deer density by driving along roads with a spotlight using distance sampling corrections described in the literature. They conclude “Our original hypotheses that DVCs in zones should negatively correlate with deer harvest and positively correlate with deer density were not supported” [26, p. 117] an important result since reduction in deer density is commonly used as tool to mitigate deer-vehicle collisions. The researchers noted that deer density is often estimated at the county level and so took care to use more local information; this data required significantly more effort to obtain. Their conclusion was that management should focus on driver behavior and road attributes, rather than deer density. Similar results were reported in a large-scale study in the United Kingdom. They concluded that “It immediately apparent, even from superficial examination of accident distribution maps … that areas of high frequency of DVCs (deer-vehicle collisions) are not simply related in any direct way to deer density [19, S. 27].” They also concluded that only fencing has been shown to consistently reduce collisions. Other studies have found fencing and wildlife crossings to be the most effective mitigation approach [27, 21].

Kolowski and Nielson examined the similarity between bobcat habitat and road habitat to create a technique for mapping wildlife-vehicle collision risk. The authors conclude: “Unlike other modeling techniques used to identify risk of road mortality, our method requires little field data collection and relies on readily available digital spatial data [18, p. 1119].”

Following these approaches, freeway segments in the study area were categorized as high risk if the segment was in an area that appeared to provide deer habitat and if the Flickr data confirmed that there were deer in the location that could reasonably get access to the freeway segment. A freeway segment where deer access would be constrained by more than several blocks of high density suburban development or other physical features was considered to free of deer ingress. High speeds associated with freeways are generally associated with high risk for deer-vehicle collisions and seem to increase the likelihood of human fatality, so freeways became the focus for this study.
RESEARCH METHODOLOGY

Beginning in the summer of 2010, photo sharing sites were searched for postings of deer located in the San Francisco Bay area. Data collection continued through fall, 2011. In all, 80 sightings of one or more deer were collected. An interactive Google map, available online, was created using a deer icon to show the location of each deer sighting. Clicking on the icon brings up a link to the deer image and information provided by the observer. The map appears in Figure 1.

Figure 1: Google Map with Icons Showing Location of Deer Sightings Reported on Photo Sharing, Social Networks

Freeways in the study area were divided into one mile segments for analysis. Segments of freeway that were within four miles of a reported deer sighting in an area unobstructed by dense suburban development (such as several blocks of high density streets and buildings) or other physical obstacles were designated as high risk, identified as
the red freeway segments in Figure 2. One advantage of using Google Maps for this analysis was the ability to “snap” a line on a map to a specific section of road, providing a very accurate measure of the distance of the road segment, no matter how many twists and turns. Distances are also easy to measure by drawing line segments on the map. Terrain analysis is greatly facilitated by features of Google Maps. Open space is clearly identified (green areas of Figures 1 and 2). Terrain can be examined using satellite imagery or detailed street level images.

The research hypothesis to be tested is as follows:

\[ H_1: \text{Deer collisions per mile of freeway are higher in the high risk segments, where the freeway runs through open space or low density development similar to deer habitat within range of deer that have been sighted and reported on a photo sharing site.} \]

Figure 2: Predicted High Risk Freeway Segments
RESULTS

The sample of 82 deer-vehicle collisions reported in the roadkill database were entered into a Google map KML file and overlaid on the predicted high risk map of Figure 2, allowing for a convenient visual inspection to determine which deer-vehicle collisions occurred on high risk freeway segments. Of the 82 reported collisions, 29 were reported to have occurred on freeways, 35.4 percent of the total.

Table 1, Total Deer Vehicle Collisions in Study Area

<table>
<thead>
<tr>
<th>Roadway Type</th>
<th>Number of Deer-Vehicle Collisions</th>
<th>Percentage Of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freeway</td>
<td>29</td>
<td>35.4 %</td>
</tr>
<tr>
<td>Other</td>
<td>53</td>
<td>64.6 %</td>
</tr>
<tr>
<td>Total</td>
<td>82</td>
<td>100.0 %</td>
</tr>
</tbody>
</table>

Data collected for the period December, 2010 through March, 2012 from the California Roadkill Observation system: http://wildlifecrossing.net/california

Table 2, Freeway Deer-Vehicle Collisions in Study Area

<table>
<thead>
<tr>
<th>Freeway Risk Category</th>
<th>Deer-Vehicle Collision per Mile</th>
<th>Number of Deer-Vehicle Collisions</th>
<th>Miles of Freeway</th>
<th>Percent of Deer-Vehicle Collisions</th>
<th>Proportion of Freeway Miles</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0.41818</td>
<td>23</td>
<td>55</td>
<td>79.3 %</td>
<td>10.4 %</td>
</tr>
<tr>
<td>Low</td>
<td>0.01265</td>
<td>6</td>
<td>474</td>
<td>20.7 %</td>
<td>89.6 %</td>
</tr>
<tr>
<td>Total</td>
<td>0.05482</td>
<td>29</td>
<td>529</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 3, Test of Hypothesis That Deer-Vehicle Collisions Per Mile of Freeway are Higher on One Mile Segments Identified as High Risk

<table>
<thead>
<tr>
<th></th>
<th>High Risk Freeway</th>
<th>Low Risk Freeway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Deer-Vehicle Collisions per Mile</td>
<td>0.418182</td>
<td>0.012658</td>
</tr>
<tr>
<td>Variance</td>
<td>0.470034</td>
<td>0.012524</td>
</tr>
<tr>
<td>Observations, Miles of Freeway</td>
<td>55</td>
<td>474</td>
</tr>
<tr>
<td>Maximum</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Null Hypothesis: No difference in the number of deer-vehicle collisions per mile in the high risk and the low risk freeway segments

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Degrees of freedom</td>
<td>54</td>
</tr>
<tr>
<td>t-stat</td>
<td>4.379884</td>
</tr>
<tr>
<td>P(T ≤ t) one tail</td>
<td>0.0000275 *</td>
</tr>
<tr>
<td>t-Critical value for α = .05</td>
<td>1.673565</td>
</tr>
</tbody>
</table>

* The null hypothesis is rejected at well above the 0.001 confidence level
As reported in Table 2, among the 29 deer-vehicle collisions on freeways, 23 occurred on the 55 miles of road predicted as high risk areas (Figure 2). Just 10.4 percent of the freeway miles accounted for 79.3 percent of the freeway collisions. The average number of collisions per mile of freeway on all segments was .0548. On high risk segments the per mile collision rate rises to 0.418, while on the low risk freeway segments the average per mile collision rate was 0.012 – a significant difference. The main research result is reported in Table 3. The null hypothesis of no difference in the frequency of collisions is rejected at a significance level well above the 1 percent level.

![Figure 3](image-url)

**Figure 3.** Overlay of all deer-vehicle collisions in the study area and the high risk freeway segments. Blue balloons identify collision locations. Google Map.

The results of the statistical test support the conclusion which is clear from a cursory inspection of the data; the basic prediction rule applied to the GIS significantly identified freeway segments at high risk for deer-vehicle collisions. Figure 3 is an overlay of all deer-vehicle collisions in the study area and the predicted high risk freeway segments. The statistical analysis was done only for deer-vehicle collisions that occurred on freeways. Although the KML file format used by Google Maps makes it easy to overlay maps and share data with other GIS programs, analytical programs such as Excel and Google spreadsheets cannot directly import a KML file, making statistical analysis somewhat cumbersome.
Google Maps does provide some very useful features to support visual analysis. The availability of easy to use satellite imagery allows the analyst to zoom in for closer inspection. Figure 4 shows satellite imagery for freeway segments on I-280 that had among the highest number of deer-vehicle collisions per segment. There are deer sighted in the area and noticeably this complex freeway segment is completely surrounded by deer habitat, consistent with the results of several studies identifying the proximity of habitat to roadways as a prime factor increasing collision likelihood.

![Figure 4](image.jpg)

**Figure 4.** High collision segment of I-280, deer habitat on both sides of freeway. Blue balloons mark deer-vehicle collisions. Image from Google Maps

Another useful visual analysis feature of Google Maps is street view, which provides a 360 degree, street level, panoramic imagery. The analyst is able to drill down to the specific collision locations illustrated in Figure 4 and verify the specific habitat and road design conditions.

As an example of using this feature, Figure 5 shows the street view of a section of US-1 with an unusually high number of collisions for the relatively small number of deer sighted in the area. The photo imagery clearly shows that thick vegetation is growing right beside the roadway, another factor identified in previous studies as increasing the likelihood of collision. It is thought that the vegetation blocks the view of both the deer and the driver, making it more difficult for either to evade collision.
By comparison, there were no reported collisions on the high risk freeway segment of I-280 shown in the street view of Figure 6. While many deer were sighted near the area, the street view shows that both sides of the freeway are clear of vegetation for a good distance and there is a fence along the side of the freeway. These two factors have been shown in many studies to reduce the likelihood of collisions. From inspection of the visual imagery, it appears the fence is less than 6 feet high, not high enough to prevent an adult deer to jump the fence and wind up on the freeway. The fence would likely prevent most does and young deer from accessing the freeway.

Figure 5. High collision segment on U.S.1, thick vegetation growing along the side of the road obstructs vision, but few deer sighted in the area. Image from Google Maps street view.

Figure 6. Although there were several deer spotted in nearby open space, no collisions were reported along this high risk segment of I-280 where there is no vegetation to block vision or attract deer, and which is fenced. Image from Google Maps street view.
CONCLUSIONS

Postings of deer sightings on the photo sharing websites successfully predicted freeway segments with high risk of deer-vehicle collisions as reported by volunteers in a citizen science project using the internet. The result encourages further effort using this approach. Overlaying the two independently collected data sources and finding statistical correlation corroborates each data source.

The data and the visual analysis suggest that variations in deer density are less important than terrain and road conditions in determining collision risk, a finding consistent with several previous studies. One advantage of the social media data is that it allows for very specific estimates of deer locations. Studies using average deer density over a large area will fail to account for the significant variations in density based on local geography.

The ready availability of satellite images and panoramic street views in Google Maps provides useful investigative tools. Many communities and organizations are publishing maps of deer-vehicle collisions that could be analyzed with the special features of Google Maps without the need to travel in order to gather location-specific information.

Other useful features of Google Maps as a GIS platform included the ability to snap lines to roadways and to measure distances. The advantages of sharing of information through Google Maps are particularly illustrated by the California Roadkill Observation System which allows users to precisely enter observed collision locations using the familiar Google Maps interface. One drawback to the use of Google Maps as the GIS is the relative difficulty of moving the data in and out of analytical programs.

In general, this research suggests that social media can be a useful source of data. As with any convenience sample, reporting bias presents a hazard; but several factors that have been identified as increasing the reliability of the data were present in this case: data were reported along with the name of the reporter, there were a large number of reporters, and there were two independent sources of data. The low cost and availability of this data source will also make it attractive to many researchers.

REFERENCES

4. California Roadkill Observation System: Available at: http://www.wildlifecrossing.net/california/
9. Gu, L. & Wang, J. A study of exploring the “Big Five” and task technology fit in web-based decision support systems. Issues in Information Systems X(2), 210-217


31. State Farm Insurance (2011). U.S. Deer-Vehicle Collisions Fall 7 Percent. Available at: