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

1 **Evaluation of the Accuracy of Global Positioning System Coordinates for**
2 **Collision Locations in California**
3
4 **Word Count:** 3,874 words + 12 tables/figures (250 words each) = 6,874 words
5 **Submission Date:** July 25, 2014
6

7 **John Bigham, MPH**
8 University of California, Berkeley
9 Safe Transportation Research and Education Center
10 2614 Dwight Way #7374
11 Berkeley, CA 94720-7374
12 Phone: 510-643-1777
13 Fax: 510-643-9922
14 jbigham@berkeley.edu
15
16 **Garrett Strang**
17 University of California, Berkeley
18 Safe Transportation Research and Education Center
19 2614 Dwight Way #7374
20 Berkeley, CA 94720-7374
21 Phone: 510-643-1777
22 Fax: 510-643-9922
23 gwstrang@berkeley.edu
24
25 **Sang Hyouk Oum, MS ***
26 University of California, Berkeley
27 Safe Transportation Research and Education Center
28 2614 Dwight Way #7374
29 Berkeley, CA 94720-7374
30 Phone: 510-642-5553
31 Fax: 510-643-9922
32 shoum@berkeley.edu
33
34 * Corresponding author
35
36

1   **ABSTRACT**
2   Traffic collision reports typically provide descriptive locations indicating where a collision
3   occurred and referencing the nearest intersection. Global Positioning System (GPS) technology
4   can be used to provide latitude and longitude coordinates in addition to the descriptive location
5   and many states now include GPS coordinates in collision reports. However, research has shown
6   that there is potential for numerous errors when police agencies use GPS to complete traffic
7   collision reports. In California, GPS coordinates have been included in the statewide collision
8   database since 2006, but their overall accuracy has never been evaluated. The objective of this
9   paper was to review the status of GPS coordinates in California collision data from 2009 to 2011
10 and to categorize types of errors or discrepancies that were exhibited, investigate error trends,
11 and develop recommendations for use of the GPS coordinates. Instead of just classifying a GPS
12 coordinate location as correct or incorrect, eleven categorizations were developed to better assess
13 the breadth of differences between the GPS coordinate and descriptive location. Overall, 43% of
14 GPS coordinates were categorized as correct, 2.5% were unknown, and the other 54.5%
15 exhibited some type of discrepancy with the descriptive location. GPS coordinates located off the
16 roadway were the most frequent error type, comprising nearly 20% of the sample, while
17 systematic GPS errors such as truncated coordinates occurred 7% of the time. Accuracy appears
18 to be improving over time, but it is recommended to thoroughly review the coordinate locations
19 prior to conducting any spatial analyses.
20
21

1 **INTRODUCTION**
2 In the state of California, a total of 161,743 fatal and injury traffic collisions were reported in
3 2011 (*1*). For each collision, officers responding to the scene were required to complete a
4 summary report, including a description of the location using the primary road and the distance
5 and direction from the nearest secondary road. If the collision took place on a state highway, the
6 description also identified the side of the highway on which it occurred, and whether or not it
7 happened on an on/off ramp. In addition, since 2006, if the officer has access to a Global
8 Positioning System (GPS) enabled device, he or she can record the latitude and longitude
9 coordinates. GPS technology simplifies the process of utilizing collision data in Geographic
10 Information Systems (GIS) software that many local, state, and federal agencies use for spatial
11 analysis to help guide and inform engineering, policy, and educational initiatives.
12   The prevalence of GPS was a focal point of a 2007 study by Ogle (*2*) sponsored by the
13 National Cooperative Highway Research Program that reviewed technologies for improving
14 safety data. Ogle determined through a survey of 24 states that 11 incorporated GPS in their
15 collision databases and that many others were contemplating its use. However, agencies should
16 first consider the issues that can limit the accuracy of GPS coordinates. In the past, the type and
17 quality of a GPS receiver could impact its accuracy and satellite locking time was a major
18 concern, but these features have improved greatly in recent years (*3, 4*). Older handheld GPS
19 receivers were difficult to use, lowering their effectiveness, however new handheld receivers
20 provide much greater ease of use and offer acceptable accuracy (*4, 5*). One drawback of
21 handheld receivers compared with more sophisticated in-vehicle systems equipped with an
22 integrated mapping interface is that the handheld devices require users to manually transcribe
23 coordinates instead of the coordinates being entered automatically (*2*). Some states have had
24 much greater success using more expensive GPS receivers equipped with a GIS style mapping
25 interface that allows officers to visually verify locations before entering the coordinates (*2, 6, 7*).
26   The more advanced integrated systems help prevent data transcription inaccuracies and
27 reduce operator errors, the most commonly cited issues affecting GPS accuracy (*4, 5*). Ogle
28 noted that in the absence of a completely automated system, obtaining a GPS coordinate and
29 compiling the data into a central database is subject to human error (*2*). Sarasua et al. reviewed
30 South Carolina collision data and identified several systematic data entry errors including
31 truncating the number of decimals, inconsistent units and coordinate systems, transposed latitude
32 and longitude coordinates, and other missing or erroneous values (*4*). In Kentucky, significant
33 post-processing was necessary to correct improperly recorded GPS coordinates typically caused
34 by operator error (*6*).
35   To better quantify the impact of the various errors in GPS accuracy in collision databases,
36 full evaluations were conducted in South Carolina and Kentucky. Sarasua et al. found that in
37 South Carolina, approximately 80% of GPS coordinates recorded in 2004 to 2006 were located
38 within reasonable levels of accuracy (*4*). Green and Agent determined that among a sample of
39 collisions recorded in Kentucky in 2003, 55% were accurately located (*5*). In 2011, the
40 researchers conducted a follow-up study following the implementation by Kentucky police
41 agencies of a new system with an integrated mapping interface and found that 92% of collisions
42 were accurately located (*6*).
43   The focus of these studies was to determine the root cause of the inaccuracies and suggest
44 methods for resolution, such as improved officer training and built-in accuracy checks during
45 data entry. However, the practical perspective for third parties interested in utilizing collision
46 data GPS coordinates for spatial analyses is lacking. The state of California now includes GPS

1   coordinates in many collision records, but does not offer guidance on how to use the coordinates.
2   The objective of this paper is to evaluate collision data in California to: a) determine the overall
3   accuracy of GPS coordinates, b) categorize types of errors or discrepancies, c) investigate error
4   trends, and d) develop conclusions and recommendations for using GPS coordinates.
5
6   **METHODOLOGY**
7
8   **About the Data**
9   In California, all fatal and injury collision reports are entered into the Statewide Integrated
10 Traffic Records System (SWITRS) database which is maintained by the California Highway
11 Patrol (CHP) (*8*). The SWITRS database was updated in 2006 to allow inclusion of GPS
12 coordinates. Over 90% of the GPS coordinates in SWITRS come from CHP officers, while the
13 other 10% come from local allied reporting agencies. Between 2009 and 2011, approximately
14 30% of all collisions involving an injury included a GPS coordinate, as shown in Table 1.
15
16 **TABLE 1 SWITRS Fatal and Injury Collisions**

Year	2009	2010	2011
Collision Count	166,329	163,614	161,743
Include GPS Coordinates	40,341 (24.3%)	48,620 (29.7%)	52,381 (32.4%)

17
18   Figure 1 shows a map of California collisions that occurred between 2009 and 2011
19 plotted by their GPS coordinates. Obvious coordinate errors are clearly visible as locations that
20 fall outside of state boundaries.
21
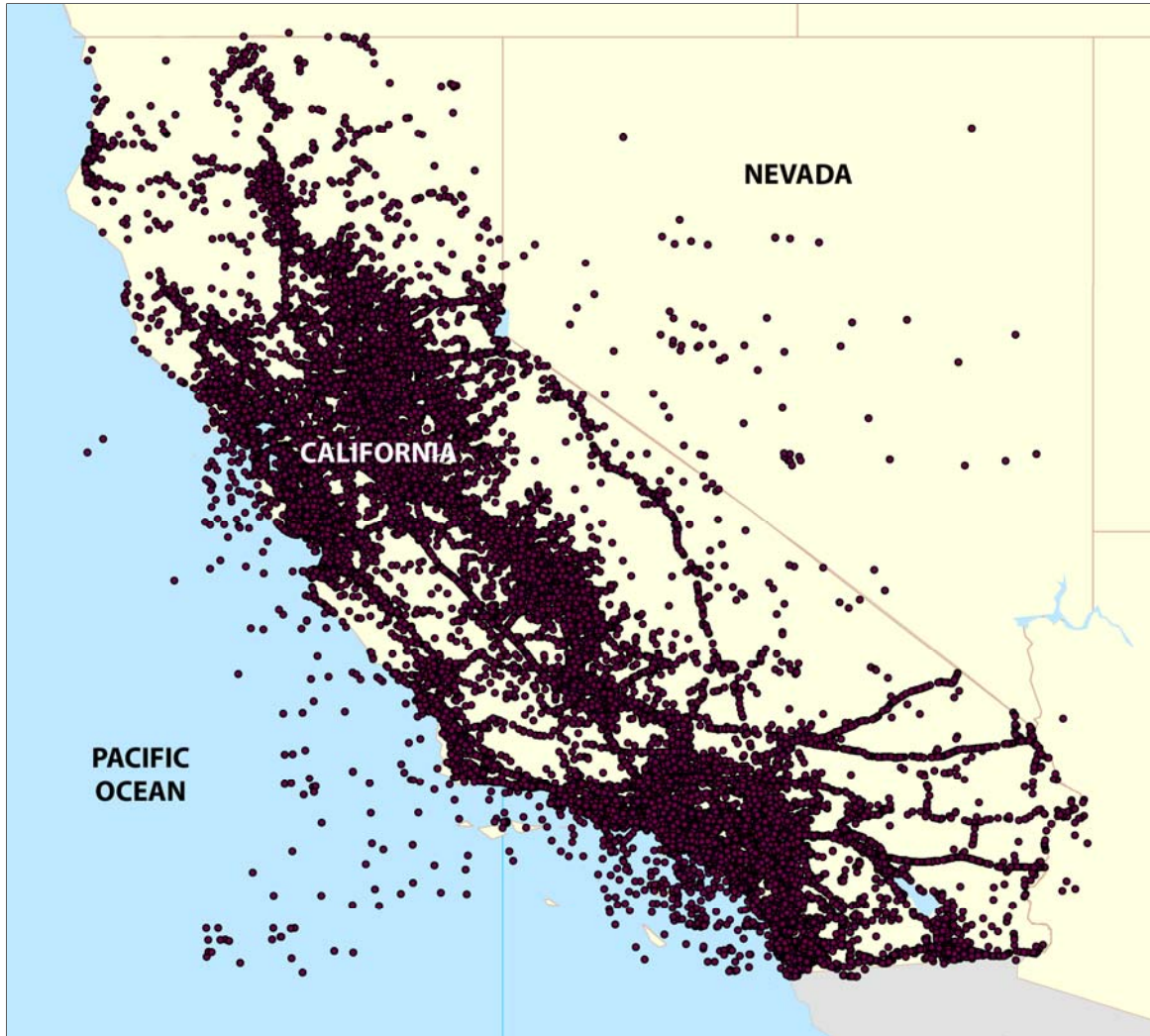
Map data ©2014 ESRI
FIGURE 1 SWITRS Collisions From 2009 to 2011 Plotted by GPS Coordinates.

1
2 **Methods**
3 All SWITRS injury collisions that occurred between 2009 and 2011 and included a GPS
4 coordinate were identified. Next, a random sample of 383 collisions from each year was
5 extracted (1,149 total). The sample size of the collisions was determined by equation (1) and (2).
6 Equation (1) was used to determine a sample size of the collisions which represent the total set of
7 the collision data. Then the finite population correction equation (2) was used because of the
8 known size of the dataset.
9

$$n = \left(z_{\alpha/2}/C\right)^2 (\bar{p})(1 - \bar{p}) \qquad (1)$$

10
11

1   where,
2   n: Required sample size for infinite population
3            $z_{\alpha/2}$: Z value at alpha significance level
4   α: Significance level
5   C: Confidence interval
6            $\bar{p}$: An estimator of the accuracy
7

$$n^* = \frac{n}{1 + \dfrac{n-1}{P}} \qquad\qquad (2)$$

8
9   where,
10   n^*: Required sample size for finite population
11   n : Required sample size for infinite population
12   P : Population
13
14   The z value is 1.96 for significance level of 0.05. The confidence interval was set to 5%
15   and the estimator of the accuracy to 0.5, which is the worst case (accurate or inaccurate) that
16   creates a statistically significant sample representation at the 95% confidence interval.
17   The randomly selected set of collisions was imported into Google My Maps using the
18   GPS coordinate. Based on the descriptive location associated with the collision, the collision
19   location was adjusted. Table 2 shows an example of several collision records and the fields used
20   to manually locate the collisions. If a collision did not occur on a state highway, as shown for the
21   Edgewater Circle collision, the extra descriptive fields are excluded from the record. If necessary,
22   county and city information could also be used to aid in location determination.
23
24   **TABLE 2 Sample Collision Records**

Primary Road	Secondary Road	Distance	Direction	State Highway	Side of Highway	Location Type
RT 87	RT 280	300	N	Y	N	Ramp
RT 118	RT 5	77	E	Y	E	Mainline
EDGEWATER CIR	ERLE RD	92	N	N	-	-

25
26   After adjusting each collision location, the distance from the original GPS coordinate was
27   calculated for each pair. A 500-foot radius was suggested by Green and Agent to evaluate the
28   accuracy of the GPS coordinates *(5)*. After the initial review, however, it was clear that relying
29   solely on distance would not fully describe the types of errors present in the sample. For example,
30   a collision within 500 feet of the descriptive location, but 300 feet off the roadway in a wooded
31   area could not be described as accurate. Therefore, collision locations were further analyzed
32   using Google Earth to review possible trends and to determine a classification system.
33
34   **Classification System**
35   The classification system was developed to better quantify the differences beyond the location
36   information simply being correct or incorrect. Each collision was assigned to one of the
37   following eleven categories, which are described in more detail below:
38

1   • Systematic GPS error
2   • GPS coordinate in parking lot
3   • GPS coordinate off roadway
4   • Offset distance mismatch
5   • Offset direction mismatch
6   • Mainline/ramp mismatch
7   • Side of highway mismatch
8   • Intersection mismatch
9   • Multiple mismatch issues
10   • Unknown
11   • Correct
12
13   *Systematic GPS Error*
14   Figure 1 illustrated many location records that could be categorized as systematic GPS errors.
15   Record locations that appear in the ocean or in Nevada are easily identified as being incorrect.
16   Figure 2 shows an example of a collision that occurred north of Sacramento, however, the GPS
17   coordinate is located due south of the actual location and in the ocean. While the longitude is
18   accurate, the latitude is off by several digits, resulting in the error. In some cases, latitude and
19   longitude may both exhibit the same magnitude and direction of error, leading to a significant
20   diagonal shift.
21
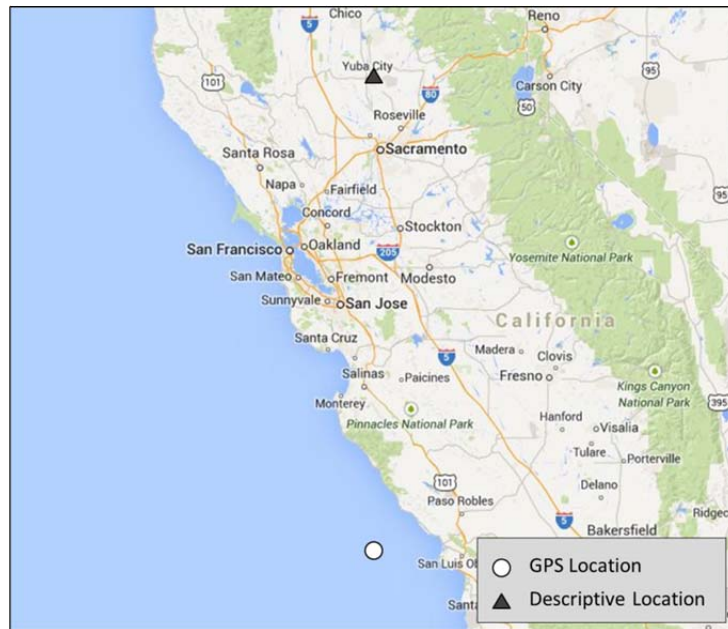
Map data ©2014 Google
FIGURE 2 Systematic GPS Error Example.

22
23   *GPS Coordinate in Parking Lot*
24   In some instances, GPS coordinates appeared to be located in parking lots. Figure 3 (a) and (b)
25   shows the GPS coordinate in a CHP office parking lot, a substantial distance from the descriptive
26   location. In other cases, the GPS coordinate was located in a parking lot in close proximity to the
27   descriptive location.
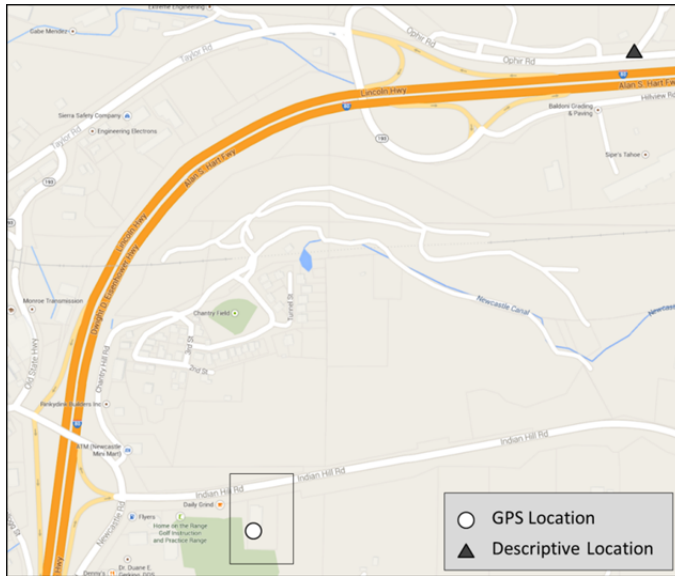
(a) GPS Location and Descriptive Location
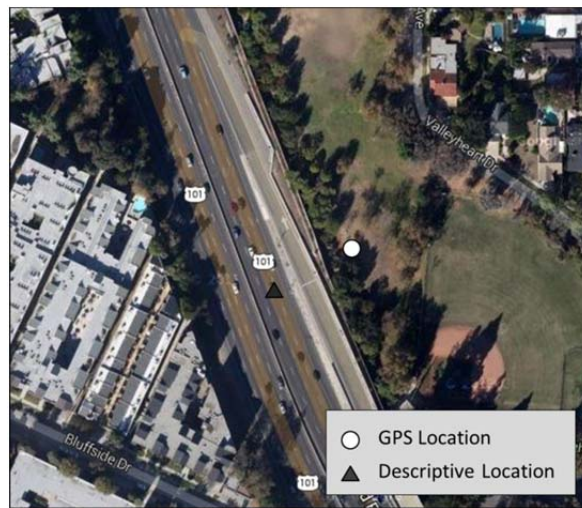Map data ©2014 Google

(b) Enlarged View of GPS Location
Imagery data ©2014 Google

FIGURE 3 An Example of GPS Coordinate in Parking Lot.

1
2   *GPS Coordinate Off Roadway*
3   Among the erroneous records, many GPS coordinates were located close to a descriptive location,
4   but off the roadway. Figure 4 shows an example of a collision that occurred on Highway 101, but
5   with a GPS coordinate located one hundred feet off the roadway.
6

Imagery data ©2014 Google
FIGURE 4 An example of GPS Coordinate Off Roadway.

7
8   *Intersection Mismatch*
9   GPS coordinates located at a different intersection, but in the vicinity of the intersection marked
10 in the descriptive location were identified as intersection mismatches. For example, if the
11 descriptive location of a collision was Main St & 1st St., but the GPS coordinate placed the
12 collision at the intersection of Main St and 2nd St., that location would be classified as an
13 intersection mismatch.

1   *Offset Distance Mismatch*
2   In SWITRS records, collisions that do not occur directly at an intersection include a specified
3   offset distance and direction from the secondary road. In some instances, the descriptive location
4   of a collision based on this distance may not match the GPS coordinate. For example, as shown
5   in Figure 5, the descriptive location for collision 1 was located 500 feet further west than the
6   GPS location.
7
8   *Side of Highway Mismatch*
9   If a collision takes place on a state highway, the description also identifies the side of the
10   highway on which it occurred. In some cases, the GPS coordinate location and descriptive
11   location are the same distance from an intersection, but on opposite sides of the highway. In
12   Figure 5, collision 4 is an example of a GPS coordinate located on the eastbound side of
13   Highway 10, while the descriptive location indicated that the collision occurred in the westbound
14   direction.
15
16   *Mainline/Ramp Mismatch*
17   Collisions that occur on freeways are assigned a ramp or mainline designation in the SWITRS
18   database. In Figure 5, collision 3 shows an example in which the GPS coordinate was located on
19   the mainline of the freeway, while the descriptive location was in the middle of the off ramp.
20
21   *Offset Direction Mismatch*
22   In some cases the GPS coordinate and descriptive location appear to be the same distance from
23   an intersection, but are offset in opposite directions. For example, in Figure 5, collision 2 was
24   shown to have occurred on Highway 10, however the GPS coordinate is west of Crenshaw Blvd.,
25   while the descriptive location is east of Crenshaw Blvd.
26
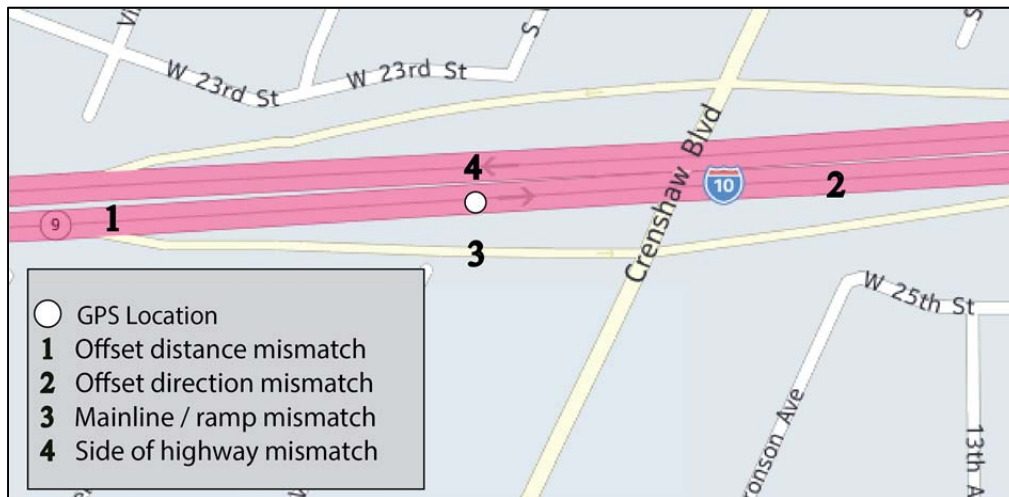
Map data ©2014 Yahoo
FIGURE 5 Examples of Mismatch Error Categories.

27
28   *Multiple Mismatch Issues*
29   If a GPS coordinate exhibited two or more of the mismatch types identified in Figure 5, then it
30   was categorized as a multiple mismatch.
31

1 *Unknown*
2 In some instances the descriptive location references an unidentifiable road or landmark (i.e.,
3 canals, streams, etc.). These locations cannot be verified and were thus categorized as unknown.
4
5 *Correct*
6 A GPS coordinate was deemed correct if the location was within 500 feet (154.2 meters) of the
7 descriptive location and could not be classified into the other categories.
8
9 **RESULTS**
10 Table 3 shows the number of collision records classified by category and year. Overall, 43% of
11 the GPS coordinates were categorized as correct, 2.5% were unknown, and the other 54.5%
12 exhibited some type of discrepancy. GPS coordinates located off the roadway were the most
13 frequent error type, comprising nearly 20% of the sample, followed by offset distance mismatch
14 in 13% of the collisions, and systematic GPS errors in 7% of the collisions. Each of the other
15 categorizations accounted for less than 4% of the sample.
16
17 **TABLE 3 Categorization Counts by Year**

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

18
19 Other categorization trends highlighted in Table 3 include the following:
20 • Systematic GPS errors and GPS coordinates in parking lots consistently represented 7%
21 and 3%, respectively, of the samples from each year.
22 • GPS coordinates off the roadway accounted for 25% of the sample for 2009 and 2010,
23 then dropped to 8.6% in 2011. This was the largest change within a single category and
24 suggests improved precision of the GPS coordinates because the percentage of correct
25 GPS coordinates increased from approximately 40% in 2009-2010 to 54% in 2011.
26 • The mismatch categories remained relatively consistent over the years with two
27 exceptions. The side of highway mismatch dropped from 6% in 2009 to 1.3% in 2011,
28 while the offset distance mismatch increased from 10% in 2009-2010 to 18.5% in 2011.
29
30   To better explain the magnitude of discrepancies in each category, Table 4 summarizes
31 the distances between GPS coordinates and the descriptive locations. As expected, systematic

1 GPS errors exhibited the largest discrepancies, while correct locations showed the smallest
2 discrepancies. Only the systematic GPS errors and the offset distance mismatches had minimum
3 distances above the 500-foot threshold. The remaining categories showed one or more records
4 within 500 feet, underscoring the need for categorizations rather than relying solely on distances
5 to measure location accuracy. Obvious errors or mismatches were detected in many of the
6 records even in cases in which the GPS coordinate was very close to the descriptive location.
7
8 **TABLE 4 Distances from Descriptive Location for All Years Combined (2009 to 2011)**

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

9 * Excludes unknown for all totals besides count
10
11   Table 5 shows the median distance by year for each category. The correct GPS
12 coordinates consistently decreased from 114 feet in 2009 to 64 feet in 2011. However, results
13 were mixed for other categories. For side of highway mismatch, ramp/mainline mismatch, and
14 GPS coordinates off the roadway, the distances actually increased each year. The median
15 distance of GPS coordinates not on the roadway exhibited the largest increase from
16 approximately 250 feet in 2009-2010 to 1,240 feet in 2011. This means that although the number
17 of records with GPS coordinates off the roadway decreased 66% from 2009 to 2011, the
18 magnitude of the distance error was five times greater.
19
20

1   **TABLE 5 Median Distance (feet) From Descriptive Location By Categorization**

Category	2009	2010	2011	2009-2011
Systematic GPS error	127,152.7	129,251.0	112,825.5	119,623.4
GPS coordinate in parking lot	4,807.2	286.5	2,428.9	394.0
GPS coordinate off roadway	250.2	262.0	1,239.9	291.7
Intersection mismatch	-	1,740.4	319.8	344.2
Offset distance mismatch	1,195.5	1,230.5	1,122.5	1,179.8
Side of highway mismatch	153.0	259.1	300.4	196.8
Ramp / mainline mismatch	204.0	219.6	548.1	219.8
Offset direction mismatch	891.2	815.6	310.4	748.2
Multiple mismatch issues	976.6	954.2	982.1	976.6
Correct	113.8	83.2	64.2	79.3

2
3   The error classifications and median distances for collision locations were also classified
4   by road type, urban/rural designation, and several additional factors, but no significant trends
5   were found. Several aspects of the systematic GPS errors and GPS coordinates in parking lots
6   were further investigated using all available records with GPS coordinate information from 2009
7   to 2011.
8   Figure 6 shows two examples of systematic GPS errors that were evident using all
9   available records. Figure 6 (a) shows collisions from a single jurisdiction in the San Francisco
10   Bay Area where a large set was correctly located in the northern portion, compared with a group
11   that was incorrectly located in the southern portion. The incorrect coordinates actually follow the
12   same freeway pattern, but in the form of a smaller transposed version that differs by a common
13   angle and distance from the correct locations. Figure 6 (b) shows another example of a
14   systematic GPS error for all collisions occurring in San Diego County. Rather than the angled
15   transposition seen in Figure 6 (a), these coordinates appear to have merely a skewed latitude or
16   longitude value. These situations occurred in multiple jurisdictions throughout the state.
17
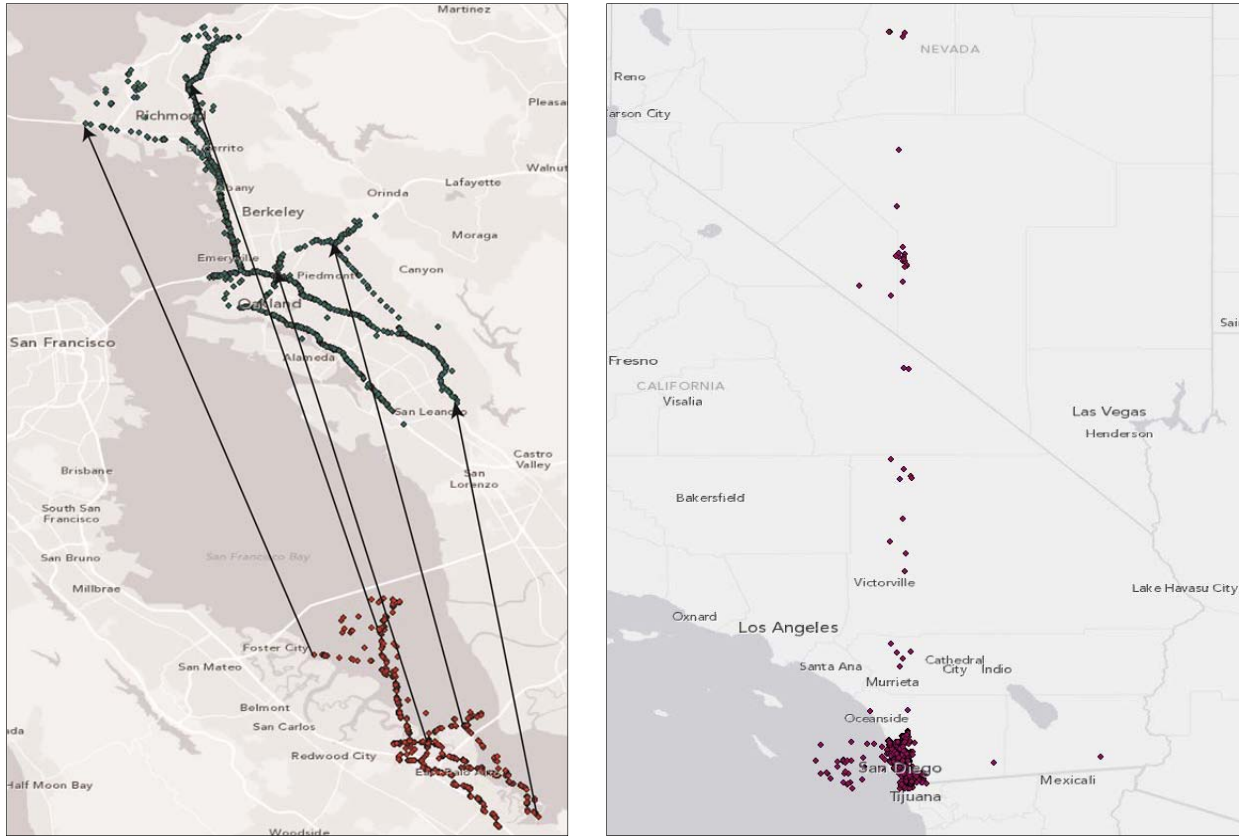
(a) Transposed Locations in the Bay Area   (b) Skewed Coordinates in San Diego County

Map data ©2014 ESRI

FIGURE 6 Additional Systematic GPS Errors.

1
2   The evaluation also showed that 3% of the GPS coordinates for each year were located in
3   parking lots. After further investigation, it was determined that these GPS coordinates were
4   typically located in CHP office parking lots. Figure 7 shows GPS coordinates for collisions that
5   were in a single CHP office parking lot. This was a recurring problem involving nearly every
6   CHP parking lot throughout the state.
7

Imagery data ©2014 Google
FIGURE 7 GPS Coordinates Located in a CHP Office Parking Lot.

1
2   **DISCUSSION**
3   Location information in crash data is inherently imprecise, especially when officers are required
4   to estimate long distances from the nearest intersection. For example, many collisions are
5   recorded at a distance of 2,640 feet (a half mile) from the intersection, as an approximation for a
6   range of distances. Using GPS can offer greater precision, however GPS coordinates for collision
7   data still suffer from numerous types of inaccuracies reflected in the various error categorizations.
8   A major outcome of this research was the inability to verify the accuracy of mismatch categories
9   given the lack of true known collision locations. For example, in 2011, combining the 27% of
10 mismatches with the 54% deemed correct would estimate GPS accuracy as high as 80%.
11 However, the remaining 20% of errors still represents a significant problem that cannot simply
12 be attributed to operator error.
13   A range of factors can affect GPS coordinate accuracy in the collision data. For example,
14 we reviewed records by individual officers (determined by the badge number included for each
15 collision in SWITRS), and found inconsistent results. In one case, an officer's reports showed no
16 systematic GPS errors in 2009, yet the same officer's error rate rose to 54% in 2010, then
17 dropped to 4% in 2011. This suggests the presence of a problem such as possible equipment
18 malfunction. There may also be different safety protocols that prevent officers from obtaining
19 exact GPS coordinates in certain circumstances. If a collision occurs on a busy or narrow stretch
20 of roadway, officers may need to enter coordinates from a distance. Officers may also be
21 required to enter coordinates after returning to the station if unable to do so at the site. Another
22 major known factor is the opportunity for GPS coordinates to be miscoded at any stage during
23 the traffic collision report entry and submission process (*9*). Ogle cited this as a problem, and it
24 is clear that this poses one of the biggest challenges to reporting accurate data (*2*).
25   Despite the deficiencies in GPS coordinate accuracy, the general trend from 2009 to 2011
26 showed steady improvement. If this trend continues, it would reduce the need for manual review
27 of the coordinates. Improved equipment and officer training, in addition to automated data input
28 and verification, are all among suggested procedures to prevent operator or systematic errors. As
29 these recommendations are implemented, the quality of the data should continue to improve.

1   However, the question still remains whether current GPS coordinates can be confidently
2   implemented in the state instead of relying on a traditional geocoding approach for locating
3   collisions.
4   Geocoding is the process of translating a descriptive location to a latitude and longitude
5   coordinate via street network reference data. This is the process typically used in GIS software to
6   locate collisions based on an intersection description. Bigham et al. conducted a large-scale
7   operation to geocode California collision data and determined that 91% of collisions could be
8   geocoded, and that 97% of those locations were accurate (*10*). Although their research used a
9   more flexible determination of accuracy (within 300 feet of descriptive location) and cannot
10   account for any mismatch between the descriptive location and true location of the collision, it
11   provides more reliable results than GPS coordinates. If one assumes that the mismatch categories
12   included in Table 3 are correct, this still results in only 80% accuracy for the GPS coordinates;
13   significantly lower than the 97% level of accuracy resulting from a geocoding process. Therefore,
14   it is difficult to recommend the use of GPS coordinates without a quality control process,
15   especially given the consistent annual 10% of collisions found to be systematically incorrect or
16   erroneously located in parking lots. This could heavily impact any spatial analysis and lead to the
17   identification of false collision clusters.
18
19   **CONCLUSION**
20   This work highlighted issues involving GPS coordinates in California collision data and the
21   difficulties in properly assessing their accuracy without a known true collision location. In many
22   cases, the descriptive location entered in SWITRS is only an estimate and cannot be assumed to
23   be accurate. Although the GPS coordinate might be far from the descriptive location, it may also
24   be the exact collision location. Without access to the original police reports, it is not possible to
25   confirm the location. Regardless, if the GPS coordinate is several hundred feet from the nearest
26   road, it cannot be deemed accurate even if it falls within a measurement threshold. It is also
27   evident that systematic errors could be corrected to improve accuracy. GPS coordinate errors are
28   typically attributed to operator or data entry error, which could be minimized through quality
29   control procedures. However, for California collision data through 2011, it is recommended to
30   thoroughly review GPS coordinates or apply traditional geocoding processes in GIS software to
31   obtain revised coordinates before conducting spatial analyses.
32
33   **ACKNOWLEDGEMENTS**
37
38

REFERENCES

1. California Highway Patrol. *2011 California Quick Collision Facts CHP SWITRS Quick Facts*. www.chp.ca.gov/switrs/pdf/2011-quick.pdf. Accessed May 30, 2014.

2. Ogle, J. H. (2007). *Technologies for Improving Safety Data*. NCHRP Synthesis 367. Transportation Research Board, Washington, D.C, 2007.

3. Sando, T., Mussa, R., Sotter, R., Sobanjo, J., and Spainhour, L. The Potential of GPS Receivers for Improving the Accuracy of Crash Location Data. Presented at the *30th International Traffic Records Forum*, Nashville, TN, 2004.

4. Sarasua, W. A., Ogle, J. H., and Geoghegan, K. (2008). Use of Global Positioning System to Identify Crash Locations in South Carolina. *Transportation Research Record: Journal of the Transportation Research Board*, *2064*(1), Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 43-50.

5. Green, E. R., and Agent, K. R. *Evaluation of the Accuracy of GPS as a Method of Locating Traffic Collisions*. Kentucky Transportation Center, 2004.

6. Green, E. R., and Agent, K. R. Evaluation of the Locations of Kentucky's Traffic Crash Data. In *3rd International Conference on Road Safety and Simulation*, Indianapolis, IN, 2011.

7. Havlicek, J. P., and Barnes, R. D. *GPS Location Data Enhancement in Electronic Traffic Records*. Oklahoma Transportation Center, 2013.

8. Statewide Integrated Traffic Records System. *California Highway Patrol*. www.iswitrs.chp.ca.gov. Accessed May 30, 2014.

9. Bigham, J., and Oum, S. (2014). *Location Error Consideration in Geocoding Police Reported Collision Data in California*. Working paper. University of California, Berkeley, 2014.

10. Bigham, J., Rice, T., Pande, S., Lee, J., Park, S., Gutierrez, N., and Ragland, D. Geocoding Police Collision Report Data from California: a Comprehensive Approach. *International Journal of Health Geographics*, Vol 8(1), 2009, pp. 72.