University of Arkansas, Fayetteville ScholarWorks@UARK

Theses and Dissertations

7-2020

# Commodity-based Freight Activity on Inland Waterways through the Fusion of Public Datasets for Multimodal Transportation Planning

Magdalena I. Asborno University of Arkansas, Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

Part of the Civil Engineering Commons, Structural Engineering Commons, and the Transportation Engineering Commons

## Citation

Asborno, M. I. (2020). Commodity-based Freight Activity on Inland Waterways through the Fusion of Public Datasets for Multimodal Transportation Planning. *Theses and Dissertations* Retrieved from https://scholarworks.uark.edu/etd/3760

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact ccmiddle@uark.edu.

# Commodity-based Freight Activity on Inland Waterways through the Fusion of Public Datasets for Multimodal Transportation Planning

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering

by

# Magdalena I. Asborno La Plata National University Bachelor of Science in Civil Engineering, 2004 La Plata National University Bachelor of Science in Transportation Infrastructure Engineering, 2004 Energy Superior Institute Master of Science in Petroleum, Gas, and Marketing, 2007

## July 2020 University of Arkansas

This dissertation is approved for recommendation to the Graduate Council

Sarah Hernandez, Ph.D. Dissertation Director

Heather Nachtmann, Ph.D. Committee Member Kevin Hall, Ph.D. Committee Member

John Kent, Ph.D. Committee Member

#### ABSTRACT

Within the U.S., the 18.6 billion tons of goods currently moved along the multimodal transportation system are expected to grow 51% by 2045. Most of those goods are transported by roadways. However, several benefits can be realized by shippers and consumers by shifting freight to more efficient modes, such as inland waterways, or adopting a multimodal scheme. To support such freight growth sustainably and efficiently, federal legislation calls for the development of plans, methods, and tools to identify and prioritize future multimodal transportation infrastructure needs. However, given the historical mode-specific approach to freight data collection, analysis, and modeling, challenges remain to adopt a fully multimodal approach that integrates underrepresented modes, such as waterways, into multimodal forecasting tools to identify and prioritize transportation infrastructure needs. Examples of such challenges are data heterogeneity, confidentiality, limitations in terms of spatial and temporal coverage, high cost associated with data collection, subjectivity in surveys responses, etc. To overcome these challenges, this work fuses data across a variety of novel transportation sources to close existing gaps in freight data needed to support multimodal long-range freight planning. In particular, the objective of this work is to develop methods to allow integration of inland waterway transportation into commodity-based freight forecasting models, by leveraging Automatic Identification System (AIS) data. The following approaches are presented in this dissertation:

 Maritime Automatic Identification System (AIS) data is mapped to a detailed inland navigable waterway network, allowing for an improved representation of waterway modes into multimodal freight travel demand models which currently suffer from unbalanced representation of waterways. Validation results show the model correctly identifies 84% stops at inland waterway ports and 83.5% of trips crossing locks.

- AIS and truck Global Positioning System (GPS) data are fused to a multimodal network to identify the area of impact of a freight investment, providing a single methodology and data source to compare and contrast diverse transportation infrastructure investments. This method identifies parallel truck and vessel flows indicating potential for modal shift.
- iii) Truck GPS and maritime Lock Performance Monitoring System (LPMS) data are fused via a multi-commodity assignment model to characterize and quantify annual commodity throughput at port terminals on inland waterways, generating new data from public datasets, to support estimation of commodity-based freight fluidity performance measures. Results show that 84% of ports had less than a 20% difference between estimated and observed truck volumes.
- iv) AIS, LPMS, and truck GPS datasets are fused to disaggregate estimated annual commodity port throughput to vessel trips on inland waterways. Vessel trips characterized by port of origin, destination, path, timestamp, and commodity carried, are mapped to a detailed inland waterway network, allowing for a detailed commodity flow analysis, previously unavailable in the public domain.

The novel, repeatable, data-driven methods and models proposed in this work are applied to the 43 freight port terminals located on the Arkansas River. These models help to evaluate network performance, identify and prioritize multimodal freight transportation infrastructure needs, and introduce a unique focus on modal shift towards inland waterway transportation. ©2020 by Magdalena I. Asborno All Right Reserved

#### ACKNOWLEDGEMENTS

To Dr. Hernandez, my advisor, who made my Ph.D. journey an incredibly enjoyable experience. Her guidance, encouragement, gratitude, flexibility, and understanding made this long-desired dream come true. Thanks to my committee members, Dr. Nachtmann, Dr. Hall, and Dr. Kent, for providing valuable comments to my dissertation proposal, and accepting necessary changes along the way.

Thanks again to my advisor, Dr. Hernandez, to the Civil Engineering Department Chair, Dr. Hale, and to the University of Arkansas Graduate School and International Education for supporting my Ph.D. studies through the Doctoral Academy Fellowship; and for trusting me to outreach our work at numerous congresses during the past four years. Thanks to the Maritime Transportation Research and Education Center (MarTREC), and to the Arkansas Department of Transportation for funding the research projects I contributed to at the Freight Transportation Data Lab; to Dr. Hernandez for securing such funding, and trusting me to act as their lead graduate research assistant.

Thanks to my co-workers at the Freight Transportation Data Lab for their professionalism and companionship at conferences, at Cato Springs Research Center, and during long data collection days: Karla Diaz-Corro, Taslima Akter, Mariah Crews, Sanjeev Bhurtyal, Madeline Giebler, Rebecca Hardwick, Sharif Mahmud, Marjan Asadinia, Fu Durandal, Josh Baker, Collin Burris, Manzi Yves, Dishoungh White, Donn Mershon, Ian Hargrove, and Kyle Sadler. Special thanks to Taslima Akter for running the map-matching algorithm to truck GPS data, and to Manzi Yves for groundtruthing vessel tracking data, necessary for this work. Thanks to all researchers and practitioners around the world who develop and share open source tools. In particular, thanks to Pedro Camargo, for developing the source code for the stop identification and map-matching algorithm that I adapted for inland waterways.

This work is dedicated to my mother, Alicia, my father, Marcelo, my brother, Pedro and my partner, Daniel, who I missed the most during the past four years. Thank you for your unconditional love and support, and for advocating through example an equitable world for people of all races, religions, gender, and nationalities.

# LIST OF PUBLISHED PAPERS

**Chapter 5:** Asborno, M.I., Hernandez, S., & Akter, T. (2020), Multi-commodity port throughput from truck GPS and lock performance data fusion. *Maritime Economics and Logistics*. https://doi.org/10.1057/s41278-020-00154-7

# TABLE OF CONTENTS

CHA	APTER 1. Introduction	1
1.1.	Multimodal Freight Planning	1
1.2.	Planning Stage: Identification of Needs	2
1.3.	Commodity-based Freight Planning and Freight Fluidity	4
1.4.	Investment Evaluation and Prioritization of Needs	7
1.5.	Research Objectives	
1.6.	References	9
CHA	APTER 2. Background	12
2.1.	Data Sources for Long-Range Freight Planning	12
2.2.	Data Fusion	22
2.3.	Data Sources Fused in This Work	23
2.4.	References	25
CHA	APTER 3. Network Mapping of AIS data to Characterize Inland Waterway Frei	ght
Trar	nsportation	29
3.1.	Abstract	29
3.2.	Introduction	30
3.3.	Background	32
3.4.	Methodology	37
3.5.	Case Study: Maritime Freight Activity in Arkansas	49
3.6.		
	Discussion	54
3.7.	Discussion	54 56
<ul><li>3.7.</li><li>3.8.</li></ul>	Discussion Conclusion Disclaimer	54 56 58
<ol> <li>3.7.</li> <li>3.8.</li> <li>3.9.</li> </ol>	Discussion Conclusion Disclaimer References	54 56 58 58
<ul><li>3.7.</li><li>3.8.</li><li>3.9.</li><li>CHA</li></ul>	Discussion Conclusion Disclaimer References APTER 4. GIS-Based Identification and Visualization of Multimodal Freight	54 56 58 58
<ul><li>3.7.</li><li>3.8.</li><li>3.9.</li><li>CHA</li><li>Tran</li></ul>	Discussion Conclusion Disclaimer References APTER 4. GIS-Based Identification and Visualization of Multimodal Freight nsportation Impact Areas	54 56 58 58 58
<ul> <li>3.7.</li> <li>3.8.</li> <li>3.9.</li> <li>CHA</li> <li>Tran</li> <li>4.1.</li> </ul>	Discussion Conclusion Disclaimer References APTER 4. GIS-Based Identification and Visualization of Multimodal Freight nsportation Impact Areas Abstract	54 56 58 58 58 62

4.3.	Background	66
4.4.	Methodology	71
4.5.	Case Study: Impact Areas of Port terminals on the Arkansas River	
4.6.	Discussion	85
4.7.	Conclusion	87
4.8.	Acknowledgements and Disclaimer	89
4.9.	References	89

# CHAPTER 5. Multi-Commodity Port Throughput from Truck GPS and Lock

Perfo	ormance Data Fusion	93
5.1.	Abstract	93
5.2.	Introduction	94
5.3.	Methodology	100
5.4.	Case Study: McClellan-Kerr Arkansas River Navigation System	111
5.5.	Results	113
5.6.	Discussion	116
5.7.	Conclusions	120
5.8.	Acknowledgements and disclaimer	121
5.9.	Authors Contributions Statement	121
5.10.	References	122
СНА	PTER 6. Commodity-based Vessel Trip Characterization on Inland Waterways	126
6.1.	Abstract	126
6.2.	Introduction and Background	127
6.3.	Methodology	130
6.4.	Case Study: Commodities Transported by Vessels on the Arkansas River	138
6.5.	Model Evaluation and Discussion	143
6.6.	Conclusion	149
6.7.	References	150

CHAPTER 7. Conclusion and Future Work 1	153
---	-----

APPENDIX A. Synthesis on Potential Data Sources for Commodity-Based, Multimodal				
Long	Long-Range Freight Planning			
A.1.	Automatic Identification System (AIS)	164		
A.2.	Data from the Waterborne Commerce Statistics Center (WCSC)	165		
A.3.	Rail Carload Waybill	168		
A.4.	Truck GPS Data	169		
A.5.	Business Data	172		
A.6.	Data from the United States Department of Agriculture (USDA)	173		
A.7.	Data from the United States Energy Information Administration (EIA)	176		
A.8.	Weigh-In-Motion (WIM) System.	177		
A.9.	Federal Motor Carrier Safety Administration (FMCSA).	178		
A.10.	National Performance Management Research Data Set (NPMRDS).	179		
A.11.	References	179		

# LIST OF TABLES

Table 2.1	Commodity Flow Survey - Sample Design	7
Table 2.2	Comparison of Transearch, CFS, and FAF	0
Table 2.3	Data Sources Fused in This Work	4
Table 3.1	AIS Data Characteristics	5
Table 3.2	Vessel Stop Identification Parameters	1
Table 4.1	Vehicle Tracking Data Characteristics	8
Table 4.2	Key Performance Indicators of Each Freight Facility, Measured From its Impact Area	
		7
Table 4.3	Multimodal Stop Identification Parameters	0
Table 5.1	Multi-Commodity Assignment Model Formulation 108	8
Table 5.2	2016 McClellan Kerr-Arkansas Upriver Freight Transloaded per Port, Commodity,	
and Mode	(Annual Tons)114	4
Table 5.3	2016 McClellan Kerr-Arkansas Downriver Freight Transloaded per Port, Commodity,	,
and Mode	(Annual Tons)115	5
Table 5.4	Model Evaluation Metric per River Section (EM) 116	б
Table 6.1	LPMS Commodity Classification	8
Table 6.2	Model Formulation – Stage 1	5
Table 6.3	Model Validation Metric V <sub>j,l</sub> MKARNS, 2016	4
Table 6.4	Validation of Ton-miles by Commodity <i>M<sub>j</sub></i> MKARNS, 2016	5
Table A.1 Planning .	Main Characteristics of Data Sources Non-Systematically Utilized for Freight	0
Table A.2	Sample GPS Data Coverage Coefficients in Arkansas	1

# LIST OF FIGURES

Figure 1.1 Project assessment framework	2
Figure 1.2 Structure of 4-step TDM with commodity-based freight component	3
Figure 3.1 AIS data preparation flowchart	39
Figure 3.2 Example of AIS data preparation	40
Figure 3.3 Sample anchoring and port terminal areas	42
Figure 3.4 Detailed navigable waterway network development	43
Figure 3.5 Stop cluster example	44
Figure 3.6 AIS stop identification flowchart	45
Figure 3.7 AIS trip identification flowchart	46
Figure 3.8 Map-matching and trip characterization algorithm for a sample vessel	48
Figure 3.9 Stop identification results from AIS data	52
Figure 3.10 Trip identification results from AIS data	53
Figure 3.11 Sensitivity of stop identification output to algorithm parameters	56
Figure 4.1 AIS and truck GPS data fusion methodology flowchart	73
Figure 4.2 Study area: 43 freight port terminals on Arkansas River.	78
Figure 4.3 Multimodal impact areas for freight port terminals located on the Arkansas Ri 2016	ver, 81
Figure 4.4 Size of multimodal and mode-specific impact areas, 2016	82
Figure 4.5 VMT and VHT of multimodal impact areas, 2016	82
Figure 4.6 Unique TAZs as origin or destination of multimodal trips to and from port terr 2016	ninals, 83
Figure 4.7 Number of registered business within the multimodal impact areas, 2016	84
Figure 4.8 Population within the multimodal impact areas, 2016	84
Figure 5.1 Locks	97
Figure 5.2 LPMS and truck GPS data fusion	102

Figure 5.3	Example of port area geographic bounding boxes 103
Figure 5.4	Multi-commodity assignment model schematics 106
Figure 5.5	Arkansas portion of the MKARNS 112
Figure 5.6	Model evaluation
Figure 6.1	Model schema
Figure 6.2	Study area
Figure 6.3	Ton-miles transported on MKARNS by commodity, 2016 142
Figure 6.4	Commodity flow on MKARNS detailed inland waterway network, 2016 143
Figure 6.5 volume of	Sensitivity of deterministic scenarios and stochastic model results to the maximum freight carried per trip
Figure A.1	WIM and ADTT stations within Arkansas

#### **CHAPTER 1. Introduction**

### 1.1. Multimodal Freight Planning

In 2018, over 5,200 billion ton-miles of freight were transported within the U.S. and by 2045, it is expected that over 7,600 billion ton-miles of freight will be shipped (Bureau of Transportation Statistics, 2020). The projected freight growth within the U.S. and its importance to the economy at a national, state, and regional level require an increased effort to improve freight data and to provide a sound basis for multimodal infrastructure asset management (in particular, for project prioritization). Federal legislation, namely the Moving Ahead for Progress in the 21<sup>st</sup> Century (MAP-21) and the Fixing America's Surface Transportation (FAST) Acts, facilitated the creation of statewide, multimodal transportation plans with freight components (Statewide Freight Plans, SFPs) (FHWA, 2013, 2017). Among others, SFPs are required to (Cornell Law School, 2015):

- identify significant freight system trends, needs and issues within the State, including bottlenecks, and congestion;
- describe freight policies, strategies, and performance measures that guide freight-related transportation investment decisions;
- list multimodal critical rural and urban freight corridors and facilities within the State;
- describe projected improvements to reduce or impede deterioration of multimodal transportation system infrastructure;
- incorporate a freight investment plan, including a list of key projects that meet the criteria to receive public funding for development.

Notably, a major criterion for a project to be eligible to receive federal funding is that it must be included in the SFP (FHWA, 2012). All 50 U.S. States have recently updated,

developed, or are in the process of developing an SFP. The successful implementation of SFPs is key to support infrastructure management through the estimation of freight performance measures, project prioritization, and guide multimodal planning (FHWA, 2017).

Even though the project assessment framework may differ by agency and state, in general, a typical project included in a SFP is subject to six key stages (Figure 1.1) (Chacon Hurtado et al., 2016).



Figure 1.1 Project assessment framework

The following sections synthesize freight planning assessment tools and their limitations to forecast multimodal freight needs, and describe broader impacts of this work in regard to multimodal freight infrastructure investment decision makings, in particular during the planning/strategy and prioritization stages.

## **1.2.** Planning Stage: Identification of Needs

Travel demand models (TDM) with freight components are typically used to estimate transportation infrastructure performance and needs at national, state, and metropolitan area level. A traditional TDM follows a sequential four-step approach: trip generation, trip distribution, mode choice and route assignment (Ortuzar et al., 2011). Each of these four steps constitutes a distinct mathematical sub-model and serves as input to subsequent steps. In most cases, TDMs have two components, a passenger model and a freight model, which are combined before route assignment (Alliance Transportation Group, 2015) (Figure 1.2).



Figure 1.2 Structure of 4-step TDM with commodity-based freight component

Freight components of TDMs were initially developed for truck transportation. At a later stage, answering applicable legislation that calls for multimodal long-range freight planning, alternative modes of transportation, such as rail, maritime, and air, were incorporated. As a result, state-of-the-practice TDMs often consist of imbalanced mode-specific networks. For example, the Arkansas State TDM (AR-STDM) incorporates abstract road and rail networks, but lacks representation of the waterway network, despite the key role that the Arkansas River plays to the state economy (Alliance Transportation Group, 2015; Nachtmann et al., 2015). Such imbalanced network representation limits the ability of TDMs to properly identify multimodal

bottlenecks, future infrastructure needs beyond those highways, and most importantly, the interaction between multiple modes that form critical supply chains. This work overcomes this limitation by creating a detailed navigable waterway network, and developing algorithms to map-match highly disaggregated vessel tracking data to it (Chapter 3). The result of this approach constitutes a dataset of trips assigned to an inland navigable waterway network, and characterized by length, duration, origin and destination, which may be used to integrate maritime freight activity into TDMs.

Traditional data sources for commodity-based 4-step TDMs are based on socio-economic activity census surveys, such as the Commodity Flow Survey (CFS), the Freight Analysis Framework (FAF), and Transearch, complemented by commodity-specific data such as the census of agriculture, and employment data from the Ministry of Labor. These sources are discussed later in Chapter 2, and limited by their relatively low collection frequency, spatial aggregations, commodity groups and modal representation, and cost. Thus, there is a need to explore the use of new data sources and their conflation potential to generate reliable data to input into TDMs. Focusing on waterways, vessel tracking data and publicly available commodity flow data are examples of sources explored in this work.

## 1.3. Commodity-based Freight Planning and Freight Fluidity

Building upon freight planning and modeling, a desired characteristic for freight forecasting tools such as TDMs is the ability to evaluate performance and forecast freight movement at the commodity level (Kam et al., 2017). By modeling at the commodity-level rather than the vehicle level, predictions of industry growth/decline driven by economic trends are directly connected to freight activity and transportation system performance. For example, commodity-based freight TDMs first estimate production and attraction of freight (in annual tons) within each zone of a larger region. Then, annual tons by commodity are distributed across zones to represent origins-destinations (OD) flows. Next, annual tons by OD pair are disaggregated by mode, generating OD matrices with annual tons of freight per mode, per commodity. Later, tons of freight are converted to number of vehicles by adopting payload factors (e.g. tons per truck, tons per rail carload), and those vehicles are assigned to the modeled transportation network (Alliance Transportation Group, 2012), provided a mode-specific network is modeled. The absence of a waterway network, and the unavailability of public, disaggregated commodity flow by water, constitute a limitation for the multimodal assignment of trips by commodity in TDMs. This limitation to identify commodity-flow mode-share by water presents an opportunity to explore and develop methods to derive inland waterway freight flows by commodity, at a sufficient level of spatial discrimination to support STDMs.

Moreover, given the complexity of the multimodal freight transportation system, there has been increased interest in developing multimodal "freight fluidity" indicators that capture end-to-end supply chain performance (Transportation Research Board, 2014). Freight fluidity measures require different types of data (e.g., movements, transactions, cost, commodity) from a variety of sources (e.g., government databases, private industry), and are intended to evaluate mobility, reliability, resiliency, cost, and quantity of freight along a multimodal transportation network (Eisele et al., 2016). For example, the FHWA National Freight Fluidity Monitoring Program combines waterborne data from the U.S. Army Corps of Engineers (USACE), railway data from TransCore and the Carload Waybill Sample, highway data from the National Performance Management Research Data Set (NPMRDS), and supply-chain data from U.S. private companies to generate a mapping tool to track the reliability, cost, and travel time (but not quantities) for multimodal freight movements across selected supply chains on a quarterly

basis (Parker, 2019). While freight fluidity has been implemented for international containerized supply chains, it is yet to be adapted to domestic transportation of bulk commodities involving inland waterways. Given the historical mode-specific approach in freight data collection and analysis, challenges remain to collect and analyze multimodal data for freight fluidity purposes, making the data fusion approaches developed in this dissertation timely and relevant (Transportation Research Board, 2018)

This work addresses the need to analyze commodity-based multimodal data for freight fluidity and TDM modeling purposes on inland waterways by developing a novel multicommodity assignment model (MCAM) solved via optimization, that fuses vehicle tracking and USACE's Lock Performance Monitoring System (LPMS) data to characterize and quantify highly-disaggregated freight flows on inland waterways. First, the MCAM fuses LPMS and truck Global Positioning System (GPS) data to output annual port throughput by commodity and mode on inland waterway port terminals. LPMS provides commodity flow aggregated at the lock-level, which is spatially disaggregated to port-terminals by observing the relative volume of trucks accessing each port from truck GPS data. To deal with the uncertainty associated to the sample that truck GPS data represents from the total truck population, relaxation of constraints to the MCAM optimization is introduced. The output of this model is the annual volume of freight transloaded between barge and rail, and between barge and truck by commodity (Chapter 5). In a second stage, the MCAM concept is used to fuse the port throughput by commodity (from Chapter 5) to the trips characterized by port of origin and destination for the same study area (from Chapter 3), resulting in the identification of volume and type of commodity carried by each vessel-trip assigned to an inland navigable waterway network (Chapter 6).

Knowledge of commodity-based port-level throughput, trip cargo characteristics, and linkage between waterborne and roadway freight flows supports the development of commodityspecific, multimodal freight fluidity performance measures, and may be used to prioritize transportation infrastructure investments.

#### **1.4.** Investment Evaluation and Prioritization of Needs

In order to match infrastructure supply with the demand for projected freight growth, continuous improvements to the multimodal transportation network and freight facilities are required. In this context, several projects compete for a limited amount of public and private funding. Only a portion of all the identified needs can be materialized at a time, and thus it is necessary to implement investment evaluation and prioritization measures which ensure a transparent and value-added expenditure of the resources available (Asborno and Hernandez, 2018).

From an analysis of the methods available to evaluate and prioritize projects (Economic Development Research Group et al., 2014), all rely in part on estimation of benefits relative to costs. The calculation of benefits necessitates a clear and consistent definition of the extent, location, and characteristics of a project's impact area (Chacon Hurtado et al., 2016; Weisbord et al., 2009). The impact area of a project affecting a multimodal freight facility can be defined as the region where the facility draws and delivers freight, or the connected origin-destination (OD) pairs served by the facility (Vadali et al., 2017).

In this context, arguably the most important opportunity for improvement is the lack of consistency in the data and procedures used to evaluate different projects affecting diverse modes, but subject to the same competition of funds. For example, planning agencies within Metropolitan Planning Organizations (MPOs) and state DoTs must use professional judgement

to define each project impact area (AASHTO, 2015) instead of following a systematic, datadriven procedure. Freight OD pairs may be obtained from project-specific data like stakeholder surveys or traffic counts, but those might not be consistent throughout diverse agencies; and/or from STDMs, provided all modes are represented with a similar level of detail, which usually are not (Alliance Transportation Group, 2015). Such lack of consistency in guidance, data, and tools to evaluate freight infrastructure investments across diverse geographies and modes potentially leads to a less-than-optimal allocation of funds.

This work adds value to the body of practice by developing data-driven methodologies that support project prioritization, such as the geospatial data fusion method to identify the impact area of multimodal freight projects using ubiquitous vehicle tracking data (Chapter 4). When compared to the state-of-the-practice, these novel prioritization tools have the advantage that all projects evaluated are subject to ubiquitous data and a systematic criteria to identify their impact, constituting a sound, common basis for proper comparison and competition of funds.

## 1.5. Research Objectives

The specific, intrinsically related objectives of this research are:

- to describe novel freight transportation and non-transportation data sources, emphasizing leveraging maritime Automatic Identification System (AIS) data, focusing on conflation potential;
- (2) to develop novel commodity-based multimodal data fusion models with the data examined in (1);
- (3) to estimate port throughput commodity flows from publicly available data with the broader impact of closing critical data gaps for inland waterway freight, by applying the models developed in (2);

- (4) to identify, characterize, and quantify commodity-based freight trips assigned to a detailed inland navigable waterway network, and
- (5) to develop a methodology to systematically identify data-driven, multimodal projectspecific freight catchment areas.

The research objectives highlighted above are in line with the marine transportation system priorities recommended by the U.S. Committee on the Marine Transportation System (CMTS), indicating the relevancy and timely of this dissertation. In particular, sample CMTS recommendations are: i) coordinate and apply big data analytics to reveal research gaps and overlap, foster potential collaboration, manage knowledge, and inform decision-making; ii) couple the newly-available vehicle probe data sets with more traditional freight data resources to quantify and contextualize travel times, dwell times, trip counts and other metrics; iii) create specific MTS system-scale performance indicators that relate to the freight flow network so they may be periodically updated and used for network calibration and validation; iv) develop and use decision support tools to identify nationally significant priority areas and project locations where agencies can leverage a variety of funding opportunities (U.S. Committee on the Marine Transportation System (CMTS), 2018).

## 1.6. References

- AASHTO. (2015). EconWorks Case Study Development Training. *Module 5 Case Study Data Needs and Sources*. Retrieved from https://planningtools.transportation.org/425/module-5---case-study-data-needs-and-sources.html
- Alliance Transportation Group. (2012). Arkansas Statewide Travel Demand Model: Freight Model Phase I.
- Alliance Transportation Group. (2015). Arkansas Statewide Travel Demand Model Documentation.

- Asborno, M.I., and Hernandez, S. (2018). Using Data from a State Travel Demand Model to Develop a Multi-Criteria Framework for Transload Facility Location Planning. *Transportation Research Record*, 2672, 9, 12-23. doi:10.1177/0361198118772699.
- Bureau of Transportation Statistics. (2020). *Pocket Guide to Transportation*. Retrieved from https://s3-us-west-2.amazonaws.com/dot-concept-menus/menu/accordion.html
- Chacon Hurtado, D., Bardaka, E., Yang, R., et al. (2016, June 26-29). Transportation and Economic Development: An Assessment of the SHRP2 C11 Tools and Their Synergies with Economic Simulation Models for the Evaluation of Wider Economic Benefits of Transportation Projects. ASCE International Conference on Transportation & Development.
- Cornell Law School. (2015). U.S. Code Ttle 49, Sbtitle IX, Chapter 702, Section 70202 State Freight Plans. Retrieved from https://www.law.cornell.edu/uscode/text/49/70202
- Economic Development Research Group, Cambridge Systematics, ICF International, et al. (2014). *Development of Tools for Assessing Wider Economic Benefits of Transportation* (978-0-309-27360-2). Retrieved from http://www.trb.org/Main/Blurbs/169524.aspx
- Eisele, W., Juster, R., Sadabadi, K., et al. (2016). Implementing Freight Fluidity in the State of Maryland. *Transportation Research Record*, 2548, 1, 62-70. doi:10.3141/2548-08
- FHWA. (2012). Federal-aid Essentials for Local Public Agencies. Funding Basics and Eligibility - Companion resource. Retrieved from https://www.fhwa.dot.gov/federalaidessentials/companionresources/44funding.pdf
- FHWA (Producer). (2013, September 12). Significant Freight Provisions. Moving Ahead for Progress in the 21st Century Act. Retrieved from www.fhwa.dot.gov/map21/factsheets/freight.cfm
- FHWA (Producer). (2017, February 8). Freight Planning and Policy. *Fixing America's Surface Transportation Act*. Retrieved from www.fhwa.dot.gov/fastact/factsheets/fpppfs.cfm
- Kam, K., Jiang, N., Bujanovic, P., Savage, K., Walthall R., Seedah, D., & Walton, M. (2017). Finding and Exploring Use of Commodity-Specific Data Sources for Commodity Flow Modeling. *Transportation Research Record*, 2646, 1, 77-83. doi:10.3141/2646-09.
- Nachtmann, H., Boudhoum, O., & Oztanriseven, F. (2015). *Regional Economic Impact Study for the McClellan-Kerr Arkansas River Navigation System*. Retrieved from https://martrec.uark.edu/research/regional-economic-impact-study-for-the-mcclellankerr-arkansas-river-navigation-system.pdf
- Ortúzar, J. d. D., & Willumnsen, L. G. (2011). *Modelling Transport, 4th Edition*. West Sussex, United Kingdon: Wiley.

- Transportation Research Board. (2014). Developing Freight Fluidity Performance Measures (Circular E-C187), Washington, DC. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec187.pdf
- Transportation Research Board. (2018). *Workshop on Implementing a Freight Fluidity Performance Measurement System (Circular E-C240)*, Washington, DC. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec240.pdf
- U.S. Committee on the Marine Transportation System (CMTS). (2020, May 6). *MTS Priorities* and CMTS Recommendations. Retrived from https://www.cmts.gov/about/priorities\_and\_recommendations
- Vadali, S., Kruse, C. J., Kuhn, K., et al. (2017). NCFRP Report 38: Guide for Conducting Benefit-Cost Analyses of Multimodal, Multijurisdictional Freight Corridor Investments. Retrieved from http://www.trb.org/Publications/Blurbs/175606.aspx
- Weisbord, G., & Reno, A. (2009). *Economic Impact of Public Transportation Investment*. Retrieved from http://onlinepubs.trb.org/onlinepubs/tcrp/docs/TCRPJ-11Task7-FR.pdf

#### **CHAPTER 2. Background**

The literature review presented below synthetizes the state-of-the-practice in terms of multimodal freight data used for long-range transportation planning, contextualizing its role in performance evaluation and investment prioritization. Next, a general introduction to data fusion is presented, closing with a list of data sources conflated in this work. Each of the following chapters introduce specific novel data fusion techniques, and elaborate on the background pertaining to specific fusion methods and datasets.

## 2.1. Data Sources for Long-Range Freight Planning

Freight planners have expressed their concern about the lack of publicly available freight data (Cambridge Systematics and GeoStats, 2010). Robust freight data is produced by private sector's logistics technologies and sensors, but there are several barriers that difficult to effectively share it among private and public sectors. Examples of those barriers are privacy laws, lack of resources for data processing, competitiveness and confidentiality concerns, institutional and coordination complexity, etc. (Cambridge Systematics et al., 2013). Another obstacle to gather business data resides in the difficulty to identify who makes transportation decisions (Ortúzar et al., 2011), given the number of players involved in the supply chain distribution (senders/consignees, freight forwarders, operators/carriers, insurance companies, etc.). For these reasons, the development of long-range freight modeling using public datasets is one of the main topics identified in NCFRP Report 8 (Cambridge Systematics et al., 2010) for additional research.

In the U.S., commodity-based freight models typically use data from the following sources: Transearch, Commodity Flow Survey (CFS), and Freight Analysis Framework (FAF). These may be complemented by local or regional surveys (Cambridge Systematics et al., 2008; Tatineni et al., 2005). The following paragraphs briefly describe each of these data tools, and compares the three databases discussed above: Transearch, CFS and FAF.

#### 2.1.1. Transearch

Transearch is a proprietary database of U.S. annual county-level freight flow data, by commodity, produced by IHS Global Insight (IHS, former Reebie Associates). More than 340 commodities are included in the database, classified by Standard Transportation Commodity Classification (STCC) 4-digit codes. From a geographic perspective, Transearch covers over 3,000 counties within 172 Bureau of Economic Analysis (BEA) regions in the U.S., and international regions. Freight flow volumes by geography and commodity are presented in tons, and translated to: shipment units (such as carloads and truck counts), vehicle miles travelled, shipment values, and ton-miles. This database considers seven major transportation modes, namely: for-hire truck, less-than-load truck, private truck, truck/rail intermodal, rail, waterborne, and air (IHS / Global Insight).

The methodology implemented by IHS to produce Transearch freight flows consists of four steps: First, the value of production and consumption of each commodity (disaggregated into North American Industry Classification System (NAICS) groups) is estimated at county level. These estimates are based on a combination of the BEA's input/output tables, and IHS's Business Market Insight (BMI) sales information. BMI does not include all commodities. Thus, alternative sources are used for agricultural products and livestock (U.S. Department of Agriculture), automobiles and coal (IHS in-house databases), and minerals (U.S. Geological Survey). Then, NAICS commodities are re-classified as per STCC 4-digit codes and a price per ton is used to translate each commodities' monetary values to tonnage. Next, commodity flows are classified into domestic, import and export by using port-level census data. Later, modal

freight flows of domestic county-to-county movements are developed for railroad, waterborne, and air cargo, per commodity (origin-destination known pairings). Lastly, truck flows are calculated, by subtracting the origin-destination known pairings from the total productions and attractions in each county. The methodology presented above is used to estimate base year freight flows. Future freight flow estimations are also provided. Forecasts are based on projections of supply and demand at county level, by 4-digit STCC commodity type, which are further constrained to a national total, for consistency. IHS proprietary services are leveraged as input to the forecasting process, such as IHS' U.S. Macroeconomic service, IHS' U.S. Agricultural service, Energy Service, etc. (IHS / Global Insight, 2011).

Transearch includes waterborne shipments, derived from state-to-state annual flows of broad commodity groups published by USACE, and disaggregated using proprietary methods. Although drayage for marine ports is captured in Transearch, drayage for inland ports is not captured (IHS / Global Insight, 2011). The work presented in this dissertation allows for the identification of truck drayage for inland waterway ports (Chapter 4).

A notable technical limitation of Transearch is its way of handling shipments made by trucks that exceed the limits for a state, specifically in the way that shipments to distribution centers are recorded. Transearch records the first portion of the trip (from the origin to a distribution center) in the National database, while the second portion of the trip (from the distribution center to its destination) is recorded as an individual movement in the State database (Alliance Transportation Group, 2012). Thus, it is not possible to identify the actual origin and destination for some commodity flows. In terms of inland waterways transportation, tugs may pick up loaded barges at a port, then "park" loaded barges at anchoring grounds, to be picked up later (possibly by another tug) to reach its final destination. Thus, anchoring grounds play a

similar role than distribution centers in Transearch. To overcome the difficulty of associating both legs of the freight flow (before and after an anchoring ground), and identify the true origin and destination of freight, this dissertation uses highly disaggregated maritime vessel tracking data to generate "trip chains", that capture the true origin and destination of freight with intermediate stops on anchoring grounds (Chapter 3).

Moreover, Transearch reports all movements made from distribution centers as a unique STCC commodity code, e.g., Secondary Traffic (Alliance Transportation Group, 2012), masking the actual commodity transported. In addition, from a commodity classification perspective, some commodities are not reported in Transearch, such as construction, retail, refined petroleum, municipal solid waste and farm-based agriculture shipments (Alliance Transportation Group, 2012). Lastly, Transearch presents other modal limitations, related with: i) the lack of international air shipments (HIS / Global Insight, 2011), ii) pipeline mode is not included (Beagan et al., 2007), and iii) incomplete and inconsistent information may be provided for multi-modal trips (Cambridge Systematics et al., 2010). On another note, when comparing to CFS and FAF, Transearch has cost and transparency limitations due to its proprietary nature. In 2010, data for a single year of a single state could cost between \$50,000 and \$100,000 (Cambridge Systematics et al., 2010). The mechanics and models used to produce estimates are proprietary, functioning as a "black box" where users do not know what happens inside. Despite its limitations, in the absence of other spatially disaggregated commodity-based datasets, Transearch is widely used for freight planning purposes by the FHWA, U.S. States, Metropolitan Planning Organizations (MPOs), and private freight carriers and shippers (Cambridge Systematics et al., 2008; Asborno and Hernandez, 2018).

### 2.1.2. Commodity Flow Survey

While Transearch is a proprietary database, CFS and FAF are publicly available. The Commodity Flow Survey (CFS) is a shipper survey of goods transported from establishments in the U.S. that provides key information about each shipment. CFS is based on a probability sample of all U.S. shipments. Among the twenty data attributes of each observation, it includes origin state, destination state, mode of transportation, shipment weight, value, commodity (as per NAICS, 3-digit classification), distance routed, whether the shipment is for export, whether it contains hazardous substances, if it is temperature-controlled, and if it is rush. The survey was initiated in 1993 and is currently conducted every 5 years, in a combined effort by the U.S. Census Bureau and the U.S. Department of Transportation. There is an approximate three-year delay between data collection and public release of aggregated data, e.g., the last collection is from 2017, but final data tables will be released in July 2020 (U.S. Census Bureau, 2019a). The latest public use microdata file available is from 2012 and it includes over 4 million records; however only 75 records correspond to shipments by water to and from Arkansas. Modes considered in this database are: truck, rail, water, air, and pipeline (single modes), and parcel, truck and rail, truck and water, rail and water (multiple modes). The industry sectors covered by this survey are manufacturing, mining, wholesale, electronic shopping, mail order, fuel dealers and publishing industries (U.S. Census Bureau, 2016a). Notably, agriculture is not included in the sample frame. Establishments classified in transportation, construction, and most retail and services industries are also excluded from the survey, as well as farms, fisheries, foreign establishments, and most government-owned establishments. In total, 43 commodities are included in the database (in NAICS, 3-digit classification) (Bureau of Transportation Statistics, 2015b). Geographically, CFS divides the 50 U.S. States and the District of Columbia territory in

132 areas, classified as metropolitan areas and "reminder of State" areas. Each metropolitan area can be comprised by portions of more than one state, as it is the case of Chicago Combined Statistical Area, spanning through Illinois, Indiana and Wisconsin. CFS does not cover shipments originating from establishments in Puerto Rico (U.S. Census Bureau, 2017).

The CFS survey method consists of a comprehensive questionnaire mailed to more than 100,000 establishments from a sample frame of over 716,000, listed in the U.S. Census Bureau's Business Register. The sample is selected using a stratified three-stage design (Table 2.1) (U.S. Census Bureau, 2016b). The stratification criteria, applied to the first design stage, is based on the establishments' location (geography), industry, and size, measured in terms of number of employees and sales. Auxiliary establishments (truck transportation facilities, warehouses, and central administrative offices) with shipping activity are included on the sampling frame.

Design Stage	Sampling Units	Sample Frame	Sample Size
1 <sup>st</sup> Stage	Establishments from the Census Bureau's Business Register.	716,000	102,565
2 <sup>nd</sup> Stage	Groups of four one-week periods (reporting weeks) within the survey year.	52 weeks	4 weeks/year, one in each quarter.
3 <sup>rd</sup> Stage	Shipments.	Total number of shipments per week	40 shipments or less (if total is less than 40)

 Table 2.1 Commodity Flow Survey - Sample Design

One of the main limitations of CFS is its untimeliness. CFS untimeliness refers to the low frequency when the survey is collected (i.e. every five years), and the long time it takes to publish the data (three years). Because of this untimeliness, the effects of an event that occurs and fully recovers in-between successive collections are missed. In case an event is captured by CFS, it will take three years for researchers and practitioners to know the effects of such events from CFS data. Thus, this lack of agility in handling rapidly changing operations makes CFS unsuitable to handle several effects, like the COVID-19 pandemic. In this context, the use of ubiquitous and continuous vehicle tracking data present an alternative to overcome the untimeliness limitation.

#### 2.1.3. Freight Analysis Framework

The Freight Analysis Framework (FAF) is a freight flow data tool produced by the Bureau of Transportation Statistics (BTS) in partnership with the Federal Highway Administration (FHWA). In its latest version (FAF4), FAF takes the 2012 CFS database and complements it with other public data sources, namely: the Census Foreign Trade Statistics, Economic Census data, USDA's Census of Agriculture, Port Import/Export Reporting Service (PIERS), Vehicle Inventory and Use Survey (VIUS), National Highway Planning Network (NHPN), Highway Performance Monitoring System (HPMS), and U.S. Energy Information Administration (EIA). FAF4 provides a comprehensive set of estimated annual freight flows for 2012-2018 period, plus long-term forecast scenarios. Freight flows are expressed in weight, weight-distance, and value, and can be disaggregated by geography area, commodity, mode, and whether they are domestic, export, or import. Geographically, FAF4 considers the same 132 domestic areas as the CFS, plus eight international regions (Bureau of Transportation Statistics, 2015a). The transportation modes include: truck, rail, water, air (including truck-air), multiple modes and mail, pipeline, other and unknown, and "no domestic mode". The latter refers to shipments that have an international mode but no domestic mode, addressing crude petroleum imports from inbound ships that enter directly to a refinery in the U.S. In terms of commodities, FAF4 considers 44 groups, following the Standard Classification of Transported Goods (SCTG) 2-digit classification. FAF4 has a forecasting tool and further assigns freight flows to the

Highway Performance Monitoring System (HPMS) network. Data can be filtered and downloaded as Microsoft Access and comma separated values (.csv) files.

The main limitation of FAF is its lack of geographic detail, not sufficiently refined to be consistent with the full array of many transportation agency's freight planning applications (Cambridge Systematics, 2013). For example, Arkansas in the FAF is represented as a single zone.

Data source / Characteristics	TRANSEARCH	CFS	FAF
Relevant data values provided	Weight, value and amount of shipments by geography and commodity; classified as outbound, inbound, intra and through shipments.	For each shipment: origin state, destination state, mode of transportation, shipment weight, value, commodity, transportation distance, whether the shipment is for export, whether it contains hazardous substances, if it is temperature- controlled, if it is rush.	Weigh, weigh-distance and value of freight flow by geography, commodity, and mode; classified in import, export, and domestic.
Geographical coverage	3,000+ counties 172 BEA regions U.S., Mexico, Canada	132 areas within U.S.	132 areas within U.S. and 8 international regions.
Modes included	for-hire truck less-than-load truck private truck truck/rail intermodal rail waterborne air	Single Modes: Truck (for hire, private), Rail, Water (inland water, great lakes, deep sea, multiple waterways), Air (including truck and air), Pipeline. Multiple Modes: Parcel, Truck and Rail, Truck and Rail, Truck and Water, Rail and Water, Other multiple modes. Other Modes.	truck, rail, water, air (including truck-air), multiple modes and mail, pipeline, other and unknown, and "no domestic mode".
Commodities considered	340+ commodities, STCC 4-digit codes	47 commodity groups, as per NAICS, 3-digit classification.	44 commodity groups, as per SCTG 2-digit classification
Update frequency	Annual	Every 5 years	Annual estimates
Data cost	\$50,000 - \$100,000 per state, per year (NCFRP Report 8).	free	free

# Table 2.2 Comparison of Transearch, CFS, and FAF

Data source / Characteristics	Transearch	CFS	FAF
Primary data collection sources	Partnership with main U.S. truck and rail (class I) carriers, Railroad Waybill Sample, U.S. Army Corps of Engineers, Airport Activity Statistics (BTS), CFS.	Survey to over 100,000 business establishments within U.S.	CFS Census Foreign Trade Statistics, Economic Census data, USDA's Census of Agriculture, Port Import/Export Reporting Service (PIERS), Vehicle Inventory and Use Survey (VIUS), National Highway Planning Network (NHPN), Highway Performance Monitoring System (HPMS), and U.S. Energy Information Administration (EIA)
Main limitations	Inability to track origin and destination of some freight flows (truck and multimodal). Lack of reporting of certain commodities. Multimodal shipments reported as main single mode.	Gaps in industry and commodity coverage (Agriculture, crude petroleum extraction). Lack of geographic detail. Lack of international flows.	Lack of geographic detail.

# Table 2.2 Comparison of Transearch, CFS, and FAF (cont.)

#### 2.2. Data Fusion

Fusion of multiple datasets has been used to overcome the limitations imposed by traditional sources, which may not be sufficiently detailed or accurate for specific commodities and spatial resolutions (Ahanotu et al., 2003) (Vieira da Silva & Almeida D'Agosto, 2013). For example, Kam et al. (2017) modeled the transportation of sorghum and corn grains from farms to grain elevators, and lastly to feed yards. They fused data on: (a) acreage planted, harvested, yield per acre, and production (bushels) by county, from the NASS Southern Plains Regional Field Office; (b) grain elevator locations and capacity, from BNSF elevator directory; and feed yard demand, derived form (c) an inventory of hogs and pigs provided by the 2012 USDA census of agriculture, and (d) cattle permit database from the Concentrated Animal Feeding Operations, Texas Commission on Environmental Quality.

Freight transportation data gaps prevent the development of a network-based freight demand model that incorporate all modes at similar levels of detail for various geographies (Schaefer, 2017). The combination of datasets is a necessary approach to solve existing data gaps, while avoiding costs associated with the development and implementation of expensive data collection techniques. However, the combination of several datasets is challenging. The main challenges occur because each data development entity follows different procedures to define, collect, process and share the data (Tok et al., 2011).

Data heterogeneity can be classified as: taxonomic (different definition for the same term), temporal (such as changes in data collection methodology over time), or methodological (for example, commodity data reported in different units across datasets) (Walton et al., 2015). Resolving heterogeneity is necessary to link data across levels of geography, topics and modes (Walton et al., 2015). In addition, the unavailability of metadata and/or data dictionaries leaves
room for user's interpretations, which may be inaccurate or incorrect. To overcome these limitations, data architectures are proposed. For example, Quiroga et al. (2011) emphasize the need to develop a list of components, to establish data-integration points, to conduct a data gap and disaggregation need analysis, and to use standardized terminology when developing a data architecture (Quiroga et al., 2011).

Building upon Quiroga's work, NCFRP Report 35 (2015) implemented the Freight Transportation Data Architecture to create a Data Element Dictionary (Walton et al., 2015). To resolve differences in data element's definitions, "bridging" (cross-walk) tables were introduced. Walton applied his proposed approach to a case study in Texas (Walton, 2013).

Moreover, Tok et al. (2011) created Cal-FRED, a user-centered, online freight datarepository tool, to gather publicly available sources of freight data for state planning and analysis. The authors' main contribution is the practical development of a standardized freight data architecture, a standardized data-quality assessment, and a physical architecture (Tok et al., 2011).

### **2.3.** Data Sources Fused in This Work

Several existing U.S. commodity databases commonly used for freight planning (CFS, FAF, Transearch) have potential for conflation with new and emerging sources to support multimodal project prioritization with focus on inland navigable waterways. This section presents the data sources incorporated into the methodologies presented in this dissertation (Table 2.3). For completeness, a synthesis of multimodal data sources, several outside the usual purview of freight data, is presented in Appendix A.

Data type	<b>Brief Description</b>	Entity	Reference	Chapter
Commodity flow data	LPMS Waterborne monthly commodity data, 2016.	USACE	(U.S. Army Corps of Engineers)	5;6
Freight vehicle tracking data	Statewide Truck GPS data, 2016.	ATRI	(American Transportation Research Institute, 2019)	4; 5; 6
Freight vehicle tracking data	Waterborne AIS data (timestamped geospatial vessel locations), 2016	U.S. Guard & USACE	(Office for Coastal Management, 2018)	3; 4; 6
Freight vehicle count by location	LPMS: number of commercial vessels per lock, 2016	USACE	(U.S. Army Corps of Engineers, 2018)	3
Infrastructure map layers	Roadway system: Statewide network line layer based on Arnold maps	Arkansas Department of Transportation (ARDOT) and Arkansas GIS Office	(Arkansas Department of Transportation)	4
Infrastructure map layers	Waterway system: i) Navigable waterway lines layer, ii) Locks, and iii) Ports point geospatial layers from the National Transportation Atlas Database (NTAD)	Geospatial at the Bureau of Transportation Statistics, U.S. Department of Transportation	(Bureau of Transportation Statistics, 2018, 2019a, 2019b)	3; 4; 5; 6
Infrastructure map layers	Location of port terminals in Arkansas, geospatial map layer	Arkansas Economic Development GIS office	(Arkansas Economic Development Commission, 2018)	3; 4; 5
Business	County Business patterns, 2016	U.S. Census Bureau	(U.S. Census Bureau, 2018)	4
Geopolitical boundaries	Arkansas state and county boundaries polygon geospatial map layer	Arkansas GIS Office	(Arkansas GIS Office, 2016)	3
Geopolitical boundaries	Arkansas TIGER/Line® Shapefiles: Census Tracts	U.S. Census Bureau	(U.S. Census Bureau, 2019b)	4
Geopolitical boundaries	Traffic Analysis Zones polygon layer	Arkansas Department of Transportation	AR-STDM zone layer (Alliance Transportation Group, 2015)	3; 4
Aerial imagery	Historical aerial imagery of the McClellan Kerr- Arkansas River Navigation System (MKARNS)	Google	(Google, 2020)	3; 4; 5

# Table 2.3 Data Sources Fused in This Work

# 2.4. References

- Ahanotu, D., Fischer, M., & Louch, H. (2003). Developing a Commodity Flow Database from Transearch Data. *Transportation Research Record*, 1955, 1, 14-21. doi:10.3141/1855-02
- Alliance Transportation Group. (2012). Arkansas Statewide Travel Demand Model Phase I: Freight Model.
- Alliance Transportation Group. (2015). Arkansas Statewide Travel Demand Model Documentation.
- American Transportation Research Institute. (2019). ATRI. Retrieved from https://truckingresearch.org/
- Arkansas Department of Transportation. ArcGIS REST Services Directory. Retrieved from https://gis.arkansas.gov/arcgis/rest/services/FEATURESERVICES/Transportation/Featur eServer/9
- Arkansas Economic Development Commission. (2018). Ports. Retrieved from https://www.arkansasedc.com/docs/default-source/map-room/waterwayportmap\_08202018.pdf?sfvrsn=5a351d51\_0
- Arkansas GIS Office. (2016). County Boundary (polygon). Retrieved from https://gis.arkansas.gov/product/county-boundary-polygon/
- Asborno, M.I., & Hernandez, S. (2018). Using Data from a State Travel Demand Model to Develop a Multi-criteria Framework for Transload Facility Location Planning. *Transportation Research Record*, 2672, 9, 12-23. doi:10.1177/0361198118772699
- Beagan, D., Fischer, M., & Kuppam, A. (2007). *Quick Response Freight Manual II*. Retrieved from http://sites.poli.usp.br/d/ptr2377/Qrfm-ii-FHWA-HOP-08-010.pdf
- Bureau of Transportation Statistics. (2015a). Freight Analysis Framework Version 4. User's Guide for Release 4.0. Retrieved from https://www.bts.gov/archive/subject\_areas/freight\_transportation/faf/users\_guide/
- Bureau of Transportation Statistics. (2015b). United States: 2012 Economic Census. Transportation. 2012 Commodity Flow Survey (EC12TCF-US). Retrieved from https://www.census.gov/content/dam/Census/library/publications/2015/econ/ec12tcfus.pdf
- Bureau of Transportation Statistics. (2018). National Transportation Atlas Database Locks. Retrieved from https://data-usdot.opendata.arcgis.com/datasets/locks
- Bureau of Transportation Statistics. (2019a). National Transportation Atlas Database Navigable Waterway Lines. Retrieved from https://datausdot.opendata.arcgis.com/datasets/navigable-waterway-lines

- Bureau of Transportation Statistics. (2019b). National Transportation Atlas Database Ports. Retrieved from https://data-usdot.opendata.arcgis.com/datasets/ports
- Cambridge Systematics. (2013). NCFRP Report 26: Guidebook for Developing Subnational Commodity Flow Data. Retrieved from https://www.nap.edu/catalog/22523/guidebookfor-developing-subnational-commodity-flow-data
- Cambridge Systematics, & GeoStats. (2010). NCFRP Report 8: Freight-Demand Modeling to Support Public-Sector Decision Making. Retrieved from https://www.nap.edu/catalog/14445/freight-demand-modeling-to-support-public-sectordecision-making
- Cambridge Systematics, Global Insight, Cohen, H., et al. (2008). NCHRP Report 606: Forecasting Statewide Freight Toolkit. Retrieved from https://www.nap.edu/read/14133/chapter/1#ii
- Cambridge Systematics, North River Consulting Group, & University of Washington. (2013). *NCFRP Report 25: Freight Data Sharing Guidebook*. Retrieved from https://www.nap.edu/catalog/22569/freight-data-sharing-guidebook
- Google. (2020). Google Earth. Retrieved from https://www.google.com/earth/
- IHS / Global Insight. Transearch Product Sheet. Retrieved from http://www.ihsglobalinsight.com/gcpath/Transearch\_Product\_Sheet.pdf
- IHS / Global Insight. (2011). OKI Regional Freight Plan. Appendix A: IHS/Global Insight TRANSEARCH Commodity Flow Report. Commodity Flows in the Cincinnati Region. . Retrieved from https://www.oki.org/wp-content/uploads/2015/11/Appendix-A-Revised-OKI-Commodity-Flow-Data-Nov-2011.pdf
- Kam, K., Jiang, N., Bujanovic, P., Savage, K., Walthall R., Seedah, D., & Walton, M. (2017). Finding and Exploring Use of Commodity-specific Data Sources for Commodity Flow Modeling. *Transportation Research Record*, 2646, 1, 77-83. doi:10.3141/2646-09.
- Office for Coastal Management. (2018). 2016 Nationwide Automatic Identification System. https://inport.nmfs.noaa.gov/inport/item/52963
- Ortúzar, J. d. D., & Willumnsen, L. G. (2011). *Modelling Transport, 4th Edition*. West Sussex, United Kingdon: Wiley.
- Quiroga, C., Koncz, N., Kraus, E., et al. (2011). *NCFRP Report 9: Guidance for Developing a Freight Transportation Data Architecture*. Retrieved from http://www.trb.org/Publications/Blurbs/164644.aspx
- Schaefer, R., Worth, M., Heilman, J., & Kehoe, N. (2017). Freight Demand Modeling and Data Improvement Implementation Handbook. Retrieved from https://ops.fhwa.dot.gov/publications/fhwahop18018/fhwahop18018.pdf

- Tatineni, V. C., & Demetsky, M. J. (2005). Supply Chain Models for Freight Transportation *Planning*. Charlottesville, VA.
- Tok, A., Zhao, M., Chow, J., et al. (2011). An On-line Data Repository for Statewide Freight Planning and Analysis. *Transportation Research Record*, 2246, 121-129. doi: https://doi.org/10.3141/2246-15
- U.S. Army Corps of Engineers (Producer). Lock Performance Monitoring System. *Monthly Tons Report*. Retrieved from https://corpslocks.usace.army.mil/lpwb/f?p=121:6:15549769738061::NO:::
- U.S. Army Corps of Engineers. (2018). USACE Digital Library Public lock reports locks by waterway, lock usage, CY 1993 2017. Retrieved from https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/2958
- U.S. Census Bureau. About the Survey. *Commodity Flow Survey*. Retrieved from https://www.census.gov/econ/cfs/about\_the\_survey.html#geography
- U.S. Census Bureau. (2016a). 2012 Commodity Flow Survey (CFS) Public Use Microdata (PUM) File. Data Users Guide. Retrieved from https://www2.census.gov/programs-surveys/cfs/datasets/2012/2012-pums-files/cfs-2012-pumf-users-guidefeb2016.pdf
- U.S. Census Bureau. (2016b). 2012 Commodity Flow Survey Methodology. *Commodity Flow Survey*. Retrieved from https://www.census.gov//econ/cfs/2012\_methodology.html
- U.S. Census Bureau. (2017). Commodity Flow Survey. Retrieved from https://www.census.gov/econ/cfs/
- U.S. Census Bureau. (2018). County Business Patterns: 2016. https://www.census.gov/data/datasets/2016/econ/cbp/2016-cbp.html
- U.S. Census Bureau. (2019a). Commodity Flow Survey. Retrieved from https://www.census.gov/programs-surveys/cfs.html
- U.S. Census Bureau. (2019b). Tiger/Line Shapefiles: Census Tracts. https://www.census.gov/cgibin/geo/shapefiles/index.php?year=2019&layergroup=Census+Tracts
- Vieira da Silva, M., & Almeida D'Agosto, M. (2013). A Model to Estimate the Origin– Destination Matrix for Soybean Exportation in Brazil. *Journal of Transport Geography*, 26, 97-107, https://doi.org/10.1016/j.jtrangeo.2012.08.011.
- Walton, C. M. (2013). Integration of Data Sources to Optimize Freight Transportation in Texas. Austin, TX. Retrieved from https://library.ctr.utexas.edu/ctr-publications/psr/0-6697-ctrs.pdf

Walton, C. M., Seedah, D. P. K., Choubassi, C., et al. (2015). *NCFRP Report 35: Implementing the Freight Transportation Data Architecture: Data Element Dictionary*. Retrieved from: http://www.trb.org/NCFRP/Blurbs/173083.aspx

# CHAPTER 3. Network Mapping of AIS data to Characterize Inland Waterway Freight Transportation

# 3.1. Abstract

To support freight growth, Travel Demand Models (TDM) with freight forecasts are employed to estimate performance metrics for competing freight infrastructure investments and policy changes. Unfortunately, freight TDMs, initially developed for highway assessment, fail to represent non-truck modes with levels of detail adequate for multimodal infrastructure and policy evaluation. Expanded public availability of maritime freight movement data introduces strong potential to expand representation of marine modes within freight TDMs. This paper focuses on a key example, the Automatic Identification System (AIS) data which tracks vessel locations as timestamped latitude-longitude points (pings) along waterways. For viable estimation, calibration, and validation of freight TDMs, AIS data must be mapped to a representative network and contain trip-level travel patterns. This work develops a detailed inland waterway transportation network and identifies vessel trips by applying network mapping (map-matching) heuristics to AIS data. Map-matching produces complete paths between stops from vessel ping data as series of connected links. Uniquely, the stop identification procedure estimates parameters to distinguish freight stops at ports from delays through locks or prestaging/anchoring areas. The methods are evaluated on 747 miles of inland waterways in Arkansas, with AIS data representing 88% of vessel activity. Manual inspection of 3,820 AIS trajectories was used to train the heuristic parameters including stop time, duration, and location. Validation results show 84.0% accuracy in detecting stops at ports and 83.5% accuracy in identifying trips crossing locks. A single set of parameters does not fit best all vessels, possibly

explaining the less-than-perfect algorithm accuracy. Since AIS data is ubiquitous in time and space, the proposed methods are transferable to any region with waterways.

Key words: Inland Waterway Transportation (IWT), Automatic Identification System (AIS), map-matching algorithm, Geographic Identification Systems (GIS), Freight.

### **3.2.** Introduction

18.6 billion tons of goods valued at 18 trillion USD were moved in 2018 along the U.S. multimodal transportation system. This is equivalent to more than 7 million trucks, 1.8 million carloads, or 124 thousand barges (FHWA, 2019). To support expected freight grow of 51% by 2045 (FHWA, 2019), federal legislation calls for multimodal freight planning- a significant distinction from the sole pursuit of highway-oriented freight planning (FHWA, 2017). In this context, it is imperative to identify future infrastructure needs for highway, rail, water, air and pipeline. Travel demand models (TDM) with freight forecasts are common tools to identify and prioritize transportation infrastructure needs by estimating performance metrics for demand, policy, and capacity scenarios. However, because freight components of TDMs were initially developed for truck transportation, they often lack the level of detail needed to evaluate multimodal freight performance metrics like freight fluidity. For example, the Arkansas State TDM incorporates road and rail networks, but lacks a waterway network (Alliance Transportation Group, 2015), despite the key role of the Arkansas River in the state economy (Nachtmann et al., 2015). Such imbalanced network representation limits the ability of TDMs to estimate freight fluidity metrics, identify bottlenecks across multimodal supply chains, and support infrastructure planning for multimodal facilities like ports. In the absence of a detailed waterway network, state-of-the-practice freight TDMs cannot assign number of vessels per draft and cargo to the network, preventing a true multimodal comparison of capacity upgrade needs

and benefits among roadways (by the addition of travel lanes) and inland waterways (by dredging).

Due to privacy and confidentiality concerns, data required for freight TDMs such as spatially disaggregated origin-destination (OD) flows and trip characteristics like commodity carried are limited for all modes and especially underrepresented for inland waterways (FHWA, 2017b). Along inland waterways, freight flows distinguished by commodity and port OD are not publicly available. The U.S. Army Corps of Engineers (USACE) collects detailed domestic waterborne traffic movements, which are mandatorily reported by all vessel operators. For cargo movements, the point of loading and unloading of each commodity is reported (U.S. Army Corps of Engineers, 2018). However, such detailed data is reserved for use by collecting agencies like the USACE while a summarized version is shared via the Waterborne Commerce of the United States (WCUS). WCUS provides statistics on foreign and domestic commerce along U.S. waterways (U.S. Army Corps of Engineers, 2016). The Manuscript Cargo and Trip File provides movements of commodities at certain ports, harbors, and inland waterways in the U.S., including the annual number of trips reported per port and waterway, by direction, vessel type, and draft (U.S. Army Corps of Engineers, 2018). Focusing on data requirements for long-range freight planning (e.g., OD volumes/tonnages and trip characteristics), the Manuscript Cargo and Trip File is limited in: i) its spatial aggregation, (e.g., it includes only three ports from the more than 40 freight port terminals in Arkansas); and ii) it is based on manually entered reports, which may contain errors. This work overcomes such limitations by developing a data-driven, reproducible map-matching methodology to identify trips on inland waterways based on data collected automatically (i.e. not prone to human errors), between ODs located at any node along a detailed inland waterway network.

The recent open availability of historical Automatic Identification System (AIS) data, which tracks vessels' location and timestamp, is a promising source to model maritime freight flows. This work develops a detailed inland waterway transportation network and identifies vessel trips by applying network mapping (map-matching) heuristics to AIS trajectory data. Map-matching produces complete paths between stops from vessel ping data as series of connected links. Uniquely, the stop identification procedure contained in the heuristic estimates parameters to distinguish freight stops at ports from delays through locks or prestaging/anchoring areas. The methods are evaluated with AIS data for 747 miles of navigable waterways in Arkansas, including the McClellan Kerr-Arkansas River Navigation System (MKARNS) and the portion of the Mississippi River along Arkansas' eastern border. Since AIS data is ubiquitous in time and space, the proposed methods are transferable to any region with waterways.

# 3.3. Background

### 3.3.1. Multimodal Freight Planning and Travel Demand Models

The projected U.S. freight growth and its importance to the economy require increased effort to improve multimodal freight demand forecasting tools (FHWA, 2017). TDMs identify future system deficiencies based on forecasted activity (demand) and infrastructure (supply) scenarios. Improvements to TDMs include more detailed representation of the behavioral models used to generate activity forecasts and of the multimodal networks that represent truck, rail, and water infrastructure. Federal legislation, Moving Ahead for Progress in the 21st Century (MAP-21) and the Fixing America's Surface Transportation (FAST) Acts, facilitated creation of statewide TDMs (STDMs) with freight components, which estimate freight demand flows,

support the estimation of freight performance measures, and guide multimodal planning (FHWA, 2013) (FHWA, 2017).

TDMs are mainly used for long-term travel forecasting, e.g. 20-40-year planning horizons. Depending on their application, TDM techniques range from simplified sketchplanning methods to complex trip- and activity-based approaches (National Academies of Sciences, Engineering, and Medicine, 2012). Trip-based models typically follow a sequential four-step approach: trip generation, trip distribution, mode choice and route assignment (Ortuzar & Willumnsen, 2011). In most cases, TDMs have passenger and freight models, which are combined before route assignment. The conventional four-step TDM is an effective method for determining network flows (when the network is represented), but lacks behavioral richness because it considers trips to occur independently rather than as trip chains. Activity Based Models (ABMs) use a disaggregate approach to incorporate relationships between trips, tours, and activities (Ortuzar & Willumnsen, 2011). The key step for activity-based models is to generate a synthetic population to represent agents (i.e. individuals, tucks, or vessels) in the study area. Most ABMs use U.S. Census data and public use microdata sample (PUMS) files to generate synthetic populations that match demographic and economic control targets for the base year (Castiglione, Bradley, & Gliebe, 2015). Then, agent activity patterns are generated. ABMs provide an intuitive and behaviorally realistic representation of travel by recognizing travel as a derived demand, but require highly disaggregated data (such as truck and vessel movements) map-matched to a transportation network to create a representative synthetic population and model its behavior.

Both trip-based and activity-based models require an accurate representation of the transportation network. However, state-of-the-practice multimodal TDMs have imbalanced

representation of mode-specific transportation networks, limiting their ability to accurately identify bottlenecks and impacts on the multimodal transportation system (Alliance Transportation Group, 2015). This work overcomes this limitation by creating a detailed navigable waterway network, and map-matching highly disaggregated maritime data. The purpose of this approach is ultimately to enable integration of maritime modes into trip-based or activity-based TDMs.

### 3.3.2. Automatic Identification System data (AIS)

AIS consists of vessel traffic data, collected for navigational safety purposes (e.g., collision avoidance) (Table 3.1). It is required for all passenger-carrying vessels and commercial vessels over 300 gross tonnage that travel internationally, by the International Maritime Organization (IMO) since December 2004. An onboard navigation device transmits location and characteristics of vessels in real time to receivers onshore in base stations, satellites, buoys, and other vessels (U.S. Coast Guard, n.d.). In the U.S., AIS is required as per Title 33, Code of Federal regulations (U.S. Coast Guard, n.d.) but not for all inland waterways (Dobbins & Langsdon, 2013), however most vessels use AIS transponders (DiJoseph & Mitchell, 2015).

Historical AIS data (2009-2017) is available for free download at (Office for Coastal Management, 2018) as geodatabases, including vessel, voyage, and broadcast information. Several vessel and voyage features are manual fields containing substantial errors and omissions while broadcasting features do not require manual entry. Each file contains point location data at 1-minute interval, per month and UMT zone.

Dataset characteristics	Waterborne AIS data
U.S. Data collection	U.S. Coast Guard
Data elements	<u>Vessel</u> : Vessel name, length, width, MMSI, IMO, and call sign <u>Broadcasting</u> : Latitude, longitude, time stamp, speed over ground, course over ground, heading <u>Voyage</u> : Cargo, draft, status
Spatial Coverage	All international and U.S. waterways
Temporal Coverage / 'ping' frequency	Discrimination to the minute of the day (for data storage and sharing purposes)
Update frequency	Since 2009, data collected in real time, shared via annual updates
Data storage format	File geodatabases (.csv) containing one month of data per UMT zone
Data sharing scheme	Open source via www.MarineCadastre.gov
Limitations	Carried by tugs and tows on inland waterways, not barges. Manually entered data may lack accuracy

 Table 3.1 AIS Data Characteristics

### 3.3.3. Challenges in using AIS Data for Freight Applications

Although AIS data collection is required for most freight vessels, publicly available datasets may contain select samples of all AIS records. For instance, some tug and tow operations might not be recorded due to smaller tugs not meeting the AIS reporting criteria (Perez et al., 2019). AIS data coverage may differ by region or port. In the Gulf Coast region, Perez at al. (2019) compared tug counts by port derived from AIS data with WCUS data concluding that AIS data accurately represented activity in the biggest port area, but overestimated or underestimated activity in smaller port areas, potentially due to the presence of fewer AIS reception points. Dobbins and Langsdon (2013) generated inland waterway one-day tow-trips from AIS data collected by a single AIS antenna and compared them to lockages reported by the USACE's Lock Performance Monitoring System (LPMS). They found that LPMS lockages were three times higher than AIS-detected lockages. LPMS records all vessels

that traverse each of approximately 200 locks and dams along the U.S. inland waterways, constituting a valuable source of data to evaluate coverage of AIS. Historical lockage data (1993-2017) is openly available in (U.A. Army Corps of Engineers, 2018c).

### 3.3.4. Map-matching Algorithm for Network Mapping

Map-matching reconstructs the trajectory of a GPS-enabled device on a network, from a series of potentially sparse, noisy position records or "pings" (Jensen & Tradišauskas, 2009). Each ping is defined by latitude, longitude, and timestamp. Map-matching algorithms iterate through timestamp-ordered pings, associate each ping to a network link based on location proximity, and store the series of links utilized by the vehicle (Camargo et al., 2017). A limitation is that, for dense networks, high-frequency pings, and large-scale data (i.e. several vehicles), the computation time can be prohibitive. Conversely, low-frequency pings and/or dense networks lead to incomplete path identification and low map matching accuracy, e.g., many links are traversed between pings. As a result, most map-matching algorithms trade-off between computation time and accuracy (Hashemi & Karimi, 2014).

Within the context of freight transportation, to overcome low-performance issues, Pinjari et al. (2014) reduced truck GPS pings to 5-minute frequencies prior to map-matching (Pinjari et al., 2014). Camargo et al. (2017) proposed a map-matching algorithm for low-frequency truck GPS data. The algorithm iterates through all pings and identifies stops based on calculated speed, stop duration, and coverage (length of the diagonal of a bounding box containing all consecutive-stopped pings). Then, the algorithm re-iterates through the pings to identify which network links are likely used by the truck along its path. Lastly, a trip is created by computing the shortest path between each pair of consecutive stops, using the links previously identified. The algorithm was applied for activity-based modeling in a large metropolitan area to conduct select link, OD and

time-of-day analysis, and trajectory visualization (Camargo et al., 2017). Akter et al. (2019) adapted Camargo's algorithm to truck GPS data for a state-wide network. Using the mapmatching algorithm output, Akter et al. derived truck operational characteristics (stop time-ofday, stop duration, trip length, trip duration, and total number of stops in a day) and fed a multinomial logit model that distinguished truck daily activity patterns into five commodity groups (Akter, Hernandez, Corro-Diaz, & Ngo, 2018). This work expands the utilization of mapmatching algorithms to waterway networks by adapting the work of Camargo et al. (2017).

### 3.3.5. Trip Identification from AIS data

Previous works reconstructed vessel trajectories from AIS data (Zhang, Meng, Xiao, & Fu, 2018; Dobrkovic, Iacob, & Van Hillegersbarg, 2018; Zhao, Shi, & Yang, 2018; Graser, 2019) but were limited either in their ability to match pings to a defined and detailed inland waterway network, or in that movements were divided per day, masking the identification of trips. DiJoseph and Mitchell (2015) overcome the latter by applying an algorithm to link time-consecutive AIS records together to generate paths on inland waterways (DiJoseph & Mitchell, 2015), but did not fuse generated vessel paths with a defined network. The inability to map vessel data to a network precludes future incorporation and integration of AIS data into multimodal, network-based models, such as TDMs. In contrast, the algorithm developed in this work allows for the identification of trips defined by origin and destination (not duration) and matched to a defined network.

# 3.4. Methodology

The methodology consists of three steps: (1) Data preparation, (2) Vessel stop identification, and (3) Vessel trip identification. All data and tools are open source.

### 3.4.1. Step 1-Data preparation

Data preparation involves three procedures: i) AIS data reduction, ii) AIS data quality control, and iii) development of the detailed inland waterway network. (Figure 3.1 and 3.2)

### Step 1.1-AIS Data Reduction

Data reduction is necessary to accelerate "big data" processing. In AIS datasets, records with zero speed outnumber the non-zero speed records (Osekowska, Johnson, & Carlsson, 2017) and, depending on the application, removal of zero speed records provides a mechanism for data reduction. For example, Fujino et al. reconstructed vessel trajectories from a reduced AIS dataset and applied unsupervised machine learning to identify vessel course and issue real-time off-course warnings. The original dataset of 5,756,438 records was reduced by 40% by removing records with zero speed (Fujino, Claramunt, & Boudraa, 2018). Following this example, in this paper, zero speed records are removed with no loss of representation of trip characteristics needed for map matching and stop identification heuristics. By removing zero-speed records from the AIS dataset, computational time is reduced while still benefiting from highly disaggregated, ubiquitous AIS characteristics.

#### Step 1.2-AIS Quality Control

AIS data contains erroneous or irrelevant records that result from transmission interference and device mishandling. Erroneous records are defined as those with unusual high speed, or records located far from inland waterways (Figure 3.2.a and b). Irrelevant records come from vehicles that emitted less than 20 records within the reporting period, and/or from vessels whose records are outside reasonable waterway boundaries. After identifying erroneous and irrelevant records as described below, they are removed from further analysis.



Figure 3.1 AIS data preparation flowchart



Figure 3.2 Example of AIS data preparation

To identify erroneous records, first, a spatial buffer is created for an inexact U.S. navigable waterway network from the National Transportation Atlas Database ("NTAD") (Bureau of Transportation Statistics, 2015), clipped to the study area. The buffer width is derived from the Global River Bankfull Width & Depth Database ("NARVIS") (Andreadis, Schumann, & Pavelsky, 2013). NARVIS and NTAD are provided as geodatabases. Because the NTAD waterway geometry is abstract, it may not follow observed and valid AIS records. Therefore, a spatial buffer should be established to exclude records grossly outside of the navigable waterways (Figure 3.2.c). Adopted buffer size of two standard deviations from the NARVIS mean width was found appropriate in this work. Records outside the buffer are removed.

Second, a forward sequential search iterates over consecutive AIS records to calculate speed (eq. 3.1). The speed (as space mean speed) is checked against a reasonableness threshold of 27.7km/h (15 knots), based on (El-Reedy, 2012). By applying the proposed speed threshold, records corresponding to non-freight vessels are discarded.

$$speed_{i} = \frac{travelled \ distance_{[i-1,i]}}{traveled \ time_{[i-1,i]}}$$
 (3.1)

Where,

speed = space-mean-speed associated with pings i-1 and i, in km/h

*travelled distance* = great-circle distance based on position (latitude, longitude) between pings *i*-1 and *i*, in kilometers

*travelled time* = time to travel between pings *i*-1 and *i*, in hours

Next, if less than 20 records are associated with one vessel, all the records for such vessel are removed. Last, spatial coverage of each remaining vessel records is calculated as the diagonal of a bounding box around all of its pings. Vessels with coverage less than 2km are removed. The coverage threshold is defined as the minimum distance between different port authorities in the study area.

#### Step 1.3- Inland Waterway Network Development

The objective of this step is to create a detailed representation of an inland waterway network as nodes and links on which to map-match the AIS vessel movements. Unlike previous work (Bureau of Transportation Statistics, 2015), the network is expanded to include nodes representing: (i) connections between links to accommodate geometry and attribute changes; (ii) locks; (iii) port terminals; and (iv) staging or anchoring areas. To identify (i) and (ii), the NTAD network layer (Bureau of Transportation Statistics, 2015) is used. To identify (iii), because the NTAD network node layer lacks sufficient detail (several port terminals are aggregated to single port authorities) it must be supplemented with local data such as state or regional port databases. If local data is not available, a review of open-source aerial imagery should be performed. To identify (iv), NTAD cannot be used as it does not specify the location of designated anchorage grounds. Thus, clusters of AIS records corresponding to tugs with low speed but that did not match the location of a port terminal can be used to locate designated or undesignated anchorage grounds. As a result, "staging" or anchoring areas are identified, i.e. areas within the waterway where tugs may leave barges to be picked up later. Polygons representing port terminals,

anchoring areas, and locks are drawn to surround clusters of low-speed records observed from AIS data and labeled (Figure 3.3).





The waterway line layer used as a basis to develop the detailed network is also obtained from the NTAD. Since the NTAD port layer is supplemented with additional port and non-port nodes, it is necessary to edit the NTAD waterways line layer to accommodate these "new" nodes. Next, AIS data is used to identify potential missing links and accommodate the network representation to the path followed by vessels (Figure 3.4). First, pings outside a buffer of the NTAD waterway line layer are selected. Buffer size is the NARVIS mean river width (Andreadis, Schumann, & Pavelsky, 2013). Second, clusters of pings outside the buffer are identified. Third, for each cluster, the two closest nodes to each cluster centroid are found, filtering out repeated node pairs. Fourth, identified nodes are connected to the cluster centroid with a new link. Lastly, the modified waterway line layer is subject to a GIS plugin to generate a routable network model (AequilibraE, 2018) with link "cost" determined from link length. Attributes added to the network link layer include length (miles), and travel time (hours). Transit travel time is calculated based on the link length and on an assumed vessel speed of 5.8mph (Yuan & Harik, 2010), except for links representing locks, were a "lockage transit travel time" is obtained from the annual average processing time provided in the LPMS (U.S. Army Corps of Engineers, 2018c).



Figure 3.4 Detailed navigable waterway network development

# 3.4.2. Step 2-Stop Identification

The purpose of this step is to identify and characterize stops made by vessels using a stop identification algorithm modified from (Camargo et al., 2017). For AIS records, zero and low speed position records both correspond to stops, and many zero or low speed position records found in close geographic proximity may correspond to the same stop, rather than to several unique stops. Even though each position record includes a point-speed estimate, point speeds may not be reliable due to transmission issues and thus cannot be used alone to define a stop. Instead, the stop identification algorithm evaluates consecutive series of position records to discover 'stop clusters' (Figure 3.5). Each stop cluster is defined by a stop time (the average timestamp of all pings within the cluster), duration, position (the centroid of the cluster), and location (e.g., at a port, lock, anchoring, or other area) (Figure 3.6).



Legend Sample vessel Stop identified by the algorithm (cluster centroid) Low-speed records cluster

### Figure 3.5 Stop cluster example

The parameters within the stop identification algorithm are: speed, minimum time stopped, and maximum stop coverage. To define values for algorithm parameters, manual verification of stop locations within port areas is performed for a sample of AIS records. Parameters are iteratively calibrated to achieve acceptable performance measured in terms of precision (eq. 3.2), e.g., the number of correctly identified stops at ports (true positives) relative to the number of identified stops at ports (true positives and false positives). By using precision as the performance metric for algorithm calibration, the occurrence of correctly identified stops is maximized while reducing the occurrence of "duplicated stops". Duplicated stops are defined as two (or more) timewise consecutive stops occurring at nearby locations that in reality should be clustered into a single stop. Precision considers both true positives (TP, stops correctly identified as a stop, or duplicated stops). Details regarding parameter estimation and performance are presented in the case study section.

$$Precision = \frac{TP}{(FP+TP)}$$
(3.2)

Where,

*True Positives (TP)*= number of stops correctly identified by the algorithm

*False Positives (FP)*= number of stops that the algorithm incorrectly identified as such, or duplicated stops



Figure 3.6 AIS stop identification flowchart

# 3.4.3. Step 3–Trip Identification

The purpose of this step is to reconstruct vessel trajectories as complete and connected paths defined by network links and nodes using map-matching heuristics, and define individual freight trips by origin and destination (Figure 3.7).



Figure 3.7 AIS trip identification flowchart

# Step 3.1-Vessel Path Identification and Network Map-Matching

To reconstruct vessel trajectories defined by network links and nodes using mapmatching heuristics, the map-matching heuristic developed by Camargo et al. (2017) was adapted as follows (Figure 3.8). For each vessel, first, stop cluster records are associated with a network node by proximity. Second, the complete path of the vessel is reconstructed by assuming that the vessel takes the shortest path between pairs of timewise-consecutive stop clusters associated with different nodes. For highway applications, the later assumption can be a challenge to meet given dense highway networks with many competing 'shortest' paths. The algorithm by Camargo et al. accounted for this by limiting the shortest path links to those that comprised the reduced set of network links associated with pings. However, for inland waterway networks, there are relatively fewer nodes and links from which to reconstruct a shortest path between stops. Therefore, the map matching algorithm can be simplified by finding the shortest path between stop clusters without the need to look at a reduced link set. The approach of searching for shortest paths between stop clusters and not between all pings, thus, serves to increase computational efficiency without reducing path identification accuracy.

Ultimately, the map-matching algorithm produces a sequence of shortest paths ("path segments") that constitute the complete paths made by all vessels. Path segments are represented as the series of nodes of the network visited by each vessel between each pair of consecutive stops, the time when the vessel arrived and left each node, and the associated network link connecting consecutive nodes.



Figure 3.8 Map-matching and trip characterization algorithm for a sample vessel

# Step 3.2–Vessel Trip Characterization

Following the map-matching procedure vessel paths are defined by OD so that individual freight trips and trip chains can be characterized. ODs are defined as freight ports and a distinction is made between freight stops (pick-up or delivery) at ports and stops due to lockage, anchoring, and other non-freight activity.

Along a vessel path, stops at locks are mandatory, traffic-related (equivalent to a truck stopping at a traffic light), and irrelevant for characterizing freight activity as purposed in this paper. Thus, trips are defined as the combination of successive path segments that share a lock as an intermediate stop. For example, in Figure 3.8, network node 83 (associated with stop cluster 6) represents a lock; so, path segments 1 and 2 are combined into a single trip with nodes 72 and

84 as origin and destination. As such, vessel trips are based on the time and location of their stops that constitute trip origins and destinations, regardless of the duration of the trip, as was assumed in prior work (Dobbins & Langsdon, 2013) (Graser, 2019). Once trips are defined, trip characteristics such as trip length (in miles) and duration (in hours) are derived by aggregating the length and transit time of all the links comprising the trip. Other trip characteristics include: trip origin and destination nodes, and location (e.g., port, staging/anchoring, lock, and other nonport).

Next, trips are combined for cases of potentially non-freight ODs. For ODs found to be non-ports (anchoring areas and other network nodes), the consecutive trips are combined such that the stop at the non-port becomes an intermediate stop (not an origin or destination) in the trip chain (Figure 3.8). All nodes in the network should be predefined as ports or non-ports so that identification of intermediate stops is facilitated.

# 3.5. Case Study: Maritime Freight Activity in Arkansas

### 3.5.1. Scope and Data

The map-matching methodology was evaluated using AIS data gathered from the MKARNS and the portion of the Mississippi River along Arkansas' eastern border for the year 2016.

In total, 7,803,151 AIS records emitted with a 5-minute frequency by 776 vessels were extracted from (Office for Coastal Management, 2018) (Figure 3.2.a). 116 of the 776 vessels were observed within the MKARNS, while the remaining 660 vessels were observed within the Mississippi River (and did not use the MKARNS). Of these records, 53% corresponded to zero speed records, which were removed (Step 1.1, Figure 3.2.b). The quality control process (Step

1.2) excluded 518,697 position records from the dataset. As a result, 3,398,279 AIS records(44% of the original sample) were subject to the map-matching procedures (Figure 3.2.c).

The data used for this work constituted a sample of the population of vessels traveling on Arkansas waterways during 2016. Thus, a coefficient of coverage was calculated by comparing unprocessed AIS traces with LPMS vessel counts. The reduced AIS data sample represents 88% of commercial vessels operating on the MKARNS during 2016. Coverage varies per lock, possibly indicating that the AIS sample excluded more vessels observed in the proximity of the locks where a lower coefficient of coverage was found, i.e. the Oklahoma portion of the MKARNS.

The development of the detailed inland waterway network (Step 1.3) was complemented with data from the Arkansas Economic Development Commission GIS office (Arkansas Economic Development Commission, n.d.). Only port terminals located on the MKARNS were considered, resulting in 43 unique freight port terminals and 11 staging/anchoring areas.

# 3.5.2. Stop Identification and Map-Matching Parameter Calibration

Tunable parameters within the stop identification and map-matching algorithms were calibrated against a manually verified dataset generated by the authors. Using a random stratified sample of eight vessels from the 2016 AIS records, 4,869 stops (3,820 trips across 352 days) were manually identified by comparing vessel position records to aerial imagery. Stops were manually identified based on the position record spot speed, location of the stop, and characteristics (speed, position) of prior and subsequent stops. The stratified random sample considered: number of pings (less than 15,000; 15,000-30,000; and over 30,000), expanded time coverage (less than 3 months; 3-9 months, and 9-12 months), and frequency of presence at ports in the study area (less than 20; 20-30, and more than 30 ports visited).

With the groundtruthed data, sample parameters were calibrated using a partial combinatorial search heuristic within the ranges expressed in Table 3.2. The search over the parameter space continued until model performance (precision) no longer improved. Over 40 combinations were evaluated; the combination of parameter values which gave highest precision was selected. The calibrated parameters showed a precision of 83% at ports, 85% at locks, 74% at staging/anchoring areas, and 32% elsewhere. The overall precision (i.e. stops at all location types) was 67%. In addition, the calibrated parameters produced the fewest duplicated stops (16%), as defined in Step 2 of the methodology.

Stop Parameter	Calibrated Value	Range Performance	Range Tested
Stopped speed (km/h)	5.3	80.0-84.0% (4.0%)	4.5-6.0
Minimum time stopped (seconds)	300	73.9-84.0% (10.1%)	300-1,200
Maximum stop coverage (km)	5.0	83.2-84.0% (0.8%)	2.0-15.0

 Table 3.2 Vessel Stop Identification Parameters

# 3.5.3. Results

The stop identification algorithm identified 120,185 stops for the 3.4 million AIS position records, of which 24% were on the MKARNS and 75% on the Mississippi River (Figure 3.9). The subsequent map-matching algorithm identified 47,555 trips (Figure 3.10), and 31,359 trip chains. The average number of annual trips per vessel was 63, with a mean trip length of 56.7 miles within a range of 0.2 to 1,085 miles, and a mean duration of 10 hours with a range of 1 to 214 hours. Vessel trips of shortest length and duration likely correspond to movements of tugs between docks within a given port, and to support construction occurring during 2016, e.g., Broadway Bridge in Little Rock. The data was processed in 485 minutes using a computer with Intel® Core™ i7-8700 processor (3.20GHz), 32GB RAM, Microsoft Windows 10, 64-bit

operating system. Open-source software was used: Python, PostgreSQL for database management, and Quantum GIS (QGIS) for geoprocessing and visualization.

### 3.5.4. Model Validation

For validation, the trip paths identified by the model (i.e. processed AIS data) on the MKARNS were compared to LPMS data (i.e. trip lockages) (eq. 3.3) following (Dobbins & Langsdon, 2013).

$$V = \frac{Processed AIS \ data \ lockages}{LPMS \ data \ lockages} \tag{3.3}$$

Where,

V= model evaluation metric,

Processed AIS data lockages = annual number of tugs observed from the processed AIS data (trips) in transit through each of the locks in the study area, and

LPMS data lockages = annual number of commercial vessels reported by LPMS for the same locks during the same time period (U.S. Army Corps of Engineers, 2018.c)

To estimate the Processed AIS data lockages, trip geometries of tugs/tows that

intersected locks (represented by screenlines) were counted as vessels in transit through the lock. Validation results show that the model is capable of correctly identifying 83.5% of trip lockages, with a range of [65.6%-96.6%] by lock. This validation is limited in that it only considers the trips that crossed a lock, thus excluding trips on the lower Mississippi River (where there are no locks).

Algorithm precision likely varies as a result of: i) the AIS dataset where the model is tested represents 88% (not 100%) of the vessel activity in the area; ii) the random stratified sample of vessels used to train the model constitutes only 1% of the vessels in the dataset; and iii) it is observed that a single set of stop identification algorithm parameters does not fit best all groundtruthed vessels.



Figure 3.9 Stop identification results from AIS data

Figure 3.10 Trip identification results from AIS data

# 3.6. Discussion

The map-matching method presented in this paper recreates vessel trips from AIS position records by first identifying the location of freight delivery stops that constitute trip ODs, and then connecting those stops as complete consecutive series of inland waterway network links. In addition, matching of vessel trips to a robust inland waterway network allows for further integration into multimodal STDMs, which typically fall short in their representation of non-truck modes.

Since AIS data is available worldwide and for various time periods (past and present), the proposed methodology has potential for spatial and temporal transferability. Tunable parameters within the stop identification and map matching heuristics such as stop duration, speed, and spatial coverage are calibrated using manually verified vessel trajectories. It is possible that for other regions and time periods, waterway geometry and vessel operational characteristics may differ, and thus tunable parameters should be recalibrated. A sensitivity analysis is performed to illustrate the impact of various model parameters on model performance (Figure 3.11) including: (a) stopped speed, (b) minimum time stopped, and (c) maximum stop coverage.

Stopped speed is varied between 4.5 and 6.0km/h (Figure 3.11.a). For the case study, 5.3km/h (2.9 knots) produces the highest precision in stops identified at ports (84.0%). For the stopped speed values tested, in general, as the stopped speed decreases below or increases above 5.3km/h (highest precision), the algorithm precision decreases, and the number of stops identified decreases by as much as 13%. This can be attributed to the number (and proportion) of duplicated stops which tend to occur at or near at ports more than at non-port areas. This increase in the number (and proportion) of duplicated stops identified by the algorithm (false positives in eq. 3.2) produces a decrease in the precision.

Minimum time stopped is varied between 300 and 2400 seconds (5 to 40 minutes) (Figure 3.11.b). For the case study, 300 seconds produces the highest precision in the stops identified at ports. In general, as minimum time stopped increases, the algorithm precision decreases from 84.0% to 73.9%. As a reference, typically it takes a tug about 30 minutes to a few hours to pick-up and deliver barges at a port terminal, depending on weather, cargo, etc. Most importantly, the selection of the minimum time stopped is dependent on the frequency of the ping data. For instance, considering that vessels may emit only one ping while stopped and that the AIS data used for analysis has a frequency of 300 seconds, stops corresponding to single ping records would not be identified if the minimum time stopped parameter was greater than 300 seconds.

Maximum stop coverage is varied between 2 and 15km (Figure 3.11.c). For the case study, 5.0km produces the highest precision in combination with the other parameters (84.0%). In general, as the stop coverage increases, the number of stops identified decreases slightly (by 0.8%), while precision to identify stops at ports does not change substantially (0.9%). This is likely due to an increase in stops identified in locations other than ports.

A notable limitation of the proposed methodology to analyze freight activity based on AIS tracking data is that AIS transponders are installed on tugs and tows, instead of on the barges that carry freight (Kruse, et al., 2018). This has several implications: i) trips made by tow/tugboats not transiting barges or transiting empty barges are included in the AIS data; and ii) a tow/tugboat may pick-up loaded barges from an origin port, and leave them in the vicinity of its destination port, to be picked up later to reach its final destination. Such movements are recorded as two separate trips, masking the true OD of the freight. Lastly, since each tow may push several barges, the amount of freight carried in each trip is unknown. This uncertainty in

freight volume would be mitigated by creating policy to mandate AIS tracking on barges, which would also support safety. Notably, the Port of Antwerp requires all barges to carry an AIS device (Port of Antwerp, 2012).



Figure 3.11 Sensitivity of stop identification output to algorithm parameters

# 3.7. Conclusion

Vessel tracking data, ubiquitous in time and space, provides a consistent source to observe freight activity on inland navigable waterways. The stop identification and map matching heuristics presented in this paper allow vessel tracking data to be used to define and characterize freight trips along the inland waterway network. The methodology presented in this paper first identifies stops made by each vessel by clustering successive AIS position records based on their location, timestamp, and calculated speed. Then, each stop is associated with a network node based on proximity. If two timewise-consecutive stops are assigned to different network nodes, they constitute the OD of a path segment. Then, a map-matching algorithm reconstructs complete vessel paths by finding the shortest path between OD pairs. Path segments through locks are joined to define freight trips, and trip chains between freight activity stops at ports. Lastly, freight trips characteristics are derived, such as trip length, duration, origin, and destination. The methodology is applied to Arkansas waterways with 84.0% precision in detecting stops at ports in the Arkansas River. Sensitivity of the model parameters like maximum stop speed and duration show that to ensure accuracy for other regions, parameter calibration is necessary. Validation results show the model correctly identifies 83.5% of trips crossing locks. Given that historical AIS data are increasingly available worldwide, the proposed methodology may be applied to any region with waterways.

Overcoming the limitations of prior analyses of AIS datasets, this work allows AIS data to be mapped to a well-defined inland waterway network (also generated from AIS data). In doing so, freight activity along the inland waterway can be integrated into travel demand model (TDM) frameworks. This is a benefit because many freight TDMs focus mainly on highways ignoring important multimodal connectivity, leading to the inability to estimate multimodal performance metrics like freight fluidity. With the availability of AIS data and the methods for freight trip identification presented in this paper it is increasingly possible to represent and integrate non-truck modes in freight TDMs.

Building upon the methodology proposed in this paper, the authors are working on further characterizing inland waterway freight movements by identifying the commodity carried in each trip. Such characterization may be realized by fusing LPMS, AIS, and truck GPS data using stochastic assignment methods. Ultimately, this paper and future work help to fill data gaps

often referenced for freight commodity flows so that freight project identification and prioritization can best leverage data driven approaches.

# 3.8. Disclaimer

This material is based upon work supported by the U.S. Department of Transportation under Grant Award Number 69A3551747130. The work was conducted through the Maritime Transportation Research and Education Center at the University of Arkansas. This work reflects the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

# 3.9. References

AequilibraE. (2018, July). AequilibraE Documentation. Retrieved from www.aequilibrae.com

- Akter, T., Hernandez, S., Corro-Diaz, K., et al. (2018). Leveraging Open Source GIS Tools to Determine Freight Activity Petterns from Anonymous GIS Data. Paper presented at the AASHTO GIS for Transportation Symposium. https://gis-t.transportation.org/wpcontent/uploads/sites/51/2018/05/GIS\_T\_Taslima\_Akter\_svh\_TA.pdf
- Alliance Transportation Group. (2015). *Arkansas Statewide Travel Demand Model Documentation*.
- Andreadis, K., Schumann, G., & Pavelsky, T. (2013). A Simple Global River Bankfull Width and Depth Database. *Water Resources Research*, 49, 7164-7198.
- Arkansas Economic Development Commission. (2017, March). Arkansas Site Selection. retrieved from http://www.arkansassiteselection.com/aedc/gis-data-download.aspx
- Bureau of Transportation Statistics. (2015). *National Transportation Atlas Database*, Geospatial at the Bureau of Transportation Statistics. Retrieved from http://osav-usdot.opendata.arcgis.com/datasets?keyword=Water
- Camargo, P., Hong, S., & Livshits, V. (2017). Expanding the Uses of Truck GPS Data in Freight Modeling and Planning Activities. *Transportation Research Record*, 2646, 1, 68-76. doi:https://doi.org/10.3141/2646-08
- Castiglione, J., Bradley, M., & Gliebe, J. (2015). *Activity-Based Travel Demand Models: A Primer*. Retrieved from https://www.nap.edu/read/22357/chapter/1
- DiJoseph, P., & Mitchell, K. (2015). *Estimating Vessel Travel Time Statistics for Inland Waterways with Automatic Identification System Data*. Paper presented at the Transportation Research Board 94th Annual Meeting, Washington, DC.
- Dobbins, J., & Langsdon, L. (2013). Use of Data from Automatic Identification Systems to Generate Inland Waterway Trip Information *Transportation Research Record*, 2330, 1, 73-79. https://doi.org/10.3141/2330-10
- Dobrkovic, A., Iacob, M. E., & Van Hillegersbarg, J. (2018). Maritime Pattern Extraction and Route Reconstruction from Incomplete AIS data. *International Journal of Data Science and Analytics, 2018, 5,* 111-136. https://doi.org/10.1007/s41060-017-0092-8
- El-Reedy, M. (2012). Chapter 5: Fabrication and Installation, in Offshore Structures, 293-381.
- FHWA. (2013, September 12). Significant Freight Provisions. *Moving Ahead for Progress in the 21st Century Act*. Retrieved from www.fhwa.dot.gov/map21/factsheets/freight.cfm
- FHWA. (2017, February 8). Freight Planning and Policy. *Fixing America's Surface Transportation Act.* Retrieved from www.fhwa.dot.gov/fastact/factsheets/fpppfs.cfm
- FHWA. (2017b). *Freight Demand Modeling and Data Improvement Handbook*. Retrieved from https://ops.fhwa.dot.gov/publications/fhwahop18018/fhwahop18018.pdf
- FHWA. (2019). *Freight Analysis Framework Version 4*. Retrieved from https://faf.ornl.gov/faf4/Extraction1.aspx
- Fujino, I., Claramunt, C., & Boudraa, A.-O. (2018). Extracting Courses of Vessels from AIS Data and Real-time Warning Against Off-course. Paper presented at the International Conference on Big Data Research, Weihai, China. https://doi.org/10.1145/3291801.3291823
- Graser, A. (2019, January 26). Movement data in GIS. *Splitting Trajectories by Date*. Retrieved from https://anitagraser.com/2019/01/26/movement-data-in-gis-19-splitting-trajectories-by-date/
- Hashemi, M., & Karimi, H. (2014). A Critical Review of Real-time Map-matching Algorithms: Current Issues and Future Directions. *Computers, Environment and Urban Systems*, 48, 153-165. https://doi.org/10.1016/j.compenvurbsys.2014.07.009
- Javor, A., Bober, S., Creemers, P., et al. River Information Services Portal. (2013). *Operational Use of Inland AIS*. Retrieved from http://www.ris.eu/docs/File/348/leaflet\_operational\_use\_of\_inland\_ais%20\_ed2013\_vtte g\_2013%20en.pdf

- Jensen, C., & Tradišauskas, N. (2009). *Encyclopedia of Database Systems*. Boston, MA: Springer.
- Kruse, J., Mitchell,K., DiJoseph, P., Kang, D., Schrank, D., and Eisele, W. (2018). Developing and Implementing a Port Fluidity Performance Measurement Methodology using Automatic Identification System Data. *Transportation Research Record*, 2672, 11, 30-40. https://doi.org/10.1177/0361198118796353
- Nachtmann, H., Boudhoum, O., & Oztanriseven, F. (2015). *Regional Economic Impact Study for the McClellan-Kerr Arkansas River Navigation System*. Retrieved from https://martrec.uark.edu/research/regional-economic-impact-study-for-the-mcclellankerr-arkansas-river-navigation-system.pdf
- National Academies of Sciences, Engineering, and Medicine. (2012). *Travel Demand Forecasting: Parameters and Techniques*. The National Academies Press, Washington, DC. Retrieved from http://www.trb.org/Publications/Blurbs/167055.aspx
- Office for Coastal Management (Producer). (2018). Vessel Traffic Data. *MarineCadastre.gov*. Retrieved from https://marinecadastre.gov/ais/
- Ortuzar, J. D., & Willumnsen, L. G. (2011). *Modelling Transport, 4th Edition*. West Sussex: Wiley.
- Osekowska, E., Johnson, H., & Carlsson, B. (2017). Maritime Vessel Traffic Modeling in the Context of Concept Drift. *Transportation Research Procedia*, 25, 1457-1476. https://doi.org/10.1016/j.trpro.2017.05.173
- Perez, H., Chang, R., Billings, R., & Kosub, T. (2019). Automatic Identification Systems (AIS) Data Use in Marine Vessel Emission Estimation. Paper presented in 18th Annual International Emission Inventory Conference.
- Pinjari, A., Zanjani, B., Thakur, A., et al. (2014). Using Truck Fleet Data in Combination with Other Data Sources for Freight Modeling and Planning. Retrieved from https://www.fdot.gov/docs/default-source/research/reports/FDOT-BDK84-977-20-rpt.pdf
- Port of Antwerp. (2012). Annual Report 2011. Retrieved from https://www.portofantwerp.com/en/ais-barges
- U.S. Army Corps of Engineers. (2018a). Manuscript. *Digital Library [2000-2018 trips] Manuscript Cargo and Trips Data Files*. Retrieved from https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/1671
- U.S. Army Corps of Engineers. (2018b). Manuscript All Regions 2016. *Digital Library [2000-2018] Manuscript Cargo and Trips Data Files*. Retrieved from https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/1689

- U.S. Army Corps of Engineers. (2018c). *Table of Contents Locks by Waterway, Lock Usage, CY 1993-2017," USACE Digital Library- Public Lock Reports.* Retrieved from https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/2958
- U.S. Army Corps of Engineers. (2016). *Waterborne Commerce of the United States*. Retrieved from https://publibrary.planusace.us/document/f202cdec-ef75-49ca-8e16-829182cfdf4c
- U.S. Coast Guard (Producer). Automatic Identification System. *Navigation Data Center*. Retrieved from https://www.navcen.uscg.gov/?pageName=AISmain
- U.S. Coast Guard (Producer). *AIS Requirements, Title 33*. Retrieved from https://www.navcen.uscg.gov/?pageName=AISRequirementsRev
- Yuan, P., & Harik, I. (2010). Equivalent Barge and Flotilla Impact Forces on Bridge Piers. *Journal of Bridge Engineering*. 15, 5, 523-532.
- Zhang, L., Meng, Q., Xiao, Z., et al. (2018). A Novel Ship Trajectory Reconstruction Approach Using AIS Data. Ocean Engineering, 159, 165-174. https://doi.org/10.1016/j.oceaneng.2018.03.085
- Zhao, L., Shi, G., & Yang, J. (2018). Ship Trajectories Pre-processing Based on AIS Data. *The Journal of Navigation*, 71, 1210-1230. https://doi.org/10.1017/S0373463318000188

# CHAPTER 4. GIS-Based Identification and Visualization of Multimodal Freight Transportation Impact Areas

# 4.1. Abstract

Measuring the area of impact of a transportation infrastructure project is necessary to estimate its impacts, monetize its benefits, support cost-benefit analyses, and project prioritization. State-of-the practice methods to identify a facility's impact area consist primarily of arbitrarily selecting a radial perimeter around it. For freight related projects, this method ignores complex interactions among freight modes and supply chains, longer travel distances, etc., and may limit the impact area. Instead, the impact area of a project affecting a freight facility can be better defined by examining the origins and destinations (OD pairs) served by the modes that use the facility, or "freight catchment" area. Such OD pairs may be obtained from a travel demand model (TDM) and/or project specific data. Unfortunately, TDMs do not typically contain robust depictions of water and rail modes, preventing the identification of multimodal freight catchment areas. In addition, project-specific data including local traffic counts and stakeholder surveys are time consuming and subjective. This work overcomes the limitations of gathering multimodal OD pairs by introducing a method to identify multimodal freight catchment areas by leveraging emerging sources of "big data". Geospatial data fusion approaches including map-matching and route identification are applied to integrate truck Global Positioning System (GPS) and maritime Automatic Identification System (AIS) data, which are continuous and ubiquitous over time and space. A case study of port terminals on the Arkansas River exemplifies the methodology. Results show that adopting an arbitrary radial impact area for different ports would lead to inaccurate project benefit estimates, and identify corridors of modal competition.

**Key words:** Multimodal Freight, Project Prioritization, Economic Impacts, Catchment Areas, Geographic Information System (GIS)

# 4.2. Introduction

In the context of public and private transportation infrastructure investment, projects compete for limited resources. Thus, evaluation and prioritization of competing projects is critical, but of the methods available, all rely in part on estimation of benefits relative to costs. Calculation of benefits necessitates detailed estimation of the types and magnitude of project impacts associated with the project impact area (Chacon Hurtado et al., 2016; Weisbord et al., 2009). Therefore, it is important to clearly and consistently define the extent, location, and characteristics of a project's impact area.

The impact area of a project affecting a multimodal freight facility can be defined as the region where the facility draws and delivers freight, or the origin-destination (OD) pairs served by the facility (Vadali et al., 2017). Despite general agreement on conceptual and qualitative identification of a project's economic impact area (Economic Development Research Group et al., 2014; Vadali et al., 2017), little published work is available to guide data-driven methods to define impact areas for freight projects. Instead, planning agencies within Metropolitan Planning Organizations (MPOs) and state Departments of Transportation (DOTs) must use professional judgement to define each project impact area (AASHTO, 2015). As a result, the lack of a shared method to determine the area of impact of competing projects may prevent proper comparison and promote unfair competition among projects from different agencies and jurisdictions. To ensure fair comparisons, impact areas should be determined by following the same, systematic methodology.

State-of-the practice methods to identify project impact areas consist primarily of selecting an arbitrary radial perimeter around the facility (Carroll et al., 2017; NADO Research Foundation, 2011; Tyndall Air Force Base, 2018). For freight related projects, however, this method ignores complex interactions among freight supply chain components (e.g. truck, rail, water) and longer travel distances, for example. Instead, impact area of a project affecting a freight facility can be better defined as the OD pairs served by freight modes using the facility. This has been referred to as the "freight catchment" area.

Freight OD pairs may be obtained from project-specific data like stakeholder surveys or traffic counts and/or statewide travel demand models (TDM). Project-specific data can be time consuming to collect, subjective, or in the case of annual traffic counts, may not be available at or near the project. If a statewide freight TDM exists, the OD pairs served by a freight facility might be found by performing a 'select link analysis' (Alliance Transportation Group, 2015). However, although TDMs contain representative models of the roadway network, they often do not provide robust depictions of water and rail networks and are thus unsuitable for multimodal freight catchment analyses (Alliance Transportation Group, 2015; Donnelly et al., 2018). For example, the statewide TDM in Arkansas (ARSTDM) contains a multimodal mode choice model, but only performs trip assignment for highway flows and not vessel flows because waterways are not part of the model network (Alliance Transportation Group, 2015). This is a notable limitation considering the key benefits of the Arkansas River to the state economy (Nachtmann et al., 2015). In addition, network representation in TDMs often lacks the level of detail necessary to represent actual roadway geometry, i.e. port access roads may not be represented.

The lack of guidance regarding multimodal catchment area definitions can be attributed in part to the heterogeneity of the data used for this purpose. In a freight supply chain, the catchment area contains several modes, freight facilities, and industries, which would be better represented (and linked together) by spatially and temporally continuous data, such as historical truck and vessel paths. To overcome the limitations in defining catchment areas for freight facilities, this paper leverages two emerging sources of "big data" to identify multimodal freight paths: truck Global Position System (GPS) and marine Automatic Identification System (AIS) data. The main contribution is a geospatial data fusion method to identify the impact area of multimodal freight projects using ubiquitous vehicle tracking data. Conceptually, multimodal freight tracking data is used to characterize spatial patterns of freight intensity. In particular, GPS tracking data from trucks and marine vessels accessing freight facilities are mined to identify stops and to find complete paths which are then mapped to a high resolution, multimodal transportation network. The geographical coverage of the truck and vessel trips define the "multimodal catchment" or impact area of the project. In this way, all competing projects are subject to identical criteria for impact area definition, providing a common basis for funding priority.

The methodology is applied to freight ports located on inland navigable waterways, although it applies to other infrastructure including bridges, railyards, and warehouses. Beyond benefit-costs analyses, quantitative definitions of multimodal impact areas further efforts to: i) quantify multimodal performance measures, ii) visualize the extent of transportation impacts of extreme weather events (such as flooding), iii) estimate population exposures to pollutants or congestion effects induced by freight facilities, and iv) identify areas of modal competition.

# 4.3. Background

#### 4.3.1. Project Evaluation and Prioritization and Catchment Areas

To mitigate negative externalities of projected freight growth, physical and operational improvements to the multimodal transportation network and freight facilities are required. In this context, several multimodal projects compete for limited public and private funding. Three analytic methods are typically utilized to prioritize, compare, and select transportation projects: i) Benefit-Cost Analysis (BCA), ii) Economic Impact Analysis (EIA), and iii) Multi-Criteria Analysis (MCA) (Economic Development Research Group et al., 2014). BCA consists of quantifying project impacts as monetary units and distributing them over time to calculate the present value of all benefits and costs. The results are expressed as a net benefit (benefit minus cost), or as a benefit/cost ratio. A broader version of the BCA accounts for social impacts, including environmental impacts that affect non-travelers. In the EIA, project impacts are measured in terms of their effect on a region's economy. Quantitative measures include business output, job generation, net business income generation, household income, and GDP. In MCA, impacts can be measured either as quantitative indices or as qualitative ratings to portray relative importance. Thus, a broader range of positive and negative impacts may be considered for decision-making (Economic Development Research Group et al., 2014).

All three methods require the impact area of the project to be defined. For instance, according to the Guide for Conducting Benefit-Cost Analyses (BCA) of Multimodal, Multijurisdictional Freight Corridor Investments (Vadali et al., 2017), the first step of a BCA is to define a project by: i) the type of facility or location to be analyzed (whether it is a corridor, a modal or intermodal facility), ii) its impact area, iii) the modes involved, and iv) the nodes involved (i.e. connections to freight network points such as ports, distribution centers, etc.). However, no methodology is proposed to identify impact areas. Moreover, the Guide highlights the need to integrate data sources to perform BCA, because individual data sources do not address all modes (Vadali et al., 2017). Similarly, the EconWorks economic impact assessment tool allows transportation agencies to estimate economic impacts of diverse project types using past projects as case studies (AASHTO, 2015). For each case study, economic impacts are measured in terms of number of jobs, sales, income, and investment. The size and location of the area where the economic impacts are calculated is critical for the study, and described as the "counties in which the project passes, or which are immediately impacted by the project" (AASHTO, 2015). However, the selection of the impact area is not a data-driven analysis; instead, it is a judgement call made by the analyst. Another BCA tool, the Freight module within the BCA tool Transportation Economic Development Impact System (TREDIS) enables users to define a project and identify affected freight flows and associated economic activities. Given a user-defined region, TREDIS profiles the area's freight flow patterns, assesses the supply chain roles of those freight flows, and calculates how emerging economic trends may change future freight flows and investment needs. As output TREDIS allows visualization of county-level, not link or corridor level, freight flows. Arguably the most powerful tool available on the market, TREDIS does not provide network-based analyses and visualization of project-specific freight catchment areas.

# 4.3.2. Vehicle Tracking Data Characteristics

Multimodal catchment areas defined in this paper are based on freight vehicle tracking data, specifically from maritime AIS and truck GPS (Table 4.1). Both sources cover wide geographies, contain population-level data or exist as large samples, and are publicly available, either directly from government sources or through data sharing agreements with private data

providers, making them viable and promising sources for data-driven catchment area

identification as described in this paper.

Dataset characteristics	Waterborne AIS data	Truck GPS data	
Data collection in the U.S.	U.S. Coast Guard.	Private data collection entities such as truck preclearance programs and Electronic Logging Device (ELD) providers.	
Data elements	<u>Vessel</u> : Vessel name, length, width, MMSI, IMO, and call sign. <u>Broadcasting</u> : Latitude, longitude, time stamp, speed over ground, course over ground, heading. <u>Voyage</u> : Cargo, draft, status.	Anonymous truck identifier, position (latitude and longitude), timestamp, heading, spot speed.	
Spatial Coverage	All international and U.S. waterways.	All U.S. territory.	
Temporal Coverage / 'ping' frequency	Aggregated to one-minute intervals (for data storage and sharing purposes)	Every 30 seconds.	
Update frequency	Since 2009, data collected in real time, shared via annual updates.	Since 2002, data collected in real time.	
Data storage format	File geodatabases (.csv) containing one month of data per time zone.	Text files (.txt) containing a time window requested by the user.	
Data sharing scheme	Open source via www.MarineCadastre.gov	Publicly available through agreements with data collection entities.	
LimitationsCarried by tugs and tows on inland waterways, not barges.May lack industries		May lack representability of industries and small fleets.	

 Table 4.1 Vehicle Tracking Data Characteristics

# 4.3.3. Automatic Identification System data (AIS)

AIS data (Table 4.1) is collected for navigational safety purposes (e.g., collision avoidance) and is required by the International Maritime Organization (IMO) for all passengercarrying vessels and commercial vessels over 300 Gross Tonnage that travel internationally. Onboard navigation devices transmit location and characteristics of vessels in real time to receivers on shore, satellite, buoy, and other vessels (U.S. Department of Homeland Security). In the U.S., AIS is mandatory along the Ohio River, between Mileposts 593 and 606, and in the Lower Mississippi River, up to Milepost 254.5 (Dobbins et al., 2013). Even though AIS is not required in all U.S. inland waterways, most vessels use the AIS transponder (DiJoseph et al., 2015). Vessel and voyage features entered to the database manually contain substantial errors and omissions. Broadcasting features, e.g., location (latitude and longitude), time stamp, speed over ground, course over ground, and heading, do not require manual intervention, thus contain few errors, and are used in this work for catchment area definition. In particular, AIS data has the ability to track a vessel's path with time stamps which is suitable to identify freight flows though inland navigable waterways. Although previous studies reconstructed vessel trajectories from AIS data (Graser, 2019; Zhang et al., 2018; Zhao et al., 2018), they are limited in the lack of an inland waterways network setting, or in that movements are divided per day, masking the identification of trips. DiJoseph and Mitchell (2015) overcome the latter by linking consecutive AIS records together to generate paths on inland waterways; however, they did not fuse generated vessel paths with a defined network. The inability to map vessel data to a network precludes future integration of AIS data into multimodal, network-based models, such as state TDMs. In contrast, the algorithm applied in this work allows for the identification of trips defined by origin and destination (not duration), and matched to a defined intermodal network.

# 4.3.4. Truck GPS data

Truck GPS data consists of vehicle positioning data (latitude and longitude) emitted by onboard GPS devices. Spatial coverage in the US is almost ubiquitous (Short, 2014). Private truck fleets typically record positioning data of their own trucks for security, route tracking, fuel cost, and other operational analyses. Data providers typically share anonymous (no identification of industry, operator, company, etc.) truck GPS data gathered from a sample of private fleets.

Truck GPS data has been used for bottleneck identification, travel time analyses, border crossings, truck parking, hours of services tracking, etc., and is a valuable source of truck routing, time-of-day usage, volume and speed data (Laranjeiro et al., 2019; Short, 2014). Truck GPS data covers every single road in the statewide network, while other truck data sources, such as static sensors like Weigh-in-Motion (WIM), inductive loop detectors, or temporary tube counters, are restricted to fixed and few locations. Like AIS data, it is necessary to employ geospatial fusion methods to map GPS traces to a defined transportation network. Methods for map matching and route identification for truck GPS data have been carried out in several prior studies (Camargo et al., 2017; Ciscal-Terry et al., 2016; Hashemi et al., 2014).

# 4.3.5. AIS and Truck GPS data fusion

The challenge in fusing truck and vessel tracking data is overcoming data heterogeneity in units of time, space, and context. Xu et al. (2017) developed a Generic Target Monitoring System (GTMS) to monitor multimodal vehicles, and tested it with AIS and truck GPS data collected at a sea port terminal. To overcome multimodal data heterogeneity, vehicle tracking data from different sources (i.e. truck, vessel) was converted to a uniform data format. A GIS web-based interface allowed users to visualize and analyze real-time and historical multimodal vehicle tracking data within a designated geographical area (Xu et al., 2017). Meyer-Larsen et al. (2015) combined real-time AIS and truck GPS data to improve the efficiency of logistics at container terminals. The system tracked container vessels positions from AIS data to estimate vessels estimated time of arrival (ETA) and compared it with the ETA manually entered by the vessel operator. The system automatically detected deviations between planned and scheduled ETA and communicated potential deviations in real time to port stakeholders (including truck operators), so they could schedule operations in response to vessels' delays (Meyer-Larsen et al., 2015). Monsreal et al. (2019) performed statistical analyses to determine vessels and truck activity correlations and causalities at coastal ports using AIS data from non-liquid carrying vessels acquired from a vendor, and truck GPS probe data from the National Performance Management Research Data Set (NPMRDS). The analysis was complemented with census and port administration datasets. The analyses produced coefficients representing changes in directional road traffic volumes corresponding to changes in import/export freight volume (measured in weight), and the time when those increments on road traffic were expected. For example, unloading of a vessel with 1,000 TEU would increase traffic along an inland highway by approximately 500 trucks during the week the vessel arrives, and decrease by approximately 400 trucks two weeks later (Monsreal et al., 2019). Overall, these studies were limited by the lack of: i) a systematic, data-driven procedure to identify multimodal freight port catchment areas, ii) network assignment procedures for AIS data, and iii) multimodal data fusion approaches applied to inland waterway transportation. The method presented in this paper overcomes these limitations by characterizing spatial patterns of freight intensity that exceed the circular perimeter (radial buffer) typically used for catchment analysis by explicitly assigning truck and vessel flows to defined multimodal networks.

## 4.4. Methodology

The "catchment" area of a multimodal freight transportation infrastructure project is defined as the region where such facility draws and delivers freight, which can be visualized as the paths followed by vehicles, vessels, railcars, etc. accessing the facility. The methodology consists of the following steps (Figure 4.1):

*Step 1: Data Preparation-* The vehicle tracking data of each of the two modes are independently subjected to a quality control process to remove erroneous or irrelevant records.

*Step 2: Data Analysis-* Each vehicle tracking dataset is independently subjected to stop identification and map-matching procedures to define locations and duration of stops, and connected paths on a defined transportation network. The stop identification and map-matching algorithm used for both vessels and trucks are similar but with mode-specific parameters. For each mode, the outputs of the map-matching algorithm are: i) a table listing the nodes of the network visited by each vehicle during its trips, and ii) a table listing the trips made by all the vehicles, identifying the origin and destination of each trip (by network node, traffic analysis zone, and port if applicable).

*Step 3: Multimodal Data Fusion: Visualization and Quantification-* The output of the map-matching algorithm applied to each mode is post-processed to identify trip paths to/from a specific freight facility, depicting its mode-specific project impact area. Then, the two mode-specific impact areas are super-imposed into a single map to depict the multimodal impact area. Several mode-specific and combined indicators are derived to quantify and compare the impact area: i) impact area size, ii) Vehicle Miles Travelled (VMT), iii) Vehicle Hours Traveled (VHT), iii) population within the impact area, iv) number of business registered within the impact area, and iv) number of unique traffic analysis zones (TAZ) as origin or destination of trips associated to each facility.

Waterway data analysis

Roadway data analysis



Figure 4.1 AIS and truck GPS data fusion methodology flowchart

## 4.4.1. Step 1- Data Preparation

The purpose of this step is to remove low-quality records from the AIS and truck GPS datasets. Unprocessed AIS and anonymous truck GPS data may contain erroneous or irrelevant records due to transponder issues, transmission obstructions, etc. Erroneous GPS records refer to records with unusual high speed (i.e. cargo vessels travelling at 70 mph), or located far from the transportation network (i.e. vessels far away from inland waterways or heavy trucks away from roadways). Irrelevant records come from vehicles that emitted less than 20 records within the reporting period. These records are removed from the datasets. For example, geospatial tools are used to remove records outside a buffer representing the navigable waterways.

# 4.4.2. Step 2- Data Analysis: Stop Identification and Map Matching

The purpose of this step is to reconstruct the vehicle and vessel paths observed from the GPS/AIS position data using mode-specific networks, also known as map-matching. From the several map-matching algorithms available (Camargo et al., 2017; Hashemi et al., 2014), Camargo et al. (2017) was used because it has the advantage of wide applicability to multimodal data sets<sup>4.1</sup>. Akter and Hernandez (Akter et al., 2018) adapted the algorithm to statewide truck GPS samples, while Asborno and Hernandez (Asborno et al., 2020) adapted it to vessel movements on inland navigable waterways.

AIS and truck GPS datasets were subjected independently to the adapted map-matching algorithm. First, stops made by each vehicle are identified by iterating through temporally consecutive location records (i.e., GPS "pings" or latitude-longitude-timestamp points). A naïve approach to find stops would be to locate all zero speed pings. However, assuming that a vehicle

<sup>&</sup>lt;sup>4.1</sup> Camargo's algorithm is written in Python 2.7 and openly available at https://github.com/pedrocamargo/map\_matching

emits a signal every few seconds even when stopped, several consecutive pings with low or zero speed likely represent a single stop and the naïve method would thus over count stops and misrepresent stop duration. The stop identification algorithm instead defines stop clusters based on parameters defining point speed thresholds, minimum duration, and maximum geospatial stop coverage. For each mode, the stop identification algorithm outputs a list of stops made by all vehicles, indicating: anonymous vehicle/vessel identification number, a generated stop identification number, time when the stop occurred, and its location coordinates (longitude and latitude).

In parallel, the algorithm identifies all links of the network that are likely used by the vehicle as it travels between stops. Using geospatial analysis, each ping is associated with a network link if its location falls within a pre-defined buffer distance from the link. Links with pings within their buffers are likely used by the vehicle as it traveled between stops. In some cases, a vehicle/vessel may traverse many links between ping recordings, thus the map-matching algorithm reconstructs the complete path of consecutive links using shortest path algorithms. For each mode, the map-matching algorithm outputs a sequenced list of network nodes visited by each vehicle, the time when the vehicle arrived and left each node, and its associated network link.

# 4.4.3. Step 3- Multimodal Data Fusion: Visualization and Quantification

The purpose of this step is to visualize the trips made by all vehicles to/from a given freight facility on a multimodal network. Freight port terminals along the U.S. inland waterways, selected from a publicly available database, are specified by location (latitude, longitude) and commodities handled (if any) (Bureau of Transportation Statistics, 2019). Additional freight port terminals not included in the database were found in a similar fashion than Joubert and Axhausen

(2013) did to identify commercial facilities visited by trucks. This work extends Joubert and Axhausen's (2013) work to inland waterway transportation facilities as follows. First, we select nation-wide facilities that serve freight from (Bureau of Transportation Statistics, 2019), and then define a bounding box delimiting the area of study to identify only those freight facilities within our study area. Later, we super-impose the results of the AIS stop identification algorithm (step 2) to the port terminal locations and identify clusters of stops not associated with any port terminal location (i.e. stop clusters located outside a buffer area around each port location). Those clusters constitute potential locations of port terminals, docks, or loading/unloading areas, which were not included in the initial port database. These potential locations are verified using aerial imagery. For the case study of Arkansas, three ports are added to the initial port database following this approach. The commodities handled at those three freight ports are deducted by observing their storage areas on the aerial imagery, and through a web-search of publicly available data about those facilities.

To visualize the vehicle trips to and from such freight facilities, first, for each mode, stop identification, and map-matching results are joined to create a list of all vehicle and vessel trips that accessed each facility. Trip data include the sequence of network nodes visited by each vehicle, the time when the vehicle arrived and left each node, its associated network link, and the trip identification number. To add geometry for visualization purposes, the trip data is joined to with the corresponding mode-specific transportation network based on network link attributes (e.g., link ID). This also allows for estimation of VMT and VHT performance measures as the network link attributes include distance and free flow travel time. As a result, for each freight facility, the geometry of vehicle and vessel trips to/from the freight facility is produced and includes length (miles), duration (hours), origin, and destination (network node, TAZ, and port

ID when applicable). GIS software is then used for visualization on catchment areas by mode, and each mode is super-imposed to visualize the multimodal catchment area.

To complement visual depictions of freight catchment areas, key quantitative indicators (Table 4.2) are calculated per mode and by combining all modes (e.g., multimodal). The indicators constitute performance of the freight activity associated with each facility. The impact area size, population, number of business within the impact area, and location of unique TAZs serving as the origin or destination of trips to/from each port are derived using statistical packages and modeling tools in GIS platforms. The VMT and VHT corresponding to all trips to and from each port are calculated by aggregating the trip length (in miles) and duration (in hours) for all the trips with origin or destination in the said port.

 Table 4.2 Key Performance Indicators of Each Freight Facility, Measured From its Impact

 Area

Key indicators	Units of measurement	
Impact area size	Acres	
Vehicles Miles Travelled (VMT) VMT <sub>multimodal</sub> = VMT <sub>truck</sub> + VMT <sub>vessel</sub>	Aggregated miles for all vehicles; Percentage of VMT per mode	
Vehicles Hours Travelled (VHT)	Aggregated hours for all vehicles;	
Population within the impact area <sup>a</sup>	Number of individuals	
Number of business registered within the impact area <sup>b</sup>	Number of businesses	
Number and location of unique TAZ as origin or destination of trips to/from each facility	Number of TAZs	

Notes: (a) Might be stratified per population characteristic. (b) Might be stratified per commodity (NAICS code).

# 4.5. Case Study: Impact Areas of Port terminals on the Arkansas River

# 4.5.1. Scope

The methodology was applied to 43 freight port terminals located on the Arkansas River (Figure 4.2) with AIS and truck GPS data samples from 2016. The Arkansas River is a 308-mile stretch of navigable waterway that plays a key role in the national economy by connecting the heartland of the U.S. to the international markets via the Mississippi River, and contributes to the national economy with 4.5 B USD in sales, 34,000 jobs, and 168MUSD in taxes (Nachtmann et al., 2015).



Figure 4.2 Study area: 43 freight port terminals on Arkansas River. Labels represent

# names of municipalities

# 4.5.2. Data

The AIS data in this case study consisted of over 3,390,000 records, emitted by a sample of 765 unique vessels observed along Arkansas during 2016. The raw data from which this sample was extracted is available for free public download in (Office for Coastal Management, 2018). Truck GPS data acquired from a non-profit trucking industry research firm corresponds to over 4 million "pings" emitted by approximately 40,000 unique trucks within a 10-mile buffer around the state of Arkansas, during four two-week periods in 2016: February, May, August, and November. Within Arkansas, the truck GPS data represents a sample of about 10% of the truck gPS pings are observed at port facilities with varied levels of concentration. Heavier concentration of truck GPS activity is observed in the Port of Little Rock and in the Port of Pine Bluff, while a relatively lower concentration is observed at all other ports.

Information about the commodities handled by each port were gathered from the National Transportation Atlas Database (Bureau of Transportation Statistics, 2019). Population data by census tract in Arkansas was obtained from the Census Bureau Topologically Integrated Geographic Encoding and Referencing (TIGER). Business location data was obtained from ESRI, and consisted of a geocoded list of more than 200,000 establishments registered in Arkansas, including name, location, and North American Industry Classification System (NAICS) code, among others.

# 4.5.3. Applied Methodology

The parameters adopted for the stop identification and map matching algorithms (Table 4.3) were obtained by comparing the number of total stops and trips identified by the algorithm to control vehicles subjected to manual verification of stops and paths. Mode-specific parameters

were identified independently for the unprocessed GPS and AIS traces (Akter et al., 2018; Asborno et al., 2020).

Stop Parameter / Mode	Maritime	Roadway
Stopped speed (km/h)	5.3	4.8
Minimum time stopped (seconds)	300	300
Maximum stop coverage (km)	5.0	0.3
Geoprocessing buffer size (degrees)	0.01-0.1	0.001

 Table 4.3 Multimodal Stop Identification Parameters

# 4.5.4. Results

Four of the 43 ports in the case study were selected for detailed analysis based on their commodity and location diversity. Commodities range from diesel fuel (liquid bulk), food and farm, and other dry bulk quarry products, to steel structures and construction equipment. The impact areas (Figure 4.3) and corresponding quantitative impact measures (Table 4.2 and Figure 4.4-Figure 4.8) show the variability in impact scale and scope observed for each port. TAZs used for this analysis correspond to the Arkansas State TDM. The 5,849 zones within the state follow the boundaries of aggregated census blocks, while out-of-state zones match the Transearch data and Business Economical Area (BEA) districts, totaling 306 U.S. BEAs (without Arkansas) (Alliance Transportation Group, 2015). Thus, TAZs within Arkansas represent areas much smaller than out-of-state TAZs.



a. Port in Pine Bluff shipping food and farm products



c. Port in North Little Rock shipping wholesale diesel fuel



b. Port in Pine Bluff shipping steel structures and construction equipment



d. Port in Van Buren shipping dry bulk quarry products



Figure 4.3 Multimodal impact areas for freight port terminals located on the Arkansas

River, 2016. Truck paths (red) were not available out-of-state. Green circles delimit an

arbitrary 100-mile diameter area around each facility



Figure 4.4 Size of multimodal and mode-specific impact areas, 2016



Figure 4.5 VMT and VHT of multimodal impact areas, 2016



a. Port in Pine Bluff shipping food and farm products. 29 unique TAZs as origin or destination of trips from/to this port



c. Port in North Little Rock shipping wholesale diesel fuel. 50 unique TAZs as origin or destination of trips from/to this port



b. Port in Pine Bluff shipping steel structures and construction equipment. 21 unique TAZs as origin or destination of trips from/to this port



d. Port in Van Buren shipping sand, gravel, and other dry bulk quarry products. 28 unique TAZs as origin or destination of trips from/to this port



Figure 4.6 Unique TAZs as origin or destination of multimodal trips to and from port

terminals, 2016



Figure 4.7 Number of registered business within the multimodal impact areas, 2016



Figure 4.8 Population within the multimodal impact areas, 2016

## 4.6. Discussion

### 4.6.1. Insights to Port Activity by Region

Multimodal freight activity per port can be used to target policies and drive investment. For example, in terms of freight corridor planning, different modes (maritime and roadway) compete along parallel shipping lines, such as the corridors connecting Little Rock (Figure 4.3.a) and Dardanelle (Figure 4.3.b) with Pine Bluff. Such areas of modal competition may indicate a potential for modal shift in those regions. Further analysis on the quantities of the specific commodities travelling on those routes may guide targeted policies to incentivize shifts from truck to water on those routes. In contrast, complementary modal interaction can also be observed in the usage of the river for local, short trip deliveries (Figure 4.3.d). Multimodal interactions observed in these visualizations provide evidence on the use of the river for domestic shipping of steel structures and sand (Figure 4.3.b and Figure 4.3.d, respectively), and the key role of the Arkansas River to connect the U.S. Midwest with international markets through the Mississippi River for shipping farm products (Figure 4.3.a). Modal competition is also observed in the performance metrics quantified by this analysis (Figure 4.4). For each port, the bigger the difference between the multimodal area (left bar) and the sum of the individual modal areas (right bar), the longer the corridor(s) where the modes compete.

In terms of VMT (Figure 4.5), port terminals are dominated by truck travel as compared to maritime. VMT is typically used to calculate emissions and other environmental impacts. To evaluate cost-benefit ratios of modal shift policies (from truck to vessels), for example, estimates of VMT by mode are a necessary externality to measure. To complement this analysis, the authors are exploring methods to estimate commodity ton-miles transported per mode and port, which would provide information to more detailed cost-benefit analyses.

# 4.6.2. Catchment Area Comparisons

The size and extent of each port terminal catchment area varies significantly by port (Figure 4.3 and Figure 4.4), and thus it would not be appropriate to adopt a generalized, arbitrary impact area as shown by the 100-radius circles around freight facilities. The difference between the arbitrary radial areas and the multimodal impact areas derived from vehicle tracking data (Figure 4.3) indicate the extent of the freight activity that would be ignored if arbitrary radial impact areas were utilized to estimate port activity. The unique impact areas could not be visualized by relying solely on surveys or static traffic data. Even though truck trip paths (and thus, areas of impact) may be visualized from the output of a travel demand model, such models are based, in large part, on survey data as well. Waterway trip paths cannot be visualized from travel demand models that do not represent the navigable waterway network. In this context, vehicle tracking data provides a viable alternative to the outputs of state travel demand models to analyze multimodal freight catchment areas for project evaluation and prioritization.

Lastly, the number and location of unique TAZs that constitute the origin and destination of trips to and from each port (Figure 4.6), derived from multimodal vehicle tracking data, can be used to support long-range transportation planning purposes, such as scenario planning. Scenarios simulating disruption of business in those zones might impact the port terminal economic activity, and vice-versa. For example, a severe weather event such as a flooding affecting a port in Little Rock (Figure 4.3.c), located in the center of the state, may have an impact on freight flows observed as far as Northwest Arkansas, encompassing a total area of 10,500 thousand acres (Figure 4.4). While an event affecting traffic flows in Northwest Arkansas, such as an accident at a highway/rail crossing, may have an impact on the economic activity of a port located as far as Little Rock.

### 4.6.3. Transferability and Future Work

In terms of transferability, the same analysis applies to any transportation infrastructure for which geospatial location is available, such as bridges, intermodal connectors, storage and warehousing facilities, rail crossings (provided rail tracking data, which is proprietary of rail operators, is available), etc. Moreover, it could be applied to any area of interest, such as parcels corresponding to a specific land use, the location of a specific industry (i.e. an inland petrol refinery, a forestry industrial area, etc.), and evaluate its area of impact.

Building upon the methodology, future work may measure the interaction between land use and freight transport by evaluating specific impacts. Overall, the novel multimodal freight data analysis constitutes a sound basis to characterize spatial patterns of freight intensity to/from specific land-use parcels. Moreover, the identification and visualization of impact areas of different freight ports (or any other piece of infrastructure) are subject to the same data and criteria to identify their multimodal area of impact, providing a common basis for proper comparison and competition of funds.

#### 4.7. Conclusion

Vehicle tracking data, namely AIS and truck GPS data, provide ubiquitous and consistent sources to identify multimodal freight paths to and from freight facilities and specific land-use parcels, such as ports. The methodology in this paper consists of multimodal freight data collection and analysis that allows characterization of spatial patterns of freight intensity. By matching vehicle tracking data to mode-specific networks, and selecting the trips with origin or destination within bounding boxes surrounding a freight facility, the resulting freight paths illustrate the impact areas of an investment in such facility. In addition, any events affecting the transportation infrastructure within the catchment area of the facility will influence the use of the

freight facility. For example, the impact area of an extreme weather event such as flooding of a road or port could be identified for resiliency evaluation. Moreover, the identification and visualization of the geographic extent of multimodal freight catchment areas can be used to estimate population exposure statistics, such as exposure to emissions, by super-imposing census and business locations to the catchment areas.

Within the context of transportation infrastructure investment, several projects compete for a limited amount of resources, based on an estimation of project benefits relative to costs. To evaluate project benefits, it is important to understand the extent, location, and characteristics of a project's impact area, or "catchment" area, which can be defined as the region where the facility draws and delivers freight, or the OD pairs served by the facility. However, little has been written regarding systematic methods to identify multimodal catchment areas. State-of-the practice methods to identify the impact area of a facility consist of arbitrarily selecting a radial perimeter around the facility, ignoring complex interactions among freight modes and supply chains. Alternatively, freight paths to and from a facility may be obtained from project specific data like surveys, which are not always comparable among projects, and/or from travel demand models which have imbalanced or non-existent multimodal network representations. The main contribution of this paper is a geospatial data fusion method to identify the impact area of a multimodal freight project by using increasingly ubiquitous vehicle tracking data. In this way, all projects evaluated are subject to the same data and criteria to identify their impact area, providing a common basis for proper comparison and competition of funds.

A case study to illustrate the value of identifying multimodal freight catchment areas highlights the differing size and shape of port impact areas, further supporting limitations presented by the naïve assumption of radial impact areas for freight facilities. Examples of inland

waterway ports in Arkansas show that modes compete on the same freight corridors, presenting an opportunity for mode shift. Since the AIS and Truck GPS data are increasingly available worldwide, the methodology has wide applicability to broad geographies and facility types.

# 4.8. Acknowledgements and Disclaimer

The authors thank the Maritime Transportation Research and Education Center (MarTREC), a tier 1 University Transportation Center by the U.S. Department of Transportation, for sponsoring this paper under "Combining truck and vessel tracking data to estimate performance of inland ports"; the Arkansas Department of Transportation for acquiring the truck GPS data under "Truck activity analysis using GPS data" (TRC1702); and Taslima Akter for processing truck GPS data.

This material is based upon work supported by the U.S. Department of Transportation under Grant Award Number 69A3551747130. The work was conducted through the Maritime Transportation Research and Education Center. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The work is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

# 4.9. References

- AASHTO (Producer). (2015, August 5). EconWorks Case Study Development Training. *Module* 5 - Case Study Data Needs and Sources. Retrieved from https://planningtools.transportation.org/425/module-5---case-study-data-needs-andsources.html
- Akter, T., Hernandez, S., Corro-Diaz, K., et al. (2018). Leveraging Open Source GIS Tools to Determine Freight Activity Petterns from Anonymous GIS data. Paper presented at the

AASHTO GIS for Transportation Symposium. https://gis-t.transportation.org/wp-content/uploads/sites/51/2018/05/GIS\_T\_Taslima\_Akter\_svh\_TA.pdf

- Alliance Transportation Group. (2015). Arkansas Statewide Travel Demand Model Documentation.
- Asborno, M., Hernandez, S., & Yves, M. (2020). Network Mapping of AIS Data to Characterize Inland Waterway Freight Transportation - working paper. *Freight Transportation Data Research Lab*. Retrieved from https://sites.uark.edu/sarahvh/
- Bureau of Transportation Statistics. (2019). *National Transportation Atlas Database Ports*. Retrieved from https://data-usdot.opendata.arcgis.com/datasets/ports
- Camargo, P., Hong, S., & Livshits, V. (2017). Expanding the Uses of Truck GPS Data in Freight Modeling and Planning Activities. *Transportation Research Record*, 2646(1), 68-76. doi:https://doi.org/10.3141/2646-08
- Carroll, M., Carlyle, C., & Seman, M. (2017). *The Economic and Fiscal Impacts of Development near DART Stations*. Retrieved from https://www.dart.org/about/economicdevelopment/May2017DARTEconomicandFiscalIm pacts.pdf
- Chacon Hurtado, D., Bardaka, E., Yang, R., et al. (2016, June 26-29). Transportation and Economic Development: An Assessment of the SHRP2 C11 Tools and Their Synergies with Economic Simulation Models for the Evaluation of Wider Economic Benefits of Transportation Projects. *ASCE International Conference on Transportation & Development*.
- Ciscal-Terry, W., Dell'Amico, M., Selini Hadjidimitriou, N., et al. (2016). An Analysis of Drivers' Route Choice Behaviour Using GPS Data and Optimal Alternatives. *Journal of Transport Geography*, 51, 119-129.
- DiJoseph, P., & Mitchell, K. (2015). *Estimating Vessel Travel Time Statistics for Inland Waterways with Automatic Identification System Data*. Paper presented at the Transportation Research Board 94th Annual Meeting, Washington, DC.
- Dobbins, J., & Langsdon, L. (2013). Use of Data from Automatic Identification Systems to Generate Inland Waterway Trip Information. *Transportation Research Record*, 2330(1), 73-79. doi:https://doi.org/10.3141/2330-10
- Donnelly, R., Upton, W., & Knudson, B. (2018). Oregon's Transportation and Land Use Model Integration Program: a Retrospective. *The Journal of Transportation and Land Use*, 11(1), 19-30.
- Economic Development Research Group, Cambridge Systematics, ICF International, et al. (2014). *Development of Tools for Assessing Wider Economic Benefits of Transportation*. Retrieved from http://www.trb.org/Main/Blurbs/169524.aspx

- Graser, A. (2019, January 26). Movement Data in GIS. #19: splitting trajectories by date. Retrieved from https://anitagraser.com/2019/01/26/movement-data-in-gis-19-splitting-trajectories-by-date/
- Hashemi, M., & Karimi, H. (2014). A Critical Review of Real-time Map-matching Algorithms: Current Issues and Future Directions. *Computers, Environment and Urban Systems*, 48, 153-165.
- Hernandez, S., Williams, B., Williamson, M., et al. (2019). *Truck Activity Analysis Using GPS Data*. Retrieved from https://cpb-use1.wpmucdn.com/sites.uark.edu/dist/e/220/files/2019/08/TRC1702-Final-Report\_v8-FINAL.pdf
- Joubert, J., & Axhausen, K. (2013). A Complex Network Approach to Understand Commercial Vehicle Movement. *Transportation*, 40, 3, 729-750. doi:10.1007/s11116-012-9439-0.
- Laranjeiro, P., Merchán, D., Godoy, L., et al. (2019). Using GPS Data to Explore Speed Patterns and Temporal Fluctuations in Urban Logistics: The Case of São Paulo, Brazil. *Journal of Transport Geography*, 76, 114-129.
- Meyer-Larsen, N., Müller, R., & Köhler, T. (2015). *Optimisation of Intermodal Transport Using Satellite-based Services*. Paper presented at the Operational Excellence in Logistics and Supply Chains.
- Monsreal, M., Kruse, J., Kang, D. H., et al. (2019). *Developing and Applying an Analysis Methodology to Identify Flow Generation Influences between Vessel and Truck Shipments*. Retrieved from https://doi.org/10.5281/zenodo.3165953
- Nachtmann, H., Boudhoum, O., & Oztanriseven, F. (2015). *Regional Economic Impact Study for the McClellan-Kerr Arkansas River Navigation System*. Retrieved from https://martrec.uark.edu/research/regional-economic-impact-study-for-the-mcclellankerr-arkansas-river-navigation-system.pdf
- NADO Research Foundation. (2011). *Inland Ports: Linking Freight and Regional Economic Development*. Retrieved from https://www.nado.org/wp-content/uploads/2011/10/Inland\_report\_FINAL.pdf
- Office for Coastal Management (Producer). (2018, 04). Vessel Traffic Data. *MarineCadastre.gov*. Retrieved from https://marinecadastre.gov/ais/
- Short, J. (2014). *Truck GPS Data for Tracking Freight Flows*. Paper presented at the Developing Freight Fluidity Performance Measures.
- TREDIS (Producer). Products. *TREDIS Freight*. Retrieved from https://www.tredis.com/products/tredis-freight

- Tyndall Air Force Base. (2018). 2017 Economic Impact Analysis. Retrieved from https://www.tyndall.af.mil/Portals/107/documents/AFD-180604-018.PDF?ver=2018-06-04-154005-160
- U.S. Department of Homeland Security (Producer). Automatic Identification System. *Navigation Data Center*. Retrieved from https://www.navcen.uscg.gov/?pageName=AISmain
- Vadali, S., Kruse, C. J., Kuhn, K., et al. (2017). NCFRP Report 38: Guide for Conducting Benefit-Cost Analyses of Multimodal, Multijurisdictional Freight Corridor Investments. Retrieved from http://www.trb.org/Publications/Blurbs/175606.aspx
- Weisbord, G., & Reno, A. (2009). *Economic Impact of Public Transportation Investment*. Retrieved from http://onlinepubs.trb.org/onlinepubs/tcrp/docs/TCRPJ-11Task7-FR.pdf
- Xu, K., Zheng, H., Li, Y., et al. (2017). Comprehensive Monitoring System for Multiple Vehicles and Its Modeling Study. *Transportation Research Procedia*, 25, 1824-1833.
- Zhang, L., Meng, Q., Xiao, Z., et al. (2018). A Novel Ship Trajectory Reconstruction Approach Using AIS Data. *Ocean Engineering*, 159, 165-174.
- Zhao, L., Shi, G., & Yang, J. (2018). Ship Trajectories Pre-processing Based on AIS Data. *The Journal of Navigation*, 71, 1210-1230.

# CHAPTER 5. Multi-Commodity Port Throughput from Truck GPS and Lock Performance Data Fusion

### 5.1. Abstract

Inland waterways ports are key elements of an efficient multimodal freight transportation system. Data on the capacity and throughput of inland waterway ports by commodity supports effective long-term freight planning and travel demand modeling. More specifically, such data can be used to estimate multimodal, commodity-based freight fluidity performance measures and to support location selection for freight transload facilities. State-of-the-practice means of obtaining commodity flows data, such as shipper/carrier surveys and vessel and vehicle movements are limited in their ability to provide monthly or seasonal statistics on individual port operations, rather, they provide annualized statistics for river segments which may contain multiple ports. Our work addresses these limitations by developing a Multi-Commodity Assignment Model to quantify commodity throughput at inland waterways ports. The model fuses waterborne Lock Performance Monitoring System (LPMS), which provides the commodity dimension, and anonymous truck Global Positioning System (GPS) data, which allows for spatial disaggregation. A goal programming approach minimizes the deviation between known and estimated truck flows at each port. The methodology was applied to the Arkansas River, a 308mile navigable waterway served by 14 locks and 43 freight ports. Overall, 84% of ports had less than 20% difference between observed and predicted truck flows. The model is applicable to any inland waterways with aggregated commodity flow data and truck GPS coverage, and fills a critical data gap by describing commodity throughput at inland waterway ports using publicly available data.

**Key words:** Multimodal Transportation, Data Fusion, Commodity Flow, Inland Waterways Transport

#### 5.2. Introduction

Along America's Marine Highway Network, inland waterway ports are critical connections among transport modes, making them key elements of an efficient multimodal freight transportation system. While data about the commodity flows through each port may be collected by port operators, it is proprietary and is not regularly shared with public agencies due to privacy concerns. Publicly available maritime port statistics, such as the Port Performance Freight Statistics Program introduced by the U.S. Bureau of Transportation Statistics, are limited to the top-25 ports in the US (U.S. Department of Transportation, 2019), which tend to exclude inland waterway ports. State-of-the-practice means of gathering port commodity flows data include economic surveys (e.g. U.S. Census Bureau's U.S. Port Data (2019)), surveys targeting freight flows (e.g. Commodity Flow Survey (CFS)), and mode specific datasets (e.g. National Performance Management Research Data Set, or Waterborne Commerce Statistics). Such sources are limited in their spatial disaggregation, temporal continuity, and multimodal integration. For example, CFS is carried out every five years, and considers only 132 geographical zones within the U.S., where the state of Arkansas constitutes a single zone. Such spatial aggregation is not suitable to describe commodity flows at port level. In addition, surveys carried out every a number of years impose the need to estimate annual volumes. In contrast, this paper produces annualized data from sources that are continuously updated in terms of temporal coverage. To address these limitations, this paper presents a methodology to quantify commodity flows through inland waterway ports by fusing two mode-specific datasets in a Multi-Commodity Assignment Model.
Our motivation to quantify and describe inland waterways port throughput stratified by commodity type was two-fold. First, given the complexity of the multimodal freight transportation system, there has been increased interest in developing multimodal "freight fluidity" indicators that capture end-to-end supply chain performance (Transportation Research Board, 2014). The term freight fluidity is a measure of the ease at which freight (in quantities of tonnage or volume) can move through the multi-modal supply chain. Freight fluidity is often measured in travel time, travel time reliability, or transportation costs. Most importantly, though, it should reflect performance of all modes within the supply chain. Freight fluidity measures require different types of data (e.g., movements, transactions, cost, commodity type) from a variety of sources (e.g., government databases, private industry). The data is intended to evaluate mobility, reliability, resilience, cost, and quantity of freight in a multimodal transportation network (Eisele et al., 2016). Multimodal freight fluidity indicators require not only mode-specific data, but an understanding of the interaction between individual modes (Transportation Research Board, 2016). To date, most modal interactions are captured by fusing mode-specific datasets via demand models, visualization tools, etc. (Hwang et al., 2016; IHS / Global Insight, 2011; Parker, 2019). For example, the FHWA National Freight Fluidity Monitoring Program combines waterborne data from the U.S. Army Corps of Engineers (USACE), railway data from TransCore and the Carload Waybill Sample, highway data from the National Performance Management Research Data Set (NPMRDS), and supply-chain data from U.S. private companies to generate a mapping tool to track the reliability, cost, and travel time (but not quantities) for multimodal freight movements across selected supply chains on a quarterly basis (Parker, 2019). Given the historical mode-specific approach to freight data collection and analysis, challenges remain to collect and analyze multimodal data for freight

fluidity purposes (Transportation Research Board, 2018). The purpose of this paper is to create methods for fusing multi-modal freight data in an effort to quantify and describe port level commodity flows along inland waterways. Knowledge of port-level throughput, and linkage between waterborne and roadway freight flows, resulting from the methodology presented in this paper, supports the development of commodity-specific, multimodal freight fluidity performance measures.

Second, a quantitative description of current and historical port throughput provides a basis for long-term port cargo projections, which are essential for port facilities, infrastructure, and location planning (de Langen et al., 2012). Previous freight facility location studies were limited by the spatial aggregation of commodity data at county level (Asborno et al., 2018). In this sense, the disaggregated approach proposed here helps to overcome such limitations. Moreover, for public agencies, policy and incentive programs for port development can be guided by information on commodity specific port growth. For the private sector, quantification of port throughput by commodity allows industry stakeholders to assess business opportunities and potential more efficient shipping solutions. This paper provides a generalizable, data-driven method to quantify commodity throughput for inland waterway ports, a measure that is currently not directly publicly available as most inland ports are not covered in the BTS top-ranked port lists.

In this work, inland waterborne commodity flows between locks were spatially disaggregated to each port by fusing publicly available USACE's Lock Performance Monitoring System (LPMS) data with anonymous truck Global Positioning System (GPS) data. Commodity flows on the U.S. inland waterways were quantified using LPMS, which contains monthly volumes carried by vessels using USACE maintained locks for each of the 40 commodities, and

is aggregated into nine commodity groups (U.S. Army Corps of Engineers, 2016). Locks and dams are located along several river navigation systems in the U.S. including the McClellan Kerr-Arkansas River Navigation System (MKARNS), the Upper Mississippi River, Illinois Waterway, and Ohio and Tennessee rivers (Figure 5.1.a) (U.S. Army Corps of Engineers). Navigation locks and dams maintain slack water pools for year-round navigation along inland waterways (Figure 5.1.b). Although LPMS provides commodity volume through the locks, commodity flows through inland waterway ports are not immediately available, as several ports are located between each pair of consecutive locks. The Arkansas portion of the MKARNS, for example, consists of 308 miles of channel controlled by 14 locks and serves 43 freight ports (U.S. Army Corps of Engineers).



a. Location of U.S. locks





To mine freight fluidity insights, LPMS data is used in conjunction with other data sources. Campo, Mayer and Rovito (2012) evaluated the resilience of inland waterways transport in terms of port (un)loading capabilities during catastrophic closures, and applied the method to a segment of the Upper Mississippi river, between six consecutive locks. Commodity volumes in the study area were collected from the LPMS. However, spatial disaggregation of commodity volumes to each port was not possible due to the limitations of the LPMS previously mentioned (e.g., LMPS corresponds to locks, not to ports) (Campo et al., 2012). Dobbins and Langsdon (2013) generated inland waterway towboat trips from Automatic Identification Systems (AIS) data, and used LPMS to compare the number of lockages reported by USACE with the number of towboats operating with AIS receivers (Dobbins et al., 2013). Thoma and Wilson (2005) proposed a model to forecast annual waterborne freight of coal, and food and farm products, using LPMS commodity volumes on key locks and historical waterborne commerce statistics data, but did not disaggregate it to port-level (Thoma et al., 2005). In all, the methodologies and applications of the above-mentioned studies were hindered by the inability to disaggregate lock to lock commodity flows to port-level flows. Our work uses LPMS commodity-specific volumes rather than vessel counts (e.g., as was done by Dobbins and Langsdon, 2013) to add the required commodity dimension to quantify port throughput, thus overcoming the lack of lock to port disaggregation in Campo et al. (2012).

The anonymous truck GPS data used in this work was gathered by the American Transportation Research Institute (ATRI). The data contained location information over time for a sample of the truck population but did not contain commodity information. Truck GPS data was used to quantify truck volumes at each port, which, in turn, allowed for distribution of LMPS commodity volumes to ports. Truck GPS data of the type used in our work has been used in similar contexts. Bartholdi et al. (2019) used truck GPS data within container terminals to measure service times (Bartholdi et al., 2019). Pinjari et al. (2016) explored the use of truck GPS data to analyze trajectories, e.g. travel paths of petroleum-tanker trucks between Florida ports and delivery locations. They applied spatial heuristics (e.g. geographic bounding boxes) to fuse two consecutive months of truck GPS trajectories with geocoded gas stations to capture trucks which stopped at these facilities (Pinjari et al., 2016). Similarly, we created bounding boxes

around waterway ports to capture truck GPS trajectories with stops in the port areas. However, instead of mapping truck trajectories with business locations, we leveraged lock and dam commodity flow data for multimodal data fusion. Further expanding on the work of Pinjari et al., we accessed eight weeks of truck GPS data, distributed throughout the year, to capture seasonality, and included nine commodity groups rather than exclusively examining petroleum products. Examples of truck GPS data fusion for freight planning purposes are found in Flaskou et al. (2015), who developed a methodology to convert truck GPS trajectories into freight performance measures and applied the methodology to freight corridors in Tennessee. Kuppam et al. (2014), used heavy-duty truck GPS data to develop a tour-based truck travel demand model which was incorporated into the Arizona Freight Demand Model (CPCS, 2019). Our work complemented the methods of Pinjari et al. (2016), Flaskou et al. (2015), and Kuppam et al. (2014) to derive insights from anonymous truck GPS data by adding a multimodal perspective to the data fusion process.

Although waterborne commodity flows from the LPMS dataset and truck trajectories from anonymous GPS data have been used separately for freight applications, methods to fuse the two distinct datasets have yet to be explored. There is significant potential in fusion of these multimodal datasets to address existing, critical gaps in port performance measurement, namely the ability to measure port-level commodity flows. This paper demonstrates that it is possible to produce port-level commodity flows by overcoming several challenges associated with multimodal freight data fusion. Our main challenge is data heterogeneity. Data heterogeneity is caused by the diverse methods to collect, process, store, and share data from different agencies and data providers (e.g. USACE, ATRI, etc.), leading to discrepancies in units, terminology, temporal and spatial coverage, etc. (National Academies of Sciences, Engineering, and

Medicine, 2015). To overcome this challenge, databases were aligned in terms of spatial aggregation, time period, and units of measurement and fed into a Multi-Commodity Assignment Model to describe and quantify port-level commodity flows.

The remainder of this paper presents the methodology to fuse truck GPS and LPMS data, necessary to feed the novel Multi-Commodity Assignment Model, its application to a case study on the Arkansas River, a discussion, and concluding remarks.

### 5.3. Methodology

Our methodology consists of three main steps. First, truck movement data and LPMS commodity flows are spatially and temporally combined. Second, the fused dataset is used to feed a Multi-Commodity Assignment Problem (MCAP) solved via goal programming. The objective is to identify, for a time window *t* and river section *s*, the number of truckloads corresponding to each commodity *j* transloaded at each port *i* to truck  $(x_{i,j}^{s,t})$  and rail  $(R_i^{s,t})$  (transload operations between barge and rail are not distinguished by commodity). Third, the results of the optimization model are post-processed to identify flow and cargo directionality (e.g. up- and down-river flows). In this section, spatial and temporal fusion are discussed, followed by formulation of the MCAP optimization model and post-processing requirements.

### 5.3.1. Fusion of Truck GPS Data and Lock and Dam Commodity Flows

The data required for MCAP included: i) the number of trucks accessing each port, derived from truck GPS data, and ii) the number of equivalent truckloads of each commodity passing through each lock, derived from LPMS, both for the same time period *t*. The truck GPS and LPMS data were heterogeneous in units (trucks vs. weight -in tons-), spatial coverage (ports vs. locks), and temporal scope (hourly vs. monthly). Fusion of the datasets required conversion of each dataset to the same units (truckloads), geographical areas (river segments), and temporal scope (annualized).

### Lock Performance Monitoring System (LPMS) Data Processing

LPMS data consists of monthly quantities (by weight) of 40 commodities transported along U.S. inland navigable waterways by direction (e.g. upriver and downriver). The USACE collects data on the quantity of commodity at each of their approximately 200 locks and dams. Historical monthly data (2009-2017) is publicly available in '.xlsx' format from the Public Lock Commodity Report (U.S. Army Corps of Engineers, 2016). For any current and previous calendar year, a summarized report by lock, for nine aggregated commodity groups, is available in '.html' format from the LPMS website (U.S. Army Corps of Engineers).

LPMS data processing (Figure 5.2.a) consisted of calculating the difference in the quantity (by weight) of each commodity between each pair of consecutive locks, per direction, per month,  $(\Delta L^{s,t}_{j,U}; \Delta L^{s,t}_{j,D})$ , referred to as 'commodity flux'. Upriver and downriver commodity flux were aggregated to quantify commodity flux per month,  $\Delta L^{s,t}_{j}$ , and then converted to equivalent truckloads by dividing commodity flux by commodity-specific truck payload factors  $f_j$ . Truck payload factors can be gathered from State Travel Demand Models (FHWA, 2019), the Freight Analysis Framework (Macks Inc., 2016), or are commonly collected through shipper and carrier surveys. Equivalent monthly truckloads  $c_j^{s,t}$  were then summed over the year to obtain the annual equivalent truckloads of each commodity flux between each pair of consecutive locks,  $c_j^{s,annual}$ .



a. LPMS data preparation

b. Truck GPS data preparation

# Figure 5.2 LPMS and truck GPS data fusion

# Truck GPS Data Processing

The anonymized truck GPS data used in this work consisted of timestamped locations (latitude and longitude) for a sample of the truck population covering a statewide region. Each truck's GPS transponder, identified by a unique but anonymous number, emits intermittent signals ("pings") over time, indicating its location. First, the anonymous GPS pings were grouped by truck into "trips" and then subjected to quality control protocols to remove inconsistent records. Inconsistencies were defined as trips of less than 20 pings, trips with geographic coverage less than 1.2 miles (e.g. length of the diagonal of the bounding box including all pings), and calculated speeds higher than 81 mph. Then, a stop identification

algorithm developed by Camargo et al. (2017) and adapted by Akter et al. (2018) was applied to identify stop locations and durations for each trip. A truck was considered to be stopped when its speed was lower than 3 mph for more than 5 minutes, and the stop coverage area was less than 0.2 miles. All pings corresponding to a single stop were clustered within a rectangular bounding box, and the location of the first ping in the cluster was assigned as the stop location. After the stops made by individual trucks were found, trucks with stops within port areas were identified. A port area was defined as a bounding box around the port facility (including the truck loading area) corresponding to a dock, identified by aerial imagery (e.g., Figure 5.3).



# Figure 5.3 Example of port area geographic bounding boxes

The GPS data used in this study contained four two-week samples, roughly capturing the start of each quarter of the year (Figure 5.2.b). Studies show the coverage of the GPS sampled data to be 10-15% of the total truck population (Diaz Corro et al., 2019). To later fuse annual commodity flows from LPMS with the truck GPS data sample, it was necessary to estimate annual, total (e.g. population level) truck volume for each of the ports. Thus, once the sample number of trucks *S* found at each port *i* during each sample period *w*,  $S_i^w$ , was found, two

expansion factors were applied to estimate the annual truck volume at each port. First, the sample was expanded to represent population-level truck volumes. Expansion factors for each sample period,  $V_{\nu}^{w}$ , were derived as the ratio of the GPS sample volume to truck counts from nearby Weigh-In-Motion (WIM) stations for the same time period<sup>5.1</sup>. The GPS-derived truck volume at each port,  $S_{i}^{w}$ , was multiplied by  $V_{\nu}^{w}$  to estimate the total population of trucks at each port, e.g. "volume-expanded". Next, temporal representation of the GPS sample was addressed by extrapolating each volume-expanded two-week period to an annual volume. Each volume-expanded, two-week period was multiplied by a temporal expansion factor,  $V_{i}^{Q}$  (e.g. number of two-week periods in a three-month quarter) (eq. 5.1). Lastly, quarterly volumes  $S_{i}^{Q}$  were summed to obtain the annual number of trucks accessing each port (eq. 5.2).

$$S'_{i}^{Q} = S_{i}^{W} \times V_{v}^{W} \times V_{t}^{Q}$$

$$5.1$$

$$T_i^{annual} = \sum_Q S'_i^Q$$
 5.2

### 5.3.2. Multi-Commodity Assignment Model Formulation

The model to quantify and describe commodity throughput at each inland waterway port from LPMS and truck GPS data was conceptualized as a Generalized Assignment Problem (GAP). The reason to select this model type (e.g. assignment) and not others (e.g., discrete choice models, gravity models, machine learning, simulation) is in the data that was available (and unavailable). For instance, the problem could be considered as a choice selection, in which vessels transporting commodities and observed at locks choose among a set of ports for transloading. However, to apply a discrete choice model would require an origin-destination matrix for each commodity to be derived from historical data or alternative sources, which do not

<sup>&</sup>lt;sup>5.1</sup> WIM are embedded roadway sensors that continuously measure truck volume and weight by axle configuration (FHWA, 2016).

exist in the public realm. Gravity, or entropy maximizing, model formulations present another option for our application. These models are utilized in traditional 4-step travel demand models to obtain origin-destination matrices by "matching" production and attraction of trips in each zone. In our case, we may consider that each port and each lock constitute a zone. In that context, the application of a gravity model would require the availability of historic and/or current production and attraction data on each commodity in each port, which also does not exist in the public realm (or likely in any aggregated private dataset). Supervised machine learning methods were also disregarded, because we do not have any labeled training instances such as historical/observed data on the number of truckloads of each commodity transloaded at each port. Lastly, simulation can also be an effective tool in this context. However, simulation would require an independent dataset of commodity flows at each port for calibration. Since such dataset does not exist in the public domain, it was not possible to adopt a simulation approach. Further, our goal was to develop a descriptive quantification of commodity flows at port level and, thus, a simulation approach (which is targeted at analyzing policy sensitive inputs) was not necessary. It was deemed untenable to consider executing an intercept survey at each port to obtain the commodity carried by each truck in an effort to gather labeled training instances for a supervised machine learning model.

A GAP seeks to optimally assign tasks to agents, subject to capacity restrictions on the agents, and an agent may be assigned many tasks (Kundakcioglu et al., 2008). A GAP objective function may minimize cost, maximize profit, etc. GAP and its extensions have been applied broadly, including job scheduling, facility location, routing, etc. (Kundakcioglu et al., 2008). In applying a GAP to the quantification of port throughput, commodities were considered "tasks" to be assigned to ports, i.e. "agents". The objective function targeted minimal deviation between

observed and predicted truck flows at each port. Such minimization was subject to the following conditions: i) port "capacity", defined as the number of trucks accessing each port, and ii) assurance that all commodities were assigned to *at least* one port. Limitations of extending the GAP to quantify port throughput were addressed. GAP assigns each task to only one agent. For the port throughput problem, however, the total commodity flux on a river section may be transloaded at several ports. To incorporate these restrictions, a Generalized Multi-Assignment Problem (GMAP) was applied. GMAP differs from GAP in that tasks may be duplicated and assigned to more than one agent (Kundakcioglu et al., 2008).

In this paper, the application of the GMAP to a multi-commodity scenario was referred to as a Multi-Commodity Assignment Problem (MCAP). In the MCAP formulation, a river section *s* was defined as a stretch of inland navigable waterway located between a pair of consecutive locks ( $L_1$ ,  $L_2$ ) (Figure 5.4.a). The objective was to assign commodity flux  $c_j$  along river section *s* to each port *i* within that river section (Figure 5.4.b), i.e. to identify the (unknown) quantity of commodity *j* transloaded between barge and truck at port *i*  $\epsilon$  *s*. The (unknown) quantity of freight transloaded between barge and rail was represented as *R* (Figure 4b). The quantity of freight was measured in equivalent truckloads. The individual commodities *j* belong to a work group of commodities *a*.



Figure 5.4 Multi-commodity assignment model schematics

The problem was formulated by equilibrating commodity flux c (expressed as equivalent truckloads) along waterways with land-side freight volume carried by truck T and rail R, during time period t (eq. 5.3):

$$\sum_{i} (T_{i}^{s,t} + R_{i}^{s,t}) = \sum_{j} c_{j}^{s,t}$$
 5.3

Equation 5.3 assumed that commodity flux (supply) in each section was transferred to truck or rail at each port (demand) such that supply and demand were balanced. However, considering the contextual, spatial, and temporal heterogeneity of the LPMS and truck GPS datasets, it is unlikely that commodity flux in each river section will balance with truck volumes at each port. Furthermore, operational characteristics of the ports such as inventory holding and storage likely weaken the assumption of balanced supply and demand over shorter time periods, e.g. monthly. Thus, the model was applied to annual commodity flows, reducing the effects of long-term commodity storage or inventory holding at each port. Further, the problem was conceptually formulated as a minimization problem in which the difference between truckloads representing supply (LPMS-derived commodity flows) and demand (GPS-derived truck volumes) was minimized (eq. 5.4).

Minimize 
$$f = \sum_{i} (T_i^{s,t} + R_i^{s,t}) - \sum_{j} c_j^{s,t}$$
 5.4

The definition of sets, parameters, variables, objective function and constraint equations involved in the model are presented in Table 5.1.

The model was solved by adopting goal programming techniques, which consist of the relaxation of conflictive conditions in an optimization formulation to allow a feasible solution (but not necessarily optimal) to be found (Colapinto et al., 2017; Gardi et al., 2014). The relaxation techniques implemented in the MCAP formulation were: i) inequalities adopted in eq.

5.7 and 5.8 instead of equalities, and ii) discrete (integer) values were replaced with continuous

values, e.g. trucks were allowed to be partially loaded.

 Table 5.1 Multi-Commodity Assignment Model Formulation

Sets										
j∈a	Set of commodities									
$i \in s$	Set of ports within each river section									
$s \in r$	Set of sections within a river									
Parame	eters									
$\alpha^{s}_{i,j}$	Coefficient to indicate whether port <i>i</i> on river section <i>s</i> handled commodity <i>j</i> , subject to loading equipment. $\alpha_{i,j}^{s} = \begin{cases} 1 \text{ if port } i \in s \text{ handles commodity } j \\ 0 \text{ otherwise} \end{cases}$									
$\beta^{s_{i}}$	Coefficient to indicate whether port <i>i</i> on river section <i>s</i> had rail access.									
	$\beta_i^s = \begin{cases} 1 \text{ if port } i \in s \text{ has access to the rail network} \\ 0 \text{ otherwise} \end{cases}$									
Variabl	les									
Decision	n variables									
$\chi^{s,t}_{i,j}$	Number of barge/truck truckloads of commodity $j$ transloaded at port $i$ during time period $t$ , on river section $s$									
$R^{s,t}{}_i$	Equivalent truckloads transloaded from barge to rail (and vice-versa) at port $i$ during time period $t$ , on river section $s$									
Input va	ariables									
$C^{s,t}_{j}$	Flux of commodity <i>j</i> on river section <i>s</i> during time period <i>t</i>									
$T^{s,t}{}_i$	Number of trucks $T$ accessing port $i$ on river section $s$ during time period $t$									
	$T_i^{s,t} = \sum_j \alpha_{i,j}^s  x_{i,j}^{s,t} \tag{5.5}$									

Model

**Objective function** 

$$Minimize \ f(x, R) = \sum_{i} \sum_{j} \alpha_{i,j}^{s} x_{i,j}^{s,t} + \sum_{i} \beta_{i}^{s} R_{i}^{s,t} - \sum_{j} c_{j}^{s,t} \ \forall j \in a, \forall i \in s \qquad 5.6$$

Subject to the following constraints

i) **Flow conservation:** The amount of each individual commodity  $c_j$  observed in river section *s* must be less or equal than the sum of the amounts of the same commodity *j* loaded/unloaded to truck at all ports  $i \in s$ 

$$\sum_{i} \alpha_{i,j}^{s} x_{i,j}^{s,t} \le c_j^{s,t} \quad \forall j \in a$$
 5.7

### Table 5.1 Multi-Commodity Assignment Model Formulation (Cont.)

ii) **Proportional truck volumes:** The commodity-specific proportion of trucks *T* accessing each port *i* must be less or equal than the total trucks observed along a river section *s* 

$$\frac{\sum_{j} \alpha_{i,j}^{s,t} x_{i,j}^{s,t}}{\sum_{j} c_{j}^{s,t} - \sum_{i} \beta_{i}^{s} R_{i}^{s,t}} \le \frac{T_{i}^{s,t}}{\sum_{i} T_{i}^{s,t}} \quad \forall i \in s$$
5.8

iii) **Non-negativity constraints:** Non-negativity constraints were placed on the truck volume (*x*), rail volume (*R*), and objective function

$$x_{i,j} \ge 0 \quad \forall j \in a, \forall i \in s$$
 5.9

$$R_i \ge 0 \quad \forall i \in s \tag{5.10}$$

$$\sum_{i} \sum_{j} \alpha_{i,j}^{s} x_{i,j}^{s,t} + \sum_{i} \beta_{i}^{s} R_{i}^{s,t} - \sum_{j} c_{j}^{s,t} \ge 0 \quad \forall j \in a, \forall i \in s$$
5.11

## 5.3.3. Results Post-processing

The decision variables obtained with the MCAP, e.g., the number of truckloads of each commodity transloaded at each port, by mode, for time period *t*, were post-processed to describe the upriver and downriver directionality, and to convert from the number of trucks (e.g., truckloads) back to commodity flows (e.g., freight by weight). The post-processed results were the freight (by weight) transloaded to rail and the freight (by weight) by commodity transloaded to truck at each port, by direction (upriver, downriver), during time period *t*. Rail transloads were quantified but not described per commodity due to rail data unavailability, such as number of railcars observed per port.

#### Throughput directionality

Directionality post-processing consisted of distributing port throughput to upriver and downriver flows. From the GPS data, it was not possible to determine whether a truck was at a port to pick-up or drop-off an upriver or downriver cargo. Thus, upriver and downriver commodity flux were aggregated to account for this discrepancy. To distribute throughput to upriver and downriver portions, the percentage of upriver and downriver commodity flux were calculated from the LPMS data for each river section and commodity. It was assumed that all

ports within the same river section had the same percentage of commodity flux with respect to total commodity flux in that section and direction. Even though this assumption might be somewhat restrictive, considering the data that is available, this may be the best approximation to distinguish the volume of commodity corresponding to each direction of travel on the inland waterways (upriver vs. downriver). Nevertheless, the total volume of commodity transloaded per port and commodity (i.e. upriver and downriver, aggregated) remains unaffected by this assumption. The output of this step was the number of truckloads at each port, per commodity, and per direction (upriver and downriver) for time period *t*:  $(x_{i,j,U}^{s,t}, x_{i,j,D}^{s,t}) \ \forall j \in a, \ \forall i \in s$  (eq. 5.12 and eq. 5.13).

$$x_{i,j,U}^{s,t} = \frac{\Delta L_{j,U}^{s,t}}{\Delta L_{j,U}^{s,t} + \Delta L_{j,D}^{s,t}} \times x_{i,j}^{s,t}$$

$$x_{i,j,D}^{s,t} = \frac{\Delta L_{j,D}^{s,t}}{\Delta L_{j,U}^{s,t} + \Delta L_{j,D}^{s,t}} \times x_{i,j}^{s,t}$$
5.12
5.13

### Commodity volumes

The MCAP described and quantified port throughput in terms of truckloads as a means to alleviate data heterogeneity issues. In particular, the LPMS commodity data (by weight) was converted to truckloads to match the truck GPS data units. To convert from the number of trucks (e.g., truckloads) back to commodity flows (e.g., freight by weight), truckloads were multiplied by payload factors *f<sub>j</sub>* specific to each commodity (eq. 5.14 and eq. 5.15). The payload factors corresponding to each LPMS commodity group can be gathered from the State Travel Demand Model (STDM), vehicle use surveys, or national TDMs, such as the Freight Analysis Framework (FAF). Since LPMS commodity groups may not match with STDM or FAF commodity group to one or more STDM commodity groups. Further, when more than one STDM or FAF commodity group

matched a unique LPMS commodity group, for example, the payload factor of the LPMS group can be calculated as an average of the payload factors of the matched STDM commodity groups. The results obtained after this step were the tons of each individual commodity transloaded per port by direction (upriver and downriver) during time period *t*:  $(X_{i,j,U}^{s,t}, X_{i,j,D}^{s,t})$  (eq. 5.14 and eq. 5.15). For rail, an average payload factor of all commodities was applied.

$$X_{i,i,U}^{s,t} = x_{i,i,U}^{s,t} \times f_j$$
 5.14

$$X_{i,j,D}^{s,t} = x_{i,j,D}^{s,t} \times f_j$$
 5.15

## 5.4. Case Study: McClellan-Kerr Arkansas River Navigation System

The multimodal data fusion methodology and MCAP were applied to the Arkansas portion of the McClellan-Kerr Arkansas River Navigation System (MKARNS), which consists of 308 miles of river divided by locks into 13 river sections. 43 freight ports are located along the waterway (Figure 5.5), which contributes to the national economy with \$4,535M in sales, \$168M in business taxes, and 33,695 jobs (Nachtmann et al., 2015). Within the next 50 years, the net present value of sales, GDP, and tax economic impacts of the MKARNS are expected to be \$232.5B, \$111.3B, and \$7.8B respectively (Oztanriseven et al., 2019). LMPS and truck GPS data were obtained for the year 2016. Truck GPS data for four two-week periods was obtained from ATRI while monthly LPMS data was downloaded from the USACE Navigation Data Center. Commodities were grouped into nine categories, as defined by the LPMS commodity grouping scheme. Truck payload factors for each of the nine categories were derived from the 40 LPMS commodity sub-groups, using the Standard Transportation Commodity Codes (STCC2) payload factors included in the Arkansas State Travel Demand Model to assist with the commodity cross-walks (Alliance Transportation Group, 2012). The proposed MCAP model was applied to each of the eight river sections where freight ports are located. Spatial data (GIS files) for the navigable waterways geometry, port and lock locations were publicly available from the National Transportation Atlas Database (NTAD) for the year 2018. Since waterborne infrastructure does not change frequently, using 2018 data (instead of 2016, matching truck GPS and LPMS data) was considered to be acceptable. Since the NTAD port data contained all ports located along U.S. inland navigable waterways, it was filtered to select only freight ports along the MKARNS. Further, to ensure no ports were missing from the NTAD dataset, or in incorrect locations, Google Earth satellite imagery was used to manually confirm port locations. When available, 2016 imagery was used to account for potential changes between 2016 and 2018 datasets. As a result, 41 freight serving ports were included in the NTAD dataset, and an additional two were added based on satellite imagery (Figure 5.5).



Figure 5.5 Arkansas portion of the MKARNS

## 5.5. Results

The model was formulated and solved using IBM ILOG CPLEX optimization studio version 12.8.0. This software has a built-in feature to relax conflictive constraints if needed to reach a feasible solution. This relaxation feature is necessary to resolve conflicts, allowing for a solution to be found outside of the boundaries imposed by the originally conflictive constraints. This is the case when even the formulation of constraints as inequalities may lead to infeasible solutions. For example, the inequality constraint on truck proportions (eq. 5.8) imposes an upper limit on the proportion of trucks accessing each port within a river section, which may be surpassed after relaxing it, allowing for a higher number of quantified trucks to transload at such port than observed. By applying the proposed model (eq. 5.6) with constraint equations 5.7-5.11, and further relaxing the proportional truck volume condition (eq. 5.8), a feasible solution for each of the river sections was found. Resulting truckloads (e.g., number of trucks) were distributed to upriver and downriver commodity flows per port, and further converted to measures of freight by weight.

Resulting quantifications of port throughput by commodity were summarized in tabular form (Table 5.2-Table 5.3), with the cells representing quantity (measured in annual tons) of each commodity assigned to each port for the upriver and downriver directions. Note that blank cells in the tables denote that the port did not handle a specific commodity or serve a given mode, while a zero value denotes no commodity was transloaded even though the port was equipped to handle that commodity. For example, for Port 3001 in river section 3, 69,414 annual tons of chemicals were transloaded to/from barges traveling upriver, while 128,109 tons of chemicals were transloaded to/from barges traveling downriver. Likewise, 179,082 annual tons of food and farm products were transloaded at Port 3002 to/from barges traveling downriver, representing the dominate movement (only 6,469 annual tons were transloaded by barges

travelling upriver).

# Table 5.2 2016 McClellan Kerr-Arkansas Upriver Freight Transloaded per Port,

Commodity, and Mode (Annual Tons)

	ation	oot	coal	ettol	micals	Materials	tadured	Farm	ninery	laste	wnown	ransload	ansbad at	otal
				1 <sup>2</sup> CH	c. crude	Nam Nam	\$000 F000	Wa		יע <sup>אי</sup>	Truck	. Rail	Pot	
	3001	15,293		69,414	31,404	67,029	3,574				186,714		186,714	
3	3002						6,469				6,469		6,469	
	3003		14,896								14,896		14,896	
4	4001						109,273				109,273		109,273	
	5001						63,534				63,534	34,887	98,421	
	5002				0	0					0		0	
	5003		27,582	0		0	0				27,582		27,582	
	5004			138,749			0				138,749	0	138,749	
	5005			0	29,344	0				1,787	31,130	0	31,130	
	5006						0				0	0	0	
5	5007					14,681		3,596			18,277		18,277	
	5008					9,576		0			9.576		9,576	
	5009			0		0	0				0		0	
	5010				12,443	0					12,443		12,443	
	5011				, -	59,955					59,955		59.955	
	5012						0				0		0	
	5013						0				0		0	
	7001-2			0	0	0					0	45.200	45.200	
	7003		0								0	0	0	
	7004			145	0	0	0			0	145	•	145	
	7005			1.0	0						0	0	0	
	7006	239			0						239		239	
7	7007	257	137 646								137 646		137 646	
•	7008		157,010	0			51 308				51 308	0	51 308	
	7009			0	0	0	01,500		0	10 251	10 251	0	10 251	
	7010-11			0	17 509	541 310	13 280		0	0	572 099		572 099	
	7012		0	192 253	182 050	511,510	0		0		374 302		374 302	
	7012		0	172,235	334		0				334		334	
	9001				33 915						33 915		33 915	
9	9002				0,015	42 419					42 419		42 419	
,	9002			64 801	0	42,419	26.060				90.862		90.862	
	10001			04,001	92.520		20,000				92 529	10.647	103 175	
10	10001			1/7 807	12,329	43 162	30.053				22,329	10,047	221 113	
	110002			147,077	0	34 262	50,055	2 109			36.460	0	36.460	
	11001				10 526	34,202		2,198			10,526		10,526	
11	11002	0		0	10,520	0		0		0	10,520		10,520	
	11003	0		0	0	0				0	0	175.057	175.057	
	12004				14.215	0			0	0	14.215	175,957	14 215	
	13001				14,315	0			0	0	14,315		14,315	
	13002	12.002		242.001	154 515	50,362					50,362	7 (27	50,362	
13	13003-4	13,993		242,891	154,515	121,982	(5.440				555,581	7,637	541,019	
	13005				0		65,449				65,449	0	65,449	
	13006		10.402		0						0	0	0	
	13007		19,422	0	0	0					19,422	0	19,422	

Note that blank cells in the tables denote that the port did not handle a specific commodity or serve a given mode.

# Table 5.3 2016 McClellan Kerr-Arkansas Downriver Freight Transloaded per Port,

	Section	Por	COST C	settol Cr	ernicals	Materials	utsoured	8 fam Ma	stinery w	Waste Ur	Wrown .ct	Iransload at T	ansload po	tuotal
$\angle$	/		/			MIL	~~~	/ `.	/		110	Ru	/	
	3001	20,755		128,109	225,177	5,353	98,946				478,340		478,340	
3	3002						179,082				179,082		179,082	
	3003		10,016								10,016		10,016	
4	4001						251,922				251,922		251,922	
	5001						158,703				158,703	31,116	189,819	
	5002				0	0					0		0	
	5003		1,335	0		0	0				1,335		1,335	
	5004			37,173			0				37,173	0	37,173	
	5005			0	98,573	0				0	98,573	0	98,573	
_	5006						0				0	0	0	
5	5007					1,524		4,074			5,598		5,598	
	5008			0		994	0	0			994		994	
	5009			0	41 707	0	0				41 707		41 707	
	5010				41,797	6 224					41,797		41,797	
	5011					0,224	0				0,224		0,224	
	5012						0				0		0	
	7001-2			0	0	0	0				0	27 775	27 775	
	7001-2		0	0	0	0					0	27,775	27,775	
	7003		0	67	0	0	0			0	67	0	67	
	7004			07	0	0	0			0	0	0	0	
	7006	270			0						270		270	
7	7007	270	7,868								7.868		7.868	
-	7008		.,	0			207.603				207.603	0	207.603	
	7009			0	0	0	0		4,785	0	4,785	0	4,785	
	7010-11			0	20,569	25,837	53,735		0	0	100,142		100,142	
	7012		0	89,639	213,874		0				303,513		303,513	
	7013				392						392		392	
	9001				84,479						84,479		84,479	
9	9002				0	2,135					2,135		2,135	
	9003			39,810			99,530				139,340		139,340	
10	10001				72,667						72,667	7,758	80,426	
10	10002			71,908	0	4,889	60,581				137,378	0	137,378	
	11001					3,666		2,997			6,663		6,663	
11	11002				158,892			0			158,892		158,892	
	11003	0		0	0	0				0	0		0	
	11004									0	0	421,510	421,510	
	13001				4,080	0			0	0	4,080		4,080	
	13002					2,795					2,795		2,795	
13	13003-4	6,897		118,462	44,040	6,049					175,448	4,108	179,556	
10	13005						136,099				136,099		136,099	
	13006				0						0	0	0	
	13007		14,840	0	0	0					14,840	0	14,840	

# Commodity, and Mode (Annual Tons)

Note that blank cells in the tables denote that the port did not handle a specific commodity or serve a given mode.

## 5.6. Discussion

Commodity flows through inland ports were not publicly available for model validation (in fact, the (in)ability to acquire such data was the main goal of this work). Therefore, a method to assess model performance using the previously defined relaxed constraints was developed. In particular, the differences between predicted and observed percentages of trucks at each port were used as an evaluation metric (EM) (eq. 5.16). Generally, lower EM corresponds to better model results.

$$EM_{i} = \left| \frac{T_{i}^{s,t}}{\sum_{i} T_{i}^{s,t}} \right|_{predicted} - \frac{T_{i}^{s,t}}{\sum_{i} T_{i}^{s,t}} \right|_{observed} \times 100\%$$
5.16

Overall, 84% of the ports (36 out of 43) show EM less than 20% (Table 5.4, Figure 5.6.a). By averaging the EM of all ports within each river section, 75% of the river sections with ports (6 out of 8) show an average EM less than 20%. Notably, the six river sections with EM less than 20% represent 80% of the ports within the river navigation system (Table 5.4).

 Table 5.4 Model Evaluation Metric per River Section (EM)

<b>River sections</b>	Number of ports	Average EM per section
3; 4; 5; 7; 10	30	< 10%
13	6	< 20%
9; 11	7	< 40%
1; 2; 6; 8	0	No ports. Algorithm not applicable

A further metric used to assess model performance was the rail-to-truck ratio (RT) of transloaded freight at each river section (eq. 5.17). This metric captured the model ability to mimic RT ratios observed in independent national datasets.

$$RT^{s} = \frac{\sum_{i} R_{i}^{s,t}}{\sum_{i} T_{i}^{s,t}} \times 100\%$$
5.17

With the exception of section 11, all river sections showed RT between 0% to 9% (Figure 5.6.b). Since the decision variable *R* captured both barge/rail transload operations and freight consumed at facilities located at the ports (e.g., refineries, power plants), the high RT observed in section 11 may be explained by commodities arriving by water and being consumed at a power plant with port access located along that river section. The overall RT considering all river sections was 13%, in line with 15% national freight mode share (U.S. Department of Transportation, 2019).



a. Evaluation metric per port (EM)



b. Rail-to-Truck mode share ratio (RT)

Figure 5.6 Model evaluation

The novel, multimodal fusion model presented in this paper closes a critical gap in the ability to quantify and describe port-level commodity flows, which is essential for estimating the demand for freight transportation facilities and services, safety risk, energy consumption, and environmental impacts per port. For example, the resulting port-level commodity flow shown in Table 5.2 and Table 5.3 could be used for estimation of commodity-specific, multimodal freightfluidity performance measures, and to support location selection for transload facilities. Specifically, a relatively high amount of food and farm products was transloaded in river sections 4 and 13, but there was only one port capable of handling such products. This may indicate an opportunity to invest in port infrastructure along those sections through direct investment of private companies or the public sector. As for measuring the resilience of the multi-modal freight supply chain, the model could be adapted to predict the impact of permanent or temporary port closures on the waterborne network. For example, by adjusting the coefficients  $\alpha_{i,j}$  which indicate the types of commodities that can be handled at each port, we could simulate a temporary port shutdown after a severe weather event. The results could then be used to highlight which and to what degree other ports on the waterway accommodate the displaced commodity flows.

Although the proposed methodology was able to quantify and describe port throughput with relatively high accuracy, the methodology could be improved to produce more accurate and robust estimates. First, the model was applied to inland waterways transport with publicly available data on the types of commodities by weight transported through locks operated by USACE. Notably, the USACE issues a Public Lock Commodity Report with data from approximately 200 Locks within the U.S., thus providing coverage of much of the 12,000 miles of U.S. commercial inland waterways (U.S. Army Corps of Engineers, 2014, 2018). On the

MKARNS, locks and dams operate 24 hours per day, 7 days per week, and the data collected is representative of all vessels utilizing the locks. However, the amount and type of commodity carried is provided in a report manually filed by the vessel operator (U.S. Army Corps of Engineers, 2013), which may contain reporting errors. Expanding data collection opportunities to wider geographies and automating the process to reduce respondent error would improve model inputs, and thus allow for more accurate model outputs.

Second, the truck GPS sample used in this study, although very large, may not represent all industries or commodities. Studies have shown the data to under-represent smaller fleets and/or private owner-operators (Pinjari et al., 2014). If lack of representation was tied to port operations, then there was potential for bias within the proposed model framework. For example, if logging trucks serving a particular port were under-represented in the truck GPS sample data, then the model may potentially assign a higher proportion of logging operations to rail or to other ports within the same river section, because trucks carrying logs were not contained in the GPS data for a particular port. Fortunately, representation of truck GPS datasets is expected to improve, as more companies are included in the sampling frame with the recent regulation requiring Electronic Logging Devices (ELDs). In addition, truck GPS data is currently only available through partnerships with private operators or data providers. This may hinder data use by public sector agencies. In future work, a sensitivity analysis will be conducted to evaluate how much truck GPS data (in terms of temporal scope) is needed to obtain acceptable throughput results. The goal would be to determine the minimal amount of truck GPS data needed to estimate stable truck flows to and from the ports to reduce data acquisition costs. Furthermore, truck GPS data was the only source of vehicle movement data used in this work, but could be supplemented by other sources such as the Marine Automatic Identification System (AIS). AIS is

a publicly available dataset providing the timestamped location of operating marine vessels. A preliminary observation of 2016 AIS data on the Arkansas River indicates that it may serve as a suitable complement or replacement to truck GPS data. AIS tracking capabilities may provide data to replace the assumption that directionality percentages per commodity are equivalent for all ports within the same river section. This assumption could be violated considering in-bound and out-bound trade imbalances, however this element of the post-processing is minor to the purpose of the study.

In terms of model improvements, the model was sensitive to assumptions of rail access at each port. Currently, the presence of a rail spur at a port was considered sufficient to allow barge/rail transload operations. However, although rail spurs were present, they may be out of operation. Stakeholder interviews may help to identify ports which do not use observed rail spurs. This would better inform the model variable  $\beta$ , which represents whether a port accesses the rail network or not. Beyond the model improvements detailed above, extensions to the model framework include: i) developing a time-expanded approach to disaggregate from annual to monthly volumes, ii) identifying paths of trucks observed at ports to better link commodity carried by each truck to port operations, and iii) to develop a predictive model that is sensitive to policy and pricing changes that affect commodity throughput at each port or along the river.

## 5.7. Conclusions

This paper presented a novel methodology to spatially disaggregate commodity flows observed at locks along the inland navigable waterways to each port located between locks. Disaggregation was based on the temporal, spatial, and contextual fusion of truck GPS and LPMS data. The methodology proposed, e.g. Multi-Commodity Assignment Problem, consisted of a multimodal data fusion approach to feed an optimization model solved via goal

programming under relaxed constraints. The methodology was evaluated with a case study of the McClellan Kerr-Arkansas river navigation system using 2016 data. Results show that 84% of ports had less than a 20% difference between estimated and observed truck volumes. The quantification of port throughput by commodity and mode obtained by adopting the proposed methodology, which was not previously publicly available, can be utilized for several purposes, such as: i) to estimate multimodal, commodity-based freight fluidity performance measures, ii) to incorporate into multimodal freight travel demand models, iii) to support location selection for waterborne freight transload facilities, and iv) to perform scenario planning and resilience evaluations. The methodology can be applied to any geography with available truck GPS data and inland navigable waterways with aggregated commodity flow data, such as that obtained from locks operated by USACE.

### 5.8. Acknowledgements and disclaimer

This material is based upon work supported by the U.S. Department of Transportation under Grant Award Number 69A3551747130. The work was conducted through the Maritime Transportation Research and Education Center at the University of Arkansas. This work reflects the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

# 5.9. Authors Contributions Statement

The authors confirm contribution to the paper as follows: study conception and design: M. Asborno and S. Hernandez; data gathering and processing: M. Asborno, T. Akter; analysis and interpretation of results: M. Asborno and S. Hernandez; draft manuscript preparation: M.

Asborno. All authors reviewed the results and approved the final version of the manuscript.

# 5.10. References

- Akter, T., Hernandez, S., Corro-Diaz, K., et al. (2018). Leveraging Open Source GIS Tools to Determine Freight Activity Petterns from Anonymous GIS data. Paper presented at the AASHTO GIS for Transportation Symposium. https://gis-t.transportation.org/wpcontent/uploads/sites/51/2018/05/GIS\_T\_Taslima\_Akter\_svh\_TA.pdf
- Alliance Transportation Group. (2012). Arkansas Statewide Travel Demand Model Phase I: Freight Model.
- Asborno, M., & Hernandez, S. (2018). Using Data from a State Travel Demand Model to Develop a Multi-Criteria Framework for Transload Facility Location Planning. *Transportation Research Record*, 2672(9), 12-23.
- Bartholdi, J., Lasso, A., Ratliff, H. D., et al. (2019). Using GPS to Measure Truck Service Times in a Container Terminal. *Maritime Economics and Logistics*, 21(1), 146–155.
- Campo, M., Mayer, H., & Rovito, J. (2012). Supporting Secure and Resilient Inland Waterways. *Journal of the Transportation Research Board*, 2273, 10-17.
- Colapinto, C., Jayaraman, R., & Marsiglio, S. (2017). Multi-criteria Decision Analysis with Goal Programming in Engineering, Management and Social Sciences: a state-of-the art review. *Annals of Operations Research*, 7-40.
- CPCS. (2019). NCFRP 49: Understanding and Using New Data Sources to Address Urban and Metropolitan Freight Challenges. Retrieved from https://www.ncfrp49newfreightdata.com/
- de Langen, P., van Meijeren, J., & Tavasszy, L. (2012). Combining Models and Commodity Chain Research for Making Long-Term projections of Port Throughput: an Application to the Hamburg-LeHavre Range. *European Journal of Transport and Infrastructure Research*, 12(3), 310-331.
- Diaz Corro, K., Akter, T., & Hernandez, S. (2019). Comparison of Overnight Truck Parking Counts with GPS-Derived Counts for Truck Parking Facility Utilization Analysis. *Transportation Research Record*, 2673(8), 377-387. doi:https://doi.org/10.1177/0361198119843851
- Dobbins, J., & Langsdon, L. (2013). Use of Data from Automatic Identification Systems to Generate Inland Waterway Trip Information. *Transportation Research Record*, 2330(1), 73-79. doi:https://doi.org/10.3141/2330-10

- Eisele, W., Juster, R., Sadabadi, K., et al. (2016). Implementing Freight Fluidity in the State of Maryland. *Transportation Research Record*, 2548, 62-70.
- FHWA. (2016). *Traffic Monitoring Guide*. Retrieved from https://www.fhwa.dot.gov/policyinformation/tmguide/
- FHWA. (2019). *Quick Response Freight Methods. Third Edition*. Retrieved from https://ops.fhwa.dot.gov/publications/fhwahop19057/fhwahop19057.pdf
- Flaskou, M., .Dulebenets, M., Golias, M., Mishra, S., and Rock, R. (2015). Analysis of Freight Corridors Using GPS Data on Trucks. *Transportation Research Record*. 2478, 113–122. https://doi.org/10.3141/2478-13.
- Gardi, F., Benoist, T., Darlay, J., et al. (2014). *Mathematical Programming Solver Based on Local Search*: Wiley.
- Hwang, H.-L., Hargrove, S., Chin, S.-M., et al. (2016). The Freight Analysis Framework Version 4 (FAF4). Retrieved from https://faf.ornl.gov/fafweb/data/FAF4%20draft%20report%20on%20data%20and%20met hod.pdf
- IHS / Global Insight. (2011). OKI Regional Freight Plan Appendix A: IHS/Global Insight Transearch Commodity Flow Report. Retrieved from https://www.oki.org/wpcontent/uploads/2015/11/Appendix-A-Revised-OKI-Commodity-Flow-Data-Nov-2011.pdf
- Kundakcioglu, E., & Alizamir, S. (2008). Generalized Assignment Problem. In *Encyclopedia of optimization*: Springer.
- Kuppam, A., Lemp, J., Beagan, D., Livshits, V., Vallabhaneni, L., and Nippani, S. (2014).
   *Development of a Tour-Based Truck Travel Demand Model Using Truck GPS Data.* Paper presented in 93rd Annual Meeting of the Transportation Research Board,
   Washington, DC. https://trid.trb.org/view/1289301
- Macks Inc. (2016). FAF4 Freight Traffic Assignment. Retrieved from https://faf.ornl.gov/fafweb/data/Final%20Report\_FAF4\_August\_2016\_BP.pdf
- Nachtmann, H., Boudhoum, O., & Oztanriseven, F. (2015). *Regional Economic Impact Study for the McClellan-Kerr Arkansas River Navigation System*. Retrieved from https://martrec.uark.edu/research/regional-economic-impact-study-for-the-mcclellankerr-arkansas-river-navigation-system.pdf
- National Academies of Sciences, Engineering, and Medicine. (2015). NCFRP Report 35: Implementing the Freight Transportation Data Architecture: Data Element Dictionary. Retrieved from http://www.trb.org/NCFRP/Blurbs/173083.aspx
- Oztanriseven, F., & Nachtmann, H. (2019). Modeling Dynamic Behavior of Navigable Inland Waterways. *Maritime Economics & Logistics*, 1-23.

- Parker, M. (2019). FHWA Freight Fluidity Supply Chain Monitoring Program. SASHTO 2019 Annual Meeting. Retrieved from http://sashto.org/Presentations/Monday,%20August%2019/A1/Mary%20Grace%20Parke r%20-%20FF%20SASHTO%20Aug%202019.pdf
- Pinjari, A., Kamali, M., Luong, T., et al. (2016). GPS Data for Truck-Route Choice Analysis of Port Everglades Petroleum Commodity Flow. Rretrieved from http://www.trb.org/Main/Blurbs/174055.aspx
- Pinjari, A., Zanjani, A., Thakur, A., et al. (2014). Using Truck Fleet Data in Combination with Other Data Sources for Freight Modeling and Planning. Retrieved from https://www.fdot.gov/docs/default-source/research/reports/FDOT-BDK84-977-20-rpt.pdf
- Thoma, M., & Wilson, W. (2005). Leading Transportation Indicators: Forecasting Waterborne Commerce Statistics Using Lock Performance Data. *Journal of the Transportation Research Forum*, 44(2), 33-44.
- Transportation Research Board. (2014). *Developing Freight Fluidity Performance Measures* (*Circular E-C187*), Washington, DC. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec187.pdf
- Transportation Research Board. (2016). Advancing Freight Fluidity Performance Measures (Circular E-C207), Washington, DC. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec207.pdf
- Transportation Research Board. (2018). *Workshop on Implementing a Freight Fluidity Performance Measurement System (Circular E-C240)*, Washington, DC. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec240.pdf
- U.S. Army Corps of Engineers. (2013). Inland Navigation Economics Webinar Series: Navigation Data Resources Webinar. Retrieved from https://planning.erdc.dren.mil/toolbox/webinars/13mar6-InlandNavEcon.pdf
- U.S. Army Corps of Engineers. (2014). Sustainable Solutions to America's Water Resource Needs. Retrieved from https://www.usace.army.mil/Portals/2/docs/civilworks/news/2014-18\_cw\_stratplan.pdf
- U.S. Army Corps of Engineers. (2016). Lock Performance Monitoring System. Public Lock Commodity Report. *Monthly Tons Report*. Retrieved from http://www.navigationdatacenter.us/lpms/Public%20Lock%20Commodity%20Report%2 0(CDM)-2016.xlsx
- U.S. Army Corps of Engineers. (2018). Digital Library. *Locks by waterway, lock commodity, CY* 1999 - 2017. Retrieved from https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/2956

- U.S. Army Corps of Engineers. Lock Performance Monitoring System. *Monthly Tons Report*. Retrieved from https://corpslocks.usace.army.mil/lpwb/f?p=121:6:15549769738061::NO:::
- U.S. Army Corps of Engineers. Navigation Locks. Retrieved from https://www.usace.army.mil/Missions/Civil-Works/Navigation-Locks/
- U.S. Census Bureau. (2019). U.S. Port Data Foreign Trade Statistics. *Schedule D District and Port Codes and descriptions*. Retrieved from https://www.census.gov/foreign-trade/schedules/d/distname.html
- U.S. Department of Transportation. (2016). *Transportation Statistics Annual Report*. Retrieved from https://rosap.ntl.bts.gov/view/dot/32787
- U.S. Department of Transportation. (2019). Port Performance Freight Statistics Program, Annual Report to Congress. Retrieved from https://www.bts.gov/ports

### CHAPTER 6. Commodity-based Vessel Trip Characterization on Inland Waterways

## 6.1. Abstract

Given the complexity of the multimodal freight transportation system, there is increased interest in developing multimodal "freight fluidity" indicators to capture end-to-end supply chain performance. Inland navigable waterways play a key role in the multimodal transportation system by connecting productive heartland areas to international gateways, while keeping transportation costs competitive. To expand the concept of freight fluidity to inland navigable waterways, which typically carry bulk freight, requires a highly disaggregated understanding of freight flow by commodity. However, publicly available commodity flow on U.S. inland waterways is limited in its spatial aggregation to the location of locks which is insufficient to identify commodity flows by port origin-destination. Automatic Identification System (AIS) has the potential to further disaggregate inland waterway commodity flows but has thus far only been used to measure general waterway performance (e.g., speed, travel time). The purpose of this work is to quantify and characterize inland waterway commodity flows at trip level from publicly available data. This is accomplished through the development of a multi-commodity assignment model which conflates vessel and vehicle movement data (from AIS and GPS) with commodity volumes (from Lock Performance Monitoring System data). Validation using data from the Arkansas River show agreement between model predictions and aggregated commodity volumes with differences between 0.00%-1.82% by commodity and lock. Detailed commodityflow estimates allow us to derive commodity-based freight fluidity measures and forecasts, which can support data-driven project prioritization and scenario planning.

**Key words:** Inland Waterway Transportation, Commodity Flow, Automatic Identification System, Freight.

## 6.2. Introduction and Background

Given the complexity of the multimodal freight transportation system, there is increased interest in developing multimodal "freight fluidity" indicators to capture end-to-end supply chain performance (Transportation Research Board, 2014). Freight fluidity measures the ease at which freight (in quantities of tonnage or volume) moves through the multi-modal supply chain. Fluidity indicators were first introduced by Transport Canada, who measured the total transit time of inbound containers from overseas markets to strategic North American inland destinations via various gateways (Transport Canada, 2017). Unlike Canada, the U.S. has not yet adopted practices of measuring multi-modal freight flows that encompass end-to-end (e.g., port-rail-highway-customer) goods movement (FHWA, 2017). Currently, the FHWA is leading national efforts to implement freight fluidity system performance measures and analysis. Examples of such efforts can be found in (Eisele, Juster, Sadabadi, Jacobs, & Mahapatra, 2016; Cambridge Systematics, Inc., 2011; I-95 Corridor Coalition, 2019). However, these efforts are currently limited to the use of only truck probe data.

Inland navigable waterways play a key role in the multimodal transportation system by connecting productive heartland areas with international gateways, while keeping transportation costs competitive (U.S. Committee on the Marine Transportation System, 2020). More than 25,000 miles of U.S. inland waterways carry about 14% of all domestic freight representing more than 600 million tons of cargo annually (American Society of Civil Engineers, 2017). Expanding the concept of freight fluidity to inland waterways where non-containerized, bulk or dimensional cargo is transported (Wiegmans, 2017) requires a highly-disaggregated understanding of freight flow by commodity. However, at best, publicly-available commodity flow data on U.S. inland waterways is limited in its spatial aggregation to the location of locks

through the Lock Performance Monitoring System (LPMS) (U.S. Army Corps of Engineers, 2018). LPMS aggregates commodities into nine groups (Table 6.1). Notably, the U.S. Army Corps of Engineers (USACE) collects disaggregated commodity flow data at vessel trip level (U.S. Army Corps of Engineers, 2018). However, such detailed data is reserved for use by collecting agencies and not made available to public agencies like State Departments of Tranportaiton. A summarized version of this data is shared via the Waterborne Commerce of the United States (WCUS) (U.S. Army Corps of Engineers, 2016). In particular, the Manuscript Cargo and Trip File provides movements of commodities at certain ports, harbors, and inland waterways in the U.S., but it is limited in: i) its spatial aggregation, (for example, it includes only three ports from the more than 40 freight port terminals in Arkansas); and ii) it is based on manually entered reports, which may contain errors. This work overcomes those limitations by leveraging geospatial Automatic Identification System (AIS) data and other publicly available databases to quantify and characterize highly disaggregated commodity flow on inland waterways.

Code – Commodity group					
10 – Coal, lignite, and coal coke					
20 – Petroleum and petroleum products					
30 – Chemicals and related products					
40 - Crude materials, inedible, except fuels					
50 – Primary manufactured goods					
60 – Food and farm products					
70 - Manufactured equipment and machinery					
80 – Waste material					
90 – Unknown or not elsewhere classified					

Table 6.1	LPMS	Commodity	Classification
-----------	------	-----------	----------------

Contributing to the measurement of multimodal transportation system performance and freight fluidity, the USACE developed several applications of fluidity using AIS data, namely: i) lock operations management (interactions between individual vessel operators and the system), ii) the Inland Marine Transportation System Travel Time Atlas (under development, will include travel time, travel time reliability, and port terminal dwell time); and iii) the Port Fluidity Performance Measurement Methodology (port system time from anchorage to exit, cycle time from entrance to channel exit, travel time, travel time indices) (Transportation Research Board, 2018). For the development of the Inland Marine Transportation System Travel Time Atlas, USACE produces travel time estimates for key waterway segments, updated quarterly. Travel time between ports or river markers is estimated from AIS historical data (DiJoseph & Mitchell, 2015). The output is presented as vessel travel time tables that summarize the 25th, 50th, and 75th percentile travel times between inland waterway ports that constitute origin-destination pairs per river segment (Kress et al., 2016). This paper supplements USACE's work by allowing for commodity-based travel-time characterization, leveraging AIS along with other data sources.

The purpose of this work is to characterize highly disaggregated commodity flows on an inland waterway network based on publicly available data. This is accomplished by conflating multimodal, ubiquitous geospatial vehicle tracking data (maritime AIS and truck Global Positioning System (GPS)) with aggregated commodity data (USACE LPMS). The proposed model can be used to quantify and describe the type of commodities carried by vessel trips mapped to a detailed inland navigable waterway network. Applications include the development of commodity-specific, multimodal freight fluidity performance measures which extend to data driven project prioritization. For example, detailed sections of an inland waterway can be prioritized for dredging based on their importance to the economy, as measured by the value and

tonnage of the commodities transported through each river section. In addition, the geospatial, timestamped trip data characterized by commodity produced by our model can support planning and scheduling of transportation infrastructure investments. In particular, traffic-disruptive maritime operations can be scheduled based on the selection of the time of year when a given commodity has its lowest traffic on the link and node of the network where the infrastructure improvements are planned, thus minimizing construction and maintenance impacts on the economy.

The remaining of this paper presents the data preparation necessary to feed the multicommodity assignment model, specification and formulation of the model, a case study to test and evaluate the model on the Arkansas River, and concluding remarks.

### 6.3. Methodology

For this work, a trip is defined as a sequence of network links and nodes visited by a vessel in transit between each pair of time-wise-consecutive stops associated with port nodes in the network (presented as "trip-chains" in Chapter 5). Intermediate stops along a vessel trip may occur at nodes characterized as non-port locations, such as locks or anchoring grounds. The methodology to characterize and quantify vessel trip cargo by commodity on an inland waterway transportation network is based on a multi-commodity assignment model, formulated as an optimization assignment model, fed by freight data derived from public databases. Details on input data and model formulation are presented in the following paragraphs.

### 6.3.1. Data Preparation

The input for the multi-commodity assignment model has two components: i) the port of origin and destination of each vessel trip; and ii) port throughput by commodity. Each input is defined for the same time period and study area.
## Vessel Trip Identification by Port of Origin and Destination

The port of origin and destination for each vessel trip is derived from AIS data by following the heuristic in Chapter 3 (Asborno, Hernandez, and Yves, 2020). Briefly, the heuristic first identifies vessel stops by clustering successive AIS records based on their location, timestamp, and calculated speed. Then, each stop is associated with a network node based on proximity. Timewise-consecutive stops constitute the origin and destination of a path segment. Later, a map-matching algorithm reconstructs complete vessel paths by finding the shortest path between origin-destination pairs. Path segments are joined to define freight trips with origin and destination in ports. Lastly, freight trips are characterized by origin, destination, length, duration, time-of-year (week, season, etc.), and path (but not by commodity). Trip origin and destination are represented by network nodes, location type (port, anchoring ground, lock, or other), and a unique location identification number. While the ports of origin and destination for trips within the study area are identified from the AIS sample pertaining to the study area, trips coming from or exiting the study area boundaries are subject to data preparation. Thus, the true origin and destination of trips coming from or destined to ports located outside of the study area are represented by "artificial" ports at the pair of locks located at the study area boundaries (Figure 6.1). Next, the trips with origin or destination in the study area boundary locks are identified by applying geoprocessing tools. First, "screenlines" are created at each of the locks. Then, vessel trips (paths) intersecting the screenlines are identified. Lastly, the information about the origin and destination of all trips is converted into a matrix where rows represent each freight port of the network, and columns represent each trip. The matrix values represent a binary variable called "visited"  $(v_{i,t})$ , which takes the unit value if port i was either the origin or destination of trip *t*, and the null value if vice-versa (eq. 6.1).

$$v_{i,t} = \begin{cases} 1 \ if \ trip \ t \ visited \ port \ i \\ 0 \qquad otherwise \end{cases}$$
(6.1)

### Port Throughput by Commodity

In the U.S., port throughput by commodity is publicly available only for the major ports based on their annual cargo handled (Bureau of Transportation Statistics, 2019). This typically excludes inland waterway ports which serve relatively smaller annual cargo quantities than coastal ports. Thus, this work derives the annual volume (in tons) of commodity transloaded at port terminals on inland waterways by following the procedure in Chapter 5 (Asborno, Hernandez, and Akter, 2020). Briefly, the method spatially disaggregates inland waterways commodity flow between locks, as observed from USACE LPMS, into port-terminal commodity flow, by observing the relative volume of trucks accessing each port terminal during the same time period and study area from truck GPS data. The method estimates the volume of commodity (in equivalent truckloads) transloaded annually between barge and truck per port terminal (in tons) and direction (upriver, downriver), and the total volume of freight (all commodities aggregated) transloaded between barge and rail or not subject to transload (produced or consumed at the facility), per port terminal and direction. Notably, the estimates do not distinguish between pick-up and drop-off freight volumes. This output serves as input to the procedure described in this paper, following several data preparation steps. First, the volume of freight transloaded between barge and rail (or not subject to transload at all, for which commodity type is not known) is summed to the LPMS commodity group labeled as "unknown or not elsewhere classified". The purpose of this step is to calibrate the estimated port level flows to aggregate observed totals. Then, volume of freight of downriver and upriver directions are aggregated, and a payload factor by commodity is applied to convert truckloads to tonnages (eq. 6.2). Lastly, a matrix is created, where rows represent each freight port *i* of the network, and

columns represent each of the nine commodity groups *j* following the LPMS aggregation (Table 6.1). The matrix values  $(a_{i,j})$  represent the annual tonnage of each commodity transloaded at each port. For the "artificial" ports represented by the locks located at the study area boundaries, the volume and type of commodities for the study period are obtained directly from the LPMS.

$$a_{i,j} = (a_{i,j,D} + a_{i,j,U}) \times f_j$$
 (6.2)

Where,

 $a_{i,j}$  = volume of commodity *j* (in tons) transloaded at port *i* 

 $a_{i,j,D}$  = volume of commodity *j* (in equivalent truckloads) transloaded at port *i*, corresponding to downriver direction

 $a_{i,j,U}$  = volume of commodity *j* (in equivalent truckloads) transloaded at port *i*, corresponding to upriver direction

 $f_j$  = truck payload factor for commodity j

#### 6.3.2. Stochastic Multi-commodity Assignment Model

Trip port of origin and destination and type and volume of commodities transloaded at each port serve as input to the two-stage multi-commodity assignment model described here. Since the inputs are derived from different sources (AIS, GPS, and LPMS), the assignment model is a tool to minimize data heterogeneity by assuming a non-integer, linear, and stochastic model formulation. The first modeling stage consist of a deterministic, linear objective function that seeks to minimize differences in the volume of commodity transloaded at ports and assigned to trips visiting such ports, for all ports, all commodities, an all trips in the study area during the study period. Decision variables in this model pertain to volume (in tons) of each commodity carried per vessel trip, and are treated as non-integer variables to resemble a continuous volume of commodity loaded in any given trip. AIS data is linked to tugs and tows pushing barges on inland waterways, but not to the barges that carry the load (Kruse, et al., 2018). Thus, the second modeling stage introduces stochasticity to reflect uncertainty in the volume of freight (all commodities aggregated) transported per trip. Stochasticity is modeled by assuming different scenarios of maximum freight carried per trip, and combining the results of all scenarios into a single model output. Further details are provided later in this section. The model is evaluated based on the difference of the distribution of commodity volumes observed at locks (LPMS) and model output aggregates.

The selection of an optimization approach to solve the assignment problem is made after considering other potential model types (gravity models, supervised machine learning) against data requirements and availability. For example, consider a gravity model where ports serve as production and attraction zones and the number of trips linking ports used as impedance factors to connect such zones. In the context of available data, from the commodity flows available at the port level (Chapter 5) we are unable to discern productions from attractions, creating a challenge to adoption of gravity models. Similarly, supervised machine learning models are not considered due to the unavailability of a public 'groundtruthed' data of trips characterized by commodity that is needed in the training stage of a machine learning model. Therefore, a generalized assignment optimization model is selected, where commodities represent "tasks" that are assigned to "agents", or trips, subject to constraints. Constraints consider port capacity, trip capacity, commodity flow conservation, and non-negativity boundaries.

The model schema (Figure 6.1) and formulation (Table 6.2 ) consider that the (known) volume *a* of commodity *j* transloaded at port *i* ( $a_{i,j}$ ) during a given timeframe is the sum of the (unknown) volume *x* of the same commodity *j* carried by all trips *t* ( $x_{t,j}$ ) that visit port *i* during the same time period (eq. 6.3). Extending this concept to the set of ports, trips, and commodities on

the network leads to its matrix form (eq. 6.4). Commodity tonnages on both sides of eq. 6.4 may not be in agreement due to data heterogeneity. Thus, the objective function minimizes the difference between the volume of cargo transloaded at ports (A) and the volume of cargo transported by trips (X) for the study area, in search for a feasible set of decision variables  $x_{t,j}$  for all ports, trips, and commodities (eq. 6.4).

$$a_{i,j} = \sum_{t} (v_{i,t} \times x_{t,j}) \tag{6.3}$$

$$A = V \times X \tag{6.4}$$



Figure 6.1 Model schema depicting a section of river with three ports between a pair of

locks

Table 6.2	Model	Formulatio	on – Stage 1
-----------	-------	------------	--------------

Sets	
$i \in P$	Set of ports
$t \in T$	Set of vessel trips
$j \in C$	Set of commodities
$s \in S$	Set of scenarios
Parame	ters
	Coefficient to indicate whether port <i>i</i> is the origin or destination of trip <i>t</i>
V <sub>i,t</sub>	$v_{i,t} = \begin{cases} 1 & if trip t visited port i \\ 0 & otherwise \end{cases}$

otherwise

#### Table 6.2 Model Formulation – Stage 1 (cont.)

 $b^s$  Maximum volume of cargo (in tons) transported per vessel trip, assumed for scenario s

#### Variables

Decision variables

 $x_{t,i}^s$  Volume (in tons) of commodity *j* transported in trip *t* in scenario *s* 

Input variables

 $a_{i,j}$  Volume (in tons) of commodity *j* loaded/unloaded in port *i* 

#### Model

Objective function

minimize 
$$f(x) = \left(\sum_{i} \sum_{j} (a_{i,j}/2) - \sum_{i} \sum_{j} \sum_{t} (v_{i,t} \times x_{t,j}^{s})\right)$$
 (6.5)

Subject to the following constraints

i) **Trip capacity:** The maximum volume of cargo (all commodities aggregated) transported per trip must not exceed the capacity of a reasonable number of barges pushed per tug, assumed for scenario *s* 

 $\sum_{j} x_{t,j}^s \leq b^s \ \forall t \in T$  (6.6)

- ii) **Port capacity:** The volume of freight (all commodities aggregated) carried by all trips visiting a port must match the total volume of freight transloaded at such port (all commodities aggregated)  $\sum_{t} \sum_{j} (v_{i,t} \times x_{t,j}^{s}) = \sum_{j} a_{i,j}/2 \quad \forall i \in P$  (6.7)
- iii) **Commodity flow conservation:** The volume of each commodity transported by all trips carrying such commodity must match the volume of the same commodity transloaded at all ports  $\sum_{t} x_{t,j}^{s} = \sum_{i} a_{i,j}/2 \quad \forall j \in C$  (6.8)
- iv) **Non-negativity constraints:** Non-negativity bounds are placed on the volume of commodity transloaded per trip  $x_{t,i}^s \ge 0 \quad \forall j \in C, \forall t \in T (6.9)$

In the second modeling stage, stochasticity is introduced to model the uncertainty associated to the maximum volume of freight transloaded per trip, *b*. Stochasticity may be modeled by representing the uncertain parameters by random variables and model the randomness by a finite set of scenarios (Seker and Noyan, 2012). Thus, the model presented above is applied for different scenarios of trip capacity (eq. 6.6). The results of each scenario are combined to an overall model result of volume of commodity (in tons) assigned per trip,  $x_{t,j}$ , considering the probability of occurrence of each scenario (Lin et al., 2018) (eq. 6.10)

$$\boldsymbol{x}_{t,j} = \sum_{\boldsymbol{p}} \sum_{\boldsymbol{s}} \boldsymbol{p}^{\boldsymbol{s}} \times \boldsymbol{x}_{t,j}^{\boldsymbol{s}} \tag{6.10}$$

Where,

 $x_{t,j}^s$  = annual volume (in tons) of commodity *j* carried by trip *t* in scenario *s*,  $p^s$  = probability of occurrence of scenario *s*, and

 $x_{t,j}$  = annual volume (in tons) of commodity *j* carried by trip *t* (model results).

# 6.3.3. Model Validation

For validation, the volume of each commodity group (in tons) assigned to trips identified by the model (i.e. model results) on the study area are compared to the volume of each commodity group observed at locks on the same area, during the same time period (i.e. "LPMS"). The comparison consists of calculating the difference of commodity volume at each lock, as an absolute value,  $V_{j,l}$ . Commodity volumes are normalized to the total freight volume on the system (all commodities, trips, and locks aggregated) to eliminate scaling effects that would prevent a direct comparison, and presented as percentages of total freight (eq. 6.11). The average of  $V_{j,l}$  for all locks and commodities constitutes the overall model evaluation metric, V (eq. 6.12). Subtotals per lock and commodity further support model evaluation and validation (eq. 6.13 and eq. 6.14).

$$V_{j,l} = \left| \binom{LPMS_{j,l}}{\sum_{j} \sum_{l} LPMS_{j,l}} \times 100 - \binom{\sum_{t} x_{t,j}^{l}}{\sum_{j} \sum_{l} \sum_{t} x_{t,j}^{l}} \times 100 \right| \quad \forall j \in C, \forall l \in L \quad (6.11)$$
$$V = \frac{\sum_{j} \sum_{l} V_{j,l}}{L \times C} \quad (6.12)$$
$$V_{l} = \sum_{j} V_{j,l} \quad \forall l \in L \quad (6.13)$$
$$V_{j} = \sum_{l} V_{j,l} \quad \forall j \in C \quad (6.14)$$

Where,

 $V_{j,l}$  = model validation metric for tonnages of commodity group *j* and lock *l*,  $LPMS_{j,l}$  = annual volume (in tons) of commodity *j* reported by LPMS for lock *l*,  $x_{t,j}^{l}$  = annual volume (in tons) of commodity *j* carried by trips *t* (model results) observed at lock *l*. To calculate  $x_{t,j}^{l}$ , a screenline approach is used such that trip path geometries of tugs/tows that intersected locks (represented by line segments) are counted as vessels in transit through the lock,

V = overall model validation metric (considering tonnages of all commodities, all locks),

C = number of commodity groups, and

L = number of locks within the study area.

### 6.4. Case Study: Commodities Transported by Vessels on the Arkansas River

#### 6.4.1. Scope, Data, and Model Parameters

The proposed methodology is applied to the Arkansas River. The study area includes a 308-mile segment of the inland navigable waterway consisting of 14 locks and 43 freight ports. The Arkansas River contributes to the national economy with \$4,535M in sales, \$168M in business taxes, and 33,695 jobs (Nachtmann et al., 2015). The case study includes a set of 43 ports and 2 external locks treated as proxy ports. The proxy ports are the W.D. Mayo lock and dam, and the Montgomery point lock & dam, constituting the western and eastern study area boundaries, respectively (Figure 6.2).

The set of vessel trips characterized by port of origin and destination on the Arkansas River for the year 2016 is derived from AIS data. The AIS data sample used for this work may be obtained from (NOAA Office for Coastal Management, 2018). For the vessels observed within the study area, AIS data mining for trip identification, including port of origin and destination, follows the procedure in Chapter 3 (Asborno, Hernandez, and Yves, 2020). As a result, 4,374 trips with either origin or destination in one of the 45 ports in the study area are identified. 3,096 of such trips transit within the study area, while the remaining traverse the study area boundaries, and 102 trips represent pass-through vessel movements. The identification of port of origin and destination of each trip allows for the extraction of the model parameters,  $v_{i,t}$ .



Figure 6.2 Study area

The set of commodities follows the commodity aggregation from LPMS (Table 6.1). The volume of each type of commodity observed during 2016 in the 14 locks within the study area is obtained from (U.S. Army Corps of Engineers, 2017).

The second piece of input data required for this work, namely the annual port throughput by commodity handled by each of the 43 ports on the network (excluding bounds) during 2016 is derived from truck GPS and LPMS data by following the procedure in Chapter 5 (Asborno, Hernandez, and Akter, 2020). Truck GPS data for four two-week periods of 2016 is acquired from a national vendor, representing seasonal movements of roughly 10% of the truck population in Arkansas (Diaz-Corro, Akter, and Hernandez, 2019). Trucks within bounding boxes around port areas are selected, time-expanded, and volume-expanded, to match with LPMS annual data. Payload factors by commodity are further applied to convert number of trucks to commodity volume (in tons). By applying the procedure in Chapter 5 (Asborno, Hernandez, and Akter, 2020), the volume of each type of commodity transloaded in each of the 43 ports of the study area is obtained.

The parameter *b*, which sets an upper bound to the volume of freight carried per trip, is derived from LPMS. In particular, the lock usage report provides the number of loaded barges and the number of commercial vessels observed at each lock operated by USACE (U.S. Army Corps of Engineers, 2018). An average of 4.72 loaded barges per vessel was observed at the locks within the study area during 2016, with a standard deviation of 0.84. To account for the uncertainty in the maximum volume of freight carried per trip, five scenarios are modeled, where *b* takes the form of a discrete variable and is varied 2 standard deviations below and above the average, with a step of one standard deviation. Considering the capacity of most barges is 1,500 tons, the average volume of freight per trip, *b* is 7,085 tons, and the set of scenarios is *S*={4,564; 5,825; 7,085; 8,345; 9,606}. In the absence of further statistical data pertaining the distribution of number of barges per vessel in the study area, the five scenarios are considered to have an equal probability of occurrence, thus *p* = 0.20.

## 6.4.2. Results

Each modeled scenario, which has 39,366 decision variables (4,374 trips and 9 commodities), is programmed and solved in less than one minute with IBM ILOG CPLEX Optimization Studio version 12.10. Due to input data heterogeneity (AIS, truck GPS, and LPMS), relaxation of conflictive constraints, namely port capacity (eq. 6.6) and commodity flow conservation (eq. 6.8), is necessary for a feasible solution to be found. Notably, under some scenarios, relaxing constraints in conflict may lead to an assignment of freight per trip that

violates the commodity flow conservation principle. Such principle dictates that the volume of commodity assigned to trips (output) should be the same than the volume of commodity transloaded at ports (input) (eq. 6.8). In particular, scenarios with an upper bound of the volume of freight carried per trip being equal or less than the average plus one standard deviation, i.e.  $b \le$  8,345 tons, result in this violation. Thus, under such relaxed constraints, the stochastic model (all five scenarios combined) results in 80% of the total freight transloaded at ports being assigned to trips. The flow conservation principle stands when the analysis is done by commodity type for all commodities except chemicals and food and farm products. To account for the un-assigned freight flow of chemicals and food and farm products, model results are post processed as follows. First, it is assumed that the distribution of volume of commodity per trip, for all the trips that carry the given commodity, stands. Then, the volume of commodity assigned per trip is increased proportionally, to match the total volume of such commodity transloaded at ports.

It is observed that 65% of the set of trips are assigned freight. This is consistent with the known presence of vessels within the study area that do not carry freight; being involved instead in repositioning empty barges, construction, dredging operations, etc. Notably, 29% of freight (in tons) is assigned to trips (chains) that have the same port of origin and destination, indicating that such cargo is transported in only a portion of the trip, most likely between a port and a barge anchoring ground, for another tug to pick them up later.

Trips derived from highly disaggregated AIS data are characterized by port of origin, destination, length (miles), duration (hours), and location (path). This work further characterizes trips by commodity carried. Commodity-based measures on the study area are derived for 2016 by aggregating the length and number of trips carrying each commodity (Figure 6.3). Moreover, commodity flow on a detailed inland waterway network is pictured by aggregating the volume of

each commodity carried by all trips transiting each network link. Thus, existing commoditybased maps, spatially aggregated to the location of locks on the network, are disaggregated to the location of ports (Figure 6.4).



Figure 6.3 Ton-miles transported on MKARNS by commodity, 2016



Figure 6.4 Commodity flow on MKARNS detailed inland waterway network, 2016

# 6.5. Model Evaluation and Discussion

# 6.5.1. Model Validation

First, the stochastic model is evaluated using the validation metrics V (eq. 6.12), and  $V_{j,l}$  (eq. 6.11). The lower the validation metrics, the better the model. Differences in the volume of commodity between LPMS and model results,  $V_{j,l}$ , range from 0.00% to 1.82% for each lock,

with an average of 0.25%. We aggregated the differences corresponding to each lock and to each commodity (eq. 6.13 and 6.14, respectively). We found that the commodity with highest difference ( $V_j$ =12.24%) is Food and Farm products, while all other commodities have  $V_j$ <5.41%. As for locks, low differences range between  $V_l$  = [0.00% to 1.95%] (Table 6.3).

Commo- dity/Lock	Coal	Petrol	Che- micals	Crude Materials	Manu- factured	Food & Farm	Mach- inery	Waste	Unk- nown	$V_l$
85	0.15	0.08	0.65	0.74	0.17	0.63	0.02	0.00	0.12	0.94
88	0.20	0.03	0.47	0.61	0.29	0.73	0.02	0.00	0.13	0.00
89	0.20	0.00	0.48	0.62	0.30	0.72	0.02	0.00	0.13	0.01
90	0.11	0.01	0.01	0.21	0.25	1.82	0.01	0.00	0.15	1.85
91	0.10	0.00	0.12	0.21	0.27	1.79	0.02	0.00	0.14	1.95
92	0.07	0.00	0.00	0.25	0.19	1.39	0.01	0.00	0.15	1.26
93	0.08	0.00	0.00	0.25	0.19	1.39	0.01	0.00	0.15	1.26
105	0.01	0.06	0.27	0.25	0.21	0.95	0.02	0.00	0.22	0.39
101	0.01	0.05	0.24	0.22	0.20	0.97	0.01	0.00	0.22	0.46
102	0.02	0.01	0.35	0.24	0.22	0.79	0.03	0.00	0.22	0.23
104	0.00	0.17	0.40	0.25	0.16	0.69	0.03	0.00	0.21	0.47
103	0.03	0.24	0.53	0.52	0.54	0.08	0.03	0.00	0.22	1.92
106	0.03	0.25	0.49	0.52	0.52	0.15	0.03	0.00	0.22	1.79
107	0.03	0.26	0.50	0.53	0.52	0.14	0.03	0.00	0.22	1.82
$V_j$	0.99	1.15	4.24	5.41	0.23	12.24	0.31	0.01	2.51	

 Table 6.3 Model Validation Metric V<sub>j,l</sub> MKARNS, 2016

A second metric used to evaluate the model is the percentage of ton-miles by commodity transported on the MKARNS during 2016, provided at (U.S. Army Corps of Engineers, 2018), and compared to aggregated model results (eq. 6.15),  $M_j$ . This comparison reveals an average difference by commodity of 3.62%, within a range of 0.0% to 15.3%, with food and farm products being the commodity group with higher difference (Table 6.4).

$$M_j = \left| \frac{\sum_t x_{t,j}}{\sum_j \sum_t x_{t,j}} - \frac{m_j}{\sum_j m_j} \right|$$
(6.15)

Where,

 $M_j$  = ton-miles model validation metric for commodity group j, and

 $m_i$  = annual volume (in ton-miles) of commodity *j* reported by USACE for the MKARNS

Commodity	Model results	USACE WCUS	$M_{j}$
Coal	3.03%	2.21%	0.82%
Petrol	3.66%	4.60%	0.94%
Chemicals	44.21%	39.98%	4.23%
Crude Materials	19.10%	10.81%	8.29%
Manufactured	7.74%	7.29%	0.46%
Food & Farm	19.66%	35.00%	15.33%
Machinery	0.13%	0.12%	0.02%
Waste	0.01%	0.00%	0.01%
Unknown	2.45%	0.00%	2.45%

Table 6.4 Validation of Distribution of Ton-miles by Commodity, Mj. MKARNS, 2016

Differences in model validation metrics and "perfect" model results may be caused by a number of reasons. First, the validation assumes that tug-trips carry freight along all their path, while freight might be carried only for a portion of the trip, e.g. between a port and an anchoring area. Second, other potential causes for the differences are related to the model input data. For example, it was observed that 2016 AIS data covers 88% of the vessel population on MKARNS (Asborno, Hernandez, and Yves, 2020). In addition, there could be issues with the commodity volumes manually reported in LPMS. Third, in terms of model characteristics, assumptions of tonnage capacities per trip plays a key role in model results, as evidenced in the sensitivity analysis discussed in the next section. Despite the adoption of a stochastic approach that adopts a commodity distribution per trip based in the distribution of commodity per trip from several scenarios that assume diverse trip capacity, the model may be improved by increasing the number of scenarios, as in a Monte Carlo simulation approach (Lin et al., 2018).

## 6.5.2. Sensitivity Analysis

This section discusses a sensitivity analysis of model scenario results to the input parameter *b*, which assumes the maximum freight volume transported per trip (in tons, all commodities aggregated). For the sensitivity analysis, *b* is varied between two fully loaded barges below and two fully loaded barges above the mean,  $\mu_b$ , with a step of one fully loaded barge. The mean  $\mu_b$  is calculated from the average number of barges pushed per tow on the study area during the study period, as observed per lock and informed through publicly available LPMS usage data, and rounded to the nearest higher integer;  $\mu_b = 5$  fully loaded barges. Notably, LPMS operators collect data of all vessels and barges observed on the locks operated and maintained by USACE, representing not a sample but the vessel population. The five sensitivity scenarios are compared based on the average distribution of commodity tonnages at locks, *V* (eq. 6.12), and the percentage of unassigned freight in each scenario (violation of commodity flow conservation principle, eq. 6.8) (Figure 6.5). As discussed, the lower these metrics, the better the model.

As the capacity of freight per trip (*b*) increases from three to seven fully loaded barges (equivalent to 4,500 to 10,500 tons), the percentage of unassigned freight decreases from 30% to 0%. This is due to the higher capacity of the set of trips on the system to carry the goods transloaded at ports in the study area (input data). For a maximum trip capacity of 5 fully loaded barges (7,500 tons) or more, the set of trips is able to absorb all the volume of commodities transloaded at ports. Scenarios with trip capacity lower than 5 fully loaded barges are not capable of absorbing all the freight transloaded annually on port terminals on the MKARNS.





# Figure 6.5 Sensitivity of deterministic scenarios and stochastic model results to the

#### maximum volume of freight carried per trip

As the capacity of freight per trip increases from five to seven fully loaded barges (7,500 to 10,500 tons), the model validation metric, *V*, which considers the distribution of commodity assigned to trips per locks, increases from 0.36% to 0.72%. Similarly, as the capacity of freight per trip decreases from five to three fully loaded barges (7,500 to 4,500 tons), the model validation metric *V* increases from 0.36% to 0.66%. This indicates that, for a deterministic approach, assuming a volume of maximum freight per trip equivalent to five fully barges or 7,500 tons, which is closest to the average number of barges per trip recorded in LPMS data (4.72 barges), leads to results that better represent commodity flow on the MKARNS.

However, as observed in Figure 6.5, the stochastic model approach, that combines the results of five deterministic scenarios, results in better evaluation metrics that each individual scenario. In particular, the average distribution of tonnages at locks by commodity, V, is significantly better (0.25%). The results post-processing further improves the model capacity to

represent commodity flow on an inland navigable waterway by assigning all freight transloaded at ports to trips.

#### 6.5.3. Applications

By conflating aggregated commodity datasets (such as LPMS) and highly spatially and temporal-disaggregated multimodal vehicle tracking data through the model presented in this paper, it is possible to characterize and quantify commodity flow on an inland waterway network per trip (i.e. not limited to the location of locks). The results of the methodology proposed in this paper, based on public data, may be used by public agencies to prioritize infrastructure investments. For example, by assigning a monetary value to each commodity flowing on a detailed waterway network, network links and corridors may be prioritized by value for dredging purposes. Ubiquitous AIS and truck GPS data permit the transferability of the proposed model to other regions with waterways and aggregated commodity-flow data.

In addition, the proposed methodology may be used as a basis for scenario planning and forecasting. For example, the results of applying the model to a scenario simulating a port closure would inform which portions of the river would experience a change in freight flow by commodity type and volume, allowing more robust dredge scheduling, resiliency analysis, or other planning of infrastructure investment needs.

Furthermore, vessel trips identified from highly disaggregated, ubiquitous, automatic data allow for seasonal commodity flow analysis. Such analysis may be used by agencies as a decision-making factor (among others) to support the selection of the time of year to conduct construction and maintenance operations on transportation infrastructure that minimizes economic disruptions by supply chain.

#### 6.5.4. Limitations and Future Work

A notable limitation of the proposed methodology is that it relies on truck GPS data, which can be expensive, and may underrepresent certain commodities and truck market segments (small fleets, independent owner operators) (Pinjari et al., 2014). The authors are exploring an alternative model where truck GPS data would not be required to estimate the volume and type of commodities carried per vessel trip. Another limitation of the model is that AIS data is linked to tugs and tows pushing barges on inland waterways, but not to the barges that carry the load (Kruse, et al., 2018). The model presented in this paper would benefit from AIS transponders being installed on barges, as implemented in the Port of Antwerp (Port of Antwerp, 2012), or by inquiring the number of barges pushed by the tow in the AIS messages, as it is used in European inland waterways (Javor et al., 2013).

#### 6.6. Conclusion

The main contribution of this paper is the development of a multi-commodity assignment model solved via optimization under relaxed constraints, to characterize and quantify commodity flow on inland waterways at vessel trip level from publicly available datasets. Uncertainty on the assumption of model input parameters is handled by introducing a stochastic scenario approach. Commodities handled by freight ports on a river corridor during a given time period are assigned to vessel trips derived from highly disaggregated maritime geospatial data (AIS) for the same time period and study area. Vessel trips characterized by port of origin, destination, path, timestamp, and commodity carried, are mapped to a detailed inland waterway network, allowing for a detailed commodity flow analysis, previously unavailable in the public domain. Moreover, by leveraging AIS data, this work improves confidential commodity flow datasets that rely on manually-entered origin-destination trip information to derive detailed commodity flows. The methodology developed in this work is tested on the Arkansas River, a 308-mile navigable waterway with 43 ports and 14 locks, with 2016 AIS, truck GPS, and LPMS data. Ubiquitous AIS and truck GPS data permit the transferability of the proposed model to other regions with waterways and aggregated commodity-flow data.

This work may be applied to derive commodity-based freight fluidity performance measures, scenario planning, and scheduling of transportation infrastructure investment. AIS data has been used to derive travel time measures of freight fluidity on inland waterways; this work expands AIS use by quantifying freight fluidity by commodity. Furthermore, the model might be applied to scenarios simulating port closures, resulting in the extent of displacement of freight flows by commodity. Moreover, timestamped trips from AIS data allow for a seasonality analysis, permitting to plan and schedule interventions on transportation infrastructure that minimize impact on a given supply chain.

# 6.7. References

- American Society of Civil Engineers. (2017). *Infrastructure Report Card Inland Waterways*. Retrieved from: https://www.infrastructurereportcard.org/wpcontent/uploads/2017/01/Inland-Waterways-Final.pdf
- Asborno, M., Hernandez, S., & Akter, T. (2020). Multicommodity Port Throughput from Truck GPS and Lock Performance Data Fusion. *Maritime Economics and Logistics*, 196-217. doi:https://doi.org/10.1057/s41278-020-00154-7
- Asborno, M., Hernandez, S., & Yves, M. (2020). Network Mapping of AIS Data to Characterize Inland Waterway Freight Transportation - White paper.
- Bureau of Transportation Statistics. (2018). *Port Performance Freight Statistics Program Handbook of Methods*. Retrieved from https://www.bts.gov/sites/bts.dot.gov/files/docs/browse-statistical-products-anddata/port-performance/217686/ppfsp-handbook-methods.pdf
- Cambridge Systematics, Inc. (2011). *Bottleneck Performance in the I-95 Corridor*. Retrieved from https://i95coalition.org/wp-content/uploads/2015/03/FreightBottleneckReport.pdf?x70560

- Diaz Corro, K., Akter, T., & Hernandez, S. (2019). Comparison of Overnight Truck Parking Counts with GPS-Derived Counts for Truck Parking Facility Utilization Analysis. *Transportation Research Record*, 2673, 8, 377-387. doi:10.1177/0361198119843851
- DiJoseph, P., & Mitchell, K. (2015). Estimating Vessel Travel Time Statistics for Inland Waterways with Automatic Identification System Data. *Transportation Research Board* 94th Annual Meeting. Washington, DC. Retrieved from https://trid.trb.org/view/1339419
- Eisele, W., Juster, R., Sadabadi, K., Jacobs, T., & Mahapatra, S. (2016). Implementing Freight Fluidity in the State of Maryland. *Transportation Research Record*, 2548, 1, 62-70. doi:10.3141/2548-08
- FHWA. (2017). *Freight Demand Modeling and Data Improvement Handbook*. Retrieved from https://ops.fhwa.dot.gov/publications/fhwahop18018/fhwahop18018.pdf
- I-95 Corridor Coalition. (2019). FHWA National Freight Fluidity Monitoring Program. Display of Supply Chain Performance Data. *TRB Innovations in Freight Data Workshop*. Irvine, CA. Retrieved from https://i95coalition.org/wp-content/uploads/2018/12/Freight-Fluidity-TRB-2019\_04\_09-Final-2.pdf?x70560
- Javor, A., Bober, S., Creemers, P., Van Gils, J., & Van der Heijden, W. (2013). *River Information Services Portal.* Retrieved from Operational Use of Inland AIS: http://www.unece.org/fileadmin/DAM/trans/doc/2013/sc3wp3/Leaflet\_Operational\_use\_ of\_Inland\_AIS\_2013.pdf
- Kress, M., Mitchell, K., DiJoseph, P., Rainey, S., Chambers, M., Hsieh, J., & Lilycrop, J. (2016). *Marine Transportation System Performance Measures Research*. USACE ERDC Coastal and Hydraulics Laboratory.
- Kruse, J. C., Mitchell, K. N., DiJoseph, P. K., Kang, D. H., Schrank, D., & Eisele, W. (2018). Developing and Implementing a Port Fluidity Performance Measurement Methodology Using Automatic Identification System Data. *Transportation Research Record*, 2672, 11, 30-40. doi:https://doi.org/10.1177/0361198118796353
- Lin, J., Magnago, F., & Alemany, M. (2018). Optimization Methods Applied to Power Systems: Current Practices and Challenges. *Classical and Recent Aspects of Power System Optimization*. Elsevier. https://doi.org/10.1016/B978-0-12-812441-3.00001-X
- Nachtmann, H., Boudhoum, O., & Oztanriseven, F. (2015). *Regional Economic Impact Study for the McClellan Kerr Arkansas River Navigation System*. Retrieved from https://martrec.uark.edu/research/regional-economic-impact-study-for-the-mcclellankerr-arkansas-river-navigation-system.pdf
- NOAA Office for Coastal Management. (2018). Vessel Traffic Data. Retrieved from https://marinecadastre.gov/ais/
- Pinjari, A., Zanjani, B., Thakur, A., Irmania, A., Kamali, M., Short, J., Park, L. (2014). Using Truck Fleet Data in Combination with Other Data Sources for Freight Modeling and

*Planning*. Florida Department of Transportation. Retrieved from https://rosap.ntl.bts.gov/view/dot/27807

- Port of Antwerp. (2012). *Annual Report 2011*. Retrieved from https://www.portofantwerp.com/en/ais-barges
- Seker, M., & Noyan, N. (2012). Stochastic Optimization Models for the Airport Gate Assignment Problem. *Transportation Research Part E*, 48, 438-459. doi:10.1016/j.tre.2011.10.008
- Transport Canada. (2017). *Gateways and Corridors*. Retrieved from https://www.tc.gc.ca/eng/policy/anre-menu-3023.htm
- Transportation Research Board. (2014). *Developing Freight Fluidity Performance Measures* (*Circular E-C187*). Washington, DC. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec187.pdf
- Transportation Research Board. (2018). *Workshop on Implementing a Freight Fluidity Performance Measurement System (Circular E-C240).* Washington, DC. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec240.pdf
- U.S. Army Corps of Engineers. (2016). *Waterborne Commerce of the United States*. Retrieved from https://publibrary.planusace.us/document/f202cdec-ef75-49ca-8e16-829182cfdf4c
- U.S. Army Corps of Engineers. (2017). *Public Lock Commodity Report*. Retrieved April 2017, from Navigation Data Center: http://www.navigationdatacenter.us/lpms/Public%20Lock%20Commodity%20Report%2 0(CDM)-2016.xlsx
- U.S. Army Corps of Engineers. (2018). *Manuscript*. Retrieved from USACE Digital Library -[2000-2018 trips] Manuscript cargo and trips data files: https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/1671 https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/1813
- U.S. Committee on the Marine Transportation System (CMTS). (2020). An Economic Analysis of Spending on Marine Transportation System (MTS) Infrastructure. Retrived from https://www.cmts.gov/downloads/CMTS\_Economic\_Analysis\_of\_Spending\_on\_MTS\_In frastructure.pdf
- Wiegmans, B., & Konings, R. (2017). Inland Waterway Transport Challenges and Prospects. doi:9781315739083

#### **CHAPTER 7. Conclusion and Future Work**

Fusion of "big data" sources not typically used for freight transportation planning, such as maritime Automatic Identification System (AIS), truck Global Positioning System (GPS), and Lock Performance Monitoring System (LPMS) data, provides a consistent and novel data source for multimodal, long-range freight planning. The methods developed for this work describe, quantify, and characterize commodity-based freight activity on a multimodal transportation system, with focus on inland waterway networks.

The main methodological contributions of this work are:

- i. The *map-matching of maritime AIS data* to identify vessel trips on a detailed inland waterway transportation network. Building upon previous research, vessel trips are defined by their origin and destination, rather than on assumed trip duration or pre-defined bounding boxes.
- ii. The *geospatial data fusion of truck GPS and AIS data* to identify multimodal origindestination pairs associated with a freight facility. This provides a systematic way to identify the area of impact of diverse multimodal freight facilities or industries.
- The development of a *novel multi-commodity assignment model* to quantify and characterize annual port throughput by commodity at inland waterway port terminals; and the temporal and spatial disaggregation of annual port throughput per vessel trips.
   Secondary methodological contributions presented in this work are:
- iv. The identification of *areas with maritime freight activity* that are not currently designated as loading/unloading areas in public databases, based on vessel stop clusters and satellite imagery;

v. The identification of *freight corridors of modal competition* by visualization of multimodal freight paths.

This dissertation also contributes to practical applications and policy tools through:

- vi. The network-based trip identification from AIS data allows for future integration into *multimodal travel demand models* (TDMs) with freight components. In particular, the highly disaggregated nature of AIS data can be leveraged for the generation of synthetic populations for *activity-based freight demand models*. Moreover, the approach to define a routable inland waterway network allows for the incorporation of currently underrepresented inland waterway networks in statewide TDMs to enable identification of multimodal bottlenecks and infrastructure needs.
- vii. The multimodal catchment areas associated with potential investments on freight facilities constitutes a sound, consistent basis to *estimate impacts and benefits of diverse transportation infrastructure investments*. The use of ubiquitous data in time and space, such as AIS and truck GPS, provides a more accurate depiction of the impact area of a freight facility (when compared to the naïve assumption of radial impact areas around the facility), and a common basis for proper comparison and competition of funds. The proposed approach improves the state-of-the-practice that utilizes static, limited traffic counts and subjective survey data. Moreover, the identification and visualization of the geographic extent of multimodal freight catchment areas can be used to estimate population exposure statistics, and to discover areas of modal competition where to target modal-shift policies.
- viii. The quantification of port throughput by commodity and mode obtained by fusing truck GPS and LPMS data fills a critical gap by providing data that was not

previously publicly available. Such data can be used to estimate freight fluidity performance measures; to perform scenario planning, simulate partial or permanent port closures, and resilience evaluations; and to support location selection for multimodal freight facilities on inland waterways.

ix. The commodity-based characterization of vessel trips on an inland waterway network constitutes a data-driven guide to *strategic investment decision-making*. In particular, the quantification and identification of the type of commodities transported on inland waterways segments from highly disaggregated data allows for the prioritization of dredging of such segments, based on the economic value of commodities transported.

Several research avenues may be developed in the future based on the work presented in this dissertation. Three avenues are presented in the following paragraphs, namely: i) further commodity disaggregation; ii) replacement of manually entered AIS fields by machine learning methods, and iii) temporal disaggregation of the annualized port throughput by commodity obtained in this work.

Focusing on commodity-based planning, the nine commodity groups defined by LMPS were used in this work. However, it may be beneficial to further disaggregate commodities. For example, with the food and farm products category it would be valuable to know the breakdown of soybeans and rice, from other grains as they have different harvesting and shipment patterns as well as different constituent groups that lobby for their consideration in freight planning and policy development. The methods (e.g., GMAP) developed in this dissertation can leverage additional, commodity-specific data sources for data fusion with the goal of commodity disaggregation. Such sources include the Agricultural Marketing Service data from Department

of Agriculture (USDA), data from the National Agricultural Statistics Service, and data from the United States Energy Information Administration (EIA) (Appendix A).

This dissertation focused primarily on assignment models. Another possible modeling tool is that of machine learning (ML). For example, ML tools may be applied to public use microdata (such as CFS PUM) to derive factors that affect freight mode choice by commodity. ML tools are adept at finding patterns and making predictions from ubiquitous, highly disaggregated data like AIS. In particular, a drawback noted in the current publicly shared version of AIS data are the manually entered fields for commodity carried. As these fields were prone to human error, they were not used in this work. However, ML may be a promising tool to replace manually entered features. One such feature is the "status" of a vessel that indicates the type of activity in which a vessel is involved. To minimize human efforts and error, an unsupervised data mining algorithm (such as K-means clustering) could be applied to derive the activity based on features derived in this work from AIS data like trip length, duration, coverage, average speed, origin, and destination.

Lastly, the multi-commodity assignment model presented in this work may be improved by adopting a time-expanded approach, in which a monthly analysis during a complete year is conducted. The results of a time-expanded multi-commodity assignment model would be the port throughput by commodity per month (instead of annual, as presented in this work). Such results would provide a more detailed input to the characterization and quantification per commodity type of cargo transported by vessel trips as identified on inland waterway networks, improving its results. Alternatively, LPMS commodity data may be directly disaggregated into the type and quantity of cargo transported by vessel trips transiting inland waterway locks (as observed from AIS data), and then use the paths, origin and destination of those trips to derive

port throughput and highly disaggregated commodity flow on a detailed inland navigable waterway network. In this way, the truck GPS data, which is the most expensive source used in this work, would not be needed.

To conclude, the novel data fusion models and methods presented in this dissertation support several long-range multimodal freight transportation planning applications, such as project prioritization. This work presents a critical step towards the broader goal of representing robust inland waterway freight activity into multimodal transportation infrastructure management and strategic decision-making.

# APPENDIX A: Synthesis on Potential Data Sources for Commodity-Based, Multimodal Long-Range Freight Planning

The literature review section synthesized data sources traditionally utilized for freight planning purposes, i.e. CFS, Transearch, and Freight Analysis Framework (FAF). This appendix synthesizes existing U.S. data sources which are not systematically utilized for freight planning, focusing on their suitability for conflation for commodity-based long-range freight planning purposes. Notably, most non-traditional datasets consider each transportation mode individually, such as waterborne lock performance and monitoring data, AIS, truck Weigh-In-Motion data, and the rail carload waybill. Very few sources include information that could be applied across all modes, such as data from the United States Department of Agriculture (USDA) and County Business Patterns. The data and sources examined in this section are:

- Automatic Identification System (AIS), collected by the U.S. Coast Guard.
- Data from the Waterborne Commerce Statistics Center, including the Lock Performance Monitoring System (LPMS), from the U.S. Army Corps of Engineers.
- Rail Carload Waybill, issued by the U.S. Department of Transportation.
- Truck GPS data from the American Transportation Research Institute (ATRI).
- Business data from ReferenceUSA, InfoUSA, ESRI, and the U.S. Census Bureau under the County Business Partners (CBP).
- Data about transportation of agricultural products, provided by the U.S.
   Department of Agriculture (USDA) through its Agricultural Marketing Service and through the National Agricultural Statistics Service.
- Data from the United States Energy Information Administration (EIA)

- Roadway Weigh-In-Motion (WIM) system.
- Data from the Federal Motor Carrier Safety Administration (FMCSA).
- National Performance Management Research Data Set (NPMRDS).

Each individual dataset is examined in the following paragraphs. For each dataset, the key characteristics, commonalities, and value-added elements are evaluated and compared by creating a data conflation matrix. Examples of data characteristics include: data values provided, units, spatial and temporal scope, frequency of data gathering, reliability, completeness, representativeness, cost, availability, format, etc. The data-conflation matrix summarizing the findings is presented in Table A.1.

Data source / Characteristics	Relevant data values provided	Geographical coverage	Temporal coverage & update frequency	Commodity disaggreg.	Modes	Data cost & availability	Data file type	Main opportunities & limitations
AIS (MARAD)	Vessel location (lat/long), timestamp, vessel characteristics (cargo, tow, passenger).	All waterways.	Temporal disaggregation to the minute of each day. Public file updated annually.	No commodity data available.	Water	Free download.	GIS point layer organized per UTM zone & month/year.	GIS-based, fusion-friendly. Good temporal & geographical coverage. No commodity data.
WCSC / LPMS (USACE)	Quantity (tons) of commodities found each month in each lock.	All U.S. inland waterways where lock & dams are located.	Monthly aggregates, updated annually.	9 commodity groups. Further disaggregation available in .pdf format.	Water	Free download.	Spreadsheet table (.csv).	Suitable for GIS-based fusion. Temporal disaggregation per day would be preferred.
Rail Carload Waybill (FRA)	Individual shipments' quantity (tons), routed distance (miles), Origin, Destination, commodity.	U.S. territory. O/D specified as BEA/State.	Shipment date identified. Updated annually.	STCC 5-digit codes.	All Rail, Rail & Truck, Rail & Water.	Free download.	Spreadsheet table (.csv).	Only railway dataset containing commodity information. Geographical aggregation too broad.

# Table A.1 Main Characteristics of Data Sources Non-Systematically Utilized for Freight Planning

Data source / Characteristics	Relevant data values provided	Geographical coverage	Temporal coverage & update frequency	Commodity disaggreg.	Modes	Data cost & availability	Data file type	Main opportunities & limitations
Truck GPS data (various)	Vehicle location (lat/long), timestamp.	All U.S. territory.	Temporal disaggregation to the minute of each day.	No commodity data available.	Truck	For purchase.	GIS point layer.	Suitable for GIS-based fusion. Good temporal & geographical coverage. Potential lack of representativity.
Business Data / CBP (U.S. Census)	Business establishments' location (lat/long), size (employment), industry served (NAICS).	U.S. territory. By State, County, Metropolitan area, ZIP Code, and Congressional District Levels.	Updated annually.	NAICS 5-digit codes	N/A	Free download.	Spreadsheet (.csv).	Good temporal & geographical coverage.
USDA / Census of Agriculture	Map layers showing number of farms, livestock, poultry, land use, crops, ownership, income, chemical use.	All U.S. territory. Statistics summarized at county, state and national level.	Survey conducted every 5 years (latest: 2012).	Crop type: corn, soybeans, wheat, hay, etc.	N/A	Free download.	.csv spreadsheets suitable to generate GIS-layers.	Information about agriculture specific supply chains. Suitable spatial coverage.

# Table A.1 Main Characteristics of Data Sources Non-Systematically Utilized for Freight Planning

Data source / Characteristics	Relevant data values provided	Geographical coverage	Temporal coverage & update frequency	Commodity disaggreg.	Modes	Data cost & availability	Data file type	Main opportunities & limitations
USDA / Grain Transportation Report	Grain transport cost indicators per mode. Rail: carloads deliveries to port; tariff rail rates, etc. Water: barge grain movements through specific locks and dams by grain type.	Water: 5 locks & dams located on the Mississippi, Ohio, and Arkansas rivers.	Weekly report. Barge shipments aggregated per quarter and annually.	Grain type: corn, wheat, and soybean.	Truck, Water, Rail.	Free download.	Report delivered in .pdf, tables provided in excel.	
EIA	Location and type of infrastructure used on the supply chain of energy products.	All U.S. territory.	Frequently updated.	by energy type: biomass, coal,electricity, fossil fuel resources, geothermal, hydroelectric, natural gas, petroleum, solar, wind, and renewable energy power plants.	Water, Rail, Pipeline, Transmission lines.	Free download.	Excel spreadsheet and google maps.	Information about energy- specific supply chains. Suitable for GIS-fusion.

 Table A.1 Main Characteristics of Data Sources Non-Systematically Utilized for Freight Planning

Data source / Characteristics	Relevant data values provided	Geographical coverage	Temporal coverage & update frequency	Commodity disaggreg.	Modes	Data cost & availability	Data file type	Main opportunities & limitations
WIM (FHWA)	Weight and axle-based classification of trucks.	WIM sensor locations, typically few, on interstates or main U.S. Highways.	Data collected daily.	No commodity data available.	Truck	Upon request	Spreadsheet (.csv).	Poor geographical coverage. No data about commodity.
FMCSA	Carrier registration information: legal name, U.S. DOT number, contact details, number of power units, fleet VMT, operation, cargo carried	U.S. territory.	Current registration information. Frequent updates.	30 commodities, but non- conclusive (several categories per carrier).	Truck	Free download.	Text document (.txt)	Non-conclusive commodity data.
NPMRDS (FHWA)	Average travel time data on the National Highway System.	U.S. National Highway System, by state or region (four regions).	Updated monthly.	No commodity data available.	Truck, passenger cars.	Free, access available only to State and MPO officials.	shapefile	Geographical limitations, non-open- source dataset, no commodity data.

 Table A.1 Main Characteristics of Data Sources Non-Systematically Utilized for Freight Planning

#### A.1. Automatic Identification System (AIS)

The AIS consists of vessel's traffic data, collected for navigational safety purposes (collision avoidance). It is required for all passengers vessels and all commercial vessels over 300 gross tonnage that travel internationally, by the International Maritime Organization (IMO), since December 2004. An onboard navigation device transmits location and characteristics of large vessels in real time. The receivers are base stations on shore, buoys, satellites, and other vessels (U.S. Department of Homeland Security). In the U.S., AIS data is mandatory collected by the U.S. Coast Guard in U.S. and international waters. For inland waterways, AIS is mandatory in the Ohio River, between Mileposts 593 and 606, when the McAlpine upper pool gauge is at approximately 13.0 ft or above, and in the Lower Mississippi River, up to 20 mi above Baton Rouge, Louisiana, at Milepost 254.5 (Dobbins et al., 2013). Even though AIS is not currently required in most U.S. inland waterways, most vessels are using the AIS transponder (DiJoseph & Mitchell, 2015). Historical AIS data (2009-2017) is organized in file geodatabases, including vessel, voyage, and broadcasting information, and it is available for free download at (NOAA Office for Coastal Management, 2018). Examples of vessel data elements are: Vessel name, length, width, and MMSI. Voyage data elements include destination, cargo, draught, ETA, etc. Notably, several of these features are entered to the database manually, and contain substantial errors and omissions. In particular, cargo details are too broad to provide any meaningful information pertaining the commodity carried by each vessel. Examples of broadcasting features are: location, speed over ground, course over ground, heading, status, etc. Each file contains point location data at 1-minute interval, per month and UMT zone (NOAA Office for Coastal Management, 2018). In addition to vessel positioning, the AIS system captures information that may be used for freight planning purposes. In particular, AIS data includes the type of vessel,

size, and the potential ability to track a vessel path with time stamps. This information may be suitable to identify freight flows though U.S. inland navigable waterways, and in combination with highway and USACE Locks data, constitute a valuable source for freight planning purposes. The main limitation of AIS data is its lack of information about the commodity carried by vessels, which is complemented by combining AIS and LPMS data. The USACE has used AIS data to evaluate travel time and reliability on waterways (Transportation Research Board, 2014), but it is yet to be integrated with truck GPS data and with commodity databases to evaluate multimodal freight fluidity.

#### A.2. Data from the Waterborne Commerce Statistics Center (WCSC)

The Waterborne Commerce Statistics Center (WCSC) is a division of the Institute of Water Resources, operated and maintained by the U.S. Army Corps of Engineers (USACE). The WCSC makes waterborne data and statistics available to the public at no charge, via the Navigation Data Center (NDC). Examples of the data found at the WCSC are: manuscript cargo data, manuscript trips data, a complete dock list, a list of the principal ports of the U.S., U.S. flagship vessel characteristics, commodity data collected through the Lock Performance Monitoring System (LPMS), and the Commodity Movements from Public Domain Database, among others. The following paragraphs provide a brief description of the most relevant of these datasets.

#### A.2.2. Lock Performance Monitoring System

Probably one of the most important pieces of WCSC data for this research, the Lock Performance Monitoring System (LPMS) is operated and maintained by USACE. The USACE collects data of a complete sample of U.S. flag vessels and foreign vessels operating in U.S. waterways that transit a USACE-owned or operated lock structure; which is managed and shared by the Navigation data Center (NDC) (U.S. Army Corps of Engineers). Publicly available records summarize annual and monthly tonnage of a series of commodities carried by vessels at each lock chamber and direction. Products shipped are classified into nine commodity groups (Table 6.1). Details on specific companies or commodities are confidential and not included in the dataset. In addition, the database provides information about the total number of different types of vessels observed at each lock chamber. Examples of vessel types are: tows, recreation, commercial, and other. The information is organized in a series of reports, available in .pdf and .xlsx format. For example, annual summaries of lock use, performance, and characteristics are available in a Commodity report, a Lock usage report, and an Unavailability report (U.S. Army Corps of Engineers, 2018c). Monthly tonnage summaries per commodity and lock chamber are available for download in excel format for the current and previous year exclusively (U.S. Army Corps of Engineers, 2016).

#### A.2.3. Commodity Movements from Public Domain Database

The Public Domain database is comprised of a series of .pdf reports which indicate the annual tonnage of more than 100 commodities by origin, destination, and commodity group. This is the only open-source, publicly available waterborne source that includes data on the origin and destination of waterborne commerce. This is also one of the most complete waterborne data source in terms of commodity disaggregation. The database is available in .pdf and, since 2011, in excel form, but not in GIS form. For the purpose of this research, unfortunately, the commodity movements are aggregated in annual tonnages, and the origins and destinations are spatially defined as broadly as U.S. States. (U.S. Army Corps of Engineers, 2020)
# A.2.4. Manuscript Cargo and Trips Data

The Manuscript Cargo Data contains the consolidated annual tonnage of more than 100 products observed at more than 500 river systems and principal ports of the U.S. The database indicates the type of movement (inbound receiving, outbound shipping, or thru), the reporting year, region (where the port or navigation system is located), and commodity code. It can be downloaded by the WCSC website in the form of a spreadsheet. Even though the annual consolidation of tonnage does not support the temporal resolution required for this research, this database may be used to further disaggregate the simple commodity breakdown of the LPMS dataset (only 9 commodity groups) into more detailed markets. In addition, the Manuscript Trips Data informs the number of annual upbound and downbound trips made on each river system and principal ports of the U.S., by vessel, during the reporting year. For confidentiality purposes, vessels are not identified on the database. Instead, the vessel's draft (in feet) and type is provided (dry cargo barge, liquid barge, self-propelled dry, tanker, towboat, or other –cranes, etc.). (U.S. Army Corps of Engineers, 2018a)

## A.2.5. Waterborne Infrastructure

The WCSC provides a list of the "Principal Ports of the U.S.". In this annually-updated list, the importance of a port is measured by the total tonnage of all commodities handled by the port during the reporting year. The spreadsheet includes the annual tonnage, name, and location (latitude and longitude) of the top-150 ports. Notably, only one port in Arkansas is included in the 2016 report: Helena port. (U.S. Army Corps of Engineers, 2018b). A much more comprehensive compilation of maritime U.S. facilities is included under the "Master Docks Plus" in the form of an access database. This database contains more than 40,000 facilities, identified as docks, fleeting areas, locks and/or dams, and milepoints. The publicly available

version of this database indicates the waterway where each facility is located, as well as its location (latitude and longitude), name and identifier code, and commodities handled (but not its volume), among others. The data is collected by survey; it is accompanied by a database schema and data dictionary, and can be downloaded form (U.S. Army Corps of Engineers, 2019).

#### A.3. Rail Carload Waybill

The Rail Carload Waybill (RCW) consists of a stratified sample of annual rail shipment data that must be filed by all railroads that operate more than 4,500 cars per year in the U.S. The sample is collected by a partnership between the Surface Transportation Board and the Federal Railroad Administration. Even though the complete RCW is confidential, a comprehensive Public Use Waybill File (PUWF) is open-sourced to the public. Among its 61 features, the PUWF includes shipment origin, destination, commodity, weight, number of cars, revenue, haul length, date, whether the cargo is hazardous, etc. (Cornell Law School). In addition, the RCW database identifies intermodal movements. The PUWF is available on a text (.txt) file and lacks any geographic references (such as latitude, longitude, or geometry), preventing users to directly map it into a GIS environment. In particular, the 2016 RCW sample includes 649,772 shipment waybills, covering more than 45 commodity groups, classified as per 5-digit Standard Transportation Commodity Codes (STCC) (Railinc, 2018). The STCC which appear more often on the 2016 RCW are: 01-Farm products, 11-Coal, 13-Crude petroleum, natural gas or gasoline, 28-Chemicals, and 37-Transportation equipment. The RCW is the only publicly available database providing commodity flow data for the railway system, and it is utilized by the Federal Railway Agency (FRA) to analyze the rail movements of hazardous materials and support safety and security (Wright et al., 2017).

One limitation of RWC is that origins and destination are broadly defined as 172 Business Economic Areas (BEA) within the U.S., complemented with 13 foreign zones within Mexico and Canada. This spatial resolution is not detailed enough to meet the objectives of this dissertation. A second challenge to meet the objective of this dissertation is associated with mapping the RCW data. As mentioned earlier, RCW lacks spatial references that enable mapping. On this matter, the FRA developed a toolkit to map the confidential RCW into a proprietary GIS software (Volpe). The toolkit consists of python scripts, available free of charge from FRA upon request, for research purposes. The toolkit requires supporting data, some of which is proprietary: i) a GIS rail network file (publicly available at the Geospatial webpage, Bureau of Transportation Statistics (Bureau of Transportation Statistics, 2020)), ii) a Centralized Station Master file, copyrighted by Railinc (Railinc, 2020), and iii) data on flow rights and weight (million gross tons), proprietary to the Class 1 railroads. Alternatively, geocoded North America Rail Stations are available at PC\*Miler|Rail, a Trimble licensed software (Trimble, 2018).

### A.4 Truck GPS Data

Truck GPS data consists of vehicle positioning data (latitude and longitude) emitted by GPS devices onboard a truck. The spatial coverage in the US is almost ubiquitous (Transportation Research Board, 2014). Private truck fleets typically record positioning data of their own trucks, for security and route tracking purposes, fuel cost and other operational optimization analysis. The American Transport Research Institute (ATRI), part of the American Trucking Association, gathers anonymous truck GPS data from a number of private fleets. In cooperation with FHWA, truck GPS data gathered by ATRI is used for diverse purposes, such as bottleneck identification, travel time analysis, border crossings, truck parking and hours of

services tracking, rerouting, etc. (Transportation Research Board, 2014). Truck GPS data can be acquired from private vendors.

Truck GPS data is a valuable source of truck routing, time-of-day corridor usage, volume and speed data. For reference, GPS data in Arkansas for 2016 represents about 35 million raw data points per week corresponding to approximate 40,000 unique trucks. Because current sources of truck GPS data are samples of the total truck population, it is important to evaluate the spatial and temporal coverage for each application (Diaz-Corro et al., 2019). The spatial and temporal analysis based on truck GPS data has several advantages over other truck data, such as Weigh-in-Motion (WIM) or Annual Average Daily Truck Traffic (AADTT data) gathered by the Federal Highway Administration (FHWA). The main advantage is the broad spatial and temporal coverage of truck GPS data. From a spatial coverage point of view, truck GPS data covers every single road in the statewide network, while AADTT is restricted to fixed and few counting stations (Figure A.1). Even though the information derived from truck GPS data is comprehensive, it lacks the commodity carried or industry served by trucks and thus it needs to be complemented with other commodity databases, such as USDA (for agriculture) or from business sources (for other commodities).



Figure A.1 WIM and ADTT stations within Arkansas

The temporally continuous nature of the GPS data allows for robust time-of-day usage analysis. Previous studies show that GPS data is a sample of roughly 10% of the total population of trucks travelling on the roads (Pinjari et al., 2014). This was confirmed for the Arkansas data sample by comparing the volume of trucks on the GPS dataset at WIM stations, with the volume of trucks counted at those WIM stations in Arkansas (Hernandez et al., 2018). Coefficients of coverage of sample locations considered in this work are shown in Table A.2.

Quarter	Van Buren	Little Rock	Pine Bluff
Q1	15.69	16.58	11.76
Q2	14.02	9.91	11.12
Q3	14.53	10.39	10.45
Q4	16.74	13.28	13.00
Average	15.25	12.54	11.11

 Table A.2 Sample GPS Data Coverage Coefficients in Arkansas

## A.5. Business Data

Two potential sources of business data are InfoUSA and ReferenceUSA, both developed and maintained by Infogroup. ReferenceUSA consists of a series of databases, namely: Business locations, U.S. New Business, U.S. Healthcare, Canadian Business, U.S. Jobs and Internships, and U.S. Historical Business, among others. The databases are compiled from "hundreds of thousands" of public information sources, analyzed and verified annually by Infogroup staff (Infogroup, 2020). Reference USA has been utilized to build customized lists of employers by location, industry as per North American Industry Classification System (NAICS), or company size (measured in sales volume and in number of employees), and includes a mapping feature, suitable to map business locations. A business' search provides company name, executives, business type (major industry group or NAICS), geography, phone, business size, ownership, financial data, etc. The search has the option to include verified, unverified and/or closed business. Search results are available on either a list or a map, downloadable to excel/.csv files. Heat maps are created by an embedded tool with business of the same category. A transportation layer can be added to the map. On the other hand, InfoUSA has been available since 1972, and provides business mailing lists, consumer mailing lists, email lists, and marketing campaign services. Consumer databases are built from real state & tax assessments, voter registration files, utility companies, etc. Both databases are available for a subscription fee. In particular, InfoUSA may offer marketing services which go beyond the needs of this dissertation. Alternatively, ESRI offers business location and summary data for a fee. ESRI's data, however, is based on the same source than ReferenceUSA and InfoUSA: a business dataset of over 12 million establishments developed by Infogroup, providing the same information.

An open-source alternative to Infogroup-based business data is the County Business Patterns (CBP) program. Under this program, the United States Census Bureau has been publishing subnational economic data by industry annually since 1964. The CBP data is obtained from various Census Bureau programs, such as the Economic Census, Current Business Surveys, and Annual Survey of Manufacturers. The CBP data consist of the number of establishments by geographic area and industry, employment during the week of March 12, first quarter payroll, and annual payroll. The CBP geographic area covers the continental U.S., Puerto Rico, and Island Areas. The geographical resolution ranges from State, County, Metropolitan Area, ZIP code and Congressional District Levels. The data for establishments is presented by industry, using 6-digit NAICS codes. For confidentiality purposes, noise infusion is applied. Before being published, the CBP data is subject to various edits for quality assurance, such as validation of geographic coding, addresses, and industry classification. The database can be downloaded at (U.S. Census Bureau, 2018). The CBP is utilized by private business with marketing purposes, such as to analyze marketing potential, measuring sales effectiveness, advertising programs, setting sales quotas, and developing budgets. Public agencies utilize the data for administration and planning. Notably, CBP covers most NAICS industries but excludes crop and animal production (which may be supplemented by USDA data) (U.S. Census Bureau, 2020). Statewide databases incorporating business location data include: Arkansas Department of Transportation (ARDOT) business establishments and structures, and Arkansas Economic Development Commission (AEDC) gas stations and convenience stores.

## A.6. Data from the United States Department of Agriculture (USDA)

The United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS), performs the Census of Agriculture every five years. The latest data available was collected in 2017, concurrently with (and complementing) the Commodity Flow Survey. Data collected in 2017 was made available for the 50 continental U.S. States in April 2019; remaining data on U.S. Territories will be completed by July 2020. This highlights the lengthy process required to process and make data publicly available. The Census of agriculture includes all places from which \$1,000 or more of agricultural products are produced or sold on the census year (U.S. Department of Agriculture, 2020b). With the census data, the NASS produces a series of statistics and reports. Statistics are summarized at county, state, and national level. One of such comprehensive reports is the Geographic Area Series, including data about the number of farms; livestock, poultry and their products; land use; crops; irrigation; farm and operations' characteristics; ownership; income; production expenses; chemical use; etc. (U.S. Department of Agriculture, 2012). In addition, the NASS offers online visualization and free downloadable map layers showing land cover and acreage per crop type (i.e. corn, soybeans, wheat, hay, etc.), at county level. These cropland data layers are based on satellite imagery and suitable to a GIS environment. They can be downloaded from (U.S. Department of Agriculture, 2020a).

Moreover, the USDA consistently reports information about transportation of agricultural products through the Agricultural Marketing Center (AMC), Transportation Research and Analysis division (U.S. Department of Agriculture, 2020c). The data is free of charge, downloadable, and organized by research subjects, some of which and briefly described as follows:

• Grain transportation report. A weekly comprehensive report including (among others): grain transport cost indicators per mode; U.S. origins to export position price spreads (\$/bushel); other U.S. grain exports and imports information; and information per mode. For rail mode: carloads deliveries to port; railcar auction

offers; bids/offers for railcars to be delivered in the secondary market; tariff rail rates for unit and shuttle train shipments; tariff rail rates for U.S. bulk grain shipments to Mexico; and railroad fuel surcharges. For water mode: Illinois river barge freight rate; weekly barge freight rates for southbound shipments; barge grain movements through specific locks and dams by grain type; grain barge movements through Mississippi River locks 27; upbound empty barges through Mississippi River locks 27, Arkansas River lock and Dam 1, and Ohio River locks and dam 52; grain barges unloaded in the New Orleans region; Gulf vessel inspections, loading activity, and rates. The report is issued in .pdf format, with tables and figures provided in excel for the current year.

- Agricultural refrigerated truck quarterly. Information compiled from the weekly fruit and vegetable truck rate report, including the following refrigerated truck data: quarterly shipment tonnages by origin and commodity; regional quarterly truck rates (\$/mile) by origin and distance; quarterly truck rates (\$/mile) for U.S. average by distance, per mileage category; quarterly truck rates (\$/mile and \$/truckload) by origin-destination pair; and weekly availability by origin and commodity. Origins and destinations are represented as U.S. States, Mexico, or Canada. Approximately 37 fruit and vegetable commodities are included. However, quarterly truck shipments for all commodities and origins are not available. Those available are reported, but do not represent complete movements of a commodity. Tables are available in excel format.
- **Transportation of U.S. grains: a modal share analysis**. This report presents the tonnage of grain (identified per type, including corn, wheat, soybeans, sorghum,

and barley) transported by truck, rail, and barge, per year from 1978 to 2013. The data is broken-down into domestic and for export. Even though the topic of this report is relevant, the geographical data aggregation approach adopted makes it unfit for the purpose of this research.

- **Barge rates and movements**. Information on the transport of grain on the Mississippi, Ohio, and Arkansas rivers. Excel tables are provided, including: grain shipments' tonnage aggregated by quarter and total annual from 2000 to 2017, detailing the weekly tonnage observed in five locks and dams located on the river systems mentioned above, per commodity (corn, wheat, and soybean); and average weekly Mississippi river barge rates by quarter, for the same period.
- **Biofuels and co-products**. Public rail tariff rates and fuel surcharges for ethanol and DDGS shipped on class I railroads, by origin-destination pair. Origins and destinations are identified as 17 cities located along class I railway lines. The data is available in excel files and updated monthly.

## A.7. Data from the United States Energy Information Administration (EIA)

The EIA independent statistics and analysis section provides a valuable interactive map of the U.S. energy system. The map locates energy infrastructure and reserves, namely: coal mines, power plants, oil & gas refineries and processing plants, natural gas and HGL market hubs, oil & gas wells and platforms, energy resources and reserves, and transportation. Transportation infrastructure includes pipelines and transmission lines, crude oil rail terminals, petroleum product terminals, petroleum ports, natural gas underground storage, LNG terminals, and waterways for petroleum movements. Map views can combine all energy types or can be filtered by energy type, namely: biomass, coal, electricity, fossil fuel resources, geothermal, hydroelectric, natural gas, petroleum, solar, wind, and renewable energy power plants. The interactive map can add county boundaries, congressional districts, administrative boundaries and demography layers. The map can be displayed for the continental U.S. of focused to a single region or state (State energy profiles) (U.S. Energy Information Administration, 2017). In its commitment to make energy data more accessible, understandable, relevant, and responsive to user needs, the EIA provides open data through and Application Programming Interface (API) which requires a free sign-up (U.S. Energy Information Administration). Data may be downloaded to excel and google sheets through web add-ins.

#### A.8. Weigh-In-Motion (WIM) System.

In the U.S., efforts to collect weight data of moving trucks started in the early 1950s. Technology has evolved since, but the operational principle remains the same. Weigh-in-Motion (WIM) sensors measure axle loads of vehicles moving at normal highway speed, through signals recorded by devices typically embedded in the road surface. Data collected at WIM sites is utilized to derive the following information pertaining to each vehicle: speed, lane, time and date, wheel load, axle load, axle group load, gross vehicle weight, individual inter-axle spacings, overall vehicle length, and axle-based vehicle classification (Quinley, 2010). WIM data has been used in combination with inductive signature data for highly detailed truck-body vehicle classification (Hernandez et al., 2016). For research purposes, WIM data is available upon request from the FHWA to State Departments of Transportation, free of charge.

In Arkansas, WIM data is collected continuously by ARDOT. These sensors consist of a single inductive loop to detect and count traffic, with two weight sensors either straddling the loops or sandwiched between two loops. Weight sensors can be piezoelectric systems (polymeric, ceramic, and quartz), bending plates or load cells (Arkansas State Highway and

Transportation Department, 2013). Figure A.1 shows the location of WIM stations within Arkansas. The number and distribution of stations is somewhat scarce, and primarily capture vehicles that are travelling on the interstate/highway and not stopping at ports or intermodal facilities. No routing or commodity information is provided by WIM data. Thus, WIM data in Arkansas may is not suitable to identify project-catchment areas. However, in the event a WIM station is located within the catchment area of a project, it might be used to evaluate the characteristics of traffic (mainly weight) occurring in such area.

### A.9. Federal Motor Carrier Safety Administration (FMCSA).

The Federal Motor Carrier Safety Administration (FMCSA) at the U.S. Department of Transportation has the purpose to reduce crashes, injuries and fatalities involving large trucks and buses. In this context, the FMCSA created and maintains the Motor Carrier Management Information System (MCMIS) to monitor the amount, severity, and location of safety incidents where commercial motor carriers are involved (Federal Motor Carrier Safety Administration, 2018). The MCMIS consist of the following files: crash, census, inspection, and investigation. The census includes carrier registration information, such as legal name, U.S. DOT number, contact details, number of power units, fleet VMT, operation classification (i.e. authorized for hire or not, U.S. Mail, private passenger, state government, etc.) and cargo carried.

Some highway traffic monitoring management products and systems used for law enforcement and truck compliance can read license plates and U.S.DOT numbers. The U.S.DOT number is then linked to the FMCSA database so officers can observe, among others, the commodities authorized to be carried by the fleet to which the truck belongs (International Road Dynamics, 2020). Within FMCSA's census "cargo carried" section, the carrier selects its fleet's cargo from a series of 30 commodities. Examples of such commodities are: building materials,

grain, paper products, refrigerated food, oil field equipment, livestock, chemicals, beverages, logs, metals, and general freight. The census file is updated monthly and is available for free download at (Federal Motor Carrier Safety Administration, 2020a). However, the downloadable version does not contain cargo carried details. These details can be obtained on a website search (Federal Motor Carrier Safety Administration, 2020b). However, for fleets that carry general freight or a large number of different commodities, the commodity identification via FMCSA is not conclusive to discriminate the commodity that each individual truck is carrying at any time.

#### A.10. National Performance Management Research Data Set (NPMRDS).

Briefly, the NPMRS consists of average travel time data on the National Highway System (NHS, as defined by FHWA), obtained by FHWA and updated monthly. The main purpose of this dataset is to support the calculation of performance measures (FHWA, 2020). Travel times include both passenger and freight activity. Data can be downloaded by state or region (the U.S. is divided into four regions). It is available in shapefile form, exclusively to State Department of transportations and metropolitan organizations since July 2013 (FHWA; National Operations Center of Excellence, 2017). Additional road coverage beyond the NHS may be requested to FHWA separately and is not a part of the NPMRDS. However, because of the geographical limitations of this non-open-source dataset, it is not used for this research.

## A.11. References

- Arkansas State Highway and Transportation Department. (2013). *Traffic Handbook*. Retrieved from http://www.arkansashighways.com/System\_Info\_and\_Research/traffic\_info/documents/tr affic\_handbook\_2014.pdf
- Bureau of Transportation Statistics. (2019). National Transportation Atlas Database Ports. Retrieved from https://data-usdot.opendata.arcgis.com/datasets/ports
- Bureau of Transportation Statistics. (2020). North American Rail Lines. Retrieved from https://osav-usdot.opendata.arcgis.com/datasets?keyword=Rail

Procedures for the Release of Waybill Data.

- Diaz Corro, K., Akter, T., & Hernandez, S. (2019). Comparison of Overnight Truck Parking Counts with GPS-Derived Counts for Truck Parking Facility Utilization Analysis. *Transportation Research Record*, 2673(8), 377-387. doi:https://doi.org/10.1177/0361198119843851
- DiJoseph, P., & Mitchell, K. (2015). *Estimating Vessel Travel Time Statistics for Inland Waterways with Automatic Identification System Data*. Paper presented at the Transportation Research Board 94th Annual Meeting, Washington, DC.
- Dobbins, J., & Langsdon, L. (2013). Use of Data from Automatic Identification Systems to Generate Inland Waterway Trip Information *Transportation Research Record*, 2330(1), 73-79. doi:https://doi.org/10.3141/2330-10
- Federal Motor Carrier Safety Administration. (2018). *Pocket Guide to Large Truck and Bus Statistics*. Retrieved from https://cms8.fmcsa.dot.gov/safety/data-and-statistics/2018pocket-guide-large-truck-and-bus-statistics
- Federal Motor Carrier Safety Administration. (2020a). Motor Carrier Census Information. Retrieved from https://ai.fmcsa.dot.gov/SMS/Tools/Downloads.aspx
- Federal Motor Carrier Safety Administration. (2020b). Search for Carrier. Retrieved from https://ai.fmcsa.dot.gov/SMS/Search/Index.aspx
- FHWA. Freight Management and Operations Performance Measurement. Retrieved from https://ops.fhwa.dot.gov/freight/freight\_analysis/perform\_meas/index.htm#data
- FHWA. (2020). Operations Performance Measurement. Retrieved from https://ops.fhwa.dot.gov/perf\_measurement/index.htm
- Hernandez, S., Tok, A., & Ritchie, S. (2016). Integration of Weigh-in-Motion (WIM) and Inductive Signature Data for Truck Body Classification. *Transportation Research Part C: Emerging Technologies*, 68, 1-21. doi:https://doi.org/10.1016/j.trc.2016.03.003
- Hernandez, H., Akter, T., & Diaz, K. (2018). *Truck Activity Analysis Using GPS Data*. Retrieved from https://www.arkansashighways.com/System\_Info\_and\_Research/TRC/Presentations/201 8%20Spring%20TRC%20Presentations%20in%20pdf/5-17-2018%20Thursday/203-204/Session%201/2%20-%20TRC1702\_update\_May2018%20v2.pdf
- Infogroup. (2020). Reference USA. Retrieved from http://resource.referenceusa.com/about-us/
- International Road Dynamics. (2020). ITS Solutions. Retrieved from https://www.irdinc.com/pages/its-solutions.html
- National Operations Center of Excellence. (2017). NPMRDS Users Quarterly Technical Assistance Webinar. Retrieved from https://transportationops.org/event/nationalperformance-management-research-data-set-npmrds-users-quarterly-technical-assistance
- NOAA Office for Coastal Management (Producer). (2018, 04). Vessel Traffic Data. *MarineCadastre.gov*. Retrieved from https://marinecadastre.gov/ais/

- Pinjari, A. R., Zanjani, A. B., Thakur, A., et al. (2014). Using Truck Fleet Data in Combination with Other Data Sources for Freight Modeling and Planning. Retrieved from
- Quinley, R. (2010). WIM Data Analyst's Manual. Retrieved from https://www.fhwa.dot.gov/pavement/wim/pubs/if10018/if10018.pdf
- Railinc. (2018). 2016 Surface Transportation Board Carload Waybill Sample Reference Guide. Retrieved from https://www.stb.gov/stb/docs/Waybill/2016%20STB%20Waybill%20Reference%20Guid e.pdf
- Railinc. (2020). Centralized Station Master. Retrieved from https://public.railinc.com/resources/centralized-station-master
- Transportation Research Board. (2014, May). *Circular E-C187* Paper presented at the Developing Freight Fluidity Performance Measures: Supply Chain Perspective on Freight System Performance, Washington, DC.
- Trimble. (2018). North America Rail Stations, FSAC, and Rail SPLC File with Latitude/Longitude Coordinates. Retrieved from https://support.pcmiler.com/en/support/solutions/articles/19000089553-north-americarail-stations-fsac-and-rail-splc-file-with-latitude-longitude-coordinates
- U.S. Army Corps of Engineers. Navigation Data Center.
- U.S. Army Corps of Engineers. (2016). Lock Performance Monitoring System. Public Lock Commodity Report. *Monthly Tons Report*. Retrieved from http://www.navigationdatacenter.us/lpms/Public%20Lock%20Commodity%20Report%2 0(CDM)-2016.xlsx
- U.S. Army Corps of Engineers. (2018a). Manuscript. *Digital Library [2000-2018 trips] Manuscript cargo and trips data files*. Retrieved from https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/1671
- U.S. Army Corps of Engineers. (2018b). Principal Ports of the United States.
- U.S. Army Corps of Engineers. (2018c). USACE Digital Library Public lock reports locks by waterway, lock usage, CY 1993 2017. Retrieved from https://usace.contentdm.oclc.org/digital/collection/p16021coll2/id/2958
- U.S. Army Corps of Engineers. (2019). *Master Docks Plus Public Extract*. Retrieved from https://publibrary.planusace.us/#/document/e3b93389-2622-4cfa-a370-3fc037c2dc6d
- U.s. Army Corps of Engineers. (2020). Commodity Movements from the Public Domain Database. Retrieved from https://publibrary.planusace.us/#/series/Commodity%20Movements%20from%20Public %20Domain%20Database
- U.S. Census Bureau. (2018). County Business Patterns: 2016. https://www.census.gov/data/datasets/2016/econ/cbp/2016-cbp.html

- U.S. Census Bureau. (2020). County Business Patterns About This Program. Retrieved from https://www.census.gov/programs-surveys/cbp/about.html
- U.S. Department of Agriculture. (2012). *Census of Agriculture*. Retrieved from https://www.nass.usda.gov/Publications/AgCensus/2012/
- U.S. Department of Agriculture. (2020a). CropSpace and Cropland Data Layer. Retrieved from https://www.nass.usda.gov/Research\_and\_Science/Cropland/Release/index.php
- U.S. Department of Agriculture. (2020b). *Guide to Products and Services*. Retrieved from https://www.nass.usda.gov/Publications/catalog.pdf
- U.S. Department of Agriculture. (2020c). Transportation Research and Analysis. Retrieved from https://www.ams.usda.gov/services/transportation-analysis
- U.S. Department of Homeland Security (Producer). Automatic Identification System *Navigation Data Center*. Retrieved from https://www.navcen.uscg.gov/?pageName=AISmain
- U.S. Energy Information Administration. Open Data. Retrieved from https://www.eia.gov/opendata
- U.S. Energy Information Administration. (2017). State Profiles and Energy Estimates. Retrieved from https://www.eia.gov/state/
- Volpe. Waybill GIS Toolkit Documentation. Retrieved from
- Wright, R., & Baker, G. (2017). Modernizing FRA's Rail Freight Routing and Analysis Tools. Paper presented at the ESRI user Conference, Boston, MA. https://proceedings.esri.com/library/userconf/proc17/papers/2185\_504.pdf