Framework for Modeling Statewide Freight Movement Using Publicly Available Data

by Subhro Mitra and Denver Tolliver

This paper presents a methodology to model statewide truck trips using publicly available data developed by federal and state organizations. This methodology is applied for the statewide freight planning of North Dakota. In the absence of ample research funding, states, counties and metropolitan planning organizations can resort to these freely available data. A state level commodity-by-industry input-output table customized from a freely available national input-output table is used to disaggregate trips to the traffic analysis zone level. These databases are available nationally; hence, the methodology discussed in this paper can be transferred to other states with relative ease.

INTRODUCTION

Transportation planners are intimidated by the lack of data whenever they embark on a project involving freight movement – whether on a national, statewide or regional scale. Statewide freight planning is undertaken by Departments of Transportation (DOTs) and research organizations at different levels of accuracy and levels of disaggregation based on the research objectives and funds allocated for the project. There is literature on data requirements and data available for freight planning. A National Cooperative Highway Research Program (NCHRP) report on requirements of freight data for transportation planning (Roger Creighton Associates and R.L. Banks & Associates 1977) presents a catalog of freight data sources for use by state and regional planners. This manual is outdated, as there are many additions and alterations in the data sources. Similarly, the *Directory of Transportation Data Sources* (Bureau of Transportation Statistics 1995) is also now obsolete, as there are many additions and alterations to its list of data sources.

The Federal Highway Administration (FHWA) has acknowledged the importance of freight modeling and has a website dedicated to improving freight models (U.S. Department of Transportation 2007b). This "Freight Model Improvement Program" is a joint effort by the U.S. Department of Agriculture, the U.S. Department of Energy and the U.S. Army Corps of Engineers and is supported by the Oak Ridge National Laboratory. This website provides information about major sources of national freight transportation data. Top on the list is the *Commodity Flow Survey (CFS)* (U.S. Census Bureau 2007b). *Freight Analysis Framework (FAF)* (U.S. Department of Transportation 2007a), *GeoFreight* (U.S. Department of Transportation 2007c), *Rail Waybill Data* (Surface Transportation Board 2007), *Maritime Statistics* (U.S. Department of Transportation 2007d), and *Waterborne Commerce of the United States* (U.S. Army Corps of Engineers 2007) are other sources referred to in this website.

The Transportation Research Board (TRB) conducted a study to gain insight into various freight data needed by different users for freight planning, as well as to identify different freight data providers (Donnelly 2003). This study identifies different attributes of freight movement that are part of freight data, namely origin and destination, commodity characteristics, modes of movement, routing and vehicle configuration. This report, published by TRB, identifies the CFS and the Transearch database from Global Insight (2006) as the most comprehensive freight flow data in the country. This report also points out that these databases have their own deficiencies. FAF overcomes some of the deficiencies of the CFS data by incorporating CFS "out of scope data,"

comprised of goods generated from imports, publishing, farms, construction, logging services and fisheries.

This paper outlines a methodology to model statewide freight flow using publicly available data, which are freely downloadable from websites hosted by federal and state entities. This methodology, along with these databases, can be successfully used to estimate a truck origin-destination (OD) matrix disaggregated to transportation analysis zone (TAZ) levels, which are counties in this project.

Most freight models lack logistics aspects (De Jong and Ben-Akiva 2007). In the absence of detailed survey data of warehouse locations and distribution centers, truck count data from the state DOT and the matrix estimation method are used to introduce distribution legs in the OD pairs. This methodology, even with its shortcomings, produces a fairly good statewide freight model and is used to assess economic impacts of infrastructure investment in the state and various other facility location studies associated with freight movement.

In this statewide freight planning project, agricultural goods and commodities other than agriculture,[1] consisting of manufactured goods and mining products, are modeled separately. The data for the manufactured goods are obtained from the second generation freight analysis framework (FAF[2]) data, developed by the FHWA. The OD data, obtained from FAF[2], has origins and destinations as states and sub-state regions.[2] This OD data are disaggregated to the TAZ level using population and employment information available from *County Business Patterns* (CBP) data, which is developed by the U.S. Census Bureau (2007a), and commodity by industry input-output (I-O) tables, which are developed by Bureau of Economic Analysis (2006). The national I-O table is scaled down to the state level using the location quotient and is explained in detail later in this paper. In some statewide freight models, as in the Montana highway reconfiguration study, disaggregation of OD data is done using *Implan Professional software* (Minnesota Implan Group 2007). In the modeling framework discussed in this paper, no such professional software is used for disaggregation.

Some of the tasks undertaken in this project include:

- Identify sources of manufactured freight movement data, as well as data required for disaggregating flows in the public domain.
- Combine data from heterogeneous sources, such as public domain data and local survey results.
- Develop a methodology to subdivide the statewide freight model into commodity-based sub-models for analytical purposes and finally merge them into one.

The CFS does not provide any information about crop movement from fields to elevators. States like North Dakota, where agricultural production dominate the state economy, trucks hauling crops from fields to elevators account for a big share of freight movement in the state. In this project, data for agricultural freight are developed from satellite imagery of crop data layers (Mitra et al. 2007). This paper does not discuss the methodology for modeling agricultural freight; it only outlines the method for modeling manufactured goods.

DATA SOURCES

In this project, the main source of data is the second generation *Freight Analysis Framework* data (FAF[2]), developed by the FHWA in cooperation with the Bureau of Transportation Statistics through Oak Ridge National Laboratory and MacroSys Research and Technology (U.S. Department of Transportation 2007a). The commodity flow data available from FAF[2] are comprised of four dimensions: origin, destination, commodity and mode of transportation. The FAF[2] data, like the CFS data, have 114 regions, 17 international gateways and seven international regions. These data have 43 two-digit standard classifications of transported goods (SCTG) classes of goods and seven modes of transport. This OD database has three categories of data, namely the CFS within scope data, auxiliary data and CFS "out-of-scope" data.

Data available from different sources, such as the *Carload Waybill sample* (Surface Transportation Board 2007), *Domestic Waterborne Commerce of the United States* (U.S. Army Corps of Engineers 2007), *International Waterborne Commerce of the United States* (U.S. Army Corps of Engineers 2007), *Transborder Surface Freight* (Bureau of Transportation Statistics 2007), and *U.S. Air Freight Movement* (Bureau of Transportation Statistics 2007), are used in a log-linear model to fill in the voids of the CFS data. Certain classes of commodities are totally absent in the CFS database; some are partially absent, whereas some are in part of the supply chain. Class of commodities that are missing in the CFS data, but included in the FAF database, are farm-based products, fisheries, crude petroleum, natural gas, municipal solid waste, logging, construction, publishing, retail, imports, petroleum products, exports and in-transits. Three approaches are used to validate these data. In the first approach, a cell of the OD matrix is removed, the commodity OD data are estimated, and then the two tables are compared. In the second approach, a comparison is made between the derived parameters and the auxiliary data consisting of waterborne commerce, rail waybill and air carrier data. In the third validation approach, the absolute values of cell values calculated from out-of-scope commodities are compared to cell values calculated from auxiliary data sources (U.S. Department of Transportation 2007a).

The I-O commodity-by-industry coefficients are available from the benchmark I-O accounts developed by the U.S. Department of Commerce/Bureau of Economic Analysis (Bureau of Economic Analysis 2006). The 1997 benchmark I-O account, though out of date, is the most detailed commodity-by-industry account of the United States economy. In this study, the interactive table is used, where the user can aggregate commodities and industries according to his or her requirements. In the 1997 benchmark I-O account, there are commodity flows from 483 commodities groups to 491 industries and 13 end users. A salient feature of the 1997 I-O account is the classification of the commodities and industries based on the North American Industry Classification System (NAICS) as an alternative to the Standard Industrial Classification (SIC) classifications used in the previous year accounts.

MODEL DESCRIPTION

Overview of the Model

The basic modeling framework is shown in Figure 1. The prime source of data is the FAF[2] data. Manufactured goods consist of commodities other than agricultural products, namely mining, construction, wholesale and retail trade, petroleum and coal products. FAF[2] is a repository of OD data of manufactured goods moving in all modes of transport. In this statewide freight flow model, as we are interested with truck traffic only, manufactured goods transported in trucks are extracted from the database. In addition to 52 internal TAZs, there are 15 external TAZs, which are connected with the state highways at possible exits at the boundary locations with dummy links as shown in Figure 2.

This modeling framework has three stages: trip generation, trip distribution and trip assignment. There is no mode choice stage in the model, as the OD data queried out from the FAF[2] database are for truck only. The statewide freight flow model is broken into internal-external, internal-internal and external-external sub-models. The concept of onion model is used here, assuming that congestion is not an issue in most highway networks of the state. The onion model assignment is done in layers for individual commodities or sub-models, and the assigned network is finally merged to get the estimated trucks in the network. In this project, truck flow, assigned networks from the manufactured goods model, is merged with that from the agricultural freight flow model to get the total truck traffic in the network.

Figure 1: Methodological Framework of the Manufacturers' Freight Model
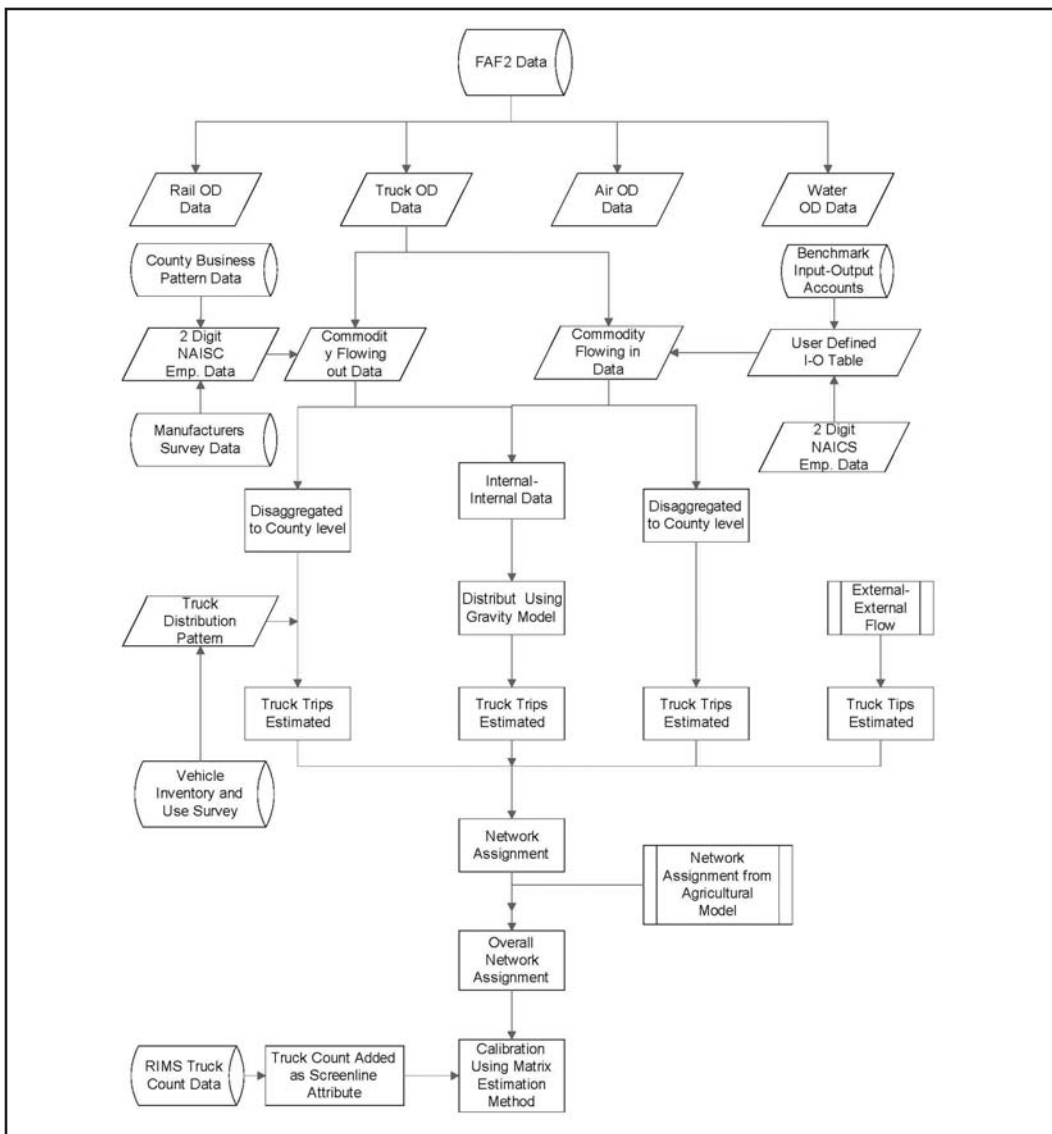
Analyzing FAF Data

The FAF[2] is the source of freight movement OD data between states, sub-state regions and major international gateways. The FAF[2] data, unlike the CFS data, provide both dollar value and tonnages for each OD pair; hence, it is not required to convert dollar value to tonnage using producer price indices (PPI). The first generation FAF data classified commodities based on Standard Transportation Commodity Code (STCC). In the second generation FAF, the goods are classified based on the Standard Classification of Transported Goods (SCTG), which is developed by the U.S. Department of Transportation, the U.S. Census Bureau and Statistics Canada. The SCTG covers goods in all modes of transport, and it results in uniformity of transportation data of the United States and Canada. SCTG also creates an integrated system of commodity classification for economic analysis and facilitates the North American Free Trade Agreement (NAFTA).

Figure 2: Internal and External TAZ for the North Dakota Freight Flow Model

Internal TAZs of ND
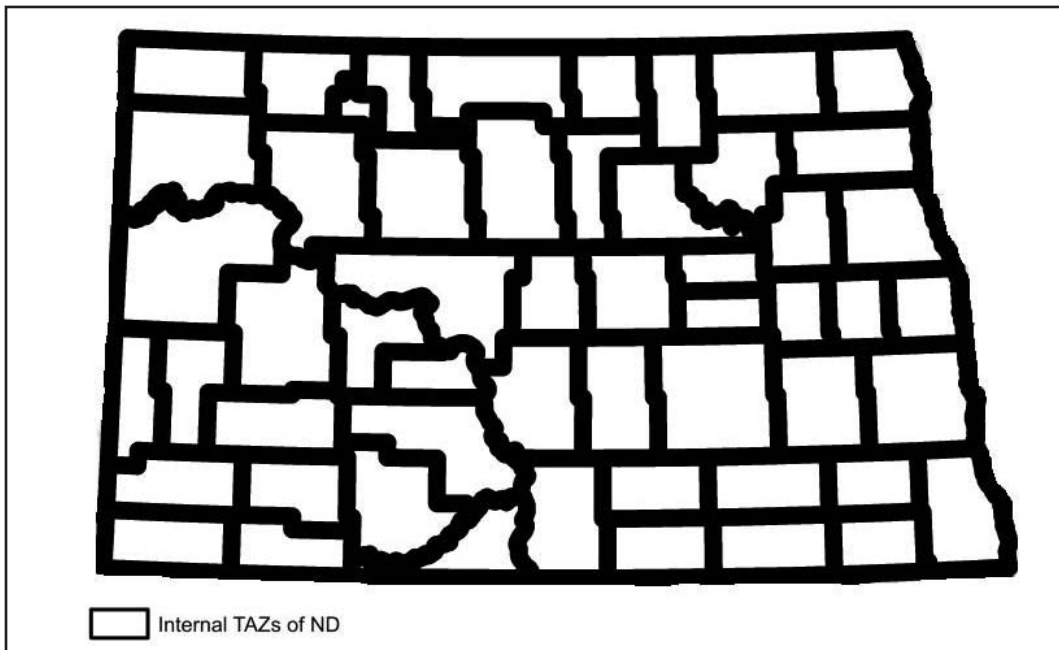
The FAF[2] OD data is a four-dimensional matrix – origin, destination, commodity and mode. The complete database is in Microsoft Access, and the name of the file is "FAFOD_2002.MDB." This file has information on commodity flows both within domestic regions and also to foreign countries. The Microsoft Assess table "FAFOD_DOM_2002" has the domestic origin and destination data; the "FAFOD_BRD_2002" table provides data for commodity movement across the border. The "FAFOD_SEA_2002" table has information about commodities transported by water. For this project, the two tables "FAFOD_2002.MDB" and "FAFOD_BRD_2002" are used, and required OD data are extracted using Structured Query Language (SQL). The first query is used to extract all OD data with origins in North Dakota and transported in trucks. From this list of commodities that moves out of North Dakota, 18 major commodities that make up approximately 97% of the total freight are selected, as shown in Table 1. In this table, it is seen that cereal grain constitutes 54% of the total freight moving out. As agricultural freight is modeled separately, cereal grain is excluded from this list. Milled grain products are included in this list, because they were not included in the agricultural freight flow model. The two-digit level SCTG classifications and their descriptions are available from the Bureau of Transportation Statistics website. The two-digit level SCTG classes of the FAF[2] commodities are added to the extracted OD data to facilitate further analysis. This SCTG classification helps in connecting a commodity to its producing industry, which is classified using North American Industrial Classification System (NAICS).

The freight moving out of the state has different domestic destinations. For the present model, 15 destinations are selected as shown in Figure 3. These destinations make up approximately 98% of aggregated freight flowing out. It is seen that 77% of the commodities generated in North Dakota have a destination within North Dakota, as shown in Table 2. These commodities moving from and to North Dakota are modeled as internal-internal flow. The commodities moving across the state boundary to other states and Canada are modeled in the internal-external flow model.

Table 1: List of Major Commodities Moving Out of North Dakota

Commodity	Quantity in Kilo-ton	Percentage of Total	SCTG Code
Cereal grains	47,011.9	54.0%	02
Coal	51,38.1	5.9%	15
Waste/scrap	4,881.0	5.6%	41
Other ag prods.	4,477.8	5.1%	03
Gravel	4,317.6	5.0%	12
Gasoline	3,847.6	4.4%	17
Nonmetal min. prods.	3,242.4	3.7%	31
Fuel oils	2,362.5	2.7%	18
Coal, n.e.c.	1,733.2	2.0%	19
Other foodstuffs	1,638.8	1.9%	07
Unknown	1,471.0	1.7%	----
Animal feed	843.3	1.0%	04
Machinery	790.3	0.9%	34
Milled grain prods.	578.9	0.7%	06
Natural sands	562.0	0.6%	11
Mixed freight	557.0	0.6%	43
Wood prods.	547.5	0.6%	26
Live animals/fish	417.6	0.5%	01

(U.S. Department of Transportation 2007a)

Figure 3: Commodities Moving to Major Destinations from North Dakota

Table 2: Destination of Commodities Moving Out of North Dakota within United States

Destinations FAF regions	Description	Quantity in Kilo-tons	Percentage of Total
ND	North Dakota	67242.063	77.27%
MN remainder	Remainder of Minnesota	11877.248	13.65%
MN Minneapolis	Minneapolis	3097.741	3.56%
SD	South Dakota	1416.059	1.63%
WI remainder	Wisconsin	790.957	0.91%
MT	Montana	439.593	0.51%
NE	Nebraska	194.658	0.22%
IA	Iowa	120.531	0.14%
IL Chicago	Illinois- Chicago	115.738	0.13%
IL remainder	Illinois	96.479	0.11%
CA remainder	Remaining of California	85.62	0.10%
PA remainder	Remaining of Pennsylvania	71.046	0.08%
IL St Louis	St. Louis Illinois	65.136	0.07%
WI Milwaukee	Milwaukee Wisconsin	57.322	0.07%
VT	Vermont	55.306	0.06%
CA Los Angles	Los Angeles - Long Beach CA	53.763	0.06%
WY	Wyoming	48.596	0.06%

(U.S. Department of Transportation 2007a)

Employment Distribution Pattern

The statewide production data are further disaggregated to the county level using two-digit level NAICS county employment data available from the U.S. Census Bureau's *County Business Patterns* database (Vachal and Tolliver 2001) and supplemented by data available from the manufacturer's survey done at the Upper Great Plains Transportation Institute (UGPTI). CBP data are available from 1988 to 2005. Until 1997, the establishments were classified on the basis of standard industrial classification (SIC); henceforth the classification is based on the NAICS. Retail trade, wholesale trade and manufacturing are included in both of these systems, but they cover somewhat different groups of industries.

CBP has employment data for industries categorized to three-digit level NAICS. However, for most industries there is data suppression to avoid disclosure. Hence, it was decided to use two-digit level NAICS for classification of industries in the state. SCTG, which is a classification of commodities, and NAICS, which is a classification of industries, need to be linked. A bridge dataset provided by BTS serves the purpose of linking two-digit level SCTG to two- and three-digit level NAICS.

Disaggregating Trip Generation

The statewide production data are disaggregated to county level using two-digit NAICS county employment data. The basic assumptions for disaggregating state production data using county employment data are as follows:
- Manufacturing plant's output is proportional to the number of employees.

- All plants in the same industry have the same productivity.

The ratio of county employment to employment in the entire state in a particular NAICS category determines the county production of that commodity.

$$(1) \quad t_{cj}^a = t_j^a \times e_c^b / \sum e^b \; \forall \; j$$

t_{cj}^a = tons of SCTG class a commodity produced in county c moving to destination j

t_j^a = total tons of SCTG classification a commodity produced in the state moving to destination j

e_c^b = employment in industry b in county c

e^b = total employment in industry b in the state

a = two-digit level SCTG commodity class

b = two-digit level NAICS industries class producing commodity a

An example of disaggregation methodology is illustrated in Table 3. The county employment distribution for NAICS category 21, which is the industry classification for mining products, is shown in row one of the table. Employment in this category is distributed within seven counties of the state of North Dakota. The second row of the table shows the estimated ratio of the county employment in NAICS 21 to the total employment in the NAICS 21 category in the state i.e. $e_c^b / \sum e^b$. The cell in the third row of the table shows the amount in Kilo-tons that moves from North Dakota to the eight major destinations. The ratios estimated in the second row are used to disaggregate the quantities in the third row to the county production, as shown in cells of row four to ten in Table 3. The amount of commodity 41.238 Kilo-tons as shown in row 4 moving from Burke county to ND is the product of 3686.06 Kilo-tons of commodity moving from ND to ND and the county employment ratio .01, as shown in row 2. The difference of the result is because of rounding of the employment ratios in row 2.

Trip Attraction

From the FAF²'s "FAFOD_DOM_2002" Microsoft Access table, data for all commodities moving into North Dakota by truck are extracted using an SQL query. From this list, 16 major FAF regions, as shown in Table 4, are selected that generate 99% of the freight moving into the state. Commodities originating and terminating in North Dakota make up the bulk of this OD data. This implies, as we have seen, that the biggest percentage of freight generated in the state has its destination within North Dakota. Those commodities, having their origins and destinations in North Dakota, are analyzed in the internal-internal module.

From this list of commodities moving in the state, 18 commodities are selected that make up 97% of the total freight flow, as shown in Table 5. According to data in Table 5, 48% of this freight is cereal grains. This is excluded from this manufactured goods model, as it is included in the agricultural freight flow model.

Input-Output Analysis

Disaggregation of trip attraction is not as simple as disaggregating trip production data. Trip attraction is based on the characteristics of the receiving industry and end users that consume the goods moving in the state. To link the goods moving in the state to the consuming industries and end users, a supply side commodity-by-industry I-O table is used. The FAF² inbound traffic is disaggregated to the TAZ level based on the number of manufacturers and other economic sectors

Table 3: Disaggregating Trip Production Data to County Level

Row	NAICS (b)	County Employment (e_c^b)							Total Employment (e_c^b)
		Burke	Burleigh	McKenzie	Mercer	Stark	Ward	Williams	For all Counties
1	21	26	89	84	1249	274	273	329	2,324
2	$e_c^b /$ e^b	0.01	0.04	0.04	0.54	0.12	0.12	0.14	

		Commodity "a" shipped to destination "j" from North Dakota (Kilo-Tons) (t_j^a)							
		IA	MN Minne	MN rem	MT	ND	SD	WI rem	WY
3	ND	0.29	73.49	185.66	4.30	3686.06	364.32	2.58	0.93

		Commodity "a" moving to destination "j" from county "c" (Kilo-Tons) (t_{cj}^a)							
		IA	MN Minne	MN rem	MT	ND	SD	WI rem	WY
4	Burke	0.003	0.822	2.077	0.048	41.238	4.076	0.029	0.010
5	Burleigh	0.011	2.814	7.110	0.165	141.161	13.952	0.099	0.036
6	McKenzie	0.011	2.656	6.711	0.155	133.231	13.168	0.093	0.034
7	Mercer	0.156	39.494	99.779	2.311	1981.017	195.796	1.388	0.500
8	Stark	0.034	8.664	21.889	0.507	434.587	42.953	0.305	0.110
9	Ward	0.034	8.632	21.809	0.505	433.000	42.796	0.303	0.109
10	Williams	0.041	10.403	26.283	0.609	521.821	51.575	0.366	0.132

Table 4: FAF² Regions Generating Freight Moving in North Dakota

Origin	Quantity in Kilo-tons	Percentage
ND	67,242	89.66%
MN remainder	3,276	4.37%
MN Minneapolis	1,212	1.62%
IL remainder	640	0.85%
SD	600	0.80%
MT	390	0.52%
WI remainder	269	0.36%
MO remainder	250	0.33%
IA	177	0.24%
IL Chicago	118	0.16%
WY	84	0.11%
ID	54	0.07%
TN Memph	47	0.06%
MI remainder	43	0.06%
NE	40	0.05%
CA remainder	39	0.05%

(U.S. Department of Transportation 2007a)

Table 5: Major Commodities Moving into North Dakota

Commodity	Quantity in Kilo-tons	Percentage
Cereal grains	36,004	48.0%
Coal	5,162	6.9%
Gravel	5,009	6.7%
Waste/scrap	4,872	6.5%
Other agri. prods.	3,924	5.2%
Gasoline	3,630	4.8%
Nonmetal min. prods.	3,516	4.7%
Fuel oils	2,090	2.8%
Coal, n.e.c.	1,834	2.4%
Other foodstuffs	1,308	1.7%
Unknown	1,285	1.7%
Natural sands	757	1.0%
Wood prods.	720	1.0%
Mixed freight	586	0.8%
Animal feed	584	0.8%
Machinery	582	0.8%
Nonmetallic minerals	487	0.6%
Live animals/fish	375	0.5%

(U.S. Department of Transportation 2007a)

in the TAZ. An I-O table gives a comprehensive snapshot of economic activities in a state or region for a period, say a year. It portrays the flow of goods from the producers to the manufacturers and end users. In the I-O table, the row sum gives the total sales of the sector, and the column sum gives the total input purchases of the sector. The basic principle of this table is that the sum of the input to an industry is equal to its output. The table can be divided into three components: inter-industry transaction, final demand and final payments. The inter-industry block includes transaction between the state's or region's industries. The final demand is the sales from industries to the end users, and final payments are the purchases of labor and capital by the industries of the region (Lawson et al. 1997).

The national I-O table is obtained from the BEA's Commodity-by-Industry 1997 benchmark I-O account. The 1997 benchmark account is more detailed than other annual accounts. This benchmark account shows the flow of 483 commodities to 491 industries and to 13 final users. The salient features of the 1997 I-O account are the introduction of the NAICS in place of SIC. This seems to be more relevant for service industries. This I-O account is more consistent with gross domestic product (GDP) by industry accounts and gross state products of industries. This benchmark I-O account has the advantage of being downloadable at different levels of aggregation. The I-O classification system is based on the NAICS classification, but it also considers special industries and government industries. The commodities in the I-O classification system are given the code of the industry that produces that commodity. The summary I-O table is aggregated to 38 commodity and industry user-defined categories (Liu and Vilain 2004). Lawson et al. (1997), in their explanation of the benchmark input-output accounts, provides the NAICS class corresponding to the I-O account's classification. For each of the 38 industries, the three-digit level NAICS class is identified. For the commodities, it is convenient to identify the I-O group in which they fall. In

the I-O accounts table, there are some negative entries in column 38. These negative entries reflect import of the commodities from other countries. The value of imports in the producer's price is the negative value entered in the final demand column. The regional I-O table is developed from the Bureau of Economic Analysis's (BEA) benchmark I-O account and location coefficients for each level of commodity aggregation. The user-defined 38 categories and their corresponding NAICS are shown in Table 6.

Disaggregating Trip Attraction

The FAF² inbound traffic is disaggregated to the TAZs of the state (Figure 2) based on the number of manufacturers and other economic sectors in the TAZ. The assumptions for using the I-O table as stated in NCHRP 260 are as follows (Memmott 1983):

- Freight shipments are proportional to the dollar outputs contained in the table.
- All plants in the same commodity group have production in proportion to the number of employees in the plant.
- All receivers in the same industry share the resulting commodity flow proportionally.

The first step of the disaggregation methodology is to download the user-defined I-O table. In this table, there are 38 rows and 38 columns. The rows represent the amounts of commodities in (million) dollar values used by the industries and final users.

(2) $U = [U_{ij}]$

U is an m by n matrix with each element representing the amount of commodity i in dollar value used by industry j as input

(3) $A = U.\mathbf{1}$

A_i = total amount of commodity i used by industry and end user

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ . \\ 1 \\ 1 \end{bmatrix} \text{Unit Vector } (i \times 1), i \text{ is the number of commodities}$$

(4) $\beta = (diag\ A)^{-1}U$

β_{ij} = proportion of commodity i sold to industry j

A cell in the β matrix is the proportion in which the commodity in the row is consumed by the industry in the column. The sum of a row is equal to "1," and a zero value in a cell implies no consumption of the commodity in the row by the industry in that column.

The national I-O coefficients must be regionalized to the state level. Location quotients, as shown in Table 7, are used to regionalize the national I-O coefficients (Liu and Vilain 2004). A location quotient of more than one implies a greater share in the economy by that industry at the state level compared to its share at the national level.

(5) $L_{state} = \dfrac{(\% \text{ of total state employment in industry } j)}{(\% \text{ of total U.S. employment in industry } j)}$

L_{state} = matrix $(1 \times j)$ ratio of state employment in industry j to national employment in industry j

Table 6: User-defined 38 Categories and Corresponding NAICS

Category	Description	Corresponding NAICS
1	Farm Products	1110, 1120
2	Forest and fishing products	1130, 1140, 1150
3	Mining (mineral) products	2110, 2121, 2122, 2123, 2130
4	Construction	2301, 2302, 2303
5	Ordnance or accessories	332A
6	Food or kindred products	3110, 3121
7	Tobacco products, excluding insecticides	3122
8	Textile mill products	3130, 3140
9	Apparel or other finished textile products	3150
10	Lumber or wood products, excluding furniture	3210
11	Furniture or fixtures	3370
12	Pulp, paper or allied products	3221, 3222
13	Printing matter	3230
14	Chemical or allied products	3251, 3253, 3259
15	Petroleum or coal products	3240
16	Rubber or miscellaneous plastics products	3252, 3260
17	Leather or leather products	3160
18	Clay, concrete, glass or stone products	3270
19	Primary metal products	331A, 331B, 3315, 3321, 3322, 3324
20	Fabricated metal products	332B
21	Machinery, excluding electrical	3331, 3332, 3335
22	Electrical machinery, equipment or supply	3341, 334A, 3353, 3359
23	Transportation equipment	3361, 336A, 336B, 3364
24	Instruments, photographic goods, optical good	3333
25	Miscellaneous products or manufacturing	3391, 3399
26	Railroad transportation	4820
27	Trucking and warehousing	4840, 4930
28	Water transportation	4830
29	Other transportation	4810, 4850, 48A0, 4920
30	Communications	5111, 5112, 5120, 5131, 5132, 5133, 5141
31	Electric, gas, and sanitary services	5620, 2211, 2212, 2213
32	Wholesale trade	4200
33	Retail trade	4A00
34	Finance, insurance, and real estate	52A0, 5230, 5240, 5250, 5310, 5321, 5324
35	Services	5411, 5412, 5414, 5415
36	Government and government enterprises	S001, S002, S005
37	Others	S003, S004, S006, S007
38	Value added / final demand	

Table 7: State and National Location Quotient for Different NAICS Class

NAICS	ND Employment	US Employment	L_j
11	242	182,121	0.62
21	2,324	470,280	2.34
22	2,201	634,734	1.64
23	13,760	6,647,641	0.98
31	20,112	13,821,976	0.68
32	5,684	3,980,178	0.67
33	15,256	7,485,582	0.96
42	15,887	5,907,051	1.27
44	41,813	15,351,431	1.29
48	8,342	4,098,870	0.96
51	6,977	3,472,427	0.95
52	13,299	6,481,304	0.97
53	3,285	2,086,085	0.74
54	10,790	7,569,981	0.67
55	2,774	2,824,787	0.46
56	10,118	8,708,052	0.55
61	2,979	2,893,346	0.48
62	49,571	15,814,812	1.48
71	2,563	1,889,044	0.64
72	25,727	10,749,811	1.13
81	13,187	5,416,193	1.15

(U.S. Census Bureau 2004a)

(6) $\beta_{state} = \beta.(diagL_{state})$

β_{state} = the supply side commodity-by-industry coefficient for the state

β_{state} is more than β for those commodities which have L_{state} values more than one, and β_{state} is less than β for those commodities that have location quotients less than one. This is justified, as the industries that are predominant in the state as compared to its predominance in the nation will have higher proportion of consumption of commodities than at the national level. This is true under certain assumptions like homogeneity of goods, spatial uniformity of demand and none of the good is exported. The rationalization process changes the row sum of β_{state}, so an adjustment is necessary to make the row sum equal to one.

(7) $\gamma = \beta_{state} . \mathbf{1}$

(8) $C = diag\{\gamma\}^{-1} \beta_{state}$

C = adjusted commodity by industry supply side I-O model for the state

(9) $D = C.C_e$

C_e = employment matrix

D = portion of commodity moving to the counties

(10) $D_{ik} = \Sigma_j C_{ij}C_{jk}$ (Co-ordinate form)

D_{ik} = portion of commodity i moving to county k
C_{ij} = adjusted commodity i by industry j supply side I-O coefficient for the state
C_{jk} = employment/population in industry/end-user j in county k

(11) $\hat{D} = diag\{D.1\}^{-1} D$

$\hat{D}$ = adjusted D matrix making the row sum equal to one.

Each element of the $\hat{D}$ matrix represents the portion of commodity i shipped to county j. This is the final matrix, which helps to disaggregate the commodities moving into the state to the county level based on the employment and population patterns of the state.

(12) $B_{ijk} = F_{li}. \hat{D}_{ik}$

F_{li} is the amount of commodity i moving into the state from origin l

i = commodity type, k = county and l = origin

B_{ijk} is the OD matrix for different commodities ranging from 1 to i. The origin l is outside state FAF region, and the destination k is a county in the state.

Internal-Internal Flow

The internal-internal flow data for North Dakota are available from the FAF[2] databases. These sets of data have their origins and destinations in North Dakota. The trip production data for this internal-internal dataset are disaggregated to the TAZ level, using the NAICS employment distribution. The freight attraction data are also disaggregated to the TAZ level, using the I-O table and the county employment pattern, as explained in detail in the previous section. Unlike the internal-external and the external-internal data, the internal-internal data have both their origin and destination disaggregated; hence, an additional step of trip distribution is required, to develop an OD matrix.

The internal-internal flows are distributed using a gravity model. Data for trip length distribution of manufactured goods in the state is available from a survey done at UGPTI (Vachal and Tolliver 2001). This trip length distribution data is used to calibrate the gravity model.

(13) $X_{ij}^c = \dfrac{P_i^c A_j^c F_{ij}}{\Sigma_j^n A_j^c F_{ij}} \; \forall \; c$

Where:

X_{ij}^c = flow of manufactured freight from TAZ i to TAZ j

P_i^c = freight c production in TAZ i

A_j^c = attraction of freight c at TAZ j

c   = waste/scrap, gravel, gasoline, nonmetal mineral products, fuel oils, coal, other foodstuffs, unknown, animal feed, machinery, natural sands, mixed freight, wood products and live animals/fish

(14) $F_{ij} = f(t_{ij}) = F_{ij} = ae^{-bt_{ij}}$; *where* t_{ij} = travel impedance based on distance, b = calibration factor, a = coefficient

The gamma function for the impedance factor used is as follows:

(15) $F_{ij} = ae^{-bt_{ij}}$

t_{ij} = travel impedance, a and b are coefficients

The log form of the function is used in a regression analysis to estimate the coefficients.

(16) $ln(F_{ij}) = ln(a) - bt_{ij}$

Coefficients *a* and *b* with values 12 and .025 give the best fit. The observed and model trip distributions are compared; the mean trip length for observed flow is 89 miles and estimated flow is 84 miles, which a is good fit. The major truck routes obtained from the internal-internal model are shown in Figure 4.

Figure 4: Major Truck Routes for the Internal-Internal Commodity Flow
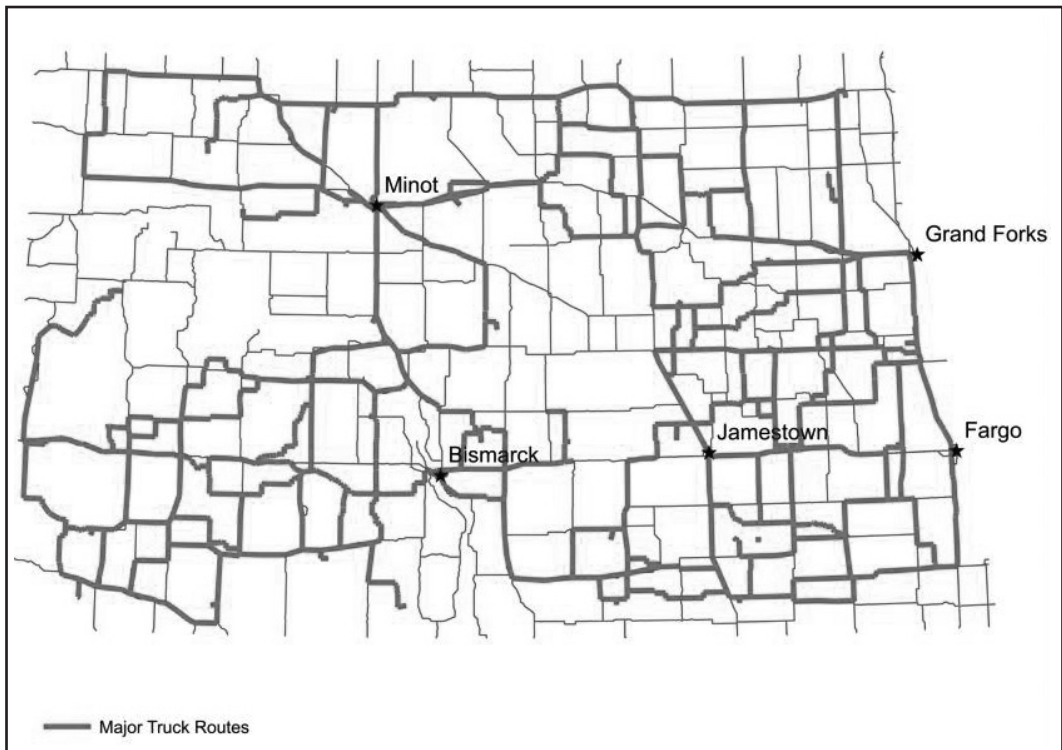
Analyzing Cross-Border Traffic

The FAF²'s Microsoft Access database "FAFOD_2002.MDB" has a table named "FAFOD_BRD_2002" that has data on cross border freight movement to neighboring countries, namely Canada and Mexico. An SQL query is done to extract freight movement between Canada and North Dakota. From the list of commodities moving to and from Canada, eight commodities and two ports of exit, as shown in Table 8, are selected. The two ports of exit are North Dakota and Detroit, Michigan. To model this flow, highways leaving the state at the northern state boundary are connected to provinces in Canada with dummy links, and for the Detroit port of exit it is assumed that the truck will travel via eastern exits of the state and then move to Canada through Detroit. This cross-border freight OD data is disaggregated to the TAZ level using county employment data, as has been done for domestic freight moving out of the state. The commodities moving in are disaggregated to the TAZs using the I-O table and the employment distribution pattern.

Table 8: Major Commodities Moving to Canada and the Port of Exit

Commodity	Port	Kilo-ton
Animal feed	ND	48.23
Animal feed	MI Detroit	10.2
Cereal grains	MI Detroit	14.31
Cereal grains	ND	67.69
Coal	ND	44.65
Coal	MI Detroit	9.44
Fertilizers	MI Detroit	3.63
Fertilizers	ND	17.16
Machinery	ND	11.31
Machinery	MI Detroit	2.39
Motorized vehicles	MI Detroit	1.57
Motorized vehicles	ND	7.42
Other ag prods.	MI Detroit	21.51
Other ag prods.	ND	101.77
Other foodstuffs	MI Detroit	1.86
Other foodstuffs	ND	8.78

(U.S. Department of Transportation 2007a)

Model Validation

The statewide freight model is validated using truck count data from different sources. The truck count data for the state of North Dakota are available from cordon surveys, Highway Performance Monitoring System (HPMS), vehicle miles traveled (VMT) estimates and Roadway Information Management System (RIMS). In this project, the truck traffic assigned to the network from the agricultural goods flow model is merged with the manufactured goods flow model. There is another set of truck count data available from the North Dakota DOT's automatic traffic recording (ATR) stations. There are 38 permanent ATR stations in the state. These stations have directional truck count data for every single day of the year. For this project, the ATR station's annual truck traffic is used for validation and calibration of the estimated truck traffic.

Statistical methods used to compare link traffic counts are absolute difference in volumes, percentage differences in volumes, average errors, average percent errors, standard deviations, R-square values, root mean square errors and correlation coefficients. Table 9 gives the observed and estimated link truck traffic. O_l is the observed traffic and E_l is the estimated traffic. The percent RMSE is calculated using equation 17, and results are shown in Table 10.

$$(17) \quad \%RMSE = \frac{\left(\sum_j \left(Model_j - Count_j\right)^2 / \left(Number\ of\ Counts - 1\right)\right)^{.5} \times 100}{\left(\sum_j Count_j / Number\ of\ Counts\right)}$$

Table 9: Absolute Percentage Error of Estimated Truck Traffic

Name of highway	O_l	E_l	$O_l - E_l$	$\left\|\dfrac{O_l - E_l}{O_l}\right\|(\%)$
S200	32,756	32,877	-121	0.4%
S200	40,243	40,424	-181	0.5%
S200	14,529	14,211	318	2.2%
S200	14,306	14,372	-66	0.5%
U2	52,985	51,791	1,194	2.3%
U2	52,373	52,061	312	0.6%
U83	95,051	96,142	-1,091	1.1%
U83	84,270	86,343	-2,073	2.5%
S13	24,071	23,913	158	0.7%
S13	28,225	28,023	202	0.7%
I94	278,715	282,127	-3,412	1.2%
I94	263,952	265,950	-1,998	0.8%
I29	437,825	404,157	33,668	7.7%
I29	439,472	432,766	6,706	1.5%
U83	94,963	96,090	-1,127	1.2%
U52	100,815	101,751	-936	0.9%
I94	758,577	742,380	16,197	2.1%
I94	691,500	675,439	16,061	2.3%
I29	244,162	245,567	-1,405	0.6%
I29	259,734	264,967	-5,233	2.0%
S1	20,063	20,072	-9	0.0%
S1	20,221	20,292	-71	0.3%

Table 10: Percent Root Mean Square

Truck counts	% RMSE
< 50,000	1.47
50,000 – 100,000	1.36
100,000 - 150,000	.47
150,000-200,000	9.04
200,000 - 250,000	8.99
250,000 – 300,000	1.42
300,000 - 350,000	.86
> 350,000	3.30

CONCLUSION

This research process, along with its findings, contributes to building a framework for integrating statewide agricultural freight flow models with that of manufactured goods flow. The micro-level disaggregation of the total flows into sub-modules helps to analyze the effect of external factors on freight travel demand in greater detail. In statewide freight modeling, availability of data is a big concern for most states. In this research, most of the freight data are generated from publicly available databases. One of the prime sources of the agricultural freight model is satellite imagery, which is freely available from the National Agricultural Statistics Service (NASS). This data can be used not only for statewide study, but also for more detailed study of smaller regions at county or sub-county levels.

FAF[2] data that fill the void of the CFS data are utilized for this statewide freight modeling purpose. This modeling framework, with the successful outcome, will help build confidence in using publicly available data in the absence of proprietary data. Implan (Minnesota Implan Group 2007) is used by many states for disaggregation of trips to the TAZ level. In this project, it demonstrated how, in the absence of this software, data from heterogeneous sources like Bureau of Transportation Statistic's I-O account data, county employment data and local survey results can be used to disaggregate trips to the TAZ level.

In freight demand modeling, the base origin/destination matrix does not always represent the actual freight movement, as there can be many intermediate legs in the OD flow. There can be a number of warehouses and distribution centers between the origin and final destinations. Most of the available data do include these shipment legs. In this project, a matrix estimation method is used to introduce these distribution legs by updating the estimated OD matrix with the available link count data. All the innovative methods introduced in this statewide freight model are a stride toward the development of robust statewide freight modeling methodologies.

Endnotes

1. Henceforth referred to as "manufactured goods/freight" for convenience.

2. FAF data is used in this project to generate OD data, and it is disaggregated and assigned using the models described in the paper. The FAF freight incorrectly assigned to highway segments in some states is not used in this project.

References

Bureau of Economic Analysis 2005. *Benchmark Input-Output Accounts, The Use of Commodities by Industries before Redefinitions*. http://www.bea.gov/bea/dn2/home/benchmark.htm, Accessed March 9, 2006.

Bureau of Transportation Statistics. *Directory of Transportation Data Sources*. U.S. Department of Transportation, Washington, D.C., 1995.

Bureau of Transportation Statistics. *Freight Data and Statistics*. U.S. Department of Transportation, Washington, D.C., 2007.

De Jong, G. and M.E. Ben-Akiva. "A Micro-simulation Model of Shipment Size and Transportation Chain Choice." *Transportation Research Part B* 41, (2007): 950-965.

Donnelly, R. *National Freight Data Program.* Transportation Research Board Special Report 276, Transportation National Research Council, Washington, D.C., 2003.

Global Insight. *Transearch*. http://www.globalinsight.com/ProductsServices/ProductDetail 2322. htm, Accessed March 9, 2006.

Lawson, A.M., K.S. Bersani, M.F. Nader, and J. Guo. *Benchmark Input-Output Accounts of the United States*. Bureau of Economic Analysis, Online articles and other documents, 1997.

Liu, L.N. and P. Vilain. "Estimating Commodity Inflows to a Substate Region Using Input-Output Data." *Journal of Transportation and Statistics* 7(1), (2004): 23-38.

Memmott, W.F. *Application of Statewide Freight Demand Forecasting Techniques*. NCHRP Report 260: Transportation Research Board, National Research Council, Washington, D.C., 1983.

Minnesota Implan Group. *Implan.* http://implan.com, Accessed January 7, 2007.

Mitra, S., D. Tolliver, A. Varma, and A. Dybing. "Analyzing the Effects of Spring Highway Load Restrictions on North Dakota's Agricultural Freight Flows." *Transportation Research Record 2008* (2007): 92-99.

Roger Creighton Associates, Inc. and R.L. Banks & Associates, Inc. *NCHRP Research Report 177: Freight Data Requirements for Statewide Transportation Systems Planning.* Transportation Research Board, National Research Council, Washington, D.C., 1977.

Surface Transportation Board. *Economic Data: Waybill.* http://www.census.gov/epcd/ cbp/ download/dwncbp04.html, Accessed January 7, 2007.

U.S. Army Corps of Engineers. *Waterborne Commerce Statistics Center*. http://www.iwr.usace. army.mil/ndc/wcsc/wcsc.htm, Accessed January 8, 2007.

U.S. Census Bureau. *County Business Patterns: 2004.* http://www.census.gov/epcd/cbp /download/ dwncbp04.html, Accessed January 7, 2007a.

U.S. Census Bureau. *2002, Commodity Flow Survey.* http://www.census.gov/svsd/www/ cfsdat/2002cfs.html, Accessed January 8, 2007b.

U.S. Department of Transportation. Federal Highway Administration. *Freight Analysis Framework*. http://ops.fhwa.dot.gov/freight/freightanalysis/faf.index.htm, Accessed January 4, 2007a.

U.S. Department of Transportation. Federal Highway Administration. *Freight Model Improvement Program.* http://www.freight.dot.gov/fmip, Accessed January 7, 2007b.

U.S. Department of Transportation. Federal Highway Administration. *Geofreight Intermodal Freight Display Tool.* http://www.fhwa.dot.gov/freightplanning /geofreight.htm, Accessed January 8, 2007c.

U.S. Department of Transportation. Federal Highway Administration. *Maritime Statistics.* http://www.marad.dot.gov/library_landing_page/data_and_tatistics/Data _and_ Statistics.htm, Accessed January 9, 2007d.

Vachal, K. and D. Tolliver. *Regional Elevator Survey: Grain Transportation and Industry Trends for Great Plains Elevators*. Upper Great Plains Transportation Institute, Fargo, ND, Report No. DP-143. 2001.

Subhro Mitra *is an associate research fellow at the Upper Great Plains Transportation Institute at North Dakota State University. He is also a faculty member in NDSU's Transportation Logistics Ph.D. program and an adjunct professor in the civil engineering department. He holds a Ph.D. in transportation and logistics from NDSU and has 14 years of diverse work experience in both operational and technical aspects of transportation and highway engineering. He is a registered professional engineer of the State of North Dakota.*

Denver Tolliver *is associate director of the Upper Great Plains Transportation Institute and director of the Transportation & Logistics program at North Dakota State University. He is also director of the Mountain-Plains Consortium (MPC)—which is the regional university transportation center for federal region 8. Tolliver is a long-time member of TRF and a former president of the Agricultural Transportation chapter. His primary research interests are highway, railroad, and inland waterway planning. Tolliver holds doctoral and master degrees from Virginia Tech and a baccalaureate degree from Morehead State. Before joining the faculty of North Dakota State University, he was a transportation planner for the North Dakota Department of Transportation.*