The Importance of Highways to U.S. Agriculture

Methodology Whitepaper

03/30/2021

U.S. Department of Transportation
Research and Innovative Technology Administration
John A. Volpe National Transportation Systems Center

Volpe

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
03/30/2021

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
02/2021 – 03/2021

4. TITLE AND SUBTITLE
The Importance of Highways to U.S. Agriculture: Methodology Whitepaper

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Wilke, Peter; Gilmore, Michelle; Baker, Gary

5d. PROJECT NUMBER
51OS96A119

5e. TASK NUMBER
UK607, UK608

5f. WORK UNIT NUMBER
V-320

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
John A. Volpe National Transportation Systems Center
U.S. Department of Transportation
55 Broadway Street,
Cambridge, MA 02142

8. PERFORMING ORGANIZATION REPORT NUMBER
DOT-VNTSC-USDA-21-02

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Agricultural Marketing Service U.S.
Department of Agriculture Stop 0266 USDA
AMS TM TSD 1400 Independence Ave SW
4523-S

10. SPONSOR/MONITOR'S ACRONYM(S)
USDA AMS TM TSD

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

No restrictions.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This white paper describes the methods and data used to develop the agricultural freight flows and performance section of the U.S. Department of Agriculture (USDA) report *The Importance of Highways to U.S. Agriculture*. The contents of this document describes how the project team prepared, processed, and analyzed the data for the commodity flows and corridor conditions and performance analysis presented in Appendix B of the report. This whitepaper is a compliment to Appendix C of the report, providing more technical details.

15. SUBJECT TERMS
agriculture, freight, highway, freight data, freight analysis

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Adam Sparger
U	U	U	UU	22	**19b. TELEPHONE NUMBER** *(include area code)* 202-205-8701

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Contents

Preface

This document describes the methods and data used to develop the agricultural freight flows and performance section of the U.S. Department of Agriculture (USDA) report *The Importance of Highways to U.S. Agriculture*. The contents of this document describes how the project team prepared, processed, and analyzed the data for the commodity flows and corridor conditions and performance analysis presented in Appendix B of the report. This whitepaper is a compliment to Appendix C of the report, providing more technical details.

1 Input Data

To identify high-volume domestic agriculture highways and to analyze infrastructure condition and agricultural freight performance for selected corridors, the project team used the following datasets and associated attributes:

- *Transearch Database – IHS Markit (2018)*
  - Commodity flow data including tonnage, market value, and truck units
- *Highway Performance Monitoring System (HPMS) – Federal Highway Administration (FHWA) (2017-2018)*
  - Average Annual Daily Traffic (AADT), urban/rural designation, pavement condition, functional class
- *All Road Network of Linear Referenced Data (ARNOLD) – FHWA (2017-2018)*
  - Shapefile which HPMS data is attached to
- *National Performance Management Research Data Set (NPMRDS) – FHWA (2018)*
  - Derived Travel Time Index (TTI) and Truck Travel Time Reliability (TTTR)
- *National Bridge Inventory (NBI) – FHWA (2019)*
  - Bridge condition
- *Fatality Analysis Reporting System (FARS) – National Highway Transportation Safety Administration (2014-2018)*
  - Crash fatalities involving trucks

2 Agriculture Commodity Flows

The U.S. highway network is extensive and contains hundreds of thousands of highway miles. Conducting the level of analysis intended for this study on the full highway network is both time and resource intensive. Instead, the project team identified High-Volume Domestic Agriculture Highways (HDAH), the highways that carry large volumes of the domestic commodities studied for this report in terms of tonnage and market value, as well as 17 analysis corridors.

2.1 Identify Baseline Network

The project team used 2018 domestic agricultural commodity flow data from the IHS Markit Transearch database to define the full highway network, from which the HDAH and the analysis corridors were identified. The full network is built on domestic (non-imports) county-to-county truck flows for a representative sample of agricultural commodities selected by USDA. The commodities included in this study, and their corresponding Standard Transportation Commodity Codes (STCC), are listed in Table 1.

Table 1. Focus Agricultural Commodities

Commodity Group	Commodity	STCC
Grain	Corn	01132
	Soybeans	01144
	Wheat	01137
Fruits	Apples	01221
	Oranges	01214
	Strawberries	01293
	Watermelons	01392
Vegetables	Dry Onions	01318
	Lettuce	01335
	Potatoes, other than sweet	01195
Milk & Dairy Products	Dairy farm products	0142
	Processed whole milk, skim, cream or fluid products	2026
Meat Perishables	Meat, fresh or chilled	2011
	Meat, fresh-frozen	2012
	Dressed Poultry, fresh or chilled	2015
	Dressed Poultry, fresh-frozen	2016
Livestock	Livestock	0141
Poultry	Live Poultry	0151

Source: Volpe Center analysis

As part of the IHS Markit Transearch database, commodity flow data for volume (tons), value (dollars), and shipment units (truck counts) were aggregated to the feature level. IHS Markit's geospatial road network, for which the commodity flow data are assigned, was developed based on the original network created by the Oak Ridge National Lab (ORNL), and has been maintained and updated overtime by IHS Markit. This same road network developed by ORNL was also used as the base network for FHWA and the Bureau of Transportation Statistics' Freight Analysis Framework (FAF). Because the Transearch database includes county-to-county flows, a limitation is that there is no within-county flow data available.

2.2 High-Volume Domestic Agricultural Highways (HDAH)

The full network defined above was subset in order to identify the HDAH. The full network was first categorized into “Interstates” (functional class 1) and “non-Interstates” (all other functional classes). Given that Interstates tend to carry higher volumes than non-Interstates, distinguishing between the two allowed for non-Interstates with high volumes of the focus commodities to be included in the analysis while excluding the less important Interstates. The project team then calculated the total value and total units for each feature by commodity.

For each classification type (Interstates and non-Interstates), the cumulative percentage for each measure was then calculated. Features which were within the top 80% of the cumulative percent type for value or units (sorting largest to smallest) for at least one commodity type were retained. The benefit of this approach is that it does not result in half of the HDAH being compiled of Interstates and the other half being of non-Interstates, but rather results in the features within the top 80% for each classification type. 88% of the features in the HDAH have two or more focus commodities within the top 80% for either value or units, and 54% have all nine commodities within the top 80%. This data-driven approach to develop the HDAH resulted in a network that was 17% of the full Transearch data, accounting for 80% of the commodity flows, and retained 95% of road features that were connected to at least one other feature.

Commodity flow data can be summarized in two ways using the IHS Markit Transearch database. The first is by feature, which assigns the county-to-county flows to individual network features. The second is by corridor, which assigns the flows for each route (origin-destination pair) that traverses the features along a corridor. Assigning flows by corridor removes any double counting of data when identifying the flow quantity across the network because it does not assign to each feature and then sum. Whereas assigning flows by feature can be used to identify the density distribution among individual features. For example, in Figure 1, Route A consists of three features – X, Y, and Z – and a total tonnage flow of 100 tons. In Figure 2, Route B consists of three features – W, Y, and V – and a total tonnage flow of 150 tons. To summarize by route, the total tonnage for each route would only be counted once. To summarize by feature, each individual feature would be assigned the tonnage value for each route. In Figure 3, each feature on Route A (X, Y, and Z) is assigned 100 tons, and each feature on Route B (W, Y, and V) is assigned 50 tons. Feature Y is part of both routes and is assigned 100 tons from Route A and 50 tons from Route B, for a total of 150 tons.

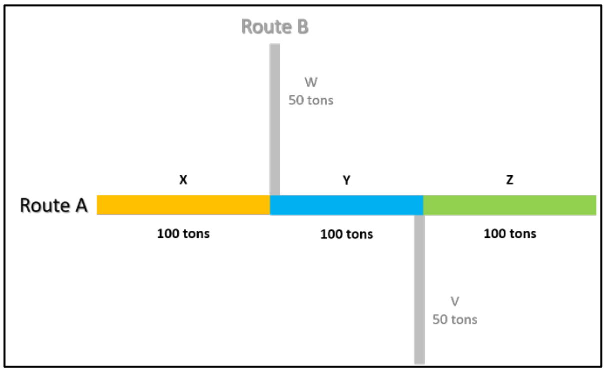

Figure 1. Route A with a total tonnage of 100 tons and constructed of three features
Source: Volpe Center analysis of 2018 Transearch data

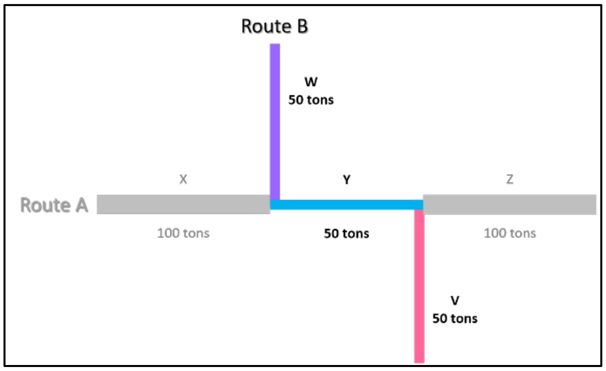

Figure 2. Route B with a total tonnage of 50 tons and constructed by three features
Source: Volpe Center analysis of 2018 Transearch data

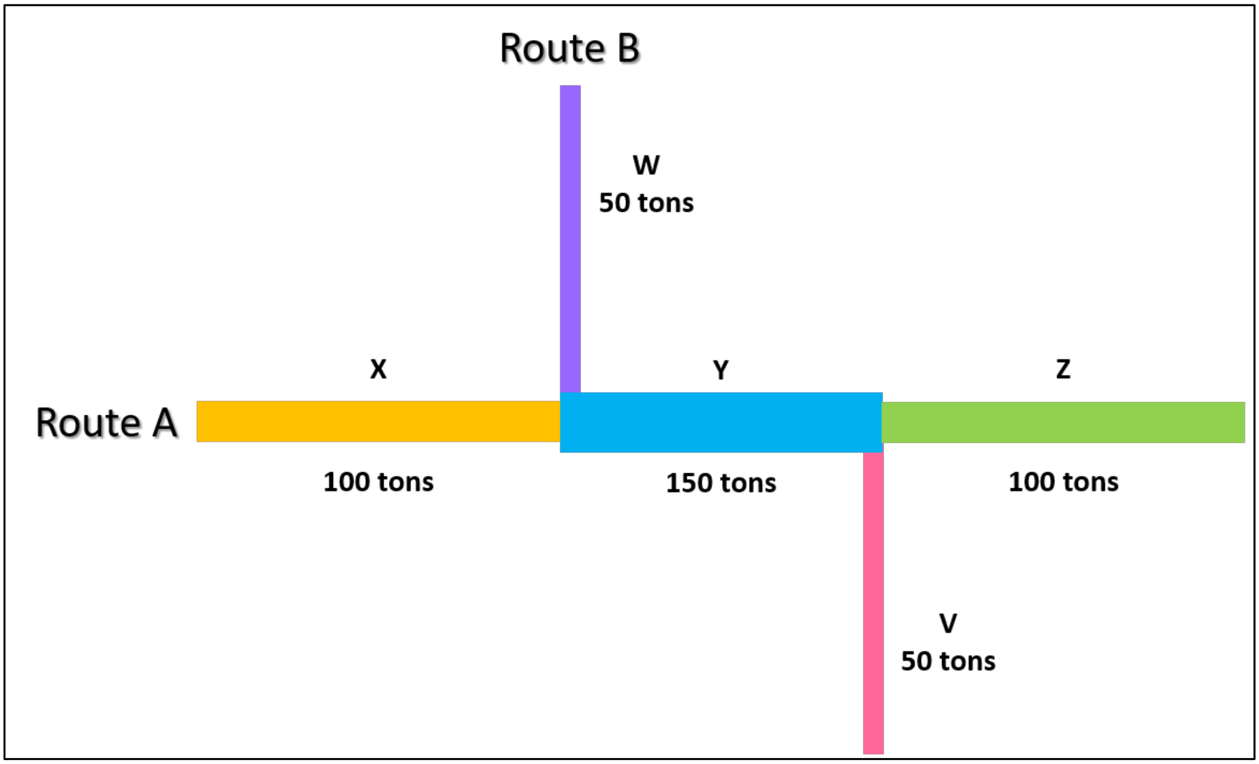

Figure 3. Total tonnage summarized by feature for Route A and Route B
Source: Volpe Center analysis of 2018 IHS Markit Transearch data

To identify HDAH, value and units were summarized by feature since the objective was to retain features of high density. The commodity flow data was summarized by route in order to compare the flows along the analysis corridors (discussed in next section) for the individual commodities and total flows without double counting, representing the actual or "real-world" flows.

2.3 Analysis Corridors

The project team identified 17 highway corridors from the HDAH for further analysis of domestic agricultural commodity flows, and corridor performance in terms of pavement and bridge conditions, congestion, reliability, and safety (discussed in Section 3).

The analysis corridors were based on the movement of the agricultural commodities across the HDAH and on availability of performance data from HPMS and NPMRDS. Because the NPMRDS performance data is only available for roadways on the National Highway System (NHS), features which are not part of the NHS were removed. The remaining features accounted for 88% of the HDAH.

From the remaining HDAH features, the project team identified the top 5% of HDAH features by volume, value, and units (again, based on cumulative sum), for each commodity type, and then compiled all of these sets of features across all commodities. This combined set was then overlaid on a single map, which identified areas that are highly dense in agricultural commodity flows but disconnected from one another. To account for breaks along the corridors, other HDAH features (not within the top 5%) were added back in manually in order to connect dense commodity flows. Corridor endpoints were chosen using natural breaks in the commodity flow data, such as major cities, freight hubs, or transfer points, and compared to highway infrastructure and relevant intermodal and processing facilities, including production values from the 2017 USDA Agricultural Census, meat processing facilities, grain elevators, ports handling agricultural tonnage, and in-land waterways.

Two additional corridors that fell just shy of the 5% threshold were also added. The first was Corridor #7, along I-95 from Florence, SC to Jacksonville, FL, which was added in order to incorporate better geographic balance. The second was Corridor #15, along State Route-99 from Stockton, CA, to Los Angeles, CA, which was added based on stakeholder engagement that indicated that this was a critical corridor for transporting agricultural freight.

3 Conditions & Performance Analysis

After creating the analysis corridors layer, the project team used the following steps to calculate and analyze corridor conditions and performance.

3.1 Network Analysis Setup

Since the analysis corridors spanned multiple states and ARNOLD/HPMS and NPMRDS data are provided at the state level, portions of corridors in different states were processed separately and later assembled. Once corridors were identified they were assigned unique numeric identifiers and clipped at state boundaries. For example, a single corridor from western Iowa to Chicago, Illinois would consist of two segments, henceforth referred to as “corridor state segments,” one spanning the length of Iowa and the second from the western border of Illinois to Chicago. Each corridor state segment was given a unique identifier consisting of the corridor ID and the state postal code abbreviation separated by an underscore. If western Iowa to Chicago was corridor 3, then the two corridor state segments would be named 3_IA and 3_IL. This example is illustrated in Figure 4. For automated processing purposes a corridor-to-corridor state segment mapping text file was created listing the corridors by ID, the origins and destinations, and the corridor state segments which make up each corridor. An excerpt of this is included in Table 2.

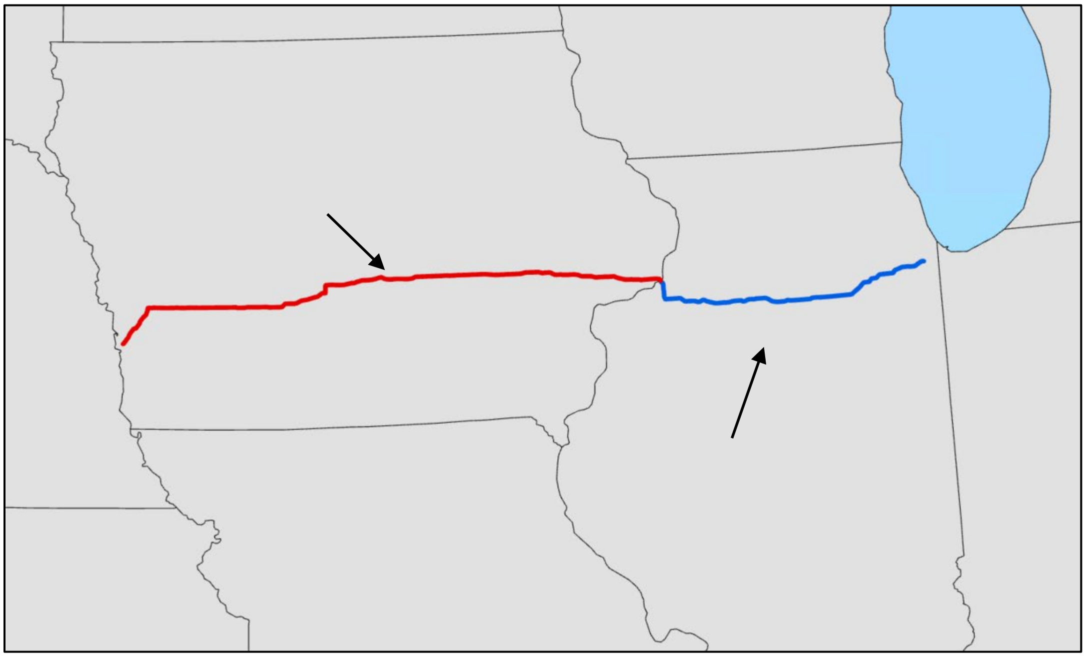

Figure 4. Analysis corridor 3 spanning Iowa and Illinois
Source: Volpe Center analysis of 2018 ARNOLD/HPMS, NPMRDS, and Transearch data

Table 2. Excerpt of corridor to corridor state segment mapping text file

Corridor	Corridor State Segments	Origin	Destination
1	1_CA	CA/OR Border	I-5/SR-99 Interchange
2	2_MS,2_AL,2_GA,2_SC,2_NC	Jackson, MS	Charlotte, NC
3	3_IA,3_IL	Omaha, NE	Chicago, IL

Source: Volpe Center

Setup and processing steps were automated with Python scripts which allowed for repeatability and scalability. Esri's ArcPy package was used extensively to automate geospatial processing steps in these scripts. The first Python script in the process performed the initial setup of the networks. A corridor ID was passed in as a parameter and the corridor-to-corridor state segment mapping text file was used to determine the corridor state segment components and performed the setup on each one. The national analysis corridors layer was subset to the corridor state segment using the corridor ID and state attribute fields. The endpoints were determined for the corridor state segment to be used as an origin-destination pair later in the process. A field was added to the points layer to indicate flow order that was manually updated after this step to indicate which end was the origin and which the destination for the network flowing tools. For the Iowa corridor state segment of the Iowa to Chicago corridor example, the point on the western end of Iowa would receive a value of 1 as the origin and the eastern end a 2 as the destination. The ARNOLD and NPMRDS layers both contain dual carriageways, one for each side of the roadway. To allow for an accurate comparison on the strip charts found in Appendix B of the main report, both directions along a corridor, origin to destination and destination to origin, were flowed in the same direction but on opposite sides of these dual carriageways. This allowed both sides of the highway to be aligned on the charts for easier comparisons.

A 500 meter buffer was then created around the ARNOLD and NPMRDS layers for the given state with the "[ArcPy Buffer]" tool (Figure 5). The layers were clipped to this buffer using the "[ArcPy Clip]" tool and projected to the USA Contiguous Albers Equal Area Conic projection that was being used for the project with the "[ArcPy Project]" tool. This subset allowed for faster processing time in subsequent steps. There were cases in which a corridor state segment was unable to flow because the network diverged from the corridor layer by greater than 500 meters and a larger buffer distance was implemented.

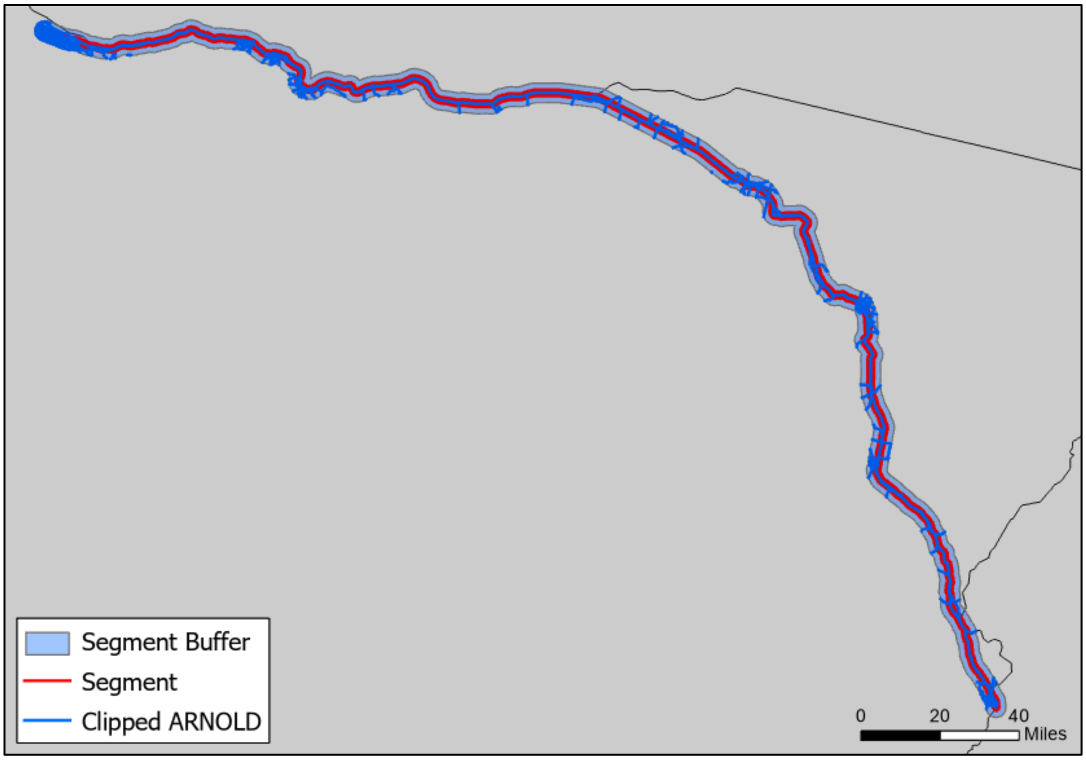

Figure 5. Example of corridor state segment with buffer used to clip ARNOLD geometries
Source: Volpe Center analysis of 2018 ARNOLD/HPMS data

Each state's ARNOLD layer contains geometries of all public roadways. Each roadway is a single polyline geometry but the only attribute included is a unique route ID. For example, the eastbound side of the Massachusetts Turnpike (I-90) is a single feature spanning from the New York border to Boston with a route ID of "I90 EB". This route ID is used in combination with the internal measure values of the geometries to locate the HPMS attributes along the roadways with a method called linear referencing.

Once the network geometries were set up, the tabular HPMS data needed to be prepared. A spreadsheet of all HPMS attributes listing their index, name, data type, and a column to indicate inclusion was created. A subset of this is shown Figure 6 in which the year, state code, route ID, HPMS record start and endpoints, AADT, and combination truck AADT attributes are designated for inclusion.

Table 3. Excerpt of spreadsheet used to select HPMS attributes to keep

Index	Name	Data Type	Include
0	YEAR_RECORD	Integer	x
1	STATE_CODE	String(2)	x
2	ROUTE_ID	String(50)	x
3	BEGIN_POINT	Float	x
4	END_POINT	Float	x
5	AADT_VN	Integer	x
6	AADT_VT	Integer	
7	AADT_VD		
8	AADT_COMBINATION_VN	Integer	x

Source: Volpe Center analysis of 2018 ARNOLD/HPMS data

An ArcGIS table of the selected HPMS attributes and start and end mileposts of all records for route IDs contained in the clipped ARNOLD layer was created (Table 4). To reduce processing time, and as functional class 7 features will be untraversed on these routes, this clipped ARNOLD layer was subset to functional classes 1 through 6. Geometries to keep were determined by creating a Python list of route IDs in the layer and iterating over the state's input HPMS file to check if any of the associated records for each route ID were of functional class 1 through 6. It should be noted that when reviewing flow results during the course of this project, there were cases of features that were required for an accurate flow but did not have HPMS data and were removed during this step. In these scenarios, overrides were added to ensure these features were retained by specifying their HPMS route ID.

Table 4. Example of a raw HPMS table in ArcGIS for attributes selected for inclusion

YEAR_ RECORD	STATE_ CODE	ROUTE_ ID	BEGIN_ POINT	END_ POINT	AADT	CRACKING_ PERCENT	F_ SYSTEM	FAULTING	IRI
2018	41	064000I00	4.6	4.66	86800	0	1	Null	45
2018	41	004000I00	185.39	185.4	4300	28	3	Null	72
2018	41	00500I00	179.2	179.3	720	0	3	Null	82
2018	41	8297	0	0.86	1280	Null	6	Null	Null
2018	41	00400I00	1.26	1.28	3700	Null	4	Null	Null
2018	41	00400I00	1.28	1.33	3700	Null	4	Null	Null
2018	41	00600I00	370.8	370.9	11600	0	1	Null	37

Source: Volpe Center analysis of 2018 ARNOLD/ HPMS data

Once subset, the "ArcPy Integrate" tool was run at a .02 meters tolerance to fix very small gaps in the ARNOLD layer. The "ArcPy Feature To Line" tool was then run to split geometries at intersections, something that is necessary because, as mentioned above, each roadway is submitted as a single polyline geometry. In that form, the network traversal tools would only be able to flow from one roadway to another at the start or end as opposed to at intersections. The clipped NPMRDS layer was also split using the "Feature To Line" tool. Start and end "mileposts" based on the geometries' internal measure values were assigned to each feature. Using the route ID, milepost values, and the HPMS records' begin and endpoints, the functional class attribute was added to the geometries through linear referencing. For features with multiple overlapping HPMS records of different functional class values, the longest value was used.

Figure 6 shows an example feature and Table 5 shows its attribute table. Based on its measure values, it runs from 32.765 to 33.92. As seen in Table 6, this route ID has 15 HPMS records that overlap those values. They are all functional class 1, resulting in this feature being assigned a functional class value of 1. Had some of the records been a different value, the total mileage of the records for each would have been calculated and the longest would have been used. The selected HPMS attributes of the associated ARNOLD geometries were stored in a Python dictionary to be accessed after the network was flowed.

At this stage, the National Bridge Inventory (NBI) dataset, initially in text file format, was imported by writing records for the current state to a comma separated text file and running the "ArcPy XY Table To Point" tool to create a point layer. This layer was also clipped with the above buffer to create a geographic subset to review later.

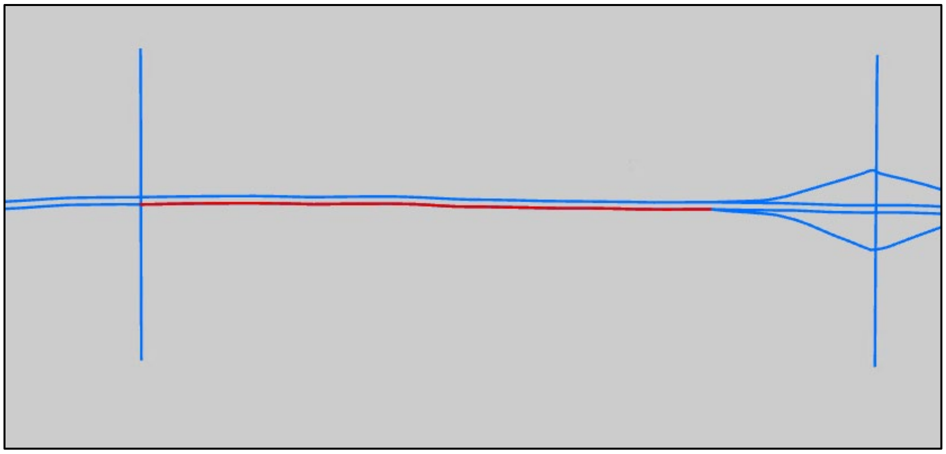

Figure 6. Example HPMS feature (shown in red) in ArcGIS
Source: Volpe Center analysis of 2018 ARNOLD

Table 5. Example HPMS feature attribute table

YEAR_RECOR	STATE_CODE	ROUTE_ID	FR_MP	TO_MP	FUNC_CLASS
2018	19	S001910080E	32.765	33.92	1

Source: Volpe Center analysis of HPMS data

Table 6. Raw HPMS data for example feature

Year_Record	State_Code	Route_ID	Begin_Point	End_Point	F_SYSTEM_VN
2018	19	S001910080E	32.7	32.766	1
2018	19	S001910080E	32.766	32.8	1
2018	19	S001910080E	32.8	32.9	1
2018	19	S001910080E	32.9	33	1
2018	19	S001910080E	33	33.1	1
2018	19	S001910080E	33.1	33.2	1
2018	19	S001910080E	33.2	33.3	1
2018	19	S001910080E	33.3	33.4	1
2018	19	S001910080E	33.4	33.5	1
2018	19	S001910080E	33.5	33.6	1
2018	19	S001910080E	33.6	33.7	1
2018	19	S001910080E	33.7	33.76	1
2018	19	S001910080E	33.76	33.8	1
2018	19	S001910080E	33.8	33.9	1
2018	19	S001910080E	33.9	33.921	1

Source: Volpe Center analysis of 2018 ARNOLD/ HPMS data

3.2 Flow ARNOLD/HPMS

With the networks prepared, the next step was to manually assign route endpoints with a 1 or 2 to indicate the start and end of the corridor respectively. ArcGIS's Network Analyst tools were then used to traverse the corridor from the origin to the destination and determine the corridor state segments in the correct order. This was an iterative process as adjustments were sometimes needed to reconcile the start and endpoints of the different networks as well as to ensure the traversed paths along the networks were accurate. To fix this, small features were added to each network connecting the start and endpoints to the line geometries as needed to ensure connectivity for traversing the corridor state segment. Adopting this solution rather than moving the points for differently segmented networks ensured each corridor state segment started and ended at the same location.

For ARNOLD/HPMS, if a corridor state segment started or ended on a dual carriageway, a feature would be manually added connecting the point to each side of the carriageway and the route ID attribute would be manually edited to "o_to_d" (origin to destination) or "d_to_o" (destination to origin) depending on direction to restrict the flow in the correct direction. Then these route IDs would be added to a text file for this corridor state segment, as discussed shortly.

The example in Figure 7 shows the start of a corridor segment and the clipped ARNOLD geometries. This corridor state segment is heading east (upper right of figure), therefore the southernmost carriageway would be the origin to destination direction and the northernmost carriageway would be the destination to origin direction. The features shown in red were added to ensure the flow begins on the correct side of the dual carriageway with the northern feature receiving a route ID of "d_to_o" and the southern "o_to_d". These manual edits represented a very small percentage of the overall mileage and were necessary to enable cross network comparisons while maintaining accuracy at the regional and national levels.

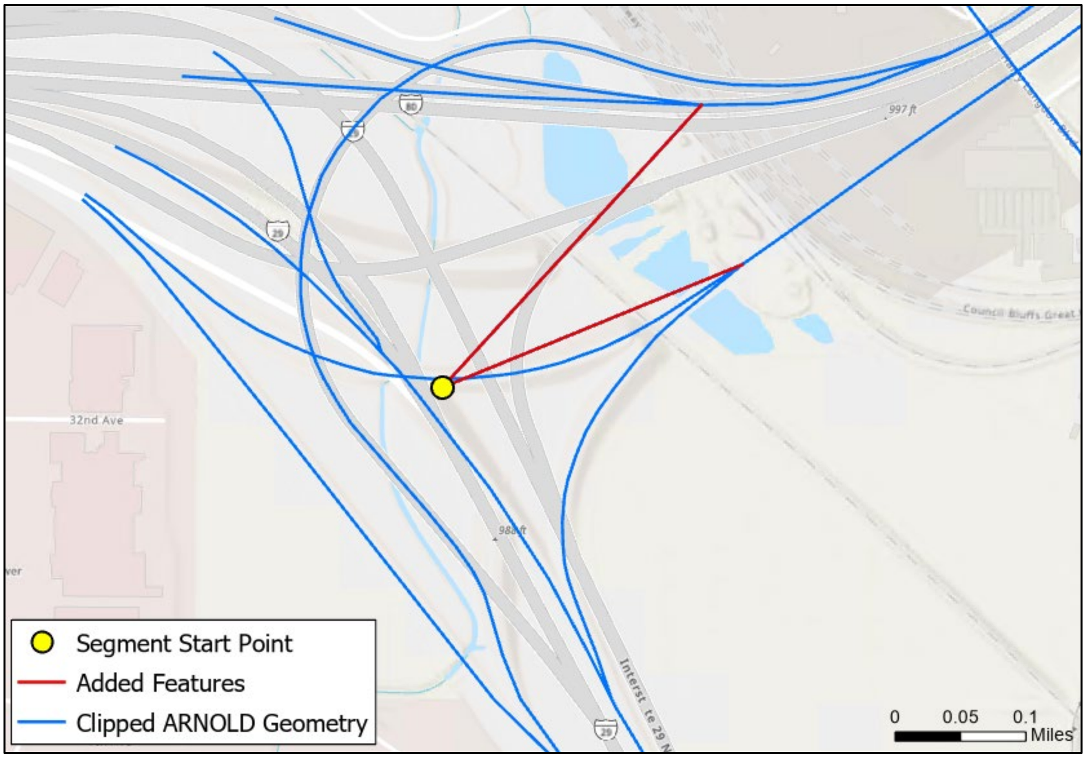

Figure 7: Example of features added at start of corridor state segment to ensure correct ARNOLD/HPMS flow
Source: Volpe Center analysis of 2018 ARNOLD/ HPMS data

Impedance values for flowing on ARNOLD/HPMS were calculated using the values in the Table 7 to favor higher functional class roads during the routing process. The functional class “Other” refers to features that lacked HPMS attribution.

Table 7. Impedance Values for Flowing ARNOLD/HPMS

Functional Class	Impedance
1	Length * 0.50
2	Length * 0.60
3	Length * 0.70
4	Length * 0.80
5	Length * 0.90
Other	Length * 2.00

Source: Volpe Center analysis of 2018 ARNOLD/ HPMS data

Each route flowed along the ARNOLD/HPMS network was reviewed to ensure that each flow direction remained on the correct side of dual carriageway roads. If, for a given direction, the flow traveled on the incorrect side of a dual carriageway, the route ID of the correct feature to be flowed was added to a text file for the corridor state segment. Table 8 shows an example of this text file. The route IDs shown in

Figure 8 will have their lengths multiplied by .01 for the given directional flow, resulting in them having a lower impedance value for that corridor state segment and direction to help correct the flow. Figure 8 shows an example of this incorrect flowing on an eastbound route where it switched from route ID STRUIR00080C to STRUIR00080N. After these adjustments were made, the corridor state segment was reflowed and checked for accuracy once again until the flow was correct for both directions.

Table 8: Feature route IDs in a text file to have their impedance values reduced

DIRECTION	ROUTE IDs
o_to_d	o_to_d,STRUIR00080C,SMASHIR00080KC,SSUMIR00080KC
d_to_o	d_to_o,STRUIR00080N

Source: Volpe Center analysis of 2018 ARNOLD/HPMS data

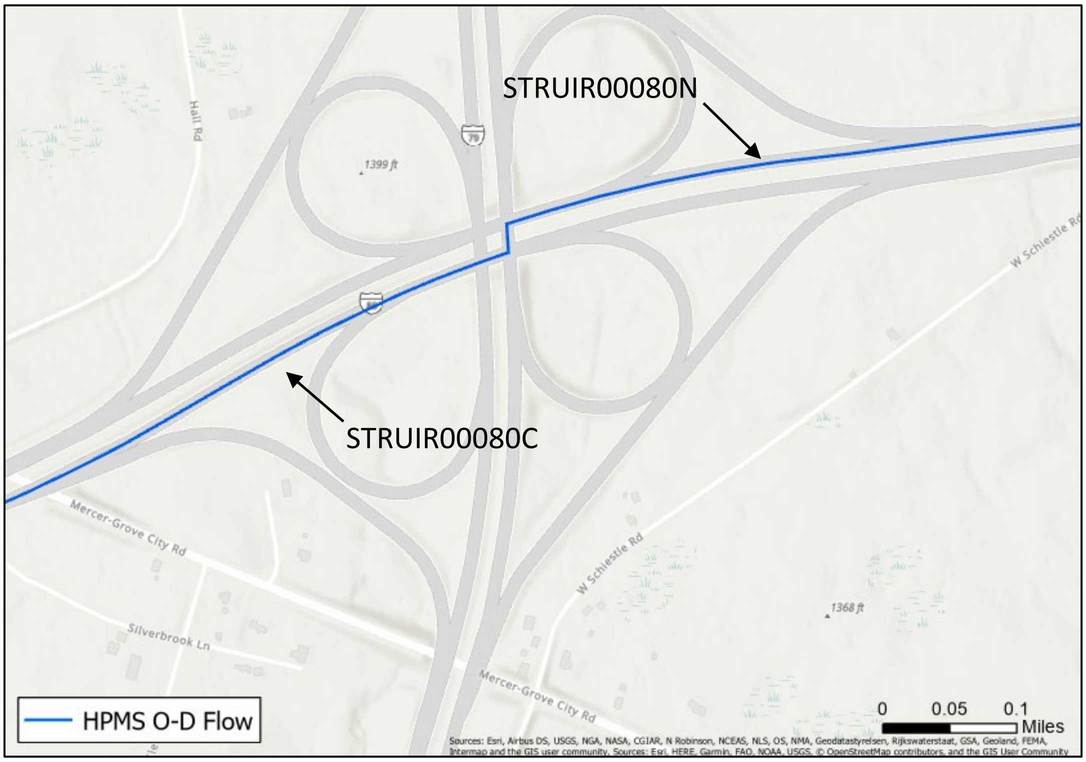

Figure 8. Example of incorrect HPMS flow on a dual carriageway
Source: Volpe Center analysis of 2018 ARNOLD/ HPMS data

Once a corridor state segment was flowed, each traversed feature was given start and end values based on their length and position along the route. For example, a road feature one mile long starting 7.8 miles into the corridor state segment would have a start value of 7.8 and end value of 8.8. The table of HPMS attributes for each feature was accessed to write the attribution to a text file alongside each feature's assigned start and end value to be used to chart attribution along each corridor later in the process. An example excerpt is shown in Table 9.

Table 9. Example HPMS output text file

START	END	THROUGH_LANES	AADT	PAVEMENT_CONDITION	FUNC_CLASS	NHS	URBAN_RURAL
0.017	0.117	4	31518	F	1	1	U
0.117	0.187	4	31518	F	1	1	U
0.187	0.217	4	31518	G	1	1	U
0.217	0.287	4	31518	F	1	1	U
0.287	0.317	4	26531	F	1	1	U

Source: Volpe Center analysis of 2018 ARNOLD/HPMS data

3.3 Flow NPMRDS

The next script performed the flowing of NPMRDS data. The impedances for this network were calculated based on the direction of the network geometry. Unlike with HPMS data, the directionality of the NPMRDS features for dual carriageways correspond to the direction of the flow of traffic. Therefore, in the origin to destination direction, dual carriageway features in the from-to direction were favored and vice-versa for destination to origin. Figure 9 shows the directionality of the features flowed at the start of the example corridor state segment depicted in Figure 7.

When adding the geometries to connect to the start point, it was important to match the correct directionality to each direction. The origin to destination feature was drawn from the point to the line feature, whereas the destination to origin feature was drawn from the line feature to the point. Similar to HPMS, once the corridor state segment was flowed, each feature was assigned start and end values based on their location along the corridor state segment. The attribution of the traversed Traffic Message Channels (TMCs) alongside feature start and end values were output to a text file to chart later in the process.

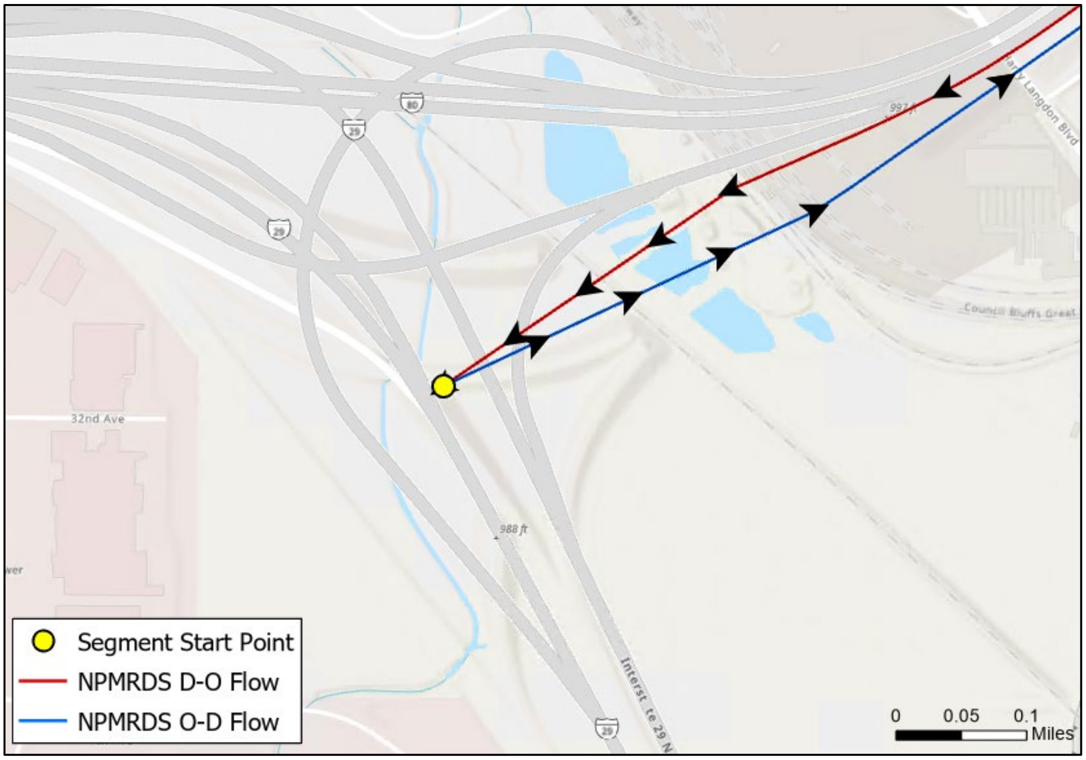

Figure 9. NPMRDS flow with feature directionality
Source: Volpe Center analysis of 2018 NPMRDS data

3.4 Flow Transearch

The final network to be flowed was Transearch. The Transearch layer for each corridor had two copies of each geometry, one for each direction with the associated commodity flow information. This is not the original format of the data but is an artifact of how the data was processed, as detailed in Section 2.

Similar to the previous scripts, a corridor ID was passed in as a parameter and using the input corridor text file. The corridor state segments were obtained based on corridor ID and state, and run one at a time. Origin and destination fields were used to select the correct features for each direction. Since the Transearch network was used to define the corridors it was not necessary to add features at the start and end of the corridor state segments to connect to the points unless they were particularly far away from the ARNOLD and NPMRDS geometries and were moved to accommodate this. As with HPMS and NPMRDS, feature start and end values were added based on their location along the corridor state segment and these values and their associated commodity flow information were written to a text file for each direction to be used for charting later in the process.

3.5 Snap NBI & FARS

After the corridor state segments had been determined on each of the networks, the NBI and FARS datasets were processed. In the first script, a layer of bridges within 500 meters of each corridor state segment had been created. A manual review of this layer was performed to identify bridges that would be driven on while traversing the corridor (i.e., ignoring bridges passing over the highway) based on the route number and record type attributes in the NBI. An attribute was added to this layer and populated with a 1 for the bridges of interest to allow the code to select these bridges and snap them to the corridor to determine their mileage along the corridor for charting.

Similarly, the FARS data was imported for the state of the current corridor state segment, clipped to the 500 meter buffer, and snapped to the nearest road feature within 75 meters, if one existed. The location information for NBI and FARS was written to a text file for each dataset for charting, with the NBI file also including bridge condition attribution.

3.6 Calibration

Due to real variations in divided highways (e.g., when the two different directions are temporarily separated due to terrain) and differences in the Geographic Information System (GIS) layers as exemplified in Figure 10, cumulative route distances vary by direction (e.g., southbound versus northbound) and across networks (HPMS vs NPMRDS or Transearch.) For this reason, it was necessary to go through a process to calibrate the GIS geometries so that they can be accurately compared to each other.

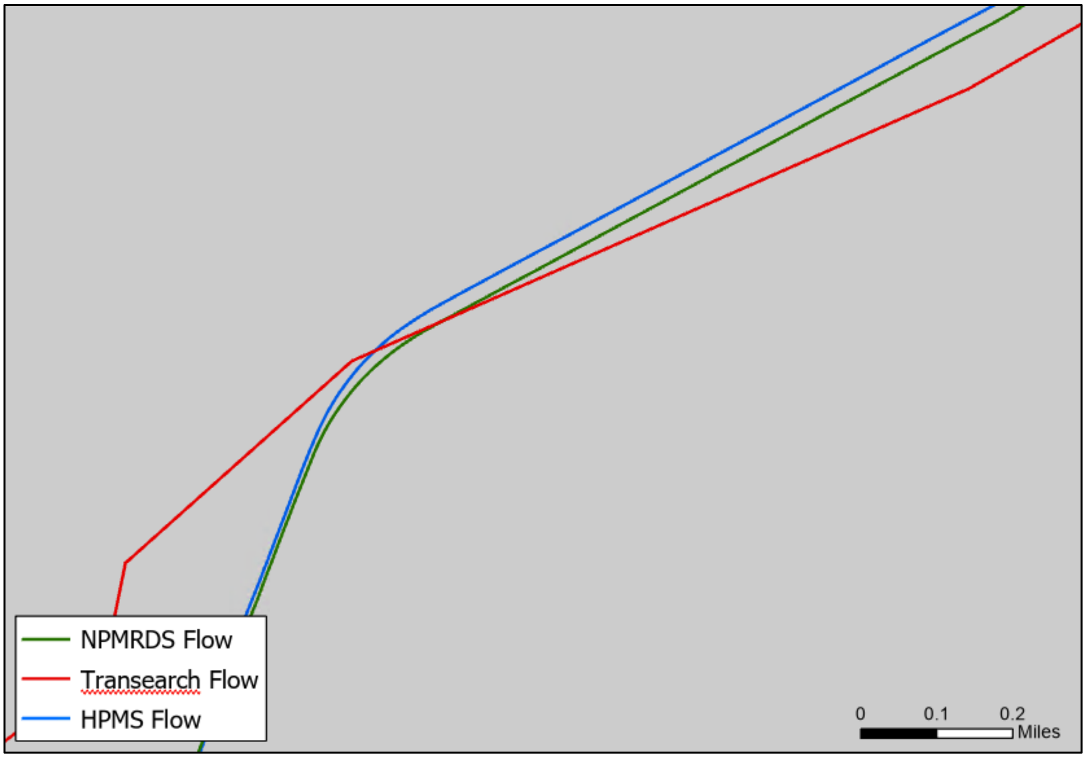

Figure 10. Example of differences in network geometries
Source: Volpe Center analysis of 2018 ARNOLD/HPMS, NPMRDS, and Transearch data

For example, a section of roadway may span from corridor mile 124.1 to 124.3 in HPMS but from corridor mile 125.6 to 125.8 in NPMRDS because of geometry differences in the layers. When developing charts from the different network sources, it was important that the features were as aligned as possible for an accurate comparison.

To accomplish this, the HPMS origin to destination direction was used as the baseline to which the other networks and directions were calibrated to. Vertices for each feature were generated using the "ArcPy Feature Vertices to Points" tool. The position of each vertex along the feature which it belongs to was determined and using this information and the start and end mileposts of the feature, the milepost of each vertex was assigned. For example, a vertex halfway along a feature with start and end mileposts of 15.7 and 16.7 would have a milepost value of 16.2. This process was repeated for each of the three networks in both directions with the exception of our baseline of HPMS origin to destination. Once this step was complete, for each of the network and direction combinations, the nearest HPMS origin to destination direction vertex within 350 feet was found using the "ArcPy Generate Near Table" tool. For each network and direction that was compared to the HPMS origin to destination baseline, a calibration text file was created that listed the corridor milepost of the vertex and the equivalent HPMS origin to destination corridor milepost.

After the calibration file was created, the flow result text file for the network and direction created earlier was accessed to read in the start and end mileposts and attribution of each traversed feature in the corridor state segment. For each feature, the start milepost was compared to the values in the calibration file. An NPMRDS feature that started at corridor milepost 0.180 would be calibrated to .177 based on the example in Table 10. This was calculated by first identifying which two values in the file a given milepost is between. In this case .180 is between .175 and .194. The milepost to be calibrated, .180 here, is divided by the lower milepost, .175 and the result is multiplied by the lower HPMS milepost, .172. This yields .177 which is the new calibrated starting value for this feature. If the original value was .175 it would be assigned a new value of .172. This is repeated for the end value of each feature as well and the new calibrated start and end mileposts and associated attribution for the feature is written to a new text file. This process enables a more accurate comparison of attribution on different geometric networks at a regional and national scale.

Table 10. Example records from NPMRDS O-D calibration file

HPMS_O_TO_D	NPMRDS_O_TO_D_EQUIV
0.172	0.175
0.195	0.194
0.229	0.222

Source: Volpe Center analysis of 2018 ARNOLD/HPMS and NPMRDS data

3.7 Generate Final Products

Until this point outputs were created at the corridor state segment level and the calibrated text files with mileposts and attribution and traversed layers started at zero for each corridor state segment. For each corridor, the segment files were merged with cumulative milepost values to allow for a multi-state corridor-level analyses.

Once all datasets were calibrated to the same corridor length and corridor state segments were merged, they could be charted on the same X-axis showing location along the corridor. The Python library Matplotlib was used to create the corridor-level strip charts found in Appendix B of the main report. This was done by reading each corridor's post-calibration text file containing the full corridor milepost information and attribution. The attributes plotted on line graphs – AADT, urban/rural designation, TTTR, TTI, and commodity data – were visualized by plotting coordinate pairs of each record's attribute value and start point and attribute value and endpoint and then connecting the points. Pavement condition was shown with three horizontal broken bar charts – one for each condition rating – with the same Y-value and passing in the start value and length of each HPMS record as parameters. The point layers – bridges and FARS crashes involving trucks – used consistent Y-values and plotted their standardized corridor milepost values along the X-axis.

Figure 11 shows an example of HPMS, NPMRDS, FARS, and NBI attribution charts for a corridor from Appendix B.

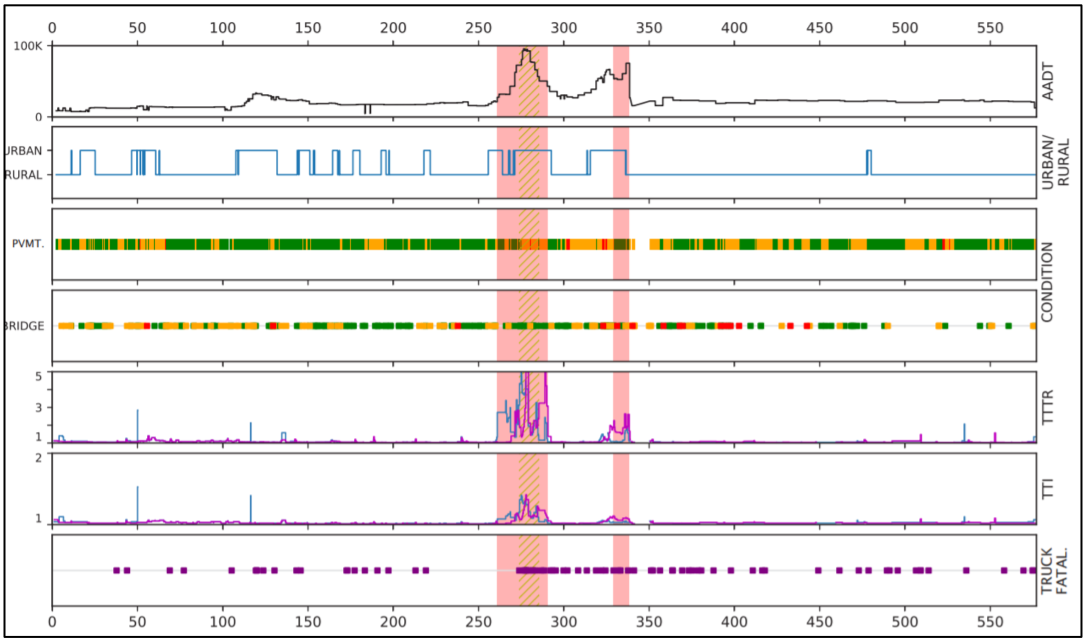

Figure 11. Example of HPMS, NPMRDS, FARS, and NBI attribution charts
Source: Volpe Center analysis of 2018 ARNOLD/ HPMS, NPMRDS, FARS, and NBI data

4 Conclusion

The approach detailed above enabled the analysis of commodity flows and corridor conditions and performance using attribution from an array of data sources. GIS tools and Python programming allowed for the identification of corridors of interest and the processing and comparison of multiple networks with differing geometries in a largely automated workflow. In addition to the preparation and processing of the data, a novel approach for attribute visualization was adopted. Beyond the results that this effort yielded, this process serves as a framework for how these tools and technologies may be employed for similar projects using the same or other datasets in the future.

The Importance of Highways to U.S. Agriculture - Methodology Whitepaper
http://dx.doi.org/10.9752/TS330.06-2021