

Multi-Modal Freight Transportation System Capacity and Diversion Assessment for Middle Tennessee

**Project #: RES2016-29
Final Report**

**Submitted to
Tennessee Department of Transportation**

Janey Camp, Paul Johnson, Jordan Williams, Hiba Baroud
Vanderbilt Center for Transportation and Operational
Resiliency (VECTOR)

July 1, 2019

Technical Report Documentation Page

1. Report No. RES2016-29	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Multi-Modal Freight Transportation System Capacity and Diversion Assessment		5. Report Date May 2018	
7. Author(s) Janey Camp, Paul Johnson, Jordan Williams, Hiba Baroud		6. Performing Organization Code	
9. Performing Organization Name and Address Vanderbilt Center for Transportation and Operational Resiliency, Vanderbilt University, PMB 351831, 2301 Vanderbilt Place, Nashville, TN 37235		8. Performing Organization Report No. TDOT PROJECT # RES2016-29	
12. Sponsoring Agency Name and Address Long Range Planning Division, Tennessee Department of Transportation, James K. Polk Building, Suite 900, 505 Deaderick Street, Nashville, TN 37243-0344		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. RES2016-29	
		13. Type of Report and Period Covered Final Report 08/2016-05/2018	
		14. Sponsoring Agency Code	
15. Supplementary Notes			
16. Abstract The objective of this study was to perform a data-driven, comprehensive analysis of multi-modal freight system capacity to understand the capacity of the system overall and be able to simulate and analyze impacts of disruptions to the system. To accomplish this, a multi-modal network of highway and rail networks in the middle Tennessee region was created in ArcGIS and then ran through a range of disruptive scenarios for analysis. Ultimately, simulation results were intuitive with real-life expectations. Without any disruption, network metrics aligned with empirical FAF and Waybill data. During disruptions, rail freight will often prefer switching to the road network, but vice-versa does not hold. For a set of disruption scenarios, disruption was most severe when a minimally redundant section of highway was impacted. Lesser impacts occurred during rail disruptions and in areas where multiple alternate routes existed. Overall, this project provides proof of concept for an approach to simulate and predict regional impacts to freight flow within Middle Tennessee's multimodal network in the event of disruptions using a unique, multi-faceted approach that combines ArcGIS Network Analyst simulations coupled with machine learning techniques. This approach provides a method for identifying critical areas for planned response and an understanding of the impacts of disruption on a combined intermodal (highway and rail) network in a very active freight region.			
17. Key Words Freight, capacity, rail, highway, diversion			18. Distribution Statement
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages	22. Price NA

DISCLAIMER

This research was funded through the State Research and Planning (SPR) Program by the Tennessee Department of Transportation and the Federal Highway Administration under RES2016-29, Multi-Modal Freight Transportation System Capacity and Diversion Assessment.

This document is disseminated under the sponsorship of the Tennessee Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Tennessee and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the views of the author(s) who are solely responsible for the facts and accuracy of the material presented. The contents do not necessarily reflect the official views of the Tennessee Department of Transportation or the United States Department of Transportation.

TABLE OF CONTENTS

TECHNICAL REPORT DOCUMENTATION PAGE	II
LIST OF TABLES	V
LIST OF FIGURES.....	V
EXECUTIVE SUMMARY.....	1
1. INTRODUCTION.....	3
2. LITERATURE REVIEW.....	4
3. METHODOLOGY	6
3.1. Data Preparation and Mapping the Networks	6
3.2. Use of ArcGIS Network Analysis Tool to Calculate Capacity and Other Metrics	9
3.3. Use of ArcGIS Network Analysis Tool to Simulate Disruptions.....	10
3.4. Statistical Learning to Predict Future Disruption Impacts	13
3.5. Modeling Specific Disruption Scenarios.....	13
4. RESULTS.....	15
4.1. <i>Testing the Multi-Modal Connection Approach</i>	15
4.2. Impacts of Specific Disruptive Scenarios	17
5. CONCLUSIONS AND RECOMMENDATIONS.....	19
6. REFERENCES	22
APPENDIX 1: DATA PROCESSING.....	25
APPENDIX 2: NETWORK SIMULATION.....	29
APPENDIX 3: STATISTICAL MODELING USING R SOFTWARE.....	32

LIST OF TABLES

Table 1: OD routing options during disruption.....	11
Table 2: Results from modeled simulations	18

LIST OF FIGURES

Figure 1: Middle Tennessee area under consideration.....	8
Figure 2: US rail tonnage passing through Tennessee.	9
Figure 3: Map showing areas for specific disruption scenarios.	14
Figure 4: Atlanta multi-modal network as pilot test of methodology	16
Figure 5: Simulation Example.....	17
Figure 6: Comparison of travel time distributions across scenarios.....	18
Figure 7: Statistical model results for disruption scenarios.	19

EXECUTIVE SUMMARY

The objective of this study was to perform a data-driven, comprehensive analysis of multi-modal freight system capacity (highway and rail) in the middle Tennessee region to understand the system's capability to: 1) serve the anticipated demand for freight (and possibly passenger) traffic and 2) be able to accommodate additional commodity flow in the event that diversion is needed. To accomplish this, a multi-modal network of highway and rail networks in the middle Tennessee region was created in ArcGIS and then subjected to a range of disruptive scenarios for analysis. FAF and Waybill data were utilized to estimate parameters to build the Network Analyst model simulations and develop metrics such as volume and capacity for OD pairs throughout the system.

Ultimately, simulation results were intuitive. Without any disruption, results aligned with those of the empirical data in FAF and Waybill. During disruptions, rail freight will sometimes prefer switching to the road network, but vice-versa does not hold. This notion makes sense when considering the structure of each network within the Middle Tennessee area. Tennessee's road network is highly redundant while the rail network consists of only a few rail lines with few alternative routes. This finding limits the prospect of adding capacity for passenger transport to the rail lines. In fact, for most cases, it is more efficient for freight that is already traveling on road within the Middle Tennessee area to remain on road.

The main takeaway from this analysis is that highly complex, interdependent factors involved with transportation network disruptions can be modeled effectively with simulation and machine learning techniques. Impacts to the system scale intuitively with a disruption's severity and range of impact but are also highly dependent upon where the event takes place along the network. This insight can also help identify critical areas for Middle Tennessee's intermodal network.

Middle Tennessee's redundant road network offers alternative route options that are just not available on rail, at least at this scale. For a set of four disruptions simulated, disruption was most severe when a minimally redundant section of highway was impacted. Lesser impacts occurred during rail disruptions and in areas where multiple alternate routes existed.

Overall, this analysis provides proof of concept for an approach to simulate and predict regional impacts to freight flow within Middle Tennessee's multimodal network in the event of disruptions.

1. INTRODUCTION

Tennessee's economic vitality relies heavily on the timely and efficient flow of commodities supported by a multi-modal freight transportation system. There is growing concern regarding the ability of such an interconnected system to withstand disruptive events and maintain the necessary level of commodity flow to support a growing economy and mitigate potential cascading impacts. Disruptive events such as the middle Tennessee flooding in May of 2010, which shut down portions of both Interstates 24 and 40 for extended periods, as well as the Atlanta I-85 bridge collapse in 2017 only highlight the interconnectedness and importance of such networks (OEM 2011, Straight 2010). Ideally, multi-modal transportation systems include redundancy in the network to ensure effective diversion and minimal impacts to supply chains and freight flows in the event of a disruption to one or more modes.

Added complexity is introduced when additional, sometimes competing, interests and demands are being considered such as a commuter rail service that would utilize portions of the freight rail network. A case in point is the mobility challenge facing middle Tennessee, where the passenger rail option is likely constrained by the near capacity existing rail lines.

The Tennessee Department of Transportation (TDOT) recognized the need to better understand such a complex system's capabilities and limitations both under current freight flows and during disruptions. To assist TDOT in achieving this goal, this project focused on developing and applying an appropriate methodology to estimate capacity and capabilities of the middle Tennessee rail and highway interconnected system for a range of operating conditions.

The middle Tennessee region serves as a logical candidate for developing and testing the methodology to address these issues due to the area serving as a key east-west and north-south corridor, housing the third-largest switching hub in the CSX network while experiencing considerable population and business growth. This region provided an adequately complicated and important case study application with anticipation that the methodology utilized could be replicated or expanded to other areas of the state and beyond.

The research activities included the following primary tasks: (1) obtain and prepare necessary data for analysis, (2) determine capacity and identify "hot spots", (3) develop and model scenarios for disruption. In the following sections, the methodology employed and findings from the analysis are presented.

2. LITERATURE REVIEW

A review of the literature was performed to identify the best approaches for defining and quantifying capacity for the study.

In general, the capacity of a transportation network can be defined as the maximum loading that can be achieved on that system in terms of demand supplied (i.e., train cars or loaded train cars, commuters, etc. supplied by track or road space) or the maximum throughput of demand from origin to destination over a given time while maintaining an acceptable level of service. Alternatively, capacity can also be thought of as (1) the amount of a commodity that can be transported past a single point of measure in one direction in a given time or (2) the value of some controlling parameter at which unacceptable delays (threshold delays) begin to manifest and propagate (Pouryousef, et al. 2015). It is commonly accepted that there are two types of capacity, or measures of capacity, and those are theoretical capacity, and practical capacity. The first is an upper bound; the latter is a more reliable measure. There is a possible third measure, which is reserve capacity, which constrains traffic assignment to the existing traffic pattern ratios and can be taken as a lower bound.

Another consideration in evaluating capacity and potential shifts of traffic or freight during periods of disruption is flexibility. Flexibility (in the context of a transportation network) can be thought of as the amount of change the network (roads, bridges, rail, etc.) can handle or absorb and still perform within intended parameters. Change can be in terms of system degradation, demand volume, demand pattern and demand type. In other words, flexibility is the range of different states a system can occupy, link loading, trip production volumes, link accessibility, so forth, while still providing a viable transportation network. Reliability is an important concept to incorporate when discussing flexibility. Allan and Billington (1992) state reliability to be the “probability of a device [or system] performing its purpose adequately for the period of time intended under operating conditions encountered”. As the capacity available to fill demand is the primary measure of a transportation network’s performance, flexibility of a transportation network is often translated to capacity flexibility.

System capacity and flexibility are important measures for planning, design and administration purposes and consequently for expanding the capabilities of both freight and passenger transportation networks which often overlap and utilize the same resources. Meanwhile, demand for freight transportation on rail networks has risen 64% domestically in the last three decades (Morlok and Chang 2004). Along with freight needs, there is increased interest in commuter rail to relieve congestion and pollution from automobiles on highways. Over a

decade ago, railway experts came to the conclusion that flexibility is of greater importance than speed for assuring that transportation demands are met reliably (Morlok and Chang 2004).

Standard approaches for measuring the capacity of a transportation network include the following:

1. The bi-level optimization algorithm proposed in "Modeling the Capacity and Level of Service of Urban Transportation Networks", which utilizes an adapted ETDA-VDC model to explicitly consider level of service and route choice while distributing trip units over a network of links and nodes to satisfy origin destination demands and other physical constraints (Yang et. al 2000). The result of this algorithm is the maximum number of trips that can be accommodated by the system, (i.e., its capacity)
2. Penalty ADDVOL methods proposed in "Measuring Capacity Flexibility of Transportation System," which also use bi-level algorithms to load links on a transportation network to produce a lower bound and upper bound capacity respectively (Morlok and Chang 2004).
3. Network Capacity with Limited Flexibility algorithm proposed in "Modeling Capacity Flexibility of Transportation Networks" which allows for route choice for all trips assigned in addition to the base, existing traffic pattern and produces a practical capacity (Chen and Kisikitwiwat 2010).

Ideally, all of these methods can be used given limited information about the transportation network of focus. A final method that can be used, given sufficient information, is a simple but involved string plot method where link characteristics are examined along with physical principles to determine maximum theoretical capacity along said link, then load the network accordingly. Trips can be assigned in this manner using TransCAD or other software (AAR 2007). This method constitutes a rough estimate, but can be an efficient approach. Note that according to the Association of American Railroads (AAR) National Rail Study (2007), practical capacity is roughly equal to the theoretical capacity scaled by a factor of 0.7. Ultimately, the best method for determining capacity will be the one that both planners and railroad representatives can agree on and that best reproduces real world existing conditions.

The differences that arise between highway and rail transportation largely stem from some limiting factors. For instance, it is more difficult to accurately control for jerk and stopping distance for highway analysis as there are many individuals deciding what parameters are acceptable for themselves. Additionally, whereas it is practical to assume that for every trip demand there is an

automobile available (at least, road supply will likely run out, before auto supply), train car supply and station access are far more limited and require careful consideration. Furthermore, train cars can be removed from or added to the network at multiple stations sited at discrete locations, autos are assigned to trip producing and receiving zones that generalize input and output and node interactions such as free flow speed and congestion on ramps, time spent parking, and possible flow restrictions due to capacity conflicts. Most of these physical constraints can be interpreted in algorithms, or be accounted for via careful consideration of data and application of adjustment factors.

For the purposes of this study, capacity of the networks is calculated utilizing ArcGIS software for the roadways and rail corridors passing through the Middle Tennessee region and primarily Interstate and state highways in the region that are most capable of accommodating freight movements.

3. METHODOLOGY

The methodology applied for this project consists of integrating GIS, simulation, and machine learning techniques to estimate and simulate freight movement and impacts that disruptions have on multimodal transportation networks. Data on transportation disruptions are scarce, and their resulting impacts are largely uncertain. The chosen approach attempts to address these issues. Simulation techniques allow for the generation of many different scenarios and statistical learning techniques help establish bounds of uncertainty on potential impacts.

First, GIS data relevant to the road and rail networks in Middle Tennessee were assimilated. The Middle Tennessee area is the focus of this project, but these techniques are scalable to other regions of the United States. Second, network disruptions for a range of scenarios are simulated via ArcGIS' Network Analysis Tool. Lastly, statistical models are trained on these simulated outcomes to allow for intuitive predictions of impacts from future scenarios.

3.1. Data Preparation and Mapping the Networks

In order to construct simulations for Middle Tennessee's multimodal transportation network, assimilating data for US road and rail networks was necessary. While inland waterways are also part of Tennessee's multimodal system, that form of transportation is currently out of scope for this project but may be considered in future research.

The Highway Network

For the highway network, data for use in this project was obtained from the US Bureau of Transportation Statistics' Freight Analysis Framework (FAF). The key files

included a roadway network shapefile, CVS files with corresponding traffic flow metrics, and a shapefile outlining the FAF zones. The network shapefile provides a georeferenced collection of highways and roads across the US, with identifying characteristics defined at various intervals specified by FAF such as name, speed limit, miles, number of lanes, and others (Ford 2017). The traffic flow metrics consist of items such as freight tonnage, traffic volume, and peak capacity for each road segment defined by FAF. These metrics are mapped to the accompanying road segment in the network shapefile. The FAF zone file was then used to isolate the middle Tennessee region of interest.

Figure 1 below depicts an overview of the assimilated road data, and Appendix 1 contains details regarding the data processing. The Middle Tennessee FAF region was utilized as the focus region for this project (shown as a light yellow region in Figure 1). An additional 50km buffer was created around the region (i.e., the light green region in Figure 1) and used for network considerations to allow for more realistic routing scenarios within the region (i.e., a truck may reasonably travel along a road that is outside Middle Tennessee even if its origin and destination lie within the region). As expected and indicated in the map in Figure 1, the interstate highways carry the largest volume of freight traffic in the region.

The Rail Network

For the rail network, data is amalgamated from several sources. First, private Waybill data containing origin-destination (OD) rail records was provided by the US Department of Transportation (USDOT) via TDOT. A public sample exists for general users, but the private data is necessary for in-depth analysis. The Waybill data contains rail shipping information such as carloads, tons shipped, and commodity type identifiers, but since each record corresponds to an individual shipment, records need to be aggregated for each leg of the network in order to look at overall traffic flows. US DOT's Freight Routing Tool was leveraged for this task (Wright and Baker 2017). Data processing steps are provided in Appendix 1, but essentially, the output consists of a shapefile of aggregated Waybill metrics as queried by the user. Figure 2 below shows an example output, depicting all freight tonnage that pass through the state of Tennessee.

Shapefiles of the US rail network and rail stations, the latter of which is used later to create the intermodal connections were also obtained from the US DOT. Lastly, Oak Ridge National Laboratory provides information on rail capacity, which was utilized for this project (Peterson 2014). The sources of information were all joined into one rail network. Appendix 1 contains details on creating this network.

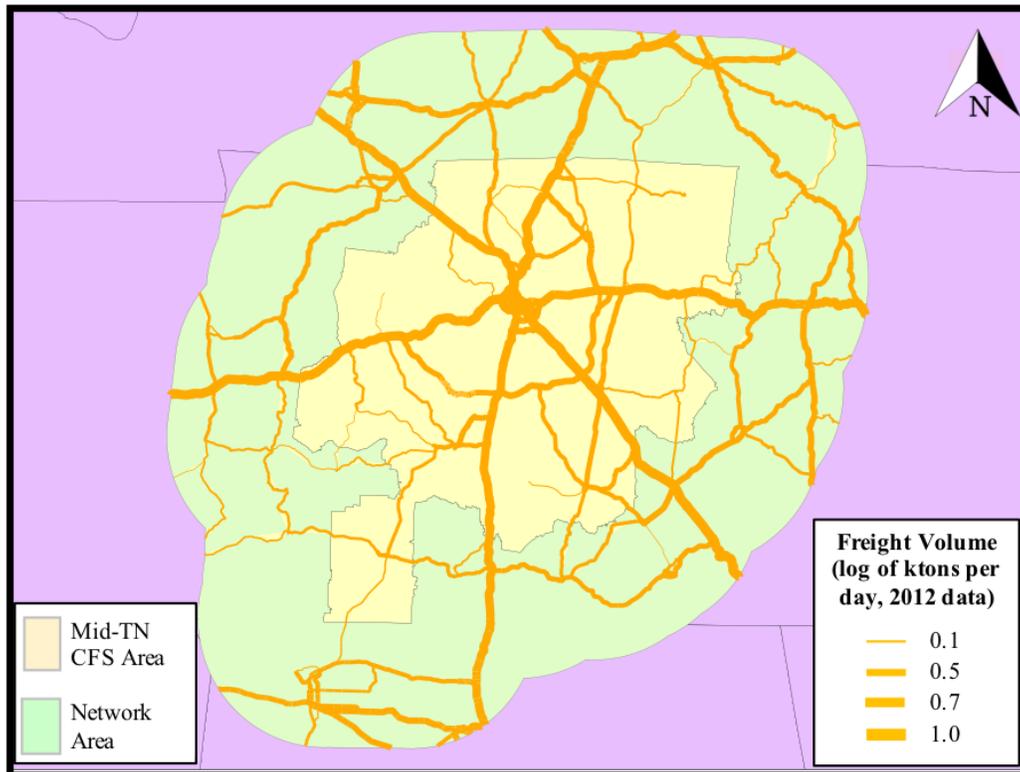


Figure 1: Middle Tennessee area under consideration including the commodity flow service (CFS) area, the roadway network area considered and roadways used in the analysis sorted based upon freight volume.

Developing a Multi-Modal Network

The road and rail networks were then interconnected within ArcGIS into a single multimodal network. Compatible metrics for average ton-miles per hour (volume) and allowable ton-miles per hour (capacity) were derived for each transportation mode. The two networks were then connected topologically via the rail station junctions for use in simulations considering shifts between modes. Appendix 1 contains the exact procedures for creating the multimodal network.

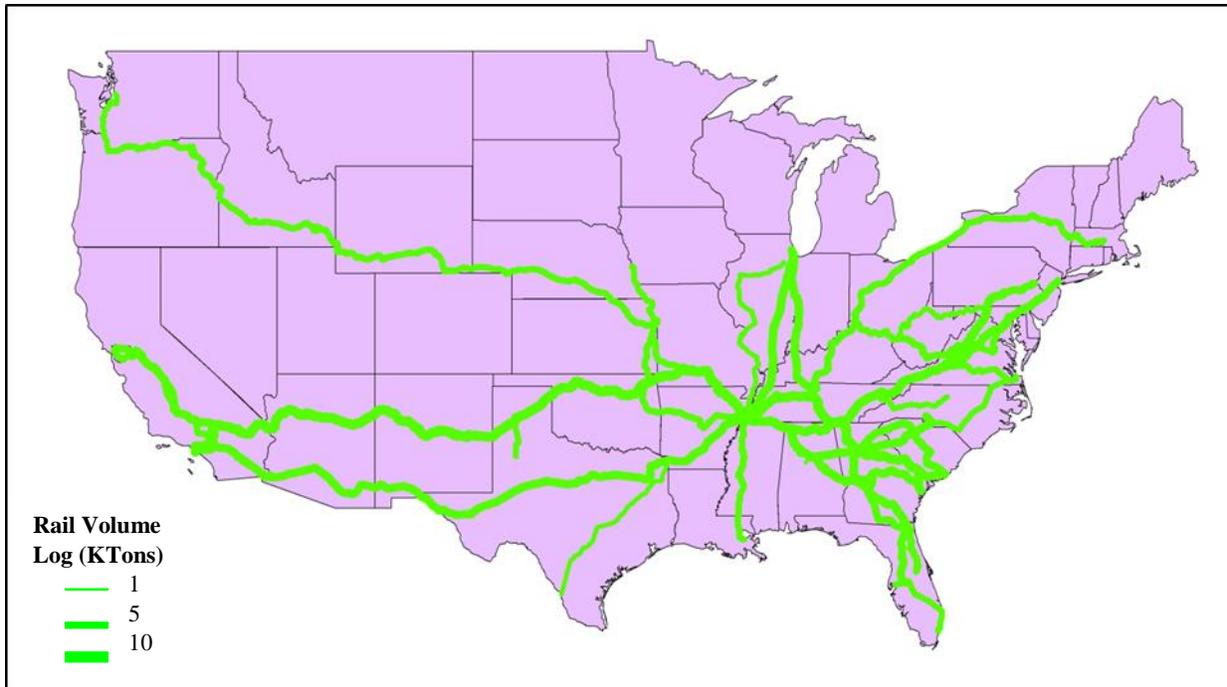


Figure 2: US rail tonnage passing through Tennessee.

3.2. Use of ArcGIS Network Analysis Tool to Calculate Capacity and Other Metrics

Using ArcGIS' Network Analysis Tool, the multimodal network was then leveraged to estimate capacity and volume across the entire region as well as simulate impacts from disruption scenarios. Utilizing methods from existing literature as a premise for our approach (Asakura 1999; Matisziw et al. 2007; Kurauchi et al. 2009), the process utilized in this study included the following tasks:

- OD pairs were constructed along the multimodal network
- Routes between each OD pair were simulated based on fastest travel time
- Volume and capacity measurements were accumulated for each route for a baseline (no disruption) condition
- Disruptions of varying size, severity, and location were simulated, thereby affecting the route choice between implicated OD pairs
- The impacts to each OD pair's accumulated metrics were determined under disruption scenarios
- The aggregated OD pair impacts served as a system-wide assessment of the disruption's effect compared to other disruptions and the baseline condition.

Route choice is determined by fastest travel time. For each part of the network, travel time, measured in hours, is estimated by multiplying the average speed

limit of the segment, determined by FAF metrics for roadways and track class for railways, by the length of that segment. Transitions between roadways and railways are allowed at rail stations, but a time penalty of one hour is incurred for an intermodal transfer. This time penalty is based on empirical studies of intermodal transfer times (Verma and Verter 2010).

It is important to note that routes are modeled independently of other simulated routes. This assumption is not realistic, as everyday traffic jams inconveniently illustrate how travel time depends on other vehicles' route choice. However, ArcGIS is currently not capable of handling interdependent routing. Instead, impacts from routing interdependencies are calculated in a separate manner, discussed later.

The OD pairs are divided into two disparate groups, one for each mode of transport. The roadway ODs consist of all the road segments defined in the FAF shapefiles, and the railway ODs consist of the rail stations provided by US DOT. For each disruption simulation, routes for all OD pairs within each group are calculated (i.e. road ODs to road ODs are calculated and rail ODs to rail ODs are calculated). Intermodal transport is allowed, but road ODs to rail ODs are not explicitly modeled because that choice doesn't reflect real life circumstances (i.e. a train station origin would not likely have a destination in the middle of a highway). There are a total of $370^2=136,900$ road OD pairs and $232^2=53,824$ railway OD pairs for the area of middle Tennessee under consideration, amounting to 190,724 total routes calculated for each simulation.

For each OD route, accumulated metrics for volume, capacity, travel distance, and travel time are measured. The aggregated comparison of volume to capacity metrics over all OD pairs serves as an overall indicator for the capacity constraints of the entire system. Metrics derived when no disruption is present provide a base case for comparison to disruptions. Differences in travel distance infer if a reroute takes place in the event of a disruption. Differences in travel time are used for several calculations, discussed later. The derivations for all these measurements are identified in Appendix 2.

3.3. Use of ArcGIS Network Analysis Tool to Simulate Disruptions

To test the robustness of the multi-modal network under disruption, disruptions of randomly varying size, severity, and location are simulated within the Middle Tennessee FAF region. The size of the event ranges from 1 to 65km, as measured by the radius of a circle. In actuality, disruptions will likely be asymmetric and/or not resemble a circle, but the chosen approach enables efficient simulations and accessible interpretability without detracting from main inferences. The severity of the event is modeled via a scaled time cost according to how long it would take a vehicle to pass through the impacted area: 2 (twice as long), 3, 4, and 99 (a complete closure). For simplification purposes, the event is assumed to

have an equivalent severity impact on roadways and railways, such as what would likely happen during natural disasters. This model assumption could be changed at a later time if there is a desire and empirical basis to explore impacts of mode-specific scenarios. The centroid of each disruption is constrained to the Middle Tennessee region, but the range of impact can extend outside this area depending on the radius of impact. In total, 99 different random scenarios were simulated.

If a simulated disruption falls onto a route, that route will be affected in one of three ways: the route will remain the same but take a longer time to complete, a reroute along the same mode of transportation will occur, a reroute involving a different mode of transportation will occur. The outcome is based on whatever minimizes travel time. For example, a reroute from the rail network to road network occurs. The depicted OD pair's original route is a direct rail route. However, the disruption simulated with a severity of 99 closes the original rail route, and an intermodal option becomes the most viable alternative.

All affected routes' accumulated volume and capacity measurements will also change accordingly. The ensuing calculations and assumptions are meant to address the aforementioned lack of interdependent routing capabilities within ArcGIS (i.e. impacts to traffic volumes are manually derived instead of explicitly simulated). Table 1 below provides an overview for the three cases.

Table 1: OD routing options during disruption.

	Road Volume	Road Capacity	Rail Volume	Rail Capacity
Same Route	$\text{Volume}_{\text{old}} * \frac{\text{Time}_{\text{new}}}{\text{Time}_{\text{old}}}$	$\text{Capacity}_{\text{old}}$	$\text{Volume}_{\text{old}} * \frac{\text{Time}_{\text{new}}}{\text{Time}_{\text{old}}}$	$\text{Capacity}_{\text{old}}$
Reroute – No Intermodal	$\text{Volume}_{\text{new}} + 0.5 * \text{Volume}_{\text{old}}$	$\text{Capacity}_{\text{new}}$	$\text{Volume}_{\text{new}} + 0.5 * \text{Volume}_{\text{old}}$	$\text{Capacity}_{\text{new}}$
Reroute – Rail to Road	$\text{IMVolume}_{\text{old}}$	0	$\text{Volume}_{\text{old}} - \text{IMVolume}_{\text{old}}$	$\text{Capacity}_{\text{old}}$
Reroute – Road to Rail	$\text{Volume}_{\text{old}} - \text{IMVolume}_{\text{new}}$	$\text{Capacity}_{\text{old}}$	$\text{IMVolume}_{\text{new}}$	0

For a given OD pair that is affected by the disruption, if the route follows the same path, then its baseline accumulated volume is multiplied by the ratio of

the disrupted travel time to the baseline travel time. Capacity remains the same. This calculation assumes that routes contained within the disruption will incur a volume to capacity strain proportional to that of the relative increase in travel time (i.e., double the travel time implies doubling the volume to capacity ratio).

If the simulation determines a reroute is the best if not the only viable alternative for a given OD pair, then it must be determined if an intermodal transfer takes place. If the rerouting takes place exclusively on the same mode of transportation, then the impacted volume equals the disruption simulated volume (that accumulated in the alternative route) plus one-half that of the baseline path. Capacity is equal to the new route's accumulated capacity. These assumptions are meant to model the real-life situations when traffic gets diverted to alternative routes. The new, alternative routes will be subject to their original traffic volumes plus additional volume from diverted OD trips. The one-half comes from a simplifying assumption that, on average, half of the traffic from the disrupted route would be transferred to the new route. In actuality, this percentage can range from 0 to 100 for each route segment, and ideally a more detailed measure could be obtained. The challenge with that granular approach is that there is not a one to one mapping between OD pairs and network segments, as one OD route comprises one or many segments. The segment-specific information is lost during the simulations. It is possible to indirectly acquire this information via visual inspection of each new OD route but becomes impractical when looking at 190,724 routes for 99 different scenarios.

Lastly, if an intermodal reroute occurs, the volumes and capacities of each mode of transport need to be taken into account. In simulations where rail reroutes to road, the new road volume is equivalent to the intermodal tonnage available in the rail's baseline route. It is assumed that only intermodal cargo can be transferred and that all of it will. The new road route's capacity is set to zero, so that the aggregated impacts reflect additional volume placed on the roads without also changing road capacity. The rail volume becomes whatever the baseline rail volume was minus the baseline intermodal cargo. Rail capacity remains the same. The assumption that only intermodal cargo is transferred serves as an upper bound for transfers to reflect real-life settings because in the simulation, if intermodal routes are found preferable, all of the cargo tends to be transferred. Additionally, the amount of intermodal cargo relative to the total cargo is small, so the assumption that all of it gets moved in the event of a transfer does not skew aggregated results.

For road to rail reroutes, the same concepts are applied. However, because the intermodal measurements are derived from the rail network and not the road network, due to intermodal data being available in Waybill but not FAF, the equations are slightly altered. The implementation of all these calculations is shown in Appendix 3.

A system-wide impact of the disruption is then estimated by aggregating its impacts on all OD pairs. Ultimately, each of the 99 disruptions contains an aggregated measurement for volume to capacity, and this measurement serves as the dependent variable for statistical modeling.

3.4. Statistical Learning to Predict Future Disruption Impacts

Statistical models are trained on the simulation data in order to predict the impact of a future disruption based on its severity, range, and location. The dependent variable (DV) is the system-wide volume to capacity measure, which can be compared to the baseline scenario to assess a disruption's impact.

Explanatory factors include 68 predictors: radius of disruption (km), severity of impact (scaled cost of time), latitude, longitude, and one binary variable for each of the 64 grid points. The grid boxes overlaid on top of the Nashville region are used as binary variables within the statistical models (i.e., if grid point 32 is impacted by scenario X, then for that model run the variable is coded as a 1). Although the exact latitude-longitude coordinates of the disruption are also included in the prediction models as explanatory variables, the grid points allow for a direct location-based network parameter without relying on an interaction of three variables (latitude, longitude, and radius of disruption). Impacts are heavily dependent on the network's structure, and grid-based approaches can help identify critical locations (Jenelius and Mattsson 2012; Kim, Chen, and Linderman 2015).

Several statistical models and machine learning techniques were attempted to minimize prediction errors. Of the 99 simulations, 70 were included in the training process and the remainder served as the testing sample.

Models attempted to include multiple linear regressions, lasso regressions, principal component regressions, general additive models, random forests, and boosted decision trees. It should be noted that all of the models were fit to the natural log of the DVs in order to normalize the spread of the residuals. The analyses were first conducted on the untransformed DVs, but the magnitude of residual errors systemically increased with larger impacts. This bias violated underlying assumptions of several of the modeling techniques. Appendix 3 contains details on the machine learning process employed here.

3.5. Modeling Specific Disruption Scenarios

Following creation of the multi-modal network and model testing under a range of random disruptions, specific disruptive scenarios were developed for testing. The scenarios were selected (1) to represent different types of disruption and direct modal impacts, (2) to comprise a range of disruption types, and (3)

aimed at distributing the impacts across the middle Tennessee region. Four specific scenarios were simulated as listed below and shown in Figure 3. The scenarios consisted of the following:

1. A highway flooding shutdown of Interstate 40 in the vicinity of the Harpeth River, West of Nashville (impact radius of 3.57 km).
2. A CSX rail line hazmat incident near the Tennessee State Capitol and Bicentennial Mall in downtown Nashville near the intersection of Rosa Parks Boulevard (impact radius of 0.875 km).
3. A bridge failure disruption at the I-40 and I-65 overpass in South West Nashville (impact radius of 0.187 km).
4. A manmade, targeted rail disruption at Radnor Yard which is located south of Nashville near I-65 and Harding Place (impact radius of 0.436 km).

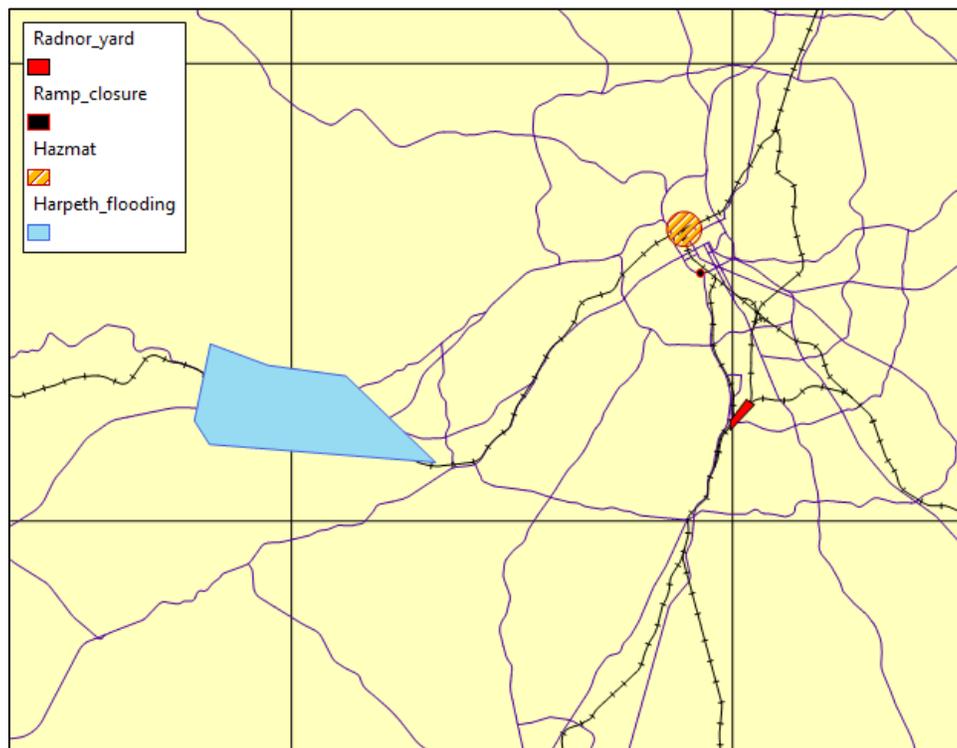


Figure 3: Map showing areas for specific disruption scenarios.

For the disruption scenarios, the regions of the map were “blocked off” in the ArcGIS Network Analyst as described above and then routing performed across the network for all OD pairs. For each simulation the travel time, travel distance, volume, and capacity for each OD route was determined for analysis and compared to the base case with no disruptions.

4. RESULTS

The results from the multi-modal network creation, simulation, and disruption modeling are discussed in the following sections.

4.1. Testing the Multi-Modal Connection Approach

The viability of using a multi-modal network within ArcGIS to simulate interconnectivities across modes and during times of disruption was a unique challenge in this project. Therefore, the project team chose to first explore and test the approach with a recent, actual event that resulted in shifts across modes although the shift was in passenger movement as opposed to freight movement. The test case application chosen was a model and simulation of the I-85 highway bridge collapse in Atlanta, GA, in 2017. Following the I-85 bridge collapse, a reported 25% increase in ridership was observed by the Metropolitan Atlanta Rapid Transit Authority (MARTA) as many commuters participated in a multi-modal shift from car to train for about six weeks (Khan 2017). Traffic flows in the Atlanta region were used to test the multi-modal connectivity approach and its use with the ArcGIS Network Analyst Tool before it was utilized for modeling freight movements in middle Tennessee.

In the Atlanta application, the downtown “Five Points” area was considered the destination with all traffic originating at county centers in the surrounding area. Consecutive links that serviced more than two routes were then singled out as key corridors. Key links that serviced diverted traffic in the secondary shortest paths were singled out as critical. The links of the four MARTA lines were added to the layer by digitizing from GoogleMaps™, where MARTA stations with parking served as transfer points between highway and MARTA lines. In the test model application, all trips originating from outside the Atlanta area destined to downtown shifted to MARTA rail lines in the modeled disruption simulation. From network demand patterns produced by first order skim-tree analysis, it was determined that the rail lines present the most time efficient mode for transportation from within the I-285 loop to the city center, when only reasonable speed estimates are considered. In short, we were able to successfully shift movements in a disruption simulation across modes. Figure 4 shows the modeled multi-modal network for Atlanta.

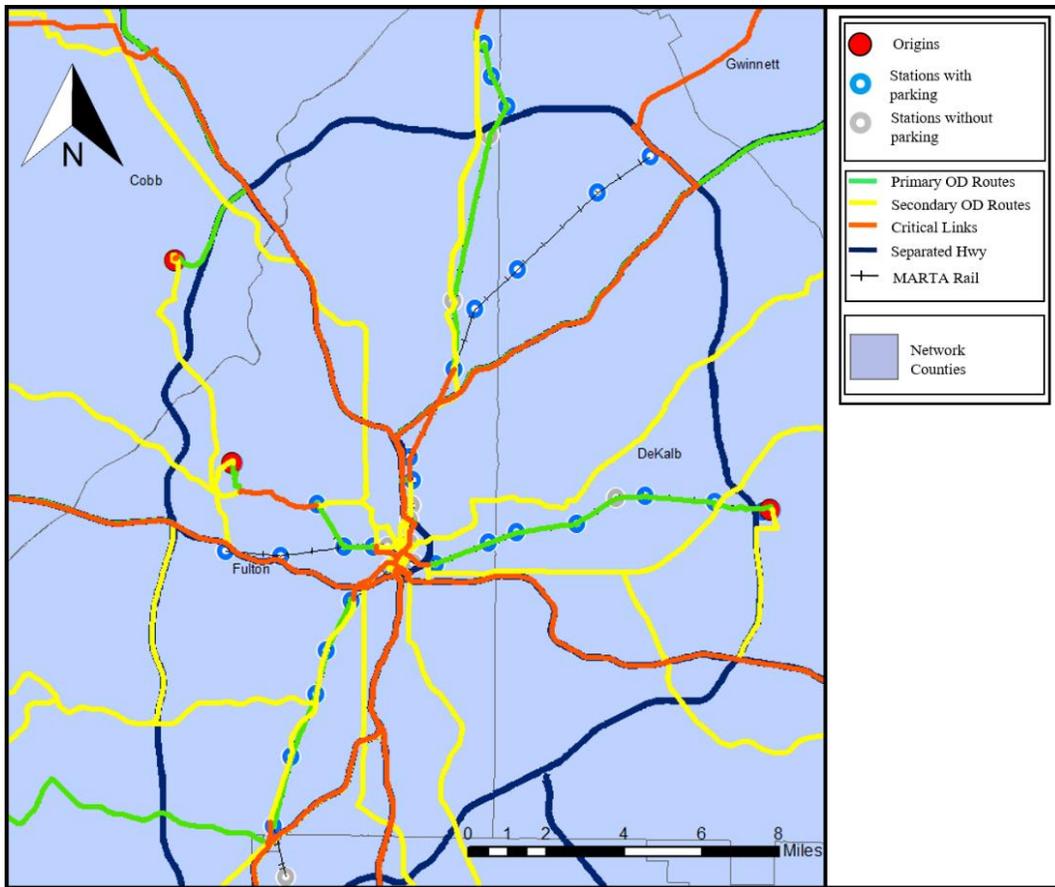


Figure 4: Atlanta multi-modal network as pilot test of methodology

Figure 5 provides an illustration of the simulation process for middle Tennessee. In this, the gray lines comprise the multimodal network itself. The blue and orange dots entail road and rail OD pairs respectively. The rail OD pairs also correspond to rail station locations, allowing for connections between the roadways and railways. The large red circle depicts an example disruption area. The weighted purple line represents the preferred route for an example OD pair, with labels “1” and “2” marking the start and end points respectively. Lastly, the grid boxes are used by statistical models to help predict impacts.

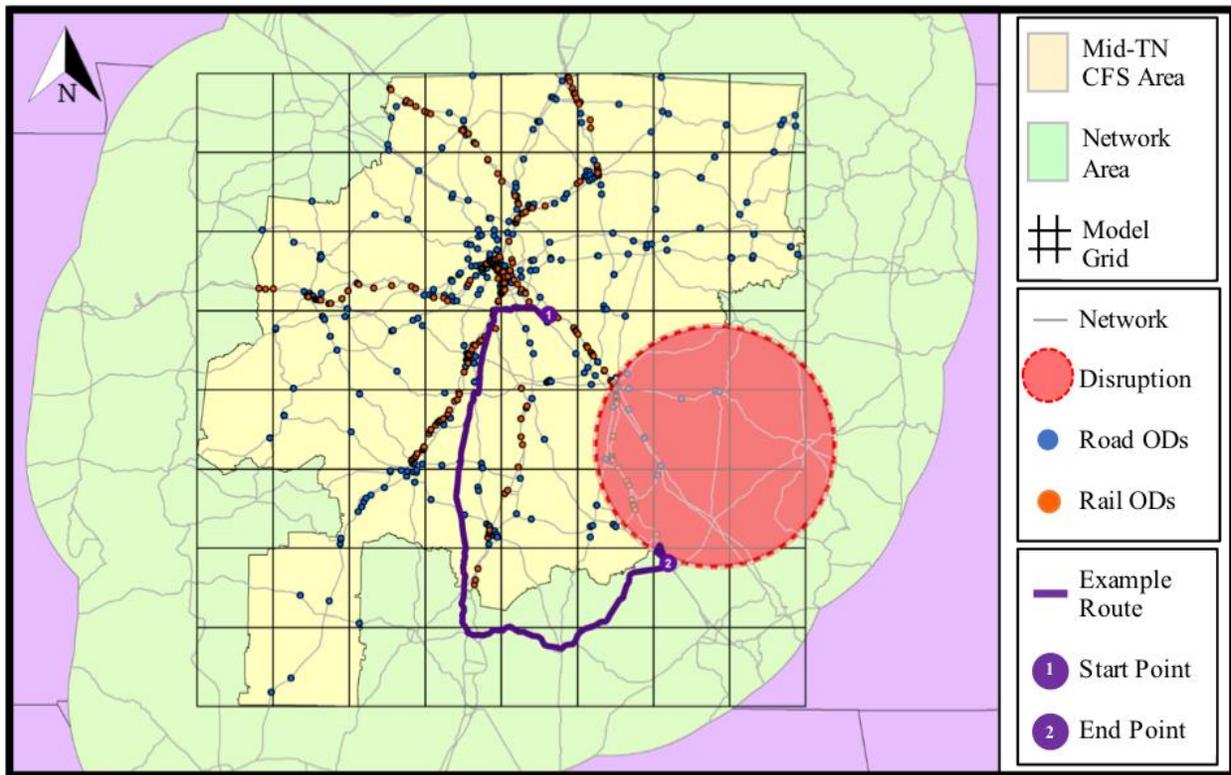


Figure 5: Simulation Example: (light-yellow) Middle Tennessee FAF region, (light-green) multimodal network extent, (gray lines) multimodal network, (large red circle) example disruption, (blue points) road ODs, (orange) rail ODs/stations, (purple line) example route for one OD pair given the disruption depicted.

4.2. Impacts of Specific Disruptive Scenarios

For the baseline and the four simulated disruptive scenarios of interest, a comparison of the network analyst outputs for each scenario was performed. The metrics under consideration (travel time, distance traveled, volume, and capacity) were all averaged across the OD pairs and are presented in Table 2. In comparison to the base condition (according to the Network Analyst simulation) all disruptive scenarios increased the average travel time and distance traveled with the exception of the disruption of the I-65 and I-40 which only increased minimally and is not reflected in the table. The greatest disruption occurred with the flooding of I-40 in the Bellevue area west of Downtown Nashville. Figure 6 shows the distribution of travel times across the OD pairs for the scenarios and the base situation. The lack of redundancy in major highways in that region is a key contributor to the impact of disruption on I-40 in that scenario. Capacity also decreased most under that scenario.

Table 2: Results from modeled simulations

Scenario	Average Travel Time (hours)	Average Distance Traveled (Miles)	Average Volume (tonnage/hour)	Average Capacity (tonnage/hour)
Base (no disruptions)	0.84	45.58	100,505	89,638
Flooding of Harpeth River Impacting I-40 West of Nashville	1.01	45.89	109,884	79,998
Hazmat Incident on Rail near Downtown Nashville	0.92	45.69	107,867	89,925
Disruption at Radnor Rail Yard	0.89	45.79	101,636	90,080
Overpass Collapse/Failure at the Intersection of I-40 and I-65 in Southwest Nashville	0.84	45.67	114,647	88,814

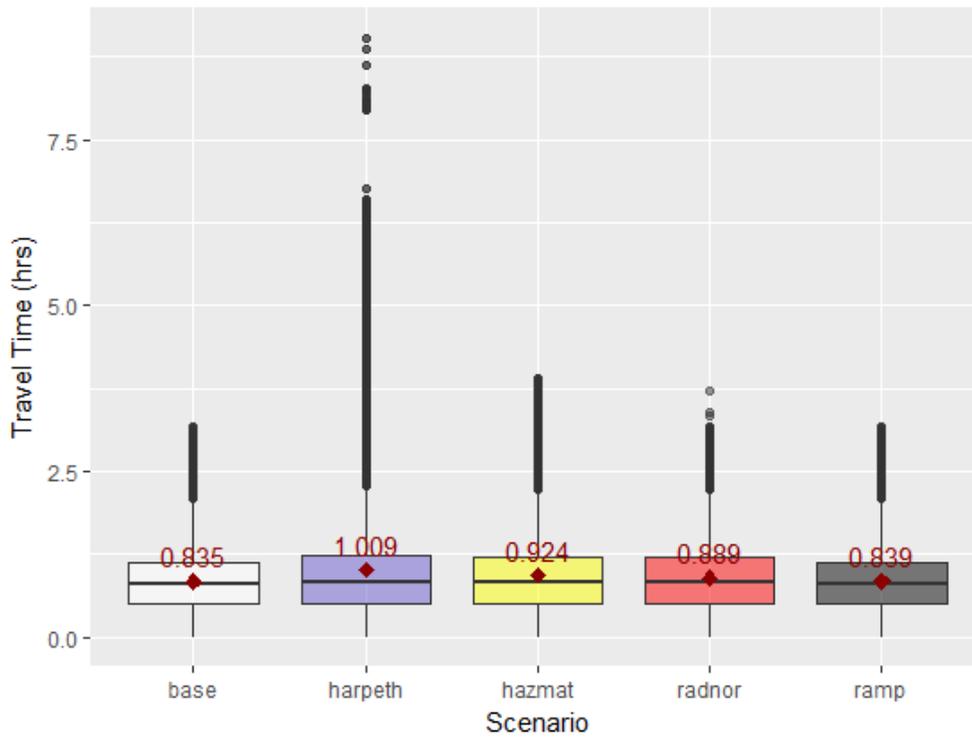


Figure 6: Comparison of travel time distributions across scenarios.

When applying the statistical learning model to the four disruptive scenarios, the predictive model most closely replicated the disruptions of the Harpeth River flooding on I-40 and the I-65 and the I-40 intersection closure (i.e., “ramp” in the results outputs). Both of these were highway-centric disruptions. Figure 7 shows the ArcGIS Network Analyst results (black dots) for each disruption scenario as well as the model estimated values and 95% model confidence interval bounds in red.

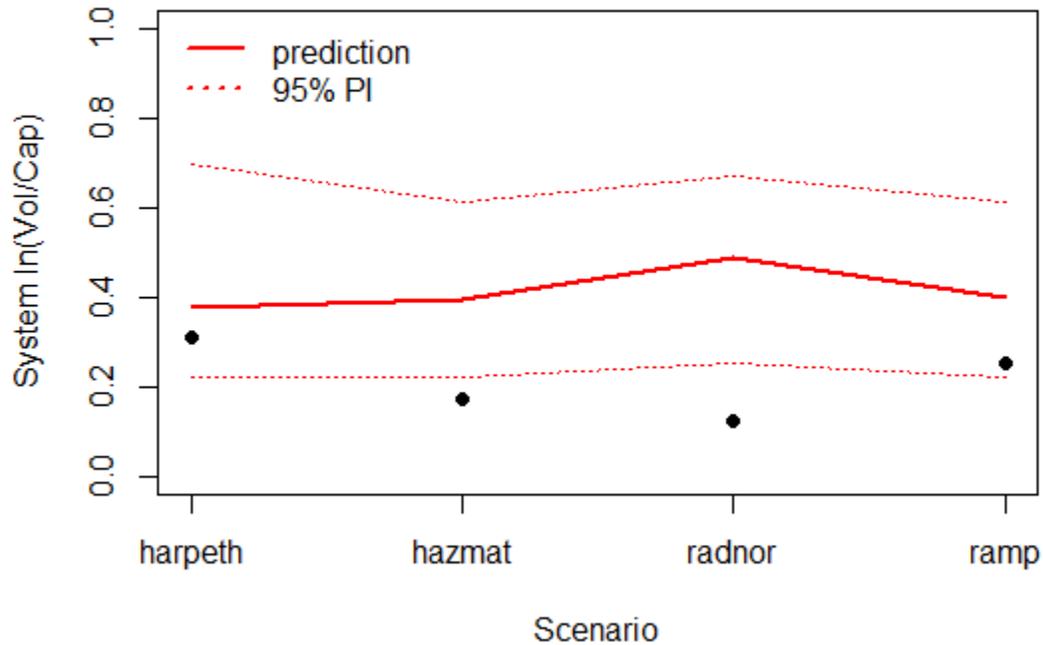


Figure 7: Statistical model results for disruption scenarios.

This high level of agreement between the statistical model utilizing machine learning and the Network Analyst results demonstrates that using a combined, multi-faceted approach in this way could prove beneficial to model additional scenarios for disruption to identify “hot spots” or overall system impacts for a range of disruptive events within one or across multiple modes.

5. CONCLUSIONS AND RECOMMENDATIONS

The objective of this study was to perform a data-driven, comprehensive analysis of multi-modal freight system capacity (highway and rail) in the middle Tennessee region to understand the system’s capability to: 1) serve the anticipated demand for freight (and possibly passenger) traffic and 2) be able to accommodate additional commodity flow in the event that diversion is needed. To accomplish this, a multi-modal network of highway and rail networks in the middle Tennessee region was created in ArcGIS and then ran through a range of disruptive scenarios for analysis. FAF and Waybill data were utilized to estimate parameters to build the Network Analyst model simulations and develop metrics such as volume and capacity for OD pairs throughout the system.

Ultimately, simulation results were intuitive. Without any disruption, results aligned with those of the empirical data in FAF and Waybill. Granular route checks also showed that chosen alternatives in the event of a disruption did indeed follow the shortest possible travel time.

An interesting finding is that when looking at intermodal preferences during disruptions, rail freight will often prefer switching to the road network, but vice-versa does not hold. This notion makes sense when considering the structure of each network within the Middle Tennessee area. Tennessee's road network is highly redundant, offering many alternative pathways, as seen in Figure 1. However, at this same local scale, its rail network consists of only a handful of rail lines with very few alternatives. This finding curtails the idea of using excess space on passenger rail to transport freight. In almost all cases, it is more efficient for freight that is already traveling on road within the Middle Tennessee area to remain on road. In fact, it appears that there is little flexibility and capacity available in the rail network to support freight or passenger diversion, while the highway network has the flexibility to support freight diversion from rail as well as from disrupted highway links.

The main takeaway from this analysis is that highly complex, interdependent factors involved with transportation network disruptions can be modeled effectively with simulation and machine learning techniques. Impacts scale intuitively with a disruption's severity and range of impact but are also highly dependent upon where the event takes place along the network. Grid-based modeling approaches can help identify critical areas, and for Middle Tennessee's intermodal network, northwest Nashville is one of these spots.

A second major takeaway from this analysis is that the intermodal preference in the wake of a disruption is one-sided. Middle Tennessee's redundant road network offers alternative route options that are just not available on rail, at least at this scale. Modeling road-specific disruptions may change this outcome but most likely in a significant manner because of the time lost during intermodal transfers. For a set of disruptions simulated, disruption was most severe when a minimally redundant section of highway was impacted. Lesser impacts occurred during rail disruptions and in areas where multiple alternate routes existed.

Immediate extensions of this analysis include creating more scenario simulations, so the statistical models can have more data to learn from, which would be especially helpful in predicting extreme events. Additionally, a main, inherent complication with this analysis is the difficulty in validating results against historical impacts. This issue is common when modeling network disruptions, as real life disruptions are relatively lacking compared to the amount of data necessary for establishing precise predictions (He and Liu 2012). Ideally, more

granular historical data, perhaps on a daily time-scale, can be obtained and used to further tune predictive accuracy.

Additional considerations to expand this work include incorporating Tennessee's inland waterway network within the intermodal framework and/or expanding the scale of analysis to include all of Tennessee or perhaps the entire Southeast region. Additionally, impacts can be integrated with economic models, so policy makers can obtain reasonable fiscal estimates of these disruptions. Furthermore, performing the routing using other techniques such as Python script beyond the ArcGIS Network Analyst could allow for providing more organic interdependent routing and potentially enable better modeling of temporal effects such as outage durations, etc.

Overall, this analysis provides a method for simulating and predicting impacts to traffic flow within Middle Tennessee's multimodal network in the event of disruptions. The ability to intuitively model impacts of disruptions is crucial, as real-life events are limited in both occurrence and data and their impacts are highly uncertain.

6. REFERENCES

- Asakura, Y. (1999). Reliability Measures of an Origin and Destination Pair in a Deteriorated Network with Variable Flow. In M. G. H. Bell (Ed.), *Transportation networks: recent methodological advances : selected proceedings of the 4th EURO Transportation Meeting*. United Kingdom: Emerland.
- Association of American Railroads (AAR). (2007) *National Rail Freight Infrastructure Capacity and Investment Study*. Cambridge, Massachusetts. Cambridge Systematics, Inc.
- Billington, R., and R.N. Allan. (1992). *Reliability Evaluation of Engineering Systems, Concepts and Techniques*. Plenum Press, New York.
- Bing, Alan J., Eric W. Beshers, Megan Chavez, David P. Simpson, Emmanuel S. Horowitz, Walter E. Zullig, Jr. (2010) *NCHRP Report 657: Guidebook for Implementing Passenger Rail Service on Shared Passenger and Freight Corridors*. Transportation Research Board.
- Chen, A. and P. Kasikitwiwat. (2010) *Modeling Capacity Flexibility of Transportation Networks*. *Transportation Research Part A: Policy and Practice* .Vol. 45, Issue 2, Elsevier Ltd.
- C. Matisziw, T., Murray, A., & H. Grubestic, T. (2007). *Bounding Network Interdiction Vulnerability Through Cutset Identification* (pp. 243–256). https://doi.org/10.1007/978-3-540-68056-7_12
- Ford, C. (2017, July 13). *Freight Analysis Framework (FAF)*. Retrieved November 27, 2017, from <http://faf.ornl.gov/faf4/Default.aspx>
- He, X., & Liu, H. X. (2012). Modeling the day-to-day traffic evolution process after an unexpected network disruption. *Transportation Research Part B: Methodological*, 46(1), 50–71. <https://doi.org/10.1016/j.trb.2011.07.012>
- Jenelius, E., & Mattsson, L.-G. (2012). Road network vulnerability analysis of area-covering disruptions: A grid-based approach with case study. *Transportation Research Part A: Policy and Practice*, 46(5), 746–760. <https://doi.org/10.1016/j.tra.2012.02.003>
- Kahn, M. (2017). MARTA saw no lasting increase in ridership from the I-85 bridge collapse. *Curbed Atlanta*. July 7, 2017. Available at <https://atlanta.curbed.com/2017/7/7/15930922/marta-ridership-increase-interstate-85-bridge-collapse>.

- Kim, Y., Chen, Y.-S., & Linderman, K. (2015). Supply network disruption and resilience: A network structural perspective. *Journal of Operations Management*, 33–34(Supplement C), 43–59.
<https://doi.org/10.1016/j.jom.2014.10.006>
- Kurauchi, F., Uno, N., Sumalee, A., & Seto, Y. (2009). *Network evaluation based on connectivity vulnerability*. Springer. https://doi.org/10.1007/978-1-4419-0820-9_31
- Morlok, E.K. and D.J. Chang. (2004) *Measuring Capacity Flexibility in a Transportation System*. *Transportation Research Part A: Policy and Practice*. Vol. 38, Issue 6. Elsevier Ltd.
- Parkinson, T. and I. Fisher. (1996) *Transit Cooperative Research Program Report 13: Rail Transit Capacity*. Washington, D.C. National Academy Press.
- Peterson, B. (2014, October). CTA Railroad Network. Retrieved May 15, 2018, from <https://cta.ornl.gov/transnet/RailRoads.html>.
- Pouryousef, H., P. Lautala, and T. White. (2015) *Railroad Capacity Tools and Methodologies in The U.S. and Europe*. Springerlink.com, *Journal of Modern Transportation*.
- Straight, B. 2010. "Tennessee flooding closes I-40, other major roads." *FleetOwner*. Available at <http://www.fleetowner.com/management/news/i40-tennessee-flooding-0503>.
- The Secretary Delay Attribution Board. (2015). *Delay Attribution Guide*. London, U.K. Available at <http://www.delayattributionboard.co.uk/delayattributionguides.htm>.
- Verma, M., & Verter, V. (2010). A lead-time based approach for planning rail–truck intermodal transportation of dangerous goods. *European Journal of Operational Research*, 202(3), 696–706.
<https://doi.org/10.1016/j.ejor.2009.06.005>
- Wright, R. and G. Baker (2017). *Modernizing FRA's Rail Freight Routing and Analysis Tools*. Retrieved from http://proceedings.esri.com/library/userconf/proc17/papers/2185_504.pdf
- Yang, H., M. Bell, and Q. Meng. (1999) *Modeling Capacity and Level of Service of Urban Transportation Networks*. Elsevier Ltd.

Yung-Chen, L. and C. Barkan. (2014) Enhanced Parametric Railway Capacity Evaluation Tool. Transportation Research Record: Journal of the Transportation Research Board. Vol. 2117. Elsevier Ltd.

Appendix 1: Data Processing

Data collection and processing is discussed below.

Regarding the roadway network, the following data was collected from FAF (<http://faf.oml.gov/faf4/Default.aspx>):

- FAF4 network shapefile (*faf4_esri_arcgis.zip*)
- FAF4 zone shapefile (*CFS_area_shapefile.zip*)
- FAF4 zone IDs (*CFS Area Code – FAF Zone ID.xlsx*)
- Segment flow metrics (*assignment_results.zip*)

The network shapefile contains the freight roadway network across the US. The FAF4 zone shapefile depicts the different regions, as determined by FAF4, across the US. Both files are projected in ArcGIS. Based on the zone IDs, region 400 corresponds to the Nashville/Middle-Tennessee area of interest for this project. Within ArcGIS, the FAF4 zone shapefile is then selected to where the CFS12_Area field is equal to 400; all other regions are removed from the map. Note this field corresponds to the most recent FAF framework; the 2007 field will give slightly different regions.

Using ArcGIS Dissolve tool, these regions are then merged into one polygon. Using the Buffer tool, a 50km buffer region is extended beyond the merged polygon. Using the Clip (Analysis) tool, the network shapefile is then selected to include only the roadway intervals that lie within the 50km buffered region. All other parts of the roadway network are removed. The clipped network is then joined with segment flow metrics via the FAF4_ID field.

The network is now populated with valuable traffic flow metrics for each roadway interval. An additional field, time (hours), is created by dividing the miles field by the speed limit (miles/hour). Additional fields are also created by converting existing metrics to units desired for the analysis. These fields are CAP12 * Miles (peak capacity of volume of traffic, number of vehicles, per hour multiplied by length of the highway segment), SF12 * Miles (average hourly volume of traffic, in number of vehicles, multiplied by length), and Ton_{Freight} * Miles (tons of freight per hour multiplied by length). There is no metric related to the capacity of freight specifically, so we use the assumption that the ratio of total service flow to total peak capacity is equivalent to the ratio of total freight throughput to total freight capacity.

The finalized network from above is imported into a geodatabase as a feature class. This feature class must be planarized, essentially splitting all lines at intersections to help ensure connectivity throughout the network. Next, topological rules are formulated for the planarized feature class to ensure the network analysis will perform as expected. They include the following rules: must

be a single part (to ensure that all line segments correspond to one highway section), must not overlap (to ensure that two line segments do not occupy the same space), must not self-overlap (to ensure that a road does not overlap with itself), must not self-intersect (to ensure that a given road interval does not form a loop with itself), must not intersect (to ensure the planarization worked correctly), and must not have dangles (to ensure connectivity). Upon inspection of errors, only the last rule is violated. Segments that give rise to these errors that are not on the boundary of the network are removed.

The corrected feature class is then exported as a shapefile, and a network dataset is created based on this shapefile. Another shapefile is created based on the network junction layer that is automatically produced from the network creation. The junction layer is dissolved by the SIGN1 and SPD_LIMIT fields to produce a point shapefile that will act as the OD input to assess system-wide performance

For the rail network, data is amalgamated from several sources, with exact references in the main paper:

- Private Waybill data for rail shipments through Tennessee (from TDOT)
- Freight Routing Toolkit (from US DOT)
- US rail network shapefile (from US DOT)
- US rail station shapefile (from US DOT)
- Rail capacity shapefile (from Oak Ridge National Laboratory)

Once loaded into SQL Server, or an equivalent relational database management system, the Waybill data need to be aggregated for each leg of the network in order to look at overall traffic flows. US DOT's Freight Routing Toolkit is leveraged for this task. As the toolkit is proprietary to US DOT, we refer the readers to that reference in the main paper for more information on the tool and its operating procedures. Essentially, several Python and SQL scripts are used to query and aggregate the Waybill data. Results are then projected into ArcGIS as a shapefile. The outputs for this analysis are segment flows of Expanded Car Tons (that statistical extrapolation of the waybill samples to estimate yearly freight volume, in tons), filtered by freight which had an origin or destination within Tennessee and by freight which had an origin or destination within Tennessee and the cargo had rail-to-road intermodal capabilities. The former filter is used for the assumption that in the event that a reroute is necessary due to a disruption in the Middle Tennessee area, the rail shipment will reroute to another track of the US network prior to reaching Tennessee. Therefore, cargo flows of only Tennessee origin or destination are included in this analysis. Comparing overall flows to this filter suggest that the proportion of through-Tennessee cargo to the proportion of OD-Tennessee cargo is reasonable. The latter, additional intermodal criteria is used to assess the amount of freight that can realistically be routed to road, given a disruption

occurs. Outputs are then converted to hourly ton miles, to align with road volume and capacity metrics.

The segment flow outputs are then spatially joined to the US rail network shapefile, and similarly with the rail capacity shapefile. There is currently no publicly available metric for rail capacity; Bruce Peterson at Oak Ridge National Laboratory formulated a segment-level traffic density based on his experience. His metric is used to derive segment capacity measurements for this analysis in the following equation:

$$\text{capacity}_{\text{segment}} = \text{flow}_{\text{segment}} * (\text{density}_{\text{median}} + 1) / (\text{density}_{\text{segment}} + 1).$$

Here, rail capacity is proportional to the amount of volume multiplied by the ratio of the median density metric, since it is an ordinal value from 0 to 7, to the segment-defined density. The plus ones offer an easy way to avoid dividing by zero.

Lastly, the speed limit of all the rail segments is defined at 50mph, the midpoint of the speed limits for Class 3 and Class 4 freight tracks. We were not able to find publicly available data for granular speed limits on tracks, so this simplifying assumption serves as a good average approximation for system-wide speeds. In the network simulations, discussed in Appendix 2, simulated routing is ultimately governed by average expected travel times, estimated by speed limits. Segment level rail speed limits would be ideal, but since the most mainline freight tracks are Class 3 or Class 4, this simplifying assumption still allows for intuitive interpretations.

Like the road network, this rail network is then clipped by the 50km buffer and then subjected to the same topological rule checks.

The data now consist of two shapefiles, a road network and a rail network that comprise the Middle Tennessee region. The next step is to connect the two networks.

Connecting lines within ArcGIS must first be created from the rail stations to each of the networks. ESRI provides good instruction for accomplishing this task (<https://support.esri.com/en/technical-article/000011815>). The end product of this procedure is two sets of connection lines and corresponding nodes. All will be utilized in the intermodal network assembly.

To assemble the final, interconnected network, seven shapefiles are imported into a feature dataset:

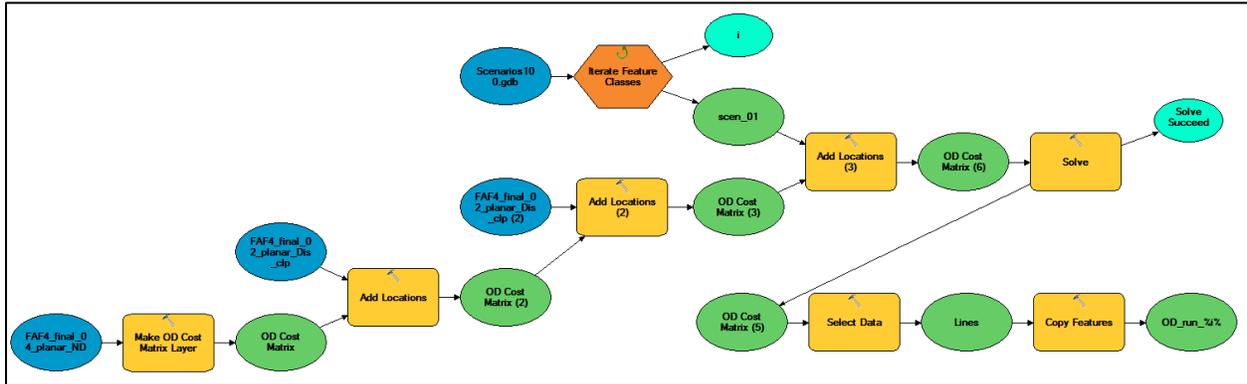
- Road network
- Rail network

- Rail stations
- Connecting lines 1 (road network to rail station)
- Connecting lines 2 (rail network to rail station)
- Nodes corresponding to connecting lines 1
- Nodes corresponding to connecting lines 2

An ArcGIS intermodal network is then created with the above feature classes. The road network is connected to the nodes 1, connecting lines 1 are connected to nodes 1 and rail stations, connecting lines 2 are connected nodes 2 to rail stations, and the rail network is connected to nodes 2 and the rail stations. The latter part of this last connection is necessary because some connecting lines 2 have zero distance because the rail stations are located on the rail lines themselves. Since all intermodal traffic will have to pass through connecting lines 1, this part of the network is used to enforce the 1 hour time penalty associated with switching modes of transportation. The network is then configured to use time as an impedance metric (minimizing travel time as an objective function) while also accumulating volume, capacity, time, and distance metrics. More details on these variables and their role in the simulation process are discussed in the main text and Appendix 2.

Appendix 2: Network Simulation

Using ArcGIS's Model Builder and OD Cost Matrix within the Network Analysis Tool, routing simulations are iteratively run on the intermodal network. A model process diagram is presented below.



As mentioned in the main text, the OD pairs are divided into two distinct sets, one for each mode of transport, so this modeling process is conducted for each grouping. The chosen OD pairs are added to the network, and OD Cost Matrix solves a routing problem between all pairs, calculating total travel time (hours), total distance traveled (miles), total expected freight throughput on road (hourly ton-miles), total capacity of the road route (hourly ton-miles), total expected freight throughput on rail (hourly ton-miles), total capacity of the rail route (hourly ton-miles), and total expected intermodal freight throughput (hourly ton-miles). The corresponding equations are provided below.

- Road
 - Volume

$$\text{YKTON}_{12} / (365 * 24) * \text{MILES} / 1000 = \text{Freight ton miles per hour (volume for each segment)}$$
 - Capacity

$$\text{Freight Ton miles per hour (above)} * (\text{CAP}_{12} * \text{MILES}) / (\text{SF}_{12} * \text{MILES}) = \text{Freight Ton miles per hour (capacity for each segment)}$$
- Rail
 - Volume

$$\text{EXPANDED_TONS} / (365 * 24) = \text{Freight ton miles per hour (volume for each segment)}$$
 - Capacity

$$\text{Freight Ton miles per hour (above)} * (\text{DENSITY}_{\text{median}} + 1) / (\text{DENSITY}_{\text{segment}} + 1)$$

Simulated disruptions of randomly varying size, range, and location are then added to the iterative process. To accomplish this task, the Random Points tool is used to create 99 random points within the Middle Tennessee region. This point feature is then updated with two new fields, radius and severity, which are randomly populated with values using Python code below.

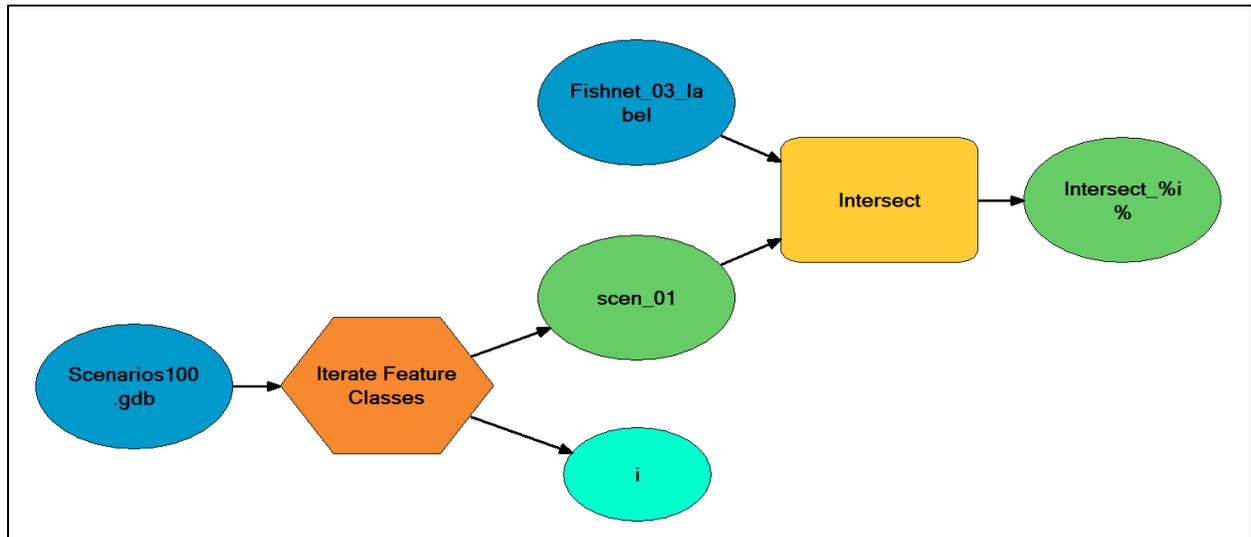
```
#Create two fields, "severity" and "radius"  
#Random number for radius (km)  
import arcpy, random  
fc="C:/Users/Paul  
J/Desktop/VECTOR/04TDOT_RouteDiversion/TDOT_GIS_final_routing/TDOT_GIS_final_rout  
ing.gdb/random_disruptions_04"  
with arcpy.da.UpdateCursor(fc, "radius") as cur:  
    for row in cur:  
        row[0] = random.randint(1, 65)  
        cur.updateRow(row)
```

```
#Random number for severity 2:5 ... will replace 5 w/ 99 after code is run  
import arcpy, random  
fc="C:/Users/Paul  
J/Desktop/VECTOR/04TDOT_RouteDiversion/TDOT_GIS_final_routing/TDOT_GIS_final_rout  
ing.gdb/random_disruptions_04"  
with arcpy.da.UpdateCursor(fc, "severity") as cur:  
    for row in cur:  
        row[0] = random.randint(2, 5)  
        cur.updateRow(row)
```

Take note to replace severity "5" with "99" after the code is run. Each point feature is then buffered by the specified radius and incorporated within the simulation. A base case, no disruption, is also added to the list. The severity fields of this disruption are used as an input to the scaled cost of time travel within the network options. The location and radius of the events dictate which routes are impacted.

Solutions are saved as line features, which can be exported to CSV files. Figure 6 below depicts the model. In total, 190,724 routes (136,900 OD pairs plus 52,824 rail OD pairs) are solved for each scenario, and there are a total of 100 scenarios (99 disruptions plus the base case). Accumulated network metrics for each route for each simulated disruption are compared to those of the base case to establish estimates for the total system-wide impacts. These comparisons are discussed in the main text.

Lastly, grid-based location variables are set up to help train the statistical models. These variables are created by overlaying a grid, using the Create Fishnet Tool, on top of the Middle Tennessee region. Each scenario is then iteratively applied to the grid, flagging which boxes are affected by each disruption, outlined below.



Appendix 3: Statistical Modeling using R Software

The following code shows how outputs from the network simulation are processed and used to train statistical models that predict impacts of future disruptions.

```
#####  
#####  
#LOAD O/D NETWORK ANALYSIS RESULTS FROM ARCGIS  
  
#Batch import csv files from ArcGIS output  
setwd("C:/Users/Paul  
J/Desktop/VECTOR/04TDOT_RouteDiversion/00_Final_Report_TDOT/data_out_road")  
my_files_road <- list.files(pattern = "\\\\.csv$")  
my_data_road = lapply(my_files_road, read.csv)  
my_data_road = lapply(my_data_road, function(df){ #sort data by OD to make sure files  
aligned  
  df[order(df$OriginID,df$DestinationID),]  
})  
  
#repeat for rail  
setwd("C:/Users/Paul  
J/Desktop/VECTOR/04TDOT_RouteDiversion/00_Final_Report_TDOT/data_out_rail")  
my_files_rail <- list.files(pattern = "\\\\.csv$")  
my_data_rail = lapply(my_files_rail, read.csv)  
my_data_rail = lapply(my_data_rail, function(df){ #sort data by OD to make sure files aligned  
  df[order(df$OriginID,df$DestinationID),]  
})  
  
#save workspace checkpoint 1  
setwd("C:/Users/Paul J/Desktop/VECTOR/04TDOT_RouteDiversion/00_Final_Report_TDOT")  
save.image(file = "checkpoint01_b.RData")  
  
#need to trim road and rail ODs that prefer intermodal in base case  
omit_road=which(my_data_road[[1]]$Total_FLOW_rail>0) #empty  
omit_rail=which(my_data_rail[[1]]$Total_Flow_road>0) #4981 pairs out of 38512 ODs  
my_data_rail=lapply(my_data_rail,function(df){  
  df=df[-omit_rail,]  
})  
  
#base case shorthand  
road_00 = my_data_road[[1]]  
rail_00 = my_data_rail[[1]]  
  
#add outputs part 1
```

```

"/" <- function(x,y) ifelse(y==0,0,base::"/"(x,y))
my_data_road = lapply(my_data_road, function(df){
  data.frame(df
    , "Time_diff"=df$Total_Time-road_00$Total_Time
    , "Miles_diff"=df$Total_Miles-road_00$Total_Miles
    , "Flag_reroute"=as.numeric((abs(df$Total_Miles-road_00$Total_Miles))>0.0001)
    , "Flag_road"=as.numeric(1)
    , "Flag_IM"=as.numeric((abs(df$Total_FLOW_rail_tn-
road_00$Total_FLOW_rail_tn))>0.0001)
    , "Cap_road_tons"=df$Total_Flow_road/df$Total_Cap_road*df$Total_TON_road
  )
})

my_data_rail = lapply(my_data_rail, function(df){
  data.frame(df
    , "Time_diff"=df$Total_Time-rail_00$Total_Time
    , "Miles_diff"=df$Total_Miles-rail_00$Total_Miles
    , "Flag_reroute"=as.numeric((abs(df$Total_Miles-rail_00$Total_Miles))>0.0001)
    , "Flag_road"=as.numeric(0)
    , "Flag_IM"=as.numeric((abs(df$Total_Flow_road-rail_00$Total_Flow_road))>0.0001)
    , "Cap_road_tons"=df$Total_Flow_road/df$Total_Cap_road*df$Total_TON_road
  )
})

#save workspace checkpoint 2
setwd("C:/Users/Paul J/Desktop/VECTOR/04TDOT_RouteDiversion/00_Final_Report_TDOT")
save.image(file = "checkpoint02_b.RData")

road_00 = my_data_road[[1]]
rail_00 = my_data_rail[[1]]

#add outputs part 2
my_data_road = lapply(my_data_road, function(df){
  data.frame(df
    , "Road_Vol"=(1-
df$Flag_reroute)*(df$Total_TON_road*df$Total_Time/road_00$Total_Time)
    + df$Flag_reroute*(1-df$Flag_IM)*(df$Total_TON_road +
0.5*road_00$Total_TON_road)
    +
df$Flag_reroute*df$Flag_IM*(road_00$Total_TON_road*df$Total_Time/road_00$Total_Time
  )
    , "Road_Cap"=(1-df$Flag_reroute)*(df$Cap_road_tons)
    + df$Flag_reroute*(1-df$Flag_IM)*(df$Cap_road_tons)
    + df$Flag_reroute*df$Flag_IM*road_00$Cap_road_tons
    , "Rail_Vol"=(1-
df$Flag_reroute)*(df$Total_FLOW_rail_tn*df$Total_Time/road_00$Total_Time)
  )
})

```

```

    + df$Flag_reroute*(1-df$Flag_IM)*(df$Total_FLOW_rail_tn +
0.5*road_00$Total_FLOW_rail_tn)
    + df$Flag_reroute*df$Flag_IM*(df$Total_FLOW_rail_tn_im)
    , "Rail_Cap"=(1-df$Flag_reroute)*(df$Total_CAP_rail_tn)
    + df$Flag_reroute*(1-df$Flag_IM)*(df$Total_CAP_rail_tn)
    + df$Flag_reroute*df$Flag_IM*road_00$Total_CAP_rail_tn
  )
})

my_data_rail = lapply(my_data_rail, function(df){
  data.frame(df
    , "Road_Vol"=(1-
df$Flag_reroute)*(df$Total_TON_road*df$Total_Time/rail_00$Total_Time)
    + df$Flag_reroute*(1-df$Flag_IM)*(df$Total_TON_road +
0.5*rail_00$Total_TON_road)
    + df$Flag_reroute*df$Flag_IM*(rail_00$Total_FLOW_rail_tn_im)
    , "Road_Cap"=(1-df$Flag_reroute)*(df$Cap_road_tons)
    + df$Flag_reroute*(1-df$Flag_IM)*(df$Cap_road_tons)
    + df$Flag_reroute*df$Flag_IM*rail_00$Cap_road_tons
    , "Rail_Vol"=(1-
df$Flag_reroute)*(df$Total_FLOW_rail_tn*df$Total_Time/rail_00$Total_Time)
    + df$Flag_reroute*(1-df$Flag_IM)*(df$Total_FLOW_rail_tn +
0.5*rail_00$Total_FLOW_rail_tn)
    + df$Flag_reroute*df$Flag_IM*((rail_00$Total_FLOW_rail_tn -
rail_00$Total_FLOW_rail_tn_im)*df$Total_Time/rail_00$Total_Time)
    , "Rail_Cap"=(1-df$Flag_reroute)*(df$Total_CAP_rail_tn)
    + df$Flag_reroute*(1-df$Flag_IM)*(df$Total_CAP_rail_tn)
    + df$Flag_reroute*df$Flag_IM*rail_00$Total_CAP_rail_tn
  )
})

```

```

road_00 = my_data_road[[1]]
rail_00 = my_data_rail[[1]]

```

```

#sum over all rows (O/D pairs)
sum_data_road = lapply(my_data_road, function(df){
  apply(df[,c("Road_Vol", "Road_Cap", "Rail_Vol", "Rail_Cap")], 2, sum)
})
sum_data_rail = lapply(my_data_rail, function(df){
  apply(df[,c("Road_Vol", "Road_Cap", "Rail_Vol", "Rail_Cap")], 2, sum)
})

```

```

#put data in matrix

```

```

data_out_road = matrix(0,nrow=length(my_data_road),ncol=4
                      ,dimnames =list(NULL,c("Road_Vol","Road_Cap","Rail_Vol","Rail_Cap"))
                      )
for (i in 1:length(my_data_road)){
  data_out_road[i,]=sum_data_road[[i]]
}

data_out_rail = matrix(0,nrow=length(my_data_rail),ncol=4
                      ,dimnames =list(NULL,c("Road_Vol","Road_Cap","Rail_Vol","Rail_Cap"))
                      )
for (i in 1:length(my_data_rail)){
  data_out_rail[i,]=sum_data_rail[[i]]
}

#combine in data frame
data_out=data.frame(data_out_road+data_out_rail)

#calculate Volume to Capacity metrics for each
data_out = data.frame(data_out
                      , "VCR_road"=data_out$Road_Vol/data_out$Road_Cap
                      , "VCR_rail"=data_out$Rail_Vol/data_out$Rail_Cap
                      , "VCR_comb"=(data_out$Road_Vol+data_out$Rail_Vol)/(data_out$Road_Cap+data_out$Rail_Cap)
                      )

#strip out null run and convert to dataframe
data_out_fr = data.frame(data_out[-1,])

setwd("C:/Users/Paul J/Desktop/VECTOR/04TDOT_RouteDiversion/00_Final_Report_TDOT")
save.image(file = "checkpoint03_c.RData")

#####
#####
#LOAD DISRUPTION SCENARIO FLAGS FROM ARCGIS

#Batch import csv files from ArcGIS output
setwd("C:/Users/Paul J/Desktop/CEE5999 Data Analytics R/Project/flags_out")
my_files2 <- list.files(pattern = "\\\\.csv$")
my_data2 <- lapply(my_files2, read.csv)
flag_arr=matrix(0,nrow=length(my_data2),ncol=64) #64 is number of grids

#find where there are no observations
ind=t(sapply(my_data2, function(df) {dim(df)}))

```

```

which(ind[,1]==0) #25 28 31 38 51 55 56 59 72 85 86 88 94

#Within list, convert dataframes to matrices and then create flags in flg_arr for each gridbox
referenced by each list element
my_data2_a = lapply(my_data2, function(df) {data.matrix(df)})
for (i in 1:99){
  if (i %in% which(ind[,1]==0))
    {print("No row inserts")}
  else
    {flag_arr[i,my_data2_a[[i]][,11]]=1}
}
colnames(flag_arr) = paste("grid_",1:64,sep="")

#convert flg_arr to dataframe
flag_out_fr=data.frame(flag_arr)

#####
#####
#LOAD DISRUPTION SCENARIO INFO FROM ARCGIS

library(readxl)
random_disruptions_04_a <- read_excel("C:/Users/Paul J/Desktop/CEE5999 Data Analytics
R/Project/random_disruptions_04_a.xls")

#####
#####
#APPEND ALL DATA TOGETHER

data_final=data.frame(random_disruptions_04_a,flag_out_fr,data_out_fr[,-c(1:4)])
head(data_final)
setwd("C:/Users/Paul J/Desktop/VECTOR/04TDOT_RouteDiversion/00_Final_Report_TDOT")
#save.image(file = "checkpoint04.RData")
#save(data_final, file="checkpoint04_data_final.RData")

#####
#####
#TRAIN STATISTICAL MODEL

#FIELD KEY
#(1)OID, (2)radius, (3)severity, (4)POINT_X, (5)POINT_Y
#(6)-(69) grid_1-grid_64
#(70)VCR_road (71)VCR_rail (72)VCR_comb

library(gbm)

```

```

library(reshape2)
library(lattice)
rmse <- function(resid) {sqrt(mean(resid^2))}

##Take log of DVs (analysis with raw data shows residual errors increase with event
magnitude)
data_final$VCR_road_ln=log(data_final[, "VCR_road"])
data_final$VCR_rail_ln=log(data_final[, "VCR_rail"])
data_final$VCR_comb_ln=log(data_final[, "VCR_comb"])

#split into train/test
set.seed(2)
train= sample(seq(99), 70, replace = FALSE)
data_train=data_final[train,]
data_test=data_final[-train,]

#Various statistical methods employed: MLR, PCR, LASSO, GAMs, RFs ... Boosted Trees have
by far the best errors
#BOOSTED
TREES#####
#####
cv.boost=function(data,k,t,h,d){
  set.seed(2)
  start_time=Sys.time()
  folds=sample(rep(1:k,length=nrow(data)))
  tune.trees=seq(from=1000,to=t,by=1000)
  t_length=t/1000
  tune.shrink=seq(from=.001,to=h,by=.001)
  h_length=h/.001
  tune.depth=seq(from=1,to=d,by=1)
  cv.err.array=array(NA,dim=c(k,t_length,h_length,d))
  in.err.array=array(NA,dim=c(k,t_length,h_length,d))

  #iterate through all tuning parameters for CV_MSE and InSample_MSE
  for (w in 1:k){
    for (x in 1:t_length){
      for (y in 1:h_length){
        for (z in 1:d){
          boost.fit=gbm(VCR_comb_ln~.
            ,data=data[folds!=w,]
            ,distribution="gaussian"
            ,n.trees=tune.trees[x]
            ,shrinkage=tune.shrink[y]
            ,interaction.depth=tune.depth[z]

```

```

        ,n.minobsinnode = 2)
    boost.pred=predict(boost.fit,newdata=data[folds==w,],n.trees=tune.trees[x])
    cv.err.array[w,x,y,z]=mean((data$VCR_comb_ln[folds==w]-boost.pred)^2)
    in.err.array[w,x,y,z]=mean((data$VCR_comb_ln[folds!=w]-boost.fit$fit)^2)
    print(paste(w,x,y,z))
  }
}
}
}

avg.cv.err=apply(cv.err.array,c(2,3,4),mean) #average across k folds
avg.in.err=apply(in.err.array,c(2,3,4),mean) #average across k folds
print(Sys.time()-start_time) #~1hr to run
list.err=list("cv"=avg.cv.err,"in"=avg.in.err)
return(list.err)
}

data_train_clean=data_train[,-c(1,70,71,72,73,74)] #remove Object ID and variables that
perfectly predict ln(DVs)
bt.1=cv.boost(data_train_clean,5,6000,.01,4)

#run-time: ~10min
sqrt(min(bt.1$"cv")) #0.337
which(bt.1$"cv"==min(bt.1$"cv"),arr.ind = TRUE) #1 4 4
bt.1.melt=melt(bt.1$"cv")
colnames(bt.1.melt)=c("trees","lamda","depth","MSE")
wireframe(bt.1.melt$MSE~bt.1.melt$trees*bt.1.melt$lamda|bt.1.melt$depth,screen=list(z=245,x
=-75,y=3)
  ,drape=TRUE
  ,col.regions = colorRampPalette(c("blue", "red"))(80))

bt.1.best=gbm(VCR_comb_ln~.
  ,data=data_train_clean
  ,distribution="gaussian"
  ,n.trees=1000
  ,shrinkage=.004
  ,interaction.depth=4
  ,n.minobsinnode = 2)
summary(bt.1.best) #severity, radius, grid_52, POINT_Y, grid_37
plot(data_train$VCR_comb_ln)
lines(bt.1.best$fit,col="red")
rmse_bt.1.fit=rmse(bt.1.best$fit-data_train_clean$VCR_comb_ln) #.104
bt.1.pred=predict(bt.1.best,newdata = data_test, n.trees=1000, shrinkage=.004,
interaction.depth=4, n.minobsinnode=2)
rmse_bt.1.pred=rmse(bt.1.pred-data_test$VCR_comb_ln) #0.278; Mean=0.531 SD=0.629

```

```

##BOOTSTRAP GBM FOR PREDICTION
INTERVAL#####
####
library(boot)
booty.pred=matrix(NA,nrow=29,ncol=100)
set.seed(2)
for (b in 1:100){
  booty=sample(1:70,70,replace=TRUE)
  booty=data_train_clean[booty,]

booty.gbm=gbm(VCR_comb_ln~.,data=booty,distribution="gaussian",n.trees=1000,shrinkage=
.004,interaction.depth=4,n.minobsinnode = 2)
  booty.pred[,b]=predict(booty.gbm,newdata = data_test, n.trees=1000, shrinkage=.004,
interaction.depth=4, n.minobsinnode=2)
}

booty.quantile=matrix(NA,nrow=29,ncol=2)
for (r in 1:29){
  booty.quantile[r,1]=quantile(booty.pred[r,],.05)
  booty.quantile[r,2]=quantile(booty.pred[r,],.95)
}

df.gbm=data.frame(1:29,data_test$VCR_comb_ln,bt.1.pred,booty.quantile)
df.gbm=df.gbm[order(df.gbm$data_test.VCR_comb_ln,df.gbm$bt.1.pred),]

plot(df.gbm$data_test.VCR_comb_ln, xlab="Scenario",ylab="System
ln(Vol/Cap)",pch=19,ylim=c(0,4))
lines(df.gbm$bt.1.pred,col="red",lwd=2,type="l")
lines(df.gbm$X1,col="red",lty=3)
lines(df.gbm$X2,col="red",lty=3)
legend("topleft",legend=c("prediction","95% PI"),col="red",
      lwd=2,bty="n",cex=1,lty=c(1,3))

setwd("C:/Users/Paul J/Desktop/VECTOR/04TDOT_RouteDiversion/00_Final_Report_TDOT")
save.image(file = "checkpoint05.RData")

```