# 4-step commodity freight forecasting models

May 13
2012

This document presents a summary of the 4-step commodity model, accompanied by case studies for each section.

**CEE 290A**

**Professor Ritchie**

**By:**

**Neda Masoud**

**Kyung Hyun**

## Table of Contents:

# 1 Introduction

The 4-step commodity freight models follow the same steps as the travel-demand passenger models, although the details involved in the two models are not exactly the same.

In the trip generation step, which is the first component of the 4-step commodity models, total productions and attractions of each commodity at each Freight Analysis Zone (FAZ) is obtained for the forecast year. Trip distribution converts total productions and attractions to OD tons between FAZs. In the mode split component the total tonnage is distributed to different modes, and finally at the network assignment step, which is usually applied only to the truck network and sometimes to the rail network, the network is loaded. Figure 1 shows a schematic display of the 4-step commodity model.

Since 4-step commodity models usually are multimodal models, the commodity flow units common to all modes, such as tons, should be used.

In section 2 of this chapter, the commodity and vehicle classification schemes, as well as the zoning structures are discussed. Section 2 presents a brief review of the common data sources used in 4-step commodity models, along with their advantages and shortcomings. Section 4 talks about each of the components of the 4-step commodity models, along with the case studies for each component. Section 5 discusses the validation procedures of the 4-step commodity models. Section 6 lists some of the limitations of these models, along with some examples of the ways different model developers have tried to overcome these problems. Section 7 concludes this chapter with a brief comparison of the 4-step commodity models, with other state-of-the-practice freight forecasting models.

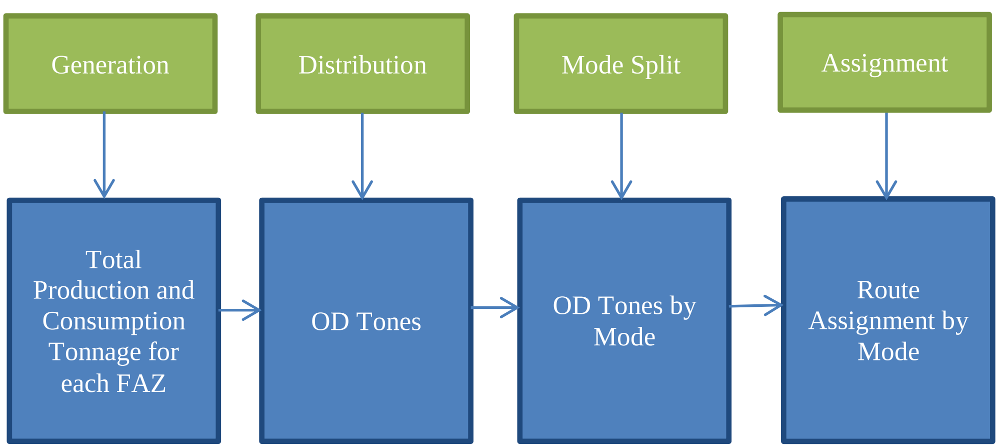


Figure 1. 4-step commodity model

# 2 Model Preparation

## 2.1 Commodity Classification

### 2.1.1 Standard Transportation Commodity Classification (STCC)

STCC code was developed in 1960s by American association of railroads. This commodity classification scheme is the reporting system is the STB's carload waybill sample, CFS prior to 1997, TRANSEARCH commodity flow, VIUS, and FAF databases. STCC is compatible with the Standard Industrial Classification (SIC) system.

STCC coding system classifies commodities to up to 7 digits. The first digit indicates the major economic division of the commodity. For example, 2 is the class of nondurable manufacturing products. The second digit shows the economic major group. For example, 20 indicates food. The third digit shows the industry group. For example, 202 shows dairy products. The fourth digit shows specific industry. For example, 2024 shows ice-cream and frozen desserts. And this commodity classification system provides up to 7 digits. In freight forecasting models, usually up to 2 digits are considered to classify commodities.

### 2.1.2 Standard Classification of Transportation Goods (SCTG)

SCTG is a commodity classification scheme developed jointly by U.S. and Canadian government. This classification system is based on the harmonized schedule classification of commodities, which is the predominant product coding throughout the world. This commodity classification scheme is compatible with the North American Industry Classification System (NAICS). SCTG code is used in 1997 and subsequent CFSs.

### 2.1.3 Comparison of STCC (and SIC) and SCTG (and NAICS)

Most available data on freight flow are based on the STCC commodity classification and SIC industrial classification codes, since SCTC and NAICS came into use only after 1997. Private commercial data sets such as TRANSEARCH still use STCC and SIC coding systems.

SCTG code is not directly transferable to STCC. There is a correspondence between SIC and NAICS only at the four-digit SIC and six-digit NAICS codes.

NAICS codes in general cover a larger range of industries, including emerging industries, advanced technology industries, and product and service industries.

### 2.1.4 Identifying considered commodity groups

A review of the state-wide freight model suggests that most states do not use all the 43 SCTG commodity groups, and further group the commodities before using them in the freight models. Using all the commodities could be very costly and increase the dimension of the problem.

Furthermore, separate explanatory variables may not exist at the disaggregate level for each commodity. Therefore, the commodity groups that share the same explanatory variables may be grouped into one group.

There are no strict criteria on the number of commodities, however a rule of thumb suggest that between 10-20 commodity groups should be good enough.

Some recommended criteria to consider in grouping commodities include:

- Combining commodities with little tonnages
- Keeping major commodities as a single group
- Keeping commodities with specific mode of transportation as a separate group
- Keeping homogeneity of products and related industries in one group

As table 1 shows different states have selected different commodity groups to consider in state-wide freight models. Some states like Nebraska consider 24 commodity groups, while others, like Tennessee and Pennsylvania group their commodities into only 10 groups.

Also, the state of Wisconsin has selected different commodity groups to consider in the production and attraction models. This selection is based on the interaction of industries within the state. For example, if the state produces specific products that use other imported products, considering different commodity groups in the production and attraction models could be an option if there are not enough resources (or no need is sensed) to consider all the commodity group in the study.

Table 1.Number of commodity groups considered in different studies

| State | Number of commodity groups |
|---|---|
| Florida | 14 |
| Indiana | 19 |
| Nebraska | 24 |
| Wisconsin | 16 production, 16 attraction (Not entirely the same) |
| Pennsylvania | 10 |
| Tennessee | 10 |
| Texas | 11 |
| Michigan | 12 |

### 2.1.5 Case study

#### 2.1.5.1 Florida

Table 2 shows the commodity group used in the state of Florida freight model. Some commodities that are more important to the economy of the state are considered separately as a group, and other commodities are further grouped.

Table 2.Commodity groups in the Florida freight model

| No. | commodity group | STCC code |
|---|---|---|
| 1 | Agricultural | 1, 7, 8, 9 |
| 2 | Nonmetallic Minerals | 10, 13, 14, 19 |
| 3 | Coal | 11 |
| 4 | Food | 20 |
| 5 | Nondurable Manufacturing | 21, 22, 23, 25, 27 |
| 6 | Lumber | 24 |
| 7 | Chemicals | 28 |
| 8 | Paper | 26 |
| 9 | Petroleum Products | 29 |
| 10 | Other Durable Manufacturing | 30, 31, 33-39 |
| 11 | Clay/Concrete/Glass | 32 |
| 12 | Waste | 40 |
| 13 | Miscellaneous Freight | 41-47, 5020, 5030 |
| 14 | Warehousing | 5010 |

Source: Cambridge Systematics, Inc., 2008. Forecasting statewide freight toolkit, National Cooperative Highway Research Program (NCHRP) Report 606.Transportation Research Board, Washington, DC.

## 2.2 Zoning Structure

Typically no more than 25% of the freight flow is within the state, and the remaining 75% is the flow to, from and within the state. Therefore in order to properly forecast the freight traffic the geographical region considered in the freight studies usually includes the whole North America (Quick Response II). Zone structure is usually detailed within the study area and less detailed as the distance from the study area increases.

For example, in the state of Indiana's state-wide freight model 92 zones are considered inside the state. The adjacent states to Indiana including Ohio, Illinois, Kentucky and Michigan are grouped into 9 zones, and the rest of the states in the United States are each considered as a separate zone. Not surprisingly, the major destinations in terms of both value and tonnage is Indiana itself, which is the case for most states, and the major destinations in terms of value and tonnage include the neighbor states each divided into multiple FAZs.

Table 3 presents some of the zoning structures considered in various freight models. As this table shows, the number of zones considered within the study area is usually at least as high as the number counties inside the study area. Models that consider finer zones than county level usually try to respect the county boundaries in further dividing them. Other boundaries that freight models try to respect include but are not limited to:

- air-basin
- land use
- natural barriers
- zone size and density

- number of employment
- location of connectors to road and rail networks
- location of major ports and road facilities

Another important boundary is the passenger model TAZs. Respecting this boundary will enable the combining of the freight and travel demand models, and therefore regional analysis. Since most available data sources are at best detailed at the county level, states that try to integrate passenger and freight models, such as Florida and Oregon, have to use either commercial data sources, or do surveys to collect their own data.

Table 3.Zoning schemes considered in different states

| Sate | No. of counties | No. of zones Inside the state | No. of zones outside the state | Total zones | Note |
|---|---|---|---|---|---|
| Iowa | 99 | 99 | 14 | 113 | - |
| Indiana | 92 | 92 | 9+44 | 145 | 9 nodes or terminals representing adjacent states of Ohio, Illinois, Kentucky and Michigan, and a single zone for other states and the District of Columbia |
| Cross-Cascade Corridor Washington | 39 | 54 | 7 | 61 | 7 counties within the corridor were rather subdivided into 2 to 4 zones. 1 zone is for Idaho, and 6 external zones |

In addition to FAZs, gateways are other types of zones, which can include ports and border crossing points that allow the flow inside the study area. Gateways are differentiated from FAZ because in the FAZs the commodity flow is derived by production and attraction activities. However, gateways are only intermodal facilities and there is little or no production or attraction related to the zone within them. The commodity flows to these special zones is usually assumed to be exogenously given.

## 2.3 Vehicle classification

To forecast the truck movement, a vehicle should be identified in terms of size, weight and class. Classified vehicle counts are required in several steps of modeling, for instance, to compare model estimates and truck count data by class in the model validation.

Vehicle classification is normally based on the number of axles, axle spacing and vehicle weight. Even though vehicle classification is basically categorized by characteristics of the trucks' axle, it varies across the model. For example, Caltrans categorized the vehicles into 15 classes by axle including motorcycles and passenger vehicles, and FHWA has the 13 class scheme as shown in table 4. VIUS has more general categories by classifying vehicles into 4 classes; however, it also provides additional 72 classes associated with truck and axle type on specific truck models.

Table 4.FHWA Vehicle Classification Description

| Category | Description |
|---|---|
| 1 | Motorcycles |
| 2 | Passenger cars, Light vans, Light pick-ups |
| 3 | 2 Axle 4 Tire , Full size Pick-ups, Full size Vans, Limos |
| 4 | Buses |
| 5 | 2 Axle, 6 Tire single-unit trucks |
| 6 | 3 Axle single-unit truck |
| 7 | 4 or More Axle single-unit trucks |
| 8 | 3 or 4 Axle single-trailer trucks |
| 9 | 5 Axle single-trailer trucks |
| 10 | 6 or more Axle single-trailer trucks |
| 11 | 5 or less Axle multi-trailer trucks |
| 12 | 6 Axle multi-trailer trucks |
| 13 | 7 or more Axle multi-trailer trucks |

Source: Federal Highway Administration, traffic monitoring guide, vehicle classification monitoring, http://www.fhwa.dot.gov/ohim/tmguide/tmg4.htm#chap1

## 3 Data sources

The Commodity Flow Database (CFD) is the backbone of the commodity flow data used in different models, since it is the only survey that carriers are obligated to response to. Freight models usually use this database directly or indirectly by using Freight Analysis Framework (FAF) or TRANSEARCH data. CFS has some limitations that FAF and TRANSEARCH try to overcome using multiple other data sources. These limitations include but are not limited to:

- CFS does not report imports
- CFS reporting of export flows is also subject to data quality issues resulting from limited sample size
- Flow estimates that have high variability (sampling error)
- On the following areas, CFS either does not collect data or collects insufficient data:
    - Truck, rail and pipeline flows of crude petroleum and natural gas
    - Truck freight shipment of farm-based, fishery, logging, construction, retail, municipal solid waste, household and business moves
    - Imported and exported goods transported by ship, air, and trans-border land (truck, rail) modes

FAF is a more complete version of CFS where the data that is missing from CFS is obtained from other data sources, including international waterborne freight data, Trans-Border freight data, international air freight data, and crude petroleum import and natural gas imports and exports. CFS and FAF are publically available data sources

TRANSEARCH is another source of freight commodity flows. This data source is a proprietary source, which improves CFS data using other data sources. Due to the complexity of the data sources and the proprietary nature of TRANSAEARCH all the data used in this database are not publically available. TRANSEARCH data is at the county level, and the price depends on the number of records purchased.

The main data sources that can be useful in developing, calibrating, or forecasting a 4-step commodity model are presented in table 5.

# 4 4-step commodity model components

## 4.1 Trip Generation

Trip generation is the first component in the 4-step commodity model. Trip generation produces a set of daily or annual flows for each commodity that are originated from and terminated at each FAZ. The productions and attractions of commodities in each zone are assumed to be a function of the economic activity inside that zone.

The production of a commodity at each zone is usually considered to be a function of the employment by the industry that produces that commodity. Although this can be a reasonable assumption for most commodities, there are commodities whose productions cannot be fully or at all explained by employment.

The attraction models are based on the assumption that each commodity at a given zone has two main markets: direct consumers and industrial consumers. Therefore, the attraction of a commodity in a given zone can be explained by the zone's population and the employment of the industrial consumers. Most states use employment of the industrial consumers to calibrate attraction models. The industrial consumers of a given commodity can be identified by examination of an input-output model.

Regression models are usually used to explain productions and attractions for each commodity and at each zone, based on the variables described above. The observed commodity flows to calibrate the production and attraction regression models are usually obtained either from the publically available data sources such as Commodity Flow Survey (CFS), and Freight Analysis Framework (FAF), or private commercial data sources such as Reebie's TRANSEARCH database. The employment data is usually collected from County Business Pattern (CBP) database of the US census bureau, and the population is obtained from the census data. Usually the intercept of the regression line is forced to be zero, since no economic activity is expected if there are no resources available.

Table 5.Main data sources used in 4-step commodity models

| Data Source | Description |
|---|---|
| CFS (Commodity Flow Survey) | A national survey, conducted every 5 years as part of the economic census. Provides flow data by commodity, origin, destination, and mode. The reported flows are in two forms: state to state, and between 86 Metropolitan areas. Combination of modes in CFS is defined as a separate model.<br>Advantages:<br>• Is the only survey shippers are obligated to reply to and therefore is less likely to be biased<br>Disadvantages:<br>• Is not consistent from year to year (methodology, sample size, etc.) making it hard to build time series<br>• Is a sparse matrix (lots of zeros) that should be filled out<br>• Does not include some commodities |
| FAF (Freight Analysis Framework) | A national database, providing freight flow by origin, destination, commodity group, and mode of transport. FAF3 provides aggregated data for 123 zones inside the united states.<br>Advantages:<br>• A more complete version of CFS where the data that is missing from CFS is obtained from other data sources |
| TRANSEARCH | A proprietary national database of freight flows that uses mode-specific data sources to provide flows by origin, destination, commodity, and mode at the county aggregation level. TRANSEARCH is an unlinked trip table that reports trip proportions by each mode.<br>Advantages:<br>• Accepted and used by FHWA and many states and MPOs.<br>Disadvantages:<br>• Is a proprietary database and price depends on the number of records purchased<br>• The complexity of the sources and the proprietary nature of the database prohibit its owners to reveal all the data sources used to build this database |
| Census population | Obtained every 5 year as part of the US census. The population estimates are provided at the state, county, city and town level. |
| CBP (County Business Pattern) | CBP annual reports show the economic activity by industry in county level. Employment by industry used in development of the generation models in the first step of the 4-step commodity model is obtained from CBP.<br>Disadvantages:<br>• Does not include all employment data |
| VIUS (Vehicle Inventory and Use Survey) | A survey that is conducted every 5 years, as a part of the US Economic Census. This survey contains a 2000 sample of truck data per state, including physical and operational characteristics. VIUS is usually used to convert tons to truck trailer for each commodity class in the network assignment step of the 4-step commodity model. |

| | |
|---|---|
| WIM | Data provides 24hour traffic information including the speed, weight, and count on the highway at the state level, and the traffic count can be used for validating the freight forecast model.<br>Disadvantages:<br>• Does not provide the arterial or local traffic condition and has erroneous records due to the WIM stations' performance. |
| PMA | Obtained by the Pacific Maritime Association every year. This database provides waterborne cargo loaded and discharged in California, Washington and Oregon ports.<br>Advantages:<br>• The data includes international flows and domestic trades assessed by the tonnages and values<br>Disadvantages:<br>• The data are obtained only from the membership carriers of PMA and does not include liquid or LPG carrier cargo.<br>• Cargo tonnage can be over-estimated by reporting all loaded and discharged cargo, regardless of actual transport. |
| Waterborne | A national data source that reports point of loading and unloading for each commodity. The reports are usually made to the Army Corps of Engineers on the basis of vessel movements.<br>Disadvantages:<br>• Military cargo that are moved in department of defense vessels are reported |
| Trans-border | Is obtained from documents collected by the U.S. Customs and Border Protection, and shows flows by commodity and mode between united states and its NAFTA partners.<br>Advantages:<br>• The data is obtained from a complete enumeration of documents obtained from US customs and border protection and therefore is free of sampling errors.<br>Disadvantages:<br>• Other sources of error, including underestimation of low-valued transactions, undocumented shipments, and reporting errors are present |
| STB Carload Rail Waybill sample | Is a national stratified sample of Carload Waybill, containing origin, destination, commodity, tons, etc. from railroad terminating more than 4500 cars per year.<br>Disadvantages:<br>• Is not available for public use. However, there is an aggregated publically available version |
| surveys | Local surveys where necessary. |
| Interviews | Interviews where necessary. |

The assumption behind the development of trip generation regression models is that productions and attractions for each zone are based on the economic activity of that zone. This does not hold true in case that there are intermodal facilities in a zone. In this case there is freight activity in a zone where there is little or no economic activity associated with it. Neglecting intermodal facilities could manifest itself in the form of outliner zones in regression models. This can be resolved by treating intermodal facilities as special generators.

Whether special generators are required or not depends on the type of commodity flow databases used. In databases where the commodity flows are unlinked, such as TRANSEARCH, the flow of a commodity will begin or end not only at the ultimate producing or attracting zones, but also at intermodal facilities. Therefore, special generators are required. On the other hand, if commodity flow databases that are used contain linked trip flows, such as CFS, then special generators for intermodal facilities are not required. In case of CFS, the linked modal trips are presented in forms such as “rail-truck” mode, and the freight flows begin or end at the ultimate producing or consuming zones. Even is this case, special generators for international gateways, such as border crossings and ports, might be required.

Once the regression models are developed, the only data required to forecast freight is the population and employment by industry at the zonal level for the forecast year. The commodity flow of the intermodal facilities, such as seaports or airport, if required, can be obtained directly from those facilities. The zonal productions and attractions used in the distribution step will then be the sum of regression model outputs, and the intermodal facility volumes.

One limitation of using the previously mentioned data sources on commodity flow is that commodity flows are usually available at best at the county level, and inter-county movements are not captured in these databases. Doing independent surveys or developing regression models to obtain flow rates at intra-county zonal level are some of the potential approaches to tackle this problem.

Table 6 below displays some of the approaches by different sates in developing the state-wide freight models. As table 6 shows regression models are often used to explain productions and attractions. Also, the independent variables used are mostly employment and population. The data sources used in each cased in also presented in the table.

### 4.1.1 Case study:

#### 4.1.1.1 Florida

Tables 7 and 8 show the production and attraction equations for the state of Florida freight forecasting model. Both production and attraction equations use a zero intercept in the generation equations. The production equations mainly use the employment by industry and the attraction equations use the employment by the industrial consumer for each commodity group. Population is also used as an explanatory variable is some of the attraction equations.

Table 6.Generation models in different states

| State | Method | Independent variables | Data source |
|---|---|---|---|
| Florida | Regression | Employment by the SIC<br>Population | • TRANSEARCH freight database<br>• Intermodal ports or terminals<br>• CFS |
| Indiana | Regression | Employment by the NAICS<br>Population (to represent the consumer market) | • 1977 CFS<br>• 1977 census<br>• 1977 county business pattern<br>• 1993 CFS<br>• 1993 Census<br>• 1992 Woods & Poole (population and employment forecast-future year) |
| Nebraska | Regression | Employment by the SIC<br>Population | • 1993 CFS (production)<br>• IMPLAN (attraction coefficients) |
| Vermont | NA | NA | • TRANSEARCH<br>• Road-side surveys<br>• Motor carrier surveys<br>• Interviews with shippers |
| Iowa | Fratar | Employment by the SIC<br>Population | • TRANSEARCH |
| Wisconsin | Regression | Employment by SIC | • 2001 TRANSEARCH |

Table 7.Production equation in the Florida freight model

| No. | Commodity Groups | Coefficient Variable | (Employment) |
|---|---|---|---|
| 1 | Agricultural | 45.597 | SIC07 |
| 2 | Nonmetallic | 6,977.77 | SUM(SIC10-14) |
| 3 | Coal | | No Production Employment |
| 4 | Food | 245.464 | SIC20 |
| 5 | Nondurable Manufacturing | 90.12 | SUM(SIC21,22,23,25,27) |
| 6 | Lumber | 241.464 | SIC24 |
| 7 | Chemicals | 678.583 | SIC28 |
| 8 | Paper | 190.814 | SIC26 |
| 9 | Petroleum Products | 795.117 | SIC29 |
| 10 | Other Durable Manufacturing | 212.202 | SUM(SIC30,31,33-39) |
| 11 | Clay, Concrete, Glass | 1498.501 | SIC32 |
| 12 | Waste | 0.5 | TOTEMP |
| 13 | Miscellaneous Freight | 0.599 | TOTEMP |
| 14 | Warehousing | 314.852 | SIC50 + SIC51 |

Source: Cambridge Systematics, Inc., 2008. Forecasting statewide freight toolkit, National Cooperative Highway Research Program (NCHRP) Report 606.Transportation Research Board, Washington, DC.

Table 8.Attraction equations in the Florida freight model

| No. | Commodity Groups | Coefficient | Variable | Coefficient | Variable |
|---|---|---|---|---|---|
| 1 | Agricultural | 23.537 | SIC20 | 0.448 | Pop |
| 2 | Nonmetallic Minerals | 1461.302 | SIC28 | | |
| 3 | Coal | 178.639 | SIC49 | | |
| 4 | Food | 109.51 | SIC51 | | |
| 5 | Nondurable Manufacturing | 24.698 | SIC51 | | |
| 6 | Lumber | 147.624 | SIC25 | | |
| 7 | Chemicals | 83.247 | SIC51 | | |
| 8 | Paper | 23.924 | SIC51 | | |
| 9 | Petroleum Products | 0.228 | Pop | | |
| 10 | Other Durable Manufacturing | 46.762 | SIC 50 | | |
| 11 | Clay, Concrete, Glass | 2.964 | Pop | | |
| 12 | Waste | 68.089 | SIC33 | | |
| 13 | Miscellaneous Freight | 2.886 | SUM (SIC42,44,45) | | |
| 14 | Warehousing | 2.926 | Pop | | |

Source: Cambridge Systematics, Inc., 2008. Forecasting statewide freight toolkit, National Cooperative Highway Research Program (NCHRP) Report 606.Transportation Research Board, Washington, DC.

### 4.1.1.2 Indiana

The Indiana statewide freight forecasting models uses 19 commodity groups, and uses linear regression to estimate 19 production and 19 attraction equations.

As table 9 shows, the production equations are mainly functions of the employment by the industry that produces the products. Only in the case of “waste and scrap material” commodity group, the production equation is a function of population.

The attraction equations are mainly functions of the consumer industries and population. However, there are cases that the attraction equation for a commodity is simply a linear function of the production equation for the same commodity.

Another important point is that the intercept of all the production and attraction equations except for the “farm products” commodity group is set to zero. This is based on the fact that if the value of the explanatory variables that cause production and attraction of a commodity within a FAZ are zero, then that commodity should not be produced or attracted at that FAZ.

Table 9.Production and attraction equations, and the corresponding adjusted $R^2$s by state of Indiana statewide freight project

| Commodity group | production model | attraction model | Pro –Adj $R^2$ | Att.-Adj $R^2$ |
|---|---|---|---|---|
| Farm Products | Prod01 = 1445 –.523 Agser +.0048 Cash | Attr01 = .819 *Prod01* | 0.562 | 0.66 |
| Coal | Prod11 = 7.6 Coal | Attr11 = 3.1 Coal + 5.3 Min | 0.65 | 0.657 |
| Nonmetallic Minerals | Prod14 = .078 Man | Attr14 = .997 *Prod14* | 0.658 | 0.977 |
| Food and Kindred Products | Prod20 = .282 Food | Attr20 = .832 Pop + .162 Food | 0.965 | 0.965 |
| Basic Textiles | Prod22 = .016 Tex | Attr22 = .003 App + .0001 All | 0.931 | 0.743 |
| Apparel | Prod23 = .004 App | Attr23 = .002 App + .011 Pop | 0.919 | 0.926 |
| Lumber and Wood Products | Prod24 = .668 Lum | Attr24 = .728 *Prod24* | 0.808 | 0.805 |
| Furniture and Fixtures | Prod25 = .017 Furn | Attr25 = .033 Pop + .002 Furn | 0.906 | 0.96 |
| Pulp and Paper Products | Prod26 = .103 Pulp + .056 Lum | Attr26 = .085 Pulp + .002 Furn | 0.886 | 0.953 |
| Chemicals and Allied Products | Prod28 = .150 Chem + 1.164 Pet | Attr28 = .077 Chem + .455 Pet + .683 Pop | 0.758 | 0.851 |
| Petroleum and Coal Products | Prod 29 = 6.857 Pet | Attr29 = 4.007 Pet + 1.881 Pop | 0.945 | 0.938 |
| Stone, Clay and Glass Products | Prod32 = 2.882 Pop | Attr32 = 2.914 Pop | 0.851 | 0.871 |
| Primary Metal Products | Prod33 = .085 Met | Attr33 = .093 Met + .061 Fab | 0.982 | 0.923 |
| Fabricated Metal Products | Prod34 = .013 Met + .034 Fab | Attr34 = .035 Fab | 0.927 | 0.861 |
| Machinery (except Electrical) | Prod35 = .013 Mac | Attr35 = .010 Mac | 0.883 | 0.878 |
| Electrical Machinery | Prod36 = .004 Met + .004 Fab + .003 Elec | Attr36 = .005 Fab + .034 Pop | 0.826 | 0.915 |
| Transportation Equipment | Prod37 = .040 Tran | Attr37 = .027 Tran | 0.753 | 0.837 |
| Waste and Scrap Material | Prod40 = .00048 **Pop** | Attr40 = .0067 Man | 0.704 | 0.791 |
| Other Manufactured Products | Prod50 = 1.097 Attr50 | Attr50 = .245 Pop | 0.858 | 0.857 |

Source: W.R. Black, Transport Flows in the State of Indiana: Commodity Database Development and Traffic Assignment, Phase 2, Bloomington, Indiana: Transportation Research Center, Indiana University, 1997.

## 4.2 Trip Distribution:

Trip distribution is the second component of the 4-step commodity model. The output of the trip generation component, which is the daily or annual trip rates by commodity to and from each zone, is the input to the trip distribution component.

In trip distribution a trip table for each commodity is created. Trip tables show flow rates from each zone to all other zones. The flow rates are usually in formats that are common to all modes, such as tonnages.

Gravity models are usually used for trip distribution. To create the trip tables by a gravity model, total productions and attractions by zone (outputs of the trip generation step), and relative impedances between zones in the form of friction factors are required. Friction factors are measures of zonal desirability, and are usually expressed in form of travel time or transport cost. The trips generated by a gravity model are presented in the form of :

$$T_{ij} = \frac{P_i A_j F_{ij}}{\sum_j A_j F_{ij}}$$

$$F_{ij} = e^{-\frac{c_{ij}}{k}}$$

Where $F_{ij}$ is the friction factor between zones $i$ and $j$, and
$T_{ij}$ =amount of a given commodity (tons) shipped from origin zone $i$ to destination zone $j$
$P_i =$ amount of a given commodity (tons) available for shipment from origin $i$
$A_j =$ amount of a given commodity demanded at destination $j$
$c_{ij} =$ a measure of the travel impedance between zones $i$ and $j$ (usually in miles or time)
$k =$ average distance between all zones

Note that the average distance for travel is commodity specific and therefore the gravity model for each commodity will have a different constant coefficient. The models, impedances, and data sources used by different freight forecasting models in freight forecasting models are displayed in table 10.

Table 10.Trip distribution models by different states

| State | Method | impedance | Data source |
|---|---|---|---|
| Florida | Gravity model | Distance and time (time for short distance trips) | TRANSEARCH (average trip length) |
| Indiana | Gravity model | Straight line distance | 1993 CFS |

Note that travel time or distance are not always the best impedances for trip distribution. Shipping rate data, if available, could be more appropriate measures of impedance. Also, travel time or cost might not be able to explain the freight distribution as assumed by the gravity model.

Long-standing relationships or product differentiation may affect the freight distribution. (Iowa-2)

Another approach to do the trip distribution is using fractional split models, which have shown to give better results compared to gravity models (Sivakumar and Bhat). A fractional split distribution model estimates the fraction of commodity at the destination zone from all the origin zones. The reason these models render better results than gravity models lies in the fact that they consider the impact of the location of zones on the amount of commodities produced at each zone.

### 4.2.1 Case study:

The Indiana statewide freight forecasting model uses a standard gravity model in the distribution step presented below:

$$S_{jk} = A_j B_K O_j D_k \exp\left(-\beta c_{jk}\right)$$

$$A_j = \left[\Sigma\, B_k\, D_k \exp(\beta c_{jk})\right]^{-1}$$

$$B_j = \left[\Sigma\, A_j\, O_j \exp(\beta c_{jk})\right]^{-1}$$

Where
$S_{jk}$ = the amount of a given commodity shipped from origin zone $j$ to destination zone $k$
$O_j$ = the amount of a given commodity available for shipment from origin $j$
$D_k$ = the amount of a given commodity demanded at destination $j$
$c_{jk}$ = a measure of the cost

The real data from Indiana was then used to refine the estimate of flows.

Table 11 shows the average distance per ton for each commodity resulted from the trip distribution step, along with the actual average distance for Indiana and the whole US.

## 4.3 Mode choice

The mode choice model component forecasts freight flows in tonnage and by commodity group between the FAZ for all freight modes; highway-based, rail, the water-based, and the air mode. The total tonnage of each OD pair from the trip distribution is disaggregated to O-D tons by mode in the mode choice step. According to the Quick Response Freight Manual II (FHWA, 2007), the goods characteristics are strongly related to mode selection. The air and truck which are the fastest and the most reliable modes ship the time-sensitive and highest priced but the lowest weight products. The rail and water which are the slower, less reliable and less visible modes take charge of the less time-sensitive, highest weight, and lower value commodities. The mode sharing of rail, air, and water sometimes take little effort compared to trucks due to the limitation on the data availability. However, those modes need to be handled significantly because they not only take up considerable part in transporting freight but also serve important role in estimating truck forecasts. At the same time, they could be future competitors of the truck freight. Nevertheless, in many statewide forecast models, estimating truck volumes on highway is a primary goal.

Table 11.Actual and modeled average distance travelled by commodity for the state of Indiana and the whole US

| Commodity group | US average | | Indiana average | |
|---|---|---|---|---|
| | actual | modeled | actual | modeled |
| Farm Products | 434 | 434 | 435 | 432 |
| Coal | 432 | 432 | 85 | 436 |
| Nonmetallic Minerals | 87 | 116 | 44 | 122 |
| Food and Kindred Products | 315 | 311 | 333 | 311 |
| Basic Textiles | 458 | 445 | 236 | 489 |
| Apparel | 658 | 420 | 391 | 397 |
| Lumber and Wood Products | 182 | 190 | 220 | 222 |
| Furniture and Fixtures | 591 | 592 | 794 | 563 |
| Pulp and Paper Products | 464 | 313 | 313 | 314 |
| Chemicals and Allied Products | 434 | 345 | 280 | 294 |
| Petroleum and Coal Products | 152 | 153 | 89 | 140 |
| Stone, Clay and Glass Products | 105 | 202 | 124 | 189 |
| Primary Metal Products | 365 | 365 | 356 | 361 |
| Fabricated Metal Products | 359 | 358 | 342 | 345 |
| Machinery (except Electrical) | 559 | 500 | 472 | 473 |
| Electrical Machinery | 649 | 505 | 481 | 483 |
| Transportation Equipment | 560 | 487 | 449 | 446 |
| Waste and Scrap Material | 211 | 211 | 181 | 243 |
| Other Manufactured Products | 560 | 507 | 426 | 465 |

Source: W.R. Black, Transport Flows in the State of Indiana: Commodity Database Development and Traffic Assignment, Phase 2, Bloomington, Indiana: Transportation Research Center, Indiana University, 1997.

Multimodal trip table which is the output of the mode choice step serves as an input to the assignment model component. However, the flow units are commonly represented by the tons, conversion to vehicle is needed for assignment step. It will be discussed in following chapter. Table 12 presents the commodity tonnage by mode and by commodity flow for Florida Intermodal Statewide Highway Freight Model.

Table 12.Commodity tonnage, Florida Intermodal Statewide Highway Freight Model (unit: 1000 annual tons)

| Commodity Group | Truck | % | Carload | % | Intermodal Rail | % | Air | % | Water | % | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Agricultural | 320 | 6.4 | 710 | 14.3 | 3,934 | 79.3 | - | - | - | - | 4,964 |
| Non-metallic Minerals | - | - | 67,817 | 100.0 | - | - | - | - | - | - | 67,817 |
| Coal | - | - | 32,270 | 100.0 | - | - | - | - | - | - | 32,270 |
| Food | 22,740 | 82.8 | 2,762 | 10.1 | 1,959 | 7.1 | 13 | - | - | - | 27,474 |
| Non-durable manufacturing | 4,764 | 92.8 | 312 | 6.1 | 17 | 0.3 | 11 | 0.2 | 30 | 0.6 | 5,134 |
| Lumber | 12,342 | 85.6 | 366 | 2.5 | 1,663 | 11.5 | 49 | 0.3 | - | - | 14,420 |
| Chemicals | 27,419 | 87.1 | 2,684 | 8.5 | 135 | 0.4 | 47 | 0.1 | 1,210 | 3.8 | 31,495 |
| Paper | 11,806 | 66.0 | 1,464 | 8.2 | 1,046 | 5.8 | 14 | 0.1 | 3,564 | 19.9 | 17,894 |
| Petroleum Products | 7,757 | 70.8 | 2,485 | 22.7 | 502 | 4.6 | 10 | 0.1 | 204 | 1.9 | 10,958 |
| Other Durable Mfg | 16,484 | 81.6 | 2,136 | 10.6 | 623 | 3.1 | 12 | 0.1 | 940 | 4.7 | 20,195 |
| Clay/Concrete/Glass | 26,010 | 42.2 | 124 | 0.2 | 34,166 | 55.5 | 1,088 | 1.8 | 178 | 0.3 | 61,566 |
| Waste | 4,141 | 100.0 | - | - | - | - | - | - | - | - | 4,141 |
| Miscellaneous Freight | - | - | - | - | 10,399 | 100.0 | - | - | - | - | 10,399 |
| Warehousing | 83,020 | 100.0 | - | - | - | - | - | - | - | - | 83,020 |
| TOTAL | 216,803 | 815.3 | 113,130 | 283.2 | 54,444 | 267.6 | 1,244 | 2.7 | 6,126 | 31.2 | 391,747 |

Source: Cambridge Systematics, Inc., 2002, Technical memorandum; model specification for FISHFM (Freight Intermodal State wide Highway Freight)

Discrete Logit choice model is the most common and widely used in the mode choice step. However, due to the limitation of the information or the underlying assumption of a model, developing a Logit choice model is very complicated. The Logit model is based on the random utility maximization theory. The utility is defined as a function of the generalized cost which is based on the distance, time, cost and reliability for mode. However, those are not easy to obtain and might vary for different analysis.

In addition, the fundamental assumption of the Logit model is that each shipping unit should serve as a decision maker considering utility of a mode. However, in reality, the decisions are involved with a few of individuals even though freight movement provides the millions of tons of commodities. This assumption makes it difficult to apply Logit model in freight forecasts.

Therefore, many states have applied a simple approach for mode choice. They assumed that the existing mode splits remain in future, therefore the current mode shared by commodity groups can be applied to estimate the future forecasts. In general, freight movement may not be sensitive to the travel time, costs and reliability which are the key component of the utility function for Logit model. From this perspective, many statewide freight forecasting use the current mode shares in the mode choice step.

CARGO model in CUBE provides the Logit model as follow;

$$z_{cm}(d,t,x) = \frac{e^{-G_{cm}(d,t,x)}}{\sum_m e^{-G_{cm}(d,t,x)}}$$

Where:
$z_{cm}$ = proportion of commodity group c by mode m
$G_{cm}(d,t,x)$ = Generalized cost for commodity group c by mode m

Generalized cost can be calculated as follows:

$$G_m(d,t,x) = k_{ocm} + k_{1cm} * d + k_{2cm} * t + k_{3cm} * x$$

Where:
$k_{\mathrm{ocm}}$=constant term, k1distance coefficient, k2cm=time coefficient, k3cm=cost coefficient
$d$ = distance
$t$ = time
$x$=cost

In CARGO, multinomial Logit model is utilized for mode splits to assign commodity and the utility is calculated as a function of distance, time, and the cost.

### 4.3.1 Case Study

#### 4.3.1.1 Indiana

The mode choice model of Indiana, NEWMODE, provides 17 mode categories(9 individual modes and 8 multi-modes) as shown in table (). Those modes categories are specified with 9 distance-based sub-categories; less than 50 miles, 50 to 99 miles, 100 to 249 miles, 250 to 499 miles, 500 to 749 miles, 750 to 999 miles, 1,000 to 1,499 miles, 1,500 to 1,999 miles, and 2,000 miles or more. The future forecasts are estimated in the distance categories under the assumption that the mode choices of the base year remain the same.

Table 13.Modal Categories by Indiana freight model

| Single Modes | Multiple Modes |
|---|---|
| Parcel/Courier | Private Truck and For-Hire |
| U.S. Portal Service | Truck and Air |
| Private Truck | Truck and Rail |
| For-Hire Truck | Truck and Water |
| Air | Truck and Pipeline |
| Rail | Rail and Water |
| Inland Water | Inland Water and Great Lakes |
| Great Lakes | Inland Water and Deep Sea |
| Deep Sea Water | |

Source: William R. Black, 1993, Indiana commodity transport model Transportation Research Center, Indiana University.

#### 4.3.1.2 Florida

In the Florida state-wide freight model, mode split model assumes that the air and the waterborne mode splits are maintained in the future due to the limitation on network information of these two modes. Instead, future forecast is focused on the truck and rail for producing the accurate tonnage. Based on the TRANSEARCH data as base year pivots, changes in mode share is calculated with an incremental Logit model. To overcome the restriction on the basic assumption of the Logit model, Florida applies the incremental Logit mode split model only considering the changes in mode split percentages based on the historical data without estimating the existing mode shares.

$$S'_i = \frac{S_i * \exp(\triangle U_i)}{\sum_{j=1}^{J} S_j * \exp(\triangle U_j)}$$

Where:

$S'_i$ = new share of mode i

$S_i$ =original share of mode i

$\triangle U_i$ = utility of mode i in the choice set J (j=1,2,3……J)

= Modal constant + $b^v$ * $\triangle$Explanatory variable $i^v$

$b^v$ = coefficient for explanatory variable

Explanatory variables = ln (travel time * commodity value per ton * travel cost)

The travel time and cost which are functions of the distance are selected as explanatory variables. The coefficient for the model is taken from the survey in NEWYORK and calibrated with the Florida database.

#### 4.3.1.3 Wisconsin

As table 14 shows, Wisconsin applies a simple assumption on the mode share. The existing mode share is assumed to be identical to the future. The TRANSEARCH database of the tonnage matrices for each mode is used as a base to forecast the future.

Table 14.Trendlining freight forecast method in Wisconsin

| Year | Origin | Destination | SIC | Tons | Rail Share | Truck Share | Rail Tons | Truck Tons |
|---|---|---|---|---|---|---|---|---|
| 1992 | ABC | DEF | XX | 100 | 40% | 60% | 40 | 60 |
| | | | YY | 200 | 60% | 40% | 120 | 80 |
| | | | TOTAL | 300 | 53% | 47% | 160 | 140 |
| | | | | | | | | |
| 2020 | ABC | DEF | XX | 200 | 40% | 60% | 80 | 120 |
| | | | YY | 500 | 60% | 40% | 300 | 200 |
| | | | TOTAL | 700 | 54% | 46% | 380 | 320 |

Source: Wilbur Smith Associates in association, 1996. Multimodal freight forecasts for Wisconsin The rail and the truck share remain the same but the tonnage of each mode might vary by estimated total tons.

## 4.4 IV. Data Conversion

To assign the commodity flow on the network, the conversion from tonnage to daily freight truck trips is required. Moreover, the unit of freight flow should correspond to the passenger flow when the freight flow is combined with the passenger flow on the same network. The freight annual flow by commodity group in tonnage can be converted to the vehicles with the commodity specific factor or payload factor. Since each commodity has different handling characteristics and weights, the payload factor should be computed by the commodity groups. Several data sources such as VIUS from U.S. Census Bureau or carrier survey provide the average payload factors by commodity group for the long-haul trips. It includes the information on the percentage of miles traveled in terms of commodity group which can be used to develop the payload factors with the weight of the truck records. For the short-haul trips, each state model estimates the factor relevant to each state.

Holguin-Veras and GopalPatil introduced a model that combined a commodity-based model with empty trips. They compared a model considering the loaded and the empty truck trips together (model 1) with a second model with two sub-models, one considering the loaded and the other considering the empty traffic (model 2). The data and network from Guatemala City are applied to a case study.

Two objective functions are applied to each model, respectively.

$$\min(\beta, p) \quad F_v = \sum_l (V_i^o - V_i^e)^2 \qquad \text{model 1}$$

$$\min(\beta, p) \quad F_v = \sum_l (V_i^o - V_i^e)^2 = \sum_l (W_i^o - W_i^e)^2 \sum_l (U_i^o - U_i^e)^2$$

$$V_i^e = \sum_i \sum_j z_{ij}^e \overline{p}_{ij}^l = \sum_i \sum_j (x_{ij}^e + y_{ij}^e) \overline{p}_{ij}^l = \sum_i \sum_j x_{ij}^e \overline{p}_{ij}^l + \sum_i \sum_j y_{ij}^e \overline{p}_{ij}^l \qquad \text{model 2}$$

$$= W_l^e + U_l^e$$

Where:

$V_l^e$: Estimated and Observed truck traffic on link i

$W_l^o, W_l^e$: Observed and estimated loaded link volumes

$U_l^o, U_l^e$: Observed and estimated empty link volumes

The second model outperformed to reduce errors associated with estimated traffic and observed counts. This result can explain the pattern of empty tripsin an urban area that comprises 30% to 40% of the total truck trips. The formulation introduced in this paper can be applied to a statewide model to improve replicating the travel pattern.

In general, waterborne and air freight are not converted because the assignment for those modes are not performed.

Besides, conversion from annual tons to daily tons or annual trips to daily trips should be considered for assignment. Typically, each state provides the working days that trucks are expected to move as a conversion factor.

### 4.4.1 Case Study

#### 4.4.1.1 Florida

Florida developed the payload factor based on the Vehicle Inventory and Usage Survey (VIUS) data by commodity groups and distance classes. The percentages of the empty trucks are also considered to develop the conversion factors by commodity and the distance. Table 15 presents the average payload by commodity in Florida. The payload factor by the commodity and the distance class can be calculated by average pound miles divided by average miles. Each commodity has a different payload factor, as different densities, shipment sizes and handling characteristics determine the conversion factor. The distances are categorized to five class; less than 50 miles, 50-100 miles, 100-200 miles, 200-500 miles, and over 500miles trip. As table16 shows, empty truck percentages are combined with the tonnage conversion factors to produce annual tons to annual truck conversion factor. After the conversion, daily truck trip tables are generated using the annual to daily factor of 306 working days per year.

Table 15. Tonnage to truck conversion factors for Florida freight model

| Commodity Group | Average payload in pounds | | | | | |
|---|---|---|---|---|---|---|
| | On road average | < 50 miles | 50-100 miles | 100-200 miles | 200-500 miles | 500+ miles |
| Agricultural | 32,725 | 18,408 | 36,286 | 43,901 | 38,956 | 35,572 |
| Minerals/Coal | 41,637 | 41,237 | 35,000 | 42,138 | - | 46,000 |
| Food Products | 36,456 | 17,283 | 37,194 | 44,574 | 42,209 | 42,465 |
| Non-durable manufacturing | 17,358 | 7,155 | 10,105 | 36,208 | 12,441 | 29,579 |
| Lumber | 28,052 | 9,405 | 50,378 | 44,780 | 56,639 | 48,314 |
| Paper | 30,218 | 22,630 | 19,924 | 39,723 | 34,003 | 36,960 |
| Chemicals | 33,170 | 23,215 | 41,506 | 39,240 | 46,916 | 37,329 |
| Petroleum Products | 42,082 | 39,091 | 51,042 | 54,648 | 43,708 | 34,653 |
| Other Durable Mfg | 22,761 | 10,237 | 13,944 | 37,440 | 38,416 | 34,464 |
| Clay/Concrete/Glass | 36,931 | 31,647 | 40,617 | 39,934 | 45,413 | 44,802 |
| Waste | 25,801 | 20,565 | 34,060 | 32,295 | 46,132 | 42,066 |
| Miscellaneous Freight | 24,878 | 13,796 | 14,416 | 41,777 | 38,575 | 36,853 |
| Warehousing | 18,137 | 18,039 | 13,068 | 47,820 | 6,685 | 23,125 |
| Average | 28,429 | 19,931 | 24,034 | 41,130 | 39,219 | 37,602 |

Source: Cambridge Systematics, Inc., 2002, Technical memorandum; model specification for FISHFM (Freight Intermodal State wide Highway Freight)

Table 16.Annual tons to annual trucks conversion factors for Florida freight model

| Commodity group | < 50 miles | 50-100 miles | 100-200 miles | 200-500 miles | 500+ miles |
|---|---|---|---|---|---|
| Agricultural | 0.1001 | 0.0494 | 0.0376 | 0.0461 | 0.0542 |
| Minerals/Coal | 0.0421 | 0.0564 | 0.0451 | 0.0443 | 0.0435 |
| Food Products | 0.1109 | 0.0516 | 0.0431 | 0.0453 | 0.0458 |
| Non-durable manufacturing | 0.2706 | 0.1960 | 0.0535 | 0.1583 | 0.0647 |
| Lumber | 0.1856 | 0.0373 | 0.0423 | 0.0340 | 0.0397 |
| Paper | 0.0816 | 0.0920 | 0.0496 | 0.0569 | 0.0518 |
| Chemicals | 0.0815 | 0.0439 | 0.0459 | 0.0331 | 0.0504 |
| Petroleum Products | 0.0440 | 0.0336 | 0.0360 | 0.0436 | 0.0535 |
| Other Durable Mfg | 0.1831 | 0.1315 | 0.0509 | 0.0481 | 0.0532 |
| Clay/Concrete/Glass | 0.0507 | 0.0428 | 0.0456 | 0.0414 | 0.0430 |
| Waste | 0.0897 | 0.0537 | 0.0619 | 0.0409 | 0.0473 |
| Miscellaneous Freight | 0.1398 | 0.1193 | 0.0454 | 0.0455 | 0.0523 |
| Warehousing | 0.1097 | 0.1328 | 0.0342 | 0.1613 | 0.0859 |

Source: Cambridge Systematics, Inc., 2002, Technical memorandum; model specification for FISHFM (Freight Intermodal State wide Highway Freight)

#### 4.4.1.2 Indiana

In Indiana, the density factors by commodity were developed for rail and truck flows. The Carload waybill sample is used for the rail density factor, and the truck is assumed to carry 40% of the rail-load based on the different capacity in rail and truck. Table () represents the traffic density factors for rail and truck by commodity. In addition, conversion factor of annual tons to daily truck flows is applied from the Highway Capacity Manual Special Report 209, which provides an annual to daily factor of 306 days and weekend truck traffic of multiplying 0.4 by weekday truck traffic.

Table 17.Traffic density factors in Indiana

| Commodity STCC | Import rail traffic | Export rail export | Weighted rail density (tons) | Weighted truck density (tons) |
|---|---|---|---|---|
| 1 | 94.9 | 96.2 | 96.13 | 38.44 |
| 11 | 100.6 | 99.1 | 100.42 | 40.17 |
| 14 | 97.1 | 97.4 | 97.2 | 38.88 |
| 20 | 77.35 | 80.36 | 79.52 | 31.81 |
| 22 | 25 | 15 | 18.33 | 7.33 |
| 23 | ------- | ------- | *10.00 | *4.00 |
| 24 | 73.88 | 55.5 | 72.27 | 28.91 |
| 25 | ------- | 15 | 15 | 6 |
| 26 | 64.82 | 50.64 | 62.1 | 24.84 |
| 28 | 85.11 | 90.11 | 87.58 | 35.03 |
| 29 | 63.2 | 77.16 | 65.9 | 26.36 |
| 32 | 86.7 | 77.1 | 81.15 | 32.46 |
| 33 | 87.48 | 85.21 | 85.82 | 34.33 |
| 34 | 28.4 | 16.16 | 19.76 | 7.9 |
| 35 | 68.75 | 21.7 | 28.42 | 11.37 |
| 36 | 18.8 | 16.25 | 16.69 | 6.68 |
| 27 | 19.93 | 23.4 | 22.5 | 9 |
| 40 | 75.4 | 82.6 | 78.47 | 31.39 |
| 21,27,30,31 38 and 39 | 92.85 | 14.88 | 86.56 | 34.62 |

*Estimated values

Source: Wilbur Smith Associates in association, 1996. Multimodal freight forecasts for Wisconsin

### 4.5 Network Assignment

The last step of the 4-step commodity freight model is the network assignment component. Freight truck trips by mode between FAZ are provided through the former 3 components and the conversion step. In general, there are three assignment models including rules-based, freight truck only, and multiclass network assignment model. For trucks, the assignment step finds an optimum path and estimates the flow on the highway. However, it should be noted that the freight model cannot be developed separately with the auto trips because the two modes would be correlated in highway and affect each other during congestion periods. Considering the congestion effect, several techniques can be applied in terms of multimodal analysis including all-or-nothing, equilibrium, capacity-restrained equilibrium, and stochastic multi-path capacity restraint techniques.

The rules-based assignment (fixed-path assignment) techniques are normally developed for rail, water, and air assignments. Those modes have a fixed route that is not flexible in the new facility or changes in performance, therefore analysts should rely on the existing network and routing procedure. The technique is considered to assign long-haul traffic as those traffic patterns are not easily affected by changes in network or performance and may remain fixed over a long period of time.

Dynamic path assignment including freight-truck-only network assignment and multiclass network analysis are typically used for freight trucks. Since the information on the travel time or distance between OD is available, the rules-based assignment is not suitable for the freight truck assignment. If the table of the passenger vehicle is not available, the freight-truck-only network assignment can be developed. Otherwise, multiclass network analysis will be used for assignment. It can respond to the congestion on utilizing highway and may alter routes based on travel time.

In the freight truck only network assignment, all-or-nothing technique is commonly used. One of the key assumptions of the freight truck only network assignment is that the long-haul trips of large trucks are dominant in freight flows, and the congestion on the specific segment of highway cannot play an important role in route choice for the long-haul trip. Although all-or-nothing technique loads too many vehicles than capacity and produces high travel time, it can be used for long-haul trip. In some cases, an additional step for adjusting speed is added to respond to the congestion. An equilibrium technique is applied for the multiclass network analysis because the passenger vehicles and the freight trucks share highway. Multiclass assignment assumes that the truck drivers are familiar with the alternatives and can change the route in congested conditions. In addition, the truck volumes are converted into the passenger car equivalents (PCE). PCE indicates the occupied spaces on the highway and the performance factor such as acceleration, deceleration, and braking times of each mode which can significantly affect the travel time on the network. In user-equilibrium method, all the vehicles travelling the same OD should experience the same travel time. All-or-nothing assignment could be used for the first loading, then iteration based on recalculated link impedance should be implemented to obtain the same travel time.

### 4.5.1 Case study

#### 4.5.1.1 Indiana

All or nothing technique is used for assignment. To overcome the all-or-nothing's limitation of over-assignment on the minimum cost path, adjustment procedure is added to calculate the new speed.

$$New\ Speed\ =\ Old\ Speed\ +\ (2\times\ (65-Old\ Speed)^{\frac{1}{2}})$$

The rail assignment is applied with the new variables defined as the index of spatial separation not only to minimize the distance but also maximize the usage of rail-line.

$$I\ =\ L(\frac{1}{D\ +\ 1})$$

Where
I = the index of spatial separation
L = the length of the line segment of the network;
D = the traffic density of the line in millions of gross ton-miles-per year

#### 4.5.1.2 Florida

While Florida statewide freight model is a multimodal commodity 4 step model, the assignment step is only performed on the highway network. The assignment is based on the all-or-nothing technique using the free flow path, and trucks are assigned prior to loading the passenger vehicles. This is because trucks are not significantly affected by the congestion, unlike the passenger vehicles.

# 5 Model Validation

The reliability of future forecasts is considered in model validation. It includes validation of the results of the model using data sets that are not used in model establishment. Freight models need to be examined for the accuracy of the forecasts while maintaining the balance between the accuracy and economic feasibility.

In the model calibration, the parameters estimated in the regression models in the trip generation step and impedance factors used in distribution will be estimated using the socioeconomic data. For mode choice step, the parameter estimates in Logit model are calibrated with the socioeconomic data and commodity flow dataset. For example, calibration processes examine whether the total flows in the gravity model were equal to actual traffic production or generated flows by the model were equal to actual flow reported in the commodity census.

In model validation estimated and observed values are matched to determine the reasonableness of the model. For example, comparison between the observed and estimated truck trip lengths in trip distribution validation is one of the standard methods in model validation. Vehicle classification counts and estimated truck traffic volumes on highway are compared in assignment validation step.

## 5.1 Case study

### 5.1.1 Indiana

The 1993 CFS data is the base validation data for commodity flow tables of every zone, and InDOT truck count is used to validate the truck volumes generated by the model. Rail flows were validated by visual examination in terms of reasonableness due to the limitation in data availability.

### 5.1.2 Florida

Truck count data from 1999 AADT Report for Florida and Truck Weight study data for the U.S. is used for validation. Since these counts include all trucks including non-freight truck, percentage of freight truck from the FAF's loaded highway network is used to estimate the freight truck trips.

# 6 4-step Model Limitations

## 6.1 Data limitations

One of the important issues when developing a 4-step commodity model is data availability at the FAZ level. Data is usually available at the state, or at best at the county level. For example, the truck payload factors obtained from VIUS are at the state level. The payload factors available at this database are therefore the average of all trucks across the state. Usually adjustments based on locally available data, such as WIM, need to be made to the state-level data before they can be used at the FAZ level. Conducting local surveys to obtain data is another way to tackle this problem. SCAG, for instance, used intercept-based cordon surveys to obtain local data.

Another place where data availability could be an issue is the commodity flows inside a state (or county). This can be a problem when both the origin and destination of a FAZ are located inside a state (or county). To obtain these commodity flows, the intra-state (county) flows should be captured. The truck model for Seattle, for instance, uses a land use-based trip rate method to generate truck trips.

Another important data-related limitation is the validation of the available data. There is usually not much data available to validate the trip tables by TRANSEARCH or CFS. This accuracy can be confirmed by comparing data from different data sources to make sure that the accuracy of the data is within an acceptable range. Carrying out vehicle counts or cordon surveys is another way of checking the accuracy of the data.

Another data-related issue is that the forecast year data is required for various socio economic variables, and/or import/export commodity flows. The accuracy of these estimates would impact the accuracy of the model.

## 6.2 Modeling limitations

One important limitation of the 4-step commodity models is neglecting the relations between suppliers and manufacturers and relying on cost only to perform trip distribution.

Another limitation of the commodity models is on their inability to handle empty trips. Although some practitioners have tried to deal with this problem by considering empty trips as a separate commodity, doing so is inconsistent with the concept of commodity models that assume the logistics of commodity movement as the initiator of a trip.

Some theoretical methods have been developed to deal with this problem, such as the one by Hautzinger (1984) that suggests using probabilistic models to calculate the probability of truck backhauling as a function of variables such as commodity volumes, backhauling costs, truck capacity, and distance. However, such models have not yet been used in practice to the knowledge of the writers.

In model validation, there is one limitation in terms of data availability. Typically intra-zonal trips of trucks or non-freight flows are not included or generated in modeling. However, it is difficult to identify non commodity truck trips in traffic count data which is the major source of validation. There is few validation data available distinguish between freight trucks and service trucks because the major source for data validation is from the survey or the observation. This problem causes large variation and inconsistency between the actual count and estimated flow.

## 6.3 Other issues

There are also minor issues regarding the conversion from the annual tons to daily tons for assigning the vehicle trips. Daily tons of all commodities are converted based on the working days per year which is 312 days per year (6 working days per week) or 250 days per year (5 working days per week). However, it should be noted that certain commodities have seasonality trend in transporting or producing, and applying a fixed factor might overestimate (or underestimate) the actual flow.

# 7 Comparison of Freight Models

Vehicle-based models are not policy sensitive and are not able to reflect the changes in growth rate in commodity groups, while commodity models emphasize commodity flows as the underlying force of freight traffic, and consider the sensitivity of commodity flows to transportation system forces. Vehicle-based models, however, are appropriate models to be used in estimating short distance service vehicle trips.

The economic activity models use an economic/land use model to forecast employment or economic activity prior to the trip generation step. The rest of the steps in the economic activity models are similar to those of the commodity models. Therefore, the economic activity class of models has a component to predict the forecast year employment and economic activity, while in the commodity model class of models these data are assumed exogenous to the model.

## 8 References

Beagan, D., Fischer, M., Kuppam, A. Quick response freight manual II FHWA-HOP-08-010, Federal Highway Administration, 2007.

Black, W. R, Indiana commodity transport model, Transportation Research Center, Indiana University, 1993.

Cambridge Systematics, Inc., 2002, Technical memorandum; model specification for FISHFM (Freight Intermodal State wide Highway Freight).

Cambridge Systematics, Inc., 2008. Forecasting statewide freight toolkit, National Cooperative Highway Research Program (NCHRP) Report 606.Transportation Research Board, Washington, DC.

Citilab, A Technical Description of CUBE CARGO, The Modeling of freight; The commodity approach and associated data sources, 2009.

Hautzinger, H. The Prediction of Interregional Goods Vehicle Flows: Some New Modelling Concepts. Ninth International Symposium on Transportation and traffic Theory, VNU Science Press, 1984, pp. 375-396.

Holguín-Veras, J., Patil, G. R. A multi-commodity integrated freight, 2008.

Lu Q., Harvey, J., Le, T., Lea, J., Quinley, R., Redo, D., and Avis, J.Truck traffic analysis using Weigh-In-Motion (WIM) data in California. University of California, Berkeley, Institute of Transportation Studies Pavement Research Center, 2002.

Origin–destination synthesis model. Network and Spatial Economics, pp. 309-326.

Sivakumar, A; Bhat, C. FRACTIONAL SPLIT-DISTRIBUTION MODEL FOR STATEWIDE COMMODITY-FLOW ANALYSIS. Transportation Research Record, Issue 1790, 2002, p. 80-88

Wilbur Smith Associates in association, 1996. Multimodal freight forecasts for Wisconsin

Xiong, D., Integrating data and models for analysis of freight movements on multimodal transportation systems for Florida, Oak Ridge National Laboratory National Transportation Research Center, 2007.

## Exercises:

1- Describe two commodity classification schemes currently in use in freight data sources and the associated industrial classification schemes. Which classification scheme is in accordance with the Harmonized Schedule which is the predominant commodity classification scheme currently in use worldwide?

2- An MPO is trying to develop a 4-step commodity model to forecast freight movement in 2040. The MPO decides that the freight model needs to be divided into 10 freight analysis zones (FAZ), each FAZ being a country. Also, the MPO is using the 2-digit STCG code to group the products.

Based on the input/output analysis done for the STCG group 23 (chemical products) the MOP comes to the conclusion that the production of chemical products in each FAZ is a function of the employment at the "chemical" and "fabricated metal" industries. Also, they think that the attraction of the chemical products at a FAZ is a function of its population, and the employment by "printing", "textile mill", and "plastic and rubber" industries.

To come up with a regression model that can forecast the tonnage of chemical products "produced" in and "attracted" to each zone in 2040, they need the production and attraction of chemical products for each zone for a base year (2010), the population and the employment in he mentioned industries at each zone for both 2010 and 2040.

They obtained production and attraction of chemical products from TRANEARCH for 2010. The population for 2010 is obtained from Census, and the employment for NIACS codes of 332 (chemical), 325 (fabricated metal), 324 (printing), 314(textile mill), and 326 (plastic and rubber) is obtained from CBP. All these data along with the prediction of population and employment in each sector in 2040 is provided in the table below. The production and attractions are in 1000 tons, employment is in 100 person, and population is in 10,000 person.

a. Using a linear regression model, forecast the productions and attractions for the chemical products for each zone in year 2040. Use statistics such as R-square and standard errors to predict how reliable your forecast is.
b. Explain, based on the results, that how good your regression model is? How good the forecasted productions and attractions would be? Is a well-developed regression model the only important factor in obtaining reliable production and attraction forecasts?
c. One proposal by the developers of the freight forecasting model was that the intercept in the production and attraction regression models be set to zero. Do you think this is a good proposal, and why?

| Zone | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Production (2010) | 710 | 450 | 650 | 400 | 750 | 610 | 350 | 670 | 700 | 500 |
| Attraction (2010) | 850 | 550 | 600 | 615 | 900 | 550 | 500 | 750 | 638 | 570 |
| 332 emp (2010) | 245 | 140 | 220 | 190 | 249 | 200 | 60 | 221 | 185 | 170 |
| 325 emp (2010) | 95 | 50 | 90 | 75 | 101 | 85 | 40 | 85 | 80 | 70 |
| 324 emp (2010) | 61 | 40 | 51 | 70 | 60 | 50 | 35 | 50 | 40 | 40 |
| 314 emp (2010) | 343 | 215 | 280 | 120 | 365 | 270 | 202 | 350 | 260 | 235 |
| 326 emp (2010) | 800 | 510 | 560 | 678 | 990 | 725 | 550 | 830 | 260 | 545 |
| Population (2010) | 105 | 120 | 160 | 30 | 150 | 50 | 100 | 80 | 80 | 90 |
| 332 emp (2010) | 1457 | 949 | 1030 | 1044.5 | 1524 | 942 | 848 | 1299 | 1093.6 | 997 |
| 325 emp (2040) | 406.5 | 243 | 401 | 328 | 429.3 | 336 | 97 | 378.7 | 339.5 | 304 |
| 324 emp (2040) | 160.5 | 93 | 180 | 152.5 | 197.7 | 141.5 | 80 | 141.5 | 165 | 144 |
| 314 emp (2040) | 97.7 | 68 | 95.7 | 114 | 114 | 114 | 70.5 | 77 | 79 | 83 |
| 326 emp (2040) | 593.1 | 382.5 | 474 | 219 | 645.5 | 470 | 365.4 | 614 | 470 | 418.5 |
| Population (2040) | 1359 | 878 | 954 | 1149.6 | 1713 | 1242.5 | 957 | 1401 | 437 | 920.5 |

3- Describe two methods that can be used to distribute productions and attractions to the OD trip tables. Which method do you recommend?

4- According to Florida mode choice component, incremental Logit model is represented as follow:

$$S'_i = \frac{S_i * \exp(\triangle U_i)}{\sum_{j=1}^{J} S_j * \exp(\triangle U_j)}$$

Where:

$S'_i$ = new share of mode i

$S_i$ = original share of mode i

$\triangle U_i$ = utility of mode i in the choice set J (j=1,2,3……J)

= Modal constant + $b^v$ * $\triangle$ Explanatory variable $i^v$

$b^v$ = coefficient for explanatory variable

The table is shown the original share of modes and coefficient of utility function by commodity group. The utility function includes 3 explanatory variables of travel time, travel cost, and commodity value per ton in the form of linear function. Freight travel time and cost is the function of distance.

| Commodity Group | Original share of mode | coefficient of utility | | |
|---|---|---|---|---|
| | | time | cost | value per ton |
| Agricultural | truck(Pt) : 40%<br>rail(Pr) : 30%<br>water(Pw) : 30% | -0.5 | -0.2 | -0.10 |
| Minerals/Coal | truck(Pt) : 20%<br>rail(Pr) : 50%<br>water(Pw) : 30% | -1.00 | -0.1 | -0.01 |
| Food Products | truck(Pt) : 70%<br>rail(Pr) : 20%<br>water(Pw) : 10% | -0.8 | -0.4 | -0.20 |

a. What is the utility function of the agricultural group?
b. Calculate the new share of modes if the current truck time of 1min increases 1.5 min (per mile) and rail cost of 20 decreases by 18 (per mile)?

5- Describe the payload factor.

6- Find one example of the model validation.

7- Compare commodity-based models with truck-based models

8- What are data related limitations of the 4-step commodity models?

# Solution:

## Question 1:

1. Standard Transportation Commodity Classification (STCC)

STCC code was developed in 1960s by American association of railroads. This commodity classification scheme is the reporting system is the STB's carload waybill sample, CFS prior to 1997, TRANSEARCH commodity flow, VIUS, and FAF databases. STCC is compatible with the Standard Industrial Classification (SIC) system.

2. Standard Classification of Transportation Goods (SCTG)

SCTG is a commodity classification scheme developed jointly by U.S. and Canadian government. This classification system is based on the harmonized schedule classification of commodities, which is the predominant product coding throughout the world. This commodity classification scheme is compatible with the North American Industry Classification System (NAICS). SCTG code is used in 1997 and subsequent CFSs.

## Question 2:

a.

The linear regression models for production and attraction are as follows:

$$Production = b_0 + b_1 e_{332} + b_2 e_{325}$$

$$attraction = b_0 + b_1 e_{324} + b_2 e_{314} + b_3 e_{326} + b_4\, pop$$

Running the regression models with 2010 production and attraction tonnages as the observed values, and employment and population data as explanatory variables, the coefficients for the production and attraction regression models, as well as R-square statistics and standard errors are obtained and tabulated below:

| Production Model | | |
|---|---|---|
| | coefficients | se |
| $b_0$ | 71.357 | 119.412 |
| $b_1$ | -0.417 | 1.824 |
| $b_2$ | 7.602 | 5.290 |
| R-square | | 0.76 |

| Attraction Model | | |
|---|---|---|
| | coefficients | se |
| $b_0$ | -9.873 | 145.413 |
| $b_1$ | 6.811 | 3.051 |
| $b_2$ | 1.244 | 0.454 |
| $b_3$ | -0.037 | 0.195 |
| $b_4$ | 0.197 | 0.715 |
| R-square | | 0.84 |

Based on the forecast employment and population for 2040, the productions and attractions for 2040 are calculated using the regression model developed above, and the results are tabulated below:

| year 2040 | | |
|---|---|---|
| Zone | production | attraction |
| 1 | 2553.412 | 1450.627 |
| 2 | 1522.525 | 866.970 |
| 3 | 2689.838 | 1505.575 |
| 4 | 2128.836 | 1389.012 |
| 5 | 2698.772 | 1792.074 |
| 6 | 2232.438 | 1323.107 |
| 7 | 454.786 | 797.734 |
| 8 | 2408.027 | 1302.962 |
| 9 | 2195.764 | 1280.933 |
| 10 | 1966.214 | 1240.033 |

b. The R-square statistics are relatively good, so the regression model is well developed. Other important factors affecting the accuracy of the forecasted productions and attractions would include correct forecast of the employment and population data for year 2040. It is also assumed here that the relationship that exists between employment and population and productions and attractions in 2010, will continue to be correct in the year 2040. This assumption is a big assumption, since changes like emergence of new technology, automation of factories, etc., can change this relationship.

c. Yes! It is logical to assume that once there are no resources, the production would be zero. Similarly, lack of consumption in a FAZ will lead to zero attraction for that zone.

**Question 3:**

Gravity models are usually used for trip distribution. To create the trip tables by a gravity model, total productions and attractions by zone (outputs of the trip generation step), and relative impedances between zones in the form of friction factors are required. Friction factors are measures of zonal desirability, and are usually expressed in form of travel time or transport cost. Another approach to do the trip distribution is using fractional split models, which have shown to give better results compared to gravity models (Sivakumar and Bhat). A fractional split distribution model estimates the fraction of commodity at the destination zone from all the origin zones.

**Question 4:**

a.

$$U_a = -0.5\, time - 0.02 \cos t - 0.10\, value\ \ per\ ton$$

b.

[Agricultural]

$$U_t - U^0{}_t = -0.5 * (1.5 - 1.0) = -0.25$$
$$U_r - U^0{}_r = -0.2 * (18 - 20) = 0.4$$

$$P_t = \frac{0.4 \exp(-0.25)}{0.4 \exp(-0.25) + 0.3 \exp(0.4) + 0.3} = 29.41\%$$

$$P_t = \frac{0.3 \exp(0.4)}{0.4 \exp(-0.25) + 0.3 \exp(0.4) + 0.3} = 42.25\%$$

$$P_w = 1 - 0.2941 - 0.4225 = 28.34\%$$

[Minerals/Coal]

$$U_t - U^0{}_t = -1.00 * (1.5 - 1.0) = -0.5$$
$$U_r - U^0{}_r = -0.1 * (18 - 20) = 0.2$$

$$P_t = \frac{0.2 \exp(-0.5)}{0.2 \exp(-0.5) + 0.5 \exp(0.2) + 0.3} = 11.75\%$$

$$P_t = \frac{0.5 \exp(0.2)}{0.2 \exp(-0.5) + 0.5 \exp(0.2) + 0.3} = 59.18\%$$

$$P_w = 1 - 0.1175 - 0.5918 = 29.07\%$$

[Food products]

$$U_t - U^0{}_t = -0.8*(1.5-1.00) = -0.4$$
$$U_r - U^0{}_r = -0.4*(18-20) = 0.8$$

$$P_t = \frac{0.7\exp(-0.4)}{0.7\exp(-0.4)+0.2\exp(0.8)+0.1} = 46.26\%$$

$$P_t = \frac{0.2\exp(0.8)}{0.7\exp(-0.4)+0.2\exp(0.8)+0.1} = 43.88\%$$

$$P_w = 1-0.4626-0.4388 = 9.86\%$$

| Commodity Group | Original share of mode | New share of mode |
|---|---|---|
| Agricultural | truck(Pt) : 40%<br>rail(Pr) : 30%<br>water(Pw) : 30% | truck(Pt) : 29.41%<br>rail(Pr) : 42.25%<br>water(Pw) : 28.34% |
| Minerals/Coal | truck(Pt) : 20%<br>rail(Pr) : 50%<br>water(Pw) : 30% | truck(Pt) : 11.75%<br>rail(Pr) : 59.18%<br>water(Pw) : 29.07% |
| Food Products | truck(Pt) : 70%<br>rail(Pr) : 20%<br>water(Pw) : 10% | truck(Pt) : 46.26%<br>rail(Pr) : 43.88%<br>water(Pw) : 9.86% |

**Question 5:**

To assign the commodity flow on the network, the payload factor is required to convert the tonnage to vehicle trips. It is defined as the average tonnage of freight carried by commodity groups. The key feature of the payload factor is that payload factor depends on the commodity group because each commodity has different handling characteristics and weights. Most commonly used data sources are VIUS from U.S. Census Bureau or carrier survey.

**Question 6:**

In Validation part, estimated and observed values are matched. Vehicle classification counts and the estimated truck traffic volumes on highway are compared in assignment validation step. AADT, or truck count data from survey or WIM data can be used as data sources.

**Question 7:**

Vehicle-based models are not policy sensitive and are not able to reflect the changes in growth rate in commodity groups, while commodity models emphasize commodity flows as the underlying force of freight traffic, and consider the sensitivity of commodity flows to transportation system forces. Vehicle-based models, however, are appropriate models to be used in estimating short distance service vehicle trips.

**Question 8:**

One of the important issues when developing a 4-step commodity model is data availability at the FAZ level. Data is usually available at the state, or at best at the county level. For example, the truck payload factors obtained from VIUS are at the state level. The payload factors available at this database are therefore the average of all trucks across the state. Usually adjustments based on locally available data, such as WIM, need to be made to the state-level data before they can be used at the FAZ level. Conducting local surveys to obtain data is another way to tackle this problem. SCAG, for instance, used intercept-based cordon surveys to obtain local data.

Another place where data availability could be an issue is the commodity flows inside a state (or county). This can be a problem when both the origin and destination of a FAZ are located inside a state (or county). To obtain these commodity flows, the intra-state (county) flows should be captured. The truck model for Seattle, for instance, uses a land use-based trip rate method to generate truck trips.

Another important data-related limitation is the validation of the available data. There is usually not much data available to validate the trip tables by TRANSEARCH or CFS. This accuracy can be confirmed by comparing data from different data sources to make sure that the accuracy of the data is within an acceptable range. Carrying out vehicle counts or cordon surveys is another way of checking the accuracy of the data.

Another data-related issue is that the forecast year data is required for various socio economic variables, and/or import/export commodity flows. The accuracy of these estimates would impact the accuracy of the model.