Prediction methods for the actual duration of Greek highway projects

MAGDALINI TITIRLA
Structural Mechanics and Coupled Systems Laboratory (LMSSC)
Conservatoire National des Arts et Métiers (CNAM)
2 rue Conté, 75003 Paris
FRANCE

WALID LARBI
Structural Mechanics and Coupled Systems Laboratory (LMSSC)
Conservatoire National des Arts et Métiers (CNAM)
2 rue Conté, 75003 Paris
FRANCE

GEORGIOS ARETOULIS
Department of civil engineering
Aristotle University of Thessaloniki (AUTh)
Egnatia 137, Thessaloniki
GREECE

Abstract: - This study presents an overview of estimation methods to predict the actual project duration of Greek highway projects. Emphasis is given to the selection of the appropriate parameters that correlate with the actual project duration and to compare the performances of the main two methods, the linear regression (LR) with the neural network models (NN) based on data available at the bidding stage. In the context of the current research, thirty-seven highway projects were examined, constructed in Greece with similar available data like the extent, the type of work packages and the significance. Selection and ranking variables through correlation analyses using SPSS 25 has been carried on, in order to identify the most significant project variables. These include archeological findings, type of terrain, land expropriation, the existence of bridge, tunnel and embankment. Next step was the use of WEKA application, that highlighted the most efficient subset of variables. After the definition and grouping of the variables for actual duration prediction, these were used as input data for linear regression models (LR) and neural network models (NN). Various models have been created from each investigated method. While their performance and the comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects are presented in this paper. Results' discussion and conclusions along with limitations and further research are appropriately analyzed.

Key-Words: - Actual duration, highway projects, linear regression, neural networks, predicting models

1 Introduction

A significant number of studies that focus on the estimation of highway projects' duration exist in the bibliography, where each study focuses on various characteristics of the project and the location where the project is implemented. There are several variables that influence the project implementation in each country like the unique legal framework, the project procurement systems and the type of the contracts, as well as the cost and the schedule planning, both internal and external [1]. Variation in project costs usually results in exceeding the initial estimated cost [2]. Each highway project is unique as there are a lot of parameters which play a role in the differentiation of projects. According to Shah and Dawood [3], the construction site accessibility and materials' procurement play an important role in construction [4]. Antoniou [5] has classified and standardized all the factors that cause delay in road

construction projects based on delay data from 120 projects constructed in a similar environment with akin procurement and contract policies.

Arun and Rao [6] proposed an innovative prediction tool using generating stochastic algorithm for the duration overrun, cost overrun and activities associated with any specific delay in highway construction projects. In the proposed tool, the models were focused on the nature of the delay, the activities connected with the delay and stochastic nature of the duration and cost overruns. In the same concept, Long and Ohsato [7] proposed a new method based on genetic algorithms for scheduling construction projects with several variables including project cost and project duration. Furthermore, Williams et al. [8] used data from 311 projects from Texas and Virginia in USA and found out that construction time is independent of the cost. Irfan et al. [9] used a mathematical approach to estimate the duration of highway project based on variables which were known at the planning phase, such as planned cost, project type and contract type. The main findings proposed a relationship among highway project duration and the magnitude of the planned cost, project type and contract type.

Anastasopoulos et al. [10] studied data from 1,722 projects from Indiana. The selected variables included project time delay, project cost, project type, weather, level of competition, difference between the winning bid amount and the engineer's estimation and finally difference between the winning bid amount and the next lowest bid. Han et al. [11] focused on the way to increase the accuracy of projects duration of highways during construction period. The proposed method for road duration risk is based on the risk analysis theory. The following parameters were considered: allowable stage duration, expected according to the characteristic of road construction stage duration, risk rate of stage duration, weight of stage duration and stage time risk. The study of Kaleem et al. [12] aiming to the estimation of the highway project duration. The investigated variables were planned cost and project type which are known at the planning phase, as well as pavement construction, improvement, rehabilitation and bridge construction projects of National Highway Authority, Pakistan. The outcomes of their study was a mathematical relationship between highway project duration, planned cost and project type. Antoine et al. [13] studied the relationship among project duration, project intensity and timing of cost certainty in highway project delivery methods. Alternative contracting methods of construction manager/ general contractor CM/GC and Design Build are superior to the traditional Design Bid Build method for the project performance.

Aziz et al. [14] produced a mathematical approach, based on statistical regression analysis, to predict the cost and the time of a project and to facilitate the evaluation of the project's performance during construction. In this study three methods of analysis have been used: ridge regression analysis, general regression analysis and non-linear partial least square regression analysis. Furthermore, Waziri et al. [15] developed through regression analysis models to predict highway construction duration. They incorporated relevant variables, with statistically significant relationship with highway completion time, based on historical data of highway projects initiated and completed between 2007 and 2012.

Wang et al. [16] built neural network (NN) models linking pre-project planning and project success. In this context, Wang et al. [17] developed artificial neural networks (ANN) and support vector machines learning models to estimate project cost and schedule success. Pewdum et al. [18] focused their research on the factors affecting project final budget and duration before developing the forecasting models using Artificial Neural Network (ANN). Furthermore, Glymis et al. [19] using FANN Tool proposed three neural network models to predict the actual project duration for highways, based on tender budget, length of highway project, number of lanes, number of technical projects, number of bridges, tunnels and road total length. Gkovedarou and Aretoulis [20] proposed Neural Network models for time and cost prediction of urban redevelopment projects, while Titirla and Aretoulis [21] proposed Neural Network models for the time prediction of highway projects. Finally, Aretoulis [21] used data from selected public Greek highway projects in order to create models to estimate their actual construction cost, based on ranged variables with the aid of SPSS.

The aim of this study is to present an overview of estimation methods to predict the actual project duration of Greek highway projects. For our research, thirty-seven highway projects constructed in Greece with similar available data like the extent, the type of work packages and the significance have been examined. The research focuses on the selection of the appropriate parameters that correlate with the actual project duration, creating a ranged list through correlation analysis and WEKA application. After that, the variables were used as input data for linear regression models (LR) and neural network models (NN). Various models have been created from each investigated method, while

their performance and the comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects are presented in this paper. Finally, a comparison of the performances of the main two methods, the linear regression (LR) with the neural network models (NN) was carried out. Results' discussion and conclusions along with limitations and further research are appropriately analyzed.

2 Description of the highway projects and the available data

The 37 investigated highway projects come from Egnatia Odos' sub-projects, which is one of the most modern highways in the Southeast Europe and part of the TEN-T corridors. The database consists of 12 variables for each one of the 37 projects, which cover common available data across them and according to expert opinion greatly affect the project duration. The variables, presented in Table 1, were recorded as quantitative and 6 of them were transformed into qualitative ones to facilitate additional statistical analyses.

Table 1. Variables of the investigated data.

Variables	
Quantitative variables	**Qualitative variables**
Project length *(#1)*	Landfill *(#7)*
Initial duration *(#2)*	Geotechnical project *(#8)*
Initial cost *(#3)*	Embankment *(#9)*
No. of lanes *(#4)*	Tunnel *(#10)*
No. of technical projects *(#5)*	Bridge *(#11)*
Tender offer *(#6)*	Land requirement *(#12)*

The quantitative variables take on numerical values, the length (#1) in meters, the duration (#2) in days; the cost (#3) in euros, as well as the number of lanes (#4) is the number per direction. The qualitative variables take on binary values (Yes/No) or ordinal values. The following variables: landfill (#7), geotechnical project (#8), tunnel (#10), bridge (#11) and land requirement (#12), can only take the values of "0" and "1" (if there has been no landfill, the variable's value is 0, in the opposite case, the variable's value is 1), while the value for the variable of embankment ranges from 1 to 3, according to the height of the embankment (up to 4 m = 1, between 4 and 10 m = 2, over 10 m = 3).

3 Methodology

The present study focuses on the production of reliable and efficient models (linear regression and neural networks) for the prediction of actual construction duration of highway projects located in Greece. The methodological approach includes the following steps (Fig. 1):

Step 1: Consideration of 37 selected highway projects of similar extent and content.

Step 2: Collection of available variables for actual duration prediction.

Step 3: Construction of an appropriate SPSS database. Analysis was conducted one time for quantitative variables and a second time for mixed variables, both quantitative and qualitative variables.

Step 4: Application of the "Attribute Selection" module of WEKA application for selecting subsets of variables

Step 5: Creation of a ranked list of variables.

Step 6: Models creation.

Step 6-1: Creation of linear regression models, based on SPSS.

The models were created starting with the highest correlating variable according to correlation analysis and WEKA and then adding/removing one more variable each time from the ordered list.

Step 6-2: Creation of neural network models, based on FANN tool.

The models were created starting with the highest correlating variable and then adding one more variable each time from the ordered list based on correlation analyses.

Step 7: Comparison of the performance of the produced models (LR and NN).

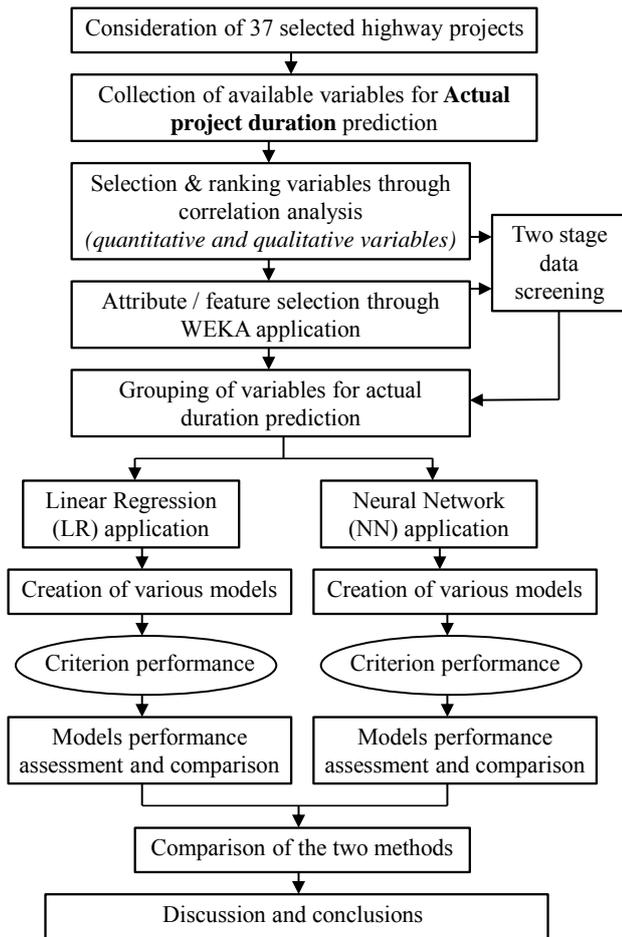**Step 8:** Comparison of the performance between the LR model and the NN model.

Fig. 1. Methodological approach.

4 Proposed Models

In this section the creation of the main two models are presented, the linear regression (LR) and the neural network models (NN) based on data available at the bidding stage of highway projects.

4.1 Linear Regression (LR)

Linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quintile is used. Linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate

analysis. Having identified the potential correlations among the selected common variables, the current paper applies the linear regression analysis. The selected method is the stepwise. The aim is to predict the final duration.

In the Step 3, a database in SPSS was organized, taking into account the variables both qualitative and quantitative in nature. In the linear regression model only the quantitative variables have been used and a brief look on the correlation analyses' results revealed that the actual project duration was associated in descending correlation order with the following variables: Initial Cost (#3), Tender Offer (#6), Initial Duration (#2), Number of Lanes (#4) and Length (#1), while there was no correlation with the number of technical projects (#5). In total 7 models were presented in this study, the 6 of them according to the ranking list of correlation analysis and one of WEKA ranking list. Fig. 2 shows the methodological approach of linear regression models, and a comparison of the mean squared error (MSE).
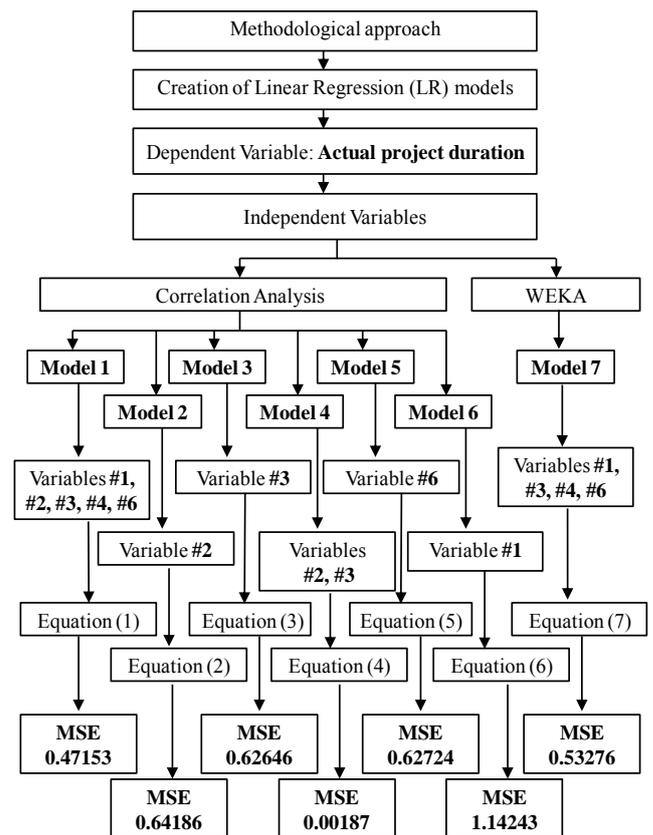

Fig. 2. Methodological approach creating the linear regression models.

After the analysis of the 37 data cases, the proposed linear regression equations are presented in the Table 2. Regarding Figure 2 we can notice that the model with the smaller MSE is the model 4 which

consist of the variables Initial cost (#3) and initial duration (#2) (equation 4).

Table 2. Linear regression models.

(1) $Actual\ Duration = 2.885^{-6} \cdot (Initial\ Cost) + 0.794 \cdot (Initial\ Duration) + 0.009 \cdot (Length) + 7.557 \cdot (Lanes) - 4.436^{-07} \cdot (Tender\ Offer) + 326.804$

(2) $Actual\ Duration = 1.458 \cdot (Initial\ Duration) + 200.51$

(3) $Actual\ Duration = 4.632^{-06} \cdot (Initial\ Cost) + 841.539$

(4) $Actual\ Duration = 2.818^{-06} \cdot (Initial\ Cost) + 0.843 \cdot (Initian\ Duration) + 424.729$

(5) $Actual\ Duration = 4.106^{-06} \cdot (Tender\ offer) + 801.688$

(6) $Actual\ Duration = 0.020 \cdot (Highway\ Length) + 844.079$

(7) $Actual\ Duration = 0.009 \cdot (Length) + 2.645^{-06} \cdot (Initial\ Cost) + 52.797 \cdot (Lanes) + 9.177^{-07} \cdot (Tender\ Offer) + 499.152$

Furthermore, the equation 4 (Model 4 of LR) which includes two variables, the initial cost and the initial duration and it is able to predict the actual project duration with predictability equal to 72%. Figure 3 compares the actual duration in days versus the prediction duration of the linear regression Model 4.
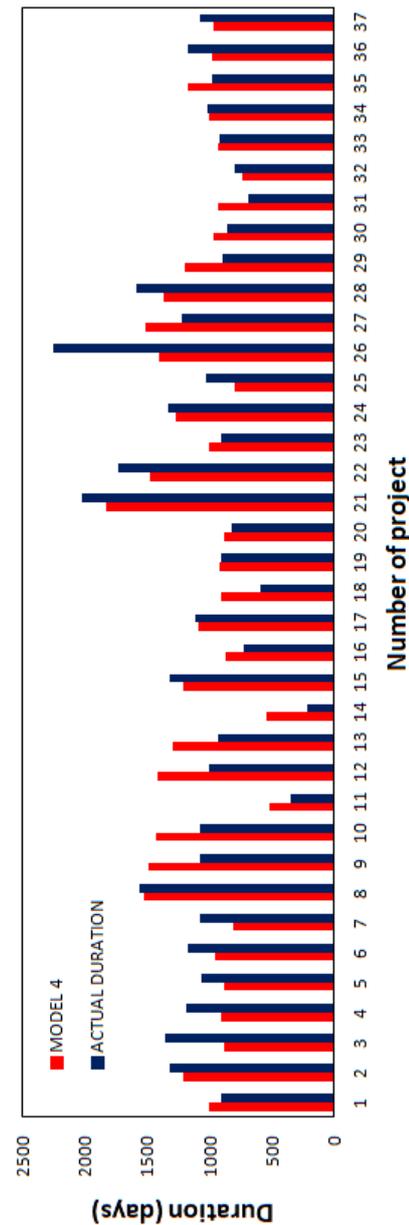

Fig. 3. Linear regression prediction duration versus actual final duration.

4.2 Neural Network (NN)

FANN Tool application was implemented to produce the neural network. It should be mentioned that the FANN library (FANN) is a free open source neural network library. The specific library implements multilayer artificial neural networks in C and provides support for both fully and sparsely connected networks. Furthermore, it includes a framework for easy handling of training data sets. It is indeed an easy to learn and use application and at the same time versatile, well documented, efficient and fast. In essence, FANN Tool is a Graphical User Interface (GUI) to the FANN library. Finally, it facilitates easy usage without the need of

programming. Summarizing, this specific tool facilitates a number of relevant process [23]:

- preparation of the data in FANN library standard;
- design of an ANN;
- training of the designed ANN;
- testing of the trained ANN; and
- applying of the trained ANN.

Neural networks produced several models. The dependent variable was "Actual Project Duration" in days. Twenty-one projects were used for training the neural network and sixteen projects were used for testing the produced neural network. The methodology involving the application of neural networks for the prediction of actual duration is based on multiple neural networks designed and applied specifically for the problem at hand. The current research is contributing and focusing on identifying the best possible combination of input variables for the optimum prediction. Figure 4 shows the methodology approach of Neural Network models, and a comparison of the mean squared error (MSE).
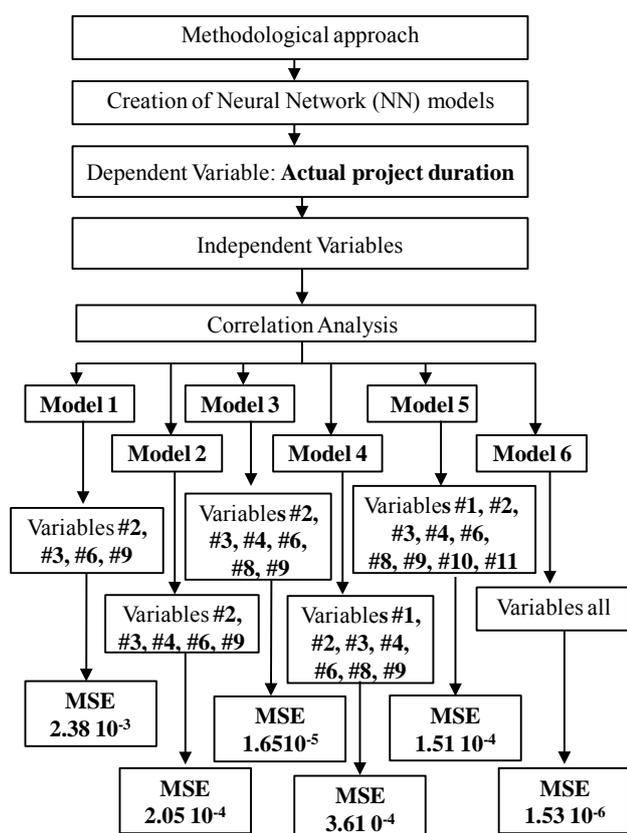

Fig. 4. Methodological approach creating the Neural Network models.

The best Neural network model has the smaller minimum squared error (MSE). The proposed model is the one that includes the top 12 variables, namely,

initial cost (#3), initial duration (#2), length (#1), lanes (#4), technical projects (#5), bridges (#11), tunnels (#10), geotechnical projects (#8), embankment (#9), landfill (#7), land requirement (#12) and tender offer (#6). This neural network model has a MSE equal to $1.53 \cdot 10^{-6}$.

5 Conclusion and further research

This paper presents the important variables that correlate with the actual project duration and compare linear regression with neural network models to predict a reliable final duration for highway construction projects based on data available at the bidding stage. In total, 37 highway projects constructed in Greece with similar extent and content, were investigated. The corresponding models (linear regression and neural networks) are presented along with their performance. Finally, this paper presents the comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects.

The proposed model of linear regression (LR) includes only 2 variables, the initial cost and the initial duration, while for the neural networks the corresponding model includes the top 12 variables, namely, initial cost, initial duration, length, lanes, technical projects, bridges, tunnels, geotechnical projects, embankment, landfill, land requirement and tender offer. That means that LR model is able to predict the actual duration of Greek highway projects using less selected variables (only 2 compared to 12 at NN), as a result easier and faster data collection.

The linear regression model is able to predict the actual project duration with a predictability of 72% and with MSE equal to $1.87 \cdot 10^{-3}$, while the neural network model produces a MSE equal to $1.53 \cdot 10^{-6}$.

A further research in this area could aim at the development and the comparison of the results with other methods and tools, like the creation of Neural Network using Radial Basis Functions. In addition, a wider number of projects should be included, so that correlations and trends that are not observed by this research could be identified in a more extensive research and at the same time increase models' reliability and accuracy.

References:
[1] Antoniou F., Aretoulis G.N., Konstantinidis D. and Papathanasiou J., Choosing the Most Appropriate Contract Type for Compensating Major Highway Project Contractors, *Journal of Computational Optimization in Economics and*

Finance (JCOEF), Vol. 6, No. 2, 2014, pp. 77-95.

[2] Aretoulis G. N., Angelides D. C., Kalfakakou G. P., Fotiadis G. S. and Anastasiadis K. I., A prototype system for the prediction of final cost in construction projects. International Journal of Operational Research, Special Issue, Vol. 6, No. 3, 2006, pp. 423–432.

[3] Shah, R.K. and Dawood, N., An innovative approach for generation of a time location plan in road construction projects. *Construction Management and Economics,* Vol. 29 No. 5, 2011, pp. 435-448.

[4] Aretoulis, G.N., Kalfakakou, G.P. and Striagka, F.Z., Construction material supplier selection under multiple criteria, *Operational Research,* Vol. 10 No. 2, 2010, pp. 209-230.

[5] Antoniou F., Delay Risk Assessment Models for Road Projects, Systems, Vol.9, No.3, 2021, pp.70.

[6] Arun, C. and Rao, B.N., Knowledge based decision support tool for duration and cost overrun analysis of highway construction projects, *Journal of the Institution of Engineers (India): Civil Engineering Division*, Vol. 88, 2007, pp. 27-33.

[7] Long, L.D. and OHSATO, A., A genetic algorithm-based method for scheduling repetitive construction projects, *Automation in Construction*, Vol. 18 No. 4, 2009, pp. 499-511.

[8] Williams, R.C., Hildreth, J.C. and Vorster, M.C., Highway construction data collection and treatment in preparation for statistical regression analysis, *Journal of Construction Engineering and Management*, Vol. 135 No. 12, 2009, pp. 1299-1306.

[9] Irfan, M., Khurshid, M.B., Anastasopoulos, P., Labi, S. and Moavenzadeh, F., Planning-stage estimation of highway project duration on the basis of anticipated project cost, project type, and contract type, *International Journal of Project Management*, Vol. 29 No. 1, 2011, pp. 78-92.

[10] Anastasopoulos, P.C., Labi, S., Bhargava, A. and Mannering, F.L., Empirical assessment of the likelihood and duration of highway project time delays, *Journal of Construction Engineering and Management*, Vol. 138 No. 3, 2011, pp. 390-398.

[11] Han, S.Y., Li, H. and Hong, B.N., Risk analysis method for project duration of highways during construction period and its application, *Yantu Gongcheng Xuebao/Chinese Journal of Geotechnical Engineering*, Vol. 35, 2013, pp. 264-268.

[12] Kaleem, S., Irfan, M. and Gabriel, H.F., Estimation of highway project duration at the planning stage and analysis of risk factors leading to time overrun, T*&DI Congress 2014: Planes, Trains, and Automobiles*, 2014, pp. 612-626.

[13] Antoine, A.L., Alleman, D. and Molenaar, K.R., Examination of project duration, project intensity, and timing of cost certainty in highway project delivery methods, *Journal of Management in Engineering*, Vol. 35 No. 1, 2018, pp. 04018049.

[14] Aziz, A.M.A., Mcdanie, C., Engineer, C., Dayton, K., Baker, T.E., Smith, R. and Region, O., Performance Analysis and Forecasting for WSDOT Highway Projects (No. WA-RD 675.1), *Washington State Department of Transportation*, 2007.

[15] Waziri, B.S., Jibrin, A.T. and Kadai, B., Functional duration models for highway construction projects in Nigeria, *Arid Zone Journal of Engineering, Technology and Environment*, Vol. 10, 2014, pp. 13-23.

[16] Wang, Y.-R., Chen, Y.-J. and Huang, C.J., Applying Neural Network Ensemble Concepts for Modelling Project Success. *International Symposium on Automation and Robotics in Construction (ISARC 2009)*, University of TX at Austin, Austin, TX, 2009, pp. 199-204.

[17] Wang, Y.-R., Yu, C.-Y. and Chan, H.-H., Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models, *International Journal of Project Management*, Vol. 30 No. 4, 2012, pp. 470-478.

[18] Pewdum, W., Rujirayanyong, T. and Sooksatra, V. (2009), Forecasting final budget and duration of highway construction projects, *Engineering, Construction and Architectural Management*, Vol. 16 No. 6, 2009, pp. 544-557.

[19] Glymis, E., Kanelakis, A., Aretoulis, G. and Matoras, T., Predicting highway projects' actual duration using neural networks, *Joint Conference on Computing in Construction (JC3)*, 2017, Heraklion, Greece.

[20] Gkovedarou M and Aretoulis G., Neural Networks and Statistical Analysis for Time and Cost Prediction Models of Urban Redevelopment Projects, *IJISSC* vol.8, no.4, 2017, pp.37-52.

Magdalini Titirla, Walid Larbi, Georgios Aretoulis

[21] Titirla M. and Aretoulis G., Neural network models for actual duration of Greek highway projects. *Journal of Engineering, Design and Technology*, Vol. 17, No. 6, 2019, pp. 1323-1339.

[22] Aretoulis G.N., 2019. Neural network models for actual cost prediction in Greek public highway projects. *International Journal of Project Organization Management*, Vol.11, No.1, 2019, pp.41 – 64.

[23] FANN Tool Users Guide (2018), available at: www.slideshare.net/bluekid/fann-tool-usersguide.