THE PROPAGATION OF UNCERTAINTY THROUGH TRAVEL DEMAND MODELS

By

Yong Zhao

Kara Kockelman

Research Report SWUTC/00/167804-1

Southwest Region University Transportation Center

Center for Transportation Research

The University of Texas at Austin

Austin, Texas 78712

October 2000

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SWUTC/00/167804-1		

4. Title and Subtitle	5. Report Date / 6. Performing Organization Code
The Propagation of Uncertainty Through Travel Demand Models	October 2000
7. Author(s)	8. Performing Organization Report No.
Yong Zhao and Kara Kockelman	Research Report 167804-1
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS) / 11. Contract or Grant No.
Center for Transportation Research University of Texas at Austin 3208 Red River, Suite 200 Austin, Texas 78705-2650	10727
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered / 14. Sponsoring Agency Code
Southwest Region University Transportation Center Texas Transportation Institute Texas A&M University System College Station, Texas 77843-3135	

15. Supplementary Notes

Supported by general revenues from the State of Texas.

16. Abstract

The future operations of transportation systems involve enormous uncertainty – in both input and model parameters. This work investigates the stability of contemporary transport demand model output by quantifying variability in model input, such as zonal socioeconomic data and trip generation rates. It simulates the propagation of their variation through a series of common demand models over a simplified twenty-five-zone network. The results suggest that uncertainty may be compounded over a series of models and highly correlated across output. The propagated uncertainty varies remarkably in link flows, thus, some link flows may be more uncertain than others. The final step model, trip assignment, may reduce prior increased uncertainty through the first three steps, but generally could not lessen the uncertainty lower than the input uncertainty. Mispredictions at early stages (e.g., trip generation) in multi-stage models appear to amplify across later stages; thus, improvements to these models and their estimates are sorely needed.

17. Key Words	18. Distribution Statement
Demand Models, Planning, Sequential Modeling, Trip Generation	No Restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	46	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DISCLAIMER

ABSTRACT

The future operations of transportation systems involve enormous uncertainty – in both input and model parameters. This work investigates the stability of contemporary transport demand model output by quantifying variability in model input, such as zonal socioeconomic data and trip generation rates. It simulates the propagation of their variation through a series of common demand models over a simplified twenty-five-zone network. The results suggest that uncertainty may be compounded over a series of models and highly correlated across output. The propagated uncertainty varies remarkably in link flows, thus, some link flows may be more uncertain than others. The final step model, trip assignment, may reduce prior increased uncertainty through the first three steps, but generally could not lessen the uncertainty lower than the input uncertainty. Mispredictions at early stages (e.g., trip generation) in multi-stage models appear to amplify across later stages; thus, improvements to these models and their estimates are sorely needed.

ACKNOWLEDGMENTS

This publication was developed as part of the University Transportation Centers Program which is funded 50% with general revenue funds from the State of Texas.

EXECUTIVE SUMMARY

The future operations of the transportation system involve enormous uncertainty. Generally, large-scale transport demand models are estimated sequentially, with the results or estimates of one model acting as inputs to subsequent models. In almost all cases, only point estimates are passed forward, rather than estimates of variation and covariation. Such modeling processes limit the final results to point estimates, so comparisons of plans or scenarios based on the results may be incorrect. This work investigates the stability of contemporary transport demand output by simulating four-step model over a 25-zone network. Sensitivity analysis is adopted to target resources to overall uncertainty.

Current travel-demand-modeling practice does not acknowledge the stochastic uncertainty, especially input uncertainty. For example, structured statistical models produce variance and covariance matrices with their point estimates. However, only point estimations of parameter mean values are carried forward through travel demand models. The covariance information is generally lost. Many variables used as input in transport demand models come from other models, whose associated uncertainty is not known or incorporated. If point estimates of these future variables (such as population, housing, and automobile ownership) are to be used in travel demand models, an appreciation of variability in all results requires distributional information on the input..

To assess some forms of uncertainty in model predictions, most transport modeling processes employ "model validation" to test a model's forecast ability. Uncertainty in future forecast due to input and inherent uncertainty, however, changes over time. Thus, there is no guarantee that future predictions will be bounded by the acceptable range of uncertainty. A "before and after" study is another method used to assess a model's predictive accuracy. It also is difficult to draw useful conclusions from an individual study (Aitken and White, 1972).

There is a fair amount of transportation research really focused on modeling uncertainties. Most of them adopted a simulation technique to capture uncertainty patterns. Using simulation techniques, Ashley (1980) studied the probability distribution of various outputs from an interurban highway scheme forecast model. His approach simultaneously accounted for uncertainty from a variety of sources. Correlated input variables were drawn from multivariate probability distributions. However, Ashley's study gave no information about which error source contributes most to the overall uncertainty. Due to cost and computing limitation, stochastic simulation was not widely applied in other modeling studies. Pell (1984) suggested a method to estimate variability in travel demand forecasts based on identifying those sources of input uncertainty and error that make the largest contributions to forecast uncertainty. He proposed two criteria for selecting the most important error sources. One is the sensitivity of forecasts to input

errors, as measured by elasticity. Another is the magnitude of forecast errors, as measured by coefficient of variation. Nevertheless, his study did not employ correlated input variables.

Simulation techniques are suggested as one of the most useful methods in this field because one can simulate uncertainty from a variety of sources simultaneously. Furthermore, simulation methods are capable of setting multivariate distributions with covariation. This study adopts Monte Carlo simulation. In such a simulation, a model is run repeatedly, using different data drawing from their distribution assumptions. In this approach, the study focuses on covariation simulation. Sensitivity analysis is another effective method to study uncertainty. It traces output uncertainty back to input; thus, it can reveal both the linear relationship and the non-linear relationship between input uncertainty and output uncertainty. Two correlation coefficients are calculated in this study. The sample correlation coefficient provides an initial estimate of any linear relationship between the input and output. The rank correlation coefficient more appropriately suggests any strong non-linear relationship.

This work considers the traditional UTPP model paradigm via its four primary components: trip generation, trip distribution, mode choice, and route selection. To simplify the model structure, this study uses a cross-classification model to calculate the home based work trips (HBW) in trip generation. In this study, three types of employment are used: basic employment, retail employment, and service employment. Five types of zones are specified based on the population and employment density. A common gravity model form is used for trip distribution, with a production constraint. A simple exponential function is used as the impedance function for the gravity model This study simplifies the travel mode choice between public transit and drive-alone and uses a binary logit model for mode split. It employs a user equilibrium method in its trip assignment model and incorporates a standard Bureau of Public Roads link capacity function in searching for convergence to an equilibrium state. This study uses the larger BPR parameter values as NCHRP report 365 suggested.

All together, this sequence of four sub-models produces a set of link-flow estimates. These are the model outputs of greatest interest in this work. The four-step model approach is applied into a road network with 25 zones and 818 links, which is separated from the Dallas-Fort Worth highway system. For outside input, this study uses the demographic data associated with the data. For model parameters, this study uses mean values from DFW area travel model description report (NCTCOG 1999). Necessary simplification and modification has been made based on two published formal manuals (NCHRP Report 187 and 365).

The modeling software used here for the first three sub-model steps (i.e., trip generation, trip distribution, and mode choice) is @Risk (Palisade 1998), which loads through Microsoft Excel software. TransCAD (Caliper Co., 1996) is used here for the final, trip assignment sub-model in order to apply its commercialized UE algorithm.

The sequence of four-step sub-models produces a set of link-flow estimates. The study simulates the forecasting approach by repeating running the four-step models for 100 times. Two arcs are chosen as the critical links. Critical link one presents the general pattern of links with congested volumes, while critical link two presents others with non-congested volumes. All the uncertainty in terms of coefficient of variation are set to be 0.30. As evident in the overall uncertainty results, the variability of the selected link flows is sizable. Both coefficients of variation of the two link flows are larger than 0.30, which suggests the final uncertainty may be compounded to be higher than any input or parameter uncertainty. The flow uncertainty appears not to have a strong relation with congestion. However, the average travel time on the link shows a relatively strong relation with congestion. The travel time uncertainty of the congested link, 1.899, is much higher than that of the uncontested link, 0.127. The total vehicle mile traveled (VMT) of the network is a weighted sum of individual link flows. The uncertainty of VMT, in terms of the coefficient of variance, 0.236 is relatively low. The link flows show great correlation between one another. For probabilistic simulations, correlations greater than 0.5 between input and output indicate substantial dependence. Since the total VMT is the weighed sum of all link flow volumes, there is a strong correlation between total VMT and individual link flows. That explains why the uncertainty of total VMT is generally lower than that of most links.

In each model step, there is a finite amount of input and output. Given the distribution assumption of the input and parameters of the model, the simulation yields 100 observations of each output. Although the amount of output of each step is different, the average coefficient of variation can be collected to track the changes in uncertainty through each step. The five percentile and ninety-five percentile of the uncertainty among each step is also collected to indicate the variability of the uncertainty.

As can be seen, the increasing average uncertainty in the first three step models suggests significant uncertainty propagation through those models. Nevertheless, the final step assignment model reduces the previous compounded uncertainty, but it generally cannot lessen the uncertainty to a level lower than the input uncertainty. The expanding 5% and 95% bound suggests that through the four-step model, the variability of final uncertainty extends. Thus, some link flows' uncertainty may be reduced substantially while others may increase considerably, which indicates the possibility for wide swings in system. However, one can improve the UTPP model forecasting if he cannot avoid this problem. One passable solution for that is to provide the information regarding uncertainty in final results. Then the policymakers will be aware of the uncertainty when comparing scenarios.

One can compare the output's sensitivity to parameters in each model step. Not surprising, the parameter which has the strongest correlation with link flows is the trip generation rate. This is partly consistent with Smith and Cleveland's results (1976). Also, the overall output is sensitive to the demographic input. Most zonal demographic inputs contribute substantially to the overall uncertainty in link flows. Given the linear function pattern of the trip generation model, it is not surprised that the demographic inputs and the trip generation parameters show strong linear correlation with the overall outputs. The pattern is more obvious for the link flows nearby, which is consistent with Mackinder and Evans' study (1981). However, the overall uncertainty evinces a relatively small sensitivity to other three models (trip distribution, mode split, and trip assignment).

Future work on this and related topic is needed. For example, applications on more realistic networks may be examined. A variety of common model types can be tested. Moerover, one may compare different uncertainty level in input resulting overall uncertainty. Modeler then will be able to know which part of the modeling approach needs more work and how intense the effort should be. Additionally, feedback from travel-time estimates to destination, mode, and route choices is needed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THE PROPAGATION OF UNCERTAINTY THROUGH TRAVEL DEMAND MODELS

CHAPTER 1. INTRODUCTION

The future operations of the transportation system involve enormous uncertainty. Modeling these complicated systems requires many variables and behavioral components whose variability may be poorly identified or simply ignored. Without explicit and rigorous statistical recognition of uncertainty in transportation demand forecasts, transportation planning of towns, cities, and metropolitan areas represents a veritable gamble. Transportation plans and polices based on these forecasts may be inaccurate and even misleading. Thus, a lot of money may be wasted in transport facility investment.

Generally, large-scale transport demand models are estimated sequentially, with the results or estimates of one model acting as inputs to subsequent models. In almost all cases, only point estimates are passed forward, rather than estimates of variation and covariation. Such modeling processes limit the final results to point estimates, so comparisons of plans or scenarios based on the results may be incorrect. First, the estimates for different plans or scenarios may overlap with one another. Moreover, the difference between the comparisons may not be statistically significant.

This work investigates the nature of uncertainty propagation in contemporary transport demand models. The objective of this study is to quantify variability in model output and to track the sources of the variability back into input. Monte Carlo simulation and sensitivity analysis are used to investigate the propagation of this variation over a simplified network.

The rest of this paper is organized as follows. The next section describes the study background and literature review. Section 3 specifies the model structure and assumptions. Section 4 discusses the simulation results and sensitivity analysis. The final section provides a summary of the research findings and identifies possible extensions of the current research.

CHAPTER 2. BACKGROUND

First, the technique background and review related literature on the study of uncertainty in travel demand modeling are discussed. There are many sources that produce forecast errors. Modelers can do relatively little to the errors due to measurement, sampling, computation, model specification, and aggregation. (Barton-Aschman,1997). In contrast, purely stochastic errors can be accommodated statistically. Components of these stochastic errors arise from three sources, which here are termed “inherent uncertainty”, “input uncertainty”, and “propagated uncertainty”. Since travel demand model parameters are random variables, estimated from samples of the population, model estimates are associated with variations and covariations. This outcome constitutes inherent uncertainty. Also, the use of predictions of future demographic data (e.g., employment and land use) as input to traffic demand forecasting models contribute input uncertainty. Moreover, since transport demand models are generally estimated and applied sequentially, the results or estimates of one model act as input to subsequent models. Their uncertainty is passed forward, producing propagated uncertainty. The cumulative impact of these three forms of uncertainty is the focus of this research.

Unfortunately, current travel-demand-modeling practice does not acknowledge all these sources of uncertainty, especially input uncertainty. For example, structured statistical models produce variance and covariance matrices with their point estimates. However, only point estimations of parameter mean values are carried forward through travel demand models. The covariance information is generally lost. Many variables used as input in transport demand models come from other models, whose associated uncertainty is not known or incorporated. If point estimates of these future variables (such as population, housing, and automobile ownership) are to be used in travel demand models, an appreciation of variability in all results requires distributional information on the input. As an illustration of this, Smith and Shahidulla (1995) have suggested that the predictive value of census tract projections is quite limited over a ten-year period.

Modeling methods based on point estimates dramatically constrain all final results into point estimates, and the point estimates may be highly biased. For example, the expected value of a linear function of independent variables requires only mean values of the input variables. However, non-linear functions and any functions involving correlated variables require distributional information in order to avoid bias when estimating the function’s mean value (See, e.g., Rice, 1995). Comparisons of alternative transportation plans or scenarios based on these do not convey information regarding uncertainty in estimates – or the statistical significance of differences. Neglect of data and parameter uncertainties and their correlation ultimately mean that transportation planning, policy-making, and infrastructure decisions are much more of a gamble than they need to be.

To assess some forms of uncertainty in model predictions, most transport modeling processes employ "model validation" to test a model's forecast ability. Although validation compares model predictions to observations using the data that are not used in model estimation, this step can only assess the model's predictive strength for contemporary behavior. Uncertainty in future forecast due to input and inherent uncertainty, however, changes over time. Thus, there is no guarantee that future predictions will be bounded by the acceptable range of uncertainty.

Barton-Aschman *et al.* (1997) have provided a set of specific guidelines for model validation, but they also have recognized that input error (and inherent uncertainty) propagates to overall uncertainty. There is the concern that each step in the Urban Transportation Planning System (UTPS) models could possibly increase the overall error. They write: "while there is a potential for the errors to offset each other, there is no guarantee that they will." (pp. 12) No attempt is made to quantify the propagated uncertainty.

A "before and after" study is another method used to assess a model's predictive accuracy. It also is difficult to draw useful conclusions from an individual study (Aitken and White, 1972). Such examples can be seen in the following studies: Horowitz and Emsile,1978, ITE,1980, and Mackinder and Evans,1981. Additionally, without sensitivity analysis, one does not know which resource contributes the main part of the uncertainty. Mackinder and Evans (1981) have suggested that the errors in socioeconomic variables might dominate highway volume forecast errors. Moreover, percent root mean square error (%RMSE) is being estimated to validate traffic assignment models by comparing predicted and observed flow volume. Practical results (e.g. NCHRP Report No. 365) suggest that average hourly or daily flow forecasts come with %RMSE of 30 to 50 percent, and links with low flows tends to have higher %RMSE than those with high flows.

There is a fair amount of transportation research really focused on modeling uncertainties. Most of them adopted a simulation technique to capture uncertainty patterns. For example, Robbins (1978) estimated the possible error in each of the four step models. However, several of his assumptions were naïve. Bonsall (1977) proposed a more systematic approach with sensitivity analysis, but no particular distribution was specified.

There are some more sophisticated approaches. Using simulation techniques, Ashley (1980) studied the probability distribution of various outputs from an interurban highway scheme forecast model. His approach simultaneously accounted for uncertainty from a variety of sources. Correlated input variables were drawn from multivariate probability distributions. However, Ashley's study gave no information about which error source contributes most to the overall uncertainty. Due to cost and computing limitation, stochastic simulation was not widely applied in other modeling studies. Pell (1984) suggested a method to estimate variability in travel demand forecasts based on identifying those sources of input uncertainty and error that make the largest contributions to forecast uncertainty. He proposed two criteria for selecting the most important error sources. One is the sensitivity of forecasts to input errors, as measured by elasticity. Another is the magnitude of forecast errors, as measured by coefficient of variation. He recommended fewer simulation runs after identifying the influence of a small number of uncertain sources. Nevertheless, his study did not employ correlated input variables.

There are many other less relevant studies in uncertainty analysis. For example, Rose's network study (1986) focused on flow predictions and did not account for correlated input. Leurent (1997) developed a sensitivity and uncertainty analysis method for the equilibrium solution of a dual criteria model on a small-scale network.

In summary, many researchers have seriously studied the propagation of uncertainty through travel demand models. Simulation techniques are suggested as one of the most useful methods in this field because one can simulate uncertainty from a variety of sources simultaneously. Furthermore, simulation methods are capable of setting multivariate distributions with covariation. Sensitivity analysis is another effective method to study uncertainty. It traces output uncertainty back to input; thus, it can reveal both the linear relationship and the non-linear relationship between input uncertainty and output uncertainty. Due to cost and computational limitation, prior studies have several common weaknesses. First, few large-scale data applications have been accomplished in the past. Another limitation is that no firm conclusion has been approached.

CHAPTER 3. MODEL APPLICATION

This study adopts the effective methods suggested by prior research. It investigates the stability of contemporary transport demand model's output. The models studied here are the traditional, four-step urban transportation planning process (UTPP) models. There are two main methods, Monte Carlo simulation and sensitivity analysis to be used here. Hahn and Shapiro (1967) illustrated the general approach using Monte Carlo simulation. In such a simulation, a model is run repeatedly, using different data drawing from their distribution assumptions. In this approach, the study focuses on covariation simulation. Another technique used here is sensitivity analysis (Cullen and Frey, 1999). Two correlation coefficients are calculated in this study. The sample correlation coefficient provides an initial estimate of any linear relationship between the input and output. The rank correlation coefficient more appropriately suggests any strong non-linear relationship.

This work considers the traditional UTPP model paradigm via its four primary components: trip generation, trip distribution, mode choice, and route selection. The following is a discussion of model specification.

3.1 Trip Generation

Trip generation models have two basic structures: (1) regression equations at an aggregate (zonal) or disaggregate (household/person) level, and (2) cross-classification of trip rates at an aggregate level. Kassoff and Deutschman (1970) suggested that disaggregate models represent the true correlation between and variation within variables and produce better results in comparison with aggregate models. However, the stability of trip rate model parameters may not be strong across different data sets. Kannel and Heathington's (1973) study results indicated temporal stability of trip rates, but tests by Smith and Cleveland (1976) rejected this stability in trip rates from Detroit survey data. To simplify the model structure, this study uses the following simplified cross-classification models to calculate the home based work trips (HBW).

Trip Production:

$$T_i = \alpha HH_i \tag{1}$$

where T_i is the number of HBW trips produced in zone i,

HH_i is the total number of households in zone i,

and α is the trip producation rate.

Trip Attraction:

$$A_i = \sum_{k,l} \beta_{kj} EMP_{ik} x_{il} \tag{2}$$

where A_i is the number of HBW trips attracted in zone i,

EMP_{ik} is the total number of employment type k in zone i,

x_{il} is the dummy varaiable which equals 1if zone i is zone type l, 0 otherwise ,

and β_{kj} is the trip attraction rate of employment type k in zone type l.

In this study, three types of employment are used: basic employment, retail employment, and service employment. Five types of zones are specified based on the population and employment density.

3.2 Trip Distribution

In general, trip distribution models used in practice come from one of two basic structures: the growth factor (Fratar) model and the gravity model. The most common model form used for trip distribution is the gravity model, and this is the model used here, with a production constraint. This model form, subject to a production constraint, is defined as follows:

$$T_{ij} = T_i \left(\frac{A_j F(t_{ij})}{\sum_i A_k F(t_{ik})} \right) \tag{3}$$

where T_{ij} is the number of trips from zone *i* to zone *j*,

T_i is the number of trip productions in zone *i*,

A_j is the number of trip attractions in zone *j*,

t_{ij} is the impedance (time or generalized cost) from *i* to *j*,

and $F(t_{ij})$ is the impedance function recognizing travel cost between zones *i* and *j*.

The impedance function should be inversely related to zonal separation. Gamma, power, or exponential functions usually are used. Here a simple exponential functions is used, as follows:

$$F(t_{ij}) = t_{ij}^{\gamma} \tag{4}$$

where γ is the impedance parameter.

Equation 3 yields a trip matrix consistent with the number of productions in each zone but not with the number of attractions. Thus, this form of the gravity model is “singly constrained”. This study applies three iterations switching between the attraction constrained calculation and the production constrained calculation to balance the trip matrix.

Murchland (1978) has suggested, via extensive calculation, that for small errors in both trips generated and impedance matrix values, the relative variance (i.e., the coefficient of variation squared) of the resultant cell values is approximately the sum of the relative variances of the input.

3.3 Mode Split

Multinomial and nested logit models are very common models of mode choice. A multinomial logit (MNL) specification essentially assumes equal competition across alternatives. Using this model the proportion of trips made by mode *m* between zones *i* and *j* is the following:

$$\Pr_{m|ij} = \frac{e^{V_{m|ij}}}{\sum_{l} e^{V_{l|ij}}} \qquad (5)$$

where $V_{m|ij}$ is the utility of mode *m* given origin *i* and destination *j.* $V_{m|ij}$ is specified to be a linear function of trip time, cost, and other variables. Here a simple linear function is used:

$$V_m = \theta_m + TT_m \delta + \varepsilon_m \qquad (6)$$

where TT_m is total travel time by mode *m*,

ε_m represents unobserved heterogeneity(assumed to be *iid* GEV),

and θ_m, δ are parameters.

So the total number of trips by mode *m* from zone *i* and *j* are:

$$T_{ijm} = T_{ij} \Pr_{m|ij} \qquad (7)$$

This study simplifies the travel mode choice between public transit and drive-alone.

3.4 Route Choice

Network assignment of trip can include several common features. For example, an all-or-nothing method assigns all traffic flows between an O-D pair to the shortest path. Capacity-restrained assignments attempt to approximate an equilibrium solution by iterating between all-or-nothing traffic loading and recalculating link travel times based on link capacity functions. User equilibrium (UE) methods utilize an iterative process to achieve a convergent solution ("equilibrium") in which no traveler can improve his/her travel time by shifting routes.

The uncertainty in assignment models appears to be small if equilibrium techniques are used. Leurent (1997) suggested that an equilibrium network assignment method is very stable, given well-defined criteria and constraints. Indeed, in congested networks the equilibration process may reduce the magnitude of uncertainties from the interchange models, in the reproduction of link flows.

This study employs a user equilibrium method in its trip assignment model. UE algorithms incorporate link capacity functions in their search for convergence to an equilibrium state. This function describes the relationship between link flow and link impedance. A common functional form developed by the Bureau of Public Roads is:

$$t = t_f \left[1 + \alpha_0 \left(\frac{q}{q_{\max}} \right)^{\beta_0} \right] \tag{8}$$

where t is impedance of a given link at flow q,

t_f is free flow impedance of the link,

$q_{\max}$ is link capacity,

and α_0, β_0 are volume/delay coefficients.

The standard BPR values for α_0 and β_0 are 0.15 and 4.0. However, *NCHRP report* 365 suggests larger values. The larger values, 0.84 and 5.5, are used here.

All together, this sequence of four sub-models produces a set of link-flow estimates. These are the model outputs of greatest interest in this work. However, any variability in the output is due solely to uncertainties in input and parameters. These input and parameter uncertainties are simulated by first specifying their distributions and then generating values randomly from these distributions. To impose sign constraints on many of these variables (for example, trip generation rate cannot be less than zero), lognormal distributions are used. To accommodate covariation across input and parameter values, multivariate distributions were specified, including the multivariate lognormal distribution.

The four-step model approach is applied into a road network (see Figure 1 in Appendix) with 25 zones and 818 links, which is separated from the Dallas-Fort Worth highway system. For outside input, this study uses the demographic data associated with the data. For model parameters, this study uses mean values from DFW area travel model description report (NCTCOG 1999). Necessary simplification and modification has been made based on two published formal manuals (NCHRP Report 187 and 365). However, there are several arbitrary variation and covariation assumptions such as the equal coefficient of variance for all input and parameters. More reliable estimates of variation and covariation are likely to require model estimation using actual travel data since estimates of variation and covariation are rarely reported in the literature.

The modeling software used here for the first three sub-model steps (i.e., trip generation, trip distribution, and mode choice) is @Risk (Palisade 1998), which loads through Microsoft Excel software. This is a very flexible and user-friendly software for Monte Carlo simulation and risk analysis; however, many standard programming languages and other software packages are viable for such techniques. TransCAD(Caliper Co., 1996) is used here for the final, trip assignment sub-model in order to apply its commercialized UE algorithm.

The results of greatest interest are mean values of link flows and their matrices of covariation, across model simulations. These are discussed in the following section.

CHAPTER 4. SIMULATION RESULTS

The sequence of four-step sub-models produces a set of link-flow estimates. The study simulates the forecasting approach by repeating running the four-step models for 100 times. Final link flows are obtained from the converged UE assignment results. Most of the ratios of volume versus capacity are relatively low (85% of them are less than 0.76), which indicates the assignment equilibrium is not under a congestion situation. In fact the result is a portion of a general assignment, it only includes morning peak hour house-based work auto trip assignment.

The flow volumes from one assignment are shown in Figure 2. Two arcs are chosen as the critical links. Critical link one (Rochelle Blvd. between Northgate and Rochelle) presents the general pattern of links with congested volumes, while critical link two (SH183 eastbound passed Story Road ramp) presents others with non-congested volumes. The flow distribution of 100 simulation results for these two links are shown in Figure 3. Not surprisingly, given the normal and lognormal distribution assumptions of input and parameters, the result distributions are approximately normal.

The input and parameters with uncertainty of all four-step models used in simulation are shown in Table 1. All the uncertainty in terms of coefficient of variation are set to be 0.30. As can be shown, the inverse of the coefficient of variation is equal to the T-statistic value of a parameter. The 0.30 uncertainty value of a parameter suggests that the corresponding T-statistic value is larger than 3, which indicates the parameter is significant at 0.05 level in the estimation.

The overall uncertainty results can be shown in terms of coefficient of variation in Table 2. As evident in these results, the variability of the selected link flows is sizable. Both coefficients of variation of the two link flows are larger than 0.30, which suggests the final uncertainty may be compounded to be higher than any input or parameter uncertainty. The flow uncertainty appears not to have a strong relation with congestion. A more elaborated illustration is shown in Figure 3, which represents the uncertainty of all loaded links respected to their volume/capacity ratios. As can be seen, most link flow uncertainties are larger than 0.30, not matter what their v/c ratio values are. Some points in the left lower area provide a possibility that under an extremely low flow/capacity level, the overall uncertainty may be reduced to some degree. However, the average travel time on the link shows a relatively strong relation with congestion. The travel time uncertainty of the congested link, 1.899, is much higher than that of the uncontested link, 0.127.

The total vehicle mile traveled (VMT) of the network is a weighted sum of individual link flows. The uncertainty of VMT, in terms of the coefficient of variance, 0.236 is relatively low. 14. Same as the uncertainty in the total vehicle hour traveled (VHT) of network. However, the VHT of a congested network is less reliable; thus, it is seldom used in assessing transportation plans. As can be seen in Table 3, the link flows show great correlation between one another. For probabilistic simulations, correlations greater than 0.5 between input and output indicate substantial dependence. Since the total VMT is the weighed sum of all link flow volumes, there is a strong correlation between total VMT and individual link flows. That explains why the uncertainty of total VMT is generally lower than that of most links.

Overall, the uncertainty propagation process through the four-step travel demand forecast model is shown in Figure 5. In each model step, there is a finite amount of input and output. Given the distribution assumption of the input and parameters of the model, the simulation yields 100 observations of each output. Although the amount of output of each step is different, the average coefficient of variation can be collected to track the changes in uncertainty through each step. The five percentile and ninety-five percentile of the uncertainty among each step is also shown in Figure 5 to indicate the variability of the uncertainty. Even though all the input uncertainties are set to be the same value, 0.30, the actual simulation data drawn from certain distributions may contain uncertainties slightly different from this value. Thus, the 5% and 95% of demographic input uncertainty are 0.2592 and 0.3397 respectively.

As can be seen, the increasing average uncertainty in the first three step models suggests significant uncertainty propagation through those models. Nevertheless, the final step assignment model reduces the previous compounded uncertainty, but it generally cannot lessen the uncertainty to a level lower than the input uncertainty. The expanding 5% and 95% bound suggests that through the four-step model, the variability of final uncertainty extends. Thus, some link flows' uncertainty may be reduced substantially while others may increase considerably, which indicates the possibility for wide swings in system. However, one can improve the UTPP model forecasting if he cannot avoid this problem. One passable solution for that is to provide the information regarding uncertainty in final results. Then the policymakers will be aware of the uncertainty when comparing scenarios.

The simulation results suggests the trip assignment equilibrium technique may reduce the overall uncertainty, which is partially consistent with Leurent's (1997) study. Leurent suggests that in congested networks the equilibration process may reduce the magnitude of uncertainties in the reproduction of link flows. One possible explanation is the capacity constraint restricts the variability of link flows. However, in this study, the assignment is not under congested situation. One may suggests that link flows are somehow the sum of related O-D trip pairs. With strong correlation among those O-D trips, the combined uncertainty in link flow may be less than the uncertainty coming from the prior model results.

In order to promote improved understanding and interpretation of the four-step model approach, sensitivity analysis is useful for identifying model input that are key contributors to uncertainty in model output. The results of sensitivity analysis can be shown in Table 4. The "*" marks indicate the significant correlations at 0.05 level in the table. Since there are a lot of demographic input variables (household, employment of each zone), only the sums of those variable across zones are presented. One can compare the output's sensitivity to parameters in each model step. Not surprising, the parameter which has the strongest correlation with link flows is the trip generation rate. This is partly consistent with Smith and Cleveland's results (1976). Also, the overall output is sensitive to the demographic input. Most zonal demographic inputs contribute substantially to the overall uncertainty in link flows. Given the linear function pattern of the trip generation model, it is not surprised that the demographic inputs and the trip generation parameters show strong linear correlation with the overall outputs. The pattern is more obvious for the link flows nearby, which is consistent with Mackinder and Evans' study (1981). However, the overall uncertainty evinces a relatively small sensitivity to other three models (trip distribution, mode split, and trip assignment), which may be contributed by the complex calculation of those models, e.g., the trip balance iteration in trip distribution and the equilibrium technique used in trip assignment.

CHARTER 5. CONCLUSIONS

This work investigates the stability of contemporary transport demand output. By simulating four-step model over a 25-zone network. Sensitivity analysis is adopted to target resources to overall uncertainty.

The results of this work suggest that uncertainty is compounded over travel demand model and highly correlated across output. The propagated uncertainty varies remarkably in link flows, thus, some link flows may be more uncertain than others. The final step model, trip assignment, may reduce prior increased uncertainty through the first three steps, but generally could not lessen the uncertainty lower than the input uncertainty. Mispredictions at early stage (e.g., trip generation) of multi-stage models appear to be amplified across later stages. Flows on various links show high correlation. That suggests they are resulted from a equilibrium assignment. Overall, these results indicate that predictions from many travel demand models may be highly uncertain and improvements to the models and their estimates are sorely needed. Transportation modelers should be able to recognize, calculate, and show uncertainty. On the other hand, policy make should be able to know uncertainty and make policies based on uncertainty.

Future work on this and related topic is needed. For example, applications on more realistic networks may be examined. A variety of common model types can be tested. Moreover, one may compare different uncertainty level in input resulting overall uncertainty. Modeler then will be able to know which part of the modeling approach needs more work and how intense the effort should be. Additionally, feedback from travel-time estimates to destination, mode, and route choices is needed.

REFERENCE

1. Aitken, J.M. and White, R. (1972) " A Comparison between a Traffic Forecast and Reality." *Traffic Engineering an Control*. August, pp. 174-177.
2. Ashley, D.J. (1980) "Uncertainty in the Contest of Highway Appraisal." Transportation 9(3), pp. 249-267.
3. Barton-Aschman Associates, Inc., and Cambridge Systematics, Inc. (1997) Model Validation and Reasonableness Checking Manual. Federal Highway Administration Report, Washington, D.C.
4. Ben-Akiva, M. and Lerman, S. (1985) Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge, Massachusetts: MIT Press.
5. Bonsall, P. W., Champerowne, A.F., Mason, A.C., and Wilson, A.G. (1977) "Transport Modeling: Sensitivity Analysis and Policy Testing." Progress in Planning. 7(3).
6. Criscenti, L. J., Laniak, G. F. , and Erikson, R. L. (1996) "Propagation of Uncertainty Though Geochemical Code Calculations." Geochimica et Cosmochimica Acta, 60 (19), pp. 3551-3568.
7. Caliper Co. (1998) "TransCAD User's Guide." Windows Version 3.0. Newton, MA: Caliper Corporation.
8. Cullen, A. C. and Frey, H. C. (1998) Probabilistic Techniques In Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Model and Inputs. Plenum Press, New York.
9. Domencich, T.A., and McFadden, D.(1975). Urban Travel Demand: A Behavioral Analysis. Amsterdam, North-Holland.
10. Hahn, G. L. and Shapiro, S. S. (1967) Statistical Models in Engineering, Wiley Classics Library, John Wiley and Sons, New York.
11. Hakanson, L. (1998) "Error Propagation in Step-by-Step Predictions: Examples for Environmental Management Using Regression Models for Lake Ecosystems." Environmental Modelling and Software, 14(1), pp. 49-58.
12. Horowitz, J. and Emslie, R. (1978) "Comparison of Measured and Forcast Traffic Volumes on Urban Interstate Highways." Transportation Research. 12(1), pp. 29-32.
13. ITE (1980) "Evaluation of the Accuracy of Past Urban Transportation Forecasts." Institute of Transportation Engineers Journal. 50(2), pp. 24-34.

14. Laosirihongthong, T. (1997) "Updating Judgments of Uncertainty in Travel Demand Forecasts with Model Output: Empirical Results and Analytical Implication." Ph.D. Dissertation, The Ohio State University.
15. Liu, H. (1993) "Tests of Predictive Accuracy of Experimentally-Based Choice Models." Ph.D. Dissertation, The University of Iowa.
16. Leurent, F. (1997) "Sensitivity and Error Analysis of the Dual Criteria Traffic Assignment Model." Transportation Research (B), 32(3), pp. 189-204.
17. Leurent, F. (1996) "An Analysis of Modeling Error, With Application to a Traffic Assignment Model with Continuously Distributed Values of Time." 24th European Transportation Forum, Proceedings of Seminar, D1, PTRC, London.
18. Mackinder, I.H. and Evans, R. (1981) "The Predictive Accuracy of British Transport Studies in Urban Areas." SR 699, Transport and Road Research Laboratory. Crowthorne, Berkshire.
19. Martin, W. (1998) Travel Estimation Techniques for Urban Planning. NCHRP Report 365, Transportation Research Board, National Research Council, Washington, D.C..
20. MTC (1998) Travel Forecasting Assumptions '98 Summary. [Online]. Metropolitan Transportation Commission, Oakland, California. http://www.mtc.ca.gov/datamart/forecast/assume98.htm. (Accessed: October 20, 1999).
21. NCTCOG (1999) Dallas-Fort Worth Regional Travel Model (DFWRTM): Description Of The Multimodal Forecasting Process. Transportation Department, North Central Texas Council of Governments, Texas.
22. Ortúzar, J. de D and Willumsen, L.G. (1995) Modelling Transport, Second Edition. John Wiley & Sons, London.
23. Palisade (1998) "@Risk: Simulation Add-In For Miscrosoft Excel." Windows Version Release 4.0. Newfield, NY: Palisade Corp. (November)
24. Pell, C. M. (1984) "The Analysis of Uncertainty in Urban Transportation Planning Forecasts." Ph.D. Dissertation, Cornell University.
25. Rice, J. (1995) Mathematical Statistics and Data Analysis, Second Edition. Belmont, California: Duxbury Press.
26. Robbins, J. (1978) "Mathematical Modelling – the Error of Our Ways." Traffic Engineering and Control. January, pp. 32-35.
27. Rose, G. (1986) "An Analysis of Error Propagation in Transportation Network Equilibrium Models." Ph.D. Dissertation, Northwestern University.

28. Smith, S. K. and Shahidullah, M. (1995) "An Evaluation of Population Projection Errors for Census Tracts." Journal of the American Statistical Association, 90(429), pp12-25.
29. Sosslau, A., et al (1978) Quick Response Urban Travel Estimation Techniques and Transferable Parameters: User Guide. NCHRP Report 187, Transportation Research Board, National Research Council, Washington, D.C.
30. Tobin, R. L. and Friesz, T. L. (1988) "Sensitivity Analysis for Equilibrium Network Flow." Transportation Science, Vol. 22, No. 4, pp. 245-250.
31. Wilson, R. and Crouch, E. A. (1981) "Regulation of Carcinogens." Risk Analysis, 1(1), pp. 47-57.

APPENDIX

TABLE 1. SIMULATION SET-UP FOR THE 25-ZONE NETWORK

Forecasting Input

The mean values of input (household and different types of employment) come from the data set associated with the DFW area travel model. The coefficients of variation are all set to 0.30. The S.D.s are then determined by multiplying the mean values by corresponding coefficient of variation. The distribution is assumed to be multivariate normal with a same correlation coefficient (0.30) among each other.

Model Parameters[*]

Model	Parameter	Mean	S.D.	Coefficient of Variation	Distribution	Covariance
Trip Generation	α	2.303	0.691	0.30	Lognormal	-
	$\beta_{1,2}$	1.389	0.417	0.30	Lognormal	-
	$\beta_{1,3}$	1.328	0.398	0.30	Lognormal	-
	$\beta_{1,4}$	1.309	0.393	0.30	Lognormal	-
	$\beta_{1,5}$	1.476	0.443	0.30	Lognormal	-
	$\beta_{2,2}$	1.396	0.419	0.30	Lognormal	-
	$\beta_{2,3}$	1.530	0.459	0.30	Lognormal	-
	$\beta_{2,4}$	1.448	0.434	0.30	Lognormal	-
	$\beta_{2,5}$	1.386	0.416	0.30	Lognormal	-
	$\beta_{3,2}$	1.304	0.391	0.30	Lognormal	-
	$\beta_{3,3}$	1.371	0.411	0.30	Lognormal	-
	$\beta_{3,4}$	1.369	0.411	0.30	Lognormal	-
	$\beta_{3,5}$	1.392	0.418	0.30	Lognormal	-
Trip Distribution	γ	1.16E-3	3.48E-4	0.30	Lognormal	-
Model Split	$\theta_{transit}$	-0.549[**]	0.165	0.30	MVLognormal[*]	ρ=0.67
	δ	-0.0297	0.0089	0.30	MVLognormal[*]	
Traffic Assignment	α_0	0.84	0.252	0.30	Lognormal	-
	β_0	5.50	1.65	0.30	Lognormal	-

* The mean values of parameters are from DFW area travel model report. (NCTCOG 1999).
**To be negative, the parameter is drawn from a lognormal distribution with |mean| and SD, and add a minus sign.

TABLE 2. NETWORK FLOW SIMULATION RESULTS*

Variable	Description	Mean	S.D.	Coefficient of Variation	Avg. V/C Ratio
f_1	Main direction flow on link 1	1172	363	0.310	1.116
f_2	Main direction flow on link 2	1522	489	0.322	0.235
T_1	Average travel time on link 1 (hour)	0.1058	0.201	1.899	-
T2	Average travel time on link 2 (hour)	0.0137	0.0017	0.127	-
Total VMT	Total vehicle-miles traveled on the network	129518	30579	0.236	-
Total VHT	Total vehicle-hours traveled on the network	3347	777	0.232	-

* All the results are based on converged UE assignments for 100 runs. The total demand (morning peak hour HBW auto trips) distributes with a mean of 23856 and S.D. of 5503.

TABLE 3. CORRELATION COEFFICIENTS BETWEEN LINK FLOWS

	f_1	f_2	*Total VMT*	*Total VHT*
f_1	1.000	0.601	0.849	0.862
f_2	0.601	1.000	0.724	0.725
Total VMT	0.849	0.724	1.000	0.983
Total VHT	0.862	0.725	0.983	1.000

TABLE 4. SAMPLE CORRELATION BETWEEN INPUT AND OUTPUT

Model	Parameter	f_1	f_2	*Total VMT*	*Total VHT*
Trip Generation	A	0.0589	0.1280	0.1024	0.0990
	$\beta_{1,2}$	0.0345	0.0133	-0.0399	-0.0283
	$\beta_{1,3}$	0.2150*	0.3182*	0.3396*	0.3204*
	$\beta_{1,4}$	-0.0274	-0.0594	-0.0262	-0.0269
	$\beta_{1,5}$	0.0467	0.0343	-0.0008	0.0035
	$\beta_{2,2}$	0.0869	-0.0248	0.0549	0.0562
	$\beta_{2,3}$	-0.1094	0.0394	-0.0086	-0.0004
	$\beta_{2,4}$	0.0091	-0.0123	-0.0023	-0.0076
	$\beta_{2,5}$	0.1270	0.2089	0.1500	0.1483
	$\beta_{3,2}$	0.1013	0.1582	0.0326	0.0488
	$\beta_{3,3}$	0.6052*	0.3646*	0.5944*	0.5987*
	$\beta_{3,4}$	-0.0356	-0.0226	-0.0636	-0.0555
	$\beta_{3,5}$	-0.1701	-0.1753	-0.1259	-0.1297
Trip Distribution	Γ	0.0244	0.0099	0.0084	0.0049
Model Split	$\theta_{transit}$	0.0711	0.1558	0.1121	0.1075
		0.0457	0.1651	0.1327	0.1271
Traffic Assignment	α_0	-0.0431	-0.0427	-0.0793	-0.0628
	β_0	-0.0409	0.0305	0.0223	0.0080
Input	*Total Household*	0.4419*	0.3354*	0.4719*	0.4791*
	Total Basic Employment	0.4511*	0.3230*	0.5639*	0.5706*
	Total Retail Employment	0.5212*	0.3244*	0.5347*	0.5427*
	Total Service Employment	0.6055*	0.3872*	0.6427*	0.6517*

TABLE 5. RANK CORRELATION BETWEEN INPUT AND OUTPUT

Model	Parameter	f_1	f_2	*Total VMT*	*Total VHT*
Trip Generation	α	0.0698	0.0959	0.1558	0.1596
	$\beta_{1,2}$	0.0191	0.0220	-0.0433	-0.0291
	$\beta_{1,3}$	0.1471*	0.2296*	0.3019*	0.2827*
	$\beta_{1,4}$	0.0594	-0.0509	0.0585	0.0602
	$\beta_{1,5}$	0.0713	0.0387	0.0211	0.0248
	$\beta_{2,2}$	0.1254	-0.0109	0.0930	0.1001
	$\beta_{2,3}$	-0.1326	-0.0495	-0.0485	-0.0474
	$\beta_{2,4}$	-0.0254	-0.0053	0.0178	0.0050
	$\beta_{2,5}$	0.1982	0.2266*	0.1897	0.1909
	$\beta_{3,2}$	0.0291	0.1156	0.0031	0.0155
	$\beta_{3,3}$	0.5879*	0.3360*	0.5517*	0.5531*
	$\beta_{3,4}$	-0.0836	-0.0899	-0.1048	-0.1050
	$\beta_{3,5}$	-0.1582	-0.1437	-0.1548	-0.1625
Trip Distribution	γ	0.0057	-0.0184	-0.0327	-0.0399
Model Split	$\theta_{transit}$	0.0963	0.1187	0.1227	0.1139
	δ	0.0815	0.1530	0.1303	0.1282
Traffic Assignment	α_0	-0.0068	-0.0534	-0.0469	-0.0308
	β_0	-0.0430	0.0641	0.0045	-0.0053
Input	*Total Household*	0.4408*	0.3548*	0.4679*	0.4727*
	Total Basic Employment	0.4276*	0.3172*	0.5327*	0.5391*
	Total Retail Employment	0.4950*	0.3334*	0.4924*	0.5010*
	Total Service Employment	0.5680*	0.3867*	0.6093*	0.6141*

Note: * indicates the correlation is significant at 0.05 level(2-tailed).

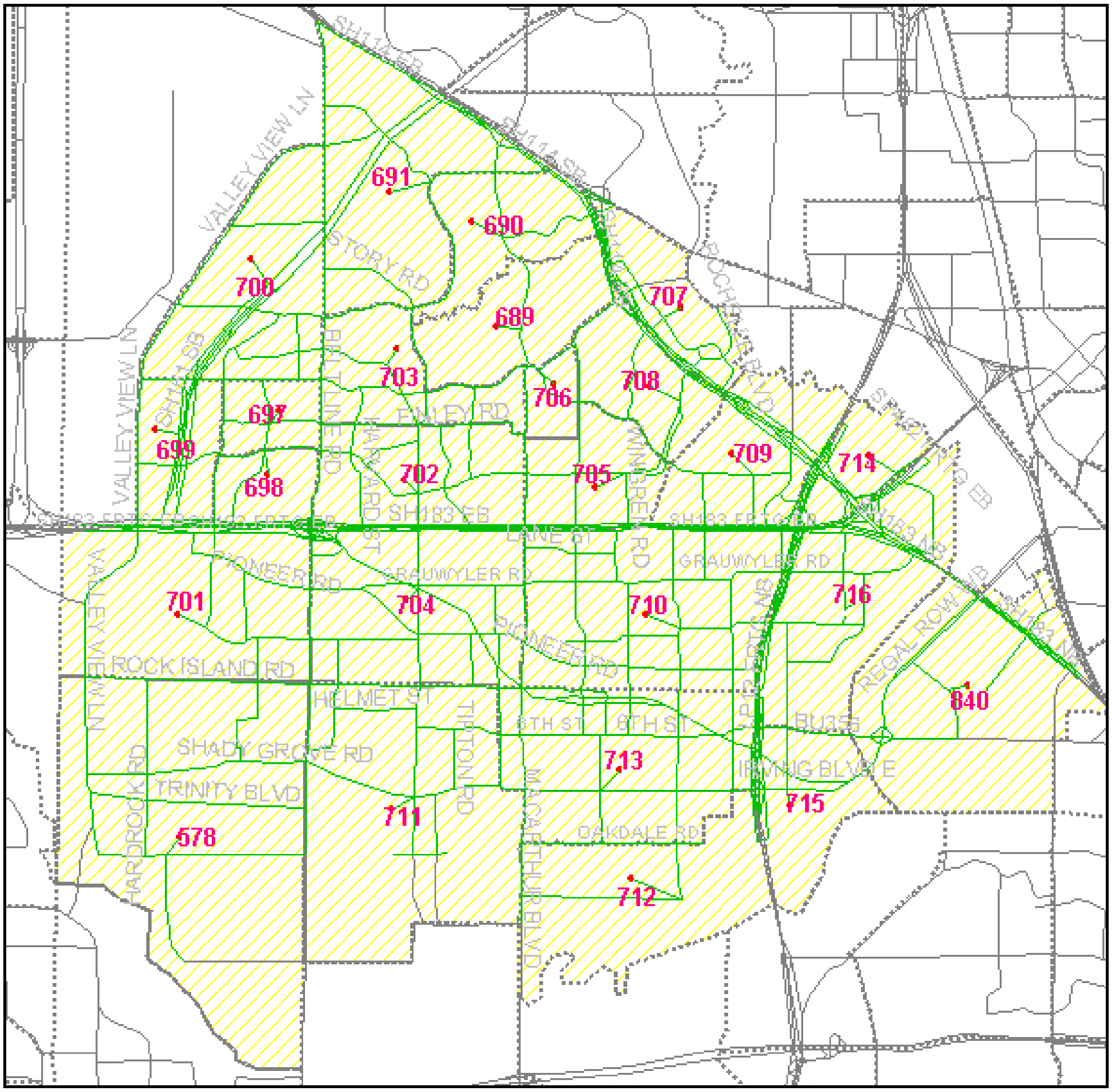


Figure 1. The 25 zones subnet from DFW highway network

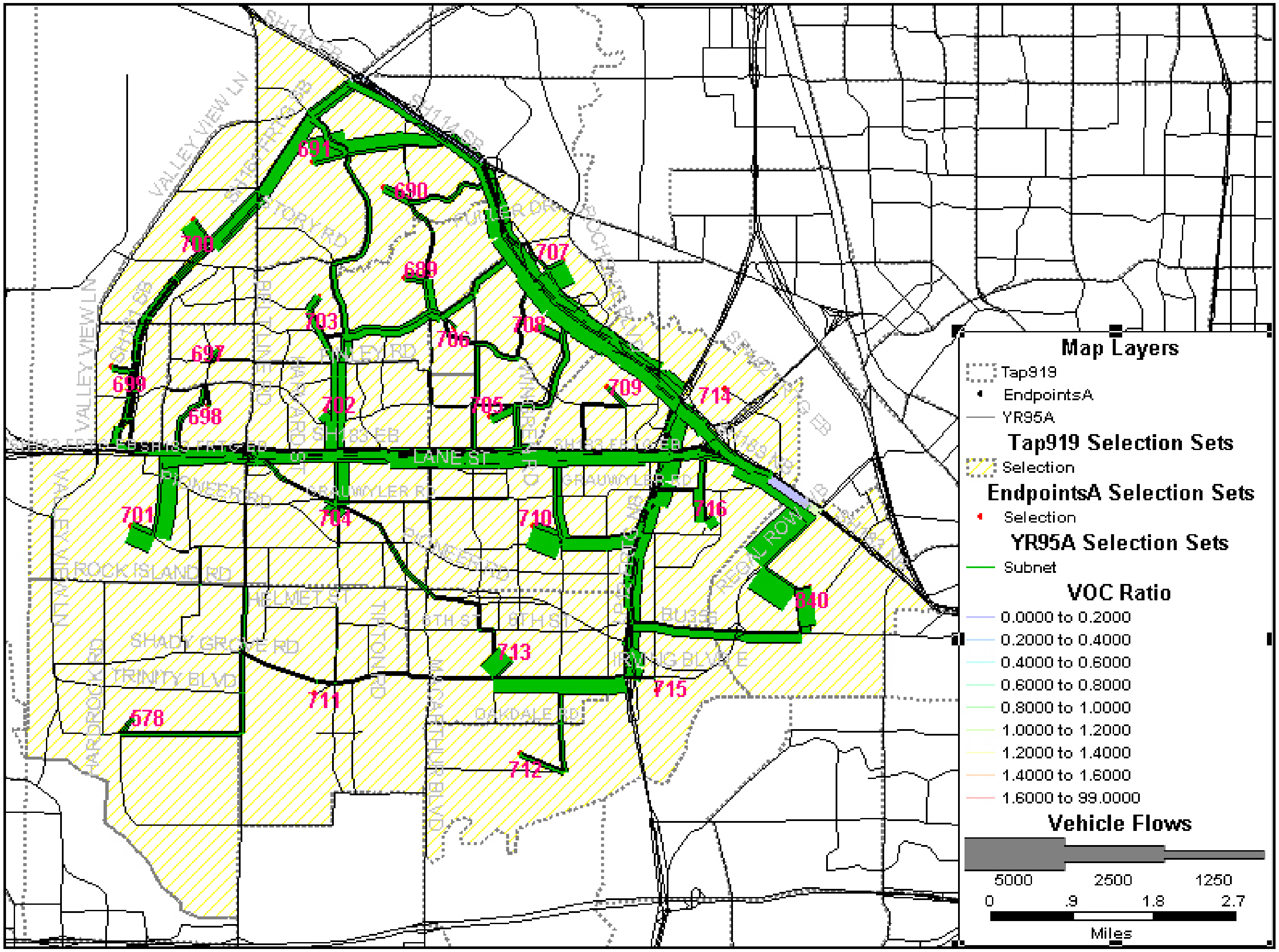
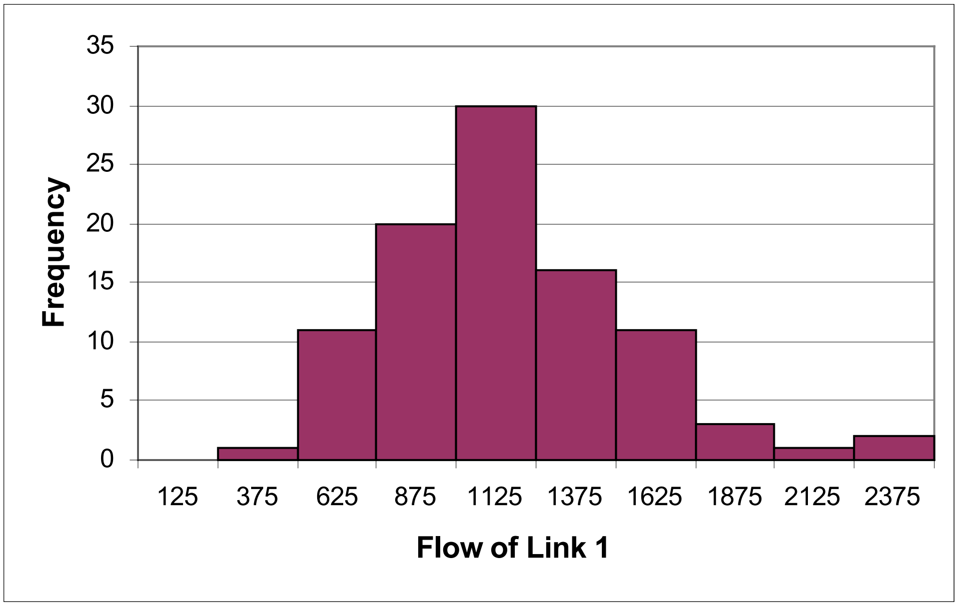


Figure 2. One UE assignment result of the 25 zones subnet

Link 1 flow distribution



Link 2 flow distribution

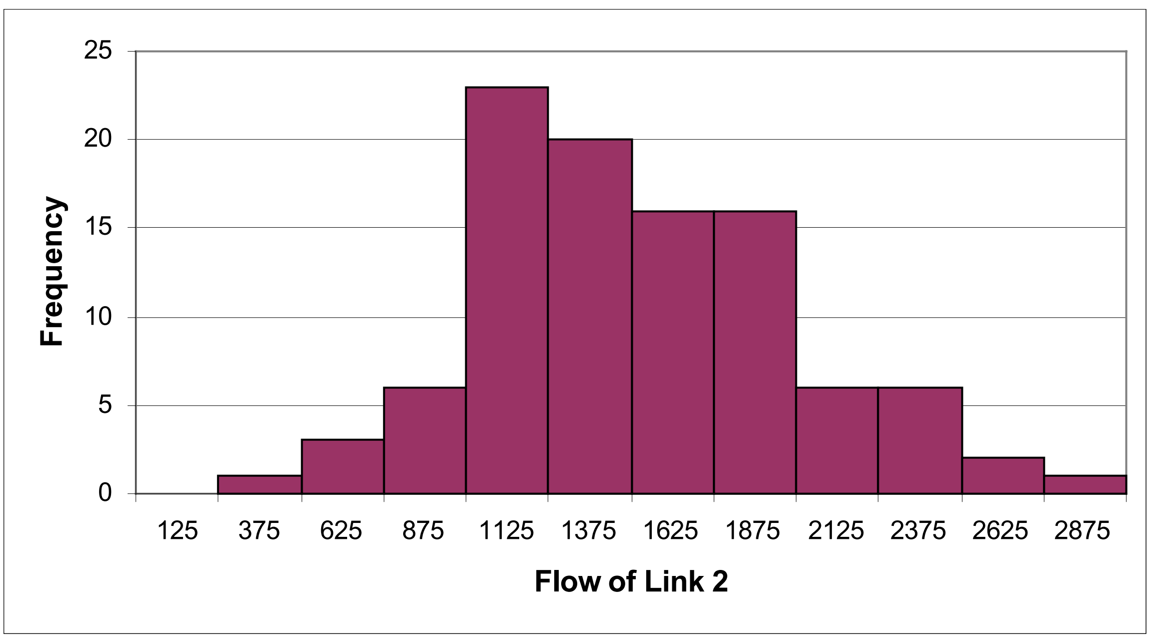


Figure 3. The distribution of 100 assignment results of selected links

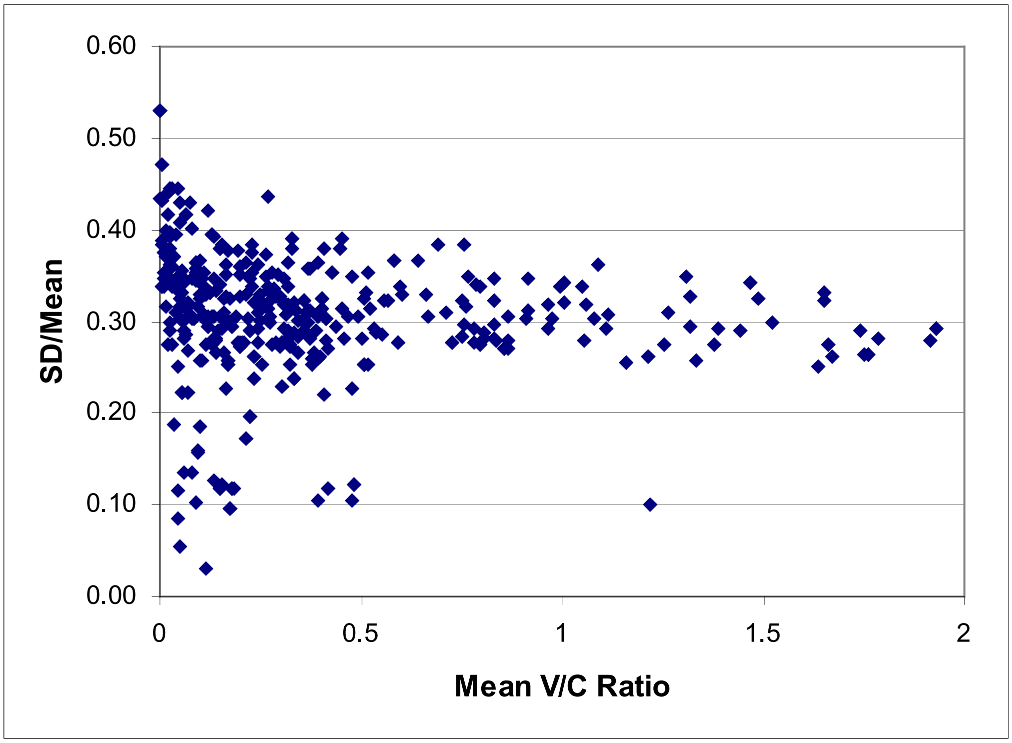


Figure 4. The scatter plot of uncertainty and volume/capacity ratio

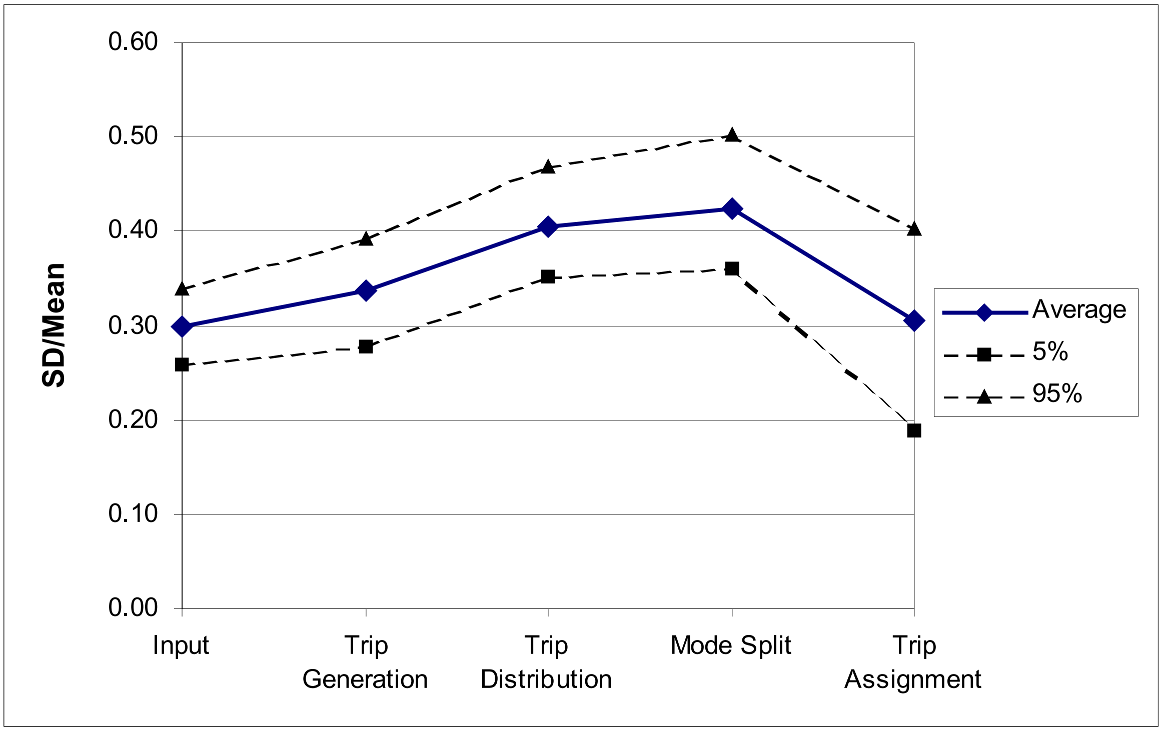


Figure 5. Uncertainty propagation through 4-step models