# EVALUATION OF THE LOCATIONS OF KENTUCKY'S TRAFFIC CRASH DATA

Eric R. Green
Transportation Research Engineer, Kentucky Transportation Center,
Lexington KY, USA, e-mail: egreen@engr.uky.edu

Kenneth R. Agent
Transportation Research Engineer, Kentucky Transportation Center,
Lexington KY, USA, e-mail: kagent@engr.uky.edu

*Submitted to the 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA*

## ABSTRACT

This research study evaluated the current accuracy of the location information provided in Kentucky's crash data. Since the year 2000, the Kentucky Open Portal System's (KYOPS) eCrash form has included global positioning system (GPS) data as well as the more traditional County, Route and Milepoint data (CRMP). A previous study found that both the CRMP and GPS location data had substantial problems resulting in the location data only being accurate about 50 percent of the time using either method. This research aimed to evaluate Kentucky's crash data since the addition of a mapping system integrated into KYOPS.

An evaluation of a random sample of crashes from 2009 was performed to assess the current accuracy of the crash data's location information. A second evaluation was performed on crash data since the MapIt system was introduced (in late 2007) through 2009. This analysis calculated the distance between the locations plotted by CRMP and by the GPS data and summarized the data by several factors.

The evaluation of the random sample of crash data revealed that the location information is accurate in a substantially higher number of records compared to a study five years ago (92 percent accuracy compared to about 50 percent). Furthermore, the evaluation of the 2007 through 2009 crash data yielded results that aided in making recommendations that will address the most egregious and frequent errors related to location data. Finally a literature review was conducted to better describe where a crash should be located when a police report is completed.

**Keywords:** Crash locations, GPS, GIS, collision report, crash mapping

*This paper has been modified from its original format to satisfy length limitations. Several tables and appendices have been removed.*

## INTRODUCTION

An important component of a traffic crash report is the proper documentation of the crash location. This can be difficult, and somewhat subjective, if the crash occurs over a large distance or if there were multiple events. One of the most important reasons for accurate crash locations is so that a traffic engineer can properly identify where a crash occurred in an effort to avoid future crashes. Prevention can be in the form of educational campaigns, enforcement efforts or highway safety improvements. Moreover, properly locating crashes to the road it occurred (and the correct sections of road) will allow the crash data to be linked to roadway traffic volumes so that rates can be calculated. Rates allow a way for researchers to find roadway sections that may not have a large number of crashes, but rather a higher crash or fatality rate when based on the number of vehicles that travel on that roadway section.

The objective of this research was to evaluate the current accuracy of the crash data since the implementation of the MapIt system. An evaluation of a random sample of crashes from 2009 was performed to assess the current accuracy of the crash data's location information. A second evaluation was performed on crash data since the MapIt system was introduced (in late 2007) through 2009. This analysis calculated the distance between the locations plotted by CRMP and by the GPS data and summarized the data by several factors.

## BACKGROUND

The police agencies in Kentucky have two methods of reporting traffic crashes: using the eCrash component of Kentucky Open Portal System's (KYOPS), an electronic reporting system, or manually using paper reports. The location information for either method comes in two forms: latitude and longitude coordinates (referred to as global positioning system, GPS, regardless of how the coordinates are obtained) and county, route and milepoint (CRMP). These two formats have distinct advantages over one another with the ideal solution being the preservation of both formats.

The paper report contains fields for county, route and milepoint and GPS coordinates. The CRMP data is typically entered using the ‘distance from’ field where the distance and the direction from a reference point are entered (a milepoint from a known location). Additionally, the county and route are also entered. CRMP data is only requested for state-maintained roadway crashes. This has been the primary source of location data prior to 2000. In 2000, GPS coordinates were added to the form. On June 1$^{st}$, 2002, GPS coordinates were required on all reports. The coordinates are reported in degrees, minutes and seconds. An example portion of a paper report is shown below.

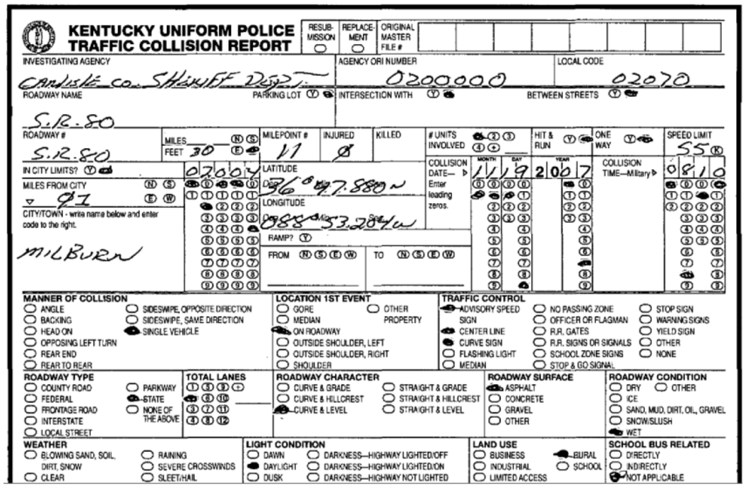
KENTUCKY UNIFORM POLICE TRAFFIC COLLISION REPORT
RESUB-MISSION | REPLACE-MENT | ORIGINAL MASTER FILE #
INVESTIGATING AGENCY: Carlisle Co. Sheriff Dept. | AGENCY ORI NUMBER: 0200000 | LOCAL CODE: 02070
ROADWAY NAME: S.R. 80 | PARKING LOT Y N | INTERSECTION WITH Y N | BETWEEN STREETS Y N
ROADWAY #: S.R. 80 | MILES / FEET 30 N S E W | MILEPOINT #: 11 | INJURED: Ø | KILLED | # UNITS INVOLVED 1 2 3 4 + | HIT & RUN Y N | ONE WAY Y N | SPEED LIMIT 55
IN CITY LIMITS? Y N | 07004 | LATITUDE Deg 36° Min 47.880 N | COLLISION DATE— Enter leading zeros. MONTH 11 DAY 19 YEAR 2007 | COLLISION TIME—Military 0810
MILES FROM CITY N S E W: Ø1 | LONGITUDE Deg 088° Min 53.284 Sec W
CITY/TOWN - write name below and enter code to the right.: MILBURN | RAMP? Y | FROM N S E W | TO N S E W
MANNER OF COLLISION: ANGLE; BACKING; HEAD ON; OPPOSING LEFT TURN; REAR END; REAR TO REAR; SIDESWIPE, OPPOSITE DIRECTION; SIDESWIPE, SAME DIRECTION; SINGLE VEHICLE
LOCATION 1ST EVENT: GORE; MEDIAN; ON ROADWAY; OUTSIDE SHOULDER, LEFT; OUTSIDE SHOULDER, RIGHT; SHOULDER; OTHER PROPERTY
TRAFFIC CONTROL: ADVISORY SPEED SIGN; CENTER LINE; CURVE SIGN; FLASHING LIGHT; MEDIAN; NO PASSING ZONE; OFFICER OR FLAGMAN; R.R. GATES; R.R. SIGNS OR SIGNALS; SCHOOL ZONE SIGNS; STOP & GO SIGNAL; STOP SIGN; WARNING SIGNS; YIELD SIGN; OTHER; NONE
ROADWAY TYPE: COUNTY ROAD; FEDERAL; FRONTAGE ROAD; INTERSTATE; LOCAL STREET; PARKWAY; STATE; NONE OF THE ABOVE
TOTAL LANES: 1 2 3 4 5 6 7 8 9 10 11 12 +
ROADWAY CHARACTER: CURVE & GRADE; CURVE & HILLCREST; CURVE & LEVEL; STRAIGHT & GRADE; STRAIGHT & HILLCREST; STRAIGHT & LEVEL
ROADWAY SURFACE: ASPHALT; CONCRETE; GRAVEL; OTHER
ROADWAY CONDITION: DRY; OTHER; ICE; SAND, MUD, DIRT, OIL, GRAVEL; SNOW/SLUSH; WET
WEATHER: BLOWING SAND, SOIL, DIRT, SNOW; CLEAR; RAINING; SEVERE CROSSWINDS; SLEET/HAIL
LIGHT CONDITION: DAWN; DAYLIGHT; DUSK; DARKNESS—HIGHWAY LIGHTED/OFF; DARKNESS—HIGHWAY LIGHTED/ON; DARKNESS—HIGHWAY NOT LIGHTED
LAND USE: BUSINESS; INDUSTRIAL; LIMITED ACCESS; RURAL; SCHOOL
SCHOOL BUS RELATED: DIRECTLY; INDIRECTLY; NOT APPLICABLE

Figure 1 The top portion of a paper report showing location data

The paper report shown indicates that the crash occurred 30 feet west of milepoint 11 (likely based on the presence of a mile marker). This is calculated in the database as 10.994 (11 – 30/5,280). The county is identified as Carlisle County (code 020) and the route is KY-80. The CRMP data is therefore 020-KY-0080 at 11.994. The GPS data is shown in this case in degrees and decimal minutes, despite the fact that the form expects the data in degree-minutes-seconds. This could be entered erroneously into the database because of this discrepancy.

The eCrash system dramatically increases the effectiveness of the crash reporting system. Initially, the same system as in the paper report was used to locate crash data. That is, CRMP using the 'distance from' form and degree-minute-second GPS data. However, a major update to KYOPS included the implementation of a mapping system called MapIt which was added on October 1, 2007. This system allows the officer to click on a point on a map and the GPS coordinates and CRMP data would be automatically filled out. A screenshot of the MapIt system is shown below.

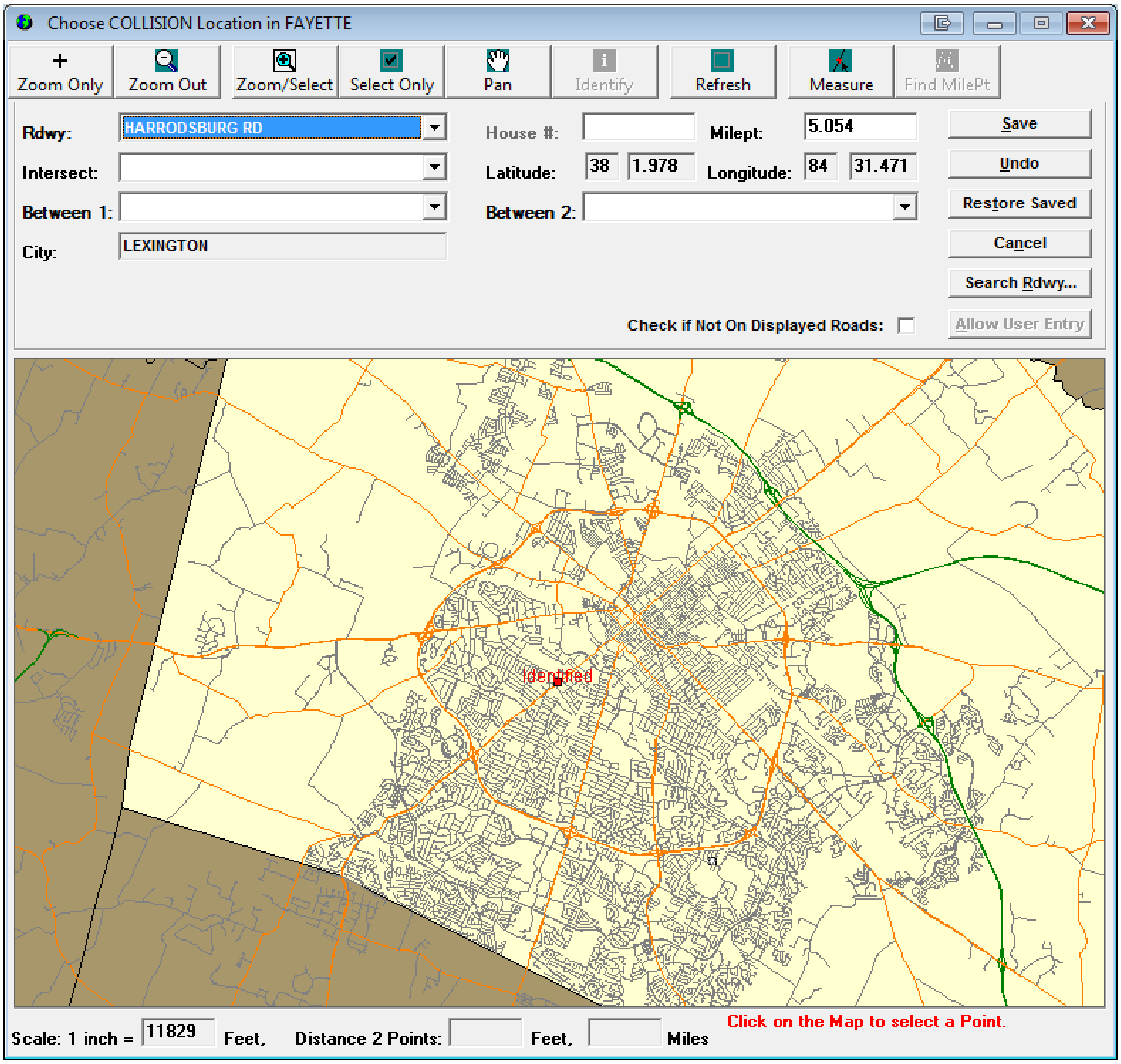

Figure 2 The MapIt module used to generate location data in eCrash

In this example the identified location is shown by milepoint and by degree and decimal-minutes GPS. Additionally, the proper route identifier (034-US-0068) is provided in the location form.

## LITERATURE SEARCH

An important aspect of improving the safety of roadways is the accurate identification of crash locations. Historically, police officers reported the location of a crash by estimating its distance from a known reference point. This estimation is subject to error and has the potential to be highly variable from one officer to another reducing the uniformity in crash location data. Global Positioning Systems greatly reduce the potential for error when used to determine the location of a crash. A recent NCHRP Synthesis of Highway Practice report indicates that GPS and GIS crash location are being used by approximately 48% to 54% of responding states, with varying levels of coverage; however, the primary method of crash location is still through traditional route and distance methods. A study by Sarasua of Clemson University evaluated South Carolina crash location data and found that, once reporting officers were equipped with

GPS units to determine coordinate location data, 80 percent of that state's crash records were located within reasonable levels of accuracy. Graettinger et al. evaluated several portable GPS units and found that even inexpensive units are capable of locating a crash within an eight meter radius.

Spatial analysis technology is quickly changing and improving. In 1998, a study conducted in Iowa, Washington, and Wisconsin reported an average satellite locking time of 8 minutes. Two years later, a study conducted in Alabama reported a 3 minute satellite locking time, and in a 2004 study done in South Carolina, the average locking time was 1 minute, 49 seconds. With reporting officers under time constraints, the increased speed of GPS satellite locking will likely lead to a more streamlined process of completing crash reports and result in fewer errors of omission. With higher quality spatial data, GIS-based analysis tools, such as Iowa's GIS-ALAS, are able to more accurately determine problem areas and the data can be more efficiently shared and used by other agencies to improve safety.

## PROCEDURE

### Random Sample of Crashes

A representative sample of crashes was needed to be reviewed to assess the accuracy of the location data. The required sample size was calculated for a population of 126,237 reported crashes in 2009 (not including private property and parking lot crashes). The following equation was used to find the sample size:

$$n = \frac{N}{1+N(e)^2} \quad (1)$$

where:
*n : sample size*
*N : population*
*e : level of precision (+/-7%)*

This equation returns that a sample size of 204 was needed to produce an estimate at the 95% confidence level with a confidence interval of +/- 7 percent. The number of samples needed would need to be nearly doubled to reach a confidence interval of +/-5 percent. Unfortunately, the process is too time-consuming to reach that level of confidence for this study.

A sample of 222 random crashes (an extract with a slightly larger sample size than needed was conveniently available) from the 2009 crash database was obtained. Only crashes with a GPS value were included in the initial dataset. Crashes occurring on local and state-maintained roadways were used. The police report for each crash was reviewed. A data field was used to verify that all reports had a GPS location that was generated by the MapIt system. Therefore, the GPS of each report was used as the “Report Location.”

An attempt to determine if the report location was accurate was made for each report. A “yes” was assigned to the crash if its report location was within approximately 500 feet from the presumed location of the crash. Conversely, a “no” was given to crashes where the presumed location was outside of a 500 foot radius of the reported location. The presumed location was based on any information in the report that helped identify its location. Report narratives, addresses or intersecting or between roads were typically used to identify the presumed location. Each crash was assigned a location type:

- Intersection – an intersecting road was given
- Between Streets – two streets were listed as reference points
- None – neither of the above

A value of "unknown" was given to reports where a presumed location could not be pinpointed. Google® Maps was the primary tool used for plotting addresses and GPS coordinates. In some cases, ArcMap® was used to plot milepoints for comparisons.

It should be noted that a distance of 500 feet is reasonable for network screening or high crash locations but it can be too small for individual crash analysis. This distance was used to account for errors in data projection or address approximations.

Each crash was given a range that defined the distance between the presumed and reported locations. For example, if the crash was determined to be approximately 300 feet from the actual location then the range would be: Min=300, Max=300. If the actual location could not be precisely pinpointed then a range of realistic values was used. This could be based on the length of the road or presence of locatable landmark (intersection, bridge, road character, etc.). For instance, it is unlikely to assume that an interstate crash is 10 miles from the reported location if there are exits every mile. The average offset is the midpoint of Max and Min values. This value can be used to quantify the accuracy of the location. When possible, a description for why the location was incorrect was given.

**Crash Database Analysis**

Crash data were obtained from October 1, 2007 to December 31, 2009. This time frame was used as it was the start of KYOPS's mapping system and included data up until the last full yearly extract. This resulted in a database of 334,354 crash records.

All of the records with a valid county, route and milepoint were plotted against the state's Allrds_m shapefile using ESRI's ArcMap®. The crash fields RSEUniqueGPS (county-route) and CurrentDerivedMiepointNumber (milepoint) were used to plot along the LRS_ID field of Allrds_M. The county-route field identifies the county number and the route prefix, number, suffix as well as the route type (mainline, ramp, non-cardinal, etc.). Some of the data reported had a mis-formated RSE field. For instance, some had leading zeros in the route number and some had a dash after the county. For plotting purposes, the format required was: 001 KY-55 or 001 KY-55-20. The latter indicates that the route is an auxiliary system to KY-55 (such as a channelized right turn lane). The RSE field was modified where necessary to plot properly.

A GPS coordinate was added to all successfully plotted data. These coordinates were compared to the crash GPS data by calculating the distance between them using the following formula:

$$D = R\cos^{-1}\left(\cos(long_1 - long_2)\cos(lat_1)\cos(lat_2) + \sin(lat_1)\sin(lat_2)\right) \quad (2)$$

where:
D : distance in miles
R : radius of Earth (3,963.19 miles)
$lat_1$ : latitude from crash report
$long_1$ : longitude from crash report
$lat_2$ : latitude of plotted location
$long_2$ : longitude of plotted location

The difference in the longitude from the reported GPS and the longitude created from the plotted CRMP data was used to approximate the horizontal component of the distance between the two locations (equation 3). For such relatively small distance (as compared to the radius of the Earth), the spherical formula used above was unnecessary. The vertical distance was similarly calculated from the latitude values (equation 4). Crashes that had a horizontal component that was more than 90 percent or less than 10 percent of the straight-line distance (equation 5) between the two coordinates were flagged as a 90/10 error (90 percent of the distance is in either the horizontal or vertical direction) (equation 6). The following formulas were used to calculate these values.

$$D_{horz} = Long_{CRMP} - Long_{GPS} \quad (3)$$

$$D_{vert} = Lat_{CRMP} - Lat_{GPS} \quad (4)$$

$$D = \sqrt{D_{horz}^2 + D_{vert}^2} \quad (5)$$

$$E_{90/10} = \frac{D_{horz}^2}{D^2} \times 100 \quad (6)$$

where:
$D_{horz}$ : horizontal distance between CRMP and GPS locations in miles
$D_{vert}$ : vertical distance between CRMP and GPS locations in miles
D : straight-line distance in miles
$E_{90/10}$ : percent of distance contributed by the horizontal component

As an example, if the straight-line distance was five miles, the horizontal distance was three miles and the vertical distance was four miles, then the horizontal percentage would be 36 percent (9/25*100).

Crashes with the horizontal component higher than 90 percent are possibly due to GPS recording errors in longitude. Likewise, crashes with a horizontal component lower than 10 percent are possibly due to GPS recording errors in latitude. A misleading instance of this error can exist in cases where roads are oriented either east-west or north-south. In these cases, either the GPS reading could have been measured down the road from the actual crash location or the milepoint may have been reported at an incorrect distance from the actual crash location. In either case, the results could be perceived as a latitudinal/longitudinal type of error. Errors in the CRMP data that could contribute to this error occur when the wrong direction is given from the reference milepoint.

## RESULTS

### Random Sample of Crashes

Of the 222 crashes, 71 percent were shown to be in the accurate position, 6 percent were not, and the remaining crashes were unable to be confidently assessed. It should not be assumed that the unknown locations are incorrect. These locations typically had a lack of reference that made them harder to pinpoint. The accuracy percentage is 92 percent when excluding those records with unknown accuracy (152 of the 167 with determinable locations). Table 1 shows the percentage by accuracy and type of crash location.

Table 1 Percent of crashes by accuracy level and by location type

| Was location Accurate (Percent)? | Type of Location | | | |
|---|---|---|---|---|
| | Between streets | Intersection | None | Grand Total |
| Yes | 100.0 | 100.0 | 54.6 | 71.2 |
| No | 0.0 | 0.0 | 9.2 | 5.9 |
| Unknown | 0.0 | 0.0 | 36.2 | 23.0 |
| **Grand Total** | **100.0** | **100.0** | **100.0** | **100.0** |

In Table 1 percentages are shown by columns. It is clear that unknown and inaccurate data is only a factor for crashes that do not have intersecting or between streets. This is largely due to the fact that there are much fewer reference points in the MapIt system when crashes are reported in more rural areas.

As discussed earlier, an average offset was calculated for all crash records in this sample. This value represents an average distance between the presumed location and the reported location. When the location is known, this value represents the distance along the road that the presumed location is from the reported location. For unknown locations, this value represents the distance between the reported location and the midpoint between the minimum and maximum presumed location. This value is subjective, but is controlled by the presence of logical reference points. Intersection crashes have the shortest offset, which is explainable as intersection crashes offer the best reference system. The second shortest offset was for accurate crashes with location type 'none'. Between street crashes had the longest offset for accurate crashes which could be attributed to distance between the streets. These crashes have reference points but they are not nearly as definitive as intersection crash.

The maximum offset of the crash records was over eight miles. The maximum offset for records with an unknown location was about 3.5 miles. This value, although calculated from a small sample of only 51 crashes, represents a logical maximum for crashes based on available reference points.

The thirteen crashes that were coded as inaccurate had two explanations:

- GPS is not consistent with address/reference (12 crashes)
- No reference point (1 crash)

The former explanation was indicated when an address or specific reference point (presumed location) was given in the narrative or otherwise on the report and that location is not within 500 feet of the GPS location (report location). In some cases this could be due to errors in the addressing system used by Google® Maps. However, this is unlikely as those types of errors would be smaller than 500 feet. The latter explanation indicated that there was not a nearby reference point for the officer to use. This implies that the officer needed to "guess" where to click on the mapping system.

**Crash Database Analysis**

The crash data were summarized by several factors in order to evaluate the accuracy of the data and to recommend improvements. Of the 334,354 crashes, only 186 crashes had no GPS data. Figure 3 shows the crash data as plotted by the reported GPS coordinates.

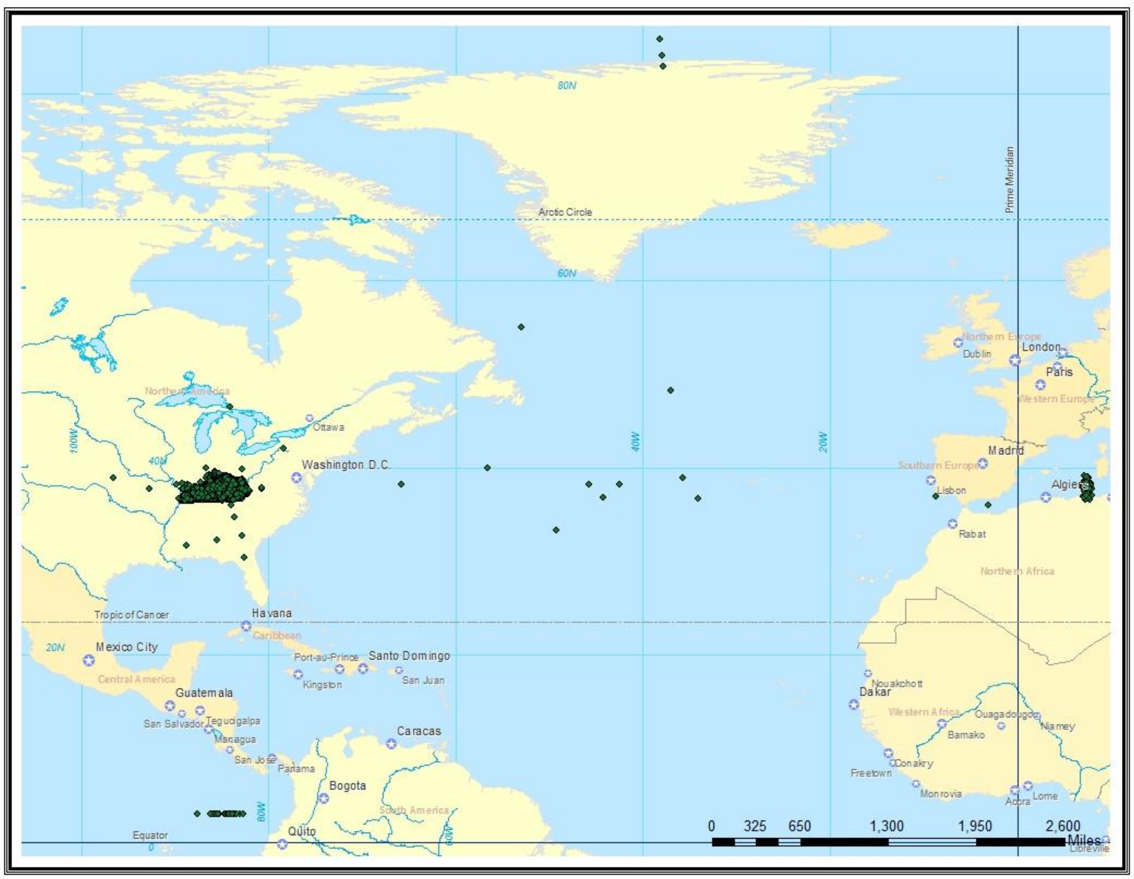


Figure 3 All crashes from October 1st 2007 to December 31st 2009plotted by reported GPS

It should be noted that Figure 3 does not indicate the percentage of crashes that were plotted out of Kentucky. In fact, as discussed below, this is a very small percentage. However, Figure 3 was included to show the scale of some of the GPS errors.

Of the 334,354 crashes, 261,973 (78 percent) had values for RSE (county and route) and milepoint. Of those, 258,376 (77 percent of all crashes) were able to be plotted (along local and state-maintained roads). For a majority of the analysis below, crashes that had a GPS and were able to be plotted were referred to as usable. These represent the crashes that could have a distance calculated between the GPS and CRMP locations. Seventy-seven percent of all crashes (258,374) were flagged as usable.

The "CityCountyCode" field from the crash database was used to obtain the county in which each crash was reported. The reported county was compared to county in which the GPS was plotted. Figure 4 shows all the crashes near Kentucky as red circles or gray squares. Red circles represent crashes that were not plotted in their reported county. Obviously, any crashes that were plotted outside of the view frame would also be red. The red circles are larger and have been promoted to the top of the map to indicate where the mis-plotted crashes are being plotted. Ninty-nine percent of all crashes were plotted in the reported county for the study period.

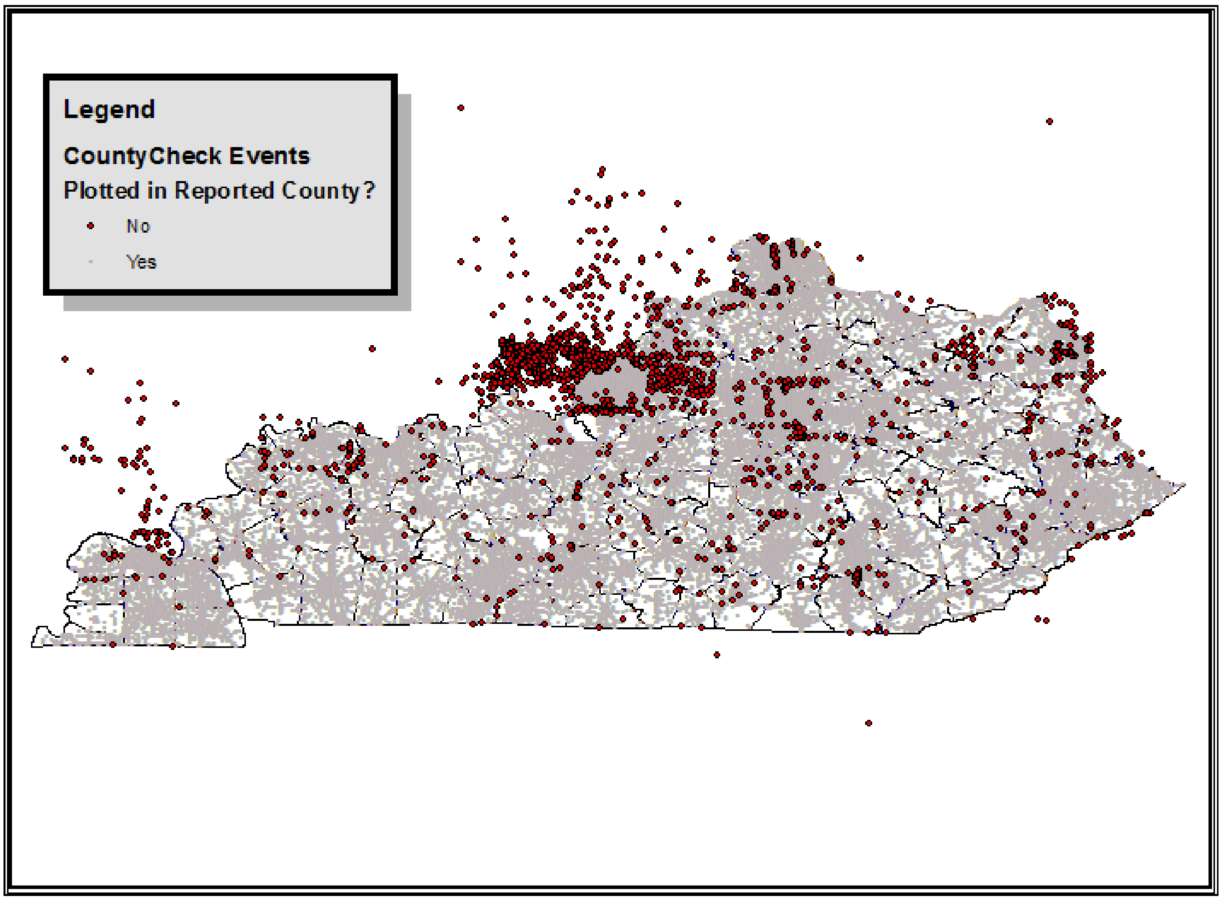


Figure 4 Plotted crashes shown by county check

The percentage of crashes that were plotted outside of their county was summarized by reporting county. Twenty-three counties had more than one percent of the crashes plotted outside of the reported county. Hancock and Whitley counties had the highest percentages (12.8 and 7.1, respectively). Only seven counties had a percentage of zero indicating that all of the crashes for that county were plotted inside of that county. Those counties are: Bracken, Graves, Hickman, Metcalfe, Owsley, Spencer and Wolfe.

The same analysis was repeated for the crash city code and the agency code (indicates the police agency responsible for the crash report). It should be noted that several of these cities and agencies had very small sample sizes resulting in very misleading percentages.

Every year there are fewer paper reports submitted. This trend is in favor of the newer and more user-friendly eCrash system through KYOPS. Only 2.8 percent (4,067 of 148,010) of all submitted reports were paper in 2009 (up from 11.3 percent in 2008). This is the lowest percentage of paper reports year-to-date. Paper reports were fairly evenly distributed by agency type: Kentucky State Police (KSP), Sheriff or Local Police and by roadway type (local or state routes). Nine counties had a paper report submission rate of twenty percent of more.

A similar summary was produced for cities and agencies. Lists of these cities and agencies have been omitted in order to satisfy word count limitations.

The crash database contains a field called UserEnteredRdwysInd that served a dual-purpose in this analysis. This field is populated with a "Y" if the user chooses to modify the location data provided by the MapIt system. By default the field is "N" indicating that the user did not modify the MapIt location data. However, if the MapIt system was not used, the field was blank. Table 2 shows the percentages for electronic and paper submission types by year as well as percentages of each mapping type.

Table 2 Percent of crashes by year and MapIt usage

| Percent | Year | | | |
|---|---|---|---|---|
| **Submission Type** | **2007** | **2008** | **2009** | **Grand Total** |
| **Electronic** | **82.9** | **88.7** | **97.3** | **91.8** |
| No MapIt | 44.3 | 14.0 | 0.0 | 11.5 |
| Used MapIt | 35.6 | 70.7 | 90.3 | 75.1 |
| Used MapIt But Changed | 3.0 | 4.0 | 6.9 | 5.2 |
| **Paper** | **17.1** | **11.3** | **2.7** | **8.2** |
| No MapIt | 17.1 | 11.3 | 2.7 | 8.2 |
| **Grand Total** | **100.0** | **100.0** | **100.0** | **100.0** |

Less than three percent of all crashes did not use MapIt in 2009 compared to over 25 percent in 2008.

The remainder of this analysis involved looking at the difference between the plotted GPS location versus the plotted CRMP location. Both the GPS and CRMP locations had to be useable. All records with usable location data were flagged and referred to as "usable." Crashes with no GPS were not usable. Very few crashes had no GPS. There were 186 crashes in the timeframe with no GPS. Crashes with no RSE, MP or crashes that had an invalid RSE or MP were not usable. That is to say, any crash that did not plot on the state's allrods_m shapefile was not usable. Table 3 shows the percentage of each plot type by year.

Table 3 Percent of crashes by year and plot type

| All Records (Percent) | Year | | | |
|---|---|---|---|---|
| Plot Type | 2007 | 2008 | 2009 | Grand Total |
| OKAY | 61.3 | 77.2 | 81.8 | 77.3 |
| No RSE | 6.1 | 2.4 | 1.7 | 2.6 |
| Did Not Plot | 1.2 | 1.3 | 0.8 | 1.1 |
| No RSE and MP | 25.7 | 8.0 | 4.4 | 8.5 |
| No MP | 5.8 | 11.2 | 11.3 | 10.6 |
| **Grand Total** | **100.0** | **100.0** | **100.0** | **100.0** |

A large majority (over 80 percent) of all paper reports didn't have CRMP data; however, a large portion of these could be local roads that do not require this data for paper submissions. Electronic submissions, however, generate a CRMP for local roads when MapIt is used. Table 4 shows the same data as above for only electronic submissions.

Table 4 Percent of electronic crashes by year and plot type

| | Percent | | | |
|---|---|---|---|---|
| Electronic | Year | | | |
| Plot Type | 2007 | 2008 | 2009 | Grand Total |
| OKAY | 70.0 | 84.9 | 83.6 | 82.6 |
| No RSE | 7.0 | 2.5 | 1.7 | 2.6 |
| Did Not Plot | 0.7 | 1.1 | 0.8 | 0.9 |
| No RSE and MP | 22.2 | 3.1 | 2.8 | 5.1 |
| No MP | 0.1 | 8.4 | 11.1 | 8.8 |
| **Grand Total** | **100.0** | **100.0** | **100.0** | **100.0** |

Some of these percentages may include crashes that had no GPS. Only crashes with GPS and a plot type of 'okay' were considered usable. Table 5 shows the number and percentage of crashes by agency and whether it is usable.

Table 5 Count and percent of usable crashes by agency

| | Usable | | | | Total | |
|---|---|---|---|---|---|---|
| | Yes | | No | | | |
| Reporting Agency | Count | Percentage | Count | Percentage | Count | Percentage |
| KSP | 35,208 | 92.6 | 2,811 | 7.4 | 38,019 | 100.0 |
| Local | 169,352 | 72.2 | 65,322 | 27.8 | 234,674 | 100.0 |
| Sheriff | 53,814 | 87.3 | 7,847 | 12.7 | 61,661 | 100.0 |
| **Grand Total** | **258,374** | **77.3** | **75,980** | **22.7** | **334,354** | **100.0** |

Kentucky State Police had the highest percentage of usable crashes, followed closely by County Sheriff. Local police had the lowest percentage of usable crashes (72.7 percent); however, they handled the majority of crashes (70 percent).

When “distance” is referred to in the remainder of this report, the distance is between CRMP and GPS locations. The “distance” was calculated for all usable crashes. The average “distance” for all usable crashes ranged from 3.1 miles in the last quarter of 2007 (24,874 crashes) to 0.3 miles in 2009 (121,066 crashes). It should be noted that calculating an average distance is largely affected by outliers. Furthermore, the “distance” should not be viewed as a measure of crash location accuracy but rather a discrepancy of the locations which might be an indicator of an inaccurately located crash.

The crash database was sorted by “distance” and several of the most egregious errors were examined. There were two types of electronic reporting errors that yielded large “distance” between the CRMP and GPS locations.

One error resulted in the longitude being off by 96 degrees. Subtracting this value from these crashes re-plotted the crash very close to the CRMP location. For example, the crash record with a master file number of 70651040 had a reported longitude of 7.759983063 but a longitude based on the CRMP of -88.2399139 (7.76 - 96 = -88.24). Interestingly, the erroneous GPS value is reported as 7.759983063 on the printable crash report and in the online database, however, the raw database (used in this analysis) shows a value of “07.759983-01” whereas a typical value would be “-86.43655862.” There was a similar error for latitude. It should be noted that this only occurred for electronic reports. It was also only observed in 252 crashes, most of which were in 2008. The error was not observed since January of 2009. It is assumed to have been fixed.

The second error was harder to quantify its frequency as the “distance” was not nearly as egregious as the former error, however it was more prevalent. It was noticed in about 700 crashes that the GPS coordinates were modified by the ‘distance from’ field. Prior to the existence of the MapIt system, the user could enter a reference milepoint (typically from a milepoint log book or reference marker) and modify this location with a distance and direction. For instance, if an officer knew he or she was a half of a mile south (based on the cardinal direction of the roadway) from a logged intersection at milepoint 1.750 then they could enter 0.5 miles South into the ‘distance from’ form and the new milepoint would be 1.250. It seems that officers can still use the ‘distance from’ form even if they use the MapIt system. For instance, it was noticed that the GPS was spatially offset from the CRMP location for several crashes that had values in the ‘distance from’ fields (even if the value in this field was zero). It is assumed that this is by design so that the user can offset their location by a distance if they know the GPS of a reference location. Unfortunately, the user can only offset due north, south, east or west, which is a very unlikely alignment. It is likely that this affected more than 700 crashes, however the error is presumed to be much smaller. This error is currently repeatable on a KYOPS installation with the latest version as of this writing.

There were also several errors noticed related to the paper reporting system. Several of these errors were due to the officer entering the data wrong or the data being keyed into the system incorrectly. There were 731 crashes with a GPS minute value greater than or equal to 60 (which is not a valid value). There were also over 400 crashes with noticeably bad GPS. Furthermore, there were several crashes that had no minutes or seconds or suspicious values such as 30 minutes. Although these represent a small number of crashes, they are a much higher proportion when only paper reports (about 27,000) are considered.

As shown earlier, the average "distance" of all usable crashes was 3.8 miles. The effect of some preventable errors can be shown despite the fact that this average "distance" is largely affected by the outliers. For instance, if all crashes that were plotted outside of their reported county are removed, this "distance" reduces from 3.8 to 0.2 miles. This implies that if a county check was performed before accepting a crash that the most egregious errors could be eliminated. It should be noted that most of the worst errors occurred prior to 2009. That is to say, the average "distance" in 2009 would reduce from 0.3 miles to 0.1 miles by requiring this check.

Another unexpected error was seen in crashes indicating that the officer used the MapIt system and did not edit its location. It would be expected that the "distance" between the CRMP and the GPS locations would be ostensibly zero. A small distance could be attributed to rounding and map projection errors. Unexpectedly, however, there were 5,267 (4.4%) usable crashes in 2009 that used an unedited MapIt location, yet had a "distance" of 500 feet or more. There were six crashes with a "distance" over 4,000 miles; however this can be explained by the latitude/longitude errors discussed above. There were 12 crashes with a "distance" of over 20 miles. The average "distance" of this dataset was 6.6 miles (1.3 miles excluding the six with latitude/longitude errors). The $50^{th}$ and $85^{th}$ percentiles were 0.2 miles and 1.8 miles respectively. Several of these crashes were reviewed to determine the reason for these errors. For 127 of the crashes the error seemed to be related to the 'distance from' error discussed above. Some of the largest "distances" seemed to be related to errors in the MapIt milepoint database.

A summary was performed on all of the crashes that had an unmodified MapIt location in 2009 by county-route combinations. The number of each county-route (for example KY-36 in Bath County) was counted in the unmodified MapIt location database and this number was compared to that county-route's occurrence in all 2009 crashes. A percentage was calculated based on these two counts. The unmodified MapIt location database represents only 4.4 percent of all 2009 crashes. Therefore, any county-route combination with a percentage higher than 4.4 percent is overrepresented. Several county-routes had very high percentages implying that there may be something wrong with these routes in the MapIt database.

The same analysis was performed on the reporting agency (KSP, local police or sheriff). Again, these percentages can be compared to 4.4 percent. Agencies with higher percentages may indicate that training issues may prevent some of these errors.

Although most of the location errors have been addressed and mitigated in KYOPS, the following analysis only looks at paper reports or electronic reports with modified MapIt locations (only crashes flagged as usable). The "distance" between the GPS and CRMP should be zero for all unmodified MapIt locations (the previous analysis investigated those that were not); therefore these were removed from the analysis so that the averages were not driven down by the frequency of such crashes. The average “distance” ranged from 2.2 miles in the last quarter of 2007 (11,565 crashes) to 1.2 miles in 2009 (4,709 crashes). The data from 2009 has the lowest proportion of modified locations and paper crashes (3.9 percent of all 2009 usable crashes) and the lowest average “distance.”

As discussed previously, a check for crashes being plotted outside of their reported county would reduce the average “distance”. The average “distance” of 1.2 miles is reduced to 0.8 miles when only looking at crashes plotted within their county. The remainder of the error is likely related to milepoint errors as seen in a previous study where errors were caused by poor milepoint references, inaccurate estimations of distance or incorrect ‘direction from’ values (Green & Agent, 2004). These types of errors are typically seen on paper reports but the proper county would still be reported.

The average “distance” was 0.1 miles for electronic submissions with a modified MapIt location (3,929 crashes) and 6.4 miles for paper reports (780 crashes). Paper reports produce the largest “distance” yet make up a very small percentage of submission type.

As seen in previous research, and as one might expect, the average “distance” is improved for crashes at intersections or with identified between streets (Green & Agent, 2004). Such indicators offer a better reference system than, for example, a rural two-lane road. More reference points allow the reporting officer to be able to better locate a crash. Moreover, between street and intersection locations are more prevalent in urban areas where the crash location is easier to pinpoint. Almost all of the ramp crashes in 2009 were electronic reports (only 10 of 2,502 ramp crashes were paper). The MapIt system introduced a better ramp classification system by adding indicators to the RSE field. Even though these MapIt locations were modified, ramps are typically small in length; therefore there is little room for error. This explains the lower average “distance” for ramp crashes (0.1 miles compared to 2.4 miles for non-ramp crashes). Table 6 shows the average “distance” for all crashes by functional classification.

Table 6 Average "distance" and sample by function class for all usable paper and edited electronic crashes in 2009

| Functional Classification | Average “Distance” (Miles) | Sample Size |
|---|---|---|
| Rural Interstate | 1.1 | 92 |
| Rural Principal Arterial (non-interstate) | 3.4 | 198 |
| Rural Minor Arterial | 1.7 | 149 |
| Rural Major Collector | 3.8 | 325 |
| Rural Minor Collector | 3.2 | 127 |
| Rural Local | 60.8 | 40 |
| Urban Interstate | 0.1 | 1,005 |
| Urban Principal Arterial (Freeway & Expressways) | 0.0 | 211 |
| Urban Principal Arterial (other) | 0.2 | 452 |
| Urban Minor Arterial | 0.2 | 224 |
| Urban Collector | 0.3 | 67 |
| Urban Local | 0.1 | 17 |
| Unknown | 0.1 | 1,802 |
| **Grand Total** | **1.2** | **4,709** |

As expected, rural locations tend to have a higher average "distance" (5.5 miles) compared to urban (0.1 miles). The highest "distance" also has one of the lowest sample sizes. Rural locations have fewer reference points making it more difficult for an officer to pinpoint their location on a map.

The location data is better for fatal crashes as compared to all crashes in the 2009 data. Compared to an average "distance" of 0.34 miles for all crashes there was an average "distance" of 0.31 miles for injury crashes and 0.07 miles for fatal crashes. More care is usually given to the reporting of serious crashes especially those resulting in a fatality.

The 50$^{th}$ and 85$^{th}$ percentile "distances" were also summarized in an effort to reduce the impact of outliers on the results. Table 7 shows the 50$^{th}$ and 85$^{th}$ percentiles for several subsets of the crash data. All years of crash data were used (October 1$^{st}$ 2007 to December 31$^{st}$ 2009).

Table 7 50th and 85th percentiles for several subsets of the crash data (2007-2009)

| | "Distance" (Miles) | | "Distance" (Feet) | |
|---|---|---|---|---|
| **Subset** | **50th** | **85th** | **50th** | **85th** |
| All | 0.0031 | 5,788 | 16 | 30,561,059 |
| All Usable | 0.0023 | 0.0369 | 12 | 195 |
| All Usable - Intersections | 0.0028 | 0.0440 | 15 | 232 |
| All Usable - Between | 0.0025 | 0.0443 | 13 | 234 |
| All Usable - eCrash | 0.0022 | 0.0323 | 12 | 170 |
| All Usable - Paper | 0.2774 | 3.2183 | 1,465 | 16,993 |
| All Usable - Intersections - Paper | 0.2668 | 4.4773 | 1,409 | 23,640 |
| All Usable - Intersections - eCrash | 0.0012 | 0.0136 | 6 | 72 |
| All Usable - Sheriff | 0.0049 | 0.0886 | 26 | 468 |
| All Usable - Local | 0.0016 | 0.0209 | 9 | 110 |
| All Usable - KSP | 0.0050 | 0.0754 | 26 | 398 |
| All Usable - Intersections - Sheriff | 0.0020 | 0.0974 | 11 | 514 |
| All Usable - Intersections - Local | 0.0012 | 0.0144 | 7 | 76 |
| All Usable - Intersections - KSP | 0.0011 | 0.0362 | 6 | 191 |
| All Usable - In County | 0.0023 | 0.0356 | 12 | 188 |

The first subset shows all of the data (even data with explainable errors or no CRMP or GPS data). Unfortunately, the prevalence of very small "distance" values confounds this type of analysis. In addition, large "distances" (typically seen on paper reports) also confound the results. Some of the trends seen in the previous report were not seen because of these two factors.

The crashes with 90/10 errors where 90 percent of the "distance" between CRMP and GPS is in either the horizontal or vertical direction were summarized. Similarly, the crashes with 95/5 errors were summarized. Again, data from 2009 were used in order to minimize errors. There were no major differences between all crashes and intersections crashes.

**Locating a Traffic Crash**

As discussed previously, it can be difficult and subjective to properly locate crashes. According to the American National Standards, the definition of a crash location is the location on a roadway to be documented as where the first harmful event of the crash occurred (American National Standard, 2007). A harmful event is characterized as the first injury or damage-producing event. There are several types of crashes that require specific attention. Such crash types are discussed below. Recommendations to handling these crash types are also given in the subsequent section.

Intersection Crashes

Intersections crashes can be easy to define when the crash is clearly intersection related (an angle crash as a result of a red light running). However, intersection crashes would also include crashes occurring as a result of a sight distance issue related to the environment around an intersection or a rear-end collision as a result of stopped traffic at an intersection.

The location alone does not fully provide all of the crash location data, particularly at intersections. According to results from research at the Northwestern University's Center for Public Safety, the road in which the most harmful event occurred should be listed as the primary road (Baker, 2001). When the crash involves two or more vehicles travelling on different roads then the road with the higher importance should be used. The research study defines the following categories of roads in order of importance.

1. Interstate Highway
2. Other U.S. route numbered highway
3. Other state route numbered highway
4. County road
5. City street
6. All others

If the two intersecting roads are of equal importance, the collision should then be located on the traffic way "with the lowest route, number, or street name nearest the beginning of the alphabet" (Baker, 2001).

Ramp or Interchange Crashes

Ramp crashes have previously been difficult to properly locate. With the addition of the MapIt system in KYOPS, ramp identifiers (provided by the transportation cabinet) are available for all roads.

Crashes Over Large Distances

For crashes that cover a large distance or involve several vehicles, the location of the first harmful event should be used as the crash location and not the final resting place of any of the vehicles.

## CONCLUSIONS

This evaluation has determined that the accuracy of traffic crash location data is substantially better than the previous study completed in 2004 (Green & Agent, 2004). A random sample of crashes indicated that 92 percent of all crashes were accurate compared to around 50 percent in the previous study. Most of this improvement can be attributed to the implementation of the MapIt system in eCrash. The crashes that were determined to be incorrectly located were largely due to a lack of reference points. The MapIt system requires the user to click a location on a map which requires the user to know their current location. Since there is not a GPS receiver integrated with the system, the user must use intersecting roadways, roadway geometry and mile-markers to help pinpoint their location.

Several components could be added to the MapIt system to help improve the mapping component, particularly for rural or interstate locations where reference points are not as frequent. The following is a list of additions that could be implemented.

<u>Aerial Photos</u>

Integrating the MapIt system with aerial photos would largely increase the number of reference points to more accurately locate a crash. Buildings, driveways and many other features could be used as reference points. Aerial photos are admittedly very large but several techniques can be employed to minimize the space needed. For example, aerials can only be shown when needed and could be tied to the Kentucky Transportation Cabinet's mapping systems or one of many free mapping services. Furthermore, some agencies could download only the aerials they need.

<u>GPS Receivers</u>

A previous study evaluated the use handheld GPS receivers for locating crashes (Green & Agent, 2004). Unfortunately, human error contributed largely to the use of these devices. GPS receivers are cheaper and a lot more user friendly today. The data show that the mapping system is much more effective in locating crashes. However, the ideal solution would be to integrate the two systems: GIS of MapIt and GPS of receivers. A cheap USB receiver that requires no user interface could be integrated into eCrash to provide an indication of the user's current location. The "distance" between the MapIt location can be compared to an integrated GPS and flagged if it is above a certain level. The GPS data from the already-purchased GPS units could be used to verify the MapIt-based GPS coordinates, but is not ideal due to the propensity for user error.

<u>Reference Logs</u>

The transportation cabinet provides a route log inventory of the roadways in Kentucky that is spatially-enabled. This inventory could be added as a layer to the MapIt system that would indicate many features that can be used as reference points such as: bridges, culverts, mile markers, etc. Mile marker information is already provided in the MapIt system but it requires the user to continually click points along a road until a mile point ending in a zero is found, which can be tedious.

Training

Training could be provided to encourage users to use other tools available to help pinpoint their location. Other mapping sources such as Google® Maps could be used. The importance of a proper location should be emphasized, particularly its role in highway safety. Preference should always be given to state routes over local roads so that the data can be tied to inventory data and be used to calculate crash rates. Additionally, the measure tool in the MapIt system can be used to help gauge distances.

Other Recommendations

Several of the major issues found have been addressed or minimized by the use of the eCrash system. Agencies still using paper reports should be encouraged to move to eCrash or additional training should be provided to minimize some of these errors. The following agencies, cities and counties should be contacted.

- Hancock, Nicholas, Johnson, Jefferson and Whitely Counties all had nearly 30 percent or more of their reports submitted as paper
- The following cities had paper submission rates over 90 percent:
  - Plantation
  - Villa Hills
  - Thealka
  - Sitka
  - Volga
  - Tutor Key
  - Meally
  - Hawesville
- Several agencies had very high paper submission rates. Louisville Metro Police Department had the highest submission rate when considering the number of crashes filed by agency (37 percent paper with 56,189 reports in the study period).

Most of the plotting issues have been minimized; however, the following counties, cities and agencies had the highest percentages of crashes plotted outside of their counties.

- Hancock, Whitely and Fulton Counties had over five percent of their crashes plotted outside of their county
- The cities of Volga, Carter and Hickman had over 20 percent of their crashes plotted outside of their county
- The following agencies had over 10 percent of their crashes plotted outside of their county:
  - Breathitt County Sheriff Dept.
  - Hickman Police Department
  - Lewisport Police Department
  - Johnson County Sheriff Dept.
  - Hancock County Sheriff Dept.
  - Burkesville Police Department
  - Whitley County Sheriff Dept.

As discussed earlier, in 2009 there were 392 crashes plotted outside of their reported county and 5,267 crashes using MapIt (unmodified) that had “distances” over 500 feet when compared to the CRMP data. Both of these should be software edits that prevent or flag the location data as potentially being incorrect. A more advanced county check could be used to ensure that the crash is within the county boundary instead of checking if it is within a box surrounding the county. Furthermore the “distance” between the CRMP data and the GPS location should always be checked and flagged if above 500 feet.

When querying crash data, the proper RT_UNIQUE suffix (e.g. “-000”) should be used to avoid returning undesired results. For instance, a query of Roadway ID = I 0064 between milepoints 0.0 and 5.0 would also return any ramp crash on I-64 throughout the entire state since ramp crashes have very low milepoints. This can be avoided by filtering all ramp crashes by the RT_UNIQUE field. Ramp crash data is substantially better than previous years due to the implementation of the MapIt system. It should be noted that ramps take advantage to the RT_UNIQUE’s suffix information.

Training and possible KYOPS edits should include the importance of roadway selection priority at intersections. That is, choosing a state-maintained road over a local road when available. The following priority table could be used (the highest priority listed first):

1. Interstates*
2. Parkways*
3. US routes*
4. KY routes*
5. All other roads

*The lower number should be used if both roads are the same priority

Other issues and potential solutions are listed below.

- Attention should be given to the glitch linked to the ‘distance from’ field. Several cases seem to imply that there is an issue with the ‘distance from’ field. This should be investigated by KYOPS programmers.

- Between streets should always be selected when a crash occurs in midblock to assist in crash locating. This could be addressed through training and by using a check in KYOPS.

- Attention should be given to how traffic crashes should be properly located. National consensus defines the crash site as the location of the first harmful event. Training should underscore this definition. The Model Minimum Uniform Crash Criteria (MMUCC) has diagrams that could be included in training or in KYOPS to better exemplify this standard.

## REFERENCES


American National Standard. (2007). Manual of Classification of Motor Vehicle Traffic Accidents. (7th, Ed.) Itasca, Illinois: National Safety Council.

Baker, K. S. (2001). Traffic Collision Investigation. Evanston, IL: Northwestern University Center for Public Safety.

Delucia, B. H., and R. A. Scopatz. (2005). *NCHRP Synthesis of Highway Practice 350: Crash Records Systems*. Washington, D.C.: Transportation Research Board of the National Academies.

Graettinger, A. (2000). *Improving Crash Location, Display, and Analysis by Combining CARE and GPS Technologies.* Tuscaloosa, AL. University Transportation Center for Alabama.

Green, E. R., & Agent, K. R. (2004). *Evaluation of the Accuracy of GPS as a Method of Locating Traffic Collisions.* Lexington, KY: Kentucky Transportation Center.

Ogle, J. H. (2007). *NCHRP Synthesis of Highway Practice 367: Technologies for Improving Safety Data*. Washington, D.C.: Transportation Research Board of the National Academies.

Pfefer, Ronald C., & Raub, Richard A. (1998). *Highway Safety Data: Costs, Quality, and Strategies for Improvement.* Evanston, IL. Northwestern University.

Sando, T. (2005). *Quantification of the Accuracy of Low-Priced GPS Recievers for Crash Location.* Fargo, ND. Journal of the Transportation Research Forum.

Sarasua, Wayne A. (2008). *Use of Global Positioning System to Identify Crash Locations in South Carolina*. Clemson, SC: Clemson University.

Pawlovich, Michael D. & Souleyrette, Reginald R. (1996). *A GIS-based Accident Location and Analysis System*. Ames, IA. Center for Transportation Research and Education.