Final Report

Evaluation of Origin-Destination Matrix Estimation Techniques to Support Aspects of Traffic Modeling

Submitted by:
Dr. Michael Anderson, Dr. Jeffrey Wilson and Tahmina Khan
Department of Civil and Environmental Engineering
University of Alabama in Huntsville
Huntsville, AL 35899

Submitted to:
Alabama Department of Transportation
Bureau of Research and Development
1409 Coliseum Boulevard
Montgomery, AL 36110

May 5, 2014

ACKNOWLEDGEMENT

This project was sponsored by Alabama Department of Transportation.

DISCLAIMER

Table of Contents

List of Tables

List of Figures

Executive Summary/Abstract

Travel demand forecasting models are used to predict future traffic volumes to evaluate roadway improvement alternatives. Each of the metropolitan planning organizations (MPO) in Alabama maintains a travel demand model to support planning efforts. For smaller areas, non-MPOs, no travel demand models exist to support planning efforts. The goal of this project is to examine the use of a methodology that can utilize existing traffic counts to develop origin/destination matrices in an attempt to improve existing travel demand models and support the development of new travel demand models in areas that are too small to maintain their own models. Applying the methodology to develop the origin/destination matrix for a community from existing traffic counts has the potential to be used to evaluate area friction factors, be used to support the development/refinement of trip generation models and determine the level of external-internal and external-external traffic.

1 Chapter Introduction

1.1 Background

Travel demand forecasting models predict the current or future traffic volume on roadway sections that can help planners to make decisions on issues such as roadways requiring an increase in capacity and the location of new roadway infrastructure. A travel demand model utilizes the Four Step Planning Process; the first step is trip generation, in which the number of trips within a planning area is calculated by the use of a combination of socioeconomic and household survey data. The second step is trip distribution in which the trips that are generated from the trip generation algorithms are distributed using a gravity model that takes into account the number of trips produced and attracted between origin/destination pair and a factor that relates to the distance, or travel time, between the pair. The third step is mode choice analysis in which a decision is made regarding the mode of transportation used to travel between origin and destination. The mode choice step is commonly complete through the use of a logit choice model with utilities values (travel time and cost) to evaluate the alternatives. The fourth step is traffic assignment in which the automobile trips are assigned to the roadway system to generate the daily traffic expected on the roadways.

Ideally, the trips assigned to each roadway in the model should match the existing traffic count for the roadway. If so, the model is said to be validated to base conditions and can serve as an appropriate tool for future forecasting. Unfortunately, the time and resources required to collect socioeconomic, land use, household survey data or area specific friction factor data to develop a quality travel demand model frequently result in a these steps being done with limited effort as the focus is generally on the forecasting phase and the development of a plan. Finding ways to increase the speed and reduce the costs associated with developing these travel demand forecasting models would be beneficial to all communities, but especially small- and medium-sized communities where funds are limited or nonexistent and models are not always available.

The methodology examined in this study is intended to reduce the time and cost associated with developing travel models in communities by evaluating the potential to estimate an accurate origin/destination matrix from existing traffic counts and determining the further applications and benefits to build an accurate origin/destination matrix of a certain size through aggregated or non aggregated zonal systems [1].

The benefit of this research to the Alabama Department of Transportation (ALDOT) is the evaluation of a methodology that can potentially be used to improve travel demand models. The improvements to travel demand models can be made in the areas of:

- Identified friction factors by which the models can be calibrated,
- Development/refinement of trip generation factors, and
- Improved ability to model external-internal and external-external trips.

The research effort will determine if the methodology is appropriate for future application within the MPOs and by personnel in ALDOT when considering infrastructure improvement decisions in areas that lack a travel demand model.

1.2 Research main goal

The objective of this research proposal is to evaluate the effectiveness of applying a methodology to develop an origin/destination matrix from existing traffic counts.

1.3 Research Breakdown Tasks

The following tasks were proposed and done in the course of time:

- Reviewed relevant literatures,
- Implemented different scenarios for the methodology,

- Tested the methodology for several small/large communities regarding external-internal and external-external trip making where several actual datasets are available for validation,
- Evaluated friction factors for small/large MPOs using the trip distribution step of the model,
- Evaluated the time necessary for a large MPO to implement the methodology,
- Evaluated the parameters of trip generation for small communities,
- Determined the number of traffic counts required for a small community to accurately estimate an origin/destination matrix and developed a relationship between the number of traffic counts needed and the number of zones in the network, and
- Verified and compared the trip generation models by using statistical analysis.

1.4 Document Organization

This document is divided into seven chapters and can be presented as follows. The first chapter presents the general introduction. The second chapter summarizes the literature and the resources studied to provide insight on past and current practices for building travel demand forecasting models. The third chapter provides an overview of the objective and research tasks associated with the application of the methodology in two small communities namely Arab and Roanoke by estimating an origin/destination matrix from traffic counts to test external-external trip patterns. The fourth chapter provides an overview of the objective and research tasks associated with the application of the methodology in Brazos County, Texas, by estimating an origin/destination matrix from traffic counts to test external-external trip patterns and evaluate friction factors. The fifth chapter provides an overview of the objective and research tasks associated with the application of the methodology in a small community, Hartselle, by estimating an origin/destination matrix from traffic counts to test external-external trip patterns, evaluate friction factors and trip generation parameters, determine the required number of counts, and develop the relationship between number of counts and number of zones. The sixth chapter provides an overview of the objective and research tasks associated with the application of the methodology in Anniston by estimating an origin/destination matrix from traffic counts to test external-external trip patterns, evaluate friction factors and trip generation parameters through aggregated or non aggregated zonal systems. The seventh chapter lays out the model description coded in Visual C and how it can be implemented in any community. The eighth chapter demonstrates the application of the code in several cities of Alabama to estimate an origin/destination matrix using any network with the traffic counts available and investigates the output to perform the above tasks. The ninth chapter concludes with the overall outcomes of applying the methodology and provides brief recommendations.

2 Literature Review

2.1 Detail Review of Old and Recent Studies

There are currently two proposed approaches for creating an origin/destination matrix from traffic counts. They are traffic modeling based and statistical inference approaches. Some of the principles proposed by researchers within these methods require an outdated origin/destination matrix to supplement the traffic counts for target purposes while others do not [2]. The focus of this research paper is to build an origin/destination matrix by only using traffic counts because small-sized communities do not have an existing origin/destination matrix to base findings.

The method described by Ortuzar and Willumsen in Modelling Transport gives insight to creating an origin/destination matrix from traffic counts with a basis in travel modeling. They highlight entropy maximizing work by Wilson [3] which was applied by Van Zuylen and Willumsen [4] for origin/destination matrix estimation as the basis for their methodology.

Their methodology does not require an existing outdated origin/destination matrix to supplement the traffic counts. It is also static in nature due to the fact that all the traffic counts are assumed to be collected from one single time period.

Ortuzar and Willumsen view traffic counts as the combination of a trip matrix and a route choice pattern [5]. This means that traffic counts give direct information about the sum of origin/destination pairs that use a particular roadway in the network [5].

When building the model from traffic counts the most important part is to identify the roadways used by the trips from origin to destination [5]. The flows for roadways are expressed mathematically in the following equation:

$$V_a = \sum_{ij} T_{ij}\, p_{ij}^{a}, \quad 0 \le p_{ij}^{a} \le 1$$

Where the flow (Va) for link a is the summation for all the trips in that link [5]. The variable T_{ij} is the traffic count between Zone i and Zone j [5]. Also, p_{ij}^{a} is the proportion of trips from Zone i to Zone j travelling through the link [5]. Now, the proportion variable p_{ij}^{a} is determined by the type of trip assignment technique being used for a particular model [5].

When it comes to classifying assignment methods for origin/destination matrix estimation from traffic counts there are two main ways of processing the information [5]. They are proportional and non-proportional assignment [5].

In proportional assignment the proportion of drivers choosing each route are independent from the flow levels [5]. This approach has the ability to associate a proportion for each link flow for each origin/destination pair and calculate the probability that this portion of traffic comes from a particular origin and destination [6]. The proportional assignment tactic was used in the study by Van Zuylen and Willumsen [4]. The most common use of the proportional assignment is through the use of the all-or-nothing assignment method [5]. The following equation illustrates the all-or-nothing assignment proportion technique.

$$p_{ij}^{a} \begin{cases} 1 \; if\ trips\ from\ origins\ i\ to\ destinations\ j\ use\ link\ a \\ 0 \; otherwise \end{cases}$$

The variable p_{ij}^{a} is given a value of one if the origin/destination pairs use that particular roadway and traffic count, otherwise, it is simply given a value of zero [5]. Beyond the all-or-nothing assignment, there are some stochastic methods that will give values that range from zero to one to variable p_{ij}^{a}.

Non-proportional assignment may work better under congested conditions because this would allow for trips to take paths other than the shortest travel time [5]. It causes the proportion of travelers on each link to not depend on link flows [5]. It works by taking an iterative approach where a set of route choice proportions are assumed then a trip matrix is estimated [5].

This trip matrix is loaded onto the network and a new set of route choice proportions are calculated [5]. This process is repeated until the route choice proportions and trip matrix are similarly consistent [5]. The main downfall to this technique is it makes the route choice proportions interdependent to the trip matrix [5].

One way of implementing non-proportional assignment is through the use of incorporating equilibrium to the traffic flows which was proposed by Nguyen [7] and further studied by Farhangian and LeBlanc [8] and Iida et al. [9]. This approach assigns link cost functions, link travel cost, and path travel cost to the network as a way to minimize travel costs [10] [11]. The equilibrium approach requires the use of a target trip matrix to reproduce the observed traffic counts [2] [10]. These models were developed on small test networks and their applicability on a large network is not ensured [11].

The use of user-equilibrium to assign the trips through the network was proposed by Fisk [12]. It uses a model with a bi-level structure that maximizes the entropy on an upper level then solves the user-equilibrium assignment on a lower level [2]. The output of this method will work if the traffic counts are available for each roadway in the network and they are consistent with user-equilibrium [2]. This method is mostly theoretical and it requires high computing requirements which reduce any practical value [5] [10].

A recent study by Kockelman [10] used the maximum entropy method as a means to estimate the origin/destination matrix for a small subnetwork of a large city. The purpose of the study was to test the use of a newly devised linearization algorithm of the Frank-Wolfe type to estimate the origin/destination matrix by only using traffic counts [10].

The linearized algorithm was stated to be a more efficient process for computing the origin/destination matrix [10]. The methodology was tested on a subnetwork that contained 100 percent of the traffic counts and proved successful in replicating the origin/destination matrix with a root mean square error of about 11 to 13 percent [10]. The results of this study show that maximum entropy is still a viable method for estimating an origin/destination matrix.

A Path Flow Estimation technique has been proposed as a way of melding both the origin/destination matrix estimation and traffic assignment steps into one process. A study by Bell et al. [13] for congested networks uses stochastic user equilibrium to determine the paths that vehicles will take between origin and destination; then assigned traffic counts to the network links to estimate path travel times [13]. Like all stochastic models it uses a logit formulation. It calculates the path flows and travel times for each link by assigning a cost based on delay and capacity [13].

Path flow estimation was used in a study by Chen et al. [14] to estimate the number of trips between O/D pairs on a small freeway network. The study used centroid connectors as well as freeway traffic counts in estimating the flows.

In another study by Chen et al. [15] path flow estimation was used to estimate the flow of traffic in a small town that was congested with traffic from tourists. The town proposed adding a roadway to alleviate congestion on a nearby highway [15]. The study used path flow estimation to develop a forecast of link and turning movement intersection volumes for the network [15]. Level of service analysis was employed to determine lane configurations for the new roadway [15].

Path flow estimation examines delay and capacity of links and is meant to be used in congested networks. The small community analyzed in this study is located in rural Alabama and congestion is not a problem. Therefore, path flow estimation will not be used to estimate the O/D matrix.

The statistical inference techniques refer to using the methods of maximum likelihood, generalized least squares, or Bayesian approach for estimating the origin/destination matrix. All of these approaches attempt to find the origin/destination matrix based on an exchange between a target origin/destination matrix and traffic counts [10].

The maximum likelihood approach is one of the most widely used methods for statistical estimation [16]. It attempts to find the most likely value for a parameter based on the collected data set [17]. In terms of estimating an origin/destination matrix it attempts to find the true origin/destination matrix by observing a target origin/destination matrix and traffic counts [2] [11].

This method was explored by Spiess [18] who used small examples to test the practice [10] [2]. Maximum likelihood is useful with simple normal data but with complicated models it may not result in a closed solution [17]. Also, when it is used with all the information available it can produce inconsistent results [16]. Because this approach requires the use of a target origin/destination matrix it will not be used in the development of methodology for this research paper.

The generalized least squares approach is an attempt to connect random variables and parameters through a linear equation in order to find the best fit model for a given data set [19] [17]. In origin/destination matrix estimation it evaluates survey data related to a target origin/destination matrix combined with traffic counts [2] [11]. This method was explored by Cascetta [20] and Bell [21] [22]. Both studies use proportional assignment [2].

The study by Cascetta [20] used a small sample network to test the practice and found that the results had a lower mean square error when compared to maximum entropy. The study also found that the method is sensitive to variations and accuracies in the traffic counts and the target origin/destination matrix [2]. This is because the generalized least squares method is susceptible to outliers [17].

In the study by Bell [21] it was found that if there is a high number of traffic counts then the generalized least squares approach is approximately the same as the maximum entropy method [2]. The study by Bell [22] sought to improve the algorithm used to estimate the origin/destination matrix by considering a non-negativity constraint [2].

The application of the generalized least squares methodology has not been applied to a real network [11]. In general the models are difficult to solve due to high computational requirements [11]. Because this approach requires the use of a target origin/destination matrix it will not be used in the development of the methodology for this research paper.

Bayesian inference is derived from Bayes' theorem of conditional probability which is a way of revising predictions based on new evidence [17] [23]. Bayesian inference is a method of adjusting the function of a parameter to be estimated that starts with little to no information then is adjusted as new data is collected [17].

In origin/destination matrix estimation it refers to making the target origin/destination matrix as a prior probability function and combining it with the traffic counts as another source of information [2] [6]. This method is also applicable to estimating turning movements at an intersection [6]. It also has the advantage of being able to balance the target matrix with other sources of information besides traffic counts [6].

The Bayesian inference technique was used in a study by Maher [24] who tested the method on a very small transportation network [2]. The disadvantage of the methodology is that it causes a linear relationship between origin/destination pairs and traffic counts [6]. Because this approach requires the use of a target origin/destination matrix it will not be used in the development of the methodology for this research paper.

2.2 Summary

In a perfect scenario there would be an independent and consistent traffic count for each roadway segment throughout the network. This is the case required to determine the unique origin/destination matrix that is at work in the network. However, having a traffic count for each segment of roadway in the network rarely happens in reality. Without all traffic counts, there will be more than one, but a limited number, of trip matrices that can satisfy the number of traffic counts given in the study area [5].

Most of the studies reviewed in the literature focused on optimization techniques where an old/target origin/destination matrix was updated by analyzing traffic counts using maximum likelihood, generalized least squares, or Bayesian inference techniques.

Research was also conducted that examined making the iterative process of the maximum entropy technique more efficient.

The methodology based on Van Zuylen and Willumsen [4] presented in this section will be applied to the research in this dissertation because it has the ability to estimate an origin/destination matrix by only using traffic counts.

The Van Zuylen and Willumsen [4] technique will be applied to traffic counts in sample networks and actual case study cities in this report. This method may not represent traffic flow the best during heavily congested conditions due to proportional assignment but traffic congestion is very minimal in small communities [6]. The main advantage of this technique is it allows origin/destination matrices to be estimated by only using traffic counts. The other approaches rely on existing origin/destination matrix information to build a viable model. The technique will be a supplement to current traffic models. It will reduce the dependency transportation planners have on socioeconomic data and household surveys.

3 Chapter Arab and Roanoke

This chapter provides an overview of the procedure and the application of the methodology in two small communities namely Arab and Roanoke by estimating an origin/destination matrix from traffic counts to test external-external trip patterns.

3.1 Motives for the External-External Trips Determination

Through trips, or pass through trips, are a great concern for small and medium sized communities as these trips contribute to congestion on roadway infrastructure of the community that must be accommodated, but for which there are not simple methods to determine. The complete construction of a through trip table also involves determining the fraction of trips at an external station that are through trips, as opposed to trips having one end internal to the study area [25]. Trips can be termed as through trip or external-external (E-E), where origin and destination of trip fall outside the community [26]. External surveys, conducted at external stations, obtain through trip information through license plate surveys, roadside handout surveys, roadside interview surveys, and roadside interview combined with handout surveys [26]. However, the use of surveys has diminished due to rising costs, traffic delays and safety issues. Attempts have been made to use cell phone records or Bluetooth capture devices to collect traveler information. Unfortunately, these have not always been accepted by the general public and are often seen as intrusions into driver privacy.

The main goal in estimating through trips is to predict the total number, or percent, of trips that would be passing through and the distribution of trips between external stations. To remove the issues associated with surveys, regression models or synthetic procedures have been developed. It should be noted that through trip results are not always transferable between areas as they are heavily depend on the specific location, size, roadway network and relationship to other communities.

The principle methodology for determining through-trip rates cited in recent literature involves the application of a series of regression models that were developed based on external station surveys. The models predict the external trip exchange based on highway functional classification, the average daily traffic (ADT) at the external station, the percentage of trucks (excluding vans and pickups), the percentage of vans and pickups, route continuity and the population of the study area. As an alternative, Anderson presented a spatial economic model to synthesize a through trip table using surrounding communities and their impact on trip making, the study was shown to be more accurate than the common regression-based model for limited applications [27]. Han updated through trip estimation procedures with new survey data and include geographic and economic explanatory factors that can be applied in any small and medium sized urban area [26]. The drawback of the current through trip model is that, while applicable in small and medium sized communities, there is still an element of borrow from other studies and hope that the equations/methods will be transferable to the community being studied, not just the community where the model was developed [27].

Transportation professionals need a methodology to estimate through trips to improve the transportation planning process and better allocate resources for roadway infrastructure investment. This report presents research performed that utilizes roadway connectivity and traffic count data to estimate through trip patterns. The methodology calculates travel origin-destination locations from the actual roadway counts, which are collected during routine traffic monitoring procedures. The methodology has been tested against and performed similarly to a community that underwent a Bluetooth data collection study. The outcome of this research will benefit any community with a traffic monitoring program where through trip patterns could be used to improve resource allocation.

3.2 Selection of study area

Defining the territory of a study area depends on where the external zones/stations will be located that are the main generator or attractor of through trips. Based on the road network, boundary and location of external zones internal and external zones have to be identified and labeled by number on a printed map. The following figures show the zonal structure, and the numbering of external and internal zones of Arab and Roanoke respectively.

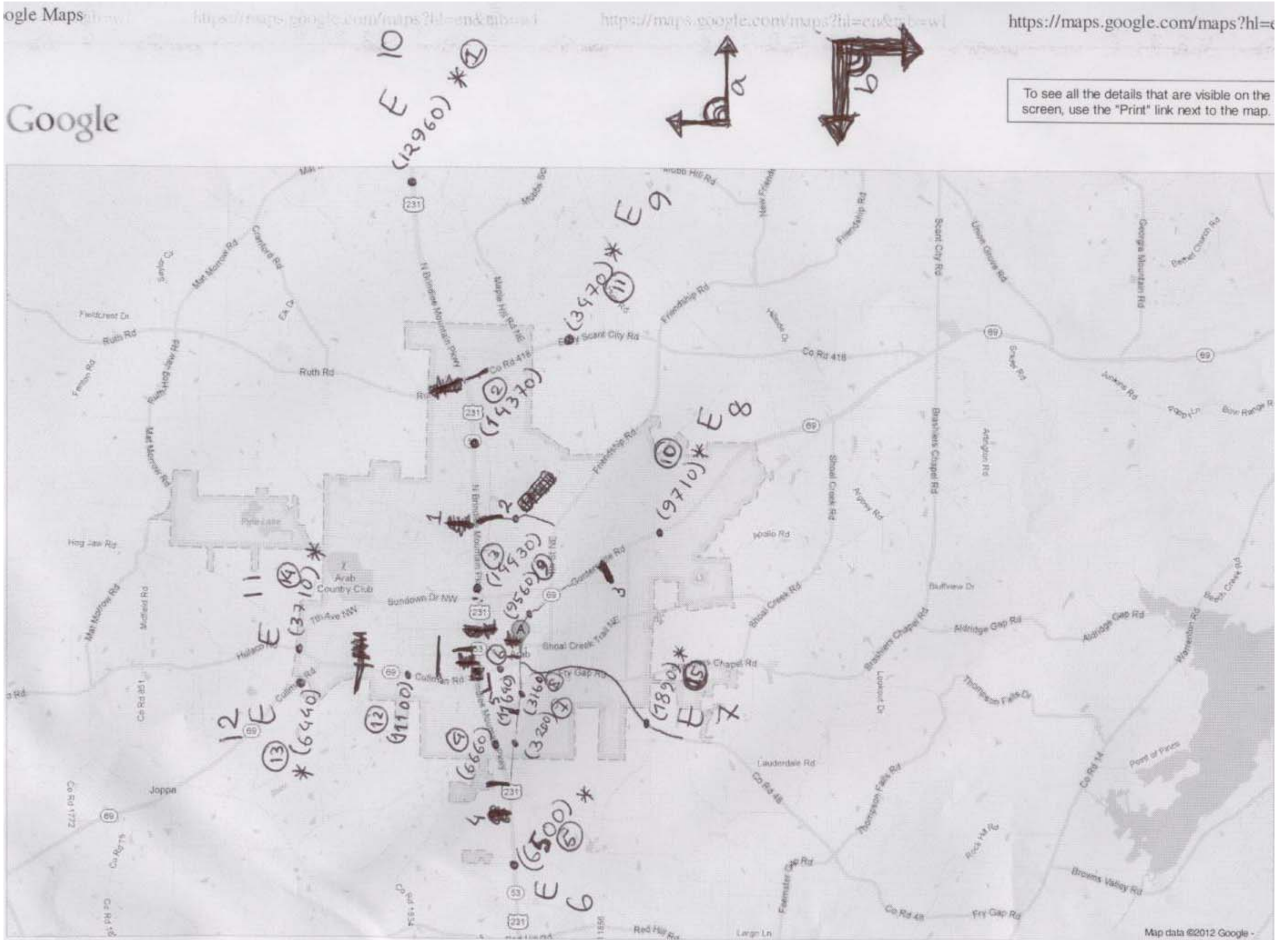


Figure 1 Arab Internal and External Zonal Layout

Figure 2 Roanoke Internal and External Zonal Layout

3.3 Required Data for Analysis and Validation Process

Traffic counts of the aforementioned road network maps were collected from the Alabama Department of Transportation's website to know the traffic volume on each link. Also, the actual percent through trips and total trips of the external stations will be required for validation purposes.

3.4 Determining Shortest Route between Zones

Because of the simplicity of road networks, the shortest routes between any pair of origin and destination were determined manually, and presented in a tabular form. The final output was formed to produce an array/matrix with all movements between zones in rows and the links in columns as value 1 if it's used along their shortest routes, a value of 0 was entered if the roadway was not on the shortest path.

3.5 Developing OD Spread Sheets for Initialization and Running Iterations

The proportional assignment was conducted through the use of the all-or-nothing assignment method in the previous step that identifies origin/destination pair on the shortest routes where traffic counts occur. Any origin/destination pair which would have the shortest path use the roadway for which the traffic count exists will be set to one, if the origin/destination pair does not use the roadway, the value will be set to zero.

During the initialization phase, a starting initialization is done by assigning a value of one trip for each OD pair (can be called as standard initialization). This first traffic count column is summed and is used to calculate initial trip volumes that can be called as "Current Volume" column [1]. The equation for each cell can be presented as follows:

$$\text{Current Volume} = \frac{Traffic\ count\ value\ for\ the\ cell\ *\ Actual\ count\ of\ link\ a}{\sum Traffic\ count}$$

The summation of Current volume column will be equal to the Actual Count of link a. Updated OD trips column compares the current volume column with the initialized OD trips column and keeps the current volume when current volume is not zero, otherwise pull the value from the initialized OD trip value. This action can be carried out by using a set of if-else procedures [1].

The next link b column is made up of proportion values for the origin/destination pairs that use the traffic count. In the column labeled “b*” the proportion values in the column “Traffic Count b” are multiplied individually by the values in the “Updated OD trips” column. In the next column labeled “Current Volume” is a ratio of the observed/actual traffic count of link b compared to the sum of the “b*” column that can be shown in the following equation [1].

$$\text{Current Volume} = \frac{cell\ value\ of\ \text{"b*"}column\ *\ Actual\ count\ of\ link\ b}{\sum \text{"}b * \text{"}column}$$

The current volume column can be summed to yield a value which is the same as the observed traffic count. The Updated OD trips column can be calculated as stated above for link a. In similar fashion “*” column followed by “Current Volume” and “Updated OD trips” columns can be developed for the other links. The “Updated OD trips” column brings forward the previous and new values and creates a new potential solution for the trip matrix [1].

The iteration phase builds on what was done in the first initialization section. It starts off with the “Traffic Count a” column which is still populated with the same proportion values for the origin/destination pairs as before. The new column labeled “a*” is calculated by multiplying each cell in the corresponding proportion value from the “Traffic Count a” column by the final “Updated OD trips” column found in the initialization spread sheet. The rest of the calculation will be the same as the initialization phase [1].

If the number of iterations increase, the summation of “*” column for a specific link is approaching closer to the observed traffic count for that link and the difference between two final Updated OD trip values of consecutive iterations is getting lesser.

3.6 Results and Evaluation

The whole OD trip matrix can be grouped into external-internal (EI), internal-external (IE), external-external (EE) and internal-internal (II) movements. As our study focuses on the external movements, the portion of the origin/destination matrix that contain external movements (EI, IE and EE), is shown and the trips at the external stations are compared with the actual and percent through trips. The following tables present the actual trips and the percentages made by external stations for Arab and Roanoke respectively. It can be noted that the diagonal values are the summation of trips from internal zones to the specific external zone.

Table 1 Actual External trips for Arab

O/D	6	7	8	9	10	11	12	Total
6	96	183	925	240	1175	230	400	3250
7	183	127	0	75	364	71	124	945
8	925	0	723	377	1842	361	627	4855
9	240	75	377	218	515	113	197	1735
10	1175	364	1842	515	1067	554	963	6480
11	230	71	361	113	554	519	6	1855
12	400	124	627	197	963	6	902	3220

Table 2 Percent External trips for Arab

O/D	6	7	8	9	10	11	12
6	2.95%	5.63%	28.46%	7.40%	36.16%	7.09%	12.32%
7	19.36%	13.49%	0.01%	7.89%	38.56%	7.56%	13.14%
8	19.05%	0.00%	14.89%	7.76%	37.94%	7.44%	12.92%

9	13.85%	4.30%	21.72%	12.58%	29.67%	6.53%	11.35%
10	18.13%	5.62%	28.42%	7.94%	16.47%	8.55%	14.86%
11	12.42%	3.85%	19.46%	6.11%	29.86%	27.97%	0.34%
12	12.43%	3.86%	19.49%	6.12%	29.91%	0.19%	28.01%

Table 3 Actual External trips for Roanoke

O/D	**10**	**11**	**14**	**15**	**Total**
10	1462	1	1570	325	3358
11	1	1264	296	61	1622
14	1570	296	363	0	2229
15	325	61	0	867	1254

Table 4 Percent External trips for Roanoke

O/D	**10**	**11**	**14**	**15**
10	43.54%	0.02%	46.75%	9.69%
11	0.04%	77.92%	18.26%	3.78%
14	70.41%	13.28%	16.30%	0.00%
15	25.95%	4.90%	0.00%	69.16%

4 Analysis in Brazos County, TX

This chapter provides an overview of the procedure and the application of the methodology in a medium sized county, Brazos, TX by estimating an origin/destination matrix from traffic counts to test external-external trip patterns and evaluate friction factors.

To test the methodology presented for building an OD matrix to determine through trips, a case study was conducted using Brazos County, Texas. The location was selected as there was a recent E-E study preformed that used Bluetooth data collected from cell phones, which was used as a validation data set.

4.1 Selection of Study Area

Based on the road network, boundary and location of the zones, both internal and external, have to be identified and labeled by number on a printed map. Brazos's external zone ranges from 1 through 13 and internal zone ranges from 14 through 58 while intersected nodes range from 100 to 219.

4.2 Required Data for Analysis and Validation Process

The aforementioned road network map preferably needs to have the traffic count for all links and a uniform scale because if the distance or travel time between zones is not available then distance between nodes has to be measured manually from the map and tabulated properly. Also, the actual percent through trips and total trips of the external stations will be required for validation purposes. It can be noted that the year must be same for the traffic counts of the study area, and the actual percent through trips and total trips of the external stations. District traffic data of Texas Department of Transportation (TxDOT) supplies the necessary traffic counts and street network map for Brazos County [28], [29]. Figure 3 shows the study area namely Brazos County with 13 external stations (ES) [30]. Table 5 presents 2011 Bluetooth output of 13 external stations where the Count column has been collected from TxDOT traffic count, was placed and aligned with Bluetooth output for comparison purposes [30], [28]. It can be seen that there are discrepancies between Total trips and Count Columns.

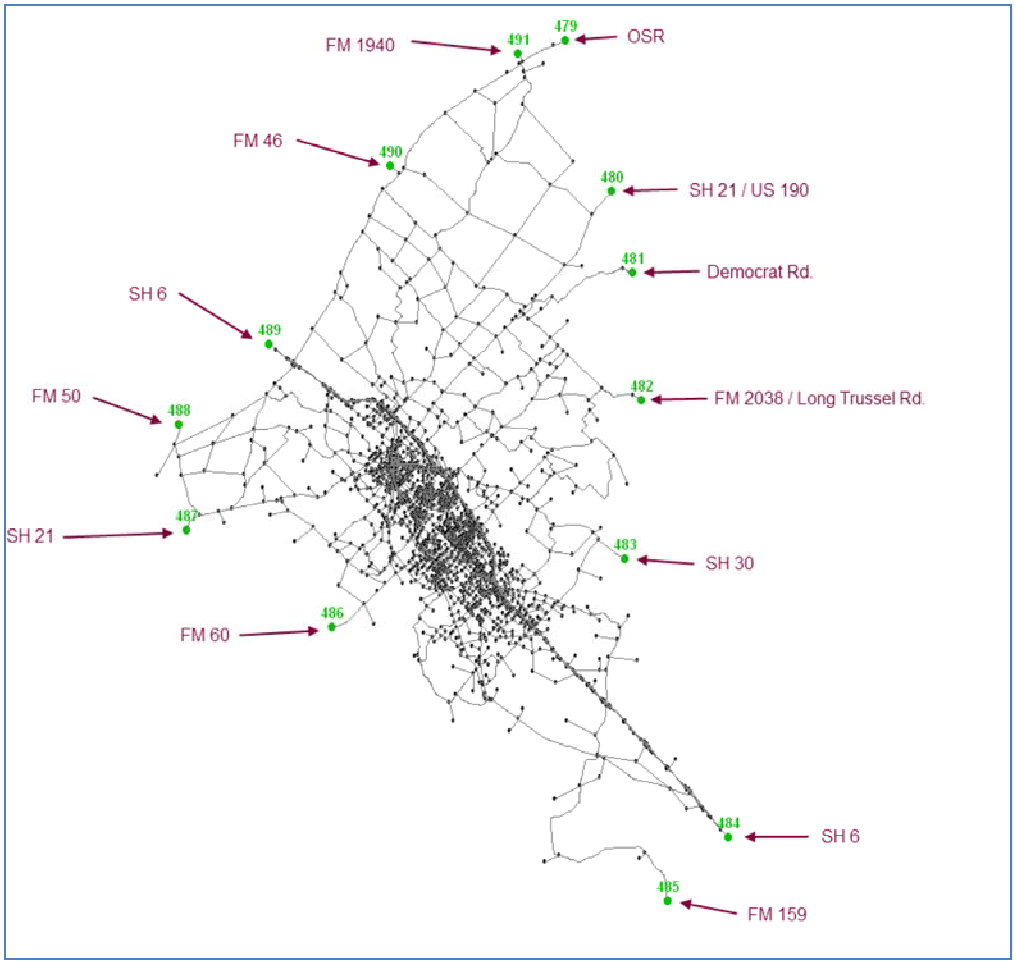


Figure 3 Brazos study area with 13 external stations [30]

Table 5 2011 Bluetooth Test Output [30], [28]

Station	ES No	Through Trips	Local Trips	% Percent Through	Total Trips	Count
479	1	91	1,028	8.1%	1,119	1,200
491	2	21	808	2.5%	829	1,050
490	3	65	1,925	3.3%	1,990	2,300
489	4	3,506	21,156	14.2%	24,662	19,000
488	5	109	871	11.1%	980	1,250
487	6	863	12,146	6.6%	13,009	12,400
486	7	216	8,474	2.5%	8,690	9,300
485	8	14	418	3.2%	432	450
484	9	3,405	22,476	13.2%	25,881	26,000
483	10	361	6,410	5.3%	6,771	6,000
482	11	0	207	0.0%	207	490
481	12	4	152	2.3%	156	700
480	13	538	6,812	7.3%	7,350	6,900

4.3 Determining Shortest Route between Zones

Preparing input database:

Links with traffic count are presented as two adjoining points - NODE A and NODE B. The nodes representing centriods or zone locations are labeled with a different range of number level so that zones can be differentiated from the intersected link nodes easily. The corresponding link distance is recorded in column named DISTANCE. Brazos study area was divided into 58 zones and access nodes were given where there is a change in traffic count. Access nodes and intersected nodes were numbered 100 through 219. Input file with three columns thus contains link, zone and distance information of a study area.

Building shortest route algorithm:

The movement from one specific zone to the other was coded in MATLAB script file in such a way that draws a tree diagram with definite number of levels until it reaches the desired destination. Number of routes need to be compared between two zones and it is considered as an input variable.

MATLAB script reads the excel input file and calculate shortest paths between all zones in the community; the shortest route from zone 1 to the other 58 zones, such as 1 to 2, 1 to 3 and so on as incremented by one in for loop statement which is nested by another for loop to find the same for zone 2, 3 and so on. To draw one route from one zone to other, first level defined as a structure array variable saves the value of starting zone to help storing the locations of it from the input table; similarly second level stores the values of node that matches the locations of first level with either NODE A or NODE B. In third layer and further, locations of previous level matching with input node columns are stored and based on those locations; next level saves the nodal information from either NODE A or NODE B. In addition, corresponding distances are stored in the specific structural array level. A for loop starts executing these actions at third layer and ends till the layer of destination zone can be found. Moreover, the whole process can be repeated until desired number of routes can be found. At last, another for loop back tracks the whole route from bottom till second layer, and each route can be presented as the connecting nodes beginning from the origin zone and ending at the destination zone and total distance of corresponding linked nodes. This output of all routes and their distances was saved as mat file and fed into again to compare available routes between two zones based on shorter distance. After sorting out those routes, the route for one certain movement that possesses the minimum distance was saved in a row of an array where each column contains the associated links presented as “NODE A – NODE B”. The final output was formed to produce an array/matrix with all movements between zones in rows and the links in columns as value 1 if it’s used along their shortest routes; if not on the shortest route the value is 0. For Brazos network, number of routes has been kept 50, and the way this script file draws the tree diagram is to have fewer layers for adjoining links, and follow the order of the node columns created in input file for picking one of those. It is possible using this methodology to select a path that is not truly the shortest path, however in terms of total distance this discrepancy is negligible.

Running and exporting output table:

This algorithm can be run by feeding in the input database for any study area and desired 0 and 1 output table can be exported and pasted in Excel spread sheet.

4.4 Developing OD Spread Sheets for Initialization and Running Iterations

These steps were completed following the above formulas mentioned in the same header section. In larger network, more than 200 zones, it is more difficult to satisfy every observed count and it is wise to keep track of the difference of two consecutive iterated final Updated OD values. It can be noted that the count columns at the end of the spread sheet are likely to satisfy the observed counts. As this study is concern about the external stations, link columns must be rearranged in such a way so that all external counts will be placed at the end of the initialization and iteration spread sheets.

4.5 Formulation of Different Options

The standard initialization values, where everything was considered equal, showed impractical results, leading to the values in the origin/destination table to be overestimated for certain origin destination pairs who are geographically located near each other for external zones and showed negligible development inside the territory to justify the large movements for internal zones. Therefore, based on the assessment of trip matrix developed under standard initialization, the following options can be formulated and can be reflected in the new initialization columns which can be identified as constrained initialization columns:

- External OD pairs which are geographically closely located can be constrained by initializing their trip value as zero and will be unchangeable till the end of the last iteration. It is reasonable to hypothesize that two external stations in close proximity to each other would be less likely to exchange through trips than two external stations at opposite sides of the urban area [25].
- Internal zones attracting or generating excessive trips, that are causing huge deviation for other OD pair trips can be constrained by initializing all trips related to the Internal zone as zero and will be unchangeable till the end of the last iteration
- Any combination of above two options can create other viable option

Running initialization and iteration spread sheets for each option can be done and final trip matrices for each option can be stored for further analysis.

The following options were selected to make a comparison among them for Brazos County, as the results from standard initialization shows the aforementioned unrealistic situations:

- Trips between external stations located close together and not passing through the study area (shown in Figure 3), are constrained as zero trips. The following table details the clusters of external stations based on geographical setting:

Table 6 The Clusters of External Stations and Their Movements

EE Origin/Destination (O-D) Clusters	**O-D Movements**
1, 2, 3, 4, 5	1-2, 2-1, 1-3, 3-1, 1-4, 4-1, 1-5, 5-1, 2-3, 3-2, 2-4, 4-2, 2-5, 5-2, 3-4, 4-3, 3-5, 5-3, 4-5, 5-4
1, 2, 12, 13	1-12, 12-1, 1-13, 13-1, 2-12, 12-2, 2-13, 13-2, 12-13, 13-12
5, 6, 7	5-6, 6-5, 5-7, 7-5, 6-7, 7-6
8, 9, 10, 11, 12, 13	8-9, 9-8, 8-10, 10-8, 8-11, 11-8, 8-12, 12-8, 8-13, 13-8, 9-10, 10-9, 9-11, 11-9, 9-12, 12-9, 9-13, 13-9, 10-11, 11-10, 10-12, 12-10, 10-13, 13-10, 11-12, 12-11, 11-13, 13-11

- In addition to the previous constraint, trips between any zone and internal Zone 52 (shown in Figure 4) are also constrained as zero trips to avoid the unreasonable amount of attracting/producing trips (total about 44,000 trips and highest among all zones) within the zone.

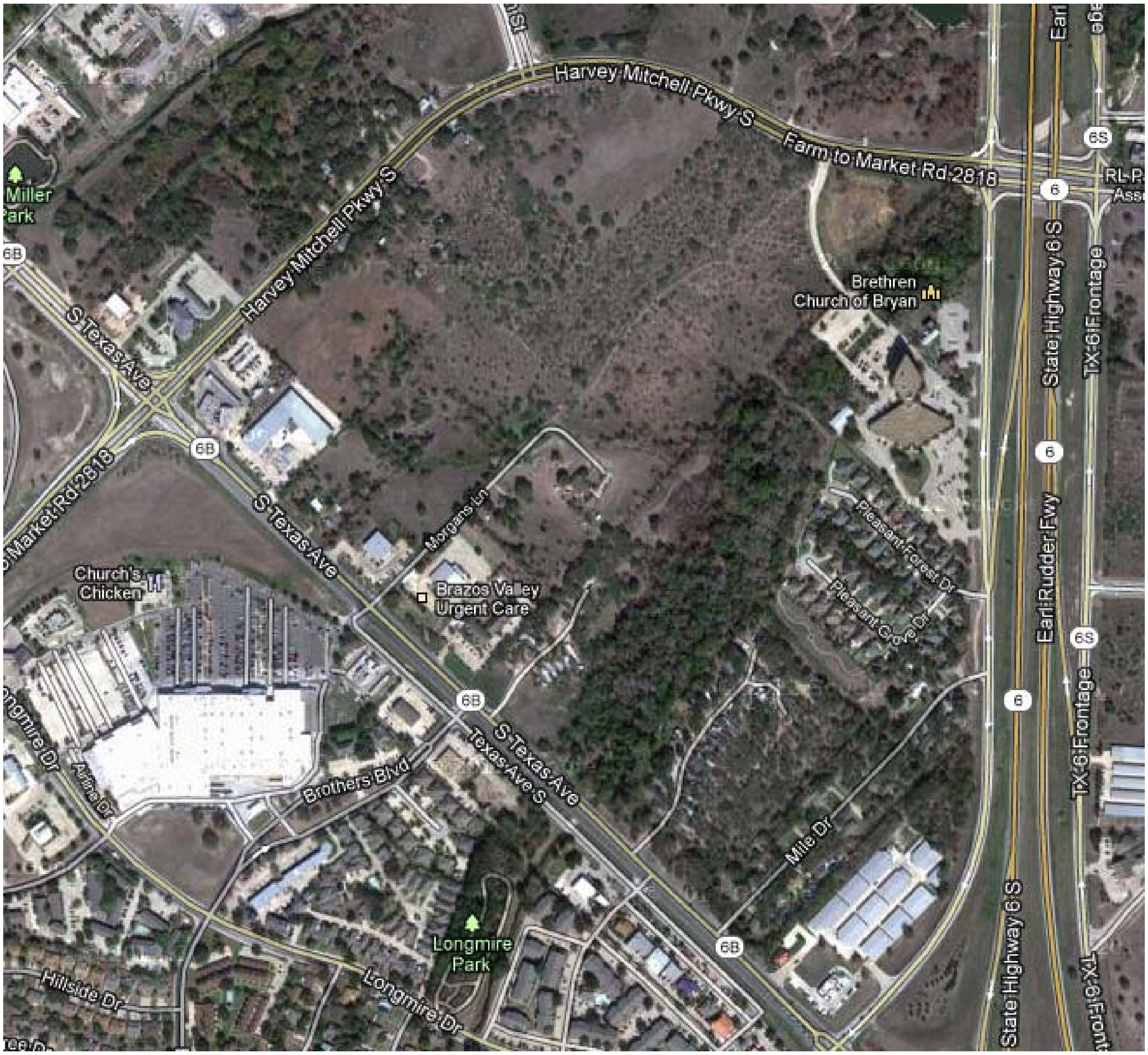


Figure 4 Actual layout of household or employment for internal zone 52

- In addition to the above constraint, trips between any zone and internal Zone 57 and 58 (shown in Figures 5 and 6 respectively) are also constrained as zero trips to avoid the unreasonable amount of attracting/producing trips (total about 10,000 trips individually) within the zone.

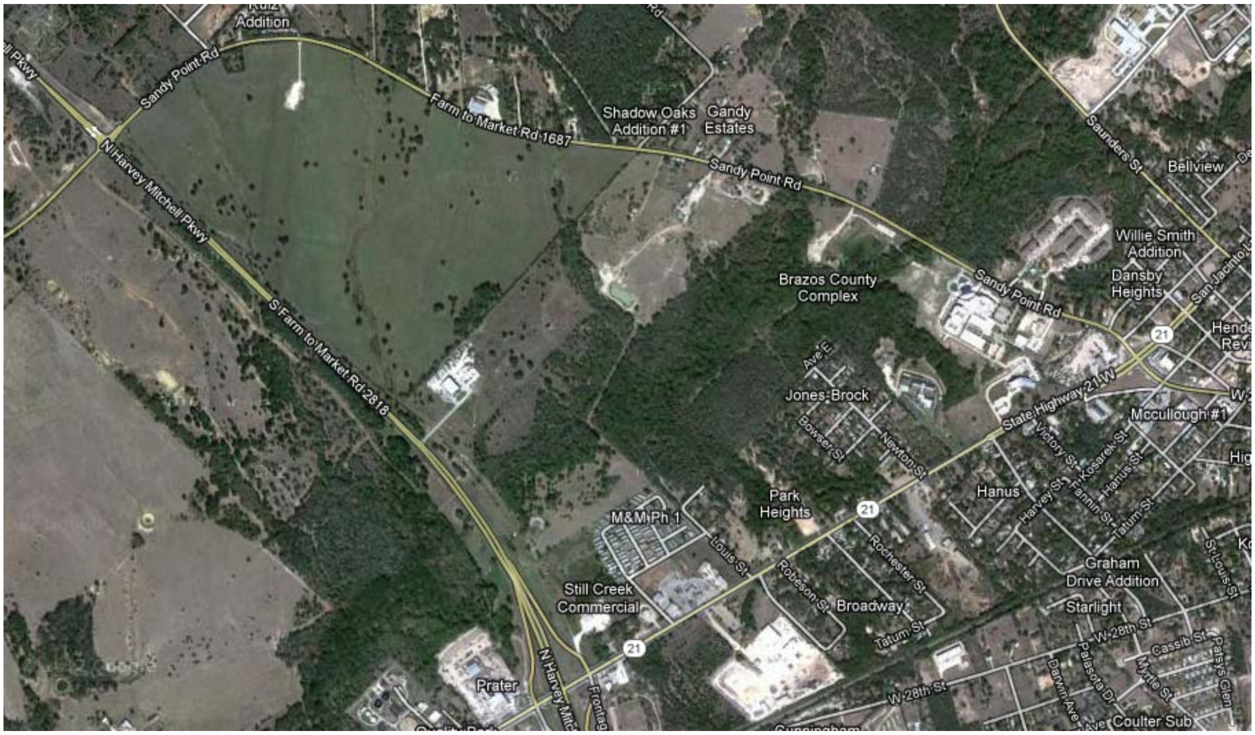


Figure 5 Actual layout of Household or Employment for Internal Zone 57

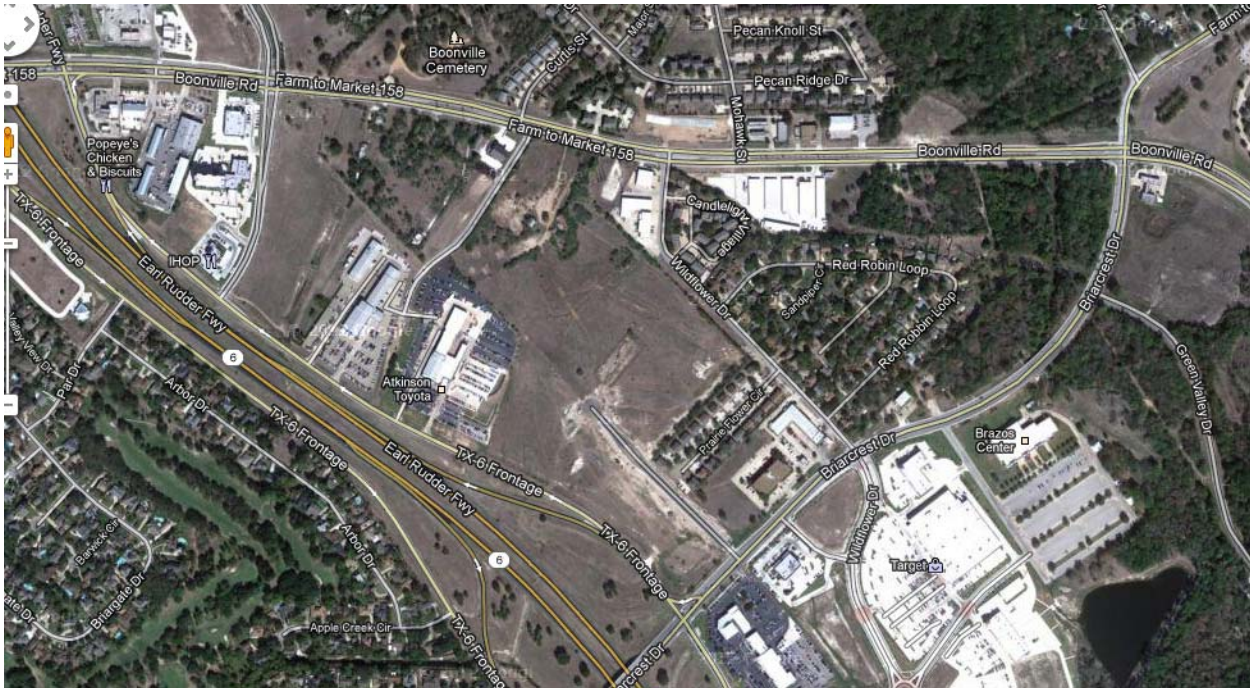


Figure 6 Actual layout of Household or Employment for Internal Zone 58

4.6 Evaluation/Validation of Options

The whole OD trip matrix can be grouped into EI, IE, EE and II movements. As our study focuses on the external movements, therefore the origin/destination matrix was compressed to show only external movements. The external movements were summed by external stations and compared with the actual percent through trips and total trips of the external stations. Total differences are calculated by using the following equations:

Total Differences in % through trips

$$= \sum_{\mathrm{N}} (\mathrm{ABS}(\mathrm{Actual}\ \%\ \mathrm{through\ trips} - \mathrm{modeled}\ \%\ \mathrm{through\ trips}))$$

$$\text{Total Differences in total trips} = \sum_{N} (ABS(Actual\ total\ trips - modeled\ total\ trips))$$

Where, N = number of external stations, ABS = Absolute value

Each option can be examined after running a certain number of iterations and following the aforementioned procedures and the best route can be selected based on the minimum values of the total differences. To assess the trend of total differences by increasing the number of iterations, a scatter plot can be drawn for the above two parameters. Parallel-constant lines will be found that represent no need of further iterations and thus the number of iterations can be identified to reach a steady state.

4.7 Results

The last iteration has been selected as 151 where the slope of lines shown in Figure 7 tends to zero. Among the three options, results of best two options are presented in the following tables.

Table 7 Results on applying initialization based on constraining Internal Zone 52 and specific EE movements

Station	**EE No**	**Total Trips Difference**	**% difference**	ABS of % difference	ABS of Total trips difference
479	1	21.67	0.42%	0.42%	21.67
491	2	169.09	6.02%	6.02%	169.09
490	3	117.20	11.69%	11.69%	117.20
489	4	-5723.84	-13.51%	13.51%	5723.84
488	5	269.80	-10.94%	10.94%	269.80
487	6	-829.10	-3.51%	3.51%	829.10
486	7	579.77	0.99%	0.99%	579.77
485	8	17.01	-1.37%	1.37%	17.01
484	9	56.25	-8.67%	8.67%	56.25
483	10	-771.00	-4.66%	4.66%	771.00
482	11	283.00	12.65%	12.65%	283.00
481	12	544.00	-2.30%	2.30%	544.00
480	13	-450.55	-6.69%	6.69%	450.55
			Total	83.42%	9832.27

Table 8 Results on applying initialization based on constraining Internal Zones 52, 57 & 58 and specific EE movements

Station	**EE No**	**Total Trips Difference**	**% difference**	ABS of % difference	ABS of Total trips difference
479	1	16.13	1.25%	1.25%	16.13
491	2	164.24	6.85%	6.85%	164.24
490	3	100.37	13.11%	13.11%	100.37
489	4	-5730.59	-12.97%	12.97%	5730.59
488	5	268.97	-8.00%	8.00%	268.97
487	6	-784.22	-3.60%	3.60%	784.22
486	7	580.40	0.87%	0.87%	580.40
485	8	11.73	5.69%	5.69%	11.73
484	9	61.26	-8.81%	8.81%	61.26
483	10	-771.00	-4.66%	4.66%	771.00
482	11	283.00	12.53%	12.53%	283.00
481	12	544.00	-2.30%	2.30%	544.00
480	13	-453.49	-4.27%	4.27%	453.49
			Total	84.91%	9769.39

After looking at the Tables 7 and 8, above options are very close to each other in terms of the maximum and total amount discrepancies, and errors in percent difference are in a range of around -15% to +15% for both cases.

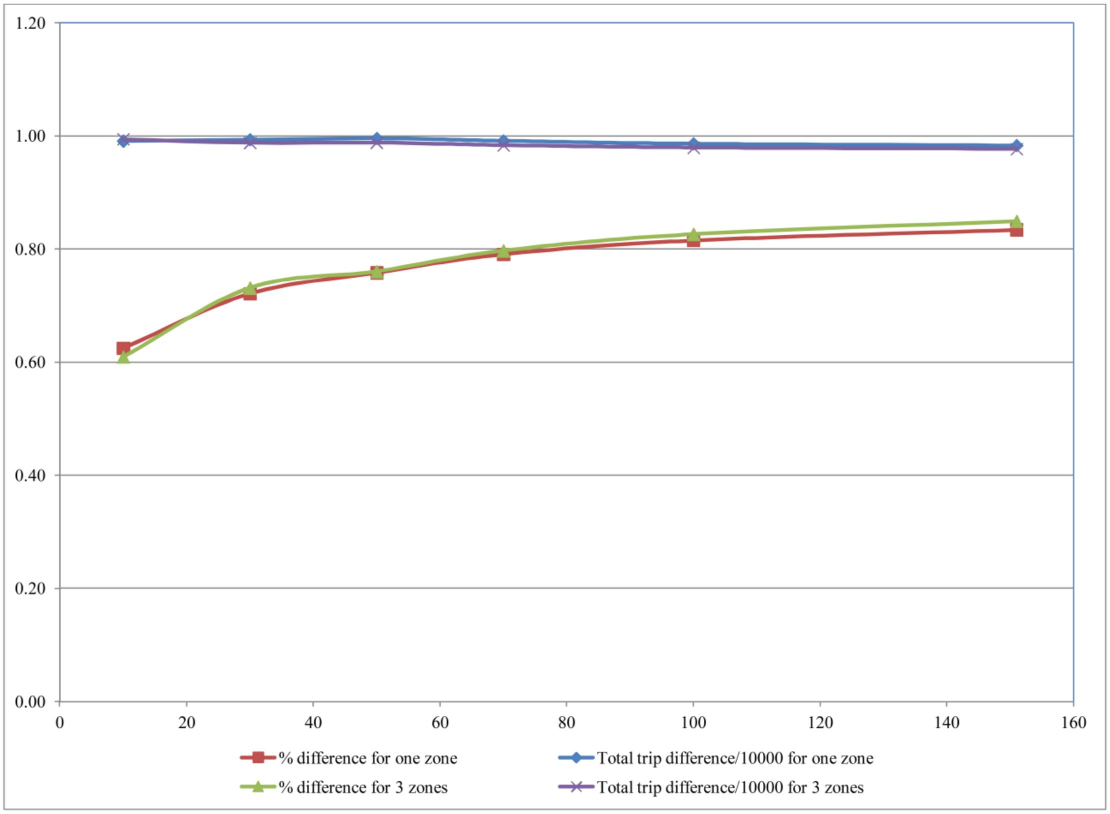


Figure 7 Scatter plots of total errors in % difference and total trips difference/10000

Figure 7 shows how the numbers of iteration influence the slope of lines that were drawn to check the pattern of total errors of two options along the horizontal axis. The option constrained by one internal zone with minimal total errors can be selected as the top best among all of them. Finally it can be stated that future trend of through trips can be estimated by using this option followed by the above iterative procedure for 13 external stations of Brazos County. Since there are discrepancies between Total trips and Count of 13 external stations, a comparison of percent through and through trips between blue tooth and the best option from this study after replacing those 13 counts by total trips (results shown in the following Table).

Table 9 Results after replacing 13 external counts

EE No	Through Trips from this study	Through Trips from Bluetooth	% Percent Through from this study	% Percent Through from Bluetooth	Difference in percent
1	115	91	11.01%	8.1%	-2.91%
2	85	21	11.01%	2.5%	-8.51%
3	272	65	14.97%	3.3%	-11.67%
4	631	3,506	2.56%	14.2%	11.64%
5	1	109	0.13%	11.1%	10.97%
6	885	863	6.93%	6.6%	-0.33%
7	273	216	3.15%	2.5%	-0.65%
8	8	14	1.79%	3.2%	1.41%
9	1,136	3,405	4.40%	13.2%	8.80%
10	40	361	0.59%	5.3%	4.71%
11	25	0	12.23%	0.0%	-12.23%
12	0	4	0.00%	2.3%	2.30%
13	21	538	0.28%	7.3%	7.02%

As shown in Table 9, the two methodologies produce similar results for the majority of external stations. The external stations with the higher traffic volume, the traffic count methodology performs worse as it seems these trips are more likely distributed as E-I trips versus pass through. However, the difference in percent between the two studies is within +/- 13%. This discrepancy is similar to the values found when using regression models or synthetic procedures [26], [27].

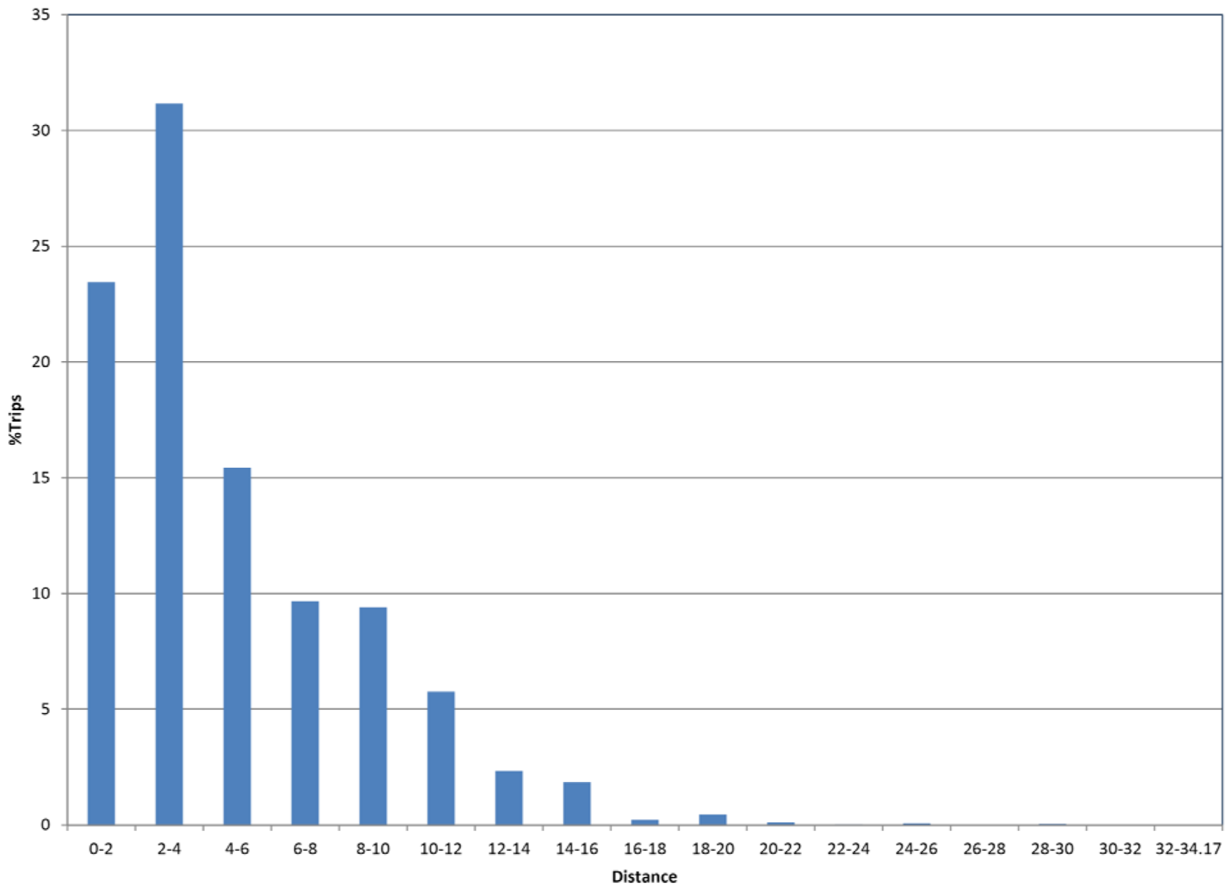


Figure 8 Bar Diagram of Distance vs Percentage of trips between OD pairs of Brazos, TX

Figure 8 represents the distribution of trip lengths, essentially reflecting the friction factors of Brazos County. It shows few or no trips generated at the tail while most trips occurred at the first quarter.

4.8 Discussion

Modlin's and Anderson's methods employ linear regression equations to estimate through trips. The discrete choice based model proposed by Martchouk and Fricker can ensure that the through trip percentages add up to100 percent [31]. Recent research done by Talbot and Burris developed a set of two logit models to estimate through trips for a wide range of study area sizes that requires a significant amount of data, including external survey data, traffic data, roadway data, demographic data, interaction score data, and measures of external station separation [32].

Unlike many previous methodologies, this step by step process does not require obtaining sufficient amount of data for the study area. In this study only traffic counts and actual through trips are necessary database to find the best option with a cutoff point to run the number of iterations for estimating through trips, that is the uniqueness of this research. And traffic counts are accessible for any year that can easily be corresponded with the year of actual through trips. There is no doubt of its transferability for a wide range of study area sizes and this methodology can be applied in smaller or larger areas.

4.9 Conclusion

This research developed a step by step procedure to determine through trips pattern for any kind of study area by using minimal amount of existing data. The outcome of this research will be very useful for the urban areas where external survey data cannot be conducted due to lack of resources.

5 Chapter Hartselle, AL

5.1 Hartselle Analysis

The methodology will be examined on a real world network as a case study. Doing so will show that the methodology is viable for use on a real world network. Also, any issues or challenges associated with analyzing a real world network can be examined.

This chapter will begin by showing the location of the real world network chosen for the case study. The chapter will demonstrate how the network was setup and how the methodology was utilized. Finally, the chapter will culminate with a section showing that a statistically significant origin/destination matrix for the region in question was created.

5.2 Case Study Location

The city of Hartselle, Alabama was used as the location for the case study. It is located within North Central Alabama about 10 miles south of the city of Decatur. The total area of the community is 16.26 square miles [33]. It has a population of 14,255 people with 5,177 households [34] [33].

The city of Hartselle is serviced by two major highways: United States Highway 31 and Alabama Highway 36.

The city of Hartselle's roadway network comes under the control of the Decatur Metropolitan Planning Organization. The organization provided the information required to conduct the research for this study.

Hartselle was chosen as the location for the case study because data was easily obtainable. Also, the roadway network of the city was the right size to show that the methodology will work for a small community.

5.3 Network Analysis - Plan and Setup

The plan and thought process used to analyze the Hartselle roadway network will be explained in this section.

The Decatur Metropolitan Planning Organization provided this research study with information for their entire network which included the city of Decatur, Hartselle, and other outlying areas. The information given included a travel demand forecasting model based on the Four Step Planning Process which was setup in CUBE software along with an origin/destination matrix.

The plan to analyze the real world Hartselle roadway network begins by looking at the existing traffic analysis zones, land use, and socioeconomic data that is currently accepted and used by the transportation planners. The model will be run using these existing parameters to assign traffic to the roadways using the all-or-nothing assignment technique. These traffic counts populating each roadway segment in the network will be used for implementation of the methodology. A scenario employing all the traffic counts will be run to show that the methodology estimates an origin/destination matrix that is statistically similar to the origin/destination matrix constructed by the CUBE model.

The thought process for using simulated traffic counts is that traffic assigned by an all-or-nothing technique to the roadway network will cause the methodology to yield results replicating the origin/destination matrix produced by CUBE, thereby, showing that the methodology is viable for an actual community.

The setup of the Hartselle network in CUBE took several steps. Because the area given in the CUBE model included all of the communities under the authority of the Decatur Metropolitan Planning Organization, the first step in the setup process was to subdivide the roadway network. The vicinity of Hartselle within the larger network was identified. The traffic analysis zones and roadways important to the flow of traffic within Hartselle were selected. Then the Hartselle region was cutout and set apart for modeling purposes. The Hartselle network as displayed in CUBE can be seen in Figure 9.

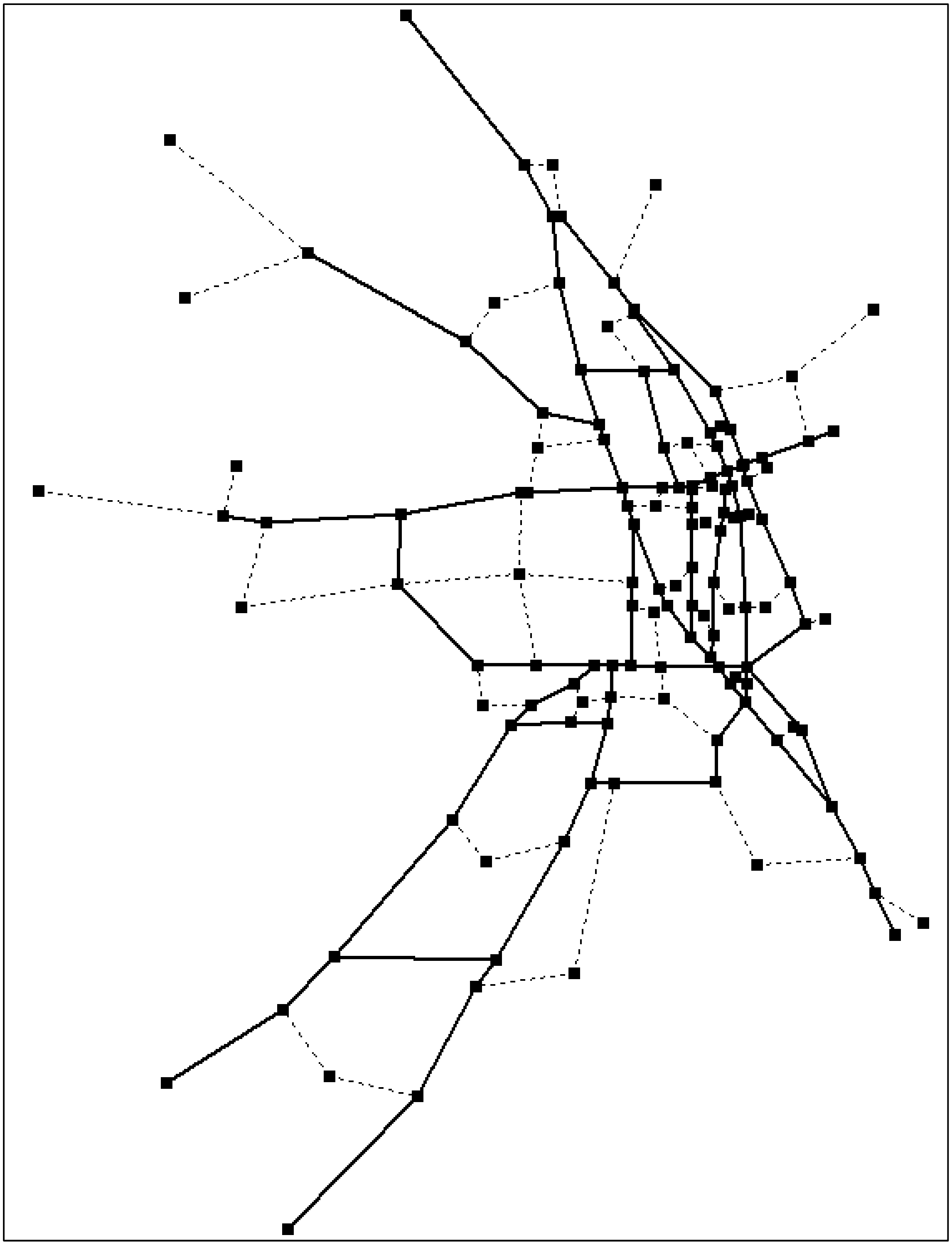


Figure 9 Hartselle Network as shown in CUBE

The network in CUBE contained many short links in a connected sequence to represent the arrangement of the roadway system. These links were reworked so that some of them were removed and others were lengthened to simplify the network. The information on the links, which included traffic volume, capacity, and travel time was maintained on the reworked links. A before and after picture showing the links in CUBE is exhibited in Figure 10. The final count on the number of bidirectional links in the network is 188. In addition, there are 42 zones in the network.

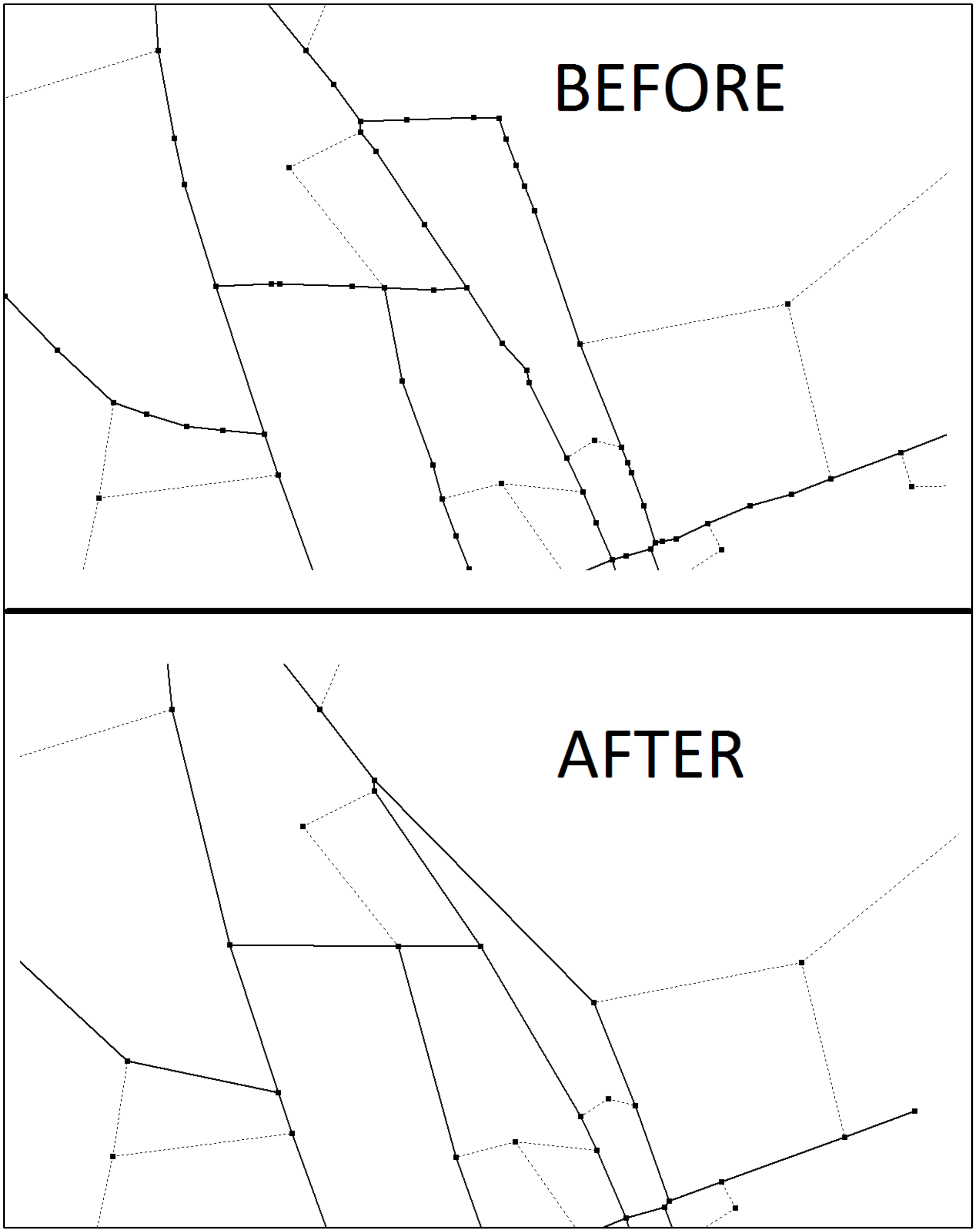


Figure 10 Before and After Shot of Reworked CUBE Links

Reducing the number of links in the network was done to make the process of identifying the shortest path between origin/destination pairs more straightforward. The number of links in the CUBE network is directly proportional to the number of columns and rows that have to be utilized in the program. Minimizing the number of links accomplishes two things. First, it makes setting up and entering the data into a text file easier. Secondly, it cuts back on the computer resource requirements needed to run the program.

At this point the city of Hartselle's roadway network is setup and ready to be used. Following the methodology the traffic was assigned to the network using the shortest path algorithm and all-or-nothing assignment in CUBE. Each link was analyzed to see if it was on the shortest path between origin/destination pairs.

5.4 Complete Traffic Count Scenario

The network was analyzed according to the methodology. An origin/destination matrix was estimated using 100 percent of the traffic counts in the network. The development and results of which are shown in this section.

The Hartselle network contained many centroids that were directly aligned at intersections with other centroid connectors. This required that they be included with the roadways in the text file because trips would otherwise be lost which would result in fewer trips between origin/destination pairs. An example of this is shown in Figure 11 where zone 10 is directly connected with zones 9 and 11.

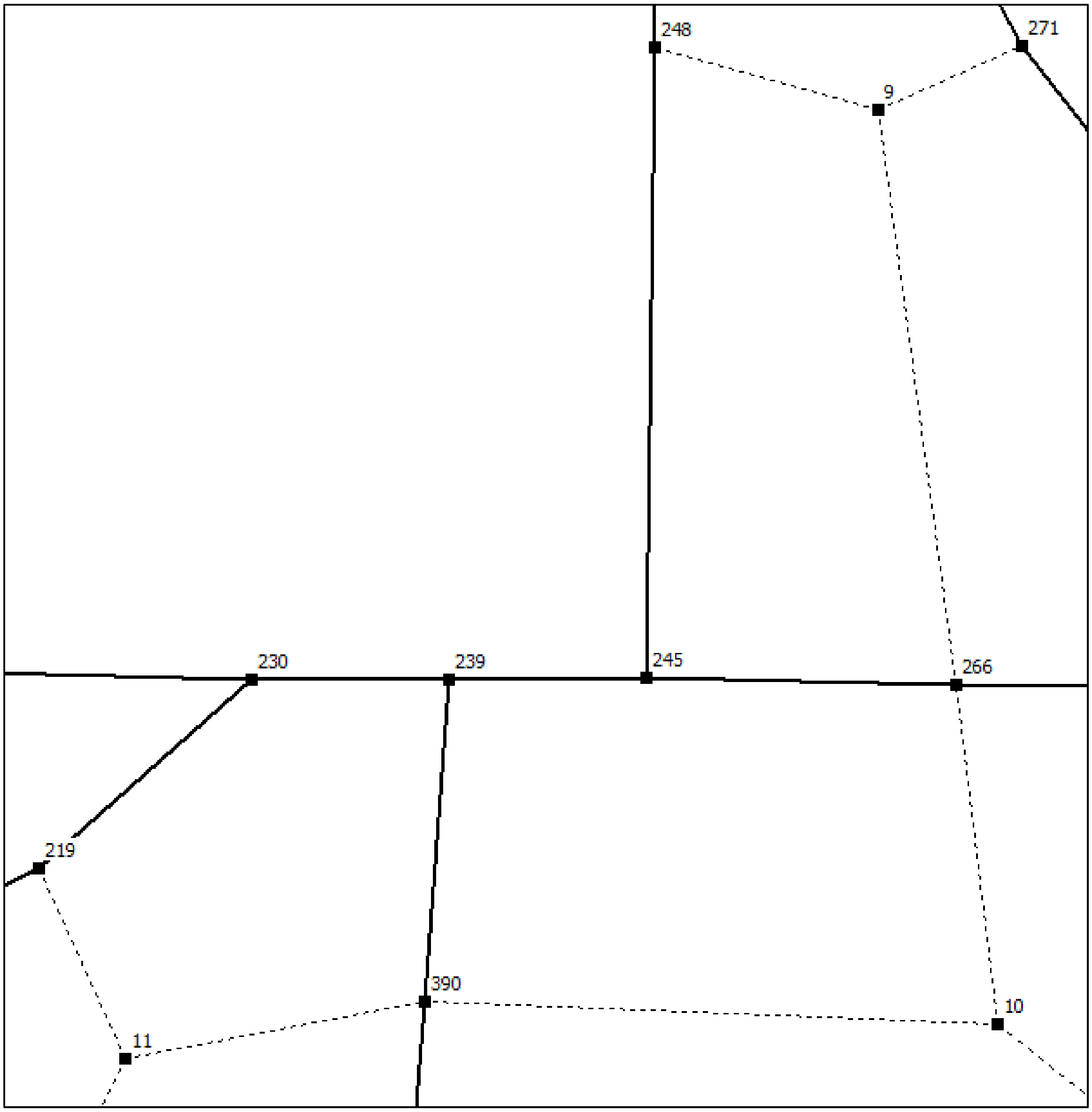


Figure 11 Direct Links between Centroids

There were a total of 109 bidirectional roadways and 79 bidirectional centroid connectors in the network. Each roadway link and centroid connector were given two columns in the text file; one column for each direction. The roadway links were placed in the first part of the spreadsheet and the centroid connectors were placed in the last. There are a total of 42 zones in the network which yielded a total of 1,764 origin/destination pairs. When the number of origin/destination pairs is combined with the number of columns used to process the convergence of the traffic counts for the roadway links and the centroid connectors a matrix of size 1,128 columns and 1,764 rows is created. After constructing the initial text file it took a process of 31 iterations for convergence within five percent of the traffic counts to be reached and 200 iterations were run in total to further reduce the convergence point.

To illustrate that the program was working correctly zone 40 was connected directly to zone 30 by a centroid connector. It did not connect to any roadway. The only trips that were shown to pass to zone 40 were from zone 30. This is supporting evidence of the trips being distributed throughout the network appropriately. A screenshot of the layout of zone 40 and zone 30 in the network is shown in Figure 12.

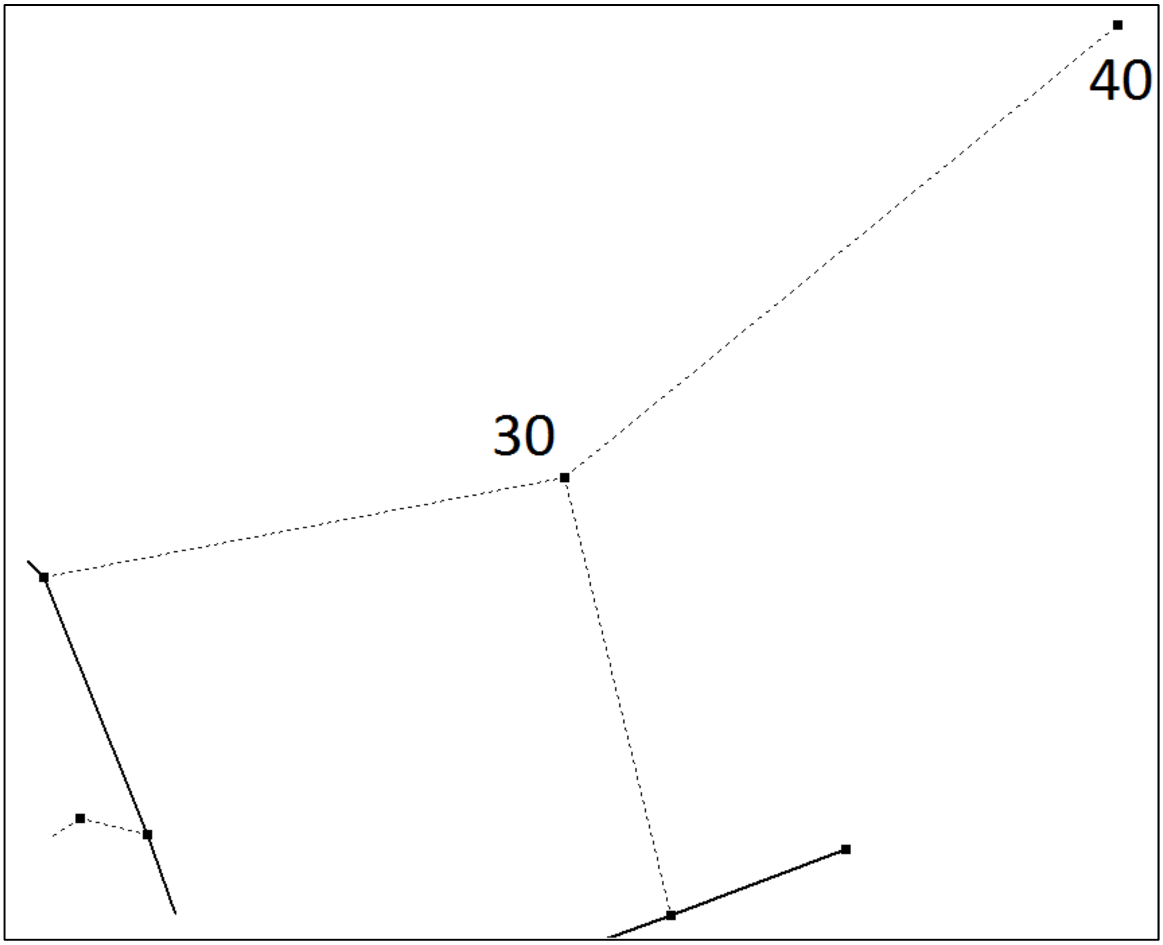


Figure 12 Snapshot of Zone 40 and Zone 30 Location

The origin/destination table in CUBE contained intrazonal pairs. These pairs had to be removed from the analysis because the methodology has no way of estimating trips within zones. With the editing process complete it is now time to analyze the results of the origin/destination matrix estimated from 100 percent of the traffic counts. The resulting table of origin/destination pairs is too large to show in this paper, therefore descriptive statistics are used to illustrate its performance.

Not all of the estimated trips between origin/destination pairs matched perfectly to the CUBE trip table. It was found that over 80 percent of the estimated trips between origin/destination pairs were within 15 trips of the CUBE trip table. Along those same lines over

90 percent of the estimated trips between origin/destination pairs were within 30 trips of the CUBE trip table.

The total number of trips recorded in CUBE's trip table is 63,974. For comparison purposes the total number of trips for the estimated origin/destination matrix is 63,885. These values are within two-thousandths of a percent, therefore they are reasonably close. The sum of the estimated trips for each origin zone is also close to the CUBE output. The root mean square error is 51.98 while the mean absolute error is 14.33 and the mean percentage error is 0.994. With 80 percent of the estimated trips being within 15 trips of the CUBE output it makes sense that the mean absolute error is less than a value of 15.

To ensure that the estimated origin/destination matrix is not statistically significantly different from the CUBE origin/destination matrix a Wilcoxon Signed Rank test was performed at an alpha value of five percent. The results for the test are shown in Figure 13.

Wilcoxon Signed Rank Test: 100% Traffic Count Scenario

```
Test of median = 0.000000 versus median not = 0.000000

                  N for   Wilcoxon         Estimated

               N   Test  Statistic      P     Median

Difference  1502   1502   549947.0  0.391    -0.1133
```

Figure 13 Wilcoxon Signed Rank Test Results for a Complete Traffic Count Scenario

The results for the Wilcoxon Signed Rank test in Figure 13 show a P-value of 0.391. This means that the estimated origin/destination pairs are not significantly different from CUBE's origin/destination pairs because the P-value is larger than the alpha value of five percent.

Since the estimated origin/destination matrix passed the Wilcoxon Signed Rank test, the methodology is seen as working validly on the real world network of Hartselle.

5.5 Conclusion

The results of the roadway network for the city of Hartselle, Alabama have shown that the methodology can be viably used in a real world network.

The information revealed in this chapter will benefit small and medium sized communities that want to estimate an origin/destination matrix without using the traditional Four Step Planning Process. It will especially benefit communities that do not have adequate funding to conduct household surveys as part of the trip generation step. This methodology can also be used to supplement the traditional process.

With ample traffic counts in a network it is possible to successfully use the estimation technique to develop travel demand forecasting models.

6 Chapter Anniston, AL

6.1 Introduction

It was desired to further test the capability of the program on a real world community. The newly developed software program will be used to estimate an O/D matrix for a real world community, then the O/D will be compared to an accepted CUBE O/D matrix for the community and the results will be statistically analyzed. After the initial analysis the zones of the community will be aggregated to see if acceptable results can be achieved by using larger zones to correspond to a lower number of available traffic counts.

The community that was chosen for the analysis was Anniston, Alabama. This city lies in the east-central portion of the state. It has a population of 24,276 people and an area of 52.3 square kilometers [35].

The process this paper will follow for analyzing the community will begin by running the Anniston CUBE network using all-or-nothing traffic assignment. The assigned roadway volumes will be exported to a properly formatted text file so that the program can generate an estimated O/D trip table. The O/D matrix used by CUBE will be compared to the O/D matrix created by the estimation program and statistical tools will be employed for descriptive purposes and to identify if the matrices are significantly different.

The next phase of the analysis will be to use actual traffic counts around the city of Anniston to create an O/D matrix using the estimation program. Due to the low number of available traffic counts in the community the existing zones will have to aggregated to create larger ones

6.2 Anniston Network

The project began by analyzing the Anniston network in its current state. The Anniston Network in CUBE software was comprised of 142 traffic analysis zones and 1,732 links. The CUBE network for Anniston can be seen in Figure 14.

Traffic assignment using the shortest path algorithm and all-or-nothing assignment for the CUBE network was run. The processed network was exported to a database file. The database file was edited to only retain columns for “Node A,” Node B,” “Time,” and “Volume.” Then it was saved as a plain text tab delimited file.

The tab delimited text file was used by the estimation software to generate an O/D matrix based on the shortest path between zones/nodes and the assigned traffic volumes. The results of which are examined in the next section.

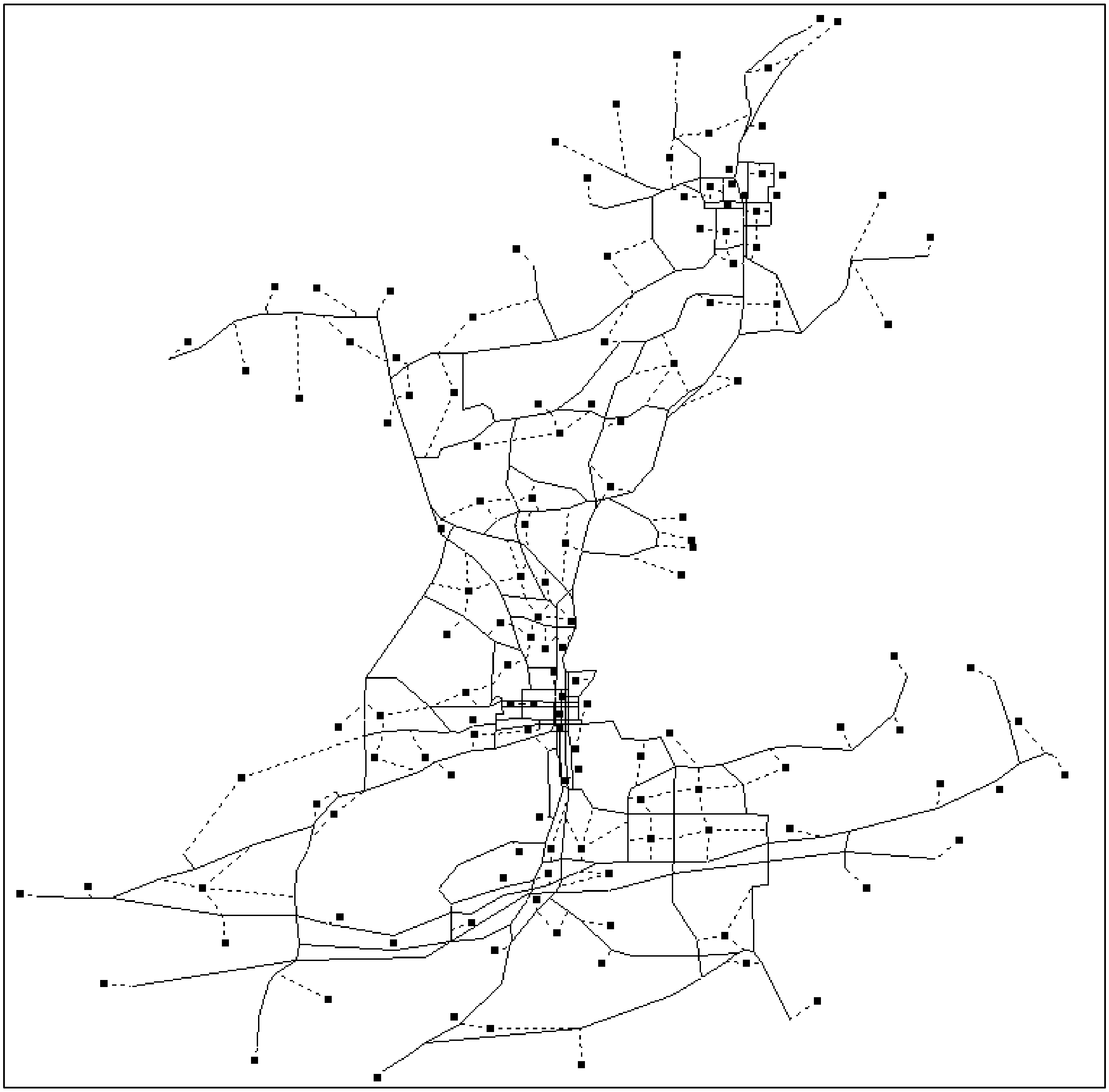


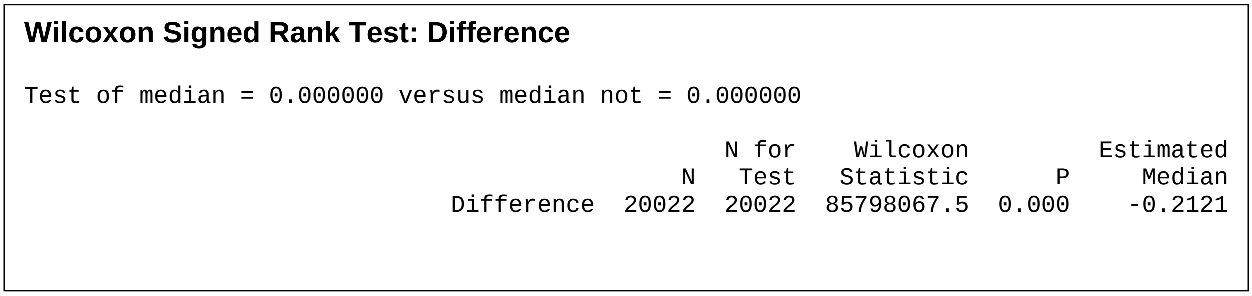
Figure 14 Anniston CUBE Network with 142 Zones

6.3 Anniston Network Results

The CUBE O/D matrix and the estimated O/D matrix were each formatted into a three column trip table so that a side by side comparison could be made. Doing so allowed the data to be analyzed statistically using root mean square error and mean absolute error.

The results of the comparison between the CUBE O/D and the estimated O/D trip table reveal that 91 percent of the trips are within a value of 15 and 95 percent are within 30 trips. There are a total of 20,022 O/D pairs in the Anniston network. The root mean square error is 40.06 and the mean absolute error is 6.81.

Additional statistical analysis will be run on the two data sets to test that the estimated matrix does not significantly differ from the CUBE O/D matrix. To do so the difference between the CUBE O/D matrix and the estimated O/D matrix were found for each O/D trip pair using the three column trip tables. Then a test known as a Wilcoxon Signed Rank Test can be performed on the differences. The test was executed at an alpha value of five percent. The Wilcoxon Signed Rank Test compares the median of the CUBE origin/destination matrix to the estimated origin/destination matrix. The H_0 for the test is that the difference between the medians is zero where H_1 states that the differences between the medians are different from zero [36]. The Wilcoxon Signed Rank Test was performed using Minitab® Statistical Analysis Software, Release 16 on each simulation run [37]. The results of the Wilcoxon Signed Rank test are shown in Figure 15.



```
Wilcoxon Signed Rank Test: Difference

Test of median = 0.000000 versus median not = 0.000000

                                N for   Wilcoxon          Estimated
                          N      Test   Statistic      P     Median
Difference            20022     20022  85798067.5  0.000    -0.2121
```

Figure 15 Wilcoxon Signed Rank Test for the Anniston Network O/D Trips Comparison

The results of the Wilcoxon Signed Rank test reveal a P-value of zero which is less than the alpha value of five percent, therefore, it must be concluded that the estimated O/D matrix is significantly different from the CUBE O/D matrix.

Further examination of the data divulges that two O/D pairs had a difference of approximately 3,300 trips. While 12 more O/D pairs had a difference of more than 400 trips. It can be surmised that these larger differences were enough to push the median of the differences away from zero.

6.4 Link Volume Comparison

Further analysis on why the some of the O/D pairs had a large difference is needed. To get a better understanding of this difference the link volumes for the Anniston network will be analyzed. The estimated O/D matrix was input to CUBE for the Anniston network. The network was run using shortest path and all-or-nothing assignment.

The link volumes for the original Anniston network run in CUBE can now be compared to the link volumes generated through the estimated O/D matrix. The assigned traffic between each node from each network was formatted into a three column table and the differences were determined. These differences were plotted using ArcGIS which graphically shows the roadway network of Anniston and can be seen in Figure 16. It should be noted that negative differences less than -350 vehicles were highlighted in cyan and positive differences greater than 350 vehicles were highlighted in purple.

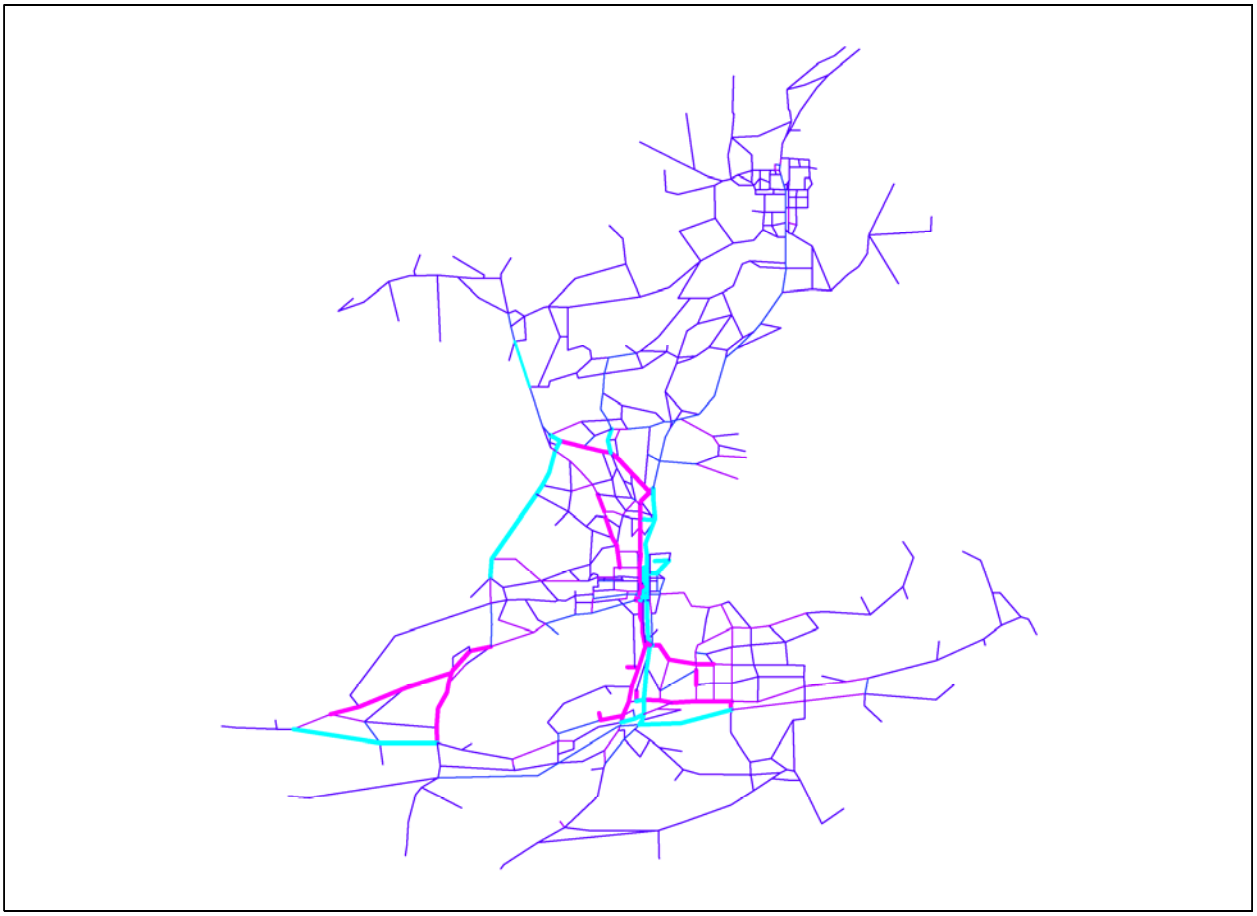


Figure 16 Link Volume Comparison for Anniston

When examining Figure 16 the locations with large link differences were centered near the downtown area and next to the army base. The rest of the link volumes look good.

The downtown area will be explored further. Figure 17 is a zoomed in shot of downtown Anniston. From the figure it can be seen that the two roads are almost a mirror opposite of each other difference wise. Upon further investigation, CUBE was routing more trips along the bottom roadway which is an interstate roadway and the estimation program directed them to the highway above.

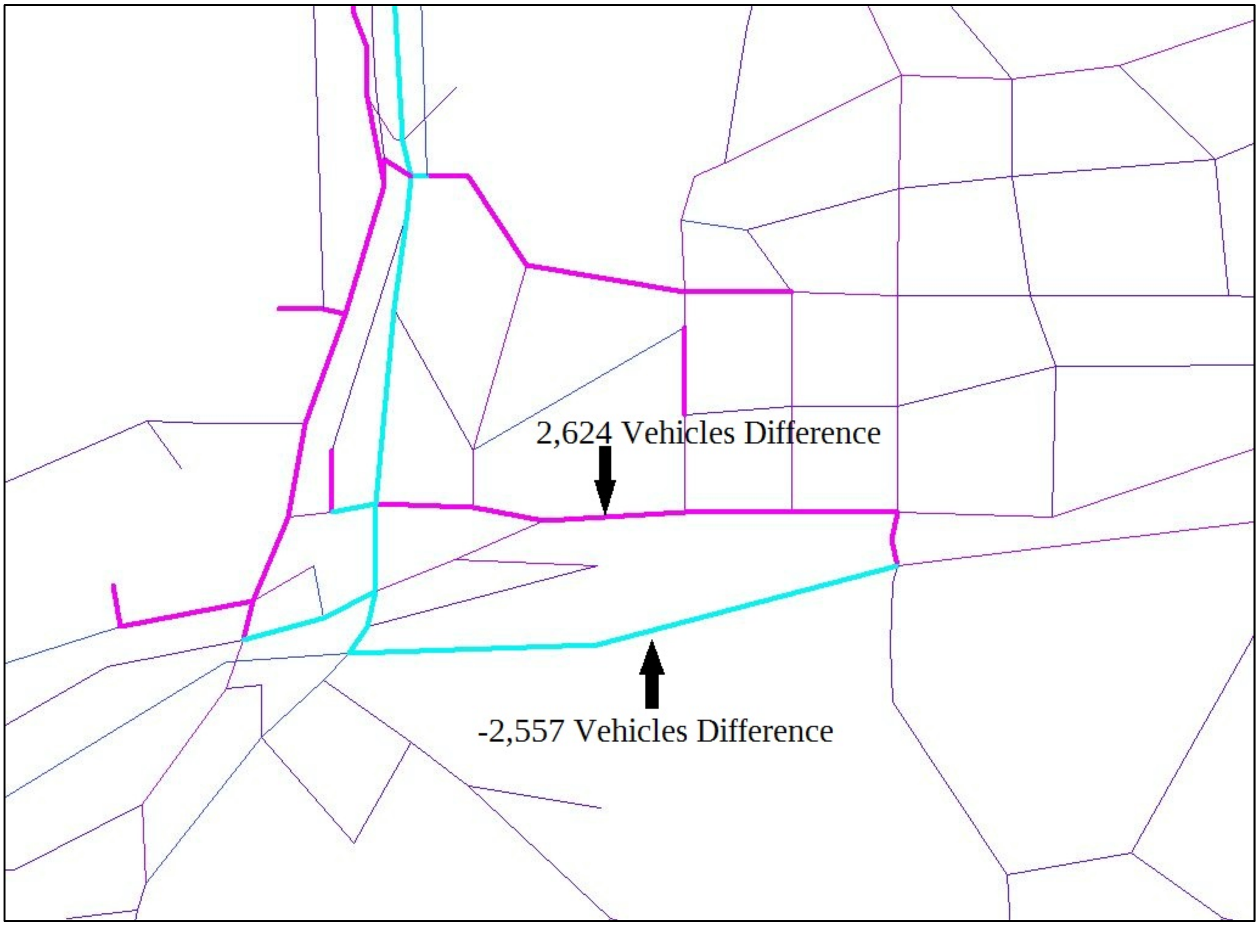


Figure 17 Downtown Link Volume Comparison East/West

Another zoomed in snapshot of the downtown area is shown in Figure 18. Again a near mirror image of vehicle differences appears. The O/D estimation software has routed vehicles along the left road where CUBE has placed them on the right which is a major arterial running through the center of town.

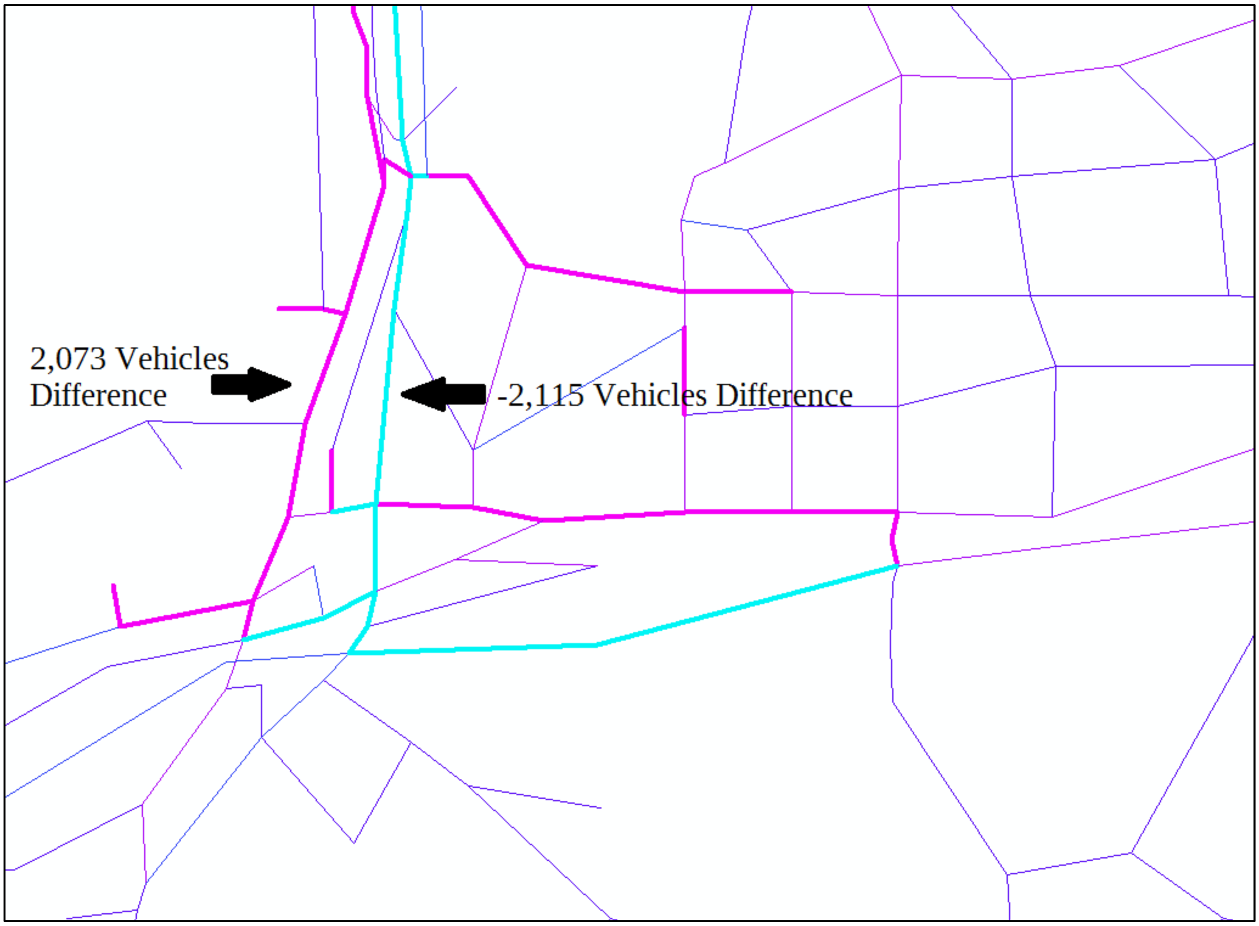
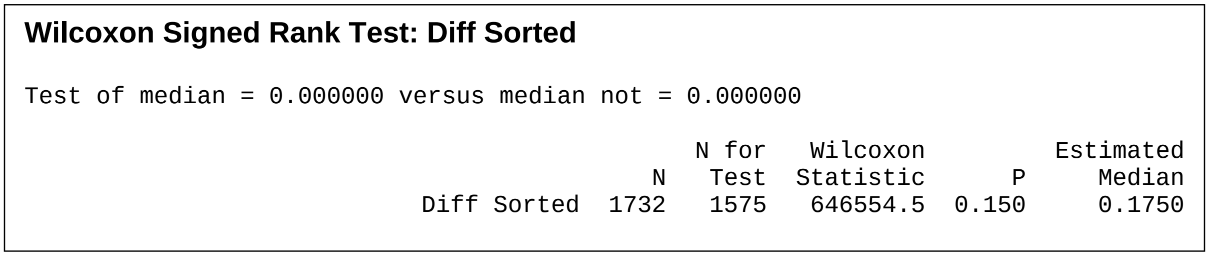


Figure 18 Downtown Link Volume Comparison North/South

Based on the finding shown in Figures 17 and 18 the O/D estimation software does not always route vehicles in the same way CUBE does. The O/D estimation software can only process shortest path based on the distances between nodes. It appears that there might be an underlying setting in CUBE that is routing the trips differently. CUBE is a more intelligent program in that capacity for roadways can be taken into account during traffic assignment. It is possible that this or some other functionality was causing traffic to be diverted in a fashion different from the O/D estimation software in these instances.

Outside the areas previously mentioned the overall forecast for the Anniston network looks good. A Wilcoxon Signed Rank test was run to determine if the original CUBE assigned traffic values were significantly different from the traffic assigned in CUBE using the estimated

O/D matrix. The Wilcoxon Signed Rank test for the link volume comparison is shown in Figure 19.



Wilcoxon Signed Rank Test: Diff Sorted

```
Test of median = 0.000000 versus median not = 0.000000

                                     N for   Wilcoxon         Estimated
                                 N    Test  Statistic      P     Median
Diff Sorted  1732    1575   646554.5  0.150     0.1750
```

Figure 19 Wilcoxon Signed Rank Test for 142 Zone Link Volume Comparison

The results of the Wilcoxon Signed Rank test reveal that the P-value is 0.15 which is above the alpha value of five percent. Therefore, it can be stated that the link volumes produced by CUBE in the original Anniston network do not differ significantly from the link volumes generated in CUBE after inputting the estimated O/D matrix. This means that overall the link volumes generated by this technique are acceptable.

6.5 Summary of Results

The Wilcoxon Signed Rank test showed that CUBE O/D matrix and the estimated O/D matrix were significantly different. This was probably due to the two O/D pairs with differences in trips larger than 3,300 because 95% of the estimated trips between O/D pairs were within 30 trips of the CUBE trips. Also, the mean absolute error was only 6.8 trips meaning that overall the estimated trips fell to within seven trips of the CUBE O/D matrix.

The investigation into the link volumes seems to show that issue may be the way CUBE is routing the traffic through the network. It may not be assigning traffic strictly based on the shortest path/all-or-nothing. When comparing the link volumes from the original CUBE network to the CUBE network based on the estimated O/D matrix the Wilcoxon Signed Rank test shows that the two data sets are not significantly different.

Estimating an O/D matrix from traffic counts using the estimation software is a viable option for small to medium sized communities and has been proven in other test cities.

6.6 Anniston - Zone Aggregation

It is desired to estimate the O/D matrix based solely on the available traffic counts in the city of Anniston. The Alabama Department of Transportation has 167 traffic counts for arterial roadways in the city. However, there are 142 zones in the Anniston network. Based on previous research this means that 568 traffic counts would be required to accurately estimate an O/D matrix. Since this is not the case, the Anniston network of zones will be aggregated together based on the location of traffic counts which are the only source of information for the estimation software.

Based on the locations of the traffic counts two levels of zone aggregation will be attempted. First, a network comprised of 38 zones will be attempted and secondly, a network of only 20 zones will be run in the O/D estimation software. For comparison purposes, the 142 zone CUBE matrix will be aggregated and statistical analysis will be run.

6.7 38 Zone Aggregated Network

The majority of the Alabama Department of Transportation's traffic counts were located along the arterial roadways. There were some collectors as well. The 38 zone network shown in Figure 20 was the largest network that could possibly be made.

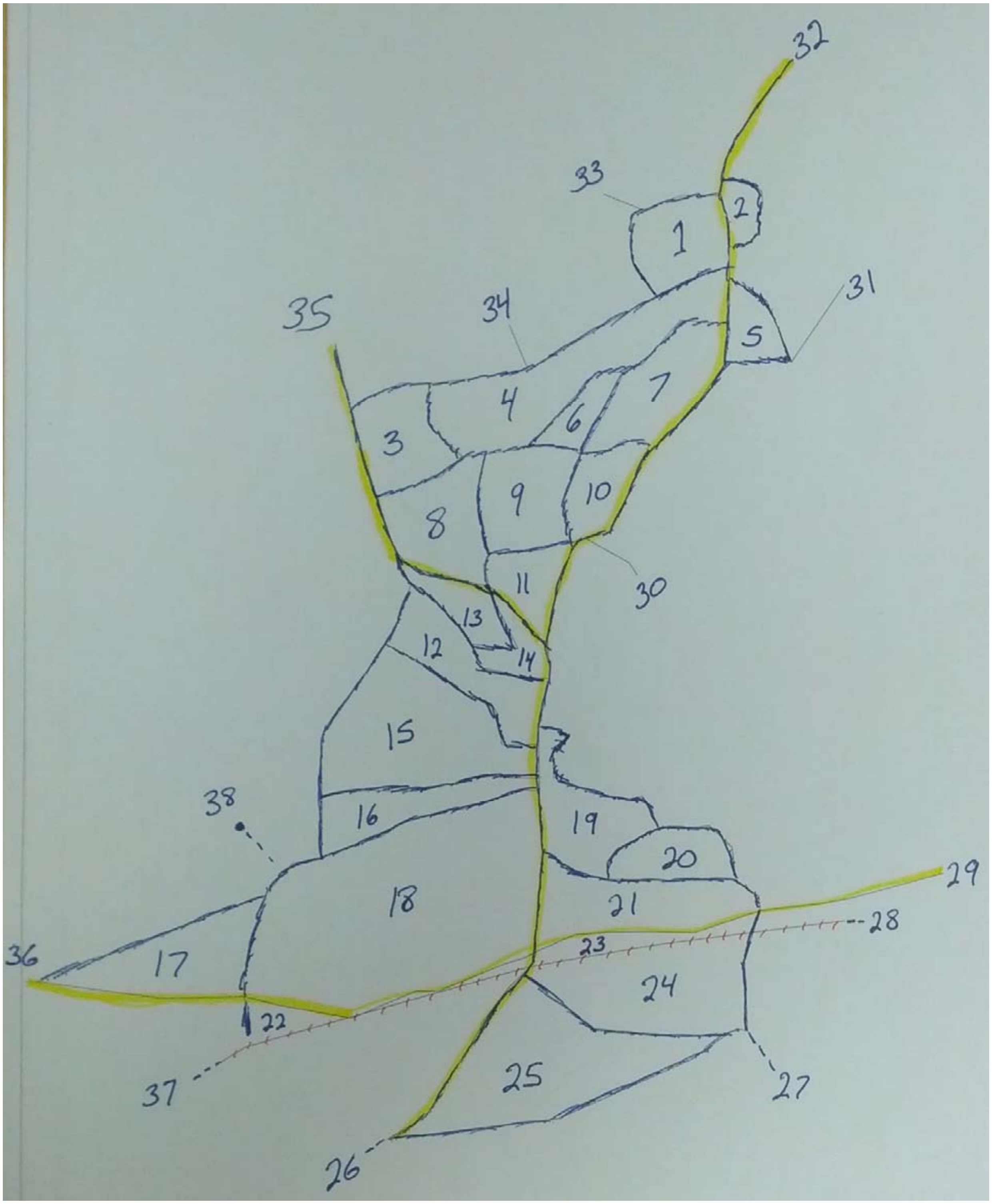


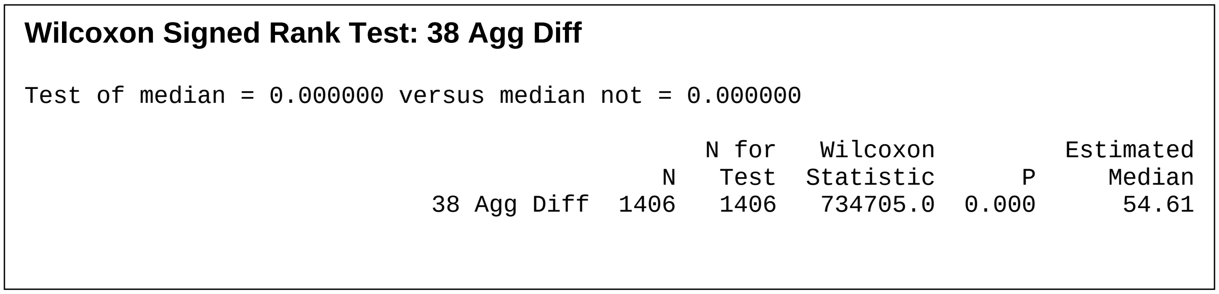
Figure 20 38 Zone Aggregated Network

The network shown in Figure 20 was drawn by hand. Each link in the network was given a starting and ending node number and all roads were assumed to be two-way. A tab delimited text file was prepared based on the nodes, distance between nodes, and the available traffic count. The text file was input to the O/D estimation software and ran.

The CUBE O/D matrix for the 142 zone network was aggregated to compare it to the 38 zone network for statistical analysis. The aggregation began by locating the zones that were within a boundary of the 38 zone network. Then each zone was categorized based on its location and then each origin and destination zone was joined to its new zone name. Finally, all the newly joined O/D trip pairs were summed.

Both the CUBE aggregated O/D matrix and the estimated O/D matrix were formatted into a 3 column O/D trip table for purposes of statistical analysis.

The results of the comparison between the CUBE O/D matrix and the estimated O/D matrix reveal that 22% of the trips are within a value of 15 and 35% are within 30 trips. The root mean square error is 668.41 and the mean absolute error is 264.83. A Wilcoxon Signed Rank test was run and is shown in Figure 21.



```
Wilcoxon Signed Rank Test: 38 Agg Diff

Test of median = 0.000000 versus median not = 0.000000

                                   N for   Wilcoxon          Estimated
                              N    Test   Statistic      P      Median
            38 Agg Diff    1406    1406    734705.0  0.000       54.61
```

Figure 21 38 Zone Aggregated Comparison

The results of the Wilcoxon Signed Rank test disclose the P-value to be zero which is less than the alpha value of five percent. Therefore, it must be concluded that the two aggregated zone networks are significantly different.

The results for the 38 zone aggregation were poor with the datasets being shown to be significantly different and only 35 percent of the estimated trips being within 30 trips of the corresponding CUBE O/D trip pairs.

Perhaps aggregating larger zones will yield better results.

6.8 20 Zone Aggregated Network

The Anniston network was aggregated down to 20 zones based on the traffic counts and the location of industry and residences. The 20 zone network is shown in Figure 22.

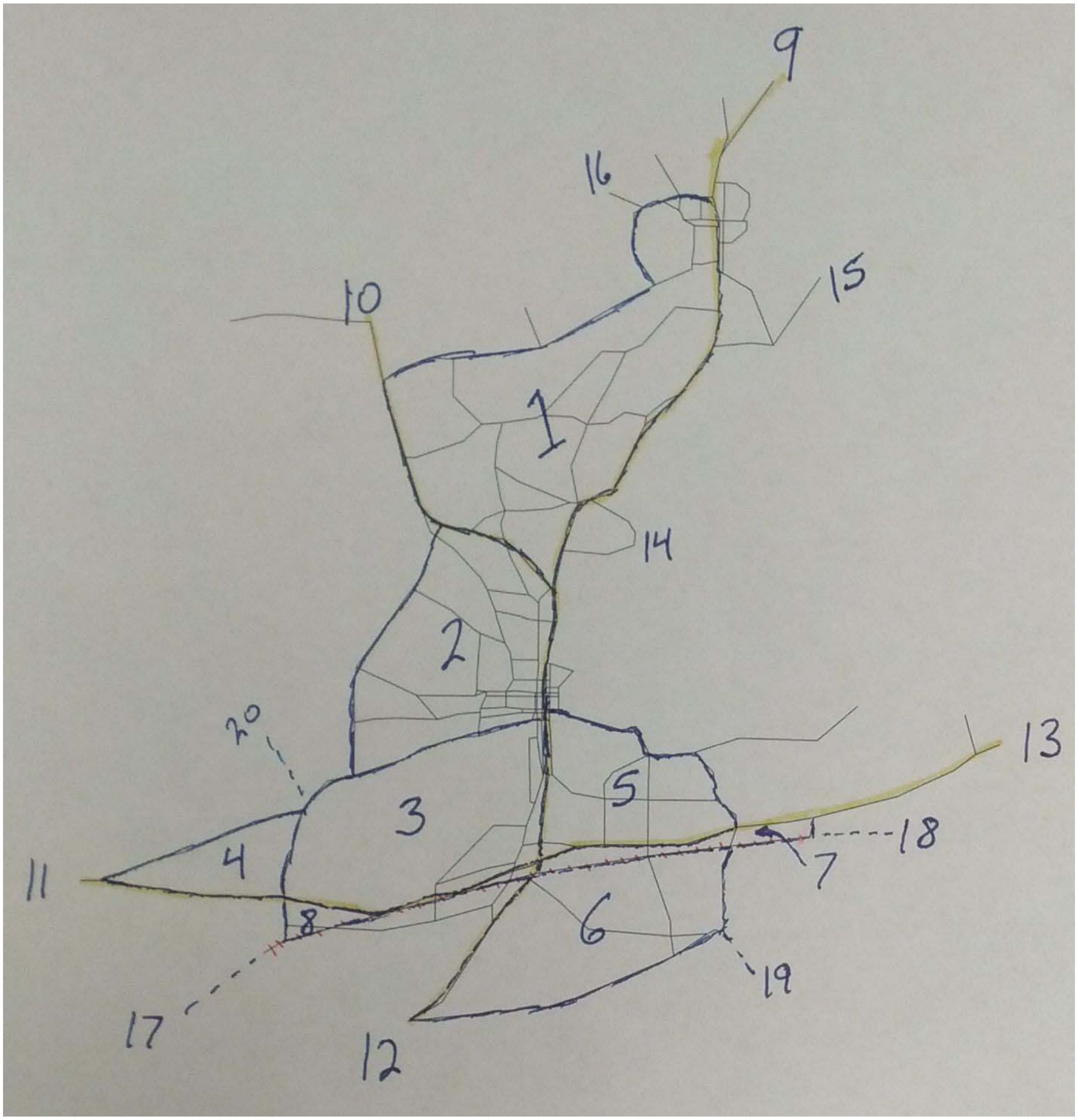
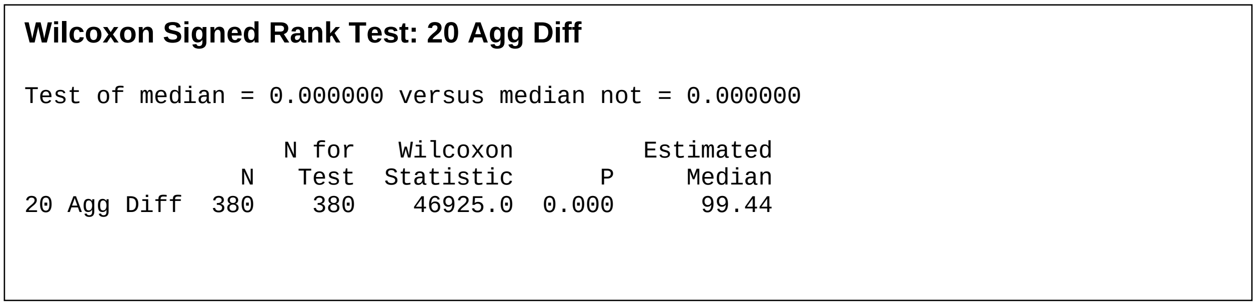


Figure 22 20 Zone Aggregated Network

The 20 zone network shown in Figure 22 was designed by hand. Each link in the network was given a starting and ending node number and all roads were assumed to be two-way. A tab delimited text file was prepared based on the nodes, distance between nodes, and the available traffic count. The text file was input to the O/D estimation software and ran.

The CUBE O/D matrix for the 142 zone network was aggregated to compare it to the 20 zone network in a fashion similar to the 38 zone aggregated network. Then the CUBE aggregated O/D matrix and the estimated O/D matrix were formatted into a 3 column O/D trip table for purposes of statistical analysis.

The results of the comparison between the aggregated CUBE O/D matrix and the estimated O/D matrix reveal that 12 percent of the trips are within a value of 15 and 18 percent are within 30 trips. The root mean square error is 1,220.81 and the mean absolute error is 558.58. The Wilcoxon Signed Rank test was run on the differences and the results are shown in Figure 23.



Wilcoxon Signed Rank Test: 20 Agg Diff

```
Test of median = 0.000000 versus median not = 0.000000

                  N for   Wilcoxon         Estimated
               N   Test  Statistic      P     Median
20 Agg Diff  380    380    46925.0  0.000      99.44
```

Figure 23 20 Zone Aggregation Comparison

The results of the Wilcoxon Signed Rank test disclose the P-value to be zero which is less than the alpha value of five percent. Therefore, it must be concluded that the two aggregated zone networks are significantly different.

The results for the 20 zone aggregation were poor with the datasets being shown to be significantly different and only 18 percent of the estimated trips being within 30 trips of the corresponding CUBE O/D trip pairs.

6.9 Summary

The results for the aggregated zone networks were shown to be significantly different from the CUBE aggregated network. This may be due to the fact that there were only limited traffic counts and most of them were on major streets. This means that there will be less information for the estimation software to generate an O/D matrix.

The results for the 20 zone aggregated network were worse than the 38 zone aggregated network. This can probably be explained by the fact that larger aggregated zones had less traffic count information because one zone had the potential to engulf a large number of traffic counts.

The estimation software does not have the ability to estimate intrazonal trips. This can lead to lost trips in the network.

7 Chapter Model development in Visual C

This chapter lays out the model description coded in Visual C and how it can be implemented in any community.

7.1 Model Description

The model developed in Visual C programming environment, can be run under LINUX or WINDOWS operating system. The model first reads the input file and arranges the order so that bi directional links will be one after another while one way will have single row information. Then model scans arranged input file and reads the starting zone till the ending zone to find the shortest route following the Dijkstra's algorithm where each route presents one origin destination pair with connecting nodes of minimum distances.

Dijkstra’s algorithm is a graph search algorithm that solves the single source shortest path problem for a graph with nonnegative edge path costs, producing a shortest path tree. This algorithm finds the path with lowest cost (i.e. the shortest path) between the vertex and every other vertex. For example, if the vertices of the graph represent cities and edge path costs represent driving distances between pairs of cities connected by a direct road, Dijkstra’s algorithm can be used to find the shortest route between one city and all other cities. For the calculation infinite value is given to all Vertices. Distances between the connecting nodes and the starting node replace the predefined infinite values. Nodes with minimum distance can be selected and the total distance calculation that is related to the connectivity with starting node. This process will be continued till the destined vertex can be reached. [38]

However, there is some variation in this algorithm to accommodate the requirement of avoiding the shortest route passing through other zones. The program than implements Ortuzar and Willumsen methodology to create OD matrix from traffic counts by following all or nothing assignment.

7.2 How to execute the model

Steps to execute the program correctly under LINUX operating system can be given as follows:

1) Must have GCC compiler. To determine if the computer system has GCC compiler, typing "GCC -V" command in command prompt. This displays the version of GCC compiler installed in your system.

2) Compile the source code "FINAL.C' using command "GCC -O FINAL FINAL.C". It creates an executable file "FINAL".

3) To execute program type "./FINAL <Input text file>".

Steps to execute the program correctly under WINDOWS operating system:-

To run the code, click final.exe file. Then type the name of input file.

Properties of input file:

The input text file should have both links present in it if link is bidirectional. Given a link from "a to b" is not implied as there is also a link from "b to a". The user has to explicitly mention both the links in input file.

For example,

```
20      1 2 3 6 8 9 10 11 23 25 26 27 28 29 31 32 33 35 36 37
1       142
301     836
A       B       Dist            VOLUME
354     355     0.68008         19500
```

355	354	0.68008	19500
351	364	2.65984	19100
364	351	2.65984	19100
489	495	0.62000	17550
495	489	0.62000	17550
316	351	5.28999	16650
351	316	5.28999	16650
599	830	1.40000	14550
830	599	1.40000	14550
629	630	0.24000	12250
629	632	0.64000	12250
630	629	0.24000	12250
632	629	0.64000	12250
633	637	2.16000	11350
637	633	2.16000	11350
651	652	1.92000	11350
652	651	1.92000	11350
755	758	0.24000	11300
758	755	0.24000	11300
680	681	2.08000	10950
681	680	2.08000	10950
607	675	3.44000	10550
675	607	3.44000	10550

353	354	0.36996	10300
354	353	0.36996	10300

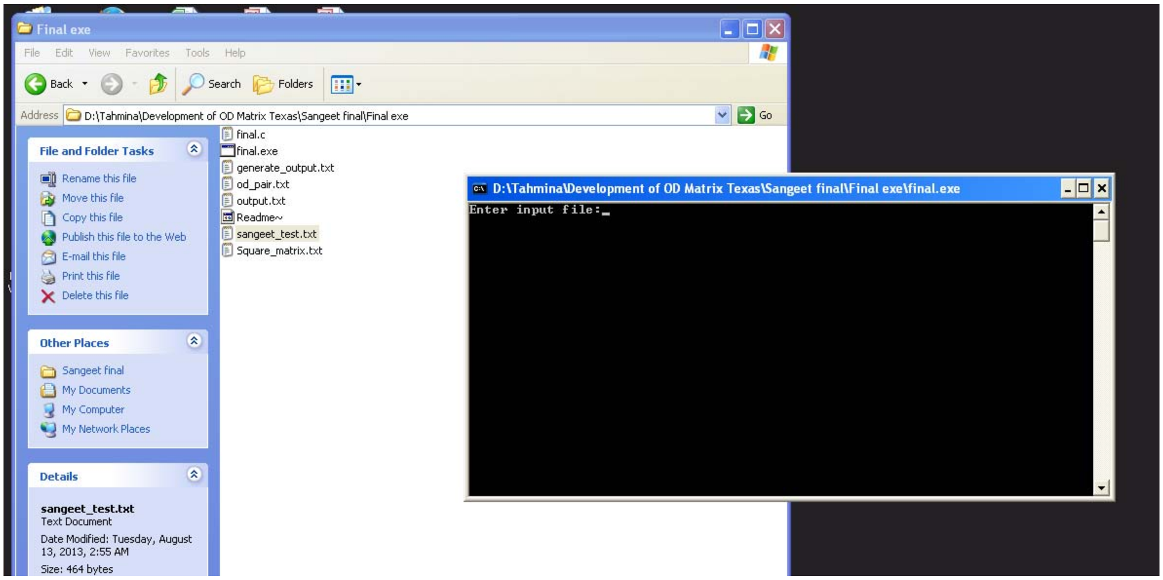
Where, the value of 20 in the first line represents the total number of external zones followed by the actual number of the external station for the study area. It can be noted that the link columns at the end of the spread sheet are likely to satisfy the observed counts. As this study is concerned with the external stations, link columns must be rearranged in such a way so that all external counts will be placed at the end of the initialization and iteration spread sheets.

The values of 1 to 142 are used to declare the starting and ending of zone number that must be continuous. The values of 301 to 836 are used to declare the starting and ending of node numbers that may not have continuous numbering.

A is the starting node/zone; B is the ending node/zone. Dist is the distance between nodes; Distance can be replaced by travel time. Volume is the traffic count of the link A-B.

Steps to run the model are as follows:

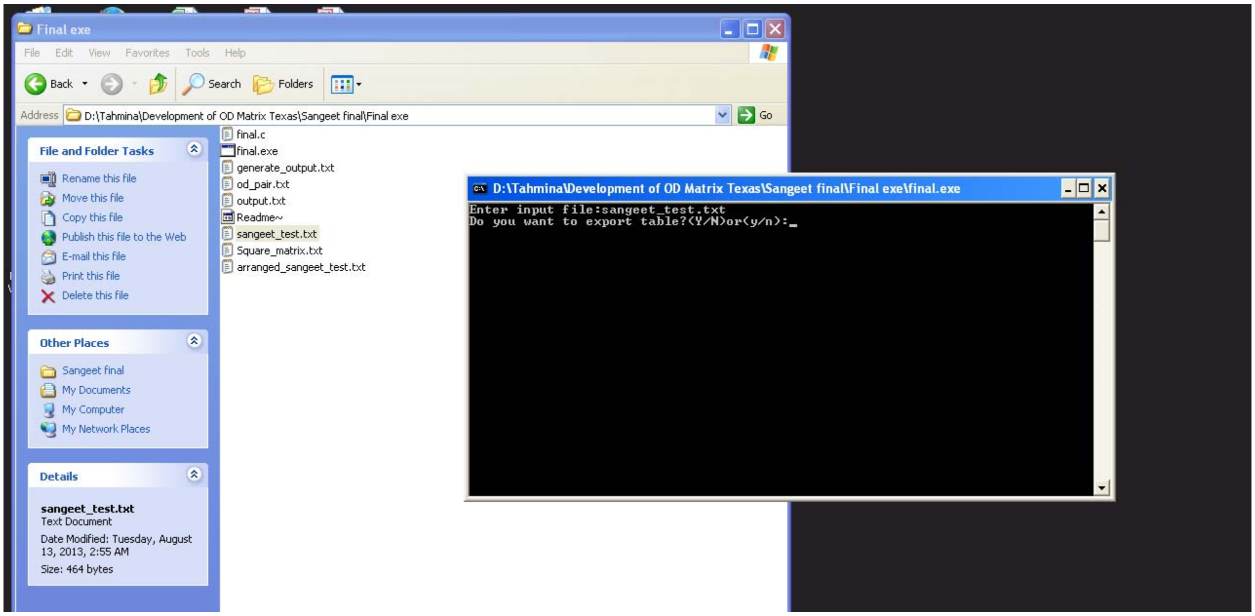
1) The user is asked to type the name of the input file they are interested in (shown in the following Figure). The user should type the name with proper extension. Everything needs to be in same directory or folder.



```
Enter input file:_
```

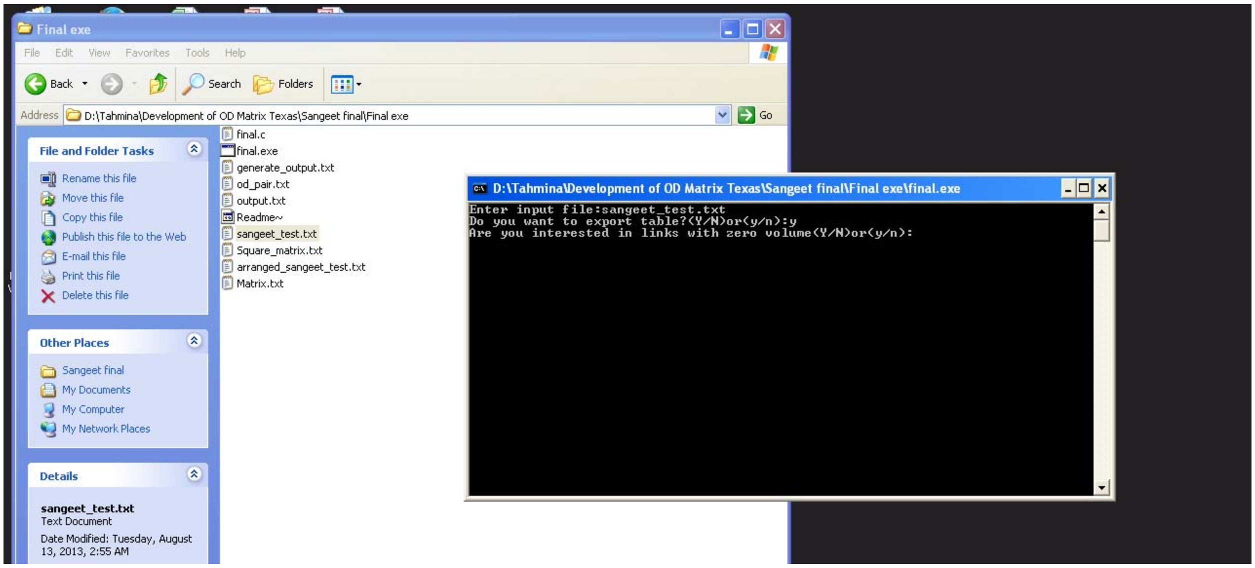
Figure 24 Input file Window

2) In the second step of the program, the user is asked if he/she wants to export the zeroes and ones matrix in "matrix.txt file". If option "y/Y" is typed, user is asked whether he/she is interested in links with zeroes volume to be exported. This steps creates matrix.txt file which could be copied to an excel file for further analysis. The steps bring up the following windows.



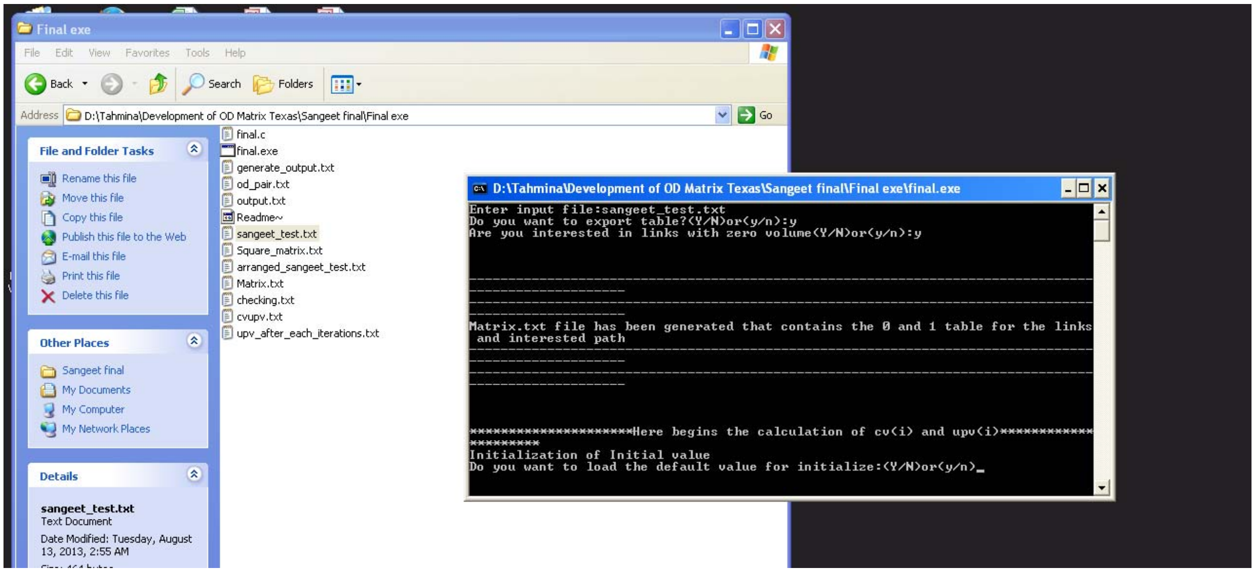
```
Enter input file:sangeet_test.txt
Do you want to export table?(Y/N)or(y/n):_
```

Figure 25 Window to permit exporting of 0 & 1 table



```
Enter input file:sangeet_test.txt
Do you want to export table?(Y/N)or(y/n):y
Are you interested in links with zero volume(Y/N)or(y/n):
```

Figure 26 Window asking to use links with zero volume



```
Enter input file:sangeet_test.txt
Do you want to export table?(Y/N)or(y/n):y
Are you interested in links with zero volume(Y/N)or(y/n):y



--------------------------------------------------------------------------------
--------------------
--------------------------------------------------------------------------------
--------------------
Matrix.txt file has been generated that contains the 0 and 1 table for the links
 and interested path
--------------------------------------------------------------------------------
--------------------
--------------------------------------------------------------------------------
--------------------



*********************Here begins the calculation of cv(i) and upv(i)************
*********
Initialization of Initial value
Do you want to load the default value for initialize:(Y/N)or(y/n)_
```

Figure 27 Window to permit the using of default initialization values

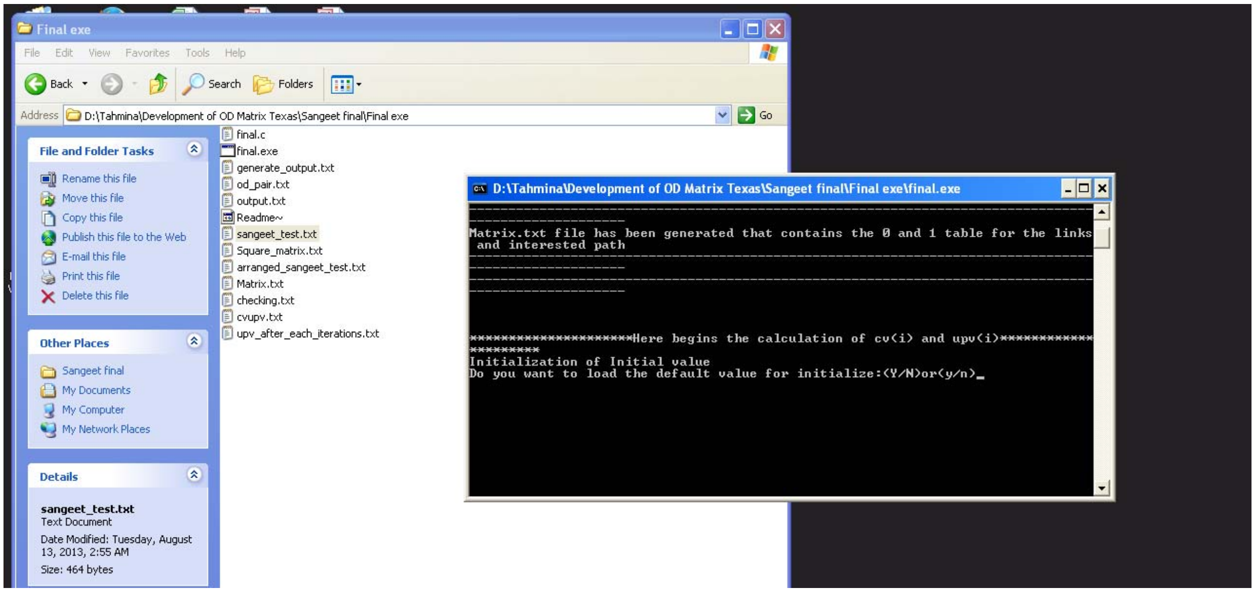
3) In the third step of the program, the user id asked if default value of initialization (1 for all OD pair) or values specified in some file should be used for the calculation of CV and UPV. The user should give the proper filename for the second option. And the file must have the following order of OD pair to define their own initialization value (shown in the following Figures).

For example, a 4 by 4 zonal movement can be presented as follows:

Table 10 4 by 4 sample initialization other than standard

OD pair	Initialization
1to2	1
1to3	0
1to4	0
2to1	1
2to3	0
2to4	0
3to1	1
3to2	1
3to4	1

The text file should contain one column that's the rightmost column of the above table.



D:\Tahmina\Development of OD Matrix Texas\Sangeet final\Final exe

final.c
final.exe
generate_output.txt
od_pair.txt
output.txt
Readme~
sangeet_test.txt
Square_matrix.txt
arranged_sangeet_test.txt
Matrix.txt
checking.txt
cvupv.txt
upv_after_each_iterations.txt

sangeet_test.txt
Text Document
Date Modified: Tuesday, August 13, 2013, 2:55 AM
Size: 464 bytes

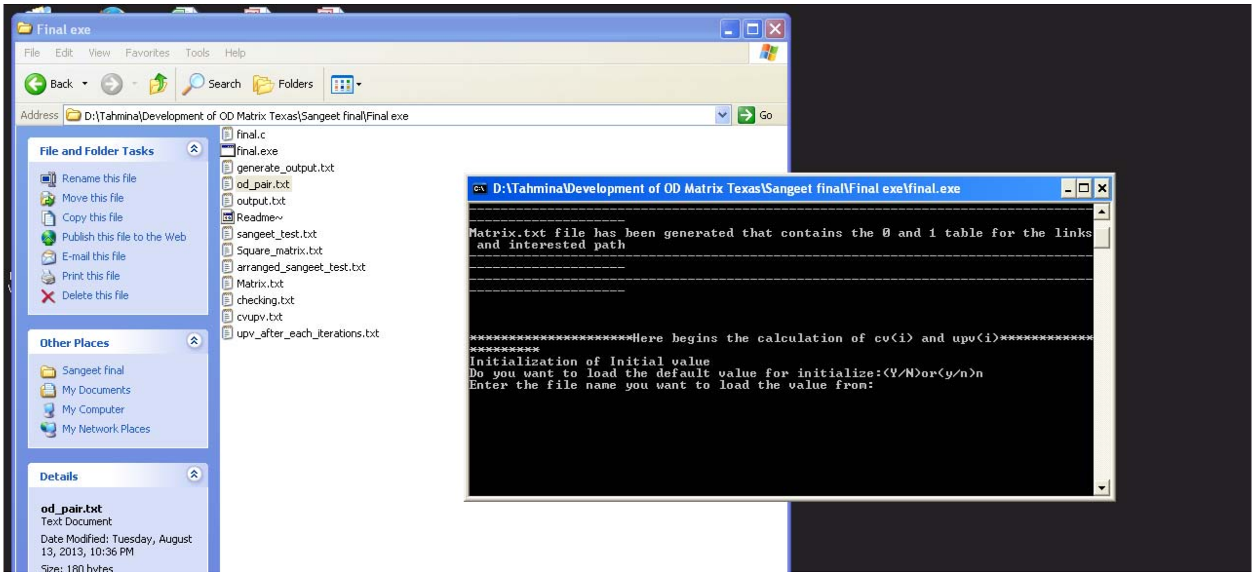
D:\Tahmina\Development of OD Matrix Texas\Sangeet final\Final exe\final.exe

```
--------------------------------------------------------------------------------
--------------------
Matrix.txt file has been generated that contains the 0 and 1 table for the links
 and interested path
--------------------------------------------------------------------------------
--------------------
--------------------------------------------------------------------------------
--------------------



*********************Here begins the calculation of cv(i) and upv(i)*************
*********
Initialization of Initial value
Do you want to load the default value for initialize:(Y/N)or(y/n)_
```

Figure 28 Window to permit the using of different initialization values



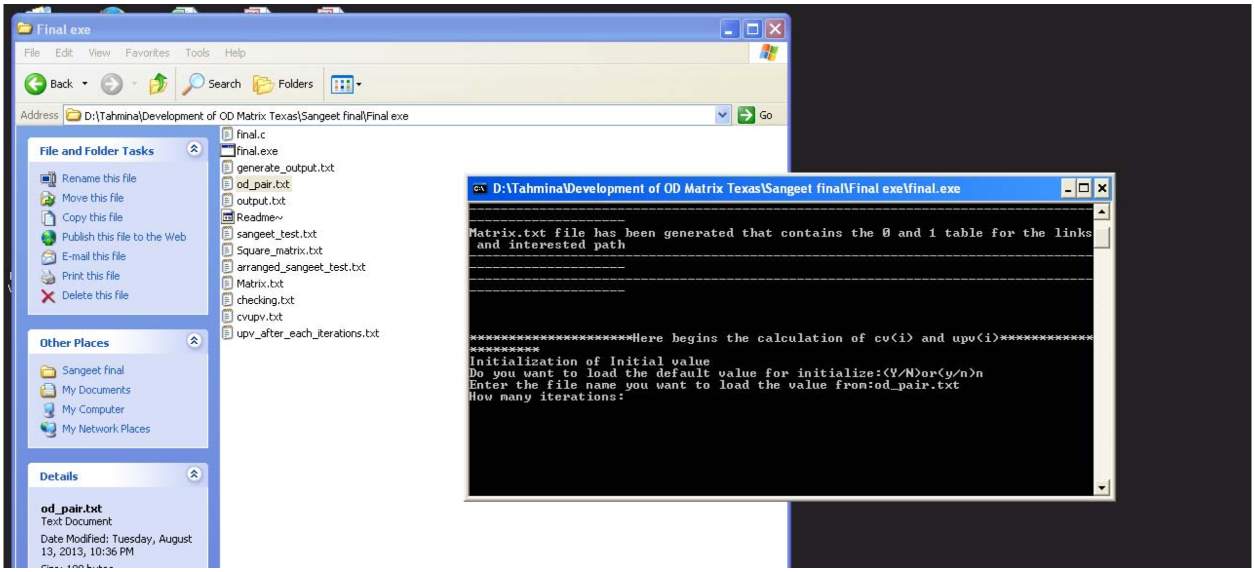
```
--------------------------------------------------------------------------------
--------------------
Matrix.txt file has been generated that contains the 0 and 1 table for the links
 and interested path
--------------------------------------------------------------------------------
--------------------
--------------------------------------------------------------------------------
--------------------



*********************Here begins the calculation of cv(i) and upv(i)************
**********
Initialization of Initial value
Do you want to load the default value for initialize:(Y/N)or(y/n)n
Enter the file name you want to load the value from:
```

Figure 29 Window to provide the name of the file

4) The final step asks user for the total number of iterations the program should run for the calculation of final UPV values.



```
--------------------------------------------------------------------------------
--------------------
Matrix.txt file has been generated that contains the 0 and 1 table for the links
 and interested path
--------------------------------------------------------------------------------
--------------------
--------------------------------------------------------------------------------
--------------------



*********************Here begins the calculation of cv(i) and upv(i)************
**********
Initialization of Initial value
Do you want to load the default value for initialize:(Y/N)or(y/n)n
Enter the file name you want to load the value from:od_pair.txt
How many iterations:
```

Figure 30 Window to provide the number of iterations

Files generated after the completion of running model can be presented as follows:

1) arrange_<Input text file>:- This file contains the opposite links arranged side by side when its two way and leaves one way link as it is.

2) generate_output.txt:- This file contains the shortest path starting from origin zone to destination zone. Destination zone is the second last field. The last field is the total distance of the complete path. These information are presented by line basis.

3) Matrix.txt:- This file contains zero and one table which includes all movements between zones in rows and the links in columns as value 1 if it's used along their shortest routes or else 0.

4) cvupv:- This file contains CV and UPV values for the last iteration of end link. "Current Volume" is the portion of observed/actual traffic count for a certain link that is multiplied by the weighting factor from previous UPV value. Thus the current volume column can be summed to yield a value which is the same as the observed traffic count. Updated volume column compares the current volume column with the last UPV column and keeps the current volume when current volume is not zero; otherwise the program will pull the value from the previous UPV. "Updated Volume" column is shaping up and refining across the link columns to produce a potential solution for the trip matrix.

5) output.txt:- similar to generate_output.txt but in a different format that is required for the program to work correctly. This is intermediate file generated at the mid of program execution.

6) Square_matrix.txt:- This file contains UPV values in square matrix form.

7) upv_after_each_iterations.txt:- This file contains a summation of UPV values after each iteration to keep the track of convergence. The values should be displayed as the total summation of UPV against the number of iteration in a tabular or graphical form.

8) checking.txt:- It contains the difference between actual and modeled volume for all links separately. After assigning the final UPV across the link columns for all movements, the summation of the link column volume can be compared with the actual volume.

7.3 Limitation of the model

The model coded in Visual C can handle the maximum array size of 40,000 in a 32 bit Windows environment. A network with a total number of zones close to 200 can be run in this environment. However, under the LINUX operating system it is possible to run over 500 zones, with a total run time of about 24 hours.

7.4 Caution of preparing input file

The heading of columns must be entered without having any space in between words, or can be presented as a single word. The input file must not contain any blank lines at the end.

8 Chapter Application of Coding in Other Cities

This chapter demonstrates the application of the code in several cities of Alabama to estimate an origin/destination matrix using any network with the assigned traffic supplied by CUBE and investigates the output to determine external trips, attraction production models and pattern of friction factor.

8.1 Selection of Cities

The following cities have been tested in the program under WINDOWS and LINUX operating system depending on their size. Those cities selected had an available CUBE network and traffic counts for selected roadways.

- Anniston
- Gadsden
- Auburn
- Huntsville

It has been found that actual traffic counts are less in number comparing to the total number of zones of a study area. Therefore, initialization and iteration were done using the assigned traffics of each link that were provided by CUBE. Moreover, socio economic data are also available for those cities at zonal level in CUBE network.

8.2 Application

Input files generated from CUBE contain node, time and volume information as per described in the above chapter. Information related to external station, starting and ending of total zones, and nodes were entered at the beginning of the columns. Then the program was executed to find the shortest route, do the initialization with default values and run 150/200 iterations based on the size of the cities consecutively for each city. Finally, eight files were generated as mentioned in the above chapter after the completion of running model.

8.3 Results

External trips:

To determine the external trips, square matrix output file was opened in excel and summarized by EE and EI trips by external zone. Percentage of trips from any external zones is shown in Appendices 1a through 1d. Values presented in Appendices along the diagonal are trips from external to all internal zones. The following Table shows the percent comparison between program and cube outputs of EE and EI movements for Anniston. It can be seen that the difference of two values is within -15% to +15%.

Table 11 Comparison of percent EE and EI values for Anniston

External Stations	Program	Cube	Difference	Program	Cube	Difference
	External-External trips percent			External-Internal trips percent		
122	35.13%	50.13%	15.00%	64.87%	49.87%	-15.00%
123	35.65%	33.74%	-1.91%	64.35%	66.26%	1.91%
124	35.80%	41.75%	5.95%	64.20%	58.25%	-5.95%
125	11.78%	11.58%	-0.20%	88.22%	88.42%	0.20%
126	13.78%	13.29%	-0.49%	86.22%	86.71%	0.49%
127	9.19%	10.30%	1.11%	90.81%	89.70%	-1.11%
128	10.76%	10.68%	-0.08%	89.24%	89.32%	0.08%
129	8.44%	8.57%	0.13%	91.56%	91.43%	-0.13%
130	10.26%	11.37%	1.11%	89.74%	88.63%	-1.11%
131	10.19%	10.17%	-0.02%	89.81%	89.83%	0.02%
132	18.28%	16.92%	-1.36%	81.72%	83.08%	1.36%

External Stations	Program	Cube	Difference	Program	Cube	Difference
	External-External trips percent			External-Internal trips percent		
133	14.44%	17.10%	2.66%	85.56%	82.90%	-2.66%
134	16.02%	24.27%	8.25%	83.98%	75.73%	-8.25%
135	17.03%	25.40%	8.37%	82.97%	74.60%	-8.37%
136	15.82%	24.04%	8.23%	84.18%	75.96%	-8.23%
137	16.47%	17.08%	0.61%	83.53%	82.92%	-0.61%
138	36.59%	51.12%	14.54%	63.41%	48.88%	-14.54%
139	50.70%	50.35%	-0.35%	49.30%	49.65%	0.35%
140	45.70%	47.79%	2.08%	54.30%	52.21%	-2.08%
141	30.14%	27.37%	-2.77%	69.86%	72.63%	2.77%
142	43.34%	46.46%	3.12%	56.66%	53.54%	-3.12%

Friction Factors:

Among eight output files, cvupv and generate_output were used to determine the pattern between percent trips and certain time ranges. Travel time and trip info for any origin destination combination can be found in generate_output and cvupv file respectively. Percent trips can be calculated with respect to the total trips of the study area. Based on a certain time range, summary tables can be created between travel time and percent trips for each city and can be presented as bar diagrams (shown in the following Figures).

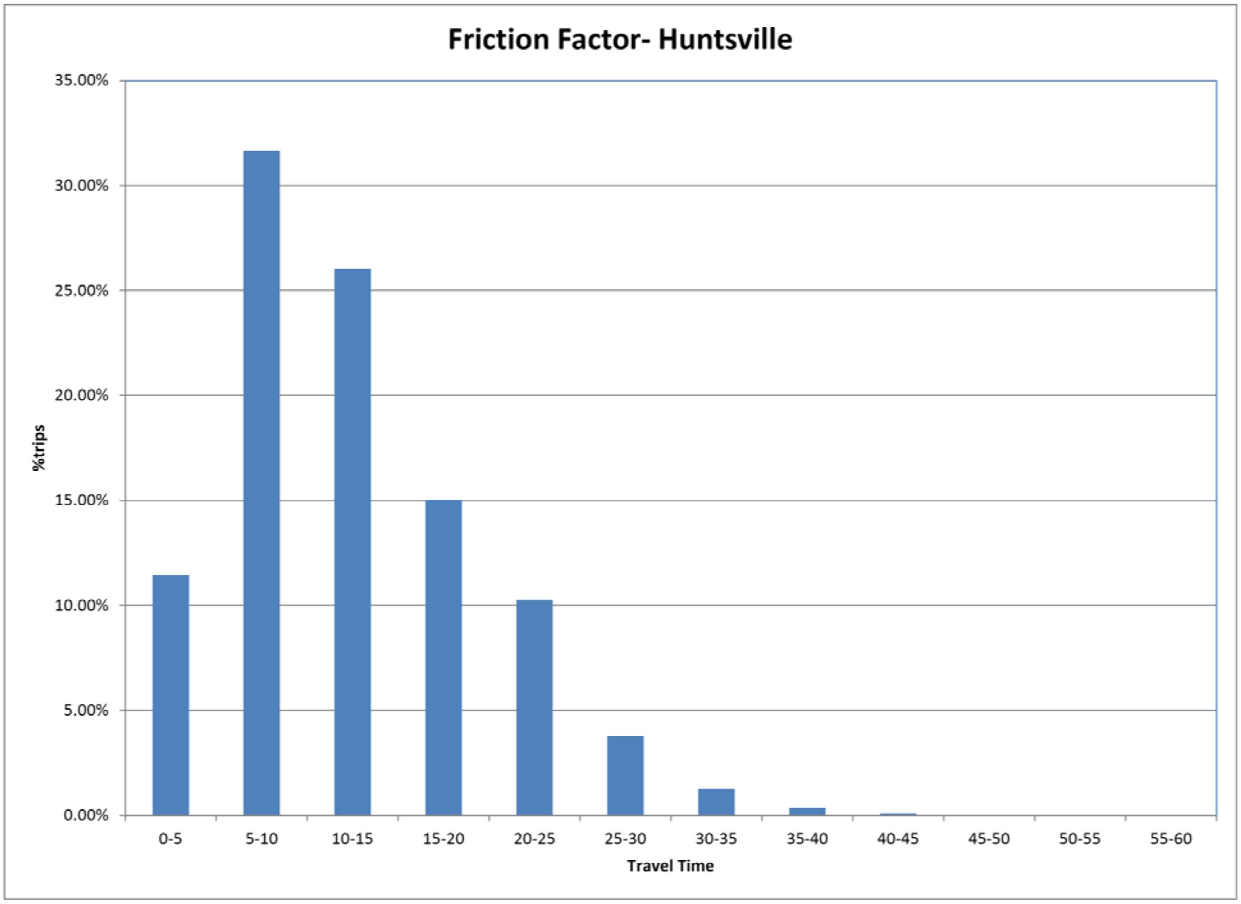


Figure 31 Friction Factor - Hunstville

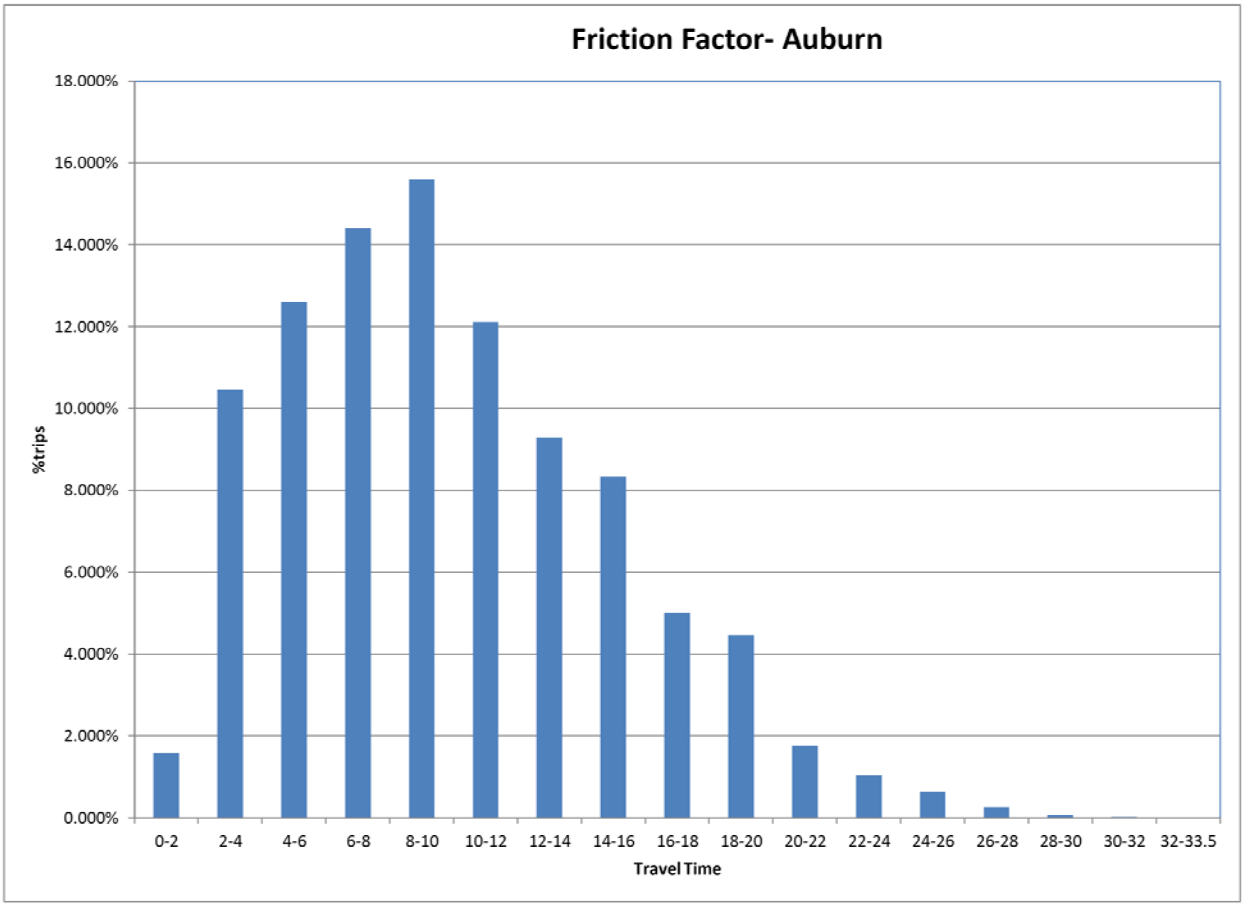


Figure 32 Friction Factor - Auburn

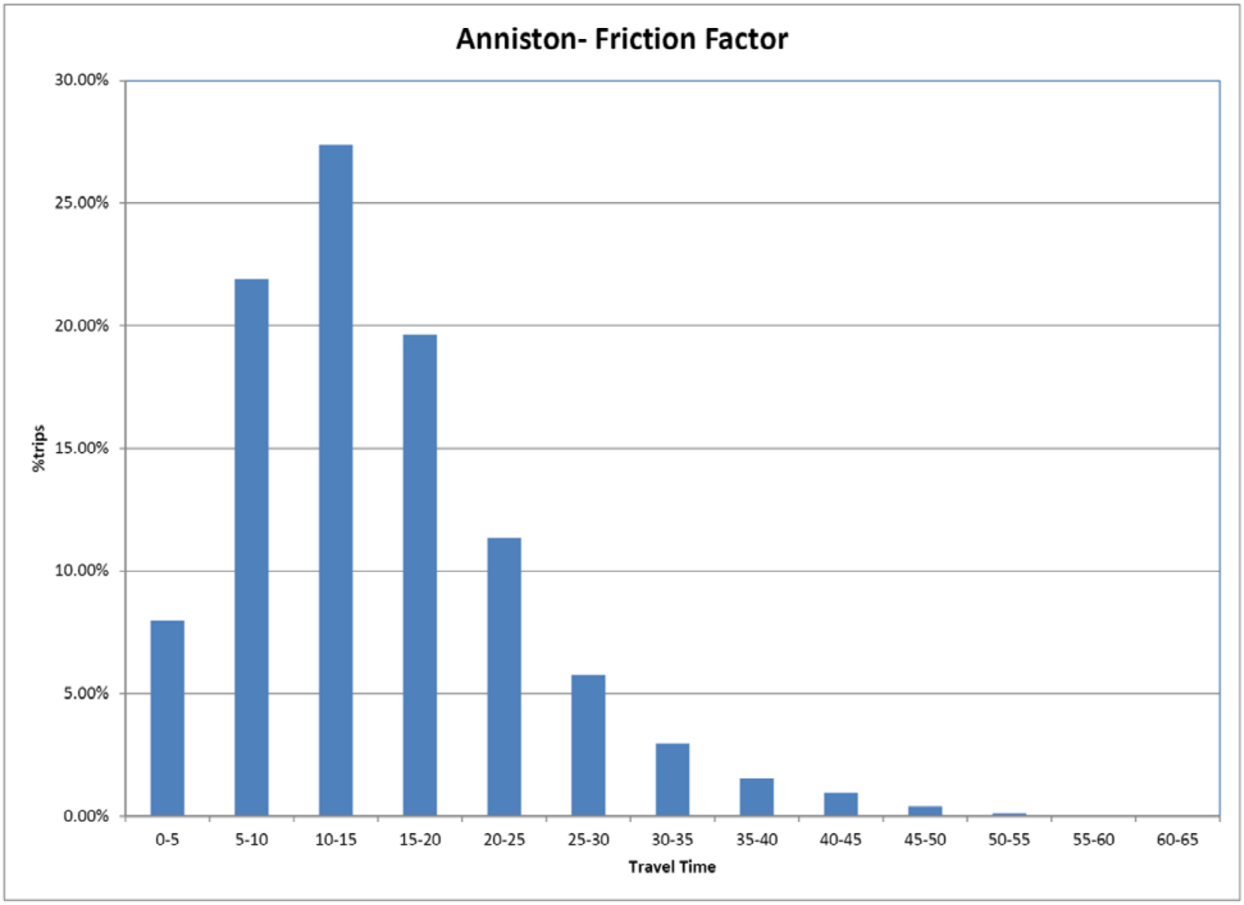


Figure 33 Friction Factor - Anniston

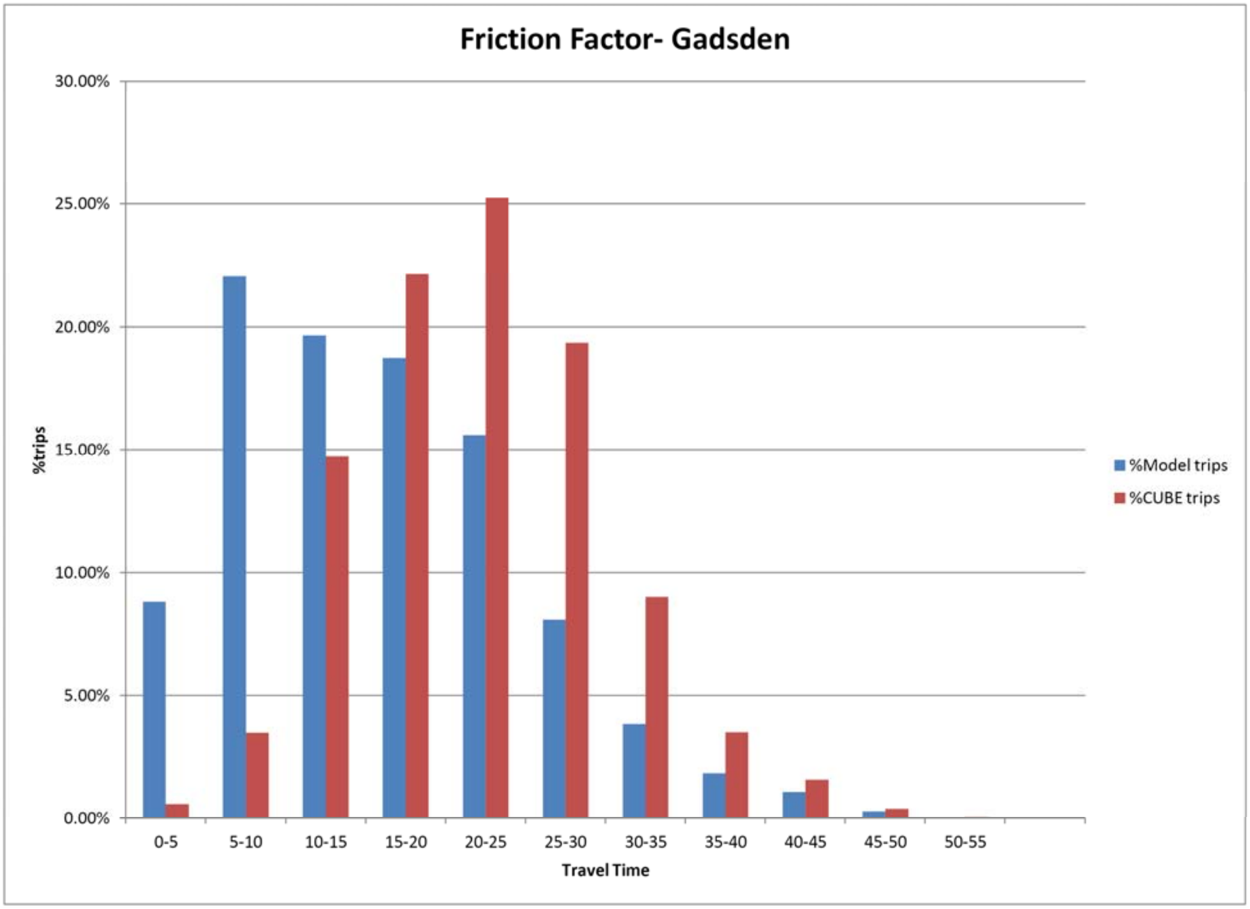


Figure 34 Friction Factor - Gadsden

Figure 34 shows the comparison of two kinds of trips where the positions of peaks are apart from each other. This discrepancy occurs due to the fact that the assignment method being employed in the two models differ.

Trip generation models:

To determine attraction and production models, square_matrix and socio economic files were used. Trips along the row and column by internal zone is summed up to be used as the dependent variable, where row summation means number of trip production and column summation means the number of trip attraction by that particular zone. Trip information can be found in square_matrix file. Independent variables provided from CUBE by internal zone for each city are household (HH), income (I), retail employment (RE), non retail employment (NRE) and school enrollment (SE) information. Regression analysis was done in excel and the regression output has three components:

- Regression statistics table
- ANOVA table
- Regression coefficients table

The parameters of greatest importance are R Square, Significance F and P-value that are the part of above three tables respectively. R Square value is the percentage of variation of dependent variable that can be explained by the regressors or independent variables. Significance F has the associated P-value of F-test, for example if it is less than 0.05, at least one of the coefficients of regression model is not equal to zero. Any coefficient can be statistically insignificant if the corresponding P-value from regression coefficients table is more than α. The following tables present the summary of statistics, ANOVA and regression model tables.

Table 12 Regression output summary

Name of models	Name of Statistical Parameters	Value of Statistical Parameters		
		Huntsville	Auburn	Gadsden
Trip Production	R Square	0.4127	0.7097	0.6241
	Significance F	0.0000	0.0000	0.0000
Trip Attraction	R Square	0.9913	0.9613	0.9497
	Significance F	0.000	0.000	0.000

Table 13 Regression models summary

Name of Models	Name of Regressors	Value of coefficients and associated P-value					
		Huntsville		Auburn		Gadsden	
Trip Production	HH	5.681	0.00	5.817	0.000	7.653	0.00
Trip Attraction	HH	2.837	0.00	3.032	0.000	4.776	0.00
	RE	3.476	0.00	3.964	0.000	6.654	0.00
	NRE	1.560	0.00	2.419	0.000	2.797	0.00
	SE	0.0726	0.09	0.102	0.001	0.259	0.40

It can be seen that production and attraction models are quite different from original models used in CUBE models. The main reason behind this difference is the model square matrix cannot generate any diagonal trip values and the other reason is the assignment method implemented in this study is all or nothing and in CUBE model is equilibrium assignment. However, the models found under this study present strong R Square value for production and school enrollment as insignificant variable in three cases at significance level α of 0.05. The same regression analyses were performed for Anniston cube output and model output. It can be noted that the diagonal values of square matrix from cube were changed into zero to match the model ones. The following tables show the parameters are very close after changing the diagonal values and income is not a significant variable at significance level α of 0.05.

Table 14 Anniston Regression output summary

Name of models	Name of Statistical Parameters	Value of Statistical Parameters	
		Anniston-Cube	Anniston-Model
Trip	R Square	0.440055	0.448828

Production	Significance F	0.00000	0.00000
Trip Attraction	R Square	0.979442	0.978751
	Significance F	0.00000	0.00000

Table 15 Anniston Regression models summary

Name of Models	Name of Regressors	Value of coefficients and associated P-value			
		Anniston-Cube		Anniston-Model	
Trip Production	HH	4.144618	0.00000	4.196502	0.00000
	Income	0.009149	0.21529	0.00924	0.20818
Trip Attraction	HH	2.222489	0.00000	2.303467	0.000000
	RE	6.182059	0.00000	6.114957	0.000000
	NRE	2.378697	0.00000	2.356671	0.000000
	SE	0.53268	0.00000	0.546225	0.000000

Wilcoxon test:

To detect the existence of significant difference of two similar outputs, in our case they are cube square matrix and model square matrix, the Wilcoxon signed-ranks test for the median difference was performed in MINITAB for each city. It has been found that p-value is less than α = 0.05 and the decision is to reject the null hypothesis. Thus it can be concluded that there is evidence of a significant median difference in two trip matrices for any city mentioned in this chapter.

9 Chapter Conclusion

It has been found that estimating the ability to predict through trips using the origin/destination estimation methodology is decent. The study using the data from Texas showed discrepancies within +/- 13%, which is similar to values found when using regression models or synthetic procedures.

The production and attraction models are not that different when regression analyses were performed for Anniston cube output and model output, where that the diagonal values of square matrix from cube were changed into zero to match the model ones. It has been found that the parameters are very close after changing the diagonal values and income is not a significant variable at significance level α of 0.05.

The distribution patterns are same but Figure 34 shows the positions of peaks are apart from each other. This discrepancy occurs due to the different assignment method employed in those two models. However, it can be concluded that friction factors do not match with the actual friction factors.

The resolution of aggregation is important to model a large network as larger aggregated zones are likely to have less traffic count information.

The results of the Wilcoxon Signed Rank test disclose the P-value to be zero which is less than the alpha value of five percent. Therefore, it must be concluded that almost all cities's O/D values are significantly different from that of CUBE except Hartselle.

10 Works Cited

[1] J. P. Wilson, "Evaluation of Origin/Destination Matrix Estimation Techniques to Support Aspects of Traffic Modeling," Huntsville, 2012.

[2] T. Abrahamsson, "Estimation of Origin-Destination Matrices Using Traffic Counts - A Literature Survey," Laxenburg, Austria, 1998.

[3] A. G. Wilson, Entropy in Urban and Regional Modelling, London: Pion, 1970.

[4] H. J. Van Zuylen and L. G. Willumsen, "The Most Likely Trip Matrix Estimated from Traffic Counts," *Transportation Research Board,* vol. 14B, pp. 281-293, 1980.

[5] J. d. D. Ortuzar and L. G. Willumsen, Modelling Transport, 2nd ed., New York, New York: John Wiley and Sons, 1994.

[6] F. Viti, "State-of-Art of O-D Matrix Estimation Problems Based on Traffic Counts and Its Inverse Network Location Problem: Perspectives for Application and Future Developments," 2008, 2008.

[7] S. Nguyen, "Estimation an O-D Matrix from Network Data: A Network Equilibrium Approach," Montreal, 1977.

[8] K. Farhangian and L. LeBlanc, "Selection of a Trip Table Which Reproduces Observed Link Flows," *Transportation Research,* vol. 16B, no. 2, pp. 83-88, 1982.

[9] Y. Iida, T. Sasaki and H. Yang, "The Equilibrium-Based Origin-Destination Matrix Estimation Problem," *Transportation Research,* vol. 28B, no. 1, pp. 23-33, 1994.

[10] K. M. Kockelman, W. S. Travis and C. Xie, "A Maximum Entropy Method for Subnetwork Origin-Destination Trip Matrix Estimation," in *89th Annual Meeting of the Tranportation Research Board*, Washington D.C., 2010.

[11] S. Bera and K. V. Krishna Rao, "Estimation of Origin-Destination Matrix from Traffic Counts: The Stat of the Art," *Eurpean Transport \ Transporti Europei,* no. 49, pp. 3-23, 2011.

[12] C. Fisk, "On Combining Maximum Entropy Trip Matrix Estimation with User-Optimal Assignment," *Transportation Research,* vol. 22B, 1988.

[13] M. G. Bell, C. M. Shield, F. Busch and G. Kruse, "A Stochastic User Equilibrium Path Flow Estimator," *Transportation Research Part C: Emerging Technologies,* vol. 5, no. 3/4, pp.

197-210, 1997.

[14] A. Chen, P. Chootinan and W. M. Recker, "Examining the Quality of Synthetic Origin-Destination Trip Table Estimated by Path Flow Estimator," *Journal of Transportation Engineering,* vol. 131, no. 7, pp. 506-513, 1 July 2005.

[15] A. Chen, P. Chootinan, W. Laabs, L. M. S. and W. Recker, "Modeling Network Traffic for Planning Applications in a Small Community," *Journal of Urban Planning and Development,* vol. 132, no. 3, pp. 156-159, 1 September 2006.

[16] L. Le Cam, "Maximum Likelihood; An Introduction," *International Statistics Review,* vol. 58, pp. 153-171, 1990.

[17] A. A. Evans, "Maximum Likelihood Estimation," San Francisco, 2008.

[18] H. Spiess, "A Maximum Likelihood Model for Estimating Origin-Destination Matrices," *Transportation Research,* vol. 21B, no. 5, pp. 395-412, 1987.

[19] McGraw-Hill, Dictionary of Scientific and Technical Terms, 6th ed., New York, New York: McGraw-Hill Companies, Inc., 2003.

[20] E. Cascetta, "Estimation of Trip Matrices From Traffic Counts And Survey Data: A Generalized Least Squares Estimator," *Transportation Research ,* vol. 18B, no. 4/5, pp. 289-299, 1984.

[21] M. Bell, "Log-linear Models for the Estimation of Origin-Destination Matrices from Traffic Counts," in *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*, Delft, 1984.

[22] M. Bell, "The Estimation of an Origin-Destination Matrix from Traffic Counts," *Transportation Science,* vol. 17, no. 1, pp. 198-217, 1991.

[23] Encyclopedia Britannica, "Bayes's Theorem," Encyclopedia Britannica Inc., 2012. [Online]. Available: http://www.britannica.com/EBchecked/topic/56808/Bayess-theorem. [Accessed 12 September 2012].

[24] N. J. Maher, "Inferences on Trip Matrices from Observations on Link Volumes: A Bayesian Statistical Approach," *Transportation Research,* vol. 17B, no. 6, pp. 435-447, 1983.

[25] A. J. Horowitz and M. H. Patel, "Through Trip Tables for Small Urban Areas: A Method for Quick Response Travel Forecasting," 1998.

[26] Y. Han, "Synthesized Through Trip Model For Small and Medium Urban Areas," Raleigh, NC, 2007.

[27] M. D. Anderson, Y. M. Abdullah, S. E. Gholston and S. L. Jones, "Development of a Methodology to Predict," 2005.

[28] T. D.-T. P. a. P. Divison, "District Traffic Maps - 2011," 2012. [Online]. Available: http://ftp.dot.state.tx.us/pub/txdot-info/tpp/traffic_counts/2011/bry_base.pdf. [Accessed 19 October 2012].

[29] T. D.-T. P. a. P. Divison, "District Traffic Maps - 2011," 2012. [Online]. Available: http://ftp.dot.state.tx.us/pub/txdot-info/tpp/traffic_counts/2011/bry_supp.pdf. [Accessed 19 October 2012].

[30] S. Farnsworth, T. Voigt and D. Borchardt, "Alternative Methods for External Travel Surveys," 2011.

[31] M. Martchouk and J. D. Fricker, "Through-trip Matrices Using Discrete Choice Models: A Planning Tool For Smaller Cities," *TRB,* 2009.

[32] E. S. Talbot, M. W. Burris and S. Farnsworth, "Estimating Through Trip Travel without External Surveys," *Transportation Research Board,* 2011.

[33] United States Census Bureau, "State and County QuckFacts," 2010. [Online]. Available: http://quickfacts.census.gov/qfd/states/01/0133448.html. [Accessed 10 October 2012].

[34] City of Hartselle, "Community Profile," 2012. [Online]. Available: http://www.hartselle.org/wp/about/demographics/. [Accessed 10 October 2012].

[35] Calhoun County Chamber of Commerce, "Anniston: The Model City," [Online]. Available: http://www.annistonal.gov/pages/?pageID=25. [Accessed 1 October 2013].

[36] W. W. Hines, D. C. Montgomery, D. M. Goldsman and C. M. Borror, Probablilty and Statistics in Engineering, 4th ed., New York, New York: John Wiley and Sons, Inc., 2004.

[37] Minitab Inc., *Minitab Statistical Software, Release 16,* State College, Pennsylvania, 2010.

[38] A. Arora, "slideshare," [Online]. Available: http://www.slideshare.net/ami_01/dijkstras-algorithm-7716656. [Accessed 22 August 2013].

[39] A. G. Wilson, Entropy in Urban and Regional Modelling, London: Pion, 1970.

[40] T. Abrahamsson, "Estimation of Origin-Destination Matrices Using Traffic Counts - A Literature Survey," International Institute for Applied Systems Analysis, Laxenburg, Austria, 1998.

[41] S. K. R. K. V. Bera, "Estimation of Origin-Destination Matrix from Traffic Counts: The Stat of the Art," *Eurpean Transport \ Transporti Europei(49),* pp. 3-23, 2011.

[42] J. d. Ortuzar and L. G. Willumsen, Modelling Transport, 2nd ed., New York,, New York, United States of America: John Wiley and Sons., 1994.

[43] K. M. Kockelman, W. S. Travis and C. Xie, "A Maximum Entropy Method for Subnetwork Origin-Destination Trip Matrix Estimation," *Transportation Reserach Board,* 2010.