Regression Analysis-Based Cost Estimation of Simulation Models in Model Cost Database

Hae Young Lee¹, So Jin Lee², Hyung-Jong Kim¹, and SeungHyun Byun¹

¹ Seoul Women’s University
01797 Seoul, Republic of Korea

² unaffiliated
06900 Seoul, Republic of Korea
whichmeans@gmail.com

Abstract. The effectiveness of simulation-based acquisition, which is a robust, collaborative use of modeling and simulation (M&S) technology throughout acquisition phases and programs, should be estimated quantifiably, even in a rough order of magnitude manner, with the consideration of M&S overhead. This paper presents a simulation model cost database that can estimate the development costs of simulation models based on regression analysis. When simulation model that have been newly developed throughout acquisition programs are inserted into the database, size-to-cost trends are automatically computed for all data sets in the database. Based on these trends, the development costs of simulation models to be developed can be estimated if their sizes can be roughly estimated. The database may be used in parts of the effectiveness estimation of the use of modeling and simulation technologies in acquisition programs. Moreover, it may be used to verify various methods of simulation model size measurement and find out the most promising ones. We have implemented a prototype of the database and are working on the implementation of a regression analysis based estimation module.

Keywords: Simulation based acquisition, modeling and simulation, model development cost estimation, model size estimation, software sizing

1 Introduction

Simulation based acquisition (SBA) is a robust, collaborative use of modeling and simulation (M&S) technology that is integrated across acquisition phases and programs [1], which could enable, by making design alternatives analyzed within risk-free virtual environments, defense systems, e.g., weapon systems, military aircrafts, and so on, to be acquired within limited budget as well as other constraints in time, quality, safety, and so on [2]. However, such benefits from the use of SBA are intuitive. Therefore, in order to clearly show the benefits, the effectiveness of SBA should be estimated quantifiably, even in a rough order of magnitude manner [1].

¹ This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract UD110006MD, Korea.
Another consideration in the quantifiable estimation of SBA effectiveness is overhead of using M&S, i.e., extra costs of simulation model development, those of simulation infrastructure construction (e.g., run-time infrastructure), and so on [3].

The development cost of a simulation model can be calculated based on its size, which can be measured using all or some of its components, e.g., sets of I/O, state set, transition relation or function, and so on [4]. For example, the size of a simulation model developed based on Discrete Event System Specification (DEVS) [5] can be computed using the cardinality of its state set, and/or the number of phases, and so on [4]. For a simulation model for a hybrid system, which has both continuous and discrete components, both of the components can be considered in its size measurement [6]. Possibly, we may use software development effort estimation methods, such as function point analysis [7], to estimate the sizes of and the development costs of simulation models. These efforts, however, need to be verified thoroughly and compared with each other.

In this paper, we present a simulation model cost database (SMCDB), which is based on our first concept of SMCDB [8], with a regression analysis based estimation of simulation model development costs. Simulation models developed throughout an acquisition program are inserted into SMCDB, together with their development costs and sizes that can be measured by various methods of simulation model size measurement, e.g., [3,4,6]. By using simple linear regression, SMCDB computes size-to-cost trends based on all data sets in SMCDB. Then, the development costs of simulation models to be developed can be estimated using these size-to-cost trends if the sizes of the models could be ‘roughly’ estimated. SMCDB may be employed in parts of the estimation of SBA effectiveness since it can estimate extra costs of simulation model development. Also, it may be used to find out which of methods of model size measurement is the most promising one, i.e., we may use it to verify them. We have implemented a prototype of SMCDB and are working on the implementation of the regression analysis-based estimation.

2 SMCDB with Regression Analysis Based Cost Estimation

This section describes SMCDB with the simple linear regression based estimation of simulation model development costs in detail.

2.1 Overview

Fig. 1 shows a conceptual model of the proposed SMCDB. While a system has been developed through an acquisition program, multiple simulation models used throughout the program are developed. Each of these simulation models is then inserted into SMCDB, along with its development costs, in terms of multiple factors, such as development expense, development time, human resources, and so on, and sizes measured by various methods of simulation model (or software) size measurement, such as a method for discrete or continuous system models [4], that for hybrid system models [6], and so on. These data sets in SMCDB can be reviewed through tables or graphically. For all data sets in SMCDB, it uses simple linear
regression to compute size-to-cost trends. Also, using these size-to-cost trends, the development costs of simulation models to be developed can be estimated if the estimated sizes of them can be given in a rough order of magnitude manner.

![Diagram](image)

**Fig. 1.** Basic concept of the simulation model cost database (SMCDB)

### 2.2 Data Insertion

Once a simulation model has been developed through parts of an acquisition program, the user inserts the model into SMCDB, along with its development costs and sizes. The development costs of a single model can be comprised of multiple costs factors, such as development expense, development time, human resources, and so on. The sizes of a model can be also measures using various methods of software size measurement, such as function point analysis [7], object points [9] and so on, or of simulation model size measurement, such as a method for discrete or continuous system models [4], that for hybrid system model [6], and so on. The size measurement of a model can be automated; i.e., when the user inserts a model into SMCDB, the sizes of the model can be automatically calculated by add-ons for SMCDB and then inserted into SMCDB.

Then, SMCDB uses simple linear regression (SLR) to fit a straight line through each size-to-cost data set in SMCDB in such a way that makes vertical distances between the points of the data set and the fitted line as small as possible. SLR is used for each type of model sizes and for each type of model development costs. Suppose that a row (tuple) in SMCDB includes two types of mode sizes (say, SIZE1 and SIZE2) and two types of model development costs (COST1 and COST2). When a model is inserted into SMCDB along with these types of sizes and costs, SMCDB uses SLR to fit four lines through four data sets (SIZE1-to-COST1, SIZE1-to-COST2, SIZE2-to-COST1, and SIZE2-to-COST2) in SMCDB. In order to increase performance, computations related to SLR can be done when SMCDB is not accessed (e.g., in the middle of every night), rather than whenever a new data set has been added.
2.3 Data Retrieval and Cost Estimation

Data sets in SMCDB can be reviewed through tables or graphically. Lines that have been computed using SLR and show size-to-cost trends, referred to as ‘size-to-cost trend lines (SCTLs)’ hereafter, can appear on graphical reviews. These SCTLs can be used to roughly estimate the development costs of models to be developed. The development costs of models to be developed can be estimated more accurately by specifying the estimated sizes of the models. Then, SMCDB computes the estimated costs of the models based on the estimated sizes and SCTLs.

2.4 Applications of SMCDB

Although SMCDB is basically designed to manage and estimate the development costs of and the sizes of simulation models, it may be also used, as shown in Fig. 2(a), to determine whether M&S technologies can be applied to an acquisition of a system, i.e., it may be used as parts of SBA effectiveness estimation. For an acquisition of a system, if the estimate costs of simulation models that should be newly developed throughout the acquisition program do not exceed the estimated benefits from the use of modeling and simulation techniques within the program (i.e., the estimated benefits from an SBA-based acquisition), it will be better to apply SBA to the program.

![Diagram](image)

**Fig. 2.** Examples of SMCDB applications.

Also, SMCDB may be used to evaluate methods of software or simulation model size estimation measurement; once SMCDB has got a large number of data sets, it may be used to find out which of the methods was the most promising one.
3 Implementation

We have implemented a prototype of SMCDB. The prototype allows models developed based on DEVS using a DEVS-based modeling environment [10]. It parses simulation model files produced by [10], chosen by the users through its UI, and computes the sizes of the models using four types of methods of simulation model size measurement: I/O-based, state-space-based, transition-based, and phase-based. Thus, the user needs not to input the sizes.

Data sets inserted into SMCDB can be reviewed through its UI shown in Fig. 3. As shown in the figure, they can be shown in a table or as graphs. Currently, we are working on the implementation of simple linear regression for the cost estimation. Once it has been implemented, SCTLs will be shown on the graphs.

![Query interfaces of our SMCDB prototype.](image)

5 Conclusions and Future Work

This paper presented SMCDB that supports the estimation of simulation model development costs based on SCTLs. Simulation models, their sizes and costs are
inserted into SMCDB. SMCDB then uses SLR to obtain SCTLs from data sets. Based on SCTLs, the development costs of simulation models to be developed can be estimated. SMCDB may be used in the effectiveness estimation of SBA and to verify simulation model size measurement methods. A prototype of SMCDB has been implemented, while a regression analysis based estimation module is under development.

The future version of SMCDB will support other types of simulation models, e.g., Simulink [11], and additional methods of simulation model size measurement. Also, we will perform some case studies with SMCDB in order to show its feasibility. Finally, we will try to accumulate data sets for SMCDB, so that it can be used in parts of SBA effectiveness evaluation as well as verification of model size measurement methods.

References