Mechanics Transport Communications

ISSN 1312-3823 (print)
ISSN 2367-6620 (online)
volume 16, issue 3/1, 2018

Academic journal http://www.mtc-aj.com article № 1644

TRAFFIC VOLUME FORECAST USING REGRESSION ANALYSIS AND ARTIFICIAL NEURAL NETWORK BASED ON PRINCIPAL COMPONENTS

Ramadan Duraku[1], Vaska Atanasova[2]
ramadan.duraku@uni-pr.edu, vaska.atanasova@tfb.uklo.edu.mk

[1]University of Prishtina "Hasan Prishtina" Faculty of Mechanical Engineering
Sunny Hill nn, 10 000 Prishtina, KOSOVO
[2]University "St.Kliment Ohridski", Faculty of Technical Sciences
Makedonska Falanga 33, Bitola, MACEDONIA

Key words: *traffic volume, model, forecasting, principal component, multiple regression analysis, artificial neural network.*

Abstract: *Aim of this Paper is to explore which factors have the greatest impact on generating traffic volume and to set up a suitable model for forecasting it for the main road network of Anamorava region. In this regard, several demographic and social economic variables were taken into account for the period 2004-2016. Variability between variables is done through correlative analysis (multi-co linearity problem) between them. After this, models were developed using original data via multiple regression analysis (MLR) and artificial neural network (ANN) methods. However, with the aim of involving as many variables as possible and eliminating the multi co-linearity respectively improving forecasting model capability, a new methodology was used using the principal component analysis (PCA) as an input. These models then are compared based on prediction errors also verified on the basis of some performance indicators. Results show that the use of principal components as inputs improved both models forecasting by reducing their complexity and eliminating data co-linearity.*

1. INTRODUCTION

Sustainable transport planning begins with traffic volume forecasting for the future period. In the period 2004-2016, traffic demand on the main road network of Anamorava region marked an average annual increase of 4.13 % [1]. This as such is key element to be taken in account in achieving appropriate planning as well as in decision making for investments by competent authorities in transport sector to get an impact in establishing a balance between demands and supply [2]. Nowadays, traffic is one of essential elements for economic and social development in any modern country [3]. Traffic demand increase is essentially related to demographic and social economic development [4]. Thus, traffic forecasting is imperative for better functioning as well as for a reliable transport system and for functioning of economy as a whole [5]. Underestimation of traffic demand would for sure constitute possible limitations to life and economy, while over estimation would result in increase of unnecessary capacities and as such would create unjustified financial losses. Therefore, traffic demand forecasting is crucial to policy makers of competent authorities in order to eliminate unnecessary expenses. In this regard, the period and long term forecasting accuracy play considerable effects in road infrastructure planning, respectively in constructing new roads with higher capacities to cope with future traffic demand [6]. In order to achieve this objective there is necessity of being familiar with the topic as well as in application of various techniques and methods of forecasting. This paper contains some conventional and non-conventional methods for traffic demand forecasting which will be addressed accordingly.

1.1 Study Area

This paper is about traffic demand in the main road network of Anamorava region. The region as such is situated in Kosova valley and from the point of view of therritory and administrative organization is one of seven regions of Kosova, covering six municipalities (Gjilan, Partesh,Vitia, Kllokot, Kamenica and Ranillug) with total 166 settlements and with the surface of 1331 km^2 [7]. It surrounds with other regions of Kosova which is with Prishtina at north-west, Ferizaj at the south, with south-west of Serbia (Medvegja, Bujanoc and Presheva municipalities) and with Macedonia at the south . Nowadays, in this region there is significant number of main, regional and local roads. The study is conducted on the main road network, more specifically in four locations in which automatic traffic counting (ATC) were installed, which is: 1-Slivovo, 2-Sojevo, 3-Ranillug and 4-Pasjan as presented through Fig 1.

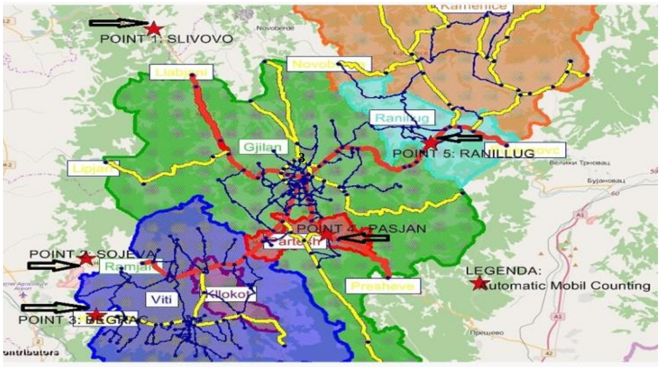

Fig.1. Study area-location of ATC in Anamorava region

2. METHODOLOGY

The main objective of this paper is to identify variables and to assess suitable models according to various methods. Apart from this, the study will also test and assess relation between dependent variable (Y_i) which means the traffic volume expressed like AADT and independent variable (X_{in}) which will enable identification of demographic and social economic variables on region and country level. Primary data are gained by secondary data, which are mainly taken from annual reports and statistics published by competent authorities responsible for recording, structuring and maintaining of them [8]. Data for variables are presented in time format for the period of thirteen years (2004 – 2016), in which relation between them resulted is setting up some models according to various methods and approaches. The study is relied on the use of deductive method and is based on two approaches. In first place, according to aggregate approach the analysis of factors important to generate traffic volumes an macro level is done in Anamorava region and in second place, an analysis of factors according to disaggregate approach at the level of each of four locations separately for the given region. The analysis and comparison of results show that disaggregate approach provide much better results in forecasting and due to this is taken into account for the further analysis is conducted at location 1-Slivovo as well as representative location.

2.1 Modelling Technique

So far, there are many methods developed for traffic demand forecasting in the long term [9].These methods may in general be split according to conventional and non-conventional approaches [10]. Conventional approach contains mainly statistical methods, the most popular known like multiple linear regressions (MLR). On the other hand, as regard non-conventional method the most popular method used is artificial neural network (ANN). Principal Component Analysis (PCA) is used with intention to get as accurate as possible results in forecasting, respectively to eliminate high co-relation (co-linearity problem) between variables. PCA is a multivariate statistical method which is broadly used in analysing data in many areas because as such is simple and is qualified like non-parameter method. PCA is the procedure to reduce variables. It means mathematical procedure, transforming co-related number of variables to the non-correlated number of variables Principal Component (PCs). Thus, applying PCs, apart from specific models MLR and ANN in forecasting of traffic demand, there are also combined models PCA-MLR known as Principal Component Regression (PCR) and PCA+MLP model. Following is explanation for these two combined or hybrid models.

2.1.1 Multiple Regression Analysis

MLR is one of modelling techniques to find out relation between one dependent variable (Y_i) and some other independent variables (X_{in}), when i=1,2,...13, n=1,2,...13. The general form of model according to MLR is given through Eq. 1 [11]:

$$Y_i = \beta_0 + \beta_1 \cdot x_{1i} + ... + \beta_k \cdot x_{ki} + \varepsilon_i \qquad (1)$$

In which: Y_i-dependent variable, β_0-intercept, β_i-are regression coefficients, x_{in}-are independent variables and ε_i-is an error associated during regression. In MLR model, this error term ε_i is supposed to be normal distribution around zero and constant variance. It is also supposed the residuals should not correlate. In order to calculate the values of parameters the least square method is used. In this regard, statistical tests are used for assessing of the model from the point of view of relevance of other elements (F-test, t-test, VIF, DW etc.)

2.1.2 Artificial Neural Network

Appearance of significant non-linearity in statistical models cause the fact that conventional models are not enough suitable to be used or that given models represent the reality to the appropriate level. The proposed alternative for completion of forecasting is non-conventional method which is relied on the artificial intelligence.

Lately, models relied on artificial neural network (ANN) became attractive to many authors, which have been qualified like strong computer tools to resolve the problem of forecasting in many areas [12]. ANNs have great ability to detect and reproduce *linear* and *non-linear* relation between variables contained in the model. They are classified like *non-linear* models because they could give better explanation on data compared to conventional models. The advantage compared to statistical conventional models is that their structure is easily changeable. The process of learning associated with adjustability enable that network may modify and update its structure adjusting it to the dynamics of environment. Based on their taxonomy, there are many types of them based on the way of functioning, the number of layers, transformation function etc [13]. In this study we have applied ANN-Multilayer Perceptron (MLP) type. The joint feature of conventional and nonconventional models through ANN is that both function based on use of data from the past.

2.1.3 Multiple Regression Analysis based on Principal Component

An analysis through PCR model enables that MLR and PCA be together in order to create the relation between dependent variable and PCs, selected like input variables as presented in figure 2 [14]. Therefore, PCR is a hybrid of PCA and MLR.

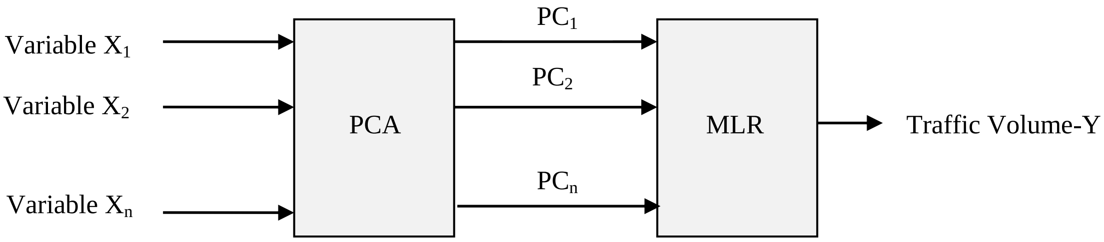

Fig.2 Scheme of functioning of combined method PCA-MLR (PCR)

In this case points per PCs gained from PCA are taken like independent variables in the equation of MLR in order to have PCR analyze functioning. The general expression of PCR model is given through Eq. 2.

$$Y = \beta_0 + (\beta_1 \cdot PC_1 + \beta_2 \cdot PC_2 + \cdots + \beta_k \cdot PC_k) = \beta_0 + \sum_{k=1}^{n} \beta_j \cdot PC_j + u_i \tag{2}$$

2.1.4 Artificial Neural Network based on Principal Components

With intention to improve model performance established through ANN network by all original input data, analysis is used through PCA in order to get Principal Components (PCs) which can be used afterwards as an input to neural network [15]. This combination will gain hybrid model known as PCA-ANN. In this case the number of input variables can also be reduced which will also reduce the complexity in treating network. As explained above, through PCA analysis there is possibility to eliminate correlation between independent variables, respectively to avoid multiple co-linearity problems. The modelling strategy in this case is the result of an idea to get the most efficient model of ANN which will be capable to forecast traffic volumes with higher accuracy but with minimum number of variables used like input data to the model. It should also be verified whether through combined model PCA-MLP (hybrid) there is possibility to forecast traffic volumes for a region compared to conventional statistical methods and to neural network method according to MLP. The way of functioning is reflected in figure 3.

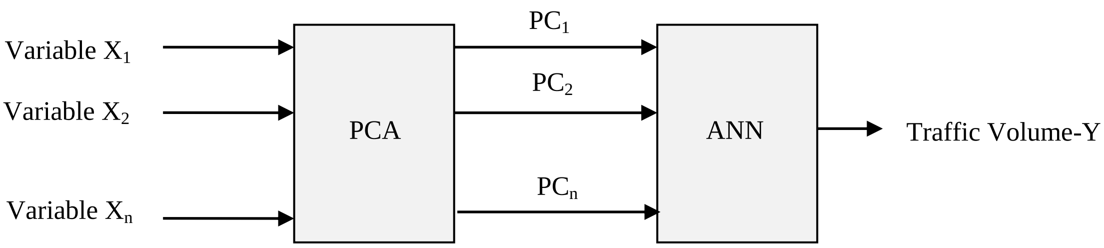

Fig.3 Scheme of functioning of combined method PCA-ANN

3. RESULT AND DISCUSSION

The analysis starts with normality of variables involved in model. For all variables based on Shapiro Wilk test the result is normal distribution fulfilling the condition that the value is Sig=0.823>0.05. Since the condition is fulfilled, there is no need to apply required transformations to these data (such is log, sqrt etc), and the data in the data set could be directly used for further analysis in setting up the models. Further, it is continued with descriptive statistics for variables taken in to the presented analysis according to table 1. In this regard, the values for all other independent variables are given with exception of X_1 variable which indicates years and all other variables provided according to years. Identification of these variables with impact in generating traffic demand is done relied on the literature availiable in this area [16].

Table 1. Basic descriptive statistic for variables reviewed

Variables		N	Min	Max	Mean	Std. dev
Traffic volume	Y	13	6325	10439	7449	1240
State Population	X_2	13	1786282	1891906	1846303	37473
State Household	X_3	13	278915	338618	311062	17863
State Employment	X_4	13	236181	340911	290450	33631
State Vehicle registration	X_5	13	179157	336942	249102	52192
Consumer Price Index	X_6	13	77	101	90	9
Gross Domestic Product	X_7	13	3006100	5984900	4431790	1072114
Per capita Income	X_8	13	1763	3356	2507	562
Gasoline price	X_9	13	0.840	1.160	1.015	0.027
Region population	X_{10}	13	240502	254723	248583	5045
Region household	X_{11}	13	48999	51442	50504	685
Region employment	X_{12}	13	32270	43692	37302	3983
Reg. vehicle registration	X_{13}	13	29031	53806	39419	8514

Once each of four methods mentioned above were applied using SPPS software version 22 after verification of many analysis and statistical tests, significant models were identified on the control level $p< 0.05$. Results are issued and reflected as follows.

3.1 Results according to MLR

By regression analysis, the relation between independent variables and traffic volume is found out. The MLR model is developed through 13 observations by 12 independent variables (X_{in}) against dependent variable (Y_i). In this regard, by application of stepwise technique, the best model is found with highest values adjusted R^2=0.859, which contain only one significant dependent variable (X_{13}) compared to other tested and proven models. Summarized model data are presented in table 2.

Table 2. Model summary of statistical parameters using MLR

Model Summary[a]						
Model	R	R^2	Adjusted R^2		Std. Error	D.Watson
1	0.933[a]	0.871	0.859		465.53011	1.338
ANOVA						
Model	SumSquares	Df.	Mean Square		F	Sig.(p<0.05)
Regression	16071147.64	1	16071147.643		74.157	0.000[b]
Residual	2383901.12	11	216718.284			
Total	18455048.76	12				
Coefficients						
	B	Std.Err	t	Tolerance	VIF	Sig.(p<0.05)
Constant	2091.432	635.437	3.291			0.007
X_{13}	0.136	0.016	8.611	1.000	1.00	0.000
a. Dependent Variable: Traffic volume (Y)						

With an aim to eliminate multi co-linearity values of Variance Inflation Factor (VIF) and statistical test of Durbin Watson (DW) is used. The value of independent variable was VIF=1, which show that values up to 10 indicate that there is no multi co-linearity between dependent variables. In this regard, statistical test DW=1.338 shows that the model does not face problem with first rule of auto correlation. The results indicate right orientation for determination of traffic volume which is 85.9 % of dependent variable of variation. The regression model resulted statistically significant by F-test (F= 74.157) for the control level 0.05 (since p < 0.000) and relation between independent variable and the dependent one is expressed by Eq.3:

$$Y = 2091{,}432 + 0{,}136 \cdot X_{13} \quad (3)$$

The regression coefficient is positive (0,136). It means that by increasing the level of variable, the value of dependent variable "traffic volume" is also increasing. The similar conclusion is gained also by statistical t-test for controlling individual regression coefficients, indicating that these coefficients are different from zero. Which is t-test (t=8.611 and p<0.000). In other words, X_{13} variable gives significant contribution to the model. In this regard, the residual analysis indicates that residuals are distributed in normal way by zero mean and constant variance.

3.2 Results by MLR based on Principal Component

When applying PCA analysis, PC1 and PC2 resulted as significant (p<0.05) with linear relation to the traffic volume (Y_i), reflected in table 3. Results in this table show clearly there is no problem with multi co-linearity because the value of VIF is much smaller than 10 (VIF=2.713) and the value of tolerance is bigger than 0.1 (TOL=0.369). Durbin-Watson statistical test show that the model does not suffer from the problem of the first scale of auto correlation (DW=1.486). Residual analysis show that residuals are distributed normally around zero as average and constant variance.

Table 3. Model summary of statistical parameters using PCA

Model Summary[a]						
Model	R	R^2	Adjusted R^2		Std. Error	D.Watson
1	0.948[a]	0.899	0.879		431.58135	1.486
ANOVA						
Model	SumSquares	df	Mean Square		F	Sig.(p<0.05)
Regression	16592424.14	2	8296212.07		44.540	0.000[b]
Residual	1862624.62	10	186262.46			
Total	18455048.76	12				
Coefficients[a]						
	B	Std.Err	t	Tolerance	VIF	Sig.(p<0.05)
Constant	7449.308	119.699	62.234			0.000
PC1	1516.286	205.210	7.389	0.369	2.713	0.000
PC2	-473.241	205.210	-2.306	0.369	2.713	0.044
a. Dependent Variable: Traffic volume (Y)						

Based on the table above, both PC resulted significant as independent variables in setting up MLR analysis. The coefficient of determination of PCR is R^2=0.879 which have improved compared to the model set up according to MLR. PC_1 has a positive impact while PC_2 has negative impact in forecasting traffic volume. The relation between these two components of traffic volume is given through these two components and expressed through the equation Eq.4:

$$Y = 7449{,}308 + 1516{,}286 \cdot PC1 - 473{,}241 \cdot PC2 \quad (4)$$

3.3 Results achieved by ANN

The aim here is to verify whether through ANN (type of MLP) there is possibility to improve forecasting of traffic volume on the main road network of any given region compared to conventional statistical methods. Data used to set up the model are taken randomly on 10 observations or 76.9%, testing for 3 observations or 23.1%. Many trials are done according to “trial and error” technique to gain suitable percentage between training and testing of data so that the model provides the best performance in forecasting. Data set of training is used to find out weights and to set up the model while data on testing are used to find errors and to prevent overtraining in the process of training. Selection of network structure is done automatically introducing the minimal value 1 while the maximum is 50, which resulted to be 12 neurons as input, 7 neurons in a hidden layer and 1 neuron at the exit layer which means also exit variable of traffic volume. On the hidden layer “hyperbolic tangent” activation function is used, while for exit layer “identity” is applied. Error function applied is “sum of squares”. Summary model is shown in table 5, provide information related to the results of training and testing sample. The sum of errors in square is used in both samples of observations both on training and tested one. Thus, error function is used to show that network goes to its minimum in the testing phase.

Table 5 Model summary

Training	Sum of Squares Error	.184
	Relative Error	.041
	Stopping Rule Used	1 consecutive step(s) with no decrease in error [a]
	Training Time	0:00:00.02
Testing	Sum of Squares Error	.012
	Relative Error	.020

Dependent Variable: Y
a. Error computations are based on the testing sample.

The value of “sum of squares error” of training (= 0.184) show the strength of the model to forecast the result. The lowest value of the sum of square in the testing case (=0.012) is lower than in the case of testing, which means that network model is not over trained by training data and as such is used to summarize the trend. The results gained justify the role of testing sample which is dedicated to prevent overtraining. Based on results in table 5, the percentage of inaccuracy to forecast the sample of training is 0.041 (4.1%) while at testing it drops to 0.020 (2%). Thus, calculation of errors is based on the testing sample. The value R^2=0.963, expressed by equation 5 means an extraordinary good result.

$$y=4.5e2+0.94\cdot x \qquad (5)$$

3.4 Results by ANN based on Principal Components

PCA is applied in many researches in pre-processing phase in neural network of MLP type. Results show that PCs as inputs improve forecasting by MLP models by eliminating co linearity of data and reducing the number of forecasting variables. In this regard, explanation variables according to PCA technique are analysed based on co-variance matrix and as such principle components PC1 and PC2 are gained. Kaiser criterion on selection of the number of components is applied in this study according to which PCs with higher than 1 eigenvalue is kept and are taken into account for further analysis. Once PCs are gained, rotation is completed so that factors be independent or dependent remain linear. In this case values for principal components gained of PCA are considered like independent variables in multilayer neural network type MLP so that the model may function. In order to prove validity of PCs the sample of dataset is divided in two parts, the one of training and the other one of testing.

The data on setting up the model are fixed on random basis on training, nine observations or 69.2%, and testing 4 observations or 30.8%. Determination of neurons of hidden layers is done by selection in automatic way and by selection of 2 input neurons, 2 neurons of hidden layer and one neuron of exit layer which means also exit variable for traffic volume setting the minimal value 1 and the maximum 50. As regards hidden layer “hyperbolic tangent” activating function is used while for exiting layer the "identity" function is used. Error function used is "sum of squares". The summary model shown in table 6, provide information related to the results of training and testing samples. The sum of square errors is used in both samples of observation both on training and testing one. Thus, error function is used showing that network is minimizing in the phase of testing.

Table 6 summary model

Training	Sum of Squares Error	.033
	Relative Error	.008
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.00
Testing	Sum of Squares Error	.026
	Relative Error	.020

Dependent Variable: Y
a. Error computations are based on the testing sample.

The value of "sum of squares error" on training (= 0.033) show the strength of the model to forecast the result. The lowest value of squares at the testing case (=0.026) is lower than at training case which means that network model is not over trained by training data and as such it tends to summarize. The results justify the role of testing sample which is dedicated to prevent overtraining. Based on results according to table 6, the percentage of inaccuracy in forecasting of training sample is 0.008 (0.8 %) while on testing is 0.020 (2%). In this case calculation of mistakes is based on the testing sample. The value R^2= 0.990, in equation 6 means extraordinary good result.

$$y=1.5e2+1.02\cdot x \tag{6}$$

4. PERFORMANCE INDICATORS

One of the most important rules in selecting the forecasting method is its accuracy or as else expressed to what extend forecasting data (F_i) comply with observation data (A_i) so that errors are as low as possible [17]. In this case the comparison and assessment of models according to some indicators of performance to determine the accuracy of forecasting of models established according to various conventional methods and techniques (MLR), and those non-conventional (ANN) as well as their combination (PCA-MLR and PCA-ANN) is shown in table 7.

Table 7. Performance indicators between MLR, MLP, PCA-MLR and PCA-MLP models

Performance Indicators	MLR	MLP	PCA-MLR	PCA-MLP
Coefficient of Corelation (Adjusted R^2)	0.99839	0.99953	0.99874	0.99986
Mean Absolute Deviation (MAD)	361	141	336	102
Mean Squared Error (MSE)	183377	52954	143279	15855
Root Mean Squared Error (RMSE)	428	230	370	126
Mean Absolute Percentage Error (MAPE)	4.25	1.67	4.46	1.39
Root Mean Square Percentage Error (RMSPE)	5.09	2.47	4.93	1.68
Coefficient of Variation (CV)	0.057485	0.03089	0.050813	0.01690

Comparing results based on performance indicators it results that the model which performs better in forecasting or which provide less error is combined "hybrid" PCA-MLP model. Therefore, in forecasting traffic volumes for any given period it is preferred to take this model.

5. CONCLUSION

The study as such is of macro level because it takes into account demographic and social economic variables in national and regional level. All variables identified resulted to have normal distribution with impact in generating traffic demand in regional level. In this regard, based on them, the model for forecasting traffic volume is established for the main road network for Anamorava region according to aggregate and disaggregates approaches and multiple linear regression methodology (MLR). Following detailed analysis it resulted that MLP model according to disaggregate approach provide the best performance in forecasting. As a consequence of appearance of non-linearity of variable data and with intention to improve forecasting ability of the model, the model was applied according to artificial neural network (ANN), in which based on the way of functioning, respectively the algorithm used to set up the model known like Multilayer Perceptron (MLP). Further, with intention to eliminate high correlation expressed between independent variables (multi co-linearity phenomenon) taking part in setting up the model as well as improvement of its forecasting ability PCA is applied in which principle components (PCs) are applied like input for MLR and ANN methods. According to this logic, combined or "hybrid" models are developed which are known like PCA-MLR (or by an abbreviation PCR) and the other one PCA-ANN. Thus, four models are set up, in which, based on detailed statistical analysis it resulted that all models are relevant and may be used for forecasting. Nevertheless, with intention to find out which of them provide the best forecasting performance, comparison is applied according to performance indicators. Comparison show that the model with better performance is the combined PCA-MLP model, and as such is preferred to be used in forecasting the demand, respectively to forecast the traffic volumes in the main road network of Anamorava region.

REFERENCES:

[1] Ministry of Infrastructure-Directorate of Roads. „Data for traffic volumes for the period 2004-2016", Prishtina,2016

[2] Mihailo M., Planiranje Saobracaj i Prostora. Gredevinski Fakultet, Beograd, 2004

[3] Gjevori SH., Traffic Engineering-First Part (Book in Albanian). Polytechnic University of Tirana, Faculty of Mechanical Engineering, Tirana, 2011

[4] Neveu, A.J., Quick response procedure to forecast rural traffic. Transportation research record 944, June 1982

[5] Gjevori SH., Transport Systems. (Book in Albanian). Polytechnic University of Tirana-Faculty of Mechanical Engineering. Tirana, 2010

[6] Hutchinson, B.G. Principle of urban transport system planning. Script Book Company, Washington D.C.

[7] Agency Statistics of Kosovo. Kosovo Census Atlas, Geographic and Administrative Division of Kosovo. Prishtina, 2013.

[8] Agency Statistics of Kosovo,"General Statistics of Kosovo".Prishtine, 2016.

[9] Slipunas T., Annual average daily traffic forecasting using different techniques. Transport and Road Research Institute. VolXXI,No.1,38-43, Lithuania, 2006.

[10] Mirko, Č. Modeliranje u Železnickom Saobračaju. Saobracajni Fakultet-Universitet u Beogradu. Beograd, 2003

[11] Washington S.P, Karlaftis M.G, Mannering F.L.,"Statistical and Econometric methods for transportation data". Second Edtion, Taylor and Francis Group, LLC, 2011

[12] Sousa et al., "Multiple linear regression and ANN based on principal components to predict ozone concentrations". Elsevier, 2006

[13] Živojinović, D. Z. Razvoj i Primena Hemometrijskih Metoda za Klasifikaciju i Procenu Kvaliteta Vode. Doktorska Disertacija. Univerzitet u Beogradu-Tehnološko-Metalurški Fakultet. Beograd.2013.

[14] P.B Mistry, "Principal regression for crop yield estimation". Springer, 2016.
[15] Ali, G.A., Awdalla, T. Characteristics and predictions of traffic accident causalities in Sudan using statistical modelling and ANN. IJTST, Vol 1. N0.4,2012-pg.305-317.
[16] Fricker J.D, Saha K.Sunil,. Traffic volume forecasting methods for rural state highways-Final report. FHWA/IN/JHRP-86.20, JHRP, Purdue University, 1987
[17] Long range Forecasting From Crystal Ball to Computer. www.forecastingprinciples.com/files/LRF-ch13.pdf

ПРОГНОЗИРАНЕ НА ОБЕМА НА ПРЕВОЗИТЕ ЧРЕЗ ИЗПОЛЗВАНЕ НА РЕГРЕСИОНЕН АНАЛИЗ И ИЗКУСТВЕНА НЕВРОННА МРЕЖА

Рамадан Дураку[1], Васка Атанасова[2]
ramadan.duraku@uni-pr.edu, vaska.atanasova@tfb.uklo.edu.mk

[1]Университет Прищина "Хасан Прищина" Факултет по инженерна механика
Съни Хил, 10 000 Прищина, КОСОВО
[2]Университет "Св. Климент Охридски", Факултет по технически науки
Македонска фаланга 33, Битоля, МАКЕДОНИЯ

Ключови думи: *обем на превозите, модел, прогнозиране, основни компоненти, множествена регресия, изкуствена невронна мрежа.*

Резюме: *Целта на доклада е да се изследва кои фактори оказват най-голямо влияние върху генерирането на превози, както и да се създаде подходящ модел за прогнозиране на превозите по пътно-шосейната мрежа на Анаморавския регион. В тази връзка са изследвани няколко демографски и социални показатели за периода 2004-2016. Променливостта на показателите е изследвано чрез корелационен анализ (проблем на мултиколинеарността). По нататък в разработката се извършват анализи на база реални данни посредством множествена регресия и методите на изкуствената невронна мрежа. Поради факта, че при анализа се добавят възможно най-много променливи и с цел да се избегне явлението мултиколинеарност и да се подобри надеждността на прогнозата, приложение намира и метода за анализ на основните компоненти. Посочените методи се сравняват на база статистическата грешка. Резултатите от проведеното изследване показват, че използването на основни компоненти като входни данни при анализа водят до по-надеждни прогнози и елиминиране на явлението мултиколинеарност.*