# Using a Federal Database and New Factors for Disaggregation of Freight to a Local Level

**Michael Anderson[1], Gregory Harris[2], Sandeep Jeereddy[3], Sampson Gholston[4], James Swain[5], Niles Schoening[6]**

**ABSTRACT.** Transport professionals at the local level often have difficulty incorporating freight into transport models and plans because freight data is proprietary at local levels requiring extensive aggregation to national levels before being released to the public. Understanding freight activity and factors affecting freight activity are extremely important for modeling infrastructure supply to transport demand and for assessing potential investment and operational strategies. This paper presents research into a national freight origin/destination database and attempts to develop disaggregation techniques using a collection of local factors: population, employment, personal income, and value of shipments. A case study of the disaggregation is performed using the Federal Analysis Framework Version 2 Database, which is a national freight database for the United States and attempts to disaggregate the data for use in a statewide or local transport model. A case study is presented that addresses the disaggregation for Alabama, comprising two zones at the national level into 67 counties at the state level. The case study uses Cube/TRANPLAN to model

[1] Associate Professor of Civil Engineering, Department of Civil and Environmental Engineering, The University of Alabama in Huntsville, Huntsville, AL, USA. email: mikea@cee.uah.edu
[2] Director of Alabama Technology Network. The University of Alabama in Huntsville, Huntsville, AL, USA. email: harrisg@uah.edu
[3] Graduate Student, Department of Industrial Systems Engineering and Engineering Management, The University of Alabama in Huntsville, Huntsville, AL, USA. email: jeereds@email.uah.edu
[4] Associate Professor of Industrial Systems Engineering and Engineering Management, Department of Industrial Systems Engineering and Engineering Management, The University of Alabama in Huntsville, Huntsville, AL, USA. email: gholston@ise.uah.edu
[5] Professor and Chair of Industrial Systems Engineering and Engineering Management, Department of Industrial Systems Engineering and Engineering Management, The University of Alabama in Huntsville, Huntsville, AL, USA. email: swain@ise.uah.edu
[6] Professor, Department of Economics and Information System, The University of Alabama in Huntsville, Huntsville, AL, USA. email: schoenn@uah.edu

disaggregrated freight data on a statewide network. The results of the cast study indicate that personal income and value of shipment provide slightly better results than using population and employment.

## INTRODUCTION

Transport professionals at the local level often have difficulty incorporating freight into transport models and plans because freight data is proprietary at local levels, requiring extensive aggregation to national levels before being released to the public. Understanding freight activity and factors affecting freight activity are extremely important for modeling infrastructure supply to transport demand and for assessing potential investment and operational strategies. In the United States, many national freight databases aggregate information to the individual states, or major communities in the states. For example, the Freight Analysis Framework, Version 2 Database (FAF2) developed and distributed by the Federal Highway Administration (FHWA) contains freight flows for 114 zones at the national level, as shown in Figure 1 (FAF2 2007). The benefit of using the FAF2 database for transport analysis of freight is related to the inclusion of future freight forecasts within the database. Currently, the database includes freight flow data for base-year, 2002, as well as forecasts for 2010 through 2035 in 5-year increments (FAF2 2007).

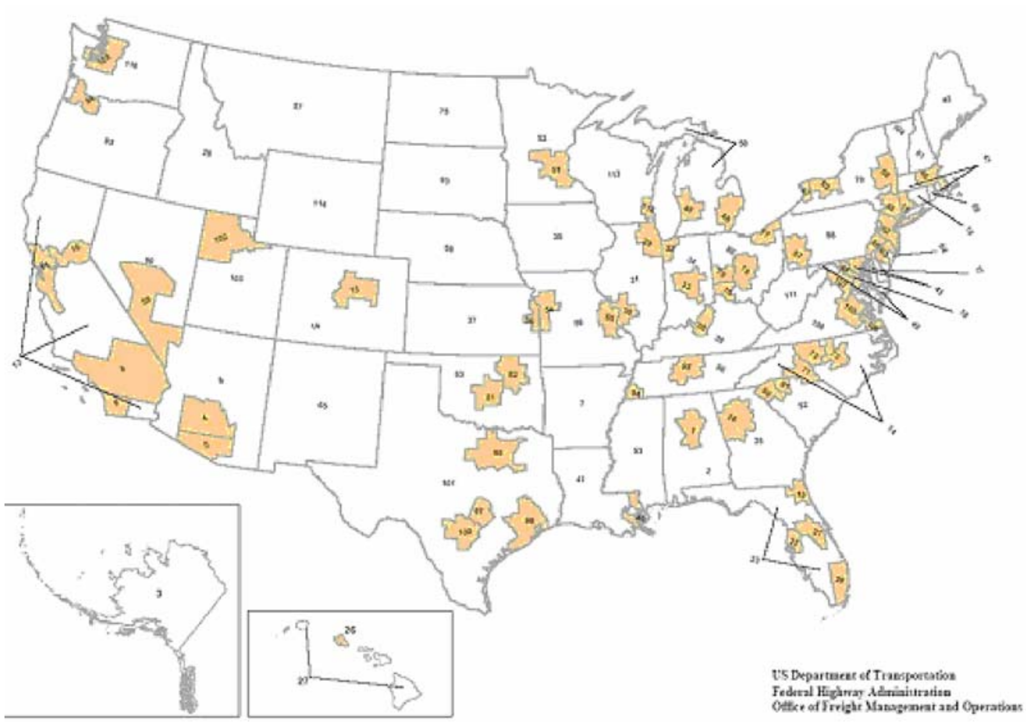


Figure 1. Geographic locations for FAF2 data (FAF2 Areas 2007)

The application of freight data to the local level is challenging due to the high level of aggregation placed on the data. Often, the disaggregation freight from national levels for use in local areas has been based on the relative employment in the local area to the total employment in the zone. This disaggregation technique has come under scrutiny lately due to the limitation that productivity improvements allow manufacturers to produce more product, requiring more freight shipments into and out of the facilities or region, using fewer employees (UAH 2005).

This paper presents research into using a national freight origin/destination database and various socio-economic factors to perform the disaggregation. The factors considered in this work include population, employment, personal income, and value of shipments. A case study of the disaggregation is performed using the FAF2 applied to a statewide transport model. A case study is presented that addresses the disaggregation for Alabama, comprising only 2 zones at the national level and disaggregating into 67 individual counties at the state level. The case study uses a CUBE/TRANPLAN to model disaggregrated freight data on a statewide network with a variety of weighting factors placed on the four socio-economic data elements. The objective of this paper is to define the importance level for each disaggregating factor, so that a better forecast is achieved by the modeling software.

**FAF2 DATA**

The FAF2 database is a continuation of the original Freight Analysis Framework developed by the U.S. Department of Transportation, FHWA. Whereas the original FAF provided the public with generalized freight movement and highway congestion maps without disclosing the underlying data, FAF2 provides commodity flow origin-destination (O-D) data and freight movement data on all highways within the FAF2 highway network. The O-D data covers both the base year (2002) and future years between 2010 and 2035 in 5-year intervals (FAF2 2007).

The FAF2 database contains a 114 X 114 origin/destination value for tonnage and value of shipment, identified for six unique transport modes and 42 individual commodities identified using the Standard Classification for Transported Goods (SCTG) (FAF2 2007). As stated previously, there are two identified zones for Alabama in the FAF2 database. The disaggregation of this data is not merely a reduction of data; there is a process of defining the data into nine unique trips purposes.

- Internal-Internal for Zone 1 and Zone 2. The internal trips for the individual zones are defined as the total trips that are both produced and attracted in the zone of interest. These trips are disaggregated into production and attraction values for the individual zones using the socio-economic factors.
  - Internal to Zone 1
  - Internal to Zone 2
- Values exchanged between Zone 1 and Zone 2. The freight values produced in one Alabama zone and attracted to the other Alabama zone are handled by applying the disaggregation factors to both the counties as a function of the total trips produced or attracted.
  - From Zone 1 to Zone 2
  - From Zone 2 to Zone 1
- Values exchanged between Alabama and the Remainder of the Country. The freight values are disaggregated through the use of the socio-economic factors for Alabama counties.
  - From Zone 1 to locations outside Alabama
  - From Zone 2 to locations outside Alabama
  - From outside Alabama to Zone 1
  - From outside Alabama to Zone 2
- Alabama pass through. The final purpose is those freight values that neither originate nor terminate in Alabama, but travel on Alabama roadways because of the location of Alabama. These trips are defined using the following relationship:

$$FAF2(ee) = [\ FAF2 - FAF2\ (origin\ AL) - FAF2\ (to\ AL) - FAF2\ (not\ AL)] \quad (1)$$

Where:

FAF2(ee) = pass through on Alabama Roadways
FAF2 = entire database
FAF2 (origin AL) = values originating in Alabama
FAF2 (to AL) = values terminating in Alabama
FAF (not AL) = values that do not travel through Alabama.

- Pass through values

**MODELING TOOLS**

A travel demand model network was developed in CUBE/TRANPLAN and used to assign the trips obtained from the FAF2 database. The model contains all Interstates, U.S. Highways and many Alabama Highways totaling nearly 5,000 miles of roadway in the state. The roadways are attributed with posted speed limits and capacities, using approved ALDOT capacities for travel modeling purposes, see Figure 2. As mentioned, the model contains 67 internal zones, representing each county in Alabama and has 15 external roadways connecting Alabama with the remainder of the nation. The counties are also shown in figure 2. A gravity distribution model has been incorporated to distribute the trips between the counties using the nine trip purposed previously described. The assignment is performed using an all-or-nothing assignment as the assumption is made that freight will not deviate from the shortest path because there is not necessarily knowledge regarding shortest path alternative when assigning trips for potential out-of-town shippers.

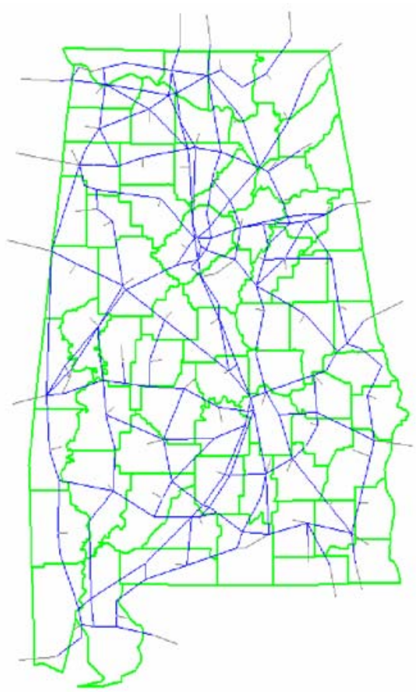

Figure 2. Modeling Network.

**EXPERIMENTAL PROCEDURE**

The entire procedure for this analysis is to analyze how much each factor is contributing to the input of the modeling software so that a better modeled freight flow can be deduced. This can be achieved by varying the input data of the software and then see if this had an impact over the final output. For a better understanding, the procedure is displayed as a Figure 3.

Figure 3. Experimental Procedure

As listed in the Figure 3, the methodology for this research consists of three major tasks, namely generating the input (INPUT), running the modeling software (PROCESS) and analyzing the output (OUTPUT). Since generating the input was thought to be the most crucial step in this process, it is discussed below in detail followed by the succeeding two tasks.

The input which is accepted and required for the modeling software is the number of freight carrying trucks visiting each county $PA_i$. Here, productions are referred as the number of trucks going out and attractions are the number of trucks coming inside each county. These productions and attractions are a function of initially assumed factors that were to affect the freight traffic.

Zonal truck counts in AL and the factor amounts for each county that are population, employment, personal income and value of shipment are available. Based on the data available, the equation below was used to disaggregate the zonal truck counts to county level.

The equation used for generating the productions and attractions (truck counts) for each county:

$$PA_i = (NFD_{ab}) * \frac{(WF) * Factor_i}{\sum Factor_{ij}} \qquad (2)$$

Where,

$PA_i$ = Truck passing County i
$NFD_{ab}$ = Truck Counts from Zone-a to Zone-b taken from the National Freight Flow
WF = Weight of the factor (or) importance of the factor (or) proportion of the factor considered for disaggregating
$Factor_i$ = Factor level for county i
$\sum Factor_{ij}$ = Total Factor level for the corresponding Zone of county i
i= county number (1, 2, 3, 4……67)
j= Zone number (1, 2)

When the factors population, personal income, employment, and value of shipment were substituted in place of '$Factor_i$' in the above equation, it is of the form:

$$PA_i = (NFD) * \left[ \frac{W_1 * P_i}{\sum P_j} + \frac{W_2 * PI_i}{\sum I_j} + \frac{W_3 * E_i}{\sum E_j} + \frac{W_4 * VOS_i}{\sum VOS_j} \right] \quad (3)$$

P = population, PI = Personal Income, E = Employment, VOS = Value of shipment

$W_1$, $W_2$, $W_3$, $W_4$ are the weights or contribution levels of population, personal income, employment and value of shipment respectively in calculating the county level truck counts (input). The amount of each factor used for disaggregating the National Freight Flow data is given by the above weights. This can be better explained by an example below.

For example, if the contribution of each factor is considered to be the same in calculation of truck counts, then $W_1=W_2= W_3=W_4=0.25$.

This equation aids us in disaggregating or distributing the zonal truck counts from the National Freight data to the county level. Therefore, the total number of trucks before and after disaggregating must be the same.

$$\sum PA_i = \sum NFD_{ab} \quad (4)$$

For satisfying criterion, there are two constraints in the equation involving $W_1$, $W_2$, $W_3$ and $W_4$

1. $\sum_{i=1}^{4} W_i = 1$ (5)

2. $W_i = Range(0,1)$ (6)

These levels sum to 1 because if $\sum W_i > 1$, the total number of modeled trucks would exceed the total actual trucks. This would add more trucks to the model than are actually present. For example, if $\sum W_i = 2$, the modeling software would forecast double the amount of actual total freight traffic inside Alabama. Therefore, $\sum W_i = 1$ and range of $W_i = (0, 1)$. By assigning a number within the range (0, 1) to these weights, we are actually choosing the contribution level or the importance of each factor in generating the input, which is then entered into the modeling software.

The next task after generating the input (disaggregated zonal truck counts) is to enter the data into the modeling software and extract the output. Deducing the output and to format it into a usable form is the next step in this experiment.

The output is the freight truck traffic generated on Alabama roadways. This is displayed in the form of an excel file containing various roadways numbered from 1 to 383, the total number of roadways modeled in the software. The assignment of the forecasted truck counts for each roadway is contingent to the input $PA_i$ entered in the modeling software. On varying the input to the model, truck counts assigned to all the roadways change randomly thus making it more difficult to measure the impact or variation on output with the variation in input.

One way to measure the impact on output of this model is to measure the deviation or difference of each data point with respect to the actual counts. The yard stick in this case is the actual truck traffic in the Alabama network provided by the Alabama Department of Transportation (ALDOT). Closer the actual counts are to the modeled values, the better is our forecast and thus the factor contributions. Minimizing the difference between actual counts and modeled values can be achieved by varying the factor contribution in disaggregating the zonal truck counts. By this analysis we can deduce a combination of factor contributions that aid us in forecasting the truck counts. The remainder of this paper is devoted to analyzing the deviation of output (modeled truck counts) from the counts by varying the factor contributions.

## ANALYSIS

### Metrics

Three metrics were identified that could give a measure of accuracy of the forecast. Brief description of all the metrics is give below:

Root Mean Square Error (RMSE) is a common measure of the variability in the error (difference between model and actual counts) of any model. As a result, this was used as one of the metric. Greater the RSME, less accurate is our model.

$$RMSE = \frac{\sqrt{\sum (Model_i - Ground_i)^2 / (NumofCounts - 1) * (100)}}{\sum Ground_i / NumofCounts}, \text{ (Monsere 2001)} \qquad (7)$$

Where:
RMSE = root mean square error
$Model_i$ = Modeled Value for the roadway i
$Ground_i$ = Actual Counts for the roadway i.

The next measure used in this analysis was the Nash Sutcliffe's (NS) coefficient which can range from -∞ to 1. An efficiency of 1 ($E$=1) corresponds to a perfect match of forecasted counts to the actual counts. An efficiency of 0 ($E$=0) indicates that the forecasted values are as accurate as the mean of the actual counts, whereas an efficiency less than zero ($-\infty<E<0$) occurs when the forecasted mean is less than the actual values. In other words, this coefficient gives us a measure of scatter variation from the 1:1 slope line of modeled truck counts Vs the actual counts. More the deviation of points from the slope line, lesser will be the coefficient. Greater the NS-value better is our forecast. It can be calculated using the formula:

$$\text{NS-Coefficient} = 1 - \frac{\sum_1^n (ModeledCounts - GroundCounts)^2}{\sum_1^n (GroundCounts - MeanGoundCounts)^2} \text{ (Monsere 2001)} \qquad (8)$$

The Nash Sutcliffe's statistic is considered the best measure of deviation between two data sets and used in many similar instances. In this paper, this statistic is used as a primary measure for the balance of the analysis.

Another measure used was the percent error between the forecasted and the actual counts. It has given the percentage of difference between both the data sets.

$$Percenterror = \left(\frac{Model(i) - Ground(i)}{Ground(i)}\right)(100)/N \text{ (Monsere 2001)} \qquad (9)$$

Where,
$Model_i$ = Modeled Value for the roadway i
$Ground_i$ = Actual Counts for the roadway i

N = Total number of modeled values.

**Setting up the Experiment**

Since the aim of this paper is to deduce relevant factors for disaggregating the zonal truck values, various combinations of factor contributions or factor importance levels were executed as shown in Table 1 for which all the three metrics have been calculated.

Below is a set of runs containing a combination of factor proportions ($W_i$) for each run in the experimental design which was generated from the Minitab 14.0 under the Mixture Experiments Section. The column under each factor represents the contribution level of each factor in disaggregating the zonal truck counts. Since all the weights must sum up to one, all the run totals are equal to one and no single factor exceeds this value. Note: The values under each column of P, PI, E, VOS are the corresponding Weights ($W_i$): (P=population, PI=personal income, E = employment and VOS = value of shipment)

Table 1. Set of Runs containing various factor levels

| **RUN** | **P** | **PI** | **E** | **VOS** | **NS-Value** | **RMSE** | **%Error** |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0.195821 | 105.92 | 86.44 |
| 2 | 0 | 1 | 0 | 0 | 0.197551 | 105.8 | 87.97 |
| 3 | 0 | 0 | 1 | 0 | 0.195821 | 105.92 | 86.36 |
| 4 | 0 | 0 | 0 | 1 | 0.193561 | 105.92 | 86.44 |
| 5 | 0.5 | 0.5 | 0 | 0 | 0.195821 | 105.85 | 86.39 |
| 6 | 0.5 | 0 | 0.5 | 0 | 0.196825 | 105.84 | 86.31 |
| 7 | 0.5 | 0 | 0 | 0.5 | 0.196985 | 105.86 | 86.26 |
| 8 | 0 | 0.5 | 0.5 | 0 | 0.196642 | 105.8 | 87.97 |
| 9 | 0 | 0.5 | 0 | 0.5 | 0.197551 | 105.82 | 87.12 |
| 10 | 0 | 0 | 0.5 | 0.5 | 0.197239 | 105.87 | 86.61 |
| 11 | 0.33333 | 0.33333 | 0.33333 | 0 | 0.1965 | 105.85 | 86.04 |
| 12 | 0.33333 | 0.33333 | 0 | 0.33333 | 0.196835 | 105.92 | 86.36 |
| 13 | 0.33333 | 0 | 0.33333 | 0.33333 | 0.195821 | 105.91 | 86.42 |
| 14 | 0 | 0.33333 | 0.33333 | 0.33333 | 0.195952 | 105.9 | 86.11 |
| 15 | 0.25 | 0.25 | 0.25 | 0.25 | 0.19606 | 105.89 | 86.18 |
| 16 | 0.625 | 0.125 | 0.125 | 0.125 | 0.196219 | 106.07 | 85.57 |
| 17 | 0.125 | 0.625 | 0.125 | 0.125 | 0.193561 | 106 | 85.67 |
| 18 | 0.125 | 0.125 | 0.625 | 0.125 | 0.194573 | 105.83 | 85.96 |
| 19 | 0.125 | 0.125 | 0.125 | 0.625 | 0.197182 | 105.82 | 86.53 |

An experiment in which the response is assumed to depend only on the relative proportions of the factors present is a mixture experiment. In a mixture experiment, the total amount of mixture is held constant and the value of the response changes when changes are made in the relative proportions of those ingredients making up the mixture (Cornell 1990). Analogous to this statement, the total number of trucks used for as the input is always constant even in this case. The total number of trucks used for disaggregating the zonal values is always constant and the ingredients to make up this constant value are the factor contribution levels in this case. A factorial experiment would not apply for this situation since any design would not confine to the assumptions such as dependency, and orthogonality.

A simplex mixture design experiment was chosen for this problem; since it had a greater number of runs than the simplex lattice design since a much accurate results can be achieved with more number of runs. This design covers almost all the combinations of factor contributions that are to be tested initially. It has 4 input factors and 19 runs. The first four runs gave held each of the four factors at highest importance. The next six observations took combinations of two out of four in each run, giving them a weight of zero and 0.5 to the remaining variables. Four trials were also allotted to a scheme, giving a weight of zero to each variable in turn and equal weights to the remaining four. A run was also apportioned equal weight to all the four input variables.

The three metrics were then calculated for each run after it was executed from the modeling software. Given the experimental design and the response, which is Nash-Sutcliffe's coefficient, the immediate step would be to trace a combination of factor contributions that have a positive effect over the response.

All coefficients of $W_i$ are approximately equal to 0.19 exemplifying that the effect of all the variables is almost the same and choosing any one of them for calculating the county level truck counts does not really impact the modeled output. When considering the interactions in this experiment, the magnitudes of their coefficients are close to zero, indicating a similar conclusion that none of the interactions are significant. In order to validate these inferences, it would be necessary to perform tests of hypotheses (F-test) on the model's parameters (Cornell 1990).

We chose to test the null hypothesis ($H_0$) which states:
$H_0$ = Response does not depend on the mixture components
Against the alternative hypothesis ($H_1$) which states:
$H_A$ = Response does depend upon the mixture components

When the null hypothesis is true, all four linear coefficients $\beta_1$, $\beta_2$, $\beta_3$ and $\beta_4$ are equal to some constant value say ($\beta_1$) and the remaining terms in the regression equation are equal to zero. This implies that the null hypothesis is of the form:

$H_0$: $\beta_1=\beta_2=\beta_3=\beta_1=\beta_0$ and $\beta_p\beta_q = \beta_p\ \beta_q\ \beta_r = \beta_p\ \beta_q\ \beta_r\ \beta_s$ (Where each of p, q, r, s take values 1, 2, 3, 4)

An F-test was employed to test the hypothesis and Minitab® yielded a p-value of 0.279 which is greater than the 5% significance level. The null hypothesis could not be rejected in this case according to 5% significance. As a result, no matter whatever the importance level we assign to each of the factors, population, personal income, employment and value of shipment, it is not varying modeled truck traffic inside the Alabama network.

When the scatter plots were graphed between the modeled trucks counts versus the actual counts from the ALDOT, for all the runs, there was not much of a difference in the scatter pattern for all the runs. Even this shows that there is not much impact by varying the factor levels for generating the input. In addition, the RSME (root mean square error) and the percent error showed constant results, giving very less scope for variation.

This could have been the conclusion for this investigation whether the population, personal income, employment and value of shipment of each county effected the disaggregation of zonal truck counts, and thus the modeled truck traffic. However, when observed carefully, there was not a single combination of factor contributions that yielded a NS-value close to 1, which means that regardless of which combination of factors were considered, the modeled truck traffic is not matching with the actual counts (actual freight carrying truck counts provided by ALDOT). If the modeling software was built appropriately, there must have been some point where the modeled truck counts were close to the actual counts. As a result, speculation was aroused as to whether some of the factors did really have an impact over the freight flow.

Before it could be concluded that none of the initially assumed factors influenced the freight flow in Alabama, there is one plausible argument which negates the previous inference that none of the factors help in disaggregating the zonal truck counts and the modeled truck traffic. All the above runs were executed for a freight transfer of 30 tons. The truck capacity is an attribute in the modeling software that is changeable. Until this point it was to disaggregate the trucks to the Alabama network which had an assumed freight carrying capacity of 30 tons. This is an attribute setting in the forecasting software where we can vary the tonnage from 0 to 30 tons with a fixed interval of 5 tons. As a subsequent step to the initial analysis, the rest of the paper presents a similar methodology to test whether the same inferences hold good even for a different capacity of the trucks. With the help of this subsequent analysis, we can bolster our conclusion regarding the factors helping the desegregation of truck counts and have an idea about the results of modeling software.

Since the varying of coefficients had a limited impact on the final truck counts for the 30 tons/vehicle, one combination of the coefficients was predefined and a set of runs were carried by varying the tonnage of the trucks. Below is the graph of how the Nash Sutcliffe's coefficient varied with the change in tonnage. In other words, as an initial step for the second phase of analysis, the Nash-Sutcliffe's coefficient was calculated for all the truck capacities for the same combination of factors levels. Figure 4 shows the variation of this coefficient with the tons/vehicle.

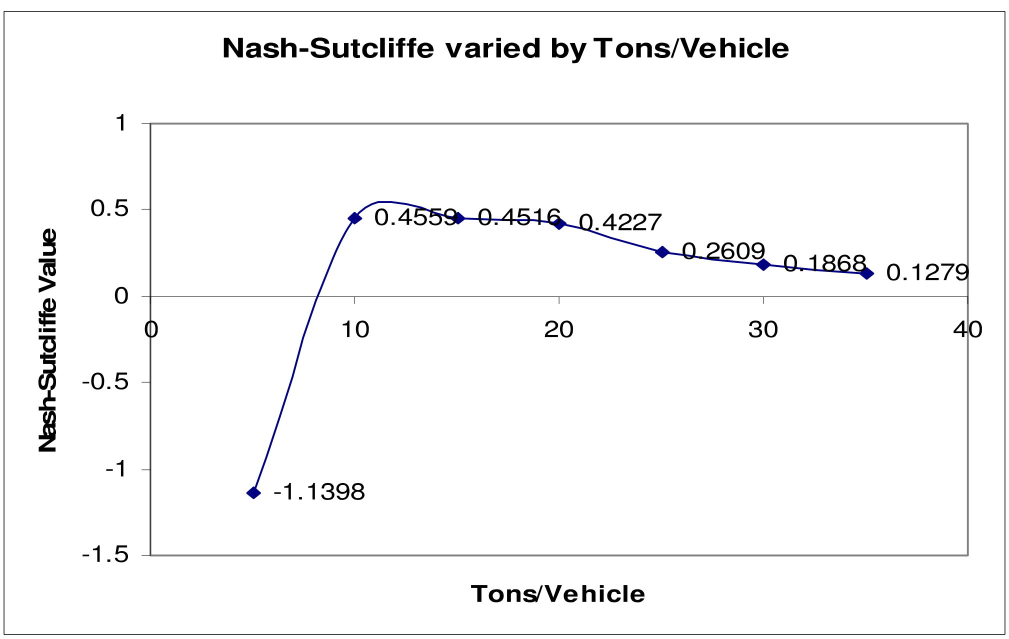


Figure 4. Nash Sutcliffe's values for various tonnages

From the above graph, it is evident that this macro level change in the software impacted the network and a highest value was recorded for the 10 tons. When trucks with a capacity of 10 tons/vehicle were used for the modeling network, it yielded the truck counts closest to the actual truck counts (actual counts) provided by the Alabama Department of Transportation (ALDOT).

A similar micro level analysis which was done for the 30 tons/vehicle was again performed for the 10 tons/vehicle model and analyzed if disaggregating based on county level factors impacted the response.

**Second Phase of Analysis**

Even in this case, a similar design was setup as that of the initial analysis. A simplex centroid design with 19 runs was setup in the Minitab® 14.0 and a regression equation for this design was setup, see Table 2. From this regression equation, we can make some initial conclusions as to what factors really impact the modeled traffic flow. If the same conclusion as that of the previous analysis was pictured, a hypothesis test would follow to validate our inferences. The scatter plots, percent error and root mean square error for all the runs would give us a better picture of our analysis and hence bolster our conclusions.

Table 2. Run results for the 10 tons/vehicle data

| RUN | P | PI | E | VOS | NS-Value | RMSE |
|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0.4559 | 77.4425 |
| 2 | 0 | 1 | 0 | 0 | 0.460142 | 77.1463 |
| 3 | 0 | 0 | 1 | 0 | 0.462018 | 75.5991 |
| 4 | 0 | 0 | 0 | 1 | 0.47171 | 74.5071 |
| 5 | 0.5 | 0.5 | 0 | 0 | 0.459501 | 77.1895 |
| 6 | 0.5 | 0 | 0.5 | 0 | 0.472021 | 76.3102 |
| 7 | 0.5 | 0 | 0 | 0.5 | 0.469717 | 75.7614 |
| 8 | 0 | 0.5 | 0.5 | 0 | 0.463931 | 76.1724 |
| 9 | 0 | 0.5 | 0 | 0.5 | 0.481489 | 75.6334 |
| 10 | 0 | 0 | 0.5 | 0.5 | 0.469971 | 74.9775 |
| 11 | 0.33333 | 0.33333 | 0.33333 | 0 | 0.468871 | 76.5275 |
| 12 | 0.33333 | 0.33333 | 0 | 0.33333 | 0.464233 | 76.1502 |
| 13 | 0.33333 | 0 | 0.33333 | 0.33333 | 0.468771 | 75.6167 |
| 14 | 0 | 0.33333 | 0.33333 | 0.33333 | 0.469662 | 75.5556 |
| 15 | 0.25 | 0.25 | 0.25 | 0.25 | 0.467459 | 75.9246 |
| 16 | 0.625 | 0.125 | 0.125 | 0.125 | 0.467388 | 76.6304 |
| 17 | 0.125 | 0.625 | 0.125 | 0.125 | 0.469834 | 76.4629 |
| 18 | 0.125 | 0.125 | 0.625 | 0.125 | 0.469302 | 75.7198 |
| 19 | 0.125 | 0.125 | 0.125 | 0.625 | 0.46922 | 75.1712 |

The Minitab® output, it indicates that all the coefficients of $W_i$ produce NS-Values that are approximately equal to 0.45, exemplifying that the effect of all the variables is almost the same and choosing any one of them for calculating the county level truck counts does not really impact the modeled output. When considering the interactions in this experiment, their magnitudes of coefficients in the regression equation have little variation, indicating a similar conclusion that none of the interactions are significant. When a hypothesis test (F-test) was conducted for this equation with the null and alternative hypotheses as:

$H_0$ = Response does not depend on the mixture components

$H_A$ = Response does depend upon the mixture components

Minitab® yielded a P-value 0.385 which was greater than the 5% significance indicating that the null hypothesis could not be rejected with that significance level. This means that changing the factor proportions did not impact the modeled truck traffic even for the 10 tons/vehicle capacity. Even for this case, scatter plots were graphed between the modeled trucks counts versus the actual counts from the ALDOT, for all the runs and similarly, there was not much of a difference in the scatter pattern for all the runs. Even this shows that there is not much impact by varying the factor levels for generating the input. Also, the RSME (root mean square error) of all the runs resulted almost constant values and very less scope for variation.

When the residual plots (difference between the actual counts and model values versus actual counts) were graphed for both the 30 tons/vehicle runs and 10 tons/vehicle runs, even they indicated the same inferences asserted from the scatter plots. There is little difference in the residual patterns. For the 30 tons/vehicle runs, the plotted data points had an increasing trend indicating that the error was higher for larger actual values. In other words, the error in the modeled values was larger for busier roadways. When the 10 tons/vehicle residual plots were plotted, there was a different scatter when compared to original runs, but the same conclusion that changing factor proportions did not matter holds good even for this case. As a result, varying the factors in the calculation of county level truck counts did not have an impact over the modeled truck traffic in the state of Alabama.

## CONCLUSION

This paper presented a research as to what factors were the best considerations for disaggregating the national freight flow data, which can be used as input for Alabama freight flow modeling software.  After the initial analysis, it was deduced that factors that were considered for disaggregating the national freight flow data (zonal truck counts) did not impact the modeled freight flow inside AL. An attribute regarding tonnage was changed and the Nash-Sutcliffe's efficiency was calculated for different tonnages. The tonnage yielding the best Nash-Sutcliffe's efficiency was considered an additional analysis was done using the same methodology applied for the runs with 30 tons/vehicle.  The best Nash Sutcliffe's value was recorded for the attribute 10 tons/vehicle and the experiment was redone for this case. No change was observed even in this case and the same conclusion that population, personal income, employment and value of shipment did not affect the desegregation of freight flow to the county level was concluded.

In both the cases with different tonnage capacities, the Nash Sutcliffe's statistic was different but a higher value was derived for the 10 tons/vehicle runs.  This indicated that the modeled values were much closer to the actual counts meaning a better batch of runs containing the 10 tons/vehicle trucks.  By varying the factors and the attribute within software, the approximate maximum Nash-Sutcliffe Value was around 0.47.  The highest achieved values were when Value of Shipment and Personal Income were used in the analysis.  Even by adjusting so many variables, the modeled truck traffic was not close to actual traffic which is the actual counts. This indicates that none of the initially assumed factors influenced the modeled traffic flow or a there might be some speculation for error within the software itself.  If the modeling software was the problem, an immediate correction is needed and a similar analysis could be performed in the future to deduce what factors might be impacting the freight flow.  This could be the suggested future work after all the analysis and inferences.

## REFERENCES

FAF2 Internet Page. http://ops.fhwa.dot.gov/freight/freight_analysis/faf/index.htm (August 1, 2007).

FAF2 Areas. http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/cfs_faf_areas.htm (August 1, 2007)

UAH Office for Economic Development (2005), *Transportation Infrastructure In Alabama - Meeting the Needs for Economic Growth,* Final Report on the Requirements for Infrastructure and Transportation to Support the Transformation of the Alabama Economy. Prepared for the Office of the Secretary, U.S. Department of Transportation, Grant No. DTTS59-03-G-00008.

Monsere, C.M.(2001). "A GIS based multi-commodity freight model: Typology, model refinement and field validation" Dissertation for Degree in Civil Engineering, Iowa State University.

Cornell, J.A. (1990). "Experiments with Mixtures" Second Edition, John Wiley and Sons, Inc. 1990.