Proceedings of the 2005 Winter Simulation Conference
M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.

THE USE OF SIMULATION FOR PRODUCTIVITY ESTIMATION BASED ON MULTIPLE REGRESSION ANALYSIS

Seungwoo Han

Construction Management Program
Georgia Southern University
Forest Drive, Carruth Bldg.
Statesboro, GA 30460-8047, U.S.A.

Daniel W. Halpin

Division of Construction Engineering & Management
Purdue University
550 Stadium Mall Drive
West Lafayette, IN 47907-2051, U.S.A.

ABSTRACT

Productivity estimation has been fundamental subject investigated in academia and industry. There are two common methods for estimation of productivity: (1) deterministic and (2) simulation methods. The deterministic method does not reflect actual conditions, such as randomness of work duration, whereas simulation method can overcome this limitation. However, the user without a background in simulation may struggle with implementation due to the difficulty of modeling. The presented productivity estimation model in this research was created using multiple regression analysis with data generated by WebCYCLONE. The model representing the mathematical relations between conditions and productivity allows planners or site personnel to estimate productivity by simply entering input data reflecting actual site conditions. In academia, the research methodology utilized in this research provides a framework for the user to establish other application models for estimating or evaluating the performances of new technologies.

1 INTRODUCTION

In general, one of the most important tasks confronting planners in the construction industry is performance estimation of operations prior to commencement of construction. Productivity has been used as one criterion for explaining operational performance. Earthmoving is a fundamental construction operation and productivity estimation of earthmoving operations has provided people in both academe and industry with an important subject for research.

Planners have relied upon three methods to estimate productivity based on: (1) historical data; (2) references, such as R.S. Means and equipment handbooks; (3) particular methods such as construction simulation or statistic analysis. The method based on historical data or references is typically referred to deterministic analysis.

2 CONVENTIONAL PRODUCTIVITY ESTIMATION METHODS

2.1 Deterministic Analysis

Deterministic analysis was developed for simple calculation of the productivity of an earthmoving operation based on the equipment characteristics, equivalent grades, and the haul distance provided by performance handbooks published by most manufacturers. A deterministic model primarily focuses on the use of time durations that are fixed or constant values, with the assumption that any variability in the task duration is assumed to be ignored (Halpin et al. 1992).

Halpin et al. (1992) describes an example of a simple deterministic model for earthmoving operations, consisting of a scraper for a hauling and a pusher dozer for a loading. The deterministic durations for the scraper travel times to and from the fill location are available by using simple nomographs. Deterministic analysis tends to overestimate actual field productivity.

2.2 Simulation Methods

With rapid advances in computer technologies, researchers have tried to create simulation models to help construction engineers estimate construction productivity prior to commencing actual activities. Simulation models have been extensively developed and broadly used as a management tool within the manufacturing and business industries.

The CYCLONE (CYCLic Operation Network) system approach was developed in the early 1970s. This system demonstrated the potential for modeling and simulation of repetitive construction processes. In 1982, Lluch and Halpin developed a microcomputer version of CYCLONE named MicroCYCLONE. Many improvements to MicroCYCLONE have been developed in the past two decades.

Construction simulation has been broadly adapted for analysis of repetitive construction processes ranging from a mason supply system to a real project, such as the Isle of Palm Connector Bridge. In general, construction simulation is conducted in several steps (i.e., site observation, duration and resource data collection, modeling using CYCLONE, running simulation, and sensitivity analysis) (Kannan 1997; Wang 2004).

3 MULTIPLE REGRESSION ANALYSIS

In compliance with the need for a new productivity estimation tool to overcome the limitations of both deterministic and simulation models, a multiple regression model has been developed.

Regression analysis is the most commonly performed statistical procedure for prediction of certain tendencies based on observed datasets. The ultimate goal of regression analysis is not only to find the values of parameters, but also which type of mathematical function fits best. Using this tool, researchers have been able to investigate and understand the relationships between the so-called explanatory variables and a result called a response variable.

In the linear regression model, the response variable is assumed to be a linear function of one or more explanatory variables associated with error. The response variable can also be estimated by curvilinear functions interacting with multiple explanatory variables in a nonlinear regression model. Several examples of equations are shown below:

1. Linear regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

2. Nonlinear regression models:
   - Quadratic model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i22} + \cdots + \beta_p X_{ip2} + \varepsilon_i$$

   - Exponential model:

$$Y_i = \beta_0 + \beta_1 \exp(X_{i1}) + \cdots + \beta_p \exp(X_{ip}) + \varepsilon_i$$

   - Periodic model:

$$Y_i = \beta_0 + \beta_1 \sin(X_{i1}) + \cdots + \beta_p \sin(X_{ip}) + \varepsilon_i$$

- Y_i is the response variable corresponding the explanatory variables $x_1, \cdots, x_p$ at the *i* th observation.
- β is the coefficient of each explanatory variable. In the single linear regression model, β_0 indicates the intercept and β_1 does the slope.
- ε_i indicates a normally distributed random error (Neter et al. 1996).

In order to create a best-fitted multiple regression model, several concerns must be taken into account: (1) correlation between explanatory variables; (2) relationships between the predicted variable and the residuals; (3) residual variance and R-square; and (4) correlation coefficient R (Devore 2000).

4 DATA GENERATION BASED ON EXPERIMENTAL DESIGNS

The acquisition of the large input datasets associated with actual site conditions is fundamental to creating a new productivity estimation model based on a multiple regression model. Actual data acquisition from construction sites is limited due to the characteristics of the construction industry. Each construction site produces different and sometimes uneven data because of the uniqueness of construction site conditions. Generated data reflecting actual situations can be used in a multiple regression model. With reference of Kelton (2003) and Wang (2004) recommend the use of simulation to generate input datasets.

Simulation allows users to find and estimate outputs by considering various inputs. To do so effectively, careful planning of the model design is necessary. This careful planning of how models are to be used is important (Kelton et al. 2003).

A variety of approaches, methods, and analysis techniques, known as experimental design, have been introduced and documented by many researchers over the past 30 years. According to Kelton and Barton (2003), *one of the principal goals of experimental design is to estimate how changes in input factors affect the results, or responses, of the experiment.*

To illustrate this technique, it is supposed that two values, or levels of each input factor, should need to be identified. If there are *k* input factors, 2^k different combinations of the input factors and each defining a different configuration of the model can be reviewed, which is called a 2^k factorial design. A design matrix is then formed including "+" level (i.e., representing the optimistic conditions) and "-" level (i.e., representing the pessimistic conditions). In the case of three input factors to be studied, there are $2^3 = 8$ configurations (Kelton et al. 2003; Wang et al. 2004).

The main effect of Factor 2 in the above example is defined as the average difference in response when this factor moves from its "-" level to its "+" level. Consequently, the main effect of Factor 2 is described as $(-R_1 - R_2 + R_3 + R_4 - R_5 - R_6 + R_7 + R_8)/2^{k-1}$. The main effect of interaction can be produced with the same procedure as that of an individual factor. For instance, the main effect of interactions of factors 1 and 3 can be achieved by the formula that multiplies the columns of factors 1 and 3 and adds them and then divides by 2^{k-1}. That is, $(+R_1 - R_2 + R_3 - R_4 - R_5 + R_6 - R_7 + R_8)/2^{k-1}$ (Kelton et al. 2003).

The problem of conducting full factorial experimental design is that if the number of factors becomes even moderately large, the number of runs extremely increases as 2^k. Thus, it would be most helpful to identify early in the course of experimentation which factors are important and which are not. The unimportant factors can then be fixed at reasonable values and dropped from consideration, and further experimentation can be conducted on the important factors, which is called factor-screening design. When experimental design is successfully completed, the validation of the model by experimental design is considered. In general, an algebraic regression-model is utilized. The best and the worst combinations are tested using a regression model that is created based on the experimental design and compared with the result by responses, which is conducted by a simulation model in this research (Kelton et al. 2003; Wang et al. 2004).

5 IMPLEMENTATION

5.1 Data Acquisition

As the first phase, data collection was conducted in construction sites where earthmoving operation was executed in Indiana from May through November in 2003. Table 1 describes six construction projects where data collection was conducted.

Table 1: Descriptions of Earthmoving Projects

Project Name	Location	Fleet Organization	Haul Distances (miles)
Purdue Nano Technology Center Phase 1	W.Lafayette, IN	1 Excavator, 7 Trucks	3
Stadium Avenue Reconstruction	W.Lafayette, IN	1 Excavator, 1 Dozer 2 Trucks	2.9
Discovery Park Road Construction	W.Lafayette, IN	1 Excavator, 4 Trucks	15.8
Purdue Nano Technology Center Phase 2	W.Lafayette, IN	1 Excavator, 10 Trucks	4.8
West Lafayette Trail Road Construction	W.Lafayette, IN	1 Excavator, 2 Trucks	1.1
Ace Hardware Building Construction	Lafayette, IN	1 Excavator, 7 Trucks	9.4

From the projects described in Table 1, data per hour were collected for four or five hours in two or three days at each jobsite. Total 23 separate hourly data were collected Accordingly, each hourly dataset covered a period of multiple cycles. Each dataset represents a remarkable sample of earthmoving operations in both a two-link system composed of an excavator and trucks and a three-link system composed of an excavator, a dozer, and trucks due to its project size, different fleet managements, and varied conditions.

The pictures that were recorded from the jobsites provided consistent observations for analyzing the event times of each piece of equipment using stop watch analysis. The basic information about sieve analysis was conducted using soil samples that were taken from the jobsites in order to investigate the soil characteristics. Through this observation and analysis, the travel time, loading time, machine break time and surveying time were acquired. The basic conditions of the jobsite, such as haul distance, capacities of excavator buckets and trucks, and the number of pieces of equipment, were established as well. Table 2 is a summary of the data collected from the selected jobsites.

Table 2: Summary of Data Characteristics Collected from the Jobsites

Site Observation		Field Measurement	Calculation
Stopwatch Analysis based on Videotaping	Interviews		
Machine break time	Bucket capacity	Soil condition	Hauling speed
Survey time	Truck capacity	Hauling distance	Productivity
Loading time	Number of equipment		
Travel time	Operators' experience		
Number of loading	Age of equipment		

Figure 1 demonstrates simulation modeling based on one dataset collected from Ace Hardware Construction Project. This simulation model is designed to measure the productivity in terms of truck-dump per hour. During excavation process, 8.93 % of interruption by on-site surveyor was observed. This interruption was due to restaking the knock down stacks by surveyors. This kind of interruption was usually observed in all sites where earthmoving was operated. Conducting simulation reflecting this actual conditions and interviews with site personnel indicate that this interruption causes the delay of cycle time and eventually provides lower productivity. All the durations of each task were obtained by the combination between video taping and stop watch analysis. The durations associated with various cycle times, such as loading the earth to truck, trucks' traveling to dump location and returning were assumed to be fit to beta distribution. AbouRizk's paper (1992) recommended that the beta distribution be used in modeling random input processes of construction durations for simulation studies.

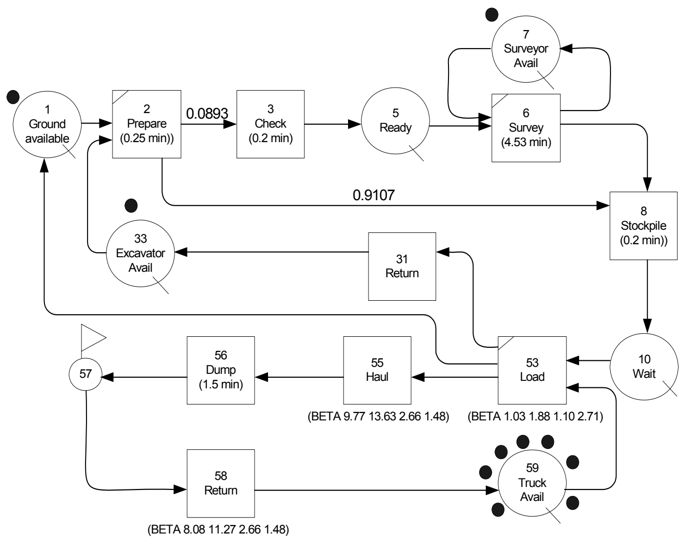

Figure 1: CYCLONE network of the Simplified Earthmoving Operation in Ace Hardware Construction Project

5.2 Data Generation

5.2.1 Major Input Factors and Estimated Effects

The fractional factorial experimental design is used for providing guidelines to generate input datasets for application models. One dataset collected from Ace Hardware Construction as jobsite is used as a sample experiment.

In the first phase, the main factors significantly impacting productivity were determined. In this research, a maximum of four main factors was considered since the number of experimental combination was increased at the rate of 2^k. Through interviews and site observations, the four main factors were determined: (1) surveying probability, (2) number of trucks, (3) number of excavators, and (4) surveying time. Table 3 shows the main factors and the low and high levels of each factor. The low and high levels of each factor were determined by checking the availability of resources on jobsites based on interviews, and the actual values presented the real value associated with one dataset for the experiment.

Based on Table 4, the effects of each main factor and each interaction associated with the main factors were estimated, following the equation suggested in the experimental design manual. The estimated effects are shown below in Table 5.

Table 3: Main Factors for the 2^k Experimental Design

Factor		Low (-)	High (+)	Actual Value
Surveying probability	A	0.05	0.3	0.0893
Number of trucks	B	5	10	7
Number of excavators	C	1	3	1
Surveying time (min.)	D	4	20	4.53

A design matrix and response for experimental design based on identification of main factors were conducted. Table 4 describes the design matrix and response, and note that the response actually the productivity derived from the simulation results using each specified main factor.

Table 4: Design Matrix and Responses

Run	A	B	C	D	Prod.
1	-	-	-	-	11.71
2	+	-	-	-	11.46
3	-	+	-	-	22.19
4	+	+	-	-	18.74
5	-	-	+	-	11.72
6	+	-	+	-	11.48
7	-	+	+	-	22.34
8	+	+	+	-	19.06
9	-	-	-	+	10.00
10	+	-	-	+	6.70
11	-	+	-	+	15.48
12	+	+	-	+	7.55
13	-	-	+	+	10.04
14	+	-	+	+	6.75
15	-	+	+	+	15.64
16	+	+	+	+	7.65

Table 5: Estimated Effects

Factors	Effects	Factor	Effects
D	6.1123	BC	0.0748
B	6.0974	BCD	0.0364
A	3.7184	ABCD	0.0307
BD	2.8916	ACD	0.0302
AB	1.9439	CD	0.0197
AD	1.9116	AC	0.0176
ABD	0.3840	ABC	0.0119
C	0.1041		

The results from checking the estimated effects indicate that factor D (surveying time) provide the most impact on the productivity and factor B (number of trucks) and factor A (probability of surveying) followed.

5.2.2 A Regression Model for Experimental Designs

Most experimental designs, including those mentioned above, are based on an algebraic regression model assumption about the way the input factors used as main factors affect the outputs used as the productivity. The SAS program using input data and output data mentioned above allowed finding the best fitted regression model as follows:

$$Y = 4.02129 + (1.81254 \cdot B) + (-0.18042 \cdot D) + (0.96546 \cdot AD) + (-0.28228 \cdot ABD)$$

The capability and availability of this regression model created based on the specific range between the low and high levels through interviews and site observations, need to be investigated in two cases: 1) conducting in an inbound range and 2) conducting in an outbound range.

Table 6 presents the input data and the outputs and comparisons between the regression outputs using either inbound or outbound data and simulation outputs using the same input data. The comparison rates shown in Table 6 represent the percentage rate of regression output to simulation output.

Table 6: Test Results of Data Range in a Regression Model

Data Range	A	B	C	D	Reg. (cycles /hr)	Sim. (cycles /hr)	Comp rate (%)
In-bound data	0.25	7	2	12	11.51	11.30	101.85
Out-bound data	0.5	15	5	3	25.77	18.54	138.99

Accordingly, the experimental design and the data range test in the regression model provide several guidelines for input data generation using simulation models.

Several main factors significantly influence productivity: surveying time, the probability of surveying, and the number of resources. Low and high levels for each dataset can be determined by analysis of the datasets collected from jobsites. The range of low and high levels can be determined by the actual value of data collected and the mean value of distribution of all datasets collected from six construction sites. Each number of datasets generated based on one actual dataset must be the same in order to achieve proper applications where all datasets are evenly reflected.

5.2.3 Input Data Generation Based on Simulation Models

To determine the low and high levels of the probability of surveying and surveying time, the data distributions were investigated in order to find the mean value of each system. Figures 2 and 3 show the best fitted distributions, using the BestFit program, of surveying time and the probability of surveying.

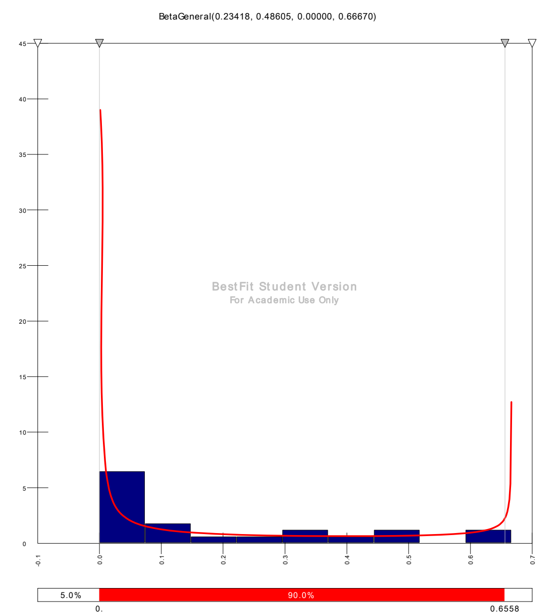

Figure 2: Beta Distribution of Datasets for the Probability of Surveying

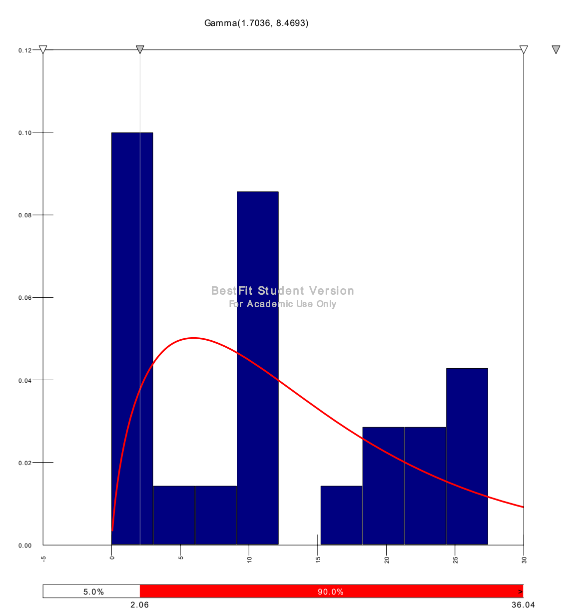

Figure 3: Gamma Distribution of Datasets for Surveying Time

The number of resources associated with simulation models is within the range of availability on jobsites. This information was determined through interviews with site personnel. If a two-link system was used, the numbers of excavators and trucks were subject to change from low level to high level. In the cases where a three-link system was used, the numbers of dozers, excavators, and trucks were subject to change from low level to high level.

Based on the guidelines provided by experimental designs, one dataset collected from actual jobsites can generate 192 datasets (i.e., combinations of $2 \times 2 \times 3 \times 16$ in cases of two-link systems or $2 \times 2 \times 2 \times 3 \times 8$ for three-link systems). Accordingly, a total of 4,416 datasets were generated based on 23 actual datasets (i.e., 23×192).

Table 7: Information of Best Fitted Distribution and Mean Values

Factor	Distribution	Mean
Probability of surveying	Beta	21.68 %
Surveying time	Gamma	14.43 min.

5.3 A New Productivity Estimation Model Based on Multiple Regression Analysis

5.3.1 Model Configurations

In accordance with the variables, denoted as factors, collected from jobsites, there were three models to be considered and examined: (1) Model I: a full model with 17 variables, (2) Model II: a reduced model with 10 variables, and (3) Model III: a reduced model with 7 variables. Model I is associated with all variables considered, including information for the probability of surveying/checking and surveying time.

All of the variables associated with model I were obtained from all the data collected from jobsites where work tasks were conducted. In order to create a productivity estimation model allowing users to predict results and to select a system, models II and III were considered with the limited information that can be found prior to actual operations. For instance, information about the probability of surveying and surveying time cannot be obtained before starting actual operations.

The reduced model can be separated into the model with sufficient information, named model II, and the model with insufficient information, named model III. The difference between models II and III involves consideration of three variables: the experience of the excavator's operator, the age of excavator, and the age of trucks. Through interviews and site observations, the variables used in each model were determined. The variables used in each of models are shown in Table 8.

Table 8: Variables Used in Each Model

Variables		I	II	III
Haul distance	A	O	O	O
Hauling speed	B	O	X	X
Bucket capacity of	C	O	O	O
Number of loading	D	O	O	O
Probability of machine break	E	O	X	X
Machine break	F	O	X	X
Prob. of surveying	G	O	X	X
Surveying time	H	O	X	X
Soil conditions	I	O	O	O
Loading duration	J	O	X	X
Travel duration	K	O	X	X
Number of trucks	L	O	O	O
Number of dozers	M	O	O	O
Number of excavators	N	O	O	O
Experience of ex-	O	O	O	X
Age of excavator	P	O	O	X
Age of trucks	Q	O	O	X
Productivity by simulation models		O	O	O

5.3.2 A Multiple Regression Model

A multiple regression model provides estimations of specific results, demonstrating the relationship between a response variable, which is the productivity of each dataset in this study, and the explanatory variables, which are the factors affecting productivity (i.e., travel times, loading times, and hauling distance).

In order to achieve the best-fitted regression model, three steps were conducted in this research. These are (1) step regression, (2) transformations, and (3) ridge regression.

Table 9 shows the finalized multiple regression models I, II, and III of each model through three steps mentioned above.

They present mathematical relationships between the explanatory variables, denoted as predictors, and a response variable. These mathematical relationships allow the user to estimate productivity when input data which reflect actual situations are provided prior to actual commencement site work.

Table 9: Variables and Coefficients of Each of Multiple Regression Models

Model	Variables and Coefficients				
I	G	L	AI	AL	BG
	1.2702	0.1018	-0.0729	0.0081	-0.0646
	BI	BM	CK	CL	CM
	0.0443	0.0260	-0.0045	0.0185	-0.5733
	DH	DI	EF	FN	GG
	-0.0042	-0.0252	-0.1777	0.0028	1.5072
	GH	GI	GK	GL	HH
	-0.1593	-0.5425	0.0420	-0.1206	0.0003
	HJ	HK	HL	HO	HQ
	0.0051	0.0007	-0.0014	-0.0013	0.0080
	IL	IM	JK	JP	LL
	0.0088	-0.4720	-0.0087	-0.0247	-0.0075
	LM	LN	LO	JM	INT.
	0.0087	0.0017	0.0642	-0.0776	2.0584
II	L	AC	AI	AL	CL
	0.0179	-0.0049	0.0272	-0.0018	-0.0138
	DL	DP	IO	LL	LM
	-0.0013	-0.0055	-0.0195	0.0013	-0.0014
	LN	LO	MM	MO	INT.
	-0.0003	0.0016	0.0156	-0.0230	0.6912
III	N	AD	AI	AL	AM
	-0.0028	-0.0004	0.0314	-0.0018	-0.0137
	CC	CL	DD	DI	DL
	-0.0078	-0.0042	-0.0044	-0.0430	-0.0009
	DM	IL	LL	LM	INT.
	0.0018	-0.0043	0.0013	-0.0020	0.6984

6 VALIDATION OF NEW PRODUCTIVITY ESTIMATION MODEL BASED ON MULTIPLE REGRESSION ANALYSIS

The best-fitted models were determined through the procedures based on these statistical criteria in previous sections in this research. In this section, comparison of the results by between the simulation models and the multiple regression models is conducted. As previously mentioned, the reliability of regression models is determined by two criteria: R-square and MSE; however, the comparison of the results between the simulation models and the regression models demonstrate how close the results of the regression models are to the simulation model. The comparison rates shown in Table 10 represent the percentage rate of regression output to simulation output.

According to Table 10, the factors included in model I and excluded in models II and III (i.e., the probability of surveying, surveying durations, the probability of machine breaking, and repair durations of machine) functioned as significant factors impacting the results. However, the factors included in model II and excluded in a model III (i.e., age of equipment and experience of operators) appear not to impact the estimation results significantly. Based on the results comparisons, the interruption by surveying for re-staking or checking stakes during operations influenced productivity significantly.

Table 10: Comparison of Productivity between Simulation Models and Multiple Regression Models

Data sets	Sim. Prod	I		II		III	
		Reg. Prod	Comp. Rate (%)	Reg. Prod	Comp. Rate (%)	Reg. Prod	Comp. Rate (%)
1	19.10	18.60	97.38	13.71	71.78	13.52	70.79
2	14.22	13.84	97.33	13.71	96.41	13.52	95.08
3	26.00	22.32	85.85	13.71	52.73	13.52	52.00
4	16.13	15.84	98.20	13.71	85.00	13.52	83.82
5	19.81	17.85	90.11	13.71	69.21	13.52	68.25
6	5.08	4.63	91.14	3.85	75.79	3.81	75.00
7	2.62	2.51	95.80	3.85	146.95	3.81	145.42
8	3.40	3.18	93.53	3.85	113.24	3.81	112.06
9	4.17	3.62	86.81	3.27	78.42	3.22	77.22
10	3.74	3.63	97.06	3.27	87.43	3.22	86.10
11	17.48	16.63	95.14	12.72	72.77	12.83	73.40
12	8.38	9.46	112.89	12.72	151.79	12.83	153.10
13	15.52	15.01	96.71	12.72	81.96	12.83	82.67
14	18.56	17.50	94.29	12.72	68.53	12.83	69.13
15	16.12	14.55	90.26	12.72	78.91	12.83	79.59
16	4.57	4.59	100.44	3.91	85.56	3.93	86.00
17	3.37	3.23	95.85	3.91	116.02	3.93	116.62
18	3.88	3.62	93.30	3.91	100.77	3.93	101.29
19	16.14	15.77	97.71	11.84	73.36	11.79	73.05
20	15.94	15.45	96.93	11.84	74.28	11.79	73.96
21	16.78	16.25	96.84	11.84	70.56	11.79	70.26
22	12.75	11.82	92.71	11.84	92.86	11.79	92.47
23	15.48	14.52	93.80	11.84	76.49	11.79	76.16
Avg.			95.22		87.86		87.54
Std.			5.30		24.22		24.32

7 CONCLUSIONS

This study provides a methodology to establish a productivity estimation model combining actual data collection, input data generation using experimental designs and multiple regression analysis.

The first issue in a productivity estimation model is how to design the model with configurations of input data and output data. The model enables the user to estimate productivity prior to site works. Thus, the input of application models must be designed to record all available information, such as the number of resources, working conditions, soil conditions, and etc. The precision and accuracy of how well input data reflect actual situation in construction sites determine the reliability of the model enabling to estimate productivity.

The second issue is how to collect data and establish a database based on the actual data collected from various jobsites. This issue is related to the difficulties in obtaining constant results from the actual earthmoving operation. As described in previous sections, input data were collected through the videotape recording in conjunction with the stopwatch study and interviews. However, the productivity resulting from each actual data has several limitations: (1) the low number of input data, (2) the large variance, and (3) the difficulties of acquiring of actual data. This suggests the use of a simulation methodology as an alternative to resolve the limitations of actual data. This research demonstrates the use of WebCYCLONE as one of simulation programs in order to generate datasets. A large number of datasets was generated by WebCYCLONE using sensitivity analysis, and the datasets was used as constant, precise, and abundant resources that provide input datasets to a multiple regression model.

The proposed productivity estimation model eventually provides the mathematical relations between conditions denoted as variables and productivity denoted as a predictor. It contributes in industry that planners or site personnel who are struggling with the limitations of deterministic and simulation methods enable to estimate productivity with simply entering input data reflecting actual site conditions. In addition, the represented model presented in this research is currently confined to earthmoving operations; however, it presents a framework available for application to other operations if the target operations provide reliable input data. The research methodology utilized in this research would be beneficial for the user to establish other application models for estimating or evaluating the performances of new technologies that are being newly applied in construction.

REFERENCES

AbouRizk, S. M. and Halpin D. W. (1992). Statistical properties of construction duration data. *Journal of Construction Engineering and Management*, ASCE, 118(3), pp525-544.

Devore, J. L. (2000). *Probability and statistics (5th Edition).* Duxbury, Pacific Grove, CA, U.S.A.

Halpin, D. W. (1973). An investigation of the use of simulation networks for modeling construction operations. *Ph.D. Dissertation of University of Illinois at Urbana-Champaign,* IL, U.S.A.

Halpin, D. W. and Riggs, L. S. (1992). *Planning and analysis of construction operations*. John Wiley & Sons, Inc., New York, NY, U.S.A.

Halpin, D. W., Jen, H., and Kim, J. (2003). A construction process simulation web service. *Proceedings of the 2003 Winter Simulation Conference*, S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds., 1503 – 1509. Institute of Electrical and Electronics Engineers, Piscataway, NJ, U.S.A.

Kannan, G. (1999). A methodology for the development of a production experience database for earthmoving operations using automated data collection. *Ph. D. Dissertation of Civil Engineering in Virginia Polytechnic Institute and State University* at Blacksburg, VA. U.S.A.

Kelton, W. D., and Barton, R. R. (2003). Experimental design for simulation. *Proceedings of the 2003 Winter Simulation Conference*, S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds., 59 – 65. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Law, A. M., and Kelton, W. D. (2000). *Simulation modeling and analysis*. McGraw Hill, New York, NY, U.S.A.

Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996). *Applied linear statistical models (4th Edition).* WCB/McGraw Hill, Boston, MA, U.S.A.

Wang, S., and Halpin, D. W. (2004). Simulation experiment for improving construction processes. *Proceedings of the 2004 Winter Simulation Conference,* R.G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds., 1252-1259. Institute of Electrical and Electronics Engineers, Piscataway, NJ, U.S.A.

AUTHOR BIOGRAPHIES

SEUNGWOO HAN is Assistant Professor in Construction Management Program at Georgia Southern University in Statesboro, Georgia. He received a B.S.E. and M.S.E. degrees from Korea University at Seoul, South Korea in 1994 and 1996, respectively, and M.S.E. and Ph.D. Degrees in Civil Engineering from Purdue University in 2002 and 2005, respectively. His email address is <shan3purdue@yahoo.co.kr>.

DANIEL W. HALPIN is Professor and Head of the Division of Construction Engineering and Management at Purdue University in West Lafayette, Indiana. He received a B.S. degree from the U.S. Military Academy at West Point in 1961. He received M.S. and Ph.D. degrees in Civil Engineering from the University of Illinois in Champaign in 1969 and 1973 respectively. He was on the faculty at Georgia Tech (1973-1985), and held the A. J. Clark Chair Professor Position at University of Maryland (1985-1987). He developed the CYCLONE methodology for construction simulation. He has received many awards, including the Huber and Peurifoy Awards, for his contributions to computer applications in civil and construction engineering. His email address is <halpin@ecn.purdue.edu>.