Predicting the Effectiveness of Commute Reduction Plans Using Neural Networks

by

Monica R. Rush

SUBMITTED TO THE DEPARTMENT OF MECHANICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

BACHELOR OF SCIENCE IN MECHANICAL ENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 6th, 2005 [June 2005]

Author____
Monica Rush
Department of Mechanical Engineering
May 6th, 2005

Certified by____
David Wallace
Associate Professor of Mechanical Engineering
Thesis Supervisor

Accepted by____
Ernesto Cravalho
Mechanical Engineering Undergraduate Officer

Predicting the Effectiveness of Commute Reduction Plans Using Neural Networks

By
Monica R. Rush

Submitted to the
Department of Mechanical Engineering
May 6th, 2005

In Partial Fulfillment of the Requirements for the Degree of
Bachelor of Science in Mechanical Engineering

ABSTRACT

Commute trip reduction plans are being implemented at an increasing number of worksites. In order to be able to structure the most effective plan for a specific worksite, it is necessary to understand the factors that determine commuter response to company incentives to change commuting habits. This study compares the predictive ability of neural networks compared to linear regression models in calculating the change in vehicle trip rate (VTR) at a given worksite over a year long travel plan period. Using a Los Angeles area dataset ($n = 3439$), linear regression and neural network models were constructed and optimized using input variables including worksite incentives, business type and number of years of incentives at the worksite. Significant differences(p=*0.006, 0.007*) in program effectiveness were discovered between the results of local authority worksites ($\overline{\Delta VTR} = -1.495$) and both businesses ($\overline{\Delta VTR} = -0.986$) and hospitals ($\overline{\Delta VTR} = -0.728$). It was determined that the neural network ($R^2 = 0.229$) performed better than the linear regression models ($R^2 = 0.038$) when evaluated by R^2, representing an improvement of *0.014* on previous models. However, when measuring the ability of the models to predict within a certain interval around the output the linear regression models outperformed the neural network model by a factor of 35 percentage points. The lack of strong linear correlations between the inputs and the outputs of these models suggests that the most significant factors in creating successful transportation demand management programs are not currently being tracked. Given the statistically significant superior performance at local authority worksites it is suggested that more worksite demographics are tracked.

Thesis Supervisor: David Wallace
Title: Associate Professor of Mechanical Engineering

ACKNOWLEDGEMENTS

There are many individuals who have given assistance in one form or another over the course of the past several months. I would like to extend a heartfelt thanks to Professor David Wallace and Sittha Sukkasi as well as all of the members of CADlab for their support and guidance during the research process. In the provision of data and offering insight Phil Winters at the Center for Urban Transportation Research and Terry McCall at the South Coast Air Quality Management District were invaluable. Finally, I would like to thank my parents for their continuous encouragement and understanding.

Table of Contents

1.0 Introduction

The global dependence on the diminishing supply of easily accessible oil and gas resources and the resulting pollution and climate change effects have serious implications for the earth's ecosystems and our well-being. Some governments have begun to react to the growing concern over these issues with initiatives on both the international and local levels, such as the Kyoto Protocol and the Cambridge Climate Protection Plan [1]. Research is being conducted to ascertain the best methods of implementing renewable energy systems, as well as to assess the impact of policy on greenhouse gas emissions.

Such protocols contain several strategies for reducing emissions. Their goals are to attack greenhouse gas emissions on many levels, from reducing industrial pollution to improving facilities for pedestrians and cyclists. They often include plans of attack that would not only reduce emissions through regulation but also efforts to change the mentality of emission regulation, encouraging citizens to feel a personal responsibility for emissions reduction.

Many localities provide incentives to reduce emissions on a smaller scale through programs with local employers. These policies target employers as agents of change, mandating the creation of programs that encourage employees to switch modes of transportation to work, which in turn reduces the number of single-occupancy vehicles on the road. Transportation is important as it accounts for one third of the United States' greenhouse gas emissions [2]. In the transportation sector cities such as Los Angeles and Tucson as well as the state of Washington have implemented policies that rely on change that begins at the company level [3, 4].

It is important to understand how these incentives impact commuter behavior, both in the sense of traffic control as well as emissions reduction. Researchers have been working on building models to predict how transportation modes change as a result of incentives such as bicycle facility improvements, subsidized transit passes or discounted parking for ride-sharers. Work done by the Center for Urban Transportation Research at the University of Southern Florida established a vehicle trip reduction model for the public use distributed on their website [5].

This paper builds on the model established by the Center for Urban Transportation Research by adding new classifications to the study programs. The resulting model will further the understanding of the factors that shape transportation mode choice. An artificial neural network will be trained using an existing database with further classifications coded by the researchers. It is hoped that due to the nonlinearity of the relationships in transportation mode choice that these new classifications will add more explanatory power to the model.

2.0 Background & Theory

2.1 Prior Work

Research has been performed in several areas of relevance to commuter trip plans. Aside from the work done by the Center for Urban Transportation Research [6], investigators in Europe have also been working on the determination of major factors in transportation demand management. A series of papers presented at the 2001 European Conference of Ministers of Transport gives a general background of the options available to companies in transportation demand management and lays down the groundwork for more technical research in the area [7].

A review of the literature on company incentives revealed that much of the research in this area has focused on reducing traffic instead of emissions reduction [8, 9]. This is not surprising -- traffic reduction has a more immediate economic and societal impact than emissions reduction. Public awareness of traffic congestion is much higher than awareness of greenhouse gas emissions. Papers focused on emissions reduction also seemed to be only preliminary papers in the area – case studies of specific programs, or general summaries of trends seen in company travel plans and results [6, 10, 11]. These papers show that there are a variety of different methods being employed to encourage commuters to change their modes of transportation, as tabulated in Table 1.

Table 1: Summary of Incentives Used in Transportation Plans

CATEGORY	INCENTIVE	DESCRIPTION
Bicycle Use	Facility Improvements	Addition of bike racks, lockers, showers
	Discounted Cycles	Subsidized bicycle purchases
Public Transportation	Facility Improvements	Increase proximity to public transportation stops, improvements to comfort level of bus stops
	Subsidies	Monetary subsidies for public transportation passes
Parking Related	Decrease in Parking Available	Employer reduces number of parking spots

Parking Related (continued)	Cash-Out Programs	Monetary Incentives to not bring a car to work
	Increase in Price of Parking	Employer raises cost of parking
Marketing	Public Information Campaign	Includes flyers, ads, commuter fairs, etc.
	Focus Groups	for ride sharers
	Company recognition	For employees who use alternative methods of commuting to work
Carpooling Incentives	Guaranteed Ride Home	Program that secures a ride home (via taxi, public transportation or other) for carpoolers/transit users should their regular transportation be unavailable
	Preferential Parking	Parking closer in location or better sheltered for employees who ride-share
	Flextime	Ability for ride sharers to choose their own hours
	Ride share matching systems	Programs that connect employees to each other for ride sharing purposes

Parking cash out incentives have been explored in more depth than the other incentives in table 1. Shoup shows in *The High Cost of Free Parking* that free parking and employer-paid parking discourages carpooling, and that monetary compensation for use of alternative methods for traveling to work reduce vehicle travel to work by 12% on average [13].

In theory, the changes in commuting habits over the course of a program would be directly related to the various promotions related to commuting. The Center for Urban Transportation Research has built a model that examines the impact of the combination of these incentives using a neural network. Regression analyses were performed in parallel using the same data set, however, it was found that neural network was more accurate than the linear regression models.

2.2 Statistical Methods

In order to help determine which variables were of interest, and to understand the relationships between the variables and the outputs statistical methods were employed. Linear correlations were used to determine if there were any one-to-one linear relationships within the data. The linear correlation coefficient for two variables X and Y is determined by Equation 1 [14]

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sigma_x \sigma_y} \qquad \text{Equation (1)}$$

where

$$-1 \le \rho_{x,y} \le 1 \qquad \text{Equation (2)}$$

and

$$Cov(X,Y) = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu_x)(y_j - \mu_y)\,. \qquad \text{Equation (3)}$$

The larger the absolute value of $\rho_{x,y}$, the stronger the linear relationship between the two sets of data X and Y.

Student t-tests were also conducted to determine the importance of categorizations in the data sets. T tests are used to determine the statistical significance of the difference between the means of two groups. For a comparison of samples A and B the t statistic is calculated using Equation 4 [14]

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\sigma_A^2}{n_A} - \frac{\sigma_B^2}{n_B}}} \qquad \text{Equation (4)}$$

where n_A and n_B are the number of records in samples A and B respectively. This value is then compared with values in a table of significance to determine if the difference of means is large enough to have not been a coincidental finding. After this comparison is made it is possible to determine from the table the likelihood of the difference between the two samples to have simply been a chance finding. In most social research, this likelihood is deemed significant if it is equal or less than 5% [14].

Multiple regression analyses generate linear functions to describe the relationships between multiple inputs and a singular output. Using a dataset, they create an equation of the form

$$Y = \beta_0 + \beta_1 x_x + \beta_2 x_2 + \ldots + \beta_n x_n\,, \qquad \text{Equation (5)}$$

where β_n are the coefficients of each input, and x_n represents each of the input variables. For multiple regression analyses, the R^2 statistic is indicative of the amount of variability of the output that can be explained by the inputs. The R^2 statistic is calculated using Equation 6.

$$R^2 = \sum_{i=1}^{n}\left(\frac{|r|^2 - r_i^2}{n - p - 1}\right) \qquad \text{Equation (6)}$$

where r_i is the residual for the ith data point, n is the sample size and p is the number of parameters in the model.

2.3 Neural Networks

Neural Networks are capable of learning the relationships between a set of inputs and a specified output when the dependence is unknown. They consist of several computational units, "neurons", which process their inputs in order to produce the desired output. The neurons are arranged in layers, with each layer's inputs determined by the previous layer's outputs. In this manner the inputs of the first layer are the inputs are the inputs of the network, and the outputs of the last layer are the outputs of the network. The structure of a neural network is represented in Figure 1. Each neuron is denoted by $\boxed{\Sigma}$, where f^n represents the fitting function of each layer, and a^n represents the outputs of each layer [15].

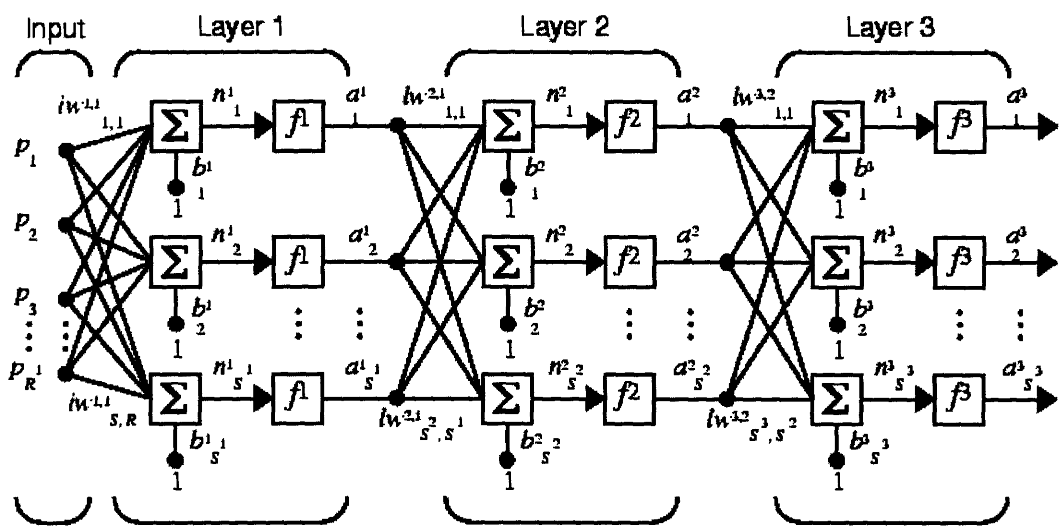

Figure 1: Representative Structure of a Multi-layer Neural Network[15]

The network represents the incorporation of the each layer's outputs through weighting each output appropriately. Improper weighting placed on any of the outputs will result in a less accurate model.

Networks are first trained using a previously collected set of data. This is in order to determine the proper weighting of each variable. A subset of the training data is set aside to be used as a testing set. This ensures that the model does not "over-fit" the training data which would result in an extremely accurate fit for the training data, but less accurate results when the model is applied to testing data. The training process is represented in Figure 2.

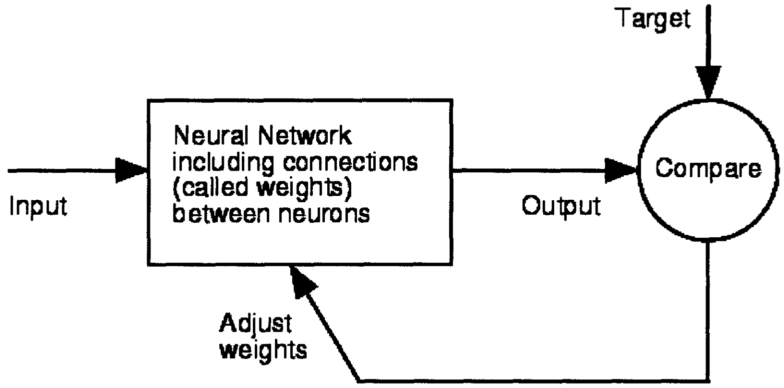

Figure 2: Block Diagram Depiction of Neural Network Training Process [15]

Before the model building can begin, the data must be formatted appropriately. This means that the builder must accurately identify the relevant inputs and outputs. This is typically done through the statistical methods outlined in the section above, linear correlations, t tests and regression analyses. After the inputs and outputs are chosen, the design of the neural network must be determined.

Design of neural networks is not the focus of this research, however it is still necessary to outline a few of the important characteristics of neural network design. The design of the neural network is shaped by the number of layers, number of neurons per layer, the transfer function used for each layer, as well as several other parameters. These parameters are all evaluated by the designer of the neural network, with a final design chosen for its accuracy to the training and validation data. Accuracy is determined through a mean squared error analysis, governed by the Equation 6 [14]

$$Mean\ \ Squared\ \ Error = \frac{\sum (x_{predicted} - x_{actual})}{n_{data}} \qquad \text{Equation (6)}$$

where $x_{predicted}$ and x_{actual} are the outputs predicted by the neural network and the outputs in the data sets respectively and n_{data} is the number of data points in the data set.

3.0 Methodology

There were three main stages of the development of the artificial neural network in this study. The first step is acquiring the data and formatting the variables in the most descriptive manner. The second step is running statistical analyses on the data set to see what relationships exist between the variables. The third step is to use the dataset to train and optimize a neural network.

3.1 Data Acquisition & Formatting

Initially, data were collected through literature review. Relevant sources identified included several papers from the United Kingdom in response to the UK Department for Transports endorsement of worksite and school travel plans [18] as well as several case studies available from Washington State Department of Transport [19]. However, difficulties with finding a sufficient number of data points, as well as differences in measured outcome variables prevented a large cohesive data set from being constructed. The dataset available from the Center for Urban Transportation Research at the University of South Florida was both sizeable and relevant, and thus was chosen for use in this study.

The data used in this study were provided by CUTR in the form of a Microsoft Excel spreadsheet. It has been used already by CUTR to create models that predict changes in vehicle trip rate [6]. The data set was given to CUTR by the South Coast Air Quality Management District, the air pollution control agency for all of Orange County and the urban portions of Los Angeles, San Bernardino and Riverside Counties of California [16]. The data was collected to supplement the Air Quality Management Plan for the area, in order to keep track of the vehicle trip reduction programs required of companies in the Los Angeles Area. All data in the data set was self-reported by the companies.

The data set consisted of 16,303 data points, with each data point representing a given year of a travel plan at a specific company. Each record indicated the specific incentives implemented by the transportation plan, as well as the number of years the transportation plan was in effect. Other characteristics indicated in the data set were target average vehicle rate, which is an indication of the locale of the worksite – urban, suburban or rural. Each record had 45 input variables associated with it. The output variable of interest was the change in vehicle trip rate (VTR). Vehicle trip rate is calculated as the number of cars commuting to a worksite per the number of employees at that worksite. VTR is important for traffic reduction, however it is not the ideal variable for looking at emissions reduction, which would be to look at single occupancy vehicle share. Unfortunately, a large data set was not available that tracked this variable.

CUTR compiled several of the characteristics into grouped variables to reduce the incidence of shared variance, lessen the complexity of the models created, and give more explanatory power to the model[16]. This grouping was repeated in this study for the same reasons, resulting in 9 grouped variables described in table 2.

Table 2: Grouped Variables Used in Data Set

COMPRESSED	Compressed work week schedules, allowing employees to work fewer, but longer days each month
ONSITE	Onsite Services, such as childcare, food services, ATMs, post office, fitness center that enable employees to complete errands at worksite.
MARKETING	Public Information Campaigns at worksite, including posters, fairs, presentations during new hire orientation, company recognition, etc.
RS_MATCH	Rideshare matching programs, that connect carpoolers with other commuters in their region
DIRECT_NONFINANCIAL	Gift certificates and other services as rewards for using alternatives to commuting alone
TELE	Allowing Employees to telecommute
FACILITY_AMENITIES	Passenger loading areas, bicycle facilities, showers
RIDE_HOME	Guaranteed ride home programs for ride sharers
PARKMGT$	Financial Incentives to not park at worksite – subsidized parking for ride sharers, increased charge for single occupancy vehicle commuters

In addition to these variables, the number of years that a travel plan had been in place at the worksite was included, as well as the target average vehicle ridership. After the outliers in the dataset were removed and entries with no recorded year were eliminated, the dataset was reduced to 14,753 data points.

Finally, the South Coast Air Quality Management District provided a document that linked the identification numbers in the data set with the names of the businesses and each record was categorized as a business, local authority, higher education or hospital. This system of grouping worksites was established in a 2002 paper by Rye that gives an overview of take-up levels of travel plans at UK establishments[17]. Due to missing ID numbers in the document provided by SCAQMD, the data set was reduced to 3,439 records.

3.2 Data Analysis

During the formatting of the data, statistical analyses were run to determine the variables of interest using both Microsoft Excel and MATLAB.

3.2.1 Linear Correlations

Linear correlations were run on each of the grouped variables with the change in vehicle trip rate using Microsoft Excel. The values are shown in Table 3.

Table 3: Correlation Coefficients of Input Variables

Variable	Correlations
Number of Years Travel Plan at Worksite	0.139176
Onsite Amenities	0.05774
Marketing	0.034581
Hospital Flag	0.028126
Business Flag	0.021136
Telecommuting	0.010167
Rideshare Matching Programs	0.008296
Higher Education Flag	-0.00423
Compressed Work Week Schedules	-0.00656
Guaranteed Ride Home Program	-0.01821
Parking Disincentives	-0.02743
Direct Non-financial Rewards	-0.03631
Local Authority Flag	-0.04584

None of the correlation coefficients are large enough to be considered significant. The lack of linear correlations means that it is unlikely that the relationships between these variables and the change in vehicle trip rate are linear. It is notable that the variable with the highest correlation coefficient is the number of years that a travel plan was in place at a worksite. Since the change in VTR is calculated from year to year, this shows that companies often see stronger results at the beginning of implementation of a plan, with the number of people changing modes of transport tapering off over the years.

3.2.2 Students' T-Tests

Microsoft Excel was used to run t-tests comparing the mean change in vehicle trip rate of each business type to each other. The mean change in vehicle trip rate is shown in figure 3, with the p values for the differences in mean in table 4.

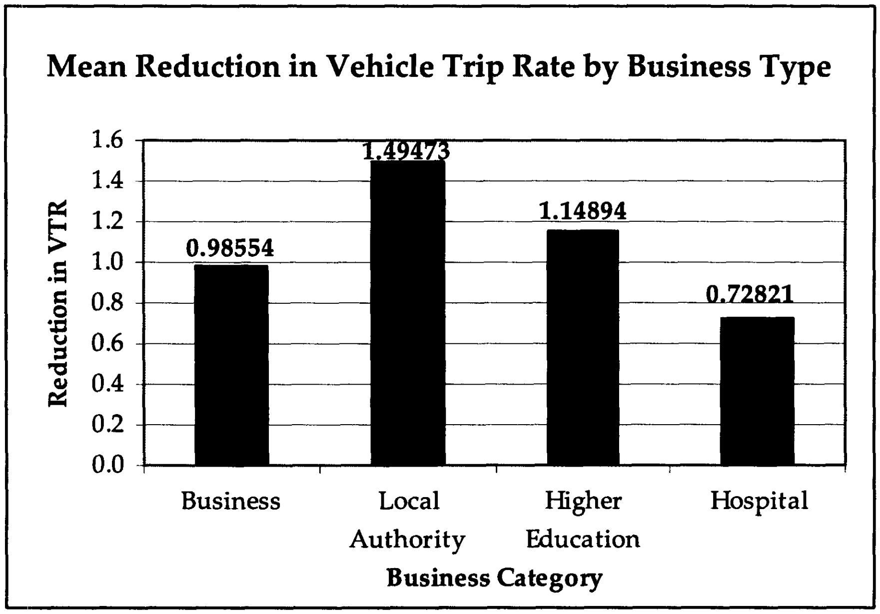

Figure 3: Mean Reduction in VTR (Cars/Person) by Type of Business

Table 4: P Values For Differences in Mean Change in Vehicle Trip Rate

p values, two tail	Business	Local Authority	Higher Education	Hospital
Business		**0.006163**	0.602217	0.319218
Local Authority	**0.006163**		0.287798	**0.007556**
Higher Education	0.602217	0.287798		0.255763
Hospital	0.319218	**0.007556**	0.255763	

These values show that there is a statistically significant difference between the average decrease in VTR of local government worksites as compared to both businesses and hospitals. None of the other differences could be deemed statistically significant.

It can be seen from the table that local authorities and higher education establishments are achieving the largest reductions in vehicle trip rate. This could be explained by several factors. One, commute trip reduction plans in this data set are in response to L.A. regional policy, required by the government. It is possible that government worksites have more accessibility to planners who could help develop the most appropriate plan for their worksites. Two, it could also be that civil servants and employees of higher education institutions are more likely to participate in commute trip reduction plans because of a greater awareness of the reasons behind the policy. Worksite characteristics factor into this analysis as well – hospital employees as well as many business employees may have less predictable schedules, thus requiring personal vehicles to be brought to the worksite. For this reason, greater differentiation within the business type classification may prove to be informative.

3.2.3 Regression and Stepwise Regression Analysis

Using MATLAB, regression models and stepwise regression models were run on the data set. The multiple linear regression analysis used all 15 input variables and the change in VTR as the output variable. The R^2 value was found to be

$$R^2 = 0.0375.$$

This means that the input variables in the multiple linear regression analysis account for 3.75% of the variability in the output variable. This shows that the linear regression model is capable of fitting a line to the dataset inputs and outputs, but each of the coefficients is negligible, resulting in a model that depends mostly on the constant factor in the linear regression equation. Although the results may seem informative, the model does not reveal anything significant about the relationships between the inputs and outputs. The plot of the predicted change in VTR versus the actual change in VTR is shown in Figure 4.

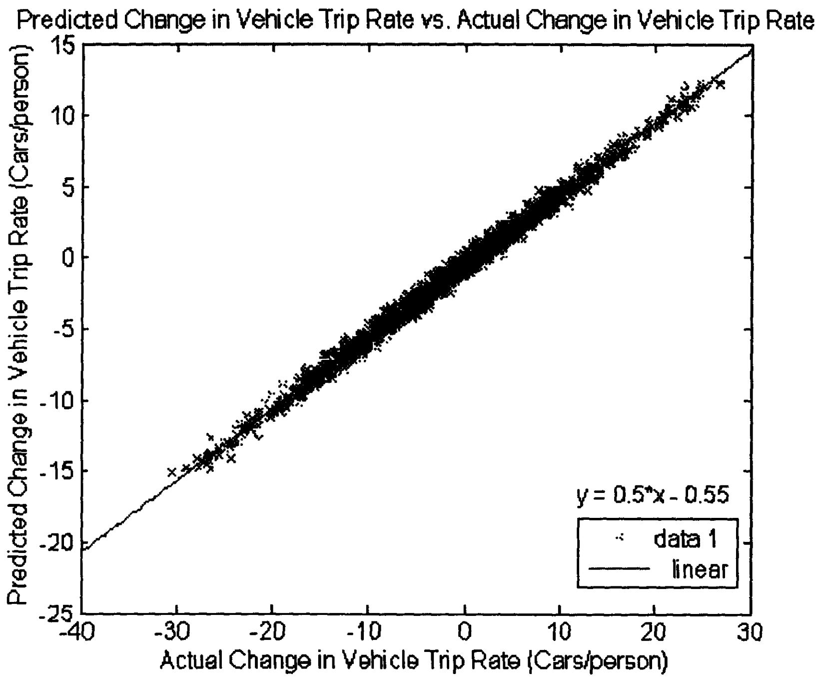

Figure 4: Predicted Change in VTR vs. Actual Change in VTR Multiple Linear Regression Model

Forward stepwise regression, which allows users to input statistically significant variables one at a time until there are no statistically significant variables remaining, was also run on this data set. The p values required to enter the model and to be removed from the model were $\alpha_{in} \leq 0.05$ and $\alpha_{out} \geq 0.1$. The stepwise regression model's inputs in the order of introduction to the model are in table 5.

Table 5: Stepwise Regression Inputs by Order of Insertion to Model

Variable	t statistic	p value
No. of Years of Travel Plan	6.9039	0.0000
Parking Disincentives	-2.8628	0.0042
Onsite Amenities	2.5689	0.0103
Direct Non-financial Rewards	-2.2448	0.0249

The R^2 value of the stepwise regression model was found to be

$$R^2 = 0.032,$$

which shows that the relationship between the inputs and the variation of the output of this model is not very strong. The plot of the predicted change in VTR versus the actual change in VTR for the stepwise regression model is shown in Figure 5.

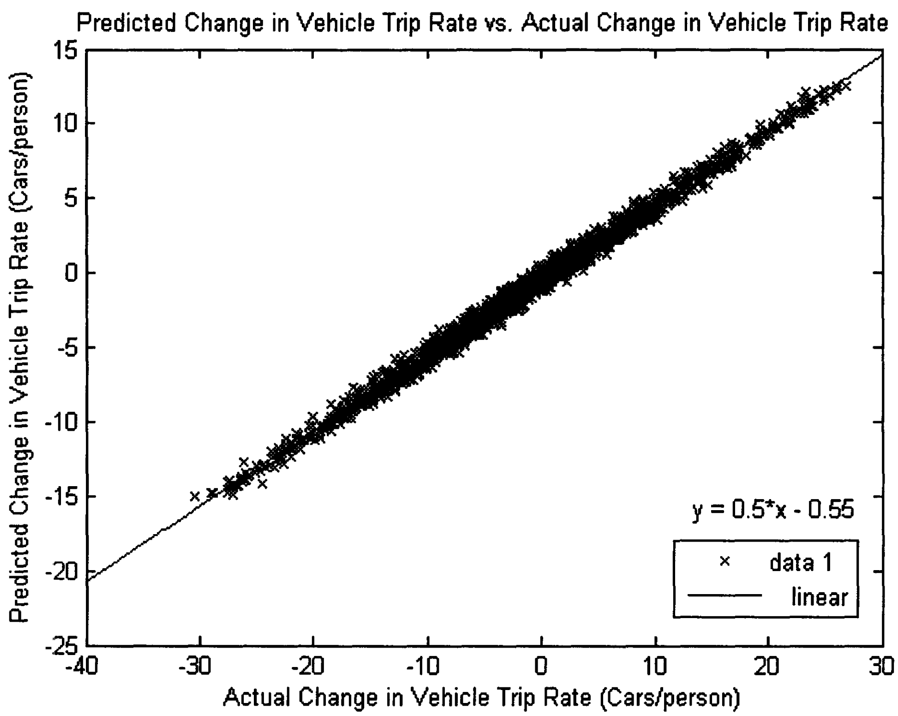

Figure 5: Actual Change in VTR vs. Predicted Change in VTR Stepwise Linear Regression Model

As can be seen from this plot, the results of the stepwise linear regression model are essentially the same as the results of the multiple linear regression model despite the use of only 4 input variables for the stepwise linear regression model.

Even though the relationship between the predicted and the actual data seems linear, the residual plot in figure 6 reveals that the model accuracy is poor.

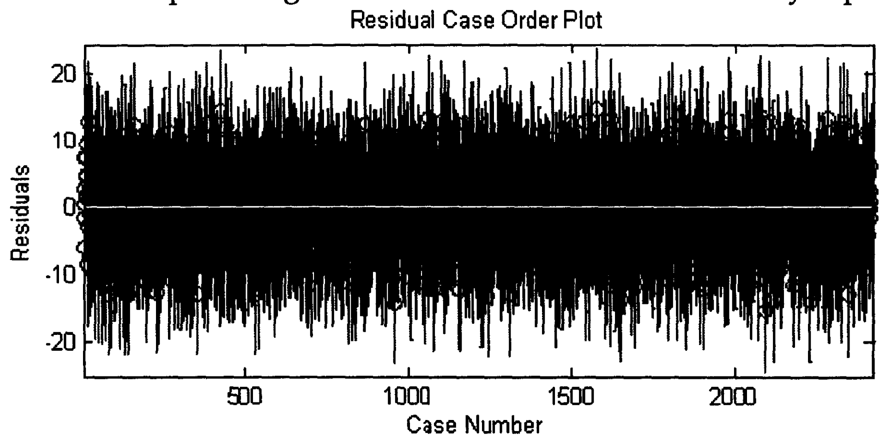

Figure 6: Residual Plot for Regression Analyses

This could mean that several of the input variables in this data set are correlated to each other or simply not linearly related to the change in vehicle trip rate, so linear regression cannot clearly model the outputs. This residual plot shows the high level of uncertainty in the predictions of the linear regression model: with 90% confidence intervals between -10 and 10, the model is not capable of providing accurate predictions for the change in vehicle trip rate.

3.3 Neural Network Procedure

The neural networks were built using MATLAB's Neural Network Toolbox, Version 7.0.4 on a 1.8 GHz Pentium IV PC with 1 GB of RAM. An algorithm developed by Mike Nolan at MIT was used to optimize the neural network, with Mean Squared Errors being tracked to assess the performance of each network. This algorithm cycled through the parameters shown in table 6 for a 4 layer artificial neural network.

Table 6: Parameters Varied in Neural Network Optimization

Parameter	**Values**
Number of Neurons in Layers 2 and 3	1 – 15
Transfer Functions for each Layer	Log sigmoid Hyperbolic tangent sigmoid Pure Linear
Training Function (all backpropagation)	Gradient descent with momentum and adaptive lr Levenberg-Marquardt Resilient Scaled Conjugate Gradient

The number of neurons in the first and fourth layers was kept constant at 15 and 1 respectively, in addition to the maximum number of epochs each iteration was allowed to run (170). The design of the neural network in this study is depicted in Figure 7.

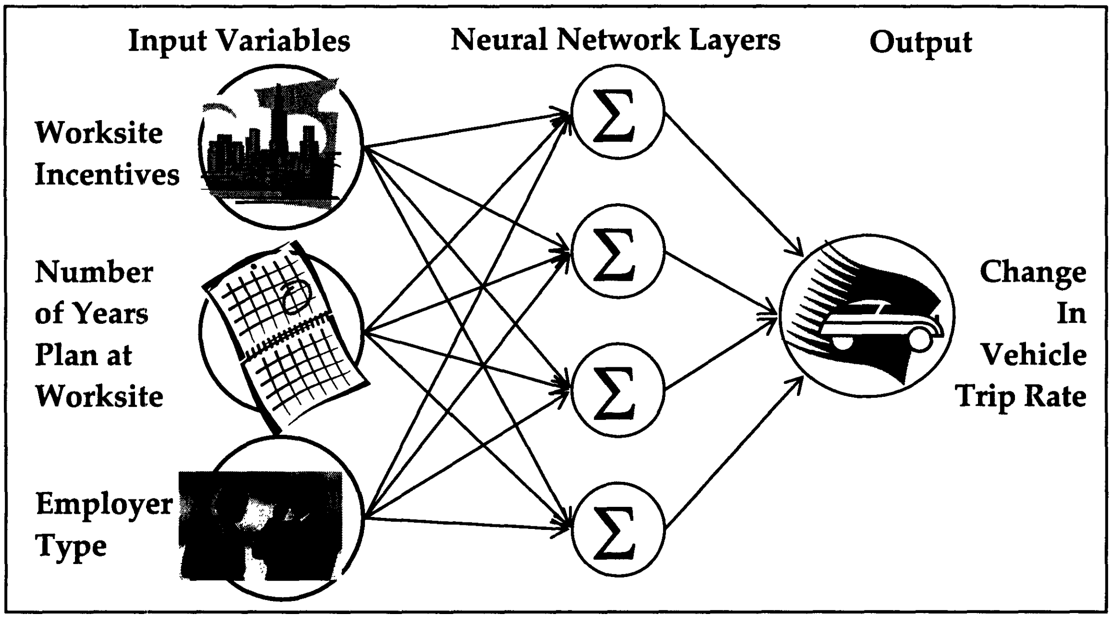

Figure 7: Design of Neural Network for Predicting Vehicle Trip Rate Changes Using Employer Incentives and Worksite Characteristics

In deciding between the different neural network architectures, the mean squared error of each network was tracked during the iterative process. The network with the smallest mean squared error for the training, validation and testing data was chosen as the optimal network.

4.0 Results

The neural network with the best results was determined to be the network with 3 layers, 15 neurons in the first layer, 7 in the second and 1 in the third. The first and second layers used log sigmoid transfer functions, while the third used a hyperbolic tangent sigmoid transfer function. The network was trained using Levenberg-Marquardt back propagation over13 epochs.

The mean squared error during the artificial neural network (ANN) training process is shown in Figure 8.

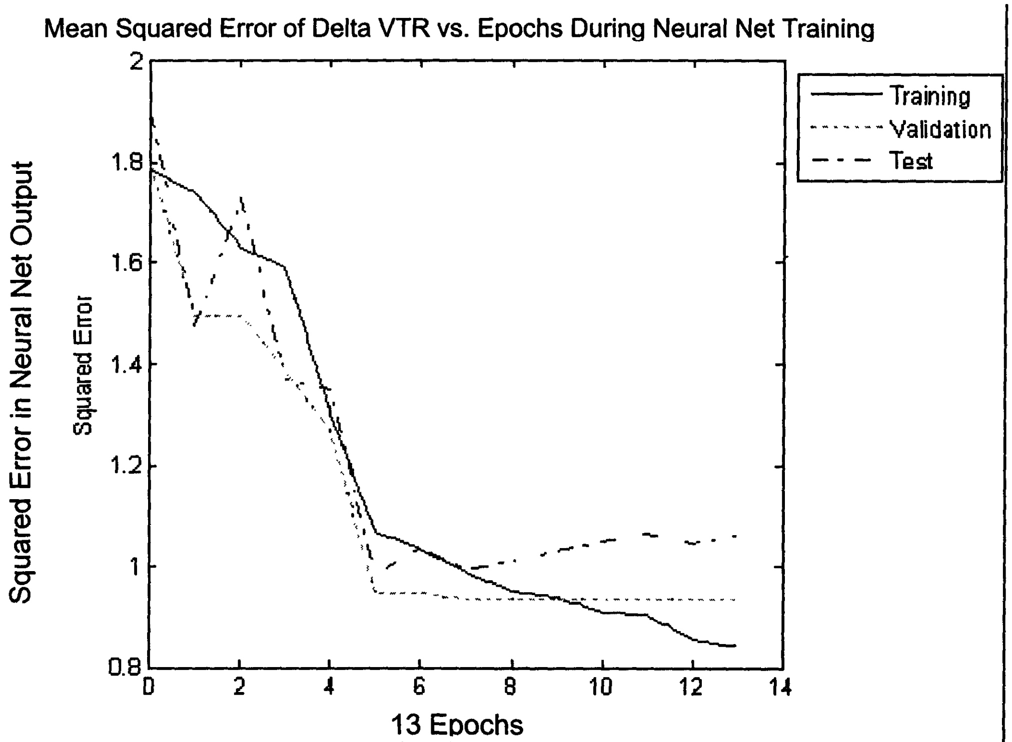

Figure 8: Mean Squared Error During Neural Network Training Process –Before Reversing Normalization of Data

The diagram of predicted change in VTR plotted against the actual change in VTR is shown in Figure 9.

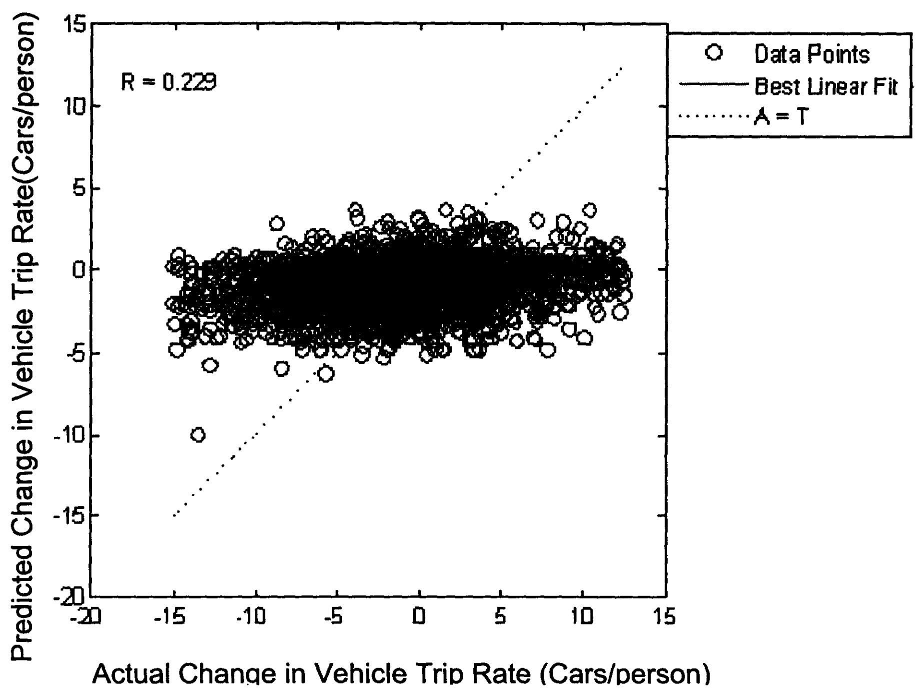

Figure 9: Regression Analysis Between ANN Output VTR and Actual VTR

It can be seen from figure 9, that the Neural Network is only capable of predicting outputs in a certain range [-5, 5]. The networks inability to predict large values of change in vehicle trip rate is reflected in the R square value for this network,

$$R^2 = 0.229.$$

This value means that the inputs into this network account for 22.9 percent of the variability in the change in the vehicle trip rate. Although this seems like a relatively high R^2 value compared to the linear regression model, it is clear from figure 9 that no matter the inputs to the ANN, it is predicting outputs in the same range. This diagram shows that the ANN is having difficulty fitting the inputs to the outputs of the data set – as well as that the predicted outputs are not very accurate to the actual outputs used to train and test the network. This means that additional input variables need to be determined in order to improve the predictive power of the network.

5.0 Discussion

5.1 Comparison of MSE and R^2 Values

The mean squared errors for the artificial neural network, stepwise regression and multiple linear regression models are shown in Figure 10.

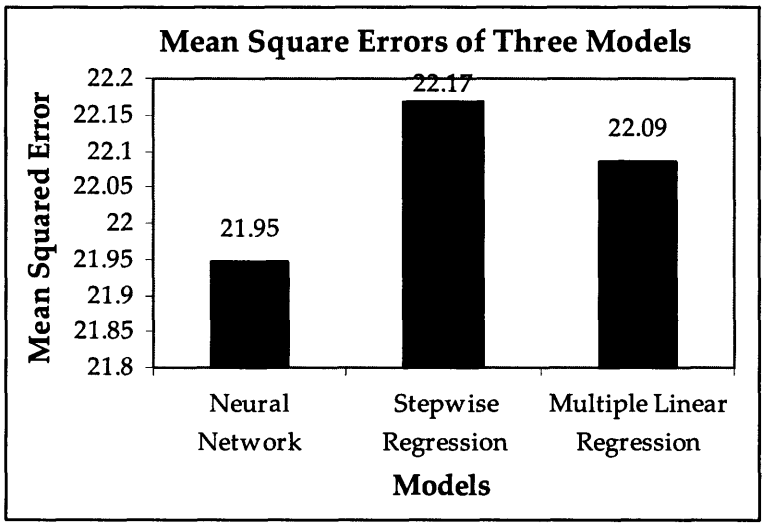

Figure 10: Mean Square Error Values for Neural Network, Stepwise Regression and Multiple Linear Regression Models

It can be seen that in terms of mean square error, the neural network outperforms the other two models. The R^2 statistic is an indicator of the fit of the for the multiple linear regression model, the stepwise regression model and the artificial neural network. The value for each network is plotted in Figure 11.

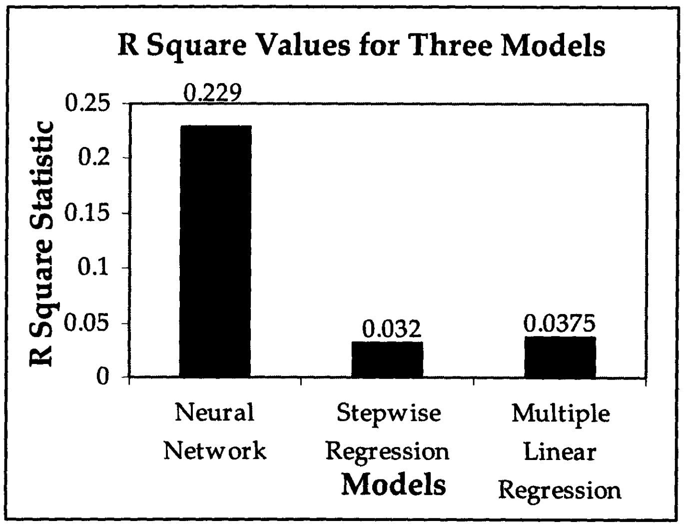

Figure 11: R Square Statistic for the Neural Network, Stepwise Regression and Multiple Linear Regression Models

The neural network again outperforms the stepwise regression and multiple regression models. This could be a result of the nonlinearity of the relationship between the inputs and the output of the models. However, the R^2 values are generally poor for all three of the models, and this shows that there could be considerable improvement in input variables to these models.

5.2 Error Comparison - Bins

In order to compare the models prepared in this study to the models prepared previously by CUTR using the same dataset, it is necessary to restructure the outputs in terms of bins. Using CUTR's method of placing each output in a bin, with bin sizes determined by the distribution of the data [6], the outputs and errors were reformatted using Microsoft Access. Error formatting involved creating two flags for accuracy – "Correct Bin" and "Off One Bin." "Correct Bin" was flagged if the model predicted a value for change in vehicle trip rate that was within the same bin as the actual measurement of change. "Off One Bin" indicates records for which the model predicted a value for change in vehicle trip rate that was either in the correct bin or in one of the neighboring bins. The bin intervals used are $(-\infty, -7)$, $[-7, -4.5)$, $[-4.5,-3)$, $[-3,-1.5)$, $[-1.5,0)$, $[0,1.5)$, $[1.5, 3.5)$, $[3.5, \infty)$. It was found that the bin accuracies for the linear regression models were significantly better than the bin accuracies for the neural network, as shown in figure 12.

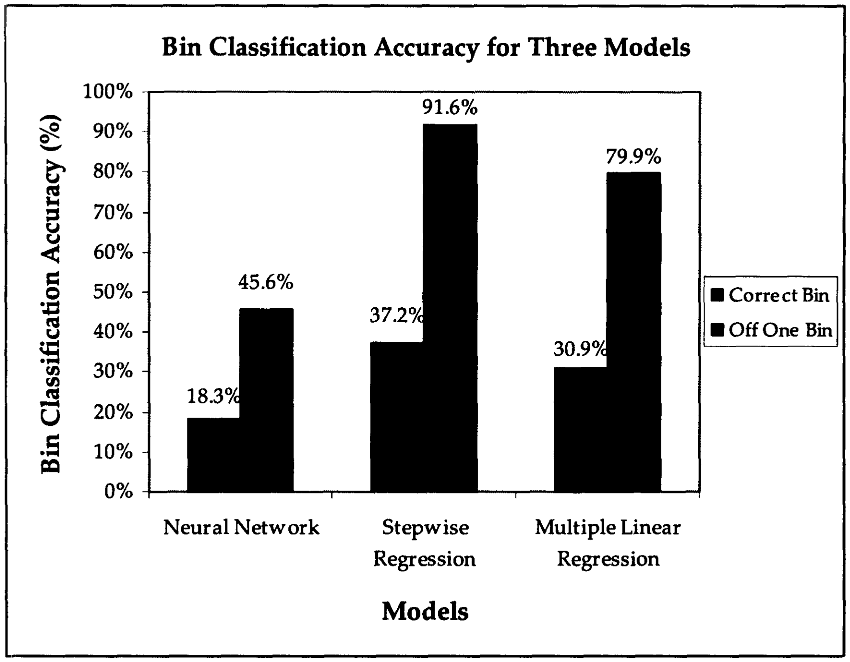

Figure 12: Bin Classification for Neural Network, Stepwise Regression and Multiple Linear Regression Models Using Complete Data Set

Looking at the research from CUTR for the full California Data Set in figure 13, the models reached peak values for bin classification data around 17% [6].

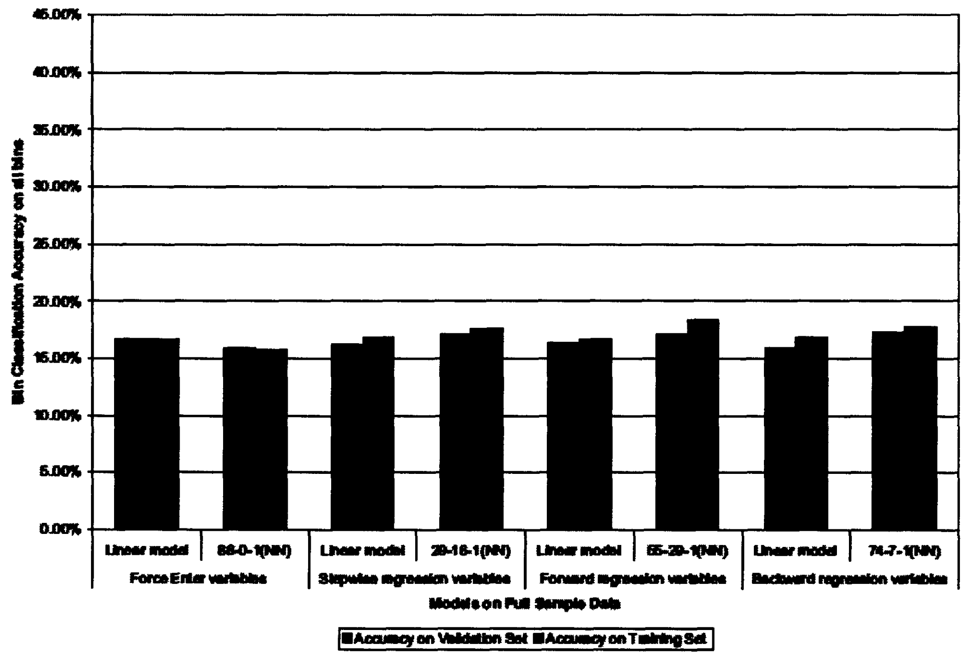

Figure 13: Bin Classification Accuracy on Full Range of Change in VTR for Validation and Training Set (Different Models on Full Sample Data) from CUTR Research [6]

The models created in this study, with the additional variables of Target Average Vehicle Ridership and the categorizations for business type are able to at least match – if not double or triple – the bin classification accuracy. These additional variables clearly have a significant impact on the predictive power of models for change in vehicle trip rate. This increased predictive power is most likely due to the inclusion of variables that describe the work setting and attempt to capture the culture of each worksite.

6.0 Conclusion

The research conducted in this study shows that government establishments see significantly better results in the reduction of vehicle trip rate as compared to businesses and hospitals. It also shows that there is not a clear linear correlation between the inputs tracked and change in vehicle trip rate. Despite the lack of correlation, the models built in this study still outperforms current models, but with limited improvement. This improvement can be attributed to the inclusion of the new parameters – classification of business type, target average vehicle ridership, and the number of years a plan has been in effect at a worksite. It is unclear whether the model has transferability to other data sets for other regions of the country other than the Los Angeles area because it has not yet been tested on other regions.

From this study it is clear that there is still much work to be done in order to accurately predict changes in vehicle trip rate at worksites. The characteristics of travel reduction plans currently being collected by local authorities simply do not contain enough information to describe a considerable amount of the variance of changes in commuter habits. In order to see higher R^2 values, better descriptive variables will have to be used as inputs, which requires better tracking of commute trip reduction plans. One variable that may be of interest is a more refined categorization of business type. Blue collar jobs tend to have more consistent workday hours than white collar jobs, which would facilitate carpooling. Also, the cultures of various businesses are likely to have an impact on how employees react to incentives to change commuting habits.

It is difficult to determine whether a neural network is preferable to linear regression models for predicting changes in commuting habits at worksites. While neural networks perform better on the level of mean squared error and R^2 values, the linear regression models have better accuracy when outputs are grouped into bins. The binning of data removes the weighting on the errors that the R^2 value represents, which explains why the neural network has better values of R^2 while the linear regression models perform better in terms of bins. Looking at the performance of models, the neural network performs better for the middle range of change in vehicle trip rate, values between -5 and 5. However, when predicting large changes in vehicle trip rate the linear regression model is the optimal choice.

Future work should begin to explore what parameters have higher statistical relationships with the output of vehicle trip rate. In conducting the literature review for this paper, it became clear that Washington state is recording variables that other data sets or not. It would be useful to investigate some of the parameters in that data set such as the increased detail of classification of employer type as well as clear indications of the locale of the workplace and thus the ease of commuting to the worksite without a car. The increased precision of employer type, looking at categories such as nonprofit, manufacturing, construction and agriculture, is likely to reveal that ridesharing is more suitable for certain commuting habits than others. Also, a closer examination of public information campaigns would most likely reveal the efficacy of specific elements of marketing efforts. It is important to understand these factors, as well as others, before continuing to build models in order to ensure that the factors of change in vehicle trip rate are being represented as accurately as possible.

References

[1] City Council, Climate Protection Task Force, City Staff. (2003). *City of Cambridge Climate Protection Plan.* Cambridge, MA.

[2] Greene, D. L. and A. Schafer. (2003, May). *Reducing Greenhouse Gas Emissions From U.S. Transportation.* Pew Center on Global Climate Change. Arlington, VA.

[3] Whitely, P. (1999, July 26). Companies Go the Extra Mile to Cut Commuting. *Seattle Times.* Monday Final Version. p. A11.

[4] Polakovic, G. (1995, December 9). Car Pool Program Replaced; Rideshare Trial Dies After 8 Years. *The Press-Enterprise.* p. A01.

[5] Winters, P. (2004, April 13). Combined Worksite Trip Reduction Model. (Online), May 5, 2005. http://www.nctr.usf.edu/worksite/

[6] Winters, P. et al. (2004, April). *Worksite Trip Reduction Model and Manual.* (Florida DOT Report No. BC-137-45) Tampa, FL: Center For Urban Transportation Research.

[7] European Conference of Ministers of Transport. (2001, November 29 -30). *Managing Commuters' Behavior: A New Role for Companies.* Paris, France: OECD Publications Service.

[8] Hurdle, D. (2004, October). Introducing Travel Plans at Business Parks. *Journal of the Institute of Logistics and Transport, 6*(8), 39 – 41.

[9] Washington State Department of Transportation. (2000, July). *TDM Success Story: University Village Employee Program Eases Urban Shopping Center Parking Problems.* Seattle, WA.

[10] Schipper, L. (1996, June). People in the Greenhouse: Indicators of Carbon Emissions from Households and Travel. *Revue de l'Énergie, 47,* 285-298.

[11] Henry, G. and C. Gordon. (2003). Driving Less for Better Air: Impacts of a Public Information Campaign. *Journal of Policy Analysis and Management, 22,* 45 – 63.

[12] Blok, K. (1997 May/June). Reducing Carbon Dioxide Emissions in Europe. *The Courier,* 53 -55.

[13] Shoup, D.C. (2002, March). The High Cost of Free Parking. Published in [7], 41 – 173.

[14] DeGroot, M. & M. Schervish. (2002). *Probability and Statistics*. 3rd Edition. Boston, MA: Addison-Wesley

[15] MATLAB. (2005, January 29). *Help Files.* Version 7.0.4.365. Natick, MA: The MathWorks.

[16] Wallerstein, Barry. (2004, April 28). *Introducing AQMD.* (Online) May 5, 2005. http://www.aqmd.gov/aqmd/intraqmd.html

[17] Rye, T. (2001, May). Company Management of Staff's Travel Choices in the UK. Published in [7], 175 – 200.

[18] United Kingdom Department for Transport. (2004, July 20). *The Future of Transport.* White Paper CM 6234. TSO. Norwich, UK.

[19] Washington State Department of Transportation. (2002) *Commute Trip Reduction (CTR) Program.* (Online) May 5, 2005. http://www.wsdot.wa.gov/tdm/program_summaries/ctr_summ.cfm