Science and Technology, 3 (1998) 11-20
© 1998 Sultan Qaboos University

A Comparative Analysis and Prediction of Traffic Accident Casualties in the Sultanate of Oman Using Artificial Neural Networks and Statistical Methods

Galal A. Ali*, Saleh M. Al-Alawi# and Charles S. Bakheit+

**Department of Civil Engineering, College of Engineering, #Department of Electrical and Electronics Engineering, College of Engineering, Sultan Qaboos University, P.O.Box 33, Al-Khod 123, Muscat, Sultanate of Oman; and +Department of Mathematics and Statistics, College of Science, Sultan Qaboos University, P.O.Box 36, Al-Khod 123, Muscat, Sultanate of Oman.*

التحليل المقارن والتنبوء بإصابات حوادث المرور في سلطنة عُمان بإستخدام الشبكات العصبية الإصطناعية وعلم الإحصاء

جلال عبدالله علي ، وصالح محمد العلوي ، وشارلس ساكي بخيت

خلاصة : تشكل حوادث الطرق في سلطنة عمان أحد الاسباب الرئيسية للوفاة لاسيما وسط الشباب إذ أن معدل الوفيات لكل ١٠٠٠٠ سيارة من النسب العالية في العالم . مما يوجب دراسة حوادث المرور وأسبابها والتنبؤ بحجم المتضررين منها وبالتالي التوصل الى الوسائل التى تحد من هذه الحوادث .ويهدف البحث الى تطبيق نظم الشبكات العصبية الاصطناعية ، بالاضافة الى قوانين الاحصاء ، لمعرفة عدد الوفيات والمصابين بإستخدام المعلومات المتوفرة من عام ١٩٧٦م إلى ١٩٩٤م عن العوامل ذات الصلة مثل عدد السكان والسيارات . وقد دلت نتائج البحث على أن الشبكات العصبية الاصطناعية تقدم النموذج الامثل للبيانات بالاضافة الى التنبؤ بعدد المتضررين من الحوادث في الفترة من عام ١٩٩٥م إلى عام ٢٠٠٠م ، مما يدل على أنها تمثل آلية فعالة لدراسة حوادث المرور .

ABSTRACT: Traffic accidents are among the major causes of death in the Sultanate of Oman. This is particularly the case in the age group of 16 to 25. Studies indicate that, in spite of Oman's high population-per-vehicle ratio, its fatality rate per 10,000 vehicles is one of the highest in the world. This alarming situation underlines the importance of analyzing traffic accident data and predicting accident casualties. Such steps will lead to understanding the underlying causes of traffic accidents, and thereby to devise appropriate measures to reduce the number of car accidents and enhance safety standards. In this paper, a comparative study of car accident casualties in Oman was undertaken. Artificial Neural Networks (ANNs) were used to analyze the data and make predictions of the number of accident casualties. The results were compared with those obtained from the analysis and predictions by regression techniques. Both approaches attempted to model accident casualties using historical data on related factors, such as population, number of cars on the road and so on, covering the period from 1976 to 1994. Forecasts for the years 1995 to 2000 were made using ANNs and regression equations. The results from ANNs provided the best fit for the data. However, it was found that ANNs gave lower forecasts relative to those obtained by the regression methods used, indicating that ANNs are suitable for interpolation but their use for extrapolation may be limited. Nevertheless, the study showed that ANNs provide a potentially powerful tool in analyzing and forecasting traffic accidents and casualties.

Every day the world witnesses increasing traffic accidents and casualties (fatalities and injuries). Approximately half a million people are killed and 10 million injured in road traffic accidents (RTAs) in the world every year. These are in addition to more than 10 million who are disabled (Petrucelli, 1991; Pattnaik and Sreedar, 1993; Oman Daily, 1993). Developing countries represent 67% of RTA fatalities in the world although they own only about 11% of the vehicles. Treatment of traffic accident injuries accounts for 5 to 10% of hospital costs, amounting to about 90 billion US dollars in the USA, Germany and UK alone. These alarming statistics underline the importance of continually updating our methods of analyzing traffic accident data and of providing better predictions of accident casualties. These efforts will contribute to the assessment of new traffic regulations and safety measures for safer driving.

A review of the situation on road traffic accidents in some Arab countries and the Middle East underscores the magnitude of the problem in the Arab world. The 1993 traffic casualties in Egypt amounted to 5,000 fatalities and 23,000 injuries. The main causes of accidents were attributed to speeding, driver negligence and violation of traffic regulations, a pattern observed in many countries (Ali et al. 1995). In 1994 Jordan experienced about 27,000 accidents, 70% of which occurred in the capital city. The casualties were 443 fatalities and 12,516 injuries, with 40% of these involving pedestrians. The 1990-1993 annual average accident statistics in Tunisia showed that fatalities and injuries were 1270 and 12,430, respectively, resulting from 9,650 accidents. In Morocco more than 3,000 people are killed and 60,000 injured in traffic accidents every year. The number of accidents in 1994 was 43,681, resulting in 3,605 fatalities, which was an increase of 7.3% from the 1993 figure.

Road Traffic Accident Casualties in the Gulf Region

Trends similar to those in the above North African Arab countries were also observed in the Gulf region. The 1995 RTA casualties in the Kingdom of Saudi Arabia were 4,000 fatalities and 31,000 injuries. Additional accident costs were more than half a billion US dollars. Speeding was the main cause for 45% of the 147,000 accidents in the Kingdom (Al-Khaleej, 1996). Eighty per cent of those involved were in the age group of 18-40 years. The significant involvement of young drivers, taxis and animals is a special feature of RTAs in the United Arab Emirates (UAE). In 1993 alone, 90% of the 521 fatalities in UAE were in the youth age group. In Bahrain, the number of persons per car in the past three decades dropped from 12 to 3, making Manama one of the most congested capital cities in the world. In the middle of 1994, traffic police had to be assigned close to schools. The result was that, the number of children involved in pedestrian accidents reduced drastically, from 43% in the first half of 1994 to 23% during the second half (Asharq Al-Awsat, 1994). Although improved road network and advanced traffic control systems characterize the traffic environment in Kuwait, more accidents occurred due to poor driving and to pedestrians' lack of awareness of traffic rules. The 1993 RTA tolls were among the highest in the world with 19,785 accidents, resulting in 290 fatalities and 2,018 injuries (Al-Damyan, 1995).

In the Sultanate of Oman, injuries and fatalities are about 6 to 8 times more than in many developed countries despite the lower car ownership of 11 vehicles per 100 population. More than 60% of the casualties are in the age group of 16-50 years, the major cause of death for 47% of this age group being RTAs (Ali et al., 1994). Pedestrian fatalities in Oman amount to about 32% compared to 18% in the USA. In 1993 alone, property damage from RTAs cost 2.225 million Omani rials (about 6 million US dollars). The high rate of population growth, the large proportion of young drivers and the dramatic increase in

TABLE I

Accident and Casualty Rates in Oman – 1995 and 1997.

Accident/Casualty Parameter	1995	1997
Accidents per 10^4 vehicles	367	236
Accidents per 10^5 population	512	380
Accidents per day	30	23
Fatalities per 10^3 accidents	43.5	65
Fatalities per 10^5 population	22	24
Fatalities per month	40	46
Injuries per 10^3 accidents	500	862
Injuries per 10^5 population	309	314
Injuries per day	18	20
Traffic violations per day	500	537
Total population	2161732	2315701
Total vehicle fleet	300238	357880

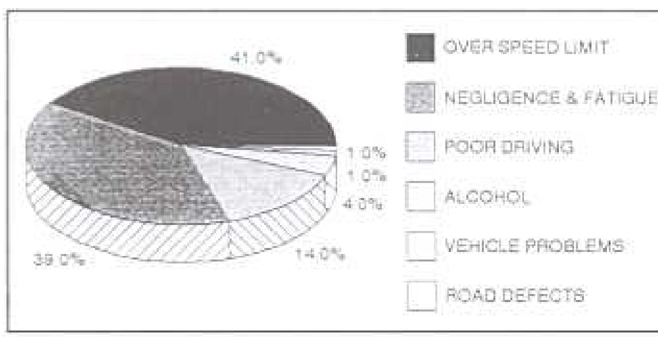

Figure 1. Causes of accidents in Oman according to ROP 1995 records.

the number of vehicles over the years have contributed to the high accident rates. Table 1 depicts and compares the various 1997 accident and casualty rates in the Sultanate with those of 1995. The table indicates that casualties (fatalities and injuries) have increased despite the significant reduction in the number of accidents. Generally, drivers have been the main cause in about 98-99% of accidents for the past several years, as shown in Figure 1. This is more than the global value of 95% (Ergun, 1988). The various studies in the Arab world and the Gulf region indicated that the main contributors to RTAs are high speeds and the pedestrians.

Key ingredients for successful traffic improvement programs and prerequisites for traffic safety management are the availability of sufficient and reliable data, and the capability to predict traffic accident casualties and safety situations. Improvement schemes and effective safety management programs can then be developed for implementation and assessed. A crucial requirement, therefore, for the development of successful traffic improvement programs is the availability of a reliable predictive model that incorporates the essential factors related to accidents and casualties.

Accident casualty prediction today relies on various models and methods, based largely on statistical techniques and expert systems. They predict accidents and casualties either as a function of registered vehicles per population (Smeed and Jeffcoate, 1970; Jacobs and Hutchinson, 1973; Al-Suleiman and Al-Masaeid, 1992), or traffic flow and road type (Jadaan and Nicholson, 1992). Others use time (Ali et al., 1994; Wong-Toi, 1994), or many more exogenous variables (Pattnaik and Sreedar, 1993). Causes of traffic accident casualties are numerous and complex and safety may be related to several factors such as road geometry, traffic characteristics and enforcement. Sometimes the developed models do not fit the data (Jacobs and Fouracre, 1977; Wong-Toi, 1994), or only address linear relationships between the dependent and independent variables. A more pragmatic approach is required to account for the important variables and their relationships in analyzing and predicting traffic accident casualties. The main objective of this study was to predict traffic accident casualties for Oman up to the year 2000, applying both regression analysis techniques and ANNs, and to present a comparative analysis of the results.

Artificial Neural Networks

Artificial Neural Networks (ANNs) are computer models that are designed to emulate human information processing capabilities such as knowledge processing, speech, prediction, and control. The ability of ANN systems to spontaneously learn from examples, to reason over inexact and fuzzy data, and to provide adequate and quick responses to new information not previously stored in memory has generated increasing interest in and acceptance for this technology in the different engineering fields. As a result of numerous applications in engineering, this new tool has demonstrated remarkable success.

Artificial Neural Networks are relatively new in the traffic engineering field. This new approach has only been sparsely demonstrated in areas such as traffic management (Goodman et al., 1993), traffic congestion forecasting (Gilmore and Abe, 1993; Taber et al., 1995), estimating traffic characteristics (Elizandro et al., 1993), determining truck attributes (Gagarin et al., 1994), traffic control (Nakatsuji et al., 1993), and a few other applications. Recently, Wu and Heydecker (1995) and Al-Alawi et al. (1996) investigated the application of computer-based natural language understanding techniques and artificial neural networks, respectively, to the analysis and prediction of road traffic accidents.

Artificial Neural Networks are now also being used in problems that have been traditionally solved by statistical methods. A number of studies have been carried out to compare the methods from statistics with ANNs. Ripley (1993) has written an extensive review paper with examples on ANNs in a statistical context. In another of his papers on the subject (Ripley 1994), he set up a general framework for comparing the statistical methods of classification with ANNs. Other papers that have made comparative studies of statistical methods with ANNs include Weiss and Kulikowski (1991), Ripley and Hjort (1994) and Chen and Titterington (1994). More recently, Stern (1996) has discussed the relevance of ANNs models to applied statistics and used a time series prediction problem as an example. These studies are all of the view that ANNs, if trained on large data sets, can be quite good for prediction, as the construction requires no assumptions concerning the functional form of the relationship between predictor variables and response variables. However, this can be a severe drawback if investigations for which an understanding of relationships between variables are required.

Data Preparation and Rehabilitation for Model Building

The most important factors considered to contribute to annual casualty figures included annual population size, the gross domestic product of Oman (GDP), the number of cars on the road, the number of driving licenses issued, and the total length of paved roads. Many other factors could have been incorporated, such as road and vehicle conditions, driver negligence, speeding, poor driving skills, etc. However, these are either difficult to quantify or they do not have complete historical data over the period of study. Historical data on traffic and accident casualty statistics were readily obtained from the Directorate General of Traffic (DGT), Royal Oman Police (DGT, 1997). Data on population, GDP and the length of paved roads were obtained from the Oman Statistical Yearbook of the Oman Development Council (1994 and 1996). The preparation and rehabilitation of the data are described in Al-Alawi et al. (1996).

With all candidate variables analysed and corrected, correlation analysis was performed in order to determine the linear significance of variables. Results of the correlation analysis (Table 2), using the coefficient of correlation (R^2), indicated that the population and the number of licenses issued were found to have the highest correlation to the number of accident casualties. Next were the number of cars on the road and the number of kilometres of paved roads. The gross domestic product

TABLE 2

Correlation Matrix.

Variable	POPG	GDP	NOCR	NOLI	NKPR	No. of Casualties
POPG	1.0	0.941	0.993	0.996	0.973	0.985
GDP		1.0	0.958	0.940	0.981	0.826
NOCR			1.0	0.996	0.981	0.971
NOLI				1.0	0.970	0.980
NKPR					1.0	0.949
No. of Casualties						1.0

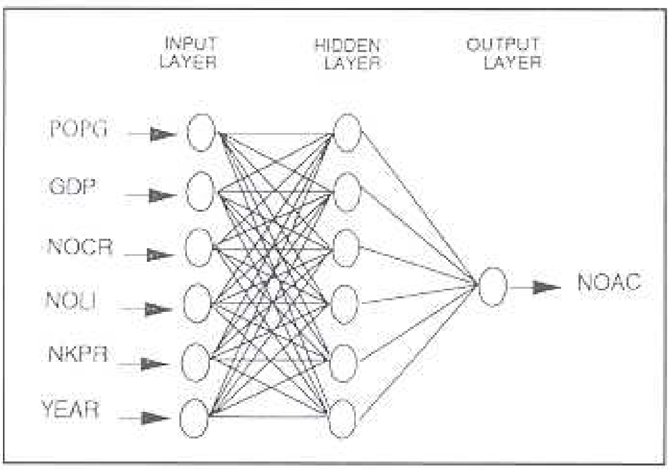

Figure 2. The architecture of the proposed ANN model.

was also found to be highly correlated to the number of accident casualties.

Developing The Artificial Neural Networks Model

As Figure 2 shows, an ANN-based model was developed for the modeling and prediction of the number of accident casualties in the Sultanate of Oman. The variables selected for developing the ANN models were as mentioned previously, namely: the population growth (POPG), the gross domestic product (GDP), the number of cars on the road (NOCR), the number of licenses issued (NOLI), and the number of kilometres of paved roads (NKPR). All these variables were assumed to be functions of the year (Y), and were chosen as input parameters to the proposed ANN architecture. The number of car accident casualties (NOAC), the dependent variable, was chosen as output in the model. Since the relationship of the above variables to the number of casualties may not be linear, the following nonlinear model was proposed:

$$NOAC_Y = f(POPG_Y,\ GDP_Y,\ NOCR_Y,\ NOLI_Y,\ NKPR_Y) \tag{1}$$

where Y=1,2,3,......n; n being the number of years which the ANN model was to be trained, and $NOAC_Y$ the predicted number of accident casualties for year Y.

As illustrated by the ANN architecture in Figure 2, each network comprises many simple processing elements that are organized into a sequence of layers. These are the input layer, the hidden layer, and the output layer. The neurons in the input layer receive six input signals representing the above input variables. Hence, six neurons were used for the input layer in the ANN architecture. The output layer, on the other hand, consists of one output neuron representing the number of casualties. Between the input and output layers, generally, there is one or more hidden layers. Since there is no direct and precise way of determining the number of hidden layers to use and the exact number of neurons to include in each hidden layer, one hidden layer containing five neurons was found adequate to develop the model. Research by Lapeds and Farber (1988) and Hecht-Nielsen (1989) indicated that one or two hidden layers with an adequate number of neurons is sufficient to model any solution surface of practical interest. The number of hidden neurons is a function of the complexity of the problem, of the number of input and output parameters, and of the amount of training cases available. Hence, a trial and error process was used to determine the number of hidden neurons. After trying a number of different configurations of hidden neurons, it was found that five neurons in one hidden layer yielded the best results for the model.

The multilayer feedforward networks used in this study were trained using the Backpropagation (BP) paradigm developed by Rumelhart and McClelland (1986). The BP algorithm uses the supervised training technique. In this technique, the interlayer connection

TABLE 3

Sample of Training Patterns Used to Develop the ANN Model.

Year	$POPG_Y$ (persons)	GDP_Y (10^6 RO)	$NOCR_Y$ (vehicles)	$NOLI_Y$ (licenses)	$NKPR_Y$ (km)
1976	1,124,437	884	28,603	20,140	653
1980	1,290,318	2,064	68,767	119,643	2,052
1984	1,480,669	3,047	138,792	295,628	3,002
1987	1,641,644	3,003	178,638	407,916	3,506
1990	1,820,120	4,051	219,275	533,615	4,362

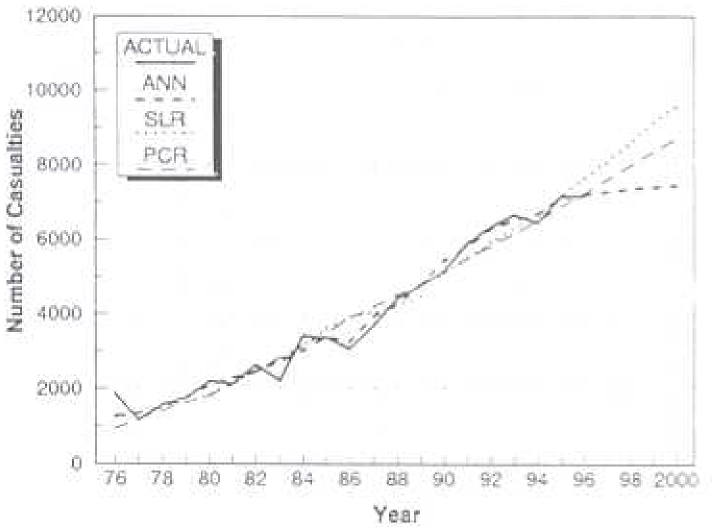

Figure 3. ANN, PCR and SLR casualties development and prediction models (1976 - 2000).

weights and the processing elements' thresholds were first initialized to small random values. The network is then presented with a set of training patterns, each consisting of an example of the problem to be solved (the input) and the desired solution to this problem (the output). Historical data covering the period from 1976 to 1994 were used in training the proposed model. Typical examples of the different training patterns used as part of the training data set are shown in Table 3. These training patterns were presented repeatedly to the ANN model, and weights were adjusted by small amounts that were dictated by the General Delta Rule (Rumelhart and McClelland, 1986). This adjustment is performed after each iteration when the network's computed output is different from the desired output. This process continues until weights converge to the desired error level or the output reaches an acceptable level. The system of equations that provides a generalized description of how the learning process is performed by the BP algorithm is described by Simpson (1990). The training process of the ANN models was performed using the NeuroShell® simulator. The network converged to a threshold of 0.0001. The R^2 value for the model was 0.990. This indicates that approximately 99.0% of the variability in the number of car accident casualties (the dependent variable) could be explained by the selected independent variables and the historical data used. Having trained the network successfully, the next step was to test the network in order to judge its performance and to examine its generalization capabilities.

Network Testing, Validation and Prediction

The numbers of casualties predicted by the ANN model were compared with the actual observations. It was found that the model predictions were good. In fact, the 1995 prediction of 7,117 car accident casualties differs from the actual observation (7,164) by less than 0.7%, and the 1996 prediction of 7,203 casualties differs from the recorded value of 7,166 by about 0.5%. Nonetheless, Figure 3 shows that ANN does not appear to smoothen sufficiently the stochastic components of the empirical data.

Once the model was developed and it produced accurate results, the contributions of the different independent variables to the variation in the values of the dependent variable were obtained using the NeuroShell® utility. Examination of these contributions (Table 4) revealed that the number of licenses issued ($NOLI_Y$) and the gross domestic product (GDP_Y) had substantial influence on the increasing number of car accidents (20.6% and 20.0%, respectively). The increase in GDP_Y reflects the country's prosperity and economic growth, resulting in consumer purchasing power of cars. The number of cars on the road ($NOCR_Y$) accounted for 16.5% of the variation in the number of accident casualties, while 15.5% could be explained by the change in population ($POPG_Y$). These contributions were computed by the NeuroShell® program as measures of each input

TABLE 4

The Contribution of the Input Parameters to the Output Results.

Parameter	$POPG_Y$	GDP_Y	$NOCR_Y$	$NOLI_Y$	$NKPR_Y$
Contribution	15.5	20.0	16.5	20.6	6.8

strength in relation to those of the other input parameters of the model being developed. The determination of the contribution of a predictor variable, for instance 20.6 % for NOLI, is based on how this variable is related to the other variables in the model, and not how the variable is related to the predicted NOAC. Examination of the historical data indicated that the motorization level increased from 25 cars per 1,000 people in 1976 to 139 cars per 1,000 people in 1995. This growth in the motorization level can be explained by the increase in the population's purchasing power and the large percentage increase in the number of young drivers. It is known that approximately 25% of the population is in the age group of 16-29 years and this group is the largest contributor to the number of accident casualties. Other factors, such as the number of kilometres of paved roads ($NKPR_Y$), also contributed, although marginally (6.8%), to the variation in the number of car accident casualties.

In order to use the developed model to predict the number of car accident casualties for the years to come, it was necessary first, to forecast the model input variables. With the exception of the population growth, a linear regression analysis, with year as the independent variable, was used to project these variables by curve-fitting the historical data (Eqs. 3 - 5). The best fit for the number of licenses issued was a cubic function of year (Eq. 6). For the population growth, an annual growth rate of 3.5%, provided by the Oman Development Council (Eq. 2), was used. Needless to mention, these projections are one of many plausible scenarios that could evolve in the future. With the many developments in road construction and upgrading, in addition to the continuous improvement in traffic management in the Sultanate, it would be difficult to make more accurate projections of these variables.

Equation	R^2	
$POPG_{Y+1} = POPG_Y\,[\,1.035\,]$		(2)
$GDP_Y = 835.4+198.5Y$	0.918	(3)
$NOCR_Y = 6902 + 14144Y$	0.995	(4)
$NKPR_Y = -1031.8+1379.8\ Y$	0.972	(5)
$NOLI_Y = 19247.3+3872.3Y^2-112.2Y^3$	0.995	(6)

Using these equations, forecast data for the years 1995 to 2000 were generated for each variable. The projected values of the predictor variables were then fed as input to the ANN model in order to predict the number of car accident casualties for each year until the year 2000.

The reduction in the rate of increase in the length of the paved roads and the number of licenses issued with rigorous licensing procedures and the strict enforcement of traffic regulations would tend to reduce the number of accidents. On the other hand, the constant population growth rate and continuous increase in GDP and the number of vehicles on the road (traffic volume) would result in a counter-effect to increase the number of accidents. The net result is a slow growth in the predicted accidents over time due to the combined effects, as demonstrated by Figure 3. However, this slow growth could also be attributed to overemphasis by the model of the relatively small increase in NOAC observations during the years 1995 and 1996.

Linear Regression Prediction

For comparative purposes, the same data set was modeled using regression analysis techniques. The classical linear regression model consists of a dependent variable, Y, which depends linearly on a set of r predictor (explanatory) variables, $Z_1, Z_2, \ldots, Z_r$, and a random error term ϵ. This error term accounts for both measurement error and the possible effects of other variables not explicitly included in the model. The values of the predictor variables are considered fixed. It is also assumed that the random errors are independent and normally distributed with mean equal to zero and a constant variance denoted by σ^2. In a real life situation usually these assumptions are only approximately satisfied (Myers, 1990).

As shown in Table 2, preliminary examinations of the predictor variables indicated that they were highly correlated (0.91 - 0.99). It is well known that, in the presence of such strong multicollinearity, ordinary least squares regression fitting yields unstable coefficients of regression, with inflated standard errors (Myers 1990). Even when a model adequately fits the data, the prediction quality of the model is severely curtailed outside the range of the explanatory variables used to fit the model. There are a number of modified regression procedures available to handle the problem of multicollinearity. These include ridge regression, principal component regression and latent roots regression methods. These methods produce stable but biased estimates of coefficients of regression with reduced variances (Jolliffe, 1986). In this study we used principal component regression. The choice was more of convenience than for any technical superiority of one method over the other. The other two methods were also used for the analysis, but the results were not that much different from the principal component regression results.

The principal component method involves the transformation of the standardized predictor variables into an equivalent number of new variables called principal components, so that these new variables are orthogonal to each other. The principal components are ordered such that the first one accounts for the largest proportion of variability in the original data set, and the last one accounts for the least proportion of the variability (Myers, 1990).

After the predictor variables were transformed, the least squares procedure was applied to obtain a prediction equation, with the standardized casualty figures as a function of the principal components. The best fit was an equation based on the first two principal components that accounted for 99.6% of the variation in the original standardized predictor variables. The result showed that 97.26% of the variation in the number of casualties was explained by the regression equation. The equation was transformed back to a function of the original variables, resulting in the following model:

$$\mathrm{NOAC} = -1674.1 + 0.0021786\ \mathrm{POP} - 0.1614965\ \mathrm{GDP} + 0.0065351\ \mathrm{NOCR} + 0.097412\ \mathrm{NKPR} + 0.0030712\ \mathrm{NOLI} \qquad (7)$$

The above principal component regression (PCR) model, includes all the explanatory variables, and is based on the statistically significant coefficients of the principal components used and their goodness of fit. However, a test of significance of the coefficients in equation (7) showed that the only statistically significant regression coefficient at the 5 % level was that of the number of cars on the road (NOCR). This meant that, with the exception of NOCR, the other four predictor variables were probably redundant and could therefore be dropped from the model. The PCR model was still retained as a viable model for comparison on the strength of the significance of the coefficients of the principal components used.

A number of simple linear regression (SLR) models were also fitted to the data, using only the predictor variable NOCR. Several combinations of different transformations of both the dependent and the predictor variables were attempted, and the best fit was obtained by the following regression model:

$$\sqrt{\mathrm{NOAC}} = 29.99 + 0.000189\ \mathrm{NOCR} \qquad (8)$$

This SLR model explained 95.99% of the variation in the number of casualties, and both regression coefficients were highly significant ($p = 0.000$). In addition, no serial correlation of the residuals was indicated, implying a good fit.

Both models showed that the casualty figures for 1983 and 1986 were atypical. Apparently 1983 was the year that the upgrading of one of the main country's highways (the Batinah dual carriageway) was completed, leading to a significant drop in casualties on that busy stretch of road. On the other hand, the low casualty figure for 1986 could be attributed to the worldwide economic recession experienced that year, which meant low economic activities and therefore less traveling. The data were fitted with these observations as missing and the results did not differ significantly from those obtained by both models.

According to the PCR based model, except for GDP, an increase in any of the regressor variables would raise the number of casualties. On the other hand, higher GDP would mean lower casualty figures. However, as these regressors are all highly correlated, a change in either of them precipitates changes in the others. Thus, an increase in the national GDP would contribute to more construction of paved roads, the potential of owning more vehicles and the acquiring of more driving licenses. As for the SLR model, casualties are related quadratically to the number of cars on the roads.

The regression models were used to extrapolate casualty figures from 1995 up to the year 2000, using the projections of the predictor variables provided by equations 2 to 6, and the results are shown in Table 5.

Comparison of the ANN and Linear Regression Models

The main objective of each of the methods was to fit an accurate model of the casualty figures and to use it for prediction of future figures. The adequacy of such models is typically measured either by the coefficient of determination (R^2) of the predictions against actual values or by the mean squared errors of the estimates (MSE). The model with the smaller MSE and the higher R^2 is normally considered the better model. However, it has been shown that the models that fit best by these criteria do not necessarily provide the best prediction (Myers, 1990).

Based on these criteria, the ANN model gives the lowest MSE of 76,243.8 as compared to 152,033.1 and 124,988.2 for PCR and SLR, respectively. It also has higher R^2 (0.990) than either PCR ($R^2 = 0.973$) or the SLR

TABLE 5

Models predictions for the number of casualties for the years 1995-2000.

YEAR	1995	1996	1997	1998	1999	2000
ANN Forecast	7117.3	7202.5	7293.5	7366.3	7405.1	7443.7
PCR Forecast	6889.4	7247.1	7610.4	7979.2	8354.1	8734.9
SLR Forecast	7184.1	7644.4	8119.0	8607.8	9111.0	9628.5
Observed NOAC	7164	7166	7827			

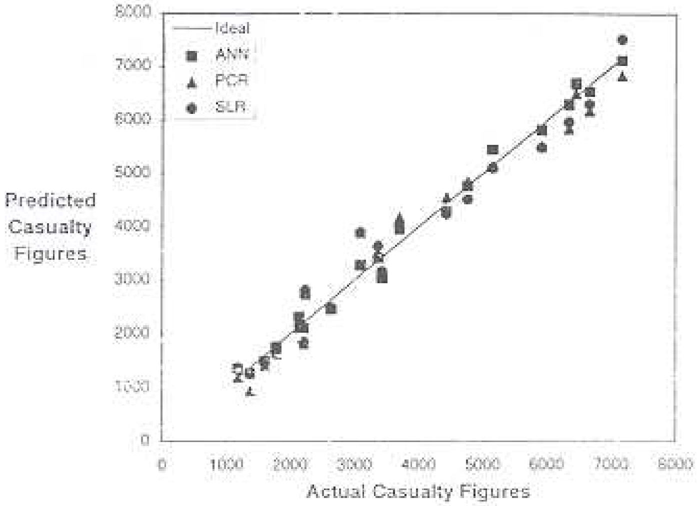

Figure 4. Scattergram comparison of predictions by ANN, PCR and SLR (1976 - 1994).

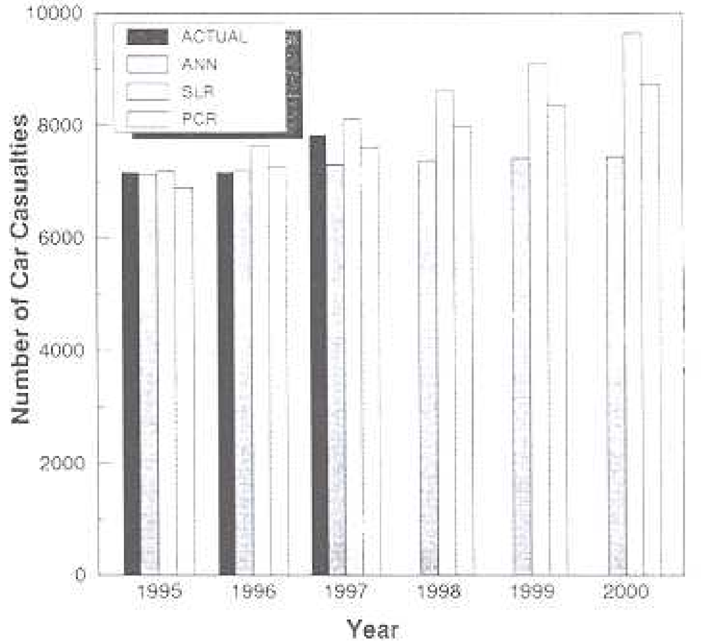

Figure 5. Comparative results of ANN, PCR and SLR predictions (1995 - 2000).

(R^2 = 0.960). However, the differences in R^2 are not significant. It can therefore be said that, for this data set, the regression based models are very much comparable to ANN model in goodness of fit. Figure 4 shows the scattergram depicting the observed against the fitted number of casualties for ANN, PCR and SLR models. The ideal shape in all scattergrams of this type would be a straight line with a gradient of 45° passing through the origin. For the regression models the graph shows marked deviations from the ideal as represented by the straight line. The ANN estimates, on the other hand, are much closer to the line, a reflection of its small MSE and high R^2 values.

The forecasts provided by the three models upto the year 2000 in Table 5 and Figure 5 also differ significantly. While ANN forecasts a very slow growth in the annual number of casualties, the SLR model forecasts high rates of growth. Thus, the ANN predicts the figure to be just below 7,500 by the year 2000, while the SLR model predicts 9,628. On the other hand, the PCR model falls in between (8,735). From the history of the growth of the casualty figures, it would seem ANN predictions are very much to the lower side, while those of the SLR model could be considered to be on the high side. The PCR model predictions do appear to be more reasonable. Examination of the ANN and PCR model predictions against actual records for 1995 and 1996 (Table 5) revealed that while the ANN model predictions of these records were more accurate, the long term predictions by the PCR model are more realistic as demonstrated by the 1997 prediction (7610) and observation (7827). The ANN prediction for 1997 was much lower (7294). This indicates that while ANN models may be suitable for interpolation, their use to extrapolate may be limited by their inability to handle the stochastic effects.

Summary and Conclusions

In this paper the authors have attempted to investigate and compare the predictive capabilities of ANN with multiple linear regression models for annual car accident casualties in Oman. Traffic accidents are among the major causes of death in the Sultanate, and so the need for such investigations contribute to the understanding of the underlying features of the problem and the development of better methods of analysis and assessment of new safety measures.

The main predictor variables used to model car accident casualties were annual population of the country, the annual GDP, the number of registered cars on the road, the number of kilometers of paved roads and the cumulative number of driving licenses issued. Other factors that could also significantly contribute did not have data available.

The response variable (NOAC) was fitted using ANNs. For comparative purposes, the NOAC was also modeled using regression techniques. Preliminary examination of the data indicated that the predictor variables were highly collinear. This suggested the use of principal component regression to fit the data. Further analysis showed that a simple linear regression model also fitted the model adequately after taking the square root of the response variable, and with the number of cars on the road as the only predictor variable. It was found that, on the basis of MSE, the ANN model fit was better than that of either of the regression models. The forecasts obtained from the three models were also different from one another. While ANN forecasts tended to be low, the SLR forecasts were quite high. Forecasts from the PCR model, on the other hand, were in between and appeared the more realistic.

In general, classical multiple linear regression does have a number of limitations, among which are the assumptions about the error term and the functional form for fitting the data. While nonlinearities can be handled in theory by transformation of the data, achieving this in complex problems may be difficult. ANNs, on the other hand, require no assumption about the relationship among the variables, and are therefore amenable to model more complex relationships. However, in an investigation for which an understanding of the relationships among variables are required, the absence of a functional form can be a drawback. Moreover, while ANNs are suitable for interpolation, their use to extrapolate is limited when the empirical data have stochastic components. Nonetheless, the results presented in this paper show that ANNs have the potential to be a viable prediction tool to complement classical statistical methods in the study of traffic accidents and casualties.

References

AL-ALAWI, S., ALI, G.A. and BAKHEIT, C.S. 1996. A Novel Approach for Traffic Accident Analysis and Prediction Using Artificial Neural Networks. *Journal of Road & Transport Research, Australian Research Board,* **5(2)**, pp. 118 - 128.

AL-DAMYAN, A., 1995. Traffic Accidents in Kuwait, Asharq Al-Awsat, the *International Newspaper of Arabs*, No. 5906, January **29**, 1995, p. 15.

AL-SULEIMAN, T. I. and AL-MASAEID, H. R. 1992. Descriptive Model for Fatality Rates of Traffic Accidents in Jordan. *ITE Journal*, **62 (4)**, pp. 37-39.

ALI, G.A., BAKHEIT C.S., and. SIVAKUGAN, N. 1994. Traffic Accidents in Oman: Characteristics and Comparative Analysis of Fatality Rates. *World Safety Journal.* Volume VIII/1-**94**, pp 15-19.

ALI, G.A., MATWANI, T.M. and KOUSHKI, P.A. 1995. Urban Traffic and Safety Management: Case Studies from the Gulf States. *Proceedings of the Second Regional Conference and Exhibition on "Save the Environment", the American Society of Civil Engineers - Saudi Arabian Section (ASCE-SAS)*, Beirut, Lebanon, 16-18 November 1995, Vol. **1**, pp. 327 - 340.

AL-KHALEEJ DAILY NEWSPAPER 1996. Cost of Traffic Accidents in the Kingdom of Saudi Arabia, No. 6102, January 30, 1996.

ASHARQ AL-AWSAT 1994. Traffic Accidents and their Victims in the Arab World. The Arab International Newspaper, Nos. 5827-5833, November 11-17, 1994.

BURDEN, R.L. and FAIRIES, J.D. 1985. *Numerical Analysis*, Prindel, Webber and Schmidt.

CHEN, B. and TITTERINGTON, D. M. 1994. Neural Networks: A Review From a Statistical Perspective, (with discussion). *Statistical Science*. **9**, pp 2-54.

DIRECTORATE GENERAL OF TRAFFIC 1976-1997. General Traffic Statistics. Royal Oman Police, Muscat, Sultanate of Oman.

ELIZANDRO, D.W., STARKS, S.A., LERMA-SANCHEZ, O. and VERN W. 1993. Nonparametric Approach to Classification Using Neural Networks. *Proceedings of SPIE - The International Society for Optical Engineering*. V. 826. Publ. by International Society for Optical Engineering, Bellingham, WA, USA, pp. 114-121.

ERGUN, G. 1988. A Comparative Analysis of Highway Safety Problems in Arabian Gulf, Developing and Developed Countries. *The Arabian Journal of Science and Engineering*, **13(1)**, pp. 3-15.

GAGARIN, N., FLOOD, I. and ALBRECHT, P. 1994. Computing Truck Attributes with Artificial Neural Networks. *Journal of Computing in Civil Engineering*, **8(2)**, pp. 179-200.

GILMORE, J. and ABE, N. 1993. Neural Networks System for Traffic Congestion Forecasting. *Proceedings of the International Joint Conference on Neural Networks*, V. **2**. Publ. by IEEE, IEEE Service Center, Piscataway, NJ, USA, pp. 2025-2028.

GOODMAN, R.M., AMBROSE, B., LATIN, H. and FINNELL, S. 1993. Hybrid Expert System/Neural Network Traffic Advice Systems. *Proceedings of the IFIP TC6/WG6.6 3rd International Symposium on Integrated Network Management.* San Francisco, CA, USA, IFIP Transactions C: Communication Systems NC-12, pp. 607-616.

HECHT-NIELSEN, R. 1989. Theory of Backpropagation Neural Network. *Proceedings of International Joint Conference on Neural Networks*, IEEE Washington, D.C., Vol. 1, pp. 593-605.

JACOBS, G.D. and FOURACRE, P.R. 1977. Further Research on Road Accident Rates in Developing Countries. *TRRL Report SR 270*, Transport and Road Research Laboratory, Crowthorne, Berkshire.

JACOBS, G.D. and HUTCHINSON, P. 1973. A Study of Accident Rates in Developing Countries. *TRRL Report L546*, Transport and Road Research Laboratory, Crowthorne, Berkshire.

JADAAN, K. S. and NICHOLSON, A.J. 1992. Relationships between Road Accidents and Traffic Flows in an Urban Network. *Traffic Engineering and Control*, **33(9)**, pp. 507-511.

JOLLIFFE, I.T. 1986. *Principal Component Analysis. Springer - Verlag*, New York, NY.

LAPEDS, A. and FORBER, R. 1988. How Neural Network Works. *Neural Information Processing Systems.* American Institute of Physics, pp. 442-456.

MYERS, R.H. 1990. *Classical and Modern Regression with*

Applications.(2nd Edition). Duxbury Press, Belmont, CA.

NAKATSUJI, T., SEKI, S. and T. KAKU, T. 1993. Development of Self-Organizing Traffic Control System Using Neural Network Model. *Proceedings of the 4th International Conference on Microcomputers in Transportation*, ASCE, New York, NY, pp. 332-343.

OMAN DAILY NEWSPAPER 1993. Traffic Accident Casualties in the World. No. 4485, 4 Sept. 1993, p. 16.

OMAN DEVELOPMENT COUNCIL 1994, 1996. *Oman Statistical Yearbook*. Ministry of Development, Muscat, Sultanate of Oman.

PATTNAIK, S.B. and SREEDAR, M.A. 1993. Accident Analysis with an Expert System for Remedial Measures-Madras as Case Study. *IE(I) Journal*, **74**, pp. 122-127.

PETRUCELLI, E. 1991. Traffic Injuries: A Public Health Concern Worldwide. *ITE Journal*, **61(7)**, pp. 15-20.

RIPLEY, B.D. 1993, Statistical Aspects of Neural Networks. *Networks and Chaos*. Eds. Berndorff-Nielsen, O.E., Jensen, T. L. and Kendall W.S, pp. 40-123.

RIPLEY, B.D. 1994. Neural Networks and Related Methods for Classification (with discussion). *J. R. Statist. Soc. B*, **56(3)**, pp. 409-456.

RIPLEY, B.D. and HJORT, N.L. 1994. *Pattern Recognition and Neural Networks: a Statistical Approach*. Cambridge University Press, Cambridge.

RUMELHART, D.E. and MCCLELLAND J.L. 1986. Parallel Distribution Processing: *Exploration in the Microstructure of Cognition*. Vol. 1, Foundations, MIT Press.

SIMPSON, P.K. 1990. *Artificial Neural Systems: Foundations, Paradigms, Applications, and Implementations. (1st Edition)*. Pergamon Press Inc., Elmsford, New York, NY.

SMEED, R.J. and. JEFFCOATE, G.O. 1970. Effects of Changes in Motorization in Various Countries on the Number of Road Fatalities. *Traffic Engineering and Control*, **12**, pp. 150-151.

STANLEY, J. 1990. *Introduction to Neural Networks.(3rd Edition)*. Sierra Madre.

STERN, H.S., 1996. Neural Networks in Applied Statistics (with discussion). *Technometrics*, **38(3)**, 205-220.

TABER, J.T. GRENNEY, W.J. and SENTI, T.E. 1995. A Knowledge-base and Simulation Algorithm Hybrid Model For Traffic Congestion At Intersections. In *Developments in Artificial Intelligence for Civil and Structural Engineering*. Ed. B.H.V. Topping, Civil-Comp Press, Edinburgh, UK.

WEISS, S.M. and KULIKOWSKI, C.A. 1991. Computer Systems that Learn: *Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems*, Morgan Kaufman, San Mateo.

WONG-TOI, D. 1994. Macroscopic Models Relating Traffic Volumes and Fatalities-A Case Study of New Zealand and Australia. *Road & Transport Research, Australian Road Research Board*, **3(4)**, pp. 28-40.

WU, J. and HEYDECKER, B.G. 1995. Natural Language Understanding in Road Accident Data Analysis. In *Developments in Artificial Intelligence for Civil and Structural Engineering*. Ed. B.H.V. Topping, Civil-Comp Press, Edinburgh, UK.

Received 6 April 1997
Accepted 14 May 1998