MARITIME TRANSPORTATION RESEARCH AND EDUCATION CENTER
TIER 1 UNIVERSITY TRANSPORTATION CENTER
U.S. DEPARTMENT OF TRANSPORTATION

MarTREC
Maritime Transportation Research & Education Center

Mississippi Multimodal Freight Analysis Model

October 2021 – December 2023

Principal Investigator: Dr. Tzusheng Pei, Jackson State University, Tel: 6019798252, email: tzusheng.pei@jsums.edu
Co-Principal Investigator: Dr. Ali Abu-EL Humos, Jackson State University, Tel: 6019793319, email: ali.a.humos@jsums.edu

FINAL RESEARCH REPORT

Prepared for:
Maritime Transportation Research and Education Center

University of Arkansas
4190 Bell Engineering Center
Fayetteville, AR 72701
479-575-6021

ACKNOWLEDGEMENT
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.

DISCLAIMER
The contents of this report 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 report 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.

1. Project Description

The Mississippi Department of Transportation [1] strives to facilitate a safe and efficient movement of freight within the state due to the growing need to understand what is causing congestion [2] and improve safety measures. The scope of this project is to develop a multimodal freight transportation analysis model for Mississippi for assessing the demand for transportation facilities and services. This research will enable policymakers, transportation planners, and logistic analysts in various federal, state, and local agencies to assess the demand for transportation facilities and services, energy use, safety, risks, and environmental concerns. Implementation of state-level strategies that support efficient freight movement is therefore essential not only for attracting new industries to move freight into and out of the state, but also for addressing the needs of existing state businesses. This work provides an analysis model on shipments to and from Mississippi. We use origins and destinations of shipments and their business addresses (zip code to zip code) to calculate route path, mileage, and total travel time. PC*MILER [3] and ArcGIS [4] are used to calculate routed and geodesics distances. This will generate a huge amount of routing records which will be filtered out to determine the optimal routing record for each shipment. Only business address data from the states of Mississippi and California are collected to calculate routed distance information.

2. Methodological Approach

Commodity flows are a defining aspect of the Mississippi regional economics' viability and livability. Mississippi's regional and local economy substantially benefits from increased intra-, inter-and through regional freight activities between different trading partners and intermodal facilities. Implementation of state-level strategies that support efficient freight movement is therefore essential not only for attracting new industries to move freight within, into, or out of the region but also for addressing the needs of existing businesses.

The first step was to collect necessary shipments data. The bulk of the data came from the Bureau of Transportation Statistics, U.S. Department of Transportation, and U.S. Census Bureau, U.S. Department of Commerce. Ascertaining the required data was a major part of the research. Initially, we were using the 2017 CFS (Commodity Flow Survey) PUF (Public Use File) [5, 6] and its associated businesses address information [7] where the necessary data was extracted. Once the data from the CFS file was filtered to the domain of interest, PC*MILER and ArcGIS were used to calculate the routed and geodesics distances between origins and destinations for each shipment.

Primarily, the data from the CFS file were so huge (numbering in the millions), so it was decided that the focus point should be shipments to or from California and Mississippi. Then, the shipment data was matched to businesses address information using NAICS (North American Industry Classification System) code [8]. Finally, routed and geodesics distances were calculated, and optimal records were extracted based on the matched routed and great circle distances from PUF.

The Data

1. Individual shipment data

The individual shipment data was found in the 2017 CFS PUF. This file includes Origin State (ORIG_STATE), Origin Metropolitan Area (ORIG_MA), Destination State (DEST_STATE), Destination Metropolitan Area (DEST_MA), NAICS Codes (NAICS), Shipment Great Circle Distance (SHIPMT_DIST_GC), and Shipment Routed Distance (SHIPMT_DIST_ROUTED). Also, there are no metropolitan areas in Mississippi (only see MA code 99999 when the state is Mississippi), and that was an issue when calculating the distances.

	SHIPMT_ID	ORIG_STATE	ORIG_MA	DEST_STATE	DEST_MA	NAICS	MODE	SHIPMT_VALUE	SHIPMT_WGHT	SHIPMT_DIST_GC	SHIPMT_DIST_ROUTED
1	0000001	06	99999	06	260	326	05	4380	391	54	60
2	0000002	49	482	47	314	4541	14	56	4	1524	1810
3	0000003	06	348	06	348	4231	05	255	440	2	5
4	0000004	06	260	06	99999	212	05	250	44912	30	35
5	0000005	45	273	45	273	45431	05	46	73	9	11
6	0000006	48	12420	36	160	4541	14	605	4	1348	1561

Fig. 1*: The actual shipment data (key columns) from CFS dataset*

2. Businesses addresses data

The objective of this research was to focus on shipments between California and Mississippi. The business addresses in the Mississippi Business Report were used [7]. This report included all current registered businesses in Mississippi (Most businesses are in Mississippi, and some are in other states such as California). After we checked the California businesses website, it was noted that the data only included company addresses, but not any relevant business identification or classification information like NAICS code, so the data was concluded as ineffective for the research.

3. Other Data

The data in PUF file has FIPS (Federal Information Processing Standards) to represent origin and destination states (column ORIG_STATE & DEST_STATE). To get the state names for the codes and the USPS (United States Postal Service) State Abbreviations, the FIPS codes table [9] was needed.

Methodology

The origin and destination in a shipment in the 2017 CFS PUF data were from a metropolitan area to a metropolitan area. The method is to try finding all possible routed distances and geodesics distances for all potential ZIP Codes pairs. Then, compare them with the values given in PUF as a reference to get the optimal result.

Firstly, according to the state and the metropolitan area information of every origin and destination pair, all potential businesses in Mississippi Business Report are assigned to this shipment pair. After this, several potential zip code pairs and coordinate pairs are assigned for each individual shipment.

Secondly, PC*MILER and ArcGIS were used to find the routed and geodesic distances for each coordinate pair obtained in the first step.

Then we use the following formulas to calculate the required distances:

$$\text{Routed Distance Difference} = \text{Routed Distance by PC*MILER} - \text{Routed Distance in 2017 CFS PUF} \quad (1)$$

$$\text{Geodesic Distance Difference} = \text{Geodesic Distance by ArcGIS} - \text{Great Circle Distance in 2017 CFS PUF} \quad (2)$$

$$\text{Weighted Distance Difference} = \text{Abs (Difference of Routed Distance)} * 50\% + \text{Abs (Difference of Geodesic Distance)} * 50\% \quad (3)$$

Finally, the smallest weighted distance difference for each shipment is obtained, and it is considered the optimal result, for the purpose of narrowing down a shipment from metropolitan-metropolitan pair to zip code - zip code pair.

1. Geodesic distance

A geodesic distance is the shortest distance between two points on the surface of a sphere, measured along the surface of the sphere (as opposed to a straight line through the sphere's interior). The distance between two points in Euclidean space is the length of a straight line between them, but on the sphere, there are no straight lines. In a space with curvature, straight lines are replaced by geodesic line. Geodesics on the sphere are circles on the sphere whose centers coincide with the center of the sphere and are called 'great circles'.

First, ArcGIS is used to calculate great circle distances of some ZIP code pairs, but a more flexible alternative way was found called Geopy [10]. Geopy is a Python library that simplifies the calculation of geographic distances between two points. It makes it easier for developers/programmers to retrieve coordinates of various locations using third-party geocoders. It contains the function geodesic () for geodesic distance and the great_circle() or great circle distance. The results from ArcGIS and Geopy were compared. It was found that the results from ArcGIS were closer to the results from geodesic(), but not those from great_circle(). The difference between geodesic() and great_circle() is about 0.18% to 0.21%. The difference between using ArcGIS and geodesic () is about 0.000183% to 0.099%.

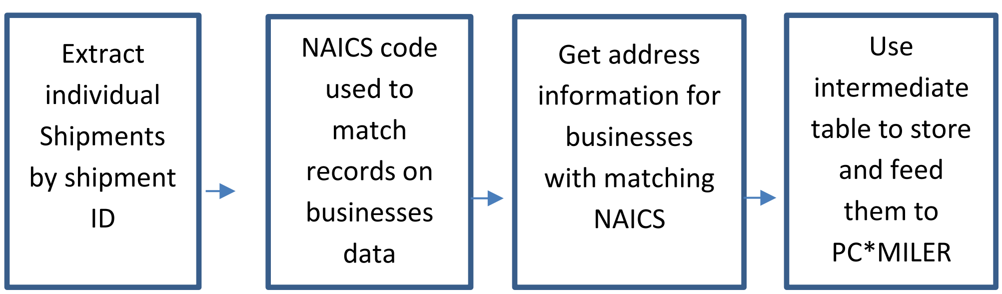
The “great_circle ()” considers Earth as a sphere, but the Earth is an ellipsoid. If the Earth is treated as a sphere, the geodesics are great circles (all of which are closed) and the problems reduce to ones in spherical trigonometry. However, Newton (1687) showed that the effect of the rotation of the Earth results in its resembling a slightly oblate ellipsoid. So, geodesic() in Geopy was chosen instead of ArcGIS.

2. Routed Distance

The PC*MILER software provides accurate routed distance, mileage, and travel time information with address-to-address routing. It also provides transit times and estimated times of arrivals using predictive and real-time traffic data specifically designed for commercial vehicles. The PC*MILER Web services [11] API (Application Programming Interface), such as geocoder API, route matrix API were used to convert address line information into coordinate data and to calculate routed distance, travel time between given source and destination pairs. Once the calculated values were obtained, it was observed that the values were not very close to the routed distances given in 2017 CFS PUF data. Once this understanding was established, it was collectively thought of the following scenario:

The truck taking this shipment was not travelling directly from the origin to the destination, instead, it might have gone to a hub close to the origin, for loading or unloading goods, then might have gone to a hub close to the destination. Due to this intermediate transit, there may be a variation in the number of miles and travel time and the PUF does not provide any of these intermediate stop's information. So, to check more shipment data and see if there are any shipments that have a nearby matching is considered for the next phase of the research.

CFS and Businesses (Mississippi) Dataset preprocessing:



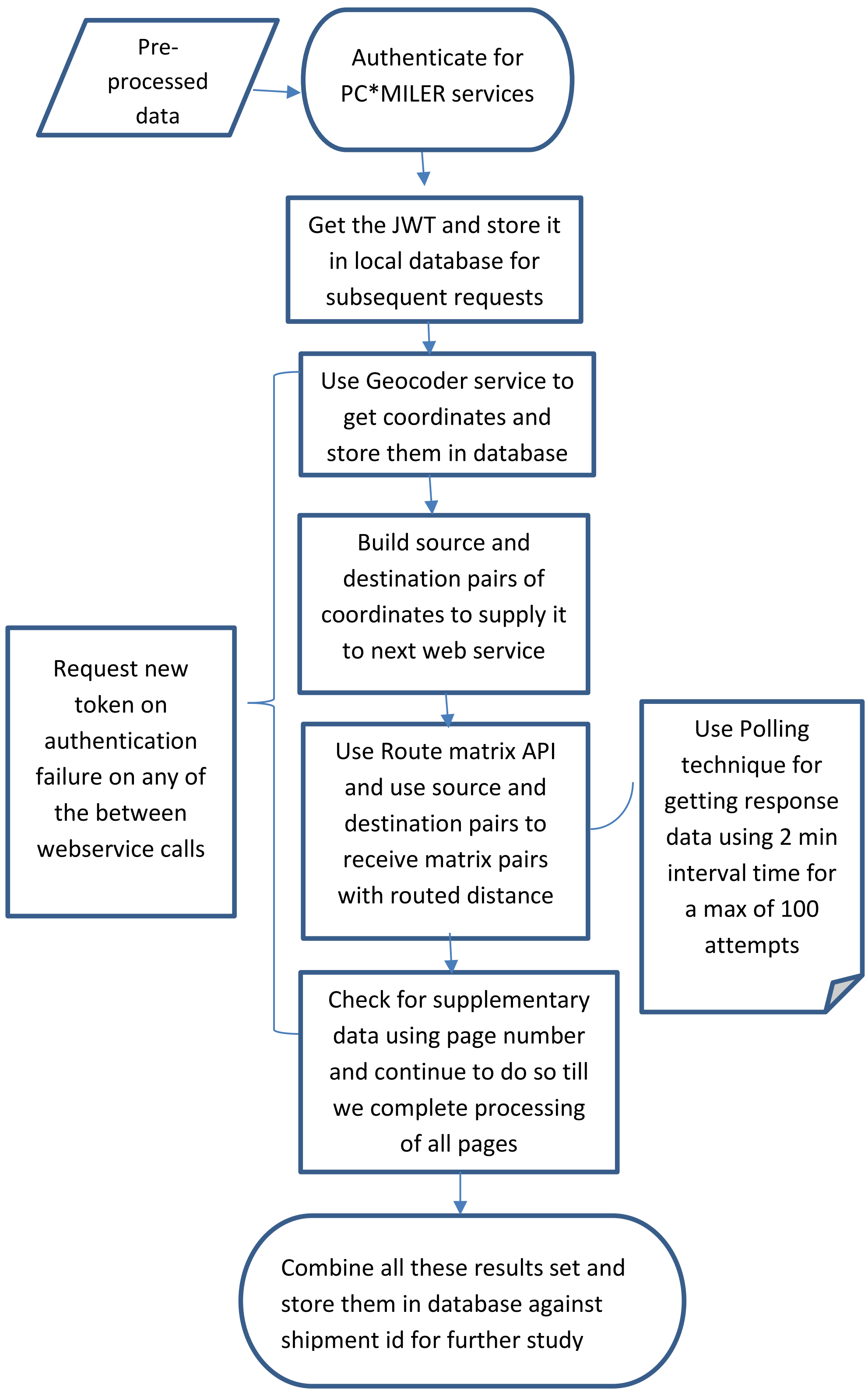
The above process flow shows the data pre-processing steps on how to manage shipment data by ID and to use them for processing on further steps

PC*MILER Mapping service:

The PC*MILER Web services API was used to get the routed, travel time information, and the data required for this service was taken from the final step in the above preprocessing data procedure as the input. The application processes the data by shipment_id as discussed above to feed them to the integrated PC*MILER's Web services API. And currently, geocoder and matrix routed distance API services were used for this purpose.

Each of the PC*MILER services requires the API key to get the token and the received JWT (JSON Web token) was authenticated first. The JWT is used until its expiry to access the required Web services. Here the geocoder API service was used to translate the address line information into coordinates data associated with businesses for each shipment and was stored for use in a further step. While extracting the data from the database, it was categorized as the origin and destination pairs, and the focus was on shipments to and from Mississippi and California. In the next step, the location information, i.e., the coordinates (latitude and longitude), was used to build a source-destination matrix pairs to be the input to PC*MILER Matrix route API. The data used here has unique origin-destination coordinates. This format was used for bulk processing data. Suppose if there are 300 source and 40 destination pairs, it was expected to have 12000 (300 times 40) records in the total result set.

Flow chart for accessing PC*MILER Web services API



3. Results/Findings

Six - Shipments with actual shipmentID and corresponding miles that were optimal with CFS dataset on calculated routed distance and great circle distance dataset were plotted in Figure 2 and Figure 3.

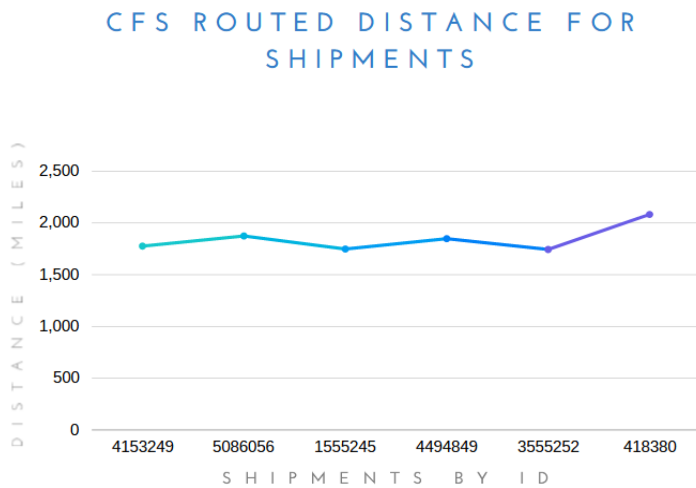


Fig. 2*: Shipments and their distance travelled in miles extracted from CFS dataset for Mississippi and California*

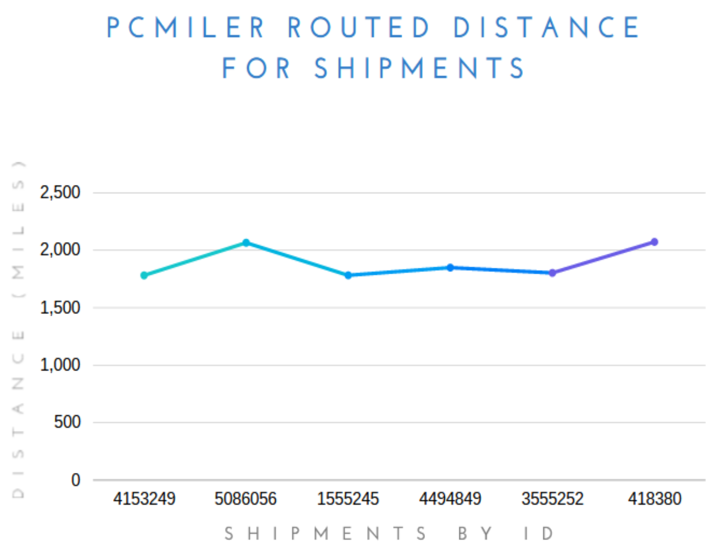


Fig. 3*: Shipments and their distance travelled in miles calculated from PC*MILER for Mississippi and California*

The ratio of CFS routed distance / CFS Great Circle distance was calculated as listed in Fig. 4:

Ratio of Routed Distance/Great Circle Distance		
From	To	Count
<=1		0
1	1.1	67
1.1	1.2	2348
1.2	1.3	288
1.3	1.4	179
1.4	1.5	83
1.5	1.6	10
1.6	1.7	0
1.7	1.8	0
1.8	1.9	0
1.9	2	0
2	2.1	0
2.1	2.2	4
>2.2		0

Fig. 4*: Summary of the Ratio of Routed Distance/Great Circle Distance related to Mississippi & California*

In the above figure, most shipment records were in the range 1.1~1.2. These ratios suggest that there may not be hubs between the origin and destination places. When the ratios are in 1 -1.2, or the respective hubs were very close to the origin or destination. For the range 1.2-1.6, there may be one or more hubs between the origin and the destination, and the hub may be far from the direction of the normal route from the origin to the destination. In this ratio range, the transporter may need to take a special route to the hub on the origin end and then route the truck to the destination, possibly through a hub on the destination end. In the ratio range 2.1-2.2, the truck may be routed through a hub which appears to be far from the destination and might have been routed to the destination from there.

4. Impacts/Benefits of Implementation

- Developed a CFS dataset at higher geographical resolution (zipcode) to serve as the foundation for freight-related analysis: The main objective of this study was to use the attributes provided in public-use data, specifically great circle distance and routed distance, to estimate the origin and destination at the zip code level.
- One publication:
    - Vivek Sunchu, Canhao Wang, Terrence Conley, Mercy Chebet, Shihmiao Chin,Tzusheng Pei, and Ali Abu-El Humos, "Mississippi Multimodal Freight Analysis Model", The 2022 World Congress in Computer Science, Computer Engineering, and Applied Computing (CSCE'22) July 25-28, 2022, Luxor (MGM), Las Vegas, USA.
- Two presentations:
    - Vivek Sunchu, Canhao Wang, Terrence Conley, Yu Wu, Shihmiao Chin,Tzusheng Pei, and Ali Abu-El Humos, "Mississippi Ground Freight Analysis Model", Research Engagement Week at Jackson State University, October 24-28, 2022, Jackson, MS.
    - Vivek Sunchu, Canhao Wang, Mercy Chebet and Terrence Conley: "Mississippi Multimodal Freight Analysis Model". Jackson State University Engineering Week February 22, 2022, invited project presentation by the College Dean, Jackson State University.
- One African American and two female students were funded through this project.

5. Recommendations and Conclusions

This research provides insights and data availability for demand routes, travel time with real-time traffic information, and mileage information for optimized route planning specifically for commercial vehicles. It also helps to enhance existing infrastructure, and facilities on demand travel routes for better freight movement inbound and outbound of the Mississippi.

Some future work activities may include:

- Extend the scope of research to other states to and from Mississippi.
- Build a comprehensive web interface and dashboard where the routing information for all shipments can be shown on a real time interactive map with attributes such as miles, travel tim, etc. It should also include the routing plan of a shipment being taken, estimated travel time to reach destination and mileage information.
- Finding the ratio and difference between the initial CFS data. The shipments that made progress stops in between origin and destination were identified to be affecting the accuracy of data.

References

[1] Mississippi Department of Transportation, "Mississippi online maps," 2022. [Online]. Available: https://mdot.ms.gov/portal/maps, January 2022.

[2] Mississippi Department of Transportation, "MDOT traffic resources," 2022. [Online]. Available: https://www.mdottraffic.com, December 2021.

[3] Trimble Inc., "Platforms," 2022. [Online]. Available: https://www.pcmiler.com/platforms/#web-api, May 2022.

[4] Esri ArcGIS, "Online map viewer," 2022. [Online]. Available: https://maps.arcgis.com/apps/mapviewer/index.html, January 2022.

[5] U.S. Census Bureau, "2017 commodity flow survey datasets," 2021. [Online]. Available: https://www.census.gov/data/datasets/2017/econ/cfs/historical-datasets.html, January 2022.

[6] U.S. Census Bureau, "Commodity flow survey (CFS)," 2022. [Online]. Available: https://www.census.gov/programs-surveys/cfs.html, January 2022.

[7] Mississippi Secretary of State, "Mississippi business reports," 2022. [Online]. Available: https://www.sos.ms.gov/business-services/business-reports, March 2022.

[8] U.S. Census Bureau, "North American industry classification system," 2022. [Online]. Available: https://www.census.gov/naics, March 2022.

[9] U.S. Bureau of Labor Statistics, "USPS State abbreviations and FIPS codes," 2005. [Online]. Available: https://www.bls.gov/respondents/mwr/electronic-data-interchange/appendix-d-usps-state-abbreviations-and-fips-codes.htm, April 2022.

[10] GitHub, Inc., "Geocoding library for Python," 2022. [Online]. Available: https://github.com/geopy/geopy, March 2022.

[11] Trimble Inc., "PC*MILER web services," 2022. [Online]. Available: https://developer.trimblemaps.com/restful-apis/developer-guide/introduction, May 2022.