Commodity flow estimation models incorporating land use characteristics and economic factors

T. G. Jin[1], M. Saito[1] & D. L. Eggett[2]
*[1]Civil and Environmental Engineering Department,
Brigham Young University, USA*
*[2]Center for Statistical Consultation and Collaborative Research,
Brigham Young University, USA*

Abstract

Reliably forecasting freight demand has become essential in freight transportation management in order to systematically plan for future transportation facility needs. Land use is one of the key factors that affect commodity flow. Other factors include location factors, physical factors, operational factors, dynamic factors such as seasonal variations in demand and changes in customers' tastes, and pricing factors. Modeling commodity flows with land use is one possible first step; however, it has some limitations because land use data lacks detailed information on economic activities in particular land use. Hence, models that use land use characteristics instead of aggregate land use type data were developed in this study using the commodity flow data available from the 2002 Commodity Flow Survey and other data available to the public via the Internet. These models were used to estimate commodity flow in 2007 and their estimates were compared with commodity flow values found in the 2007 Commodity Flow Survey. Results of this comparison showed that the models could produce good estimates of commodity flows and are of practical use because the model's input data are available free via the Internet.
Keywords: commodity flow modeling, land use characteristics, economic factors.

1 Introduction

Commodity flow has become a prominent consideration as freight movement has rapidly increased and affected various aspects of state transportation systems. Especially, the increase in truck traffic on highways produces many problems and challenges to surface transportation planning, design, operation, and management. Hence, developing a reliable freight demand forecasting model has become an important task for state transportation agencies.

The most significant hurdle to actively including freight demand modeling in freight transportation planning is that most of the existing demand forecasting models were developed for passenger trips and do not reflect the characteristics of commodity flow. Ogden reported that these existing models usually assume that freight trips follow a behavioral mechanism similar to passenger trips—that is, truck traffic is estimated as a function of passenger-car traffic—and suggested two major freight modeling platforms: commodity-based modeling and trip-based modeling However, in order to overcome the weakness of using a typical four-step demand forecasting modeling process and more accurately capture the fundamental economic mechanisms that drive freight movements, which are largely determined by the attributes of the cargo being carried, the concept of commodity flow modeling was recommended for developing a freight demand model [1].

Characteristics of commodity flow have been determined by contributing factors including location factors, physical factors, operational factors, dynamic factors including seasonal variations in demand and changes in customers' tastes, and pricing factors [2]. Among these factors, location factors are considered the most important factors for estimating production or attraction of commodity flow. The characteristics of location factors can be determined by land use type. Iding et al. [3] developed sample indicators of freight trip generation by land use type in Germany through a large-scale postal and telephone survey.

Direct modeling of commodity flow using land use type performed by the authors showed close correlation between these two variables [4]; however, the application of such models still has some limitations because land use alone cannot capture the effect of the characteristics associated with land use on commodity flow. Also, land use data are not collected as often as are land use characteristics. Hence, development of a new model is recommended, which would help transportation planners estimate commodity flow using selected characteristics of land use, for which data are updated and available more frequently than land use data itself.

In order to fill the modeling gap that currently exists in freight demand modeling, a study has been conducted with three objectives in mind: (1) Build a state-wide commodity-based demand model with various factors describing land use characteristics as independent variables, using the 2002 commodity flow survey (CFS) data; (2) Build a commodity-based demand model using employment data, which is the current popular model used by several states, using the 2002 CFS data; and (3) Compare the accuracy of these two models using the CFS 2007 data.

This paper presents a brief summary of existing freight demand forecasting models, results of the development of two commodity-flow-based freight demand forecasting models, and results of a comparison of commodity flow estimates by the two models with the commodity flows reported in the CFS 2007.

2 Characteristics of existing state-wide freight demand models

An extensive literature review of existing state-wide freight demand models was made to categorize the concepts used for developing those models. In this section, existing commodity flow models are briefly discussed.

There are currently two prototype national commodity flow models: the Freight Analysis Framework (FAF) [5] and the Quick Response Freight Manual (QRFM) [6, 7]. FAF helps the user to forecast commodity flow on various interstate highway corridors. FAF integrates data from a variety of sources to estimate commodity flows and related freight transportation activities among states, regions, and major international gateways. FAF version 2 (FAF 2.2) [8] provides estimates of commodity flows for years 2002 and 2006, with forecasts for years 2010 through 2035. The QRFM developed by Cambridge Systems Inc. provides a simple method of forecasting urban truck trips. The Federal Highway Administration (FHWA) has an updated QRFM to allow simple, low-cost, and relatively accurate truck traffic forecasting at a regional or local level.

A variety of freight commodity models have been developed in order to cover a single state or multiple states, or an urban area or a region. The existing freight flow models deal with the movement of commodity transported by single modes (truck, rail, water, air, pipeline, and so on) or multiple modes (truck and rail, truck and water, rail and water, and so on) or the combined modes. Freight demand models are categorized by the type of demand forecasting methodology, such as commodity-based models, vehicle-trip-based models, origin-destination (O-D) matrix-based estimation models, and so on [9–19].

The models summarized in this section involve complex models that are the fruit of efforts of many researchers and practitioners. These freight demand models, however, often require involvement of highly skilled professionals to implement. When funds and personnel are limited, it is desirable to have a simpler method to get rough estimates of commodity flow using publicly available data.

3 Methodology

In order to build commodity flow estimation models using land use characteristics and economic factors, it was necessary to build a database. This section presents the procedure followed to collect and analyze necessary data, which are publicly available via the Internet.

3.1 Collection and reduction of freight, land use, and employment data

Data necessary for this research were collected from commodity- or freight-related websites open to public, including the Freight Analysis Framework (FAF) [5], the Commodity Freight Survey (CFS) [21], Utah GIS Portal [22], and the official website for the State of Utah [23]. CFS-related websites have provided results of a commodity flow survey, including shipment characteristics by mode of transportation for origin state in year 2002. The year 2002 County Business Pattern (CBP) data were drawn from the FAF website [5]. The Utah GIS Portal provided geographic map files and a geodetic database organized by county. The year 2002 population and social and economic characteristics of Utah were obtained from the official website of the State of Utah. All raw data were transformed by natural log in order to normalize data distributions for statistical analyses. State-wide commodity flow totals produced from (production), attracted to (attraction), and distributed within Utah (within) by two-digit standard industrial classification (SIC) code were allocated to each county using the proportion of the county total employment to the state total employment because CFS data are state-level summaries. Previous research studies found that employment data were closely related to the amount of commodity flow generated [3, 13].

3.2 Classification of data by industry code

Commodity flows were classified into 44 industries by the SIC code. Commodity flow data obtained from the CFS website were classified into three groups: 2002 commodity produced from Utah (CFS2P), 2002 commodity attracted to Utah (CFS2A), and 2002 commodity distributed within Utah (CFS2I). Here, the term *production* means commodities produced in Utah and sent to other states as well as Utah itself, *attraction* means commodities transported to Utah from other states as well as Utah itself, and *within* means a sum total of commodities produced in one county in Utah and transported to other counties in Utah and attracted to the county from the other counties in Utah. Table 1 shows a list of commodities classified by the SIC code; their variable names used in the analysis were $CFS2P_i$, $CFS2Ai$, and $CFS2Ii$ where i is an SIC code.

Employment data by area and by industry (EMPBI) were classified into 26 types by the North American Industry Classification System (NAICS) code. Table 2 shows a list of EMPBI classifications by the NAICS code; their variable name was $EMPBI_j$ where j is an NAICS code.

The year 2002 county business pattern (CBP) data were divided into paid employees for paid period including March 12 (CBPE), first-quarter payroll (CBPFQP), annual payroll (CBPAP), and total establishments (CBPTE). CBP data were classified into 23 types by the NAICS code. Table 3 shows a list of CBP classifications by the NAICS code; their variable names were $CBPE_k$, $CBPFQP_k$, $CBPAP_k$, and $CBPTE_k$ where k is an NAICS code.

Number of jobs (NOJ) and wage (W) data were classified into 12 types by the NAICS code. Table 4 shows a list of classifications of NOJ and W by the

Table 1:   Commodity classifications.

SIC Code	Detailed Industries	SIC Code	Detailed Industries
	Total (Code T was used)	23	Chemical products and preparations, nec*
1	Live animals and live fish	24	Plastics and rubber
2	Cereal grains	25	Logs and other wood in the rough
3	Other agricultural products	26	Wood products
4	Animal feed and products of animal origin, nec*	27	Pulp, newsprint, paper, and paperboard
5	Meat, fish, seafood, and their preparations	28	Paper or paperboard articles
6	Milled grain products and preparations, and bakery products	29	Printed products
7	Other prepared foodstuffs and fats and oils	30	Textiles, leather, and articles of textiles or leather
8	Alcoholic beverages	31	Non-metallic mineral products
9	Tobacco products	32	Base metal in primary or semi-finished forms and in finished basic shapes
10	Monumental or building stone	33	Articles of base metal
11	Natural sands	34	Machinery
12	Gravel and crushed stone	35	Electronic and other electrical equipment and components and office equipment
13	Non-metallic minerals, nec*	36	Motorized and other vehicles (including parts)
14	Metallic ores and concentrates	37	Transportation equipment, nec*
15	Coal	38	Precision instruments and apparatus
17	Gasoline and aviation turbine fuel	39	Furniture, mattresses and mattress supports, lamps, lighting fittings
18	Fuel oils	40	Miscellaneous manufactured products
19	Coal and petroleum products, nec*	41	Waste and scrap
20	Basic chemicals	43	Mixed freight
21	Pharmaceutical products	——	Commodity unknown (SIC code 99 was assigned for analysis)
22	Fertilizers		* nec = not elsewhere classified

NAICS code; their variable names were NOJ_l and W_m, where l and m are NAICS codes.

Table 2:   Employment data by area and by industry (EMPBI) classifications.

NAICS Code	Detailed Industry	NAICS Code	Detailed Industry
	Total (Code T was used)	13	Management of companies
1	Natural resources	14	Administrative and waste services
2	Mining	15	Educational services
3	Utilities	16	Health and social services
4	Construction	17	Arts, entertainment, and recreation
5	Manufacturing	18	Accommodation and food services
6	Wholesale trade	19	Other services
7	Retail trade	20	State and local government
8	Transportation and warehousing	21	Federal civilian
9	Information	22	Federal military
10	Finance and insurance	23	Farm
11	Real estate, rental and leasing	24	State government
12	Professional and technical services	25	Local government

Table 3:   County business pattern (CBP) classifications.

NAICS Code	Detailed Industry	NAICS Code	Detailed Industry
	Total (Code T was used)	53	Real estate & rental & leasing
11	Forestry, fishing, hunting, and agriculture support	54	Professional, scientific, and technical services
21	Mining	55	Management of companies and enterprises
22	Utilities	56	Admin., support, waste mgt., remediation services
23	Construction	61	Educational services
31	Manufacturing	62	Health care and social assistance
42	Wholesale trade	71	Arts, entertainment, and recreation
44	Retail trade	72	Accommodation and food services
48	Transportation and warehousing	81	Other services (except public administration)
51	Information	95	Auxiliaries (except corporate, subsidiary, and regional mgt.)
52	Finance and insurance	99	Unclassified establishments

Table 4:   Number of jobs (NOJ) and wages (W) classifications.

NAICS Code	Detailed Industry	NAICS Code	Detailed Industry
	Total (Code T was used)	6	Financial activity
1	Natural resources and mining	7	Professional and business services
2	Construction	8	Education and health services
3	Manufacturing	9	Leisure and hospitality
4	TTU	10	Other services
5	Information	11	Government

3.3 Data sorting and transportation

Ramsey and Schafer state that if the ratio of the largest to the smallest measurement in a group is greater than 10, the data are probably more conveniently expressed on the log scale [24]. Also, if the graphical display of raw data shows a skewed distribution and if the data with a larger average also has a larger spread, log transformation is likely to be a good choice to normalize the data for subsequent analysis [24]. Distributions of the collected raw data did show the trends described above; hence, all raw data were transformed by natural log (*ln*) in order to normalize data distribution.

3.4 Stepwise regression analysis

Stepwise multiple regression analysis was performed by SPSS on transformed data for all contributing factors describing land use characteristics. Two models were developed. Equation (1) and Equation (2) show the general format of multiple regression models of the two models. CFS2P in the model means year 2002 commodity flow values for production. For attraction and within, commodity values P is replaced by A and I respectively.

$$\text{Model 1: } \ln (CFS2P_i) = Constant_i + \sum_{j=1}^{25} a_{ij}\ln (EMPBI_j) \tag{1}$$

$$\text{Model 2: } \ln (CFS2P_i) = Constant_i + \sum_{j=1}^{25} a_{ij}\ln (EMPBI_j) +$$
$$\sum_{k=1}^{21} b_{ik}\ln (CBPTE_k) + \sum_{l=1}^{11} c_{il}\ln (NOJ_l) + \sum_{m=1}^{11} d_{im}\ln (W_m) \tag{2}$$

where, i = commodity classification number (see Table 1),
j = employment classification number (see Table 2),
k = business pattern classification number (see Table 3),
l = industry classification number for number of jobs see Table 4), and
m = industry classification number for wage (see Table 4).

As independent variables of Model 1, 25 classifications of EMPBI by the NAICS code (see Table 2) were entered. Employment-related factors have been

used by existing state-wide freight demand forecasting models. Commodity flows by the SIC code became dependent variables.

In Model 2, 68 contributing factors were used as independent variables, including 25 EMPBI variables, 21 CBPTE variables, 11 NOJ variables, and 11 W variables. Out of the four CBP independent variables, CBPE, CBPAP, and CBPFQP variables were excluded because they had missing data.

As a method of choosing the best model, Schwartz's Bayesian Information Criterion (BIC), which is a numerical test for the fit of a model, was used because R^2 values were not suitable for this task [24]. The smaller the BIC value, the better the model.

4 Results of analysis

This section presents the coefficients of the two multiple regression models, a comparison of the 2007 commodity flows estimated by the two models, a comparison of the 2007 commodity flows estimated by the two commodity-based models and the CFS 2007 by BTS and U.S. Census Bureau, and discussions on the accuracy of the two commodity-based models.

4.1 Variables selected for the two multiple regression models

In Model 1, five to six independent variables among the 25 employment-related variables were selected, including EMPBI7, EMPBI19, EMPBI18, EMPBI20, EMPBI23, EMPBI25, EMPBI5, and EMPBI3 (see Table 2).

In Model 2, six to ten independent variables among the 68 variables were selected, including EMPBI7, EMPBI19, EMPBI18, EMPBI23, EMPBI3, EMPBI15, CBPTE 21, CBPTE22, CBPTE31, CBPTE95, NOJ9, W8, and W11, (see Table 1 through Table 4).

4.2 Comparison of commodity flows estimated by the models and the CFS2007

In order to validate the two models developed in this study, the commodity flows estimated by the two models were compared with the commodity flows in the CFS 2007. Due to space limitation, only total commodity values are compared in this paper. The total commodity values presented in this section are a sum total of individual commodity values estimated by the two models. Table 5 shows a side-by-side comparison of the total commodity flows estimated by the two models and the CFS 2007. In general the commodity flows estimated by the two models were lower than the values found in the CFS 2007; in other words, the models tend to underestimate. The total commodity flow for Utah in the CFS 2007 was 305,414 thousand tons, with production of 123,245 thousand tons, attraction of 104,616 thousand tons, and within Utah of 77,553 thousand tons. The estimates by Model 1 were 83% to 84% of the estimates in the CFS 2007, while the estimates by Model 2 were 86% to 88% of the estimates in the CFS 2007.

Table 5:   Comparison of total commodity flows, in thousand tons.

2007CFS	The Models		CFS 2007
	Model 1	Model 2	
Production	102,939	108,556	123,245
Attraction	87,066	91,072	104,616
Within	64,089	67,036	77,553
Total	254,094	266,664	305,414

Table 6:   Comparison of commodity flows by county, in thousand tons.

County Name	The Models		CFS 2007	County Name	The Models		CFS 2007
	Model 1	Model 2			Model 1	Model 2	
Beaver	571	517	617	Piute	84	96	132
Box Elder	4,412	4,420	5,240	Rich	208	196	252
Cache	11,378	10,452	12,244	Salt Lake	104,366	117,517	138,294
Carbon	2,152	2,171	2,491	San Juan	929	955	1,102
Daggett	81	103	131	Sanpete	1,906	1,792	1,978
Davis	23,690	25,820	28,422	Sevier	2,026	1,999	2,078
Duchesne	1,614	1,779	1,917	Summit	5,691	4,832	6,143
Emery	976	950	1,289	Tooele	3,597	3,716	4,029
Garfield	565	563	651	Uintah	2,936	2,881	3,601
Grand	1,036	1,106	1,233	Utah	42,112	42,096	46,473
Iron	4,143	4,318	4,448	Wasatch	1,789	1,730	1,980
Juab	813	768	961	Washington	12,718	12,732	13,974
Kane	1,043	821	841	Wayne	214	283	319
Millard	1,319	1,350	1,316	Weber	21,203	20,100	22,561
Morgan	519	600	694	**Total**	**254,094**	**266,664**	**305,414**

Table 6 shows a comparison of commodity flows by county, by the two models, and by the CFS 2007. The average difference expressed by the ratio between the commodity flows by county, by the two models, and by the CFS 2007 were 0.86 and 0.87 of the commodity flows in the CFS 2007. The standard deviation of difference in the ratio between the commodity flows by county, by the two models, and by the CFS 2007 was 0.12 for Model 1 and 0.07 for Model 2, indicating that Model 2 results in less variation in the estimated commodity flow values and is therefore considered more reliable than Model 1.

4.3 Evaluation of the two models

The comparison of commodity flows estimated by the two models and by the CFS 2007 showed that Model 2, which shows the relationship between commodity flow and major contributing factors describing land use characteristics, is better than Model 1, which shows the relationship between commodity flow and the employment related factors. Model 2 is more reliable than Model 1 because Model 2 has a narrower range of standard deviations and a higher ratio between the commodity flows by Model 2 and by the CFS 2007. Even though Model 2 requires more input data, extra work for using Model 2 will not be overwhelming because all data of these variables are publicly available via the Internet. Hence, Model 2 is considered superior overall to Model 1.

5 Conclusions

In this study, two commodity-based freight demand models were developed for the State of Utah, using various factors describing land use characteristics. Model 1 is a commodity-based model relating commodity flow to employment-related factors only. Model 2 is a commodity-based model using factors describing land use characteristics, including employment, business pattern factors, job availability, and wage related factors. The commodity flows estimated by the two models were compared with the 2007 CFS values. It was found that overall the models tend to underestimate the commodity flow when compared to the 2007 CFS values and that between the two, Model 2 performed better than Model 1.

This model development effort demonstrated that commodity flows can be estimated by using multiple regression models for practical purposes. The data needed for the independent variables are available to the public via the Internet, thus making the task of commodity flow estimation practical and economical.

In order to accurately reflect commodity flow, it is necessary to use complex methods that can describe the interactions of factors affecting commodity flow; however, when funds and personnel are limited, simpler macro-level models like those developed in this study will be useful.

References

[1] Ogden, K. W. "Urban goods movement: A guide to policy and planning." Ashgate, Aldershot, Hampshire, United Kingdom, 1992.
[2] Ortuzar, J. de D. & Willumsen, L. G. "Modeling transport - 3rd ed.", Wiley & Sons, New York, 2002.
[3] Iding M. H.E., Meester, W. J., & Tavasszy, L. A. "Freight trip generation by firms." *Paper for the 42nd European Congress of the Regional Science Association*, Dortmund, Germany, August 27-31, 2002.
[4] Jin, T. G., Saito, M., Eggett, D. L., & Nelson, J. "Finding relationships between commodity flow and land use." Submitted for presentation and

publication in the XXIVth World Road Congress, Mexico City, September 26-30, 2011.

[5] Federal Highway Administration. FAF website, http://ops.fhwa.dot.gov/freight/freight_analysis/faf/. Accessed in October 12, 2009.

[6] Cohen, H. S., Alfelor, R. M., Rhoades, K. L., Horowitz, A. J., Chatterjee, S., McAdams, M., Matherly, D.W., & Sosslau, A. "Quick response freight manual." U.S. Department of Transportation, Washington, DC, 1996.

[7] Cambridge Systematics, Inc. "Quick response freight manual." Report No. DTFH61-93-C-00075, U.S. Department of Transportation, Washington, DC, 1996.

[8] U.S. Census Bureau & Bureau of Transportation Statistics. "Freight analysis framework (FAF) Version 2.2, User Guide, Commodity Origin-Destination Database 2002-2035." Washington, DC, 2006.

[9] Al-Deek, H., Klodzinski, J., El-Helw, A., Sarareddy, P., & Emam, E. "Development of a statewide model for heavy truck freight movement on external networks connecting with Florida ports." Florida Department of Transportation, Tallahassee, FL, 2002.

[10] Proussaloglou, K., Popuri, Y., Tempesta, D., Kasturirangan, K., & Cipra, D. "Wisconsin passenger and freight statewide model." *Transportation Research Record,* 2003, Transportation Research Board (TRB), National Research Council (NRC), Washington, DC, pp. 120-129, 2007.

[11] List, G. F., Konieczny, L. A., Durnford, C. L., & Papayanoulis, V. "Best-practice truck flow estimation model for the New York City region." *Transportation Research Record,* 1790, TRB, NRC, Washington, DC, pp. 97-103, 2002.

[12] Jackson, R. W., Schwarm, W. R., Okuyama, Y., & Islam, S. "A method for constructing commodity by industry flow matrices." *Research Paper,* 2004-5, Regional Research Institute, West Virginia University, WV, 2004.

[13] Fisher, M., Ang-Olson, J., & La, A. External Urban Truck Trips Based on Commodity Flow. *Transportation Research Record: Journal of the Transportation Research Board,* no. 1707, TRB, NRC, Washington, D.C., pp. 73–80, 2000.

[14] Gliebe, J. "The North-South transportation initiative: truck travel forecasting methodology report." *Technical Report,* Ohio-Kentucky-Indiana Regional Council, OH, 2001.

[15] Ambite, J. L., Giuliano, G., Gordon, P., Pan, Q., & Bhattacharjee, S. "Integrating heterogeneous data sources for better freight flow analysis and planning." Digital Government Society of North America, Columbia University, NY, 2002.

[16] Roberts, P. & Chiang, Y. "Development of a policy-sensitive model for forecasting freight demand." Report No.DOT-P-30-81-04, U.S. Department of Transportation, Washington, DC, 2981.

[17] Black, W. R. "Commodity flow modeling." *Transportation Research Circular*, E-C011, TRB, NRC, Washington, D.C., pp 136-155, 1999.

[18] Cambridge Systematics, Inc. "FASTrucks Corridor Needs Study: Truck Model Documentation." *Technical report*, Washington State Department of Transportation, WA, 2000.

[19] Mitra, S., Tollive, D., Varma, A., & Dybing, A. "Analyzing effects of spring highway load restrictions on North Dakota's agricultural freight flows." *Transportation Research Record,* 2007, TRB, NRC, Washington, D.C., pp. 92-99, 2007.

[20] SPSS Inc. Version 1.0 © Copyright 2004, IBM, Chicago, IL, 2004.

[21] U.S. Census Bureau & Bureau of Transportation Statistics. CFS website, http://www.census.gov/econ/www/ se0700.html. Accessed in October 15, 2009.

[22] Utah Geographic Information Council. Utah GIS Portal website. http://agrc.its.state.ut.us/. Accessed Sept. 12, 2009.

[23] State of Utah. The official website of the State of Utah, <http://www.utah.gov/government/ governor.html>. Accessed in Sept. 16, 2009.

[24] Ramsey, F. L. & Schafer, D. W. "The statistical sleuth – A course in methods of data analysis, Second Edition." Duxbury, Pacific Grove, CA, 2002.