UC Berkeley
Research Reports

Title
Evaluation of the Accuracy of Global Positioning System Coordinates for Collision Locations in California

Permalink
https://escholarship.org/uc/item/2kh3z760

Authors
Bigham, John
Strang, Garrett
Oum, Sang Hyouk

Publication Date
2014-07-25

Peer reviewed
Evaluation of the Accuracy of Global Positioning System Coordinates for Collision Locations in California

Word Count: 3,874 words + 12 tables/figures (250 words each) = 6,874 words

Submission Date: July 25, 2014

John Bigham, MPH
University of California, Berkeley
Safe Transportation Research and Education Center
2614 Dwight Way #7374
Berkeley, CA 94720-7374
Phone: 510-643-1777
Fax: 510-643-9922
jbigham@berkeley.edu

Garrett Strang
University of California, Berkeley
Safe Transportation Research and Education Center
2614 Dwight Way #7374
Berkeley, CA 94720-7374
Phone: 510-643-1777
Fax: 510-643-9922
gwstrang@berkeley.edu

Sang Hyouk Oum, MS *
University of California, Berkeley
Safe Transportation Research and Education Center
2614 Dwight Way #7374
Berkeley, CA 94720-7374
Phone: 510-642-5553
Fax: 510-643-9922
shoum@berkeley.edu

* Corresponding author
ABSTRACT
Traffic collision reports typically provide descriptive locations indicating where a collision occurred and referencing the nearest intersection. Global Positioning System (GPS) technology can be used to provide latitude and longitude coordinates in addition to the descriptive location and many states now include GPS coordinates in collision reports. However, research has shown that there is potential for numerous errors when police agencies use GPS to complete traffic collision reports. In California, GPS coordinates have been included in the statewide collision database since 2006, but their overall accuracy has never been evaluated. The objective of this paper was to review the status of GPS coordinates in California collision data from 2009 to 2011 and to categorize types of errors or discrepancies that were exhibited, investigate error trends, and develop recommendations for use of the GPS coordinates. Instead of just classifying a GPS coordinate location as correct or incorrect, eleven categorizations were developed to better assess the breadth of differences between the GPS coordinate and descriptive location. Overall, 43% of GPS coordinates were categorized as correct, 2.5% were unknown, and the other 54.5% exhibited some type of discrepancy with the descriptive location. GPS coordinates located off the roadway were the most frequent error type, comprising nearly 20% of the sample, while systematic GPS errors such as truncated coordinates occurred 7% of the time. Accuracy appears to be improving over time, but it is recommended to thoroughly review the coordinate locations prior to conducting any spatial analyses.
INTRODUCTION
In the state of California, a total of 161,743 fatal and injury traffic collisions were reported in 2011 (1). For each collision, officers responding to the scene were required to complete a summary report, including a description of the location using the primary road and the distance and direction from the nearest secondary road. If the collision took place on a state highway, the description also identified the side of the highway on which it occurred, and whether or not it happened on an on/off ramp. In addition, since 2006, if the officer has access to a Global Positioning System (GPS) enabled device, he or she can record the latitude and longitude coordinates. GPS technology simplifies the process of utilizing collision data in Geographic Information Systems (GIS) software that many local, state, and federal agencies use for spatial analysis to help guide and inform engineering, policy, and educational initiatives.

The prevalence of GPS was a focal point of a 2007 study by Ogle (2) sponsored by the National Cooperative Highway Research Program that reviewed technologies for improving safety data. Ogle determined through a survey of 24 states that 11 incorporated GPS in their collision databases and that many others were contemplating its use. However, agencies should first consider the issues that can limit the accuracy of GPS coordinates. In the past, the type and quality of a GPS receiver could impact its accuracy and satellite locking time was a major concern, but these features have improved greatly in recent years (3, 4). Older handheld GPS receivers were difficult to use, lowering their effectiveness, however new handheld receivers provide much greater ease of use and offer acceptable accuracy (4, 5). One drawback of handheld receivers compared with more sophisticated in-vehicle systems equipped with an integrated mapping interface is that the handheld devices require users to manually transcribe coordinates instead of the coordinates being entered automatically (2). Some states have had much greater success using more expensive GPS receivers equipped with a GIS style mapping interface that allows officers to visually verify locations before entering the coordinates (2, 6, 7).

The more advanced integrated systems help prevent data transcription inaccuracies and reduce operator errors, the most commonly cited issues affecting GPS accuracy (4, 5). Ogle noted that in the absence of a completely automated system, obtaining a GPS coordinate and compiling the data into a central database is subject to human error (2). Sarasua et al. reviewed South Carolina collision data and identified several systematic data entry errors including truncating the number of decimals, inconsistent units and coordinate systems, transposed latitude and longitude coordinates, and other missing or erroneous values (4). In Kentucky, significant post-processing was necessary to correct improperly recorded GPS coordinates typically caused by operator error (6).

To better quantify the impact of the various errors in GPS accuracy in collision databases, full evaluations were conducted in South Carolina and Kentucky. Sarasua et al. found that in South Carolina, approximately 80% of GPS coordinates recorded in 2004 to 2006 were located within reasonable levels of accuracy (4). Green and Agent determined that among a sample of collisions recorded in Kentucky in 2003, 55% were accurately located (5). In 2011, the researchers conducted a follow-up study following the implementation by Kentucky police agencies of a new system with an integrated mapping interface and found that 92% of collisions were accurately located (6).

The focus of these studies was to determine the root cause of the inaccuracies and suggest methods for resolution, such as improved officer training and built-in accuracy checks during data entry. However, the practical perspective for third parties interested in utilizing collision data GPS coordinates for spatial analyses is lacking. The state of California now includes GPS
coordinates in many collision records, but does not offer guidance on how to use the coordinates. The objective of this paper is to evaluate collision data in California to: a) determine the overall accuracy of GPS coordinates, b) categorize types of errors or discrepancies, c) investigate error trends, and d) develop conclusions and recommendations for using GPS coordinates.

METHODOLOGY

About the Data
In California, all fatal and injury collision reports are entered into the Statewide Integrated Traffic Records System (SWITRS) database which is maintained by the California Highway Patrol (CHP) (8). The SWITRS database was updated in 2006 to allow inclusion of GPS coordinates. Over 90% of the GPS coordinates in SWITRS come from CHP officers, while the other 10% come from local allied reporting agencies. Between 2009 and 2011, approximately 30% of all collisions involving an injury included a GPS coordinate, as shown in Table 1.

TABLE 1 SWITRS Fatal and Injury Collisions

<table>
<thead>
<tr>
<th>Year</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collision Count</td>
<td>166,329</td>
<td>163,614</td>
<td>161,743</td>
</tr>
<tr>
<td>Include GPS Coordinates</td>
<td>40,341 (24.3%)</td>
<td>48,620 (29.7%)</td>
<td>52,381 (32.4%)</td>
</tr>
</tbody>
</table>

Figure 1 shows a map of California collisions that occurred between 2009 and 2011 plotted by their GPS coordinates. Obvious coordinate errors are clearly visible as locations that fall outside of state boundaries.
Methods

All SWITRS injury collisions that occurred between 2009 and 2011 and included a GPS coordinate were identified. Next, a random sample of 383 collisions from each year was extracted (1,149 total). The sample size of the collisions was determined by equation (1) and (2). Equation (1) was used to determine a sample size of the collisions which represent the total set of the collision data. Then the finite population correction equation (2) was used because of the known size of the dataset.

\[ n = \left( \frac{z_{\alpha/2}}{C} \right)^2 \left( \bar{p} \right) \left( 1 - \bar{p} \right) \]  

(1)
where, 
\[ n: \text{Required sample size for infinite population} \]
\[ z_{\alpha/2}: \text{Z value at alpha significance level} \]
\[ \alpha: \text{Significance level} \]
\[ C: \text{Confidence interval} \]
\[ \bar{p}: \text{An estimator of the accuracy} \]

\[
n^* = \frac{n}{1 + \frac{n-1}{\bar{p}}} \quad (2)
\]

where,
\[ n^*: \text{Required sample size for finite population} \]
\[ n: \text{Required sample size for infinite population} \]
\[ P: \text{Population} \]

The z value is 1.96 for significance level of 0.05. The confidence interval was set to 5% and the estimator of the accuracy to 0.5, which is the worst case (accurate or inaccurate) that creates a statistically significant sample representation at the 95% confidence interval.

The randomly selected set of collisions was imported into Google My Maps using the GPS coordinate. Based on the descriptive location associated with the collision, the collision location was adjusted. Table 2 shows an example of several collision records and the fields used to manually locate the collisions. If a collision did not occur on a state highway, as shown for the Edgewater Circle collision, the extra descriptive fields are excluded from the record. If necessary, county and city information could also be used to aid in location determination.

<table>
<thead>
<tr>
<th>TABLE 2 Sample Collision Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Road</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>RT 87</td>
</tr>
<tr>
<td>RT 118</td>
</tr>
<tr>
<td>EDGEOATER CIR</td>
</tr>
</tbody>
</table>

After adjusting each collision location, the distance from the original GPS coordinate was calculated for each pair. A 500-foot radius was suggested by Green and Agent to evaluate the accuracy of the GPS coordinates (5). After the initial review, however, it was clear that relying solely on distance would not fully describe the types of errors present in the sample. For example, a collision within 500 feet of the descriptive location, but 300 feet off the roadway in a wooded area could not be described as accurate. Therefore, collision locations were further analyzed using Google Earth to review possible trends and to determine a classification system.

**Classification System**

The classification system was developed to better quantify the differences beyond the location information simply being correct or incorrect. Each collision was assigned to one of the following eleven categories, which are described in more detail below:
Systematic GPS Error

Figure 1 illustrated many location records that could be categorized as systematic GPS errors. Record locations that appear in the ocean or in Nevada are easily identified as being incorrect. Figure 2 shows an example of a collision that occurred north of Sacramento, however, the GPS coordinate is located due south of the actual location and in the ocean. While the longitude is accurate, the latitude is off by several digits, resulting in the error. In some cases, latitude and longitude may both exhibit the same magnitude and direction of error, leading to a significant diagonal shift.

GPS Coordinate in Parking Lot

In some instances, GPS coordinates appeared to be located in parking lots. Figure 3 (a) and (b) shows the GPS coordinate in a CHP office parking lot, a substantial distance from the descriptive location. In other cases, the GPS coordinate was located in a parking lot in close proximity to the descriptive location.
Among the erroneous records, many GPS coordinates were located close to a descriptive location, but off the roadway. Figure 4 shows an example of a collision that occurred on Highway 101, but with a GPS coordinate located one hundred feet off the roadway.

GPS coordinates located at a different intersection, but in the vicinity of the intersection marked in the descriptive location were identified as intersection mismatches. For example, if the descriptive location of a collision was Main St & 1st St., but the GPS coordinate placed the collision at the intersection of Main St and 2nd St., that location would be classified as an intersection mismatch.
Offset Distance Mismatch
In SWITRS records, collisions that do not occur directly at an intersection include a specified offset distance and direction from the secondary road. In some instances, the descriptive location of a collision based on this distance may not match the GPS coordinate. For example, as shown in Figure 5, the descriptive location for collision 1 was located 500 feet further west than the GPS location.

Side of Highway Mismatch
If a collision takes place on a state highway, the description also identifies the side of the highway on which it occurred. In some cases, the GPS coordinate location and descriptive location are the same distance from an intersection, but on opposite sides of the highway. In Figure 5, collision 4 is an example of a GPS coordinate located on the eastbound side of Highway 10, while the descriptive location indicated that the collision occurred in the westbound direction.

Mainline/Ramp Mismatch
Collisions that occur on freeways are assigned a ramp or mainline designation in the SWITRS database. In Figure 5, collision 3 shows an example in which the GPS coordinate was located on the mainline of the freeway, while the descriptive location was in the middle of the off ramp.

Offset Direction Mismatch
In some cases the GPS coordinate and descriptive location appear to be the same distance from an intersection, but are offset in opposite directions. For example, in Figure 5, collision 2 was shown to have occurred on Highway 10, however the GPS coordinate is west of Crenshaw Blvd., while the descriptive location is east of Crenshaw Blvd.

Map data ©2014 Yahoo

FIGURE 5  Examples of Mismatch Error Categories.

Multiple Mismatch Issues
If a GPS coordinate exhibited two or more of the mismatch types identified in Figure 5, then it was categorized as a multiple mismatch.
**Unknown**
In some instances the descriptive location references an unidentifiable road or landmark (i.e., canals, streams, etc.). These locations cannot be verified and were thus categorized as unknown.

**Correct**
A GPS coordinate was deemed correct if the location was within 500 feet (154.2 meters) of the descriptive location and could not be classified into the other categories.

**RESULTS**
Table 3 shows the number of collision records classified by category and year. Overall, 43% of the GPS coordinates were categorized as correct, 2.5% were unknown, and the other 54.5% exhibited some type of discrepancy. GPS coordinates located off the roadway were the most frequent error type, comprising nearly 20% of the sample, followed by offset distance mismatch in 13% of the collisions, and systematic GPS errors in 7% of the collisions. Each of the other categorizations accounted for less than 4% of the sample.

**TABLE 3 Categorization Counts by Year**

<table>
<thead>
<tr>
<th>Category</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2009-2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic GPS error</td>
<td>31 (8.1%)</td>
<td>24 (6.3%)</td>
<td>27 (7%)</td>
<td>82 (7.1%)</td>
</tr>
<tr>
<td>GPS coordinate in parking lot</td>
<td>12 (3.1%)</td>
<td>11 (2.9%)</td>
<td>13 (3.4%)</td>
<td>36 (3.1%)</td>
</tr>
<tr>
<td>GPS coordinate off roadway</td>
<td>99 (25.8%)</td>
<td>94 (24.5%)</td>
<td>33 (8.6%)</td>
<td>226 (19.7%)</td>
</tr>
<tr>
<td>Intersection mismatch</td>
<td>0 (0%)</td>
<td>1 (0.3%)</td>
<td>4 (1.3%)</td>
<td>5 (0.4%)</td>
</tr>
<tr>
<td>Offset distance mismatch</td>
<td>37 (9.7%)</td>
<td>38 (9.9%)</td>
<td>71 (18.5%)</td>
<td>146 (12.7%)</td>
</tr>
<tr>
<td>Side of highway mismatch</td>
<td>23 (6%)</td>
<td>13 (3.4%)</td>
<td>5 (1.3%)</td>
<td>41 (3.6%)</td>
</tr>
<tr>
<td>Ramp/mainline mismatch</td>
<td>9 (2.3%)</td>
<td>17 (4.4%)</td>
<td>6 (1.6%)</td>
<td>32 (2.8%)</td>
</tr>
<tr>
<td>Offset direction mismatch</td>
<td>10 (2.6%)</td>
<td>8 (2.1%)</td>
<td>7 (1.8%)</td>
<td>25 (2.2%)</td>
</tr>
<tr>
<td>Multiple mismatch issues</td>
<td>9 (2.3%)</td>
<td>18 (4.7%)</td>
<td>6 (1.6%)</td>
<td>33 (2.9%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>12 (3.1%)</td>
<td>12 (3.1%)</td>
<td>5 (1.3%)</td>
<td>29 (2.5%)</td>
</tr>
<tr>
<td>Correct</td>
<td>141 (36.8%)</td>
<td>147 (38.4%)</td>
<td>206 (53.8%)</td>
<td>494 (43%)</td>
</tr>
<tr>
<td>Total</td>
<td>383</td>
<td>383</td>
<td>383</td>
<td>1149</td>
</tr>
</tbody>
</table>

Other categorization trends highlighted in Table 3 include the following:
- Systematic GPS errors and GPS coordinates in parking lots consistently represented 7% and 3%, respectively, of the samples from each year.
- GPS coordinates off the roadway accounted for 25% of the sample for 2009 and 2010, then dropped to 8.6% in 2011. This was the largest change within a single category and suggests improved precision of the GPS coordinates because the percentage of correct GPS coordinates increased from approximately 40% in 2009-2010 to 54% in 2011.
- The mismatch categories remained relatively consistent over the years with two exceptions. The side of highway mismatch dropped from 6% in 2009 to 1.3% in 2011, while the offset distance mismatch increased from 10% in 2009-2010 to 18.5% in 2011.

To better explain the magnitude of discrepancies in each category, Table 4 summarizes the distances between GPS coordinates and the descriptive locations. As expected, systematic
GPS errors exhibited the largest discrepancies, while correct locations showed the smallest discrepancies. Only the systematic GPS errors and the offset distance mismatches had minimum distances above the 500-foot threshold. The remaining categories showed one or more records within 500 feet, underscoring the need for categorizations rather than relying solely on distances to measure location accuracy. Obvious errors or mismatches were detected in many of the records even in cases in which the GPS coordinate was very close to the descriptive location.

Table 4 Distances from Descriptive Location for All Years Combined (2009 to 2011)

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Mean (feet)</th>
<th>Median (feet)</th>
<th>SD (feet)</th>
<th>Min (feet)</th>
<th>Max (feet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic GPS error</td>
<td>82</td>
<td>190,006</td>
<td>119,623</td>
<td>312,050</td>
<td>630</td>
<td>1,893,345</td>
</tr>
<tr>
<td>GPS coordinate in parking lot</td>
<td>36</td>
<td>13,708</td>
<td>394</td>
<td>26,096</td>
<td>44</td>
<td>106,608</td>
</tr>
<tr>
<td>GPS coordinate off roadway</td>
<td>226</td>
<td>3,181</td>
<td>291</td>
<td>12,193</td>
<td>25</td>
<td>112,978</td>
</tr>
<tr>
<td>Intersection mismatch</td>
<td>5</td>
<td>955</td>
<td>344</td>
<td>948</td>
<td>186</td>
<td>2,211</td>
</tr>
<tr>
<td>Offset distance mismatch</td>
<td>146</td>
<td>2,885</td>
<td>1,180</td>
<td>6,748</td>
<td>501</td>
<td>64,250</td>
</tr>
<tr>
<td>Side of highway mismatch</td>
<td>41</td>
<td>237</td>
<td>197</td>
<td>151</td>
<td>42</td>
<td>529</td>
</tr>
<tr>
<td>Ramp/mainline mismatch</td>
<td>32</td>
<td>349</td>
<td>220</td>
<td>375</td>
<td>78</td>
<td>2,083</td>
</tr>
<tr>
<td>Offset direction mismatch</td>
<td>25</td>
<td>850</td>
<td>748</td>
<td>762</td>
<td>64</td>
<td>3,476</td>
</tr>
<tr>
<td>Multiple mismatch issues</td>
<td>33</td>
<td>1,260</td>
<td>977</td>
<td>927</td>
<td>145</td>
<td>3,514</td>
</tr>
<tr>
<td>Unknown</td>
<td>29</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Correct</td>
<td>494</td>
<td>118</td>
<td>79</td>
<td>125</td>
<td>0</td>
<td>498</td>
</tr>
<tr>
<td>Total*</td>
<td>1,149</td>
<td>15,520</td>
<td>227</td>
<td>97,382</td>
<td>0</td>
<td>1,893,345</td>
</tr>
</tbody>
</table>

* Excludes unknown for all totals besides count

Table 5 shows the median distance by year for each category. The correct GPS coordinates consistently decreased from 114 feet in 2009 to 64 feet in 2011. However, results were mixed for other categories. For side of highway mismatch, ramp/mainline mismatch, and GPS coordinates off the roadway, the distances actually increased each year. The median distance of GPS coordinates not on the roadway exhibited the largest increase from approximately 250 feet in 2009-2010 to 1,240 feet in 2011. This means that although the number of records with GPS coordinates off the roadway decreased 66% from 2009 to 2011, the magnitude of the distance error was five times greater.
<table>
<thead>
<tr>
<th>Category</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2009-2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic GPS error</td>
<td>127,152.7</td>
<td>129,251.0</td>
<td>112,825.5</td>
<td>119,623.4</td>
</tr>
<tr>
<td>GPS coordinate in parking lot</td>
<td>4,807.2</td>
<td>286.5</td>
<td>2,428.9</td>
<td>394.0</td>
</tr>
<tr>
<td>GPS coordinate off roadway</td>
<td>250.2</td>
<td>262.0</td>
<td>1,239.9</td>
<td>291.7</td>
</tr>
<tr>
<td>Intersection mismatch</td>
<td>-</td>
<td>1,740.4</td>
<td>319.8</td>
<td>344.2</td>
</tr>
<tr>
<td>Offset distance mismatch</td>
<td>1,195.5</td>
<td>1,230.5</td>
<td>1,122.5</td>
<td>1,179.8</td>
</tr>
<tr>
<td>Side of highway mismatch</td>
<td>153.0</td>
<td>259.1</td>
<td>300.4</td>
<td>196.8</td>
</tr>
<tr>
<td>Ramp / mainline mismatch</td>
<td>204.0</td>
<td>219.6</td>
<td>548.1</td>
<td>219.8</td>
</tr>
<tr>
<td>Offset direction mismatch</td>
<td>891.2</td>
<td>815.6</td>
<td>310.4</td>
<td>748.2</td>
</tr>
<tr>
<td>Multiple mismatch issues</td>
<td>976.6</td>
<td>954.2</td>
<td>982.1</td>
<td>976.6</td>
</tr>
<tr>
<td>Correct</td>
<td>113.8</td>
<td>83.2</td>
<td>64.2</td>
<td>79.3</td>
</tr>
</tbody>
</table>

The error classifications and median distances for collision locations were also classified by road type, urban/rural designation, and several additional factors, but no significant trends were found. Several aspects of the systematic GPS errors and GPS coordinates in parking lots were further investigated using all available records with GPS coordinate information from 2009 to 2011.

Figure 6 shows two examples of systematic GPS errors that were evident using all available records. Figure 6 (a) shows collisions from a single jurisdiction in the San Francisco Bay Area where a large set was correctly located in the northern portion, compared with a group that was incorrectly located in the southern portion. The incorrect coordinates actually follow the same freeway pattern, but in the form of a smaller transposed version that differs by a common angle and distance from the correct locations. Figure 6 (b) shows another example of a systematic GPS error for all collisions occurring in San Diego County. Rather than the angled transposition seen in Figure 6 (a), these coordinates appear to have merely a skewed latitude or longitude value. These situations occurred in multiple jurisdictions throughout the state.
The evaluation also showed that 3% of the GPS coordinates for each year were located in parking lots. After further investigation, it was determined that these GPS coordinates were typically located in CHP office parking lots. Figure 7 shows GPS coordinates for collisions that were in a single CHP office parking lot. This was a recurring problem involving nearly every CHP parking lot throughout the state.
DISCUSSION
Location information in crash data is inherently imprecise, especially when officers are required to estimate long distances from the nearest intersection. For example, many collisions are recorded at a distance of 2,640 feet (a half mile) from the intersection, as an approximation for a range of distances. Using GPS can offer greater precision, however GPS coordinates for collision data still suffer from numerous types of inaccuracies reflected in the various error categorizations. A major outcome of this research was the inability to verify the accuracy of mismatch categories given the lack of true known collision locations. For example, in 2011, combining the 27% of mismatches with the 54% deemed correct would estimate GPS accuracy as high as 80%. However, the remaining 20% of errors still represents a significant problem that cannot simply be attributed to operator error.

A range of factors can affect GPS coordinate accuracy in the collision data. For example, we reviewed records by individual officers (determined by the badge number included for each collision in SWITRS), and found inconsistent results. In one case, an officer’s reports showed no systematic GPS errors in 2009, yet the same officer's error rate rose to 54% in 2010, then dropped to 4% in 2011. This suggests the presence of a problem such as possible equipment malfunction. There may also be different safety protocols that prevent officers from obtaining exact GPS coordinates in certain circumstances. If a collision occurs on a busy or narrow stretch of roadway, officers may need to enter coordinates from a distance. Officers may also be required to enter coordinates after returning to the station if unable to do so at the site. Another major known factor is the opportunity for GPS coordinates to be miscoded at any stage during the traffic collision report entry and submission process (9). Ogle cited this as a problem, and it is clear that this poses one of the biggest challenges to reporting accurate data (2).

Despite the deficiencies in GPS coordinate accuracy, the general trend from 2009 to 2011 showed steady improvement. If this trend continues, it would reduce the need for manual review of the coordinates. Improved equipment and officer training, in addition to automated data input and verification, are all among suggested procedures to prevent operator or systematic errors. As these recommendations are implemented, the quality of the data should continue to improve.
However, the question still remains whether current GPS coordinates can be confidently implemented in the state instead of relying on a traditional geocoding approach for locating collisions.

Geocoding is the process of translating a descriptive location to a latitude and longitude coordinate via street network reference data. This is the process typically used in GIS software to locate collisions based on an intersection description. Bigham et al. conducted a large-scale operation to geocode California collision data and determined that 91% of collisions could be geocoded, and that 97% of those locations were accurate (10). Although their research used a more flexible determination of accuracy (within 300 feet of descriptive location) and cannot account for any mismatch between the descriptive location and true location of the collision, it provides more reliable results than GPS coordinates. If one assumes that the mismatch categories included in Table 3 are correct, this still results in only 80% accuracy for the GPS coordinates; significantly lower than the 97% level of accuracy resulting from a geocoding process. Therefore, it is difficult to recommend the use of GPS coordinates without a quality control process, especially given the consistent annual 10% of collisions found to be systematically incorrect or erroneously located in parking lots. This could heavily impact any spatial analysis and lead to the identification of false collision clusters.

CONCLUSION

This work highlighted issues involving GPS coordinates in California collision data and the difficulties in properly assessing their accuracy without a known true collision location. In many cases, the descriptive location entered in SWITRS is only an estimate and cannot be assumed to be accurate. Although the GPS coordinate might be far from the descriptive location, it may also be the exact collision location. Without access to the original police reports, it is not possible to confirm the location. Regardless, if the GPS coordinate is several hundred feet from the nearest road, it cannot be deemed accurate even if it falls within a measurement threshold. It is also evident that systematic errors could be corrected to improve accuracy. GPS coordinate errors are typically attributed to operator or data entry error, which could be minimized through quality control procedures. However, for California collision data through 2011, it is recommended to thoroughly review GPS coordinates or apply traditional geocoding processes in GIS software to obtain revised coordinates before conducting spatial analyses.

ACKNOWLEDGEMENTS

Funding for this research was provided by a grant from the California Office of Traffic Safety through the National Highway Traffic Safety Administration. Special thanks to Grace Felschundneff for general editing and revisions.
REFERENCES


