COMPARATIVE ANALYSIS AND PREDICTION OF TRAFFIC ACCIDENTS IN SUDAN USING ARTIFICIAL NEURAL NETWORKS AND STATISTICAL METHODS

DR GALAL A ALI, and *DR CHARLES S. BAKHEIT

Professor & ConsultEng, Sudan University of Science & Technology (SUST), P O Box 12281, Khartoum 11111, Sudan Cell Tel.: +249 912345507 Fax: +249 183 463614
E-mail:ga03ali@yahoo.com galalabdalla@sustech.edu
***Associate Professor, Department of Mathematics and Statistics, College of Science, Sultan Qaboos University, Al-Khod 123, Sultanate of Oman

ABSTRACT

Road traffic accidents (RTAs) are one of the major causes of death in Sudan, notably in the age group of 20 to 40 that constitutes 44% of the population. Fatality rate per 10,000 vehicles is one of the highest in the world, in spite of Sudan's low vehicle-per-capita ratio of 125 persons per car (average value over the last 20 years). Thus, it signifies the importance of properly analyzing traffic accident data and predicting casualties. Such studies will explore the underlying causes of RTAs and thereby develop appropriate safety measures to reduce RTA casualties. In this paper, analysis and prediction of RTAs in Sudan were undertaken using Artificial Neural Networks (ANNs). ANN is a powerful technique that has demonstrated considerable success in analyzing historical data to predict future trends. However, the use of ANNs in the area of traffic engineering and accidents analysis is relatively new and rare. Input variables to ANN model were carefully selected through examining the strength of the correlation between the annual number of accidents and related variables such as annual population growth, gross domestic product, number of driving licenses issued annually, etc. For further validation of the model, principle component regression (PCR) technique was used to fit the same data. Both approaches attempted to model accidents using historical data on related factors, such as population, number of cars on the road and so on, covering the period from 1991 to 2009. Forecasts for the years 2005 to 2012 were made using ANNs and principle component regression method. Analysis using ANNs resulted in the best fit for the data with high R^2. However, both methods provided forecasts that were very similar in values. The study showed that ANNs are more suitable for interpolation than extrapolation. Nevertheless, it demonstrates that ANNs provide a potentially powerful tool in analyzing and forecasting traffic accidents and casualties.

Keywords: accident characteristics and causes; comparative analysis; casualties; fatality rates; safety measures

1 INTRODUCTION

The world is experiencing increased traffic accidents and casualties (fatalities and injuries) particularly in developing countries. Annually, over three-quarter million people are killed while injured and disabled victims in road traffic accidents (RTAs) together surpass 40 million (Ali, 2010). Developing countries alone represent 67% of RTA fatalities in the world although they own only about 11% of the vehicle fleet. The fatality rates per 10^4 vehicles in some African and Asian countries range between 15 and 65 (Ali and Shigidi, 2002). In contrast, in the USA, the 2008 accident records indicated that only 3 out of the 51 states reported fatality increase from 2007 figures, with an average of -10 %. In fact, the fatality rate per 100 million vehicle-miles-traveled (MVMT) reduced from 1.36 to 1.25 (US DOT-NHTSA, 2010 cited in TRB E-NL, May 2010). It is worth mentioning here that due to lack of MVMT data, fatality rate per 10^4 vehicles closely estimates the corresponding best measure of fatality rates (Ali et al., 1994). These alarming statistics underline the importance of continually updating and improving accident records as well as methods of analyzing traffic accident data. Thus, better understanding of traffic accident data and casualties will assist policy makers device better traffic regulations and safety measures to enhance safety.

An overview 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 main causes of accidents were attributed to speeding, driver negligence and violation of traffic regulations, a pattern observed in many countries (Abdel-Aty and Abdel-Wahab, 2003; Ali et al. 1995). In many cases, majority of these occur in urban areas, while in some countries about 40% of the casualties involve pedestrians. In Kuwait, a reduction of up to 15% in total fatalities was observed after installation of traffic cameras (Al-Jassar et al., 2004).

2 TRAFFIC ACCIDENT CHARACTERISTICS IN SUDAN

In Sudan, fatalities and injuries are about 10 times more than in many developed countries despite the current low car ownership of one vehicle per 100 population. More than 60% of the casualties are in the age group of 21-60 years as, shown in Figure 1, the major cause of death for 49% of this age group being RTAs (Ali et al., 2010). The high rate of population growth, the large proportion of young drivers, the dramatic increase in the number of vehicles over recent years compounded with less strict law enforcement, not to mention the very poor road conditions in places, have contributed to the high accident rates. Table 1 summarizes the historical data for Sudan (Directorate General of Traffic, 2010), while Table 2 depicts and compares the various 2004 accident and casualty rates in Sudan with those of 2008 (Directorate General of Traffic, 2010). Table 2 indicates that except for two rates, all other parameter rates of accidents and casualties (fatalities and injuries) have increased. The only rates that dropped were fatalities and injuries rates per 10^3 accidents. This was largely because the increase in fatality from 2004 to 2008 was low compared to that of the accidents (Table 1). This may be attributed to more accidents occurring in urban areas with less severity combined with more improvements of highways. Generally, drivers have been the main cause in about 90 % of accidents primarily in terms of negligence, high speed and poor driving, as shown in Figure 2. This is less than the global value of 95 - 98%. Poor road and vehicle conditions contribute to the remaining 10 %. Various studies in the Arab world and Africa indicated that the main contributors to RTAs are high speeds and the pedestrians (Ref. ?). 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.

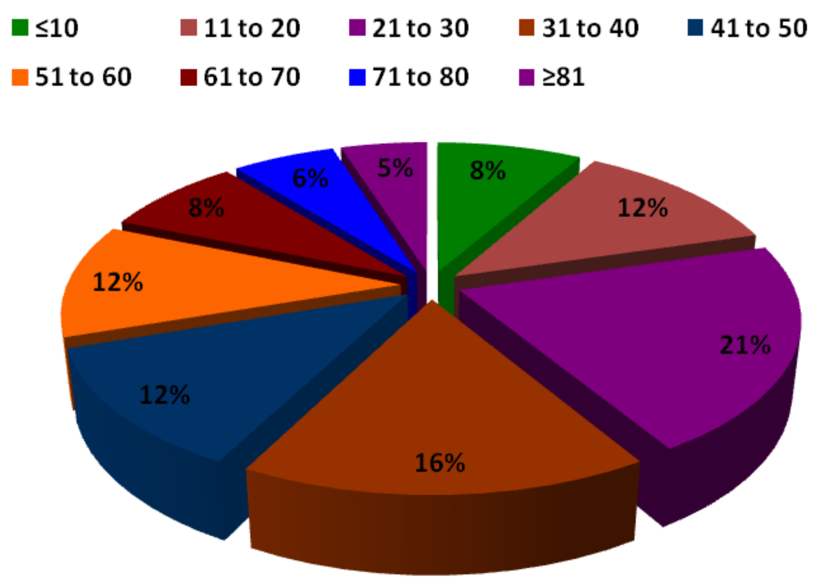

Figure1. Causalities (Fatalities and Injuries) in Sudan by age group (2009)

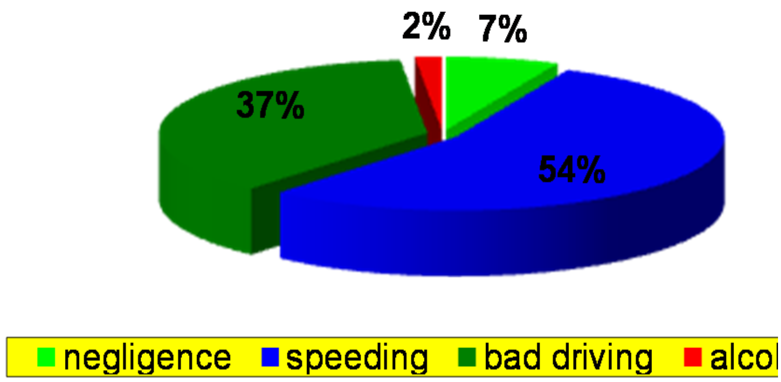

Figure 2. Major driver-related causes of accidents in Sudan (2008)

Table 1. Sudan historical data to determine casualty rates and modeling

Year	Pop*10^3	GDPy*10^3 (SDG)	NOCR (Vehicles)	NOLI (licenses)	NKPR (km)	Injuries	Fatalities	Casualties
1991	24366	18.0	124720	3520	950	2912	395	3307
1992	25040	18.7	151053	3650	1010	4444	413	4857
1993	25734	20.4	172933	3780	1057.2	3543	404	3947
1994	26445	22.1	163053	3865	1126.8	3641	395	4036
1995	27177	22.4	123389	4126	1196.4	4532	580	5112
1996	27489	26.6	164929	4956	1243.3	5245	589	5834
1997	28701	28.5	175722	5125	1298.2	5452	608	6060
1998	29495	29.1	182587	5814	1339.5	5725	715	6440
1999	30281	30.4	190000	7163	1687.9	7113	864	7977
2000	31088	32.6	198000	64686	2036.3	8115	905	9020
2001	31913	35.7	220000	34527	2384.7	8820	952	9772
2002	32592	49.3	259852	49465	2733.1	6207	490	6697
2003	33393	52.9	325897	63077	3081.5	6222	784	7006
2004	34203	71.0	330804	49943	3429.9	5858	651	6509
2005	35009	76.2	364107	25231	3778.3	7198	751	7949
2006	35814	85.9	275350	18338	4126.7	8998	810	9808
2007	36620	97.5	368954	19015	4475.1	10193	870	11063
2008	38242	107.8	395720	27058	4562.3	9307	810	10117
2009	39800	111.0	397318	34226	4740.4	9307	899	10206

Sources: Directorate General of Traffic. Annual Statistical Books 1991-2009, Khartoum Sudan, 2010

Table 2. Accident and casualty rates in Sudan – 2004 and 2008

Accident/Casualty Rates	2004	2008
Accidents per 10^4 vehicles	664	990
Accidents per 10^5 population	64.2	102.4
Accidents per day	60	107
Fatalities per 10^3 accidents	30	21
Fatalities per 10^5 population	1.9	2.1
Fatalities per month	54.3	67.5
Injuries per 10^3 accidents	267	238
Injuries per 10^5 population	17.1	24.3
Injuries per day	16	25

Source: Table 1

Prediction of accidents and casualties today relies on various models and relationships, based largely on expert systems and statistical techniques. They predict accidents and casualties either as a function of registered vehicles per population (Al-Suleiman and Al-

Masaeid, 1992), or traffic flow and road type (Jadaan and Nicholson, 1992). Others used time (Ali et al., 1994) or several exogenous variables (Pattnaik and Sreedar, 1993). Causes of traffic accidents are numerous and complex, and safety may be related to several factors such as road geometry, traffic characteristics, user behaviour and enforcement (Petredou and Moustaki, 2001). Sometimes the developed models may not fit the data (Wong-Toi, 1994), or only address linear relationships between the dependent and independent variables. More comprehensive approaches are required to account for the important variables and their relationships in analyzing and predicting traffic accident casualties. The main objective of this investigation was to predict accidents for Sudan up to the year 2012, applying both ANNs and regression techniques, along with a presentation of a comparative analysis of the results.

3 ARTIFICIAL NEURAL NETWORKS

ANNs are a computer models that are designed to emulate human information processing capabilities such as knowledge processing, prediction and control. The ability of ANN systems to spontaneously learn from examples, to reason over fuzzy data, and to provide adequate responses to new information not previously stored in memory has generated increasing interest in this technology (Lee et al., 2005; Mussone and Oneta, 1999). As a result of numerous applications in various engineering fields, this new technique has gained growing acceptance and demonstrated remarkable success.

ANNs are relatively new in the fields of traffic engineering and accident analysis. This new approach has only been sparsely demonstrated in areas such as traffic congestion forecasting (Taber et al., 1995), determining truck attributes (Gagarin et al., 1994), and a few other applications. Al-Alawi et al. (1996) applied computer-based techniques and artificial neural networks, respectively, to the analysis and prediction of road traffic accidents.

Artificial Neural Networks have also been used in problems that were traditionally solved by statistical methods. A number of researchers have conducted comparative studies of statistical methods with ANNs (Ripley and Hjort, 1994; Stern, 1996). Their studies have shown that, if trained on medium to large data sets, ANNs can be quite useful in prediction, and our date set does satisfy this condition to a certain extent. No assumptions are required concerning the functional form of the relationship between predictor and response variables as the case is with the statistical methods.

3.1 Data preparation and rehabilitation for model building

The crucial factors considered to contribute to annual accident figures included annual population size (POPG), the gross domestic product (GDP) of Sudan, the number of registered cars (NORC), the number of driving licenses issued (NOLI) and the total length of paved roads (NKPR). Other factors could have been incorporated, such as road and vehicle conditions, driver negligence, speeding, or poor driving skills. However, some of these are difficult to quantify while others are poorly documented or have incomplete historical data over the period of study. Data on traffic and accident casualty statistics were readily obtained from Sudan Directorate General of Traffic (DGT), Ministry of Interior (DGT, 2010). Data on population and GDP, and the length of paved roads were obtained, respectively, from the Ministry of Finance and Economic Planning and the Ministry of Roads and Bridges of the Government of Sudan.

After carefully scrutinizing the values of candidate variables, correlation analysis was performed in order to assess the linear association between the variables. Results of the correlation analysis are shown in Table 3. On the strength of the correlation coefficients between the variables the results indicated that the number of kilometers of paved roads (0.96) and the population (0.91) were found to have the highest positive correlation to the number of casualties. Next were the number of registered cars and the gross domestic product. The number of licenses issued was not highly correlated to the number of casualties probably because of confounding with the related factors POPG and NORC

Table 3. (Directorate General of Traffic, 2010) Correlation matrix

Variable	POPG	GDP	NORC	NOLI	NKPR	No. of casualties
POPG	1.0	0.94	0.94	0.996	0.57	0.91
GDP		1.0	0.94	0.38	0.98	0.88
NORC			1.0	0.54	0.96	0.90
NOLI				1.0	0.51	0.31
NKPR					1.0	0.96
No. of Casualties						1.0

3.2. Development of the ANN model

An ANN-based model was developed as shown in Figure 3, for modeling and prediction of the number of casualties in Sudan. The variables selected for developing the ANN model were as mentioned previously, namely: the population growth (POPG), the gross domestic product (GDP), the number of registered cars (NORC), the number of licenses issued (NOLI), and the number of kilometers 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 (NOCA), 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:

$$NOCA_Y = f(POPG_Y, GDP_Y, NORC_Y, NOLI_Y, NKPR_Y) \quad (1)$$

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.

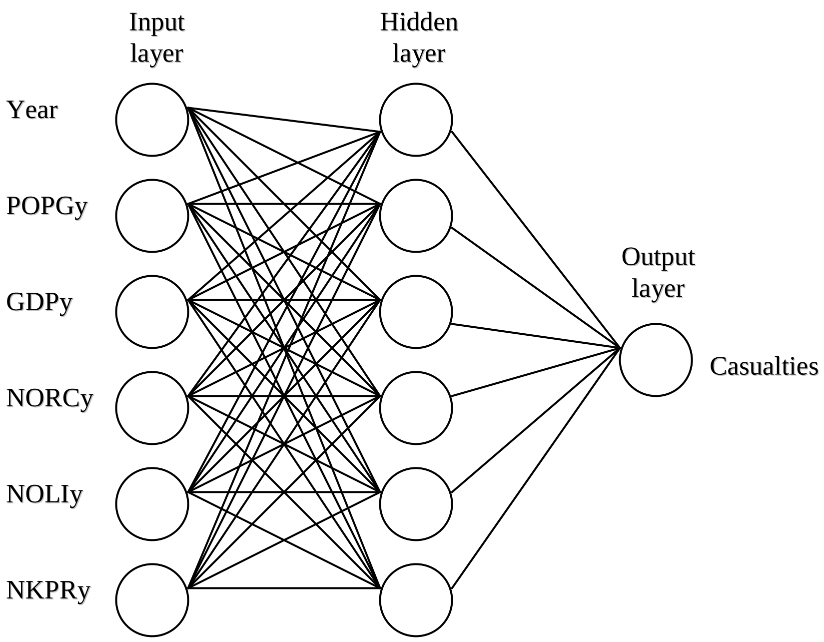

Figure 3. The architecture, input and output of the proposed ANN model

As illustrated by the ANN architecture in Figure 3, 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 accidents. 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. 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 multilayer feed forward networks used in this study were trained using the back propagation (BP) paradigm developed by Rumelhart and McClelland (1986). The BP algorithm uses the supervised training technique. In this technique, the interlayer connection 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 1991 to 1999 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 4. 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. 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 may be found elsewhere (Simpson, 1990).

Table 4. Sample of training patterns used to develop the ANN model

Year	Pop*10^3	GDP*10^3 (SDG)	NORC	NOLI	NKPRy (km)
1991	24366	18	124720	3520	950
1992	25040	18.7	151053	3650	1010
1993	25734	20.4	172933	3780	1057.2
1994	26445	22.1	163053	3865	1126.8
1995	27177	22.4	123389	4126	1196.4
1996	27489	26.6	164929	4956	1243.3
1997	28701	28.5	175722	5125	1298.2
1998	29495	29.12	182587	5814	1339.5
1999	30281	30.41	190000	7163	1687.9

The training process of the ANN models was performed using the NeuroShell® simulator. After thousands of iterations the network converged to a threshold of 0.0001. The high value of R^2 for the model indicates that the variability in the number of car accident casualties (the dependent variable) could be very satisfactorily 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 assess its performance and to examine its generalization capabilities.

3.3. Network testing, validation and prediction

The numbers of accidents predicted by the ANN model were compared with the actual observations. It was found that the model predictions were quite satisfactory (Table 5). In fact, the 2005 prediction of 27,054 accidents differs from the recorded value of 27,712 by less than 2.5 %. Likewise, the 2007 forecast of 37,547 car accidents was even better and differs from the actual observation (37,402) by about 0.4 %.only. Nonetheless, Figure 4 shows that ANN does not appear to smoothen sufficiently the stochastic components of the empirical data but rather attempts to project to higher levels at future years.

Table 5. Model predictions of the number of accidents for the years 2005-2012

Year	2005	2006	2007	2008	2009	2010	2011	2012
Actual Observed	27712	34029	37402	39176	44668	NA	NA	NA
ANN Forecast	27054	35108	37547	40253	41591	47671	53611	59730
PCR Forecast	27995	35427	37292	39872	41094	46738	52236	57901

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®. Examination of these contributions presented in Table 6 revealed that the population growth (POP_Y).and the gross domestic product (GDP_Y) had substantial influence on the increasing number of car accident casualties (55 % and 22 %, respectively). The increase in GDP_Y reflects the country's prosperity and economic growth, resulting in consumer purchasing power of vehicles.

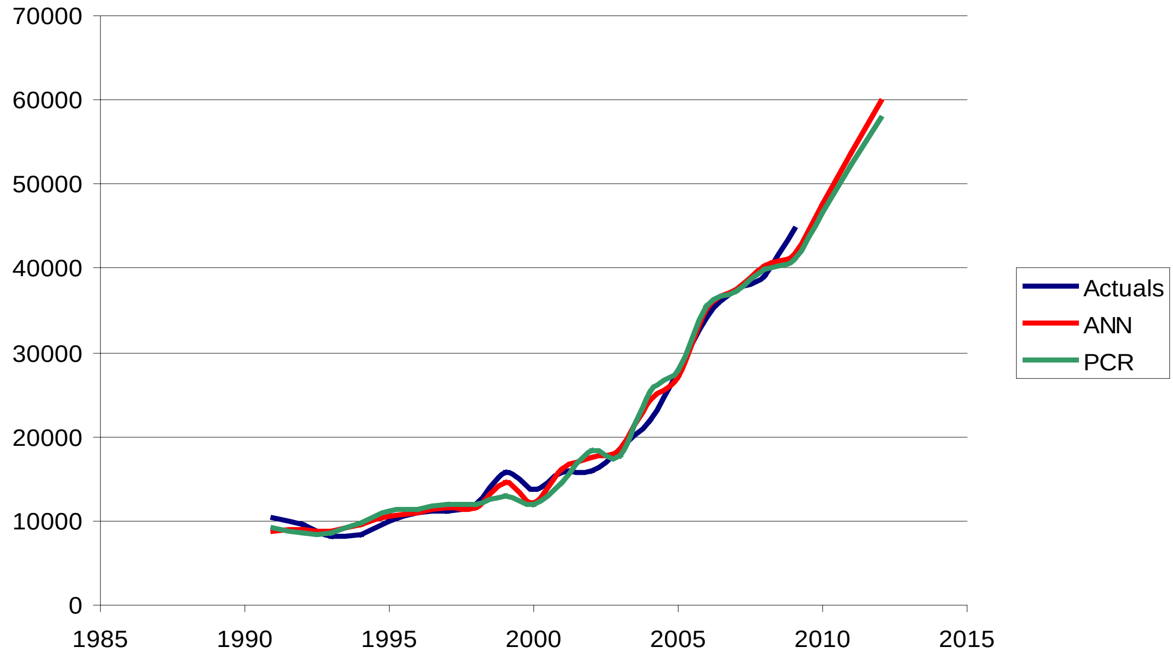

Figure 4. ANN and PCR accidents development and prediction models (1991- 2012)

Table 6. The contribution of the input parameters to the output results

Parameter	$POPG_Y*10^3$	GDP_Y	$NORC_Y$	$NOLI_Y$	$NKPR_Y$
Contribution, %	55	22	13	8	3

-

The number of registered cars ($NORC_Y$) accounted for 13 % of the variation in the number of accidents; while 8 could be explained by the number of licenses issued ($NOLI_Y$) and the length of paved roads ($NKPR_Y$) only marginally (3 %). These contributions were computed by the NeuroShell® utility as measures of the variable input 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 22 % for GDP_Y, 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 NOCA.

4 MULTIPLE REGRESSION AND PRINCIPAL COMPONENT ANALYSIS

Multiple regression analysis is one of the most widely used methodologies for expressing the dependence of a response variable on several independent (predictor) variables. In spite of its evident success in many applications, however, the regression approach can face serious difficulties when the independent variables are correlated with each other (McAdams et al., 2000). Multicolinearity, or high correlation between the independent variables in a regression equation, can make it difficult to correctly identify the most important contributors to a physical process. One method of removing such multicolinearity is to use the multivariate data analysis (MDA) techniques. MDA have been an effective tool in analyzing voluminous data for trends and relationships (Statheropoulos et al., 1998). One such method is principal component analysis (PCA) which has been employed to separate interrelationships into statistically independent basic components (Vaidya et al., 2000). They are equally useful in regression analysis for mitigating the problem of multicollinearity. Essentially, PCA is a special case of factor analysis which transforms the original set of inter-correlated variables into a new set of an equal number of independent uncorrelated variables or principal components (PCs) that are linear combinations of the original variables.

Formally, assuming that there are p original variables, V_i, i = 1, 2,…., p, then the PCs are expressed by the following p linear combinations.

$$\boldsymbol{PC_i = a_{1i}V_1 + a_{2i}V_2 + \ldots\ldots + a_{pi}V_p, \qquad i = 1, 2, \ldots, p,} \qquad (2)$$

where PC_i is the ith principal component and a_{ji} is the loading or correlation coefficient of the original variable V_j and PC_i. In addition, PCA can also be used to identify outliers and filter the data to separate the factors. More details on these and other methods can be found elsewhere (Statheropoulos et al., 1998).

After the predictor variables were transformed, the least squares procedure was applied to obtain a prediction equation. 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 98.9% 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 as follows:

$$\begin{aligned} \text{NOAC} = {} & 4693.77 - 0.0071\ \text{POP}*10^3 + 344.85\ \text{GDP} - 0.0323\ \text{NOCR} \\ & + 2.676\ \text{NKPR} - 0.0422\ \text{NOLI} \end{aligned} \qquad (3)$$

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. Regression models were used to extrapolate casualty figures from 2005 up to the year 2012, using the projections of the predictor variables and the results are shown in Table 5.above.

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 accidents for use to predict future trends. 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). It can be seen that, for this data set, the regression based model is very much comparable to ANN model in goodness of fit. Figure 5 shows the scattergram depicting the observed against the fitted number of casualties for ANN, and PCR model. The ideal shape would be a straight line with a gradient of 45^0 passing through the origin. For the regression model 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.

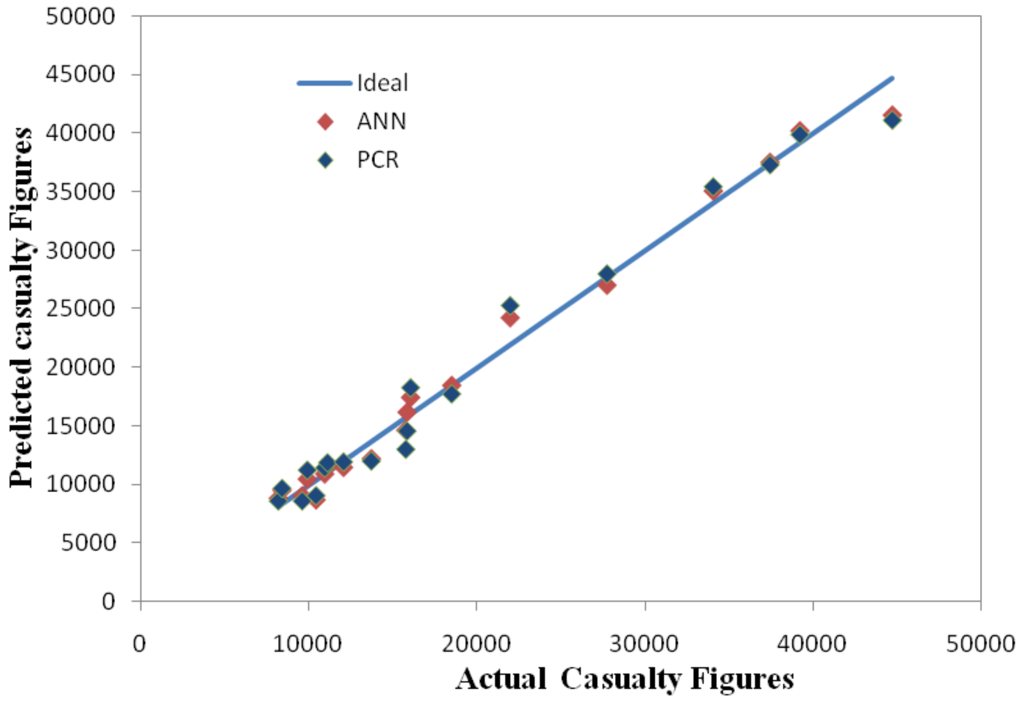

Figure 5 Scattergram Comparison of Prediction By ANN and PCR (19991 -2009)

The forecasts provided by the two models up to the year 2012 in Table 4 and Figure 6 differ only slightly, and compare well with observed data. While ANN forecasts a slightly higher growth in the annual number of accidents, the differences with PCR forecasts are not statistically significant. Thus, the ANN predicts the figure to be very close to 60,000 by the year 2012, the prediction for the PCR model is slightly lower, at around 58,000 (a 3% discrepancy). From the history of the growth of the accident figures, it would seem ANN predictions are relatively much to the higher side.

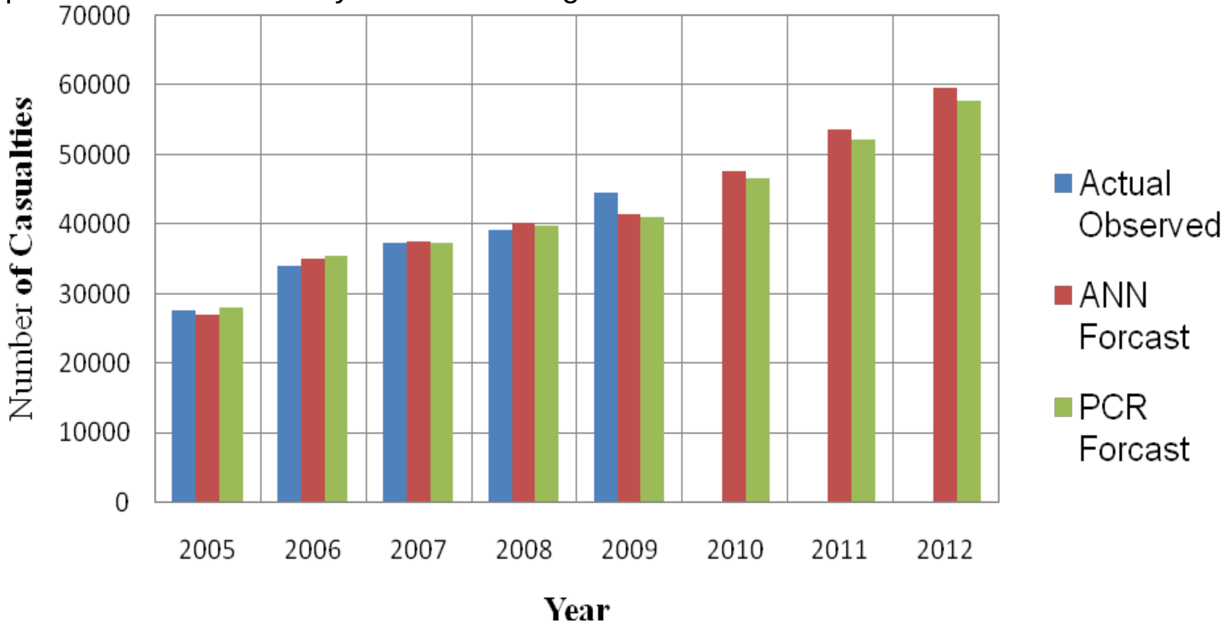

Figure 6. Comparison of ANN and PCR predictions with observed data (2005–2012)

6 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 Sudan where traffic accidents are among the major causes of death in spite of its low motorization level. Thus, the need for such investigations contributes to the understanding of the underlying features of the problem and to the development of better methods of analysis and assessment of new safety measures.

The models used the number of car accident casualties as the dependent variable and the annual population of the country, GDP, the number of registered cars, the length of paved roads and the number of driving licenses issued as the independent variables. The response variable (NOCA) was fitted using ANNs. For comparative purposes, the NOCA 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. It was found that the forecasts obtained from the two models compared favorably with observed data. However, the ANN forecasts tended to be relatively high, the PCR model, on the other hand, were slightly lower and appeared more realistic. Overall, the study shows that as from the year 2000, the annual casualty figures are growing more steeply than in the past. This is possibly due to the country's recent economic boom due to its new oil wealth, which has meant more vehicles on the roads, more paved roads and higher GDP.

REFERENCES

Abdel Aty, M and Abdel-Wahab, H, 2003. Analysis and prediction of traffic fatalities resulting from angle collisions including the effect of vehicles' configuration and compatibility. *Accidents Analysis and Prevention, 2003*

Al-Alawi, S, Ali, G and Bakheit, CS, 1996. A novel approach for traffic accident analysis and prediction using Artificial Neural Network. *Journal of Road & Transport Research, Australian Research Board,* Vol 5, No 2, June 1996, pp. 118 - 128.

Ali, Galal A., "Traffic Accidents and Road Safety Management: A Comparative Analysis and Evaluation in Industrial, Developing and Rich-Developing Countries", the 29th Southern African Transport Conference (SATC 2010), Pretoria, South Africa, 16 - 19 August 2010, Proc website: www.satc2010.za.co, pp. 530-540,

Ali, G.A**.,** Bakheit CS 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, Galal A, Bakri, M and Abdalla, M, Aug 2010. Analysis and prediction of traffic accident casualties using statistical methods and ANN, unpublished Research Report, Sudan University of Science and Technology, Khartoum, Sudan, 73 pp.

Ali, Galal A. and Shigidi, A.M.T. "A Comparative Analysis of Traffic Accidents and Road Safety Management," SORIC' 02 (Safety on Roads: International Conference), 21-23 October 2002, Manama, Bahrain

Al-Suleiman, TI and Al-Masaeid, HR, 1992. Descriptive model for fatality rates of traffic accidents in Jordan. *ITE Journal*, 62(4), pp. 37-39.

Aljassar, AH, Ali, MA and Al-Anzi, MS, 2004. An investigation on the effect of traffic cameras on road safety in Kuwait, 2nd Gulf Road Conf Paper No. E5 AD, UAE,

Directorate General of Traffic, (2010). Annual Statistics Books 1991-2009. Ministry of Interior, Khartoum, Sudan

DOT- NHTSA and TRB, 2010. 2008 Accidents data summary. US Department of Transportation, cited in TRBE-Newsletter, TRB@newsletters.nas.edu, May 2010

Gagarin, N, Flood, I. and Albrecht, P, 1994. Computing truck attributes with Artificial Neural Networks. *Journal of Computing in Civil Engineering*, 8(2), p 179.

,

Hecht-Nielsen, R, 1989. Theory of Back-propagation Neural Network. *Proc of Int Joint Conf on Neural Networks,* IEEE*.* Washington, D.C., Vol. 1, pp. 593-605.

Jadaan, KS. and Nicholson, AJ, 1992. Relationships between road accidents and traffic flows in an urban network. *Traffic Eng and Control,* 33(9), pp. 507-511.

Lee, S, Lee, T, Kim, HJ and Lee, YK, 2005. Development of accident prediction model with intel system theory. Gervasi et al.(Eds) ICCSA, LNCS 3481, pp. 880-888

Mussone, L., Ferrari, A., & Oneta, M., An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention*, Vol. 31, 1999, pp. 705-718.

Pattnaik, SB and Sreedar, MA. 1993. Accident analysis with an expert system for remedial measures. *IE(I) Journal*, 74, pp. 122-127.

Petredou, E and Moustaki, M, 2001. Human factors in the causation of road traffic crashes. *European J of Epidemio*, **16:** 819-826, Kluwer Acad Pub, the Netherlands

Ripley, BD. and Hjort, NL,. 1994., Pattern recognition and Neural Networks: A Statistical Approach*.* Cambridge University Press, Cambridge.

Rumelhart, DE and Mcclelland JL,1986. Parallel distribution processing: exploration in the microstructure of cognition*. Vol. 1, Foundations,* MIT Press.

Simpson, PK, 1990. Artificial Neural Systems. *Foundations Paradigms, Applications, and Implementations. 1st Ed.* Pergamon Press Inc., Elmsford, NY.

Stern, HS, 1996. Neural Networks in applied statistics (with discussion), *Technometrics,* 38 (3), 205 - 220.

Taber, JT, Grenney, WJ and Senti, TE.. 1995. A knowledge-base and simulation algorithm hybrid model for traffic congestion at intersections.. *Developments in Artificial Intelligence for Civil & Struc Eng.* Ed. Topping, Civil-Comp Press, Edin, UK.

Wong-Toi, D, 1994. Macroscopic models relating traffic volumes and fatalities - A case study of NZ & Australia. *Road & Transport Research, ARRB*, 3(4), pp. 28-40.