KENTUCKY TRANSPORTATION CENTER
176 Raymond Building, University of Kentucky
859. 257. 4513 www.ktc.uky.edu

Develop a Highway Inventory Data Quality Control Scheme (PL-26)

Kentucky Transportation Center Research Report — KTC-16-27/PL-26-1F
DOI: http://dx.doi.org/10.13023/KTC.RR.2016.27

KTC's Mission

We provide services to the transportation community through research, technology transfer, and education. We create and participate in partnerships to promote safe and effective transportation systems.

Kentucky Transportation Center
176 Oliver H. Raymond Building
Lexington, KY 40506-0281
(859) 257-4513

www.ktc.uky.edu

Research Report
KTC-16-27/PL-26-1F

Develop a Highway Inventory Data Quality Control Scheme (PL-26)

Final Report

By

Mei Chen
Associate Professor
Department of Civil Engineering

Obaidur Kazi
Graduate Research Assistant
Department of Civil Engineering

Tony Fields
Kentucky Transportation Center

College of Engineering
University of Kentucky
Lexington, Kentucky

in cooperation with
Transportation Cabinet
Commonwealth of Kentucky

and

Federal Highway Administration
U.S. Department of Transportation

February 2017

Table of Contents

List of Tables

List of Figures

EXECUTIVE SUMMARY

Since passage of the Moving Ahead for Progress in the 21st Century (MAP-21), transportation agencies have become increasingly reliant on data for planning and programming, asset management, and performance tracking. An efficient asset inventory database with accurate data is critical for states to comply with MAP-21 (and now FAST Act) requirements.

The Kentucky Transportation Cabinet's (KYTC) Division of Planning maintains the Highway Information System (HIS), which is a primary source of this information for transportation professionals at the Cabinet (KYTC). Because highway system is continually changing, it is a challenging task to keep the information in the HIS up to date. Further, the level of accuracy and precision of the initial data collection activities vary among the agencies tasked with data collection.

This study developed a quality check program of the HIS to ensure key items in the database are accurate. KYTC identified a list of priority data items to include in this assessment. They were: (1) Number and width of through lanes; (2) Type and width of medians; (3) Presence and width of auxiliary lanes (e.g., truck climbing, turning, two-way left turn); and (4) Type and width of shoulders.

For this study, Kentucky Transportation Center (KTC) researchers developed a process to systematically sample highway segments and generate data collection points. These points were located on Google Earth and Google Street View so that those programs' native tools could be used to measure the relevant attributes. Researchers then compared the HIS records and observed data to identify possible errors.

This study's analysis verified that the vast majority of the sampled segments have accurate entries in the HIS database. For Through Lane and Shoulder asset types, inconsistencies between the observed and archived data were mostly on roads in small urban areas and rural minor arterials or lower functional groups. Since they account for about 86% of the total mileages in Kentucky, additional sampling of roads in these groups is recommended. Median width was identified as another item that warrants further review. The Auxiliary Lane asset type appears to have the least inconsistency between the observed and archived data, according to random samples.

CHAPTER 1 BACKGROUND

Since passage of the Moving Ahead for Progress in the 21st Century (MAP-21), transportation agencies have become increasingly reliant on data for planning and programming, asset management, and performance tracking. An efficient asset inventory database with accurate data is critical for states to comply with MAP-21 (and now FAST Act) requirements.

The Kentucky Transportation Cabinet's (KYTC) Division of Planning maintains the Highway Information System (HIS), which is a primary source of asset inventory data for transportation professionals at KYTC. KYTC staff strive to ensure that the HIS receives updates when roads are added or modified during construction and maintenance activities. However, the ongoing transformations of the highway system make it a challenging task to keep the information in the HIS accurate and current. Further, the level of accuracy and precision of the initial data collection activities vary among the agencies tasked with its collection.

The goal of this study is to develop a quality check program of the HIS to ensure that key items in the database are accurate. The program will be able to (1) identify where data problems exist, and (2) determine the confidence of the data set.

KYTC identified a list of priority data items to include in this assessment:

1) Number and width of through lanes
2) Type and width of medians
3) Presence and width of auxiliary lanes (e.g., truck climbing, turning, two-way left turn)
4) Type and width of shoulders

These items are currently listed in the HIS database as the LN, MD, AL, and SH asset types, respectively.

CHAPTER 2 RESEARCH APPROACH

A spatial sampling scheme was designed to collect data for this study. The basic steps involved were:

(1) Categorize the roadways by area type and functional classification.
(2) Perform spatial clustering analysis based on the feature in question to identify segments that are considered as outliers.
(3) Collect data using iRAP tools and Google Maps and Google Earth tools.
(4) Perform spatial analysis to detect potential pattern of errors.

2.1 Roadway Categorization

Design standards vary according to roadway type, which is often reflected in functional classifications. Area type and functional classification are combined to create roadway groups. The underlying idea is that roadways in the same group (i.e., in the same functional class and area type groups) are likely to have similar geometric features. For example, interstates are more likely to have 12-foot lanes than roads that fall into other classifications, and roadway sections located in urban areas are likely to have more lanes than those in rural areas.

For this study, the functional groups identified below were created to support the stratified sampling process. This was largely based on the functional classification coding in HIS (FUNCT and FC).

- Interstates and Other Limited Access Roads (i.e., those segments with FUNCT = 1, 2, 11, and 12) with the exception of rural principal arterials (FUNCT =2 and FC = 3)
- Rural Principal Arterials (i.e., those segments with FUNCT = 2 and FC = 3)
- Rural Mountainous Minor Arterials and Major Collectors (FUNCT = 6, 7)
- Rural Non-Mountainous Minor Arterials and Major Collectors (FUNCT = 6, 7)
- Rural Mountainous Minor Collectors and Locals (FUNCT = 8, 9)
- Rural Non-Mountainous Minor Collectors and Locals (FUNCT = 8, 9)
- Large Urban Arterials (FUNCT = 14, 16)
- Large Urban Collectors and Locals (FUNCT = 17, 19)
- Small Urban Arterials (FUNCT = 14, 16)
- Small Urban Collectors and Locals (FUNCT = 17, 19)

Rural Minor Arterials, Collectors, and Locals were further divided by their area type based on consideration of the impact of terrain type on highway design parameters. Mountainous counties include Lawrence, Elliott, Morgan, Menifee, Powell, Estill, Jackson, Laurel, Knox, Bell, Harlan, Letcher, Pike, Martin, Johnson, Magoffin, Floyd, Wolfe, Breathitt, Knott, Lee, Perry, Owsley, Leslie, and Clay Counties. They were identified previously in a separate study (Speed Estimation for Air Quality Analysis) by KYTC.

Urban surface roads were further divided by the size of the urban area because the size is used as a surrogate of travel demand, which in turn is often used to determine number of lanes and other design parameters. For this study, the urban segments within Kenton, Campbell, Boone, Jefferson, and Fayette Counties were considered as in a large urban area. All other urban segments were considered in a small urban area.

2.2 Initial Sampling

For each functional class group, spatial clustering analysis was performed to identify outliers. The purpose of this analysis was to identify the outliers of the data item (e.g., lane width) based on the level of dissimilarity between the value of a given segment and its neighboring segments in the same functional class group. The justification for this procedure was that roadways with similar functional classifications and located in similar types of area tend to have similar design characteristics (e.g., lane width).

Spatial clustering analysis was performed in ArcGIS using overlaid HIS LN, MD, and SH asset types downloaded from the KYTC HIS website. The version was current as of 11/20/2015. Spatial clustering analysis produced a list of segments whose attribute in question (e.g., lane width) was identified as a statistically significant outlier compared to its neighboring segments. These outliers were labeled as HH, LL, HL, or LH. The first letter indicates whether the attribute value of the given segment was high (H) or low (L) compared to its neighbors, and the second letter indicates whether the given segment's neighborhood average was higher (H) or lower (L) than the entire population of the functional classification group. For example, HH describes an attribute value (e.g., lane width) of a given segment which was higher than that of its neighbors while its neighborhood average was higher than the entire population; HL describes an attribute value of a given segment that was higher than those of its neighboring segments while the entire neighborhood's average was lower than the entire population. Figure 2-1 illustrates a spatial clustering analysis based on lane width. For qualitative data items such as median and shoulder types, the data are nominal rather than ratio. As such, it would not be meaningful to perform this analysis.

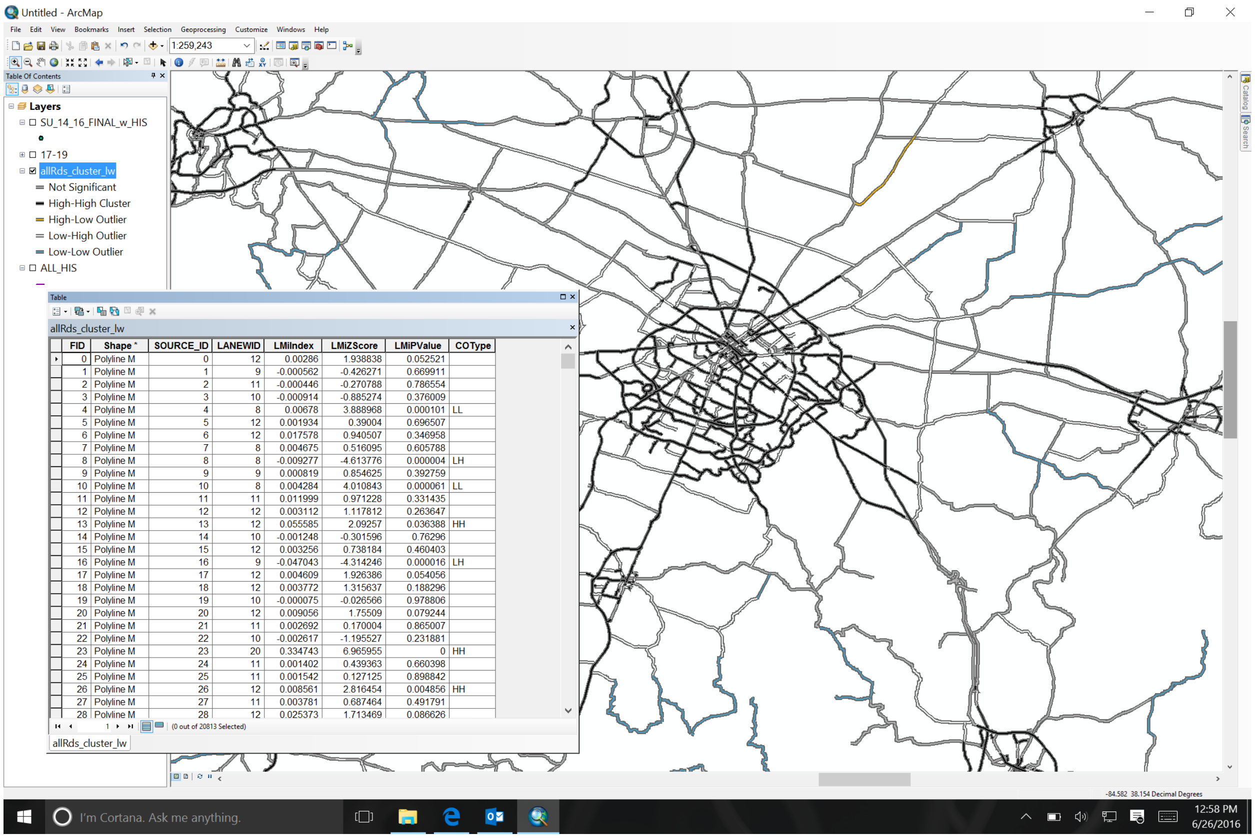


Figure 2-1 Spatial clustering analysis

When samples in addition to what has been identified above are needed, sample segments can be chosen from those with low Moran's I values. Moran's I value is an indicator of the disagreement in attribute values between the segment and its neighboring segments.

2.3 Data Collection Plan

2.3.1 Development of Data Collection Points

Once the network was divided into functional class subsets and focus segments were chosen, point locations were generated at which to collect attribute data. All focus segments for a functional class group were plotted in ArcGIS and the segments were dissolved into a single feature class. The *Create Random Points* tool was used to generate points along the lines at a random interval. The number of points created for each functional class group was a function of the total length within the group. Consideration was also given to the area types in which urban areas had a higher point density than rural areas. In general, more points were generated than were needed for data collection so that extraneous points could be omitted as necessary. Examples of extraneous points are those that were located within the limits of an intersection, in a transition zone between segments with differing values of the attributes to be measured, or very close to another data collection point. Data collectors used their judgement to omit any points they deemed problematic. When points are constrained by the lines on which they are created the output format for the location information of the *Create Random Points* tool is a distance measure along a line (a combination of route and mile point). Route and mile point combinations were plotted in ArcGIS, and latitude and longitude coordinates were appended to each point using the *Add XY Coordinates* tool. Ultimately, the geographic coordinates were used to locate data collection points in Google Earth in order to assess their HIS attributes.

2.3.2 Joining Archived HIS Data to Collection Points

The newly created data collection points had no inherent attribution so it was necessary to append the attribute table to include all attribution from the corresponding HIS line files. This was done to facilitate comparison between the stored HIS data and the information assessed by the data collectors. First, the applicable HIS shapefiles were added to an ArcGIS map document. The exclusive assets (through lanes and medians) were joined by using the *Route Overlay* tool to create a new table that combined the segmentation of both line files. The non-exclusive assets (shoulders and auxiliary lanes) were similarly joined but first the overlapping lines (e.g., cardinal right shoulder and non-cardinal right shoulder) were combined to create one line with multiple attributes using the segmentation method described above. Figure 2-2 illustrates the premise of combined segmentation.

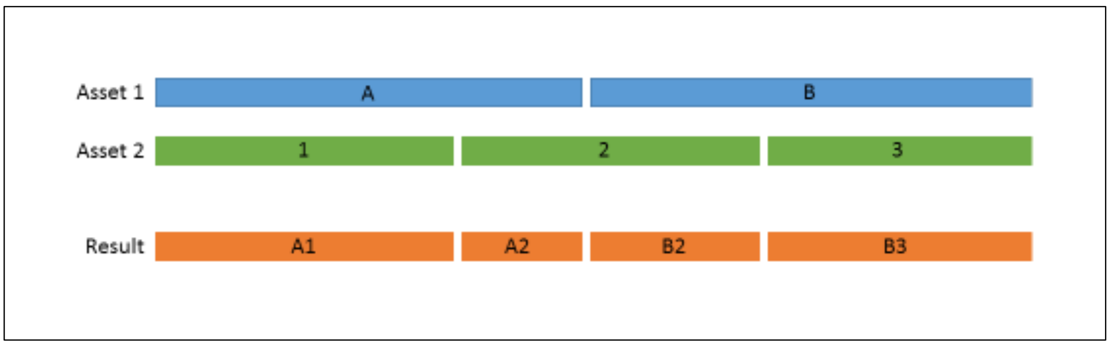


Figure 2-2 Generalized example of combined segmentation

Once all applicable lines and their attributes were incorporated into one combined line, it was possible to append the attribute table of the data collection points with the information associated with the combined line layer. The points were joined to the appropriate line segment based on spatial location. A check was performed to ensure that the route field in the joined line segment matched the route field the data collection point was created from. Any point with conflicting route information was deleted from the database as it would likely lead to a comparison between observed data and archived HIS data from different locations.

2.3.3 Data Collection

Once the data collection points and their corresponding HIS attributes were generated, the database files (.dbf) were converted to Excel and divided into input files containing between 500 and 1,000 points each to distribute to the data collectors. Figure 2-3 provides an overview on the locations of the data collection points.

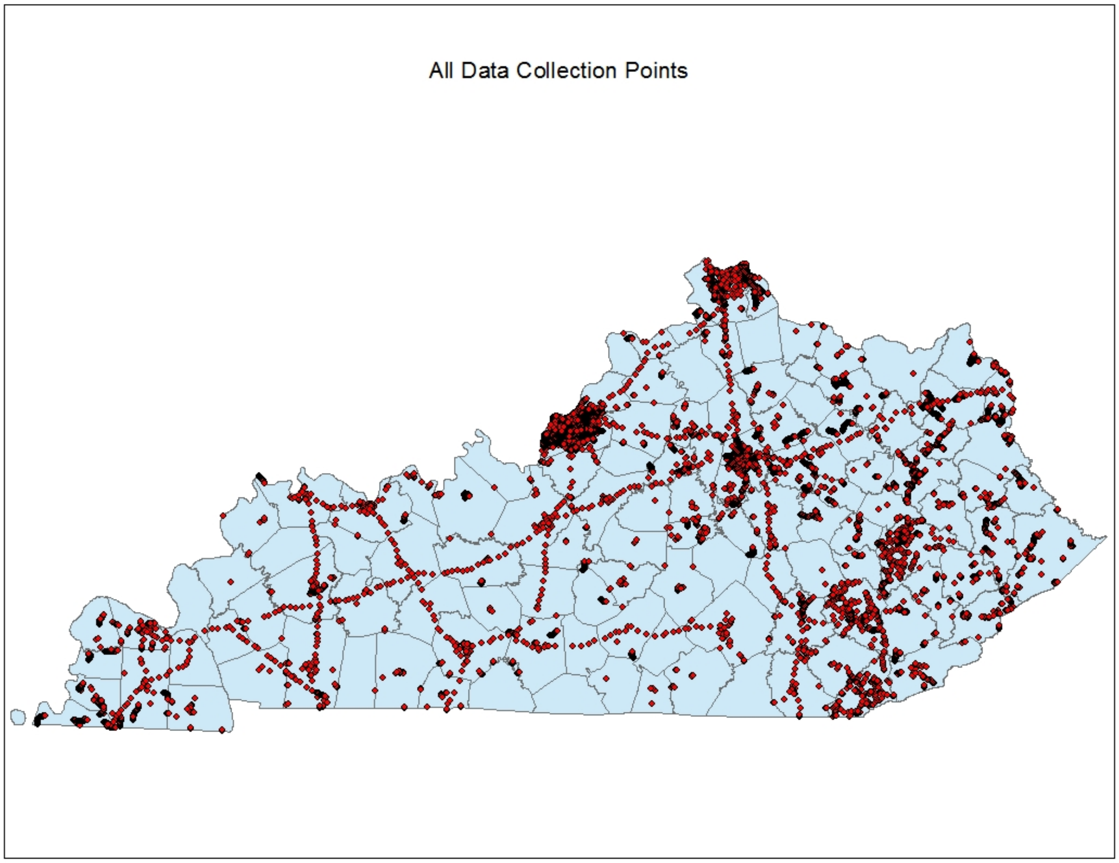


Figure 2-3 Data collection points

All data collectors were trained on KYTC Division of Planning's HIS data guidelines. They also received training on the procedure used to assess all necessary roadway attribution at each data collection location. Data collection began with the Excel-based input file pre-populated with geo-located points. The collector copied the coordinates from the input file and pasted them into the search field in Google Earth Pro. With the map centered on the data collection point, the collector assessed the number of lanes and, using the ruler tool (), measured the width of lanes, medians and shoulders at the location. It was generally necessary to enter the Street View to assess the shoulder type and median type. In the event that an auxiliary lane was present at the location, the type and width were noted. All information was recorded in the database.

Several challenges were encountered during the data collection process. The Google images in some rural area were blurry or outdated, which affected the accuracy of measured widths. In addition, a data collector's judgment was necessary in several cases, such as identifying shoulder type from Google Street View, and whether to consider an extended climbing lane as a through lane. These could potentially complicate the data collection task and require extensive post-processing.

Many data collection points generated by the process were in the vicinity of intersections, where the value of roadway attributes may differ from the dominant values for the entire segment.

While they provide useful information on certain features, such as on auxiliary lanes, they were removed from the evaluation of through lanes, median, and shoulder data.

CHAPTER 3 DATA ANALYSIS

3.1 Data Conversion

Because all data were collected at a point location where the exact feature value may differ from the HIS attribute stored in linear format, point-based highway attribute data were aggregated to the segment level. The goal was to desensitize the point-based data to local changes in attribute values and therefore allow direct comparison with linear data. Using a combination of route and milepoint, each collected point was assigned to its corresponding route segment. The route segments were generated by overlaying the Through Lanes (LN), Median (MD), and Shoulder (SH) attribute tables from the HIS, which resulted in shorter segments as the line file was broken at the endpoints of each HIS segment (see Figure 2-2). Once the point-based data were joined to their corresponding line segments, basic statistics such as average, maximum, and minimum were calculated in ArcGIS for each segment. To avoid having the errors cancel each other out, these statistics were based on the absolute difference between the observed and archived values. Segments with inconsistencies between HIS entries and observations were identified and submitted to KYTC for review.

3.2 Comparison with HIS Record

3.2.1 Asset Type LN

The collected data revealed very few segments (11 out of 3,108) that differed from the HIS data in terms of number of through lanes. In total, 2,371 segments (76.3%) had an average through lane width within 1 foot of the HIS value. Those segments, whose average measured lane width was found to be between 1 and 2 feet off of the archived value, were considered to be in minor disagreement. A total of 591 segments (19%) fell into this category. Only 146 (4.7%) segments were considered in major disagreement with their HIS counterparts. These segments had an average measured lane width differing more than 2 feet. Table 3-1 summarizes the statistics of the differences based on number of segments and mileages.

Table 3-1 Summary statistics of through lanes data

	Number of Through Lanes		Lane Width		
	Agreement	Disagreement	Agreement	Minor Disagreement (1-2 ft)	Major Disagreement (2+ft)
No. of Segments	3097	11	2371	591	146
Percent of Segments (%)	99.6	0.4	76.3	19	4.7
Mileage (mile)	4513.806	11.05	3658.612	741.831	124.413
Percent of Mileage (%)	99.8	0.2	80.9	16.4	2.7

Further details of how the two data sets compare for each functional group with respect to number of segments and mileages are shown in Table 3-2 and Table 3-3, respectively. It appears that the majority of the difference in number of lanes came from urban surface streets and collectors in non-mountainous area. Overall, few disagreements existed between the observed and archived values on number of through lanes. Figure 3-1 shows how the two data sets compare spatially with respect to the number of through lanes.

Table 3-2 Comparison of through lanes data in percent of segments

Functional Group	Total No. of Sampled Segments	Number of Through Lanes		Lane Width		
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)
14_16_LU	947	99.8	0.2	94.5	5.1	0.4
14_16_SU	296	98.6	1.4	57.4	36.5	6.1
17_19_LU	506	99.8	0.2	88.1	7.1	4.7
17_19_SU	140	100	0	24.3	37.9	37.9
6_7_M	301	100	0	69.8	28.9	1.3
6_7_NM	103	96.1	3.9	67	33	0
8_9_M	251	100	0	75.7	19.9	4.4
8_9_NM	330	100	0	40.6	50	9.4
Int_Pkwy	185	100	0	98.4	1.6	0
Non_Pkwy_RPA	49	100	0	83.7	14.3	2
Total	3108	99.6	0.4	76.3	19	4.7

Table 3-3 Comparison of through lanes data in percent of mileage

Functional Group	Total Mileage Sampled (miles)	Number of Through Lanes		Lane Width		
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)
14_16_LU	608.628	99.8	0.2	95.2	3.8	1
14_16_SU	293.55	98.3	1.7	59.9	36.6	3.5
17_19_LU	332.117	100	0	88.1	8.5	3.4
17_19_SU	78.347	100	0	23.4	39.1	37.6
6_7_M	515.809	100	0	81.3	18.5	0.2
6_7_NM	387.508	98.8	1.2	67.8	32.2	0
8_9_M	486.279	100	0	83.2	12.4	4.5
8_9_NM	505.754	100	0	41.7	50.9	7.4
Int_Pkwy	1141.464	100	0	99.4	0.6	0
Non_Pkwy_RPA	175.4	100	0	91.3	4.5	4.2
Total	4524.856	99.8	0.2	80.9	16.4	2.7

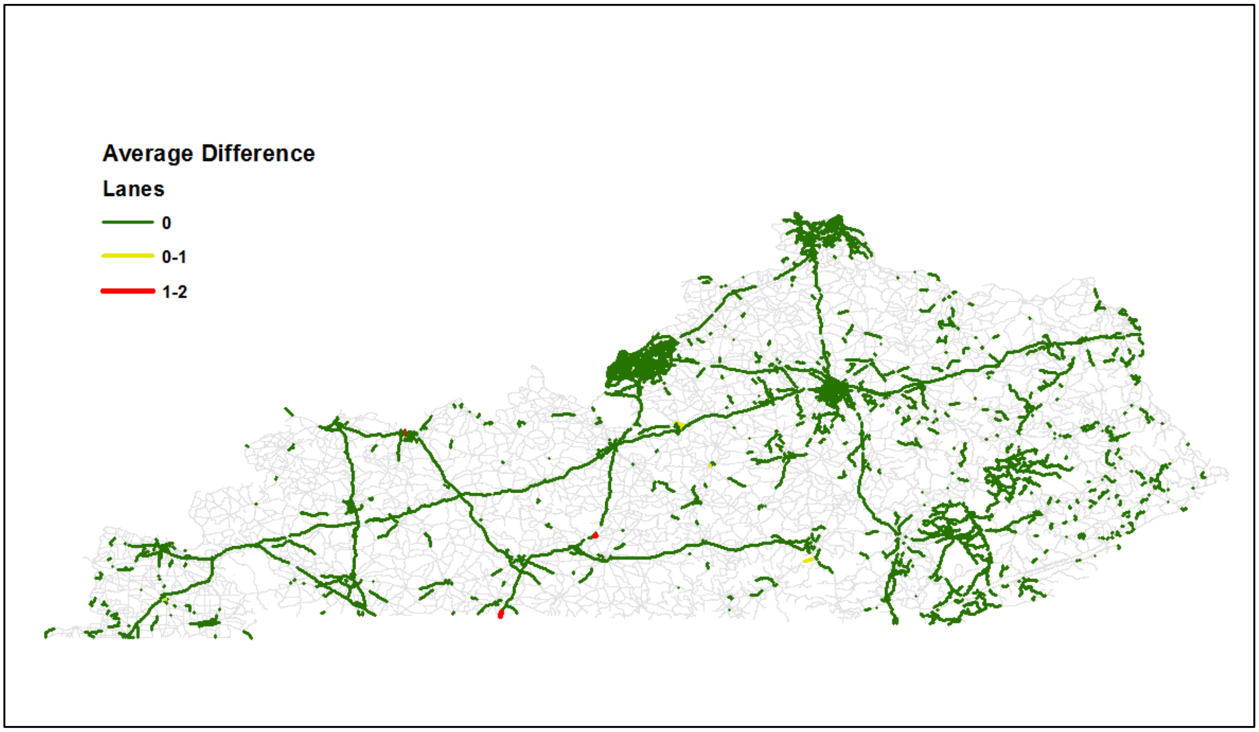


Figure 3-1 Difference in number of through lanes

For lane width, disagreement between the observed and archived values would be more obvious. Overall, about 24% of the segments (or 19% of the mileages) sampled showed some differences (Table 3-2 and Table 3-3). The difference was rather visible on rural and small urban area surface streets except on non-parkway rural principal arterials. For example, for rural collectors and local roads in non-mountainous region (the 8_9_NM group) 50.9% of the 505 total miles sampled showed an average difference of 1-2 feet in lane width, while 9.4% of the miles sampled indicated an average difference greater than 2 feet. In general, it seems that lower functional class roads tend to show more disagreement. However, majority of the differences were in the minor disagreement category. Figure 3-2 shows the location and extent of these differences.

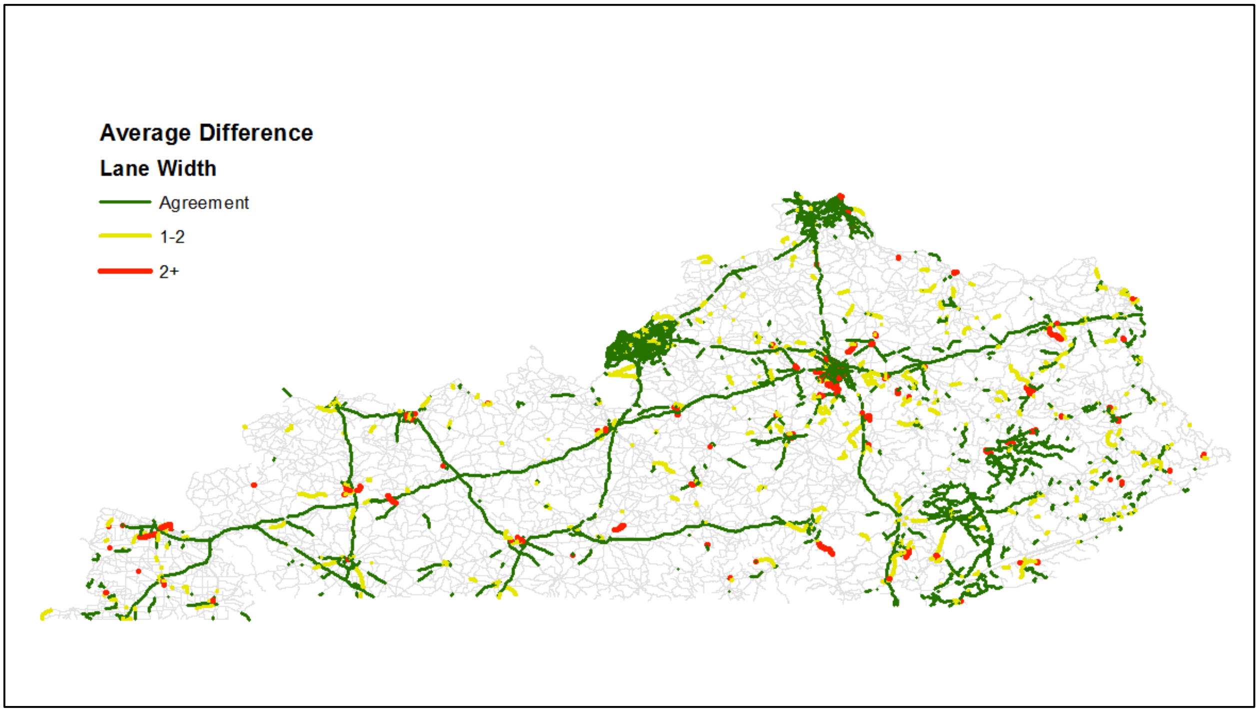


Figure 3-2 **Difference in average lane width**

3.2.2 Asset Type MD

The median type as originally collected had a large proportion of segments that differed from the corresponding HIS data. However, following the guidance from the Division of Planning, several rules were put into place during post-processing of the data that changed most of the collected median type data, resulting in widespread agreement. For example, median type codes 4, 5, 6, and 7 refer to *Raised non-mountable*, *Raised mountable*, *Flush*, and *Depressed*, respectively. As these can be difficult to distinguish from imagery (and also rely on a subjective assessment), the Division of Planning instructed researchers to only report disagreement where there was no question that the archived median type was incorrect. Another rule was that undivided routes (i.e., those with only a painted stripe to separate opposing flows of traffic) must have a median type code 8, representing *None*. Therefore it was ensured that all data collected on undivided roads had a median type of 8. Once these and other similar rules were applied, there were almost no segments whose HIS median type could be definitively identified as incorrect.

Median widths were less subjective and could be more directly compared. Of the 3,108 segments studied, 2,889 (93%) were found to be within 2 feet of the archived value. Minor disagreement, or widths between 2 and 6 feet off from the archived value, occurred found in 116 (3.7%) of the segments. And finally, major disagreement, or a difference greater than 6 feet, was found in 103 (3.3%) segments. Table 3-4 summarizes segment and mileage statistics.

Table 3-4 Comparison of median data

	Median Type		Median Width		
	Agreement (%)	Disagreement (%)	Agreemen t (%)	Minor Disagreement (2-6ft) (%)	Major Disagreement (6+ft) (%)
No. of Segments	3107	0	2889	116	103
Percent of Segments (%)	100	0	93	3.7	3.3
Mileage	4522.691	0	3493.73	441.832	589.294
Percent of Mileage (%)	100	0	77.2	9.8	13

Further details of how median type and width compared between observed and archived values for each functional group with respect to number of segments and mileages are shown in Table 3-5 and Table 3-6, respectively. There were no differences in median type. Figure 3-3 shows where median data were collected.

Table 3-5 Comparison of median data in percent of segments

Functional Group	Total No. of Sampled Segments	Median Type		Median Width		
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (2-6ft) (%)	Major Disagreement (6+ft) (%)
14_16_LU	947	100	0	99.5	0.5	0
14_16_SU	296	100	0	89.9	6.8	3.4
17_19_LU	506	100	0	98.8	1	0.2
17_19_SU	140	100	0	98.6	0.7	0.7
6_7_M	301	100	0	92.7	5.3	2
6_7_NM	103	100	0	100	0	0
8_9_M	251	100	0	99.2	0.8	0
8_9_NM	330	100	0	98.8	0.9	0.3
Int_Pkwy	185	100	0	27	32.4	40.5
Non_Pkwy_RPA	49	100	0	73.5	8.2	18.4
Total	3108	100	0	93	3.7	3.3

Table 3-6 Comparison of median data in percent of mileage

Functional Group	Total Mileage Sampled (miles)	Median Type		Median Width		
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (2-6ft) (%)	Major Disagreement (6+ft) (%)
14_16_LU	608.628	100	0	99.7	0.3	0
14_16_SU	293.55	100	0	91.2	3.2	5.6
17_19_LU	332.117	100	0	99.5	0.5	0
17_19_SU	78.347	100	0	99.7	0.1	0.2
6_7_M	515.809	100	0	97.3	2.1	0.6
6_7_NM	387.508	100	0	100	0	0
8_9_M	486.279	100	0	99.9	0.1	0
8_9_NM	505.754	100	0	99.8	0.2	0
Int_Pkwy	1141.464	99.8	0	19.1	35.5	45.4
Non_Pkwy_RPA	175.4	100	0	64.2	6.8	29
Total	4524.856	100	0	77.2	9.8	13

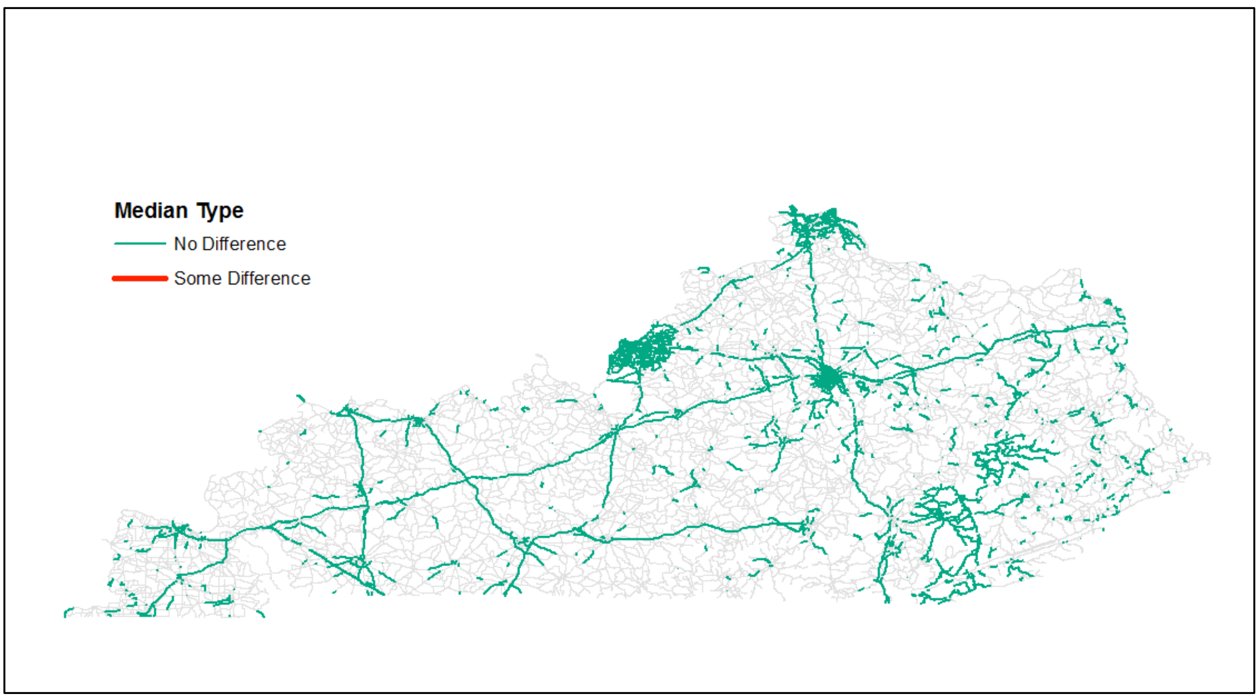


Figure 3-3 Difference in median type

For median width, overall 7% of the segments (or 23% of the mileages) sampled differed, as shown in Table 3-5 and Table 3-6. Among them, majority of the difference in median width came from interstates and other limited access roads, which is the *Int_Pkwy* group, 81% of the 1141 total miles sampled showed some disagreement where 35.5% of the miles showed minor disagreement and 45.4% of the miles show major disagreement which was above 6 feet average difference. There was also significant disagreement in median widths for non-parkway rural principal arterials, where 36% of the 175 total miles sampled showed disagreement while 29% of the miles sampled showed an average difference greater than 6 feet. Other functional groups had a closer alignment between the observed and archived values. However, a median may not be present on many of those roads. Figure 3-4 shows the location and extent of the disagreement.

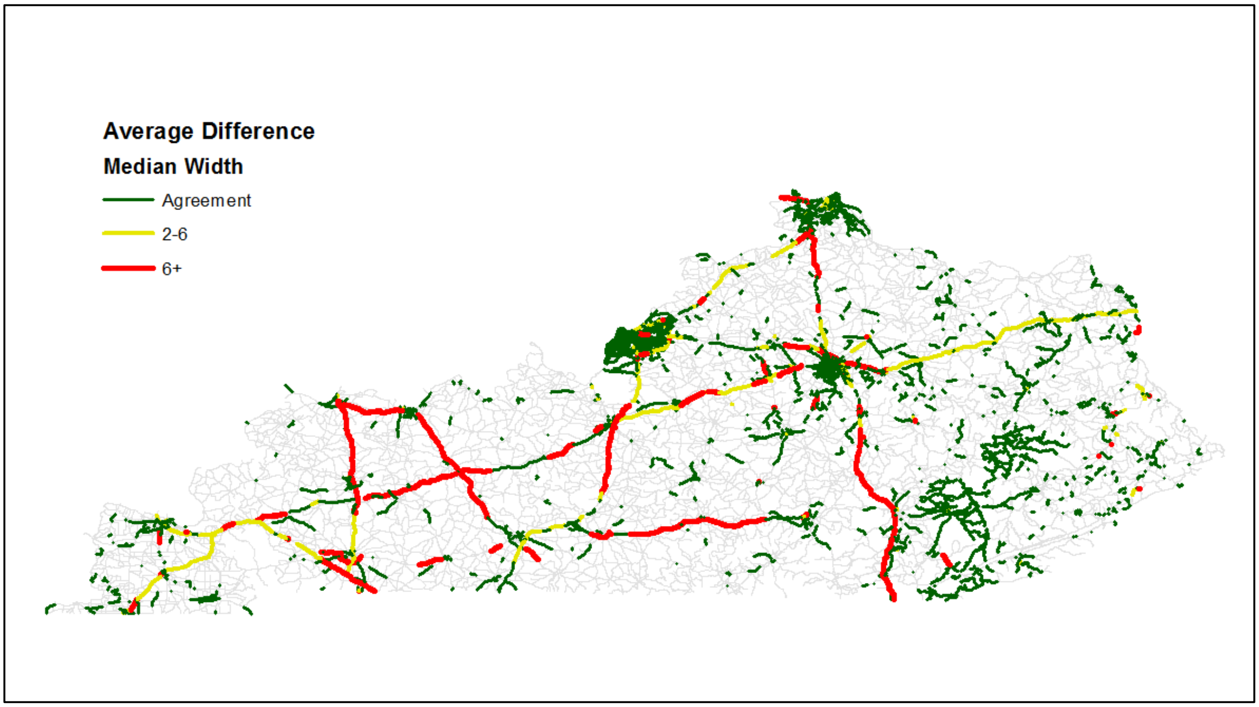


Figure 3-4 Difference in median width

3.2.3 Asset Type SH

The shoulder asset in the HIS accounts for four separate shoulder entries: *cardinal right*, *non-cardinal right*, *cardinal left*, and *non-cardinal left*. These four shoulders were assessed as unique features, and therefore the collected data could be compared individually to the archived HIS data. For both right shoulders, data exist for all segments and the percentages were based on a total number of 3,108. Left shoulders, however, do not exist on all segments, so the percentages were based on the 734 divided segments (1515 miles of roads) that left shoulder data were collected on.

Using satellite imagery to accurately measure the widths of certain shoulder types is challenging. Accordingly, the following categorization was used in the analysis of shoulder width. When the segments had an average shoulder width within 1 foot of the HIS value, they were considered to be in agreement, regardless of shoulder type. Segments with shoulder types 2, 3, or 8 (paved, concrete, and curbed, respectively) whose average measured shoulder width was between 1 and 2 feet off of the archived value were considered to be in minor disagreement. Differences over 2 feet between the observed and archived value were treated as a major disagreement. For segments with shoulder types 5, 6, or 7 (stabilized, combination, and earth, respectively) the minor disagreement range was 1 to 4 feet, and any difference greater than 4 feet was categorized as a major disagreement. Summary statistics for the comparisons are shown in Table 3-7. The locations and extent of the disagreement between observed and HIS shoulder type and width data are shown in Figure 3-5 and Figure 3-6.

Table 3-7 Summary statistics of shoulder data

	CR_Shoulder Type		CR_Shoulder Width				
				Types 2,3,8		Types 5,6,7	
	Agreement	Disagreement	Agreement	Minor Disagreement (1-2ft)	Major Disagree-ment (2+ft)	Minor Disagree-ment (1-4ft)	Major Disagree-ment (4+ft)
No. of Segments	3060	48	2541	135	134	248	50
Percent of Segments (%)	98.5	1.5	81.8	4.3	4.3	8	1.6
Mileage	4461.852	63.004	3633.18	246.456	100.216	490.542	54.462
Percent of Mileage (%)	98.6	1.4	80.3	5.4	2.2	10.8	1.2
	NR_Shoulder Type		**NR_Shoulder Width**				
No. of Segments	2952	156	2436	136	143	333	60
Percent of Segments (%)	95	5	78.4	4.4	4.6	10.7	1.9
Mileage	4149.678	375.178	3482.066	235.879	101.039	649.835	56.037
Percent of Mileage (%)	91.7	8.3	77	5.2	2.2	14.4	1.2
	CL_Shoulder Type		**CL_Shoulder Width**				
No. of Segments	725	9	657	29	46	1	1
Percent of Segments (%)	98.8	1.2	89.5	4	6.3	0.1	0.1
Mileage	1515.794	10.24	1177.037	135.736	212.408	0.172	0.681
Percent of Mileage (%)	99.3	0.7	77.1	8.9	13.9	0	0
	NL_Shoulder Type		**NL_Shoulder Width**				
No. of Segments	724	10	662	27	45	0	0
Percent of Segments (%)	98.6	1.4	90.2	3.7	6.1	0	0
Mileage	1515.608	10.426	1205.64	110.468	209.926	0	0
Percent of Mileage (%)	99.3	0.7	79	7.2	13.8	0	0

An in-depth analysis of the four different shoulder categories comparing the agreement and disagreement of the shoulder type and width among the functional group are presented below.

For the *cardinal right* (CR) shoulder type and width, the difference between observed and archived values along each functional group with respect to the number of segments and mileages is shown in Table 3-8 and Table 3-9, respectively. Most of the difference in CR shoulder type came from the small urban surface streets and rural minor arterials, collectors, and local roads in non-mountainous areas. In general, roads in large urban area and higher functional classes tended to have better agreement with the HIS classification. Figure 3-5 (a) shows the location and extent of the agreement and disagreement.

For the CR shoulder width, the disagreement between the observed and HIS values were more apparent. Overall, approximately 18% of the segments (or 19% of the mileages) sampled showed disagreement, most of which were on small urban roads, rural surface roads, as well as interstates and parkways. For example, in the *17_19_SU* functional group, 68.6% of the 140 segments were in disagreement, with a little less than half in major disagreement. Roads in large urban area and the non-parkway rural principal arterials agreed more with the archived value. The location and extent of the agreement and disagreement are shown in Figure 3-6 (a).

Table 3-8 Comparison of cardinal right shoulder data in percent of segments

Functional Group	Total No. of Sampled Segments	CR_Shoulder Type		CR_Shoulder Width				
					Types 2,3,8		Types 5,6,7	
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)	Minor Disagreement (1-4ft) (%)	Major Disagreement (4+ft) (%)
14_16_LU	947	99.4	0.6	98	0.4	0.8	0.4	0.3
14_16_SU	296	95.6	4.4	74	5.7	4.1	12.2	4.1
17_19_LU	506	100	0	99.8	0	0	0.2	0
17_19_SU	140	88.6	11.4	31.4	17.9	25.7	18.6	6.4
6_7_M	301	100	0	62.5	12	13.6	8.6	3.3
6_7_NM	103	97.1	2.9	59.2	2.9	5.8	30.1	1.9
8_9_M	251	100	0	78.5	2.8	5.6	9.6	3.6
8_9_NM	330	97.6	2.4	63	2.7	3	30	1.2
Int_Pkwy	185	98.9	1.1	78.9	17.3	3.8	0	0
Non_Pkwy_RPA	49	100	0	91.8	4.1	0	2	2
Total	3108	98.5	1.5	81.8	4.3	4.3	8	1.6

Table 3-9 Comparison of cardinal right shoulder data in percent of mileage

Functional Group	Total Mileage Sampled (miles)	CR_Shoulder Type		CR_Shoulder Width				
					Types 2,3,8		Types 5,6,7	
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)	Minor Disagreement (1-4ft) (%)	Major Disagreement (4+ft) (%)
14_16_LU	608.628	99.1	0.9	98.2	0.4	0.3	1	0.1
14_16_SU	293.55	96.8	3.2	66.3	3.4	3.3	22.7	4.3
17_19_LU	332.117	100	0	99.8	0	0	0.2	0
17_19_SU	78.347	91.1	8.9	38.2	12.8	18.1	26.2	4.7
6_7_M	515.809	100	0	78.8	6.9	5.6	7.3	1.4
6_7_NM	387.508	97.5	2.5	60.4	1.1	0.6	34.1	3.8
8_9_M	486.279	100	0	84.4	1.9	2.1	9	2.7
8_9_NM	505.754	97	3	61.2	3	0.7	35	0.1
Int_Pkwy	1141.464	98.6	1.4	83.5	13.9	2.7	0	0
Non_Pkwy_RPA	175.4	100	0	95	0.8	0	3.3	0.9
Total	4524.856	98.6	1.4	80.3	5.4	2.2	10.8	1.2

The difference between the observed and archived values for the *non-cardinal right* (NR) shoulder width and type along each functional group with respect to the number of segments and mileages are shown in Table 3-10 and Table 3-11. For shoulder type, non-mountainous 6 & 7, and small urban roads there was significant disagreement with HIS classification. For the NR shoulder width, similar trends were observed with the CR shoulder width. Most of the disagreement was observed in small urban area, rural collectors, and interstates and parkways. Figure 3-5(a) and Figure 3-6(a) shows the location and extent of the disagreement for the type and width, respectively.

Table 3-10 Comparison of non-cardinal right shoulder data in percent of segments

Functional Group	Total No. of Sampled Segments	NR_Shoulder Type		NR_Shoulder Width				
					Types 2,3,8		Types 5,6,7	
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)	Minor Disagreement (1-4ft) (%)	Major Disagreement (4+ft) (%)
14_16_LU	947	99.2	0.8	94.8	0.7	2	1.5	1
14_16_SU	296	82.4	17.6	74.3	6.1	4.1	11.8	3.7
17_19_LU	506	100	0	100	0	0	0	0
17_19_SU	140	86.4	13.6	31.4	16.4	25	21.4	5.7
6_7_M	301	97.3	2.7	61.8	11	14.3	9.3	3.7
6_7_NM	103	54.4	45.6	59.2	2.9	5.8	30.1	1.9
8_9_M	251	98.8	1.2	75.7	3.6	4.4	13.1	3.2
8_9_NM	330	95.2	4.8	41.5	3.3	3.3	48.8	3
Int_Pkwy	185	100	0	80.5	16.2	3.2	0	0
Non_Pkwy_RPA	49	93.9	6.1	91.8	4.1	0	2	2
Total	3108	95	5	78.4	4.4	4.6	10.7	1.9

Table 3-11 Comparison of non- cardinal right shoulder data in percent of mileage

Functional Group	Total Mileage Sampled (miles)	NR_Shoulder Type		NR_Shoulder Width				
					Types 2,3,8		Types 5,6,7	
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)	Minor Disagreement (1-4ft) (%)	Major Disagreement (4+ft) (%)
14_16_LU	608.628	99	1	96.6	0.6	1	1.4	0.4
14_16_SU	293.55	79.7	20.3	65.7	4	3.6	22.7	4
17_19_LU	332.117	100	0	100	0	0	0	0
17_19_SU	78.347	78.2	21.8	36.8	12.7	18.2	28	4.3
6_7_M	515.809	93.5	6.5	78.8	6.8	5.6	7.4	1.4
6_7_NM	387.508	47.7	52.3	60.4	1.1	0.6	34.1	3.8
8_9_M	486.279	99.2	0.8	81.8	2	1.8	12.9	1.4
8_9_NM	505.754	91.2	8.8	32.6	3.1	0.7	62	1.6
Int_Pkwy	1141.464	100	0	85	12.6	2.4	0	0
Non_Pkwy_RPA	175.4	95.4	4.6	95	0.8	0	3.3	0.9
Total	4524.856	91.7	8.3	77	5.2	2.2	14.4	1.2

Table 3-12 and Table 3-13 show the comparison of each functional group with respect to number of segments and mileages for the *cardinal left* (CL) shoulder type and width. For CL shoulder type, the only disagreement observed was for non-parkway rural principal arterials and small urban arterials. Other functional groups had 100% agreement, mostly because they do not have left shoulder. For the shoulder width, most of the disagreement was observed on interstates and parkways, non-parkway rural principal arterials, and small urban arterials. For example, for the *Int_Pkwy* group, 33% of the 185 total segments showed disagreement, and about 2/3 of these indicated major disagreement. The location and extent of the disagreement for the type and width are shown in Figure 3-5(c) and Figure 3-6(c), respectively.

Table 3-12 Comparison of cardinal left shoulder data in percent of segments

Functional Group	Total No. of Sampled Segments	CL_Shoulder Type		CL_Shoulder Width				
					Types 2,3,8		Types 5,6,7	
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)	Minor Disagreement (1-4ft) (%)	Major Disagreement (4+ft) (%)
14_16_LU	343	99.7	0.3	99.7	0	0.3	0	0
14_16_SU	65	90.8	9.2	80	9.2	7.7	1.5	1.5
17_19_LU	54	100	0	100	0	0	0	0
17_19_SU	7	100	0	100	0	0	0	0
6_7_M	39	100	0	100	0	0	0	0
6_7_NM	5	100	0	100	0	0	0	0
8_9_M	3	100	0	100	0	0	0	0
8_9_NM	11	100	0	100	0	0	0	0
Int_Pkwy	185	100	0	67	11.9	21.1	0	0
Non_Pkwy_RPA	22	90.9	9.1	90.9	4.5	4.5	0	0
Total	734	98.8	1.2	89.5	4	6.3	0.1	0.1

Table 3-13 Comparison of cardinal left shoulder data in percent of mileage

Functional Group	Total Mileage Sampled (miles)	CL_Shoulder Type		CL_Shoulder Width				
					Types 2,3,8		Types 5,6,7	
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)	Minor Disagreement (1-4ft) (%)	Major Disagreement (4+ft) (%)
14_16_LU	201.136	99.9	0.1	99.9	0	0.1	0	0
14_16_SU	51.978	93.8	6.2	89.6	6.1	2.8	0.3	1.3
17_19_LU	16.665	100	0	100	0	0	0	0
17_19_SU	4.747	100	0	100	0	0	0	0
6_7_M	18.436	100	0	100	0	0	0	0
6_7_NM	1.503	100	0	100	0	0	0	0
8_9_M	1.348	100	0	100	0	0	0	0
8_9_NM	2.323	100	0	100	0	0	0	0
Int_Pkwy	1141.464	100	0	70.2	11.4	18.5	0	0
Non_Pkwy_RPA	86.434	92	8	96.3	3.5	0.2	0	0
Total	1526.034	99.3	0.7	77.1	8.9	13.9	0	0

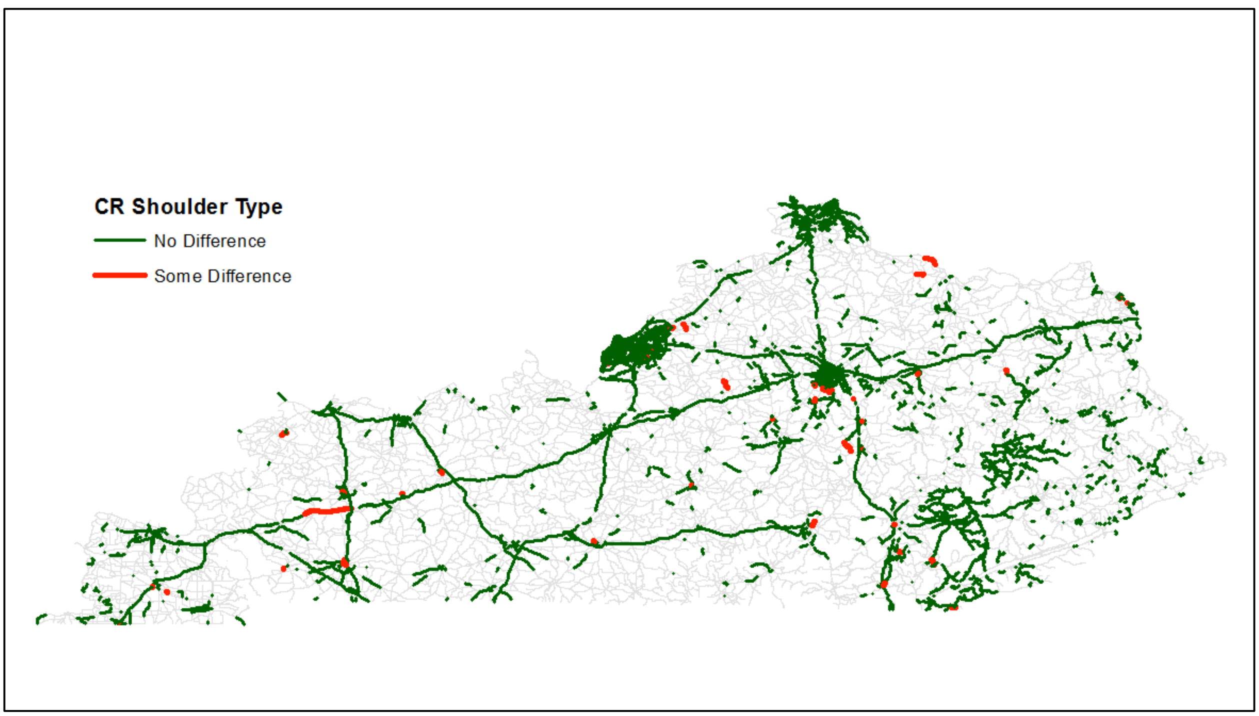
Table 3-14 and Table 3-15 show a similar comparison for the *non-cardinal left* (NL) shoulder type and width. The observations were almost identical to those for the CL shoulder type and width. In general, interstates and parkways, non-parkway rural principal arterials, and small urban arterials showed greater disagreement. The location and extent of the disagreement for the NL shoulder type and width are shown in Figure 3-5(d) and Figure 3-6(d), respectively.

Table 3-14 Comparison of non-cardinal left shoulder data in percent of segments

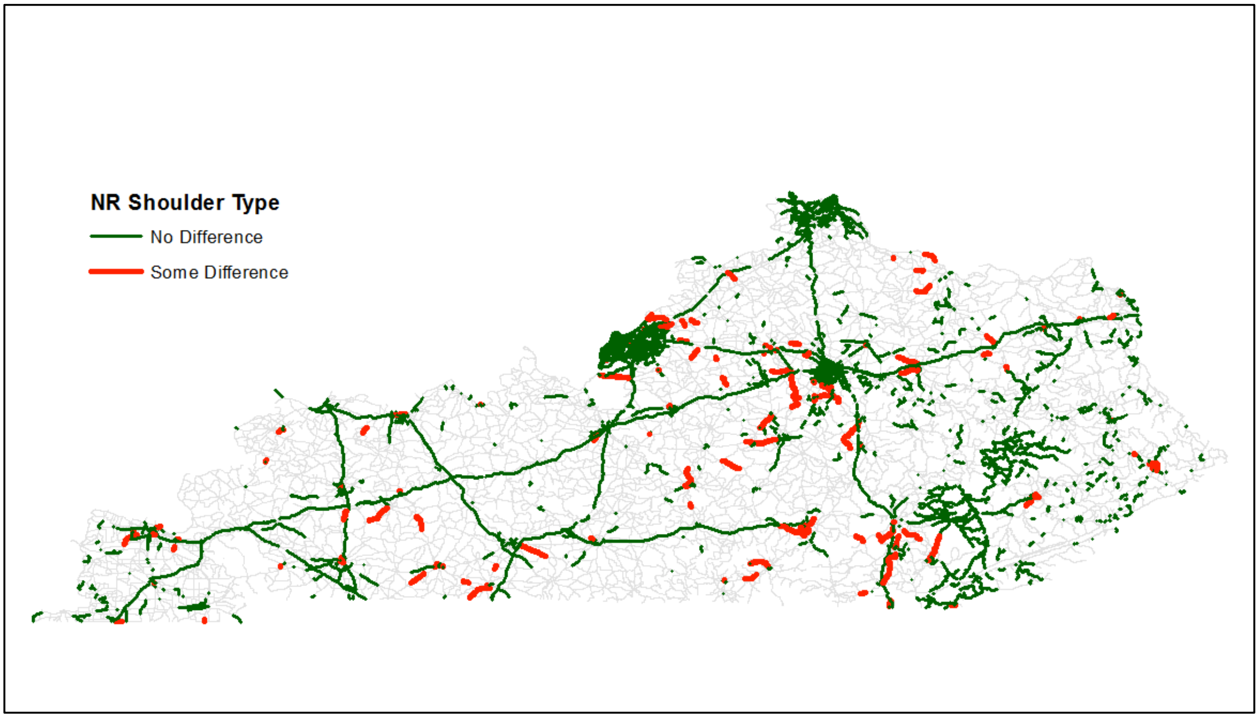
Functional Group	Total No. of Sampled Segments	NL_Shoulder Type		NL_Shoulder Width				
					Types 2,3,8		Types 5,6,7	
		Agreement (%)	Disagree-ment (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)	Minor Disagreement (1-4ft) (%)	Major Disagreement (4+ft) (%)
14_16_LU	343	99.4	0.6	100	0	0	0	0
14_16_SU	65	90.8	9.2	84.6	7.7	7.7	0	0
17_19_LU	54	100	0	100	0	0	0	0
17_19_SU	7	100	0	100	0	0	0	0
6_7_M	39	100	0	100	0	0	0	0
6_7_NM	5	100	0	100	0	0	0	0
8_9_M	3	100	0	100	0	0	0	0
8_9_NM	11	100	0	100	0	0	0	0
Int_Pkwy	185	100	0	67.6	11.4	21.1	0	0
Non_Pkwy_RPA	22	90.9	9.1	90.9	4.5	4.5	0	0
Total	734	98.6	1.4	90.2	3.7	6.1	0	0

Table 3-15 Comparison of non-cardinal left shoulder data in percent of mileage

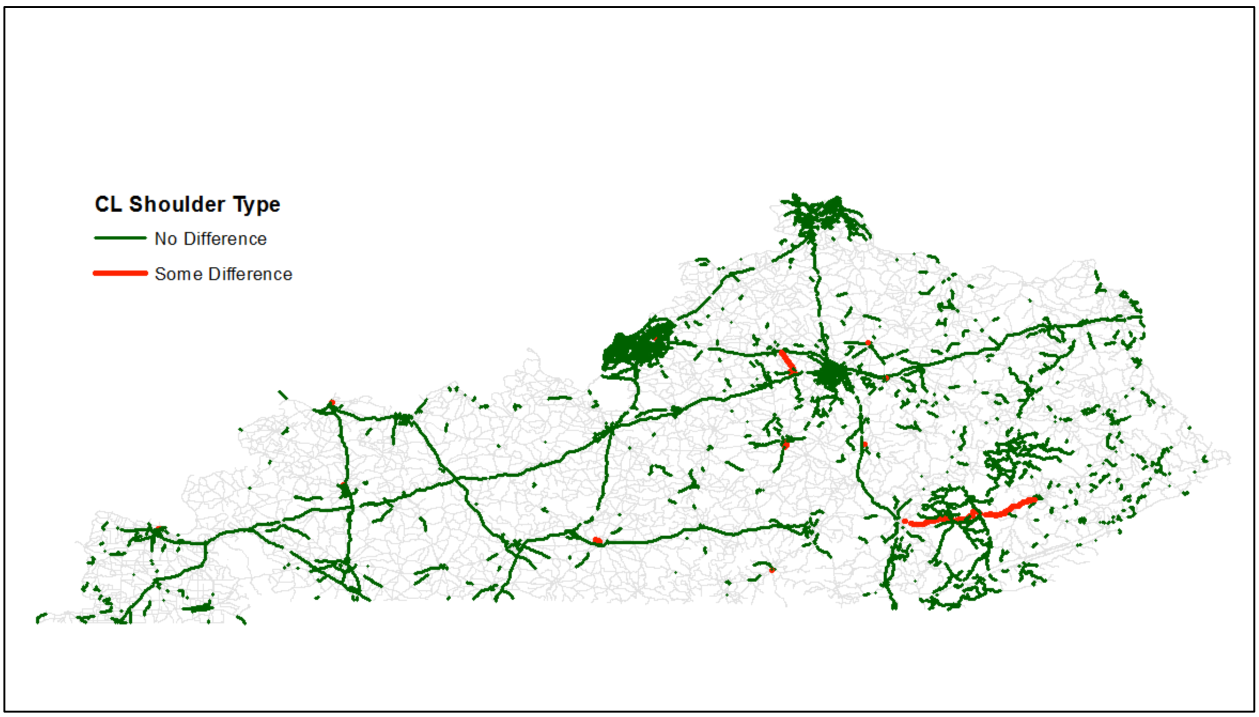
Functional Group	Total Mileage Sampled (miles)	NL_Shoulder Type		NL_Shoulder Width				
					Types 2,3,8		Types 5,6,7	
		Agreement (%)	Disagreement (%)	Agreement (%)	Minor Disagreement (1-2ft) (%)	Major Disagreement (2+ft) (%)	Minor Disagreement (1-4ft) (%)	Major Disagreement (4+ft) (%)
14_16_LU	201.136	99.9	0.1	100	0	0	0	0
14_16_SU	51.978	93.8	6.2	91.6	5.7	2.8	0	0
17_19_LU	16.665	100	0	100	0	0	0	0
17_19_SU	4.747	100	0	100	0	0	0	0
6_7_M	18.436	100	0	100	0	0	0	0
6_7_NM	1.503	100	0	100	0	0	0	0
8_9_M	1.348	100	0	100	0	0	0	0
8_9_NM	2.323	100	0	100	0	0	0	0
Int_Pkwy	1141.464	100	0	72.6	9.2	18.3	0	0
Non_Pkwy_RPA	86.434	92	8	96.3	3.5	0.2	0	0
Total	1526.034	99.3	0.7	79	7.2	13.8	0	0



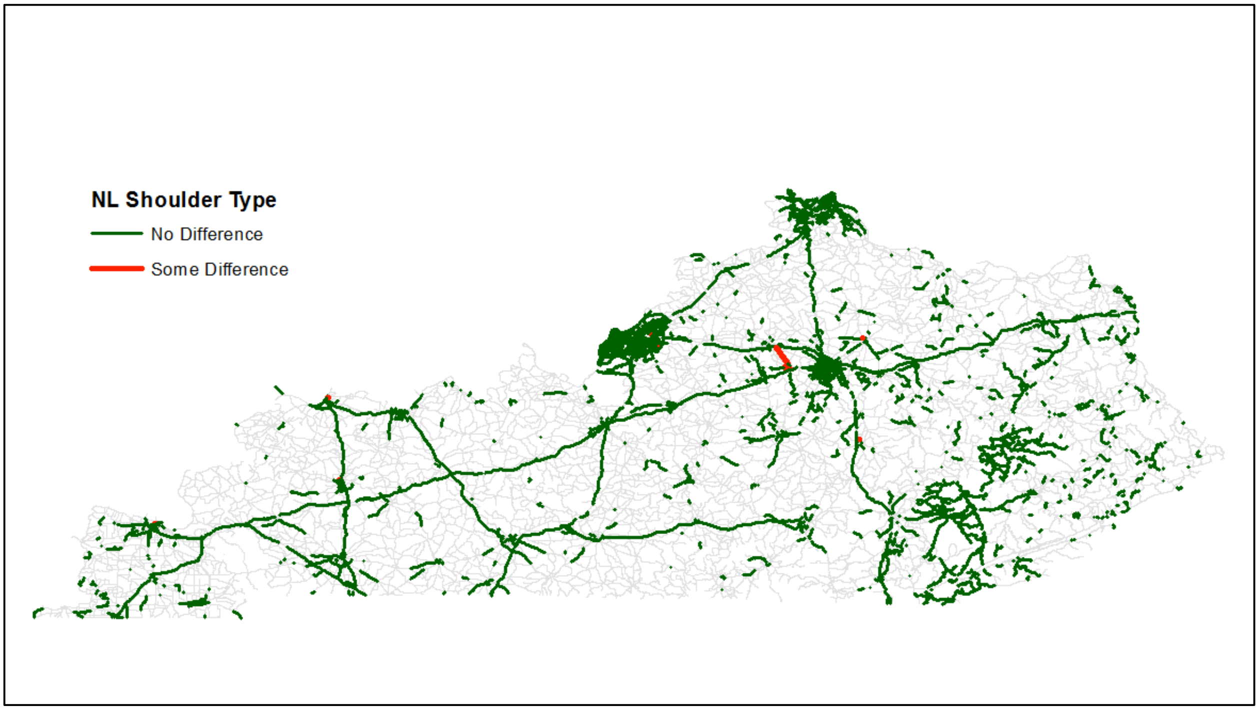
(a) Cardinal right shoulder type



(b) Noncardinal right shoulder type

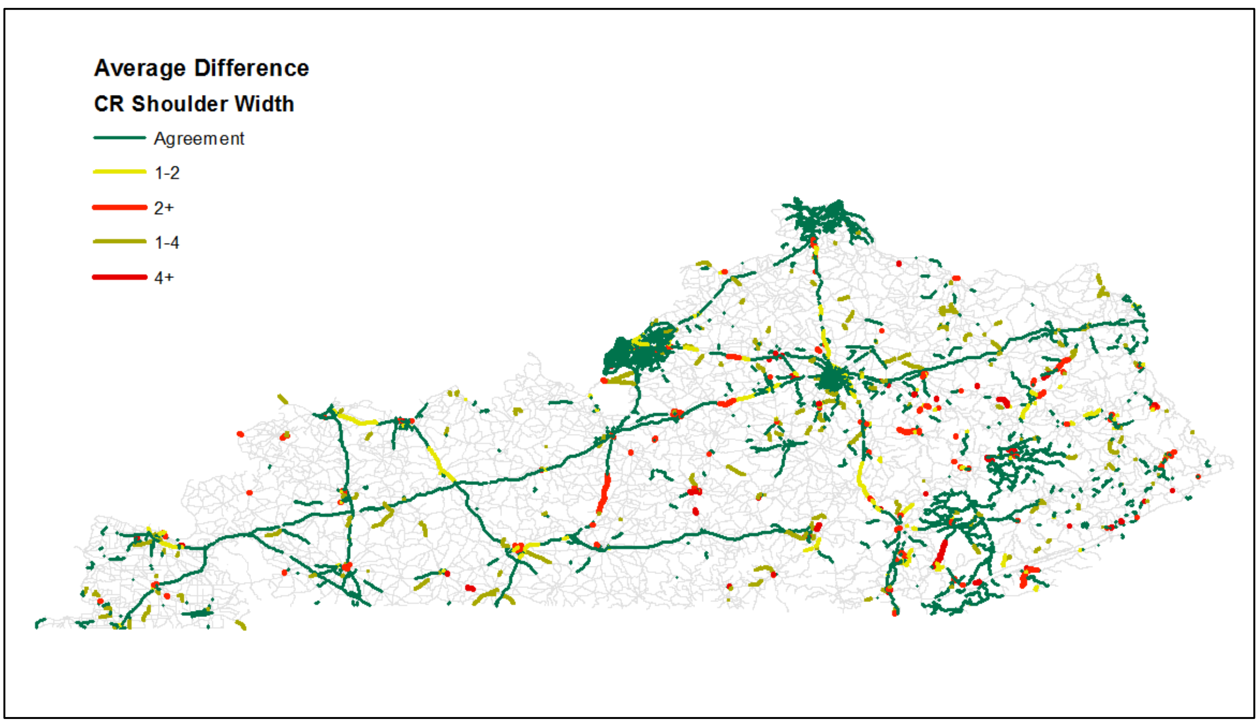


(c) Cardinal left shoulder type

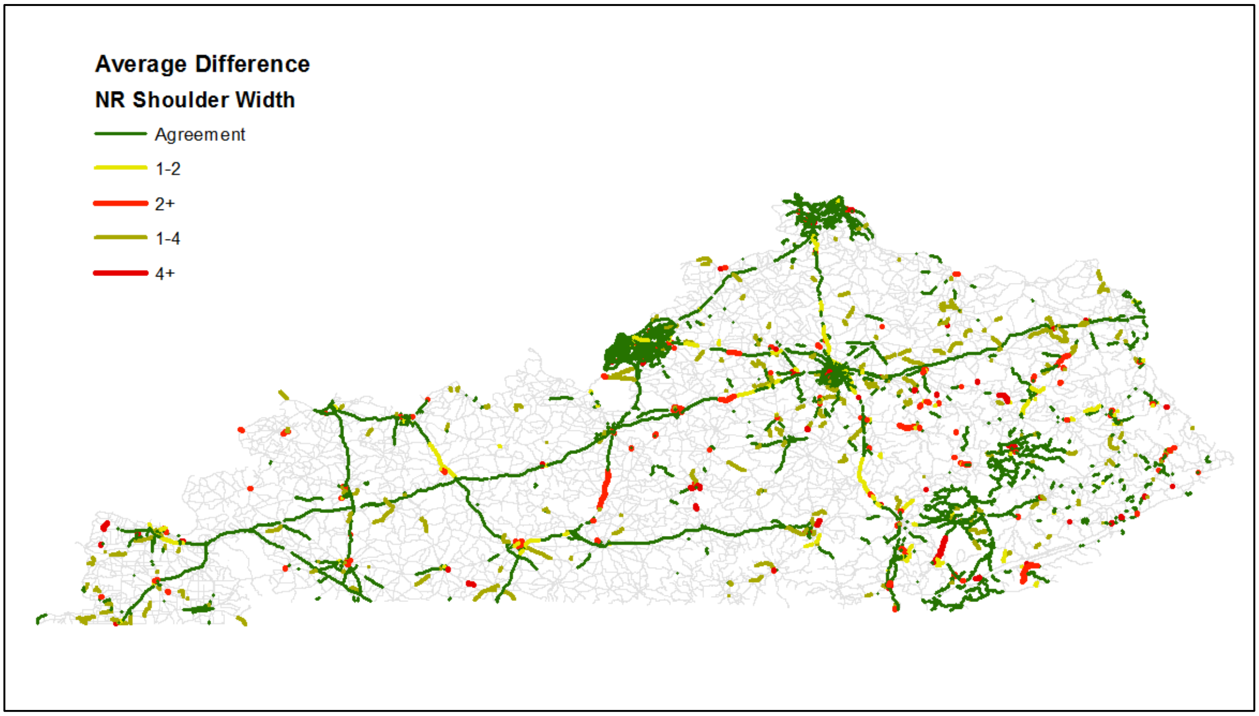


(d) Noncardinal left shoulder type

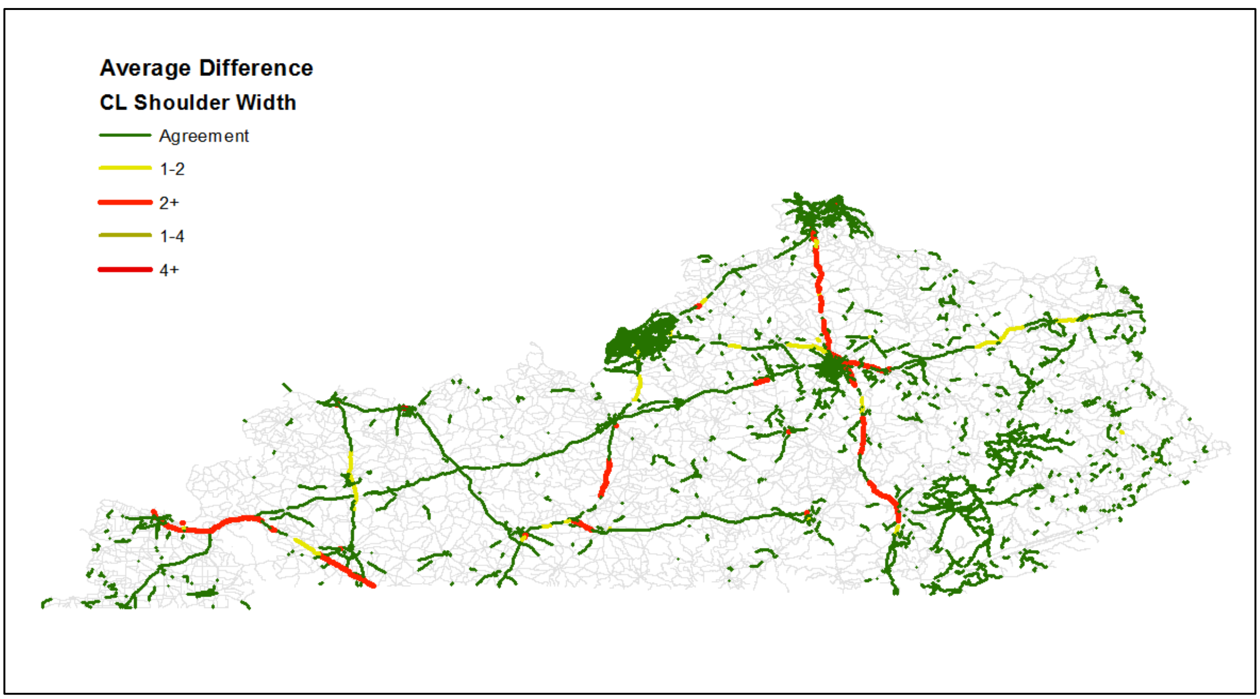
Figure 3-5 Difference in shoulder type



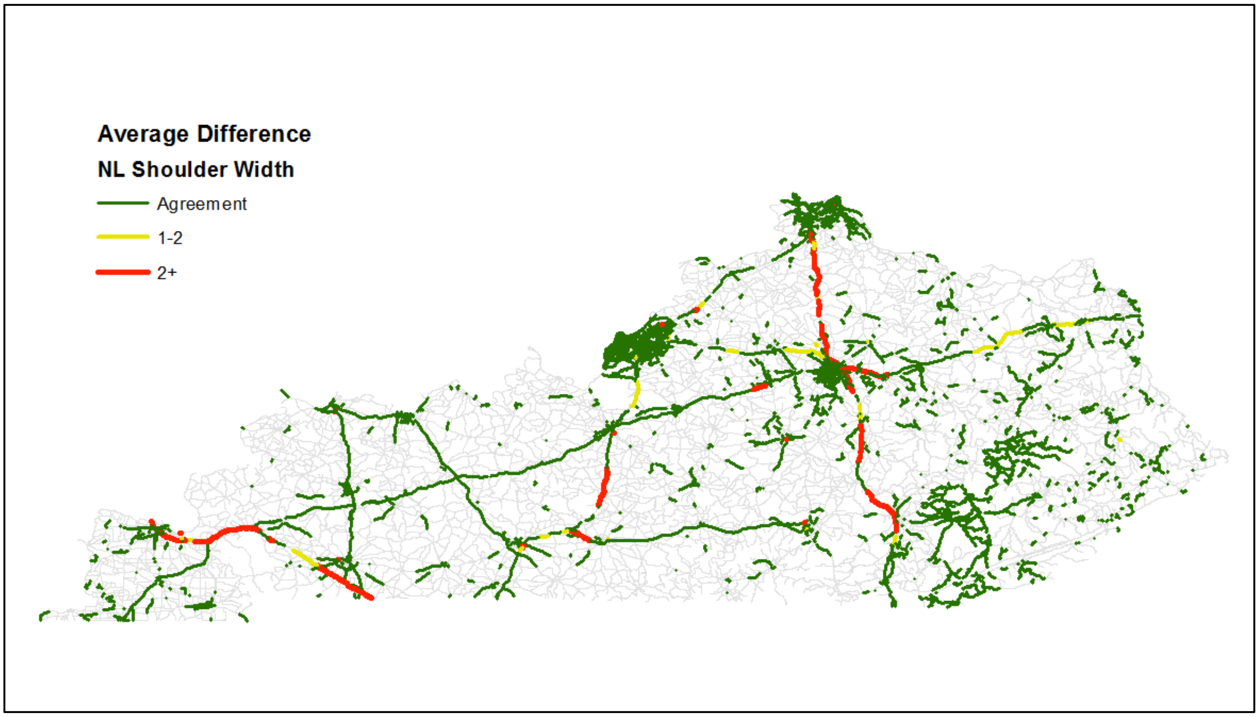
(a) Cardinal right shoulder width



(b) Noncardinal right shoulder width



(c) Cardinal left shoulder width



(d) Noncardinal left shoulder width

Figure 3-6 Difference in shoulder width

3.2.4 Asset Type AL

The auxiliary lane asset is not continuous throughout the highway network and was therefore treated more like a point feature than a line feature. Two types of analyses were performed: 1) data collectors recorded the type and width of any auxiliary lane that was present at the data collection points for which they were already collecting data on the other three assets, and 2) additional data collection points were generated at the midpoint of randomly selected auxiliary lane features in the HIS database.

In the first type of analysis, a total of 525 original points had at least one auxiliary lane present. These data were compared to the HIS records, and all 525 observed auxiliary lanes existed in the HIS database. The second type of analysis was similar to that of the other assets in that an input file was generated and data collectors accessed the given points in order to evaluate the type and width of the auxiliary lane. In total, 1,348 auxiliary lanes were examined. Of those, 28 locations (2.1%) had a width that did not match the HIS record, and at 12 locations (0.9%) there was no auxiliary lane identified. Table 3-16 shows the summary statistics while Figure 3-7 shows the location where disagreement between HIS and observation were found.

Table 3-16 Comparison of auxiliary lane data

Type	Number of Location	Percent of Location
No Difference	1308	97
Inconsistent Width	28	2.1
No Auxiliary Lane	12	0.9

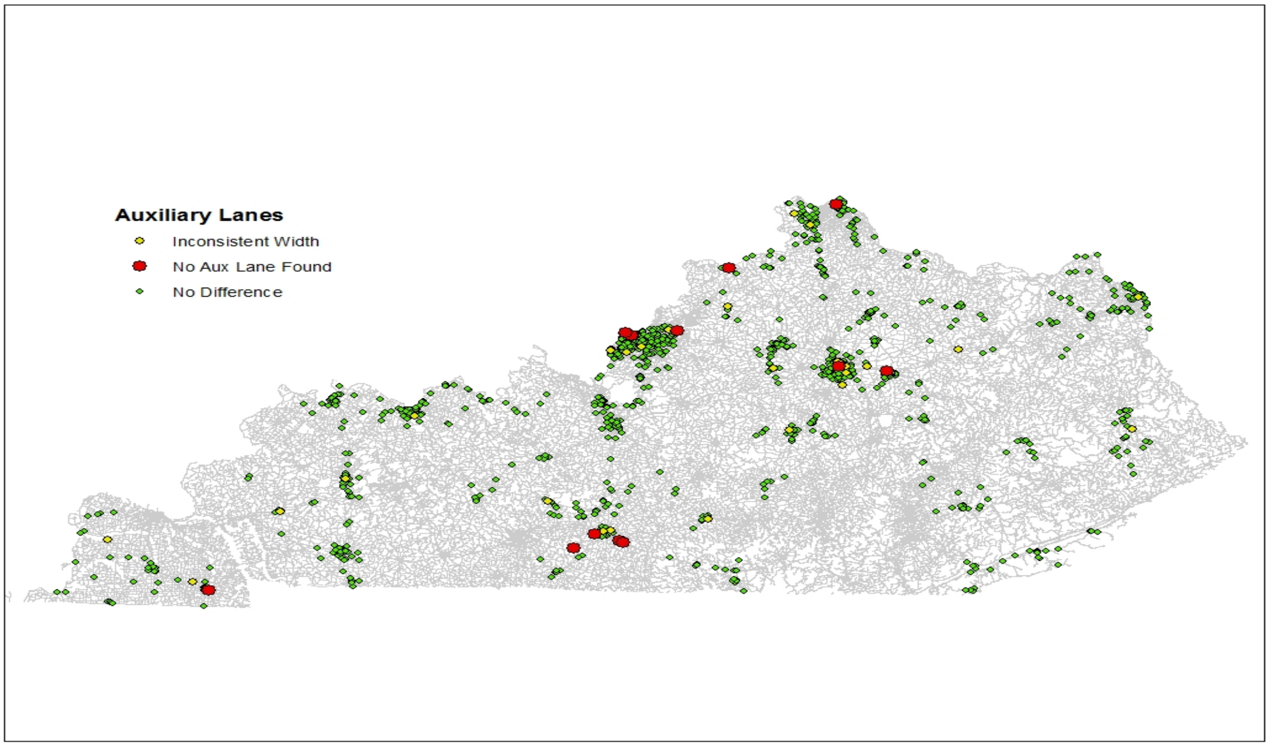


Figure 3-7 **Difference in auxiliary lane data**

CHAPTER 4 SUMMARY

This study used imagery and tools available from Google Maps and Google Earth to verify entries of the HIS asset types Through Lanes, Median, Shoulder, and Auxiliary Lanes. KYTC's photo log images, where available, were also used for additional verification when Google images were over three years old.

While the measurement tool in Google Map can reasonably measure distance (e.g., width of lanes and median), examining attributes such as shoulder type mostly relied on street views available from Google Maps or Photo Log. Subjective judgment was sometimes needed to determine the value of such items. In this case, only when the observed data amounted to a significant departure from the HIS record was the item be flagged as suspicious.

The study's stratified spatial sampling approach yielded a total of 4,525 miles of roads (or about 15% of the total statewide mileages) being selected for attribute verification. Table 4-1 shows the distribution of mileages among functional groups. The vast majority of the segments sampled showed consistent attribute values with those in HIS. For LN, MD, and SH asset types, approximately 20% of the mileages sampled had inconsistent widths, most of which fell into the minor disagreement category. However, it should not be inferred that the 20% inconsistency rate applies uniformly for all roads in the state. This is because the selection of sampling locations was not entirely random. Instead, most were selected using spatial clustering analysis with an embedded prior knowledge (or more accurately, suspicion) aiming to identify discrepancies in attribute values (e.g., lane width) among neighboring roads.

Table 4-1 Percentage roads sampled

Functional Group	Statewide Mileage	State Maintained Mileage	Sampled Mileage	Percent Sampled
14_16_LU	738.732	553.731	608.628	82.39
14_16_SU	1260.871	1218.371	293.55	23.29
17_19_LU	564.087	198.479	332.117	58.88
17_19_SU	1143.86	855.601	78.347	6.85
6_7_M	2002.781	2002.781	515.809	25.76
6_7_NM	6142.598	6132.69	387.508	6.31
8_9_M	2767.098	2765.194	486.279	17.58
8_9_NM	11338.98	11325.03	505.754	4.47
Int_Pkwy	1365.471	1365.471	1141.464	83.6
Non_Pkwy_RPA	1216.69	1216.69	175.4	14.42
Total	28541.17	27634.04	4524.856	15.86

Figure 4-1 shows the distribution of disagreements, where present, based on combined LN, MD, and SH width comparison. The color reflects the highest level of inconsistency in any of the width attributes in these three asset types. For example, if a segment has major disagreement in median width, but is in agreement with respect to both lane width and all shoulder widths, it would be categorized as a *Major Disagreement in* the map. A segment was classified as an *Agreement* when none of its width attributes in all three asset types showed any disagreement with archived values in HIS.

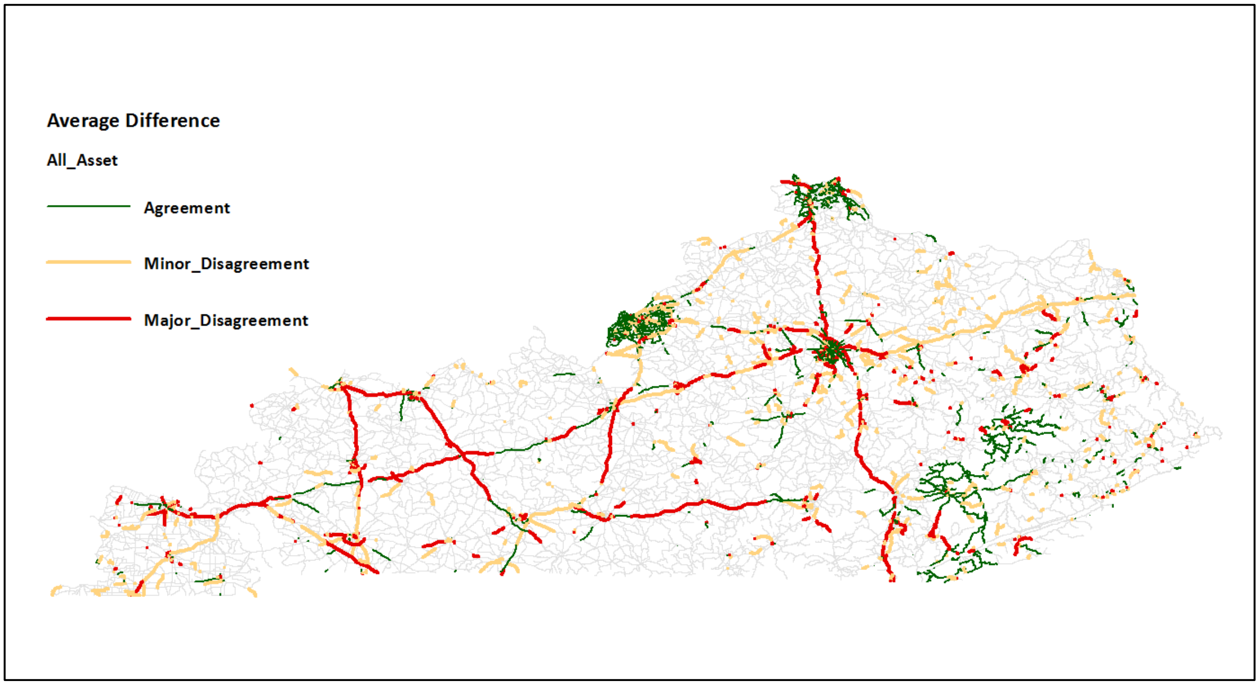


Figure 4-1 Combined LN, MD, SH width distribution

The major disagreement on interstates and parkways was mostly associated with median width. Further investigation revealed some correlation between the median width disagreement and the presence of left shoulders. For example, at location (37.3325811677, -86.9607140355) shown in Figure 4-2, the median width was measured at 30 feet (shown in the photo on the right side) according to KYTC guideline. However, the archived median width in HIS is 24 feet, which is roughly the distance between the edges of the pavement (shown in the photo on the left side). The CL and NL shoulder widths at this location were both 4 feet, and they agreed with those in HIS. Most of the discrepancies on median width appeared to be of similar scenario. It is recommended that further review of median width entries be conducted.

Figure 4-2 **Median width issue at (37.3325811677,-86.9607140355)**

Segments that had no inconsistency for any width attributes appear mostly in large urban areas. Several (Bell, Breathitt, Clay) counties in southeastern Kentucky that were heavily sampled also had most of its width-related entries validated.

Analyses in Chapter 3 showed that for Through Lanes and Shoulder asset types, most of the disagreements were observed on surface streets in small urban areas and rural roads in minor arterial and lower functional groups. Although hundreds of miles were sampled within each of these groups, except for collectors and local roads in small urban area, the percentage of miles sampled remained relatively low compared to other functional groups. It is recommended that additional mileages should be sampled from roads in these categories.

Data associated with the Auxiliary Lane asset type revealed few cases of disagreement between the HIS's observed and archived values. Beyond the locations and extent of the disagreement shown in Figure 4-1, as well as the details for each asset type, no other spatial pattern could be discerned.

This project is the first of its kind to use satellite imagery to verify highway inventory data using both qualitative and quantitative measures. Training data collectors is a critical step toward minimizing error due to judgment inconsistencies. Obsolete satellite imagery encountered from time to time, especially in rural areas, complicated the tasks. Nevertheless, the workflow on generating locations and processing data has become more efficient after several rounds of data collection.