5-2006

A Comparison of Four-Step Model and Path Flow Estimator for Forecasting Network Flow: A Case Study of Cache County in Utah

Siddareddy B. Pedaballi
Utah State University

A Comparison of Four-Step Model and Path Flow Estimator for Forecasting Network Flow: A Case Study of Cache County in Utah

by

Siddareddy B. Pedaballi

A master's report submitted in partial fulfillment

of the requirements for the degree

of

MASTER OF SCIENCE

in

Civil and Environmental Engineering

APPROVED:

UTAH STATE UNIVERSITY
Logan, UT

2006

ABSTRACT

Development of alternative methodologies for travel demand modeling has become important in recent years due to the lack of resources for small and medium communities to adopt conventional four step travel models. Many researchers have proposed alternative tools of travel demand modeling for these communities. But majority of them still require large amount of data and technical sophistication.

In this study, the Path Flow Estimator (PFE) is used to estimate the network traffic of Cache County. PFE estimations are based on the collected traffic counts, vehicle production and attractions of zones estimated using Institute of Transportation Engineers (ITE) trip generation rates. The daily trips of the four step model are converted into peak hour trips using a common peak hour factor. The PFE estimations are then compared with the estimations of the four-step model.

The differences in vehicle trip production and attraction estimates by both models are due to the use of a common peak hour factor to reduce the daily trips of the four step model to peak hour trips. These differences got dis-aggregated spatially in the trip table estimations, leading to similar trip table estimations. The trip length frequency distributions of both models accounting for congestion have similar distribution. The average trip length estimated by PFE is 2.36 minutes more than the four-step model. Further, PFE has estimated link flows better than the four step model on the links having ground counts due the traffic counts constraint in PFE's formulation. Low percentage root mean square errors between PFE and the four step model estimated link flows show that PFE has replicated the link flows satisfactorily. The four-step model has overestimated the link flows on Main Street due to the use of daily trips V-C ratio in the zones with high demand due to the slight overestimation of trips by ITE trip generation rates for those zones. On the screenlines identified in the original four step model also PFE has performed better than the four step model.

The results of this study show that Path Flow Estimator can be a useful tool in traffic demand modeling for small and medium communities. But PFE should not be used for large communities because of the difficulty in arriving proper trip generation rate for complex land uses. PFE does not distinguish between an internal zone and an external zone.

(124 pages)

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction and purpose

1.1 Introduction

The life of an urban area depends on its transportation system, and a healthy urban economy requires that transportation be smooth and efficient. Recognizing the importance of this fact most urban areas have developed travel demand models (TDMs) to forecast traffic conditions and to plan for the transportation systems accordingly. Travel demand models (TDMs) are used to forecast traffic flows on the transportation system. TDMs are used by consulting firms, metropolitan planning organizations (MPOs), and state and local departments of transportation (DOT) to identify probable deficiencies of the transportation systems in a future year. These agencies also use these models to evaluate the impact of alternative transportation solutions for development of long and short-range transportation plans. TDMs are also used for pollution emission estimates and for congestion management system.

Current practice in travel demand modeling is through the conventional four-step models. The traffic forecasts of the four-step models are based on forecasted land use, demographics, and travel patterns unique to the region. Regions applying the travel demand models range in size from cities, counties to metropolitan areas. The models could cover an entire state. The inputs to the travel demand model include the transportation network and variables such as population, employment, households, dwelling units, trip rates, transit fares, etc. Among other statistics and reports, outputs from the model are maps of the transportation system with traffic volumes for every roadway segment included in the model.

With the four-step approach, the number of trips produced from and attracted to each sub-area of the study area is estimated from the regional travel surveys and land use data. The estimated trip production and attraction of each sub-area are then "distributed" to other areas in the region, based on a gravity model, to form a trip table, also known as the origin-destination (O-D) matrix. In cases where other modes of transportation are present in the study area, a trip table is split into multiple ones, each for a specific mode. Each trip table is "assigned" to the corresponding network to estimate and forecast traffic volume on each link. Observed link volumes are collected to calibrate and validate the modeling results.

As explained above, the four-step models require a large amount of data, travel surveys, and technical staff for operation and maintenance. For large communities, the collection of data and availability of experts are not a big hindrance but for small and medium communities where the resources are scarce the development and maintenance of travel demand model is a challenge. Yan (1998) noted that many smaller communities usually do not have sufficient resources to conduct travel surveys, nor to house technical staff for model development and maintenance. Without data from a travel survey in the study area, trip generation rates of various land use zones are often "borrowed" from published data such as Trip Generation (ITE, 1998) of the Institute of Transportation Engineers (ITE) or reports of travel surveys performed in other areas. The unavailability of data on

Trip Length Frequency Distribution (TLFD) of local travelers often forces modelers to skip the calibration of trip distribution models. Instead, calibration and validation of the overall model are often carried out by altering the friction factors and adding k-factors, in a trial-and-error fashion, to the trip distribution model so the results of traffic assignment would match traffic counts on selective screen lines and critical links. The calibration process is usually a lengthy process and the resultant models often contain many factors that do not have the necessary behavioral foundation established from travel surveys. The practical difficulty for small to medium-sized communities to develop 4-step models prompted Schutz (2000) to note the deficiency of resources in small communities and added that, for these communities to meet the planning requirements, development of innovative methodologies is urgent and necessary.

1.2 Purpose

Many researchers have proposed over years simplified methods for travel demand modeling. But most of them either still require large data and efforts in development of travel demand models in forecasting the traffic network flows. Lee, Chootinan, Chen and Recker (2006) have developed a procedure in which Path Flow Estimator is used to estimate origin-destination table with traffic counts, productions and attractions, and target trip table as estimation constraints. City of St. Helena with a population of 6019 (as of January1, 2002) with in an area of 4 square miles was used as a case study. The model was calibrated and validated. This study is inspired from the work done by Lee, Chootinan, Chen and Recker (2006) and an effort is made to see how their proposed methodology of modeling network flows performs when compared to the conventional four step model estimates of a county size area. The major thrust of this work is to perform O-D trip estimation and traffic assignment simultaneously using PFE (path flow estimator) with input field data such as demographic, socio-economic data/land uses (converted to zonal trip production and attraction), traffic counts and target trip table as estimation constraints and compare it with the estimates of a four-step model.

1.3 Objectives

The main goal of the study is to estimate the O-D matrices and traffic flows using PFE based on the procedure developed by Lee, Chootinan, Chen and Recker (2006) for the Cache County transportation network and to compare and analyze trip table estimations and link flow estimations to the four-step model estimations.

The specific objectives of the study are listed below:

- To understand the problems faced by the small and medium communities.
- To understand the four-step modeling process in general and Cache County travel demand model in specific.
- To estimate trip table and link flows using PFE and examine the differences in four-step model and PFE in their estimates of the following.
    - Productions and attractions between regression analysis estimates of four step model and estimates using ITE trip generation rates.
    - Trip table estimations
    - Link flows of both observed and unobserved
    - Screen line flows

1.4 Limitations

- The study is limited to Cache County that comes under Cache County Metropolitan Planning Organization.
- The procedure is more beneficial for small and medium communities than for larger sized communities due to the limitation in the use of ITE trip generation rates.
- The vehicle productions and attraction estimates are based on the Institute of Transportation Engineers trip generation rates.
- Dispersion parameter used in PFE is not calibrated with the inter-zonal travel times due to non-availability of origin-destination survey data.
- Only automobile trips are considered and transit trip share is not considered.

1.5 Organization of report

This report is organized into eight chapters. The first chapter explains the need of the study, objectives and limitations of the study. The second chapter explains the four step model and problems of small communities in general. The third chapter reviews the existing research done by different researchers in development of alternate methodologies for travel demand modeling. The fourth chapter explains the profile of Cache County and its process of conventional travel demand model development. The fifth chapter explains the modeling approach used in the study. In the sixth chapter the procedure adopted to model the network flow using ITE trip generation rates and PFE is explained. The seventh chapter discusses and compares and analyses the results of PFE and the conventional model. The eighth chapter gives conclusions of the study and scope of further research.

Chapter 2 Four-step model and problems of small communities

2.1 Conventional four-step model[1]

The conventional four-step model separates demand functions into trip generation, trip distribution, mode choice, and traffic assignment. These steps are diagrammatically shown in the Figure 2.1.

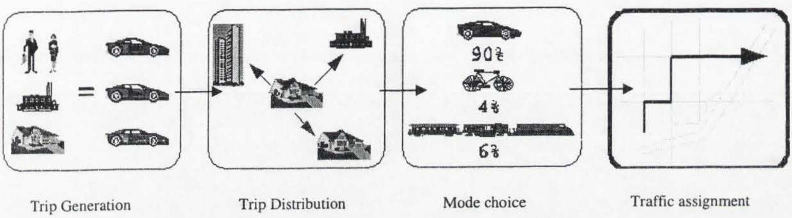


Figure 2-1: Pictorial representation of four-step model.
(http://www.dot.state.oh.us/urban/AboutUs/TravelDM.htm)

2.1.1 Trip generation

Trip generation is the first step in the conventional four-step transportation planning process. It predicts the number of trips originating in or destined for a particular traffic analysis zone.There are two kinds of trip generation models (a) production models and (b) attraction models. Trip production models estimate the number of home-based trips to and from zones where trip makers reside. Trip attraction models estimate the number of home-based trips to and from each zone at the non-home end of the trip. Different

[1] Source: http://www.dot.state.oh.us/urban/AboutUs/TravelDM.htm

production and attraction models like cross-classification, multiple regressions etc., are used for each of the trip purposes. Special trip generation models are used to estimate non-home based, truck, taxi, and external trips.

2.1.2 Trip distribution

Once the trip productions and attractions for each zone are computed, the trips can be distributed among the zones using trip distribution models. Trip distribution has traditionally been based on the gravity model which will be discussed below.

The Gravity Model

The gravity model is much like Newton's theory of gravity. The gravity model assumes that the trips produced at an origin and attracted to a destination are directly proportional to the total trip productions at the origin and the total attractions at the destination and indirectly proportional to the distance between them. The calibrating term or "friction factor" (F) represents the reluctance or impedance of persons to make trips of various duration or distances. The general friction factor indicates that as travel times increase, travelers are increasingly less likely to make trips of such lengths. Calibration of the gravity model involves adjusting the friction factor.

Standard form of gravity model

$$T_{ij} = \frac{A_j F_{ij} K_{ij}}{\sum_{allzones} A_x F_{ij} K_{ix}} x P_i \qquad \textbf{2-1}$$

where T_{ij} = trips produced at i and attracted at j

P_i = total trip production at i

A_j = total trip attraction at j

F_{ij} = a calibration term for interchange *ij*, (friction factor) or travel time

factor ($F_{ij} = C/t_{ij}^{n}$)

C= calibration factor for the friction factor

K_{ij} = a socio-economic adjustment factor for interchange ij

i = origin zone

n = number of zones

Before the gravity model can be used for prediction of future travel demand, it must be calibrated. Calibration is accomplished by adjusting the various factors within the gravity model until the model can duplicate a known base year's trip distribution.

2.1.3 Mode choice

Mode choice models estimate how many people will use public transit and how many will use private automobiles. The most common form of the mode choice model is the logit model. The logit mode choice relationship states that the probability of choosing a particular mode for a given trip is based on the relative values of a number of factors such as cost, level of service, and travel time. In regions where there are several alternative modes available, the mode choice model may require a special form called the "nested" logit. The general logit formulation is given below.

Logit Model

$$P_{it} = \frac{e^{U_{it}}}{\sum_{all j} e^{U_{jt}}} \qquad \text{2-2}$$

where P_{it} = probability of individual t choosing mode i

U_{it} = utility of mode i to individual t

U_{jt} = utility of mode j to individual t

2.1.4 Trip assignment

This step involves assigning traffic to a transportation network such as roads and streets or a transit network. Traffic is assigned to available transit or roadway routes using a mathematical algorithm that determines the amount of traffic as a function of time, volume, capacity or impedance factor. There are many methods for trip assignment. The common assignments used are All-Or-Nothing, Capacity-Restrained, Modified Capacity-Restrained, Incremental, Method of Successive Averages, Convex combinations etc., (Sheffi, 1984).

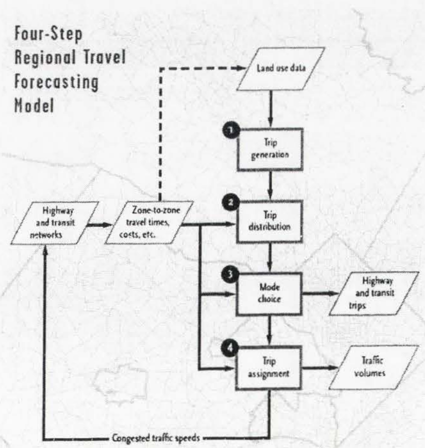


Figure 2-2: Flow chart of conventional four-step model.

Figure 2.2 above shows the flow chart of the four-step model. The four step model described above has few drawbacks. First, the four-step approach is not based on any unifying rationale that would explain all the aspects of demand jointly. All the steps involved are based on different behavioral rationale (Oppenheim, 1995). The second drawback is that all steps are treated independently and each step's output is fed as an input to the other. For free flow conditions this kind of approach may work fine. But under congested conditions, this kind of approach fails to replicate the real world. To overcome this, "feedback" process is used (Figure 2.2) in which generally the travel times obtained after traffic assignment are feedback into trip distribution until a stable distribution of travel times is obtained. But this iterative technique is also another drawback in view of the computational effort involved specially for large networks (Oppenheim, 1995).

2.2 Problems faced by the small communities

Before looking into the alternate models for travel demand modeling, it is important to look into and understand problems of small and medium communities. The problems faced by the small and medium communities can be classified under two basic issues, resource issues and specific issues. Schutz (2000) describes these two issues faced by the small and medium communities in great detail.

2.2.1 Resource issues

Planning requirements

The planning requirements include maintaining a financially constrained metropolitan transportation plan, a three-year transportation improvement plan, and a public involvement process. In addition, each year's unified planning work program emphasizes on a number of focus areas such as, integrating intelligent transportation systems (ITS) into planning, access to jobs, and social justice. These communities often address this planning task by prioritizing requirements and focusing resources on top priorities, while addressing less important priorities with less emphasis. The least important issues may receive only cursory discussion without any real analysis. Sometimes discussion of lesser issues may be grouped with analysis of higher-priority issues.

Education

Education has particularly important aspects in the transportation planning field. Education necessary to deal with the changing and increasingly complex nature of the transportation field and to communicate to target groups is a challenge.

Staff

Staffs of small and medium-sized communities are generalists, knowing something of many areas. High costs involved in educating these staff, in required specializations is an issue.

Technology

Nearly all technologies pose issues like communications overload for these communities. Other issues include deciding when to update software and hardware, selecting a software that is compatible with larger nearby communities when less expensive packages will fill the needs of the smaller community, and maintaining computer applications once they are established. The latter problem is particularly relevant to traffic forecasting models. The original model is often developed by a consultant as part of a contract. The staff may be trained to run the model, but they do not have the time or resources to maintain and update it. Thus the model is useful for a short time and then becomes obsolete if growth occurs, as it usually does where models are needed.

2.2.2 Specific issues

Data collection

Data collection is an expensive and time-consuming process. Small and medium communities do not have good ways to collect, store and update the data regularly. Some state transportation agencies are considering the needs of local jurisdictions in developing their own data collection programs and thereby reducing burden on these communities. Coordinating and sharing resources and data is another issue.

Intelligent Transportation Systems (ITS)

In small and medium-sized communities, ITS applications focus on communications, such as weather or road-condition advisories; incident management; traffic management; and traveler information.

Because of their limited resources, these communities must make good decisions about investing in new technologies. Some of the critical factors are the ability of the new technology to be viable over time; to be compatible not only with other technologies in the agency, but also with those of other agencies; and to solve problems facing the agency cost-effectively. Because of the rapid changes in communications technology, viability over time is an issue that the agency may have to rationalize. Cost-effectiveness is addressed by incorporating the investment decision into the prioritization process for other projects.

All the above stated and explained problems faced by the small and medium communities are posing problems and difficulties to them directly or indirectly in their development of conventional travel demand models. The need of innovative methodologies to provide an alternative to the conventional model for small and medium sized communities is essential in view of the problems stated above.

Chapter 3 Literature Review

Over years different strategies and techniques have been proposed by researchers to model network flows for the small and medium sized communities. Turnquist and Gur (1979) have formulated a non-linear programming problem to estimate origin-destination trip tables using traffic counts or volumes on the networks. This approach was developed as a cost-effective alternative to the conventional model which is also resource intensive. Since all small and medium sized communities regularly collect traffic counts, this approach's input requirements are not additional burden. But this approach cannot take changes in land use and transportation networks into consideration in estimation of trip tables. Another problem with this approach is non-uniqueness of trip table generated. (Lee, Chootinan, Chen and Recker, 2006).

Khisty and Rahi (1987) identified three alternative modeling methodologies applicable for smaller communities. The first method was Low's approach to transportation system modeling that uses a regression model with socio-economic data as the explanatory variables, to determine traffic volumes. In this model, the production and attraction values associated with each zone were assigned to the network to develop a trip probability value for each link in the network. Then, provided quality traffic counts data exists, traffic volumes and trip probability values were used to develop a relationship through the use of linear regression. Using each link's regression equation, forecasts were determined by assigning forecast year's trip probabilities. Then using the regression equations the forecast year traffic counts were predicted. The second methodology proposed by Elrazaz, Halkias, and Neumann distributes and assigns zonal socio-economic variables directly to the study area network. In this model, external traffic count volumes were removed from traffic count data, and then the remaining traffic count volumes were used to develop a regression model as a function of socio-economic variables. As with Low's model, the regression model was again used to forecast horizon year traffic. The third methodology proposed by Khisty and Rahi, referred to as an internal volume forecasting (IVF) model, follow the work of Low while incorporating a few changes to reduce errors. One difference was the use of employees and job positions per zone as the basis for trip generation and the other is the use of total employment available in the study area as a factor in developing the trip interchange index assigned to the network. All three models require high quality, system-wide traffic counts and assume stability of mathematical relationships between development and horizon years. Furthermore, trip generation equations and distribution parameters are likely to change significantly for small areas, with infrastructure changes negating previously developed regression equations.

Anderson, Sharfi and Sampson (2005) have proposed a direct demand forecasting model which is a regression equation with randomly chosen traffic counts (average daily traffic) as the dependent variable. The independent variables chosen are functional class of road, number of lanes, population within 0.5 buffer mile of the count station, employment within 0.5 mile buffer of the count station, and a binary variable reflecting mobility. Functional classes of the road and number of lanes have come out as significant variables. Validation is done for randomly chosen traffic count stations. Even though this model is simple and easy to use its universal appeal may be questionable. Moreover, the model's significant variables are functional class of road and number of lanes, with little significance given to demographic and economic variables. This means that the predicted ADT is not influenced by the increase in population or employment. So if the functional class and number of lanes do not change, even a high increase in the population and employments will have a negligible impact on the predicted ADT.

There are other methodologies that do not rely on regression equations using traffic count data. The National Highway Cooperative Research Program (NCHRP) developed Report 187 that provides methods for quick response urban travel estimation. *NCHRP Report 187* contains a wide range of trip generation rates, based on population, income, automobile ownership and employment segregated by type. The resulting productions and attractions which form the trip generation equations are distributed and assigned to the transportation network using typical urban planning techniques (Anderson, 2000).

Wegmann, Chatterjee, and Gregory (2000) studied three small communities in Tennessee as case studies in their work. They adopted a mid-range procedure incorporating a simplified travel demand model using user friendly software like QRS II and TransCAD. The number of households and the auto ownership per household were used to determine average person daily trips per household and percentage of daily person trips by purpose. Person trips and trip purpose figures were taken from NCHRP Report 365. Trip attractions were calculated from "windshield survey[2]" and local sources. The equations

[2] Windshield survey- Two people drive and walk through the cities and estimate number of employees at each business place.

used to calculate trip attractions were based on QRS manuals. Gravity model was used for trip distribution. Traffic assignment was done using All-or-Nothing assignment, since small communities have isolated or limited congestion. Changes in the placement of centriod connectors and changes in the link speeds were done during calibration. The availability of a GIS coded network, transferable parameters, local knowledge, and a windshield survey of local land uses are important parameters in this kind of modeling. The network coding has to be very precise since this is a GIS based approach. Windshield data collection is a very approximate estimation of trip attraction and its results may be questionable.

Schute and Ayash (2004) have proposed a Joint Estimation Model (JEM) on the argument that "a single prototypical model of a small urban area could be estimated for joint travel behavior survey data and then this model could be calibrated to reflect local conditions within specific cities and can be used for any small community model development". The model is entirely implemented through a series of script files written in the R statistical programming language, with the exception of traffic assignment, which is carried out in EMME/2. The model followed the four-step process consisting of pre-generation, trip generation, trip attraction and traffic assignment. Within pre-generation, all the necessary inputs for trip generation are produced. Trip assignment is performed using a single-class, equilibrium capacity constrained assignment. The input for this model is, socio-economic data, travel time data, time-of-day, directional factors, average vehicle occupancy etc. This model's basic assumption of having a common model for all small communities in a specific area may not be replicable in all areas. The requirement of input data is also quite substantial and is almost as much as the data requirement for the conventional model.

Mann and Dawoud (2004) have proposed a simplified modeling process for any area size using software called "TP/ 4 in 1[3]" which can run all four steps in one execution on a personal computer. The main difference or limitation is its assumption of ten vehicle trips per detached house. However they argue that it satisfies all the sparsely populated areas. In the trip distribution model, work trips were calibrated to match census jurisdiction to jurisdiction trip tables, and non-work trips were calibrated to match ground counts. Measured household trip generation rates were given more confidence than home-interview trip generation rates or trip length frequency distributions on the argument that travel surveys under-report short trips rather than long trips and may even inflate the long trips. Mode split was applied to work trips only. The model was run for daily trips. "TP/4in 1" uses "incremental capacity restraint trip assignment". Even this model requires substantial data and a calibrated regional model. The applicability of 10 vehicle trips per detached house may not be applicable for all communities.

Lee, Chootinan, Chen, and Recker (2006) have proposed an alternate methodology to model and forecast traffic using path flow estimator (PFE) using land use data (converted to zonal trip production and attraction using ITE trip generation rates), traffic counts, and a target trip table as estimation constraints. The level of satisfaction for each constraint can be defined differently based on the confidence of measured data collected. The forecast was done by dropping the traffic count constraint. For forecasting, the target trip

[3] TP/4in1 is software developed by Dawoud and Mann for Virginia DOT.

table is found by scaling the baseline trip table so that it matches the total future demand. This kind of scaling of the trip table will keep the relationship of travel impedance and trip interchange between O-D pairs established from traffic counts in the forecasting process. Acceptable root mean square of errors between traffic counts and model estimated link flows for the City of Helena in Napa valley have shown that the model has performed satisfactorily. The same process except forecasting is replicated in this study and is explained in detail in Chapter 6.

Chapter 4 CACHE COUNTY AND TDM

4.1 Cache County [4]

Cache County is a high mountain valley located in the northeastern corner of Utah known as Bridgerland, the northern end of the Wasatch Front. It is nestled between the Wellsville Mountains to the west and the Bear River Mountain Range to the east. Cache County has been traditionally rural with agricultural influence. Yet, recent growth has added an urban flavor. Utah State University is located in Logan, the largest city in the county. Cache County's population is close to 100,000 as per 2000 census. Other towns in the county are North Logan, Hyde Park, Smithfield, Richmond, Cove, Lewiston, Cornish, Trenton, Amalga, Benson, Clarkston, Newton, Petersboro, Mendon, Wellsville, River Heights, Providence, Millville, Nibley, Hyrum, Avon, and Paradise. Figure 4.1 shows Cache Metropolitan Planning Organization's (CMPO) urbanized area.

---

[4] Source: Cache Valley Long Range Transportation Plan, 2030 and CMPO.

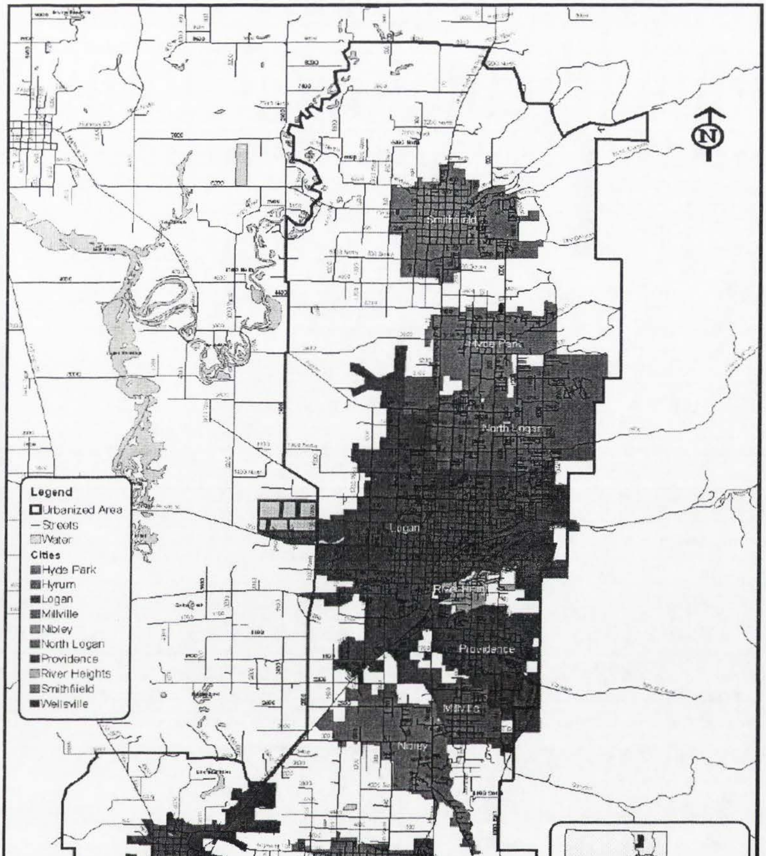


Figure 4-1: Cache County and urbanized area *(Source: CMPO)*

The CMPO was designated by the federal government following 1990 census, when the urbanized area surpassed 50,000 in population. The main center for Cache County is the city of Logan which is home for half of the county's population. The economy of the area has historically depended on agriculture. Due to the academic setting of Utah State University, there has been diversity in employment and population. Table 4.1 gives the details of socio-economic data.

Table 4-1: Socioeconomic profile of Cache MPO, city wise

City	Population				Households				Employment		
	1990	2000	2003	% change 90-03	1990	2000	2003	% change 90-03	2001	2003	% change 01-03
Hyde Park	2,190	2,955	3,237	47.8	544	779	864	58.8	350	328	-6.3
Hyrum	4,829	6,316	7,176	48.6	1,260	1,744	1,877	49	2,232	2,241	4
Logan	32,762	42,670	44,372	35.4	11,034	14,692	15,746	42.7	26,153	26,329	0.7
Millville	1,202	1,507	1,430	19	287	405	381	32.8	27	33	22.2
Nibley	1,167	2,045	2,881	146.9	314	580	819	160.8	238	221	-7.1
North Logan	3,768	6,163	6,910	83.4	961	1,778	1,956	103.5	4,542	5,475	20.5
Providence	3,344	4,377	5,268	57.5	873	1,290	1,526	74.8	807	1,114	38
River Heights	1,274	1,496	1,276	0.2	387	492	424	9.6	129	92	-28.7
Smithfield	5,566	7,261	8,029	44.3	1,513	2,159	2,383	57.5	1,068	1,595	49.3
Wellsville	2,206	2,728	2,912	32	603	815	865	43.4	311	463	48.9
Total	**58,308**	**77,518**	**83,491**	**43.20**	**17,776**	**24,734**	**26,841**	**51.00**	**35,857**	**37,891**	**5.70**

(source: CMPO)

CMPO area population according to 1990 census data was 58,308, which grew to 77,518 in 2000, a 33 % increase in ten years. Between 2000 and 2003, the population increased by another mainly with the addition of Nibley, Hyrum and Wellsville into CMPO area.

Logan is the main employment center for CMPO area. The largest employer in Logan is Utah State University (USU) with approximately 6000 employees. With a stable employment base due to employers such as USU and good business potential (CBD), many work trips are generated within the City of Logan.

Land use in CMPO is a mix of residential and commercial. Higher densities of both residential and commercial activities are concentrated in Logan and the densities are decreasing from center to the edges of urbanized area. Figure 4.2 shows the land use patterns of the CMPO area.

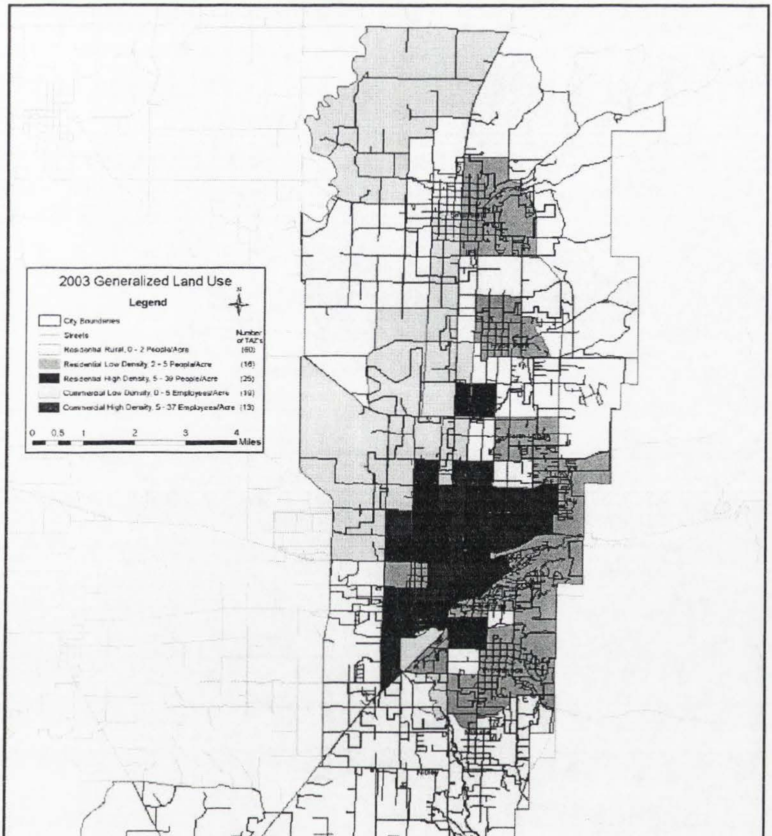


Figure 4-2: Existing land use pattern *(Source: CMPO)*

The region's main road US-91 runs north and south serving as City of Logan's Main Street and is connects to I-15 in Box Elder County. The US 89/91 corridor carries bulk of the traffic in this region. Other major road facilities in the region are SR-89, SR-165, SR-101, and SR-30. Figure 4.3 shows the road network with functional classification.

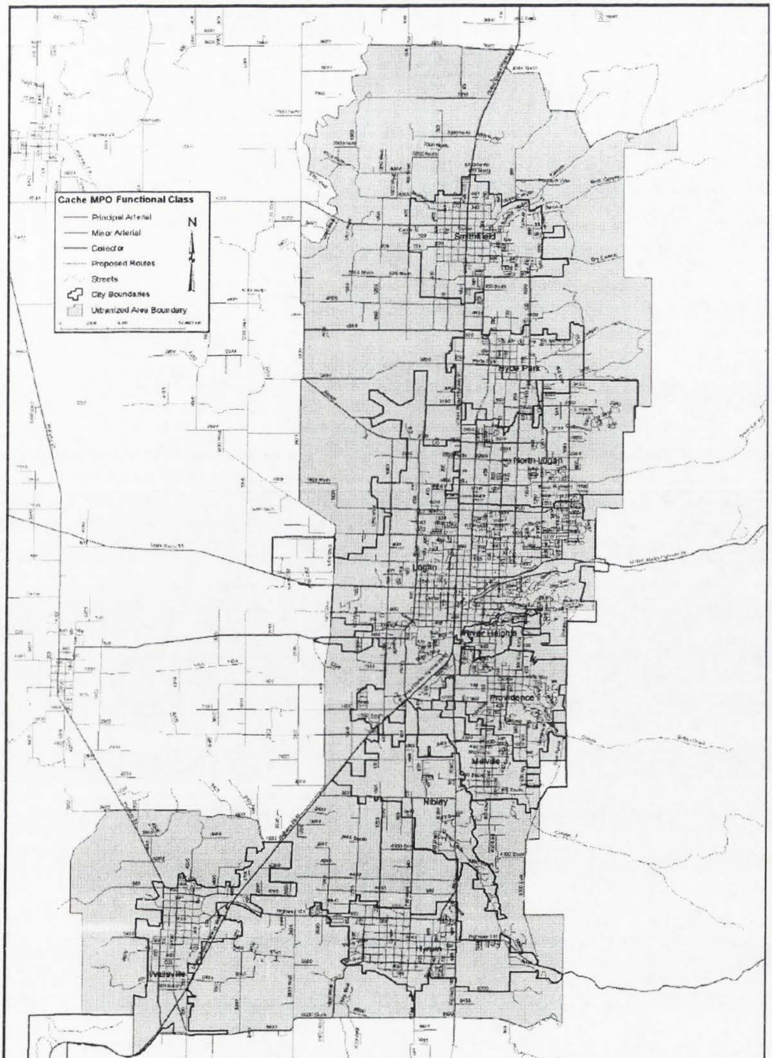


Figure 4-3: Cache MPO area roads functional classification *(Source: CMPO)*

Cache County is served by two public bus transit services- Cache Valley Transit District (CVTD) which serves entire Cache valley and Logan Transit District (LTD) which serves Logan and North Logan. No fares are charged in the transit service.

U.S Highways 89 and 91 along with State Route 30 are the main freight routes linking the county with I-15 and I-84 to the west. It is estimated per week 2600 trucks traverse through Cache County mainly in the peak hours.

4.2 Cache County TDM

4.2.1 General outline

The methodology followed for Cache County TDM is identical to Wasatch Front Regional Council's (WFRC) model which is developed for Salt Lake City and Ogden areas. Three trip purposes were used in trip generation, trip distribution and mode split sub models: Home-based Work (HBW), Home-based Others (HBO), and Non-Home Based (NHB).

The roadway network data and traffic volume data were provided by Utah Department of Transportation (UDOT). Socio-economic data was provided by Cache Countywide Planning and Development Office. This data was supplemented by substantial data collection effort conducted for Cache valley corridor to obtain trip making characteristics.

4.2.2 Data collection

Data collection and surveys are important for travel demand modeling (TDM) importantly in TDM's calibration and validation. Generally, the metropolitan planning organizations (MPOs) coordinate and collect information from the local communities, workforce services, state government and federal government agencies.

Cache County MPO coordinates gathers, edits data, expands dwelling unit information to estimate population, and generates reports documenting the collected information also. The data requirements for a travel demand model are transportation network data, census boundaries, travel analysis zone data, demographic, socio-economic data, land use data, and corridor studies. The data sources and its development processes are explained in Annual Report of Socio-Economic Characteristics, Cache County, 2004.

4.2.2.1 Transportation network and TAZs

The network was developed using precise scale base mapping obtained from the U.S. Geological Survey (USGS) for the central part of Cache County, and from the official Cache County map for outlying areas not covered by the USGS mapping. Digital orthophoto quadrangle data was also used to verify roadway connections and lane configurations wherever necessary. Most of this work is done by Utah Department of Transportation (UDOT).

Traffic analysis zones (TAZs) developed by Cache Countywide Planning and Development Office are adopted in consistent with US Census tract boundaries for Cache County. Cache Countywide Planning and Development Office have divided the county into 131 TAZs, 101 in urbanized area and 32 in non-urbanized area. TAZs 1 to 101 are in urbanized area and 201 to 233 are in non-urbanized area. U.S. Census Bureau has divided Cache County into 16 census tracts and that makes eight TAZs per census tract on average. Six external zones were identified, representing the gateways where vehicles enter and leave Cache County, these are:

- U.S 91, North of Lewiston, at Utah border;
- Utah State Road 23, North of Cornish, at Utah Border;
- Utah State Road 30, west of Peterboro, at the county line;
- Utah State Road 89/91, south of Wellsville, at the county line;
- Utah State Road 101, east of Hyrum, at the Fox canyon; and
- Utah State 89, east of Logan, at Logan canyon.

The identified traffic analysis zones of Cache County are shown in Figure 4.4 and external zones are shown clearly in Figure 4.5.

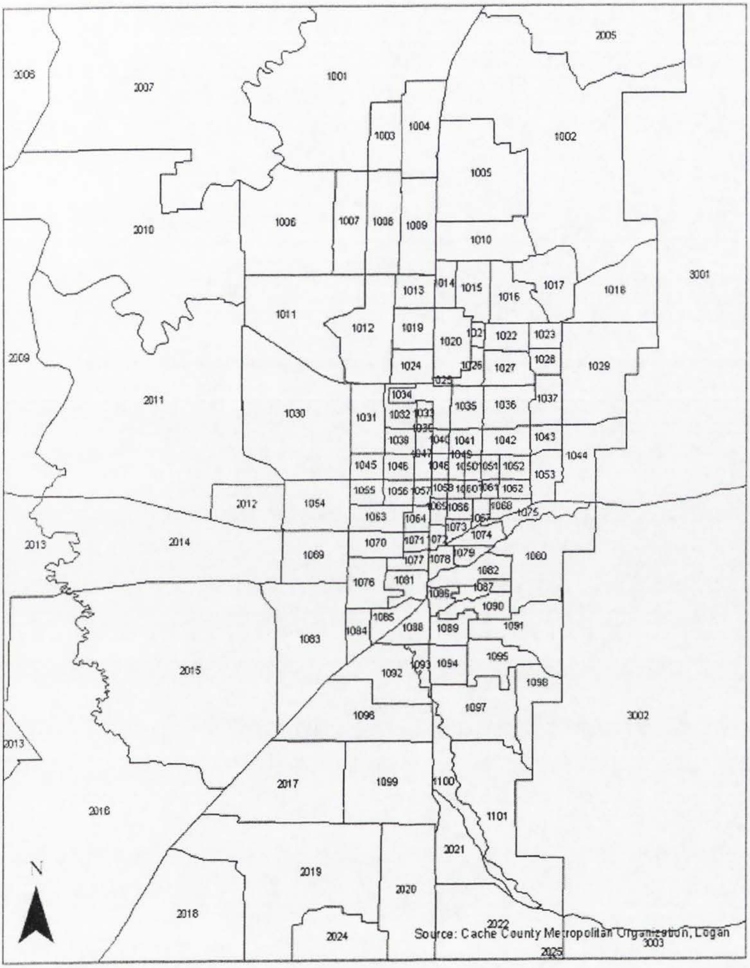


Figure 4-4: Cache County travel analysis zones

(Note: Detail TAZ maps and census tract maps are given in Appendix 1)

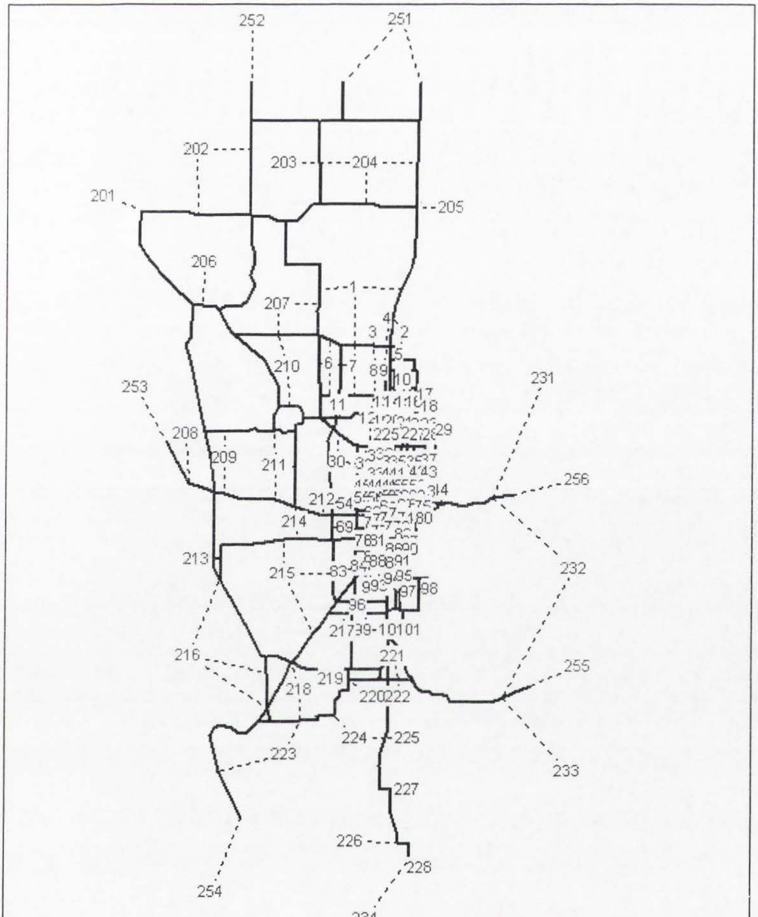


Figure 4-5: Cache County travel demand model's identified TAZ especially external zones

(Note: To compare with Figure 4.4 TAZ ids, Zone id 216 in this figure should be read as 2016 in Figure 4.4)

Zone 251- U.S 91, North of Lewiston, at Utah border
Zone 252- Utah State Road 23, North of Cornish, at Utah Border
Zone 253- Utah State Road 30, west of Peterboro, at the county line
Zone 254- Utah State Road 89/91, south of Wellsville, at the county line
Zone 255- Utah State Road 101, east of Hyrum, at the Fox canyon
Zone 256- Utah State 89, east of Logan, at Logan canyon

Nodes were divided into regular nodes representing intersections and into centroids representing center of activity of zone. The average speeds and capacity values assumed for the links were as per Highway Capacity Manual (HCM). Table 4.2 and Table 4.3 show the average speeds and lane capacities assumed by facility and area types.

Table 4-2: Average speed (mph) by facility type and area type

Area Type	Facility type				
	Principal Arterial	Minor Arterial	Urban/Rural Collector	Residential street	Centriod collector
CBD	22	22	22	20	10
Outer CBD	24	22	22	20	10
Rural/Residential	32	30	30	28	15

Table 4-3: One-way lane capacities by facility type and area type (vph)

Area Type	Facility type				
	Principal Arterial	Minor Arterial	Urban/Rural Collector	Residential street	Centriod collector
CBD	650	550	550	450	10000
Outer CBD	750	600	500	450	10000
Rural/Residential	850	750	500	450	10000

In this model, level of service "C" was assumed since level of services "D" and "E" would over estimate speeds under congested conditions on the fact that Logan has very little history of any severe traffic congestion.

4.2.2.2 Demographic and socio-economic data

(Source: annual report of socio-economic characteristics by CMPO)

Population

This describes number of persons residing in a zone or tract. The zonal population is derived from the census 2000. For future forecasts, the housing units are factored by family size derived from the 2000 census. Group quarters such as nursing homes, prisons, and college dormitory's population are added by zone.

Total housing units

This represents the number of housing units within a zone including single family homes, multi-family and mobile homes. The dwelling unit data is updated annually by the local cities based on the building permits issued. The census dwelling total is derived using 2000 census data is adjusted for the end of the year 2002.

Multi-family -This represents the housing units with two or more units in structure, including apartments and condominiums. Group quarter's population is not included. The multi-family units are included in the total dwellings total. This data was also provided by the local city authorities.

Total employment

This represents the total number of non-agricultural workers employed in a zone including wage and salary workers, self-employed workers, and government workers. The Utah Department of Workforce Services assembles this data using the unemployment compensation coverage file as the primary basis, and does considerable supplemental work. Construction and agricultural employment have been removed from the total because workers in these sectors often work at sited locations away from the reported place of business.

Total service employment

This is the portion of total employment primarily engaged in service industries. This data is extracted from the database used to generate total employment and includes the North

American Industry Classification System (NAICS) sectors 70 through 81. (For explanations see NAICS coding and sectors explained in Appendix 2).

Total retail employment

This is the portion of the total employment primarily engaged in retail trade. This data is extracted from the database used to generate total employment and includes the North American Industry Classification System (NAICS) sectors 41 through 46. (For explanations see NAICS coding and sectors explained in Appendix 2).

Total industrial employment

This is the portion of the total employment primarily engaged in manufacturing, transportation, communication, electric, gas, and sanitary services. This data is extracted from the database used to generate total employment and includes the North American Industry Classification System (NAICS) sectors 31 through 33. (For explanations see NAICS coding and sectors explained in Appendix 2).

4.2.2.3 Corridor studies

Corridor studies include internal road surveys, external roadside surveys, home travel surveys, on-board transit surveys and cordon traffic counts. Table 4.4 shows the use of field data in the process of travel demand model development.

Table 4-4: Use of field data collection in model development

Travel demand model module	Field data collection effort				
	Internal road surveys	External road surveys	Home travel surveys	On-board transit surveys	Cordon traffic counts
Trip Generation			•		
Internal-Internal trip distribution	•		•	•	
Internal-External trip distribution	•	•	•		•
External-External trip distribution		•			•
Model Split			•	•	
Traffic Assignment					•

(Source: Model development and validation report, Cache County)

The data collection is an extensive process and requires a lot of resources. The Cache valley corridor study is done by Wilbur Smith Associates (consultant hired by the CMPO) in conjunction with local cities officials.

4.2.3 Trip generation

The data inputs to a travel demand model are based on the latest set of demographic and socio-economic estimates developed by Cache County's Countywide Planning and Development Office by adjusting U.S. Census block and tract information. The information used in this step is given in Appendix 5.

The distribution of traffic between home based work (HBW), home based others (HBO), and non-home based (NHB) is approximately 19.7%, 43.8%, and 36.4 % based on recent home travel survey.

For trip generation linear regression models are deduced based on the demographic and socio-economic data. The following data was considered for trip production model:

- Total population
- Single family household
- Multi-family household
- Average persons per household
- Number of vehicles per household.

The following socio-economic data was considered for trip attraction model:

- Total dwelling units
- Retail employment
- Industrial employment
- Other employment
- Total employment

After regression and correlation analysis, the refined and final models were deduced. The external-internal and internal-external trips were estimated by techniques of internal trip estimation and external station counts collected. The external-external trip estimation is done based on origin-destination data collected as part of road side survey. Utah State

University was considered as a separate trip generator since standard equations do not provide satisfactory trip generation and distribution estimates.

Table 4.5 presents the regression equations models developed for different trip purposes.

Table 4-5: Trip Generation regression model's equations

Trip purpose	Trip production	Trip attraction
HBW	3.00 x SFH + 2.00 x MFH - 0.7	1.9 x TOT_EMP.
HBO	7.00 x SFH + 3.5 x MFH +13.60	0.6 x POP + 14.4 x R_EMP
NHB	3.5 x TH + 4.0 x R_EMP	3.5 x TH + 4.0 x R_EMP

where POP- Population
SFH- Number of single family households
MFH- Multi family households
TOT_EMP- Total number of employees
R_EMP- Total number of retail employees
TH- Total number of households

The regression analysis results for home based productions are given the Table 4.6.

Table 4-6: Regression analysis results for home based trips (trip generation)

Measure	Home Base Work	Home Based Other
Multiple R	0.916	0.851
R Square	0.839	0.729
Adjusted R Square	0.821	0.695
Standard Error	6.017	19.91

(Source: Model development and validation report, Cache County)

The validation of the trip generation model is done by comparing the model's estimated values against observed values from internal and external O-D surveys, as well as home surveys conducted as part of Cache valley corridor survey. Future trip generation models were developed using these regression equations and external traffic was estimated by increasing the base year traffic at an annual growth rate of four percent estimated by rural traffic forecasting model developed by Utah Department of Transportation.

4.2.4 Trip distribution

The assumption in central trip distribution model 'each traveler making a trip chooses a destination from all the available destinations on the basis of the characteristics of each destination and the relative impedance associated with traveling to each destination'.

The trip distribution model for Cache County follows the formula below:

$$T_{ij} = P_i \frac{A_j * F_{ij} * K_{ij}}{\sum_j A_j * F_{ij} * K_{ij}} \qquad \textbf{4-1}$$

where T_{ij} = The number of trips produced in zone i and attracted to zone j
P_i = the number of trips in zone i and for a specific trip purpose
Aj = the number of attractions in zone j for a specific trip purpose
F_{ij} = an empirically derived "friction factor", which expresses the average area wide effect of spatial separation on the trip interchanges between the two zones, i and j
K_{ij} = Empirically zone-to-zone adjustment factor, which takes into account the effects on travel patterns of defined social and economic linkages not otherwise incorporated into the model

The impedance measure expressed as friction factor is the average travel time. Special adjustment factors called K- Factors were not used. Intra-zonal trip impedance is calculated as half the travel impedance to the nearest zone. For each trip purpose, the gravity model is calibrated to estimate the friction factors. These are calculated from the comparison of observed to model-estimated trip length frequency distributions, that were obtained from the internal and external origin-destination surveys, as well as on-board transit and home surveys. Two distribution functions were tested to estimate the friction factors.

Gamma function- $F = a * I^b * e^{cI}$ **4-2**

Negative exponential distribution- $F = a * e^{-bI}$ **4-3**

where

F- friction factor
a,b,c- equation coefficients
I- Impedance

Friction factor curves have been adjusted till the modeled trip length is within 16 percent of the household survey data. The validation of the trip distribution model is done with reasonableness and consistency check with available data sources like survey data collected from corridor study, state averages, and the model developed by WFRC for Salt Lake City and Ogden areas. The level of reasonableness identified and used in this model is not specified in the Calibration and Validation report.

4.2.5 Mode choice model

Mode choice model developed for Cache County is a daily binary choice regression model, a form of discrete choice model. For each trip purpose, separate mode choice models were developed.

The variables considered for regression analysis of mode choice model are:

- Average household income
- Average vehicle ownership
- Total population
- Accessibility to transit
- Number of vehicles per household

Two separate linear equations were estimated for Home-based Work (HBW) and Home-based Other (HBO) trip purposes which are of the form:

HBW Automobile percentage share:

$$\text{HBW Auto} = -17.17 * \ln(TA) + 39.89 * (AOWN) - 120.14$$

HBO Automobile percentage share:

$$\text{HBO Auto} = 0.0265 * TVEH + 0.0117 * POP + 99.81$$

where TA – Transit accessibility
AOWN – Average vehicle ownership
TVEH – Total Vehicle Ownership
POP – Total population

Based on the approach followed by Wasatch Front Regional Council (WFRC), in the urban areas where there is a possibility of walk or transit trips, Non-Home Based trips (NHB) automobile trips are assumed as 40 percent of the HBW share. For zones outside the urban fringe, automobile share for NHB trips is taken to be same as that of HBW trips.

Calibration of mode choice model is done by using goodness-of-fit measures to compare predicted trips with observed data for both auto and transit trips. The t-tests were conducted to check the slope co-efficient. The results of the regression and t-tests are given below in Table 4.7.

Table 4-7: Regression analysis results for Home based trips (mode choice)

Measure	Home-based Work	Home-based Other
Multiple R	0.921	0.881
R square	0.849	0.777
Adjusted R square	0.845	0.767
Standard Error	3.865	2.732
T-Statistic (1^{st} variable)	6.117	7.834
T-Statistic (2^{nd} variable)	11.34	5.203

(Source: Model development and validation report, Cache County)

The criteria for acceptance of the regression model is by looking at R-square value typically between 0.85 to 1 and t-statistic equal or greater than two. Closer the value of R-square to 1, the higher will be the expectancy powers of the equation. Higher the value of t-statistic, higher would be the probability that the slope of the regression equation is not equal zero.

4.2.6 Trip assignment

Auto trips were loaded onto the highway network using an "incremental, capacity-constrained assignment" process. The network traffic is loaded in the following eight increments: 50 percent, 20 percent, 5 percent, 5 percent, 5 percent, 5 percent, 5 percent, and 5 percent, for a total of 100 percent trips. The reason given for decreasing size of increments is that the disadvantage of not assigning trips to the ultimate shortest path will be minimized by decreasing the percentage of vehicles, being loaded onto the highway network in each of iterations. The rationality behind the selection of number and size of increments is not explained. The standard BPR function is modified to calculate the congested travel time. The maximum volume-capacity ratio in the BPR function is limited to three.

$$t(v) = t_f\,(1 + .015(\min(3, v/c))^4 \qquad \textbf{4-4}$$

where $t(v)$= congested travel time
t_f= free flow travel time
v/c= volume-capacity ratio

The volume-capacity ratio used in BPR function is for the daily traffic in contrast to the general practice of using peak hour traffic. This will estimate the travel times less than the actual due to which there is a possibility of assignment of high flows on certain links.

The travel impedance is adjusted at every iteration using the following equation.

$$ADJIMP() = t_f + 0.25(\,t(v) - t_f\,) \qquad \textbf{4-5}$$

At the end of each iteration, the travel time on the network is updated by adding 25% of the difference between the congested travel time and free flow travel time to the free flow travel time to reflect effect of congestion in assigning the flows in this step.

4.2.7 Calibration and validation

For accuracy and evaluation of the assignment model, three techniques were used.

- Screen line comparisons
- Percent root mean square error
- Facility-specific validation targets

The calibration and validation for this model was done in the original model which was developed in 1995. For screen line comparisons six screenlines (A to F) were setup and Figure 4.6 shows their locations. The screenlines were set up at Cache County borders to validate the total amount of traffic in and out of the county (screenlines A, B, C, and D). Screenline D was setup to capture state route 30 alone. Screen lines E and F were set up to validate the traffic within the County in such a way that they bisect County both vertically and horizontally.

Observed and modeled traffic volumes were compared both at a link level and an entire screen level for each of the twelve runs. In the first run maximum variation to the ground counts was 37% excess (screen line E). This was reduced to 14% by the end of twelfth run where the calibration of screen lines in conjunction with facility-specific measures was stopped. Results of screen line analysis can be found in the Cache County's model development and validation report.

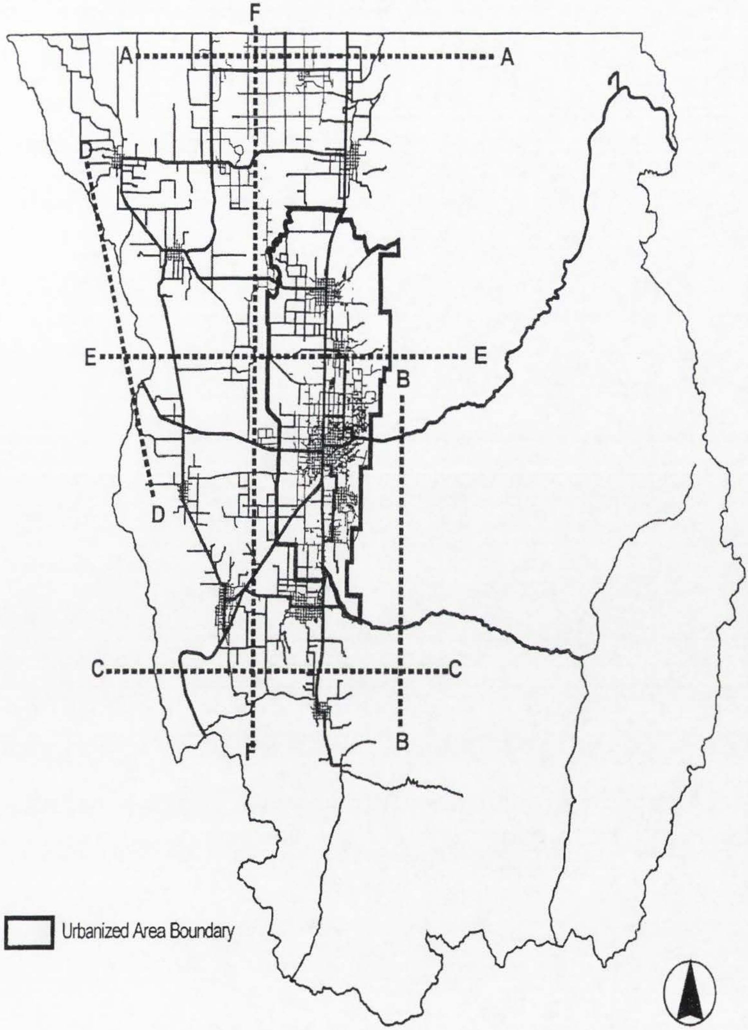


Figure 4-6: Screen lines selected for calibration in Cache County travel demand model

Percent root mean square error (PRMSE) provides an indication of relative closeness of estimated and observed traffic volumes for individual observations, while putting great weight on large differences. Values of PRMSE have not been documented in the model's calibration and validation report.

$$PRMSE = \frac{\sqrt{\sum_{i=1}^{N}(O_i - E_i)^2 / N}}{\sum_{i=1}^{N} O_i / N} * 100 \qquad \textbf{4-6}$$

where O = Observed traffic volumes, from traffic counts
E = Estimated traffic volumes, from the model
N = Number of traffic counts.
i = Link

Facility-specific measures of model performance involve, comparing assigned traffic volumes, vehicles miles of travel (VMT), and vehicle hours of travel (VHT) with same quantities calculated based on actual field traffic counts. Comparisons were done on an area wide basis, on a roadway classification basis, and on a sub-area basis, with expectations for accuracy at each level. In the principal arterials, the variation of VMT, VHT and volumes were reduced to 15% in the twelfth run from 50% in the first run. In the central business district (CBD), the variation of VMT, VHT and volumes were reduced to with in 5% from 13.9%. For details refer Cache valley corridor model development and validation report.

4.2.7.1 Overall model validation process

The model was validated iteratively in the following steps:

1. Proof of highway network
   a. Plot the network to spot the obvious errors.
   b. Check the shortest paths between zones for reasonableness.
   c. Check zone-to-zone travel times.
2. Trip generation
   a. Identify isolated sub-areas with uniformly too high or too low traffic assigned to streets
   b. Review land use assumptions for these areas.
   c. Review trip generation rates.
3. Trip distribution
   a. If totals of all screen lines are too high, then friction factors may be too flat.
   b. Review directional flows across screen lines
   c. Use "K_{ij}" adjustment factors if necessary.
4. Mode choice
   a. Check estimated travel times used as input to the mode-split process for reasonableness.
   b. If final travel times are significantly different from the initial estimates, feed them back into trip distribution.
5. Trip assignment
   a. Check individual screen lines

b. Calculate percent mean square error for each facility

c. Compare assignment with facility-specific validation targets

d. If there are substantial differences between the observed and the model, revise the model assumptions from the beginning.

In every run, the screen lines and facility specific measures were checked and relevant parameters were adjusted. Majority of the iterations were run to achieve facility specific parameter- local streets estimation within the reasonable limits (40%). In total, twelve runs were made to finalize the model and after every run one or more above stated evaluations and related adjustments were done to achieve the final model.

Chapter 5 Path Flow Estimator and Visual PFE

5.1 Path Flow Estimator[5]

Path Flow Estimator (PFE) was originally proposed by Bell and Shield in 1995. PFE estimates path flows and travel times based on the traffic counts in the given network. It follows Turnquist and Gur's (1979) approach of estimating base line trip table from traffic counts as a non-linear programming. Unlike equilibrium assignment in Turnquist & Gur's approach, PFE uses logit stochastic user equilibrium in assigning the flows. PFE is an optimization problem that which minimizes the network travel time under observed volume constraints, vehicle production, attraction constraints and target trip table constraints. It requires data on network characteristics, observed volumes, zonal vehicle productions, attractions, and a target trip table if available (Lee, Chootinan, Chen and Recker, 2006).

In PFE, the path flow estimates are unique. These unique path flows can be aggregated to obtain O-D flows. By the nature of trip table estimation problem, multiple trip tables can reproduce same set of link flows. PFE selects the O-D table that is consistent with the logit stochastic user equilibrium (SUE) assumption. The advantage of PFE is its assumption of stochastic user equilibrium (Sheffi, 1984) which says that users may travel in non-equal travel time routes due to their imperfect knowledge of the network.

[5] reference: Visual PFE manual and Lee, Chootinan, Chen, Recker, 2006.

Based on the equivalent formulation for a logit-based SUE problem (Fisk, 1980), the PFE formulation is formulated as follows:

Minimize: $$\frac{1}{\theta}\sum_{rs}\sum_{k} f_k^{rs}(\ln f_k^{rs} - 1) + \sum_{a}\int_0^{x_a} t_a(w)dw \qquad \text{5-1}$$

Subject to:

$$(1-\varepsilon_a)\cdot v_a \le x_a \le (1+\varepsilon_a)\cdot v_a, \quad \forall a \in M, \qquad \text{5-2}$$

$$x_a \le C_a, \quad \forall a \in U, \qquad \text{5-3}$$

$$(1-\varepsilon_{rs})\cdot z_{rs} \le q_{rs} \le (1+\varepsilon_{rs})\cdot z_{rs}, \quad \forall\, rs \in RS, \qquad \text{5-4}$$

$$\sum_{s}\sum_{k\in K_{rs}} f_k^{rs} = P_r, \quad \forall r \in R, \qquad \text{5-5}$$

$$\sum_{r}\sum_{k\in K_{rs}} f_k^{rs} = D_s, \quad \forall s \in S, \qquad \text{5-6}$$

where

$$x_a = \sum_{rs}\sum_{k} f_k^{rs}\delta_{ka}^{rs}, \quad \forall a, \qquad \text{5-7}$$

$$q_{rs} = \sum_{k\in K_{rs}} f_k^{rs}, \quad \forall\, rs \in RS \qquad \text{5-8}$$

Notation:

A	=	Set of all network links ($A = M \cup U$)
M	=	Set of measured links
U	=	Set of unmeasured links
R	=	Set of origin nodes
S	=	Set of destination nodes
RS	=	Set of O-D pairs
K_{rs}	=	Set of paths connecting origin r and destination s
θ	=	Dispersion parameter
$t_a(\cdot)$	=	Link cost function of link a

f_k^{rs} = Estimated flow on path k connecting origin r and destination s

δ_{ka}^{rs} = Path-link indicator: 1 if link a is on path k between origin r and destination s, and 0 otherwise

v_a = Observed flow on link a

ε_a = Measurement error, (0,1), for observation on link a

C_a = Capacity of link a

x_a = Estimated flow on link a

z_{rs} = Observed O-D flows from origin r to destination s

ε_{rs} = Measurement error, (0,1), for a priori O-D flow from origin r to destination s

q_{rs} = Estimated O-D flows from origin r to destination s

P_r = Trip production of origin r

D_s = Trip attraction of destination s

In this formulation, the logit SUE objective function aims minimize travel time by satisfying traffic constraints on the network links (equation 5.2) and target trip table (equation 5.4), if available. The link flows of unobserved links are constrained by their capacities as upper bound (Equation 5.3). The productions and attractions calculated from land use plans are used as production and attraction constraints (Equation 5.5 and Equation 5.6).The level of satisfaction of each constraint (Equation 5.2 and Equation 5.4) can be defined based on the confidence level of collected data. So the high level of confidence in the data collected will constrain the estimation within smaller tolerance. Path flows can be derived analytically as a function of path costs and dual variables associated with constraints (5.2), (5.3) and (5.4) by Langrangian function.

$$f_k^{rs} = exp\left(\theta \cdot \left(-\sum_a t_a \delta_{ka}^{rs} + \sum_{a \in M} \left(u_{a^-} \delta_{ka}^{rs} + u_{a^+} \delta_{ka}^{rs} \right) + \sum_{a \in U} d_a \delta_{ka}^{rs} + o_{rs^+} + o_{rs^-} \right)\right), \quad \forall k, rs$$

5-9

where, u_a^+ = Dual variable of observation constraint on link a (upper limit)

u_a^- = Dual variable of observation constraint on link a (lower limit)

d_a = Dual variable of capacity constraint on link a

o_{rs}^+ = Dual variable of target trip table for O-D pair rs (upper limit)

o_{rs}^- = Dual variable of target trip table for O-D pair rs (lower limit)

The dual variables representing upper and lower bounds in equation 5.2 and equation 5.4 will be inactive if estimated link flows and OD flows fall with in acceptable ranges. The dual variable of equation 5.3 will become active when a link flow approaches its capacity. The solution procedure iterates between solutions of primal and dual variables until the convergence is achieved.

5.1.1 Forecasting future traffic condition with PFE

Forecasting can also be done using PFE. The traffic counts constraint will be removed due to non-availability of future traffic counts. The production and attraction constraints can be calculated from the forecasted land use data. Base line trip table is also used as target trip table to reflect the relationship of travel impedance and trip interchange between OD pairs established for the base year. The target trip table is estimated by scaling the base trip table to match the future total demands which are trip productions and attractions forecasted for the future year. The results of the PFE forecasting program are future path flow estimates which can be aggregated to form future OD trip table and future link flow estimates. The path flows in the forecasted year are derived using the equation given below:

$$f_k^{rs} = exp\left(\theta \cdot \left(-\sum_a t_a \delta_{ka}^{rs} + o_{rs^+} + o_{rs^-} + \rho_r + \tau_s\right)\right), \quad \forall k, rs. \qquad \textbf{5-10}$$

where, ρ_r = Dual variable of zonal constraint at origin *r*

τ_s = Dual variable of capacity constraint at destination *s*

5.2 Visual PFE[6]

Visual PFE is developed by Utah State University with the funding from California Department of Transportation under the PATH program. It uses Path Flow Estimator to model network flows. The user interface is done using Visual Basic .NET and the output is presented in the form of spreads and graphics using MapObjects. The Visual PFE software's graphical user interfaces (GUI) are presented in Figure 5.1 and Figure 5.2.

---

[6] Manual of Visual PFE

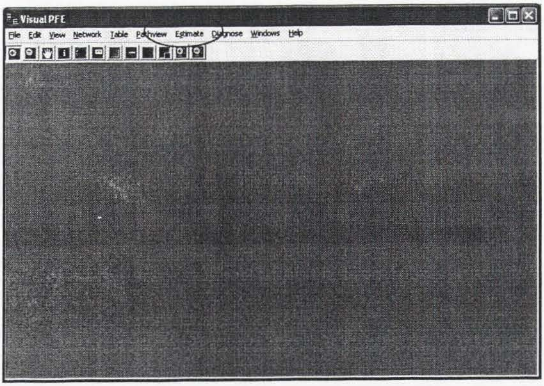


Figure 5-1: Visual PFE main graphical user interface

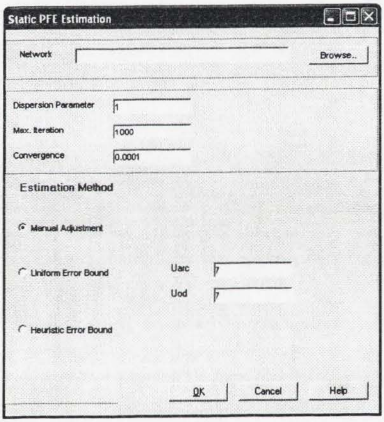


Figure 5-2: Visual PFE's dialog box

5.2.1 Data preparation

Visual PFE requires three input files as listed below to perform the estimation of O-D trip table.

- **pfe-file** – describes network topology (i.e., connectivity), contains link characteristics and observed link volumes,
- **pfx-file**– describes the skeleton (structure) of O-D trip table to be estimated as well as target O-D volumes (if available), and
- **pfp-file** – defines a certain set of routes that the user may particularly want to include into the estimation.

Common network data listed below is utilized to prepare these Visual PFE input files.

Node-related information

- Node's ID
- X-coordinate of node
- Y-coordinate of node

Link-related information

- Starting node of link
- Ending node
- Functional class of link

  Note: *the highest (maximum) functional class, defined by the user, must be reserved for centroid connectors. This is required for path-building purpose.*

- Link capacity
- Free-flow speed
- Link length
- Two parameters of BPR link cost function, α and β respectively

$$\text{travel time} = \text{free - flow travel time} \times \left(1.0 + \alpha \cdot \left(\frac{\textit{link volume}}{\textit{link capacity}} \right)^{\beta} \right)$$

- Observed link volume
- Measurement error of traffic count (e.g., percentage error – ±5%)

Figure 5.3 shows the sample format of the .dbf file prepared for Cache county network.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	LINK ID	A	B	AX	AY	BX	BY	FUNCLASS	CAPACITY	SPEED	DISTANCE	ALPHA	BETA	CNT	ERR
2	1	1	1130	1906	2642	1907	2487	63	20000	15.00	2.51	0.15	4.00	-1	0.00
3	2	1	1696	1906	2642	2032	2639	63	20000	15.00	2.05	0.15	4.00	-1	0.00
4	3	2	1044	2043	2522	2009	2557	63	20000	15.00	0.78	0.15	4.00	-1	0.00
5	4	2	1840	2043	2522	2028	2486	63	20000	15.00	0.62	0.15	4.00	-1	0.00
6	5	3	1838	1957	2523	1958	2487	63	20000	15.00	0.59	0.15	4.00	-1	0.00
7	6	4	1044	1995	2557	2009	2557	63	20000	15.00	0.23	0.15	4.00	-1	0.00
8	7	4	1836	1995	2557	1991	2486	62	20000	10.00	1.14	0.15	4.00	-1	0.00
9	8	5	1377	2027	2463	2036	2448	63	20000	15.00	0.27	0.15	4.00	-1	0.00
10	9	5	1840	2027	2463	2028	2486	63	20000	15.00	0.39	0.15	4.00	-1	0.00
11	10	6	1490	1837	2440	1838	2502	63	20000	15.00	1.01	0.15	4.00	-1	0.00
12	11	6	1844	1837	2440	1838	2356	63	20000	15.00	1.35	0.15	4.00	-1	0.00
13	12	6	1846	1837	2440	1810	2439	63	20000	15.00	0.44	0.15	4.00	-1	0.00
14	13	7	1130	1906	2432	1907	2487	63	20000	15.00	0.88	0.15	4.00	-1	0.00
15	14	7	1848	1906	2432	1904	2356	63	20000	15.00	1.22	0.15	4.00	-1	0.00
16	15	8	1850	1959	2421	1959	2356	63	20000	15.00	1.04	0.15	4.00	-1	0.00
17	16	9	1046	1990	2422	2004	2421	63	20000	15.00	0.24	0.15	4.00	-1	0.00
18	17	9	1836	1990	2422	1991	2486	62	20000	10.00	1.05	0.15	4.00	-1	0.00
19	18	9	1852	1990	2422	1989	2356	63	20000	15.00	1.06	0.15	4.00	-1	0.00
20	19	10	1856	2041	2397	2004	2397	63	20000	15.00	0.59	0.15	4.00	-1	0.00
21	20	10	1858	2041	2397	2041	2373	63	20000	15.00	0.39	0.15	4.00	-1	0.00

Figure 5-3: DBF file for creating input files for Visual PFE

O-D related information

- ID of origin node
- ID of destination node
- Target O-D flow (if available)
- Confidence of target value (if available)

User-specified path information

- List of paths
- Number of nodes defining path
- A sequence of nodes defining paths (from the origin to the destination)

A MS-Excel macro is used to run the link information file (Figure 5.3) to get the input files for Visual PFE. The user needs to give information about the name of the output file (which is the input file for Visual PFE), location where output file needs to be placed, location of the link information file, and number of zones in the macro. On running the macro successfully, the three input files as discussed above (*.pfe, *.pfx, *.pfp) are created. The OD structure file (.pfx) is edited to include the trip productions and attractions of all zones. For zones with no trip generation information (ex: external zones) the trip production and attractions are given as -1. This informs Visual PFE that it needs to estimate that value (all unknown values should be assigned -1). Figure 5.4 presents the flow chart on the operation of Visual PFE.

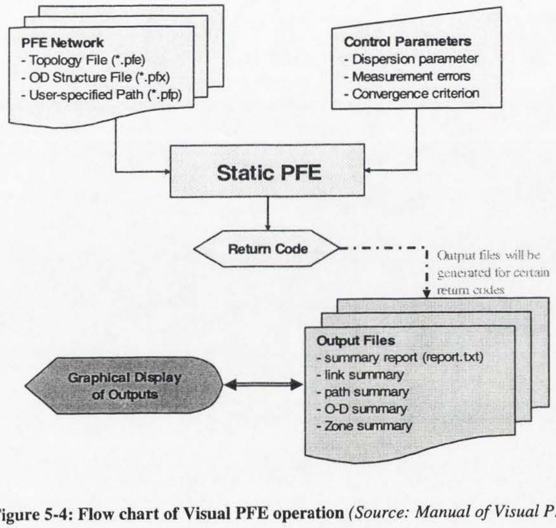


Figure 5-4: Flow chart of Visual PFE operation *(Source: Manual of Visual PFE)*

The first control parameter 'dispersion parameter' is the cost-sensitivity in the logit SUE model. It indicates how sensitive the road users are to the travel cost. A small value of "dispersion parameter" (e.g., between 0.01 and 1.00) is recommended to prevent numerical problem of the algorithm. It is however important to calibrate this parameter to properly reflect the behavior of road users (before the estimation). The second parameter 'max.iteration' is the maximum number of iterations allowed for the adjustment of path flows to match observed link volumes. The default value (recommended) is set at 1,000 iterations and the maximum value allowed to set is 5,000 iterations. The algorithm can terminate before the specified iterations if it reaches the target values. The third parameter 'convergence', is the criterion used for terminating the Visual PFE. The program will check for the maximum change of balancing factors. If the maximum change is less than the convergence criterion, the program will be terminated. The convergence value of 0.01 is recommended to obtain a reasonable solution within a reasonable computational time. A more strict convergence criterion (e.g., smaller value – <0.01) can also be used, of course, with the expense of higher computational time.

There are three methods of adjusting the inconsistencies in the traffic count data namely, manual, uniform and heuristic. The details about this information are given in the Visual PFE manual. For this work, manual adjustment is used in which error bounds are specified by the user for each link count.

The return codes are numbered from 0 to 6. The return codes 1 to 3 do not create any output files. Estimation is meaningful only when the return code '0' is given even though return codes 4 to 6 can create the output files.

The output files are - link summary file, OD summary file, path summary file, zone summary file along with a summary report. Visual PFE can graphically display these files to the user.

Chapter 6 Procedure

6.1 Preparation of data

The requirements of PFE (as explained in Chapter 5) are network characteristics, traffic count data, productions and attractions of different zones and target trip table if available.

6.1.1 Network characteristics and traffic counts

The transportation network characteristics of Cache County are retrieved in the form of a link shape file and a node shape file from the CUBE/ Voyager network file (source: CMPO) by exporting the network file using CUBE/TP+. Using the information in the shape file, base input file for MS-Macro is prepared (see Figure 5.3) to generate three input files - topology file (.pfe), OD structure file (pfx), and path specification file (pfp) for Visual PFE.

6.1.2 Productions and attractions data

The demographic and socio-economic data used are from Cache County travel demand model (source: CMPO). There are 133 identified TAZs, numbered from 1 to 101, 201 to 228 and 230 to 234. Appendix-5 shows the demographic and socio-economic data estimated by the year 2002 for Cache County.

For finding the productions and attractions for each of the zones, the Institute of Transportation Engineers (ITE) published trip generation rates are used. The published trip generation rates represent 3,750 weighted averages of traffic studies conducted throughout United States and Canada since 1960's. Trip generation rates and curve fitting equations have been developed for an average weekday, Saturday and Sunday, for weekday morning and evening peak hours of the generator, and for the weekday morning and evening peak hours on the adjacent street traffic, i.e., between 7 A.M. and 9 A.M. and 4 P.M and 6 P.M. ITE trip generation rates are given as factors of 1000 square feet of area, or per employee. The trip generation rates published by ITE are in the form of vehicle trips (Trip Generation, 6^{th} edition).

For estimation in PFE, the peak hour productions and attractions need to be calculated. So the rates used in this study are peak one hour factor for A.M peak between 7 A.M to 9 A.M. The demographic and socio economic data used are number of single family households, number of multifamily households, number of retail business employees, number of industrial employees, and number of other business employees. For all except retail business establishments the rates per employee are used. For retail business the conversion rates are not available in the form of per employee. Retail business employee numbers are converted in terms of area using a conversion factor of 400 sq. feet per employee based on the standards specified in the Urban Land Institute's Development Handbook (Report on Demographic Data and Development Projections). Different retail activities are mentioned in the ITE trip generation rates. After reviewing all the retail activities available in ITE published trip generation it is concluded that the activities "shopping center" and "supermarket" better represent the general retail activity. The average trip generation rate per 1000 sq. feet for "shopping center" and "supermarket" is taken. Table 6.1 gives the peak hour trip generation factors for Cache County.

Table 6-1: Peak hour ITE trip generation rates used

Sl No.	Socio-economic/land use	ITE peak hour TG factor	Trip Production	Trip Attraction
1	Single family house hold	0.75 per unit	75%	25 %
2	Multi family house hold	0.51 per unit	84%	16%
3	Retail employment	2.14 per 1000 sq feet	39%	61%
4	Industrial employment	0.44 per employee	17%	83%
5	Other employment	0.48 per employee	12%	88%

(Source: Trip Generation, 6th Edition, Volumes 1 and 2)
Single family household - Single Family Detached Housing, Land use code 210
Multi family household - Apartment, Land use code 220
Retail employment - Average of Shopping Center, land use code 820 and Super market, land use code 850
Industrial employment - General Light Industrial, Land use code 110
Other employment - General Office Building, Land use code 710

Using the above trip generation factors, demographic and socio-economic data are converted into trip productions and attractions. The final step in trip generation is to match the production ends and attraction ends, since the logic demands that the total productions should be equal to total attractions. Based on the approach in general practice, the attraction ends are scaled or normalized, to equal the total number of production ends.

6.1.3 Peak hour factors:

The selection of an appropriate hour for planning, design, and operational purposes is a compromise between providing an adequate level of service (LOS) for every (or almost every) hour of the year and economic efficiency. Customary practice in United States is to base highway design on an hour between the 30th- and the 100th-highest hour of the year (HCM, 2000).

The K-factor is the design hour volume (30^{th} highest hour) as a percentage of the annual average daily traffic. An automatic traffic recorder (ATR) for continuous traffic monitoring is needed to identify which hour is the 30^{th} highest hour of travel during the year at a given location (www.fhwa.gov). The state of Utah has around 100 such ATR stations that are spread all over the state. The locations of ATR stations are shown in the Appendix 4. These stations provide average traffic counts by weekday (Monday through Friday) and by weekends (Saturday and Sunday) for each month of the year. They also provide the 1st, 10th, 20th, 30th, 50th, and 100th highest peak hour traffic count and their corresponding percentage of the annual average daily traffic count. Presently, the data for the years from 1994 to 2004 is available for most of the stations. Cache County has three ATR stations numbered as 363 located at 0.8 miles North of SR 101, Wellsville, 510 located at 100 North 319 West, Smithfield and 511 located at 2416 North 800 East, North Logan. The Annual Average Daily Traffic (AADT) and the 30^{th} hourly highest peak hour percentages for the year 2002 for these stations are shown in Table 6.2. The locations of the ATRs are shown in Figure 6.1.

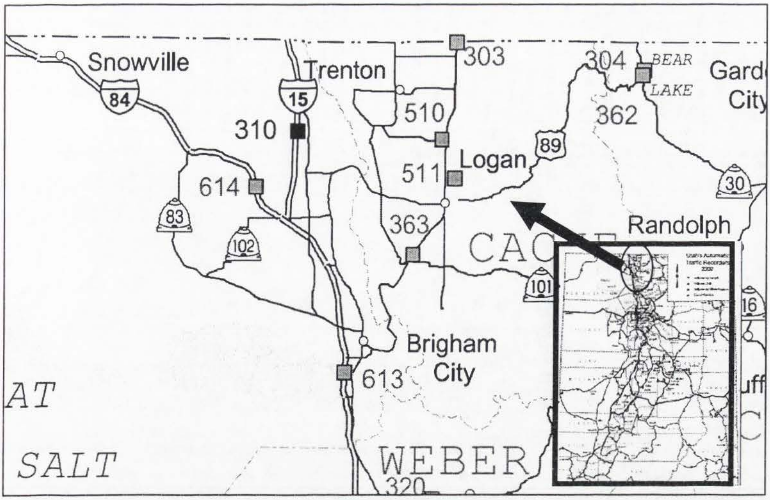


Figure 6-1: Permanent automatic traffic record stations near Cache county maintained by UDOT
(Source: www.udot.utah.gov)

Table 6-2: ATR stations near Cache County and their peak hour percentiles

Station ID	Location	2002 AADT	30 percentile peak hour % AADT
363	0.8 miles North of SR 101, Wellsville	18531	10.6%
510	100 North 319 west, Smithfield	3805	11.1%
511	2416 North 800East, North Logan	5764	12.1%

(source: www.udot.utah.gov)

A simple average of the above peak hour percentage AADTs is taken to convert the AADT on the links of the Cache County transportation network to peak hour counts. The simple average peak hour percentage of AADT is 11.27%. This is within the limits (7-

12%) specified by Highway Capacity Manual for urban roads (Roess, Prassas, and Mc Shane, 2004).

6.1.4 Preparation of input files to Visual PFE

To prepare the input files for the VisualPFE, the data needs to be converted into appropriate form. First, using the peak hour factor from permanent automatic traffic recording stations (ATRs), the traffic counts (AADT) on the observed links are converted into peak hour volumes. The OD input file to PFE (.pfx) is edited with the peak hour productions and attractions as shown in Figure 6.2.

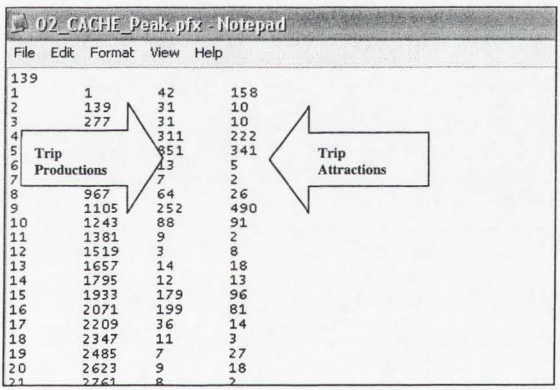


Figure 6-2: Editing of OD structure file with zonal productions and attractions

The prepared input files are loaded into Visual PFE. As explained in Chapter 5, the input parameters need to be given. As a starting point, dispersion parameter as 0.1, number of iterations as 5000 and convergence as 0.1 are given.

The dispersion factor is generally defined as the inverse of average travel time between all zones. The origin-destination travel time survey conducted as part of development of travel demand model of Cache County is not available to calibrate the dispersion factor. In the absence of the data, a unique dispersion parameter is deduced in this case by iterating the value of dispersion parameter as inverse of average inter zonal travel time till the average travel time of previous and present iterations match. Figure 6.3 shows the flow chart of the process.

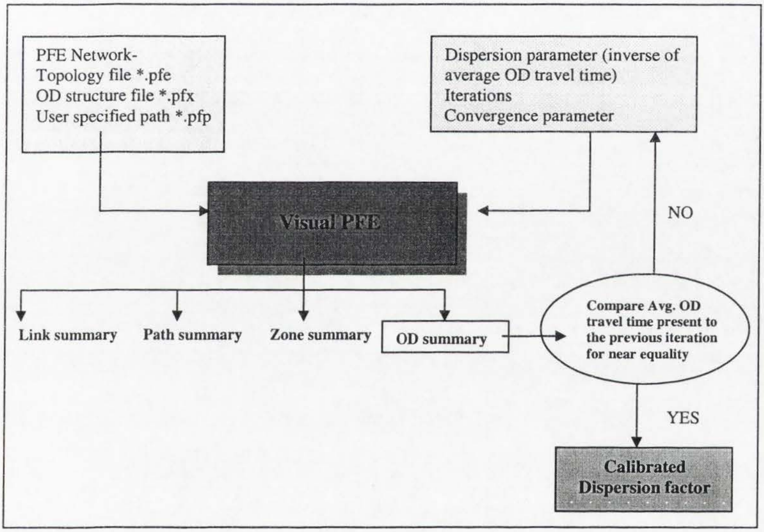


Figure 6-3: Calculation of dispersion factor

The results using this dispersion factor as control input parameter, are used in the analysis are presented in the Chapter 7. The average zonal trip length estimated by PFE is within 10% of the modeled average trip length by the conventional model (shown in Chapter 7).

Chapter 7 Analysis

7.1 Trip generation comparison

The analysis starts with the comparison of trip generation estimations of ITE trip generation rates and regression models of four step model. Then the trip table estimations are compared along with trip length frequency distributions from both the models. Link flows estimated by both models are then compared on the observed links and the unobserved links. Finally, screenline check is done for the estimations of both model link flows.

The components of trip generation, trip productions and trip attractions are estimated using ITE trip generation factors (as explained in Chapter 6) for VisualPFE and using regression analysis for the conventional four step model. The ITE trip generation rates estimate vehicle trips and the regression analysis of the conventional model estimate person trips.

The conventional four step model was modeled for daily trips and the productions and attractions calculated are for daily demand. In the trip distribution and mode choice steps, person trips are converted into vehicle trips. The four step model also considered the external-external trips, internal-external trips, external-internal trips and Utah State University as a special traffic generator. PFE in its present version does not distinguish the difference between an internal zone and an external zone. Some of the internal zones traffic may be assigned to the external stations, to satisfy the constraints. Figure 7.1 and

Figure 7.2 show the vehicle productions and attractions estimated by the two models in the form of a line graph. The zones identified are the ones with high demand.

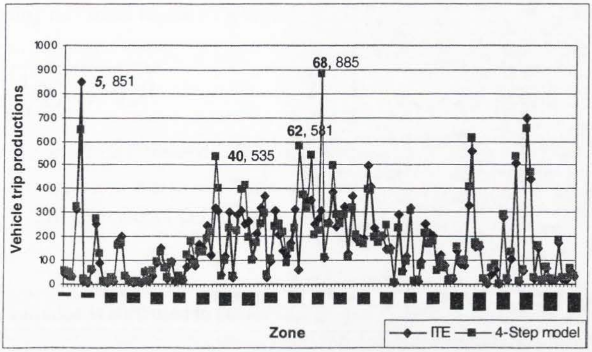


Figure 7-1: Vehicle trip production estimates- 4 step model vs. ITE

(Numbers in BOLD show TAZ ID, regular numbers show the ITE estimated demand)

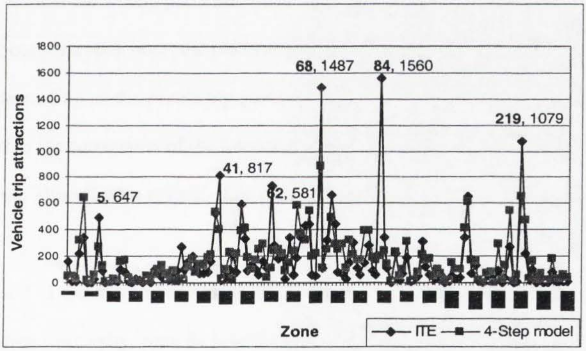


Figure 7-2: Vehicle trip attraction estimates- 4 step model vs. ITE

(Numbers in BOLD show TAZ ID, regular numbers show the ITE estimated demand)

The total number of vehicle trips estimated using ITE trip generation factors is 21000 and by conventional four-step model is 22353. The vehicle trips estimated using ITE generation rates are about 6.05% less than the vehicle trips estimated by the four-step model for the peak hour. The root mean square error (Equation 7.1) calculated for the vehicle production estimates is 83.06 and the attraction estimates is 180.56. The equation for calculating root mean square error is given below.

$$RMSE = \sqrt{\frac{1}{N}\sum_{rs}(q_{rs} - q_{rs}^{'})^2} \qquad 7.1$$

where q_{rs} - PFE estimated O-D volume

$q_{rs}^{'}$ - four step model estimated O-D volume

N – Number of O-D pairs

This large variation is attributed to the use of the common peak hour factor to convert the daily trips of the four step model to peak hour trips in this study. In the estimation of daily trips the trip productions and attractions of each zone are equal. For a daily trip, the starting point and the ending point are the same at the end of the day which is not a case in a peak hour. In a peak hour, the productions and attractions (P-A) for each zone are different depending on the zone's trip generation characteristic. By using a common peak hour factor in the conversion of the four step model daily trips, peak hour characteristic is not captured. This effect is magnified in certain zones with high peak hour demand.

Zones 40 and 41 are central business district zones on the main street, zones 62 and 68 are Utah State University, zone 84 is an industrial area (ICON factory) and zone 219 is Hyrum's industrial area (meat industry). These zones attract high number of trips in the morning peak hour and the differences in the estimation of the peak hour trips in these zones is obvious because of the use of the common peak hour factor.

The trip generation rates published by ITE are contributed on voluntary basis by different sources. ITE has only cleaned and validated the data (Trip Generation, User's Guide, pp 13). The data is collected over years and periodically updated. ITE advises to use these rates, keeping in view the geographical location and characteristics of land use. Moreover, generally the peak hour trip data collection will be more accurate than the daily trip data collection. Historically, work trips dominate the morning peak hour and most of the work trips are not shared ride. ITE advises modification of trip generation rates in presence of the shared ride and public transport systems (Trip Generation, User's Guide, pp1). For Cache County, ITE trip generation rates have overestimated the daily trips by more than 50% over the four step model. In a day, non-work trips which are dominated with shared ride are influential in vehicle trip estimations. If the shared ride is not properly considered in ITE trip generation rates, the daily trips will be overestimated. These issues in the estimation of trip generation rates might be the reason for high estimation of daily trips in this case.

7.2 Trip table comparison

The OD tables of both models are compared to see the differences in the spatial distribution. Conventional model's final trip table (daily trips) is factored with the peak hour factor of 11.27% (explained in Chapter 6) to get the peak hour trip table. Figure 7.3 is a ramp color plot on gray scale showing the differences in the trip tables estimated by PFE and the four step model.

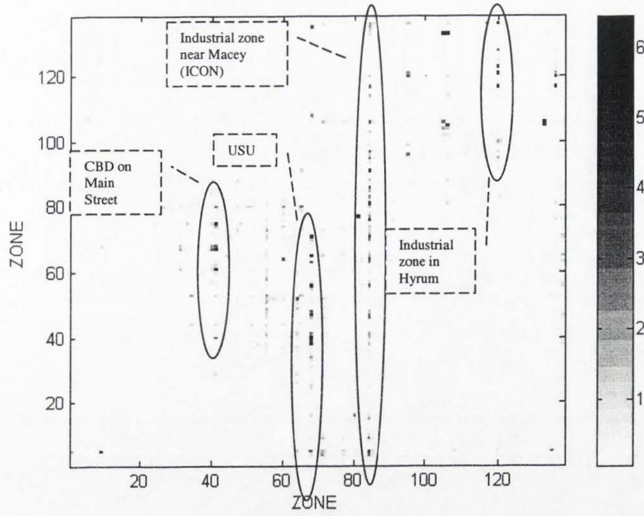


Figure 7-3: Estimated OD trip table differences – PFE vs. 4-step model

The above color plot shows that the estimations of both models are quite close except in zones 40, 41, 62, 68, 84 and 219. The use of the common peak hour factor to reduce the daily trips to peak hour trips is the reason for the variations in these zones. Root mean square error (RMSE) of the trip table estimations between the two models is found to be 4.13, which is much less than the RMSE of P-A estimates. The large differences in the P-A estimations in few zones have smoothened out spatially in the trip table causing the differences in the OD flows of these zones not showing the same level of difference. These differences even though got smoothened out in trip table estimation, yet get reflected in link flow estimation. Section 7.4 shows the effect of this in the link flow estimates.

The following observations are made in the trip table comparison:

- PFE has estimated 0.1 % of the OD pair's demand, 50 vehicles more than the conventional model estimate.
- PFE has estimated 0.3% of the OD pair's demand, 25 vehicles more than the conventional model estimate.
- PFE has estimated 2.64% of the OD pair's demand, 5 vehicles more than the conventional model estimate.

The above observations show that certain OD pairs demand is estimated higher by PFE than the four step model due to the use of common peak hour factor in conversion of daily trips to peak hour trips.

7.3 Travel time estimation

Trip distribution is calibrated based on the trip length frequency distribution of that area. This is generally done based on free-flow travel time. Figure 7.4 shows the trip length frequency distribution for free-flow travel time estimated by the four step model.

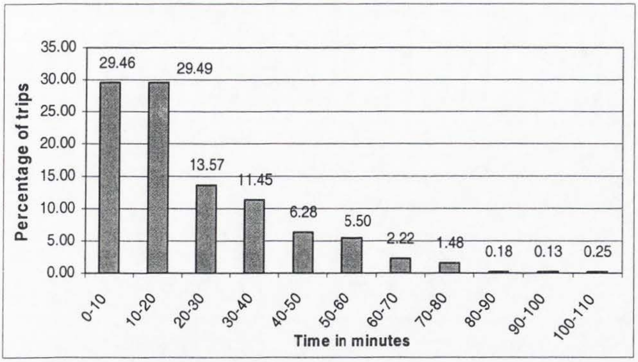


Figure 7-4: Trip length frequency distribution for free-flow travel time in the four step model

It is interesting to see the trip length distributions at congested travel times of PFE and the four step model. Figure 7.5 and Figure 7.6 present the trip length frequency distributions of both models.

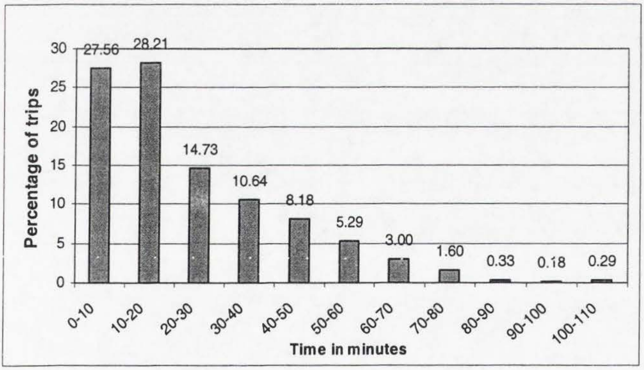


Figure 7-5: Trip length frequency distribution for peak hour estimated by the four step model

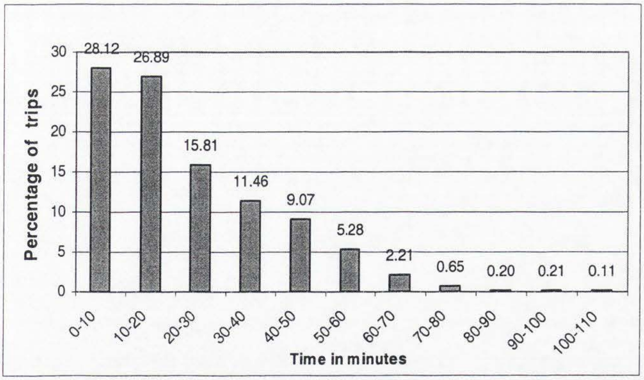


Figure 7-6: Trip length frequency distribution for peak hour estimated by the PFE

The trip length frequency distribution under the free-flow conditions and under the congested conditions does not differ much in the four step model. Figure 7.7 shows the shift of the trips towards higher travel time due to peak hour congestion. For example, 462 trips from travel time range of 0-10 minutes and 313 trips from travel time range of 10-20 minutes moved to higher travel time ranges.

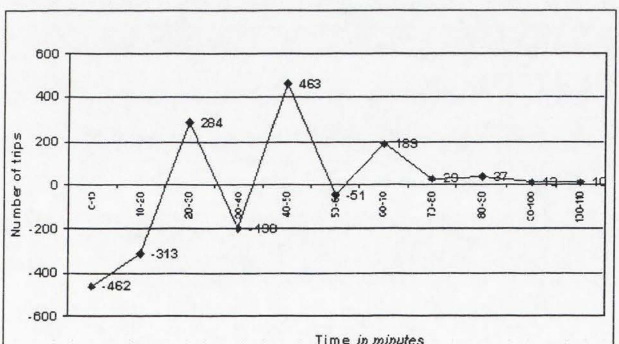


Figure 7-7: Shift in travel times of OD trips on congestion

(congested travel time – free-flow travel time)

The trip length distributions estimated by both the models on congestion are also similar. Kolmogorov-Smirnov goodness-of-fit test is conducted to determine whether two empirical distributions differ significantly. For this test, the two datasets are normalized to get a cumulative distribution that ranges from 0 to 1. The maximum absolute difference between the two cumulative distributions is taken and multiplied with square root of the size of dataset to get the test statistic ($D_{calculated}$). If $D_{calculated}$ is less than the critical value, D_{crit} (Davis, 2002), null hypothesis fails to get rejected, implying that the given datasets have similar distributions. In this case, $D_{calculated}$ =0.06 is less than D_{crit} =0.35 (Davis, 2002), resulting in the failure to reject null hypothesis which says that both the datasets follow same distribution. The small variations in the two distributions are attributed to the conversion of daily trips in the four step model and the influence of external trips on trip table estimation. The Kolmogorov-Smirnov test for trip length frequency distributions under free-flow conditions and congested conditions also fails to reject the null hypothesis, inferring that they also follow same distribution. Appendix 6 shows the details of this test.

The average O-D trip length estimated by the four-step model is 23.68 minutes under free-flow conditions and under congested conditions is 24.91 minutes. PFE has estimated the average O-D trip length as 27.27 minutes under congested conditions. As per the four step model estimate the average trip length has increased by 1.23 minutes due to congestion. PFE has estimated the average trip length 2.36 minutes more than the four-step model under congested conditions.

The reason for PFE's high estimation of trip length is attributed to the route choices used in these models. The route choice assignment used in the four step model works similar to User Equilibrium (Sheffi, 1984). But the route choice assignment used in the PFE is logit SUE. The User equilibrium assignment assigns all the flow to the shortest route and the logit SUE assignment assigns the flows based on the knowledge of the traveler (through the specified dispersion factor). More the knowledge of the traveler about the network, more closely the traffic assignment will follow the user equilibrium assignment. Another factor is the use of daily trip volume-capacity ratio in the BPR function, which results in the underestimation of travel time in the four step model. This effect is explained in detail in Section 7.4.

7.4 Link flow estimations

In the four step model the network traffic is loaded by decreasing the size of the increments in every iteration using "incremental capacity restrained assignment". The rationality behind the selection of size and number of increments was not explained in the original model's calibration and validation report.

Logit SUE assignment used in PFE takes into consideration the imperfect knowledge of the traveler (through dispersion factor) in assigning the flows to the network. This may result in the assignment of a few trips to the non-shortest routes. The link flows estimated by PFE and the four step model are shown in Figure 7.8 and Figure 7.9 respectively.

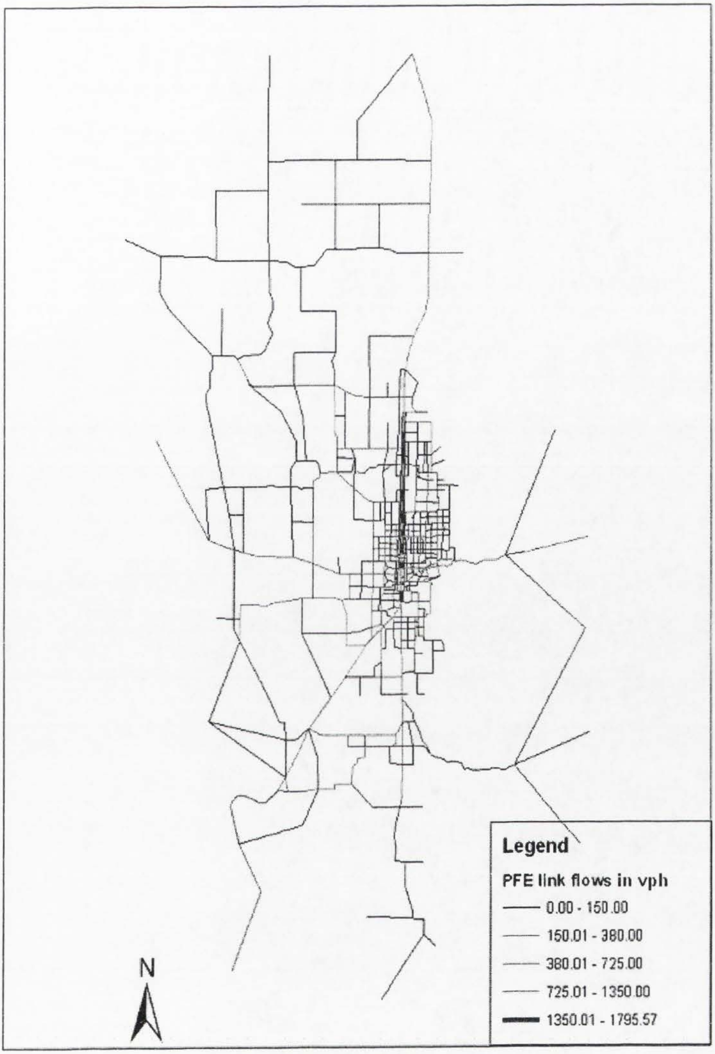


Figure 7-8: Links flows estimated by PFE- Cache County

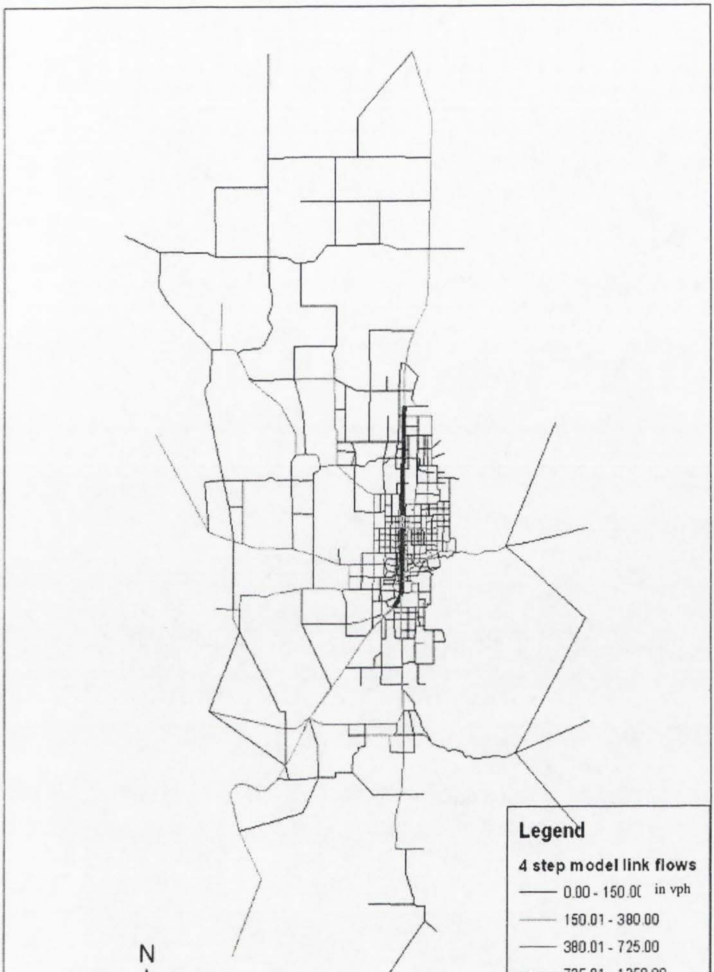


Figure 7-9: Link flows estimated by the four step model- Cache County

Link flows estimated in both models show similar patterns of traffic intensities. The Main Street (US 89/91) is carrying heavy traffic in the peak hour. But the link flows estimated by the four step model are higher than that of PFE on the Main Street. On the links connecting the high attraction zones PFE has estimated higher than that of the four step model. This is quite evident on the links connecting the zones 68 (USU) and 219 (Hyrum).

The reasons for high estimation of flows by the four step model are the route choice used in this model and the differences in the trip table estimations. Another reason is the modification done in the Bureau of Public Roads (BPR) function in the four step model. The general BPR function calculates the travel time for the peak hour. The volume-capacity (V-C) ratio used in the BPR function is for the peak hour. But in the Cache County four step model, the V-C ratio used was for daily trips[7] due to which the travel time distribution got flattened there by giving lower travel times than the actual in the peak hour. This imparts less congestion than the actual, resulting in high estimation of flows on a few links. On the links connecting a few high attraction zones with multiple routes the differences are not visible due to the distribution of trips on multiple links. On observation of the link flows, it is found that the ITE trip generation rates tend to slightly overestimate the trips in the high attraction zones. Figure 7.10 shows the link-wise differences in the estimates of PFE and the four step model. Appendix 7 shows the detailed maps showing the differences in link flows estimated by both the models.

[7] Daily capacity is calculated on the assumption that total traffic demand in a day will be served within 10 hours of day. So Daily capacity = 10 x number of lanes x vphpl.

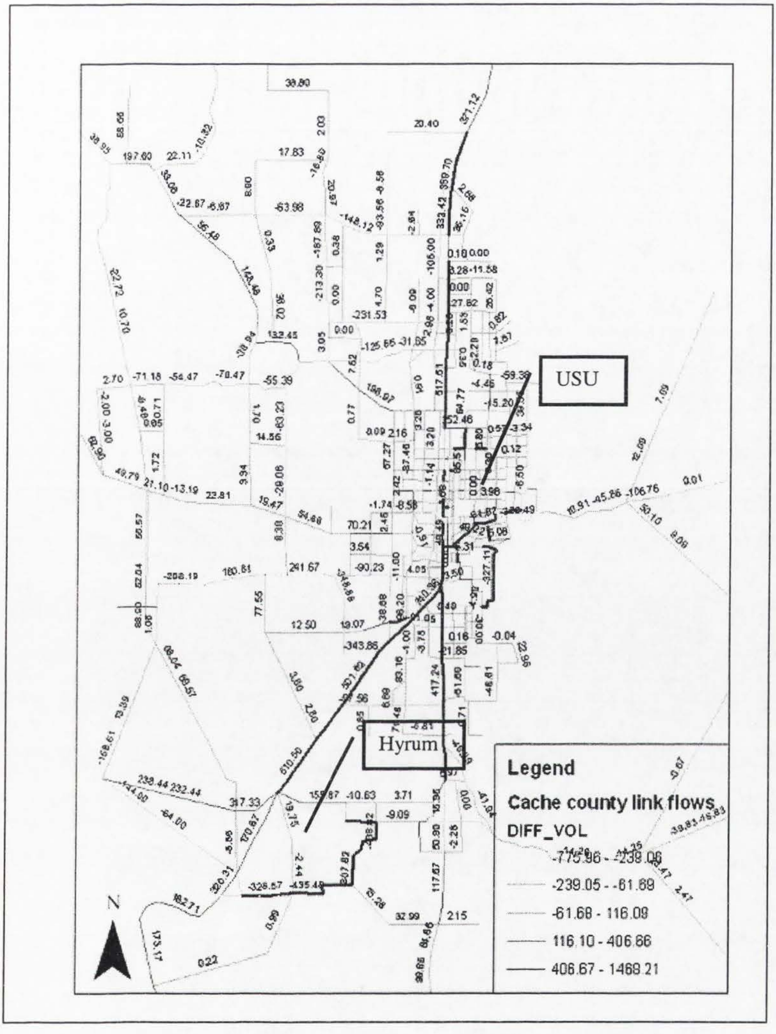


Figure 7-10: Link flow differences in the estimates of PFE and conventional model

(The thick red lines represents the link flow estimates, differ more than 400 vph.)

(Figures showing the link flows estimate differences in detail and color ramp map are given in appendix 7)

The estimated link flows of PFE and the four step model are compared with observed traffic volumes. Out of 2176 links, the traffic counts are available on 205 links, which is about 9.42%. The quality of the trip table depends on the number of traffic counts and its locations. The number of traffic counts locations and their positions on the network are vital in the accuracy of estimate in PFE (PFE Manual). Generally more the traffic count observations, better is the trip table estimated by PFE. This is because, traffic counts help in shaping (identifying) the trip table distribution (PFE manual). Figure 7.11 and Figure 7.12 show the scatter plots of the link flows estimated by both the models vs. traffic counts on the corresponding links.

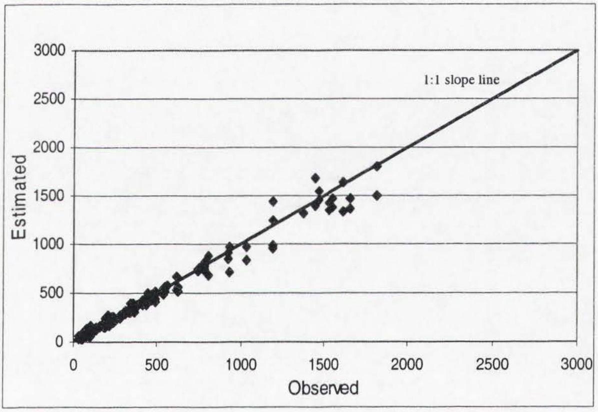


Figure 7-11: PFE estimated link flows vs. observed link flows

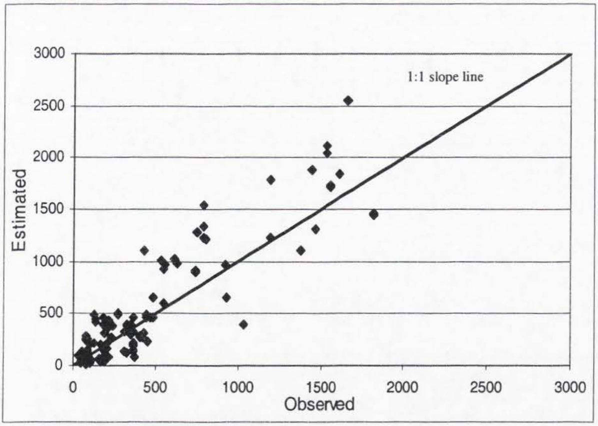


Figure 7-12: Four step model estimated link flows vs. observed link flows

The accuracy of the estimates is measured by percentage root mean square error (Equation 4.6). The percentage root mean square error between traffic counts and PFE estimation is 16.02% and that of the four step model is 62.02%. On the primary arterials, PRMSE between traffic counts and PFE estimation is 12.17% and that of the four step model is 40.29%. This shows that the PFE has estimated the link flows nearer to ground counts than the four step model. PFE assigns the link flows by trying to match the traffic counts, zonal productions and attractions through its constraints in its formulation. In contrast, in the four step model the calibration and validation of the model was done using screen line count comparisons and also facility specific measures. Due to this reason, PFE estimates of the link flows are closer to the traffic counts than the four step model.

Graphically the data points that are close to the 1:1 slope line indicate the accuracy of the estimation to the real world flow. To compare the efficiency in estimation of the actual flows by both models, a statistic called Nash-Sutcliffe statistic (Nash and Sutcliffe, 1970) is used. This statistic is proposed by Nash and Sutcliffe (1970) to compare the modeled river flows to the actual flows. This statistic uses regression analysis sum of squares concept and this statistic is given in Equation 7.2.

$$\text{Nash-Sutcliffe statistic } E = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum(O_i - \overline{O})^2} \qquad 7.2$$

where O_i - is the observed flow in that link i

P_i - is the model estimated flow on link i

$\overline{O}$ - mean of observed flows on all links.

The numerator term is analogous to residual variance of the regression analysis and the denominator is analogous to initial variance or no model value of numerator. Use of these terms enables the efficiency of model to be defined by E (analogous to the coefficient of determination) as the proportion of the initial variance accounted for, by that model (Nash and Sutcliffe, 1970). The range of the statistic is between $-\infty$ to one, $-\infty$ indicating no relation between model estimation and actual values and, one indicating exact estimation by the model.

Computing Nash-Sutcliffe statistic for the PFE link flow estimation, gives a value of 0.98 which is an indicator for "closeness-to-fit". The same statistic calculated for the four step model link flow estimation gives a value of 0.64.

Finally, the link flows estimated by PFE and the four step model are compared using percentage root mean square error. The percentage root mean square error between PFE and the four step model estimates of all link flows is 77.28 %. Table 7.1 shows the percentage root mean square errors on the observed and unobserved links.

Table 7-1: Percentage root mean square errors for link flow estimations

PRMSE		**Four step model**	
		Observed	Unobserved
PFE	Observed	55.26	-
	Unobserved	-	81.4

This increase in PRMSE on unobserved links in the estimations of PFE and the four step model is obvious because, PFE estimates the link flows on traffic count constraints and the four step model is calibrated based on screenline analysis. Nash-Sutcliffe statistic is also calculated to see how close the PFE estimation is with the link flows estimated by the four step model with respect to a statistic.

Nash-Sutcliffe statistic of all link flows is found to be 0.728 which is an indication that PFE has estimated the link flows similar to that of estimated by the four step model. On observed links the Nash-Sutliffe statistic between the two models is 0.89 and on unobserved links it is 0.71. Generally, Nash-Sutcliffe statistic anything greater than 0.6 is considered to be good in hydrological models. (Engel and Lim, 2004)

The four step model estimates the link flows more than 2000 vph on certain links on main street (links on main street are 2 lane urban roads). This is practically very high value for an area like Cache County. The capacities of the links on the Main Street are 1300 to1700 vph. As already discussed, PFE has estimated the link flows closer to traffic counts than four step model on these links. As an example, link between nodes 1016 and 1018 is taken on the main street which has got traffic count data. The location and link flow estimation details are shown in Figure 7.13.

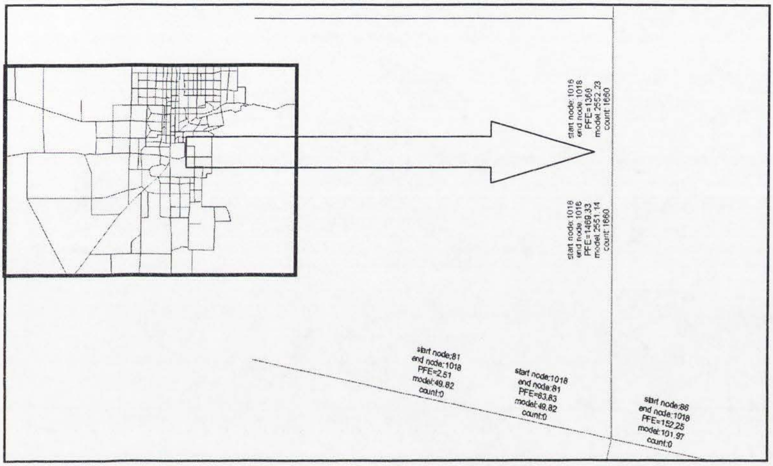


Figure 7-13: Link estimates on a link on the Main Street, Logan

The link traffic counts on peak hour are 1660 vph, in both the directions. The PFE has estimated the flow on the link as 1368 vph for south bound traffic and 1469 vph for north bound traffic. The four step model has estimated the flow on this link as 2552 vph in both the directions. This shows that the four step model is over estimating the flows very highly on Main Street where as PFE estimations are more close to the real world. The peak hour reduction factor cannot be a reason for this high estimation since it is within the limits (7-12% for urban roads) specified by Highway Capacity Manual (discussed in Chapter 6).

7.5 Screenline comparison

On the screenlines identified in the four step model, a comparison of link flow estimates is done between PFE and the four step model. Table 7.2 shows the 2002 base year screen line check between the two models.

Table 7-2: 2002 base year screenline-check between PFE and four step model

Screenline	Link ids	Observed	4-step model	% variation	PFE	% variation
A1	1724-2278	57	72	26.32	64	12.28
A2	2014-2296	102	63	-38.24	104	1.96
A3	1706-2276	364	465	27.75	342	-6.04
Screenline A Total		**523**	**600**	**14.72**	**510**	**-2.49**
B1	1762-1764	205	222	8.29	268	30.73
B2	1754-1756	57	41	-28.07	55	-3.51
Screenline B Total		**262**	**263**	**0.38**	**323**	**23.28**
C1	1738-1740	739	908	22.87	735	-0.54
C2	1814-1994	125	215	72.00	130	4.00
Screenline C Total		**864**	**1123**	**29.98**	**865**	**0.12**
D1	1334-2280	352	393	11.65	330	-6.25
Screenline D Total		**352**	**393**	**11.65**	**330**	**-6.25**
E1	1202-1950	91	67	-26.37	90	-1.10
E2	1116-1874	216	442	104.63	244	12.96
E3	1902-1904	1615	1846	14.30	1642	1.67
E4	1388-1908	193	320	65.80	188	-2.59
Screenline E Total		**2115**	**2675**	**26.48**	**2164**	**2.32**
F1	1932-2014	102	63	-38.24	124	21.57
F2	1800-1934	57	102	78.95	61	7.02
F3	1124-1948	205	164	-20.00	185	-9.76
F4	1100-1330	341	360	5.57	305	-10.56
F5	1168-1520	216	365	68.98	209	-3.24
Screenline F Total		**921**	**1054**	**14.44**	**884**	**-4.02**
Subtotal external		**2001**	**2379**	**18.89**	**2028**	**1.35**
Subtotal internal		**3036**	**3729**	**22.83**	**3048**	**0.40**
Total all screenlines		**5037**	**6108**	**21.26**	**5076**	**0.77**

On all the screenlines, PFE has estimated closer to the observed counts than the four step model except on the screenline B. This is due to the high estimation of trips around zone 68 by ITE trip generation rates. The total of all the screenlines is also better estimated by PFE than the four step model. In the original model's calibration done in 1995, the maximum variation of 14% was allowed in the screenline check. It seems that in the model's update for year 2002, the screenline check is not performed. Use of the common peak hour factor in conversion of traffic counts has slightly overestimated the variation.

Overall, PFE has satisfactorily performed in this study. The differences in P-A estimations between the ITE trip generation rates and regression models is due to the use of the common peak hour factor. These differences have smoothened in the trip table estimations. The trip length frequency distributions estimated by both models are similar. The link flow estimations by PFE have satisfactorily replicated the link flows estimated by the four step model. PFE has better estimated the flows on the observed links than the four step model. On the screenlines also PFE has performed better than the four step model.

Chapter 8 Conclusion and Future Research

8.1 Conclusion

A methodology developed by Lee, Chootinan, Chen and Recker (2006) for modeling network traffic for planning applications is implemented to see the efficiency of PFE in estimating the network flows for Cache County. The results of PFE and its comparison with real world data and with four step model show that PFE is a useful tool for estimating the link flows in small and medium communities. The major differences in the estimates of both models are attributed to conversion of daily trips to peak hour trips in the four step model. In trip generation, trip attraction estimates of both models significantly differ, evident from high root mean square error values (83.06 and 180.56) due to the missing of peak hour trip character in reducing daily trips to peak hour trips. But the trip table estimations from both models are quite comparable with a RMSE of 4.13. This is due to the dis-aggregation of differences in trip estimates over the trip table based on the traffic counts constraint. Even though the total difference in total productions and total attractions remains intact, due to spatial distribution of trips, the difference becomes less evident giving rise to closeness between link flow estimates of PFE and the four step model. PFE has estimated link flows close to the traffic counts than the four step model. On the primary arterials also PFE has estimated closer to the actual counts than the four step model. On both observed and unobserved links PFE has performed better than the four step model.

In Cache County travel demand model the V-C ratio used in BPR function is for daily capacities. Due to this the travel times are underestimated than the true values, giving rise to high estimation of flows on certain links. On the links connecting zones of high attraction PFE has estimated higher than the four step model due to the overestimation of vehicle trips to these zones by ITE trip generation rates.

Using PFE, trip lengths between origin and destinations and the link flows are efficiently estimated with only socio-economic data and traffic counts. Socio-economic data like population, employee data, household numbers and land use data are available with all the small communities since it is mandatory to maintain the data for the purpose of Census. For four step model, the requirement of expensive travel surveys, corridor studies is necessary to calibrate its parameters in all its steps.

PFE estimations can be improved if its limitation of inability of not distinguishing difference between internal and external stations is removed. For complex land uses or larger communities, use of ITE trip generation rates is not advisable due to the difficulty in arriving proper trip generation factor. PFE is more sensitive to the quality, number and locations of the traffic counts than the four step model. PFE by its nature can only estimate for peak hour flows. Future years flows also can be forecasted based on the methodology shown in Chapter 5 after scaling the trip table for future trip productions and attractions.

8.2 Future research

Although this report advocates the procedure developed for travel demand modeling for small and medium communities using Path Flow Estimator, there are several areas that need further investigation. Potential areas of investigation are outlined below:

Comparison with travel demand model developed for peak hour trips

In this report, travel demand model of Cache County developed for daily trips is normalized with a common peak hour factor so that the results can be compared with the estimations of PFE. This has resulted in differences in the estimates of both the models. A better approach would be to compare it with a travel demand model which is modeled for peak hour trips than daily trips.

Forecasting with future condition with PFE

In this report, the results of base year of both the models have been compared. To validate the use of PFE in travel demand modeling is to do forecasting using PFE and compare the results. In forecasting, the projected socio-economic data should be used in production-attraction constraint. The target trip table can be estimated by scaling the base year trip table to match the future demand (sum of total productions and attractions) so that, the relationship of travel impedance and trip interchange between OD pairs is preserved in forecasting process.

References

Anderson, D.M., Sharfi, K., and Sampson, E. G., 2005, A Small Urban Community Direct Demand Model Using Multiple Linear Regression, Presented at 85th annual meeting of *Transportation Research Board.*

Anderson, M. D., 2000, Travel Modeling for Smaller Urban Areas Using a Single Trip Purpose, Research Paper for Federal Highway Administration and Transportation Environmental Research Program.

Annual Report of Socio-Economic Characteristics, Cache County, prepared by Cache County-wide Planning and Development Office, June 2004.

Davis, J.C., 2002, *Statistics and Data Analysis in Geology*, 3rd edition, Wiley Publications, pp. 107-110.

Engel, A.B. and Lim, J.J. 2004, WHAT,(Web-Based Hydrograph Analysis Tool) manual. http://pasture.ecn.purdue.edu/~abe526/resources2/what_abe526/what_howto.doc

Institute of Transportation Engineers, *Trip Generation*, Sixth Edition, Washington DC, 1998.

Khisty, C.J., and Rahi, M.Y., 1987, Evaluation of Three Inexpensive Travel Demand Models for Small Urban Areas, *Transportation Research Record* No. 1283, pp. 70-78.

Lee, M., Chen, A., Chootinan, P., Laabs,W., Recker, W., 2006, Modeling Network Traffic for applications in a Small Community, *Journal of Urban Planning and Development* (forth coming).

Lee, M., Chootinan, P.,Chen, A., Recker, W., 2006, Forecasting Network Traffic for small and medium size communities using Path Flow Estimator. Presented at 85th annual meeting of *Transportation Research Board.*

Mann, W., and Dawoud, M., 2004, Simplified 4- Step Transportation Planning Process for Any Sized Area, Presented at *9th Conference of Transportation Planning for Small and Medium-sized Communities*, Colorado Springs, Colorado.

Manual of Visual PFE, developed by Utah State University for Caltrans.

Model Development and Validation Report, Technical Memorandum- Cache Valley Corridor Study, prepared for Utah Department of Transportation, May 1999.

Nash, J.E., and Sutlicliffe, J.V., 1970, River Flow Forecasting Through Conceptual Models, Part I- A Discussion on Principles. *Journal of Hydrology* Vol.10, Issue 3, 1970, pp 282-290.

Oppenheim, N., 1995, Urban Travel Demand Modeling, John Wiley Publications.

Reeves, P. Integrating Transportation Modeling and "Desktop GIS': A Practical and Affordable Analysis Tool for Small and Medium Sized Communities, *Presented at 9th Conference of Transportation Planning for Small and Medium-sized Communities*, Colorado Springs, Colorado, 2004.

Report on Demographic Data and Development Projections by Tischler & Associates, Inc for City of Casa Grande, Arizona, 2003 pp 6.

Roess, R. P, Prassas, E.S., and McShame, W.R., *Traffic Engineering*, 3rd edition, Princeton Hall Publications, PP109.

Schutz, J.B. Transportation Planning Needs for Small and Medium-sized Community. Millennium Papers, *Transportation Research Board*, 2000.

Schute, B., and Ayash, S. Small Urban Travel Demand Modeling in Oregon, Presented at *9th Conference of Transportation Planning for Small and Medium-sized Communities*, Colorado Springs, Colorado, 2004.

Sheffi, Y., 1985, Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods, Prentice-Hall Publications.

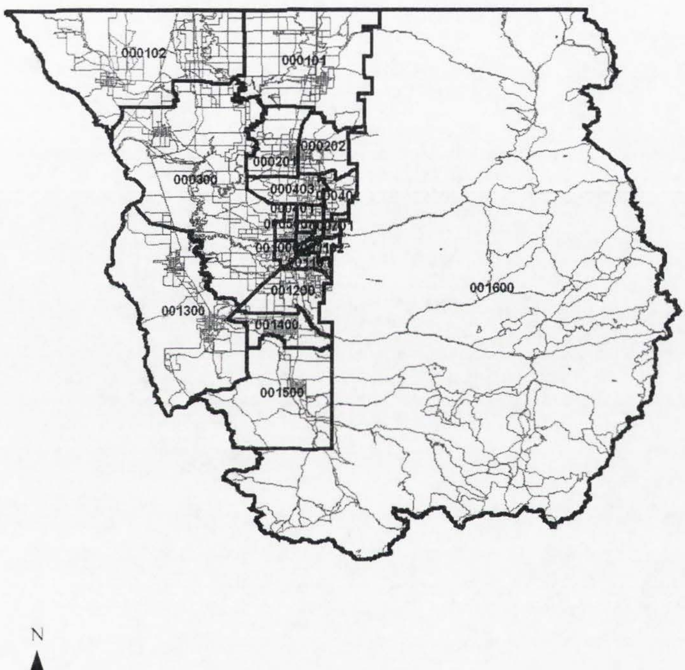
Walker, T., and Reader, P., Travel Demand Model Development for Small Urban Areas, Presented at *9th Conference of Transportation Planning for Small and Medium-sized Communities*, Colorado Springs, Colorado, 2004.

Wegmann, F., Chatterjee, A., and Gregory, R., Transportation Planning for Urban Areas, Tennessee Department of Transportation, 2000.

Yan. S. A Practical Approach to Travel Demand Modeling in Small and Medium Communities, September 1998, Spoken, Washington.
http://www.wsdot.wa.gov/fasc/engineeringpublications/Manuals/TRB.pdf

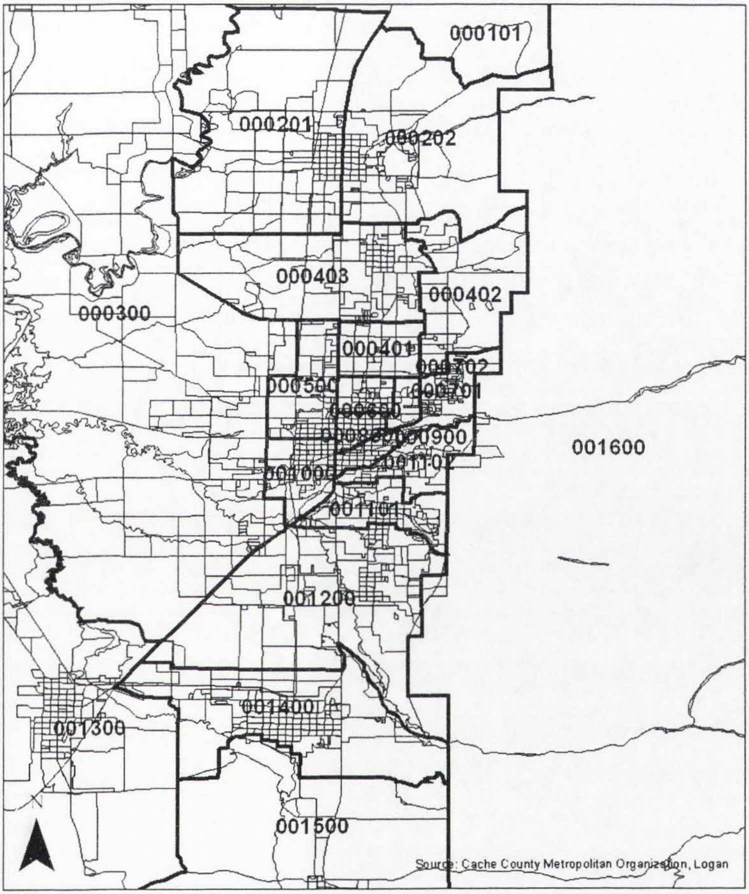
Appendix 1

Map of Census Tracts- Cache county

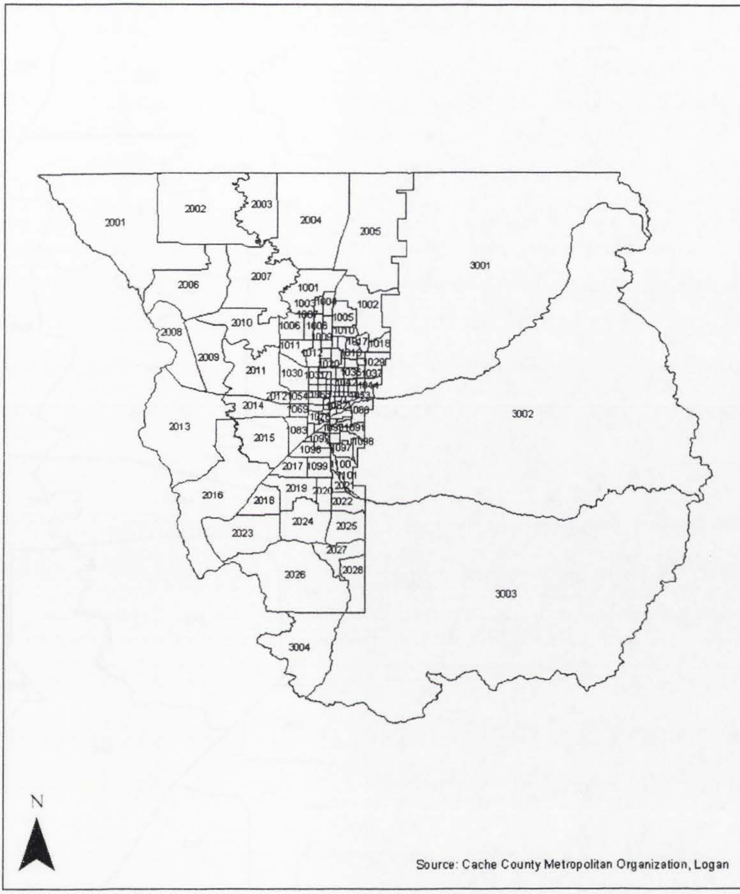


Source: Cache County Metropolitan Organization, Logan

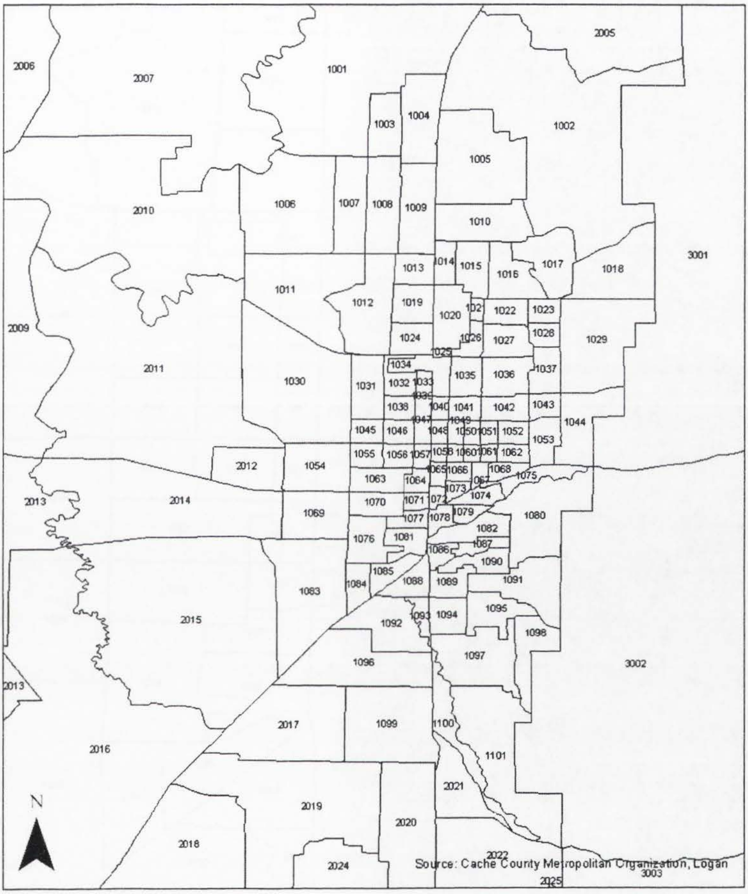
Map of Census Tracts- Cache county



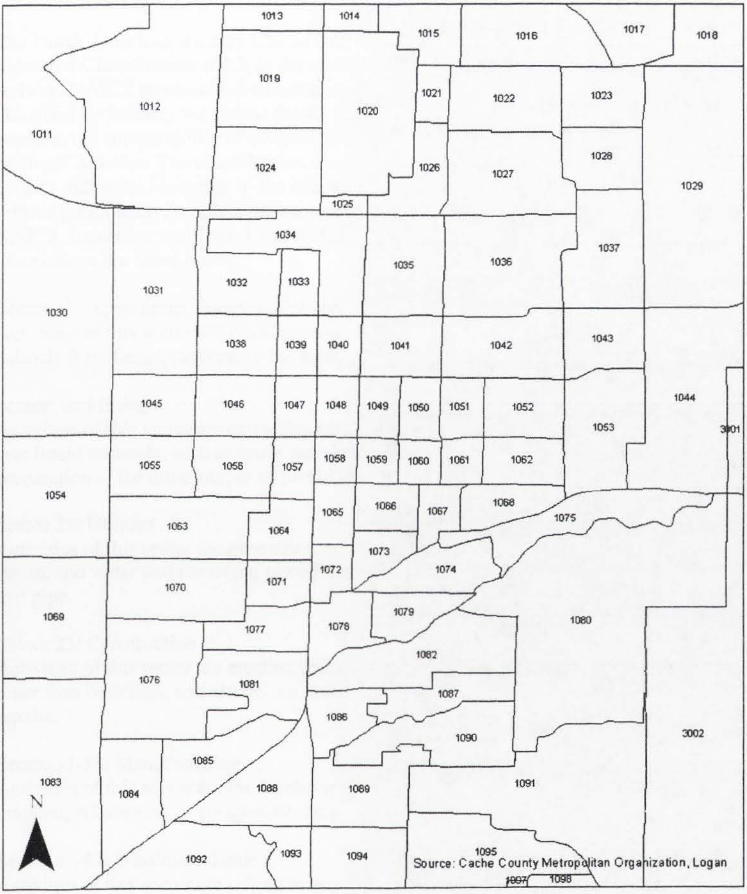
Map of Travel Analysis Zones- Cache county



Map of Travel Analysis Zones- Cache county



Map of Travel Analysis Zones- Cache county



Appendix - 2

NORTH AMERICAN INDUSTRY CLASSIFICATION SYSTEM (NAICS) CODES

The North American Industry Classification System (NAICS) replaced the Standard Industrial Classification (SIC) as the statistical classification standard for economic activity. NAICS standardized all establishment-based Federal economic statistics classified by industry for United States, Canada and Mexico. The NAICS replaced SIC to promote the comparability of establishment data describing various facets of economies of North America. The classification covers the entire field of economic activities and defines industries according to the compensation and structure of the economy. It is revised periodically to reflect the economy's changing industrial organization. Under NAICS, industries are divided into sectors and sub-sectors. Two-digit sector and brief descriptions are listed below.

Sector 11: Agriculture, Forestry, Fishing, and Hunting
Activities of this sector are growing crops, raising animals, harvesting fish, and other animals from farms, ranches, or the animals' natural habitats.

Sector 21: Mining
Activities of this sector are extracting naturally occurring mineral solids, such as coal and ore; liquid minerals, such as crude petroleum; and gases, such as natural gas; and preparation at the mine site, or as part of other mining activity.

Sector 22: Utilities
Activities of this sector are generating, transmitting, and/or distributing electricity, gas steam, and water and removing sewage through permanent infrastructure of lines, mains and pipe.

Sector 23: Construction
Activities of this sector are erecting buildings and other structures; heavy construction other than buildings; and alterations, reconstruction, installation, and maintenance and repairs.

Sector 31-33: Manufacturing
Activities of this sector are the mechanical, physical, or chemical transformation of material, substances, or components into new products.

Sector 41-43: Wholesale Trade
Activities of this sector are selling or arranging for the purchase or sale of goods for resale; capital or durable non-consumer goods; and raw and intermediate materials and supplies used in production, and providing services incidental to the sale of the merchandise.

Sector 44-46: Retail Trade
Activities of this sector are retailing merchandise generally in small quantities to the general public and providing services incidental to the sale of the merchandise.

Sector 48-49: Transportation and Warehousing
Activities of this sector are providing transportation of passengers and cargo, warehousing and storing goods, scenic and sightseeing transportation, and supporting these activities.

Sector 51: Information
Activities of this sector are distributing information and cultural products, providing the means to transmit or distribute these products as data or communication, and processing data.

Sector 52: Finance and Insurance
Activities of this sector involve the creation, liquidation, or change in ownership of financial assets and/or facilitating financial transactions.

Sector 53: Real Estate and Rental and Leasing
Activities of this sector are renting, leasing, or otherwise allowing the use of tangible or intangible assets (except copyrighted works), and proving related products.

Sector 54: Professional, Scientific, and Technical Services
Activities of this sector are providing preofessional, scientific, and technical services for the operations of the oraganizations.

Sector 55: Management of Companies and Enterprises
Activities of this sector are holding of securities of companies and enterprises, for the purpose of owning controlling interest or influencing their management decision, or administrating, overseeing, and managing other establishments of the same company or enterprise and normally undertaking the strategic or organizational planning and decision making of the company or enterprise.

Sector 56: Administrative and Support and Waste Management and Remediation Services
Activities of this sector are performing routine support activities for the day-to-day operations of other organizations.

Sector 61: Educational Services
Activities of this sector are providing instruction and training in a wide variety of subjects.

Sector 62: Health Care and Social Assistance
Activities of this sector are providing health care and social assistance for individuals

Sector 71: Arts, Entertainment, and Recreation
Activities of this sector are providing health care and social assistance to individuals.

Sector 72: Accommodation and Food Services
Activities of this sector are providing customers with lodging and/or preparing meals, snacks, and beverages for immediate consumption.

Sector 81: Other Services (except Public Administration)
Activities of this sector are providing services not elsewhere specified, including repairs, religious activities, grant making, advocacy, laundry, personal care, death care, and other personal services.

Sector 91-93: Public Administration
Activities of this sector are administration, management, and oversight of public programs by Federal, State, and local governments.

Appendix 3

Trip productions and attractions estimated using ITE trip generation rates and by regression analysis in 4- step model.

TAZ	ITE TP	ITE TA	Model TP	Model TA
1	42	158	54	54
2	31	10	48	48
3	31	10	29	29
4	311	222	324	324
5	851	341	647	647
6	13	5	22	22
7	7	2	13	13
8	64	26	60	60
9	252	490	276	276
10	88	91	126	126
11	9	2	11	11
12	3	8	5	5
13	14	18	26	26
14	12	13	17	17
15	179	96	164	164
16	199	81	172	172
17	36	14	36	36
18	11	3	11	11
19	7	27	10	10
20	9	18	13	13
21	8	2	8	8
22	58	21	50	50
23	10	3	10	10
24	36	100	58	58
25	87	67	96	96
26	151	42	135	135
27	72	30	68	68
28	23	16	29	29
29	94	35	89	89
30	12	90	13	13
31	33	273	33	33
32	18	98	18	18
33	73	159	124	124
34	104	197	181	181
35	81	106	85	85
36	165	71	141	141
37	152	66	132	132
38	244	84	192	192
39	123	159	216	216
40	321	518	535	535

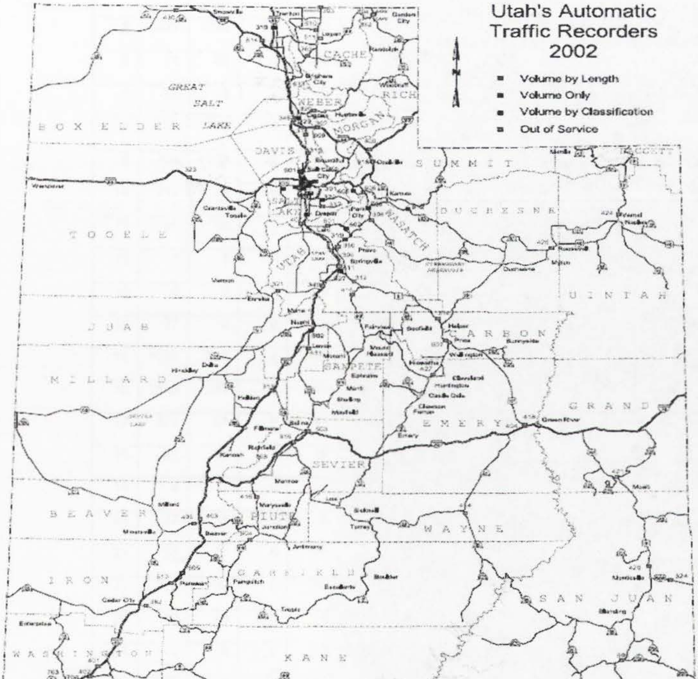
TAZ	ITE TP	ITE TA	Model TP	Model TA
41	305	817	402	402
42	37	19	32	32
43	118	38	100	100
44	301	101	233	233
45	28	211	31	31
46	289	84	226	226
47	305	589	395	395
48	250	328	415	415
49	265	95	193	193
50	108	132	100	100
51	214	101	171	171
52	316	56	250	250
53	369	162	296	296
54	27	109	39	39
55	97	737	108	108
56	306	281	238	238
57	209	184	253	253
58	141	206	217	217
59	119	33	91	91
60	171	337	150	150
61	315	61	237	237
62	63	191	581	581
63	375	345	373	373
64	346	429	320	320
65	352	446	542	542
66	249	62	206	206
67	273	55	224	224
68	305	1487	885	885
69	109	106	113	113
70	256	317	249	249
71	388	668	497	497
72	238	447	291	291
73	287	82	254	254
74	326	72	291	291
75	128	34	112	112
76	367	151	320	320
77	191	298	208	208
78	192	112	171	171
79	188	56	168	168
80	498	154	397	397
81	406	278	392	392
82	233	90	195	195
83	206	61	175	175
84	203	1560	202	202
85	147	339	246	246
86	149	115	137	137

TAZ	ITE TP	ITE TA	Model TP	Model TA
87	7	2	7	7
88	290	217	236	236
89	55	18	48	48
90	116	67	97	97
91	319	195	309	309
92	13	7	11	11
93	12	18	19	19
94	95	55	81	81
95	252	311	212	212
96	210	124	170	170
97	203	58	176	176
98	59	25	54	54
99	124	38	103	103
100	80	56	68	68
101	15	4	14	14
102	23	6	20	20
201	138	46	155	155
202	86	30	100	100
203	80	41	101	101
204	330	344	408	408
205	558	650	613	613
206	155	72	171	171
207	154	133	163	163
208	22	10	24	24
209	2	1	4	4
210	40	21	69	69
211	62	44	83	83
212	1	6	2	2
213	279	90	292	292
214	16	5	22	22
215	104	29	132	132
216	510	272	539	539
217	11	4	12	12
218	55	14	60	60
219	697	1079	655	655
220	440	222	468	468
221	24	16	28	28
222	149	105	160	160
223	16	3	17	17
224	65	22	71	71
225	18	5	20	20
226	15	4	16	16

TAZ	ITE TP	ITE TA	Model TP	Model TA
227	173	77	186	186
228	23	9	26	26
231	19	14	27	27
232	54	14	65	65
233	32	9	36	36
234	0	0		
	21000	21000	22353	22353

Appendix 4

ATR stations in Utah maintained by UDOT.



Appendix -5

TAZ wise Demographic and socio-economic data of Cache County.

TAZ	POP	SFH	MFH	RET	R_EMP	I_EMP	O_EMP	TOT_EMP
1	157	43	1	44	0	227	4	231
2	179	55	0	55	0	0	6	6
3	184	55	0	55	0	0	3	3
4	1418	415	8	423	180	7	237	424
5	5021	1370	92	1462	77	49	193	319
6	74	22	0	22	0	1	3	4
7	52	13	0	13	0	0	0	0
8	371	109	0	109	2	0	28	30
9	799	233	32	265	187	511	106	804
10	219	34	46	80	134	0	81	215
11	55	16	0	16	0	0	0	0
12	3	1	0	1	4	0	18	22
13	0	0	0	0	42	0	0	42
14	27	9	0	9	19	0	11	30
15	1002	284	4	288	37	22	68	127
16	1345	337	10	347	1	10	67	78
17	275	61	0	61	2	0	10	12
18	87	19	0	19	0	0	0	0
19	6	1	0	1	10	33	4	47
20	27	6	0	6	10	0	39	49
21	56	15	0	15	0	0	0	0
22	373	102	0	102	0	0	16	16
23	77	18	0	18	0	0	2	2
24	6	2	0	2	79	70	59	208
25	372	102	5	107	73	25	13	111
26	928	250	21	271	1	0	10	11
27	511	102	21	123	13	0	22	35
28	104	27	0	27	23	0	8	31
29	581	159	1	160	9	0	23	32

TAZ	POP	SFH	MFH	RET	R_EMP	I_EMP	O_EMP	TOT_EMP
30	8	3	0	3	0	119	30	149
31	0	0	0	0	0	361	97	458
32	40	12	0	12	0	138	15	153
33	4	1	0	1	181	45	158	384
34	2	1	0	1	274	54	137	465
35	479	114	0	114	5	0	259	264
36	980	280	6	286	1	1	81	83
37	938	254	9	263	3	5	67	75
38	928	340	89	429	38	1	32	71
39	9	6	0	6	354	0	31	385
40	262	0	76	76	769	16	538	1323
41	599	55	134	189	312	30	1922	2264
42	195	57	8	65	0	0	30	30
43	688	201	3	204	8	0	13	21
44	1606	527	0	527	2	0	63	65
45	1	1	0	1	0	221	191	412
46	1659	232	357	589	7	16	37	60
47	620	78	161	239	421	400	372	1193
48	92	7	30	37	677	4	118	799
49	1316	111	448	559	7	0	137	144
50	551	68	100	168	27	0	308	335
51	1618	191	219	410	10	0	158	168
52	2664	124	575	699	0	0	0	0
53	1870	615	9	624	30	0	167	197
54	11	3	0	3	35	62	155	252
55	0	1	0	1	46	1065	31	1142
56	1471	291	268	559	7	309	35	351
57	671	99	141	240	250	1	162	413
58	118	41	3	44	322	48	99	469
59	491	192	24	216	3	0	3	6
60	652	76	174	250	0	0	930	930
61	2488	102	594	696	8	0	9	17
62	679	1	71	72	0	0	552	552

TAZ	POP	SFH	MFH	RET	R_EMP	I_EMP	O_EMP	TOT_EMP
63	1611	379	209	588	115	113	447	675
64	489	207	340	547	90	89	828	1007
65	435	80	120	200	698	3	393	1094
66	1233	227	279	506	0	0	29	29
67	1557	127	469	596	0	0	16	16
68	597	0	120	120	7	0	4369	4376
69	516	144	0	144	52	5	175	232
70	1093	268	131	399	36	170	415	621
71	808	115	227	342	454	31	1259	1744
72	491	26	231	257	200	31	959	1190
73	1340	205	389	594	4	11	49	64
74	1688	316	345	661	0	3	1	4
75	635	156	91	247	1	0	12	13
76	2281	581	64	645	12	17	129	158
77	673	110	135	245	103	45	589	737
78	830	241	89	330	33	63	43	139
79	945	247	102	349	12	0	22	34
80	3054	870	7	877	7	4	51	62
81	1899	269	410	679	176	88	233	497
82	1219	392	17	409	0	0	89	89
83	1235	354	10	364	7	0	12	19
84	87	26	0	26	32	2216	214	2462
85	0	0	0	0	354	91	388	833
86	747	112	134	246	50	5	194	249
87	50	11	2	13	0	0	1	1
88	1186	383	94	477	46	101	197	344
89	435	95	2	97	0	0	13	13
90	641	176	24	200	0	0	116	116
91	1711	481	26	507	65	13	253	331
92	64	22	0	22	0	0	12	12
93	6	3	0	3	27	0	19	46
94	573	157	1	158	6	13	60	79
95	1303	383	8	391	8	305	140	453

TAZ	POP	SFH	MFH	RET	R_EMP	I_EMP	O_EMP	TOT_EMP
96	1239	355	2	357	0	61	89	150
97	1364	351	11	362	0	0	16	16
98	366	102	0	102	0	0	29	29
99	788	217	0	217	2	0	16	18
100	462	134	0	134	0	51	7	58
101	107	26	0	26	0	0	0	0
102	115	41	0	41	0	0	0	0
201	768	237	5	242	3	0	26	29
202	468	144	3	147	10	1	9	20
203	444	126	0	126	21	0	38	59
204	1575	456	21	477	108	277	134	519
205	2739	827	30	857	56	640	224	920
206	931	264	3	267	0	4	87	91
207	843	253	0	253	2	128	26	156
208	101	38	0	38	0	0	15	15
209	8	4	0	4	0	0	0	0
210	232	68	0	68	1	8	14	23
211	316	96	0	96	16	29	10	55
212	0	0	0	0	0	0	20	20
213	1666	480	15	495	1	9	32	42
214	95	29	0	29	0	0	2	2
215	681	155	38	193	0	0	8	8
216	2810	835	9	844	45	10	357	412
217	59	19	0	19	0	0	4	4
218	371	98	0	98	0	0	0	0
219	3431	984	23	1007	70	1123	459	1652
220	2635	635	44	679	162	0	168	330
221	122	24	10	34	16	0	14	30
222	936	250	0	250	0	80	41	121
223	107	12	22	34	0	0	0	0
224	426	114	0	114	0	0	14	14
225	117	31	2	33	0	0	0	0
226	88	26	0	26	0	0	0	0

TAZ	POP	SFH	MFH	RET	R_EMP	I_EMP	O_EMP	TOT_EMP
227	1014	298	0	298	1	4	87	92
228	127	41	0	41	0	0	7	7
231	2	22	0	22	17	0	11	28
232	4	93	3	96	0	0	0	0
233	18	56	0	56	0	0	0	0
234	0	0	0	0	0	0	0	0

Appendix 6

KS Test: Results

Kolmogorov-Smirnov Comparison of Two Data Sets
(http://www.physics.csbsju.edu/stats/KS-test.html)

The results of a Kolmogorov-Smirnov test performed at 21:43 on 13-JUN-2006

The maximum difference between the cumulative distributions, *D*, is: 0.1818 with a corresponding *P* of: 0.985.

Ho: Null Hypothesis: Both the data sets follow same distribution.

H1: Alternate Hypothesis: The data sets do not follow same distribution.

Reject the null hypothesis of no difference between your datasets if *P* is "small". Since the value of P is high, the null hypothesis cannot be rejected.

Data Set 1 (Visual PFE):

11 data points were entered

Mean = 2.216E+03

95% confidence interval for actual Mean: 511.8 thru 3919.

Standard Deviation = 2.536E+03

High = 6.874E+03 Low = 44.4

Third Quartile = 3.591E+03 First Quartile = 80.6

Median = 1.290E+03

Average Absolute Deviation from Median = 1.859E+03

KS finds the data is consistent with a normal distribution: *P*= 0.53 where the normal distribution has mean= 2639. and sdev= 2921.

KS finds the data is consistent with a log normal distribution: *P*= 0.94 where the log normal distribution has geometric mean= 743.1 and multiplicative sdev= 8.798

Items in Data Set 1:

44.4 70.0 80.6 390. 730. 1.290E+03 1.993E+03 2.592E+03 3.591E+03 6.717E+03 6.874E+03

Data Set 2 (Conventional Model):

11 data points were entered

Mean = 2.087E+03

95% confidence interval for actual Mean: 468.4 thru 3706.

Standard Deviation = 2.410E+03

High = 6.456E+03 Low = 25.2

Third Quartile = 3.630E+03 First Quartile = 47.5

Median = 1.212E+03

Average Absolute Deviation from Median = 1.837E+03

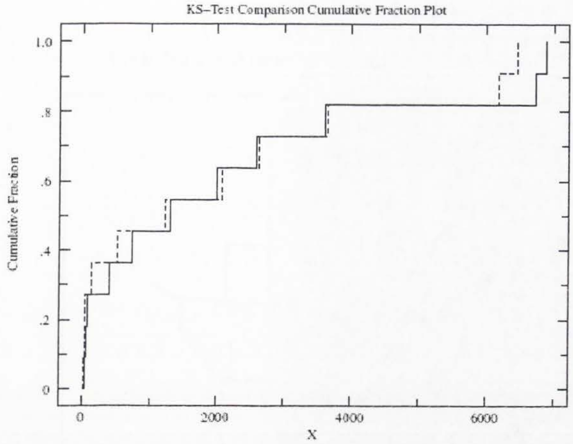
KS finds the data is consistent with a normal distribution: *P*= 0.59 where the normal distribution has mean= 2474. and sdev= 2759.

KS finds the data is consistent with a log normal distribution: *P*= 0.88 where the log normal distribution has geometric mean= 564.2 and multiplicative sdev= 11.39

Items in Data Set 2:

25.2 45.8 47.5 148. 507. 1.212E+03 2.082E+03 2.632E+03 3.630E+03 6.175E+03 6.456E+03

Comparison Cumulative Fraction Plot

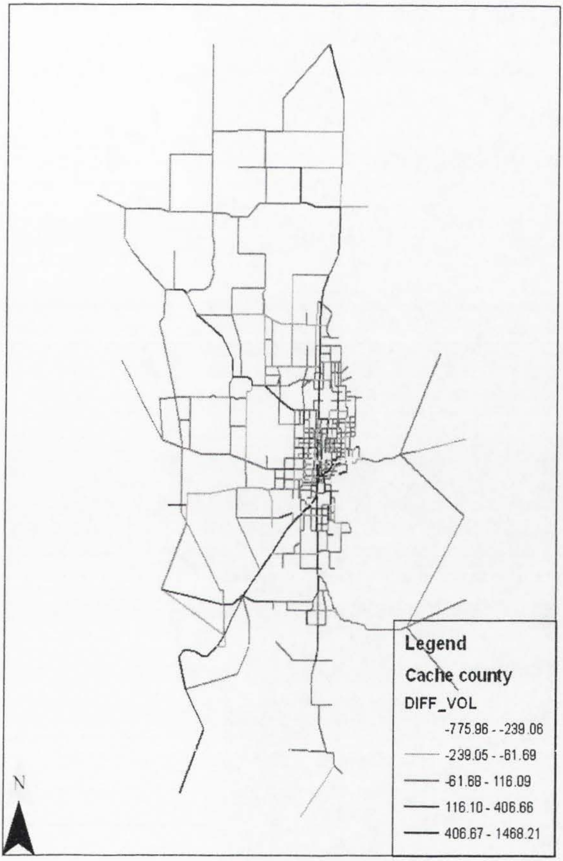


Note:

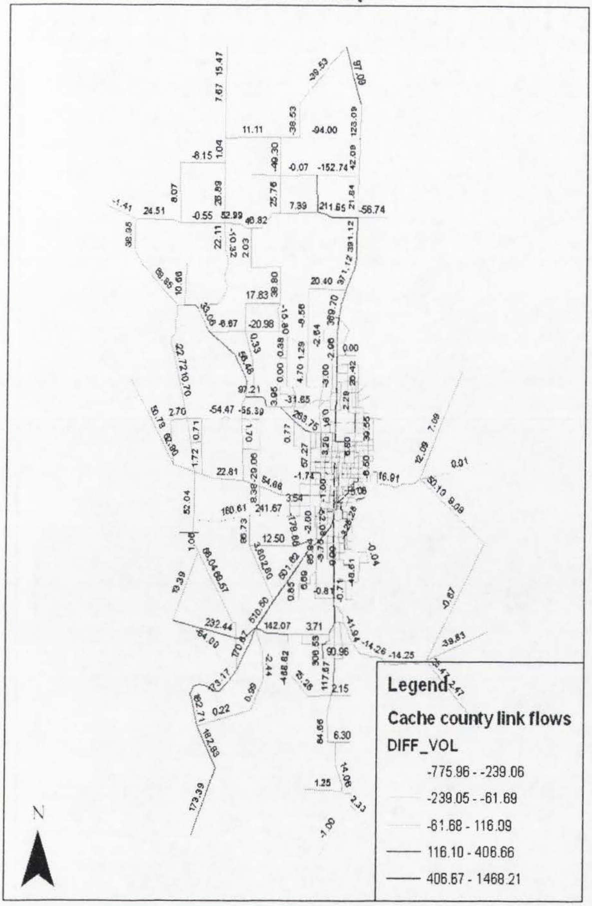
Thick line shows the data of Visual PFE estimation.
Dotted line shows the data of conventional model.

Appendix 7:

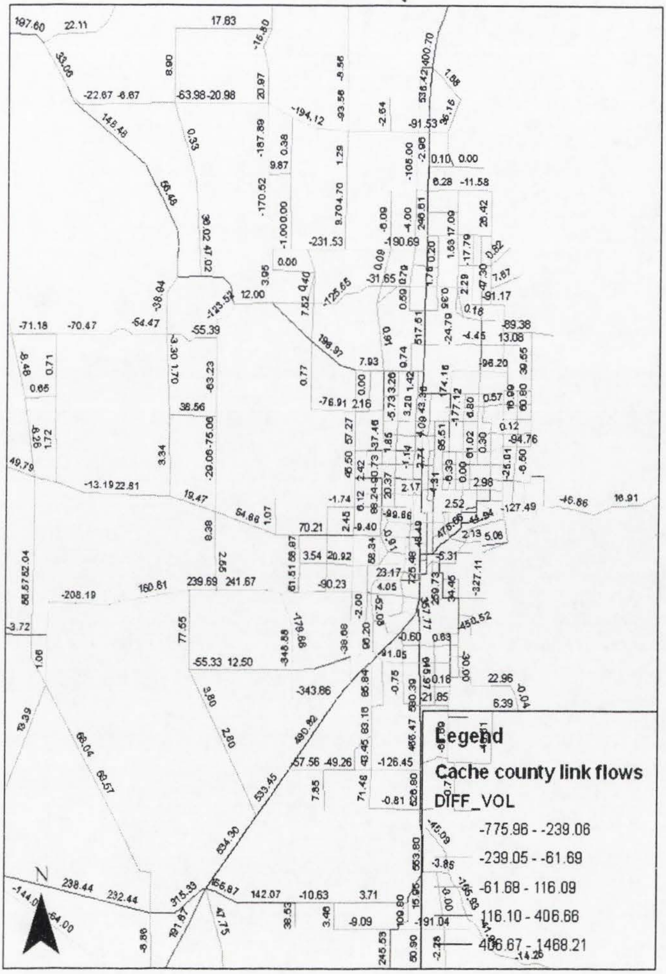
Link flow estimate differences-Cache county



Differences in estimates of Link Flows (Visual PFE-four step model)-Cache County



Differences in estimates of Link Flows (Visual PFE-four step model)-Cache County



Differences in estimates of Link Flows (Visual PFE-four step model)- Cache County