International Journal of the Physical Sciences Vol. 6(2), pp. 244-254, 18 January, 2011
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS10.606
ISSN 1992 - 1950 ©2011 Academic Journals

Full Length Research Paper

Modeling of monthly traffic accidents with the artificial neural network method

Halim Ferit Bayata[1]*, Fatih Hattatoglu[2] and Neslihan Karsli[3]

[1]Vocational High School, Erzincan University, Erzincan, Turkey.
[2]Civil Department, Vocational of High School, Ataturk University, Erzurum, Turkey.
[3]General Directorate of State Hydraulic Works, DSI 8 Regional Directorate, Erzurum, Turkey.

Accepted 04 January, 2011

Averagely, over 1.3 million traffic accidents with deaths, wounds and property losses occur in a year in Turkey. In the last 30 years, 140000 people have died in traffic accidents. The financial losses of the accidents in the last ten years are over 100 billion dollars. In this study, monthly accidents occurring in Turkey were investigated with Artificial Neural Network (ANN) method. With monthly accident data occurring between the years 1994 to 2007, seasonal fluctuations in the months to come and the number of accidents were predicted. Therefore, an attention is attracted to accidents problems, more reliable analyses about the number of traffic accidents are made and it is emphasized that precautions are increased to prevent the accidents. What sort of the accidents trend will follow in the next months was cleared in the ANN model. The best model was chosen according to the Akaike Information Criteria (AIC), Mean Squared Error and the square of correlation of coefficient (R^2). The reliability of the model was also tested by statistical analyses on the error terms.

Key words: Traffic accident analyses, Artificial Neural Network, statistical analyses.

INTRODUCTION

The cause that affects traffic accidents can be a factor or combination of many factors. The basic factors, which cause the severity of probable accidents, are driver's behavior, vehicle features, highway characteristics, environmental effects and traffic characteristics (Ozgan, 2003). The costs of fatalities and injuries since traffic accidents have an important impact on society. In the world, 3.840 people die in a day because of a simple urban activity that is transportation (WHO-World Health Organization, 2004). In Turkey, 4.228 people were died and 183.841 people were injured since traffic accidents, in the year 2008. The material loss of the traffic accidents to the national economy is about 350.000 TL in a year (TÜİK, Turkey Foundation of Statistics). In recent years, researchers have been utilizing real-life data in studying various aspects of driver injuries resulting from traffic accidents. Related important works can be summarized as follows. Delen et al. (2005) modeled the role of driver's characteristics, environmental factors and road conditions during a traffic accident with the method of Artificial Neural Network (ANN). Ferrer et al. (2006) modeled with analyses of time series and traffic accidents in their study.

Iyinam and Ogut (1998) modeled monthly traffic accidents by using a 20-year-old data and one variable time series analysis. Quddus (2008) explained that methods of Poisson and Negative Binominal (NB) regression are mostly used in accidents analysis. In this study, in the applications of accident analyses with the method of Box-Jenkins (ARIMA), seasonal time series (SARIMA), Poisson and NB were used. In the study of Altun and Dündar (2005) modeling of traffic current control, variable such as characteristic features of the drivers (age, sex, alcoholic condition, seatbelt), type of the vehicle, speed, conditions of the region and light take part in input layer. For output layer, it was stated whether the accident is with no damage, slightly damaged, medium damaged, heavily, deadly or not. Data were modeled by ANN, fuzzy logic and probit model and these methods were compared.

*Corresponding author. E-mail: hfbayata@atauni.edu.tr, hfbayata@hotmail.com. Tel: +90 (446) 226 66 06. Fax: +90 (446) 226 66 01.

Table 1. Statistical features of monthly accident numbers between 1994 to 2007

Variable Q	January	February	March	April	May	June	July	August	September	October	November	December
Min. value	19801	17318	21020	16406	17792	16564	19441	20047	19937	21845	21594	22038
Max. value	55175	51053	56139	56282	60098	62494	64523	65276	68934	62594	66788	69520
Average	35970.92	32483.21	34441.89	33339.2	35678.85	36378.39	37455.71	37989.14	39373.5	41059.25	41257.036	43558.714
Standard deviation	10011.55	8979.7585	9864.5787	10627.22	11272.186	11898.075	11926.673	12155.05	13134.945	12061.538	12397.642	12388.54
Skewness	0.21267	0.250145	0.912305	0.80375	0.728095	0.68942	0.913045	0.88075	0.942431	0.488124	0.57653	0.472504
Kurtosis	-0.022866	0.613978	0.587149	0.737508	0.708608	0.922003	1.000668	0.914874	0.898790	-0.344502	0.172605	0.80333

In this study, the investigation of the change in the seasonal accidents for modeling monthly accidents and formation of the most appropriate model by ANN were aimed. Additionally, ANN was successfully performed based modeling of monthly traffic accidents in terms of number of accidents, injuries and dead.

MATERIALS AND METHODS

Data

Models were based on monthly accident statistics taken by Turkish Statistical Institute (TSI) between 1994 to 2007. The detailed qualified information taken from Director General of Police Traffic Educational and Investigations Department was compared to data of TSI. Statistics of monthly accident numbers between 1994 to 2007 were charted in Table 1.

Method

ANN method is generally used for modeling the variables which are in nonlinear relation and good results are obtained (Bayata, 2010; Kisi, 2009; Firat et al., 2009). Since the variables used in this study are in nonlinear relation, ANN method is preferred. In the application of ANN, a program code was written by using Matlab software. In the program, a software was developed being able to do cycling among transferring functions (tansig-purelin-logsig), training functions (trainbr-trainlm) and in the number of neuron (1,2,3,4…..) in hidden layer. Other features of this developed software are that the networks's taking the Mean Squared Error (MSE) as a criterion of determination performance, it is being able to change iteration number dependent on desires and the network's putting an end to its training as it is desired.

In the structures of alternative network input layer and neuron number in hidden layer, output layer, training function, transferring functions among the layers, R^2, MSE and Akaike Information Criteria (AIC) values were provided with its being able to be read in Microsoft Excel. Therefore, the most suitable network structure can be easily determined.

ANN

ANN can be defined as a system designed to model the method to fulfill as a function of brain. An ANN is formed by being connected in various shapes of artificial neural cells with each other. ANN has an ability of gathering information, saving and generalization this information with connection weights between the cells after passing learning algorithm and learning process (Saraç, 2004; Bayata and Hınıslıoğlu, 2009). Generally, an artificial cell model consists of five components. These are; input, weights, transfer function, activation function and output (Figure 1). Features and superior sides of ANN are being nonlinear, parallelism, easiness of being reality, processing of local information, error tolerance, generalization, adaptation, hardware speed, learning, analysis and easiness of design (Öztemel, 2006). Outputs of network are calculated by spending from an activation function of input data (Equation 1).

$$v = \sum x_i w_i + \theta, \qquad y = F(v) \tag{1}$$

Where,
w: Weight matrices of cell,
x: Input vector cell,
v: Net input cell,
y: Output cell.

According to the features of transfer functions, data used in ANN must be calculated to a defined scale. Therefore, there must be minimum and maximum values existing in data set and must be dependent on according to the scales given below (Equation 2).

$$Y_{new} = \frac{Y - y_{\min}}{y_{\max} - y_{\min}} \tag{2}$$

Where,
Y: original series
Y_{min}: The minimum value of the original series
Y_{max}: The maximum value of the original series

At the time of training process, normalized input and output values are used. At the end of the process, change must be provided to real values by making a contrast change. One of the most important elements effecting learning algorithm is performing function. The most common function is MSE (Equation 3) (Sağıroğlu et al., 2003).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(t_i - td_i)^2 \tag{3}$$

Where,
N: Number of outlayer neurons, which is equal to the

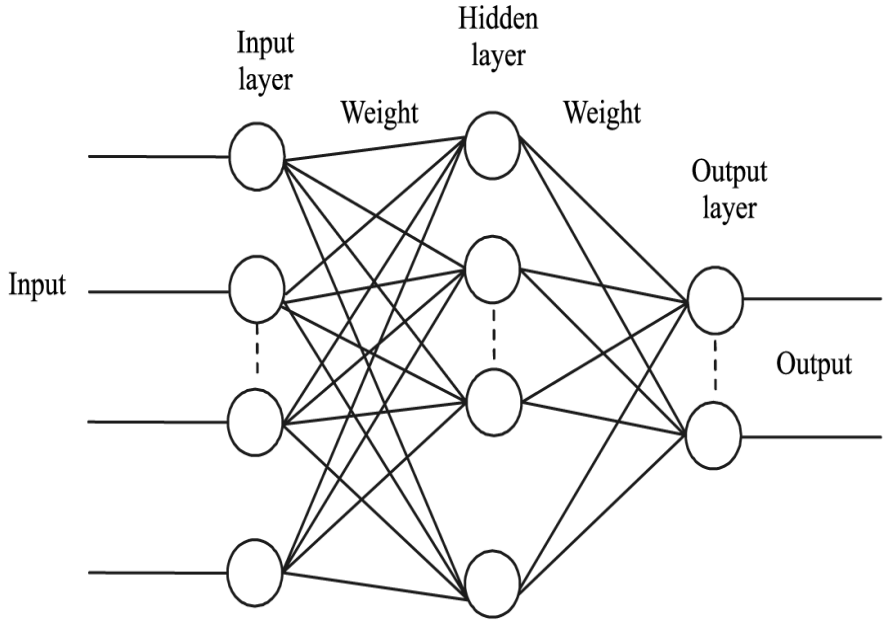

Figure 1. Network Architecture for ANN.

number of target vector elements,
t_i: *i*th element of target vector t,
td_i: *i*th element of output vector td.

Learning algorithm in ANN

Learning algorithm consists of the main frame of an ANN. In the many learning algorithm, the most common ones are trainlm and trainbr. These two algorithms are called as Levenberg – Marguardt. Tainlm, which has an algorithm that reduce memory requirements, is used when a set of learning is very large. This algorithm is the fastest of modern learning algorithm. However, trainbr is a form of enhanced Tainlm algorithm. Trainbr algorithm is a learning method regulating bayessian. This method is a learning algorithm developed to improve the generalization skill of Levenberg – Marguardt. Trainbr algorithm partly reduces the problem of how much the structure of optimum network must be (Sağıroğlu et al., 2003).

Application of ANN

The positive values of skewness coefficient of monthly results show that data are skewed to the right side. A symmetric distribution is out of question. Before starting the analysis of the data, Kolmogorov-Smirnovnormalization test was done and it was realized that the series was not suitable to normal distribution. By being applied of Y= Log (Q) change, test was tried again and it was realized that the data were suitable to normal distribution. To solve periods in lognormal series, it was turned into dimensionless condition with Equation 4.

$$Z_t = \frac{(y(t)_m - \overline{\mu_m})}{\sigma_m} \quad (4)$$

Where,

$Y(t)_m$: Logaritmed monthly accidents number for *m*th month
μ_m : Average of accident number for *m*th month
σ_m : Standard Deviation of accident number for *m*th month

The architecture of example network was presented in Figure 2 to ANN model dependent on time of the number of monthly traffic accidents occurring in 1994 to 2007. By using the numbers of different input neuron in this study, 6 different methods were applied. The results related to these applications were showed in Table 2. Simple scatter and harmonized graphics related to network architectures which are different input neurons were presented in Figures 3 to 8.

RESULTS AND DISCUSSION

In this study, the prediction of the accidents occurring in coming months with ANN method was aimed by using accident data in recent months. When compared 6 models for ANN analysis, the best model is 2 neurons in input layer model which has minimum MSE value, maximum R^2 value and AIC. In this study, the neuron number 6 in the hidden layer and the activation function trainbr and network structure for MSE's (0.004) was accepted as the best network. At the end of the application, ANN -2 (two cells input model) model was determined as the most significant model as statistical. With monthly accident data occurring between the years 1994 to 2007, seasonal fluctuation in the months to come and the number of accidents were predicted. Therefore, an attention is attracted to accidents problems, more reliable analyses about the number of traffic accidents are made and it is emphasized that precautions are increased to prevent the accidents.

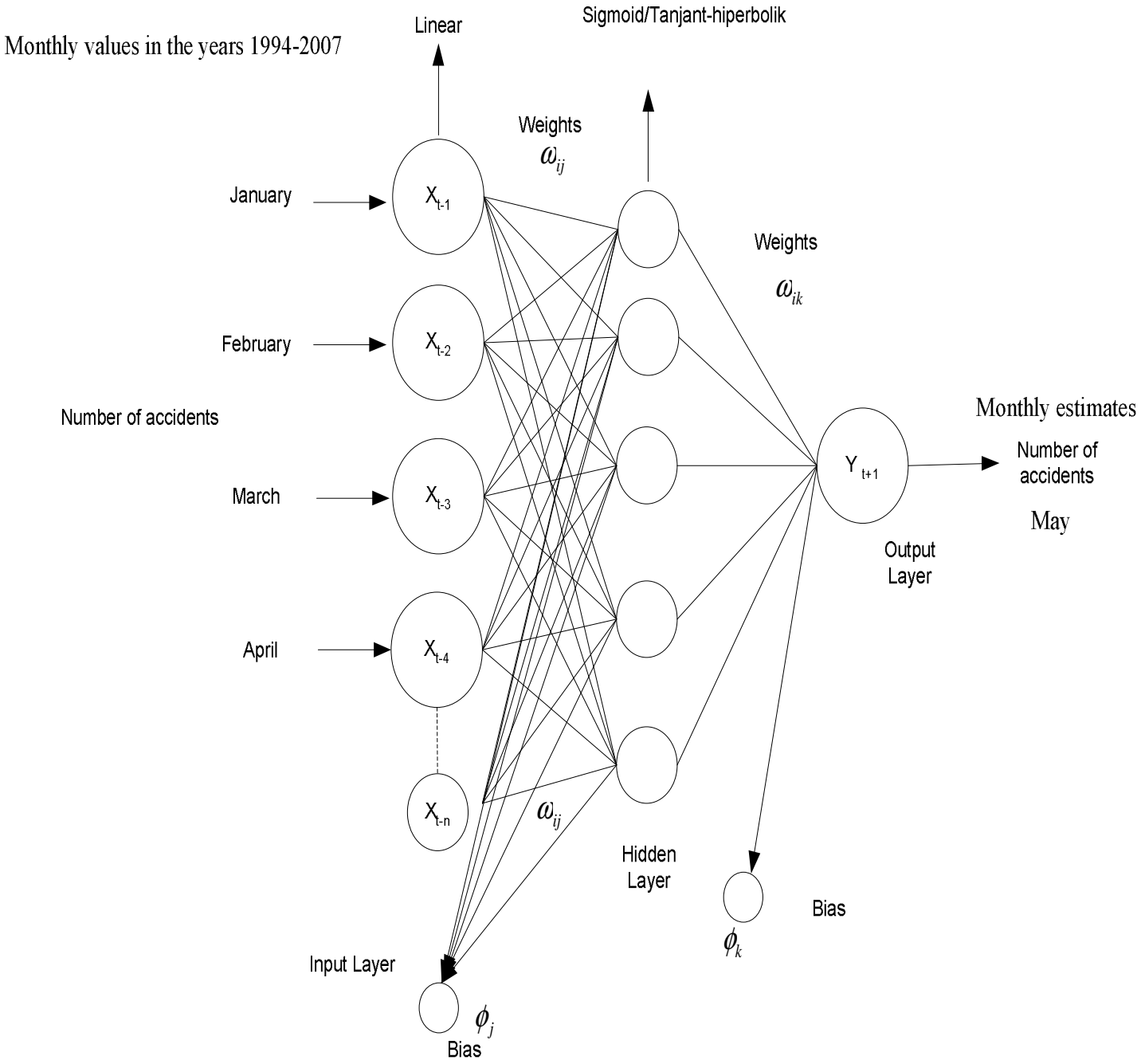

Figure 2. Network architecture sample used in this study.

Table 2. ANN application results.

Input layer neurons	Input layer neuron numbers	AIC	MSE	R^2
Q_{t-1}	1,2,3,4,5,6	-2.31	0.512	0.96
Q_{t-1} Q_{t-2}	1,2,3,4,5,6	-8.81	0.004	0.98
Q_{t-1} Q_{t-2} Q_{t-3}	1,2,3,4,5,6	-2.86	0.049	0.97
Q_{t-1} Q_{t-2} Q_{t-3} Q_{t-4}	1,2,3,4,5,6	-3.12	0.014	0.99
Q_{t-1} Q_{t-2} Q_{t-3} Q_{t-4} Q_{t-5}	1,2,3,4,5,6	-3.30	0.043	0.97
Q_{t-1} Q_{t-2} Q_{t-3} Q_{t-4} Q_{t-5} Q_{t-6}	1,2,3,4,5,6	-2.99	0.013	0.99

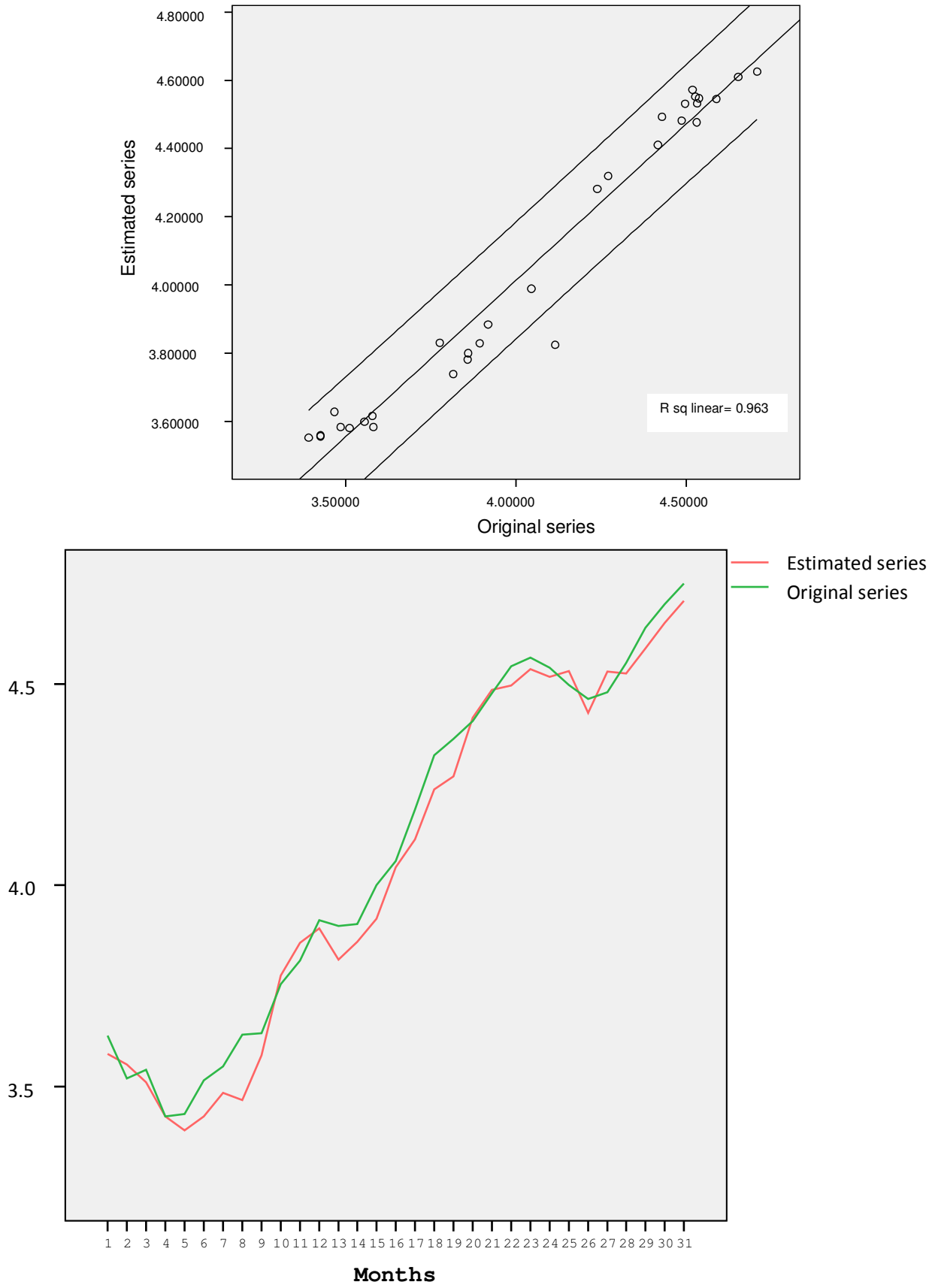

Figure 3. One cell input model results (ANN-1).

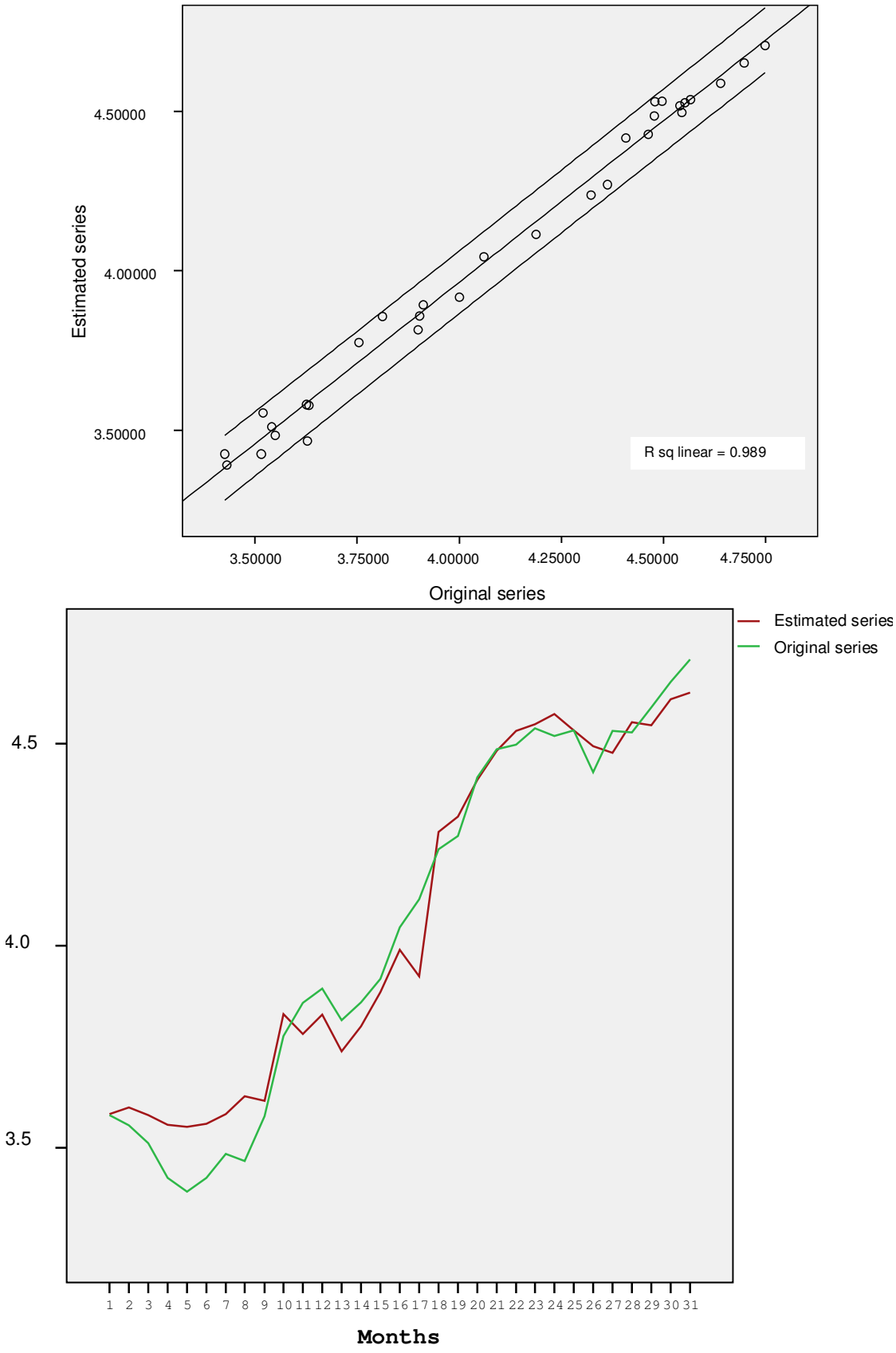

Figure 4. Two cells input model results (ANN-2).

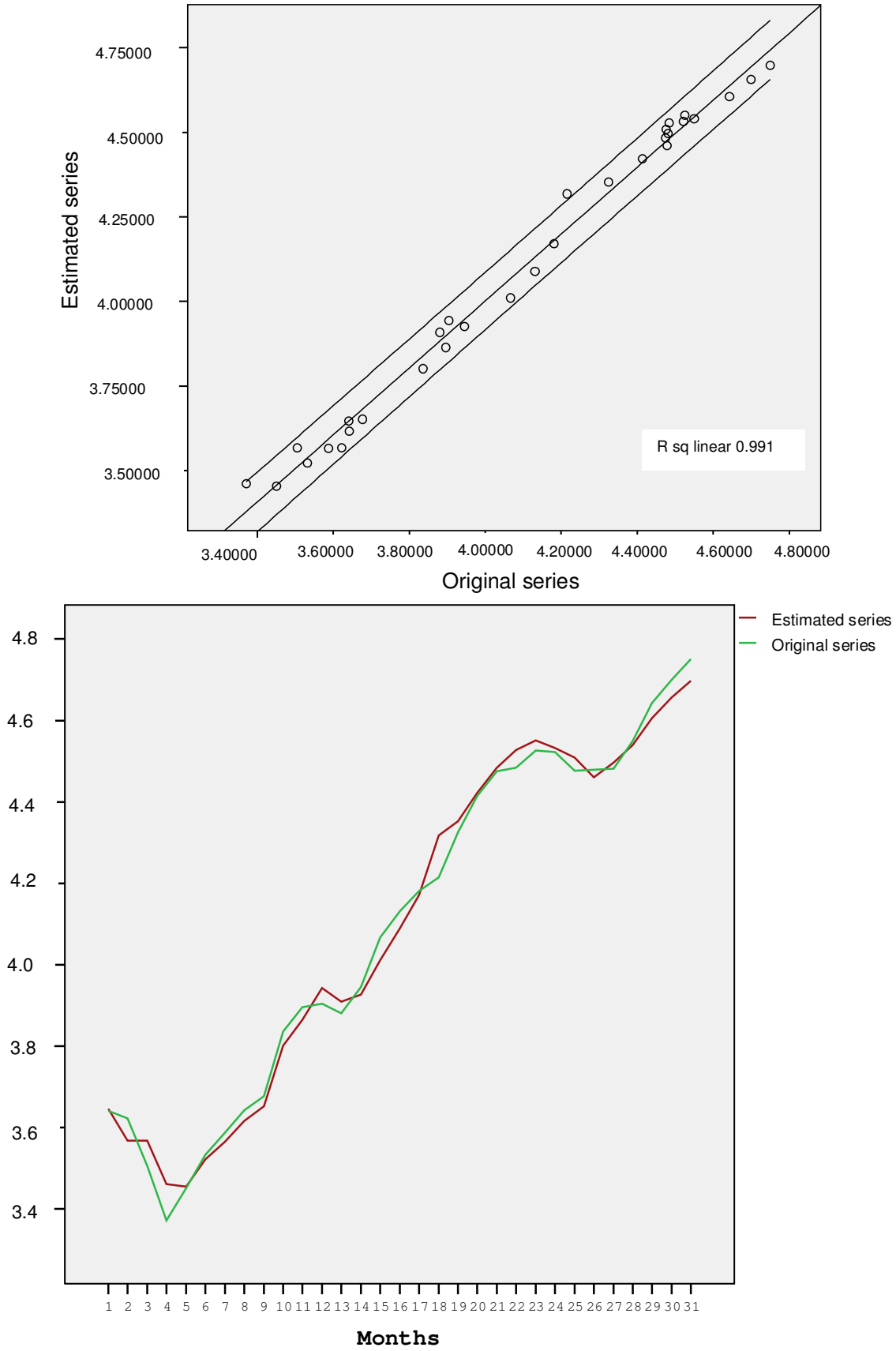

Figure 5. Three cells input model results (ANN-3).

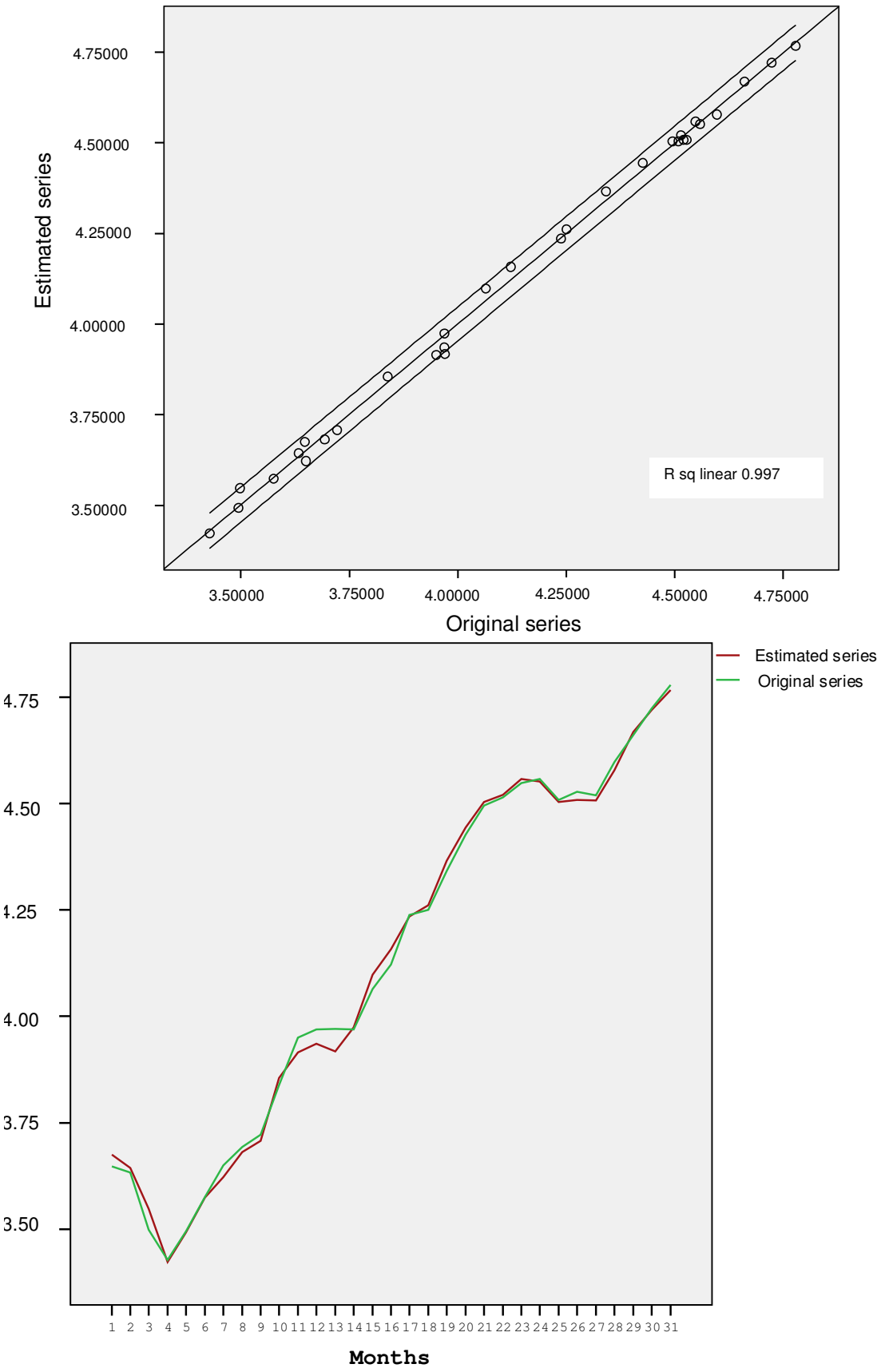

Figure 6. Four cells input model results (ANN-4).

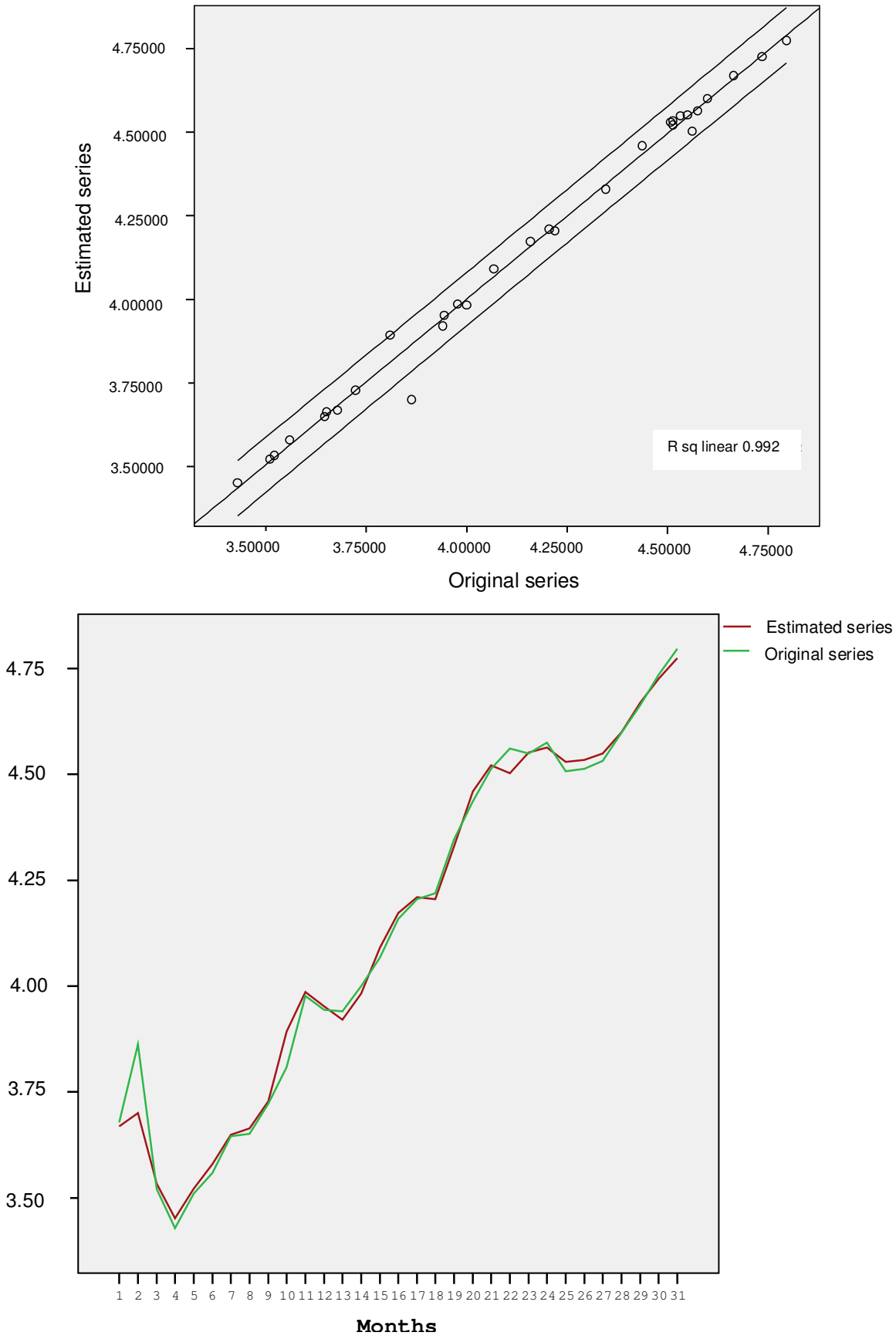

Figure 7. Five cells input model results (ANN-5).

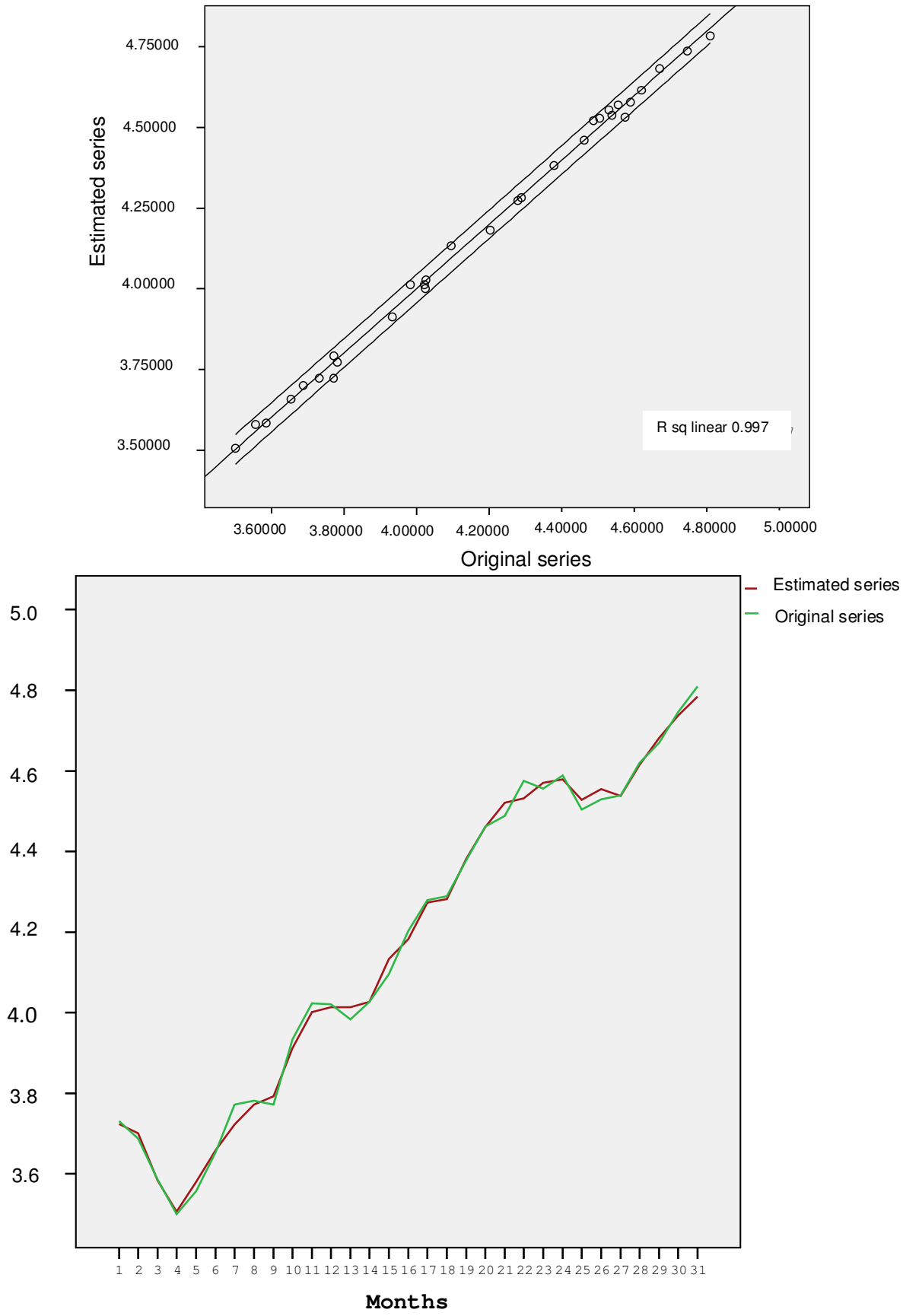

Figure 8. Six cells input model results (ANN-6).

Conclusions

In the last 30 years, 140000 people have died in traffic accidents. The material loss of the traffic accidents to the national economy is about 350.000 TL in a year. The basic factors, which cause the severity of probable accidents, are driver's behavior, vehicle features, highway characteristics, environmental effects and traffic characteristics. In this study, monthly accidents occurring in Turkey were investigated with Artificial Neural Network (ANN) method. By using the numbers of different input neuron in this study, 6 different methods were applied. Additionally, Artificial Neural Network (ANN) was successfully performed based modeling of monthly traffic accidents in terms of number of accidents, injuries and deaths. Consequently, although the accident characteristic with months is quite complex, the proposed ANN approach has been able to model them.

REFERENCES

Altun İ, Dündar S (2005). Traffic flow control with Artifical Neural Network. Earthquake Symposium. Kocaeli.

Bayata HF (2010). Modeling of traffic accidents with multi variable statistical methods. PhD thesis. Ataturk University. Erzurum, Turkey.

Bayata HF, Hinislioğlu S (2009). Modeling of monthly accident numbers in Turkey by Box-Jenkins method, 10.Econometric and Statistics Symposium, Erzurum, Turkey.

Delen D, Sharda R, Bessonov M (2005). Identifying significant predictors of injury severty in traffic accidents using a series of artifical neural ntworks. Accid. Anal. Prev., 38: 434-444.

Ferrer A, Juan G, Poncela P (2006). Forecasting Traffic Accidents Using Diaggregated Data. Int. J. Forecast., 22: 203-222.

Firat M, Yurdusev MA, Turan ME (2009). Evaluation of Artificial Neural Network Techniques for Municipal Water Consumption Modeling. Water Resour. Manage., 23(4): 617-632.

Iyinam FA, Ogüt SK (1998). Modeling of accident numbers in Turkey, İTÜ.

Kisi O (2009). Neural networks and wavelet conjunction model for intermittent streamflow forecasting. ASCE J. Hydrol. Eng., 14(8): 773-782.

Ozgan E (2003). Multi-variate clinical aspects of traffic accidents at Sivas environmental state highway. PhD thesis, Gazi University, Turkey.

Öztemel E (2006). Artificial neural networks, Papatya Publications, İstanbul.

Quddus AM (2008). Time Series Count Models: An Empirical Applicaiton To Traffic Accidents. Accid. Anal. Prev., 40: 1732-1741.

Sağıroğlu Ş, Beşdok E, Erler M (2003). Fuzzy logic applications in Engineering- 1. Artificial Neural Networks. Ufuk yayın. Kayseri.

Saraç T (2004). Artificial Neural Networks. Seminar Study. Gazi University.

WHO (World Report on Road Traffic Injury Prevention) (2004). TUİK, www.tuik.gov.tr (2008-2009).