

Research Article

Implementation of Process-Based and Data-Driven Models for Early Prediction of Construction Time

Silvana Petruseva ¹, Valentina Zileska-Pancovska,¹ and Diana Car-Pušić ²

¹“Ss Cyril and Methodius” University, Faculty of Civil Engineering, Partizanski Odredi 24, 1000 Skopje, Macedonia

²University of Rijeka, Faculty of Civil Engineering, Radmile Matejčić 3, Rijeka 51 000, Croatia

Correspondence should be addressed to Silvana Petruseva; silvana@gf.ukim.edu.mk

Received 5 October 2018; Accepted 23 December 2018; Published 23 January 2019

Guest Editor: Moacir Kripka

Copyright © 2019 Silvana Petruseva et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The need of respecting the construction time as one of the construction contract elements points out that early prediction of construction time is of crucial importance for the construction project participants' business. Thus, having a model for early prediction of construction time is useful not only for the participants involved in the construction contracting process, but also for other participants in the construction project realization. Regarding that, this paper aims to present a hybrid method for predicting construction time in the early project phase, which is a combination of process-based and data-driven models. Five hybrid models have been developed, and the most accurate one was the BTC-GRNN model, which uses Bromilow's time-cost (BTC) model as a process-based model and the general regression neural network (GRNN) as a data-driven model. For evaluating the quality of the models, the 10-fold cross-validation method has been used. The mean absolute percentage error (MAPE) of the BTC-GRNN is 3.34% and the coefficient of determination R^2 , which reflects the global fit of the model, is 93.17%. These results show a drastic improvement of the accuracy in comparison to the model when only data-driven model (GRNN) has been used, where MAPE was 31.8% and R^2 was 75.64%. This model can be useful to the investors, the contractors, the project managers, and other project participants for construction time prediction in the early project phases, especially in the phases of bidding and contracting, when many factors, that can determine the construction project realization, are unknown.

1. Introduction

Construction time is one of the key elements at the early phases of the construction project, particularly in bidding and contracting processes [1]. The problem arises precisely from the need to make a more accurate estimation of building time in those early phases of the project. There is, then, a high degree of uncertainty and often a lack of information required for a satisfactory accurate time assessment. Two elements are important for solving problems, collecting and systematically storing information, and developing new and improving existing models of time estimates, which will use such information. The accumulated legacy information in the information systems could improve decision-making [2]. Business decisions regarding the duration of construction projects can be considered among the most important because of the potential impact

on the final business results of the investor and the contractor.

While the problem of systematic storage and use of information can be solved by an appropriate information system [2], the development and improvement of appropriate assessment models are the result of scientific research. Moreover, addressing modern challenges and developmental trends in civil engineering [3–5], the information system should also include prediction time models that would use data system resources and supply it with new data. The next step further is that this information system time frame should be a segment of the unique construction information management system. Watson [5] classifies the “fragmented structure” into one of the “underlying, inherent construction industry problems.” Solutions that bring integrated management information systems have a synergistic potential with the ability to

enhance significantly the operational, functional, economic, management, and quality dimensions of the construction.

Problem of construction time prediction is a very complex and demanding task because the construction time is influenced by numerous factors such as project sector, building type, procurement route, construction materials, machines, and equipment; resources that will be used; methods used for work performance; project complexity and cost; site conditions; and many other factors [1, 6]. As a result, construction time prediction is a serious, difficult, and complex process [7].

During the construction period, the construction time is influenced by changes of its determining factors. Thus, many construction projects have not met the contracted time and have been finished with significant construction time overrun [6, 8–17]. Consequently, reaching the contracted construction time is an important problem worldwide. Project closure in the contracted time has become a project participant's challenge and an important factor that should be considered in each construction project [8, 9, 18].

Regarding the above-mentioned, construction time prediction is among the issues that are highly pronounced in the construction practice [19, 20]. As stated by Kenley [21], there are investigations focusing on predicting construction duration and cost, mostly due to cash flow modelling. However, there are also limited investigations that cover all aspects of their relation. Thus, Kenley [21] stated that the focus of researcher investigations are relations between project time and value in order to develop models for a rapid prediction of project time (using budgeted cost and current indices of cost as inputs). Additional investigations are related to exploiting and modelling time-cost relationship and its impact on important industry issues such as productivity improvement and industry efficiency.

As stated in [13], Bromilow, from Australia, was the first who investigated time and financial execution for buildings constructed in Australia between 1963 and 1967. Using simple linear regression analysis as the mathematical method, the research resulted in establishing the so-called “time-cost” model or BTC model: $Y = aX^b$, where Y is the construction time, X is the construction price, and a and b are constants. a is the coefficient that expresses the average time needed for construction of a monetary value, and b is the coefficient that indicates the sensitivity of the project duration regarding its value. BTC model has been tested and proven in additional investigations [17, 22].

The study [22] tested the plausibility of BTC model on construction projects realized in Australia in the period from 1991 to 1998. The suitability of BTC model was tested and proven with its implementation on different types of structures. It was proven that different types of projects need the explicit estimation of parameters. The investigation has also shown that the construction time for small industrial structures is shorter than the construction time for small educational and residential buildings. Two models were developed: the first for industrial and the second for non-industrial projects. It was concluded that different client sectors, contractor selection methods, and contractual

arrangements did not have influence on parameter change. At the same time, constants a and b were influenced by the characteristics of the region, economic characteristics, structure type, etc. Hence, the construction time prediction is more accurate if the model is developed for structures with similar characteristics, for a particular country or even a region.

Regarding the above-mentioned, researchers worldwide have developed regression models for different types of structures for different countries and even regions [17, 23–30]. For Malaysia, Chan [23] developed a “time-cost” model for building projects and proved its plausibility. Kaka and Price [12] developed such models for buildings and road construction in the UK, and the “time-cost” relation was confirmed once again.

Kumaraswamy and Chan [31] proved that the model can be applied in Hong Kong for buildings and construction structures. Similarly, Car-Pušić and Radujković [14] proved that the “time-cost” function can be used also in Croatia and consequently developed respective models for buildings, roads, and road structures.

The relationship between the initial estimated and the final achieved construction time was investigated by Ayodeji et al. [32]. The authors used linear regression analysis to investigate the relationship between the initial and actual construction time for public and private building projects in South Africa. It was determined that approximately 35% additional time needs to be added to the amount of the initial contract time in order to estimate the final, real contract time.

The next stage of research was to develop models based on the BTC, but with different predictors, as well as with two predictors. The main motivational factor was to obtain a more accurate time prediction because the time is influenced by numerous factors and not only by the project cost. Regarding this, Chan and Kumaraswamy [15] analyzed the government and private buildings construction duration in Hong Kong. Using BTC model as an initial model, they developed and tested new construction time prediction models as a function of one different independent variable, i.e., the total gross floor area in m^2 as well as the number of floors. Additionally, they developed and tested a model with two independent variables: the cost and the total gross floor area.

Following further model development, an increase in predictor numbers can be noticed. As stated by Car-Pušić [17], one of the starting problems for developing more variable models was the appropriate method selection. Numerous studies have proven the suitability of multilinear regression analysis. [12, 20, 29, 33, 34].

The models described above can be grouped into two main groups. The first group consists of models oriented to groups of activities, whereas the second group consists of models oriented to project characteristics [17]. Thus, in [10], the developed regression model for time prediction is oriented on groups of activities and their sequential start-start lag times. In the model, 12 independent variables (i.e., gross floor area, area of ground floor, approximate excavated volume, building height, etc.) have been used. A similar

model was developed by Chan and Kumaraswamy [15, 33] for public housing projects in Hong Kong by modelling work packages and their respective sequential start-start lag times. Chan APC and Chan DWM [29] developed similar benchmark time prognostic model for public housing projects in Hong Kong. What is interesting to point out is that the model was developed in order to formulate “benchmark measures of industry norms for construction period of public housing projects” in Hong Kong.

Regarding models oriented to project characteristics, there are numerous useful studies worth attention. Khosrowshahi and Kaka [34] stated that the project time and cost were influenced by different variables, separate or in combination. Their research was oriented on housing projects in the UK. The most influential variables were determined. The relationship between these variables and project duration, as well as the total cost, was defined, and finally, prognostic models developed.

Dissanayaka and Kumaraswamy [11] developed a time index regression model for building projects in Hong Kong considering a set of procurement and nonprocurement variables. They concluded that the project complexity representative value, programme duration, and client type, i.e., nonprocurement variables, are more significant than the procurement ones.

Žujo and Car-Pušić [35] developed regression models for construction time overrun. The construction time of building projects was considered as the function of risk factors based on the data for the buildings constructed in Federation BiH. Models for two groups of buildings, new construction and reconstruction, were established. It has been concluded that for new buildings, the most significant risk factors are weather conditions, technical documentation disadvantages, and law aspects, which means local regulations. Regarding reconstructions, the most significant risk factors are contractual and also technical documentation disadvantages. These models are applicable when an increased influence of risk factors is expected.

Similarly, Abu Hammad et al. [36] developed prediction models for construction duration for private and governmental projects in Jordan, classified according to the type of object. With 95% of probability, the proposed models predict the project duration with a precision of $\pm 0.35\%$ of the mean time.

Skitmore and Ng [27] developed several prediction models for the actual construction time based on the data of Australian construction projects. The cross-validation regression analysis was used to develop models when client sector, contractor selection method, contractual arrangement, project type, contract period, and contract sum are known. They also investigated models with the estimated contract period and contract cost.

Artificial neural networks (ANNs) are also used for time prediction modelling. In fact, ANNs have the prediction ability to solve numerous problems that appear in the construction industry [37].

Vahdani et al. [38] developed ANNs prediction model for construction project time prediction. The model is based on a new neuro-fuzzy algorithm. Furthermore, Petruseva

et al. [39] developed a multilayer perceptron (MLP) neural network model for construction time prediction. The model is based on real data. Bromilow’s “time-cost” model was implemented in two predictive models: linear regression (LR) and MLP. The results showed that using MLP significantly improves the accuracy of MLP model compared with the LR model.

Naik and Radhika [37] developed ANNs models for predicting the construction time duration for highway road construction using two completed projects. They obtained excellent results using the neural network fitting tool (Nftool) and neural network data manager (Nntool) approaches with the software MATLAB R2013a. In addition, they propose this approach to contractors for making much easier decisions.

Attarzadeh and Ow [40] propose the ANN model that improves the prediction accuracy of time by applying novel soft computing model. The model is characterized by good generalization and adaptation capability. It was shown that applying the good features of ANNs on algorithmic estimation model results in improvement of time prediction accuracy.

A model for construction time prediction using general regression neural network is presented by Petruseva et al. [41]. The coefficient of correlation between the predicted and the actual time values is around 0.999, and the error of the model is about 2.19%.

Neural networks have been used by Mensah et al. [42] to develop a hybrid model for predicting the duration of bridge construction projects in Ghana, using artificial neural network (ANN) and multilayer perceptron (MLP). Data were collected from the department of feeder roads for 18 completed bridge construction projects and included the number of lanes of the bridge components, their weights, and bridge span (20 to 54 m). The authors have shown that bridge project duration strongly depends on the bridge span and formwork used for reinforced in situ concrete. They have obtained good accuracy of the model with MAPE 4.05% and coefficient of determination $R^2 = 0.998$, making it suitable for predicting the duration of bridge construction projects.

Yousefi et al. [43] proposed a neural network model to predict time and cost claims in construction projects. By using the proposed model, the rate of possible claims in a particular construction project can be obtained.

Gab Allah et al. [44] developed the ANN time predictive model for building projects using MATLAB program as a model development environment. They used data for 130 building projects constructed in Egypt. Obtained accuracy of the model was with a maximum error of 14%.

Related to computing in civil and building engineering, many authors [3–5] point out the importance of digital information modelling as one of the guidelines and challenges for further research and development.

A relatively new predictive modelling research area is hybrid modelling, which combines two or more techniques, resulting in improvement in strength and performances of the model. The point is in using the good characteristics of each technique involved.

Regarding the above-mentioned, Roberts et al. [45] compared predictions of a process-based crop model, a data-driven model, and a hybrid model, i.e., a combination of both models, and found that the hybrid model performs much better than the other two models.

The authors in [46] proposed a hybrid model combining process-based and data-driven models in order to predict system remaining useful life applied on lithium ion battery. Drastically better accuracy was obtained when compared with the classical particle filter method. The conclusion was that by using the strengths of the data-driven and process-based methods together, the proposed hybrid prognostic framework bridges the gap between data-driven and process-based prognostics when abundant historical data and knowledge about the physical degradation process are available.

In this paper, a hybrid model, which includes process-based and data-driven models, is presented. As the result, a drastic improvement of the accuracy of the time prediction is obtained.

Issues discussed above lead to the conclusion that fast prediction of construction time, particularly at the early project phase when accurate and adequate information is limited, is not only difficult but also an important and necessary process.

Considering the fact that the construction time is influenced by a range of parameters that could not be accurately predicted, it is impossible to acknowledge all of them during the time prediction in early project phase [41]. This results in little accuracy of construction time [40] and points out the need and significance of future research in early construction time prediction. For this reason, the research has been carried out and its results are presented in this paper. The development of a more reliable model for early and fast construction time prediction, which would be used as a decision support tool at an early phase of the project, was the focus of this research.

Regarding that, data about construction duration from previous projects have been collected. Two methods, a process-based model and a data-driven model, have been used for predicting the construction duration. It should be highlighted that this is a relatively new research approach, demonstrating better results than using only one of these methods. Bromilow's time-cost model is used as a process-based model because of its simplicity and worldwide usage for time prediction. As the data-driven model, GRNN (general regression neural network) is used. Hybrid models usually combine the best characteristics of different tools in order to improve the performance of the hybrid model.

2. Materials and Methods

2.1. Data Collection. A questionnaire was developed by the authors of the paper to provide historical data relevant for the purpose of the research. The questionnaire was distributed in construction companies through personal visits of authors, meetings with company representatives, and construction site visits. Historical data, relevant for construction time prediction, were collected for a total of 116

different types of structures built since 2000. Database consists of road sections (27), petrol stations (4), bridges (7), education facilities (5), business buildings (28), residential-business buildings (10), sport halls (5), water tanks (4), residential buildings (4), water supply system sections (7), bridges (7), an overpass (1), a tunnel (1), traffic arteries (5), and other (8). The collected data refer to the structure type—purpose, year of construction, region of its location, contracted construction time and cost, and realized construction time and cost.

2.2. Process-Based and Data-Driven Models. *Process-based models* describe the process by examining two basic phases: *mathematical modelling* (mathematical equations) and *numerical solution*. In the mathematical modelling phase, the process is described by mathematical equations. After that, an accurate and efficient *numerical solution* of these equations follows. Process-based models have a wide range of applications because they are based on the theoretical understanding of relevant process and offer explicitly stated assumptions about the functioning of the process; therefore, they are used to guide the management decision under conditions of rapid global changes [47].

In order to develop process-based models, a very good understanding of the process is needed, together with accurate data that describe the process. In situations when process-based models cannot be built due to the lack of knowledge about the process which should be modelled, then *data-driven* models can be built. In such cases, some of the variables, which characterize the process are measured, and the data representing the input-output relationship describing the process should be available. Data-driven models (DDMs) can make it possible to predict some output variables. DDMs do not require a priori knowledge about the process and the laws under which the variables included in the process are connected. The only required knowledge is the factors that influence the process in order to identify which variables are relevant for the analysis. DDMs can supply important information extracted from the available data about the relationship of the variables in the process.

Recent developments in artificial intelligence, particularly computation intelligence and machine learning, have widened the capabilities of data-driven (empirical) modelling. Other research fields, which have contributed very much to improving conventional empirical (data-driven) modelling, are soft computing, data mining, and intelligent data analysis.

When process-based models are not adequate to model a particular situation or when the estimation of the parameters in the process-based models is difficult to obtain or is not precise enough, and when there are not enough data to train data-driven models, then a combination of models of different types can be an efficient solution. The research of hybrid modelling is trying to develop algorithms to obtain an efficient combination of data-driven and process-based models. This is a relatively new area of research, which has been examined in the last several years, giving important results.

Corzo et al. [48] have used a combination of process-based and data-driven models. They have obtained model performance improvement by reducing the error and increasing the model efficiency in hydrology at river flow simulation.

The authors Zhou et al. [49] have made a comparison of data-driven and process-based models for simulating HVAC (heating, ventilation, and air-conditioning) systems, analyzing their differences, and showing that the both models perform almost equally well for energy efficient control.

Computational intelligence and machine learning methods have developed the data-driven models, making them suitable for complementing or replacing the process-based models. The authors in [50] have shown that the data-driven method can sometimes outperform the process-based method because of the fact that in reality each process-based model is an approximation of the reality. The authors Rajabi et al. [51] have also shown in their research that the data-driven model RBFLN (based on RBFNN) has demonstrated best predictive accuracy in comparison with two knowledge-driven methods: Fuzzy AHP_OWA and Fuzzy GIS-based method.

Machine learning algorithms are used to determine the relationship between input and output of the system (predictors and target variable) using data set for training, which should be most representative for the behavior of the system. After training the model, it is tested on independent data set for validating how well it will generalize to new unknown data. The most important way for providing generalization to unknown data is choosing the most representative sample from the data set, which will present the whole behavior of the process [52]. In the last several years, the methods of artificial intelligence modelling have been used for improving and generating new and better process-based models from empirical data [53]. The combination of process-based and data-driven models, used in this paper for predicting construction time, is presented below.

2.2.1. Process-Based Model. Bromilow's "time-cost" model is used as a process-based model, giving the relation between the construction price and construction time (Eq. (1)) [54].

$$Y = a \cdot X^b, \quad (1)$$

Y is a construction time, X is a construction price, a is a parameter that expresses the average time needed for construction of a monetary value, and b is a model parameter that expresses the dependence of the time on the cost change [54].

This model has been tested, verified, and confirmed by many authors from many different countries around the world [13–15, 25]. According to Žujo et al. [13], one of the significant "time-cost" model limitations is that it can be applied only in the area or country of its origin because of specific economic characteristics, which are reflected on the value of model constants. Therefore, the existing models are not universally applicable and must be defined according to structure categories for each country separately. Consequently, similar studies have been conducted in many

countries in order to obtain adequate and corresponding time assessment models [15, 18, 35, 55, 56].

The most representative variables that influence the construction duration are as follows: type of structure, contracted time, real construction time, contracted price, and real construction price; these were chosen for building the model. Bromilow's "time-cost" model (Eq. (1)) is applied for contracted (planned) time and contracted (planned) price (Eq. (2)), and also for real price and real time of construction (Eq. (3)):

$$Y_1 = a_1 \cdot X_1^{b_1}, \quad (2)$$

$$Y_2 = a_2 \cdot X_2^{b_2}, \quad (3)$$

Y_1 is the planned (contracted) time for construction and X_1 is the contracted construction price, and Y_2 and X_2 are the real construction time and real construction price, respectively.

These equations ((2) and (3)) shall be logarithmized (Eq. (4) and Eq. (5)):

$$\ln(Y_1) = \ln(a_1) + b_1 \ln(X_1), \quad (4)$$

$$\ln(Y_2) = \ln(a_2) + b_2 \ln(X_2), \quad (5)$$

and Y_2 shall be expressed from Eqs. (4) and (5), by their summing (Eq. (6)):

$$\ln(Y_1) + \ln(Y_2) = \ln(a_1) + b_1 \ln(X_1) + \ln(a_2) + b_2 \ln(X_2), \quad (6)$$

$$\ln(Y_2) = \ln(a_1) + b_1 \ln(X_1) + \ln(a_2) + b_2 \ln(X_2) - \ln(Y_1). \quad (7)$$

Eq. (7) was used as a basic idea for the implementation of Bromilow's model in this research because it is linear in terms of the coefficients b_1 and b_2 and more suitable for the implementation in comparison with equations (2) and (3). Consequently, as input variables for the general regression neural network (GRNN), $\ln Y_1$, $\ln Y_2$, $\ln X_1$, and $\ln X_2$ were used, and not their actual values Y_1 , Y_2 , X_1 , and X_2 . In this way, Bromilow's time-cost model was implemented as input in the GRNN.

The importance of using Bromilow's time-cost model as a process-based model, which significantly improved the accuracy of the new hybrid model, should once more be pointed out.

2.2.2. Data-Driven Model. General regression neural network (GRNN) is used as a data-driven model using the process-based Bromilow model.

Neural networks (NN), as data-driven models, have proven their applicability in civil engineering in the last almost three decades, demonstrating very good solutions to many problems of civil engineering. NN are computational, biologically inspired models. Simulating the way of brain functioning, they learn from the experience. Using interconnected neurons, they perform input-output mapping.

The data enter the network through the neurons from the input layer. Then, they are fed forward through the middle (hidden) layer to the last output layer. The inputs are the variables that are the most representative for the process. NN capture the relationship between the actual input and output variables. NN are successful in solving a specific problem or modelling a particular process if substantial amount of data that describes the problem is available. Moreover, there should be no significant changes to the system or process that is being modelled [54]. For solving any problem with NN, the appropriate type of NN architecture should be selected because for different types of problems and available data, different types of NN architecture or data-driven model will be applicable.

For our investigation, several NN and other data-driven models were tried: linear regression, multilayer perceptron (MLP), support vector machine (SVM), RBFNN (radial basis NN), and GRNN. The most appropriate for our data was GRNN, performing the most accurate predicting.

General regression neural network (GRNN) can be applied for solving control problems, prediction, mapping, and any nonlinear regression problem [57]. The main characteristics that make it very applicable in practice are that in most of the cases, it is very accurate and it needs only several training samples to converge to the optimal solution. However, to store the model, it takes quite a lot of memory space. Some of the advantages of GRNN in comparison with other nonlinear regression models are that GRNN can generalize from the input data as soon as they are stored, learning in one pass through the data. Simulation of the GRNN algorithm is very simple, and because of the local minima of the error criterion, GRNN can converge to good solutions [58]. Figure 1 shows the architecture of GRNN.

The input layer has the same number of neurons as the number of predictors. The values of the predictors from the input neurons are fed to the neurons in the next pattern layer. The neurons from the pattern layer store the data for the rows from the training data set, each neuron for one row (case), and in this layer, for each new test case, Euclidean distance from the neuron's center is computed, RBF kernel function is applied, and that value is fed to the summation layer. The summation layer has two neurons: numerator summation unit and denominator summation unit. The numerator unit adds up the weight values multiplied by the actual value of the dependent (target) variable from each neuron from the pattern layer, and the denominator unit adds up the weight values from the neurons from the pattern layer. In the final decision layer (output unit), the predicted value of the target variable is computed by dividing the value from the numerator unit with the value from the denominator unit [57, 59].

GRNN implements the following equation (Eq. (8)) [57]:

$$y(X) = E\left[\frac{y}{X}\right] = \frac{\int_{-\infty}^{\infty} y f(X, y) dy}{\int_{-\infty}^{\infty} f(X, y) dy}. \quad (8)$$

$E[y/X]$ is the conditional expectation of the output (target variable) y for the given input X and $f(X, y)$ is the joint probability density function (jpdf) of the input vector X

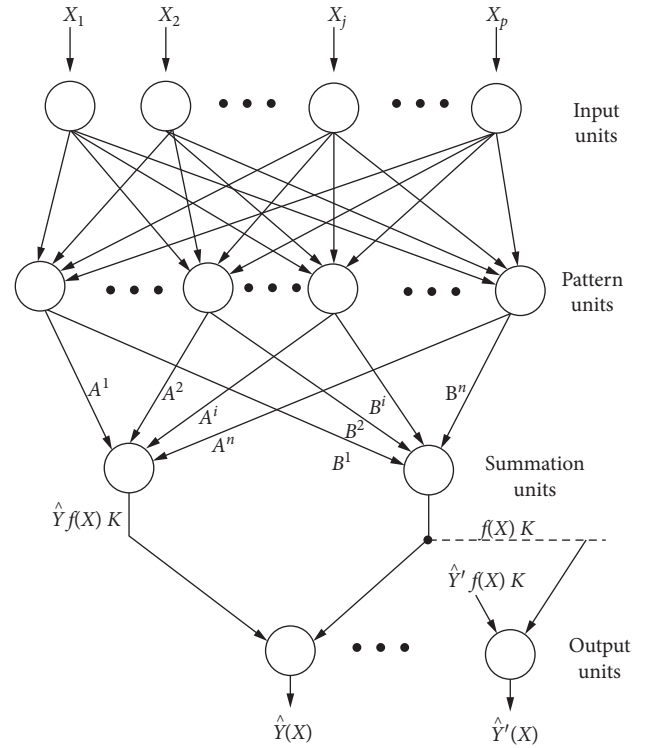


FIGURE 1: GRNN [57].

and the output y . When the function $f(X, y)$ is not known, Parzen estimators [60] are used for its estimation from the set of observations of X and y .

Specht [61], the author of GRNN, has made improvement of his first GRNN version, applying hybrid combination of three techniques: clustering, kernel regression with adaptive parameters, and the second level of clustering with the formation of a binary decision tree. These techniques have greatly improved the speed for training, the speed for readout and testing, and also the accuracy. These improvements have contributed to making GRNN useful for high-dimensional problems and for noisy data, too.

Hybrid modelling became very significant last several years, demonstrating very good predictive results.

Xiaojun et al. [58] proposed the Tree-Structure Ensemble GRNN (TSE GRNN), which consists of ensemble modelling methods using GRNN for predicting molten steel temperature in a ladle furnace, solving the larger scale issue. They obtained very good predictive results in comparison with other temperature models.

Lee et al. [62] present hybrid model developed for the classification of noisy data. The model, called GRNNFA, unites GRNN and a fuzzy adaptive (FA) resonance theory. It has been used for predicting the occurrence of the flashover in compartment fires and has demonstrated a very accurate prediction in comparison with other ANN models.

3. Results

Predicting modelling software DTREG [59, 63] has been used for modelling and predicting the construction time.

For building the predictive models in this research, among all available data, the purpose of structure, real time of construction, planned time of construction, planned price, and real price of construction have been chosen as the most representative. “Purpose of structure” is the categorical variable and the others are numerical ones.

At first, a data-driven model using only GRNN for predicting real time of construction has been developed. The real time of construction has been used as a target variable and the rest of the variables have been used as predictors. The accuracy of the model expressed by MAPE was around 30%.

After that, five new hybrid models implementing Bromilow’s TCM have been developed, and the most accurate among them was BTC-GRNN, combining Bromilow’s TCM and GRNN. According to the discussion in Section 2.2.1 (Eq. (7)), for all these five hybrid models, as numerical variables (predictors and target variable), logarithm of their values: \ln (real time), \ln (planned time), \ln (real costs), and \ln (planned costs) have been used, not their actual values. Equation (7) is used because it is more appropriate for implementation than Eqs. (2) and (3) because of its linearity in terms of the coefficients b_1 and b_2 which express the dependency of the time of construction (planned and real time, accordingly) on cost change (planned and real cost, accordingly).

Target variable is \ln (real time) and predictors are the remaining variables. The prices of all structures have been converted into euros, and the planned and real construction time for all structures has been expressed in working days. DTREG software operates with categorical variables, considering them as strings.

Usually, the data should be normalized before making the model and running the NN, but there was no need for normalizing the input data because the software DTREG does it for each predictive model.

DTREG software offers three methods for validation and testing of the model: the standard V-fold cross-validation, random percent validation, and “leave one out” (LOO) validation. DTREG also has an option for model optimization by removing unnecessary neurons from the NN. In this research, the model BTC-GRNN has been tested by three methods: first, by 10-fold cross validation, second, by using the option of “reducing number of neurons” with random percent, and third, by using LOO validation.

The accuracy of the model BTC-GRNN for the training and validation data using the 10-fold cross-validation method is presented in Table 1. The most frequently used estimators of the accuracy are MAPE, i.e., mean absolute percentage error and the coefficient of determination R^2 , which reflects the global fit of the model. For the BTC-GRNN model, MAPE is 3.34% and the coefficient of determination R^2 is 0.9317, which means that around 93.17% of the variation of the predicted target variable can be explained by the chosen predictors, whereas the remaining around 7% can be ascribed by some unknown variables or inherent variability. The coefficient of correlation between the actual and the predicted target values is 0.97.

The dependence between the actual and the predicted values for this model is presented in Figure 2 [63].

TABLE 1: Model accuracy results for BTC-GRNN model for (a) training data and (b) validating data (using 10-fold cross-validation) [63].

(a) Training data	
Mean target value for input data	5.1452765
Mean target value for predicted values	5.1456292
Variance in input data	1.1319931
Residual (unexplained) variance after model fit	0.0031972
<i>Proportion of variance explained by model (R^2)</i>	0.99718 (99.718%)
Coefficient of variation (CV)	0.010989
Normalized mean square error (NMSE)	0.002824
<i>Correlation between actual and predicted target values</i>	0.998588
Maximum error	0.2234627
RMSE (root mean squared error)	0.0565438
MSE (mean squared error)	0.0031972
MAE (mean absolute error)	0.0310945
<i>MAPE (mean absolute percentage error)</i>	0.6133842
(b) Validation data	
Mean target value for input data	5.1452765
Mean target value for predicted values	5.1156255
Variance in input data	1.1319931
Residual (unexplained) variance after model fit	0.0772845
<i>Proportion of variance explained by model (R^2)</i>	0.93173 (93.173%)
Coefficient of variation (CV)	0.054030
Normalized mean square error (NMSE)	0.068273
<i>Correlation between actual and predicted target values</i>	0.965664
Maximum error	1.2041075
RMSE (root mean squared error)	0.2780008
MSE (mean squared error)	0.0772845
MAE (mean absolute error)	0.1727287
<i>MAPE (mean absolute percentage error)</i>	3.34113

The model BTC-GRNN has also been tested by using the optimization option by reducing the number of neurons. In this case, for validation, the model random percent (16%) and the LOO method have been used. The optimal model has been obtained with only 70 neurons. The results for the estimators, i.e., MAPE and R^2 , using these 3 validation methods are presented in Table 2.

The accuracy of the other four hybrid models and the data-driven GRNN model shall be discussed in the next section.

For the numerical variables, DTREG computes their minimal, maximal, mean value and their standard deviation (Table 3) [63].

For each predictor, DTREG computes its importance for the quality of the model (Table 4) [63].

DTREG computes the target mean value for each different value of the predictors. Table 5 [63] shows the mean target values for only three different values of 4 predictors: purpose of facility, \ln (planned costs), \ln (real costs), and \ln (planned time). The first row for the predictor, “purpose of facility,” means that 4 rows, which is 3.45% of all 116 rows, have the same mean target value of 3.8491. The target value is \ln (real time), and the mean value for the real time in days can be easily computed. For the predictor \ln (planned time),

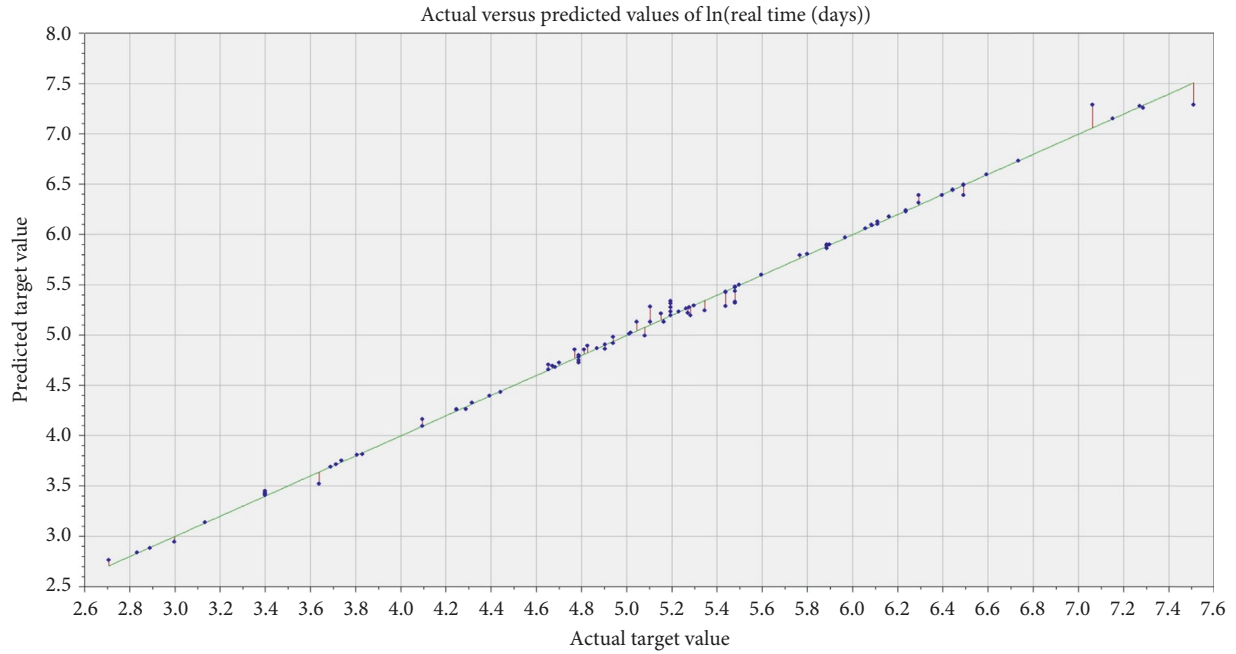


FIGURE 2: Dependence between actual and predicted values for the BTC-GRNN model.

TABLE 2: Accuracy results for the hybrid model BTC-GRNN using three validation methods.

Method for validation and testing	MAPE (%)	R^2 (%)	Coefficient of correlation between actual and predicted target values
10-fold cross-validation	3.34	93.17	0.966
Random percent (16%) (when using the optimization option by reducing the number of neurons)	3.76	93.72	0.970
LOO (when using the optimization option by reducing the number of neurons)	2.596	95.70	0.979

TABLE 3: Minimal, maximal, mean value, and standard deviation of input numerical variables for the BTC-GRNN model.

Variable	Rows	Minimum	Maximum	Mean	Standard deviation
ln(planned costs, euro)	116	9.05358	18.36511	13.27770	2.06661
ln(real costs, euro)	116	9.09459	18.51599	13.35837	2.11371
ln(planned time, days)	116	2.70805	6.81619	5.07076	1.04090
ln(real time, days)	116	2.70805	7.50934	5.14528	1.06395

TABLE 4: Overall Importance of variables for the BTC-GRNN model [63].

Variable	Importance
ln(planned time, days)	100.000
ln(real costs, euro)	28.396
Purpose of the facility	3.432
ln(planned costs, euro)	0.010

the first row means that two rows, which is 1.72%, of all 116 rows, have the same value of 2.70805, and their mean target value (ln(real time)) is 2.8519. For each predictor, only the first three rows are shown.

DTREG has a separate file from which the predicted values for the target variable can be read for each value of the predictors.

4. Discussion

As other neural networks, the GRNN learns from the input data, so their quality and quantity influence the prediction error. In this paper, the emphasis is on the importance of using the process-based Bromilow time-cost model combined with the data-driven model (GRNN). The reason is in the accuracy of the model that has been drastically improved. Moreover, for the model BTC-GRNN proposed in this paper, there was no need for computing the parameters of Bromilow's model.

Without using Bromilow's model and using only the actual values of the input variables: real time and planned time of construction and real price and planned price of construction as input data to GRNN, the model accuracy was tested by using three validation methods. Using the 10-

TABLE 5: Target mean values for three different predictor values [63].

Predictors			Target variable: ln(real time, days)
Purpose of the facility			Target mean
4	3.45%	Petrol station	3.8491
5	4.31%	Education	6.6249
7	6.03%	Bridge	4.5354
ln(planned costs, euro)			Target mean
1	0.86%	9.05358	3.6889
1	0.86%	9.35037	2.9957
1	0.86%	9.44293	2.8332
ln(real costs, euro)			Target mean
1	0.86%	9.09459	3.6889
1	0.86%	9.35037	2.9957
1	0.86%	9.44293	2.8332
ln(planned time, days)			Target mean
2	1.72%	2.708050	2.8519
2	1.72%	2.995732	2.8618
1	0.86%	3.218876	3.1355

fold cross-validation method, MAPE was 31.8, $R^2 = 75.64\%$, and the coefficient of correlation was 0.879 (Table 6). The other two validation methods, random percent (16%) validation and LOO validation, were used when the option for optimizing the model with reducing the number of neurons was applied. The results for the model accuracy, expressed by the most used estimators MAPE and R^2 , obtained by using these three validation methods are summarized in Table 7.

After developing the data-driven GRNN model, five hybrid models which implement process-based and data-driven models have been developed: BTC-SVM, BTC-LR, BTC-RBFNN, BTC-MLPNN, and BTC-GRNN, which combine Bromilow's TCM and SVM (support vector machine), LR (linear regression, RBFNN (radial basis function NN), MLPNN (multilayer perceptron NN), and GRNN, respectively.

The results for the obtained accuracy, expressed by MAPE and R^2 , using 10-fold cross-validation, are presented in Table 8. The most accurate was the BTC-GRNN model, as discussed in the previous section.

The model proposed in this paper has some limitations. Namely, the model is not applicable for higher, more intensive risk factors impact during the construction period (e.g., longer period with bad weather conditions, economic crisis, and high inflation). The project documentation is expected to be completed and corrected before the construction begins. It should be noted that some researches [17, 25] have shown that the problems with technical documentation (e.g., incompleteness and inaccuracy) sometimes cause delays in the construction process. It can be said that the model is applicable for the "normal" level of expected risk factors.

5. Conclusion

Construction time is one of the key elements in the bidding process and decision-making at the early phase of the construction project. However, at the same time, in this phase, the construction time prediction is a complex,

TABLE 6: Model accuracy results for (a) training and (b) validating data by using only data-driven GRNN model (using the 10-fold cross-validation method) [63].

(a) Training data	
Mean target value for input data	286.90647
Mean target value for predicted values	286.19978
Variance in input data	99430.392
Residual (unexplained) variance after model fit	2319.0624
<i>Proportion of variance explained by model (R^2)</i>	0.97668 (97.668%)
Coefficient of variation (CV)	0.167848
Normalized mean square error (NMSE)	0.023323
<i>Correlation between actual and predicted target values</i>	0.988279
Maximum error	300.05113
RMSE (root mean squared error)	48.156644
MSE (mean squared error)	2319.0624
MAE (mean absolute error)	21.681144
<i>MAPE (mean absolute percentage error)</i>	19.471364
(b) Validation data	
Mean target value for input data	286.90647
Mean target value for predicted values	285.08662
Variance in input data	99430.392
Residual (unexplained) variance after model fit	24222.689
<i>Proportion of variance explained by model (R^2)</i>	0.75639 (75.639%)
Coefficient of variation (CV)	0.542464
Normalized mean square error (NMSE)	0.243615
<i>Correlation between actual and predicted target values</i>	0.879069
Maximum error	1002
RMSE (root mean squared error)	155.6364
MSE (mean squared error)	24222.689
MAE (mean absolute error)	74.291454
<i>MAPE (mean absolute percentage error)</i>	31.79981

demanding task for project participants. Available project information is limited. Hence, using data from previous projects is of particular interest.

This paper presents research results of hybrid model development for early and fast construction time predicting using historical data. The model implements a combination of process-based model (Bromilow's time-cost model) and data-driven model (GRNN). Using 10-fold cross-validation, the mean absolute percentage error (MAPE) of the model is 3.34% and the coefficient of determination R^2 , which reflects the global fit of the model, is 93.17%. These results point to the drastic improvement of the accuracy when using only the data-driven GRNN model when MAPE is 31.8% and R^2 is 75.64%.

Such improved model can be successfully used at early project phases for a preliminary prediction of project duration with satisfactory accuracy. As such, it is not a substitution of detailed construction time planning.

For future researches, it is suggested to develop separate models for different types of structures and different projects characteristics (e.g., type of client, procurement characteristics, and type of contract). Homogeneous database will probably lead to improving the accuracy of the models. Moreover, the latest research and achievements in the area of

TABLE 7: Model accuracy results for validation data by using only data-driven GRNN model (using three methods).

Method for validation and testing	MAPE (%)	R^2 (%)	Coefficient of correlation between actual and predicted target values
10-fold cross-validation	31.8	75.64	0.879
Random percent (16%) (when using the optimization option by reducing the number of neurons)	32.47	83.75	0.952
LOO (when using the optimization option by reducing the number of neurons)	31.59	83.37	0.924

TABLE 8: Comparison of the accuracy of the five new hybrid models using 10-fold cross-validation.

Type of the predictive hybrid model	MAPE (%)	R^2 (%)	Coefficient of correlation between actual and predicted target values
BTC-SVM	3.9	91.63	0.958
BTC-LR	4.5	89.39	0.950
BTC-RBFNN	4.31	90.65	0.953
BTC-MLPNN	4.28	89.91	0.948
BTC-GRNN	3.34	93.17	0.966

the artificial intelligence, obtained by combining the process-based and data-driven models, can be of great significance for improving the accuracy of the predictive models.

Additionally, such models should be, in future research, considered as the part of integral building management information system. The reason lies in their characteristics and developmental potential. This would significantly solve the problems of wrong project decisions as the result of the initial faulty project time estimation.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors want to express their gratitude to Prof. Vojislav Kecman from the University of Virginia (USA) and Phil Sherrod, the author of the software DTREG, for their valuable consultations. This work was partly supported by the University of Rijeka (Grant number: 13.05.1.3.10).

References

- [1] H. M. Günaydin and S. Z. Doğan, "A neural network approach for early cost estimation of structural systems of buildings," *International Journal of Project Management*, vol. 22, no. 7, pp. 595–602, 2004.
- [2] Z. Ma, N. Lu, and S. Wu, "Identification and representation of information resources for construction firms," *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 612–624, 2011.
- [3] W. Tiziani and M. J. Mawdesley, "Advances and challenges in computing in civil and building engineering," *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 569–572, 2011.
- [4] R. J. Scherer and S.-E. Schapke, "A distributed multi-model-based Management Information System for simulation and decision-making on construction projects," *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 582–599, 2011.
- [5] A. Watson, "Digital buildings—challenges and opportunities," *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 573–581, 2011.
- [6] M. Joe, T. B. Keoughan, and I. Pegg, "Predicting construction duration of building projects," in *TS 28—Construction Economics I, Shaping the Change*, pp. 1–13, XXIII FIG Congress, Munich, Germany, 2006.
- [7] B. Dimitrov and V. Zileska-Pančovska, "Structure of price elements for construction works on water engineering systems," *Journal Gradjevinar*, vol. 67, pp. 363–368, 2015.
- [8] P. González, V. González, K. Molenaar, and F. Orozco, "Analysis of causes of delay and time performance in construction projects," *Journal of Construction Engineering and Management*, vol. 140, no. 1, 2014.
- [9] Y. A. Olawale and M. Sun, "Cost and time control of construction projects: inhibiting factors and mitigating measures in practice," *Construction Management and Economics*, vol. 28, no. 5, pp. 509–526, 2010.
- [10] R. N. Nkado, "Construction time information system for the building industry," *Construction Management and Economics*, vol. 10, no. 6, pp. 489–509, 1992.
- [11] S. M. Dissanayaka and M. M. Kumaraswamy, "Comparing contributors to time and cost performance in building projects," *Building and Environment*, vol. 34, no. 1, pp. 31–42, 1998.
- [12] A. Kaka and A. D. F. Price, "Relationship between value and duration of construction projects," *Construction Management and Economics*, vol. 9, no. 4, pp. 383–400, 2006.
- [13] V. Žujo, D. Car-Pušić, V. Zileska-Pančovska, and M. Čeček, "Time and cost interdependence in water supply system construction projects," *Technological and Economic Development of Economy*, vol. 23, pp. 895–914, 2017.
- [14] D. Car-Pušić and M. Radujković, "Modeli za brzu procjenu održivog, vremena građenja," *Gradjevinar*, vol. 58, pp. 559–568, 2006.
- [15] D. W. M. Chan and M. M. Kumaraswamy, "A study of the factors affecting construction durations in Hong Kong," *Construction Management and Economics*, vol. 13, no. 4, pp. 319–333, 1995.

- [16] K. Petlíková and Č. Jarský, "Modeling of the time structure of construction processes using neural networks," *Organization, Technology and Management in Construction: an International Journal*, vol. 9, no. 1, pp. 1559–1564, 2017.
- [17] D. Car-Pušić, "Methodology of planning the sustainable construction time," Dissertation, Faculty of Civil Engineering, University of Zagreb, Zagreb, Croatia, 2004, in Croatian.
- [18] H. P. Moura, J. C. Teixeira, and B. Pires, "Dealing with cost and time in the Portuguese construction industry," in *Proceedings of CIB World Building Congress*, pp. 1252–1265, Cape Town, South Africa, May 2007.
- [19] Y. Zhang and S. T. Ng, "An ant colony system based decision support system for construction time-cost optimization," *Journal of Civil Engineering and Management*, vol. 18, no. 4, pp. 580–589, 2012.
- [20] A. P. C. Chan and D. W. M. Chan, "Developing a benchmark model for project construction time performance in Hong Kong," *Building and Environment*, vol. 39, no. 3, pp. 339–349, 2004.
- [21] R. Kenley, *Financing Construction: Cash Flows and Cash Farming. The Time-Cost Relationship* Routledge, pp. 138–160, Taylor and Francis Group, London, UK, 2013.
- [22] S. Ng Thomas, M. M. Y. Mak, R. M. Skitmore, K. C. Lam, and M. Varnam, "The predictive ability of Bromilow's time-cost model," *Construction Management and Economics*, vol. 19, no. 2, pp. 165–173, 2001.
- [23] A. P. C. Chan, "Time-cost relationship of public sector projects in Malaysia," *International Journal of Project Management*, vol. 19, no. 4, pp. 223–229, 2001.
- [24] I. Choudhury and S. S. Rajan, *Time-Cost Relationship for Residential Construction in Texas*, Construction Informatics Digital Library, University of Ljubljana, Ljubljana, Slovenia, 2003.
- [25] D. Car-Pušić and M. Radujković, "Construction time-cost model in Croatia," *International Journal for Engineering Modelling*, vol. 22, pp. 63–70, 2009.
- [26] D. Mačková and R. Bašková, "Applicability of Bromilow's time-cost model for residential projects in Slovakia," *Selected Scientific Papers-Journal of Civil Engineering*, vol. 9, no. 2, pp. 5–12, 2014.
- [27] R. M. Skitmore and S. T. Ng, "Forecast models for actual construction time and cost," *Building and Environment*, vol. 38, no. 8, pp. 1075–1083, 2003.
- [28] R. M. Skitmor and S. T. Ng, "Australian project time-cost analysis: statistical analysis of intertemporal trends," *Construction Management and Economics*, vol. 19, no. 5, pp. 455–458, 2001.
- [29] A. P. C. Chan and D. W. M. Chan, "A benchmark model for construction duration in public housing developments," *International Journal of Construction Management*, vol. 3, no. 1, pp. 1–14, 2003.
- [30] J. F. Shr and W. T. Chen, "Functional model of cost and time for highway construction projects," *Journal of Marine Science and Technology*, vol. 14, pp. 127–138, 2006.
- [31] M. M. Kumaraswamy and D. W. M. Chan, "Determinants of construction duration," *Construction Management and Economics*, vol. 13, pp. 209–217, 1995.
- [32] A. O. Ayodeji, J. Smallwood, and W. Shakantu, "A linear regression modelling of the relationship between initial estimated and final achieved construction time in South Africa," *Acta Structilia*, vol. 19, pp. 39–56, 2012.
- [33] D. W. M. Chan and M. M. Kumaraswamy, "Forecasting construction durations for public housing projects: a Hong Kong perspective," *Building and Environment*, vol. 34, no. 5, pp. 633–646, 1999.
- [34] F. Khosrowshahi and A. P. Kaka, "Estimation of project total cost and duration for housing projects in UK," *Building and Environment*, vol. 31, no. 4, pp. 373–383, 1996.
- [35] V. Žujo and D. Car-Pušić, "Prekoračenje ugovorenog roka građenja kao funkcija rizičnih faktora," *Građevinar: Časopis Hrvatskog Saveza Građevinskih Inženjera*, vol. 61, pp. 721–729, 2009, in Croatian.
- [36] A. A. Abu Hammad, S. M. Alhaj Ali, J. S. Ghaleb, and A. Bashir, "Prediction model for construction cost and duration in Jordan," *Jordan Journal of Civil Engineering*, vol. 2, pp. 250–266, 2008.
- [37] M. G. Naik and V. S. B. Radhika, "Time and cost analysis for highway road construction project using artificial neural networks," *Journal of Construction Engineering and Project Management*, vol. 5, no. 1, pp. 26–31, 2015.
- [38] B. Vahdani, S. M. Mousavi, M. Mousakhani, and H. Hashemi, "Time prediction using a neuro-fuzzy model for projects in the construction industry," *Journal of Optimization in Industrial Engineering*, vol. 19, pp. 97–103, 2016.
- [39] S. Petruseva, V. Zujo, and V. Zileska-Pancovska, "Neural network prediction model for construction project duration," *International Journal of Engineering Research and Technology*, vol. 2, pp. 1646–1654, 2013.
- [40] I. Attarzadeh and S. H. Ow, "Software development cost and time forecasting using a high performance artificial neural network model," in *Intelligent Computing and Information Science: Communications in Computer and Information Science*, R. Chen, Ed., pp. 18–26, Springer, Berlin, Germany, 2011.
- [41] S. Petruseva, D. Car-Pusic, and V. Zileska Pancovska, "Model for predicting construction time by using general regression neural network," in *Proceedings of People Buildings and Environment, International Scientific Conference (PBE 2016)*, p. 31, Luhačovice, Czech Republic, September–October 2016.
- [42] I. Mensah, G. Nani, and T. Adjei-Kumi, "Development of a model for estimating the duration of bridge construction projects in Ghana," *International Journal of Construction Engineering and Management*, vol. 5, pp. 55–64, 2016.
- [43] V. Yousefi, S. Haji Yakhchali, M. Khanzadi, E. Mehrabanfar, and J. Šaparauskas, "Proposing a neural network model to predict time and cost claims in construction projects," *Journal of Civil Engineering and Management*, vol. 22, no. 7, pp. 967–978, 2016.
- [44] A. A. Gab Allah, A. H. Ibrahim, and O. A. Hagra, "Predicting the construction duration of building projects using artificial neural networks," *International Journal of Applied Management Science*, vol. 7, no. 2, pp. 123–141, 2015.
- [45] M. J. Roberts, N. O. Braun, T. R. Sinclair, D. B. Lobell, W. Schlenker, and W. Schlenker, "Comparing and combining process-based crop models and statistical models with some implications for climate change," *Environmental Research Letters*, vol. 12, no. 9, article 095010, 2017.
- [46] L. Liao and F. Köttig, "A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction," *Applied Soft Computing*, vol. 44, pp. 191–199, 2016.
- [47] K. Cuddington, M. J. Forth, L. R. Gerber et al., "Process-based models are required to manage ecological systems in a changing world," *Ecosphere*, vol. 4, no. 2, pp. 1–12, 2013.
- [48] G. A. Corzo, D. P. Solomatin, M. Hidayat et al., "Combining semi-distributed process-based and data-driven models in flow simulation: a case study of the Meuse river basin," *Hydrology and Earth System Sciences*, vol. 13, no. 9, pp. 1619–1634, 2009.

- [49] D. Zhou, Q. Hu, and C. Tomlin, "Model comparison of a data-driven and a physical model for simulating HVAC sustems," 2016, <https://arxiv.org/abs/1603.05951>.
- [50] S. Formentin, K. Heusden, and A. Karimi, "A comparison of model-based and data-driven controller tuning," *International Journal of Adaptive Control and Signal processing*, pp. 1–16, 2012.
- [51] M. Rajabi, A. Mansourian, P. Pilesjö, F. Hedefalk, R. Groth, and A. Bazmani, "Comparing knowledge-driven and data-driven modelling methods for suscepribility mapping in spatial epidemiology: a case study in Visceral Leishmaniasis," in *Proceedings of the AGILE 2014, Internatioanl Conference on Geographic Information Science*, Castellon, Spain, June 2014.
- [52] R. J. Abrahart, L. M. See, and D. P. Solomatine, *Practical Hydroinformatics: Water Science and Technology Library*, Springer-Verlag, Berlin, Germay, 2008.
- [53] K. Manhart, *Artificial Intelligence Modelling: Data Driven and Theory Driven Approaches (Revised Version of a Contribution Published in, Social Science Micro Simulation*, U. Troitzsch, U. Muller, G. Nigel, and J. E. Doran, Eds., Springer, Berlin 1996, Springer, Munchen, Germany, 2007.
- [54] F. J. Bromilow, "Contract time performance expectations and reality," *Building forum*, vol. 1, pp. 70–80, 1969.
- [55] O. Durson and C. Stoy, "Time-cost relationship of building projects: statistical adequacy of categorization with respect to project location," *Construction Management and Economics*, vol. 29, no. 1, pp. 97–106, 2011.
- [56] C. Sun and J. Xu, "Estimation of time for Wenchuan Earthquake reconstruction in China," *Journal of Construction Engineering and Management*, vol. 137, no. 3, pp. 179–187, 2011.
- [57] D. F. Specht, "A general regression neural network," *IEEE Transactions on neural networks*, vol. 2, no. 6, pp. 568–576, 1991.
- [58] W. Xiaojun, Y. Mingshuang, M. Zhizhong, and Y. Ping, "Tree-structure ensemble general regression neural networks applied to predict the molten steel temperature in Ladle Furnace," *Advanced Engineering Informatics*, vol. 30, no. 3, pp. 368–375, 2016.
- [59] P. Sherrod, "Predictive modelling software-tutorial," 2013, <http://www.dtreg.com>.
- [60] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.
- [61] D. F. Specht, "GRNN with double clustering," in *Proceedings of the International Joint Conference on Neural Networks*, Vancouver, Canada, July 2006.
- [62] E. W. M. Lee, Y. Y. Lee, C. P. Lim, and C. Y. Tang, "Application of a noisy data classification technique to determine the occurrence of flashover in compartment fires," *Advanced Engineering Informatics*, vol. 20, no. 2, pp. 213–222, 2006.
- [63] P. Sherrod, "Predictive Modelling Software," 2013, <http://www.dtreg.com>.

