Journal of Cleaner Production xxx (2015) 1–10

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Estimating water treatment plants costs using factor analysis and artificial neural networks

Mohamed Marzouk*, Mostafa Elkadi

Structural Engineering Department, Faculty of Engineering, Cairo University, Egypt

ARTICLE INFO

Article history:
Received 16 January 2015
Received in revised form
3 September 2015
Accepted 5 September 2015
Available online xxx

Keywords:
Parametric cost estimation
Water treatment plant
Cost driver
Descriptive statistics ranking
Principal component analysis
Artificial neural networks

ABSTRACT

The cost of construction project is a fundamental input for decision making process set by owner during procurement stage. The paper identifies the cost drivers that are used in parametric cost estimation model for water treatment plants. Cost estimation at planning stage of projects is important for the success of the next stages in the projects. It is also very useful at the design stage of a project when information is incomplete and detailed designs are limited in such an early stage. Literature has been reviewed and interviews were conducted with experts and officials in water treatment plants to explore all variables that influence the construction cost of water treatment plants. A questionnaire survey was then conducted to assess the impact of the identified factors on construction costs of water treatment plants. Datasets that consist of 160 water treatment plant projects in Egypt were collected. Construction cost drivers have been nominated through two different procedures. The first technique is descriptive statistics ranking of variables by evaluating Mean Score and Relative Importance Index based on respondents' feedback in conducting questionnaire. The second technique utilizes exploratory factor analysis on the collected dataset. These cost drivers are used to construct two predictive models for estimating the construction cost of water treatment plants models using artificial neural networks. Analysis of results was performed to validate the models and demonstrate their effectiveness. The proposed methodology aids public authorities to perform comparative analysis and evaluate the different alternatives of water treatment plant projects.

1. Introduction

Water treatment plants are classified as infrastructure projects that are usually administrated by public authorities. Such class of projects is critical and has different components including buildings, underground piping and equipment. In recent decades, Egypt witnessed construction of many water treatment plants. Early stage cost estimate plays a significant role in any initial water treatment plants projects decisions, despite the project scope has not yet been finalized and still very limited information regarding the detailed design is available. Major problems faced are lack of preliminary information, lack of database of water treatment plants construction costs, missing data, lack of appropriate cost estimation methods, and the involvement of uncertainties.

Water treatment plants' stakeholders in Egypt often need to estimate the construction costs of these plants at early stage readily and approximately to secure the required fund. Therefore, it is important to find a reasonable cost estimate tool for water treatment plants projects with acceptable level of accuracy. In Egypt, the estimating variability is high and can reach an average range of −37.8% to +28.56 within a contractor's company (Khorshid and Abdel-Razek, 1991). Due to the observed high value of estimating variability, this research was carried to develop a fast and reliable model using artificial neural networks.

Cost estimation is a heavily experience-based process that involves the evaluation of several complex relationships of cost-influencing factors, largely based on professional judgment (Alex et al., 2010). Parametric cost estimations technique is implemented in the early stage of a project. According to Project Management Body of Knowledge (PMBOK, 2013), parametric estimating is defined as a technique using a statistical relationship between

Abbreviations: EFA, Exploratory Factor Analysis; ANN, Artificial Neural Network; EFCBC, Egyptian Federation for Construction and Buildings Contracts; PCA, Principal Component Analysis; MSA, Measures of Sampling Adequacy; KMO, Kaiser-Meyer-Olkin; MLP, Multilayer Perceptron; SPSS, Statistical Package for the Social Sciences.

* Corresponding author.
E-mail address: mm_marzouk@yahoo.com (M. Marzouk).

http://dx.doi.org/10.1016/j.jclepro.2015.09.015
0959-6526/

relevant historical data and other variables to predict a cost estimate for project work. The parametric cost estimation models are used to express a dependent variable (cost) in terms of independent variables (parameters). Levels of accuracy for this technique can be highly depending upon the sophistication and underlying data built into the model. Conceptual cost estimates, also referred to as pre-design cost estimates, are prepared at the very early stages of a project, and generally before the construction drawings and specifications are available. At this stage, cost estimates are needed by the owner, contractor, designer, or lending organization for several purposes, including determination of the feasibility of a project, financial evaluation of a number of alternative projects, or establishment of an initial budget. Conceptual cost estimates are not expected to be precise, since project scope is not finalized and very limited design information is available during the pre-design stages of a project. A quick, inexpensive, and reasonably accurate estimate is needed, however, based on the available information. Several techniques have been suggested for conceptual cost estimation. Regression analysis, simulation, and neural networks are among these cost estimation techniques that are used during the early project stages.

Artificial Neural Network (ANN) model is a computational model that is inspired by the structure and (the) functionality of biological neurons. It is used as non-linear statistical data modeling tool to model complex relationships between inputs (independent variables) and outputs (dependent variables). ANNs are the preferred tool for many predictive data mining applications because of their power, flexibility and ease of use. Predictive neural networks are particularly useful in applications where the underlying process is complex, such as cost estimation. Therefore, it is an appropriate tool to solve various estimation and prediction problems in the construction engineering and management field. It can model uncertainties such as cost estimation at the early stage. Many researches applied ANN in construction cost estimation field.

There are large number of factors that affect the cost of water treatment plant projects as well as their life cycle assessment (Mahgoub et al., 2010; Renou et al., 2008, Klemeš and Huisingh, 2008). The aim of this research is to identify the cost drivers of water treatment plant projects using two procedures; Descriptive Statistics Ranking (DSR) and Exploratory Factor Analysis (EFA). Subsequently, ANN modeling is conducted to estimate construction cost of water treatment plants in Egypt based on the results of two abovementioned procedures of cost factors derivation. Different architectures of ANN are investigated to select the one that achieves best accuracy.

This paper is structured as follows; Section 2 briefly reviews the literature on ANN to predict the cost of construction projects. Section 3 presents a detailed overview of the methodology. Section 4 discuses variables identification, data gathering and preparation for next stages of this research. Section 5 covers the process of identifying the cost drivers for Water treatment plants using descriptive statistics ranking through questionnaire survey. Section 6 introduces Exploratory Factor Analysis (EFA) using principal component analysis (PCA) with VARIMAX rotation that adopted in the research methodology to isolate statistically the large number of variables that influence the construction cost of water treatment plants. Section 7 describes the neural network model that has been developed to predict the construction costs of water treatment plants. Finally, the paper concluded remarks and results are discussed in Section 8.

2. Literature review

Several research efforts have been conducted that use ANN to predict the cost of construction projects. Shehab and Farooq (2013) developed an ANN cost-estimating model for utility rehabilitation projects using data of 54 sewer and water rehabilitation projects that were built in the city of San Diego, California, USA. The study identified a set of 23 factors that highly influence the construction cost of water and sewer network rehabilitation projects and presented an ANN model that predicts the cost of these projects with high accuracy. Ahiaga-Dagbui and Smith (2012) incorporated a number of factors including site conditions, procurement, risks, price changes, likely scope changes or type of contract. ANN for modeling the final target cost of water projects and data was collected on ninety-eight water-related construction projects completed in Scotland from 2007 to 2011. A model for cost estimating of Road Tunnel Construction was proposed by Petroutsatou et al. (2012) using two types of neural networks: the multilayer feed-forward network and the general regression neural network; they collected data from 33 tunnels of 46 km total length constructed for the Egnatia Motorway in northern Greece from 1998 to 2004. In 2012, the factors that influence contingency networks were identified by Lhee et al. (2012) who proposed a new method for predicting the owner's financial contingency on transportation construction (asphalt resurfacing projects) using ANN. The data were collected from transportation projects sponsored by the Florida Department of Transportation and completed from 2004 to 2006.

ElSawy et al. (2011) used an ANN approach to develop a parametric cost-estimating model for site overhead cost in Egypt. Fifty-two actual real-life cases of building projects constructed in Egypt during the seven year period 2002–2009 were used as training materials. The ANN architecture is presented for the estimation of the site overhead costs as a percentage from the total project price. A model to estimate the cost of building construction projects at early stages using ANN was developed by Arafa and Alqedra (2011). The model was constructed using a database of 71 buildings in the Gaza Strip. Several significant parameters for the structural skeleton cost of the project can be obtained from available engineering drawings and data at the pre-design stage of the project. The input layer of the ANN model comprised seven parameters, namely: ground floor area, typical floor area, number of stores, number of columns, type of footing, number of elevators and number of rooms. Another model for life-cycle maintenance planning of deteriorating sewer network was presented by Marzouk and Omar (2013) as a multi-objective optimization problem that treats the sewer network condition and service life as well as life-cycle maintenance cost as separate objective functions. This model utilizes Markov chain model for the prediction of the deterioration of the network. Marzouk and Amin (2013) presented a methodology to be used in modifying the contract price with an attempt to predict the amount of future change in materials prices using ANN technique. Cost items were classified into four different components (building materials, equipment, labor, and administrative expenses.

Alex et al. (2010) provided a detailed analysis of all activities involved in the installation of the water and sewer services. They proposed a methodology that was based upon the analysis of past data obtained from the City of Edmonton's drainage. The methodology has been incorporated into a computer module, which integrated the concept of ANN with the current estimating system used by the City of Edmonton while Bouabaz and Hamami (2008) presented an estimation model for repair and maintenance of bridges in developing countries using ANN. Kim et al. (2005) applied hybrid models of ANN and Genetic Algorithms in order to predict cost estimation of residential buildings to predict preliminary cost estimates. Using data of residential buildings constructed from 1997 to 2000 in Seoul, Wilmot and Mei (2005) predicted the escalation of highway construction costs over time. An ANN model was

Please cite this article in press as: Marzouk, M., Elkadi, M., Estimating water treatment plants costs using factor analysis and artificial neural networks, Journal of Cleaner Production (2015), http://dx.doi.org/10.1016/j.jclepro.2015.09.015